

**## To do**

1) obtain most frequent customers by grouping transaction id by customer id through count (frequent buyers)

2) define list\_price - standard\_cost  as profitability

3) Identify high value customers for approved orders profit(summed after grouping the dataset through customer\_id) vs customer\_id

4) which brand has high profitability

5) customer segmentation(Demographic, Geographic, psychographic, Behavioural)

6) Trgeting (segmentsize, Expected growth , competitors offering)

7) Association analysis

**Sales channel analytics**

**Customer segmentation analytics**

**Web analytics**

**Social media analytics**

**Customer engagement analytics**

**Customer churn analytics**

**Customer acquisition analytics**



1) There are some missing values in the following columns. The number of missing values per column is shown below. These missing values are needed to be removed or if possible, filled with suitable values for further analysis

2) The column "product\_first\_sold\_date" has values that cannot be interpreted as a date. Example values are 41245.0 ,37659.0 etc.  
This column should be removed

3) The column gender represents the gender class with different notations  
Females are given values as "F", "Femal" and "Female". similarly males are "M", "Male". This has to be corrected by using same value for a given gender for a consistent representation. similarly new south whales and victoria states have other instances with values like 'VIC' and 'NSW'

4) The column "default" has many absurd values and should be removed from the dataset  
This column has values that cannot be interpreted or used for any analysis. This column has to be removed from the data set

5) The column country has zero variance and is not usefull for analysis  
As your all your customers are from the same country it can be safely removed.

6) The customer with customer\_id = 34 was born in the year 1843. This is absurd  
This could be an error and needs to be checked