**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.

1. **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()

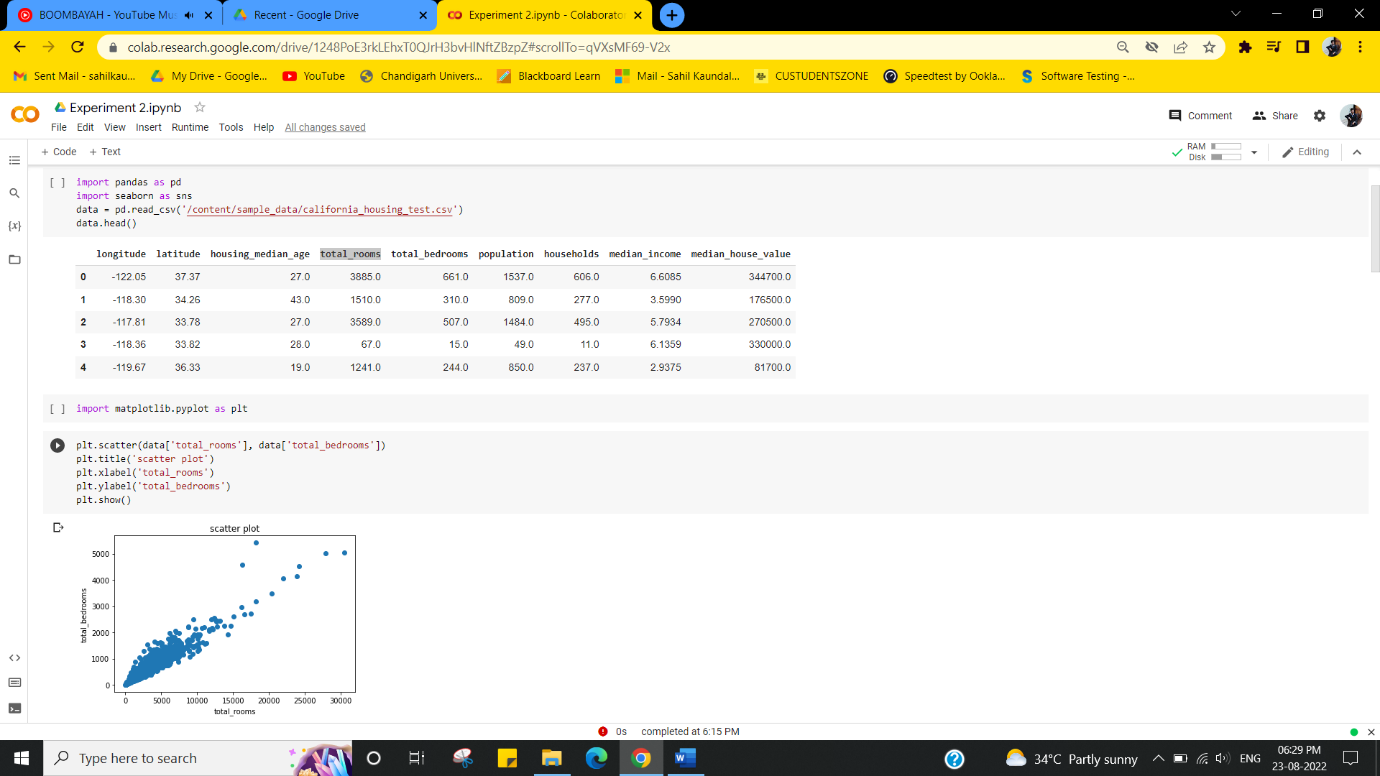
sns.histplot (x='longitude', y='median\_house\_value', data = data, hue='median\_income')

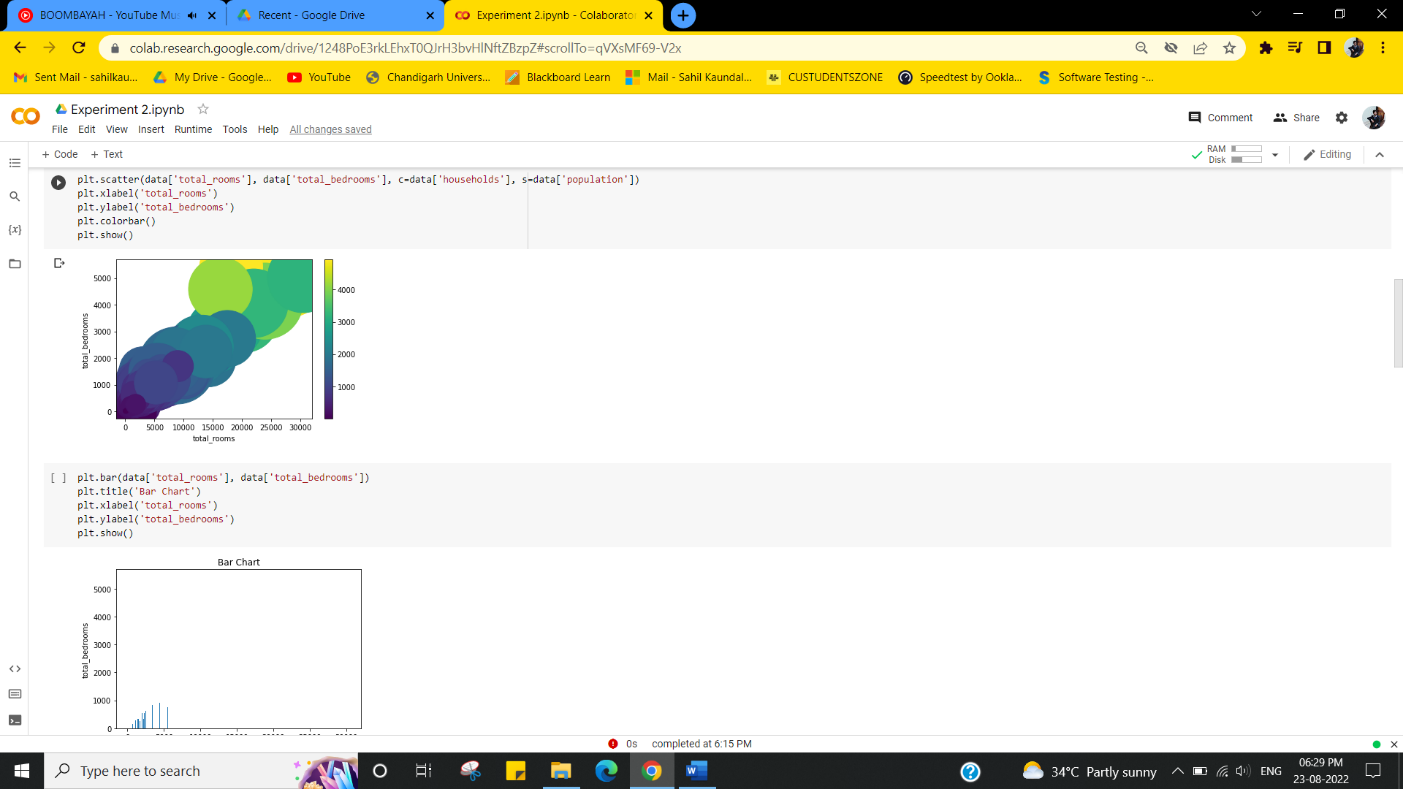
plt.show()

