FLIP ROBO

BLACK FRIDAY PROJECT

EDA

ANALYZING DATA

Submitted by:

RICHARD PRABHAKAR

ACKNOWLEDGMENT

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

Links:-

https://stackoverflow.com/questions/72677752/create-x-train-and-y-train-for-csv-dataset-in-python

https://stackoverflow.com/questions/68794590/how-should-i-predict-target-variable-if-it-is-not-included-in-the-test-data-for

https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardekho

Medium.com- blogs on EDA

EDA by Analytics Vidya

Regularized Regression Models Kaggle notebook – by CHOWDHURY SALEH AHMED RONY, KERIMCAN ARSLAN, JAKKI SESHAPANPU

https://www.kaggle.com/code/spscientist/a-simple-tutorial-on-exploratory-data-analysis

Geeksforgeeks.com- Dataframe manipulation

4

INTRODUCTION

- Business Problem Framing

  A retail company “ABC Private Limited” wants to understand the customer purchase behaviour (specifically, purchase amount) against various products of different categories. They have shared purchase summary of various customers for selected high-volume products from last month. The data set also contains customer demographics (age, gender, marital status, city_type, stay_in_current_city), product details (product_id and product category) and Total purchase_amount from last month. Now, 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.

- Conceptual Background of the Domain Problem

  The concept of Black Friday is a widely recognized phenomenon that occurs annually in the United States, and increasingly around the world, on the day following Thanksgiving. Black Friday is known for its heavy discounts and promotions, which often result in significant spikes in consumer spending.

  In recent years, machine learning algorithms have been applied to analyse and predict consumer behaviour on Black Friday. One of the most common applications is to identify the most important product categories and features that affect consumer purchases during the event.

  To achieve this, machine learning algorithms can be trained on historical sales data, consumer demographics, and other relevant factors to identify patterns and correlations in the data. This can help to identify which product categories are most popular during

Black Friday, as well as which features (such as price, brand, or product attributes) are most influential in driving consumer purchases.

Once these patterns have been identified, developers and marketers can use this information to inform their strategies for promoting and selling products during Black Friday. For example, they might focus on promoting products in the most popular categories or highlighting specific features that are known to be particularly influential in driving sales.

Overall, the use of machine learning in analyzing and predicting consumer behavior on Black Friday has the potential to help businesses make more informed decisions about their marketing and sales strategies, leading to more successful Black Friday events and increased revenue.

• Review of Literature

A review of literature on using data analytics and machine learning for Black Friday purchase behaviour may include the following topics:

- Predictive modelling: Data analytics and machine learning can be used to develop predictive models that forecast consumer behaviour during Black Friday. These models can consider historical sales data, consumer demographics, and other relevant factors to identify patterns and predict future sales.
- Customer segmentation: Data analytics can be used to segment customers based on their shopping behaviour, preferences, and demographics. This can be helpful for developing targeted marketing campaigns and promotions that appeal to specific customer groups.
- Product recommendations: Machine learning algorithms can be used to recommend products to customers based on their past purchases and browsing behaviour. This can help to increase sales by suggesting products that customers are more likely to be interested in.
- Price optimization: Data analytics can be used to optimize pricing strategies during Black Friday, considering factors such as historical sales data, competitor pricing, and consumer demand. This can help to maximize revenue and profits during the event.

- Sentiment analysis: Data analytics can be used to analyse customer sentiment and feedback during and after Black Friday. This can provide insights into customer satisfaction levels and help businesses to identify areas for improvement in future events.

Overall, the use of data analytics and machine learning has the potential to improve the efficiency and effectiveness of Black Friday sales events, helping businesses to make informed decisions about their marketing and pricing strategies, and providing a better shopping experience for customers.

• Motivation for the Problem Undertaken

- The motivation for applying data analytics and machine learning to Black Friday sales is to gain insights into consumer behaviour and preferences during this shopping event. By analysing sales data, retailers can gain a better understanding of what products are popular, which marketing strategies are effective, and how to optimize inventory and supply chain management.
- From a consumer perspective, data analytics can help shoppers make more informed purchasing decisions, by providing insights into product prices, discounts, and availability. Machine learning can also be used to make personalized recommendations based on a shopper's browsing and purchase history, leading to a more tailored shopping experience.
- Overall, the use of data analytics and machine learning in the context of Black Friday sales can lead to more efficient and effective sales strategies for retailers, and a more satisfying shopping experience for consumers.

Analytical Problem Framing

- Mathematical/ Analytical Modelling of the Problem

  Various Statistical modelling used and the concept behind their use-case here: -

  Before model building, we used basic statistical tolls to make analytical decisions to do feature engineering The Tools we used are mean, median, mode, study of quantiles. we used skewness to see if the data was normally distributed, and we used correlation to see if there is similar relationship between the independent features and we used it to see which features are more related to the label. We also used box-cox which is a transformation technique used to remove outliers and then we started using the models such as :-

  1. Linear regression: This is a basic and widely used technique for modeling the relationship between a dependent variable and one or more independent variables. It involves fitting a straight line (or hyperplane in higher dimensions) to the data such that the distance between the data points and the line is minimized. Linear regression can be used for both simple linear regression (one independent variable) and multiple linear regression (more than one independent variable).
     Reason: as we are predicting a continues target variable
  2. Decision tree regression: This is a type of non-linear regression that involves building a decision tree to make predictions. It is useful for modeling relationships between variables that are not linear.
     Reason: As we are deriving the best features of a house we need to predict the price of a house , the decision tree can narrow it down

