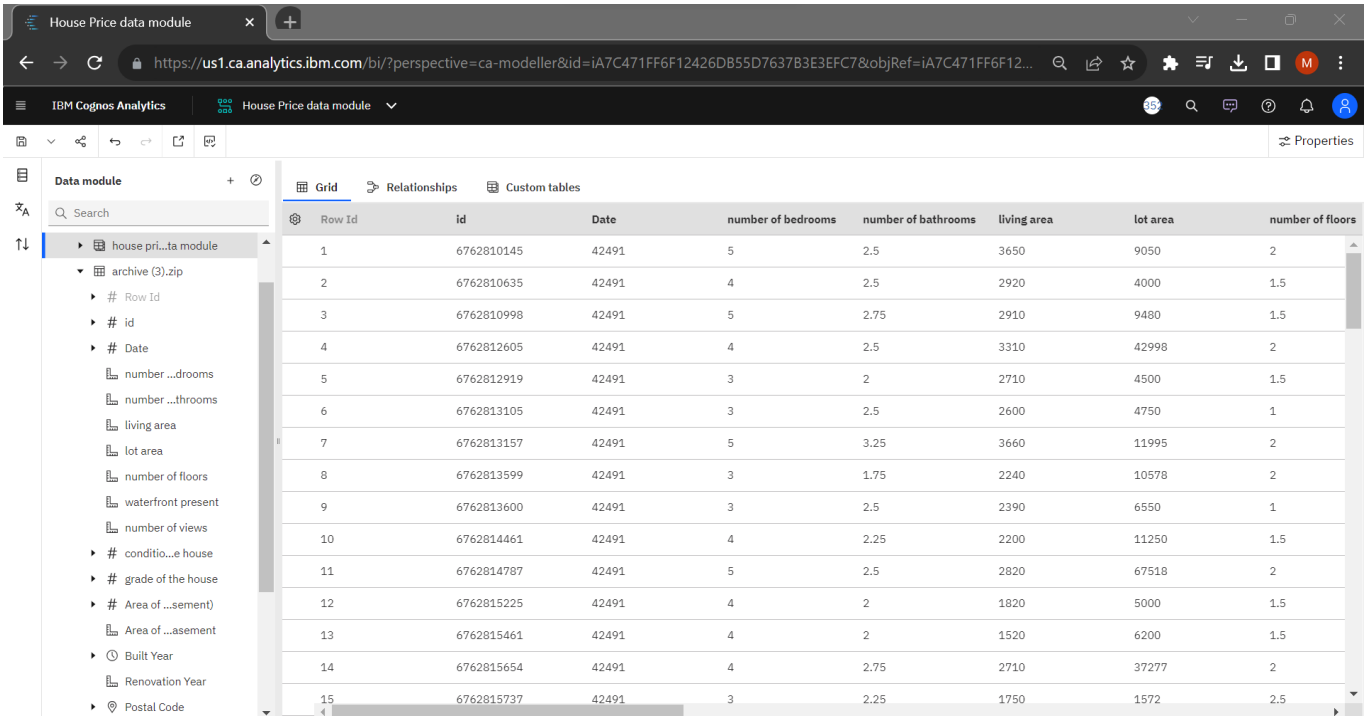
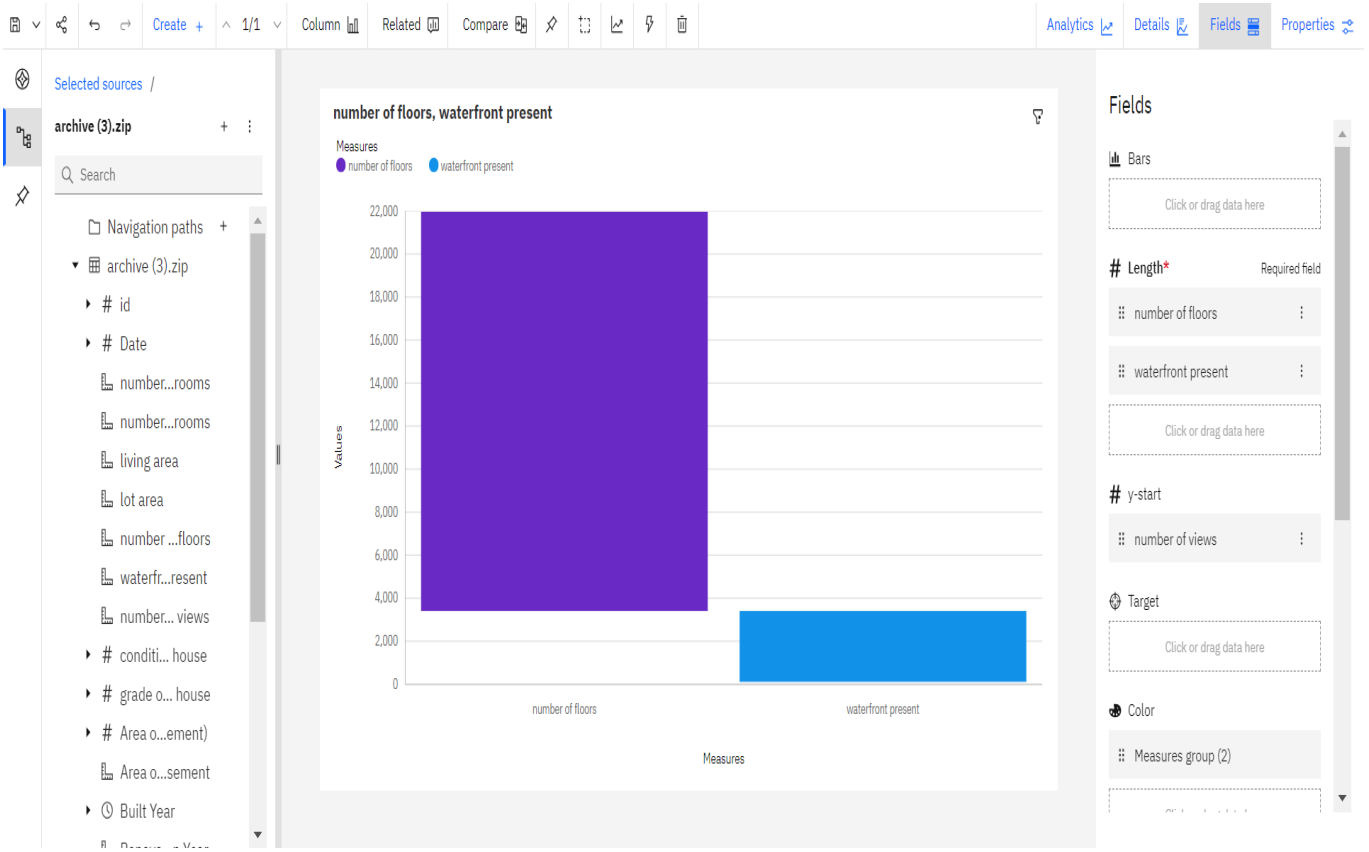
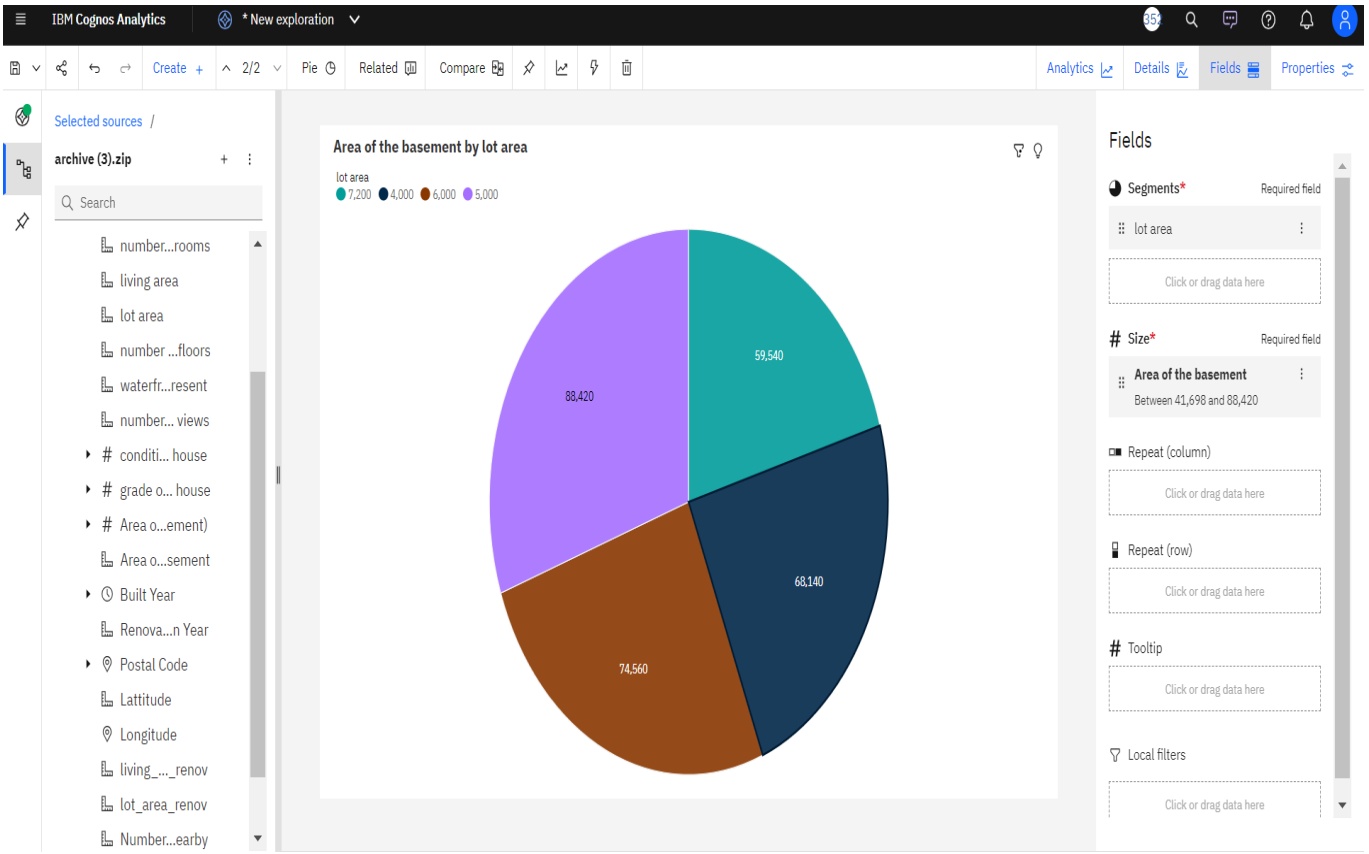
ASSIGNMENT 3

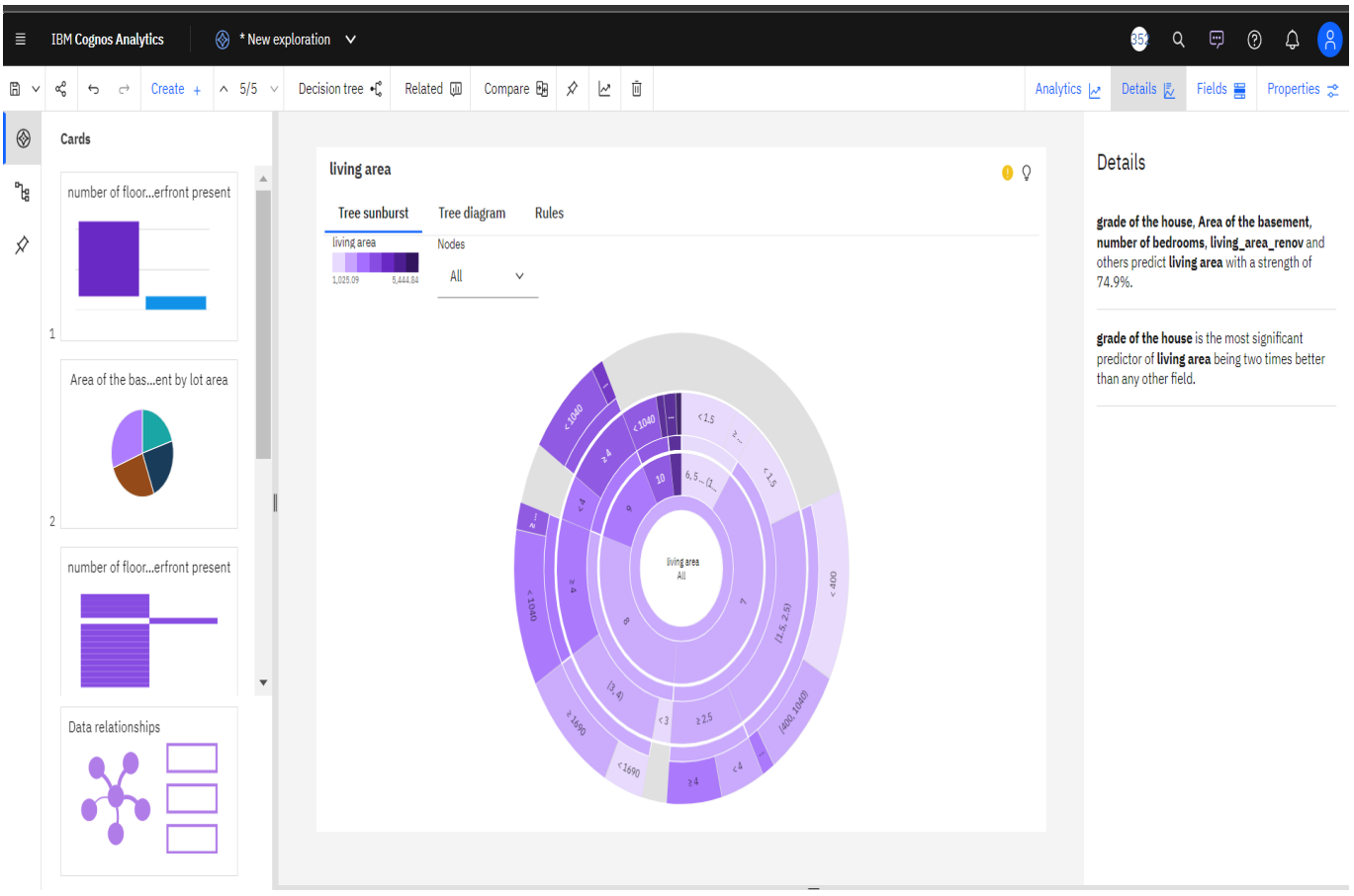
Univariate, Bivariate and Multivariate

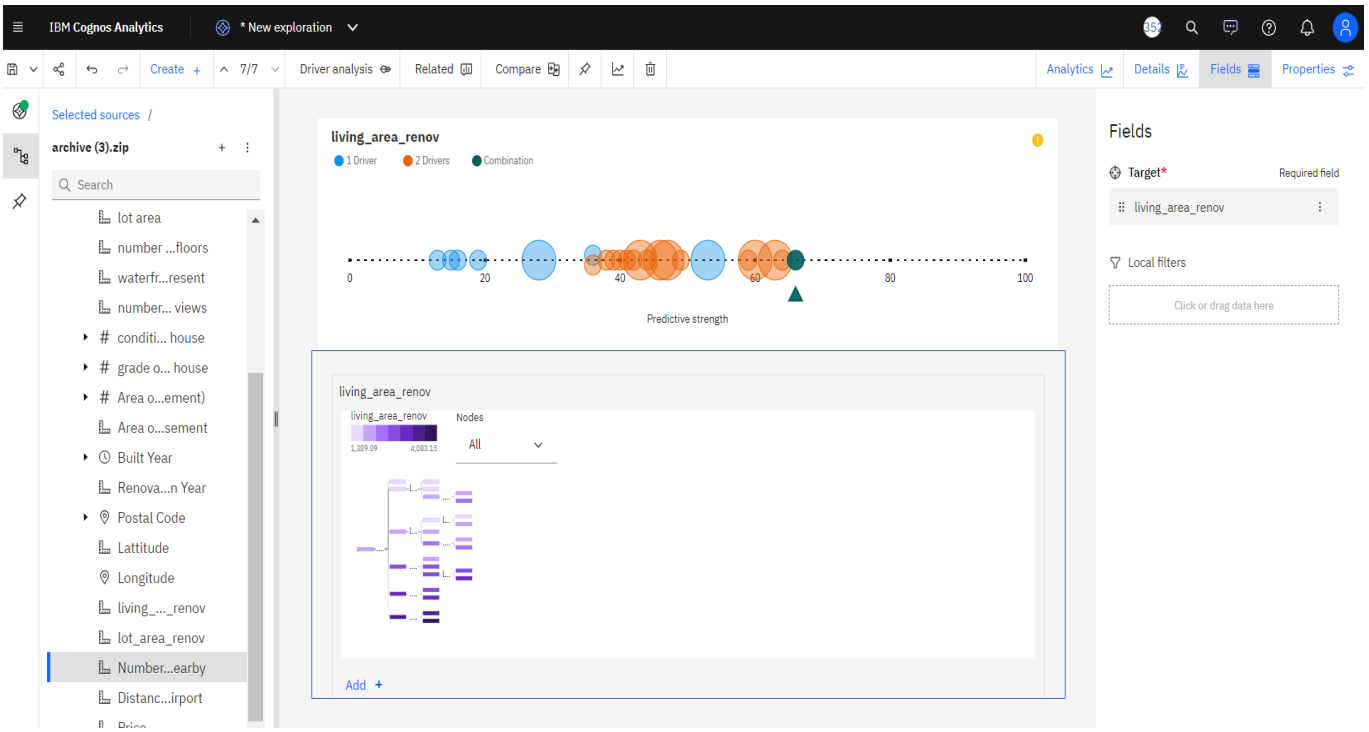


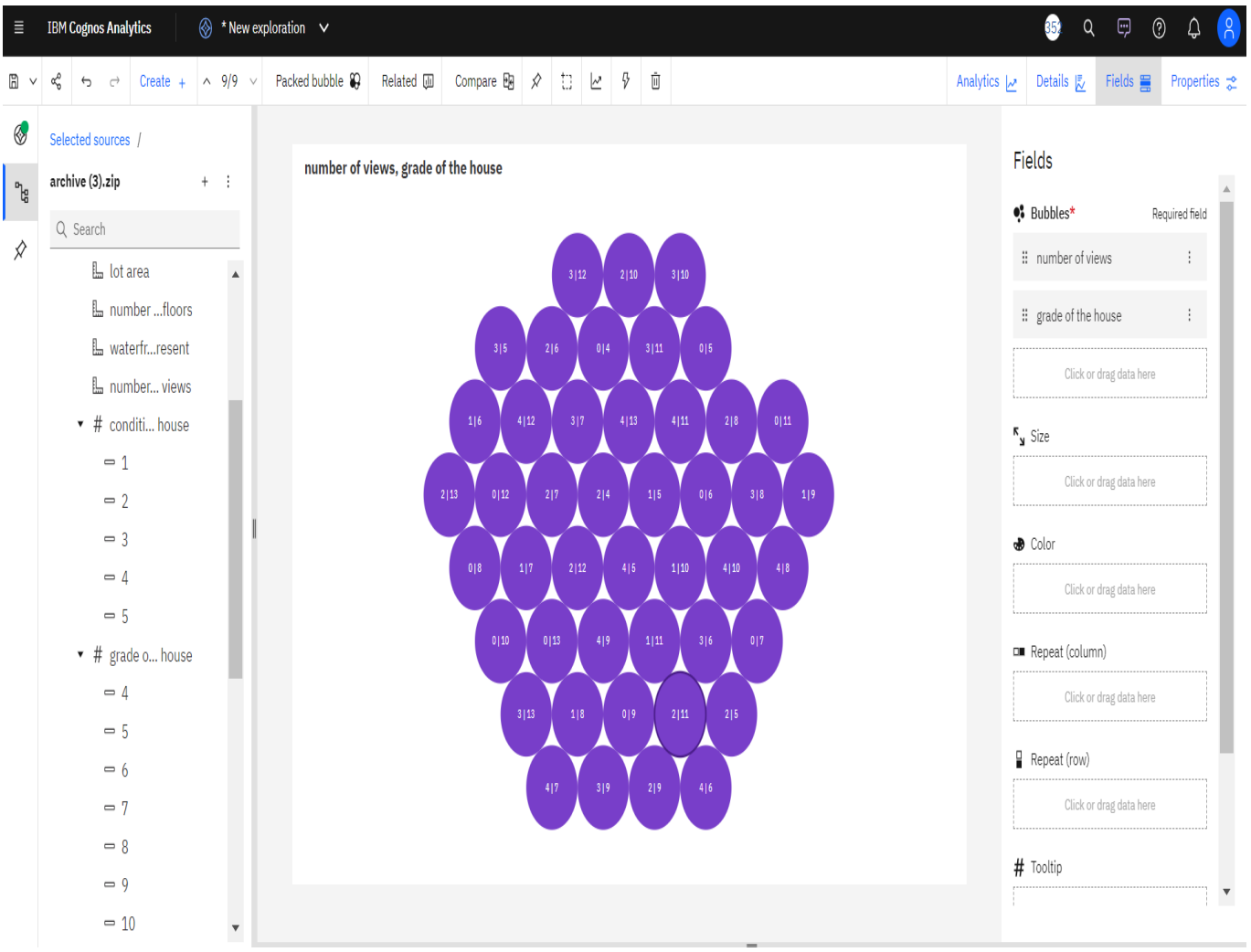
Row Id	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors
1	6762810145	42491	5	2.5	3650	9050	2
2	6762810635	42491	4	2.5	2920	4000	1.5
3	6762810998	42491	5	2.75	2910	9480	1.5
