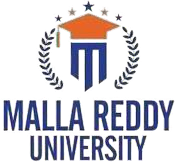
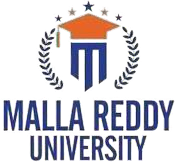
DATA ANAL YTICS WITH PYTHON



# Department of Computer Science &Engineering

## MALLA REDDY UNIVERSITY, HYDERABAD

**2023-2024**



**CO2 EMISSIONS OF PETROL AND DIESEL-BASED CARS**

**Designed and Developed by**

|  |  |
| --- | --- |
| 1.G. DEEPAK REDDY | 2111CS010114 |
| 2.B. DAKSHITHA REDDY | 2111CS010119 |
| 3.N. GANESH | 2111CS010141 |
| 4.D. GOVARDHAN | 2111CS010155 |
| 5.M. HARISH | 2111CS010166 |
|  |  |

**Guided by**

**Mr.D.I.P.MANIKUMAR**

# Department of Computer Science &Engineering

## MALLA REDDY UNIVERSITY, HYDERABAD

**2023-2024**



**CERTIFICATE**

This is to certify that this is the Data Analytics with Python lab report entitled “**CO2 EMISSIONS OF PETROL AND DIESEL-BASED CARS**”, submitted by **G.DEEPAKA REDDY- 2111CS010114, B.DAKSHITHA REDDY-2111CS010119, N.GANESH - 2111CS010141,**

**D.GOVARDHAN - 2111CS010155, M.HARISH - 2111CS010166** B. Tech **III** year II semester, Department of CSE during the year 2022-23. The results embodied in this report have not been submitted to any other university or institute for the award of any degree or diploma.

**GUIDE**

**Mr.D.I.P.MANIKUMAR**



# DECLARATION

I declare that this project report titled **CO2 EMISSIONS OF PETROL AND DIESEL-BASED CARS** submitted in partial fulfillment of the degree of B. Tech in CSE is a record of original work carried out by me under the supervision of **MR.D.I.P.MANIKUMAR** and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgements have been made wherever the findings of others have been cited.

|  |
| --- |
|  |

|  |  |
| --- | --- |
| G.DEEPAK REDDY | 2111CS010114 |
| B.DAKSHITHA REDDY | 2111CS010119 |
| N.GANESH | 2111CS010141 |
| D.GOVARDHAN | 2111CS010155 |
| M.HARISH | 2111CS010166 |

# ACKNOWLEDGEMENT

We would like to express my deepest gratitude to my guide, for her valuable guidance, consistent encouragement, personal caring, timely help and providing us with an excellent atmosphere for doing research. All through the work, in spite of her busy schedule, she has extended cheerful and cordial support to us for completing this research work. She has been helpful at every step from beginning till the end of this project. We are thankful to her for the encouragement she had given us in completing the project. We express our heart full thanks to the Head of the Department, Departmentof Computer Science and Engineering, for all the help and infrastructure provided to us to complete the project successfully and her valuable guidance. We are also thankful to all other faculty and staff members of our department for their kind co- operation and help.

|  |  |
| --- | --- |
| G.DEEPAK REDDY | 2111CS010114 |
| B.DAKSHITHA REDDY | 2111CS010119 |
| N.GANESH | 2111CS010141 |
| D.GOVARDHAN | 2111CS010155 |
| M.HARISH | 2111CS010166 |

# ABSTRACT

This data analytics project focuses on analyzing the CO2 emissions of petrol and diesel-based cars. The dataset comprises emissions data collected from various models of conventional combustion engine vehicles. Through thorough data exploration, statistical analysis, and visualization techniques, the project aims to uncover patterns and trends in CO2 emissions, considering factors such as engine size, fuel efficiency, vehicle weight, and emission control systems. By understanding the determinants of CO2 emissions in traditional internal combustion engine vehicles, this study seeks to provide insights for policymakers, automotive manufacturers, and consumers aiming to mitigate environmental impact and promote more sustainable transportation practices.

**TOOLS USED:** python, Jupyter notebook, pandas, numpy, scikit-learn

|  |  |
| --- | --- |
| **TABLE OF CONTENT** | |
| **DESCRIPTION** | **PAGE NUMBER** |
| **CERTIFICATE** | **iii** |
| **DECLARATION** | **iv** |
| **ACKNOWLEDGEMENTS** | **v** |
| **ABSTRACT** | **vi** |
| **LIST OF FIGURES** | **ix** |
| **LIST OF TABLES** |  |
| **Chapter 1 Introduction** |  |
| **1.1Introduction** | **1** |
| **1.2Problem Statement** | **2** |
| **1.3Objective of Project** | **3** |
| **1.4Goal of Project** | **5** |
| **Chapter 2 Problem Identification** |  |
| **2.1Existing System** | **7** |
| **2.2Proposed System** | **7** |
| **Chapter 3 Requirements** |  |
| **3.1 Software Requirements** | **9** |
| **3.2 Hardware Requirements** | **10** |
| **Chapter 4 Design and Implementation** |  |
| **4.1Design** | **12** |
| **4.2Implementation** | **13** |

|  |  |
| --- | --- |
| **Chapter 5 Code** |  |
| **5.1Source Code** | **15** |
| **5.2Screenshot of Application** | **20** |
| **Chapter 6 Results & Conclusion** |  |
| **6.1Results** | **21** |
| **6.2Conclusion** | **22** |
| **REFERENCES** |  |

# LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **FIGURE** | **TITLE** | **PAGE NUMBER** |
| 4.1. | Architecture | 12 |
| 4.2. | Flow chart | 13 |

* 1. **INTRODUCTION**

# CHAPTER-1 INTRODUCTION

## In the contemporary discourse surrounding environmental sustainability, the analysis of carbon dioxide (CO2) emissions from vehicles stands as a critical endeavor. With the automotive industry being a significant contributor to global emissions, understanding the dynamics of CO2 emissions from petrol and diesel-based cars is paramount in addressing climate change and fostering a transition towards greener transportation solutions. This data analytics project aims to delve into the intricacies of CO2 emissions from petrol and diesel-based cars, exploring various factors that influence emissions levels and analyzing trends over time.

## The motivation behind this project stems from the urgent need to mitigate the environmental impact of vehicular emissions and promote sustainable practices within the automotive sector. Petrol and diesel-based cars are among the primary sources of CO2 emissions, contributing to air pollution, greenhouse gas emissions, and climate change. By conducting a comprehensive analysis of CO2 emissions data, we seek to identify patterns, trends, and insights that can inform policy decisions, technological innovations, and consumer choices aimed at reducing emissions and promoting cleaner alternatives.

## The objectives of this project are twofold: firstly, to analyze the factors influencing CO2 emissions from petrol and diesel-based cars, including engine specifications, fuel efficiency, vehicle age, and driving patterns; and secondly, to develop predictive models that can forecast CO2 emissions based on these factors. By leveraging data analytics techniques and machine learning algorithms, we aim to gain a deeper understanding of the complexities surrounding emissions levels and provide valuable insights for stakeholders ranging from policymakers to automotive manufacturers and consumers.

## The dataset for this project encompasses a wide range of variables related to CO2 emissions, including car make, model, year of manufacture, engine size, fuel type (petrol or diesel), and corresponding emissions levels. Data will be collected from reputable sources such as governmental databases, industry reports, and academic studies, ensuring the accuracy and reliability of the analysis. Preprocessing steps will be undertaken to clean, format, and prepare the data for analysis, including handling missing values, outliers, and inconsistencies.

## The analysis will involve exploratory data analysis (EDA) to uncover patterns, trends, and correlations within the dataset. Descriptive statistics, visualizations, and statistical techniques will be employed to gain insights into the distribution and variability of CO2 emissions from petrol and diesel-based cars. Additionally, predictive modeling techniques such as regression analysis and machine learning algorithms (e.g., random forest, gradient boosting) will be utilized to develop models that can forecast emissions levels based on input variables.

## The findings of this project hold implications for various stakeholders involved in the automotive industry and environmental policymaking. By elucidating the factors influencing CO2 emissions from petrol and diesel-based cars, policymakers can formulate regulations and incentives to incentivize cleaner technologies and promote fuel-efficient vehicles. Automotive manufacturers can utilize the insights gained to design and produce greener vehicles that minimize emissions and align with sustainability goals. Consumers can make informed choices regarding vehicle purchases, opting for models with lower emissions and reduced environmental impact.

## PROBLEM STATEMENT

## The objective of this data analytics project is to analyze carbon dioxide (CO2) emissions from petrol and diesel-based cars and develop predictive models to assess the condition of a car based on its emissions history. The project aims to address the following key questions:

## CO2 Emissions Analysis:

## - How do CO2 emissions vary among petrol and diesel-based cars?

## - What are the main factors influencing CO2 emissions, such as car make, model, year of manufacture, and engine specifications?

## - Can we identify trends and patterns in CO2 emissions over time, and how do these vary between different types of vehicles?

## Predictive Modeling for Car Condition Assessment:

## - Given a car's previous emissions data, can we predict its present condition or health status?

## - How accurately can we assess the condition of a car based on its emissions history, and what factors contribute to this assessment?

## - What insights can we gain from the predictive models regarding the relationship between emissions patterns and car condition?

## By addressing these questions, this project aims to provide valuable insights for various stakeholders, including policymakers, automotive manufacturers, and consumers. The predictive models developed will assist in assessing the environmental impact of petrol and diesel-based cars and informing decisions related to vehicle maintenance, emissions regulations, and sustainability initiatives. Additionally, the analysis will contribute to a deeper understanding of the factors influencing CO2 emissions and their implications for car condition and overall environmental health.

## OBJECTIVE OF PROJECT

The primary objective of this data analytics project is to analyze carbon dioxide (CO2) emissions from petrol and diesel-based cars and develop predictive models to assess the condition of a car based on its emissions history. The project aims to achieve the following specific goals:

1. CO2 Emissions Analysis:

- Explore and analyze CO2 emissions data from a diverse range of petrol and diesel-based cars, including attributes such as car make, model, year of manufacture, engine specifications, and corresponding emissions levels.

- Identify trends, patterns, and correlations within the dataset to gain insights into the variability of CO2 emissions among different types of vehicles.

- Determine the main factors influencing CO2 emissions and their relative importance in predicting emissions levels.

