# Transactional And Behavioral Patterns In Credit Card Fraudulent Victimization

Max Wong and Team
Spring 2025

01

## RESEARCH MOTIVATION

Why Fraud Detection Matters?

# Why Fraud Detection Matters?

1 Digital payment systems increase vulnerability to cyber threats

2 Fraud affects billions globally

3 Companies need better fraud detection tools

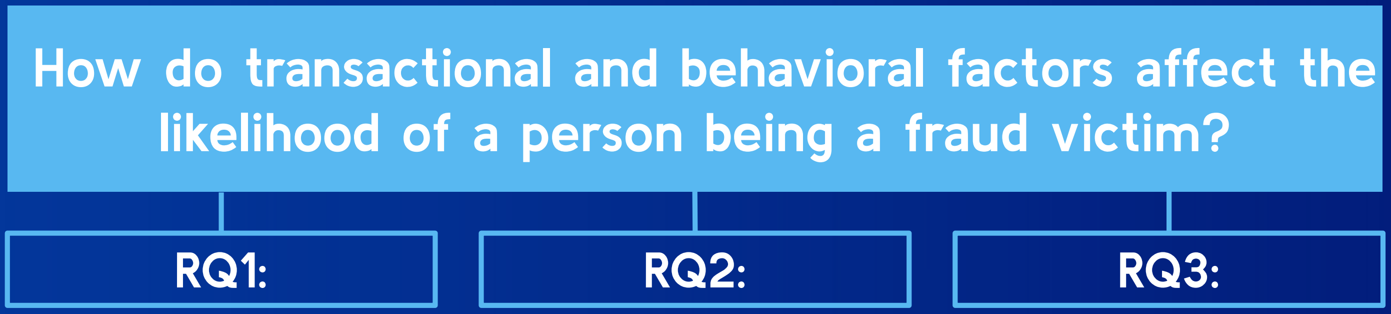
4 Machine learning offers a scalable, data-driven solution

02

# RESEARCH QUESTIONS

What Are We Trying to Study?

# What Are We Trying to Study?



How do transactional and behavioral factors affect the likelihood of a person being a fraud victim?

RQ1:

RQ2:

RQ3:

Would certain **types of merchants** have a higher possibility of fraudulent transactions?

Would merchants in more **populated cities** be more likely to be fraudulent compared to those in less populated cities?

Would the **transaction hour** of the day affect the probability of fraud in digital transactions?

03

# DATA OVERVIEW

Our Dataset

## 1.8 Million

Credit card transactions
(2019–2020)

## 23 Features

Merchant, time, location,
cardholder demographics, etc.

