Big mountain resort report

**Problem Introduction**

Big Mountain Resort recently has installed an additional chair lift to help increase the distribution of visitors across the mountain. But this increases the operating costs by $1,540,000 this season. This business profit margin is 9.2% and the investors would like to keep it there. To recoup the increased operating costs from the new chair, we analysed the data of 330 resort with similar market. We found out that increasing the ticket price of adult weekend could solve this problem.

**Dataset**

We were given a single CSV file from the database manager. The file contains information from 330 resorts in the US that can be considered part of the same market share and we have these same data columns for Big Mountain Resort as well.

**Features and Processing**

Here I listed the main steps that I processed to explore the solution:

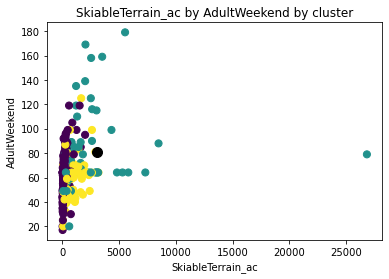
|  |  |
| --- | --- |
| Processing | Keys |
| Print out the percent of each feature that is missing | Fill up the missing and NA values |
| Review boxplots for every column | Apply outlier removal using the Interquartile range |
| Create pairplots | Explore the data relationships |
| Create a Pearson correlation heatmap | Identify features that suffer from multi-collinearity |
| Clustering with K-means algorithm | Puts the data into groups which can then become features for further analysis |
| Build models with sklearn linear regression | Review the explained variance, mean absolute error and the features relationship to choose a better model for getting the findings |

**Results and Discussion**

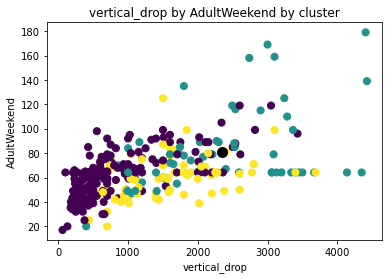
We built 3 models with sklearn linear regression. The different between 3 models is that the irrelevant features we dropped off. To determine a better model for further analysis, we reviewed the model performances ---- explained variance (also known as R-squared value) and mean absolute error. The Explained Variance is a measure of the variation explained by the model. Explained Variance can take any values between 0 to 1. A higher Explained Variance indicates a better fit for the model. Mean Absolute Error tells us how big of an error we expected from the models. The lower values of Mean Absolute Error indicate better fit. But there is no big different for explained variance and Mean Absolute Error among 3 models (showed in table 1). Then we compared the feature dropping condition. The model 3 dropped more irrelevant features that can provide more proving insights. Then we chose model 3 to do the further analysis and predicted the chairlift ticket price of adult weekend. The predicted price is $89, $8 higher than actual price. We also reviewed the results with cluster scatter plots showed in figure 1 and figure 2. The figure 1 is the scatter plots for visualizing the cost of an adult weekend chairlift ticket (AdultWeekend) compared to total skiable area in square acres (SkiableTerrain\_ac). The bigger size with black color is for big mountain resort. It showed that the bigger size of total skiable area has higher ticket price, and the price of big mountain resort is in the middle. It indicates to increasing the price of an adult weekend chairlift ticket is reliable. The same observation from figure 2, the bigger vertical change resort has higher price, and the price of big mountain resort is in the middle. This also prove that our recommendation about increasing the chairlift ticket price is practical suggestion.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Explained Variance** | **Mean Absolute Error** | **Features Dropped** |
| Model 1. | 0.9347 | 5.07 | - |
| Model 2. | 0.9243 | 5.50 | 'state' |
| Model 3. | 0.9257 | 5.46 | 'state','summit\_elev','base\_elev' |

*Table 1: model performances*



*Figure 1.* *The scatter plots for visualizing the Adult Weekend values compared to SkiableTerrain\_ac. The AdultWeekend is the cost of an adult weekend chairlift ticket, and SkiableTerrain\_ac is total skiable area in square acres. The bigger size with black color is for big mountain resort.*



*Figure 2. The scatter plots for visualizing the Adult Weekend values compared to vertical\_drop. The Adultweekend is the cost of an adult weekend chairlift ticket, and vertical\_drop is* *vertical change in elevation from the summit to the base in feet. The bigger size with black color is for big mountain resort.*

**Conclusion**

The predicted price is $8 more than the price of last season. This should allow Big Mountain Resort to recoup the increased operating costs. Future, we could investigate other features like ticket price of adult weekday and the number of days the resort will be open each season.