2. Predictive Modeling for Car Condition Assessment:

- Develop predictive models using machine learning algorithms to assess the condition of a car based on its emissions history.

- Utilize historical emissions data as input variables to predict the present condition or health status of the car.-

- Evaluate the accuracy and reliability of the predictive models in assessing the condition of petrol and diesel-based cars.

- Investigate the relationship between emissions patterns and car condition to identify key indicators of vehicle health.

3. Insights and Recommendations:

- Provide actionable insights and recommendations based on the findings of the analysis to stakeholders, including policymakers, automotive manufacturers, and consumers.

- Inform decision-making processes related to emissions regulations, vehicle maintenance, and sustainability initiatives.

- Enhance understanding of the environmental impact of petrol and diesel-based cars and promote evidence-based strategies for reducing emissions and improving overall environmental health.

By achieving these objectives, the project aims to contribute to the broader discourse on environmental sustainability and promote informed decision-making in the automotive sector. Through data-driven insights and predictive modeling techniques, the project seeks to pave the way towards a cleaner, greener future for transportation, mitigating the adverse effects of vehicular emissions and fostering a transition towards sustainable transportation solutions.

## GOAL OF PROJECT

The overarching goal of this data analytics project is to leverage CO2 emissions data from petrol and diesel-based cars to develop predictive models that assess the condition of a car based on its emissions history. The primary objectives are to:

1. Understand Emissions Patterns: Analyze CO2 emissions data to identify trends, patterns, and correlations among different types of vehicles. By exploring factors such as car make, model, year of manufacture, and engine specifications, the project aims to gain insights into the variability of emissions levels and the factors influencing them.

2. Develop Predictive Models: Utilize machine learning algorithms to develop predictive models that assess the condition of a car based on its emissions history. Historical emissions data will serve as input variables to predict the present condition or health status of the vehicle. The goal is to develop accurate and reliable models that can effectively assess car condition.

3.Provide Insights and Recommendations: Translate the findings of the analysis into actionable insights and recommendations for stakeholders, including policymakers, automotive manufacturers, and consumers. These insights will inform decision-making processes related to emissions regulations, vehicle maintenance, and sustainability initiatives, ultimately contributing to environmental sustainability and promoting greener transportation solutions.

By achieving these goals, the project aims to advance our understanding of the relationship between CO2 emissions and car condition and provide practical tools for assessing and managing the environmental impact of petrol and diesel-based cars. Through data-driven insights and predictive modeling techniques, the project seeks to empower stakeholders to make informed decisions that contribute to a cleaner, healthier, and more sustainable future for transportation.

# CHAPTER-2 PROBLEM IDENTIFICATION

## EXISTING SYSTEM

In the existing system, the assessment of a car's condition typically relies on manual inspection by mechanics or technicians. This process involves visual examination, diagnostic tests, and sometimes the use of specialized equipment to evaluate various components of the vehicle, such as the engine, transmission, brakes, and exhaust system. While this approach can provide valuable insights into the condition of individual car parts, it often lacks a comprehensive analysis of the vehicle's overall health.

Limitations of the existing system include:  
- Reliance on manual inspection methods, which can be time-consuming, subjective, and prone to human error.  
- Limited integration of emissions data into the assessment process, leading to a partial understanding of the vehicle's overall condition.  
- Reactive rather than proactive approach to maintenance, resulting in potential breakdowns or costly repairs that could have been prevented with early detection of issues.

## PROPOSED SYSTEM

The proposed system aims to enhance the assessment of a car's condition by integrating CO2 emissions data into the analysis process. The system will leverage machine learning algorithms to develop predictive models that assess the health of a car based on its emissions history. The proposed system consists of the following components:

- Data Collection: CO2 emissions data, along with other relevant attributes such as car make, model, year of manufacture, and engine specifications, will be collected from various sources, including emissions testing facilities, vehicle inspections, and automotive databases.

- Data Analysis: The collected data will be analyzed to identify patterns, trends, and correlations between emissions levels and car condition. Exploratory data analysis (EDA) techniques will be employed to gain insights into the relationship between emissions data and mechanical integrity.

- Predictive Modeling: Machine learning algorithms, such as regression analysis or ensemble methods like random forest, will be used to develop predictive models that assess the condition of a car based on its emissions history. Historical emissions data will serve as input variables to predict the present condition or health status of the vehicle.

- Validation and Evaluation: The predictive models will be validated and evaluated using appropriate metrics to assess their accuracy, reliability, and generalization performance. Cross-validation techniques and model validation protocols will be employed to ensure robustness and validity.

- Integration and Deployment: Once validated, the predictive models will be integrated into existing automotive systems or diagnostic tools for real-time assessment of car condition. The system will be deployed in automotive service centers, inspection facilities, or as part of onboard vehicle diagnostics systems.

The proposed system aims to revolutionize the assessment of car condition by providing a proactive and data-driven approach to maintenance and diagnostics. By leveraging CO2 emissions data, the system will enable early detection of mechanical issues, reduce the risk of breakdowns, and optimize vehicle performance and longevity. Additionally, the integration of emissions data into the assessment process will contribute to environmental sustainability efforts by promoting cleaner and more efficient transportation solutions

# CHAPTER-3 REQUIREMENTS

## SOFTWARE REQUIREMENTS

To develop a co2 emissions of cars using machine learning, you'll need a combination of software tools and libraries for various tasks such as data preprocessing, model training, and evaluation. Here's a list of essential software requirements:

1. **Programming Language:**

* Python: A versatile and widely used programming language for machine learning, widely used in the machine learning community, with extensive libraries and frameworks.

