



# PREVENTING FRAUD

*USYD TEAM 1*



01. SUMMARY

02. PROCESS

03. DATA CLEANING & EDA

04. MODEL

05. INSIGHTS

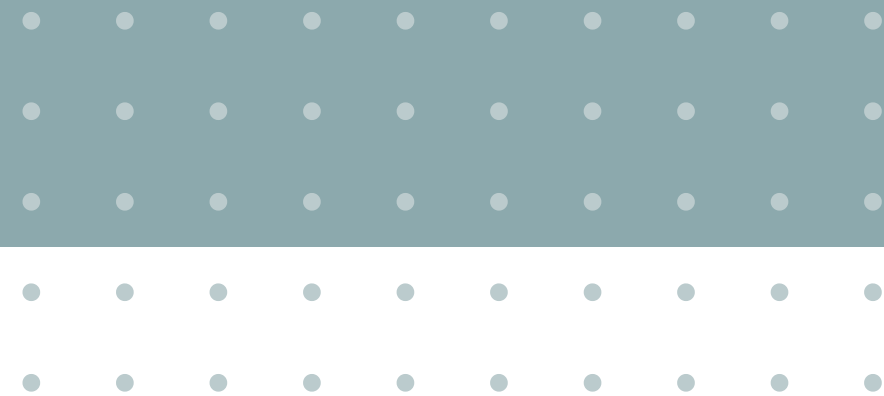


# TABLE OF CONTENT



01.

# SUMMARY





# EXECUTIVE SUMMARY



Skyline Financial Services has been suffering from revenue loss due to an increase in fraudulent activity.



Detect fraudulent users using machine learning models.



YOUNGER USERS, STUDENTS, AND THOSE WITH DEPENDENTS SHOWED HIGHER FRAUD RATES. SMALLER TRANSACTION AMOUNTS WERE MORE LIKELY TO BE FRAUDULENT.



LOGISTIC REGRESSION: ACCURACY: 0.721  
BASE RR: 0.987  
XGBOOST: ACCURACY: 0.997

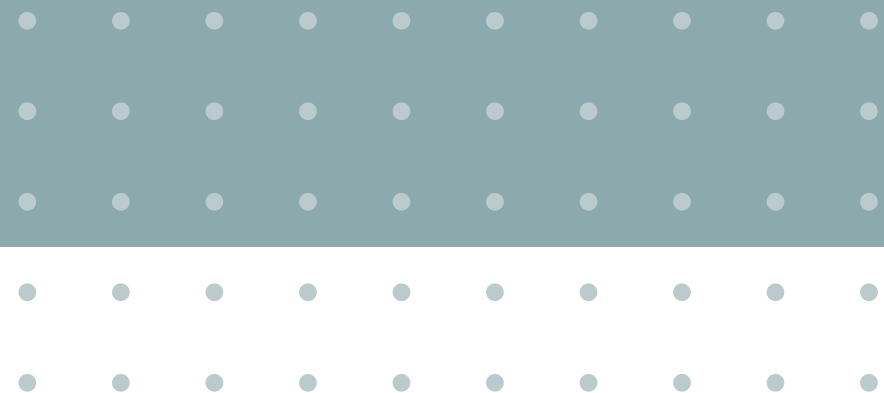


Deploy the XGBoost model for high-accuracy fraud detection,

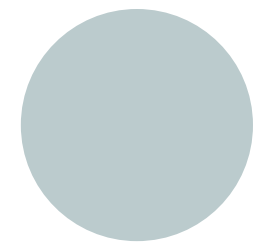


02.

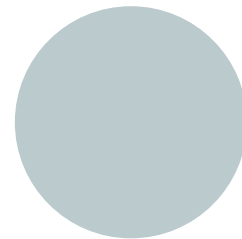
# PROCESS



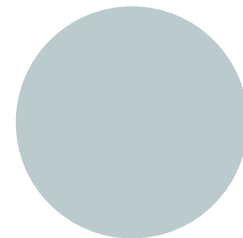
# OVERVIEW OF THE PROCESS



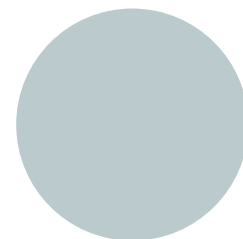
Data Understanding



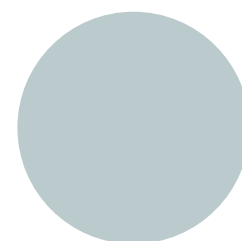
Data Cleaning



Data Modelling



Data Analysis

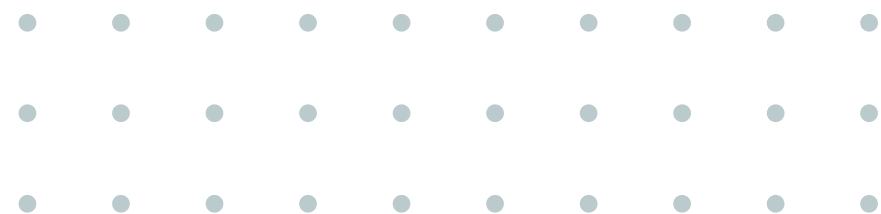


Uncover Insights



# KEY CHALLENGES

- Pinpointing the demographic groups most vulnerable to committing fraud.
- Uncovering behaviours strongly associated with fraudulent activities.
- Determining where and when fraud is most likely to occur.



# STAKEHOLDERS

Skyline  
Financial  
Services

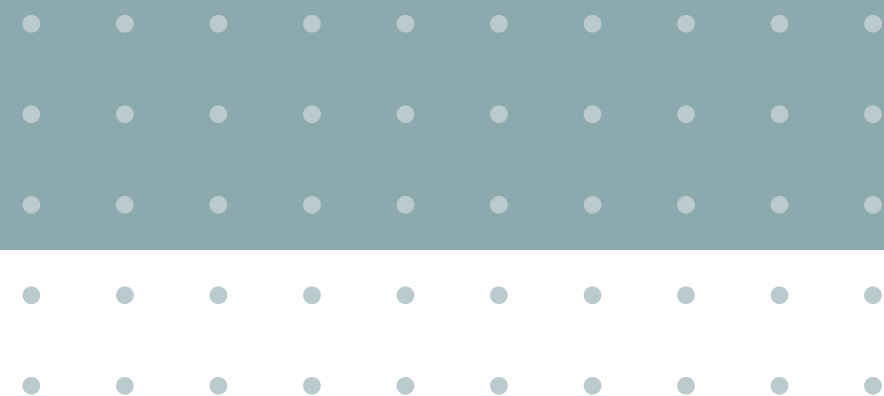
Customers

Investors



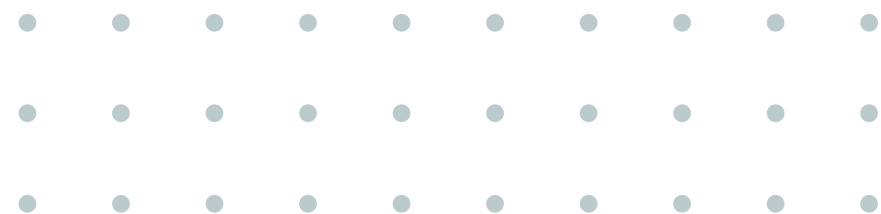
03.

