

Objetivo

La comprensión del negocio es esencial para que las empresas puedan conseguir sus objetivos, así como para el desarrollo de sus planes de futuro. La visualización de datos es imprescindible para entender el negocio (clusterización de clientes, paquetización de productos, identificación de zonas de expansión, segmentación comportamental en ventas...).

Te retamos a que crees la mejor aplicación y/o visualización sobre los datos anonimizados facilitados por Atida | Mifarma, con la información de negocio del periodo transcurrido entre el 01/01/2017 y el 31/12/2018. Puedes realizar un cuadro de mando, un ejercicio analítico exploratorio, una infografía, una web, un análisis gráfico avanzado... Sorpréndenos.

Los datasets

Ponemos a tu disposición 3 datasets con la información necesaria para poder realizar tu proyecto. Todos los ficheros están en formato CSV, separados por comas y usando el punto como indicador decimal en las variables numéricas.

Dataset 1

El primer dataset "items_ordered_2years.csv" incluye la información de las ventas realizadas a través de la web desde enero de 2017 a diciembre de 2018 en España. Se presentan los datos de los tickets de compra, mostrando las siguientes variables:

NOMBRE	DEFINICION	TIPO
<i>num_order</i>	Identificador de pedido	CATEGÓRICA
<i>item_id</i>	Identificador del objeto	CATEGÓRICA
<i>created_at</i>	Fecha de la compra	TEMPORAL
<i>product_id</i>	Identificador del producto	CATEGÓRICA
<i>qty_ordered</i>	Número de productos en el pedido	NUMÉRICA
<i>base_cost</i>	Precio base	NUMÉRICA
<i>price</i>	Precio de venta	NUMÉRICA
<i>discount_percent</i>	Descuento aplicado	NUMÉRICA
<i>customer_id</i>	Identificador del cliente	CATEGÓRICA
<i>city</i>	Municipio del cliente	CATEGÓRICA
<i>zipcode</i>	Código postal del cliente	CATEGÓRICA

Dataset 2

Por su parte, el segundo dataset “products.csv” presenta la información del catálogo de productos disponible en la web. Incluye las siguientes variables de cada producto:

NOMBRE	DEFINICION	TIPO
<i>product_id</i>	Identificador del producto	CATEGÓRICA
<i>sku</i>	Código del producto	CATEGÓRICA
<i>name</i>	Nombre del producto	CATEGÓRICA
<i>marca_value</i>	Marca del producto	CATEGÓRICA
<i>short_description</i>	Descripción del producto	CATEGÓRICA
<i>analytic_category</i>	Categoría del producto	CATEGÓRICA
<i>picture</i>	Url con la imagen del producto	CATEGÓRICA

Dataset 3

Por último, el tercer dataset “products_cat.csv”, contiene la clasificación de los productos a distintos niveles, presentando las siguientes variables:

NOMBRE	DEFINICION	TIPO
<i>sku</i>	Código del producto	CATEGÓRICA
<i>cat1</i>	Categorización primer nivel	CATEGÓRICA
<i>cat2</i>	Categorización segundo nivel	CATEGÓRICA
<i>cat3</i>	Categorización tercer nivel	CATEGÓRICA

Se valorará

- Análisis exploratorio de los datos (20 pts)
- Que el proyecto realice, permita y obtenga conclusiones de los datos facilitados.
- Enriquecimiento de datos (10 pts)
- Que el Proyecto, además de utilizar los datos procedentes de los Datasets, contenga datos de otras fuentes.
- Utilidad (20 pts)
- Que el Proyecto permita una mejor toma de decisiones.
- Grado de innovación (20 pts)
- Originalidad de las hipótesis planteadas, así como de las técnicas de visualización, modelización y comunicación utilizadas.
- Comunicación (30 pts)
- Que los resultados del Proyecto se transmitan de forma clara y concisa, la documentación interna aportada (códigos, comentarios) esté correctamente expresada y estructurada, y sea reproducible. Esta puntuación se dividirá entre:
 - Presentación resultados mediante app y/o web funcional (15 pts)
 - Valoración de la interactividad y la experiencia del usuario (15 pts)