Mini Project – 1

Name : Umesh Padul

Roll No. : COTB45

import matplotlib.pyplot as plt

import seaborn as sns

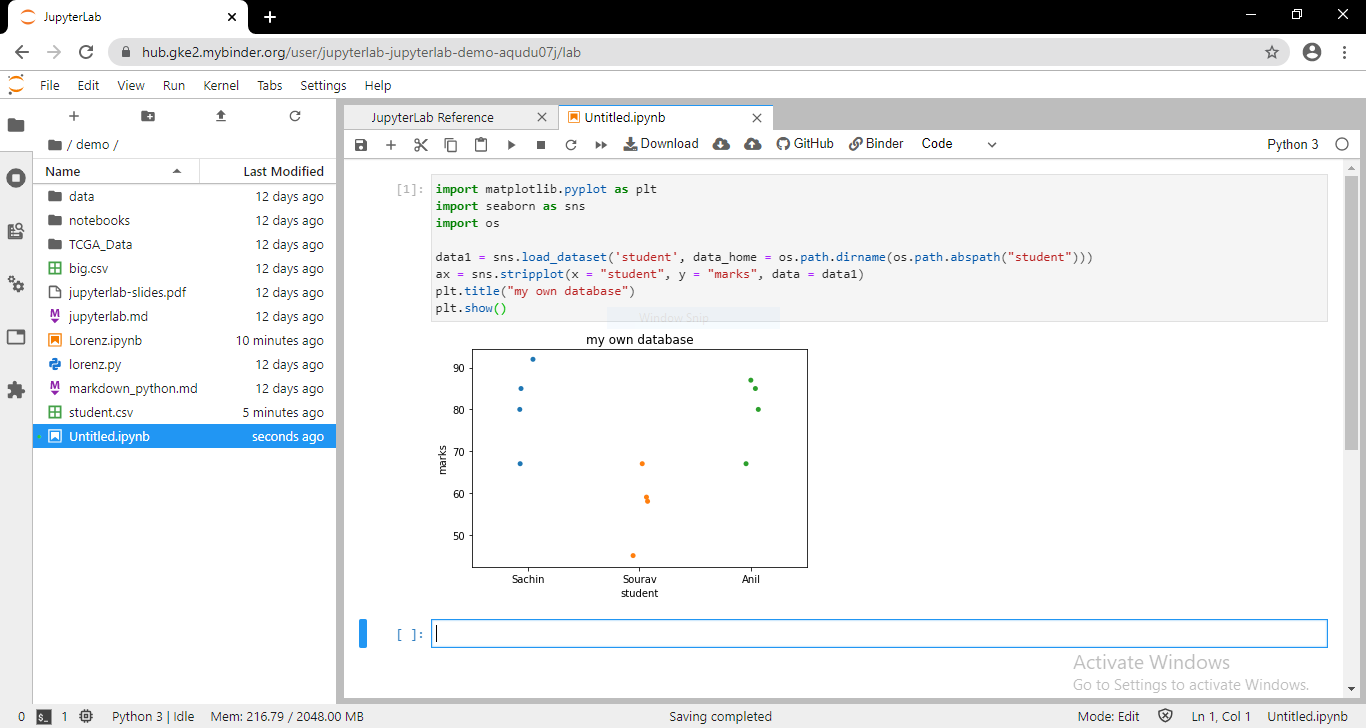
import os

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

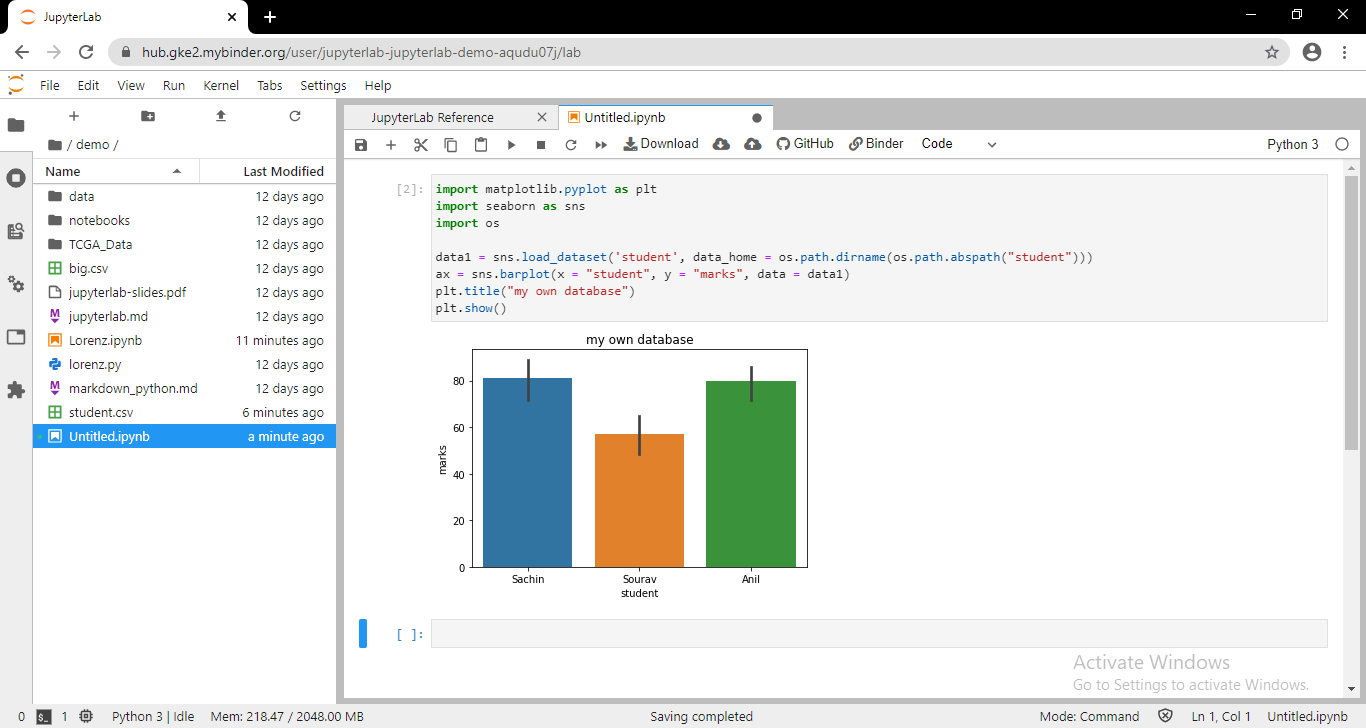
import os

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()

