**E-Commerce Consumer Behaviour Analysis and Satisfaction Prediction**

Vishal Menaria

****

Github Link - <https://github.com/vishalmenaria>

E-mail – [vishalkun530@gmail.com](mailto:vishalkun530@gmail.com)

Mobile No.- 8114494484

Date – 15 February 2025

**PROJECT OVERVIEW**

This project analyses consumer behaviour in the e-commerce industry using demographic and transactional data. The goal is to identify key patterns and build a predictive model to forecast customer satisfaction.

**OBJECTIVE**

1.Analyse customer behaviour based on gender, city, and purchase amounts.  
2. Identify key factors influencing customer satisfaction.  
 3. Build a machine learning model to predict customer satisfaction

**Dataset Overview**

* **Source:** Kaggle
* **Size:** 350 rows, 11 columns
* **Key Features:**
  + **Customer ID** – Unique identifier for each customer
  + **Gender** – Male or Female
  + **City** – City where the customer resides
  + **Purchase Amount** – Total purchase value per transaction
  + **Average Rating** - Rating given by customer
  + **Satisfaction** – Customer satisfaction level (target variable)

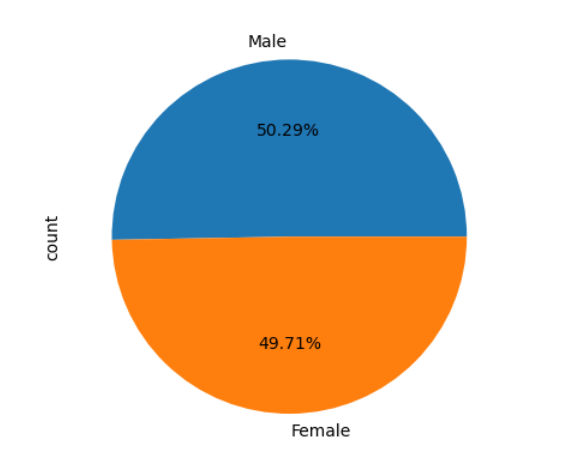
**Data Cleaning and Preprocessing**

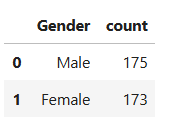
* Handled missing values using mean/median imputation.
* Removed duplicates and irrelevant columns.
* Encoded categorical variables using **LabelEncoding.**
* Scaled numerical data for better model performance.

