

ARJUN COLLEGE OF TECHNOLOGY  
**ASSIGNMENT – 3**  
NAAN MUDHALVAN

NAME : G.ANVESH

REG NO: 723920104022



My IBM | \*New story | New report | Technical Training Session | Data Analytics Session 4 | Untitled4.ipynb - Colaboratory

colab.research.google.com/drive/13oU07TPPh3DO7ryVK8bopENFXKdF9fyb?authuser=0#scrollTo=M1-i-3fLTXa

Google | Gmail | YouTube | Maps | top programs | New Tab | (727) YouTube

Untitled4.ipynb | File | Edit | View | Insert | Runtime | Tools | Help | All changes saved | Comment | Share | Settings | Profile

Files | + Code | + Text

[1] import pandas as pd

df = pd.read\_csv('/content/House Price India.csv')  
df.head()

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0

5 rows x 23 columns

House Price India.csv | 1 to 10 of 14620 entries | Filter |

id	Date	number of bedrooms	number of bathrooms
6762810145	42491	5	2.5
6762810635	42491	4	2.5
6762810998	42491	5	2.75
6762812605	42491	4	2.5
6762812919	42491	3	2
6762813105	42491	3	2.5
6762813157	42491	5	3.25
6762813599	42491	3	1.75
6762813600	42491	3	2.5
6762814481	42491	4	2.25

Show 10 per page | 1 | 2 | 10 | 100 | 1000 | 1400 | 1460 | 1462

0s completed at 11:06 AM

Type here to search | archi... | archi... | Hou... | Unti... | Turb... | Scre... | Wha... | AI A... | ENG | 11:07 04-10-2023

My IBM | \*New story | New report | Technical Training Session | Data Analytics Session 4 | Untitled4.ipynb - Colaboratory

colab.research.google.com/drive/13oU07TPPh3DO7ryVK8bopENFXKdF9fyb?authuser=0#scrollTo=weGhJ-CnPg2w

Google | Gmail | YouTube | Maps | top programs | New Tab | (727) YouTube

Untitled4.ipynb | File | Edit | View | Insert | Runtime | Tools | Help | Comment | Share | Settings | Profile

Files | + Code | + Text

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14620 entries, 0 to 14619
Data columns (total 23 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                     14620 non-null  int64
1   Date                                  14620 non-null  int64
2   number of bedrooms                   14620 non-null  int64
3   number of bathrooms                 14620 non-null  float64
4   living area                          14620 non-null  int64
5   lot area                            14620 non-null  int64
6   number of floors                    14620 non-null  float64
7   waterfront present                  14620 non-null  int64
8   number of views                     14620 non-null  int64
9   condition of the house              14620 non-null  int64
10  grade of the house                  14620 non-null  int64
11  Area of the house(excluding basement) 14620 non-null  int64
12  Area of the basement                14620 non-null  int64
13  Built Year                          14620 non-null  int64
14  Renovation Year                     14620 non-null  int64
15  Postal Code                         14620 non-null  int64
16  latitude                            14620 non-null  float64
17  longitude                           14620 non-null  float64
18  living_area_renov                   14620 non-null  int64
19  lot_area_renov                      14620 non-null  int64
20  Number of schools nearby             14620 non-null  int64
21  Distance from the airport           14620 non-null  int64
22  Price                               14620 non-null  int64

