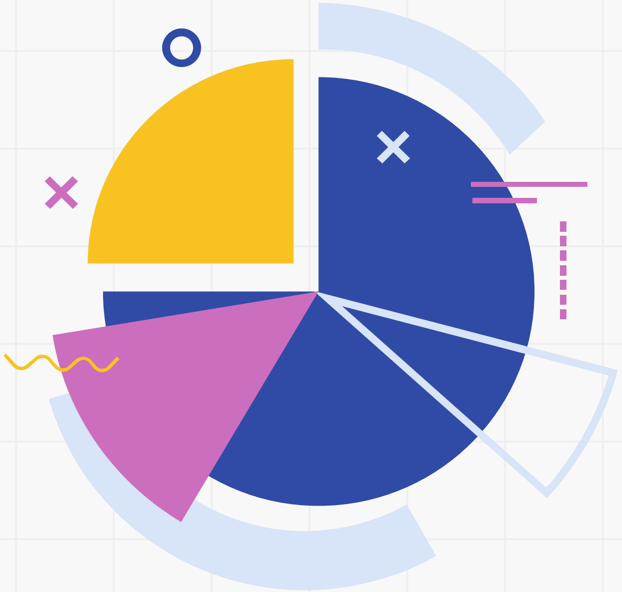
Transaction Fraud Detection



Malak Abdalla, Amy Cai, Tracy He, Charlotte Wan, Cynthia Xu

Introduction

Transaction Fraud

- Financial crime that involves using someone's financial information to make unauthorized transactions.
- Can involve credit card fraud, account takeovers, and more.

Why does it matter?

- Fraud affects millions globally — both individuals and businesses.
- Detection is crucial for maintaining security and trust in digital transactions.

It is challenging to detect fraud

- Fraudsters constantly evolve tactics, making detection difficult.
- Fraud patterns often resemble legitimate transactions.

Data & Variables

A Credit Card Transactions Dataset provides detailed records of credit card transactions, including information about transaction times, amounts, and associated personal and merchant details. This dataset has over 1.85M rows.

Variables:

- Timestamp
- Credit card number
- Merchant information
- Transaction information
- Cardholder information
- Population of the city where the transaction occurred

Methods

- Logistic Regression
- Random Forest Classification
- k-NN Classification
- Gradient Boosting Classification

Reducing the Size & Optimizing Result

```
Dataset size after sampling: (100000, 24)
```

```
Target variable distribution after undersampling:
is_fraud
0     579
1     579
```

```python
models = {
    'Logistic Regression': LogisticRegression(max_iter=1000, solver='liblinear'),
    'Random Forest': RandomForestClassifier(n_estimators=100, random_state=42),
    'k-NN': KNeighborsClassifier(n_neighbors=5),
    'Gradient Boosting': GradientBoostingClassifier(n_estimators=100, random_state=42)
}
```

Logistic Regression

Training model: Logistic Regression
Classification Report:

	precision	recall	f1-score	support
0	0.79	0.97	0.87	87
1	0.96	0.75	0.84	87
accuracy			0.86	174
macro avg	0.87	0.86	0.85	174
weighted avg	0.87	0.86	0.85	174

ROC AUC: 0.8614

Random Forest

Training model: Random Forest
Classification Report:

	precision	recall	f1-score	support
0	0.88	0.92	0.90	87
1	0.92	0.87	0.89	87
accuracy			0.90	174
macro avg	0.90	0.90	0.90	174
weighted avg	0.90	0.90	0.90	174

ROC AUC: 0.9719

k-NN

Training model: k-NN
Classification Report:

	precision	recall	f1-score	support
0	0.45	0.46	0.45	87
1	0.45	0.44	0.44	87
accuracy			0.45	174
macro avg	0.45	0.45	0.45	174
weighted avg	0.45	0.45	0.45	174

