AIR QUALITY ANALYSIS IN TAMIL NADU

DATA ANALYSIS WITH COGNOS

TEAM LEADER - S.YOGESHWARI

TEAM MEMBER - D.ABINAYA

TEAM MEMBER - M.ABINAYA

TEAM MEMBER - M.AARTHI

PHASE-3

DEVELOPMENT PART

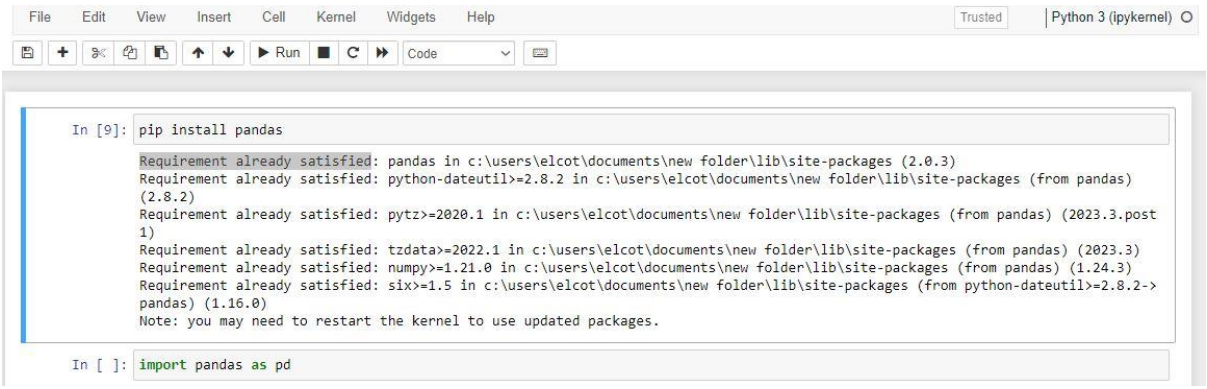
1.Import Libraries:

- First, make sure you have Python and pandas installed. You can install pandas using pip if you don't have it already

```bash
pip install pandas
```

```python
import pandas as pd
```

- A module can be imported into an interactive console environment .

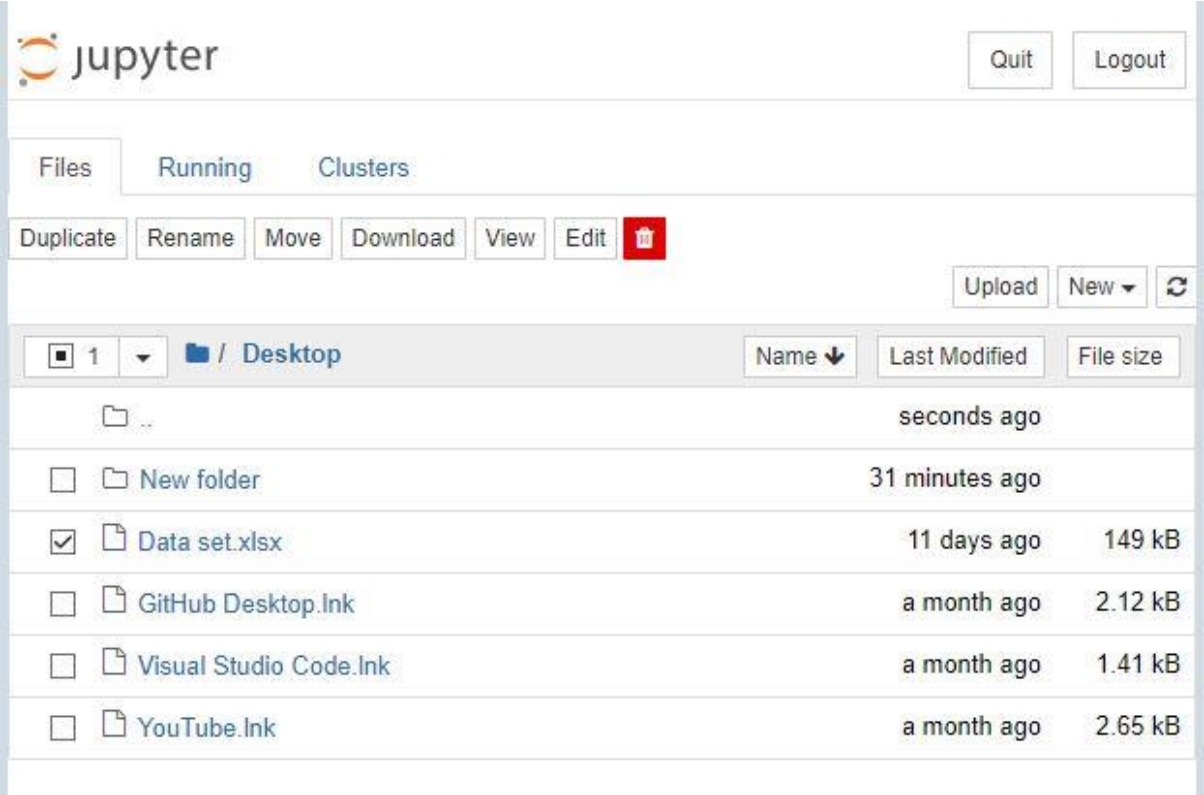


2.Load the Dataset:

- You need to have your air quality dataset in a compatible format, such as a CSV, Excel, or other tabular format. Let's assume your dataset is in a CSV file. You can load it using the' `read_ csv'` function:

```python
# Replace 'your_dataset.csv' with the actual file path of your dataset.
df = pd.read_csv('your_dataset.csv')
```

- Then click the blue Upload button displayed in the file's row to add the file to the project.

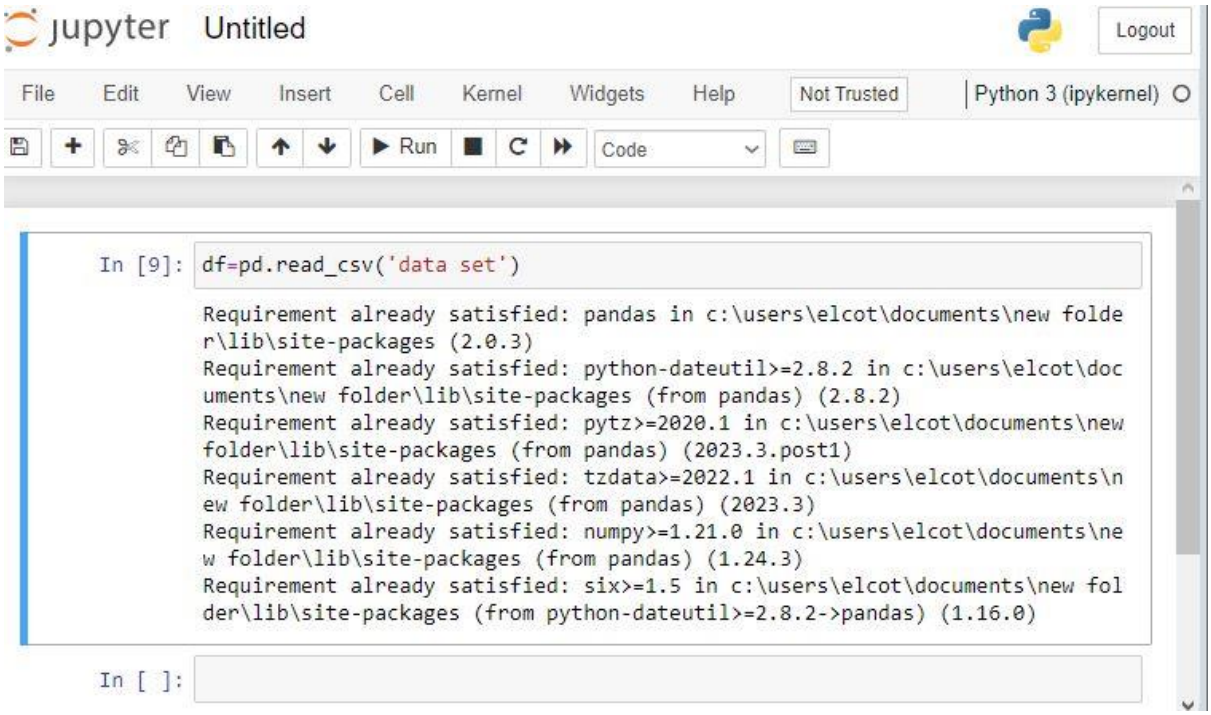


3.Explore the Dataset:

- Once you've loaded the dataset, you should explore it to understand its structure. You can start by checking the first few rows to get a glimpse of the data.

```python
# Display the first 5 rows of the dataset.
print(df.head())
```

- Explore data set the through the Object Storage Explorer and Table Explorer options on the left toolbar of the this Lab interface.



4.Data pre-processing:

- Data pre-processing is essential to ensure the data is clean and ready for analysis. Common pre-processing steps include handling missing values, removing duplicates, and dealing with data types.

Handling Missing Value:

- You can check for missing values in your dataset using the ' isna ()' function and then decide how to handle them. For example, to count missing values in each column.

```python
missing_values = df.isna().sum()
print(missing_values)
```

Date type conversions:

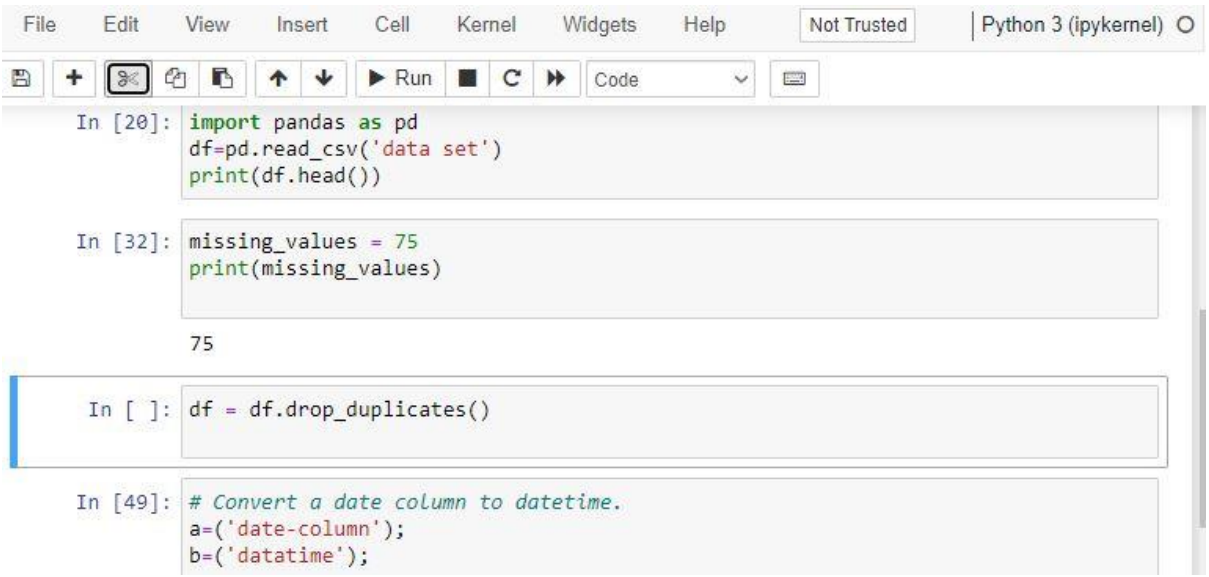
- Ensure that columns have the correct data types. For instance, dates should be converted to datetime objects if they are represented as strings.

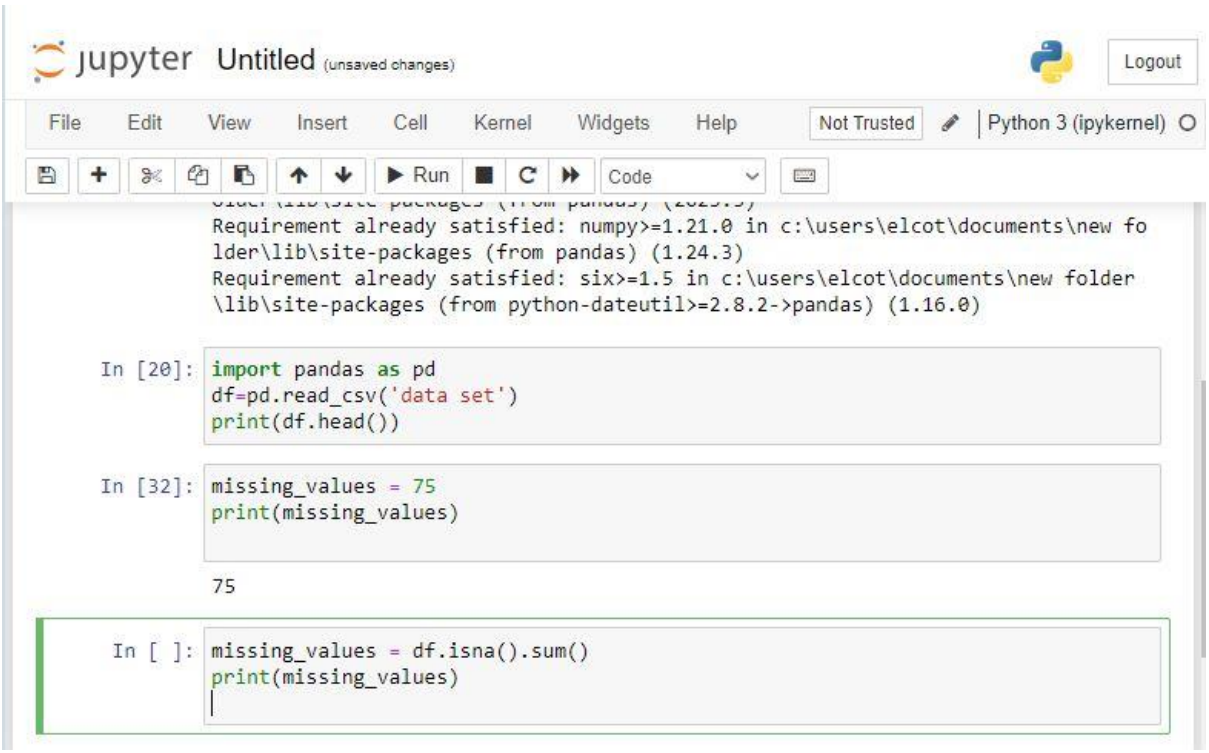
```python
# Convert a date column to datetime.
df['date_column'] = pd.to_datetime(df['date_column'])
```

