10Alytics
data & strategy

ONLINE PAYMENTS FRAUD DETECTION

BY

VICTOR ETIM

(DATA ANALYST)

CAPSTONE PROJECT

2022

CONTENT TABLE

SECTION 1

- About Client
- Problem Description

SECTION 2

Data visualization and Analysis

SECTION 3

Fraud detection model analysis

SECTION 4

Action Points

ABOUT CLIENT

Blossom Bank also known as BB PLC is a multinational financial services group, that offers retail and investment banking, pension management, asset management and payments services, headquartered in London, UK.

10Alytics
data & strategy

PROBLEM DESCRIPTION

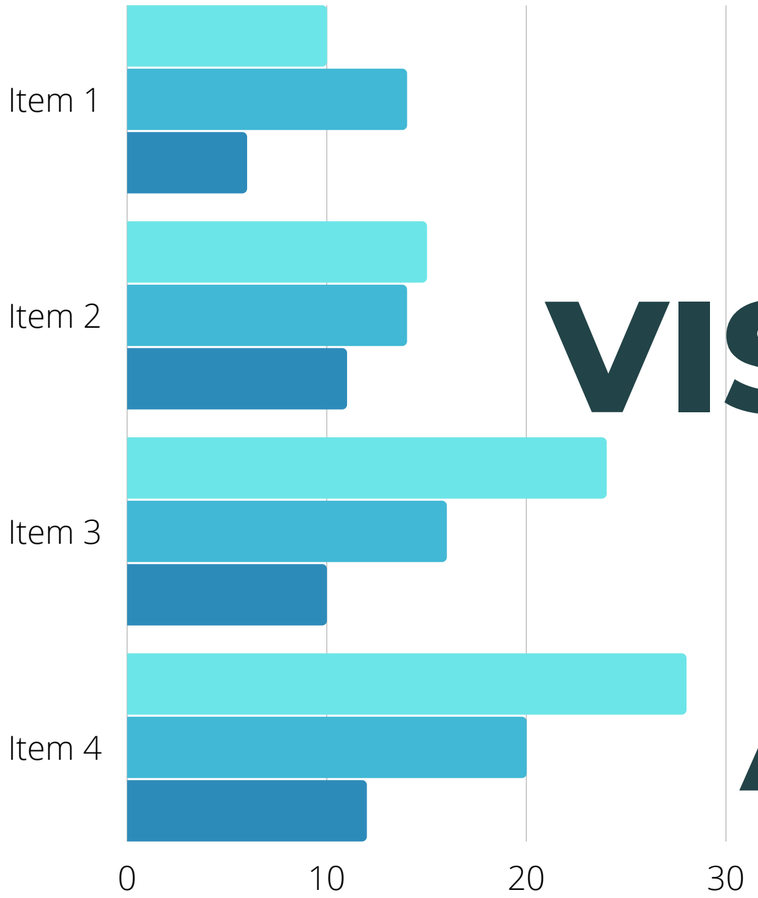
Problem Statement

- Blossom Bank wants to build a Machine Learning model to predict online payment fraud.
- With this model they are expected to make less False Negatives with regards to detecting fraudulent transactions.

Project Objective

- Visualize the relationships of the columns within the dataset (EDA).
- Find out what machine learning (ML) model will detect frauds better.

