

DETECTING FRAUDULENT CREDIT CARD TRANSACTIONS USING MACHINE LEARNING

by Tolu Oludemi



PROJECT IMPORTANCE



\$3.3B

Lost to fraud in 2021
in the United States

\$800M

Lost to fraud in 2021
in Canada

2.2M

Cases of fraud
were reported in
the United States

39%

Canadians were scared of
using their credit card due to
fear of their account being
compromised.

Introduction

PROBLEM STATEMENT:

Using machine learning on credit card transactions, can I create a model that predicts fraudulent transactions?



VALUE ADD:

In order to allow financial institutions to provide protection to their customers' finances, improving their service, and overall business growth.



Data Collection and Description

- Simulated credit card transaction dataset
- Generated using a Sparkov Data Generation tool on Github
- Contains transactions from January 2019 to December 2020
- 1,000 customers with a pool of 800 merchants
- Binary Classification (Supervised Learning)

Target



trans_date_trans_time	cc_num	merchant	category	amt	first	last	gender	street	city	...	zip	lat	long	city_pop	job	dob	unix_time	merch_lat	merch_long	is_fraud
2019-01-01 00:00:18	2703186189652095	Rippin, Kub and Mann	misc_net	4.97	Jennifer	Banks	F	561 Perry Cove	Moravian Falls	...	28654	36.0788	-81.1781	3495	Psychologist, counselling	1988-03-09	1325376018	36.011293	-82.048315	0
2019-01-01 00:00:44	630423337322	Heller, Gutmann and Zieme	grocery_pos	107.23	Stephanie	Gill	F	43039 Riley Greens Suite 393	Orient	...	99160	48.8878	-118.2105	149	Special educational needs teacher	1978-06-21	1325376044	49.159047	-118.186462	0
2019-01-01 00:00:51	38859492057661	Lind-Buckridge	entertainment	220.11	Edward	Sanchez	M	594 White Dale Suite 530	Malad City	...	83252	42.1808	-112.2620	4154	Nature conservation officer	1962-01-19	1325376051	43.150704	-112.154481	0
2019-01-01 00:01:16	3534093764340240	Kutch, Hermiston and Farrell	gas_transport	45.00	Jeremy	White	M	9443 Cynthia Court Apt. 038	Boulder	...	59632	46.2306	-112.1138	1939	Patent attorney	1967-01-12	1325376076	47.034331	-112.561071	0
2019-01-01 00:03:06	375534208663984	Keeling-Crist	misc_pos	41.96	Tyler	Garcia	M	408 Bradley Rest	Doe Hill	...	24433	38.4207	-79.4629	99	Dance movement psychotherapist	1986-03-28	1325376186	38.674999	-78.632459	0

