

Credit Card Fraud

Group 5

Breona Lewis, Chacon Lucas, Hasanain Alghanimi,
Matusevich Alena

Credit Card Fraud Intro

Credit_card_fraud.csv dataset consists of credit card transactions in the western United States. It includes information about each transaction including customer details, the merchant and category of purchase, and whether or not the transaction was a fraud. It has 339,607 credit card transaction rows with 1782 fraud transactions.

[Source]

<https://www.kaggle.com/kartik2112/fraud-detection?select=fraudTrain.csv>

Data Dictionary

transdate	trans_time	Transaction DateTime
merchant		Merchant Name
category		Category of Merchant
amt		Amount of Transaction
city		City of Credit Card Holder
state		State of Credit Card Holder
lat		Latitude Location of Purchase
long		Longitude Location of Purchase
city_pop		Credit Card Holder's City Population
job		Job of Credit Card Holder
dob		Date of Birth of Credit Card Holder
trans_num		Transaction Number
merch_lat		Latitude Location of Merchant
merch_long		Longitude Location of Merchant
is_fraud		Whether Transaction is Fraud (1) or Not (0)

Data Cleaning

```
# Check for empty values and columns data type
```

```
data.info()
```

```
# Check for missing or NaN values
```

```
data.isnull().values.any()
```

```
# Check for duplicates
```

```
duplicate_trans_num =  
data[data.duplicated('trans_num')]
```

Discussion about the data cleaning: The credit card transaction data looks clean because there is not any empty cells or missing values. Columns data type looks as expected for numerical and string values. As expected, transaction number is unique for every transaction. The data is clean and ready for further analysis.

Data Analysis and Visualizations

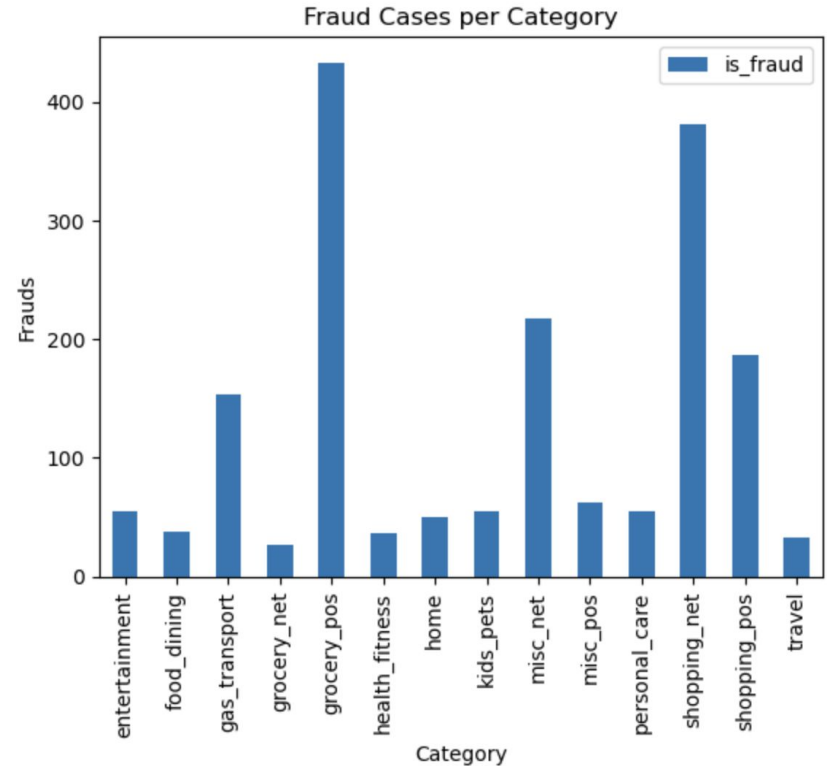
A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

