

Tackling Credit Card Fraud in US

Vertical Institute Data Analytics Capstone Project

Neo Si Yang

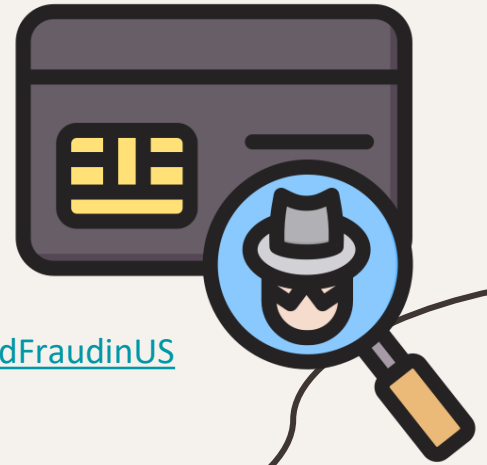


Tableau link:

https://public.tableau.com/views/CapstoneProject_NeoSiYang881G/TacklingCreditCardFraudinUS?:language=en-US&:sid=&:display_count=n&:origin=viz_share_link

Agenda

01

Background

02

**Problem Statement
& Objectives**

03

Dataset

04

Analysis

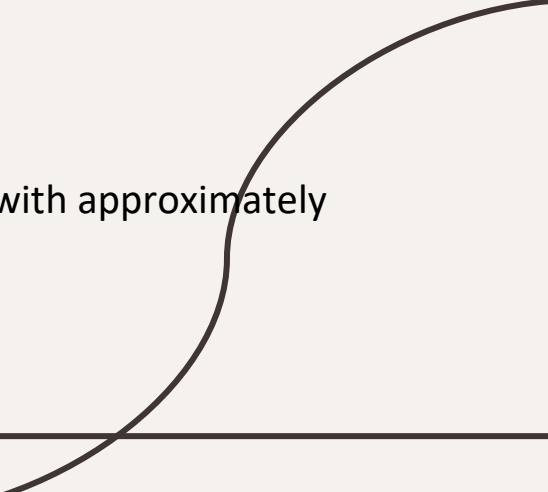
05

Recommendations

01 

Background

Background

- Credit card fraud - **Unauthorized** use of someone else's credit card to make transactions without the **cardholder's consent**.
 - **USA** has the highest susceptibility to credit card fraud globally, accounts for **>1/3** of total global loss [1].
 - **\$28.43B** lost globally due to credit card fraud in 2020 [2].
 - **Covid-19** is responsible for the surge in credit card fraud activity, with approximately **393,378** reported cases in 2020 [1].
- 

02



Problem Statement & Objectives

Problem Statement

A bank manager from Remora, a multinational bank from US, received feedback that there has been a rise of complaints from customers due to **unauthorised credit card transactions**. The bank manager instructed a data analyst from the fraud department **to investigate the underlying patterns behind credit card frauds and what are the possible measures to protect their customers from being a credit card fraud victim.**

Objectives

- Investigate distribution of fraud cases across gender and age groups.
 - Identify regions in the US that are susceptible to credit card fraud.
 - Determine the typical times when credit card fraud incidents commonly occur.
 - Examine what are the risky merchant categories.
 - Propose business recommendations based on previous findings.
-

03 
Dataset

Dataset Selection

- The chosen dataset is a simulated credit card transaction dataset that contains both legitimate and fraudulent transactions from June to December 2020.
 - Only the test dataset (fraudTest.csv) is used for the analysis.
 - Contains 555719 rows, 23 fields (columns)
 - Kaggle dataset URL: <https://www.kaggle.com/datasets/kartik2112/fraud-detection?select=fraudTest.csv>
 - Acknowledgement goes to Brandon Harris for the creation of dataset.
-

Metadata

Column Name	Description	Data Type
-	Unique ID	Integer
trans_date_trans_time	Transaction date & time	Date / String
cc_num	Credit card number	Integer
merchant	Merchant name	String
category	Category of merchant	String
amt	Transaction amount	Float
first	First name of credit card holder	String
last	Last name of credit card holder	String
gender	Gender of credit card holder	String

Metadata

Column Name	Description	Data Type
street	Street address of credit card holder	String
city	City of credit card holder	String
state	State of credit card holder	String
zip	Postal code of credit card holder	Integer
lat	Latitude location of credit card holder	Geographical / Float
long	Longitude location of credit card holder	Geographical / Float
city_pop	City population of credit card holder	Integer
job	Occupation of credit card holder	String

Metadata

Column Name	Description	Data Type
dob	Card holder's date of birth	Date / String
trans_num	Transaction number	Integer
unix_time	Unix time of transaction	Integer
merch_lat	Latitude location of merchant	Geographical / Float
merch_long	Longitude location of merchant	Geographical / Float
is_fraud	Whether a transaction is fraudulent (1 = Yes, 0 = No)	Boolean / Integer

Dataset Cleaning & Preprocessing

- CSV file is converted to Excel file to preserve data format.
 - A column name “index” was added, and this represents the unique identifier for each row.
 - Null values were checked – No missing data under each field.
 - Renaming of column name for better clarity:
 - amt to trans amt (transaction amount)
 - In the “merchant” column, removed the “fraud_” in front of all merchant names using Excel’s find & replace all function.
 - Added 3 extra columns:
 - age (age of credit card holders)
 - full name (full name of credit card holders)
 - outcome (label 1 as fraudulent, 0 as not fraudulent)
-

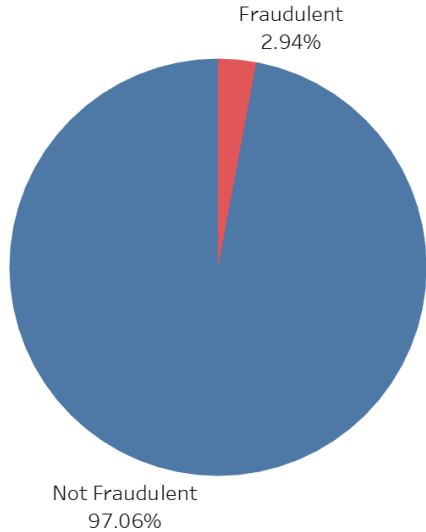
04 Analysis

Tableau link:

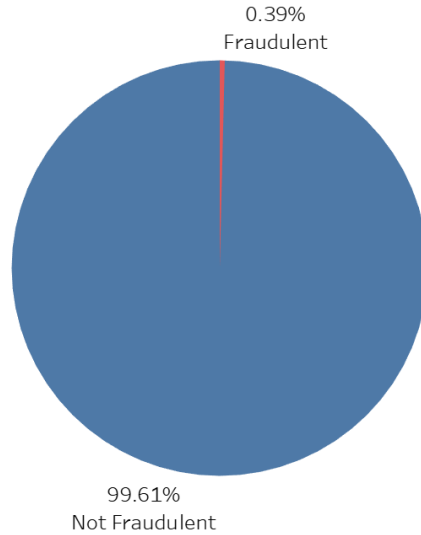
https://public.tableau.com/views/CapstoneProject_NeoSiYang881G/TacklingCreditCardFraudinUS?:language=en-US&:sid=&:display_count=n&:origin=viz_share_link

Overview

Fraud Amt Percentage



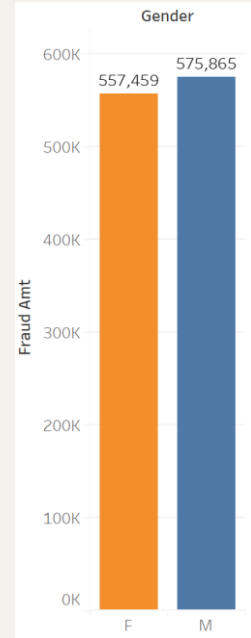
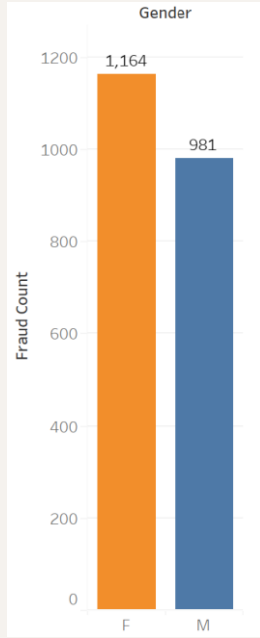
Fraud Count Percentage



2145 fraudulent transactions
from Jun to Dec 2020

\$1,133,325 lost in total

Fraud Victims Demographics (Gender)