EXPLORATORY DATA ANALYSIS

01

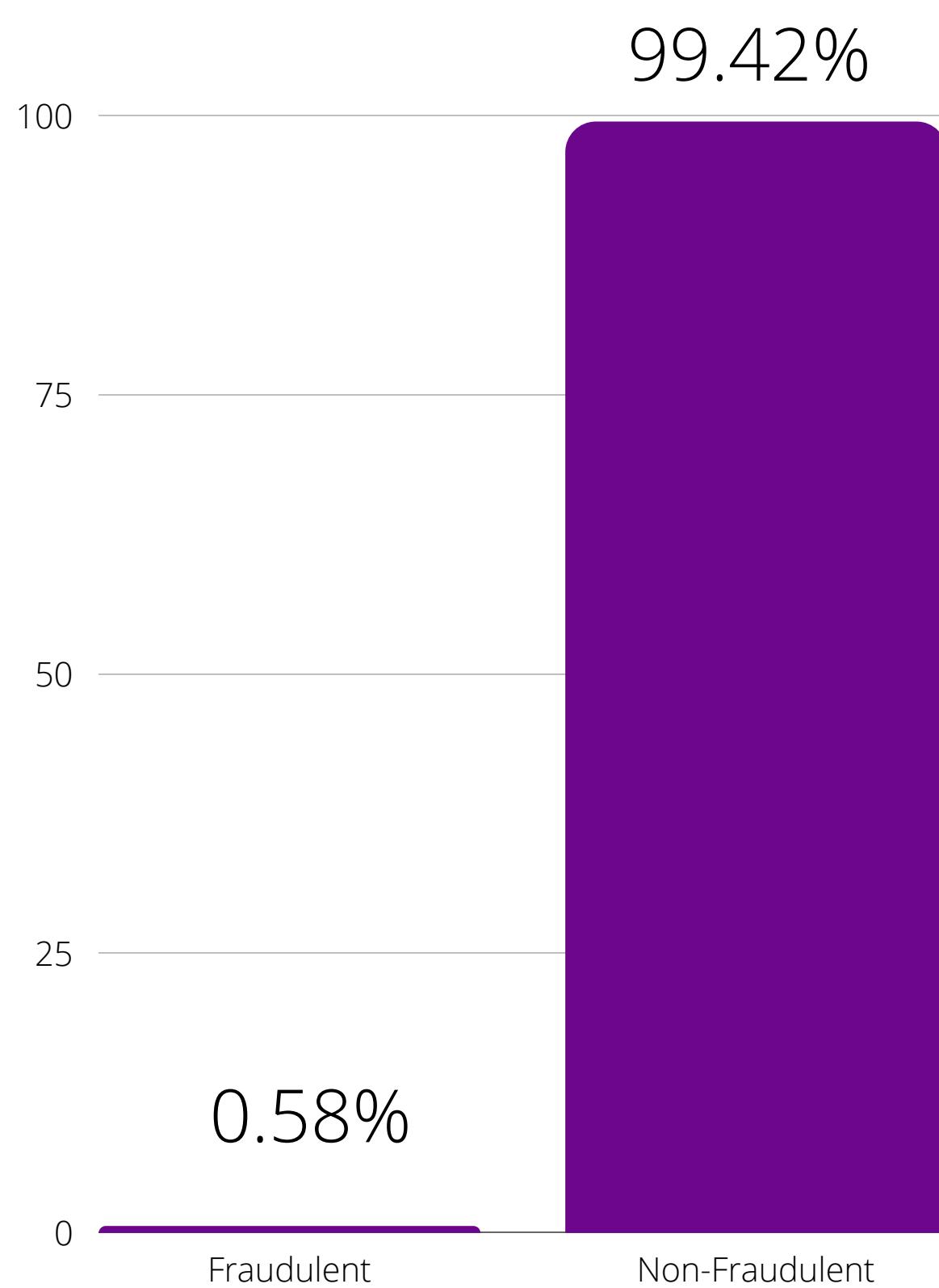
CLEANED
DATA

- Checked for duplicates, null values
- Changed Data Types to Appropriate Types

02