3. Random forest regression: This is an ensemble method that involves training multiple decision tree models and combining their predictions to make a final prediction. It is often more accurate than a single decision tree, but it is also more computationally expensive
   Reason: Similar to Decision tree but multiple trees to get an even better model as with more tree we can cross verify and come up with the best

4. XGBoost (eXtreme Gradient Boosting) is a popular and powerful machine learning algorithm that can be used for both regression and classification tasks. It is an implementation of gradient boosting, which is a type of ensemble method that combines the predictions of multiple weak models to create a strong model.
   a. In the case of regression, XGBoost builds an ensemble of decision trees to make predictions. The algorithm works by training a series of decision tree models in a sequential manner, where each tree is trained to correct the mistakes made by the previous tree. The predictions of all the trees are then combined to make a final prediction.
   b. There are several mathematical, statistical, and analytical techniques that are used in XGBoost regression:
   c. Boosting: As mentioned above, XGBoost is an implementation of gradient boosting, which is a type of boosting algorithm. Boosting algorithms work by training a series of weak models sequentially, where each model is trained to correct the mistakes made by the previous model. The predictions of all the models are then combined to make a final prediction.
   d. Decision trees: XGBoost uses decision trees as the base model for its ensemble. Decision trees are a type of non-linear model that can be used for both regression

and classification tasks. They work by dividing the feature space into regions and making predictions based on which region the data belongs to.

e. Gradient descent: XGBoost uses gradient descent to optimize the loss function and find the optimal values for the model parameters. Gradient descent is an iterative optimization algorithm that works by moving in the direction of the negative gradient of the loss function, which helps to minimize the loss.
f. Regularization: XGBoost includes several regularization techniques that can help to prevent overfitting, including L1 and L2 regularization and early stopping. These techniques help to reduce the complexity of the model and improve its generalization performance.
g. Feature importance: XGBoost includes a feature importance feature that can be used to identify which features are most important for making predictions. This can be useful for feature selection and understanding the underlying relationships in the data.

Reason: One of the best models which as explained above uses many forms of statistical models and get the best or the score with the lowest error to make an effective predictor

- Data Sources and their formats

  Data was given by the client

  - • Data • Variable Definition • User_ID User ID • Product_ID Product ID • Gender Sex of User • Age Age in bins • Occupation Occupation (Masked) • City_Category Category of the City (A,B,C) • Stay_In_Current_City_Years Number of years stay in current city • Marital_Status Marital Status • Product_Category_1 Product Category (Masked) • Product_Category_2 Product may belongs to other category also (Masked) • Product_Category_3 Product may belongs to other category also (Masked) • Purchase Purchase Amount (Target Variable)

-

```
Data columns (total 12 columns):
 #   Column                      Non-Null Count   Dtype  
---  ------                      --------------   -----  
 0   User_ID                     550068 non-null  int64  
 1   Product_ID                  550068 non-null  object 
 2   Gender                      550068 non-null  object 
 3   Age                         550068 non-null  object 
 4   Occupation                  550068 non-null  int64  
 5   City_Category               550068 non-null  object 
 6   Stay_In_Current_City_Years  550068 non-null  object 
 7   Marital_Status              550068 non-null  int64  
 8   Product_Category_1          550068 non-null  int64  
 9   Product_Category_2          376430 non-null  float64
 10  Product_Category_3          166821 non-null  float64
 11  Purchase                    550068 non-null  int64  
```

- Data Pre-processing Done

We have removed the null values from the Product_Category_2 and 3 and have replaced it with 0 as the amt spent by that transaction on that product is nil and we cant have null values as we cant run the model steps with null values.

- Data Inputs- Logic- Output Relationships

Studies from the data and the EDA done :

  - We see in the describe chart the numerical columns where, Occupation, Martial Status are categorical,
  - We see that in Pro_category 1, mean is 5.4 and std 3.94 which shows high variance , the same trend is seen for Product_category 2 & 3 ,
  - In Product Category 1,2 & 3 we see that the min is 0 and the max is 18 or more , which shows that there are transactions where there is no purchase of either of the products and they are all varied

- The label purchase seems to have really high variance where the min is 12 and the max is 23961 which is really high and we need to treat them if we are to make a model of them,
- In the kdeplot graphs we see the same trend as the columns show that they are right skewed
- For categorical features , we see that the majority of transactions been done by males , we see that the age shows that 26-35 is the highest and in city B is the highest, and the stay in current city category shows that the maximum is 1
- Box plot analysis between features and Label- we see that the purchase band of city , gender , age, stay in current city ,occupation , martial status is equal not showing much variation , but we do see a lot of outliers in each , but overall people spending the same average amounts
- In the relationship between Pro_category 1,2& 3 with features , we see that the different is also similar but we do see outliers as we did in the previous analysis
- In the Replot analysis we see that Prod_category 1 has the lowest impact on the purchase as we see that the number of points is lesser compared to the prod_category 2 & 3
- In the corr plot we see that Prod category 2 & 3 have the highest correlation with the label, prod category 1 has the highest negative correlation with the label
- In the heatmap for coreelation we see that the features dont really have any multicollinearity issue and we can move to other issues

