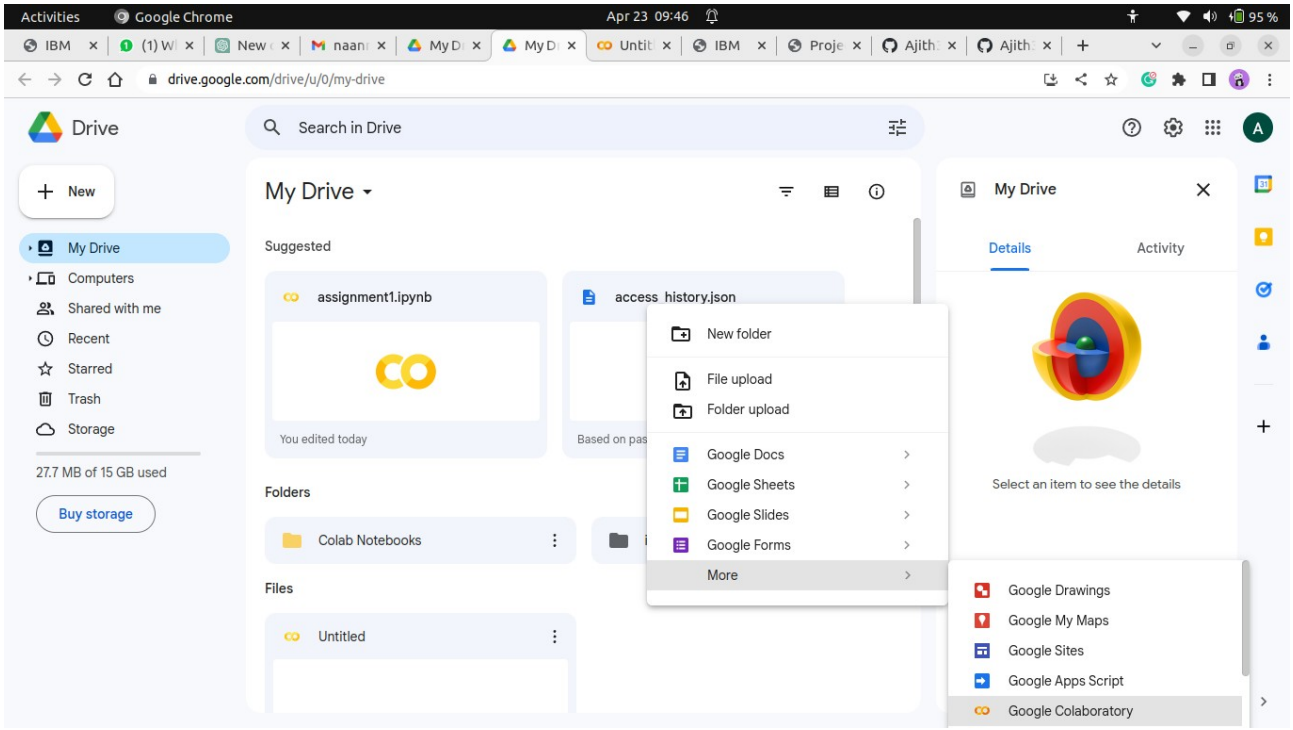
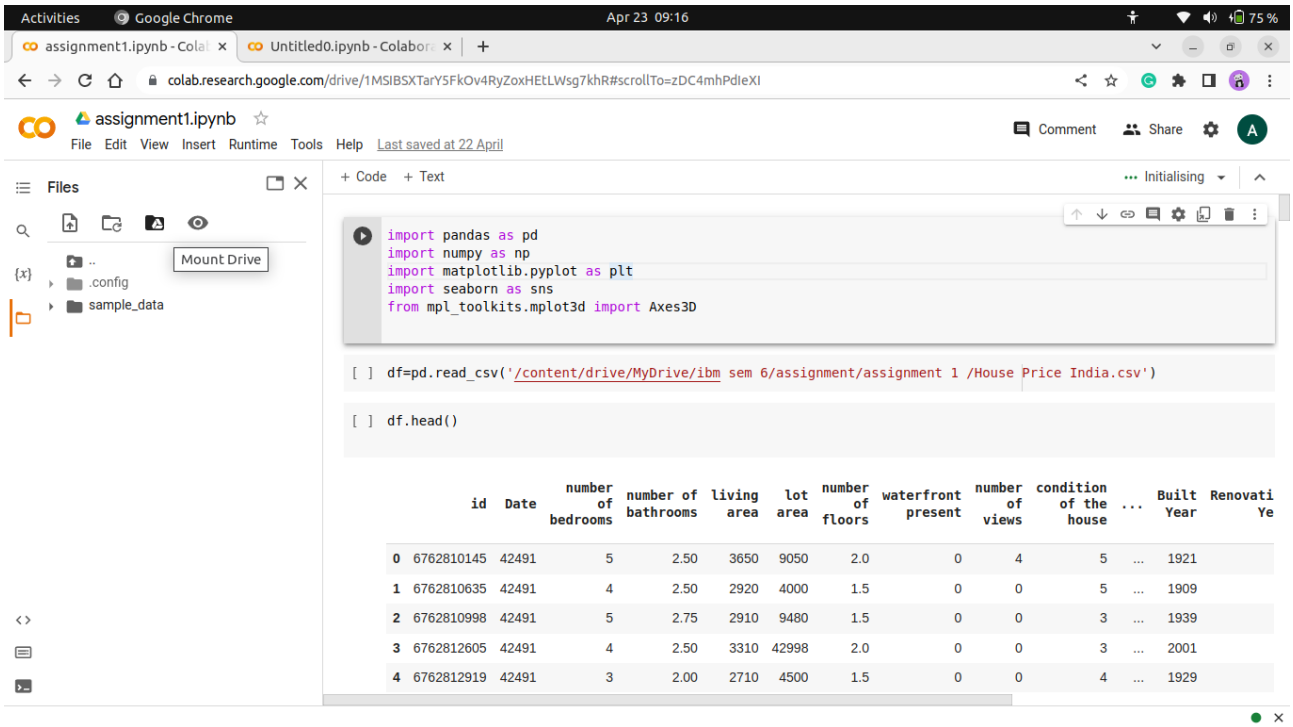
First thing open your google drive and create a folder and download the data set
save it on the folder
step2 :open google colab note book by right click your mouse or else go and search google colab also there are many ways to open it

