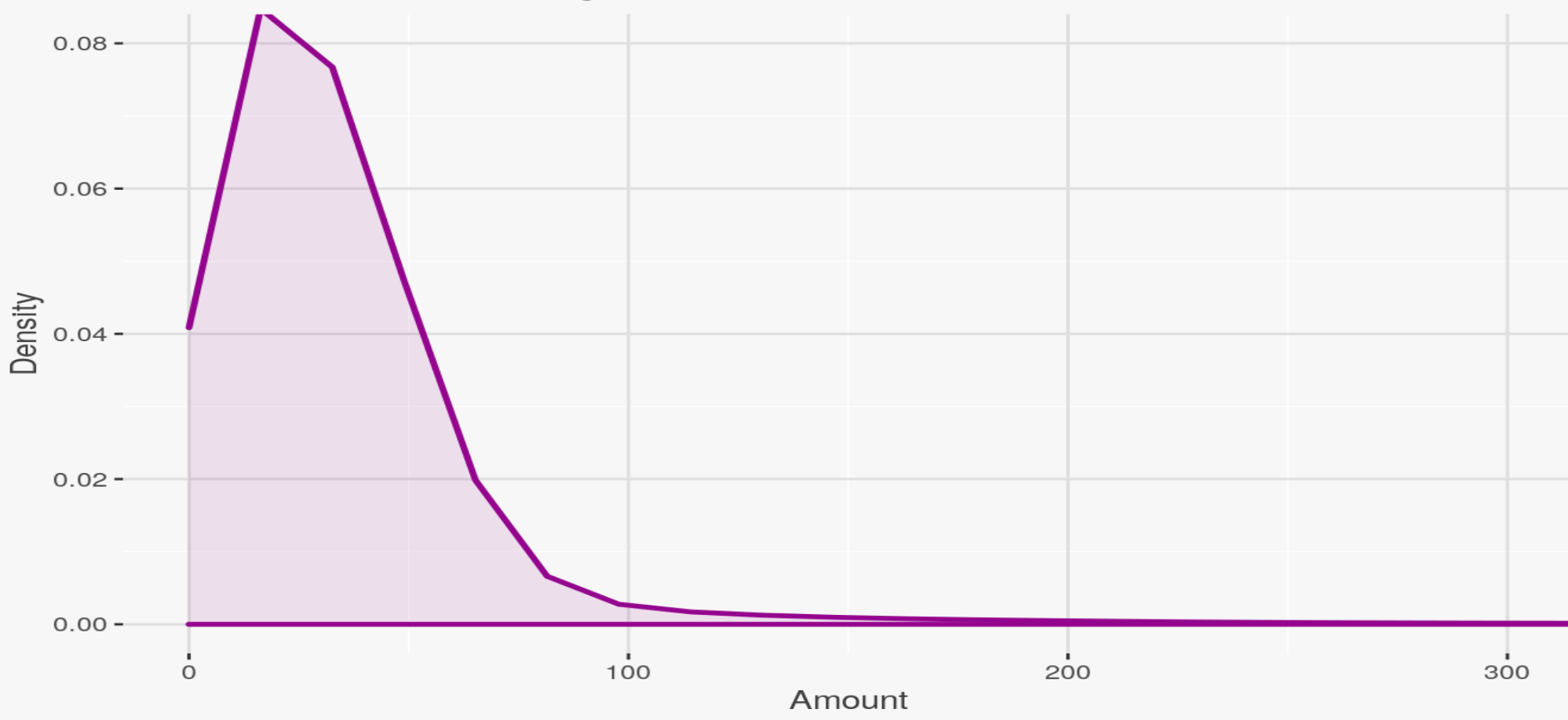


## Problem To be Solved

Due to rapid growth in field of cashless or digital transactions, credit cards are widely used in all around the world. Credit cards providers are issuing thousands of cards to their customers. Providers have to ensure all the credit card users should be genuine and real. Any mistake in issuing a card can be reason of financial crises. Due to rapid growth in cashless transaction, the chances of number of fraudulent transactions can also increasing. A Fraud transaction can be identified by analyzing various behaviors of credit card customers from previous transaction history datasets. If any deviation is noticed in spending behavior from available patterns, it is possibly of fraudulent transaction.

