Melbourne Housing Data

Appendix

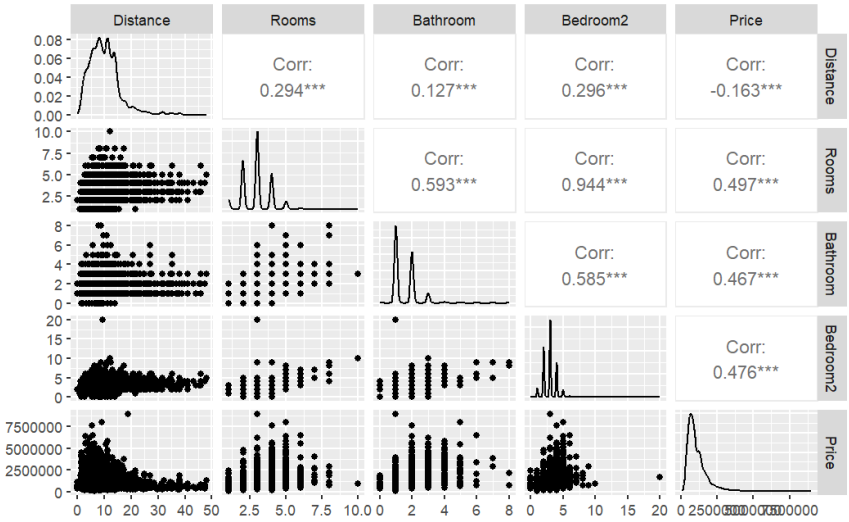
Appendix -1

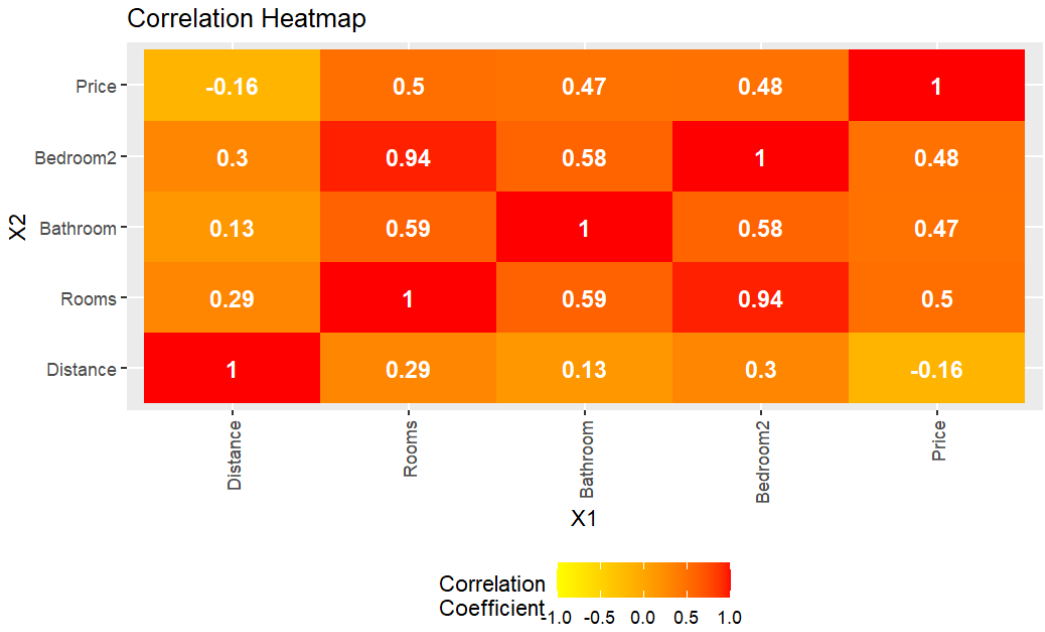
```
'data.frame':  13580 obs. of  21 variables:
 $ Suburb       : Factor w/ 314 levels "Abbotsford","Aberfeldie",..: 1 1 1 1
1 1 1 1 1 1 ...
 $ Address      : chr  "85 Turner St" "25 Bloomburg St" "5 Charles St" "40
Federation La" ...
 $ Rooms        : int  2 2 3 3 4 2 3 2 1 2 ...
 $ Type         : Factor w/ 3 levels "h","t","u": 1 1 1 1 1 1 1 1 3 1 ...
 $ Price        : num  1480000 1035000 1465000 850000 1600000 ...
 $ Method       : Factor w/ 5 levels "PI","S","SA",..: 2 2 4 1 5 2 2 2 2 2
...
 $ SellerG      : Factor w/ 268 levels "@Realty","Abercromby's",..: 24 24 24
24 165 114 165 165 24 24 ...
 $ Date         : chr  "3/12/2016" "4/02/2016" "4/03/2017" "4/03/2017" ...
 $ Distance     : num  2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 ...
 $ Postcode     : num  3067 3067 3067 3067 3067 ...
 $ Bedroom2     : num  2 2 3 3 3 2 4 2 1 3 ...
 $ Bathroom     : num  1 1 2 2 1 1 2 1 1 1 ...
 $ Car          : num  1 0 0 1 2 0 0 2 1 2 ...
 $ Landsize     : num  202 156 134 94 120 181 245 256 0 220 ...
 $ BuildingArea : num  NA 79 150 NA 142 NA 210 107 NA 75 ...
 $ YearBuilt    : num  NA 1900 1900 NA 2014 ...
 $ CouncilArea  : Factor w/ 34 levels "","Banyule","Bayside",..: 33 33 33 33
33 33 33 33 33 33 ...
 $ Lattitude    : num  -37.8 -37.8 -37.8 -37.8 -37.8 ...
 $ Longtitude   : num  145 145 145 145 145 ...
 $ Regionname   : Factor w/ 8 levels "Eastern Metropolitan",..: 3 3 3 3 3 3 3
3 3 3 ...
 $ Propertycount: num  4019 4019 4019 4019 4019 ...NULL
```

No cleaning was required for the set of variables I choose as there were no inconsistent/ missing values for the filtered dataset I am using.

Appendix -2

The correlation plots/ significance used for picking the initial predictors.

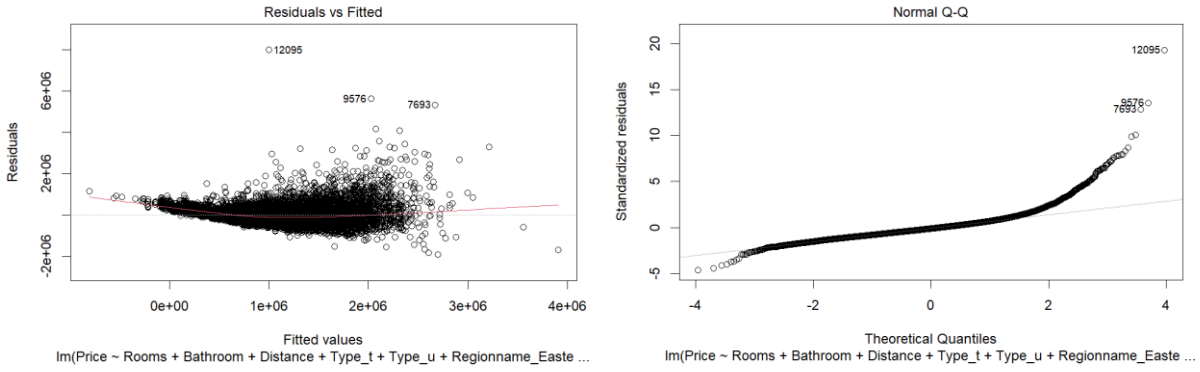


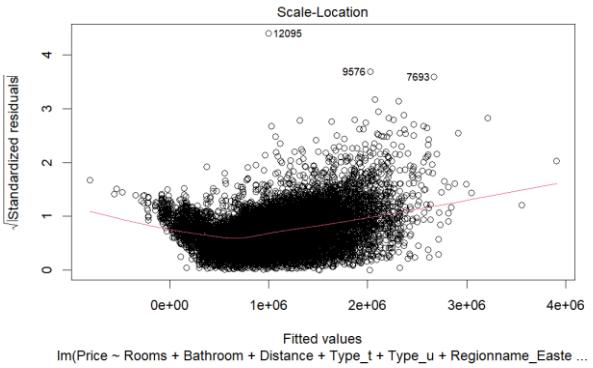


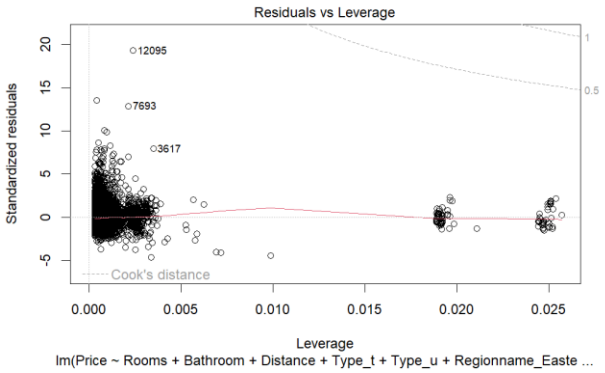
Appendix -3

The residual plots of the 3 models for testing the assumptions for multiple linear regression

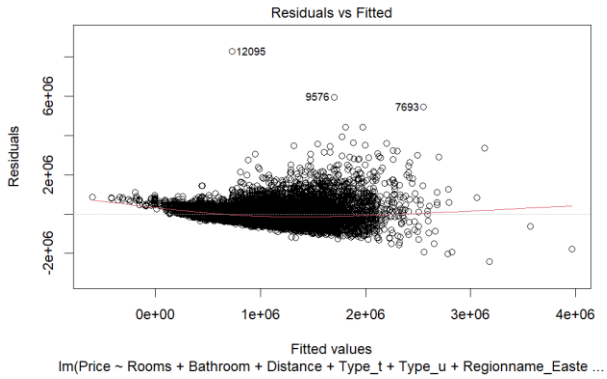
Model 2

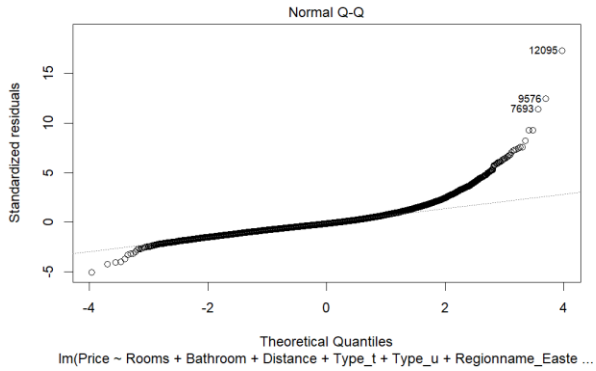


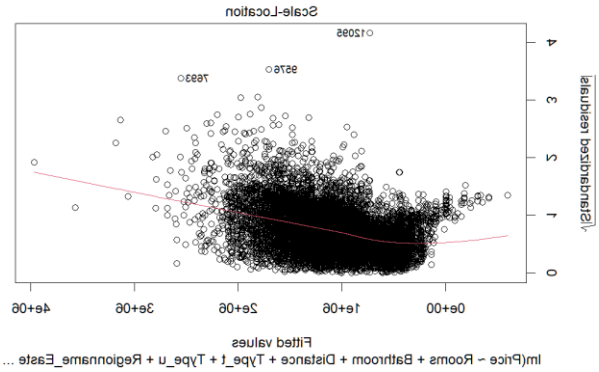

Scale-Location
√|Standardized residuals|
12095
9576
7693
0e+00
1e+06
2e+06
3e+06
4e+06
Fitted values
lm(Price ~ Rooms + Bathroom + Distance + Type_t + Type_u + Regionname_Easte ...

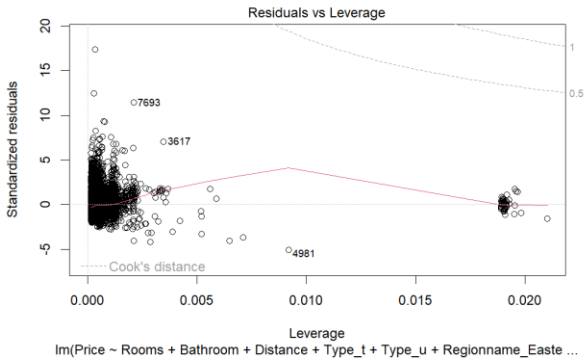

Residuals vs Leverage
Standardized residuals
12095
7693
3617
Cook's distance
0.000
0.005
0.010
0.015
0.020
0.025
Leverage
lm(Price ~ Rooms + Bathroom + Distance + Type_t + Type_u + Regionname_Easte ...

Model 1

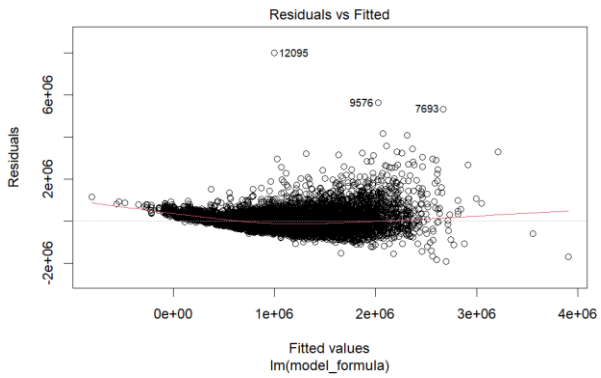

Residuals vs Fitted
Residuals
12095
9576
7693
-2e+06
2e+06
6e+06
0e+00
1e+06
2e+06
3e+06
4e+06
Fitted values
lm(Price ~ Rooms + Bathroom + Distance + Type_t + Type_u + Regionname_Easte ...

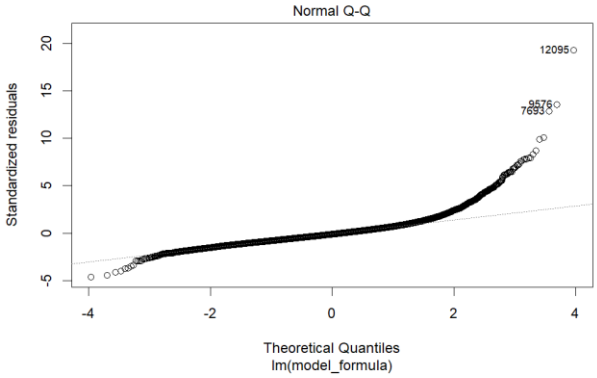

Normal Q-Q
Standardized residuals
12095
9576
7693
-4
-2
0
2
4
Theoretical Quantiles
lm(Price ~ Rooms + Bathroom + Distance + Type_t + Type_u + Regionname_Easte ...

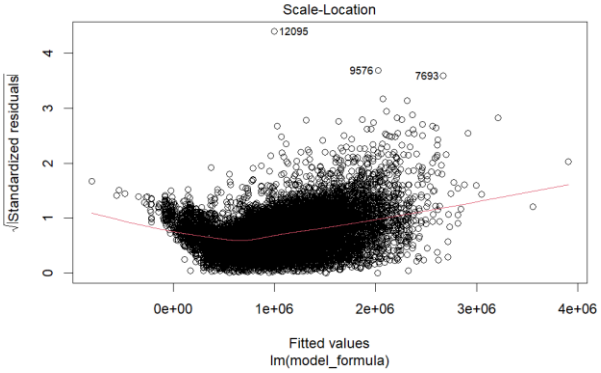

Scale-Location
12095
9576
7693
0e+00
1e+06
2e+06
3e+06
4e+06
Fitted values
lm(Price ~ Rooms + Bathroom + Distance + Type_t + Type_u + Regionname_Easte ...

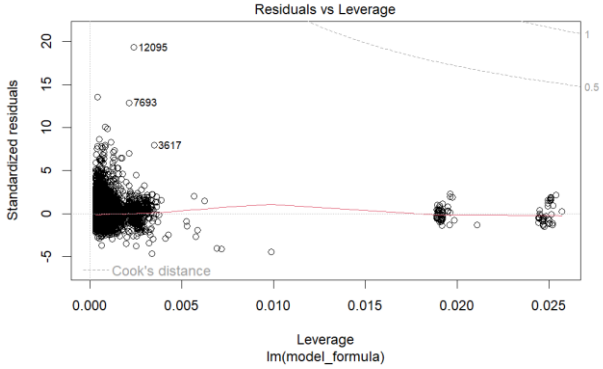

Residuals vs Leverage
Standardized residuals
7693
3617
4981
Cook's distance
0.000
0.005
0.010
0.015
0.020
Leverage
lm(Price ~ Rooms + Bathroom + Distance + Type_t + Type_u + Regionname_Easte ...

Model 3


Residuals vs Fitted
12095
9576
7693
Residuals
-2e+06
2e+06
6e+06
0e+00
1e+06
2e+06
3e+06
4e+06
Fitted values
lm(model_formula)


Normal Q-Q
12095
9576
7693
Standardized residuals
-5
0
5
10
15
20
-4
-2
0
2
4
Theoretical Quantiles
lm(model_formula)


Scale-Location
12095
9576
7693
√|Standardized residuals|
0
1
2
3
4
0e+00
1e+06
2e+06
3e+06
4e+06
Fitted values
lm(model_formula)


Residuals vs Leverage
12095
7693
3617
Standardized residuals
-5
0
5
10
15
20
1
0.5
Cook's distance
0.000
0.005
0.010
0.015
0.020
0.025
Leverage
lm(model_formula)