**MACHINE LEARNING BASED VEHICLE PERFORMANCE ANALYZER**

**TEAM ID:PNT2022TMID22329**

BACHELOR OF ENGINEERING

COMPUTER SCIENCE AND ENGINEERING

VEL TECH MULTI TECH DR.RANGARAJAN DR.SAKUNTHALA ENGINEERING COLLEGE, AVADI

INDUSTRY MENTOR : NIDHI

FACULTY MENTOR : SATHISH KUMAR P

Team Members:

MOHANA KRISHNAN R - 113119UG03058

NIKHIL B - 113119UG03062

SUNIL AHAMED - 113119UG03105

KANAGASABAI R - 113119UG03045

**Introduction:**

**1.1 Project overview:**

The monitoring of car performance, especiallygas consumption, has so far been approached only very superficially. A typical fuel gauge, when closely monitored,shows an extremelynon-linear relationship betweenneedle movement and fuel consumption. Inaccuracies occur especially in the range of critical low fuel values of 5-10% or more. In the past, due to this limitation, some luxury cars had an audible and flashing light alarm functionto indicate a low fuel condition. These systems, which add to the existingfuel level, have no more accuracy than the fuel level monitor alone.

In recent years, with the availability of computer techniquesand reliable and less expensivecomputer equipment, a number of systems have been developedto provide somewhatmore accurate information about vehicle performance.

**1.2 PURPOSE :**

The solution mainly aim on to predict the miles per gallon value based on the given input values that effect the performance of  the vehicle

**2.LITERATURE SURVEY**

**2.1 Existing problems :**

predicting the performance level of cars is an important and interesting problem. The main goal of the current study is to predict the performance of the car to improve certain behavior of the vehicle. This can significantly help to improve the systems fuel consumption and increase the efficiency. The performance analysis of the car based on the engine type, no of engine cylinders, fuel type and horsepower etc.These are the factors on which the health of the car can be predicted. It is an on-going process of obtaining, researching, analyzing and recording the health based on the above three factors. The performance objectives like mileage, dependability, flexibility and cost can be grouped together to play a vital role in prediction engine and engine management system.

This approach is the very important step towards understanding the vehicles performance

**2.2 References :**

Singh D, Singh M., "Internet of Vehicles for Smart and Safe Driving", International Conference on Connected Vehicles and Expo (ICCVE), Shenzhen,19 -23 Oct.,2015. (This paper has discussedabout smart transportation services in cloud (Cloud-STS) for safety and convenience. STS provide driver centric board services in the cloud networks. STS composed of Vehicle to WiFi networks(V to Wi-Fi), Vehicle to Cloud Network(V to CN), Vehicle to Vehicle (V to V), and Cloud Network to service provider (CN to SP). The idea is to utilize the (Wi-Fi enabled) Smart Highways and 3D camera enabled dash board navigationdevice to enhanceaccident prevention / monitoring and control.)

Zhang, Y., Lin, W., and Chin, Y., "Data -Driven Driving Skill Characterization: AlgorithmComparison and DecisionFusion," SAE TechnicalPaper 2009 -01 -1286, 2009, https://doi.org/10.4271/2009 -01 -1286.Azevedo, C. L Cardoso. (By adaptingvehicle control systemsto the skill level of the driver, the overall vehicle active safety providedto the driver can be further enhanced for the existing active vehicle controls,such asABS, Traction Control,Vehicle Stability Enhancement Systems. As a follow-up to the feasibility study in, this paper providessome recent results on data-driven driving skill characterization. In particular, the paper presentsan enhancement of discriminant features,the comparison of three differentlearning algorithms for recognizer design, and the performance enhancement with decision fusion. The paper concludes with the discussions of the experimental results and some of the future work.

J. E. Meseguer, C. T. Calafate,J. C. Cano and P. Manzoni, "Driving Styles: A smartphone application to assess driver behaviour," 2013 IEEE Symposiumon Computers and Communications (ISCC),Split, 2013, pp.000535 -000540.doi:10.1109/ISCC.2013.6755001. (The Driving Styles architecture integrates both data mining techniquesand neural networks2 to generate a classification of driving styles by analyzingthe driver behavioralong each route. In particular, based on parameterssuch as speed, acceleration, and revolutions per minute of the engine (rpm), we have implemented a neural network-based algorithm that is able to characterize the type of road on which the vehicle is moving, as well as the degree of aggressiveness of each driver. The final goal is to assist drivers at correcting the bad habits in their driving behaviour, while offering helpful tips to improve fuel economy. In this work we take advantage of two key-points: the evolution of mobile terminalsand the availability of a standard interface to access car data).

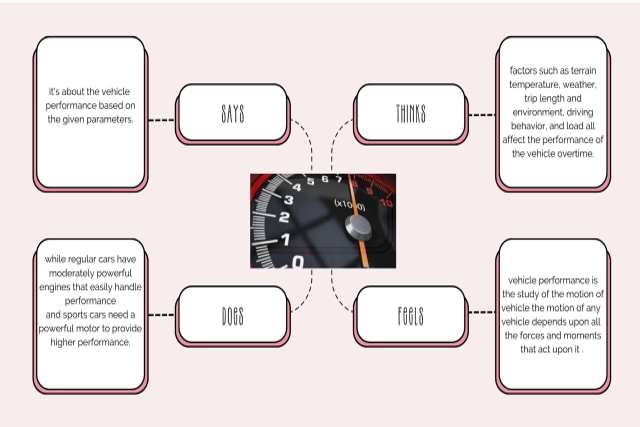
Kenneth L. Clarkson. 1985. Algorithms for Closest -Point Problems (Computational Geometry). Ph.D. Dissertation. Stanford University, Palo Alto, CA. UMI Order Number: AAT 8506171. (This dissertation reports a variety of new algorithms for solving closestpoint problems. The input to these algorithms is a set or sets of points in d- dimensional space, with an associated L (, p) metric. The problems consideredare: (1) The all nearest neighbours’ problem. For point set A, find the nearest neighbours in A of each point in A. (2) The nearest foreign neighbour problem. For point sets A and B, find the closest point in B to each point in A. (3) The geometricminimum spanning tree problem).

Goszczynska H., Kowalczyk L., Kuraszkiewicz B. (2014) Correlation Matrices as a ool to Analyze the Variability of EEG Maps. In: Piętka E., Kawa J., WieclawekW. (eds) Information Technologies in Biomedicine, Volume 4. Advancesin Intelligent Systems and Computing, vol 284. Springer.(The aim of this paper is to present the selected examplesof possible applications of image of correlation coefficients matrix of EEG map seriesin the analysis of variationof the topography of the isopotential areas in EEG maps, and thus in the assessment of stationarity, spatio-temporal variability and trends of changes of bioelectric activity of the brain. The image of correlation coefficients matrix shows similarityof all pairs of maps in a series. The choice of segmentation thresholdof characteristic areas in images of the correlation coefficients matrix of EEG map series corresponding to the sequencesimilarity relationships in a series of maps was based on the results of research conducted on test series).

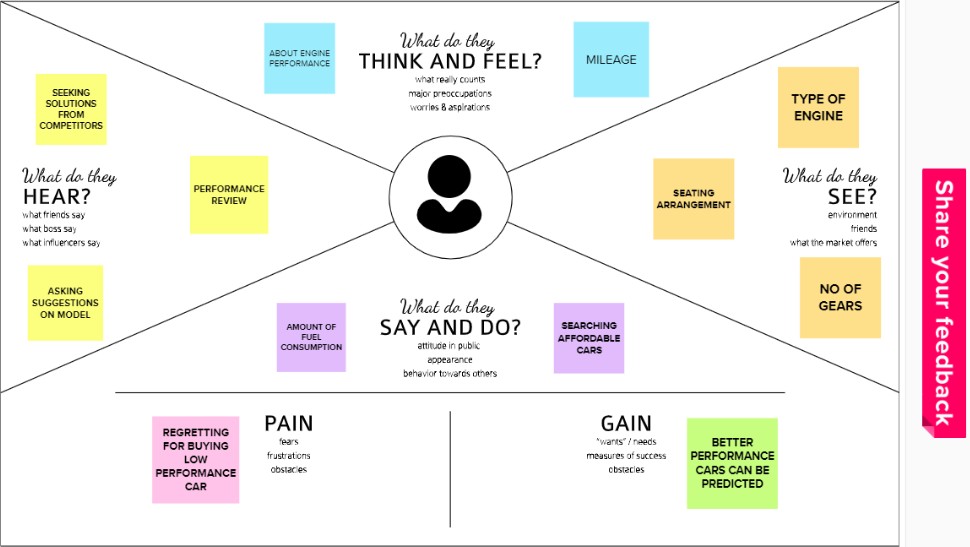
**2.3 Problem statement defination:**

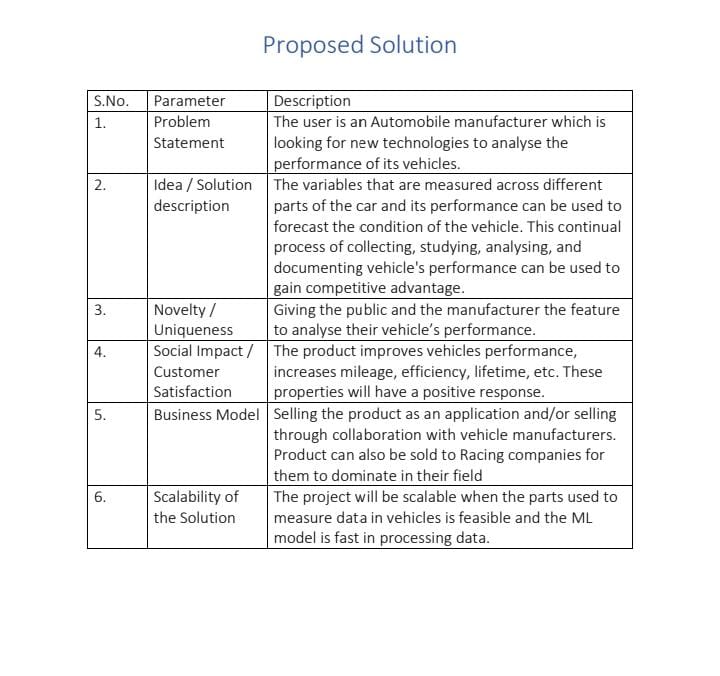
It is an important to analyse the factors using number of well-known approaches of machine learning algorithms like linear regression, decision tree and random forest to improve the vehicle performance efficiency.The range, durability and longevity of automotive traction batteries are ‘hot topics’ in automotive engineering. And here we consider a performance in mileage. To solve this problem, we will develop the models, using the different algorithms and neural networks. We will then see which algorithm predicts car performance(Mileage) with higher accuracy.

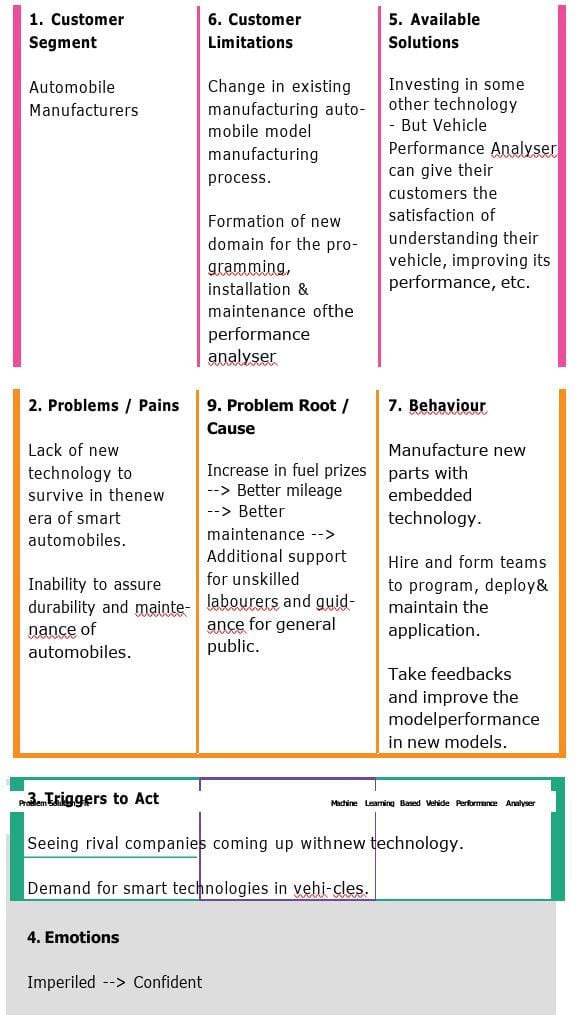
**3.Ideation and proposed solution:**

***3.1 Empathy map canvas*** 

**3.2 IDEATION ANDBRAIN STORMING :**



**3.3 PROPOSED SOLUTION**

**3.4 PROPOSED SOLUTION I**

**4. REQUIREMENT ANALYSIS :**

**Functional Requirements:**

                            Following are the functional requirements of the proposed solution.

|  |  |  |
| --- | --- | --- |
| **FR No.** | **Non-Functional Requirement** | **Description** |
| NFR-1 | Usability | • The system doesn't require any prior technical knowledge from the user, thus even a novice user can access it.  • The user interface would prioritize recognition over recall.  • User friendly  • Pay attention to internal sources of control  • It wouldn't take long for the content to load and show (30 seconds).  • The fields in the site would be selfexpla |
| NFR-2 | Security | • Only the authenticated user will be able to use the site's services. • The database should be backed up every hour. |

|  |  |  |
| --- | --- | --- |
| **FR No.** | **Functional Requirement (Epic)** | **Sub Requirement (Story / Sub-Task)** |
| FR-1 | User Registration | Registration through Form Registration through Gmail |
| FR-2 | User Confirmation | Confirmation via Email Confirmation via OTP |
| FR-3 | User Details | No of cylinders, Displacement, Horsepower,  Weight, Model year, and Origin |
| FR-4 | User Requirements | • Upload all essential details to the website's appropriate. • The system would extract all essential data based on the uploads. • Based on the information that was scraped, a list of every potential potential results will be delivered. |

**Non-functional Requirements:**

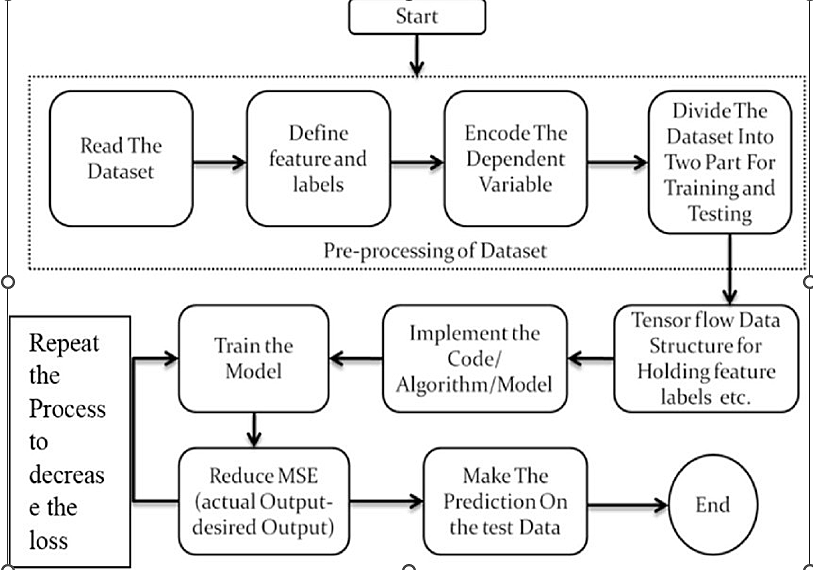
                     Following are the non-functional requirements of the proposed solution.

