Customer Behavior Analysis Project

Project Name: Customer Behavior
Domain: Data Analytics / Business Intelligence
Tools Used: Python (Pandas), PyCharm, MySQL, Power BI

1. Introduction

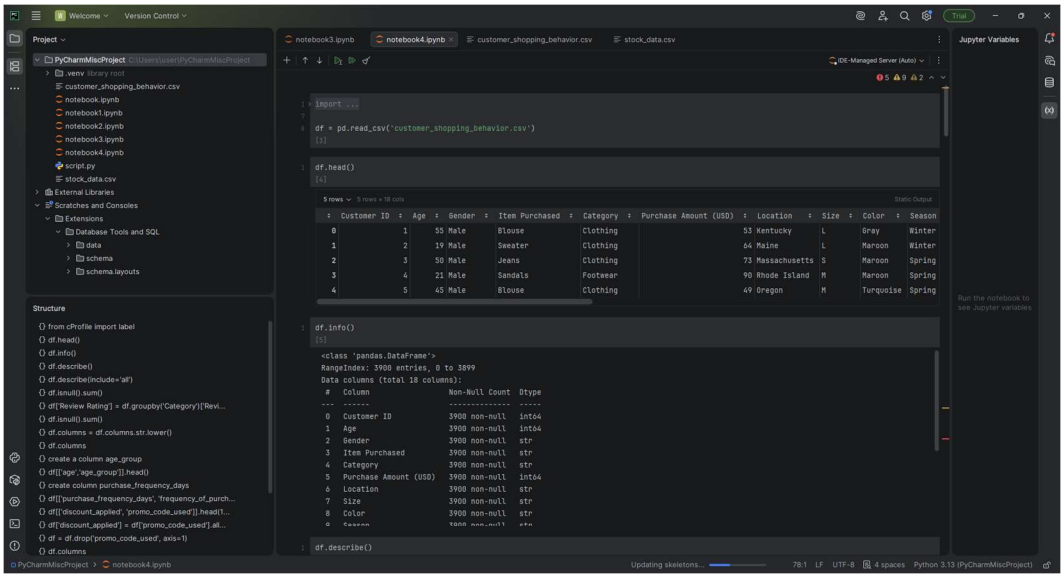
This project focuses on analyzing customer shopping behavior to understand purchasing patterns, demographics, payment preferences, and the impact of discounts and promotions. The insights generated help businesses improve decision-making, customer targeting, and revenue growth.

