

Proyecto de curso Análisis de Datos con Python 445

Instrucciones:

- Por favor, complete todas las secciones.
- Utilice Python y sus librerías de análisis de datos para resolver los problemas planteados.
- Comente su código para explicar su razonamiento.

Este examen tiene como objetivo practicar las habilidades de Python aprendidas en un conjunto de datos que incluye cargar, inspeccionar, discutir, explorar y sacar conclusiones de los datos.

1. Obtenga los datos y realice un primer acercamiento a estos. Aplique todo lo aprendido, visualizar las primeras columnas, nombre de las columnas, tipo de datos, cantidad de filas y columnas, medidas estadísticas, etc.
2. Data Cleaning: evalúe datos nulos y duplicados. Si encuentra filas donde todos los valores se repiten, elimine dichas filas.

Con respecto a los nulos, en el caso de país rellene con el valor correspondiente.

En el caso de código postal:

Para el estado de California complete con : 90036

Para el estado de Ohio complete con: 44256

Para el estado de Pennsylvania complete con : 19140

Para el resto complete con 0 (cero)

¿La columna Sales tiene valores negativos, le parece esto correcto? En caso de que su respuesta sea negativa, realice la corrección correspondiente.

3. Renombre la columna 'Order ID' por 'Numero de Orden' y cambie el tipo de datos de la columna Postal code a int.
 4. Crear dos nuevas columnas que se llamen: Year_orderdate y month_orderdate, desglosando el año y mes de la columna Order Date.
 5. Realice el código necesario para responder a las siguientes preguntas
 6. Cual es el número de consumidores por segmento?
 7. Identifique el ID de consumidor, su nombre, segmento y total de ordenes solo para aquellos clientes que tienen mas de una orden.
 8. Identifique los ID de clientes de aquellos 10 que gastaron mas (columna sales)
 9. Cual es el número de consumidores por ciudad?
 10. Cual es el total vendido por cada año?
-
11. Visualización
 - a. Grafique las ventas por categoría en un grafico de torta, cada categoría deberá figurar con el porcentaje correspondiente que representa del total.
 - b. Grafique las ventas por región utilizando barras.
 - c. Grafique las ventas del año 2018 colocando la fecha en el eje x y las ventas en el eje y. Utilice un gráfico de líneas.
 12. Se quiere realizar un descuento a un número reducido de clientes.
 - a. Utilizando el dataset costumer_discount muestre el siguiente detalle: el ID del cliente, su nombre, categoría, el total de ventas, el descuento y las ventas con el descuento aplicado.