- Bar Plot for variable fraud_bool
- Based on Fig 1, the fraud_bool variable exhibits an imbalanced distribution. The percentage of instances labeled as 0 (non-fraudulent) is 98.8971%, while the percentage of instances labeled as 1 (fraudulent) is 1.1029%

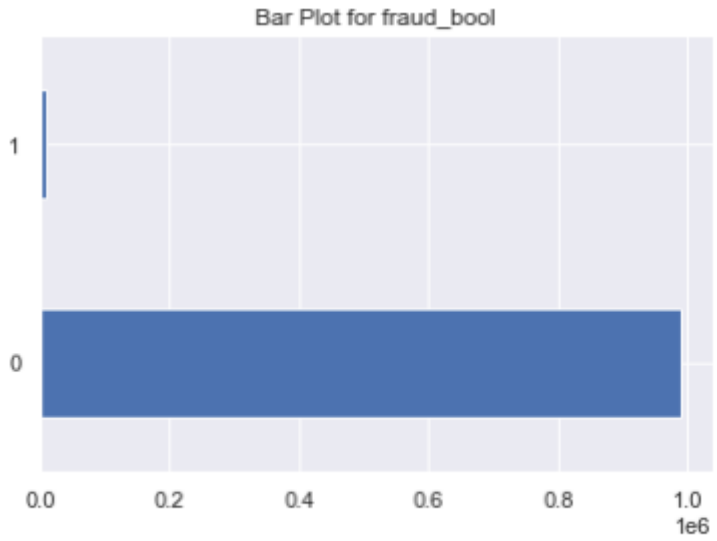


Fig 1. Bar Plot for variable fraud_bool

- Bar Plot for customer age
- This describes the age distribution of applicants,more than 93% of the applicants fall within the age range of 20-50. Less than 1% of the applicants are aged 60 or above. Approximately 2% of the applicants are aged 10.

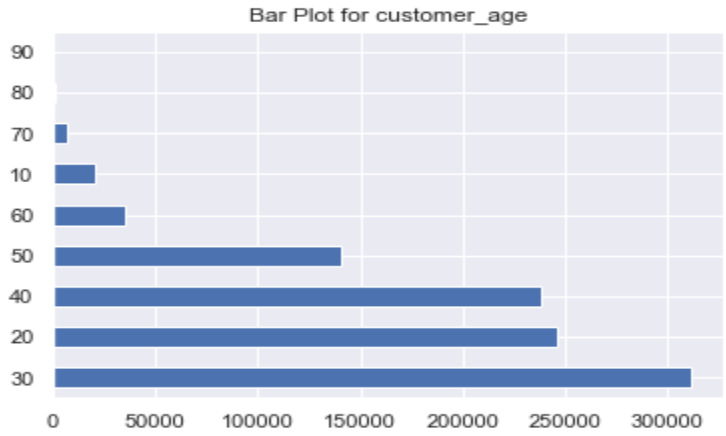


Fig 2 Bar Plot of customer age

- Box Plot of credit_risk_score with Outliers
- As depicted in Figure 3, the column "credit_risk_score" initially displayed outliers. However, after outlier treatment, as shown in Figure 4, the outliers were successfully eliminated. Similar outlier treatment techniques were applied to other features in the dataset that also contained extreme values.

Outlier Analysis:

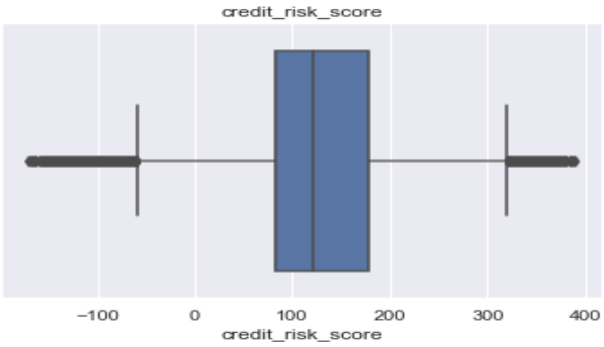


Fig 3 Box Plot of credit_risk_score with Outliers

- Box plot of credit_risk_score without outliers

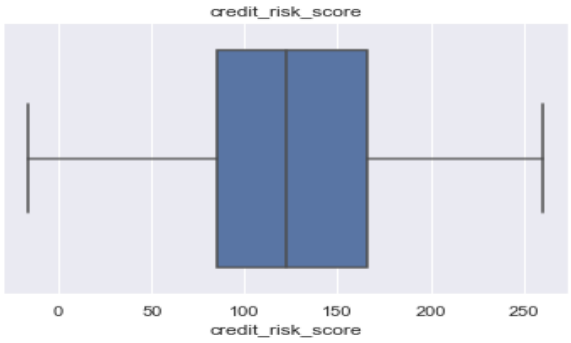


Fig 4 Box Plot of credit_risk_score without Outliers

Correlation Graph(Numerical vs Numerical)

Between the month and velocity_4w, we have the maximum correlation, which can be observed from Figure 5.