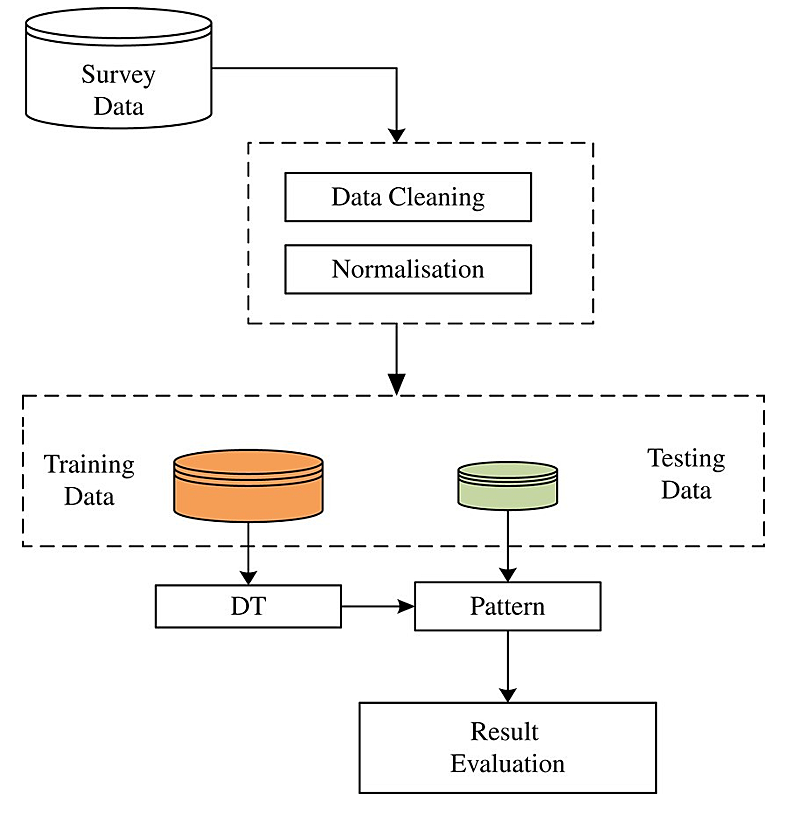
|  |  |  |
| --- | --- | --- |
| NFR-3 | Reliability | • Due to the value of data and the potential harm that inaccurate or incomplete data could do, the system will always strive for optimum reliability.  • The system will be operational every day of the week, 24 hours a day. |
| NFR-4 | Performance | • The website can efficiently handle traffic by responding to requests right away.  • A 64-kbps modem connection would take no longer than 30 seconds to see this webpage (quantitatively, the mean time) |
| NFR-5 | Availability | • Low data redundancy  • reduced error risk, quick and effective |
| NFR-6 | Scalability | • A significant number of users must be able to access the system simultaneously because an academic portal is essential to the courses that use it.  • The system will likely be most stressed during the admissions season.  • Therefore, it must be able to handle several users at once. |

1. **PROJECT DIAGRAM :**

**5.1 DATA FLOW DIAGRAM:**

A data flow diagram (DFD) illustrates how data is processed by a system in terms of inputs and outputs. As its name indicates its focus is on the flow of information, where data comes from, where it goes and howit gets stored





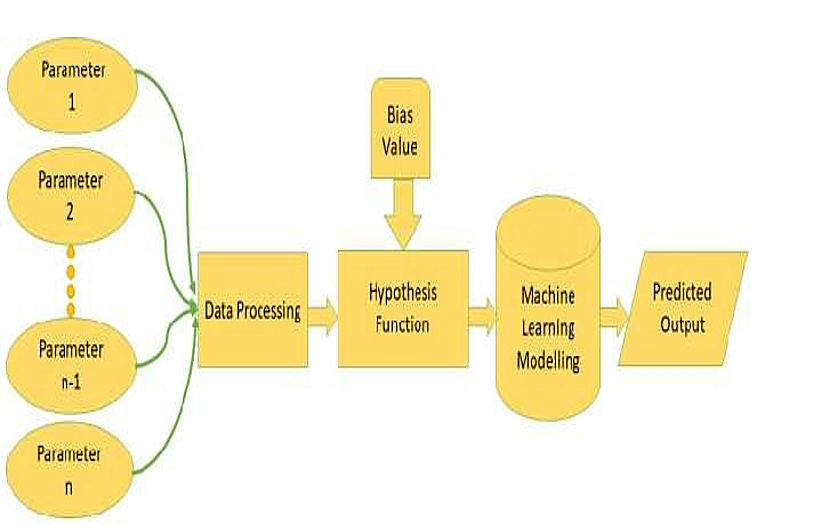
**5.2 SOLUTION AND TECHNICAL ARCHITECTURE :**

**Solution:**

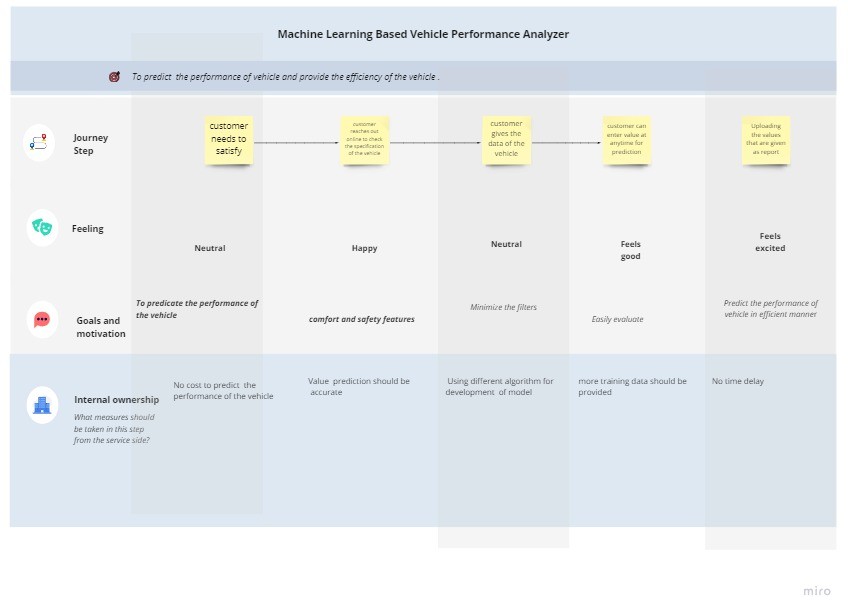
predicting the performance level of cars is an important and interesting problem. The main goal of the current study is to predict the performance of the car to improve certain behavior of the vehicle. This can significantly help to improve the systems fuel consumption and increase the efficiency. The performance analysis of the car based on the engine type, no of engine cylinders, fuel type and horsepower etc.These are the factors on which the health of the car can be predicted. It is an on-going process of obtaining, researching, analyzing and recording the health based on the above three factors. The performance objectives like mileage, dependability, flexibility and cost can be grouped together to play a vital role in prediction engine and engine management system.

This approach is the very important step towards understanding the vehicles performance

It is an important to analyse the factors using number of well-known approaches of machine learning algorithms like linear regression, decision tree and random forest to improve the vehicle performance efficiency.The range, durability and longevity of automotive traction batteries are ‘hot topics’ in automotive engineering. And here we consider a performance in mileage. To solve this problem, we will develop the models, using the different algorithms and neural networks. We will then see which algorithm predicts car performance(Mileage) with higher accuracy.



**5.3 USER STORIES :**



**6. PROJECT PLANNING AND SCHEDULING :**

**6.1 SPRINT PLANNING**

**StepsTo Perform Predictive Analysis:**

Some basicsteps should be performed in order to perform predictive analysis.

**Define Problem Statement:**

Define the project outcomes, the scope of the effort, objectives, identifythe data sets that are going to be used.

**Data Collection:**

Data collection involves gathering the necessary details required for the analysis. It involves the historical or past data from an authorized source over which predictive analysisis to be performed.

**Data Cleaning:**

Data Cleaning is the process in which we refine our data sets. In the process of data cleaning, we remove un-necessary and erroneous data. It involvesremoving the redundantdata and duplicatedata from our data sets.

**Data Analysis:**

It involves the exploration of data. We explore the data and analyze it thoroughly in order to identify some patterns or new outcomesfrom the data set. In this stage, we discoveruseful information and conclude by identifying some patterns or trends.

**Build Predictive Model:**

In this stage of predictiveanalysis, we use various algorithmsto build predictivemodels based on the patternsobserved. It requiresknowledge of python, R, Statistics and MATLAB and so on. We also test our hypothesis using standard statisticmodels.

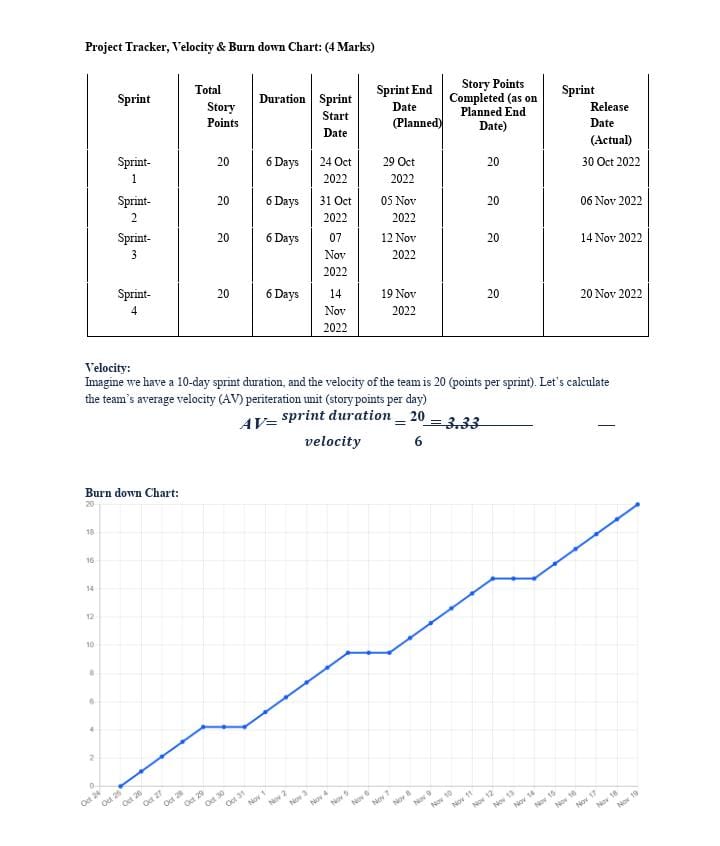
**Validation:**It is a very importantstep in predictive analysis. In this step, we check the efficiencyof our model by performingvarious tests. Here we providesample input sets to check the validityof our model. The model needs to be evaluatedfor its accuracy in this stage.

**Deployment:**

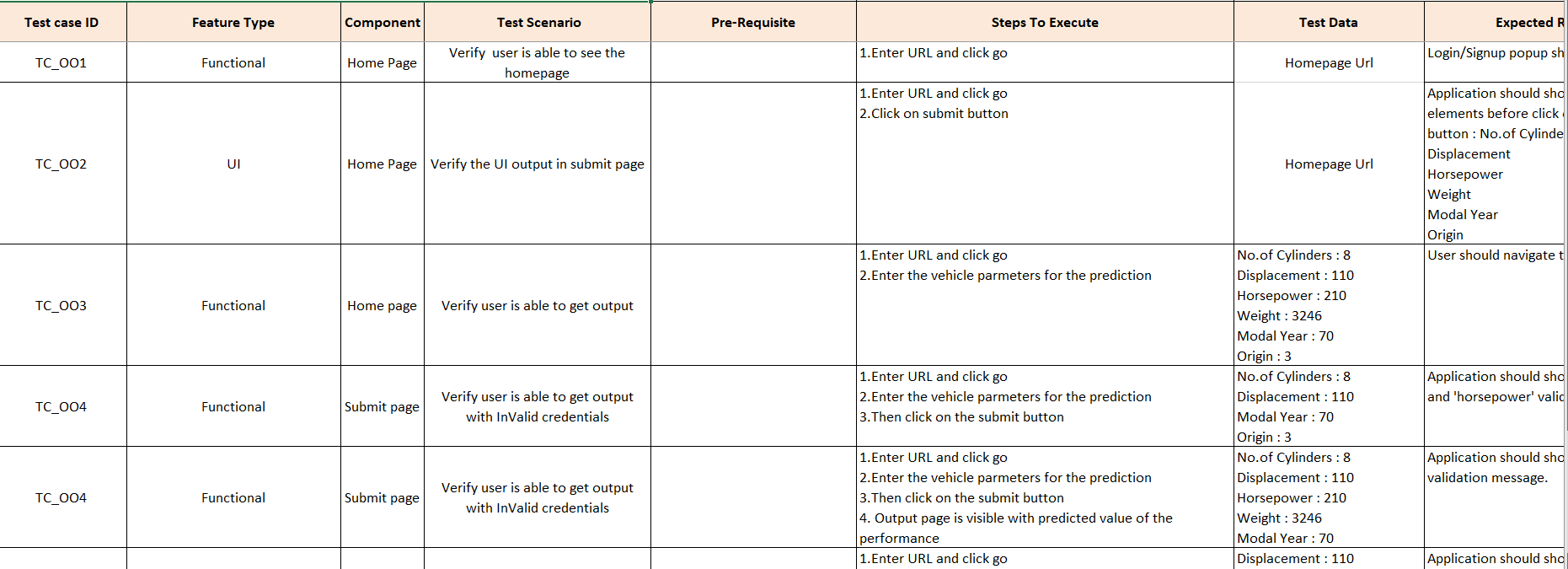
In deployment we make our model work in a real environment and it helps in everyday discussion making and make it available to use.

