Tableau:

Dimensión: siempre vienen marcadas, son datos no numéricos, pueden ser numéricos, pero en general son string o texto, son dato que dan cierta característica a las métricas, nos dan posibilidad de dar mas detalle a esa métrica.

Ejemplo: el producto, como silla oficina = x, silla gamín = y, se puede diferenciar el producto, se puede diferenciar en diferentes tiendas.

Métricas: las métricas siempre van a ser un número, una compra, una venta, un precio.

Ejemplo compras y ventas

Kpi: define una característica propia de nuestros datos en un objetivo en concreto que queremos buscar en cuanto a nuestras dimensiones y métricas, es decir es un campo calculado que nosotros realizamos a partir de nuestras métricas y nuestras dimensiones que nos da nuestra base de datos y nos va a dar algo de valor de acuerdo a nuestra información.

Ejemplo: en Airbnb, cuando hacemos una búsqueda muestra el precio, pero cuando marco la vivienda, sale el precio, la limpieza, la fianza, un kpi sería el precio final del apartamento, el cual contiene, el precio, la fianza y el precio de limpieza, al usuario final, le saldría este dato como precio final.

En las sillas el precio del ratio del beneficio de mi empresa, dividimos las compras y las ventas, obtenemos un ratio de beneficio que podemos obtener, un ratio de ganancia.

Al final, son diferentes operaciones con los datos que puedo realizar a partir de las dimensiones y las métricas.

El =# significa que es un campo calculado

Veremos las diferentes conexiones de Tableau, que son conexión al servidor y a un archivo.

Finalmente, la parte de visualización de datos, que es la parte donde se nos muestran los datos.

1. Introducción a Tableau

Arquitectura para el análisis de datos:

* ETL: Extracción, transformación y carga del dato.
* BIG DATA: almacenaje y modernización del dato.
* BUSINESS INTELLIGENCE: Consulta y presentación del dato.
* ETL + BIG DATA + BI = MACHINE LEARNING: Descubrir y Enriquecer el dato.

Con tableau estamos en la ultima parte, donde ya deben estar los datos limpios, donde toda la información es tratada, donde se han transformado todos los campos, se han hecho todos los cálculos necesarios para mi análisis, ya tengo toda esta información en un datawerehouse limpio y ya puedo consultar esa información y pintarla con mi gráfico.

Cuando la empresa no tiene los recursos para todo lo anterior, aquí ya debemos mirar como nos podemos apañar con Tableau para organizar la información, pero esto es a coste del rendimiento, cuantas mas tareas tenga que hacer Tableau, mayor será el coste de mi dashboard, en este caso Tableau estaría al nivel de BI= consulta y presentación del dato.

Tenemos el eje x, eje y, el eje x, Garner lo define como la integridad de visión, representa el conocimiento de distintos proveedores sobre como se puede aprovechar el momento del mercado actual para generar un valor para sus clientes.

El eje y, Garner lo define como la capacidad de ejecución, se mide la habilidad de los fabricantes de los proveedores de software, de ejecutar con éxito su particular visión del mercado, es decir, al final como se usan las habilidades de marketing y como se captan clientes.

Te tienen 4 cuadrantes: los lideres, es el máximo del eje y y x, aquí se encuentran los que tienen la mayor parte del mercado.

Tenemos los proveedores o aspirantes, están centrados en una única solución, es decir una única demanda del mercado, no están centrados al público en general.

Los visionarios son parecidos a los lideres, pero no disponen de los medios suficientes para disponer de la mayor cantidad del mercado.

Los jugadores de nicho: son los que están empezando a hacer cosas en el mundo de la tecnología pero aun falta habilidades de ejecución, y una integridad de visión y aun falta camino para poder llegar a cuadrantes más elevados.

La gran diferencia entre Tableau y power bi, es que al power bi tener office integrado, hace que la gran mayoría use este servicio.

Tableau quiere

* Liberar datos
* Capacitar personas
* Diseñar a las personas

Debe tener una buena experiencia de usuario, y que al final se pueda compartir con diferentes perfiles técnicos o perfiles de la empresa y puedan ser utilizados y todos puedan colaborar de esos informes y se puedan sacar conclusiones de los datos de negocio que tenemos mostrados en los gráficos.

Tableau desktop es la herramienta de los diseñadores, permite extraer y consultar información de mis bases de datos crearla en tableau mediante un datasourse y una vez cargada la información, ya puedo crear mis informes.

Yo creo mi conexión a la base de datos y me traigo la información a Tableau.

En tableau

podemos conectarnos a Tableau server, solo si TENGO UN Tableau cloud.

