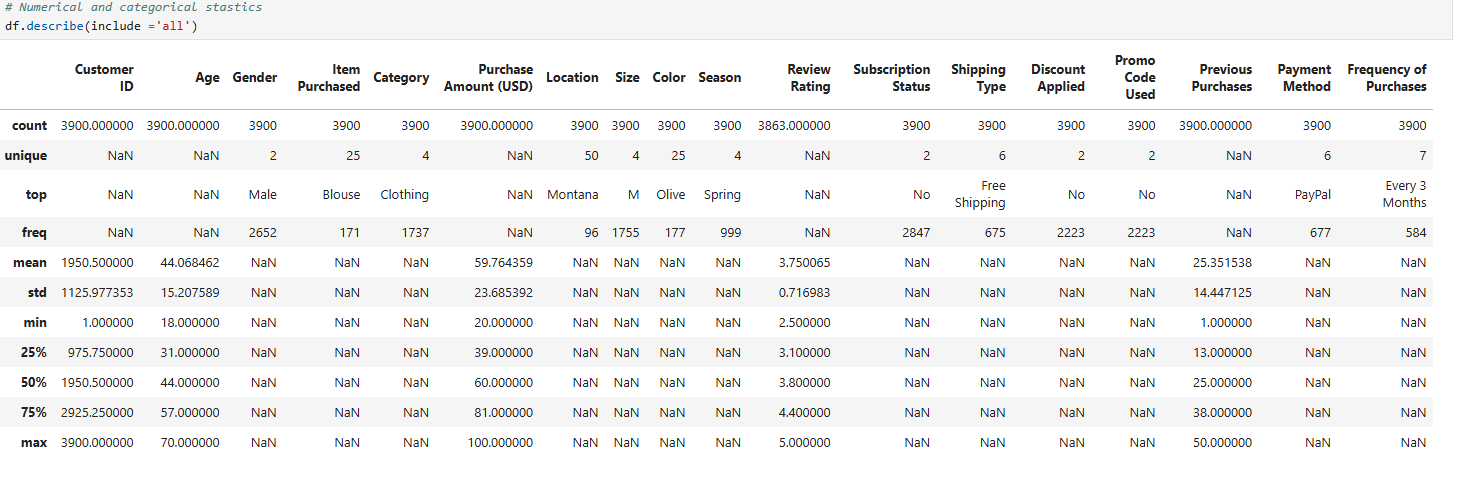
Customer Shopping Behavior Analysis

1. **Project Overview:** This project analyzes customer shopping behavior using transactional data from 3,900 purchases across various categories. The goal is to uncover insights into spending patterns, customers segments, product preferences and subscription behavior to guide strategic business decisions
2. **Dataset Summary:**

* Rows: 3900
* Columns: 18
* Key Feature
* Customer demographics (Age, gender, Location, Subscription Status)
* Purchase behavior (Discount Applied, Promo code, previous Purchase, Frequency of Purchase, Review Rating, Shipping Type)
* Missing Dat: 37 values in Review Rating column

1. **Exploratory Data Analysis using Python**

* I began with data preparation and cleaning in python:
* Data Loading: Imported the data set using pandas,
* Initial Exploration**:** Used df.info() to check data structure and .describe() for summary statistics:

****

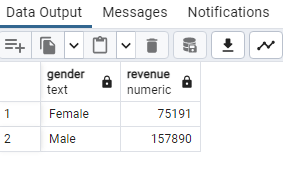
--

* **Missing Data Handling**: Checked for null values and imputed missing values in the Review Rating Column using the median rating of each product category.
* **Column Standardization**: Rename columns to **snake case** foe better readability and documentation
* **Feature Engineering**
* Created **age\_group** column by binning customer ages
* Crated **purchase\_frequecny\_days** column from purchase data.
* Data Consistency Check: Verified if discount\_applied and promo\_code\_used were redundant : dropped promo\_code\_used.
* Database integration: Connected Python script to PostgreSQL and loaded the cleaned DataFrame into the database for SQL analysis

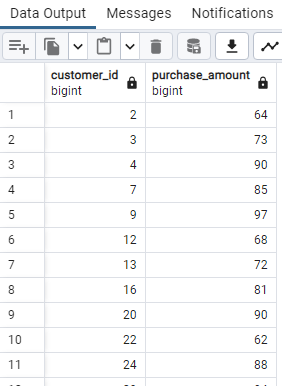
1. Data Analysis using SQL

Performed structured analysis in PostgreSQL to answer key business question

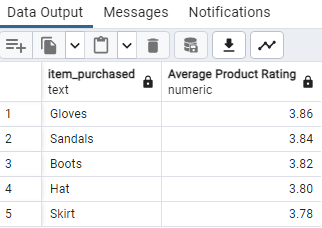
1. **Revenue by Gender**: Compared total revenue generated by male vs female customers.



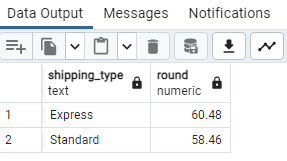
1. High-Spending Discount Users: Identified customers who used discounts but still spent above the average purchase amount.



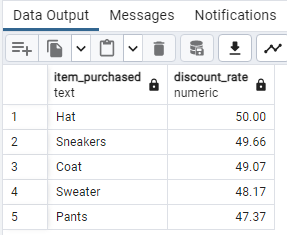
1. Top 5 Products by Rating: Found products with the highest average review ratings.



