Project Report - Big Mountain Resort

Problem Statement

Every year about 350,000 visitors come to ski or snowboard at Big Mountain Resort located in Montana. Recently, an additional chair lift has been installed which increases operating costs by $1,540,000 this season. The resort's pricing strategy is to charge a premium above the average ticket price in its market segment. Big Mountain Resort wants to find out whether its facilities can support a better price and needs some guidance on how to change the current facilities to cut costs or to support a higher price.

Big Mountain Resort wants to find the optimized ticket price this season considering the current Big Mountain's facilities and the strategy to increase revenue by either cutting costs without undermining the ticket price or charging an even higher ticket price.

Dataset and the Features

The original dataset includes information about ticket prices, location, features, and facilities from 330 ski resorts in the USA. We add population and area data for the US state from Wikipedia and it has replaced the ratio of resorts serving a given population or a given area in the data cleaning process. The final dataframe contains 277 ski resorts and 36 columns and the target feature is the weekend ticket price.

Exploratory Data Analysis

Name, State, and Region are the only categorical features. Other than that, the rest of the features are numerical.

The state-wide summary statistics indicate that the population and the area of the state are not necessarily linked to the total number of resorts or the total skiing area in terms of the absolute value. However, when plotting the ratio of resorts serving a given population or a given area, we can check that most of the states are distributed on the left side of the histogram as shown in figure 1. Few states fall far away from all the other states in both plots 'Number of resorts per 100k square miles' and 'Number of resorts per 100k population. These long skewed tails are caused by the states such as New Hampshire and Vermont. The fact that these states ranked 6 and 7 in the total number of resorts but their population and area are at the lower end leads to relatively high ratios.

