

**INTRODUCTION**

The term air quality refers to the degree to which the air in a particular place is free from pollutants.Air pollutants are substances present in the atmosphere at concentrations above their normal background levels which can have a measurable effect on humans, animals and vegetation.Air quality is measured with the Air Quality Index, or AQI. The AQI works like a thermometer that runs from 0 to 500 degrees.However, instead of showing changes in the temperature, the AQI is a way of showing changes in the amount of pollution in the air.

**What is in the air?**

The air in our atmosphere is mostly made up of two gases that are essential for life on Earth: nitrogen and oxygen. However, the air also contains smaller amounts of many other gases and particles.

**MAJOR AIR POLLUTANTS:**

1.Ground level ozone

2.Carbon monoxide

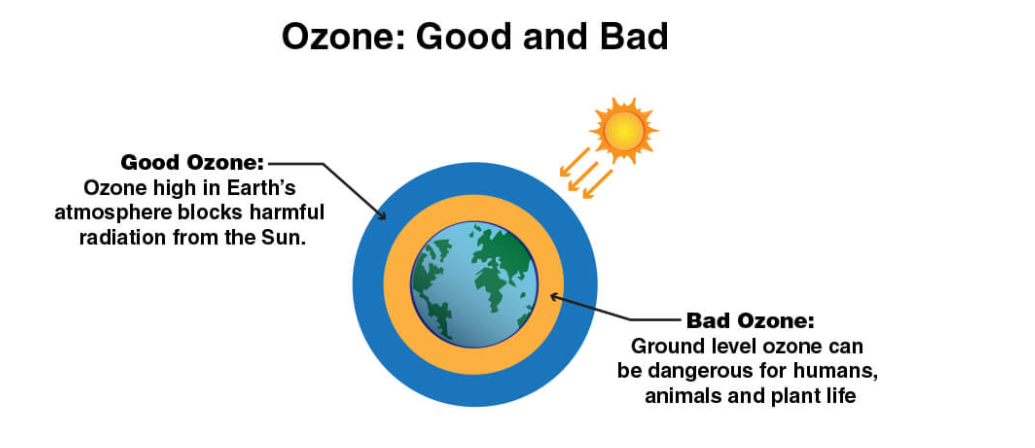
3.Sulfur dioxide

4.Nitrogen dioxide

5.Airborne particles, or aerosols

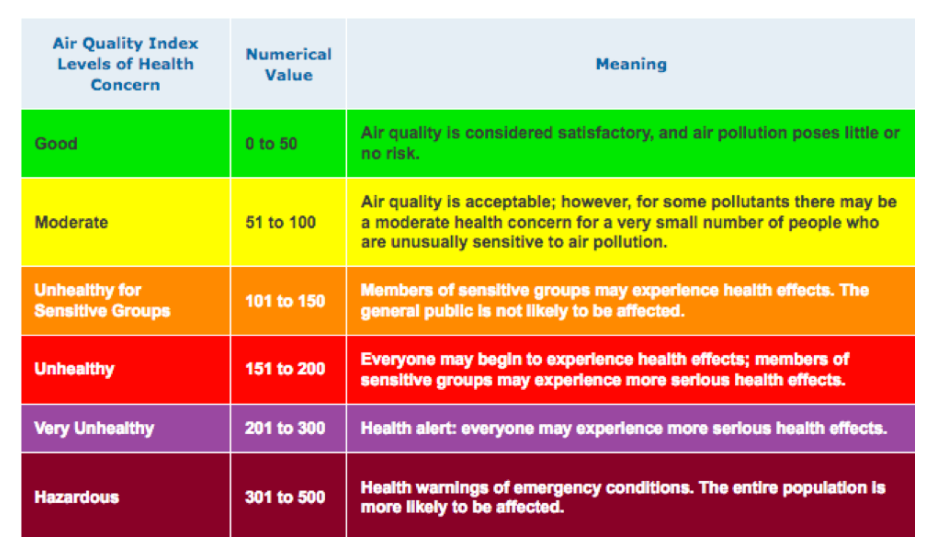
**What are some things that cause bad air quality?**

Ozone is a gas you’ve probably heard of as a layer high up in Earth’s atmosphere. This ozone layer is a good thing—it helps block us from the Sun’s harmful radiation. However, ground level ozone is bad for human health. It is created when sunlight reacts with certain chemical emissions (for example nitrogen dioxide, carbon monoxide and methane). These chemicals can come from industrial facilities, car exhaust, gasoline vapors and other sources.

****

**How is air quality communicated?**

Air quality is reported using the Air Quality Index (AQI). The AQI has six categories that communicate the level of health concern using specific colors. Code Green and Yellow means the air is generally safe for everyone. Code Orange is unhealthy for sensitive groups, including children, senior citizens, and people with heart and lung diseases. Code Red and Purple means the air is unhealthy for everyone, and Code Maroon is a health warning of emergency conditions.

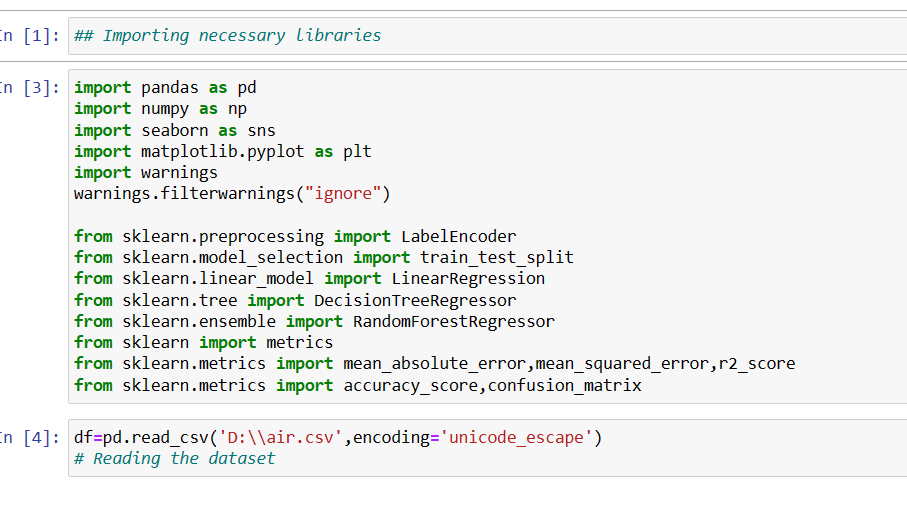


Instruments on the ground and satellites orbiting Earth collect information about what is in our air. For example, satellites in NOAA’s GOES-R (short for Geostationary Operational Environmental Satellites-R) Series monitor the particle pollution in our atmosphere.

JPSS can also observe the movement of aerosols from one side of the planet to the other. JPSS can also measure carbon monoxide which is associated with poor air quality resulting from wildfires.

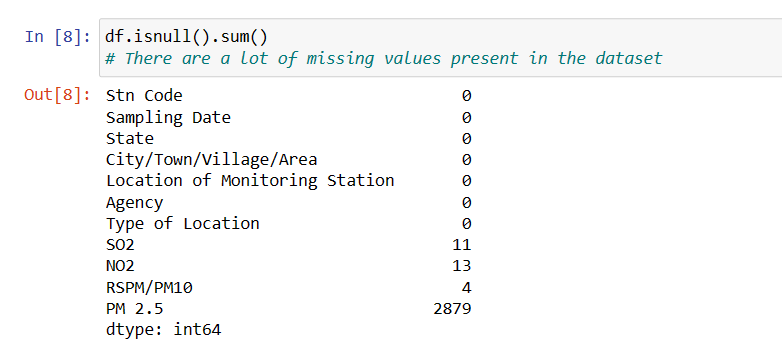
**DATA EXTRACTION**

Data extraction is the act or process of retrieving data out of (usually unstructured or poorly structured) data sources for further data processing or data storage (data migration). The import into the intermediate extracting system is thus usually followed by data transformation and possibly the addition of metadata prior to export to another stage in the data workflow. Extracting data from these unstructured sources has grown into considerable technical challenge where as historically data extraction has had to deal with changes in physical hardware formats, the majority of current data extraction deals with extracting data from these unstructured data sources, and from different software formats.



**DATA PRE-PROCESSING**

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task.

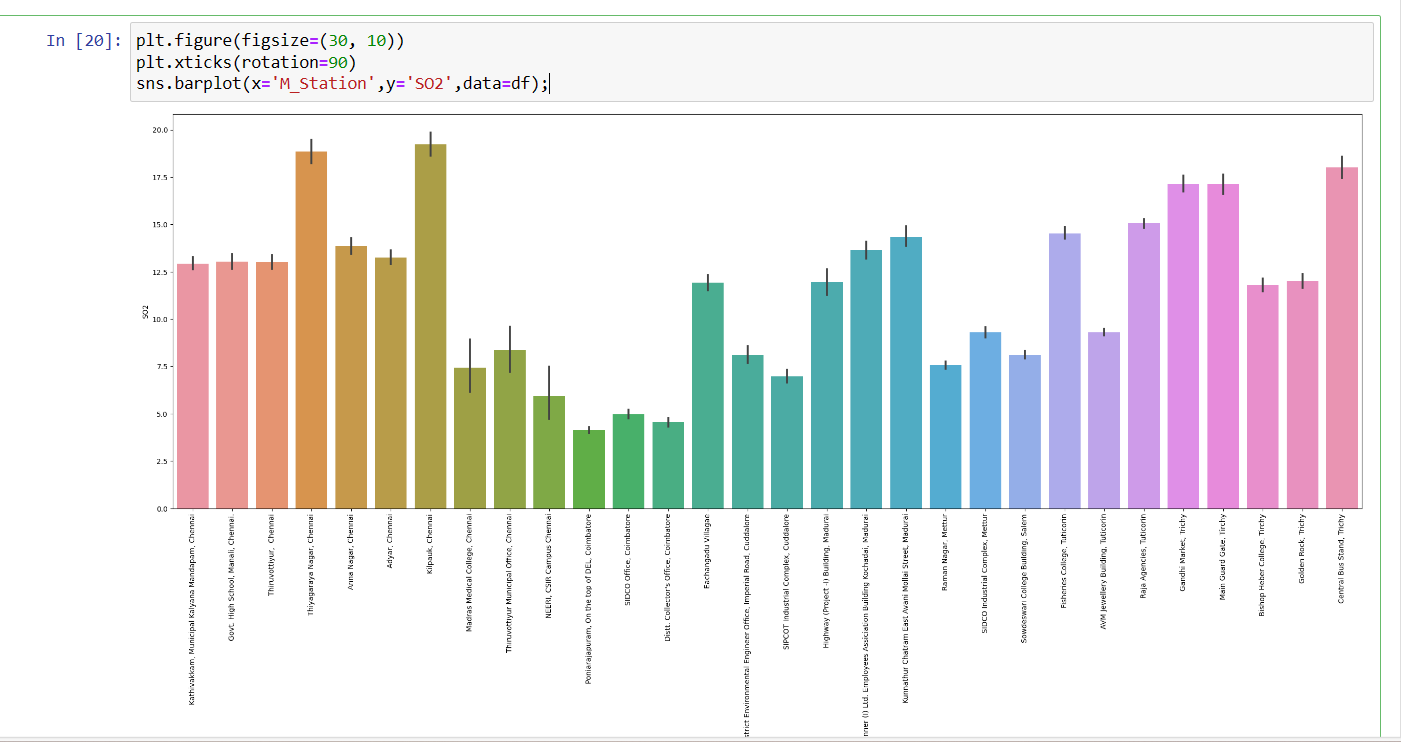


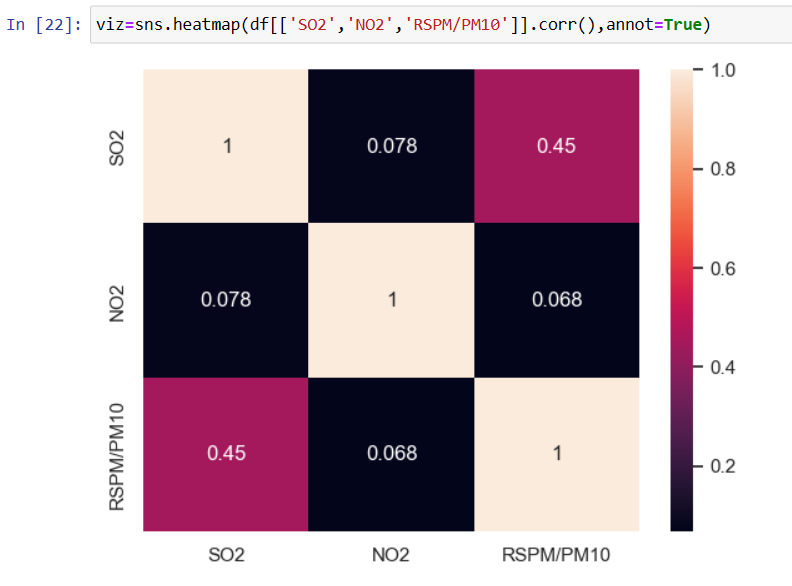
HANDLING THE MISSING VALUES

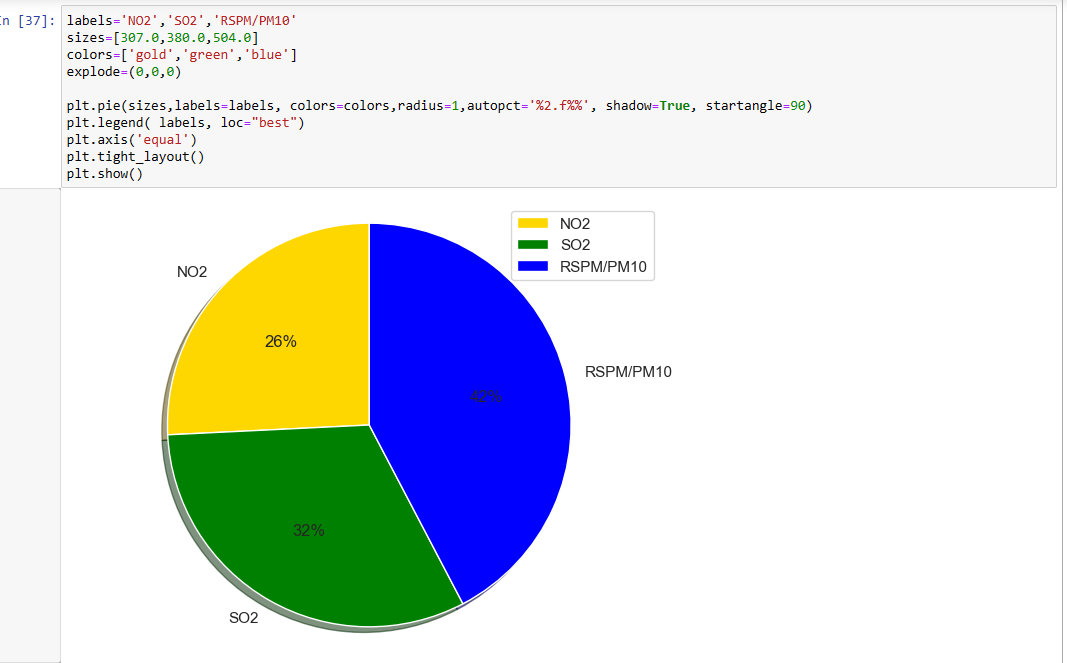


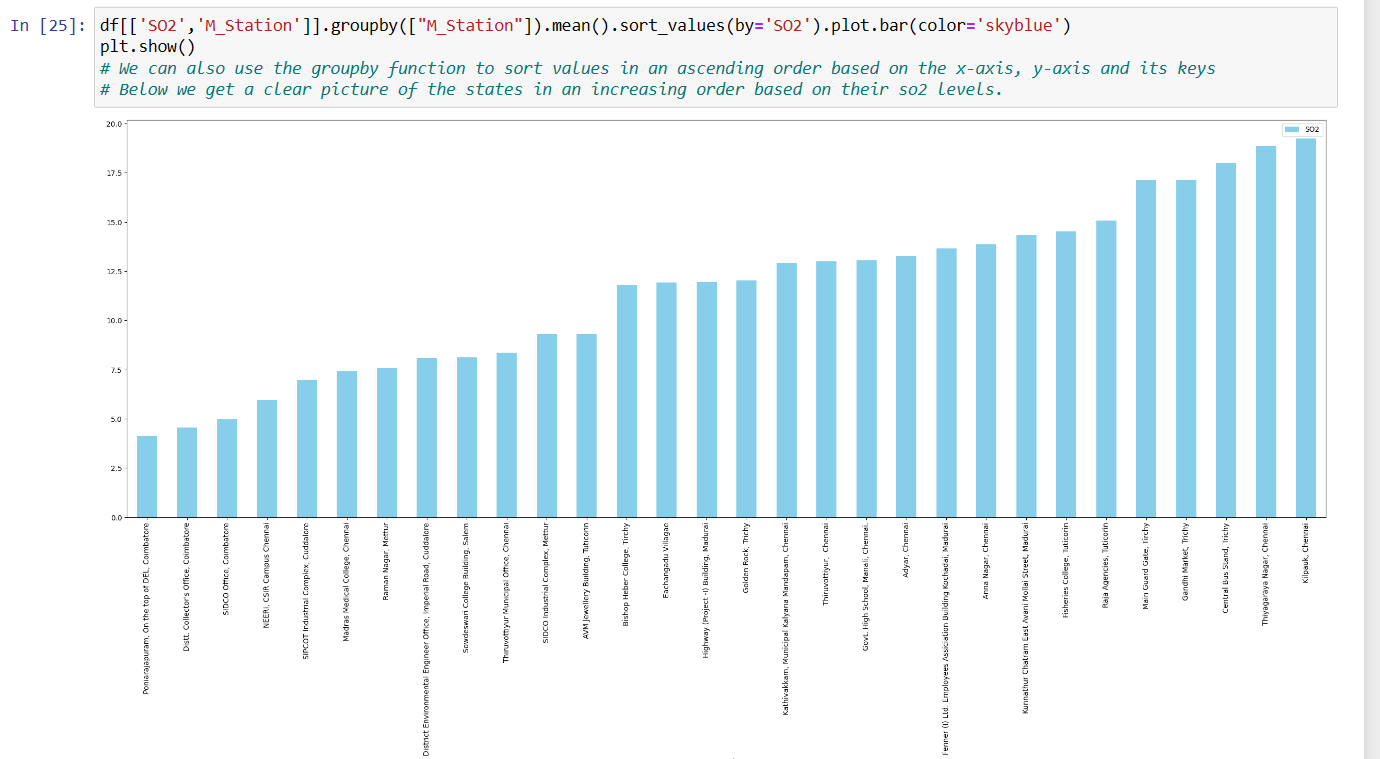
**DATA VISUALIZATION**

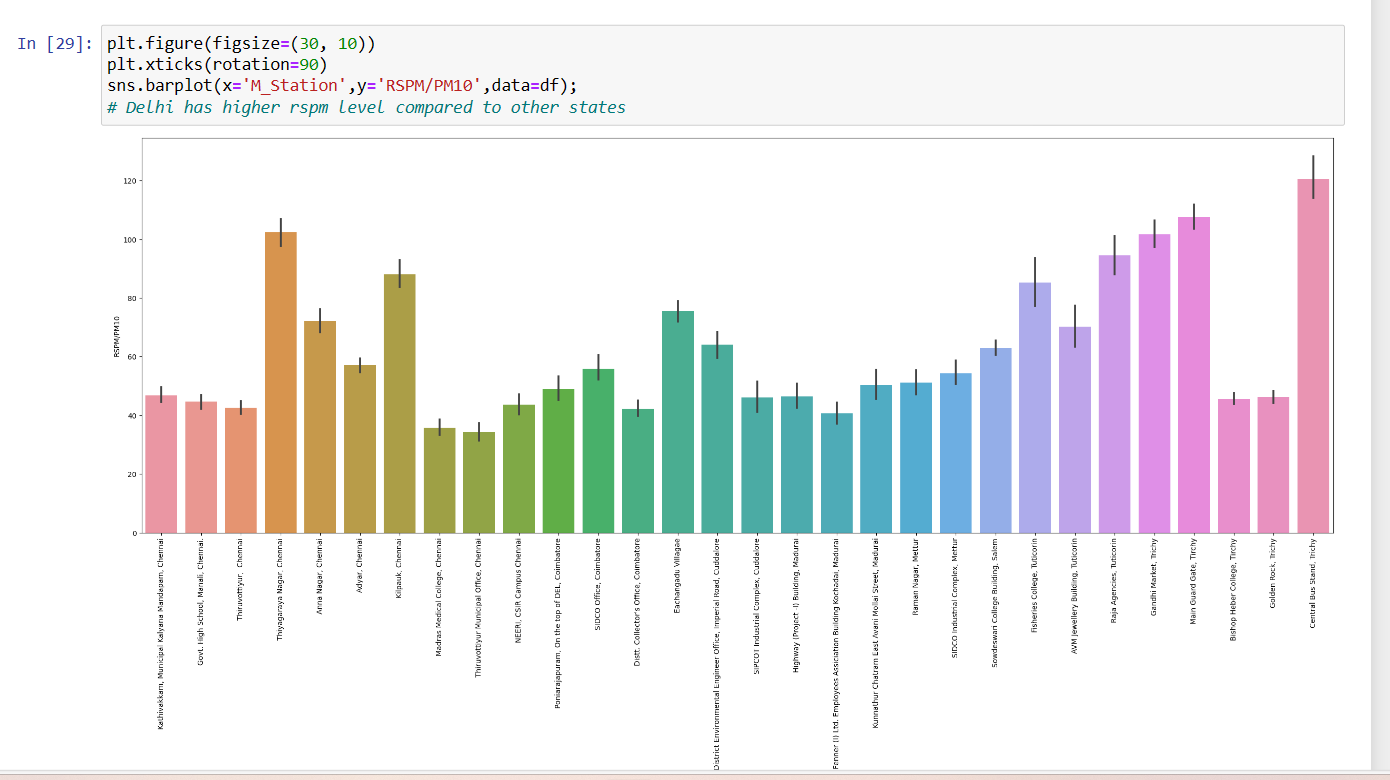
Data visualization is the practice of designing and creating easy-to-communicate and easy-to-understand graphic or visual representations of a large amount of complex quantitative and qualitative data and information with the help of static, dynamic or interactive visual items.











**PREDICTIVE ACCURACY**

Predictive accuracy, in the context of data analysis and machine learning, refers to how well a predictive model performs in making accurate predictions on new, unseen data. It is a crucial metric for evaluating the effective of a model in solving a particular problem. Several metrics can be used to assess predictive accuracy, depending on the type of problem.

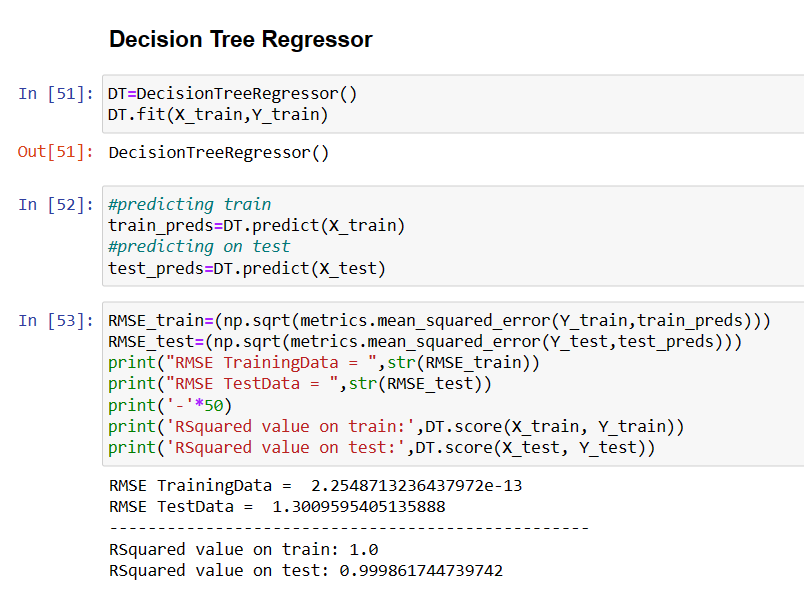
Before that we have to split the data into dependent and independent variables to achieve the target.

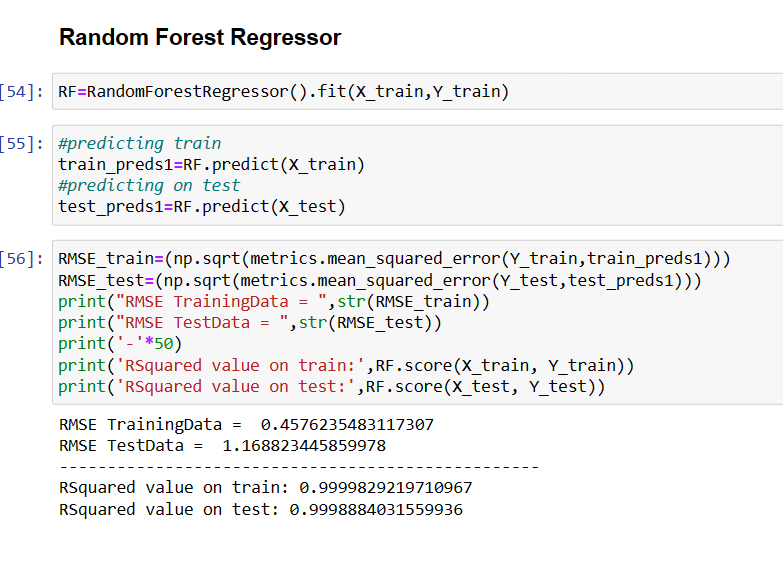


**DECISION TREE AND RANDOM FOREST REGRESSOR**:

Decision tree regression is a machine learning algorithm used for solving regression problems, where the goal is to predict a continuous numeric target variable based on input features.

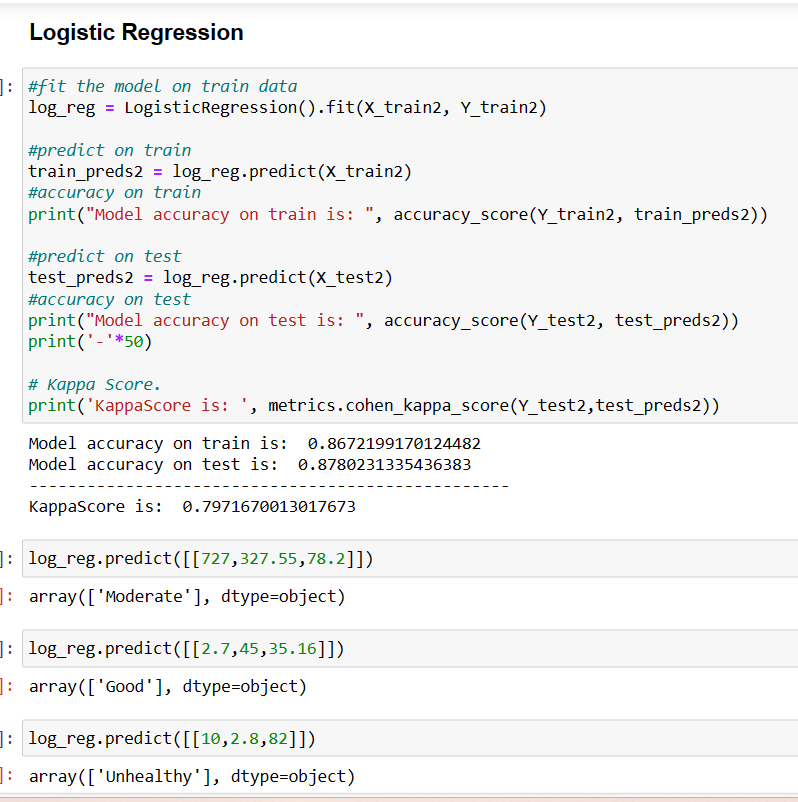
Random Forest Regression, often referred to as Random Forest for regression, is an ensemble machine learning method used for regression tasks. It is an extension of decision tree regression.





**CLASSIFICATION ALGORITHMS**

A classification algorithm is a type of machine learning algorithm used in supervised learning tasks. Its primary purpose is to categorize or classify input data into predefined categories or classes. In classification tasks, the algorithm takes a set of features or attributes associated with each data point and assigns it to one of several possible classes or categories based on patterns it has learned from the training data.





**CONCLUSION**

In light of the ongoing concern about air quality due to industrialization, urbanization, and climate change, air quality analysis remains an essential field that not only informs public policy but also enables individuals to take action to protect their health and the environment. Continued advancements in data analytics and technology will further enhance our ability to monitor and improve air quality in the future.