

1.In this question 1, we will use attached dataset as FDA.csv.

Find the output of the following sections:

i)Write a syntax to load given data

Answer:

```
import pandas as pd
```

```
data = pd.read_csv("FDA.csv")
```

```
data.head()
```

OUTPUT

The screenshot shows a Jupyter Notebook window titled "Untitled - Jupyter Notebook". The toolbar includes standard options like File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Run, and Cell Type. A Python 3 logo indicates the kernel. The code cell (In [2]) contains three lines of Python code to import pandas, read the FDA.csv file, and display its head. The output cell (Out[2]) displays the first five rows of the dataset as a Pandas DataFrame. The columns are: RowNumber, CustomerId, Surname, CreditScore, Geography, Gender, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, and EstimatedSalary. The data includes various demographic and financial details for different customers.

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary
0	1	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88
1	2	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58
2	3	Orilo	502	France	Female	42	8	159660.80	3	1	0	113931.57
3	4	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63
4	5	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10

ii) Summarize the data

Answer:

```
import pandas as pd
```

```
data = pd.read_csv("FDA.csv")
```

```
data.describe()
```

OUTPUT

The screenshot shows a Jupyter Notebook window titled "Untitled - Jupyter Notebook". The code in cell In [3] is:

```
In [3]: 1 import pandas as pd
2 data = pd.read_csv("FDA.csv")
3 data.describe()
```

The output in Out[3] is a DataFrame showing descriptive statistics for the columns:

	RowNumber	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	...
count	10000.00000	1.00000e+04	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000
mean	5000.50000	1.569094e+07	650.528800	38.921800	5.012800	76485.889288	1.530200	0.70550	0.515100	100090.239881	...
std	2886.89568	7.193619e+04	96.653299	10.487806	2.892174	62397.405202	0.581654	0.45584	0.499797	57510.492818	...
min	1.00000	1.556570e+07	350.000000	18.000000	0.000000	0.000000	1.000000	0.00000	0.000000	11.580000	...
25%	2500.75000	1.562853e+07	584.000000	32.000000	3.000000	0.000000	1.000000	0.00000	0.000000	51002.110000	...
50%	5000.50000	1.569074e+07	652.000000	37.000000	5.000000	97198.540000	1.000000	1.00000	1.000000	100193.915000	...
75%	7500.25000	1.575323e+07	718.000000	44.000000	7.000000	127644.240000	2.000000	1.00000	1.000000	149388.247500	...
max	10000.00000	1.581569e+07	850.000000	92.000000	10.000000	250898.090000	4.000000	1.00000	1.000000	199992.480000	...

The Windows taskbar at the bottom shows the search bar, pinned icons for File Explorer, Microsoft Edge, and others, and the system tray with date and time.

iii) To select columns CreditScore and Balance.

Answer:

```
import pandas as pd
data = pd.read_csv("FDA.csv")
first = data[["CreditScore","Balance"]]
first
```

OUTPUT

The screenshot shows a Jupyter Notebook interface running on a Windows desktop. The browser tab is titled "Untitled - Jupyter Notebook". The notebook cell In [6] contains Python code to import pandas and read a CSV file, followed by a command to select specific rows. The resulting output, Out[6], is a table with columns "CreditScore" and "Balance". The table has 9999 rows, with the first few rows shown.

```
In [6]:  
1 import pandas as pd  
2 data = pd.read_csv("FDA.csv")  
3 first = data[["CreditScore", "Balance"]]  
4 first  
Out[6]:  
   CreditScore  Balance  
0            619     0.00  
1            608  83807.86  
2            502 159660.80  
3            699     0.00  
4            850 125510.82  
...          ...    ...  
9995           771     0.00  
9996           516  57369.61  
9997           709     0.00  
9998           772  75075.31  
9999           792 130142.79
```

iv) To select row Bearce, Henderson and Lombardo with all columns in a data.

Answer:

```
import pandas as pd  
data = pd.read_csv("FDA.csv",index_col ="Surname")  
first = data.loc[["Bearce","Henderson","Lombardo"]]  
first
```

OUTPUT

Jupyter Notebook - Untitled

localhost:8888/notebooks/19BCD7143/Untitled.ipynb

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

jupyter Untitled Last Checkpoint: 05/05/2021 (unsaved changes) Logout

RowNumber CustomerId CreditScore Geography Gender Age Tenure Balance NumOfProducts HasCrCard IsActiveMember EstimatedSalary

Surname												
Bearce	11	15767821	528	France	Male	31	6	102016.72	2	0	0	80181.12
Henderson	18	15788218	549	Spain	Female	24	9	0.00	2	1	1	14406.41
Henderson	1092	15784092	732	France	Male	36	7	126195.81	1	1	1	133172.48
Henderson	1144	15605965	630	France	Male	43	9	0.00	2	1	1	34338.04
Henderson	2044	15642544	723	France	Male	34	5	0.00	2	0	1	12092.03
Henderson	2317	15624388	649	Germany	Female	50	5	155393.32	1	1	1	87351.42
Henderson	3283	15806901	584	France	Female	39	2	112687.69	1	1	1	127749.61
Henderson	4498	15758901	713	Spain	Female	47	1	0.00	1	1	0	107825.08
Henderson	5087	15667289	719	Spain	Male	50	2	0.00	2	0	0	10772.13
Henderson	5735	15596647	768	France	Male	54	8	69712.74	1	1	1	69381.05
Henderson	5753	15697948	752	Spain	Female	36	3	0.00	2	1	1	48505.10
Henderson	6428	15668943	746	France	Male	37	2	0.00	2	1	0	143194.05
Henderson	6513	15795429	487	France	Male	24	7	133628.09	2	1	1	98570.01
Henderson	6547	15633840	781	France	Male	20	0	125023.10	2	1	1	108301.45
Henderson	8478	15685947	556	Germany	Male	42	0	115915.53	2	0	1	125435.47
Henderson	9084	15711396	427	Spain	Male	40	8	0.00	2	1	1	82870.75

Type here to search 21:07 07-06-2021

Jupyter Notebook - Untitled

localhost:8888/notebooks/19BCD7143/Untitled.ipynb

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

jupyter Untitled Last Checkpoint: 05/05/2021 (unsaved changes) Logout

RowNumber CustomerId CreditScore Geography Gender Age Tenure Balance NumOfProducts HasCrCard IsActiveMember EstimatedSalary

Surname												
Henderson	8478	15685947	556	Germany	Male	42	0	115915.53	2	0	1	125435.47
Henderson	9084	15711396	427	Spain	Male	40	8	0.00	2	1	1	82870.75
Henderson	9139	15771668	578	France	Male	59	10	185966.64	1	0	0	9445.42
Lombardo	36	15794171	475	France	Female	45	0	134264.04	1	1	0	27822.99
Lombardo	1728	15734491	676	Spain	Female	36	4	0.00	2	1	1	3173.31
Lombardo	2577	15673529	645	Spain	Male	36	4	59893.85	2	1	0	43999.64
Lombardo	2897	15606836	782	France	Female	33	2	94493.03	1	0	1	101866.39
Lombardo	3769	15685920	599	Spain	Male	34	2	101506.66	1	0	0	198030.24
Lombardo	5238	15702566	554	Spain	Male	26	8	149134.46	1	1	1	177966.24
Lombardo	5360	15801832	684	Germany	Male	42	1	117691.00	1	1	1	23135.65
Lombardo	5884	15574795	495	France	Female	38	2	63093.01	1	1	1	47089.72
Lombardo	5885	15706036	552	Germany	Male	38	10	132271.12	2	1	1	46562.02
Lombardo	5919	15742609	600	Germany	Male	28	2	116623.31	1	0	1	59905.29
Lombardo	8685	15783097	813	Germany	Male	27	6	111348.15	1	1	0	46422.46
Lombardo	8834	15760873	594	France	Male	50	7	81310.34	1	1	1	183868.01
Lombardo	9309	15728683	742	France	Male	27	0	0.00	2	0	1	131534.96
Lombardo	9390	15756954	538	France	Female	32	2	0.00	1	1	1	80130.54
Lombardo	9425	15666200	689	France	Female	40	1	0.00	2	1	1	119446.64

