

ARJUN COLLEGE OF TECHNOLOGY
ASSIGNMENT – 3
NAAN MUDHALVAN

NAME : JAGADESSH S

REG NO: 723920243015

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

colab.research.google.com/drive/13oU07TPPh3DO7ryVK8bopENFXKdF9Yb3?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 Saving...

Comment | Share | Settings | Profile

Files | + Code | + Text

sample_data | House Price India.csv

```
[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

completed at 11:06 AM

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

colab.research.google.com/drive/13oU07TPPh3DO7ryVK8bopENFXKdF9Yb3?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

sample_data | House Price India.csv

```
[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

completed at 11:06 AM

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

colab.research.google.com/drive/13oU07TPPh3DO7ryVK8bopENFXKdF9Yb3?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 | RAM | Disk

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... | arch... | 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/13oU07TPPh3DO7ryVK8bopENFXKdF9Yb3?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 | RAM | Disk

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
```

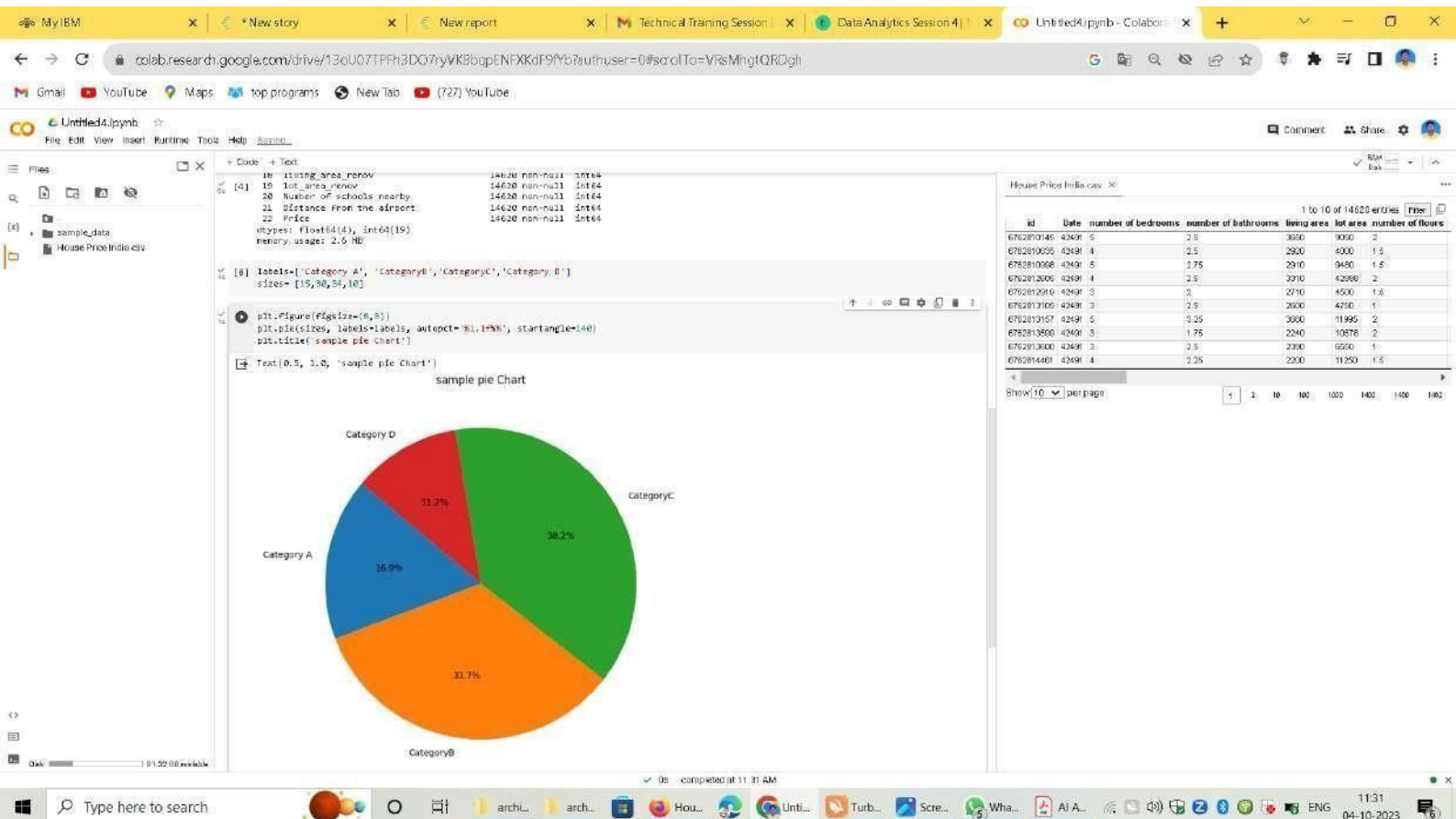
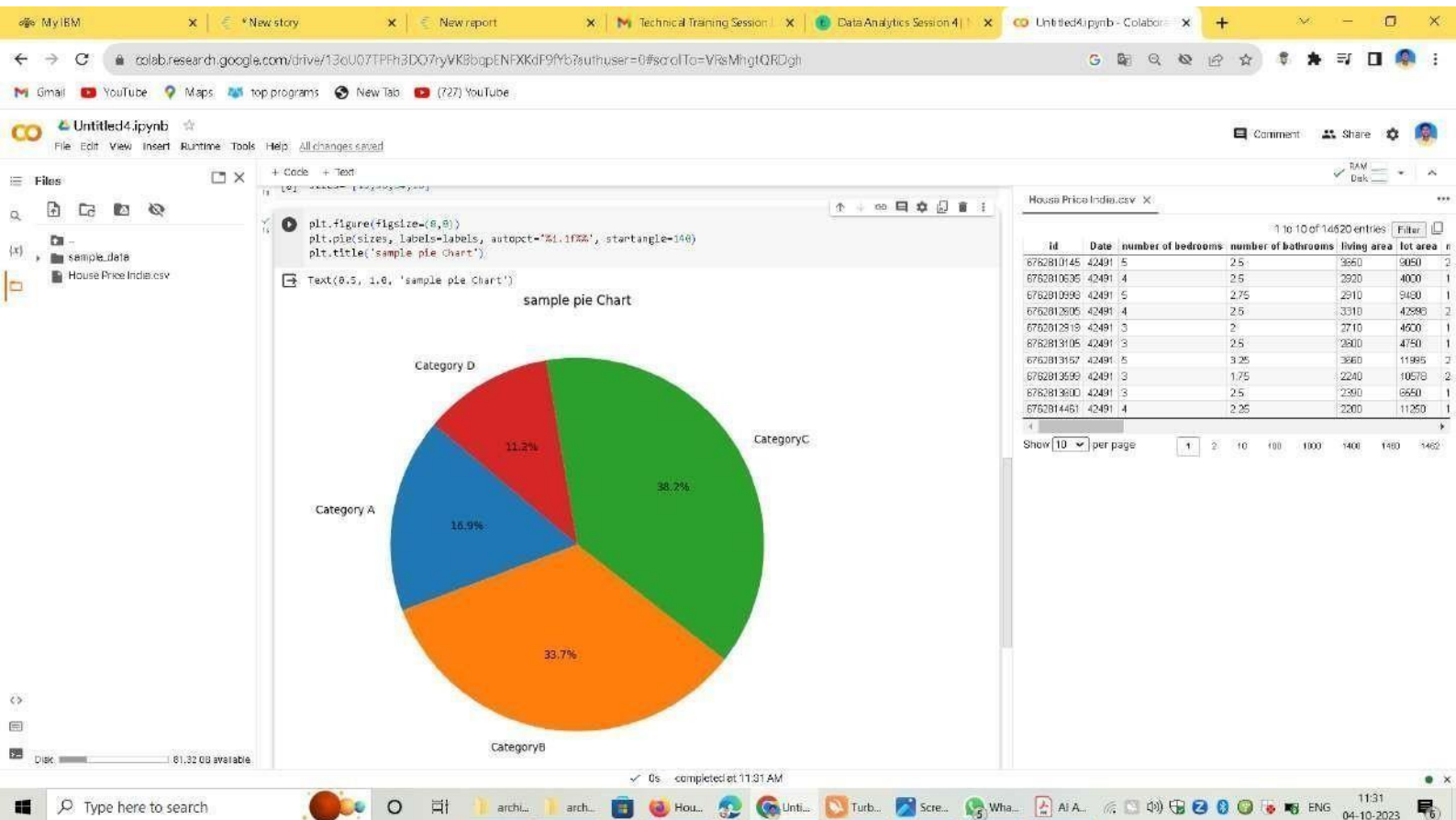
House Price India.csv | 1 to 10 of 14620 entries | Filter

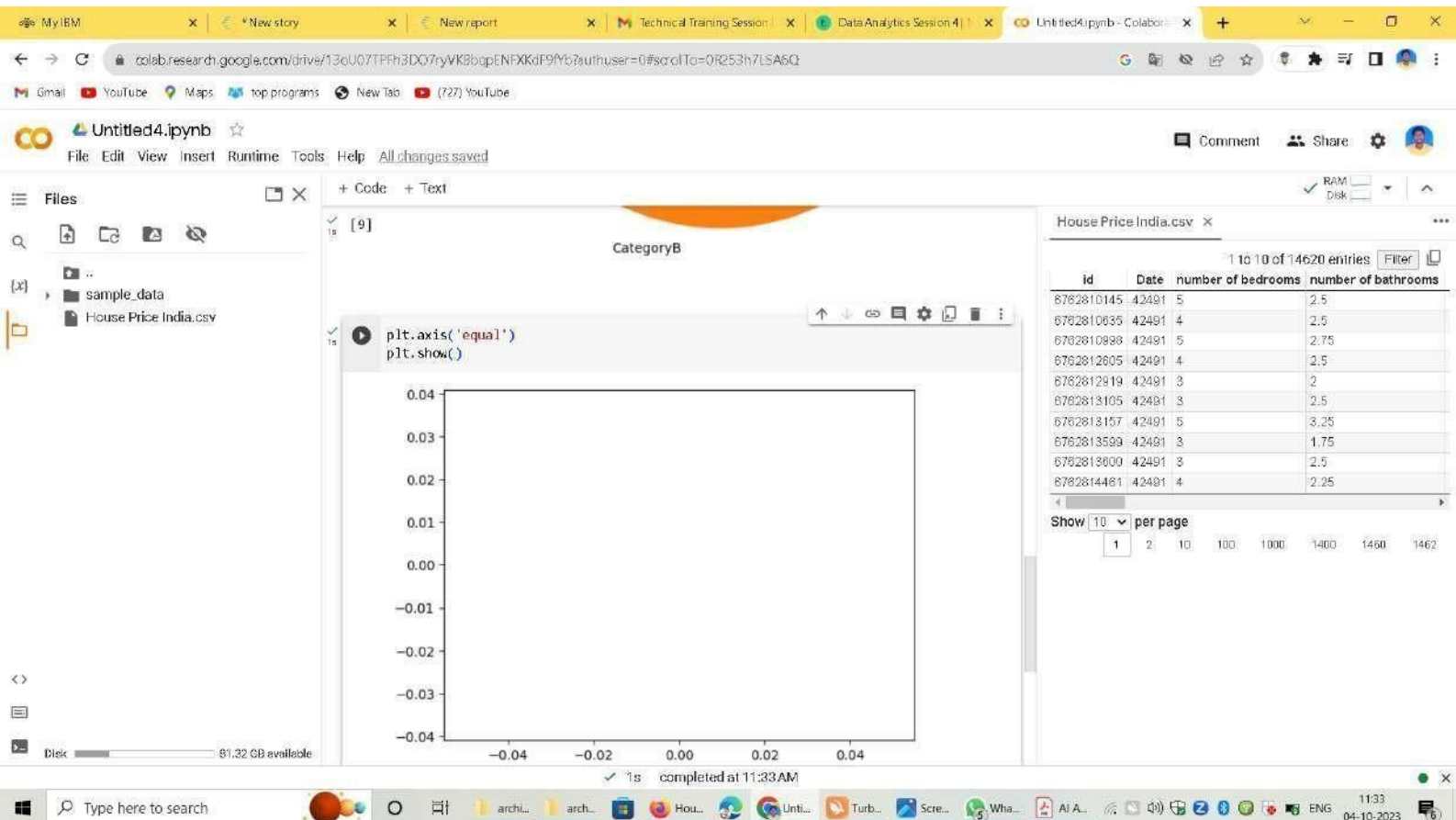
