# Bank Card Fraud Detection Analysis

Zhemin Xie(Ayrie)

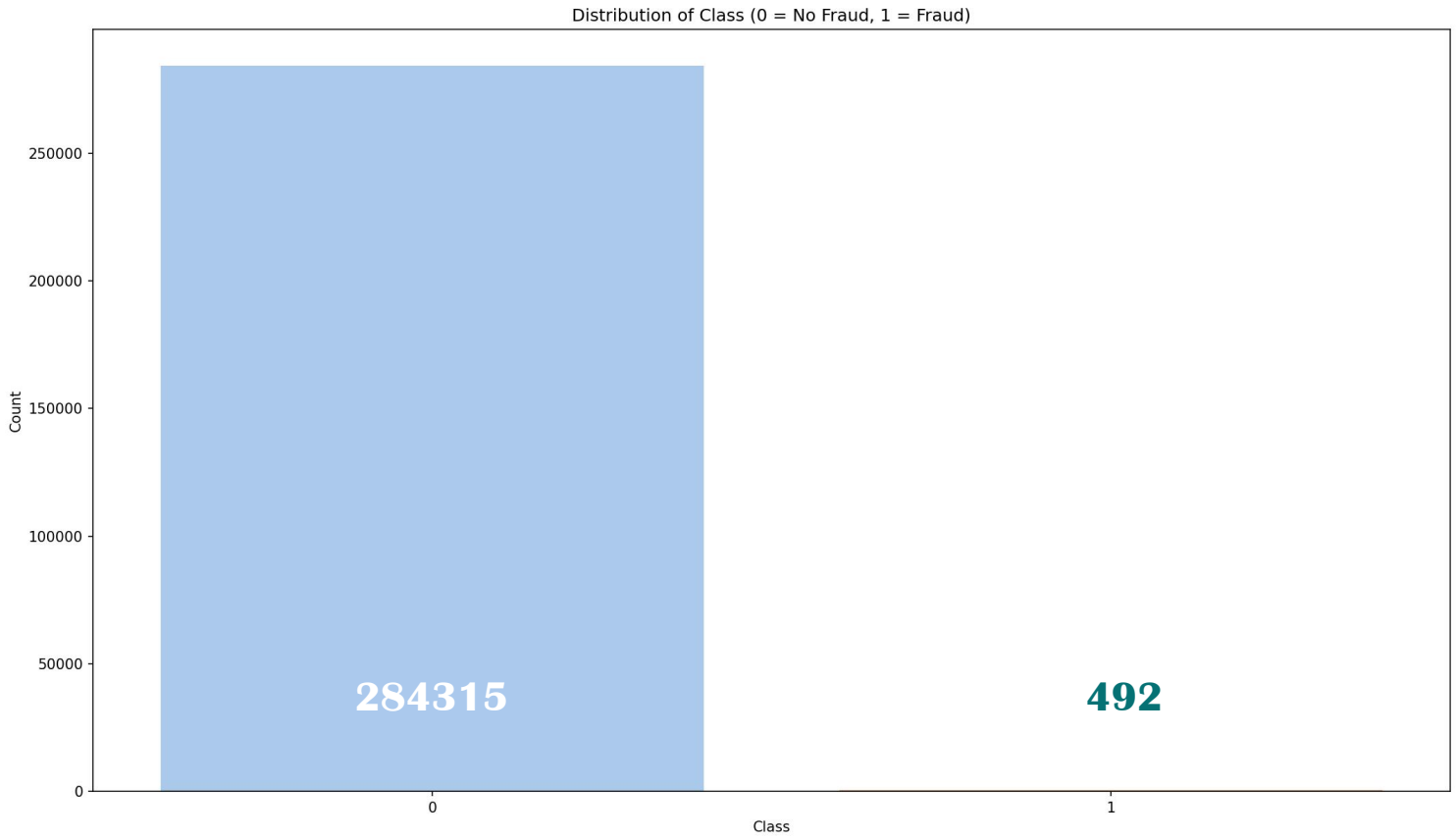
# 01 Dataset Description

| | Time | V1 | ... | Amount | Class |
|---|---|---|---|---|---|
| count | 284807.000000 | 2.848070e+05 | ... | 284807.000000 | 284807.000000 |
| mean | 94813.859575 | 1.168375e-15 | ... | 88.349619 | 0.001727 |
| std | 47488.145955 | 1.958696e+00 | ... | 250.120109 | 0.041527 |
| min | 0.000000 | -5.640751e+01 | ... | 0.000000 | 0.000000 |
| 25% | 54201.500000 | -9.203734e-01 | ... | 5.600000 | 0.000000 |
| 50% | 84692.000000 | 1.810880e-02 | ... | 22.000000 | 0.000000 |
| 75% | 139320.500000 | 1.315642e+00 | ... | 77.165000 | 0.000000 |
| max | 172792.000000 | 2.454930e+00 | ... | 25691.160000 | 1.000000 |

