Lupe Covarrubias

How can the Big Mountain Ski Resort increase ticket prices while maintaining the national average within the market, while capitalizing on their facilities and/or decreasing operational costs without undermining the ticket price before the next skiing season begins?

The data that was given for the ski resorts started with 330 rows and 27 columns. The dataset did contain the desired resort, Big Mountain Resort in Montana. In the process one column was removed which was fastEight because there was little to no data for almost every resort. There were also a few rows removed including the resort that was either brand new or had not yet been open and the rows where both ticket prices were missing (weekday and weekend). There are also some columns that have suspicious outliers such as trams and fastSixes, but there was no immediate action taken; this is just something that the team should be wary of in the future. As shown in figure 1, many states have a large variation ini ticket prices across resorts while other states have little to no variation. Also, not all the ski data states were accounted for, therefore they were compared to the web data and immediately replaced so that there were no missing states in the dataset. So far in this project, most of the undesired rows and columns have been removed and the last thing that was reviewed was the percentage of missing values in the dataset for each row, leaving with 277 rows. Once the amount of missing values has been narrowed down to the finest degree necessary, the data could be better represented to convey the best range for a desired ticket price for the Big Mountain Resort. By selecting a better ticket price the resort can decide whether they need to reduce operating costs or capitalize more on other facilities.