dtypes: float64(4), int64(19)
memory usage: 2.6 MB
```

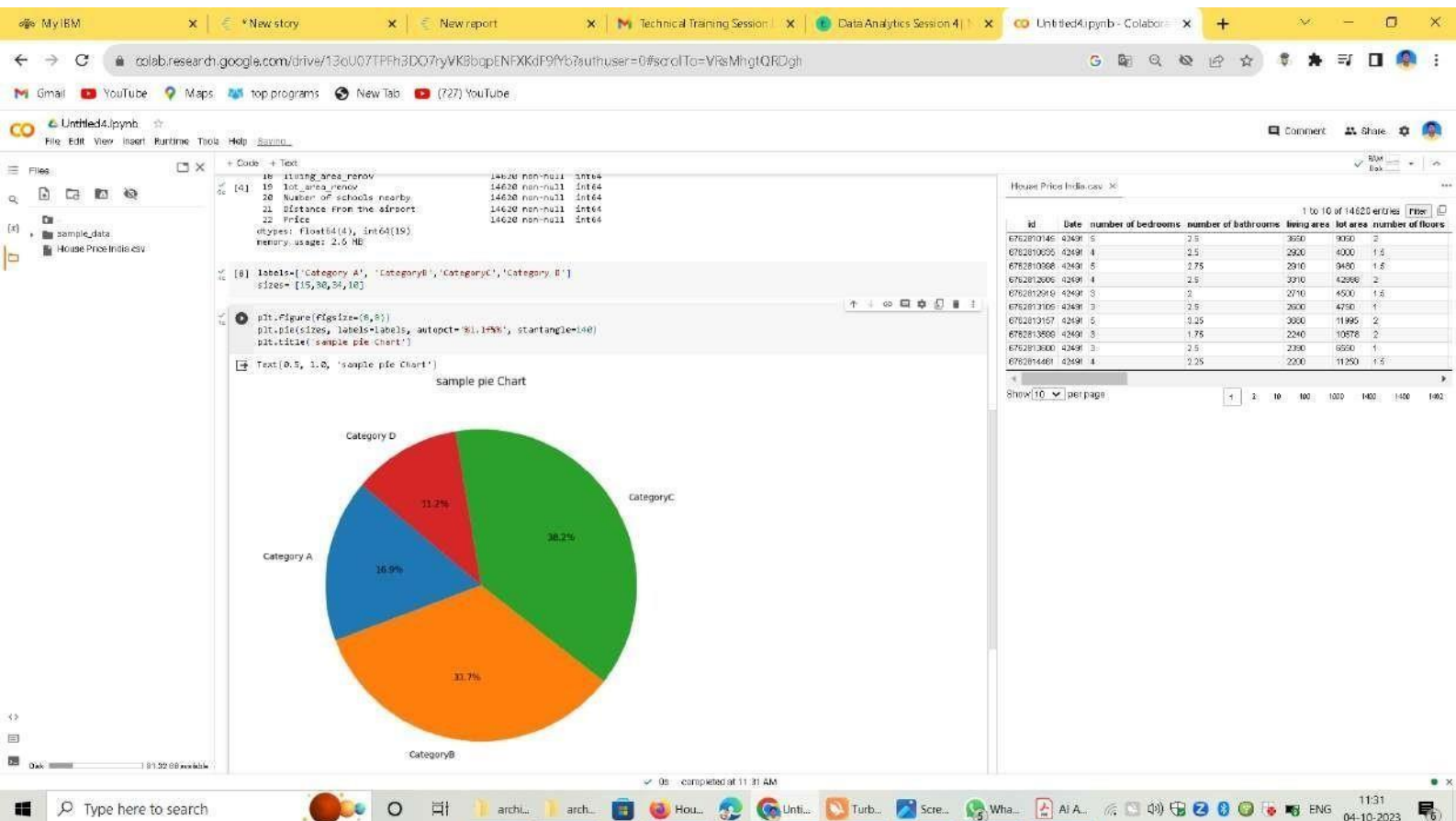
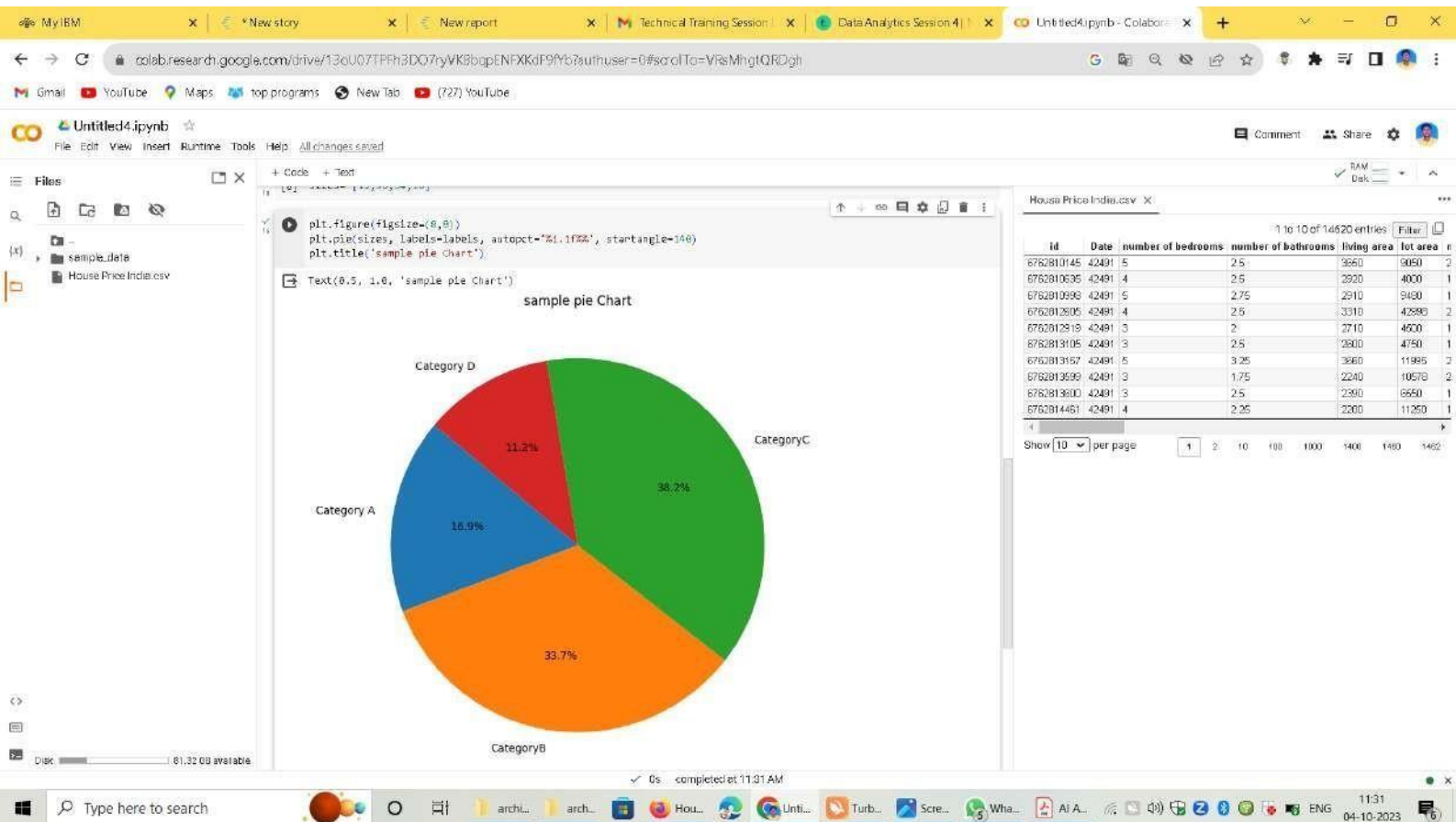
House Price India.csv | 1 to 10 of 14620 entries | Filter |

id	Date	number of bedrooms	number of bathrooms
6762810145	42491	5	2.5
6762810635	42491	4	2.5
6762810998	42491	5	2.75
6762812605	42491	4	2.5
6762812919	42491	3	2
6762813105	42491	3	2.5
6762813157	42491	5	3.25
6762813599	42491	3	1.75
6762813600	42491	3	2.5
6762814481	42491	4	2.25

Show 10 per page | 1 | 2 | 10 | 100 | 1000 | 1400 | 1460 | 1462

0s completed at 11:22 AM

Type here to search | archi... | archi... | Hou... | Unti... | Turb... | Scre... | Wha... | AI A... | ENG | 11:22 04-10-2023





My IBM | \*New story | New report | Technical Training Session | Data Analytics Session 4 | Untitled4.ipynb - Colaboratory

colab.research.google.com/drive/13ou07TPPh3DO7ryVK8bopENFXKdF9fyb?authuser=0#scrollTo=VRsMhgtQRDgh

Google | YouTube | Maps | top programs | New Tab | (727) YouTube

100% - + Reset

Comment | Share | Settings | Profile

Files | Code | Text

sample\_data | House Price India.csv

