# Customer Behavior Analysis

## Project Overview

This project analyzes customer shopping behavior data to uncover purchasing trends, spending habits, and the impact of demographic and marketing factors on sales. The analysis integrates Python for data cleaning, SQL for querying insights, and Power BI for visualization.

## Objectives

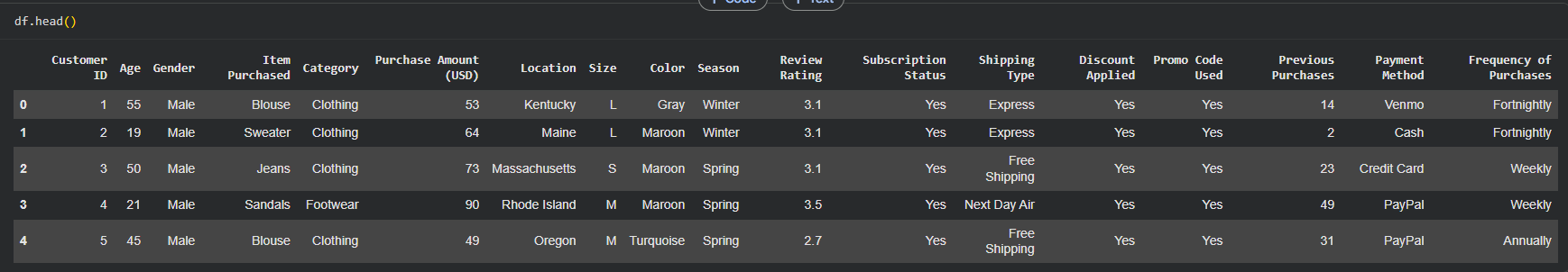
* Clean and preprocess the raw dataset for analysis
* Identify top-performing products and customer segments
* Compare spending behavior of subscribers vs. non-subscribers
* Examine the relationship between discounts, ratings, and loyalty
* Visualize insights interactively through a BI dashboard

## Data Cleaning & Transformation (Python)

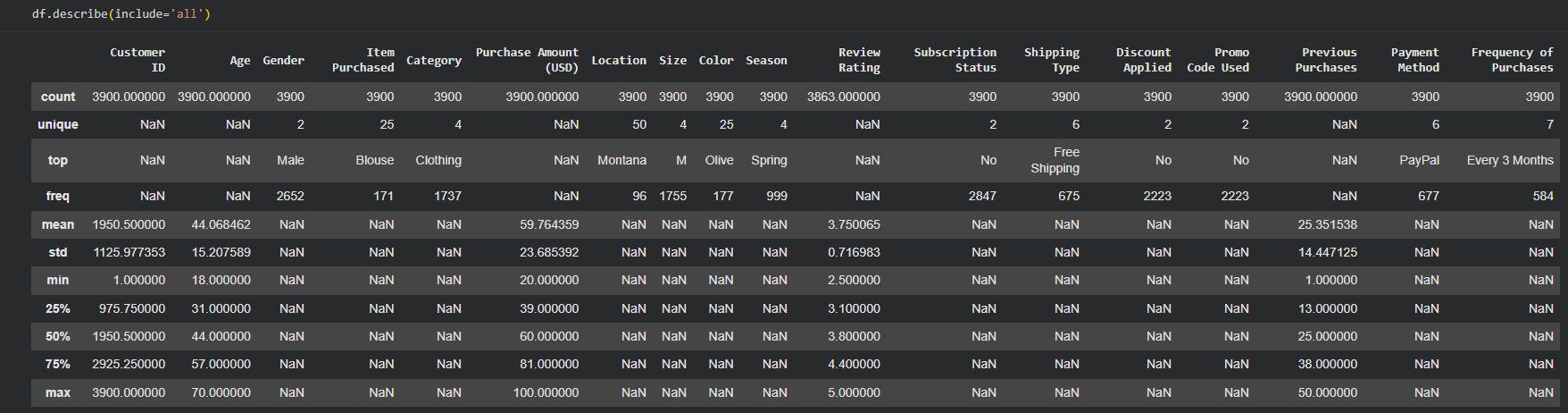
The dataset was cleaned and prepared using Pandas in Python. Missing review ratings were filled with median values by category, columns were standardized, and age groups were created.

