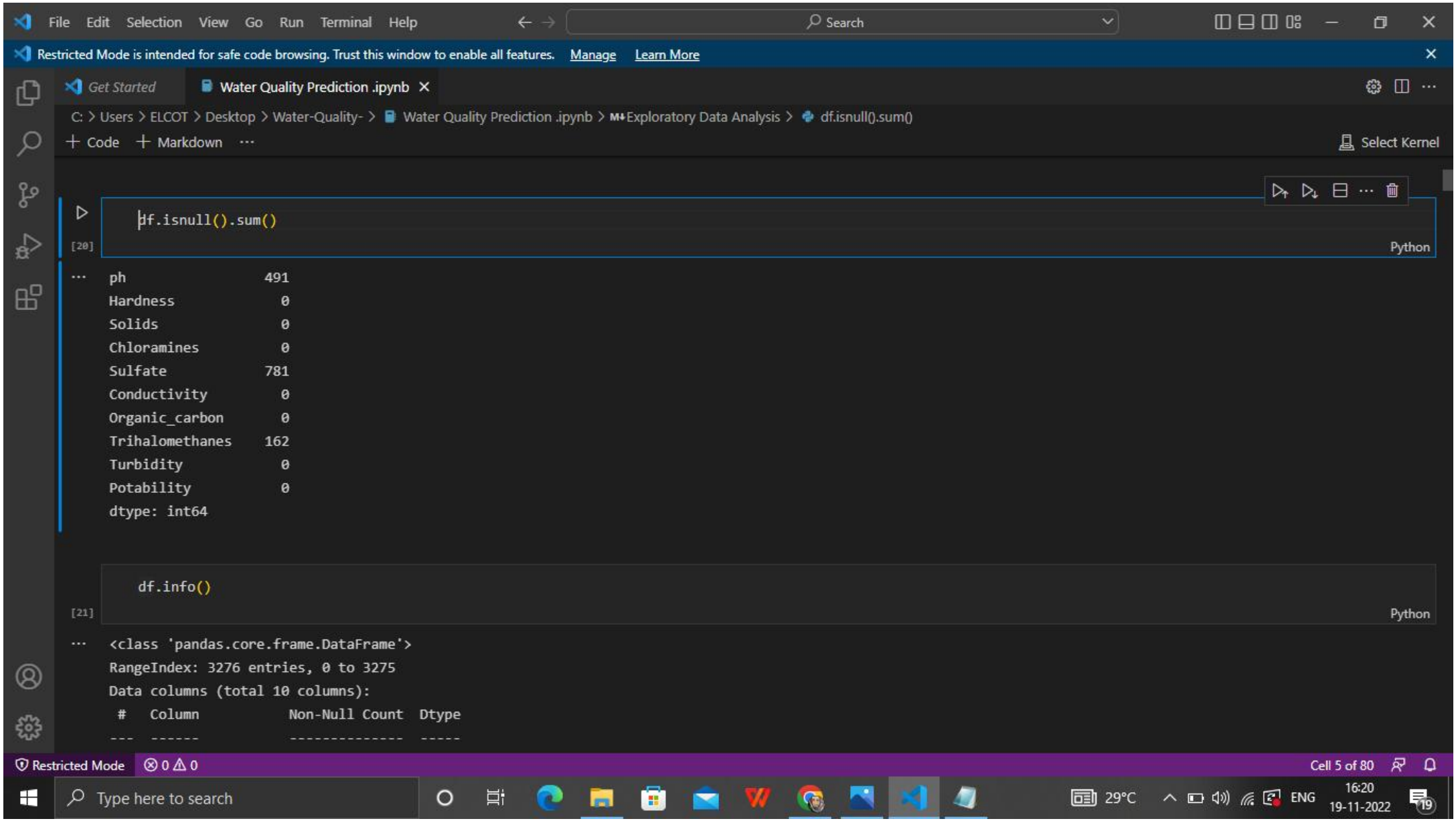
SPRINT 3

Date	15 November 2022
Team ID	PNT2022TMID49672
Project Name	Efficient Water Quality Analysis and Prediction Using Machine Learning
Maximum Marks	8 Marks