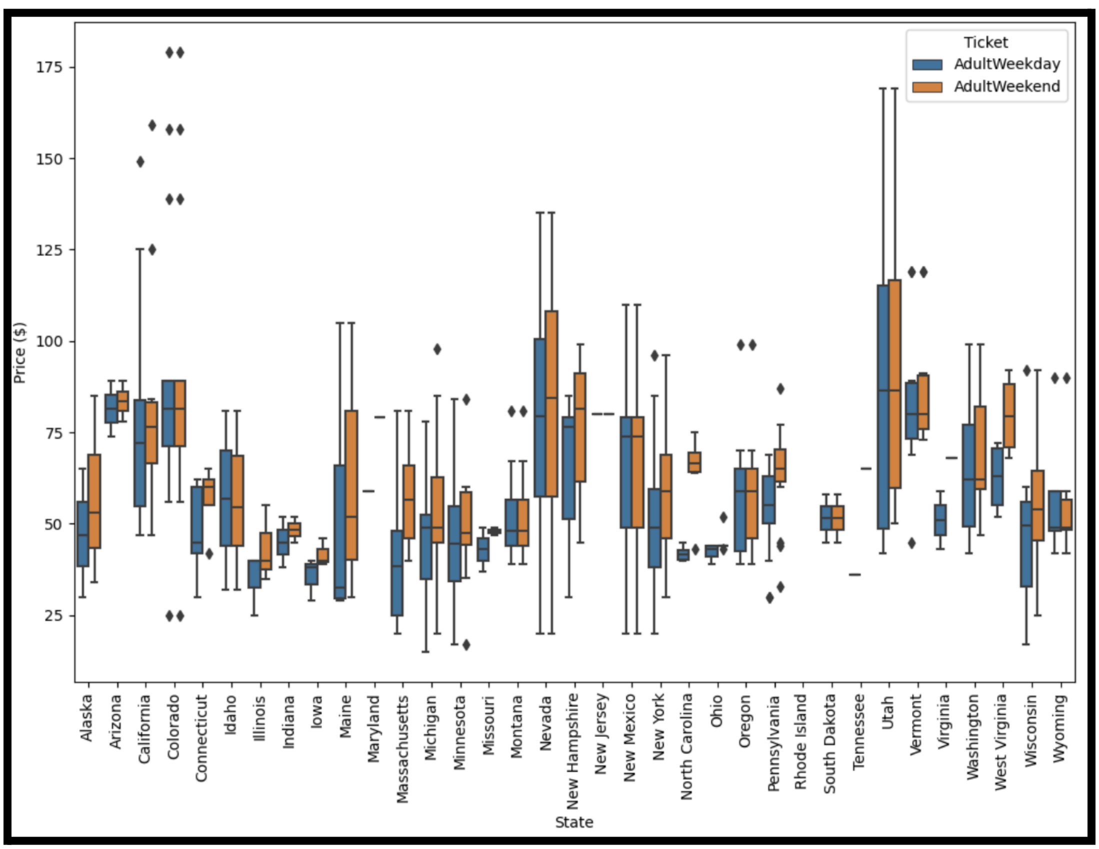


Figure 1: box plot of states mean adult weekend and weekday ticket prices

In the ski and state summary data there were a lot of numerical features including resorts_per_state, state_total_skiable_area_ac, state_total_days_open, state_total_terrain_parks, state_total_nightskiing_ac, resorts_per_100kcapita, resorts_per_100ksq_mile. The categorical features consisted of US states. Although, there has not been a clear correlation between state and ticket price, there are other correlations identified via heatmap and scatterplots between ticket price and other variables such as summit elevation, vertical drop, runs, total chairs, and fastQuads.The PCA conveyed that the PC1 is the most important in terms of describing how the states vary in data, as shown in Figure 2. The PCA graph also shows that the more clustered states such as south dakota, nevada, ohio, arizona, etc. are similar in their data. Therefore Montana which is not clustered around any other states is not very similar to other states in terms of data but may be somewhat similar to Idaho, Minnesota, and maybe Massachusetts and Wisconsin too. The heatmap (Figure 3) conveys that adult weekend ticket prices are correlated to fastquads, runs, snow_making_ac, resort_night_skiing_state_ratio, and total_chairs. All these aspects need to be kept in mind when going forward in the project. There are a number of features that could be used as targets including number of resorts, runs, total chairs, vertical drop, etc, which are highly correlated with ticket price as shown in figures 4-11. These will all need to be looked at and taken into account when choosing the target feature.