## 693

Different Merchants

## 999

Credit Card Holders

# 04

# DATA PREPARATION

Preparing the Data

# Preparing the Data

1

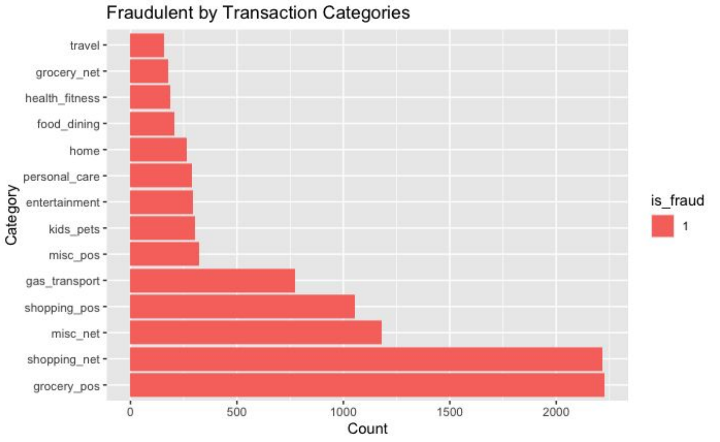
Cleaned and merged 2 datasets in R

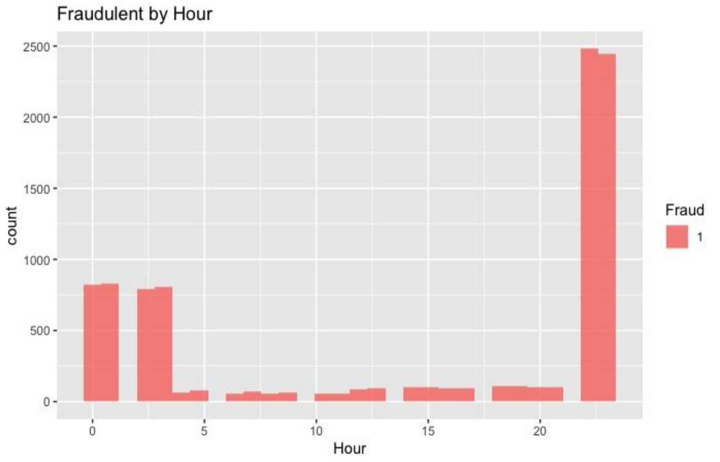
2

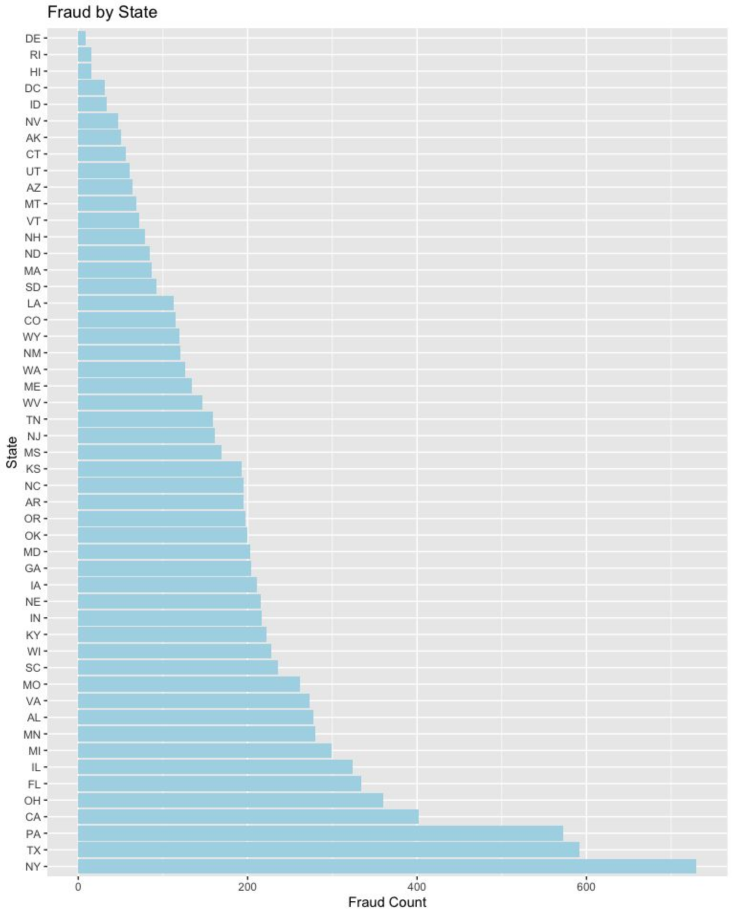
Imputed missing values (using mice)

3

Converted formats, removed duplicates, fixed zip codes







05

# METHODS

Clustering + Prediction Models

## Clustering + Predictive Models

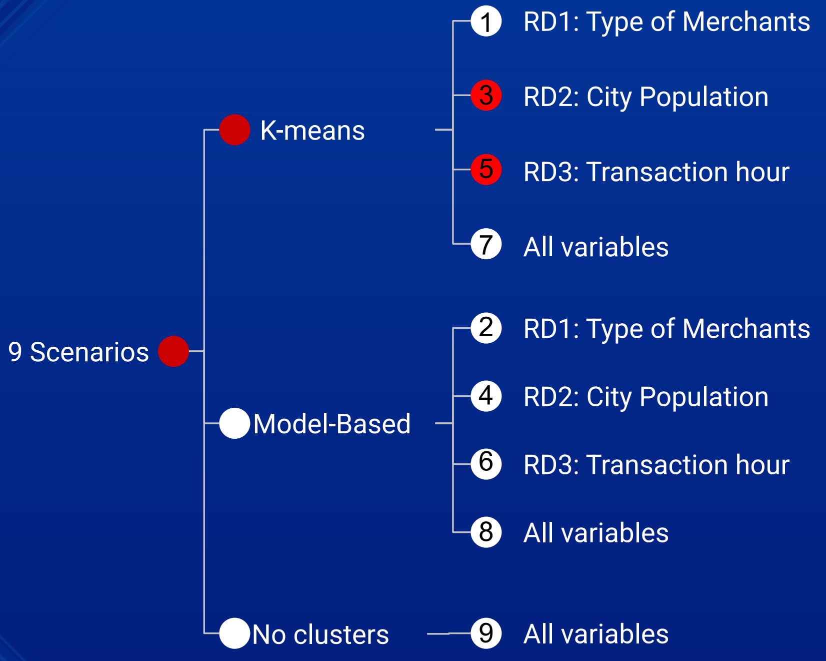
| Clustering | Scenarios | Predictive Models |
|---|---|---|
| K-means | 8 clusters +<br>1 baseline<br>(no clustering)<br>= 9 total scenarios | Logistic Regression |
| | | Random Forest |
| Model-based | | Neural Network<br>(1 hidden layers) |
| | | Deep Learning<br>(2 hidden layers) |

06

# RESULT & ANALYSIS

Can Clusters Help Isolate Fraud?

# Clustering Results



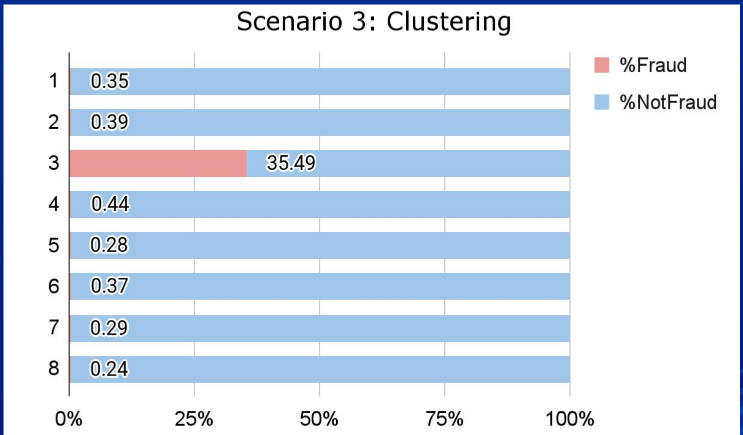
9 Scenarios

K-means
- 1 RD1: Type of Merchants
- 3 RD2: City Population
- 5 RD3: Transaction hour
- 7 All variables

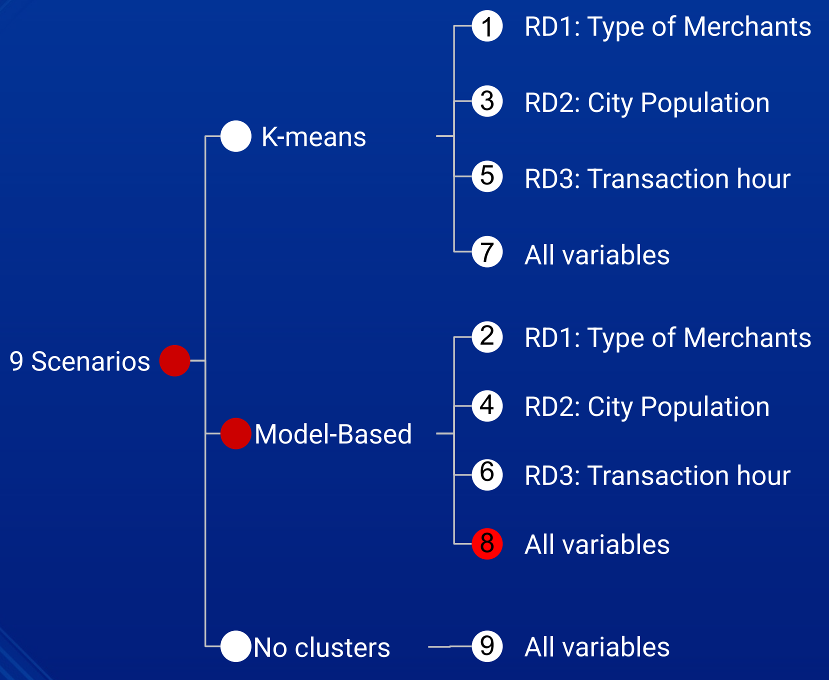
Model-Based
- 2 RD1: Type of Merchants
- 4 RD2: City Population
- 6 RD3: Transaction hour
- 8 All variables

