Project 5-Cars Case Study

This project requires you to understand what mode of transport employees prefers to commute to their office. The dataset "Cars-dataset" 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.

Following is expected out of the candidate in this assessment.

EDA (15 Marks)

- Perform an EDA on the data - (7 marks)
- Illustrate the insights based on EDA (5 marks)
- What is the most challenging aspect of this problem? What method will you use to deal with this? Comment (3 marks)

Data Preparation (10 marks)

- Prepare the data for analysis

Modeling (30 Marks)

- Create multiple models and explore how each model perform using appropriate model performance metrics (15 marks)
    - 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. (15 marks)

Actionable Insights & Recommendations (5 Marks)

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

Please note the following:

1. There are two parts to the submission:
    1. The output/report in any file format - the key part of the output is the set of observations and insights from the exploration and analysis
    2. Commented R code in .R or .Rmd
2. Please don't share your R code and/or outputs only, we expect some verbiage/story too - a meaningful output that you can share in a business environment

Scoring guide (Rubric) - Project 5 Cars Rubric (1)

Criteria	Points
Perform an EDA on the data	7
Illustrate the insights based on EDA	5

What is the most challenging aspect of this problem? What method will you use to deal with this? Comment	3
Prepare the data for analysis	10
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	15
Apply both bagging and boosting modeling procedures to create 2 models and compare its accuracy with the best model of the above step	15
Summarize your findings from the exercise in a concise yet actionable note	5