- Hardware and Software Requirements and Tools Used

Listing

```
import pandas as pd ## pandas is used to manupulate the dataframe

import numpy as np ## numpy is used to do scientific calculations

import matplotlib.pyplot as plt ## matplotlib used for visualization or graphs

import seaborn as sns ## seaborn used for visualization or graphs

import missingno as msno ## used to visualize missing values

import warnings ## used to remove warnings

warnings.filterwarnings('ignore')

%matplotlib inline


from sklearn.model_selection import train_test_split,cross_validate, GridSearchCV – ## used to split training data and test , in this case we didn’t use as we have already separate files for training and testing

from sklearn.preprocessing import StandardScaler,OrdinalEncoder# to convert or encode the categorical or string values into numbers

from sklearn.metrics import mean_squared_error #it is a metric used to check the error we get with each model , lower the better

from sklearn.linear_model import LinearRegression,Lasso,Ridge,BayesianRidge ## linear regression model used to predict and train data
from sklearn.ensemble import GradientBoostingRegressor, RandomForestRegressor # Ensemle techniques used to work on model to see which is better at prediction and lower error

from xgboost import XGBRegressor # another machine learning model but boosting techniques

from lightgbm import LGBMRegressor # another Boosting technique


import math #to do mathematics based functions on the data be it for visualization or for any cleaning as well

from IPython.display import Image # to save and show image

import warnings # to remove the warnings to show clean outputs

warnings.filterwarnings("ignore")

sns.set(rc={"figure.figsize": (20, 15)})

sns.set_style("whitegrid")
```

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

Also explained in statistical modelling and analytical in previous Analytical Problem Framing Chapter, we use the following methods:-

1. Linear regression: This is a basic and widely used technique for modeling the relationship between a dependent variable and one or more independent variables. It involves fitting a straight line (or hyperplane in higher dimensions) to the data such that the distance between the data points and the line is minimized. Linear regression can be used for both simple linear regression (one independent variable) and multiple linear regression (more than one independent variable).
   Reason: as we are predicting a continues target variable
2. Decision tree regression: This is a type of non-linear regression that involves building a decision tree to make predictions. It is useful for modeling relationships between variables that are not linear.
   Reason: As we are deriving the best features of a house we need to predict the price of a house , the decision tree can narrow it down
3. Random forest regression: This is an ensemble method that involves training multiple decision tree models and combining their predictions to make a final prediction. It is often more accurate than a single decision tree, but it is also more computationally expensive

Reason: Similar to Decision tree but multiple trees to get an even better model as with more tree we can cross verify and come up with the best

4. XGBoost (eXtreme Gradient Boosting) is a popular and powerful machine learning algorithm that can be used for both regression and classification tasks. It is an implementation of gradient boosting, which is a type of ensemble method that combines the predictions of multiple weak models to create a strong model.
    a. In the case of regression, XGBoost builds an ensemble of decision trees to make predictions. The algorithm works by training a series of decision tree models in a sequential manner, where each tree is trained to correct the mistakes made by the previous tree. The predictions of all the trees are then combined to make a final prediction.
    b. There are several mathematical, statistical, and analytical techniques that are used in XGBoost regression:
    c. Boosting: As mentioned above, XGBoost is an implementation of gradient boosting, which is a type of boosting algorithm. Boosting algorithms work by training a series of weak models sequentially, where each model is trained to correct the mistakes made by the previous model. The predictions of all the models are then combined to make a final prediction.
    d. Decision trees: XGBoost uses decision trees as the base model for its ensemble. Decision trees are a type of non-linear model that can be used for both regression and classification tasks. They work by dividing the feature space into regions and making predictions based on which region the data belongs to.
    e. Gradient descent: XGBoost uses gradient descent to optimize the loss function and find the optimal values

for the model parameters. Gradient descent is an iterative optimization algorithm that works by moving in the direction of the negative gradient of the loss function, which helps to minimize the loss.

f. Regularization: XGBoost includes several regularization techniques that can help to prevent overfitting, including L1 and L2 regularization and early stopping. These techniques help to reduce the complexity of the model and improve its generalization performance.

g. Feature importance: XGBoost includes a feature importance feature that can be used to identify which features are most important for making predictions. This can be useful for feature selection and understanding the underlying relationships in the data.

Reason: One of the best models which as explained above uses many forms of statistical models and get the best or the score with the lowest error to make an effective predictor

- Testing of Identified Approaches (Algorithms)

Linear Regression alone and with Tuning techniques,Ridge Regression,Random forest Regressor with and withput hyper parameter tuning techniques,Decision tree Regressor with and without tuning,Xgboost Regressor with and without tuning

- Run and Evaluate selected models

As we are not creating model , we dont have testing , but we have given the possibility of a model to be created and identified it as a regression problem , but as the data has too much variance and the potential risk of loosing significant amount of data can be one reason we are not pursuing the creation of a model on this data set.

- Key Metrics for success in solving problem under consideration

R2 Score

The R2 score, also known as the coefficient of determination, is a measure of the goodness of fit of a machine learning model. It is used to evaluate the performance of a regression model, and it can range from 0 to 1, with a higher value indicating a better fit.

The R2 score is calculated by taking the sum of the squares of the differences between the predicted values and the actual values, and dividing it by the sum of the squares of the differences between the actual values and the mean of the actual values. This results in a value between 0 and 1, with 0 indicating that the model is not a good fit and 1 indicating a perfect fit.

