

Prueba Data Science

Requerimientos

- Python
- Enviar la prueba en Google Colab u otro Notebook
- Herramientas open source

Caso Ventas de autos y motocicletas

Este conjunto de datos corresponde a las ventas de una tienda minorista. La información contiene pedidos, clientes, envíos, etc. Este conjunto hace parte originalmente de Pentaho DI Kettle.

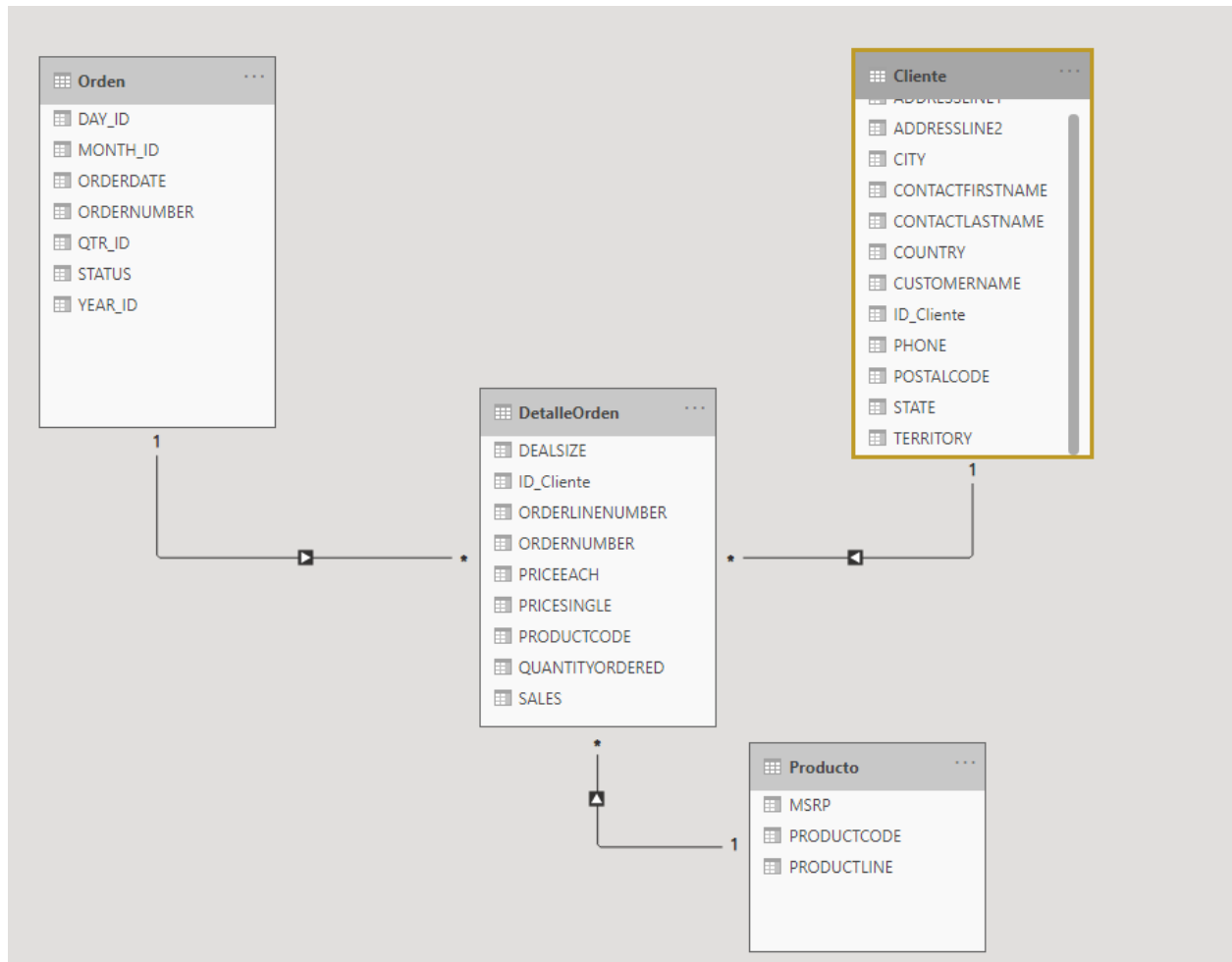
Modelado de datos

Dentro de las variables se identifica cuatro tablas que son:

- Cliente: ID Cliente, Nombre del cliente, ciudad, país, dirección, teléfono, territorio, etc.
- Producto: Código del producto, precio base del producto.
- Orden: Número de orden de la venta, ID Cliente, Fecha de la venta, Año, Trimestre, Mes y Estado de la venta.
- Detalle de la venta: ID Cliente, Número de Orden de la venta, Código del producto, Cantidad ordenada, precio unitario del producto, venta del producto.

En el Gráfico 1 se puede observar cómo quedaría modelada la información.

Gráfico 1 Modelado de datos para las ventas de productos



Fuente: Elaboración propia

1. Análisis descriptivo

Realizar un análisis donde incluyan los gráficos explicando el consumo de los clientes, los productos más vendidos, fecha, etc.

