Experiment 2

Student Name: Sahil Kaundal
UID: 21BCS8197
Branch: BE CSE (Lateral Entry)
Section/Group: 616/A
Semester: 5th
Date of Performance: 23/08/2022
Subject Name: ML Lab
Subject Code: 20CSP-317

1. Aim/Overview of the practical:

Implement Data Visualization.

2. Task To Be Done:

To analyze the data for the certain trends, patterns may become difficult if the data is in its raw format. To overcome this data visualization comes into play. Data visualization provides a good, organized pictorial representation of the data which makes it easier to understand, observe and analyze.

3. Apparatus / Simulator Used:

1. Windows 7 or above.
2. Google Collab.

Python provides various types of libraries that comes with different types of features which can support various types of graphs. These libraries are:
Matplotlib, Seaborn, Bokeh, Plotly

Matplotlib: It is easy to use low level library built on Numpy arrays. It consists of various plots like scatter plot, line plot, histogram etc. Matplotlib provides a lot of flexibility.

4. Program / Commands:

```
#Sahil Kaundal
#21BCS8197
import pandas as pd
import seaborn as sns
```

```
#Load the data
data = pd.read_csv('/content/sample_data/california_housing_test.csv')
#View the data
data.head()
import matplotlib.pyplot as plt
plt.scatter(data['total_rooms'], data['total_bedrooms'])
plt.title('scatter plot')
plt.xlabel('total_rooms')
plt.ylabel('total_bedrooms')
plt.show()
plt.scatter(data['total_rooms'], data['total_bedrooms'], c=data['households'], s=data
['population'])
plt.xlabel('total_rooms')
plt.ylabel('total_bedrooms')
plt.colorbar()
plt.show()
plt.bar(data['total_rooms'], data['total_bedrooms'])
plt.title('Bar Chart')
plt.xlabel('total_rooms')
plt.ylabel('total_bedrooms')
plt.show()
plt.bar(data['population'], data['households'])
plt.title('Bar Chart')
plt.xlabel('population')
plt.ylabel('households')
plt.show()
plt.hist(data['population'])
plt.title("Histogram")
plt.show()
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
sns.scatterplot(x='population', y='households', data=data,)
plt.show()
sns.scatterplot(x='population', y='households', data=data, hue='median_income')
plt.show()
sns.scatterplot(x='median_income', y='households', data=data, hue='population')
plt.show()
sns.lineplot(x='median_income', y='households', data=data)
```

```
plt.show()
sns.barplot(x='total_rooms', y='households', data=data, hue='population')
plt.show()
sns.histplot(x='population', y='households', data=data, hue='median_income')
plt.show()
import plotly.express as px
import pandas as pd
data = pd.read_csv('/content/sample_data/california_housing_test.csv')
fig = px.scatter(data, y='latitude', color='total_rooms')
fig.show()
import plotly.express as px
import pandas as pd
data = pd.read_csv('/content/sample_data/california_housing_test.csv')
fig = px.bar(data, x='housing_median_age', y='latitude', color='total_rooms')
fig.show()
import plotly.express as px
import pandas as pd
data = pd.read_csv('/content/sample_data/california_housing_test.csv')
fig = px.histogram(data, x='housing_median_age', color='total_rooms')
fig.show()
sns.histplot (x='longitude', y='median_house_value', data = data, hue='median_income'
)
plt.show()
sns.barplot (x='longitude', y='median_house_value', data = data, hue='housing_median_
age')
plt.show()
```

5. Result/Output/Writing Summary:

Experiment 2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

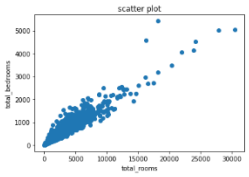
RAM Disk Editing

```
[ ] import pandas as pd
    import seaborn as sns
    data = pd.read_csv('/content/sample_data/california_housing_test.csv')
    data.head()
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value
0	-122.05	37.37	27.0	3885.0	661.0	1537.0	606.0	6.6085	344700.0
1	-118.30	34.26	43.0	1510.0	310.0	809.0	277.0	3.5990	176500.0
2	-117.81	33.78	27.0	3589.0	507.0	1484.0	495.0	5.7934	270500.0
3	-118.36	33.82	28.0	67.0	15.0	49.0	11.0	6.1359	330000.0
4	-119.67	36.33	19.0	1241.0	244.0	850.0	237.0	2.9375	81700.0

```
[ ] import matplotlib.pyplot as plt
```

```
plt.scatter(data['total_rooms'], data['total_bedrooms'])
plt.title('scatter plot')
plt.xlabel('total_rooms')
plt.ylabel('total_bedrooms')
plt.show()
```



0s completed at 6:15 PM

Experiment 2.ipynb

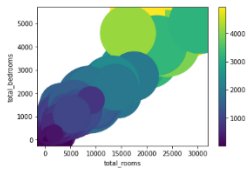
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

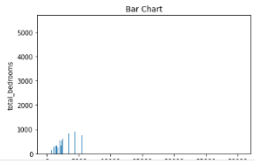
+ Code + Text

RAM Disk Editing

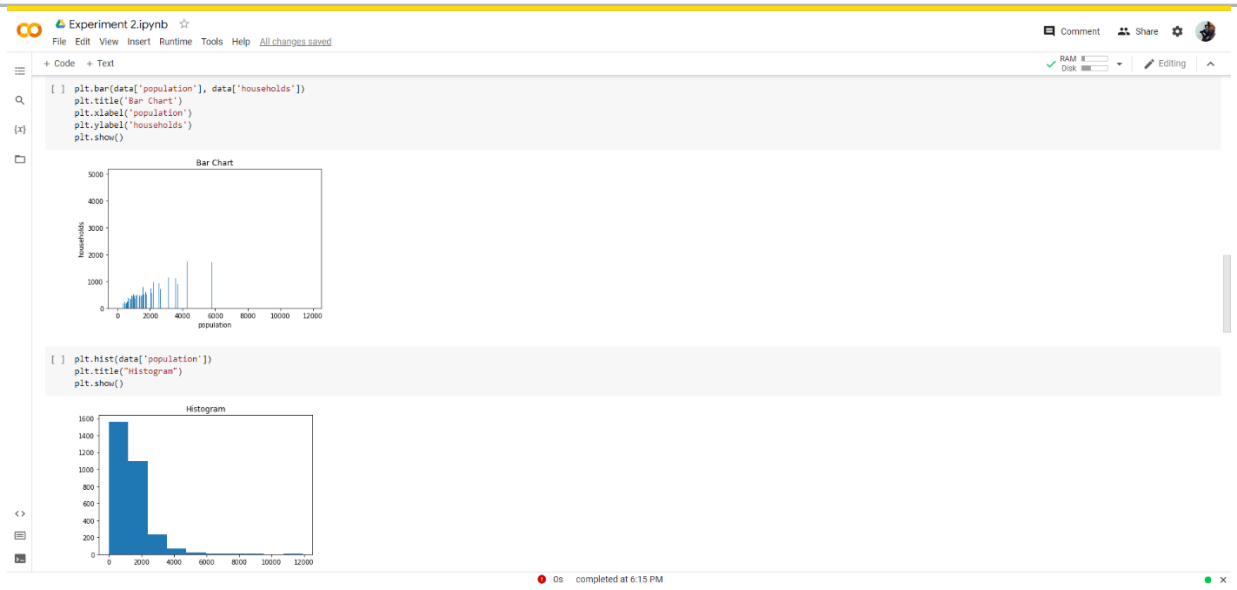
```
plt.scatter(data['total_rooms'], data['total_bedrooms'], c=data['households'], s=data['population'])
plt.xlabel('total_rooms')
plt.ylabel('total_bedrooms')
plt.colorbar()
plt.show()
```

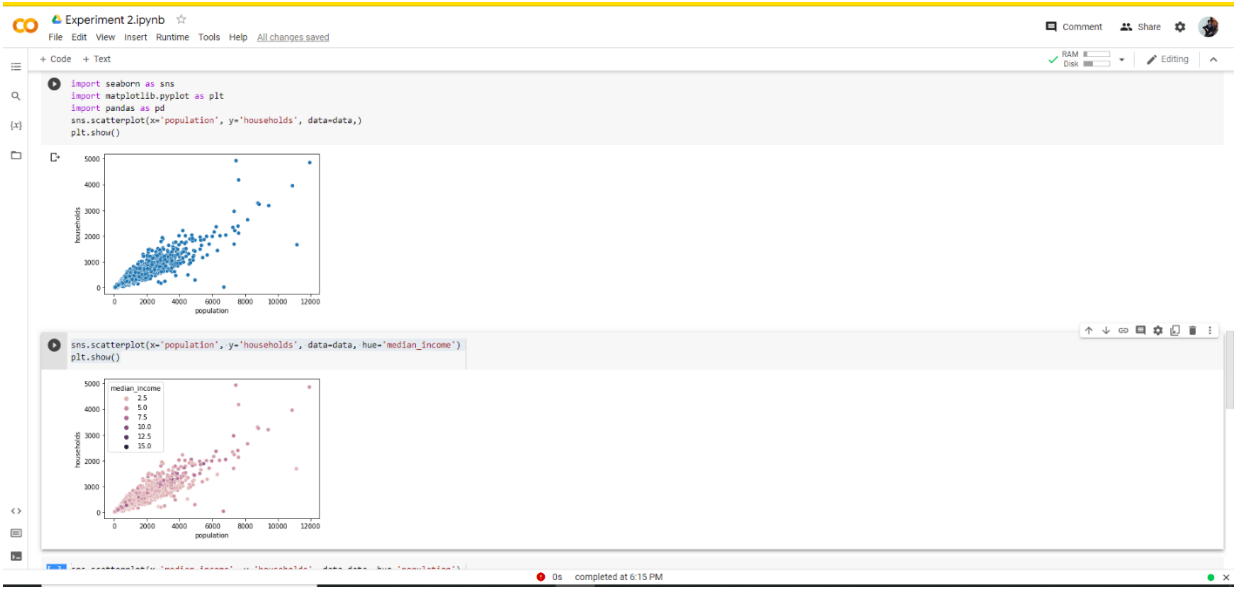


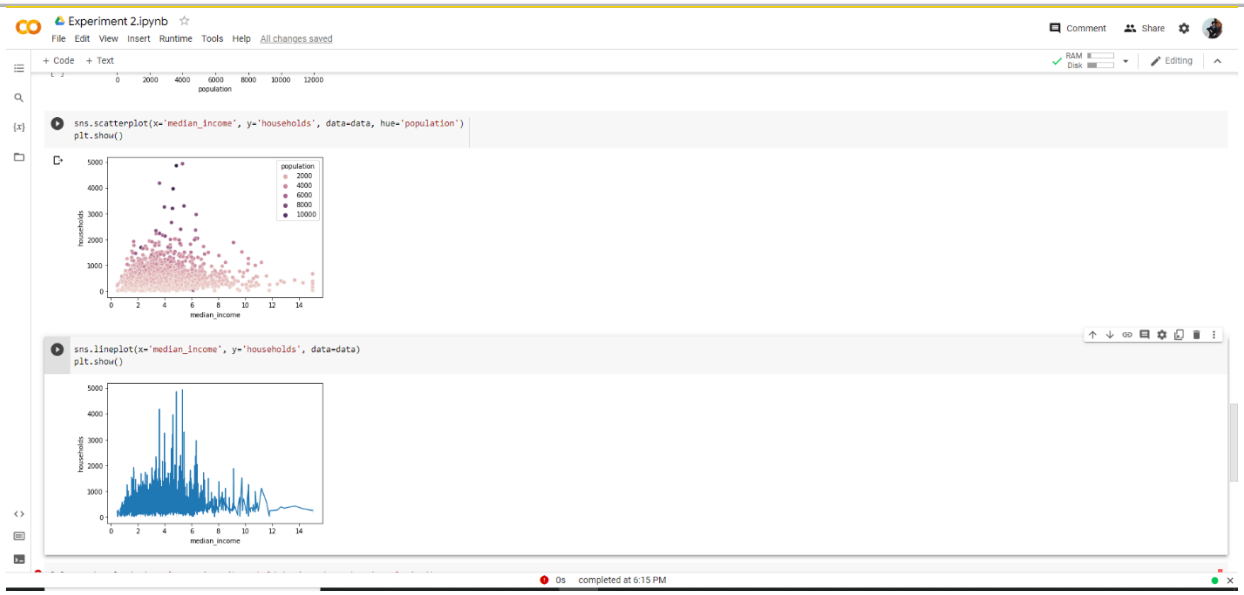

```
[ ] plt.bar(data['total_rooms'], data['total_bedrooms'])
    plt.title('Bar Chart')
    plt.xlabel('total_rooms')
    plt.ylabel('total_bedrooms')
    plt.show()
```



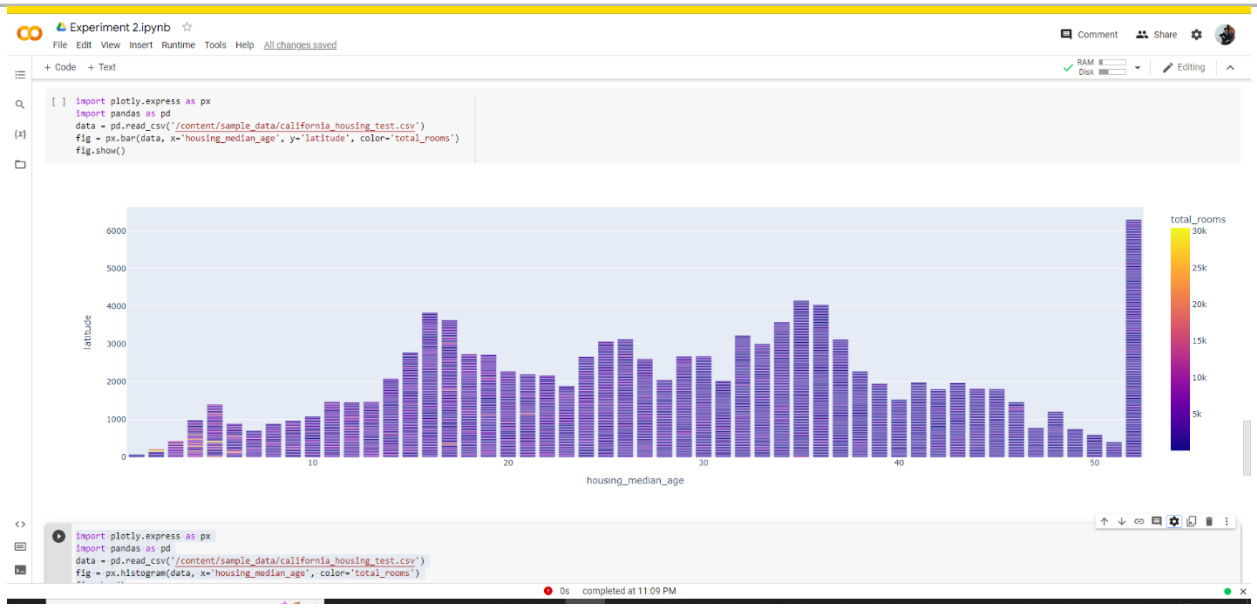
0s completed at 6:15 PM

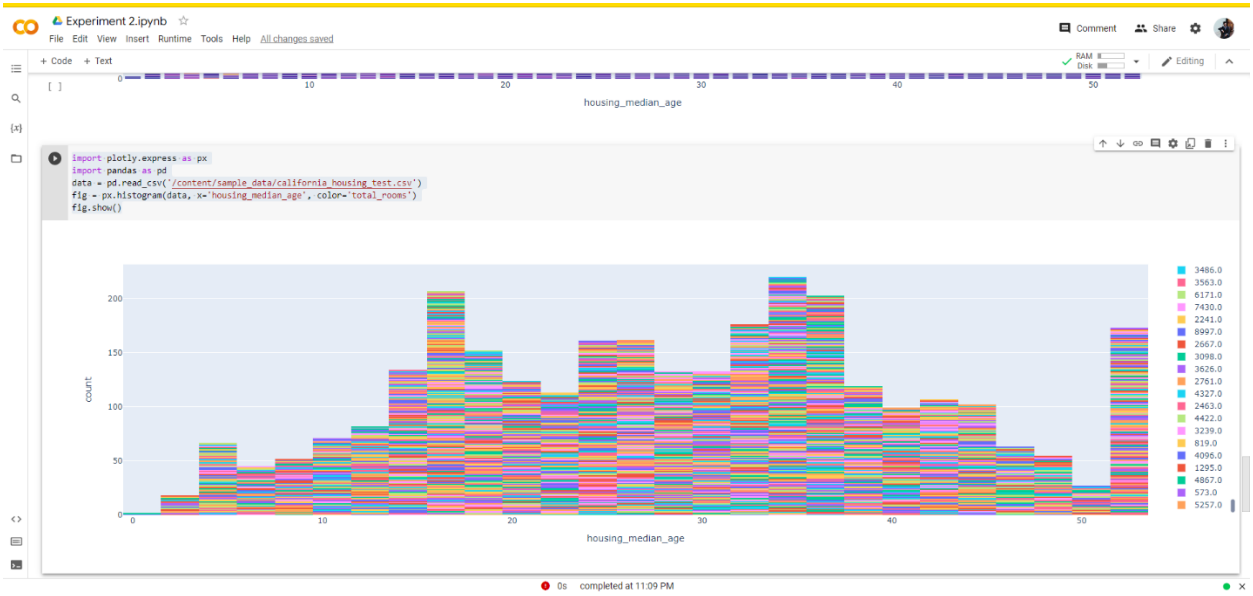

Experiment 2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk Editing
plt.bar(data['population'], data['households'])
plt.title('Bar Chart')
plt.xlabel('population')
plt.ylabel('households')
plt.show()
Bar Chart
households
population
plt.hist(data['population'])
plt.title("Histogram")
plt.show()
Histogram
0s completed at 6:15 PM

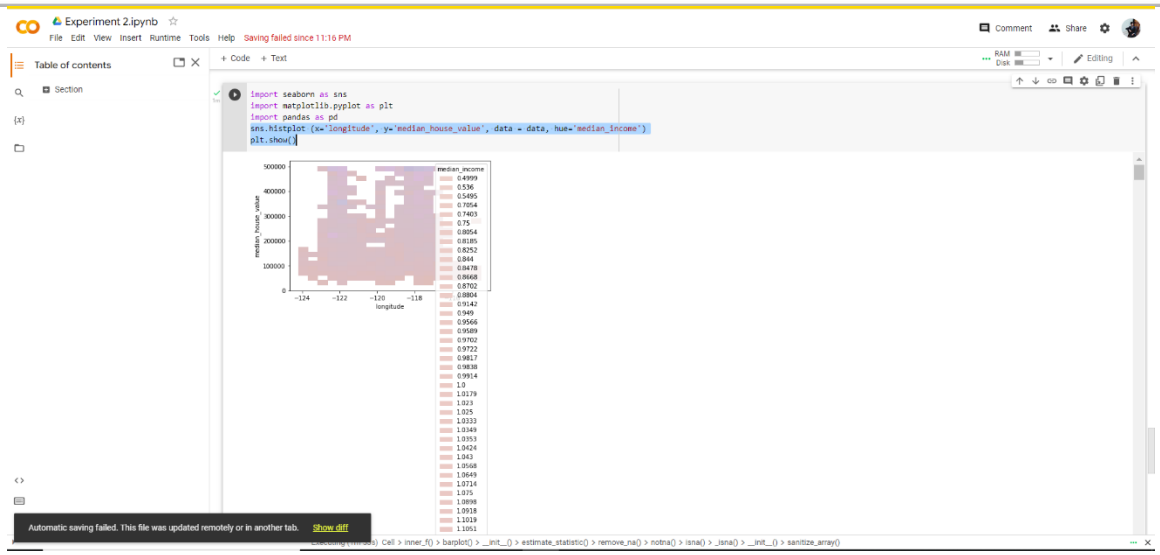

Experiment 2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk Editing
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
sns.scatterplot(x='population', y='households', data=data,)
plt.show()
households
population
sns.scatterplot(x='population', y='households', data=data, hue='median_income')
plt.show()
median_income
2.5
5.0
7.5
10.0
12.5
15.0
households
population
0s completed at 6:15 PM


Experiment 2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
population
sns.scatterplot(x='median_income', y='households', data=data, hue='population')
plt.show()
households
median_income
sns.lineplot(x='median_income', y='households', data=data)
plt.show()
0s completed at 6:15 PM


Experiment 2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
import plotly.express as px
import pandas as pd
data = pd.read_csv('/content/sample_data/california_housing_test.csv')
fig = px.scatter(data, y='latitude', color='total_rooms')
fig.show()
latitude
index
total_rooms
30k
25k
20k
15k
10k
5k
import plotly.express as px
import pandas as pd
data = pd.read_csv('/content/sample_data/california_housing_test.csv')
fig = px.bar(data, x='housing_median_age', y='latitude', color='total_rooms')
fig.show()
0s completed at 11:09 PM


Experiment 2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
[ ] import plotly.express as px
import pandas as pd
data = pd.read_csv('/content/sample_data/california_housing_test.csv')
fig = px.bar(data, x='housing_median_age', y='latitude', color='total_rooms')
fig.show()
latitude
housing_median_age
total_rooms
30k
25k
20k
15k
10k
5k
import plotly.express as px
import pandas as pd
data = pd.read_csv('/content/sample_data/california_housing_test.csv')
fig = px.histogram(data, x='housing_median_age', color='total_rooms')
0s completed at 11:09 PM


Experiment 2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
housing_median_age
import plotly.express as px
import pandas as pd
data = pd.read_csv('/content/sample_data/california_housing_test.csv')
fig = px.histogram(data, x='housing_median_age', color='total_rooms')
fig.show()
count
housing_median_age
0s completed at 11:09 PM



Learning outcomes (What I have learnt):

Implement Data Visualization.

Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.			