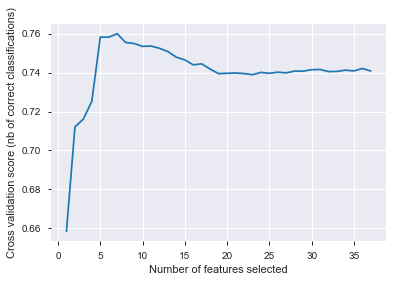
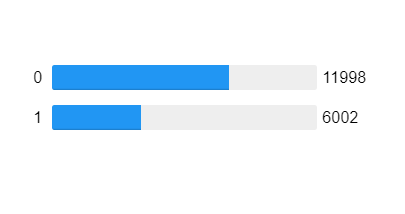
**Steps:**

* Imported Dataset using pandas
* Applied Pandas Profiling to perform EDA.
* Generated three reports and observed their distributions and frequency:
  + Using whole dataset named as Dataset\_EDA.html
  + On Customers Who has churned (Label==1) names as EDA\_label=1.html
  + On Customers Who has not Churned (Label==0) names as EDA\_label=0.html
* Divided dataset into test and train set with ratio 7:3
* Applied Random Forest on train set and then calculated the accuracy at test set which was 74.70%
* Applied Recursive Feature Elimination Cross-Validated (RFECV) feature selection which selects the best subset of features using recursive feature elimination and then selecting the best subset based on the cross-validation score of the model.

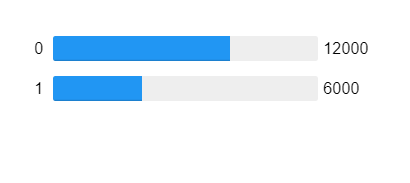


* Optimal number of features is 7 and following are the features that contribute most to the accuracy of the model: **['B1', 'B2', 'C1', 'C2', 'J12', 'J16', 'K']**

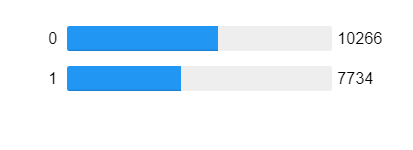
B1:



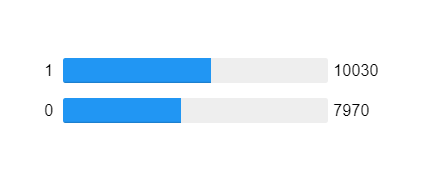
B2:



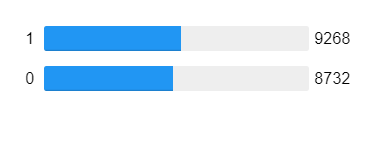
C1:



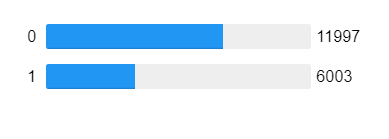
C2:



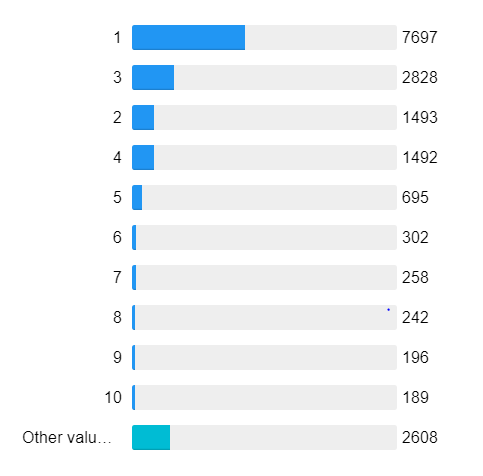
J12:



J16:



K:



* Now using above features divided dataset into test and train set with ratio 7:3.
* Applied **SVM, Random Forest and XGB Classifier** and got **76.30%, 76.76% and 76.74%** accuracy respectively. As you can see there is a gain of 2% in accuracy.
* Applied Feature Elimination with RFE to get top n features.
* “K” is the most discriminating and important factor as it alone gives an accuracy of 65.76%.
* “B2” is the second most discriminating and important factor as it boosts the accuracy to 71.51% and on its own gives 49.92% accuracy.
* Third Most important and discriminating feature is “C2” at it boosts the accuracy to 71.88% and on its own gives 63.94% accuracy.
* Top 5 features are ['K' 'B2', 'C2', 'J12', 'J16'] which gives an accuracy of 76.31%.