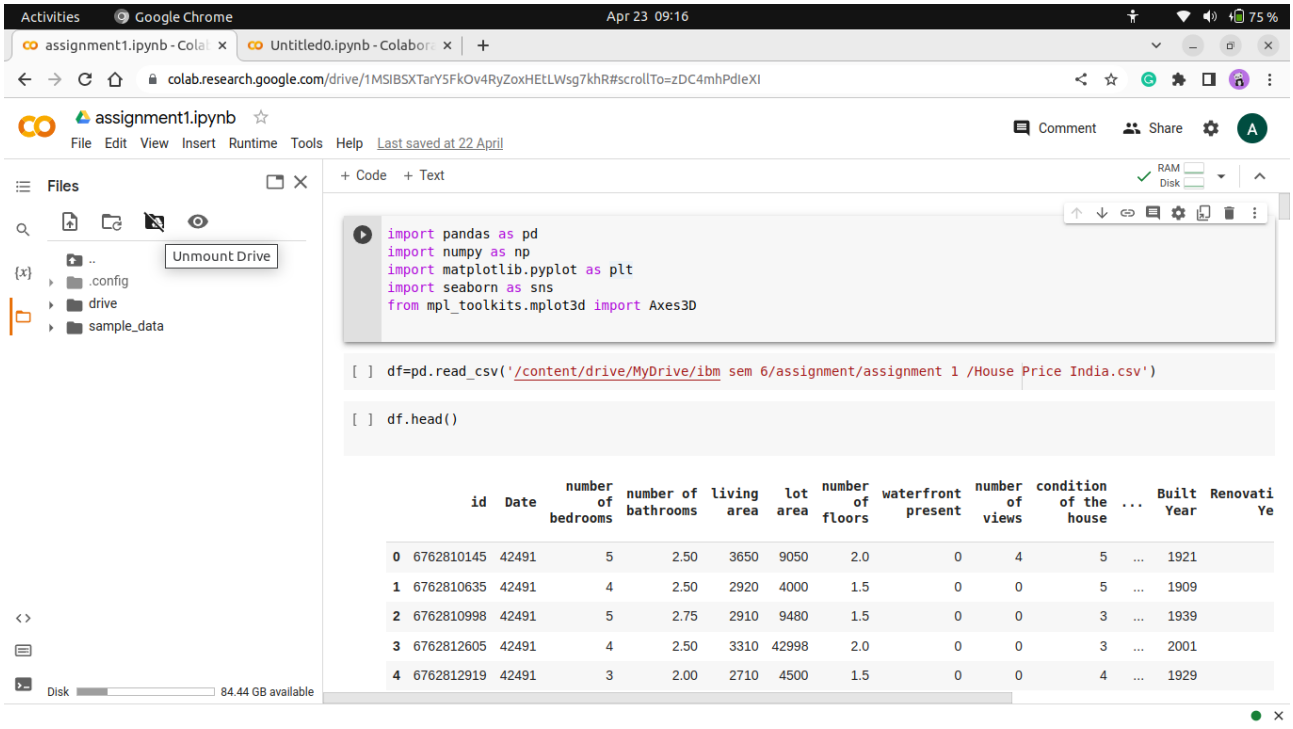


step 3: After open your google colab then change the name at initial your file name is untitled iam using assignemnt 1 as my file name and its ur wish

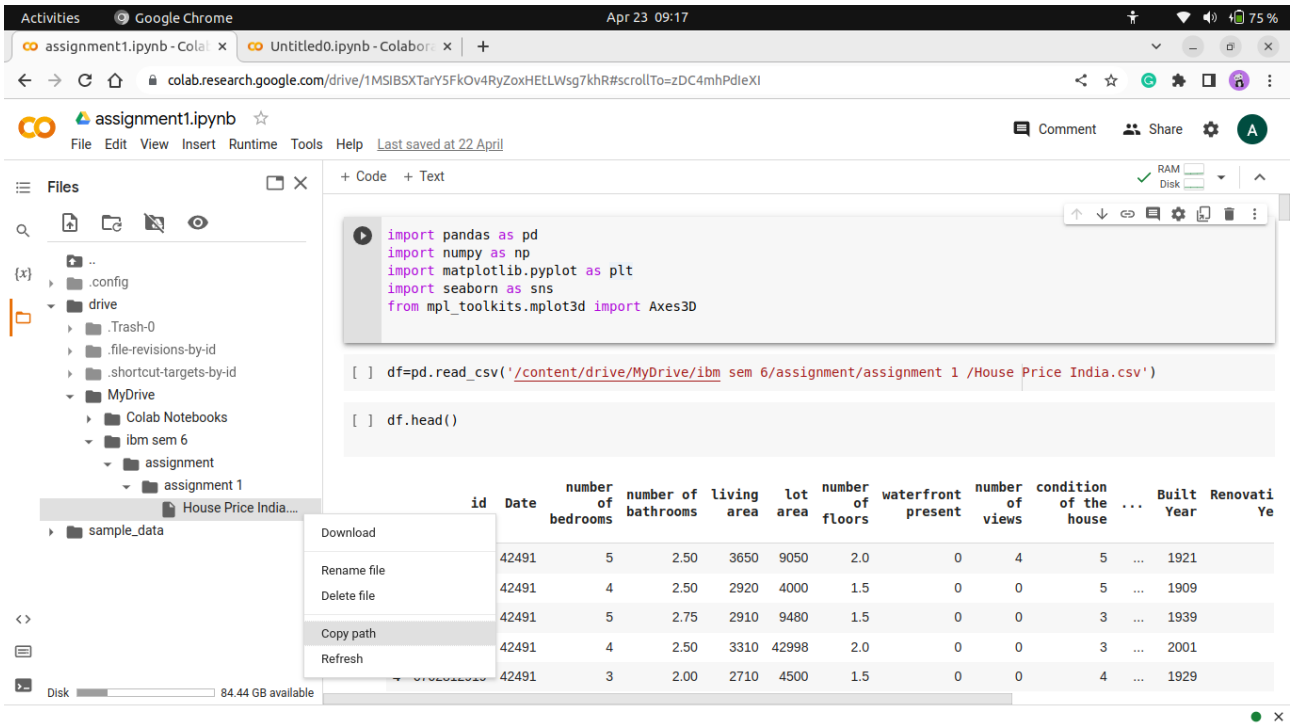


A

step 4: after name your file then you should add the data set by click the folder icon in the top left before doing that you mount your drive by simple click the mount drive folde icon and then refresh the folder now you can able to see your drive



step5; after mount your drive then open your data set and right click mouse then copy path



c

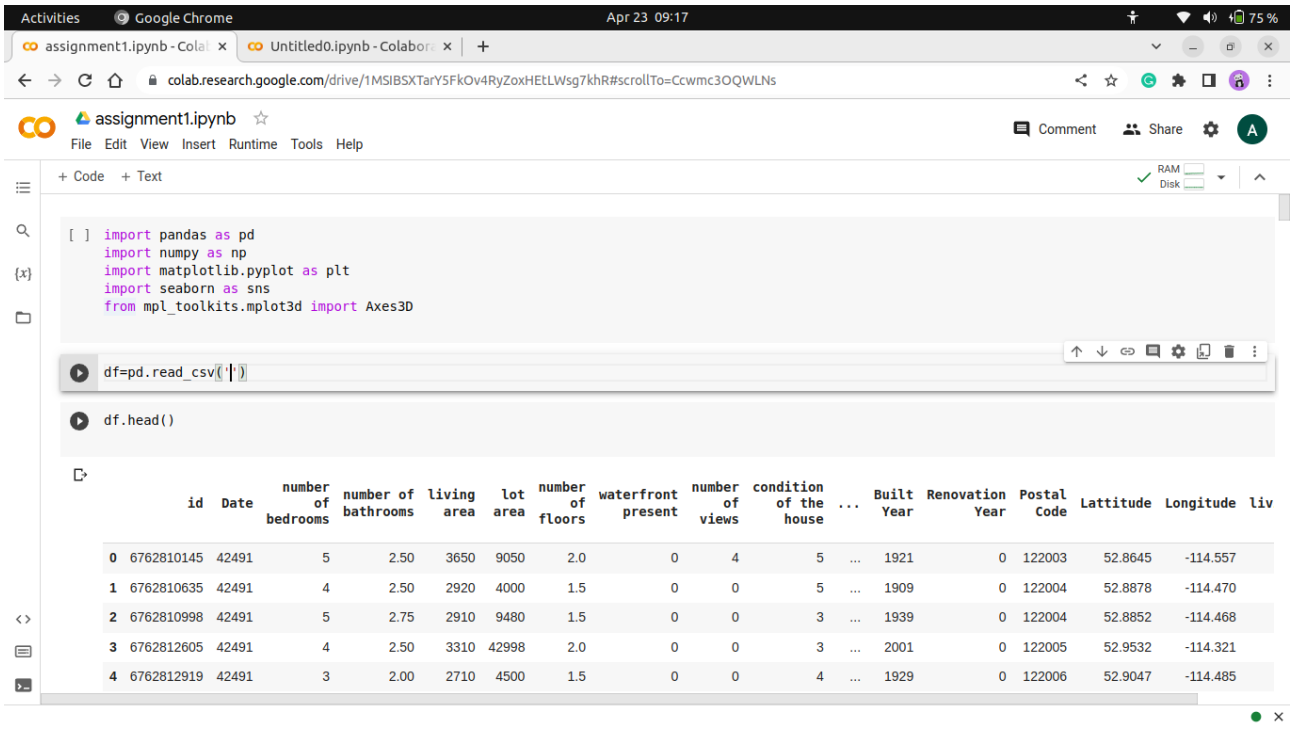
step 6: after that you import the all libraries and packeges you need to run your project

then use df=pd.read_csv(‘’) here df is variable name pd is pandas object name csv if the file formate insed the single cotation you enter your path

by create the code block press(ctrl+M+B)