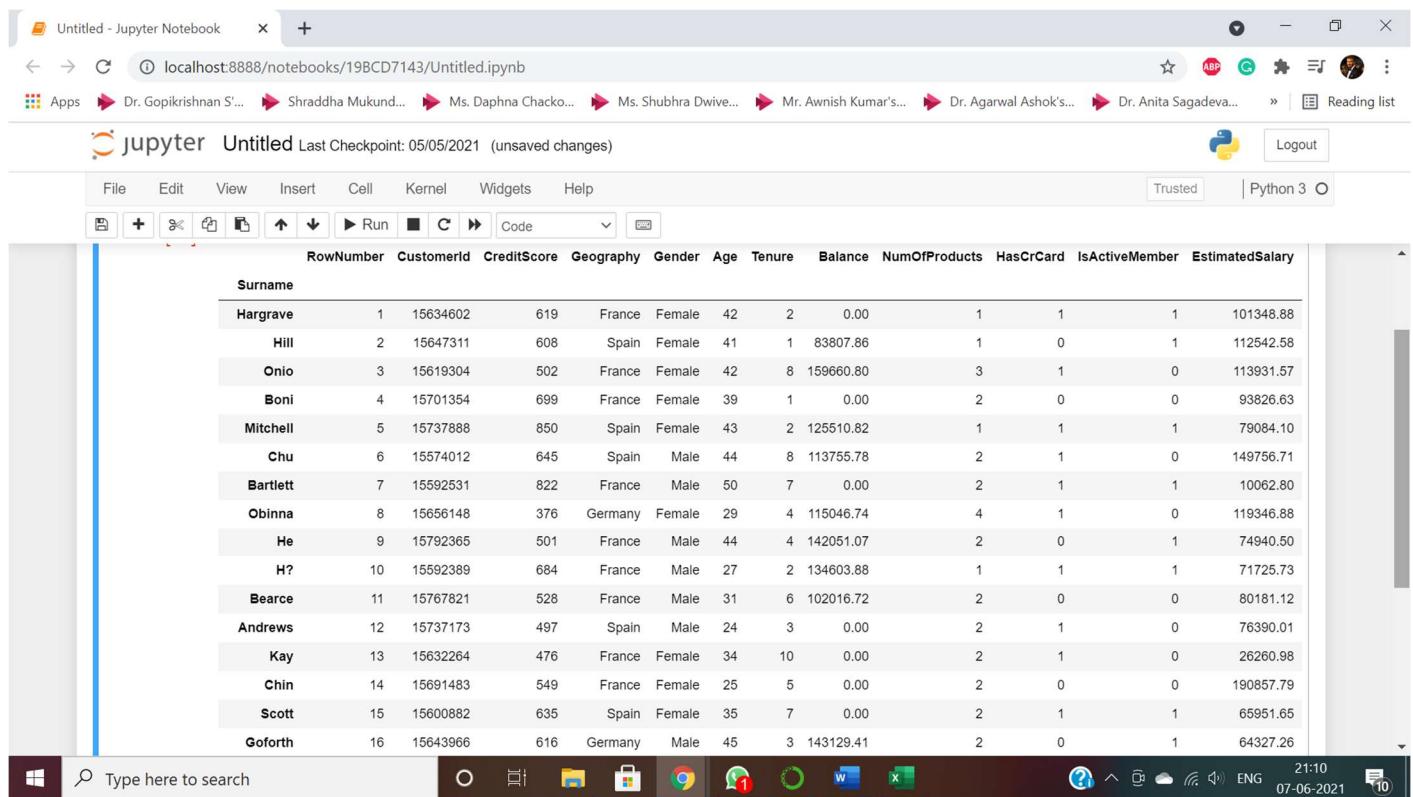
Type here to search 21:07 07-06-2021

v) print the first 20 rows of data

Answer:

```
import pandas as pd  
  
data = pd.read_csv("FDA.csv",index_col ="Surname")  
  
data.head(20)
```

OUTPUT



Surname	RowNumber	CustomerId	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary
Hargrave	1	15634602	619	France	Female	42	2	0.00	1	1	1	101348.88
Hill	2	15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58
Onio	3	15619304	502	France	Female	42	8	159660.80	3	1	0	113931.57
Boni	4	15701354	699	France	Female	39	1	0.00	2	0	0	93826.63
Mitchell	5	15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.10
Chu	6	15574012	645	Spain	Male	44	8	113755.78	2	1	0	149756.71
Bartlett	7	15592531	822	France	Male	50	7	0.00	2	1	1	10062.80
Obinna	8	15656148	376	Germany	Female	29	4	115046.74	4	1	0	119346.88
He	9	15792365	501	France	Male	44	4	142051.07	2	0	1	74940.50
H?	10	15592389	684	France	Male	27	2	134603.88	1	1	1	71725.73
Bearce	11	15767821	528	France	Male	31	6	102016.72	2	0	0	80181.12
Andrews	12	15737173	497	Spain	Male	24	3	0.00	2	1	0	76390.01
Kay	13	15632264	476	France	Female	34	10	0.00	2	1	0	26260.98
Chin	14	15691483	549	France	Female	25	5	0.00	2	0	0	190857.79
Scott	15	15600882	635	Spain	Female	35	7	0.00	2	1	1	65951.65
Goforth	16	15643966	616	Germany	Male	45	3	143129.41	2	0	1	64327.26

vi) Sort the data in descending order of Age.

Answer:

```
import pandas as pd  
data = pd.read_csv("FDA.csv")  
data.sort_values(by=['Age'], ascending=False)
```

OUTPUT

The screenshot shows a Jupyter Notebook window titled "Untitled - Jupyter Notebook". The code cell (In [47]) contains three lines of Python code to read a CSV file and sort it by age in descending order. The output cell (Out[47]) displays a table with 10,000 rows and 14 columns, including columns for RowNumber, CustomerId, Surname, CreditScore, Geography, Gender, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, and Estimated. The table shows data for various customers across different countries like France, Germany, and Spain, with ages ranging from 18 to 92.

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	Estimated	
6443	6444	15764927	Rogova	753	France	Male	92	3	121513.31	1	0	1	195
6759	6760	15660878	T'len	705	France	Male	92	1	126076.24	2	1	1	34
2458	2459	15813303	Rearick	513	Spain	Male	88	10	0.00	2	1	1	52
3033	3034	15578006	Yao	787	France	Female	85	10	0.00	2	1	1	116
3387	3388	15798024	Lori	537	Germany	Male	84	8	92242.34	1	1	1	186
...
9782	9783	15728829	Weigel	509	France	Male	18	7	102983.91	1	1	0	171
2141	2142	15758372	Wallace	674	France	Male	18	7	0.00	2	1	1	55
9501	9502	15634146	Hou	835	Germany	Male	18	2	142872.36	1	1	1	117
9520	9521	15673180	Onyekaozulu	727	Germany	Female	18	2	93816.70	2	1	0	126
1619	1620	15770309	McDonald	656	France	Male	18	10	151762.74	1	0	1	127

2.

- a) Draw the number of customers in each country in given data based on bar plot.**
- b) In the bar plot adjust x limit and y limit in the range of 0 to 7000 by gap 1000.**
- c) Draw a notched box plot of the CreditScore.**

Answer:

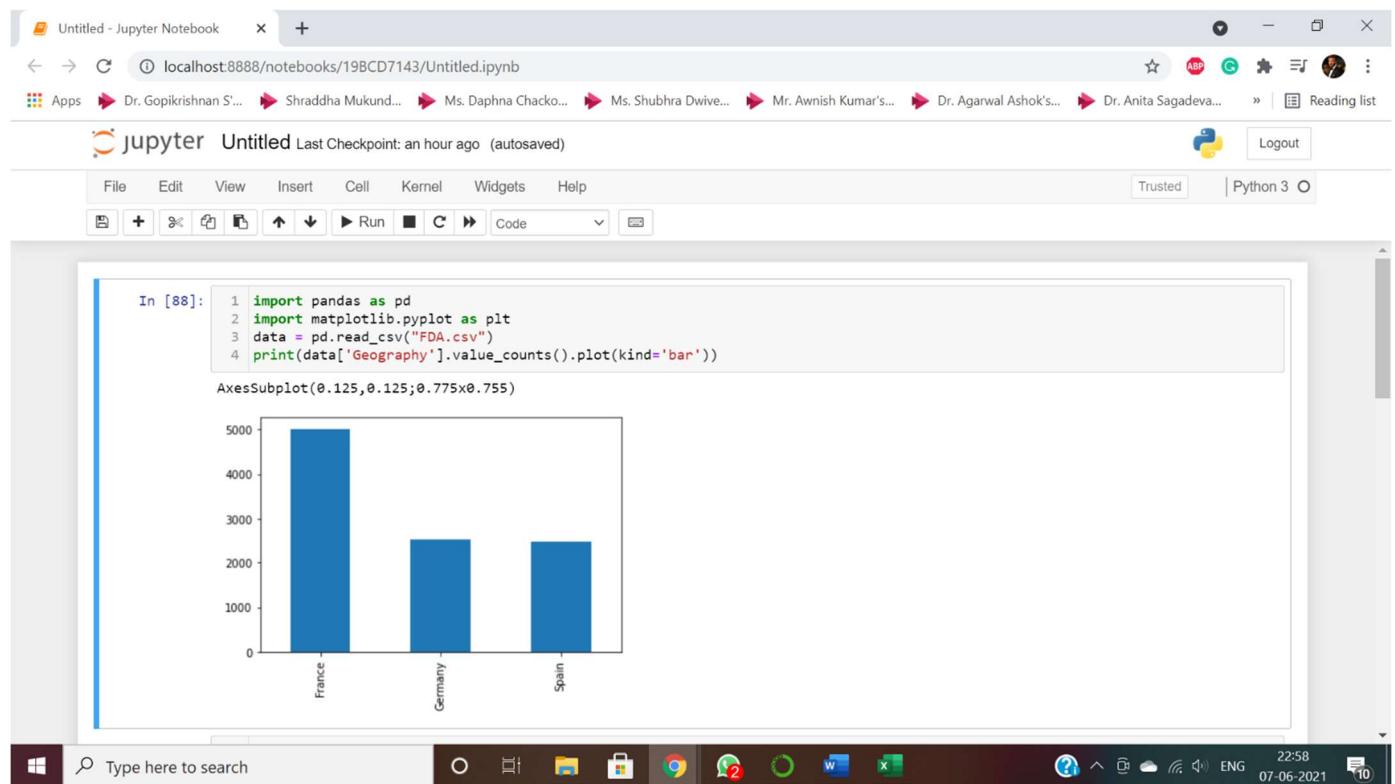
a) import pandas as pd

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

```
data = pd.read_csv("FDA.csv")
```

```
print(data['Geography'].value_counts().plot(kind='bar'))
```

OUTPUT



The screenshot shows a Jupyter Notebook interface. In the code cell (In [88]), the following Python code is run:

```
In [88]: 1 import pandas as pd
2 import matplotlib.pyplot as plt
3 data = pd.read_csv("FDA.csv")
4 print(data['Geography'].value_counts().plot(kind='bar'))
```

Below the code, the output is displayed as a bar chart titled "AxesSubplot(0.125,0.125;0.775x0.755)". The y-axis ranges from 0 to 5000 with increments of 1000. The x-axis categories are France, Germany, and Spain. The bars show approximately 5000 for France, 2500 for Germany, and 2500 for Spain.