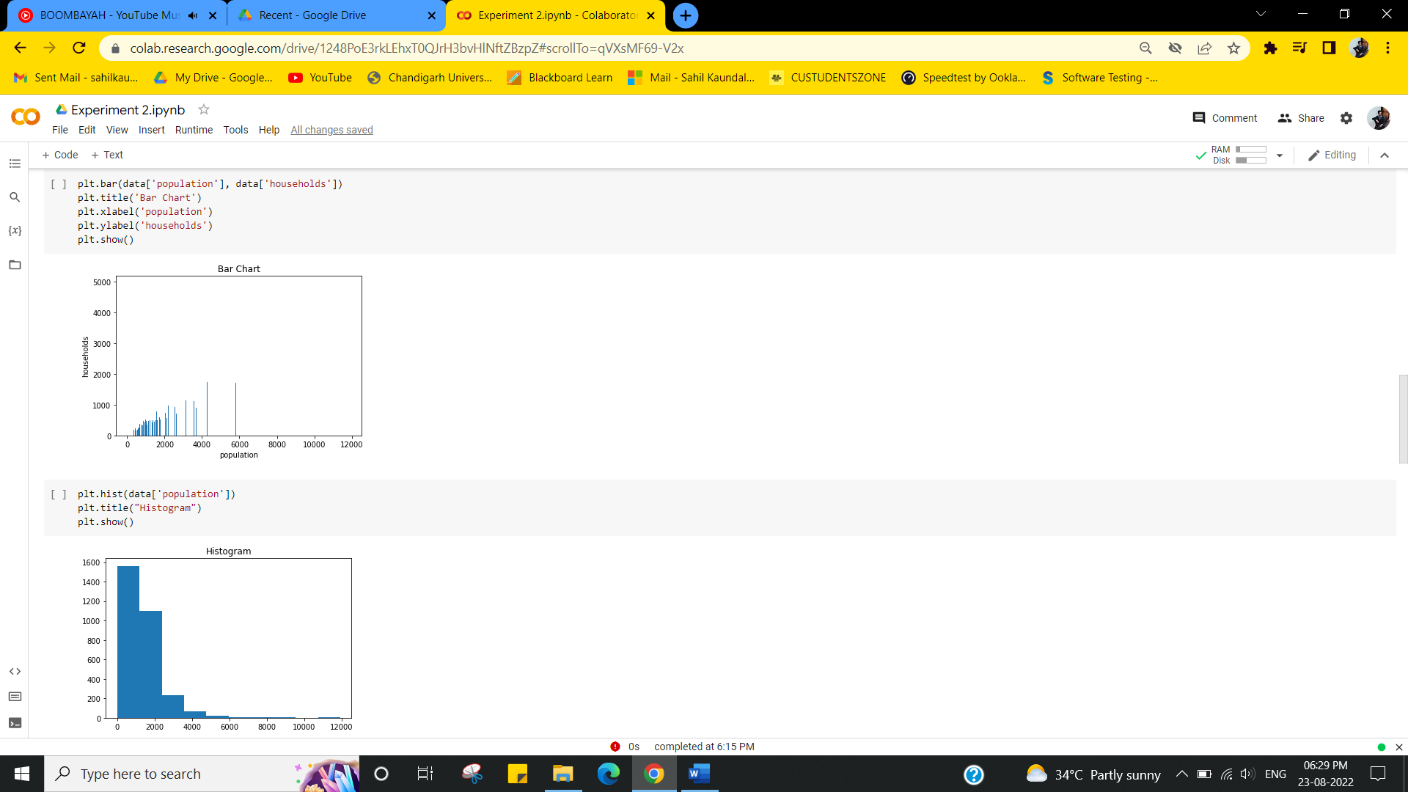
sns.barplot (x='longitude', y='median\_house\_value', data = data, hue='housing\_median\_age')

