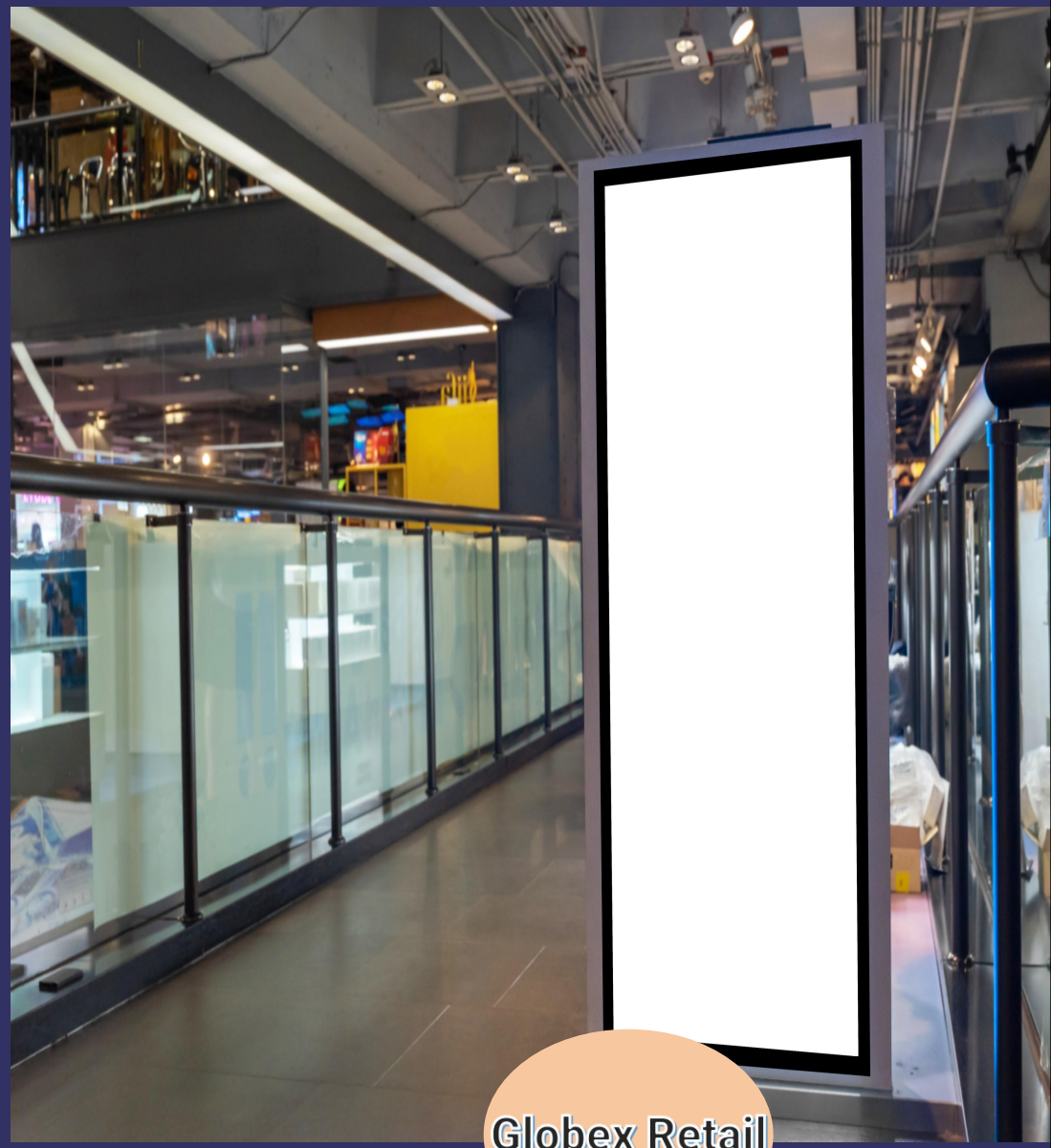
CAPSTONE PROJECT ON PANDAS

Globex Retail

Exploring how Globex Retail harnessed data engineering to transform customer data into actionable insights, driving business growth and innovation



By Chinwendu Okafor

Contents

Business Introduction

Globex Retail, a dynamic leader in the retail industry, seamlessly integrates the convenience of e-commerce with the traditional appeal of brick-and-mortar stores. As the company continues to expand its reach, it recognizes the critical role that data engineering plays in driving success.

By harnessing the power of advanced data engineering techniques, Globex Retail is committed to transforming raw data into actionable insights, optimising inventory management, and ultimately boosting revenue.

This data-driven approach positions Globex Retail at the forefront of innovation, ensuring that they not only meet but exceed the evolving demands of their diverse customer base.

Problem Statement

As a new Data Engineer at Globex Retail, a company specialising in e-commerce and brick-and-mortar sales, your first task involves analysing a dataset to understand customer purchase behavior. The goal is to identify patterns in customer purchases that could lead to targeted marketing strategies and improved inventory management.

Rationale for the Project

Globex Retail seeks to:

01

Enhance customer retention.

02

Optimise inventory levels to meet customer demand.

03

Increase revenue through targeted marketing.

04

Make data-driven decisions that improve the overall business performance.

Data Description

The dataset includes information on customer purchases, such as:

1. Customer ID: Unique identifier for customers.
2. Order ID: Unique identifier for each order.
3. Order Date: Date the order was placed.
4. Product Category: Main category of the product.
5. Product Sub-Category: More specific category of the product.
6. Quantity: Number of units ordered.
7. Price: Price per unit of the product.
8. Discount: Discount applied to the product.
9. Customer Location: Location (state code) of the customer.

Tech Stack

The Analytical Powerhouse

Python

Pandas

Jupyter

The Core Engine

The Interactive Environment

Project Workflow

The project followed a structured data engineering approach:

1. Familiarisation & Requirement Analysis:

- Understanding the dataset variables, including Order IDs, Product Categories, and Customer Locations.
- Understanding the structure and schema of the retail records.

2. Data Cleaning:

- Handling missing values and ensuring data integrity for accurate analysis.

3. Feature Engineering:

- Creating new metrics to drive deeper analysis and better understand value and behaviour.

4. Exploratory Data Analysis (EDA):

- Performing deep dives into revenue drivers and customer segments.

5. Insight Generation:

- Using Pandas to extract specific answers for stakeholderssuch as to answer critical business questions regarding geographic concentration and discount impacts.

Insights

Generating Insights with Pandas:

- Which product category has generated the most revenue?
- Which product subcategories generate the most revenue?
- Which customer segment has the highest average order value?
- What is the average discount applied to high-value customers?
- Which geographic locations have the highest concentration of high-value customers?
- How does the discount rate impact the total sales across different product categories?
- Identify trends in purchase behaviour over time (e.g., monthly).

Conclusion

This project demonstrates how a structured, data-driven approach can support the retail industry by improving customer retention and revenue. The solution provides a strong foundation for future scalability and data-led innovation at Globex Retail.

This project was completed as part of my training at 10Alytics, where I am building hands-on, job-ready experience in Data Engineering.

THANK YOU