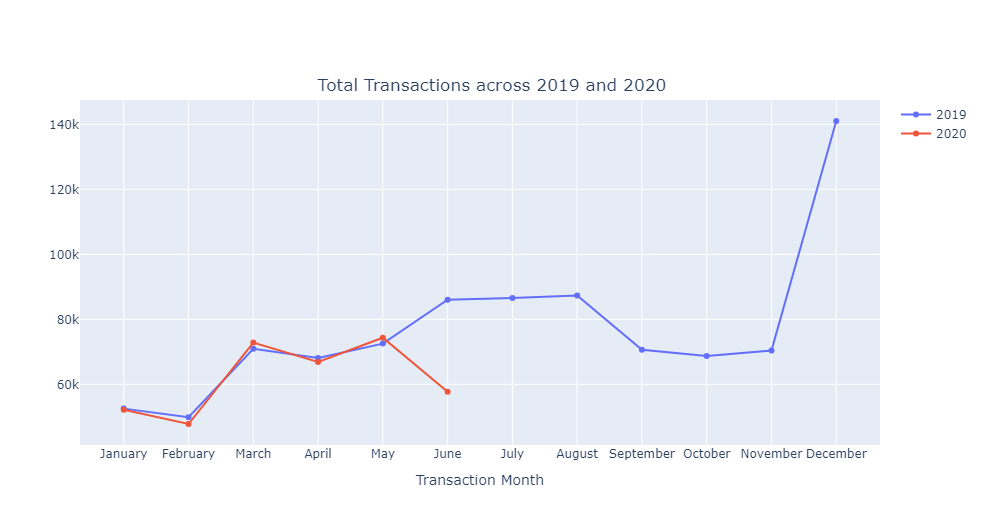
**Analysis of Credit Card Transactions**

Throughout the analysis of the dataset, I have taken a question-answer based approach to understanding the data. Below are the insights I gained from the analysis:

The dataset contains roughly **1.3 million** transactions from **983 distinct individuals** which dates from the 1st of January, 2019 through 21st June, 2020**.** A total of 924,850 transactions were completed in 2019 and 371,825 in 2020.

**Barbara Taylor and Monica Cohen performed the highest number of transactions**, each completing a total of 3123 transactions. The number of distinct credit card number also equals 983 which matches the number of distinct individual - this means that no person has more than one credit card.

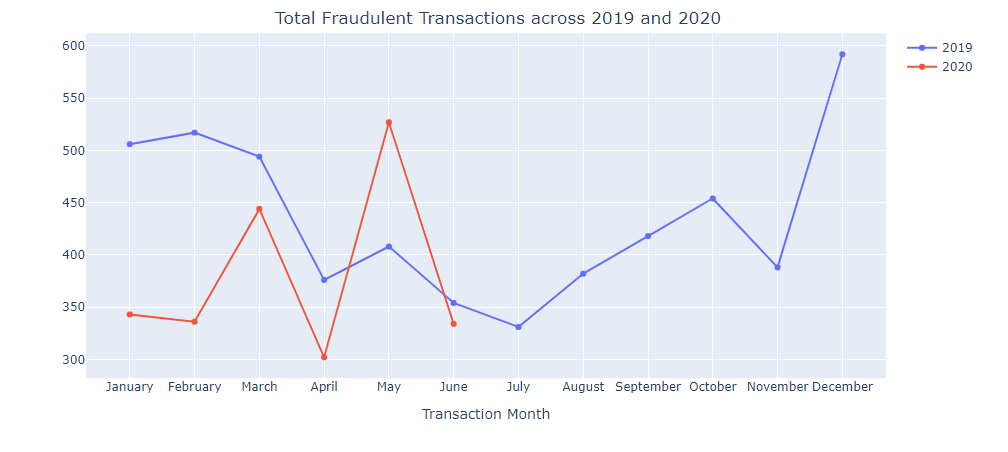
The highest amount in a single transaction is 28948.9 - performed by Kelly Zimmerman on the 1st of December 2019.



The number of transactions performed each month from January to June in both year differ by a very minor amount relatively. In fact, roughly 52,000 transactions were completed in both January 2019 and 2020. Also, **the highest number of transactions was completed in December 2019 (a total of 141,060).** A somewhat interesting observation was also seen in the total transactions completed on June, July and August of 2019. The transactions completed on these months are almost equal hence the ‘straight line’ in the graph. In addition, the total amount across these months is rougly **6M** each.

The distribution of legitimate transactions closely resembles that of total transactions, with slight variations per month which is expected due to a highly imbalance proportion of both classes. However, fraudulent transactions follow a different trend. **In 2019, there was a gradual decrease in fraudulent transactions from January to November, followed by a sharp high increase in December.**

Additionally, throughout 2019, the number of fraudulent transactions exceeded that of 2020, except for May, where 527 fraudulent transactions occurred in 2020 compared to 408 in 2019.



Contrary to what we might expect, there appears to be no correlation between people’s age and whether they get defrauded or not.

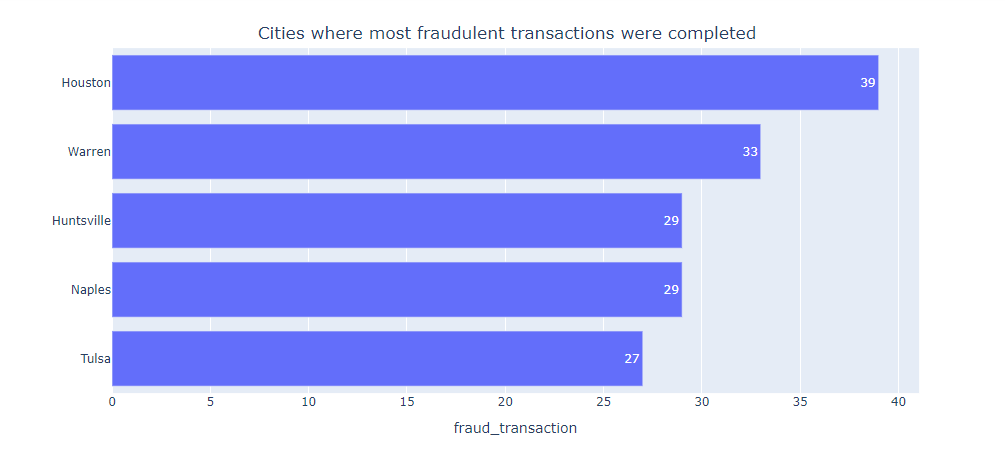
Despite females having a significantly higher number of completed transactions, the number of fraud victims is higher among males than females.

The table below shows cities with most legit transactions completed

|  |  |
| --- | --- |
| **City** | **Legit Trans** |
| Birmingham | 5606 |
| San Antonio | 5105 |
| Utica | 5080 |

The table below shows cities where most fraudulent transactions are completed

|  |  |
| --- | --- |
| **City** | **Legit Trans** |
| Houston | 39 |
| Utica | 33 |
| Huntsville | 29 |



Other insights could be gained by exploring the job, state and other columns and whether they determine if a transaction is legit or fraudulent. The process however follows a similar approach used in the notebook.