**Model Monitoring:**

Regularly monitor your models to check performance and ensure that we have proper results.It is seeing how model predictions are

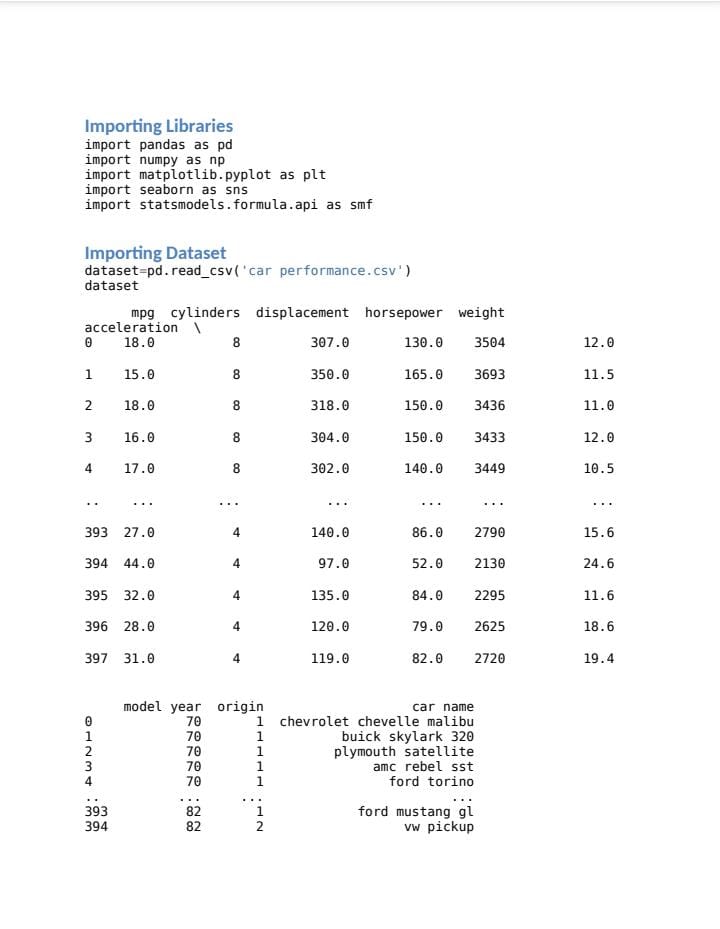
**6.2 SPRINT DELIVERY SCHEDULE :** 

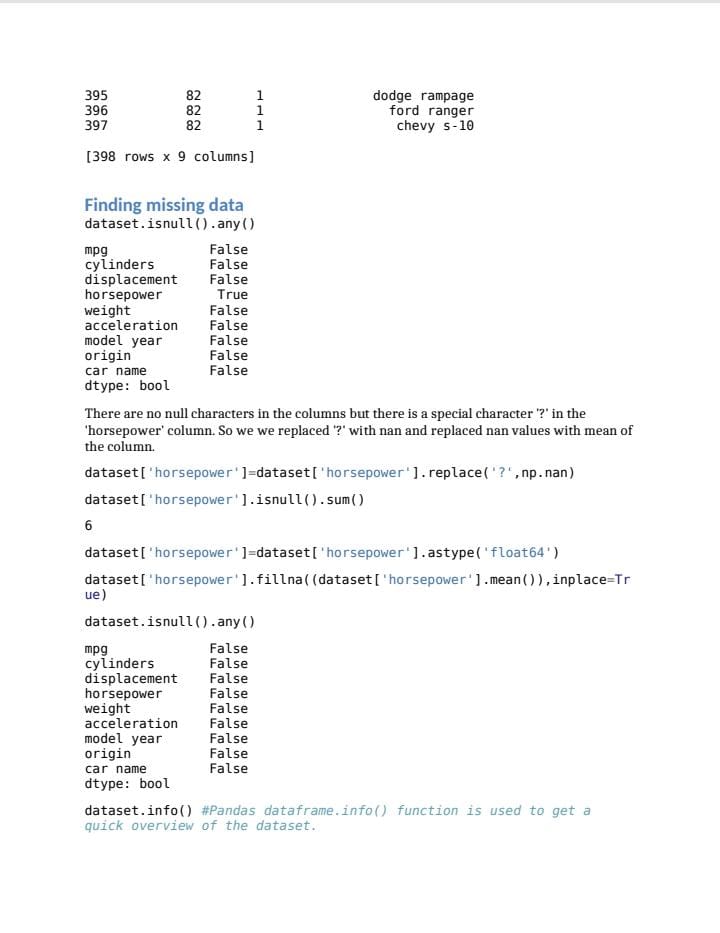
**6.3 Reports from JIRA**

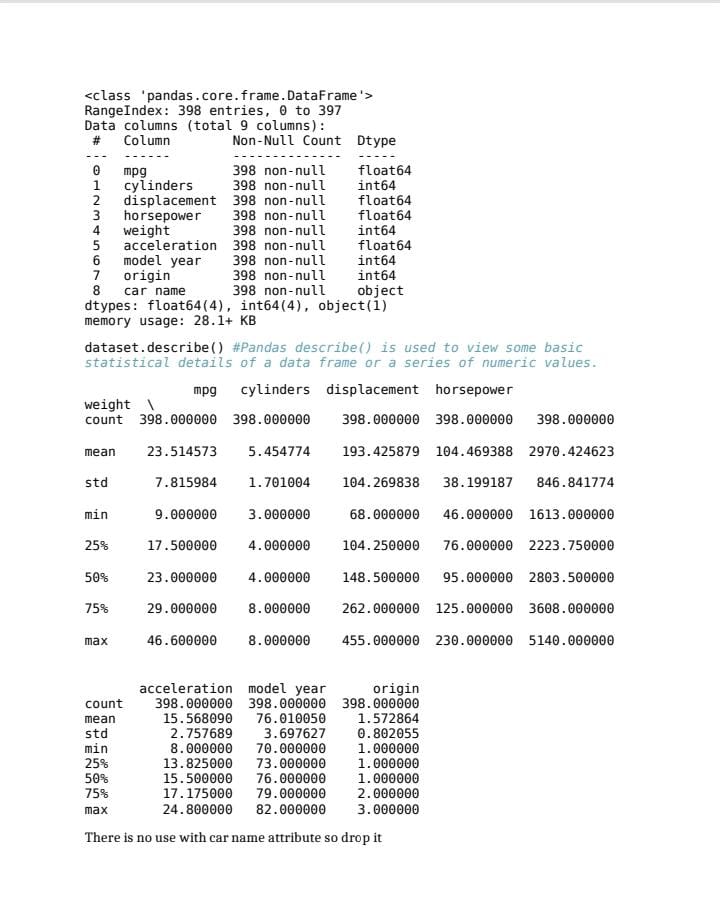


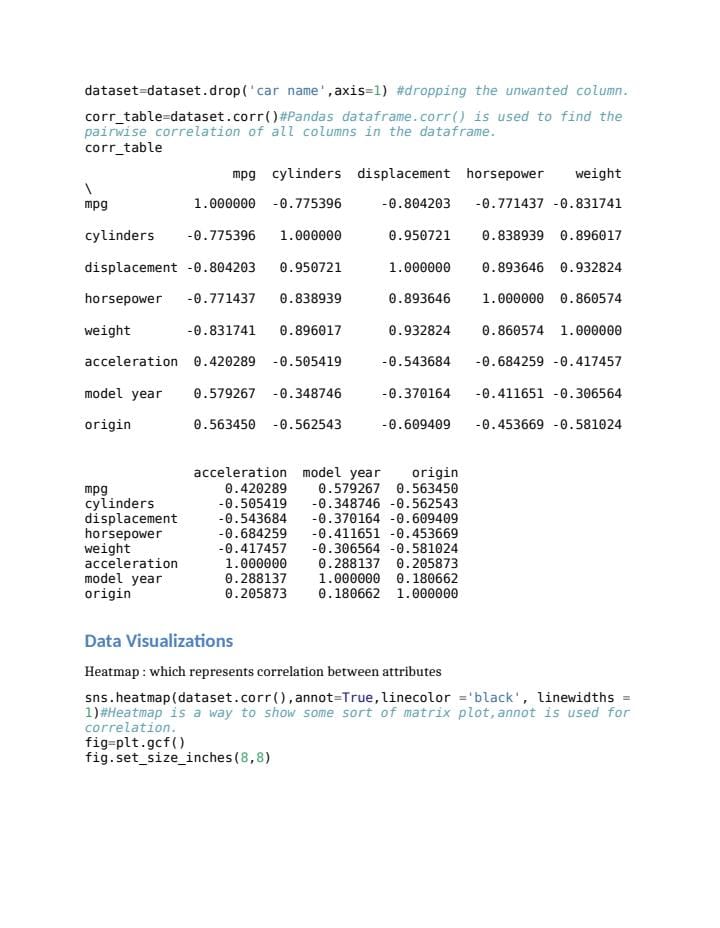
**7.CODING & SOLUTIONING  :**

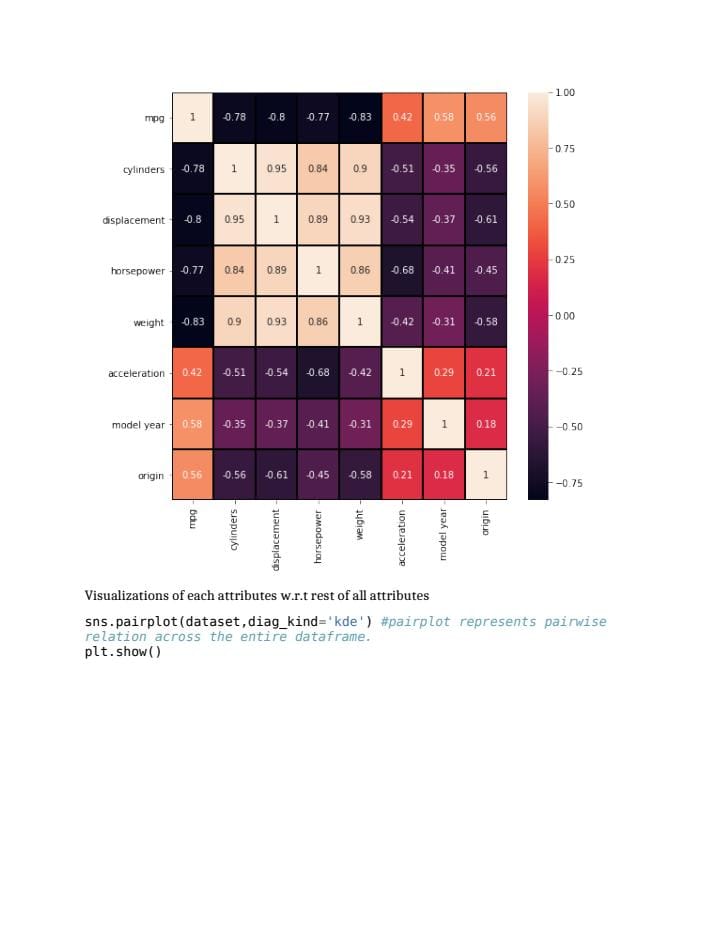
**Feature 1 :**

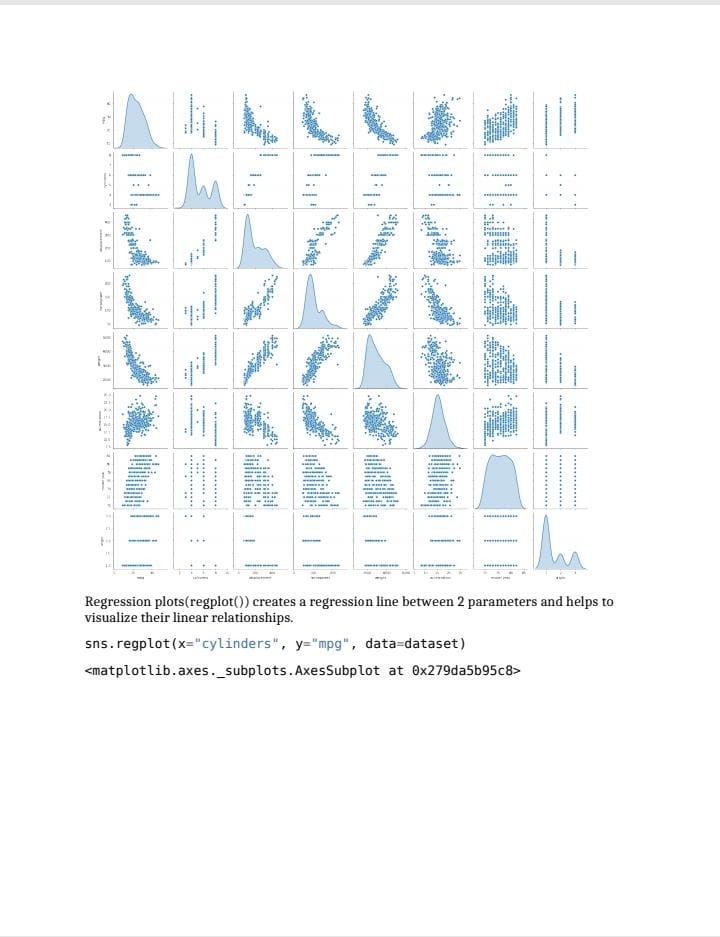


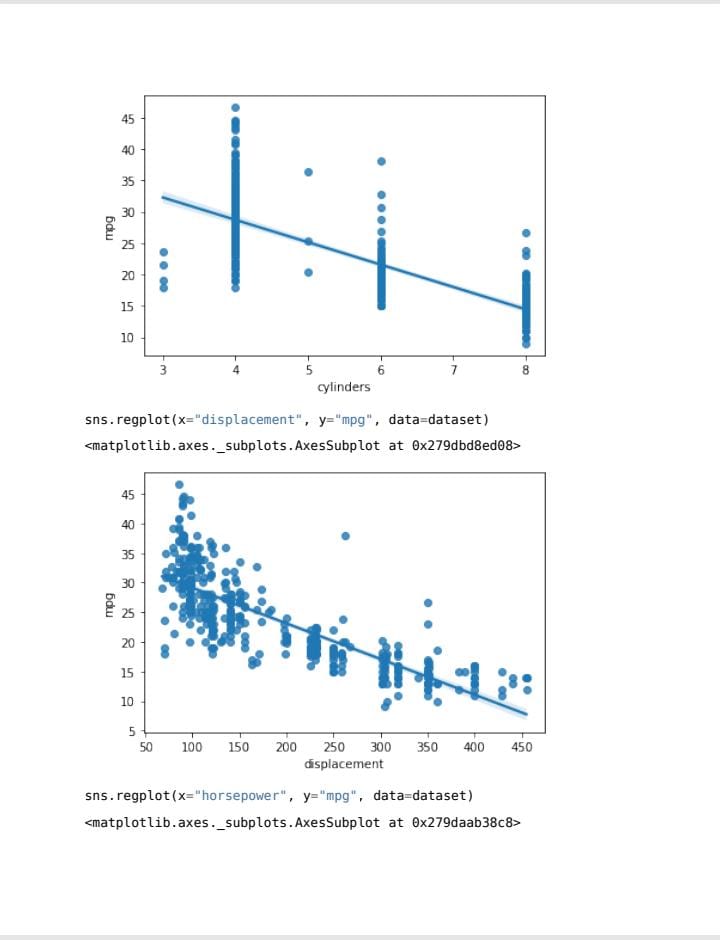


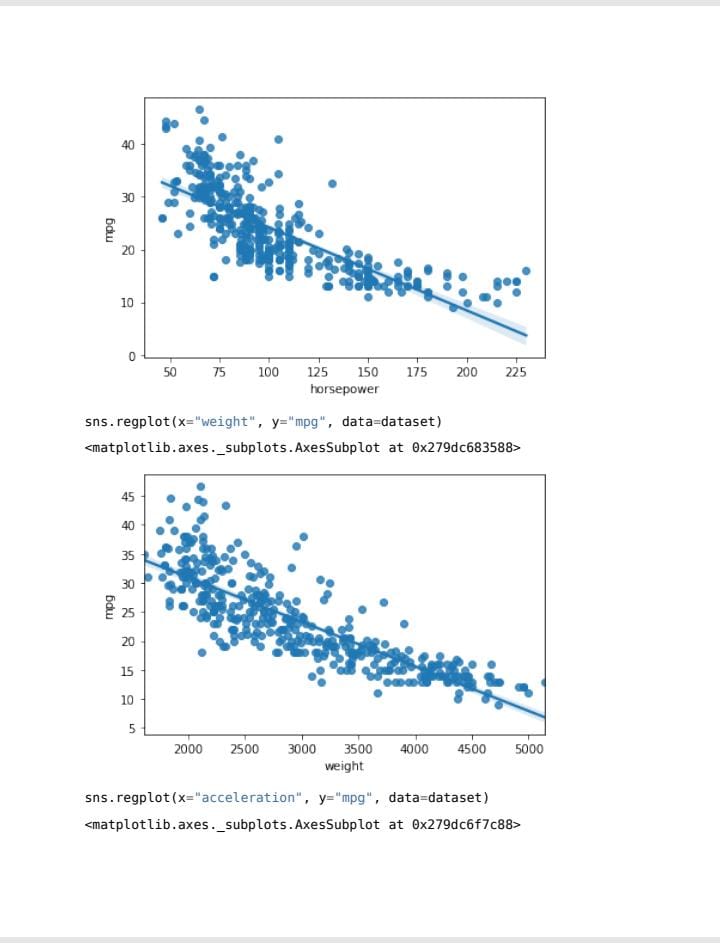


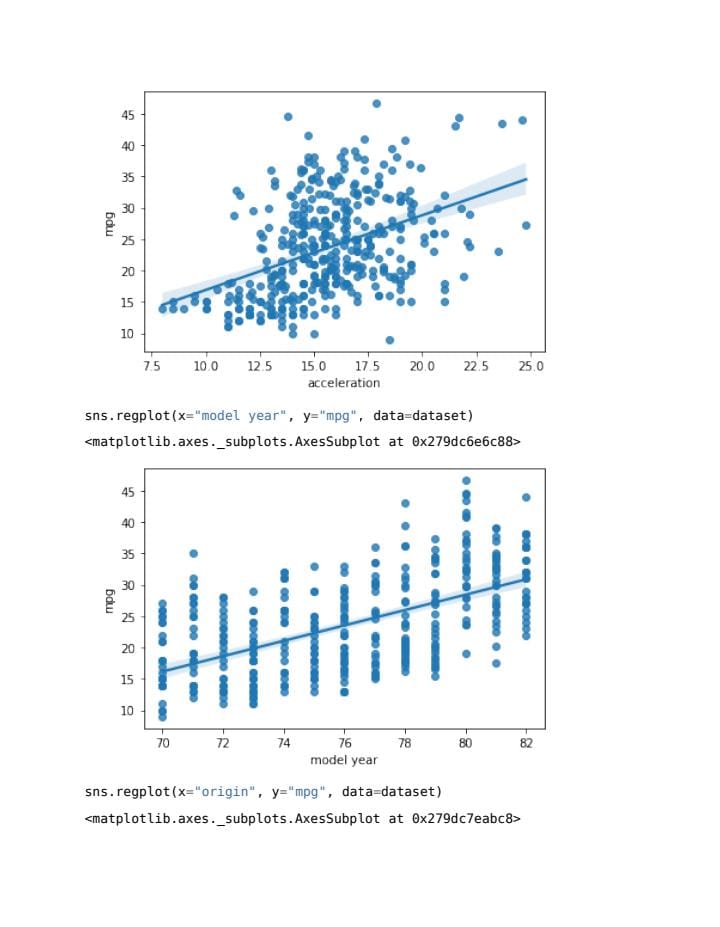


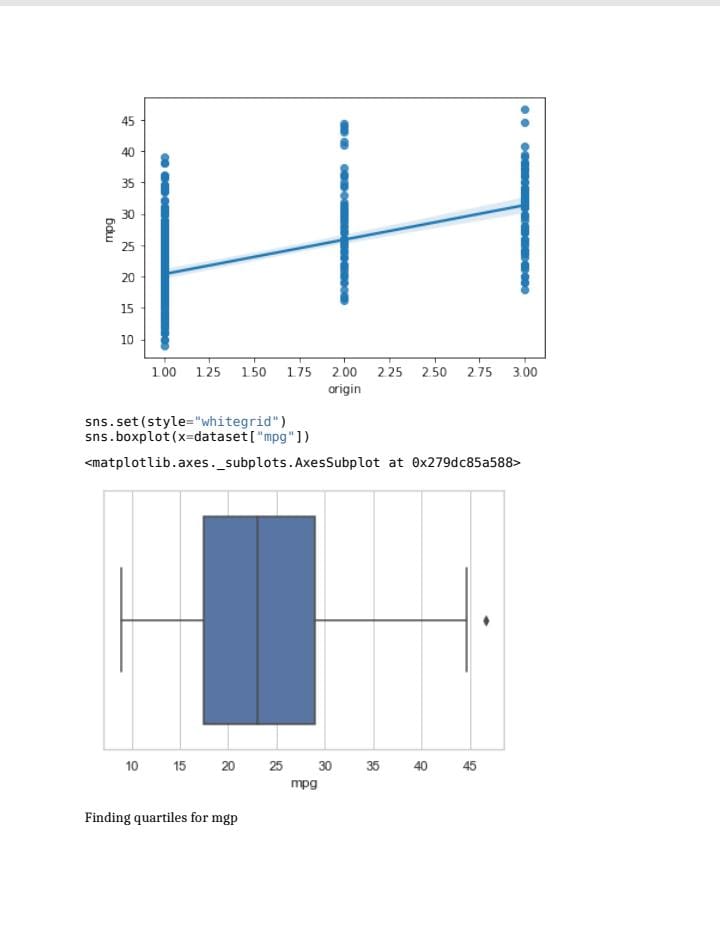