1. **Integrated Development Environment (IDE):**

* Jupyter Notebooks: Great for interactive development and data exploration.

1. **Data Manipulation and Analysis:**

* Pandas: For data manipulation and analysis.

1. **Data Visualization:**

* Matplotlib and Seaborn: For creating visualizations of the data.

1. **Machine Learning Libraries:**

* Scikit-learn: Provides simple and efficient tools for data mining and data analysis.

## HARDWARE REQUIREMENTS

### CPU (Central Processing Unit):

A multi-core CPU with good clock speed is essential for preprocessing data and training machine learning models. An Intel i5 or i7 processor, or an equivalent AMD processor, isa good starting point.

### RAM (Random Access Memory):

Adequate RAM is crucial, especially when working with large datasets and complex models. A minimum of 8 GB is recommended, but 16 GB or more is preferable for more extensive projects.

### GPU (Graphics Processing Unit):

For faster training of deep learning models, especially if using frameworks like TensorFlow or PyTorch, having a GPU is highly beneficial. NVIDIA GPUs, such as GeForce GTX or RTX series, or professional GPUs like NVIDIA Quadro or Tesla, are commonly used for machine learning tasks.

### Storage:

A solid-state drive (SSD) is recommended for faster data access during preprocessing and training. Ensure sufficient storage capacity for your datasets and model files. If working with large datasets, consider having multiple drives or using cloud storage solutions.

### Internet Connection:

A reliable internet connection is essential for downloading datasets, model libraries, and updates. Additionally, if you plan to use cloud-based services for training or deployment, a stable internet connection is crucial.

### Motherboard:

Ensure compatibility with the selected CPU and GPU. A motherboard with multiple Pistols may be beneficial if considering multiple GPUs for parallel processing.

### Power Supply:

Adequate power supply is essential, especially if using a dedicated GPU, to ensure stable and consistent performance.

### Cooling System:

Depending on the intensity of your machine learning tasks, consider additional cooling solutions to prevent overheating, especially if using a high-performance GPU.

### Backup System:

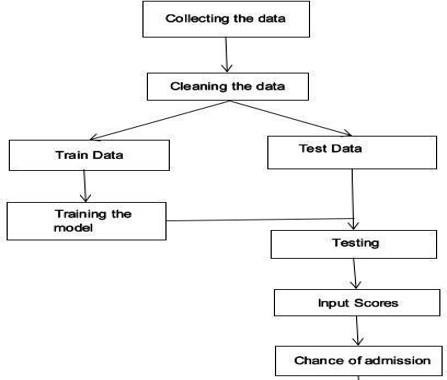
Regularly backup your work to prevent data loss, especially if working with valuable datasets.

* 1. **DESIGN**

# CHAPTER- 4

**DESIGN AND IMPLEMENTATION**

Designing a system for telecom churn prediction involves combining various factors to assess the likelihood of a candidate being able to churn away.



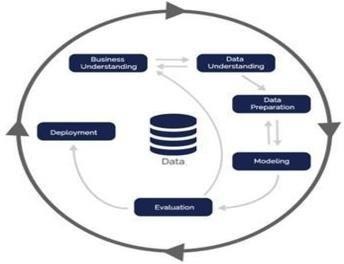
## Figure 4.1: Architecture

* 1. **IMPLEMENTATION Methodology**

Cross-Industry Standard Process (CRISP) methodology was followed in this research.

* **Business Understanding-**It is about understanding the requirements needed for the project like datasets ,different ML algorithms**.**
* **Data Understanding-I**t is analyzing of the different features present in dataset and their importance.
* **Data preparation-**It is integration of data taken by removing excessive features and data cleaning.
* **Modelling-**Multiple machine learning models were used to predict the popularity of song.
* **Evaluation:** Models developed were evaluated based on their performance and accuracy.
* **Deployment:** Once the models were evaluated they were integrated with code developed for user interface .

## Figure 4.2: Flow chart



**Algorithms Used**

Multiple machine learning algorithms were used for this research, Logistic Regression, Random forest algorithms were used to predict the likelihood of the students getting admission into university based on their profile

**Linear Regression:** It is a supervised machine learning algorithm that uses available data to create the model and further that model can be applied to predict the target variable. Linear regression quantifies the relationship between one or more predictor variable(s) and one outcome variable.

**Support Vector Regressor:** Support Vector Regression (SVR) is a type of machine learning model used for regression tasks. It is an extension of Support Vector Machines (SVM), a popular algorithm for classification. SVR, however, is designed for predicting continuous outcomes rather than class labels.

**Random Forest:** Random Forest is a powerful and versatile supervised machine learning algorithm that grows and combines multiple decision trees to create a “forest.”

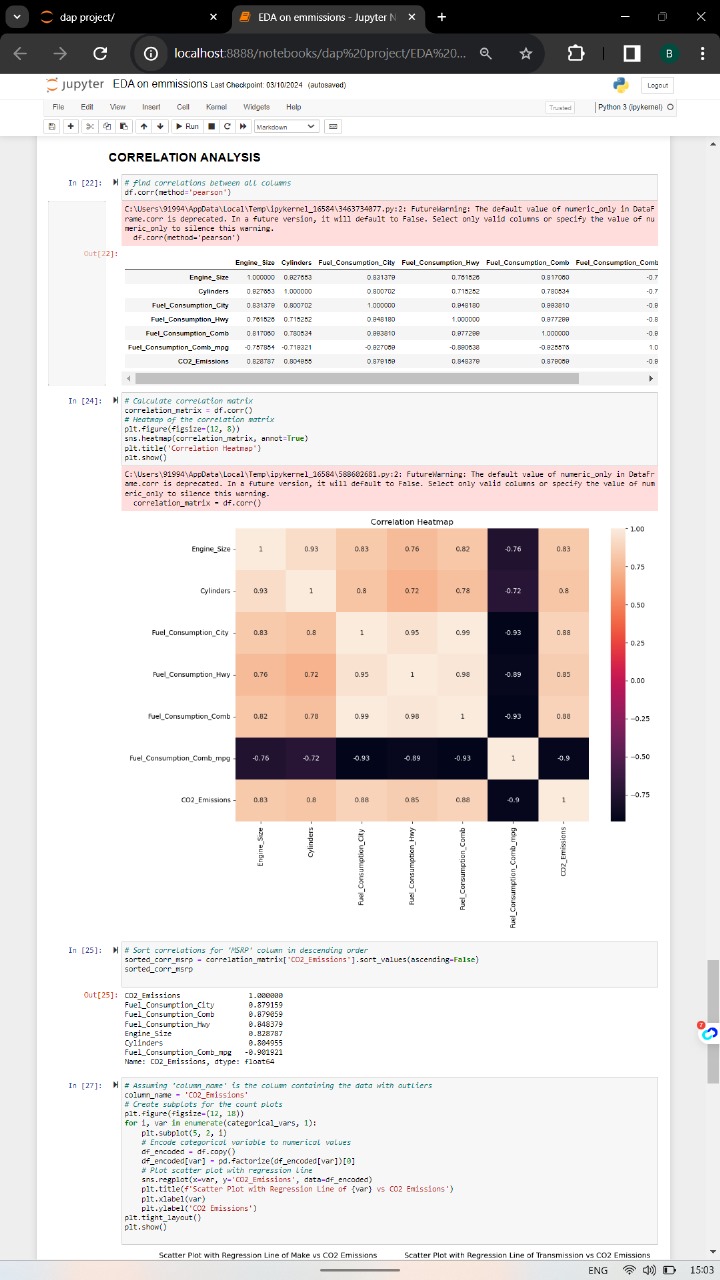
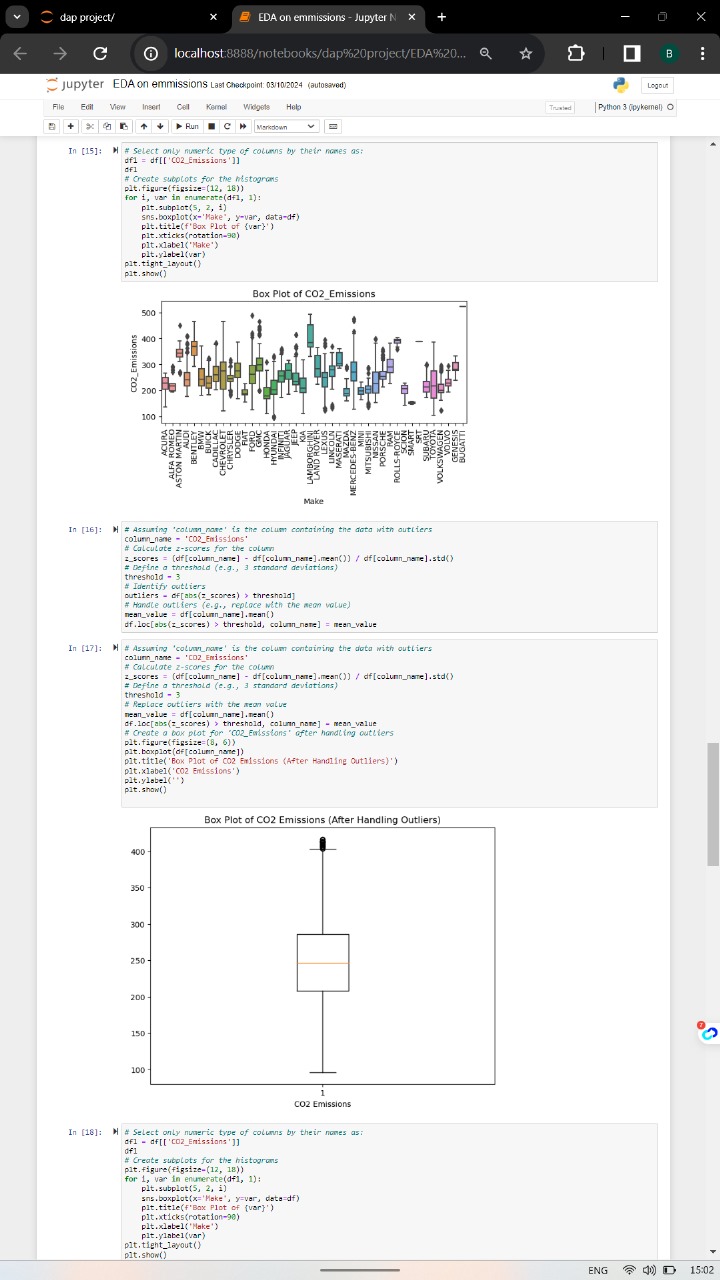
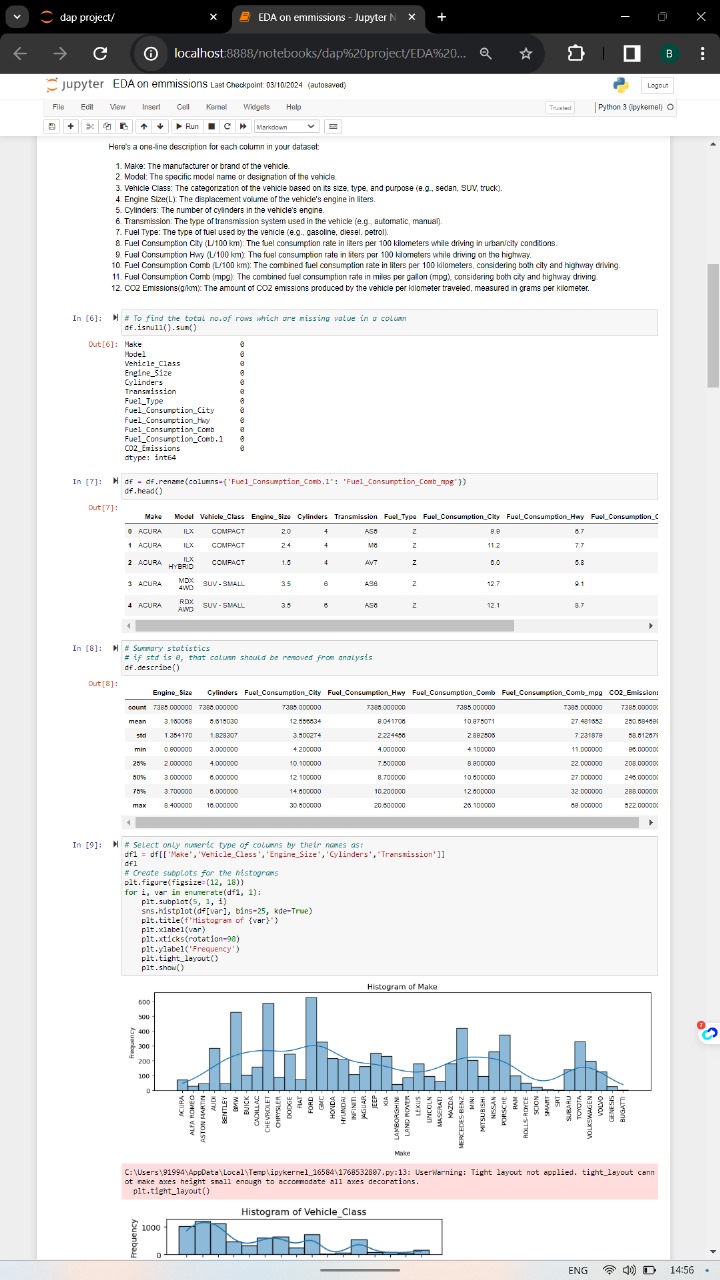
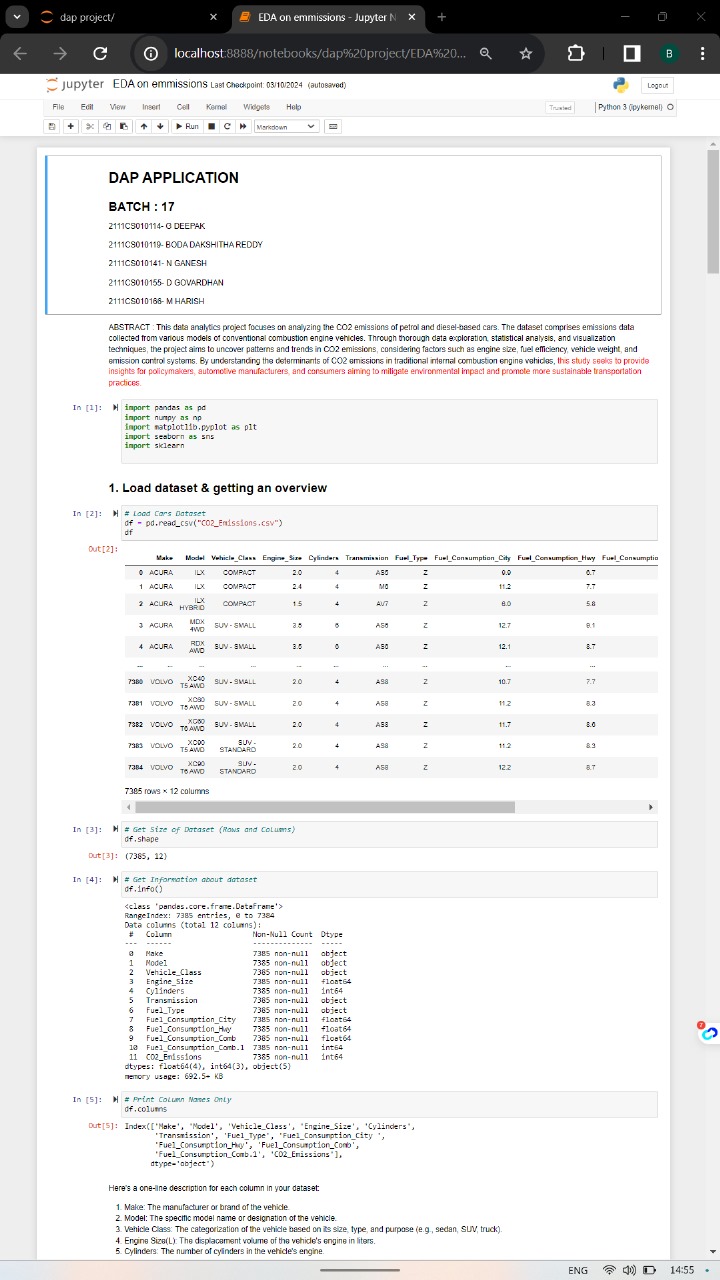
**Logistic Regression:** Logistic Regression is a statistical method used for binary classification, predicting the probability of an instance belonging to one of two classes. Despite its name, logistic regression is employed for classification tasks rather than regression tasks.

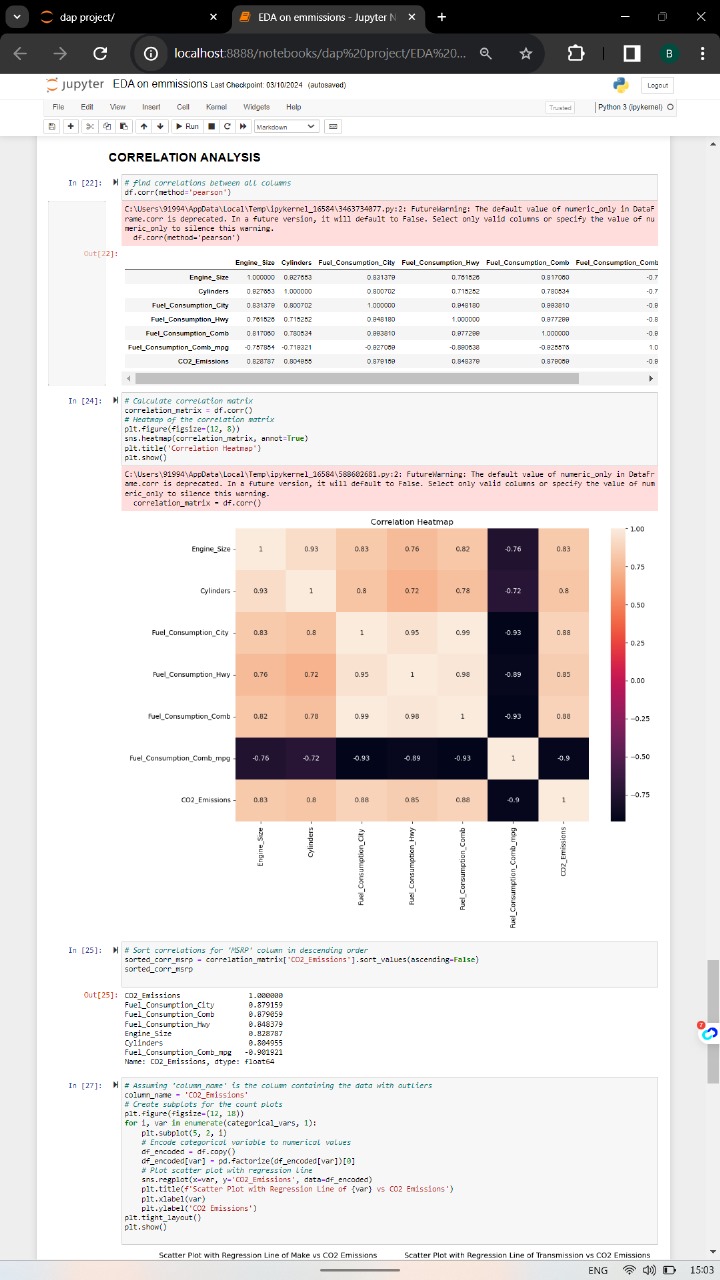
**k-Nearest Neighbors Classifier:** A k-Nearest Neighbors (k-NN) classifier is a type of instance-based learning algorithm used for classification tasks in machine learning. It is a simple and intuitive method that classifies a data point based on the majority class of its k nearest neighbors in the feature space.

