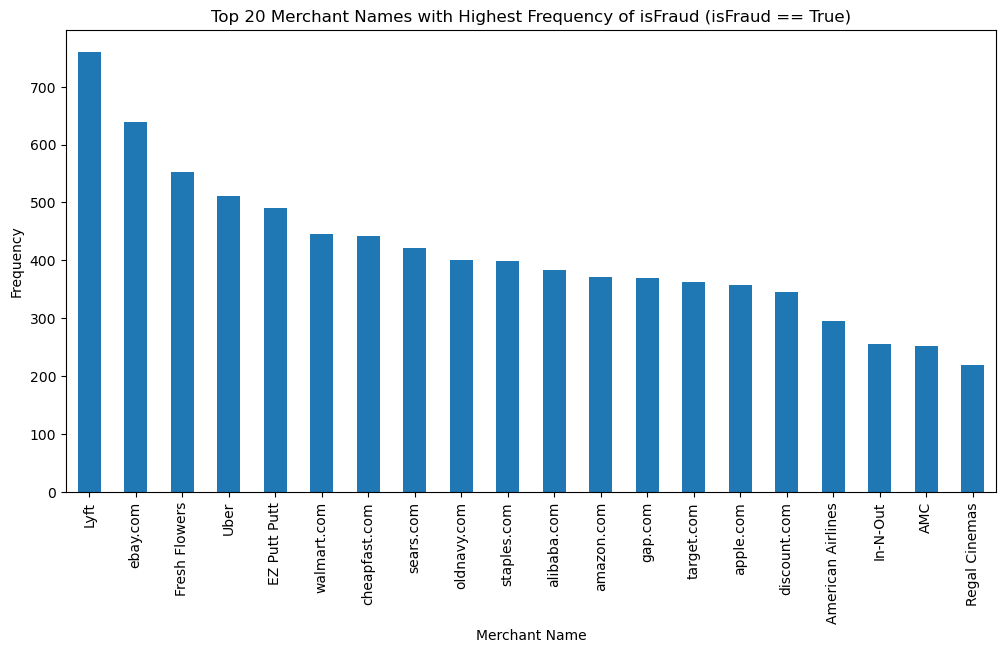
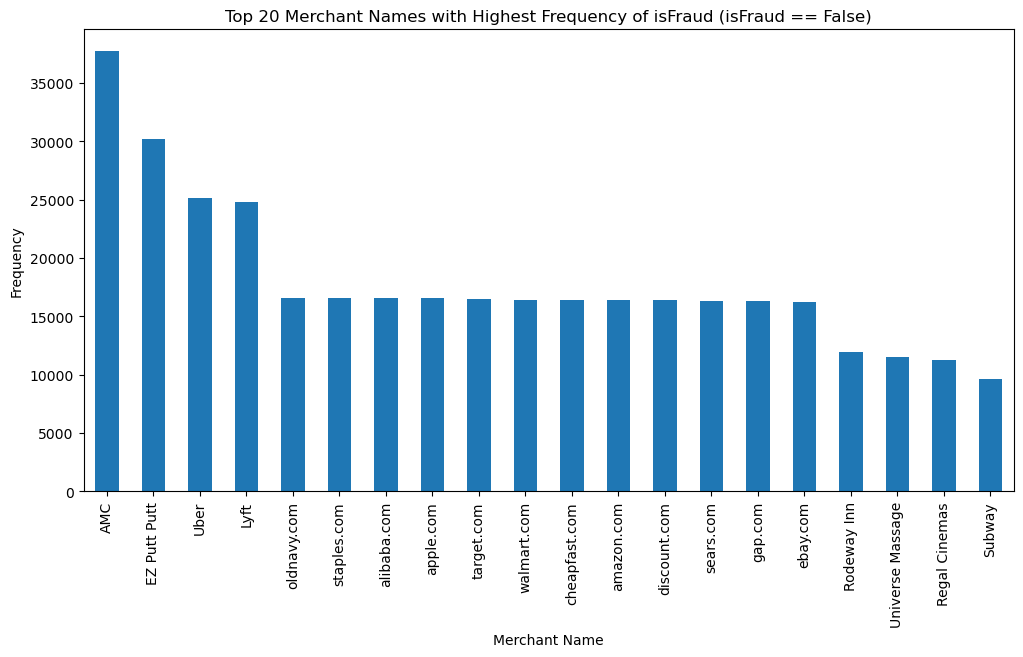
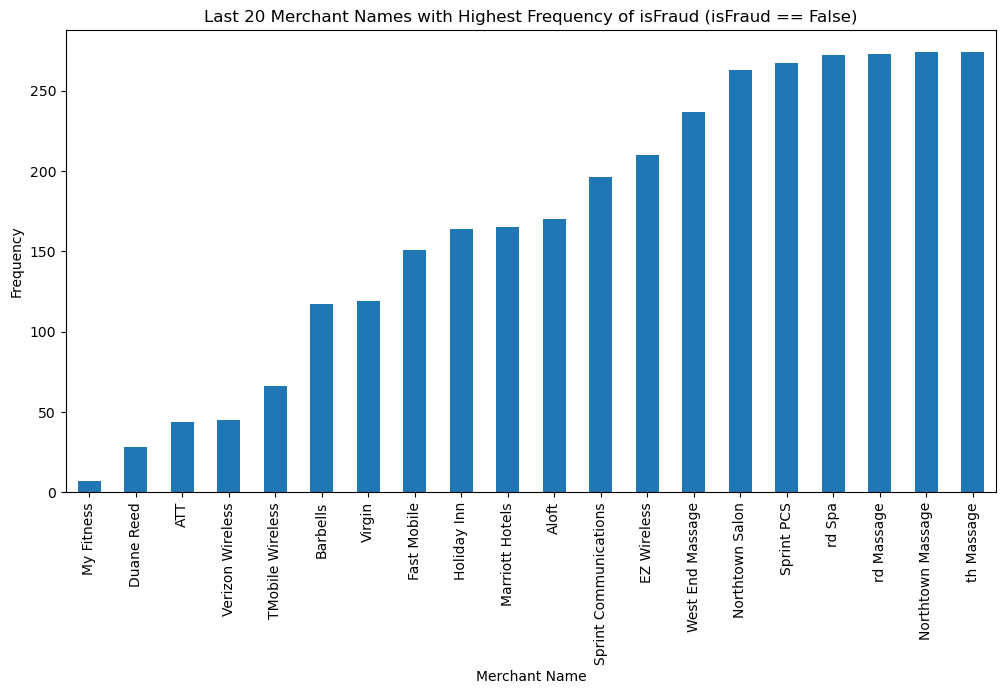
Inferences



