Data Dictionary

Variable Name	Description
ID	Identifier ID
title	Title from listing
subtitle	Neighborhood and city
sq_mt_built	Square meter built
sq_mt_useful	Square meter useful
n_rooms	Number of rooms
n_bathrooms	Number of bathrooms
sq_mt_allotment	Square meter allotment
latitude, longitude	Latitude, Longitude (Since portals hide the location mainly NAs)
raw_address	Address
is_exact_address_hidden	Boolean values
buy_price	Target Value

In [1]:

```
import pandas as pd
import numpy as np
import plotly.express as px
from matplotlib import pyplot as plt
import seaborn as sns

pd.options.display.float_format = '{:,.2f}'.format
pd.options.display.max_rows = None
pd.options.display.max_columns = None
```

```
sns.set()
```

In [2]:

```
data = pd.read_csv('houses_Madrid.csv')
```

In [3]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21742 entries, 0 to 21741
Data columns (total 58 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Unnamed: 0                    21742 non-null  int64
 1   id                            21742 non-null  int64
 2   title                         21742 non-null  object
 3   subtitle                      21742 non-null  object
 4   sq_mt_built                   21616 non-null  float64
 5   sq_mt_useful                  8228 non-null   float64
 6   n_rooms                       21742 non-null  int64
 7   n_bathrooms                   21726 non-null  float64
 8   n_floors                      1437 non-null   float64
 9   sq_mt_allotment               1432 non-null   float64
 10  latitude                      0 non-null      float64
 11  longitude                     0 non-null      float64
 12  raw_address                   16277 non-null  object
 13  is_exact_address_hidden       21742 non-null  bool
 14  street_name                   15837 non-null  object
 15  street_number                 6300 non-null   object
 16  portal                        0 non-null      float64
 17  floor                         19135 non-null  object
 18  is_floor_under                20572 non-null  object
 19  door                          0 non-null      float64
 20  neighborhood_id               21742 non-null  object
 21  operation                     21742 non-null  object
 22  rent_price                    21742 non-null  int64
 23  rent_price_by_area            0 non-null      float64
 24  is_rent_price_known           21742 non-null  bool
 25  buy_price                     21742 non-null  int64
 26  buy_price_by_area             21742 non-null  int64
 27  is_buy_price_known            21742 non-null  bool
 28  house_type_id                 21351 non-null  object
 29  is_renewal_needed             21742 non-null  bool
 30  is_new_development            20750 non-null  object
 31  built_year                    10000 non-null  float64
 32  has_central_heating           13608 non-null  object
 33  has_individual_heating        13608 non-null  object
 34  are_pets_allowed              0 non-null      float64
 35  has_ac                        11211 non-null  object
 36  has_fitted_wardrobes          13399 non-null  object
 37  has_lift                      19356 non-null  object
 38  is_exterior                   18699 non-null  object
 39  has_garden                    1556 non-null   object
 40  has_pool                      5171 non-null   object
 41  has_terrace                   9548 non-null   object
 42  has_balcony                   3321 non-null   object
 43  has_storage_room              7698 non-null   object
 44  is_furnished                  0 non-null      float64
 45  is_kitchen_equipped           0 non-null      float64
 46  is_accessible                 4074 non-null   object
 47  has_green_zones               4057 non-null   object
 48  energy_certificate            21742 non-null  object
 49  has_parking                   21742 non-null  bool
 50  has_private_parking           0 non-null      float64
 51  has_public_parking            0 non-null      float64
 52  is_parking_included_in_price  7719 non-null   object
 53  parking_price                 7719 non-null   float64
 54  is_orientation_north          11358 non-null  object
 55  is_orientation_west           11358 non-null  object
 56  is_orientation_south          11358 non-null  object
 57  is_orientation_east           11358 non-null  object
dtypes: bool(5), float64(17), int64(6), object(30)
memory usage: 8.9+ MB
```

In [4]:

```
data = data.dropna(axis=1, how='all') #remove column(axis=1), where all values are
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21742 entries, 0 to 21741
Data columns (total 48 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Unnamed: 0                    21742 non-null  int64
 1   id                            21742 non-null  int64
 2   title                         21742 non-null  object
 3   subtitle                      21742 non-null  object
 4   sq_mt_built                   21616 non-null  float64
 5   sq_mt_useful                  8228 non-null   float64
 6   n_rooms                       21742 non-null  int64
 7   n_bathrooms                   21726 non-null  float64
 8   n_floors                      1437 non-null   float64
 9   sq_mt_allotment               1432 non-null   float64
 10  raw_address                   16277 non-null  object
 11  is_exact_address_hidden       21742 non-null  bool
 12  street_name                   15837 non-null  object
 13  street_number                 6300 non-null   object
 14  floor                         19135 non-null  object
 15  is_floor_under                20572 non-null  object
 16  neighborhood_id               21742 non-null  object
 17  operation                     21742 non-null  object
 18  rent_price                    21742 non-null  int64
 19  is_rent_price_known           21742 non-null  bool
 20  buy_price                     21742 non-null  int64
 21  buy_price_by_area             21742 non-null  int64
 22  is_buy_price_known            21742 non-null  bool
 23  house_type_id                 21351 non-null  object
 24  is_renewal_needed             21742 non-null  bool
 25  is_new_development            20750 non-null  object
 26  built_year                    10000 non-null  float64
 27  has_central_heating           13608 non-null  object
 28  has_individual_heating        13608 non-null  object
 29  has_ac                        11211 non-null  object
 30  has_fitted_wardrobes          13399 non-null  object
 31  has_lift                      19356 non-null  object
 32  is_exterior                   18699 non-null  object
 33  has_garden                    1556 non-null   object
 34  has_pool                      5171 non-null   object
 35  has_terrace                   9548 non-null   object
 36  has_balcony                   3321 non-null   object
 37  has_storage_room              7698 non-null   object
 38  is_accessible                 4074 non-null   object
 39  has_green_zones               4057 non-null   object
 40  energy_certificate            21742 non-null  object
 41  has_parking                   21742 non-null  bool
 42  is_parking_included_in_price  7719 non-null   object
 43  parking_price                 7719 non-null   float64
 44  is_orientation_north          11358 non-null  object
 45  is_orientation_west           11358 non-null  object
 46  is_orientation_south          11358 non-null  object
 47  is_orientation_east           11358 non-null  object
dtypes: bool(5), float64(7), int64(6), object(30)
memory usage: 7.2+ MB
```

In [5]:

```
data.duplicated().sum()
```

Out[5]:

