Guided Capstone Project Report

By Anthony Medina

* Problem Statement

Big Mountain Ski Resort has recently added a new chair lift and has tasked us with determining a new ticket price that will keep their profit margins within 8%, after recouping expenses for their 1.5 million dollar lift.

In order to accomplish this, we used a data set that had information about a variety of other resorts and used their feature sets and prices to make a comparison and a new ticket price recommendation.

* Data Wrangling

When we started wrangling the data we were motivated by a couple of questions.

* Do we know what value we’re looking to predict?
  + From advice by the business, we were looking at ticket prices.
* Do we have any potential useful features?

We began by loading the data and taking an initial look at what values we were provided with, and how many of what values were missing. We saw that ticket prices had quite a few missing values, so we had to make a decision about using Weekday or Weekend prices. We started looking into ticket prices by state, and dug into each type. We looked at all of the numeric features and looked for variance and skew of the data. We looked further into skiable area, days open, terrain, and available night skiing. We saw that about 14% of rows were missing both ticket prices, so we dropped them. See Figure 1

* Exploratory Data Analysis

Our exploration started with summary statistics. Which state led in area, population, resorts per state, skiable area, night skiing area, total days open, and resort density. Then we scaled the data for Principal Component Analysis. After scaling the data, we looked at the mean and standard deviation of all of the features that we had. We examined average ticket price by state.

* Model Preprocessing with feature Engineering

We added some of our features to the data based on imported state data, and the features we had in our data set. We added, ratio of resort skiable area to total state skiable area, ration of resort days open to total state days open, ration of resort terrain park count to total state terrain park count, ratio of resort night skiing area to total state night skiing area, resorts per state, resorts per 100K sq miles, total chair ratio, fast quads runs, and a fast quads skiable ratio. We ended up exploring all of the numerical values with ticket prices and looked for correlations. To do this we looked at a heat map and scatter plots (Figure 2 and 3 respectively)

* Algorithms used to build the model evaluation metric

We initially used the mean to see if it was a good predicting value. In this case, it was not. We checked using R-squared, Absolute Mean Error, and the Mean Squared Error. Once we then tried the median. It was very similar to using the mean. We continued modeling and found using Cross Validation that 8 was the optimal number of features to use for modeling. (Figure 4). We used a random forest model and it preformed better than our initial linear models. We found that our top features were the same as our linear model.

* Winning Model and Scenario Modelling

We ended up going with the random forest model using the features: vertical drop, snow making, total chairs, fast quads, runs, longest run, trams, and skiable terrain.

We explored additional scenarios for cost savings.

* Scenario 1 (Close up to 10 of the least used runs)
  + We found that closing 5 runs wouldn’t change the price of the ticket very much at all is our recommendation.
* Scenario 2 (Add a run that increases the vertical drop by 150 ft, and install a new chair lift)
  + This would increase the ticket price by $2, but would see really small increase in revenue. We do not recommend this.
* Scenario 3 (Scenario 2, but also increasing snow making)
  + Similarly to scenario 2, this would be a lot of effort for no difference in revenue.
* Scenario 4 (Increase the longest run by .2 miles and guaranteeing snow coverage by adding 4 acres of snow making capability)
  + We saw no difference whatsoever. This is also not a recommendation.
* Pricing Recommendation

We concluded that a price of $95.87 was more reasonable than $81.

* Conclusion

We were tasked with finding a more reasonable price for the Big Mountain Ski Lodge to stay competitive in the market, with the addition of their new ski lift. We were able to adjust the price of the ticket and find some additional cost savings by making some recommendations about shutting down a few ski runs.

* Future Scope of Work

With the model complete, we constructed a function that can be used to further scope changes to the resort. If Big Mountain is considering the construction of an additional run, ski lift, add more snow coverage etc, you can update the parameters in the function and use that to predict future revenue. We will need to train someone on the Big Mountain staff to be able to use this self sufficiently.

Figure 1 (Initial distribution of features)

