

Trabajo de Fin de Máster
Máster Universitario en Ingeniería Industrial

Análisis de Datos y Toma de Decisiones Utilizando
una Herramienta de Business Intelligence: Tableau.

Autor: Amalia García-Vellido Santías.

Tutor: Vicente González-Prida Díaz

Dpto. Organización Industrial y Gestión de Empresas I
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2020



Trabajo de Fin de Máster
Máster Universitario en Ingeniería Industrial.

Análisis de Datos y Toma de Decisiones Utilizando una Herramienta de Business Intelligence: Tableau.

Autor:

Amalia García-Vellido Santías

Tutor:

Vicente González-Prida Díaz

Profesor Interino

Dpto. de Organización Industrial y Gestión de Empresas I

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2020

Trabajo de Fin de Máster: Análisis de Datos y Toma de Decisiones Utilizando una Herramienta de Business Intelligence: Tableau.

Autor: Amalia García-Vellido Santías

Tutor: Vicente González-Prida Díaz

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2020

El secretario del Tribunal

A mi familia por su apoyo incondicional durante estos años de estudio

A mis profesores, por darme las herramientas necesarias para culminar con éxito esta etapa.

Resumen

En la actualidad, los datos manejan el mundo, estos se usan en todos los sectores y en cualquier tipo de empresa. El principal objetivo: mejorar el rendimiento y la estrategia de las compañías.

Dentro del sector Retail, los supermercados manejan una gran cantidad de datos. El hecho de conocer lo que éstos esconden, es un punto de partida interesante para desarrollar ventajas competitivas que permitan mantener y aumentar la base de clientes y, más importante aún, fidelizar al comprador. Por ese motivo, en este Trabajo de Fin de Máster se utiliza una herramienta de Business Intelligence para poner en práctica la capacidad de transformar datos en conocimiento, obtener conclusiones y descubrir, en definitiva, qué se esconde tras ellos siendo capaces de tomar decisiones a partir de los datos estudiados.

En este documento se presentan, en primer lugar, los objetivos, la estructura del trabajo y una introducción en la que se pone en contexto el mundo del Business Intelligence y los principales softwares disponibles en este campo. A continuación, se valida la herramienta con una aplicación a la Matriz BCG y, finalmente, se realiza un estudio completo, con una base de datos real, en el que se analizan diversos aspectos del negocio, dando una visión completa a la materia mediante su aplicación al caso de la base de datos de un supermercado.

Por último, se exponen las conclusiones extraídas tras la realización de este trabajo, y se incluye un Anexo donde se explican en detalle todas las características y potencialidades de Tableau, la interfaz utilizada.

Índice

| | |
|---|--------------|
| Resumen | ix |
| Índice | xi |
| <i>Índice de Tablas</i> | <i>xiii</i> |
| <i>Índice de Figuras</i> | <i>xv</i> |
| Prólogo. | xviii |
| Objetivo y estructura del trabajo. | 1 |
| 1 La inteligencia de negocios. | 3 |
| 1.1 Principales Softwares de BI | 4 |
| 1.2 Razones para la elección de Tableau. | 9 |
| 2 Análisis de datos con Tableau. | 11 |
| 2.1. El ciclo del análisis | 11 |
| 2.2. Conexión a una base de datos | 12 |
| 2.2.1 Unir tablas y combinar datos | 15 |
| 2.2 Dashboards | 17 |
| 2.2.1 La interfaz de los Dashboards. | 17 |
| 2.2.2 Cuadros de Mando en Dashboards. | 23 |
| 3 Aplicación de la herramienta a un proceso de gestión estratégica | 25 |
| 3.1 La Matriz BCG | 26 |
| 3.2 La matriz BCG utilizando Tableau. | 28 |
| 4 Caso de Estudio: Breakfast at the Frat | 31 |
| 4.1 Justificación elección base de datos. | 31 |
| 4.2 Escenario de estudio y limpieza de datos. | 32 |
| 4.3 Conexión a la base de datos. | 41 |
| 4.4 Análisis de los Datos. | 45 |
| 4.4.1 Análisis de las unidades vendidas por categorías y marcas. | 45 |
| 4.4.2 Análisis del rango de precios. | 51 |
| 4.4.3 Análisis precios y descuentos. | 52 |
| 4.4.4 Análisis de las promociones | 55 |
| 4.4.5 Análisis respuesta consumidor a descuentos en precio | 57 |
| 4.4.6 Análisis de los Supermercados | 58 |
| 4.4.7 Análisis en función de las fechas de compra. Análisis estacional. | 61 |
| 4.4.8 Análisis geográfico. | 62 |
| 5 Conclusiones | 65 |

| | | |
|----------|--|-----------|
| 6 | <i>Bibliografía</i> | 67 |
| 7 | <i>Anexo: La interfaz de Tableau.</i> | 69 |
| 7.1 | Tipos de uniones | 69 |
| 7.2 | Tipos de campos en Tableau | 69 |
| 7.2.1 | Medidas y Dimensiones | 69 |
| 7.2.2 | Campos Discretos y Continuos. | 70 |
| 7.3 | Tipos de Gráficos | 71 |
| 7.4 | Filtros de datos | 75 |
| 7.4.1 | Filtros para campos discretos | 75 |
| 7.4.2 | Filtros para campos continuos | 76 |
| 7.5 | Visualizaciones Avanzadas | 76 |
| 7.5.1 | Comparar valores | 76 |
| 7.5.2 | Fechas y horas | 80 |
| 7.5.3 | Relacionar una parte con el total | 83 |
| 7.5.4 | Visualizar distribuciones | 86 |
| 7.5.5 | Ejes múltiples | 88 |
| 7.6 | Cálculos en Tableau | 89 |
| 7.6.1 | Campos Calculados. | 90 |
| 7.6.2 | Cálculos de Tabla. | 95 |

Índice de Tablas

| | |
|---|----|
| Tabla 1: Base de datos Superstore. | 13 |
| Tabla 2: Base de datos en Excel con dos hojas. | 15 |
| Tabla 3: Tipos de producto en la Matriz BCG | 27 |
| Tabla 4: Base de datos para matriz BCG. | 28 |
| Tabla 5: Base de datos. Hoja 1 Glosario. | 33 |
| Tabla 6: Base de datos. Hoja 2 | 33 |
| Tabla 7: Base de datos. Hoja 3 | 34 |
| Tabla 8: Base de datos. Hoja 4 | 34 |
| Tabla 9: Base de datos tras la limpieza. | 35 |
| Tabla 10: Descripción de variables. | 36 |
| Tabla 11: Diferencia conceptos UNITS,VISITS y HHS. | 37 |
| Tabla 12: Hoja PRODUCTS tras limpieza. | 38 |
| Tabla 13: Categoría y Subcategoría. | 38 |
| Tabla 14: Hoja STORES base de datos. | 39 |
| Tabla 15. Desglose de la hoja Excel de productos. | 40 |
| Tabla 16: MSA Code como campo numérico. | 42 |
| Tabla 17: Cambio a función geográfica. | 42 |
| Tabla 18: Campos tipo numérico. | 43 |
| Tabla 19: Campos calculados booleanos. | 43 |
| Tabla 20: Marca Blanca en categorías. | 50 |
| Tabla 21: Porcentaje ventas de la marca blanca en cada categoría. | 50 |
| Tabla 22: Rango de Precios por Categoría | 51 |
| Tabla 23: Listado número de tiendas en cada ciudad. | 63 |
| Tabla 24: Ciudades con más Supermercados | 64 |
| Tabla 25: Ciudades con más Facturación. | 64 |

Índice de Figuras

| | |
|---|-----|
| Figura 1: Curso Análisis de datos Tableau. (udemy, s.f.) | xix |
| Figura 2: Certificación Oficial Tableau | xix |
| Figura 3: Cuadrante de Gartner B.I. (Gartner, 2019) | 4 |
| Figura 4: Evolución Power BI. | 5 |
| Figura 5: Evolución Tableau. | 6 |
| Figura 6: Evolución Qlik. | 7 |
| Figura 7: Google trends. Data Visualization. | 8 |
| Figura 8: Google trends. Inteligencia Empresarial. | 8 |
| Figura 9: Comparación experiencia usuario. (Insights, s.f.) | 9 |
| Figura 10: El ciclo del análisis. | 11 |
| Figura 11: Webs con ejemplos de bases de datos. | 12 |
| Figura 12: Conexión a la base de datos en Tableau. | 13 |
| Figura 13: Hoja de trabajo en Tableau. | 14 |
| Figura 14: Unión de base de datos. | 15 |
| Figura 15: Combinación de bases de datos. | 16 |
| Figura 16: La interfaz de los Dashboards. | 18 |
| Figura 17: Beneficios por categoría. | 18 |
| Figura 18: Top 10 elementos menos rentables. | 19 |
| Figura 19: Beneficio por Estado. | 19 |
| Figura 20: Pronóstico. | 19 |
| Figura 21: Análisis del beneficio por producto. | 20 |
| Figura 22: Acciones en Dashboards I. | 21 |
| Figura 23: Acciones de Dashboard II. | 21 |
| Figura 24: Dashboard. | 22 |
| Figura 25: Filtro para Dashboard. | 22 |
| Figura 26: Filtro para Dashboard II. | 22 |
| Figura 27: Filtro para Dashboard III. | 23 |
| Figura 28: Dashboard de análisis del beneficio elementos menos rentables. | 23 |
| Figura 29: Dashboard con KPI 15%. | 24 |
| Figura 30: Dashboard con KPI 11%. | 24 |
| Figura 31: Ejemplo matriz BCG (Mañez, s.f.). | 26 |
| Figura 32: Matriz BCG | 29 |

| | |
|--|----|
| Figura 33: Dashboard matriz BCG | 29 |
| Figura 34: Justificación elección Dataset | 31 |
| Figura 35: Carpeta con la base de datos y pdf. | 32 |
| Figura 36: Unión de las tres tablas de datos. | 32 |
| Figura 37: Conexión a la base de datos. Parte 2. | 41 |
| Figura 38: Conexión a la base de datos. Parte 2. | 41 |
| Figura 39: Clasificación de los descuentos. | 44 |
| Figura 40: Asignación paleta colores. | 45 |
| Figura 41: Distribución de unidades vendidas en las categorías. | 46 |
| Figura 42: Unidades vendidas por subcategoría. | 47 |
| Figura 43: Unidades vendidas según marca y categoría. | 47 |
| Figura 44: Dashboard. Análisis categorías y marcas I | 48 |
| Figura 45: Dashboard Análisis categorías y marcas II | 48 |
| Figura 46: Dashboard Análisis categorías y marcas III | 49 |
| Figura 47: Dashboard Análisis categorías y marcas IV | 49 |
| Figura 48: Desempeño marca blanca en las categorías. | 50 |
| Figura 49: Rango precios entre categorías. | 51 |
| Figura 50: Unidades vendidas según el rango de descuento. | 52 |
| Figura 51: Promedio unidades vendidas según rango de descuento. | 52 |
| Figura 52: unidades vendidas según descuento y categoría. | 53 |
| Figura 53: Dashboard promedio unidades vendidas según rango de descuento. | 54 |
| Figura 54: Dashboard filtrado para productos sin descuento. | 54 |
| Figura 55: Dashboard filtrado para productos con más de un 50% de descuento. | 55 |
| Figura 56: Efectos combinados de los tres efectos. | 56 |
| Figura 57: Dashboard efecto TPR, DISPLAY, FEATURE. | 56 |
| Figura 58: Efecto Descuento. | 57 |
| Figura 59: Efecto Descuento en cada categoría. | 57 |
| Figura 60: Filtro por campo en un gráfico. | 58 |
| Figura 61: Top 10 Ítems vendidos por tienda | 58 |
| Figura 62: Análisis por ítem: Tendencia y pronóstico I. | 59 |
| Figura 63: Análisis por ítem: Tendencia y pronóstico II. | 59 |
| Figura 64: Interacción con el Dashboard. | 60 |
| Figura 65: Análisis ítems menos vendidos, tendencia y pronóstico. | 60 |
| Figura 66: Dashboard facturación mes y categoría. | 61 |
| Figura 67: Dashboard facturación mes y categoría con filtro. | 61 |
| Figura 68: Gráfico tiendas por estado. | 62 |
| Figura 69: Top 10 ciudades mayor facturación. | 63 |

| | |
|---|----|
| Figura 70: Dashboard análisis geográfico con filtro. | 64 |
| Figura 71: Representación en Tableau de medidas y dimensiones | 70 |
| Figura 72: Gráfico de barras II | 71 |
| Figura 73: Gráfico de barras apiladas. | 72 |
| Figura 74: Gráfico de líneas en Tableau II. | 72 |
| Figura 75: Mapa relleno. | 73 |
| Figura 76: Mapa de símbolos. | 74 |
| Figura 77: Mapa de densidad. | 74 |
| Figura 78: Filtros I. | 75 |
| Figura 79: Diagrama de barras en orden descendente. | 77 |
| Figura 80: Diagrama de barras resaltado. | 77 |
| Figura 81: Base datos hospital, tiempo objetivo por especialidad. | 78 |
| Figura 82: combinación fuentes datos. | 78 |
| Figura 83: Diagrama de bala. | 79 |
| Figura 84: Variación diagrama de barras. | 80 |
| Figura 85: Fechas. | 80 |
| Figura 86: Parte de fecha. | 81 |
| Figura 87: Valor de Fecha. | 81 |
| Figura 88: Combinación campos de Fechas. | 82 |
| Figura 89: Mapa de calor. | 82 |
| Figura 90: Barras apiladas. | 83 |
| Figura 91: Mapa de árbol. | 84 |
| Figura 92: Diagrama de área. | 84 |
| Figura 93: Diagrama de área relleno. | 85 |
| Figura 94: Diagrama circular. | 85 |
| Figura 95: Diagrama con círculos. | 86 |
| Figura 96: Diagrama de caja y bigotes. | 87 |
| Figura 97: Histograma. | 87 |
| Figura 98: Diagrama de dispersión. | 88 |
| Figura 99: Eje doble. | 89 |
| Figura 100: Base de datos. | 90 |
| Figura 101: Crear un campo calculado. | 90 |
| Figura 102: Base de datos alquiler vacacional. | 91 |
| Figura 103: Cálculo a nivel de fila. | 92 |
| Figura 104: Cálculo a nivel fila en la base de datos. | 92 |
| Figura 105: Cálculo agregado. | 92 |
| Figura 106: Cálculo agregado. | 93 |

| | |
|---|----|
| Figura 107: Cálculo agregado. | 93 |
| Figura 108: Nivel fila vs agregado. | 94 |
| Figura 109: Cálculo a Nivel fila | 94 |
| Figura 110: Cálculo Agregado. | 95 |
| Figura 111: Cálculos de tabla rápidos. | 95 |
| Figura 112: Cálculos de tabla rápidos II. | 96 |
| Figura 113: tabla, panel y celda | 97 |
| Figura 114: Direccionamiento y partición. | 97 |

El Software que se utiliza en este documento despertó mi curiosidad sobre el campo del Business Intelligence (BI), ya que a través de él fui consciente tanto de la importancia que está cobrando este sector en la actualidad, como de que la cantidad de datos producidos, utilizados y almacenados crece día tras día. Esto se debe, en primer lugar, a que cada vez son más usuarios a nivel mundial los que tienen acceso a Internet desde ordenadores, tabletas o teléfonos y, en segundo lugar, a que todos los dispositivos conectados a dicha red son capaces de enviar, recibir y generar datos. Es por ello que se deben crear nuevas técnicas analíticas para que las grandes cantidades de datos generados puedan ser administrados de manera eficiente.

Por este motivo, decidí formarme en este ámbito y he realizado dos cursos online.

El primero de ellos, lo realicé con el objetivo de aprender a utilizar el software Tableau, a través de Udemy.com, una plataforma de aprendizaje en línea (udemy, s.f.), que ofrece una gran variedad de cursos de distintas temáticas.

El curso se llama: “Análisis de datos y visualizaciones con Tableau” y fue impartido por Sergio Llobet, Analista y experto en visualización de datos.



Figura 1: Curso Análisis de datos Tableau. (udemy, s.f.)

Adicionalmente, realicé un curso oficial impartido por Tableau Software: “Desktop I: Fundamentals for Tableau” 10/2018.

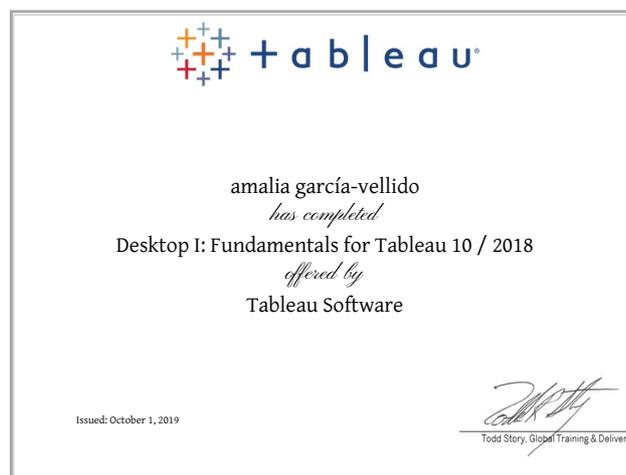


Figura 2: Certificación Oficial Tableau

En resumen, al elegir una herramienta de BI desconocida para mí, ha sido necesario realizar una formación específica para poder llevar a cabo algunos de los objetivos que se mencionan en el siguiente capítulo.

OBJETIVO Y ESTRUCTURA DEL TRABAJO.

El propósito de este Trabajo de Fin de Máster es utilizar una de las herramientas más populares en el ámbito de la Inteligencia Empresarial, con idea de validar y poner en práctica la capacidad de ésta para transformar datos, obtener conclusiones y descubrir qué se esconde tras dichos datos, de modo que se puedan tomar decisiones a partir de ellos.

Dado lo general de este propósito, la finalidad de este Trabajo se puede traducir en los siguientes **Objetivos Particulares**:

1. Analizar las herramientas de BI que hay en el mercado.
2. Justificar la selección de Tableau para el análisis de datos.
3. Revisar las potencialidades de la herramienta.
4. Desarrollar un caso práctico de validación de la herramienta para el análisis de datos y la toma de decisiones.

El documento se ha estructurado como sigue:

- En primer lugar, se sitúan el análisis de datos y la inteligencia empresarial en la sociedad, poniendo en valor la importancia de saber modelar y transmitir los descubrimientos realizados mediante visualizaciones. En esa misma sección, se realiza una breve comparación a modo de revisión de la literatura, entre los principales softwares de visualización de datos, y se exponen los motivos que han llevado a elegir Tableau frente a los otros mencionados. (**Capítulo 1: La Inteligencia de Negocios**).
- Tras poner en contexto los motivos para utilizar Tableau, se analiza cómo es el ciclo del análisis y la conexión a los datos con este software, un paso muy importante al trabajar con Tableau, poniendo el foco en los Dashboards o Cuadros de Mando, que son la herramienta de comunicación de la información y los descubrimientos realizados a partir de los datos. Es importante desarrollar cierta habilidad creando Dashboards, ya que éstos deben ser claros y limpios para que la información se transmita de la mejor forma. (**Capítulo 2: Análisis de datos con Tableau**).
- El siguiente apartado pretende señalar las potencialidades que ofrece Tableau, aplicándolo a un proceso de crecimiento estratégico, usando la matriz BCG (**Capítulo 3: Aplicación de la herramienta a un proceso de gestión estratégica**).
- Una vez se ha visto como es el software y las funcionalidades y opciones que aporta, se lleva a cabo un caso de estudio con una base de datos real de una cadena de supermercados en EE. UU. (**Capítulo 4: Caso de Estudio: Breakfast at the Frat**).
- Para finalizar este Trabajo de Fin de Máster, se exponen las conclusiones de los resultados obtenidos tras el análisis de la base de datos y posibles mejoras o línea de trabajo futuro. (**Capítulo 5: Conclusiones**).
- Se ha creído conveniente incluir, además, un apartado de Anexo que contiene todo lo relacionado con la interfaz de Tableau, los tipos de gráficas que se pueden obtener y las funcionalidades y aplicaciones para sacar el máximo partido al software. (**Anexo: La interfaz de Tableau**).

Objetivo y estructura del trabajo.

1 LA INTELIGENCIA DE NEGOCIOS.

“Todo lo que se hace se puede medir, sólo si se mide se puede controlar, sólo si se controla se puede dirigir y sólo si se dirige se puede mejorar”

- Pedro Mendoza -

La Inteligencia de negocios (o Inteligencia Empresarial) es un término que incluye las aplicaciones, la infraestructura y las herramientas, así como las prácticas recomendadas que otorgan acceso a la información y que permiten analizarla.

La inteligencia de negocios emergió junto con algunos modelos informáticos para la toma de decisiones y la planificación como una manera de transformar datos en conclusiones. Este término abarca datos, informática y análisis en el contexto de operaciones de negocios. Por tanto, es un término general que engloba procedimientos y métodos para reunir, guardar y analizar datos de actividades o procesos de negocios de manera que se optimice el rendimiento. La combinación de todo esto aporta una vista integral de la empresa y ayuda a que la toma de decisiones sea mejor y más útil.

La importancia de la inteligencia de negocios radica en que es de gran ayuda en la toma de decisiones, al mostrar datos históricos y del presente en un escenario de negocios. También puede ofrecer pruebas para comparar el rendimiento de modo que las empresas puedan así operar de forma más ágil y eficiente. Otra característica es que ayuda a la identificación de tendencias del mercado para aumentar ventas o ingresos. Si se usa de manera eficiente se puede mejorar casi todos los aspectos de una empresa.

Con la transformación digital se creó una afluencia muy grande de información, en tanto que los datos están en todas partes y en todo momento. Y más aún, enraizados en procesos de negocios en empresas de cualquier tamaño.

La inteligencia de negocios ayuda por tanto a la empresa:

- Identificando áreas o sectores donde aumentar las ganancias
- Analizando el comportamiento de clientes
- Comparando datos con la competencia
- Haciendo un seguimiento del rendimiento de la empresa
- Optimizando las operaciones de la empresa
- Identificando las tendencias en el mercado
- Identificando cualquier tipo de problema de negocios

La inteligencia de negocios funciona de la siguiente manera: Las organizaciones tienen preguntas y objetivos. En la búsqueda de respuestas y en el seguimiento de la consecución de dichos objetivos, se recopilan datos, se analizan, y se determinan las acciones que sería necesario efectuar para la evolución de la empresa.

Existen múltiples herramientas de inteligencia de negocios. Una de las formas más frecuentes de ver la inteligencia de negocios es a través de la visualización de datos, ya que para el ser humano es fácil responder a estímulos visuales o identificar un patrón o cambios de colores. Por ello, las visualizaciones presentan los datos de una forma accesible y fácil de comprender.

Los datos hacen más sencillo el respaldo de un argumento, pero para ello deben ser relevantes y comprensibles. ¿Qué hace que unos datos sean relevantes? El análisis. ¿Cómo hacer más comprensible el análisis? Con una visualización.

1.1 Principales Softwares de BI

Existen muchos softwares dentro del campo del análisis de datos, aunque Tableau, Power BI y Qlik son los líderes del mercado en cuanto a soluciones de Business Intelligence (BI) (Arribas, 2019) según los analistas tecnológicos, como se puede ver en el cuadrante mágico para plataformas de análisis e inteligencia empresarial:



Figura 3: Cuadrante de Gartner B.I. (Gartner, 2019)

En el informe elaborado por Gartner, (Gartner, 2019) se afirma que hoy en día es muy frecuente utilizar en las empresas un software de Business Intelligence. Se destacan a continuación las características de cada uno de los tres considerados como líderes del mercado:

1. **Microsoft Power BI:** Power BI ofrece, en un solo producto, preparación y descubrimiento de datos, cuadros de mando interactivos y analítica aumentada (enfoque que automatiza los conocimientos mediante el aprendizaje automático y la generación de lenguaje natural).

Fortalezas:

- Precio: si se compara con los demás competidores, su precio es más bajo. De manera que se reduce la barrera de entrada de las empresas al Business Intelligence.
- Simplicidad de uso y apariencia.
- Visión del producto: Microsoft tiene un RoadMap muy claro y lo cumple añadiendo mejores y nuevas funciones a sus plataformas.
- Experiencia del usuario: tiene una puntuación muy alta en este punto, cuenta con una fuerte comunidad de partners, distribuidores y usuarios particulares que aportan nuevas funcionalidades, visualizaciones adicionales, vídeo tutoriales y contenido adicional.

Evolución: La evolución de este producto ha sido notable, partía de una posición óptima en cuanto a facilidad de uso, y ha empleado sus esfuerzos en aportar funcionalidades, consiguiendo ser más rápida y efectiva que los otros dos en este aspecto.

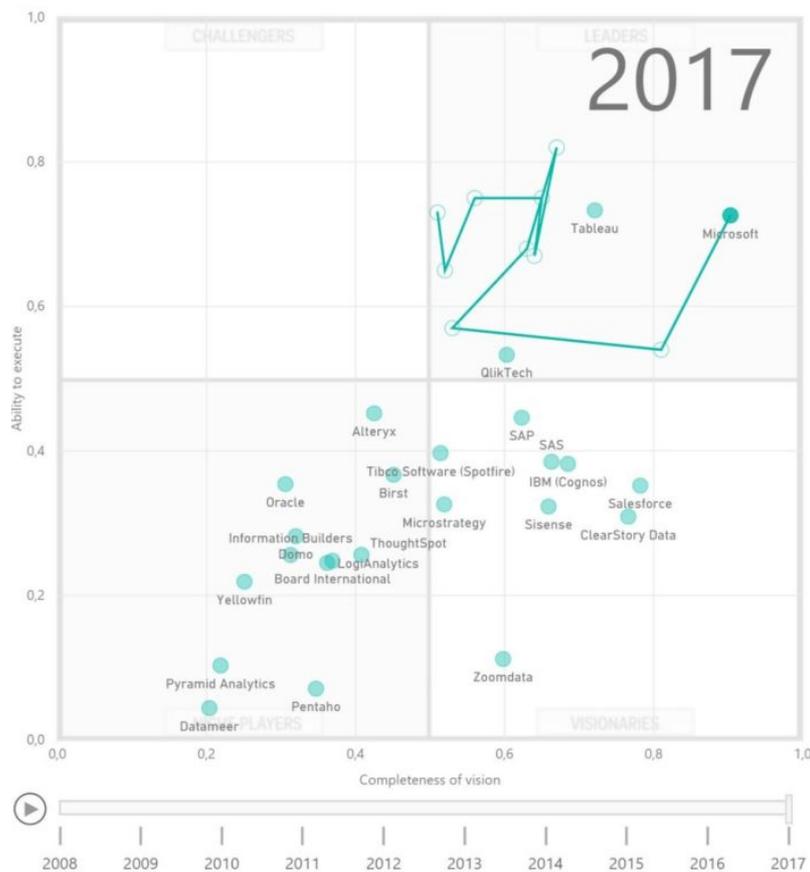


Figura 4: Evolución Power BI.

2. **Tableau:** Sirve para la visualización interactiva de los datos, con los que los usuarios pueden interactuar de varias maneras: comparando datos, filtrándolos o creando una conexión entre unas variables y otras. (Monzón, 2018) Ofrece un producto muy intuitivo que permite a los usuarios crear sus propios análisis sin un gran conocimiento de códigos o programación. Cuenta con tres productos; Tableau desktop, Tableau Server y Tableau Online.

Fortalezas:

- Es el mejor en exploración interactiva: gracias a sus funcionalidades, permite a los usuarios hacer profundas exploraciones y manipular los datos de una forma más rápida y sencilla que los demás.
- Enfoque hacia el éxito y la experiencia del usuario: está situado en el primer cuartil de soluciones BI respecto a experiencia de uso. Además, cuenta con una comunidad publica que aporta opciones de formación y el propio fabricante publica videos de capacitación en la web.
- Flexibilidad en opciones de implementación.

La evolución que ha seguido Tableau muestra un camino ideal, ha pasado de ser un producto desafiante a un producto líder por una vía muy directa, es el que presenta una mejor evolución:

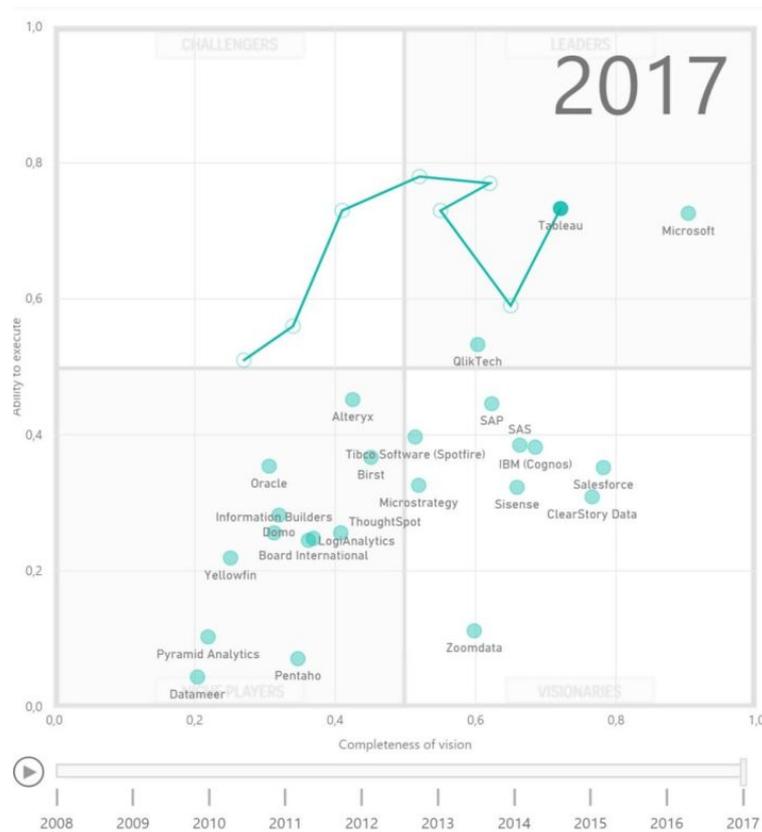


Figura 5: Evolución Tableau.

- Qlik:** Es una plataforma enfocada al análisis visual de datos y aplicaciones interactivas que tiene por objetivo mejorar el proceso de acceso a los datos de cara al usuario. Como, por ejemplo, acceder a ciertas visualizaciones ‘limpias’ y fáciles de comprender, diseños de gráficos llamativos, entre otros. (Monzón, 2018). Se ha posicionado como uno de los líderes debido a mejoras en su estrategia de marketing y la facilidad de uso de sus herramientas. Sus productos son Qlik Sense, Qlik View, o Qlik Core.

Fortalezas:

- Producto escalable: muchos clientes lo usan por su compatibilidad con un gran número de orígenes bases de datos, capacidad para modelar datos complejos y ejecutar cálculos avanzados.
- Marketing: Qlik fue reconocida como una de las empresas en el top 10 de compañías innovadoras para el bien social según la revista *Fast Company*.
- Visión de producto: cuenta con un Marketplace que permite a los desarrolladores monetizar las extensiones creadas de la plataforma.
- Partners: cuenta con una red de más de 500 integradores y 1700 partners en el mundo.

En la evolución de este producto, se caracteriza por un proceso de evolución técnica más que un avance en funcionalidad:

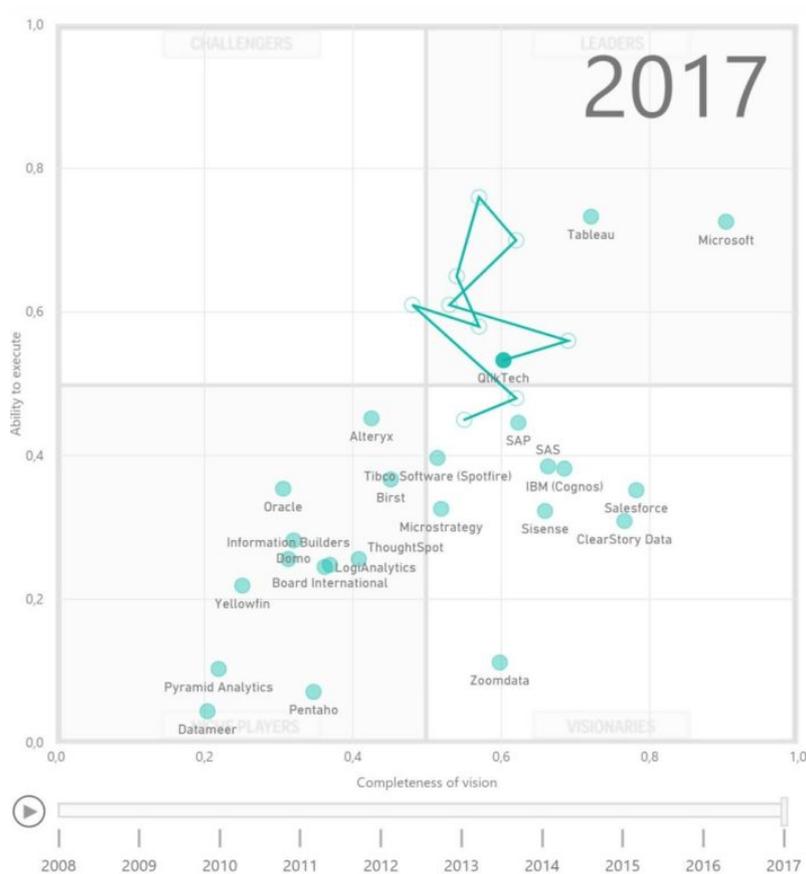


Figura 6: Evolución Qlik.

Se han buscado los términos “Data Visualization” e “Inteligencia Empresarial” en Google trends, que es una herramienta de acceso libre y gratuito brindada por Google, la cual permite comparar la popularidad de búsqueda de varias palabras o frases; de esta manera se puede conocer el nivel de búsqueda de un determinado término durante un período de tiempo determinado. (TRENDS)

Las gráficas referentes a ambos términos muestran una evolución temporal creciente en los últimos años (Figuras 7 y 8).

Además, se ha hecho una revisión de otros TFM cuyo tema principal esté relacionado con la inteligencia empresarial. No se ha encontrado ninguno que utilice Tableau, hecho que se añade a los motivos que se enumeran en el apartado 1.2 respecto a razones de la elección del software.

- El Business Intelligence en las pymes. Herramienta Power BI. (Barrionuevo)
- Análisis del diseño de visualización interactiva de información. (Attardi, 2016)
- Análisis de los sistemas de ayuda a la toma de decisiones y su aplicación a los procesos empresariales (Ruedas, 2015)
- La utilización de herramientas Business Intelligence para la gestión empresarial. (Jordá, 2016)
- Implantación de un sistema de Business Intelligence en empresa de cursos a distancia. (Martínez, 2018)

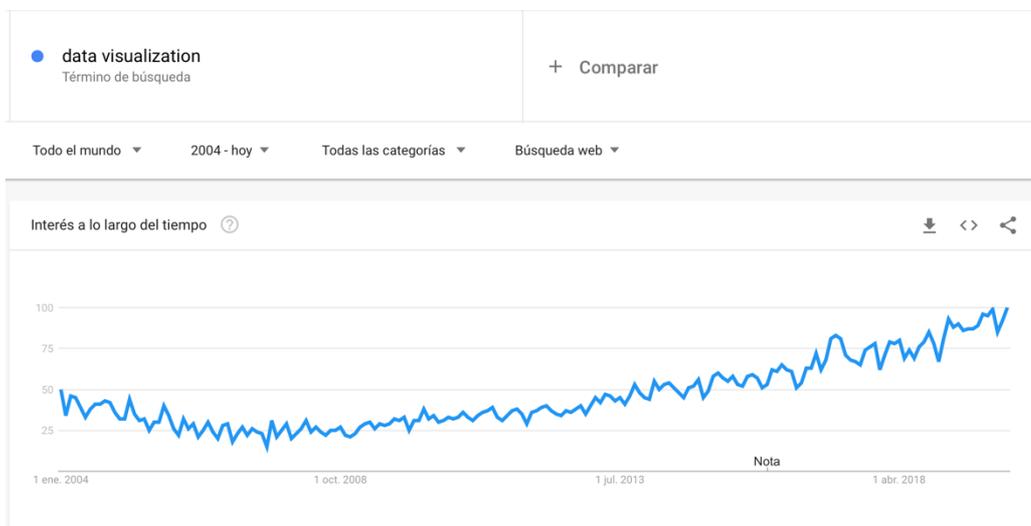


Figura 7: Google trends. Data Visualization.

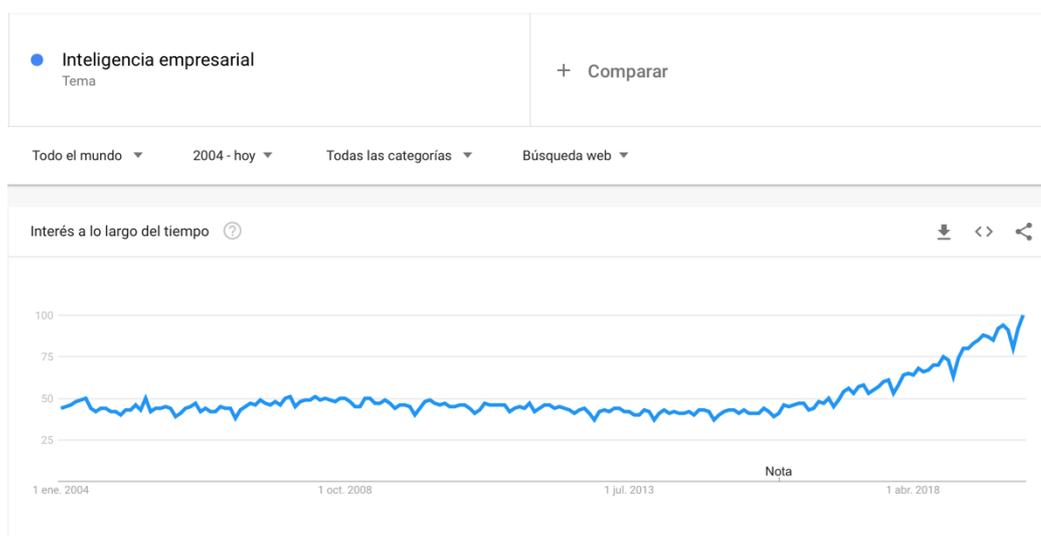


Figura 8: Google trends. Inteligencia Empresarial.

1.2 Razones para la elección de Tableau.

En un informe publicado por Gartner Peer Insights, se ve que, a pesar de las pequeñas diferencias, las 3 soluciones están actualmente muy igualadas en valoración (Insights, s.f.):

| Microsoft + Show Products (3) | Qlik + Show Products (3) | Tableau + Show Products (4) |
|----------------------------------|-----------------------------|--------------------------------|
| 4.4 ★★★★★ (1073 reviews) | 4.4 ★★★★★ (527 reviews) | 4.4 ★★★★★ (737 reviews) |
| 87% Yes 👍 1073 Reviewer(s) | 81% Yes 👍 527 Reviewer(s) | 83% Yes 👍 737 Reviewer(s) |

Figura 9: Comparación experiencia usuario. (Insights, s.f.)

Como se puede ver en la Figura 9, cualquiera de las tres sería una buena solución, pero se va a justificar la elección de Tableau frente a las otras dos por los siguientes motivos:

- Perspectiva de futuro: Existen actualizaciones y mejoras continuas del producto y cada vez más popular y usado en empresas de cualquier tipo.
- Precio: aunque el precio es bastante elevado, ofrece una licencia de un año para todos sus productos, si eres estudiante y te registras con la cuenta de la universidad
- Recursos disponibles: En su pagina web publican un gran número de informes, con novedades, aplicaciones de este software, además de videos de capacitación, para aprender a utilizar y sacar el mayor provecho a esta herramienta
- Certificación: También por ser estudiante se puede realizar un curso online gratuito y obtener una certificación emitida por tableau.

La inteligencia de negocios.

2 ANÁLISIS DE DATOS CON TABLEAU.

Cuando se ve por primera vez una base de datos, lo que se aprecia es una gran cantidad de datos sin procesar; números, fechas, texto etc. Hay mucha historia e ideas detrás de esos datos, así como decisiones y acciones que se pueden llevar a cabo, pero para ello hay que encontrar la manera de entender el significado y ser capaz de ver a través de ellos. Tableau es una plataforma óptima para ver, comprender y tomar decisiones basadas en los datos.

Utilizando Tableau, se puede descubrir, analizar y narrar la información tras los datos de una manera muy visual, ya que la interfaz está diseñada para un flujo natural y continuo tanto de pensamiento como de trabajo. En el entorno de Tableau, los ítems que se arrastran y sueltan en la pantalla de trabajo se traducen en representaciones gráficas muy potentes, donde se puede experimentar de qué manera tiene más sentido mostrar los datos.

Este software es flexible y capaz de ayudar a cada persona a resolver preguntas de una hoja de cálculo, que no pueden responderse con las funcionalidades de Excel. Tableau ayuda a responder preguntas con datos, y pueden utilizarse tanto para ayudar a incrementar las ventas en una compañía, para mejorar el servicio al cliente, gestionar inversiones, conseguir calidad y seguridad, e incluso abordar problemas medioambientales. Es una herramienta útil en cualquier tipo de empresa, entendiendo como empresa desde una multinacional, un banco, hospital, pyme... cualquier lugar donde sea necesario ver y entender datos.

Hoy en día, muchas grandes empresas utilizan Tableau porque permite crear informes con relativa facilidad y también en empresas pequeñas ya que es una forma de traducir datos en información necesaria con un bajo coste. El mundo de los datos está cambiando rápidamente. Se puede hablar de machine learning o Inteligencia artificial (A.I) pero siempre se necesitará la capacidad de comunicar ideas. Solo un ser humano puede hacer esto de manera efectiva.

2.1. El ciclo del análisis

Cualquier persona que trabaje con datos y con su entendimiento, se encontrará trabajando dentro del ciclo de análisis, el ciclo del análisis tiene en cuenta las etapas asociadas al acceso de datos descubrimiento, exploración y uso compartido de la información, este ciclo se podría ilustrar de la siguiente manera:

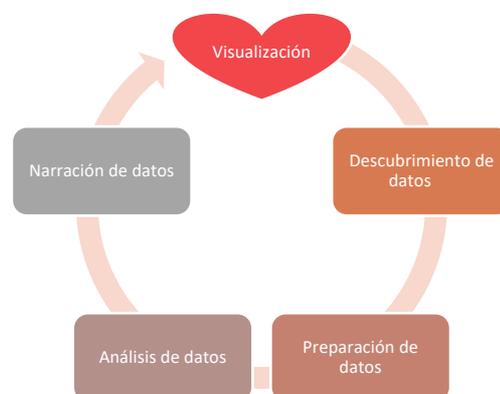


Figura 10: El ciclo del análisis.

Tableau permite saltar a cualquier paso del ciclo, moverse libremente entre los pasos e iterar a través del ciclo muy rápidamente. En particular, Tableau tiene la capacidad para:

- **Descubrimiento de datos:** Es posible explorar fácilmente un conjunto de datos con y editar los datos para su mejor comprensión.
- **Preparación de datos:** Tableau permite conectarse a datos de muchas fuentes diferentes y, si es necesario, crear una estructura que funcione mejor para su análisis. Cuenta con las herramientas para reunir datos complejos y desordenados de múltiples fuentes.
- **Análisis de datos:** Las visualizaciones de datos ayudan a que se puedan ver y comprender tendencias, valores atípicos y relaciones. Además de esto, Tableau tiene un conjunto cada vez mayor de funciones analíticas que permiten profundizar en la comprensión de relaciones complejas, patrones y correlaciones en los datos.
- **Narración de datos:** Tableau permite crear cuadros de mando e historias totalmente interactivos con las visualizaciones e ideas para que se pueda compartir la historia de datos con otros.

La visualización de datos es el corazón de Tableau, permite extraer significado de la información. Es una plataforma compatible con el ciclo de análisis, ofreciendo una representación visual y ayudando a responder preguntas, independientemente de su complejidad.

2.2. Conexión a una base de datos

Tableau se conecta a los datos almacenados en una amplia variedad de archivos y bases de datos. Esto incluye archivos planos, como documentos de Excel, archivos espaciales y archivos de texto; bases de datos relacionales, como SQL Server y Oracle; fuentes de datos basadas en la nube, como Google Analytics y Amazon Redshift; y fuentes de datos OLAP, como Microsoft Analysis Services. Con muy pocas excepciones, el proceso de análisis y creación de visualizaciones será el mismo, sin importar la fuente de datos que utilice.

En esta sección se verá como es el proceso de conexión un archivo de texto, específicamente, un archivo de valores separados por comas (.csv). En internet se pueden encontrar muchas páginas de fuentes de datos tanto económicos, geográficos, o incluso de algunas empresas como Spotify o Airbnb.

The screenshot shows the 'data.world' interface. At the top, there's a search bar and navigation icons. The main content area displays a dataset titled 'Superstore Sales EU' by LORNA EDEN. Below the title is a table with 5 rows and 8 columns. The table contains sales data for different regions and dates. Below the table, there are navigation options like 'Showing 1-5 of 10,000 rows' and 'Switch to column overview'. At the bottom, there are 'Related datasets' including '2019 Superstore', 'Modified Superstore with HI and AK', and 'Conferences'.

| id | rowid | store | date | time | class | product | region |
|----|-------|-----------------|------------|------------|----------|---------|----------|
| 1 | 1 | ES-2018-1311038 | 2018-02-07 | 2018-02-11 | Standard | Class | AS-10045 |
| 2 | 2 | ES-2018-1311038 | 2018-02-07 | 2018-02-11 | Standard | Class | AS-10045 |
| 3 | 3 | ES-2018-1311038 | 2018-02-07 | 2018-02-11 | Standard | Class | AS-10045 |
| 4 | 4 | ES-2018-1311038 | 2018-02-07 | 2018-02-11 | Standard | Class | AS-10045 |
| 5 | 5 | ES-2018-1311038 | 2018-02-07 | 2018-02-11 | Standard | Class | AS-10045 |

Figura 11: Webs con ejemplos de bases de datos.