```
np.int64(0)
```

In [6]:

```
df = data[['id', 'subtitle', 'sq_mt_built', 'sq_mt_useful', 'n_rooms', 'n_bathrooms
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21742 entries, 0 to 21741
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   id                 21742 non-null  int64  
 1   subtitle           21742 non-null  object 
 2   sq_mt_built        21616 non-null  float64
 3   sq_mt_useful       8228 non-null   float64
 4   n_rooms            21742 non-null  int64  
 5   n_bathrooms        21726 non-null  float64
 6   n_floors           1437 non-null   float64
 7   sq_mt_allotment    1432 non-null   float64
 8   rent_price         21742 non-null  int64  
 9   buy_price          21742 non-null  int64  
 10  buy_price_by_area  21742 non-null  int64  
 11  built_year         10000 non-null  float64
 12  parking_price      7719 non-null   float64
dtypes: float64(7), int64(5), object(1)
memory usage: 2.2+ MB
```

In [7]:

```
df=df.rename(columns={'subtitle' : 'city'})
```

In [8]:

```
df = df[['id', 'city', 'sq_mt_built', 'n_rooms', 'n_bathrooms', 'rent_price', 'buy_
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21742 entries, 0 to 21741
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   id                 21742 non-null  int64  
 1   city               21742 non-null  object 
 2   sq_mt_built        21616 non-null  float64
 3   n_rooms            21742 non-null  int64  
 4   n_bathrooms        21726 non-null  float64
 5   rent_price         21742 non-null  int64  
 6   buy_price          21742 non-null  int64  
 7   buy_price_by_area  21742 non-null  int64  
dtypes: float64(2), int64(5), object(1)
memory usage: 1.3+ MB
```

In [9]:

```
df.n_rooms.describe()
```

```
Out[9]:  count   21,742.00
         mean         3.01
         std          1.51
         min          0.00
         25%          2.00
         50%          3.00
         75%          4.00
         max         24.00
         Name: n_rooms, dtype: float64
```

```
In [10]: df.drop(df.index[df.n_rooms == 0], axis=0, inplace=True)
```

```
In [11]: df.n_bathrooms.mode()
```

```
Out[11]: 0   1.00
         Name: n_bathrooms, dtype: float64
```

```
In [12]: df.n_bathrooms.fillna(1, inplace=True)
         df.n_bathrooms = df.n_bathrooms.replace(0,1)
```

```
C:\Users\alhef\AppData\Local\Temp\ipykernel_7008\3396593470.py:1: FutureWarning: A v
alue is trying to be set on a copy of a DataFrame or Series through chained assignme
nt using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because
the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method
({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform
the operation inplace on the original object.


  df.n_bathrooms.fillna(1, inplace=True)
```

```
In [13]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 21303 entries, 0 to 21741
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   id                 21303 non-null  int64  
 1   city               21303 non-null  object 
 2   sq_mt_built        21177 non-null  float64
 3   n_rooms            21303 non-null  int64  
 4   n_bathrooms        21303 non-null  float64
 5   rent_price         21303 non-null  int64  
 6   buy_price          21303 non-null  int64  
 7   buy_price_by_area  21303 non-null  int64  
dtypes: float64(2), int64(5), object(1)
memory usage: 1.5+ MB
```

```
In [14]: df.sample(10)
```

Out[14]:

	id	city	sq_mt_built	n_rooms	n_bathrooms	rent_price	buy_price
18259	3483	Malasaña-Universidad, Madrid	43.00	1	1.00	936	206000
4774	16968	Palomeras Bajas, Madrid	53.00	1	1.00	621	119000
2254	19488	Almendrales, Madrid	102.00	2	1.00	508	93000
21630	112	Chueca-Justicia, Madrid	45.00	1	1.00	1275	330000
6328	15414	Águilas, Madrid	80.00	3	2.00	964	215000
12923	8819	Atalaya, Madrid	169.00	3	3.00	2378	825000
9143	12599	Pilar, Madrid	54.00	1	1.00	851	180000
9292	12450	Fuencarral, Madrid	74.00	3	1.00	837	176000
6650	15092	Lucero, Madrid	48.00	1	1.00	625	119900
14058	7684	Bernabéu-Hispanoamérica, Madrid	330.00	5	4.00	-2214	1495000

In [15]:

```
df.n_bathrooms = df.n_bathrooms.astype(int)
```

In [16]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 21303 entries, 0 to 21741
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   id                 21303 non-null  int64  
 1   city               21303 non-null  object 
 2   sq_mt_built        21177 non-null  float64
 3   n_rooms            21303 non-null  int64  
 4   n_bathrooms        21303 non-null  int64  
 5   rent_price         21303 non-null  int64  
 6   buy_price          21303 non-null  int64  
 7   buy_price_by_area  21303 non-null  int64  
dtypes: float64(1), int64(6), object(1)
memory usage: 1.5+ MB
```

In [17]:

```
#drop sq_mt_built NaN rows
df.dropna(subset='sq_mt_built', inplace=True)
```

In [18]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 21177 entries, 0 to 21741
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   id                 21177 non-null  int64  
 1   city               21177 non-null  object 
 2   sq_mt_built        21177 non-null  float64
 3   n_rooms            21177 non-null  int64  
 4   n_bathrooms        21177 non-null  int64  
 5   rent_price         21177 non-null  int64  
 6   buy_price          21177 non-null  int64  
 7   buy_price_by_area  21177 non-null  int64  
dtypes: float64(1), int64(6), object(1)
memory usage: 1.5+ MB
```

In [19]:

```
df.sample(10)
```

Out[19]:

	id	city	sq_mt_built	n_rooms	n_bathrooms	rent_price	buy_price	bu
1234	20508	Vicálvaro, Madrid	88.00	3	2	1174	290000	
13403	8339	Ventas, Madrid	62.00	2	1	666	130000	
8915	12827	Las Tablas, Madrid	93.00	3	2	1644	490000	
570	21172	Los Ángeles, Madrid	70.00	3	1	621	119000	
17298	4444	Lavapiés-Embajadores, Madrid	57.00	1	1	1035	239000	
5572	16170	Numancia, Madrid	65.00	2	1	662	129000	
6031	15711	Los Cármenes, Madrid	80.00	3	1	686	135000	
76	21666	Butarque, Madrid	117.00	3	2	1052	244798	
534	21208	San Cristóbal, Madrid	62.00	3	1	574	108000	
20869	873	Legazpi, Madrid	140.00	3	2	1428	395000	

In [20]:

```
#change negative values in Rent price column to Zero
for value in df.rent_price.values:
    if value < 0:
        df.rent_price = 0
```

In [21]:
```
df[df.rent_price < 0]
```

Out[21]:

id	city	sq_mt_built	n_rooms	n_bathrooms	rent_price	buy_price	buy_price_by_area

In [22]:
```
#split City Column to (district, city)
df['district'] = df.city.str.split(',', expand=True)[0]
```

In [23]:
```
df.head()
```

Out[23]:

	id	city	sq_mt_built	n_rooms	n_bathrooms	rent_price	buy_price	buy_price_b
0	21742	San Cristóbal, Madrid	64.00	2	1	0	85000	
1	21741	Los Ángeles, Madrid	70.00	3	1	0	129900	
2	21740	San Andrés, Madrid	94.00	2	2	0	144247	
3	21739	San Andrés, Madrid	64.00	2	1	0	109900	
4	21738	Los Rosales, Madrid	108.00	2	2	0	260000	

In [24]:
```
df = df[['id', 'district', 'sq_mt_built', 'n_rooms', 'n_bathrooms', 'rent_price',
         'buy_price', 'buy_price_by_area']]
df.head()
```

Out[24]:

	id	district	sq_mt_built	n_rooms	n_bathrooms	rent_price	buy_price	buy_price_b
0	21742	San Cristóbal	64.00	2	1	0	85000	
1	21741	Los Ángeles	70.00	3	1	0	129900	
2	21740	San Andrés	94.00	2	2	0	144247	
3	21739	San Andrés	64.00	2	1	0	109900	
4	21738	Los Rosales	108.00	2	2	0	260000	

In [25]:

```
data.head()
```

Out[25]:

	Unnamed: 0	id	title	subtitle	sq_mt_built	sq_mt_useful	n_rooms	n_bathroom
0	0	21742	Piso en venta en calle de Godella, 64	San Cristóbal, Madrid	64.00	60.00	2	1.0
1	1	21741	Piso en venta en calle de la del Manojo de Rosas	Los Ángeles, Madrid	70.00	NaN	3	1.0
2	2	21740	Piso en venta en calle del Talco, 68	San Andrés, Madrid	94.00	54.00	2	2.0
3	3	21739	Piso en venta en calle Pedro Jiménez	San Andrés, Madrid	64.00	NaN	2	1.0
4	4	21738	Piso en venta en carretera de Villaverde a Val...	Los Rosales, Madrid	108.00	90.00	2	2.0

In [28]:

```
data[['raw_address','is_exact_address_hidden','street_name']].info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21742 entries, 0 to 21741
Data columns (total 3 columns):
 #   Column                  Non-Null Count  Dtype 
---  ------                  --------------  ----- 
 0   raw_address             16277 non-null  object
 1   is_exact_address_hidden 21742 non-null  bool  
 2   street_name             15837 non-null  object
dtypes: bool(1), object(2)
memory usage: 361.1+ KB
```

In [29]:
```
data['operation'].unique()
```

Out[29]:
```
array(['sale'], dtype=object)
```

In [30]:
```
# Remove rent_price
df.drop(columns=['rent_price'], inplace=True)
```

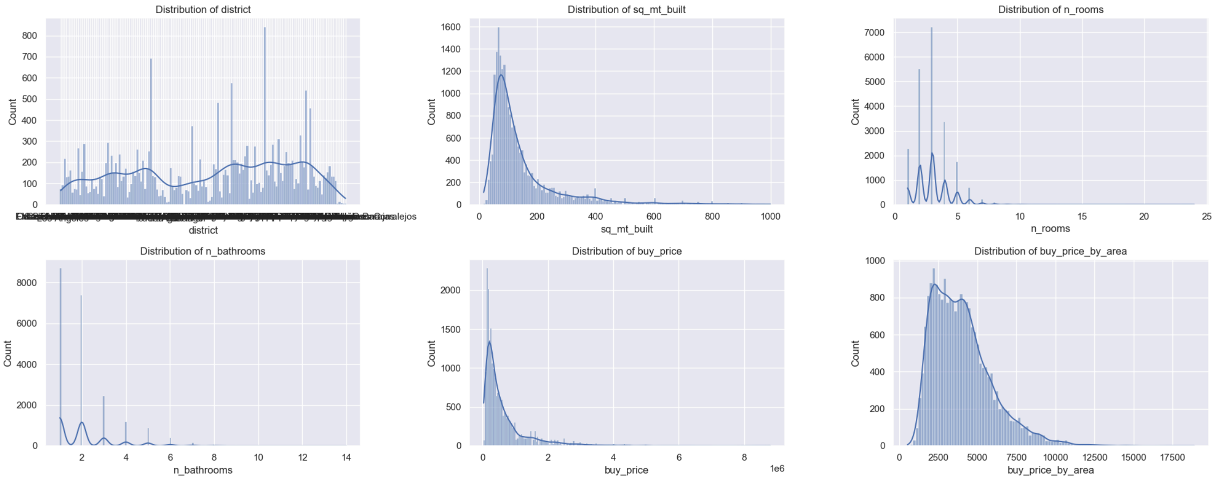
In [31]:
```
df.to_csv('Madrid_cleaned_data.csv')
```

In [32]:
```
df = pd.read_csv('Madrid_cleaned_data.csv')
```

In [33]:
```
df.info()
```

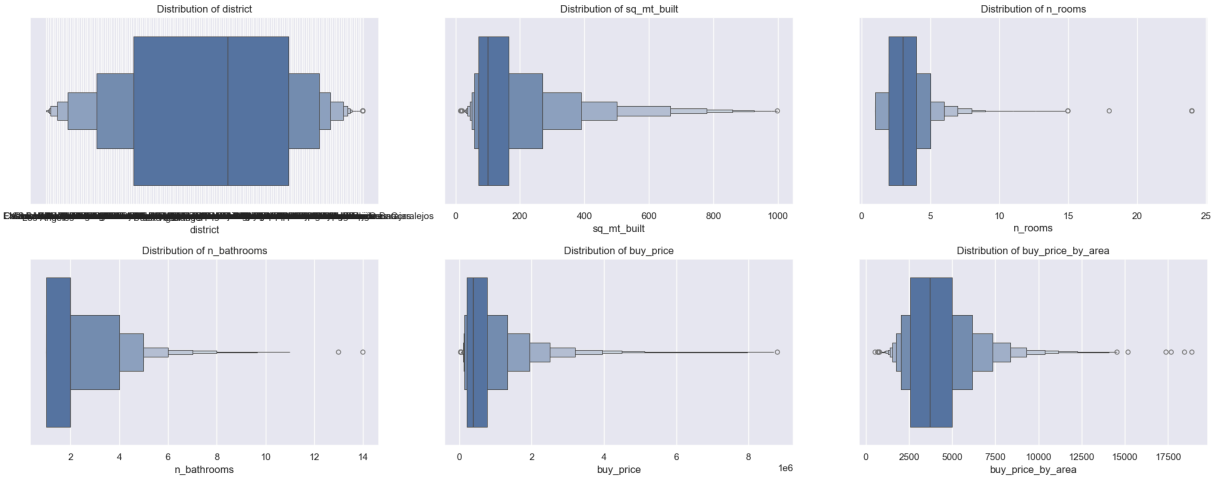
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21177 entries, 0 to 21176
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Unnamed: 0         21177 non-null  int64  
 1   id                 21177 non-null  int64  
 2   district           21177 non-null  object 
 3   sq_mt_built        21177 non-null  float64
 4   n_rooms            21177 non-null  int64  
 5   n_bathrooms        21177 non-null  int64  
 6   buy_price          21177 non-null  int64  
 7   buy_price_by_area  21177 non-null  int64  
dtypes: float64(1), int64(6), object(1)
memory usage: 1.3+ MB
```

