Guided Capstone Project Report

Pricing Model for Big Mountain Ski Resort Tickets

Fentaw Abitew
August 2021

Problem Statement

The purpose of this data science project is to come up with a pricing model for ski resort tickets in their market segment. Big Mountain suspects it may not be maximizing its returns, relative to its position in the market. It also does not have a strong sense of what facilities matter most to visitors, particularly which ones they're most likely to pay more for. This project aims to build a predictive model for ticket price based on a number of facilities, or properties, boasted by resorts (at the resorts). This model will be used to provide guidance for Big Mountain's pricing and future facility investment plans.

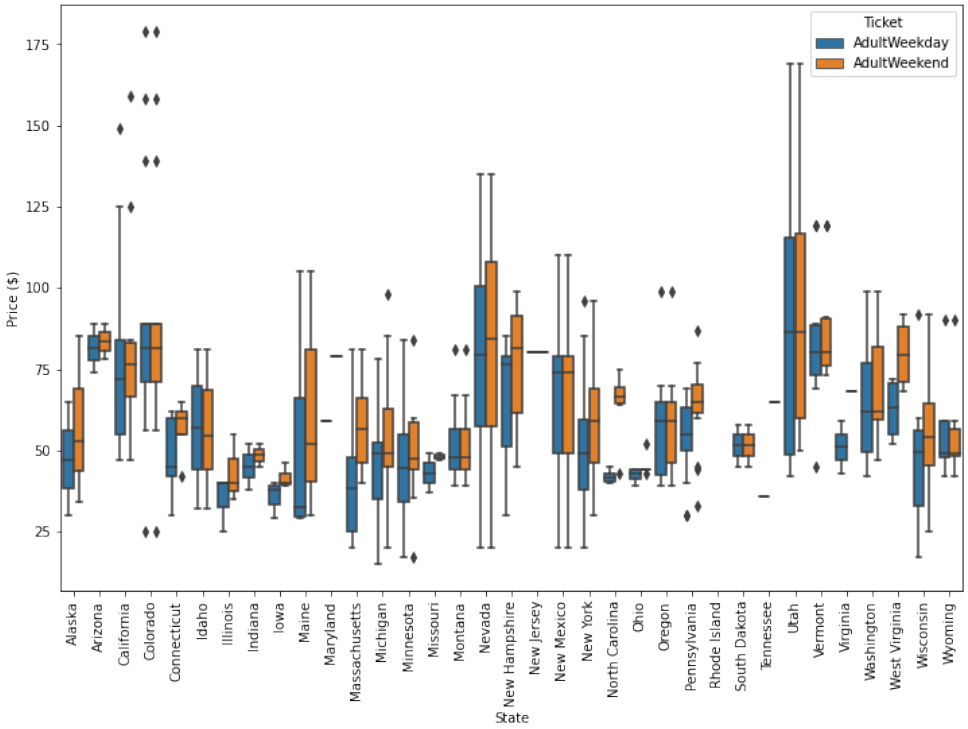
1 Data and Assumptions

In this project, we use data from over 300 Ski and resorts in the United States with more than 25 relevant features (explanatory variables). The Big Mountain Resort also available in the data, interestingly enough with no missing values on its variable.

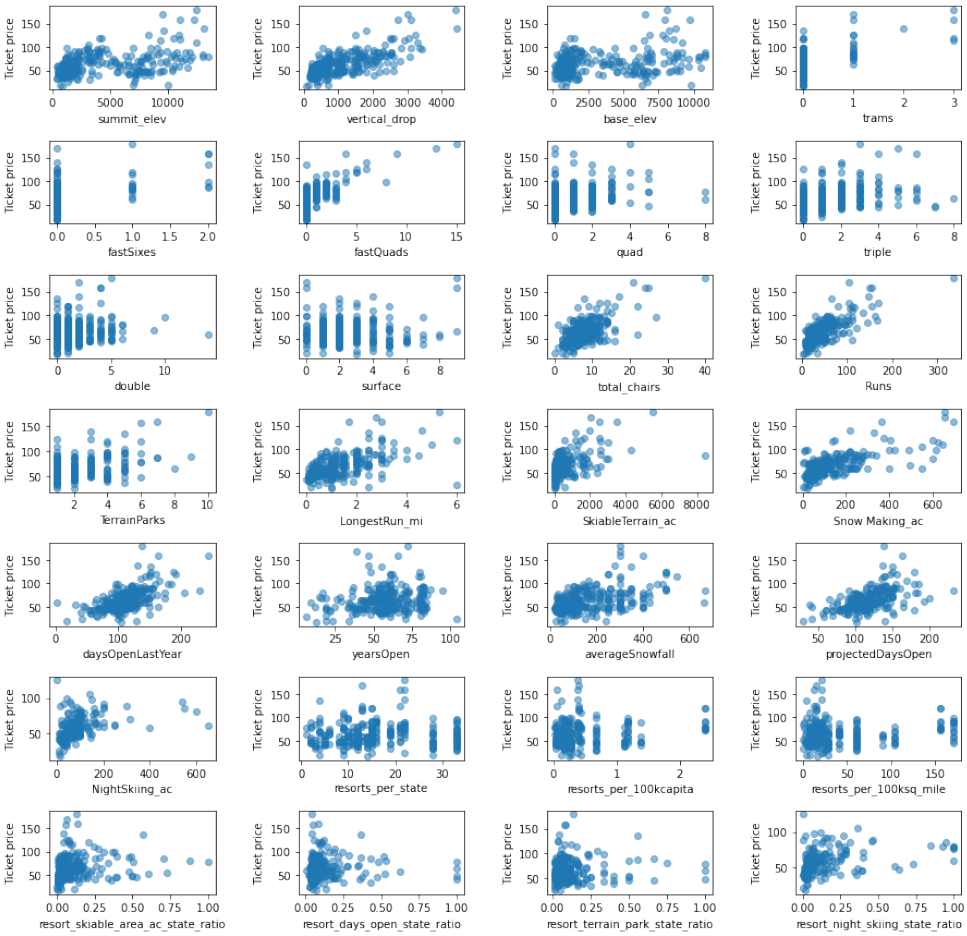
It is assumed that other resorts pricing model is based on the features presented and it is given that all resorts are part of the same market share.

2 Data Exploration and Pre-processing

While we explore the data by checking the distribution of each features, we found wrong information as outliers and missing information's. Also obtained some additional US state population and size data with which to augment the data set, which also required some cleaning. We handled few of the missing data and corrected the wrong information. Given Weekend prices have the least missing values of the two target feature, weekend price is taken as a dependent variable. With few exception, most prices appear to lie in a broad band from around 25 to over 100 dollars.



We constructed some potentially useful and business relevant features, derived from summary statistics, for each of the states. We've explored many of these features in turn and found various trends. Some states are higher in some but not in others. Some features will also be more correlated with one another than others. We us Principal Component Analysis(PCA) to disentangle this interconnected web of relationships.(scale, fit, transform).



3 Modeling Ticket Price

i. Mean as a predictor

Before we started with working on machine learning model, we considered the mean as a potential predictor (not a model). The mean predication results highest mean absolute error; around 19 off on average.

ii. Simple Linear Regression Model

In this simple linear model, we performed the following steps: 1. replace missing values with the median for each feature 2. scale the data to zero mean and unit variance 3. train a linear regression model 4. and all these steps were trained on the train split and then applied to the test split for assessment.

The more refined linear model, however, with hyper parameter search using ski learn tool called GridSearchCV simplified the process and made the performance assessment easy. It contains a pipeline that imputes missing values, scales the data, selects the k best features, trains a linear regression model, and last but not least a technique (cross-validation) for estimating model performance.

The result of the model suggest that the following eight features are more important than others for the Big mountain Resort pricing:

vertical_drop	10.767857
Snow Making_ac	6.290074
total_chairs	5.794156
fastQuads	5.745626
Runs	5.370555
LongestRun_mi	0.181814
trams	-4.142024
SkiableTerrain_ac	-5.249780

The negative coefficient for the two variables at the end shows that holding the other variables constant, increasing the two may worse off the resort.

iii. Random Forest Model

Random Forest Model has a number of hyperparameter and help us to go straight from defining the pipline to assessing performance. In line with the above linear regression model, interestingly enough our analysis confirm that the dominant four features are: FastQuads, Runs, Snow making$_a$*candVertical*$_d$*rop*.

We choose Random Forest Model for our pricing model for two reasons:

1. has lower cross-validation MAE
2. less variability

Conclusion and Remarks

Our Analysis confirmed that it is in the best interest of Big Mountain Resort not to engage in more data gathering. The learning$_c$*urveshowsthatwehaveplentyofdataathand*.

According to this price model (Random Forest), Big Mountain Resort price is 95.87 with the expected mean absolute error of 10.39. The current ticket price of Big Mountain Resort is 81. Thus, even with the MAE of 10.39, surely there resort facilities support price increase. From suggested scenarios (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) would be the best option given the fact that it increases support for ticket price by 1.99 with expected return of 3474638. Given only 1.5 million cost of installation of an

additional chair lift, the resort still be in a position of positive return of around 2 million given no other costs.