CHURN PREDICTION

Outline

- Introduction
- Methodology
- Results
- Conclusion

Introduction

- In this assignment I have performed churn prediction using the customer dataset provided.
- The dataset consisted of customer data like their name, location, monthly bill, and the information about whether the customer churn or not.
- Using the data and Jupyter Lab I prepared and tested two models on the given data and based on evaluation metrices decided to go with Logistic regression model.
- I even simulated its deployment and tested its prediction on sample data.

Methodology

Executive Summary

•Data Loading:

- Data was provided in form of excel file which I loaded on my Jupyter Lab.

•Performed data wrangling

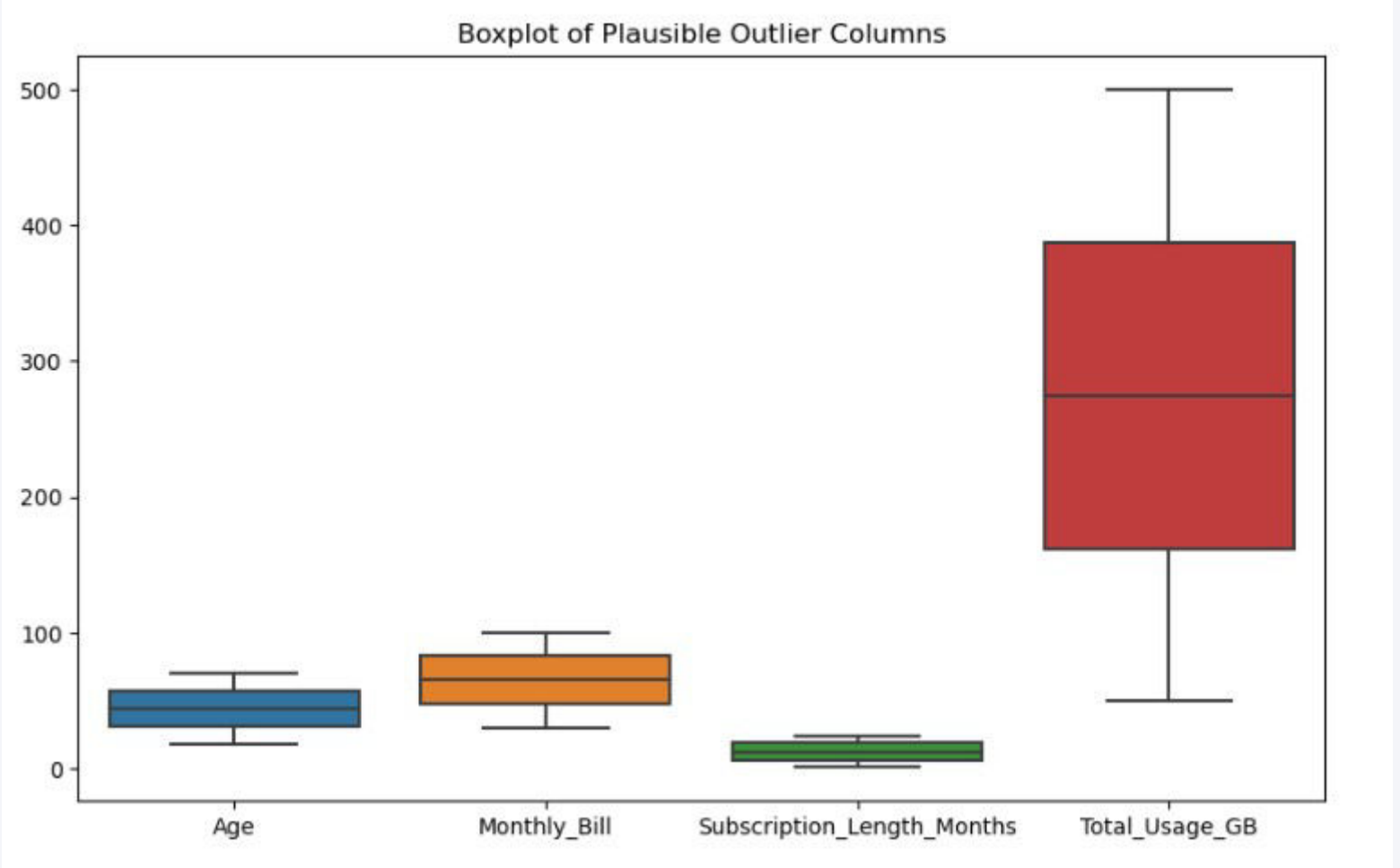
- I used .info() fuction to check for null values and boxplot to check for outliers, the data didn’t have any null value or outliers.

•Performed exploratory data analysis (EDA), on the basis of it decided to drop customer ID and name , as they will not contribute in model prediction. Encoded the categorical variables like gender and Location.

•Performed predictive analysis using classification models

- Initially I used Random Forest model, but the metrices didn’t give optimal values(even after fine tuning hyper-parameters), so I went with Logistic regression which was predicting output with 50% accuracy.

EDA with Data Visualization



Predictive Analysis (Classification)

```
from sklearn.linear_model import LogisticRegression
# Build and train the Logistic Regression model
logreg_model = LogisticRegression(random_state=42)
logreg_model.fit(X_train, y_train)

# Predictions on the test set
y_pred = logreg_model.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1-score:", f1)
```

```
Accuracy: 0.5036
Precision: 0.49953488372093025
Recall: 0.3788932567281524
F1-score: 0.43092972601169327
```

Results

```
# Simulating predictions for new customer data
new_customer_age = 30
new_customer_gender = 'Male'
new_customer_location = 'Los Angeles'
new_customer_subscription_length = 12
new_customer_monthly_bill = 80

preprocessed_data = preprocess_new_customer_data(new_customer_age, new_customer_gender, new_customer_location,
                                                 new_customer_subscription_length, new_customer_monthly_bill)
prediction = loaded_model.predict(preprocessed_data)
print("Churn Prediction for New Customer:", prediction)
```

```
Churn Prediction for New Customer: [0]
```

That is, a customer with above specification will not churn.