Big Mountain Resort Ticket Price Strategy Report

Big Mountain Resort offers spectacular views of Glacier National Park and Flathead National Forest which hosted about 350,000 people for skiing or snowboarding. Big Mountain Resort has recently installed an additional chair lift, which increases their operating costs by $1,540,000 this season. The company want to change current pricing strategy to capitalize its facilities as much as possible.

Therefore the task for data team is to come up with a pricing model for ski resort tickets in United States’ market segment. The predictive model will be based on a number of facilities and properties which will also explain what facilities matter most to ticket price. This model will be used to provide guidance for Big Mountain's pricing and future facility investment plans.

In order to find the most precise price adjusting strategy, I collected data including ticket price for resorts across United States and informational features that could be used to predict the price. The whole project including data wrangling, data exploration, pre-processing and training data and modeling.

1. Data wrangling

I ran the standard data quality check, data filtering and cleaning procedure. In the original ski\_data, we have 27 columns and 330 rows. Our resort Big Mountain Resort is present in ski\_data. There are two price column in the data AdultWeekday and AdultWeekend. Since all resorts in Montana have the same AdultWeekday and AdultWeekend and AdultWeekend has less NA values, I select AdultWeekend as target feature. After removing columns which are most NA and/or 0, correcting items which screw the distribution, rows containing apparently wrong information and rows containing no AdultWeekend information, we got 277 rows. I also downloaded public ski resorts data across all states and run similar data cleaning and filtering.

1. Data exploration

**I ran data exploration on the cleaned data to find trends in the data that are very important for building the model. The first question I checked is whether the ticket price associated with states. Figure 1 indicate the PCA plot for state summary resort data. From the figure we can find that most of states cluster together except** New Hampshire and Vermont. After double check the data, resorts\_per\_100ksq\_mile and resorts\_per\_100kcapita are found contribute a lot to the principle features. At the same time, Vermont and New Hampshire have high value on resorts\_per\_100ksq\_mile and resorts\_per\_100kcapita. Therefore, I conclude that all states together will be treated equally. Secondly, I checked the correlations between ticket price and all the features. As shown in Figure 2, there's a strong positive correlation between ticket price and vertical\_drop. fastQuads, Runs and total\_chairs appear quite similar and also useful. These trend give us confidence that the features in the data have great information which can be used to build model for price prediction.