Fraud Cases Per Category

Grocery had the highest fraud cases with 433

Internet Shopping had the a second highest with 381

Miscellaneous came in third with 217 cases

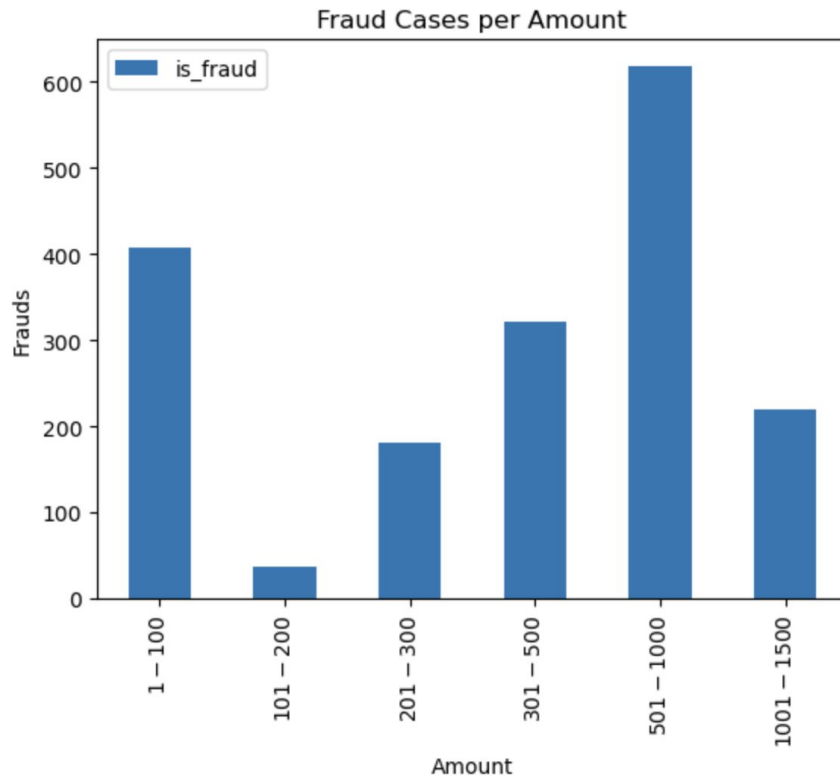


Fraud Cases per Amount

\$501 - \$1000 had the most frauds with 618 cases