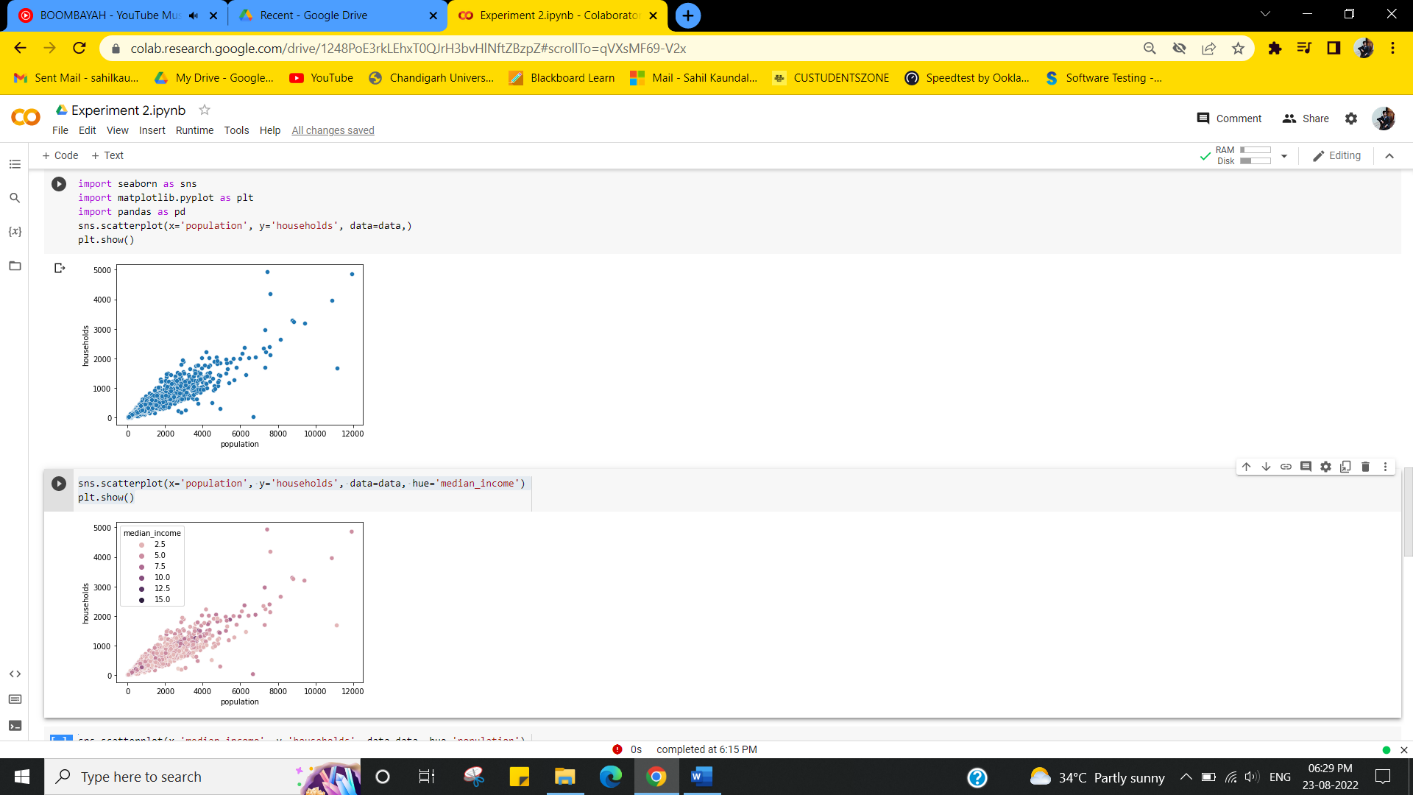
plt.show()

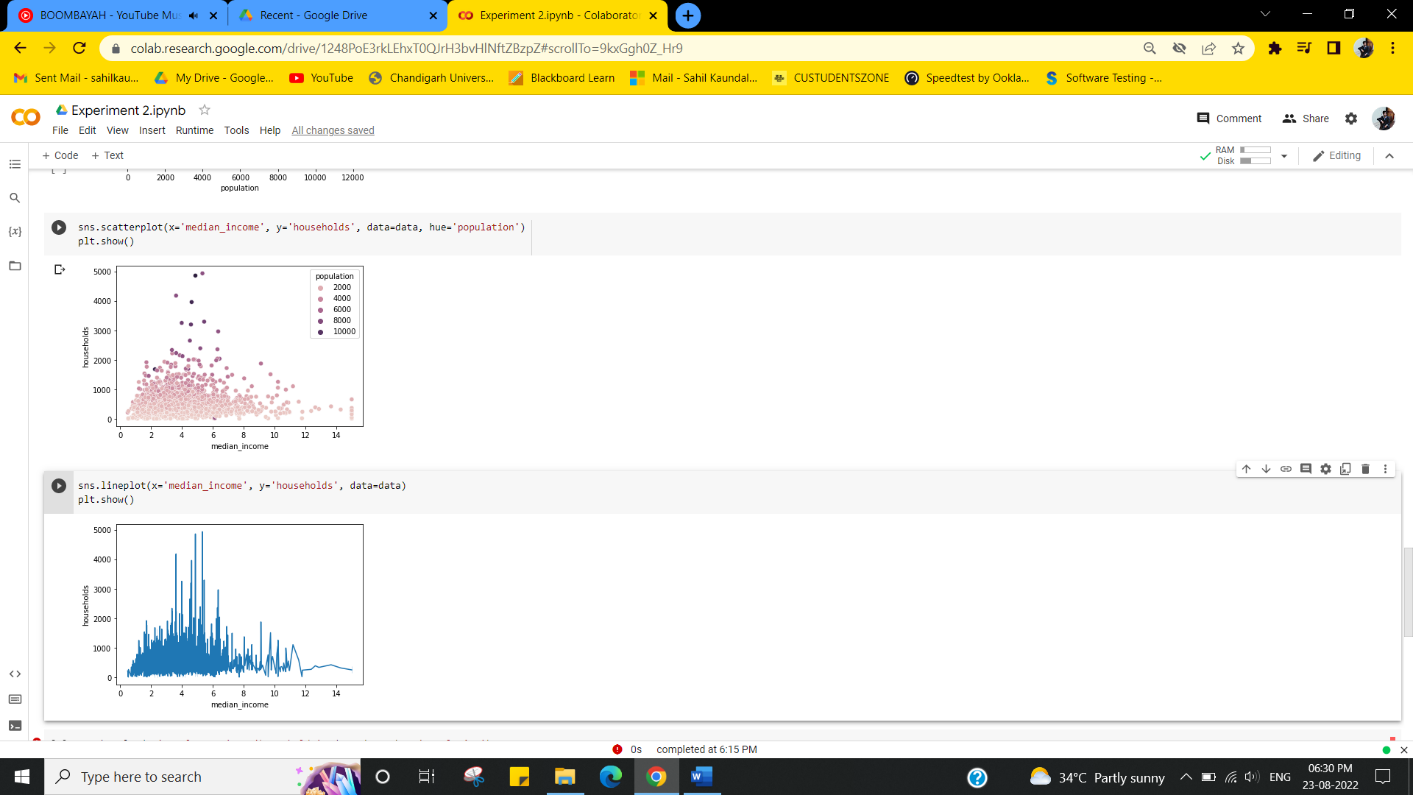
