**1. Introduction and Problem statement**

Assignment was to build a Multi linear regression model from dataset of Toyota corolla sales.

* Building model for predicting Price of a startups.
* Dataset include large number of features that can influence the price.

**2. Solution architecture and Methodology**

* Dataset were cleaned by checking null and Nan values in dataset. Duplicates were dropped.
* Features were divided into discrete and continuous types of features.
* Continuous features were explored using histogram and box-plot. Box-plot was used to identify outliers in dataset.
* Discrete features were visualised using bar charts.
* Correlation factors were checked to gauge the relation of all the features with profit.
* Only continuous features were used for modelling since discreate features showed very small correction factor with price.
* Also, different plots were used to explore various features of dataset.
* A Multi linear regression model of price with continuous features was build and validated using Sklean python module.

**3. Results and Analysis**

r2: 0.8755

MAE: 999.4099

MSE: 1933065.5847

RMSE: 1390.3473

**4. Challenges Faced**

It was slightly easier to perform since I already had idea of the process. 50 startups task helped me get broad process of building models. Difficult part about this assignment was to understanding what features are important and how to justify those.

**5. Complexity**

* It was slightly complex assignment.
* Getting good r2 score took some time.

**6. Time invested**

On and off it took **3 hours**