Finding most important features that affect the ticket price strategy for Big Mountain Resort

Abstract: Using the dataset provided we built a simple model using Random Forest regressor in order to find the resort's features which affect the ticket price the most. Our model revealed 9$ confidence interval.

Introduction: Creating a good pricing strategy is an important problem in financial optimization. A ski resort has to be able to understand what features will increase its income and how to utilize them fully. Data Science and Predictive Analytics techniques can be used to find the features that affect ticket pricing strategy. In this project, we are provided with 'ski_resort_data.csv' dataset containing 300 observations (each observation is a ski resort) and 27 variables (each variable is a particular feature).

Project objective: The goal of this project is to use data science techniques in order to estimate what resort features have the most influence over the ticket price.

Exploratory Data Analysis: The dataset was imported and explored in Python 3, and all calculations were performed using Python 3. We plot the following figures:

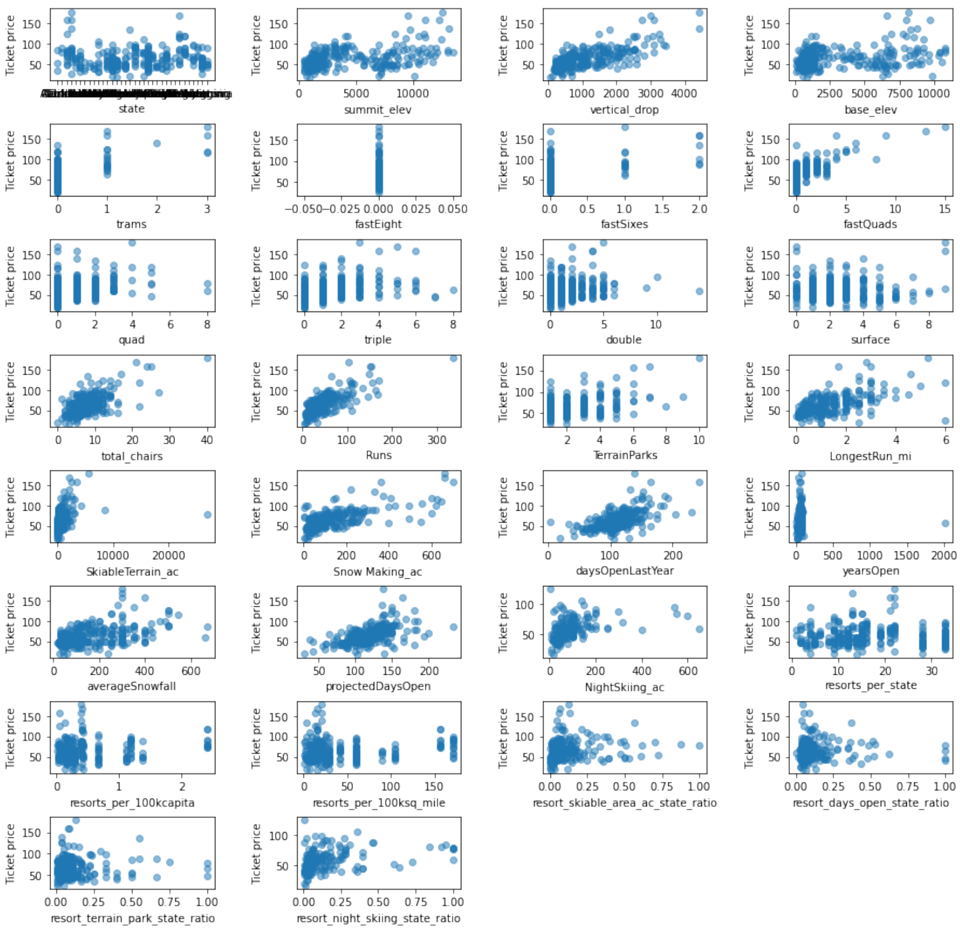

Figure 1: All features vs. ticket price scatterplots.

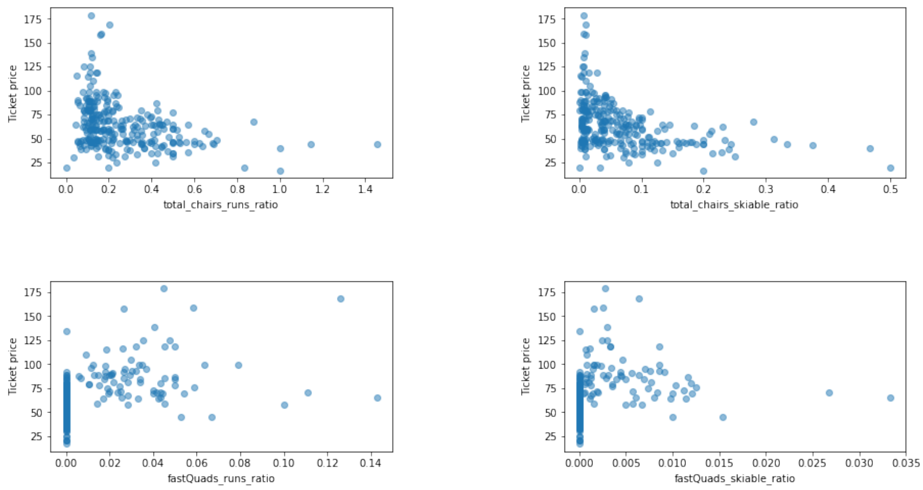

Figure 2: Transportation vs. Ticket Price scatterplots.

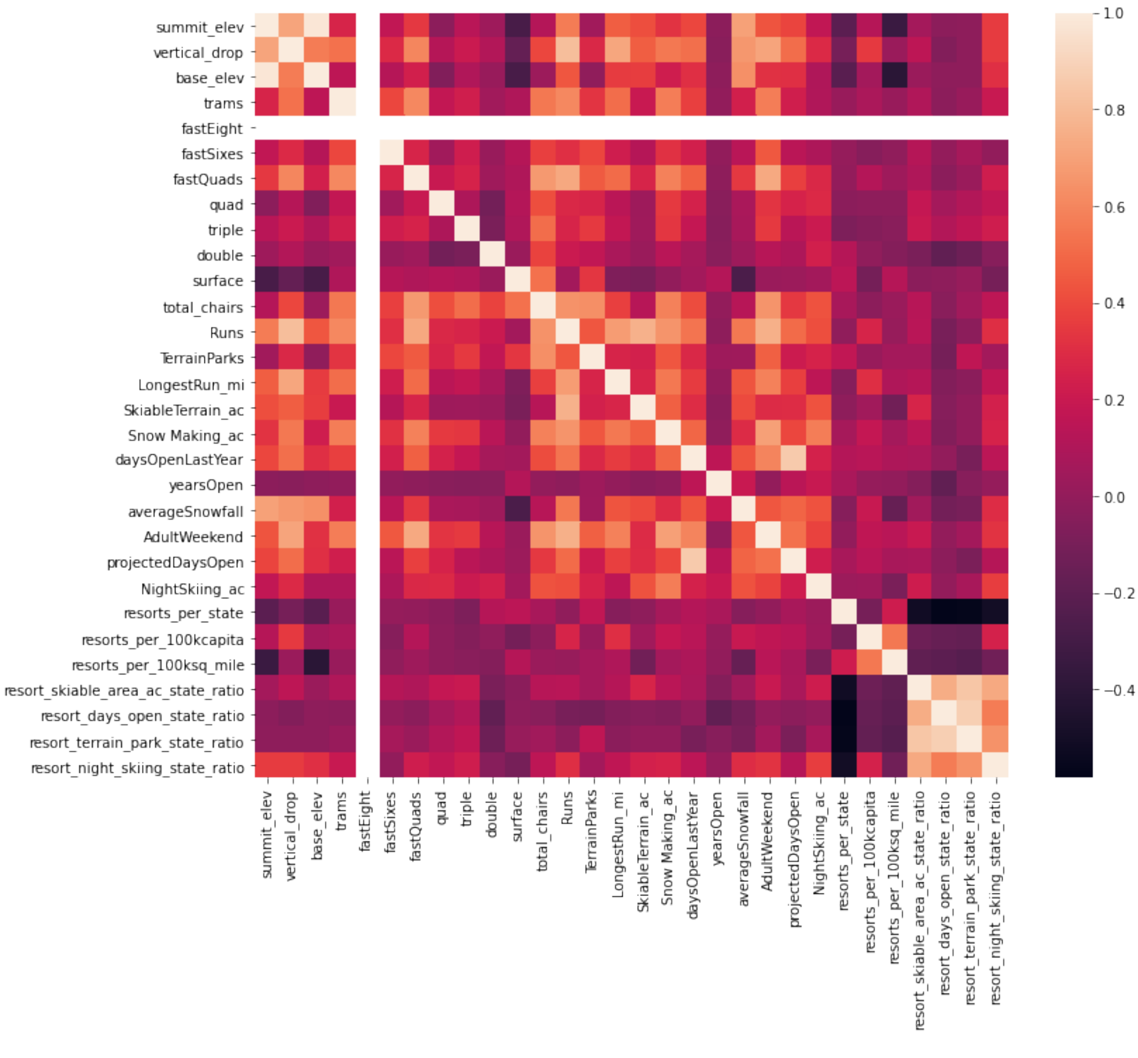

Figure 3: Features correlation heatmap.

