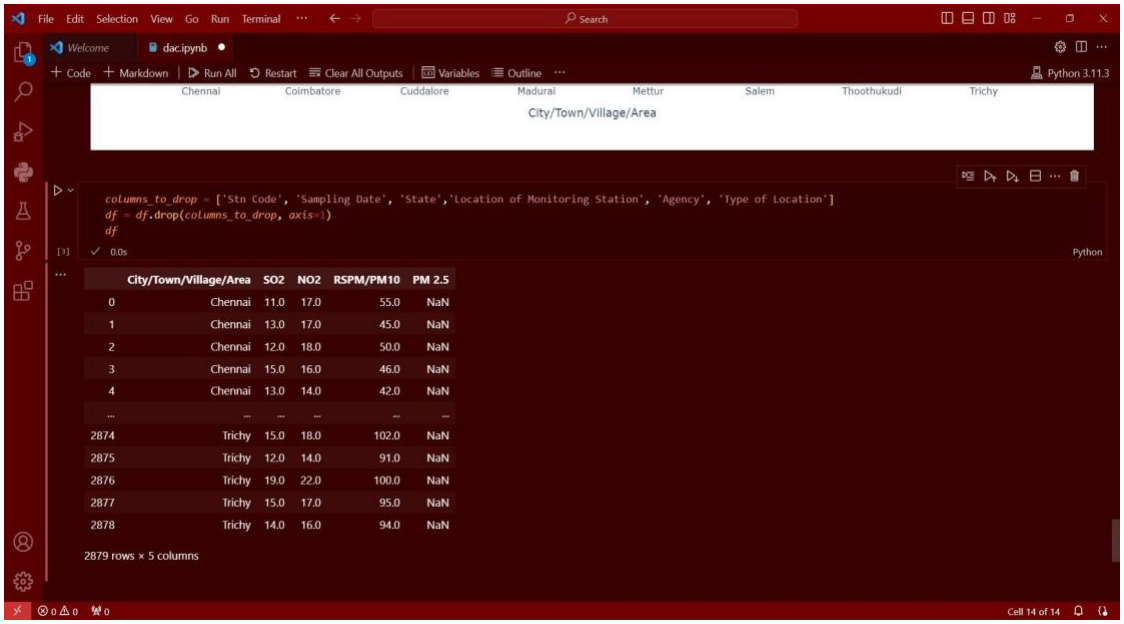
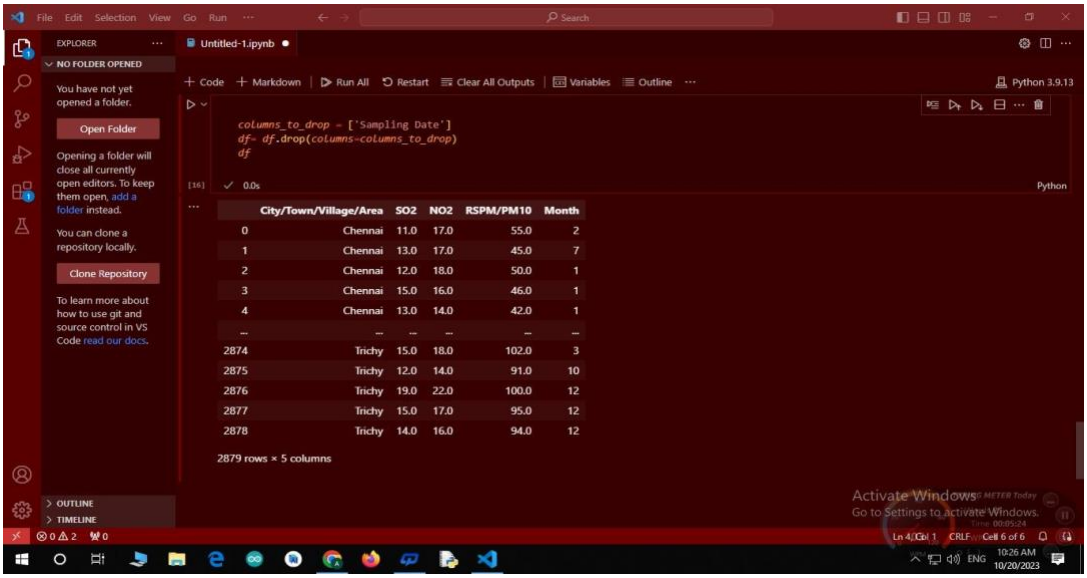
PHASE 4- DEVELOPMENT PART 2

