

Sales Performance Analysis of Amazon

Using Advanced SQL Techniques

Introduction

Amazon is one of the world's largest e-commerce platforms, operating across multiple cities and states while offering a wide range of product categories. The dataset provided contains transactional sales data including order details, fulfilment methods, shipment locations, order status, customer type (B2B / Non-B2B), and revenue information.

This project uses **advanced SQL techniques** to analyze Amazon's sales data and extract actionable insights related to sales performance, customer behavior, fulfilment efficiency, and revenue trends.

Business Problem

Amazon wants to optimize its sales and operational strategies by analyzing historical order data across different cities, states, fulfilment methods, and product categories. By leveraging SQL-based analysis, the company aims to understand revenue drivers, customer segments, order patterns, and performance trends.

Advanced SQL queries are used to answer business-critical questions related to order distribution, revenue contribution, customer segmentation, and time-based sales trends.

Dataset

- **Dataset Name:** Amazon Sales Dataset
- **Dataset Description:**

The dataset contains order-level transactional data with information such as order status, fulfilment type, shipment city and state, product category, order amount, date of purchase, and customer type (B2B or Non-B2B).

Tasks

Task 1: Finding the Total Number of Orders

Amazon wants to determine the total number of orders placed in the dataset to understand overall order volume.

Task 2: Identifying Distinct Order Statuses

Analyze the dataset to list all distinct order statuses available, helping Amazon understand the order lifecycle and cancellation patterns.

Task 3: Orders by Fulfilment Type

Amazon wants to analyze how many orders are fulfilled by:

- Amazon
- Merchant

This helps evaluate operational dependency on fulfilment methods.

Task 4: Cancelled Orders Analysis

Identify the total number of cancelled orders to assess cancellation trends and potential operational issues.

Task 5: Orders Distribution by Ship City

Analyze the total number of orders placed per ship city and rank them based on order volume to identify high-demand regions.

Task 6: Revenue Analysis by Ship State

Calculate the total revenue generated by each ship state, considering only orders with **Shipped** status to ensure accurate revenue measurement.

Task 7: Sales Performance by Product Category

Determine the total sales amount generated by each product category to understand category-level performance.

Task 8: Top 5 Revenue-Generating Cities

Identify the top 5 ship cities that contribute the highest total revenue to Amazon.

Task 9: Average Order Value by Category

Calculate the average order amount for each product category to analyze customer spending behavior across categories.

Task 10: B2B vs Non-B2B Orders Analysis

Compare the total number of orders placed by:

- B2B customers
- Non-B2B customers

This helps understand the contribution of business customers.

Task 11: High-Value Orders Detection

Identify orders with values higher than the overall average order amount using a subquery to detect high-value transactions.

Task 12: Identifying Repeat Orders by Ship City

Using a **Common Table Expression (CTE)**, identify ship cities where more than one order was placed.

(Assumption: Repeat behavior is inferred based on ship city due to the absence of a unique customer identifier.)

Task 13: Fulfilment-Wise Revenue Contribution

Calculate total revenue generated by each fulfilment type and determine the percentage contribution of each fulfilment method to overall revenue using **CTEs**.

Task 14: Ranking Orders by Order Amount

Use a **window function** to rank orders based on their order amount from highest to lowest.

Task 15: Running Total of Revenue Over Time

Calculate the cumulative revenue over time by ordering transactions by date using a window function.

Evaluation Criteria

The project is evaluated based on:

- Correctness of SQL queries
- Use of advanced SQL concepts
- Clarity of problem understanding
- Business relevance of insights
- Proper documentation and assumptions