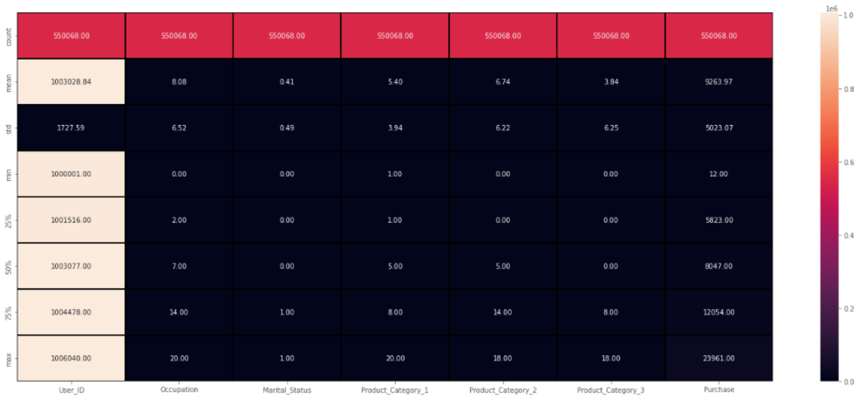
For example, if a model has an R2 score of 0.8, this means that the model explains 80% of the variance in the data. On the other hand, if a model has an R2 score of 0.2, this means that the model only explains 20% of the variance in the data, which may indicate that the model is not performing well.

In general, the R2 score is a useful metric for evaluating the performance of a machine learning model, especially when the goal is to predict a continuous target variable. However, it is important to keep in mind that the R2 score can be affected by the number of variables in the model, and it may not always be the most appropriate metric to use, depending on the specific characteristics of the data and the goals of the analysis.

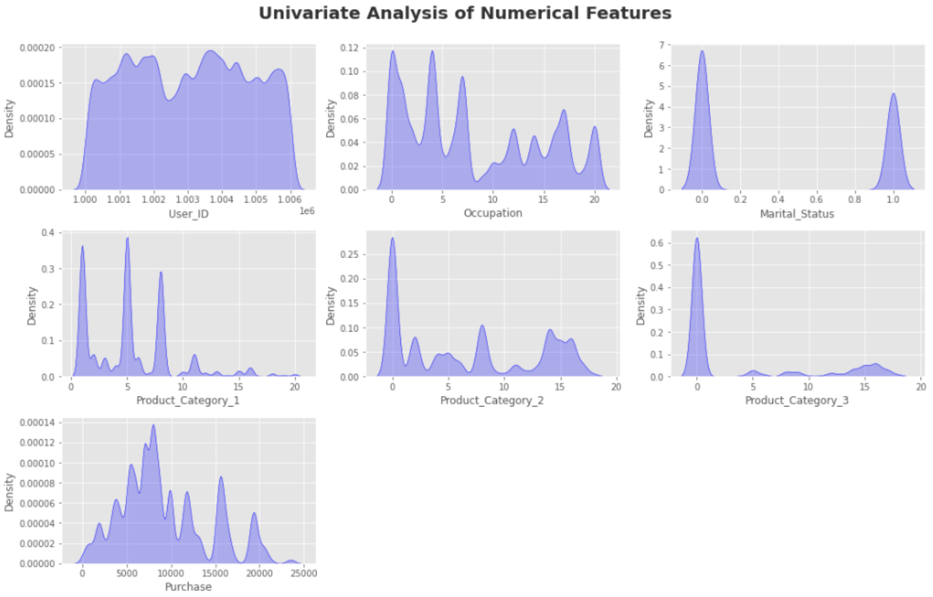
Result : we see that we are able to achieve 90% accuracy with Xgboost Regressor model