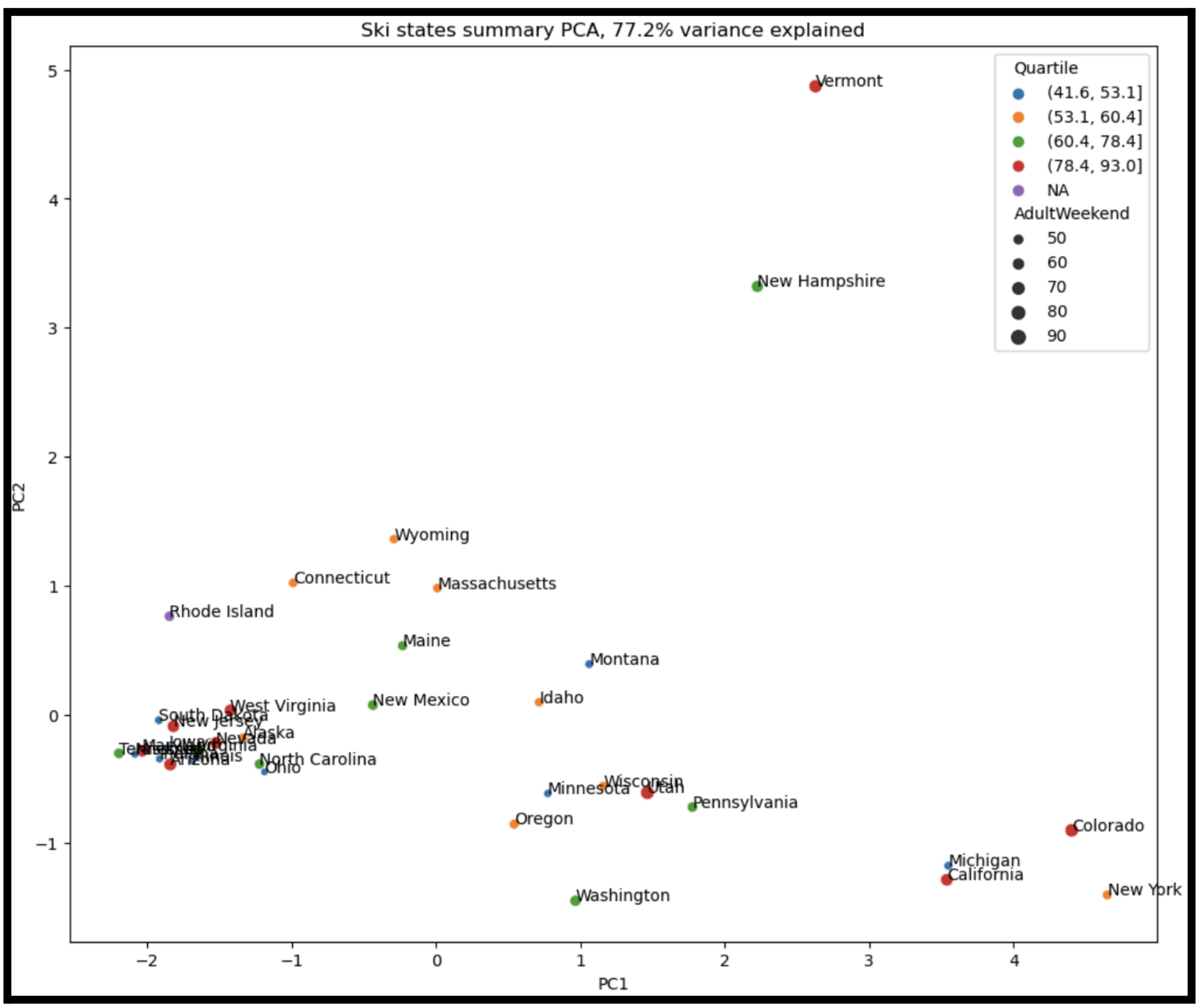


Figure 2: Principal component analysis with 7 features

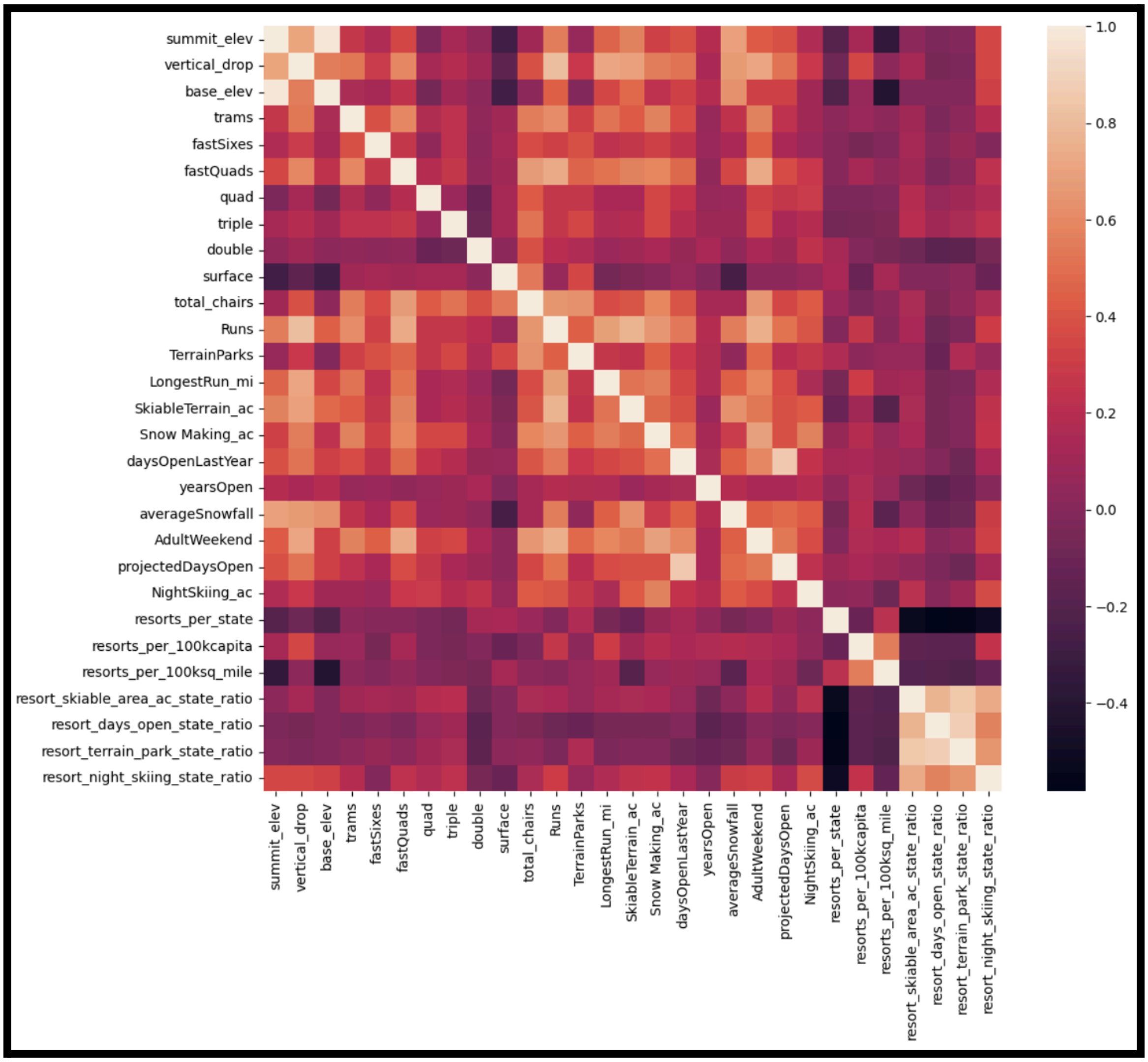
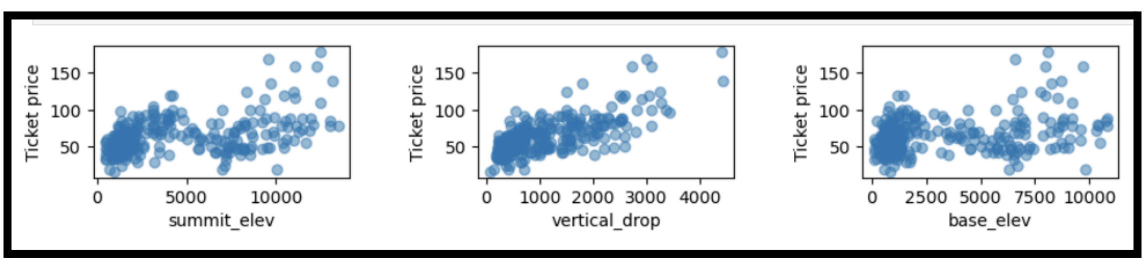