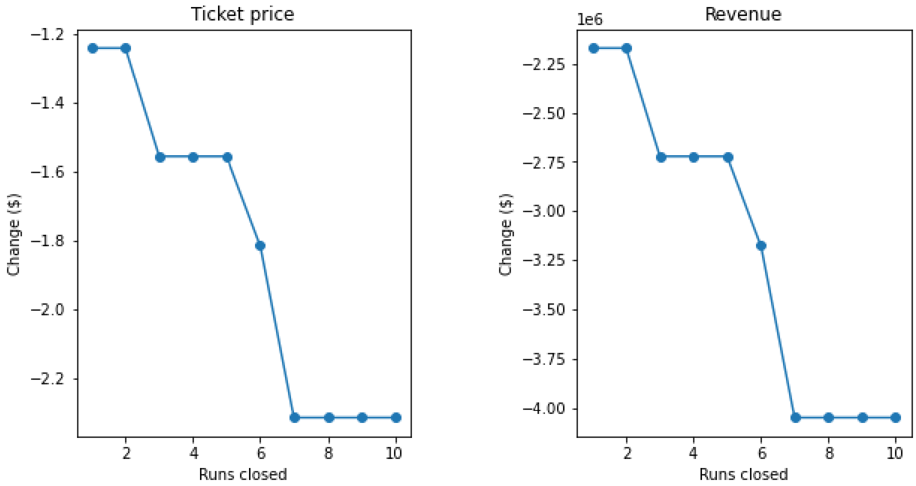

Figure 4: Runs closed scenario line plots.

Figure 1 shows all features and their relations to a ticket price.

Figure 2 shows the features related to resort's transportation and how it affects the price. It appears that the most chairs a resort has, the lower the ticket prices are. It seems counterintuitive, but might be explained by an exclusive vs. mass market resort effect - resorts with fewer chairs will have less people so they charge more for a ticket.

Figure 3 is a heatmap showing the degree of correlation between all features. It is clear that the features that ticket price is correlated with the most are *Runs*, *vertical_drop*, *fastQuads, Snow Making_ac, Skiable_Terrain_ac* and *total_chairs.*

Figure 4 shows the result of modeling the scenario where the company decides to close some runs. The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop.

Model Selection: Since our dataset contains mostly numeric data, both linear regressor and random forest regressor were used. The latter had less variance and smaller mean squared error so we chose to use it for training the model.

Predictions: Using Random Forest regressor we can predict the features which affect the ticket price. All of them have some influence on the price; however we are only interested in those that are the most important for making a decision on how to improve the ticket pricing strategy.

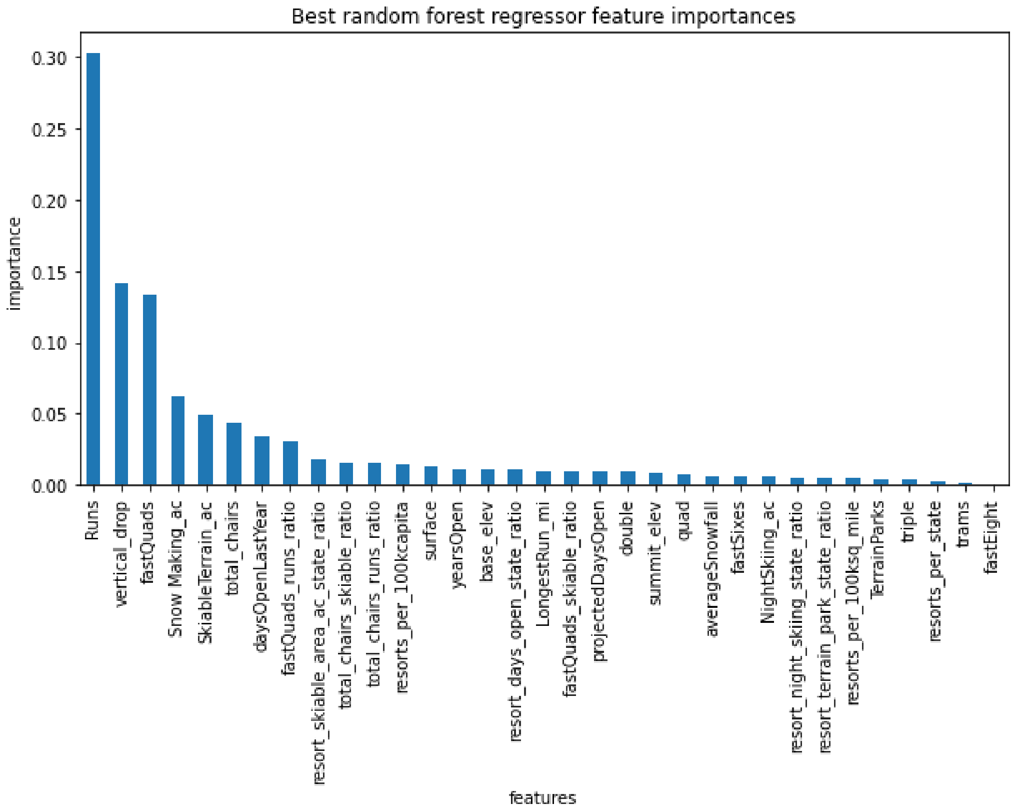

Figure 5: Feature importance for ticket price.

Based on our calculations the most important are *Runs, vertical_drop, fastQuads, Snow Making_ac, Skiable_Terrain_ac* and *total_chairs*, with first three contributing to almost 60% of the importance. The confidence interval for a model is 9$.

Conclusion: We built a simple Random Forest model for finding the importance of different features. As we can see on Figure 5, in order to improve the ticket price strategy the company's management has to pay close attention to the features with the highest level of importance.