\$1 - \$100 had the second most fraud with 407 cases

\$301 - \$500 had the third most cases with 321 cases

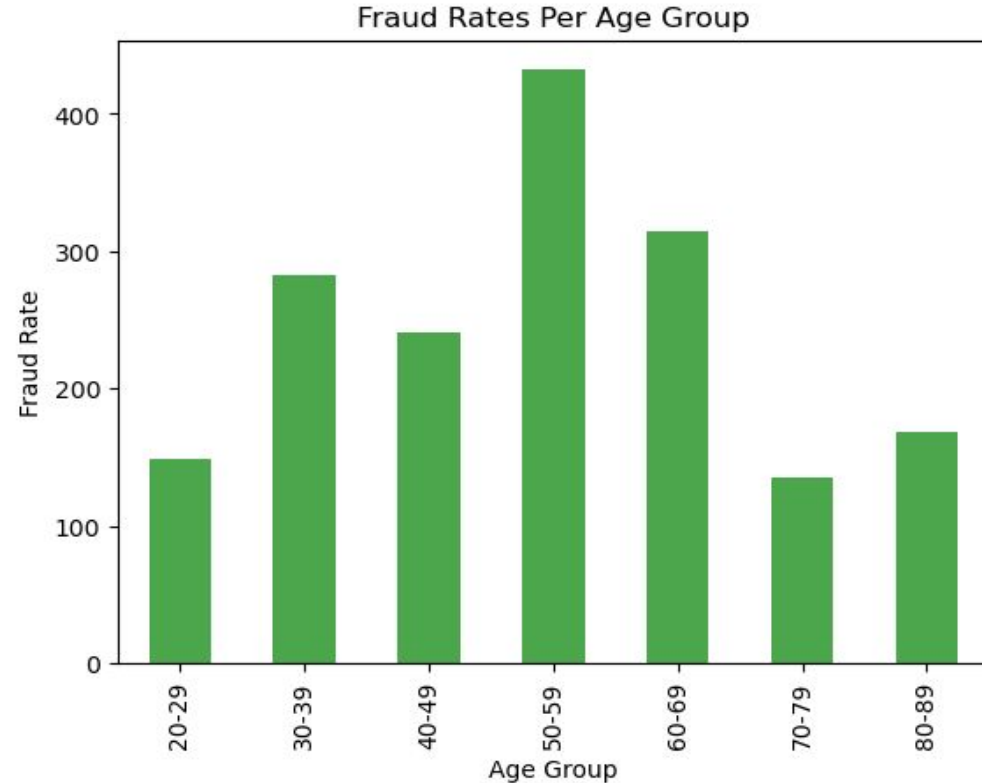


Fraud Rates per Age Group

Age group 50-59 had the highest fraud cases with 432 cases

Followed by 60-69 with 315 cases

