Dataset: Reading Property Data December 2022

Data source: Zoopla (webscraping)

Brief description:

The original dataset contained information on 950 properties which were within a 10-mile radius of the Reading (Berkshire, UK) area. Data was obtained through webscraping on Zoopla. The data included listing price, listing info, room info, and location.

Area of interest: 2-bed and 3-bed properties in this area

Key findings (properties under £1m):

- Of the 887 properties on the market which were under £1m, 29% were 3-bed properties, 28% were 2-bed, 22% were 4-bed, and 14% were 1-bed.

- The median list price of a flat is: £258,475, whilst the median list price of a house is: £515,000.

- 2-bed flat vs. 2-bed house:

  - The median list price of a 2-bed flat is £290,000 (lower quartile = £250,000; upper quartile = £327,000).
  - The median list price of a 2-bed house is £350,000 (lower quartile = £300,000; upper quartile = £412,000).
  - This represents an increase of 20.69%.

- 3-bed flat vs. 3-bed house:
  - The median list price of a 3-bed flat is £325,000 (lower quartile = £290,000; upper quartile = £335,000).
  - The median list price of a 3-bed house is £460,000 (lower quartile = £400,000; upper quartile = £550,000).
  - This represents an increase of 41.54%.

Preliminary steps

Import necessary Python packages

```
In [ ]:  import numpy as np
         import pandas as pd
         import seaborn as sns
         import matplotlib.pyplot as plt
```

Import dataset (read in csv file) and view the first 10 rows

```
In [ ]:   df = pd.read_csv("/Users/andreakoko/Documents/Reading_housing_data_Dec_20
          df.head(10)
```

Out[ ]:

	Unnamed: 0	Listing price	Listing info	All rooms	Location
0	0	£850,000	4 bed detached house for sale	Bedrooms4Bathrooms3Living rooms3	Bluebell Crescent, Woodley, Reading RG5
1	1	£290,000	2 bed flat for sale	Bedrooms2Bathrooms1Living rooms1	Kennet Walk, Reading RG1
2	2	£325,000	2 bed flat for sale	Bedrooms2Bathrooms1Living rooms1	46 Bath Road, Reading RG1
3	3	£200,000	1 bed flat for sale	Bedrooms1Bathrooms1Living rooms1	Russell Street, Reading RG1
4	4	£499,000	3 bed semi-detached house for sale	Bedrooms3Bathrooms2Living rooms1	Woolhampton Way, Reading RG2
5	5	£525,000	3 bed semi-detached house for sale	Bedrooms3Bathrooms1Living rooms2	Southcote Farm Lane, Reading RG30
6	6	£700,000	6 bed semi-detached house for sale	Bedrooms6Bathrooms3Living rooms1	Beech Lane, Reading RG6
7	7	£625,000	4 bed detached bungalow for sale	Bedrooms4Bathrooms1Living rooms2	Nash Grove Lane, Finchampstead, Wokingham RG40
8	8	£625,000	4 bed detached house for sale	Bedrooms4Bathrooms2Living rooms2	Anvil Close, Spencers Wood, Reading, Berkshire...
9	9	£750,000	5 bed detached house for sale	Bedrooms5Bathrooms4Living rooms3	Cressingham Road, Reading, Berkshire RG2

View a random sample of the dataset

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

Out [ ]:

	Unnamed: 0	Listing price	Listing info	All rooms	Location
209	209	£400,000	4 bed terraced house for sale	Bedrooms4Bathrooms2Living rooms2	Swainstone Road, Reading RG2
771	771	£225,000	1 bed maisonette for sale	Bedrooms1Bathrooms1Living rooms1	Maiden Place, Reading, Berkshire RG6
273	273	£335,000	2 bed terraced house for sale	Bedrooms2Bathrooms1Living rooms1	Mannock Way, Woodley, Reading RG5
614	614	£500,000	3 bed semi-detached house for sale	Bedrooms3Bathrooms1Living rooms1	Nightingale Road, Woodley, Reading RG5
669	669	£420,000	3 bed end terrace house for sale	Bedrooms3Bathrooms1Living rooms2	Kennedy Drive, Pangbourne, Reading, Berkshire RG8
70	70	£350,000	2 bed detached bungalow for sale	Bedrooms2Bathrooms1Living rooms1	Tuxford Mews, Reading RG30
378	378	£330,000	2 bed flat for sale	Bedrooms2Bathrooms2Living rooms1	The Parma At Renaissance, 2 Bed Apartment, Por...
833	833	£350,000	3 bed terraced house for sale	Bedrooms3Bathrooms1Living rooms2	Bourton Close, Tilehurst, Reading RG30
288	288	£525,000	4 bed end terrace house for sale	Bedrooms4Bathrooms2Living rooms1	Reading RG1,
106	106	£695,000	4 bed semi-detached house for sale	Bedrooms4Bathrooms2Living rooms2	Old Bath Road, Calcot, Reading RG31

View dataset info

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 950 entries, 0 to 949
Data columns (total 5 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   Unnamed: 0     950 non-null     int64
 1   Listing price  950 non-null     object
 2   Listing info   950 non-null     object
 3   All rooms      950 non-null     object
 4   Location       950 non-null     object
dtypes: int64(1), object(4)
memory usage: 37.2+ KB
```

Data Preparation

Steps:

1. Remove Unnamed column
2. Rename columns for cleaner code
3. Remove £ symbol from listing price and convert column data type to int64
4. Split 'Listing info' to extract property type information (split between 'bed' and 'for')
5. Split 'All rooms' and create separate columns for 'bedrooms', 'bathrooms', 'living rooms'
6. Create a new column and group all property types under two categories: either 'house' or 'flat'

Drop unnamed column (duplicate of index)

```
In [ ]: df = df.drop("Unnamed: 0", axis=1)
        df.head(10)
```

Out [ ]:

	Listing price	Listing info	All rooms	Location
0	£850,000	4 bed detached house for sale	Bedrooms4Bathrooms3Living rooms3	Bluebell Crescent, Woodley, Reading RG5
1	£290,000	2 bed flat for sale	Bedrooms2Bathrooms1Living rooms1	Kennet Walk, Reading RG1
2	£325,000	2 bed flat for sale	Bedrooms2Bathrooms1Living rooms1	46 Bath Road, Reading RG1
3	£200,000	1 bed flat for sale	Bedrooms1Bathrooms1Living rooms1	Russell Street, Reading RG1
4	£499,000	3 bed semi-detached house for sale	Bedrooms3Bathrooms2Living rooms1	Woolhampton Way, Reading RG2
5	£525,000	3 bed semi-detached house for sale	Bedrooms3Bathrooms1Living rooms2	Southcote Farm Lane, Reading RG30
6	£700,000	6 bed semi-detached house for sale	Bedrooms6Bathrooms3Living rooms1	Beech Lane, Reading RG6
7	£625,000	4 bed detached bungalow for sale	Bedrooms4Bathrooms1Living rooms2	Nash Grove Lane, Finchampstead, Wokingham RG40
8	£625,000	4 bed detached house for sale	Bedrooms4Bathrooms2Living rooms2	Anvil Close, Spencers Wood, Reading, Berkshire...
9	£750,000	5 bed detached house for sale	Bedrooms5Bathrooms4Living rooms3	Cressingham Road, Reading, Berkshire RG2

In [ ]:
```python
# df["All rooms"].value_counts()
```

Rename and standardise column names

In [ ]:
```python
df1 = df.rename({"Listing price": "list_price", "Listing info":"list_info
df1.head()
```

Out [ ]:

	list_price	list_info	rooms	location
0	£850,000	4 bed detached house for sale	Bedrooms4Bathrooms3Living rooms3	Bluebell Crescent, Woodley, Reading RG5
1	£290,000	2 bed flat for sale	Bedrooms2Bathrooms1Living rooms1	Kennet Walk, Reading RG1
2	£325,000	2 bed flat for sale	Bedrooms2Bathrooms1Living rooms1	46 Bath Road, Reading RG1
3	£200,000	1 bed flat for sale	Bedrooms1Bathrooms1Living rooms1	Russell Street, Reading RG1
4	£499,000	3 bed semi-detached house for sale	Bedrooms3Bathrooms2Living rooms1	Woolhampton Way, Reading RG2

Remove '£' sympbol and ',' from 'list_price'

```
In [ ]:  df1.list_price = df1.list_price.str.strip("£").str.replace(",", "")
         df1.head()
```

Out[ ]:

	list_price	list_info	rooms	location
0	850000	4 bed detached house for sale	Bedrooms4Bathrooms3Living rooms3	Bluebell Crescent, Woodley, Reading RG5
1	290000	2 bed flat for sale	Bedrooms2Bathrooms1Living rooms1	Kennet Walk, Reading RG1
2	325000	2 bed flat for sale	Bedrooms2Bathrooms1Living rooms1	46 Bath Road, Reading RG1
3	200000	1 bed flat for sale	Bedrooms1Bathrooms1Living rooms1	Russell Street, Reading RG1
4	499000	3 bed semi-detached house for sale	Bedrooms3Bathrooms2Living rooms1	Woolhampton Way, Reading RG2

Check the data type of each column/variable

```
In [ ]:  df1.dtypes
```

```
Out[ ]:  list_price    object
         list_info     object
         rooms         object
         location      object
         dtype: object
```

'List_price' column contains string elements, hence showing as 'object' datatype. Column contains instances of "POA" (price on asking).

Return a new dataframe without rows with "POA".

```
In [ ]:  df1 = df1[df1.list_price != "POA"]
```

Convert the data type of'list_price' column. Change from 'object' to 'integer'.

