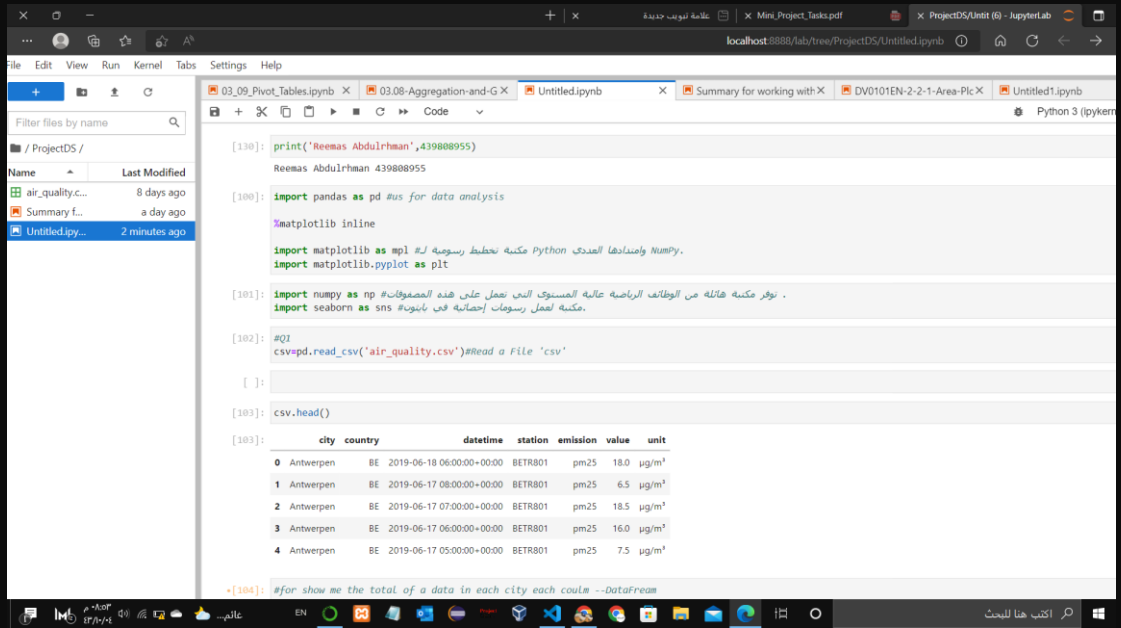


Reemas Abdulrhman 439808955



```
[130]: print('Reemas Abdulrhman',439808955)
Reemas Abdulrhman 439808955

[100]: import pandas as pd #us for data analysis
import matplotlib inline

import matplotlib as mpl #مكتبة Python مخطيط رسومية لـ
import matplotlib.pyplot as plt

[101]: import numpy as np #هذه المكتبة عالية المستوى التي تعمل على هذه المكتبات
import seaborn as sns #مكتبة لعمل رسومات إحصائية في بايثون

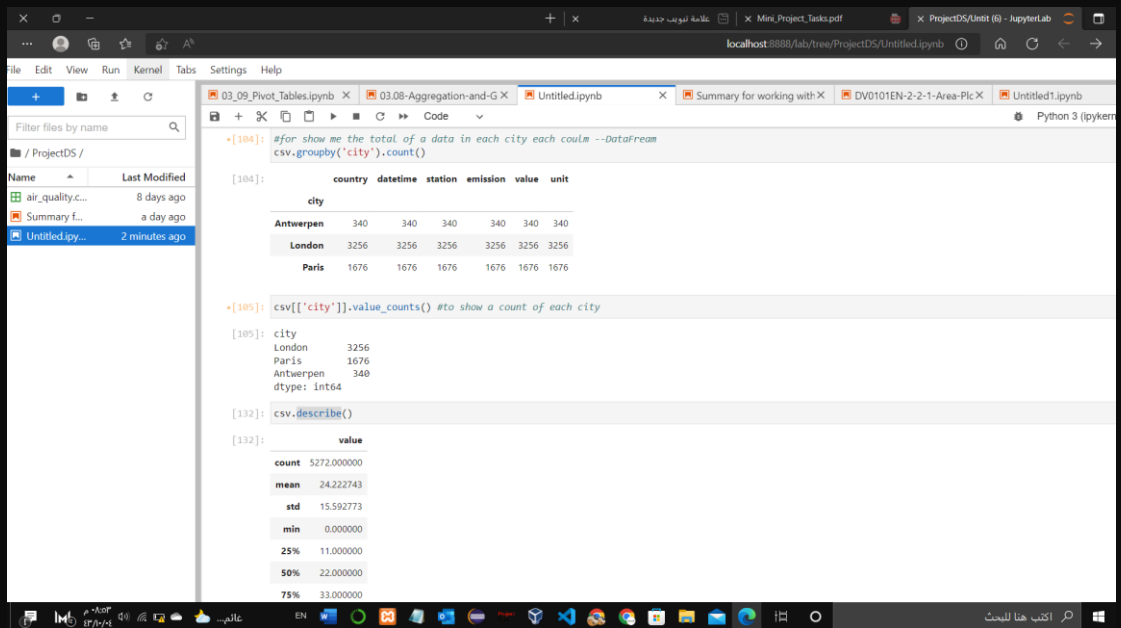
[102]: #Q1
csv=pd.read_csv('air_quality.csv')#Read a File 'csv'

[ ]:

[103]: csv.head()
```

