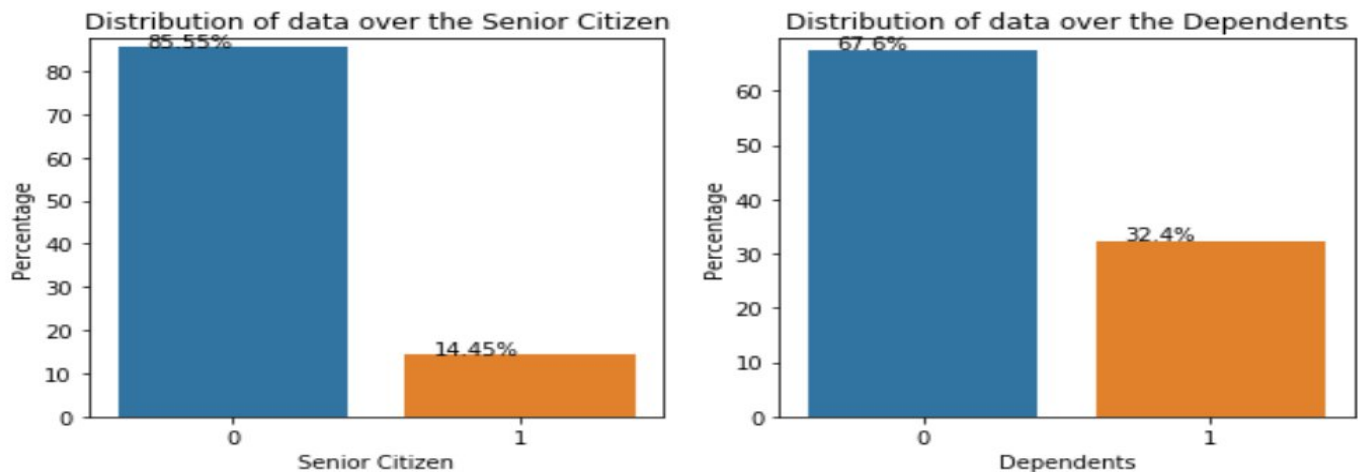


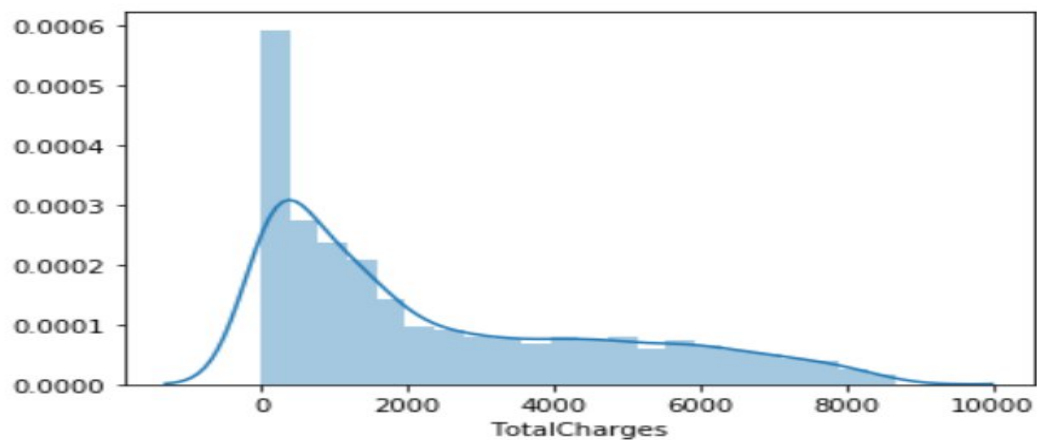
HA4 Report

1) Distribution of data over the Senior Citizen and Dependents



Only 14.45% customers are senior citizen in the data-set and rest of 85.55% customers are not, this observation is quite reasonable. Most of customers are not dependents, only 32.4% are dependents (i.e. a person who relies on another as a primary source of income).

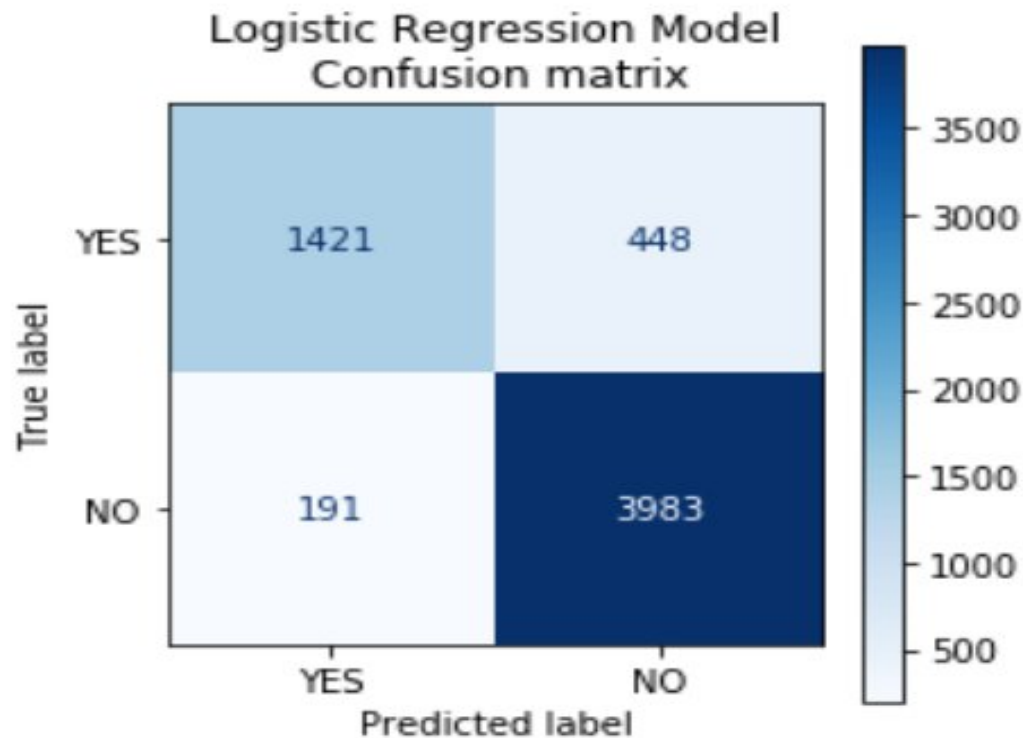
2) Distribution of total charges in a data



Distribution of total charges follows the power law distribution, means number of customers decrease with respect to the increase in total charges.

Power law distribution is quite in Social Network.

3) Confusion Matrix using Logistic Regression Model



Confusion matrix expresses the predictive results of the classification model. It is quite useful when we have imbalanced dataset. True Positive(1421): model predicted positive and it's true, True Negative(3983) model predicted negative and it's true. False Positive(191) Type 1 Error: model predicted positive and it's false. False Negative(448) Type 2 Error mode predicted negative and it's false.

4) In case of both strategy A and B, thresholds for the highest profit and highest profit per customer coincide. Total not churned customers decrease with respect to the thresholds. So, it is quite obvious that profit and profit per customer will attend the highest value at the same time.

5) Number of clients we will lose in case of both strategies:

Churned Customers has been calculated as $TP \cdot (1 - \text{Acceptance_Rate}) + FN$

Count of Churned Client	
Threshold	
0.1	2187
0.2	2080
0.3	1961
0.4	1881
0.5	1782
0.6	1675
0.7	1545
0.8	1383
0.9	1093

Table 2 Strategy-A

Count of Churned Client	
Threshold	
0.1	1925
0.2	1830
0.3	1725
0.4	1655
0.5	1568
0.6	1474
0.7	1359
0.8	1217
0.9	962

Table 1 Strategy-B

Strategy-B seems to be well promising compare to the Strategy-A as we have less number of churned customers in case of strategy-B.