Para explicar el proceso de conexión a base de datos se utilizará el archivo de Excel *Superstore* con datos de una tienda ficticia como se ve en la Tabla 1:

Tabla 1: Base de datos Superstore.

Para conectarse a una base de datos hay que seguir los siguientes pasos:

1. Abrir Tableau. En la pantalla de inicio se ve una lista de opciones de conexión a la izquierda y, si corresponde, vistas previas en miniatura de libros editados recientemente en el centro, junto con libros de muestra en la parte inferior.
2. Haciendo clic en **conectar y a un archivo**, se selecciona **archivo Excel**.
3. En el cuadro de diálogo **Abrir**, se selecciona el archivo que se desee.
4. A continuación, se verá la pantalla de conexión de datos, que permite crear visualmente conexiones a fuentes de datos, como se puede ver en la Figura 12, al incluir una base de datos Tableau de una vista previa del archivo para la conexión:

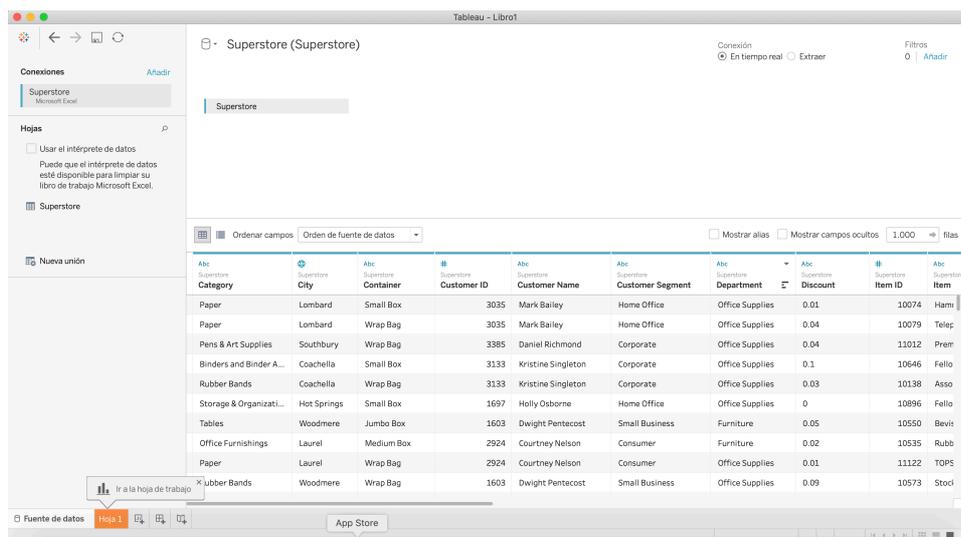


Figura 12: Conexión a la base de datos en Tableau.

Análisis de datos con Tableau.

Una vez realizada la conexión y, tras hacer clic en **Ir a la hoja de trabajo**, se ve en pantalla lo que aparece en la Figura 13, que tiene los principales elementos:

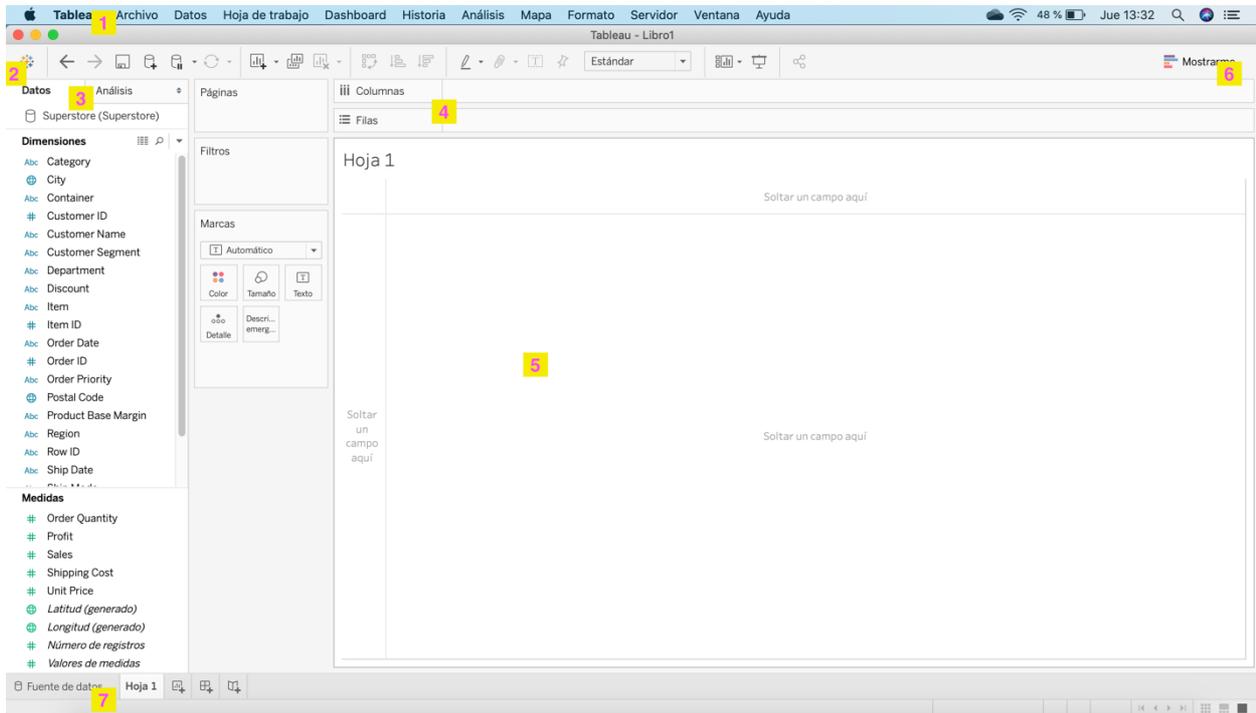


Figura 13: Hoja de trabajo en Tableau.

1. El menú contiene las pestañas de menú de tableau, que permiten realizar una gran variedad de funciones.
2. La barra de herramientas tiene las opciones típicas de guardar, rehacer, deshacer, guardar, añadir una fuente de datos etc.
3. En la barra lateral, hay dos pestañas: Datos y Análisis. Cuando está activa la pestaña datos, la barra lateral es el “panel de datos”, que contiene la base de datos y una serie de campos divididos entre Dimensiones y Medidas.
4. Existen estantes como Columns, Filas, Páginas, Filtros, Marcas, que son áreas donde se arrastran y sueltan los campos del panel de datos, dentro de las Marcas, hay estantes adicionales como Color, Tamaño, Texto, Detalle y Descripción emergente. La visualización se creará en función de estos campos.
5. La visualización se creará en esta zona de la pantalla.
6. La pestaña mostrarme permite crear visualizaciones rápidas en función de los campos que estén activos.
7. Las pestañas de la parte inferior de la pantalla dan opciones para editar la fuente de datos o navegar entre hojas o Dashboards.

Una hoja contiene una sola visualización, como podría ser un gráfico de barras o de líneas.

Un Dashboard es una presentación de varias visualizaciones relacionadas y objetos como texto o imágenes, organizadas de manera que comuniquen un mensaje o información.

2.1.1 Unir tablas y combinar datos

Tableau es capaz de trabajar a la vez con dos fuentes de datos distintas o, por ejemplo, dentro de un archivo Excel con dos hojas, es capaz de trabajar con ambas a la vez. Esto es posible gracias a la uniones o combinaciones de datos.

Las uniones de tablas y combinaciones de datos son dos formas distintas de vincular los datos en Tableau. Las uniones se realizan para vincular dos tablas de datos juntas, fila por fila, y las combinaciones para vincular fuentes de datos a nivel agregado.

○ UNIONES DE TABLAS DE DATOS

La mayoría de las bases de datos tienen múltiples datos de datos que están relacionadas de alguna manera. Por ejemplo, el siguiente archivo Excel *Superstore* tiene dos hojas, una para los pedidos y otra para las devoluciones, y se relacionan entre sí a través del campo ORDER ID, que aparece en ambas. La tabla primaria, que es la que tiene los datos de todos los pedidos, clientes, región etc. es ORDERS y la secundaria es la de RETURNS.

Tabla 2: Base de datos en Excel con dos hojas.

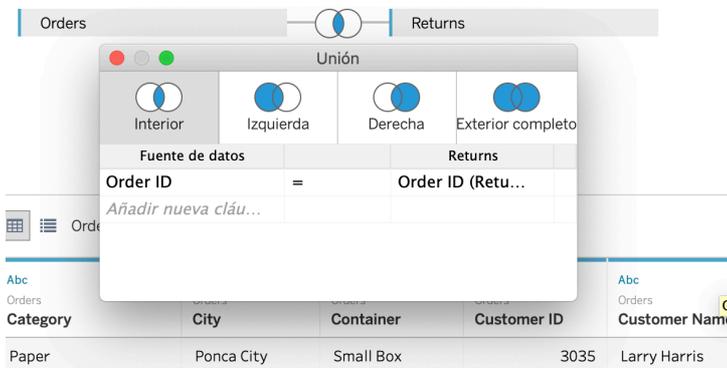


Figura 14: Unión de base de datos.

Existen cuatro tipos de uniones que se detallan en el Anexo.

○ **COMBINACIONES DE DATOS:**

Las combinaciones de datos permiten utilizar datos de múltiples fuentes en una misma visualización. La combinación de datos se realiza a nivel agregado e involucra diferentes consultas enviadas a cada fuente de datos, a diferencia de la unión, que se realiza a nivel de fila e involucra una sola consulta a una única fuente de datos.

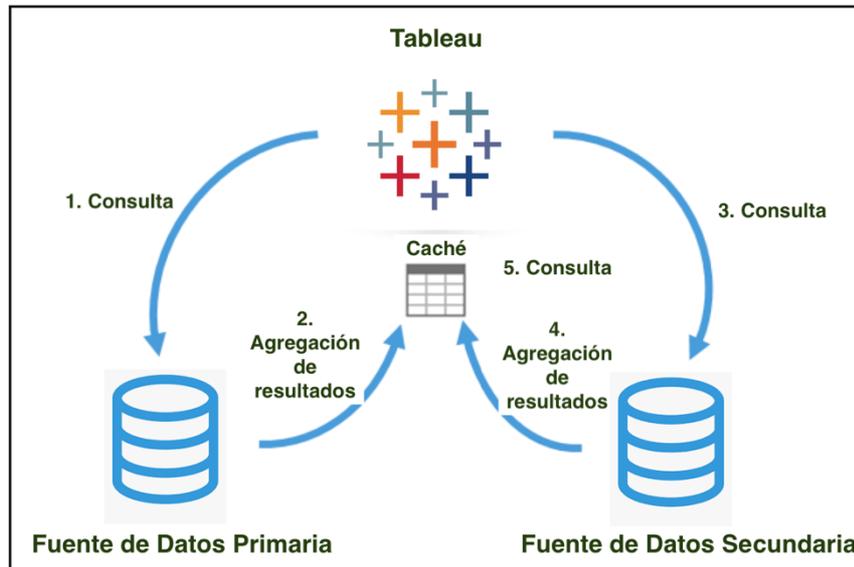


Figura 15: Combinación de bases de datos.

Se ejecuta de la siguiente manera:

1. Tableau ejecuta una consulta a la fuente de datos primaria.
2. Se devuelven los resultados agregados.
3. Tableau ejecuta otra consulta a la fuente de datos secundaria, esta consulta es filtrada en función del conjunto de valores devueltos por la fuente de datos primaria para las dimensiones que vinculan las dos fuentes de datos.
4. Se devuelven los resultados agregados.
5. Los resultados agregados de la fuente de datos primaria y los resultados agregados de la fuente de datos secundaria se combinan en la memoria caché.

La combinación de datos es diferente de la unión. Las uniones se logran en una sola consulta y los resultados coinciden fila por fila. La combinación de datos se produce al emitir dos consultas separadas y luego combinar los resultados agregados.

Los campos de enlace son dimensiones que se utilizan para unir datos combinados entre fuentes de datos primarias y secundarias. Los campos de enlace definen el nivel de detalle para la fuente secundaria. Estos se asignan automáticamente si los campos coinciden por nombre y tipo entre las fuentes de datos, aunque también se pueden asignar manualmente.

2.2 Dashboards

La idea con la que se concibió Tableau, fue para realizar Dashboard o Cuadros de Mando e interactuar con ellos.

En Tableau se pueden crear muchos tipos de gráficos o visualizaciones distintas, pero para poder ver la historia que hay detrás de los datos hace falta integrar las diferentes visualizaciones en un Dashboard.

Los Dashboards son una colección de visualizaciones, filtros, parámetros, imágenes o texto que trabajan juntos para extraer información y analizar un conjunto de datos. Los Dashboard pueden ser interactivos y hacer que el usuario explore los datos según su interés.

La interfaz de creación de un Dashboard es diferente a la que aparece al crear una visualización en una hoja de trabajo.

La pestaña de Dashboard tiene opciones de tamaño y desde donde se pueden añadir las hojas o los objetos al Dashboard. Los objetos que se pueden añadir son contenedores horizontales o verticales, donde se ubicarán las hojas de trabajo, además se pueden incluir imágenes o texto entre otros.

También puede seleccionar si los nuevos objetos que se agregarán serán en mosaico o flotante. Los objetos en mosaico encajan en un diseño en mosaico junto a otros objetos en mosaico o dentro de contenedores de diseño. Los objetos flotantes flotarán en la parte superior del tablero en capas sucesivas. Cuando una hoja de trabajo se agrega por primera vez a un tablero, cualquier leyenda, filtro o parámetro que fuera visible en la vista de la hoja de trabajo se agrega también al Dashboard.

Cuando se trabaja en el campo del Análisis de Datos, se crean numerosas visualizaciones. Cada una de estas visualizaciones aporta “una parte de la fotografía completa” en relación con la historia que existe dentro de los datos. Tableau permite reunir distintas visualizaciones en un solo Dashboard. Este Dashboard puede ser tanto una vista estática que muestre diferentes aspectos de los datos o un entorno interactivo, donde sea posible filtrar, profundizar e interactuar dinámicamente cada visualización del Dashboard.

El concepto de Dashboard en Tableau hace referencia a la disposición de visualizaciones individuales, junto con otros componentes como leyendas, filtros, parámetros, texto, imágenes, botones u objetos web que se organizan en un solo lienzo. Las visualizaciones y los componentes trabajan juntos para contar una historia de datos completa y real.

El analista de datos debe comprender muy bien la base de datos con la que está trabajando y, ser capaz de extraer la información y mostrarla gráficamente en el Dashboard.

2.2.1 La interfaz de los Dashboards.

Al hacer clic en la pestaña “Nuevo Dashboard” aparece un lienzo en donde se creará el Dashboard. En la parte de la izquierda hay opciones de diseño de este; como opciones de visualización (se puede configurar el tamaño para que sea compatible con la pantalla de un móvil, Ipad, o un ordenador), y una lista con las hojas de trabajo (las visualizaciones) que se hayan creado.

Arrastrando las hojas y cualquiera de los objetos como: texto, imágenes, botones al tablero, se crea el Dashboard.

Análisis de datos con Tableau.

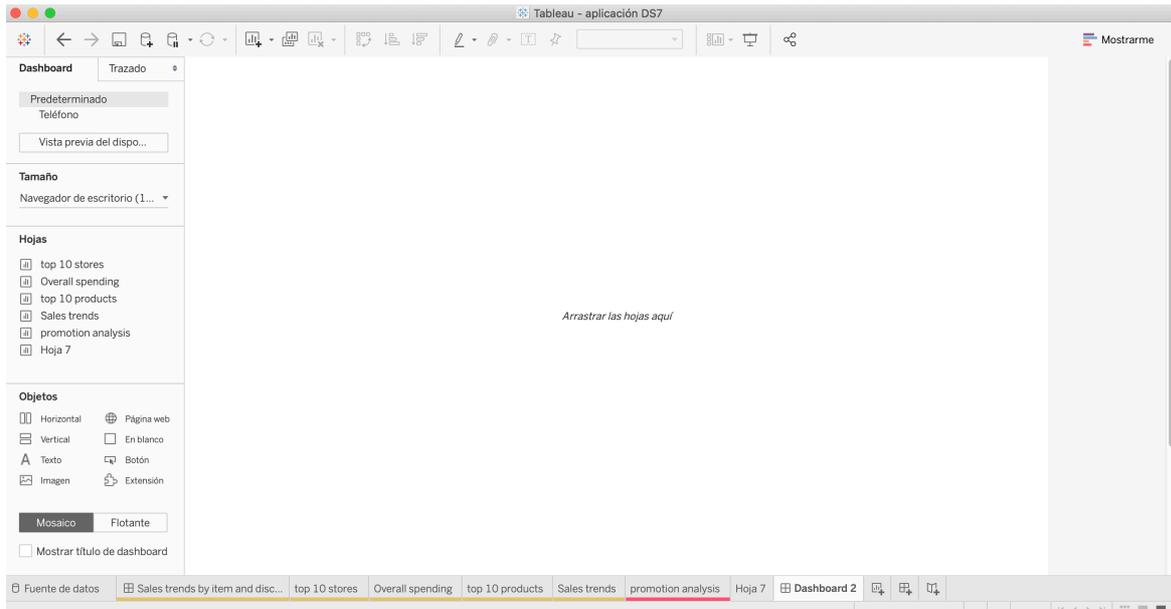


Figura 16: La interfaz de los Dashboards.

Utilizando la base de datos de Excel: Superstore, que es una base de datos distinta a la que se va a utilizar en el **Capítulo 4**, se va a realizar un análisis sencillo para ilustrar cómo se crea un Dashboard. La idea es identificar los productos menos rentables dentro de la base de datos.

Lo primero que hace es definir las hojas que van a integrarse dentro del Dashboard, pueden ser cualquiera de los tipos de gráficos que se explican en detalle en el **Capítulo 7: Anexo**. Las siguientes figuras muestran las cada una de las hojas con su gráfico correspondiente.

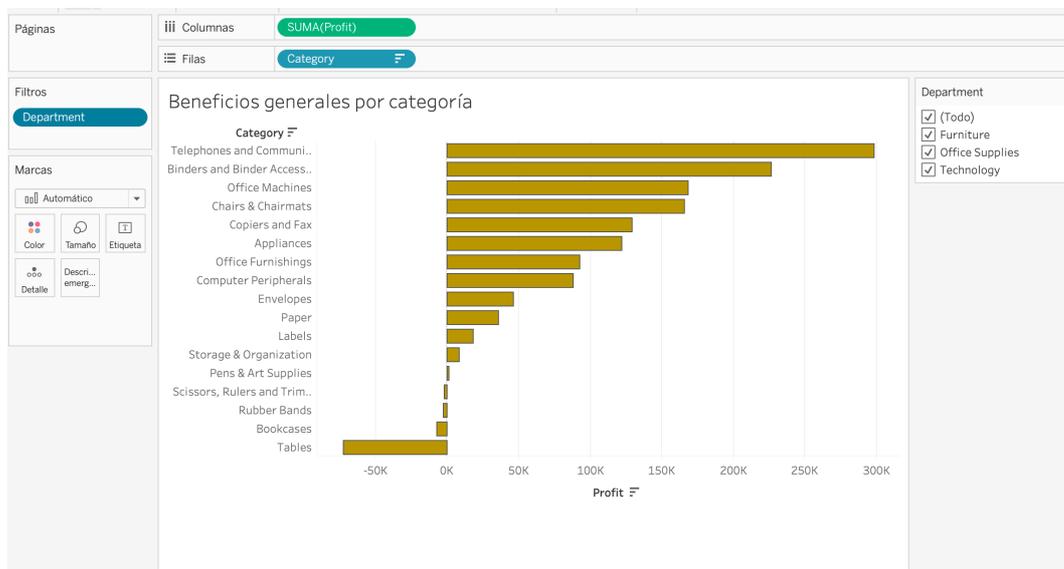


Figura 17: Beneficios por categoría.



Figura 18: Top 10 elementos menos rentables.



Figura 19: Beneficio por Estado.

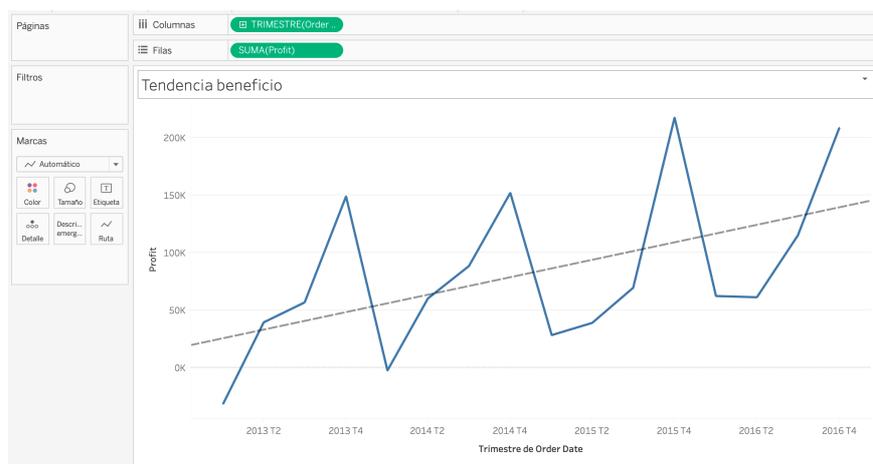


Figura 20: Pronóstico.

A continuación, se crea un Dashboard que contenga las cuatro visualizaciones que se han visto hasta ahora. Es necesario ver como se pueden relacionar cada una de estas gráficas para que tenga sentido el análisis, y la información que se muestre sea valiosa.

La Figura 17 tiene aplicado un filtro de departamento. Dicho filtro se aplica al resto de las visualizaciones, lo que hará que, para cada departamento seleccionado, se modifique el Dashboard mostrando los datos correspondientes, es decir, que los 10 elementos menos rentables, en vez pertenecer a toda la base de datos, sean los 10 menos rentables en la categoría seleccionada. El mismo razonamiento se sigue para las visualizaciones de “beneficio por estado” y “tendencia del beneficio”.

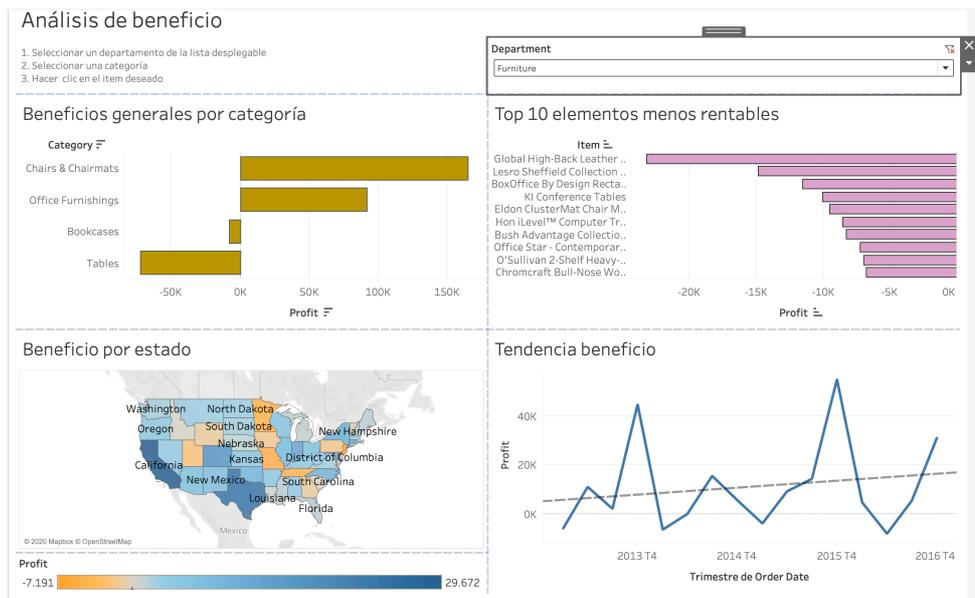


Figura 21: Análisis del beneficio por producto.

Para profundizar en el análisis hay que añadir lo que se conoce como “Acciones de Dashboard” que relacionan visualizaciones entre sí, utilizando algunas como visualización fuente y otras como objetivo. Con esto se favorece la exploración, los descubrimientos y las conclusiones.

Recapitulando; se tiene un tablero donde todas las gráficas tienen aplicado el filtro departamento. Además, se va a filtrar todo el Dashboard según la visualización de “Beneficios generales por categoría”.

Las acciones que se implementan son: Una acción con hoja de origen es “beneficios generales por categoría” y con hojas de destino las otras tres.

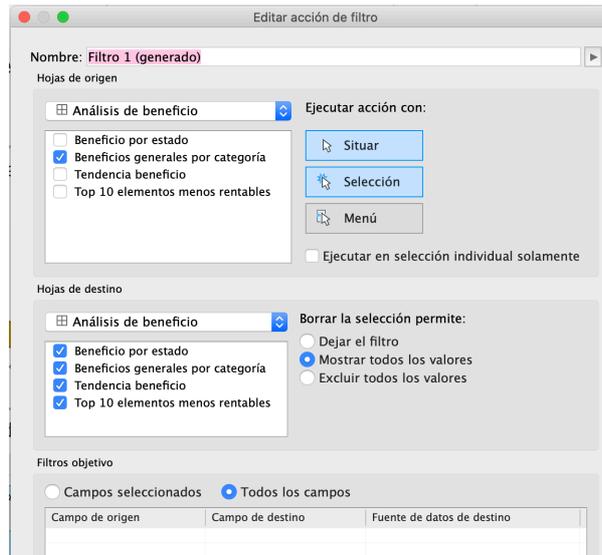


Figura 22: Acciones en Dashboards I.

Una segunda acción que, una vez se haya seleccionado un departamento y una categoría de ese departamento, al hacer clic en un ítem de la visualización “Top 10 elementos menos rentables”, permita analizar el beneficio por estado y la evolución temporal de ese ítem en particular.

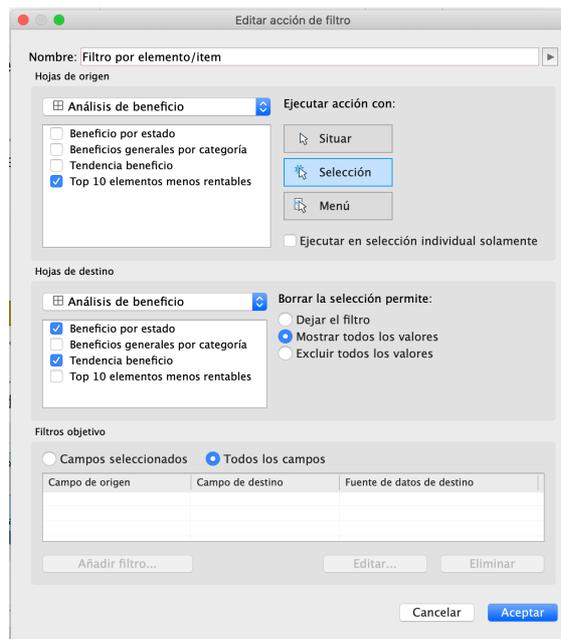


Figura 23: Acciones de Dashboard II.

Los títulos de las visualizaciones se pueden modificar, para hacer más sencillo el análisis se va a incluir el nombre del ítem que se haya seleccionado en la visualización “Top 10 elementos menos rentables” en la gráfica que muestra la evolución temporal del beneficio.

Análisis de datos con Tableau.

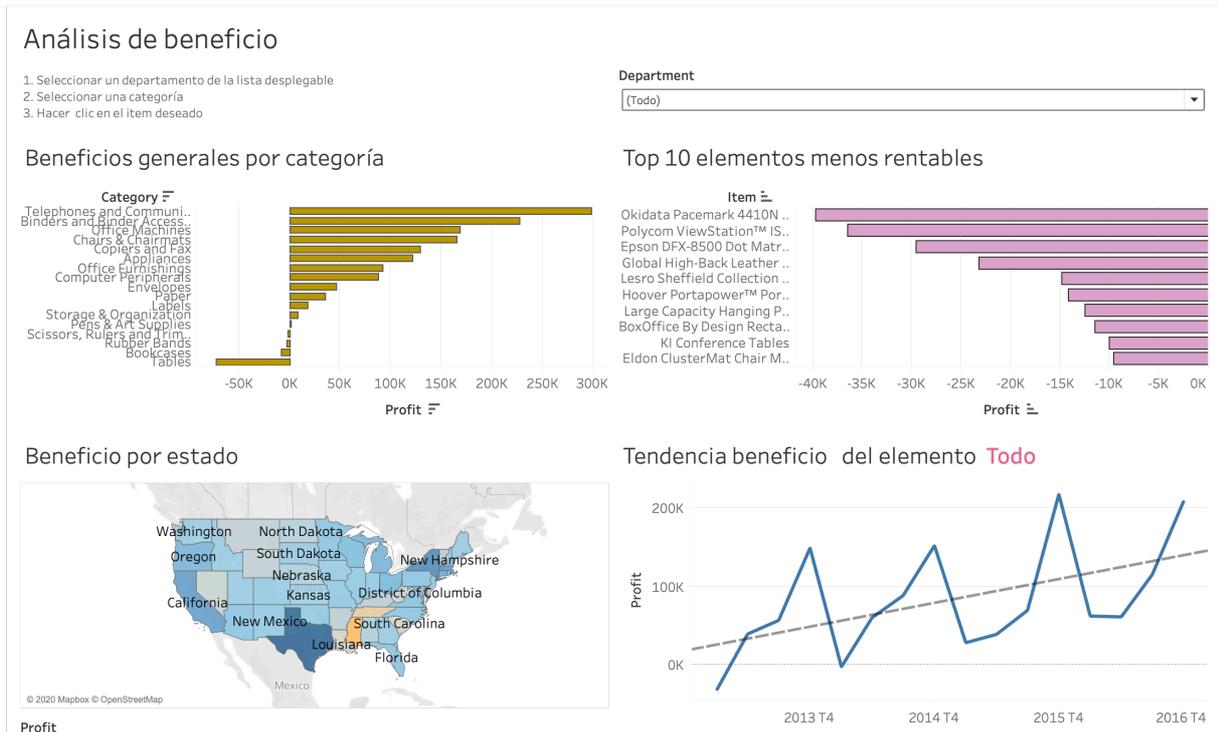


Figura 24: Dashboard.

Este Dashboard incluye un texto con las instrucciones para hacer análisis. Se puede ver el funcionamiento en la siguiente Figura:

- Se elige el departamento tecnología en la lista desplegable:



Figura 25: Filtro para Dashboard.

- Se selecciona la categoría de teléfonos:

Beneficios generales por categoría

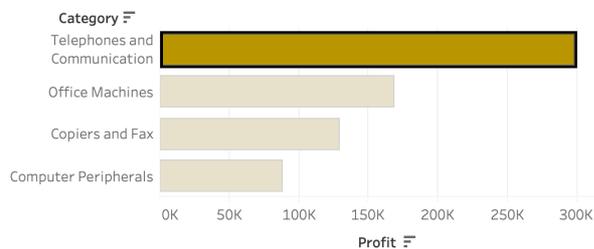


Figura 26: Filtro para Dashboard II.

- Se elige, por ejemplo, el ítem 10631, (el que menos beneficio tiene):

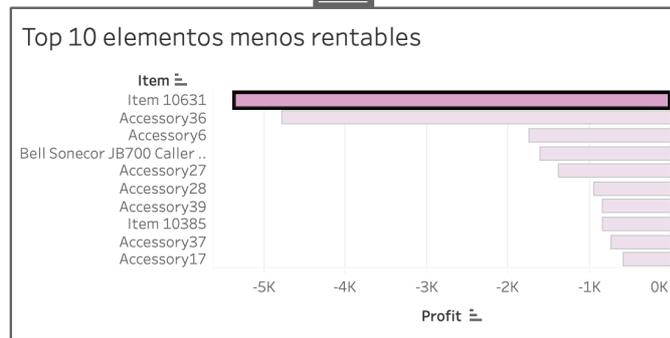


Figura 27: Filtro para Dashboard III.

Se ve que el Dashboard es interactivo y permite el análisis de la base de datos, tras este primer análisis el Dashboard quedaría como se ve en la Figura 28:

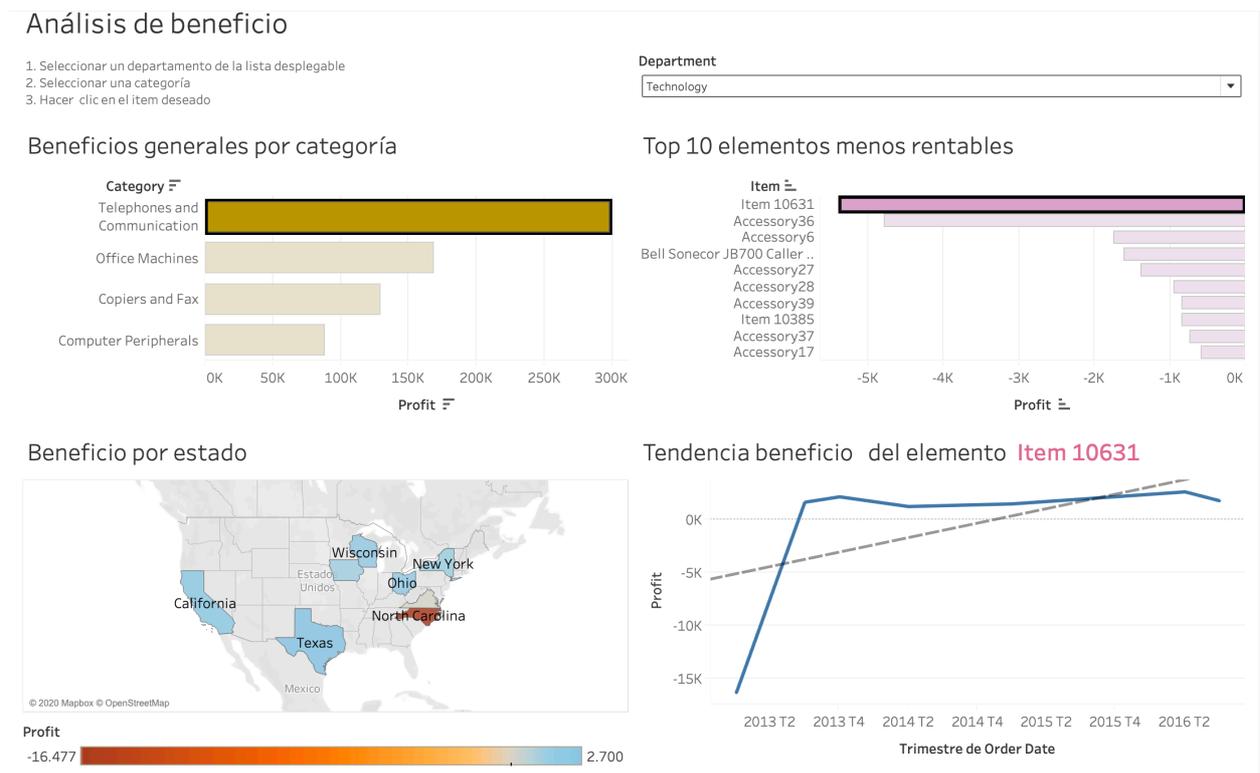


Figura 28: Dashboard de análisis del beneficio elementos menos rentables.

2.2.2 Cuadros de Mando en Dashboards.

Otra de las potencialidades que ofrece Tableau, es la opción de crear un Dashboard que funcione como un cuadro de mando integral (CMI). Utilizando la misma base de datos que en el apartado previo, se va a realizar un cuadro de mando que analice si cada una de las cuatro regiones que aparecen cumple con los objetivos establecidos para un KPI, se puede crear uno que mida la relación de beneficio (beneficio/ventas) y marcar como aceptable un valor superior al 15% , estableciendo este porcentaje como un parámetro, se puede modificar su valor para ver como cambia la situación al variar el mismo.

Análisis de datos con Tableau.

Como en la base de datos no aparece un campo con el dato relación de beneficio, se crea un campo calculado, para que se tenga el valor disponible.

El Dashboard contiene dos visualizaciones, la primera es un mapa que muestra las cuatro regiones coloreadas en función del valor de este KPI. La otra, es una tabla que desglosa mes a mes para cada una de las regiones que aparecen si este valor fue o no aceptable.

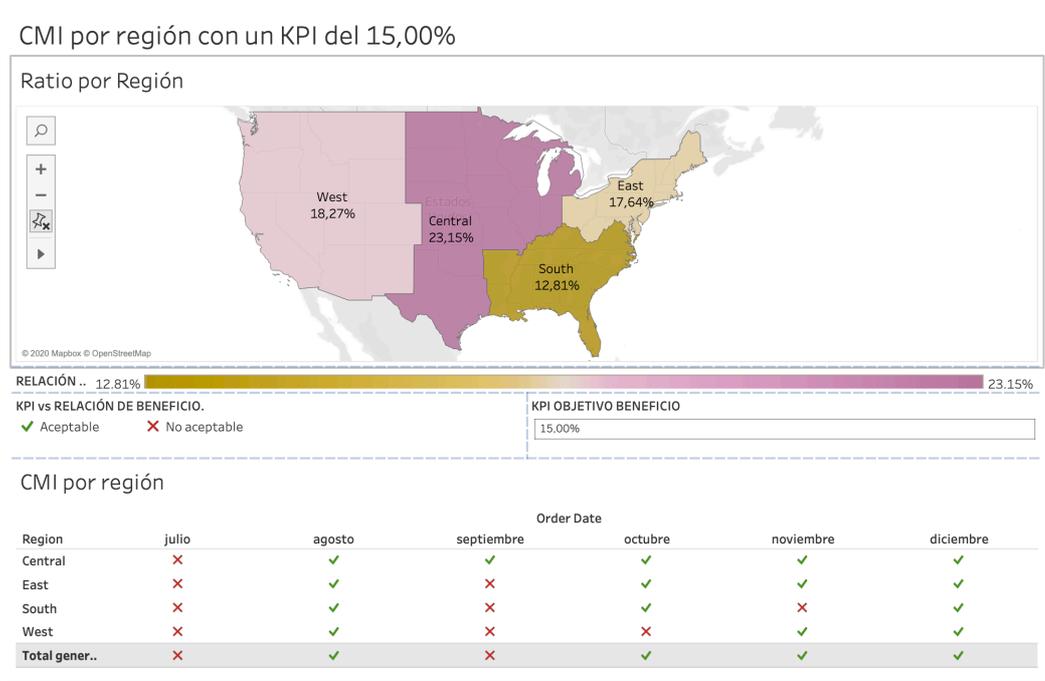


Figura 29: Dashboard con KPI 15%.

Si se quiere explorar como serían los resultados modificando el valor del KPI, se cambia el valor del parámetro en el Dashboard y se puede ver el comportamiento que desencadena este cambio de valor.

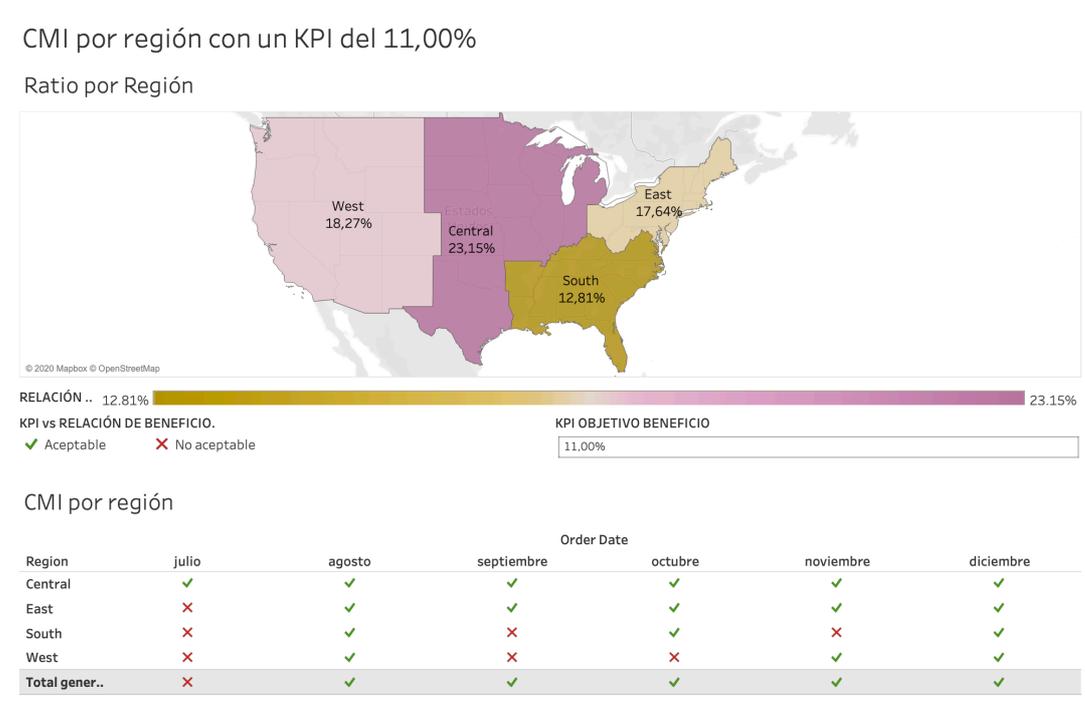


Figura 30: Dashboard con KPI 11%.

3 APLICACIÓN DE LA HERRAMIENTA A UN PROCESO DE GESTIÓN ESTRATÉGICA

Este capítulo se centra en la revisión de las potencialidades de la herramienta Tableau, y su aplicación a un proceso concreto, en este caso un proceso de crecimiento estratégico, utilizando la matriz BCG, también conocida como modelo Boston Consulting Group.

La matriz BCG es una matriz estratégica, eso significa que permite obtener un análisis de la situación de los factores que se estén analizando, y gracias a ese análisis, es posible definir las estrategias. En particular, la Matriz Boston Consulting Group es una matriz de crecimiento que se utiliza para evaluar el atractivo de la cartera de productos de una empresa y analizar qué productos son los más rentables para la misma y así determinar las estrategias que se quieran llevar a cabo. Fue desarrollada en la década de los 70 por el grupo Boston Consulting Group y su estructura es perfectamente adaptable a la actualidad.

Esta herramienta de análisis de negocio consiste en analizar la cartera de productos de una empresa en basándose en dos dimensiones.

→ La tasa de crecimiento del mercado: mantiene una fuerte relación con el atractivo de la industria en donde compite dicha empresa. Es decir, consiste en evaluar la evolución de la demanda de un producto en el mercado.

→ La tasa de participación del mercado: consiste en valorar la cuota de mercado de la empresa frente a la de la competencia. Es decir, que porcentaje de las ventas del mercado corresponde a la empresa.

La cartera de negocios de una compañía se compone de los productos que comercializa para un mercado concreto, uniéndolo a la matriz BCG se forma una herramienta de análisis interno de la empresa, que permite determinar qué productos deben seguir en la cartera y cuáles no.

El objetivo de la matriz es ayudar a determinar en qué productos se debe invertir más recursos. De esta manera se puede determinar cuáles son los mejores y peores recursos en cuanto a retorno de la inversión. Con esto se puede llevar a cabo 4 tipos de estrategias de inversión en marketing:

- Aumento de la cuota de mercado
- Defensa de la cuota de mercado
- Estrategia de cosecha
- Estrategia de eliminación

3.1 La Matriz BCG

La matriz se divide en cuatro cuadrantes divididos por dos ejes: El eje vertical, está formado por la tasa de crecimiento del mercado y el horizontal por la cuota de mercado de la empresa. Como se ve en la Figura 31, en cada uno de los cuatro cuadrantes existen cuatro tipos de productos diferentes: perro, vaca, interrogante y estrella.

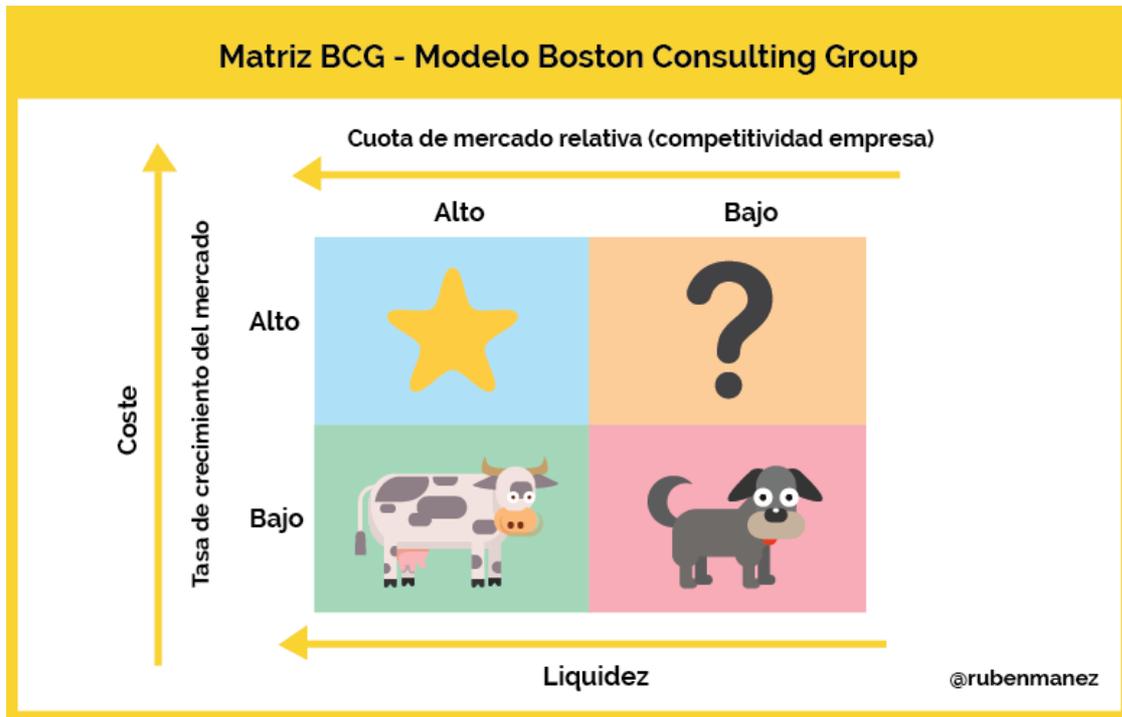


Figura 31: Ejemplo matriz BCG (Mañez, s.f.).

Cualquier empresa tendrá productos de los cuatro tipos, pero hay que tener claro que los productos vaca son los que generan más liquidez, con ese dinero se puede invertir en productos interrogantes, de los que se desconoce cuál va a ser la evolución, son aquellos que están en fase de desarrollo, pero que podrían convertirse en estrella. Por último, los productos vaca con el paso del tiempo terminan convirtiéndose en productos perro. Por tanto, la principal función de esta Matriz es analizar la cartera de productos de una empresa para colocarlos en cada cuadrante. En la tabla 3 se ven las características que presenta cada uno de los cuatro tipos de productos:

Tabla 3: Tipos de producto en la Matriz BCG

| | |
|---|---|
|  | <p>Generan liquidez, pero necesitan inversión para que se consoliden en el mercado, serían los productos líderes en el mercado con un gran potencial de crecimiento</p> |
|  | <p>Se conocen también como “generadores de caja”, ya que generan liquidez y no requieren de una gran inversión. Serían productos que ya están consolidados en un mercado que tiene poco crecimiento. La idea es mantener la situación competitiva de estos ya que son los que generan dinero que se puede invertir en otros</p> |
|  | <p>Generan poca liquidez y su cuota de mercado es baja, tienen un crecimiento limitado. Con estos productos se pueden plantear las siguientes situaciones:</p> <ul style="list-style-type: none"> - Quitarlos - Disminuir costes para sacar liquidez. |
|  | <p>Estos productos plantean el inconveniente de que se desconoce su evolución, necesitan inversión y su cuota de mercado es pequeña, por lo que en función de la inversión y estrategia se convertirán en productos estrella o perro.</p> |

En base al análisis realizado, y tras situar los productos en los cuadrantes se pueden realizar cuatro estrategias de inversión:

- Estrategia de crecimiento
- Estrategia de sostenimiento
- Estrategia de cosecha
- Estrategia de eliminación

Con los productos más fuertes (vaca y estrella) se realizarán estrategias de inversión y sostenimiento. Con los ingresos que se obtengan de las vacas se puede invertir en los interrogantes y convertirlos a estrella. De aquellos interrogantes cuyos niveles de ventas sean bajos, se eliminarán o se hará una estrategia de cosecha. Por otro lado, los productos perro deberán de eliminarse.

Esta matriz se ha relacionado a menudo con el ciclo de vida de un producto, representando su vida en el mercado al pasar por cuatro fases relacionadas con cada elemento de la matriz.

Introducción – Interrogante

Crecimiento – Vaca

Madurez – Estrella

Declive – Perro

3.2 La matriz BCG utilizando Tableau.

Para la validación de Tableau, se va a utilizar dicha herramienta para realizar este análisis de mercado con la Matriz BCG, para ello se necesita una base de datos con los productos a analizar y que además tenga datos de periodos (años, meses trimestres) una medida de ventas o ganancias, y categoría del producto y tipo de producto.

La base de datos utilizada se llama EU-Superstore, y se encuentra en el sitio web de DataWorld. (Dataworld, s.f.)

Tabla 4: Base de datos para matriz BCG.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|----|--------|------------|------------|-----------|-------------|------------|-------------------------|------------|-------------|-------------|---------|------------|------------|------------|-------------|----------|----------|----------|---------|---|
| 1 | Row ID | Order ID | Order Date | Ship Date | Ship Mode | Customer I | Customer I Segment | City | State | Country | Region | Product ID | Category | Sub-Categ | Product Ne | Sales | Quantity | Discount | Profit | |
| 2 | 1 | ES-2018-1 | 7/2/18 | 11/2/18 | Standard C | AS-10045 | Aaron Sme Corporate | Leeds | England | United King | North | OFF-ST-10 | Office Sup | Storage | Fellowes Fi | 79,2 | 3 | 0 | 39,6 | |
| 3 | 2 | ES-2018-1 | 7/2/18 | 11/2/18 | Standard C | AS-10045 | Aaron Sme Corporate | Leeds | England | United King | North | TEC-AC-1C | Technology | Accessorie | SanDisk Ni | 388,92 | 7 | 0 | 0 | |
| 4 | 3 | ES-2018-1 | 7/2/18 | 11/2/18 | Standard C | AS-10045 | Aaron Sme Corporate | Leeds | England | United King | North | OFF-LA-10 | Office Sup | Labels | Avery Legs | 35,19 | 3 | 0 | 16,11 | |
| 5 | 4 | ES-2018-1 | 7/2/18 | 11/2/18 | Standard C | AS-10045 | Aaron Sme Corporate | Leeds | England | United King | North | OFF-ST-10 | Office Sup | Storage | Fellowes Fi | 50,94 | 2 | 0 | 13,2 | |
| 6 | 5 | ES-2018-1 | 7/2/18 | 11/2/18 | Standard C | AS-10045 | Aaron Sme Corporate | Leeds | England | United King | North | TEC-AC-1C | Technology | Accessorie | Memorex M | 307,44 | 3 | 0 | 73,71 | |
| 7 | 6 | ES-2018-1 | 7/2/18 | 11/2/18 | Standard C | AS-10045 | Aaron Sme Corporate | Leeds | England | United King | North | OFF-ST-10 | Office Sup | Storage | Rogers Shi | 122,4 | 2 | 0 | 37,92 | |
| 8 | 7 | ES-2018-1 | 7/2/18 | 11/2/18 | Standard C | AS-10045 | Aaron Sme Corporate | Leeds | England | United King | North | TEC-PH-1C | Technology | Phones | Apple Sign | 413,82 | 3 | 0 | 20,61 | |
| 9 | 8 | ES-2018-5 | 2/8/16 | 7/8/16 | Second Cl | EB-13840 | Ellis Ballarc Corporate | West Brom | England | United King | North | TEC-CO-1C | Technology | Copiers | Canon Per | 428,22 | 3 | 0 | 192,69 | |
| 10 | 9 | ES-2018-5 | 2/8/16 | 7/8/16 | Second Cl | EB-13840 | Ellis Ballarc Corporate | West Brom | England | United King | North | OFF-AP-1C | Office Sup | Appliances | Hoover Stc | 3979,29 | 7 | 0 | 1989,54 | |
| 11 | 10 | ES-2018-5 | 2/8/16 | 7/8/16 | Second Cl | EB-13840 | Ellis Ballarc Corporate | West Brom | England | United King | North | OFF-FA-10 | Office Sup | Fasteners | Accos Pusi | 43,56 | 3 | 0 | 12,6 | |
| 12 | 11 | ES-2018-2 | 20/6/16 | 20/6/16 | Same Day | LH-17155 | Logan Hau Consumer | Le Bousca | Aquitaine-I | France | Central | OFF-LA-10 | Office Sup | Labels | Hon Color | 25,26 | 2 | 0 | 0,48 | |
| 13 | 12 | ES-2018-2 | 20/6/16 | 20/6/16 | Same Day | LH-17155 | Logan Hau Consumer | Le Bousca | Aquitaine-I | France | Central | OFF-AP-1C | Office Sup | Appliances | Hamilton B | 2443,905 | 5 | 0,1 | 760,305 | |
| 14 | 13 | ES-2018-2 | 20/6/16 | 20/6/16 | Same Day | LH-17155 | Logan Hau Consumer | Le Bousca | Aquitaine-I | France | Central | OFF-AR-1C | Office Sup | Art | Binney & S | 12,21 | 1 | 0 | 0 | |
| 15 | 14 | ES-2018-2 | 20/6/16 | 20/6/16 | Same Day | LH-17155 | Logan Hau Consumer | Le Bousca | Aquitaine-I | France | Central | TEC-PH-1C | Technology | Phones | Samsung | 2167,296 | 4 | 0,15 | 790,416 | |
| 16 | 15 | ES-2018-2 | 20/6/16 | 20/6/16 | Same Day | LH-17155 | Logan Hau Consumer | Le Bousca | Aquitaine-I | France | Central | OFF-ST-10 | Office Sup | Storage | Rogers Fol | 138,105 | 5 | 0,1 | -12,345 | |
| 17 | 16 | ES-2017-4 | 16/2/17 | 20/2/17 | Standard C | KC-16540 | Kelly Collis Consumer | Strasbourg | Alsace-Chi | France | Central | OFF-ST-10 | Office Sup | Storage | Smead Tra | 128,385 | 3 | 0,1 | 4,275 | |
| 18 | 17 | ES-2017-4 | 16/2/17 | 20/2/17 | Standard C | KC-16540 | Kelly Collis Consumer | Strasbourg | Alsace-Chi | France | Central | TEC-AC-1C | Technology | Accessorie | Belkin Men | 690,12 | 6 | 0 | 0 | |
| 19 | 18 | ES-2017-4 | 16/2/17 | 20/2/17 | Standard C | KC-16540 | Kelly Collis Consumer | Strasbourg | Alsace-Chi | France | Central | OFF-LA-10 | Office Sup | Labels | Hon Remo | 8,16 | 1 | 0 | 1,14 | |
| 20 | 19 | ES-2018-2 | 3/11/18 | 8/11/18 | Standard | CBF-11170 | Ben Ferrer Home Offic | Unna | North Rhin | Germany | Central | TEC-MA-1C | Technology | Machines | Okidata PR | 347,88 | 4 | 0 | 76,44 | |
| 21 | 20 | ES-2018-2 | 3/11/18 | 8/11/18 | Standard | CBF-11170 | Ben Ferrer Home Offic | Unna | North Rhin | Germany | Central | OFF-ST-10 | Office Sup | Storage | Eldon File | 575,505 | 5 | 0,1 | 12,705 | |
| 22 | 21 | ES-2018-1 | 14/8/18 | 16/8/18 | First Class | BF-11275 | Beth Fritzik Corporate | Thiais | Ile-de-Fran | France | Central | OFF-BI-10I | Office Sup | Binders | Acco Hole | 14,04 | 2 | 0 | 7,02 | |
| 23 | 22 | ES-2018-3 | 18/6/18 | 24/6/18 | Standard | DR-12680 | Dani Reich Corporate | Kassel | Hesse | Germany | Central | FUR-BO-1C | Furniture | Bookcases | Satco Stac | 266,164 | 2 | 0,1 | 5,904 | |
| 24 | 23 | IT-2018-32 | 2/5/18 | 5/5/18 | Second Cl | LB-16795 | Laurel Belt Home Offic | Milan | Lombardy | Italy | South | OFF-ST-10 | Office Sup | Storage | Rogers Shi | 36,72 | 1 | 0,4 | -5,52 | |
| 25 | 24 | IT-2018-32 | 2/5/18 | 5/5/18 | Second Cl | LB-16795 | Laurel Belt Home Offic | Milan | Lombardy | Italy | South | TEC-AC-1C | Technology | Accessorie | Memorex N | 28,17 | 1 | 0 | 1,11 | |
| 26 | 25 | IT-2017-12 | 12/8/17 | 16/8/17 | Standard | RO-19780 | Rose O'Bri Consumer | Brandenbu | Brandenbu | Germany | Central | OFF-ST-10 | Office Sup | Storage | Rogers Shi | 279,72 | 5 | 0,1 | -6,33 | |
| 27 | 26 | IT-2017-12 | 12/8/17 | 16/8/17 | Standard | RO-19780 | Rose O'Bri Consumer | Brandenbu | Brandenbu | Germany | Central | OFF-AR-1C | Office Sup | Art | BIC Pens, | 108,78 | 7 | 0 | 35,7 | |
| 28 | 27 | ES-2018-2 | 28/2/18 | 4/3/18 | Standard C | SP-20920 | Susan Pist Consumer | Heilbronn | Baden-Wu | Germany | Central | TEC-MA-1C | Technology | Machines | Panasonic | 332,01 | 7 | 0 | 112,77 | |
| 29 | 28 | ES-2018-2 | 28/2/18 | 4/3/18 | Standard C | SP-20920 | Susan Pist Consumer | Heilbronn | Baden-Wu | Germany | Central | OFF-ST-10 | Office Sup | Storage | Rogers Bo | 110,025 | 5 | 0,1 | -12,225 | |
| 30 | 29 | ES-2018-2 | 28/2/18 | 4/3/18 | Standard C | SP-20920 | Susan Pist Consumer | Heilbronn | Baden-Wu | Germany | Central | OFF-FA-10 | Office Sup | Fasteners | Stockwell F | 41,94 | 3 | 0 | 7,92 | |
| 31 | 30 | ES-2018-4 | 11/9/18 | 15/9/18 | Standard C | AZ-10750 | Annie Zype Consumer | Herten | North Rhin | Germany | Central | TEC-AC-1C | Technology | Accessorie | Belkin Mou | 192,6 | 5 | 0 | 32,7 | |
| 32 | 31 | ES-2018-4 | 11/9/18 | 15/9/18 | Standard C | AZ-10750 | Annie Zype Consumer | Herten | North Rhin | Germany | Central | TEC-PH-1C | Technology | Phones | Samsung | 336,24 | 2 | 0 | 43,68 | |
| 33 | 32 | ES-2015-2 | 2/6/15 | 6/6/15 | Standard | DB-13270 | Deborah B Home Offic | Barcelona | Catalonia | Spain | South | OFF-BI-10I | Office Sup | Binders | Cardinal Bi | 102,54 | 2 | 0 | 35,88 | |
| 34 | 33 | ES-2017-5 | 14/8/17 | 19/8/17 | Standard C | LB-16795 | Laurel Belt Home Offic | Swindon | England | United King | North | OFF-PA-1C | Office Sup | Paper | Green Bar | 63,27 | 3 | 0 | 30,96 | |

Como se puede apreciar, cuenta con los campos necesarios para la realización de la matriz, ya que tiene más de dos periodos de fecha para hacer la comparación.

Se eligen dos años para comparar, en este caso 2016 y 2017 y se definen dos campos calculados a partir de las ventas: Market Share (cuota de mercado) y Growth (crecimiento). Para definir los cuadrantes se crean dos parámetros: ejes vertical y horizontal de la matriz. Se van a comparar los productos de la empresa entre estos dos años y la participación en ventas de cada uno en la cartera de productos de la empresa.

Se crean dos parámetros, cuyo valor es variable, para definir los cuadrantes de la matriz. En función del valor que tomen dichos parámetros, se modificarán los iconos según el producto sea estrella, interrogante, vaca o perro.

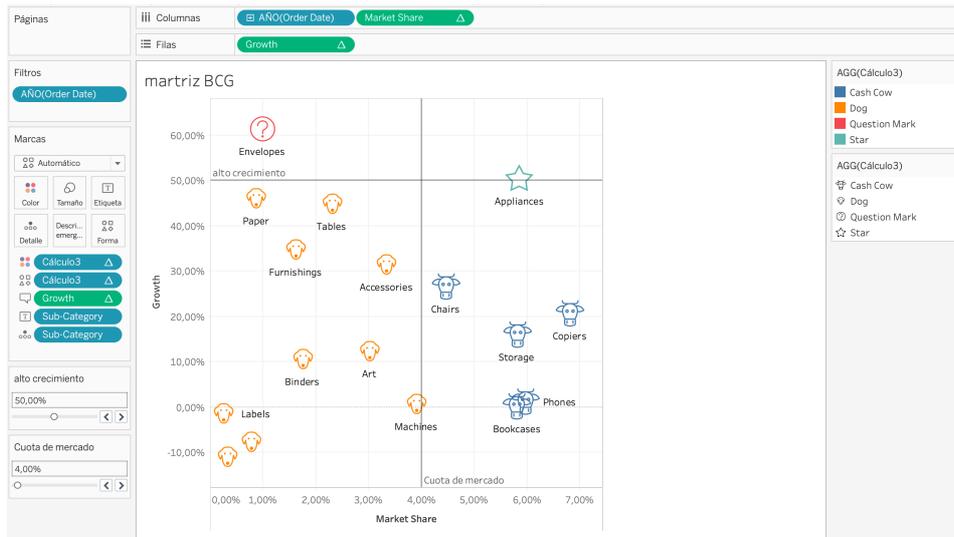


Figura 32: Matriz BCG

Para aportar más información, se crean dos visualizaciones más que añadir al Dashboard. Una que muestre los diez productos con mayores ventas dentro de cada categoría (y el beneficio de cada uno con un gradiente de color) y otra que muestre la tendencia en ventas a lo largo del tiempo.

Si se fija el valor del parámetro del crecimiento en un 50% y el que hace referencia a la cuota de mercado en un 4%, se obtiene que la única categoría estrella es la de accesorios y, dentro de ésta, el producto con más ventas, que a su vez coincide con el mayor beneficio es *Hover Stone, Red*. También se puede afirmar gracias a este Dashboard, que este producto ha tenido una evolución temporal en ventas creciente, a lo largo de los periodos estudiados.

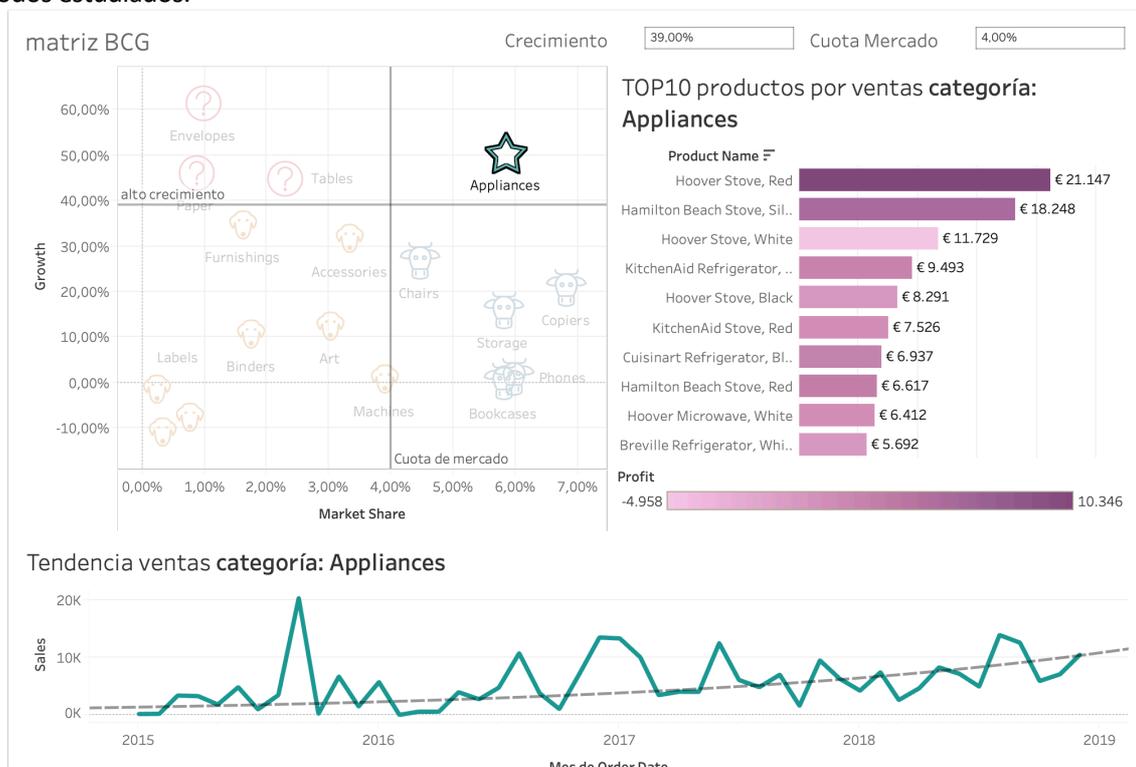


Figura 33: Dashboard matriz BCG

Con Tableau es posible crear un gráfico con la matriz BCG, que sitúe las categorías en cada uno de los cuadrantes definidos y, además añadirle valor realizando un análisis temporal de cada una de las categorías y clasificando los 10 productos con mayores ventas de cada categoría de la matriz.

Con este capítulo se demuestran que Tableau es una herramienta muy útil y se puede utilizar para diversas aplicaciones y en cualquier sector industrial.

En el **Capítulo 4: Caso de Estudio: Breakfast at the Frat** se profundiza en el análisis utilizando una base de datos de una cadena de supermercados, analizando distintos puntos como son la situación geográfica de las tiendas, o las categorías (o marcas) que más beneficio tienen.

4 CASO DE ESTUDIO: BREAKFAST AT THE FRAT

4.1 Justificación elección base de datos.

El caso de estudio pretende responder al título del documento: Análisis de Datos y Toma de Decisiones Utilizando una Herramienta de Business Intelligence en el sector Food Retail. Para ello, hace falta disponer de una base de datos amplia y completa, con gran cantidad de registros y con varias hojas con tablas de datos para poder llevar a cabo todas las potencialidades de la herramienta y así realizar un análisis y toma de decisiones completo.

La base de datos elegida es real, es decir, contiene los registros de las transacciones llevadas a cabo en una cadena de supermercados en EE. UU durante 156 semanas. Se ha obtenido en sitio web de Dunnhumby (Dunnhumby).

Dunnhumby es una empresa cuyo negocio principal es analizar los datos de casi mil millones de clientes en todo el mundo con el objetivo de ayudar a las corporaciones a tomar decisiones y maximizar sus beneficios. En España cuenta con sede en Madrid y Barcelona.

El sitio web de Dunnhumby ofrece recursos para ingenieros que quieran trabajar con bases de datos, existen cuatro bases de datos disponibles, y el análisis puede realizarse con cualquier herramienta, Python, SQL, R etc. En este caso, la conexión del archivo Excel será a Tableau, que es la herramienta que se está estudiando.

El campo del análisis de datos está en auge actualmente, cada vez son más datos los que se generan, por ello se ha hecho necesaria una nueva figura en las compañías encargada de su análisis y procesamiento.

En particular, en España el 94% de los retailers de alimentación los que tienen claro que quieren incrementar sus ingresos para 2020. Para lograrlo han afirmado que lo harán a través de la mejora del uso de los insights que extraigan de su análisis de datos para desarrollar estrategias de clientes y tomar decisiones comerciales (Ortiz, 2020).

Por esta razón se ha elegido una base de datos del sector Food Retail para sobre los campos que aparecen que se describen más adelante en este capítulo extraer la mayor cantidad de información y dirigir la estrategia de la empresa



Figura 34: Justificación elección Dataset

4.2 Escenario de estudio y limpieza de datos.

Como se ha indicado en el apartado anterior del documento la base de datos se encuentra en los recursos de ingeniería del sitio web de Dunnhumby. Al descargar el archivo, además de la base de datos en Excel, se incluye un PDF con detalles de la base de datos, como el significado de cada campo y su ubicación se dentro de las tres tablas que aparecen. (Dunnhumby)

Como se tienen tres tablas distintas, se van a utilizar los campos de las tres tablas hay que buscar la manera de relacionarlos, esto se hace identificando la fuente primaria, es decir, aquella que contiene campos que aparecen en las otras tablas (Dunnhumby). En este caso la primaria es: dh Transactional Data:

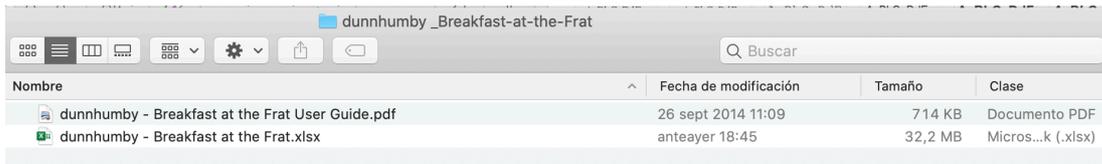


Figura 35: Carpeta con la base de datos y pdf.

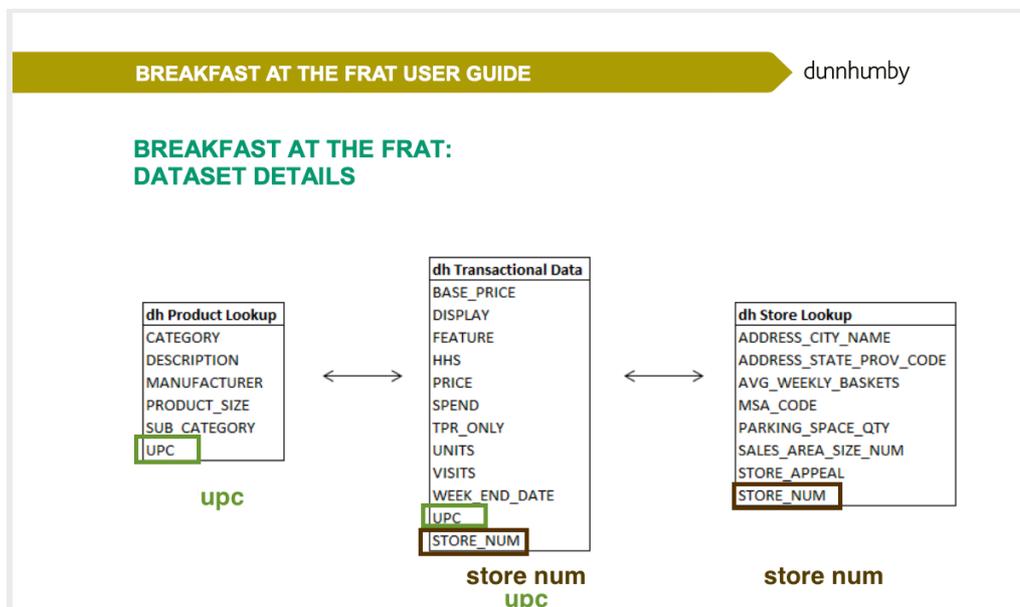


Figura 36: Unión de las tres tablas de datos.

Antes de realizar la conexión a los datos y comenzar con el análisis es necesario explorar el archivo en Excel, para ver que es cada variable, si existen campos nulos o duplicados y si se va a cambiar el nombre de las variables por otros más intuitivos o incluso, prescindir de alguna.

El archivo de Excel consta de cuatro hojas, la primera es un Glosario de los términos utilizados (los nombres de las variables) acompañados de la tabla en la que se encuentran, de las tres tablas posibles, y junto con una descripción de a lo que hace referencia.

Tabla 5: Base de datos. Hoja 1 Glosario.

| dunnhumby SOURCE FILES | | Breakfast at the Frat: A Time Series Analysis | | | | | | | | | |
|---|-------------------------|---|--|--|--|--|--|--|--|--|--|
| The data contained in this file includes sales and promotional information from a sampling of stores over 156 weeks, beginning January 2009 through December 2011. These are the sales metrics of products in 4 categories: mouthwash, pretzels, frozen pizza, and boxed cereal, the top 5 products from each of the top 3 brands in each category. | | | | | | | | | | | |
| 4 | VARIABLE NAME | TABLE | DESCRIPTION | | | | | | | | |
| 5 | ADDRESS_CITY_NAME | store lookup | city | | | | | | | | |
| 6 | ADDRESS_STATE_PROV_CODE | store lookup | state | | | | | | | | |
| 7 | AVG_WEEKLY_BASKETS | store lookup | average weekly baskets sold in the store | | | | | | | | |
| 8 | BASE_PRICE | data | base price of item | | | | | | | | |
| 9 | MANUFACTURER | products lookup | manufacturer | | | | | | | | |
| 10 | CATEGORY | products lookup | category of product | | | | | | | | |
| 11 | DESCRIPTION | products lookup | product description | | | | | | | | |
| 12 | DISPLAY | data | product was a part of in-store promotional display | | | | | | | | |
| 13 | FEATURE | data | product was in in-store circular | | | | | | | | |
| 14 | HHS | data | # of purchasing households | | | | | | | | |
| 15 | MSA_CODE | store lookup | (Metropolitan Statistical Area) geographic region with a high core population density and close economic ties throughout the surrounding areas | | | | | | | | |
| 16 | PARKING_SPACE_QTY | store lookup | number of parking spaces in the store parking lot | | | | | | | | |
| 17 | PRICE | data | actual amount charged for the product at shelf | | | | | | | | |
| 18 | WEEK_END_DATE | data | week ending date | | | | | | | | |
| 19 | SALES_AREA_SIZE_NUM | store lookup | square footage of store | | | | | | | | |
| 20 | STORE_APPEAL | store lookup | Retailer's designated store appeal | | | | | | | | |
| 21 | SPEND | data | total spend (i.e., \$ sales) | | | | | | | | |
| 22 | STORE_NUM | data, store lookup | store number | | | | | | | | |
| 23 | SUB_CATEGORY | products lookup | sub-category of product | | | | | | | | |
| 24 | TPR_ONLY | data | temporary price reduction only (i.e., shelf tag only, product was reduced in price but not on display or in an advertisement) | | | | | | | | |
| 25 | UNITS | data | units sold | | | | | | | | |
| 26 | UPC | data, products lookup | (Universal Product Code) product specific identifier | | | | | | | | |
| 27 | VISITS | data | number of unique purchases (baskets) that included the product | | | | | | | | |
| 28 | PRODUCT_SIZE | products lookup | package size or quantity of product | | | | | | | | |

Las tres siguientes ilustraciones muestran las hojas tres hojas de la base de datos.

Tabla 6: Base de datos. Hoja 2.

| dunnhumby SOURCE FILES | | Breakfast at the Frat: A Time Series Analysis | | | | | | | | | |
|--------------------------|----------|---|-------------------|-------------------------|----------|----------------|-------------------|---------------------|--------------------|--|--|
| 2 | STORE_ID | STORE_NAME | ADDRESS_CITY_NAME | ADDRESS_STATE_PROV_CODE | MSA_CODE | SEG_VALUE_NAME | PARKING_SPACE_QTY | SALES_AREA_SIZE_NUM | AVG_WEEKLY_BASKETS | | |
| 3 | 389 | SILVERLAKE | ERLANGER | KY | 17140 | MAINSTREAM | 408 | 46073 | 24767 | | |
| 4 | 2277 | ANDERSON TOWNE CTR | CINCINNATI | OH | 17140 | UPSCALE | | 81958 | 54053 | | |
| 5 | 4259 | WARSAW AVENUE | CINCINNATI | OH | 17140 | VALUE | | 48813 | 31177 | | |
| 6 | 6379 | KINGWOOD | KINGWOOD | TX | 26420 | MAINSTREAM | | 50237 | 20620 | | |
| 7 | 6431 | AT WARD ROAD | BAYTOWN | TX | 26420 | VALUE | 350 | 43698 | 24322 | | |
| 8 | 13853 | RICHARDSON | RICHARDSON | TX | 19100 | MAINSTREAM | | 53586 | 15787 | | |
| 9 | 13859 | MCKINNEY | MCKINNEY | TX | 19100 | MAINSTREAM | | 47824 | 15345 | | |
| 10 | 15755 | KROGER JUNCTION S/C | PASADENA | TX | 26420 | VALUE | 351 | 34373 | 18291 | | |
| 11 | 23327 | HOUSTON | HOUSTON | TX | 26420 | MAINSTREAM | | 50722 | 30258 | | |
| 12 | 23345 | FRISCO | FRISCO | TX | 19100 | UPSCALE | | 46369 | 21947 | | |
| 13 | 25001 | PINEWOOD | DAYTON | OH | 19380 | VALUE | 351 | 34177 | 18035 | | |
| 14 | 25983 | DAYTON | DAYTON | OH | 19380 | VALUE | | 48782 | 31461 | | |
| 15 | 623 | HOUSTON | HOUSTON | TX | 26420 | MAINSTREAM | | 46930 | 36741 | | |
| 16 | 2541 | NORWOOD | ARLINGTON | TX | 19100 | MAINSTREAM | 300 | 29231 | 11019 | | |
| 17 | 4489 | SWEETWATER PLAZA | SUGAR LAND | TX | 26420 | UPSCALE | | 49030 | 21345 | | |
| 18 | 10019 | AT EASTEX FRWY | HOUSTON | TX | 26420 | VALUE | 221 | 35675 | 19068 | | |
| 19 | 13609 | VANDALIA | VANDALIA | OH | 19380 | MAINSTREAM | | 55663 | 21232 | | |
| 20 | 15531 | GOSHEN | GOSHEN | OH | 17140 | VALUE | | 33667 | 13344 | | |
| 21 | 17599 | KEARNEY | MESQUITE | TX | 19100 | VALUE | 300 | 36994 | 15134 | | |
| 22 | 19265 | LATONIA | COVINGTON | KY | 17140 | MAINSTREAM | | 55685 | 31578 | | |
| 23 | 21485 | HOUSTON | KATY | TX | 26420 | MAINSTREAM | | 46369 | 26472 | | |
| 24 | 26981 | SPRINGFIELD | SPRINGFIELD | OH | 44220 | MAINSTREAM | | 42081 | 25122 | | |
| 25 | 4503 | ROCKWALL | ROCKWALL | TX | 19100 | MAINSTREAM | | 56230 | 25289 | | |
| 26 | 2513 | HOUSTON | HOUSTON | TX | 26420 | UPSCALE | | 61833 | 32423 | | |
| 27 | 19523 | DUNCANVILLE | DUNCANVILLE | TX | 19100 | VALUE | 400 | 48594 | 24568 | | |
| 28 | 19533 | DENTON | DENTON | TX | 19100 | MAINSTREAM | | 48330 | 28243 | | |
| 29 | 25021 | HAMILTON | HAMILTON | OH | 17140 | VALUE | 1441 | 62961 | 34191 | | |
| 30 | 367 | 15TH & MADISON | COVINGTON | KY | 17140 | VALUE | 196 | 24721 | 12707 | | |
| 31 | 2505 | DICKINSON VILLAGE | DICKINSON | TX | 26420 | MAINSTREAM | 319 | 41644 | 14460 | | |
| 32 | 6187 | ST. MARYS | SAINT MARYS | OH | 47540 | MAINSTREAM | | 55474 | 20015 | | |
| 33 | 8035 | OVER-THE-RHINE | CINCINNATI | OH | 17140 | VALUE | 17 | 10788 | 10435 | | |
| 34 | 8041 | LANDEN | MAINEVILLE | OH | 17140 | MAINSTREAM | 1200 | 49074 | 14849 | | |

Caso de Estudio: Breakfast at the Frat

Tabla 7: Base de datos. Hoja 3.

| UPC | DESCRIPTION | MANUFACTURER | CATEGORY | SUB_CATEGORY | PRODUCT_SIZE |
|------------|----------------------------|---------------|-----------------------|-----------------------------|--------------|
| 1111009477 | PL MINI TWIST PRETZELS | PRIVATE LABEL | BAG SNACKS | PRETZELS | 15 OZ |
| 1111009497 | PL PRETZEL STICKS | PRIVATE LABEL | BAG SNACKS | PRETZELS | 15 OZ |
| 1111009507 | PL TWIST PRETZELS | PRIVATE LABEL | BAG SNACKS | PRETZELS | 15 OZ |
| 1111035398 | PL BL MINT ANTSPTC RINSE | PRIVATE LABEL | ORAL HYGIENE PRODUCTS | MOUTHWASHES (ANTISEPTIC) | 1.5 LT |
| 1111038078 | PL BL MINT ANTSPTC RINSE | PRIVATE LABEL | ORAL HYGIENE PRODUCTS | MOUTHWASHES (ANTISEPTIC) | 500 ML |
| 1111038080 | PL ANTSPTC SPG MNT MTHWS | PRIVATE LABEL | ORAL HYGIENE PRODUCTS | MOUTHWASHES (ANTISEPTIC) | 500 ML |
| 1111085319 | PL HONEY NUT TOASTD OATS | PRIVATE LABEL | COLD CEREAL | ALL FAMILY CEREAL | 12.25 OZ |
| 1111085345 | PL RAISIN BRAN | PRIVATE LABEL | COLD CEREAL | ADULT CEREAL | 20 OZ |
| 1111085350 | PL BT SZ FRSTD SHRD WHT | PRIVATE LABEL | COLD CEREAL | ALL FAMILY CEREAL | 18 OZ |
| 1111087395 | PL SR CRUST SUPRM PIZZA | PRIVATE LABEL | FROZEN PIZZA | PIZZA/PREMIUM | 32.7 OZ |
| 1111087396 | PL SR CRUST 3 MEAT PIZZA | PRIVATE LABEL | FROZEN PIZZA | PIZZA/PREMIUM | 30.5 OZ |
| 1111087398 | PL SR CRUST PEPPRN PIZZA | PRIVATE LABEL | FROZEN PIZZA | PIZZA/PREMIUM | 29.6 OZ |
| 1600027527 | GM HONEY NUT CHEERIOS | GENERAL MI | COLD CEREAL | ALL FAMILY CEREAL | 12.25 OZ |
| 1600027528 | GM CHEERIOS | GENERAL MI | COLD CEREAL | ALL FAMILY CEREAL | 18 OZ |
| 1600027564 | GM CHEERIOS | GENERAL MI | COLD CEREAL | ALL FAMILY CEREAL | 12 OZ |
| 2066200530 | NWMN OWN PEPPERONI PIZZA | KING | FROZEN PIZZA | PIZZA/PREMIUM | 13.2 OZ |
| 2066200531 | NWMN OWN 4 CHEESE PIZZA | KING | FROZEN PIZZA | PIZZA/PREMIUM | 13.3 OZ |
| 2066200532 | NWMN OWN SUPREME PIZZA | KING | FROZEN PIZZA | PIZZA/PREMIUM | 14.7 OZ |
| 2840002333 | RLDGLD BRAIDED HONEY WHT | FRITO LAY | BAG SNACKS | PRETZELS | 10 OZ |
| 2840004768 | RLDGLD TINY TWISTS PRTL | FRITO LAY | BAG SNACKS | PRETZELS | 16 OZ |
| 2840004770 | RLDGLD PRETZEL STICKS | FRITO LAY | BAG SNACKS | PRETZELS | 16 OZ |
| 3000006340 | QKER LIFE ORIGINAL | QUAKER | COLD CEREAL | ALL FAMILY CEREAL | 13 OZ |
| 3000006560 | QKER CAP N CRUNCH BERRIES | QUAKER | COLD CEREAL | KIDS CEREAL | 13 OZ |
| 3000006610 | QKER CAP N CRUNCH | QUAKER | COLD CEREAL | KIDS CEREAL | 14 OZ |
| 3120505000 | HMRN CLSC SAUSAGE PIZZA | HOME RUN | FROZEN PIZZA | PIZZA/PREMIUM | 30 OZ |
| 3120505100 | HMRN CLSC SSG PEPP PIZZA | HOME RUN | FROZEN PIZZA | PIZZA/PREMIUM | 31 OZ |
| 3120506000 | HMRN CLSC CHS PIZZA | HOME RUN | FROZEN PIZZA | PIZZA/PREMIUM | 27 OZ |
| 3500068914 | COLG SPEARMINT WISP | COLGATE | ORAL HYGIENE PRODUCTS | MOUTHWASH/RINSES AND SPRAYS | 16 CT |
| 3700019521 | CREST PH WHITG CLN MINT TP | P & G | ORAL HYGIENE PRODUCTS | MOUTHWASH/RINSES AND SPRAYS | 4.2 OZ |
| 3700031613 | SCOFM ORIG MINT MOUTHWASH | P & G | ORAL HYGIENE PRODUCTS | MOUTHWASH/RINSES AND SPRAYS | 1 LT |
| 3700044982 | CREST PH CLN MINT RINSE | P & G | ORAL HYGIENE PRODUCTS | MOUTHWASHES (ANTISEPTIC) | 1 LT |
| 3800031829 | KELL BITE SIZE MINI WHEAT | KELLOGG | COLD CEREAL | ALL FAMILY CEREAL | 18 OZ |

Tabla 8: Base de datos. Hoja 4.

| WEEK_END_DATE | STORE_NUM | UPC | UNITS | VISITS | HHS | SPEND | PRICE | BASE_PRICE | FEATURE | DISPLAY | TPR_ONLY |
|---------------|-----------|-------------|-------|--------|-----|--------|-------|------------|---------|---------|----------|
| 14-ene-09 | 367 | 1111009477 | 13 | 13 | 13 | 18,07 | 1,39 | 1,57 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 1111009497 | 20 | 18 | 18 | 27,8 | 1,39 | 1,39 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 1111009507 | 14 | 14 | 14 | 19,32 | 1,38 | 1,38 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 1111035398 | 4 | 3 | 3 | 14 | 3,5 | 4,49 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 1111038078 | 3 | 3 | 3 | 7,5 | 2,5 | 2,5 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 1111038080 | 2 | 2 | 2 | 5,18 | 2,59 | 2,59 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 1111085319 | 14 | 13 | 13 | 26,32 | 1,88 | 1,88 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 1111085345 | 29 | 26 | 25 | 54,52 | 1,88 | 1,88 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 1111085350 | 35 | 27 | 25 | 69,3 | 1,98 | 1,98 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 1111087395 | 50 | 40 | 40 | 168 | 3,36 | 3,94 | 0 | 1 | 0 |
| 14-ene-09 | 367 | 1111087396 | 61 | 43 | 42 | 203,13 | 3,33 | 4 | 0 | 1 | 0 |
| 14-ene-09 | 367 | 1111087398 | 72 | 52 | 49 | 238,32 | 3,31 | 4 | 0 | 1 | 0 |
| 14-ene-09 | 367 | 1600027527 | 12 | 10 | 10 | 38,28 | 3,19 | 3,19 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 1600027528 | 31 | 26 | 19 | 142,29 | 4,59 | 4,59 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 1600027564 | 56 | 48 | 42 | 152,32 | 2,72 | 3,07 | 1 | 0 | 0 |
| 14-ene-09 | 367 | 2840004768 | 8 | 6 | 6 | 23,12 | 2,89 | 2,89 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 3000006340 | 19 | 17 | 17 | 38,76 | 2,04 | 2,74 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 3000006560 | 87 | 76 | 72 | 174 | 2 | 2,59 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 3000006610 | 48 | 41 | 41 | 96 | 2 | 2,59 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 3700019521 | 2 | 2 | 2 | 5 | 2,5 | 3,49 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 3700044982 | 2 | 2 | 2 | 10 | 5 | 5 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 3800031829 | 14 | 11 | 3 | 43,96 | 3,14 | 3,14 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 3800031838 | 36 | 35 | 34 | 106,56 | 2,96 | 2,96 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 3800039118 | 59 | 58 | 56 | 157,53 | 2,67 | 3,29 | 1 | 1 | 0 |
| 14-ene-09 | 367 | 4116709428 | 1 | 1 | 1 | 3,99 | 3,99 | 4,59 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 7027316204 | 13 | 8 | 8 | 26,26 | 2,02 | 3,29 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 7027316404 | 11 | 8 | 8 | 13,2 | 1,2 | 1,92 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 7192100336 | 12 | 9 | 9 | 71,88 | 5,99 | 6,29 | 1 | 0 | 0 |
| 14-ene-09 | 367 | 7192100337 | 23 | 22 | 22 | 137,77 | 5,99 | 6,29 | 1 | 0 | 0 |
| 14-ene-09 | 367 | 7192100339 | 17 | 15 | 15 | 101,83 | 5,99 | 6,29 | 1 | 0 | 0 |
| 14-ene-09 | 367 | 7218063052 | 1 | 1 | 1 | 6,46 | 6,46 | 6,46 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 7218063979 | 1 | 1 | 1 | 6,46 | 6,46 | 6,46 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 7218063983 | 1 | 1 | 1 | 6,46 | 6,46 | 6,46 | 0 | 0 | 0 |
| 14-ene-09 | 367 | 7797508004 | 1 | 1 | 1 | 2,5 | 2,5 | 2,89 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 7797508006 | 1 | 1 | 1 | 2,5 | 2,5 | 2,89 | 0 | 0 | 1 |
| 14-ene-09 | 367 | 31254742835 | 1 | 1 | 1 | 4,99 | 4,99 | 4,99 | 0 | 0 | 0 |

Se ha dado un formato más visual al glosario y se ha traducido la descripción al castellano, aunque los nombres de los campos se mantendrán en inglés ya que es el idioma por defecto de la base de datos. Como el campo DESCRIPTION en realidad hace referencia al producto o ítem, se ha cambiado el nombre para que no lleve a confusión. De igual manera el campo MANUFACTURER se cambiará a BRAND.

Los campos PRODUCT_SIZE, SALES_AREA_SIZE_NUMBER Y PARKING_SPACE_QTY y SEG_VALUE_NAME (que tomaba tres valores: Mainstream, Upscale o Value y no se entendía bien el concepto), se han excluido del análisis.

Tabla 9: Base de datos tras la limpieza.

| VARIABLE NAME | TABLE | DESCRIPTION | DESCRIPCIÓN |
|--------------------|-----------------------|---|---|
| CITY NAME | store lookup | city | Ciudad |
| STATE | store lookup | state | Estado |
| AVG_WEEKLY_BASKETS | store lookup | average weekly baskets sold in the store | Promedio de cestas de compra vendidas por semana en cada tienda |
| BASE_PRICE | data | base price of item | Precio base del ítem |
| BRAND | products lookup | manufacturer | Marca |
| CATEGORY | products lookup | category of product | Categoría |
| ITEM | products lookup | product description | Ítem |
| DISPLAY | data | product was a part of in-store promotional display | Producto en exhibición en tienda |
| FEATURE | data | product was in in-store circular) | Producto en el folleto publicitario de la tienda |
| HHS | data | # of purchasing households | Número de hogares compradores |
| MSA_CODE | store lookup | (Metropolitan Statistical Area) geographic region with a high core population density | Similar a un código postal |
| PRICE | data | actual amount charged for the product at shelf | Precio de venta final |
| WEEK_DATE | data | week ending date | fecha compra SEMANA = Semana de compra |
| STORE_NAME | store lookup | Retailer's designated store appeal | Nombre de la tienda |
| SPEND | data | total spend (i.e., \$ sales) | Gasto total, es decir \$ vendidos |
| STORE_ID | data, store lookup | store number | ID tienda (identificador) |
| SUB_CATEGORY | products lookup | sub-category of product | Sub categoría |
| TPR_ONLY | data | temporary price reduction only | Reducción precio en el estante; si TPR vale 1 -> descuento |
| UNITS | data | units sold | Unidades vendidas |
| UPC | data, products lookup | (Universal Product Code) product specific identifier | Identificador de un producto |
| VISITS | data | number of unique purchases (baskets) that included the product | Número de cestas de compra unicas que incluyen al producto |

Escenario de estudio: Los datos contenidos en este archivo incluyen información de ventas y promociones de una muestra de tiendas en Estados Unidos durante 156 semanas, en el que se analizan productos entre 4 categorías: Enjuagues Bucales, Pretzels, Pizza Congelada y Cajas de Cereales.