	city	country	datetime	station	emission	value	unit
0	Antwerpen	BE	2019-06-18 06:00:00+00:00	BETR801	pm25	18.0	µg/m³
1	Antwerpen	BE	2019-06-17 08:00:00+00:00	BETR801	pm25	6.5	µg/m³
2	Antwerpen	BE	2019-06-17 07:00:00+00:00	BETR801	pm25	18.5	µg/m³
3	Antwerpen	BE	2019-06-17 06:00:00+00:00	BETR801	pm25	16.0	µg/m³
4	Antwerpen	BE	2019-06-17 05:00:00+00:00	BETR801	pm25	7.5	µg/m³

```
+ [104]: #for show me the total of a data in each city each coulm --DataFream
```



```
+ [104]: #for show me the total of a data in each city each coulm --DataFream
csv.groupby('city').count()
```

	country	datetime	station	emission	value	unit
city						
Antwerpen	340	340	340	340	340	340
London	3256	3256	3256	3256	3256	3256
Paris	1676	1676	1676	1676	1676	1676

```
+ [105]: csv[['city']].value_counts() #to show a count of each city

[105]: city
London      3256
Paris       1676
Antwerpen   340
dtype: int64

[102]: csv.describe()

[102]:
```

	value
count	5272.000000
mean	24.222743
std	15.592773
min	0.000000
25%	11.000000
50%	22.000000
75%	33.000000

localhost:8888/lab/tree/ProjectDS/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/ ProjectDS /

Name	Last Modified
air_quality.c...	8 days ago
Summary f...	a day ago
Untitled.ipynb	2 minutes ago

```
[107]: #q2
#Null value to --> Zero pivot
csv.replace(np.nan,0,inplace=True)
csv.head(5)

[107]:
```

	city	country	datetime	station	emission	value	unit
0	Antwerpen	BE	2019-06-18 06:00:00+00:00	BETR801	pm25	18.0	µg/m³
1	Antwerpen	BE	2019-06-17 08:00:00+00:00	BETR801	pm25	6.5	µg/m³
2	Antwerpen	BE	2019-06-17 07:00:00+00:00	BETR801	pm25	18.5	µg/m³
3	Antwerpen	BE	2019-06-17 06:00:00+00:00	BETR801	pm25	16.0	µg/m³
4	Antwerpen	BE	2019-06-17 05:00:00+00:00	BETR801	pm25	7.5	µg/m³

```
+ [108]: csv.isnull().any()#check if any have null value

[108]:
```

	city	country	datetime	station	emission	value	unit
city	False	False	False	False	False	False	False
country	False	False	False	False	False	False	False
datetime	False	False	False	False	False	False	False
station	False	False	False	False	False	False	False
emission	False	False	False	False	False	False	False
value	False	False	False	False	False	False	False
unit	False	False	False	False	False	False	False
dtype:	bool	bool	bool	bool	bool	bool	bool

```
[109]: #q2
#pivot table that shows the maximum value of emission for each city
#Null value to --> Zero pivot
csv.groupby('city')[['emission']].max()

[109]:
```

	emission
city	
Antwerpen	pm25
London	pm25
Paris	no2

localhost:8888/lab/tree/ProjectDS/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/ ProjectDS /