```
plt.figure(figsize=(8,8))
plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=140)
plt.title('sample pie chart')
Text(0.5, 1.0, 'sample pie chart')
```

sample pie Chart

Category A: 16.9%  
Category B: 33.7%  
Category C: 38.2%  
Category D: 11.2%

House Price India.csv | 1 to 10 of 14620 entries | Filter

id	Date	number of bedrooms	number of bathrooms
6762810145	42491	5	2.5
6762810635	42491	4	2.5
6762810998	42491	5	2.75
6762812605	42491	4	2.5
6762812919	42491	3	2
6762813105	42491	3	2.5
6762813157	42491	5	3.25
6762813599	42491	3	1.75
6762813600	42491	3	2.5
6762814481	42491	4	2.25

Show 10 per page | 1 2 10 100 1000 1400 1460 1462

completed at 11:31 AM

Type here to search | arch... | arch... | Hou... | Unti... | Turb... | Scre... | Wha... | AI A... | 11:32 04-10-2023

My IBM | \*New story | New report | Technical Training Session | Data Analytics Session 4 | Untitled4.ipynb - Colaboratory

colab.research.google.com/drive/13ou07TPPh3DO7ryVK8bopENFXKdF9fyb?authuser=0#scrollTo=0R253h7LSA6Q

Google | YouTube | Maps | top programs | New Tab | (727) YouTube

100% - + Reset

Comment | Share | Settings | Profile

Files | Code | Text

sample\_data | House Price India.csv