ROC AUC: 0.4570

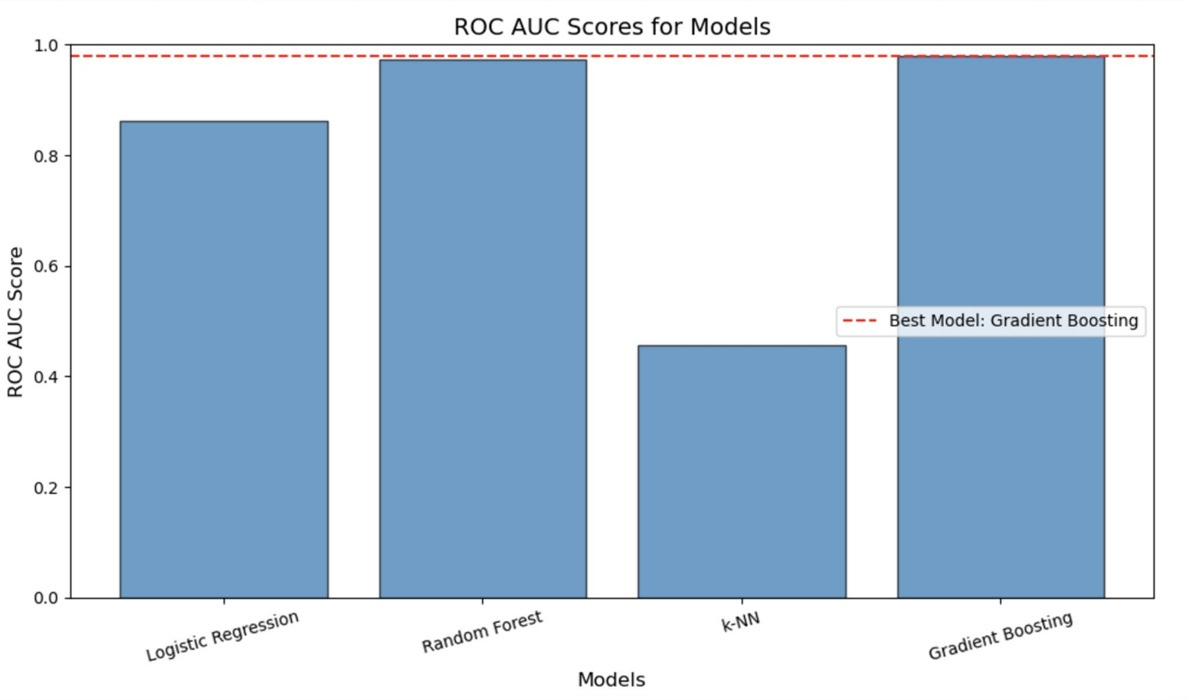
Gradient Boosting

Training model: Gradient Boosting
Classification Report:

	precision	recall	f1-score	support
0	0.93	0.90	0.91	87
1	0.90	0.93	0.92	87
accuracy			0.91	174
macro avg	0.91	0.91	0.91	174
weighted avg	0.91	0.91	0.91	174

ROC AUC: 0.9804

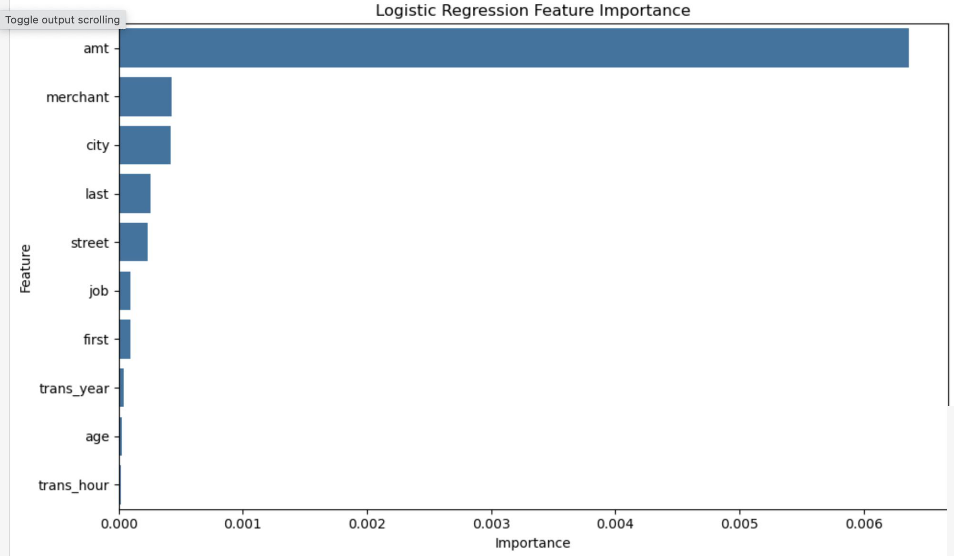
Which model is best?



Gradient Boosting is the most effective model for detecting fraud.

Logistic Regression

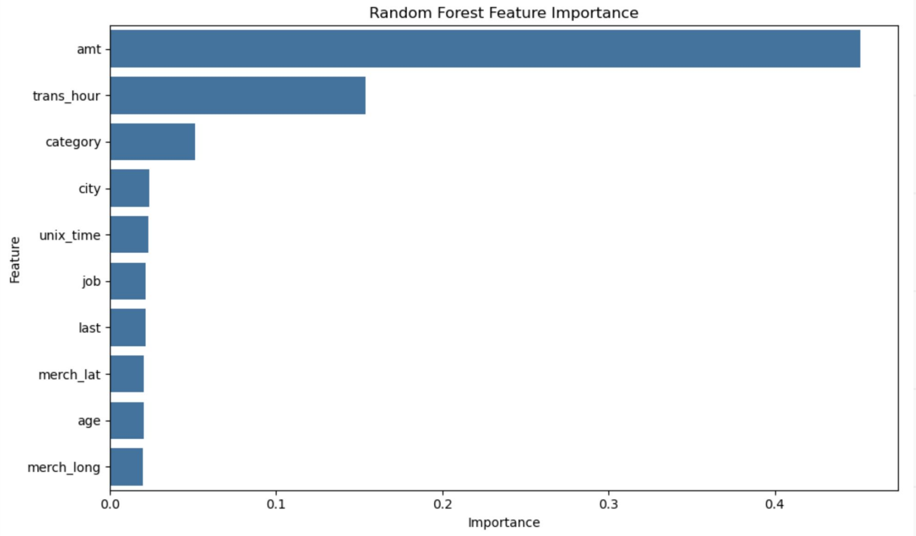
Toggle output scrolling



Feature Importance Ranking for Model: Logistic Regression

Feature	Importance
amt	0.006362
merchant	0.000424
city	0.000420
last	0.000255
street	0.000233
job	0.000093
first	0.000093
trans_year	0.000041
age	0.000021
trans_hour	0.000019