Name	Last Modified
air_quality.c...	8 days ago
Summary f...	a day ago
Untitled.ipynb	2 minutes ago

```
[109]: #q2
#pivot table that shows the maximum value of emission for each city
#Null value to --> Zero pivot
csv.groupby('city')[['emission']].max()

[109]:
```

	emission
city	
Antwerpen	pm25
London	pm25
Paris	no2

```
[110]: #Q3 Create a new data frame that contains the data of 'London' city and explore your data.
#London DataFrame
csv

[110]:
```

	city	country	datetime	station	emission	value	unit
0	Antwerpen	BE	2019-06-18 06:00:00+00:00	BETR801	pm25	18.0	µg/m³
1	Antwerpen	BE	2019-06-17 08:00:00+00:00	BETR801	pm25	6.5	µg/m³
2	Antwerpen	BE	2019-06-17 07:00:00+00:00	BETR801	pm25	18.5	µg/m³
3	Antwerpen	BE	2019-06-17 06:00:00+00:00	BETR801	pm25	16.0	µg/m³
4	Antwerpen	BE	2019-06-17 05:00:00+00:00	BETR801	pm25	7.5	µg/m³
...
5267	London	GB	2019-04-09 06:00:00+00:00	London Westminster	no2	41.0	µg/m³
5268	London	GB	2019-04-09 05:00:00+00:00	London Westminster	no2	41.0	µg/m³
5269	London	GB	2019-04-09 04:00:00+00:00	London Westminster	no2	41.0	µg/m³

localhost:8888/lab/tree/ProjectDS/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/ ProjectDS /

Name	Last Modified
air_quality.c...	8 days ago
Summary f...	a day ago
Untitled.ipynb	2 minutes ago

```

[111]: #function-->Do filter the London from 'city'
def filter_func(x):
    return (x['city']=='London').any()

[112]: london=csv.groupby('city').filter(filter_func)#function-->Do filter the London from 'city'--Save in a veribel
London

[112]:
  city country      datetime      station  emission  value  unit
177  London    GB  2019-06-21 00:00:00+00:00  London Westminster    pm25    7.0  µg/m³
178  London    GB  2019-06-20 23:00:00+00:00  London Westminster    pm25    7.0  µg/m³
179  London    GB  2019-06-20 22:00:00+00:00  London Westminster    pm25    7.0  µg/m³
180  London    GB  2019-06-20 21:00:00+00:00  London Westminster    pm25    8.0  µg/m³
181  London    GB  2019-06-20 20:00:00+00:00  London Westminster    pm25    8.0  µg/m³
...
5267 London    GB  2019-04-09 06:00:00+00:00  London Westminster    no2    41.0  µg/m³
5268 London    GB  2019-04-09 05:00:00+00:00  London Westminster    no2    41.0  µg/m³
5269 London    GB  2019-04-09 04:00:00+00:00  London Westminster    no2    41.0  µg/m³
5270 London    GB  2019-04-09 03:00:00+00:00  London Westminster    no2    67.0  µg/m³
5271 London    GB  2019-04-09 02:00:00+00:00  London Westminster    no2    67.0  µg/m³

3256 rows x 7 columns

[113]: #London=csv[csv['city']=='London'] #Another way to creat a DataFream
#London.head()

[ ]:

```

localhost:8888/lab/tree/ProjectDS/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/ ProjectDS /

Name	Last Modified
air_quality.c...	8 days ago
Summary f...	a day ago
Untitled.ipynb	3 minutes ago

```

[114]: csv['city']=='London'#check if thear any index have 'London' in 'csv DataFream'

[114]:
0      False
1      False
2      False
3      False
4      False
...
5267     True
5268     True
5269     True
5270     True
5271     True
Name: city, Length: 5272, dtype: bool

[115]: london['city']=='London'#check if thear any index have 'London' in 'London DataFream'

[115]:
177     True
178     True
179     True
180     True
181     True
...
5267     True
5268     True
5269     True
5270     True
5271     True
Name: city, Length: 3256, dtype: bool

[116]: #q4 Set the appropriate data type for the column datetime and make it as index.
#DateTime-->Index
london.info()

<class 'pandas.core.frame.DataFrame'>

```

localhost:8888/lab/tree/ProjectDS/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

ProjectDS /

Name	Last Modified
air_quality.c...	8 days ago
Summary f...	a day ago
Untitled.ipynb	3 minutes ago

```
[116]: #4 Set the appropriate data type for the column datetime and make it as index.
       london['datetime'] = pd.to_datetime(london['datetime'])

       london.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 3256 entries, 177 to 5271
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  --
 0   city        3256 non-null   object
 1   country     3256 non-null   object
 2   datetime    3256 non-null   object
 3   station     3256 non-null   object
 4   emission    3256 non-null   object
 5   value       3256 non-null   float64
 6   unit        3256 non-null   object
dtypes: float64(1), object(6)
memory usage: 203.5+ KB

[117]: # convert the 'datetime' column to datetime format
       london['datetime'] = pd.to_datetime(london['datetime'])

[118]: # Check the format of 'Date' column
       london.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 3256 entries, 177 to 5271
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  --
 0   city        3256 non-null   object
 1   country     3256 non-null   object
 2   datetime    3256 non-null   datetime64[ns, UTC]
 3   station     3256 non-null   object
 4   emission    3256 non-null   object
 5   value       3256 non-null   float64
 6   unit        3256 non-null   object
dtypes: datetime64(1), object(6)
memory usage: 203.5+ KB
```

localhost:8888/lab/tree/ProjectDS/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

ProjectDS /

Name	Last Modified
air_quality.c...	8 days ago
Summary f...	a day ago
Untitled.ipynb	3 minutes ago

```
[119]: #datetime to index the DataFrame
       london.set_index('datetime', inplace=True)

[120]: london.head()

[120]:
```

	city	country	station	emission	value	unit
datetime						
2019-06-21 00:00:00+00:00	London	GB	London Westminster	pm25	7.0	µg/m³
2019-06-20 23:00:00+00:00	London	GB	London Westminster	pm25	7.0	µg/m³
2019-06-20 22:00:00+00:00	London	GB	London Westminster	pm25	7.0	µg/m³
2019-06-20 21:00:00+00:00	London	GB	London Westminster	pm25	8.0	µg/m³
2019-06-20 20:00:00+00:00	London	GB	London Westminster	pm25	8.0	µg/m³

```
[121]: #45 Use resample () to show the weekly total value of emission that produced at London city.-->value-->emission ** النوع الانمات
       london.resample('W').sum()

# لإظهار القيمة الإجمالية الأسبوعية للانمات الناتجة في مدينة لندن
# مجموع الانمات في لندن
```

localhost:8888/lab/tree/ProjectDS/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

03_09_Pivot_Tables.ipynb 03_08-Aggregation-and-G- Untitled.ipynb Summary for working with DV0101EN-2-2-1-Area-PlcX Untitled1.ipynb

Python 3 (pyker

```
[121]: #q5 Use resample () to show the weekly total value of emission that produced at london city.-->value-->emission '' النوع الانبعاثات
#resample () لإظهار القيمة الإجمالية الأسبوعية للانبعاثات الناتجة في مدينة لندن
#مجموع الانبعاثات في لندن

[122]: لإعادة تشكيل بيانات السلاسل الزمنية. طريقة ملاحظة لتحويل التردد وإعادة تشكيل السلاسل الزمنية.
weekly_total=london.resample("W")['value'].sum()#w->week #D->Day #M->month

[123]: weekly_total

[123]: datetime
2019-04-14 00:00:00+00:00    7291.0
2019-04-21 00:00:00+00:00    13769.0
2019-04-28 00:00:00+00:00    8177.0
2019-05-05 00:00:00+00:00    6942.0
2019-05-12 00:00:00+00:00    6296.0
2019-05-19 00:00:00+00:00    7150.0
2019-05-26 00:00:00+00:00    6581.0
2019-06-02 00:00:00+00:00    4806.0
2019-06-09 00:00:00+00:00    3920.0
2019-06-16 00:00:00+00:00    4501.0
2019-06-23 00:00:00+00:00     544.0
Freq: W-SUN, Name: value, dtype: float64

[ ]:

[139]: list(weekly_total)

[139]: [7291.0,
13769.0,
8177.0,
6942.0,
6296.0,
7150.0,
6581.0,
```

localhost:8888/lab/tree/ProjectDS/Untitled.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

03_09_Pivot_Tables.ipynb 03_08-Aggregation-and-G- Untitled.ipynb Summary for working with DV0101EN-2-2-1-Area-PlcX Untitled1.ipynb

Python 3 (pyker

```
[127]: #q6 Finally, create a clear bar chart that shows the result of tasks
#bar chart -->For q5(sum for London)

[128]: weekly_total.plot(kind="bar")# a clear bar chart for weekly_total'value'

[128]: <AxesSubplot: xlabel='datetime'>

[129]: london.resample('W')['value'].sum().plot() # لإظهار التردد الإحصائي على رسم بياني
[129]: <AxesSubplot: xlabel='datetime'>
```