Figure 3: Heatmap of correlation of all columns in the ski and state data



Figures 4-6: scatter plot correlation of ticket price vs summit elevation, vertical drop, base elevation

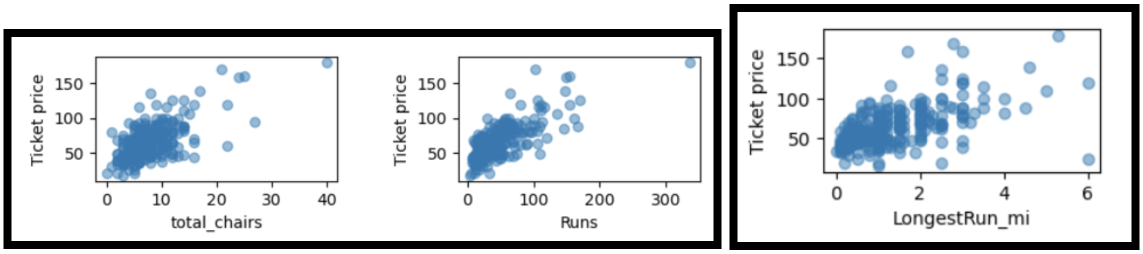


Figure 7-9: scatter plot correlation of ticket price vs total chairs, runs, longest run

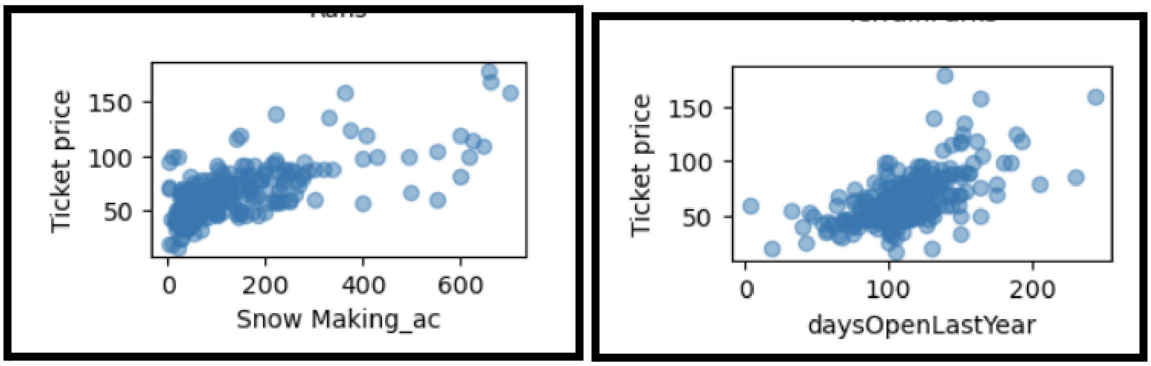


Figure 10, 11 : scatter plot correlation of ticket price vs snow making, days open last year