Total Fraud Count

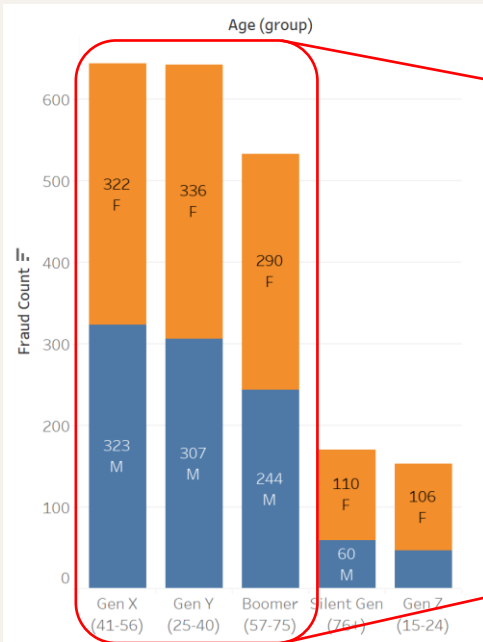


Total Fraud Amount



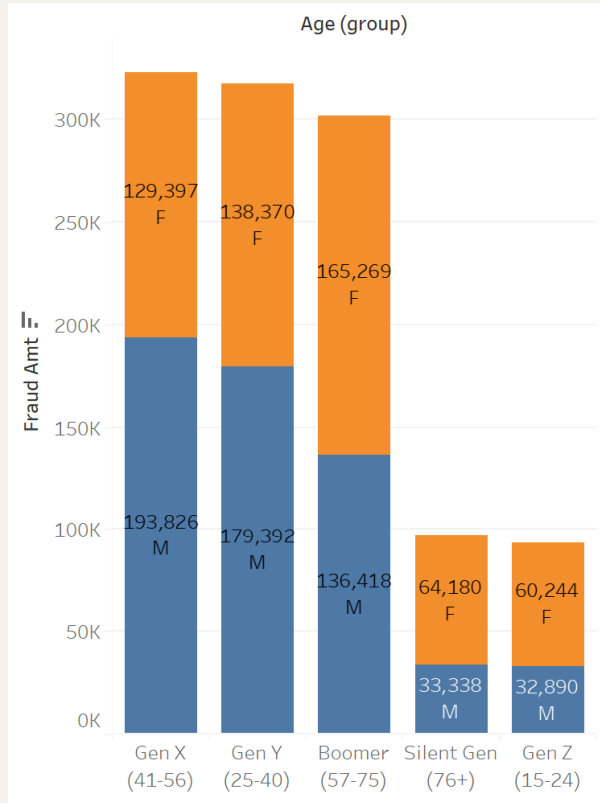
While there were more female victims than male victims, males experienced higher losses in fraud transactions due to their higher average fraud loss.

Fraud Victims Demographics (Age Group & Range)



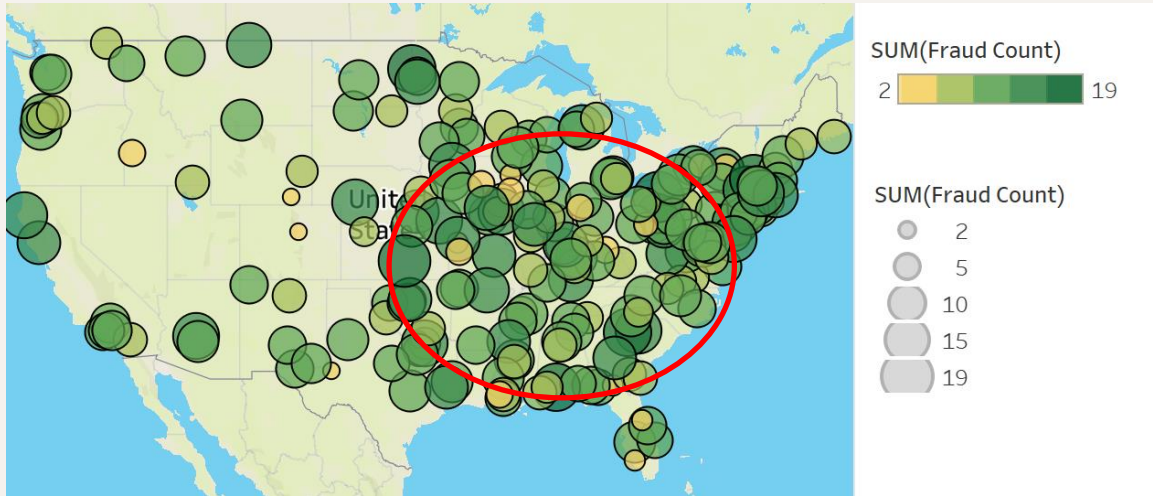
- Credit card fraud impacts individuals across all age groups, with the majority of victims belonging to **Generation X, Y, and Baby Boomers**.
- The losses incurred in fraudulent transactions were also higher within these age groups, likely due to their higher credit card possession and spending tendencies, making them more susceptible to credit card fraud.
- Breaking down the Gen X, Y, and Baby Boomers groups further into specific age ranges revealed that most victims fall within the age range of **30 to 59**.
- About **69.6%** of the fraud cases within the Generations X, Y, and Baby Boomers group were concentrated among individuals falling within the age range of 30 to 59.
- About **68.2%** of the total transaction losses within the Generations X, Y, and Baby Boomers group were contributed by victims falling within the age range of 30 to 59

Fraud Victims Demographics (Age Group & Range)



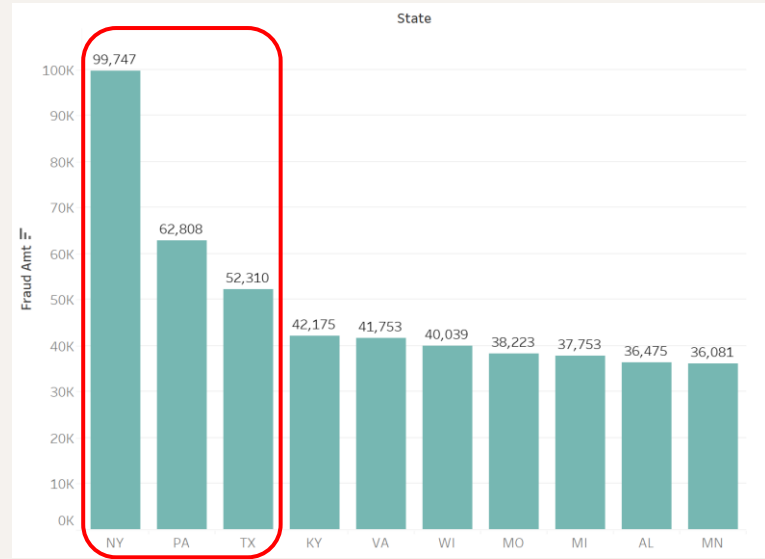
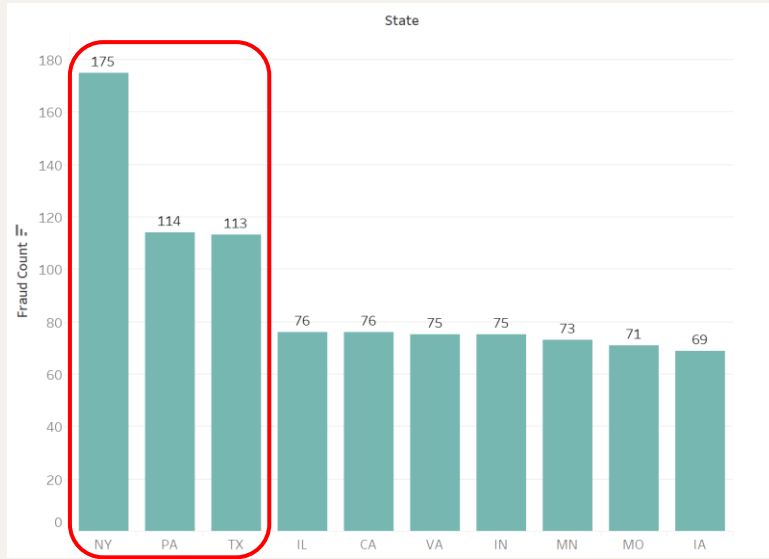
- Among Generation X, Y, and Boomers, there was no significant difference in the composition of fraud count and fraud amount between genders. Overall, both genders were equally susceptible to credit card fraud.

Fraud Rate Distribution By Location



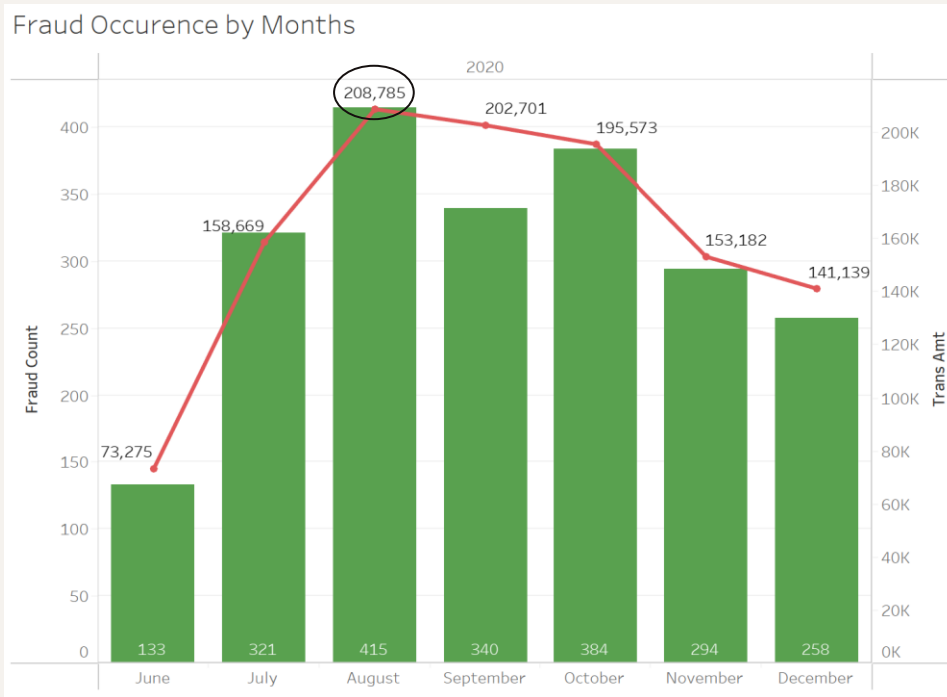
- Majority of fraud incidents occurred on **East Coast** of the USA, particularly in states known for their wealth and high population density.

Fraud Rate Distribution By Location



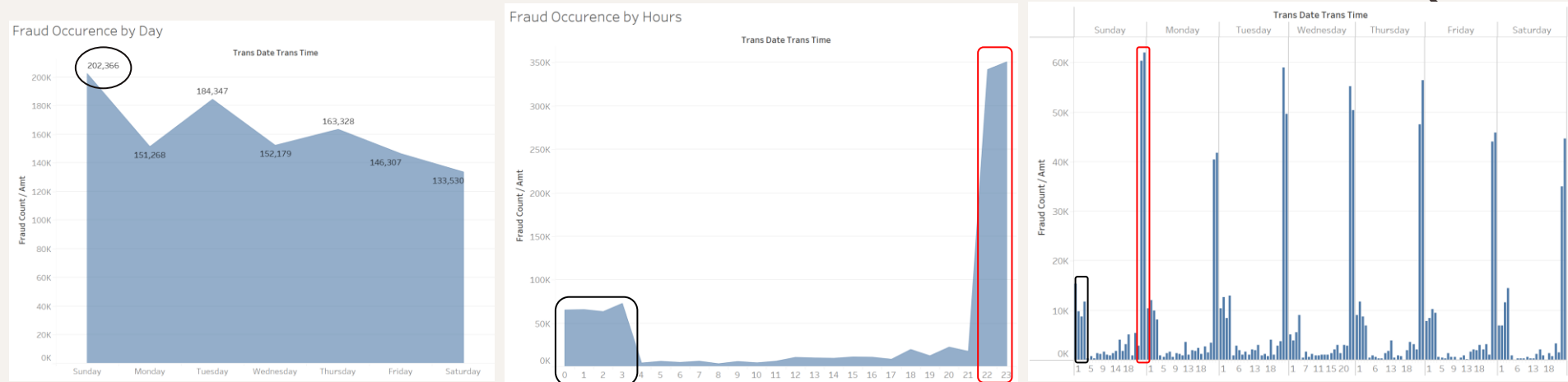
- Top 3 states with the highest instances of fraud and total fraud transactions: **New York, Pennsylvania, and Texas (NY, PA, TX)**.
- New York contributed to **8.8%** of total fraud amount loss (\$1,133,325).
- Top 3 states with the highest average fraud transactions: **Kansas, Washington and Kentucky**.
- Residents in these states often possess more than one credit card with a higher credit limit, enabling them to make substantial transaction volumes. However, this also makes them susceptible targets, as small charges on credit cards can be easily overlooked.

When Do Credit Card Frauds Occur?



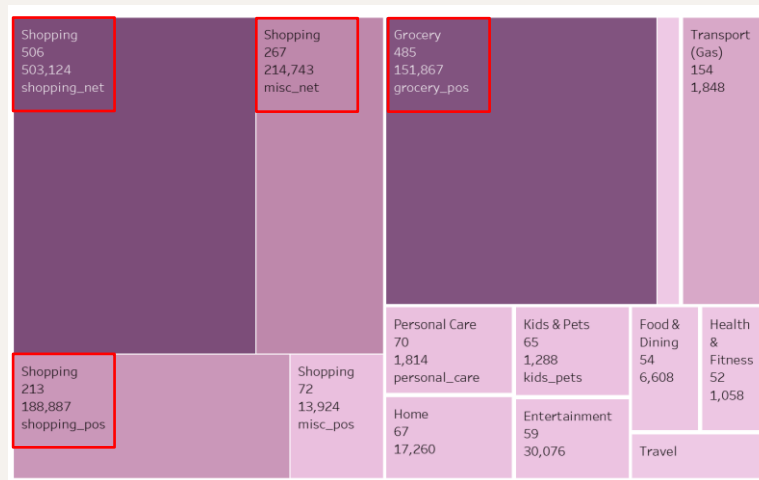
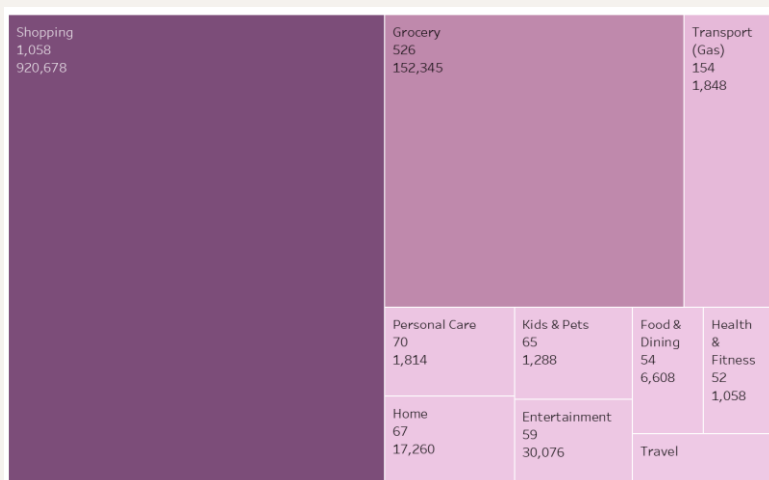
- A notable increase in fraud instances and transaction losses after June, peaking in **August**.
- This surge might be correlated with the rise in COVID-19 cases during the same period, prompting more people to stay at home and engaged in **online purchases**. This might contribute to a rise in card-not-present frauds.

When Do Credit Card Frauds Occur?



- Interestingly, data indicated that credit card fraud occurred most frequently on **Sundays**, evident from the peak in both total fraud count and amount.
- Moreover, there was a discernible pattern showing fraudulent transactions were more prominent from **22:00 to 03:00**. This suggest that monitoring systems might not be as robust during these hours.

Fraud Instances Distribution By Merchant Category

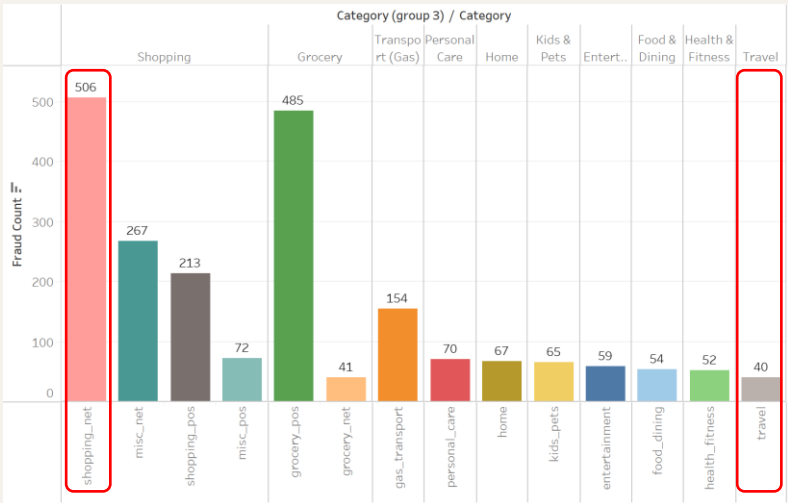


- Total fraud count and transaction losses were highest in the shopping category.
- Risky merchant subcategories: **shopping_net**, **shopping_pos**, **grocery_pos**, **misc_net**.

*Note:

- misc_net & pos are grouped together with shopping_net & pos as they are all related to shopping.
- pos – point of sale, transaction done at outlet
- net – online transaction

Fraud Instances Distribution By Merchant Category



Average Fraud Amt By Category Over Months

Category	Subcategory	June	July	August	September	October	November	December	Grand Total
Shopping	shopping_net	\$1,017.97	\$1,000.74	\$992.58	\$998.12	\$994.93	\$990.42	\$973.41	\$994.32
	shopping_pos	\$864.06	\$866.91	\$902.68	\$904.60	\$897.19	\$850.15	\$895.25	\$886.79
	misc_net	\$785.64	\$791.28	\$801.95	\$820.27	\$799.88	\$803.66	\$812.75	\$804.28
	misc_pos	\$28.94	\$139.31	\$371.50	\$143.17	\$111.28	\$19.14	\$400.36	\$193.39
Entertainment	entertainment	\$536.46	\$545.03	\$510.34	\$479.56	\$530.37	\$486.03	\$495.69	\$509.77
Grocery	grocery_pos	\$315.66	\$312.13	\$316.99	\$311.44	\$310.96	\$308.71	\$318.48	\$313.13
	grocery_net	\$13.25	\$11.84	\$10.69	\$11.59	\$13.28	\$11.22	\$7.76	\$11.67
Home	home	\$272.46	\$247.64	\$264.47	\$293.53	\$264.50	\$278.84	\$233.21	\$257.62
Food & Dining	food_dining	\$115.67	\$123.97	\$123.69	\$120.27	\$123.24	\$118.43	\$121.29	\$122.36
Personal Care	personal_care	\$28.27	\$20.09	\$26.72	\$19.83	\$27.18	\$33.71	\$25.37	\$25.92
Health & Fitness	health_fitness	\$22.18	\$19.16	\$20.53	\$19.64	\$20.85	\$18.44	\$19.07	\$20.35
Kids & Pets	kids_pets	\$19.69	\$19.30	\$19.59	\$20.15	\$20.21	\$19.93	\$19.85	\$19.82
Transportation	gas_transport	\$11.84	\$11.19	\$13.29	\$11.47	\$12.85	\$11.67	\$11.74	\$12.00
Travel	travel	\$7.83	\$8.88	\$10.12	\$8.67	\$7.68	\$8.41	\$8.70	\$8.70

- Among various subcategories, **shopping_net** recorded the highest fraud count at **506** and the highest average transaction losses at **\$994.32**, while **travel** recorded the lowest fraud count at **40** and the lowest average transaction losses at **\$8.70**.
- Shopping_net & misc_net together contributed about **63.3%** of the total transaction losses, indicating that online shopping was a substantial contributor to the overall financial losses incurred over the seven-month period.
- This suggests a possibility of a surge in online activities during the COVID-19 period, contributing to an increase in e-commerce transactions and, consequently, providing more opportunities for card-not-present frauds.
- The occurrence of credit card fraud associated with travel is expected to be low due to the low travel volume during the pandemic.

Top 5 - 20 Victims Profile Based on Total Transaction Loss

Rank	Full N.	Age (group)	State	Category	Category	Trans Amt	Hour of Tra.	Count
1	Mary Williams	Gen Z (15-24)	KS	Shopping	shopping_net	2,800.47	23	3
					shopping_pos	1,990.69	20	2
					misc_net	1,907.9	22	2
					misc_pos	966.98	18	1
					misc_net	1,719.99	23	2
					shopping_pos	1,252.51	23	1
					misc_pos	8.45	23	1
2	Jason Johnson	Gen Z (15-24)	AK	Shopping	shopping_net	4,971.59	22	5
					shopping_pos	881.12	23	1
					shopping_pos	2,437.13	22	2
					misc_net	1,183.93	23	1
					misc_net	781.06	23	1
					misc_pos	745.05	22	1
3	Gina Grimes	Gen Z (15-24)	PA	Shopping	shopping_net	4,071.08	22	4
					misc_net	1,852.75	23	2
					misc_net	1,623.28	23	2
					misc_net	1,401	22	2
					misc_pos	826.89	21	1
4	Patrick Bates	Gen X (41-56)	NY	Shopping	misc_net	2,538.55	22	3
					misc_net	1,673.92	23	2
					shopping_pos	2,155.03	22	2
					misc_net	991	23	1
					shopping_net	1,135.26	13	1
					misc_pos	966.79	23	1
5	Edward Hensley	Gen Y (25-40)	MI	Shopping	shopping_net	2,710.31	22	3
					misc_net	957.9	23	1
					misc_net	850.57	18	1
					misc_net	2,569.23	23	3
					misc_pos	847.3	22	1
					shopping_pos	916.04	22	1

- Patterns found based on top 20 victims' profile:
- The shopping category accounted for the majority of transaction losses. More precisely, **online shopping** (shopping_net) emerged as the primary source of concern, with **17 out of 20 victims** encountering the highest transaction losses within this category.
 - Most of credit card fraud incidents took place during late-night hours, with the highest concentration observed between **2200 and 2300 hrs.**
 - Most of the victims come from densely populated and richer states such as **New York, Kansas and Pennsylvania.**
 - **15 out of 20** victims belong to Gen X, Y and Boomers category.

*Note: This is only top 5 preview. For the information about top 20, please view in tableau.

04 

Recommendations

Findings



Both genders show no significant difference in terms of fraud count and amount, indicating equal vulnerability to credit card fraud. Credit card fraud impacts individuals across all age groups, with a majority falling within the **Generation X, Y, and Baby Boomer** categories, specifically within the age range of **30 to 59** years old.

Recommendations

Educational Programs



- Conduct regular educational programs for customers from age **30-59** to raise awareness about the types of card fraud, possible causes and preventive measures.
- Provide information about **online safety practices**, such as using secure websites to make e-commerce transactions and tips to avoid phishing scams. This helps customers to protect their credit card details and other confidential data.
- Such programs not only reduce card fraud occurrence but also allow bank to demonstrate their commitment to customer's well being. Informed customers are likely to have greater **trust** in the bank, leading to **increased loyalty and retention**.

Encourage Frequent Monitoring of Credit Card Accounts



- Early intervention is crucial to overcome credit card frauds effectively. Occasionally, detection systems may overlook a fraudulent attempt. Therefore, consistent monitoring of card activity is important in preventing credit card fraud.
- To encourage customers to review their accounts regularly, banks can make this easier by developing mobile app functionality that provides **automated transaction alerts** to customers, notifying them real time transactions made on their credit cards. Additionally, banks should educate customers on the usage of such feature.
- Alerts should be customizable based on **preferences** such transaction amount thresholds, international transactions, etc.
- Another way to encourage customers is to partner with credit monitoring services that offer comprehensive monitoring of credit reports, score and protection against identity theft. Customers can opt for such monitoring service at a highly subsidised rate when they apply for a new credit card from the bank.
- Encouraging frequent monitoring of credit card accounts not only enables early intervention in potential card fraud but also serves as an attractive feature for acquiring new customers who are looking for a bank that offers robust identity protection services.

Findings



Most fraud incidents occur on the **East Coast** of the USA, primarily in states characterised by **high population density and wealth**.



The surge in fraud is closely linked to the pandemic situation, as individuals increasingly opt for **online shopping** for safety concerns. This shift has created an opportunity for cybercriminals to exploit vulnerabilities and obtain confidential card data for unauthorized transactions. Some fraudulent activities occur during **point-of-sale transactions** for groceries and transportation. This may be attributed to the presence of **skimmers**, devices capable of illicitly collecting customers' confidential credit card data during transactions.

Recommendations

Encourage the use of Secure Payment Processing Methods



- Bank can proactively encourage **shopping** merchants, particularly those located in high-fraud regions like the **eastern** states of the USA, to adopt secure payment methods.
- For online payments, merchants should consider leveraging secure payment gateways, incorporating features such as end-end encryption and tokenization to ensure the secure transfer of **online payment data**.
- As for physical card payments, merchants can enforce customers to use only credit cards that are equipped with **EMV chip technology** to mitigate the risk of skimming.
- These secure payment processing methods can effectively prevent customers' confidential credit card data from being stolen.
- To motivate merchants to enhance their payment systems, bank can extend financial assistance, alleviating the financial burden linked to system upgrades. By helping them to adopt secure payment methods, banks actively contribute to the **reduction of fraudulent transactions**. This, consequently, results in **reduced financial losses pertaining to chargebacks**.

Findings



According to the analysis, fraudulent activities are most prevalent on **Sundays, Tuesdays, and Thursdays**. However, in real-life scenarios, fraudulent activities tend to fluctuate between days, influenced by factors such as festive seasons, holidays, major sales events, etc. One noticeable trend from the analysis is the common occurrence of fraudulent activities during **late-night hours**, specifically from **22:00 to 03:00** throughout the week.

Recommendations

Advanced Fraud Detection Systems



- Consider implementing a fraud detection system that utilises artificial intelligence to flag **potential fraudulent transactions in real time**. This system will automatically block suspicious transactions, regardless of their size.
- Furthermore, such detection systems can employ **machine learning techniques** to analyze customer's behavior, identifying anomalies that could indicate fraudulent activities through deviations from typical behavior of legitimate users.
- Fraud detection systems should be actively operational during **late-night hours from 2200 to 0300 daily**. If feasible, consider extending these operating hours, especially **during special occasions such as peak shopping periods**.
- Investing in robust fraud detection systems not only protects customers from credit card fraud, but also demonstrates the bank's commitment to safeguarding its **customers' financial interests**. Customers are more likely to trust a bank that actively works to protect them from fraud.

Thank You

References

[1] J. Egan, "Credit Card Fraud Statistics," Bankrate, <https://www.bankrate.com/finance/credit-cards/credit-card-fraud-statistics/> (accessed Jan. 27, 2024).

[2] J. Lee, "Credit card fraud will increase due to the COVID pandemic, experts warn," CNBC, <https://www.cnbc.com/2021/01/27/credit-card-fraud-is-on-the-rise-due-to-covid-pandemic.html> (accessed Jan. 28, 2024).