File Edit Selection View Go Run Terminal Help
Search
Restricted Mode is intended for safe code browsing. Trust this window to enable all features. Manage Learn More
Get Started
Water Quality Prediction .ipynb
C: > Users > ELCOT > Desktop > Water-Quality- > Water Quality Prediction .ipynb > Exploratory Data Analysis > df.isnull().sum()
+ Code + Markdown ···
Select Kernel
[20]
df.isnull().sum()
Python
ph 491
Hardness 0
Solids 0
Chloramines 0
Sulfate 781
Conductivity 0
Organic_carbon 0
Trihalomethanes 162
Turbidity 0
Potability 0
dtype: int64
[21]
df.info()
Python
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3276 entries, 0 to 3275
Data columns (total 10 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
Restricted Mode
Cell 5 of 80
Type here to search
29°C
ENG
16:20
19-11-2022

File Edit Selection View Go Run Terminal Help

Search

Restricted Mode is intended for safe code browsing. Trust this window to enable all features. Manage Learn More

Get Started | Water Quality Prediction .ipynb

C: > Users > ELCOT > Desktop > Water-Quality- > Water Quality Prediction .ipynb > Exploratory Data Analysis > df.isnull().sum()

+ Code + Markdown ···

Select Kernel

```
---  ------           --------------  -----
 0   ph               2785 non-null   float64
 1   Hardness         3276 non-null   float64
 2   Solids           3276 non-null   float64
 3   Chloramines      3276 non-null   float64
 4   Sulfate          2495 non-null   float64
 5   Conductivity     3276 non-null   float64
 6   Organic_carbon   3276 non-null   float64
 7   Trihalomethanes  3114 non-null   float64
 8   Turbidity        3276 non-null   float64
 9   Potability       3276 non-null   int64
dtypes: float64(9), int64(1)
memory usage: 256.1 KB
```