BASIC
EDA

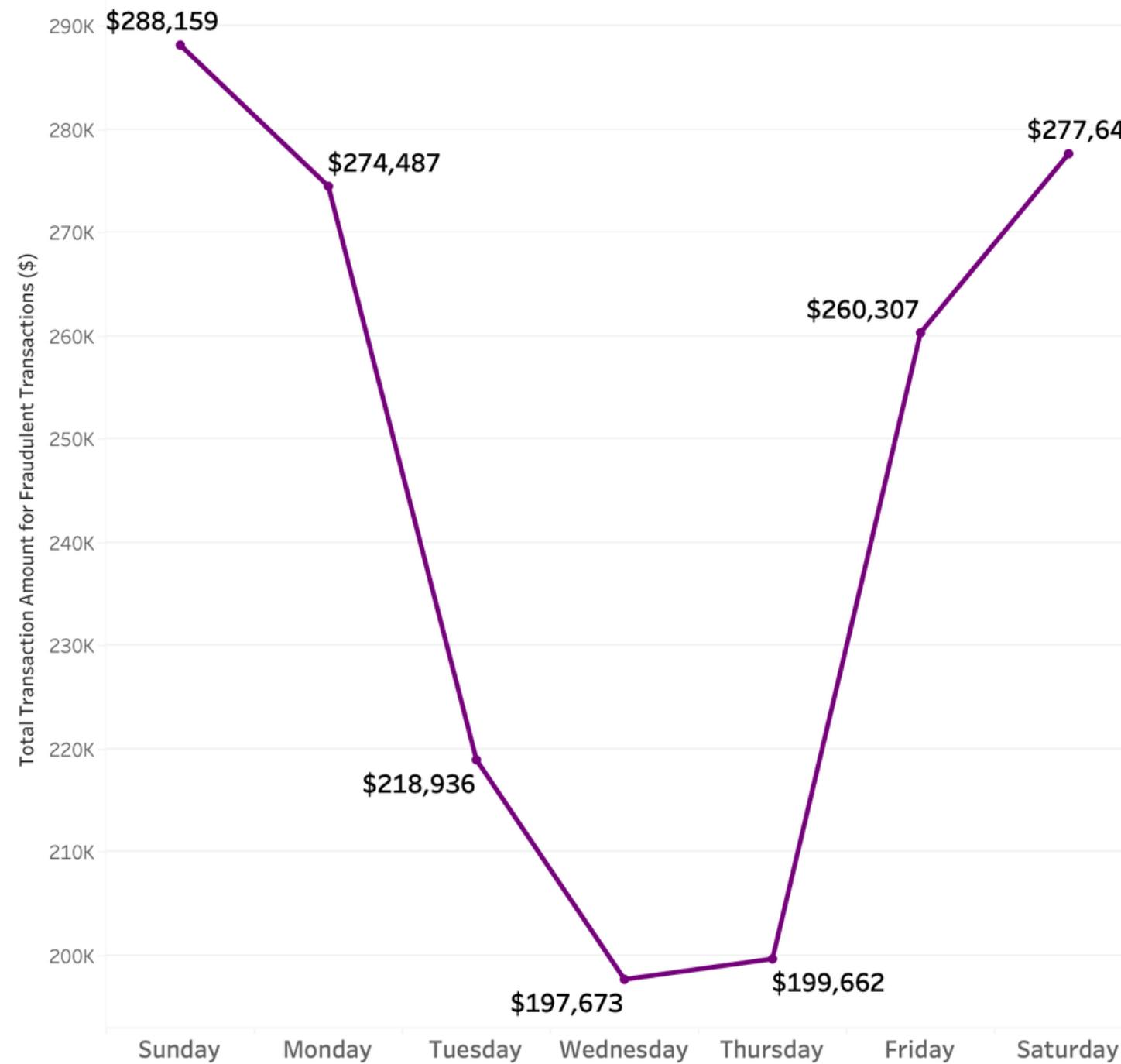
- Extreme class imbalance



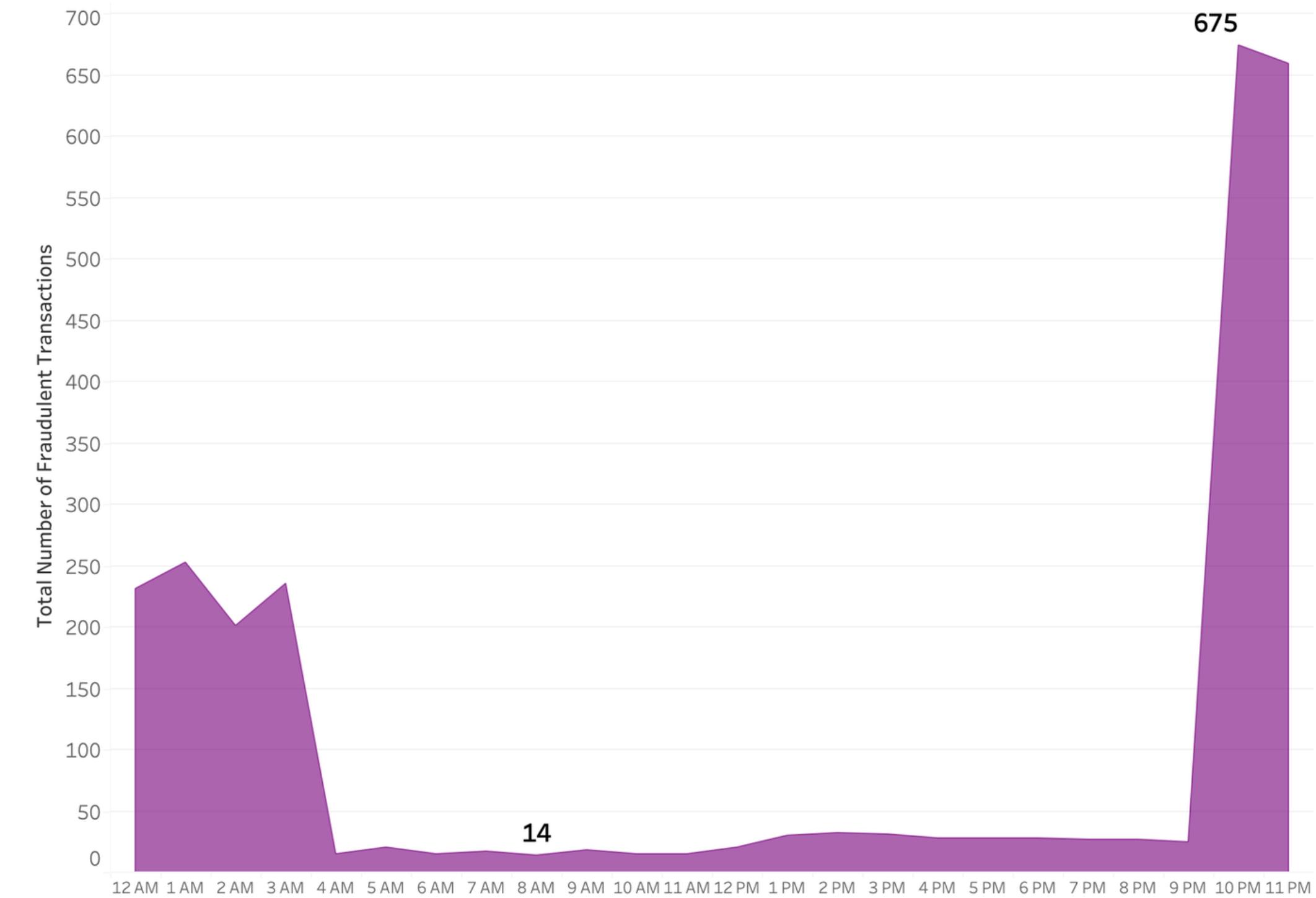
EDA INSIGHTS

Fraudulent Transactions and Time/Weekday