No clusters
- 9 All variables

- Scenario 3 and 5 showed one cluster with >35% fraud, while others were <1%.
- With transaction hour or population can isolate fraud-prone groups effectively
- Clustering helps expose hidden fraud patterns


Scenario 3: Clustering

# Prediction Results



9 Scenarios

K-means

1 RD1: Type of Merchants

3 RD2: City Population

5 RD3: Transaction hour

7 All variables

Model-Based

2 RD1: Type of Merchants

4 RD2: City Population

6 RD3: Transaction hour

8 All variables

No clusters

9 All variables

- Clustering + predictive modeling significantly improves fraud detection
  - Scenario 7 & 8 have better results than Scenario 9
- Scenario 8 (Best detection power & business impact)
  - Model-Based Clustering + Random Forest
  - 0.9830 sensitivity
  - 0.1734 precision
  - Lowest loss ($1638)

# 07

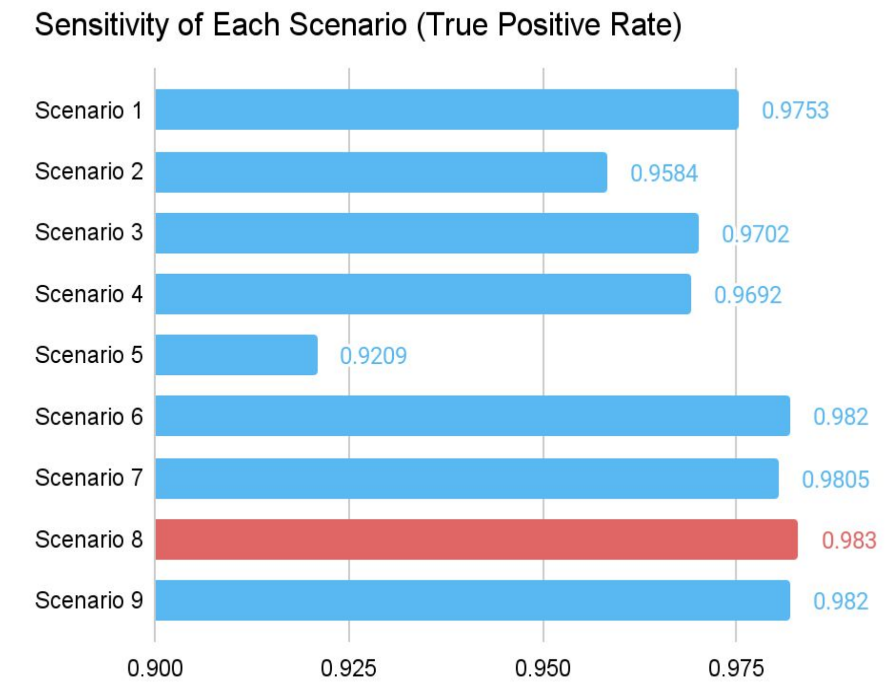
# CONCLUSIONS & RECOMMENDATIONS

Which Model Performs Best?

# Best Scenario for Business Use Case

- Goal
  - Minimize cost of loss & strengthen customer trust
  - Need high sensitivity & low false negative rate

- Scenario 8
  - All key features
  - Model-based clustering
  - Random Forest
  - ~98% sensitivity
  - ~17% precision
  - Loss amount: $1638



# Recommendations and Limitations

- Recommendations
  - A combination of
    - High-sensitivity models
    - 2-Factor Authentication (2FA)
  - Invest in continuous model updates and feature engineering

- Limitations
  - Computation power constraints
    - No hyperparameter tuning for models
  - May not reflect best performance of models

# THANK YOU!