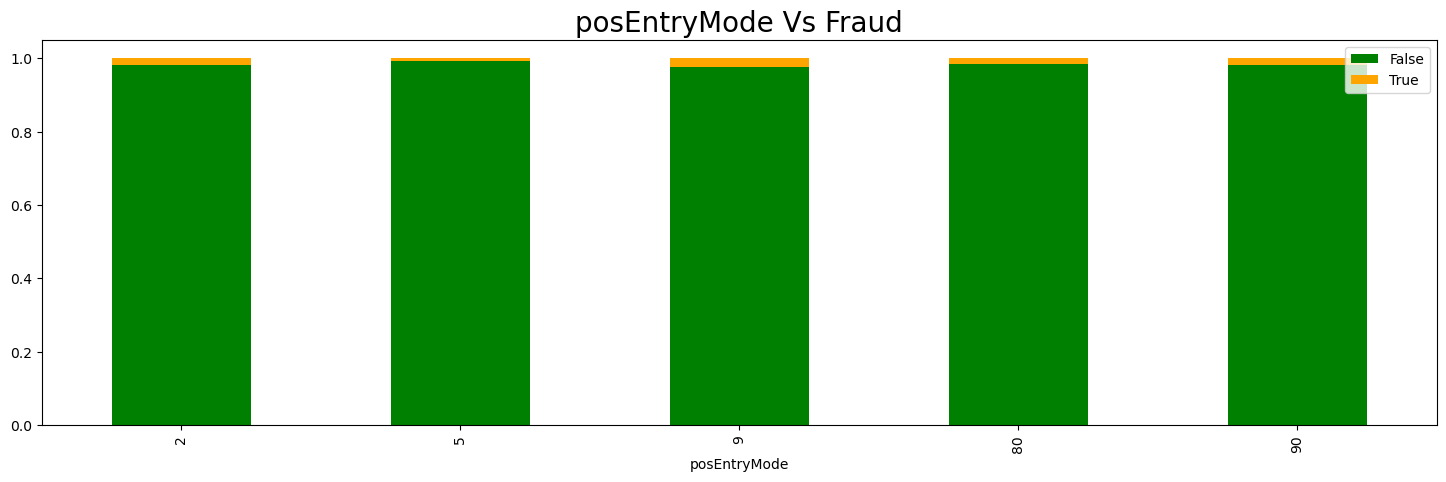
* Highest number of fraudulent transactions is done by ‘Lyft’ out of all other merchants.
* High number of merchants have done between 300-500 transactions.



* Highest number of non-fraudulent transactions is done by ‘AMC out of all other merchants.
* High number of merchants have done between 15000-20000 non fraudulent transactions.

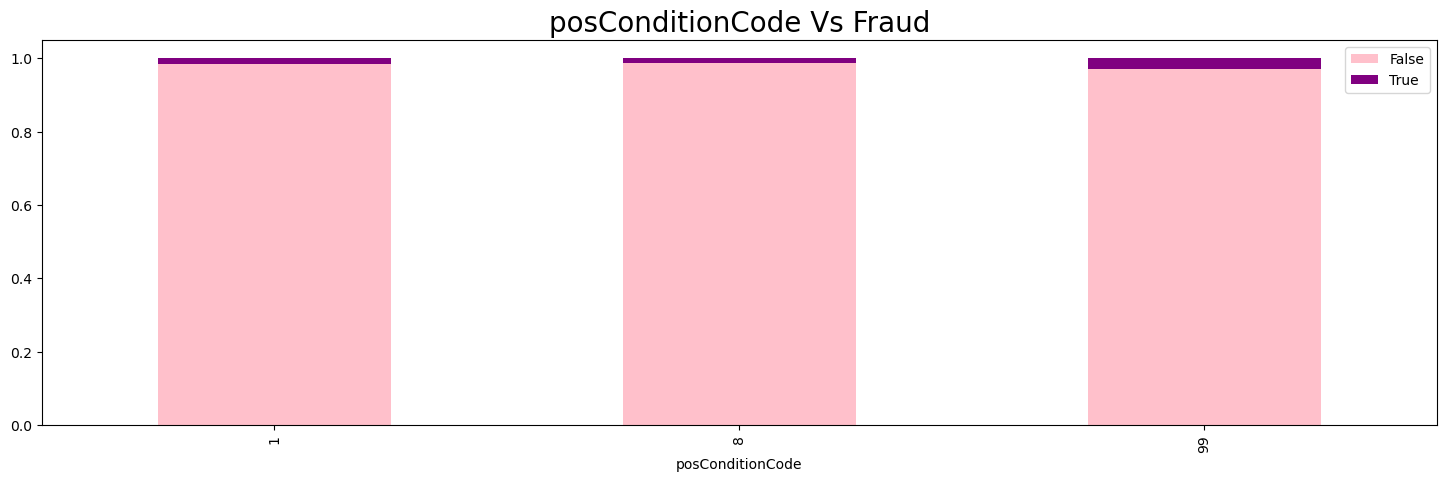


* Lowest number of non-fraudulent transactions is done by ‘MyFitness’ out of all other merchants.



* All posEntryMode has fraud transaction in it, hence transaction being fraud or not is very less related with the posEntryMode

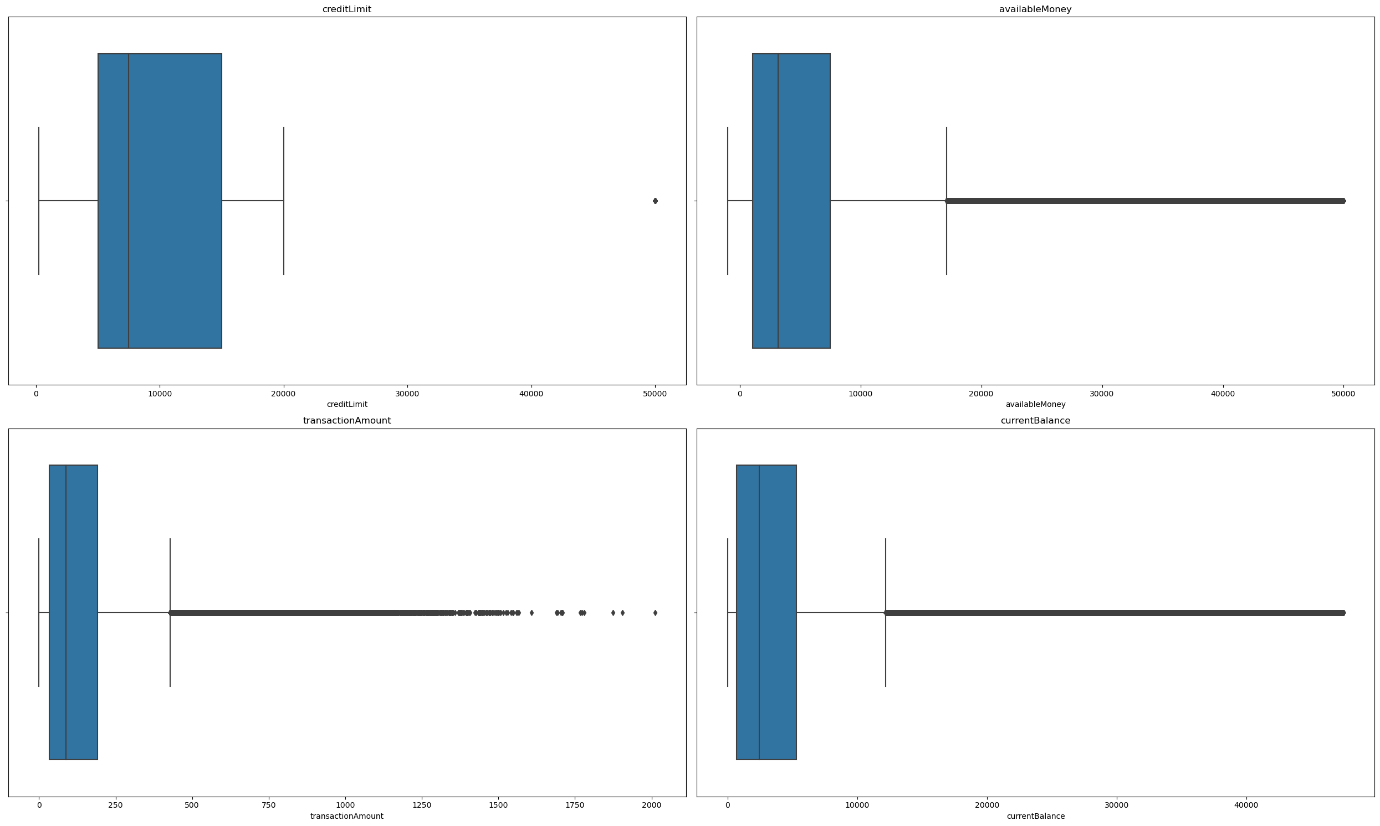
Note: This two-digit code identifies the actual method used to enter the cardholder account number and card expiration date