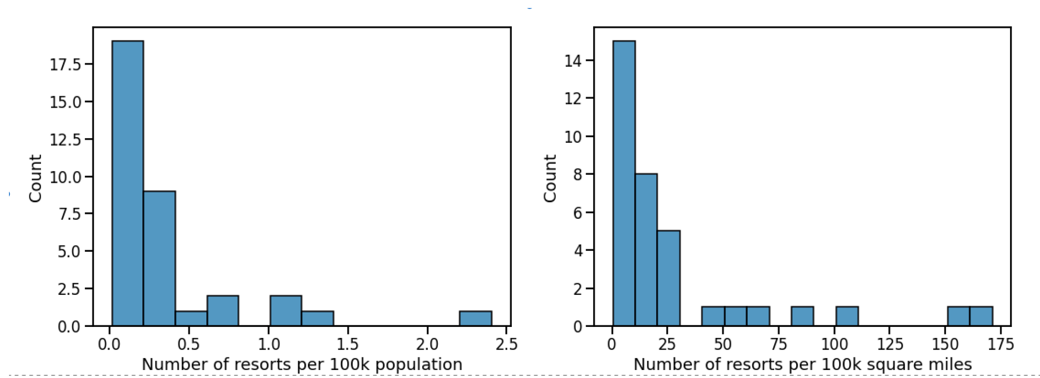

Figure 1. The density of resorts relative to the state population and size

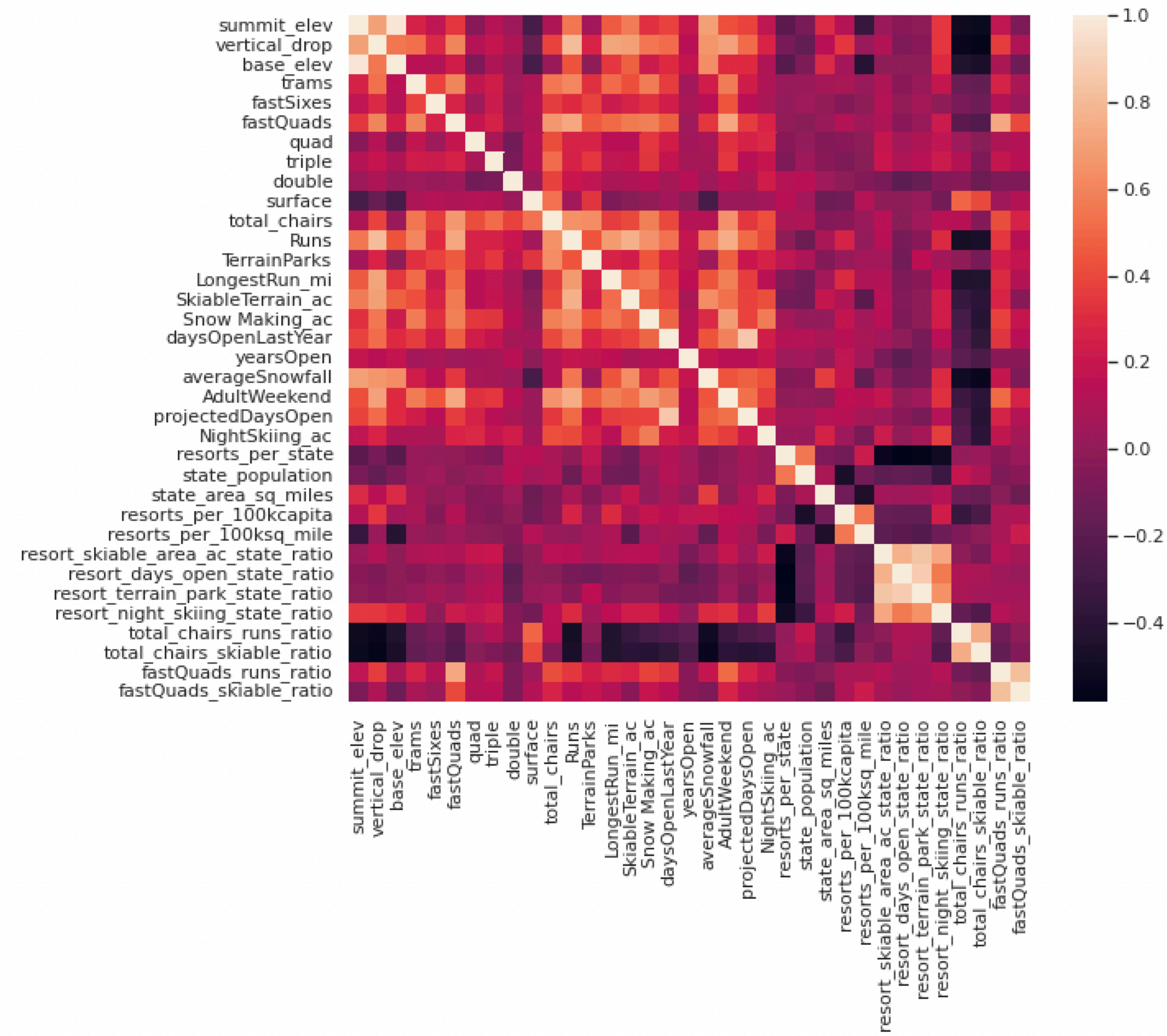

Figure 2. Feature correlation heatmap

From principal components analysis (PCA) for analyzing the relationship between the ticket price and the state-related features, no obvious pattern is shown. PCA plot shows that most states are clustered together except Vermont and New Hampshire for the same reason stated in the previous paragraph. This supports that all states can be considered equally to build a pricing model.

Figure 2, feature correlation heatmap, tells us `resort_night_skiing_state_ratio`, `fastQuads`, `runs`, `total_chairs`, `vertical_drop`, and `Snow Runs Making_ac` are highly correlated with our target feature `AdultWeekend`.

Modeling Results

We use the random forest model to analyze since it shows a better performance with a lower mean absolute error, and also less variability than the linear regression model. According to results using RandomForestRegressor, the top 4 important features are `fastQuads`, `Runs`, `Snow Making_ac`, and `vertical_drop`.

The adult weekend ticket price of Big Mountain Resort is currently $81. Compared with the national average price, $64.2, it seems that they sell a ticket at a high price. Big Mountain Resort also charges the highest price in Montana and the state average price is $51.9. However, the pricing prediction model suggests that Big Mountain resort can charge up to $95.87 with the current facilities. This price is estimated by considering all the impacts of each facility on the price put together. Since the expected mean absolute error is $10.39, this suggests consumers are willing to pay more. Though Big Mountain Resort already charges a high price compared to the market average, our analysis shows that its facilities can support further price increases. It helps that the business leaders decide the range of increase considering an accurate prediction of the costs.

The results of reviewing 4 possible scenarios for either cutting costs without undermining the ticket price or charging an even higher ticket price are as follows.

1. After analyzing the effect of closing runs on price when all other things are unchanged, the model shows that closing one run makes no difference. Therefore, Big mountain Resort can successfully keep costs down without lowering the price.
2. If Big Mountain Resort adds a run with the increased vertical drop by 150 feet, they also need to install an additional chair lift. In this case, the model supports a markup of $1.99, and the expected revenue increase is $3474683.
3. However, adding 2 acres of snow-making area on top of the previous scenario2 or increasing the longest run by .2 miles with additional 4 acres of snow-making make no difference in the ticket price.

Every year about 350,000 visitors come to ski or snowboard at Big Mountain Resort and spend five days on average. This number includes the recently installed chair lift which increases operating costs by $1,540,000 this season. Since the additional operating cost per ticket is $0.88, an increase in the price of more than $0.88 generates an increase in revenue and profit.

The Limitations of the Data and Analytics

The data includes ticket prices only for adults and the information about the total number of visitors is given. This can cause inaccurate cost-benefit analysis considering the proportion of family visitors.

The operating cost for an additional chair lift is only provided. According to scenario 2, the model suggests that Big Mountain Resort is capable of increasing the current prices. In order to build it up convincingly, the costs for managing runs and snow-making equipment are necessary.

Based on the result of pre-processing data, `fastQuads`, `Runs`, `Snow Making_ac`, and `vertical_drop` are the top 4 important features. But, in the case of `Snow Making_ac`, 13.4% of data are still missing after data cleaning. Missing values replaced with the median can cause bars in the model prediction. Also, 51 resorts that don't provide `AdultWeekend` prices are dropped. Then we can question the representativeness of the samples that might lead to the loss of prediction power.