Summary of Multiple Linear Regression Analysis

We conducted a multiple linear regression analysis to investigate the relationships between various predictor variables and the price of homes sold in top Canadian cities. Our aim was to develop a comprehensive model that captures the multifaceted influences on home prices, incorporating factors such as median family income and location-specific attributes.

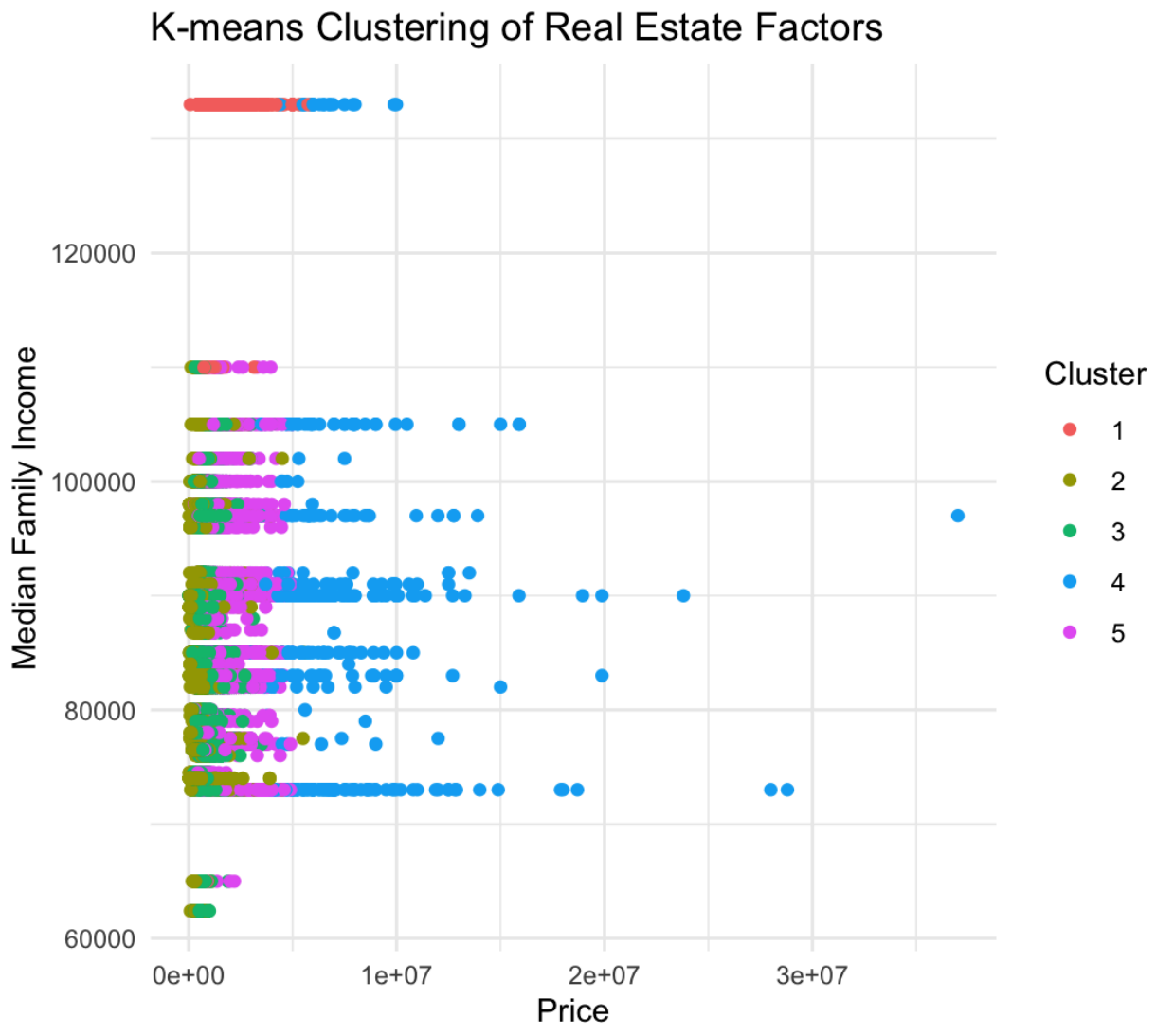
Key Findings:

- **Coefficients:**
  - The coefficient estimates indicate the direction and magnitude of the relationships between predictor variables and home prices.
  - The coefficient for "Median_Family_Income" was estimated to be 0.1699, though it was not statistically significant at the conventional levels ($p = 0.6682$).
  - Notably, the number of bedrooms ("Number_Beds") had a significant negative effect on home prices, with an estimated coefficient of -16,670 ($p < 0.001$).
  - Conversely, the number of bathrooms ("Number_Baths") had a highly significant positive effect on home prices, with an estimated coefficient of 362,400 ($p < 0.001$).
- **Model Fit:**
  - The model had a multiple R-squared value of 0.2183, indicating that approximately 21.83% of the variance in home prices was explained by the predictors included in the model.
  - The adjusted R-squared value, which accounts for the number of predictors in the model, was 0.2182.
  - The F-statistic was highly significant ($F = 3329$, $p < 2.2e-16$), suggesting that the regression model as a whole was statistically significant.
- **Residual Analysis:**
  - The residuals (differences between observed and predicted values) exhibited a wide range, with a minimum of -6,144,089 and a maximum of 33,065,556.
  - The residual standard error, an estimate of the variability of residuals around the regression line, was 902,000.

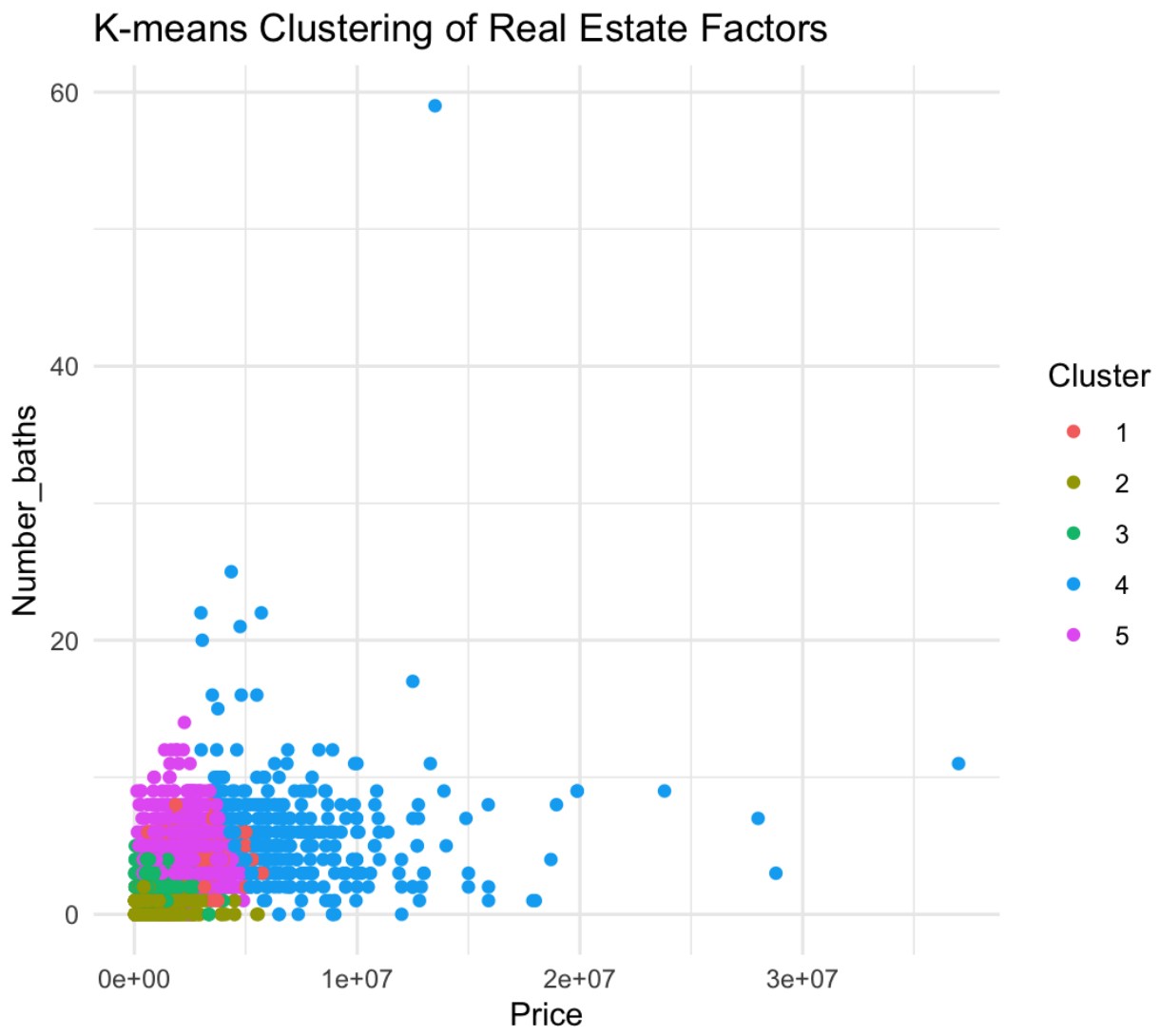
Conclusion:

Our analysis reveals several insights into the factors influencing home prices in the selected Canadian cities. While median family income did not demonstrate a significant direct association with home prices, the number of bedrooms and bathrooms emerged as key determinants. These findings underscore the importance of considering both socioeconomic factors and property-specific attributes in understanding housing market dynamics and affordability.

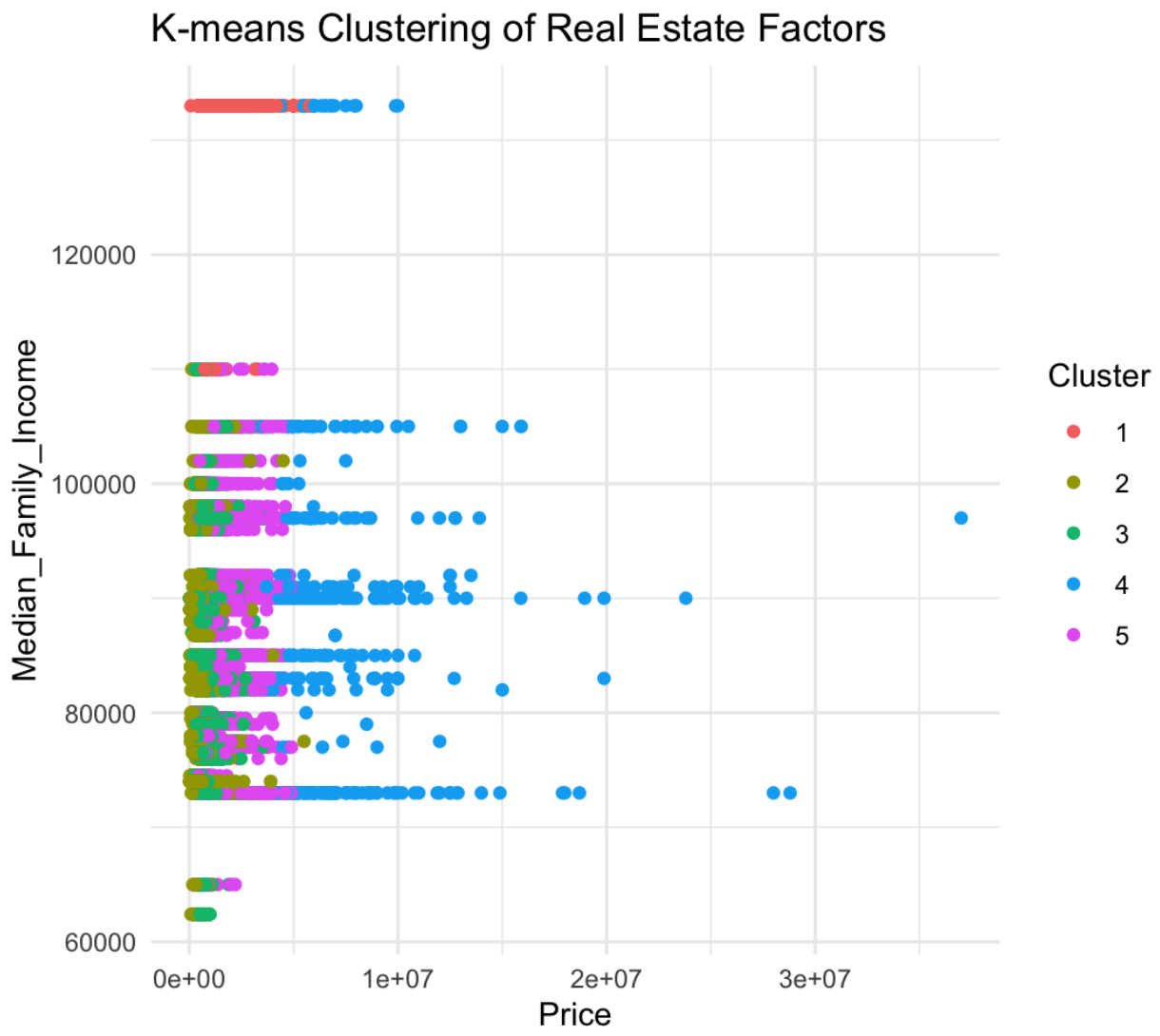
K-Means clustering:



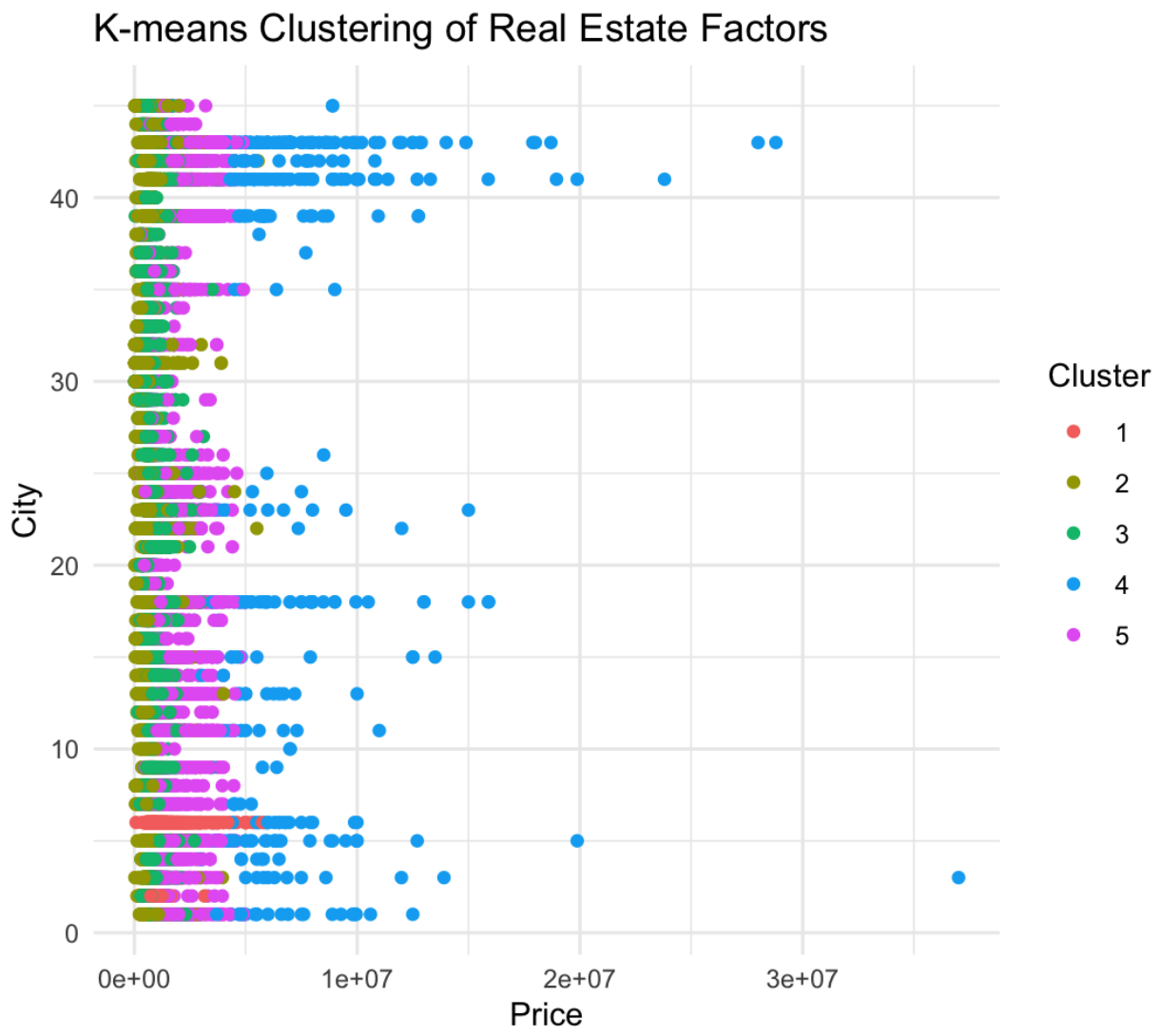
Now price / Number_Baths



Now price / median family Income



Now Price / City



Naive Bayes: very bad, may not use:

"Accuracy: 0.00391389432485323"

Cross Validation:

```
Call:
lm(formula = .outcome ~ ., data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-7408107  -296724   -84491   132520 33304196

Coefficients: (1 not defined because of singularities)
                      Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)              359948      31157  11.553  < 2e-16 ***
CityAirdrie             -703562      61434 -11.452  < 2e-16 ***
CityBarrie              -239966      38876  -6.173 6.79e-10 ***
CityBrantford           -381836      43656  -8.746  < 2e-16 ***
CityBurnaby              229137      36965   6.199 5.76e-10 ***
CityCaledon              -86954      36836  -2.361   0.0183 *
CityCalgary             -611792      36923 -16.569  < 2e-16 ***
CityEdmonton            -714552      36788 -19.424  < 2e-16 ***
CityGuelph              -195231      48638  -4.014 5.98e-05 ***
CityHalifax             -402378      89701  -4.486 7.29e-06 ***
CityHamilton            -249659      37058  -6.737 1.64e-11 ***
CityKamloops            -481601      48421  -9.946  < 2e-16 ***
CityKelowna             -228904      37132  -6.165 7.14e-10 ***
CityKingston            -434976      47024  -9.250  < 2e-16 ***
CityKitchener           -340510      37566  -9.064  < 2e-16 ***
CityLethbridge          -806666      50892 -15.851  < 2e-16 ***
CityLondon              -511818      36762 -13.923  < 2e-16 ***
`CityMaple Ridge`        146801      37103   3.957 7.62e-05 ***
`CityMedicine Hat`      -822038      56784 -14.477  < 2e-16 ***
CityMoncton             -752803      49375 -15.247  < 2e-16 ***
CityMontreal             -59758      62990  -0.949   0.3428
CityNanaimo             -181986      42790  -4.253 2.11e-05 ***
`CityNew Westminster`      6199      36864   0.168   0.8665
CityOshawa              -367319      38333  -9.582  < 2e-16 ***
CityOttawa              -352519      38297  -9.205  < 2e-16 ***
CityPeterborough        -385903      54745  -7.049 1.83e-12 ***
`CityPrince George`     -703695      54735 -12.856  < 2e-16 ***
CityQuebec              -467928      64979  -7.201 6.08e-13 ***
`CityRed Deer`          -790258      53571 -14.752  < 2e-16 ***
CityRegina              -801611      34785 -23.045  < 2e-16 ***
`CitySaint John`        -312589      48295  -6.473 9.76e-11 ***
CitySaskatoon           -742689      35075 -21.174  < 2e-16 ***
`CitySault Ste. Marie`  -653161      69509  -9.397  < 2e-16 ***
CitySherbrooke          -500706      64807  -7.726 1.14e-14 ***
`CitySt. Catharines`    -339707      37172  -9.139  < 2e-16 ***
`CitySt. John's`        -793649      43196 -18.373  < 2e-16 ***
CitySudbury             -583674      63926  -9.130  < 2e-16 ***
`CityThunder Bay`       -658900      71499  -9.216  < 2e-16 ***
CityToronto              161126      37231   4.328 1.51e-05 ***
`CityTrois-Rivieres`    -471145      69670  -6.763 1.38e-11 ***
CityVancouver            664838      36957  17.990  < 2e-16 ***
CityVictoria              21959      36939   0.594   0.5522
`CityWhite Rock`         743290      37796  19.666  < 2e-16 ***
CityWindsor             -538095      42154 -12.765  < 2e-16 ***
CityWinnipeg            -580760      45855 -12.665  < 2e-16 ***
Number_Beds              25862       3749   6.898 5.35e-12 ***
Number_Baths            306266       4771  64.198  < 2e-16 ***
Median_Family_Income        NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 808300 on 35721 degrees of freedom
Multiple R-squared:  0.373,      Adjusted R-squared:  0.3722
F-statistic:   462 on 46 and 35721 DF,  p-value: < 2.2e-16
```