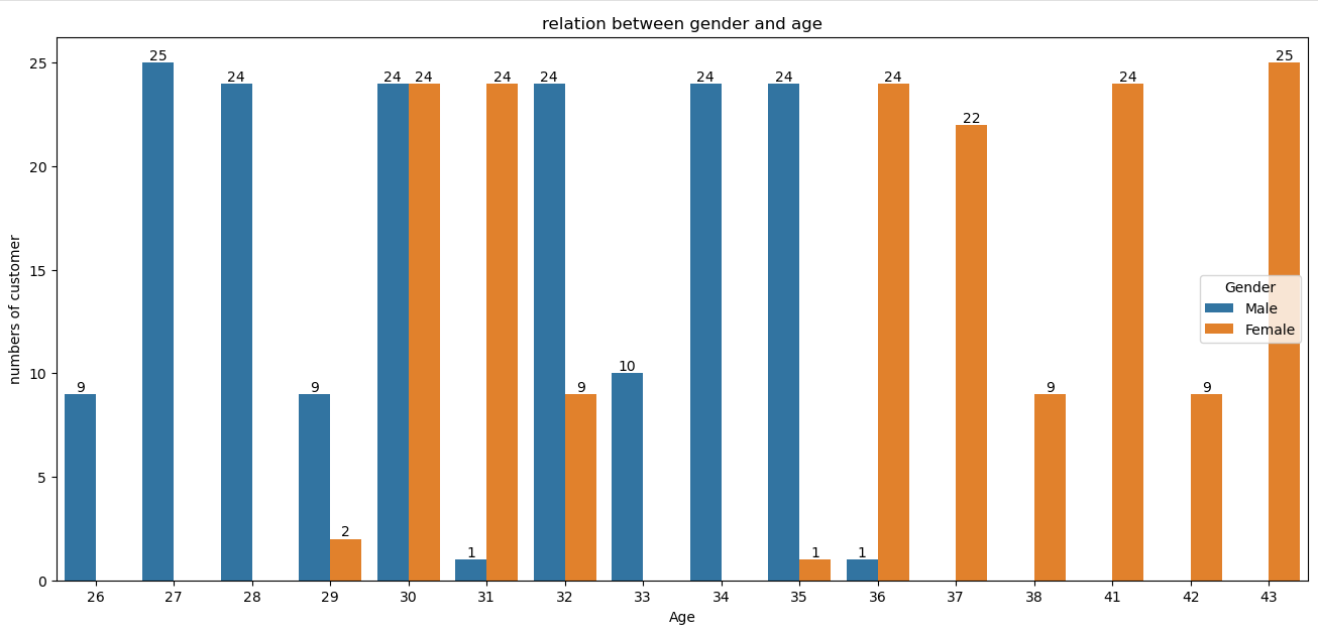
**Exploratory Data Analysis (EDA)**

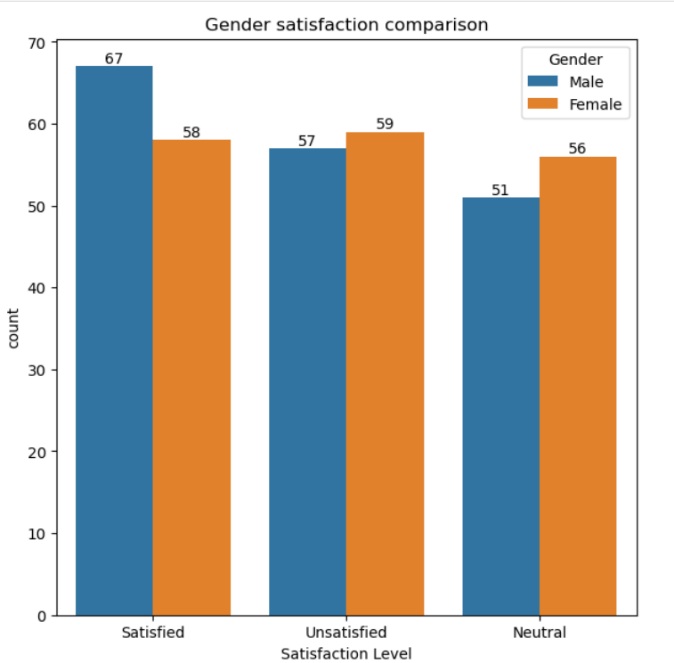
**Gender-Based Analysis:**

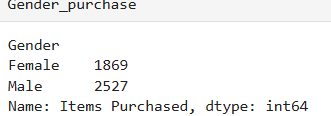
**1**.In our Dataset The male and female customers are almost equal in numbers and male consumers are rather young age then female.