posConditionCode legend

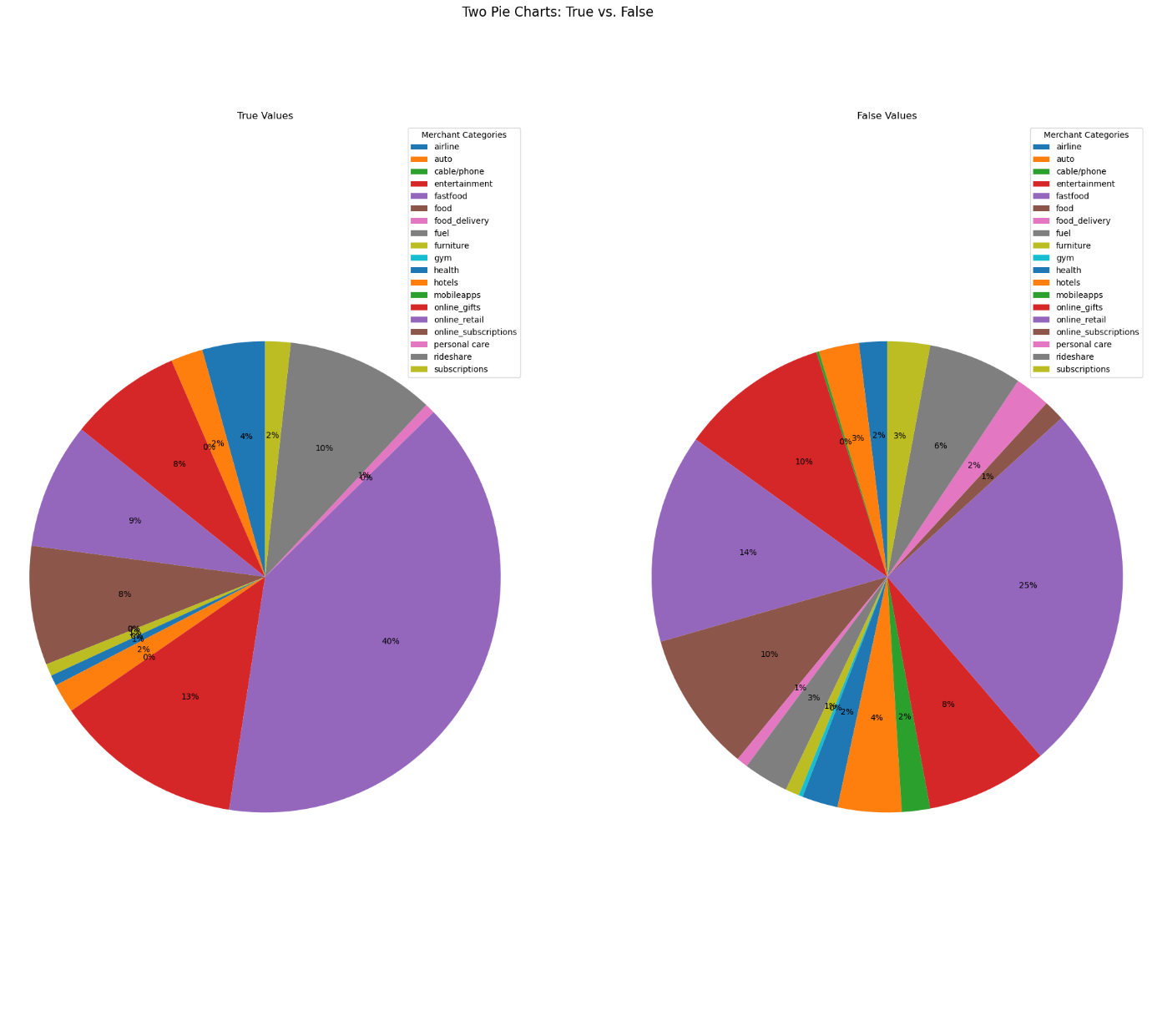
|  |  |
| --- | --- |
| 01 | Cardholder not present |
| 08 | Mail/telephone order (includes Visa phone and reoccurring transactions) |
| 99 | Other Place of Service Other place of service not identified above. |

* Fraudulent transaction has happened only in posConditionCode 01,08 and 99 (there are many other codes).
* Out of the three a slight increase in fraudulent transaction in posConditionCode 99 can be observed.

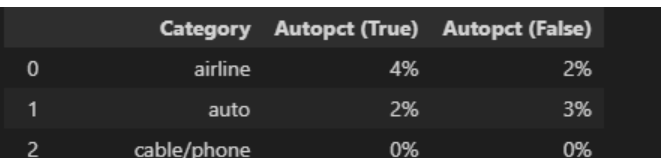


Observation regarding Outliers

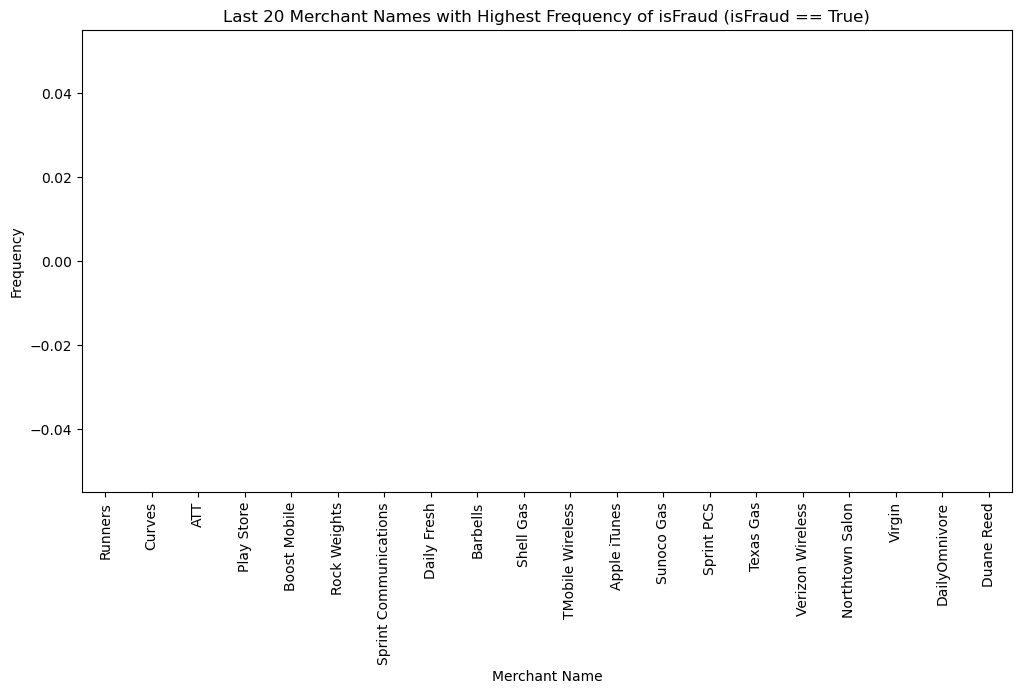
* Very less outliers are present in creditLimit column
* It can be observed that the number of outliers is comparatively very high in availableMoney, transactionAmount, currentBalance columns.



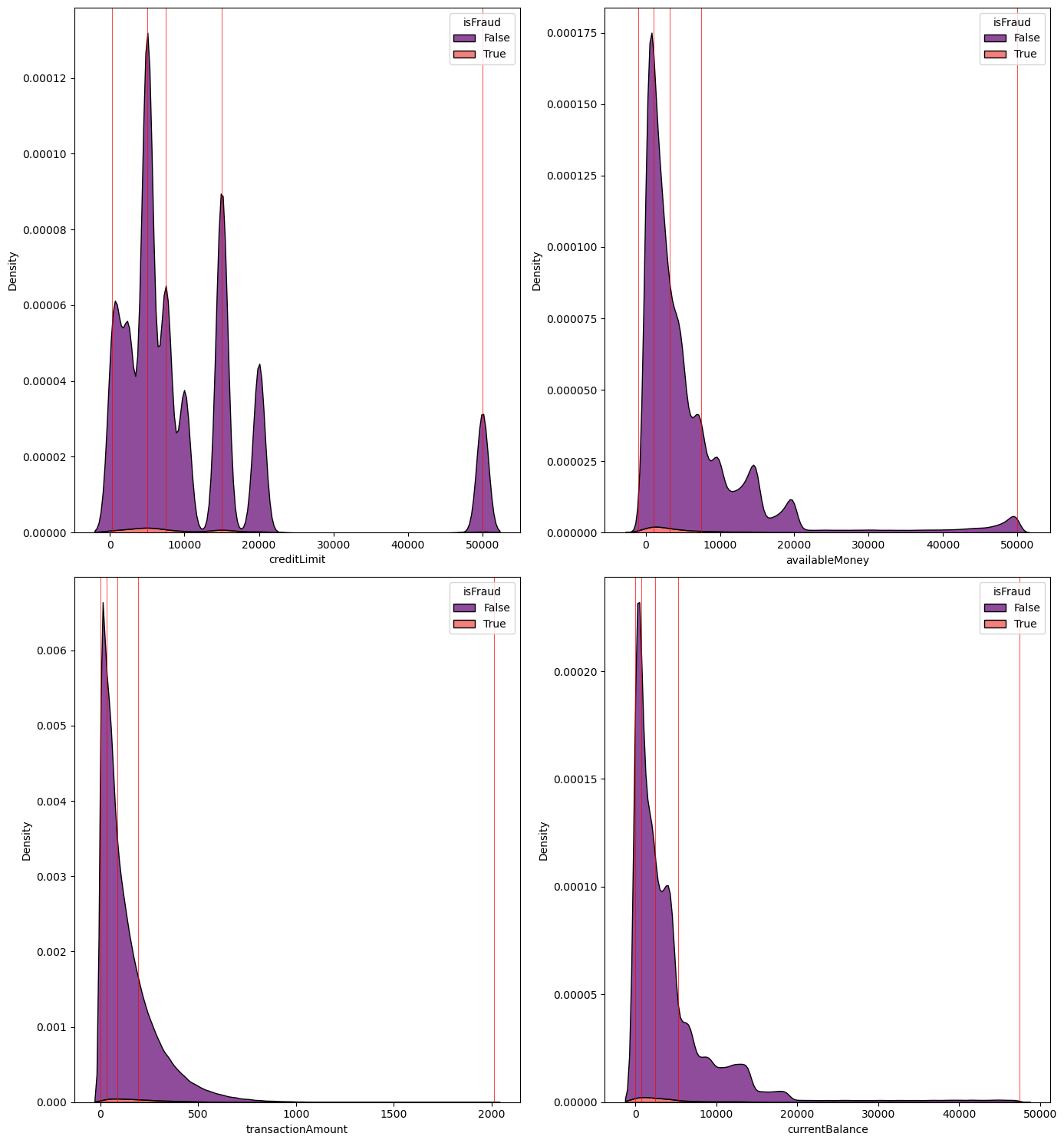
Assuming that these are the representation of number of transaction which were fraud (left)/notFraud(right) grouped categorically (because mail it is mentioned autopct true/false in mail) please check



