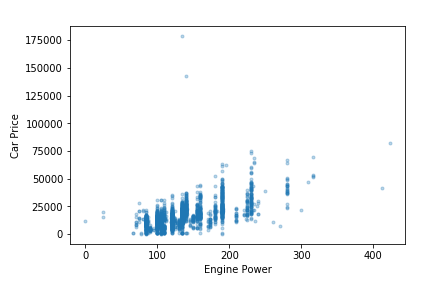
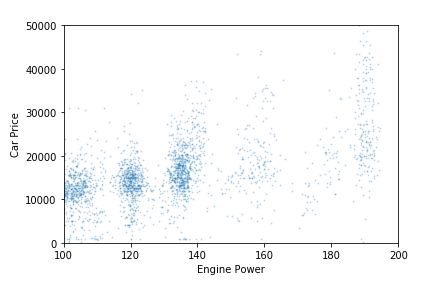
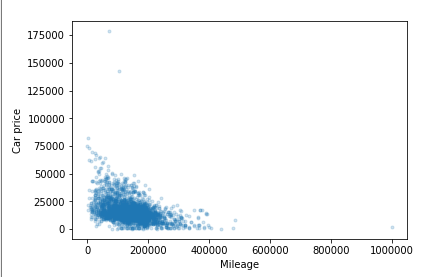
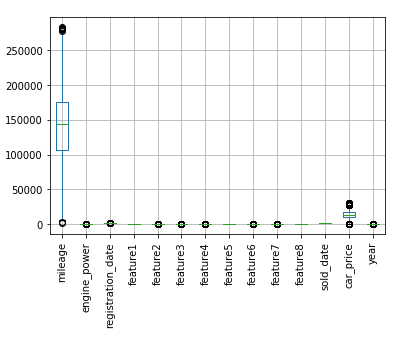
**WORKFLOW**

The solution was built in the following steps: -

* **Importing essential packages:** This step includes importing the essential modules like Pandas, NumPy, Scikit-Learn and one custom method(date\_to\_year) etc.
* **Loading the dataset:** This step includes loading the dataset from csv file into a Pandas Data Frame.
* **Data Exploration, handling and Visualisation:** This step includes *checking null values, identifying the relationship between features and label*. To make it easier visualization was used.

The visualization graphs are as follows: -

First three visuals refer to the relationship between specific feature and target variable(label) and the 4th one is used ***to detect the outliers in features***.

* **Feature and label Extraction, Encoding and Vectorization:** This step includes feature and label extraction and *encoding of some of the categorical features*. Some categorical ***features were Vectorized using Count Vectorizer***.
* **Scaling of features:** Since mileage column has data in 4 to 7 digits and other column consists data within 1 to 3 digits, therefore Standard Scaling was done.
* **Splitting the dataset into Training and Testing set:** After, all the pre-processing data is now ready for fitting therefore *splitting the features into training and testing set*. *Testing set is used for evaluating the score of model.*
* **Fitting appropriate Model:** Different models were used like Linear Regressor, LSVR, KNN Regressor, Random Forest Regressor etc. Random Forest Regressor gives the ***highest score i.e. 0.8357*** on testing set. ***Main\_code.py has the actual training part***.

**Finally, all the necessary files including model were saved using pickle.**