[8 rows x 31 columns]

The dataset contains a total of **284,807** transaction records with a large sample size and no missing values, i.e. null values. All spurious or abnormal transactions are labeled as 1, while normal transactions are labeled as 0

# 02 Distribution Check

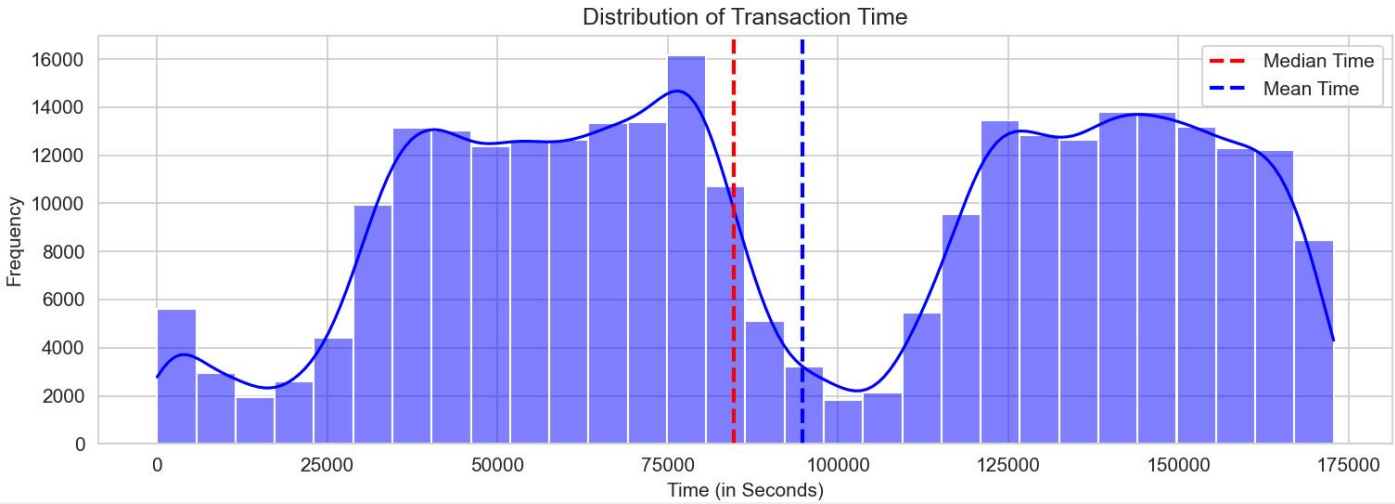



**Category imbalance:**

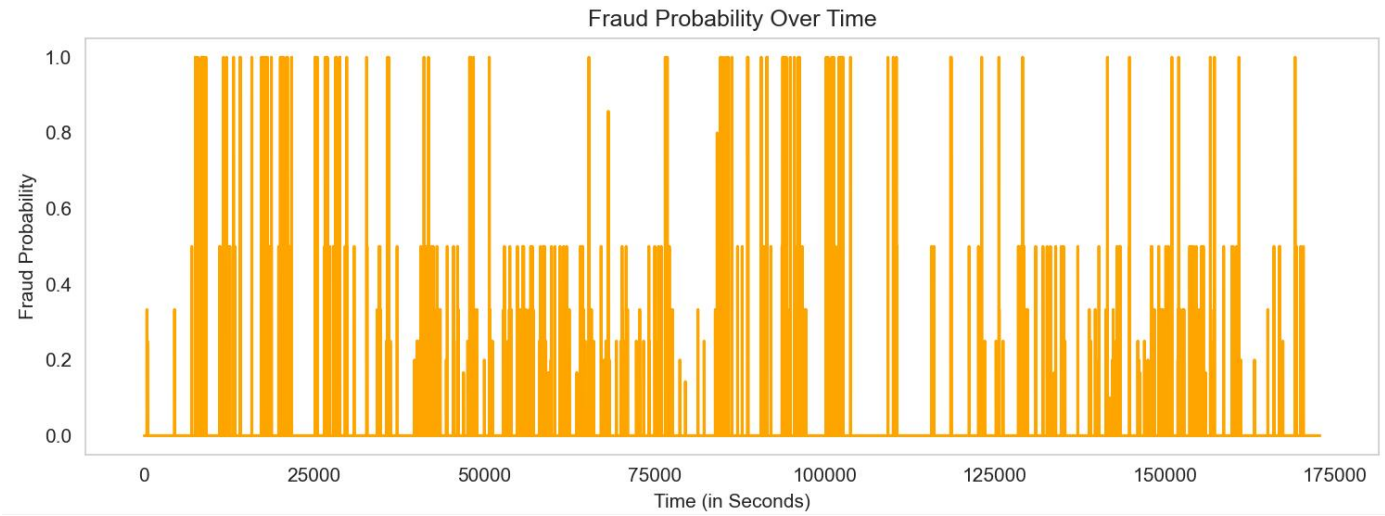
- Normal transactions (Class 0): **99.83%**.
- Fraudulent transactions (Class 1): **0.17%**.

The percentage of fraudulent transactions in the sample is extremely low and **is a highly unbalanced dataset.**
This imbalance in machine learning can lead to models that are biased towards predicting the majority of categories (normal transactions) while ignoring the minority (fraudulent transactions).
To address this problem, methods such as **oversampling, undersampling, and adjusting category weights** can be considered to improve the model's detection of fraudulent transactions

# 03 Time with Fraud Probability

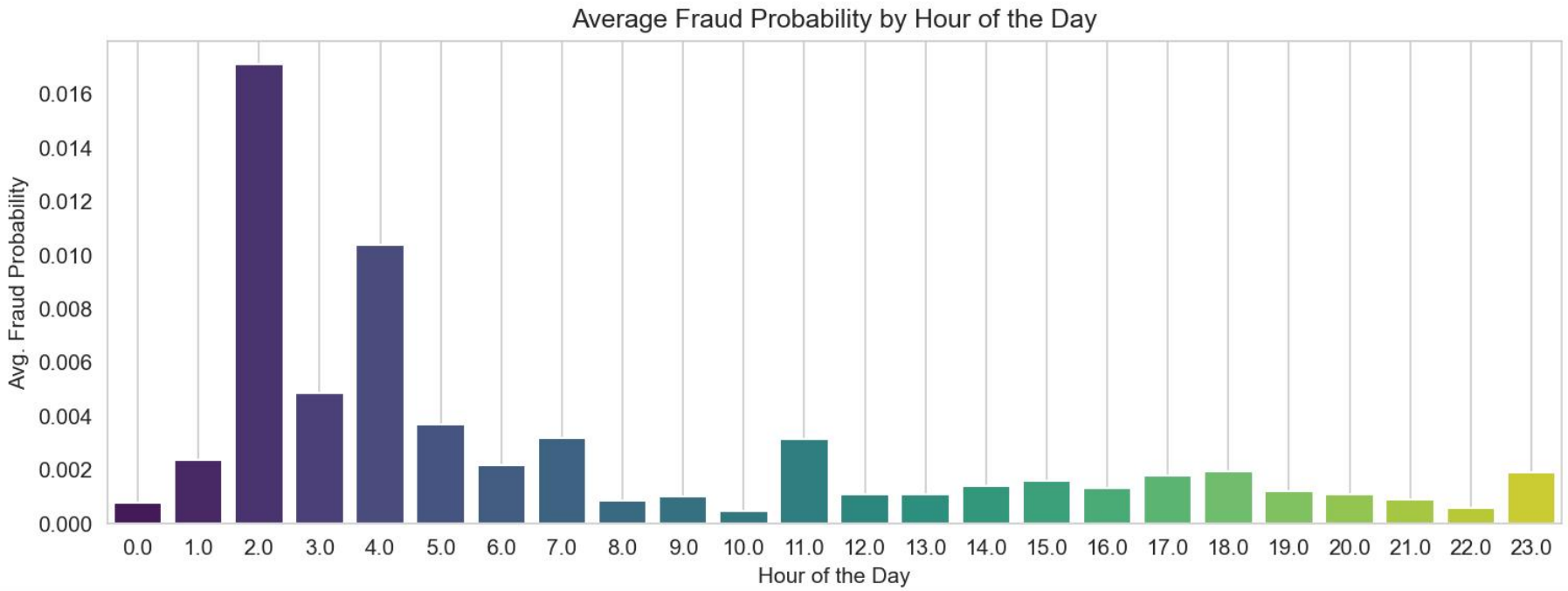



There are **significant peak hours** of trading in the sample, and there is a need to ensure that the system is stable during high volume hours. And in the low period, planned maintenance work can be arranged to avoid affecting customer business processing.




Certain time periods have a **significantly higher probability of fraud**, for which banks can add real-time monitoring and protective measures, such as **multi-factor authentication, transaction limits**, etc.

# 03.1 Time with Fraud Probability

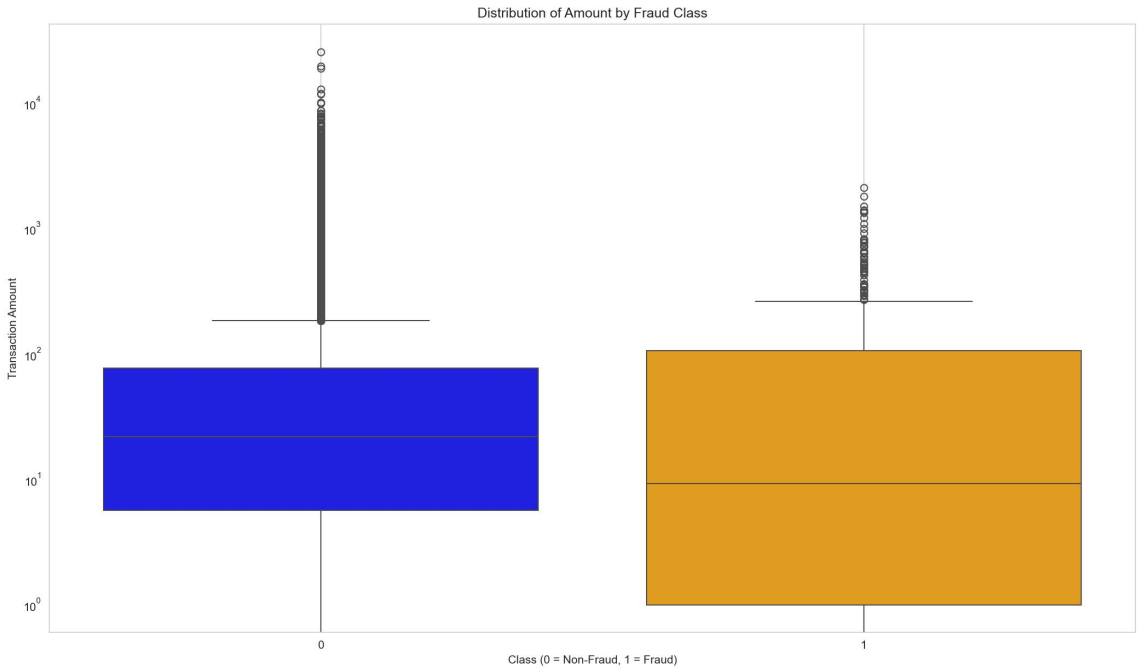


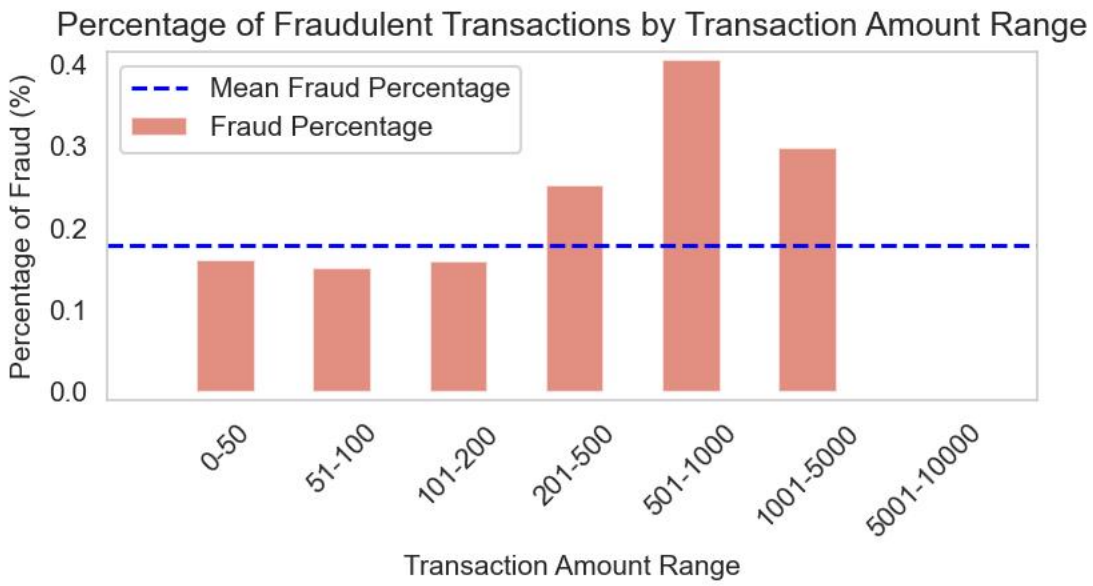

The high probability of fraud is usually concentrated **late at night (e.g., 2 a.m. to 4 a.m.)**
This may be due to:

- The nighttime hours are less regulated and there is less transactional activity, making it easier for fraudsters to commit fraud.
- Users may be in a state of fatigue late at night, increasing the risk of fraud.

Given that the probability of fraud is **significantly higher at 2 a.m**. than at other times of the day, banks should further investigate transactions during these unusually high-risk hours to identify potential fraud patterns or system vulnerabilities.

# 04 Fraud distribution with amount




- The distribution of amounts for non-fraudulent transactions is more concentrated (smaller boxes), with the majority of transactions in the lower range (lower median).
- Fraudulent transaction amounts are more widely distributed (larger boxes). And the median is significantly higher, indicating that fraudulent transactions tend to occur in transactions with larger amounts.

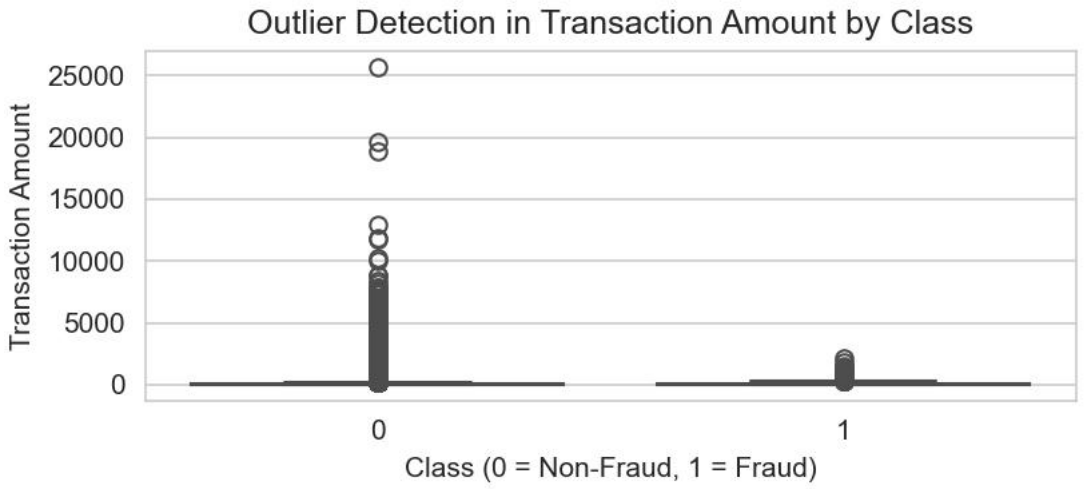
- Fraudsters are more likely to commit fraud in the mid-to-high dollar range (e.g., the 501-1000CNY, around 65-130 Euro ), potentially because such amounts are high enough to attract attention, but not so high as to immediately trigger a security alert.
- Fraud rates are lower in the small dollar range, perhaps because small-dollar transactions are not worth the risk that fraudsters put on them.

# 05 Detecting outliers in transaction amounts

The relatively small number of outliers in fraudulent transactions suggests that **fraudsters may choose a more consistent range of amounts to avoid drawing attention to themselves.**

Due to the concentration of fraudulent transaction amounts, it is recommended that **additional vigilant monitoring** be performed on transactions in these dollar ranges.

For non-fraudulent transactions, the high dollar amounts involved warrant further analysis to ensure that these outliers are not incorrectly flagged as fraudulent.



# Thank you for your attention

Zhemin Xie(Ayrie)