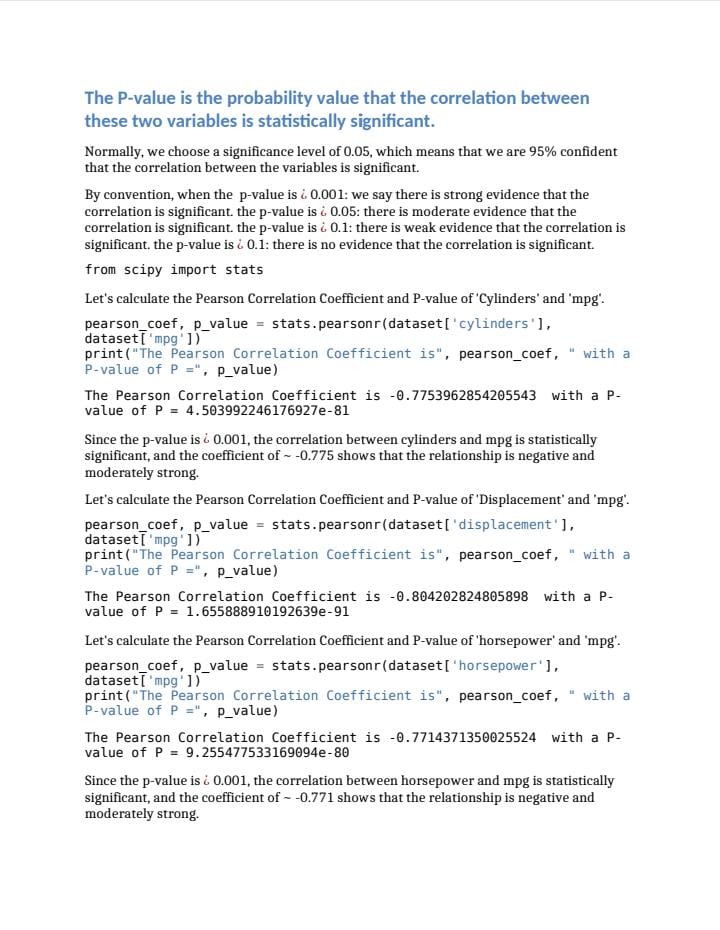




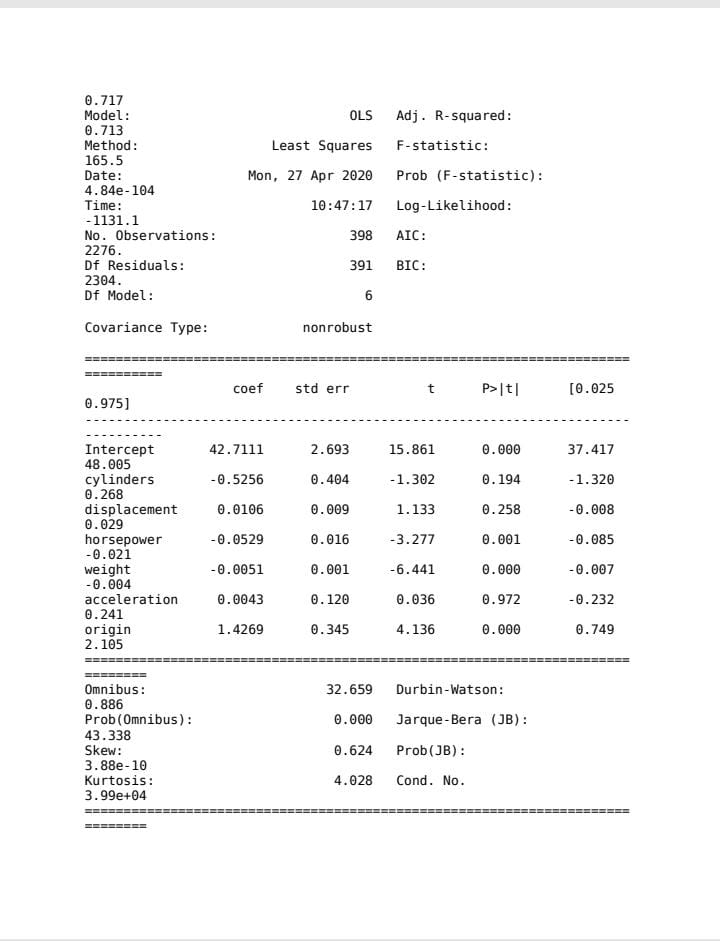


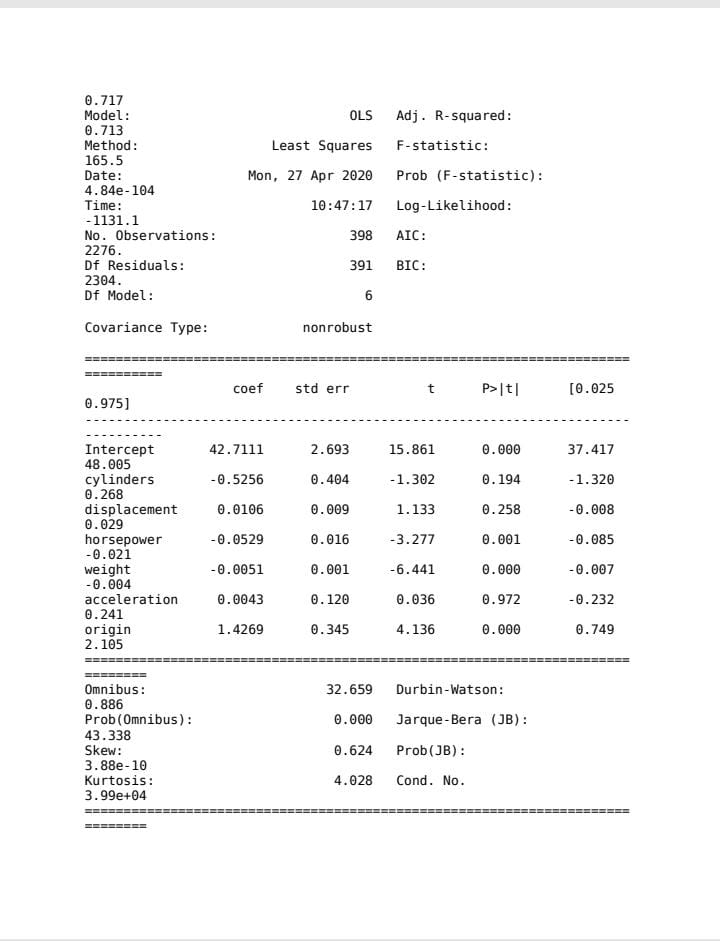


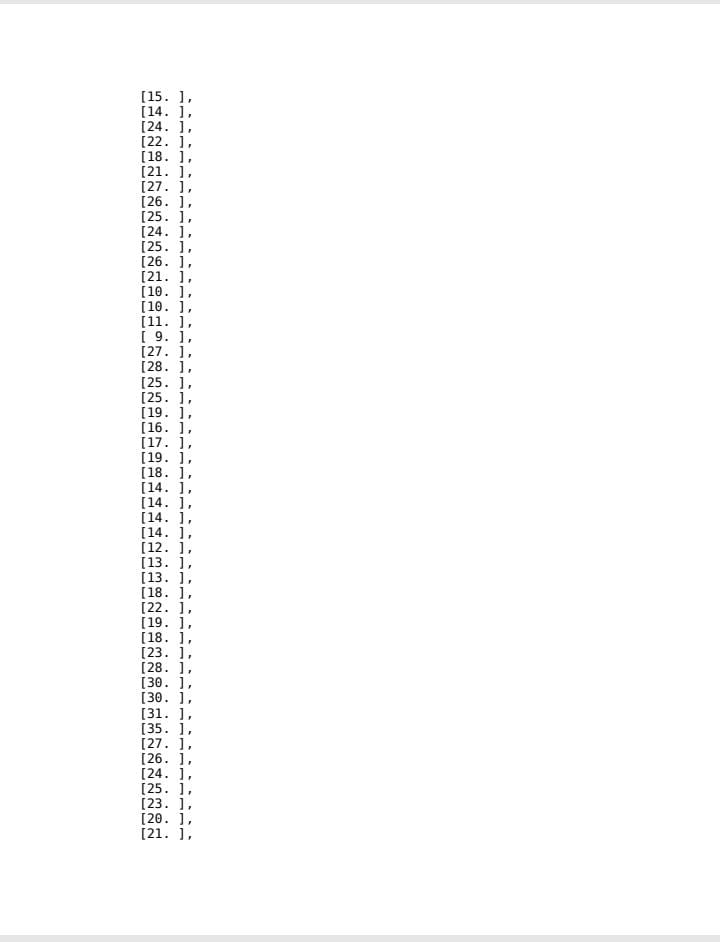


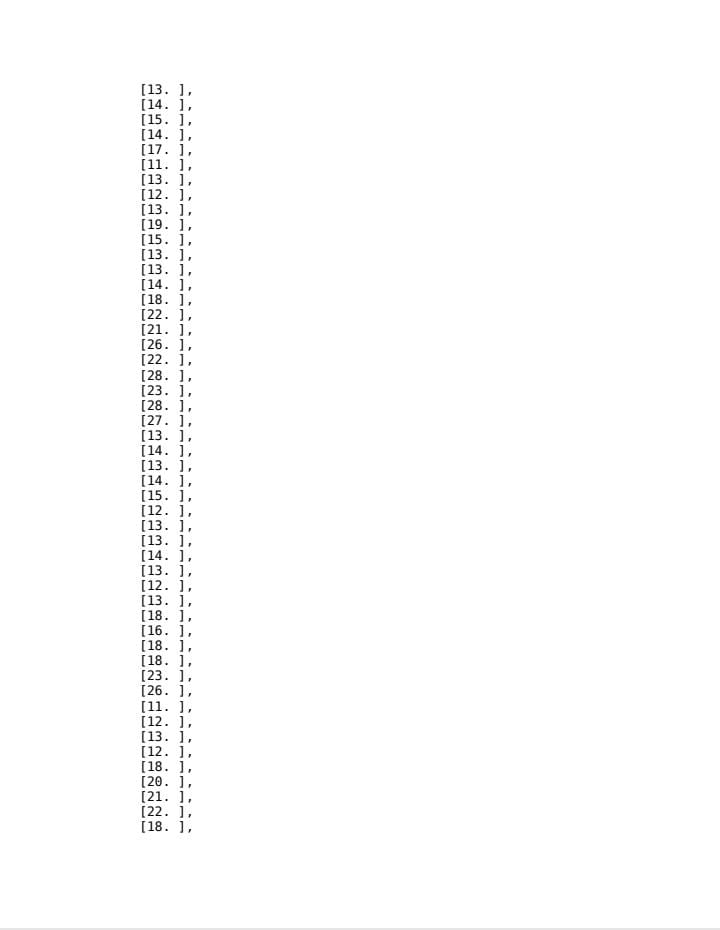


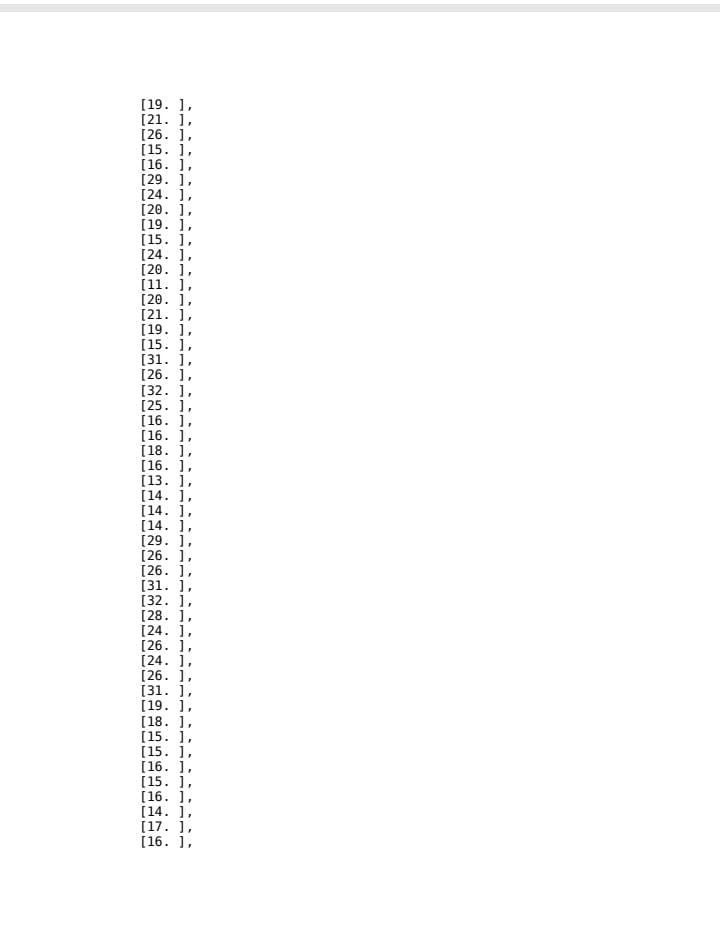


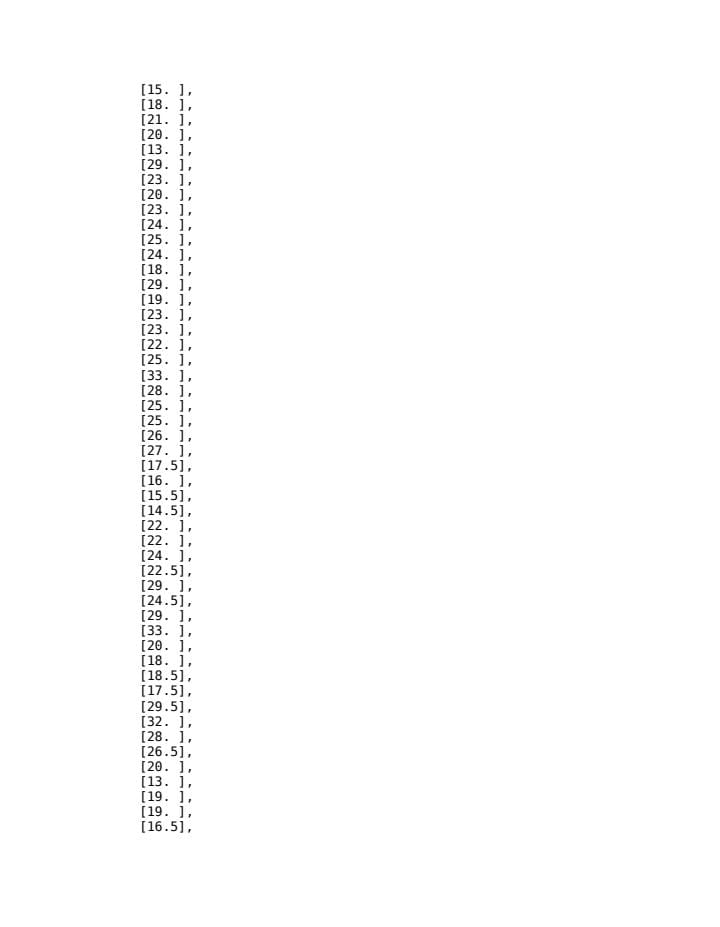


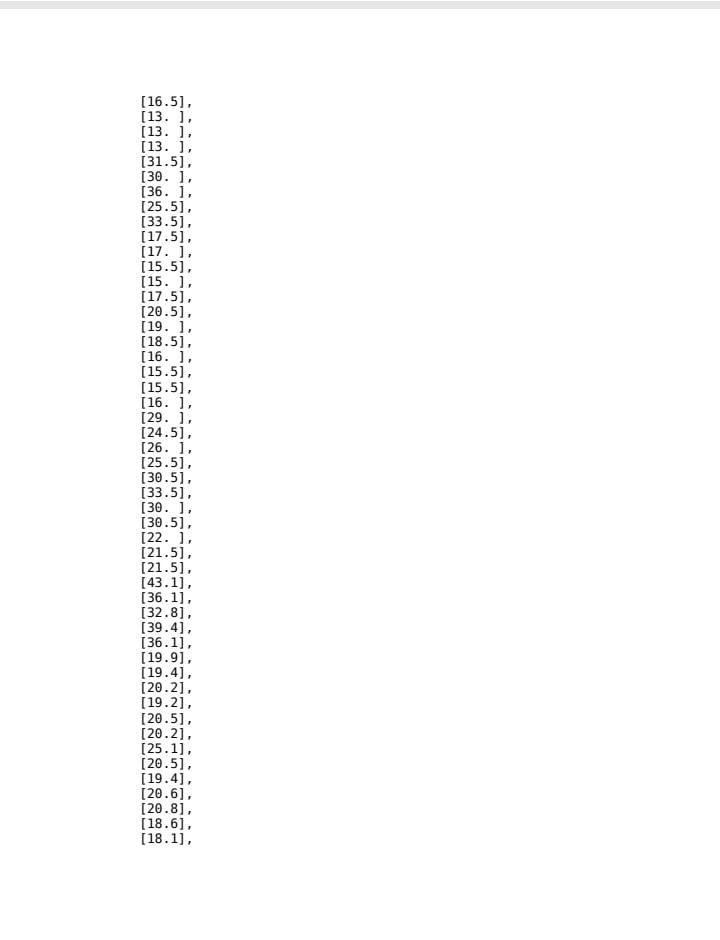


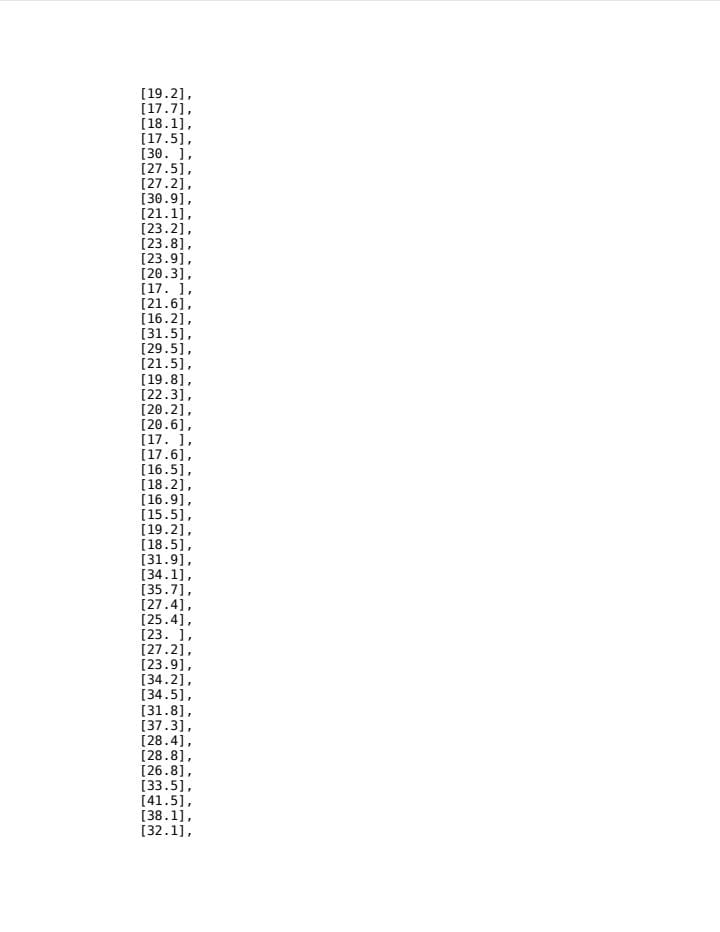


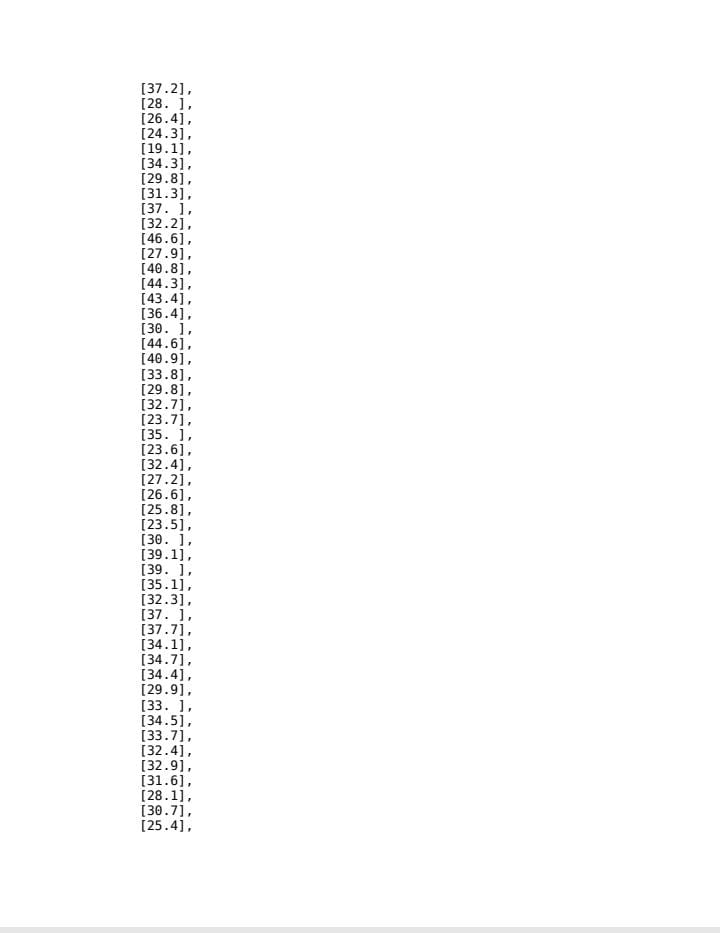


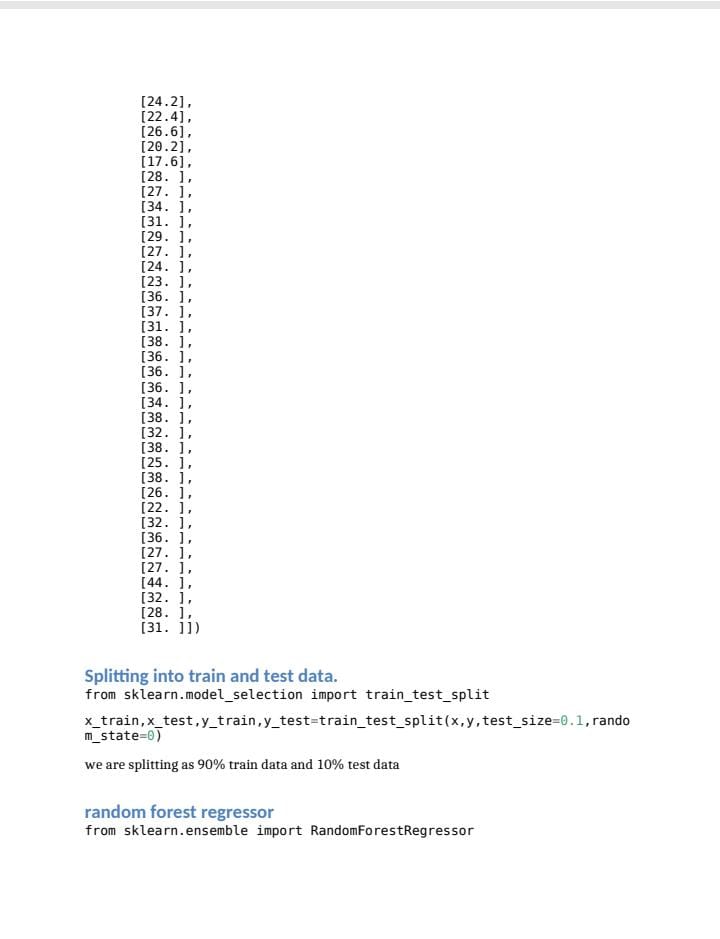


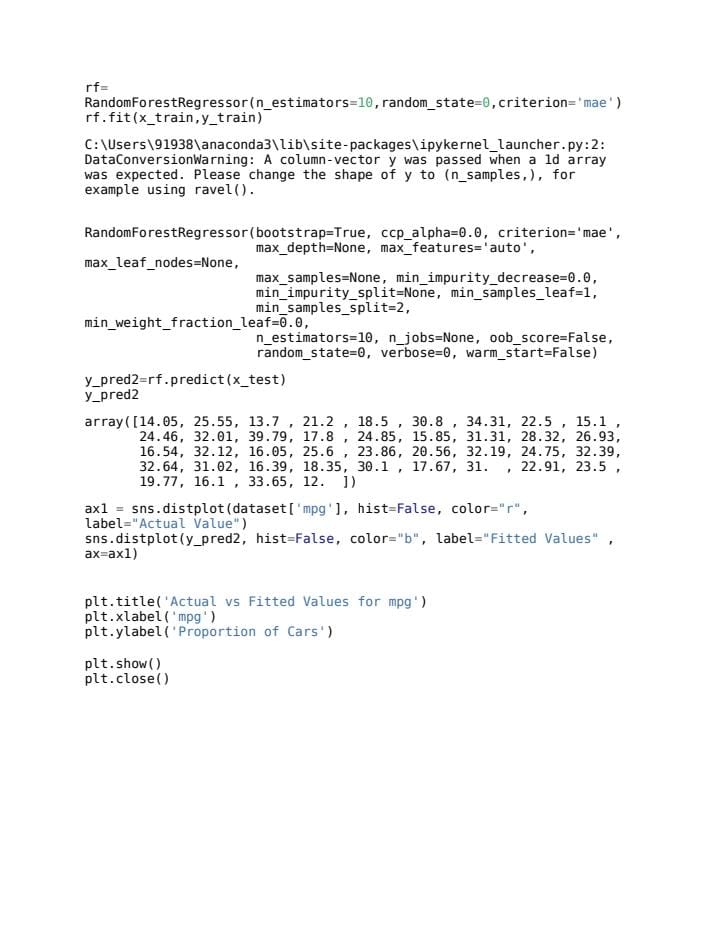


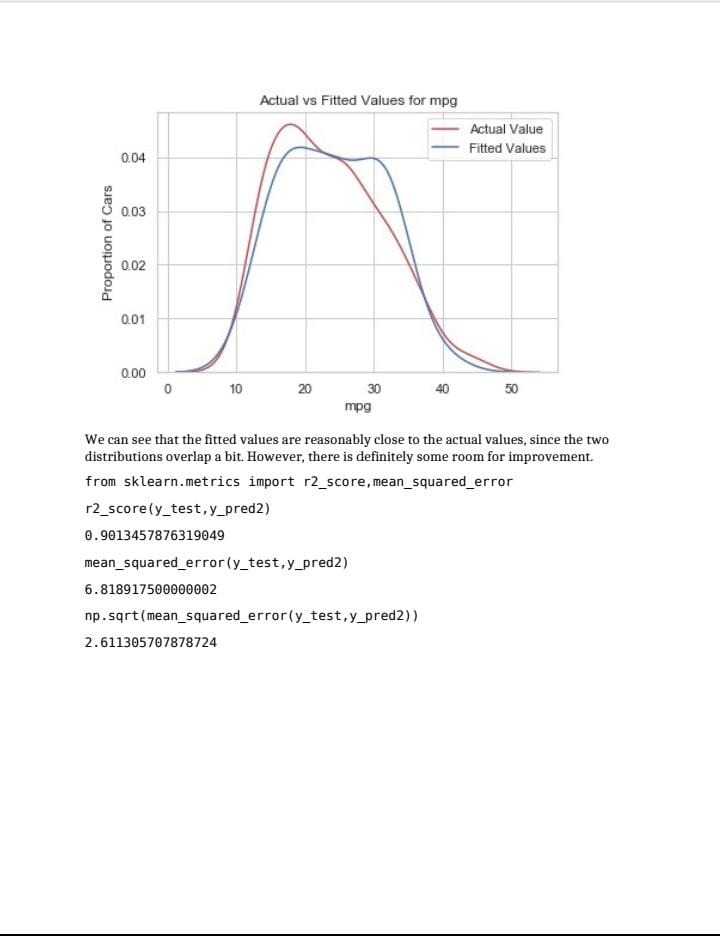






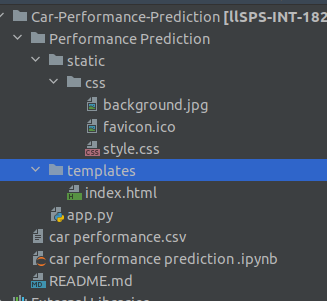




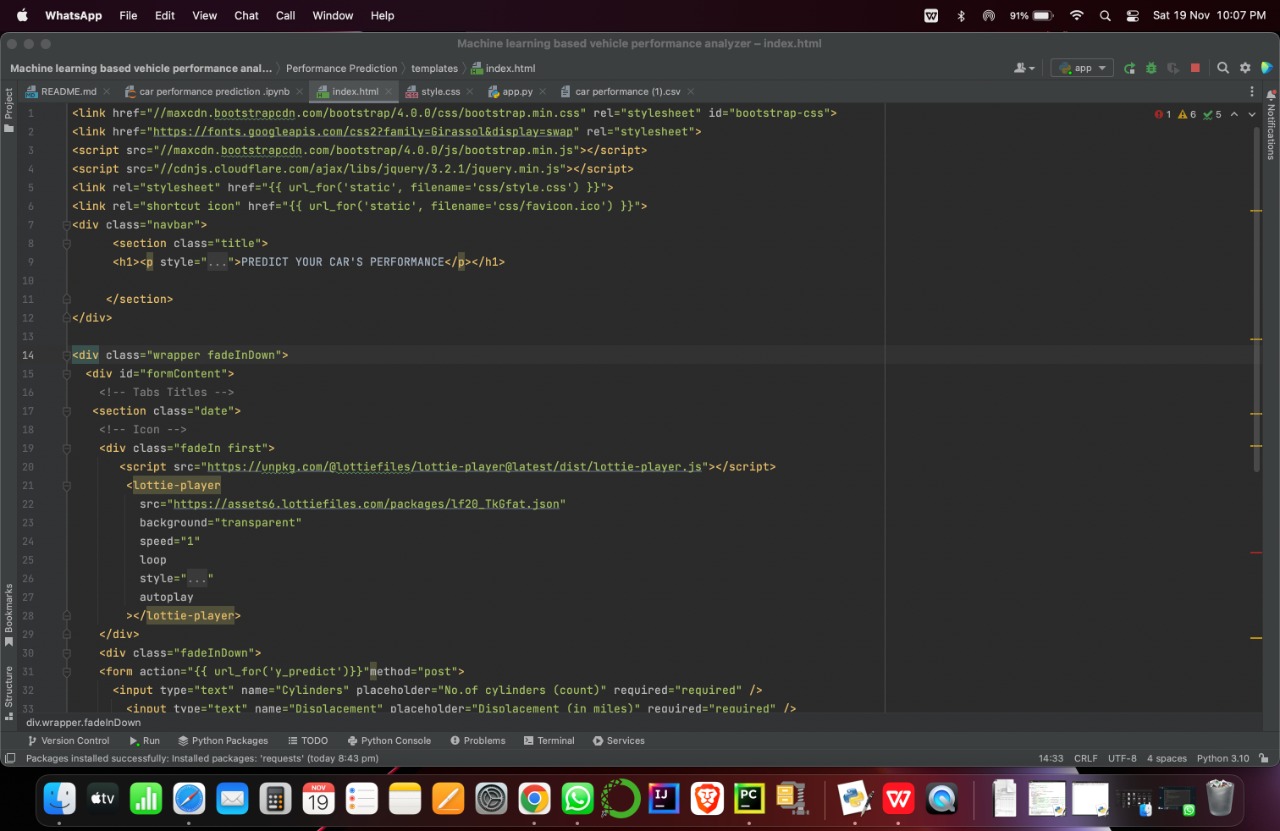


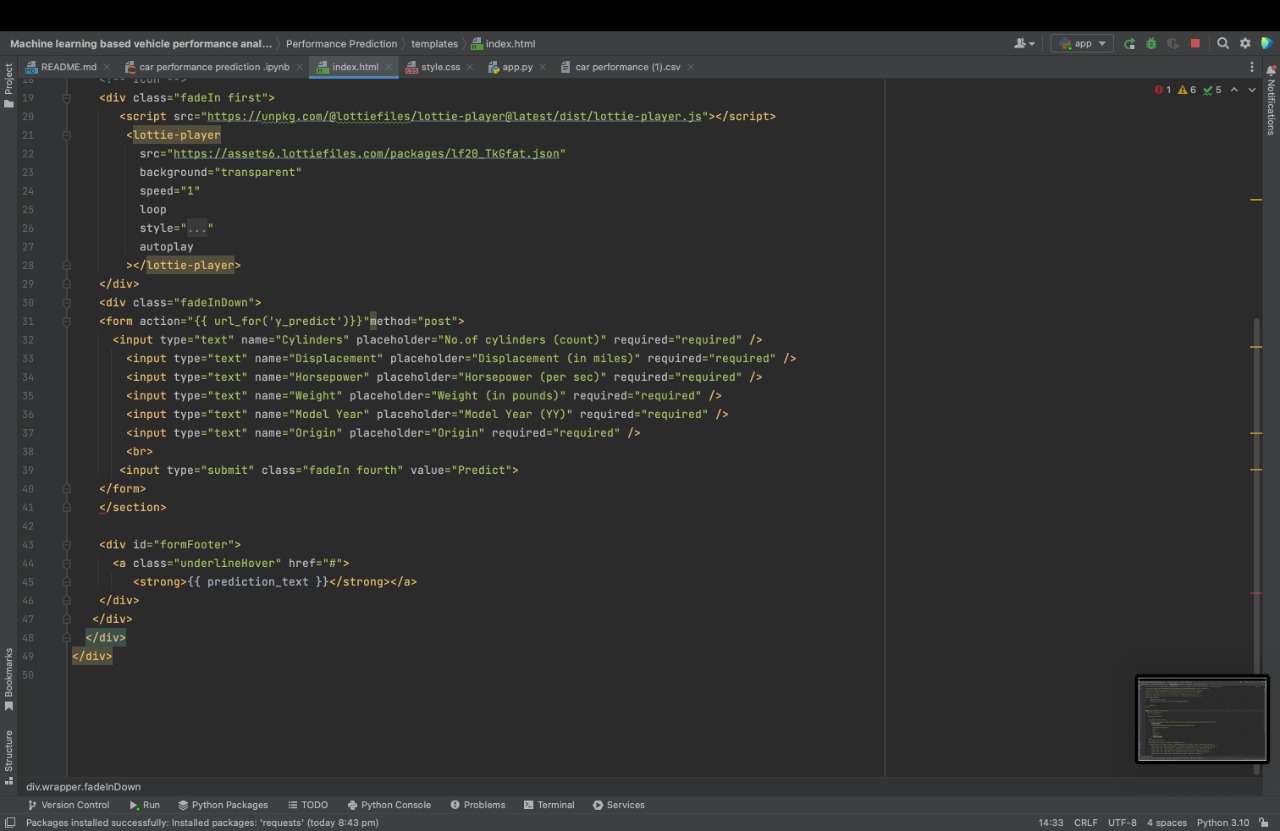
**Feature 2 :**

**Application Building :**

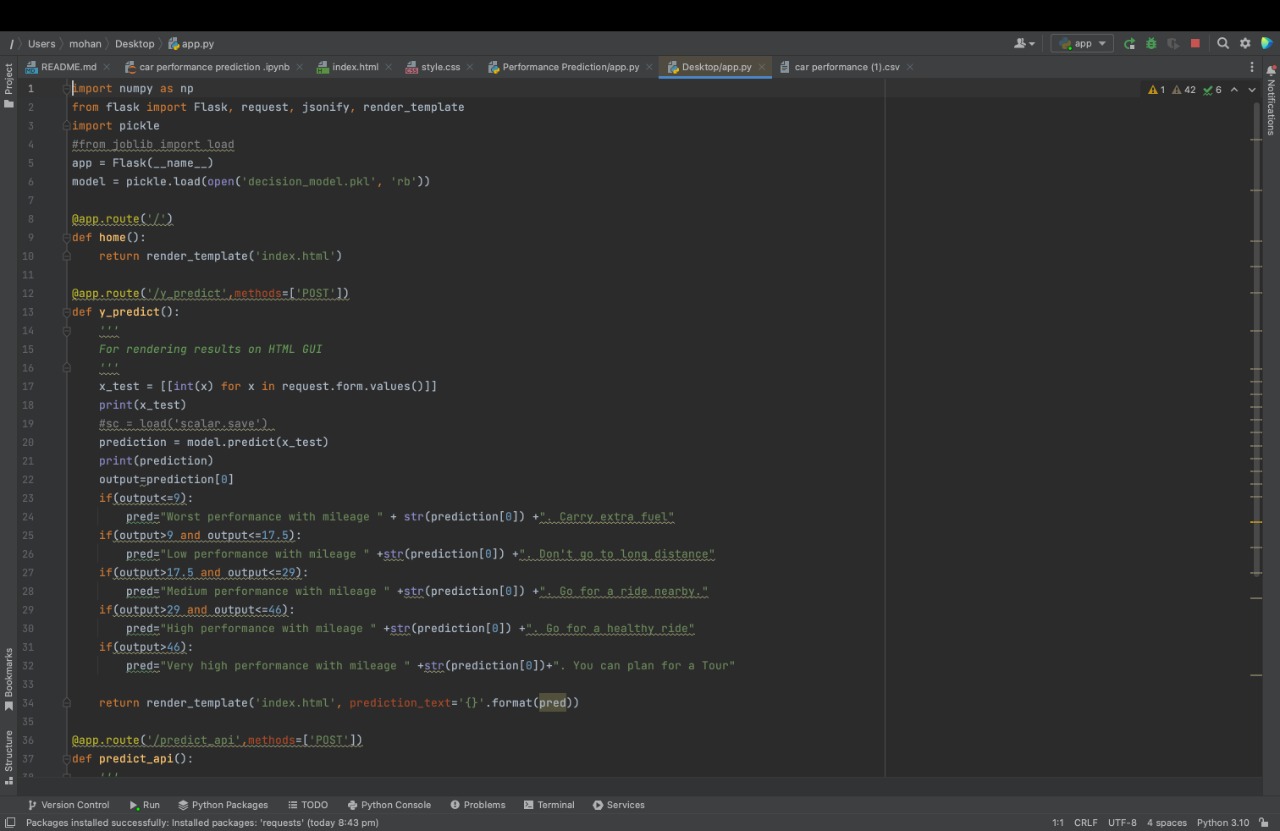


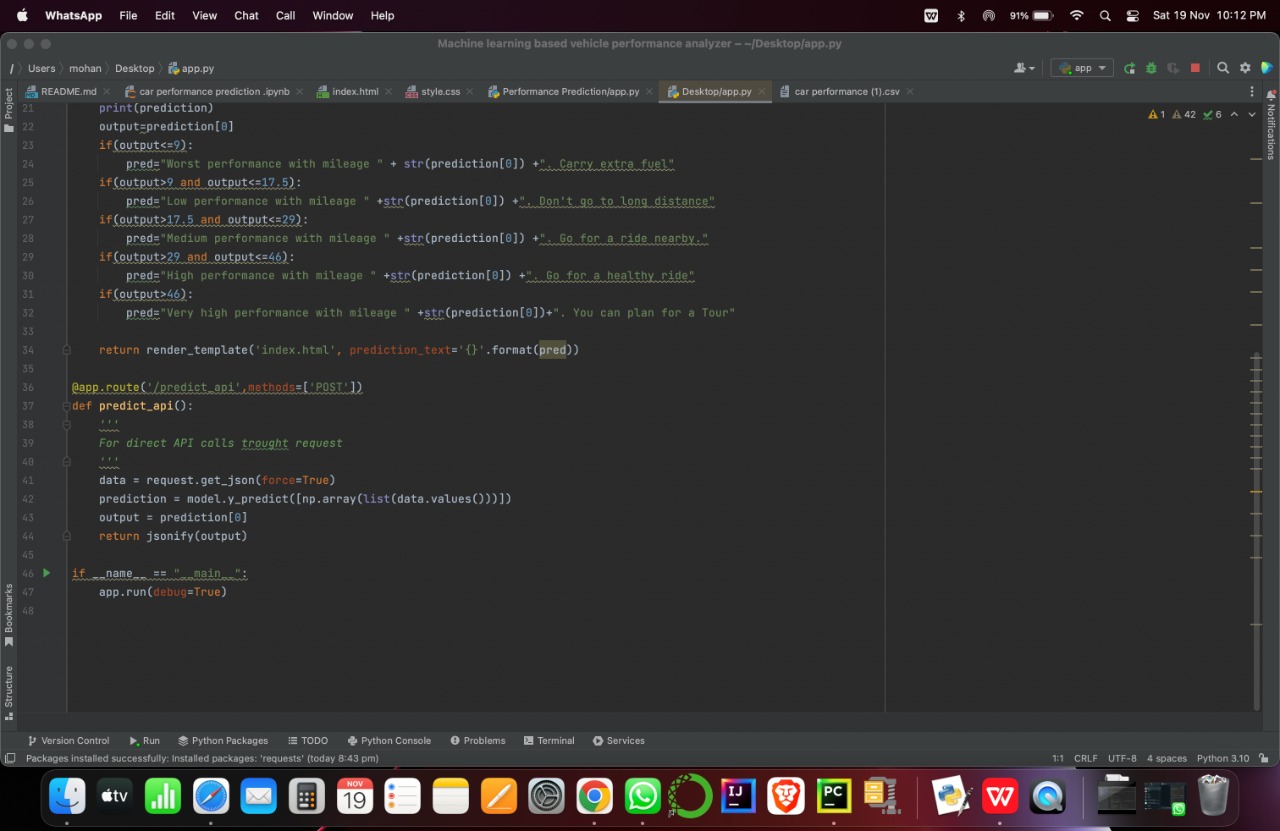
**Homepage  HTML  code :**

****

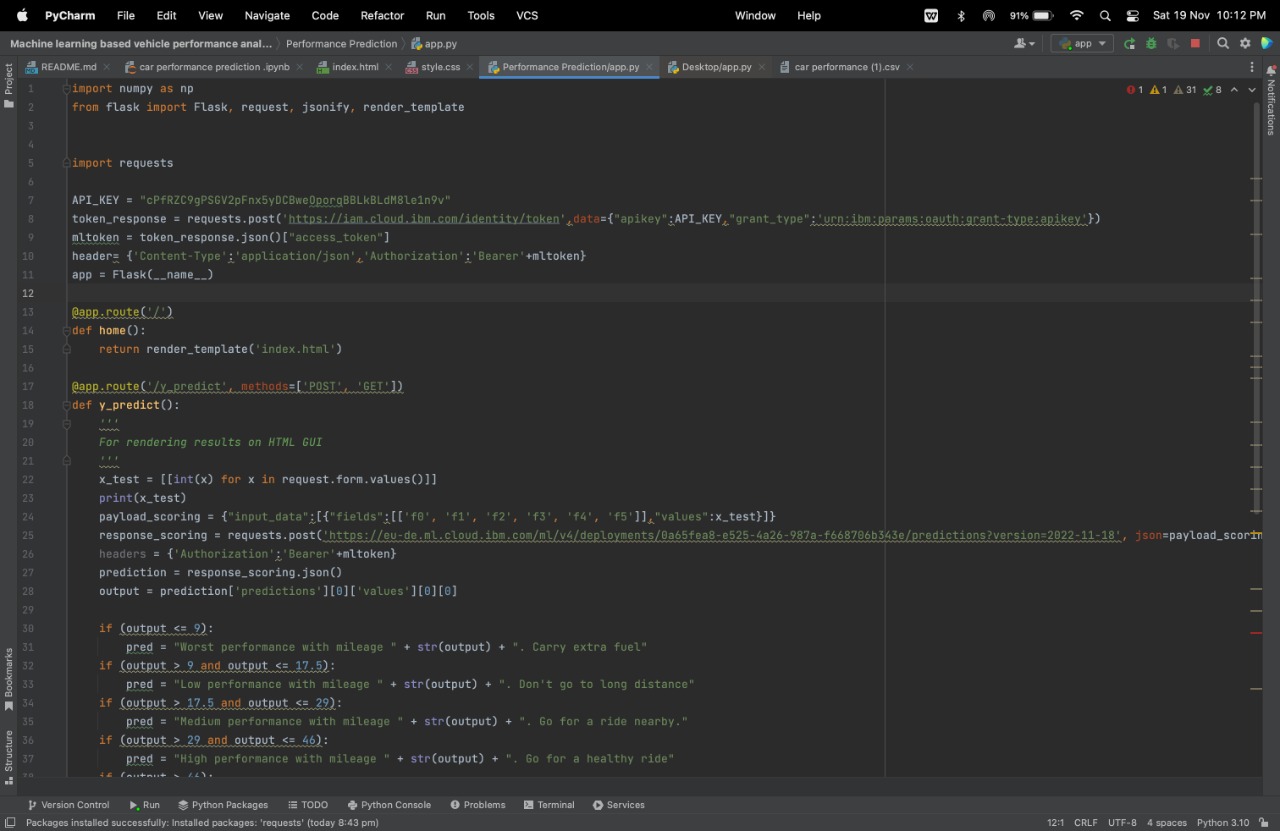
****

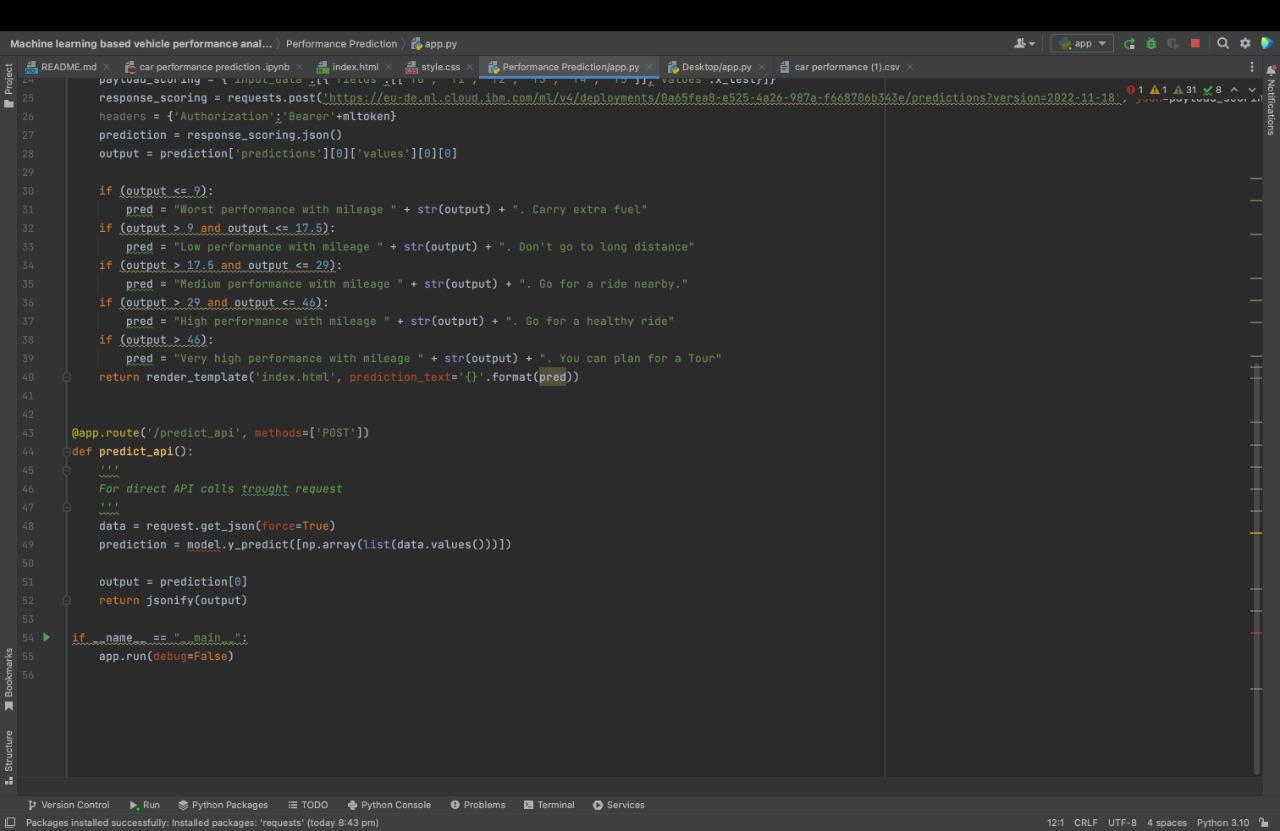
**Flask Code :**

****

****

**Integrate Flask Code :**





**TESTING**

**TESTING REPORT**

               esting of an individual software component or module is termed as Unit Testing. It is

typically done by the programmer and not by testers, as it requires detailed knowledge of the

internal program design and code.

The Code was developed in 3 separate parts-

1. AI Model developed using Jupyter Notebook

2. Web Front end was developed using VS Code

