Results and Graphs (Housing Market Prediction)

The numberings are according to the Code sections.

Section 1.

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode	Bedroom2	Bathroom	Car	Landsize	BuildingArea	YearBuilt	CouncilArea	Lattitude	Longtit
0	Abbotsford	68 Studley St	2	h	NaN	SS	Jellis	3/09/2016	2.5	3067.0	2.0	1.0	1.0	126.0	NaN	NaN	Yarra City Council	-37.8014	144.9
1	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	3/12/2016	2.5	3067.0	2.0	1.0	1.0	202.0	NaN	NaN	Yarra City Council	-37.7996	144.9
2	Abbotsford	25 Bloomburg St	2	h	1035000.0	S	Biggin	4/02/2016	2.5	3067.0	2.0	1.0	0.0	156.0	79.0	1900.0	Yarra City Council	-37.8079	144.9
3	Abbotsford	18/659 Victoria St	3	u	NaN	VB	Rounds	4/02/2016	2.5	3067.0	3.0	2.0	1.0	0.0	NaN	NaN	Yarra City Council	-37.8114	145.0
4	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	4/03/2017	2.5	3067.0	3.0	2.0	0.0	134.0	150.0	1900.0	Yarra City Council	-37.8093	144.9

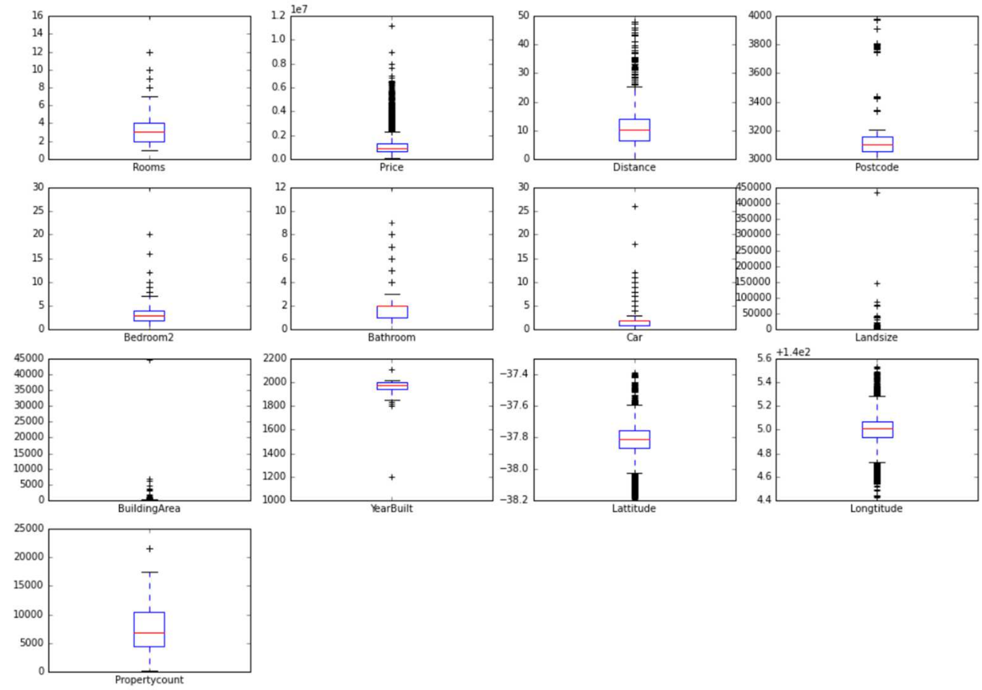
Displaying the first 5 rows of the Dataset after importing it.

Section 2.

```
[2]  df.isnull().sum()

     Suburb               0
     Address              0
     Rooms                0
     Type                 0
     Price             7610
     Method               0
     SellerG              0
     Date                 0
     Distance             1
     Postcode             1
     Bedroom2          8217
     Bathroom          8226
     Car               8728
     Landsize         11810
     BuildingArea     21115
     YearBuilt        19306
     CouncilArea          3
     Lattitude         7976
     Longtitude        7976
     Regionname           3
     Propertycount        3
     dtype: int64
```

The above shown values are the Sum of the NULL values in the respective attributes.



The Graph shown above are the outliers present in each different attribute required for the further implementations.

Section 3.

```
6] df_train.head(5)
```

	Suburb	Rooms	Method	Date	Bathroom	Car	Landsize	BuildingArea	YearBuilt	Lattitude	Longtitude	Distance	Price	year	month
0	Abbotsford	2	S	2016-04-02	1.0	0.0	156.0	79.0	1900.0	-37.8079	144.9934	2.5	1035000.0	2016	4
1	Abbotsford	3	SP	2017-04-03	2.0	0.0	134.0	150.0	1900.0	-37.8093	144.9944	2.5	1465000.0	2017	4
2	Abbotsford	4	VB	2016-04-06	1.0	2.0	120.0	142.0	2014.0	-37.8072	144.9941	2.5	1600000.0	2016	4
3	Abbotsford	3	S	2016-07-05	2.0	0.0	245.0	210.0	1910.0	-37.8024	144.9993	2.5	1876000.0	2016	7
4	Abbotsford	2	S	2016-08-10	1.0	2.0	256.0	107.0	1890.0	-37.8060	144.9954	2.5	1636000.0	2016	8

It is the result after training the data to prepare for the Random Forest Regression.

Section 4.

```
10] for i in range(4):
        print(suburban[i], '\n', replacement[suburban[i]], '\n')

    Reservoir
     {'Rooms': [3, 4, 0], 'Bathroom': [1, 3, 0], 'Car': [1, 3, 0], 'Landsize': [428, 1261, 0], 'BuildingArea': [120, 207, 11], 'YearBuilt': [1974.46, 2085.0, 1869.0], 'Lattitude': [-37.71,

    Richmond
     {'Rooms': [2, 4, 0], 'Bathroom': [1, 3, 0], 'Car': [1, 1, 1], 'Landsize': [633, 546, 0], 'BuildingArea': [101, 202, 0], 'YearBuilt': [1955.01, 2162.5, 1742.5], 'Lattitude': [-37.82, -

    Brunswick
     {'Rooms': [3, 4, 0], 'Bathroom': [1, 3, 0], 'Car': [1, 1, 1], 'Landsize': [341, 722, 0], 'BuildingArea': [129, 250, 0], 'YearBuilt': [1939.55, 2060.75, 1818.75], 'Lattitude': [-37.77,

    Bentleigh East
     {'Rooms': [3, 5, 1], 'Bathroom': [2, 3, 0], 'Car': [2, 3, 0], 'Landsize': [476, 1096, 0], 'BuildingArea': [157, 291, 14], 'YearBuilt': [1974.16, 2084.375, 1869.375], 'Lattitude': [-37
```

Result of the creating replacement values for the NULL attributes.

Section 5.

```
df_train.head(5)
```

	Suburb	Rooms	Method	Date	Bathroom	Car	Landsize	BuildingArea	YearBuilt	Lattitude	Longtitude	Distance	Price	year	month
0	Abbotsford	2	S	2016-04-02	1.0	0.0	156.0	79.0	1900.0	-37.8079	144.9934	2.5	1035000.0	2016	4
1	Abbotsford	3	SP	2017-04-03	2.0	0.0	134.0	150.0	1900.0	-37.8093	144.9944	2.5	1465000.0	2017	4
2	Abbotsford	4	VB	2016-04-06	1.0	1.0	120.0	142.0	2014.0	-37.8072	144.9941	2.5	1600000.0	2016	4
3	Abbotsford	3	S	2016-07-05	2.0	0.0	245.0	194.0	1910.0	-37.8024	144.9993	2.5	1815000.0	2016	7
4	Abbotsford	2	S	2016-08-10	1.0	1.0	256.0	107.0	1890.0	-37.8060	144.9954	2.5	1636000.0	2016	8

Here we have edited all the outliers on the training data

Section 6.

```
] new_train_x
```

```
array([[0.2       , 0.        , 0.        , ..., 0.0326087 , 0.        ,
        0.27272727],
       [0.4       , 0.2       , 0.        , ..., 0.0326087 , 0.5       ,
        0.27272727],
       [0.6       , 0.        , 0.125     , ..., 0.0326087 , 0.        ,
        0.27272727],
       ...,
       [0.2       , 0.2       , 0.125     , ..., 0.11521739, 1.        ,
        0.09090909],
       [0.2       , 0.        , 0.25      , ..., 0.11521739, 1.        ,
        0.09090909],
       [0.2       , 0.        , 0.        , ..., 0.11521739, 1.        ,
        0.09090909]])
```

```
] new_train_y
```

```
array([[0.1654617 ],
       [0.24416583],
       [0.26887526],
       ...,
       [0.13855587],
       [0.10506086],
       [0.16271621]])
```

```
error = np.sqrt(mean_squared_error(y_pred, y_valid))
error
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher
  if __name__ == '__main__':
0.04656663666167402
```

```
[87] error1= mean_absolute_percentage_error(y_valid, y_pred)
     error1*100
```

```
16.86079176678108
```

These are the MAPE and MSE values. Since the MSE and MAPE values are 0.04657 and 16.8% respectively then the prediction was not bad so we can use this model to fill the NULL price values with approximated values.

Section 8.

```
] error1= mean_absolute_percentage_error(y_valid, y_pred)
  error1*100
```

```
17.443057094149967
```

The MAPE shown above indicates that we have 17.4% of error in our train model of filling the NULL price values with approximate values. Since the error in only about 17.4% this model has worked as intended. And the data shown on the below table are the result of the Random Forest model.

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode	Bathroom	Car	Landsize	BuildingArea	YearBuilt	CouncilArea	Lattitude	Longtitude	Regior
0	Abbotsford	68 Studley St	2	House	1171442.50	SS	Jellis	2016-03-09	2.5	3067.0	1.0	1.0	126.0	108.0	1945.41	Yarra City Council	-37.8014	144.9958	Nor Metrop
1	Abbotsford	85 Turner St	2	House	1480000.00	S	Biggin	2016-03-12	2.5	3067.0	1.0	1.0	202.0	108.0	1945.41	Yarra City Council	-37.7996	144.9984	Nor Metrop
2	Abbotsford	25 Bloomburg St	2	House	1035000.00	S	Biggin	2016-04-02	2.5	3067.0	1.0	0.0	156.0	79.0	1900.00	Yarra City Council	-37.8079	144.9934	Nor Metrop
3	Abbotsford	18/659 Victoria St	3	Duplex	1067033.75	VB	Rounds	2016-04-02	2.5	3067.0	2.0	1.0	0.0	108.0	1945.41	Yarra City Council	-37.8114	145.0012	Nor Metrop
4	Abbotsford	5 Charles St	3	House	1465000.00	SP	Biggin	2017-04-03	2.5	3067.0	2.0	0.0	134.0	150.0	1900.00	Yarra City Council	-37.8093	144.9944	Nor Metrop

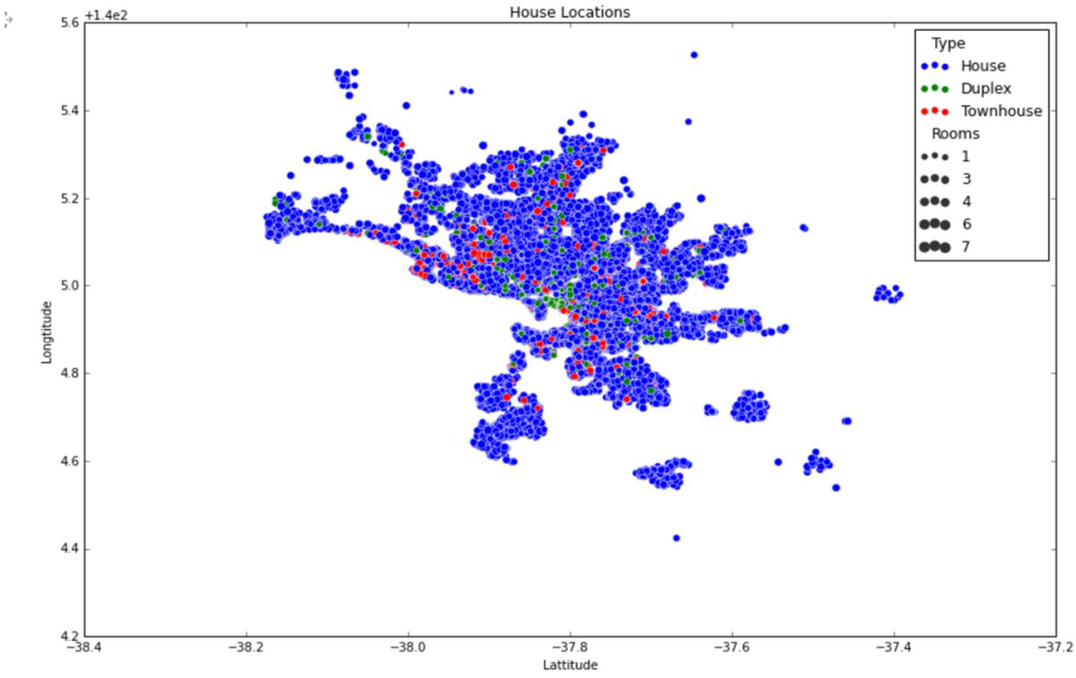
Section 9.

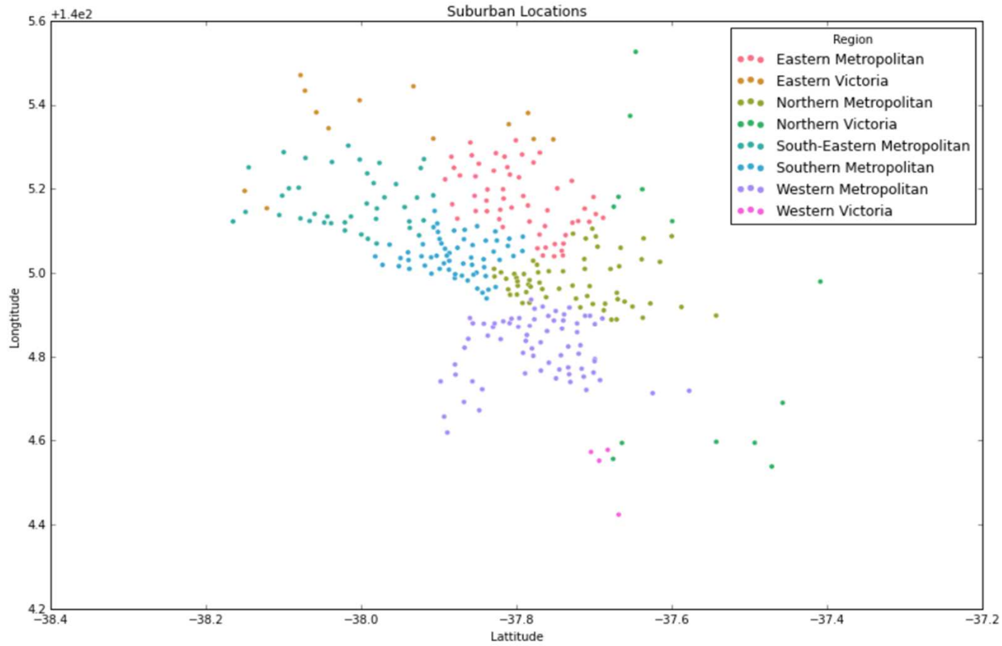
9.1 : The Price fluctuations of different type of properties listed.

As in the above graph, we can say that the Houses are the most expensive ones, then Townhouse, and then Duplex. The price of both Duplex and Townhouse were quite stable and consistent throughout the time period.

We can say that along 2017 the price is quite unpredictable for all the attributes. Also, at every beginning of each year, the prices of the properties decline.

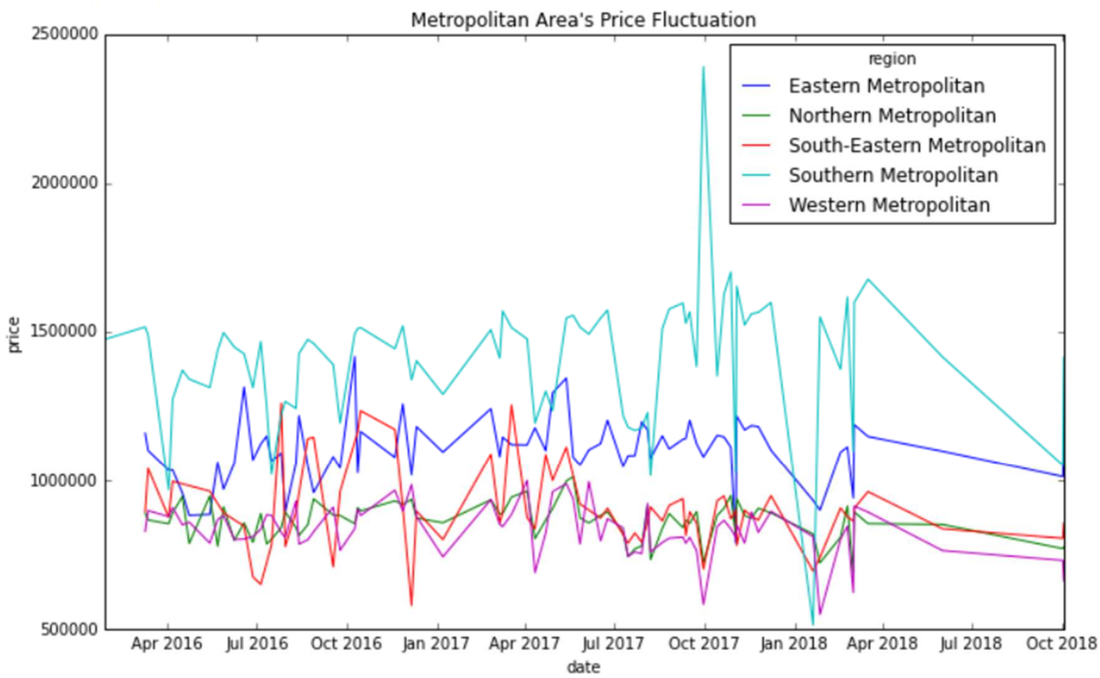
9.2 : The geographical mapping

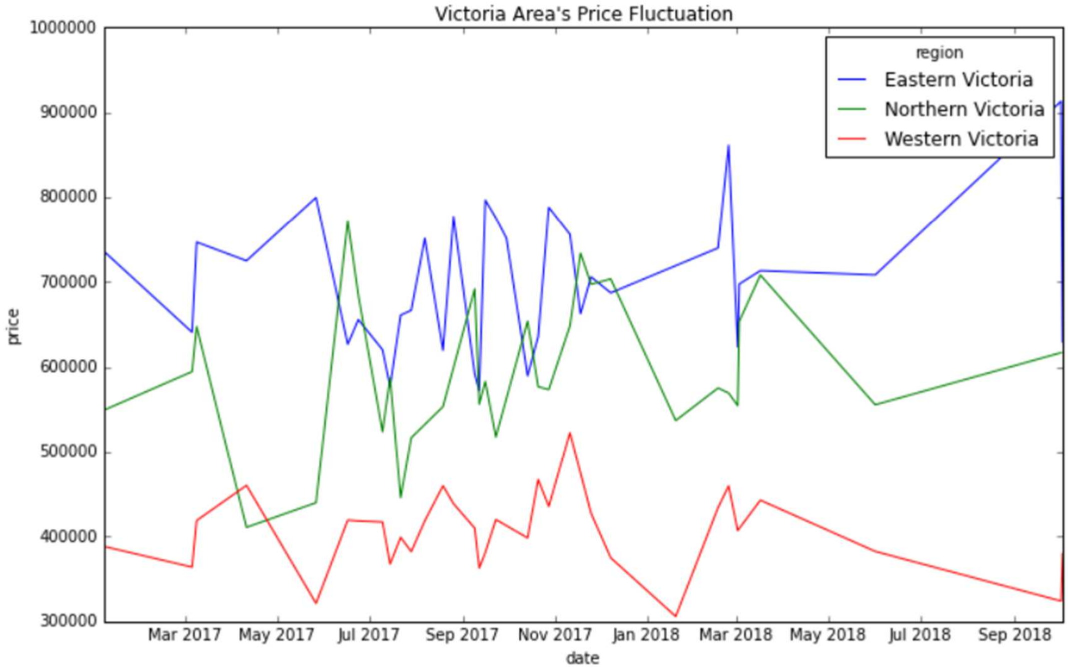




The Graphs shown above are the Housing Locations and the Suburb Locations on the geographical map. If we look at the 1st graph we can see that there are more number of house than duplex or Townhouses available in Melbourne. And in the 2nd graph we can see the representation of the number of duplex and townhouses in each region.

9.3 : The Price fluctuations in different regions





As we can see in the 1st graph, The prices of the properties at Southern Metropolitan spiked around October - November 2017 which means Buying property around that time isn't really a good idea and the prices of the properties at South-Eastern Metropolitan got lower and was similar in range as Northern and Western Metropolitan starting from June - July 2017.

In the 2nd graph shown above we can see there a more fluctuations between July to November 2017in the Victoria area of Melbourne

In Conclusion:

The most important factor that contributes to the price of a property is its distance to the city center and the area it is located at. The price of the property always declines around January or February of each year in Melbourne.

If you are looking for a cheap property located close to Melbourne City Business District (CBD), Western Metropolitan is the cheapest region. But if you are looking for the cheapest property in Melbourne and do not really have a concern about its distance to CBD, then Western Victoria will be a suited place..