Puedo conectarme a un archivo o a un servidor, cuando nos vamos a conectar a un servidor, tengo infinidad de controladores.

Seleccionamos archivo de texto, guardamos el libro de trabajo

En la parte superior derecha tenemos la tabla lógica y si damos doble click, tenemos la tabla física y en información dice una tabla física define la tabla lógica.

En la cinta que divide la parte de arriba y abajo están las opciones de tabla lógica y las relaciones

En la parte izquierda tengo el nombre de los campos en la fuente de datos

Abc: tipo string

# tipo numérico

Calendario es información tipo fecha

En la parte superior derecha da la opción para acceder a la información en tiempo real o extraer los datos y traerlos en un momento concreto.

Surgen dos escenarios,

Tengo una solución Big Data, una estructura de datos en tiempo real muy bien implementada, el rendimiento va a ser adecuado, las consultas rápidas y ver la información en tiempo real.

No tengo una solución Big Data, voy a realizar una solución mediante extractos, cuando suba mi dashboard, solicito actualizar la información para que los usuarios cuando se conecten puedan ver la información hasta el día anterior de todas las ventas, compras etc.

Esto se entiende que, si estoy trabajando en tiempo real, estoy accediendo directamente a la base de datos.

Tableau separa mi fuente de trabajo en azul las dimensiones y en verde las métricas.

Ordenamiento computerizado: La forma de ordenación anidada se usa cuando tenemos diferentes dimensiones que irá hasta el ultimo nivel y me ordenará al nivel del tipo que yo decida en orden ascendente o descendente.

Puedo crear carpetas y así poder tener datos mas ordenados y separar la información relevante para mí.

Jerarquías en tableau: Se crea a partir de arrastrar y soltar datos, uno sobre otro, estas permiten navegar sobre la información funciona como un árbol (padre con hijos), esto es útil ejemplo para saber que tiendas están vendiendo debajo del coste de mi producto.

Ejemplo, arrastramos room type sobre property type, en la columna izquierda de tablas, puedo añadir mas campos si es necesario.

Ejercicio 1:

Análisis del precio y depósitos o fianzas. ¿Cómo han evolucionado los precios y finanzas de manera mensual a lo largo de los dos últimos años? Queremos representar esta información de manera que permita contrastar los precios y finanzas con los del año anterior.

En este ejercicio comparamos el precio para 2015 y 2016 y la fianza para 2015 y 2016

* En primer lugar nos está pidiendo analizar dos métricas las cuales son:
* Precio en el dataset se llama price
* Deposito o fianza en el dataset se llama security deposit
* Y la dimensión sería el host since
* Tenemos que ver una evolución en el tiempo de estos valores de manera mensual, a lo largo de los dos últimos años.
* Entonces el precio se compara de manera mensual y de igual manera se hace con la fianza.
* Debemos interponer en el mismo grafico los dos años a evaluar.
* En métricas seleccionamos host since por año
* Luego llevamos host since a filtros, seleccionamos 2015 y 2016.
* Finalmente arrastramos de tablas al panel de todo el host since en la sección color

Todo esto hace que se desagrupe precio y fianza

Ejercicio 2:

Construir una vista que muestre claramente cuál es la relación que existe entre los precios y las fianzas a lo largo del año. Mostrar esta relación para 2015 y 2016

En este caso vamos a hacer la comparativa para dos medidas diferentes, en un mismo año comparamos precio y fianza y para el otro año, hacemos lo mismo.

* Arrastramos a métricas el host since y filtramos por mes
* A dimensiones arrastramos a Price y security deposit
* Arrastramos host since a filtros y seleccionamos 2015 y 2016
* Arrastramos host since a filas para obtener el promedio por año del precio y la fianza
* Vamos a la medida que esta mas a la derecha, damos en botón derecho y seleccionamos eje doble.
* Nos aparece el promedio de precio en azul y la fianza en naranja para cada año.
* Me voy al eje a la derecha, pulso botón derecho y sincronizar eje
* Luego nuevamente botón derecho, quitamos la selección a mostrar eje
* Vamos al eje de la izquierda, seleccionamos con botón derecho y le damos editar eje y quitamos el título, ya que necesitamos mostrar en euros las diferentes métricas.

Aquí podemos visualizar en cada año, la fianza mayor al precio.

Ejercicio 3

¿Cuál es la relación entre el precio y el tamaño de las viviendas?

