Black Friday Sales

EDA AND MACHINE LEARNING MODEL

Problem Statement

A company wants to understand the customer purchase behaviour specifically, purchase amount against various products of different categories. They have shared purchase summary of various customers. The dataset also contains customer demographics like their age, gender, marital status, city_type, stay_in_current_city, product details (product_id and product category) and Total purchase_amount from last month.

They want to build a model to predict the purchase amount of customer against various products which will help them to create personalized offer for customers against different products.

Proposed Solution

1. Data Preprocessing:

- Load the dataset and perform initial data exploration to understand its structure and contents.
- Check for missing values, outliers, and anomalies that may require cleaning or imputation.

2. Descriptive Statistics:

- Calculate summary statistics (mean, median, standard deviation, etc.) to gain an overview.

3. Data Visualization:

- Generate histograms, box plots, and scatter plots to explore the distributions and relationships among different parameters using Python libraries like Matplotlib and Seaborn

4. Model Selection:

- Choose appropriate regression models for predicting purchase amounts. Common models include Linear Regression, Decision Trees, or Random Forest.

5. Model Training and Evaluation:

- Split the dataset into training and testing sets.
- Train the selected models on the training data.
- Evaluate models using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or R-squared.
- Fine-tune hyperparameters to optimize model performance.

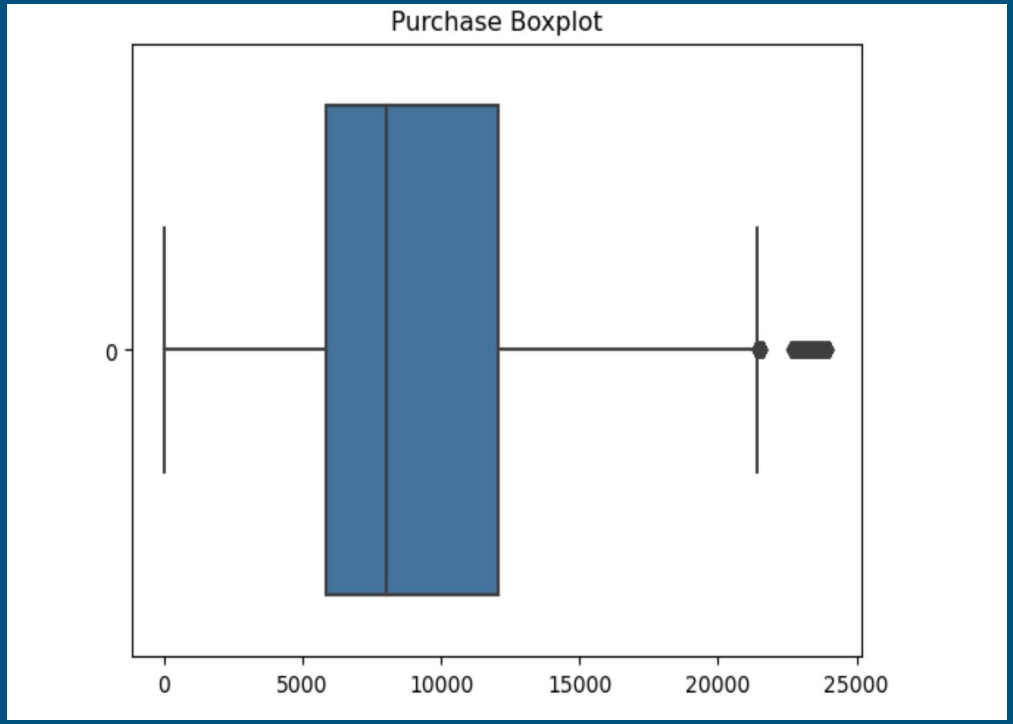
6. Model Interpretation:

- Understand the importance of features in the chosen model(s).
- Interpret coefficients or feature importances to gain insights into the factors influencing purchase amounts.

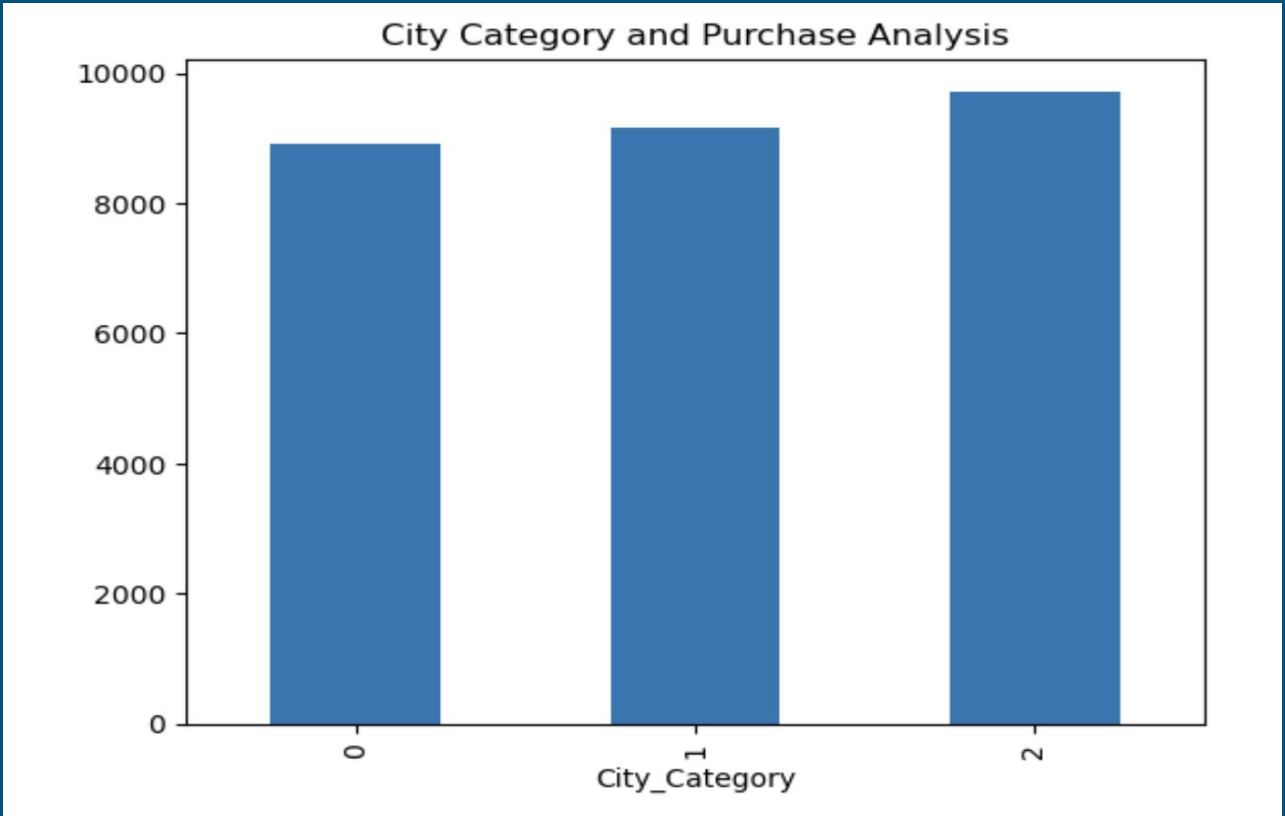
7. Personalized Offers:

- Utilize the trained model to predict purchase amounts for new customer transactions.
- Use these predictions to create personalized offers for customers based on their predicted purchase behavior.

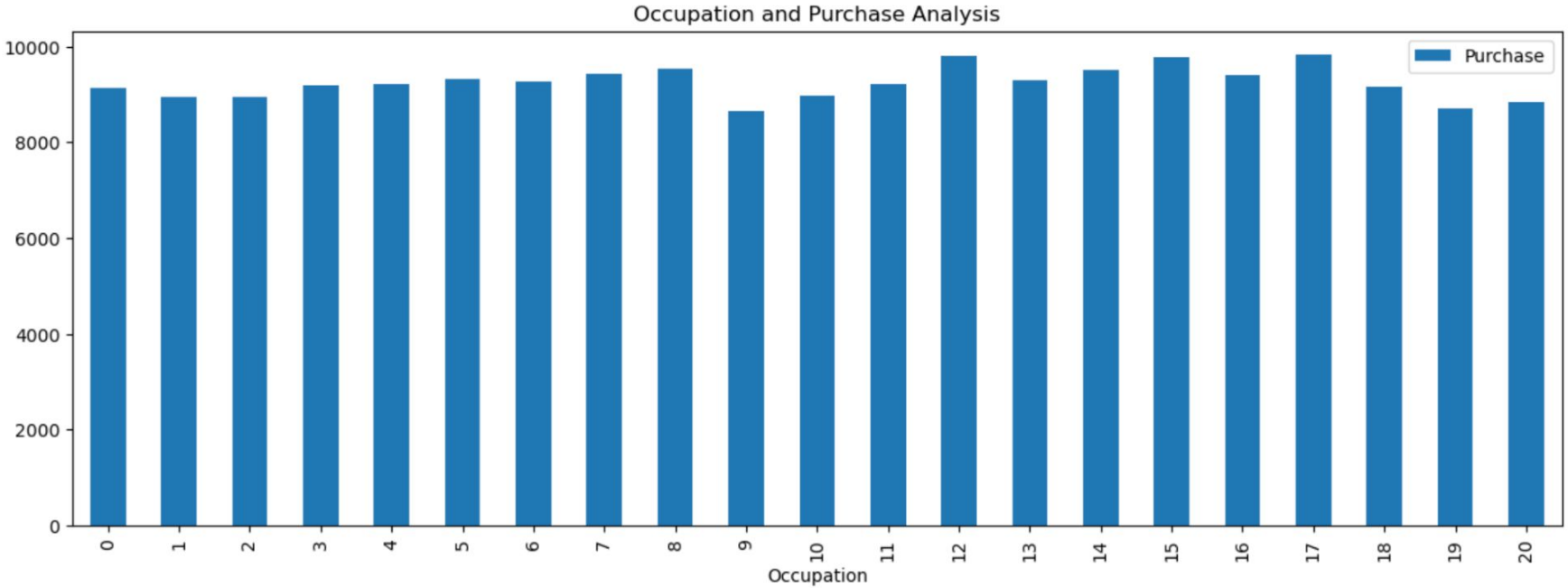
EDA



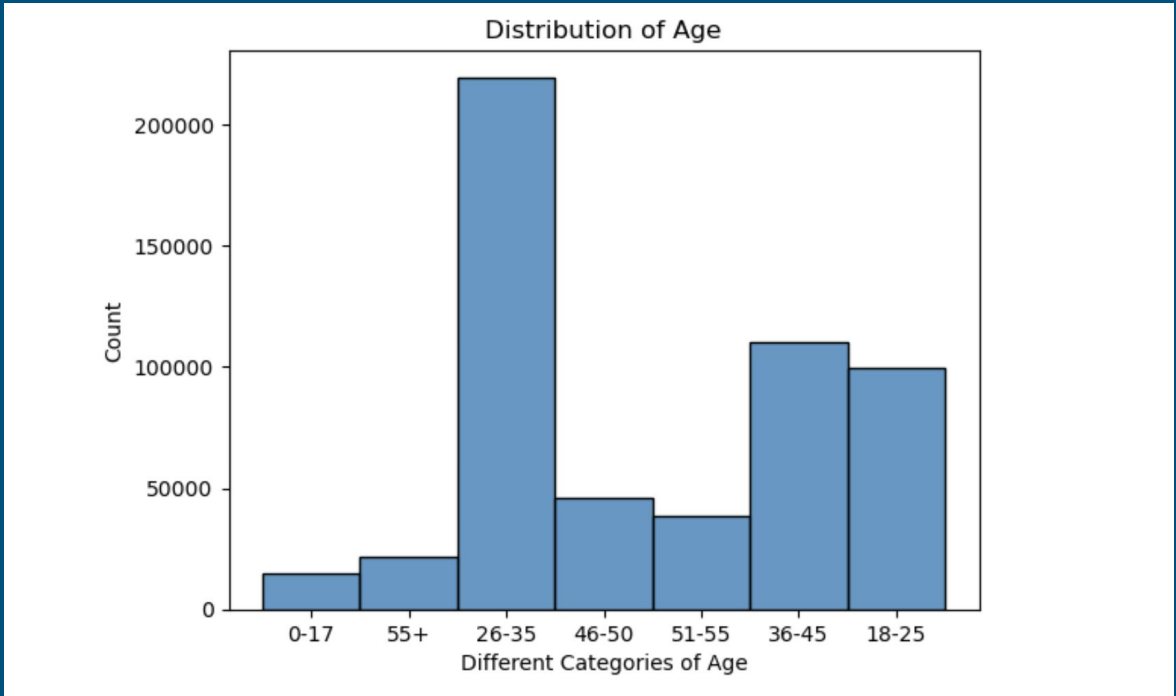
The target variable purchase_amount have outliers present.



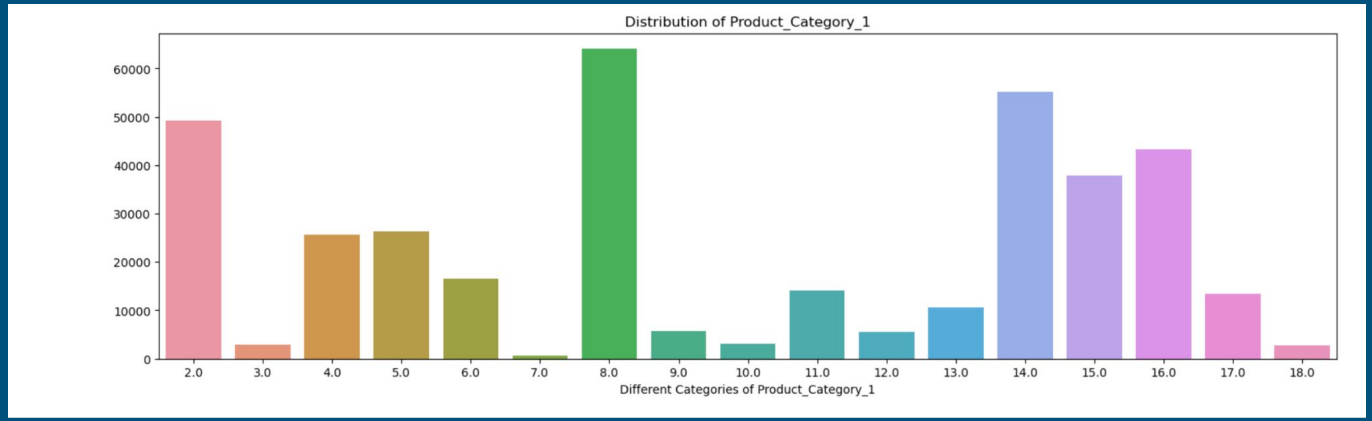
City Category 2 in the graph have more purchase amount than the other two cities.

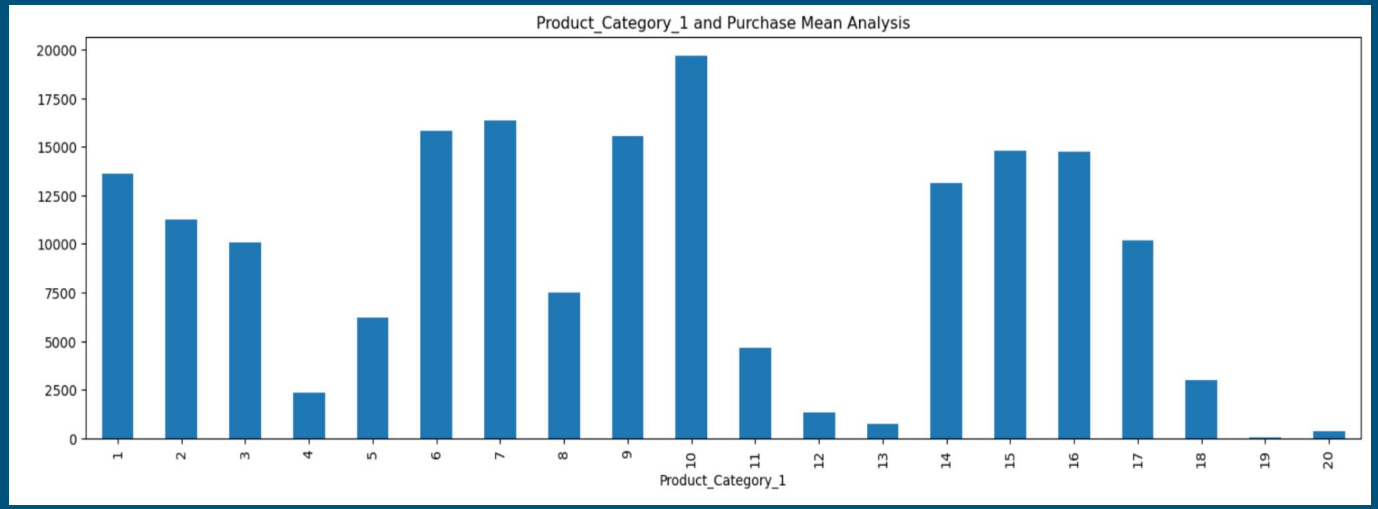


Based on Occupation we can tell that more or less each occupation has the same purchasing power.



Age 26-35 are purchasing more as compared to other age-groups.





Even though the product categories 2, 8, 9.84 shows highest in distribution categories it’s mean analysis is not the same.

Effect of Data Preprocessing

- Handling missing values : Null values can disrupt learning process and lead to inaccurate models
- Filling null values
- Converting Categorical columns into numerical columns

These are the preprocessing techniques used to handle the dataset.

In [31]:

Machine Learning Model

- We used Regression Algorithms for our dataset since the target variable was numerical.

- For applying ML Algorithms we split the dataset for training and testing purpose. Here test size data was 30% and remaining was training dataset.

- First we used, Linear Regression then Decision tree Regressor and in the end we used Random Forest Regressor.

Optimization

- We applied various ML Algorithms and checked the performance of each of them through evaluation metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), R-Squared Score (R2 Score), Root Mean Squared Error (RMSE).

- The model performance was evaluated by these metrics.

Conclusion

After performing ML Algorithms and evaluation metrics we can conclude that **Random Forest Regressor** is well suited for our dataset.

If our primary concern is accuracy and minimizing errors, we prefer Random forest regressor Model due to its lower MAE and RMSE.

If we want to prioritize explaining the variance in the target variable, Random forest regressor Model with the highest R2 score would be preferred.

Future Scope

Customer Experience Enhancement:

- Focus on improving the overall customer experience during Black Friday sales.
- Implement feedback mechanisms and incorporate customer preferences into the analysis.

Improved Targeted Marketing:

- Leverage insights from the analysis to enhance targeted marketing strategies. Refine customer segmentation for more targeted campaigns.
- Use personalized recommendations and promotions based on customer preferences.

Demand Forecasting:

- Utilize the ML models for predicting future sales demand.
- Explore the implementation of dynamic pricing based on customer behavior and market trends.