# DATA CLEANING & EXPLORATORY DATA ANALYSIS

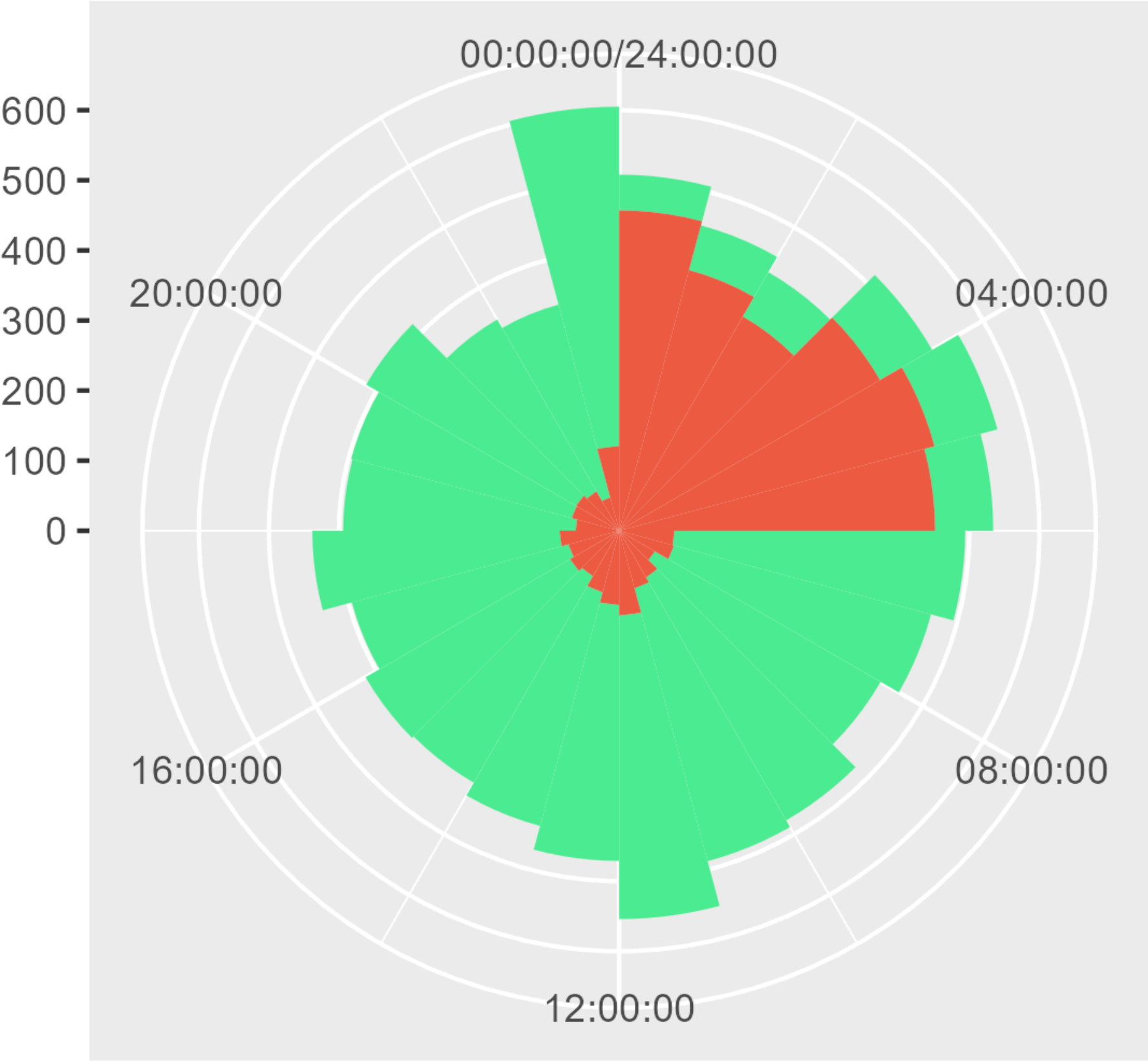
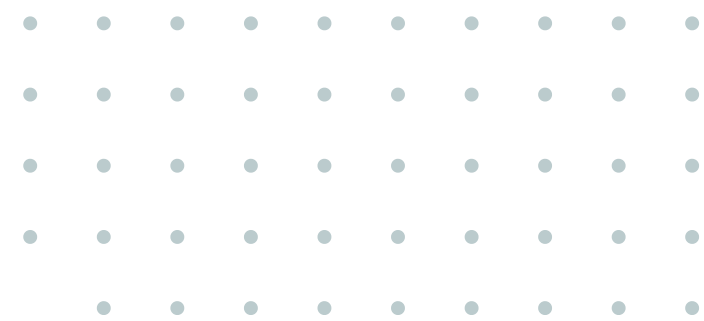


## DATA CLEANING

- Streamlined categorical variables such as location, gender and device type
- Split monetary values into 3 columns, amount, currency and amount in AUD



85% of transactions between 12-6am are fraudulent



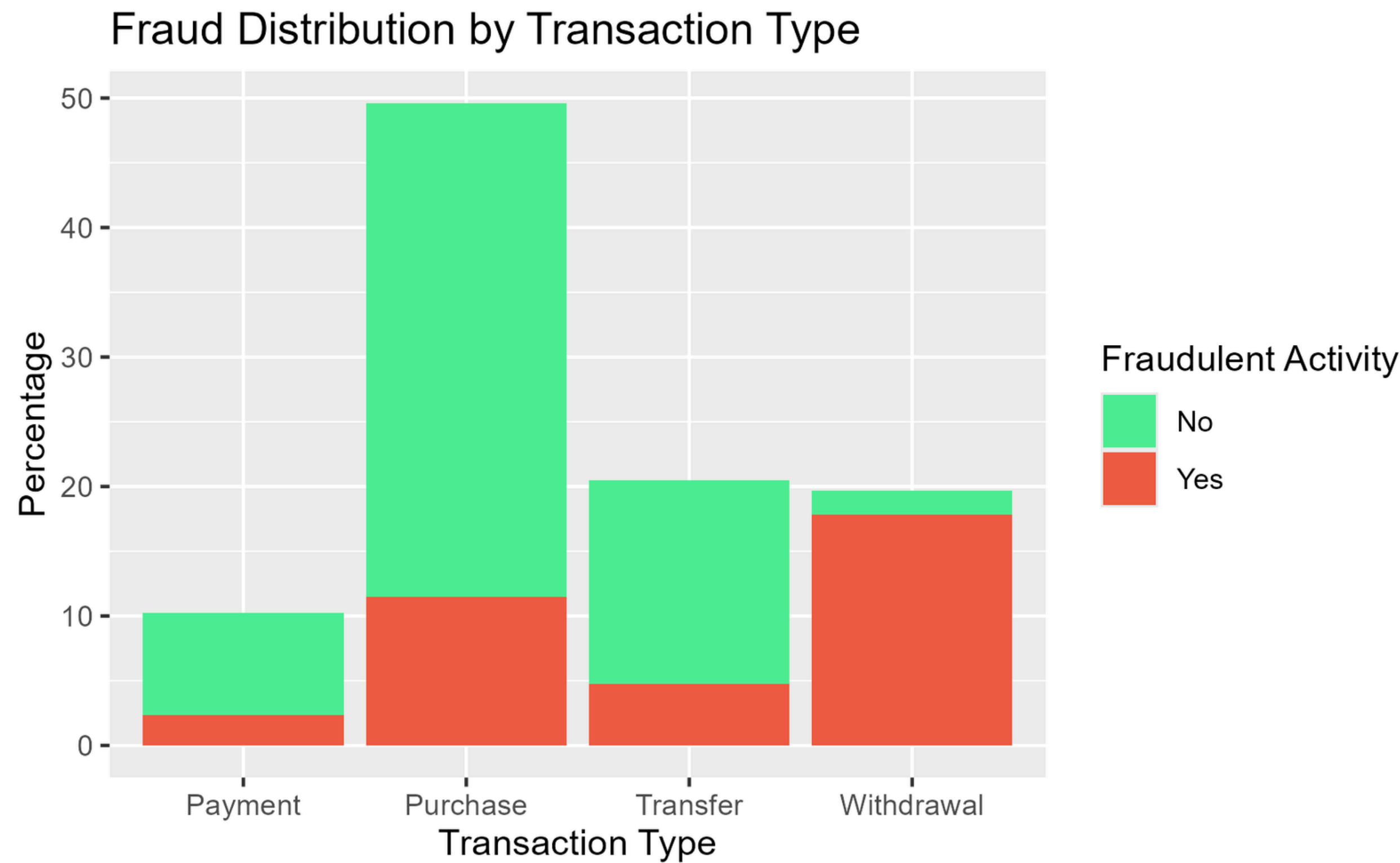
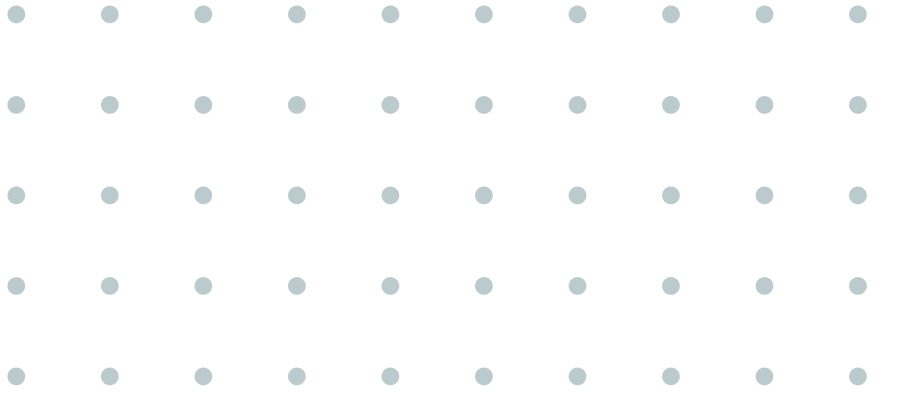
Fraudulent Activity

