

Planificación de la Visualización: Informe Final

Matheus Vieira Santos (matheus.vieira23@estudiantes.uva.es)

17/11/2023

1 RESUMEN

Este proyecto se enfoca en visualizar y comprender los patrones de compra de los consumidores a partir de un conjunto de datos sintético llamado "Customer Shopping Preferences." La visualización tiene como público objetivo a empresas interesadas en mejorar sus estrategias de marketing y logística. Se desarrollarán dos gráficos interactivos: uno que muestra las ventas por región geográfica y otro que analiza cómo se distribuyen las ventas por categoría y estación del año en un estado específico. El enfoque es pragmático y analítico, y se utilizará la API D3.js para crear la visualización.

2 INTRODUCCIÓN

Dada la profunda influencia que la era digital ha ejercido en nuestra sociedad, es evidente que los consumidores no han permanecido ajenos a dichos cambios. En la actualidad, ellos están más informados, críticos, objetivos y demandan una comunicación recíproca con las marcas. De esta forma, reconocer patrones de conducta y comprender su comportamiento durante el proceso de compra se convierte en una poderosa herramienta para las empresas [1].

En ese sentido, se eligió trabajar con el conjunto de datos "Customer Shopping Preferences" de Kaggle [2]. Una vez que este conjunto de datos representa las tendencias de compra de los clientes, lo que se busca en la visualización es representar esos comportamientos y patrones de manera visual, lo que hace con que la interpretación de los datos sea más intuitiva y sencilla.

Para más, se desea crear una visualización que no solo tenga aspectos exploratorios, revelando informaciones generales sobre las preferencias de compra de los clientes, sino que también, con solo algunos clics, pueda proporcionar información más específica, como lo que se vende, para qué público y en qué épocas del año se realizaron esas compras.

Para lograrlo, este documento está organizado de la siguiente manera: primero, tenemos una sección de análisis de los datos, donde se proporciona una descripción detallada del conjunto de datos. Después, tenemos una sección de planeamiento de la visualización, en la cual se explica más sobre la metodología utilizada para crear la visualización y por qué se la construyó de esa manera. A continuación, en la sección de diseño, se describe el gráfico desarrollado y se muestran bocetos del mismo. Por último, en la sección de referencias, se registran los autores en los que se basa el presente informe.

3 ANÁLISIS DE LOS DATOS

El conjunto de datos "Customer Shopping Preferences" de Kaggle ofrece informaciones interesantes sobre el comportamiento de consumidores y sus patrones de compra. Incluye en sus características atributos muy amplios como edad, género,

histórico de compra, métodos de pago, entre otros. El dataset tiene 3900 registros que se proponen a ayudar empresas a tomar decisiones centradas en los consumidores [2]. En la figura 1 se puede ver la estructura base del conjunto de datos y algunos de los registros presentes en ello.

Figura 1 - Estructura del conjunto de datos Customer Shopping Preferences

Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases
2896	56	Female	Hoodie	Clothing	86	Montana	L	Green	Summer	4.60	No	Standard	No	No	29	Bank Transfer	Monthly
2752	27	Female	Dress	Clothing	52	Minnesota	S	Indigo	Fall	3.10	No	Free Shipping	No	No	50	Venmo	Monthly
1224	69	Male	Pants	Clothing	24	Kansas	L	Red	Winter	3.90	No	Free Shipping	Yes	Yes	21	Bank Transfer	Weekly
2485	60	Male	Hoodie	Clothing	97	New Hampshire	M	Green	Summer	4.80	No	2-Day Shipping	No	No	50	Cash	Every 3 Months
3286	58	Female	Hat	Accessories	31	Hawaii	XL	Magenta	Fall	4.60	No	Free Shipping	No	No	11	Cash	Weekly

Fuente: [2]

Es importante mencionar como el presente conjunto de datos es compuesto de datos sintético, creado por ChatGPT para ser utilizado en estudios de análisis de datos y aprendizaje de máquina. Además, es relevante tener en cuenta como su estructura fue hecha simulando una experiencia real de compra, o sea, espejando el mundo real. Eso, por su vez, haz con que nuestra visualización, aunque ahora sea hecha sobre datos sintéticos, más tarde, si se desea, pueda ser aplicada sobre datos reales revelando informaciones concretas [2].

En la tabla 1 se describe cada uno de los atributos del conjunto de datos y también sus tipos.