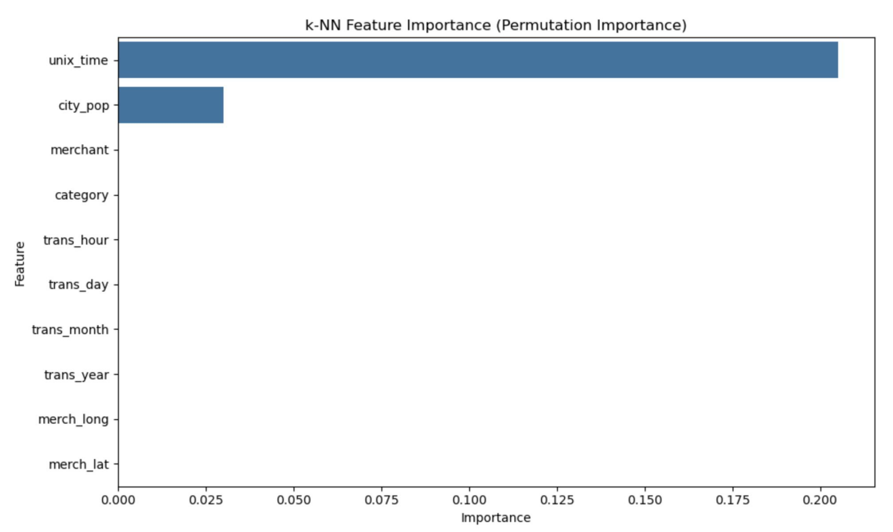
Random Forest



Feature Importance Ranking for Model: Random Forest

Feature	Importance
amt	0.451584
trans_hour	0.153767
category	0.051320
city	0.023744
unix_time	0.022910
job	0.021600
last	0.021426
merch_lat	0.020353
age	0.020165
merch_long	0.019677

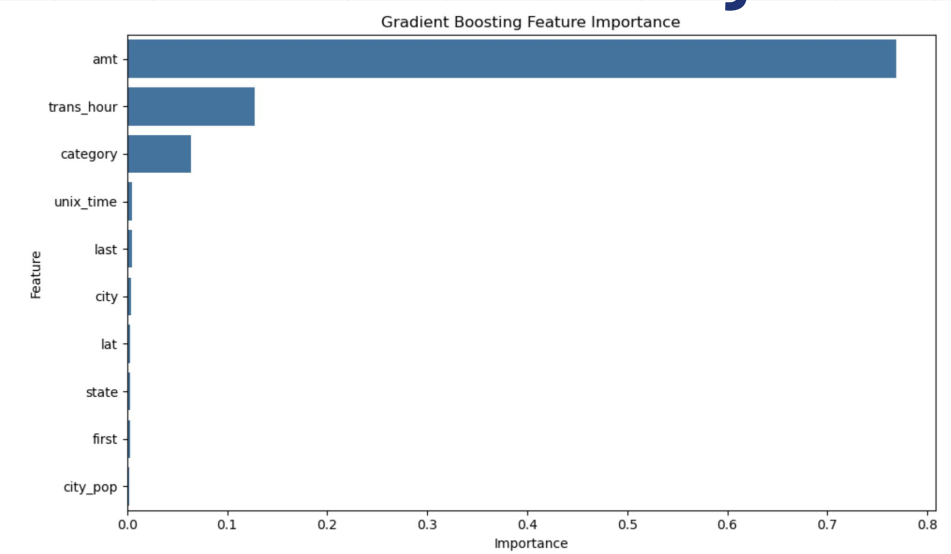
k-NN



Feature Importance Ranking for k-NN:

Feature	Importance
unix_time	0.205268
city_pop	0.030138
merchant	0.000000
category	0.000000
trans_hour	0.000000
trans_day	0.000000
trans_month	0.000000
trans_year	0.000000
merch_long	0.000000
merch_lat	0.000000

