

Mini Project – 1

Name : Umesh Padul

Roll No. : COTB45

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

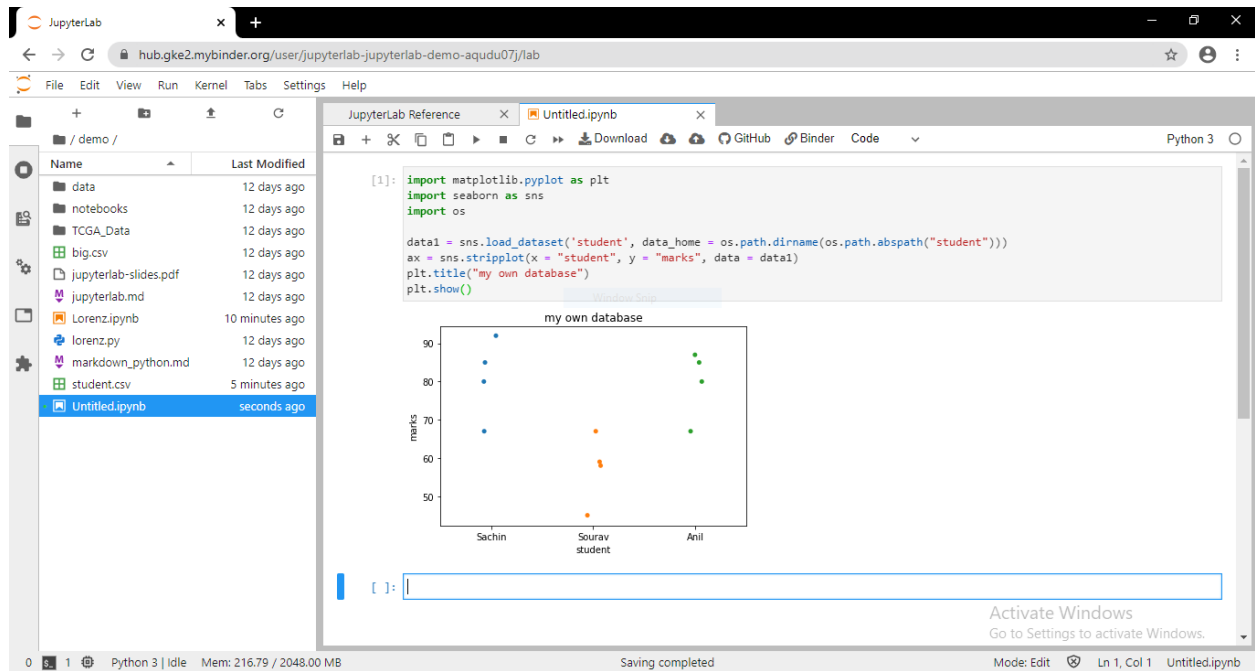
```
import os
```

```
data1 = sns.load_dataset('student', data_home = os.path.dirname(os.path.abspath("student")))
```

```
ax = sns.barplot(x = "student", y = "marks", data = data1)
```

```
plt.title("my own database")
```

```
plt.show()
```



```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

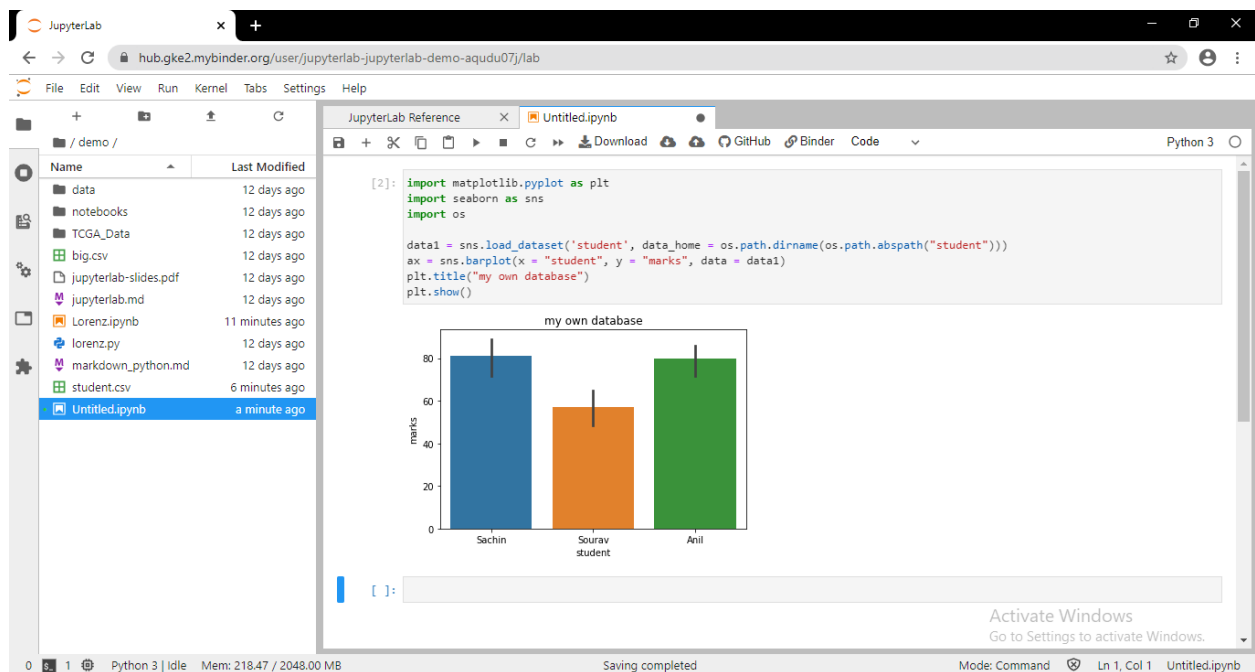
```
import os
```

```
data1 = sns.load_dataset('student', data_home = os.path.dirname(os.path.abspath("student")))
```

```
ax = sns.stripplot(x = "student", y = "marks", data = data1)
```

```
plt.title("my own database")
```

```
plt.show()
```



```
import pandas as pd
```

```
medals = pd.read_csv(r"summer.csv")
```

```
medals.head()
```

```
medals.info()
```

```
country_names = medals['Country']
```

```
country_names.head()
```

```
medal_counts = country_names.value_counts()
```

```
medal_counts.head()
```

```
medals.head(3)
```

```
counted = medals.pivot_table(index='Country', values='Athlete', columns='Medal', aggfunc='count')
```

```
counted.head(7)
```

```
counted['totals'] = counted.sum(axis='columns')
```

```
counted.head(7)
```

```
counted = counted.sort_values('totals', ascending=False)
```

```
counted.head(7)
```

```
Gender = medals[['Gender']]
```

```
Gender.head()
```

```
Gender_uniques = Gender.drop_duplicates()
```

```
Gender_uniques
```

```
medals_by_gender = medals.groupby(['Gender'])
medal_count_by_gender = medals_by_gender.count()
medal_count_by_gender
```

```
sus = (medals.Gender == 'W') & (medals.Gender == 'Men')
sus.head()
```

```
suspect = medals[sus]
suspect
```

```
print(len(medals['Sport'].unique()))
medals['Sport'].unique()
```

```
country_grouped = medals.groupby('Country')
Nsports = country_grouped['Sport'].nunique()
Nsports.head()
```

```
country_grouped['Sport'].count().sort_values(ascending = False).head()
```

```
count = medals.groupby('Country')['Medal'].count()
count.head()
```

```
count.index
```

```
count.idxmax()
```

```
count.idxmin()
```

```
import matplotlib.pyplot as plt

%matplotlib inline

usa = medals[medals.Country == 'USA']

usa_medals_by_year = usa.groupby(['Event', 'Medal'])['Athlete'].count()
usa_medals_by_year.head(10)

usa_medals_by_year = usa_medals_by_year.unstack(level='Medal')
usa_medals_by_year.head(10)

usa_medals_by_year.plot()
plt.show()

usa = medals[medals.Country == 'USA']

usa_medals_by_year = usa.groupby(['Event', 'Medal'])['Athlete'].count()

usa_medals_by_year = usa_medals_by_year.unstack(level='Medal')

usa_medals_by_year.plot.area(alpha=.5)
plt.show()

usa = medals[medals.Country == 'USA']

usa_medals_by_year = usa.groupby(['Year', 'Medal'])['Athlete'].count()

usa_medals_by_year = usa_medals_by_year.unstack(level='Medal')
```

```
usa_medals_by_year.plot.area(alpha=.5)
```

```
plt.show()
```

```
medals.Medal = pd.Categorical(values=medals.Medal, categories=['Bronze', 'Silver', 'Gold'],  
ordered=True)
```

```
medals.Medal.head()
```

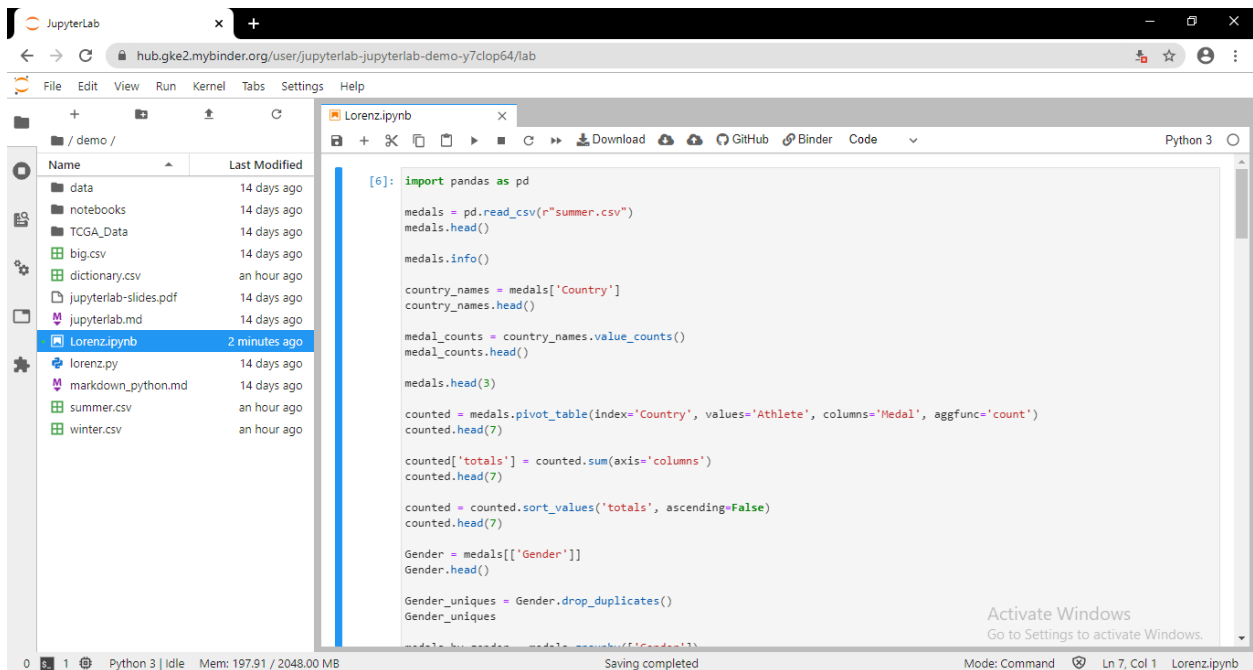
```
usa = medals[medals.Country == 'USA']
```

```
usa_medals_by_year = usa.groupby(['Event', 'Medal'])['Athlete'].count()
```

```
usa_medals_by_year = usa_medals_by_year.unstack(level='Medal')
```

```
usa_medals_by_year.plot.area(alpha=.6)
```

```
plt.show()
```



```
[6]: import pandas as pd  
  
medals = pd.read_csv(r"summer.csv")  
medals.head()  
  
medals.info()  
  
country_names = medals['Country']  
country_names.head()  
  
medal_counts = country_names.value_counts()  
medal_counts.head()  
  
medals.head(3)  
  
counted = medals.pivot_table(index='Country', values='Athlete', columns='Medal', aggfunc='count')  
counted.head(7)  
  
counted['totals'] = counted.sum(axis='columns')  
counted.head(7)  
  
counted = counted.sort_values('totals', ascending=False)  
counted.head(7)  
  
Gender = medals[['Gender']]  
Gender.head()  
  
Gender_uniques = Gender.drop_duplicates()  
Gender_uniques
```

JupyterLab interface showing a file browser on the left and a code editor on the right. The file browser displays a directory structure with files like data, notebooks, TCGA_Data, big.csv, dictionary.csv, jupyterlab-slides.pdf, jupyterlab.md, Lorenz.ipynb, lorenz.py, markdown_python.md, summer.csv, and winter.csv. The code editor shows Python code for data analysis using pandas and matplotlib.

```
Gender_uniques
medals_by_gender = medals.groupby(['Gender'])
medal_count_by_gender = medals_by_gender.count()
medal_count_by_gender

sus = (medals.Gender == 'W') & (medals.Gender == 'Men')
sus.head()

suspect = medals[sus]
suspect

print(len(medals['Sport'].unique()))
medals['Sport'].unique()

country_grouped = medals.groupby('Country')
Nsports = country_grouped['Sport'].nunique()
Nsports.head()

country_grouped['Sport'].count().sort_values(ascending = False).head()

count = medals.groupby('Country')['Medal'].count()
count.head()

count.index
count.idxmax()
count.idxmin()

import matplotlib.pyplot as plt
%matplotlib inline
```

Activate Windows
Go to Settings to activate Windows.

0 1 Python 3 | idle Mem: 197.91 / 2048.00 MB Saving completed Mode: Command Ln 7, Col 1 Lorenz.ipynb

JupyterLab interface showing a file browser on the left and a code editor on the right. The file browser displays a directory structure with files like data, notebooks, TCGA_Data, big.csv, dictionary.csv, jupyterlab-slides.pdf, jupyterlab.md, Lorenz.ipynb, lorenz.py, markdown_python.md, summer.csv, and winter.csv. The code editor shows Python code for data analysis using pandas and matplotlib.

```
import matplotlib.pyplot as plt
%matplotlib inline

usa = medals[medals.Country == 'USA']

usa_medals_by_year = usa.groupby(['Event', 'Medal'])['Athlete'].count()
usa_medals_by_year.head(10)

usa_medals_by_year = usa_medals_by_year.unstack(level='Medal')
usa_medals_by_year.head(10)

usa_medals_by_year.plot()
plt.show()

usa = medals[medals.Country == 'USA']

usa_medals_by_year = usa.groupby(['Event', 'Medal'])['Athlete'].count()

usa_medals_by_year = usa_medals_by_year.unstack(level='Medal')

usa_medals_by_year.plot.area(alpha=.5)
plt.show()

usa = medals[medals.Country == 'USA']

usa_medals_by_year = usa.groupby(['Year', 'Medal'])['Athlete'].count()

usa_medals_by_year = usa_medals_by_year.unstack(level='Medal')

usa_medals_by_year.plot.area(alpha=.5)
plt.show()
```

Activate Windows
Go to Settings to activate Windows.

0 1 Python 3 | idle Mem: 197.91 / 2048.00 MB Saving completed Mode: Command Ln 7, Col 1 Lorenz.ipynb

JupyterLab interface showing a file browser on the left and a code editor on the right. The file browser displays a directory structure with files like data, notebooks, TCGA_Data, big.csv, dictionary.csv, jupyterlab-slides.pdf, jupyterlab.md, Lorenz.ipynb, lorenz.py, markdown_python.md, summer.csv, and winter.csv. The code editor shows Python code for data analysis and visualization:

```
usa_medals_by_year = usa_medals_by_year.unstack(level='Medal')
usa_medals_by_year.plot.area(alpha=.5)
plt.show()

medals.Medal = pd.Categorical(values=medals.Medal, categories=['Bronze', 'Silver', 'Gold'], ordered=True)
medals.Medal.head()

usa = medals[medals.Country == 'USA']
usa_medals_by_year = usa.groupby(['Event', 'Medal']).count()
usa_medals_by_year = usa_medals_by_year.unstack(level='Medal')
usa_medals_by_year.plot.area(alpha=.6)
plt.show()
```

The output shows the DataFrame structure:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31165 entries, 0 to 31164
Data columns (total 9 columns):
Year      31165 non-null int64
City      31165 non-null object
Sport     31165 non-null object
Discipline 31165 non-null object
Athlete   31165 non-null object
Country   31161 non-null object
Gender    31165 non-null object
Event     31165 non-null object
Medal     31165 non-null object
dtypes: int64(1), object(8)
memory usage: 2.1+ MB
```

At the bottom, it says "Activate Windows. Go to Settings to activate Windows."