Total Transaction Amount of Fraudulent Transactions Per Weekday



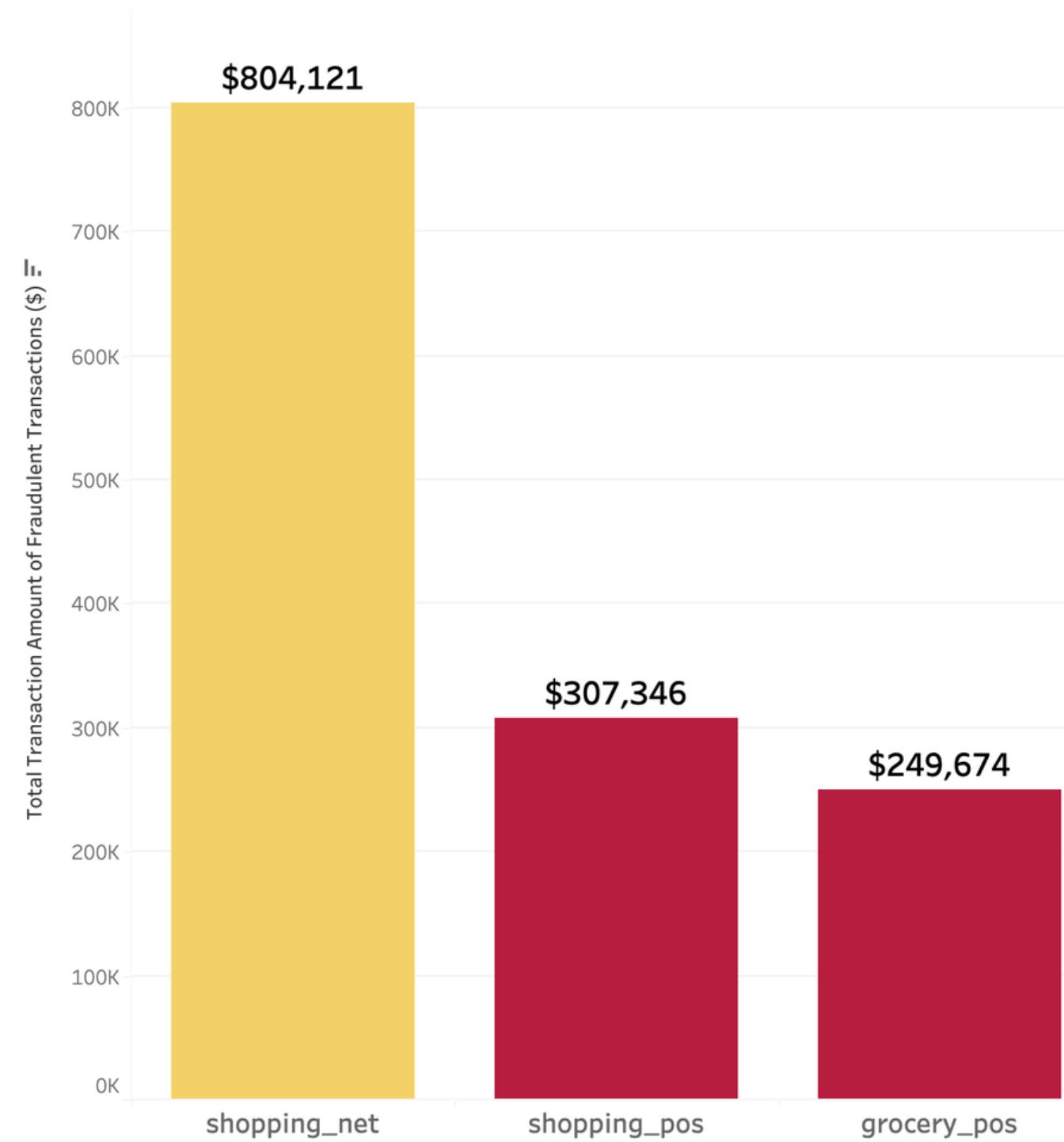
Total Number of Fraudulent Transactions Per Hour



