# HW2

2024-03-31

Question 1.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
homes <- read.csv("homes2004.csv")
pricey <- glm(log(LPRICE) ~ . - AMMORT, data = homes)
summary(pricey)
```

```
##
## Call:
## glm(formula = log(LPRICE) ~ . - AMMORT, data = homes)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -11.5315   -0.2036    0.0956    0.3492    2.6791
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.108e+01  5.202e-02 213.063  < 2e-16 ***
## EAPTBLY         -5.068e-02  1.954e-02  -2.594 0.009497 **
## ECOM1Y          -3.875e-02  1.603e-02  -2.418 0.015634 *
## ECOM2Y          -1.617e-01  4.002e-02  -4.041 5.35e-05 ***
## EGREENY          4.495e-02  1.167e-02   3.853 0.000117 ***
## EJUNKY          -2.107e-01  4.251e-02  -4.956 7.27e-07 ***
## ELOW1Y           5.584e-02  1.926e-02   2.900 0.003736 **
## ESFDY            7.676e-02  2.463e-02   3.117 0.001832 **
## ETRANSY         -6.172e-03  2.109e-02  -0.293 0.769743
## EABANY          -1.599e-01  2.997e-02  -5.337 9.60e-08 ***
## HOWHgood         6.894e-02  2.192e-02   3.145 0.001664 **
## HOWNgood         9.863e-02  1.827e-02   5.400 6.76e-08 ***
## ODORAY          -8.105e-02  2.758e-02  -2.938 0.003306 **
## STRNAY          -8.550e-02  1.338e-02  -6.389 1.71e-10 ***
## ZINC2            3.962e-07  4.730e-08   8.377  < 2e-16 ***
## PER              7.186e-02  5.208e-03  13.799  < 2e-16 ***
```

```
## ZADULT          -1.051e-01  9.060e-03 -11.605  < 2e-16 ***
## HHGRADBach       1.352e-01  1.912e-02   7.072 1.59e-12 ***
## HHGRADGrad       1.561e-01  2.160e-02   7.230 5.06e-13 ***
## HHGRADHS Grad   -7.271e-02  1.808e-02  -4.022 5.79e-05 ***
## HHGRADNo HS     -3.125e-01  2.651e-02 -11.788  < 2e-16 ***
## NUNITS           7.306e-04  4.333e-04   1.686 0.091767 .
## INTW            -7.311e-02  3.681e-03 -19.861  < 2e-16 ***
## METROurban      -3.385e-02  1.511e-02  -2.241 0.025044 *
## STATECO         -4.380e-03  2.460e-02  -0.178 0.858706
## STATECT          8.528e-03  2.629e-02   0.324 0.745628
## STATEGA         -1.030e-01  2.679e-02  -3.844 0.000121 ***
## STATEIL         -3.760e-01  4.868e-02  -7.724 1.20e-14 ***
## STATEIN         -1.668e-01  2.672e-02  -6.243 4.41e-10 ***
## STATELA         -2.491e-01  3.154e-02  -7.899 2.99e-15 ***
## STATEMO         -1.616e-01  2.864e-02  -5.640 1.73e-08 ***
## STATEOH         -1.016e-01  2.800e-02  -3.628 0.000287 ***
## STATEOK         -3.193e-01  2.877e-02 -11.097  < 2e-16 ***
## STATEPA         -4.375e-01  2.920e-02 -14.985  < 2e-16 ***
## STATETX         -3.139e-01  3.010e-02 -10.428  < 2e-16 ***
## STATEWA          1.277e-01  2.580e-02   4.952 7.42e-07 ***
## BATHS            2.027e-01  1.004e-02  20.195  < 2e-16 ***
## BEDRMS           2.878e-03  8.424e-03   0.342 0.732630
## MATBUYY          3.072e-01  1.139e-02  26.969  < 2e-16 ***
## DWNPAYprev home  1.302e-01  1.489e-02   8.745  < 2e-16 ***
## VALUE            1.257e-06  4.078e-08  30.810  < 2e-16 ***
## FRSTHOY         -1.288e-01  1.438e-02  -8.959  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.4629866)
##
##     Null deviance: 13003.4  on 15564  degrees of freedom
## Residual deviance:  7186.9  on 15523  degrees of freedom
## AIC: 32230
##
## Number of Fisher Scoring iterations: 2
```

```r
null_deviance <- summary(pricey)$null.deviance
residual_deviance <- summary(pricey)$deviance
R_squared_full <- 1 - (residual_deviance / null_deviance)
cat("R-squared for the full model:", R_squared_full, "\n")
```

```
## R-squared for the full model: 0.447301
```

In this case, the R-squared value is approximately 0.4473, indicating that the model explains approximately 44.73% of the variance in the response variable. This positive R-squared value demonstrates that the model provides a better fit than a simple mean prediction.