Handling Duplicates:

Check for and remove duplicates if they exist in the dataset.

```python
df = df.drop_duplicates()
```


File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
In [20]:
import pandas as pd
df=pd.read_csv('data set')
print(df.head())
In [32]:
missing_values = 75
print(missing_values)
75
In [ ]:
df = df.drop_duplicates()
In [49]:
# Convert a date column to datetime.
a=('date-column');
b=('datatime');

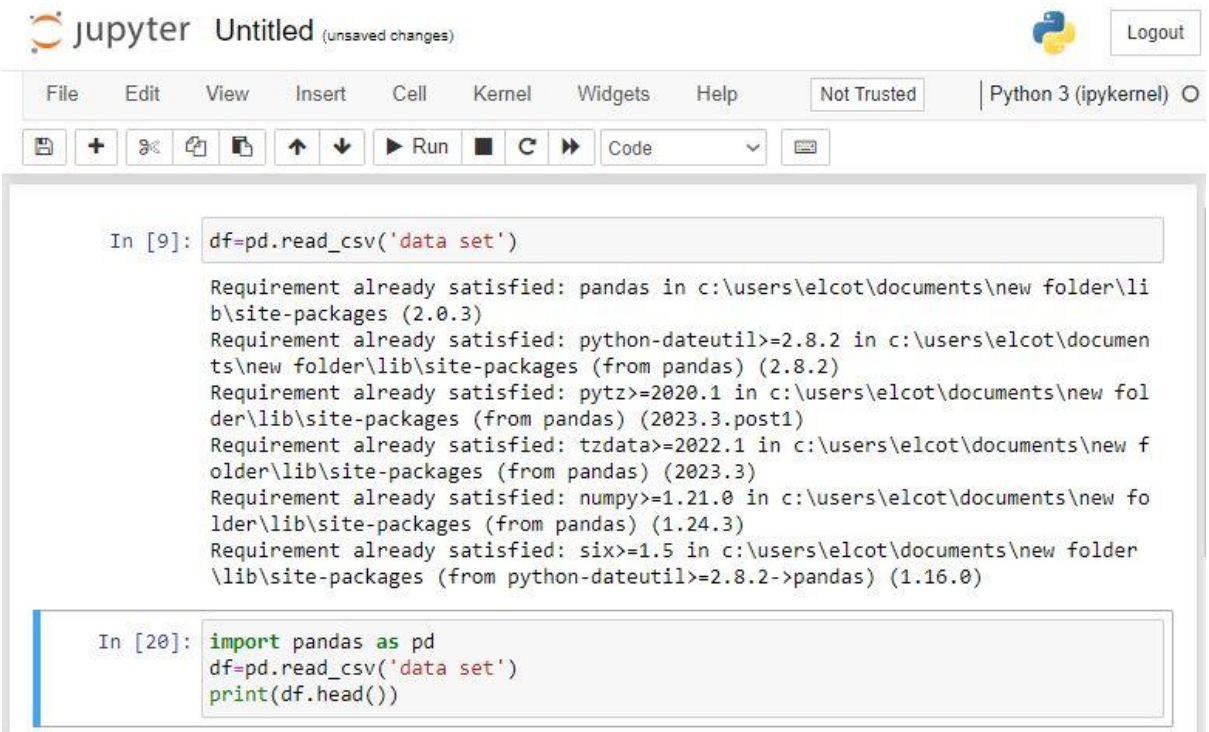

jupyter Untitled (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Requirement already satisfied: numpy>=1.21.0 in c:\users\elcot\documents\new fo
lder\lib\site-packages (from pandas) (1.24.3)
Requirement already satisfied: six>=1.5 in c:\users\elcot\documents\new folder
\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
In [20]:
import pandas as pd
df=pd.read_csv('data set')
print(df.head())
In [32]:
missing_values = 75
print(missing_values)
75
In [ ]:
missing_values = df.isna().sum()
print(missing_values)

5.Save pre-processed Data:

- After pre-processing, you may want to save the pre-processed dataset for future use:

```python
df.to_csv('preprocessed_data.csv', index=False)
```

Remember to replace 'your_dataset.csv' with the actual path to your air quality dataset and adapt the pre-processing steps to your specific data. If you encounter any issues or need further assistance, feel free to ask for help with specific aspects of pre-processing.



….End of the project Assessment….