E-Commerce Data Analysis Project

Overview

This project involves **Exploratory Data Analysis (EDA)** on an E-Commerce dataset to extract insights and trends. Using Python and SQL, various aspects of the data were explored, such as customer distribution, sales trends, product performance, and payment patterns. The goal was to derive meaningful insights that could help businesses understand their performance and improve decision-making.

Objectives

The project aimed to answer key business questions, such as:

1. Which cities have the most customers?
2. What are the sales trends across different product categories?
3. How do order and payment behaviors vary over time?
4. What is the revenue distribution across sellers and states?

Tools and Technologies

- **Python**: Pandas, Matplotlib, Seaborn, NumPy
- **SQL**: For complex data queries and transformations
- **MySQL Connector**: For connecting to the database
- **Visualization**: Used to generate meaningful charts to highlight trends and patterns

Key Features

1. **Unique Cities with Customers**
   Queried the dataset to identify all unique cities where customers are located.
2. **Order Analysis**
   - Counted the total orders placed in specific years (e.g., 2018).
   - Monthly analysis of orders placed in 2018.
3. **Sales Distribution**
   - Calculated total sales per product category.
   - Determined the percentage of revenue contributed by each category.
4. **Customer Distribution**
   Visualized customer counts by state using bar charts.
5. **Payment and Installment Trends**

    - Analyzed the percentage of orders paid in installments.
    - Examined revenue trends based on payment methods.
6. **Advanced Queries**
    - Revenue growth analysis (Year-over-Year).
    - Identified top customers and sellers by revenue.
    - Analyzed the correlation between product prices and purchase frequency.

Visualizations

- Bar plots showcasing customer distributions by state and order counts by month
- Revenue trends across sellers and product categories
- Percentage revenue contributions by product categories

Learning Outcomes

- **SQL Proficiency**: Writing complex queries to extract meaningful insights.
- **Python Skills**: Using Pandas for data manipulation and Seaborn/Matplotlib for visualization.
- **Business Insights**: Understanding e-commerce trends and customer behaviors.

Repository Contents

- `data/`: Raw and processed datasets used for analysis.
- `notebooks/`: Jupyter Notebooks containing Python code and queries.
- `visualizations/`: Images and charts generated during the analysis.
- `README.md`: This file, summarizing the project details.

Conclusion

This project was an exciting journey into data analysis, providing hands-on experience with real-world data. The findings and visualizations could serve as valuable insights for stakeholders in the e-commerce domain.

Feedback and contributions are welcome! Feel free to check out the project and share your thoughts.

***