We started with a baseline for average price which was about $64 for the ticket price. When the dummy regressor was used with the trained variables it also had a mean of $64, as well as the prediction. Then using the train_test_split linear model and training the data with a train sample size of 70% and a test sample size of 30% of the data we found that the mean absolute error changed from $19 to $9. This means that if the price were to be estimated it would be about $9 above or below average. The linear regression model was built off of r^2, mean absolute error, and mean squared error to calculate performance of the model and using the median values to replace any NaN values we can see that our simple linear regression model explains over 80% of the variance on the train set and over 70% on the test set. The cross validation model using selectkbest conveyed that the best amount of train/test folds was 8 using r^2 as the model performance; it was also portrayed that the best sample size was about 50 for the training set, as shown in Figure 13. This model had a mean r^2 value of 0.6925 with a standard deviation of 0.07 (using cross validation scores). The model stated that the feature of most importance on ticket price was vertical_drop followed by snow_making_ac. The random forest regressor was also tried using a similar pipeline to previous models (imputer(median), random forest regressor). This model had a mean r^2 value of 0.7097 with a standard deviation of 0.06 (using cross validation scores). The estimated performance via cross-validation conveyed that the features of most importance were fastQuads and then Runs (shown in Figure 12), which is inconsistent with the previous model. Between the linear regression model performance and the random forest regression model performance there was about a $1 difference in the cross validation mean absolute error with the random forest model being lower, and it had a higher value in r^2 value meaning it represented a better model of the standard performance and this is why we will be moving forward with the random forest regression model.

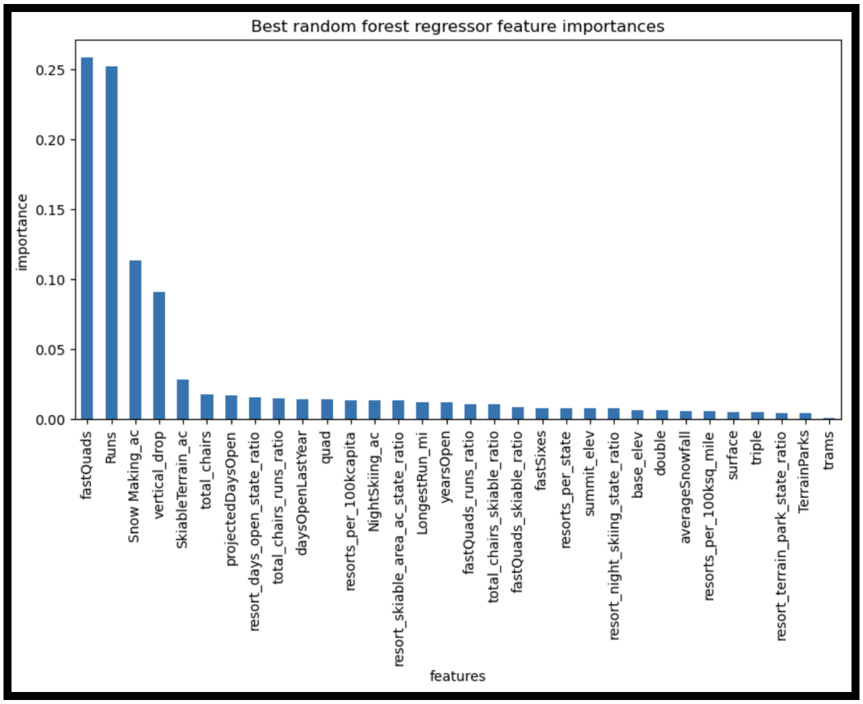


Figure 12: bar plot of features vs influence on ticket price (importance) determined by random forest regressor