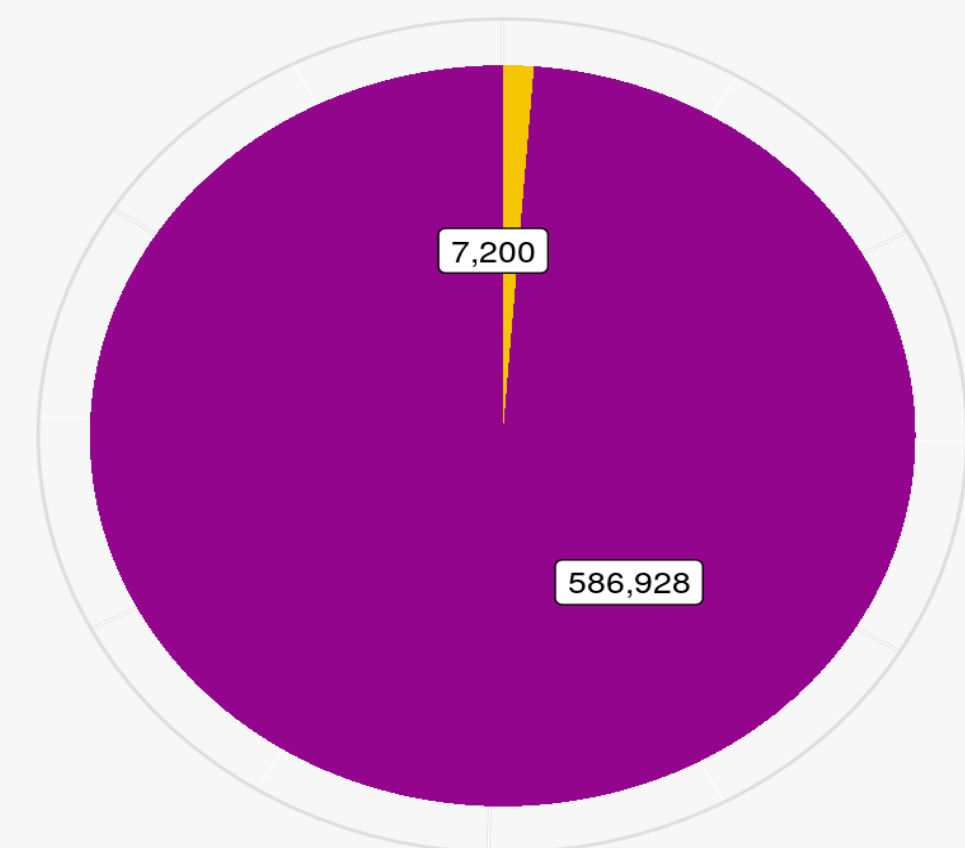
### Amount Distribution

extreme skewness to right - maximum value 8329



