- Firstly there are 80 unique first names and 100 unique accounts and 100 unique cutomer_id(confirms the number of hypothetical customers).

- The location of the customer(i.e long_lat) is always known(not null) while the other locations(i.e merchant_suburb,'merchant_state','merchant_long_lat')
  are null at times,I confirmed the reason the other locations were null is beacuse it was an online transaction hence no location for the merchant or need for one.
  (also explains why there was no merchant_id for those transactions)

- I also observed 0.0 was used instead of np.nan which made it impossible to confirm the whole column was null.

- There were no values at all for bpay_biller_code and merchant_code.

- The card_present_flag is always null for online transactions.

- I used the following formulas to calculate the outliers: low_outliers = q1 - (1.5*iqr) (any value below this value is an outlier)
high_outliers = q3 + (1.5*iqr) (any value above this value is an outlier)

- There are no low outliers for the data set because the results were negative and the data set is all positive.

- There are high outliers for the data set, the value calculated was 110.1375 from the calculations.

- I found out that the outliers are about 15% of the total transaction volume but 87% of the total transaction amount.

- The standard deviation for the data set which is the measure of the amount of variance or dispersion of a data set is very high about 590 because of the outliers.

- The standard deviation without the outliers is however relatively smaller about 19 and the mean is also smaller i.e 28.6 compared to 188 for the entire data set,
  this indicates that the outliers increases the mean and standard deviation considerably without increasing the transaction volume(just 15% of total transactions).

- The outliers will skew any attempt to analyse data, make the data more varied and also give you unreliable average transaction amounts amogst other things.