Dentro de estas cuatro categorías se incluyen los 3 mejores productos de cada una de las 5 marcas en las distintas categorías. (Había una errata en la base de datos, se decía que se incluían los 5 mejores productos de cada una de las 3 marcas principales por categoría)

En total se va a trabajar con 21 variables repartidas entre las tres hojas: Products, Data, Store. Se muestran en la Tabla 10.

Tabla 10: Descripción de variables.

| VARIABLE NAME | TABLE | DESCRIPCIÓN |
|---------------|-------------------|---|
| CITY NAME | Store | Nombre de la ciudad donde se encuentra la tienda. |
| STATE | Store | Estado donde se encuentra la tienda. |
| AVG_BASKETS | Store | Promedio de cestas de la compra vendidas por semana y tienda. |
| BRAND | Data | Marca del producto. |
| CATEGORY | Products | Categoría. |
| ÍTEM | Products | Nombre del producto. |
| DISPLAY | Products | Producto destacado en algún pasillo o en la caja. |
| FEATURE | Data | Producto en el folleto publicitario de la tienda. |
| HHS | Data | Numero de hogares compradores. |
| MSA_CODE | Data | Similar a un código postal en EE. UU. |
| PRICE | Store | Precio de venta. |
| WEEK_DATE | Data | Semana de la compra. |
| STORE_NAME | Store | Nombre de la tienda. |
| SPEND | Data | *Gasto en dólares en la compra. |
| STORE_ID | Data,Store | ID o localizador de la tienda. |
| SUB_CATEGORY | Products | Subcategoría. |
| TPR | Data | Reducción en precio en el estante donde está el producto. |
| UNITS | Data | Unidades vendidas de un producto. |
| UPC | Data, Products | Identificador de producto. |
| VISITS | Data | Numero de cestas de compras únicas que incluyen el producto. |

⇒ La diferencia que se hace entre los campos UNITS, VISITS y HHS es un poco confusa, se explica en detalle en la Tabla 11:

Tabla 11: Diferencia conceptos UNITS, VISITS y HHS.

| 1 | WEEK_DATE | STORE_ID | UPC | UNITS | VISITS | HHS | SPEND | PRICE | BASE_PRICE | TPR_ONLY | FEATURE | DISPLAY |
|----|-----------|----------|------------|-------|--------|-----|--------|-------|------------|----------|---------|---------|
| 2 | 14-ene-09 | 367 | 1111009477 | 13 | 13 | 13 | 18,07 | 1,39 | 1,57 | 1 | 0 | 0 |
| 3 | 14-ene-09 | 367 | 1111009497 | 20 | 18 | 18 | 27,8 | 1,39 | 1,39 | 0 | 0 | 0 |
| 4 | 14-ene-09 | 367 | 1111009507 | 14 | 14 | 14 | 19,32 | 1,38 | 1,38 | 0 | 0 | 0 |
| 5 | 14-ene-09 | 367 | 1111035398 | 4 | 3 | 3 | 14 | 3,5 | 4,49 | 1 | 0 | 0 |
| 6 | 14-ene-09 | 367 | 1111038078 | 3 | 3 | 3 | 7,5 | 2,5 | 2,5 | 0 | 0 | 0 |
| 7 | 14-ene-09 | 367 | 1111038080 | 2 | 2 | 2 | 5,18 | 2,59 | 2,59 | 0 | 0 | 0 |
| 8 | 14-ene-09 | 367 | 1111085319 | 14 | 13 | 13 | 26,32 | 1,88 | 1,88 | 0 | 0 | 0 |
| 9 | 14-ene-09 | 367 | 1111085345 | 29 | 26 | 25 | 54,52 | 1,88 | 1,88 | 0 | 0 | 0 |
| 10 | 14-ene-09 | 367 | 1111085350 | 35 | 27 | 25 | 69,3 | 1,98 | 1,98 | 0 | 0 | 0 |
| 11 | 14-ene-09 | 367 | 1111087395 | 50 | 40 | 40 | 168 | 3,36 | 3,94 | 0 | 0 | 1 |
| 12 | 14-ene-09 | 367 | 1111087396 | 61 | 43 | 42 | 203,13 | 3,33 | 4 | 0 | 0 | 1 |
| 13 | 14-ene-09 | 367 | 1111087398 | 72 | 52 | 49 | 238,32 | 3,31 | 4 | 0 | 0 | 1 |
| 14 | 14-ene-09 | 367 | 1600027527 | 12 | 10 | 10 | 38,28 | 3,19 | 3,19 | 0 | 0 | 0 |
| 15 | 14-ene-09 | 367 | 1600027528 | 31 | 26 | 19 | 142,29 | 4,59 | 4,59 | 0 | 0 | 0 |
| 16 | 14-ene-09 | 367 | 1600027564 | 56 | 48 | 42 | 152,32 | 2,72 | 3,07 | 0 | 1 | 0 |
| 17 | 14-ene-09 | 367 | 2840004768 | 8 | 6 | 6 | 23,12 | 2,89 | 2,89 | 0 | 0 | 0 |
| 18 | 14-ene-09 | 367 | 3000006340 | 19 | 17 | 17 | 38,76 | 2,04 | 2,74 | 1 | 0 | 0 |
| 19 | 14-ene-09 | 367 | 3000006560 | 87 | 76 | 72 | 174 | 2 | 2,59 | 1 | 0 | 0 |
| 20 | 14-ene-09 | 367 | 3000006610 | 48 | 41 | 41 | 96 | 2 | 2,59 | 1 | 0 | 0 |
| 21 | 14-ene-09 | 367 | 3700019521 | 2 | 2 | 2 | 5 | 2,5 | 3,49 | 1 | 0 | 0 |
| 22 | 14-ene-09 | 367 | 3700044982 | 2 | 2 | 2 | 10 | 5 | 5 | 0 | 0 | 0 |
| 23 | 14-ene-09 | 367 | 3800031829 | 14 | 11 | 3 | 43,96 | 3,14 | 3,14 | 0 | 0 | 0 |

Siempre se cumple que: $UNITS \geq VISITS \geq HHS$

En la fila 16 de la Tabla 11, el campo UNITS vale 56, esto quiere decir que se han vendido 56 productos, VISITS vale 48, lo que implica que algunas personas han comprado más de una unidad del producto, ya que VISITS se refiere a cestas de la compra únicas y distintas. Por último, HHS significa que de esas 48 compras distintas del producto 42 correspondían a hogares únicos, por tanto, hubo hogares que en esa misma semana compraron el producto más de una vez.

⇒ Los valores TPR, FEATURE, Y DISPLAY son del tipo Verdadero/ Falso, valen **0** o **1**. En la misma Figura se puede ver que para cada semana el producto puede estar o con una reducción de precio en el estante, o en el folleto publicitario de la tienda o, destacado en la caja o los pasillos.

En resumen, se va a definir qué contiene cada hoja:

- La hoja de PRODUCTS contiene los 58 productos distintos que se van a analizar. Tras la limpieza queda como se ve en la Tabla 12.

El término subcategoría es un poco confuso, como muestra la Tabla 13, ya que por ejemplo la categoría Bag Snacks solo tiene una subcategoría; Pretzels. Por lo que para este no tendría sentido la subcategoría ya que se solapan, pero dentro de los cereales sí que existen tres subcategorías

Caso de Estudio: Breakfast at the Frat

Tabla 12: Hoja PRODUCTS tras limpieza.

| | A | B | C | D | E |
|----|------------|---------------------------|---------------|-----------------------|-----------------------------|
| | UPC | ITEM | BRAND | CATEGORY | SUB_CATEGORY |
| 1 | | | | | |
| 2 | 1111009477 | PL MINI TWIST PRETZELS | PRIVATE LABEL | BAG SNACKS | PRETZELS |
| 3 | 1111009497 | PL PRETZEL STICKS | PRIVATE LABEL | BAG SNACKS | PRETZELS |
| 4 | 1111009507 | PL TWIST PRETZELS | PRIVATE LABEL | BAG SNACKS | PRETZELS |
| 5 | 1111035398 | PL BL MINT ANTSPCTC RINSE | PRIVATE LABEL | ORAL HYGIENE PRODUCTS | MOUTHWASHES (ANTISEPTIC) |
| 6 | 1111038078 | PL BL MINT ANTSPCTC RINSE | PRIVATE LABEL | ORAL HYGIENE PRODUCTS | MOUTHWASHES (ANTISEPTIC) |
| 7 | 1111038080 | PL ANTSPCTC SPG MNT MTHWS | PRIVATE LABEL | ORAL HYGIENE PRODUCTS | MOUTHWASHES (ANTISEPTIC) |
| 8 | 1111085319 | PL HONEY NUT TOASTD OATS | PRIVATE LABEL | COLD CEREAL | ALL FAMILY CEREAL |
| 9 | 1111085345 | PL RAISIN BRAN | PRIVATE LABEL | COLD CEREAL | ADULT CEREAL |
| 10 | 1111085350 | PL BT SZ FRSTD SHRD WHT | PRIVATE LABEL | COLD CEREAL | ALL FAMILY CEREAL |
| 11 | 1111087395 | PL SR CRUST SUPRM PIZZA | PRIVATE LABEL | FROZEN PIZZA | PIZZA/PREMIUM |
| 12 | 1111087396 | PL SR CRUST 3 MEAT PIZZA | PRIVATE LABEL | FROZEN PIZZA | PIZZA/PREMIUM |
| 13 | 1111087398 | PL SR CRUST PEPPRN PIZZA | PRIVATE LABEL | FROZEN PIZZA | PIZZA/PREMIUM |
| 14 | 1600027527 | GM HONEY NUT CHEERIOS | GENERAL MI | COLD CEREAL | ALL FAMILY CEREAL |
| 15 | 1600027528 | GM CHEERIOS | GENERAL MI | COLD CEREAL | ALL FAMILY CEREAL |
| 16 | 1600027564 | GM CHEERIOS | GENERAL MI | COLD CEREAL | ALL FAMILY CEREAL |
| 17 | 2066200530 | NWMN OWN PEPPERONI PIZZA | KING | FROZEN PIZZA | PIZZA/PREMIUM |
| 18 | 2066200531 | NWMN OWN 4 CHEESE PIZZA | KING | FROZEN PIZZA | PIZZA/PREMIUM |
| 19 | 2066200532 | NWMN OWN SUPREME PIZZA | KING | FROZEN PIZZA | PIZZA/PREMIUM |
| 20 | 2840002333 | RLDGLD BRAIDED HONEY WHT | FRITO LAY | BAG SNACKS | PRETZELS |
| 21 | 2840004768 | RLDGLD TINY TWISTS PRZL | FRITO LAY | BAG SNACKS | PRETZELS |
| 22 | 2840004770 | RLDGLD PRETZEL STICKS | FRITO LAY | BAG SNACKS | PRETZELS |
| 23 | 3000006340 | QKER LIFE ORIGINAL | QUAKER | COLD CEREAL | ALL FAMILY CEREAL |
| 24 | 3000006560 | QKER CAP N CRUNCH BERRIES | QUAKER | COLD CEREAL | KIDS CEREAL |
| 25 | 3000006610 | QKER CAP N CRUNCH | QUAKER | COLD CEREAL | KIDS CEREAL |
| 26 | 3120505000 | HMRN CLSC SAUSAGE PIZZA | HOME RUN | FROZEN PIZZA | PIZZA/PREMIUM |
| 27 | 3120505100 | HMRN CLSC SSG PEPP PIZZA | HOME RUN | FROZEN PIZZA | PIZZA/PREMIUM |
| 28 | 3120506000 | HMRN CLSC CHS PIZZA | HOME RUN | FROZEN PIZZA | PIZZA/PREMIUM |
| 29 | 3700019521 | CREST PH WHTG CLN MINT TP | P & G | ORAL HYGIENE PRODUCTS | MOUTHWASH/RINSES AND SPRAYS |
| 30 | 3700031613 | SCOPE ORIG MINT MOUTHWASH | P & G | ORAL HYGIENE PRODUCTS | MOUTHWASH/RINSES AND SPRAYS |
| 31 | 3700044982 | CREST PH CLN MINT RINSE | P & G | ORAL HYGIENE PRODUCTS | MOUTHWASHES (ANTISEPTIC) |
| 32 | 3800031829 | KELL BITE SIZE MINI WHEAT | KELLOGG | COLD CEREAL | ALL FAMILY CEREAL |
| 33 | 3800031838 | KELL FROSTED FLAKES | KELLOGG | COLD CEREAL | KIDS CEREAL |
| 34 | 3800039118 | KELL FROOT LOOPS | KELLOGG | COLD CEREAL | KIDS CEREAL |
| 35 | 4116709428 | ACT MINT A/CAV FLUOR RNS | CHATTEM | ORAL HYGIENE PRODUCTS | MOUTHWASH/RINSES AND SPRAYS |
| 36 | 4116709448 | ACT KIDS BBLGUM FLUOR RNS | CHATTEM | ORAL HYGIENE PRODUCTS | MOUTHWASH/RINSES AND SPRAYS |
| 37 | 4116709565 | ACT RSTR CL SPLSH MINT MW | CHATTEM | ORAL HYGIENE PRODUCTS | MOUTHWASHES (ANTISEPTIC) |
| 38 | 7027312504 | SHURGD PRETZEL RODS | SHULTZ | BAG SNACKS | PRETZELS |

Tabla 13: Categoría y Subcategoría.

| Category | Subcategory |
|--------------|-----------------------------------|
| Bag Snacks | Pretzels |
| Oral Higiene | -Antiseptic -Rinses and sprays |
| Cold Cereal | -Kids -All family -Adults |
| Frozen Pizza | Pizza Premium |

- o La de STORES, tiene las 79 tiendas de esta cadena de supermercados dónde se han realizado las ventas que se han registrado en las 156 semanas. Tras analizar los campos en el Excel, se ve que hay dos tiendas que aparecen duplicadas, se eliminan los duplicados y se tiene que, en realidad, el número de tiendas distintas es 77.

Tabla 14: Hoja STORES base de datos.

| | A | B | C | |
|----|-----------------|----------------------|--------------------------|--------------|
| 2 | STORE_ID | STORE_NAME | ADDRESS_CITY_NAME | ADDRI |
| 24 | 26981 | SPRINGFIELD | SPRINGFIELD | OH |
| 25 | 4503 | ROCKWALL | ROCKWALL | TX |
| 26 | 2513 | HOUSTON | HOUSTON | TX |
| 27 | 19523 | DUNCANVILLE | DUNCANVILLE | TX |
| 28 | 19533 | DENTON | DENTON | TX |
| 29 | 25021 | HAMILTON | HAMILTON | OH |
| 30 | 367 | 15TH & MADISON | COVINGTON | KY |
| 31 | 2505 | DICKINSON VILLAGE | DICKINSON | TX |
| 32 | 6187 | ST. MARYS | SAINT MARYS | OH |
| 33 | 8035 | OVER-THE-RHINE | CINCINNATI | OH |
| 34 | 8041 | LANDEN | MAINEVILLE | OH |
| 35 | 13837 | MAGNOLIA | MAGNOLIA | TX |
| 36 | 15765 | CYPRESS TOWN CENTER | HOUSTON | TX |
| 37 | 17615 | SUGARLAND | SUGAR LAND | TX |
| 38 | 21479 | WOOD FOREST S/C | HOUSTON | TX |
| 39 | 25027 | CINCINNATI | CINCINNATI | OH |
| 40 | 25253 | HIGHWAY 75 | MCKINNEY | TX |
| 41 | 29159 | CARROLLTON | CARROLLTON | TX |
| 42 | 17627 | FLOWER MOUND | FLOWER MOUND | TX |
| 43 | 613 | EAST ALLEN | ALLEN | TX |
| 44 | 2523 | LANDMARK PLACE S/C | GRAND PRAIRIE | TX |
| 45 | 11757 | INDEPENDENCE | INDEPENDENCE | KY |
| 46 | 12011 | SHERMAN | SHERMAN | TX |
| 47 | 21227 | LIBERTY TWP | MIDDLETOWN | OH |
| 48 | 21237 | LEBANON | LEBANON | OH |
| 49 | 23075 | TYLERSVILLE | WEST CHESTER | OH |
| 50 | 23349 | GARLAND | GARLAND | TX |
| 51 | 25229 | CYPRESS | CYPRESS | TX |
| 52 | 25233 | ANTOINE TOWN CENTER | HOUSTON | TX |
| 53 | 26973 | MAINEVILLE | MAINEVILLE | OH |
| 54 | 4245 | LAWRENCEBURG | LAWRENCEBURG | IN |
| 55 | 11967 | NORTHBOROUGH SQ | HOUSTON | TX |
| 56 | 11993 | DALLAS | DALLAS | TX |
| 57 | 17627 | FLOWER MOUND | FLOWER MOUND | TX |
| 58 | 21221 | MIDDLETOWN | MIDDLETOWN | OH |
| 59 | 23055 | WALNUT HILLS/PEEBLES | CINCINNATI | OH |
| 60 | 24991 | HYDE PARK | CINCINNATI | OH |
| 61 | 25261 | CROWLEY | CROWLEY | TX |
| 62 | 2281 | BLUE ASH | BLUE ASH | OH |
| 63 | 2495 | PLANTATION PLAZA | CLUTE | TX |
| 64 | 4503 | ROCKWALL | ROCKWALL | TX |
| 65 | 4521 | PARKWAY SQUARE S/C | COLLEGE STATION | TX |
| 66 | 8263 | THE WOODLANDS | THE WOODLANDS | TX |

Caso de Estudio: Breakfast at the Frat

- o La hoja TRANSACTION contiene todas las compras a lo largo de las 156 semanas desde el 14 de enero de 2009, en total tiene 524.950 registros.

Tabla 15. Desglose de la hoja Excel de productos.

| CATEGORY | SUBCATEGORY | BRAND |
|---------------------|-----------------------------------|--|
| Bag Snacks | Pretzels | 1. Private label 2. Frito lay 3. Shultz 4. MKSL 5. Snyder s |
| Oral Higiene | -Antiseptic -Rinses and sprays | 1. Private label 2. P&G 3. Chattem 4. warner 5. Colgate |
| Cold Cereal | -Kids -All family -Adults | 1. Private Label 2. General MI 3. Quaker 4. Kellogs 5. Post Food |
| Frozen Pizza | Pizza Premium | 1. Private Label 2. King 3. Home run 4. Tombstone 5. Tonys |

Llegados a este punto es posible comprender cómo está estructurada la base de datos y qué se encuentra en cada una de las tres tablas que contiene. Además, se ha prescindido de variables que no aportan datos relevantes para el análisis, se han eliminado duplicados y se ha dado un formato más estético y agradable visualmente al libro de Excel.

4.3 Conexión a la base de datos.

Una vez se ha justificado la elección de la base de datos y se ha definido el escenario de estudio se realiza la conexión de la base de datos de Excel con Tableau. Como se indicó al principio de este capítulo la base de datos primaria es la hoja Transaction Data que se relaciona con las otras dos mediante una unión interna. En la parte inferior de la ventana se puede ver una vista previa de los datos; tipo de campo, nombre, y valores.

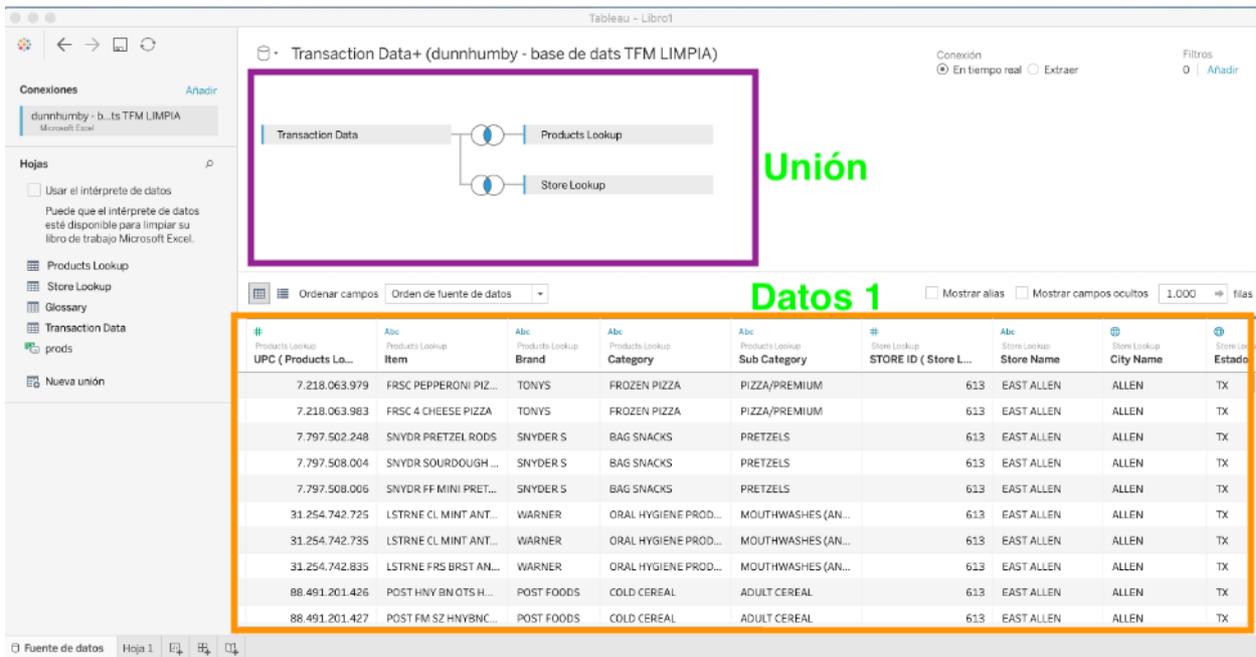


Figura 37: Conexión a la base de datos. Parte 2.

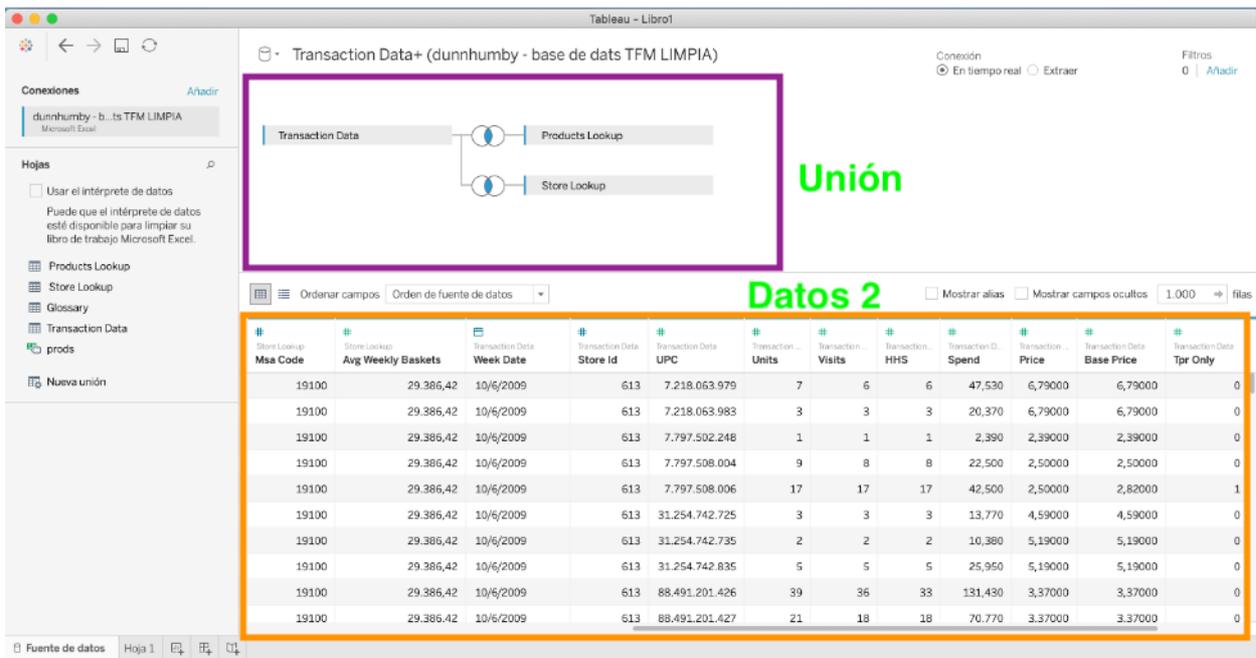


Figura 38: Conexión a la base de datos. Parte 2.

Caso de Estudio: Breakfast at the Frat

En estados unidos, el campo MSA, hace referencia un área estadística Metropolitana (Metropolitan Statistical Area); es una región geográfica con una relativa alta densidad de población y un núcleo poblacional con enlaces económicos a la región que la rodea.

Dentro de los datos geográficos con los que Tableau puede trabajar se encuentra el MSA, pero al realizar la conexión Tableau ha identificado este campo como un campo numérico (#).

Tabla 16: MSA Code I.

| # Store Lookup Msa Code | # Store Lookup Avg Weekly Baskets | Transaction Data Week Date | # Transaction Data Store Id |
|--------------------------------------|--|--------------------------------------|--|
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |

Tabla 17: MSA Code II.

| # | # | Transaction Data | # |
|--------------|--------------------|------------------|------------------|
| Store Lookup | Store Lookup | Week Date | Transaction Data |
| Msa Code | Avg Weekly Baskets | Week Date | Store Id |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |
| 19100 | 29.386,42 | 10/6/2009 | 613 |

Los tres últimos campos, que son: TPR, FEATURE, Y DISPLAY son del tipo verdadero/falso y valen 0 o 1.

Tabla 18: Campos tipo numérico.

| # Transaction Data Tpr Only | # Transaction Data Feature | # Transaction Data Display |
|-----------------------------------|----------------------------------|----------------------------------|
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| 1 | 0 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |

Tableau los ha identificado también como campos de tipo numérico (#), si en esa ventana se cambian a tipo de datos booleano, Tableau no lo detecta y escribe NULO en todos los valores, así que, para poder trabajar con ellos como si fueran Booleanos, la solución que se opta es la de crear un Campo Calculado para cada uno de estos tres conceptos, que guarde el valor como Booleano.

Se hace para los tres campos, y así se crean tres variables que valen verdadero/falso:

Tabla 19: Campos calculados booleanos.

Mostrar alias Mostrar campos ocultos 1.000 ➔ filas

| # Transaction Data Tpr Only | =T F Cálculo TPR* | # Transaction Data Feature | =T F Cálculo FEAUTURE* | # Transaction Data Display | =T F Cálculo DISPLAY* |
|-----------------------------------|-------------------------|----------------------------------|------------------------------|----------------------------------|-----------------------------|
| 0 | Falso | 0 | Falso | 0 | Falso |
| 0 | Falso | 0 | Falso | 1 | Verdadero |
| 0 | Falso | 0 | Falso | 1 | Verdadero |
| 0 | Falso | 0 | Falso | 0 | Falso |
| 0 | Falso | 0 | Falso | 0 | Falso |
| 0 | Falso | 0 | Falso | 0 | Falso |
| 1 | Verdad... | 0 | Falso | 0 | Falso |
| 0 | Falso | 0 | Falso | 0 | Falso |
| 0 | Falso | 0 | Falso | 0 | Falso |
| 0 | Falso | 0 | Falso | 0 | Falso |

Los conceptos TPR FEAUTURE o DISPLAY son independientes, el hecho de que alguno valga 1, no siempre implica un descuento en precio, el descuento solo ocurre cuando Base Price y Price son distintos. Además, estos tres campos pueden valer 1 al mismo tiempo.

Por otro lado, no existe ningún campo con una variable que se llame “Descuento” que indique en caso de que uno de los tres valores indicados anteriormente valga 1, el producto está en oferta. Para solventar este problema se crea otro Campo Calculado que se llame Discount, que compare el precio base del ítem y precio de venta y muestre si está descontado o no.

Además; como la sensibilidad de las ventas y el precio es un factor que interesa, se van a crear otros dos campos calculados más, para poder hacer un análisis más profundo de los descuentos. El primero “descuento (%)”, con el porcentaje de descuento del producto y, el segundo, con los rangos de descuento, se tienen los siguientes grupos:

Caso de Estudio: Breakfast at the Frat

- productos sin descuentos
- productos con menos de un 10 %
- productos entre 10-25 %
- productos entre 25-50 %
- productos con más de un 50 %

50\"', 'END'. At the bottom, there is a status message 'El cálculo es válido.', an 'Aplicar' button, and an 'Aceptar' button."/>

```

If [(%) descuento ]=0 THEN "0"
ELSEIF [(%) descuento ]<0.10 THEN "0-10"
ELSEIF [(%) descuento ]<0.25 Then "10-25"
ELSEIF [(%) descuento ]<0.50 THEN "25-50"
ELSE ">50"
END

```

El cálculo es válido.

Aplicar Aceptar

Figura 39: Clasificación de los descuentos.

Una vez, se tienen todos estos campos creados se está en situación de comenzar el análisis.

4.4 Análisis de los Datos.

En este punto, tras haber examinado en detalle la base de datos y haberle dado un “formato” cómodo, cambiando el nombre a variables que no eran intuitivas y eliminando las que no se van a utilizar, se está preparado para comenzar el análisis.

Dado que la base de datos es muy extensa y tiene gran cantidad de información, en los siguientes subapartados se presentan los análisis realizados desde distintos puntos de vista, los cuales se listan a continuación:

- Análisis de las unidades vendidas por categorías y marcas.
- Análisis del rango de precios.
- Análisis precios y descuentos.
- Análisis de las promociones.
- Análisis de la respuesta del consumidos a descuentos en el precio.
- Análisis de los Supermercados.
- Análisis en función de las fechas de compra. Análisis estacional.
- Análisis geográfico.

4.4.1 Análisis de las unidades vendidas por categorías y marcas.

El hilo de pensamiento que conduce a búsqueda de respuestas a preguntas en cuanto a las categorías y marcas presentes son:

A lo largo de las 156 semanas de registros de ventas, ¿Cómo se distribuyen las unidades vendidas entre las cuatro categorías? ¿Qué porcentaje corresponde a cada una?

Se asigna a cada categoría un color, este código de colores para la categoría va a ser usado en las próximas visualizaciones y Dashboards Tableau permite asignar los colores entre una gran variedad de tonalidades.

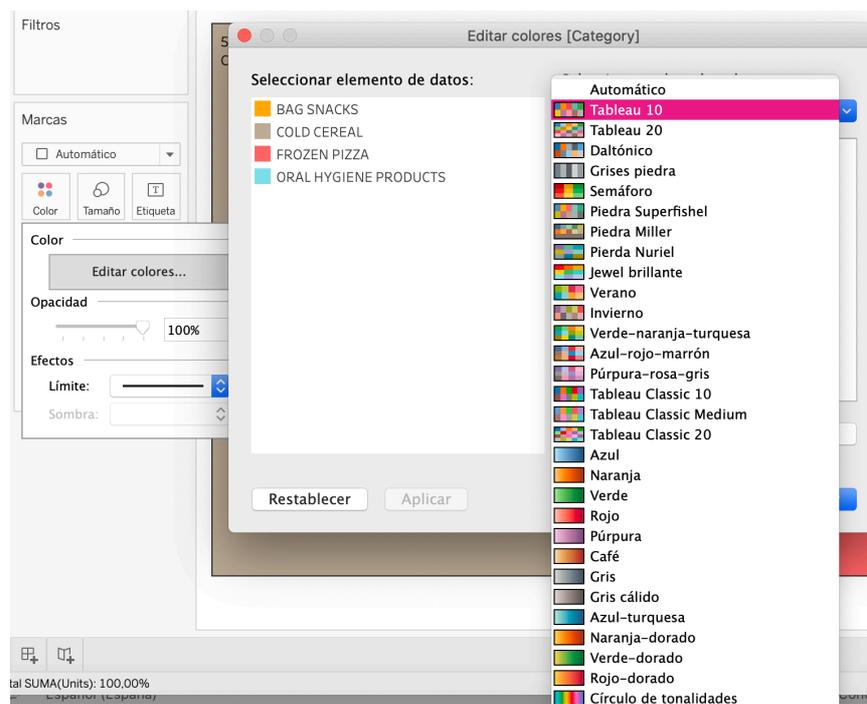


Figura 40: Asignación paleta colores.

Caso de Estudio: Breakfast at the Frat

Para poder determinar cuántas unidades del total han sido vendidas por cada categoría, y cuánto representa en porcentaje dicho valor, se va a realizar un diagrama de árbol, donde el tamaño de cada una de las partes presentes se relaciona con la medida a estudiar, en este caso Units. (Unidades Vendidas de Producto).

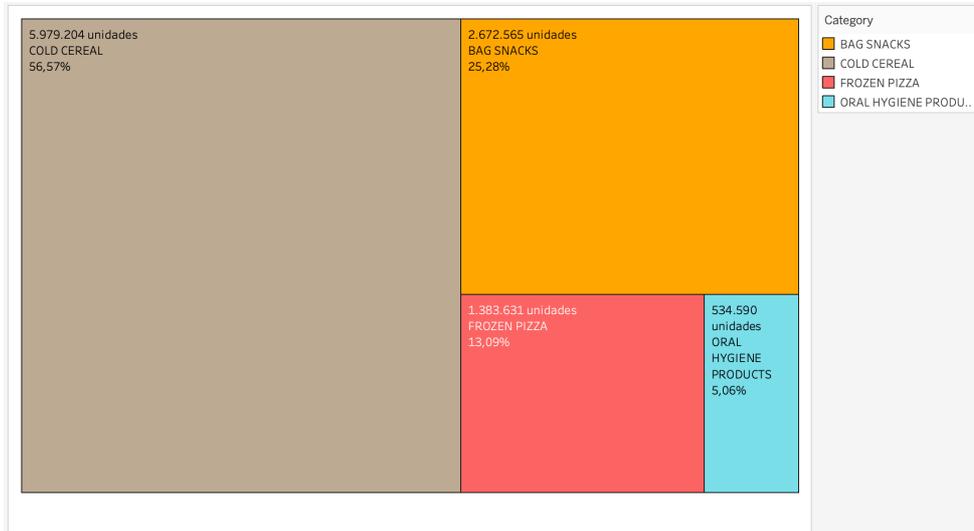


Figura 41: Distribución de unidades vendidas en las categorías.

Más de la mitad de las unidades vendidas corresponden a la venta de productos de la categoría Cold Cereal, del aproximadamente 44% restante la categoría que despunta son los productos de Bag Snacks, y los productos de la categoría de Oral Hygiene son los que menos unidades vendieron.

- La base de datos contiene productos de cuatro categorías. dos de las cuales (Cold Cereal y Oral Hygiene) cuentan con subcategorías, como se ha visto en el apartado 5.3.1. de este documento, ¿Cómo se distribuyen las unidades vendidas si se tienen en cuenta las subcategorías?

Además del gráfico de árbol de la Figura 41, se ha creado un gráfico de barras para indagar en el comportamiento al incluir las categorías en el análisis. Se tiene que los cereales que más se venden son los que pertenecen a la categoría All Family Cereal y de los productos de Oral Hygiene: Antiseptic

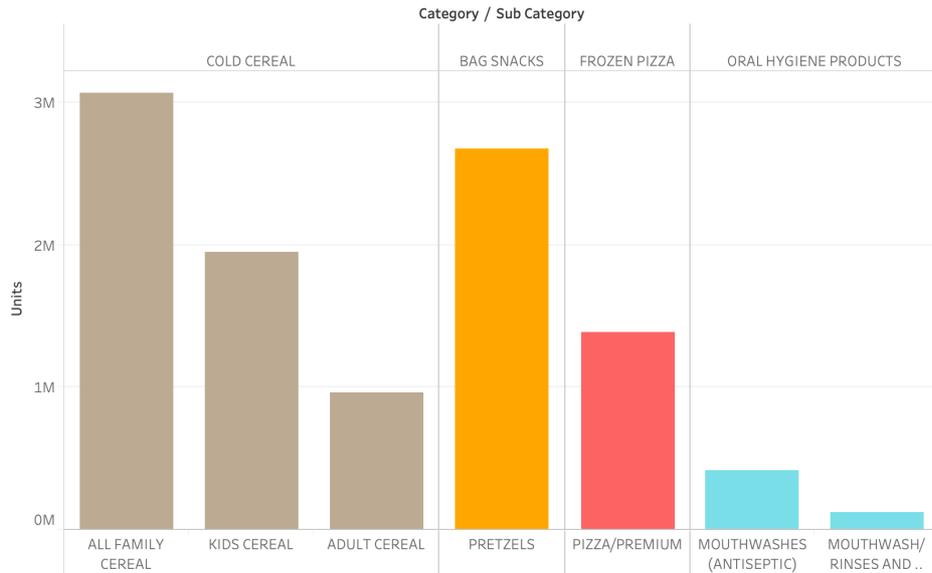


Figura 42: Unidades vendidas por subcategoría.

También se ha visto que la base de datos contiene las principales marcas que venden productos en cada categoría, en la Figura 43 se ve que GENERAL MI, KELLOGG y la Marca blanca son las que más venden. y, Colgate, MKSL, KING y CHATTEN las que menos ventas tienen.

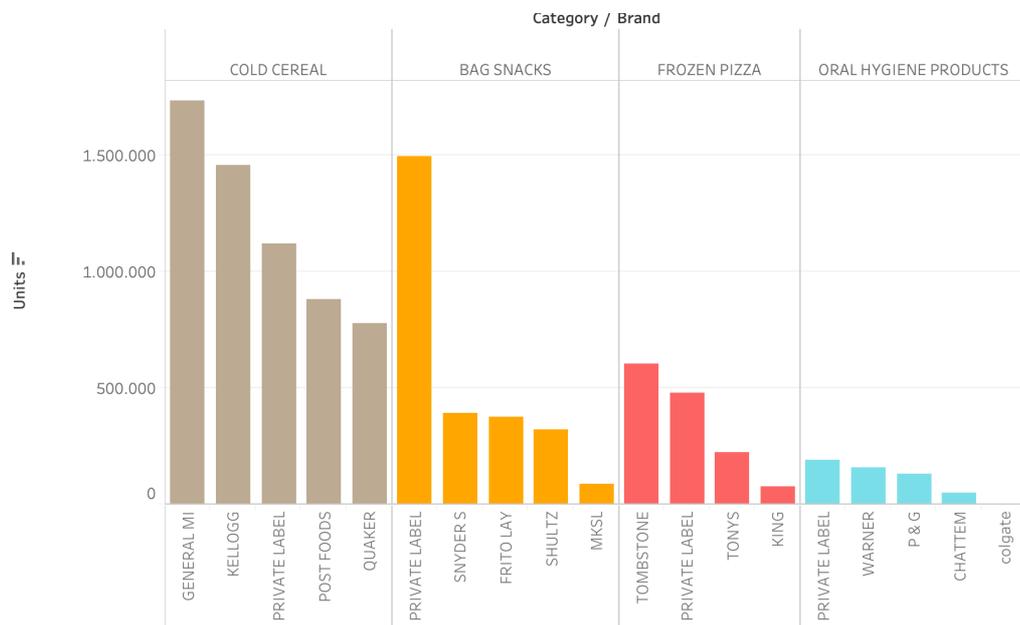


Figura 43: Unidades vendidas según marca y categoría.

Se unen las tres visualizaciones en el Dashboard y se aplican acciones para aportar interactividad y ver cómo se pueden extraer las mejores conclusiones.

Caso de Estudio: Breakfast at the Frat

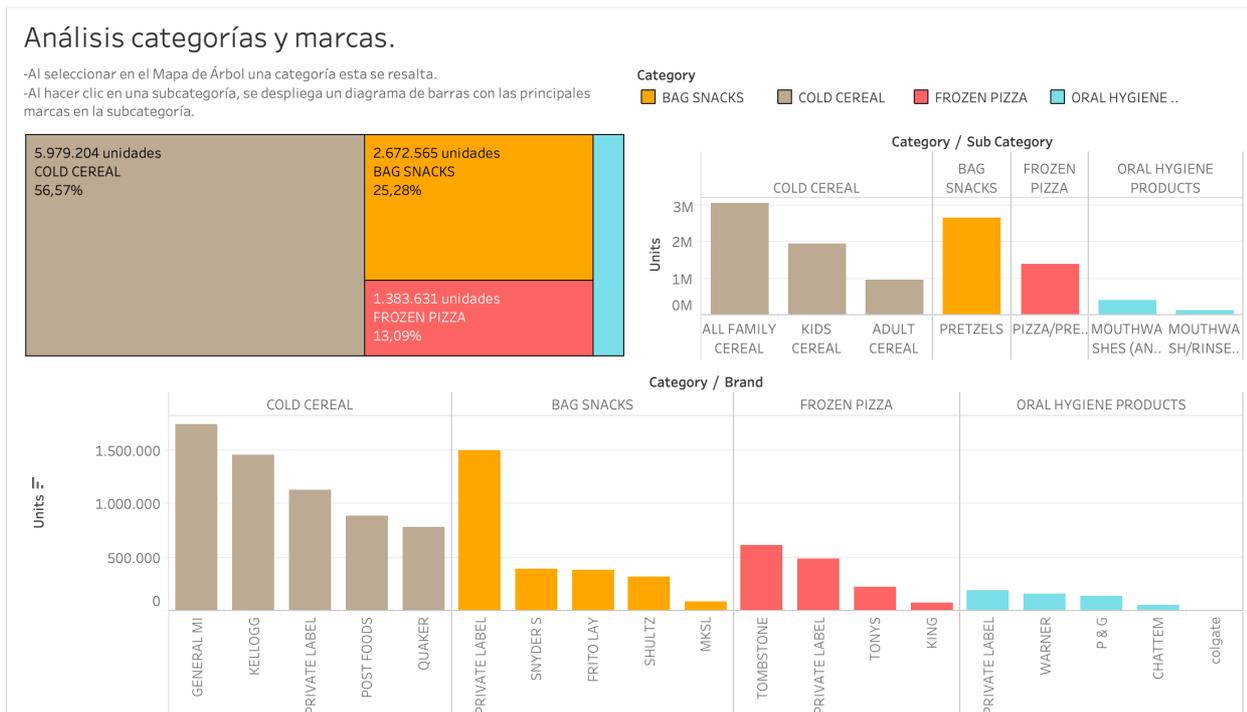


Figura 44: Dashboard. Análisis categorías y marcas I

Al hacer clic en una categoría del gráfico de árbol, ésta se ve resaltada en los otros dos diagramas de barras:

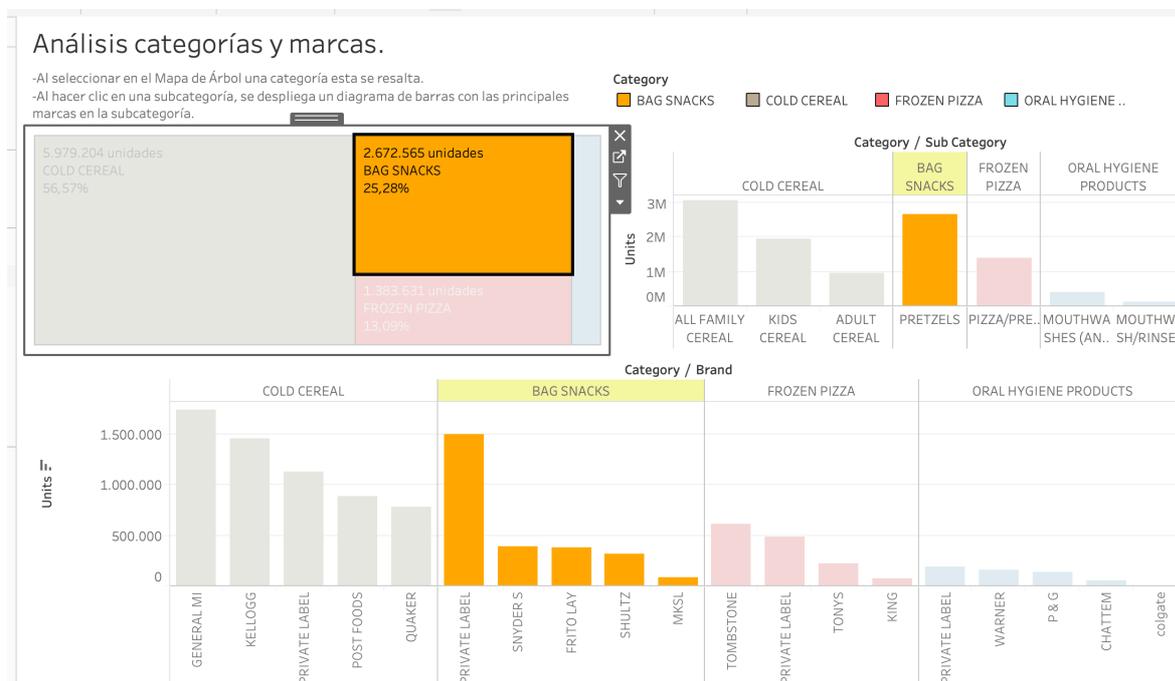


Figura 45: Dashboard Análisis categorías y marcas II

Para analizar las unidades vendidas por marca en una subcategoría, se hace clic en el gráfico de barras que contiene la subcategoría, y este actúa como filtro para ver las unidades vendidas por marca en dentro de la subcategoría.

