**Cars' MPG Prediction Using Linear Regression**

Car is related to our daily life, and there could be some interesting mechanisms behind it. This dataset represents information about 393 cars in 1970-1982. The goal is to explore and analyze this dataset and create a predictive model to predict the cars’ mpg.

**Inspiration**

* How’s the cars’ mpg distributed?
* Which year has the highest number of car models?
* Which origin has the highest number of car models?
* Which car has the highest mpg?
* Which variables are highly correlated?
* Which type of cylinder is the most popular?

**Hackathon Challenge**

The challenge is around the concept of helping car engineers to predict the cars’ mpg. There are two main tasks in this challenge, listed below:

* **Task #1**: Exploration and visualization using pandas and seaborn packages on the dataset. Creating data visualization such as histogram, barplot, boxplot, scatter plot, pie plot, etc.
* **Task #2:** Building and evaluating a regression model to predict the cars’ mpg using scikit-learn package.

**The Data**

**Auto-MPG data-set**

The data-set can be accessed separately from the UCI Machine Learning Repository along with many other interesting data-sets.

The data concerns city-cycle fuel consumption in miles per gallon, to be predicted in terms of 8 attributes. There is a total of 393 numbers of instances.

**Attribute Information:**

1. mpg
2. cylinders
3. displacement
4. horsepower
5. weight
6. acceleration
7. model year
8. origin
9. car name

Check <https://archive.ics.uci.edu/ml/datasets/auto+mpg> for more information.

**Final scoring**

The final grade of each team will be calculated based on the following:

|  |  |
| --- | --- |
| **Task** | **Weight** |
| Data Visualization | 25% |
| Model Building | 25% |
| Powerpoint Slide | 25% |
| Presentation | 25% |