10Alytics
data & strategy



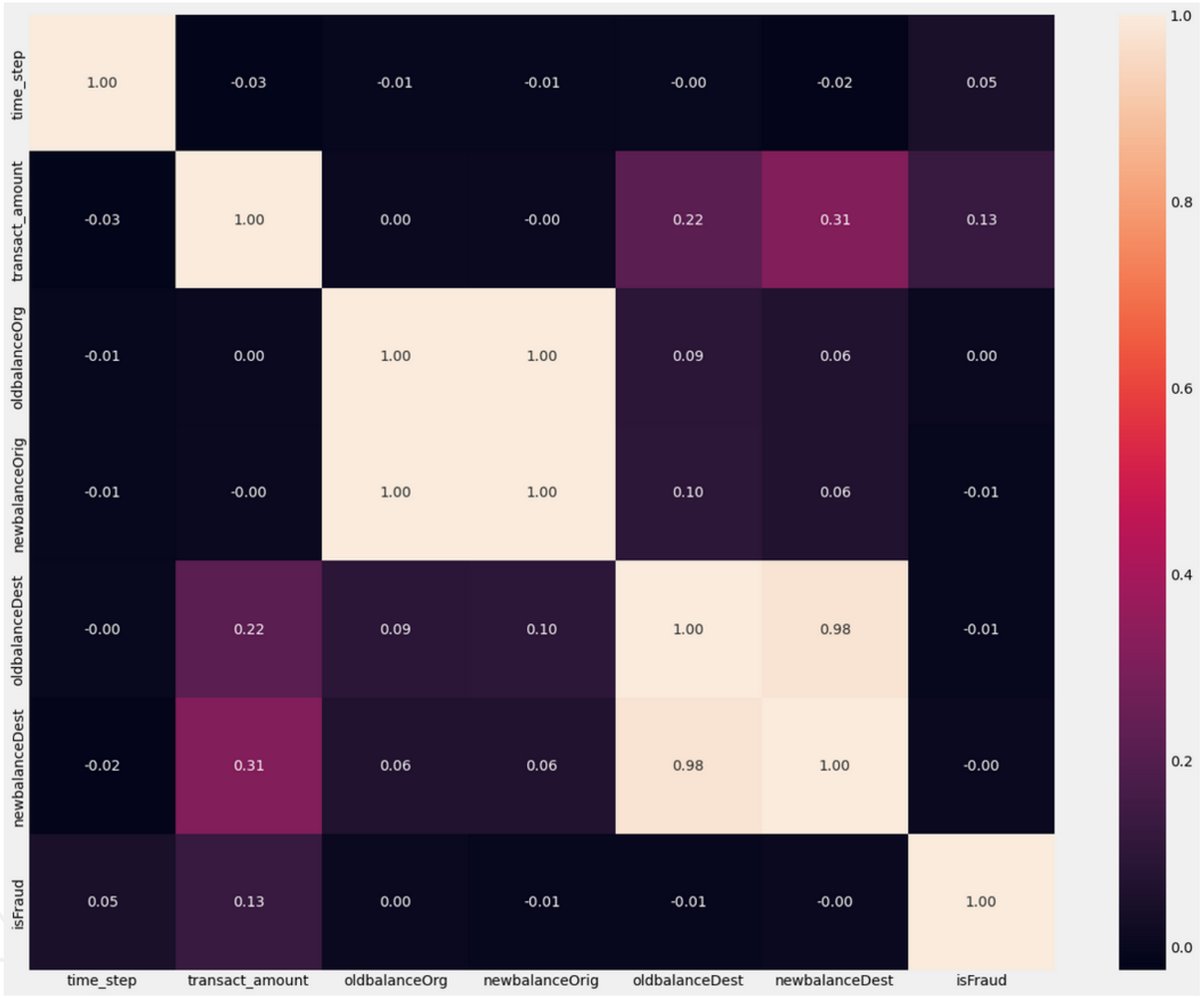
DATA VISUALIZATION & ANALYSIS

10Alytics
data & strategy

DATA INSPECTION

During data inspection, it was discovered that:

- No missing values were present
- The variable types within each column were correctly inputed.