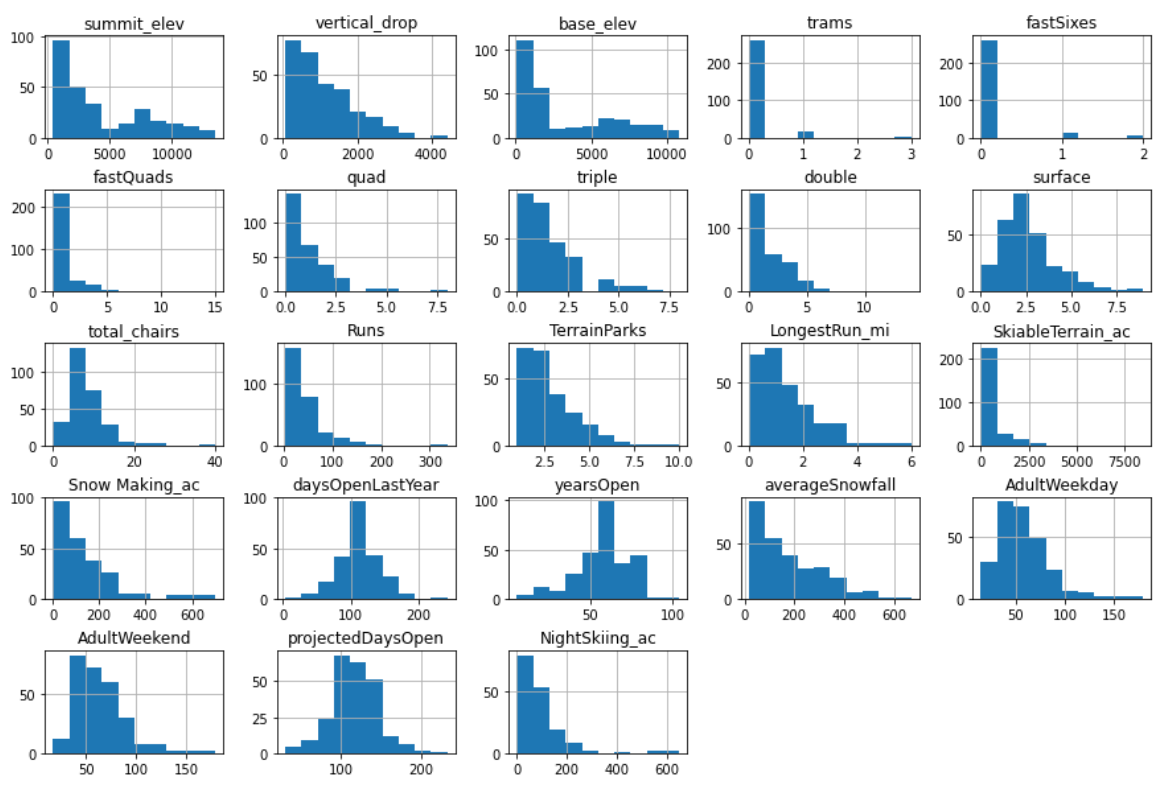


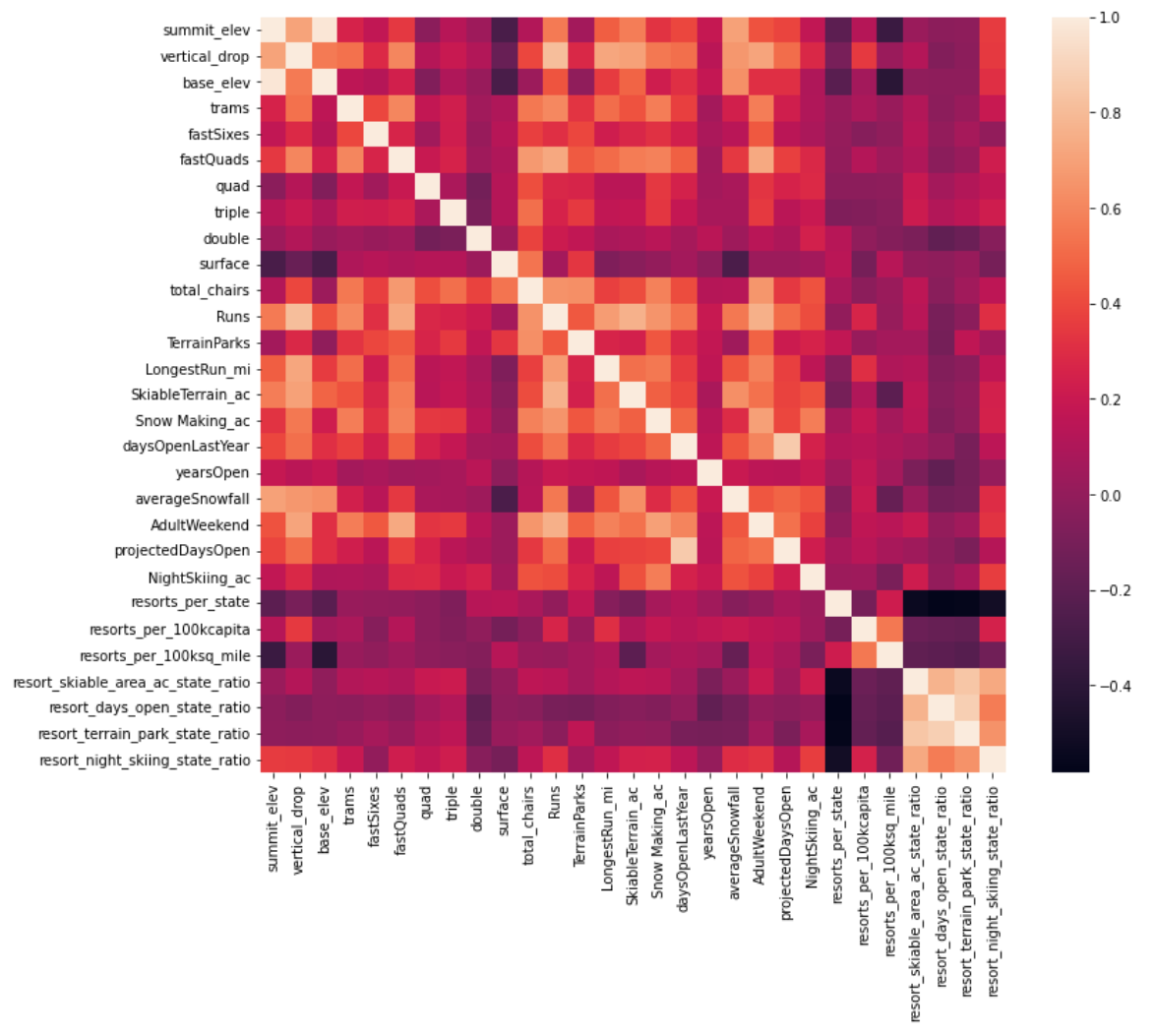
Figure 2 (Heatmap of features influence on ticket prices)