* Highest number of transactions was done by online retails, which also happens to be where the highest number of fraud transactions also taken place.
* Second highest number of fraud transaction is observed in fastfood category, followed by food,entertainment and onlinegifts.
* All other sectors seems to have similar (comparatively very less) share of number of fraud transactions.

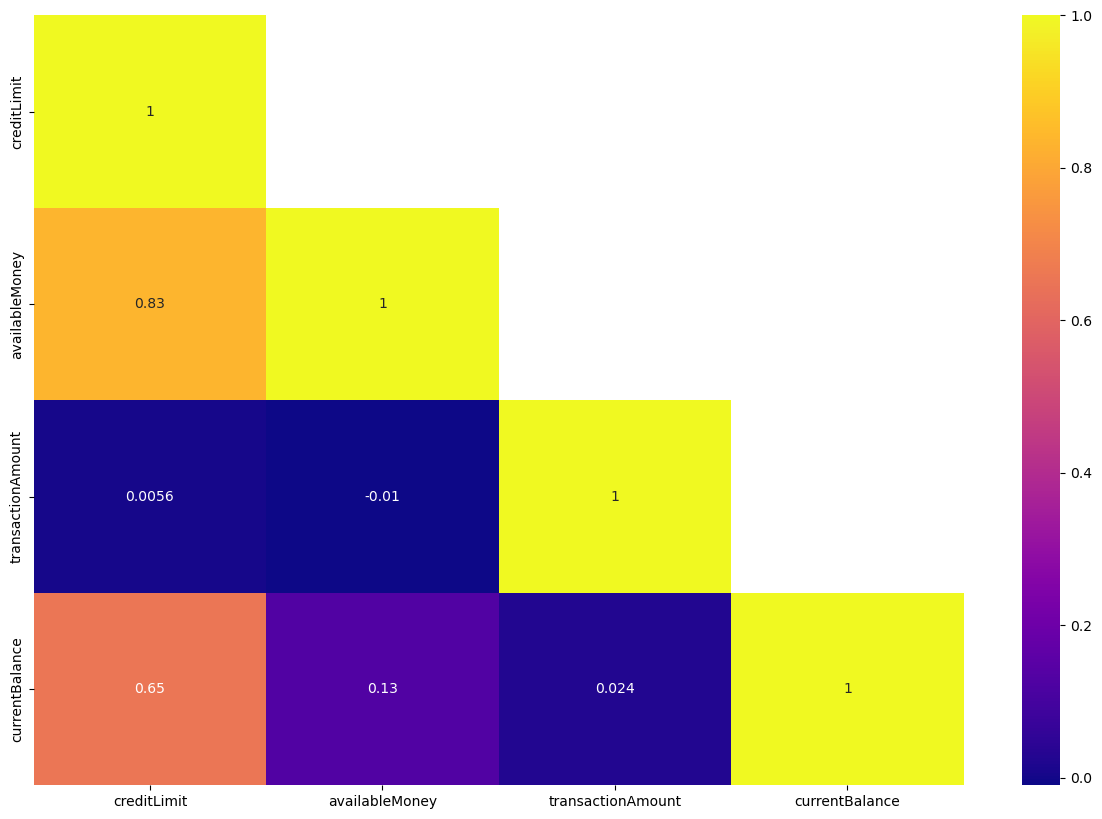


?????????????????????????????????



Inference regarding Distribution of availableMoney, transactionAmount, currentBalance,creditLimit variables.

* From the plot, it is evident that None of the features are normally distributed.
* All (availableMoney, transactionAmount, currentBalance,creditLimit) features are heavily right skewed.
* In all cases fraudulent transaction has happened in all the spectrum.
* It can be observed that creditLimit has multiple peaks

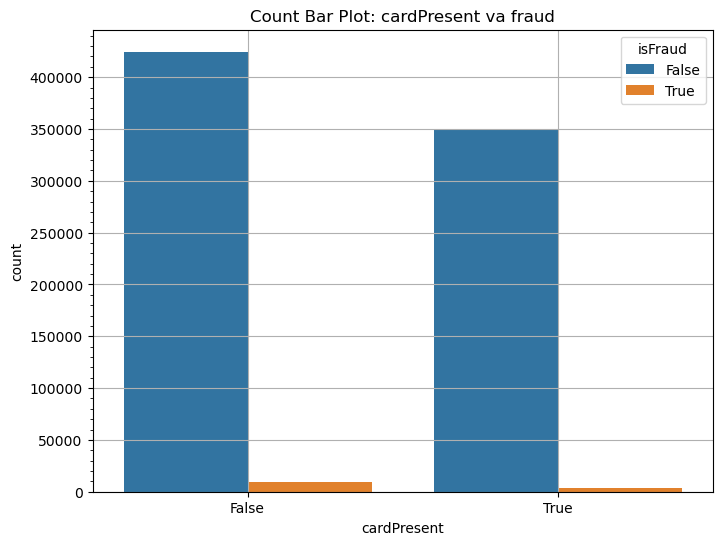


Inference regarding Correlation between (availableMoney, transactionAmount, currentBalance,creditLimit)

* It can be observed that credit limit and available money has a notable level of correlation.

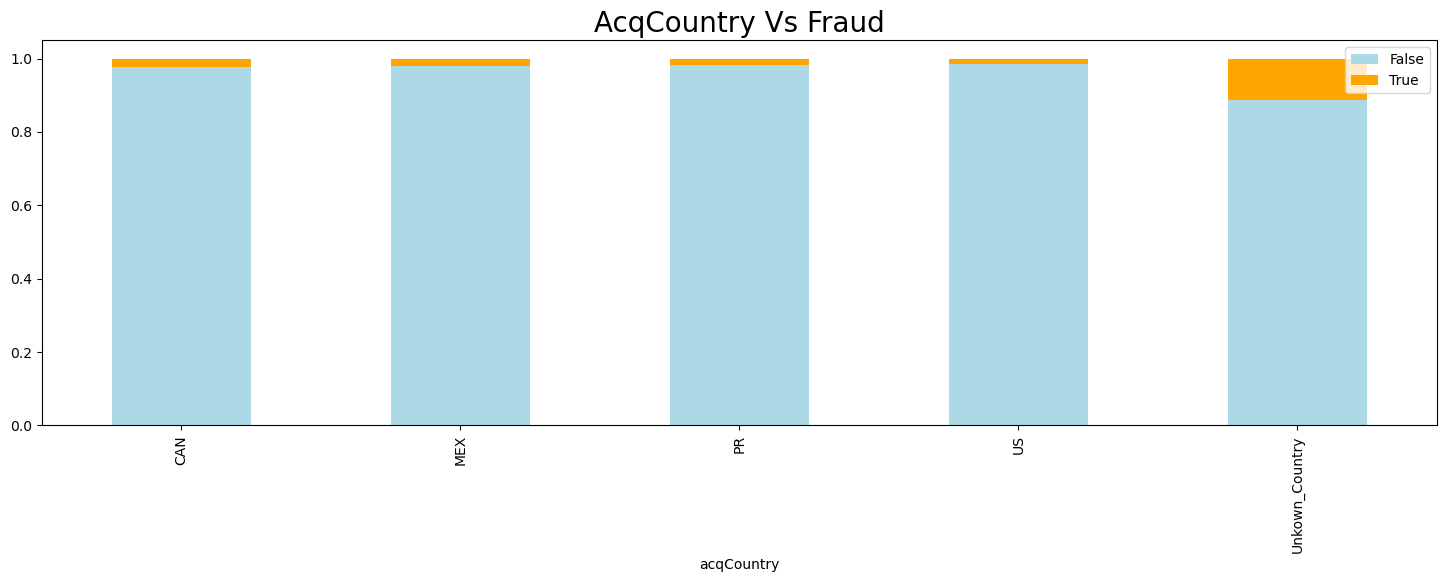
(It is because banks usually provide credit limit to an individual based on the income/available money of the individual)

* Similar trend can also be observed in tha case of creditLimit and currentBalance



* It can be observed that fraud transaction has happened irrespective of card being physically present in the transaction or not.
* Comparatively higher number of fraud transaction has happened without the physical presence of card.

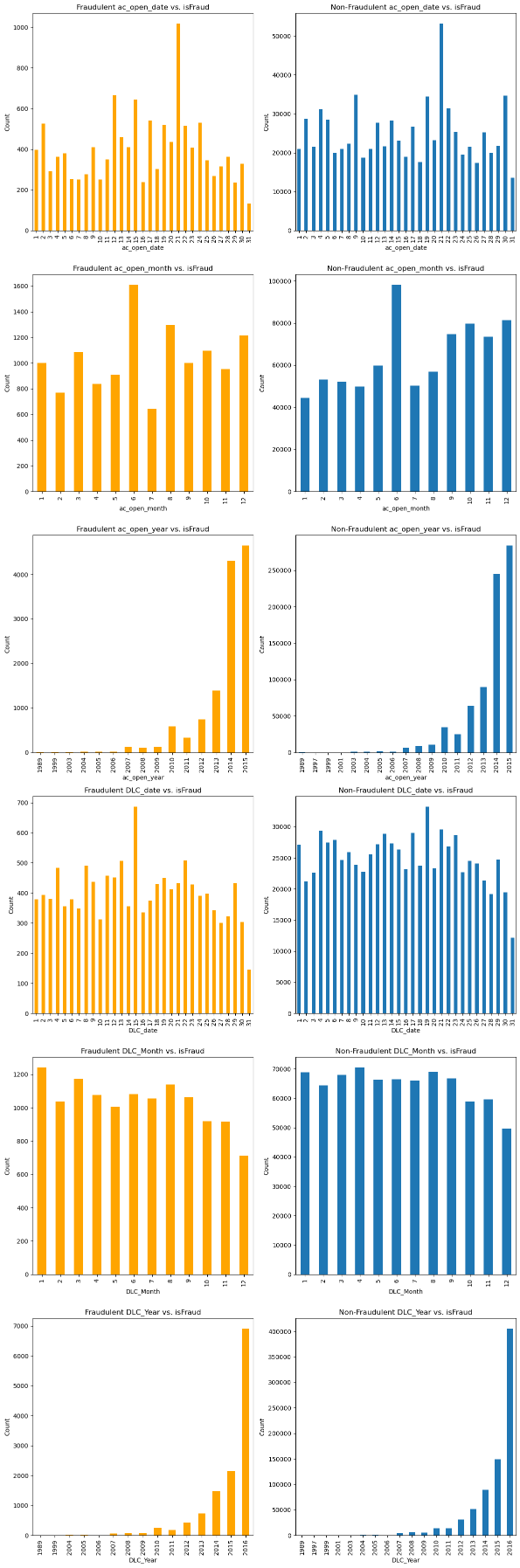
Note: cardPresent = The physical presence of the debit card in The transaction



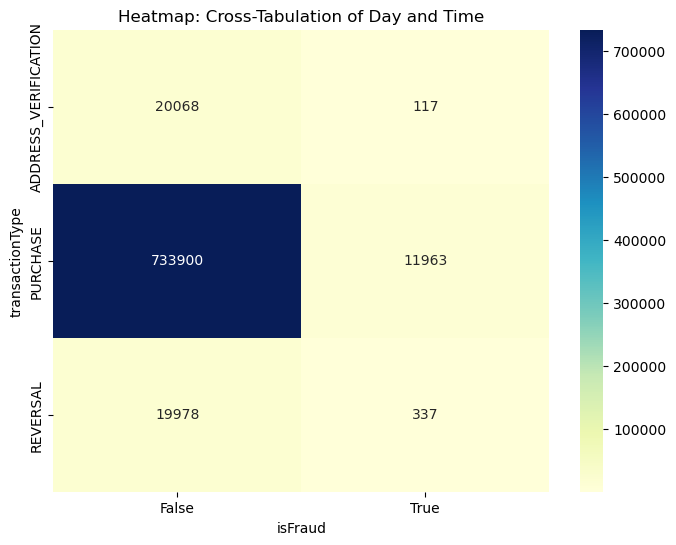
Observation based on countries.

* It can be observed that most of the merchants in the dataset are located in CANADA Mexico, US and Puerto Rico.
* There are observations in which the merchant country code in unknown.
* Merchants from all country is observed to have similar level of fraudulent transactions happening.
* Merchants from countries which are not mentioned sems to have comparatively higher fraudulent transactions.

Note: acqCountry = The country where The merchant is located



* Number of overall transaction as well as fraudulent transaction has increased severely in the year 2014 and 2015.
* Fraudulent transaction does not seem to have any relation with month in which the transaction is happening, i.e., fraudulent transaction is happening in all months.



* Highest number of transactions has happened in ‘purchase’ transaction type, which also has the highest number of Fraudulent transactions.
* Lowest number of fraudulent transactions shas happened in ‘Address Verification’ type.
* Proportion of fraudulent transaction in address verification type is comparatively less (0.5%) than other two categories(1.6%)

Note: transactionType = The types of transactions