**Project Summary**

* **Objective:** Car\_Class prediction
* **Approach:** UtilizingLogisticRegression
* **Target variable:** Car\_Class
* **Programming Language:** Python
* **Deployed Python Libraries:** Pandas, SKlearn
* **TABLE OF CONTENTS**

1. Getting acquainted with the data
2. Exploration results after loading the data in Python
3. Data Cleaning and Null Value Handling
4. Preprocessing
5. Feature Selection
6. Prediction Model with SKLEARN (LogisticRegression)
   1. Model without hyperparameter tuning
   2. Model with hyperparameter tuning
7. Result summary
8. **GETTING ACQUAINTED WITH THE DATA**:

* Upon the initial review of the dataset, it appears to be quite clean with no missing values. However, it is advisable to confirm this using Python code before proceeding with model creation.

1. **EXPLORATION RESULTS AFTER LOADING THE DATA IN PYTHON**:
2. The dataset comprises a total of 719 observations, providing a foundational understanding of the volume of data available for analysis.
3. The dataset comprises a total of 20 variables, where "Class" serves as the target variable, and the remaining variables are considered dependent features.
4. **DATA CLEANING AND NULL VALUE HANDLING**

Upon loading the dataset into Python, it is evident that the data is well-maintained with the following characteristics:

i) No missing values are observed.

ii) Consistency is maintained in data types, as all features are represented in integer data type.

1. **PREPROCESSING:**

No preprocessing is required for the dataset due to the consistent and well-maintained nature of the data.

1. **FEATURE SELECTION:**

I have tried multiple feature selection techniques. Since using the best feature in the model does not yield any better accuracy, I have included all the feature in my model.

1. Univariate Feature Selection:

*ANOVA (Analysis of Variance)*: Measures the variance between groups and can be used for feature selection in multi-class classification.

*Mutual Information*: Measures the dependency between variables and can be used for both continuous and discrete data.

1. Recursive Feature Elimination (RFE):

Recursive feature elimination involves recursively removing the least important features until the desired number of features is reached

1. **Prediction Model with SKLEARN (LogisticRegression)**

* **Model\_1 - The dataset was employed without engaging in hyperparameter tuning**

In the initial phase, the dataset underwent scaling and was input into the LogisticRegression model with default hyperparameters, yielding an accuracy of 77.78%. The code was excluded as the repository already contains the implementation with the scaled dataset, optimized hyperparameters, and improved accuracy.

* **Model\_2 - The dataset was employed with the incorporation of hyperparameter tuning.**

I used RandomizedSearchCV to determine the optimized hyperparameters for the model, resulting in an accuracy of 79.12%. For submission, I am providing the code solely for this model. Codes for additional models are available upon request.

* **Here are the results from the other models I have experimented with**:

1. Decision Tree – 74.3%
2. RandomForest – 72.2%
3. XGboost – 75.69%

Should you require corresponding codes or more comprehensive information regarding any specific model, please feel free to request it.

1. **RESULT SUMMARY**

As none of the methods yielded superior accuracy compared to Logistic Regression, I have presented my project utilizing the Logistic Regression model.