In [ ]:
```
import pandas as pd
import numpy as np
```

In [ ]:
```
data = pd.read_csv("melb_data.csv")
```

In [ ]:
```
print(data.isnull())
```

```
       Suburb  Address  Rooms   Type  Price  Method  SellerG   Date  Dist
ance  \
0       False    False  False  False  False   False    False  False     F
alse
1       False    False  False  False  False   False    False  False     F
alse
2       False    False  False  False  False   False    False  False     F
alse
3       False    False  False  False  False   False    False  False     F
alse
4       False    False  False  False  False   False    False  False     F
alse
...       ...      ...    ...    ...    ...     ...      ...    ...
...
13575   False    False  False  False  False   False    False  False     F
alse
13576   False    False  False  False  False   False    False  False     F
alse
13577   False    False  False  False  False   False    False  False     F
alse
13578   False    False  False  False  False   False    False  False     F
alse
13579   False    False  False  False  False   False    False  False     F
alse

       Postcode  ...  Bathroom    Car  Landsize  BuildingArea  YearBuilt
\
0         False  ...     False  False     False          True       True
1         False  ...     False  False     False         False      False
2         False  ...     False  False     False         False      False
3         False  ...     False  False     False          True       True
4         False  ...     False  False     False         False      False
...         ...  ...       ...    ...       ...           ...        ...
13575     False  ...     False  False     False          True      False
13576     False  ...     False  False     False         False      False
13577     False  ...     False  False     False          True      False
13578     False  ...     False  False     False         False      False
13579     False  ...     False  False     False         False      False

       CouncilArea  Lattitude  Longtitude  Regionname  Propertycount
0            False      False       False       False          False
1            False      False       False       False          False
2            False      False       False       False          False
3            False      False       False       False          False
4            False      False       False       False          False
...            ...        ...         ...         ...            ...
13575         True      False       False       False          False
13576         True      False       False       False          False
13577         True      False       False       False          False
13578         True      False       False       False          False
13579         True      False       False       False          False

[13580 rows x 21 columns]
```

In [ ]:

```
data.isnull().sum()
```

Out[ ]:
```
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 [ ]:
```
data.describe()
```

Out[ ]:

	Rooms	Price	Distance	Postcode	Bedroom2	Bat
count	13580.000000	1.358000e+04	13580.000000	13580.000000	13580.000000	13580.0
mean	2.937997	1.075684e+06	10.137776	3105.301915	2.914728	1.5
std	0.955748	6.393107e+05	5.868725	90.676964	0.965921	0.
min	1.000000	8.500000e+04	0.000000	3000.000000	0.000000	0.0
25%	2.000000	6.500000e+05	6.100000	3044.000000	2.000000	1.0
50%	3.000000	9.030000e+05	9.200000	3084.000000	3.000000	1.0
75%	3.000000	1.330000e+06	13.000000	3148.000000	3.000000	2.0
max	10.000000	9.000000e+06	48.100000	3977.000000	20.000000	8.0

In [ ]:
```
data.dtypes
```

Out[ ]:
```
Suburb            object
Address           object
Rooms              int64
Type              object
Price            float64
Method            object
SellerG           object
Date              object
Distance         float64
Postcode         float64
Bedroom2         float64
Bathroom         float64
Car              float64
Landsize         float64
BuildingArea     float64
YearBuilt        float64
CouncilArea       object
Lattitude        float64
Longtitude       float64
Regionname        object
Propertycount    float64
dtype: object
```

In [ ]:
```
data.shape
```

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

In [ ]:
```
data.Rooms.value_counts
```

Out[ ]:
```
<bound method IndexOpsMixin.value_counts of 0        2
1        2
2        3
3        3
4        4
        ..
13575    4
13576    3
13577    3
13578    4
13579    4
Name: Rooms, Length: 13580, dtype: int64>
```

In [ ]:
```
data.head(100)
```

Out[ ]:

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distar
0	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	3/12/2016	
1	Abbotsford	25 Bloomburg St	2	h	1035000.0	S	Biggin	4/02/2016	
2	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	4/03/2017	
3	Abbotsford	40 Federation La	3	h	850000.0	PI	Biggin	4/03/2017	
4	Abbotsford	55a Park St	4	h	1600000.0	VB	Nelson	4/06/2016	
...	...	...	...	...	...	...	...	...	
95	Albert Park	4a Gatehouse La	3	h	1370000.0	S	Greg	12/06/2016	
96	Albert Park	60 Brooke St	2	h	1000000.0	S	Cayzer	12/11/2016	
97	Albert Park	70 Barrett St	3	h	2575000.0	S	Greg	15/10/2016	
98	Albert Park	65 Graham St	2	h	1322500.0	S	Greg	15/10/2016	
99	Albert Park	364 Montague St	2	h	1562500.0	S	Greg	15/10/2016	

100 rows × 21 columns

In [ ]:

```
print("\n# Filling the missing values with default 0")
for i in data:
    data[i].fillna(0,inplace=True)
for i in data:
    print(f"{i}:{any(data[i].isna())}")
```

```
# Filling the missing values with default 0
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
Lattitude:False
Longtitude:False
Regionname:False
Propertycount:False
```

In [ ]:

```
# Change variable to appropriate type
print("\n# Change variable to appropriate type")
for i in "Postcode Bedroom2 Bathroom Car YearBuilt".split():
    data[i] = data[i].astype('int64')
    print(i,data[i].dtype)
```

```
# Change variable to appropriate type
Postcode int64
Bedroom2 int64
Bathroom int64
Car int64
YearBuilt int64
```

In [ ]:

```
# Change categorical to quantitative
print("\n# Change categorical to quantitative")
data=pd.get_dummies(data,columns=['Type'])
print(data)
```

```
# Change categorical to quantitative
              Suburb           Address  Rooms      Price Method   SellerG
\
0         Abbotsford      85 Turner St      2  1480000.0      S    Biggin
1         Abbotsford   25 Bloomburg St      2  1035000.0      S    Biggin
2         Abbotsford      5 Charles St      3  1465000.0     SP    Biggin
3         Abbotsford  40 Federation La      3   850000.0     PI    Biggin
4         Abbotsford       55a Park St      4  1600000.0     VB    Nelson
...              ...               ...    ...        ...    ...       ...
13575  Wheelers Hill      12 Strada Cr      4  1245000.0      S     Barry
13576   Williamstown     77 Merrett Dr      3  1031000.0     SP  Williams
13577   Williamstown       83 Power St      3  1170000.0      S     Raine
13578   Williamstown      96 Verdon St      4  2500000.0     PI   Sweeney
13579     Yarraville        6 Agnes St      4  1285000.0     SP   Village
```

```
             Date  Distance  Postcode  Bedroom2  ...  BuildingArea  YearB
uilt  \
0        3/12/2016       2.5      3067         2  ...           0.0
0
1        4/02/2016       2.5      3067         2  ...          79.0
1900
2        4/03/2017       2.5      3067         3  ...         150.0
1900
3        4/03/2017       2.5      3067         3  ...           0.0
0
4        4/06/2016       2.5      3067         3  ...         142.0
2014
...            ...       ...       ...       ...  ...           ...
...
13575  26/08/2017      16.7      3150         4  ...           0.0
1981
13576  26/08/2017       6.8      3016         3  ...         133.0
1995
13577  26/08/2017       6.8      3016         3  ...           0.0
1997
13578  26/08/2017       6.8      3016         4  ...         157.0
1920
13579  26/08/2017       6.3      3013         4  ...         112.0
1920

       CouncilArea  Lattitude  Longtitude                 Regionname  \
0            Yarra  -37.79960   144.99840       Northern Metropolitan
1            Yarra  -37.80790   144.99340       Northern Metropolitan
2            Yarra  -37.80930   144.99440       Northern Metropolitan
3            Yarra  -37.79690   144.99690       Northern Metropolitan
4            Yarra  -37.80720   144.99410       Northern Metropolitan
...            ...        ...         ...                         ...
13575            0  -37.90562   145.16761  South-Eastern Metropolitan
13576            0  -37.85927   144.87904        Western Metropolitan
13577            0  -37.85274   144.88738        Western Metropolitan
13578            0  -37.85908   144.89299        Western Metropolitan
13579            0  -37.81188   144.88449        Western Metropolitan

       Propertycount  Type_h Type_t  Type_u
0             4019.0       1      0       0
1             4019.0       1      0       0
2             4019.0       1      0       0
3             4019.0       1      0       0
4             4019.0       1      0       0
...              ...     ...    ...     ...
13575         7392.0       1      0       0
13576         6380.0       1      0       0
13577         6380.0       1      0       0
13578         6380.0       1      0       0
13579         6543.0       1      0       0

[13580 rows x 23 columns]
```