**Problem:** Big Mountain Resort has installed an additional chair lift to help increase the distribution of visitors across the mountain. The new chair lift will increase operating costs by $1,540,000 and the company would like to maintain its profit margin of 9.2%.

**Data Available**: a spreadsheet containing details of 330 ski resorts throughout the United States. 27 variables including details such as number and types of lifts, number of runs, night skiing, average snowfall and snow making acreage.

**Model**: a multiple linear regression model was set up to model the Adult Weekend chairlift price. In addition to providing a prediction of the price, the model also provides two measures of fit: the Explained Variance Score and the Mean Absolute Error. The Explained Variance Score is a measure of how much of the variance in chair lift costs is explained by the model, with 1 being a perfectly explained model. The model returned a value of .933.

**Data Preprocessing**:

* Missing data in the following columns were set to zeroes:

fastEight (166)

NightSkiing\_ac (143)

TerrainParks (51)

SnowMaking\_ac (46)

* Missing data in the following columns were set to the average of the column:

AdultWeekday (54)

projectedDaysOpen (47)

Runs (4)

AdultWeekend(51)

averageSnowfall (14)

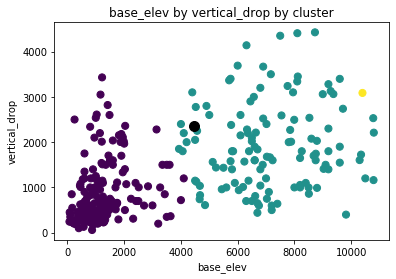
SkiableTerrain\_ac (3)

daysOpenLastYear(51)

LongestRun\_mi (5)

yearsOpen (1)

* Only data anomaly was for one resort which was listed as open for 2019 years (the value was reset to 1)
* For testing purposes, 25% of the input was preselected and not used as input while building the model. After the model was created using the other 75% of the input data, it was tested with the preselected 25% testing group.
* The linear model works best with numeric inputs, so the state variable was converted into individual flags for each state
* Resorts were grouped by base elevation and vertical drop.



**Model Development**: the first run of the model using all available inputs showed that the top determinants were the individual state variables created during preprocessing. Since location was not an actionable item as far as this analysis, the state variables were dropped from the model inputs. On the second run, both base elevation and vertical drop rose to importance. Since these two are highly correlated, the model was run a third time without base elevation. The results from each of the three runs were:

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Explained Variance** | **Mean Absolute Error** | **Features Dropped** |
| 1 | 0.940 | 4.93 |  |
| 2 | 0.931 | 5.26 | state |
| 3 | 0.933 | 5.19 | state, summit elevation, base elevation |

**Model Findings**:

* The most significant features in the model were: Adult Weekday Chairlift price and cluster (which incorporated base elevation and vertical drop).
* Predicted value of Adult Weekend Chairlift Price for Big Mountain: $88.33

**Recommendation**:

* Since the AdultWeekend and AdultWeekday are highly correlated, both should probably be changed to $88
* Based on an estimate of 350,000 visitors per year, this should result in increased revenue of $2,450,000, which would exceed the 9.2% profit goal