* Esta determinado entre price y square feet.
* Vamos a determinar ahora en barra esta Price y en burbuja esta square feet.
* En métricas Creamos un promedio de precio, un square feet
* En dimensiones arrastramos la jerarquía de tipo de estancia creada room type y bed type.
* Botón derecho en la ultima media y le damos eje doble.
* En Price seleccionamos barras
* Como las unidades de medida de los ejes Price y square feet son diferentes debemos dejarlas tal cual

**Para agrupar las métricas**

En el tablero, selecciono las barras que deseo agrupar, con click + control y selecciono el clip que significa agrupar, automáticamente se me crea una métrica que se llama bed type grupo

Los conjuntos

Permiten crear un subconjunto de datos a partir de unas condiciones, ejemplo, quiero las camas que estén por debajo de los 30 euros.

* Bueno seleccionamos como dimensión Price y como métrica bed type
* Seleccionamos los que no sean camas y le damos crear conjunto y aceptar.
* El conjunto me sirve para usarlo como un filtro.
* Ese conjunto creado lo colocamos en filtro
* Le doy click derecho selecciono mostrar conjunto dentro o fuera, si selecciono dentro, me muestra todos los elementos de esa métrica, si le digo fuera, me muestra solo las camas.

Crear conjunto de manera calculada

Significa ingresar sus miembros acordes a alguna condición.

Seleccionamos en tablas la métrica que queremos crear el conjunto, botón derecho, crear conjunto y vamos a condición

Seleccionamos en el caso del ejercicio, promedio del precio que sea menor o igual a 70 para los barrios.

Si llevo este campo a filtro y le digo mostrar los barrios que estén fuera de la condición, me muestra los que estén por encima de 70 euros.

**Ejercicio 4**

A partir del ejercicio anterior, modificar el análisis para:

Agrupar los tipos de vivienda que tienen menos de 20 reviews en una única categoría.

Mostrar los precios por tipo de habitación.

* Seleccionamos métricas promedio Price y promedio number of rewiews y en dimensiones property type
* Nos paramos al final de las métricas y damos eje doble.
* Vamos a promedio de precio y seleccionamos barra.
* Ahora seleccionamos precio, vamos a room type y arrastramos a color
* Revisamos la categoría y seleccionamos todas las que están por debajo de 20
* Editamos el nombre para distinguir la agrupación

Ejercicio 5

Realizar una visualización que muestre el precio y metro cuadrado por barrio y tipo de habitación, además debe existir la posibilidad de navegar tipo de habitación y tipo de cama.

* Barrio en las filas
* Precio y room type en las columnas,
* Square feet a color

Ejercicio 5

Construir un informe donde se muestre la tendencia del precio de cada barrio mes a mes.

* Promedio de precio en filas.
* Columnas host since
* Se desagrupa la información a nivel de room type, moviéndolo a color y así queda desagrupado.
* Para filtrar por barrio muevo a neigborhood y property type a filtros

división de datos por fecha:

por defecto Tableau a un campo de tipo fecha te lo crea como una jerarquía de tipo temporal, si se da click en el campo padre, crea automáticamente una jerarquía hija.

Tenemos fechas continuas, que es una progresión cronológica del tiempo

Fechas discretas que son, promedios mensuales, trimestrales, anuales etc.

En tableau

* las primeras fechas que nos encontramos es discreta color azul
* las de más abajo son continas de color verde.

Mapeo geográfico:

Pintar por barrio la longitud y latitud.

* Se convierte latitud y longitud a float (castear)
* Se le da formato función geográfica a cada uno según corresponda
* Se coloca siempre longitud en columnas y latitud en filas de forma cardinal (x,y)
* Finalmente se anexa jerarquía geográfica a detalle debajo de marcas.
* Price se anexa a color en forma de promedio, para pintar las bolas del código postal de un color mas intenso a menos intenso, dependiendo del precio.
* Se añade promedio de precio a tamaño, obtengo las bolas mas grandes y de color mas intenso para los precios más altos.

Para meterle más cosas al mapa

Vamos a Tableau, mapa, capas de segundo plano

Escogemos el estilo deseado, claro, oscuro o normal

En este caso seleccionaremos a nivel de calles o satélite.

Primer Problema encontrado, cuando demostramos tanta información, al final, nos cuesta distinguir nuestra información a tener en cuenta, se concluye que con esto dificultamos al usuario sacar una conclusión en concreto.

Ejercicio 7

Construir un mapa que muestre donde los pisos tienen más metros cuadrados y que muestren donde son más caros.

