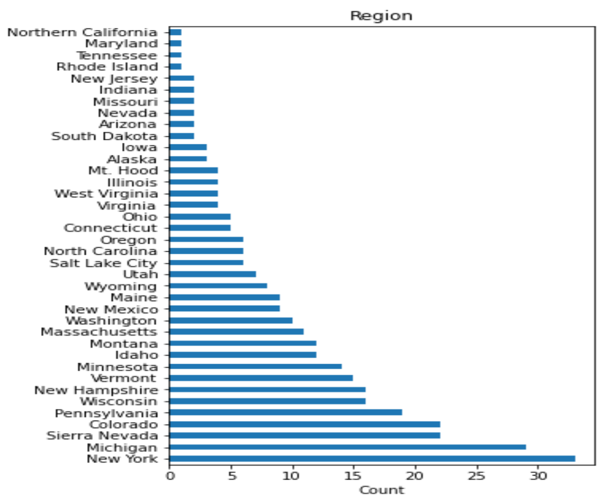
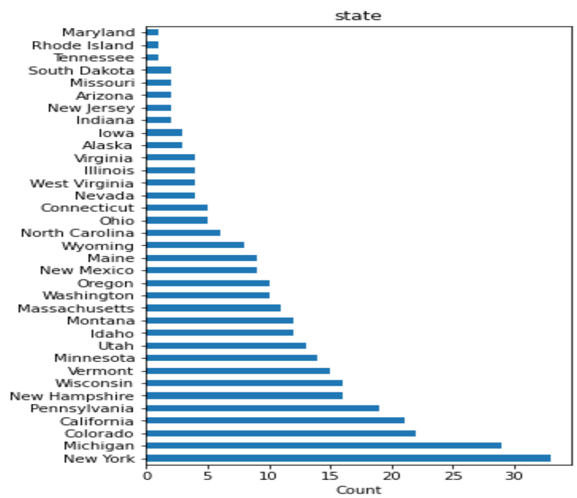
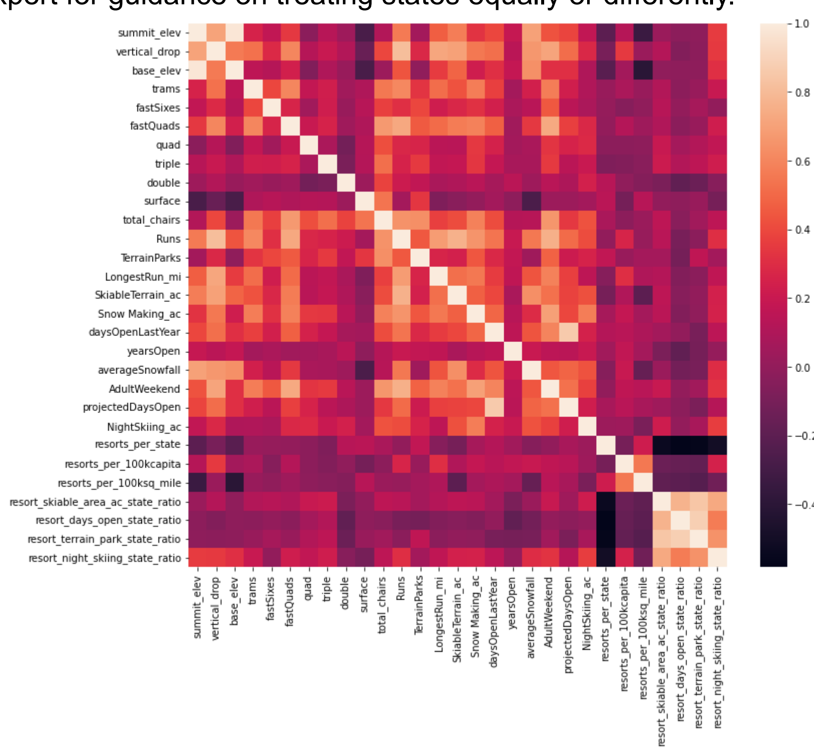
The data science team is hired to help Big Mountain Resort, a ski resort in Montana, develop a more effective pricing strategy for their tickets. The resort currently charges a premium above the average price in their market segment but suspects that they are not fully capitalizing on their facilities. They have installed an additional chair lift, which has increased their operating costs, and they want to ensure they make the most of their investments. The management has decided to implement a data-driven approach and seek guidance on selecting an optimal ticket price while considering cost-cutting measures or potential enhancements to support higher prices.
The dataset is provided in a CSV file, which includes information on various resorts. The file contains two columns, "AdultWeekday" and "AdultWeekend," representing the prices that could be considered as target variables. The remaining columns in the dataset will be treated as features. Fortunately, The Big Mountain Resort has provided complete data without any missing values. However, there are a few resorts that lack crucial information in the potential target features. As a result, these resorts had to be excluded from the analysis, after getting the most of their usable information. Since the "AdultWeekend" column has the fewest missing values, it will be utilized as the target feature for the model.





