# Guided Capstone Project Report

Big Mountain Resort is a ski resort located in Montana (USA) welcoming as much as 350,000 skiers or snowboarders per year. To accommodate the high number of visitors and increase the distribution of visitors across the mountain, to the 11 lifts, 2 T-bars and 1 magic carpet was recently added a new chair lift. This investment resulted in the increase of the operational costs by 1,540,000$ this season. The pricing strategy of the company needs to be adjusted to select a better value for the ticket price aiming to maximize the capitalization of the facilities, according to the market.

The specific problem to answer by this work is the following: “What are the options for Big Mountain Resort to adjust their pricing strategy in order to increase the capitalization rate and absorb the investment of a new chair lift (1,540,000$) this season?”

The original dataset provided by Alesha Eisen (Database Manager), composed of 330 rows and 27 columns was combined with columns from a dataset from Wikipedia containing geographic and demographic information for the US. There was no missing value for the resort of interest. Columns containing a lot of missing values, without valuable information for the actual problem to solve or with error in the values were dropped. The cleaned datasets containing the target “AdultWeekend” holding the ticket price for the weekend were saved. Figure 1 display the average ticket price by state for the weekend and the weekdays.

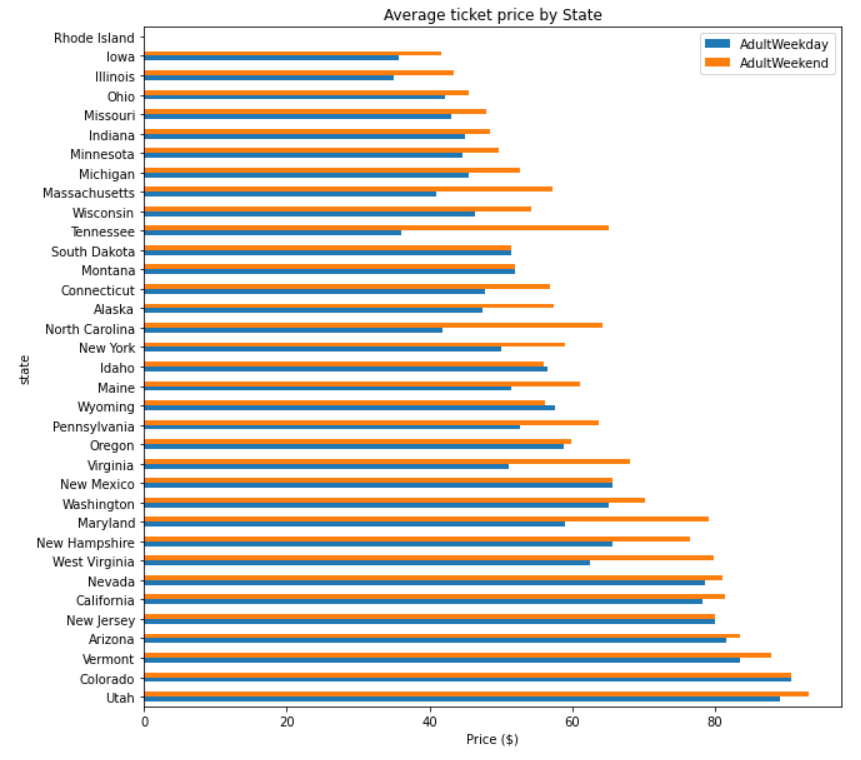


Figure 1 : Average ticket price by state.

Trends in the variable were explored during data wrangling. In order to obtain more meaningful variables, different ratios were calculated (resort per state vs state population and resort per state vs state area) and the original column were dropped. We then used PCA to explore some trend in the data. It appears that the 2 first components (figure 2) seem to account for over 75% of the variance and the first 4 components account for over 95%.