* Seleccionamos longitud float en columnas
* Seleccionamos latitud float a filas
* Arrastramos Jerarquía geográfica zipcode a detalle
* Dimensión Price, a color
* métrica square feet a tamaño

Nota: cuando agrego medidas, desagrupo la información con dimensiones, al añadir barrios, vamos a tener mas puntos en el mapa.

Ejercicio 8

Para Madrid, mostrar el precio en tamaño y como color de fondo el \_\_\_ de cada zip code. Mostrar además cómo evolucionan los precios de un trimestre a otro ¿Se podría implementar el pie chart con un tooltip con la información de cancelación?

* Seleccionamos longitud float en columnas
* Seleccionamos latitud float a filas
* Arrastramos Jerarquía geográfica zipcode a detalle
* En marca seleccionamos mapa
* Dimensión Price, a color y a tamaño
* Implementar en el tooltip el gráfico
* Creo una nueva hoja
* Busco cancellation policy la movemos a color y a texto
* Seleccionamos marcas y elegimos grafico circular
* Seleccionamos promedio de precio arrastrado a tamaño, para que cada color, sea el promedio del precio
* Arrastramos precio a etiqueta siempre sacando promedio
* Arrastramos zip code a detalle
* Vuelvo a la hoja del ejercicio
* Selecciono descripción emergente
* Me paro en el último renglón doy enter
* Insertar, hojas, ejercicio 8 pie chart
* Aceptar

Crear territorios en un mapa

Podemos agrupar diferentes áreas ejemplo a Madrid diferenciarla por Madrid sur y Madrid norte:

* Seleccionamos longitud float en columnas
* Seleccionamos latitud float a filas
* Arrastramos Jerarquía geográfica zip code a detalle
* En marca seleccionamos mapa
* En el mapa de código postal seleccionamos lo que queremos escoger, seleccionamos el clip y pinchamos en código postal

Panel de análisis

Parámetros son los valores dinámicos que pueden reemplazar varias constantes:

* Canel izquierdo en la flecha, click derecho, crear parámetro.
* Seleccionamos la opción deseada.

Ejercicio 9

Volviendo a la clasificación del ejercicio 4, con el objetivo de centrar atención en los tipos de vivienda con más volumen de mercado, completar la vista añadiendo una línea de referencia.

* Room type a columnas
* Suma precio a filas y suma number of reviews, en este final clic derecho eje doble
* En precio seleccionamos barras, en reviews circulos
* Arrastramos room type a color de Price
* Análisis, línea de referencia, arrastro y suelto a nivel de tabla y precio
* Seleccionamos línea, constante e introducimos 50.000, selecciono el formato de la línea y aceptar

Ejercicio 10

Mostrar el promedio del precio por tipo de habitación y cancelación, añadiendo líneas de referencia que señalan el máximo y el promedio.

* Arrastramos a filas el promedio del precio, cancellation policy
* A columnas coom type y bed type
* Análisis, arrastra líneas de referencia a nivel de panel, banda
* Desde promedio precio promedio, valor, ninguno
* Hasta promedio de precio máximo, valor, ninguno

Ejercicio Adicional 1

¿Cuál es el top 5 de barrios de Madrid con volumen de alquiler? ¿que política de restricción tiene cada uno?

El top n es un filtro que permite filtrar por N valores mis gráficos y mostrarlos en el orden de mayor a menor

* Suma Price a columnas
* neigborhood a filas
* Cancellation policy a color
* Click derecho en null y excluir
* En filtros editar filtro y limites
* Por campo, limite (parte superior), crear nuevo parámetro y le damos en este caso el valor 5, así nos saca los 5 más altos
* Vista completa y selecciono ver texto

Ejercicio adicional 2

¿Cuál es el top 5 dentro de cada política de cancelación separando por el tipo de alojamiento?

* Suma Price a columnas
* Cancellation policy y neigborhood a filas
* Property type a color
* Click derecho en null y excluir
* En filtros editar filtro y limites
* Por campo, limite (parte superior), crear nuevo parámetro y le damos en este caso el valor 5, así nos saca los 5 más altos
* Vista completa y selecciono ver texto

Ejercicio adicional 3

¿Cuál ha sido la evolución del precio y el depósito a lo largo de los años?

* Promedio de security deposit en filas
* Promedio de precio en columnas
* Cancellation policy a color
* Prom Price a tamaño
* Neigborhood y prom price a etiqueta
* Animation chart: Host since se arrastra a pagina (Tableau tomaría cada pagina como cada año)

Análisis

Regresión

La línea de tendencia muestra el conjunto de datos, muestra como están relacionadas una medida frente a otra.