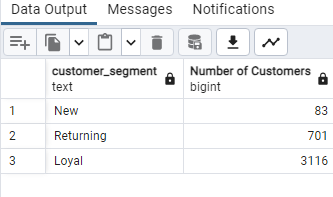
1. Compare the average purchase amounts between standard and express shipping



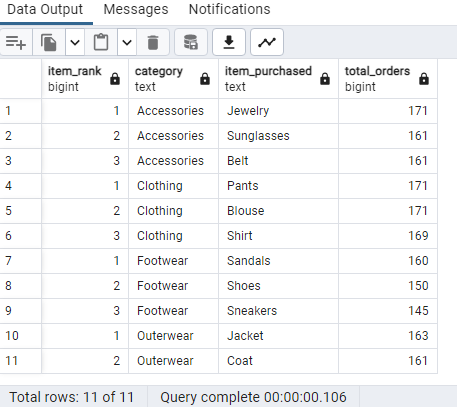
1. Do subscribed customers spend more? compare average spend and total revenue between subscribes and no subscribe



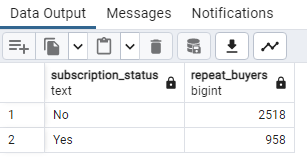
1. Segment customers into New, Returning and Loyal bases on their total number of previous purchase and show the count of each segment.



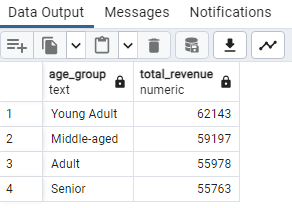
1. What are the top 3 most purchased products within each category?



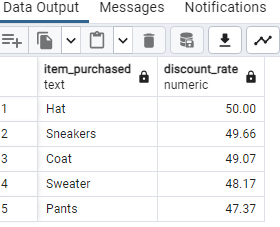
1. Are customers who are repeat buyers (more than 5 previous purchase) also likely to subscribe?

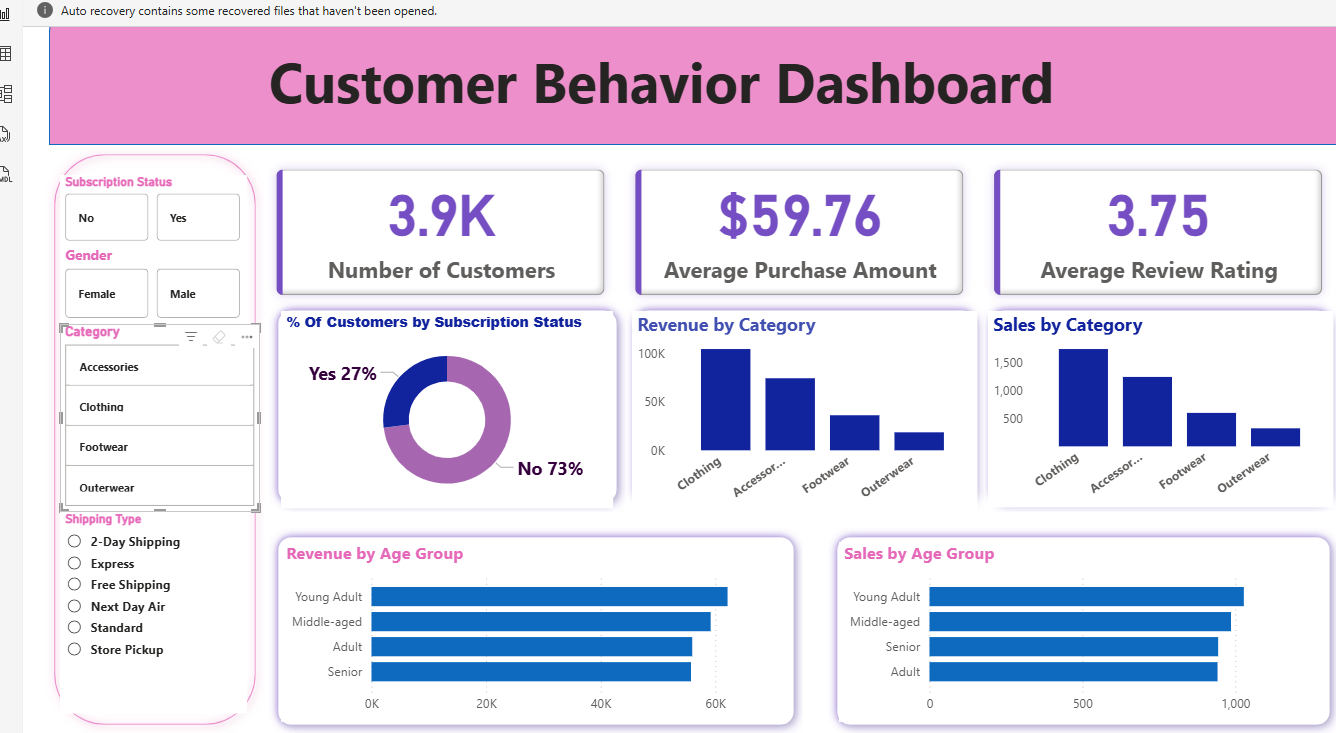


1. What is the revenue contribution of each age group?



1. Which 5 products have the highest percentage of purchase with discounts applied?



PowerBI Dash Board