### Fraud Frequency

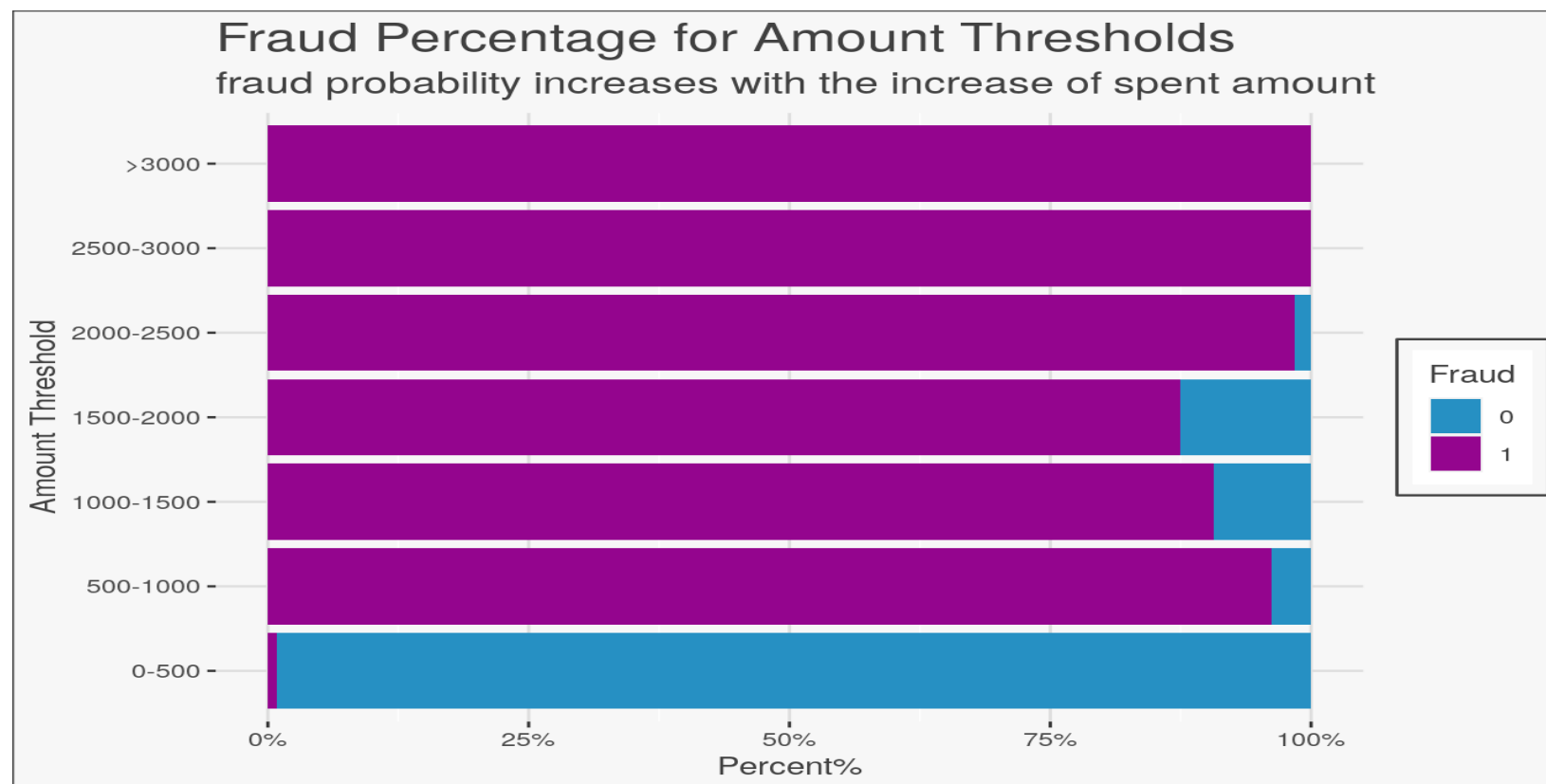
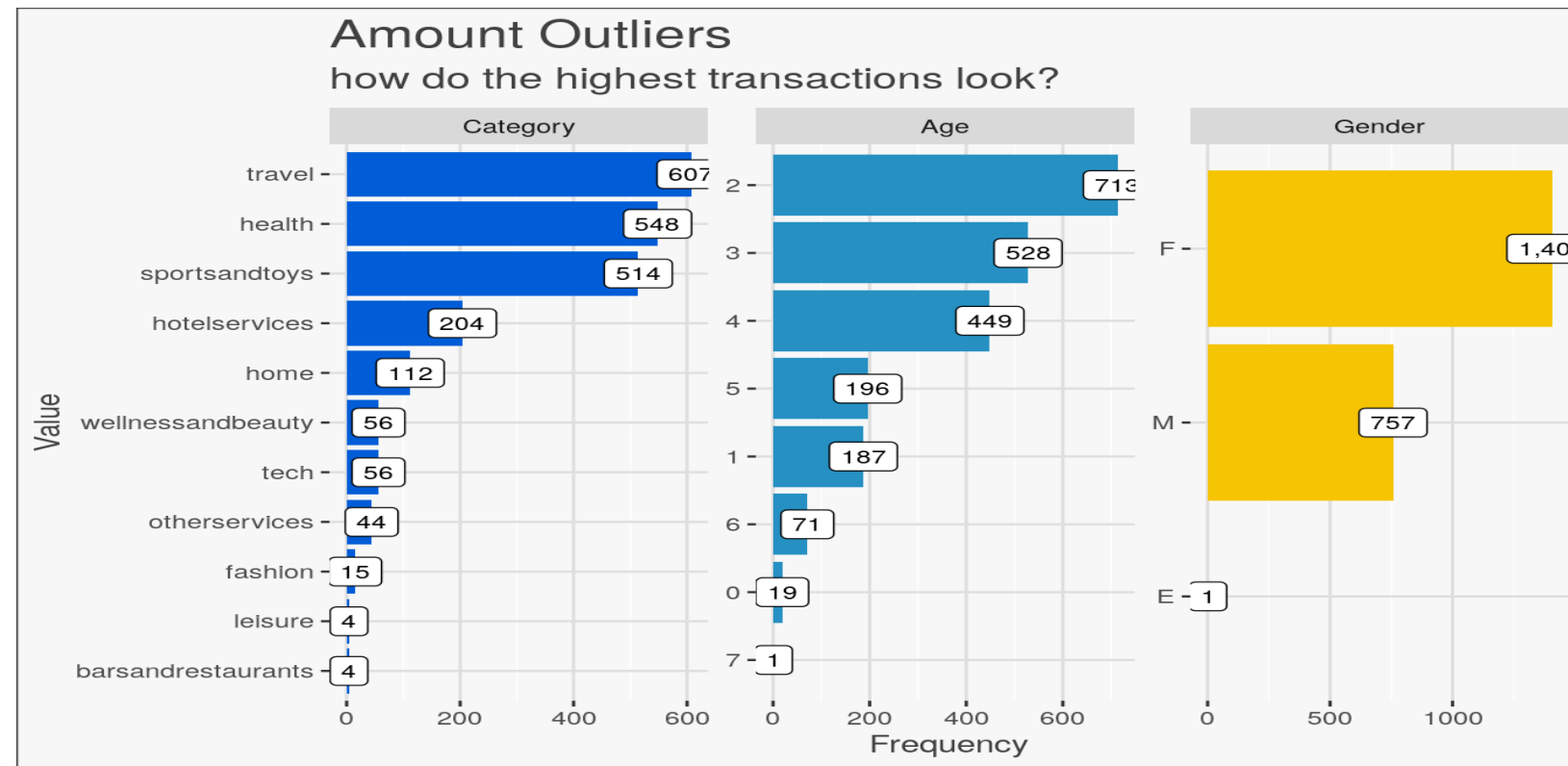
there are 1.2% fraud cases