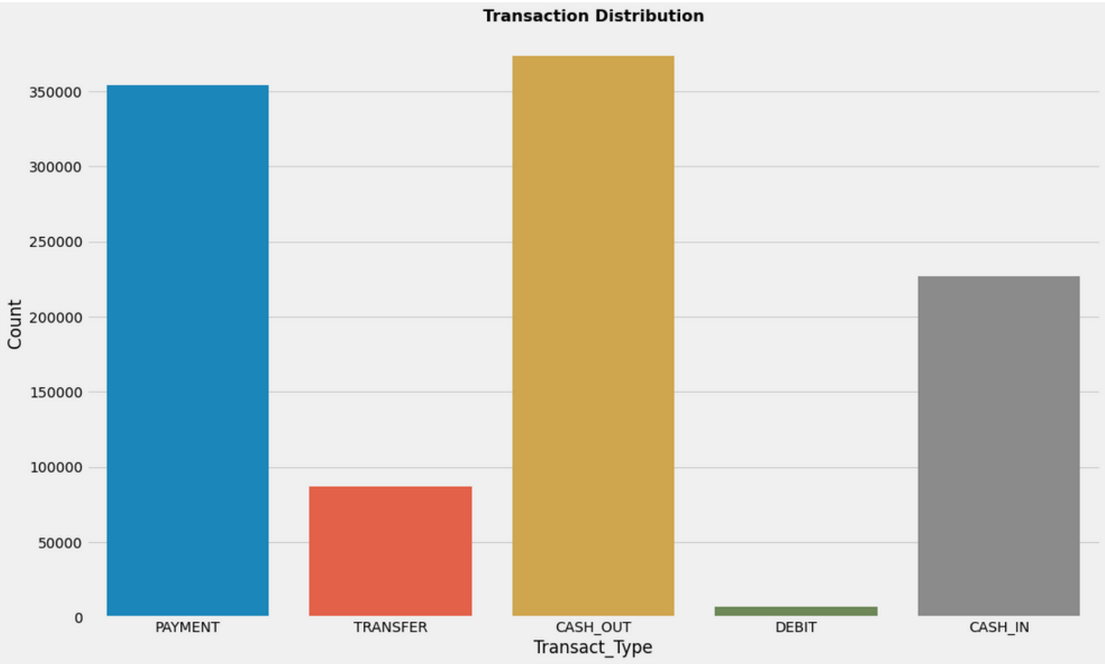
	time_step	transact_amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFraud
time_step	1.000000	-0.025996	-0.006780	-0.007180	-0.002251	-0.019503	0.045030
transact_amount	-0.025996	1.000000	0.004864	-0.001133	0.215558	0.311936	0.128862
oldbalanceOrg	-0.006780	0.004864	1.000000	0.999047	0.093305	0.064049	0.003829
newbalanceOrig	-0.007180	-0.001133	0.999047	1.000000	0.095182	0.063725	-0.009438
oldbalanceDest	-0.002251	0.215558	0.093305	0.095182	1.000000	0.978403	-0.007552
newbalanceDest	-0.019503	0.311936	0.064049	0.063725	0.978403	1.000000	-0.000495
isFraud	0.045030	0.128862	0.003829	-0.009438	-0.007552	-0.000495	1.000000