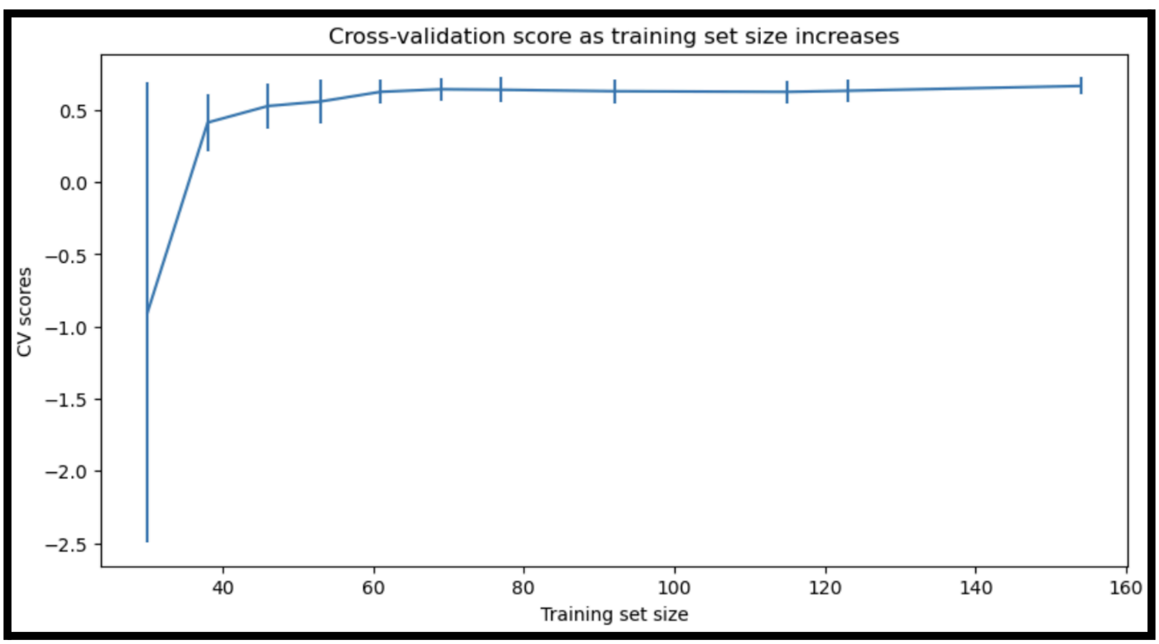
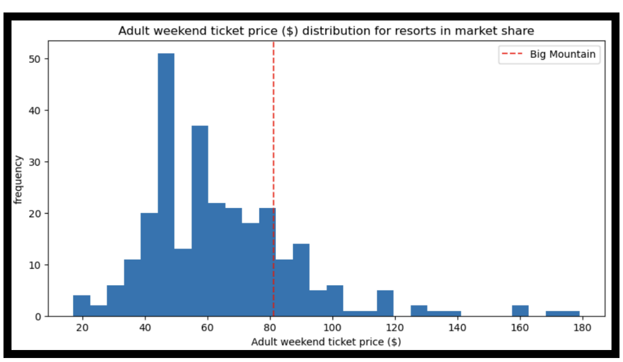
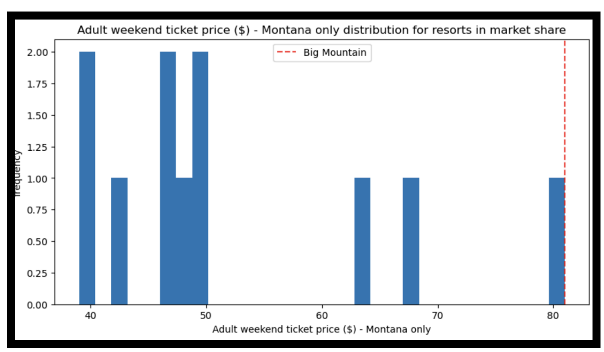


