```
In [1]:  import pandas as pd
         from pandasai import SmartDataframe
         from pandasai.llm import OpenAI
```

```
In [2]:  import numpy as np
         import matplotlib.pyplot as plt
         from sklearn.tree import DecisionTreeRegressor, export_text
         from sklearn.model_selection import train_test_split
         from sklearn.metrics import mean_squared_error
```

```
In [3]:  from sklearn.metrics import r2_score
         from sklearn.model_selection import GridSearchCV
```

```
In [59]:  !jupyter nbconvert --to webpdf openAI - Project_Final.ipynb
```

```
[NbConvertApp] WARNING | pattern '-' matched no files
[NbConvertApp] Converting notebook openAI.ipynb to webpdf
C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python310\lib\site-packages\nbco
nvert\filters\highlight.py:71: UserWarning: IPython3 lexer unavailable, falling back
on Python 3
  return _pygments_highlight(
[NbConvertApp] WARNING | Alternative text is missing on 1 image(s).
[NbConvertApp] Building PDF
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 169213 bytes to openAI.pdf
[NbConvertApp] Converting notebook Project_Final.ipynb to webpdf
[NbConvertApp] WARNING | Alternative text is missing on 15 image(s).
[NbConvertApp] Building PDF
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 527785 bytes to Project_Final.pdf
```

1 - Load and preview dataset

```
In [5]:  house_price = pd.read_csv('Real_estate _1.csv')
         house_price.head()
```

Out[5]:

	No	X1 transaction date	X2 house age	X3 distance to the nearest MRT station	X4 number of convenience stores	X5 latitude	X6 longitude	Y house price of unit area
0	1	2012.917	32.0	84.87882	10	24.98298	121.54024	37.9
1	2	2012.917	19.5	306.59470	9	24.98034	121.53951	42.2
2	3	2013.583	13.3	561.98450	5	24.98746	121.54391	47.3
3	4	2013.500	13.3	561.98450	5	24.98746	121.54391	54.8
4	5	2012.833	5.0	390.56840	5	24.97937	121.54245	43.1

```
In [6]:  llm = OpenAI(api_token="sk-proj-KtQvt_jEawaTmmBFpqxRE5fXxd7dTJfPGgsmwGYJpNW19XAv3cH
         house_price = SmartDataframe(house_price, config ={"llm": llm})
```

```
In [7]:  # Checking the shape of dataset.
         house_price.chat('how many rows and columns are in the dataset?')
```

{'type': 'string', 'value': 'The dataset contains 414 rows and 8 columns.'}

Out[7]: 'The dataset contains 414 rows and 8 columns.'

Data cleaning and Preparation

```
In [8]:  # Fixing inconsistent formatting
         house_price.chat('in column names replace whitespaces to "_" and change uppercase t
```

Out[8]:

	no	x1_transaction_date	x2_house_age	x3_distance_to_the_nearest_mrt_station	x4_nu
0	1	2012.917	32.0	84.87882	
1	2	2012.917	19.5	306.59470	
2	3	2013.583	13.3	561.98450	
3	4	2013.500	13.3	561.98450	
4	5	2012.833	5.0	390.56840	
...	...	...	...	...	
409	410	2013.000	13.7	4082.01500	
410	411	2012.667	5.6	90.45606	
411	412	2013.250	18.8	390.96960	
412	413	2013.000	8.1	104.81010	
413	414	2013.500	6.5	90.45606	

414 rows × 8 columns

```
In [9]:  # Rename 'no' column to 'transaction_id'
         house_price.chat('Rename "no" to "Transaction_ID')
```

	transaction_id	x1_transaction_date	x2_house_age	x3_distance_to_the_nearest_mrt_stati
0	1	2012.917	32.0	84.878
1	2	2012.917	19.5	306.594
2	3	2013.583	13.3	561.984
3	4	2013.500	13.3	561.984
4	5	2012.833	5.0	390.568
...	...	...	...	
409	410	2013.000	13.7	4082.015
410	411	2012.667	5.6	90.456
411	412	2013.250	18.8	390.969
412	413	2013.000	8.1	104.810
413	414	2013.500	6.5	90.456

414 rows × 8 columns

In [10]:
```python
# Checking for missing values in each column
house_price.chat('is there a missing data in the dataset?')
```

Out[10]: 'There is no missing data in the dataset.'

In [12]:
```python
# Checking for duplicate rows
house_price.chat('are the dublicates values in the dataset?')
```

Out[12]: 'There are no duplicate rows in the dataset.'

In [15]:
```python
# Checking for data types
house_price.chat('what are the data types of values?')
```

```
{'type': 'dataframe', 'value':                                  Column Data Type
0                        transaction_id     int64
1                   x1_transaction_date   float64
2                          x2_house_age   float64
3   x3_distance_to_the_nearest_mrt_station   float64
4          x4_number_of_convenience_stores     int64
5                           x5_latitude   float64
6                          x6_longitude   float64
7            y_house_price_of_unit_area   float64}
```

```
Traceback (most recent call last):
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\generate_chat_pipeline.py", line 335, in run
    ).run(input)
      ^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 137, in run
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 101, in run
    step_output = logic.execute(
                  ^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 133, in execute
    {"content_type": "response", "value": ResponseSerializer.serialize(result)},
                                          ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\responses\response_serializer.py", line 27, in serialize
    df_dict = ResponseSerializer.serialize_dataframe(result["value"])
              ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\responses\response_serializer.py", line 11, in serialize_dataframe
    json_data = json.loads(df.to_json(orient="split", date_format="iso"))
                           ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\util\_decorators.py", line 333, in wrapper
    return func(*args, **kwargs)
           ^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\generic.py", line 2721, in to_json
    return json.to_json(
           ^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\io\json\_json.py", line 210, in to_json
    ).write()
      ^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\io\json\_json.py", line 263, in write
    return ujson_dumps(
           ^^^^^^^^^^^^
OverflowError: Maximum recursion level reached
```

Out[15]: 'Unfortunately, I was not able to answer your question, because of the following e
          rror:\n\nMaximum recursion level reached\n'