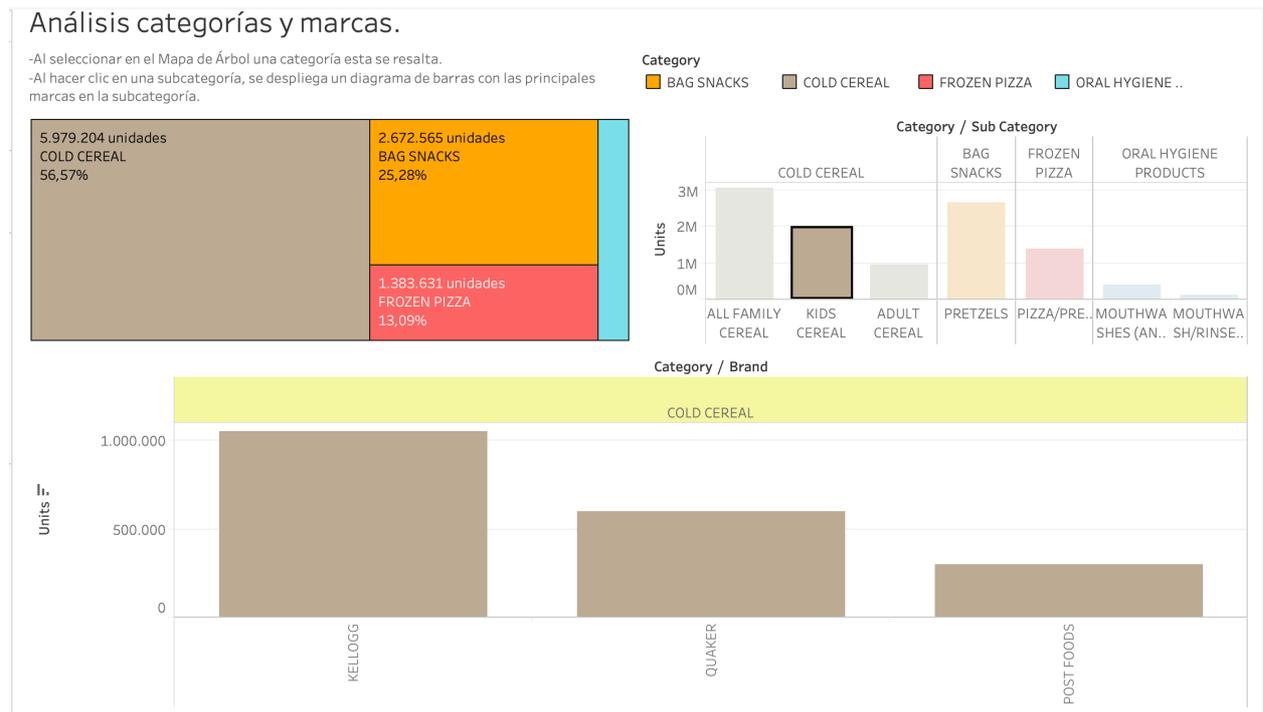


Figura 46: Dashboard Análisis categorías y marcas III

La marca con más unidades vendidas es General MI en la categoría de Cold Cereal y la que menos Colgate, dentro de la categoría Oral Hygiene. Estas barras están anotadas en el Dashboard para facilitar su identificación.

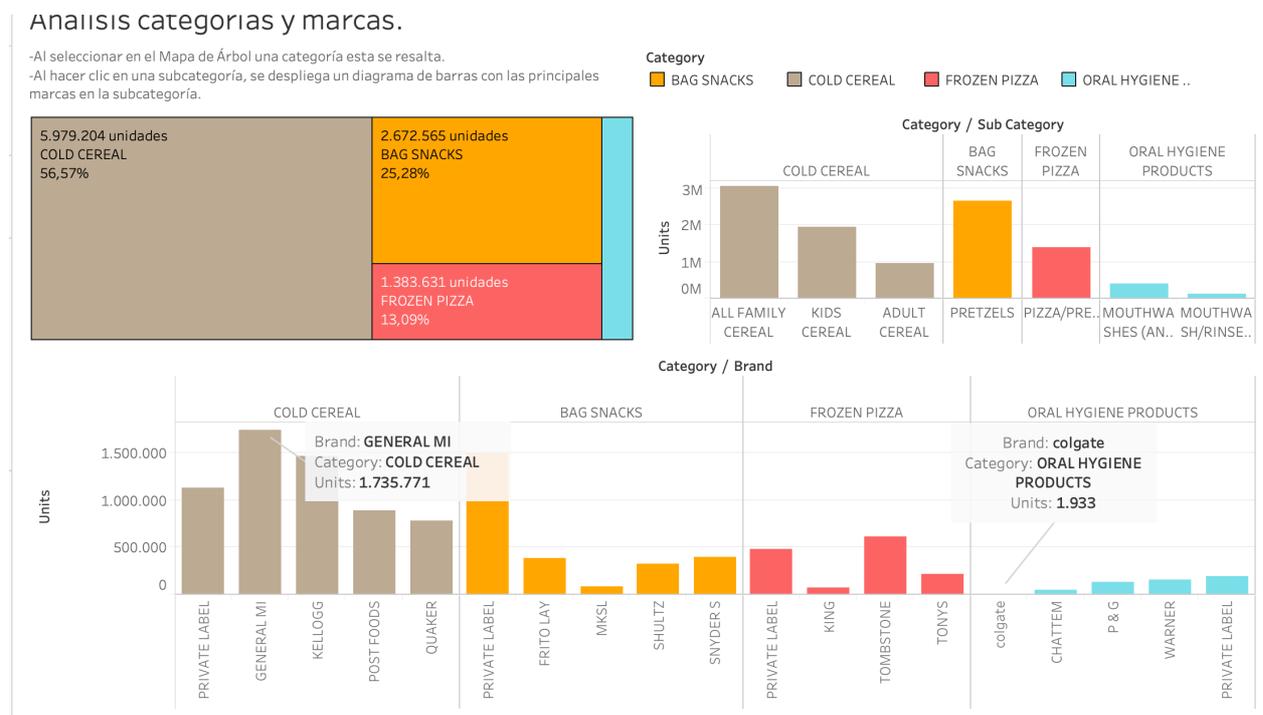


Figura 47: Dashboard Análisis categorías y marcas IV

Caso de Estudio: Breakfast at the Frat

Esta cadena de supermercados cuenta con una marca blanca: Private Label, que vende productos en las cuatro categorías que aparecen en la base de datos. A continuación, se va a analizar el desempeño de la marca blanca dentro de las categorías.

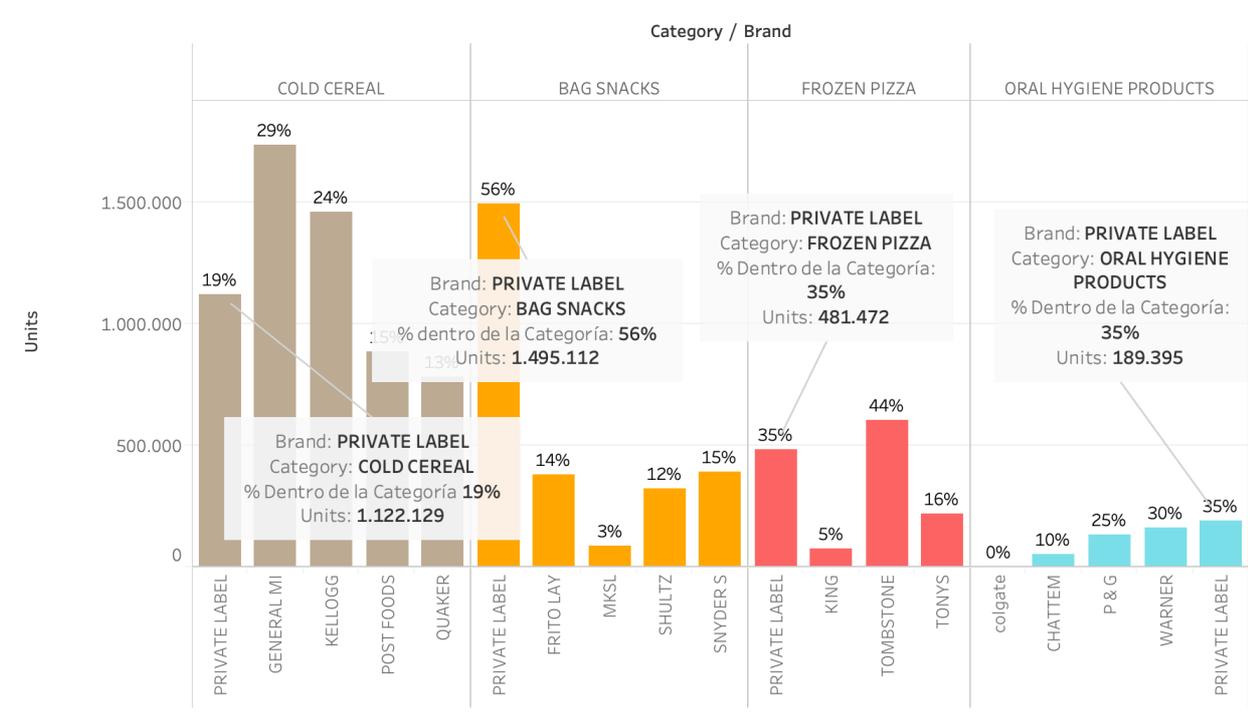


Figura 48: Desempeño marca blanca en las categorías.

Tabla 20: Marca Blanca en categorías.

Esta visualización muestra que la marca blanca es puntera en la categoría de bag snacks y, aunque el mayor número de unidades vendidas en total es dentro de la categoría de Cold Cereal, es ahí donde menor porcentaje de las unidades vendidas corresponde a la marca blanca Private Label.

Tabla 21: Porcentaje ventas de la marca blanca en cada categoría.

| Categoría Marca Blanca | % Dentro de la categoría |
|------------------------|--------------------------|
| Cold Cereal | 19% |
| Bags Snacks | 56% |
| Frozen Pizza | 35% |
| Oral Hygiene | 35% |

Este dato permite tomar decisiones estratégicas como, por ejemplo, reducir el precio de los productos en de su marca blanca Private Label en la categoría de Cold Cereal para beneficiarse del desempeño tan bueno que tiene esta categoría en la cadena de supermercado.

4.4.2 Análisis del rango de precios.

En el Tableau también se pueden visualizar distribuciones, como se ha visto en el Apartado 3.4.4 de este documento. Las distribuciones son de gran utilidad para analizar el rango de precios de los productos en las categorías.

Este análisis se ha llevado a cabo mediante la representación de un diagrama de caja y bigotes, el cual permite conocer cómo se distribuyen los precios. En este tipo de diagramas, la caja se divide por la mediana, es decir la mitad de los valores están arriba y la mitad abajo. La caja también indica los cuartiles inferior y superior, que contienen cada uno un cuarto de los valores. El intervalo de la caja constituye lo que se conoce como el Rango Intercuartil (IQR). Los bigotes se extienden a 1.5 veces el valor IQR (o la extensión máxima de los datos). Cualquier marca más allá de los bigotes son valores atípicos.

Con todo esto, en la representación se puede ver que, en promedio, los productos más caros son de la categoría de las pizzas: Frozen Pizza y los más baratos los snacks de bolsa: Bag Snacks.

En la categoría de Oral Hygiene aparecen dos valores atípicos, que podrían tratarse de alguna errata en la base de datos.

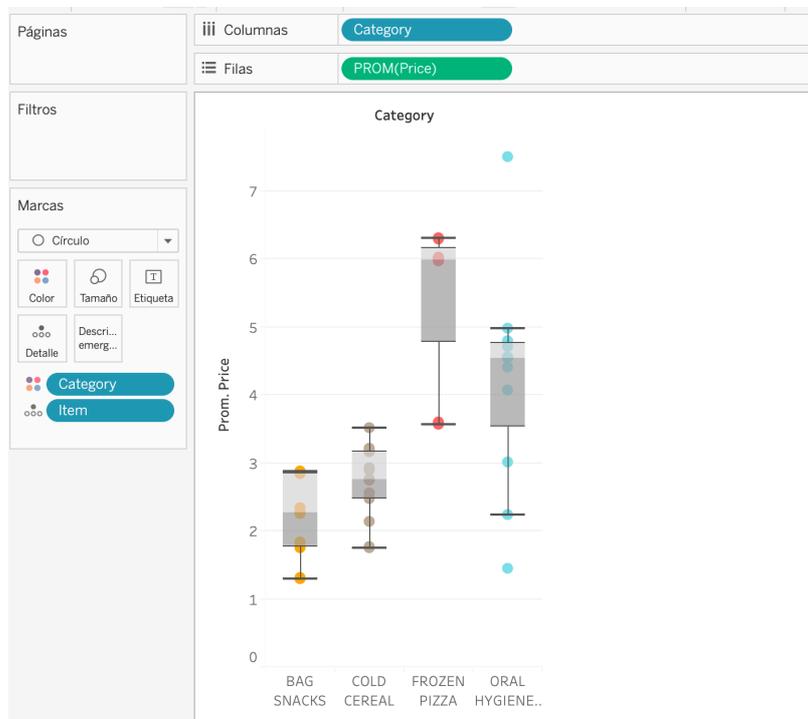


Figura 49: Rango precios entre categorías.

El rango de precios se muestra en la Tabla 22:

Tabla 22: Rango de Precios por Categoría

| Categoría | Rango de precios |
|--------------|------------------|
| Bag Snacks | 1 - 3 € |
| Cold Cereal | 1,5 - 3.5 € |
| Frozen Pizza | 3,5 - 6 € |
| Oral Hygiene | 2 - 5 € |

4.4.3 Análisis precios y descuentos.

Se ha creado un campo calculado que clasifica los descuentos en cinco rangos, se quiere ver como se han distribuido las unidades vendidas en dichos rangos de descuento. En la Figura 50 se ve el número de unidades vendidas en total a lo largo de todas las semanas que se incluyen en la base de datos segmentadas según el descuento que tenía el producto, como es lógico, la mayoría de las unidades vendidas no tenían descuento (57%), alrededor del 20% eran productos con un descuento comprendido entre (10-25%).

Estas barras están agregadas según la suma de unidades, lo cual indica que la mayoría de las unidades no tuvieron descuento ya que un supermercado no puede tener todos sus productos constantemente en oferta.

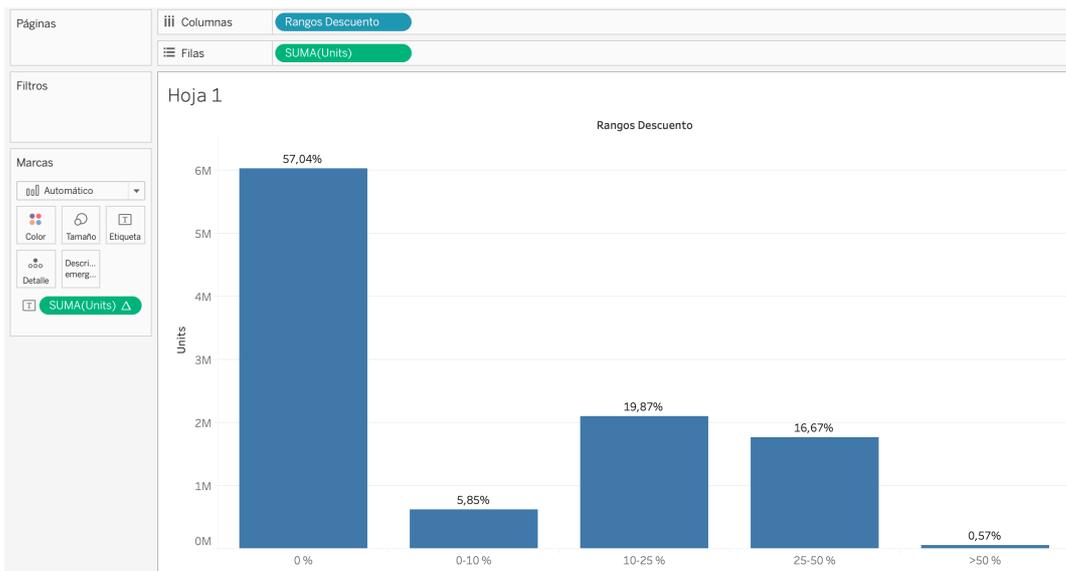


Figura 50: Unidades vendidas según el rango de descuento.

Si se representa el mismo grafico, pero en lugar de con la SUMA de unidades vendidas, con el PROMEDIO de unidades para cada segmento o rango de descuentos, se tiene que en promedio el numero de unidades vendidas aumenta con los descuentos.

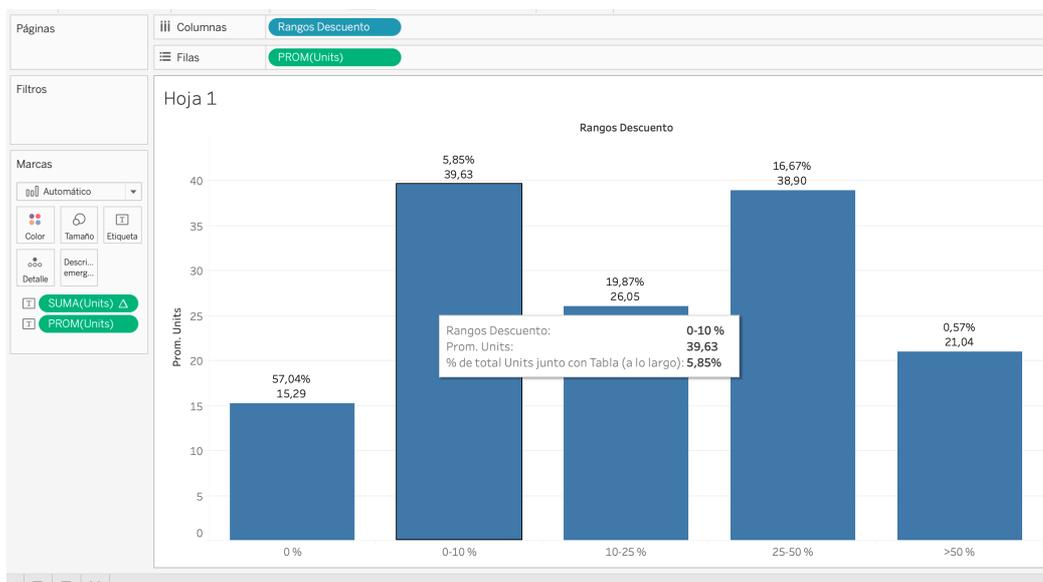


Figura 51: Promedio unidades vendidas según rango de descuento.

Se va a profundizar aún más, ampliando el nivel de detalle del análisis e incluyendo las categorías de los productos, la idea es la misma, ver cómo se han distribuido las unidades vendidas según el descuento que tenga el producto.

Las mismas apreciaciones que se han visto en análisis anteriores se mantienen en este. La categoría con mas ventas en general es Cold Cereal y con menos Oral Hygiene. Pero además se pueden extraer conclusiones nuevas. Por ejemplo, que la mayoría de los productos vendidos en la categoría de Oral Hygiene estaban a la mitad de precio o que, en la categoría de Cold Cereal, más de la mitad de los productos vendidos en promedio tenían más de un 25 % de descuento.

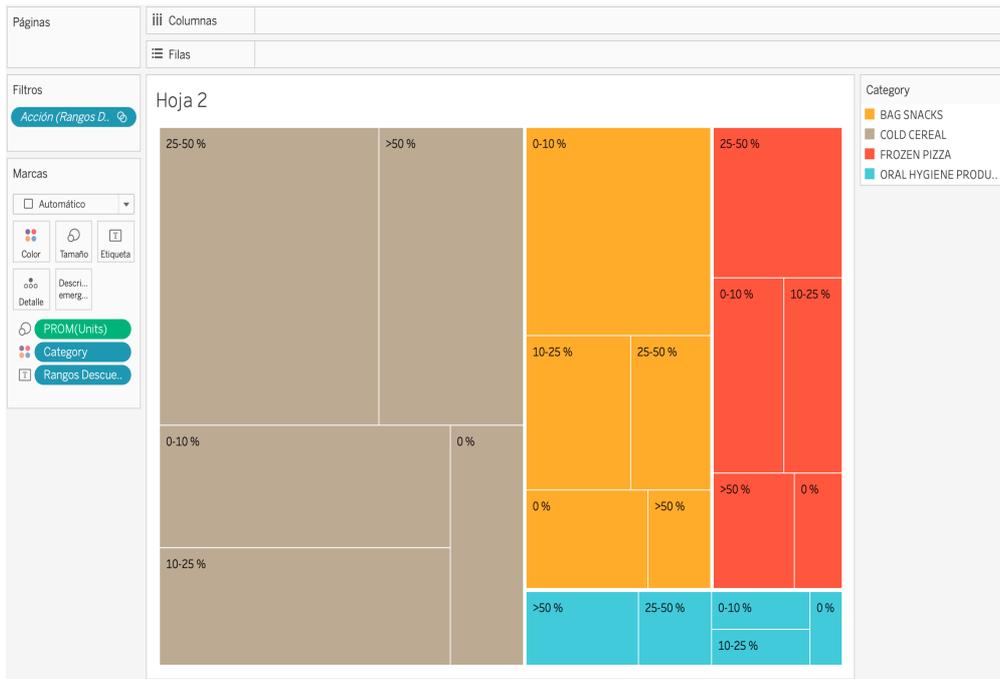


Figura 52: unidades vendidas según descuento y categoría.

Para unir la información que aportan las dos gráficas, se incluyen en un Dashboard y se implementa un filtro en el gráfico de barras, para que, al seleccionar un rango de descuento, el mapa de árbol se filtre para cada categoría y rango seleccionado.

Caso de Estudio: Breakfast at the Frat

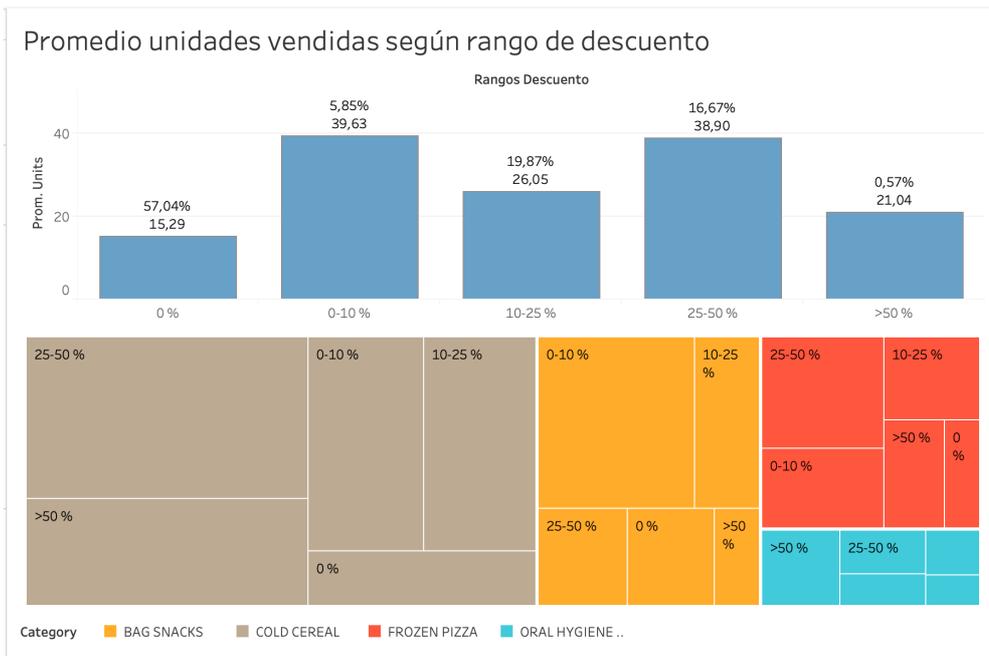


Figura 53: Dashboard promedio unidades vendidas según rango de descuento.

La Figura 53 corresponde al Dashboard antes de seleccionar algún rango de descuento del gráfico de barras y filtrar el mapa de árbol. Sí se hace clic, en el rango de productos sin descuento, el mapa de árbol se filtra para mostrar las cuatro categorías, pero con promedio de unidades vendidas sin descuento.

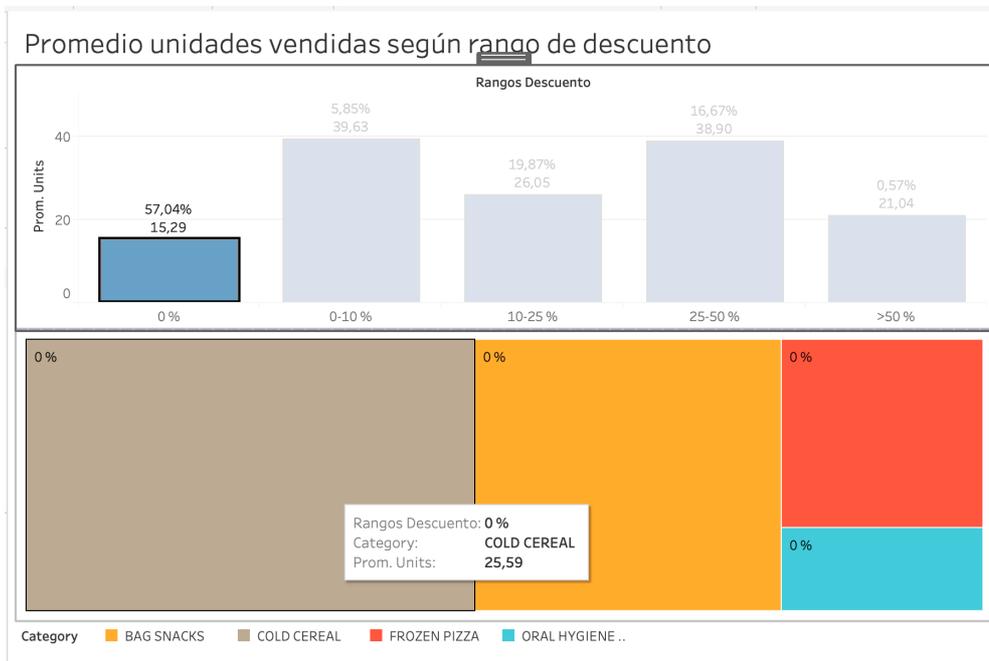


Figura 54: Dashboard filtrado para productos sin descuento.

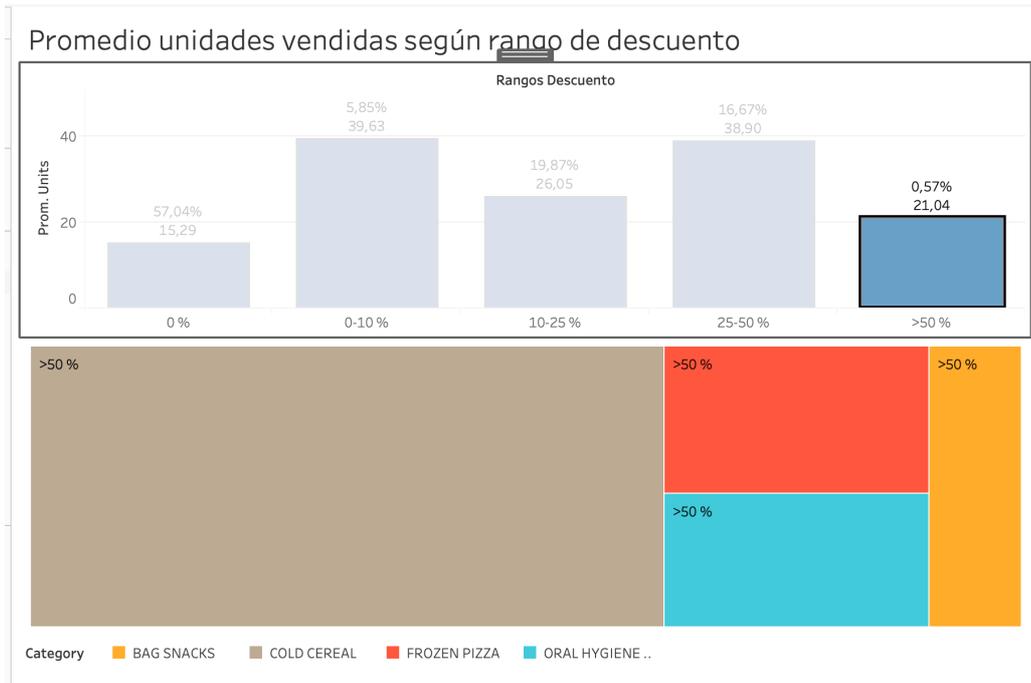


Figura 55: Dashboard filtrado para productos con más de un 50% de descuento.

4.4.4 Análisis de las promociones

En la definición de los campos se ha explicado el significado de las variables que hacen referencia a las promociones; TPR, FEATURE y DISPLAY en el Dashboard que se ilustra en este apartado se va a ver que efecto tienen estos conceptos en el comportamiento de los consumidores. Como recordatorio, se definen a continuación cada uno de estos conceptos:

- TPR: Consiste en reducir el precio de un producto que está colocado en estante o pasillos del supermercado durante un determinado periodo de tiempo. Los clientes que se guían por este parámetro no tienen fidelidad con marcas, ya que su criterio de compra va cambiando en función de que producto esté con reducción de precio. El principal hándicap es encontrar la manera de utilizar este parámetro de manera que se maximicen las ganancias. El uso efectivo de este concepto puede ayudar a disparar las ventas, con poco impacto en cuanto a pérdidas porque dura un tiempo limitado.
- Feature: Consiste en mostrar el artículo en el folleto o revista de la tienda como artículo destacado, esto puede resultar atractivo al consumidor y resultar en un aumento de las ventas para ese producto. No tiene porque implicar un descuento en el precio, aunque lo más normal es que incluya descuento.
- Display: Colocación estratégica de un producto de manera destacada, ya sea en una estantería o en la caja o escaparate. Es lo que hacen los “Visual Merchandisers”. De esta manera consiguen crear una atracción mental al producto e influenciar en la compra de este. Igual que ocurre con Feature, lo normal es que vaya acompañado de descuento, pero no siempre es así.

La Figura 56 muestra el promedio de unidades vendidas según estos campos valgan verdadero o falso, se ve que el mayor efecto en el número de productos vendidos ocurre cuando están activas las promociones DISPLAY y FEATURE.

Caso de Estudio: Breakfast at the Frat

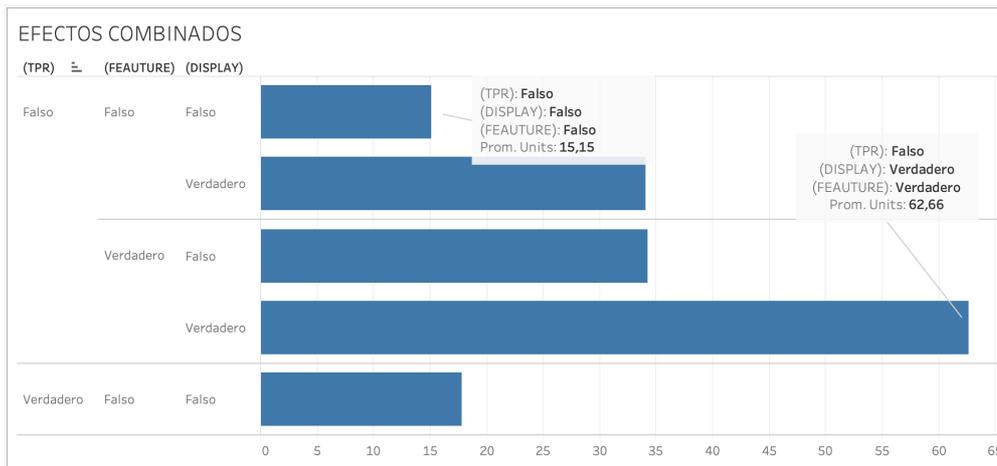


Figura 56: Efectos combinados de los tres efectos.

En la Figura 57 se ve que el TPR es el que menos influencia tiene en el número de unidades vendidas y que lo que más afecta es la combinación de destacar el producto tanto en el folleto como en los pasillos o escaparates. Además; se puede incluir un filtro de categorías, ya el comportamiento que se ve en esta Figura es general, sin distinguir a qué categoría pertenezca el producto. Pero dicho comportamiento puede variar al cambiar de una categoría a otra. Adicionalmente al efecto combinado de estos tres conceptos, se va a crear un Dashboard en el cual se tenga el efecto por separado de cada uno y además se puedan filtrar todas las visualizaciones por categoría.

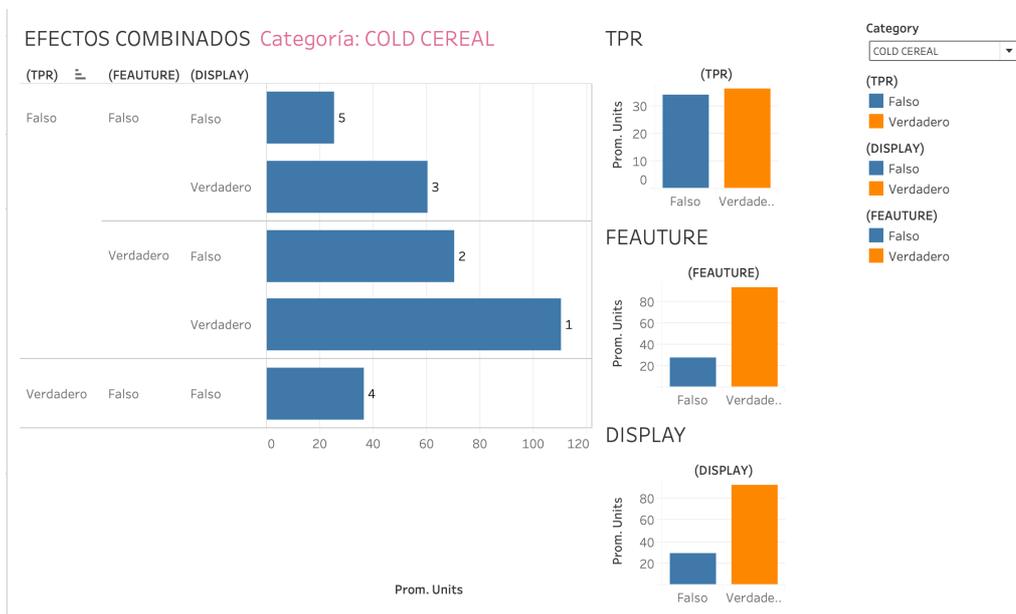


Figura 57: Dashboard efecto TPR, DISPLAY, FEATURE.

Interactuando con el filtro, se tiene que el comportamiento es el mismo en todas las categorías en cuanto al efecto positivo de FEATURE+DISPLAY. Por otro lado, el menor número de unidades vendidas ocurre cuando, o bien no hay ninguna promoción, o cuando la única promoción activa es TPR.

Con todo esto se puede afirmar que el hecho de utilizar únicamente la reducción de precio temporal no conlleva a un aumento significativo del número de unidades vendidas.

4.4.5 Análisis respuesta consumidor a descuentos en precio

La mayoría del tiempo los productos se encuentran con su precio original, es por ello que, el análisis en cuanto al dinero gastado debe hacerse con valores promedio. Se ve que, en general, el impacto del descuento en un supermercado es muy alto. En concreto, un 63% de su facturación es debida a los descuentos, lo que significa que el efecto de la reducción de precio es beneficioso para el supermercado.

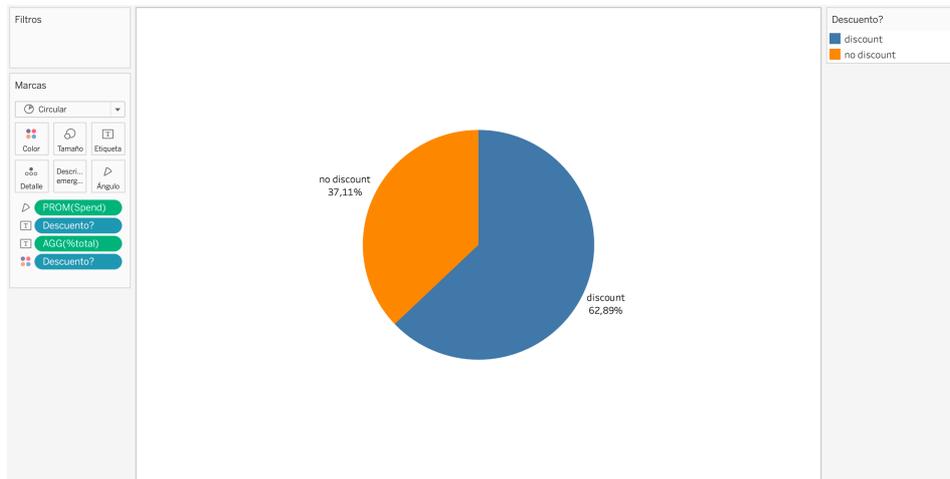


Figura 58: Efecto Descuento.

Para ver cómo es este comportamiento dentro de cada categoría se va a realizar este mismo gráfico según la categoría que corresponda. Se aprecia que, entre las cuatro categorías contempladas, en todas existe un aumento de la cantidad gastada por el consumidor cuando el artículo está en oferta.

La categoría donde mayor impacto tiene el descuento es Frozen Pizza, donde un 70% del gasto ocurre en productos con descuento. Por otro lado, donde menor impacto tiene este descuento es en bag snacks, que además son los productos que tienen un menor precio en promedio por lo que el descuento no es tan atractivo como en las pizzas congeladas.



Figura 59: Efecto Descuento en cada categoría.

4.4.6 Análisis de los Supermercados

En este análisis se busca saber cuáles son los supermercados que más han facturado de entre los más de 70 supermercados del listado. Para ello, se crea un diagrama de barras y se filtra de modo que sólo muestre las 10 tiendas con mayor suma de SPEND.

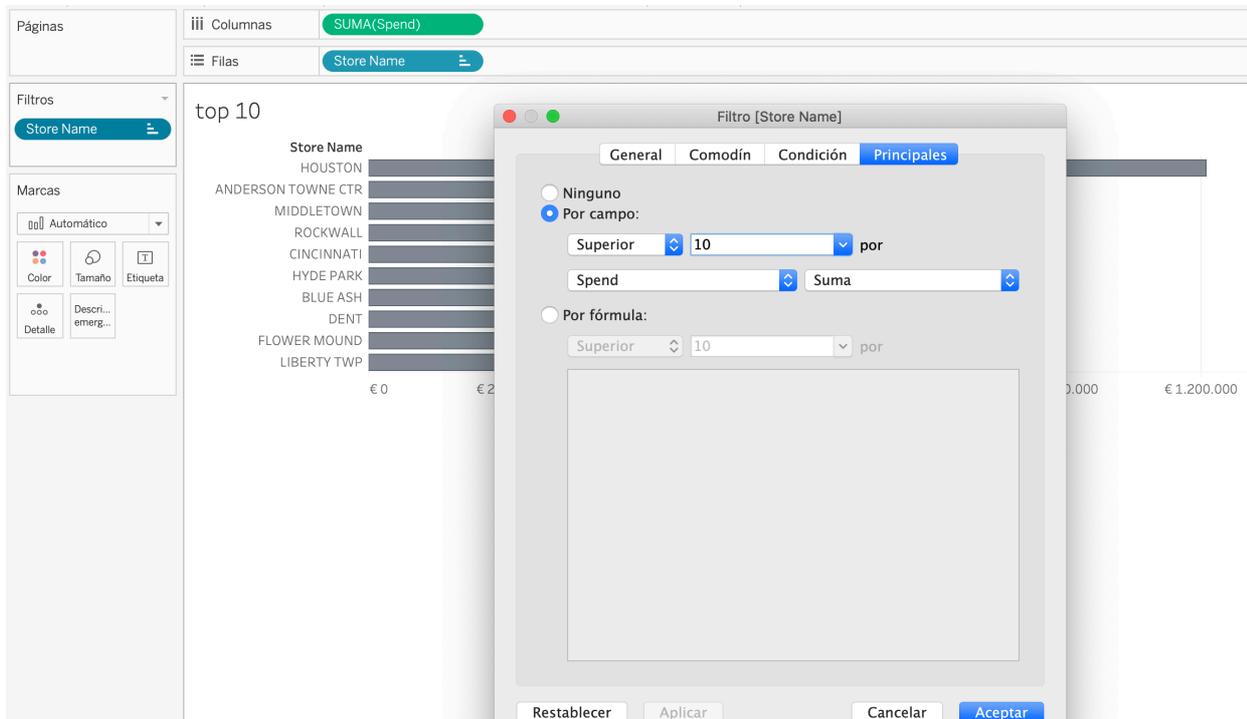


Figura 60: Filtro por campo en un gráfico.

Además, se quiere que al seleccionar una de las diez tiendas se pueda apreciar los 10 elementos más vendidos en cada una de ellas, se pone un título interactivo con el nombre de la tienda seleccionada.

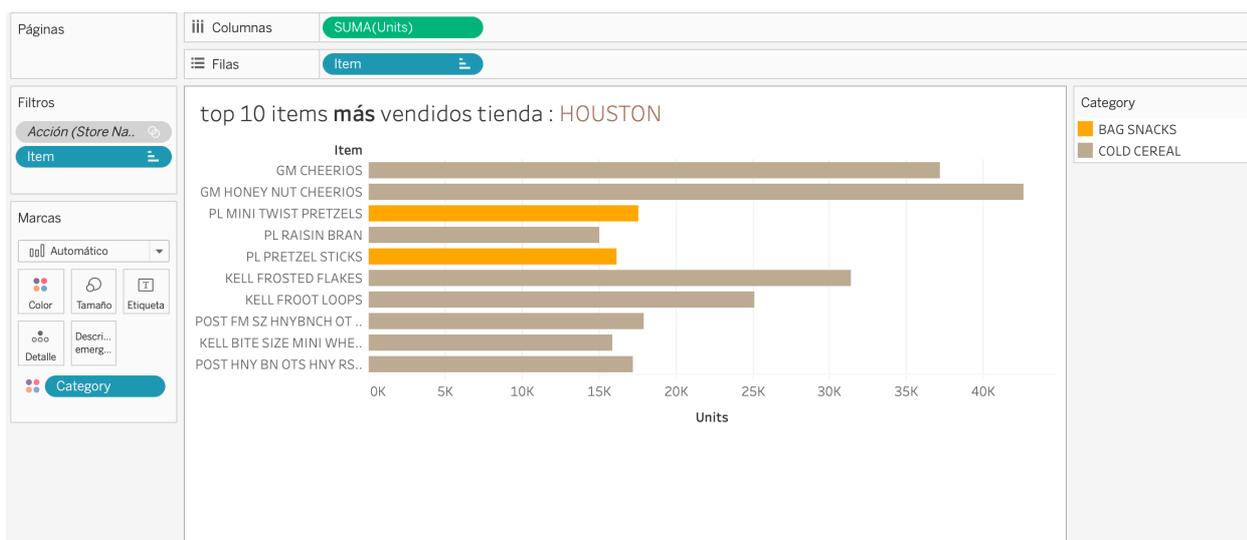


Figura 61: Top 10 Ítems vendidos por tienda

La última gráfica antes de montar el Dashboard es una serie temporal donde se relacionan las unidades vendidas de un producto seleccionado de la Figura 61 y la tienda seleccionada. Además, se ha añadido la opción de mostrar el pronóstico esperado, y la línea de tendencia.



Figura 62: Análisis por ítem: Tendencia y pronóstico I.



Figura 63: Análisis por ítem: Tendencia y pronóstico II.

Caso de Estudio: Breakfast at the Frat

De entre las diez tiendas que más facturan, se obtiene fácilmente que el ítem más vendido coincide para todas y es GM CHERRIOS, de la categoría COLD CEREAL. Al interactuar en el Dashboard se observa que el comportamiento en cuando a la tendencia en unidades vendidas y el pronóstico es creciente.

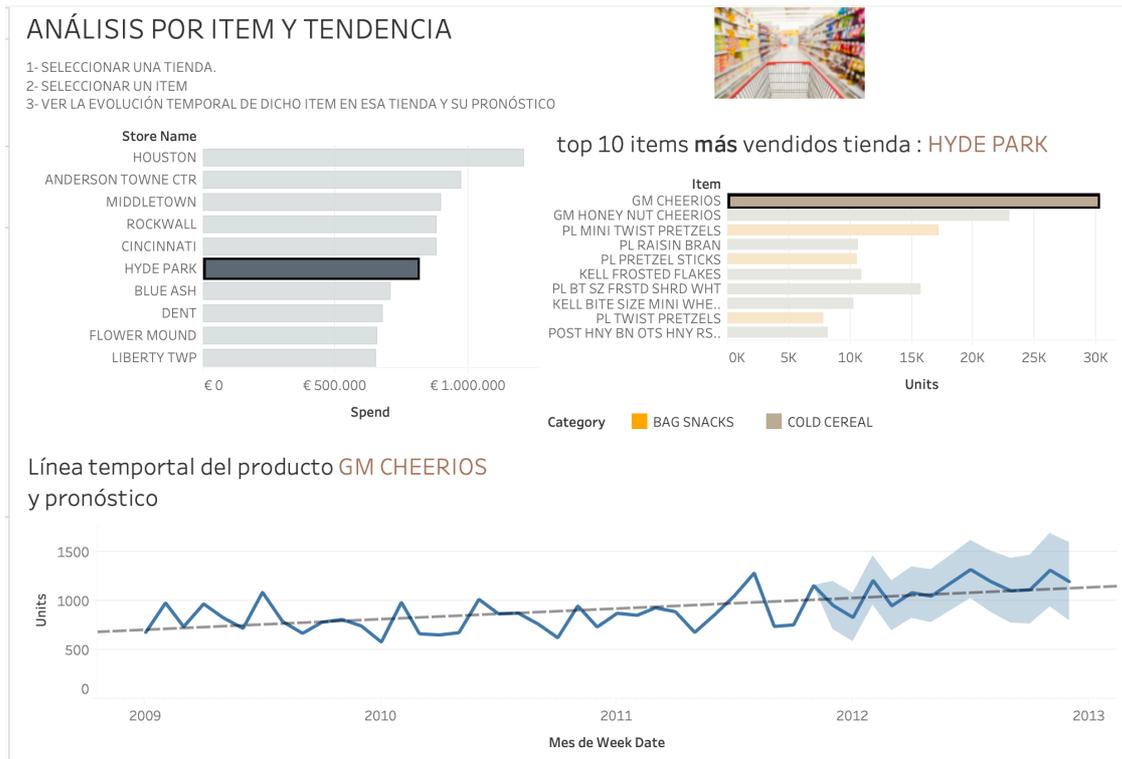


Figura 64: Interacción con el Dashboard.