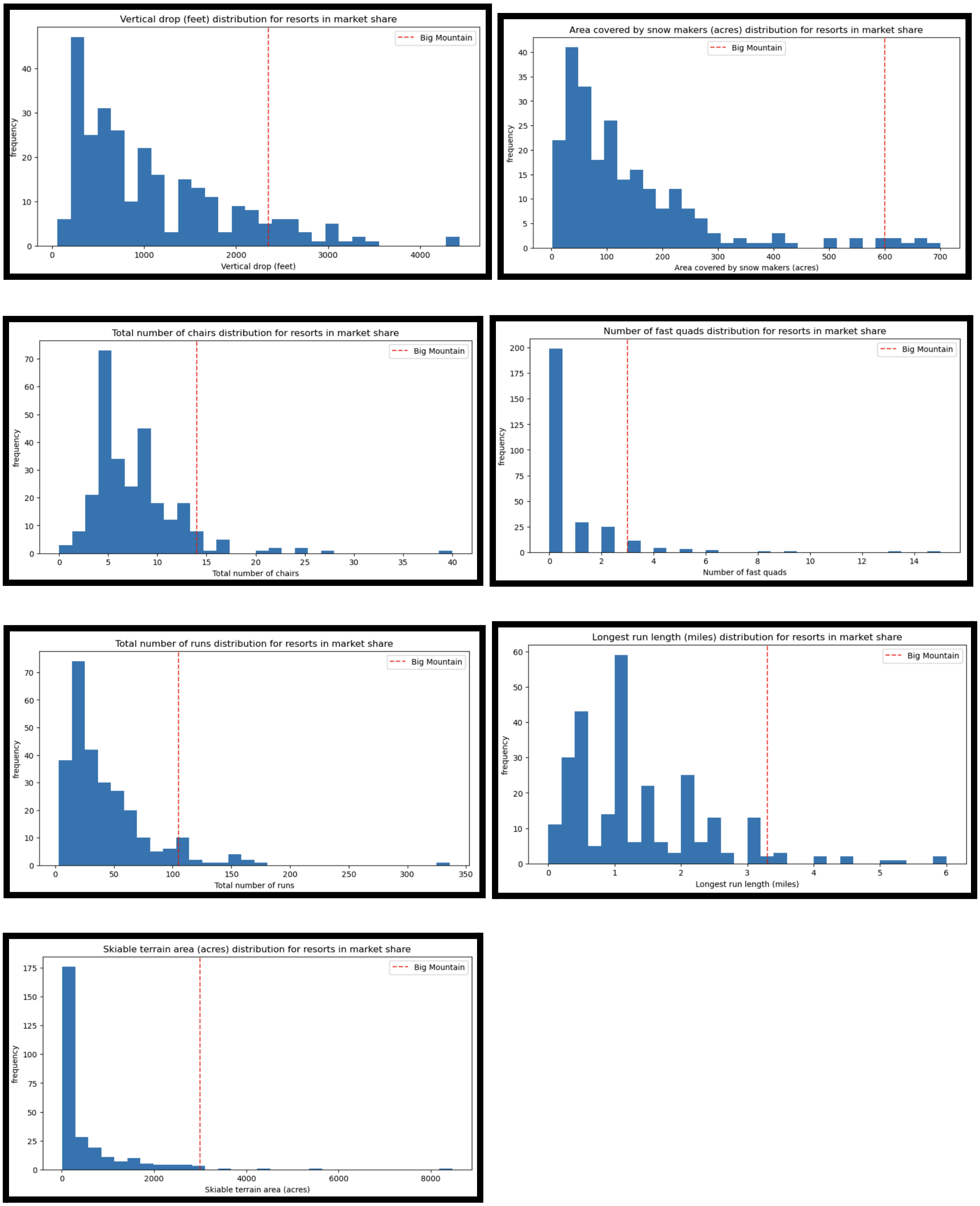
Figure 13: cross validation score determination of best size for training set

The Big Mountain Ski Resort offers a wide variety of facilities on the mountain that skiers can enjoy. As conveyed in figures 14-22 the resort is on the tail end of most distributions meaning it has a lot to offer in terms of amenities on the mountain. The Big Mountain Ski Resort in Montana currently charges $81 for a ski ticket. The Big Mountain Resort modeled price is approximately $96. Even with the expected mean absolute error of $10.39, this suggests there is room for an increase or decrease in price. Although there is no fixed price for the resort to set ticket prices to, there are certain facilities that patrons are more likely to pay a higher price for, such as higher vertical drop. The additional cost of adding the new chair lift increases operating costs by 1.54 million dollars per season. Big Mountain resort had suggested some scenarios for cutting costs or increasing revenue(based on ticket price) which include: permanently closing down up to 10 of the least used runs and this doesn't impact any other resort statistics, increase 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, same as the previous but adding 2 acres of snow making cover, increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres. After further analysis of the modeled scenarios they convey that we lose money in ticket price by further reducing the number of runs especially after reducing by 6 or more runs, if the resort were to close any runs it would have to be 5 or less because that would have the least effect on ticket price, which is conveyed in Figure 23. They also conveyed that increasing the longest run by 0.2 miles makes no difference. The best scenario was to increase the vertical drop by 150 feet, this supported ticket price increase by 2 dollars and allowed for 3,474,638 dollars additionally in profit. We also added 2 acres of snow making to see if this further supported an increase in ticket pricing but it showed no difference. Therefore, the best scenario that supports an increase in ticket price at the Big Mountain Resort in Montana would be an increase in the vertical drop by 150 feet without the additional 2 acres of snow making.