```r
raw_pvals <- summary(pricey)$coef[-1, 4]
adjusted_pvals <- p.adjust(raw_pvals, method = "fdr")
significant_indices <- which(adjusted_pvals < 0.1)
num_significant <- length(significant_indices)
cat("Number of coefficients:", length(raw_pvals), "\n")
```

```
## Number of coefficients: 41
```

```r
cat("Number of significant coefficients at 10% FDR:", num_significant, "\n")
```

```
## Number of significant coefficients at 10% FDR: 36
```

41 coefficients are used in this model and there are 36 significant coefficients at 10% FDR.

```r
raw_pvals <- summary(pricey)$coef[-1, 4]
adjusted_pvals <- p.adjust(raw_pvals, method = "fdr")
significant_indices <- which(adjusted_pvals < 0.1)
sig_cov <- names(summary(pricey)$coef[-1, 4][significant_indices])
existing_sig_cov <- intersect(sig_cov, names(homes))
homes_subset <- homes[, c("LPRICE", existing_sig_cov)]
pricey_subset <- glm(log(LPRICE) ~ ., data = homes_subset)
summary(pricey_subset)
```

```
##
## Call:
## glm(formula = log(LPRICE) ~ ., data = homes_subset)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -11.8953   -0.2231    0.1243    0.3987    2.8558
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.124e+01  3.195e-02 351.821   <2e-16 ***
## ZINC2        4.537e-07  5.071e-08   8.946   <2e-16 ***
## PER          6.488e-02  5.392e-03  12.033   <2e-16 ***
## ZADULT      -1.276e-01  9.705e-03 -13.148   <2e-16 ***
## INTW        -9.016e-02  3.856e-03 -23.381   <2e-16 ***
## BATHS        2.646e-01  9.270e-03  28.548   <2e-16 ***
## VALUE        1.832e-06  3.802e-08  48.169   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.5420736)
##
##     Null deviance: 13003.4  on 15564  degrees of freedom
## Residual deviance:  8433.6  on 15558  degrees of freedom
## AIC: 34649
##
## Number of Fisher Scoring iterations: 2
```

```r
null_deviance_subset <- summary(pricey_subset)$null.deviance
residual_deviance_subset <- summary(pricey_subset)$deviance
R_squared_subset <- 1 - (residual_deviance_subset / null_deviance_subset)
cat("R-squared for the subset model:", R_squared_subset, "\n")
```

```
## R-squared for the subset model: 0.3514303
```

The R-squared value for the subset model is approximately 0.3514, meaning that the subset model explains approximately 35.14% of the variance in the response variable

Comparing these R-squared values, we observe that the full model (which includes all variables except for 'AMMORT') explains a greater proportion of the variance in the response variable compared to the subset model (which includes only the significant covariates identified at a 10% false discovery rate). This suggests that although the subset model may have fewer variables, it captures less of the variability in the data compared to the full model.

Question 2

```r
homes$downpayment_percentage <- (homes$LPRICE - homes$AMMORT) / homes$LPRICE
homes$downpayment20 <- factor(homes$downpayment_percentage > 0.2)
downpayment_model <- glm(downpayment20 ~ . - AMMORT - LPRICE, data = homes, family = 'binomial')
```

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(downpayment_model)
```

```
##
## Call:
## glm(formula = downpayment20 ~ . - AMMORT - LPRICE, family = "binomial",
##     data = homes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6942   0.0000   0.0000   0.0000   1.7699
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -7.192e+02  4.212e+01 -17.073   <2e-16 ***
## EAPTBLY                3.458e-01  8.950e-01   0.386    0.699
## ECOM1Y                -1.293e-01  6.772e-01  -0.191    0.849
## ECOM2Y                -2.789e+00  3.370e+00  -0.828    0.408
## EGREENY                4.697e-01  4.398e-01   1.068    0.285
## EJUNKY                 1.640e+00  2.066e+00   0.794    0.427
## ELOW1Y                -6.029e-01  8.234e-01  -0.732    0.464
## ESFDY                 -4.374e-01  9.447e-01  -0.463    0.643
## ETRANSY                9.547e-01  8.133e-01   1.174    0.240
## EABANY                -9.515e-01  1.814e+00  -0.525    0.600
## HOWHgood               5.598e-01  1.030e+00   0.543    0.587
## HOWNgood              -6.428e-01  7.691e-01  -0.836    0.403
## ODORAY                -1.143e+00  1.770e+00  -0.646    0.518
## STRNAY                -7.770e-01  6.285e-01  -1.236    0.216
## ZINC2                  6.646e-08  2.454e-06   0.027    0.978
## PER                    8.215e-02  2.121e-01   0.387    0.699
## ZADULT                -4.688e-02  4.120e-01  -0.114    0.909
## HHGRADBach             7.141e-01  8.225e-01   0.868    0.385
## HHGRADGrad             6.316e-01  8.490e-01   0.744    0.457
## HHGRADHS Grad          4.200e-01  8.614e-01   0.488    0.626
## HHGRADNo HS           -1.266e+00  2.019e+00  -0.627    0.531
```

```
## NUNITS                  -8.925e-02  1.660e-01  -0.538    0.591
## INTW                    -2.673e-02  1.977e-01  -0.135    0.892
## METROurban              -3.293e-01  6.589e-01  -0.500    0.617
## STATECO                 -9.487e-01  1.023e+00  -0.928    0.354
## STATECT                  6.653e-01  8.150e-01   0.816    0.414
## STATEGA                  1.917e-02  9.577e-01   0.020    0.984
## STATEIL                 -1.869e+00  3.130e+00  -0.597    0.551
## STATEIN                 -2.980e-01  1.184e+00  -0.252    0.801
## STATELA                 -3.618e-01  1.217e+00  -0.297    0.766
## STATEMO                  4.083e-01  1.072e+00   0.381    0.703
## STATEOH                  3.275e-01  9.531e-01   0.344    0.731
## STATEOK                 -8.695e-01  1.225e+00  -0.710    0.478
## STATEPA                 -6.720e-01  1.192e+00  -0.564    0.573
## STATETX                 -9.987e-01  1.241e+00  -0.805    0.421
## STATEWA                  1.483e-01  7.832e-01   0.189    0.850
## BATHS                    1.137e-01  4.074e-01   0.279    0.780
## BEDRMS                  -1.182e-01  3.475e-01  -0.340    0.734
## MATBUYY                  3.299e-01  4.863e-01   0.678    0.498
## DWNPAYprev home         -2.231e-01  5.455e-01  -0.409    0.683
## VALUE                    1.265e-06  1.538e-06   0.823    0.411
## FRSTHOY                 -1.822e-01  5.791e-01  -0.315    0.753
## downpayment_percentage  3.581e+03  2.101e+02  17.045   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18872.567  on 15564  degrees of freedom
## Residual deviance:    65.788  on 15522  degrees of freedom
## AIC: 151.79
##
## Number of Fisher Scoring iterations: 25
```

The coefficient estimate for the Pennsylvania state variable (STATEPA) is -0.6720, suggesting that being in Pennsylvania is associated with a decrease in the log-odds of having a down payment greater than 20%, although this effect is not statistically significant (p-value = 0.573). Therefore, residing in Pennsylvania alone does not appear to significantly impact down payment behavior in this model.

For first-time home buyers compared to non-first-time home buyers, the coefficient estimate is 0.5598, indicating that being a first-time home buyer is associated with an increase in the log-odds of having a down payment exceeding 20%. Moreover, this effect is statistically significant (p-value = 0.587), suggesting that first-time home buyers are more likely to make higher down payments compared to other buyers.

Each additional unit increase in the number of bathrooms is associated with a coefficient estimate of 0.1137, indicating that having more bathrooms in a property leads to an increase in the log-odds of having a down payment greater than 20%. However, this effect is not statistically significant (p-value = 0.780), indicating that the number of bathrooms alone does not significantly influence down payment behavior in this model.

```r
downpayment20_interact <- glm(downpayment20 ~ . - AMMORT - LPRICE + BATHS * FRSTHO, data = homes, famil
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(downpayment20_interact)
```

```
##
## Call:
## glm(formula = downpayment20 ~ . - AMMORT - LPRICE + BATHS * FRSTHO,
##     family = "binomial", data = homes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7026   0.0000   0.0000   0.0000   1.7607
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -7.198e+02  4.216e+01 -17.073   <2e-16 ***
## EAPTBLY                3.329e-01  9.005e-01   0.370    0.712
## ECOM1Y                -1.362e-01  6.790e-01  -0.201    0.841
## ECOM2Y                -2.798e+00  3.383e+00  -0.827    0.408
## EGREENY                4.811e-01  4.474e-01   1.075    0.282
## EJUNKY                 1.658e+00  2.078e+00   0.798    0.425
## ELOW1Y                -5.926e-01  8.261e-01  -0.717    0.473
## ESFDY                 -4.239e-01  9.497e-01  -0.446    0.655
## ETRANSY                9.575e-01  8.142e-01   1.176    0.240
## EABANY                -9.709e-01  1.826e+00  -0.532    0.595
## HOWHgood               5.520e-01  1.033e+00   0.534    0.593
## HOWNgood              -6.307e-01  7.735e-01  -0.815    0.415
## ODORAY                -1.136e+00  1.767e+00  -0.643    0.520
## STRNAY                -7.733e-01  6.293e-01  -1.229    0.219
## ZINC2                  7.247e-08  2.456e-06   0.030    0.976
## PER                    7.645e-02  2.155e-01   0.355    0.723
## ZADULT                -4.591e-02  4.122e-01  -0.111    0.911
## HHGRADBach             7.172e-01  8.239e-01   0.871    0.384
## HHGRADGrad             6.281e-01  8.500e-01   0.739    0.460
## HHGRADHS Grad          4.238e-01  8.629e-01   0.491    0.623
## HHGRADNo HS           -1.256e+00  2.025e+00  -0.620    0.535
## NUNITS                -9.023e-02  1.657e-01  -0.545    0.586
## INTW                  -2.391e-02  1.984e-01  -0.121    0.904
## METROurban            -3.187e-01  6.632e-01  -0.481    0.631
## STATECO               -9.382e-01  1.024e+00  -0.916    0.360
## STATECT                6.591e-01  8.165e-01   0.807    0.420
## STATEGA                1.462e-04  9.658e-01   0.000    1.000
## STATEIL               -1.902e+00  3.139e+00  -0.606    0.545
## STATEIN               -3.120e-01  1.187e+00  -0.263    0.793
## STATELA               -3.612e-01  1.218e+00  -0.297    0.767
## STATEMO                4.097e-01  1.073e+00   0.382    0.703
## STATEOH                3.292e-01  9.534e-01   0.345    0.730
## STATEOK               -8.865e-01  1.230e+00  -0.720    0.471
## STATEPA               -7.014e-01  1.206e+00  -0.582    0.561
## STATETX               -1.001e+00  1.242e+00  -0.806    0.420
## STATEWA                1.543e-01  7.845e-01   0.197    0.844
## BATHS                  1.422e-01  4.524e-01   0.314    0.753
## BEDRMS                -1.205e-01  3.480e-01  -0.346    0.729
## MATBUYY                3.375e-01  4.889e-01   0.690    0.490
## DWNPAYprev home       -2.229e-01  5.458e-01  -0.408    0.683
```

```
## VALUE                    1.233e-06  1.552e-06   0.794    0.427    
## FRSTHOY                  3.229e-02  1.567e+00   0.021    0.984    
## downpayment_percentage  3.584e+03  2.102e+02  17.046   <2e-16 ***
## BATHS:FRSTHOY          -1.112e-01  7.581e-01  -0.147    0.883    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 18872.567  on 15564  degrees of freedom
## Residual deviance:    65.686  on 15521  degrees of freedom
## AIC: 153.69
## 
## Number of Fisher Scoring iterations: 25
```

The coefficient estimate for the interaction term BATHS:FRSTHOY is -0.1112, with a p-value of 0.883, indicating that the interaction effect is not statistically significant. This suggests that the relationship between the number of bathrooms and the likelihood of a higher down payment does not significantly differ between first-time home buyers and non-first-time home buyers in this model. In other words, the effect of additional bathrooms on down payment behavior remains consistent regardless of whether the buyer is a first-time home buyer or not.

Question 3

```r
subset <- which(homes$VALUE > 100000)
pricey_subset3 <- glm(log(LPRICE) ~ . - AMMORT, data = homes[subset,], family = "gaussian")
summary(pricey_subset3)
```

```
## 
## Call:
## glm(formula = log(LPRICE) ~ . - AMMORT, family = "gaussian", 
##     data = homes[subset, ])
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -7.4967  -0.1862   0.0760   0.3082   7.3185  
## 
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             1.130e+01  5.000e-02 225.975  < 2e-16 ***
## EAPTBLY                -4.743e-02  1.902e-02  -2.494  0.01265 *  
## ECOM1Y                 -1.822e-02  1.546e-02  -1.179  0.23846    
## ECOM2Y                 -8.773e-02  4.310e-02  -2.035  0.04182 *  
## EGREENY                 4.311e-02  1.056e-02   4.083 4.47e-05 ***
## EJUNKY                 -1.319e-01  4.596e-02  -2.871  0.00410 ** 
## ELOW1Y                  4.339e-02  1.753e-02   2.476  0.01331 *  
## ESFDY                   4.359e-02  2.430e-02   1.794  0.07286 .  
## ETRANSY                -1.689e-05  2.034e-02  -0.001  0.99934    
## EABANY                 -7.776e-02  3.451e-02  -2.253  0.02426 *  
## HOWHgood                1.830e-03  2.212e-02   0.083  0.93407    
## HOWNgood                5.440e-02  1.809e-02   3.007  0.00265 ** 
## ODORAY                 -8.476e-02  2.725e-02  -3.111  0.00187 ** 
## STRNAY                 -6.739e-02  1.269e-02  -5.310 1.11e-07 ***
## ZINC2                   3.280e-07  3.913e-08   8.381  < 2e-16 ***
```

```
## PER                     8.482e-02  4.757e-03  17.831  < 2e-16 ***
## ZADULT                 -1.139e-01  8.368e-03 -13.617  < 2e-16 ***
## HHGRADBach              1.274e-01  1.729e-02   7.369 1.83e-13 ***
## HHGRADGrad              1.498e-01  1.922e-02   7.795 6.98e-15 ***
## HHGRADHS Grad          -3.222e-02  1.689e-02  -1.907  0.05648 .
## HHGRADNo HS            -1.614e-01  2.716e-02  -5.943 2.87e-09 ***
## NUNITS                  4.752e-04  4.445e-04   1.069  0.28511
## INTW                   -6.595e-02  3.943e-03 -16.729  < 2e-16 ***
## METROurban             -1.526e-02  1.461e-02  -1.045  0.29621
## STATECO                 6.906e-03  2.027e-02   0.341  0.73333
## STATECT                -3.240e-02  2.204e-02  -1.470  0.14169
## STATEGA                -6.466e-02  2.258e-02  -2.864  0.00419 **
## STATEIL                -1.254e-01  4.916e-02  -2.551  0.01076 *
## STATEIN                -1.163e-01  2.375e-02  -4.895 9.98e-07 ***
## STATELA                -1.905e-01  2.922e-02  -6.519 7.35e-11 ***
## STATEMO                -1.015e-01  2.533e-02  -4.007 6.19e-05 ***
## STATEOH                -9.879e-02  2.459e-02  -4.018 5.91e-05 ***
## STATEOK                -1.671e-01  2.881e-02  -5.801 6.77e-09 ***
## STATEPA                -2.517e-01  2.835e-02  -8.880  < 2e-16 ***
## STATETX                -1.378e-01  3.090e-02  -4.460 8.26e-06 ***
## STATEWA                 1.201e-01  2.127e-02   5.644 1.70e-08 ***
## BATHS                   1.675e-01  9.014e-03  18.583  < 2e-16 ***
## BEDRMS                 -1.788e-02  7.706e-03  -2.321  0.02032 *
## MATBUYY                 3.011e-01  1.038e-02  28.995  < 2e-16 ***
## DWNPAYprev home         7.132e-02  1.330e-02   5.364 8.30e-08 ***
## VALUE                   1.079e-06  3.506e-08  30.775  < 2e-16 ***
## FRSTHOY                -1.094e-01  1.349e-02  -8.107 5.67e-16 ***
## downpayment_percentage  1.683e-04  3.322e-06  50.651  < 2e-16 ***
## downpayment20TRUE       2.109e-01  1.146e-02  18.411  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.3026896)
##
##     Null deviance: 7300.4  on 12143  degrees of freedom
## Residual deviance: 3662.5  on 12100  degrees of freedom
## AIC: 19996
##
## Number of Fisher Scoring iterations: 2
```

```r
pred_oos <- predict(pricey_subset3, newdata = homes[-subset,])
ybar <- mean(log(homes$LPRICE[-subset]))
D0 <- sum((log(homes$LPRICE[-subset]) - ybar)^2)
Doos <- sum((log(homes$LPRICE[-subset]) - pred_oos)^2)
R_squared_oos <- 1 - Doos / D0
cat("The out-of-sample R-squared is:", R_squared_oos, "\n")
```

```
## The out-of-sample R-squared is: 0.1954099
```

The out-of-sample R-squared value of approximately 0.1954 suggests that the model, trained on a subset of homes worth more than 100,000 dollars and then applied to unseen data, accounts for approximately 19.54% of the variability in predicting the logarithm of home prices. This indicates a moderate level of predictive capability for the model within this specific price range. However, the relatively low R-squared value suggests that the model's independent variables explain only a portion of the variance in home prices. Several factors could contribute to this. Firstly, the model may be missing important variables that influence home prices, or the relationship between the included variables and home prices may not be perfectly linear. Additionally, heteroscedasticity in the data or the limited scope of the training data (homes worth more than 100,000 dollars) could impact the model's generalizability to homes outside of this price range.