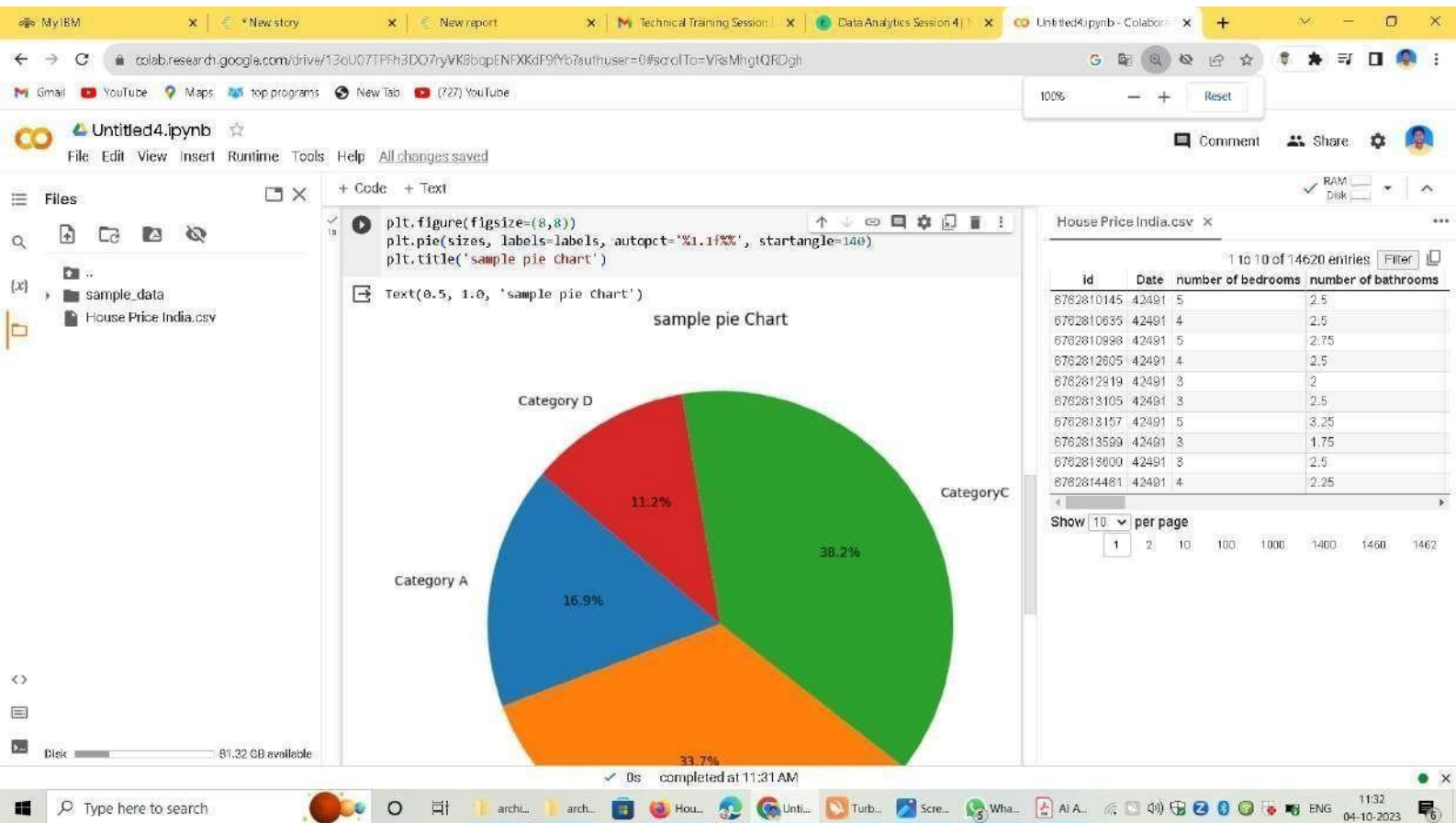
| id | Date | number of bedrooms | number of bathrooms |
|------------|-------|--------------------|---------------------|
| 6762810145 | 42491 | 5 | 2.5 |
| 6762810635 | 42491 | 4 | 2.5 |
| 6762810998 | 42491 | 5 | 2.75 |
| 6762812605 | 42491 | 4 | 2.5 |
| 6762812919 | 42491 | 3 | 2 |
| 6762813105 | 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... | arch... | 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/13cU07TPPh3DO7ryVK8bopENFXKdF9Yb3?authuser=0#scrollTo=XUOG_dHGSldP

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

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 | 9490 |
| 6762812605 | 42491 | 4 | 2.5 | 3310 | 42990 |