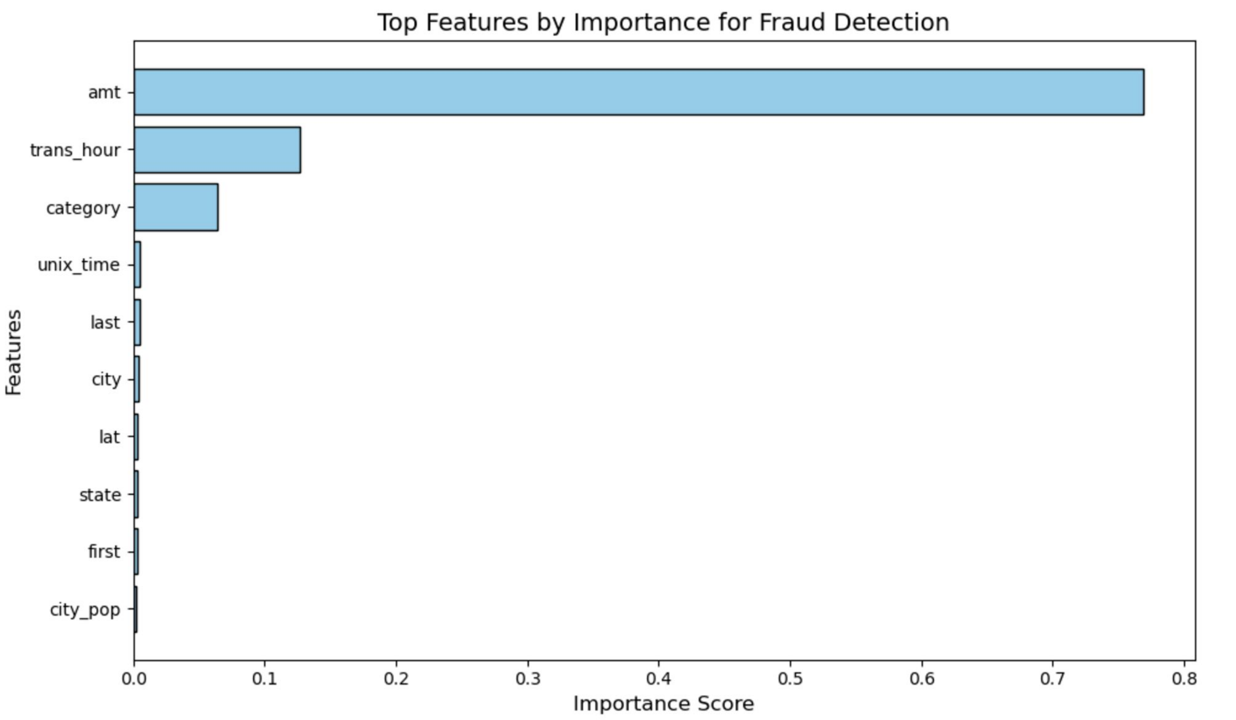
Gradient Boosting



Feature Importance Ranking for Model: Gradient Boosting

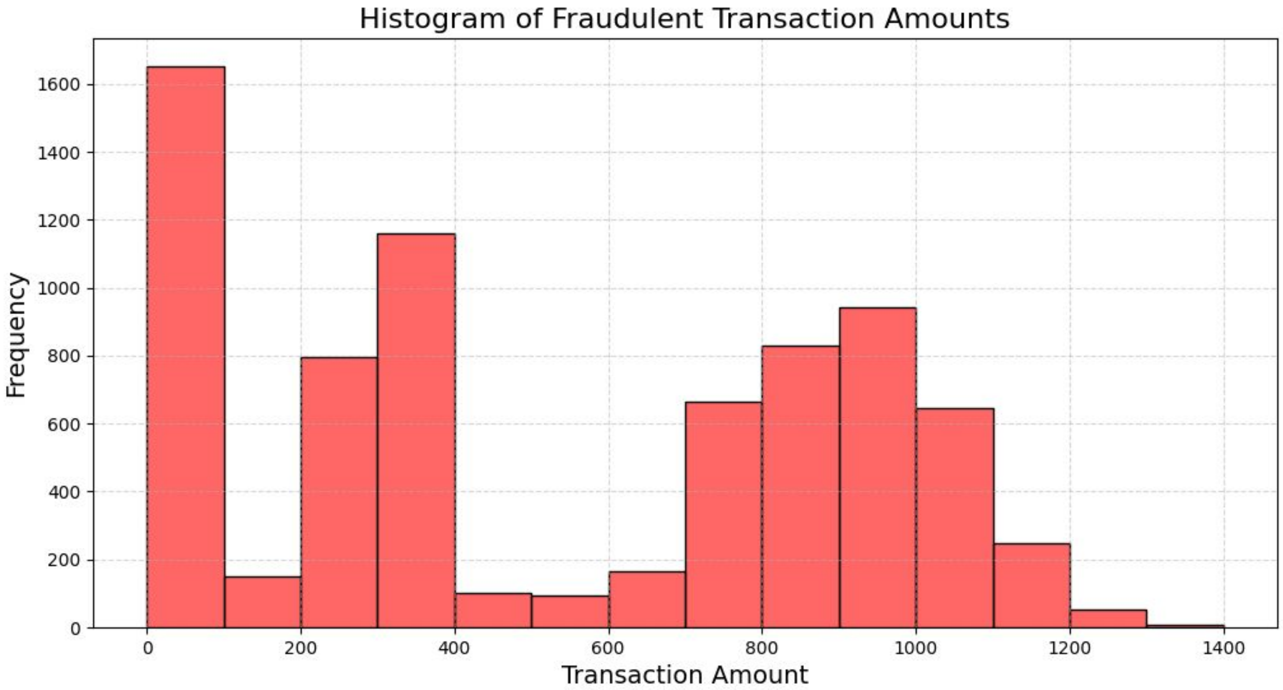
Feature	Importance
amt	0.769419
trans_hour	0.127209
category	0.063659
unix_time	0.005347
last	0.005208
city	0.003953
lat	0.003284
state	0.002935
first	0.002574
city_pop	0.002507

What feature is the most important?