4	6762812605	42491	4	2.5	3310	42998	2
5	6762812919	42491	3	2	2710	4500	1.5
6	6762813105	42491	3	2.5	2600	4750	1
7	6762813157	42491	5	3.25	3660	11995	2
8	6762813599	42491	3	1.75	2240	10578	2
9	6762813600	42491	3	2.5	2390	6550	1
10	6762814461	42491	4	2.25	2200	11250	1.5
11	6762814787	42491	5	2.5	2820	67518	2
12	6762815225	42491	4	2	1820	5000	1.5
13	6762815461	42491	4	2	1520	6200	1.5
14	6762815654	42491	4	2.75	2710	37277	2
15	6762815737	42491	3	2.25	1750	1572	2.5




IBM Cognos Analytics
* New exploration
Create
2/2
Pie
Related
Compare
Analytics
Details
Fields
Properties
Selected sources /
archive (3).zip
Search
number...rooms
living area
lot area
number ...floors
waterfr...resent
number... views
conditi... house
grade o... house
Area o...ement)
Area o...sement
Built Year
Renova...n Year
Postal Code
Lattitude
Longitude
living_..._renov
lot_area_renov
Number...earby
Area of the basement by lot area
lot area
7,200
4,000
6,000
5,000
59,540
68,140
74,560
88,420
Fields
Segments*
Required field
lot area
Click or drag data here
Size*
Required field
Area of the basement
Between 41,698 and 88,420
Repeat (column)
Click or drag data here
Repeat (row)
Click or drag data here
Tooltip
Click or drag data here
Local filters
Click or drag data here


IBM Cognos Analytics
* New exploration
Create
5/5
Decision tree
Related
Compare
Analytics
Details
Fields
Properties
Cards
number of floor...erfront present
1
Area of the bas...ent by lot area
2
number of floor...erfront present
Data relationships
living area
Tree sunburst
Tree diagram
Rules
living area
1,025.09
5,444.84
Nodes
All
Details
grade of the house, Area of the basement, number of bedrooms, living_area_renov and others predict living area with a strength of 74.9%.
grade of the house is the most significant predictor of living area being two times better than any other field.