| 6762812519 | 42491 | 3 | 2 | 2710 | 4500 |
| 6762813105 | 42491 | 3 | 2.5 | 2800 | 4750 |
| 6762813157 | 42491 | 5 | 3.25 | 3660 | 11995 |
| 6762813599 | 42491 | 3 | 1.75 | 2240 | 10570 |
| 6762813600 | 42491 | 3 | 2.5 | 2390 | 6650 |
| 6762814461 | 42491 | 4 | 2.25 | 2200 | 11250 |

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

0s completed at 11:35 AM

Type here to search | archi... | 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/13cU07TPPh3DO7ryVK8bopENFXKdF9Yb3?authuser=0#scrollTo=XUOG_dHGSldP

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

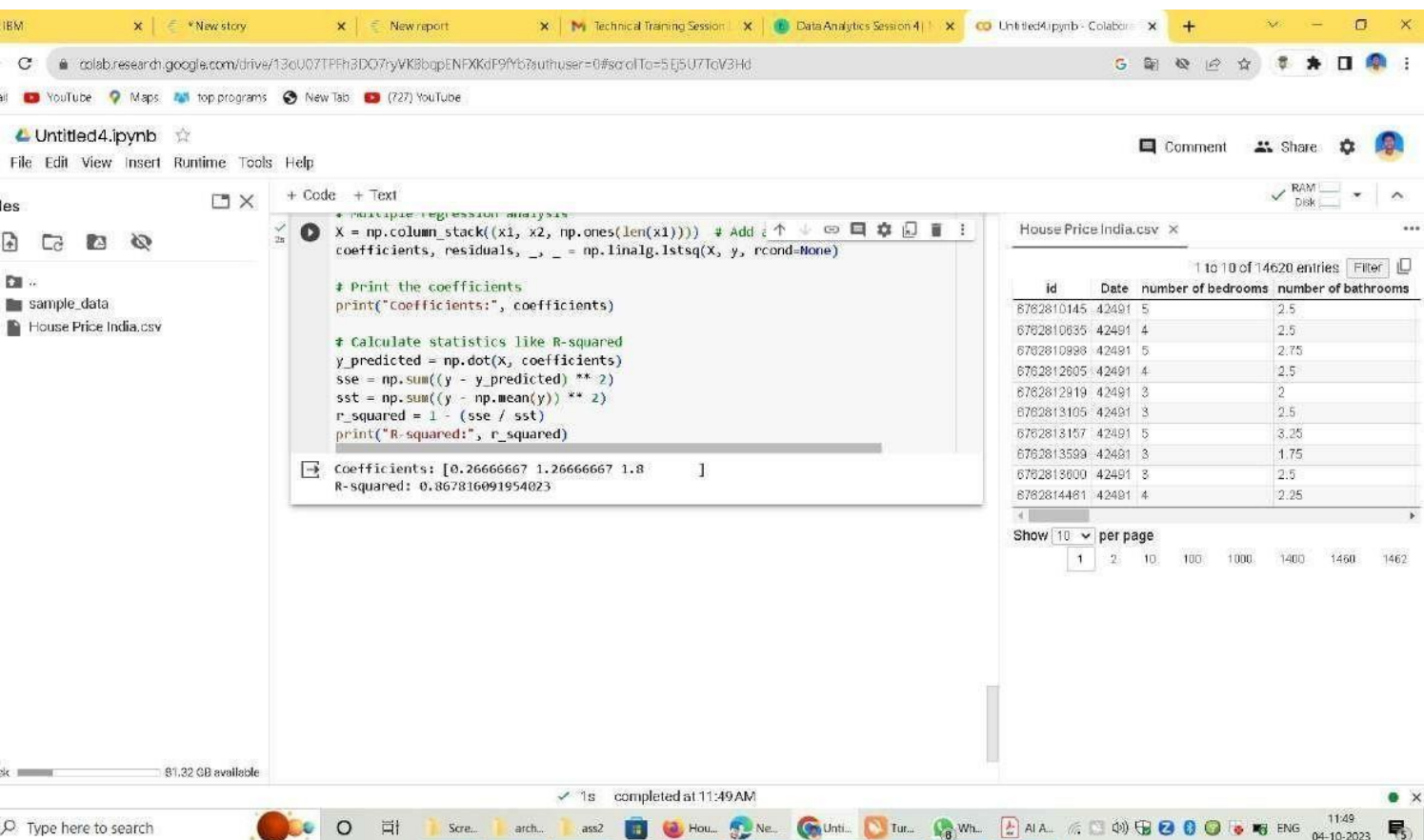
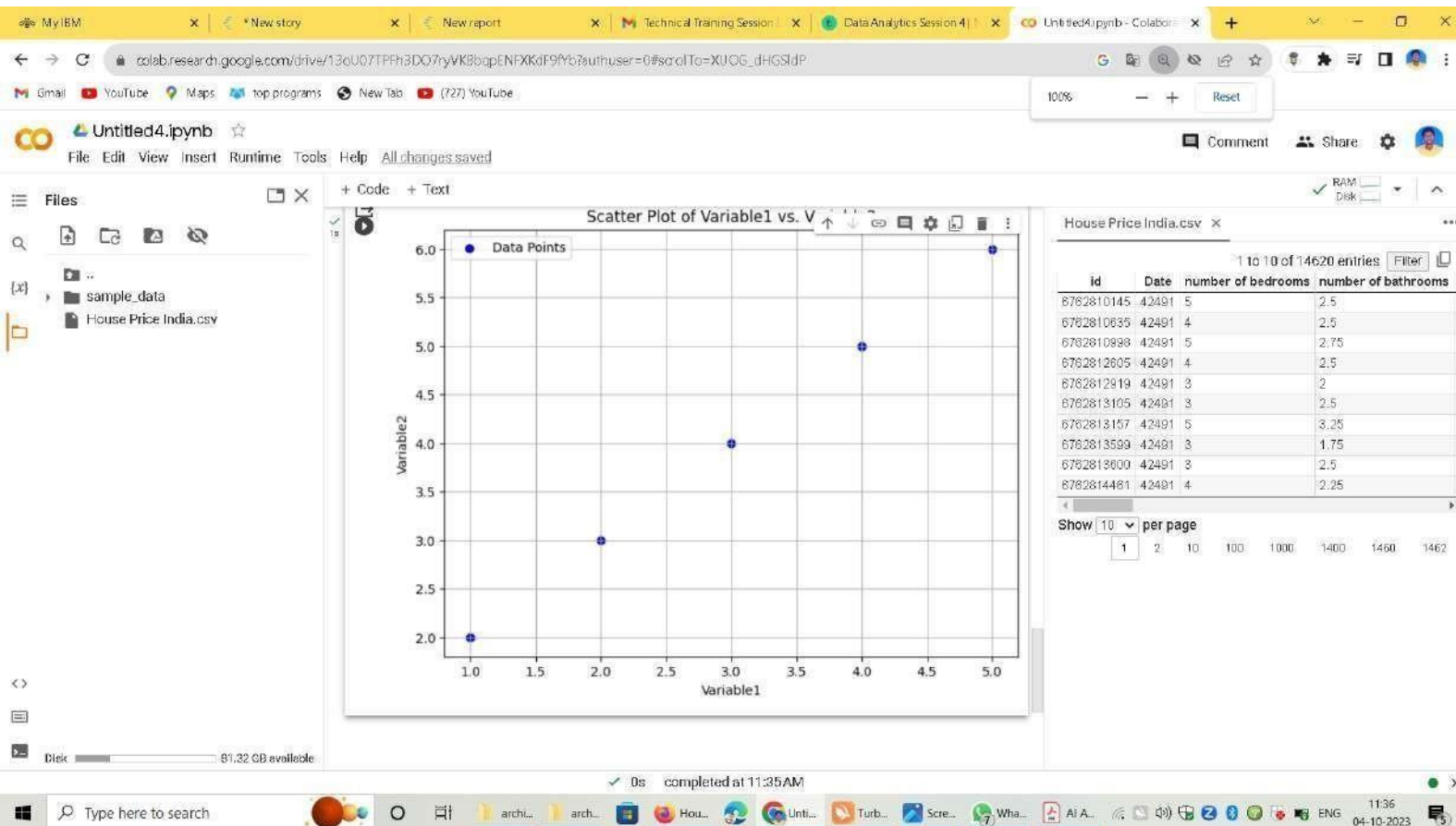
House Price India.csv | 1 to 10 of 14620 entries | Filter

