TARGET DATASET EXPLORATION: AN SQL AND PYTHON APPROACH

"NAVIGATING THE FUTURE OF ONLINE SHOPPING"

Jobaaj Learnings

DATASET OVERVIEW

Dataset Files:

1. CUSTOMERS.CSV : CUSTOMERS DEMOGRAPHIC DETAILS
2. SELLERS.CSV : SELLERS INFORMATION
3. ORDER_ITEMS : ORDER ITEMS DETAILS
4. GEOLOCATION.CSV : GEOLOCATION DETAILS
5. PAYMENTS.CSV : PAYMENTS DETAILS
6. ORDERS.CSV : ORDER ISTORY AND DETAILS
7. PRODUCTS.CSV : PRODUCTS DETAILS

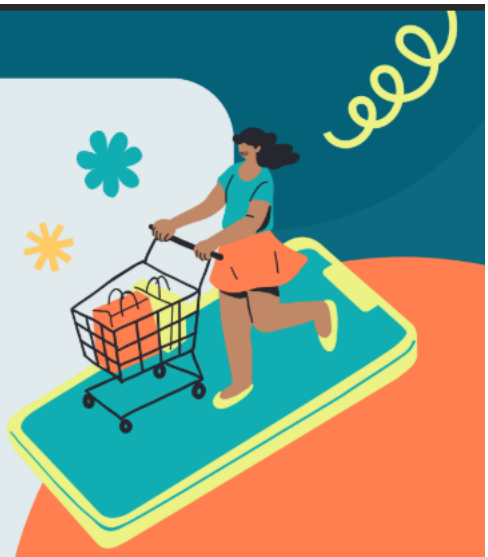
Goal: Analyze and visualize business metrics using SQL and Python.

Jobaaj Learnings

BASIC PROBLEMS:

Objective: Extract fundamental insights from the dataset.

1. List all unique cities where customers are located.
2. Count the number of orders placed in 2017.
3. Find the total sales per category.
4. Calculate the percentage of orders that were paid in installments.
5. Count the number of customers from each state.

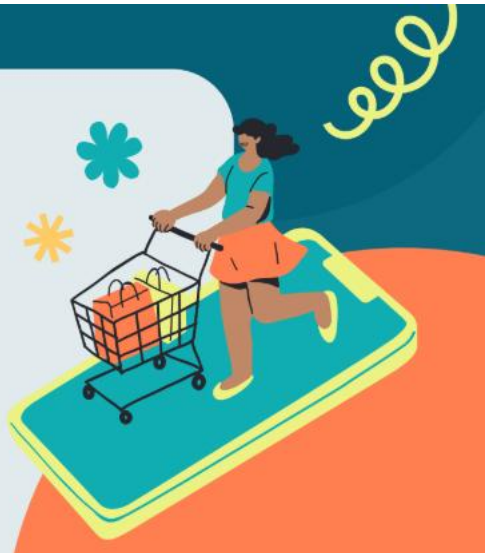


Jobaaj Learnings

INTERMEDIATE PROBLEMS:

Objective: Dive deeper into sales and order trends.

1. Calculate the number of orders per month in 2018.
2. Find the average number of products per order, grouped by customer city.
3. Calculate the percentage of total revenue contributed by each product category.
4. Identify the correlation between product price and the number of times a product has been purchased.
5. Calculate the total revenue generated by each seller, and rank them by revenue.

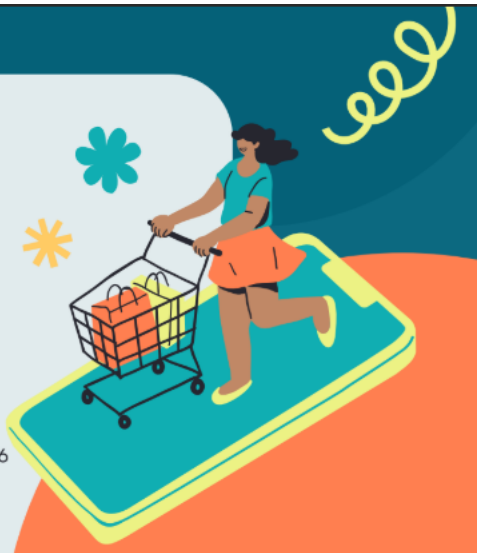


Jobaaj Learnings

ADVANCED PROBLEMS:

Objective: Generate strategic and customer-centric insights.

1. Calculate the moving average of order values for each customer over their order history.
2. Calculate the cumulative sales per month for each year.
3. Calculate the year-over-year growth rate of total sales.
4. Calculate the retention rate of customers, defined as the percentage of customers who make another purchase within 6 months of their first purchase.
5. Identify the top 3 customers who spent the most money in each year.



Jobaaj Learnings

WORKFLOW AND TOOLS

SQL: Import products.csv, orders.csv, customers.csv, sellers.csv.
Python: Use pandas to manipulate and visualize data.
Query Execution: Run SQL queries on MySQL Workbench.
Visualization: Create insightful graphs using Google Colab.

MySQL Workbench
Python Libraries: Pandas, Matplotlib

PYTHON **SQL**

Jobaaj Learnings


DELIVERABLES
SQL File: Queries for all three query levels.
Colab File: Includes Python visualizations.
Presentation: Summarizes project insights.
Feel free to reach out to your mentor for any queries via LinkedIn : Analyst-kashishagrawal/
Jobaaj Learnings