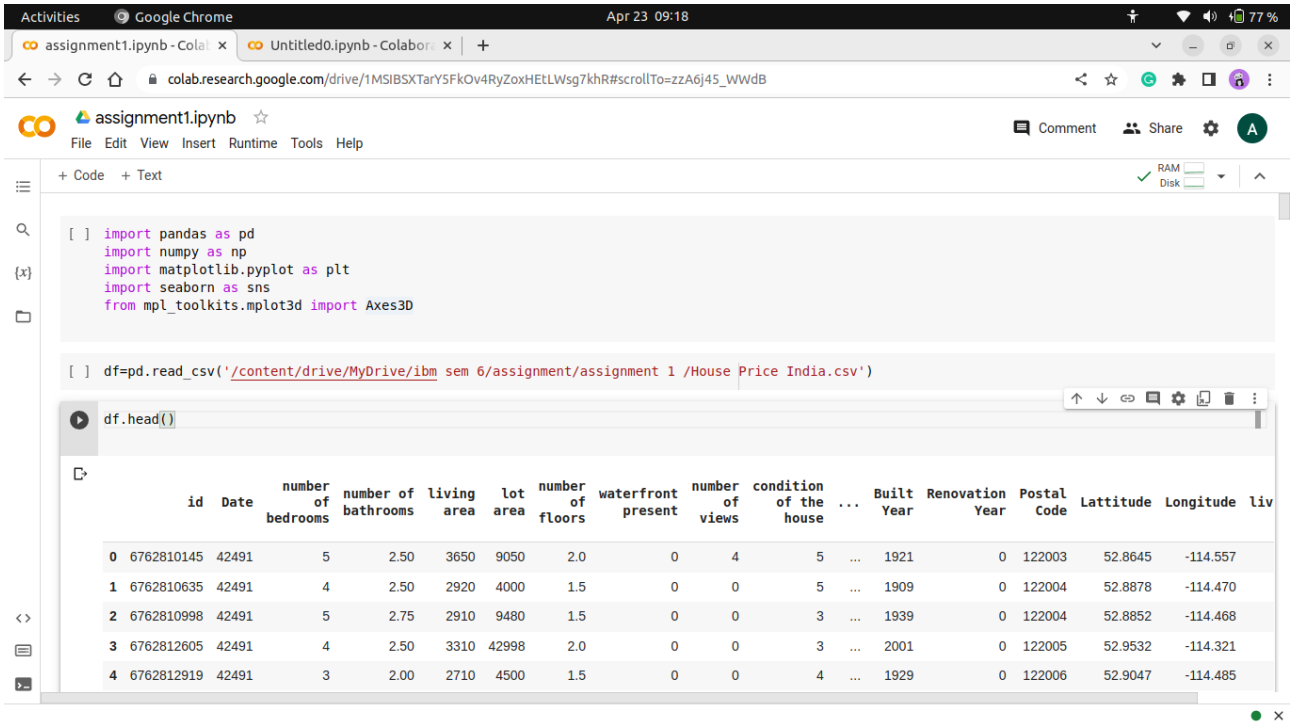
to execute the current block press(ctrl+enter)

to do both of them press (shift+enter)



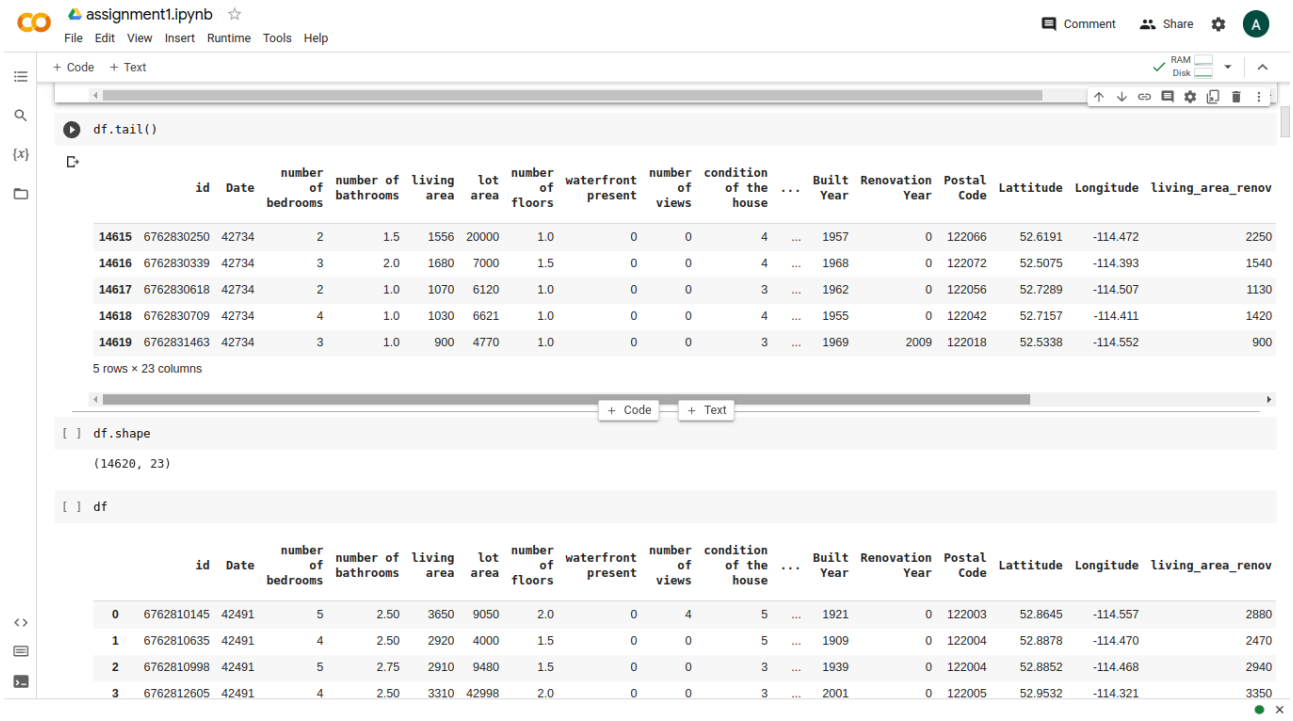
	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code	Lattitude	Longitude	liv
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.557	
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	

d



	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code	Lattitude	Longitude	liv
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.557	
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	

5



	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code	Lattitude	Longitude	living_area_renov
14615	6762830250	42734	2	1.5	1556	20000	1.0	0	0	4	...	1957	0	122066	52.6191	-114.472	2250
14616	6762830339	42734	3	2.0	1680	7000	1.5	0	0	4	...	1968	0	122072	52.5075	-114.393	1540
14617	6762830618	42734	2	1.0	1070	6120	1.0	0	0	3	...	1962	0	122056	52.7289	-114.507	1130
14618	6762830709	42734	4	1.0	1030	6621	1.0	0	0	4	...	1955	0	122042	52.7157	-114.411	1420
14619	6762831463	42734	3	1.0	900	4770	1.0	0	0	3	...	1969	2009	122018	52.5338	-114.552	900

