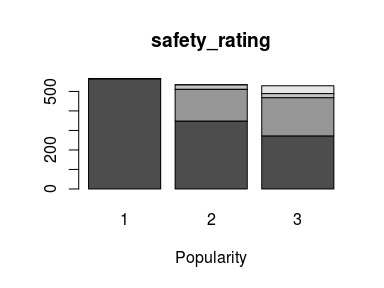
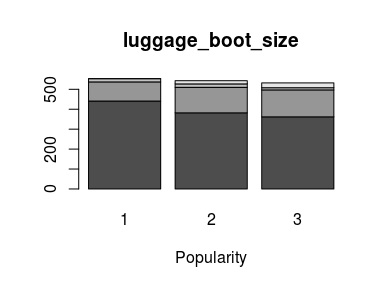
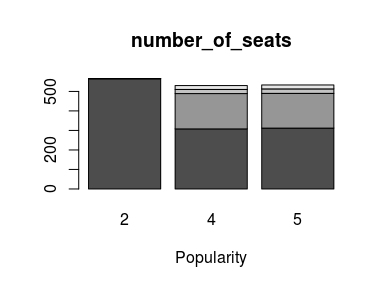
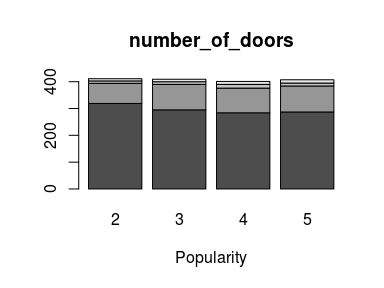
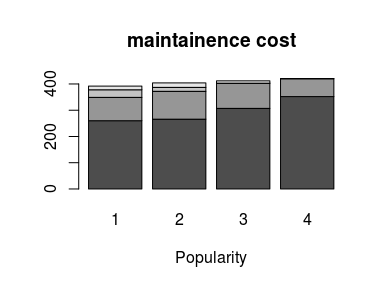
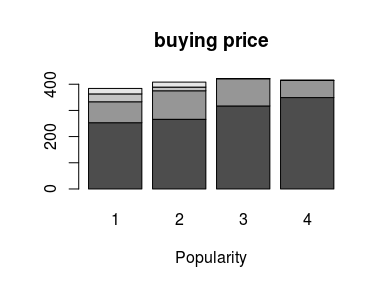
Popularity Prediction

Vidwath R Hebse

11-Feb-2017

# Data Set Description

* Bar plot respresents a table of popularity against other predictor variables on train data, to visualize counts.



## Summary

### Data set quality

* Train and Test data is clean.
* No NA or missing values in the data set.
* As per the column description it had the given key values per column.
* No outliers found in the data.
* Response variable being “Popularity” and rest variables are predictors.

### Data Prepration

* Data was supposed to be processed based on its application.
* For multinomial regression (using nnet) and random forest, data was supposed to be made as factors.
* For XGboost all the data columns should be kept as numeric.

### Model choosen

* As it is a multinomial classification, though of using decision tree, random forest and multinom (nnet) classification as well as XGboost.
* XGboost outperformed compared to other.
* XGboost, Randomforest choosen for the reason of Ensembling as the training data set was not that big, needed to convert week learner to strong learner so XGboost performed better than other.

# Request Note

* Keeping every thing in mind, I still was unable to get 1000 result. It would be helpful for me If i get to know where I went wrong, so that I can correct my approach.
* Thanks and Regards
  + Vidwath R Hebse
  + [vidwathrhebse@gmail.com](mailto:vidwathrhebse@gmail.com)
  + 9482204411