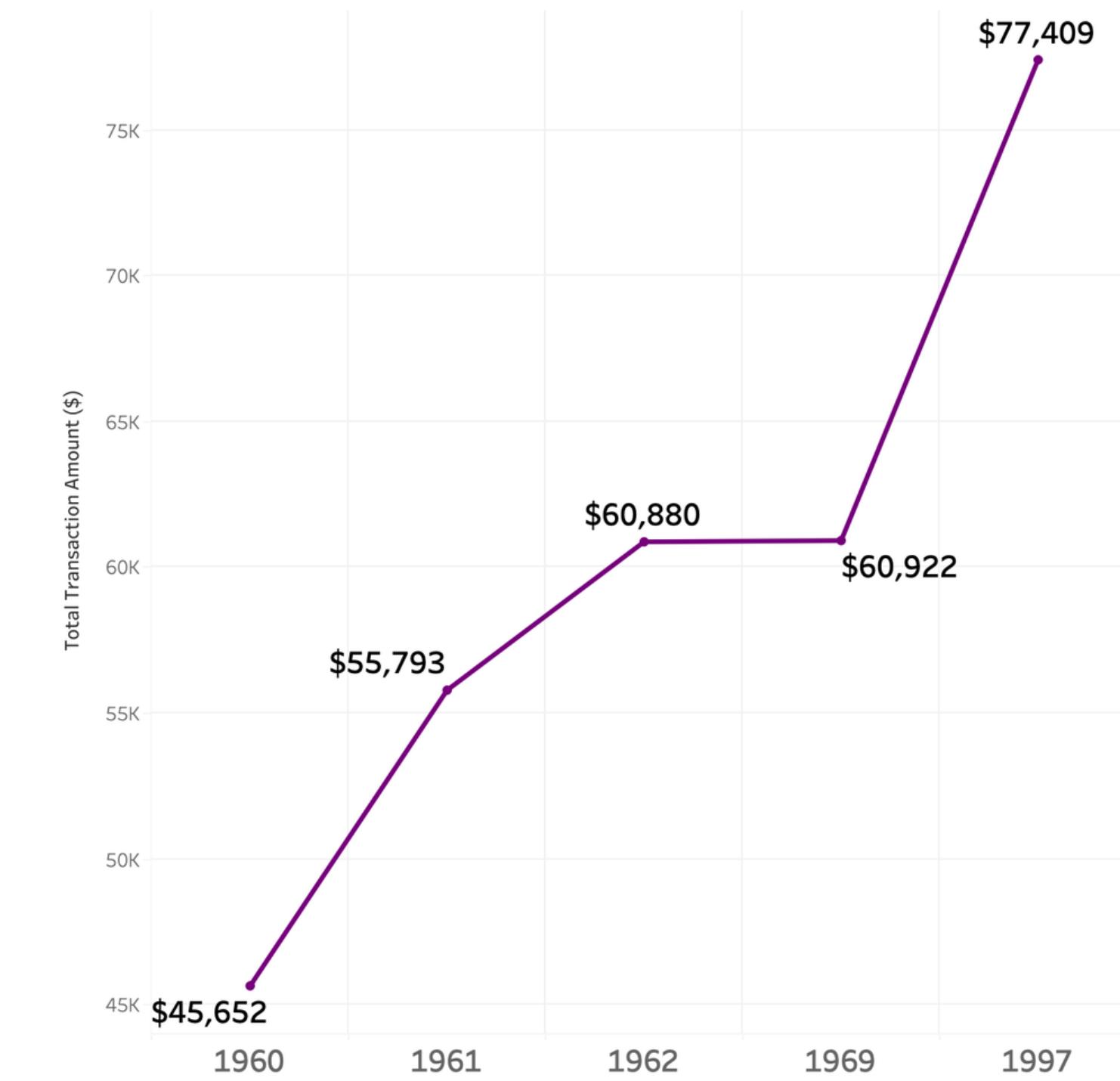
EDA INSIGHTS

Fraudulent Transactions and Categories/Date of Birth

Top Categories with the Highest Amount of Fraudulent Transactions



Top 5 Total Transaction Amount Based on the Year of Birth of Customers



PRE-PROCESSING

03

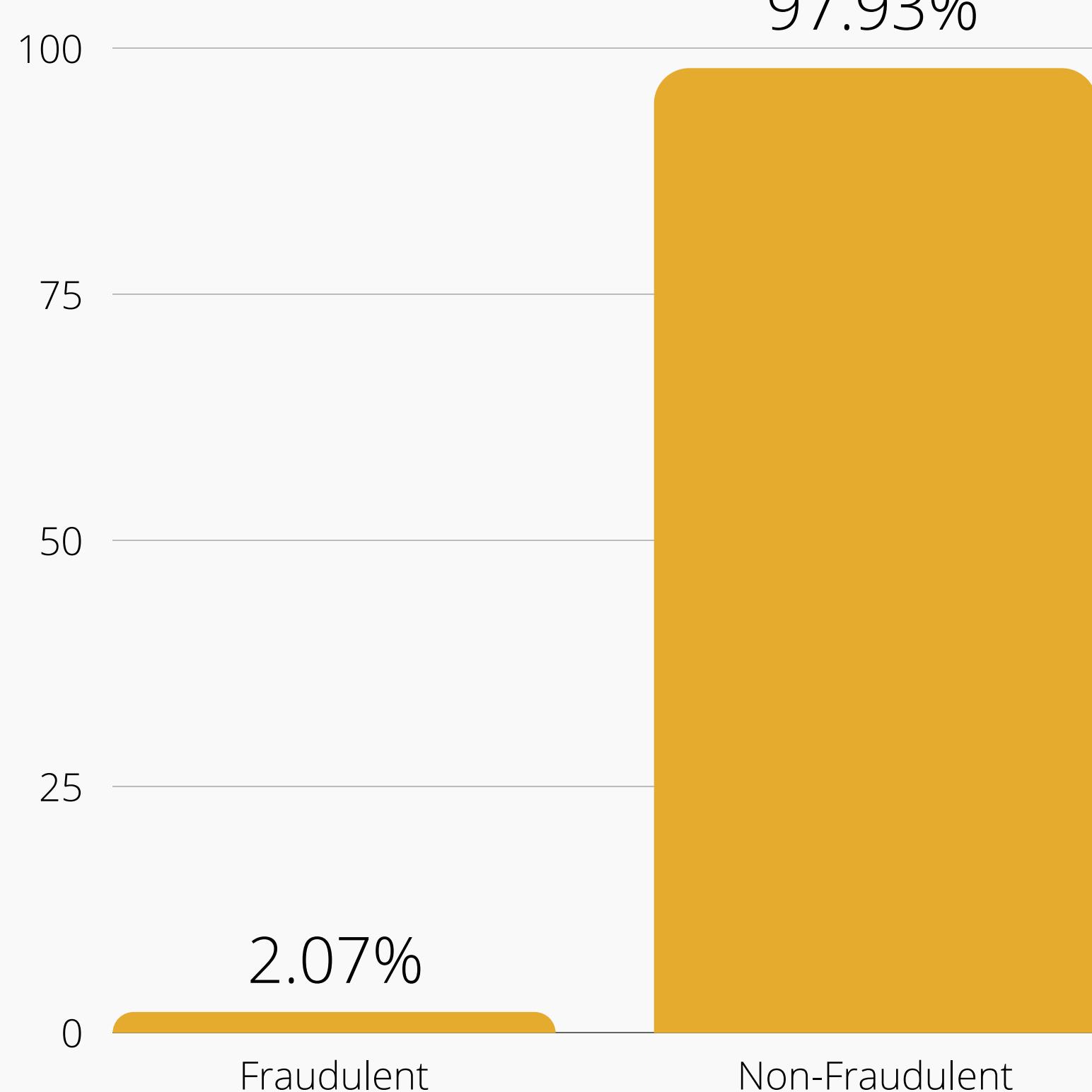
FEATURE ENGINEERING AND PRE-PROCESSING

Starting Data Points

- 1,296,675 rows
- 22 columns

Current Data Points

- 128,802 rows
- 782 columns



MODELING

04

MODELING

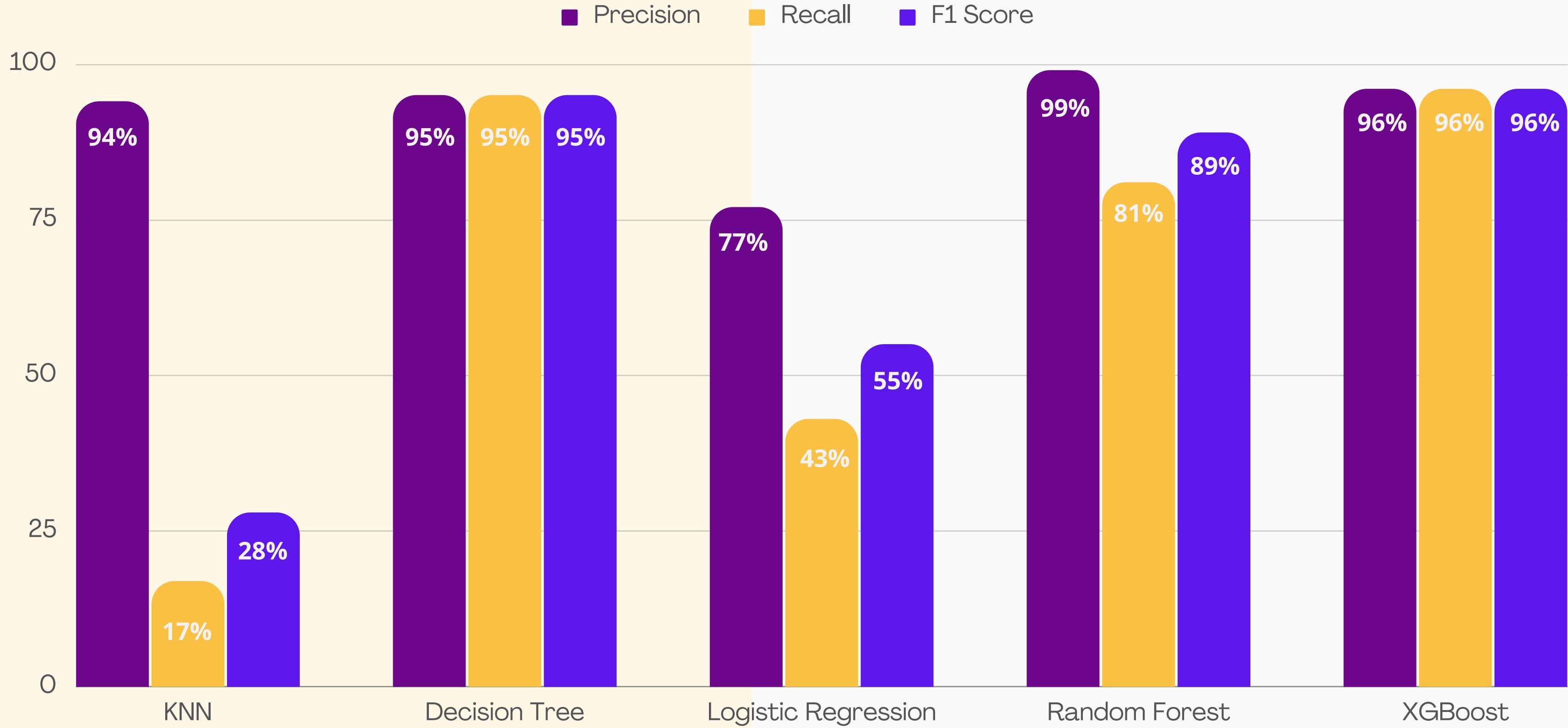
- K-Nearest Neighbor
- Decision Tree
- Logistic Regression
- Random Forest
- XGBoost

EVALUATION METRICS

- Identify as many fraudulent transactions as possible.
- Priority is Recall then having a good enough F1 Score.



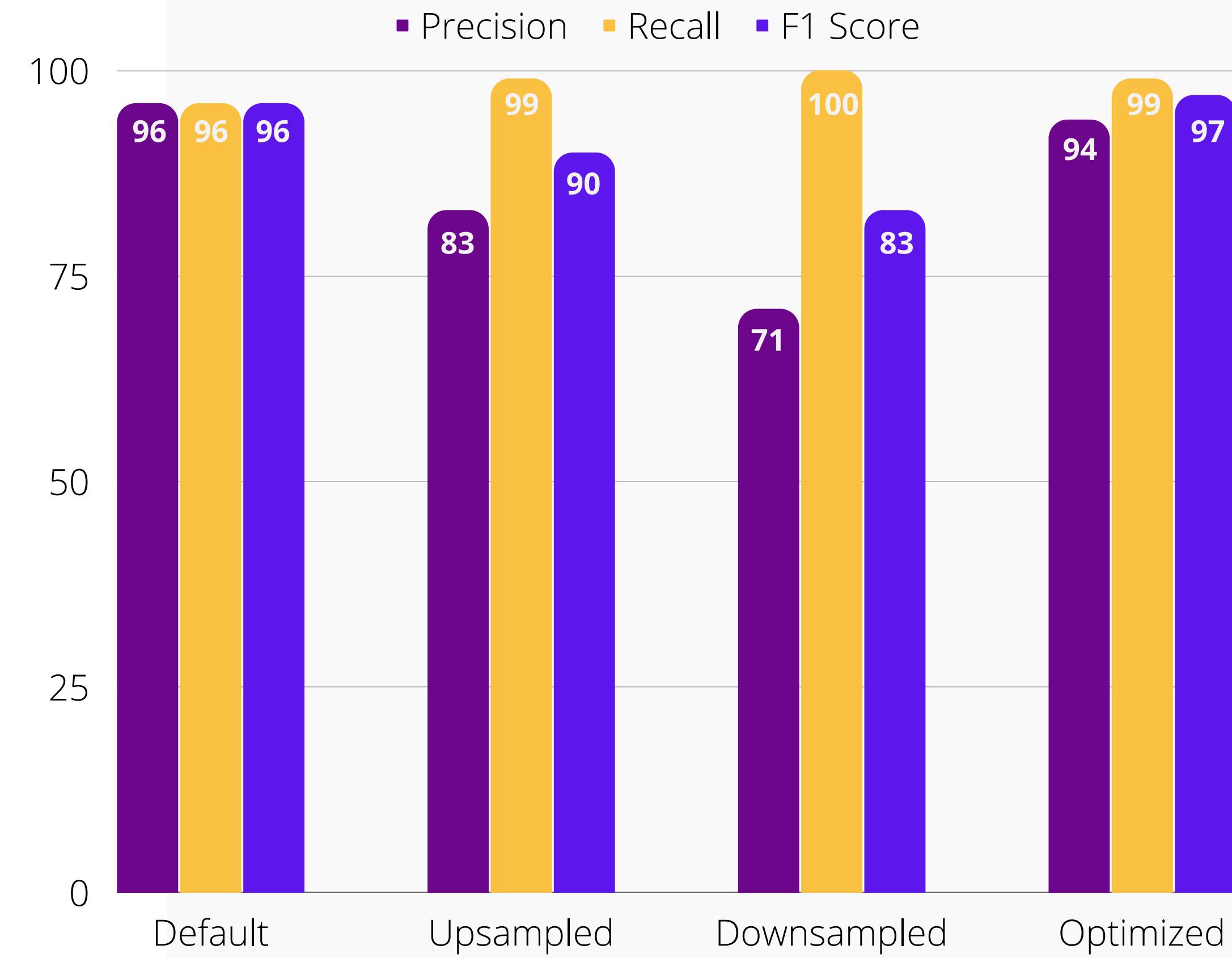
PRELIMINARY RESULTS FROM MODELS



Findings

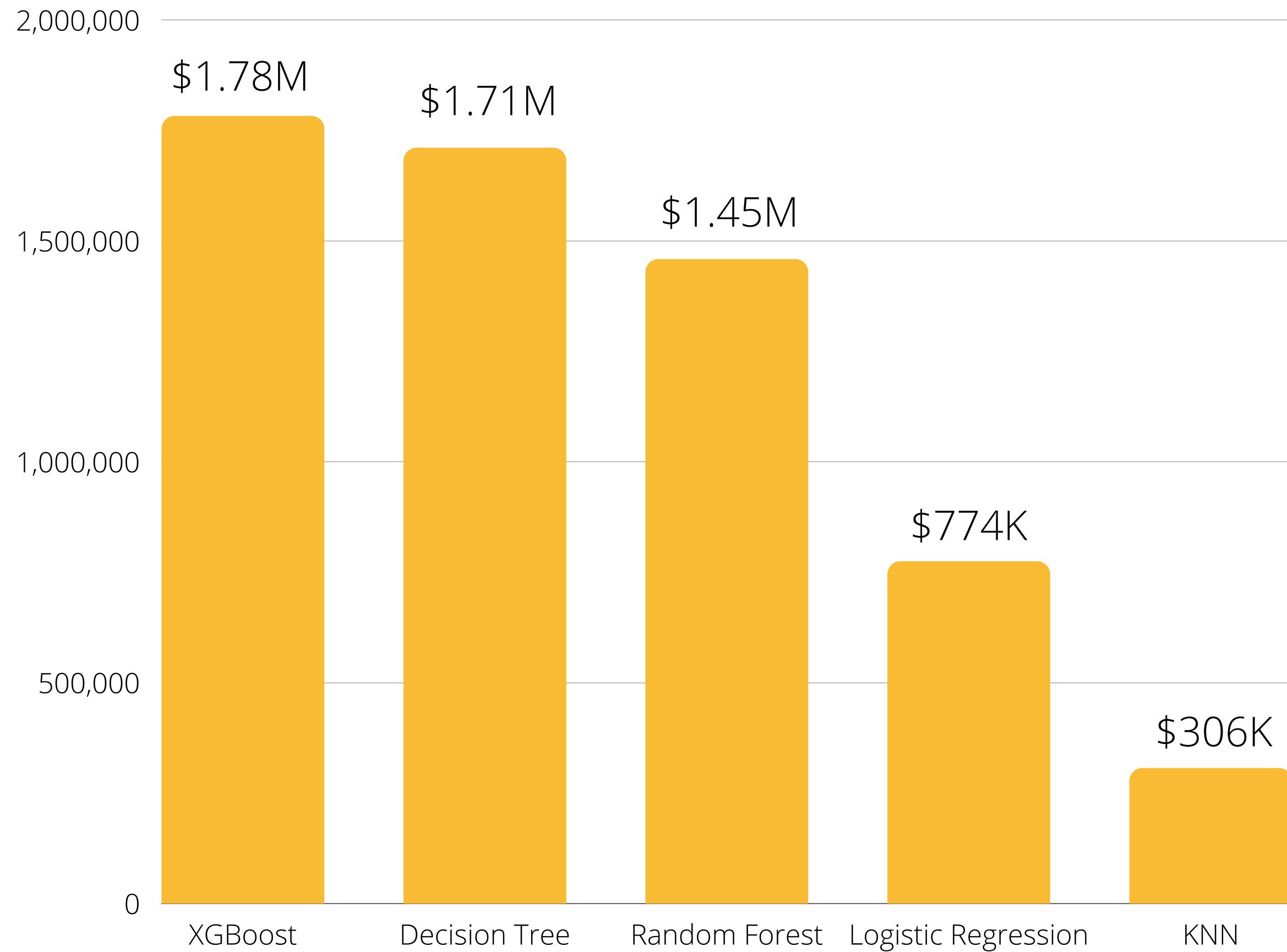
XGBOOST MODEL

- Best performing model based on the set evaluation metric.
- Corrected class imbalance using upsampling and downsampling.
- Optimized upsampled model to get the best parameters.





VALUE ADD



In the dataset, about \$1.8M was lost due to fraud. This bar plot shows the total amount of money that would have been saved by using the individual models.

Practical Applications

Financial
Institutions



Credit Card
Companies



FinTech
Companies



Potential Future Direction



Predict Customers with
Increased Likelihood of
Targeted Fraud



Further Development Outside
of Credit Card Transactions



THANKS
FOR LISTENING!