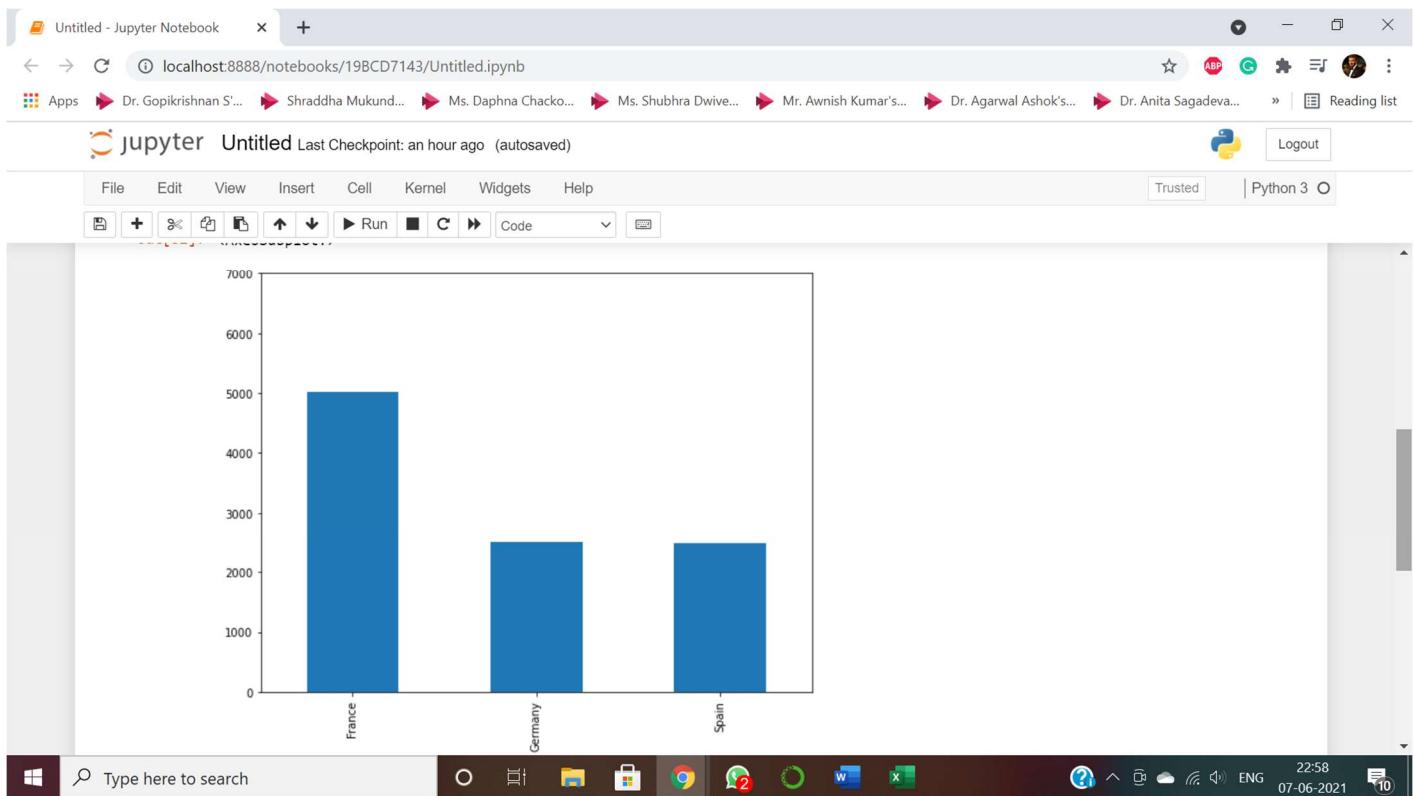
b) import pandas as pd

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

```
data = pd.read_csv("FDA.csv")
```

```
data['Geography'].value_counts().plot(kind='bar', ylim=(0,7000), figsize=(9,7))
```

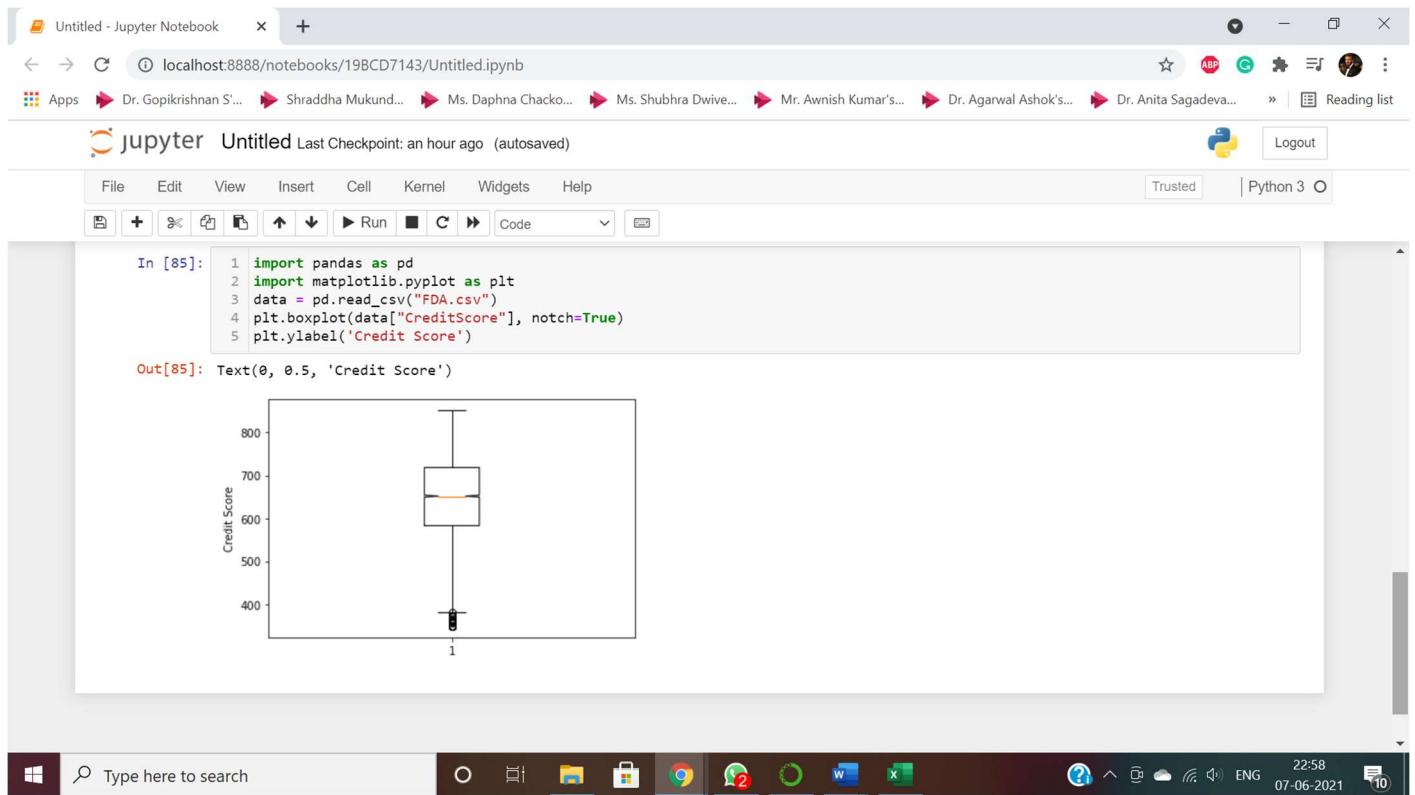
OUTPUT



c) import pandas as pd

```
import matplotlib.pyplot as plt  
data = pd.read_csv("FDA.csv")  
plt.boxplot(data["CreditScore"], notch=True)  
plt.ylabel('Credit Score')
```

OUTPUT



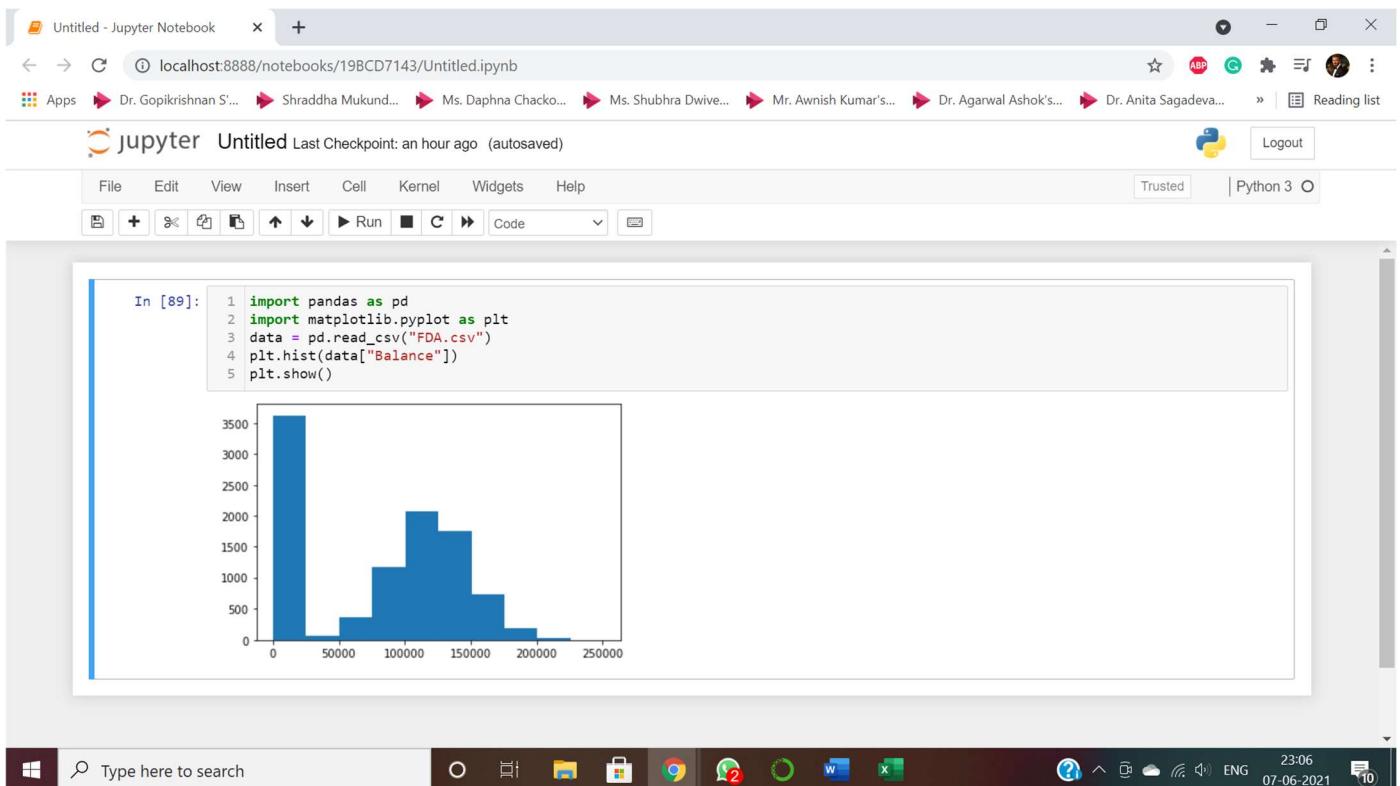
3.

a) Draw the histogram of customer Balance attribute in given data.

