|  |  |  |
| --- | --- | --- |
| NicePng_national-emblem-of-india_9441929.png | **AIR QUALITY ANALYSIS IN TAMILNADU** |  |

**NAAN MUDHALVAN PROJECT REPORT**

Submitted by

**M.ABIRAMI (730321243001)**

**M.KALAIYARASI (730321243301)**

**M.YUKESHWARI (730321243301)**

**T.SIVARANJANI (730321243024)**

**FIFTH SEMESTER**

**ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**

**BUILDERS ENGINEERING COLLEGE , KANGEYAM**

**ANNA UNIVERSITY :: CHENNAI 600 025**

**NOVEMBER 2023**



**BONAFIDE CERTIFICATE**

Certified that this is a Bonafide record of work done by **ABIRAMI.M(730321243001) , KALAIYARASI.M(730321243009)** and **YUKESHWARI.M(730321243301),SIVARANJANI.T(730321243024)** in **NAAN MUTHALVAN** during the Academic year **2023- 24** for fifth Semester.

|  |  |
| --- | --- |
|  |  |
| **STAFF -INCHARGE** | **HEAD OF THE DEPARTMENT** |
|  |  |
| Submitted for the Naan mudhalvan | viva voce held on\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  |  |
| **INTERNAL EXAMINER** | **EXTERNAL EXAMINER** |

**ABSTRACT**

Predicting air quality is a complex task due to the dynamic nature, volatility and high variability in the time and space of pollutants and particulates. Due to the presence of governing factors, varying land uses, and many sources for the elaboration of pollution, the forecast and analysis of air pollution is a difficult procedure. At the same time, being able to model, predict, and monitor air quality is becoming more and more relevant, especially in urban areas, due to the observed critical impact of air pollution on citizens’ health and the environment. In this paper, various air pollution monitoring and prediction models with respect to hardware interfacing modules and various classification approaches. The Air Quality Index (AQI) parameter is used in this paper to monitor the quality of air pollution in various regions of the world. The drawbacks of the conventional air pollution monitoring and prediction models have been stated in this paper with the methodologies used for air pollution prediction. Particulate matter is a major air contaminant in the ambient air and wide information is assessable on its concentrations from various parts of the world. Air pollutants with the strongest evidence for public health concern, include particulate matter (PM), ozone (O3), nitrogen dioxide (NO2) and sulphur dioxide (SO2). Air pollutants are reported to be harmful to human health if their concentrations exceed certain acceptable levels. Industries have contributed substantially to the air pollution problem as point source of emissions.

|  |
| --- |
| **TABLE OF CONTENT** |
| **CHAPTER TITLE PAGE NO.** |
| **NO.** |
| 1. **PROBLEM STATEMENT 5** |
| **2 DESIGN THINKING 6** |
| **3 DATASET DEFINITION 7** |
| **4 DATA PREPROCESSING AND 8**  **FEATURE EXTRACTION** |
| **5 PROPOSED ALGORITHM 13**    **6 PROPOSED INNOVATION 17**  **TECHNIQUE**  **7 CONCLUSION AND 19**  **FUTURE SCOPE** |
|  |

**CHAPTER 1**

**PROBLEM STATEMENT**

Air quality is linked to human activities, earth’s climate and ecosystems globally. Many of the drivers of air pollution (i.e. combustion of fossil fuels) are also sources of CO2 emissions and other short-lived climate pollutants, such as ozone and black carbon that greatly contribute to climate change, and affect human health.Air quality is closely related to the earth's climate and [ecosystems](https://en.wikipedia.org/wiki/Ecosystem) globally. Many of the contributors of air pollution are also sources of [greenhouse emission](https://en.wikipedia.org/wiki/Greenhouse_gas_emissions) i.e., burning of [fossil fuel](https://en.wikipedia.org/wiki/Fossil_fuel). Air pollution is a significant risk factor for a number of [pollution-related diseases](https://en.wikipedia.org/wiki/List_of_pollution-related_diseases), including [respiratory infections](https://en.wikipedia.org/wiki/Respiratory_infection), [heart disease](https://en.wikipedia.org/wiki/Heart_disease), [chronic obstructive pulmonary disease](https://en.wikipedia.org/wiki/Chronic_obstructive_pulmonary_disease) (COPD), [stroke](https://en.wikipedia.org/wiki/Stroke), and [lung cancer](https://en.wikipedia.org/wiki/Lung_cancer). Growing evidence suggests that air pollution exposure may be associated with reduced IQ scores, impaired cognition, increased risk for psychiatric disorders such as [depression](https://en.wikipedia.org/wiki/Depression_(mood)) and detrimental [perinatal](https://en.wikipedia.org/wiki/Perinatal) health. The human health effects of poor air quality are far reaching, but principally affect the body's respiratory system and the [cardiovascular system](https://en.wikipedia.org/wiki/Circulatory_system). Individual reactions to air pollutants depend on the type of pollutant a person is exposed to the degree of exposure, and the individual's health status and [genetics](https://en.wikipedia.org/wiki/Genetics).

**CHAPTER 2**

**DESIGN THINKING**

**ANALYSIS OBJECTIVES :**

* With the rapid development of economy and the frequent occurrence of air pollution incidents, the problem of air pollution has become a hot issue of concern to the whole people.
* The air quality big data is generally characterized by multi-source heterogeneity, dynamic mutability, and spatial–temporal correlation, which usually uses big data technology for air quality analysis after data fusion Empathize.
* Air quality analysis is a crucial aspect of environmental monitoring. It involves measuring the concentration of pollutants in the air and assessing their impact on human health and the environment.
* Design thinking is a problem-solving approach that emphasizes empathy, creativity, and experimentation. Applying design thinking to air quality analysis can help identify new solutions to environmental challenges.
* [One example of applying design thinking to air quality analysis is developing an Internet of Things (IoT) enabled air quality monitoring system that can measure local area air contamination and generate analyzed data based on which it alerts the people through a buzzer device integrated into the system](https://ieeexplore.ieee.org/document/9087064).
* This system is mobile in nature and can be installed in houses and small places. [Another example is building air quality models that predict the impact of pollutants released from various sources such as power plants and roadways](https://www.epa.gov/air-research/air-quality-modeling). These models are used by EPA, states, tribes, and local agencies to assess control strategies, regulate emissions, and evaluate mitigation options.
* [Design thinking can also be applied to IoT sensor-based indoor air quality monitoring systems](https://www.mdpi.com/2076-3417/12/19/9450). Researchers and building managers can use commercially available hardware and software to implement their own environmental monitoring study.

**CHAPTER 3**

**DATASET DEFINITION**

**SURVEY DATASET:**

* Survey datasets are collections of data that are gathered through the administration of surveys or questionnaires. These datasets are commonly used in various fields, including social sciences, market research, healthcare, and more, to gather information about a particular topic, population, or research question. Here are some key points about survey datasets:

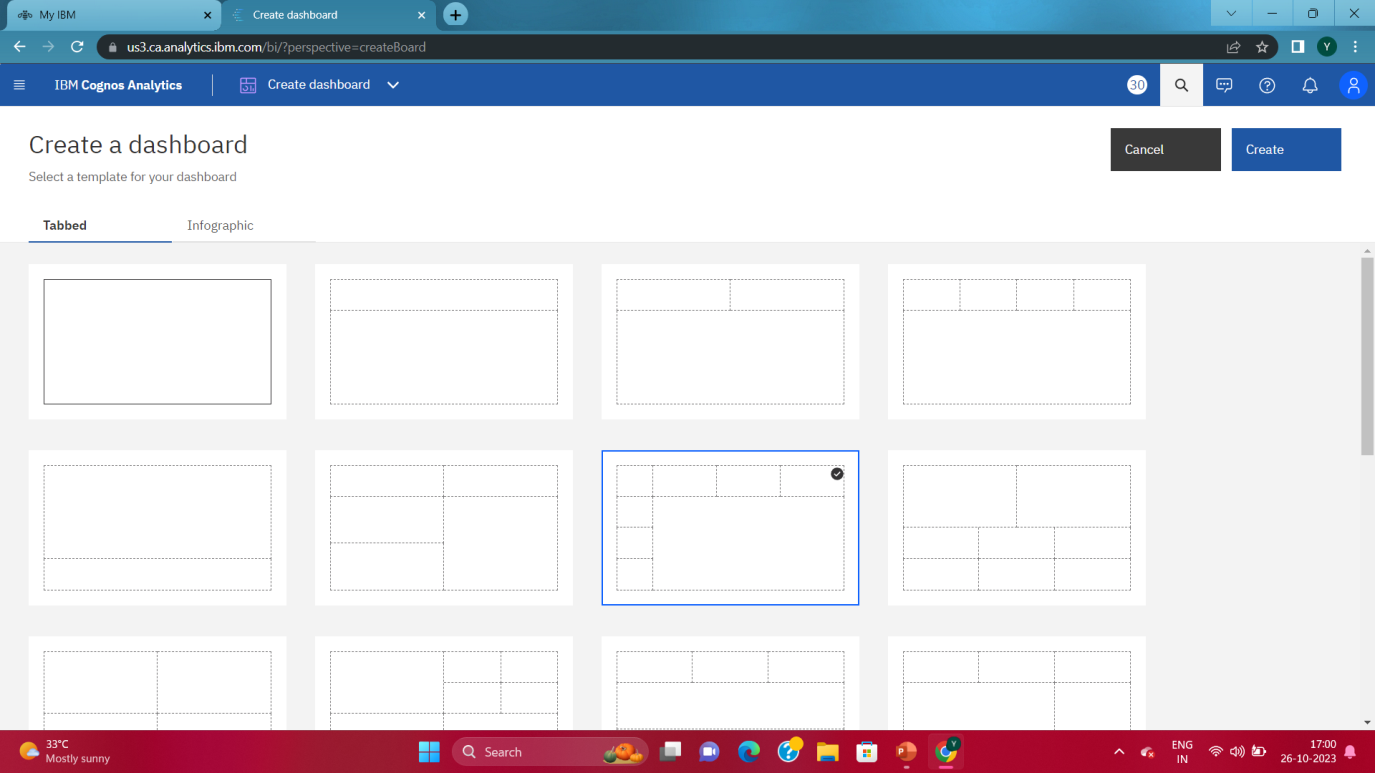
1. **Define the problem**: The first step is to define the problem you want to solve. This could be anything from predicting air pollution levels to identifying the sources of pollution.
2. **Identify the data sources**: Once you have defined the problem, you need to identify the data sources that will help you solve it. This could include government databases, research papers, or sensor data.
3. **Collect the data**: After identifying the data sources, you need to collect the data. This could involve scraping websites, downloading datasets, or collecting sensor data.
4. **Clean and preprocess the data**: Once you have collected the data, you need to clean and preprocess it. This could involve removing missing values, normalizing the data, or transforming it into a different format.
5. **Analyze and visualize the data**: After cleaning and preprocessing the data, you can start analyzing and visualizing it. This could involve creating histograms, scatter plots, or heat maps.
6. **Build and evaluate models**: Once you have analyzed and visualized the data, you can start building models to predict air pollution levels or identify pollution sources. You can use machine learning algorithms such as regression or clustering to build these models.
7. **Deploy the models**: After building and evaluating the models, you can deploy them in production environments. This could involve integrating them into web applications or mobile apps.

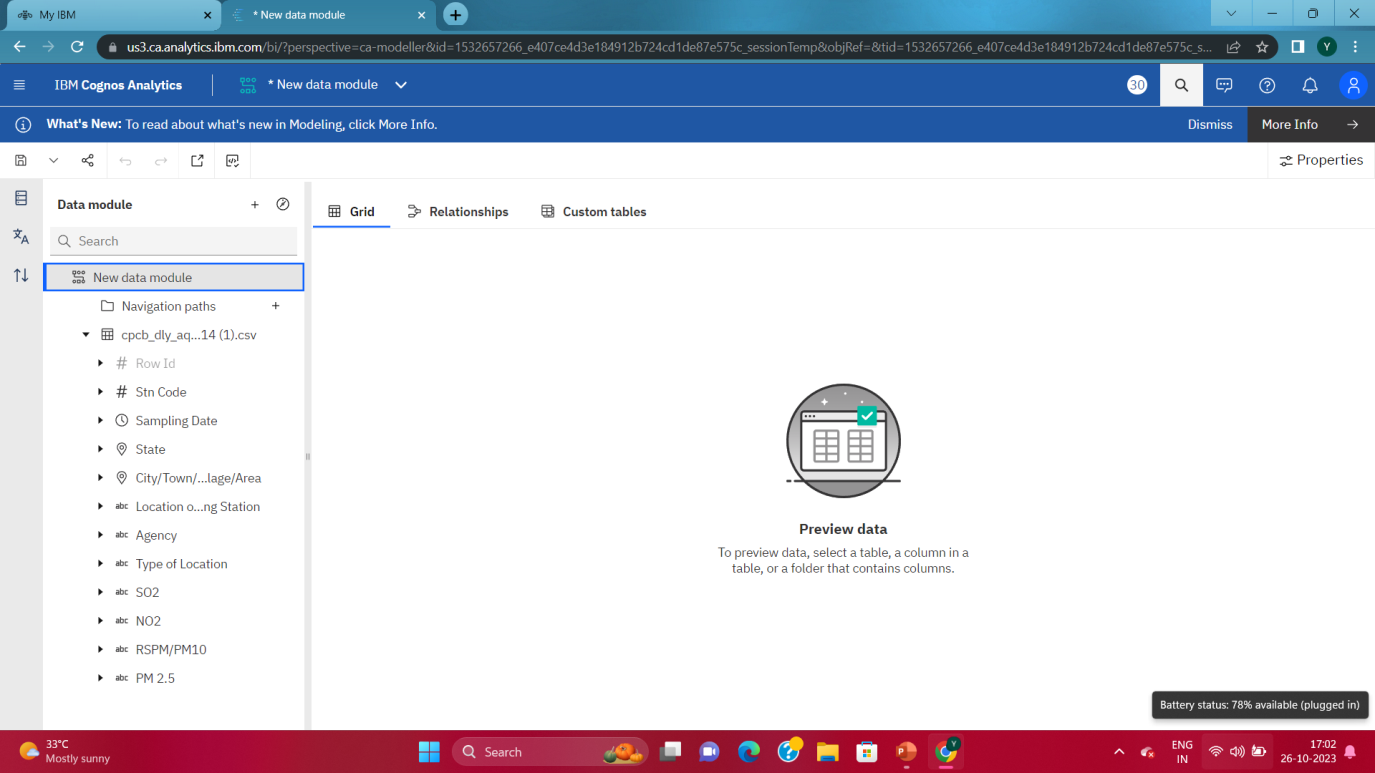
**CHAPTER 4**

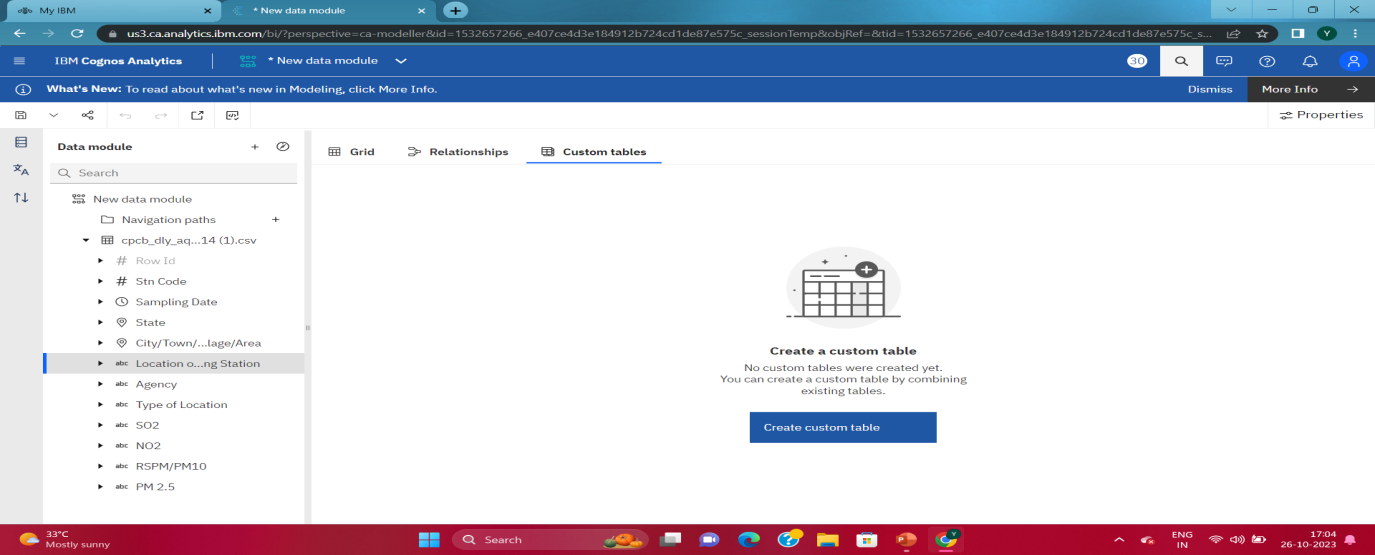
**DATA PREPROCESSING AND FEATURE EXTRACTION**

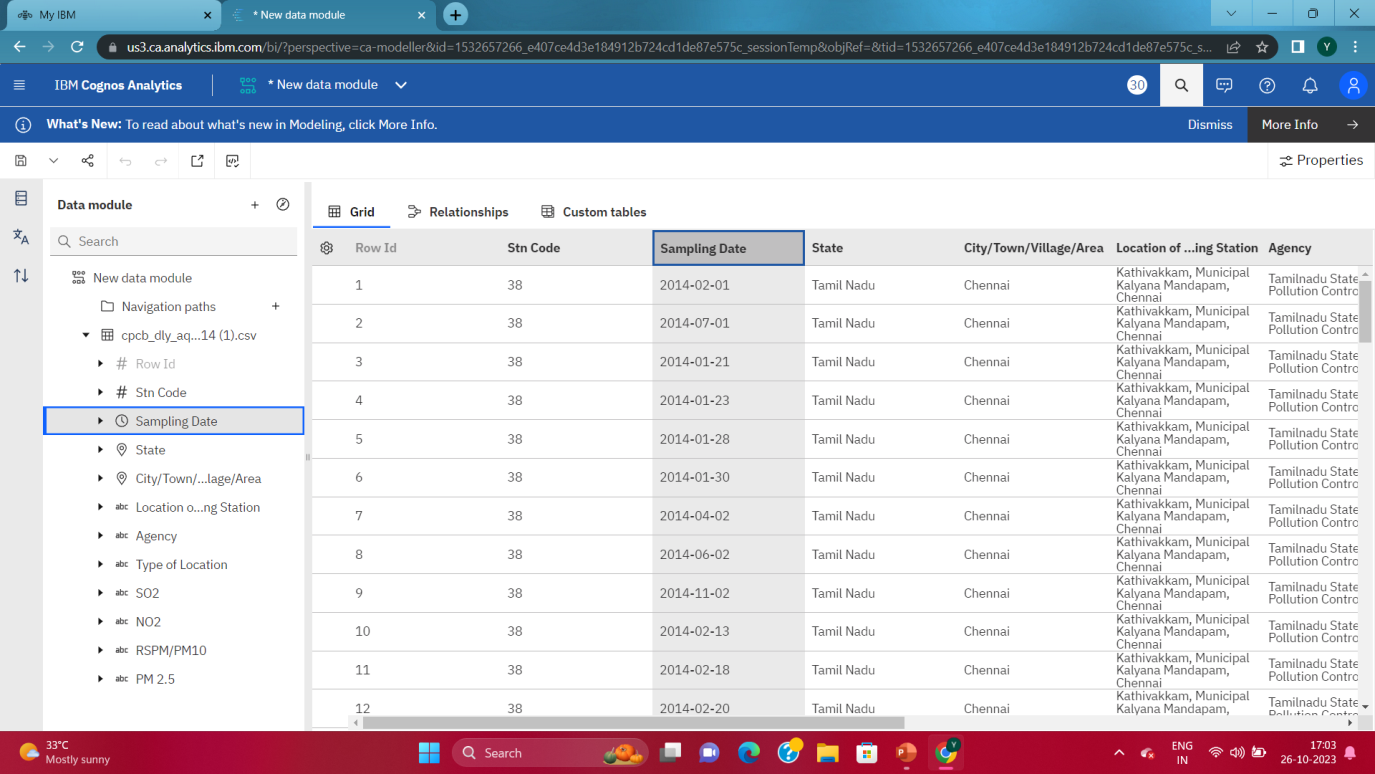
**DATA PREPROCESSING**

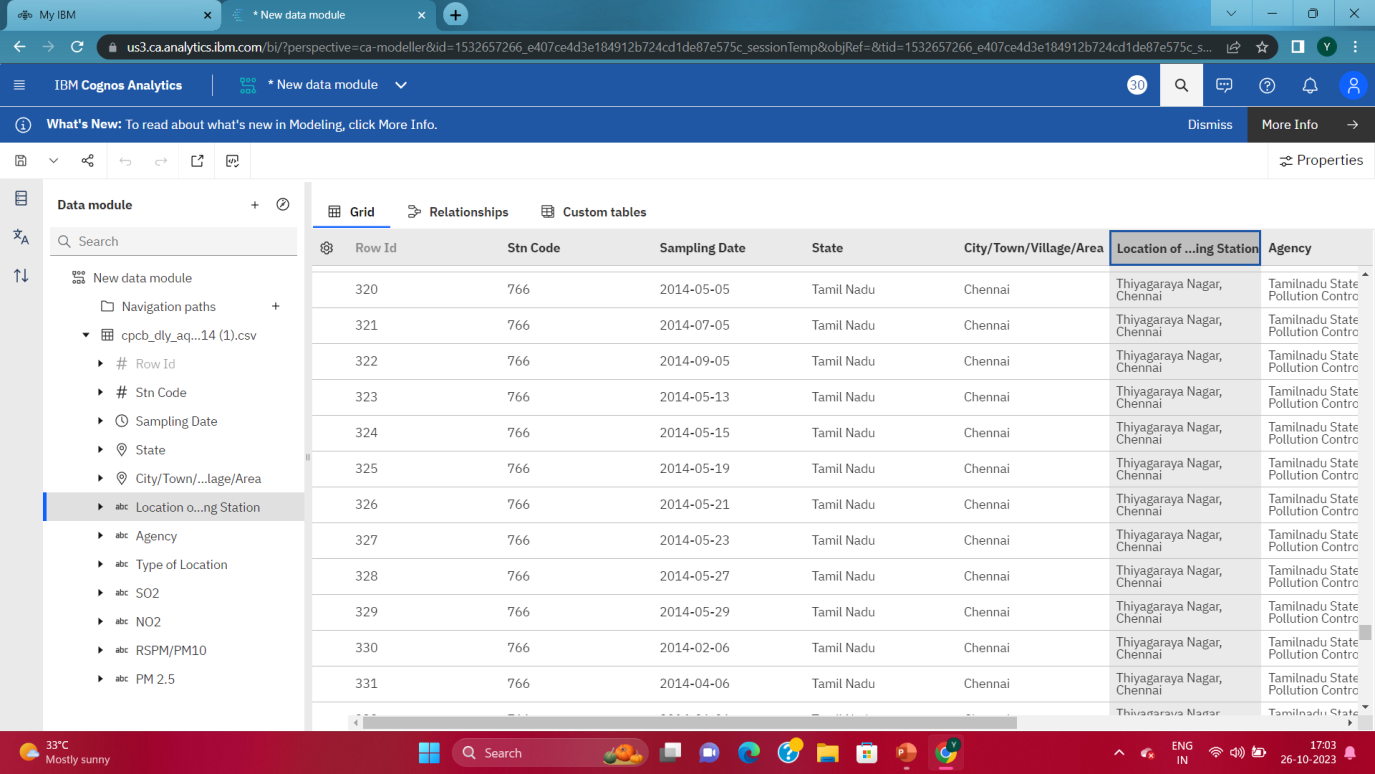
1. Data preprocessing is a crucial step in working with survey datasets. It involves cleaning and transforming the data to ensure it is accurate, complete, and suitable for analysis. Here are some common data preprocessing tasks specific to survey datasets:**Data cleaning**: The first step is to clean the data by removing any missing values, duplicates, or outliers. This ensures that the data is accurate and reliable.
2. **Data normalization**: The next step is to normalize the data to ensure that all features are on the same scale. This is important because some features may have a larger range than others, which can affect the performance of machine learning algorithms.
3. **Feature selection**: Once the data has been normalized, the next step is to select the most relevant features for analysis. This can be done using techniques such as correlation analysis or principal component analysis (PCA).
4. **Feature engineering**: Feature engineering involves creating new features from existing ones to improve the performance of machine learning algorithms. For example, we can create new features such as moving averages or time lags to capture temporal patterns in the data.
5. **Dimensionality reduction**: Dimensionality reduction techniques such as PCA can be used to reduce the number of features in the dataset while retaining most of the information.
6. **Data transformation**: Data transformation techniques such as log transformation or Box-Cox transformation can be used to transform non-normal data into a normal distribution.
7. **Data discretization**: Data discretization involves converting continuous variables into categorical variables. This can be useful when working with decision trees or other machine learning algorithms that require categorical input.

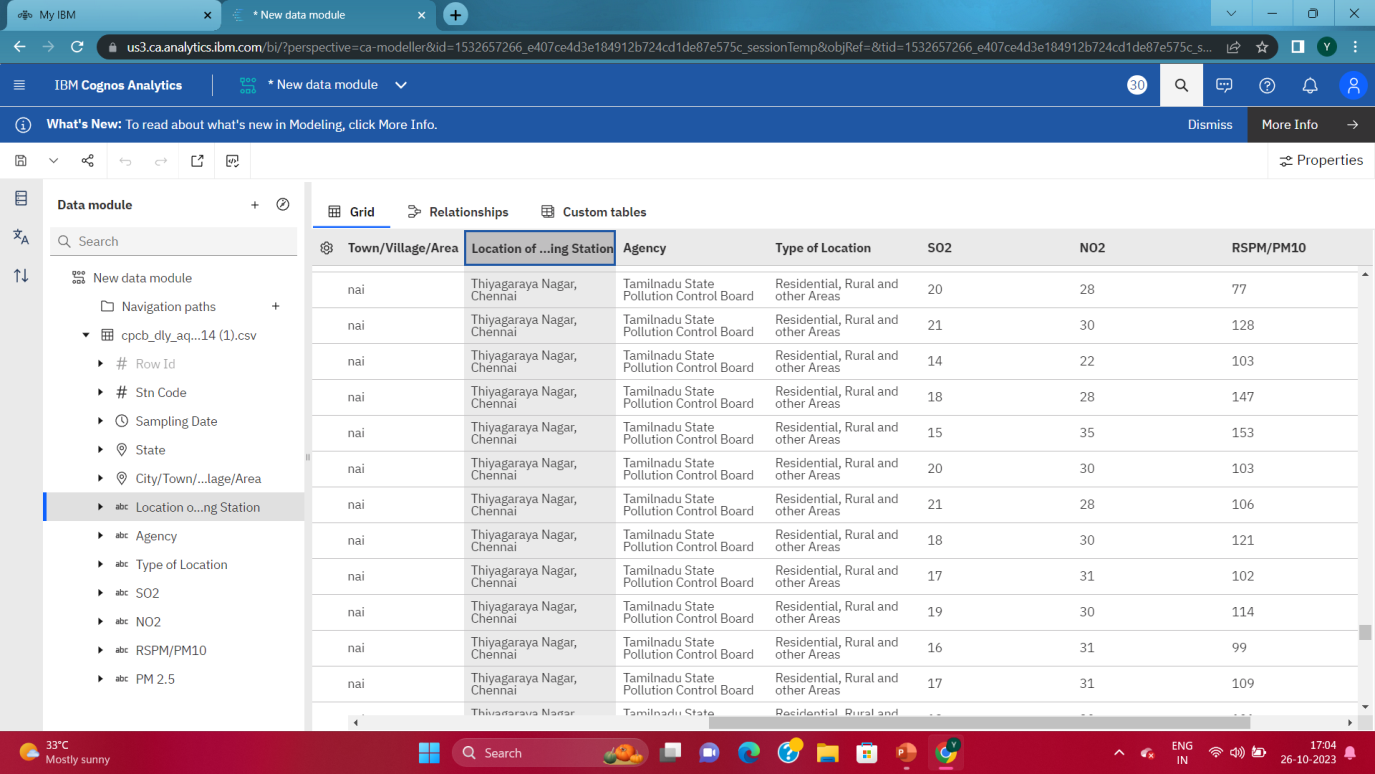


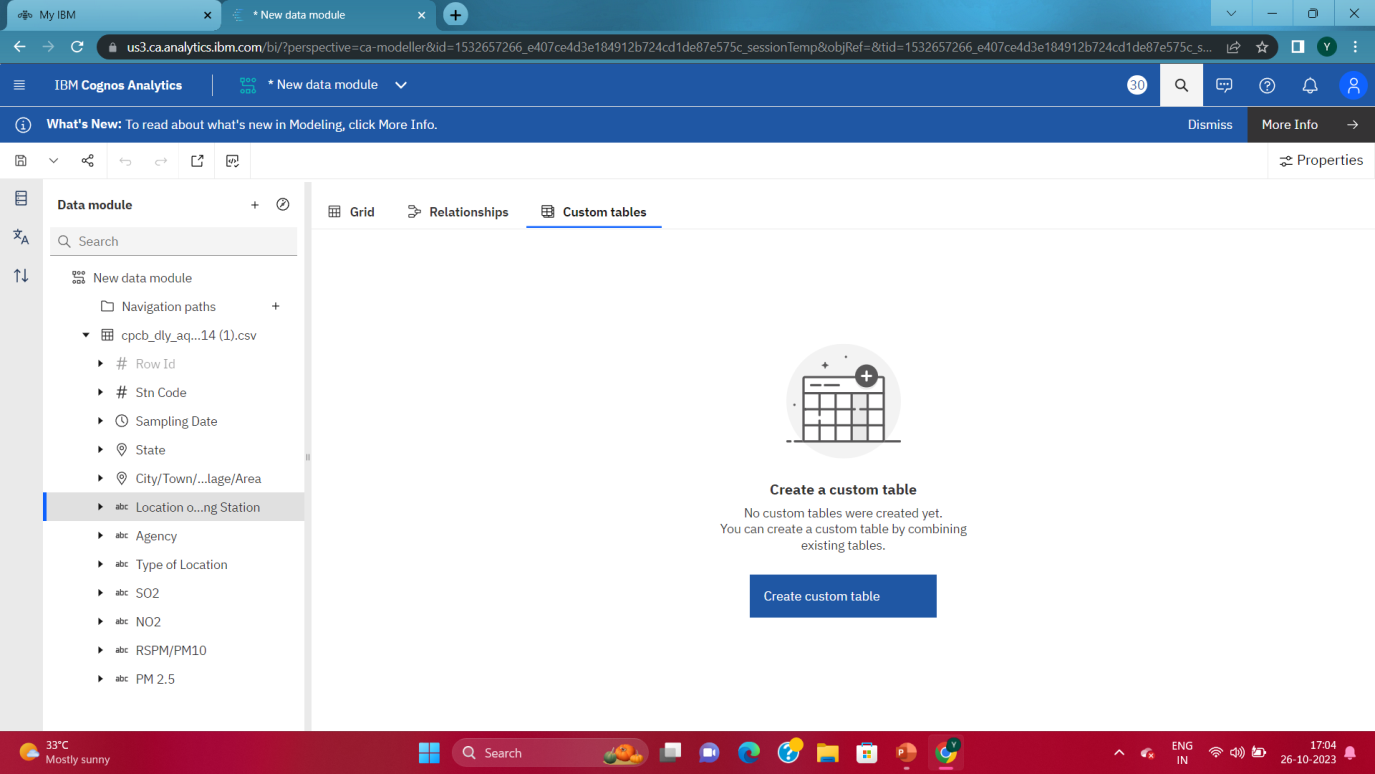
****

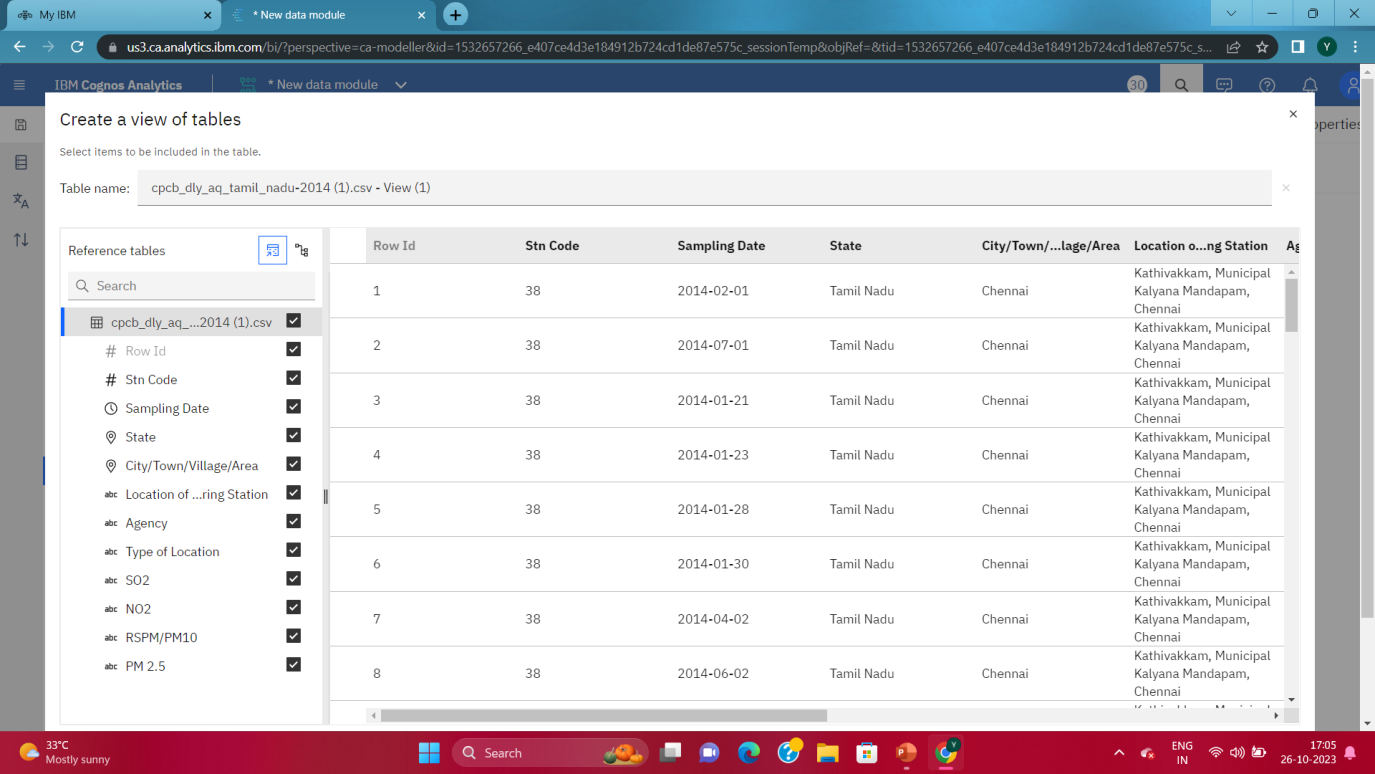
****

****

****

****

****

****

**CHAPTER 5**

**PROPOSED ALGORITHM**

**SUPPORT VECTOR MACHINES (SVM)**

Support Vector Machines (SVM) are a class of supervised machine learning algorithms used for classification and regression tasks. They are powerful tools for both linear and nonlinear data analysis. SVMs work by finding a hyperplane that best separates data points into different classes, while maximizing the margin between the two classes.

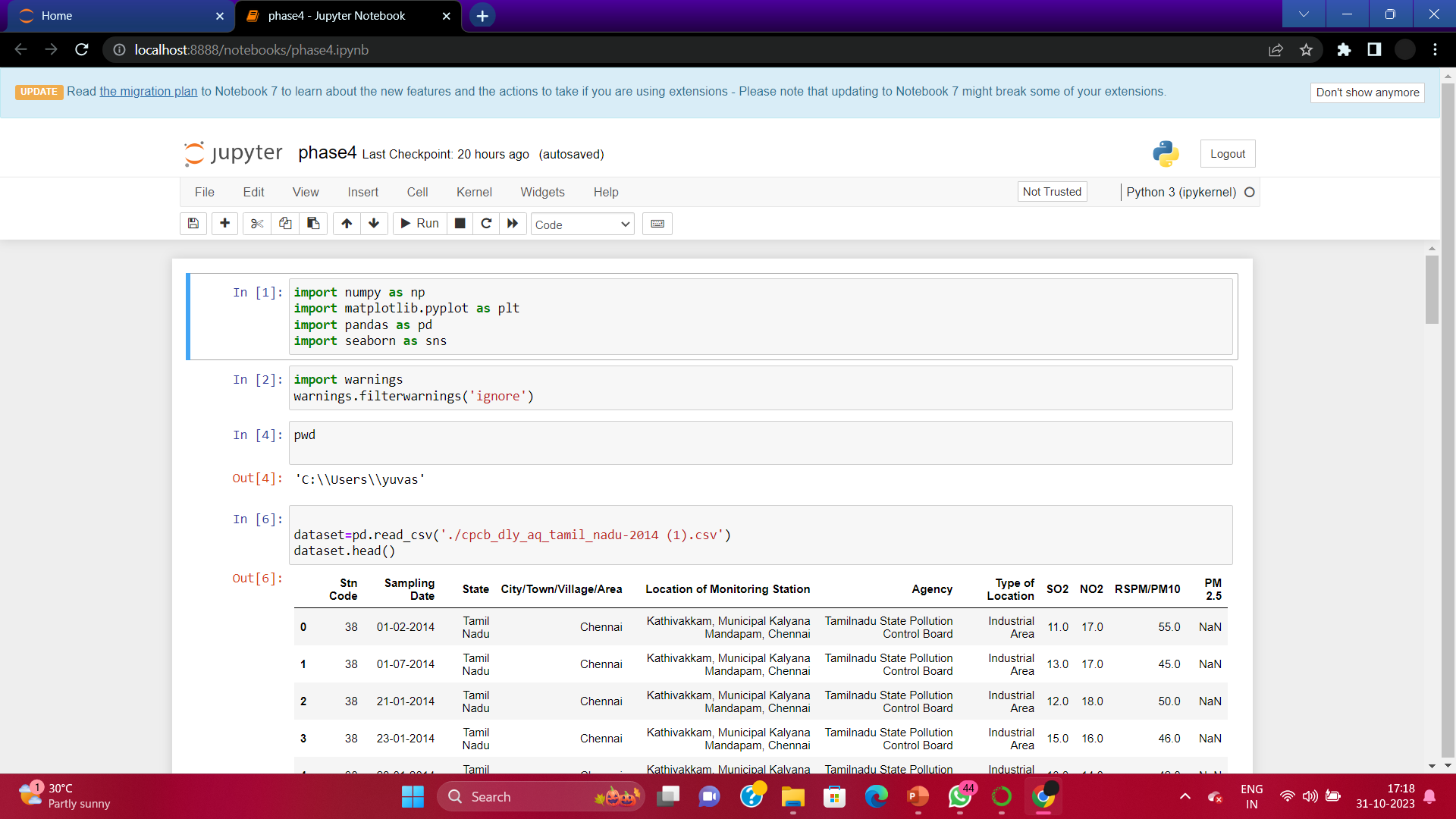
**1)APPLYING ML MODEL(REGULARIZATION)**

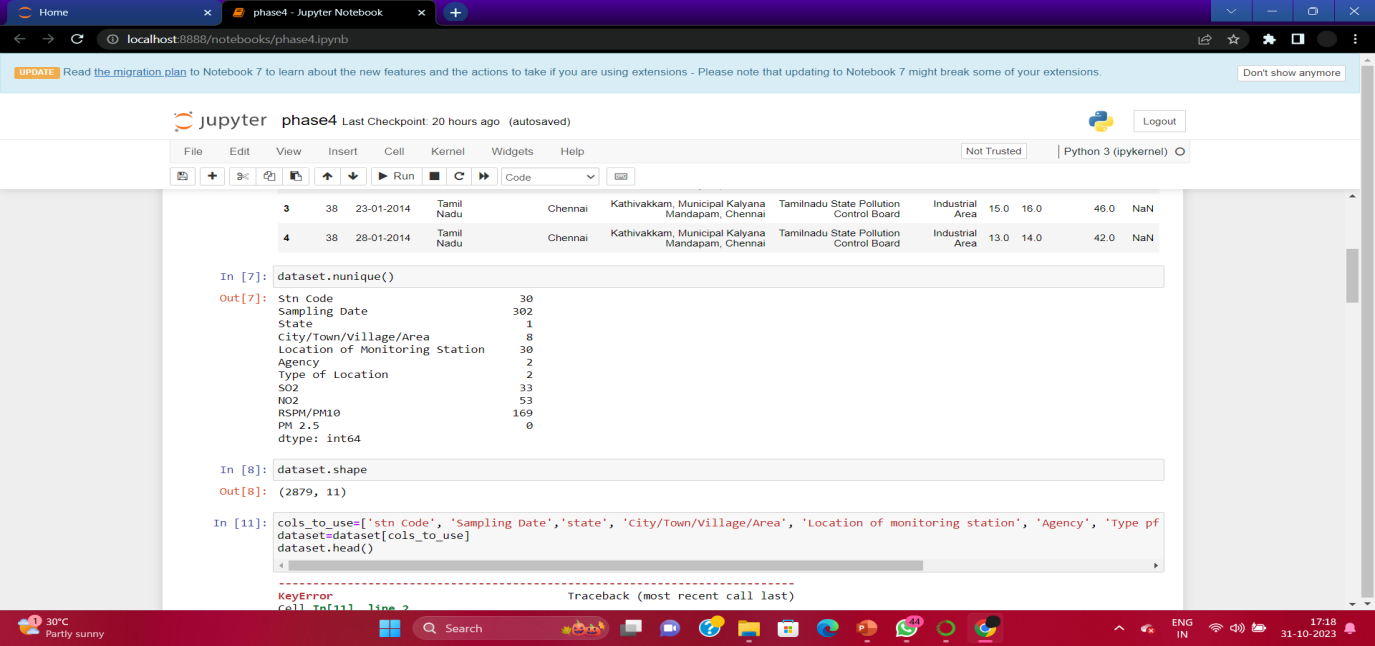
**2) SPLITTING THE TRAINING AND TESTING DATA**

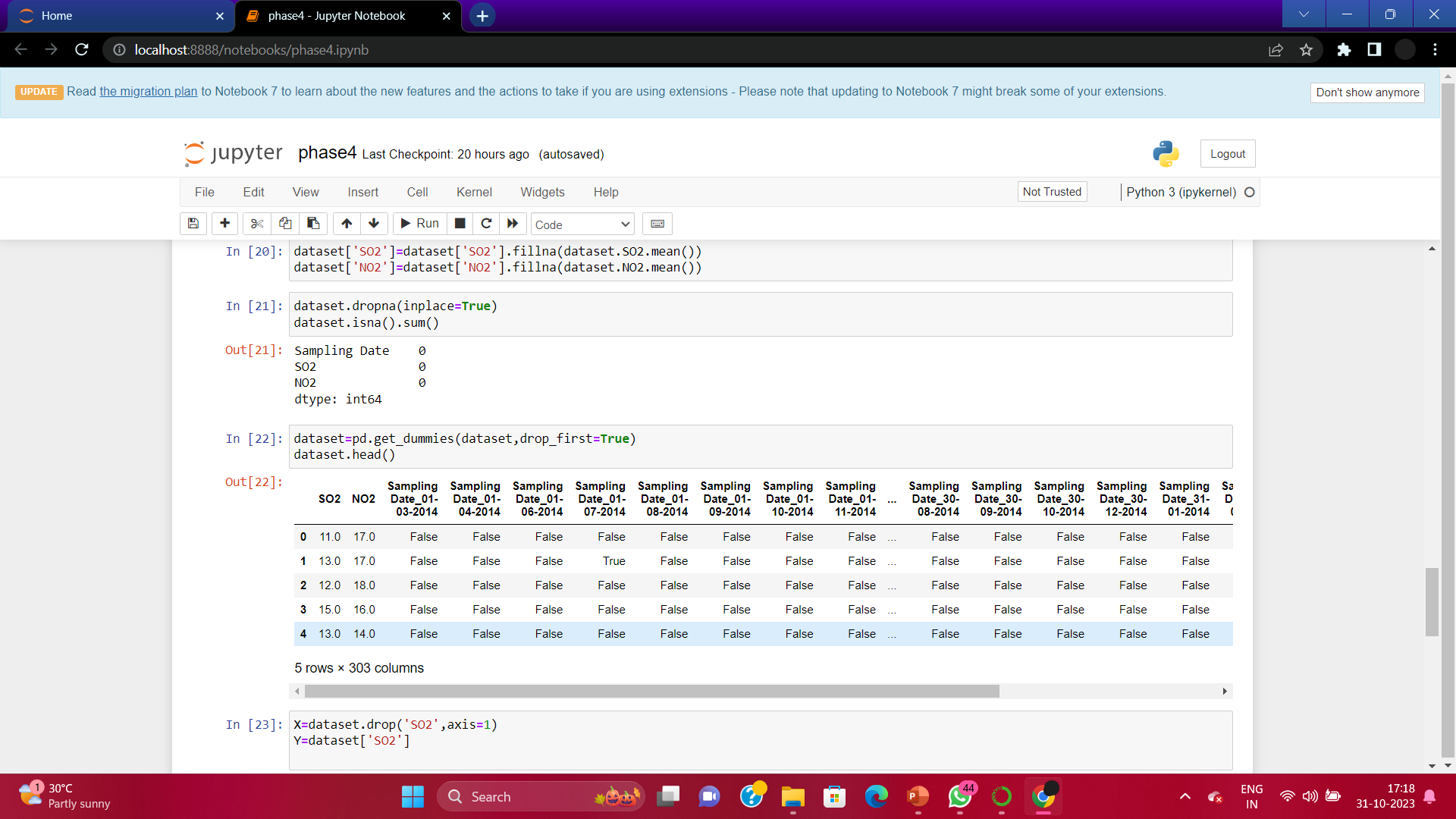
**3)CREATING DASHBOARD USING IBM COGNOS**

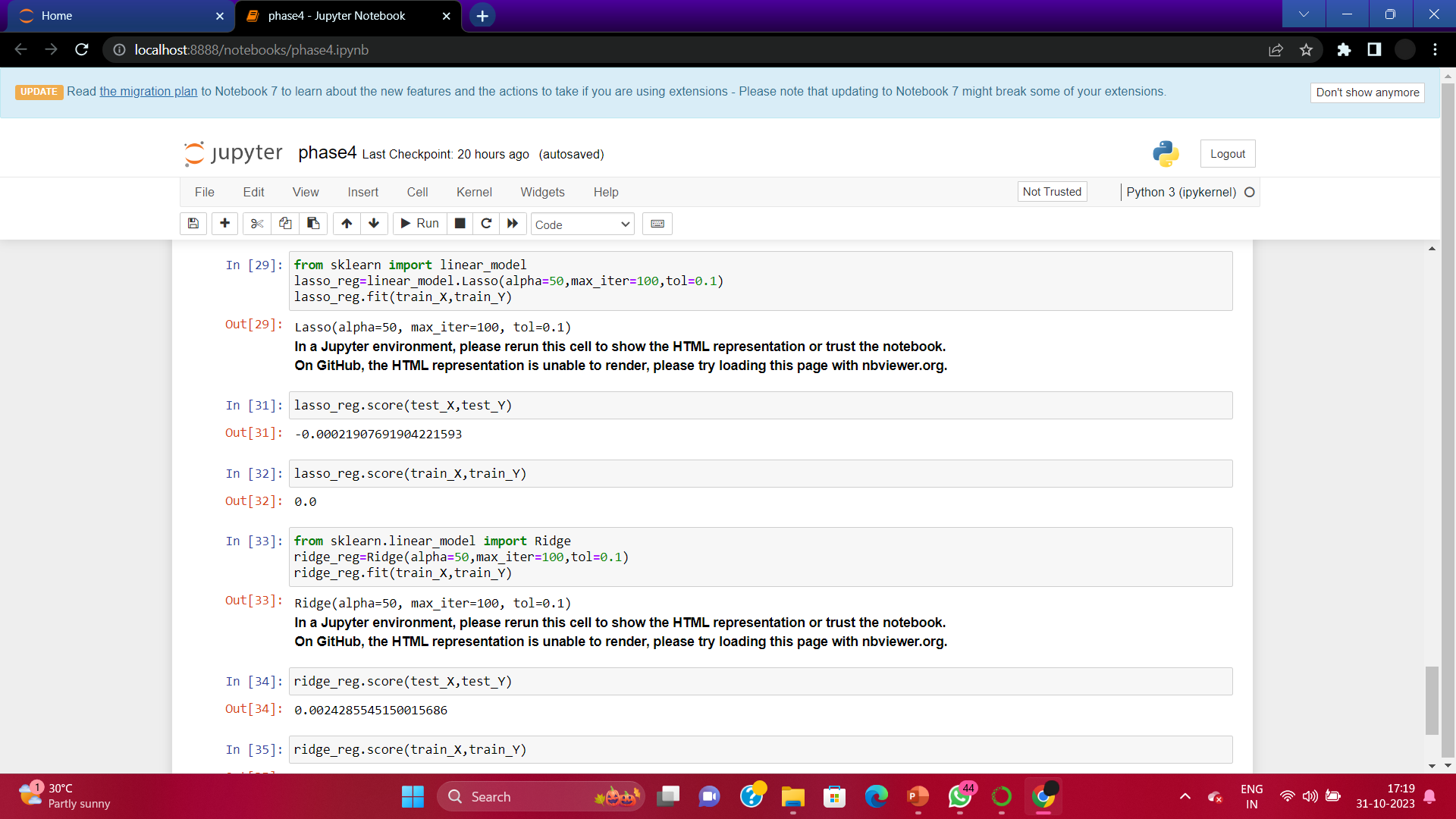
**4)CREATING REPORT USING IBM COGNOS**

**5)STATISTICAL TESTS USING TREEMAP**

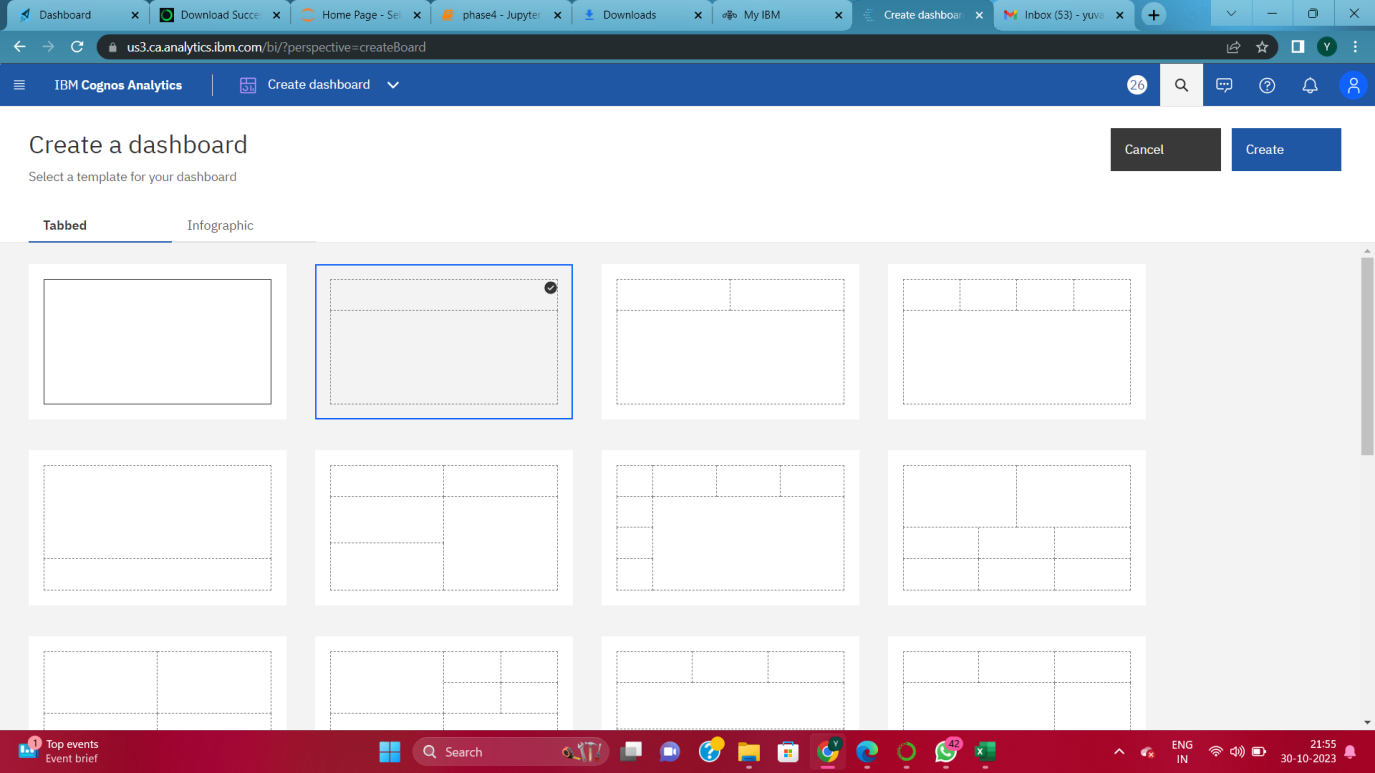
****

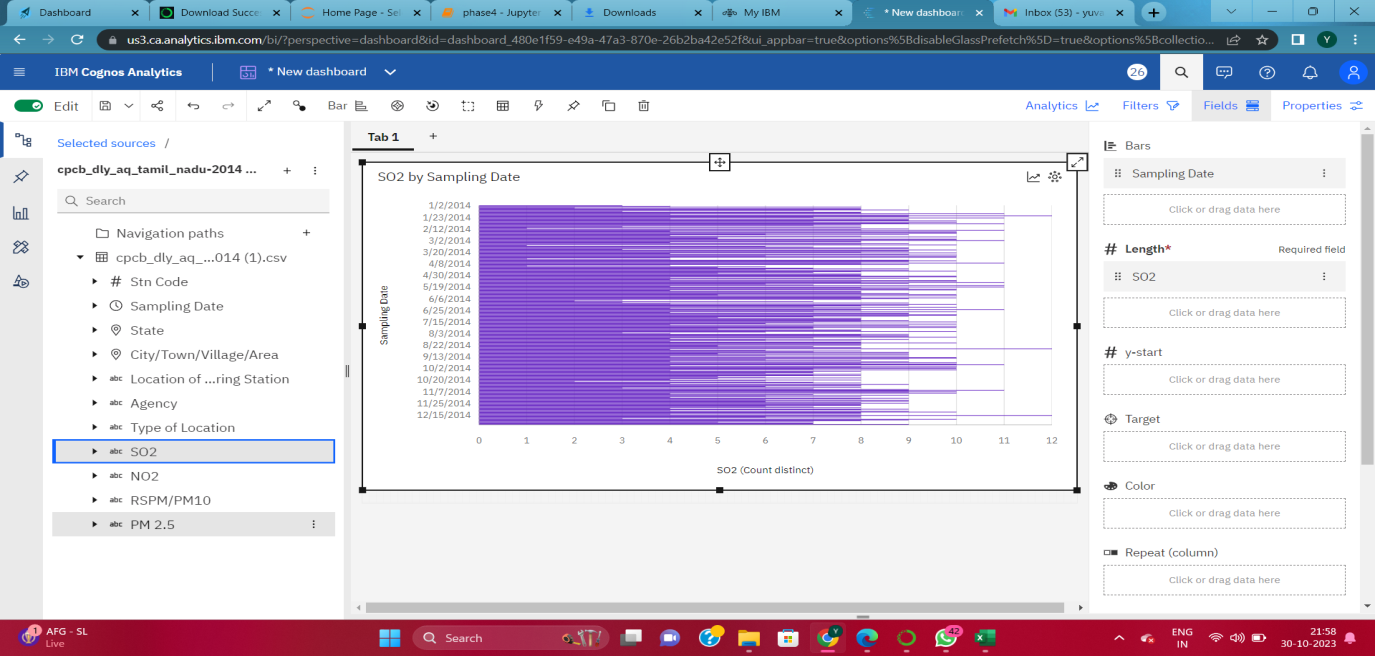
****

****

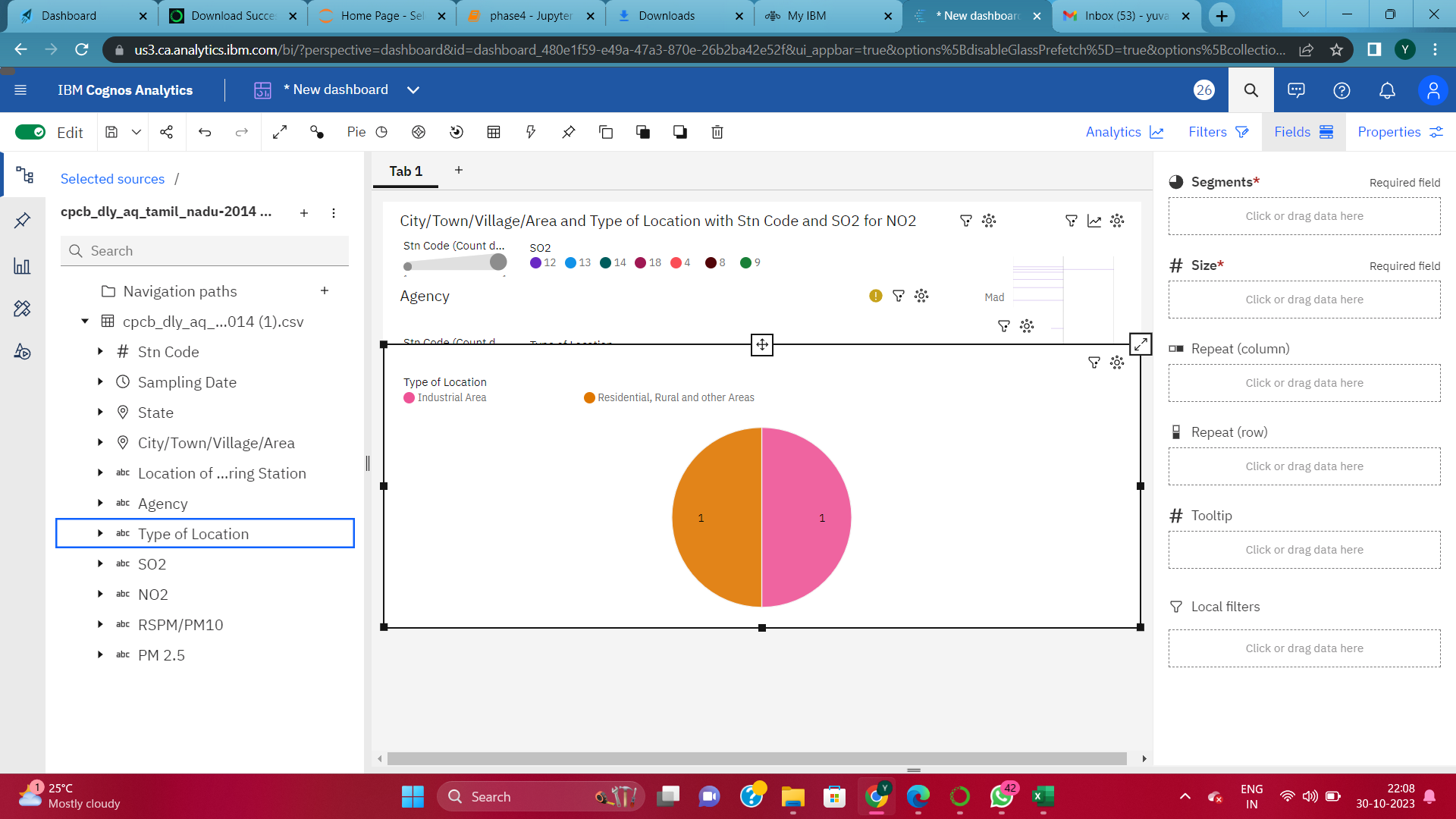
****

****

****

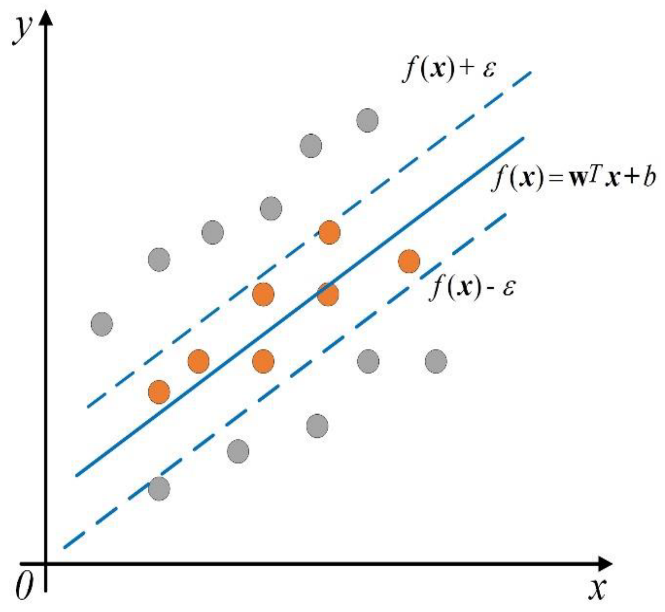
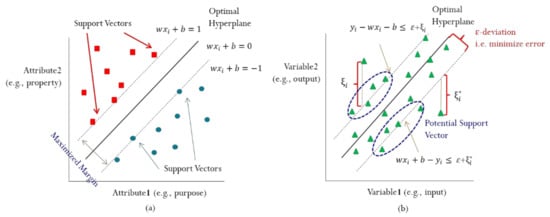
****

****

****

**CHAPTER 7**

**PROPOSED INNOVATION TECHNIQUE**

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well it’s best suited for classification. The main objective of the SVM algorithm is to find the optimal hyperplane in an N-dimensional space that can separate the data points in different classes in the feature space. The hyperplane tries that the margin between the closest points of different classes should be as maximum as possible. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

**CHAPTER 8**

**CONCLUSION AND FUTURE SCOPE**

**CONCLUSION:**

* Air pollution plays an enormous role in the environment, it exists since the ancient eras since people there were using materials that affect the quality of the air.
* As a result of these acts, there are many effects on every organism living on this earth, People’s health is most prone to pollution diseases.
* which can lead to death if the situation developed followed by the effects on the environment that can kill the crops the soil, and the water.
* Also, air pollution affects directly the atmosphere which directly results in global warming since they are dependent on each other.
* Luckily, there are some ways to reduce those effects. First is to use neat energy which does not affect the natural environment. And stop cutting down the trees since they are very helpful to produce fresh air and make it stays in its original composition.
* Finally, raising people’s awareness is also an important method and its importance is not more than others. We should do our best to reduce air pollution to make our life clean and healthy.

**FUTURE SCOPE:**

* Erosol, air quality, atmospheric chemistry and global change
* Air toxics (hazardous air pollutants (HAPs), persistent organic pollutants (POPs)) - Sources, control, transport and fate, human exposure
* Nanoparticle and nanotechnology
* Sources, combustion, thermal decomposition, emission, properties, behavior, formation, transport, deposition, measurement and analysis
* Effects on the environments
* Air quality and human health
* Bioaerosols
* Indoor air quality
* Energy and air pollution
* Pollution control technologies
* Invention and improvement of sampling instruments and technologies
* Optical/radiative properties and remote sensing