No

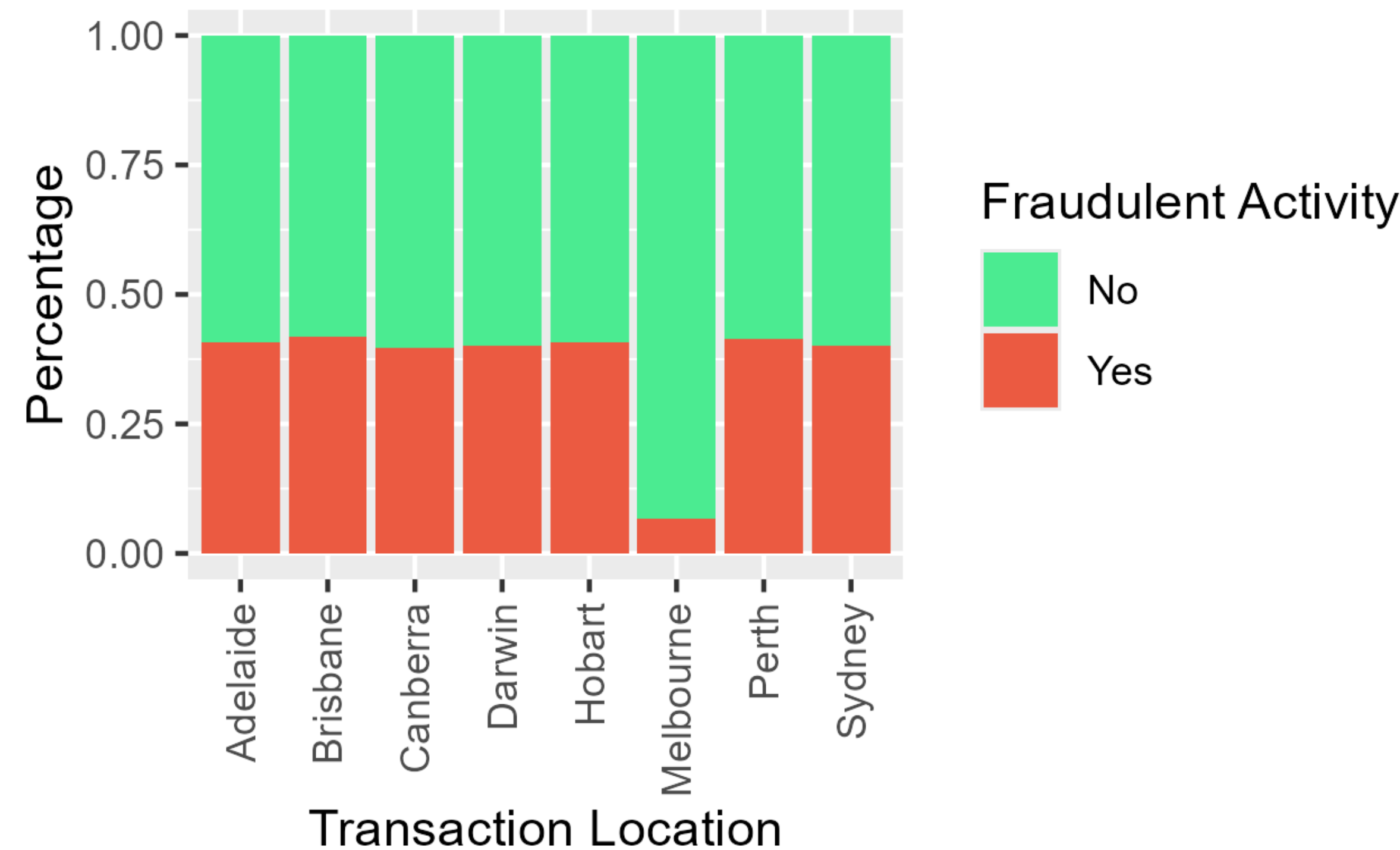
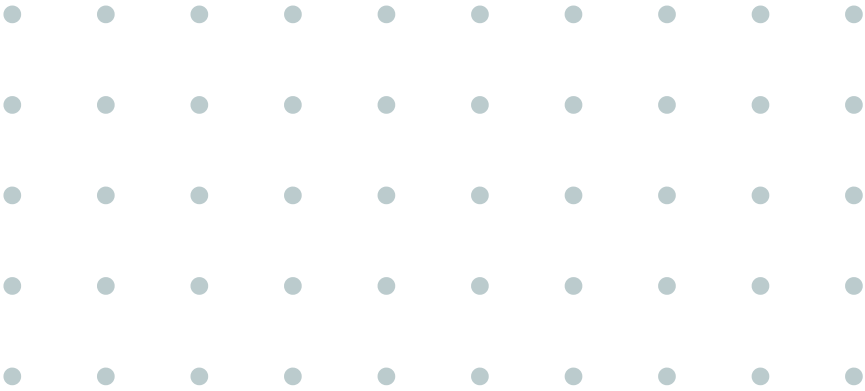
Yes



