QMM 1002 Case Study 1

# Himani Dave

# A00178443

# Due: March 15th, 2019

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 0 marks | 1 mark | 2 marks | 3 marks  /50 | Gr. |
| Data Analysis | No confidence interval is included | A confidence interval is included with more than 1 error | A confidence interval is included with 1 error | A correct confidence interval is included |  |
| No or an inappropriate t-test is included | An appropriate t-test is included with more than 1 error | An appropriate t-test is included with 1 error | A correct and appropriate t-test is included |  |
| No or an inappropriate ANOVA test | An ANOVA is included with more than 1 error | An ANOVA is included in 1 error | A correct ANOVA test is included |  |
| No plots are created | A single correct plot is created to support the data analysis | At least two plots are created but contain errors | At least two correct plots are created to support the data analysis |  |
| Report | No conditions and assumptions are checked | Conditions and assumptions are checked with 2 errors | Conditions and assumptions are checked with 1 error | All conditions and assumptions are checked correctly and in detail | **X3** |
| The Rockland housing scenario question is not answered | The housing scenario question is answered with some missing detail | The housing scenario question is clearly answered in full detail | N/A | **X4** |
| No interpretation of test/confidence intervals is included | Interpretation of test/interval is included with error | A correct interpretation is given for the test/interval | N/A | **X3** |
| No correct hypotheses are included for the test conducted | Hypotheses are included for the test conducted with at least 1 error | Correct hypotheses are included for the test conducted | N/A | **X2** |
| No interpretation of plots | Plot is interpreted with at least 1 error | Plot is correctly interpreted | N/A | **X2** |
| Format | Numerous spelling errors, incorrect punctuation, and/or severe errors in grammar so that the report is hard to understand. | Frequent spelling errors, incorrect punctuation, and grammar problems that sometimes interfere with understanding. | Occasional lapses in spelling, punctuation, grammar, but not enough to seriously distract the reader. | Very few spelling errors, correct punctuation, grammatically correct, complete sentences. |  |
| Incorrect notation and terminology used throughout | More than 2 errors in notation and terminology | 1-2 minor errors in notation and terminology | No errors in notation or terminology |  |
| There are not 3 sections of the report | There are 3 sections of the report as given in the template | N/A | N/A |  |
| Bonus | No additional test(s) are conducted | N/A | N/A | Additional test(s) are conducted and interpreted without error to support decision making |  |

# Introduction

A family of four wants to buy a house in Rockland, which was amalgamated with Clarence Town in 1988. Their budget is $350,000. In order to investigate the prices of homes, a random sample of 169 dwellings has been taken and data like number of bedrooms, number of bathrooms, the area the house is in, etc. The following questions have been answered in the data analysis:

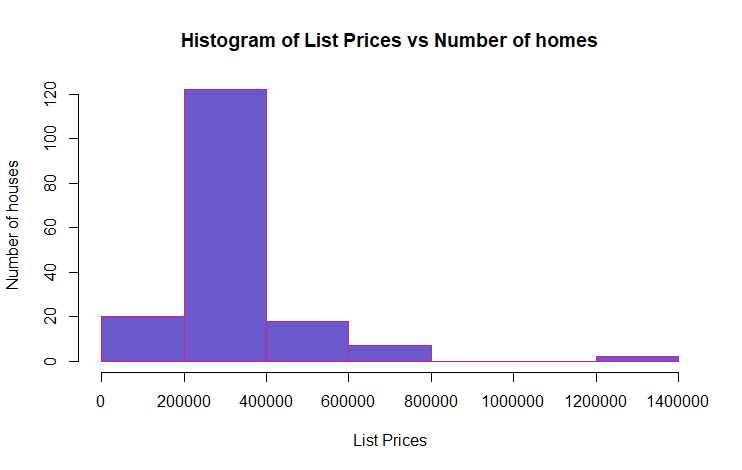
1. What should they expect to pay to buy a home in Rockland? Can they afford to buy a home in Rockland?
2. How do the list prices compare to different types of homes?
3. Is it more expensive to purchase a home in one area of the city?
4. Are there any specific patterns/differences in the types of homes available that will have enough bedrooms for the family of four?

# Methods and Analysis

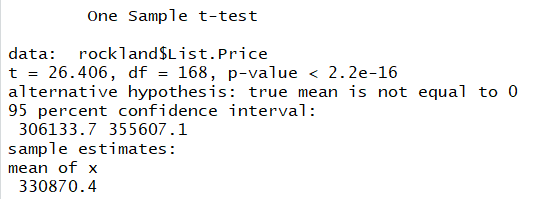
In order to find the confidence interval for the mean of list prices in Rockland, we use the basic one-sample t-test on List.Price.

Checking Assumptions:

1. Independence assumption: Since the houses are randomly selected, they are independent of each other. Thus, this condition is met.
2. Randomization condition: Since the houses are randomly selected, this condition is satisfied.
3. 10% condition: 169 homes are less than the total number of homes in Rockland. Thus, this condition is satisfied.
4. Nearly normal condition:



From the histogram, we can see that the data for List Price is skewed to the right. This indicates that there are a few outliers. From the data, we can see that the prices for two of the houses in Rockland are $1289000 and $1338000. They seem to be the outlier. However, since the sample size is large, t-tests can be applied.



From the t-test, we can be 95% confident that the real mean for the list prices in Rockland lies between $306133.7 and $355607.1.

The budget of the family is $350000.

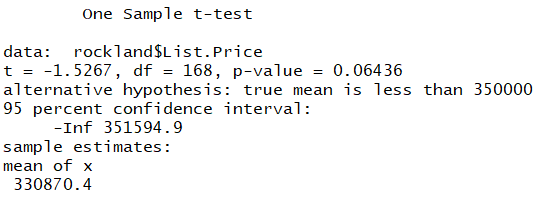
A hypothesis test can be carried out to check whether the family can afford a home in Rockland.

Hypotheses:

Null Hypothesis, H0: mu(list.price) = 350,000

Alternate Hypothesis, HA: mu(list.price) < 350,000

The conditions have already been checked.



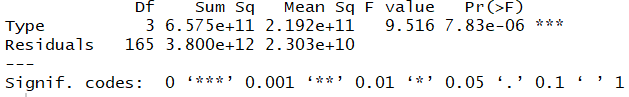
From this, we can see that the p-value is more than 0.05. So, we can reject the null hypothesis which means, there is evidence that the mean of list prices of homes in Rockland is less than 350000. Thus, the family can afford a home in Rockland.

There are four types of homes: Apartment, House, Mobile, Row/Townhouse. In order to compare the mean prices of four homes, one-way anova can be used.

Hypotheses:

Null Hypothesis, H0: There is no difference in the prices of different types of homes.

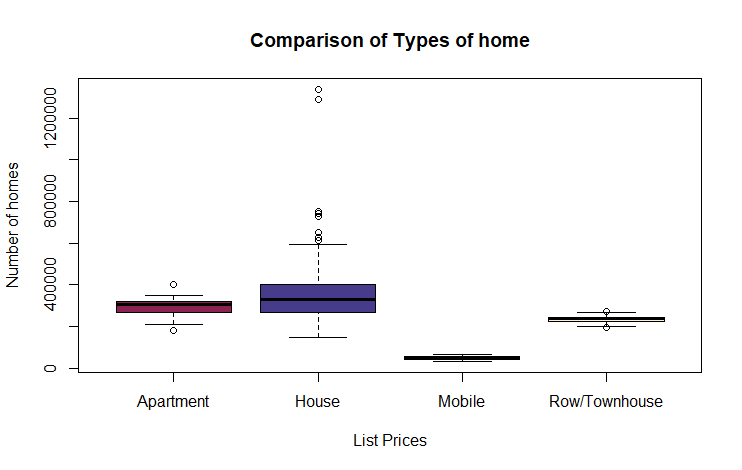
Alternate Hypothesis, HA: The list price of at least one type of home is different.



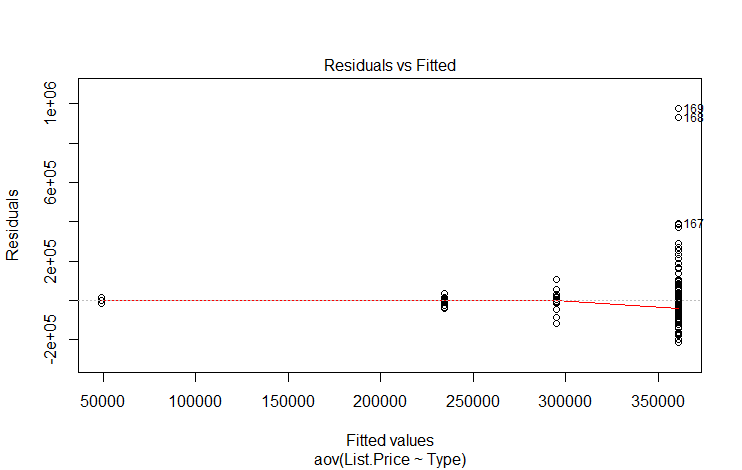
Here, the p-value is much less than alpha, so we reject the null. Thus, we can say that at least one of the types of home has a different mean list price.

Checking the conditions for one-way anova:

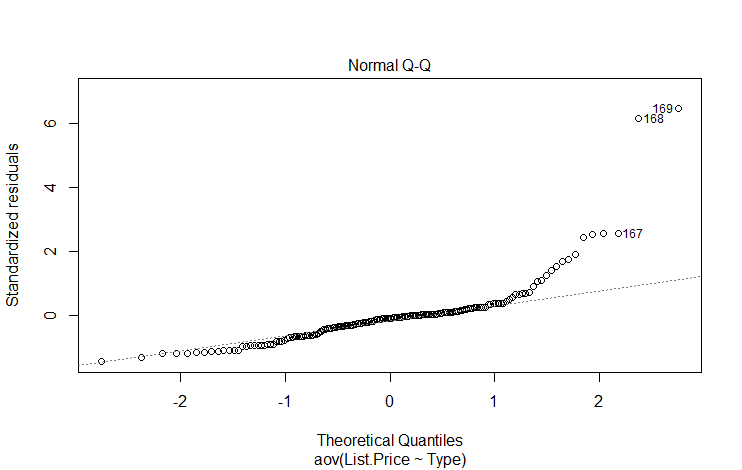
1. Independence assumption: The price of one home does not affect the price of others.
2. Randomization condition: The homes are randomly selected. Hence, this condition is met.
3. Independent group assumption: The list prices of one group do not affect that of others
4. Equal variance:



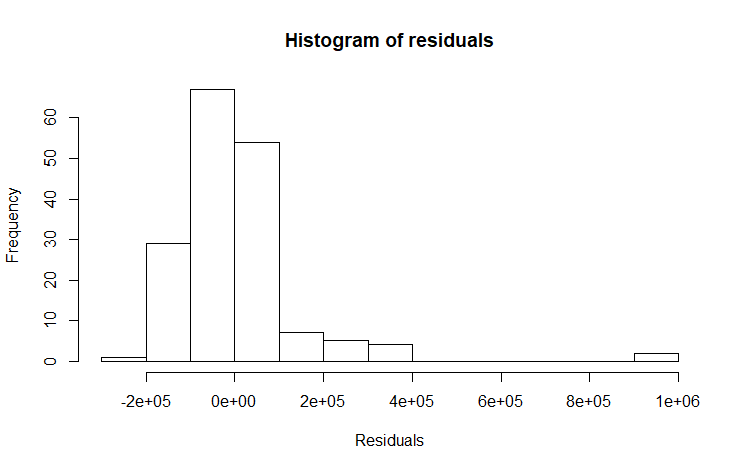
All types seem to have pretty close means. Mobile homes seem to have a lower mean than the others. There is no extreme skewness.



The Residuals vs Fitted Plot shows the red line close to the equal spread line.

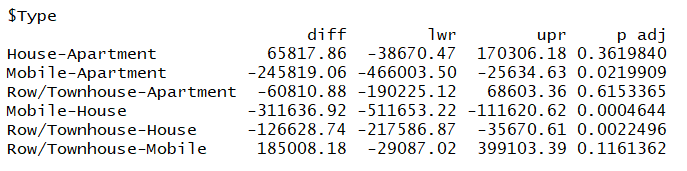


The normal Q-Q plot follows a straight line with a few outliers.

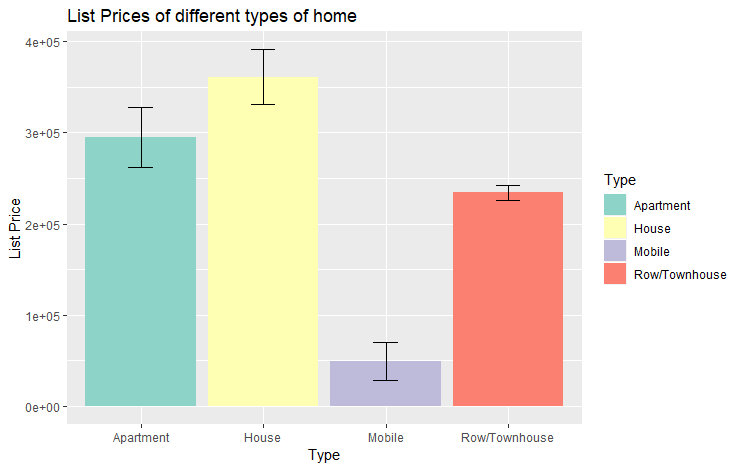


The histogram is nearly normal with a few outliers. Thus, the one-way anova holds.

Tukey's HSD test can be applied to find out which type of homes are different.



Here, we can see that the p-values for Mobile-Apartment, Mobile-House, and Row/Townhouse-House are different. Thus, we can say that the mean list price of Mobile homes is different from the rest.



In order to compare prices in different areas, we will compare the means of list prices in different areas using t-test.

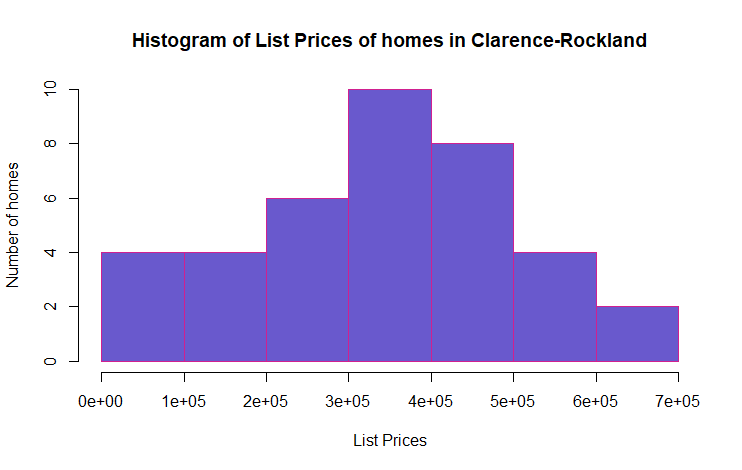
Hypotheses:

Null Hypothesis, H0: mu(List Prices of home in Clarence-Rock)-mu(List Prices of home in Rock) = 0

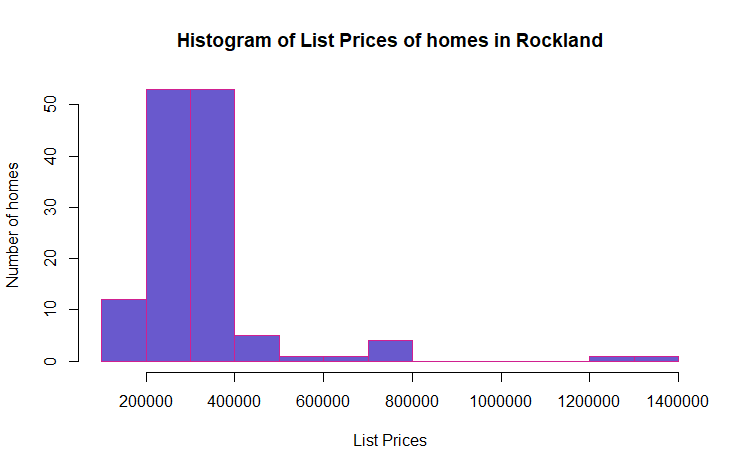
Alternate Hypothesis, HA: mu(List Prices of home in Clarence-Rock)-mu(List Prices of home in Rock) != 0

Assumptions and conditions:

1. Independence Assumption: The data is drawn independently for both areas.
2. Randomization Condition: The homes are randomly selected in both areas.
3. 10% condition: Both 38, and 131 are less than the total number of homes in both areas.
4. Nearly Normal condition:

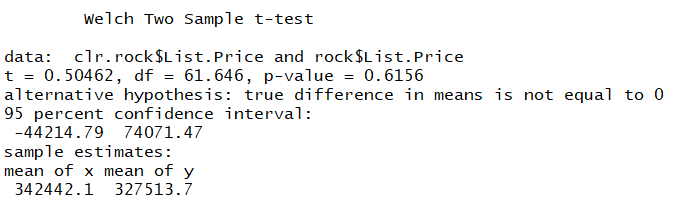


The histogram is unimodal, symmetric, and nearly normal.



The histogram is nearly normal with a few outliers.

Independent Groups Assumption: The two groups are independent of each other as the homes are randomly selected.



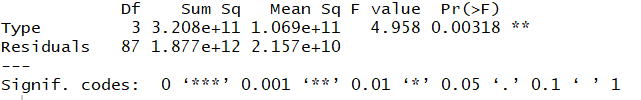
Here, p-value is greater than alpha (0.05). Thus, we fail to reject null. Thus, it can be said that the mean list price in both the areas is not significantly different from each other.

Assuming that the family of four includes parents and 2 children, they might need a three-bedroom home. We would conduct a one-way anova test for this.

Hypotheses:

Null Hypothesis, H0: There is no difference in the prices of different types of homes with three bedrooms.

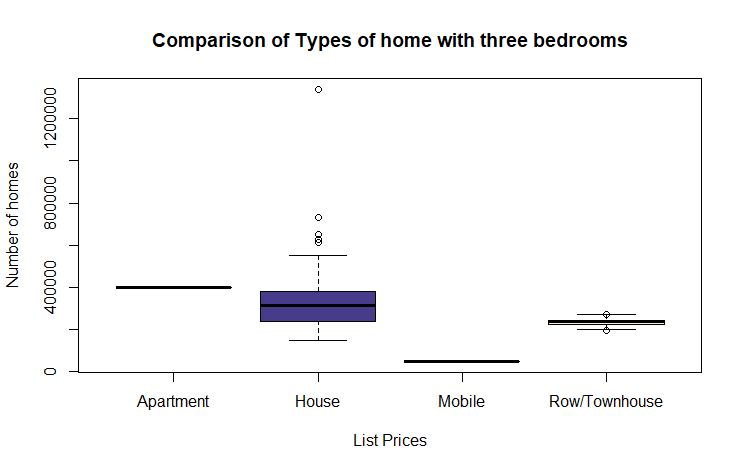
Alternate Hypothesis, HA: The list price of at least one type of home with three bedrooms is different.



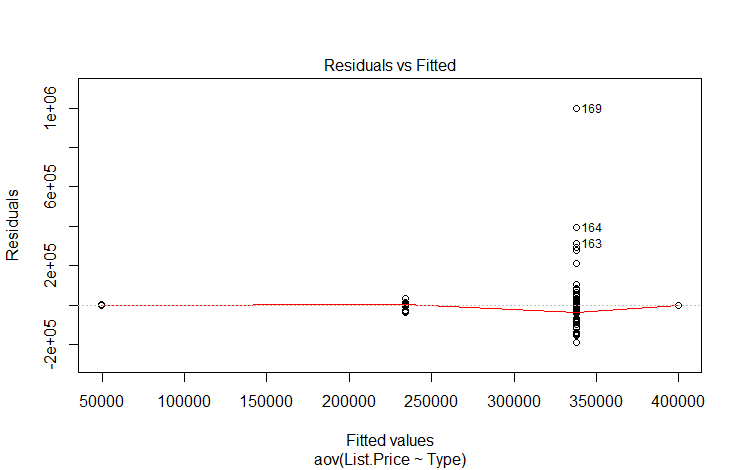
Here, p-value is less than alpha. Thus, we reject the null. We have enough evidence to say that the price of at least one type of home is different with three bedrooms.

Checking the conditions for one-way anova:

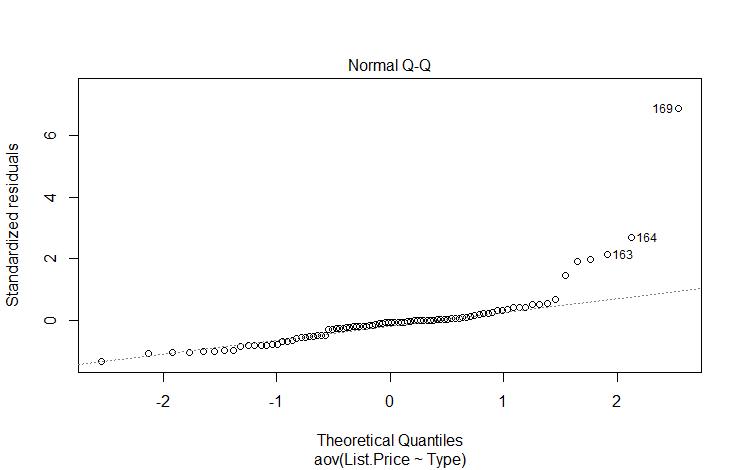
1. Independence assumption: The price of one home does not affect the price of others.
2. Randomization condition: The homes are randomly selected. Hence, this condition is met.
3. Independent group assumption: The list prices of one group do not affect that of others.
4. Equal variance:



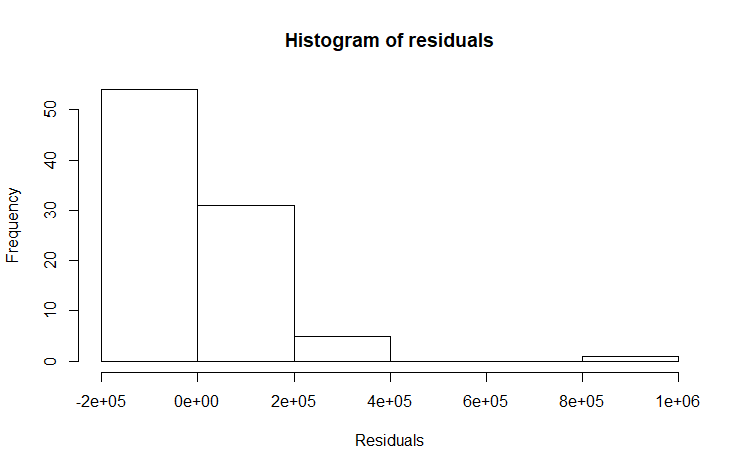
All types seem to have pretty close means. Mobile homes seem to have a lower mean than the others. There is no extreme skewness.



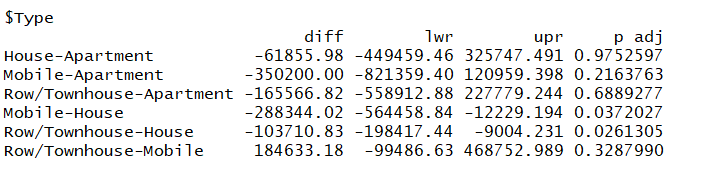
The Residuals vs Fitted Plot shows the red line close to the equal spread line.



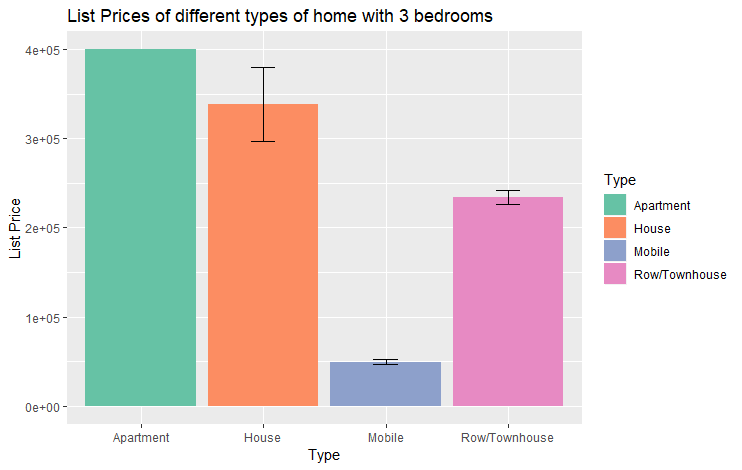
The normal Q-Q plot follows a straight line with a few outliers.



The histogram is nearly normal with a few outliers. Thus, the one-way anova holds.

Tukey's HSD test can be applied to find out which type of homes are different.

Here, we can see that the p-values for Mobile-House and Row/Townhouse-House are different. Thus, we can say that the mean list price of House with three bedrooms is different from the rest.



Thus, the houses with three bedrooms have a different price range than other types.

# Results and Discussion

From the analysis, we can say that the family can afford a house within their budget of $350,000. They should buy a three-bedroom house in any of the areas of Rockland as the prices are not much different in both areas.