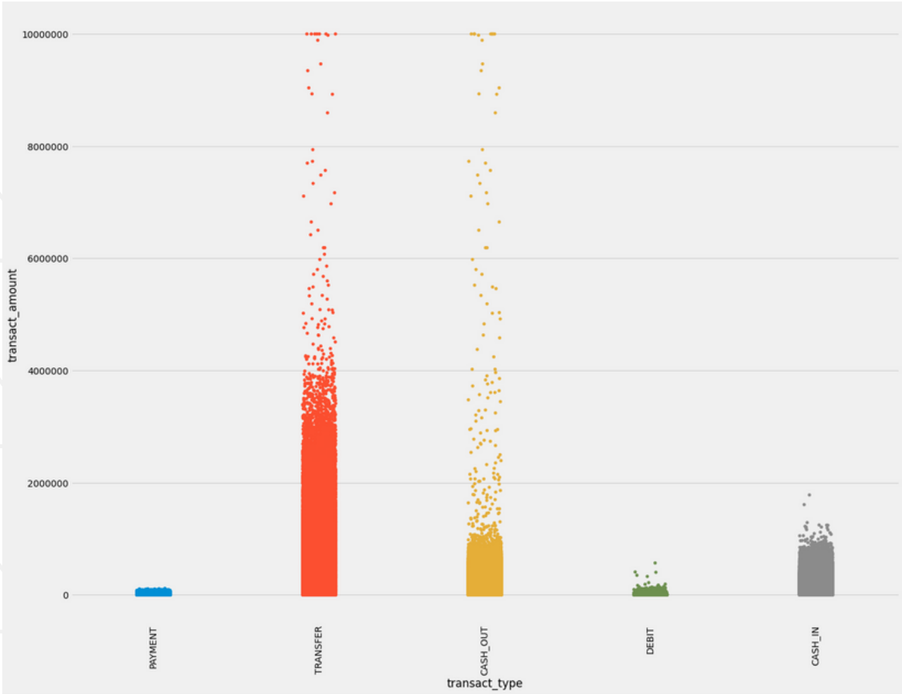
10Alytics
data & strategy

ANALYTICAL DISCOVERY

The correlation plot and table indicate that time_step (payment period) and transaction amount can be used to monitor fraudulent payments. By extension, checking the type of payment method is advisable.

"The transaction amount (transact_amount) appears to be the loudest fraud indicator."

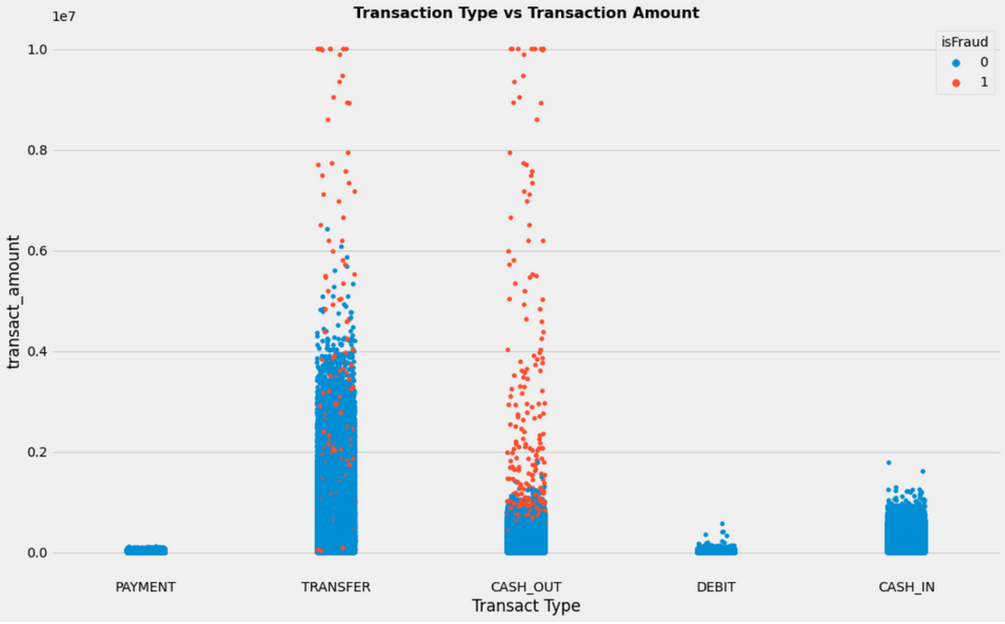




TRANSACTION AMOUNT & TRANSACTION TYPE

- Cash_Out and Payment have the highest number of recorded transactions.
- It's important to state though that, people who transact higher denominations of money prefer the Transfer or Cash_Out method.