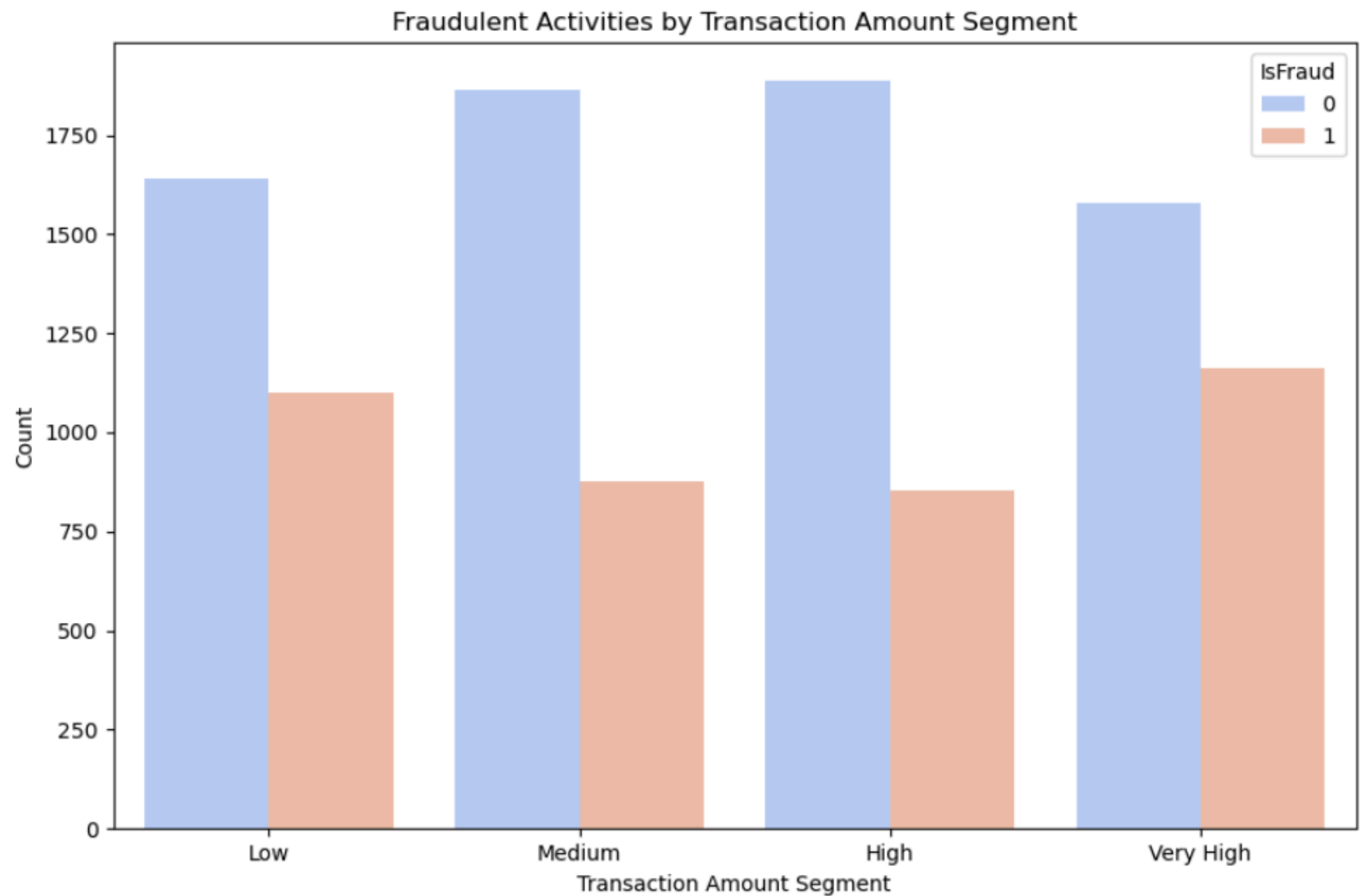
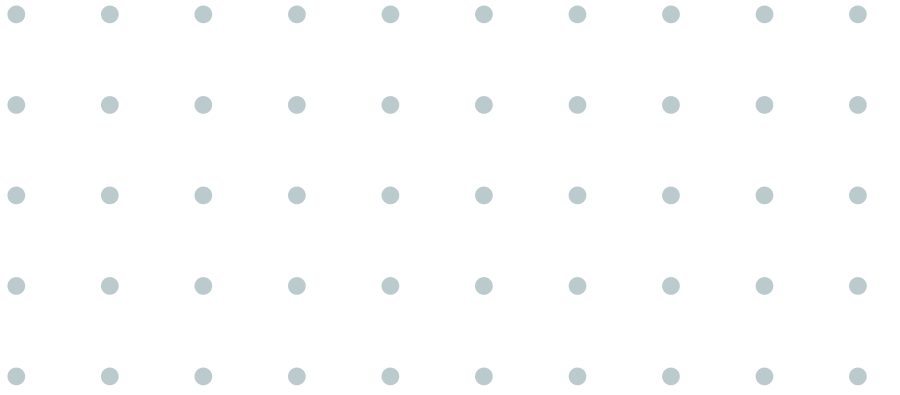
# Withdrawal is likely to be more fraudulent than other transactions



# They occur least in Melbourne



Very high transaction values are more likely to be fraudulent...



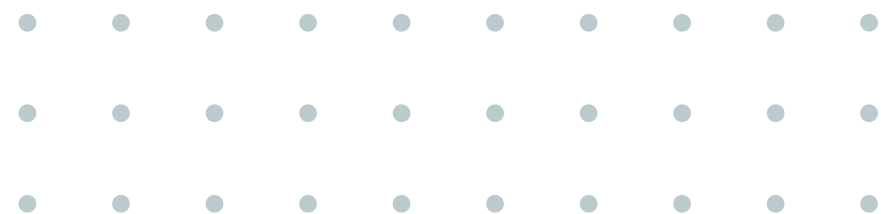
04.

# MODELS



# PREDICTIONS MODEL

- **LOGISTIC REGRESSION**
  - Accuracy: 0.721
  - Initial attempt
- **RANDOM FOREST:**
  - Accuracy: 0.987
  - Initial attempt with strong performance, but room for improvement.
- **XGBOOST:**
  - Accuracy: 0.997
  - Clear improvement, high classification performance.





# XG MODEL OUTPUT

Validation Log Loss: 0.0071

Validation AUC: 1.0000

Validation F1 Score: 0.9973

Confusion Matrix:

[[1366 2]

[ 4 822]]

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1368
1	1.00	1.00	1.00	826
accuracy			1.00	2194
macro avg	1.00	1.00	1.00	2194
weighted avg	1.00	1.00	1.00	2194

05.

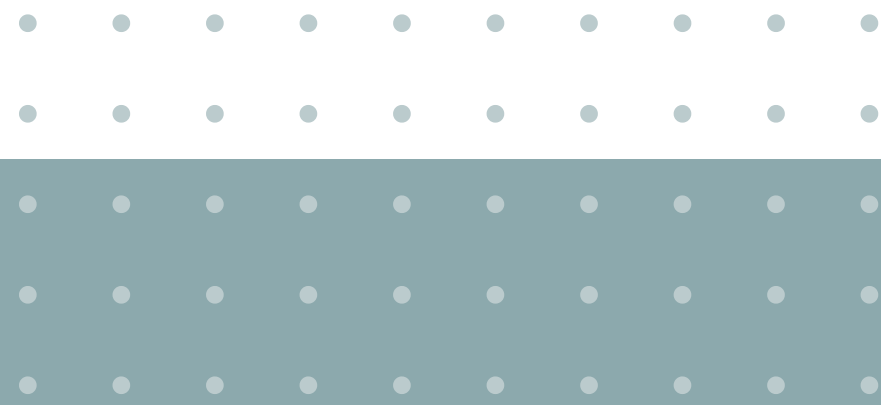
# CONCLUSION



# CONCLUSION



- 1 **XGBOOST MODEL ACHIEVED THE HIGHEST ACCURACY AT 0.997.**
- 2 **KEY INSIGHTS:** TIME OF DAY AND TRANSACTION TYPE ARE MOST RISKY
- 3 **STRATEGY:** IMPLEMENT XGBOOST FOR TARGETED FRAUD DETECTION.
- 4 **NEXT STEPS:** MONITOR AND RETRAIN MODEL WITH NEW DATA.