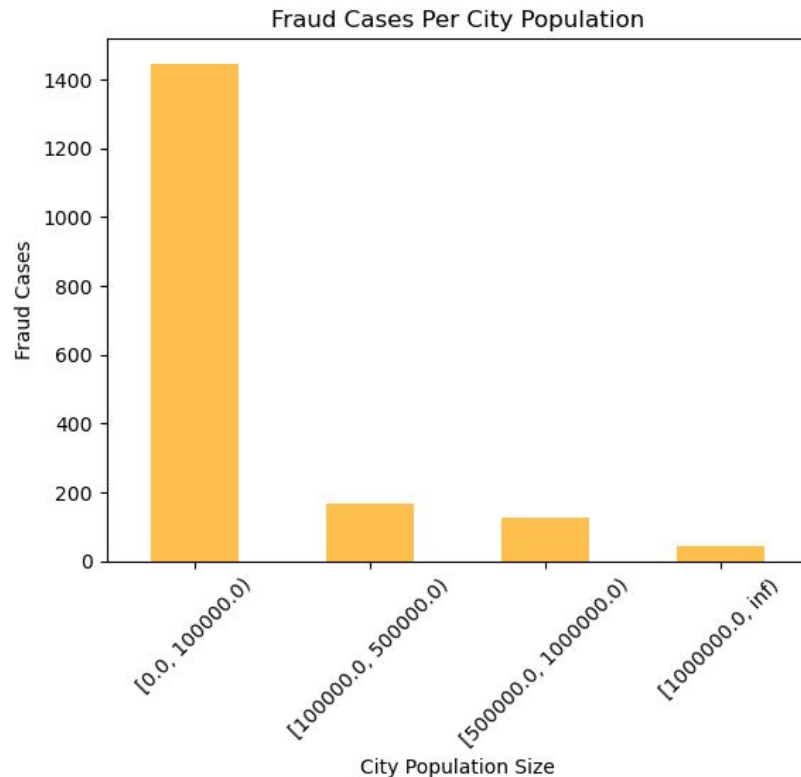
30-39 with 283 cases



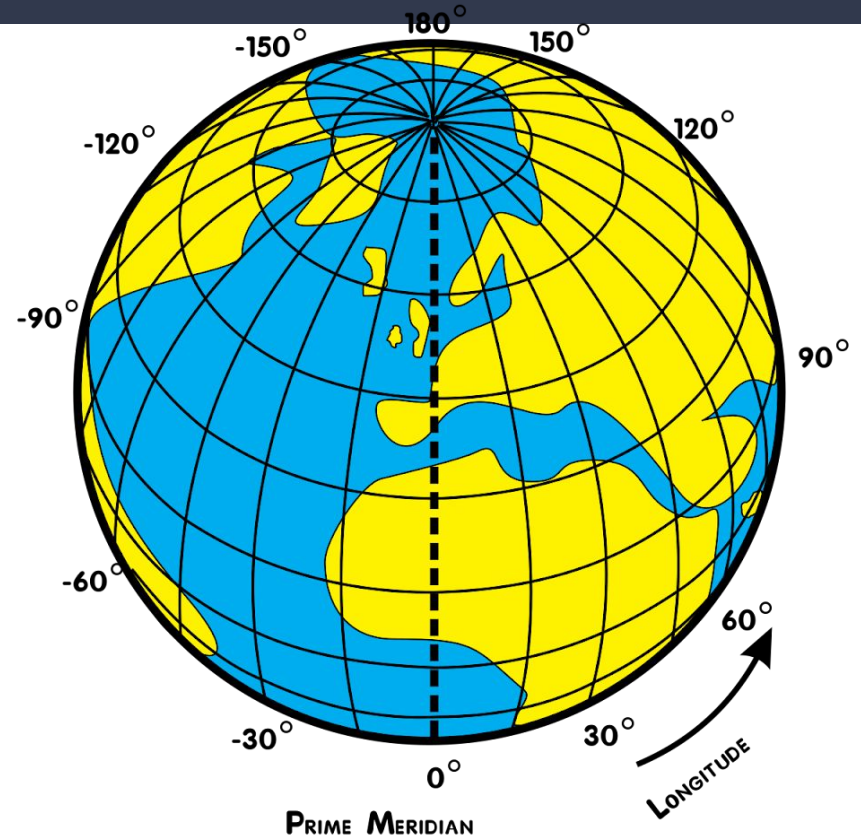
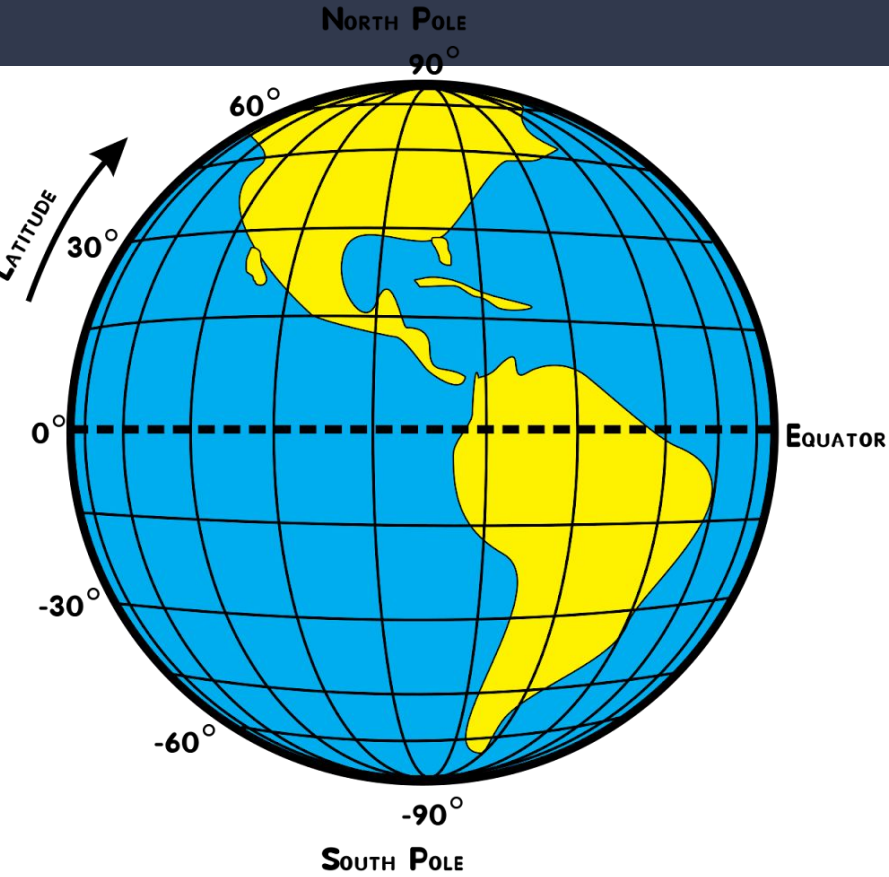
Fraud Cases by City Population

Cities with populations of 100k or less had the most fraud cases with 1446 out of 1782 total cases (81.1%)

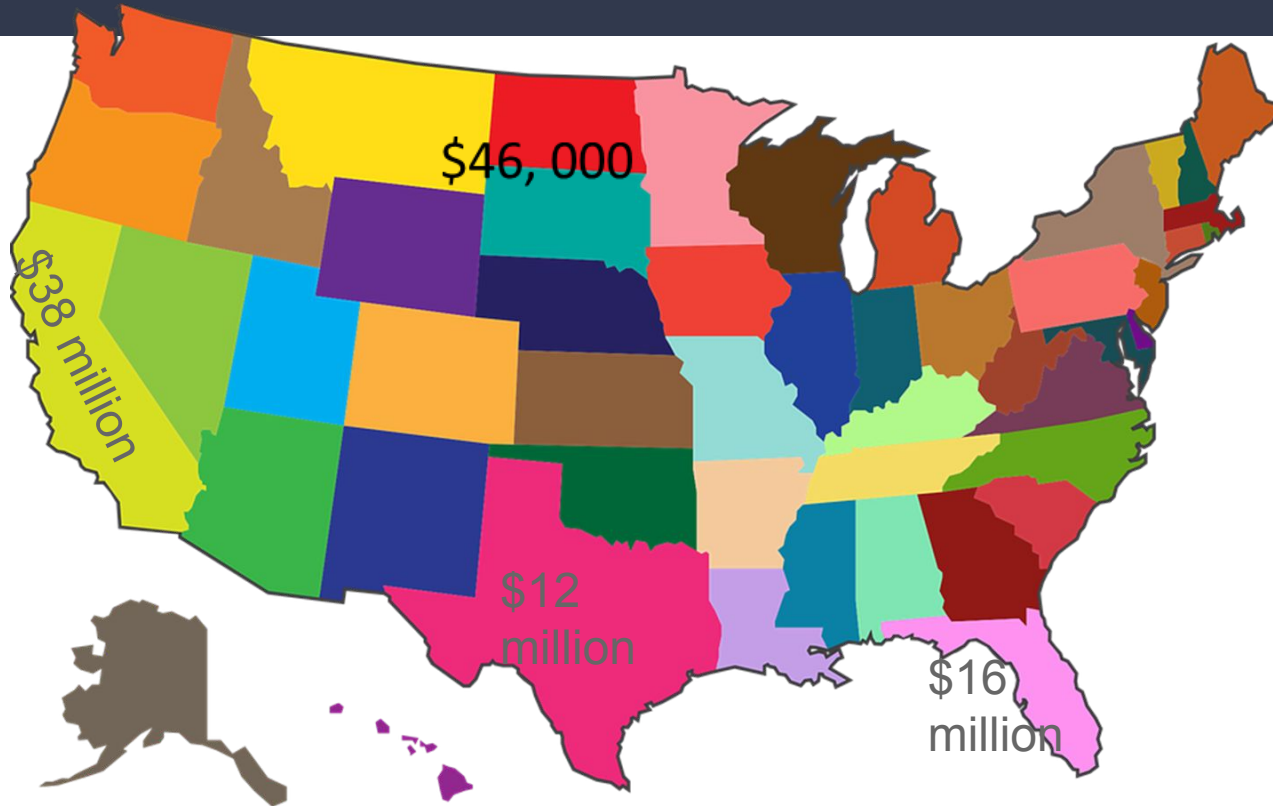
Followed by cities with 100k-500k with 168 cases, 500k- 1 million with 124 and cities over 1 million population with 44 cases



Longitude & Latitude



Fraud rates per city and state



Reverse Geocoding API

Discussion about using Reverse Geocoding API that converts lat/long coordinates to a location address: During the first call to the Reverse Geocoding API I was able to successfully complete 10470 calls before Connection Error msg shown below:

ConnectionError: ('Connection aborted.', ConnectionResetError(10054, 'An existing connection was forcibly closed by the remote host', None, 10054, None))

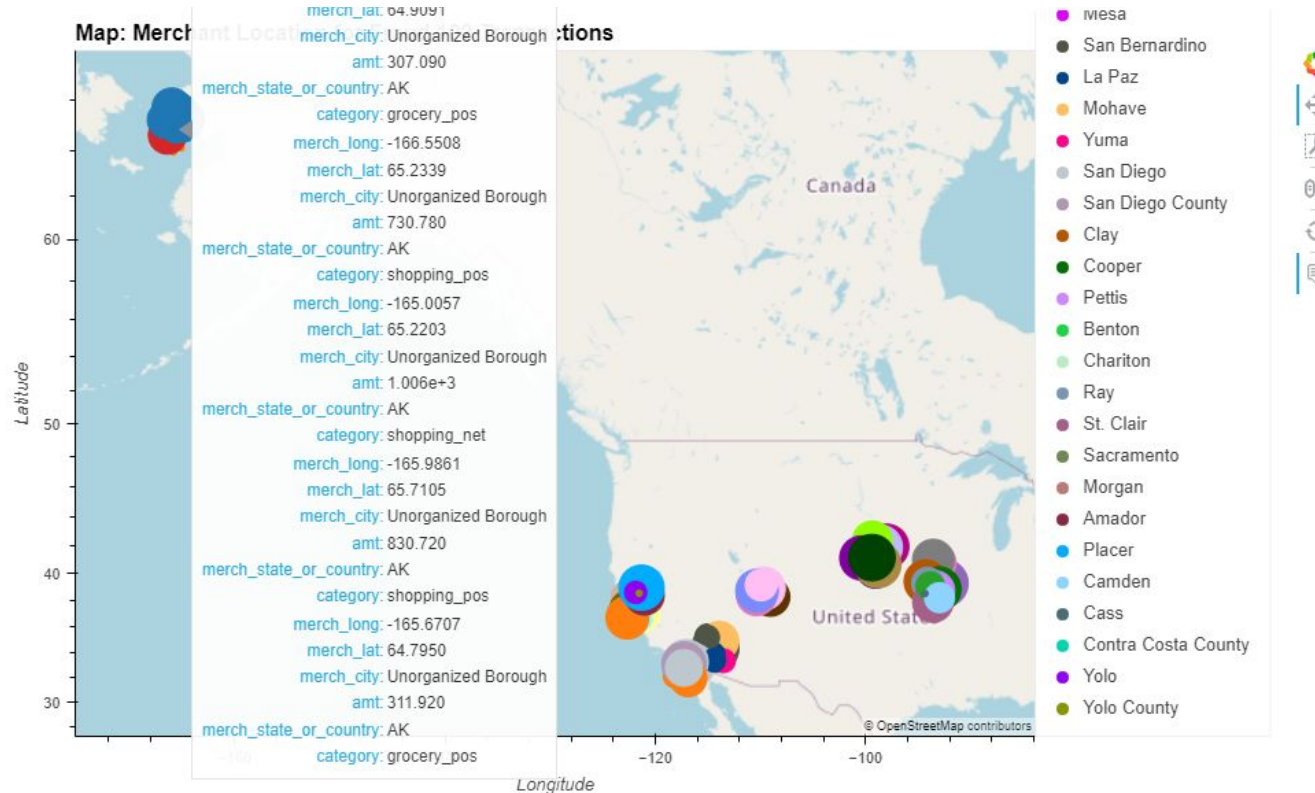
Thus, only 10470 rows with 594 rows of missing location will be used for further data analysis.

Key Learnings:

- Need to select the subset of the data to make < 10,000 calls or obtain industry license from Reverse Geocoding API
- API calls take a long time: It took 8 hr to complete 10470 calls

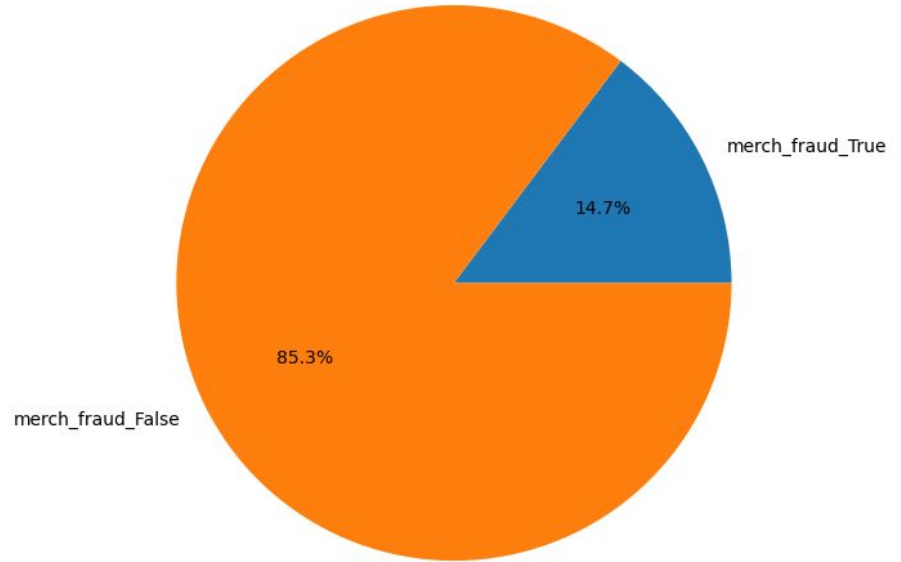
	merchant	category	amt	merch_city	merch_state_or_country	merch_lat	merch_long
10465	Cole, Hills and Jewess	home	19.66	Apache County	AZ	35.926855	-108.633895
10466	Christiansen-Gusikowski	misc_pos	671.79	Colfax	NM	36.301140	-104.973318
10467	Gislason Group	travel	7.89	Gasconade	MO	38.268495	-91.584583
10468	Hudson-Grady	shopping_pos	2.71	Perkins	NE	40.796990	-101.441473
10469	Streich, Rolfson and Wilderman	kids_pets	50.28	Solano	CA	38.425746	-122.081592

Map: Merchant Location for 102 Fraud Transaction



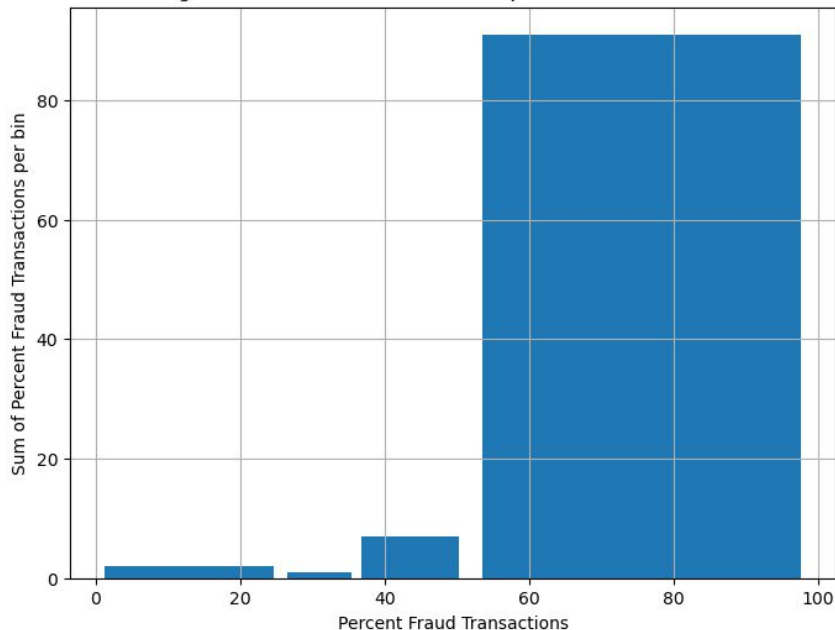
Fraud Transactions per Merchant Name Summary

Pie Plot of 692 Unique Merchant with at least 1 Fraud Transaction



Fraud Transactions per Merchant Location Summary

Histogram of 102 Fraud Transactions per Merchant Per Location



	merchant	merch_state_or_country	merch_city	count_fraud	count_total	%_fraud
66	Murray-Smitham	No country name is found	No city name found	2	4	50.000000
29	Heller-Langosh	No country name is found	No city name found	1	4	25.000000
77	Rau and Sons	No country name is found	No city name found	1	4	25.000000
46	Koepp-Witting	No country name is found	No city name found	1	3	33.333333
10	Beier-Hyatt	No country name is found	No city name found	1	2	50.000000
36	Kiehn Inc	No country name is found	No city name found	1	2	50.000000
39	Kihn, Abernathy and Douglas	CA	San Diego	1	2	50.000000
42	Kilback LLC	No country name is found	No city name found	1	2	50.000000
68	Nicolas, Hills and McGlynn	UT	San Juan	1	2	50.000000
86	Schoen, Kuphal and Nitzsche	CA	San Diego	1	2	50.000000
0	Abbott-Rogahn	MO	Camden	1	1	100.000000
1	Ankunding-Carroll	CO	Mesa	1	1	100.000000
2	Barrows PLC	MO	Grundy	1	1	100.000000
3	Bartoletti and Sons	UT	Grand	1	1	100.000000
4	Bauch-Rayno	No country name is found	No city name found	1	1	100.000000

Questions?