10Alytics

CASE STUDY

10Alytics

EXECUTIVE SUMMARY

- Haven Retail is a prominent retail player known for its diverse product range and exceptional customer service. Operating across regions like the United States, it caters to a wide customer base, from individuals to corporate clients. Despite its success, Haven Retail faces challenges in utilizing its sales data effectively for strategic decision-making and operational improvements. The role of a data engineer is crucial in designing data pipelines, optimizing databases, and developing scalable solutions for data processing. By harnessing data engineering techniques, Haven Retail aims to transform raw data into valuable insights to drive business growth.

10Alytics

BUSINESS PROBLEM STATEMENT

- Haven Retail faces challenges with its sales data utilization, including data silos, lack of real-time insights, and limited scalability. Data fragmentation impedes comprehensive analysis, while the absence of real-time processing delays decision-making. Additionally, the current infrastructure struggles to scale efficiently with growing data volumes, hindering complex analytics and future expansion.

The need for a data engineering solution that addresses these challenges head-on is clear. By leveraging advanced technologies and methodologies, the goal is to build a system that not only resolves the current issues but also sets a foundation for future innovation and growth within Haven Retail

10Alytics

YOUR TASK

As the new data engineer at Haven Retail, your core responsibility revolves around optimizing data management practices to bolster informed decision-making throughout the organization. Your primary objective is to implement a robust database system that can efficiently store, organize, and retrieve a diverse array of data, including sales transactions, customer information, product details, inventory levels, and other critical aspects of business operations. This entails designing and structuring the database schema to ensure seamless data storage and retrieval, developing mechanisms for data validation and quality assurance, and collaborating with stakeholders to understand their data needs and deliver actionable insights. Your role also encompasses implementing security measures to safeguard sensitive information and comply with data privacy regulations. Ultimately, your efforts as a data engineer are instrumental in laying the groundwork for effective data utilization and driving business growth at Haven Retail.

10Alytics

OBJECTIVES

1. Database Setup: Establish a PostgreSQL database named `retail_orders` to efficiently manage data related to sales transactions, customers, products, inventory, and business operations.

2. Data Loading: Develop Python scripts using psycopg2 to accurately load data from a CSV file into the `retail_orders` table, ensuring data integrity and completeness.

3. Data Retrieval and Analysis: Create Python functions to retrieve and display the first 5 rows of data from the `retail_orders` table, enabling initial data verification and analysis.

4. Metric Calculation: Formulate SQL queries to calculate essential metrics such as the total number of orders placed and the total sales amount for all orders, providing crucial business insights.

5. Data Modification and Maintenance: Implement Python functions for data manipulation tasks, including updating the ship mode for specific orders and deleting orders based on criteria, ensuring data accuracy and system maintenance.

10Alytics

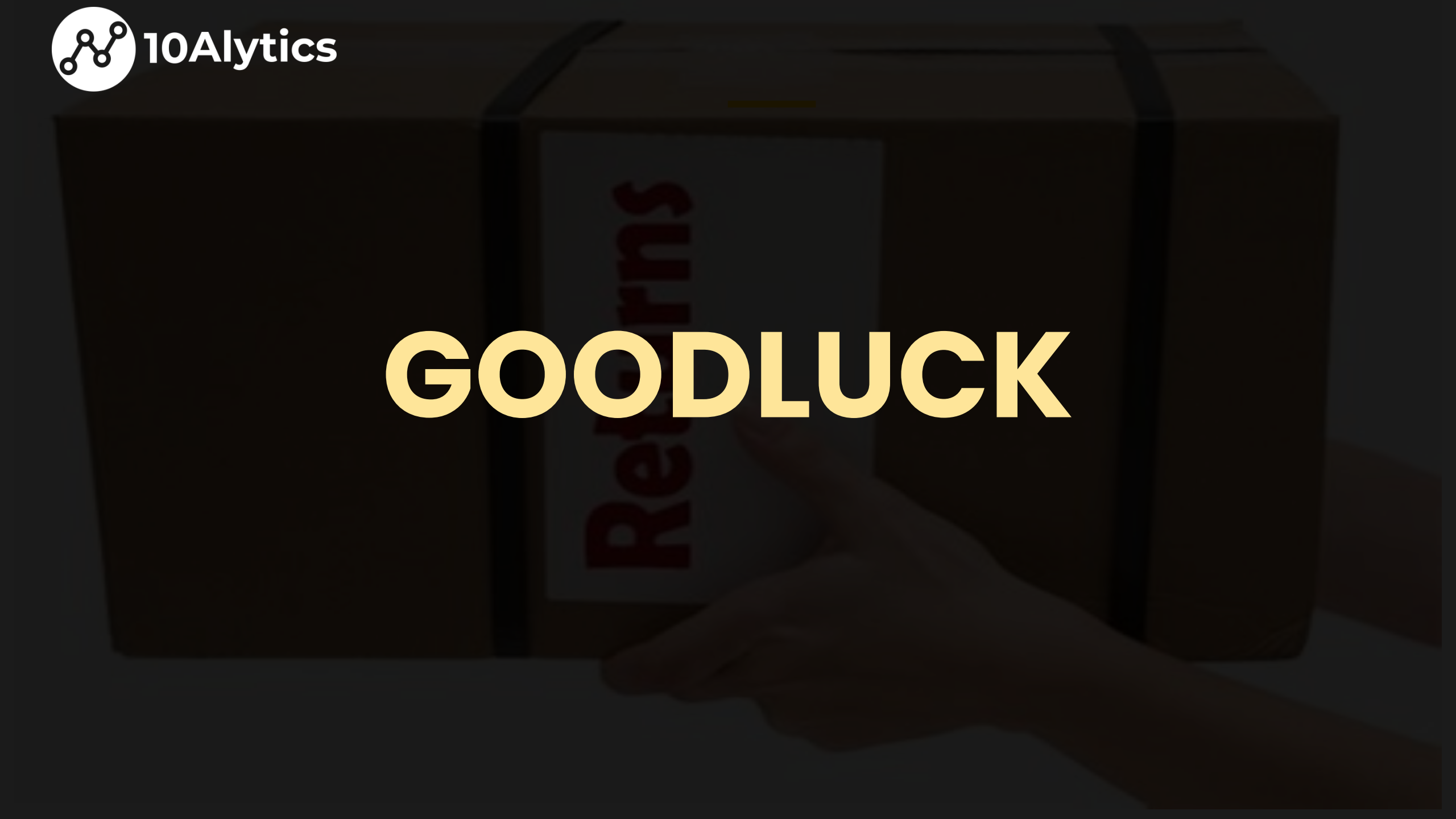
DATASET

- Here is a url to the dataset ⇒ LINK
- Here is a url to the data dictionary ⇒ LINK

10Alytics

TASKS

1. Write Python code to connect to a PostgreSQL database and create the retail_orders table.
2. Write Python code to load data from a CSV file into the retail_orders table in PostgreSQL.
3. Write Python code to check if the retail_orders table exists in the PostgreSQL database after table creation.
4. Write Python code to fetch and print the first 5 rows of data from the retail_orders table.
5. Write a SQL query to retrieve the total number of orders placed.
6. Write a SQL query to calculate the total sales amount for all orders.
7. Write a Python function to fetch all orders placed in Florida.
8. Write a SQL query to find the top 5 customers based on their total sales amount.
9. Write a Python script to insert a new order into the database.
“9800, CA-2016-128608, 12-01-16, 17-01-16, Standard Class, CS-12490, Cindy Schnelling, Corporate, United States, Toledo, Ohio, 43615, East, TEC-AC-10000487, Technology, Accessories, SanDisk Cruzer 4 GB USB Flash Drive, 10.384”
10. Write a SQL query to count the number of orders shipped using Standard Class.
11. Write a Python function to update the ship mode to “Standard Class ” for the order ID = “CA-2015-167164” .
13. Write a Python script to delete orders with the order ID = “US-2016-108966 ”
14. Write a Python function to calculate the total sales amount for the technology category.


10Alytics
GOODLUCK