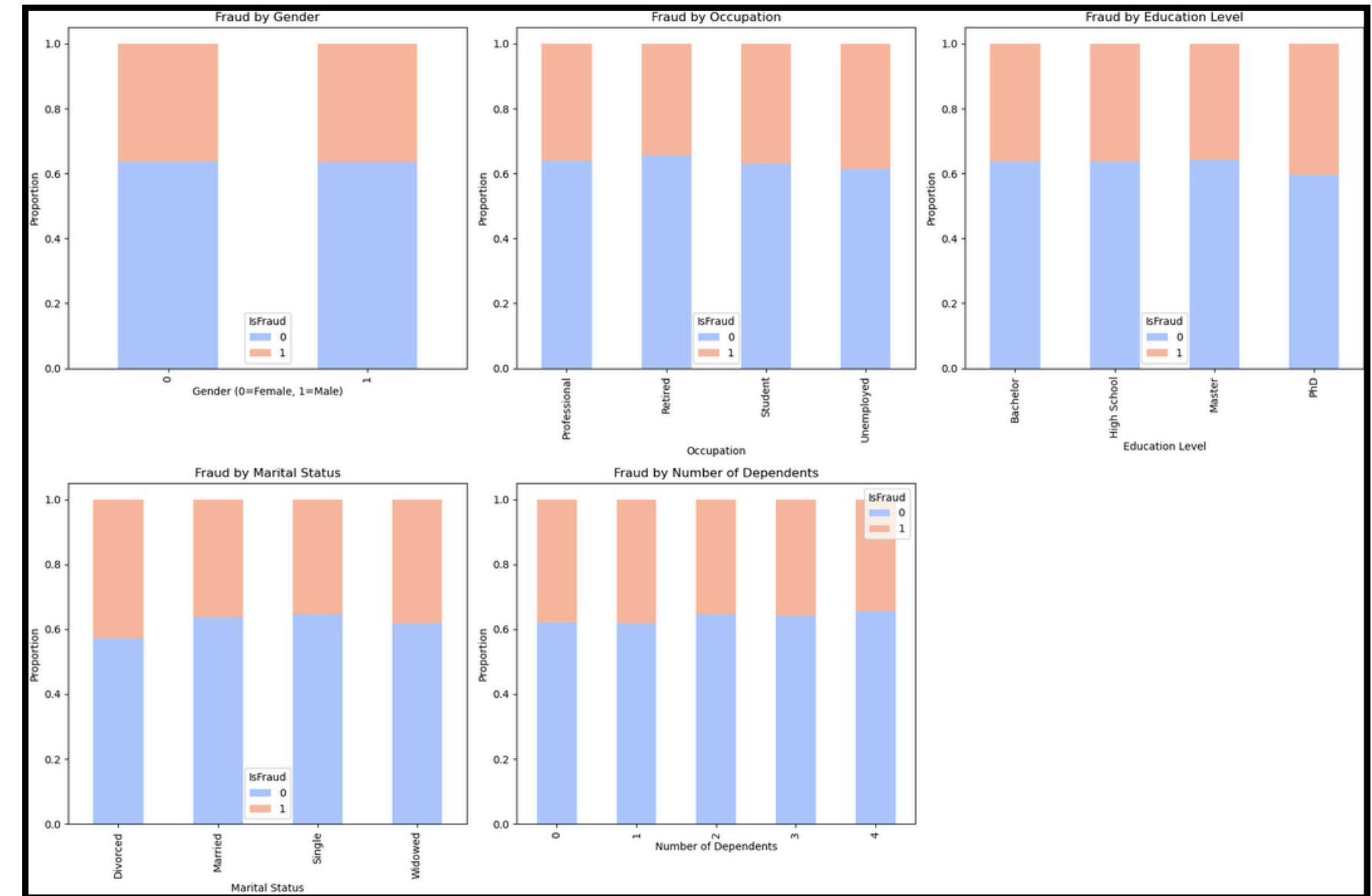
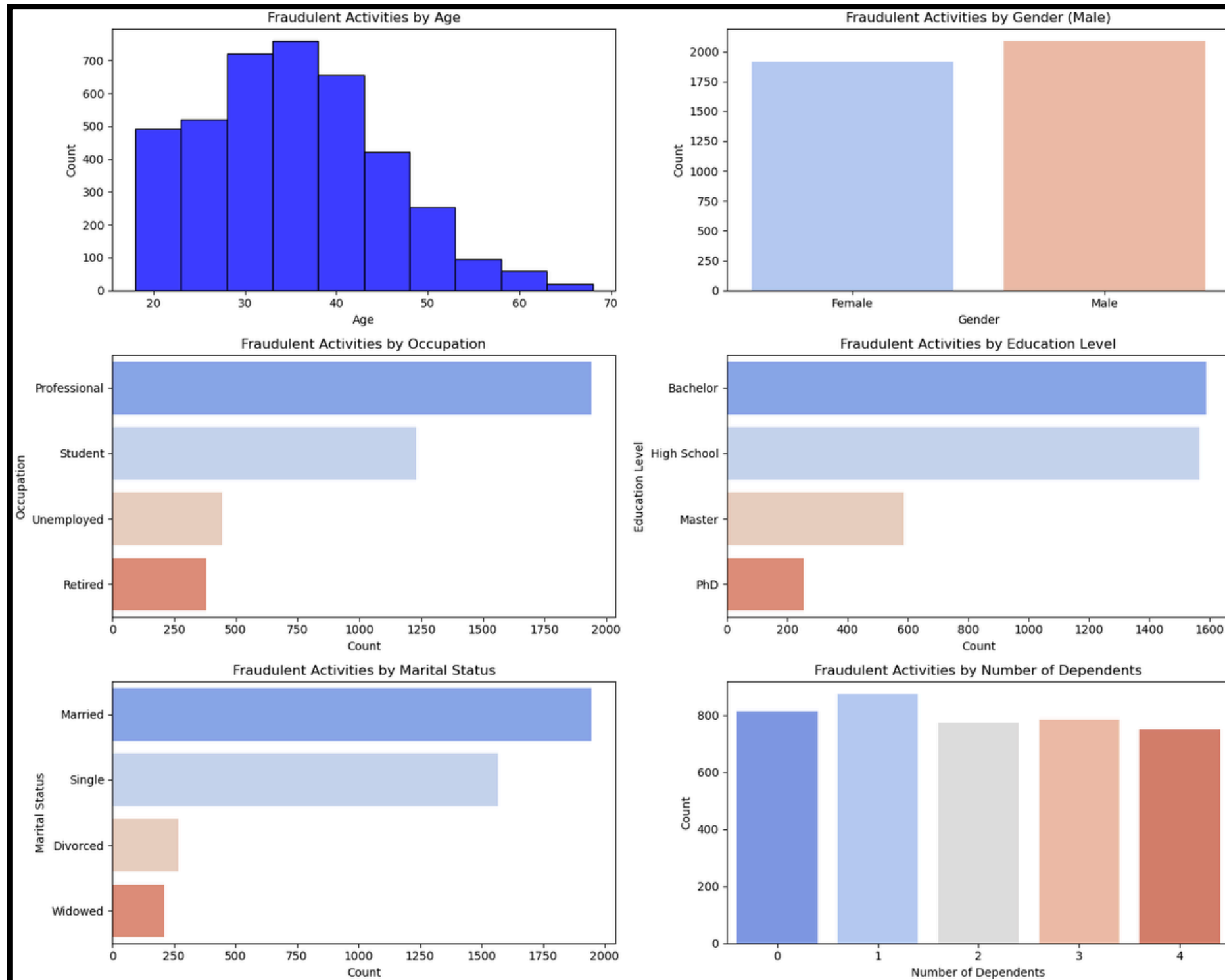
# THANK YOU

Do you have any question?

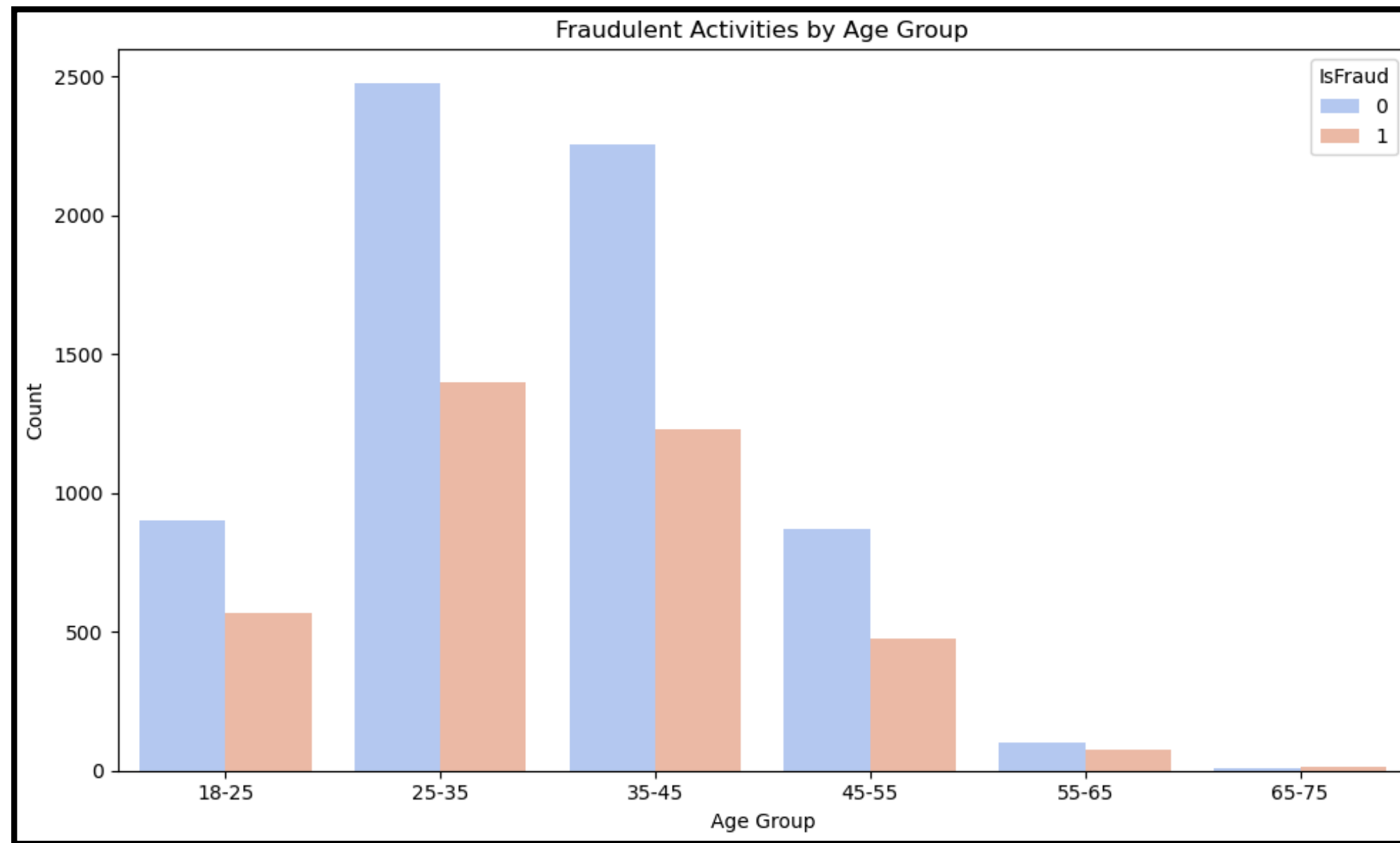


# APPENDIX

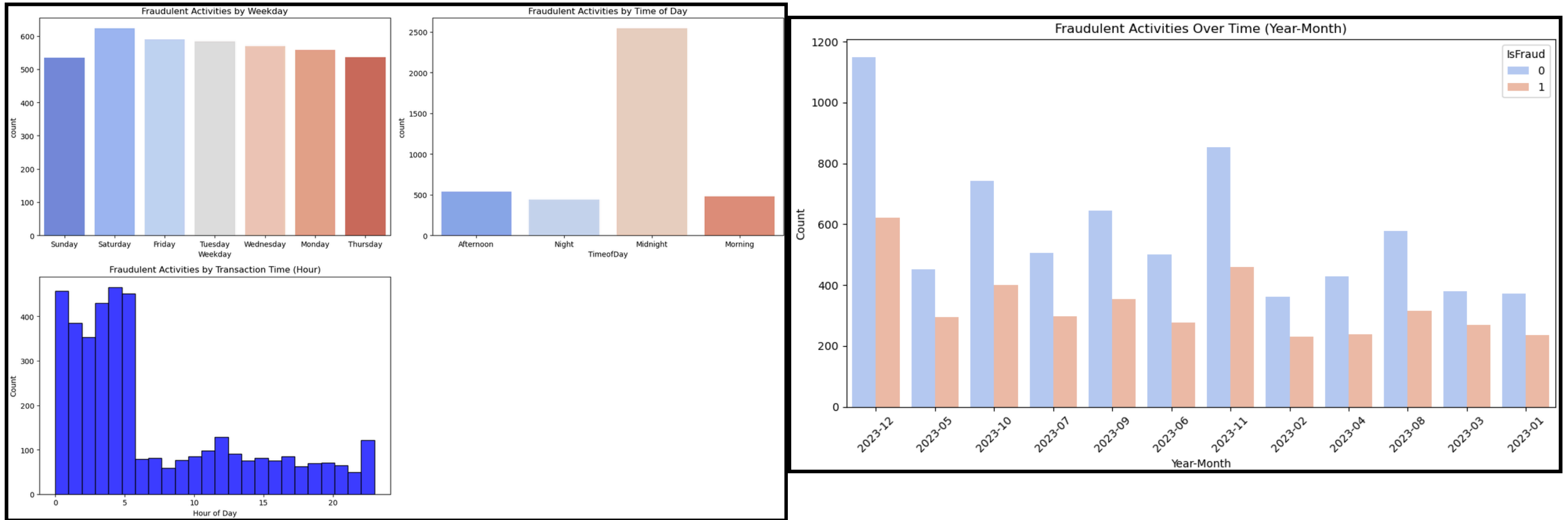
# 1.1 USER DEMOGRAPHICS



# 1.2 USER DEMOGRAPHICS

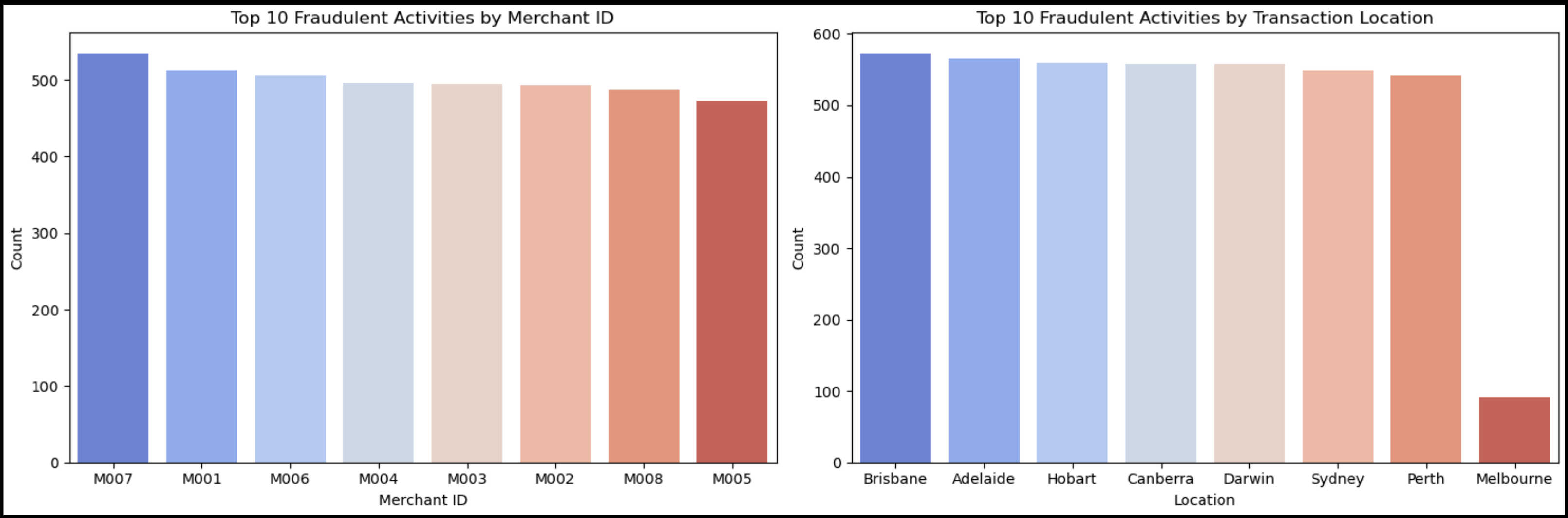


# 2.1 TIME-RELATED INFORMATION VISUALIZATION

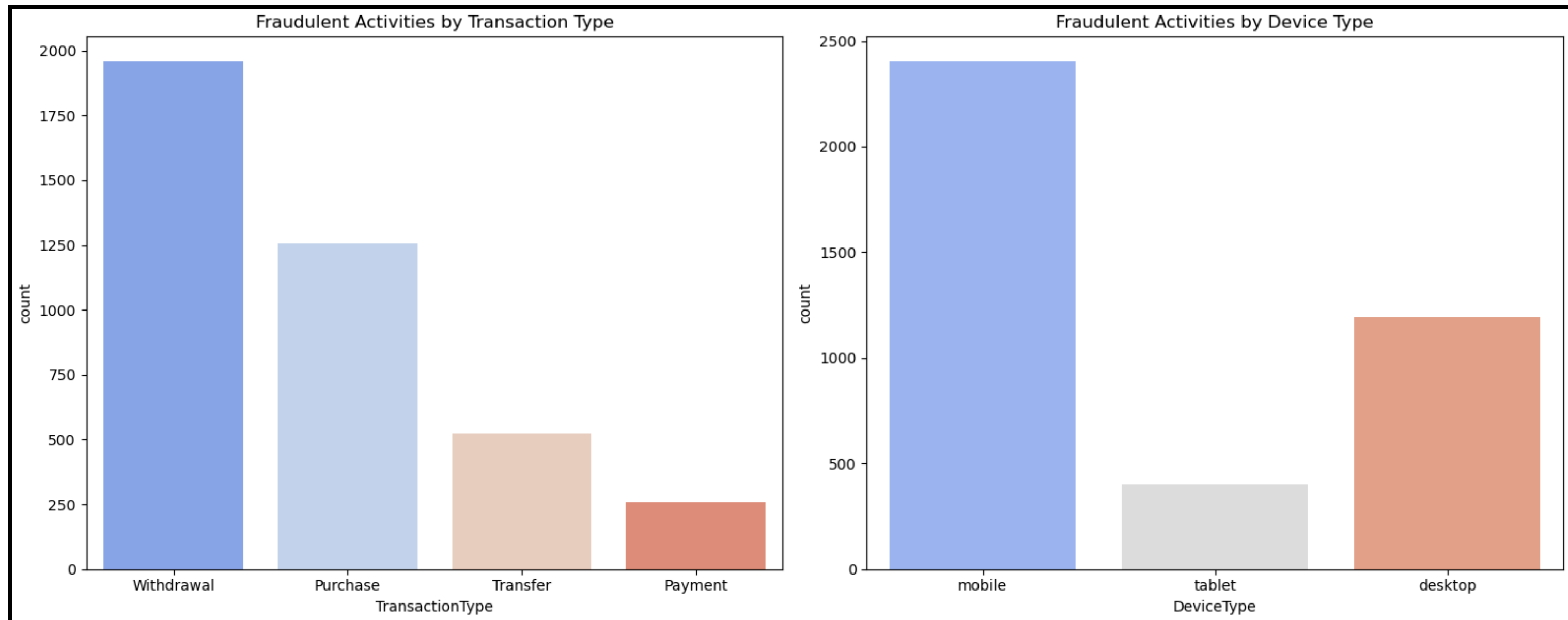




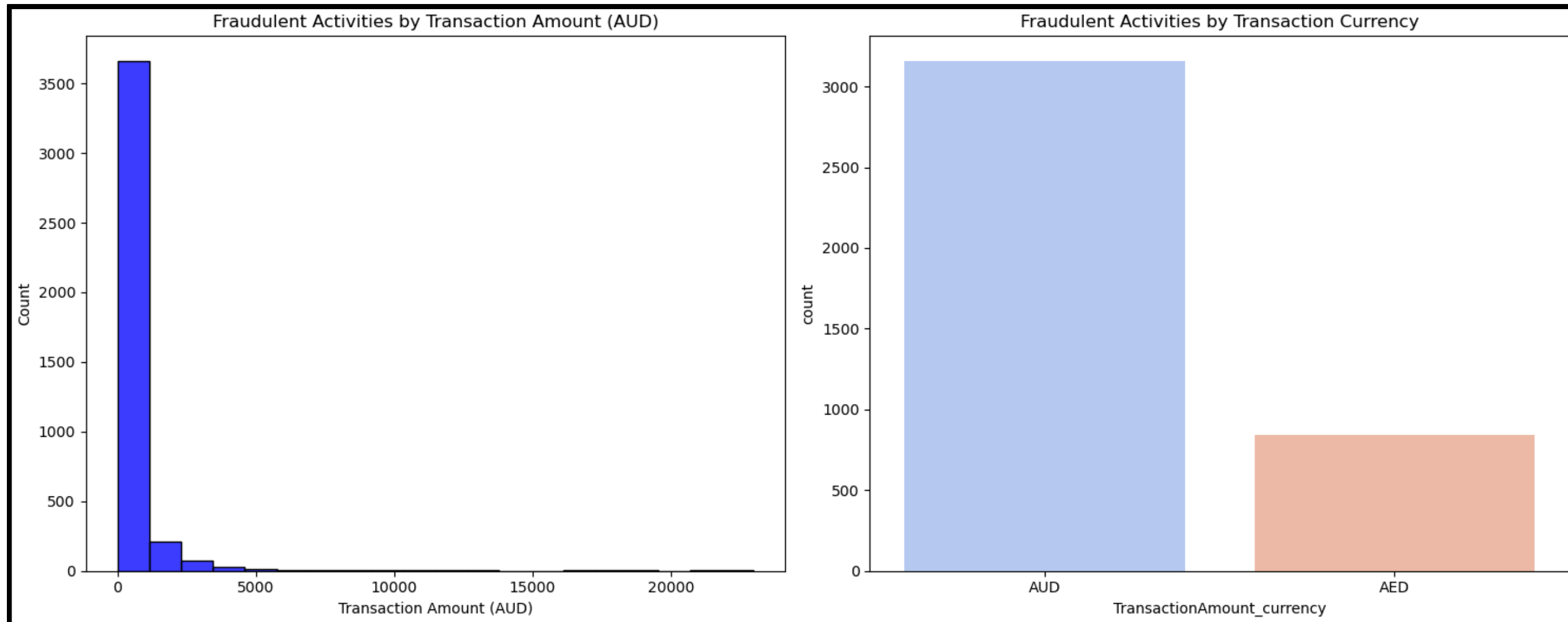
# 2.2 MERCHANT AND LOCATION-RELATED INFORMATION VISUALIZATION



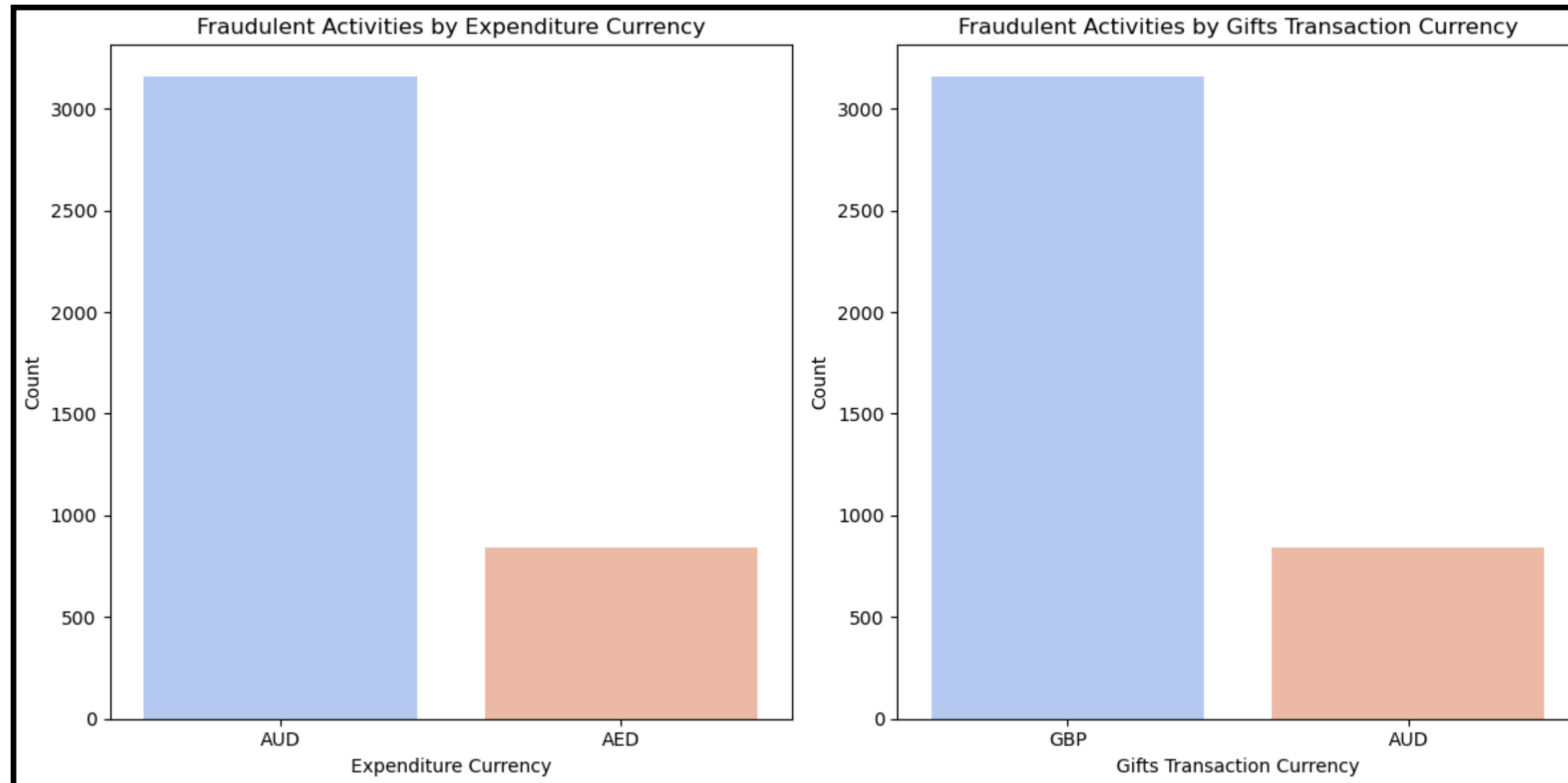
## 2.3 TRANSACTION TYPE AND MEDIUM-RELATED INFORMATION VISUALIZATION



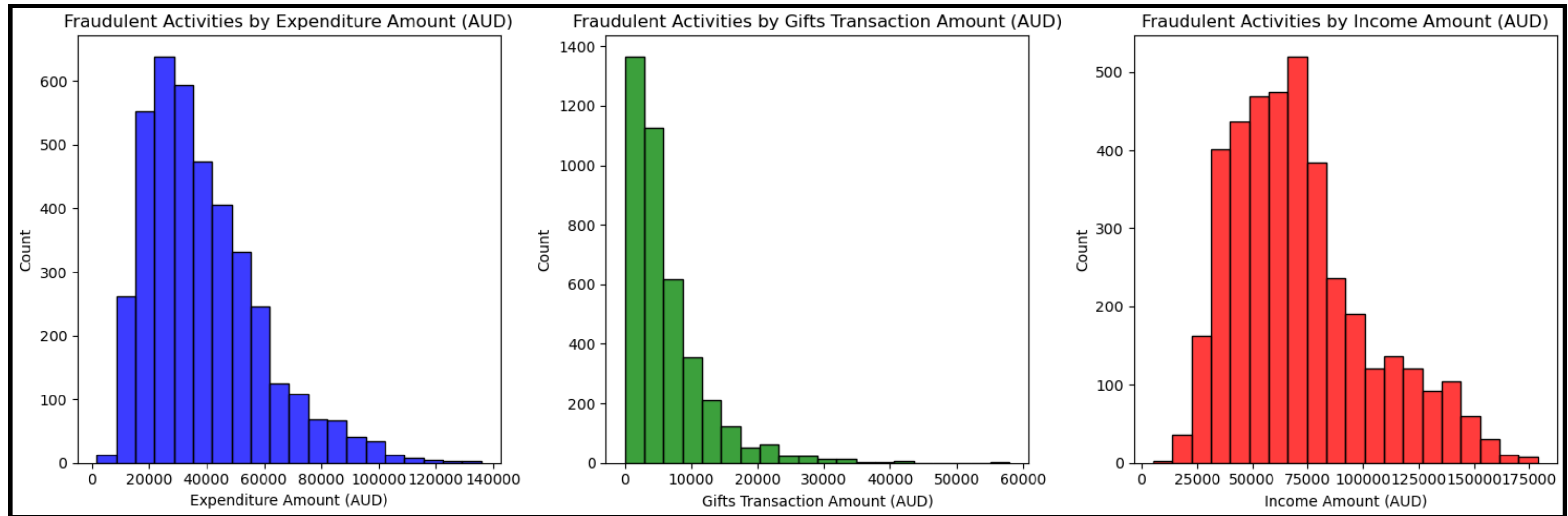