1. **pre-processing and training data.**

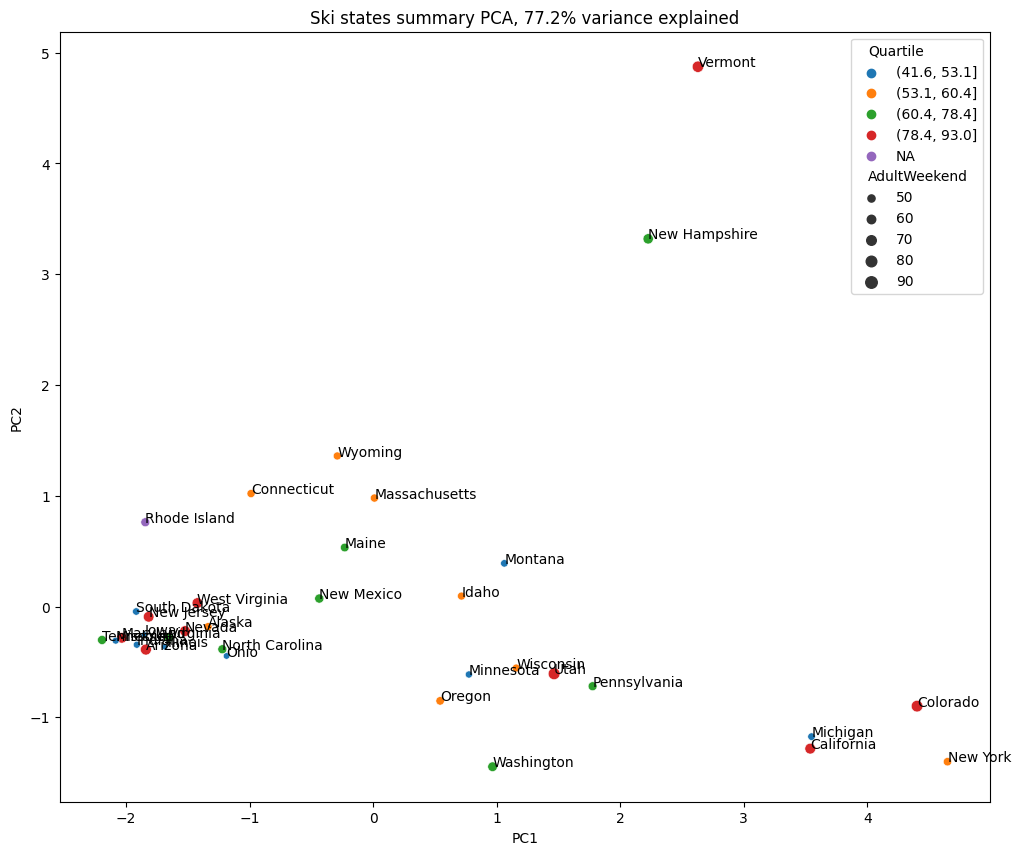


Figure 1. PCA plot of state resort summary data

Before building any model, I used mean as prediction to have a baseline idea of prediction performance. The R-sqaure is -0.0031 for test set while the mean absolute error is 19.1 which indicate the difference between using mean as prediction and real price is around $19. Start from this, I will build models to improve the performance step by step. The first model is linear regression model. The performance of using linear regression is significant better than using mean. I found the R-square has been improved to 0.72 (mean imputing). The mean absoulte error has been improved to 9.42 (mean imputing). The improved performance is encouraging, but we also found the performance on test is much worse than on train, which is a signal of overfitting. In order to handle this problem, we made use of cross-validation. I found the following 8 features have the best performance: vertical\_drop, Snow Making\_ac, total\_chairs, fastQuads, Runs, LongestRun\_mi, trams and SkiableTerrain\_ac.

A second model I built is random forest model. I made use of grid search based on cross validation to select the hyperparameters such as number of estimators and preprocessing stragety including mean vs median imputing and scale vs none scale. Grid search indicate that median imputation and none scale present the best performance. I also observe that the mean and standard deviation of performance on each fold can estimate the performance on test closely.

Since we have got best-performance parameters based on grid search for both linear regression model and random forest model, we made use of the best-performance parameters and rerun cross-validation. We also use the selected best parameters to validate the model's performance on test set. I found that both cross-validation performance and test set performance indicate that random forest has the best performance. I will select random forest to move forward.