Tabla 1 - Glosario del conjunto de datos Customer Shopping Preferences

Campo	Tipo	Descripción
Customer ID	Entero	Identificador único para cada cliente
Age	Entero	Edad del cliente
Gender	Categórico	Género del cliente (Hombre/Mujer)
Item Purchased	Categórico	El artículo comprado por el cliente
Category	Categórico	Categoría del artículo comprado
Purchase Amount (USD)	Decimal	La cantidad de la compra en USD
Location	Categórico	Lugar donde se realizó la compra
Size	Categórico	Tamaño del artículo comprado
Color	Categórico	Color del artículo comprado
Season	Categórico	Temporada durante la cual se realizó la compra
Review Rating	Decimal	Calificación otorgada por el cliente para el artículo comprado
Subscription Status	Categórico	Indica si el cliente tiene una suscripción (Sí/No)
Shipping Type	Categórico	Tipo de envío elegido por el cliente
Discount Applied	Categórico	Indica si se aplicó un descuento a la compra (Sí/No)
Promo Code Used	Categórico	Indica si se utilizó un código de promoción para la compra (Sí/No)
Previous Purchases	Entero	El recuento total de transacciones concluidas por el cliente en la tienda, excluyendo la transacción actual
Payment Method	Categórico	Método de pago preferido del cliente
Frequency of Purchases	Categórico	Frecuencia con la que el cliente realiza compras (por ejemplo, Semanal, Quincenal, Mensual)

Además, como los datos son todos sintéticos, no es necesario realizar ningún tratamiento previo a ellos. Los datos ya están todos correctamente rellenos y válidos de acuerdo con la descripción del dataset. Lo que sí será necesario hacer es transformar los datos para nuestras visualizaciones. De eso modo, realizaremos agregaciones con el valor de ventas (purchase amount) de dos formas, una por estado (location) y otra por estado (location), categoría (category) y estación del año (season).

4 PLANIFICACIÓN DE LA VISUALIZACIÓN

4.1 PÚBLICO OBJETIVO

Empresas que quieran entender sus patrones de venta online para mejorar sus estrategias de marketing o logística.

4.2 OBJETIVO

Mostrar los patrones de venta de una tienda online en los Estados Unidos.

4.3 FUNCIÓN

Exploratorio.

4.4 TONO

Pragmático y analítico.

4.5 EFECTO

No hay.

4.6 FACTORES

- Tiempo
 - Fecha límite para entrega es 17/12/2023.
- Costes
 - No se dispone de ningún recurso financiero para crear la visualización.
- Formato
 - La visualización debe ser desarrollada para web utilizando la API D3.js;
 - La visualización debe poseer interactividad;
 - La visualización debe aplicar los principios de usabilidad, accesibilidad e interactividad estudiados en la asignatura de Diseño y Evaluación de Sistemas Interactivos [3].

4.7 ENFOQUE

Se debe crear dos gráficos para la visualización. El primer debe comparar la cantidad de ventas entre los estados de Estados Unidos. En el segundo gráfico se podrá entender mejor como la venta de productos se distribuyó entre las categorías vendidas en la tienda online y por las estaciones del año.

La idea es que el segundo gráfico sirva de apoyo al primero. De eso modo, se añadirá interactividad sobre el gráfico 1 para que al hacer clic sobre un estado específico se muestre el gráfico 2. Además, otra interactividad esperada en el primer gráfico es que al pasar el cursor del ratón sobre un estado se muestre el nombre del estado y su cantidad de ventas.

5 DISEÑO

Los datos del dataset que serán utilizados son purchase amount, category, location y season. Las formas de representación a utilizar serán: para el gráfico 1,

cartograma isométrico, por su capacidad de representación concisa de los datos de venta de 50 estados; y para el gráfico 2, área apilada, por su factor estándar de mercado, al representar ventas por categoría en un determinado de tiempo.

En lo que se refiere a la organización de los gráficos en la visualización, se debe posicionar el gráfico 1 a la izquierda y el gráfico 2 a la derecha. Eso en razón de la tendencia cognitiva de lectura occidental, que sigue de izquierda a derecha. Esta elección busca mitigar posibles sesgos en la interpretación y asegurar que la información principal sea procesada primero [3].

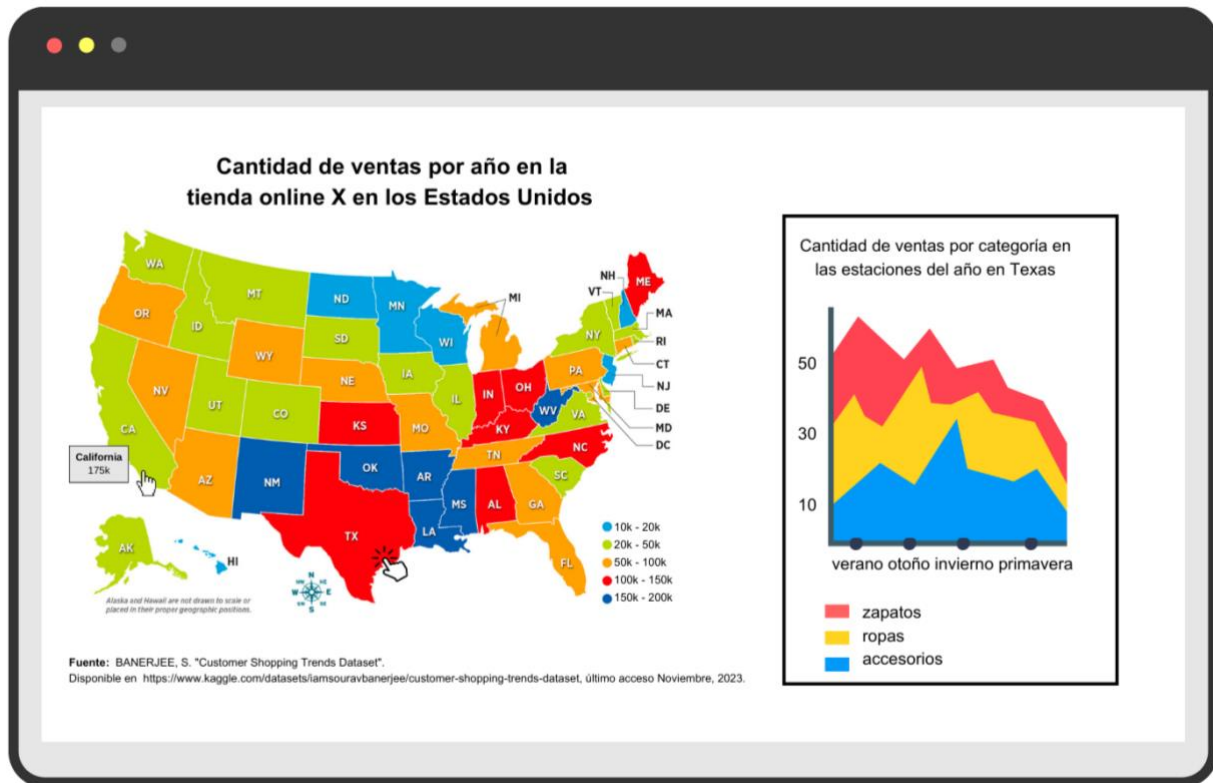
Gráfico 1

- Se mostrará la cantidad de ventas por los estados de Estados Unidos.
 - Se debe añadir el título “Cantidad de ventas por año en la tienda online X en los estados unidos”;
 - Se debe utilizar la cantidad de ventas en rangos;
 - Se debe utilizar una paleta de colores equidistante por el tono pragmático/analítico de la visualización y también para que se note bien la diferencia entre los valores;
 - Se debe añadir una leyenda indicando cuál rango de valores cada color demuestra;
 - Se debe añadir, en una nota abajo del gráfico, un enlace para la fuente de los datos.
 - Se debe mostrar la cantidad de ventas del estado al mover el ratón sobre respectivo estado;
 - Se debe abrir el Gráfico 2 al pinchar en un determinado estado.

Gráfico 2

- Se mostrará la cantidad de ventas por categoría en las diferentes estaciones del año en un determinado estado.
 - Se debe añadir el título “Cantidad de ventas por categoría en las estaciones del año en estado Y”;
 - Se debe añadir una leyenda indicando cuál categoría cada color representa;
 - Se debe utilizar una paleta de colores equidistante por el tono pragmático y analítico y también para que se note bien la diferencia entre los valores.

Figura 2 - Boceto de la visualización



6 REFERENCIAS

- [1] RICO, A. **"Patrones de compra y marketing generacional: qué estrategias emplear para vender según la edad"**. Disponible en https://www.elplural.com/el-telescopio/patrones-de-compra-y-marketing-generacional-estrategias-para-vender_207173102, último acceso Noviembre, 2023.
- [2] BANERJEE, S. **"Customer Shopping Trends Dataset"**. Disponible en <https://www.kaggle.com/datasets/iamsouravbanerjee/customer-shopping-trends-dataset>, último acceso Noviembre, 2023.
- [3] **Apuntes de la asignatura Diseño y Evaluación de Sistemas Interactivos** (2023). Universidad de Valladolid.