The data assisted us in gaining key insights on operating facilities on the mountain but not what is at the base of the mountain. The data would have been more helpful if there was more information on the lodge and what it had to offer as well as ski rental equipment. The modeled price was much higher than the current price because Big Mountain must be undercharging for some of their facilities, maybe it is not capitalizing enough on what the resort has to offer such as conveniences and services which was not included in this data. To get a better understanding of this large gap in actual versus modeled ticket price we need data on the conveniences and services. Although we don't have this information the data is still useful in other ways, it conveys that they are undershooting their ticket price and they have the room to increase it to assist in the operating costs of the new lift chair. This model can be modified if other data has been found that may better represent the facilities at the Big Mountain Resort. This model can be accessed on GitHub for others to use and explore.







Figures 14-22: Big Mountain Ski Resort mapped by dotted line on distribution to compare facilities vs other resorts facilities

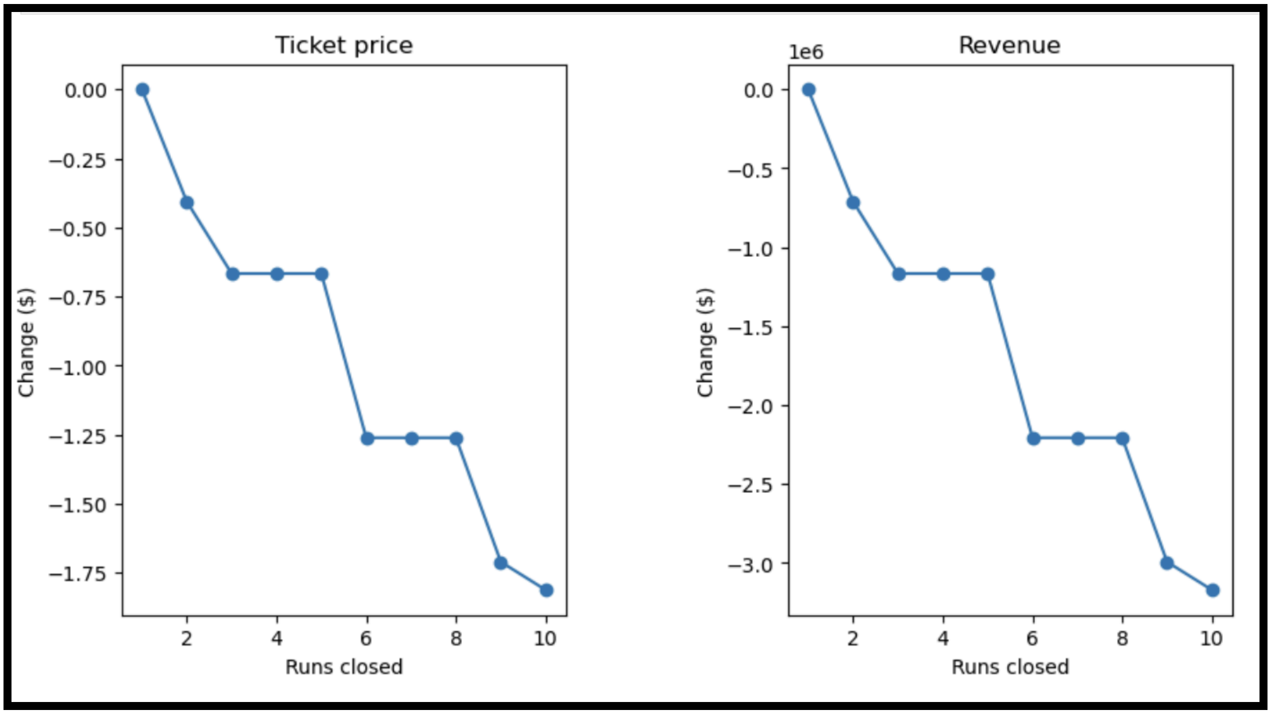


Figure 23: runs closed vs change in ticket price