AIR QUALITY ANALYSIS IN TAMILANDU

Feature Engineering:

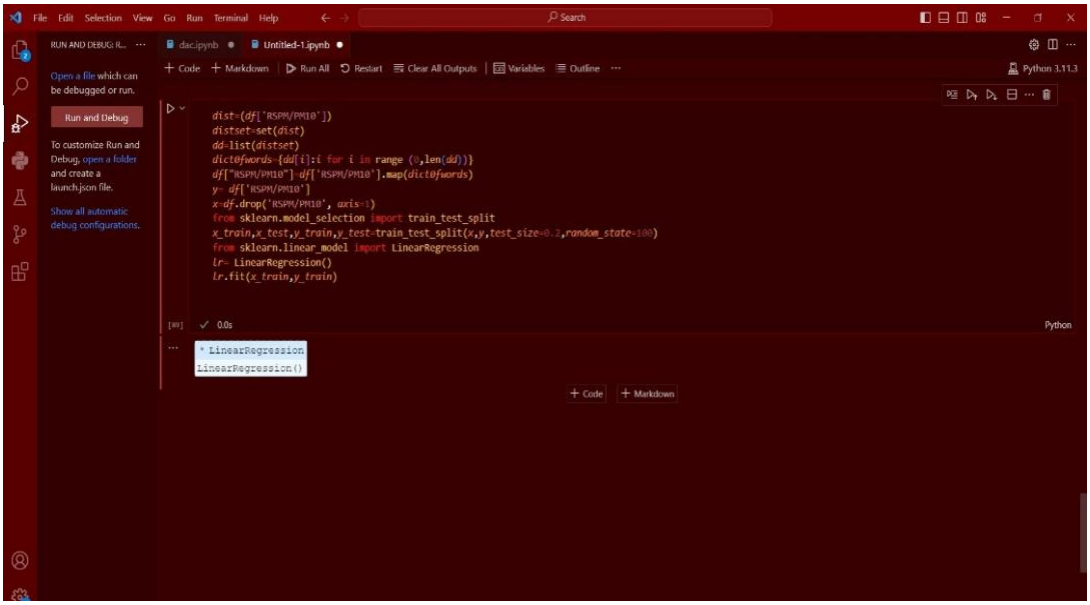
- Feature Selection: Decide which features (columns) of your dataset to include in your analysis. You may need to consider feature importance and domain knowledge to make informed choices.
- Feature Extraction: Create new features from existing ones if they can provide valuable information for your analysis.
- Handling Categorical Data: Convert categorical data into a numerical format, often using techniques like one-hot encoding or label encoding.