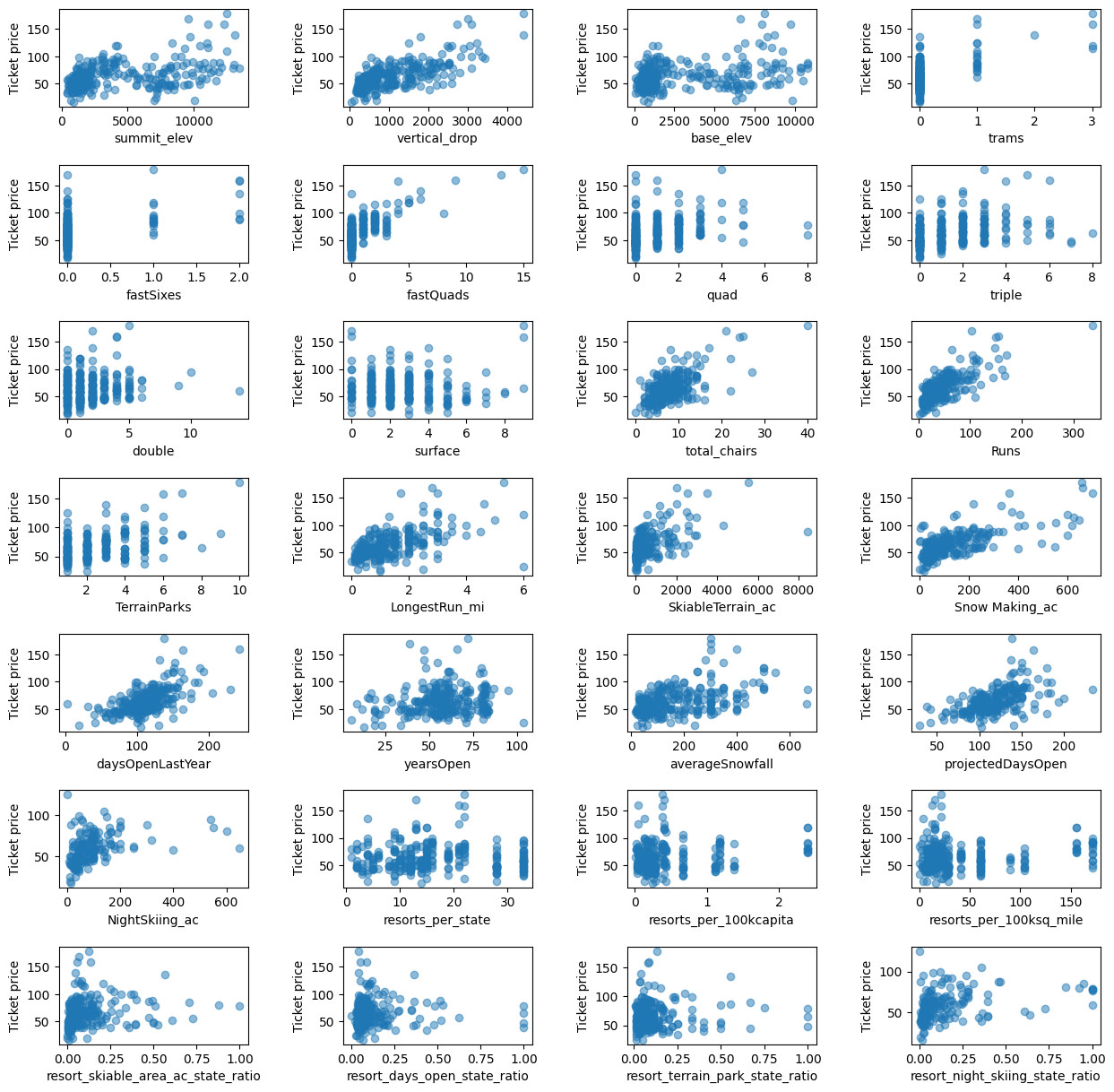


Figure 2. Correlations between ticket price and features

1. Modeling

Big Mountain currently charge $81 for ticket price. The model suggests a ticket price $95.87 based on maketplace data. Even considering expected mean absolute error of $10.39, there is still a room for increase the ticket price. Considering we will welcome 350,000 visitors and each visitor on average will buy 5 day tickets, the average cost increase per ticket due to installation of new chair lift is $1,540,000/350,000/5 = $0.88. Based on the expected price from my model, if we increase ticket price to $95.87, we will increase our revenue by $24,482,500. Even if we use the most conservative way of increasing price and only increase ticket price to $95.87-$10.39=$85.48, we can still increase revenue by $6,300,000.

Based on the senario modeling, I will suggest closing the shortest run, since the senario modeling results indicate closing the shortest run will not loose any support on current price. A second strategy is to increase the vertical drop by adding a run to a point 150 feet which requiring the installation of an additional chair lift. The senario modeling result shows that this strategy will support an price increase of $1.99 which corresponding to $3,474,638 more revenue.

Conclusion

Based on the collected data, I successfully built random forest a model to predict the price. The model indicate that current market data support a booming up of ticket price. Other than that the model also tells several scenario that the resort can boost their revenue by removing some run or increasing vertical drop.

For the next step, we can collect operation costs for each category and include these costs in the model. For example, if we know the cost of removing/adding a run, we can include these information to learn how its difference across resorts will contribute the ticket price.

Although Big Mountain was already fairly high on some of league charts, for example highest in Montana, I think current price still take full use of Big Mountain's facility. For all the important features we discovered that will support the ticket price. Big Mountain always rank high acorss resorts in the whole country, which indicate Big Mount's facility can univerally satify customer's request on all the aspects. From a single aspect, Big Mount may be not the best. But considering all these price supporting features as a whole package, Big Mount's ticket price should be much higher than its current price.

We can build an app or web server to incorporate the model as backend server. We can make all the tuningable features or parameters available in the frontend. In such a case, business leaders can got the expected price from their novel combination business strategies.