IBM Cognos Analytics
* New exploration
Create
7/7
Driver analysis
Related
Compare
Analytics
Details
Fields
Properties
Selected sources /
archive (3).zip
Search
lot area
number ...floors
waterfr...resent
number... views
conditi... house
grade o... house
Area o...ement)
Area o...sement
Built Year
Renova...n Year
Postal Code
Lattitude
Longitude
living_..._renov
lot_area_renov
Number...earby
Distanc...irport
living_area_renov
1 Driver
2 Drivers
Combination
0
20
40
60
80
100
Predictive strength
living_area_renov
living_area_renov
Nodes
All
Add
Fields
Target*
Required field
living_area_renov
Local filters
Click or drag data here


IBM Cognos Analytics
* New exploration
Create
9/9
Packed bubble
Related
Compare
Analytics
Details
Fields
Properties
Selected sources /
archive (3).zip
Search
lot area
number ...floors
waterfr...resent
number... views
conditi... house
1
2
3
4
5
grade o... house
4
5
6
7
8
9
10
number of views, grade of the house
Fields
Bubbles*
Required field
number of views
grade of the house
Click or drag data here
Size
Color
Repeat (column)
Repeat (row)
Tooltip

Code:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df =
pd.read_csv("C:\\Users\\Downloads\\HousePriceIndia.csv
")
plt.figure(figsize=(8, 6))
sns.histplot(df['living area'], kde=True)
plt.title('Univariate Analysis')
plt.show()
plt.figure(figsize=(8, 6))
sns.scatterplot(x='number of floors', y='condition of the
house', data=df)
plt.title('Bi-Variate Analysis')
plt.show()
sns.pairplot(df[['Area of the house(excluding basement)',
'grade of the house','Postal Code', 'living_area_renov']])
plt.title('Multi-Variate Analysis - Pair Plot')
plt.show()
descriptive_stats = df.describe()
print(descriptive_stats)
missing_values = df.isnull().sum()
print("Missing Values:")
print(missing_values)
```

Output:

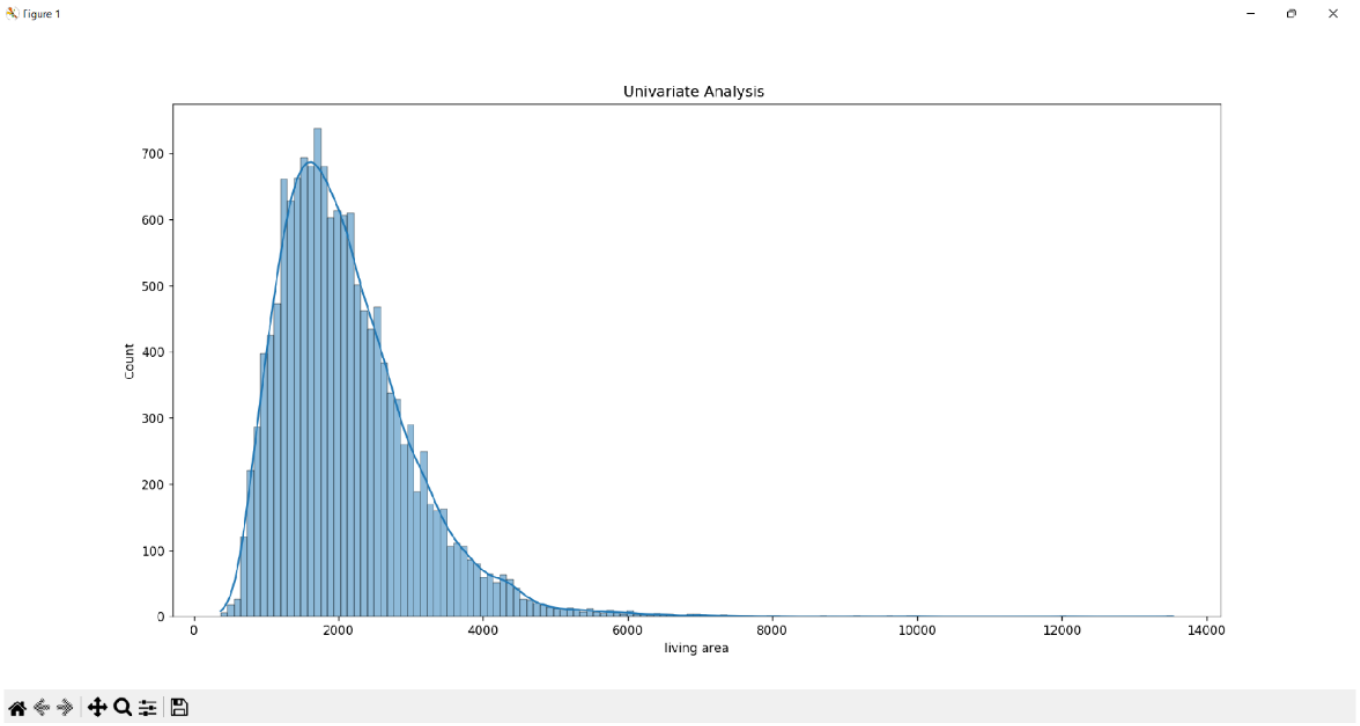


Figure 1
Univariate Analysis
Count
living area
0
100
200
300
400
500
600
700
0
2000
4000
6000
8000
10000
12000
14000

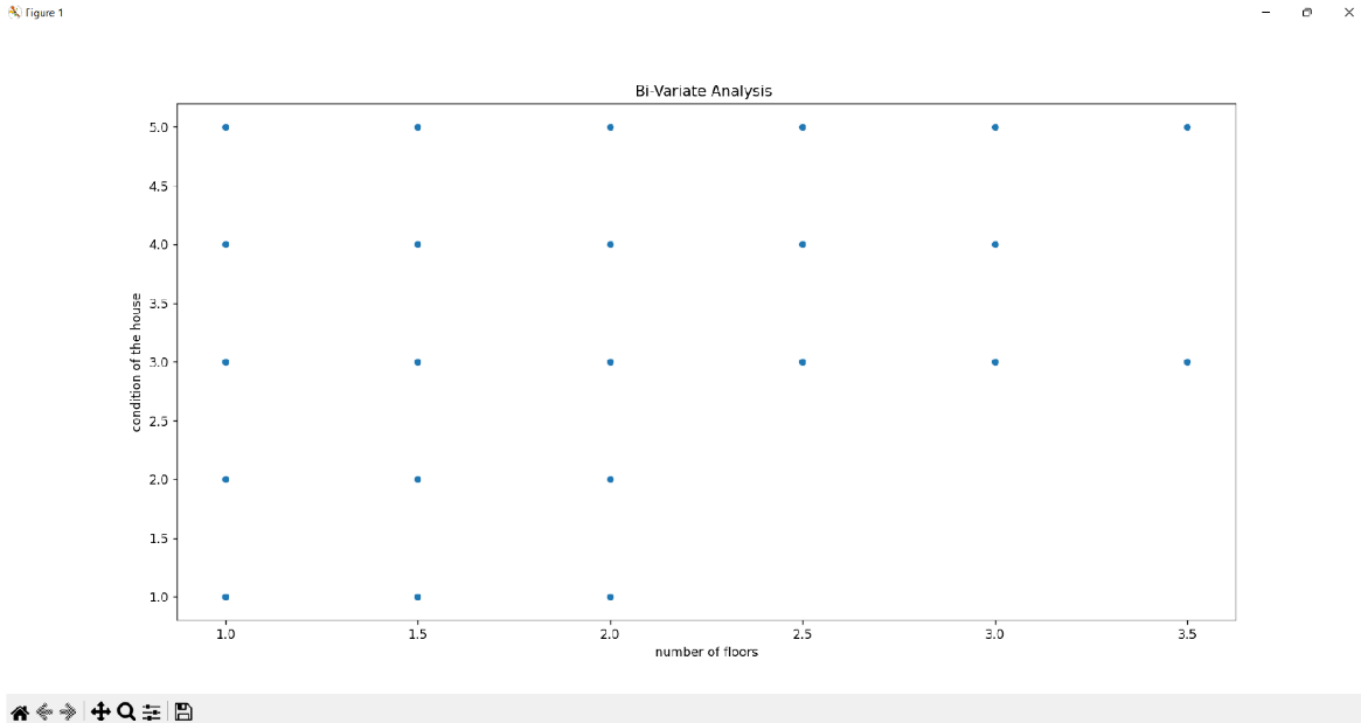


Figure 1
Bi-Variate Analysis
condition of the house
number of floors
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

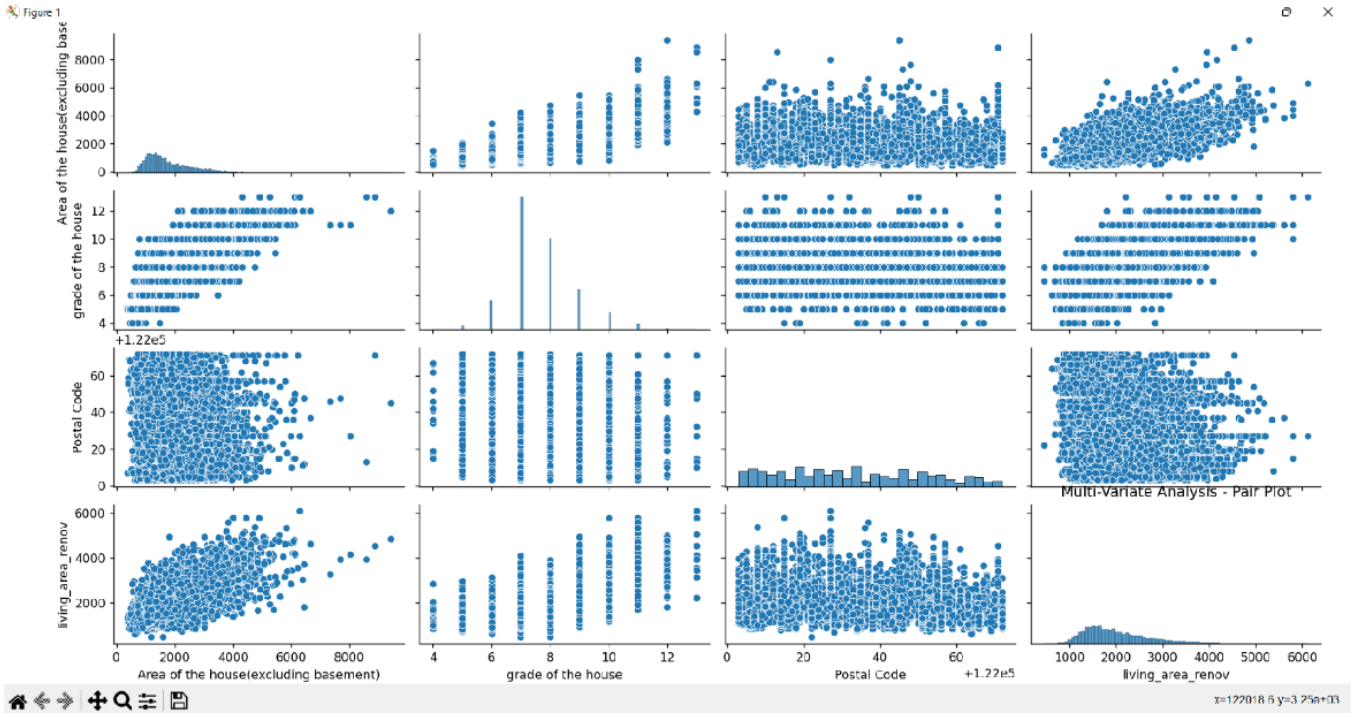