5 rows × 23 columns

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code	Lattitude	Longitude	living_area_renov
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.557	2880
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	2470
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	2940
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	3350

6

assignment1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share A

+ Code + Text

RAM Disk

```
[ ] df.shape
```

```
(14620, 23)
```

```
df
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code	Lattitude	Longitude	living_area_renov
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.557	2880
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	2470
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	2940
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	3350
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	2060
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14615	6762830250	42734	2	1.50	1556	20000	1.0	0	0	4	...	1957	0	122066	52.6191	-114.472	2250
14616	6762830339	42734	3	2.00	1680	7000	1.5	0	0	4	...	1968	0	122072	52.5075	-114.393	1540
14617	6762830618	42734	2	1.00	1070	6120	1.0	0	0	3	...	1962	0	122056	52.7289	-114.507	1130
14618	6762830709	42734	4	1.00	1030	6621	1.0	0	0	4	...	1955	0	122042	52.7157	-114.411	1420
14619	6762831463	42734	3	1.00	900	4770	1.0	0	0	3	...	1969	2009	122018	52.5338	-114.552	900

14620 rows × 23 columns

Univariate Analysis

7

assignment1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share A

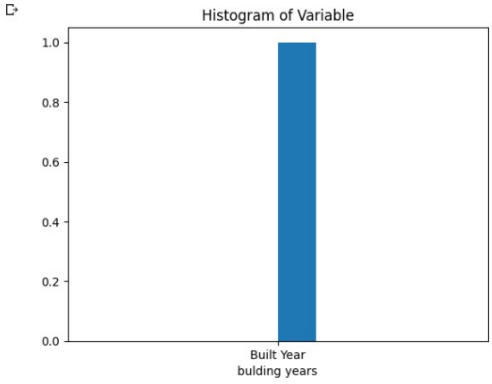
+ Code + Text

RAM Disk

Univariate Analysis

a) Histogram:

```
plt.hist(['Built Year'], bins=10)
plt.title('Histogram of Variable')
plt.xlabel('bulding years')
plt.ylabel('')
plt.show()
```



8

assignment1.ipynb

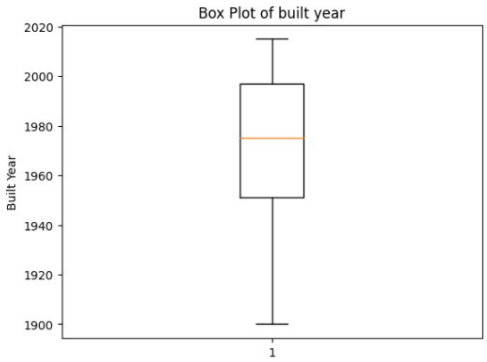
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

building years

```
[ ]
    plt.boxplot(df['Built Year'])
    plt.title('Box Plot of built year')
    plt.ylabel('Built Year')
    plt.show()
```




```
    df_01=df.loc[df['Built Year']==2001]
    df_55=df.loc[df['Built Year']==1955]
    df_69=df.loc[df['Built Year']==1969]
```

9

assignment1.ipynb

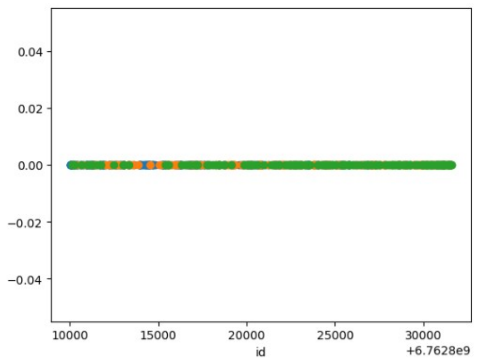
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

```
    df_01=df.loc[df['Built Year']==2001]
    df_55=df.loc[df['Built Year']==1955]
    df_69=df.loc[df['Built Year']==1969]
```

```
    plt.plot(df_01['id'],np.zeros_like(df_01['id']),'o')
    plt.plot(df_55['id'],np.zeros_like(df_55['id']),'o')
    plt.plot(df_69['id'],np.zeros_like(df_69['id']),'o')
    plt.xlabel('id')
    plt.show()
```



b) Box Plot:

10

assignment1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share A

+ Code + Text

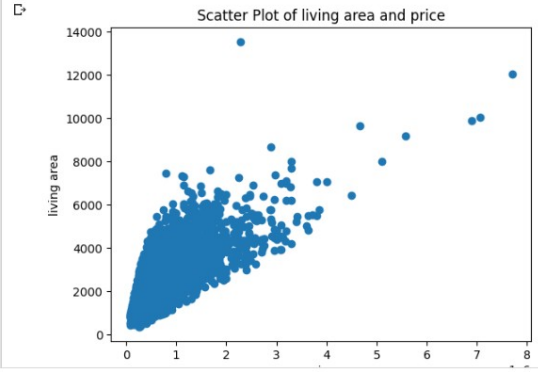
RAM Disk

id +6.7628e9

Bivariate Analysis

a) Scatter Plot:

```
plt.scatter(df['Price'], df['living area'])
plt.title('Scatter Plot of living area and price')
plt.xlabel('price')
plt.ylabel('living area')
plt.show()
```



11

assignment1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share A

+ Code + Text

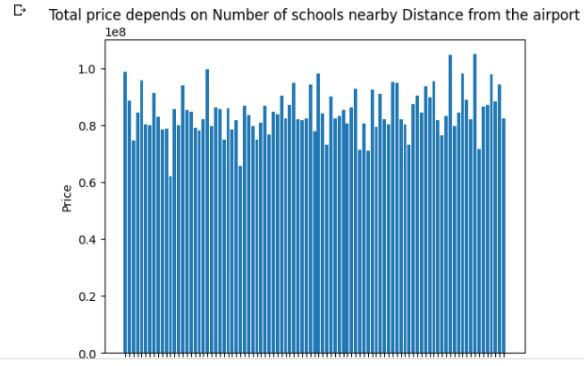
RAM Disk

0 1 2 3 4 5 6 7 8
price 1e6

b) Bar Plot:

```
sales_by_region = df.groupby(['Number of schools nearby', 'Distance from the airport'])['Price'].sum().reset_index()

plt.bar(sales_by_region['Number of schools nearby'].astype(str)  + ' - ' + sales_by_region['Distance from the airport'].astype(str) , sales_by_region['Price'])
plt.title('Total price depends on Number of schools nearby Distance from the airport')
plt.xlabel('Number of schools nearby Distance from the airport')
plt.ylabel('Price')
plt.xticks(rotation=90)
plt.show()
```



12

assignment1.ipynb

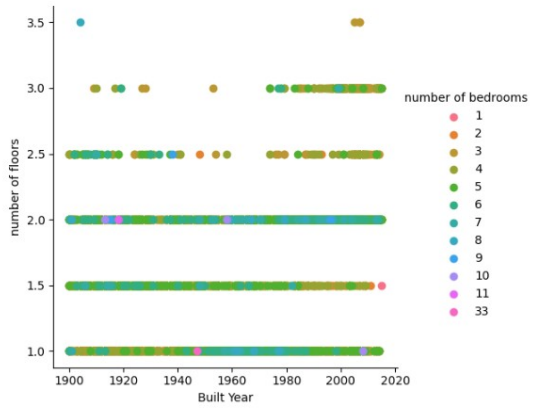
File Edit View Insert Runtime Tools Help All changes saved

Comment Share A

+ Code + Text

RAM Disk

```
sns.FacetGrid(df,hue="number of bedrooms",height=5).map(plt.scatter,"Built Year","number of floors").add_legend();
plt.show()
```



Multivariate Analysis

a) Heat Map:

13

assignment1.ipynb

File Edit View Insert Runtime Tools Help Saving...

Comment Share A

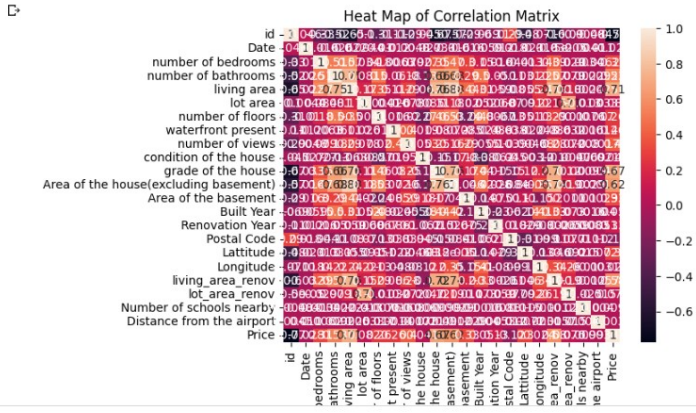
+ Code + Text

RAM Disk

Multivariate Analysis

a) Heat Map:

```
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True)
plt.title('Heat Map of Correlation Matrix')
plt.show()
```



14

assignment1.ipynb

File Edit View Insert Runtime Tools Help Saving...

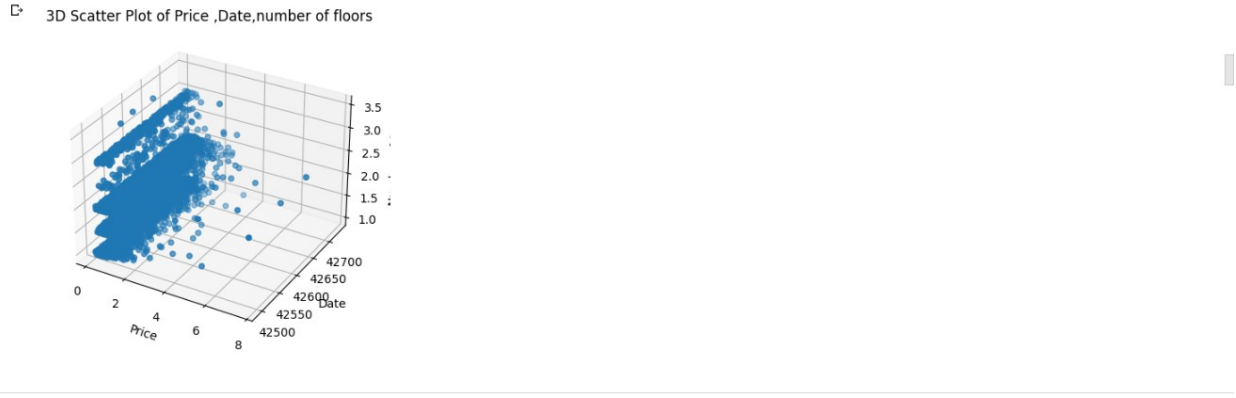
Comment Share A

+ Code + Text

RAM Disk

b) 3D Scatter Plot:

```
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(df['Price'], df['Date'], df['number of floors'])
ax.set_xlabel('Price')
ax.set_ylabel('Date')
ax.set_zlabel('Number of floors')
plt.title('3D Scatter Plot of Price ,Date,number of floors')
plt.show()
```



15

assignment1.ipynb

File Edit View Insert Runtime Tools Help Saving...

Comment Share A

+ Code + Text

RAM Disk

8

descriptive statistics

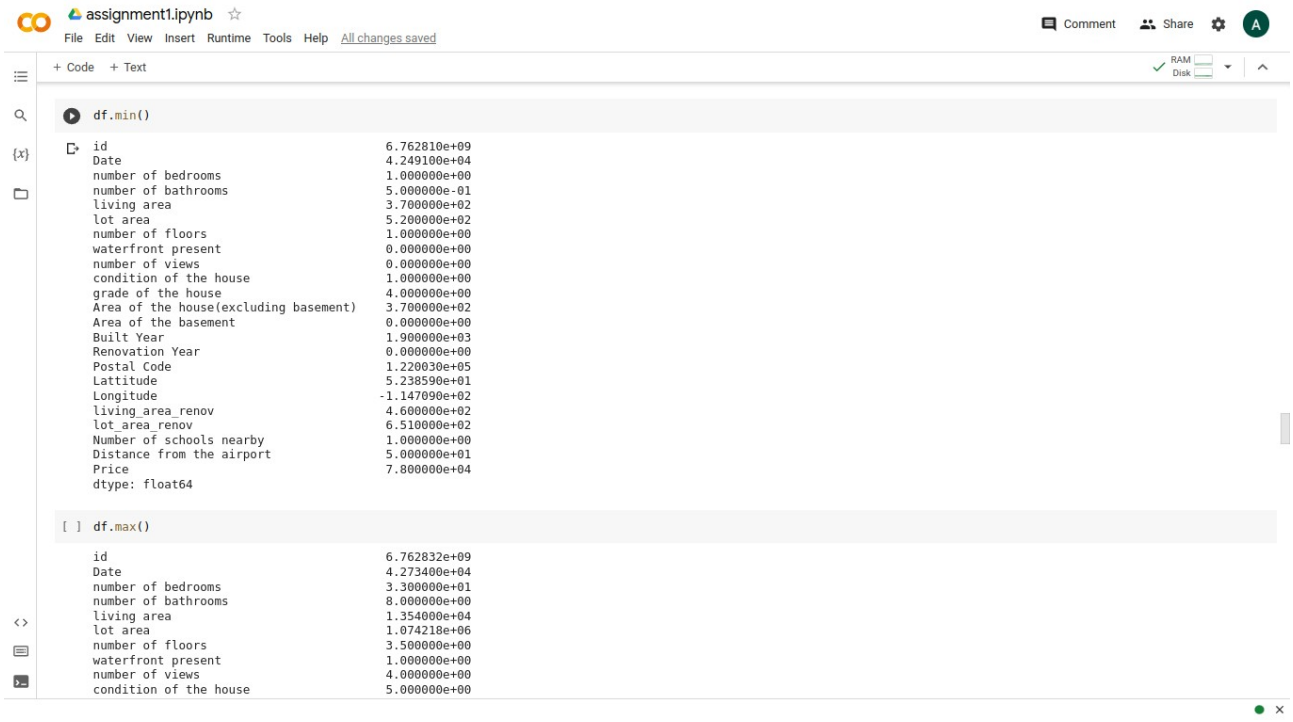
```
df.count()
```