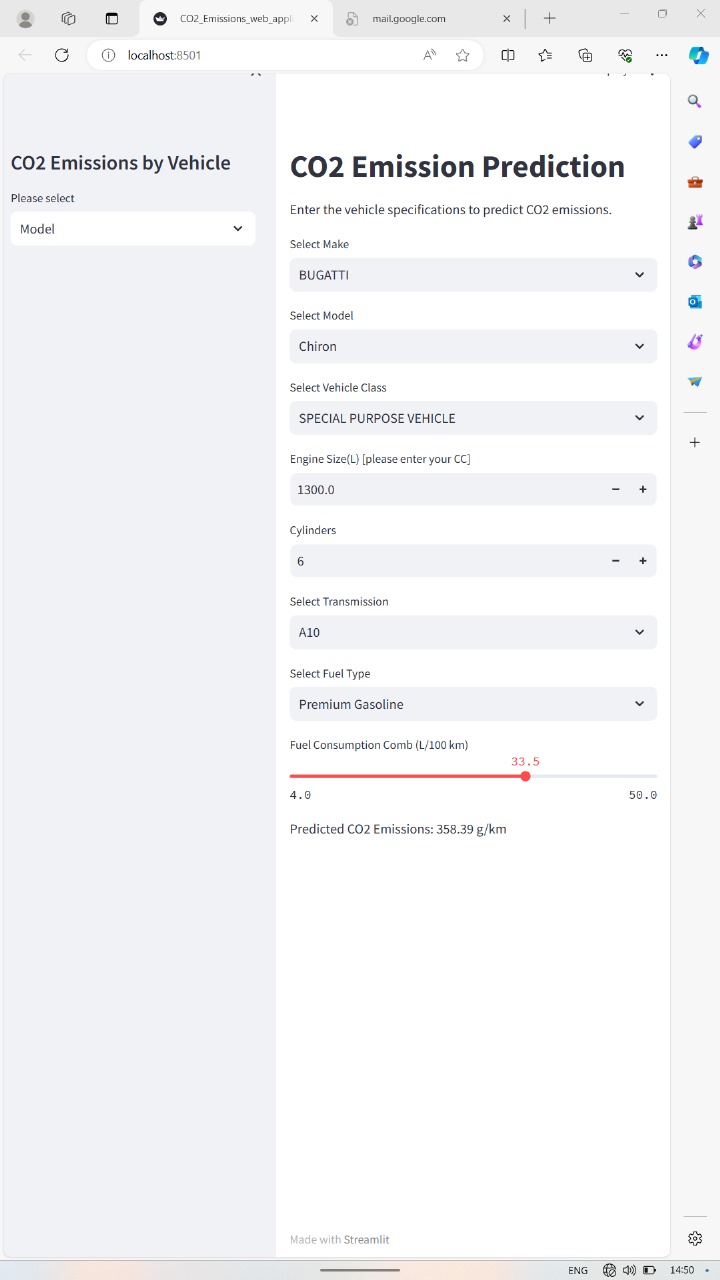
**Gradient Boosting Classifier:** Gradient Boosting is an ensemble learning technique used for both classification and regression tasks. Gradient Boosting classifiers, in particular, are popular for solving classification problems. This technique builds a strong predictive model by combining the predictions of multiple weak models, typically decision trees. The process involves sequentially adding new models that correct the errors of the previous ones.

**5.1 Source Code**

# CHAPTER – 5 CODE







**CHAPTER-6 RESULTS AND CONCLUSION**

## RESULTS

The result of a telecom churn prediction using machine learning relays on various factors, including the quality of the data, the features used for prediction, the chosen machine learning algorithm, and the overall model performance. The outcome will typically be a prediction or probability indicating the likelihood of an applicant being admitted.

## Probability Score:

* The model might provide a probability score for song, indicating the likelihood of hit.

## Binary Prediction:

* The model might provide a binary prediction, such as "churned" or "Not churned," based on a predefined threshold

## Confidence Interval:

* Instead of a single probability score, the model might provide a confidence interval, offering a range within which the true probability of hit or fail.

## Performance Metrics:

* Evaluation metrics, such as accuracy, precision, recall, and F1 score, will provide insights into how well the model is performing. These metrics help assess the model's ability to make accurate predictions.

## Feature Importance:

* The model may also provide information on the importance of different features in the prediction. This helps understand which factors contribute more significantly to the popularity.

## CONCLUSION

In conclusion, the telecom churn prediction model, built using a Logistic Regression model, demonstrates its efficacy in forecasting the churn of a telecom customer based on key features .In conclusion, the telecom churn prediction project represents a significant step forward in the quest to mitigate churn and maximize customer retention within the telecommunications industry. By leveraging advanced data analytics techniques, real-time decision-making capabilities, and personalized retention strategies, operators can enhance customer satisfaction, optimize revenue streams, and drive long-term business success.

# REFERENCES

* + 1. Abdi, F., & Abdollahzadehgan, A. (2018). Customer churn prediction in telecom using machine learning in big data platform. International Journal of Database Theory and Application, 11(10), 107-118.
    2. Fatima, I., Shafiq, M. Z., & Bashir, M. A. (2020). A Comprehensive Survey of Machine Learning Techniques for Telecommunication Churn Prediction. Journal of Network and Computer Applications, 169, 102765.
    3. Yin, P. Y., & Liao, S. H. (2010). Customer churn prediction by hybrid neural network and decision tree. Expert Systems with Applications, 37(1), 479-486.