Project Title	**Retail Order Data Analysis**
Skills take away From This Project	**Kaggle Api,Python, SQL, Streamlit**
Domain	**Data Analytics**

Problem Statement:

1. Problem Statement

- **Objective**: To analyze and optimize sales performance by identifying key trends, top-performing products, and growth opportunities using a dataset of sales transactions.
- **Goals**:
    1. Identify products and categories contributing the most to revenue and profit.
    2. Analyze year-over-year (YoY) and month-over-month (MoM) sales trends.
    3. Highlight subcategories with the highest profit margins to guide decision-making.

🌐 API Integration

—-> Data Extraction

- Source: The data is extracted from Kaggle using the Kaggle API. This step involves:
    - Creating a Kaggle profile and generating an API token to access the dataset.
    - Fetching the dataset via Python using the Kaggle API to download raw files, which are usually in CSV format.
    - Handling pagination if the dataset is large (optional, depending on API limitations).

Tools/Technologies:

- **Python: `requests`, `kaggle` library**
- **Kaggle API: For retrieving the dataset**

This the link to download the dataset directly from kaggle :

!kaggle datasets download ankitbansal06/retail-orders -f orders.csv

🧹 Data Cleaning

Raw datasets often contain inconsistencies that hinder analysis. The project resolves these using Pandas:

- **Handling Missing Values**: Replace missing numerical values with defaults like `0` or drop rows with critical missing fields.
- **Renaming Columns**: Standardize column names for clarity and compatibility with SQL databases (e.g., converting `Order ID` to `order_id`).
- **Trimming Spaces**: Remove trailing spaces from text fields.
- **#CALCULATING DISCOUNTS---derive new columns discount , sale price and profit**

🗄️ SQL Server Integration

Once cleaned, the dataset is transferred to SQL Server for efficient querying:

Data Loading into SQL Database

Create Database and move your dataframe into SQL

Tools/Technologies:

- **SQL Server**: To store and query the data

📊 Data Analysis with SQL

Business Insights through SQL Queries:

- **Top-Selling Products**: Identify the products that generate the highest revenue based on sale prices.
- **Monthly Sales Analysis**: Compare year-over-year sales to identify growth or decline in certain months.

- **Product Performance**: Use functions like `GROUP BY`, `HAVING`, `ROW_NUMBER()`, and `CASE WHEN` to categorize and rank products by their revenue, profit margin, etc.
- **Regional Sales Analysis**: Query sales data by region to identify which areas are performing best.
- **Discount Analysis**: Identify products with discounts greater than 20% and calculate the impact of discounts on sales.

NOTE: I have given you 10 queries and You need to do insights 10 queries like above mentioned. Totally you need to do data insights for 20 sql queries.

(You need to use Primary key Foreign key and join query operations)

1.# Find top 10 highest revenue generating products

2.# Find the top 5 cities with the highest profit margins

3.# Calculate the total discount given for each category

4.# Find the average sale price per product category

5.# Find the region with the highest average sale price

6.# Find the total profit per category

7.# Identify the top 3 segments with the highest quantity of orders

8.# Determine the average discount percentage given per region

9.# Find the product category with the highest total profit

10.# Calculate the total revenue generated per year

Streamlit 👍

Show the Datas in Streamlit

1. Set Up Your Environment

>>>>>>>Install Streamlit

2. Connect to the SQL Database

3.Query Data from the Database

Streamlit allows you to dynamically query data from your database and display it in real-time

4. Display Data Using Streamlit

Once you have queried the data, you can display it in the Streamlit app using different methods.

Display Data as Table or charts

5.Displaying Insights

Timeline:

1. The project must be completed and submitted within **7 days from the assigned date.**
2. Last 2 days of data will be validated.