**Introduction**

As the operational cost of the Big Mountain resort has increased for $1,540,000 due to the new chair lift, a model has been built as a guidance on (1) how to select a better value for ticket price and/or (2) either cut costs without undermining the ticket prices or will support an even higher price.

How much the ticket price should increase to cover the additional cost of the new lift? As the expected number of visitors over the season is 350,000, and each visitor is expected to get tickets for 5 days, increasing the ticket price by $0.88 will be sufficient to cover this additional cost within one season. However, the question is that whether this new ticket price is well supported.

**Data wrangling**

In the first step, I checked the data and confirmed that (1) this data can tackle the desire question with target values and useful features, and (2) the ratio of missing data is acceptable to be analyzable (< 16%). The data includes the resorts from the US nationwide, and the number of resorts per state is plotted below. In general, the states in the north and north-eastern regions have more resorts than the other regions.

Logo, icon

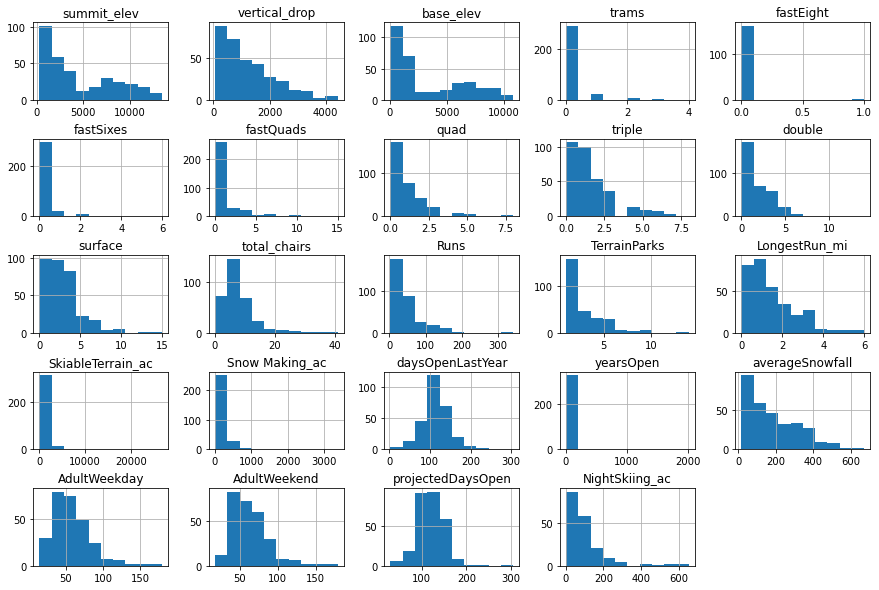
Description automatically generated

The ticket price distributions (both weekday and weekend) per state are plotted below.

Chart, bar chart

Description automatically generated

To further check whether any column contain outliers or unreasonable values, the data of each column was plotted as a histogram. It appears that “SkiableTerrain\_ac”, “Snow Making\_ac”, and “yearsOpen” columns might have unreasonable outliers and the distribution is heavily right skewed.



Further in-depth inspection identified these outliers, and they are either removed or corrected with the information found on the official website of the corresponding resort. See the updated distributions below.

A picture containing crossword puzzle, window

Description automatically generated

I also added the population and area data per state (retrieved from wikipedia) into the current data frame, so later I can include these data into the model.

In the last step of data wrangling, as the ticket price of weekday and weekend are highly similar (see figure below), and the data in the weekday column has more missing values than the weekend column, only the weekend ticket price will be analyzed in the subsequent steps.

Chart, scatter chart

Description automatically generated

In sum, 277 entries of data will be further analyzed.

**Exploratory Data Analysis**

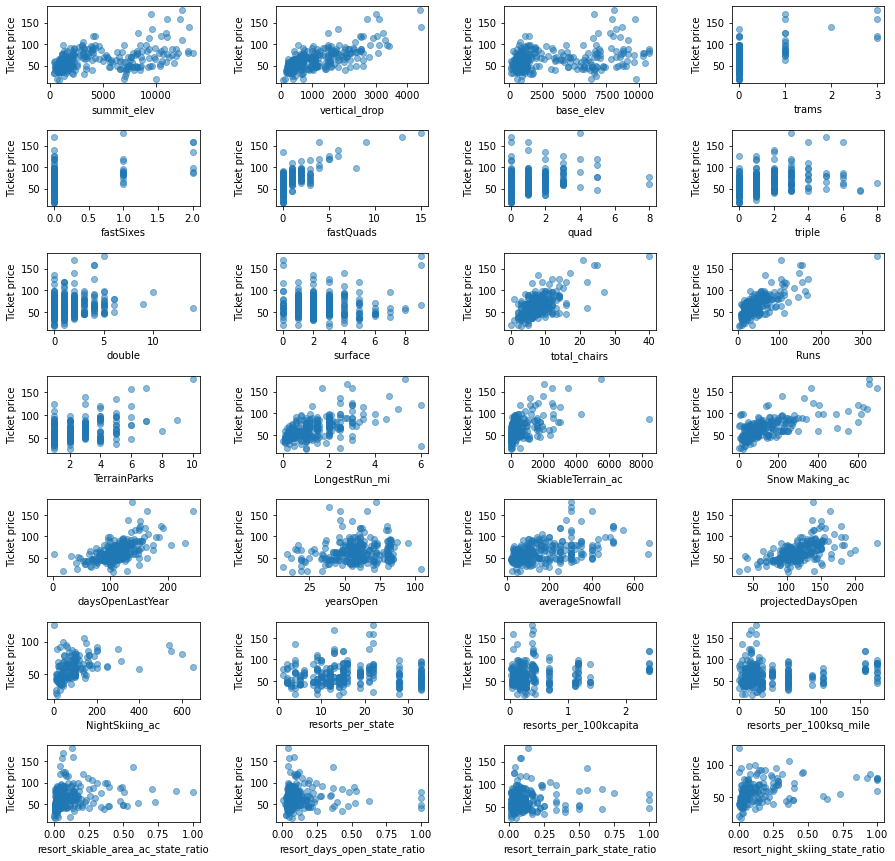
As there are many features for predicting the ticket price, I used a principle component analysis (PCA) to exploratorily reduce the dimension of the features, and it appears that the first 2 PCs account for 77% of the variance, which are mostly associated with resorts\_per\_state, state\_total\_days\_open, state\_total\_terrain\_parks (1st PC) and resorts\_per\_100kcapita, resorts\_per\_100ksq\_mile (2nd PC). Therefore, they are likely to be the crucial predictor for the ticket price in the subsequent model.

Also, as different states have very different features of resorts, as projected in the PC space (see below), estimating the ticket price with nationwide data or within the Montana state might yield different conclusions.

A picture containing chart

Description automatically generated

The scatter plots between ticket price and each feature are visualized below. There are a few features (e.g., Snow Making\_ac, Runs, total\_chairs) appeared more associated with the ticket price than the others.



**Pre-Processing & Modelling**

The rest of the missing values were filled with the median of each column. I used a 5-fold cross-validation approach. The r-squared, mean absolute error and mean squared error were used as metrics to assess the performance of the model.

Two models were developed. In first one, I used a linear model to model the relationship between features and ticket price of resorts. I used hyperparameters to find the best number of selected features for the linear model. It appears that 8 features were optimal for this linear model.

In another mode, I used a random forecast model to model the relationship between features and ticket price of resorts. The number of estimators were grid searched between 10 and 1000, and the imputer strategy of both mean and median were explored. It turned out that 69 estimators with median strategy was the best parameter set. Below is a figure of feature importance of the best random forest model.

Chart, histogram

Description automatically generated

In the end, the random forest model performed slightly better than the linear model, with mean absolute error as 9.54 vs 11.79.

To ensure that the current model was based on sufficient data, I used a learning\_curve() function to assess the relationship between the size of training data and the cross-validation scores, and it appeared that the cross-validation scores reached ceiling when the training data size is above 60 (see figure below). Therefore, the current model based on 277\*0.8 = 222 data size is very robust.

A picture containing timeline

Description automatically generated

**Recommendation**

The ticket price of a ski resort should be supported by its facilities. By analyzing the facilities across 277 ski resorts nationwide, we found that the current Big Mountain resort’s adult weekend ticket price ($81) is higher than both national and Montana average (Fig 1). On the other hand, its facilities also outperform than most of the resorts in the US, including vertical drop, area covered by snow makers, number of chairs, number of fast quads, number of runs, the length of the longest run, and the skiable terrain area. Therefore, there might be a room to further increase the ticket price.

Chart

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

I used a random forest machine learning approach to model the relationship between ticket price and facilities across all resorts. According to this model, the facilities of the Big Mountain resort worth $95.87/ticket (with a statistical estimation error of ± $10.39). Therefore, it appears that increasing the ticket price of Big Mountain resort up to $95 can be well supported by its facilities.

I further created a function to understand how adding or removing some of the facilities of the Big Mountain resort would change the supported ticket price and the revenue. For example, Fig 3 showed that closing one run makes no difference. Closing 2 and 3 runs will reduce the support of the ticket price and the revenue. Furthermore, if Big Mountain closes 3 runs, it may as well close 5 runs as it makes no further difference.

In conclusion, my suggestion would be increasing the ticket price for $1 for this season, as it sufficiently covers the additional costs. I am conservative of suddenly raising the ticket price to $95, even the model suggested so. It is because we do not have the data to understand how the customers would react to the ticket price change, and all the current model assumed that the number of tickets sold will remain constant. Nevertheless, in the next season, once the resort collects the data of the number of tickets sold per season before and after this price change, it is possible to model how the price change to $95 will affect the number of tickets sold.

Chart, line chart

Description automatically generated