Answer:

```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("FDA.csv")
plt.hist(data["Balance"])
plt.show()
```

OUTPUT

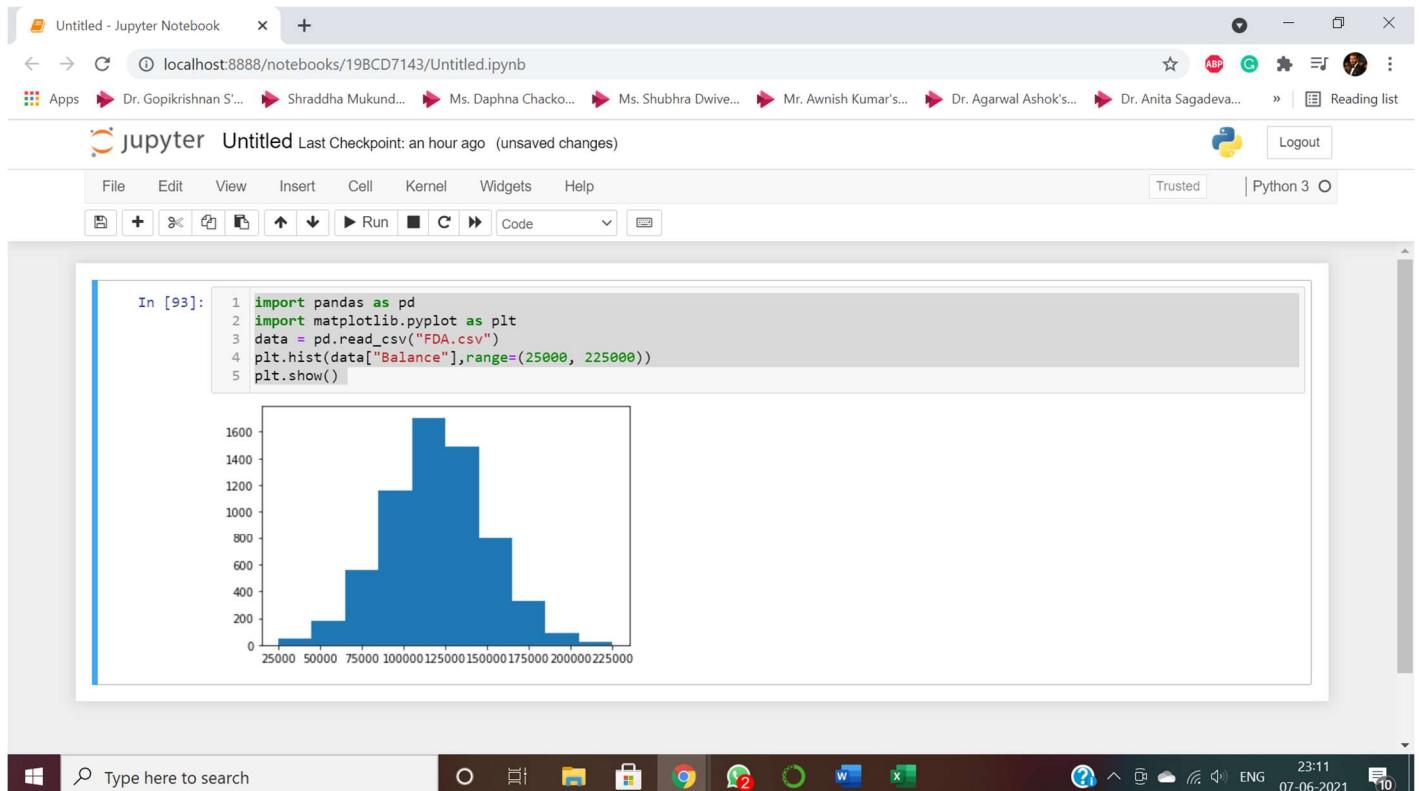


b)Draw the Customizing histogram of customer on range = (25000, 225000) by Balance attribute in given data.

Answer:

```
import pandas as pd  
import matplotlib.pyplot as plt  
data = pd.read_csv("FDA.csv")  
plt.hist(data["Balance"],range=(25000, 225000))  
plt.show()
```

OUTPUT



The screenshot shows a Jupyter Notebook window titled "Untitled - Jupyter Notebook". The URL is "localhost:8888/notebooks/19BCD7143/Untitled.ipynb". The notebook interface includes a toolbar with various icons for file operations, a menu bar with File, Edit, View, Insert, Cell, Kernel, Widgets, Help, and a code editor with a Python 3 kernel. A Python code cell (In [93]) contains the following code:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 data = pd.read_csv("FDA.csv")
4 plt.hist(data["Balance"], range=(25000, 225000))
5 plt.show()
```

The output of the code is a histogram showing the distribution of 'Balance' values. The x-axis ranges from 25000 to 225000 with major ticks every 25000 units. The y-axis ranges from 0 to 1600 with major ticks every 200 units. The histogram has several peaks, with the highest peak occurring between 75000 and 100000, reaching approximately 1600.

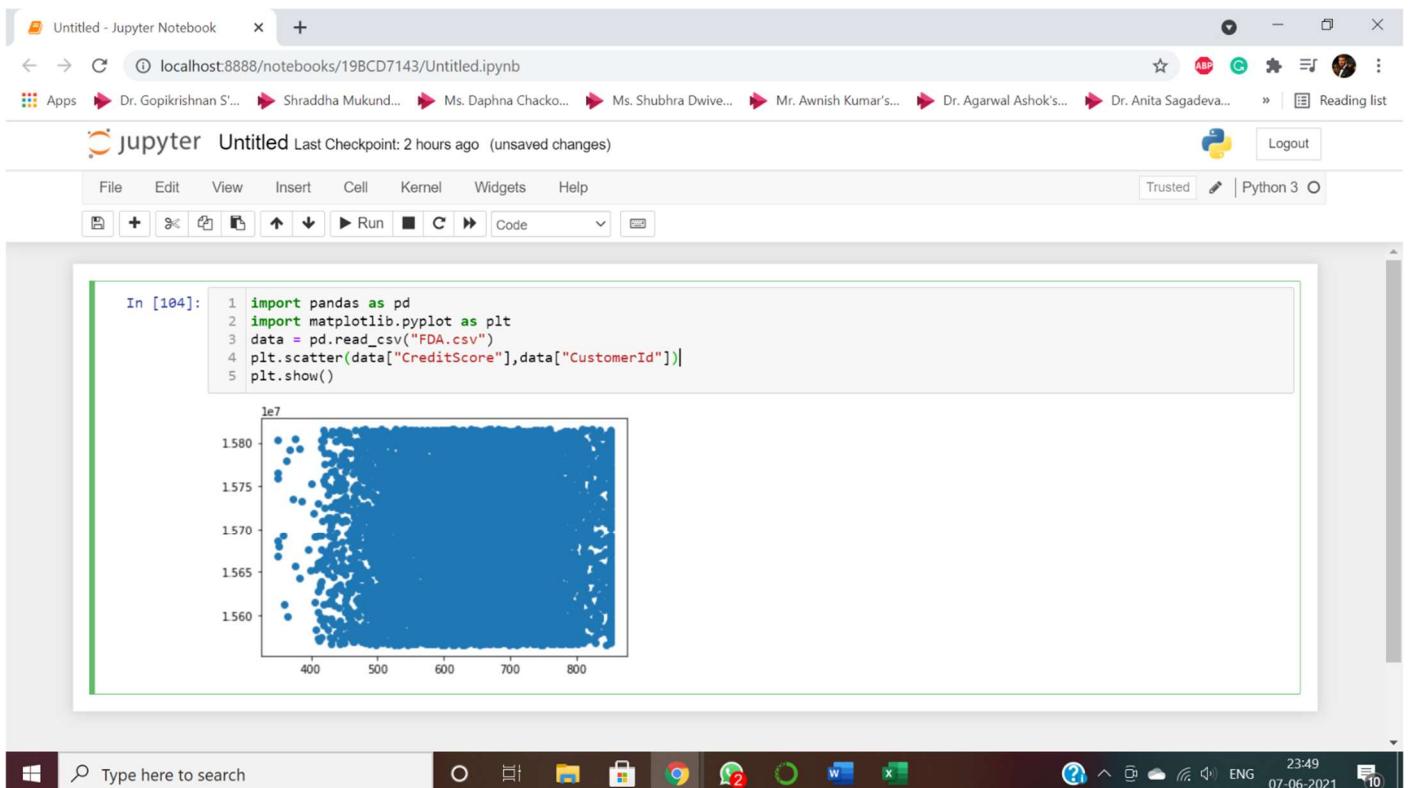
4.

a) Draw scatter plot of customer CreditScore attribute in given data.

Answer:

```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("FDA.csv")
plt.scatter(data["CreditScore"], data["CustomerId"])
plt.show()
```

OUTPUT



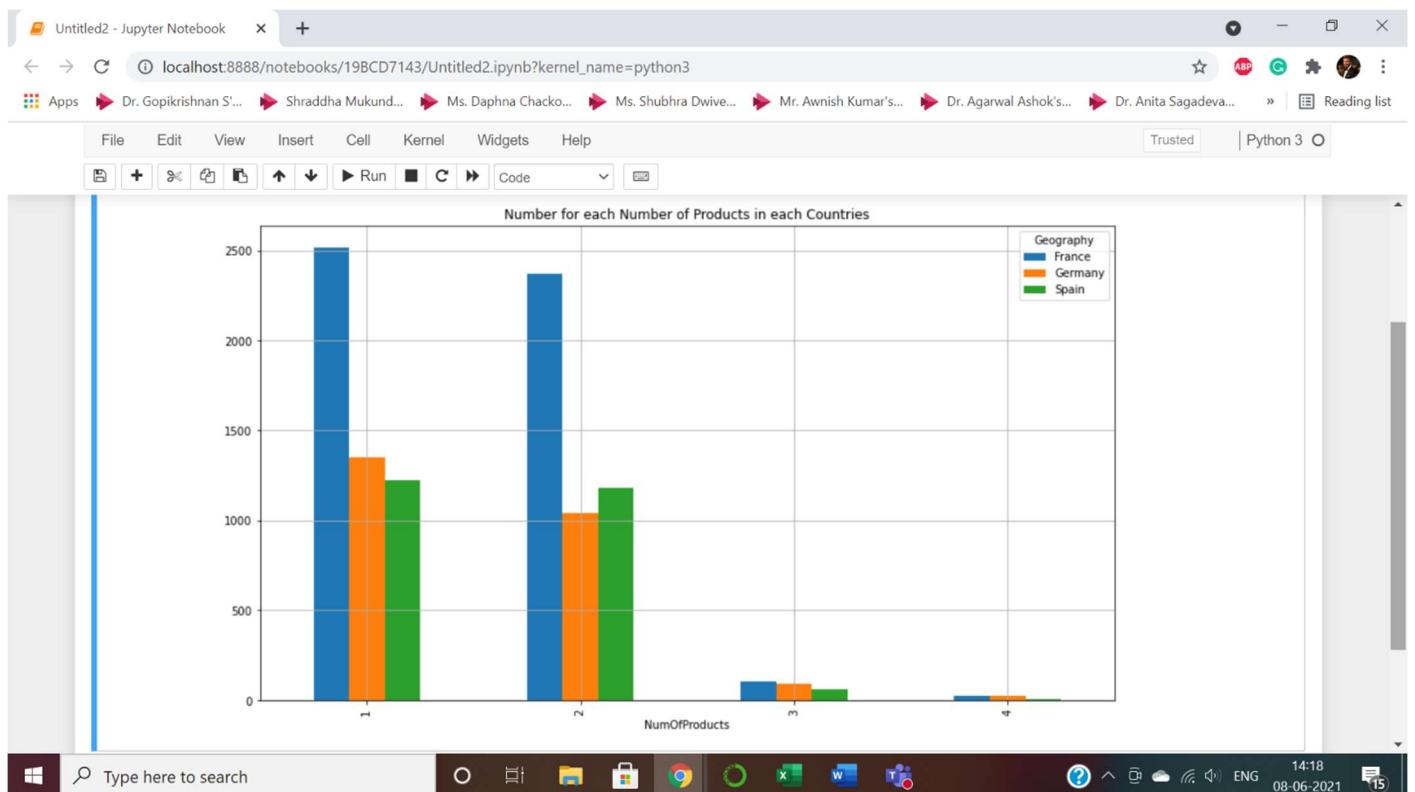
b) Draw grid of 2 columns and add bar plots attribute Countries and Number of Products.

Answer:

```
import pandas as pd  
import matplotlib.pyplot as plt  
data = pd.read_csv("FDA.csv")
```

```
ax =  
data.groupby(["Geography","NumOfProducts"])["NumOfProducts"].count().un  
stack(0).plot.bar(title="Number for each Number of Products in each  
Countries", figsize=(14,8))  
plt.grid()
```

OUTPUT



c)Draw Correlation Matrices.

Answer:

```
import pandas as pd  
import matplotlib.pyplot as plt  
data = pd.read_csv("FDA.csv")  
data.corr()
```

OUTPUT

The screenshot shows a Jupyter Notebook window titled "Untitled2 - Jupyter Notebook". The URL in the address bar is "localhost:8888/notebooks/19BCD7143/Untitled2.ipynb?kernel_name=python3". The top menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. On the right, there are buttons for Trusted and Python 3. Below the menu is a toolbar with various icons. The main area has two sections: "In [56]:" and "Out[56]:". The "In [56]" section contains the following Python code:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 data = pd.read_csv("FDA.csv")
4 data.corr()
```

The "Out[56]" section displays a correlation matrix for the dataset. The columns are labeled: RowNumber, CustomerId, CreditScore, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary, and Exited. The matrix values range from -0.016571 to 0.005988.

d) Use color #5cb85c for the Countries bar plot, color #5bc0de for the Number of Products bar plot.

Answer:

```
import pandas as pd
```

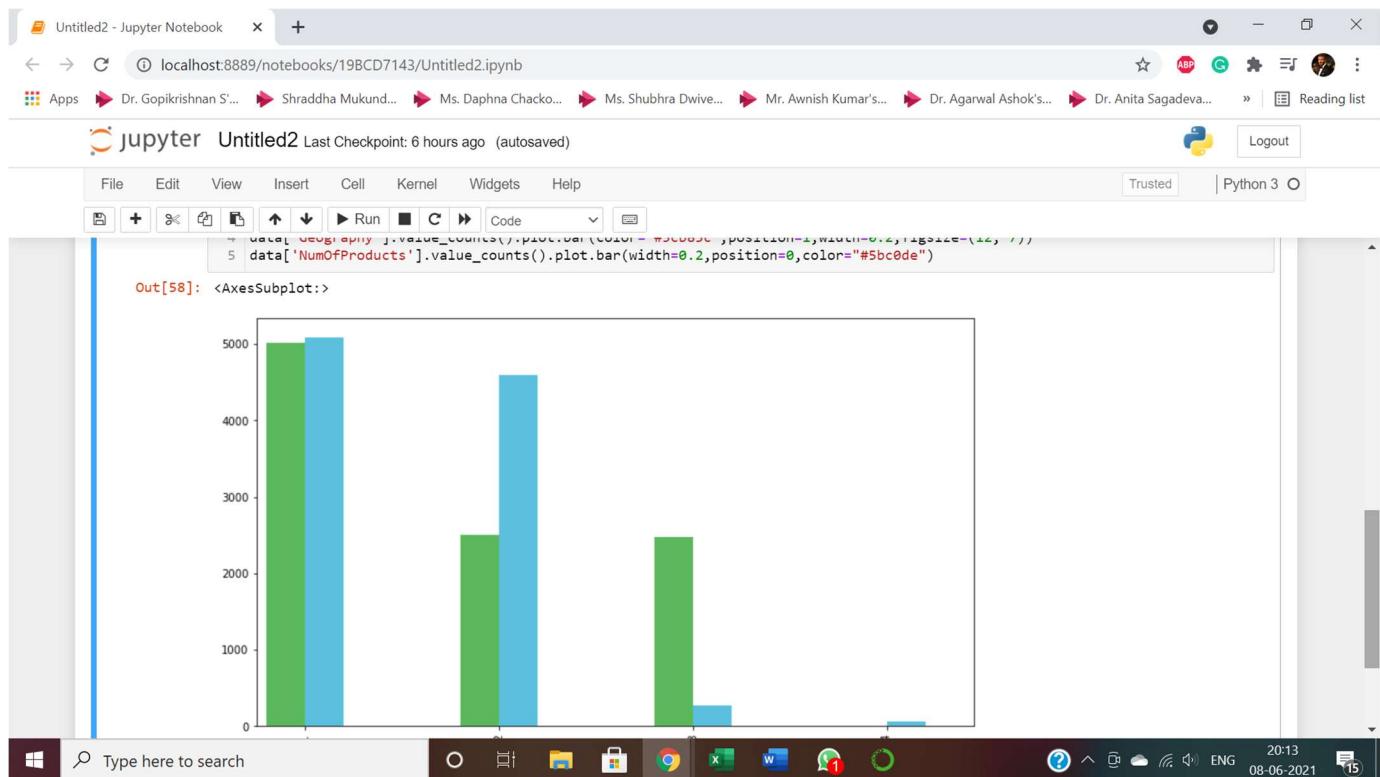
```
import matplotlib.pyplot as plt
```

```
data = pd.read_csv("FDA.csv")
```

```
data['Geography'].value_counts().plot.bar(color="#5cb85c", position=1, width=0.2, figsize=(12, 7))
```

```
data['NumOfProducts'].value_counts().plot.bar(width=0.2, position=0, color="#5bc0de")
```

OUTPUT



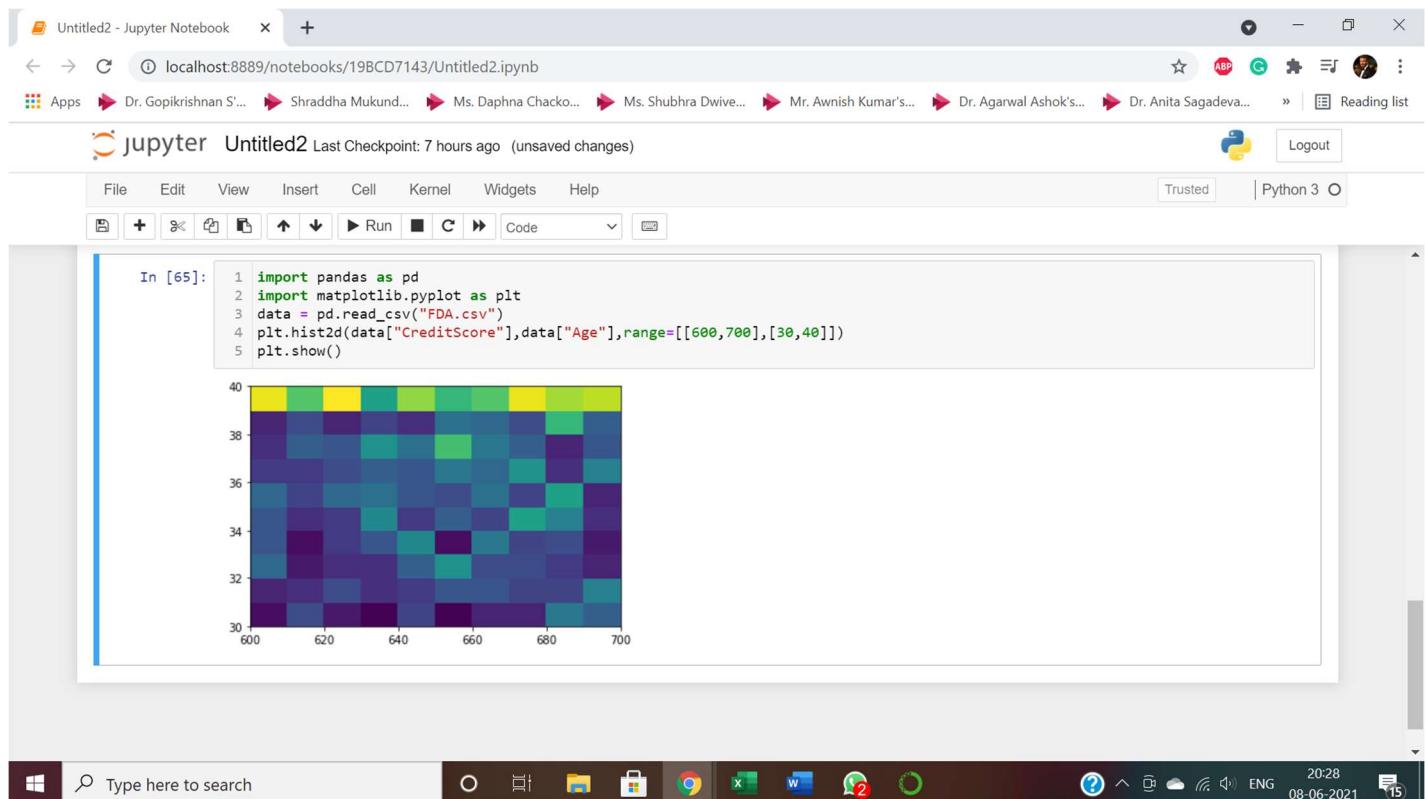
5.

a)Draw the Customizing 2 D histogram plot of CreditScore and Age attribute in given data based on between ages 30 and 40 and have credit scores between 600 and 700.

Answer:

```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("FDA.csv")
plt.hist2d(data["CreditScore"],data["Age"],range=[[600,700],[30,40]])
plt.show()
```

OUTPUT



A screenshot of a Jupyter Notebook interface. The title bar shows "Untitled2 - Jupyter Notebook". The toolbar includes standard options like File, Edit, View, Insert, Cell, Kernel, Widgets, Help, and a Trusted Python 3 button. Below the toolbar is a code cell labeled "In [65]:" containing the following Python code:

```
In [65]:  
1 import pandas as pd  
2 import matplotlib.pyplot as plt  
3 data = pd.read_csv("FDA.csv")  
4 plt.hist2d(data["CreditScore"], data["Age"], range=[[600,700],[30,40]])  
5 plt.show()
```

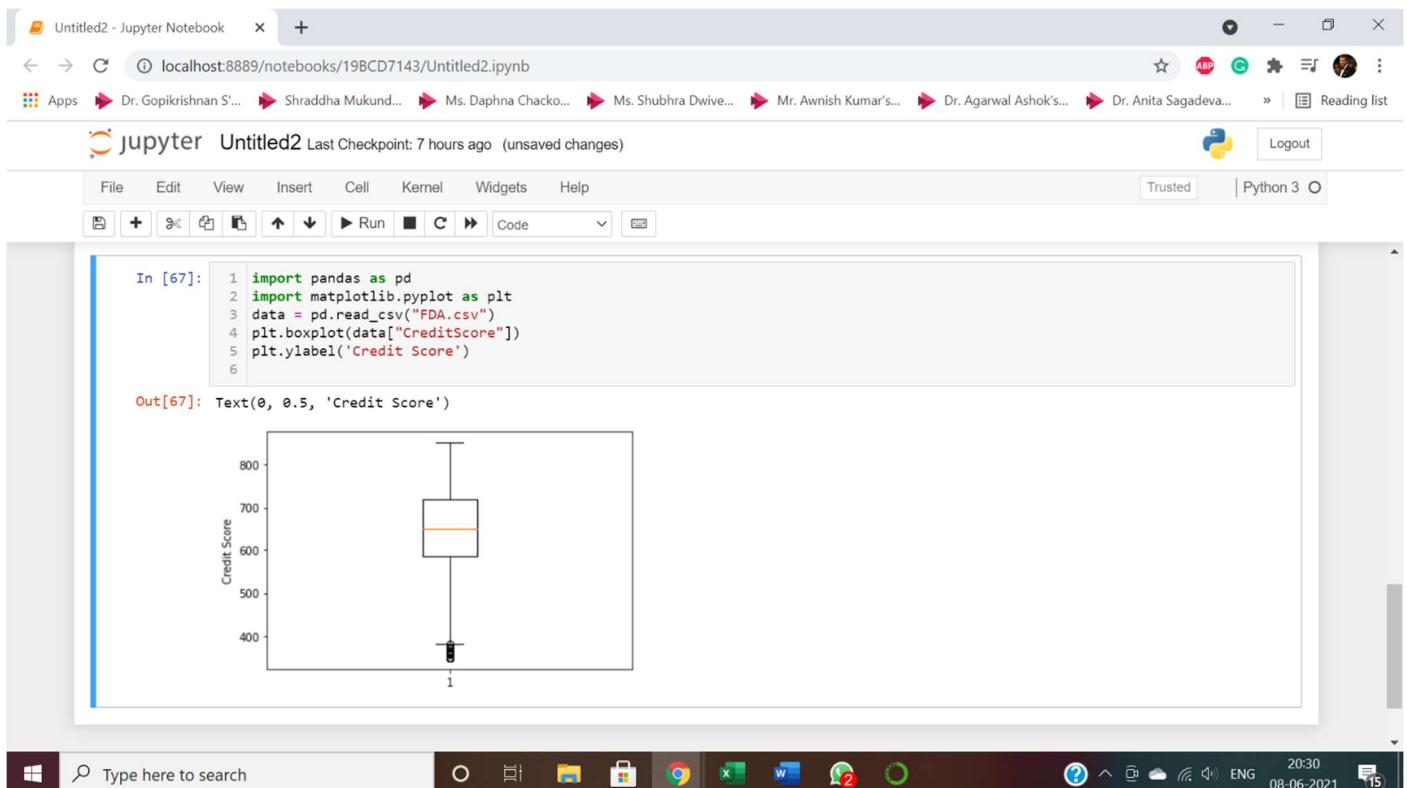
The output of the cell is a 2D histogram (heat map) showing the distribution of CreditScore (x-axis, 600 to 700) versus Age (y-axis, 30 to 40). The color scale ranges from dark purple (low density) to bright yellow (high density), indicating a positive correlation between age and credit score.

b) Draw the box plot of CreditScore attribute in given dataset.

Answer:

```
import pandas as pd  
import matplotlib.pyplot as plt  
data = pd.read_csv("FDA.csv")  
plt.boxplot(data["CreditScore"])  
plt.ylabel('Credit Score')
```

OUTPUT



c) Draw Pie Chart with Slice Percentages and Chart Legend of Geography attribute.

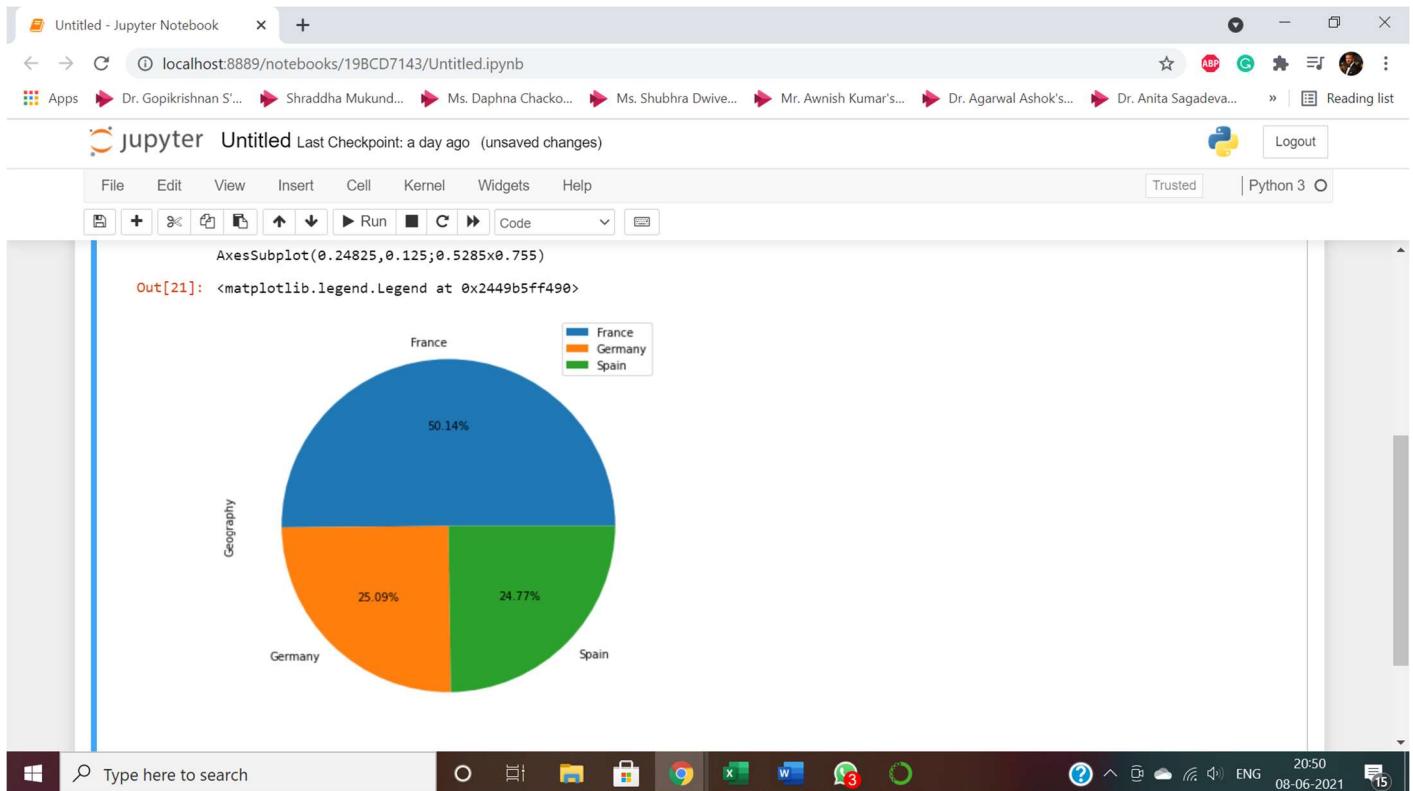
Answer:

```
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
data = pd.read_csv("FDA.csv")
```

```
print(data['Geography'].value_counts().plot(kind='pie', autopct='%.1f%%', figsize=(10, 7)))
```

`plt.legend()`

OUTPUT



6. In this question 6, we will use attached dataset as iris.csv. Find the output of the following sections:

i) Write a syntax to load given data

Answer:

```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("iris.csv")
data.head()
```

OUTPUT

The screenshot shows a Jupyter Notebook interface running on a Windows desktop. The notebook title is "Untitled1 - Jupyter Notebook". The code cell In [28] contains the following Python code:

```
In [28]:  
1 import pandas as pd  
2 import matplotlib.pyplot as plt  
3 data = pd.read_csv("iris.csv")  
4 data.head()
```

The output cell Out[28] displays the first five rows of the Iris dataset:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

The Windows taskbar at the bottom shows various open applications including File Explorer, Microsoft Edge, and Excel.

ii) Summarize the data

Answer:

```
import pandas as pd  
data = pd.read_csv("iris.csv")  
data.describe()
```

OUTPUT

The screenshot shows a Jupyter Notebook window titled "Untitled1 - Jupyter Notebook". The URL in the address bar is "localhost:8889/notebooks/19BCD7143/Untitled1.ipynb". The notebook has a single cell with the following Python code:

```
In [29]:  
1 import pandas as pd  
2 import matplotlib.pyplot as plt  
3 data = pd.read_csv("iris.csv")  
4 data.describe()
```

The output of the cell is labeled "Out[29]:" and displays a table of descriptive statistics for the "iris.csv" dataset:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

The Jupyter interface includes a toolbar with various icons for file operations, a menu bar with File, Edit, View, Insert, Cell, Kernel, Widgets, Help, and a status bar at the bottom showing the search bar, taskbar, and system tray.

iii) To select columns 'sepal_length', 'petal_length', 'petal_width'.

Answer:

```
import pandas as pd  
import matplotlib.pyplot as plt  
data = pd.read_csv("iris.csv")  
first = data[["SepalLengthCm", "PetalLengthCm", "PetalWidthCm"]]  
first
```

OUTPUT

The screenshot shows a Jupyter Notebook interface running on a Windows desktop. The notebook title is "Untitled1 - Jupyter Notebook". The code cell (In [30]) contains Python code to import pandas and matplotlib.pyplot, read an iris.csv file, and select specific columns. The resulting DataFrame (Out[30]) is displayed as a table with 150 rows and 3 columns: SepalLengthCm, PetalLengthCm, and PetalWidthCm. The table includes sample data points from index 0 to 149.

	SepalLengthCm	PetalLengthCm	PetalWidthCm
0	5.1	1.4	0.2
1	4.9	1.4	0.2
2	4.7	1.3	0.2
3	4.6	1.5	0.2
4	5.0	1.4	0.2
...
145	6.7	5.2	2.3
146	6.3	5.0	1.9
147	6.5	5.2	2.0
148	6.2	5.4	2.3
149	5.9	5.1	1.8

iv) To select row 2, 3 and 5 with all columns in a data.

Answer:

```
import pandas as pd  
import matplotlib.pyplot as plt  
data = pd.read_csv("iris.csv")  
data.iloc[[2,3,5]]
```

OUTPUT

The screenshot shows a Jupyter Notebook window titled "Untitled1 - Jupyter Notebook". The URL in the address bar is "localhost:8889/notebooks/19BCD7143/Untitled1.ipynb". The notebook interface includes a toolbar with File, Edit, View, Insert, Cell, Kernel, Widgets, Help, and a Trusted Python 3 button. Below the toolbar are buttons for cell creation, execution, and other functions. The main area displays three input cells (In [1], In [2], In [3]) and one output cell (Out[33]). The code in Out[33] imports pandas and matplotlib, reads an iris.csv file, and prints the first few rows of the dataset.

```
In [1]: 1
In [2]: 1
In [3]: 1
In [33]: 1 import pandas as pd
           2 import matplotlib.pyplot as plt
           3 data = pd.read_csv("iris.csv")
           4 data.iloc[[2,3,5]]
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa

v)print the first 10 rows of data

Answer:

```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("iris.csv")
data.head(10)
```

OUTPUT

The screenshot shows a Jupyter Notebook interface running on a Windows desktop. The notebook title is "Untitled1 - Jupyter Notebook". The code cell (In [34]) contains the following Python code:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 data = pd.read_csv("iris.csv")
4 data.head(10)
```

The output cell (Out[34]) displays the first 10 rows of the Iris dataset as a Pandas DataFrame:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa

vi) Sort the data in descending order of PetalLength.

Answer:

```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("iris.csv")
data.sort_values(by=['PetalLengthCm'], ascending=False)
```

OUTPUT

The screenshot shows a Jupyter Notebook interface running on localhost:8889/notebooks/19BCD7143/Untitled1.ipynb. The code cell In [35] contains Python code to import pandas and matplotlib.pyplot, read the iris.csv file, and sort the data by PetalLengthCm in descending order. The resulting DataFrame, Out[35], displays 150 rows of data with columns: Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm, and Species. The data includes entries for Iris-virginica, Iris-setosa, and Iris-virginica. The Jupyter interface includes a toolbar with various icons for file operations, a navigation bar with user profiles, and a status bar at the bottom showing the date and time.

```
In [35]: 1 import pandas as pd
2 import matplotlib.pyplot as plt
3 data = pd.read_csv("iris.csv")
4 data.sort_values(by=['PetalLengthCm'], ascending=False)
```

118	119	7.7	2.6	6.9	2.3 Iris-virginica
122	123	7.7	2.8	6.7	2.0 Iris-virginica
117	118	7.7	3.8	6.7	2.2 Iris-virginica
105	106	7.6	3.0	6.6	2.1 Iris-virginica
131	132	7.9	3.8	6.4	2.0 Iris-virginica
...
36	37	5.5	3.5	1.3	0.2 Iris-setosa
35	36	5.0	3.2	1.2	0.2 Iris-setosa
14	15	5.8	4.0	1.2	0.2 Iris-setosa
13	14	4.3	3.0	1.1	0.1 Iris-setosa
22	23	4.6	3.6	1.0	0.2 Iris-setosa

150 rows × 6 columns

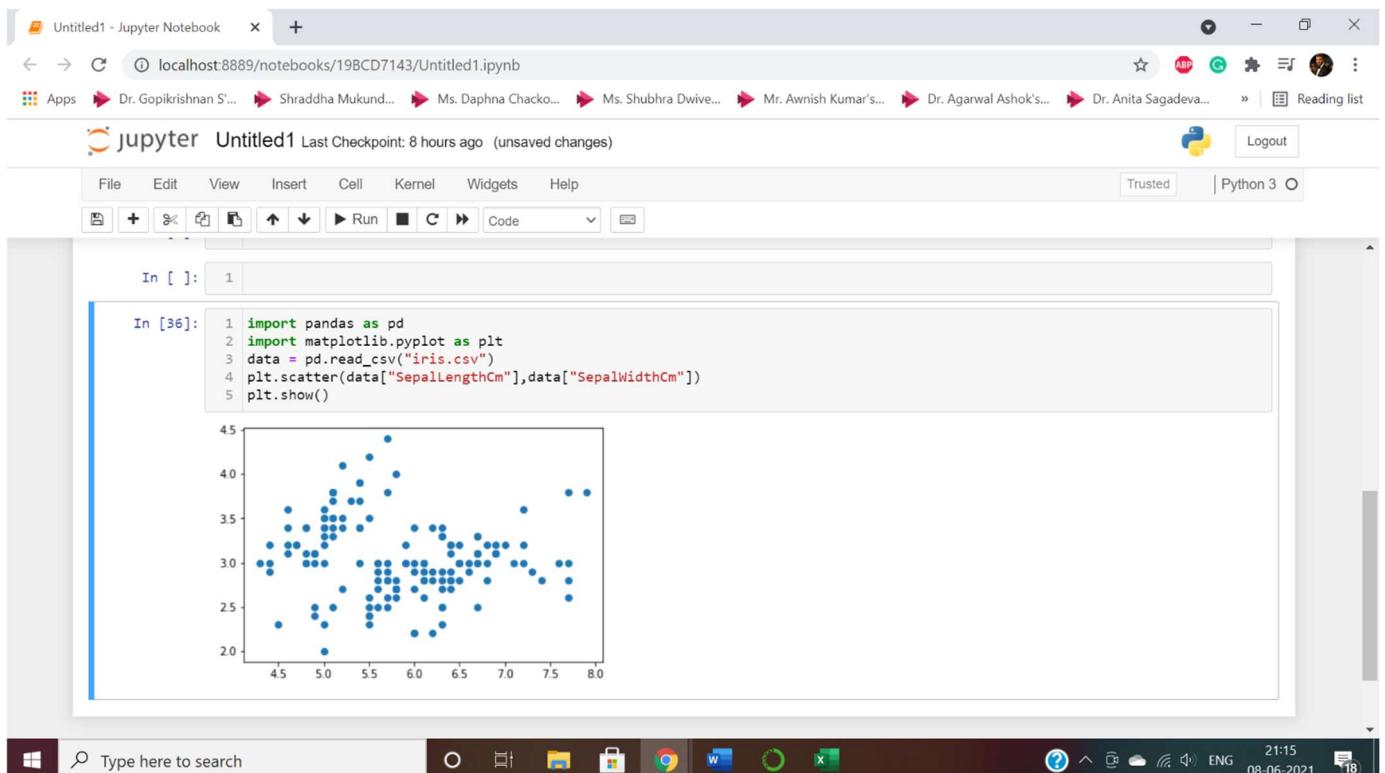
7.

a) Draw scatter between sepal_length and sepal_width.

Answer:

```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("iris.csv")
plt.scatter(data["SepalLengthCm"],data["SepalWidthCm"])
plt.show()
```

OUTPUT



b) Draw line graph of all class of iris data.

Answer:

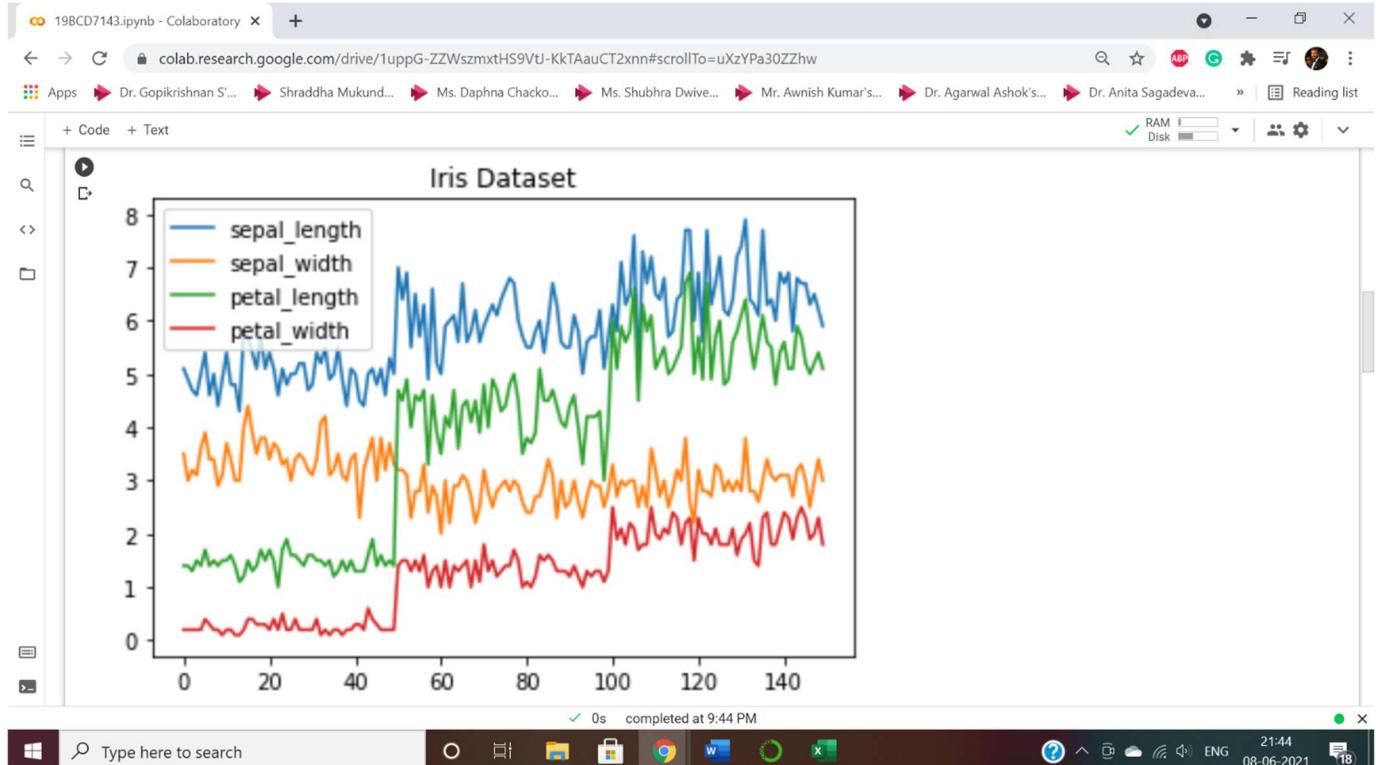
```
import pandas as pd  
  
iris = pd.read_csv('iris.csv', names=['sepal_length', 'sepal_width',  
'petal_length', 'petal_width', 'class'])  
  
columns = iris.columns.drop(['class'])  
  
x_data = range(0, iris.shape[0])  
  
fig, ax = plt.subplots()  
  
for column in columns:
```

```
ax.plot(x_data, iris[column])
```

```
ax.set_title('Iris Dataset')
```

```
ax.legend()
```

OUTPUT



c) Compute `iris.plot.hist(subplots=True, layout=(2,2), figsize=(10, 10), bins=20)`.

Answer:

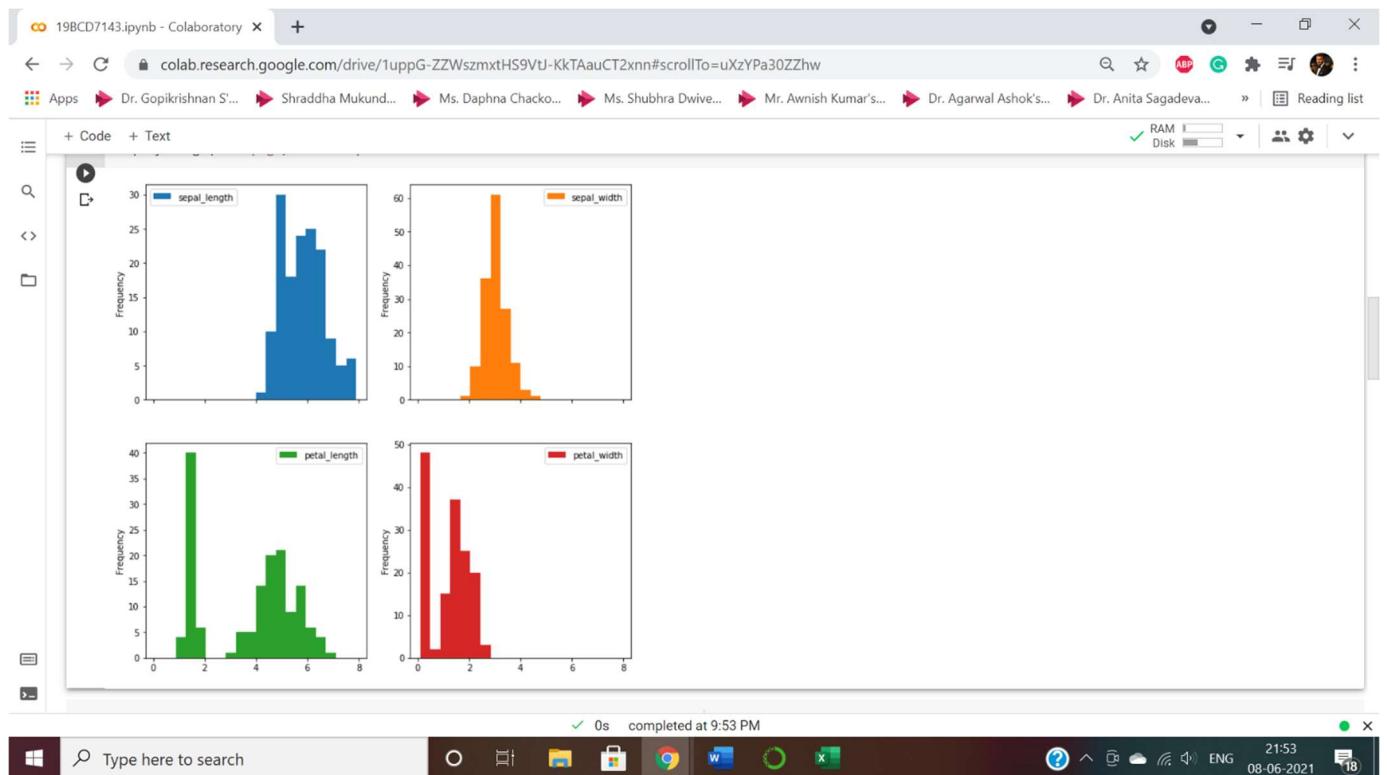
```
import pandas as pd
```

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

```
data = pd.read_csv("iris.csv")
```

```
data.plot.hist(subplots=True, layout=(5,5), figsize=(10, 10), bins=20)
```

OUTPUT



d) Draw correlation matrix between all class.

Answer:

```
import pandas as pd  
import matplotlib.pyplot as plt  
data = pd.read_csv("iris.csv")  
data.corr()
```

```
corr=data.corr()

fig, ax = plt.subplots()

# create heatmap
im = ax.imshow(corr.values)

# set labels
ax.set_xticks(np.arange(len(corr.columns)))
ax.set_yticks(np.arange(len(corr.columns)))
ax.set_xticklabels(corr.columns)
ax.set_yticklabels(corr.columns)

# Rotate the tick labels and set their alignment.
plt.setp(ax.get_xticklabels(), rotation=45, ha="right",
         rotation_mode="anchor")

# Loop over data dimensions and create text annotations.
for i in range(len(corr.columns)):
    for j in range(len(corr.columns)):
        text = ax.text(j, i, np.around(corr.iloc[i, j], decimals=2),
                      ha="center", va="center", color="black")
```

OUTPUT

A screenshot of a Jupyter Notebook interface. The top bar shows the title "Untitled1 - Jupyter Notebook" and the URL "localhost:8889/notebooks/19BCD7143/Untitled1.ipynb". The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Trusted, and Python 3. The toolbar below has icons for New, Open, Save, Run, Stop, Cell, Code, and Help. The main area shows code in In [64] and its output in Out[64].

```
In [64]:  
1 import pandas as pd  
2 import matplotlib.pyplot as plt  
3 data = pd.read_csv("iris.csv")  
4 data.corr()  
  
Out[64]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
Id	1.000000	0.716676	-0.397729	0.882747	0.899759
SepalLengthCm	0.716676	1.000000	-0.109369	0.871754	0.817954
SepalWidthCm	-0.397729	-0.109369	1.000000	-0.420516	-0.356544
PetalLengthCm	0.882747	0.871754	-0.420516	1.000000	0.962757
PetalWidthCm	0.899759	0.817954	-0.356544	0.962757	1.000000