Code Snippet:

* At first I took a look on the data



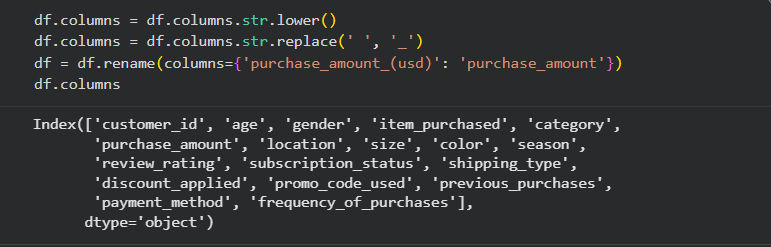
* Then I inspected it further



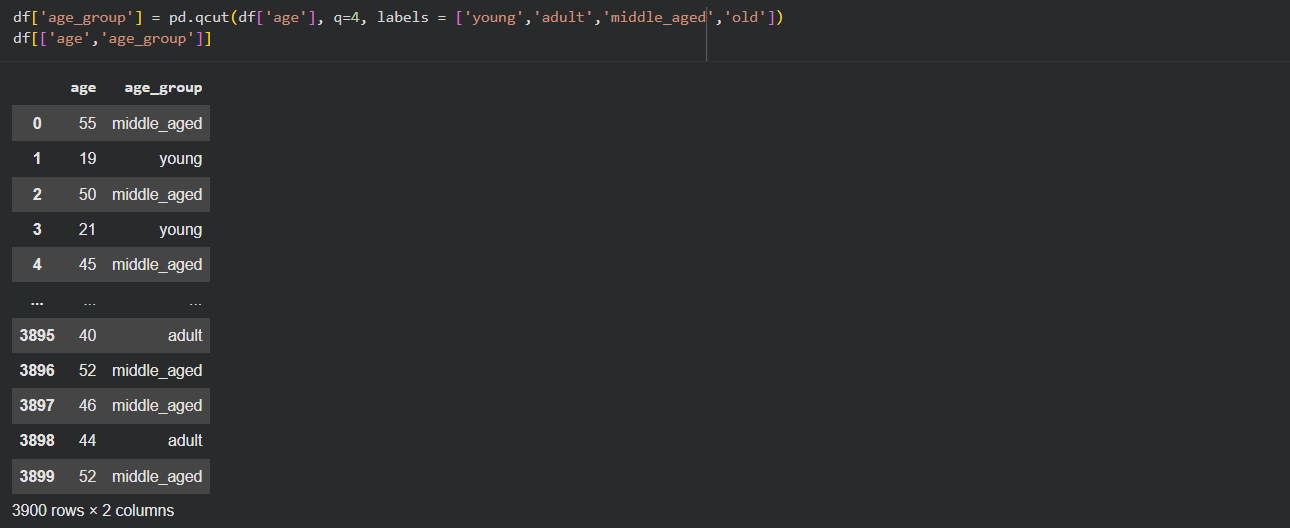
* Filled missing review ratings by the median based of each category