[22]

```
df.describe()
```

Python

···

[23]

```
df['Sulfate'].mean()
```

Python

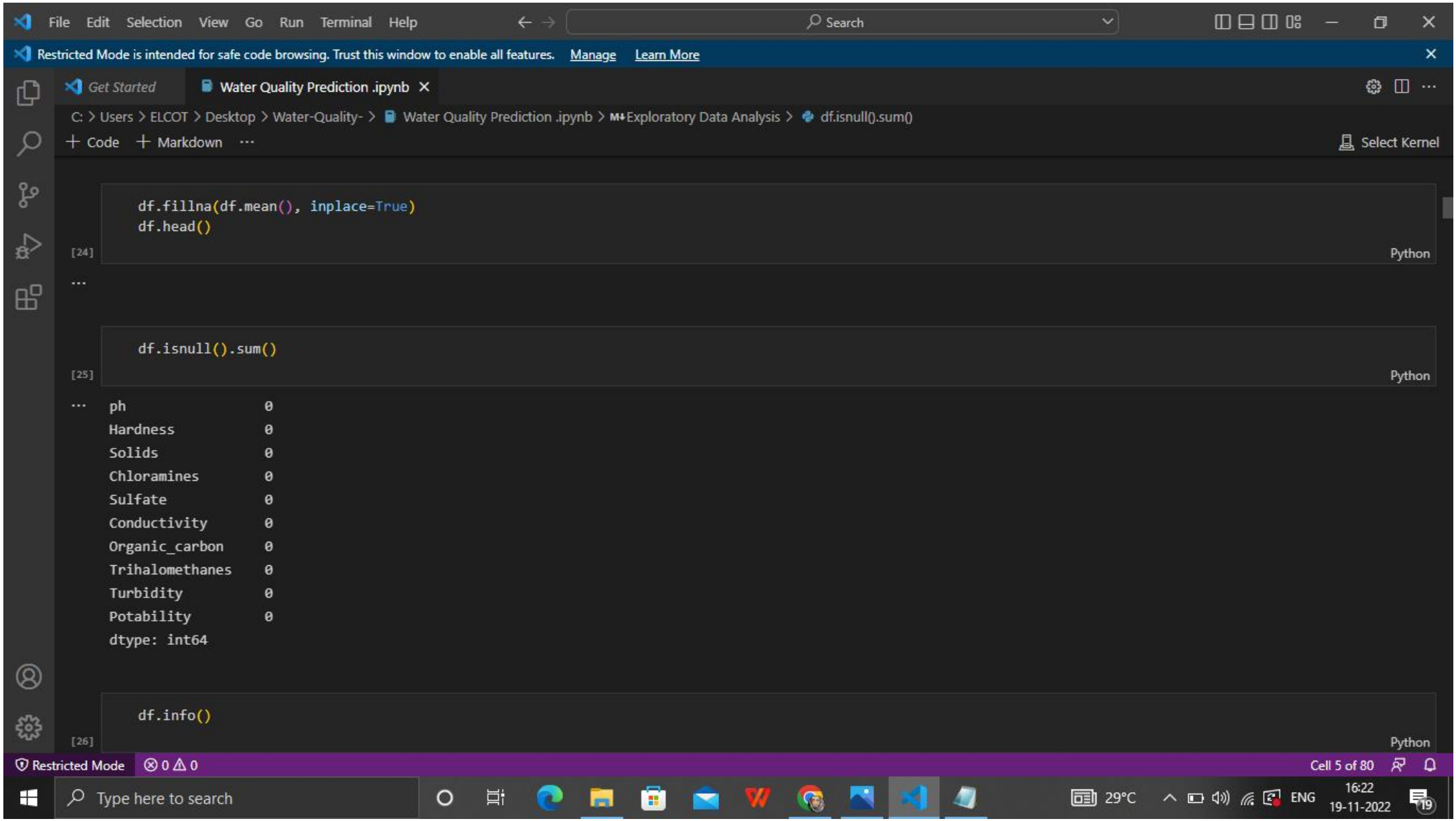
```
333.77577661081335
```

Restricted Mode ⊗ 0 ⚠ 0

Cell 5 of 80

Type here to search

29°C ENG 16:21 19-11-2022


File Edit Selection View Go Run Terminal Help
Search
Restricted Mode is intended for safe code browsing. Trust this window to enable all features. Manage Learn More
Get Started
Water Quality Prediction .ipynb
C: > Users > ELCOT > Desktop > Water-Quality- > Water Quality Prediction .ipynb > Exploratory Data Analysis > df.isnull().sum()
+ Code + Markdown ···
Select Kernel
df.fillna(df.mean(), inplace=True)
df.head()
[24]
Python
···
df.isnull().sum()
[25]
Python
··· ph 0
Hardness 0
Solids 0
Chloramines 0
Sulfate 0
Conductivity 0
Organic_carbon 0
Trihalomethanes 0
Turbidity 0
Potability 0
dtype: int64
df.info()
[26]
Python
Restricted Mode
⊗ 0 ⚠ 0
Cell 5 of 80
Type here to search
29°C
ENG
16:22
19-11-2022

File Edit Selection View Go Run Terminal Help

Search

Restricted Mode is intended for safe code browsing. Trust this window to enable all features. Manage Learn More

Get Started | Water Quality Prediction .ipynb

C: > Users > ELCOT > Desktop > Water-Quality- > Water Quality Prediction .ipynb > Exploratory Data Analysis > df.isnull().sum()

+ Code + Markdown ···

Select Kernel

```
[26]  df.info()
```

Python

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3276 entries, 0 to 3275
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   ph               3276 non-null   float64
 1   Hardness         3276 non-null   float64
 2   Solids           3276 non-null   float64
 3   Chloramines      3276 non-null   float64
 4   Sulfate          3276 non-null   float64
 5   Conductivity     3276 non-null   float64
 6   Organic_carbon   3276 non-null   float64
 7   Trihalomethanes  3276 non-null   float64
 8   Turbidity        3276 non-null   float64
 9   Potability       3276 non-null   int64  
dtypes: float64(9), int64(1)
memory usage: 256.1 KB
```