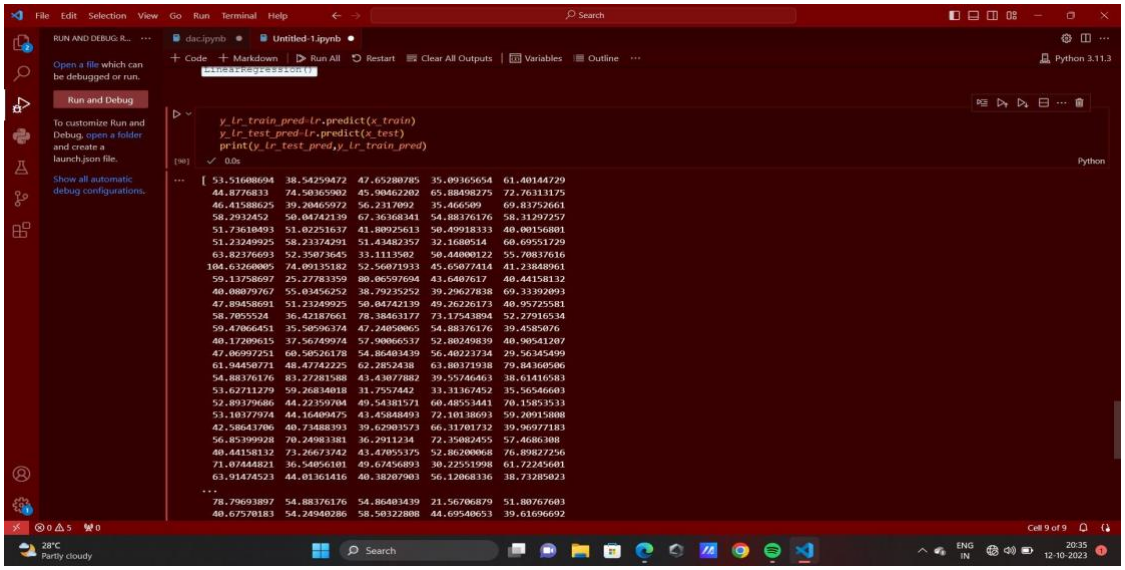




Model Selection and Training:

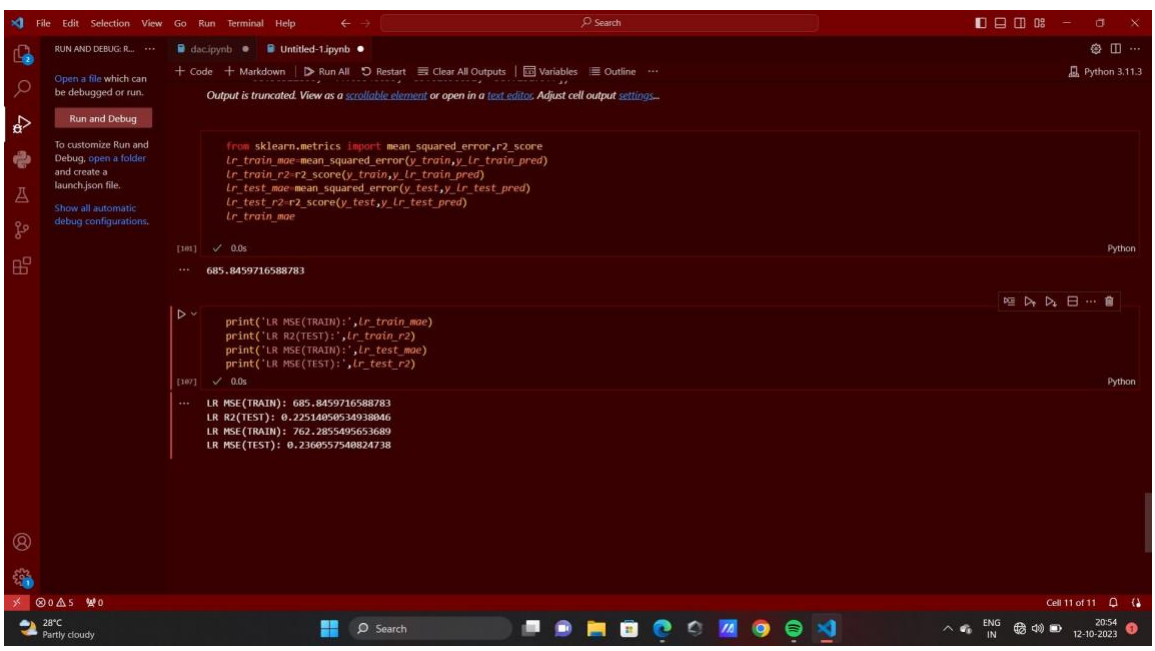
- Choose the appropriate machine learning models for your specific problem. This could include regression, classification, clustering, or other types of models.Split your data into training and testing sets to assess the performance of your models. Common methods include train-test splitting and cross-validation.
- Train your models on the training data. This involves fitting the model to the data to learn the underlying patterns.
- Hyperparameter Tuning: Optimize the model's hyperparameters to achieve the best performance.





