

Data Cleaning - Melbourne House Data



In [3]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Read CSV

In [97]:

```
data = pd.read_csv('Melbourne.csv')
OrgData = data
data
```

Out[97]:

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
0	Abbotsford	68 Studley St	2	h	NaN	SS	Jellis	03-09-2016	2.5	306
1	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	03-12-2016	2.5	306
2	Abbotsford	25 Bloomberg St	2	h	1035000.0	S	Biggin	04-02-2016	2.5	306
3	Abbotsford	18/659 Victoria St	3	u	NaN	VB	Rounds	04-02-2016	2.5	306
4	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	04-03-2017	2.5	306
...
23542	Wyndham Vale	25 Clitheroe Dr	3	u	NaN	PN	Harcourts	26-08-2017	27.2	302
23543	Wyndham Vale	19 Dalrymple Bvd	4	h	NaN	S	hockingstuart	26-08-2017	27.2	302
23544	Yallambie	17 Amaroo Wy	4	h	1100000.0	S	Buckingham	26-08-2017	12.7	308
23545	Yarraville	6 Agnes St	4	h	1285000.0	SP	Village	26-08-2017	6.3	301
23546	Yarraville	33 Freeman St	4	h	1050000.0	VB	Village	26-08-2017	6.3	301

23547 rows × 21 columns

In [98]: `data.shape`

Out[98]: (23547, 21)

In [99]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23547 entries, 0 to 23546
Data columns (total 21 columns):
Suburb          23547 non-null object
Address          23547 non-null object
Rooms            23547 non-null int64
Type             23547 non-null object
Price            18396 non-null float64
Method            23547 non-null object
SellerG           23547 non-null object
Date              23547 non-null object
Distance          23546 non-null float64
Postcode          23546 non-null float64
Bedroom2          19066 non-null float64
Bathroom          19063 non-null float64
Car               18921 non-null float64
Landsize          17410 non-null float64
BuildingArea      10018 non-null float64
YearBuilt          11540 non-null float64
CouncilArea        15656 non-null object
Latitude          19243 non-null float64
Longitude          19243 non-null float64
Regionname         23546 non-null object
Propertycount     23546 non-null float64
dtypes: float64(12), int64(1), object(8)
memory usage: 3.8+ MB
```

Summing up the missing values (column-wise)

In [100...]: `data.isnull().sum(axis=0).sort_values(ascending=False)`

Out[100...]	BuildingArea	13529
	YearBuilt	12007
	CouncilArea	7891
	Landsize	6137
	Price	5151
	Car	4626
	Bathroom	4484
	Bedroom2	4481
	Longitude	4304
	Latitude	4304
	Distance	1
	Propertycount	1
	Postcode	1
	Regionname	1
	Date	0
	SellerG	0
	Method	0
	Type	0

```
Rooms          0
Address        0
Suburb         0
dtype: int64
```

Summing up the missing values (row-wise)

```
In [101...]: data.isnull().sum(axis=1).sort_values(ascending=False).head(10)
```

```
Out[101...]: 18523    13
17100     10
16232     10
20968     10
6313      10
6310      10
6299      10
6293      10
16225     10
6289      10
dtype: int64
```

Columns having at least one missing value

```
In [102...]: data.isnull().sum() > 0
```

```
Out[102...]: Suburb      False
Address      False
Rooms       False
Type        False
Price        True
Method      False
SellerG     False
Date        False
Distance    True
Postcode    True
Bedroom2    True
Bathroom    True
Car          True
Landsize    True
BuildingArea True
YearBuilt   True
CouncilArea True
Latitude   True
Longitude  True
Regionname  True
Propertycount True
dtype: bool
```

```
In [103...]: d = data.isnull().any()
d
```

```
Out[103...]: Suburb      False
Address      False
Rooms       False
Type        False
Price        True
Method      False
SellerG     False
Date        False
Distance    True
```

```
Postcode      True
Bedroom2      True
Bathroom      True
Car           True
Landsize       True
BuildingArea   True
YearBuilt      True
CouncilArea    True
Latitude       True
Longitude      True
Regionname     True
Propertycount  True
dtype: bool
```

```
In [104... d.index
```

```
Out[104... Index(['Suburb', 'Address', 'Rooms', 'Type', 'Price', 'Method', 'SellerG',
                  'Date', 'Distance', 'Postcode', 'Bedroom2', 'Bathroom', 'Car',
                  'Landsize', 'BuildingArea', 'YearBuilt', 'CouncilArea', 'Latitude',
                  'Longitude', 'Regionname', 'Propertycount'],
                  dtype='object')
```

```
In [105... d.values
```

```
Out[105... array([False, False, False, False,  True, False, False,  True,
                  True,  True,  True,  True,  True,  True,  True,  True,
                  True,  True,  True])
```

```
In [106... d.index[d.values]
```

```
Out[106... Index(['Price', 'Distance', 'Postcode', 'Bedroom2', 'Bathroom', 'Car',
                  'Landsize', 'BuildingArea', 'YearBuilt', 'CouncilArea', 'Latitude',
                  'Longitude', 'Regionname', 'Propertycount'],
                  dtype='object')
```

By default any() operates on columns

```
In [107... data.isnull().any(axis=0)
```

```
Out[107... Suburb      False
Address      False
Rooms        False
Type         False
Price         True
Method        False
SellerG      False
Date          False
Distance      True
Postcode      True
Bedroom2     True
Bathroom      True
Car           True
Landsize      True
BuildingArea  True
YearBuilt     True
CouncilArea   True
Latitude      True
Longitude     True
Regionname    True
```

```
Propertycount      True  
dtype: bool
```

```
In [108...     d = data.isnull().any(axis=1)  
d
```

```
Out[108... 0      True  
1      True  
2    False  
3      True  
4    False  
...  
23542   True  
23543   True  
23544   True  
23545   True  
23546   True  
Length: 23547, dtype: bool
```

```
In [109... index = d.index[d.values]  
index
```

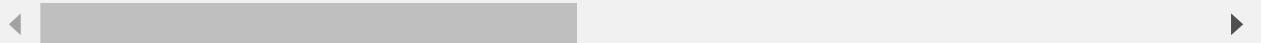
```
Out[109... Int64Index([ 0, 1, 3, 5, 7, 8, 9, 10, 12,  
13,  
...,  
23537, 23538, 23539, 23540, 23541, 23542, 23543, 23544, 23545,  
23546],  
dtype='int64', length=17351)
```

```
In [110... data.iloc[index,:]
```

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
0	Abbotsford	68 Studley St	2	h	NaN	SS	Jellis	03-09-2016	2.5	3067
1	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	03-12-2016	2.5	3067
3	Abbotsford	18/659 Victoria St	3	u	NaN	VB	Rounds	04-02-2016	2.5	3067
5	Abbotsford	40 Federation La	3	h	850000.0	PI	Biggin	04-03-2017	2.5	3067
7	Abbotsford	16 Maugie St	4	h	NaN	SN	Nelson	06-08-2016	2.5	3067
...
23542	Wyndham Vale	25 Clitheroe Dr	3	u	NaN	PN	Harcourts	26-08-2017	27.2	3024

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode
23543	Wyndham Vale	19 Dalrymple Bvd	4	h	NaN	S	hockingstuart	26-08-2017	27.2	3024
23544	Yallambie	17 Amaroo Wy	4	h	1100000.0	S	Buckingham	26-08-2017	12.7	3085
23545	Yarraville	6 Agnes St	4	h	1285000.0	SP	Village	26-08-2017	6.3	3013
23546	Yarraville	33 Freeman St	4	h	1050000.0	VB	Village	26-08-2017	6.3	3013

17351 rows × 21 columns



Columns having all missing values

```
In [111]: data.isnull().all(axis=0)
```

```
Out[111]: Suburb      False
Address     False
Rooms       False
Type        False
Price       False
Method      False
SellerG    False
Date        False
Distance   False
Postcode   False
Bedroom2   False
Bathroom   False
Car         False
Landsize    False
BuildingArea False
YearBuilt   False
CouncilArea False
Latitude   False
Longitude  False
Regionname False
Propertycount False
dtype: bool
```

Rows having all misisng values

```
In [112]: data.isnull().all(axis=1).sum()
```

```
Out[112]: 0
```

```
In [113]: data.isnull().sum(axis=1)==21
```

```
Out[113... 0      False
       1      False
       2      False
       3      False
       4      False
       ...
      23542  False
      23543  False
      23544  False
      23545  False
      23546  False
Length: 23547, dtype: bool
```

Summing up the missing values (Column-wise) : Cal in %

```
In [114... data.isnull().sum().sort_values(ascending=False)/len(data) * 100
```

```
Out[114... BuildingArea      57.455302
YearBuilt          50.991634
CouncilArea        33.511700
Landsize           26.062768
Price              21.875398
Car                19.645815
Bathroom          19.042766
Bedroom2          19.030025
Longitude         18.278337
Latitude          18.278337
Distance           0.004247
Propertycount     0.004247
Postcode           0.004247
Regionname         0.004247
Date               0.000000
SellerG            0.000000
Method              0.000000
Type               0.000000
Rooms              0.000000
Address             0.000000
Suburb              0.000000
dtype: float64
```

```
In [115... round(data.isnull().sum().sort_values(ascending=False)/len(data) * 100,2)
```

```
Out[115... BuildingArea      57.46
YearBuilt          50.99
CouncilArea        33.51
Landsize           26.06
Price              21.88
Car                19.65
Bathroom          19.04
Bedroom2          19.03
Longitude         18.28
Latitude          18.28
Distance           0.00
Propertycount     0.00
Postcode           0.00
Regionname         0.00
Date               0.00
SellerG            0.00
Method              0.00
Type               0.00
Rooms              0.00
Address             0.00
```

```
Suburb          0.00
dtype: float64
```

Removing the three columns where the max null value percentage

```
In [116...]: col = data.isnull().sum().sort_values(ascending=False).head(3).index.values
col
```

```
Out[116...]: array(['BuildingArea', 'YearBuilt', 'CouncilArea'], dtype=object)
```

```
In [117...]: data = data.drop(col, axis='columns')
data
```

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
0	Abbotsford	68 Studley St	2	h	NaN	SS	Jellis	03-09-2016	2.5	306
1	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	03-12-2016	2.5	306
2	Abbotsford	25 Bloomberg St	2	h	1035000.0	S	Biggin	04-02-2016	2.5	306
3	Abbotsford	18/659 Victoria St	3	u	NaN	VB	Rounds	04-02-2016	2.5	306
4	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	04-03-2017	2.5	306
...
23542	Wyndham Vale	25 Clitheroe Dr	3	u	NaN	PN	Harcourts	26-08-2017	27.2	302
23543	Wyndham Vale	19 Dalrymple Bvd	4	h	NaN	S	hockingstuart	26-08-2017	27.2	302
23544	Yallambie	17 Amaroo Wy	4	h	1100000.0	S	Buckingham	26-08-2017	12.7	308
23545	Yarraville	6 Agnes St	4	h	1285000.0	SP	Village	26-08-2017	6.3	301
23546	Yarraville	33 Freeman St	4	h	1050000.0	VB	Village	26-08-2017	6.3	301

23547 rows × 18 columns

```
In [118... round(data.isnull().sum().sort_values(ascending=False)/len(data) * 100,2)
```

```
Out[118]:
```

	Landsize	26.06
Price	21.88	
Car	19.65	
Bathroom	19.04	
Bedroom2	19.03	
Longitude	18.28	
Latitude	18.28	
Propertycount	0.00	
Regionname	0.00	
Distance	0.00	
Postcode	0.00	
Date	0.00	
SellerG	0.00	
Method	0.00	
Type	0.00	
Rooms	0.00	
Address	0.00	
Suburb	0.00	

dtype: float64

Check the rows where you have more than 5 missing values

```
In [119]: len(data[data.isnull().sum(axis=1) > 5])/len(data)*100
```

```
Out[119... 18.16791948018856
```

Retaining the rows having <=5 NaNs

```
In [120...]: data = data[data.isnull().sum(axis=1) <=5]
          data
```

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
23542	Wyndham Vale	25 Clitheroe Dr	3	u	NaN	PN	Harcourts	26-08-2017	27.2	302
23543	Wyndham Vale	19 Dalrymple Bvd	4	h	NaN	S	hockingstuart	26-08-2017	27.2	302
23544	Yallambie	17 Amaroo Wy	4	h	1100000.0	S	Buckingham	26-08-2017	12.7	308
23545	Yarraville	6 Agnes St	4	h	1285000.0	SP	Village	26-08-2017	6.3	301
23546	Yarraville	33 Freeman St	4	h	1050000.0	VB	Village	26-08-2017	6.3	301

19269 rows × 18 columns

In [121...]	data
--------------------	-------------

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
0	Abbotsford	68 Studley St	2	h	NaN	SS	Jellis	03-09-2016	2.5	306
1	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	03-12-2016	2.5	306
2	Abbotsford	25 Bloomberg St	2	h	1035000.0	S	Biggin	04-02-2016	2.5	306
3	Abbotsford	18/659 Victoria St	3	u	NaN	VB	Rounds	04-02-2016	2.5	306
4	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	04-03-2017	2.5	306
...
23542	Wyndham Vale	25 Clitheroe Dr	3	u	NaN	PN	Harcourts	26-08-2017	27.2	302
23543	Wyndham Vale	19 Dalrymple Bvd	4	h	NaN	S	hockingstuart	26-08-2017	27.2	302

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
23544	Yallambie	17 Amaroo Wy	4	h	1100000.0	S	Buckingham	26-08-2017	12.7	308
23545	Yarraville	6 Agnes St	4	h	1285000.0	SP	Village	26-08-2017	6.3	301
23546	Yarraville	33 Freeman St	4	h	1050000.0	VB	Village	26-08-2017	6.3	301

19269 rows × 18 columns



```
In [122...]: round(data.isnull().sum().sort_values(ascending=False)/len(data) * 100,2)
```

```
Out[122...]: Price          21.71
Landsize        9.65
Car             1.81
Bathroom        1.07
Bedroom2        1.05
Longitude       0.13
Latitude        0.13
Propertycount   0.00
Method          0.00
Address         0.00
Rooms           0.00
Type            0.00
Distance        0.00
SellerG         0.00
Date            0.00
Regionname      0.00
Postcode        0.00
Suburb          0.00
dtype: float64
```

```
In [123...]: data[data['Price'].notnull()]
```

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
1	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	03-12-2016	2.5	306
2	Abbotsford	25 Bloomberg St	2	h	1035000.0	S	Biggin	04-02-2016	2.5	306
4	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	04-03-2017	2.5	306
5	Abbotsford	40 Federation La	3	h	850000.0	PI	Biggin	04-03-2017	2.5	306

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
6	Abbotsford	55a Park St	4	h	1600000.0	VB	Nelson	04-06-2016	2.5	306
...
23540	Williamstown	8/2 Thompson St	2	t	622500.0	SP	Greg	26-08-2017	6.8	301
23541	Williamstown	96 Verdon St	4	h	2500000.0	PI	Sweeney	26-08-2017	6.8	301
23544	Yallambie	17 Amaroo Wy	4	h	1100000.0	S	Buckingham	26-08-2017	12.7	308
23545	Yarraville	6 Agnes St	4	h	1285000.0	SP	Village	26-08-2017	6.3	301
23546	Yarraville	33 Freeman St	4	h	1050000.0	VB	Village	26-08-2017	6.3	301

15086 rows × 18 columns

Another method

In [124...]

```
data[~np.isnan(data['Price'])]
```

Out[124...]

		Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
23540	Williamstown		8/2 Thompson St	2	t	622500.0	SP	Greg	26-08-2017	6.8	301
23541	Williamstown		96 Verdon St	4	h	2500000.0	PI	Sweeney	26-08-2017	6.8	301
23544	Yallambie		17 Amaroo Wy	4	h	1100000.0	S	Buckingham	26-08-2017	12.7	308
23545	Yarraville		6 Agnes St	4	h	1285000.0	SP	Village	26-08-2017	6.3	301
23546	Yarraville		33 Freeman St	4	h	1050000.0	VB	Village	26-08-2017	6.3	301

15086 rows × 18 columns



In [125...]

```
data = data[data['Price'].notnull()]
data
```

Out[125...]

		Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postco
1	Abbotsford		85 Turner St	2	h	1480000.0	S	Biggin	03-12-2016	2.5	306
2	Abbotsford		25 Bloomberg St	2	h	1035000.0	S	Biggin	04-02-2016	2.5	306
4	Abbotsford		5 Charles St	3	h	1465000.0	SP	Biggin	04-03-2017	2.5	306
5	Abbotsford		40 Federation La	3	h	850000.0	PI	Biggin	04-03-2017	2.5	306
6	Abbotsford		55a Park St	4	h	1600000.0	VB	Nelson	04-06-2016	2.5	306
...
23540	Williamstown		8/2 Thompson St	2	t	622500.0	SP	Greg	26-08-2017	6.8	301
23541	Williamstown		96 Verdon St	4	h	2500000.0	PI	Sweeney	26-08-2017	6.8	301

		Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode
23544	Yallambie	17 Amaroo Wy		4	h	1100000.0	S	Buckingham	26-08-2017	12.7	308
23545	Yarraville	6 Agnes St		4	h	1285000.0	SP	Village	26-08-2017	6.3	301
23546	Yarraville	33 Freeman St		4	h	1050000.0	VB	Village	26-08-2017	6.3	301

15086 rows × 18 columns



```
In [126...]: round(data.isnull().sum().sort_values(ascending=False)/len(data) * 100,2)
```

```
Out[126...]: Landsize      9.83
Car          1.76
Bathroom     1.07
Bedroom2     1.05
Longitude    0.15
Latitude     0.15
Propertycount 0.00
Method        0.00
Address       0.00
Rooms         0.00
Type          0.00
Price          0.00
Distance       0.00
SellerG        0.00
Date          0.00
Regionname    0.00
Postcode       0.00
Suburb        0.00
dtype: float64
```

```
In [127...]: data['Landsize'].describe()
```

```
Out[127...]: count    13603.000000
mean      558.116371
std       3987.326586
min       0.000000
25%      176.500000
50%      440.000000
75%      651.000000
max     433014.000000
Name: Landsize, dtype: float64
```

```
In [128...]: data = data[data['Landsize'].notnull()]
data
```

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode
--	--------	---------	-------	------	-------	--------	---------	------	----------	----------

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode
1	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	03-12-2016	2.5	3067.0
2	Abbotsford	25 Bloomberg St	2	h	1035000.0	S	Biggin	04-02-2016	2.5	3067.0
4	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	04-03-2017	2.5	3067.0
5	Abbotsford	40 Federation La	3	h	850000.0	PI	Biggin	04-03-2017	2.5	3067.0
6	Abbotsford	55a Park St	4	h	1600000.0	VB	Nelson	04-06-2016	2.5	3067.0
...
23537	Wheelers Hill	12 Strada Cr	4	h	1245000.0	S	Barry	26-08-2017	16.7	3150.0
23538	Williamstown	77 Merrett Dr	3	h	1031000.0	SP	Williams	26-08-2017	6.8	3016.0
23539	Williamstown	83 Power St	3	h	1170000.0	S	Raine	26-08-2017	6.8	3016.0
23541	Williamstown	96 Verdon St	4	h	2500000.0	PI	Sweeney	26-08-2017	6.8	3016.0
23545	Yarraville	6 Agnes St	4	h	1285000.0	SP	Village	26-08-2017	6.3	3013.0

13603 rows × 18 columns



In [129...]

```
round(data.isnull().sum().sort_values(ascending=False)/len(data) * 100,2)
```

Out[129...]

Car	0.46
Longitude	0.16
Latitude	0.16
Bathroom	0.01
Propertycount	0.00
SellerG	0.00
Address	0.00
Rooms	0.00
Type	0.00
Price	0.00
Method	0.00
Distance	0.00

```
Date          0.00  
Regionname   0.00  
Postcode      0.00  
Bedroom2      0.00  
Landsize       0.00  
Suburb        0.00  
dtype: float64
```

Describe Latitude and Longtitude and later imputing the values by Mean

```
In [130... data.loc[:,['Latitude','Longitude']].describe()
```

```
Out[130...  


|              | Latitude     | Longitude    |
|--------------|--------------|--------------|
| <b>count</b> | 13581.000000 | 13581.000000 |
| <b>mean</b>  | -37.809204   | 144.995221   |
| <b>std</b>   | 0.079257     | 0.103913     |
| <b>min</b>   | -38.182550   | 144.431810   |
| <b>25%</b>   | -37.856820   | 144.929600   |
| <b>50%</b>   | -37.802360   | 145.000100   |
| <b>75%</b>   | -37.756400   | 145.058320   |
| <b>max</b>   | -37.408530   | 145.526350   |


```

```
In [131... data['Latitude'].mean()
```

```
Out[131... -37.809203506369194
```

```
In [132... data.loc[:, 'Latitude'].fillna(data['Latitude'].mean(), inplace=True)
```

```
In [133... data.loc[:, 'Longitude'].fillna(data['Longitude'].mean(), inplace=True)
```

```
In [134... round(data.isnull().sum().sort_values(ascending=False)/len(data) * 100,2)
```

```
Out[134... Car          0.46  
Bathroom     0.01  
Propertycount 0.00  
Date          0.00  
Address       0.00  
Rooms         0.00  
Type          0.00  
Price          0.00  
Method         0.00  
SellerG        0.00  
Distance       0.00  
Regionname    0.00  
Postcode       0.00  
Bedroom2       0.00  
Landsize       0.00
```

```
Latitude      0.00
Longitude     0.00
Suburb       0.00
dtype: float64
```

```
In [135...]: data.loc[:,['Bathroom','Car']].describe()
```

```
Out[135...]:
```

	Bathroom	Car
count	13602.000000	13540.000000
mean	1.534921	1.610414
std	0.691834	0.962244
min	0.000000	0.000000
25%	1.000000	1.000000
50%	1.000000	2.000000
75%	2.000000	2.000000
max	8.000000	10.000000

```
In [136...]: data.groupby('Car').Car.count().sort_values(ascending=False)
```

```
Out[136...]:
```

Car	Count
2.0	5606
1.0	5515
0.0	1026
3.0	748
4.0	507
5.0	63
6.0	54
8.0	9
7.0	8
10.0	3
9.0	1

Name: Car, dtype: int64

```
In [137...]: data['Car'].astype('category').value_counts()
```

```
Out[137...]:
```

Car	Count
2.0	5606
1.0	5515
0.0	1026
3.0	748
4.0	507
5.0	63
6.0	54
8.0	9
7.0	8
10.0	3
9.0	1

Name: Car, dtype: int64

```
In [138...]: data.loc[:, 'Car'].fillna(2, inplace=True)
```

```
In [139...]:
```

```
round(data.isnull().sum().sort_values(ascending=False)/len(data) * 100,2)
```

```
Out[139... Bathroom      0.01
Propertycount  0.00
Date          0.00
Address        0.00
Rooms          0.00
Type           0.00
Price          0.00
Method          0.00
SellerG         0.00
Distance        0.00
Regionname     0.00
Postcode        0.00
Bedroom2        0.00
Car             0.00
Landsize        0.00
Latitude        0.00
Longitude       0.00
Suburb          0.00
dtype: float64
```

```
In [140... data['Bathroom'].astype('category').value_counts()
```

```
Out[140... 1.0    7517
2.0    4987
3.0    921
4.0    106
0.0    34
5.0    28
6.0    5
8.0    2
7.0    2
Name: Bathroom, dtype: int64
```

```
In [141... data.loc[:, 'Bathroom'].fillna(1,inplace=True)
```

```
In [142... round(data.isnull().sum().sort_values(ascending=False)/len(data) * 100,2)
```

```
Out[142... Propertycount  0.0
Regionname     0.0
Address        0.0
Rooms          0.0
Type           0.0
Price          0.0
Method          0.0
SellerG         0.0
Date           0.0
Distance        0.0
Postcode        0.0
Bedroom2        0.0
Bathroom        0.0
Car             0.0
Landsize        0.0
Latitude        0.0
Longitude       0.0
Suburb          0.0
dtype: float64
```

```
In [144...
```

```
len(data)/len(OrgData) * 100
```

Out[144... 57.769567248481756

In []:

In []: