**EDA Report on Retail Orders**

## A Project Report

*Submitted by*

## Siddheshwar patil 712452058

**Saurabh Gaikwad 712452054**

**M.Tech (Data Science)**

Under the guidance of

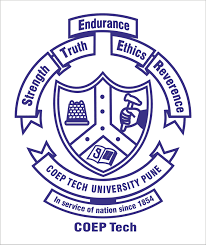
## Dr. Yashodhara Haribhakta

COEP Technological University, Pune

And

## Mr.Abhiraj Ubale

COEP Technological University, Pune



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,**

**COEP TECHNOLOGICAL UNIVERSITY, PUNE-5**

January 2025

1. **Problem statement:-**

An Exploratory Data Analysis (EDA) on a retail order dataset in CSV format, containing details of retail transactions. The goal is to analyze, clean, and visualize the data to uncover insights that can help in understanding customer behavior, sales trends, and product performance. You will leverage Python and SQL to extract, process, and analyze the data, providing meaningful insights that can support decision-making.

1. **Objectives:-**

**Data Cleaning:**

Identify and handle missing or erroneous data.

Correct inconsistent formatting and outliers.

**Data Exploration:**

Calculate summary statistics for key numerical columns (e.g., quantity, price, etc.).

Explore categorical variables like payment\_method and order\_status.

Investigate the distribution of orders over time (order\_date).

**SQL-based Analysis:**

Write SQL queries to extract key metrics such as total sales per customer, sales by product, and monthly order volume.

Perform aggregation and grouping in SQL to find trends in customer purchases and product demand.

**Insights:**

Provide insights into customer behavior, sales performance, and the overall retail operation.

Suggest potential actions or improvements based on the findings.

1. **Functional requirements:-**

**Data Ingestion, Cleaning, and Transformation:**

The system should support loading and cleaning retail order data from CSV files, handle missing or erroneous data, standardize formats, and convert columns to appropriate data types.

**Exploratory Data Analysis (EDA):**

The system should provide functions for generating summary statistics, visualizing trends and distributions, performing correlation analysis, and aggregating data via SQL queries to uncover insights on customer behavior, sales trends, and product performance.

**Data Visualization and Reporting:**

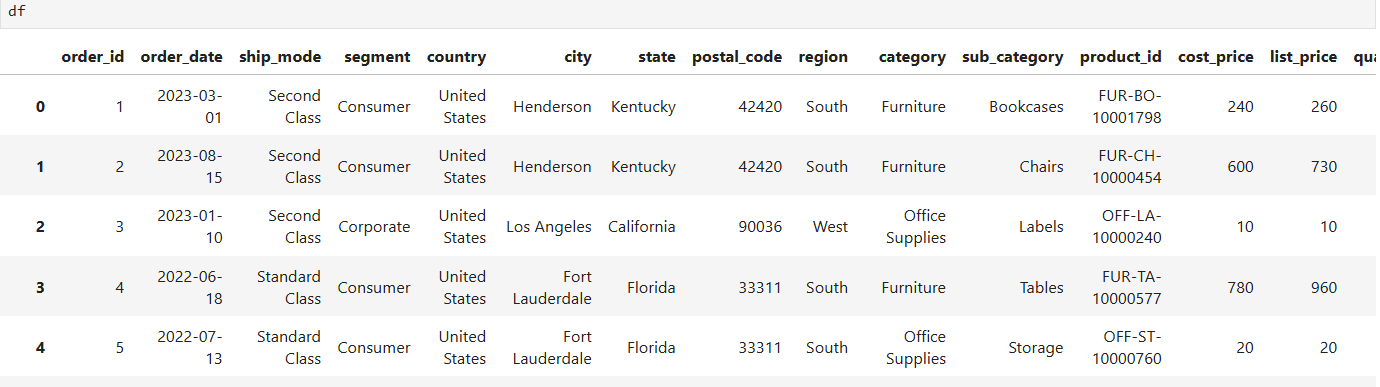
The system should support creating various visualizations (e.g., histograms, time series, scatter plots) and allow users to generate and export comprehensive reports with key findings and insights in formats like PDF, Excel, or CSV.

**User Interface and Interaction:**

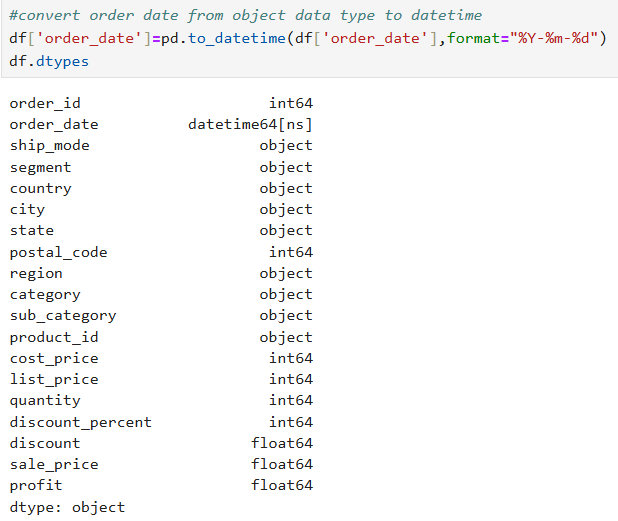
The system should feature an intuitive interface for uploading datasets, applying filters, running analyses, interacting with visualizations, and generating reports, with options for user roles and data security.

1. **Python:-**

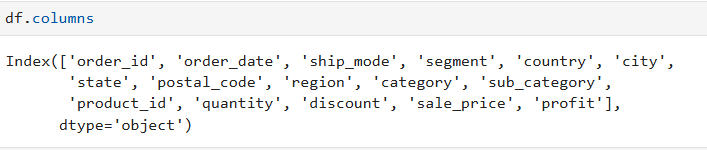
representing a DataFrame in the pandas library:

****

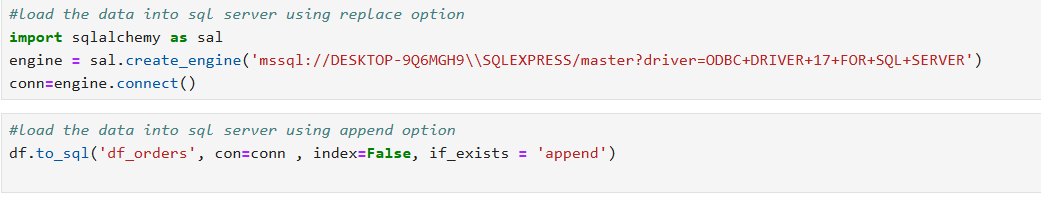
Converting order data from data type to datatime and representing the Data Types of each columns in the dataset:-

****

Representing the Columns in the Dataset:-

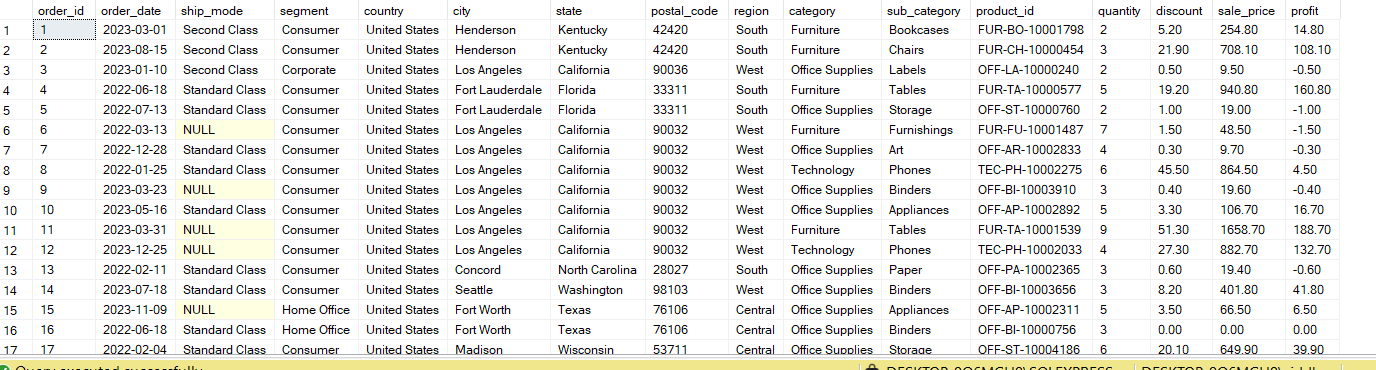
****

Loading the Data into the SQL Server:-

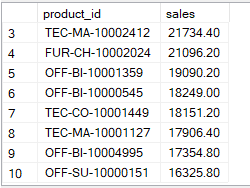
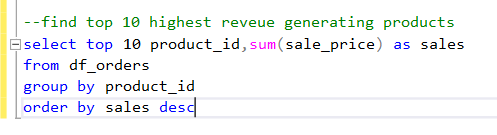
****

1. **SQL Queries:-**

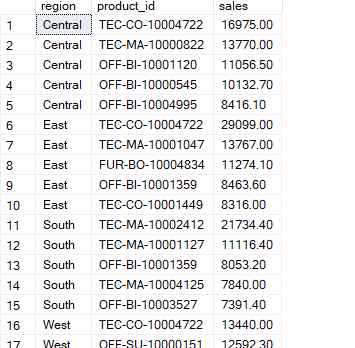
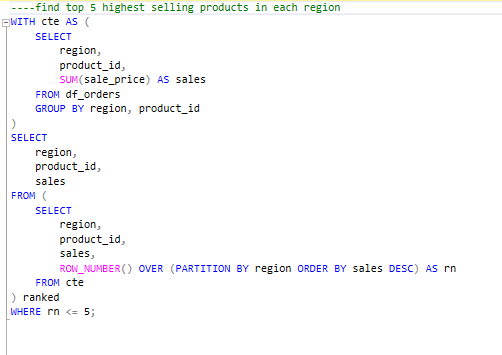
Table named Order