Se quiere hacer el mismo análisis, en este caso para los productos que menos unidades han vendido en estas mismas 10 tiendas:

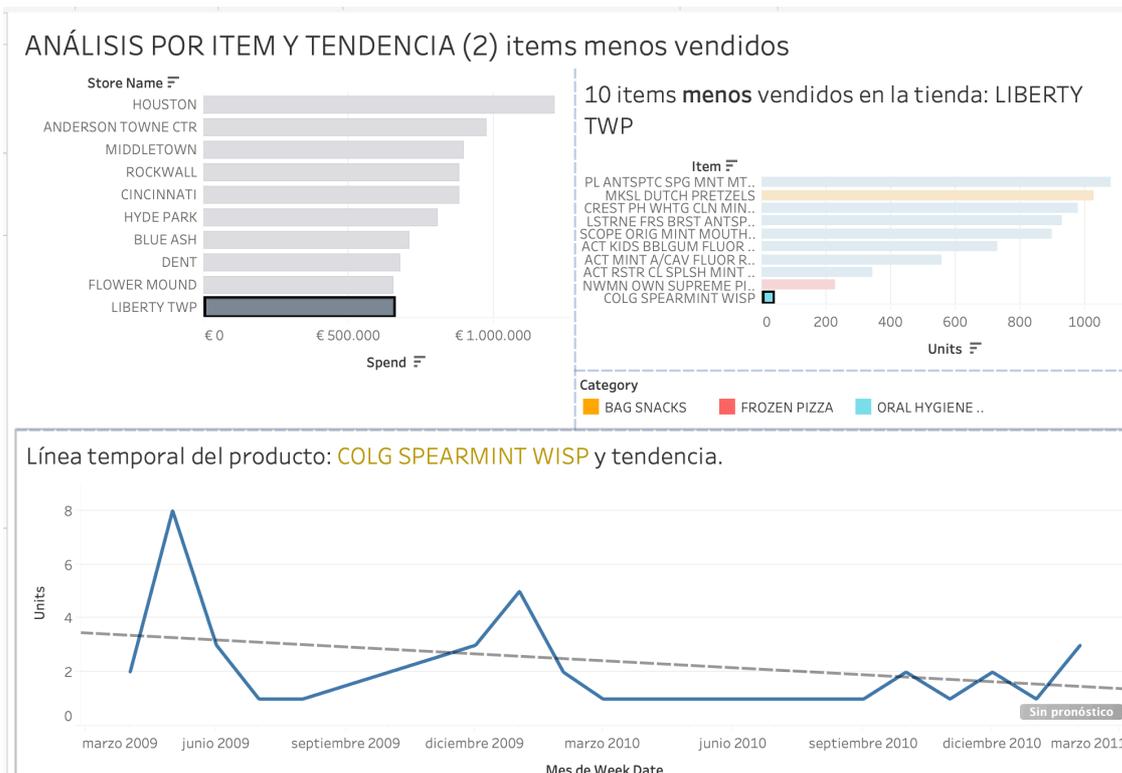


Figura 65: Análisis ítems menos vendidos, tendencia y pronóstico.

4.4.7 Análisis en función de las fechas de compra. Análisis estacional.

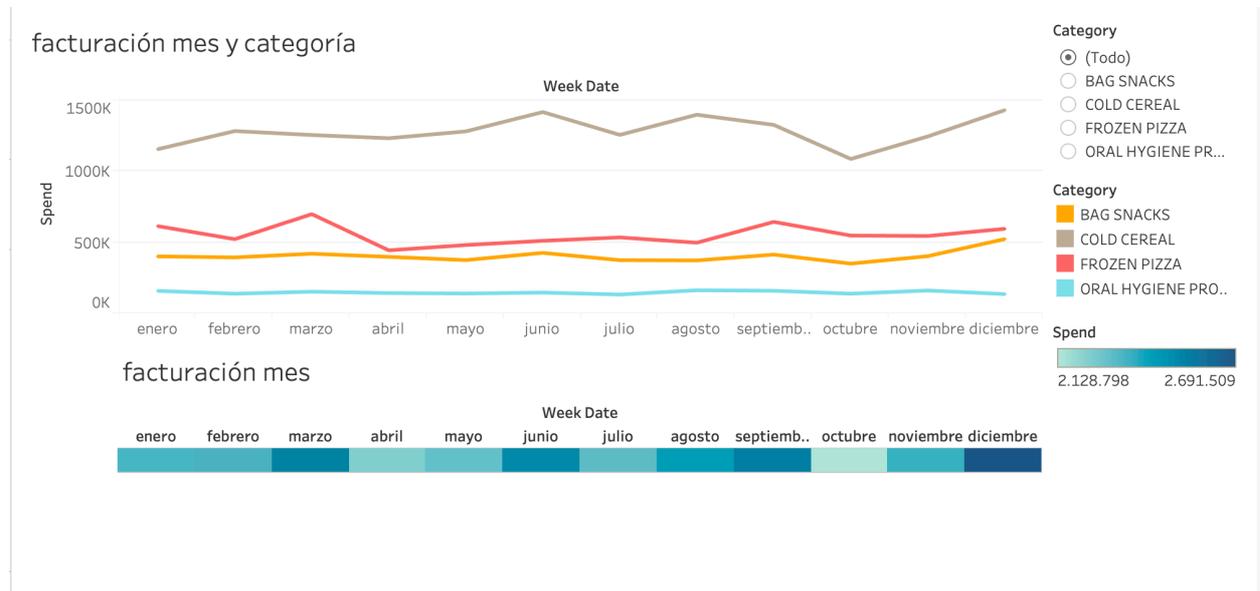


Figura 66: Dashboard facturación mes y categoría.

La estacionalidad varía en cada categoría, en general, sin aplicar un filtro a categoría, los meses con más ingresos por ventas son marzo, junio, septiembre y diciembre.

Si se selecciona una categoría en el filtro que hay a la derecha del Dashboard, el mapa de color se modifica para la categoría elegida. En este análisis se ven los mismos patrones que ya se han identificado previamente, como que la categoría con mas euros gastados es Cold Cereal y la que menos Oral Higiene, además, entre las cuatro categorías que se tienen, los mejores meses son: marzo, junio y diciembre. El mes de abril, tanto en la categoría de pizzas como es las otras, es de los meses con menores ventas

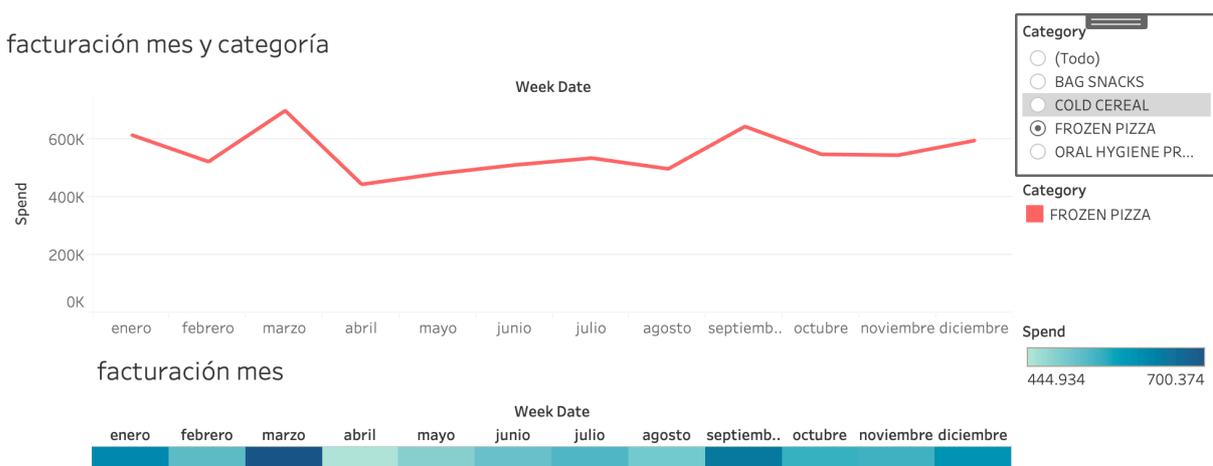


Figura 67: Dashboard facturación mes y categoría con filtro.

4.4.8 Análisis geográfico.

Es interesante entender y conocer como se distribuyen el número de tiendas que se contemplan en la base de datos en el mapa de EE. UU, existen tiendas en cuatro estados, Texas, Indiana, Kentucky y Ohio.

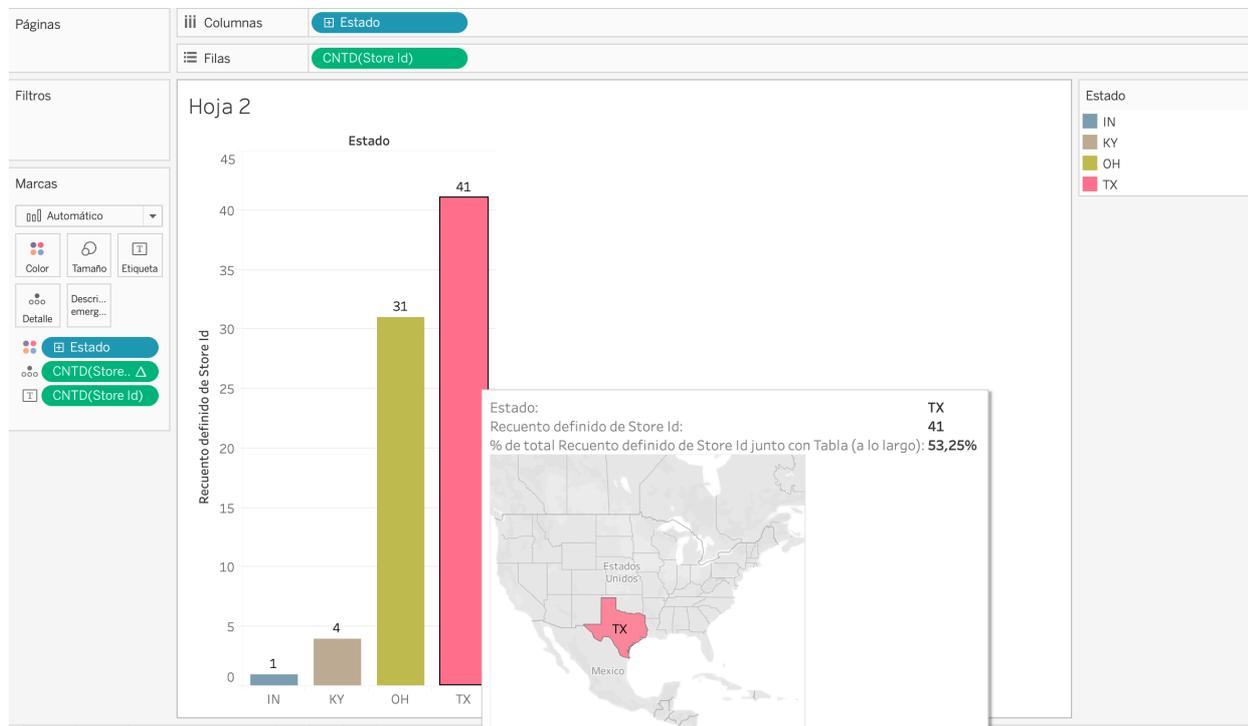


Figura 68: Gráfico tiendas por estado.

Se tiene que, alrededor del 93% de las tiendas están ubicadas en los estados de Ohio y Texas, mientras que las restantes se encuentran en Indiana y Kentucky. Cuando se hace clic en una de las barras de la Figura 68, se ve el porcentaje que representa dicho estado en el total.

Por otro lado, se crea la Tabla 23, donde aparecen todas las ciudades y el número de tiendas en cada una. En general, hay una o dos tiendas por ciudad, aunque existe el caso de ciudades que tienen 8 y 9 tiendas, el título se ha hecho interactivo para que, al aplicar un filtro según el estado, este indique a que estado pertenecen las ciudades que aparecen y el número de ciudades que se están mostrando.

En cada fila, el círculo azul indica cuantas tiendas hay por ciudad.

Tabla 23: Listado número de tiendas en cada ciudad.



Por último, también se va a utilizar en el Dashboard una gráfica de barras que indique el top 10 de ciudades que más han facturado y al estado que pertenecen:

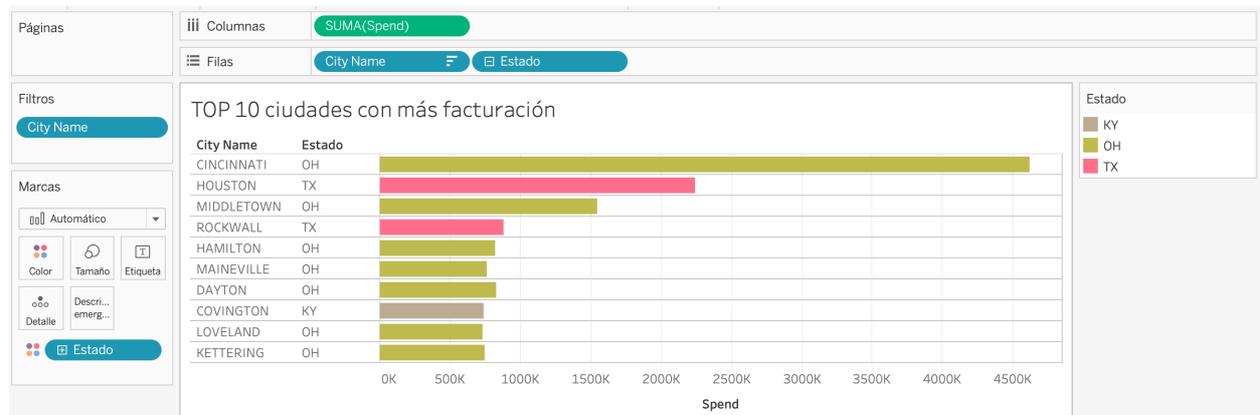


Figura 69: Top 10 ciudades mayor facturación.

Con todo esto, se tiene el Dashboard que se muestra en la Figura 70. Para realizar este análisis, al hacer clic en una de las barras se filtra el gráfico de abajo y se resalta en el Top 10, en caso de que en el estado haya algunas tiendas que pertenezcan a este ranking.

Caso de Estudio: Breakfast at the Frat

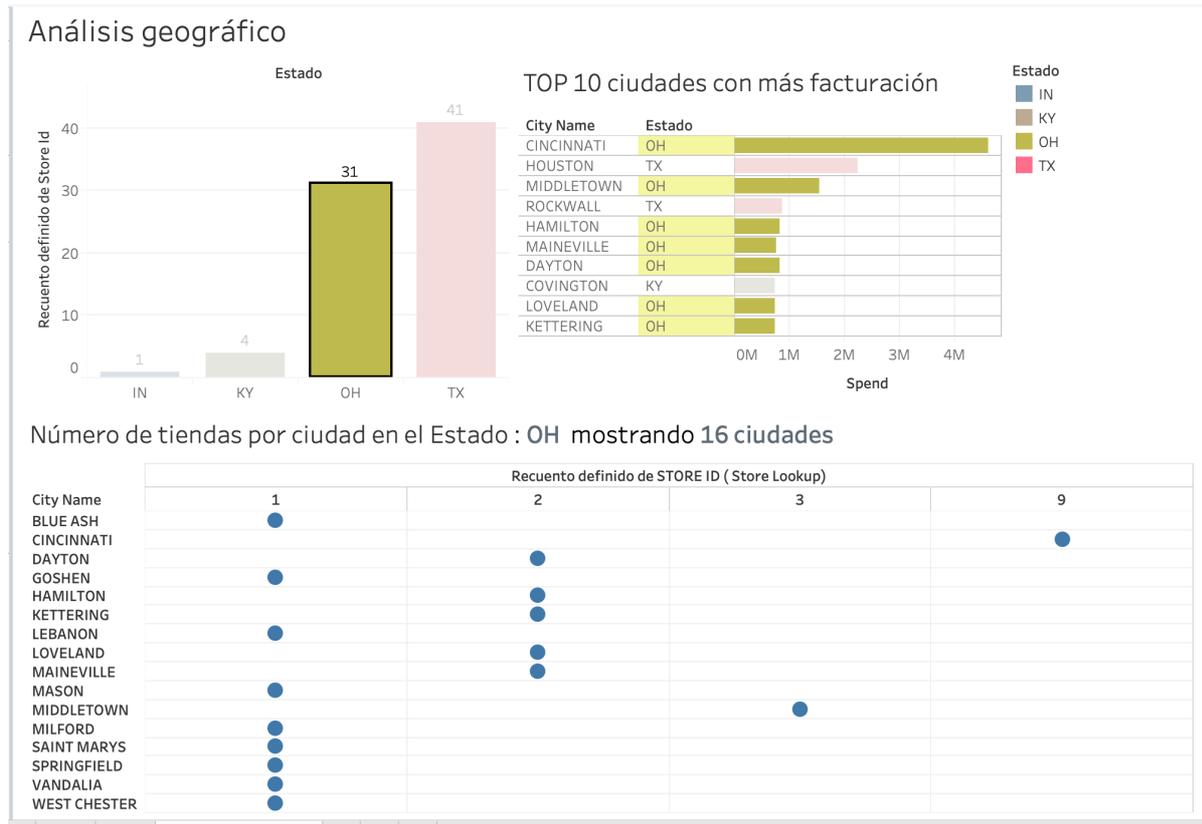


Figura 70: Dashboard análisis geográfico con filtro.

Los estados con un mayor número de tiendas son Texas y Ohio, y dentro de Texas; Houston cuenta con 8 tiendas mientras que, en Ohio, Cincinnati tiene 9 supermercados.

Tabla 24: Ciudades con más Supermercados

| Nombre Ciudad | Nº supermercados |
|---------------|------------------|
| Cincinnati | 9 |
| Houston | 8 |

De toda la base de datos las tres ciudades que más han facturado en el periodo estudiado son Cincinnati, Houston y Middletown.

Tabla 25: Ciudades con más Facturación.

| Nombre Ciudad | Facturación |
|---------------|-------------|
| Cincinnati | 4M € |
| Houston | 2,5 M€ |
| Middletown | 1,5 M€ |

5 CONCLUSIONES

Tras la realización de este Trabajo de Fin de Master, se han llegado a las siguientes conclusiones:

La importancia que está tomando el campo del Business Intelligence es todos los sectores industriales y compañías es innegable. Además, con la globalización y la gran cantidad de datos que se manejan, todas las empresas están destinando recursos al estudio de estos datos para mejorar sus estrategias, obtener más beneficios y entender a sus clientes.

La elección de Tableau frente a las otras dos herramientas mencionadas en el **Capítulo 1** se justifica debido a su gran perspectiva de futuro, precio reducido para estudiantes, recursos disponibles y simplicidad y estética en sus cuadros de mando. (Objetivos Particulares 1 y 2).

Como se ha visto en el **Capítulo 2**, el alcance de este software es muy amplio y sirve en cualquier sector, Tableau se conecta de manera sencilla a archivos en distintos formatos y participa en todas las etapas contempladas en el ciclo del análisis.

En concreto, en el **Capítulo 3** de este documento se ha utilizado como herramienta la matriz BCG para validar el software (Objetivo Particular 3), y se ha comprobado que es perfectamente válido para situar las categorías de productos y definir la estrategia según si un producto es considerado perro, vaca, estrella o interrogante.

En cuanto a las conclusiones extraídas del caso de estudio correspondiente al **Capítulo 4**, se demuestra que, partiendo de la gran base de datos sin procesar, ha sido posible descubrir y sacar conclusiones sobre cómo se comportaban tanto las categorías de productos, como las tiendas o incluso los ítems en particular. Prácticamente se puede llevar a cabo cualquier tipo de análisis y representarlo en Dashboards interactivos de forma que el usuario final pueda interactuar los con resultados y tomar decisiones estratégicas.

Dentro del caso de estudio de la base de datos del Food Retail, se ha indagado en diversos ámbitos: desde la ubicación de las tiendas, el efecto de la fecha de compra, el rango de precios etc.

Se tiene que aproximadamente el 60% de las ventas que se producen en esta cadena de supermercados corresponden a la categoría de Cereales. La marca blanca tiene el mayor número de ventas en los Snacks de bolsa, y el menor en productos que higiene oral, que además son los productos menos vendidos de toda la base de datos.

El efecto del descuento es más señalado en las Pizzas congeladas que en las otras categorías, esto se puede deber a que su rango de precios es el más caro dentro de las cuatro categorías estudiadas, con lo cual un descuento en Pizza resulta al consumidor más atractivo que un descuento en los Snacks de bolsa, cuyo rango de precio oscila alrededor de un euro.

En el transcurso de las 156 semanas de la base de datos, se aprecia un comportamiento cíclico en los meses con mejores ventas, identificándose estos como marzo, junio y diciembre. Este análisis da consistencia a las afirmaciones realizadas en cuanto a categorías con mayores ventas (Cereales y Snacks de bolsa).

También se han analizado las ciudades con mayor facturación: Cincinnati y Houston, que pertenecen a los estados de Ohio y Texas respectivamente. Centrando el análisis a nivel de cada supermercado en particular, se han clasificado las mejores tiendas y, dentro de cada una los ítems más vendidos, siendo posible ver su evolución temporal y el pronóstico.

Además, se ha podido evaluar el efecto que tienen en el consumidor los descuentos en los productos. Se ha descubierto que lo que más incita al comprador es que el producto esté destacado ya sea en la circular de la tienda o en los estantes de los pasillos.

Conclusiones

Estos insights correspondientes al caso de estudio se alinean con el Objetivo Particular 4.

Con todo esto, se afirma que este Trabajo de Fin de Máster da respuesta a los cuatro Objetivos Particulares Marcados en el inicio del documento.

Como línea para desarrollo futuro de lo elaborado en este documento, este trabajo puede verse ampliado realizando el mismo estudio, pero utilizando otro de los softwares que se mencionaron en el **Capítulo 1**, con el objetivo de poder comparar los resultados con los que se han obtenido con Tableau y ver si esas herramientas aportan más beneficios que Tableau.

Así mismo, para el caso de estudio "Breakfast at the Frat" se propone realizar el análisis eliminando los productos de higiene oral, ya que su desempeño no es bueno, y así poder enfocarse en el aumento de ventas en las otras categorías.

También se propone un análisis de las ubicaciones donde solo existe una tienda, para ver si interesa abrir alguna otra en ese estado.

6 BIBLIOGRAFÍA

- Arribas, I. (27 de enero de 2019). *Vandalytic*. Recuperado el 2020, de <https://vandalytic.com/power-bi-vs-qlik-vs-tableau-cual-es-el-mejor-software-de-analisis-de-datos/>
- Attardi, M. (2016). *Análisis del diseño de visualización interactiva de información*. Universitat Politècnica de València.
- Barrionuevo, B. M. (s.f.). *El Business Intelligence en las pymes. Herramienta Power BI*. Universidad politécnica de Cartagena.
- Dataworld*. (s.f.). Recuperado el febrero de 2020, de <https://data.world/missdataviz/superstore-sales-eu>
- Dunnhumby. (s.f.). *Real-world data to put your theory into practice* . Obtenido de <https://www.dunnhumby.com/careers/engineering/sourcefiles>
- Gartner. (2019). *Magic Quadrant for Analytics and Business Intelligence Platforms*. Obtenido de <https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb>
- Google trends*. (s.f.). Recuperado el febrero de 2020, de <https://trends.google.com/trends/explore?date=all&q=%2Fm%2F016jq3>
- google trends, data visualization*. (s.f.). Recuperado el febrero de 2020, de <https://trends.google.com/trends/explore?date=all&q=data%20visualization>
- Insights, G. P. (s.f.). *Comparing Microsoft, Qlik, Tableau* . Recuperado el 2020, de <https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms/compare/microsoft-vs-qlik-vs-tableau>
- Jordá, P. (2016). *La utilización de herramientas Business Intelligence para la gestión empresarial*. UNIVERSITAT POLITÈCNICA DE VALÈNCIA.
- Mañez, R. (s.f.). *rubenmanez*. Recuperado el febrero de 2020, de <https://rubenmanez.com/>
- Martínez, V. B. (2018). *Implantación de un sistema de Business Intelligence en empresa de cursos a distancia*.
- Monzón, L. (9 de agosto de 2018). *Arbentia*. Recuperado el 2020, de <https://www.arbentia.com/comparativa-de-herramientas-de-business-intelligence-microsoft-power-bi/>
- Ruedas, J. H. (2015). *Análisis de los sistemas de ayuda a la toma de decisiones y su aplicación a los procesos empresariales* . Universidad de Cantabria.
- TRENDS, Q. E. (s.f.). *Escuela de negocios y dirección* . Recuperado el febrero de 2020, de <https://br.escueladenegociosydireccion.com/business/marketing-digital/google-trends/>
- udemy*. (s.f.). Recuperado el octubre de 2019, de <https://www.udemy.com/course/visualizacion-de-datos-con-tableau/#instructor-1>

Bibliografia

7 ANEXO: LA INTERFAZ DE TABLEAU.

En este apartado se muestran funcionalidades de tableau que se han utilizado para la realización del análisis de la base de datos.

7.1 Tipos de uniones

- ⇒ **Interior:** Sólo se mantendrán los registros que coincidan con la condición de unión tanto de la tabla de la izquierda como de la tabla de la derecha. En el siguiente ejemplo, solo las tres filas coincidentes se mantienen en los resultados:
- ⇒ **Izquierda:** Se mantendrán todos los registros de la tabla de la izquierda. Los registros coincidentes de la tabla de la derecha tendrán valores en la tabla resultante, mientras que los registros no coincidentes contendrán valores nulos (*null*) para todos los campos de la tabla de la derecha. En el siguiente ejemplo, las cinco filas de la tabla izquierda se mantienen con resultados *NULL* para los valores de la derecha que no coinciden.
- ⇒ **Derecha:** Se conservarán todos los registros de la tabla de la derecha. Los registros coincidentes de la tabla de la izquierda generarán valores, mientras que los registros no coincidentes contendrán valores nulos (*null*) para todos los campos de la tabla de la izquierda. No todas las fuentes de datos admiten una unión correcta. Si no es compatible, la opción estará deshabilitada. En el siguiente ejemplo, las cinco filas de la tabla derecha se mantienen con resultados *null* para los valores izquierdos que no coinciden.
- ⇒ **Exterior completo:** Todos los registros de las tablas en ambos lados se mantendrán. Los registros coincidentes tendrán valores de izquierda y derecha. Los registros no coincidentes tendrán valores *null* donde no se encontró el registro coincidente izquierdo o derecho. No todas las fuentes de datos admiten una combinación externa completa. Si no es compatible, la opción estará deshabilitada. En el siguiente ejemplo, todas las filas se mantienen desde ambos lados con valores *null* donde no se encontraron coincidencias.

7.2 Tipos de campos en Tableau

7.2.1 Medidas y Dimensiones

Los campos de datos se crean a partir de las columnas de la fuente de datos. A cada campo se le asigna automáticamente un tipo de datos (por ejemplo, entero, cadena, fecha, etc.) y un rol: o bien dimensión discreta o medida continua (más habitual), o bien dimensión continua o medida discreta (menos habitual).

- Las *dimensiones* contienen valores cualitativos (por ejemplo, nombres, fechas o datos geográficos). Puede utilizar las dimensiones para categorizar, segmentar y revelar los detalles de los datos. Las dimensiones afectan al nivel de detalle de la vista.
- Las *medidas* contienen valores numéricos cuantitativos que se pueden medir. Las medidas se pueden agregar. Al arrastrar una medida a la vista, Tableau aplica una agregación a esa medida (de forma predeterminada).

La combinación de medidas y dimensiones utilizadas en la vista definen el nivel de detalle.

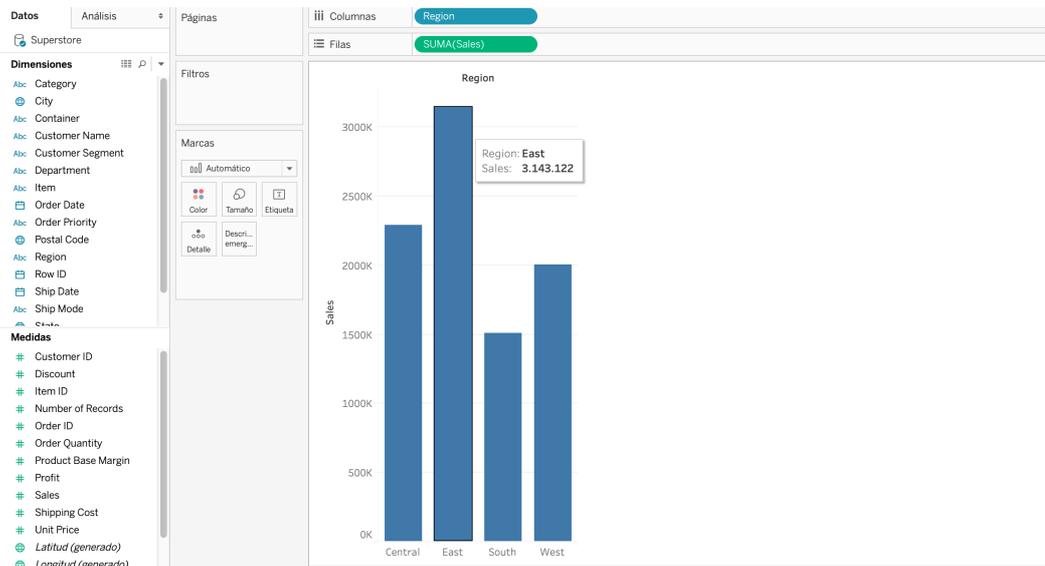


Figura 71: Representación en Tableau de medidas y dimensiones

En este ejemplo se ve que las Ventas es una medida, y que está siendo agregada como una suma: **Suma(sales)** la Región es una dimensión que tiene cuatro valores (Central, Este, Sur y Oeste). Cuando se usa una dimensión en la visualización, lo que hace es categorizar o segmentar a la medida, por ello en vez de ver el total de todas las sumas globales, se ven región a región.

7.2.2 Campos Discretos y Continuos.

El hecho de que un campo sea discreto o continuo determina como Tableau lo visualizará. Por defecto Tableau hace la distinción, aunque un campo discreto puede convertirse a continuo o viceversa.

- **Campos discretos:**

Los campos discretos, como Región en la Figura anterior, son azules. Estos campos tienen valores que se muestran como distintos y separados entre ellos, y se pueden reordenar y seguir teniendo sentido, por ejemplo, en la Figura 71 no importa que el orden sea: Central, Este, Sur, Oeste o que sea Este, Oeste, Central, Sur. El sentido y la interpretación no se ve modificado.

Cuando se arrastra un campo discreto en Filas o Columnas, se crean encabezados. Y cuando se suelta en Color se crea una paleta de colores discreta donde cada color tiene asociado un valor

- **Campos continuos:**

Los campos continuos, como Ventas en la Figura anterior, son verdes. Estos campos tienen valores que van de un primer valor a un último de forma continua, en un orden que no tendría sentido cambiar (a diferencia de lo que ocurre con los discretos) los valores numéricos y las fechas son normalmente valores continuos. Cuando se arrastra un campo continuo en filas o Columnas, se crean ejes. Y cuando se arrastra a Color, se crea un gradiente de color.

Es importante entender que continuo/discreto y dimensión/ medida son conceptos distintos, aunque normalmente las medidas son continuas y las dimensiones discretas.

7.3 Tipos de Gráficos

Tableau ofrece numerosas formas de representación, pero para llegar a las más complejas hay que entender cómo se crean los gráficos más sencillos, y qué utilidad aporta cada uno de ellos.

⇒ **Gráficos de barras:**

Los gráficos de barras hacen la comparación de valores entre distintas categorías sencilla, esto se hace principalmente por la distinta longitud entre las barras, pero, adicionalmente se puede incluir color, tamaño, o apilar las barras. Para crearlos hace falta arrastrar una dimensión a filas o columnas y una medida al estante opuesto que se haya utilizado en la dimensión:

Como se ha visto en apartados anteriores el Departamento, que es una dimensión secciona los datos (las ventas) y como es discreto define encabezados. Las ventas, que es una medida, están agregadas (en suma) y como es continuo define un eje.

En la Figura 72 se puede ver fácilmente que el departamento de tecnología es el que mayor volumen de ventas tiene. Tableau ofrece la posibilidad de profundizar aún más en este análisis y ver si al separar por región el comportamiento es el mismo.

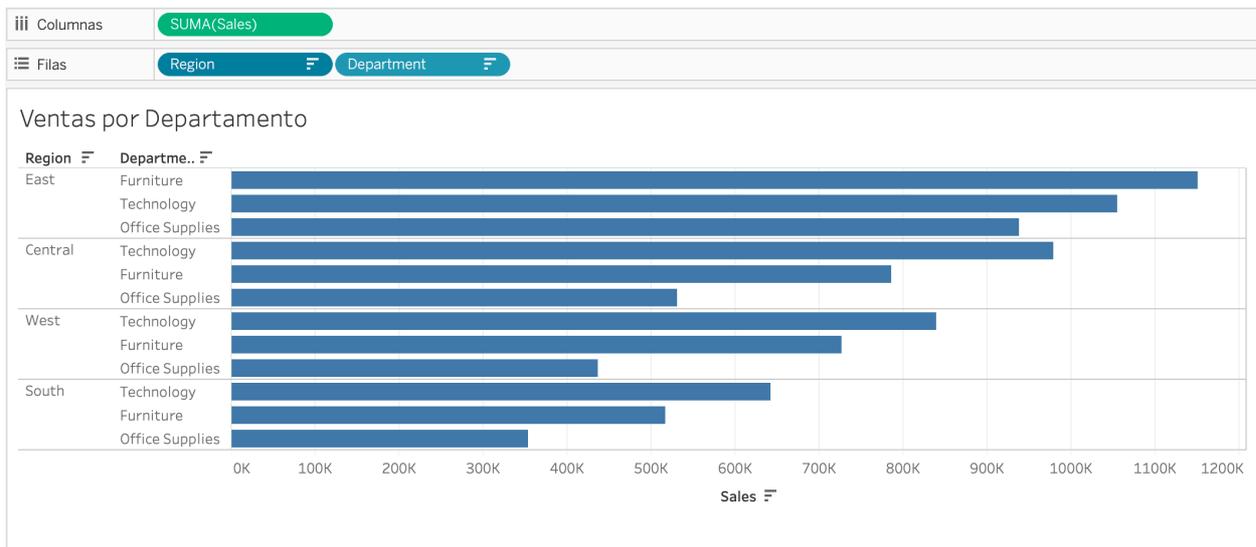


Figura 72: Gráfico de barras II

Ahora se puede ver que, aunque en total, tecnología tiene las mayores ventas, en la región Este mobiliario supera a tecnología. También se puede afirmar que, en total, la región con mas ventas es la del este y la que menos, la Sur. Y el material de oficina es el último en todas las regiones.

Utilizando los mismos campos: Ventas, Región, y departamento, pero organizándolos diferente se puede crear un gráfico de barras apiladas

Anexo: La interfaz de Tableau.

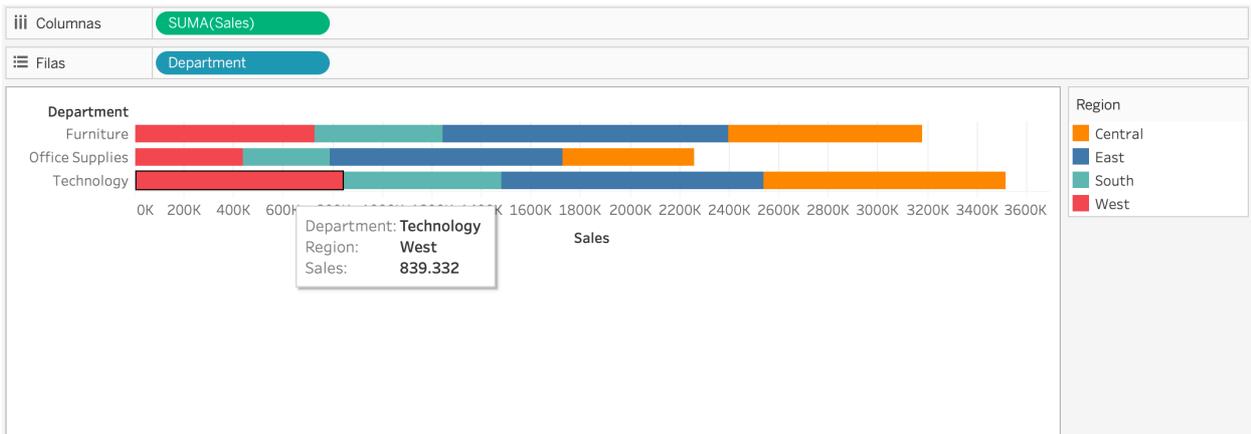


Figura 73: Gráfico de barras apiladas.

El nivel de detalle de esta visualización es el mismo; se muestra las ventas agregadas en suma para cada departamento y región.

Las barras apiladas pueden ser útiles cuando se quiere entender relaciones entre una parte y el conjunto.

⇒ Gráficos de líneas:

Los gráficos de líneas conectan marcas, y muestran movimiento o relaciones entre dichas marcas.

El tipo de gráfico de línea más común es una serie temporal. Esta muestra el movimiento de valores según pasa el tiempo.

En la Figura 74 se puede ver que las ventas tienen un comportamiento cíclico. Si se añade la dimensión Región a la vista, arrastrándolo a Color, el nivel de detalle cambia, se tiene una línea de color para cada región.

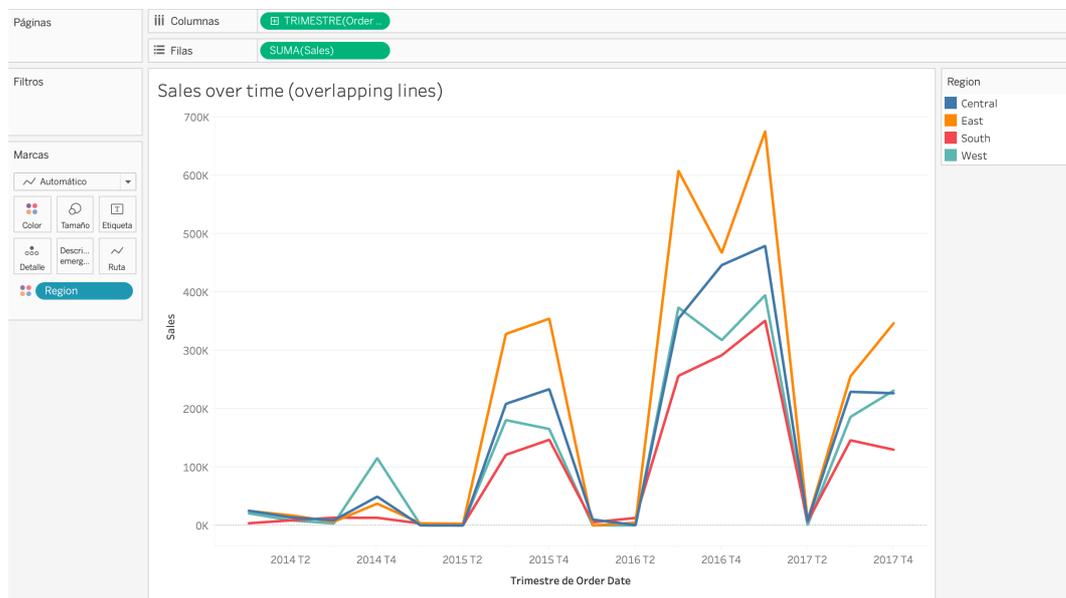


Figura 74: Gráfico de líneas en Tableau II.

⇒ **Visualizaciones geográficas:**

Si la base de datos a la que se conecta Tableau tiene datos geográficos como puede ser un país, estado, ciudad, aeropuerto o código postal, Tableau es capaz de reconocerlo como un Campo Geográfico.

Las visualizaciones geográficas son una buena forma de representación cuando se quiere entender dónde ha ocurrido algo o si la localización es de especial importancia para los datos que se estén analizando, los tres tipos de visualizaciones que ofrece Tableau son:

- **Mapas rellenos:**

Rellenan áreas como ciudades, estados o países, el color del que se rellena el área se usa para codificar valores, lo más normal es que sean medidas agregadas, como las ventas o beneficios.

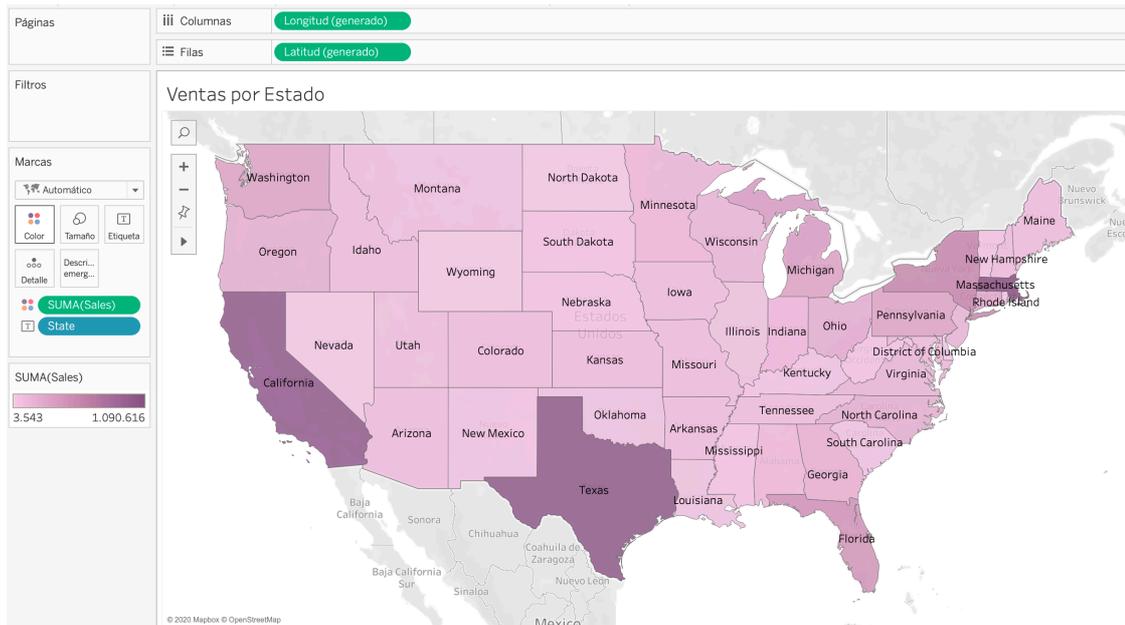


Figura 75: Mapa relleno.

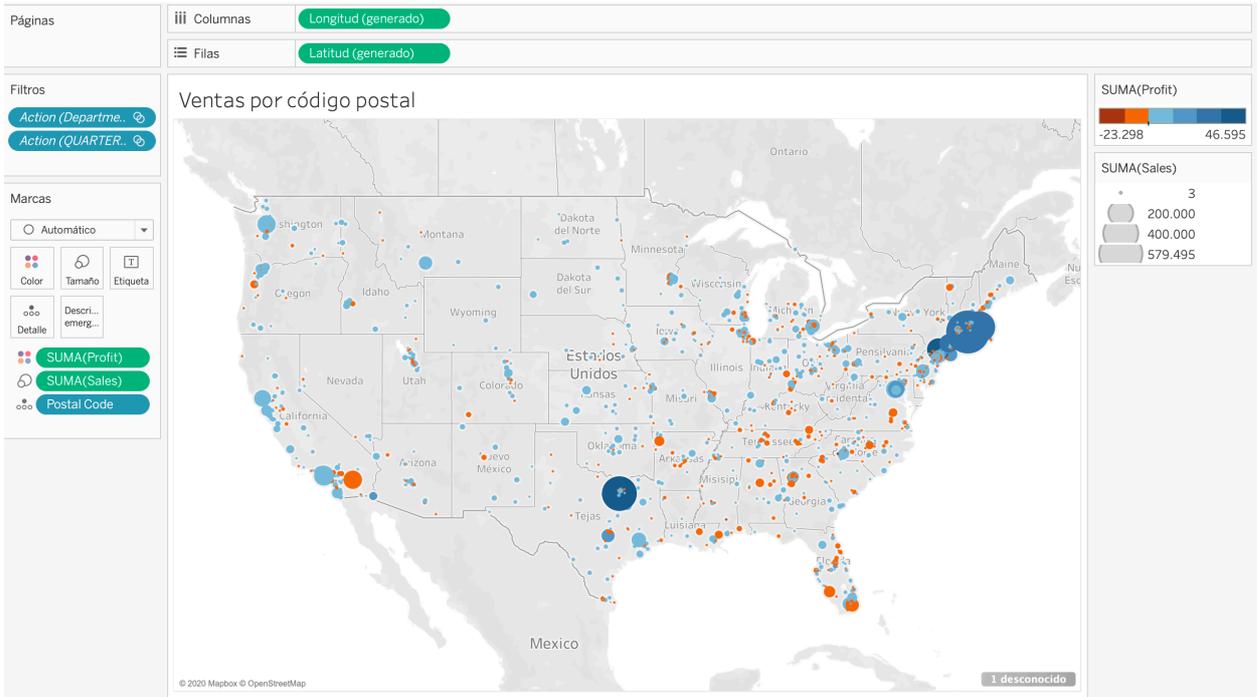
En este mapa se ve que cada estado se ha coloreado en un tono del gradiente de púrpuras para mostrar las ventas que ha tenido. En la leyenda de color se puede ver que el estado con menos ventas tuvo 3543\$y el que más 1.090.616 \$.

Si bien este tipo de mapas son muy estéticos, para cierto tipo de análisis resulta complicado trabajar con ellos. Por ejemplo, a diferencia de otras visualizaciones aquí el tamaño sólo es geográfico, si se quisiera responder a la pregunta ¿qué estado tiene mayores ventas? se vería claro que Texas y California, pero Massachusetts podría pasar desapercibido, ya que su tamaño es menor.