In [34]:
```
plt.figure(figsize=(20,8))
for e, i in enumerate(df.columns[2:]):
    plt.subplot(2,3, e+1)
    plt.title('Distribution of ' + i)
    sns.histplot(df[i],kde=True)
    plt.tight_layout()
```



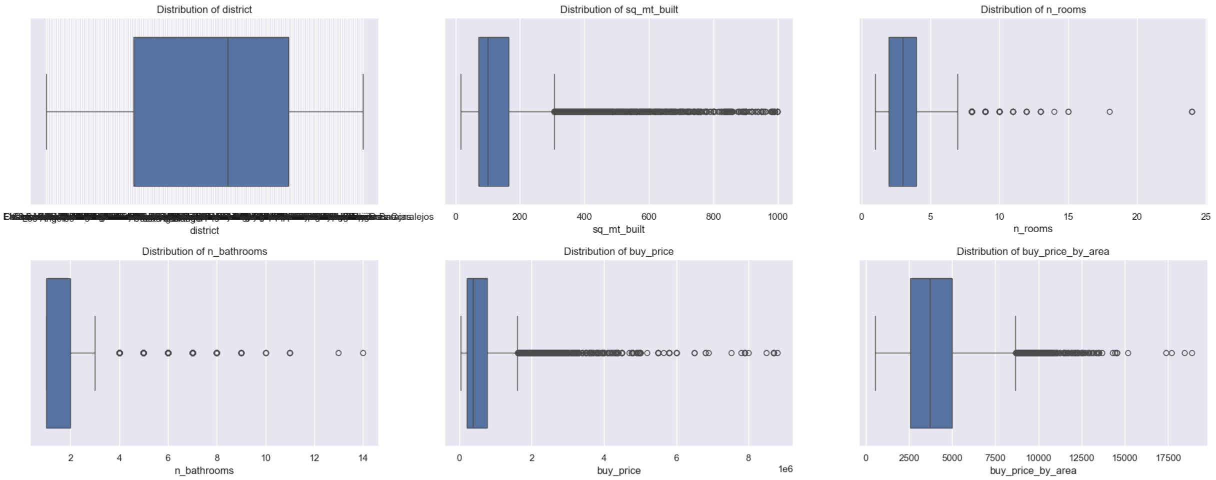
In [36]:

```
plt.figure(figsize=(20,8))
for e, i in enumerate(df.columns[2:]):
    plt.subplot(2,3, e+1)
    plt.title('Distribution of ' + i)
    sns.boxenplot(x=df[i])
    plt.tight_layout()
```

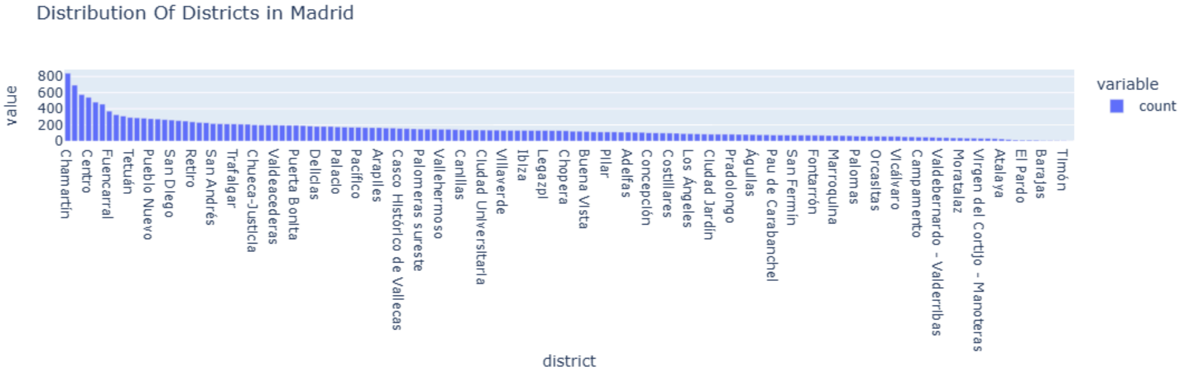


In [37]:

```
plt.figure(figsize=(20,8))
for e, i in enumerate(df.columns[2:]):
    plt.subplot(2,3, e+1)
    plt.title('Distribution of ' + i)
    sns.boxplot(x=df[i])
    plt.tight_layout()
```

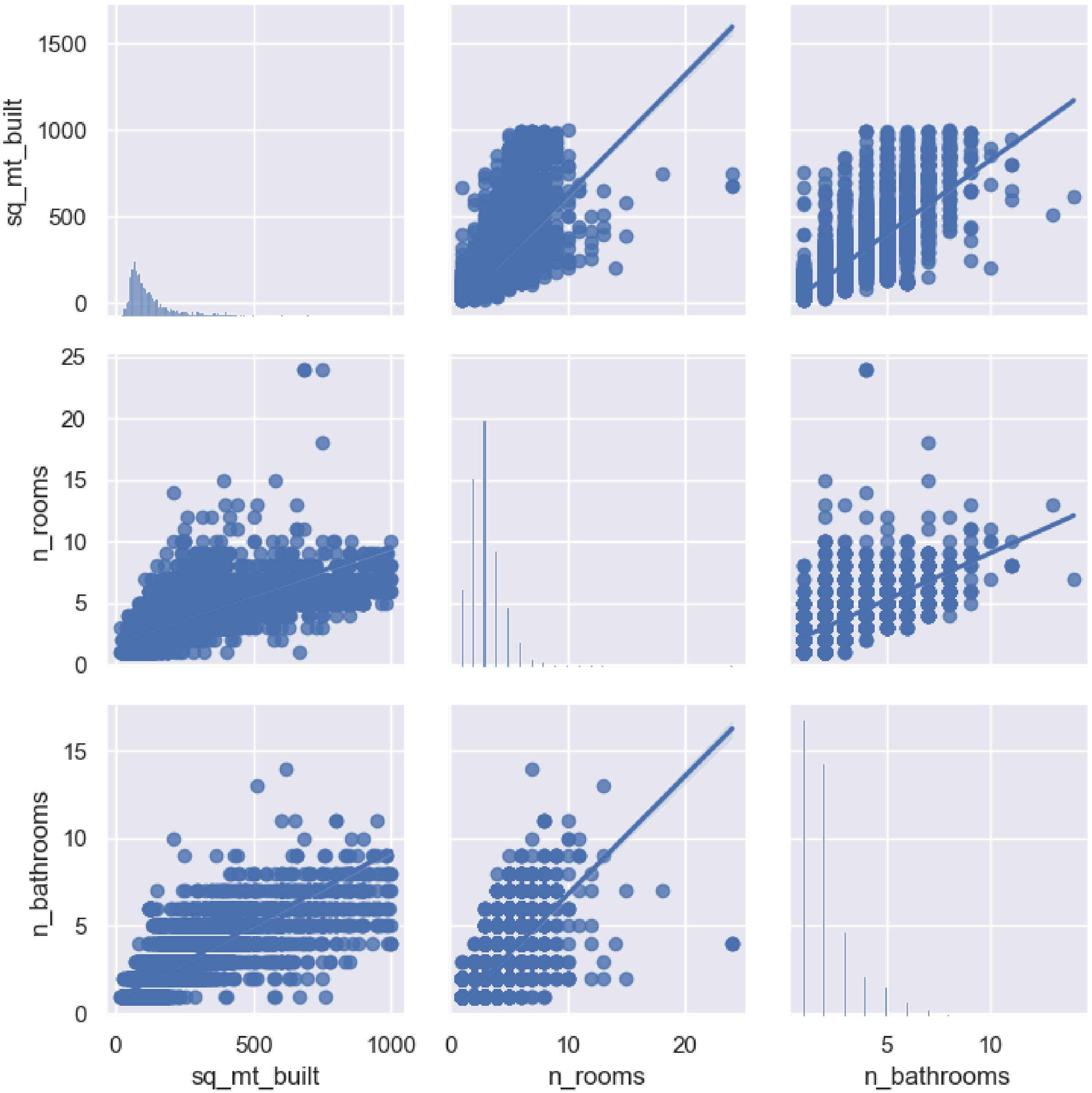


```
In [39]: px.bar(df.district.value_counts(), title= 'Distribution Of Districts in Madrid' )
```



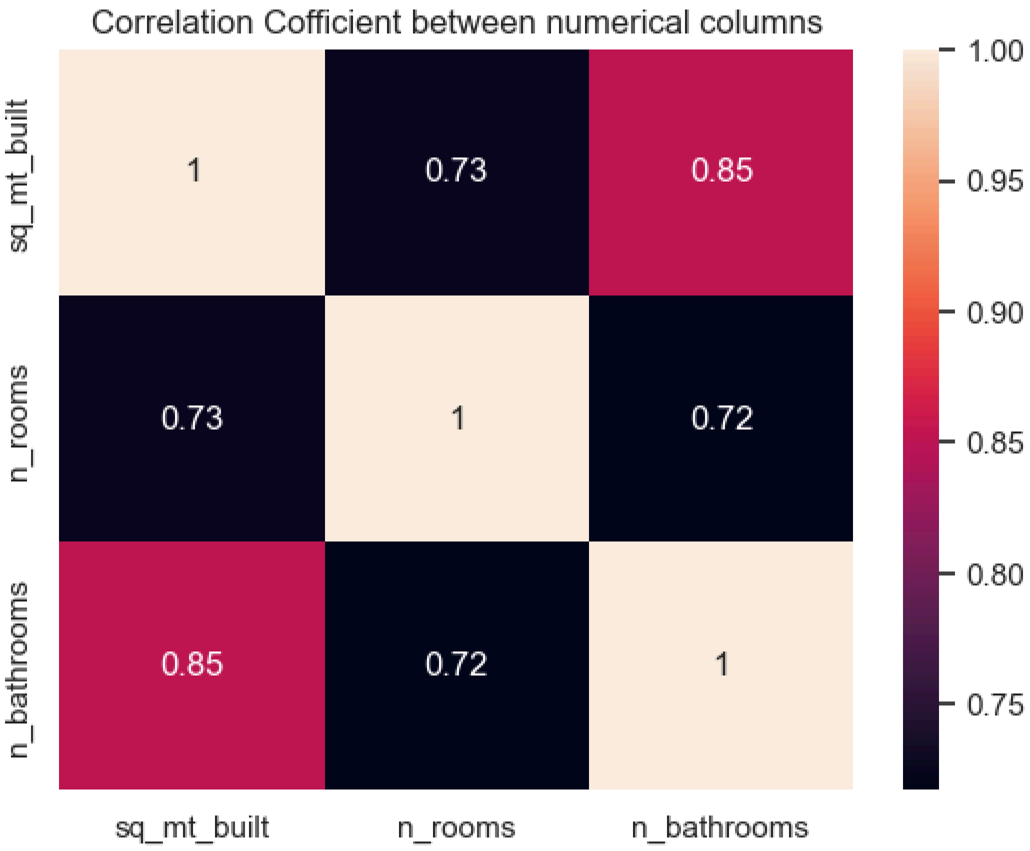
Correllation Coeffcient

```
In [40]: corr = df[['sq_mt_built', 'n_rooms', 'n_bathrooms']]
         sns.pairplot(corr,kind='reg')
         plt.tight_layout()
```



```
In [41]: sns.heatmap(corr.corr(), annot=True)
         plt.title('Correlation Cofficient between numerical columns')
```

```
Out[41]: Text(0.5, 1.0, 'Correlation Cofficient between numerical columns')
```

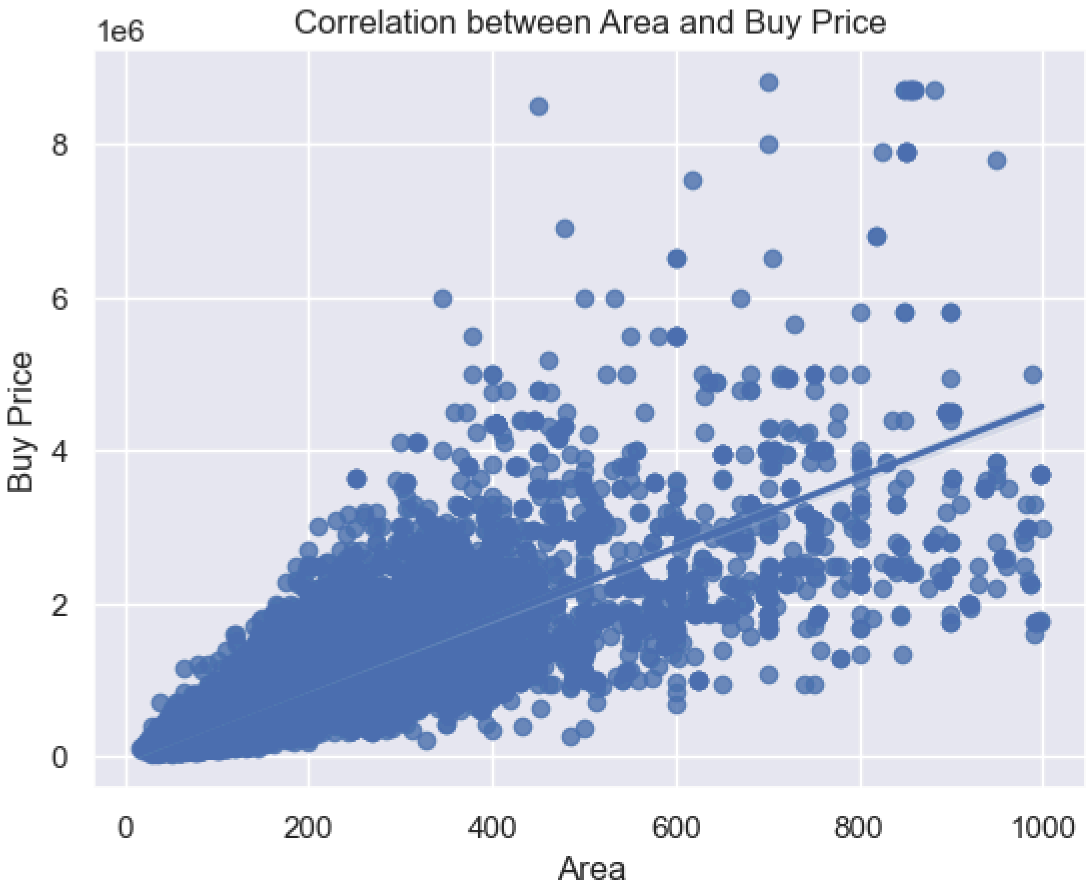


In [42]:

```
sns.regplot(data=df, x='sq_mt_built', y='buy_price')
plt.title('Correlation between Area and Buy Price')
plt.xlabel('Area')
plt.ylabel('Buy Price')
```

Out[42]:

```
Text(0, 0.5, 'Buy Price')
```



In [43]:

```
df.head()
```

Out[43]:

	Unnamed: 0	id	district	sq_mt_built	n_rooms	n_bathrooms	buy_price	buy_price_b
0	0	21742	San Cristóbal	64.00	2	1	85000	
1	1	21741	Los Ángeles	70.00	3	1	129900	
2	2	21740	San Andrés	94.00	2	2	144247	
3	3	21739	San Andrés	64.00	2	1	109900	
4	4	21738	Los Rosales	108.00	2	2	260000	

In [46]:

```
df.drop('Unnamed: 0', axis=1, inplace=True)
```

In [47]:

```
df.head()
```

Out[47]:

	id	district	sq_mt_built	n_rooms	n_bathrooms	buy_price	buy_price_by_area
0	21742	San Cristóbal	64.00	2	1	85000	1328
1	21741	Los Ángeles	70.00	3	1	129900	1856
2	21740	San Andrés	94.00	2	2	144247	1535
3	21739	San Andrés	64.00	2	1	109900	1717
4	21738	Los Rosales	108.00	2	2	260000	2407

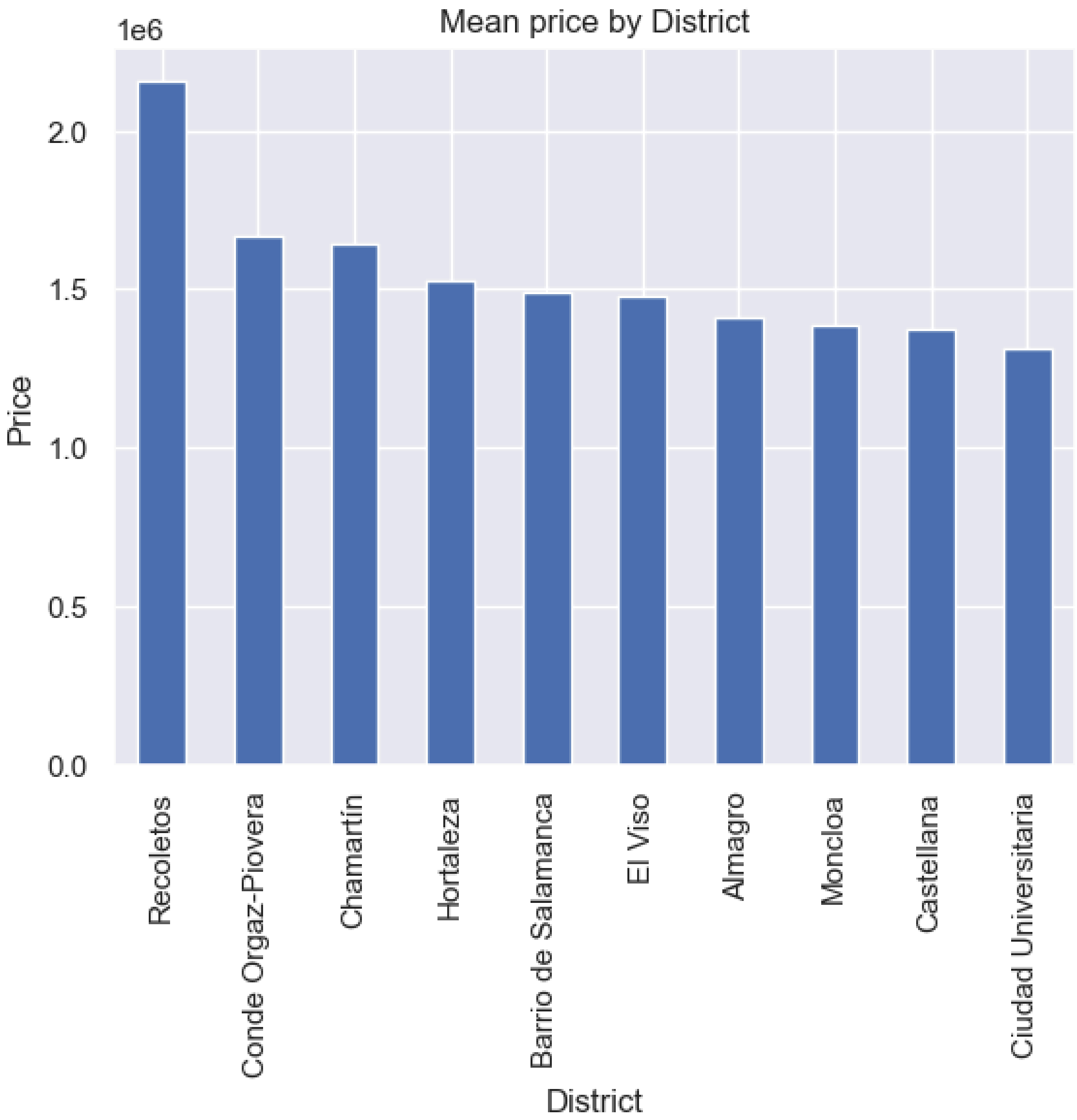
Location or Size : What influence House prices in Madrid?

