### **Problem Statement**

A Chinese automobile company Geely Auto aspires to enter the Nigerian market by setting up its manufacturing unit and producing cars locally to compete with their Nigerian, US and European counterparts.

They have contacted you, a Data Scientist to understand the factors on which the pricing of cars depends. Specifically, they want to understand the factors affecting the pricing of cars in the Nigerian market, since those may be very different from the Chinese market.

The company wants to know:

* Which variables are significant in predicting the price of a car
* How well do those variables describe the price of a car

Based on various market surveys, you have gathered a large data set of different types of cars across the Nigerian market.

### **Business Goal (Objective)**

You are required to model the price of cars with the available independent variables. The management will use it to understand how exactly the prices vary with the independent variables. They can accordingly manipulate the design of the cars, the business strategy etc. to meet certain price levels. Further, the model will be a good way for management to understand the pricing dynamics of a new market.

**Data Description**

* **Train.csv** - To be used for training and tuning of models.
* **Test.csv** - To be used only for testing the performance of the final best model.
* **variable description -** csv file carrying the variables description

#### 

#### **Submission Guidelines**

1. The Submission should be made on your GitHub repo. there are two parts to the submission:

* A well-commented Jupyter notebook
* A presentation as you would present to the top management/business leaders [format - .pdf ]

1. Any found copied/ plagiarized by another person will not be graded and is an automatic disqualification
2. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation.
3. Submission will not be evaluated if
   1. it is submitted post-deadline, or,
   2. more than 2 files are submitted.

#### **Best Practices for Notebook**

* The final notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
* The notebook should be run from start to finish in a sequential manner before submission.
* It is important to remove all warnings and errors before submission.
* It is recommended that you read the problem statement and go through the criteria and description mentioned in the rubric before starting the project.

##### 

| **Criteria** | **Points** |
| --- | --- |
| **Exploratory Data Analysis and Insights** - Overview of the data - Univariate analysis  - Bivariate analysis  -Multivariate analysis |  |
| **Data pre-processing** - Prepare the data for analysis - Missing value Treatment - Ensure no data leakage |  |
| **Model building** - Build a Regression model |  |
| **Model Performances** - evaluate the model’s performance |  |
| **Business Insights & Conclusions** - Business insights and Conclusions. |  |
| **Notebook - Overall quality** - Structure and flow. - Well commented code. |  |
| **Presentation - Overall quality** - Structure and flow - Crispness - All key insights and recommendations covered |  |
|  |  |