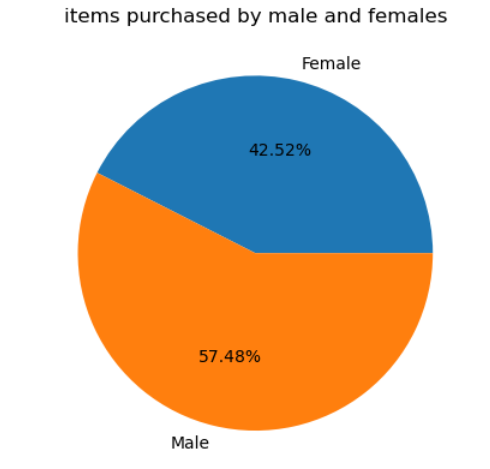
****

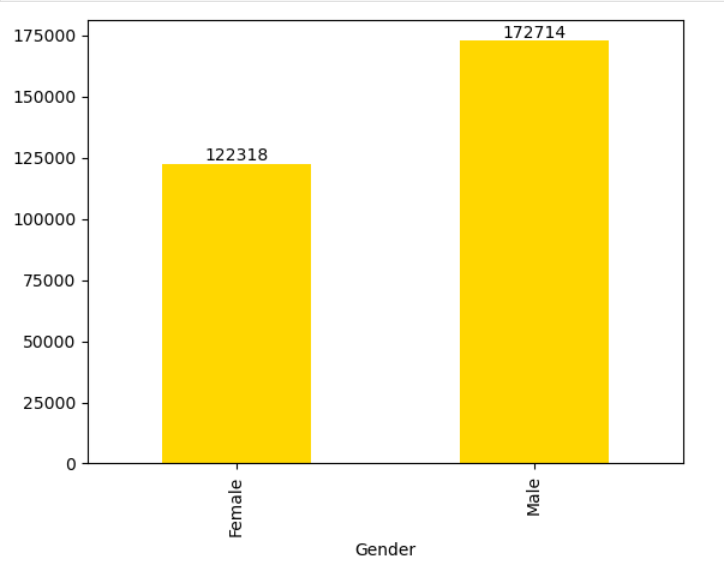
****

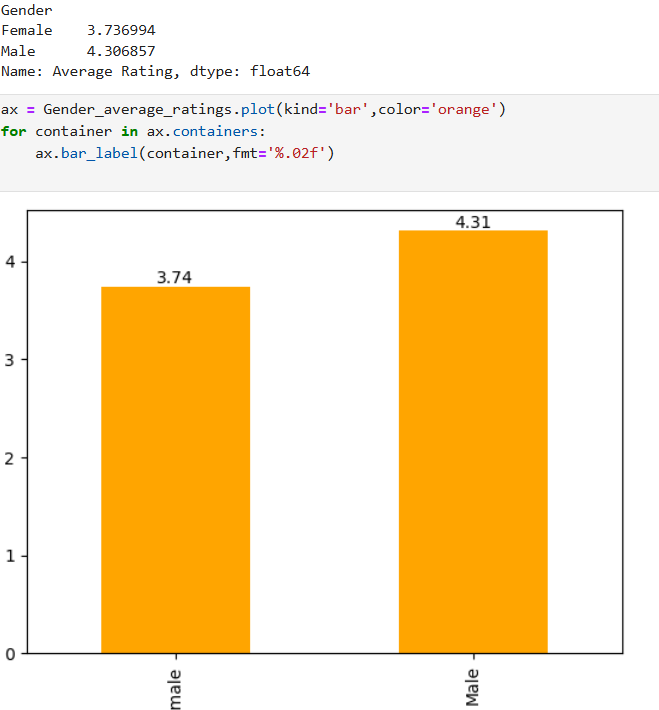
****

****2. Males are slightly more satisfied with Product and service then female consumers

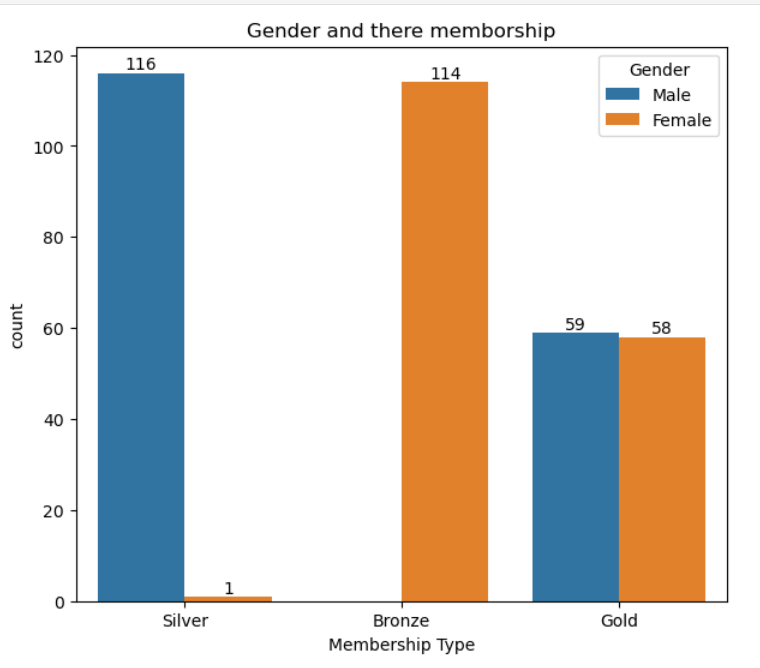
**3. Males are purchasing more product and generating more revenue than females also males are rating higher and comparatively more satisfied than females**

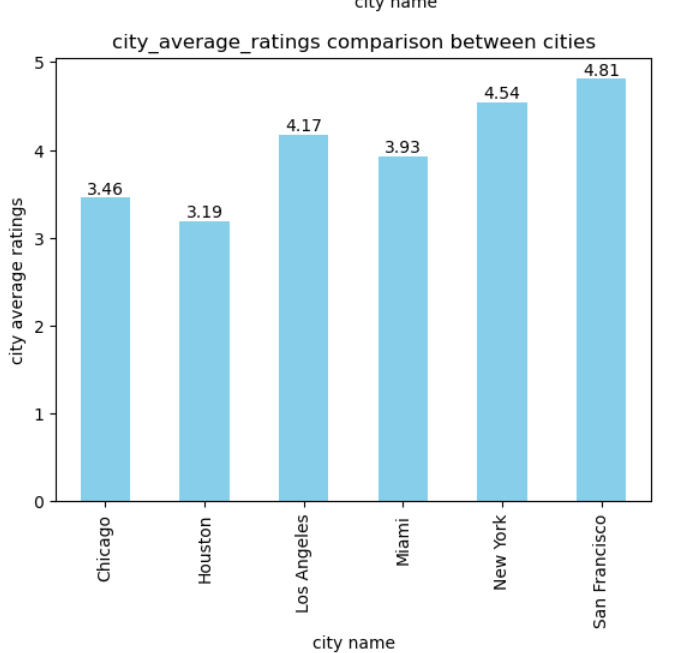
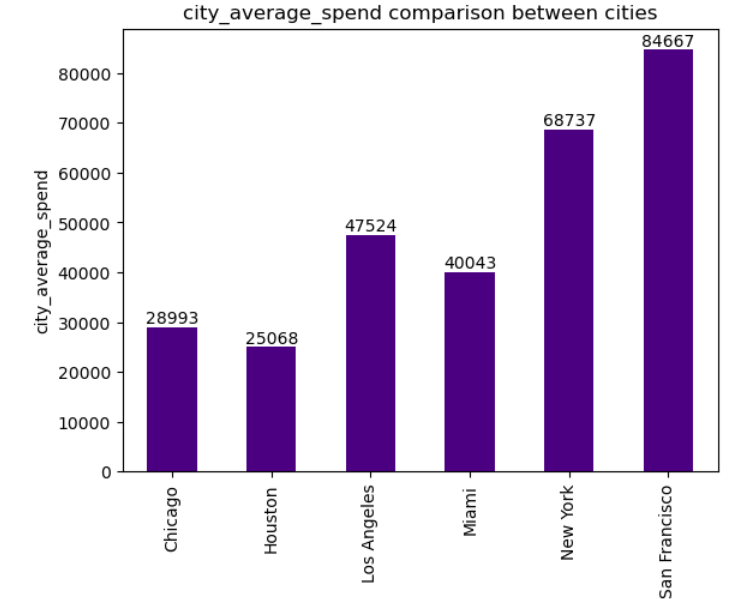
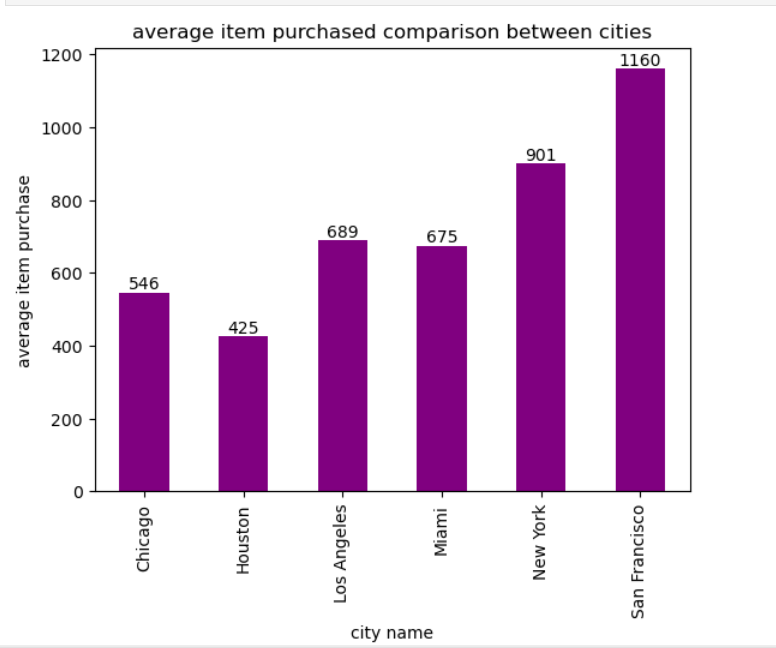
****