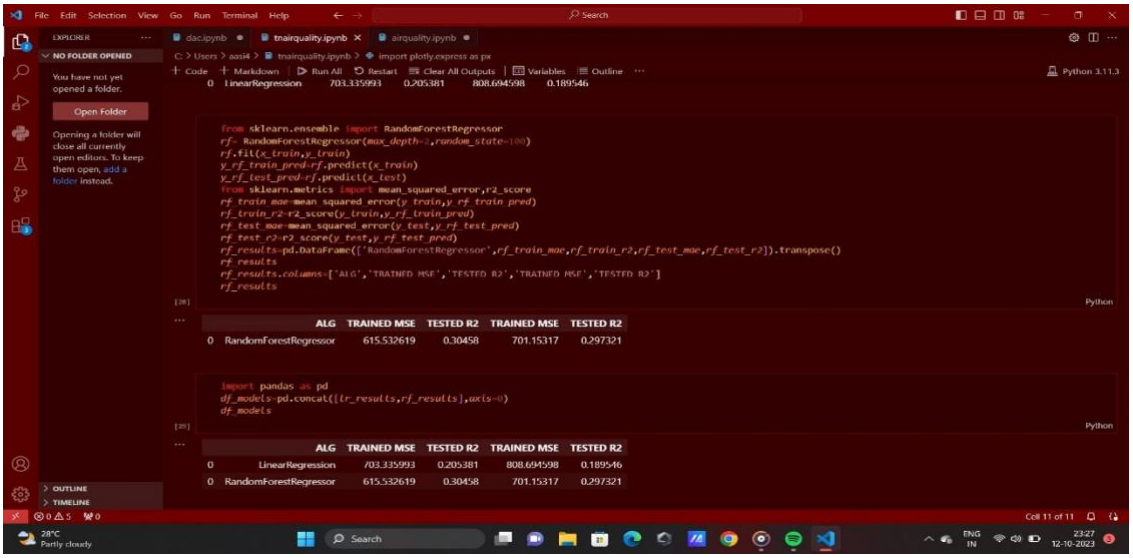
Model Evaluation:

- Evaluate the performance of your models using appropriate metrics. The choice of metrics depends on the type of problem (classification, regression, etc.).
- Common evaluation metrics include accuracy, precision, recall, F1-score, mean squared error (MSE), and others.
- Use visualizations, such as confusion matrices or ROC curves, to gain a deeper understanding of model performance.