1. **Result/Output/Writing Summary:**

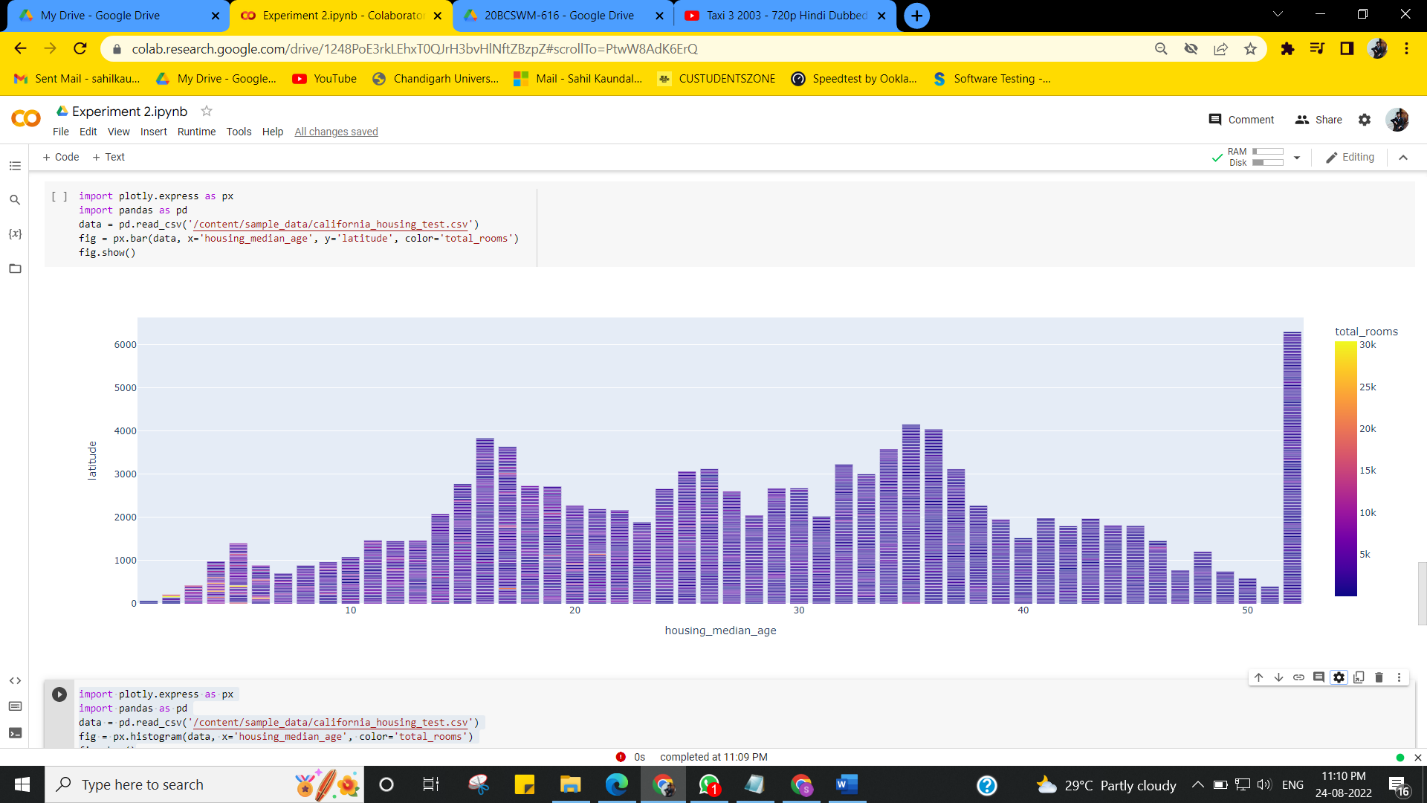


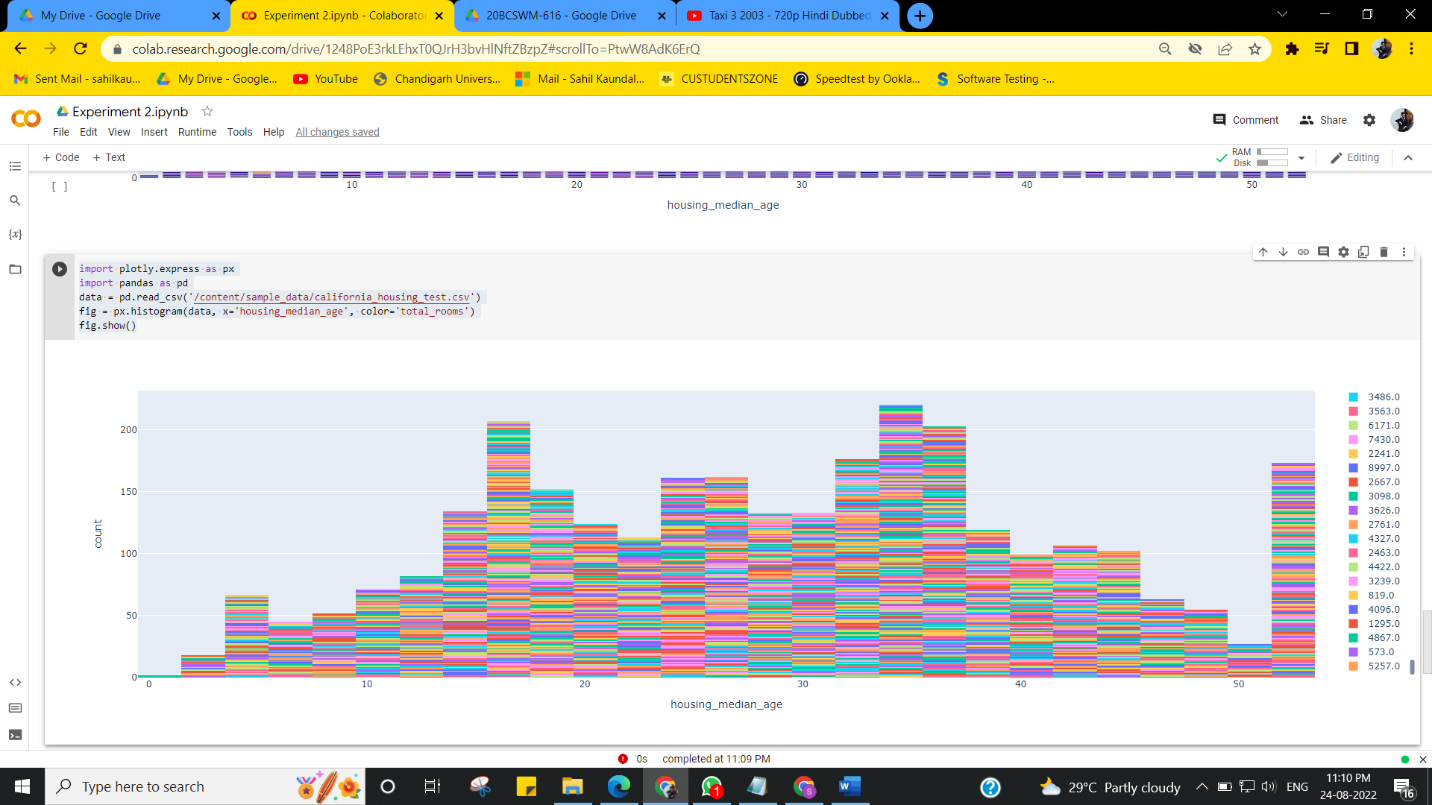


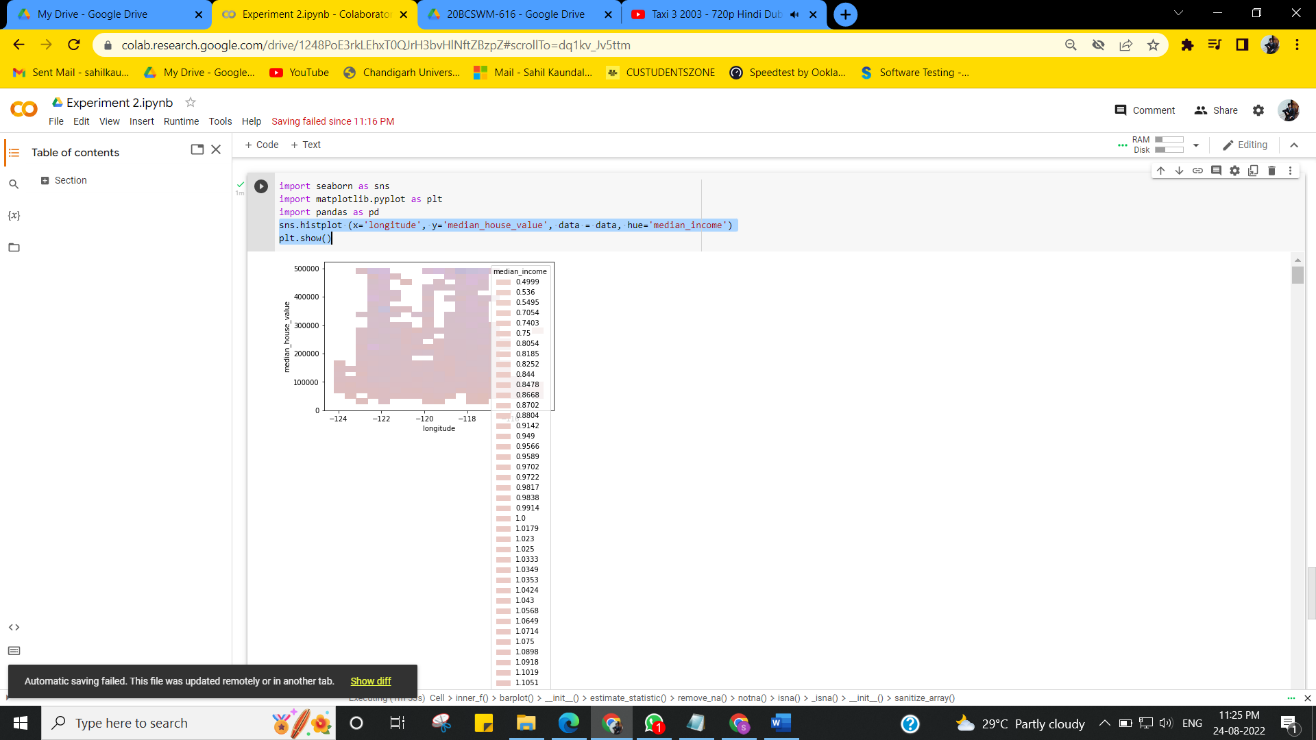


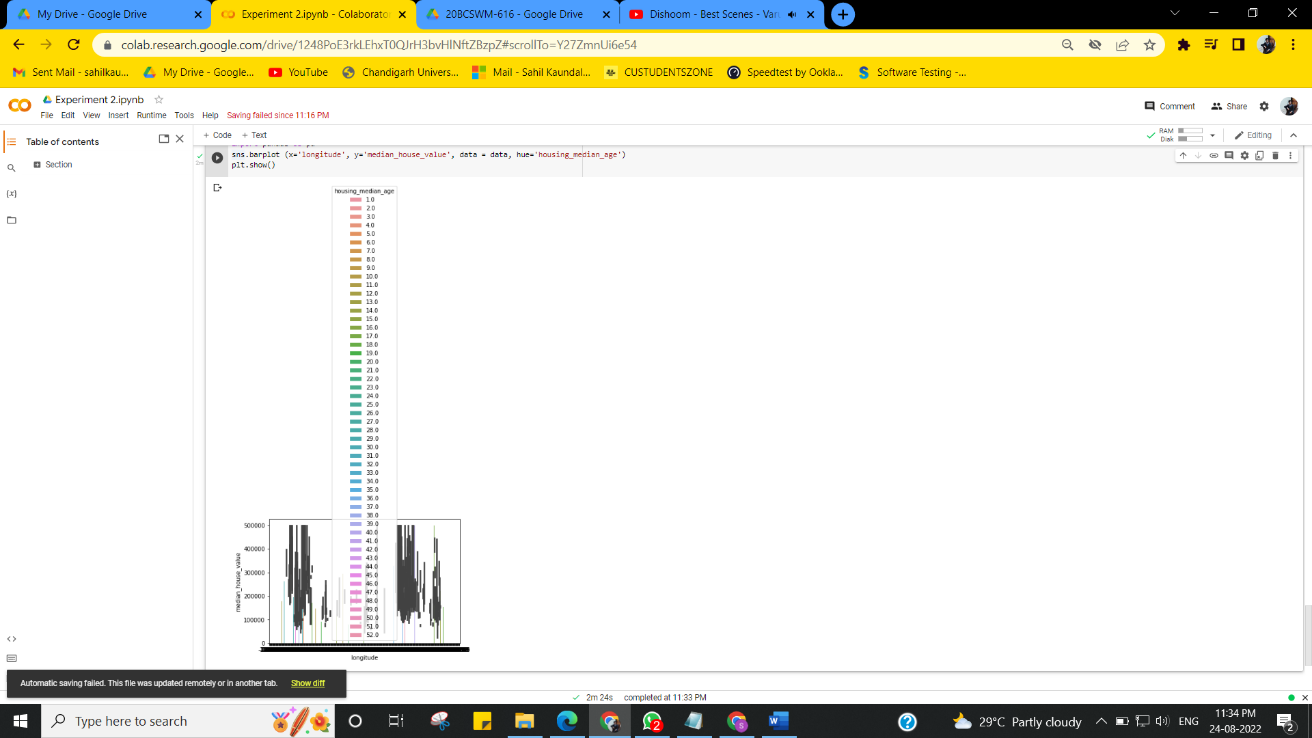












**Learning outcomes (What I have learnt):**

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. |  |  |  |