2 - Data distribution and Outliers

In [16]:
```python
# Checking Data Distribution
house_price.chat('Show data distribution of continius data. show it in tables')
```
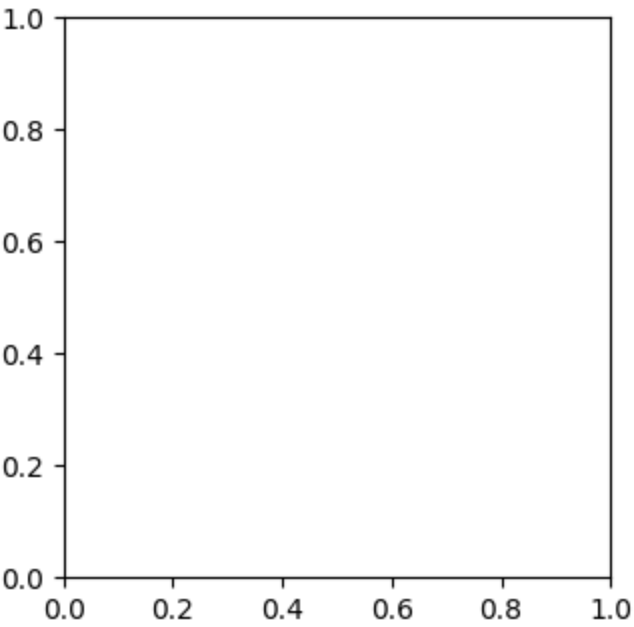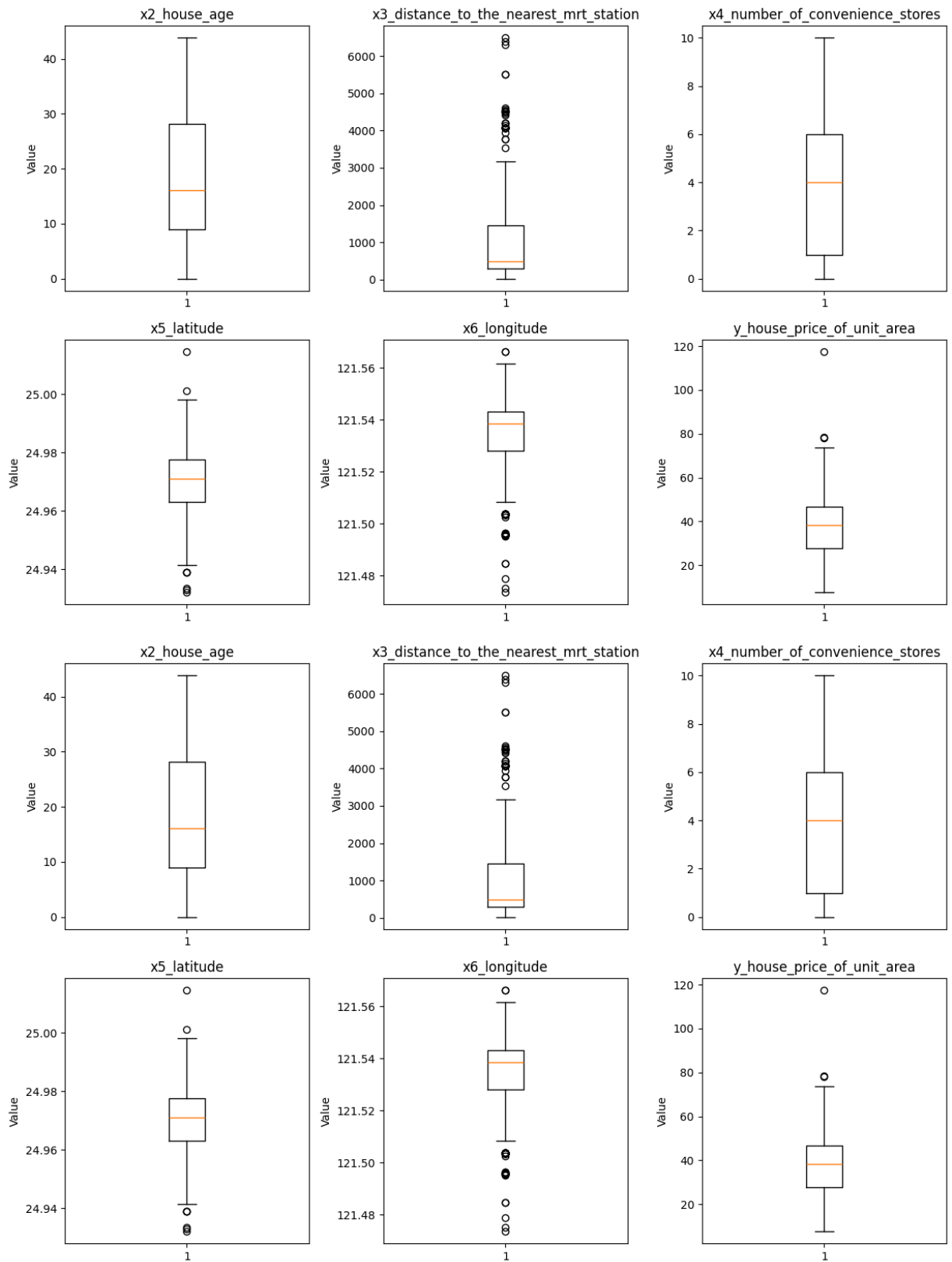
	transaction_id	x1_transaction_date	x2_house_age	x3_distance_to_the_nearest_mrt_sta
count	414.000000	414.000000	414.000000	414.000
mean	207.500000	2013.148971	17.712560	1083.88
std	119.655756	0.281967	11.392485	1262.10
min	1.000000	2012.667000	0.000000	23.38
25%	104.250000	2012.917000	9.025000	289.32
50%	207.500000	2013.167000	16.100000	492.23
75%	310.750000	2013.417000	28.150000	1454.27
max	414.000000	2013.583000	43.800000	6488.02

In [17]:
```python
# Graphical representation of outliers to confirm outlier detection for "x3_distanc
#And checking for potential outliers in other variables
house_price.chat('display all outliers of dataset in boxplots')
```

```
Traceback (most recent call last):
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\generate_chat_pipeline.py", line 335, in run
    ).run(input)
      ^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 137, in run
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 101, in run
    step_output = logic.execute(
                  ^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 113, in execute
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 85, in execute
    result = self.execute_code(code_to_run, code_context)
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 175, in execute_code
    raise NoResultFoundError("No result returned")
pandasai.exceptions.NoResultFoundError: No result returned
```

Out[17]: 'Unfortunately, I was not able to answer your question, because of the following e
          rror:\n\nNo result returned\n'

