

To: Professor. Yu Du
From: Rubina Shaik, Anamika Singh, Anurag Lahon
Subject: Ski Resort Properties Data Analysis
Date: 4/15/2020

The report explains the Exploratory data analysis and Regression analysis performed to model the factors influencing selling price of the homes at Colorado Ski Resort.

EXECUTIVE SUMMARY

Major Findings

The data listed of 39 homes with a selling price of highest \$545k and lowest \$315k is analyzed. There is no missing and categorical data. The factors influencing the selling price of the home from the correlation matrix (Exhibit 1) are number of bathrooms, sq_ft of the home, number of bedrooms, while with miles from mountain the selling price is negatively correlated. Variation in selling price per variable is listed in Exhibit 2.

Comparisons of two best possible models with validation by data splitting is presented (Exhibit 3). Model building with list price though it gives high r^2 , adjusted r^2 and low mae, rsme in validation, all the coefficients included in the model aren't significant with p values. An optimized model with factors as bedrooms, sqft, miles from mountain, lot_size and garage is built and validated with all significant p-values. The assumptions are met for regression analysis (Exhibit4).

Recommendations for Action

To get higher prices for the houses, the property should be built closer to the mountains. With an increased number of bedrooms and garage the selling price of the home can be increased. As listing price is highly related to selling price it can be used as an marketing indicator to selling price.

Analytical Overview

The data was summarized and the factors determining the selling prices of the homes identified. The fitting variables are chosen to build an optimized model explaining the means of selection based on the requirements to satisfy regression analysis. All applicable tasks are performed till the validation of the model.

Documentation

Exhibit 1:

Correlation coefficients for Selling Price	
List price	0.98992618
Bathrooms	0.74268187
Sq_Ft	0.63714824
Bedrooms	0.6109798
Garage	0.32672161
On market	0.11046973
Lot size	-0.04422749
Age	-0.19714627
Downtwon	-0.27076285
Mountain	-0.40217178

Exhibit 2:

Variable	Selling Price Per Variable
Bedrooms	121.6667
Bathrooms	174.8
Sq_Ft	0.214207
Downtwn	75.71429
Mountain	67.14286
Lot size	1106.061
Garage	228.75
Age	24.33333
On market	3.8125
List price	0.9798489

Exhibit 3:

Validation by Data Splitting		
	Model 1	Model 2
RMSE	11.7572	23.37307
MAE	9.064281	19.8473
R2	0.9562514	0.8649
Vif	<pre>> vif(Vm1) Sq_Ft Downtwon 'Lot size' 'On market' 'List price' 3.157480 3.087699 2.874854 1.401524 2.952450</pre>	<pre>> vif(Vm2) Bedrooms Sq_Ft Mountain Garage 'Lot size' 2.189031 2.389787 1.664013 1.259722 1.741934</pre>
Coefficients by Significance	<pre>Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 9.021384 11.606212 0.777 0.4437 Sq_Ft 0.002036 0.003323 0.613 0.5451 Downtwon -0.359600 0.193405 -1.859 0.0739 'Lot size' 0.222097 0.271473 0.818 0.4205 'On market' -0.013364 0.015509 -0.862 0.3964 'List price' 0.955545 0.036335 26.298 <2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</pre>	<pre>Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 266.257830 18.547033 14.356 3.71e-14 *** Bedrooms 20.648675 5.981347 3.452 0.001847 *** Sq_Ft 0.042030 0.009896 4.247 0.000229 *** Mountain -4.555631 0.563856 -8.079 1.11e-08 *** Garage 14.139980 4.689123 3.015 0.005531 ** 'Lot size' 4.160842 0.723339 5.752 4.07e-06 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</pre>

Exhibit 4: **Diagnostic Plots**

