Retail Order Data Analysis

Project Overview

The Retail Order Data Analysis project focuses on Data Analytics, leveraging various technologies such as Python, SQL, Streamlit, and Kaggle API. The goal is to analyze and optimize sales performance by identifying key trends, top-performing products, and growth opportunities from a retail sales dataset. The project uses the Kaggle API to retrieve data, Python and Pandas for cleaning and data manipulation, SQL for analysis, and Streamlit for data visualization.

Key Technologies Used:

Kaggle API: For retrieving the sales dataset.

Python: For data cleaning and manipulation using Pandas.

SQL: For querying and analyzing data in a relational database.

Streamlit: For creating interactive data visualizations and dashboards.

Project Breakdown:

PART 1: Data Retrieval and Data Preprocessing (Jupyter Notebook)

Description: In Part 1, we begin by retrieving the dataset using the Kaggle API. The dataset consists of retail sales data, and it is then cleaned and preprocessed using Python (Pandas). This includes handling missing values, transforming columns, and performing exploratory data analysis to prepare for further analysis.

Key Tasks in Part 1:

Loading the dataset using the Kaggle API: Retrieving the dataset in CSV format.

Data Cleaning: Removing or imputing missing values, handling duplicates, and normalizing data types.

Data Exploration: Performing exploratory analysis to understand the dataset’s structure and content.

Outcome: Clean and structured data ready for analysis in Part 2.

PART 2: Data Analysis using SQL and Visualization (Python Script)

Description: In Part 2, the data is imported into a relational database (SQL), where SQL queries are written to extract insights. These queries include the use of primary and foreign keys, as well as different types of joins. Additionally, Streamlit is used to create an interactive dashboard to visualize the data insights, including trends in sales, top-performing products, and potential growth opportunities.

Key Tasks in Part 2:

SQL Analysis:

Writing SQL queries to extract and analyze key metrics.

Using join operations, aggregations, and key constraints to derive insights from the data.

Data Visualization: Using Streamlit to create interactive graphs and charts to display sales trends and key product insights.

Outcome: A well-documented and interactive analysis that provides actionable insights for sales optimization.

Files in this Repository:

PART 1: retailorderdata.ipynb - The Jupyter notebook containing the code for data retrieval, cleaning, and preprocessing.

PART 2: RetailOrderStreamlit.py - The Python script that contains SQL queries and visualizations to analyze the cleaned data.

Overview of the webpage