```
plt.axis('equal')
plt.show()
```

CategoryB

House Price India.csv | 1 to 10 of 14620 entries | Filter

id	Date	number of bedrooms	number of bathrooms
6762810145	42491	5	2.5
6762810635	42491	4	2.5
6762810998	42491	5	2.75
6762812605	42491	4	2.5
6762812919	42491	3	2
6762813105	42491	3	2.5
6762813157	42491	5	3.25
6762813599	42491	3	1.75
6762813600	42491	3	2.5
6762814481	42491	4	2.25

Show 10 per page | 1 2 10 100 1000 1400 1460 1462

completed at 11:33 AM

Type here to search | arch... | arch... | Hou... | Unti... | Turb... | Scre... | Wha... | AI A... | 11:33 04-10-2023

My IBM | \*New story | New report | Technical Training Session | Data Analytics Session 4 | Untitled4.ipynb - Colaboratory

colab.research.google.com/drive/13cU07TPPh3DO7ryVK8bopENFXKdF9Yb?authuser=0#scrollTo=XUOG\_dHGSIdP

Google | Gmail | YouTube | Maps | top programs | New Tab | (727) YouTube

Untitled4.ipynb | File | Edit | View | Insert | Runtime | Tools | Help | All changes saved

Files | sample\_data | House Price India.csv

```
plt.figure(figsize=(8,6)) # set the figure size (optional)
plt.scatter(x,y, c='blue', marker='o', label='Data Points') # Scatter plot
plt.xlabel('Variable1') # X-axis label
plt.ylabel('Variable2') # Y-axis label
plt.title('Scatter Plot of Variable1 vs. Variable2') # Title (optional)
plt.grid(True) # Display grid (optional)
plt.legend() # Display legend (optional)

# Show the plot
plt.show()
```

Scatter Plot of Variable1 vs. Variable2

Variable2

Variable1

House Price India.csv | 1 to 10 of 14620 entries | Filter

id	Date	number of bedrooms	number of bathrooms	living area	lot area
6762810145	42491	5	2.5	3660	9050
6762810635	42491	4	2.5	2920	4000
6762810998	42491	5	2.75	2910	9480
6762812605	42491	4	2.5	3310	42880
6762812919	42491	3	2	2710	4600
6762813105	42491	3	2.5	2800	4750
6762813157	42491	5	3.25	3660	11985
6762813599	42491	3	1.75	2240	10570
6762813600	42491	3	2.5	2360	6650
6762814481	42491	4	2.25	2200	11250

Show 10 per page | 2 | 10 | 100 | 1000 | 1400 | 1460 | 1462

0s completed at 11:35 AM

Type here to search | arch... | arch... | Hou... | Unti... | Turb... | Scre... | Wha... | AI A... | ENG | 11:35 | 04-10-2023

My IBM | \*New story | New report | Technical Training Session | Data Analytics Session 4 | Untitled4.ipynb - Colaboratory

colab.research.google.com/drive/13cU07TPPh3DO7ryVK8bopENFXKdF9Yb?authuser=0#scrollTo=XUOG\_dHGSIdP

Google | Gmail | YouTube | Maps | top programs | New Tab | (727) YouTube

Untitled4.ipynb | File | Edit | View | Insert | Runtime | Tools | Help | All changes saved

Files | sample\_data | House Price India.csv

```
plt.ylabel('Variable2') # Y-axis label
plt.title('Scatter Plot of Variable1 vs. Variable2') #
plt.grid(True) # Display grid (optional)
plt.legend() # Display legend (optional)

# show the plot
plt.show()
```

Scatter Plot of Variable1 vs. Variable2

Variable2

Variable1

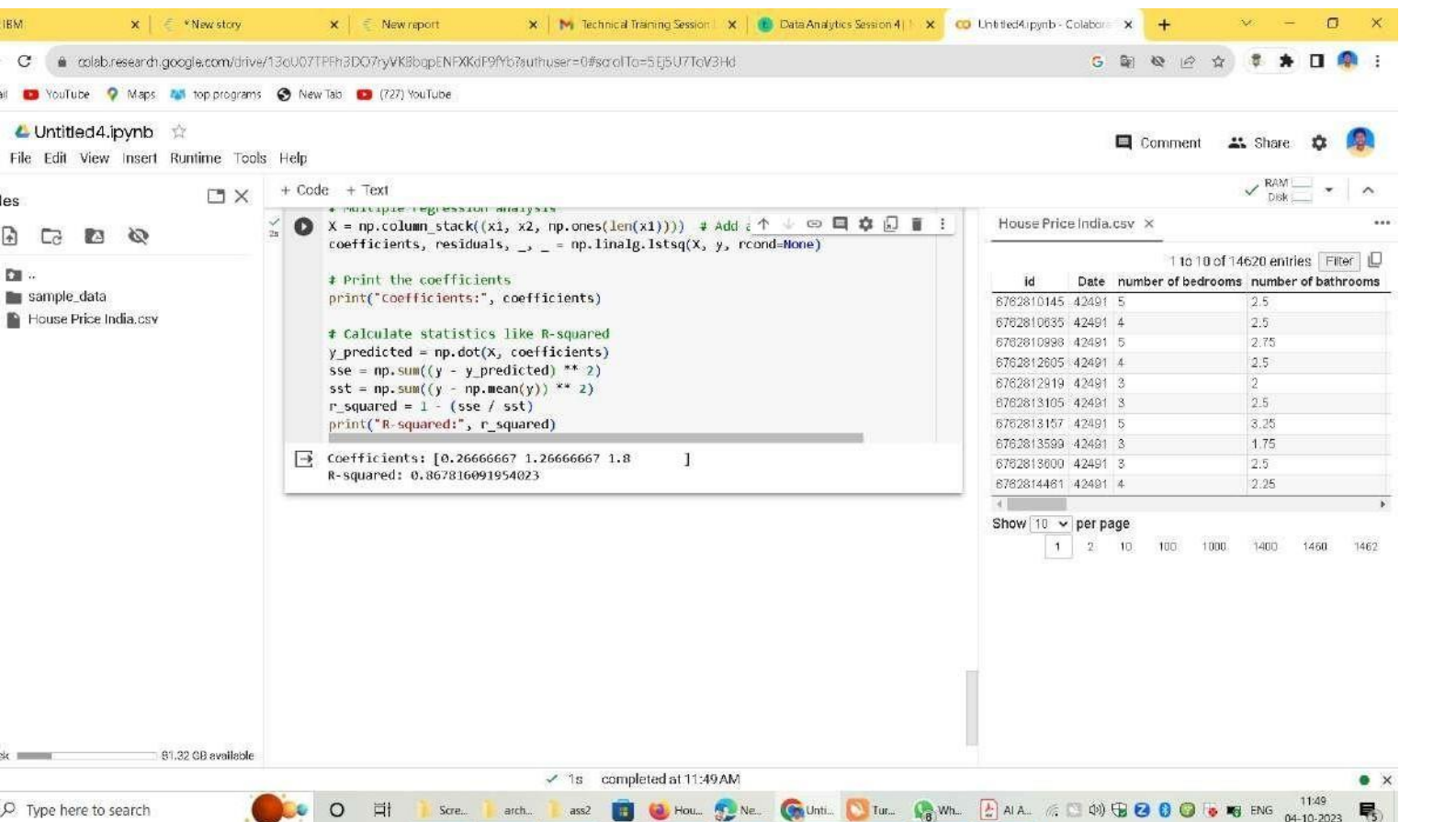
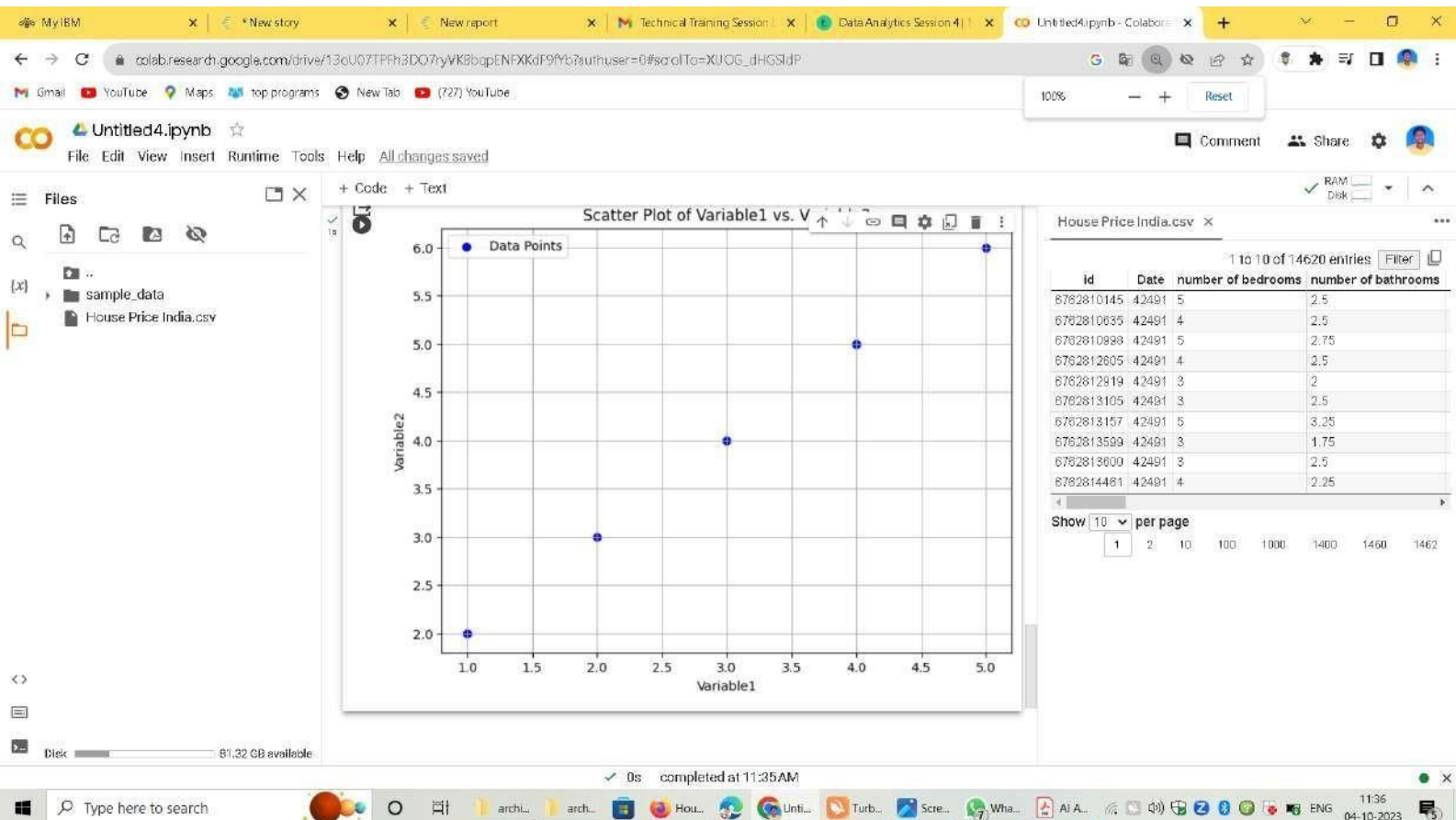
House Price India.csv | 1 to 10 of 14620 entries | Filter

id	Date	number of bedrooms	number of bathrooms
6762810145	42491	5	2.5
6762810635	42491	4	2.5
6762810998	42491	5	2.75
6762812605	42491	4	2.5
6762812919	42491	3	2
6762813105	42491	3	2.5
6762813157	42491	5	3.25
6762813599	42491	3	1.75
6762813600	42491	3	2.5
6762814481	42491	4	2.25

Show 10 per page | 1 | 2 | 10 | 100 | 1000 | 1400 | 1460 | 1462

0s completed at 11:35 AM

Type here to search | arch... | arch... | Hou... | Unti... | Turb... | Scre... | Wha... | AI A... | ENG | 11:36 | 04-10-2023





My IBM x \*New story x New report x Technical Training Session x Data Analytics Session 4 x Untitled4.ipynb - Colaboratory x +

colab.research.google.com/drive/13oU07TPPh3DO7ryVK8bopENFXKdF9fyb?authuser=0#scrollTo=5E5U7TeV3Hd

Gmail YouTube Maps top programs New Tab (77) YouTube

Untitled4.ipynb ☆

File Edit View Insert Runtime Tools Help

Files

- sample\_data
- House Price India.csv

+ Code + Text