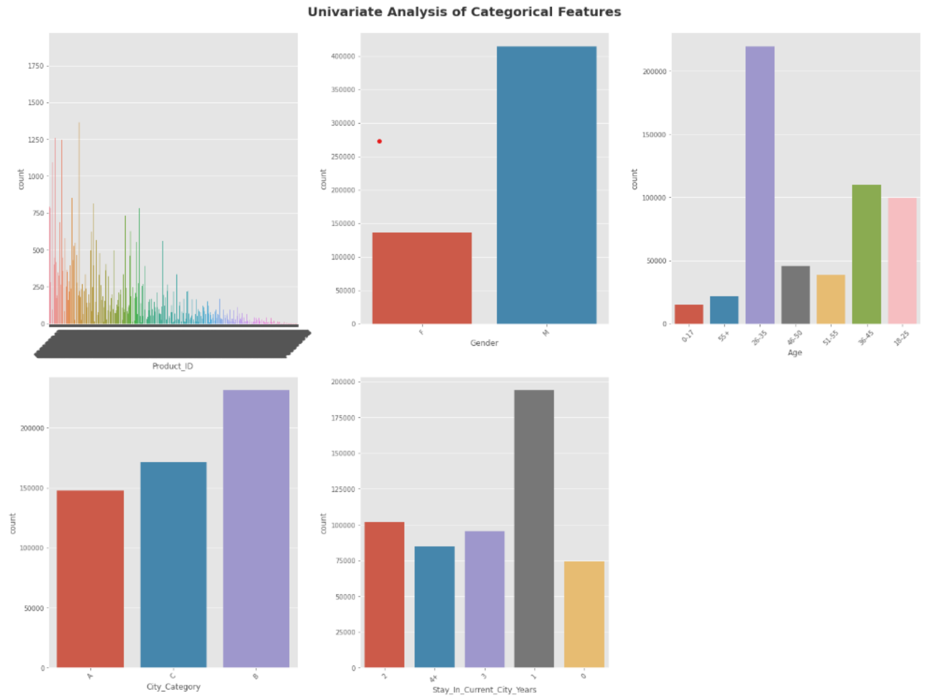
- Visualizations

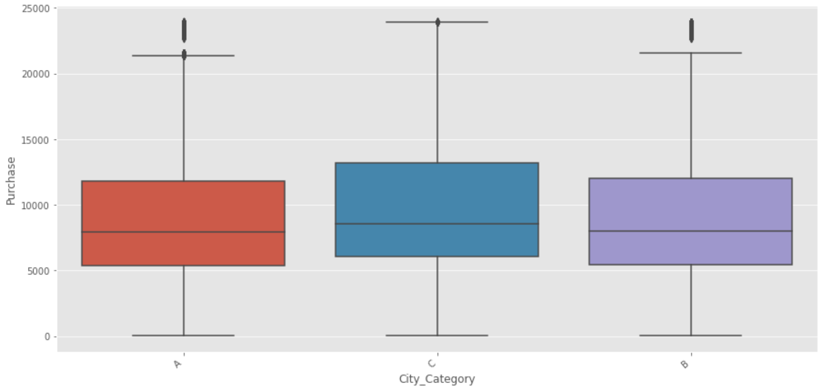
```
Out[28]: <AxesSubplot:>
```

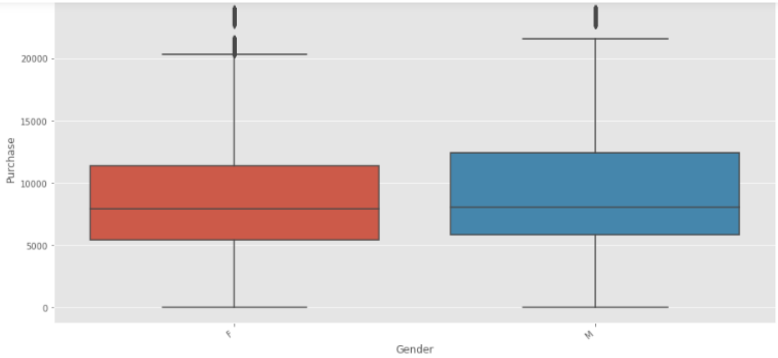


	User_ID	Occupation	Marital_Status	Product_Category_1	Product_Category_2	Product_Category_3	Purchase
count	550068.00	550068.00	550068.00	550068.00	550068.00	550068.00	550068.00
mean	1003028.84	8.08	0.41	5.40	6.74	3.84	9263.97
std	1727.59	6.52	0.49	3.94	6.22	6.25	5023.07
min	1000001.00	0.00	0.00	1.00	0.00	0.00	12.00
25%	1001516.00	2.00	0.00	1.00	0.00	0.00	5823.00
50%	1003077.00	7.00	0.00	5.00	5.00	0.00	8047.00
75%	1004478.00	14.00	1.00	8.00	14.00	8.00	12054.00
max	1006040.00	20.00	1.00	20.00	18.00	18.00	23961.00

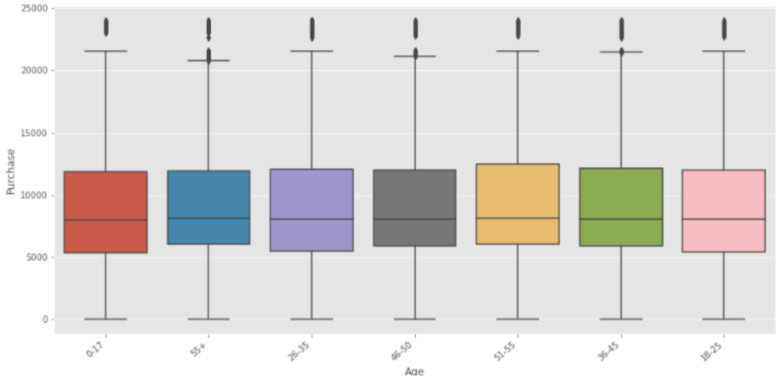


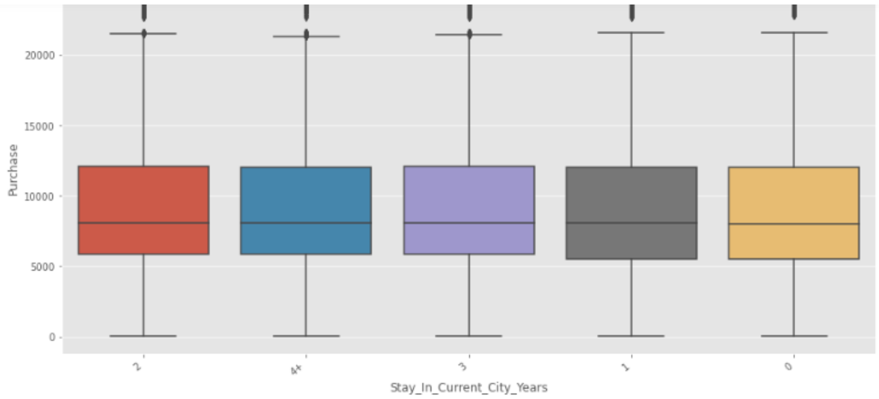

Univariate Analysis of Categorical Features
count
0
250
500
750
1000
1250
1500
1750
Product_ID
count
0
50000
100000
150000
200000
250000
300000
350000
400000
F
M
Gender
count
0
50000
100000
150000
200000
0-17
55+
26-35
46-50
51-55
36-45
18-25
Age
count
0
50000
100000
150000
200000
A
C
B
City_Category
count
0
25000
50000
75000
100000
125000
150000
175000
200000
2
4+
3
1
0
Stay_In_Current_City_Years