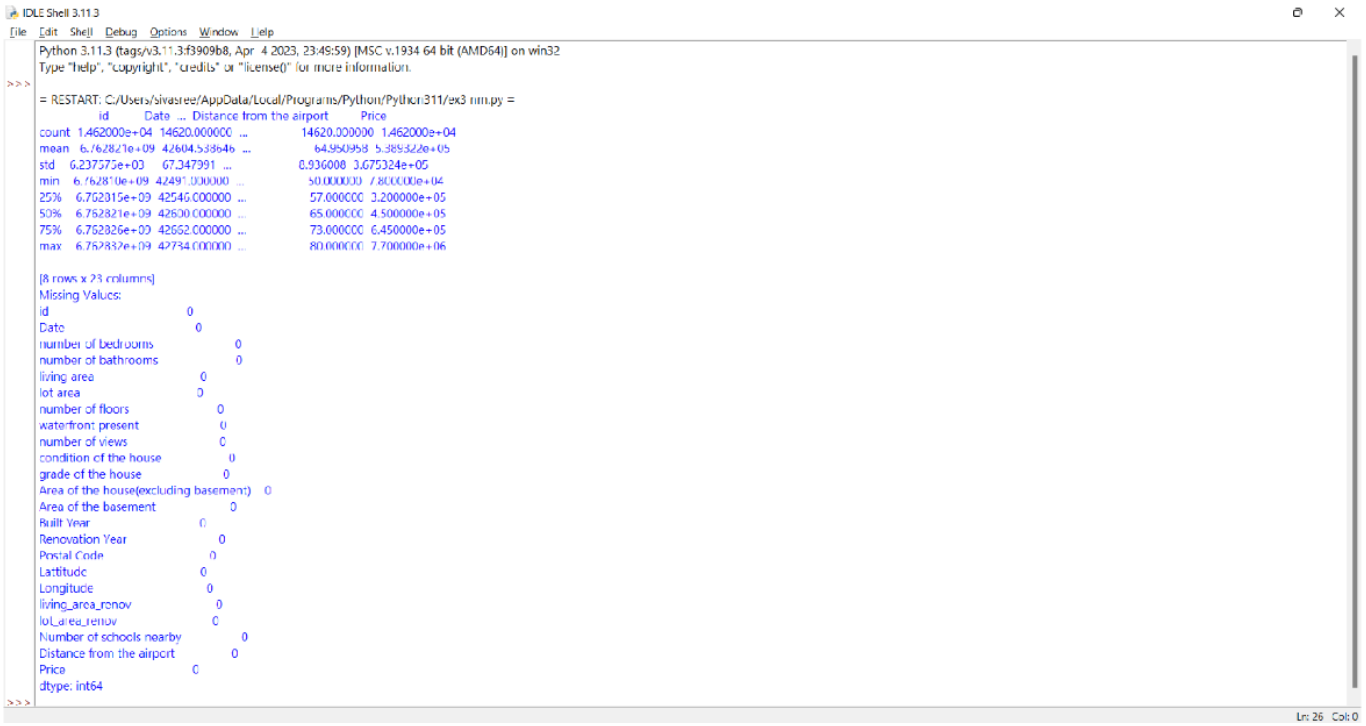


Figure 1
Multi-Variate Analysis - Pair Plot
Area of the house(excluding basement)
grade of the house
Postal Code
living_area_renov
+1.22e5
x=122018.5 y=3.25e+03


IDLE Shell 3.11.3
File Edit Shell Debug Options Window Help
Python 3.11.3 (tags/v3.11.3:f3909b8, Apr 4 2023, 23:49:59) [MSC v.1934 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: C:/Users/sivasree/AppData/Local/Programs/Python/Python311/ex3 nm.py =
id Date ... Distance from the airport Price
count 1.462000e+04 14620.000000 ... 14620.000000 1.462000e+04
mean 6.762821e+09 42604.538646 ... 64.950958 5.389322e+05
std 6.237575e+03 67.347991 ... 0.936008 3.675324e+05
min 6.762810e+09 42491.000000 ... 50.000000 7.800000e+04
25% 6.762815e+09 42546.000000 ... 57.000000 3.200000e+05
50% 6.762821e+09 42630.000000 ... 65.000000 4.500000e+05
75% 6.762826e+09 42662.000000 ... 73.000000 6.450000e+05
max 6.762832e+09 42734.000000 ... 80.000000 7.700000e+06
[8 rows x 23 columns]
Missing Values:
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
>>>
Ln: 26 Col: 0