****

****

[This image shows the average rating given by male and female]

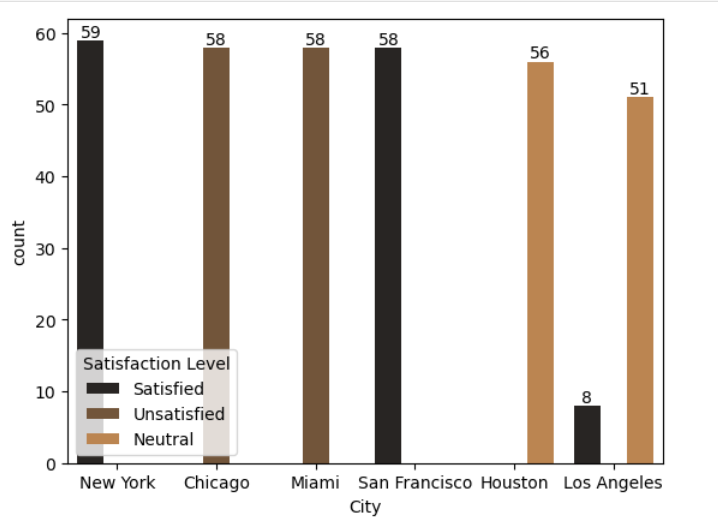
**4. males has more more silver membership and gold membership is almost equal between male and female (only diffrence of one) and females has more bronze membership than male**