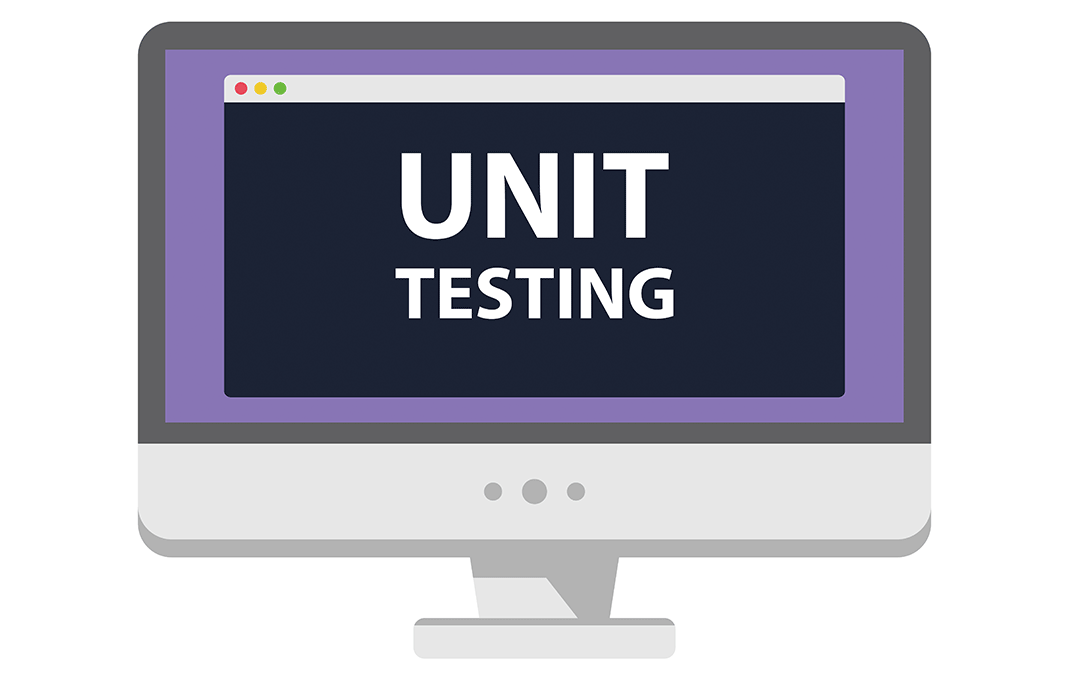
3. Backend Database was developed using MongoDB

|  |  |
| --- | --- |
| PROJECT NAME | MACHINE LEARNING VECHICLE PERFORMANCE ANALYZER |
| PROJECT TYPE | APPLIED DATA SCIENCE |

|  |  |
| --- | --- |
| DEVELOPER | KUMMITHA SIVA MOHAN REDDY |
| LANGUAGE | PYTHON,HTML,CSS,JAVA SCRIPT |
| TOTAL NUMBER OF TEST CASES | 50 |
| NUMBER OF TEST CASES EXCUTED | 49 |
| NUMBER OF TEST CASES PASSED | 45 |
| NUMBER OF TEST CASES FAILED | 4-DUE TO TECHNICAL ISSUES |

**UNIT TESTING:**

       Unit testing is carried out screen-wise, each screen being identified as an object. Attention is diverted to individual modules, independently to one another to locate errors .Thishas enabled the detection of errorsin coding and logic .Thisisthe first level of testing.In this, codes are written such thatfromone module ,we can move ontothenextmoduleaccordingtothechoiceweenter.



**SYSTEM TESTING:**

                        In this, the entire system was tested as a whole with all forms, code, modules and class modules .System testing is the stage of implementation, which is aimed at ensuring that the system works accurately and efficiently before live operation commences.

                                          It is a series of different tests that verifies that all system elements have been properly integrated and perform allocated functions.