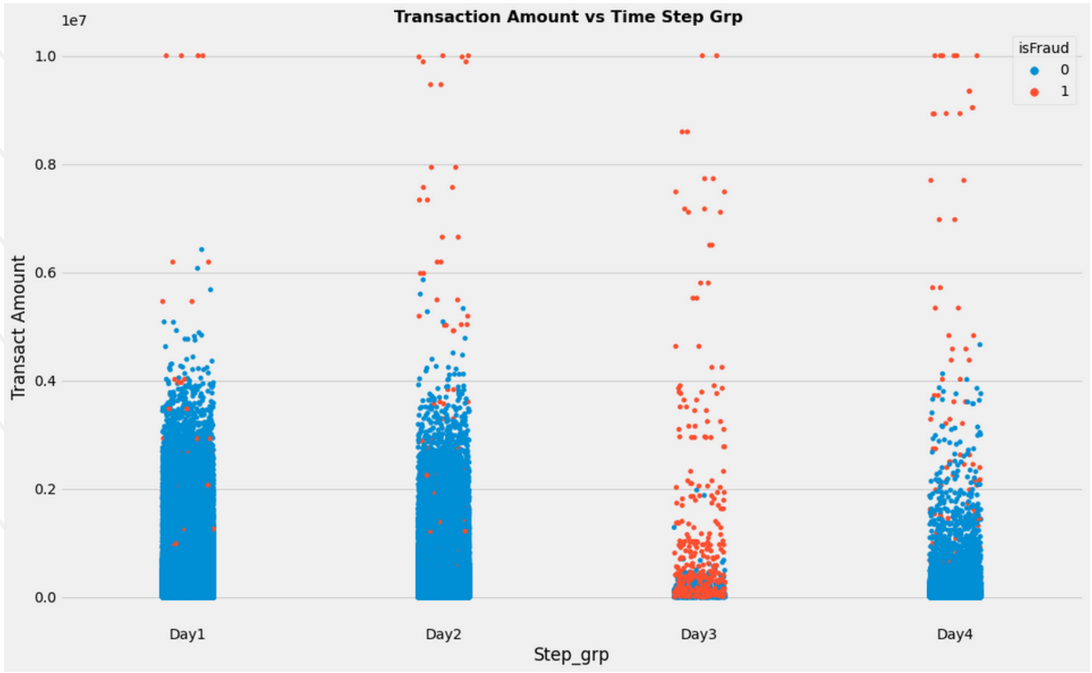
"CASH_OUT is the most frequent transaction type with a count of 373,641.
"



FRAUD DETECTION WITH TRANSACTION AMOUNT & TRANSACTION TYPE

- The strip plot helps explain the box plot above in a little more detail.
- Most of the frauds detected are within the transfer and cash_out payment types.
- All transaction amounts over 6 million were predominantly flagged as fraudulent.

"Transactions over 6 million were predominantly flagged as fraudulent.
"