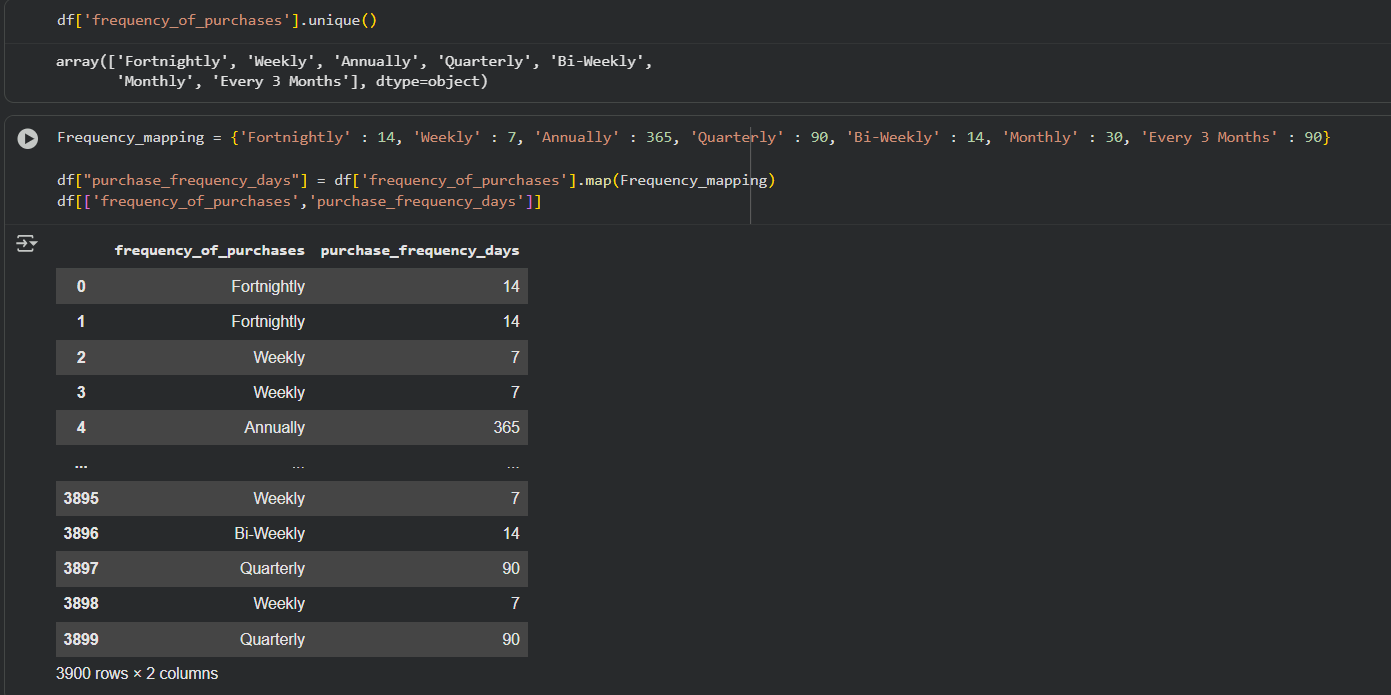
* Converted column names to normalize it



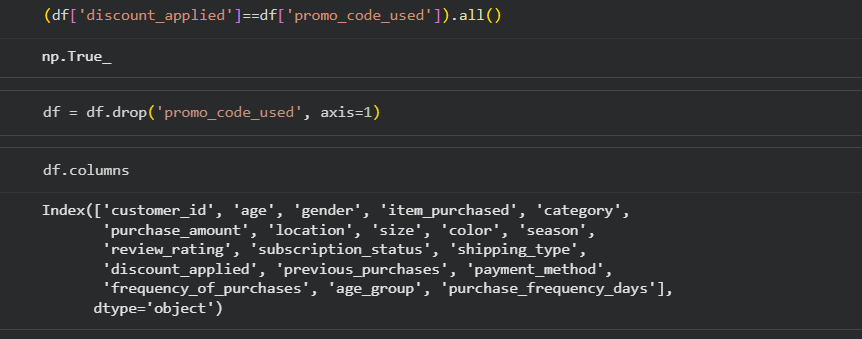
* Created age segments



* Mapped frequency of purchases to numeric days

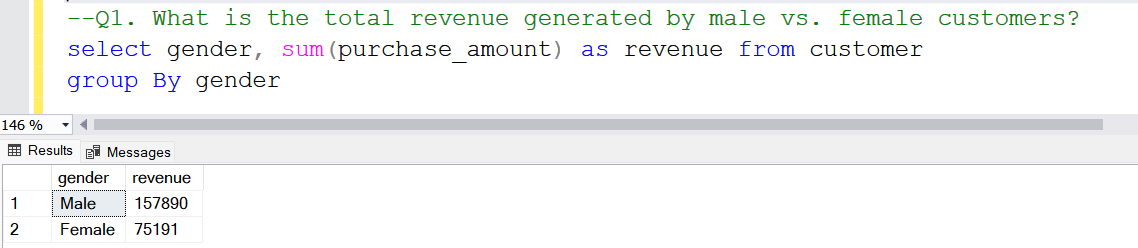


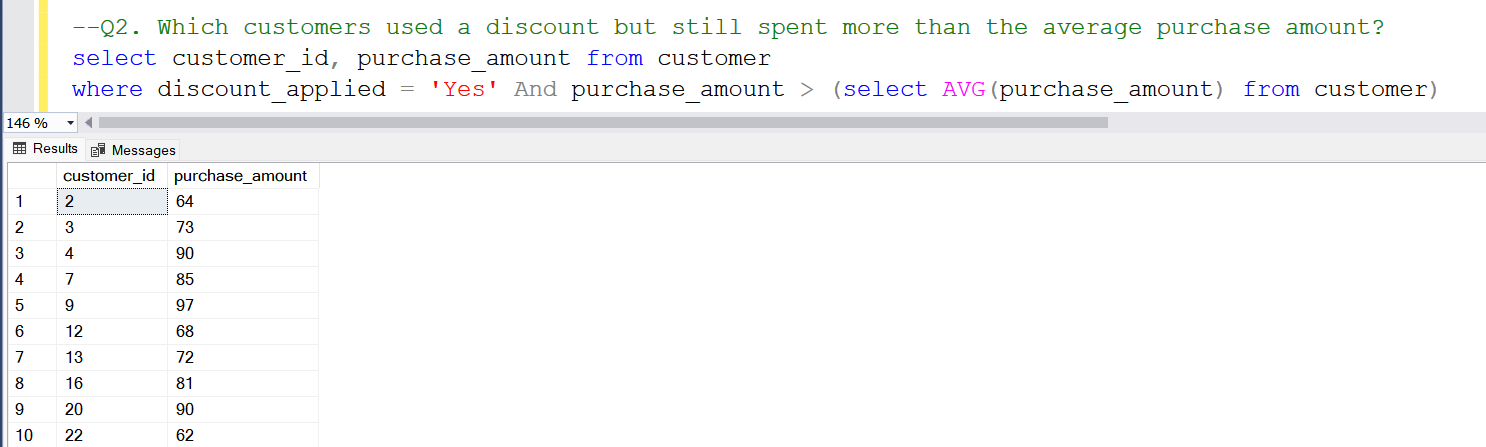
* Checked for redundant column and Removed it

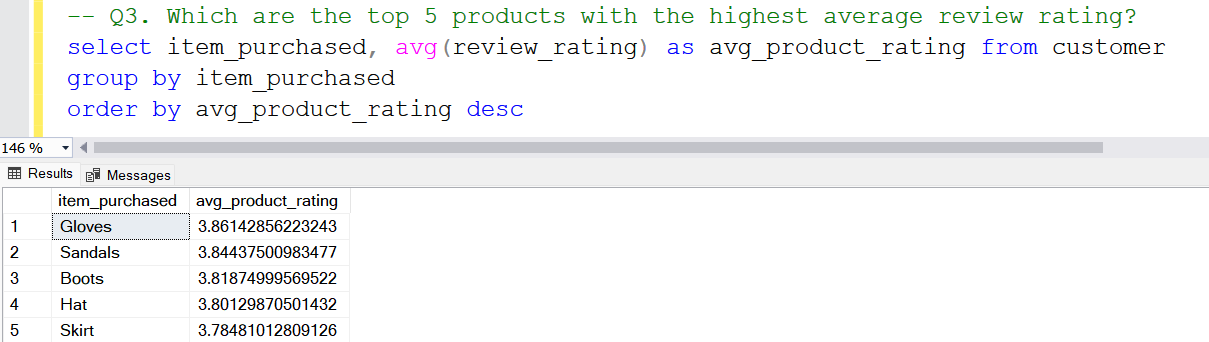


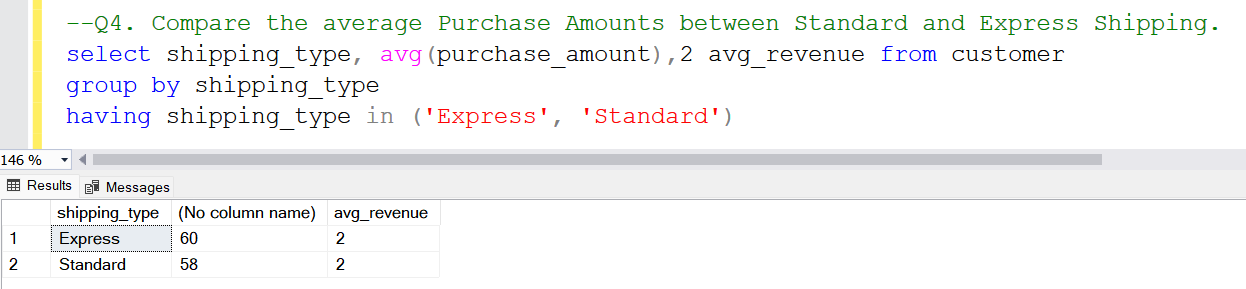
## SQL Analytical Queries

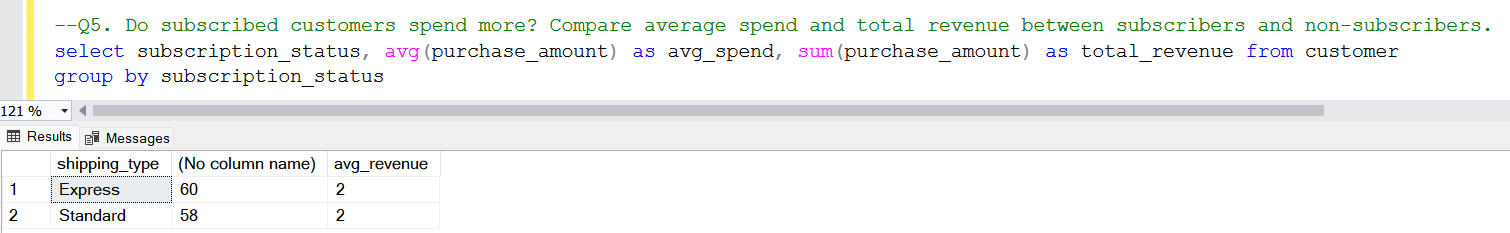
After importing my cleaned Data to MS [SQL Server](https://learn.microsoft.com/en-us/ssms/sql-server-management-studio-ssms" \t "https://search.yahoo.com/_blank) I've created a database called Customer to explor the Data Further and to make business insights to answer the business questions below