```
# Multiple regression analysis
X = np.column_stack((x1, x2, np.ones(len(x1)))) # Add 1 to the array
coefficients, residuals, _, _ = np.linalg.lstsq(X, y, rcond=None)

# Print the coefficients
print("Coefficients:", coefficients)

# Calculate statistics like R-squared
y_predicted = np.dot(X, coefficients)
sse = np.sum((y - y_predicted) ** 2)
sst = np.sum((y - np.mean(y)) ** 2)
r_squared = 1 - (sse / sst)
print("R-squared:", r_squared)
```

Coefficients: [0.26666667 1.26666667 1.8 ]  
R-squared: 0.867816091954023

House Price India.csv x

1 to 10 of 14620 entries Filter

id	Date	number of bedrooms	number of bathrooms
6762810145	42491	5	2.5
6762810635	42491	4	2.5
6762810898	42491	5	2.75
6762812605	42491	4	2.5
6762812919	42491	3	2
6762813105	42491	3	2.5
6762813157	42491	5	3.25
6762813599	42491	3	1.75
6762813600	42491	3	2.5
6762814481	42491	4	2.25

Show 10 per page

1 2 10 100 1000 1400 1460 1462

completed at 11:49 AM

My IBM x \*New story x New report x Technical Training Session x Data Analytics Session 4 x Untitled4.ipynb - Colaboratory x +

colab.research.google.com/drive/13oU07TPPh3DO7ryVK8bopENFXKdF9fyb?authuser=0#scrollTo=8Ydn7CqXLZI

Gmail YouTube Maps top programs New Tab (77) YouTube

Untitled4.ipynb ☆

File Edit View Insert Runtime Tools Help Saving...

Files

- sample\_data
- House Price India.csv

+ Code + Text

```
# Display the first few rows of the dataset
print(df.head())

# Get basic summary statistics for numeric columns
print(df.describe())

# Get information about the dataset, including data types and missing values
print(df.info())
```

```
Variable1  Variable2
0         1         2
1         2         3
2         3         4
3         4         5
4         5         6

Variable1  Variable2
count      5.000000  5.000000
mean       3.000000  4.000000
std        1.581139  1.581139
min        1.000000  2.000000
25%        2.000000  3.000000
50%        3.000000  4.000000
75%        4.000000  5.000000
max        5.000000  6.000000

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 2 columns):
#   column  Non-null count  dtype
---  ---
0  Variable1  5 non-null      int64
1  Variable2  5 non-null      int64
dtypes: int64(2)
```

House Price India.csv x

1 to 10 of 14620 entries Filter

id	Date	number of bedrooms	number of bathrooms
6762810145	42491	5	2.5
6762810635	42491	4	2.5
6762810898	42491	5	2.75
6762812605	42491	4	2.5
6762812919	42491	3	2
6762813105	42491	3	2.5
6762813157	42491	5	3.25
6762813599	42491	3	1.75
6762813600	42491	3	2.5
6762814481	42491	4	2.25

Show 10 per page

1 2 10 100 1000 1400 1460 1462

completed at 11:55 AM



The screenshot shows a Google Colab notebook titled 'Untitled4.ipynb'. The notebook contains the following code and output:

```
# Display the first few rows of the dataset
print(df.head())

# Get basic summary statistics for numeric columns
print(df.describe())

# Get information about the dataset, including data types and missing values
print(df.info())
```

The output of the first code block shows the first 5 rows of the dataset:

	variable1	variable2
0	1	2
1	2	3
2	3	4
3	4	5
4	5	6

The output of the second code block shows summary statistics for two variables:

	variable1	variable2
count	5.000000	5.000000
mean	3.000000	4.000000
std	1.581139	1.581139
min	1.000000	2.000000
25%	2.000000	3.000000
50%	3.000000	4.000000
75%	4.000000	5.000000
max	5.000000	6.000000

The output of the third code block shows the dataset's information:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 2 columns):
#   Column      Non-Null count  Dtype
---  -
0   variable1    5 non-null      int64
1   variable2    5 non-null      int64
```

On the right side of the notebook, a preview of the 'House Price India.csv' dataset is shown, displaying 1 to 10 of 14620 entries. The preview table has columns: id, Date, number of bedrooms, and number of bathrooms.