The amount distribution of the transactions in the dataset is extremely skewed to the right, with more than 95% of the transactions having a value between 1 and 100 dollars, but with the rest reaching extremely high values, with the maximum being more than 8,000 dollars.

The percentage of fraudulent transactions is 1.2%, with only 7,160 available fraud cases compared to the non-fraudulent ones, which are 584,073.

# Factors Responsible



- There was further investigation made on how these extremely high transactions look: the majority are expenses in the travel, health, sports and toys categories, usually done by people between 18 and 45 years old. Also, there are more transactions made by females than by men between these outliers.
- The fraud percentage on different amount thresholds is very interesting as well. There are close to none fraudulent cases throughout transactions between 1 and \$500, but above this value, the percentage jumps to 90% or more chance of the transaction being fraud.

# Conclusion

With the increasing complexity of fraudulent activities and the massive amounts of data available, machine learning models can help identify and prevent fraud in real-time.

Through this project, we have demonstrated the effectiveness of machine learning algorithms in detecting fraud by building and evaluating different models. Our findings suggest that ensemble methods such as Random Forest and with feature selection techniques, are effective in detecting fraudulent transactions. We also found that data preprocessing techniques such as standardization and normalization are crucial in improving model performance.

Overall, this project highlights the potential of machine learning in detecting fraud and provides insights into how to optimize fraud detection models. With ongoing advancements in technology and data availability, we expect that machine learning will continue to play a critical role in fraud detection in the future.