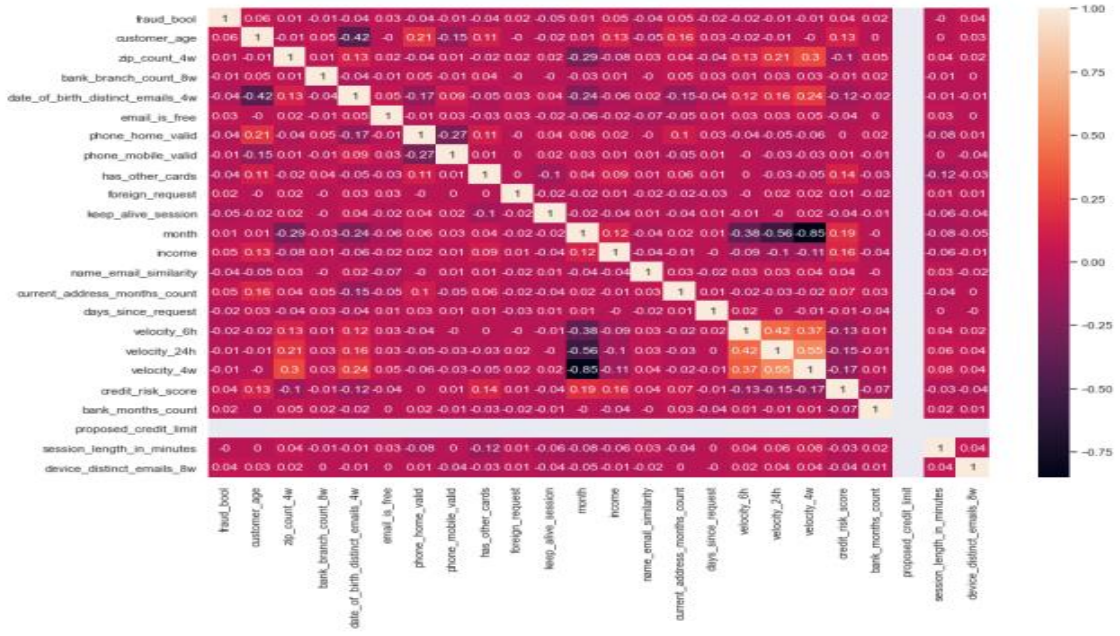
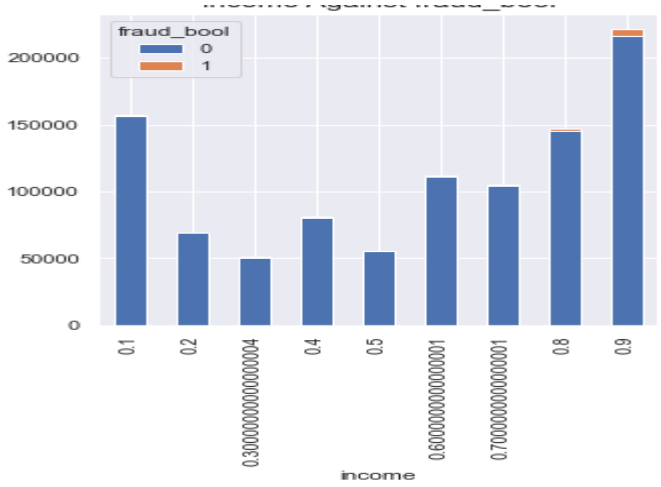


Fig 5. Correlation Graph(heat map)

- **Income again fraud_bool**



- **Housing status again fraud_bool**

Fig 7 Bar Graph housing status vs Fraud_bool

During the analysis, it was observed that when the income falls within the range of 0.9 (decimal value), the chances of fraud increase. This indicates that individuals with incomes close to 1 are more likely to be associated with fraudulent activities.

- **After evaluating various models, the Easy Ensemble model emerged as the top performer, demonstrating the following performance metrics:**

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.81      0.90    197776
           1       0.05      0.80      0.09      2224

    accuracy                           0.81    200000
   macro avg       0.52      0.81      0.49    200000
weighted avg       0.99      0.81      0.89    200000
```

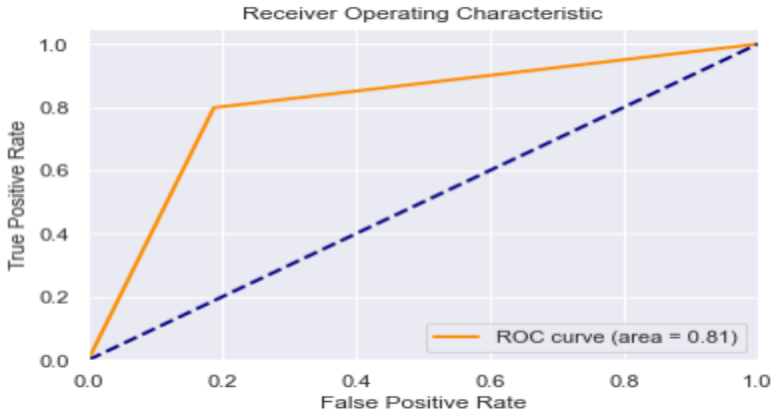


Fig 8 AUC-ROC Curve

Confusion Matrix