In [50]:

```
mean_price = df.groupby('district')['buy_price'].mean().sort_values(ascending=False
mean_price.head(10).plot(kind='bar', xlabel='District', ylabel = 'Price', title='Me
```

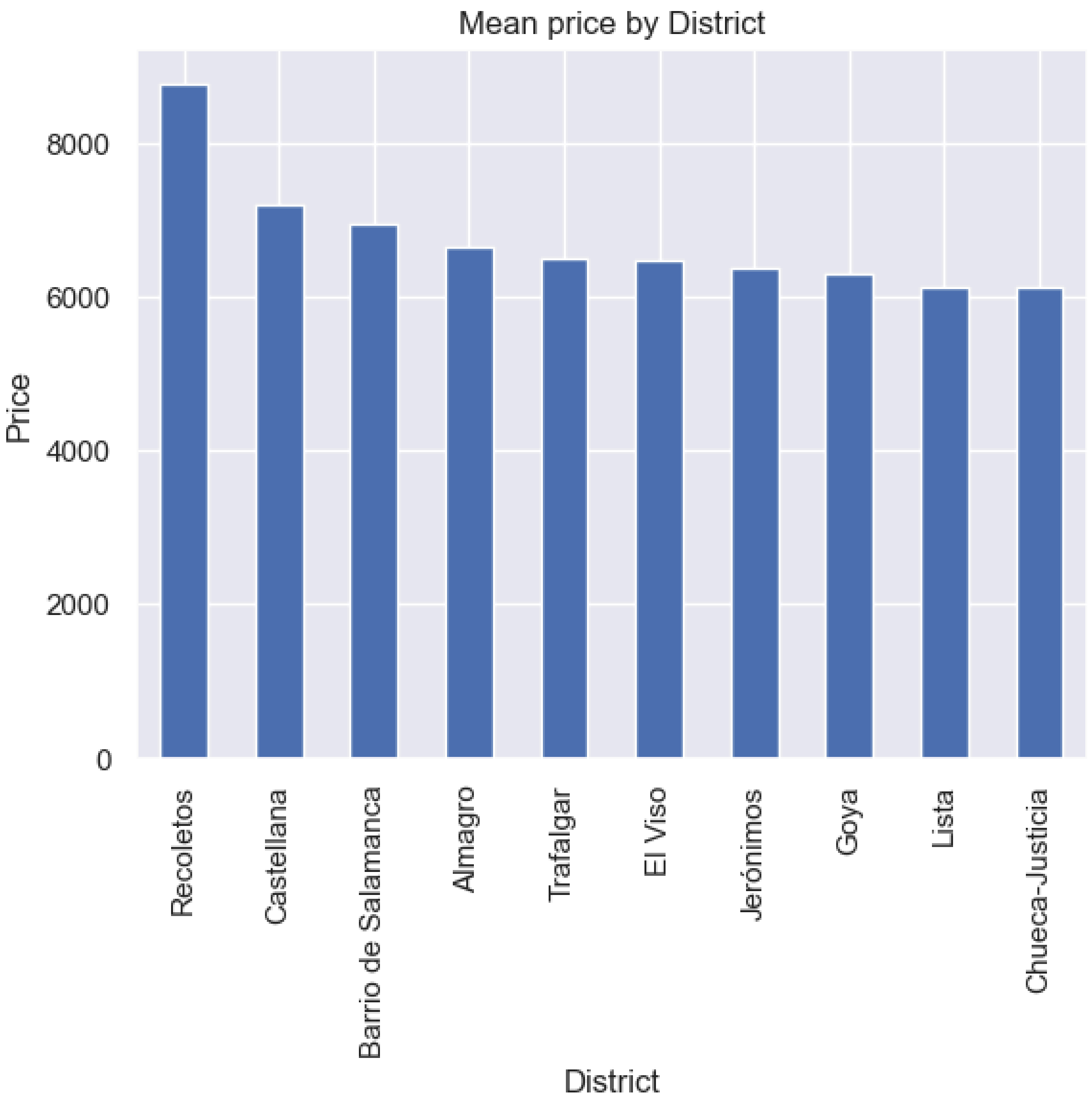
Out[50]:

```
<Axes: title={'center': 'Mean price by District'}, xlabel='District', ylabel='Pric
e'>
```



```
In [51]:  mean_price = df.groupby('district')['buy_price_by_area'].mean().sort_values(ascendi
          mean_price.head(10).plot(kind='bar', xlabel='District', ylabel = 'Price', title='Me
```

```
Out[51]:  <Axes: title={'center': 'Mean price by District'}, xlabel='District', ylabel='Pric
          e'>
```



In [53]:
```
reco = df.query('district == "Recoletos"')
```

In [54]:
```
reco.sample(10)
```

Out[54]:

	id	district	sq_mt_built	n_rooms	n_bathrooms	buy_price	buy_price_by_area
18917	2297	Recoletos	404.00	5	6	4350000	10767
18596	2620	Recoletos	92.00	2	1	378000	4109
18962	2250	Recoletos	210.00	6	3	1800000	8571
19168	2043	Recoletos	410.00	5	3	3300000	8049
19412	1795	Recoletos	404.00	5	6	4350000	10767
19160	2051	Recoletos	247.00	2	3	2200000	8907
19171	2040	Recoletos	706.00	6	7	6500000	9207
18809	2406	Recoletos	471.00	6	6	4155000	8822
18766	2449	Recoletos	365.00	4	5	3285000	9000
19137	2074	Recoletos	70.00	1	1	568000	8114

In [ ]:

```

```

In [56]:

```
pip install statsmodels
```

```
Collecting statsmodels
  Downloading statsmodels-0.14.5-cp313-cp313-win_amd64.whl.metadata (9.8 kB)
Requirement already satisfied: numpy<3,>=1.22.3 in c:\users\alhef\appdata\local\prog
rams\python\python313\lib\site-packages (from statsmodels) (2.3.2)
Requirement already satisfied: scipy!=1.9.2,>=1.8 in c:\users\alhef\appdata\local\pr
ograms\python\python313\lib\site-packages (from statsmodels) (1.16.1)
Requirement already satisfied: pandas!=2.1.0,>=1.4 in c:\users\alhef\appdata\local\p
rograms\python\python313\lib\site-packages (from statsmodels) (2.3.1)
Collecting patsy>=0.5.6 (from statsmodels)
  Downloading patsy-1.0.1-py2.py3-none-any.whl.metadata (3.3 kB)
Requirement already satisfied: packaging>=21.3 in c:\users\alhef\appdata\local\progr
ams\python\python313\lib\site-packages (from statsmodels) (25.0)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\alhef\appdata\loca
l\programs\python\python313\lib\site-packages (from pandas!=2.1.0,>=1.4->statsmodel
s) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\alhef\appdata\local\programs
\python\python313\lib\site-packages (from pandas!=2.1.0,>=1.4->statsmodels) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\alhef\appdata\local\progra
ms\python\python313\lib\site-packages (from pandas!=2.1.0,>=1.4->statsmodels) (2025.
2)
Requirement already satisfied: six>=1.5 in c:\users\alhef\appdata\local\programs\pyt
hon\python313\lib\site-packages (from python-dateutil>=2.8.2->pandas!=2.1.0,>=1.4->s
tatsmodels) (1.17.0)
Downloading statsmodels-0.14.5-cp313-cp313-win_amd64.whl (9.6 MB)
   ---------------------------------------- 0.0/9.6 MB ? eta -:--:--
   ---------------------------------------- 0.0/9.6 MB ? eta -:--:--
   -- ------------------------------------- 0.5/9.6 MB 1.8 MB/s eta 0:00:05
   ----- ---------------------------------- 1.3/9.6 MB 2.5 MB/s eta 0:00:04
   ------- -------------------------------- 1.8/9.6 MB 2.6 MB/s eta 0:00:03
   --------- ------------------------------ 2.4/9.6 MB 2.7 MB/s eta 0:00:03
   ------------- -------------------------- 3.1/9.6 MB 2.9 MB/s eta 0:00:03
   ----------------- ---------------------- 4.2/9.6 MB 3.1 MB/s eta 0:00:02
   -------------------- ------------------- 5.0/9.6 MB 3.2 MB/s eta 0:00:02
   ------------------------- -------------- 6.0/9.6 MB 3.4 MB/s eta 0:00:02
   ---------------------------- ----------- 6.8/9.6 MB 3.4 MB/s eta 0:00:01
   ------------------------------- -------- 7.6/9.6 MB 3.5 MB/s eta 0:00:01
   ---------------------------------- ----- 8.4/9.6 MB 3.5 MB/s eta 0:00:01
   -------------------------------------- - 9.2/9.6 MB 3.5 MB/s eta 0:00:01
   ---------------------------------------- 9.6/9.6 MB 3.4 MB/s  0:00:02
Downloading patsy-1.0.1-py2.py3-none-any.whl (232 kB)
Installing collected packages: patsy, statsmodels

   ---------------------------------------- 0/2 [patsy]
   ---------------------------------------- 0/2 [patsy]
   ---------------------------------------- 0/2 [patsy]
   ---------------------------------------- 0/2 [patsy]
   ---------------------------------------- 0/2 [patsy]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
```

```
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
```

```
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
-------------------- ------------------- 1/2 [statsmodels]
```

```
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   -------------------- ------------------- 1/2 [statsmodels]
   ---------------------------------------- 2/2 [statsmodels]

Successfully installed patsy-1.0.1 statsmodels-0.14.5
Note: you may need to restart the kernel to use updated packages.
```

In [58]:

```python
import statsmodels.api as sm
reco['intercept'] = 1
lm = sm.OLS(reco['buy_price'], reco[['intercept', 'sq_mt_built']])
result = lm.fit()
result.summary()
```

```
C:\Users\alhef\AppData\Local\Temp\ipykernel_7008\2760881175.py:2: SettingWithCopyWar
ning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/u
ser_guide/indexing.html#returning-a-view-versus-a-copy
```

Out[58]:

OLS Regression Results

Dep. Variable:	buy_price	**R-squared:**	0.833
Model:	OLS	**Adj. R-squared:**	0.832
Method:	Least Squares	**F-statistic:**	555.1
Date:	Sat, 16 Aug 2025	**Prob (F-statistic):**	5.30e-45
Time:	14:14:00	**Log-Likelihood:**	-1654.1
No. Observations:	113	**AIC:**	3312.
Df Residuals:	111	**BIC:**	3318.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
intercept	-4.342e+04	1.07e+05	-0.406	0.686	-2.55e+05	1.69e+05
sq_mt_built	8975.6204	380.945	23.561	0.000	8220.753	9730.488

Omnibus:	11.498	**Durbin-Watson:**	2.039
Prob(Omnibus):	0.003	**Jarque-Bera (JB):**	16.048
Skew:	0.512	**Prob(JB):**	0.000328
Kurtosis:	4.536	**Cond. No.**	574.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [59]:

```
x = 300
#YHat = b0 + b1 *x
price = 152500 + 6843*x
print("Predicted price of flat in Recoletos with area of {} m2 = {} Euro".format(x,
```

```
Predicted price of flat in Recoletos with area of 300 m2 = 2205400 Euro
```

In [ ]: