In [1]:
```
import pandas as pd
df = pd.read_csv(r'C:\Users\raj20\Downloads\melb_data.csv\melb_data.csv')
df.columns
```

Out[1]:
```
Index(['Suburb', 'Address', 'Rooms', 'Type', 'Price', 'Method', 'SellerG',
       'Date', 'Distance', 'Postcode', 'Bedroom2', 'Bathroom', 'Car',
       'Landsize', 'BuildingArea', 'YearBuilt', 'CouncilArea', 'Lattitude',
       'Longtitude', 'Regionname', 'Propertycount'],
      dtype='object')
```

In [2]:
```
df.isnull().any() #are there any missing values?
```

Out[2]:
```
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               True
Landsize         False
BuildingArea      True
YearBuilt         True
CouncilArea       True
Lattitude        False
Longtitude       False
Regionname       False
Propertycount    False
dtype: bool
```

In [3]:
```
df.isnull().sum() #no. of missing values
```

Out[3]:
```
Suburb             0
Address            0
Rooms              0
Type               0
Price              0
Method             0
SellerG            0
Date               0
Distance           0
Postcode           0
Bedroom2           0
Bathroom           0
Car               62
Landsize           0
BuildingArea    6450
YearBuilt       5375
CouncilArea     1369
Lattitude          0
Longtitude         0
Regionname         0
Propertycount      0
dtype: int64
```

In [4]:
```
#delete the columns having NaN
```

In [5]:
```
cols_with_missing = [col for col in df.columns
                     if df[col].isnull().any()]
print(cols_with_missing)
```

```
['Car', 'BuildingArea', 'YearBuilt', 'CouncilArea']
```

In [6]:
```
reduced_df = df.drop(cols_with_missing,axis =1) #axis 1 for col, 0 for rows
reduced_df.columns
```

Out[6]:
```
Index(['Suburb', 'Address', 'Rooms', 'Type', 'Price', 'Method', 'SellerG',
       'Date', 'Distance', 'Postcode', 'Bedroom2', 'Bathroom', 'Landsize',
       'Lattitude', 'Longtitude', 'Regionname', 'Propertycount'],
      dtype='object')
```

In [7]:
```
df.shape
```

Out[7]:
```
(13580, 21)
```

In [8]:
```
reduced_df.shape
```

Out[8]:
```
(13580, 17)
```

In [9]:
```
reduced_df2 = df.dropna()
reduced_df2
```

Out[9]:

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode	...	Bathroom	Car	Landsize	BuildingArea	YearBuilt	CouncilArea	Lattitude	Longtitude	Regionname	Propertycount
1	Abbotsford	25 Bloomburg St	2	h	1035000.0	S	Biggin	4/02/2016	2.5	3067.0	...	1.0	0.0	156.0	79.00	1900.0	Yarra	-37.80790	144.99340	Northern Metropolitan	4019.0
2	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	4/03/2017	2.5	3067.0	...	2.0	0.0	134.0	150.00	1900.0	Yarra	-37.80930	144.99440	Northern Metropolitan	4019.0
4	Abbotsford	55a Park St	4	h	1600000.0	VB	Nelson	4/06/2016	2.5	3067.0	...	1.0	2.0	120.0	142.00	2014.0	Yarra	-37.80720	144.99410	Northern Metropolitan	4019.0
6	Abbotsford	124 Yarra St	3	h	1876000.0	S	Nelson	7/05/2016	2.5	3067.0	...	2.0	0.0	245.0	210.00	1910.0	Yarra	-37.80240	144.99930	Northern Metropolitan	4019.0
7	Abbotsford	98 Charles St	2	h	1636000.0	S	Nelson	8/10/2016	2.5	3067.0	...	1.0	2.0	256.0	107.00	1890.0	Yarra	-37.80600	144.99540	Northern Metropolitan	4019.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
12205	Whittlesea	30 Sherwin St	3	h	601000.0	S	Ray	29/07/2017	35.5	3757.0	...	2.0	1.0	972.0	149.00	1996.0	Whittlesea	-37.51232	145.13282	Northern Victoria	2170.0
12206	Williamstown	75 Cecil St	3	h	1050000.0	VB	Williams	29/07/2017	6.8	3016.0	...	1.0	0.0	179.0	115.00	1890.0	Hobsons Bay	-37.86558	144.90474	Western Metropolitan	6380.0
12207	Williamstown	2/29 Dover Rd	1	u	385000.0	SP	Williams	29/07/2017	6.8	3016.0	...	1.0	1.0	0.0	35.64	1967.0	Hobsons Bay	-37.85588	144.89936	Western Metropolitan	6380.0
12209	Windsor	201/152 Peel St	2	u	560000.0	PI	hockingstuart	29/07/2017	4.6	3181.0	...	1.0	1.0	0.0	61.60	2012.0	Stonnington	-37.85581	144.99025	Southern Metropolitan	4380.0
12212	Yarraville	54 Pentland Pde	6	h	2450000.0	VB	Village	29/07/2017	6.3	3013.0	...	3.0	2.0	1087.0	388.50	1920.0	Maribyrnong	-37.81038	144.89389	Western Metropolitan	6543.0

6196 rows × 21 columns

In [10]:
```
reduced_df2.shape
```

Out[10]:
```
(6196, 21)
```

In [11]:
```
ml = pd.read_csv(r'C:\Users\raj20\Downloads\melb_data.csv\melb_data.csv')
```

In [12]:
```
print(ml.info())
print(ml.shape)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13580 entries, 0 to 13579
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Suburb         13580 non-null  object 
 1   Address        13580 non-null  object 
 2   Rooms          13580 non-null  int64  
 3   Type           13580 non-null  object 
 4   Price          13580 non-null  float64
 5   Method         13580 non-null  object 
 6   SellerG        13580 non-null  object 
 7   Date           13580 non-null  object 
 8   Distance       13580 non-null  float64
 9   Postcode       13580 non-null  float64
 10  Bedroom2       13580 non-null  float64
 11  Bathroom       13580 non-null  float64
 12  Car            13518 non-null  float64
 13  Landsize       13580 non-null  float64
 14  BuildingArea   7130 non-null   float64
 15  YearBuilt      8205 non-null   float64
 16  CouncilArea    12211 non-null  object 
 17  Lattitude      13580 non-null  float64
 18  Longtitude     13580 non-null  float64
 19  Regionname     13580 non-null  object 
 20  Propertycount  13580 non-null  float64
dtypes: float64(12), int64(1), object(8)
memory usage: 2.2+ MB
None
(13580, 21)
```

In [13]:
```
ml['BuildingArea'].describe()
```

Out[13]:
```
count     7130.000000
mean       151.967650
std        541.014538
min          0.000000
25%         93.000000
50%        126.000000
75%        174.000000
max      44515.000000
Name: BuildingArea, dtype: float64
```

In [14]:
```
print(ml['BuildingArea'].mean())
print(ml['BuildingArea'].median())
```

```
151.96764988779805
126.0
```

In [17]:
```
print(ml.head())
ml['BuildingArea']=ml['BuildingArea'].fillna(ml['BuildingArea'].median())
print(ml.isnull().any())
```

```
       Suburb           Address  Rooms Type      Price Method SellerG  \
0  Abbotsford      85 Turner St      2    h  1480000.0      S  Biggin
1  Abbotsford   25 Bloomburg St      2    h  1035000.0      S  Biggin
2  Abbotsford      5 Charles St      3    h  1465000.0     SP  Biggin
3  Abbotsford  40 Federation La      3    h   850000.0     PI  Biggin
4  Abbotsford       55a Park St      4    h  1600000.0     VB  Nelson

        Date  Distance  Postcode  ...  Bathroom  Car  Landsize  BuildingArea  \
0  3/12/2016       2.5    3067.0  ...       1.0  1.0     202.0         126.0
1  4/02/2016       2.5    3067.0  ...       1.0  0.0     156.0          79.0
2  4/03/2017       2.5    3067.0  ...       2.0  0.0     134.0         150.0
3  4/03/2017       2.5    3067.0  ...       2.0  1.0      94.0         126.0
4  4/06/2016       2.5    3067.0  ...       1.0  2.0     120.0         142.0

   YearBuilt  CouncilArea Lattitude  Longtitude             Regionname  \
0        NaN        Yarra  -37.7996    144.9984  Northern Metropolitan
1     1900.0        Yarra  -37.8079    144.9934  Northern Metropolitan
2     1900.0        Yarra  -37.8093    144.9944  Northern Metropolitan
3        NaN        Yarra  -37.7969    144.9969  Northern Metropolitan
4     2014.0        Yarra  -37.8072    144.9941  Northern Metropolitan

  Propertycount
0        4019.0
1        4019.0
2        4019.0
3        4019.0
4        4019.0

[5 rows x 21 columns]
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               True
Landsize         False
BuildingArea     False
YearBuilt         True
CouncilArea       True
Lattitude        False
Longtitude       False
Regionname       False
Propertycount    False
dtype: bool
```

In [18]:
```
ml.head()
```

Out[18]:

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode	...	Bathroom	Car	Landsize	BuildingArea	YearBuilt	CouncilArea	Lattitude	Longtitude	Regionname	Propertycount
0	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	3/12/2016	2.5	3067.0	...	1.0	1.0	202.0	126.0	NaN	Yarra	-37.7996	144.9984	Northern Metropolitan	4019.0
1	Abbotsford	25 Bloomburg St	2	h	1035000.0	S	Biggin	4/02/2016	2.5	3067.0	...	1.0	0.0	156.0	79.0	1900.0	Yarra	-37.8079	144.9934	Northern Metropolitan	4019.0
2	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	4/03/2017	2.5	3067.0	...	2.0	0.0	134.0	150.0	1900.0	Yarra	-37.8093	144.9944	Northern Metropolitan	4019.0
3	Abbotsford	40 Federation La	3	h	850000.0	PI	Biggin	4/03/2017	2.5	3067.0	...	2.0	1.0	94.0	126.0	NaN	Yarra	-37.7969	144.9969	Northern Metropolitan	4019.0
4	Abbotsford	55a Park St	4	h	1600000.0	VB	Nelson	4/06/2016	2.5	3067.0	...	1.0	2.0	120.0	142.0	2014.0	Yarra	-37.8072	144.9941	Northern Metropolitan	4019.0

5 rows × 21 columns

In [19]:
```
ml['CouncilArea'].unique()
```

Out[19]:
```
array(['Yarra', 'Moonee Valley', 'Port Phillip', 'Darebin', 'Hobsons Bay',
       'Stonnington', 'Boroondara', 'Monash', 'Glen Eira', 'Whitehorse',
       'Maribyrnong', 'Bayside', 'Moreland', 'Manningham', 'Banyule',
       'Melbourne', 'Kingston', 'Brimbank', 'Hume', nan, 'Knox',
       'Maroondah', 'Casey', 'Melton', 'Greater Dandenong', 'Nillumbik',
       'Whittlesea', 'Frankston', 'Macedon Ranges', 'Yarra Ranges',
       'Wyndham', 'Cardinia', 'Unavailable', 'Moorabool'], dtype=object)
```

In [20]:
```
ml['CouncilArea'].value_counts()
```

Out[20]:
```
Moreland              1163
Boroondara            1160
Moonee Valley          997
Darebin                934
Glen Eira              848
Stonnington            719
Maribyrnong            692
Yarra                  647
Port Phillip           628
Banyule                594
Bayside                489
Melbourne              470
Hobsons Bay            434
Brimbank               424
Monash                 333
Manningham             311
Whitehorse             304
Kingston               207
Whittlesea             167
Hume                   164
Wyndham                 86
Knox                    80
Maroondah               80
Melton                  66
Frankston               53
Greater Dandenong       52
Casey                   38
Nillumbik               36
Yarra Ranges            18
Cardinia                 8
Macedon Ranges           7
Unavailable              1
Moorabool                1
Name: CouncilArea, dtype: int64
```

In [ ]:
```

```