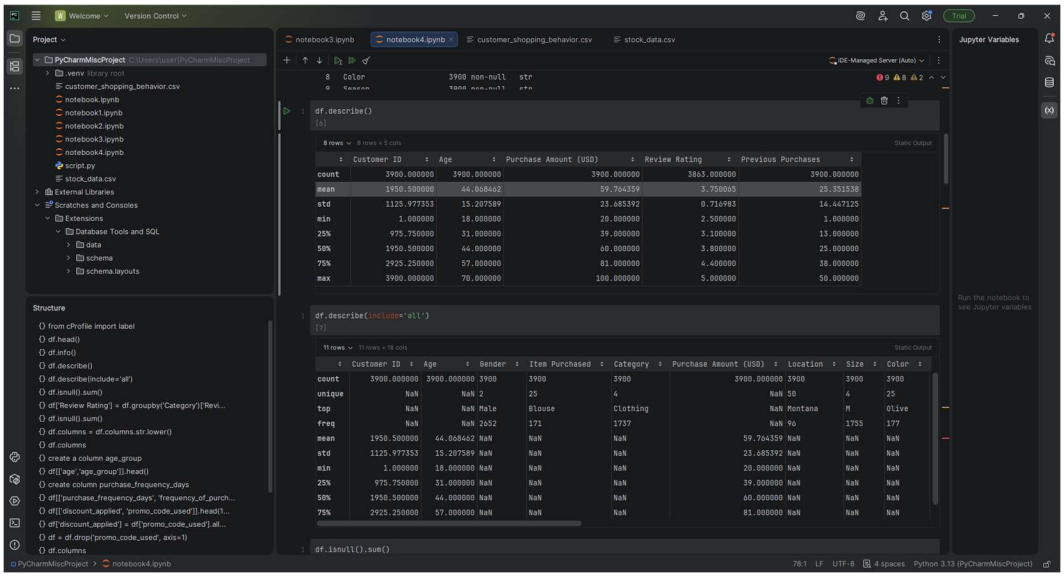
2. Dataset Description

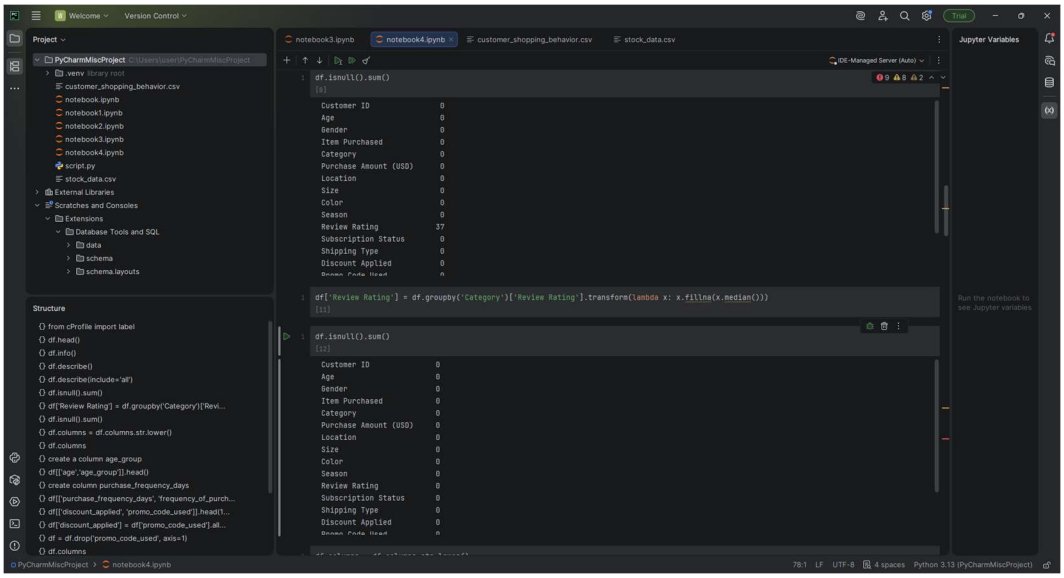
The dataset 'customer_shopping_behavior.csv' contains 3900 records with 18 attributes including:
- Customer demographics (Age, Gender, Location)
- Purchase details (Item, Category, Amount, Season)
- Behavioral metrics (Review Rating, Previous Purchases)
- Payment and promotion details (Payment Method, Discount Applied, Promo Code Used)

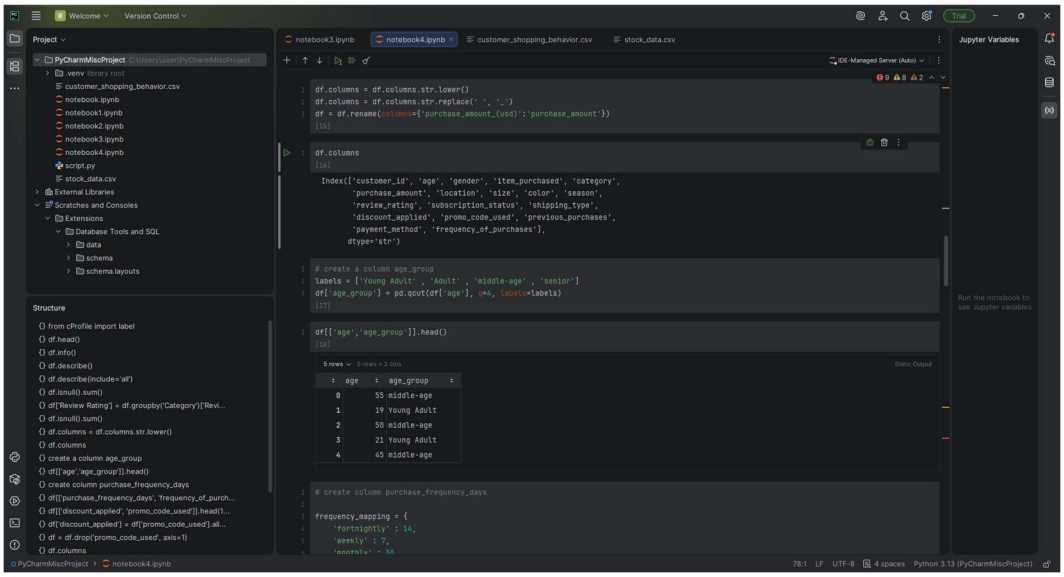
3. Data Preprocessing & Analysis (Python)

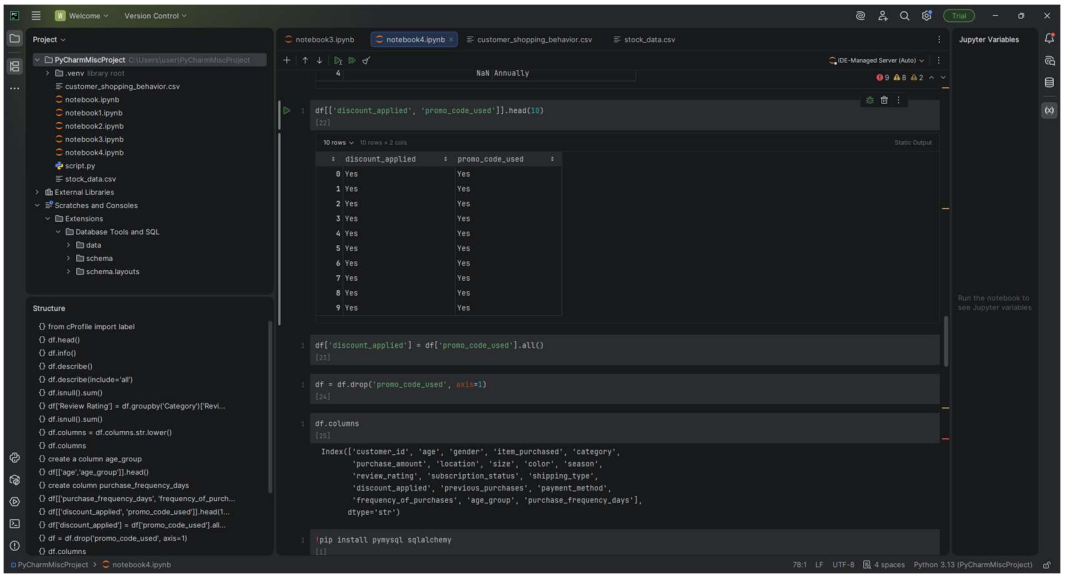
Using Python in PyCharm and Pandas, the following steps were performed:
- Loaded CSV data
- Checked data types and missing values
- Filled missing review ratings using category-wise median
- Created new features like age_group and purchase_frequency_days
- Cleaned and standardized column names

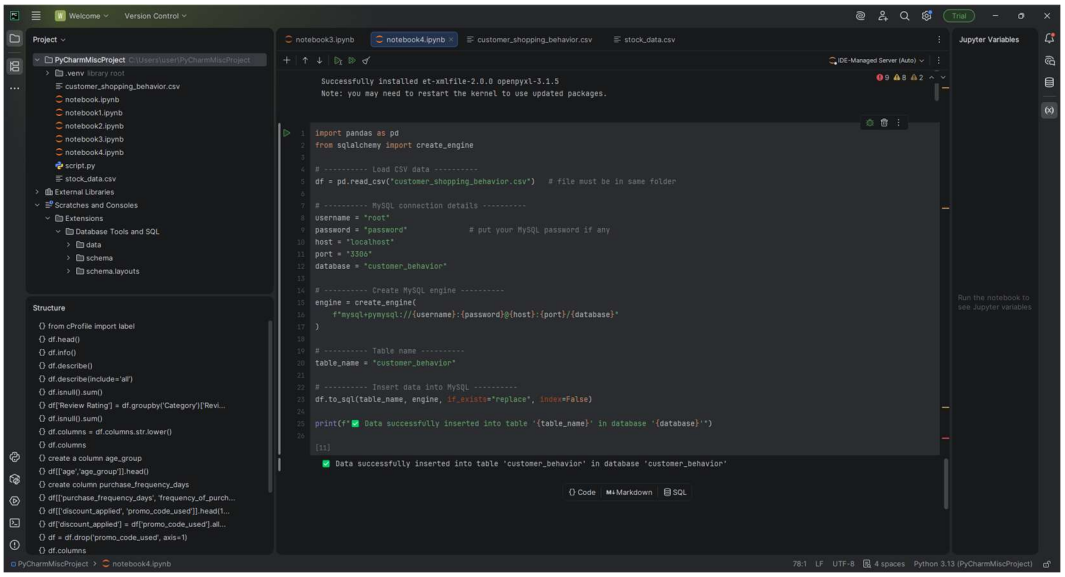

Welcome
Version Control
Trial
Project
PyCharmMiscProject C:\Users\user\PyCharmMiscProject
.venv library root
customer_shopping_behavior.csv
notebook.ipynb
notebook1.ipynb
notebook2.ipynb
notebook3.ipynb
notebook4.ipynb
script.py
stock_data.csv
External Libraries
Scratches and Consoles
Extensions
Database Tools and SQL
data
schema
schema.layouts
Structure
{} from cProfile import label
{} df.head()
{} df.info()
{} df.describe()
{} df.describe(include='all')
{} df.isnull().sum()
{} df['Review Rating'] = df.groupby('Category')['Revi...
{} df.isnull().sum()
{} df.columns = df.columns.str.lower()
{} df.columns
{} create a column age_group
{} df[['age','age_group']].head()
{} create column purchase_frequency_days
{} df[['purchase_frequency_days', 'frequency_of_purch...
{} df[['discount_applied', 'promo_code_used']].head(1...
{} df['discount_applied'] = df['promo_code_used'].all...
{} df = df.drop('promo_code_used', axis=1)
{} df.columns
notebook3.ipynb
notebook4.ipynb
customer_shopping_behavior.csv
stock_data.csv
Jupyter Variables
IDE-Managed Server (Auto)
import ...
df = pd.read_csv('customer_shopping_behavior.csv')
[3]
df.head()
[4]
5 rows
5 rows × 18 cols
Static Output
Customer ID
Age
Gender
Item Purchased
Category
Purchase Amount (USD)
Location
Size
Color
Season
0 1 55 Male Blouse Clothing 53 Kentucky L Gray Winter
1 2 19 Male Sweater Clothing 64 Maine L Maroon Winter
2 3 50 Male Jeans Clothing 73 Massachusetts S Maroon Spring
3 4 21 Male Sandals Footwear 90 Rhode Island M Maroon Spring
4 5 45 Male Blouse Clothing 49 Oregon M Turquoise Spring
df.info()
[5]
<class 'pandas.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 18 columns):
Column Non-Null Count Dtype
0 Customer ID 3900 non-null int64
1 Age 3900 non-null int64
2 Gender 3900 non-null str
3 Item Purchased 3900 non-null str
4 Category 3900 non-null str
5 Purchase Amount (USD) 3900 non-null int64
6 Location 3900 non-null str
7 Size 3900 non-null str
8 Color 3900 non-null str
df.describe()
Run the notebook to see Jupyter variables
PyCharmMiscProject
notebook4.ipynb
Updating skeletons...
78:1
LF
UTF-8
4 spaces
Python 3.13 (PyCharmMiscProject)


Welcome
Version Control
Trial
Project
PyCharmMiscProject C:\Users\user\PyCharmMiscProject
.venv library root
customer_shopping_behavior.csv
notebook.ipynb
notebook1.ipynb
notebook2.ipynb
notebook3.ipynb
notebook4.ipynb
script.py
stock_data.csv
External Libraries
Scratches and Consoles
Extensions
Database Tools and SQL
data
schema
schema.layouts
Structure
{} from cProfile import label
{} df.head()
{} df.info()
{} df.describe()
{} df.describe(include='all')
{} df.isnull().sum()
{} df['Review Rating'] = df.groupby('Category')['Revi...
{} df.isnull().sum()
{} df.columns = df.columns.str.lower()
{} df.columns
{} create a column age_group
{} df[['age','age_group']].head()
{} create column purchase_frequency_days
{} df[['purchase_frequency_days', 'frequency_of_purch...
{} df[['discount_applied', 'promo_code_used']].head(1...
{} df['discount_applied'] = df['promo_code_used'].all...
{} df = df.drop('promo_code_used', axis=1)
{} df.columns
notebook3.ipynb
notebook4.ipynb
customer_shopping_behavior.csv
stock_data.csv
Jupyter Variables
IDE-Managed Server (Auto)
8 Color 3900 non-null str
9 Season 3900 non-null str
df.describe()
[6]
8 rows
8 rows × 5 cols
Static Output
Customer ID
Age
Purchase Amount (USD)
Review Rating
Previous Purchases
count 3900.000000 3900.000000 3900.000000 3863.000000 3900.000000
mean 1950.500000 44.068462 59.764359 3.750065 25.351538
std 1125.977353 15.207589 23.685392 0.716983 14.447125
min 1.000000 18.000000 20.000000 2.500000 1.000000
25% 975.750000 31.000000 39.000000 3.100000 13.000000
50% 1950.500000 44.000000 60.000000 3.800000 25.000000
75% 2925.250000 57.000000 81.000000 4.400000 38.000000
max 3900.000000 70.000000 100.000000 5.000000 50.000000
df.describe(include='all')
[7]
11 rows
11 rows × 18 cols
Static Output
Customer ID
Age
Gender
Item Purchased
Category
Purchase Amount (USD)
Location
Size
Color
count 3900.000000 3900.000000 3900 3900 3900 3900.000000 3900 3900 3900
unique NaN NaN 2 25 4 NaN 50 4 25
top NaN NaN Male Blouse Clothing NaN Montana M Olive
freq NaN NaN 2652 171 1737 NaN 96 1755 177
mean 1950.500000 44.068462 NaN NaN NaN 59.764359 NaN NaN NaN
std 1125.977353 15.207589 NaN NaN NaN 23.685392 NaN NaN NaN
min 1.000000 18.000000 NaN NaN NaN 20.000000 NaN NaN NaN
25% 975.750000 31.000000 NaN NaN NaN 39.000000 NaN NaN NaN
50% 1950.500000 44.000000 NaN NaN NaN 60.000000 NaN NaN NaN
75% 2925.250000 57.000000 NaN NaN NaN 81.000000 NaN NaN NaN
df.isnull().sum()
Run the notebook to see Jupyter variables
PyCharmMiscProject
notebook4.ipynb
78:1
LF
UTF-8
4 spaces
Python 3.13 (PyCharmMiscProject)


Welcome
Version Control
Trial
Project
PyCharmMiscProject C:\Users\user\PyCharmMiscProject
.venv library root
customer_shopping_behavior.csv
notebook.ipynb
notebook1.ipynb
notebook2.ipynb
notebook3.ipynb
notebook4.ipynb
script.py
stock_data.csv
External Libraries
Scratches and Consoles
Extensions
Database Tools and SQL
data
schema
schema.layouts
Structure
{} from cProfile import label
{} df.head()
{} df.info()
{} df.describe()
{} df.describe(include='all')
{} df.isnull().sum()
{} df['Review Rating'] = df.groupby('Category')['Revi...
{} df.isnull().sum()
{} df.columns = df.columns.str.lower()
{} df.columns
{} create a column age_group
{} df[['age','age_group']].head()
{} create column purchase_frequency_days
{} df[['purchase_frequency_days', 'frequency_of_purch...
{} df[['discount_applied', 'promo_code_used']].head(1...
{} df['discount_applied'] = df['promo_code_used'].all...
{} df = df.drop('promo_code_used', axis=1)
{} df.columns
notebook3.ipynb
notebook4.ipynb
customer_shopping_behavior.csv
stock_data.csv
IDE-Managed Server (Auto)
Jupyter Variables
Run the notebook to see Jupyter variables
df.isnull().sum()
Customer ID 0
Age 0
Gender 0
Item Purchased 0
Category 0
Purchase Amount (USD) 0
Location 0
Size 0
Color 0
Season 0
Review Rating 37
Subscription Status 0
Shipping Type 0
Discount Applied 0
df['Review Rating'] = df.groupby('Category')['Review Rating'].transform(lambda x: x.fillna(x.median()))
df.isnull().sum()
Customer ID 0
Age 0
Gender 0
Item Purchased 0
Category 0
Purchase Amount (USD) 0
Location 0
Size 0
Color 0
Season 0
Review Rating 0
Subscription Status 0
Shipping Type 0
Discount Applied 0
PyCharmMiscProject
notebook4.ipynb
78:1
LF
UTF-8
4 spaces
Python 3.13 (PyCharmMiscProject)

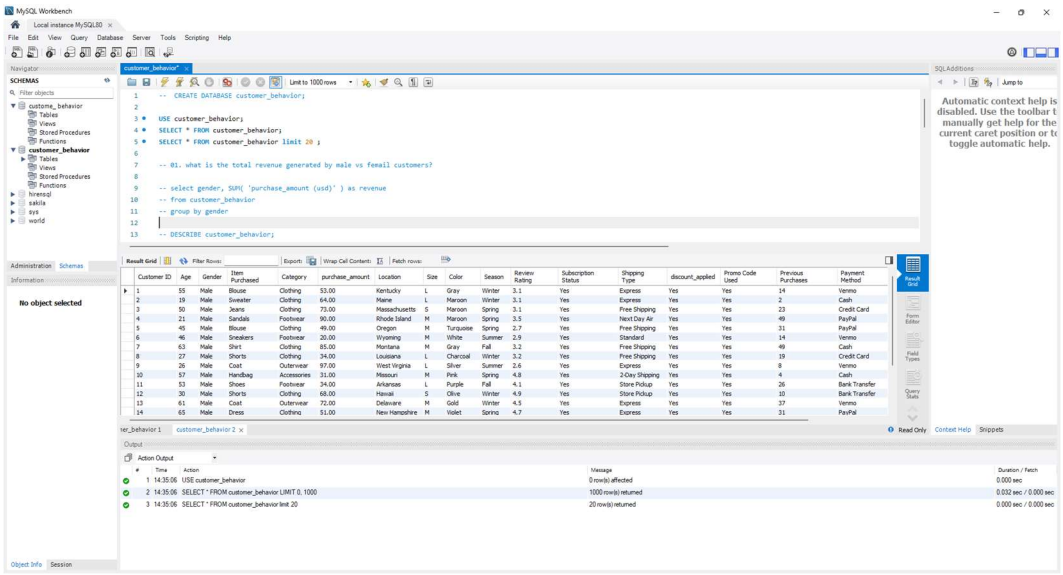

Welcome
Version Control
Trial
Project
PyCharmMiscProject C:\Users\user\PyCharmMiscProject
.venv library root
customer_shopping_behavior.csv
notebook.ipynb
notebook1.ipynb
notebook2.ipynb
notebook3.ipynb
notebook4.ipynb
script.py
stock_data.csv
External Libraries
Scratches and Consoles
Extensions
Database Tools and SQL
data
schema
schema.layouts
Structure
{} from cProfile import label
{} df.head()
{} df.info()
{} df.describe()
{} df.describe(include='all')
{} df.isnull().sum()
{} df['Review Rating'] = df.groupby('Category')['Revi...
{} df.isnull().sum()
{} df.columns = df.columns.str.lower()
{} df.columns
{} create a column age_group
{} df[['age','age_group']].head()
{} create column purchase_frequency_days
{} df[['purchase_frequency_days', 'frequency_of_purch...
{} df[['discount_applied', 'promo_code_used']].head(1...
{} df['discount_applied'] = df['promo_code_used'].all...
{} df = df.drop('promo_code_used', axis=1)
{} df.columns
notebook3.ipynb
notebook4.ipynb
customer_shopping_behavior.csv
stock_data.csv
IDE-Managed Server (Auto)
Jupyter Variables
Run the notebook to see Jupyter variables
df.columns = df.columns.str.lower()
df.columns = df.columns.str.replace(' ', '_')
df = df.rename(columns={'purchase_amount_(usd)':'purchase_amount'})
df.columns
Index(['customer_id', 'age', 'gender', 'item_purchased', 'category',
'purchase_amount', 'location', 'size', 'color', 'season',
'review_rating', 'subscription_status', 'shipping_type',
'discount_applied', 'promo_code_used', 'previous_purchases',
'payment_method', 'frequency_of_purchases'],
dtype='str')
# create a column age_group
labels = ['Young Adult' , 'Adult' , 'middle-age' , 'senior']
df['age_group'] = pd.qcut(df['age'], q=4, labels=labels)
df[['age','age_group']].head()
5 rows
5 rows × 2 cols
Static Output
age
age_group
0 55 middle-age
1 19 Young Adult
2 50 middle-age
3 21 Young Adult
4 45 middle-age
# create column purchase_frequency_days
frequency_mapping = {
'fortnightly' : 14,
'weekly' : 7,
'monthly' : 30
PyCharmMiscProject
notebook4.ipynb
78:1
LF
UTF-8
4 spaces
Python 3.13 (PyCharmMiscProject)