**City-Based Analysis:**

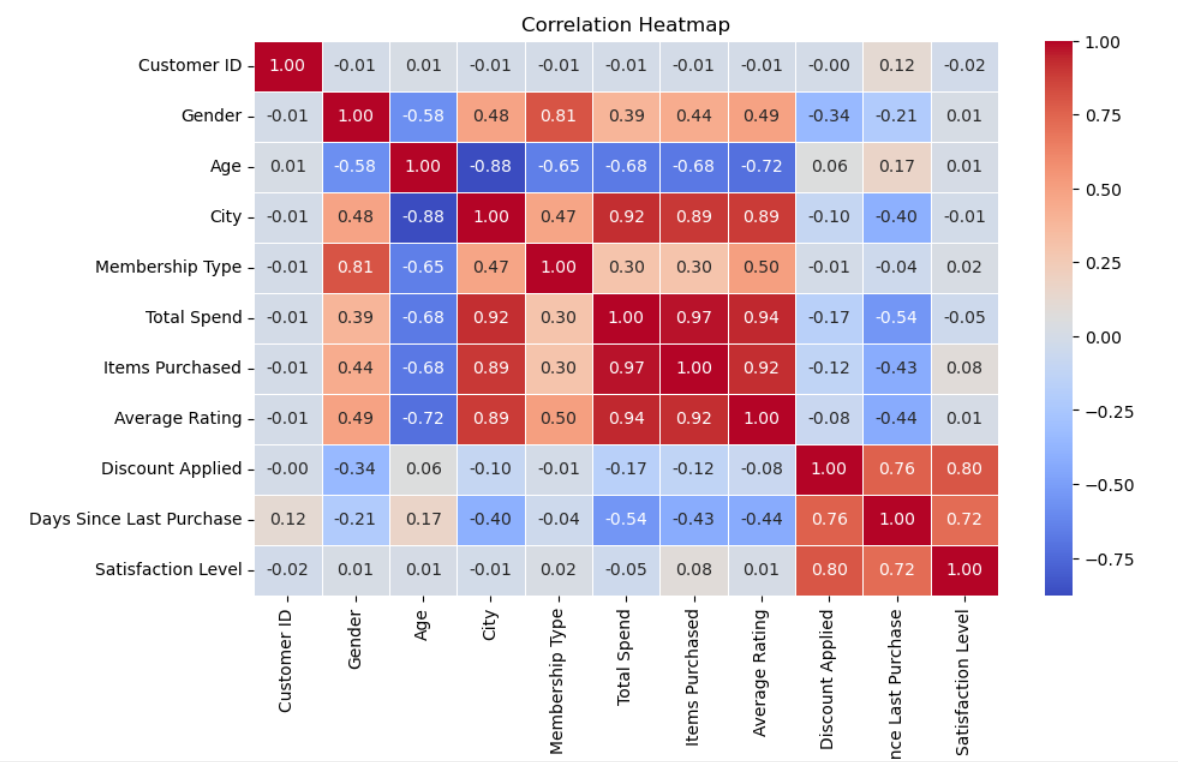
In city San Francisco maximum item sold

San Francisco city customer is spending more money while purchasing products while NewYork is doing great job too

AS expected, the maximum average ratings are from San Francisco and New York city while Los Angeles also have good average rating but with these visuals we can see that particular city Chicago and Houston there our e-commerce business is not doing well

****

**Correlation Analysis:**



* The heatmap shows a **strong positive correlation** between:
  + Discount Applied and Satisfaction Level (**0.80**) – Higher discounts lead to higher customer satisfaction.
  + Items Purchased and Total Spend (**0.97**) – More items purchased result in higher total spend.
  + Total Spend and Average Rating (**0.94**) – Customers spending more tend to give higher ratings.
* Found a **negative correlation** between:
  + Age and City (**-0.88**) – Younger customers are concentrated in certain cities.
  + Discount Applied and Total Spend (**-0.17**) – Higher discounts do not necessarily lead to higher spending.
* Satisfaction Level is moderately influenced by **Days Since Last Purchase** (**0.72**) – Recent purchases tend to result in higher satisfaction.

**Model Building**

* **Model:** Random Forest Classifier
* **Train-Test Split:** 80% training, 20% testing

**Model Evaluation**

**Model Performance:**  
The Random Forest model achieved an impressive accuracy of **98.57%** on the test data, demonstrating strong predictive capability.

**Classification Report:**  
The classification report shows high precision, recall, and f1-score across all classes, with a macro and weighted average f1-score of **0.99**, indicating balanced and effective model performance.

**Suggestions and Insights:**

* **Offer More Targeted Discounts:** Since higher discounts are clearly linked to higher satisfaction (**0.80**), running targeted promotions or personalized offers could help keep customers happy and coming back.
* **Encourage Repeat Purchases:** Customers who buy more frequently tend to be more satisfied (**0.72**). Sending friendly reminders, offering loyalty rewards, or setting up subscription models could boost repeat business.
* **Focus on High-Value Customers:** Since customers who spend more tend to leave higher ratings (**0.94**), it might be worth creating VIP programs or offering exclusive perks to keep these high-value customers engaged.
* **Personalize by City and Age:** The negative link between Age and City (**-0.88**) suggests that customer behaviour varies by location and age. Tailoring marketing campaigns based on these patterns could make them more effective.