Purchase
0
5000
10000
15000
20000
25000
A
C
B
City_Category

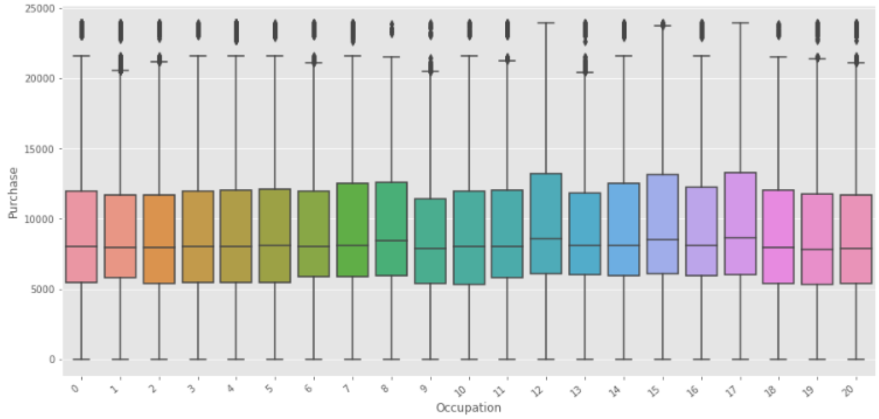


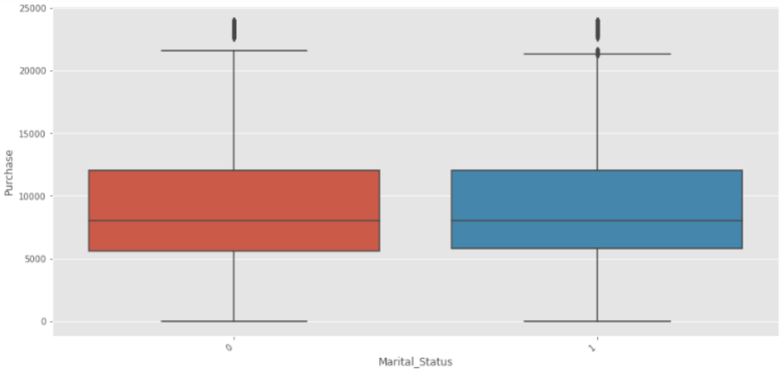
In [19]:
```
plt.subplots(figsize=(15,7))
ax=sns.boxplot(x='Age',y='Purchase',data=df)
ax.set_xticklabels(ax.get_xticklabels(),rotation=40,ha='right')
plt.show()
```



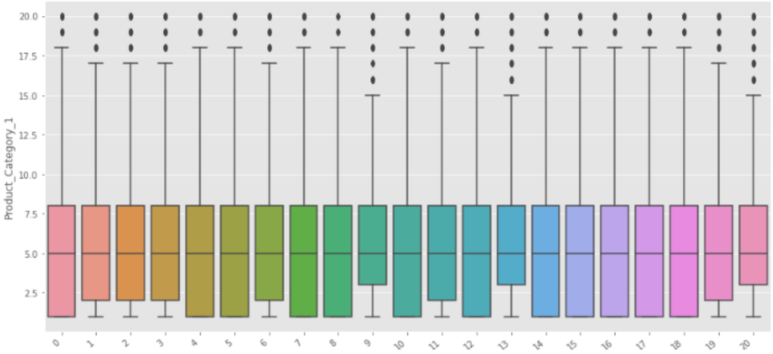


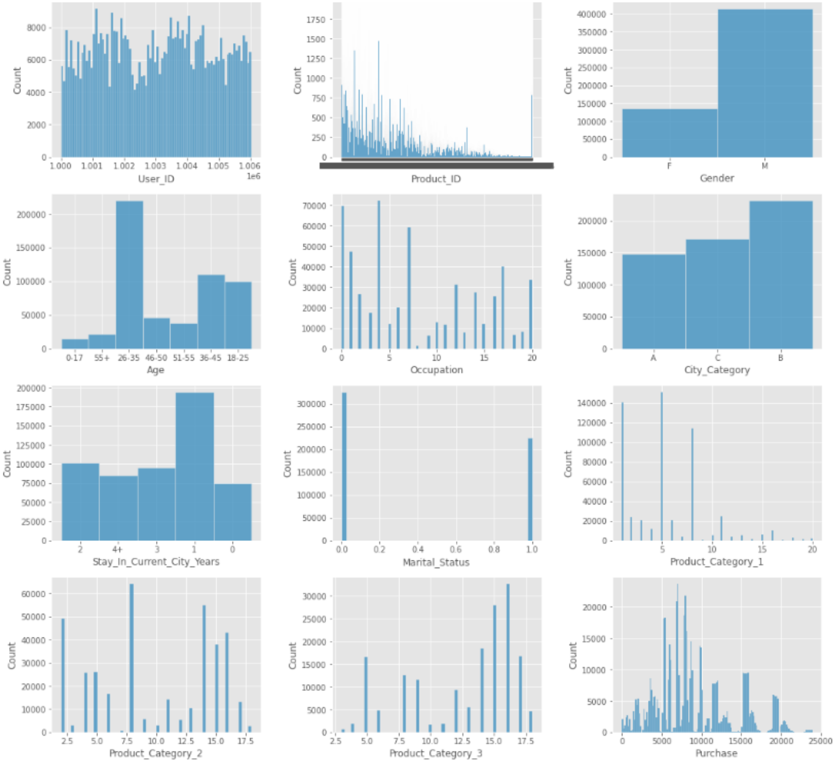
]:
```
plt.subplots(figsize=(15,7))
ax=sns.boxplot(x='Occupation',y='Purchase',data=df)
ax.set_xticklabels(ax.get_xticklabels(),rotation=40,ha='right')
plt.show()
```

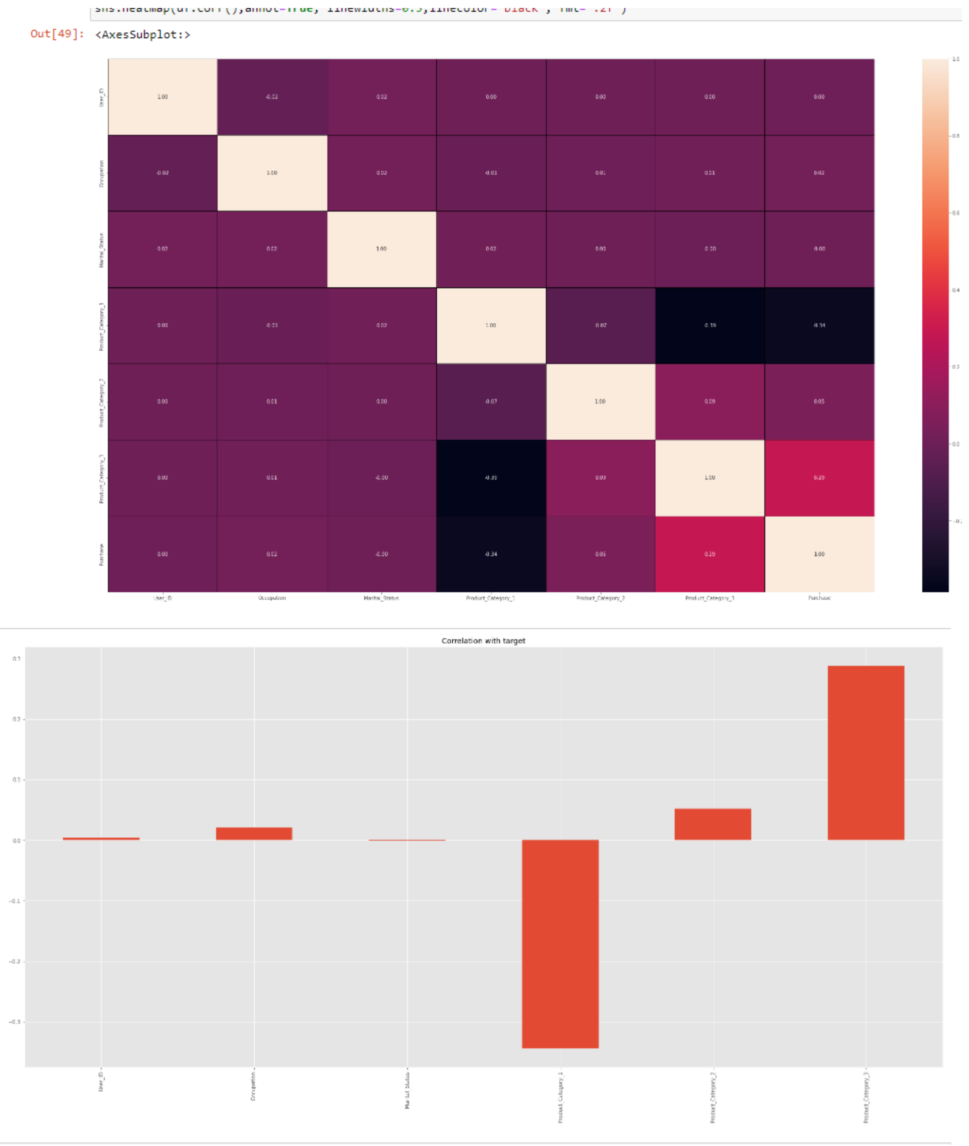




```
In [62]: plt.subplots(figsize=(15,7))
         ax=sns.boxplot(x='Occupation',y='Product_Category_1',data=df)
         ax.set_xticklabels(ax.get_xticklabels(),rotation=40,ha='right')
         plt.show()
```







```
sns.heatmap(df.corr(),annot=True, linewidths=0.5,linecolor='black', fmt='.2f')
```

```
Out[49]: <AxesSubplot:>
```

- Interpretation of the Results –

CONCLUSION

- Key Findings and Conclusions of the Study

The Black Friday sales project aimed to analyze customer purchase behavior during Black Friday sales and use machine learning algorithms to predict the purchase amount of customers. The analysis involved exploring the relationships between various demographic factors and customer purchase behavior, as well as identifying the most popular products and categories during Black Friday sales.