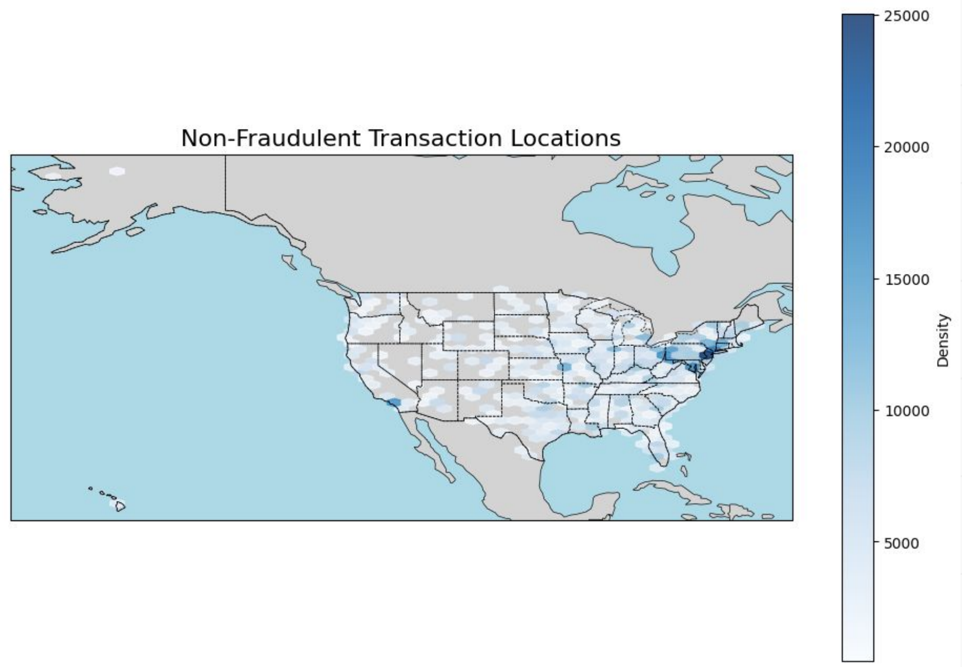


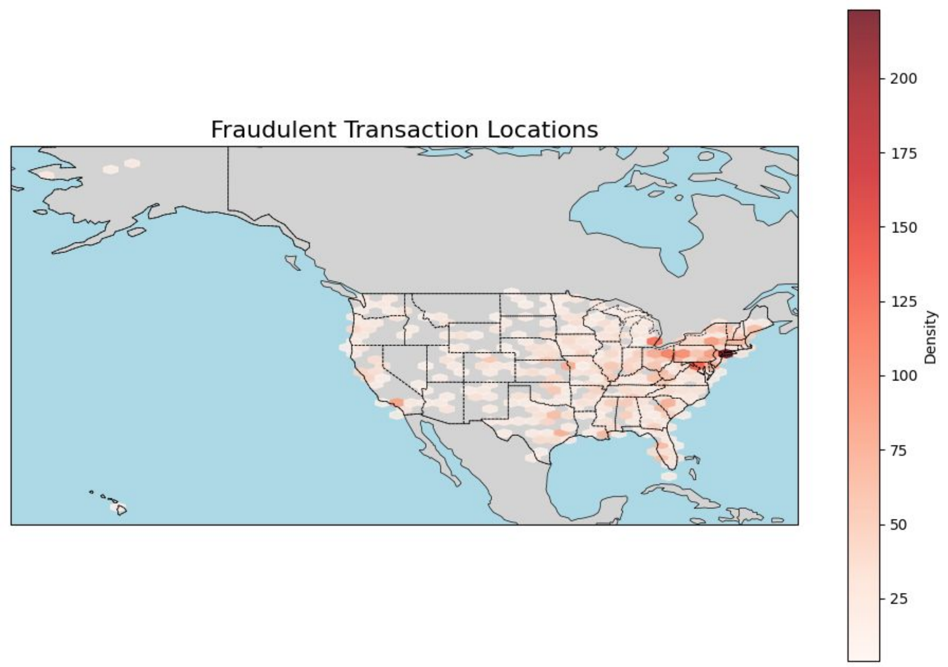
The variable 'amt' (Transaction amount) has the greatest impact on transaction fraud, with an importance score of 0.7694.（Gradient Boosting）

Typical fraudulent transaction amounts?

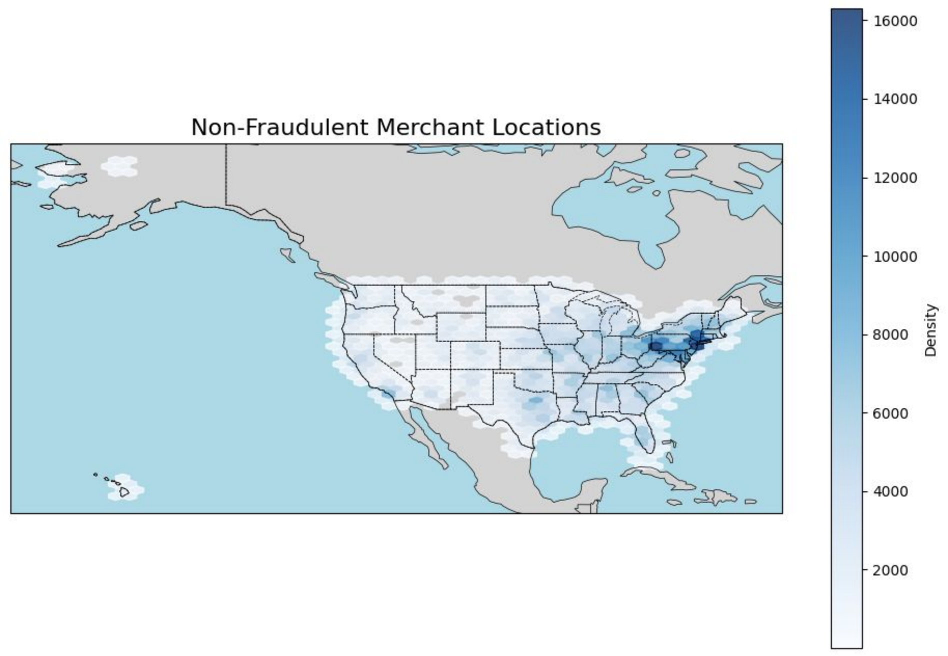


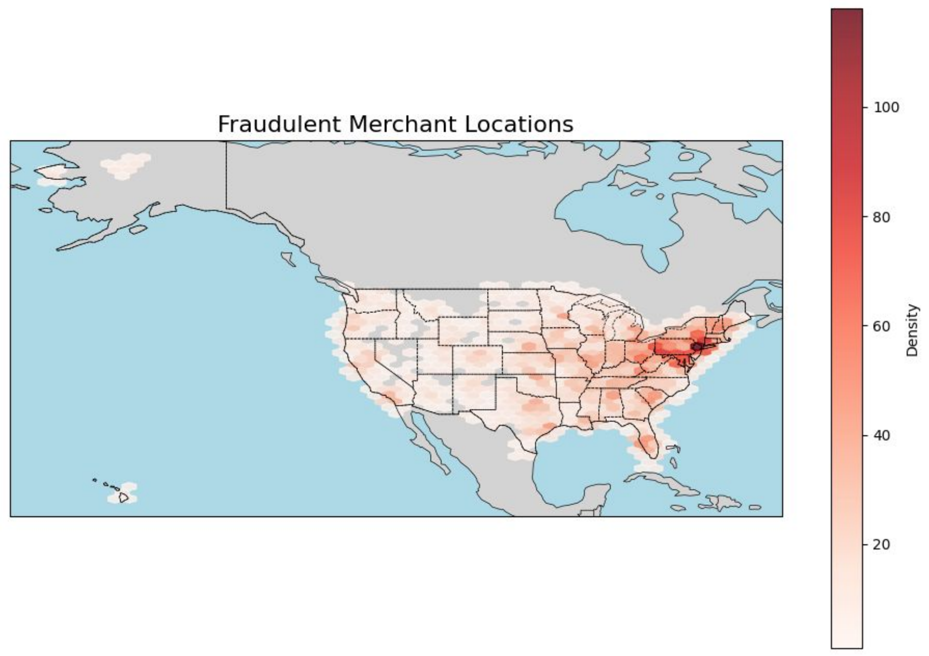
Where are the transaction locations?



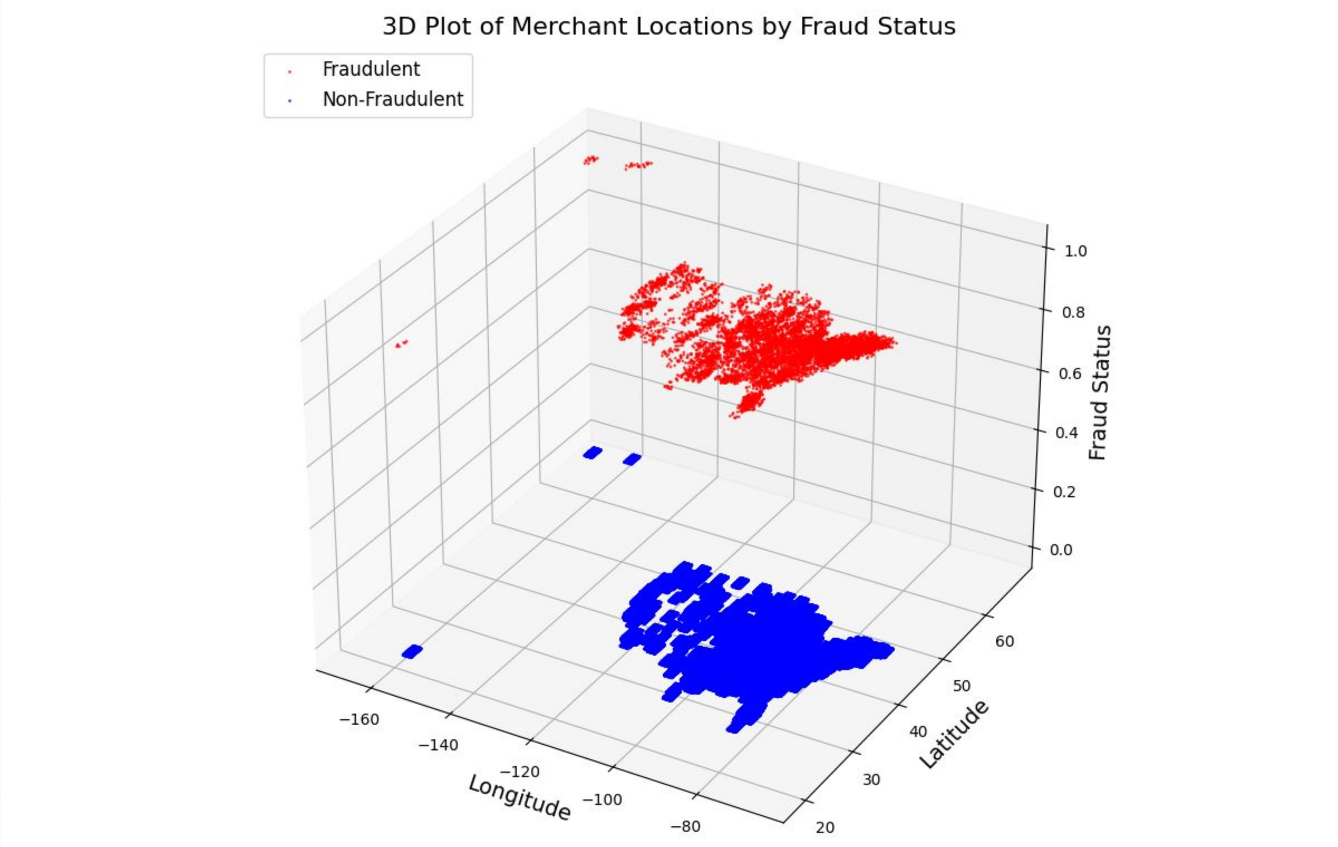


Where are the merchant locations?

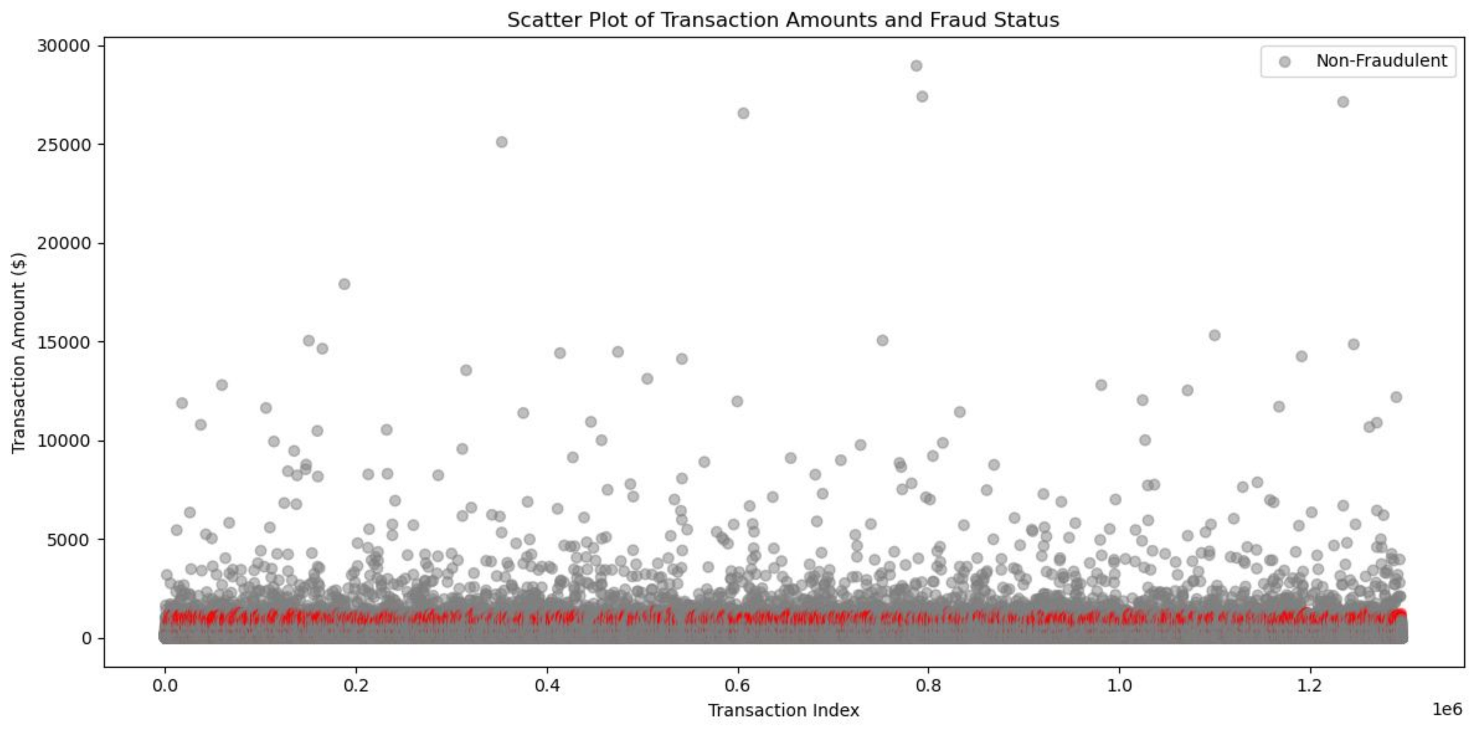




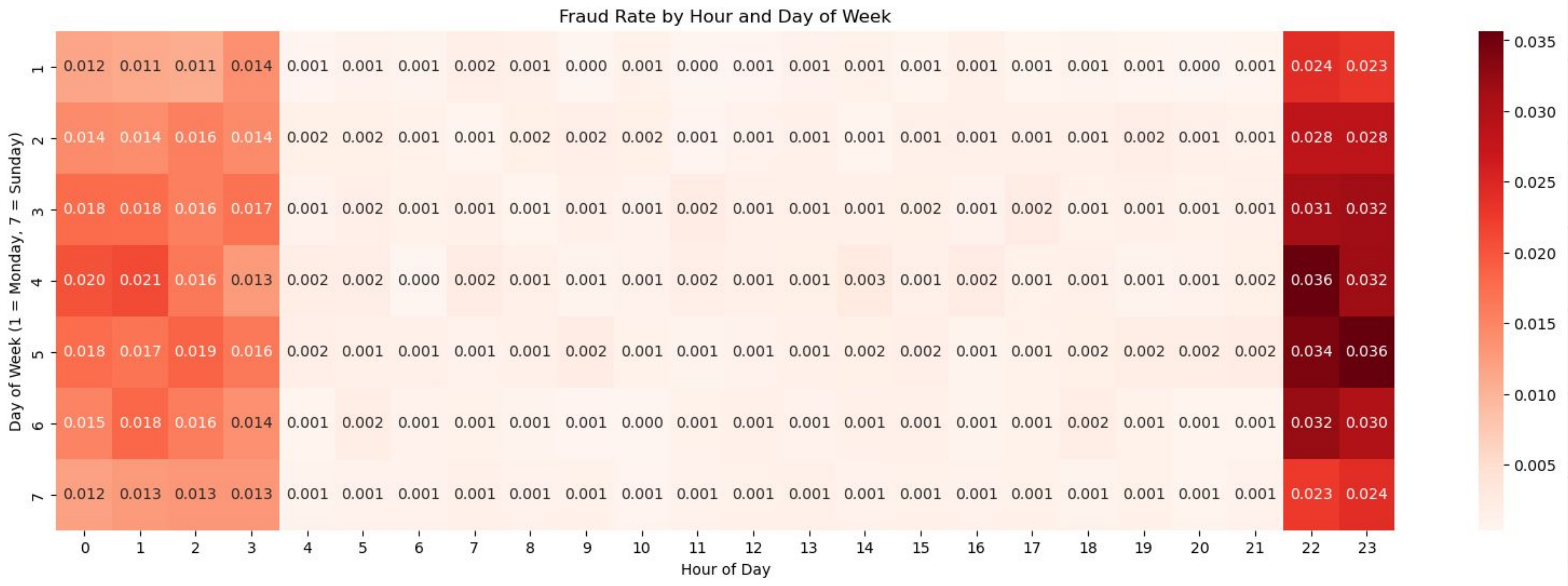
Merchant Location by Fraud Status



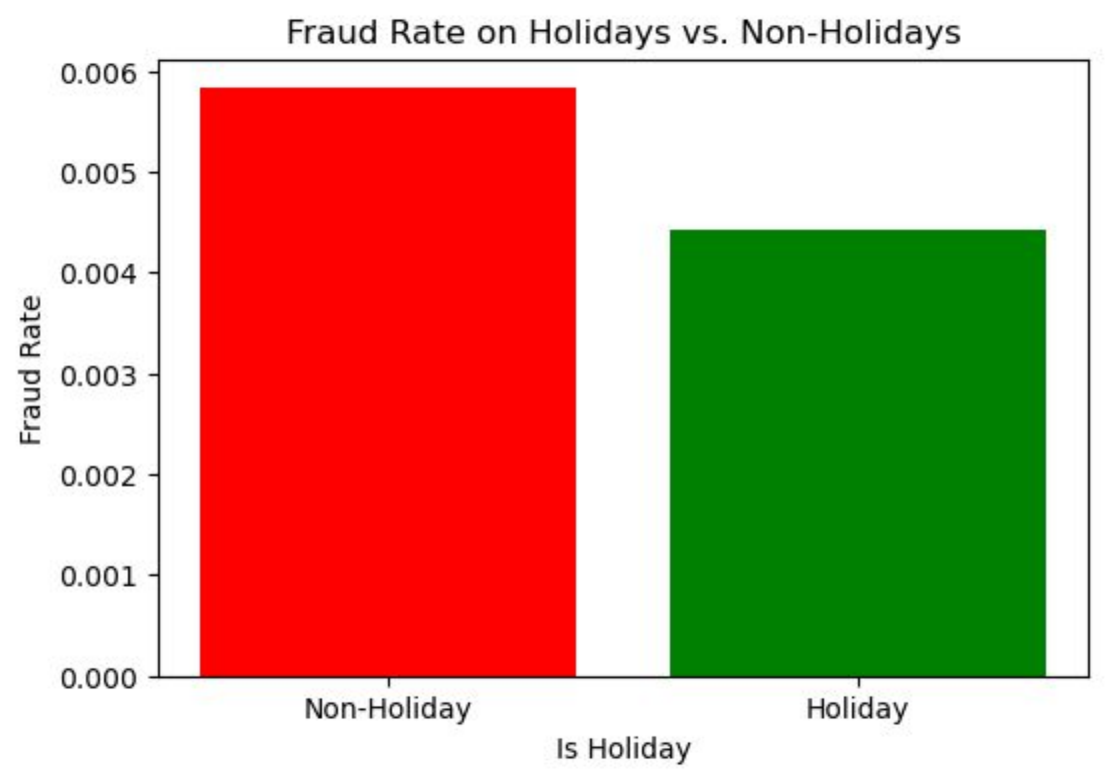
Amount vs. Fraud



Fraud by Time and Day



Fraud by Holiday and Non-Holiday



Q&A

Thank you!