Figure 3 (Scatter plot of ticket price and features)

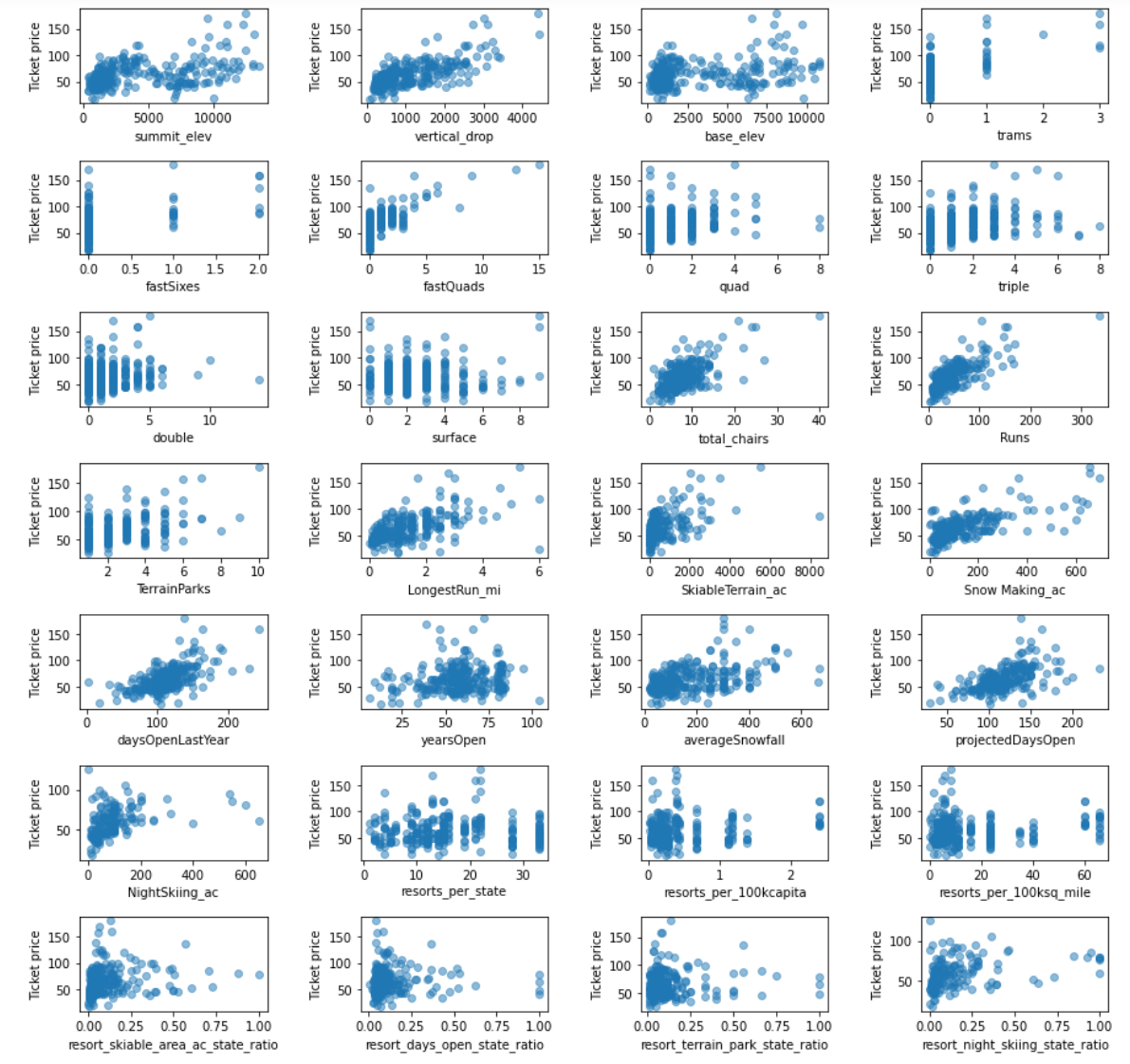


Figure 4 (Optimizing the number of features to look for)