In [18]: #Count the outliers in each of the columns
          house_price.chat('count outliers in each column')

```
                                    Outlier Count
    transaction_id                              0
    x1_transaction_date                         0
    x2_house_age                                0
    x3_distance_to_the_nearest_mrt_station     37
    x4_number_of_convenience_stores             0
    x5_latitude                                 8
    x6_longitude                               35
    y_house_price_of_unit_area                  3
```
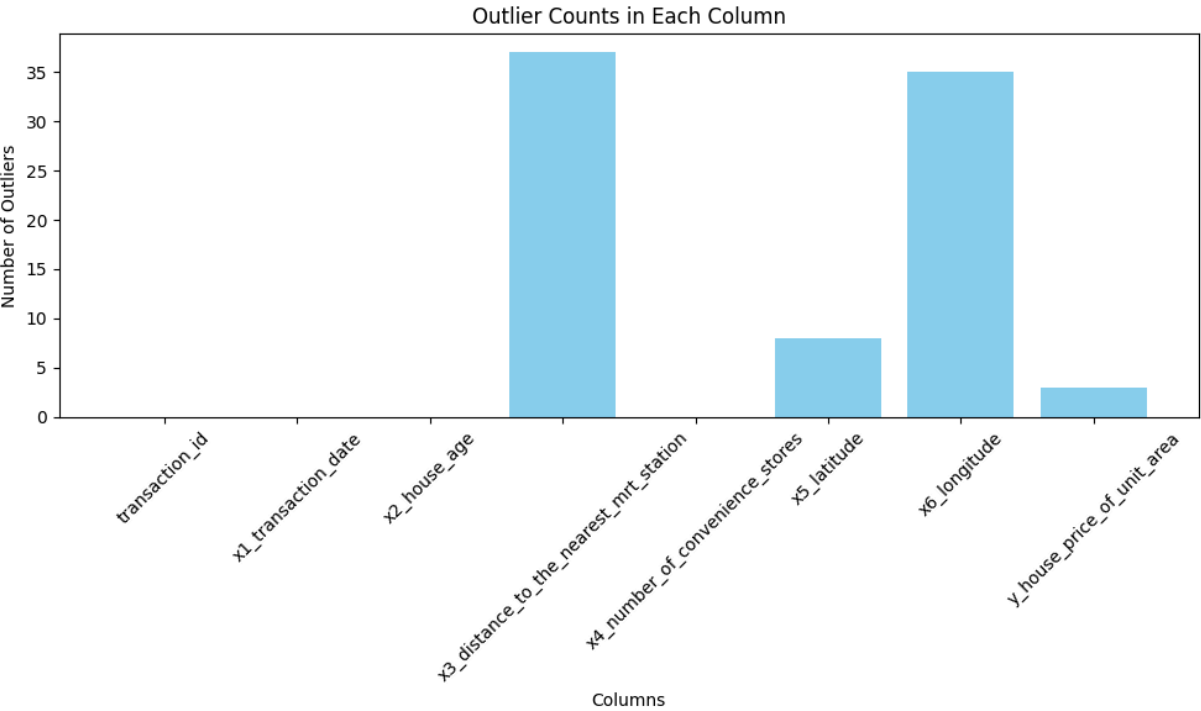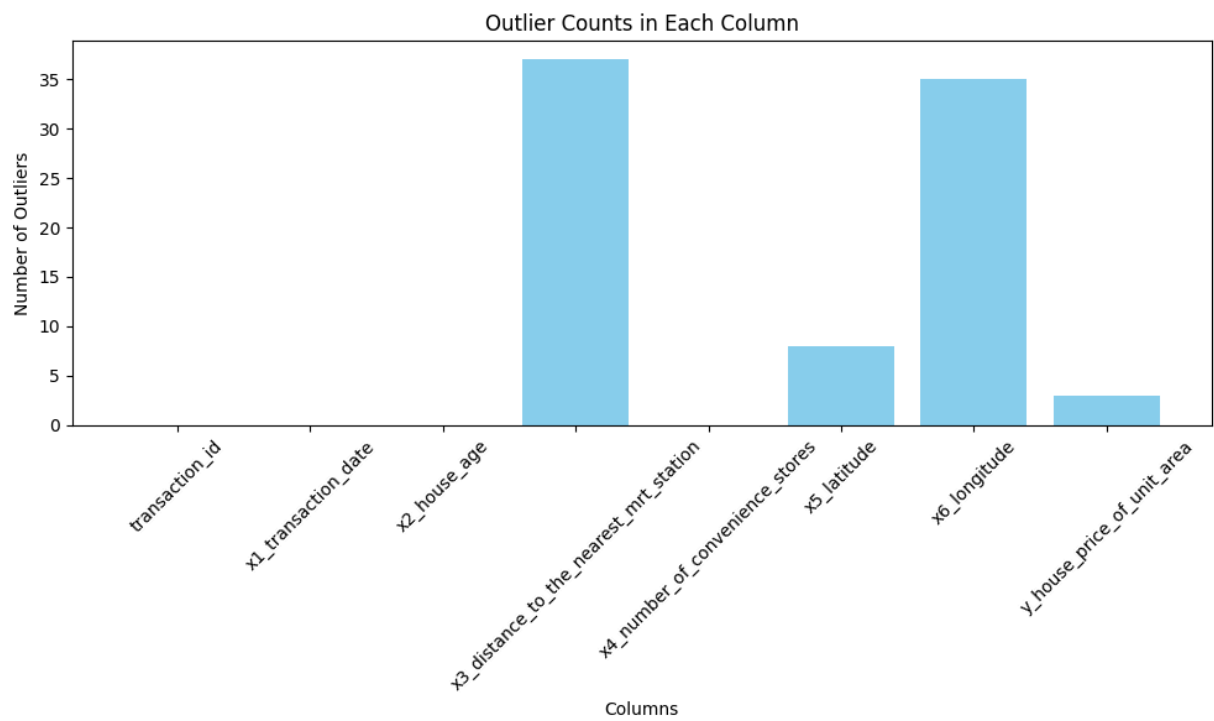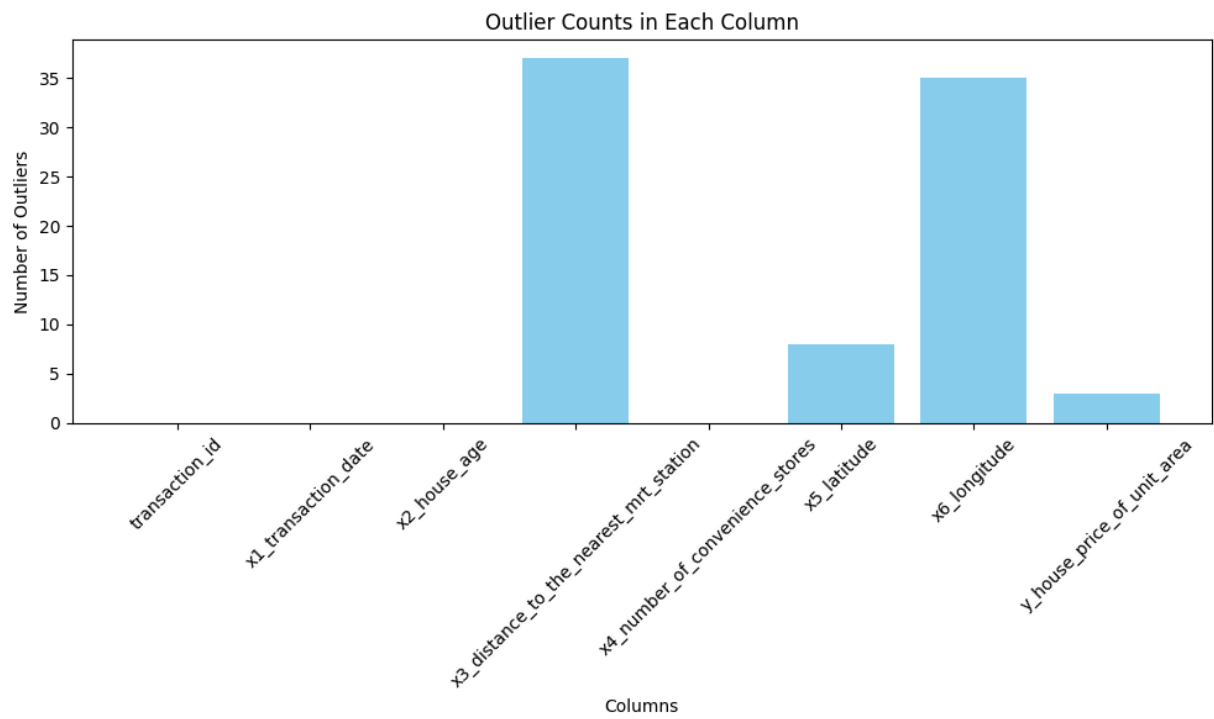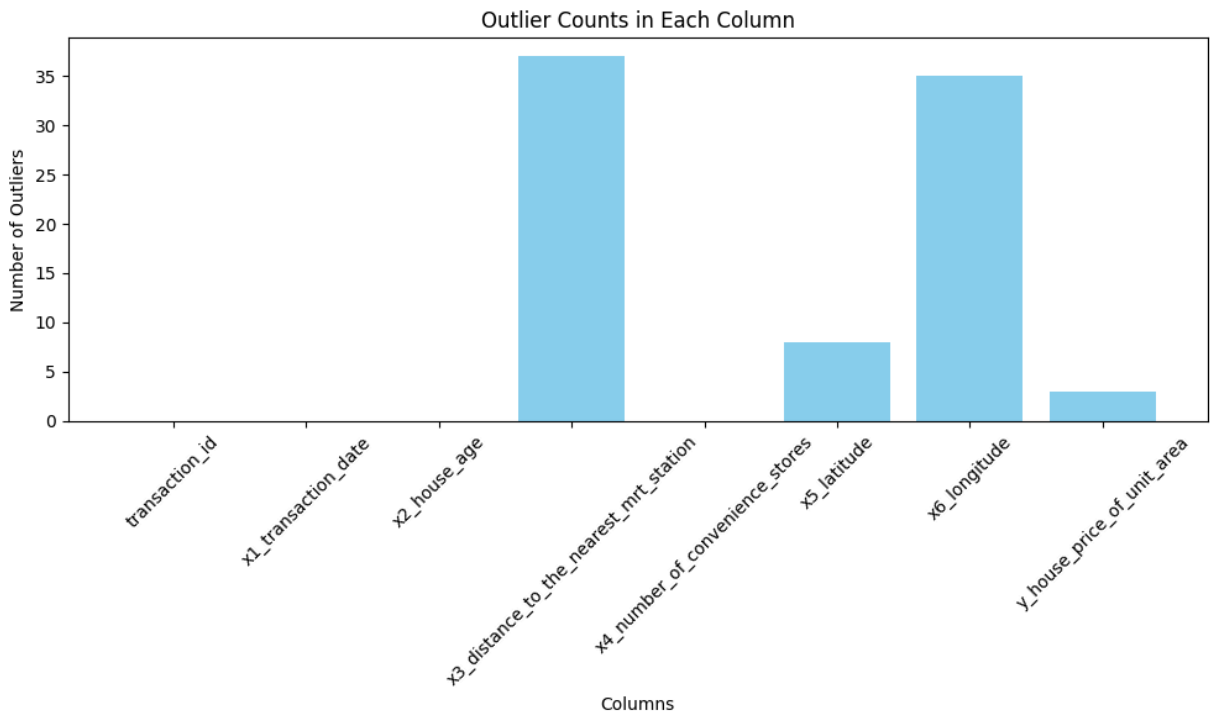
Out[18]:

	Outlier Count
transaction_id	0
x1_transaction_date	0
x2_house_age	0
x3_distance_to_the_nearest_mrt_station	37
x4_number_of_convenience_stores	0
x5_latitude	8
x6_longitude	35
y_house_price_of_unit_area	3

In [19]: house_price.chat('count outliers in each column and show it in column bars')

Outlier Counts in Each Column

Outlier Counts in Each Column

Outlier Counts in Each Column


```
Traceback (most recent call last):
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\generate_chat_pipeline.py", line 335, in run
    ).run(input)
      ^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 137, in run
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 101, in run
    step_output = logic.execute(
                  ^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 113, in execute
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 97, in execute
    raise InvalidOutputValueMismatch(
pandasai.exceptions.InvalidOutputValueMismatch: Value type <class 'str'> must match
with type plot
```

Out[19]: "Unfortunately, I was not able to answer your question, because of the following e
          rror:\n\nValue type <class 'str'> must match with type plot\n"

In [30]: house_price.chat('show 97th percentile of x3_distance_to_the_nearest_mrt_station co

          {'type': 'number', 'value': np.float64(4435.033050000001)}

Out[30]: np.float64(4435.033050000001)

In [20]: # Replacing "x3_distance_to_the_nearest_mrt_station" outliers with 97th percentile.
          house_price.chat('replace outliers in "x3_distance_to_the_nearest_mrt_station" colu

Out[20]:

	transaction_id	x1_transaction_date	x2_house_age	x3_distance_to_the_nearest_mrt_stati
0	1	2012.917	32.0	84.878
1	2	2012.917	19.5	306.594
2	3	2013.583	13.3	561.984
3	4	2013.500	13.3	561.984
4	5	2012.833	5.0	390.568
...	...	...	...	
409	410	2013.000	13.7	4435.033
410	411	2012.667	5.6	90.456
411	412	2013.250	18.8	390.969
412	413	2013.000	8.1	104.810
413	414	2013.500	6.5	90.456

414 rows × 8 columns

In [31]: `house_price.chat('replace outliers in "x3_distance_to_the_nearest_mrt_station" colu`

Out[31]:

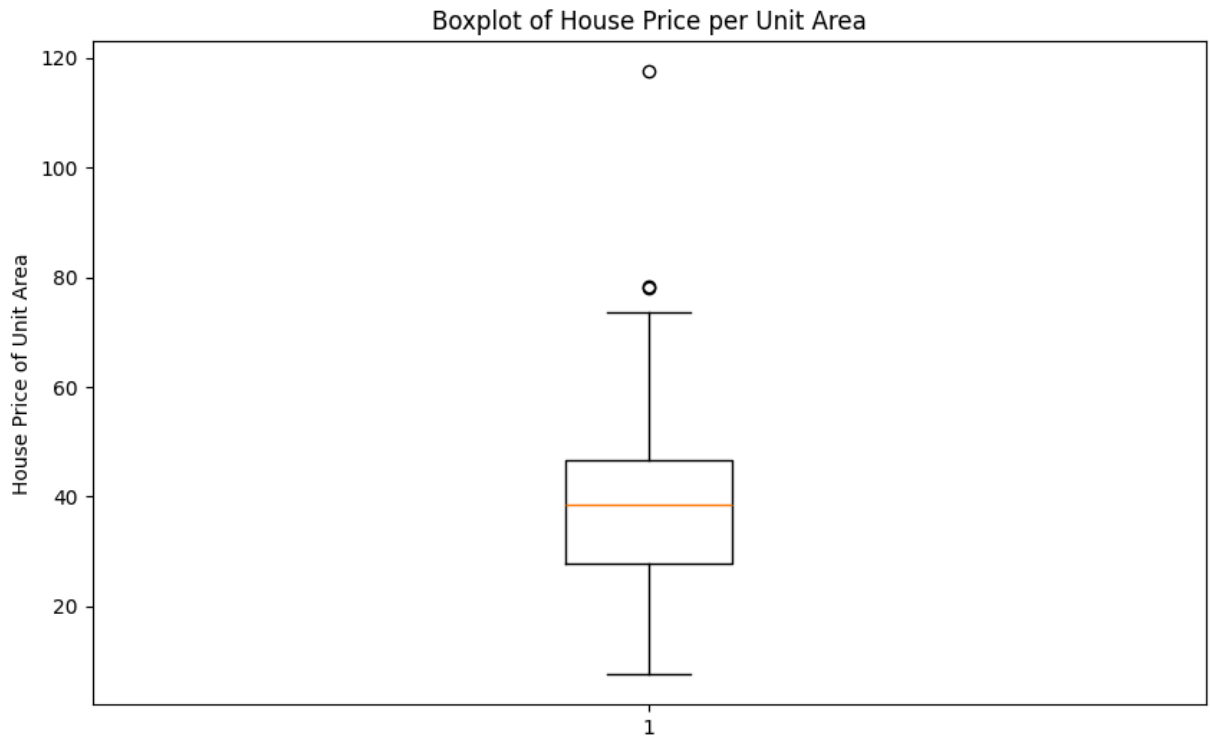|	transaction_id	x1_transaction_date	x2_house_age	x3_distance_to_the_nearest_mrt_stati
0	1	2012.917	32.0	84.878
1	2	2012.917	19.5	306.594
2	3	2013.583	13.3	561.984
3	4	2013.500	13.3	561.984
4	5	2012.833	5.0	390.568
...	...	...	...	
409	410	2013.000	13.7	4435.033
410	411	2012.667	5.6	90.456
411	412	2013.250	18.8	390.969
412	413	2013.000	8.1	104.810
413	414	2013.500	6.5	90.456

414 rows × 8 columns

In [38]:
```
#Remove records with outliers in 'y_house_price_of_unit_area' column
house_price.chat('what are the outliers of the y_house_price_of_unit_area column')
```

	transaction_id	x1_transaction_date	x2_house_age	x3_distance_to_the_nearest_mrt_static
220	221	2013.333	37.2	186.51
270	271	2013.333	10.8	252.58
312	313	2013.583	35.4	318.52

Boxplot of House Price per Unit Area

In [46]:
```python
# Removing 3 outliers from the "y_house_price_of_unit_area" column.
# We can see that number of rows reduced to 411.
house_price.chat('remove 3 outliers from the "y_house_price_of_unit_area" ')
```

|	transaction_id	x1_transaction_date	x2_house_age	x3_distance_to_the_nearest_mrt_stati
0	1	2012.917	32.0	84.878
1	2	2012.917	19.5	306.594
2	3	2013.583	13.3	561.984
3	4	2013.500	13.3	561.984
4	5	2012.833	5.0	390.568
...	...	...	...	
409	410	2013.000	13.7	4435.033
410	411	2012.667	5.6	90.456
411	412	2013.250	18.8	390.969
412	413	2013.000	8.1	104.810
413	414	2013.500	6.5	90.456

411 rows × 8 columns

Data Analysis for Data Understanding

```python
# Display descriptive statistics of the variables
house_price.chat('Display descriptive statistics of the variables')
```

	transaction_id	x1_transaction_date	x2_house_age	x3_distance_to_the_nearest_mrt_sta
count	414.000000	414.000000	414.000000	414.00
mean	207.500000	2013.148971	17.712560	1084.77
std	119.655756	0.281967	11.392485	1247.01
min	1.000000	2012.667000	0.000000	23.38
25%	104.250000	2012.917000	9.025000	289.32
50%	207.500000	2013.167000	16.100000	492.23
75%	310.750000	2013.417000	28.150000	1454.27
max	414.000000	2013.583000	43.800000	4435.03

```
<Figure size 1000x600 with 0 Axes>
```
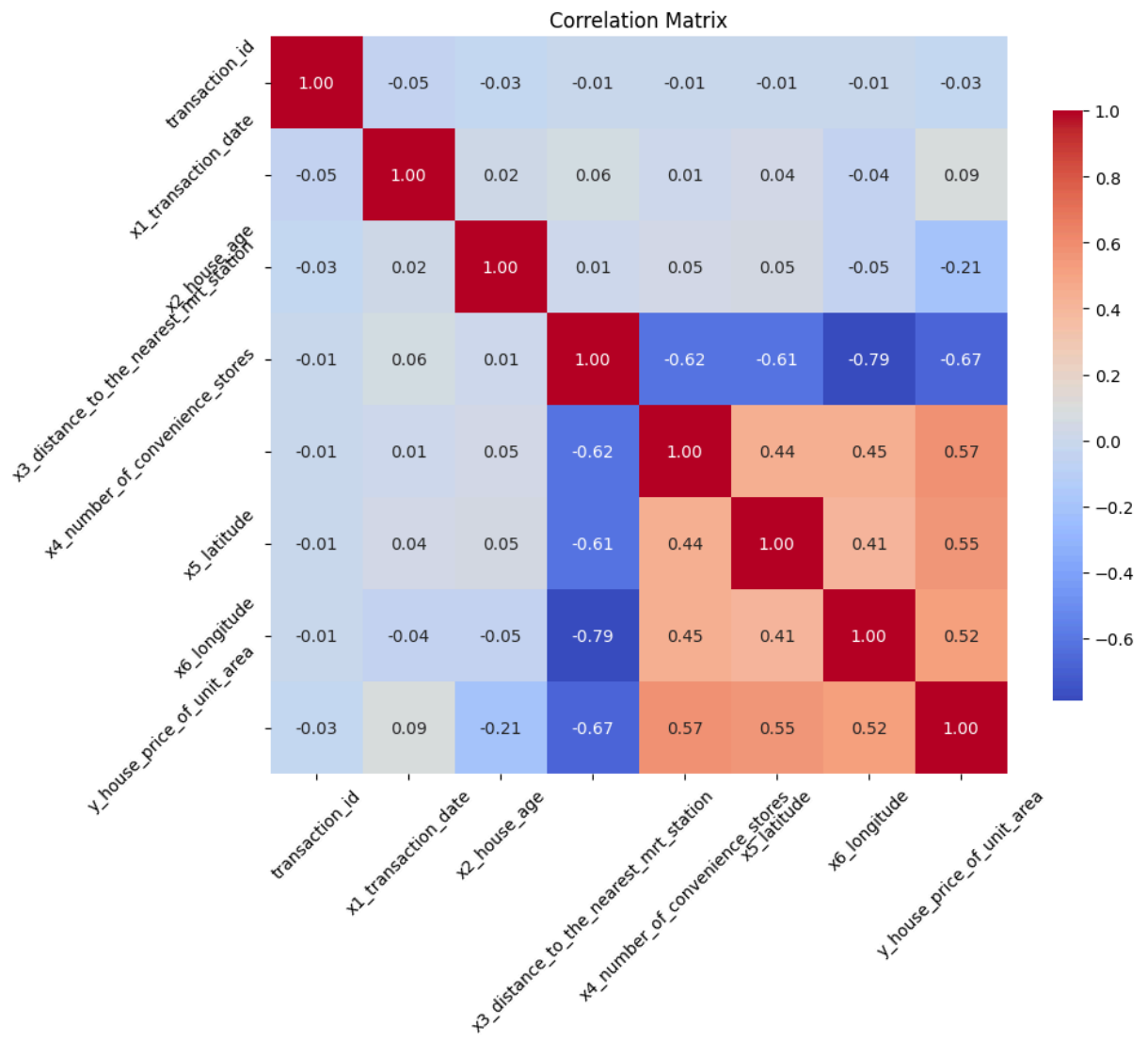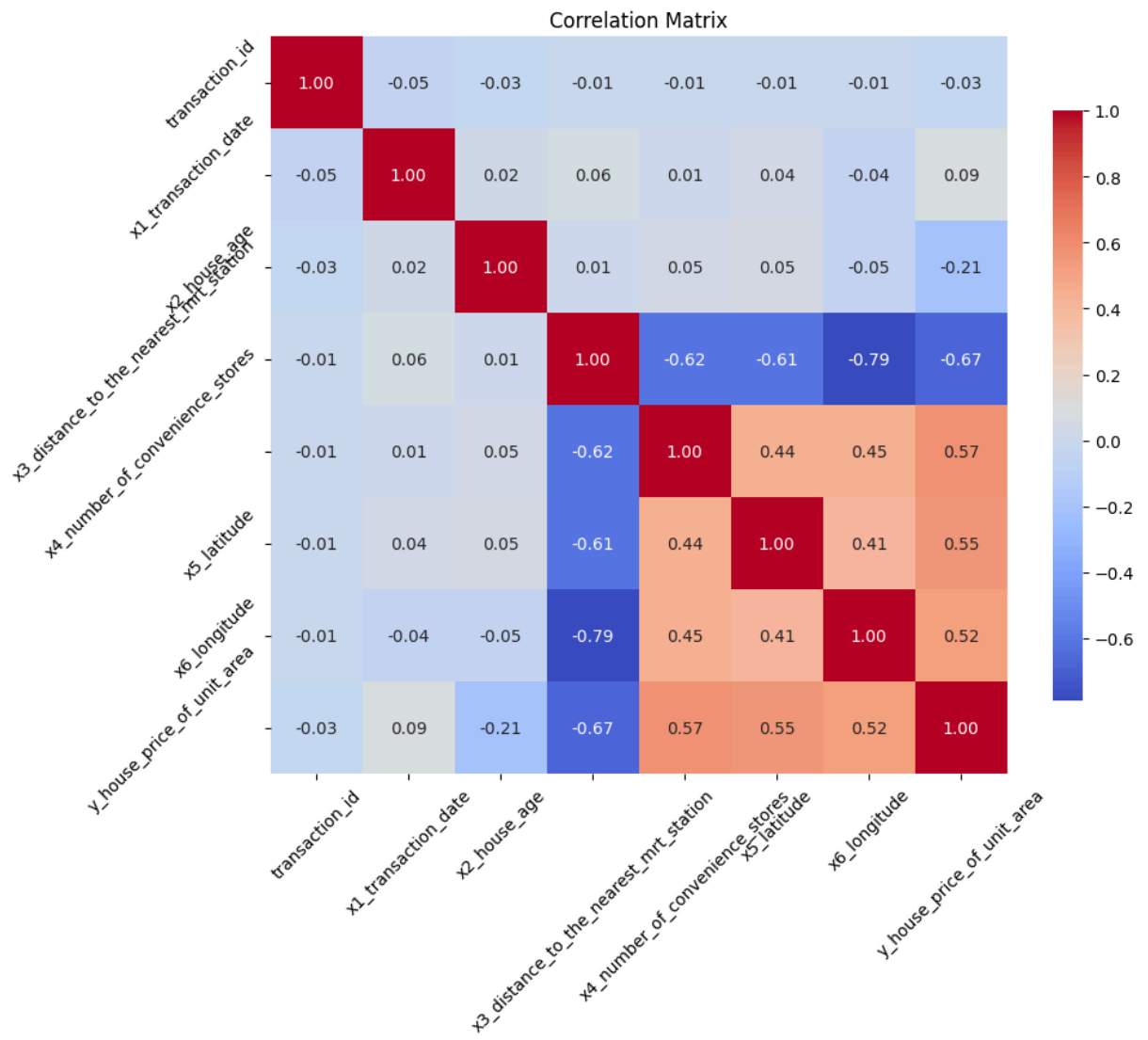
Descriptive Statistics of Variables
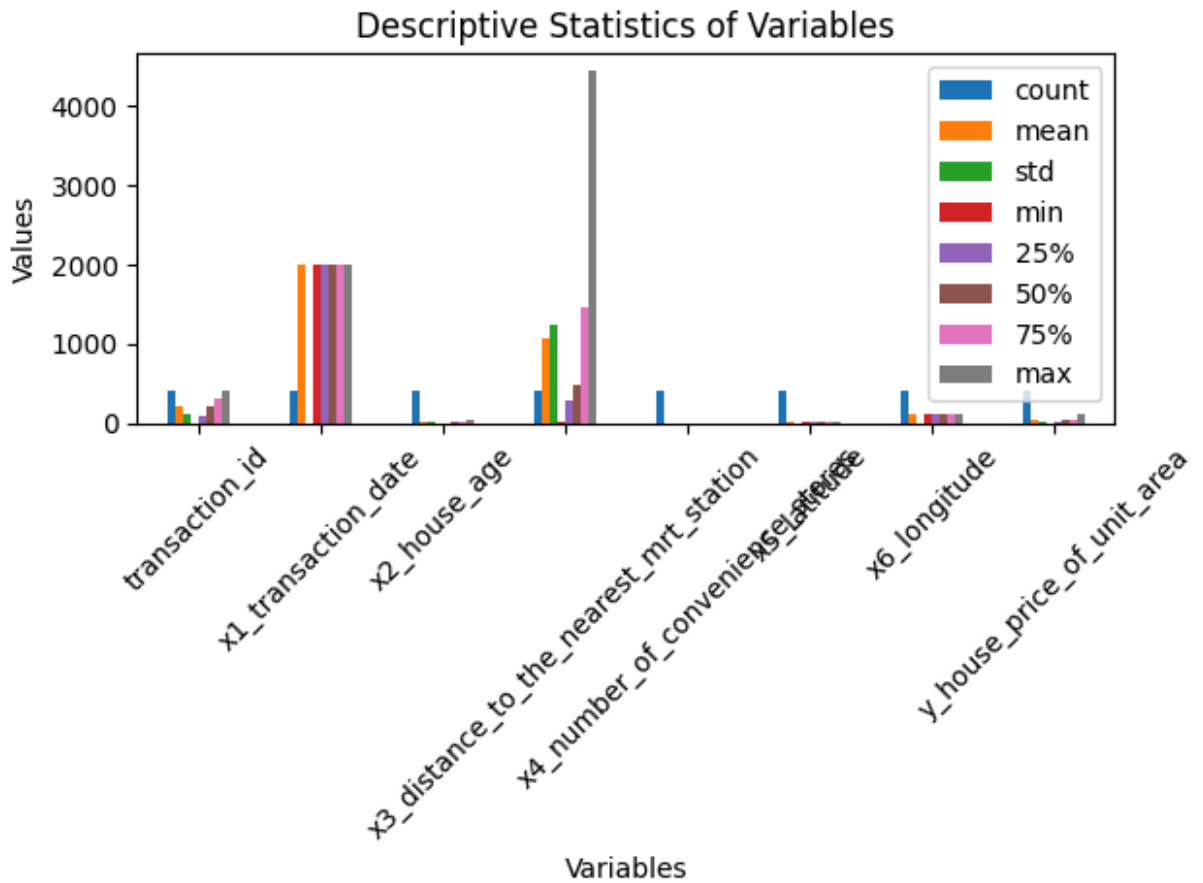
In [53]: `house_price.chat('In boxplot calculate count, mean, std, min, 25% quartile, 50% qua`

Out[53]:

	Count	Mean	Std	Min	25%	50%	75%	Max
0	414.0	37.980193	13.606488	7.6	27.7	38.45	46.6	117.5

In [54]:
```
# Generating and plotting the correlation matrix
house_price.chat('Generate and plot the correlation matrix')
```

Correlation Matrix

	transaction_id	x1_transaction_date	x2_house_age	x3_distance_to_the_nearest_mrt_station	x4_number_of_convenience_stores	x5_latitude	x6_longitude	y_house_price_of_unit_area
transaction_id	1.00	-0.05	-0.03	-0.01	-0.01	-0.01	-0.01	-0.03
x1_transaction_date	-0.05	1.00	0.02	0.06	0.01	0.04	-0.04	0.09
x2_house_age	-0.03	0.02	1.00	0.01	0.05	0.05	-0.05	-0.21
x3_distance_to_the_nearest_mrt_station	-0.01	0.06	0.01	1.00	-0.62	-0.61	-0.79	-0.67
x4_number_of_convenience_stores	-0.01	0.01	0.05	-0.62	1.00	0.44	0.45	0.57
x5_latitude	-0.01	0.04	0.05	-0.61	0.44	1.00	0.41	0.55
x6_longitude	-0.01	-0.04	-0.05	-0.79	0.45	0.41	1.00	0.52
y_house_price_of_unit_area	-0.03	0.09	-0.21	-0.67	0.57	0.55	0.52	1.00

Correlation Matrix

Correlation Matrix

Correlation Matrix

```
Traceback (most recent call last):
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\generate_chat_pipeline.py", line 335, in run
    ).run(input)
      ^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 137, in run
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 101, in run
    step_output = logic.execute(
                  ^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 113, in execute
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 85, in execute
    result = self.execute_code(code_to_run, code_context)
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 175, in execute_code
    raise NoResultFoundError("No result returned")
pandasai.exceptions.NoResultFoundError: No result returned
```

Out[54]: 'Unfortunately, I was not able to answer your question, because of the following e
rror:\n\nNo result returned\n'

In [56]: house_price.chat('is the a correlation amoung the variables?')

Out[56]: 'There is a significant correlation among the variables.'

In [57]: house_price.chat('show this correlation in a table')

Out[57]:

	transaction_id	x1_transaction_date	x2_house_age	x:
transaction_id	1.000000	-0.048658	-0.032808	
x1_transaction_date	-0.048658	1.000000	0.017549	
x2_house_age	-0.032808	0.017549	1.000000	
x3_distance_to_the_nearest_mrt_station	-0.013341	0.060272	0.013843	
x4_number_of_convenience_stores	-0.012699	0.009635	0.049593	
x5_latitude	-0.010110	0.035058	0.054420	
x6_longitude	-0.011059	-0.041082	-0.048520	
y_house_price_of_unit_area	-0.028587	0.087491	-0.210567	

Correlation Matrix
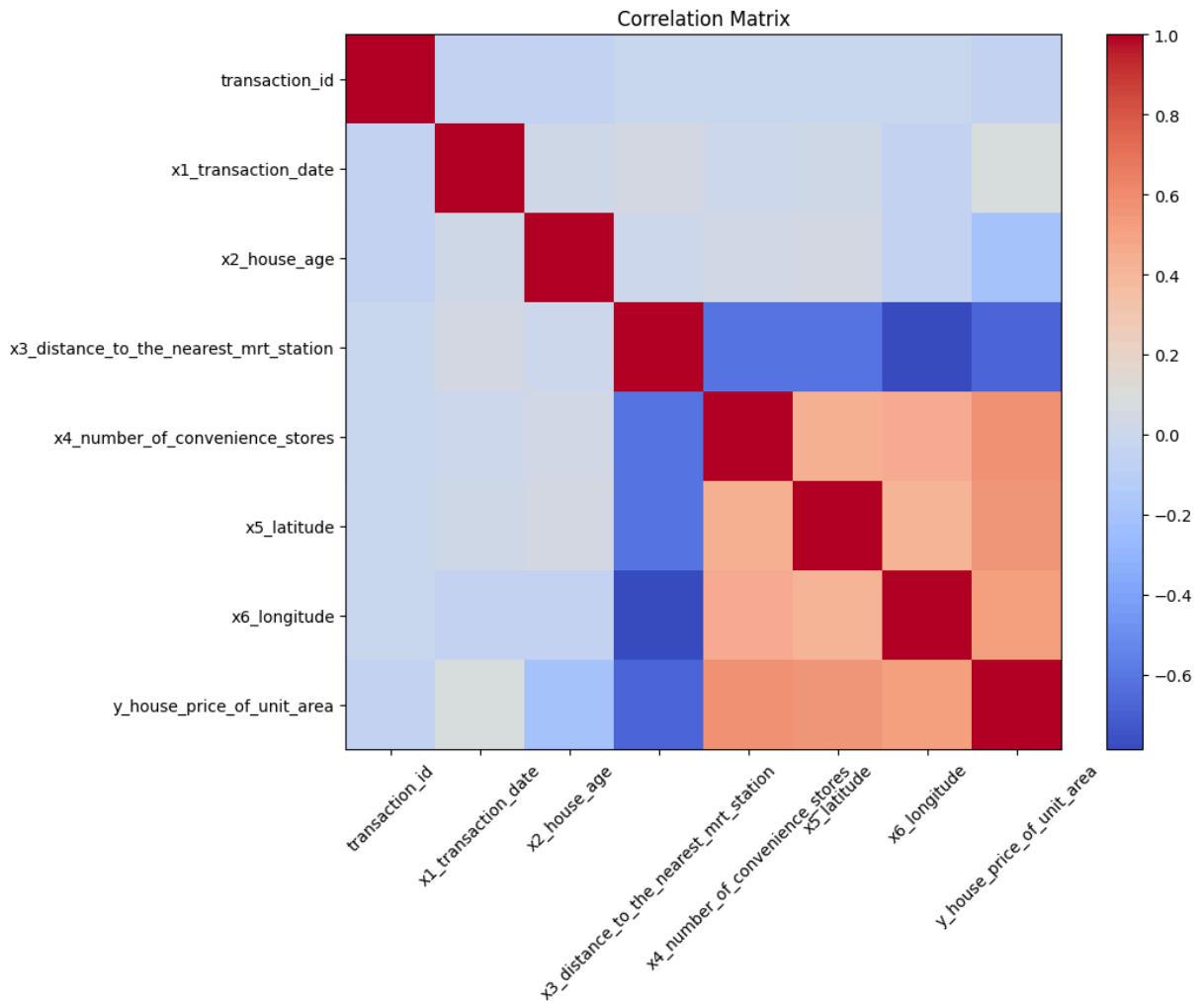
In [ ]:

3 - Decision Tree Regression Model

Defining best predictors and Target

In [61]:
```python
# Finding predictors amoung our variables.
house_price.chat('display the best predictors for the house price of unit area')
```

```
Index(['transaction_id', 'x1_transaction_date', 'x2_house_age',
       'x3_distance_to_the_nearest_mrt_station',
       'x4_number_of_convenience_stores', 'x5_latitude', 'x6_longitude',
       'y_house_price_of_unit_area'],
      dtype='object')
```

	y_house_price_of_unit_area
y_house_price_of_unit_area	1.000000
x4_number_of_convenience_stores	0.571005
x5_latitude	0.546307
x6_longitude	0.523287
x1_transaction_date	0.087491
transaction_id	-0.028587
x2_house_age	-0.210567
x3_distance_to_the_nearest_mrt_station	-0.673104

```
# Defining the best predictors for the model.
house_price.chat('which of them are the best predictors fot the house_price_of_unit
```

	Predictor	Correlation with House Price
0	x3_distance_to_the_nearest_mrt_station	-0.673104
1	x4_number_of_convenience_stores	0.571005
2	x5_latitude	0.546307
3	x6_longitude	0.523287

Correlation Matrix
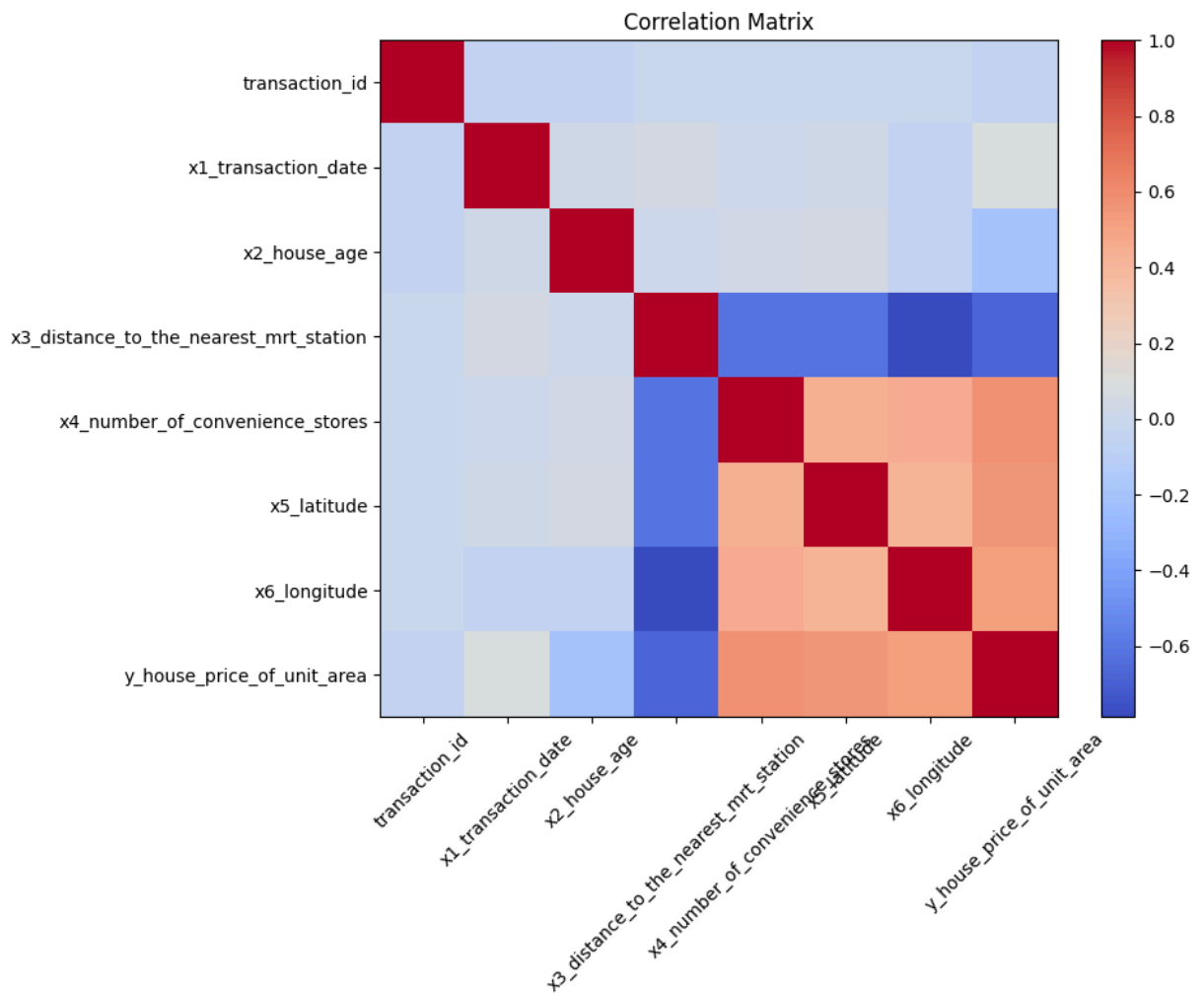
```
In [69]:  # Puting x3_distance_to_the_nearest_mrt_station, x4_number_of_convenience_stores, x
          house_price.chat('define them as a best predictors')

Out[69]:  'The best predictors for house price of unit area are: X2 house age, X3 distance t
          o the nearest MRT station, X4 number of convenience stores, X5 latitude, X6 longit
          ude.'

In [90]:  house_price.chat('show valuse of "y house price of unit area" column')
```

```
Traceback (most recent call last):
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\indexes\base.py", line 3812, in get_loc
    return self._engine.get_loc(casted_key)
           ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "pandas/_libs/index.pyx", line 167, in pandas._libs.index.IndexEngine.get_loc
  File "pandas/_libs/index.pyx", line 196, in pandas._libs.index.IndexEngine.get_loc
  File "pandas/_libs/hashtable_class_helper.pxi", line 7088, in pandas._libs.hashtab
le.PyObjectHashTable.get_item
  File "pandas/_libs/hashtable_class_helper.pxi", line 7096, in pandas._libs.hashtab
le.PyObjectHashTable.get_item
KeyError: 'Y house price of unit area'

The above exception was the direct cause of the following exception:

Traceback (most recent call last):
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\generate_chat_pipeline.py", line 335, in run
    ).run(input)
      ^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 137, in run
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 101, in run
    step_output = logic.execute(
                  ^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 113, in execute
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 85, in execute
    result = self.execute_code(code_to_run, code_context)
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 171, in execute_code
    exec(code, environment)
  File "<string>", line 2, in <module>
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\frame.py", line 4107, in __getitem__
    indexer = self.columns.get_loc(key)
              ^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\indexes\base.py", line 3819, in get_loc
    raise KeyError(key) from err
KeyError: 'Y house price of unit area'
```

Out[90]: "Unfortunately, I was not able to answer your question, because of the following e
          rror:\n\n'Y house price of unit area'\n"

In [ ]: # In my previous attempts pandasai displayed me an column of house_price_of_unit_ar

In [80]: # Putting House price of unit area under the name Target
          # !!! Pandasai displays y_hous_price_of_unit_area as it was before removing outli
          house_price.chat('define this column as a Target')

Spliting dataset into training and validation dataset.

```
In [120…    house_price.chat('crate new dataframe that consists of  target and best predictors'
```

```
Index(['transaction_id', 'x1_transaction_date', 'x2_house_age',
       'x3_distance_to_the_nearest_mrt_station',
       'x4_number_of_convenience_stores', 'x5_latitude', 'x6_longitude',
       'y_house_price_of_unit_area'],
      dtype='object')
Index(['transaction_id', 'x1_transaction_date', 'x2_house_age',
       'x3_distance_to_the_nearest_mrt_station',
       'x4_number_of_convenience_stores', 'x5_latitude', 'x6_longitude',
       'y_house_price_of_unit_area'],
      dtype='object')
Index(['transaction_id', 'x1_transaction_date', 'x2_house_age',
       'x3_distance_to_the_nearest_mrt_station',
       'x4_number_of_convenience_stores', 'x5_latitude', 'x6_longitude',
       'y_house_price_of_unit_area'],
      dtype='object')
```

```
Traceback (most recent call last):
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\generate_chat_pipeline.py", line 335, in run
    ).run(input)
     ^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 137, in run
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 101, in run
    step_output = logic.execute(
                  ^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 113, in execute
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 85, in execute
    result = self.execute_code(code_to_run, code_context)
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 171, in execute_code
    exec(code, environment)
  File "<string>", line 2, in <module>
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\frame.py", line 4113, in __getitem__
    indexer = self.columns._get_indexer_strict(key, "columns")[1]
              ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\indexes\base.py", line 6212, in _get_indexer_strict
    self._raise_if_missing(keyarr, indexer, axis_name)
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\indexes\base.py", line 6261, in _raise_if_missing
    raise KeyError(f"None of [{key}] are in the [{axis_name}]")
KeyError: "None of [Index(['X2 house age', 'X3 distance to the nearest MRT statio
n',\n       'X4 number of convenience stores', 'X5 latitude', 'X6 longitude'],\n
dtype='object')] are in the [columns]"
```

Out[120…   'Unfortunately, I was not able to answer your question, because of the following e
          rror:\n\n"None of [Index([\'X2 house age\', \'X3 distance to the nearest MRT stati
          on\',\\n   \'X4 number of convenience stores\', \'X5 latitude\', \'X6 longitud
          e\'],\\n   dtype=\'object\')] are in the [columns]"\n'

In [118… house_price.chat('split target and best predictors into training and validation dat

          Index(['transaction_id', 'x1_transaction_date', 'x2_house_age',
                 'x3_distance_to_the_nearest_mrt_station',
                 'x4_number_of_convenience_stores', 'x5_latitude', 'x6_longitude',
                 'y_house_price_of_unit_area'],
                dtype='object')

```
Traceback (most recent call last):
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\generate_chat_pipeline.py", line 335, in run
    ).run(input)
     ^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 137, in run
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\pipeline.py", line 101, in run
    step_output = logic.execute(
                  ^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 113, in execute
    raise e
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 85, in execute
    result = self.execute_code(code_to_run, code_context)
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandasai\pipelines\chat\code_execution.py", line 171, in execute_code
    exec(code, environment)
  File "<string>", line 3, in <module>
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\frame.py", line 4113, in __getitem__
    indexer = self.columns._get_indexer_strict(key, "columns")[1]
              ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\indexes\base.py", line 6212, in _get_indexer_strict
    self._raise_if_missing(keyarr, indexer, axis_name)
  File "C:\Users\ÖRNEK AİLESİ\AppData\Local\Programs\Python\Python312\Lib\site-packa
ges\pandas\core\indexes\base.py", line 6261, in _raise_if_missing
    raise KeyError(f"None of [{key}] are in the [{axis_name}]")
KeyError: "None of [Index(['X2 house age', 'X3 distance to the nearest MRT statio
n',\n       'X4 number of convenience stores', 'X5 latitude', 'X6 longitude'],\n
dtype='object')] are in the [columns]"
```

Out[118... 'Unfortunately, I was not able to answer your question, because of the following e
rror:\n\n"None of [Index([\'X2 house age\', \'X3 distance to the nearest MRT stati
on\',\\n   \'X4 number of convenience stores\', \'X5 latitude\', \'X6 longitud
e\'],\\n   dtype=\'object\')] are in the [columns]"\n'

In [ ]: