This project requires you to understand what mode of transport employees prefers to commute to their office. The attached data '[Cars.csv](https://olympus.greatlearning.in/courses/7362/files/737590/download?verifier=d51upLRepA9fEuQTEqkZhdlZKgs8Ce1eY3fwjfU2&wrap=1)' includes employee information about their mode of transport as well as their personal and professional details like age, salary, work exp. We need to predict whether or not an employee will use Car as a mode of transport. Also, which variables are a significant predictor behind this decision?

**Data Dictionary**

|  |  |
| --- | --- |
| Age | Age of the Employee in Years |
| Gender | Gender of the Employee |
| Engineer | For Engineer =1 , Non Engineer =0 |
| MBA | For MBA =1 , Non MBA =0 |
| Work Exp | Experience in years |
| Salary | Salary in Lakhs per Annum |
| Distance | Distance in Kms from Home to Office |
| license | If Employee has Driving Licence -1, If not, then 0 |
| Transport | Mode of Transport |

The following is expected out of the candidate in this assessment.

**EDA**

* Perform an EDA on the data
* Illustrate the insights based on EDA
* Check for Multicollinearity - Plot the graph based on Multicollinearity & treat it.

**Data Preparation** Prepare the data for analysis (SMOTE)

**Modeling**

* Create multiple models and explore how each model perform using appropriate model performance metrics
  + KNN
  + Naive Bayes (is it applicable here? comment and if it is not applicable, how can you build an NB model in this case?)
  + Logistic Regression
* Apply both bagging and boosting modeling procedures to create 2 models and compare its accuracy with the best model of the above step.

**Actionable Insights & Recommendations**

* Summarize your findings from the exercise in a concise yet actionable note