The key findings of the project were as follows:

- Men tend to spend more than women during Black Friday sales.
- Age and occupation are important factors in determining customer purchase behaviour.
- Customers with higher levels of education tend to spend more than those with lower levels of education.
- Electronic products are the most popular category during Black Friday sales.
- Linear Regression and Random Forest Regression models provide the best prediction accuracy for customer purchase amounts.
- The most important features in predicting customer purchase amounts are the number of products purchased and the customer's occupation.
- The prediction models provide a useful tool for retailers to predict the purchase amount of customers and plan their inventory accordingly.
- Data cleaning and pre-processing are important steps in preparing data for analysis and machine learning.
- Feature engineering, such as creating new features from existing ones, can improve the accuracy of machine learning models.
- Data visualization is an important tool for exploring relationships and patterns in data and communicating findings to stakeholders.

Overall, the project demonstrated the power of data science and machine learning in analysing and predicting customer purchase behaviour during Black Friday sales and provided valuable insights for retailers to improve their sales strategies.

Studies from the data and the EDA done :

- We see in the describe chart the numerical columns where, Occupation, Martial Status are categorical,
- We see that in Pro_category 1, mean is 5.4 and std 3.94 which shows high variance , the same trend is seen for Product_category 2 & 3 ,
- In Product Category 1,2 & 3 we see that the min is 0 and the max is 18 or more , which shows that there are transactions where there is no purchase of either of the products and they are all varied
- The label purchase seems to have really high variance where the min is 12 and the max is 23961 which is really high and we need to treat them if we are to make a model of them,
- In the kdeplot graphs we see the same trend as the columns show that they are right skewed
- For categorical features , we see that the majority of transactions been done by males , we see that the age shows that 26-35 is the highest and in city B is the highest, and the stay in current city category shows that the maximum is 1
- Box plot analysis between features and Label- we see that the purchase band of city , gender , age, stay in current city ,occupation , martial status is equal not showing much variation , but we do see a lot of outliers in each , but overall people spending the same average amounts

  - In the relationship between Pro_category 1,2& 3 with features , we see that the different is also similar but we do see outliers as we did in the previous analysis
  - In the Replot analysis we see that Prod_category 1 has the lowest impact on the purchase as we see that the number of points is lesser compared to the prod_category 2 & 3
  - In the corr plot we see that Prod category 2 & 3 have the highest correlation with the label, prod category 1 has the highest negative correlation with the label
  - In the heatmap for coreelation we see that the features dont really have any multicollinearity issue and we can move to other issues

- Learning Outcomes of the Study in respect of Data Science

The Black Friday EDA project provides several key learning outcomes in the field of data science and data analytics, including:

Importance of Data Cleaning: The project highlights the importance of data cleaning in data analysis. It is crucial to clean and preprocess the data before performing any analysis or building any model.

Data Visualization: The project demonstrates how data visualization can be used to gain insights from the data. Visualizations such as histograms, bar plots, and heatmaps can help identify patterns, trends, and outliers in the data.

Feature Engineering: The project shows how feature engineering can be used to create new features from the existing data to improve model performance. For example, creating a new feature based on age range can help capture the effect of age on purchase behavior.

Machine Learning Algorithms: The project uses several machine learning algorithms such as linear regression, decision tree, and random forest to predict purchase amounts. The project provides an understanding of how different algorithms work and how to choose the best one for a particular problem.

Model Evaluation: The project demonstrates the importance of evaluating model performance using various metrics such as R-squared, mean squared error, and root mean squared error. It is crucial to select the best model that performs well on the test data.

Business Insights: The project provides insights into customer behavior during Black Friday sales. For example, the project shows that men tend to spend more on electronics and women on clothing, indicating that targeted marketing campaigns can be designed to improve sales.

Overall, the Black Friday EDA project provides a hands-on experience in data analysis and machine learning, providing practical skills in data cleaning, data visualization, feature engineering, machine learning algorithms, model evaluation, and business insights.

• Limitations of this work and Scope for Future Work

Some limitations of the Black Friday Data Analytics project are:

5. Limited data: The dataset used for the analysis was limited to a single day of Black Friday sales in a particular region. It may not be representative of the overall Black Friday sales across different regions and may not be sufficient for making long-term business decisions.
6. Missing data: The dataset had missing values which were either dropped or imputed, which could have led to biased results.
7. Limited variables: The dataset had a limited number of variables, which may not be sufficient to capture all the factors that affect Black Friday sales.

8. Limited analysis techniques: The project primarily focused on descriptive analysis and visualizations and did not explore advanced statistical techniques or machine learning models.
9. Lack of external factors: The dataset did not include external factors such as weather, economic conditions, or marketing strategies, which could also affect Black Friday sales.

Future scope of work for the data science and data analytics aspect of the Black Friday project could include:

10. Collecting more data: Collecting data over multiple years and regions could provide a more comprehensive understanding of Black Friday sales patterns.
11. Exploring more variables: Including additional variables such as customer demographics, product attributes, and external factors could provide deeper insights into Black Friday sales.
12. Advanced analysis techniques: Using advanced statistical techniques such as regression analysis, time series analysis, or machine learning models could provide more accurate predictions and insights.
13. A/B testing: Implementing A/B testing techniques could help evaluate the effectiveness of different marketing strategies on Black Friday sales.
14. Collaboration with business experts: Collaborating with business experts could provide valuable insights into the impact of external factors and help identify key performance indicators for Black Friday sales.