2. Segmentación de Clientes RFM

Para segmentar los clientes, ud podría utilizar la técnica RFM, la cual consiste en clasificar los clientes en función de su Reciente (Recency, Días transcurridos desde su última compra), Frecuencia (Frequency, Número de compras realizadas en un periodo determinado), Valor Monetario (Money, Valor de las compras totales realizadas en un periodo de tiempo) (IMF, 2017).

Esta técnica consiste en calcular para cada cliente las variables anteriores. Para este ejercicio, nuestros clientes son las empresas que compran y se sugiere tener en cuenta las ventas realizadas entre 6/01/2003 – 31/05/2005.

Posteriormente, construya las variables RFM según la siguiente Tabla 1.

Tabla 1 Clasificación de las empresas según el RFM

Recencia	Frecuencia	Valor Monetario
5, más reciente	5, Más frecuente	5, valor más alto
4	4	4
3	3	3
2	2	2
1 más antiguo	1, menos frecuente	1, valor más bajo

Fuente: Elaboración propia.

Teniendo en cuenta estos resultados, clasifican las empresas en 5 grupos (Tabla 5).

Tabla 5. Clasificación de los grupos según los resultados del RFM

Segmentos	R	F	M	Descripción
Potenciales	4,5	4,5	4,5	Son los clientes que menos días tardan en comprar, los que más a menudo lo realizan y son los que más gastan.
Perdidos	1,2, 3	1,2, 3	1,2, 3	Son clientes que compraron hace mucho tiempo, poca cantidad y poco gasto.
Derrochadores			5	Los que más gastan
Leales		4,5		Son los que más frecuente van a comprar
Nuevos	4,5			Clientes que han comprado hace poco pero no a menudo.

Fuente: Elaboración propia

3. Realice la segmentación por Clustering

Utilice un algoritmo de clustering automático como kmeans/dbscan/EM/Agglomerative y compare ambas técnicas.

4. Modelo de Sistema de Recomendación

En este ejercicio, se desea que proponga un modelo de sistema de recomendación. Tenga en cuenta que en cada orden se compraron varios productos y en diferentes cantidades.

Por ejemplo, la siguiente imagen se puede evidenciar que en la orden 10100 se compraron 4 productos en diferentes cantidades (30, 50, 22, 49) y en la orden 10102 se compraron 2 productos también en diferentes cantidades (39, 41).

Debe construir un modelo en el cual: si un cliente compra un producto, se le pueda ofrecer otro producto. Identifique el producto que más se vendió y proponga por lo menos 3 combos. Lo mismo para el producto que menos se vendió.

ORDERNUM	QUANTITY	PRICEEAC	ORDERLIN	SALES	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCT	MSRP	PRODUCT	CUSTOMER	PHONE	ADDRESS	ADDRESS	CITY	STATE	POSTALCC	COUNTRY	TERRITOR	CONTACT	CONTACT	DEALSIZE
10100	30	100	3	5151	1/6/2003 0:00	Shipped	1	1	2003	Vintage Cars	170 \$18_1749	Online Dieca	6035558647	2304 Long Airport Avenue Nashua	NH		62005 USA	NA	Young	Valarie	Medium			
10100	50	67.8	2	3390	1/6/2003 0:00	Shipped	1	1	2003	Vintage Cars	60 \$18_2248	Online Dieca	6035558647	2304 Long Airport Avenue Nashua	NH		62005 USA	NA	Young	Valarie	Medium			
10100	22	86.51	4	1903.22	1/6/2003 0:00	Shipped	1	1	2003	Vintage Cars	92 \$18_4409	Online Dieca	6035558647	2304 Long Airport Avenue Nashua	NH		62005 USA	NA	Young	Valarie	Small			
10100	49	34.47	1	1689.03	1/6/2003 0:00	Shipped	1	1	2003	Vintage Cars	41 \$24_3969	Online Dieca	6035558647	2304 Long Airport Avenue Nashua	NH		62005 USA	NA	Young	Valarie	Small			
10102	39	100	2	4808.31	1/10/2003 0:00	Shipped	1	1	2003	Vintage Cars	102 \$18_1342	Vitachrome I	2125551500	2678 Kingsto Suite 101	NYC	NY	10022 USA	NA	Frick	Michael	Medium			
10102	41	50.14	1	2055.74	1/10/2003 0:00	Shipped	1	1	2003	Vintage Cars	53 \$18_1367	Vitachrome I	2125551500	2678 Kingsto Suite 101	NYC	NY	10022 USA	NA	Frick	Michael	Small			

Archivo adjunto con los datos.

¿Qué se debe entregar?

Tiene 4 días para enviar los resultados.

1. Código que utilizó en cada uno de los puntos
2. **Documento con los análisis para cada uno de los puntos** (incluyendo gráficas, la interpretación de los procesos y resultados, y las conclusiones).
3. Video de no más de 10 minutos explicando los puntos principales en cuanto a:
 - a. Desarrollo de la prueba (dificultades, sugerencias)
 - b. Interpretación de los resultados encontrados (hallazgos)
4. Diligenciar encuesta:

https://docs.google.com/forms/d/e/1FAIpQLSfHvYbxCm6pXBrizVbtJu6rZWY9epXW1rr-A_hVqWMUbMhrzg/viewform