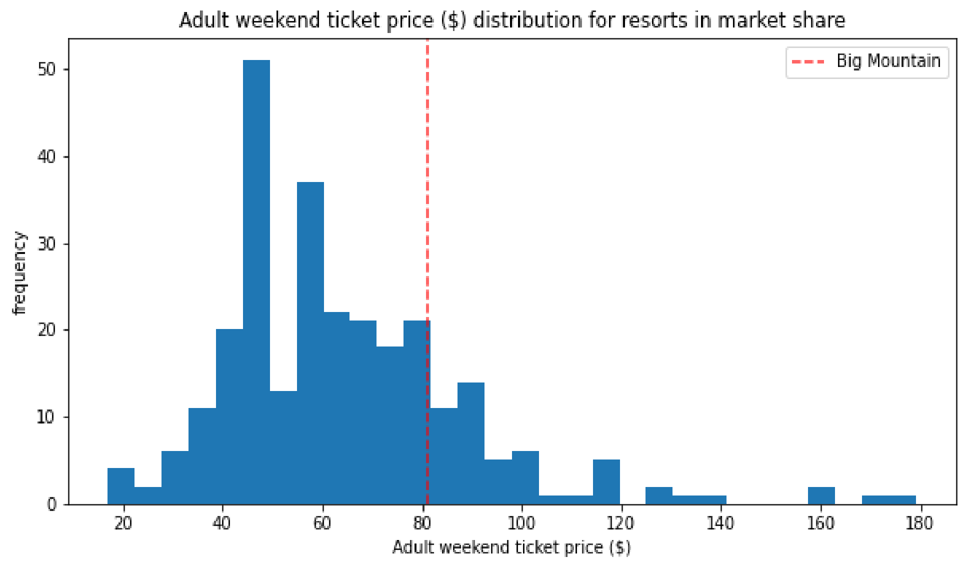
As shown in the above barplot, considering the dominance of New York in the number of resorts, while our focus is on Montana as our target location. Should we prioritize Montana and create a specific model, or include other states? The dataset implies similar prices among resorts in the same market share. It's important to consult with the client or domain expert for guidance on treating states equally or differently.

Upon further investigation of the data, a correlation heatmap was generated, uncovering interesting correlations.
The heatmap highlights a strong positive relationship between the AdultWeekend ticket price and factors such as vertical drop, fastQuads (chairlifts), the number of runs, and the presence of snowmaking facilities. This suggests that visitors are willing to pay higher prices at resorts that offer more snow making capabilities, which potentially leads to increased costs and subsequently higher ticket prices.

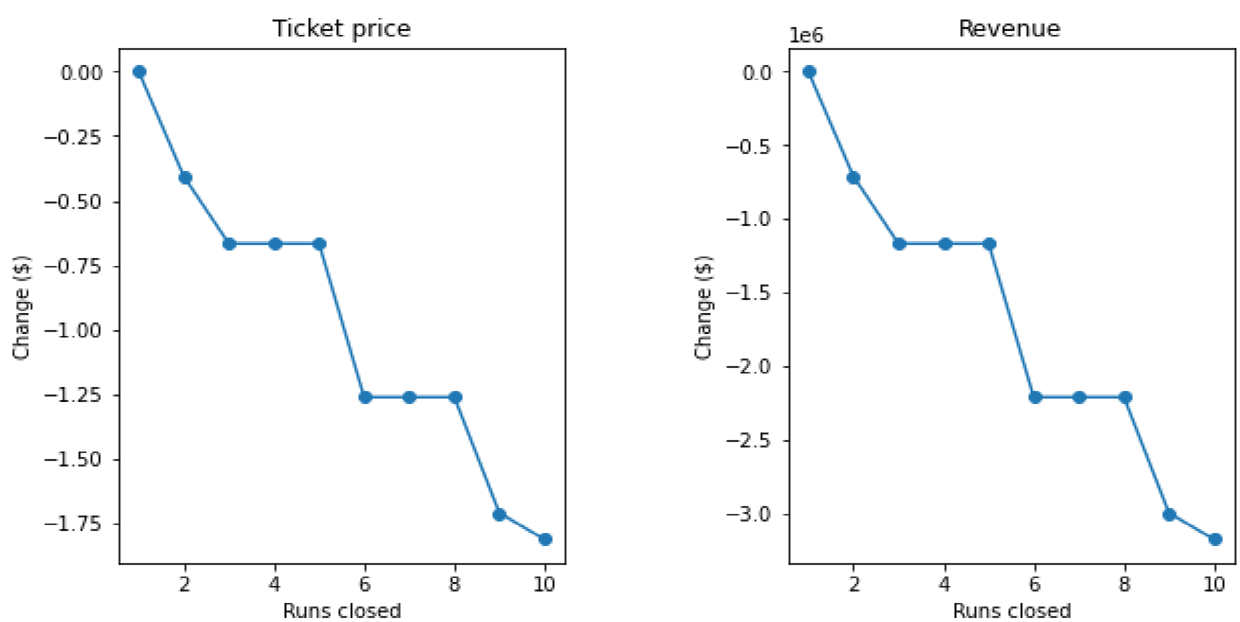


Training and testing the model with various estimators, we determined that the RandomForestRegressor model showed the best performance.

Based on the model's analysis, it is recommended that the Big Mountain Resort increases its ticket price to $95.87. Presently, the resort's price is positioned around the average range, despite offering facilities that significantly influence ticket prices, as indicated by our model.



The final model is capable of predicting outcomes in different scenarios. Based on simulations, it has been observed that the resort can potentially reduce its operating costs by closing one run without impacting the ticket price. However, closing 2 or 3 runs would have an impact on the ticket price. Interestingly, closing 4 or 5 runs would not lead to a further decrease in the ticket price. Conversely, increasing the number of snowmaking machines or the length of the longest run does not contribute to an increase in the resort's pricing value.



It is important to note that the credibility of the model's predictions is based on the underlying assumption that other resorts predominantly determine their prices based on the perceived value of specific facilities. This assumption implies that prices are established within a free market framework. In addition to the important note regarding the credibility of the model's predictions, it is worth mentioning that the estimated mean absolute error of the model is $10.39. This indicates the average difference between the predicted ticket prices and the actual prices observed in the dataset