```
id                                    14620
Date                                  14620
number of bedrooms                    14620
number of bathrooms                   14620
living area                           14620
lot area                              14620
number of floors                      14620
waterfront present                    14620
number of views                       14620
condition of the house                14620
grade of the house                    14620
Area of the house(excluding basement) 14620
Area of the basement                  14620
Built Year                            14620
Renovation Year                       14620
Postal Code                           14620
Lattitude                             14620
Longitude                             14620
living_area_renov                     14620
lot_area_renov                        14620
Number of schools nearby              14620
Distance from the airport             14620
Price                                 14620
dtype: int64
```

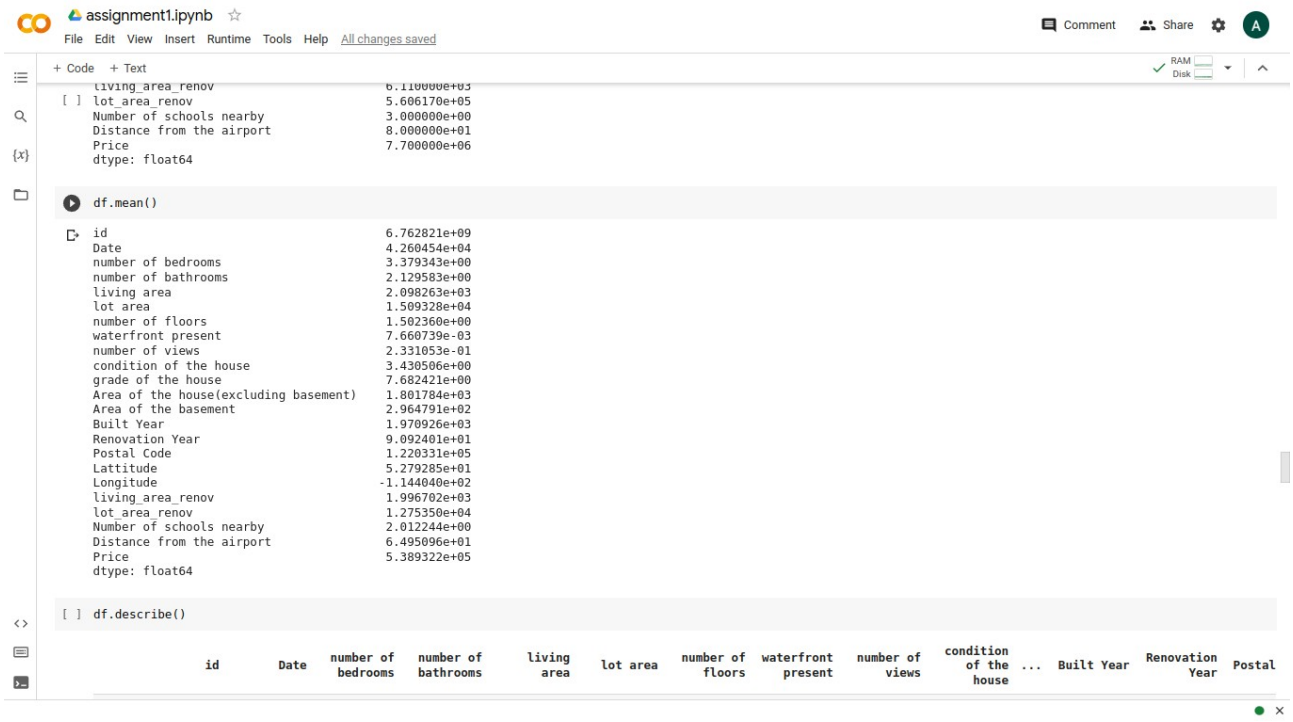
```
[ ] df.min()
```