| 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 |
| 6762812519 | 42491 | 3 | 2 |
| 6762813105 | 42491 | 3 | 2.5 |
| 6762813157 | 42491 | 5 | 3.25 |
| 6762813599 | 42491 | 3 | 1.75 |
| 6762813600 | 42491 | 3 | 2.5 |
| 6762814461 | 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 | archi... | 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/13oU07TPPh3DO7ryVK8bopENFXKdF9Yb3?authuser=0#scrollTo=5Ej5U7TeV3Hd

Gmail YouTube Maps top programs New Tab (7/7) YouTube

Untitled4.ipynb ☆

File Edit View Insert Runtime Tools Help

Comment Share

Files

- sample_data
- House Price India.csv

+ Code + Text

```
# Multiple Regression Analysis
X = np.column_stack((x1, x2, np.ones(len(x1)))) # Add a column of ones
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/13oU07TPPh3DO7ryVK8bopENFXKdF9Yb3?authuser=0#scrollTo=6Ydn7CqXIZI

Gmail YouTube Maps top programs New Tab (7/7) YouTube

Untitled4.ipynb ☆

File Edit View Insert Runtime Tools Help Saving...

Comment Share

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

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

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/13oU07TPPh3DO7ryVK8bopENFXKdF9fYb3?authuser=0#scrollTo=6Yrn7CqXLZl

Gmail YouTube Maps top programs New Tab (727) 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
```

```
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)
```

RAM Disk

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 |
| 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 11:55AM

Type here to search

Scree... arch... ass2

How... Ne... Unti... Tur... g Wh... AI A...

11:55 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/13oU07TPPh3DO7ryVK8bopENFXKdF9fYb3?authuser=0#scrollTo=1l0Q5HAsxm9-

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

+ 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
```

```
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
```

RAM Disk

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 |
| 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 12:00PM

Type here to search

Scree... arch... ass2

How... Ne... Unti... Tur... g Wh... AI A...

12:00 04-10-2023