4. Database Integration (MySQL)

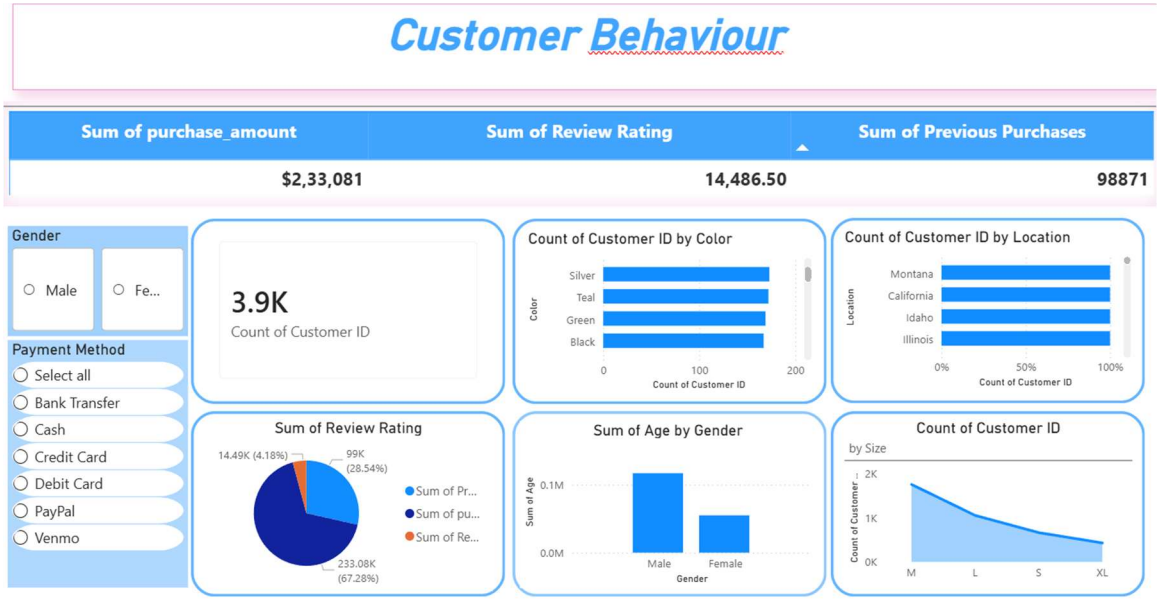
After preprocessing, the cleaned dataset was exported to MySQL using SQLAlchemy. This enabled structured querying and advanced analysis using SQL.



5. Data Visualization (Power BI)

Power BI was used to create an interactive dashboard showcasing:

- Total revenue and customer count
- Gender-wise purchase analysis
- Location and color preferences
- Payment method distribution
- Customer size and age group insights



6. Key Insights

- Male customers generated higher total revenue
- Clothing category dominated overall purchases
- Medium (M) size products were most popular
- Credit Card and PayPal were the most used payment methods
- Discounts significantly increased purchase frequency

7. Conclusion

The Customer Behavior project successfully demonstrates an end-to-end data analytics pipeline from raw data to business insights using Python, MySQL, and Power BI. This project is suitable for academic submission and showcases practical industry-relevant skills.

8. GitHub Repository Structure

- data/customer_shopping_behavior.csv
- notebooks/Python analysis notebooks
- database/MySQL scripts
- dashboard/Power BI file
- report/Customer_Behavior_Project_Report.docx