```
id                                    6.762810e+09
Date                                  4.249100e+04
number of bedrooms                    1.000000e+00
number of bathrooms                   5.000000e-01
```

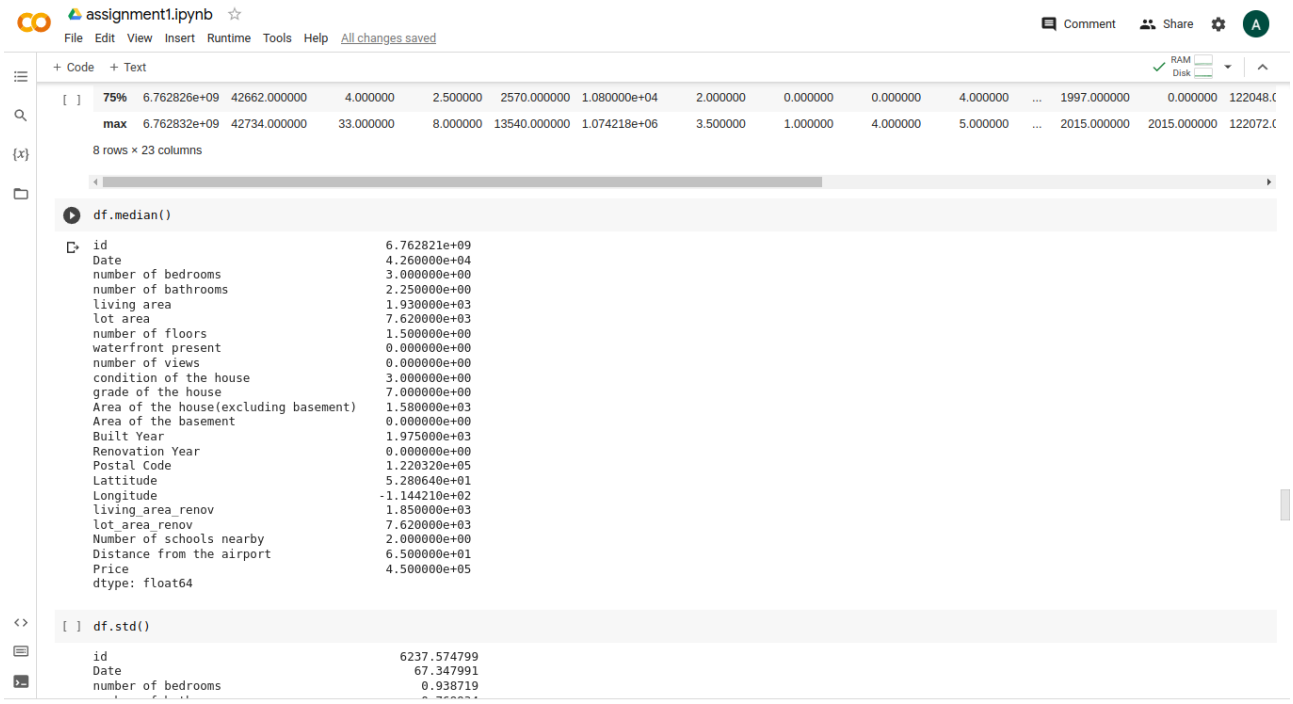
16


assignment1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
df.min()
id 6.762810e+09
Date 4.249100e+04
number of bedrooms 1.000000e+00
number of bathrooms 5.000000e-01
living area 3.700000e+02
lot area 5.200000e+02
number of floors 1.000000e+00
waterfront present 0.000000e+00
number of views 0.000000e+00
condition of the house 1.000000e+00
grade of the house 4.000000e+00
Area of the house(excluding basement) 3.700000e+02
Area of the basement 0.000000e+00
Built Year 1.900000e+03
Renovation Year 0.000000e+00
Postal Code 1.220030e+05
Lattitude 5.238590e+01
Longitude -1.147090e+02
living_area_renov 4.600000e+02
lot_area_renov 6.510000e+02
Number of schools nearby 1.000000e+00
Distance from the airport 5.000000e+01
Price 7.800000e+04
dtype: float64
df.max()
id 6.762832e+09
Date 4.273400e+04
number of bedrooms 3.300000e+01
number of bathrooms 8.000000e+00
living area 1.354000e+04
lot area 1.074218e+06
number of floors 3.500000e+00
waterfront present 1.000000e+00
number of views 4.000000e+00
condition of the house 5.000000e+00

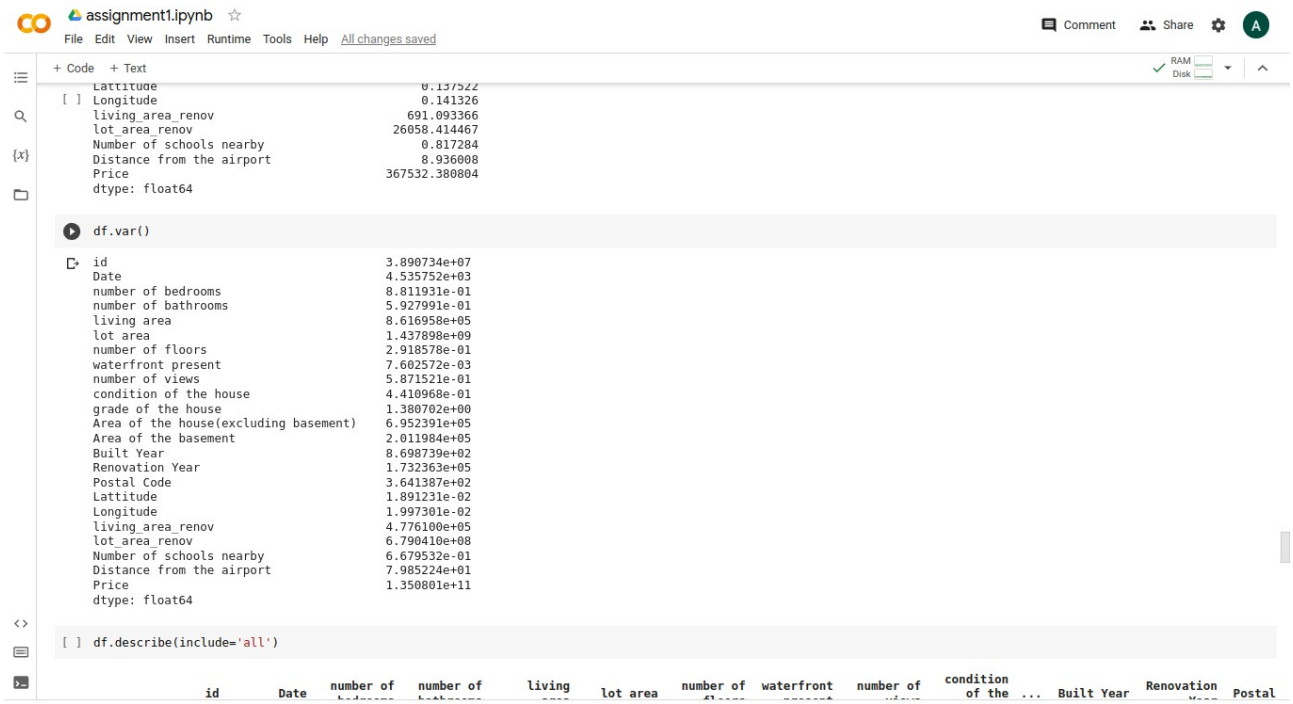
17


assignment1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
lot_area_renov 5.606170e+05
Number of schools nearby 3.000000e+00
Distance from the airport 8.000000e+01
Price 7.700000e+06
dtype: float64
df.mean()
id 6.762821e+09
Date 4.260454e+04
number of bedrooms 3.379343e+00
number of bathrooms 2.129583e+00
living area 2.098263e+03
lot area 1.509328e+04
number of floors 1.502360e+00
waterfront present 7.660739e-03
number of views 2.331053e-01
condition of the house 3.430506e+00
grade of the house 7.682421e+00
Area of the house(excluding basement) 1.801784e+03
Area of the basement 2.964791e+02
Built Year 1.970926e+03
Renovation Year 9.092401e+01
Postal Code 1.220331e+05
Lattitude 5.279285e+01
Longitude -1.144040e+02
living_area_renov 1.996702e+03
lot_area_renov 1.275350e+04
Number of schools nearby 2.012244e+00
Distance from the airport 6.495096e+01
Price 5.389322e+05
dtype: float64
df.describe()
id Date number of bedrooms number of bathrooms living area lot area number of floors waterfront present number of views condition of the house ... Built Year Renovation Year Postal

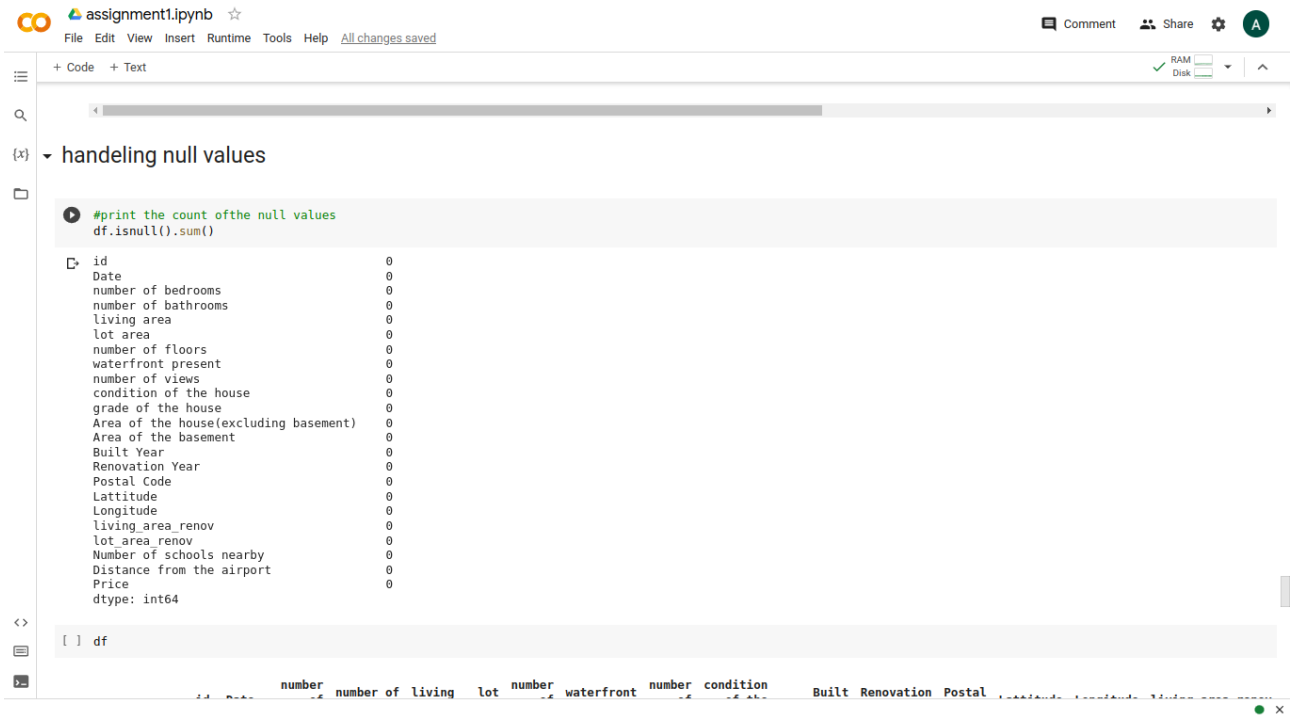
18


assignment1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
75% 6.762826e+09 42662.000000 4.000000 2.500000 2570.000000 1.080000e+04 2.000000 0.000000 0.000000 4.000000 ... 1997.000000 0.000000 122048.0
max 6.762832e+09 42734.000000 33.000000 8.000000 13540.000000 1.074218e+06 3.500000 1.000000 4.000000 5.000000 ... 2015.000000 2015.000000 122072.0
8 rows × 23 columns
df.median()
id 6.762821e+09
Date 4.260000e+04
number of bedrooms 3.000000e+00
number of bathrooms 2.250000e+00
living area 1.930000e+03
lot area 7.620000e+03
number of floors 1.500000e+00
waterfront present 0.000000e+00
number of views 0.000000e+00
condition of the house 3.000000e+00
grade of the house 7.000000e+00
Area of the house(excluding basement) 1.580000e+03
Area of the basement 0.000000e+00
Built Year 1.975000e+03
Renovation Year 0.000000e+00
Postal Code 1.220320e+05
Lattitude 5.280640e+01
Longitude -1.144210e+02
living_area_renov 1.850000e+03
lot_area_renov 7.620000e+03
Number of schools nearby 2.000000e+00
Distance from the airport 6.500000e+01
Price 4.500000e+05
dtype: float64
df.std()
id 6237.574799
Date 67.347991
number of bedrooms 0.938719

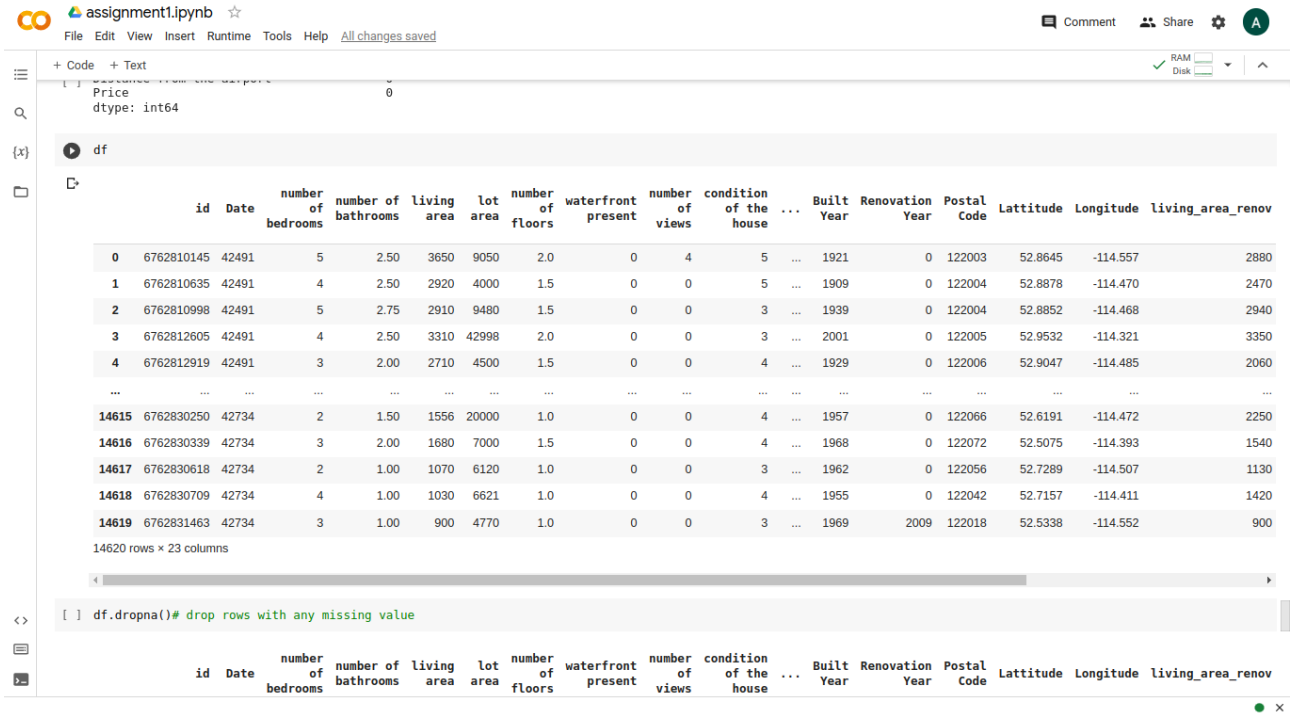
19


assignment1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Lattitude 0.137522
Longitude 0.141326
living_area_renov 691.093366
lot_area_renov 26058.414467
Number of schools nearby 0.817284
Distance from the airport 8.936008
Price 367532.380804
dtype: float64
df.var()
id 3.890734e+07
Date 4.535752e+03
number of bedrooms 8.811931e-01
number of bathrooms 5.927991e-01
living area 8.616958e+05
lot area 1.437898e+09
number of floors 2.918578e-01
waterfront present 7.602572e-03
number of views 5.871521e-01
condition of the house 4.410968e-01
grade of the house 1.380702e+00
Area of the house(excluding basement) 6.952391e+05
Area of the basement 2.011984e+05
Built Year 8.698739e+02
Renovation Year 1.732363e+05
Postal Code 3.641387e+02
Lattitude 1.891231e-02
Longitude 1.997301e-02
living_area_renov 4.776100e+05
lot_area_renov 6.790410e+08
Number of schools nearby 6.679532e-01
Distance from the airport 7.985224e+01
Price 1.350801e+11
dtype: float64
df.describe(include='all')
id Date number of bedrooms number of bathrooms living area lot area number of floors waterfront present number of views condition of the house ... Built Year Renovation Year Postal

20


assignment1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
handeling null values
#print the count ofthe null values
df.isnull().sum()
id 0
Date 0
number of bedrooms 0
number of bathrooms 0
living area 0
lot area 0
number of floors 0
waterfront present 0
number of views 0
condition of the house 0
grade of the house 0
Area of the house(excluding basement) 0
Area of the basement 0
Built Year 0
Renovation Year 0
Postal Code 0
Lattitude 0
Longitude 0
living_area_renov 0
lot_area_renov 0
Number of schools nearby 0
Distance from the airport 0
Price 0
dtype: int64
[ ] df

21


assignment1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Price 0
dtype: int64
df
id Date number of bedrooms number of bathrooms living area lot area number of floors waterfront present number of views condition of the house ... Built Year Renovation Year Postal Code Lattitude Longitude living_area_renov
0 6762810145 42491 5 2.50 3650 9050 2.0 0 4 5 ... 1921 0 122003 52.8645 -114.557 2880
1 6762810635 42491 4 2.50 2920 4000 1.5 0 0 5 ... 1909 0 122004 52.8878 -114.470 2470
2 6762810998 42491 5 2.75 2910 9480 1.5 0 0 3 ... 1939 0 122004 52.8852 -114.468 2940
3 6762812605 42491 4 2.50 3310 42998 2.0 0 0 3 ... 2001 0 122005 52.9532 -114.321 3350
4 6762812919 42491 3 2.00 2710 4500 1.5 0 0 4 ... 1929 0 122006 52.9047 -114.485 2060
14615 6762830250 42734 2 1.50 1556 20000 1.0 0 0 4 ... 1957 0 122066 52.6191 -114.472 2250
14616 6762830339 42734 3 2.00 1680 7000 1.5 0 0 4 ... 1968 0 122072 52.5075 -114.393 1540
14617 6762830618 42734 2 1.00 1070 6120 1.0 0 0 3 ... 1962 0 122056 52.7289 -114.507 1130
14618 6762830709 42734 4 1.00 1030 6621 1.0 0 0 4 ... 1955 0 122042 52.7157 -114.411 1420
14619 6762831463 42734 3 1.00 900 4770 1.0 0 0 3 ... 1969 2009 122018 52.5338 -114.552 900
14620 rows × 23 columns
[ ] df.dropna()# drop rows with any missing value
id Date number of bedrooms number of bathrooms living area lot area number of floors waterfront present number of views condition of the house ... Built Year Renovation Year Postal Code Lattitude Longitude living_area_renov

22

[ ] 14620 rows × 23 columns

```
df.dropna()# drop rows with any missing value
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code	Lattitude	Longitude	living_area_renov
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.557	2880
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	2470
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	2940
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	3350
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	2060
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14615	6762830250	42734	2	1.50	1556	20000	1.0	0	0	4	...	1957	0	122066	52.6191	-114.472	2250
14616	6762830339	42734	3	2.00	1680	7000	1.5	0	0	4	...	1968	0	122072	52.5075	-114.393	1540
14617	6762830618	42734	2	1.00	1070	6120	1.0	0	0	3	...	1962	0	122056	52.7289	-114.507	1130
14618	6762830709	42734	4	1.00	1030	6621	1.0	0	0	4	...	1955	0	122042	52.7157	-114.411	1420
14619	6762831463	42734	3	1.00	900	4770	1.0	0	0	3	...	1969	2009	122018	52.5338	-114.552	900

14620 rows × 23 columns

[ ]
```
df.dropna(axis=1)# drop columns with any missing value
```

	id	Date	number of	number of bathrooms	living area	lot area	number of	waterfront present	number of	condition of the	...	Built Year	Renovation Year	Postal Code	Lattitude	Longitude	living_area_renov