- **Mapas de símbolos:**

En estos mapas, en lugar de rellenar el área de la ubicación, se colocan marcas (que pueden ser formas o símbolos) en determinadas zonas del mapa. El tamaño, color o forma se puede usar para codificar dimensiones o medidas:

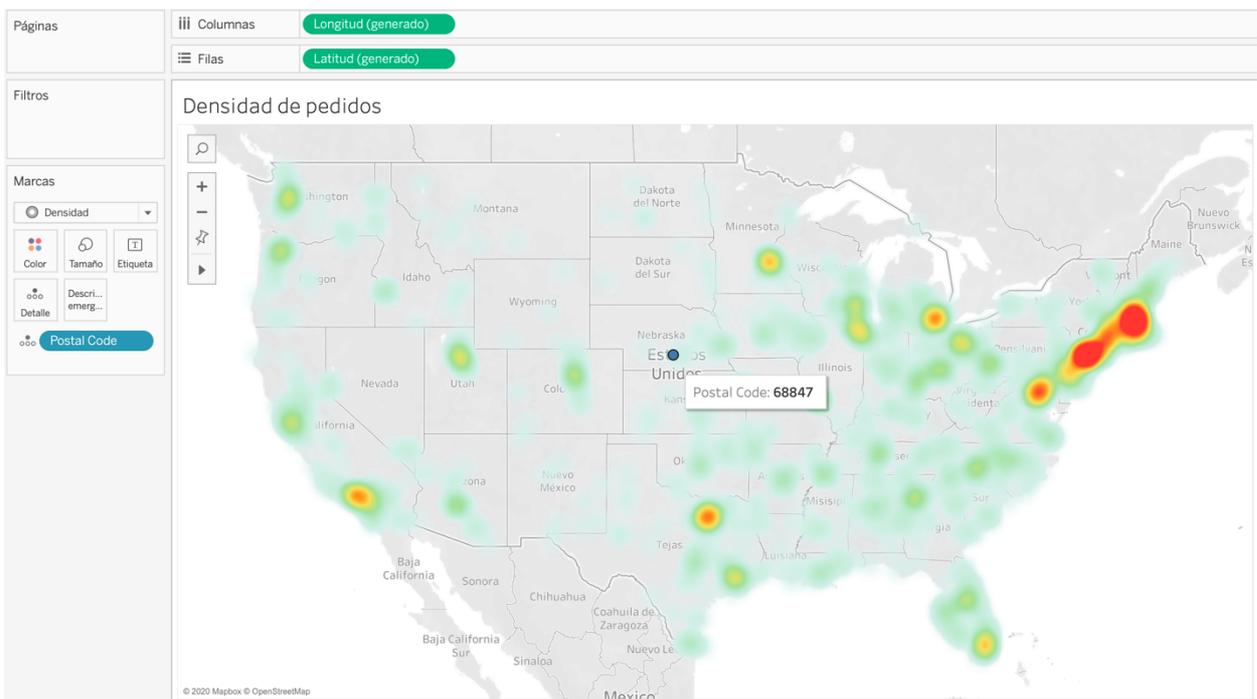
Anexo: La interfaz de Tableau.



En este tipo de mapas se puede jugar con el tamaño para el análisis, a diferencia de lo que ocurría con los mapas rellenos

o Mapas de densidad:

Los mapas de densidad muestran la expansión y concentración de valores dentro de un área geográfica, en vez de mostrarse como marcas separadas, éstas tienden a unirse con las más cercanas para mostrar intensidad en zonas con una gran concentración. Se puede controlar el color, tamaño e intensidad. Este mapa de densidad muestra una gran cantidad de pedidos de la costa este frente a la oeste.



7.4 Filtros de datos

Cuando se quiere realizar un análisis en un subconjunto de datos, reducir el enfoque o profundizar en detalle, Tableau ofrece la posibilidad de filtrar en las visualizaciones o Dashboards.

Se pueden aplicar filtros de las siguientes formas:

- Arrastrando y soltando campos del panel de datos al estante **filtros**.
- Haciendo clic en cualquier campo del panel de datos y dándole a Mostrar Filtro.
- Usando filtros de acción, que se verán más adelante.

Cuando se suelta un campo en el estante Filtros, se solicitarán opciones para definir el filtro. Las opciones de filtro diferirán en función de si el campo es discreto o continuo.

7.4.1 Filtros para campos discretos

Cuando se filtra un campo discreto, aparecen opciones para seleccionar valores individuales que mantener o excluir. Las opciones de filtros varían según la pestaña que se elija para aplicar el filtro.

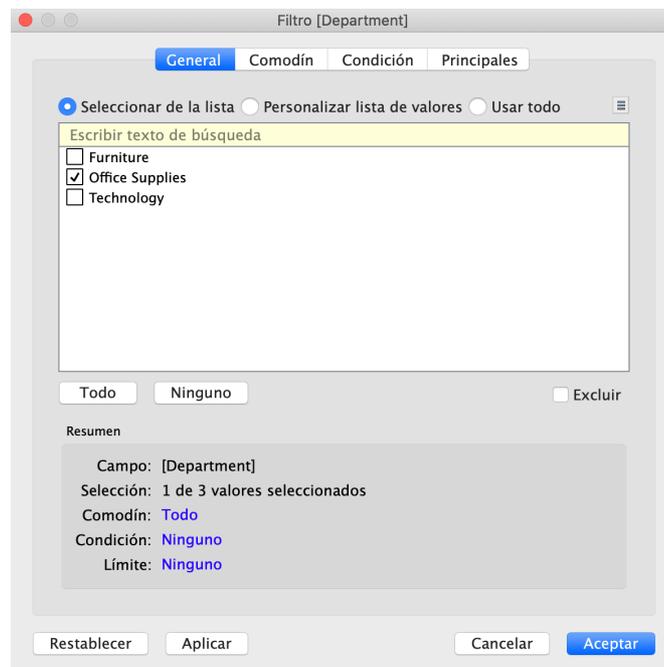


Figura 78: Filtros I.

Hay cuatro pestañas en el cuadro de diálogo:

- **General:** se esta pestaña simplemente se seleccionan los valores que se desea incluir o excluir.
- **Comodín:** se utiliza para definir un patrón de filtrado. Por ejemplo, al filtrar direcciones de correo electrónico, se recomienda incluir solo correos de un dominio específico. Se puede definir un filtro de comodín que termine en "@alum.us.es." para incluir solo las direcciones de correo electrónico de los alumnos de la Universidad de Sevilla.
- **Condición:** se usa la pestaña Condición del cuadro de diálogo Filtrar para definir las reglas de filtrado. Por ejemplo, en una vista que muestra el precio de unidad promedio de prendas de ropa, se podría estipular como condición que únicamente se muestren productos cuyo precio unitario promedio sea mayor o igual a 20€.

- **Principales:** se utilizaría, por ejemplo, en una vista que muestra el promedio de tiempo de envío de los productos de Amazon. Si se quiere filtrar mostrando solo los 10 productos principales en términos de ventas. En lugar de tener que definir un rango específico para Ventas (por ejemplo, superior a 200.000€, se puede definir un límite (los 10 primeros) relativo a los demás miembros del campo (productos).

7.4.2 Filtros para campos continuos

Al arrastrar un campo continuo al estante de filtros, las opciones que aparecen son distintas y son las siguientes:

- **Intervalo de valores:** se selecciona un rango de valores para especificar los valores mínimo y máximo del rango para incluir en la vista.
- **Mínimo:** se selecciona para incluir todos los valores mayores o iguales a un valor mínimo especificado. Este tipo de filtro es útil cuando los datos cambian a menudo, por lo que especificar un límite superior puede no ser posible.
- **Máximo:** seleccione la opción Máximo para incluir todos los valores menores o iguales a un valor máximo especificado.
- **Especial:** se utiliza para filtrar los valores nulos. Se puede elegir entre incluir solo valores nulos, valores no nulos o todos los valores.

7.5 Visualizaciones Avanzadas

Para poder analizar todo lo que esconden los datos y tomar decisiones, hay que ser capaz realizar la representación que mejor se adapte a lo que se quiere transmitir, aunque se ha visto previamente las representaciones más básicas, se puede profundizar más dentro de ellas para realizar comparaciones, representar fechas y horas, situar una parte en un conjunto, representar distribuciones o utilizar múltiples ejes. Todo esto ayudará a que se extraiga la máxima información posible del conjunto de datos.

7.5.1 Comparar valores

Se utilizan cuando se quiere comparar la diferencia entre valores medidos en diferentes categorías, o cuando surgen el siguiente tipo de preguntas:

- ¿Cuántos clientes ha atendido cada tienda?
- ¿Cuánta energía se ha producido en cada parque eólico?
- ¿A cuántos pacientes ha atendido cada doctor?

Todas tienen en común que se está haciendo una comparación, ya sea entre tiendas, energía o médicos en cuanto a algún termino que se puede medir (numero de clientes, energía, pacientes).

⇒ **Diagramas de barras:**

La siguiente Figura muestra el número de pacientes atendidos en un hospital por departamento médico, el hecho de que esté ordenado hace más fácil la comparación ya que, por ejemplo, se puede ver que microbiología tuvo más pacientes que nutrición, cosa que no sería tan fácil de ver si el orden fuera alfabéticamente según la especialidad.

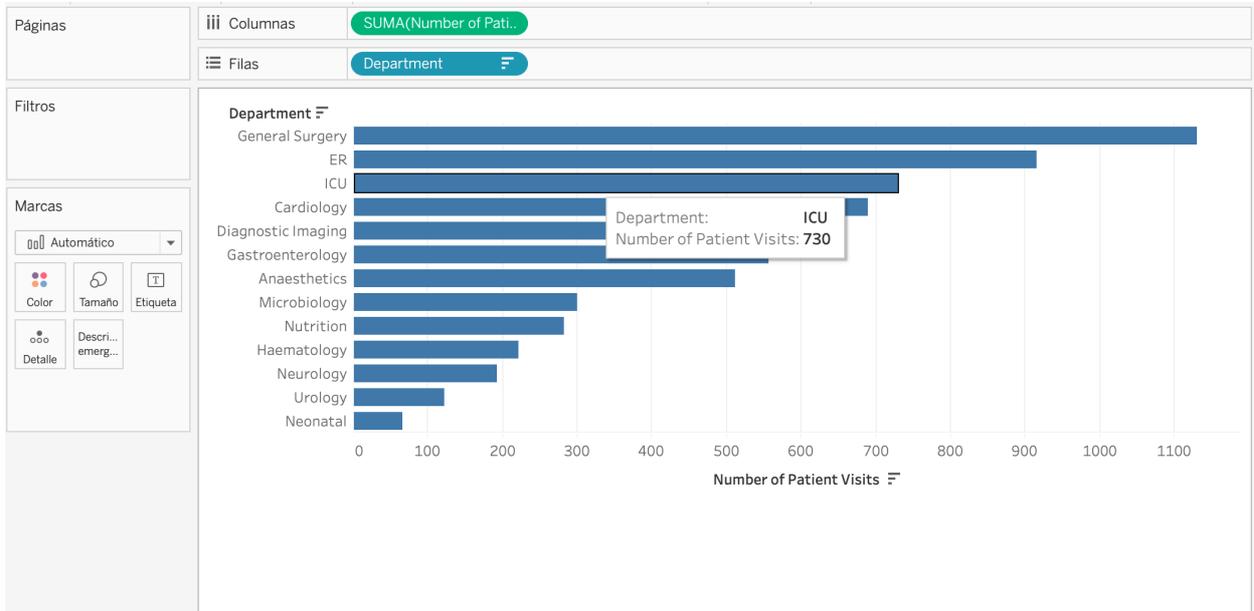


Figura 79: Diagrama de barras en orden descendente.

Especialidad es una dimensión y por eso segmenta los datos en función de la especialidad a la que pertenezca, por ser discreto (azul), crea encabezados.

Número de pacientes es una medida, que da la suma de pacientes en cada departamento, y define un eje porque es continuo (verde).

Si, por ejemplo, ocurriera que solo interesan dos categorías, de las que se quiere hacer un seguimiento, la unidad de cuidados intensivos y neonatos, se puede crear un grupo con esos dos elementos, cambiarles el color, y poner una etiqueta para ver el valor que tienen.

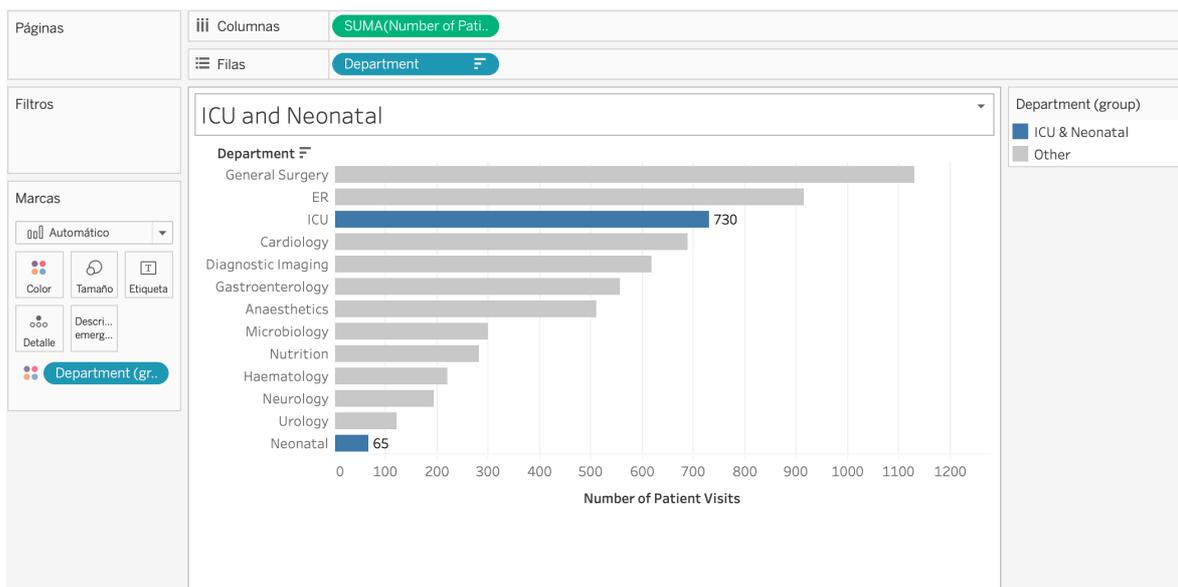


Figura 80: Diagrama de barras resaltado.

La Figura 80 muestra un grafico de barras básico, pero se puede extender este concepto para realizar otros análisis.

⇒ **Gráficos de Bala:**

Son una buena opción para cuando se quiere comparar un valor con un objetivo o con un umbral. La barra indica el valor de la medida, la línea negra vertical indica el objetivo. Y por defecto Tableau sombrea la distancia para indicar donde se ha alcanzado el 60% y 80% del valor objetivo.

Por ejemplo, si el hospital tiene un registro con el tiempo estándar u objetivo entre que cada paciente llega al hospital y es atendido por el médico de la especialidad correspondiente:

| Department | Minutes to Service Goal |
|--------------------|-------------------------|
| Anaesthetics | 38 |
| Cardiology | 30 |
| Diagnostic Imaging | 60 |
| ER | 45 |
| Gastroenterology | 40 |
| General Surgery | 40 |
| Haematology | 40 |
| ICU | 45 |
| Microbiology | 35 |
| Neonatal | 30 |
| Neurology | 45 |
| Nutrition | 30 |
| Urology | 40 |

Figura 81: Base datos hospital, tiempo objetivo por especialidad.

Por otra parte, se tiene la base de datos conocida como primaria, de cada registro por paciente, con el tiempo real que esperó hasta ser atendido por el doctor. Se puede realizar una unión entre los datos de las dos fuentes a través del campo especialidad (*department*) común a ambas.

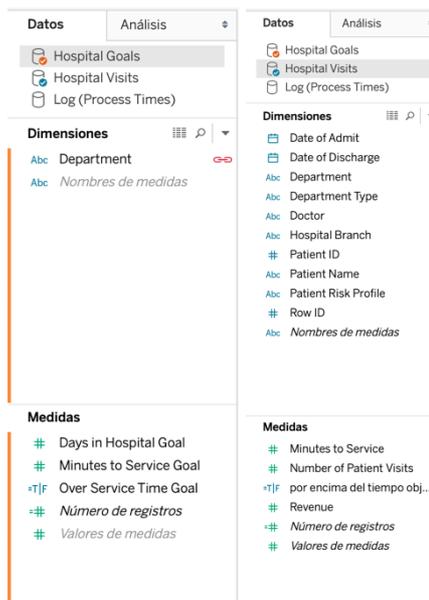


Figura 82: combinación fuentes datos.

Con esta unión de datos es posible representar gráficamente el tiempo promedio de espera por paciente y especialidad frente al objetivo o tiempo estándar.

Además, en Tableau se pueden crear campos calculados, como se verá en el capítulo 5, creando un campo calculado del tipo verdadero o falso, que compare ambos tiempos y, arrastrándolo al estante de la marca Color, se hace más sencillo la identificación de especialidades donde los tiempos son peores que los estándares.

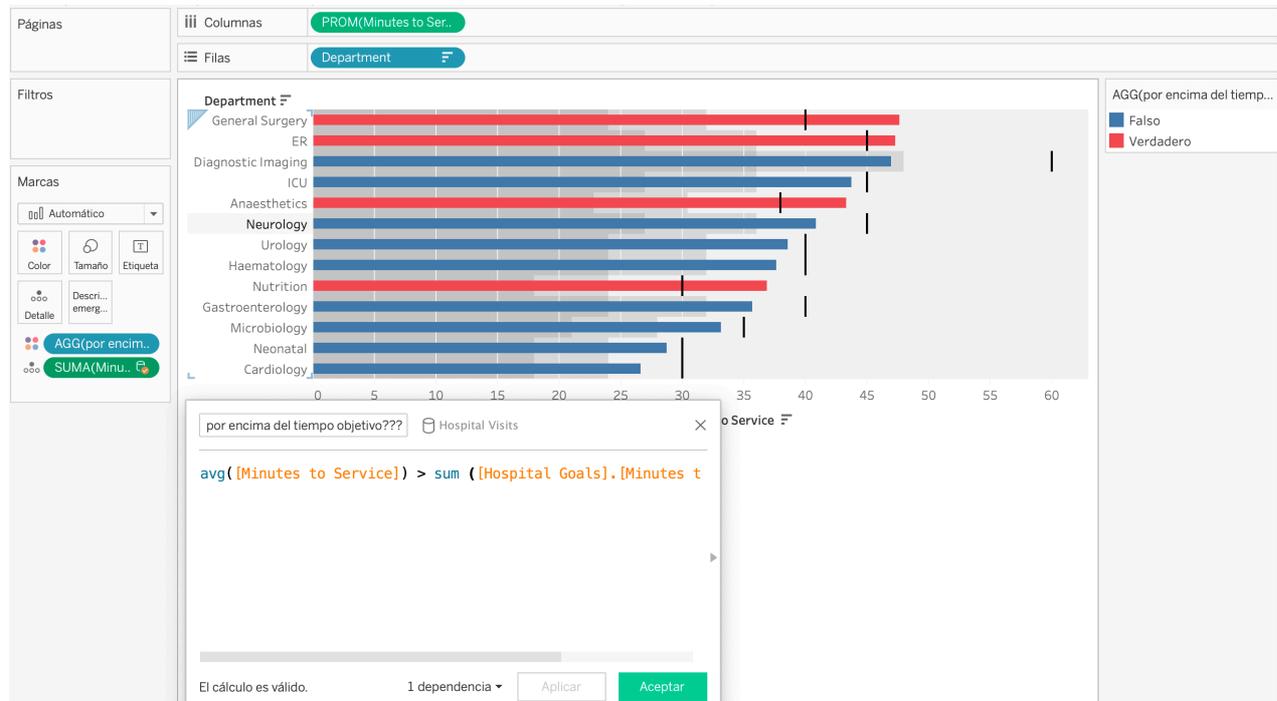


Figura 83: Diagrama de bala.

⇒ **Diagrama de barra dentro de barra**

Otra forma de mostrar relaciones entre dos valores dentro de una misma categoría es utilizando este tipo de gráficos. Sirven tanto para mostrar el progreso frente a un objetivo como para comparar dos valores entre sí.

Por ejemplo, se podría compara el beneficio obtenido entre dos años consecutivos. Es muy parecido a un gráfico de barras apiladas, pero se desactiva la opción de apilar e incluyendo el tamaño como marca para que no se solapen las dos barras y se puede entender el gráfico.

Anexo: La interfaz de Tableau.

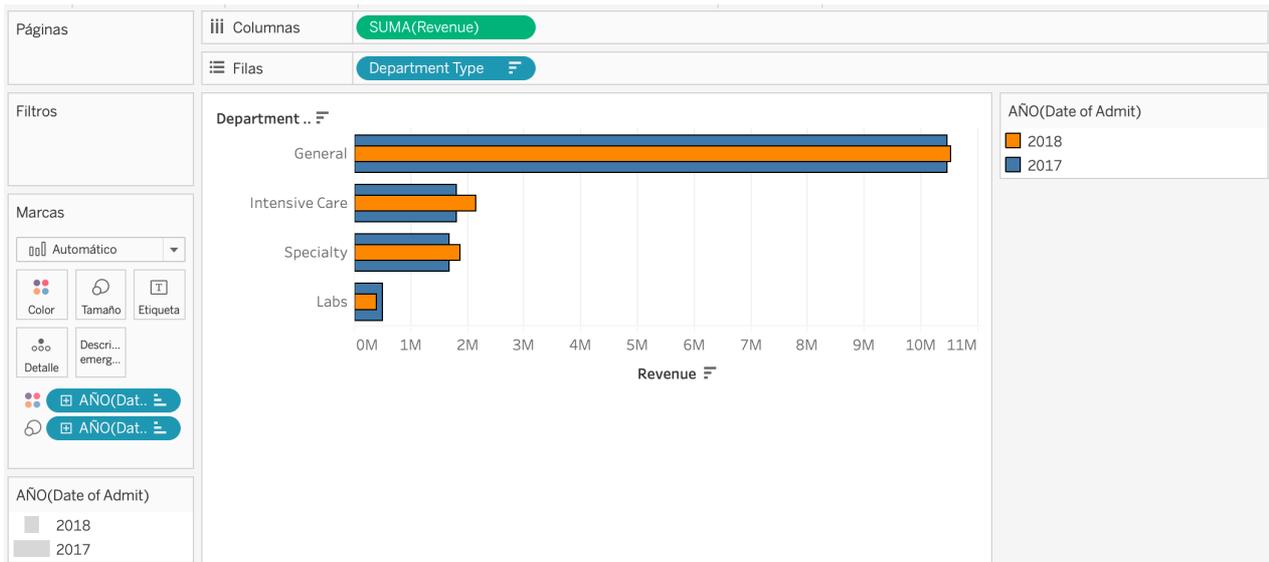


Figura 84: Variación diagrama de barras.

7.5.2 Fechas y horas

Se utilizan en situaciones en las cuales se quiere comprender cuándo ha ocurrido algo y el tipo de cuestiones que se plantean son del tipo:

- ¿Cuándo se obtuvo el máximo beneficio?
- ¿A qué hora del día hay el mayor número de clientes?
- ¿Afecta la estación del año en las ventas y el beneficio?

Es necesario comprender los conceptos de Valor de Fecha, Parte de Fecha y Fecha exacta.

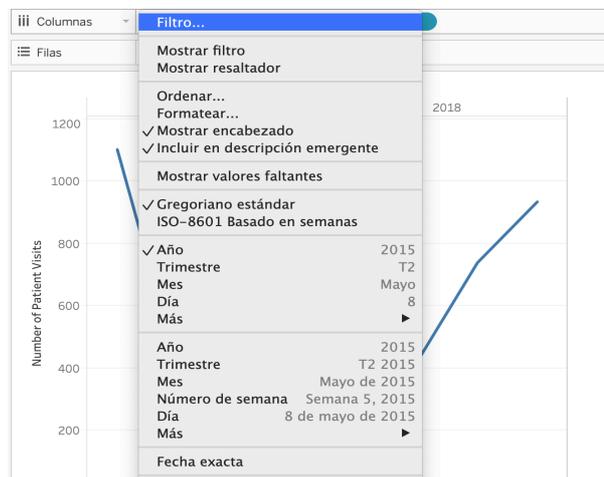


Figura 85: Fechas.

- o Parte de fecha: Representa una parte específica de la fecha, como por ejemplo el trimestre o el mes, la parte de la fecha se usa en sí misma sin tener en cuenta a lo demás, esto quiere decir que, si la fecha que se tiene es 13 de noviembre de 2018 y se elige: parte de fecha - mes, se estarán teniendo en cuenta todos los meses de noviembre que existen en la base de datos, no solo el año 2018

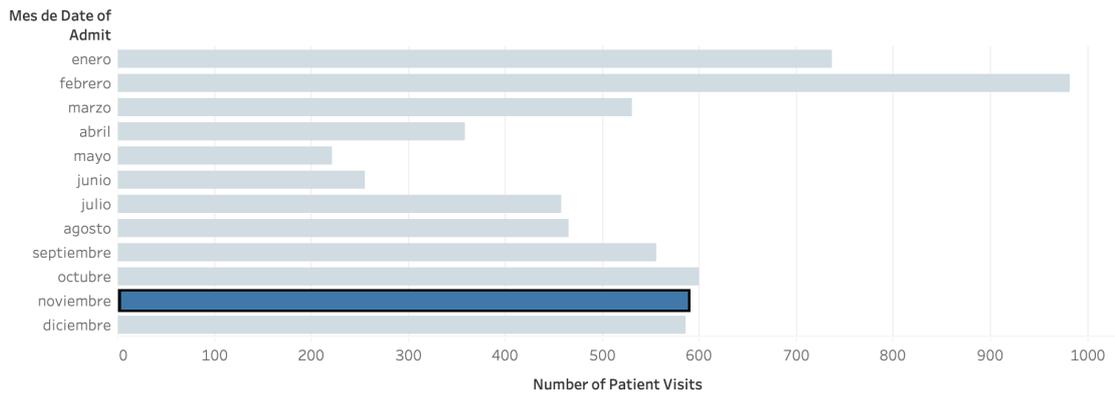


Figura 86: Parte de fecha.

- Valor de fecha: En este caso se trunca la fecha al nivel seleccionado, es decir si se tiene 13 de noviembre de 2018 y se elige: valor de fecha – mes, se tendría noviembre 2018 y noviembre 2017

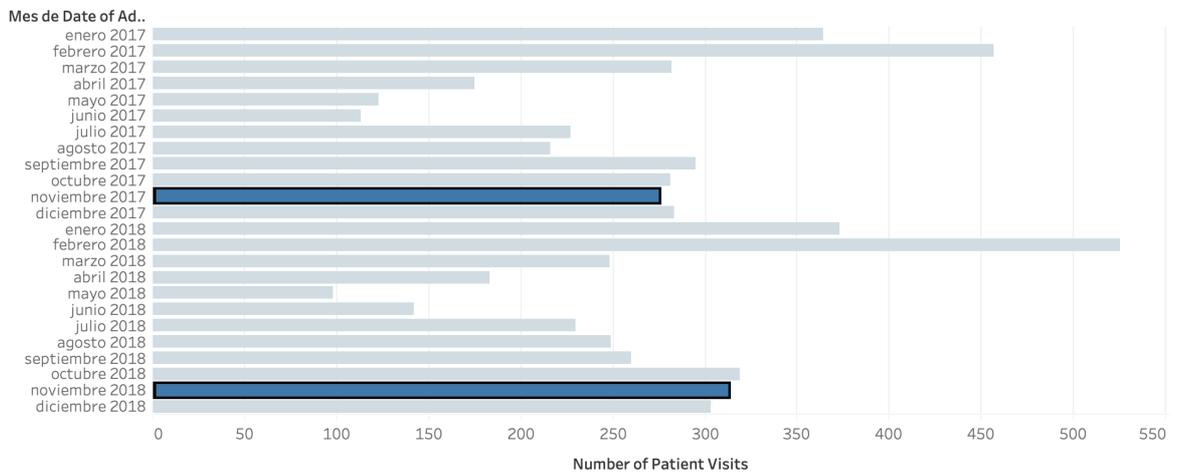


Figura 87: Valor de Fecha.

- Fecha exacta: cuando se selecciona este campo cada fecha es tratada con todo el nivel de detalle, incluyendo incluso la hora, es decir, 13 noviembre 2017 13:13 se trata distinto que 13 noviembre 2017 2:43.

Por defecto los partes de fecha son discretos y los valores de fecha y fechas exactas son continuos.

La habilidad de usar y mezclar partes y valores de fecha aporta mucha flexibilidad a la hora de crear visualizaciones interesantes.

Anexo: La interfaz de Tableau.



Figura 88: Combinación campos de Fechas.

Cuando interesa conocer patrones entre momentos temporales, ya sea entre del día semanas de mes, se pueden utilizar los mapas de calor, como el que se tiene en la Figura 89.

En este caso se representa el número de pacientes admitidos por día y mes, como el año no está incluido se está realizando el análisis para todos los años incluidos en la base de datos.

Esto sirve para sacar conclusiones como, por ejemplo, que el aumento de pacientes a mediados de febrero puede deberse a la gripe.



Figura 89: Mapa de calor.

7.5.3 Relacionar una parte con el total

Cuando el interés radica en comprender cómo una parte afecta al conjunto de datos, se podrían plantear preguntas de tipo:

- ¿Cuánto aporta cada tipo de generación de energía eléctrica (eólica solar, nuclear, carbón) al total de energía producida?
- ¿Qué porcentaje del beneficio total se ha obtenido en cada país?

En ambas preguntas se está buscando la relación entre una parte (método de producción, país) y el total (energía total, beneficio total)

En este subapartado se van a ver los tipos de visualización que pueden realizar en Tableau para realizar este tipo de análisis.

BARRAS APILADAS

Siguiendo con el caso del hospital, en la base de datos aparece que a cada paciente se le ha asignado un nivel de riesgo: alto o bajo.

Se puede analizar, para los pacientes que han sido atendidos por cada especialidad, cuántos de ellos tenían un perfil de riesgo alto y bajo. Además, para añadir más información a la vista, se puede hacer un cálculo de tablas rápido, que se explicará en detalle en el capítulo 5, para ver la misma información, pero en porcentaje respecto al total de pacientes por departamento.

Combinando estas dos barras en la Figura se añade valor a la visualización ya que se pueden extraer más conclusiones.



Figura 90: Barras apiladas.

MAPAS DE ÁRBOL

Los mapas de árbol utilizan una serie de rectángulos para mostrar cada parte respecto al total. Estos son de especial utilidad cuando se tienen relaciones jerárquicas y dimensiones con gran cantidad de valores distintos.

Anexo: La interfaz de Tableau.

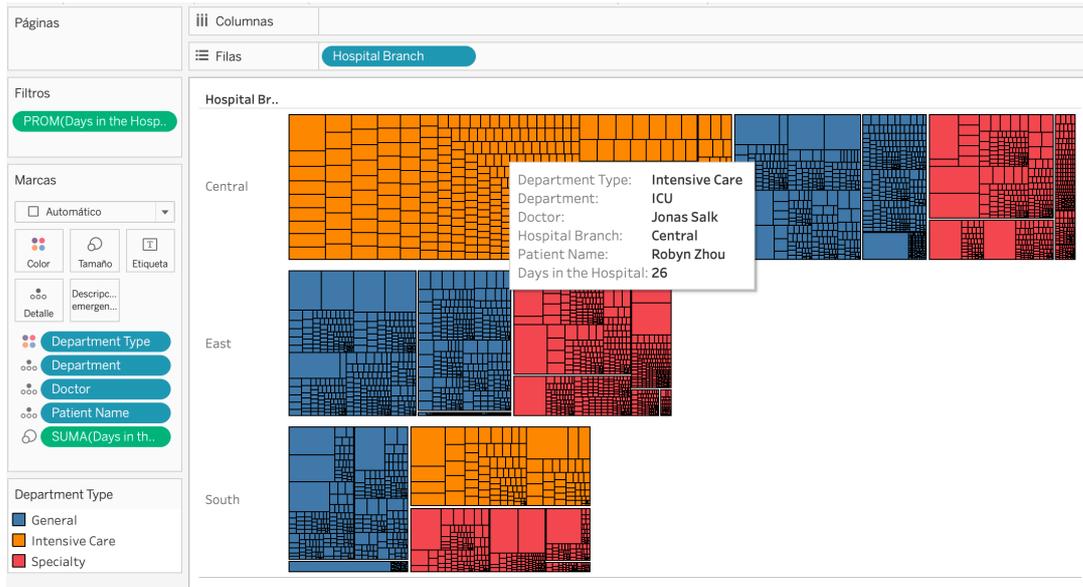


Figura 91: Mapa de árbol.

Diagramas de área

Es una variante de un gráfico de líneas, cambiando la marca de línea a una marca de área se obtiene la visualización siguiente:

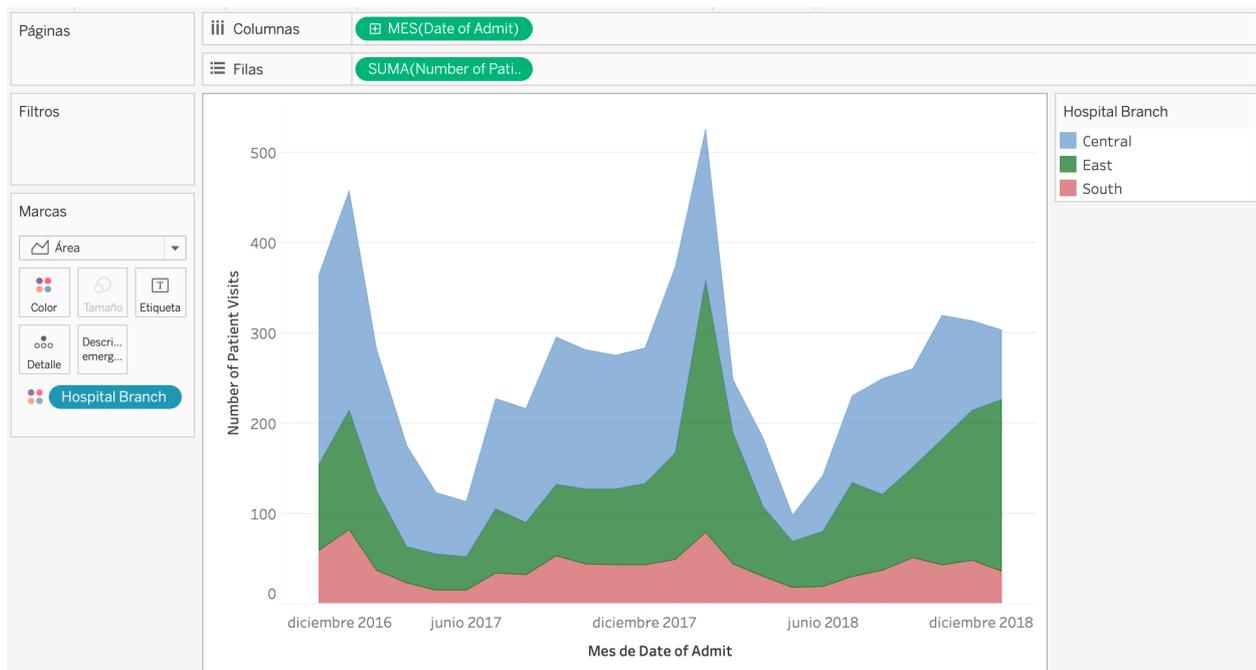


Figura 92: Diagrama de área.

Cada banda representa una sucursal del hospital y, aunque la visualización es estética, ocurre lo mismo que con los diagramas de barras apiladas, solo el área rosa, correspondiente a la sucursal Sur del hospital, puede ser leída en el eje de la izquierda. Como las otras líneas están apiladas resulta más complicado la comparación. Por ejemplo, es obvio que existe un pico en febrero en 2017 y 2016 pero no se puede afirmar con seguridad que ocurra en las tres sucursales o que sea el efecto de una de ellas y que esta afecte a las demás.

Si se realiza sobre el gráfico anterior un cálculo de tablas rápido, que se verá en detalle en el Capítulo 5, y se calcula el porcentaje del total de pacientes que había en cada sucursal por día se tendría el gráfico siguiente:



Figura 93: Diagrama de área relleno.

Ya no se ven los picos como ocurría en la Figura 92, y además se puede afirmar que alrededor de febrero/marzo de 2018 hubo un aumento significativo del número de pacientes en la sucursal del este respecto a las otras dos.

Según como se representen los datos, se enfatizará o se ocultarán algunos de estos aspectos. Por este motivo, es importante saber qué se quiere descubrir o contar y cómo aprovechar las herramientas que ofrece Tableau.

DIAGRAMAS CIRCULARES

Se usan con el mismo propósito que los anteriores, mostrar la relación entre una parte con el total.

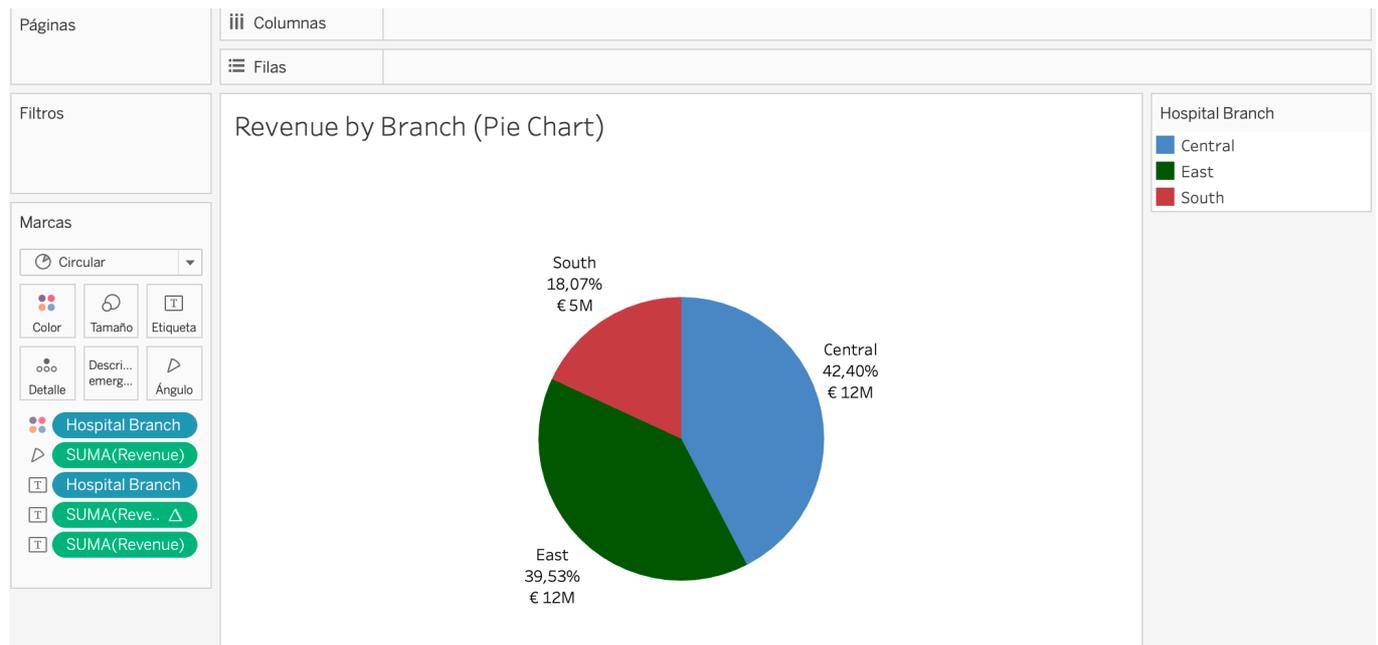


Figura 94: Diagrama circular.

Este diagrama es interesante cuando la medida que se va a arrastrar a color tiene pocos valores (entre 2 - 4) ya que sino se hace complicada la interpretación del gráfico.

7.5.4 Visualizar distribuciones

En las situaciones en las que se quiere representar donde está cada elemento individual respecto a una distribución de elementos similares. Se podrían tener preguntas como:

- ¿Cuánto gasta cada cliente en la tienda comparado con el resto de los clientes?
- ¿Cuánto tiempo están los pacientes en el hospital?
- ¿Cómo se distribuyen las notas de los alumnos respecto al aprobado?

En todas ellas se busca comprender cómo cada (cliente, paciente, alumnos) se comporta en relación con el conjunto.

Para hacer esto se pueden utilizar los siguientes gráficos:

Diagrama con círculos

Es útil si lo que se quiere ilustrar es cómo cada doctor se relaciona con los demás de su especialidad, en cuanto al promedio de minutos para atender a los pacientes.

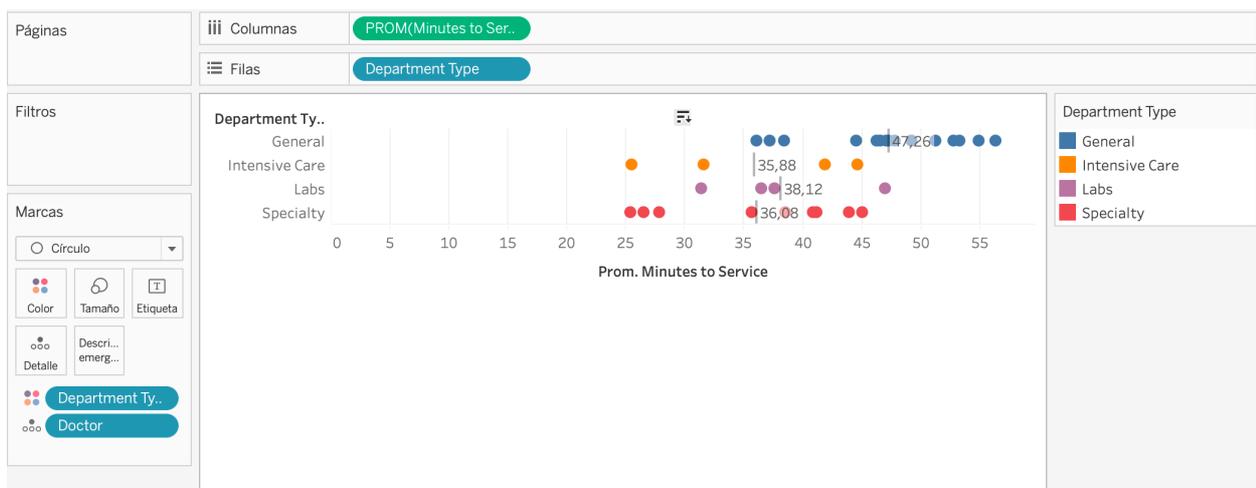


Figura 95: Diagrama con círculos.

Se ve con una línea gris el promedio de minutos por especialidad, y donde se sitúa cada doctor.

Diagramas de caja y bigotes

Añaden contexto estadístico adicional a las distribuciones. Se podría añadir al gráfico que se tenía en la Figura anterior. La caja está dividida por la mediana, eso quiere decir que la mitad de los valores están por encima y la otra mitad por debajo, también se ven los cuartiles superiores e inferiores y, las marcas que estuvieran fuera de los bigotes, en este caso ninguna, se considerarían valores atípicos.

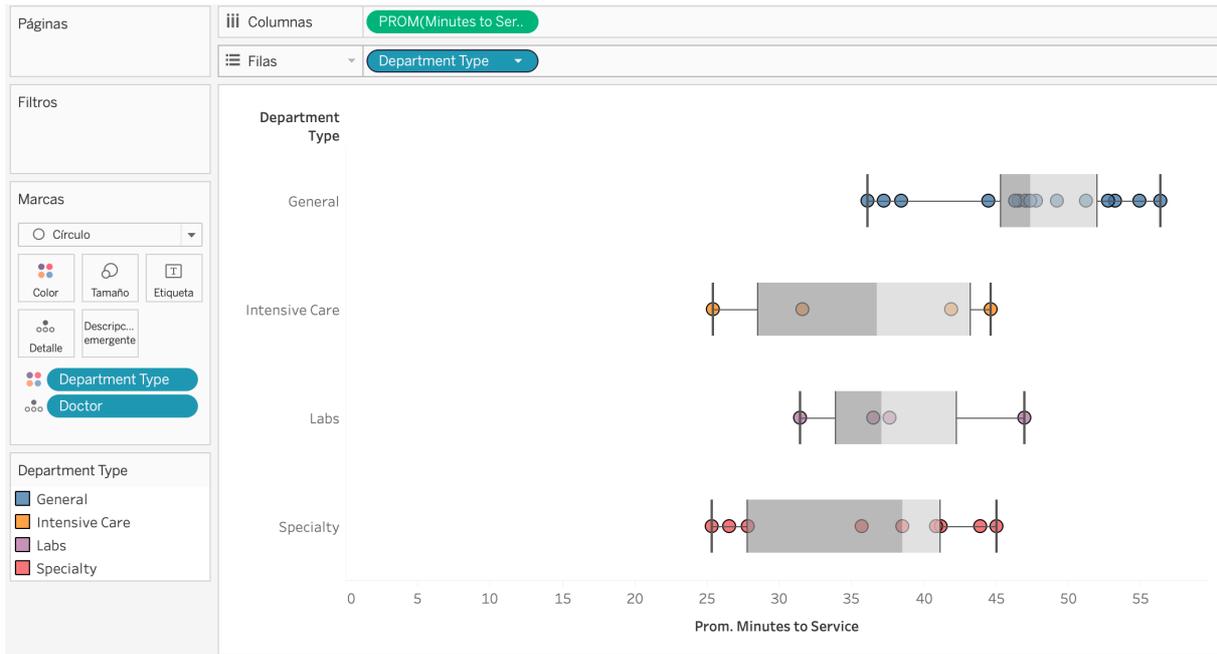


Figura 96: Diagrama de caja y bigotes.

Histogramas

Un histograma se asemeja a un diagrama de barras, con la diferencia de que en este caso las barras muestran un recuento de ocurrencia de un valor, por ejemplo, para ver la puntuación obtenida en un examen, una distribución normal sería: algunos estudiantes con notas muy bajas, otros con notas muy buenas y la gran mayoría en el medio, lo que se conoce como distribución de campana.