## 2.4 TRANSACTION AMOUNT AND CURRENCY-RELATED INFORMATION VISUALIZATION



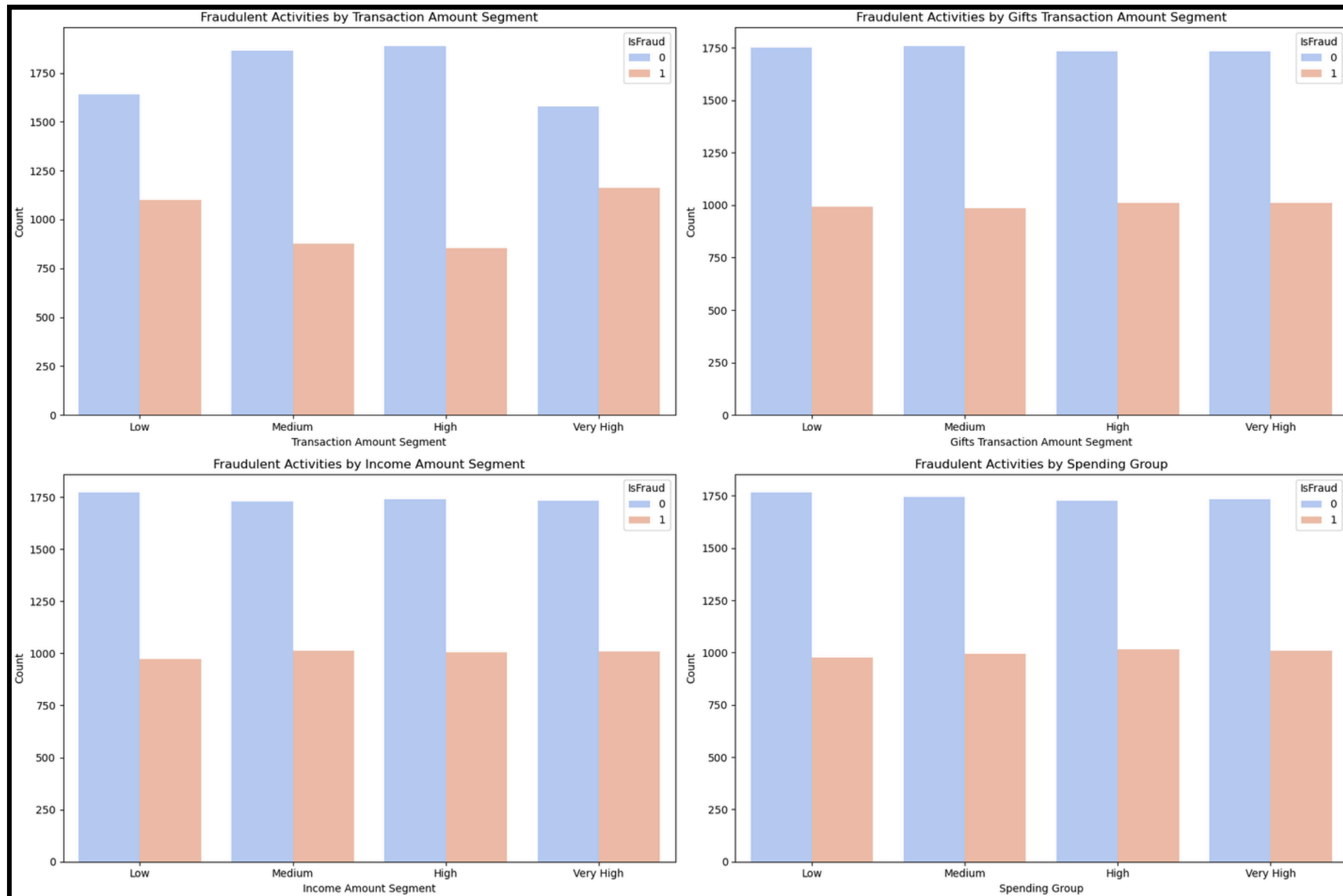
# 3.1 CURRENCY-RELATED INFORMATION VISUALIZATION



# 3.2 AMOUNT-RELATED INFORMATION VISUALIZATION



# 3.3 SEGMENTING INTO GROUPS



# 4. OTHER BEHAVIORAL/EXTERNAL INFORMATION