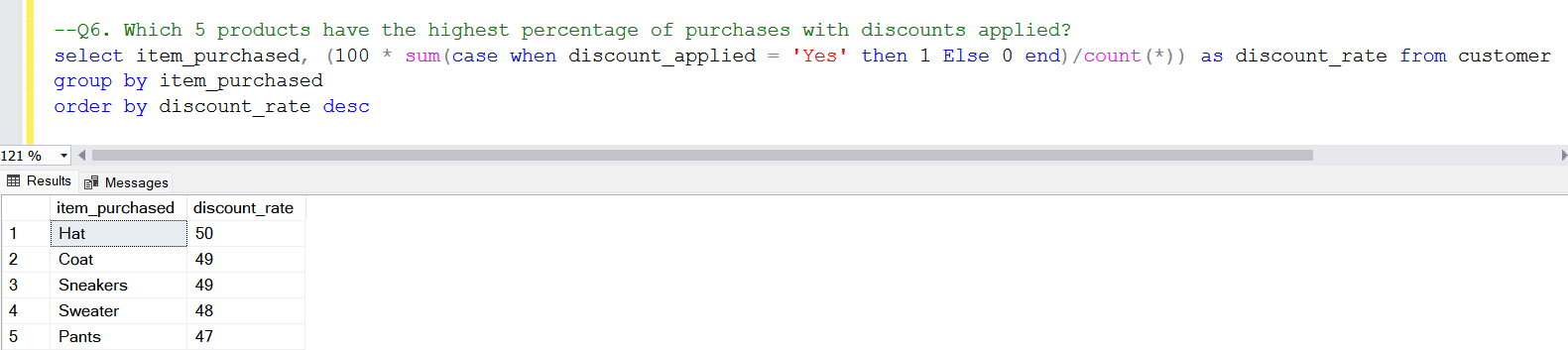


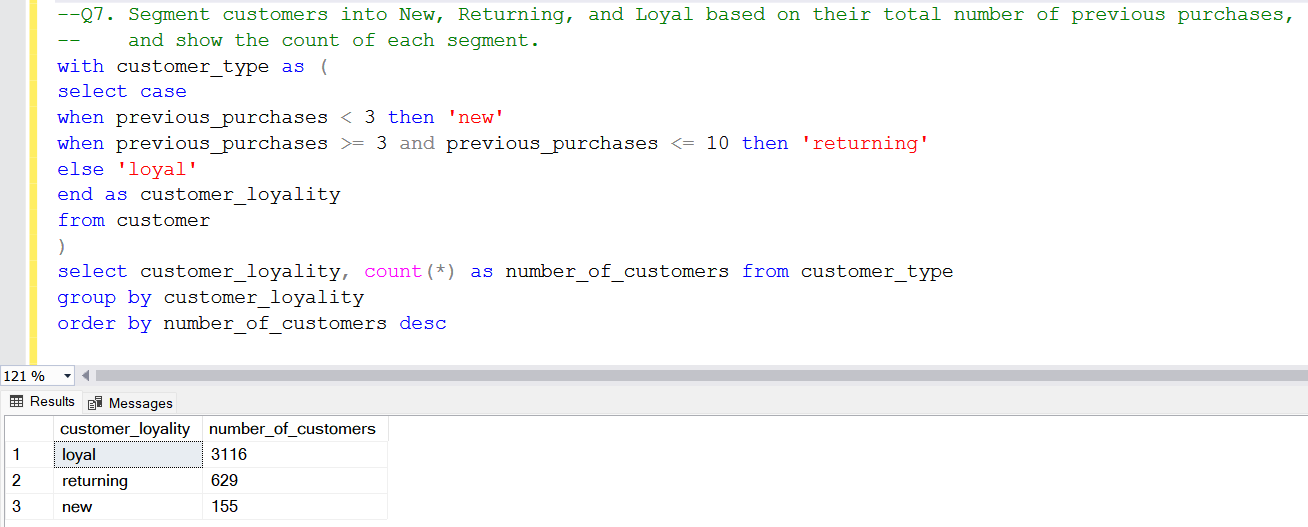


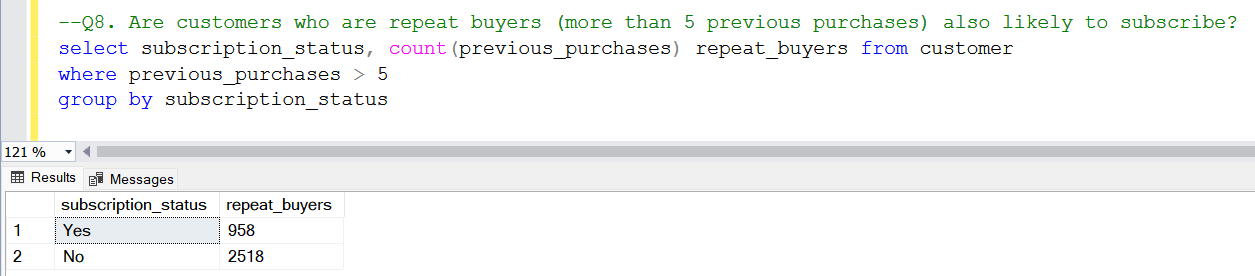












## Power BI Dashboard

The dashboard visualizes key performance metrics, customer segmentation, and sales distribution. It includes interactive filters for gender, subscription status, age group, and product category.

## customer_behaviour

## Key Insights

* Subscribers spend approximately 25% more than non-subscribers.
* Clothing and Accessories generate the highest revenue and sales volume.
* Young and middle-aged segments dominate total revenue contribution.
* Discounts increase purchase volume but not necessarily total revenue per order.
* Standard shipping has higher transaction volume; Express shipping has higher order value.

## Tools & Technologies

* Python (Pandas) for data cleaning & preprocessing
* SQL for data exploration and querying
* Power BI for data visualization and storytelling

## Future Work

* Add predictive analytics (e.g., churn prediction, LTV modeling).
* Automate ETL pipelines for data refreshes.
* Add customer segmentation clustering (e.g., K-Means).

## Links

* This project can be found on GitHub [HERE](https://github.com/abdallah22amr/Customer-Behavior-Analysis) and on my Linkedin HERE