En Tableau la más típica es el modelo lineal:

La regresión lineal (El enlace se abre en una ventana nueva) (también conocida como regresión de mínimos cuadrados ordinarios, u OLS) se suele utilizar cuando hay uno o más predictores que tienen una relación lineal entre la predicción y el destino de predicción, no se ven afectados por las mismas condiciones subyacentes y no representan dos instancias de los mismos datos (por ejemplo, ventas expresadas en dólares y euros). La regresión lineal es el modelo predeterminado para las funciones de modelado predictivo en Tableau; si no especifica un modelo, se utilizará la regresión lineal. Puede especificar explícitamente este modelo mediante la inclusión de "model=linear" como primer argumento en el cálculo de la tabla.

* Análisis
* línea de tendencia,
* lineal

esta determina la formula del algoritmo que dice que la fianza es = a dos factores a \* promedio de precio + b

Pronostico:

Los pronósticos en Tableau utilizan una técnica conocida como homogenización exponencial. Los algoritmos de pronóstico tratan de buscar un patrón regular en medidas que puedan continuarse en el futuro. Si le interesa el modelado predictivo, también disponible en Tableau, consulte Funciones de modelado predictivo en Tableau.

Normalmente se añade un pronóstico a una vista que contiene un campo de fecha y, como mínimo, una medida. Sin embargo, si no hay ninguna fecha, Tableau puede crear un pronóstico para una vista que contenga una dimensión con valores enteros y, como mínimo, una medida.

Tenemos

* Filas host since año
* Columnas Price promedio
* Análisis
* Pronostico
* Añadir un pronostico

En este caso nos pronostica los dos años siguientes con tendencia a la baja

Todos son algoritmos por defecto en Tableau, si queremos otros algoritmos, debemos importarlos o crearlos.

Clústeres:

Nos permite distinguir entre diferentes grupos en un diagrama de dispersión en función de las medidas que queramos.

El análisis de clúster divide las marcas de la vista en clústeres, donde las marcas de cada clúster son más similares entre sí de lo que lo son con respecto a las marcas de otros clústeres.

U ableau usa el algoritmo k-means para la agrupación en clústeres. Para un número dado de clústeres k, el algoritmo divide los datos en k clústeres. Cada clúster tiene un centro (centroide) que es el valor promedio de todos los puntos de ese clúster. K-means ubica centros a través de un procedimiento iterativo que minimiza las distancias entre los puntos individuales de un clúster y el centro del clúster. En Tableau puede especificar el número de clústeres que quiera o bien hacer que Tableau pruebe distintos valores de k y que sugiera el número óptimo de clústeres (consulte Criterios utilizados para determinar el número óptimo de clústeres).

K-means requiere una especificación inicial de los centros de los clústeres. Empezando con un clúster, el método elige una variable cuya media se utiliza como un umbral para dividir los datos en dos. Los centroides de estas dos partes se utilizan para inicializar k-means para optimizar la membresía de los dos clústeres. A continuación, se elige uno de los dos clústeres para dividirlo y una variable dentro del clúster, cuya media se utiliza como umbral para dividir el clúster en dos. K-means se utiliza para dividir los datos en tres clústeres, inicializado con los centroides de las dos partes del clúster partido y el centroide del clúster que ha quedado. Este proceso se repite hasta que se alcanza un número determinado de clústeres.

Tableau utiliza el algoritmo de Lloyd con las distancias cuadradas euclídeas para calcular el agrupamiento k-means para cada k. Combinado con el procedimiento de división para determinar los centros iniciales para cada k > 1, el clúster resultante es determinista, ya que el resultado depende únicamente del número de clústeres.

En primer lugar, el algoritmo selecciona los centros de clúster iniciales, el algoritmo k-means, se usa bastante.

Luego divide las marcas asignando cada una de ellas a su centro más cercano:

A continuación, acota los resultados calculando nuevos centros para cada división mediante el cálculo del promedio de todos los puntos asignados al mismo clúster:

Luego revisa la asignación de las marcas a los clústeres y reasigna cualquier marca que ahora esté más cercana a otro centro que antes.

Los clústeres se redefinen y las marcas se reasignan de forma iterativa hasta que no se produzcan más cambios

* Prom Price a columnas
* Security deposit a filas
* A tamaño Price
* A texto neigborhood prcice promedio
* Análisis, cluster, clusters

Separa grupos por promedio de precios.

Personalización de los datos por medio de crear campo calculado

Crear campo calculado:

En realidad, los campos calculados son nuevos campos, añadimos a través de tableau nuevos campos a través de medidas dimensiones etc

Flecha de tablas y crear campo calculado