Para el caso del hospital, se podría realizar un histograma para ver el tiempo que esperan los pacientes para ser atendidos.

Se puede concluir que la mayoría de los pacientes esperan entre 35 y 50 minutos para ser atendidos.

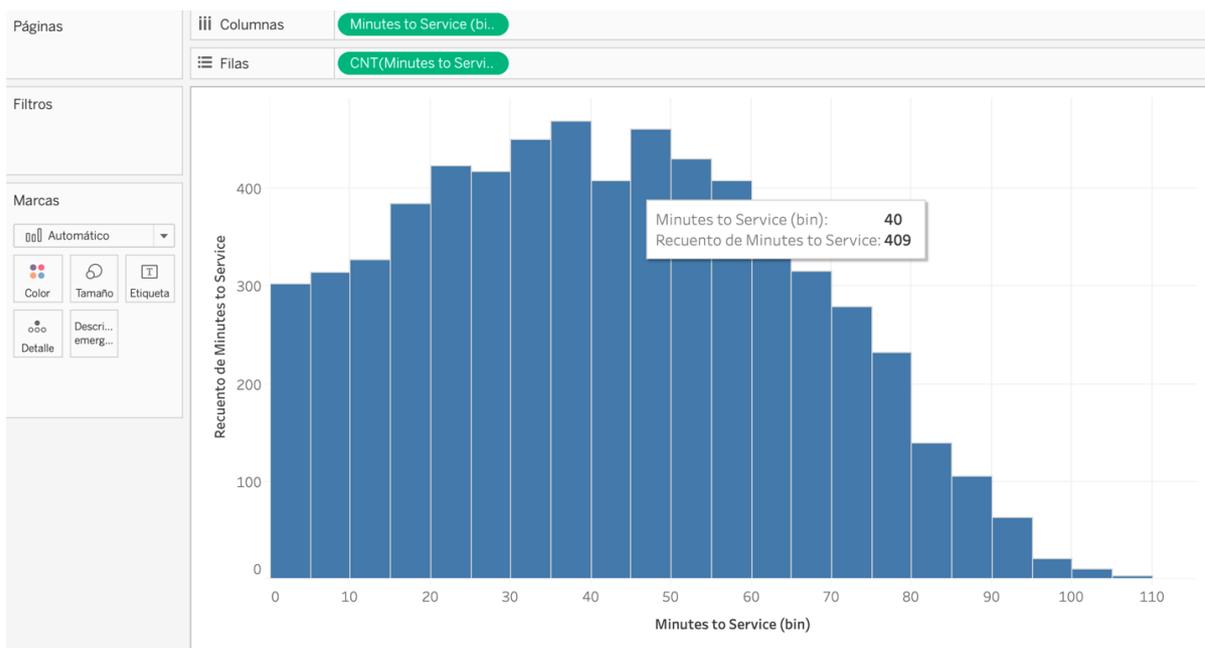


Figura 97: Histograma.

7.5.5 Ejes múltiples

Existen ocasiones en que hace falta tener más de un eje para comparar medidas, entender correlaciones o analizar la misma medida, pero en diferente nivel de detalle. Esto se puede hacer utilizando más de un eje:

Diagrama de dispersión

Este tipo de diagramas es útil para entender la relación que existe entre dos medidas; cuando se quieren responder preguntas del tipo:

- ¿Cuánto de lo que se ha gastado en publicidad ha tenido un efecto significativo en las ventas?
- ¿Existe alguna correlación entre las horas de estudio y los resultados obtenidos en el examen?

La siguiente Figura muestra la correlación entre el promedio de minutos para ser atendidos y el promedio de días que se queda un paciente ingresado en el hospital, segmentado por la especialidad médica y el doctor.

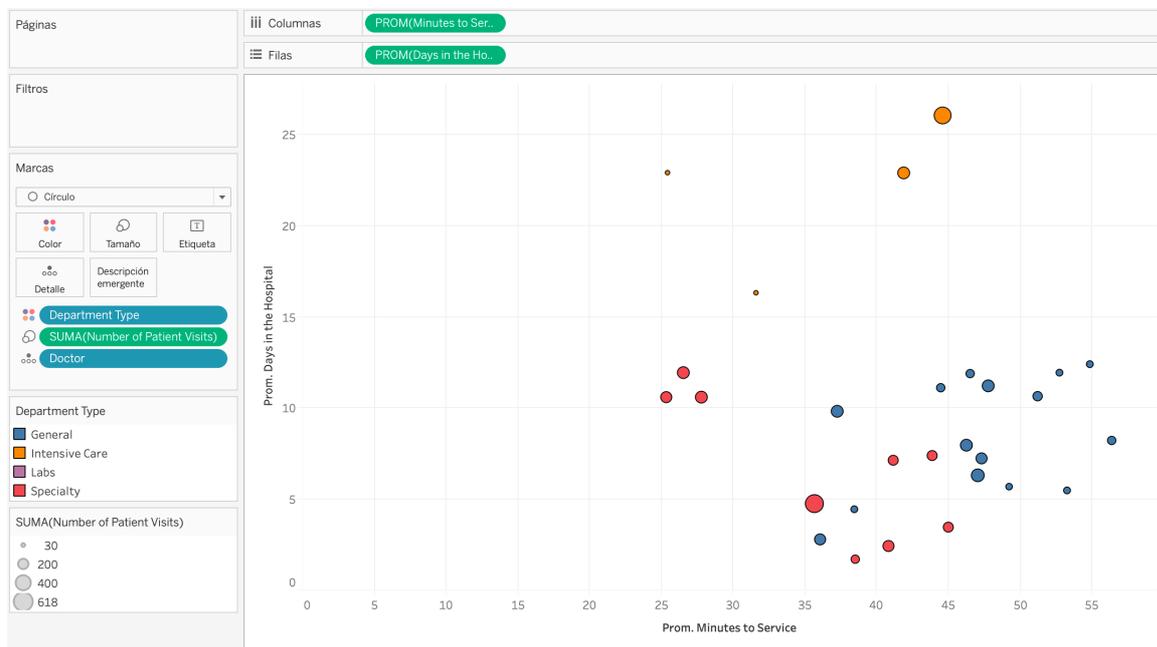


Figura 98: Diagrama de dispersión.

El nivel de detalle de esta visualización está definido por las dimensiones especialidad medica y Doctor. La marca de Color se ha usado para hacer más sencillo la distinción de especialidades, y el Tamaño para ver el número de pacientes que ha visto cada doctor.

Parece que no existe mucha relación entre los minutos de espera y los días en el hospital, pero este diagrama es útil para ver algunos patrones de agrupación entre doctores de la misma especialidad y para identificar a las marcas de la parte superior a la derecha (cuidados intensivos) como valores atípicos.

Eje doble y combinación de gráficos

Tableau permite utilizar ejes dobles, esto significa que en la visualización se están utilizando dos ejes a la vez opuestos entre sí. Para ello es necesario arrastrar dos campos continuos al estante filas o columnas y seleccionar "eje doble".

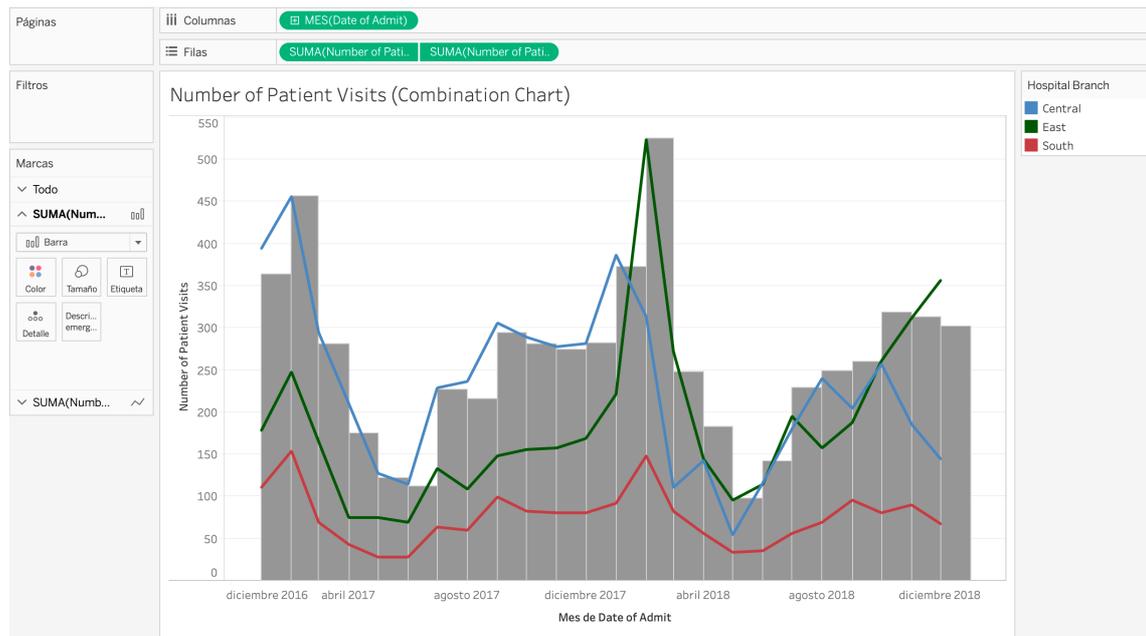


Figura 99: Eje doble.

En esta Figura se ve una combinación de gráficos con eje doble, las barras muestran el total de pacientes por mes, y las líneas lo desglosan según la sucursal del hospital donde han sido atendidos.

7.6 Cálculos en Tableau

Tableau trabaja con bases de datos con un formato similar al de la Figura 100, los títulos de cada columna se separan según sean dimensiones o medidas. Arrastrando y soltando estos campos es posible crear visualizaciones que extraen información y ayudan a realizar el análisis.

Es posible extender el alcance que se ha visto hasta el momento con Tableau, introduciendo los cálculos dentro del software, ya que se pueden crear nuevos campos de dimensiones o medidas que añaden mucha información tanto a las visualizaciones como a los Dashboards.

Dentro del software, se pueden realizar dos tipos de cálculos, los campos calculados y los cálculos de tabla, en este subapartado se va a ver en qué consiste cada uno y su aplicación.

Anexo: La interfaz de Tableau.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|----|-------------|--------|-----------|---------------|--------|------------------------|------------|----------|--------|-------|-------------|-----------|----------|-------------|--------|
| 1 | Invoice ID | Branch | City | Customer type | Gender | Product line | Unit price | Quantity | tax 5% | cogs | total price | Date | Time | Payment | Rating |
| 2 | 750-67-8428 | A | Yangon | Member | Female | Health and beauty | 74,7 | 7 | 26,1 | 522,8 | 548,9715 | 1/5/19 | 13:08:00 | Ewallet | 9,1 |
| 3 | 226-31-3081 | C | Naypyitaw | Normal | Female | Electronic accessories | 15,3 | 5 | 3,8 | 76,4 | 80,2200 | 3/8/19 | 10:29:00 | Cash | 9,6 |
| 4 | 631-41-3108 | A | Yangon | Normal | Male | Home and lifestyle | 46,3 | 7 | 16,2 | 324,3 | 340,5255 | 3/3/19 | 13:23:00 | Credit card | 7,4 |
| 5 | 123-19-1176 | A | Yangon | Member | Male | Health and beauty | 58,2 | 8 | 23,3 | 465,8 | 489,0480 | 1/27/2019 | 20:33:00 | Ewallet | 8,4 |
| 6 | 373-73-7910 | A | Yangon | Normal | Male | Sports and travel | 86,3 | 7 | 30,2 | 604,2 | 634,3785 | 2/8/19 | 10:37:00 | Ewallet | 5,3 |
| 7 | 699-14-3026 | C | Naypyitaw | Normal | Male | Electronic accessories | 85,4 | 7 | 29,9 | 597,7 | 627,6165 | 3/25/2019 | 18:30:00 | Ewallet | 4,1 |
| 8 | 355-53-5943 | A | Yangon | Member | Female | Electronic accessories | 68,8 | 6 | 20,7 | 413 | 433,6920 | 2/25/2019 | 14:36:00 | Ewallet | 5,8 |
| 9 | 315-22-5665 | C | Naypyitaw | Normal | Female | Home and lifestyle | 73,6 | 10 | 36,8 | 735,6 | 772,3800 | 2/24/2019 | 11:38:00 | Ewallet | 8 |
| 10 | 665-32-9167 | A | Yangon | Member | Female | Health and beauty | 36,3 | 2 | 3,6 | 72,52 | 76,1460 | 1/10/19 | 17:15:00 | Credit card | 7,2 |
| 11 | 692-92-5582 | B | Mandalay | Member | Female | Food and beverages | 54,8 | 3 | 8,2 | 164,5 | 172,7460 | 2/20/2019 | 13:27:00 | Credit card | 5,9 |
| 12 | 351-62-0822 | B | Mandalay | Member | Female | Fashion accessories | 14,5 | 4 | 2,9 | 57,92 | 60,8160 | 2/6/19 | 18:07:00 | Ewallet | 4,5 |
| 13 | 529-56-3974 | B | Mandalay | Member | Male | Electronic accessories | 25,5 | 4 | 5,1 | 102 | 107,1420 | 3/9/19 | 17:03:00 | Cash | 6,8 |
| 14 | 365-64-0515 | A | Yangon | Normal | Female | Electronic accessories | 47,0 | 5 | 11,7 | 234,8 | 246,4875 | 2/12/19 | 10:25:00 | Ewallet | 7,1 |
| 15 | 252-56-2699 | A | Yangon | Normal | Male | Food and beverages | 43,2 | 10 | 21,6 | 431,9 | 453,4950 | 2/7/19 | 16:48:00 | Ewallet | 8,2 |
| 16 | 829-34-3910 | A | Yangon | Normal | Female | Health and beauty | 71,4 | 10 | 35,7 | 713,8 | 749,4900 | 3/29/2019 | 19:21:00 | Cash | 5,7 |
| 17 | 299-46-1805 | B | Mandalay | Member | Female | Sports and travel | 93,7 | 6 | 28,1 | 562,3 | 590,4360 | 1/15/2019 | 16:19:00 | Cash | 4,5 |
| 18 | 656-95-9349 | A | Yangon | Member | Female | Health and beauty | 68,9 | 7 | 24,1 | 482,5 | 506,6355 | 3/11/19 | 11:03:00 | Credit card | 4,6 |
| 19 | 765-26-6951 | A | Yangon | Normal | Male | Sports and travel | 72,6 | 6 | 21,8 | 435,7 | 457,4430 | 1/1/19 | 10:39:00 | Credit card | 6,9 |
| 20 | 329-62-1586 | A | Yangon | Normal | Male | Food and beverages | 54,7 | 3 | 8,2 | 164 | 172,2105 | 1/21/2019 | 18:00:00 | Credit card | 8,6 |
| 21 | 319-50-3348 | B | Mandalay | Normal | Female | Home and lifestyle | 40,3 | 2 | 4,0 | 80,6 | 84,6300 | 3/11/19 | 15:30:00 | Ewallet | 4,4 |
| 22 | 300-71-4605 | C | Naypyitaw | Member | Male | Electronic accessories | 86,0 | 5 | 21,5 | 430,2 | 451,7100 | 2/25/2019 | 11:24:00 | Ewallet | 4,8 |
| 23 | 371-85-5789 | B | Mandalay | Normal | Male | Health and beauty | 88,0 | 3 | 13,2 | 263,9 | 277,1370 | 3/5/19 | 10:40:00 | Ewallet | 5,1 |
| 24 | 273-16-6619 | B | Mandalay | Normal | Male | Home and lifestyle | 33,2 | 2 | 3,3 | 66,4 | 69,7200 | 3/15/2019 | 12:20:00 | Credit card | 4,4 |
| 25 | 636-48-8204 | A | Yangon | Normal | Male | Electronic accessories | 34,6 | 5 | 8,6 | 172,8 | 181,4400 | 2/17/2019 | 11:15:00 | Ewallet | 9,9 |
| 26 | 549-59-1358 | A | Yangon | Member | Male | Sports and travel | 88,6 | 3 | 13,3 | 265,9 | 279,1845 | 3/2/19 | 17:36:00 | Ewallet | 6 |
| 27 | 227-03-5010 | A | Yangon | Member | Female | Home and lifestyle | 52,6 | 8 | 21,0 | 420,7 | 441,7560 | 3/22/2019 | 19:20:00 | Credit card | 8,5 |
| 28 | 649-29-6775 | B | Mandalay | Normal | Male | Fashion accessories | 33,5 | 1 | 1,7 | 33,52 | 35,1960 | 2/8/19 | 15:31:00 | Cash | 6,7 |
| 29 | 189-17-4241 | A | Yangon | Normal | Female | Fashion accessories | 87,7 | 2 | 8,8 | 175,3 | 184,1070 | 3/10/19 | 12:17:00 | Credit card | 7,7 |
| 30 | 145-94-9061 | B | Mandalay | Normal | Female | Food and beverages | 88,4 | 5 | 22,1 | 441,8 | 463,8900 | 1/25/2019 | 19:48:00 | Cash | 9,6 |
| 31 | 848-62-7243 | A | Yangon | Normal | Male | Health and beauty | 24,9 | 9 | 11,2 | 224 | 235,2105 | 3/15/2019 | 15:36:00 | Cash | 7,4 |
| 32 | 871-79-8483 | B | Mandalay | Normal | Male | Fashion accessories | 94,1 | 5 | 23,5 | 470,7 | 494,1825 | 2/25/2019 | 19:39:00 | Credit card | 4,8 |
| 33 | 149-71-6266 | B | Mandalay | Member | Male | Sports and travel | 78,1 | 9 | 35,1 | 702,6 | 737,7615 | 1/28/2019 | 12:43:00 | Cash | 4,5 |
| 34 | 640-49-2076 | B | Mandalay | Normal | Male | Sports and travel | 83,8 | 8 | 33,5 | 670,2 | 703,7520 | 1/10/19 | 14:49:00 | Cash | 5,1 |
| 35 | 595-11-5460 | A | Yangon | Normal | Male | Health and beauty | 96,6 | 2 | 9,7 | 193,2 | 202,8180 | 3/15/2019 | 10:12:00 | Credit card | 5,1 |

Figura 100: Base de datos.

7.6.1 Campos Calculados.

Cuando se crea un campo calculado en Tableau, este aparece ya sea como una medida o dimensión en el panel de datos.

Los cálculos consisten en un código compuesto de funciones, operaciones y referencias a otros campos, parámetros, constantes, grupos o conjuntos que devuelve un valor. El resultado de este código puede ocurrir por fila de datos, y otras veces, puede realizarse a nivel agregado.

Se puede crear un campo calculado de diferentes formas, desde la pestaña análisis en el menú o haciendo clic en cualquier campo del panel de datos. El campo calculado que se cree pasará a formar parte de la base de datos que se está utilizando en ese instante.

La interfaz para crear y editar campos calculados tiene este aspecto:

The screenshot shows the Tableau interface for creating a calculated field. It consists of three main parts:

- Field Name:** A text input field labeled "Nombre completo" (1) containing the formula "[Nombre] + ' ' + [Apellido]". Below it, a status indicator says "El cálculo es válido." (3).
- Function Selection:** A dropdown menu labeled "Todo" (4) is open, showing a list of functions including ABS, ACOS, AND, ASCII, ASIN, ATAN, ATAN2, ATTR, AVG, CASE, CEILING, CHAR, and COLLECT.
- Function Description:** A panel on the right shows the selected function "ABS(número)" (5) with the description "Indica el valor absoluto del número dado." and an example: "Ejemplo: ABS(-7) = 7".

At the bottom, there are two buttons: "Aplicar" and "Aceptar".

Figura 101: Crear un campo calculado.

1. Nombre del campo calculado: es el nombre con el que aparecerá el campo calculado en el panel de datos.
2. Editor de código: es donde se introduce el código que realizará el cálculo. El editor de código incluye opciones de autocompletar para campos que existan en la base de datos o funciones típicas.
3. Indicador de errores en el código.
4. Lista de funciones: contiene todas las funciones que se pueden utilizar en el código, Tableau agrupa estas funciones según su uso:
 - Número: Funciones matemáticas como el redondeo, valor absoluto, raíz cuadradas o funciones trigonométricas.
 - Cadena: funciones que son útiles para la manipulación de cadenas, como obtener una subcadena, encontrar una coincidencia dentro de una cadena, reemplazar partes de una cadena y convertir un valor de cadena en mayúsculas o minúsculas.
 - Fecha: funciones que son útiles para trabajar con fechas, como encontrar la diferencia entre dos fechas, agregar un intervalo a una fecha, obtener la fecha actual.
 - Conversión de tipos: funciones que son útiles para convertir un tipo de campo a otro, como convertir enteros en cadenas, decimales de coma flotante en enteros o cadenas en fechas.
 - Lógico: funciones de toma de decisiones, del tipo: "else" "if".
 - Agregación: funciones que se usan para agregar, como sumar, obtener los valores mínimos o máximos o calcular desviaciones o variaciones estándar.
 - Cálculo de tablas.
5. Al seleccionar una función en la lista o hacer clic en un campo, parámetro o función en el código se muestran detalles sobre la selección a la derecha.

En este capítulo se van a ver las diferencias y características de cada uno de los cálculos que se pueden realizar en Tableau.

La base de datos que se va a utilizar es de unos alquileres vacacionales, es una base de datos pequeña, para que se puedan apreciar y entender los cálculos que se realizarán sobre la misma.

Hay una columna para el nombre de la propiedad, los datos de los inquilinos, fecha de entrada y salida, descuento y el precio alquiler.

| | A | B | C | D | E | F | G |
|----|----------------|---------|----------|--------------|--------------|-----------|------------|
| 1 | Propiedad | Nombre | Apellido | Fecha Entrac | Fecha Salida | Descuento | Alquiler |
| 2 | 112-Miramar | María | Jiménez | 2-dic | 9-dic | 150,00 € | 1.500,00 € |
| 3 | 112-Miramar | Andrea | Murillo | 9-dic | 15-dic | 0,00 € | 1.500,00 € |
| 4 | 155-La palmera | Laura | Rabasco | 2-dic | 9-dic | 260,00 € | 1.300,00 € |
| 5 | 155-La palmera | Pilar | Peña | 16-dic | 23-dic | 280,00 € | 1.400,00 € |
| 6 | 207-La palmera | Jesús | Castro | 9-dic | 23-dic | 280,00 € | 2.800,00 € |
| 7 | 207-La palmera | Alfredo | Medrán | 2-dic | 9-dic | 0,00 € | 1.500,00 € |
| 8 | | | | | | | |
| 9 | | | | | | | |
| 10 | | | | | | | |

Figura 102: Base de datos alquiler vacacional.

CÁLCULOS A NIVEL DE FILA

En el campo Propiedad se tiene el nombre de la propiedad y el número de habitación, se van a separar creando dos campos calculados, uno llamado habitación y otro edificio. Esto se hace para poder realizar análisis a nivel de habitaciones y a nivel de los edificios.

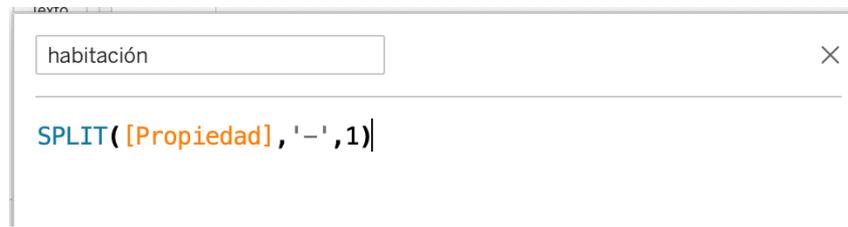


Figura 103: Cálculo a nivel de fila.

Este código quiere decir que para el campo "Propiedad", se va a utilizar el guion como separador y a la primera parte (la que está antes del guion) se le va a asignar el campo habitación y a la segunda, el campo edificio.

Tableau muestra que se trata de un campo calculado con el signo '=' antes del tipo de dato.

Como se ve, los cálculos se han realizado a nivel de fila.

| Apellido | Edificio | Fecha Entrada | Fecha Salida | habitación | Nombre | Nombre completo | Propiedad | Alquiler | Descuento | Número de registros |
|----------|------------|---------------|--------------|------------|---------|-----------------|----------------|----------|-----------|---------------------|
| Jiménez | Miramar | 2/12/2020 | 9/12/2020 | 112 | María | María Jiménez | 112-Miramar | 1.500 | 150 | 1 |
| Murillo | Miramar | 9/12/2020 | 15/12/2020 | 112 | Andrea | Andrea Murillo | 112-Miramar | 1.500 | 0 | 1 |
| Rabasco | La palmera | 2/12/2020 | 9/12/2020 | 155 | Laura | Laura Rabasco | 155-La palmera | 1.300 | 260 | 1 |
| Peña | La palmera | 16/12/2020 | 23/12/2020 | 155 | Pilar | Pilar Peña | 155-La palmera | 1.400 | 280 | 1 |
| Castro | La palmera | 9/12/2020 | 23/12/2020 | 207 | Jesús | Jesús Castro | 207-La palmera | 2.800 | 280 | 1 |
| Medrán | La palmera | 2/12/2020 | 9/12/2020 | 207 | Alfredo | Alfredo Medrán | 207-La palmera | 1.500 | 0 | 1 |

Figura 104: Cálculo a nivel fila en la base de datos.

CÁLCULOS AGREGADOS

En las gráficas que se han utilizado en los capítulos previos, se han visto las agregaciones típicas de Tableau: Suma, Máximo, Mínimo, Promedio. Estas agregaciones se pueden utilizar también para hacer cálculos más complejos, que se conocen como cálculos agregados.

Por ejemplo, si se quiere conocer el porcentaje de alquiler que ha sido descontado, ese campo no existe en la base de datos, y no se puede almacenar ya que cambia según el nivel de detalle que se tenga, es decir, no es lo mismo el porcentaje de descuento para un edificio, que para la habitación.

Se debe calcular como un agregado e ir recalculándose según cambie el nivel de detalle de la visualización.

Se va a crear un campo calculado que se llame Descuento (%) con el siguiente código:



Figura 105: Cálculo agregado.

Este código indica que la suma del descuento debe ser dividida por la suma del alquiler. Es decir, todos los valores de descuentos se añadirán y todos los de alquiler, una vez se haya hecho la suma, se realizará la división.

| Edificio | habitación | Nombre compl.. | Fecha Entra.. | Fecha Salida | Alquiler | Descuento | Descuento (%) |
|------------|------------|----------------|---------------|--------------|----------|-----------|---------------|
| La palmera | 155 | Laura Rabasco | 2/12/2020 | 9/12/2020 | 1.300 | 260 | 20% |
| | | Pilar Peña | 16/12/2020 | 23/12/2020 | 1.400 | 280 | 20% |
| | 207 | Alfredo Medrán | 2/12/2020 | 9/12/2020 | 1.500 | 0 | 0% |
| Miramar | 112 | Jesús Castro | 9/12/2020 | 23/12/2020 | 2.800 | 280 | 10% |
| | | Andrea Murillo | 9/12/2020 | 15/12/2020 | 1.500 | 0 | 0% |
| | | María Jiménez | 2/12/2020 | 9/12/2020 | 1.500 | 150 | 10% |

Figura 106: Cálculo agregado.

Si ahora se quiere ver el descuento por habitación en lugar de por huésped se puede eliminar de la visualización los campos de huésped y fechas, quedando:

| Edificio | habitación | Alquiler | Descuento | Descuento (%) |
|------------|------------|----------|-----------|---------------|
| La palmera | 155 | 2.700 | 540 | 20% |
| | 207 | 4.300 | 280 | 7% |
| Miramar | 112 | 3.000 | 150 | 5% |

Figura 107: Cálculo agregado.

Los valores de descuento han cambiado porque las agregaciones dependen de las dimensiones que estén definiendo el nivel de detalle de la visualización.

En el primer caso, el nivel de detalle está definido por los cinco campos que aparecen en el estante filas, entonces, el cálculo suma todo el alquiler para cada período de alquiler y todos los descuentos para el periodo de alquiler y luego los divide.

En el segundo caso, Edificio y Habitación redefinen el nivel de detalle. De forma que el cálculo suma todos los precios de cada edificio y habitación y todos los descuentos para cada edificio y habitación y luego los divide.

La diferencia entre los cálculos a nivel de fila y los cálculos agregados es muy importante. Para que se entienda bien esta diferencia se va a realizar el mismo cálculo del descuento (%), pero a nivel fila.

Con el código `[Descuento] / Alquiler` en lugar de `SUM([Descuento]) / SUM([Alquiler])`

Anexo: La interfaz de Tableau.

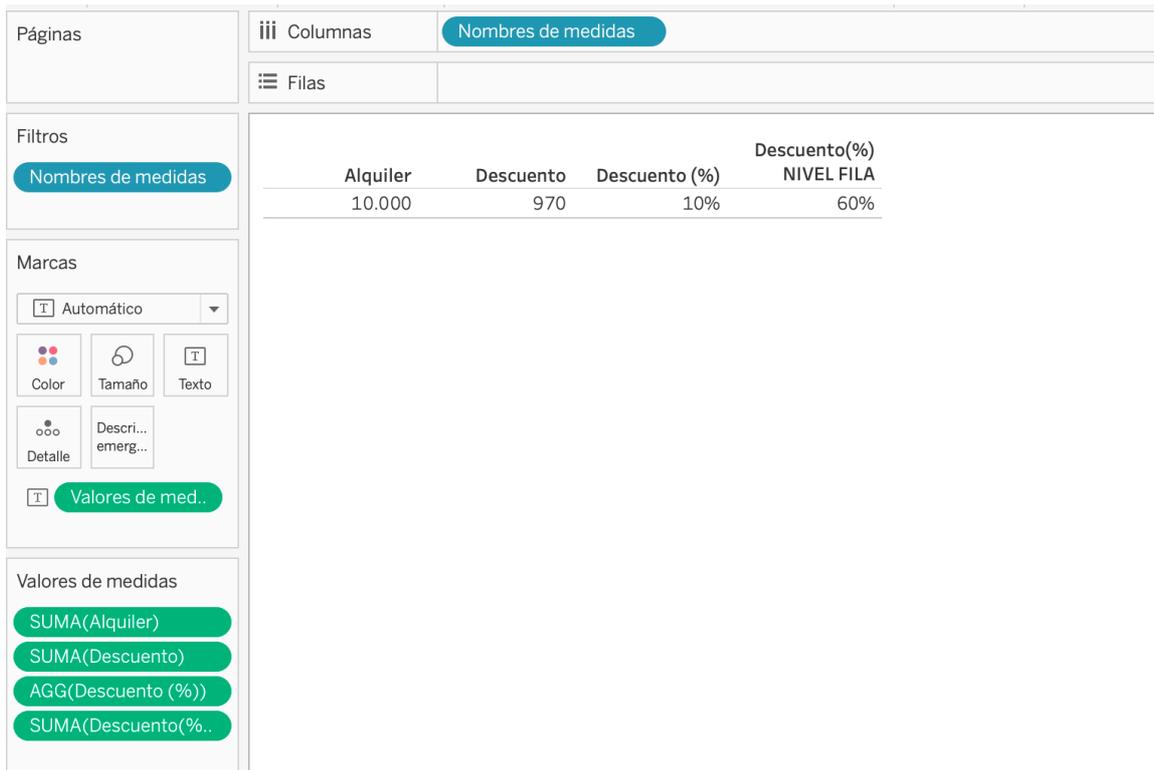


Figura 108: Nivel fila vs agregado.

En los valores de medidas, el cálculo a nivel de fila aparece con SUM en lugar de AGG. Esto significa que lo que Tableau hace es, calcular el descuento fila a fila y luego sumarlo para mostrarlo en la tabla, el 60% es la suma de cada porcentaje que se calculó fila a fila, frente al 10% que se ha calculado sumando todos los descuentos divididos por todos los alquileres.

| A | B | C | D | E | F | G | H | I |
|---|---------|----------|--------------|--------------|-----------|------------|---------------|---|
| Propiedad | Nombre | Apellido | Fecha Entrac | Fecha Salida | Descuento | Alquiler | descuento (%) | |
| 112-Miramar | María | Jiménez | 2-dic | 9-dic | 150,00 € | 1.500,00 € | 10,00% | |
| 112-Miramar | Andrea | Murillo | 9-dic | 15-dic | 0,00 € | 1.500,00 € | 0,00% | |
| 112-Miramar | Andrea | Rabasco | 2-dic | 9-dic | 260,00 € | 1.300,00 € | 20,00% | |
| 155-La palmera | Pilar | Peña | 16-dic | 23-dic | 280,00 € | 1.400,00 € | 20,00% | |
| 207-La palmera | Jesús | Castro | 9-dic | 23-dic | 280,00 € | 2.800,00 € | 10,00% | |
| 207-La palmera | Alfredo | Medrán | 2-dic | 9-dic | 0,00 € | 1.500,00 € | 0,00% | |
| | | | | | | | 60% | |
| <p>Nivel fila: se calculan los resultados fila a fila y ,por último se suman al nivel de detalle de la visualización</p> | | | | | | | | |

Figura 109: Cálculo a Nivel fila

| A | B | C | D | E | F | G | H | I |
|----------------|---------|----------|--------------|--------------|-----------|-------------|---------------|---|
| Propiedad | Nombre | Apellido | Fecha Entrac | Fecha Salida | Descuento | Alquiler | descuento (%) | |
| 112-Miramar | María | Jiménez | 2-dic | 9-dic | 150,00 € | 1.500,00 € | 10,00% | |
| 112-Miramar | Andrea | Murillo | 9-dic | 15-dic | 0,00 € | 1.500,00 € | 0,00% | |
| 155-La palmera | Laura | Rabasco | 2-dic | 9-dic | 260,00 € | 1.300,00 € | 20,00% | |
| 155-La palmera | Pilar | Peña | 16-dic | 23-dic | 280,00 € | 1.400,00 € | 20,00% | |
| 207-La palmera | Jesús | Castro | 9-dic | 23-dic | 280,00 € | 2.800,00 € | 10,00% | |
| 207-La palmera | Alfredo | Medrán | 2-dic | 9-dic | 0,00 € | 1.500,00 € | 0,00% | |
| | | | | | 970,00 € | 10.000,00 € | 10% | |

suma(descuento)
 suma(alquiler)
 $\frac{\text{suma(descuento)}}{\text{suma(alquiler)}}$

Figura 110: Cálculo Agregado.

7.6.2 Cálculos de Tabla.

Los cálculos de tabla son una de las herramientas más poderosas de Tableau, ya que permiten:

- Comparar y realizar cálculos sobre valores agregados en las filas de la tabla resultante.
- Abrir posibilidades para el análisis y enfoques creativos para resolver problemas o resaltar ideas.

La complejidad de estos varía, existen cálculos que se crean de manera muy sencilla, con un par de clics, y otros, que requieren una mayor comprensión de lo que se está haciendo y de la información que se quiere extraer. Un campo con un cálculo de tabla activo se identifica con el símbolo delta (Δ), por ejemplo: SUMA (sales) Δ .

Para ver como se usan los cálculos de tabla se usará la base de datos de Excel *Superstore*, que ya se ha utilizado en capítulos previos.

- o **Cálculos de tabla rápidos**

Los cálculos de tabla rápidos son cálculos predefinidos que pueden ser aplicados a cualquier campo que esté siendo usado como una medida en la visualización. Algunos de los cálculos de tabla rápidos disponibles son: total acumulado, diferencia, diferencia de porcentaje, o clasificación. Utilizando la base de datos *Superstore*, se muestran algunos ejemplos de los cálculos de tabla rápidos:

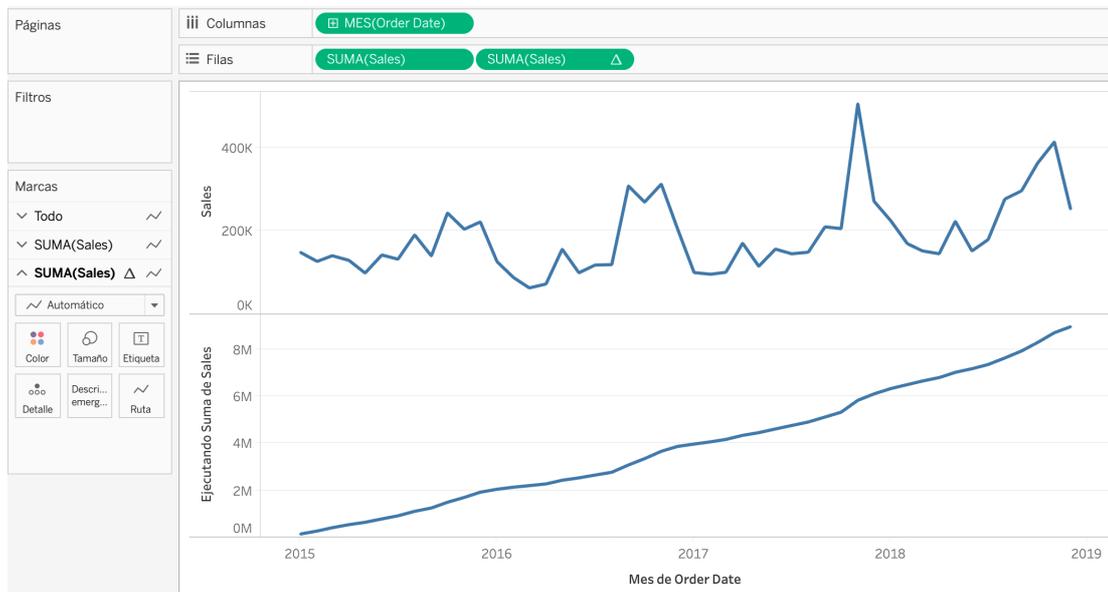


Figura 111: Cálculos de tabla rápidos.

En la Figura 111 se representan dos gráficos, en el que está arriba se ve la suma de ventas para cada fecha y en el de abajo el total acumulado de la suma de ventas según la fecha.

En el siguiente ejemplo se ven otros de los cálculos de tabla rápidos que hay disponibles

| Año de Order Date | Trimestre de Order Date | Sales | Ejecutando Suma de Sales junto con Tabla (abajo) | Diferencia en Sales desde Anterior junto con Tabla (abajo) | Clasificación de Sales junto con Tabla (abajo) |
|-------------------|-------------------------|-----------|--|--|--|
| 2015 | T1 | 417.555 | 417.555 | | 12 |
| | T2 | 372.289 | 789.844 | -45.266 | 13 |
| | T3 | 464.319 | 1.254.163 | 92.030 | 10 |
| | T4 | 670.182 | 1.924.345 | 205.863 | 5 |
| 2016 | T1 | 279.148 | 2.203.493 | -391.034 | 16 |
| | T2 | 330.269 | 2.533.762 | 51.121 | 14 |
| | T3 | 546.875 | 3.080.637 | 216.606 | 7 |
| | T4 | 788.255 | 3.868.892 | 241.380 | 3 |
| 2017 | T1 | 298.848 | 4.167.740 | -489.407 | 15 |
| | T2 | 443.764 | 4.611.504 | 144.916 | 11 |
| | T3 | 505.453 | 5.116.957 | 61.689 | 9 |
| | T4 | 982.675 | 6.099.632 | 477.222 | 2 |
| 2018 | T1 | 547.656 | 6.647.288 | -435.019 | 6 |
| | T2 | 521.650 | 7.168.938 | -26.006 | 8 |
| | T3 | 752.933 | 7.921.871 | 231.283 | 4 |
| | T4 | 1.030.156 | 8.952.027 | 277.223 | 1 |

Figura 112: Cálculos de tabla rápidos II.

○ CONCEPTOS IMPORTANTES DEL CÁLCULO DE TABLAS:

Los cálculos de tabla se pueden computar o relativos a la tabla o fijados por una de las dimensiones de la tabla:

⇒ Relativo a la tabla; alcance y dirección:

Los cálculos de tabla se computarán en función del layout o diseño de la tabla. Se moverán a lo largo o hacia abajo en la tabla, si se cambia la distribución de la tabla los cálculos se verán afectados.

Alcance y dirección: estos términos describen como se ha computado un cálculo de tabla.

El alcance define los límites en los que se ejecutará el cálculo, estos pueden ser la tabla completa, un panel o una celda, la dirección indica como se mueve el cálculo en la tabla (abajo, a lo largo, abajo y luego a lo largo, a lo largo y después abajo)

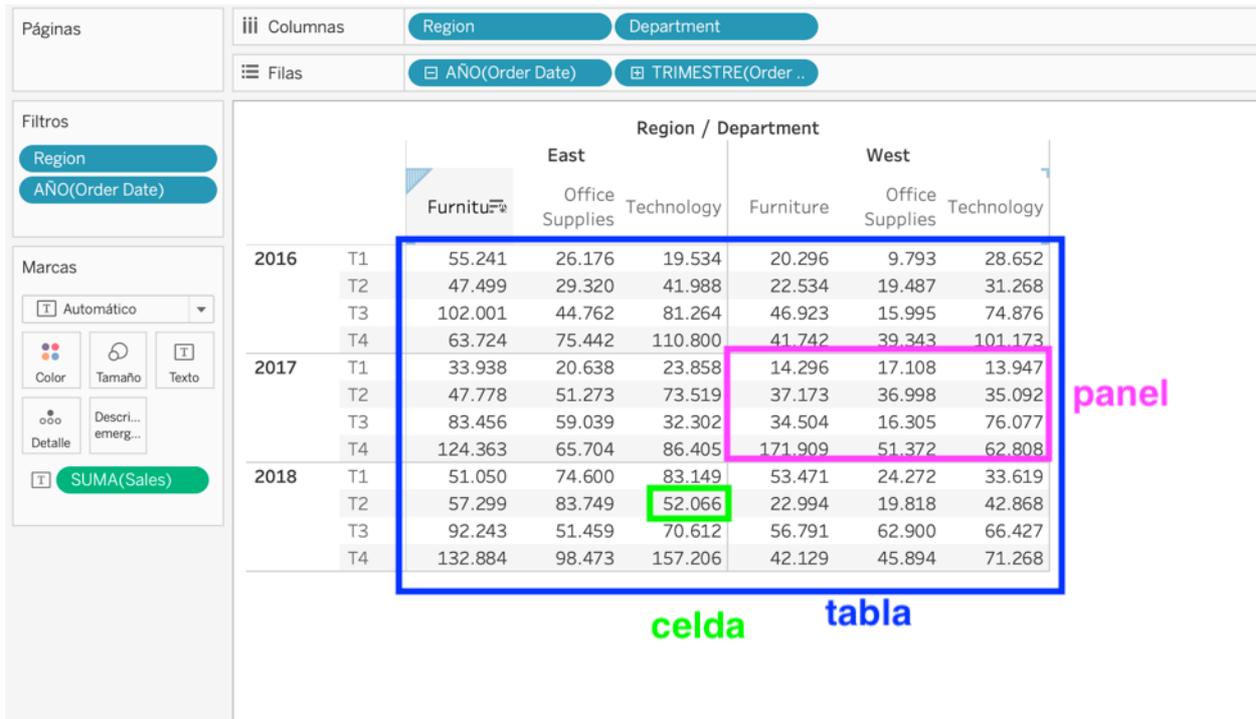


Figura 113: tabla, panel y celda

⇒ Fijado por una dimensión; direccionamiento y partición:

El direccionamiento y la partición son conceptos muy parecidos a los vistos previamente (alcance y dirección) con la diferencia en el direccionamiento y partición los cálculos no están referenciados al layout de la tabla, sino que se relacionan con ciertas dimensiones para el direccionamiento y con otras para la partición.

Estos cálculos dan un control más preciso, ya que se desligan del diseño de la tabla para realizar los cálculos, es decir, los resultados se mantienen aún cambiando la orientación de la table o el orden en que se muestran los campos.

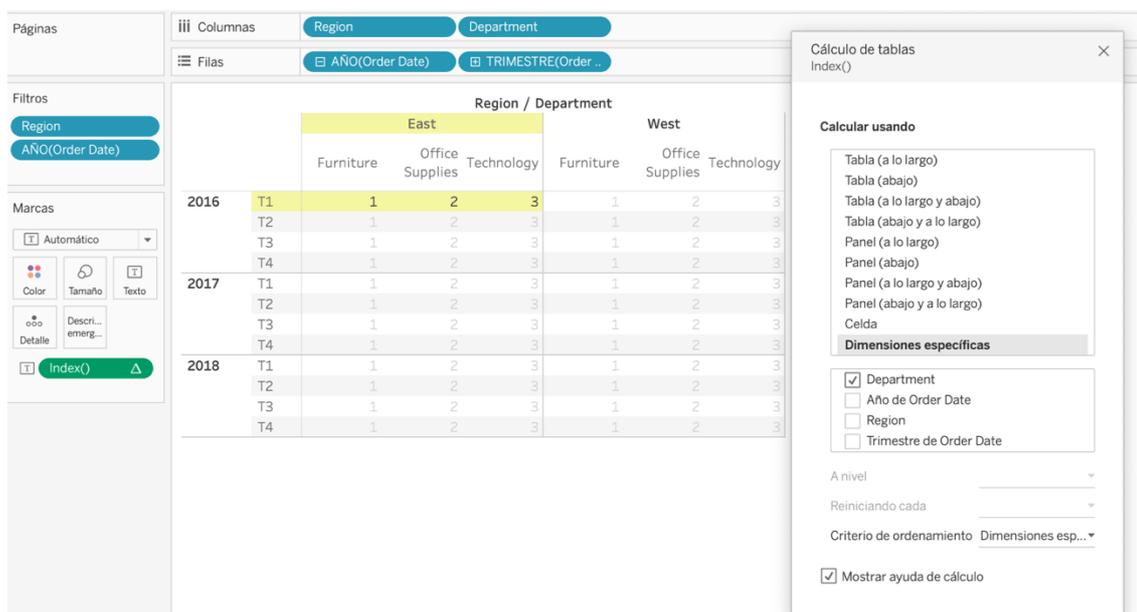


Figura 114: Direccionamiento y partición.