****

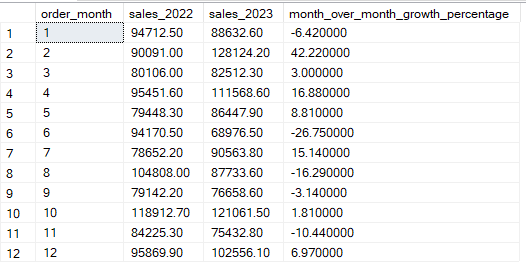
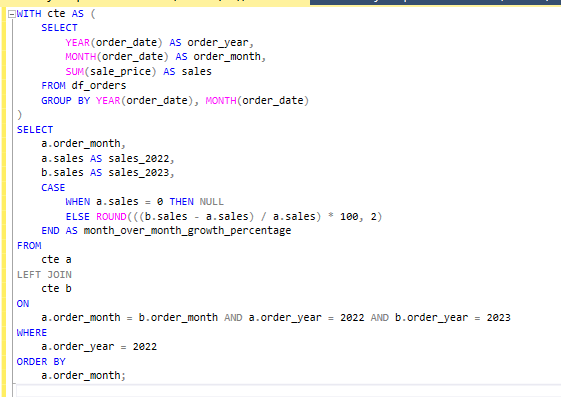
Finding the TOP 10 highest revenue generating products:-

****

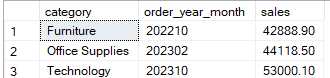
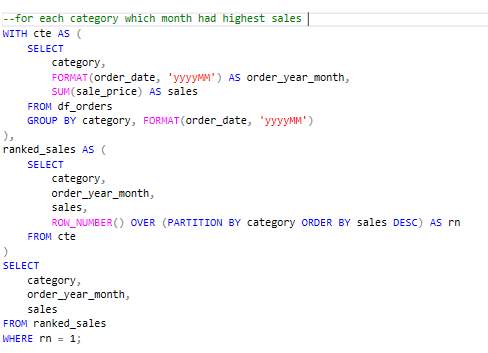
Finding the top5 highest selling products in each region:-

****

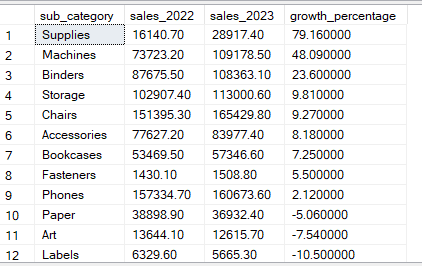
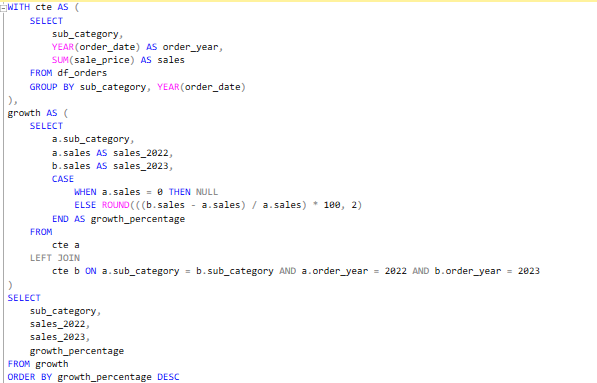
find month over month growth comparison for 2022 and 2023 sales eg : jan 2022 vs jan 2023

****

For each category which month had highest sales:-

****

which sub category had highest growth by profit in 2023 compare to 2022

****

1. **Conclusion:-**

The Exploratory Data Analysis (EDA) on the retail order dataset provides a comprehensive framework for understanding key aspects of retail transactions, such as customer behavior, product performance, and sales trends. By effectively cleaning, transforming, and visualizing the data, businesses can gain valuable insights that inform decision-making and drive strategic actions. The use of Python and SQL for data processing and analysis enhances the flexibility and scalability of the system, enabling both detailed data exploration and high-level aggregations. With interactive visualizations and customizable reports, stakeholders can easily interpret the findings and make data-driven decisions that improve operational efficiency, customer engagement, and product performance. Ultimately, this process empowers businesses to optimize their retail operations and better meet customer needs, leading to enhanced growth and profitability.

1. **Future Scope:-**

The future scope of the EDA system includes several exciting advancements. Real-time data integration could allow businesses to monitor and analyze transactions as they occur, providing immediate insights for faster decision-making. Incorporating predictive analytics would enable forecasting of sales, customer churn, and product demand, empowering businesses to proactively manage inventory and marketing strategies. The system could also evolve to offer personalized customer insights through advanced segmentation and recommendation systems. Additionally, integrating data from multiple sales channels (online, in-store) would provide a comprehensive view of customer behavior. Automation of insights, reporting, and alerts would reduce manual intervention and enable businesses to respond quickly to significant changes. Future versions could feature more interactive dashboards with drill-down capabilities, as well as integrate external data sources like social media and market trends to enhance decision-making. Finally, deploying the system on cloud platforms would provide scalability, security, and facilitate collaboration across teams, ensuring better performance and ease of access as the dataset grows.