E-Commerce Sales & Customer Behavior Analysis Dashboard using SQL & Python

Project Summary & Report

Objective

The project aims to extract, analyze, and visualize insights from a real-world e-commerce dataset using SQL and Python. The goal is to understand sales trends, customer behavior, and business performance using analytical techniques and build a mini business intelligence solution.

Dataset Overview

The dataset simulates a Brazilian e-commerce platform and includes the following CSV files:

File Name	Description
`orders.csv`	Order metadata including timestamps and statuses
`order_items.csv`	Line items within each order, including product and price
`payments.csv`	Payment methods and payment values
`customers.csv`	Customer details including city and state
`products.csv`	Product categories
`sellers.csv`	Seller information
`geolocation.csv`	Geographic coordinates of cities

Tech Stack

- **Language**: Python
- **Libraries**: Pandas, Matplotlib, Seaborn, Plotly, ipywidgets
- **SQL Engine**: SQLite via sqlite3
- **Visualization**: Static and interactive plots using Seaborn, Plotly

Key Analyses Performed

1. Customer Distribution by State

- Aggregated customer count per state using SQL.
- Visualized with bar plots (interactive with ipywidgets).

2. Monthly Sales Trend (2018)

- Extracted orders from 2018 and grouped by month.
- Visualized order volume trends with bar plots.

3. Product Category Contribution

- Calculated percentage of sales by product category.
- Used SQL window functions and bar charts for ranking.

4. Cumulative Sales Over Time

- Implemented SUM(...) OVER() to track monthly cumulative sales.

5. Year-over-Year Growth

- Used SQL window functions to compare annual revenue and calculate % growth.

6. Repeat Purchase Analysis

- Calculated how many customers made repeat purchases within 6 months.
- Provided insights on retention (0% repeat rate observed).

7. Top Customers Ranking

- Identified top 3 customers by payment per year using DENSE_RANK().
- Compared customer contributions across years visually.

Key Insights

- **Most Orders**: Concentrated in a few months (seasonal spike observed).
- **High Performing Categories**: Few product categories dominate revenue.
- **No Short-Term Repeat Purchases**: May indicate need for retention strategy.
- **Top Customers**: A small group of customers contribute significantly each year.
- **Sales Growth**: Visualized YoY sales trends and growth rates.

Features Implemented

- SQL queries using window functions (RANK, DENSE_RANK, LAG, OVER)
- Modularized Jupyter Notebook for readable step-by-step analysis
- Interactive sliders for visual control using ipywidgets
- Plotly-based interactive charts and Seaborn static charts

~ AKASH KUMAR RAJAK
~ Email Id : akashkumarrajak200@gmail.com
~ **Github** : Link