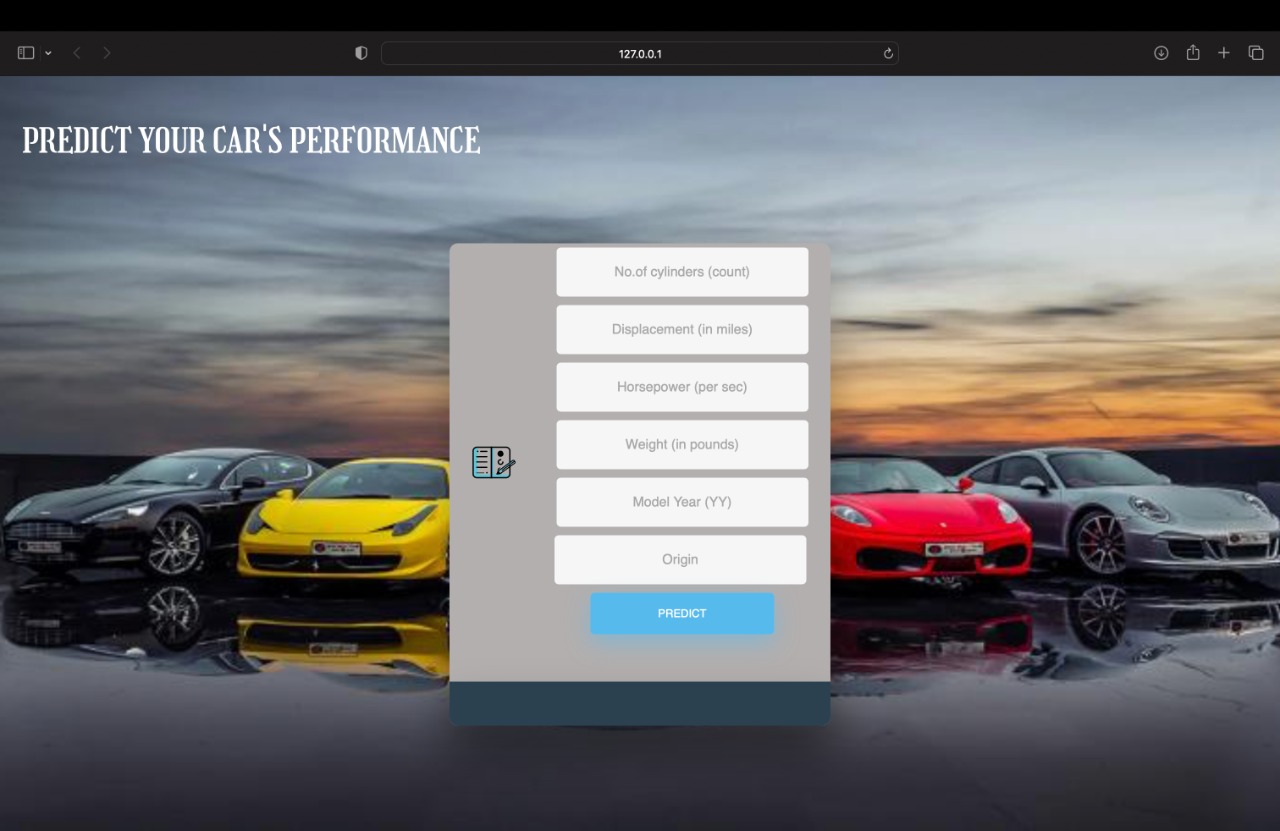
                              System testing makes logical assumptions that if all parts of the system are correct, the goal will be successfully achieved. Testing is the process of executing the program with the intent of finding errors.

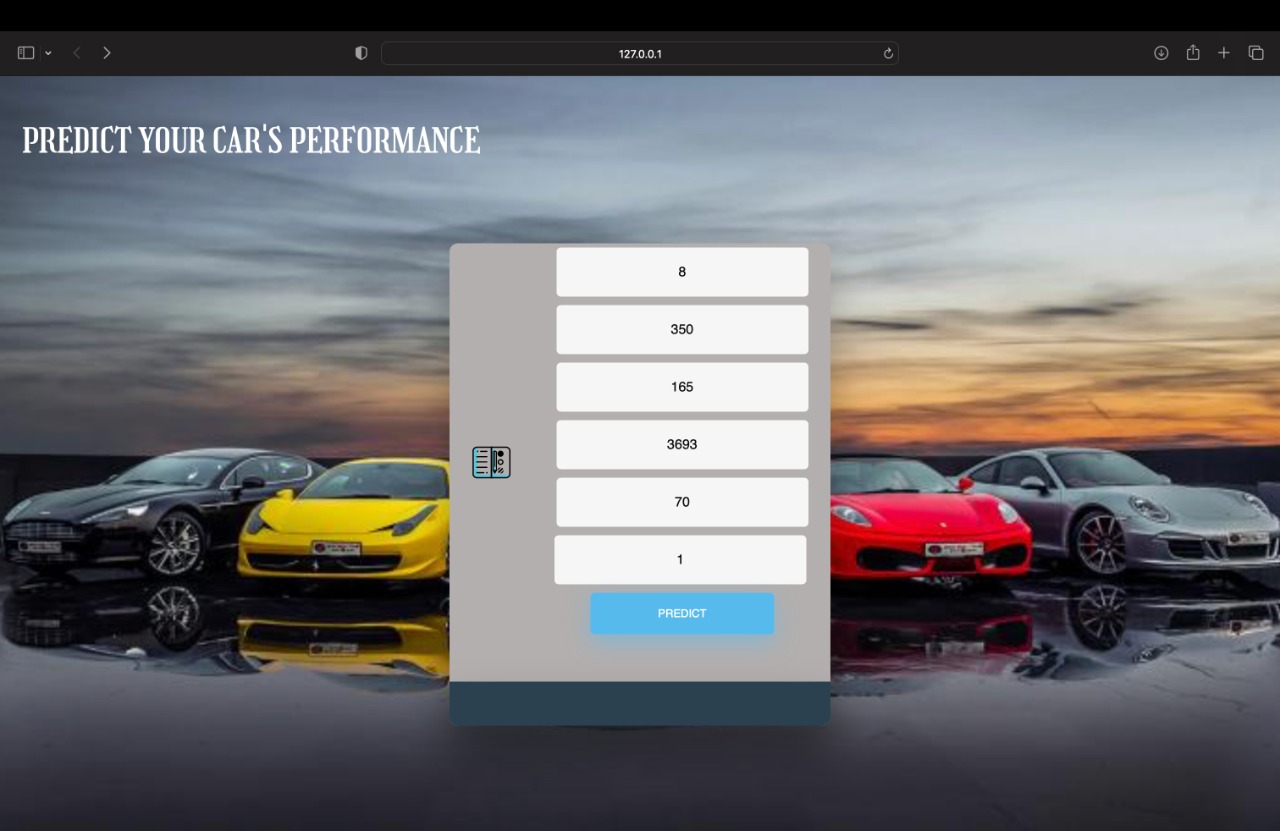
                                             Testing cannot show the absence of defects, it can only show that software errors are present.

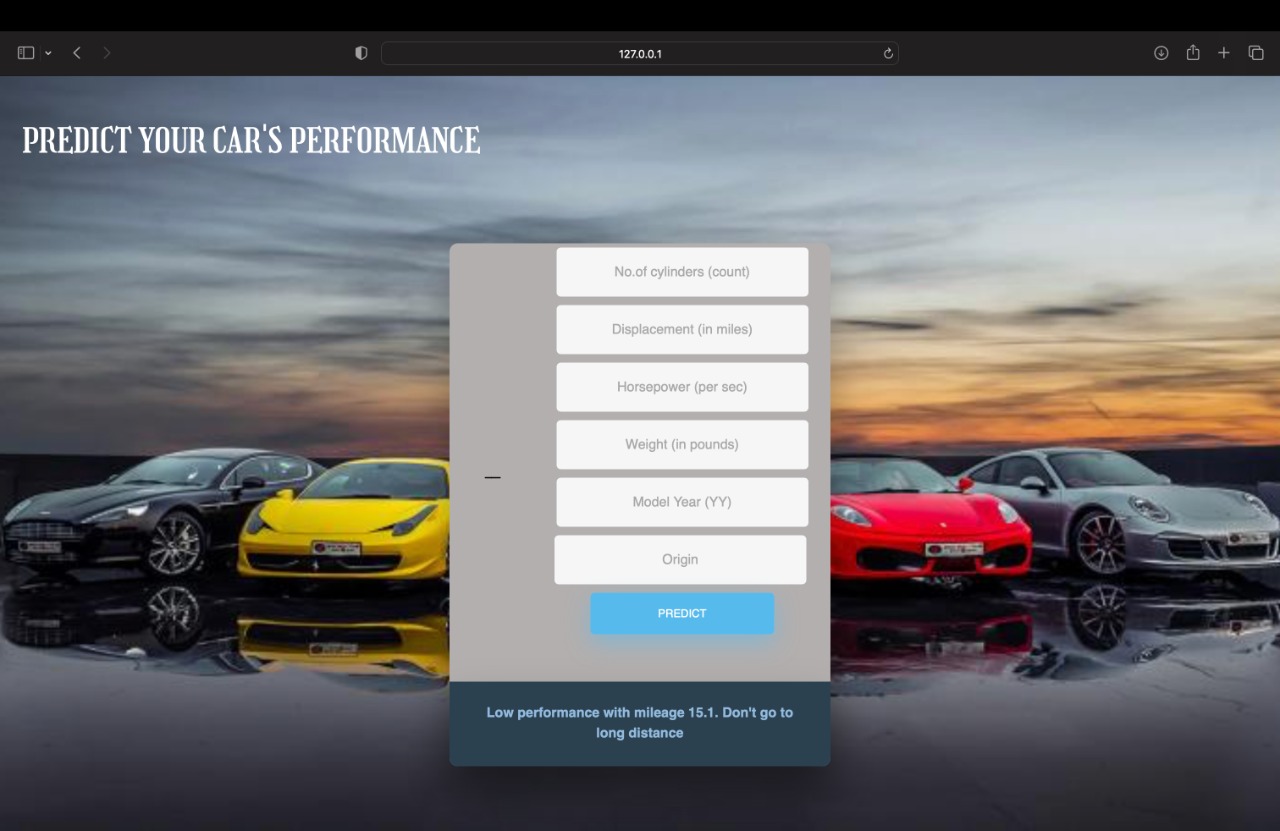


***8.1 Test cases :***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| TEST CASE | No of Cylinders | Displacement | HP | Weight | Year | Origin | Predicted Value |
| 1 | 8 | 307 | 130 | 3504 | 70 | 1 | 18.1 |
| 2 | 8 | 350 | 165 | 3693 | 70 | 1 | 15.2 |
| 3 | 4 | 130 | 95 | 2372 | 70 | 3 | 24.2 |
| 4 | 6 | 198 | 95 | 2833 | 70 | 1 | 22.3 |
| 5 | 4 | 104 | 95 | 2375 | 70 | 2 | 24.2 |







***8.2 User Acceptance Testing :***

 Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the Machine Learning-based Vehicle Performance Analyzer project at the time of the release to User Acceptance Testing (UAT).

Defect Analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Section | Total Cases | Not Tested | Fail | Pass |
| Print Engine | 9 | 0 | 0 | 9 |
| Client Application | 44 | 0 | 0 | 44 |
| Security | 2 | 0 | 0 | 2 |

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

1. Test Case Analysis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Resolution | Severity 1 | Severity 2 | Severity 3 | Severity 4 | Subtotal |
| By Design | 15 | 6 | 2 | 3 | 26 |
| Duplicate | 1 | 0 | 3 | 0 | 4 |
| External | 2 | 3 | 0 | 1 | 6 |
| Fixed | 12 | 3 | 5 | 22 | 42 |
| Not Reproduced | 0 | 0 | 1 | 0 | 1 |
| Skipped | 0 | 0 | 1 | 1 | 2 |
| Won't Fix | 0 | 4 | 2 | 1 | 7 |
| Totals | 30 | 16 | 14 | 28 | 88 |

This report shows the number of test cases that have passed, failed, and untested

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Outsource Shipping | 3 | 0 | 0 | 3 |
| Exception Reporting | 9 | 0 | 0 | 9 |
| Final Report Output | 4 | 0 | 0 | 4 |
| Version Control | 2 | 0 | 0 | 2 |

**9 . Results :**

9.1 Performance Metrics :

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

|  |  |  |  |
| --- | --- | --- | --- |
| S.No. | Parameter | Values | Screenshot |
| 1. | Metrics | Regression Model:                MAE - 1.7841,          MSE -  6.5057,          RMSE -2.5506 ,            R2 score – 0.9058 |  |
| 2. | Tune the Model | Hyperparameter  Tuning – |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| TEST CASE | No of Cylinders | Displacement | HP | Weight | Year | Origin | Predicted Value |
| 1 | 4 | 120 | 97 | 2506 | 72 | 3 | 23 |
| 2 | 4 | 98 | 80 | 2164 | 72 | 1 | 28 |
| 3 | 4 | 97 | 88 | 2100 | 72 | 3 | 27 |
| 4 | 8 | 350 | 175 | 4100 | 73 | 1 | 13 |
| 5 | 8 | 304 | 150 | 3672 | 73 | 1 | 14 |

**10. ADVANTAGES & DISADVANTAGES :**

**ADVANTAGES:**

* It helps users for predicting the vehicle performance.
* Here the chance of occurrence of error is less when compared with the existing system.
* It is fast, efficient and reliable.
* Avoids data redundancy and inconsistency.
* Very user-friendly.
* Easy accessibility of data

**DISADVANTAGES:**

* **computer literacy and network access**
* **Low Computer Literacy**
* **Security Concerns**
* **Authenticity**

**Infrastructural Requirement.**

**10 . Conclusion :**

The monitoring of car performance, especially gas consumption, has so far been approached

only very superficially. A typical fuel gauge, when closely monitored, shows an extremely non-

linear relationship between needle movement and fuel consumption. Inaccuraciesoccur

especially in the range of critical low fuel values of 5-10% or more. In the past, due to

thislimitation, some luxury cars had an audible and flashing light alarm function to indicate a

low fuel condition. These systems, which add to the existing fuel level, have no more accuracy

than the fuel level monitor alone. In recent years, with the availability of computer techniques

and reliable and less expensive computer equipment, a number of systems have been

developed to provide somewhatmore accurate information about vehicle performance.

**12. FUTURE SCOPE :**

This merits exploratory methods based on actual failures to deduct likely failure modes. This thesis presents two methods for data mining the vehicle maintenance records and vehicle usage data to learn usage or wear patterns indicative of failures. This requires detailed maintenance records where the failure root cause can be deducted with accurate date or mileages of the repair.

Further, more wide-spread adoption of predictive maintenance calls for automatic and less human-resource demanding methods, e.g. unsupervised algorithms with lifelong learning. Such methods are easier to scale up and they can thus be ubiquitously applied since much of the modelling is automated and requires little or no human interaction.

Maintenance predictions can be enhanced by combining the deviations in onboard data with off-board data sources such as maintenance records and failure statistics. This is exclusive product knowledge, only accessible to the vehicle manufacturers, which gives them an advantage in predicting maintenance. Still, data mining has yet to become a core competence of vehicle manufacturers, which makes the road to industrialisation long.

The aim of this thesis is to investigate how on-board data streams and off-board data can be used to predict the vehicle maintenance. More specifically, how on-board data streams can be represented and compressed into a transmittable size and still be relevant for maintenance predictions. Further, the most interesting deviations must be found for a given repair which requires new ways of combining semantic maintenance records with deviations based on compressed on-board data.

This can be accessed anytime anywhere, since it is a web application provided only an internet connection.

**13. APPENDIX :**

[Github Link](https://github.com/IBM-EPBL/IBM-Project-27292-1660053205" \t "undefined) : - https://github.com/IBM-EPBL/IBM-Project-21845-1659792766