```
In [ ]:  df1.list_price = df1.list_price.astype("int64")
```

```
In [ ]:  df1.dtypes
```

```
Out[ ]:  list_price     int64
         list_info     object
         rooms         object
         location      object
         dtype: object
```

Split 'list_info' values to extract property type

```
In [ ]:    # View a sample

           df1.list_info.sample(10)
```

```
Out[ ]:    528      3 bed semi-detached house for sale
           877          2 bed terraced house for sale
           542          4 bed detached house for sale
           578      3 bed end terrace house for sale
           256          5 bed detached house for sale
           520                    1 bed flat for sale
           437      3 bed semi-detached house for sale
           116      3 bed semi-detached house for sale
           310          4 bed detached house for sale
           289          2 bed terraced house for sale
           Name: list_info, dtype: object
```

Perform string manipulation to slice off the beginning (x bed) and end (for sale) to extract property type

```
In [ ]:    df1.list_info = df1.list_info.str[6:-8]
```

Check if slicing worked

```
In [ ]:    df1.head()
```

Out[ ]:

	list_price	list_info	rooms	location
0	850000	detached house	Bedrooms4Bathrooms3Living rooms3	Bluebell Crescent, Woodley, Reading RG5
1	290000	flat	Bedrooms2Bathrooms1Living rooms1	Kennet Walk, Reading RG1
2	325000	flat	Bedrooms2Bathrooms1Living rooms1	46 Bath Road, Reading RG1
3	200000	flat	Bedrooms1Bathrooms1Living rooms1	Russell Street, Reading RG1
4	499000	semi-detached house	Bedrooms3Bathrooms2Living rooms1	Woolhampton Way, Reading RG2

Check 'rooms' column by viewing a random sample

```
In [ ]:    df1.rooms.sample(10)
```

```
Out[ ]:   583      Bedrooms2Bathrooms1Living rooms2
          247      Bedrooms2Bathrooms1Living rooms1
          926               Bathrooms1Living rooms1
          432      Bedrooms4Bathrooms2Living rooms3
          76       Bedrooms2Bathrooms1Living rooms1
          103      Bedrooms3Bathrooms2Living rooms1
          212      Bedrooms4Bathrooms4Living rooms4
          66       Bedrooms3Bathrooms1Living rooms2
          783      Bedrooms3Bathrooms2Living rooms2
          796      Bedrooms2Bathrooms1Living rooms1
          Name: rooms, dtype: object
```

Split 'rooms' column in order to extract number of bedrooms, bathrooms and living rooms, and to create separate columns for each

```python
In [ ]:   import re
```

Create a function to perform split

```python
In [ ]:   def split_rooms(value):
              splitted = re.split('(\d+)', value)
              return splitted
```

Check if function works as intended by applying it to first value

```python
In [ ]:   split_rooms(df1.rooms[0])
```

```
Out[ ]:   ['Bedrooms', '4', 'Bathrooms', '3', 'Living rooms', '3', '']
```

Apply function to entire column using lambda function

```python
In [ ]:   df1.rooms = df1.rooms.apply(lambda x: split_rooms(x))
```

Check if string values have been split

```python
In [ ]:   df1.rooms.head()
```

```
Out[ ]:   0    [Bedrooms, 4, Bathrooms, 3, Living rooms, 3, ]
          1    [Bedrooms, 2, Bathrooms, 1, Living rooms, 1, ]
          2    [Bedrooms, 2, Bathrooms, 1, Living rooms, 1, ]
          3    [Bedrooms, 1, Bathrooms, 1, Living rooms, 1, ]
          4    [Bedrooms, 3, Bathrooms, 2, Living rooms, 1, ]
          Name: rooms, dtype: object
```

```python
In [ ]:   df1.rooms.sample(20)
```

```
Out[ ]:   559     [Bedrooms, 1, Bathrooms, 1, Living rooms, 1, ]
          775     [Bedrooms, 3, Bathrooms, 1, Living rooms, 1, ]
          47      [Bedrooms, 2, Bathrooms, 1, Living rooms, 1, ]
          567     [Bedrooms, 2, Bathrooms, 1, Living rooms, 2, ]
          891     [Bedrooms, 4, Bathrooms, 2, Living rooms, 2, ]
          888     [Bedrooms, 5, Bathrooms, 3, Living rooms, 2, ]
          538     [Bedrooms, 3, Bathrooms, 2, Living rooms, 1, ]
          663     [Bedrooms, 3, Bathrooms, 2, Living rooms, 1, ]
          739     [Bedrooms, 4, Bathrooms, 2, Living rooms, 2, ]
          285     [Bedrooms, 3, Bathrooms, 1, Living rooms, 1, ]
          589     [Bedrooms, 6, Bathrooms, 4, Living rooms, 3, ]
          95      [Bedrooms, 3, Bathrooms, 1, Living rooms, 1, ]
          619     [Bedrooms, 4, Bathrooms, 2, Living rooms, 2, ]
          944     [Bedrooms, 1, Bathrooms, 1, Living rooms, 1, ]
          506     [Bedrooms, 1, Bathrooms, 1, Living rooms, 1, ]
          163     [Bedrooms, 3, Bathrooms, 1, Living rooms, 1, ]
          519                             [Bedrooms, 1, ]
          772     [Bedrooms, 3, Bathrooms, 1, Living rooms, 1, ]
          40      [Bedrooms, 7, Bathrooms, 4, Living rooms, 1, ]
          171     [Bedrooms, 1, Bathrooms, 1, Living rooms, 1, ]
          Name: rooms, dtype: object
```

Write a function to return the number of bedrooms if 'bedroom' is listed under 'rooms'

```python
In [ ]: def bedrooms_num(list):
            if "Bedrooms" in list:
                bedroom_index = list.index("Bedrooms")
                bed_number_index = bedroom_index+1
                return list[bed_number_index]
            else:
                return np.nan
```

Check if function works

```python
In [ ]: bedrooms_num(df1.rooms[0])
```

```
Out[ ]: '4'
```

```python
In [ ]: bedrooms_num(df1.rooms[27])
```

```
Out[ ]: '2'
```

Create same function to return bathroom and living room numbers if they are listed under 'rooms'

```
In [ ]: def bathrooms_num(list):
            if "Bathrooms" in list:
                bathroom_index = list.index("Bathrooms")
                bathroom_number_index = bathroom_index+1
                return list[bathroom_number_index]
            else:
                return np.nan
```

```
In [ ]: def livingroom_num(list):
            if "Living rooms" in list:
                livingroom_index = list.index("Living rooms")
                livingroom_number_index = livingroom_index+1
                return list[livingroom_number_index]
            else:
                return np.nan
```

Create 3 new columns to respectively show the number of bedrooms, bathrooms, and living rooms by applying the above functions to the 'rooms' column

```
In [ ]: df1["bedrooms"] = df1.rooms.apply(lambda x: bedrooms_num(x))
        df1["bathrooms"] = df1.rooms.apply(lambda x: bathrooms_num(x))
        df1["living_rooms"] = df1.rooms.apply(lambda x: livingroom_num(x))
```

Check that the new columns have been added to the dataframe

```
In [ ]: df1.head()
```

Out[ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
0	850000	detached house	[Bedrooms, 4, Bathrooms, 3, Living rooms, 3,]	Bluebell Crescent, Woodley, Reading RG5	4	3	3
1	290000	flat	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Kennet Walk, Reading RG1	2	1	1
2	325000	flat	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	46 Bath Road, Reading RG1	2	1	1
3	200000	flat	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Russell Street, Reading RG1	1	1	1
4	499000	semi-detached house	[Bedrooms, 3, Bathrooms, 2, Living rooms, 1,]	Woolhampton Way, Reading RG2	3	2	1

In [ ]: `df1.sample(20)`

Out[ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
928	650000	detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 3,]	The Pightle, Grazeley Green, Reading, Berkshir...	4	2	3
782	550000	terraced house	[Bedrooms, 5, Bathrooms, 3, Living rooms, 1,]	Kings Road, Reading RG1	5	3	1
736	270000	flat	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Meadow Way, Caversham RG4	2	1	1
796	260000	flat	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	119 Orts Road, Reading RG1	2	1	1
			[Bedrooms,	Burghfield			

544	200000	flat	2, Bathrooms, 1, Living rooms, 1,]	Road, Reading, Berkshire RG30	2	1	1
429	450000	semi-detached house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Ashridge Road, Wokingham RG40	3	1	1
10	425000	semi-detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 2,]	Basingstoke Road, Reading, Berkshire RG2	4	2	2
39	550000	terraced house	[Bedrooms, 6, Bathrooms, 2, Living rooms, 1,]	Addington Road, Reading RG1	6	2	1
656	1100000	detached house	[Bedrooms, 4, Bathrooms, 1, Living rooms, 1,]	Henley-On-Thames, South Oxfordshire RG9	4	1	1
705	450000	semi-detached house	[Bedrooms, 3, Bathrooms, 2, Living rooms, 1,]	Coley RG1,	3	2	1
360	325000	terraced house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 2,]	Belmont Road, Reading RG30	3	1	2
130	800000	detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 2,]	Wokingham, Berkshire RG41	4	2	2
523	590000	semi-detached house	[Bedrooms, 3, Bathrooms, 2, Living rooms, 1,]	Nicholson Drive Wokingham, Berkshire RG41	3	2	1
158	180000	mobile/park home	[Bedrooms, 3, Bathrooms, 1, Living rooms, 2,]	Mereoak Park, Three Mile Cross RG7	3	1	2
788	220000	flat	[Bedrooms, 1, Bathrooms, 1, Living	Addington Road, Reading, Berkshire	1	1	1

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
			rooms, 1,]	RG1			
824	350000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Milsom Close, Shinfield, Reading, Berkshire RG2	2	1	1
153	285000	flat	[Bedrooms, 2, Bathrooms, 2, Living rooms, 1,]	Friary Court, Tudor Road, Reading, Berkshire RG1	2	2	1
205	290000	terraced house	[Bedrooms, 2, Bathrooms, 1,]	Bromley Walk, Tilehurst, Reading RG30	2	1	NaN
177	350000	semi-detached house	[Bedrooms, 2, Bathrooms, 2, Living rooms, 1,]	Ashmore Road, Reading, Reading RG2	2	2	1
225	375000	end terrace house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Appleby End, Reading RG30	3	1	1

Missing Data

In [ ]:
```python
print(df1.isnull().sum())
```

```
list_price        0
list_info         0
rooms             0
location          0
bedrooms         14
bathrooms        54
living_rooms     62
dtype: int64
```

In [ ]:
```python
# There are missing values in list_info but it's not being picked up when
# Checking the original csv file, these were originally listed as 'studio

df2 = df1.sort_values(by=["list_info", "bedrooms"])
df2.head(20)
```

Out [ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
			[Bedrooms,	Flat 3, 34 Eastern Avenue,			

484	75000	1,]	Reading, Berkshire RG1	1	NaN	NaN
900	130000	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Watersfield Close, Lower Earley, Reading RG6	1	1	1
124	150000	[Bathrooms, 1, Living rooms, 1,]	Colleton Drive, Twyford, Berkshire RG10	NaN	1	1
195	140000	[Bathrooms, 1, Living rooms, 1,]	George Street, Reading RG1	NaN	1	1
290	175000	[Living rooms, 1,]	Central Reading RG1,	NaN	NaN	1
333	475000	[Bathrooms, 1, Living rooms, 1,]	Hobbs End, Henley-On-Thames RG9	NaN	1	1
356	160000	[Bathrooms, 1, Living rooms, 1,]	Beech Lane, Lower Earley, Reading RG6	NaN	1	1
393	135000	[Bathrooms, 1, Living rooms, 1,]	Tippett Rise, Reading RG2	NaN	1	1
436	130000	[Bathrooms, 1, Living rooms, 1,]	Castle Hill, Reading, Berkshire RG1	NaN	1	1
495	160000	[Bathrooms, 1,]	Tippett Rise, Reading, Berkshire RG2	NaN	1	NaN
513	140000	[Bathrooms, 1,]	Castle Hill, Reading, Berkshire RG1	NaN	1	NaN
		[Bathrooms,	London Road,			

721	125000		1, Living rooms, 1,]	Reading, Berkshire RG1	NaN	1	1
722	177500		[Bathrooms, 1, Living rooms, 1,]	Flat 23 Cadogan House, Rose Kiln Lane, Reading...	NaN	1	1
841	175000		[Bathrooms, 1, Living rooms, 1,]	38-50 Kings Road, Reading, Berkshire RG1	NaN	1	1
926	195000		[Bathrooms, 1, Living rooms, 1,]	Bourne Close, Calcot, Reading RG31	NaN	1	1
133	1000000	semi-detached house	[Bedrooms, 12, Bathrooms, 4,]	Reading RG1,	12	4	NaN
26	399000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 2,]	Bridport Close, Lower Earley, Reading RG6	2	1	2
47	425000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Faygate Way, Lower Earley, Reading RG6	2	1	1
217	330000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Rushmoor Gardens, Calcot, Reading RG31	2	1	1
236	450000	bungalow	[Bedrooms, 2, Bathrooms, 2, Living rooms, 2,]	Birch Lane, Mortimer Common, Reading RG7	2	2	2

Check what is in the empty 'list_info' cells

```
In [ ]:  df2.list_info[484]

Out[ ]:  ' '
```

```
In [ ]:  df2.list_info[124]
```

```
Out[ ]:  ' '
```

```
In [ ]:  # A space (' ') remains in some instances following the string manipulati
         df2.list_info = df2.list_info.apply(lambda x: x.replace(" ", "studio") if
```

```
In [ ]:  df2.head(20)
```

Out [ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
484	75000	studio	[Bedrooms, 1,]	Flat 3, 34 Eastern Avenue, Reading, Berkshire RG1	1	NaN	NaN
900	130000	studio	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Watersfield Close, Lower Earley, Reading RG6	1	1	1
124	150000	studio	[Bathrooms, 1, Living rooms, 1,]	Colleton Drive, Twyford, Berkshire RG10	NaN	1	1
195	140000	studio	[Bathrooms, 1, Living rooms, 1,]	George Street, Reading RG1	NaN	1	1
290	175000	studio	[Living rooms, 1,]	Central Reading RG1,	NaN	NaN	1
333	475000	studio	[Bathrooms, 1, Living rooms, 1,]	Hobbs End, Henley-On-Thames RG9	NaN	1	1
356	160000	studio	[Bathrooms, 1, Living rooms, 1,]	Beech Lane, Lower Earley, Reading RG6	NaN	1	1
393	135000	studio	[Bathrooms, 1, Living rooms, 1,]	Tippett Rise, Reading RG2	NaN	1	1

436	130000	studio	[Bathrooms, 1, Living rooms, 1,]	Castle Hill, Reading, Berkshire RG1	NaN	1	1
495	160000	studio	[Bathrooms, 1,]	Tippett Rise, Reading, Berkshire RG2	NaN	1	NaN
513	140000	studio	[Bathrooms, 1,]	Castle Hill, Reading, Berkshire RG1	NaN	1	NaN
721	125000	studio	[Bathrooms, 1, Living rooms, 1,]	London Road, Reading, Berkshire RG1	NaN	1	1
722	177500	studio	[Bathrooms, 1, Living rooms, 1,]	Flat 23 Cadogan House, Rose Kiln Lane, Reading...	NaN	1	1
841	175000	studio	[Bathrooms, 1, Living rooms, 1,]	38-50 Kings Road, Reading, Berkshire RG1	NaN	1	1
926	195000	studio	[Bathrooms, 1, Living rooms, 1,]	Bourne Close, Calcot, Reading RG31	NaN	1	1
133	1000000	semi-detached house	[Bedrooms, 12, Bathrooms, 4,]	Reading RG1,	12	4	NaN
26	399000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 2,]	Bridport Close, Lower Earley, Reading RG6	2	1	2
47	425000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Faygate Way, Lower Earley, Reading RG6	2	1	1
			[Bedrooms, 2, Bathrooms,	Rushmoor Gardens, Calcot,			

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
217	330000	bungalow	1, Living rooms, 1,]	Reading RG31	2	1	1
236	450000	bungalow	[Bedrooms, 2, Bathrooms, 2, Living rooms, 2,]	Birch Lane, Mortimer Common, Reading RG7	2	2	2

Reset index

```
In [ ]:  df2 = df2.reset_index(drop=True)
         df2.head(20)
```

Out[ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
0	75000	studio	[Bedrooms, 1,]	Flat 3, 34 Eastern Avenue, Reading, Berkshire RG1	1	NaN	NaN
1	130000	studio	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Watersfield Close, Lower Earley, Reading RG6	1	1	1
2	150000	studio	[Bathrooms, 1, Living rooms, 1,]	Colleton Drive, Twyford, Berkshire RG10	NaN	1	1
3	140000	studio	[Bathrooms, 1, Living rooms, 1,]	George Street, Reading RG1	NaN	1	1
4	175000	studio	[Living rooms, 1,]	Central Reading RG1,	NaN	NaN	1
5	475000	studio	[Bathrooms, 1, Living rooms, 1,]	Hobbs End, Henley-On-Thames RG9	NaN	1	1
6	160000	studio	[Bathrooms, 1, Living rooms, 1,]	Beech Lane, Lower Earley, Reading RG6	NaN	1	1

7	135000	studio	[Bathrooms, 1, Living rooms, 1,]	Tippett Rise, Reading RG2	NaN	1	1
8	130000	studio	[Bathrooms, 1, Living rooms, 1,]	Castle Hill, Reading, Berkshire RG1	NaN	1	1
9	160000	studio	[Bathrooms, 1,]	Tippett Rise, Reading, Berkshire RG2	NaN	1	NaN
10	140000	studio	[Bathrooms, 1,]	Castle Hill, Reading, Berkshire RG1	NaN	1	NaN
11	125000	studio	[Bathrooms, 1, Living rooms, 1,]	London Road, Reading, Berkshire RG1	NaN	1	1
12	177500	studio	[Bathrooms, 1, Living rooms, 1,]	Flat 23 Cadogan House, Rose Kiln Lane, Reading...	NaN	1	1
13	175000	studio	[Bathrooms, 1, Living rooms, 1,]	38-50 Kings Road, Reading, Berkshire RG1	NaN	1	1
14	195000	studio	[Bathrooms, 1, Living rooms, 1,]	Bourne Close, Calcot, Reading RG31	NaN	1	1
15	1000000	semi-detached house	[Bedrooms, 12, Bathrooms, 4,]	Reading RG1,	12	4	NaN
16	399000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 2,]	Bridport Close, Lower Earley, Reading RG6	2	1	2
17	425000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living	Faygate Way, Lower Earley, Reading	2	1	1

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_roo
18	330000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Rushmoor Gardens, Calcot, Reading RG31	2	1	1
19	450000	bungalow	[Bedrooms, 2, Bathrooms, 2, Living rooms, 2,]	Birch Lane, Mortimer Common, Reading RG7	2	2	2

Investigate rows with missing values

```python
df_missing = df2[df2.isnull().any(axis=1)]
```

```python
# Set option to display all rows
pd.set_option("display.max_rows", None)

df_missing
```

Out[ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_roo
0	75000	studio	[Bedrooms, 1,]	Flat 3, 34 Eastern Avenue, Reading, Berkshire RG1	1	NaN	N
2	150000	studio	[Bathrooms, 1, Living rooms, 1,]	Colleton Drive, Twyford, Berkshire RG10	NaN	1	
3	140000	studio	[Bathrooms, 1, Living rooms, 1,]	George Street, Reading RG1	NaN	1	
4	175000	studio	[Living rooms, 1,]	Central Reading RG1,	NaN	NaN	
5	475000	studio	[Bathrooms, 1, Living rooms, 1,]	Hobbs End, Henley-On-Thames RG9	NaN	1	
6	160000	studio	[Bathrooms, 1, Living rooms, 1,]	Beech Lane, Lower Earley, Reading RG6	NaN	1	
7	135000	studio	[Bathrooms, 1, Living rooms, 1,]	Tippett Rise, Reading RG2	NaN	1	
8	130000	studio	[Bathrooms, 1, Living rooms, 1,]	Castle Hill, Reading, Berkshire RG1	NaN	1	

9	160000	studio	[Bathrooms, 1,]	Tippett Rise, Reading, Berkshire RG2	NaN	1	N
10	140000	studio	[Bathrooms, 1,]	Castle Hill, Reading, Berkshire RG1	NaN	1	N
11	125000	studio	[Bathrooms, 1, Living rooms, 1,]	London Road, Reading, Berkshire RG1	NaN	1	
12	177500	studio	[Bathrooms, 1, Living rooms, 1,]	Flat 23 Cadogan House, Rose Kiln Lane, Reading...	NaN	1	
13	175000	studio	[Bathrooms, 1, Living rooms, 1,]	38-50 Kings Road, Reading, Berkshire RG1	NaN	1	
14	195000	studio	[Bathrooms, 1, Living rooms, 1,]	Bourne Close, Calcot, Reading RG31	NaN	1	
15	1000000	semi-detached house	[Bedrooms, 12, Bathrooms, 4,]	Reading RG1,	12	4	N
40	300000	detached bungalow	[Bedrooms, 2, Bathrooms, 1,]	Thirlmere Avenue, Tilehurst, Reading RG30	2	1	N
62	550000	detached house	[Bedrooms, 3, Bathrooms, 1,]	Basingstoke Road, Reading RG2	3	1	N
70	525000	detached house	[Bedrooms, 3,]	Blanchard Close, Woodley, Reading RG5	3	NaN	N
72	525000	detached house	[Bedrooms, 3, Bathrooms, 1,]	Chaffinch Close, Wokingham, Berkshire RG41	3	1	N
87	555000	detached house	[Bedrooms, 3,]	"Drayton ? Detached" at Sheerlands Road, Wokin...	3	NaN	N
95	550000	detached house	[Bedrooms, 4, Living rooms, 2,]	Wokingham, Berkshire RG41	4	NaN	
		detached	[Bedrooms,	Shiplake Bottom,			

124	850000	house	4,]	Peppard Common, Henley-On-Tha...	4	NaN	N
126	640000	detached house	[Bedrooms, 4, Living rooms, 2,]	Bearwood Road, Wokingham, Berkshire RG41	4	NaN	
154	650000	detached house	[Bedrooms, 4, Bathrooms, 2,]	Calcot Park, Calcot, Reading RG31	4	2	N
168	675000	detached house	[Bedrooms, 4,]	Pinecroft Road, Wokingham RG41	4	NaN	N
176	675000	detached house	[Bedrooms, 4,]	"Warrington" at Sheerlands Road, Wokingham RG4...	4	NaN	N
193	875000	detached house	[Bedrooms, 4,]	Bayliss Road, Wargrave, Reading, Berkshire RG10	4	NaN	N
207	795000	detached house	[Bedrooms, 4,]	Hermitage Drive, Twyford, Berkshire RG10	4	NaN	N
216	600000	detached house	[Bedrooms, 4, Living rooms, 2,]	Wokingham RG40,	4	NaN	
262	750000	detached house	[Bedrooms, 5, Bathrooms, 3,]	Fairway Avenue, Tilehurst, Reading RG30	5	3	N
270	850000	detached house	[Bedrooms, 5,]	"Lancaster Detached" at Sheerlands Road, Wokin...	5	NaN	N
276	900000	detached house	[Bedrooms, 5, Bathrooms, 2,]	Denmark Avenue, Woodley, Reading RG5	5	2	N
278	1800000	detached house	[Bedrooms, 5,]	Mustard Lane, Sonning, Reading RG4	5	NaN	N
		detached	[Bedrooms, 5,	Chestnut Avenue,			

283	1850000	house	Bathrooms, 2,]	Wokingham RG41	5	2	N
292	1795000	detached house	[Bedrooms, 6, Bathrooms, 7,]	Broadlands Close, Calcot, Reading, Berkshire RG31	6	7	N
298	600000	detached house	[Bedrooms, 8,]	High Street, Theale, Reading RG7	8	NaN	N
299	2950000	detached house	[Bedrooms, 8,]	Loddon Drive, Wargrave, Reading RG10	8	NaN	N
300	1050000	detached house	[Bedrooms, 8,]	Reading, Berkshire RG2	8	NaN	N
302	2750000	detached house	[Bedrooms, 9, Bathrooms, 7,]	Church Lane, Grazeley, Reading RG7	9	7	N
307	425000	end terrace house	[Bedrooms, 2,]	Havelock Road, Wokingham, Berkshire RG41	2	NaN	N
328	635000	end terrace house	[Bedrooms, 3, Bathrooms, 2,]	Bearwood Road, Sindlesham, Wokingham, Berkshir...	3	2	N
330	450000	end terrace house	[Bedrooms, 3, Living rooms, 1,]	Reading RG4,	3	NaN	
353	495000	etached house	[Bathrooms, 1, Living rooms, 1,]	Wantage Road, Reading RG30	NaN	1	
357	240000	flat	[Bedrooms, 1, Living rooms, 1,]	Reading, Berkshire RG1	1	NaN	
359	190000	flat	[Bedrooms, 1, Living rooms, 1,]	Blagdon Road, Reading, Berkshire RG2	1	NaN	
382	270000	flat	[Bedrooms, 1, Bathrooms, 1,]	Eastport Apartment, Sunapee Road, Green Park V...	1	1	N
386	249950	flat	[Bedrooms, 1, Bathrooms, 1,]	Lockhart Drive, Wokingham RG40	1	1	N

390	275000	flat	[Bedrooms, 1, Living rooms, 1,]	Reading RG2,	1	NaN	
399	255000	flat	[Bedrooms, 1,]	"Darlington House, Second Floor" at Sheerlands...	1	NaN	N
400	252500	flat	[Bedrooms, 1,]	"Darlington House, First Floor" at Sheerlands ...	1	NaN	N
401	250000	flat	[Bedrooms, 1,]	"Darlington House, Ground Floor" at Sheerlands...	1	NaN	N
402	170000	flat	[Bedrooms, 1, Bathrooms, 1,]	Basingstoke Road, Riseley, Reading RG7	1	1	N
408	210000	flat	[Bedrooms, 1, Living rooms, 1,]	Reading RG1,	1	NaN	
409	210000	flat	[Bedrooms, 1, Living rooms, 1,]	Reading, Berkshire RG1	1	NaN	
410	190000	flat	[Bedrooms, 1, Living rooms, 1,]	Catherine Street, Reading, Berkshire RG30	1	NaN	
412	210000	flat	[Bedrooms, 1, Living rooms, 1,]	Caversham RG4,	1	NaN	
420	225000	flat	[Bedrooms, 1, Living rooms, 1,]	Rose Street, Wokingham, Berkshire RG40	1	NaN	
421	272995	flat	[Bedrooms, 1,]	Warren House Road, Wokingham, Berkshire RG40	1	NaN	N
422	272995	flat	[Bedrooms, 1, Bathrooms, 1,]	Warren House Road, Wokingham, Berkshire RG40	1	1	N
431	190000	flat	[Bedrooms, 1, Living rooms, 1,]	Reading, Berkshire RG1	1	NaN	

435	180000	flat	[Bedrooms, 1, Living rooms, 1,]	Reading RG1,	1	NaN
462	260000	flat	[Bedrooms, 2, Living rooms, 1,]	Reading, Berkshire RG1	2	NaN
471	400000	flat	[Bedrooms, 2, Bathrooms, 2,]	5-9 Berkeley Avenue, Reading, Berkshire RG1	2	2	N
503	314950	flat	[Bedrooms, 2,]	"Iver" at Sheerlands Road, Finchampstead RG40 ...	2	NaN	N
515	495000	flat	[Bedrooms, 2, Bathrooms, 2,]	Huntley Wharf, Palmer Street, Reading RG1	2	2	N
518	230000	flat	[Bedrooms, 2, Bathrooms, 1,]	Petworth Court, Bath Road, Reading RG1	2	1	N
522	240000	flat	[Bedrooms, 2, Bathrooms, 1,]	Westcote Road, Reading RG30	2	1	N
535	230000	flat	[Bedrooms, 2, Bathrooms, 1,]	Westcote Road, Reading RG30	2	1	N
538	310000	flat	[Bedrooms, 2,]	Chatham Place, Reading, Berkshire RG1	2	NaN	N
557	329950	flat	[Bedrooms, 2,]	"Iver" at Sheerlands Road, Finchampstead RG40 ...	2	NaN	N
564	349995	flat	[Bedrooms, 2, Bathrooms, 2,]	Warren House Road, Wokingham, Berkshire RG40	2	2	N
565	350995	flat	[Bedrooms, 2, Bathrooms, 2,]	Warren House Road, Wokingham, Berkshire RG40	2	2	N
			[Bedrooms,	Marlow Court,		

586	280000	flat	2, Bathrooms, 2,]	All Hallows Road, Caversham RG4	2	2	N
591	210000	flat	[Bedrooms, 2,]	Reading, Berkshire RG30	2	NaN	N
621	74950	lodge	[Bedrooms, 2, Bathrooms, 1,]	UK Lodges, UK RG9	2	1	N
627	200000	maisonette	[Bedrooms, 1, Living rooms, 1,]	Aquila Close, Wokingham, Berkshire RG41	1	NaN	
694	370000	semi-detached house	[Bedrooms, 3,]	Primrose Close, Purley On Thames, Reading RG8	3	NaN	N
734	590000	semi-detached house	[Bedrooms, 3, Living rooms, 2,]	Caversham RG4,	3	NaN	
744	550000	semi-detached house	[Bedrooms, 3, Bathrooms, 1,]	Amberley Drive, Twyford, Reading, Berkshire RG10	3	1	N
745	500000	semi-detached house	[Bedrooms, 3, Bathrooms, 2,]	Mylne Square, Wokingham, Berkshire RG40	3	2	N
754	519950	semi-detached house	[Bedrooms, 3,]	"Drayton Semi Detached" at Sheerlands Road, Wo...	3	NaN	N
755	519950	semi-detached house	[Bedrooms, 3,]	"Drayton Semi Detached" at Sheerlands Road, Wo...	3	NaN	N
780	389500	semi-detached house	[Bedrooms, 3, Living rooms, 1,]	Whitley, Reading RG2	3	NaN	
799	425000	semi-detached house	[Bedrooms, 4, Bathrooms, 1,]	Pierces Hill Tilehurst, Reading RG31	4	1	N
814	600000	semi-detached house	[Bedrooms, 4,]	Dalley Road, Wokingham, Berkshire RG40	4	NaN	N

845	290000	terraced house	[Bedrooms, 2, Bathrooms, 1,]	Bromley Walk, Tilehurst, Reading RG30	2	1	N
857	300000	terraced house	[Bedrooms, 2,]	St. Johns Street, Reading, Berkshire RG1	2	NaN	N
876	400000	terraced house	[Bedrooms, 3, Living rooms, 1,]	Reading, Berkshire RG30	3	NaN	
877	500000	terraced house	[Bedrooms, 3, Living rooms, 1,]	Woodley, Berkshire RG5	3	NaN	
884	499950	terraced house	[Bedrooms, 3,]	"Ashford ? Terraced" at Sheerlands Road, Finch...	3	NaN	N
886	450000	terraced house	[Bedrooms, 3, Bathrooms, 1,]	The Brookmill, Reading RG1	3	1	N
906	300000	terraced house	[Bedrooms, 3, Living rooms, 2,]	Reading RG1,	3	NaN	
907	305000	terraced house	[Bedrooms, 3,]	Kings Road, Caversham, Reading RG4	3	NaN	N
915	425000	terraced house	[Bedrooms, 4, Living rooms, 2,]	Reading, Southcote RG30	4	NaN	
916	375000	terraced house	[Bedrooms, 4, Living rooms, 2,]	Reading RG1,	4	NaN	
931	495000	terraced house	[Bedrooms, 6, Bathrooms, 2,]	Wantage Road, Reading RG30	6	2	N
933	500000	terraced house	[Bedrooms, 9, Bathrooms, 2,]	Basingstoke Road, Reading RG2	9	2	N
945	425000	town house	[Bedrooms, 4, Living rooms, 1,]	Reading, Berkshire RG30	4	NaN	

Check the total number of rows with missing values

```
In [ ]:  len(df_missing)
```

Out[ ]:   98

Most 'studios' have NaN under 'bedrooms'. As they are studio flats, we do not expect them to have bedrooms, so we can replace NaN with 0 to show 0 bedrooms.

In [ ]:   ```
df2.bedrooms = df2.bedrooms.replace(np.nan, 0)
```

In [ ]:    ```
df2.head(20)
```

Out[ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
0	75000	studio	[Bedrooms, 1,]	Flat 3, 34 Eastern Avenue, Reading, Berkshire RG1	1	NaN	NaN
1	130000	studio	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Watersfield Close, Lower Earley, Reading RG6	1	1	1
2	150000	studio	[Bathrooms, 1, Living rooms, 1,]	Colleton Drive, Twyford, Berkshire RG10	0	1	1
3	140000	studio	[Bathrooms, 1, Living rooms, 1,]	George Street, Reading RG1	0	1	1
4	175000	studio	[Living rooms, 1,]	Central Reading RG1,	0	NaN	1
5	475000	studio	[Bathrooms, 1, Living rooms, 1,]	Hobbs End, Henley-On-Thames RG9	0	1	1
6	160000	studio	[Bathrooms, 1, Living rooms, 1,]	Beech Lane, Lower Earley, Reading RG6	0	1	1
7	135000	studio	[Bathrooms, 1, Living rooms, 1,]	Tippett Rise, Reading RG2	0	1	1
				Castle Hill,			

8	130000	studio	[Bathrooms, 1, Living rooms, 1,]	Reading, Berkshire RG1	0	1	1
9	160000	studio	[Bathrooms, 1,]	Tippett Rise, Reading, Berkshire RG2	0	1	NaN
10	140000	studio	[Bathrooms, 1,]	Castle Hill, Reading, Berkshire RG1	0	1	NaN
11	125000	studio	[Bathrooms, 1, Living rooms, 1,]	London Road, Reading, Berkshire RG1	0	1	1
12	177500	studio	[Bathrooms, 1, Living rooms, 1,]	Flat 23 Cadogan House, Rose Kiln Lane, Reading...	0	1	1
13	175000	studio	[Bathrooms, 1, Living rooms, 1,]	38-50 Kings Road, Reading, Berkshire RG1	0	1	1
14	195000	studio	[Bathrooms, 1, Living rooms, 1,]	Bourne Close, Calcot, Reading RG31	0	1	1
15	1000000	semi-detached house	[Bedrooms, 12, Bathrooms, 4,]	Reading RG1,	12	4	NaN
16	399000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 2,]	Bridport Close, Lower Earley, Reading RG6	2	1	2
17	425000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Faygate Way, Lower Earley, Reading RG6	2	1	1
18	330000	bungalow	[Bedrooms, 2, Bathrooms,	Rushmoor Gardens, Calcot,	2	1	1

			1, Living rooms, 1,]	Reading RG31			
19	450000	bungalow	[Bedrooms, 2, Bathrooms, 2, Living rooms, 2,]	Birch Lane, Mortimer Common, Reading RG7	2	2	2

Although there are missing values under 'bathrooms' and 'living rooms', these two variables will not be used in our analysis so we can leave them as they are for now.

There is still a missing value under 'bedrooms' in row 353. We will remove this instance from the dataset.

```
In [ ]:  df2.drop(353)
         df2.head()
```

Out[ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
0	75000	studio	[Bedrooms, 1,]	Flat 3, 34 Eastern Avenue, Reading, Berkshire RG1	1	NaN	NaN
1	130000	studio	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Watersfield Close, Lower Earley, Reading RG6	1	1	1
2	150000	studio	[Bathrooms, 1, Living rooms, 1,]	Colleton Drive, Twyford, Berkshire RG10	0	1	1
3	140000	studio	[Bathrooms, 1, Living rooms, 1,]	George Street, Reading RG1	0	1	1
4	175000	studio	[Living rooms, 1,]	Central Reading RG1,	0	NaN	1

```
In [ ]:  df2.reset_index(drop=True)
         df2.head()
```

Out[ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
0	75000	studio	[Bedrooms, 1,]	Flat 3, 34 Eastern Avenue, Reading, Berkshire RG1	1	NaN	NaN
1	130000	studio	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Watersfield Close, Lower Earley, Reading RG6	1	1	1
2	150000	studio	[Bathrooms, 1, Living rooms, 1,]	Colleton Drive, Twyford, Berkshire RG10	0	1	1
3	140000	studio	[Bathrooms, 1, Living rooms, 1,]	George Street, Reading RG1	0	1	1
4	175000	studio	[Living rooms, 1,]	Central Reading RG1,	0	NaN	1

Double check if there are still missing values in the 'bedrooms' column

In [ ]:
```python
print(df2.isnull().sum())
```

```
list_price       0
list_info        0
rooms            0
location         0
bedrooms         0
bathrooms       54
living_rooms    62
dtype: int64
```

Bedroom numbers range from 0 - 12, with 3-bed and 2-bed properties being the most common.

In [ ]:
```python
df2.bedrooms.value_counts()
```

```
Out[ ]:   3      260
          2      244
          4      214
          1      121
          5       70
          6       15
          0       14
          8        4
          7        3
          9        3
          12       1
          Name: bedrooms, dtype: int64
```

Exploratory Data Analysis

View the distribution of property prices

```python
In [ ]:  # Suppress scientific notation
         pd.set_option('display.float_format', '{:.2f}'.format)

         sns.histplot(x="list_price", data=df2)
         plt.show()
         plt.close()
```
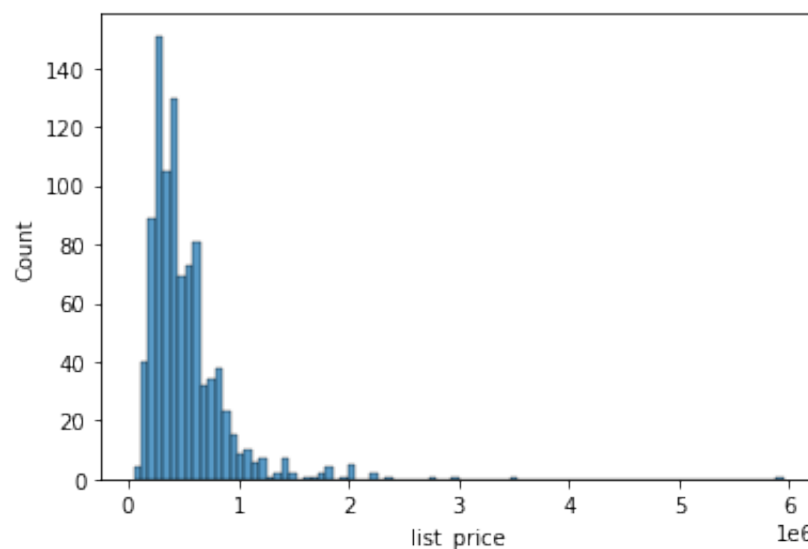

Skewed due to outliers; let's check the central tendency and max., min.

```python
In [ ]:  df2.list_price.agg(["max", "min", "mean", "median", "std"])
```

```
Out[ ]:   max        5950000.00
          min          48750.00
          mean        524398.21
          median      430000.00
          std         391340.10
          Name: list_price, dtype: float64
```

Visualise property prices in box plot format to check outliers

```
In [ ]:  sns.boxplot(y="list_price", data=df2)
         plt.show()
         plt.close()
```
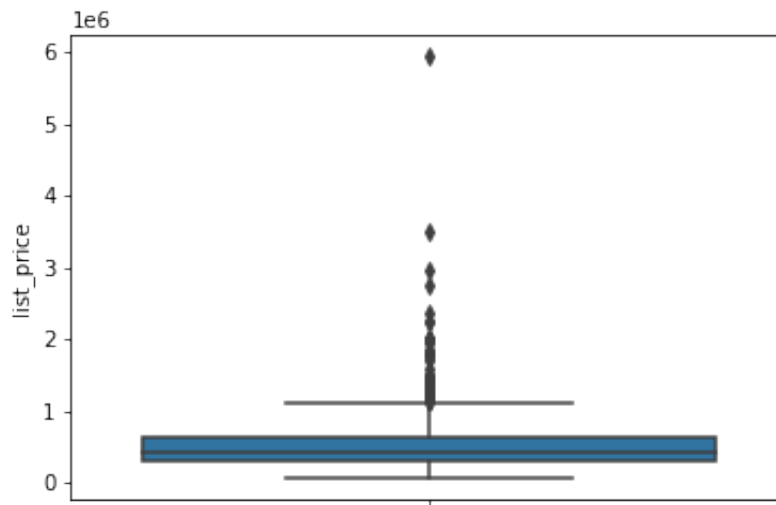


```
In [ ]:  df2.describe()
```

Out[ ]:

	list_price
count	949.00
mean	524398.21
std	391340.10
min	48750.00
25%	300000.00
50%	430000.00
75%	629950.00
max	5950000.00

```
In [ ]:  df2.dtypes
```

```
Out[ ]:  list_price       int64
         list_info        object
         rooms            object
         location         object
         bedrooms         object
         bathrooms        object
         living_rooms     object
         dtype: object
```

'Bedrooms' column is still showing as having data type 'object'; convert to interger

```
In [ ]:  df2.bedrooms = df2.bedrooms.astype("int64")
         df2.dtypes
```

```
Out[ ]:  list_price        int64
         list_info        object
         rooms            object
         location         object
         bedrooms          int64
         bathrooms        object
         living_rooms     object
         dtype: object
```

In [ ]: `df2.describe()`

Out[ ]:

	list_price	bedrooms
count	949.00	949.00
mean	524398.21	2.93
std	391340.10	1.38
min	48750.00	0.00
25%	300000.00	2.00
50%	430000.00	3.00
75%	629950.00	4.00
max	5950000.00	12.00

Create a subset of the dataset with outliers removed (properties over £1m)

In [ ]: `df_under1m = df2[df2.list_price < 1000000]`

In [ ]: `df_under1m.describe()`

Out[ ]:

	list_price	bedrooms
count	887.00	887.00
mean	451171.27	2.77
std	204216.62	1.24
min	48750.00	0.00
25%	290000.00	2.00
50%	425000.00	3.00
75%	595000.00	4.00
max	995000.00	9.00

Check median property price now that outliers (properties above £1m) have been
excluded

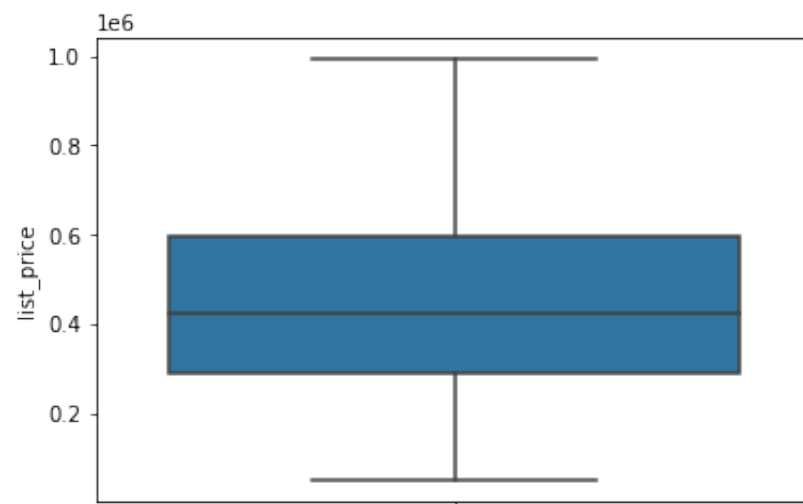In [ ]: `df_under1m.list_price.median()`

Out[ ]: 425000.0

Median house price = 425,000

Visualise distribution of property prices (under £1m)

In [ ]:
```python
sns.histplot(x="list_price", data=df_under1m, kde=True)
plt.ticklabel_format(style="plain")
plt.show()
plt.close()
```
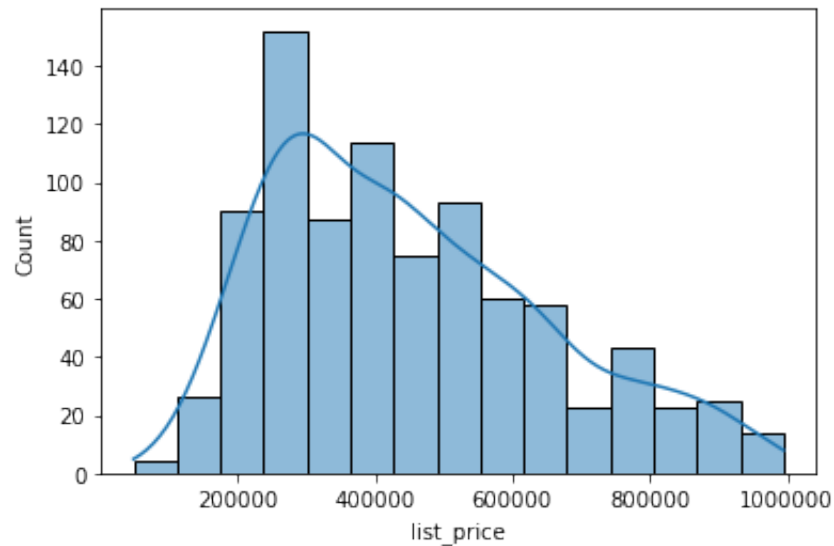

In [ ]:
```python
sns.boxplot(y="list_price", data=df_under1m)
plt.show()
plt.close()
```


Most properties are between £300,000 and £600,000.

Percentage of properties by number of bedrooms:

In [ ]:
```
percentage = pd.DataFrame(df_under1m.bedrooms.value_counts(normalize=True
percentage
```

Out [ ]:

	bedrooms
3	0.29
2	0.28
4	0.22
1	0.14
5	0.05
0	0.02
6	0.01
9	0.00
7	0.00
8	0.00

In [ ]:
```
df_under1m.bedrooms.value_counts().sum()
```

Out [ ]: 887

Visualise price vs. number of bedrooms

In [ ]:
```
sns.boxplot(data=df_under1m, x="bedrooms", y="list_price")
plt.show()
```
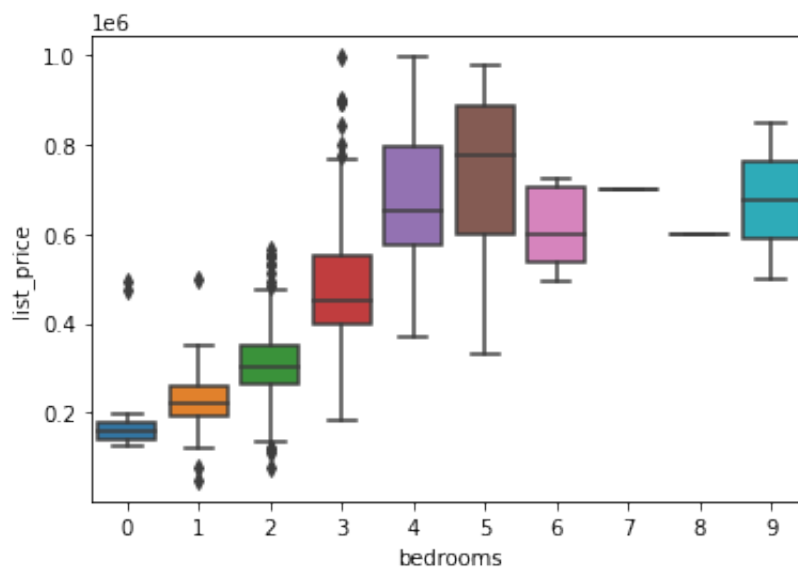
There appears to be a clear difference in list price based on the number of bedrooms. A steady upward trend can be observed for properties with 0 - 5 bedrooms, with little overlap with one another, showing a fairly uniform increase in list price as the number of bedrooms go up. From 4-bed properties onwards, there appears to be little correlation between number of bedrooms and list price, but this may be due to the low sample size for these property types and the fact that properties over £1m had been excluded. This can be confirmed by further correlation analysis if of interest.

Create a subset of the dataset which only contains properties with 1 - 5 bedrooms

```
In [ ]:  onetofivebeds = df_under1m[(df_under1m.bedrooms >= 1) & (df_under1m.bedro
         onetofivebeds.sample(20)
```

Out [ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
107	675000	detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 2,]	Nabbs Hill Close, Tilehurst, Reading, Berkshir...	4	2	2
631	295000	maisonette	[Bedrooms, 2, Bathrooms, 2, Living rooms, 1,]	Carey Road, Wokingham RG40	2	2	1
798	625000	semi-detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 2,]	Nicholson Drive, Wokingham RG41	4	2	2
344	465000	end terrace house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 3,]	Cecil Aldin Drive, Tilehurst, Reading, Berkshi...	3	1	3
747	549950	semi-detached house	[Bedrooms, 3, Bathrooms, 2, Living rooms, 3,]	Havelock Road, Wokingham RG41	3	2	3
348	500000	end terrace house	[Bedrooms, 4, Bathrooms, 1, Living rooms, 2,]	Bulmershe Road, Earley, Reading RG1	4	1	2
521	250000	flat	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Argyle Road, Reading, Berkshire RG1	2	1	1

771	420000	semi-detached house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 2,]	Grovelands Road, Reading, Berkshire RG30	3	1	2
605	650000	flat	[Bedrooms, 4, Bathrooms, 4, Living rooms, 4,]	London Road, Reading RG1	4	4	4
562	420000	flat	[Bedrooms, 2, Bathrooms, 2, Living rooms, 1,]	Danesfield, Wiltshire Road, Wokingham, Berkshi...	2	2	1
615	525000	link-detached house	[Bedrooms, 4, Bathrooms, 1, Living rooms, 2,]	Skerritt Way, Purley On Thames, Reading RG8	4	1	2
878	300000	terraced house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 2,]	Norton Road, Reading, Berkshire RG1	3	1	2
656	370000	property	[Bedrooms, 4, Bathrooms, 2, Living rooms, 1,]	Elm Park, Reading RG30	4	2	1
546	437600	flat	[Bedrooms, 2, Bathrooms, 2, Living rooms, 1,]	Highlands Lane, Rotherfield Greys, Henley-On-T...	2	2	1
106	650000	detached house	[Bedrooms, 4, Bathrooms, 1, Living rooms, 2,]	Elmley Close, Wokingham, Berkshire RG41	4	1	2
370	219000	flat	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Garrard Street, Reading RG1	1	1	1
57	412000	detached house	[Bedrooms, 2, Bathrooms, 2, Living rooms, 1,]	Maine Street, Reading, Berkshire RG2	2	2	1
			[Bedrooms, 3,	Westerham Walk,			

603	290000	flat	Bathrooms, 2, Living rooms, 2,]	Reading, Berkshire RG2	3	2	2
841	350000	terraced house	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Portway, Riseley, Reading, Berkshire RG7	2	1	1
122	675000	detached house	[Bedrooms, 4, Bathrooms, 3, Living rooms, 3,]	Mint Close, Lower Earley, Reading RG6	4	3	3

In [ ]:
```python
onetofivebeds.bedrooms.describe()
```

Out[ ]:
```
count    861.00
mean       2.76
std        1.10
min        1.00
25%        2.00
50%        3.00
75%        4.00
max        5.00
Name: bedrooms, dtype: float64
```

Find the median prices for properties based on number of bedrooms

In [ ]:
```python
median_onetofive = onetofivebeds.groupby("bedrooms").list_price.apply(lam
median_onetofive
```

Out[ ]:

	bedrooms	list_price
0	1	220000.00
1	2	300000.00
2	3	450000.00
3	4	650000.00
4	5	775000.00

Visualise the above in a scatterplot to demonstrate positive upward trajectory

In [ ]:
```python
plt.scatter(x=median_onetofive.bedrooms, y=median_onetofive.list_price)
plt.xlabel("Number of bedrooms")
plt.ylabel("List Price (£)")
plt.show()
```
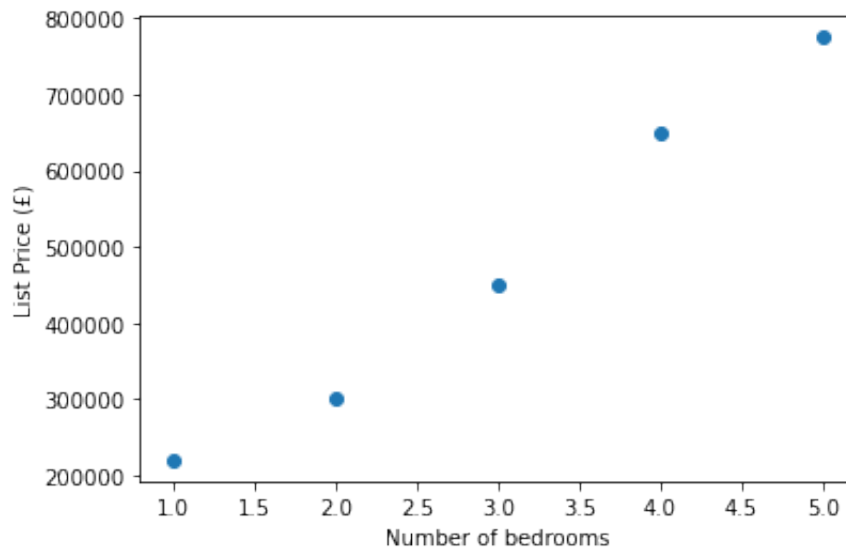
```
In [ ]:  df_under1m.list_info.value_counts()
```

```
Out[ ]:  flat                        252
         detached house              190
         semi-detached house         169
         terraced house              103
         end terrace house            50
         maisonette                   23
         detached bungalow            17
         studio                       15
         bungalow                     15
         town house                   15
         link-detached house          15
         property                     10
         cottage                       5
         mobile/park home              4
         penthouse                     1
         semi-detached bungalow        1
         etached house                 1
         lodge                         1
         Name: list_info, dtype: int64
```

Create a function to divide properties into two main categories: flat or house

```python
def house_flats(value):
    if "flat" in value:
        return "flat"
    elif "maisonette" in value:
        return "flat"
    elif value == "studio":
        return "flat"
    elif value == "penthouse":
        return "flat"
    elif value == "lodge":
        return "flat"
    else:
        return "house"

df_under1m["property_type"] = df_under1m.list_info.apply(lambda x: house_
df_under1m.sample(30)
```

```
/var/folders/n6/w301ddls6l330t92ry0gygj80000gn/T/ipykernel_4904/424341526
4.py:15: SettingWithCopyWarning:
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-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_under1m["property_type"] = df_under1m.list_info.apply(lambda x: hous
e_flats(x))
```

Out [ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_roor
847	325000	terraced house	[Bedrooms, 2, Bathrooms, 1, Living rooms, 2,]	Bedford Road, Reading, Berkshire RG1	2	1	
325	440000	end terrace house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Norreys Avenue Wokingham, Berkshire RG40	3	1	
540	220000	flat	[Bedrooms, 2, Bathrooms, 2, Living rooms, 1,]	Westgate Court, Oxford Road, Reading, Berkshir...	2	2	
363	205000	flat	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	The Green, Theale, Reading, Berkshire RG7	1	1	
670	325000	semi-detached house	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Shilling Close, Tilehurst, Reading, Berkshire ...	2	1	
			[Bedrooms,				

703	525000	semi-detached house	3, Bathrooms, 1, Living rooms, 2,]	Wroxham Road, Woodley, Reading RG5	3	1	
601	335000	flat	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Hartslock Court, Shooters Hill, Pangbourne, Be...	3	1	
728	499950	semi-detached house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Wokingham Road, Earley, Reading RG6	3	1	
172	525000	detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 2,]	Rushmoor Gardens, Calcot, Reading RG31	4	2	
800	475000	semi-detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 1,]	Boston Avenue, Reading RG1	4	2	
315	400000	end terrace house	[Bedrooms, 3, Bathrooms, 2, Living rooms, 1,]	Caribou Walk, Three Mile Cross, Reading, Berks...	3	2	
207	795000	detached house	[Bedrooms, 4,]	Hermitage Drive, Twyford, Berkshire RG10	4	NaN	N
107	675000	detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 2,]	Nabbs Hill Close, Tilehurst, Reading, Berkshir...	4	2	
462	260000	flat	[Bedrooms, 2, Living rooms, 1,]	Reading, Berkshire RG1	2	NaN	
383	275000	flat	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Reading, Berkshire RG2	1	1	
485	350000	flat	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Oak Tree Road, Tilehurst, Reading RG31	2	1	
			[Bedrooms,				

911	400000	terraced house	4, Bathrooms, 2, Living rooms, 2,]	Swainstone Road, Reading RG2	4	2	
329	450000	end terrace house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Reading RG30,	3	1	
809	550000	semi-detached house	[Bedrooms, 4, Bathrooms, 1, Living rooms, 2,]	Forest Road, Wokingham, Berkshire RG40	4	1	
503	314950	flat	[Bedrooms, 2,]	"Iver" at Sheerlands Road, Finchampstead RG40 ...	2	NaN	N
888	315000	terraced house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Drovers Way, Woodley, Reading RG5	3	1	
98	799995	detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 3,]	"Holden" at Warren House Road, Wokingham RG40	4	2	
115	800000	detached house	[Bedrooms, 4, Bathrooms, 2, Living rooms, 2,]	Swallowfield RG7,	4	2	
885	265000	terraced house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Carron Close, Tilehurst, Reading RG30	3	1	
790	550000	semi-detached house	[Bedrooms, 4, Bathrooms, 1, Living rooms, 2,]	Colemans Moor Road, Woodley RG5	4	1	
925	600000	terraced house	[Bedrooms, 5, Bathrooms, 3, Living rooms, 2,]	Cintra Close, Reading, Berkshire RG2	5	3	
624	220000	maisonette	[Bedrooms, 1, Bathrooms,	Reading Road, Winnersh RG41	1	1	

|						
			1, Living rooms, 1,]			
792	550000	semi-detached house	[Bedrooms, 4, Bathrooms, 3, Living rooms, 1,]	Champlain Street, Reading, Berkshire RG2	4	3
938	895000	town house	[Bedrooms, 3, Bathrooms, 2, Living rooms, 2,]	Easthampstead Road, Wokingham RG40	3	2
872	375000	terraced house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Pitford Road, Woodley, Reading RG5	3	1

In [ ]:
```python
df_under1m.property_type.value_counts()
```

Out[ ]:
```
house     597
flat      290
Name: property_type, dtype: int64
```

Visualise the difference in price between flats and houses

In [ ]:
```python
sns.boxplot(data=df_under1m, x="property_type", y="list_price")
plt.show()
```
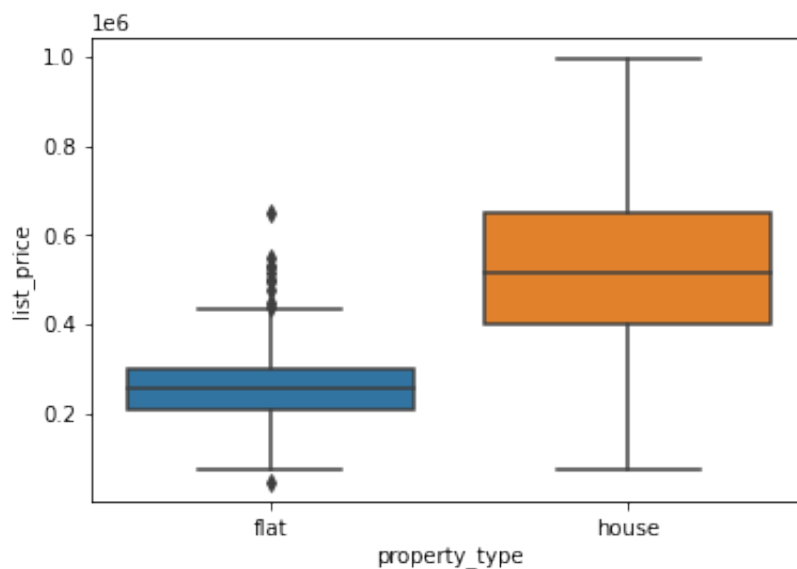

Find the median of each category: number of bed - median list price

In [ ]:
```python
flat_house = df_under1m.groupby(["property_type", "bedrooms"])["list_pric
flat_house.head(20)
```

Out[ ]:

	property_type	bedrooms	list_price
0	flat	0	160000.00
1	flat	1	220000.00
2	flat	2	290000.00
3	flat	3	325000.00
4	flat	4	650000.00
5	house	0	495000.00
6	house	1	265000.00
7	house	2	350000.00
8	house	3	460000.00
9	house	4	650000.00
10	house	5	775000.00
11	house	6	600000.00
12	house	7	700000.00
13	house	8	600000.00
14	house	9	675000.00

Use pivot function to reorganise table

In [ ]:
```
flat_house_pivot = flat_house.pivot(columns="property_type", index="bedro
flat_house_pivot
```

Out[ ]:

property_type	flat	house
bedrooms		
0	160000.00	495000.00
1	220000.00	265000.00
2	290000.00	350000.00
3	325000.00	460000.00
4	650000.00	650000.00
5	NaN	775000.00
6	NaN	600000.00
7	NaN	700000.00
8	NaN	600000.00
9	NaN	675000.00

In [ ]:
```
flat_house_pivot["percentage of increase"] = ((flat_house_pivot.house - f
flat_house_pivot
```

Out[ ]:

property_type	flat	house	percentage of increase
bedrooms			
0	160000.00	495000.00	209.38
1	220000.00	265000.00	20.45
2	290000.00	350000.00	20.69
3	325000.00	460000.00	41.54
4	650000.00	650000.00	0.00
5	NaN	775000.00	NaN
6	NaN	600000.00	NaN
7	NaN	700000.00	NaN
8	NaN	600000.00	NaN
9	NaN	675000.00	NaN

In [ ]:
```
flats = df_under1m[df_under1m.property_type == "flat"]
houses = df_under1m[df_under1m.property_type == "house"]

flats.head(10)
```

Out[ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms	prop
0	75000	studio	[Bedrooms, 1,]	Flat 3, 34 Eastern Avenue, Reading, Berkshire RG1	1	NaN	NaN	
1	130000	studio	[Bedrooms, 1, Bathrooms, 1, Living rooms, 1,]	Watersfield Close, Lower Earley, Reading RG6	1	1	1	
2	150000	studio	[Bathrooms, 1, Living rooms, 1,]	Colleton Drive, Twyford, Berkshire RG10	0	1	1	
3	140000	studio	[Bathrooms, 1, Living rooms, 1,]	George Street, Reading RG1	0	1	1	
4	175000	studio	[Living rooms, 1,]	Central Reading RG1,	0	NaN	1	
5	475000	studio	[Bathrooms, 1, Living rooms, 1,]	Hobbs End, Henley-On-Thames RG9	0	1	1	
6	160000	studio	[Bathrooms, 1, Living rooms, 1,]	Beech Lane, Lower Earley, Reading RG6	0	1	1	
7	135000	studio	[Bathrooms, 1, Living rooms, 1,]	Tippett Rise, Reading RG2	0	1	1	
8	130000	studio	[Bathrooms, 1, Living rooms, 1,]	Castle Hill, Reading, Berkshire RG1	0	1	1	
9	160000	studio	[Bathrooms, 1,]	Tippett Rise, Reading, Berkshire RG2	0	1	NaN	

```
In [ ]: print(len(flats))
```

290

```
In [ ]: houses.head(10)
```

```
In [ ]: print(len(flats))
```

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_rooms
16	399000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 2,]	Bridport Close, Lower Earley, Reading RG6	2	1	2
17	425000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Faygate Way, Lower Earley, Reading RG6	2	1	1
18	330000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Rushmoor Gardens, Calcot, Reading RG31	2	1	1
19	450000	bungalow	[Bedrooms, 2, Bathrooms, 2, Living rooms, 2,]	Birch Lane, Mortimer Common, Reading RG7	2	2	2
20	300000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Shirley Avenue, Reading RG2	2	1	1
21	350000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Highgrove Mews, Highgrove Street, Reading, Ber...	2	1	1
22	350000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Milsom Close, Shinfield, Reading, Berkshire RG2	2	1	1
23	640000	bungalow	[Bedrooms, 3, Bathrooms, 2, Living rooms, 3,]	Westview Drive, Twyford, Reading, Berkshire RG10	3	2	3
24	525000	bungalow	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	The Vines, Wokingham, Berkshire RG41	3	1	1
25	625000	bungalow	[Bedrooms, 3, Bathrooms, 2, Living rooms, 2,]	Windsor Ride, Finchampstead, Wokingham RG40	3	2	2

In [ ]:
```python
print(len(houses))
```
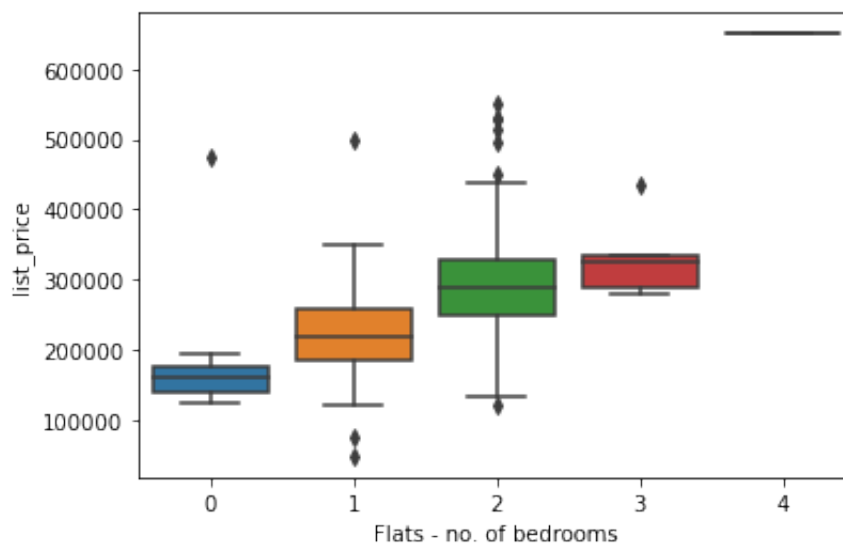
597

Visualise the difference in price for houses

In [ ]:
```python
sns.boxplot(data=houses, x="bedrooms", y="list_price")
plt.xlabel("Houses - no. of bedrooms")
plt.show()
```
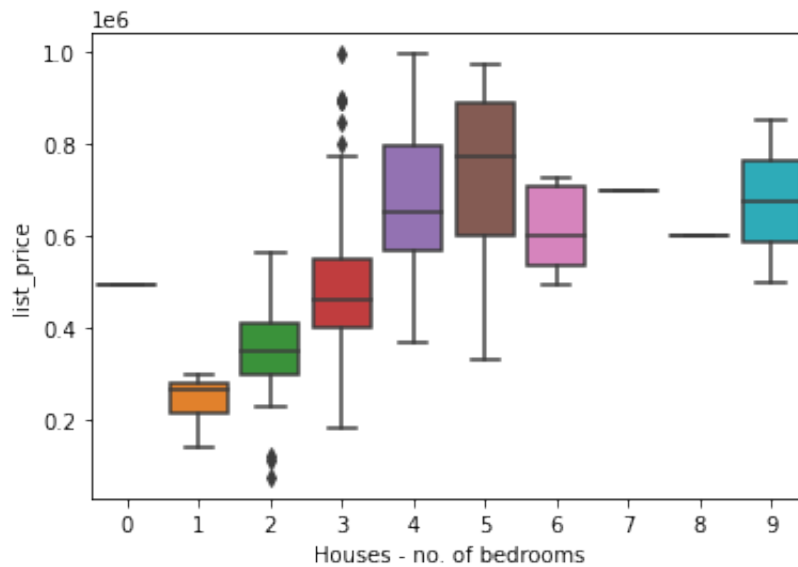

Visualise the difference in price for flats

In [ ]:
```python
sns.boxplot(data=flats, x="bedrooms", y="list_price")
plt.xlabel("Flats - no. of bedrooms")
plt.show()
```


In [ ]:
```python
flats.describe()
```

Out [ ]:

	list_price	bedrooms
count	290.00	290.00
mean	264065.55	1.56
std	80923.76	0.63
min	48750.00	0.00
25%	210000.00	1.00
50%	258475.00	2.00
75%	300000.00	2.00
max	650000.00	4.00

In [ ]:
```python
flats.list_price.median()
```

Out [ ]: 258475.0

Visualise the distrubtion of list price for flats only

In [ ]:
```python
sns.histplot(data=flats, x="list_price", kde=True)
plt.ticklabel_format(style="plain")
plt.xlabel("Flats - list price")
plt.show()
plt.close()
```
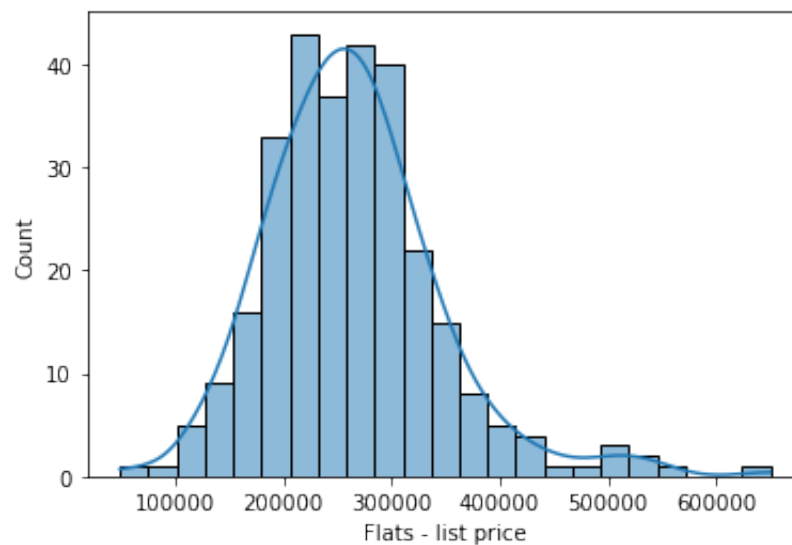

In [ ]:
```python
houses.describe()
```

Out[ ]:

	list_price	bedrooms
count	597.00	597.00
mean	542060.14	3.36
std	183063.93	1.02
min	74950.00	0.00
25%	400000.00	3.00
50%	515000.00	3.00
75%	650000.00	4.00
max	995000.00	9.00

In [ ]:
```
houses.list_price.median()
```

Out[ ]: 515000.0

In [ ]:
```
houses[houses.bedrooms == 2].describe()
```

Out[ ]:

	list_price	bedrooms
count	81.00	81.00
mean	355010.37	2.00
std	86222.88	0.00
min	74950.00	2.00
25%	300000.00	2.00
50%	350000.00	2.00
75%	412000.00	2.00
max	565000.00	2.00

In [ ]:
```
houses[houses.bedrooms == 3].describe()
```

Out[ ]:

	list_price	bedrooms
count	255.00	255.00
mean	483343.08	3.00
std	124584.80	0.00
min	180000.00	3.00
25%	400000.00	3.00
50%	460000.00	3.00
75%	550000.00	3.00
max	995000.00	3.00

In [ ]: `flats[flats.bedrooms == 2].describe()`

Out[ ]:

	list_price	bedrooms
count	163.00	163.00
mean	295189.45	2.00
std	70080.26	0.00
min	122000.00	2.00
25%	250000.00	2.00
50%	290000.00	2.00
75%	327475.00	2.00
max	550000.00	2.00

In [ ]: `flats[flats.bedrooms == 3].describe()`

Out[ ]:

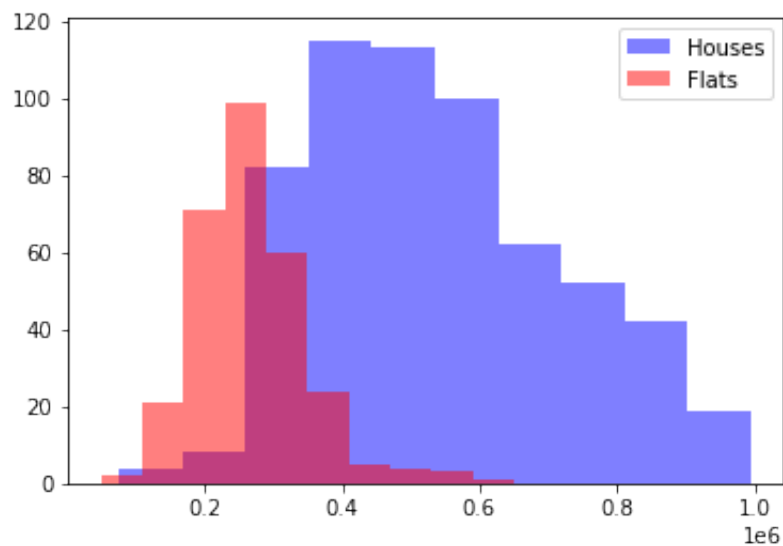|	list_price	bedrooms
count	5.00	5.00
mean	333000.00	3.00
std	61502.03	0.00
min	280000.00	3.00
25%	290000.00	3.00
50%	325000.00	3.00
75%	335000.00	3.00
max	435000.00	3.00

Visualise the distribution of list price for houses

```
In [ ]:  sns.histplot(data=houses, x="list_price", kde=True)
         plt.ticklabel_format(style="plain")
         plt.xlabel("Houses - list price")
         plt.show()
         plt.close()
```
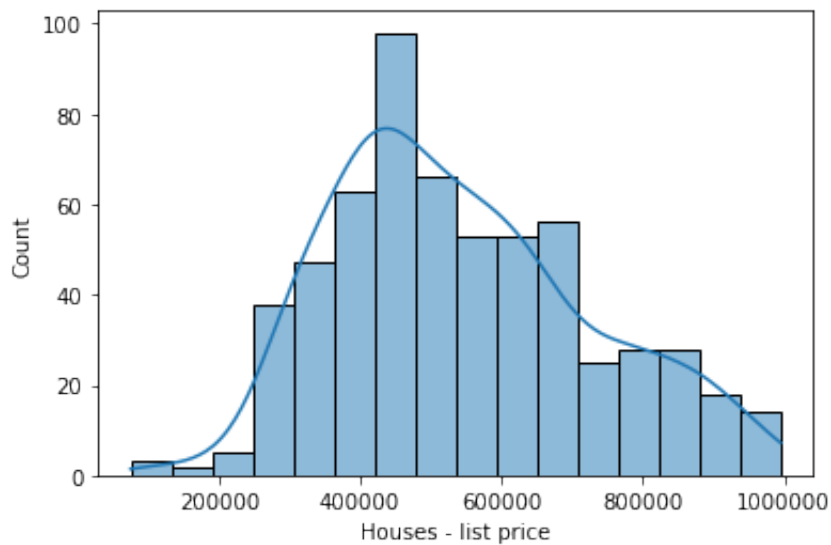

Visualise the difference in list price between houses and flats using an overlapping histogram

```
In [ ]:  plt.hist(houses.list_price, color="blue", label="Houses", alpha=0.5)
         plt.hist(flats.list_price, color="red", label="Flats", alpha=0.5)
         plt.legend()
         plt.show()
```


As we are most interested in two-bed and three-bed properties, let's compare the list price of these properties.

Create a subset of the datset which only contains two-bed and three-bed properties.

In [ ]:
```python
two_three_beds = df_under1m[(df_under1m.bedrooms == 2) | (df_under1m.bedr
two_three_beds.sample(10)
```

Out[ ]:

	list_price	list_info	rooms	location	bedrooms	bathrooms	living_roon
551	250000	flat	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Meadow Way, Amersham Road, Caversham, Berkshir...	2	1	
904	350000	terraced house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 2,]	Bourton Close, Tilehurst, Reading RG30	3	1	
873	389950	terraced house	[Bedrooms, 3, Bathrooms, 1, Living rooms, 1,]	Medina Close, Wokingham RG41	3	1	
83	425000	detached house	[Bedrooms, 3, Bathrooms, 2, Living rooms, 1,]	Mill Close, Wokingham, Berkshire RG41	3	2	
877	500000	terraced house	[Bedrooms, 3, Living rooms, 1,]	Woodley, Berkshire RG5	3	NaN	
466	265000	flat	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Northumberland Avenue, Reading, Berkshire RG2	2	1	
691	400000	semi-detached house	[Bedrooms, 3, Bathrooms, 2, Living rooms, 3,]	Curtis Road, Calcot, Reading, Berkshire RG31	3	2	
17	425000	bungalow	[Bedrooms, 2, Bathrooms, 1, Living rooms, 1,]	Faygate Way, Lower Earley, Reading RG6	2	1	
678	400000	semi-detached house	[Bedrooms, 2, Bathrooms, 2, Living rooms, 2,]	Stoneham Close, Tilehurst, Reading RG30	2	2	
520	240000	flat	[Bedrooms, 2, Bathrooms, 2, Living rooms, 1,]	Ashdene Gardens, Reading, Berkshire RG30	2	2	

Visualise the respective list price of the four types of properties: 2-bed houses, 2-bed flats, 3-bed houses, 3-bed flats.

```
In [ ]: sns.boxplot(data=two_three_beds, x="bedrooms", y="list_price", hue="prope
        plt.show()
```
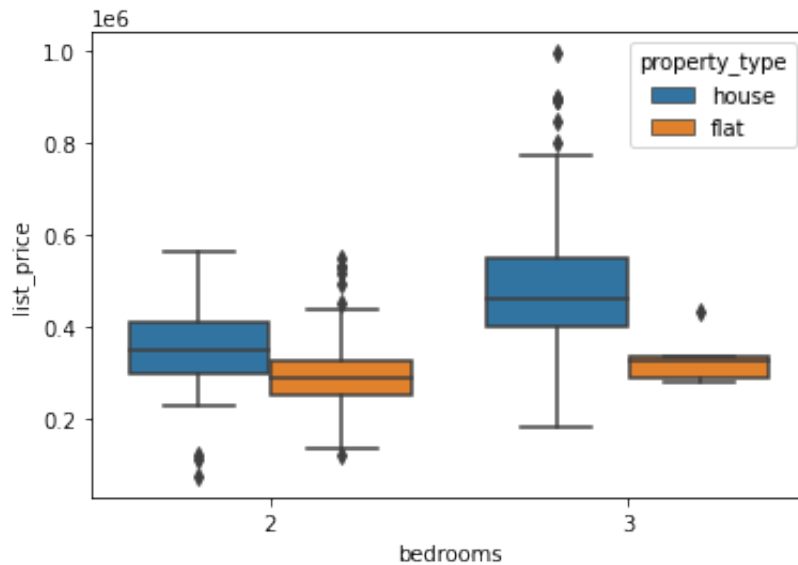

There is some overlap in list price between 2-bed flats and 2-bed houses. The difference in list price between 3-bed flats and 3-bed houses is more significant, with no overlap. Note the much smaller amount of 3-bed flats on the market.

End of analysis