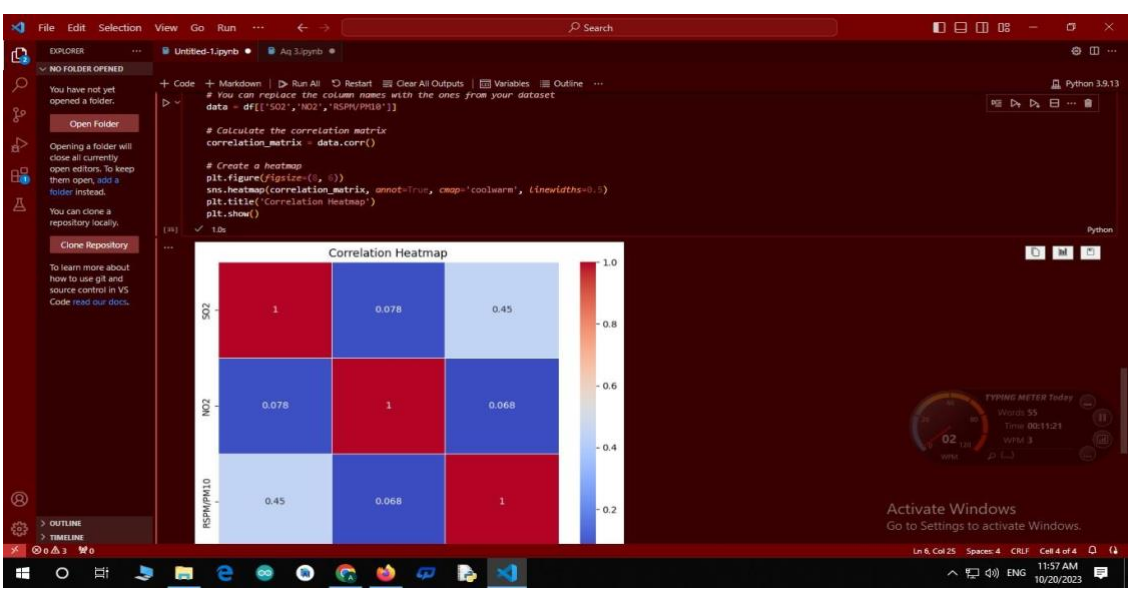


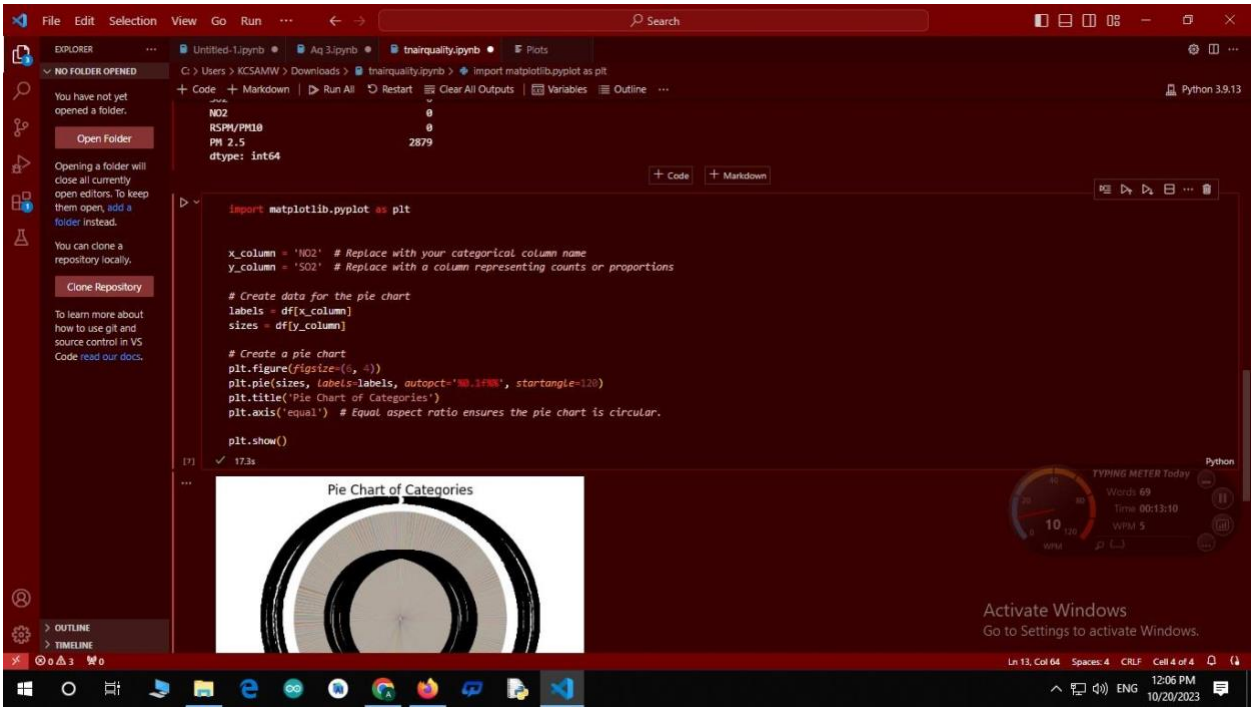
Model Comparison:

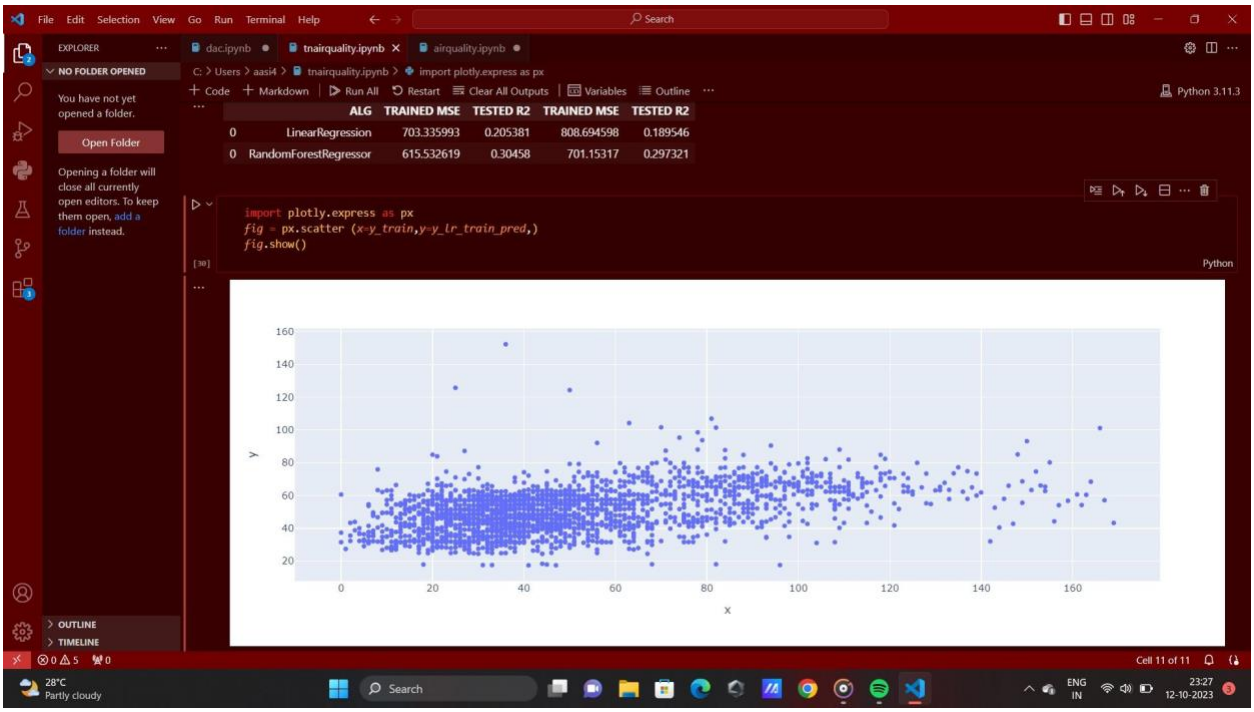
- Compare the performance of different models to determine which one works best for your problem.
- Make use of model selection techniques to choose the best-performing model.



DATA VISULIZATION TO GAIN INSIGHT:




File Edit Selection View Go Run
Untitled-1.ipynb
Aq 3.ipynb
tnairquality.ipynb
Plots
C: > Users > KCSAMW > Downloads > tnairquality.ipynb > import matplotlib.pyplot as plt
Code Markdown Run All Restart Clear All Outputs Variables Outline
Python 3.9.13
NO2 0
RSPM/PM10 0
PM 2.5 2879
dtype: int64
import matplotlib.pyplot as plt
x_column = 'NO2' # Replace with your categorical column name
y_column = 'SO2' # Replace with a column representing counts or proportions
# Create data for the pie chart
labels = df[x_column]
sizes = df[y_column]
# Create a pie chart
plt.figure(figsize=(6, 4))
plt.pie(sizes, labels=labels, autopct='%0.1f%%', startangle=120)
plt.title('Pie Chart of Categories')
plt.axis('equal') # Equal aspect ratio ensures the pie chart is circular.
plt.show()
17.3s
Pie Chart of Categories
Activate Windows
Go to Settings to activate Windows.


File Edit Selection View Go Run Terminal Help
dac.ipynb
tnairquality.ipynb
airquality.ipynb
C: > Users > aasi4 > tnairquality.ipynb > import plotly.express as px
Code Markdown Run All Restart Clear All Outputs Variables Outline
Python 3.11.3
ALG TRAINED MSE TESTED R2 TRAINED MSE TESTED R2
0 LinearRegression 703.335993 0.205381 808.694598 0.189546
0 RandomForestRegressor 615.532619 0.30458 701.15317 0.297321
import plotly.express as px
fig = px.scatter (x=y_train,y=y_lr_train_pred,)
fig.show()
y
x