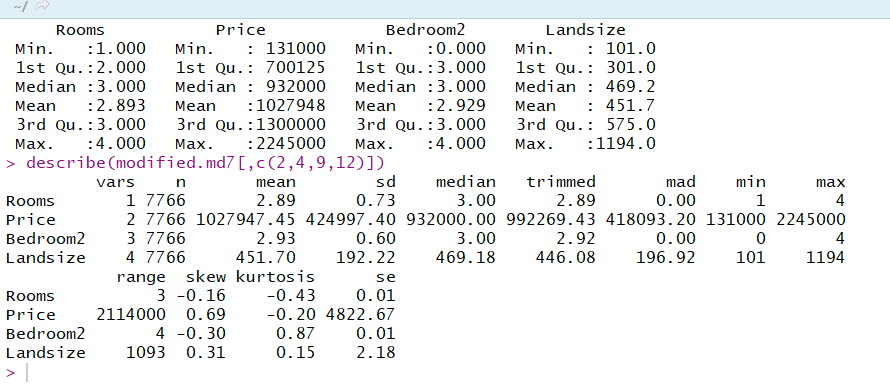
Housing Project Segmentation and Recomender System Analysis

1. What did you understand about univariate analysis?

a)

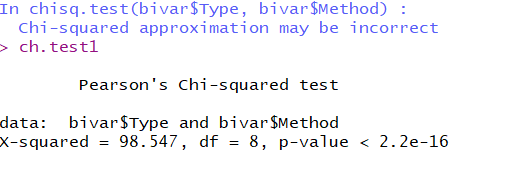


If you want the number of houses in your rooms to be 3 and if you want the bedrooms to be 3 and if you want to have the landsize to be 451, then you must have the price of 1027948.

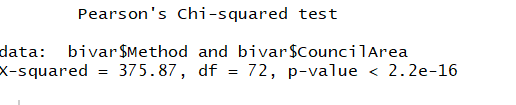
2.what is the results for anova and chi square?

a)

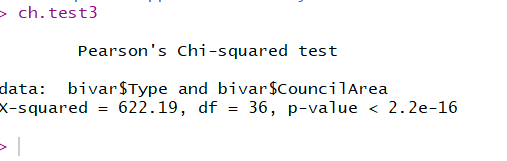
i)**chi-square between type and method**:p value is less than 0.05 which is good.



ii)**chi-square between council area and method**:p value is less than 0.05 which is good.

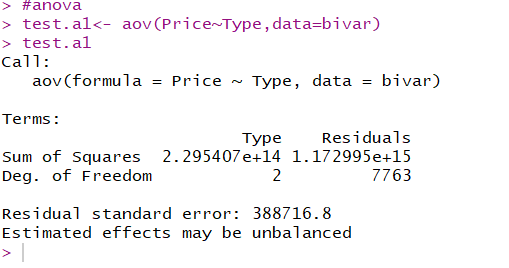


iii) **chi-square between council area and type** :p value is less than 0.05 which is good.

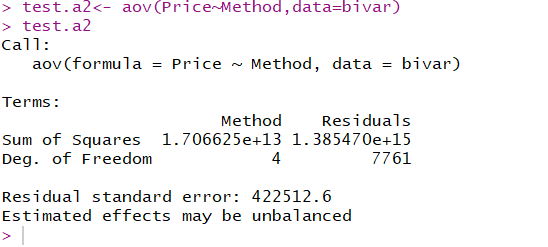


Now anova:

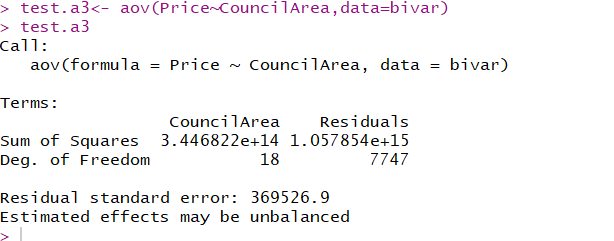
**i)Between price and Type:**



**ii)Between price and method:**



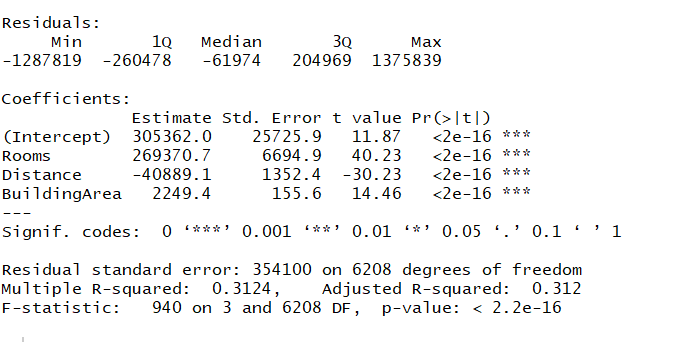
**iii)Between price and council area:**



3.What are the important variables for linear regression? What is R2 value?

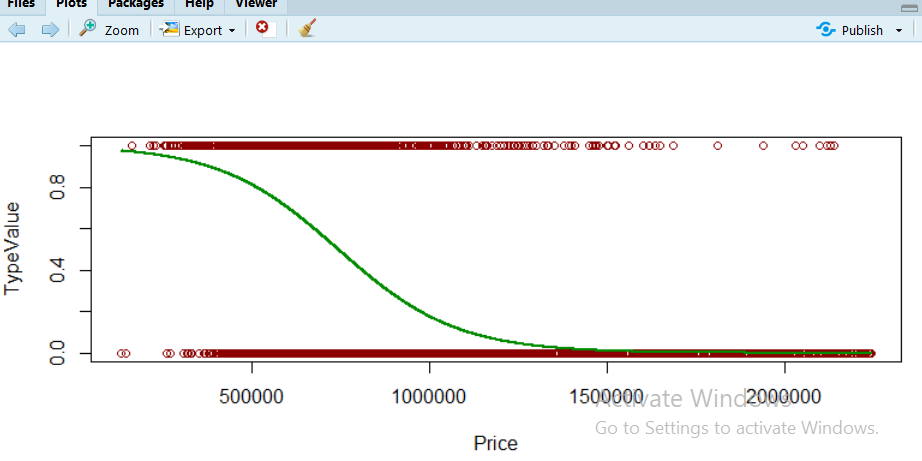
a) Rooms, distance and building area are most important variables for logistic regression.

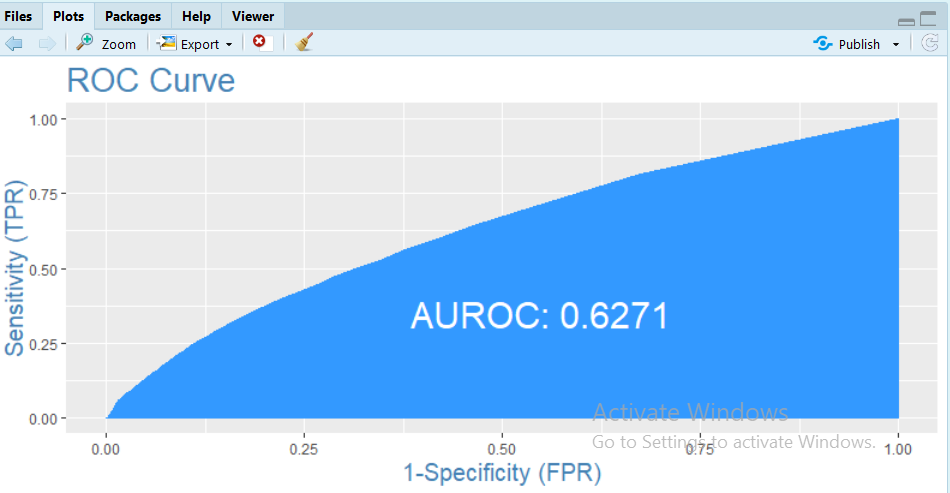
The adjusted R2 value is 31 percent



4. what is your inference on logistic regression? What is AUROC? What are the important variables for prediction?

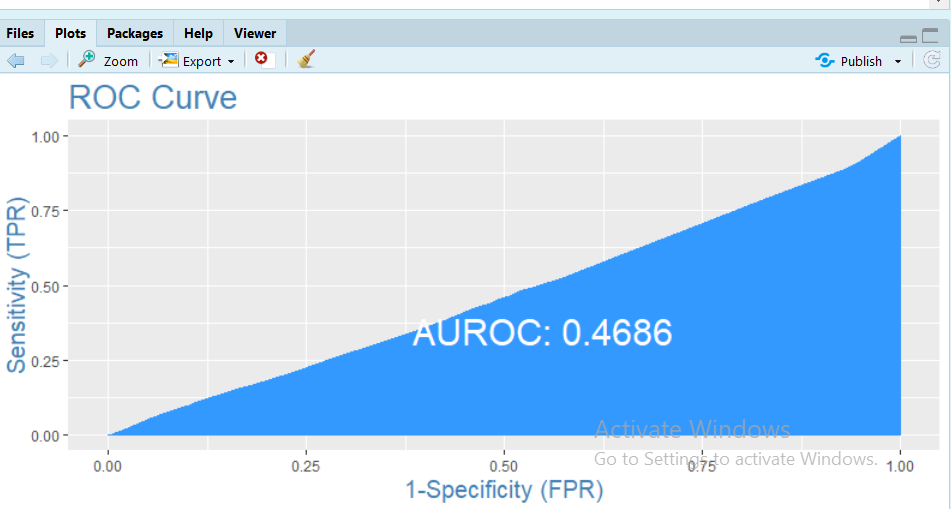
a)As price increases, we can predict the type of the house is u upto 62% for hu type.

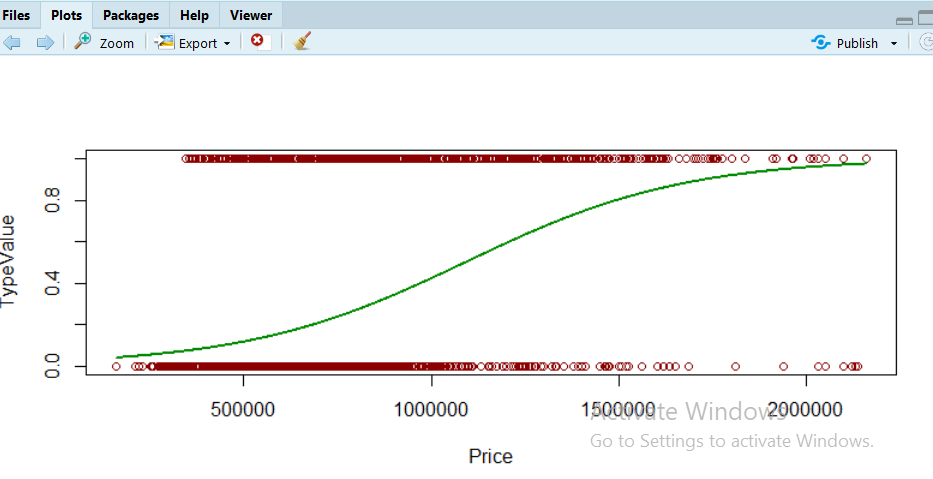




**For UT:**

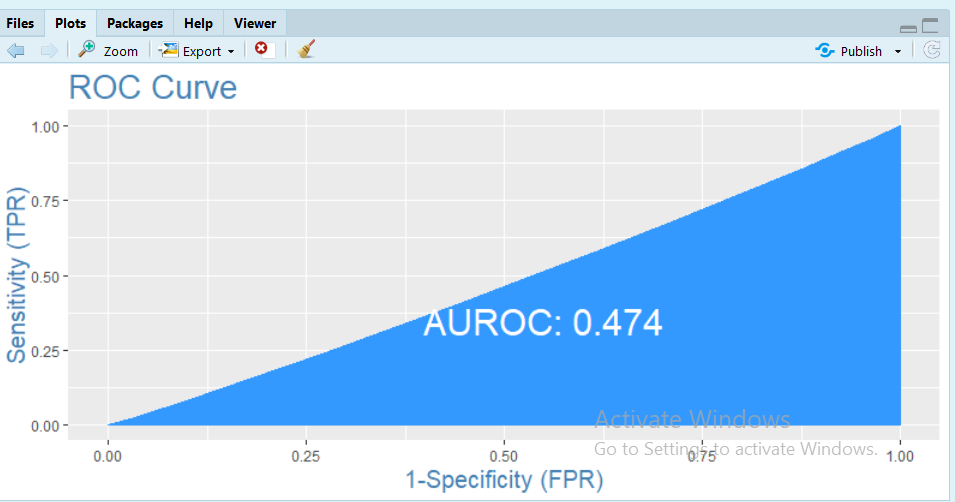
As price increases, we can predict the type of the house is t upto 46% for ut type.

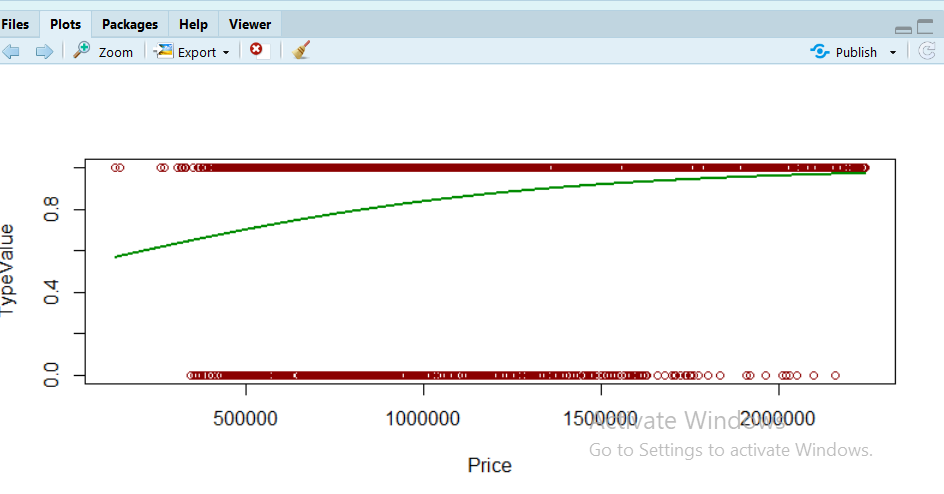




**For TH:**

AS price increases, the type of the house is h upto 47%





Price and rooms have effect on the type of the house.

5.what is your svm model inference? What is your conclusion?

a)

**Based on landsize and building area:**

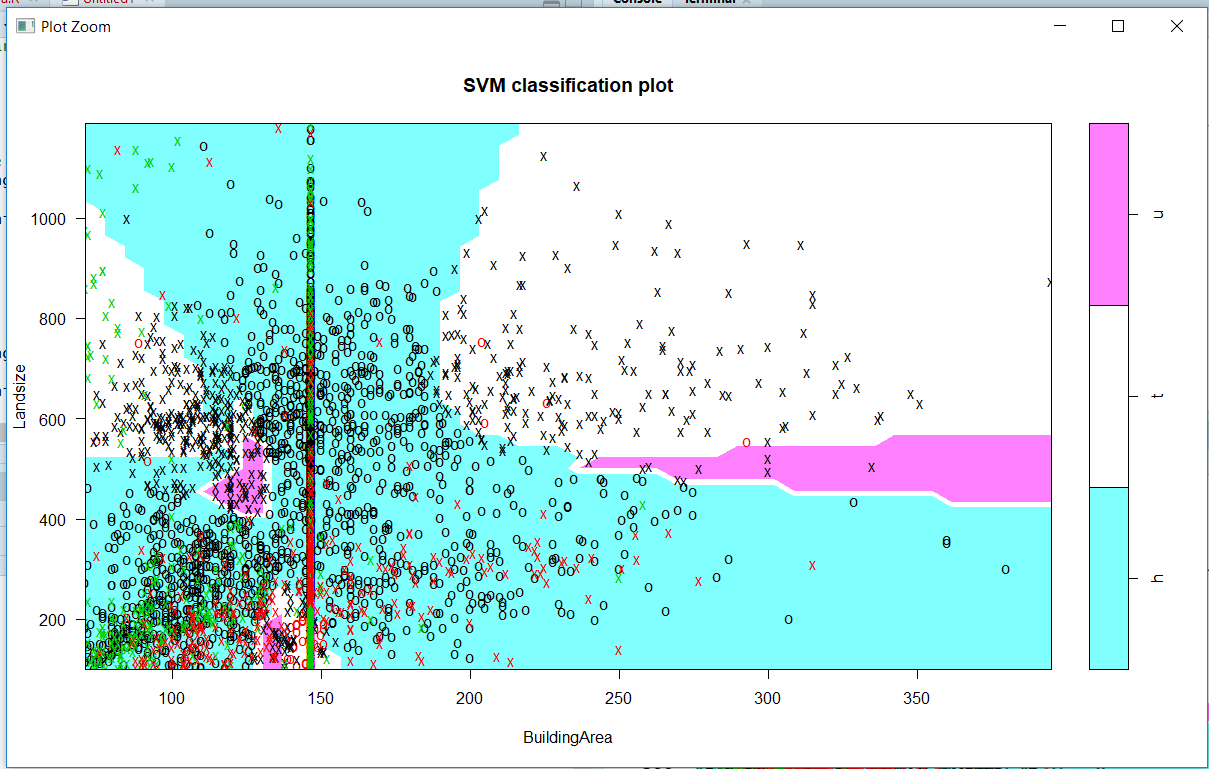
i)if the building area is less than 120 and landsize is less than 400, then we can say type of house is h.

ii) if the building area is greater than 160 and landsize is less than 400, then we can say type of house is h.

iii) if the building area is greater than 130 and landsize is less than 400, then we can say type of house is h.

iv) if the building area is greater than 200 and landsize is in between 500 and 600, then we can say type of house is u.

v) if the building area is greater than 250 and landsize is greater than 600, then we can say type of house is t.

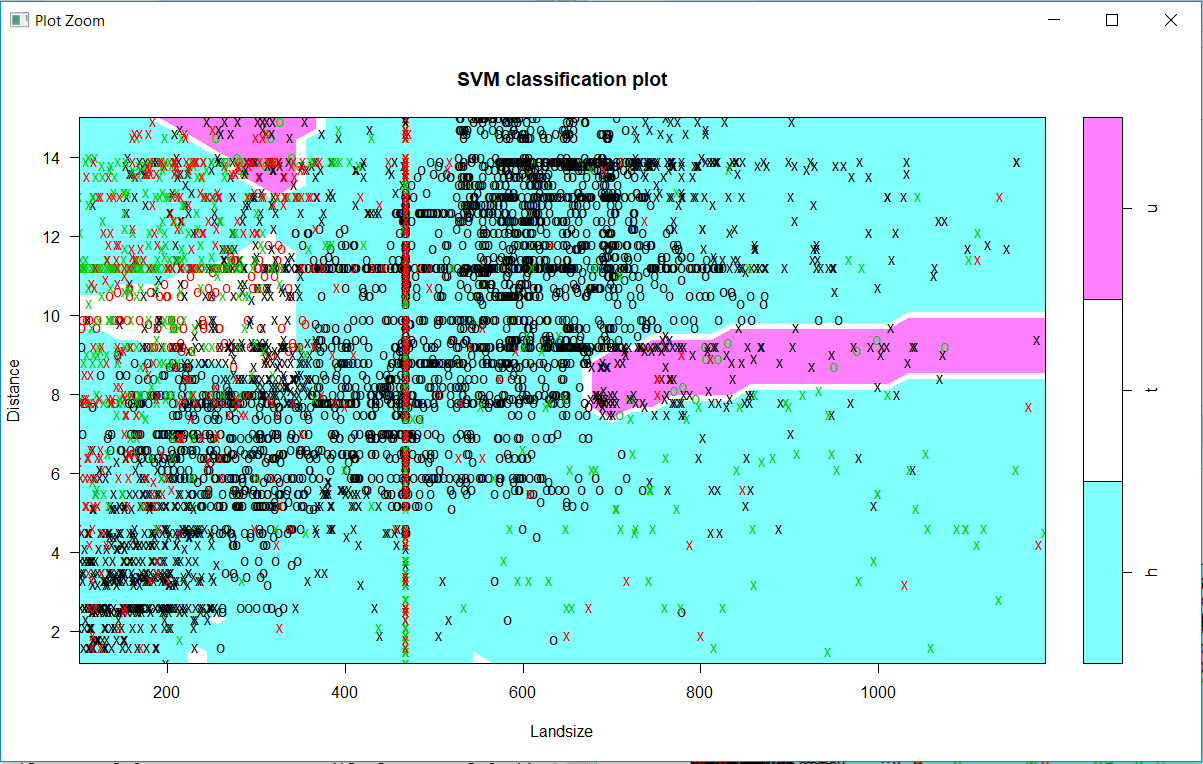


**On distance and building area:**

i)if the building area is greater than 150 and at any distance, then we can say type of house is h.

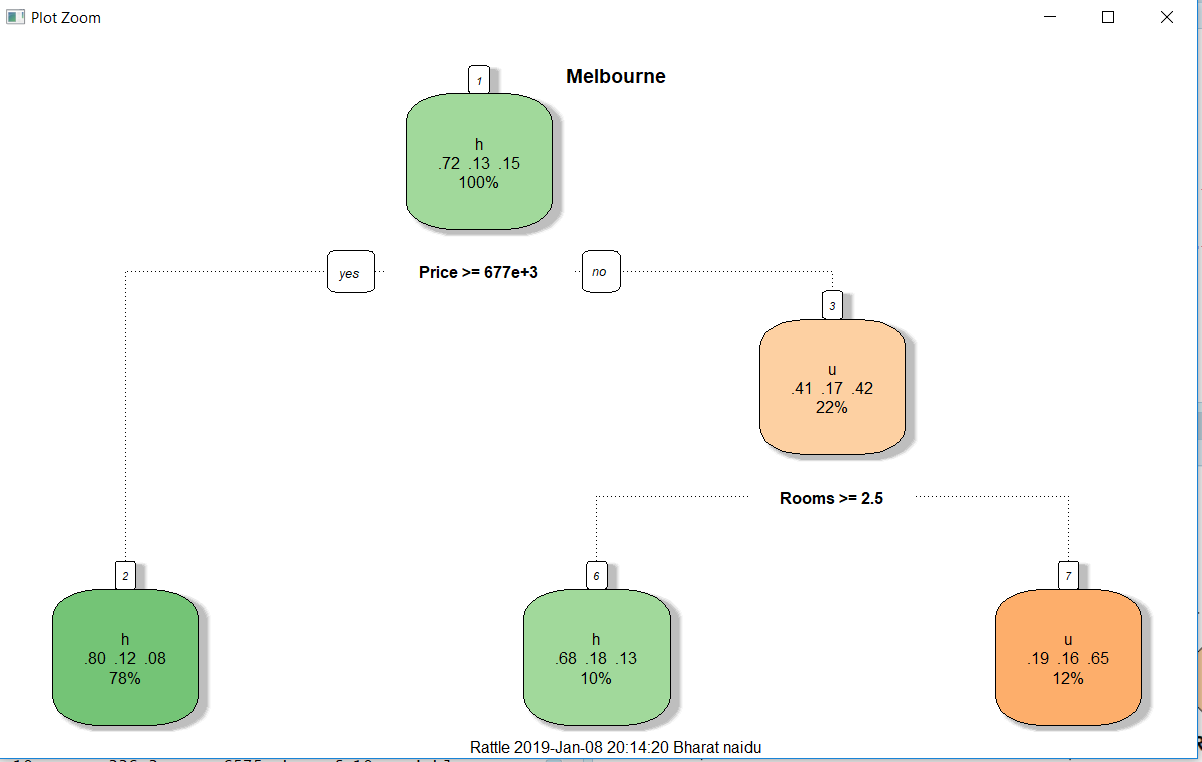
ii)if the building area is greater than 250 and at distance is less than 8, then we can say type of house is u.

iii)if the building area is greater than 220 and at distance is greater than 9, then we can say type of house is t.



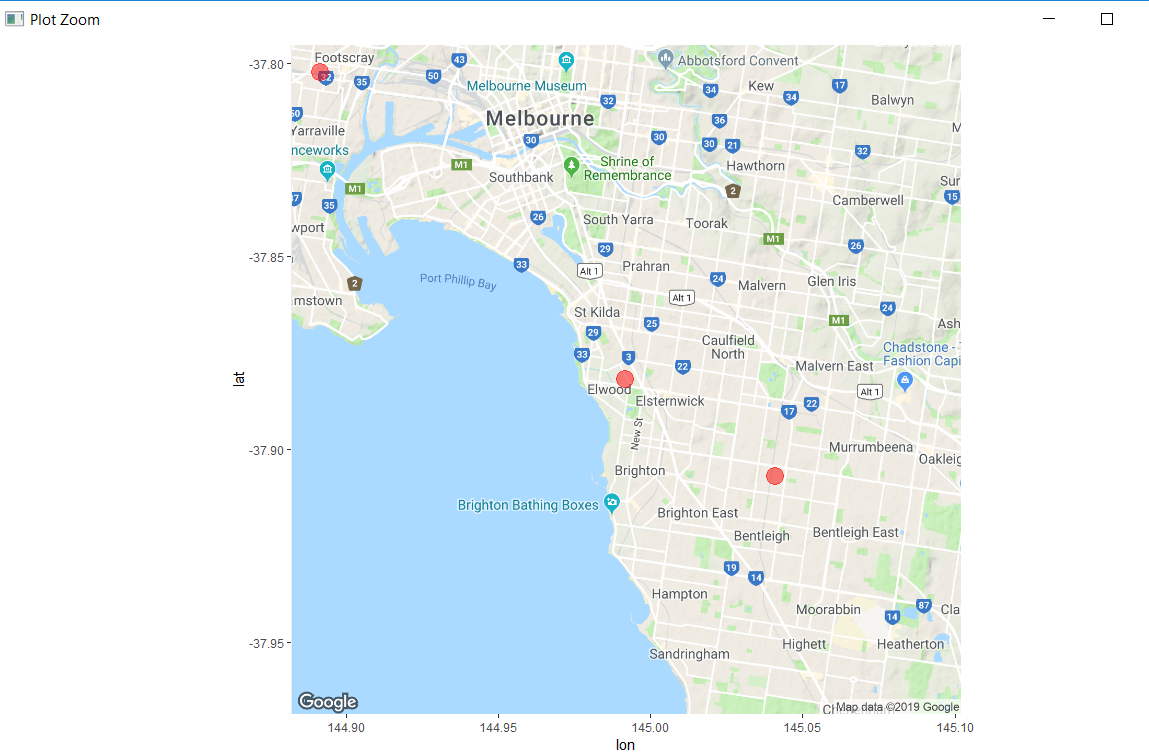
6. what is your decision tree results? Which variables are important?

a)From my decision tree results, i can infer that type of the house depends on price and number of rooms of the house.



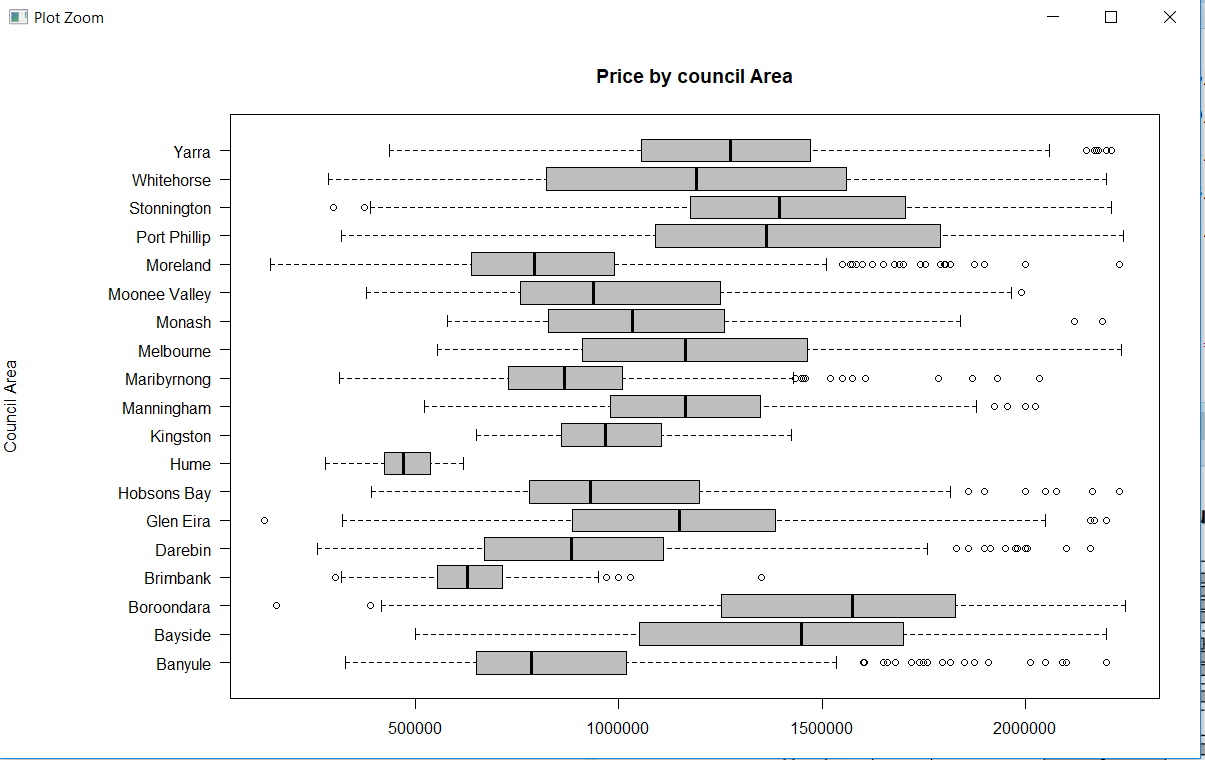
7. Are you able to draw google maps with median price?

a)yes



8)Are you able to identify any segmentation in data?

a)yes



If you have price of 2000000, you can buy 4 houses in hume.

9)Are you able to identify any recommender system?

a) yes