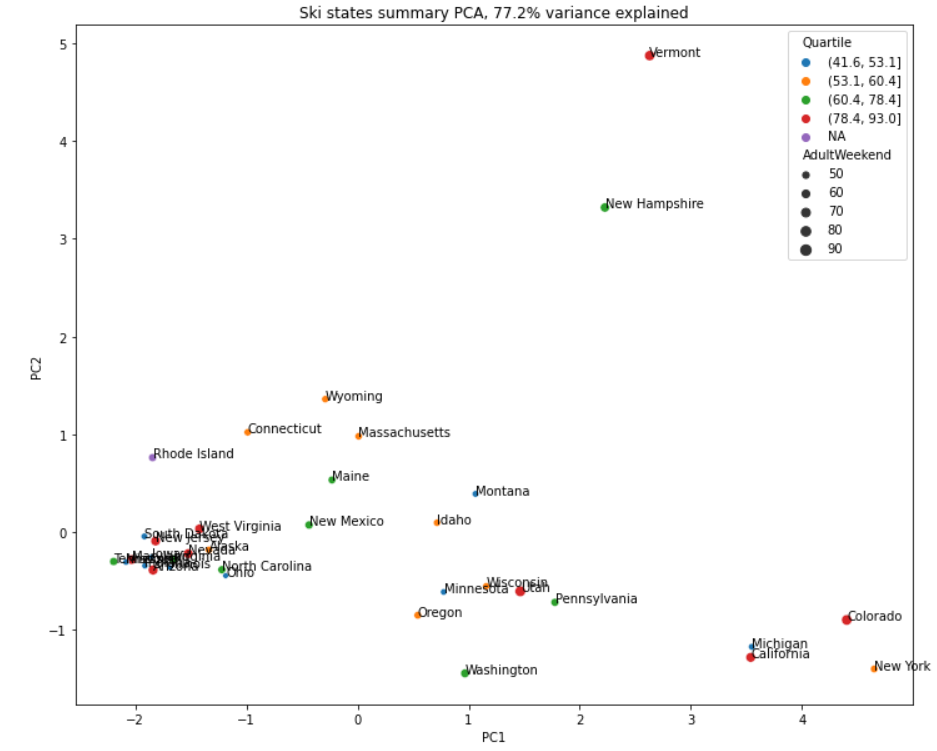


Figure 2 : Variance between PC1 and PC2 for ski state summary.

Une image contenant texte

Description générée automatiquement

Figure 3 : Feature correlation heatmap.

Correlation between variables were visualized on a heatmap (figure 3) and we can see that “resorts\_per\_100kcapita” and “resorts\_per\_100ksq\_mile” have a high correlation. We can also see that the ticket price is related to a lot of variables, including the number of chairs, the characteristics of the mountains, and the other facilities.

For pre-processing and training, the dataset was splitted into the training set (70%) and test set (30%) to have an independent assessment of how the model might perform in the future. To avoid waist of time, we verified the mean of the ticket price as a predictor for a prediction of the ticket price for Big Mountain. The result (63.81$) underestimates the actual ticket price and could not suggest any increase.

To further explain the variance in the data, many metrics were verified: R2, MAE, MSE. The results indicates that on average, the ticket price is off by 19$ if it is based on an average of known values.

Missing values were also imputed using the median and the model implies that, on average, the difference between the estimate of a ticket price is within 9$ or so of the real price. This is better that the model based on the mean. We verified the random forest model (figure 4) to check the mean and median strategies for imputing missing values, with and without scaling. The results show that imputing with the median helps, but scaling the features does not.

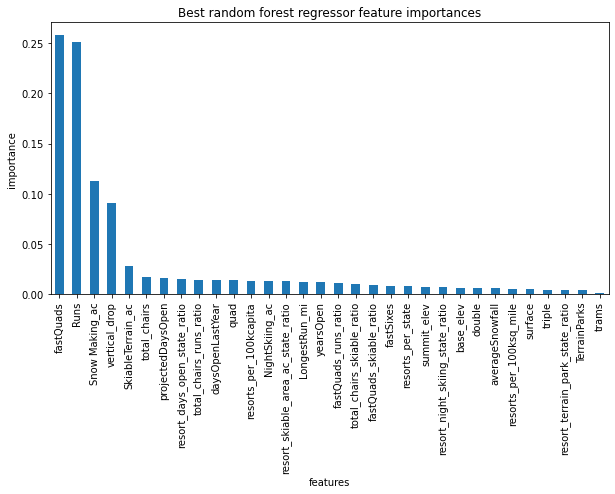


Figure 4 : Best random forest regressor feature importances.

The pre-modeling of the results show that the vertical drop and the area covered by snow making equipment of a ski resorts have the biggest correlation with ticket price. In contrary, in this model, the skiable terrain area has a negative correlation with ticket price. In addition, the results of the linear model indicate that the top four dominant features related to ticket price are: 1) fastQuads; 2) Runs; 3) Snow Making\_ac; and 4) vertical\_drop.

The modelled price for Big Mountain is 95.87$, while the actual price is 81.00$. Considering an expected mean absolute error of 10.39$, Big Mountain resort seems to be charging much less that what is predicted. By the observation of figure 5, we can see that the ticket price for Big Mountain is near the end of the distribution (without considering the more extreme values or outliers). We can also see that the ticket price for the Big Mountain resort is already the highest in Montana.

Considering the vertical drop, we can see that Big Mountain is near the end of the distribution, but there quite few resorts with a greater drop. The situation is similar for the skiable terrain area. For the snow making area, we can see that Big Mountain is in the highest category. Considering the total chairs, Big Mountain is amongst the highest number of total chairs, without considering the extreme values and the outliers. The situation is similar whit the number of fast quads distribution, as well as for the total of runs and the longest run length. However, Big Mountain do not have trams, like the majority of resorts.

There are two avenues for Big Mountain resort to solve the problem: either cutting costs or increasing revenues (ticket price). Four scenarios were analyzed to gain insight and to possibly solve the problem: 1) Permanently closing down up to 10 of the least used runs; 2) Increasing the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage; 3) Same as number 2, but adding 2 acres of snow making cover; 4) Increasing the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres.

Assuming 350,000 visitors over the season and visitors skiing for 5 days. The results for scenario 1 indicate that closing one run makes no difference but closing 2 and 3 runs successively reduces support for ticket price or revenue. If Big Mountain closes 3 runs, it makes no difference than closing 4 or 5 runs. But increasing the closures down to 6 or more leads to a large drop in revenue. For scenario 2, the result support an increase of 1.99$ for ticket price. Over a season, this could lead to a profit of 3474638$. For scenarios 3 and 4, this makes no difference in the profit. In order to upgrade instead of downgrading, scenario 2 would be support here. Considering that an increase of 14.87 ± 10.39$ is suggested by modeling, 1.99$ could be considered as a minimum increase to account for the investment of a new chairlift.

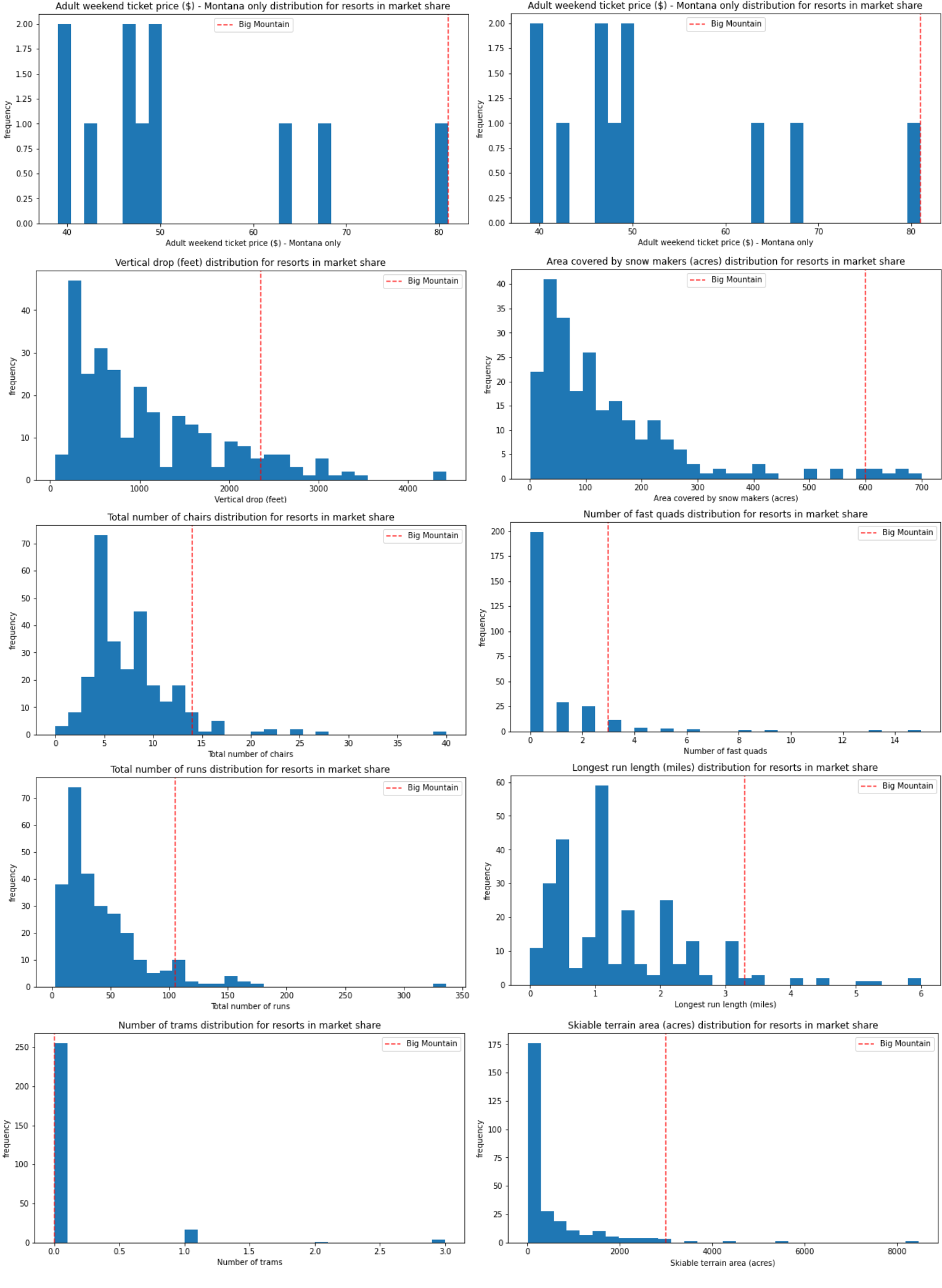


Figure 5 : Comparison of Big Mountain amongst all resorts.

Some elements need to be taken into account when considering the modeling results:

1. The validity of the model lies in the assumption that the other resorts accurately set their prices according to what the market support. There is a possibility that ticket prices of other resorts may also be biased.
2. The ticket price modeling is based on the weekend price and not the weekday due to the amount of missing data. May be there should we a reflection about a possible variation of ticket price throughout the week.
3. We should model the operational cost increase (or variation) as a function of visitor traffic, considering the investment of a new chairlift.
4. We need to take into account the position in the market of the resorts in Montana, and also more globally. Should we modulate the modeled ticket price for Big Mountain?
5. A progressive increase in ticket price for Big Mountain resort could test the market and the ability to pay of the visitors. A survey could also be achieved to know more about what facilities are directly related to the willingness to pay more by the visitors.