```
[27]  df.describe()
```

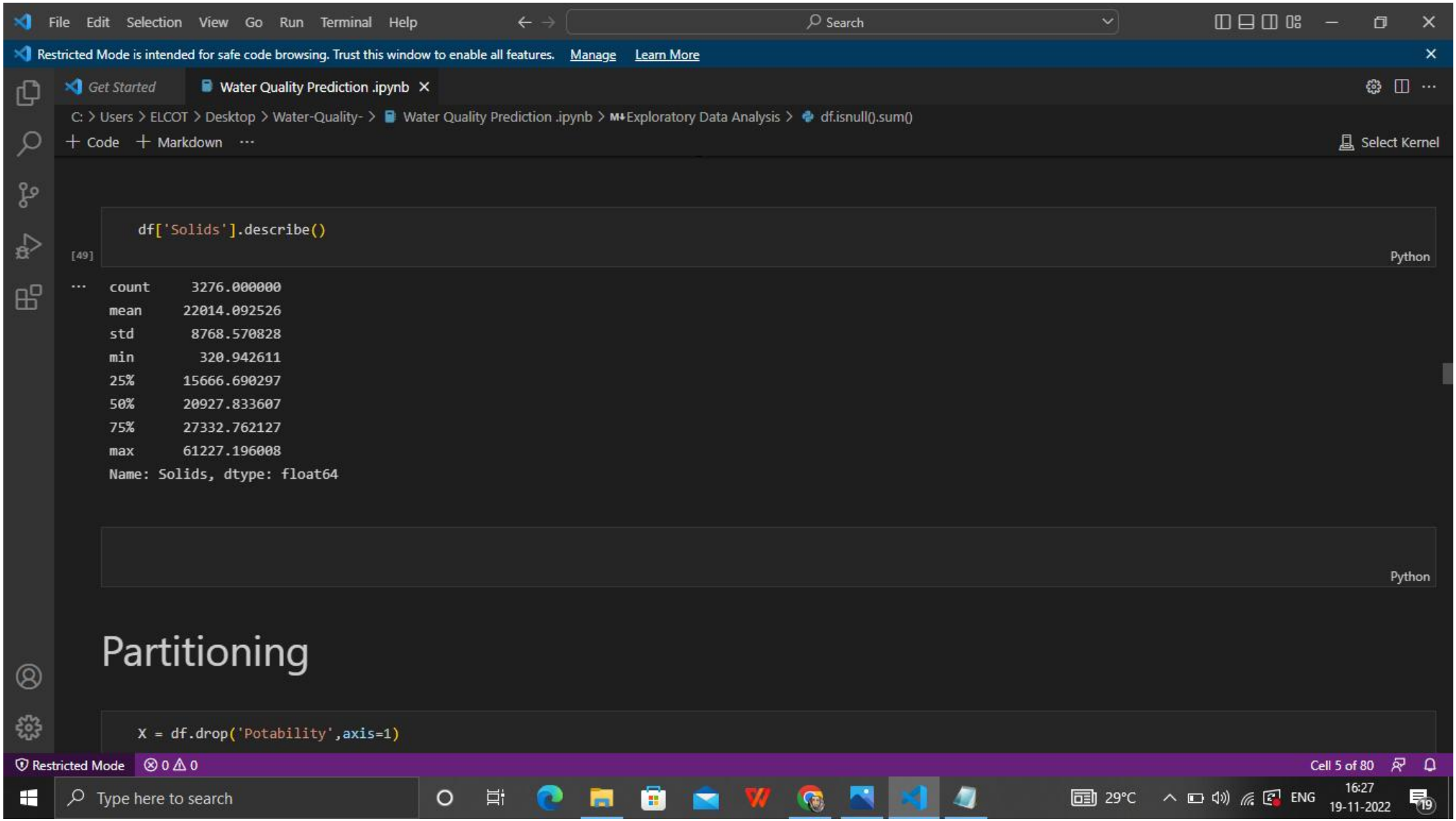
Python

Restricted Mode ⊗ 0 ⚠ 0

Cell 5 of 80

Type here to search

29°C ENG 16:22 19-11-2022


File Edit Selection View Go Run Terminal Help
Search
Restricted Mode is intended for safe code browsing. Trust this window to enable all features. Manage Learn More
Get Started
Water Quality Prediction .ipynb
C: > Users > ELCOT > Desktop > Water-Quality- > Water Quality Prediction .ipynb > Exploratory Data Analysis > df.isnull().sum()
+ Code + Markdown ···
Select Kernel
df['Solids'].describe()
[49]
Python
count 3276.000000
mean 22014.092526
std 8768.570828
min 320.942611
25% 15666.690297
50% 20927.833607
75% 27332.762127
max 61227.196008
Name: Solids, dtype: float64
Python
Partitioning
X = df.drop('Potability',axis=1)
Restricted Mode
Cell 5 of 80
Type here to search
29°C
ENG
16:27
19-11-2022