10Alytics
data & strategy

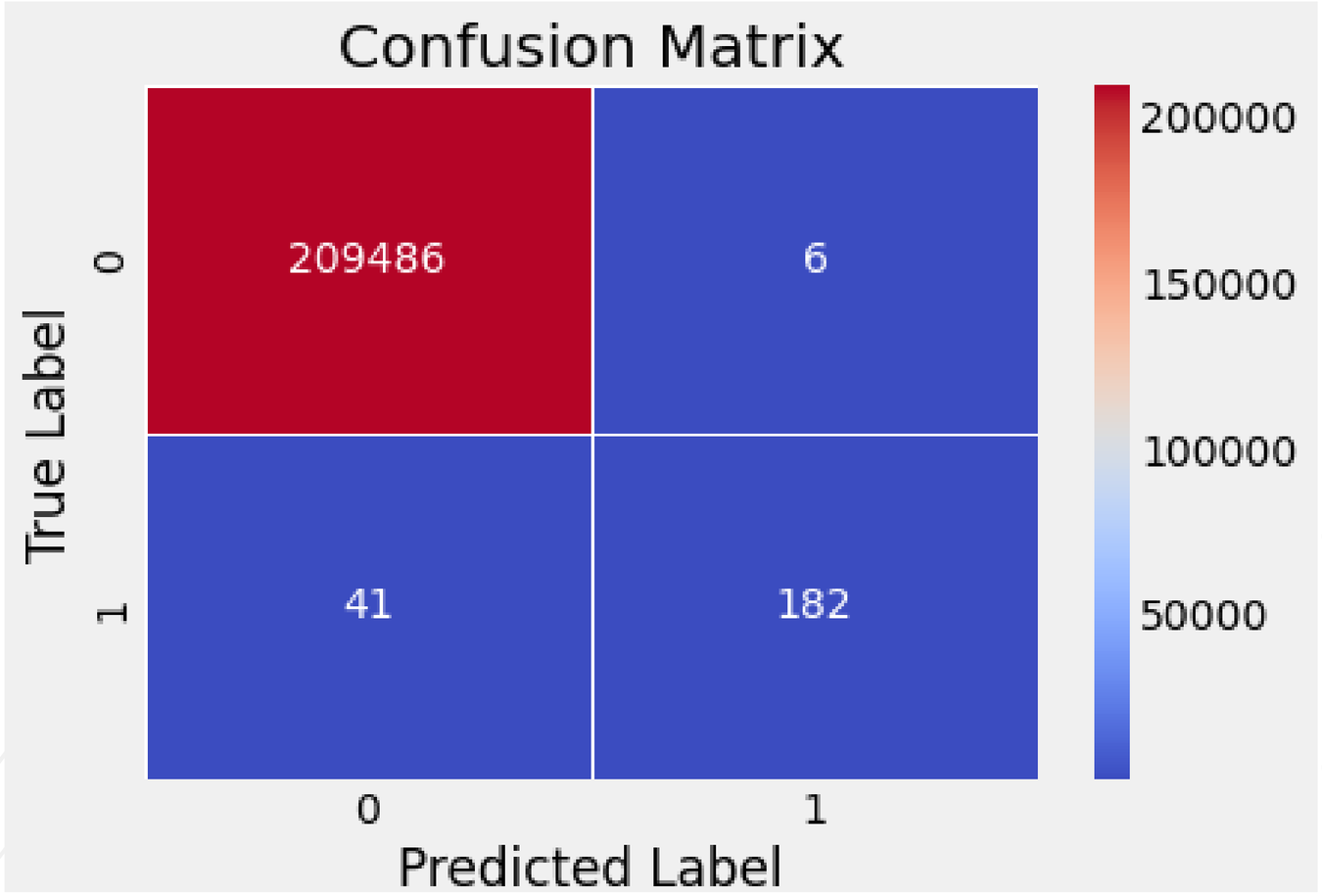
FRAUD DETECTION WITH TRANSACTION AMOUNT & TIME STEP

- The time_step was grouped into days (24-hr clock); and since the highest transaction amount was 10 million, the amount was segmented into an interval of 2 million.
- The most fraud occurred on transactions pending for 3 days before completion.
- But, there were traces of fraudulent transactions daily.

"The most fraud occurred on transactions pending for 3 days before completion."

10Alytics
data & strategy

DETECTION MODEL ANALYSIS



For RandomForestClassifier, Accuracy score is 0.999775886321913

	precision	recall	f1-score	support
0	1.00	1.00	1.00	209492
1	0.97	0.82	0.89	223
accuracy			1.00	209715
macro avg	0.98	0.91	0.94	209715
weighted avg	1.00	1.00	1.00	209715

10Alytics
data & strategy

SUMMARY

- The confusion matrix metric for this model is a RECALL, because the model would be predicting based on an event that has already occurred.
- The Random Forest model will detect fraud with the lowest number of False Negatives possible as opposed to that of the Logistic Regression, KNeighbors and Support Vector Machine models.

"The Random Forest Classifier model will give accurate target predictions."

RECOMMENDATIONS

10Alytics
data & strategy

Transaction:

Transactions over 6 million ought to be treated cautiously.

Detection Model:

Blossom Bank needs to adopt the Random Forest Classifier model as soon as possible.

10Alytics
data & strategy

CONTACT ME

vetim34@gmail.com

Mobile
+234 912 2410 887

www.linkedin.com/in/victoretim

https://github.com/vetim34

THANK YOU!

10Alytics
data & strategy