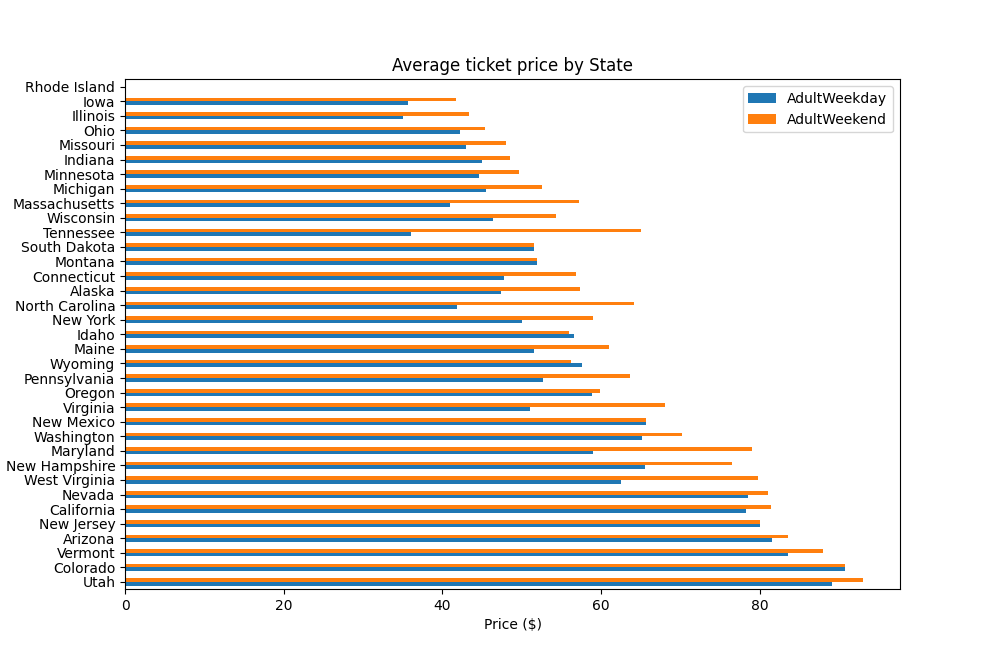
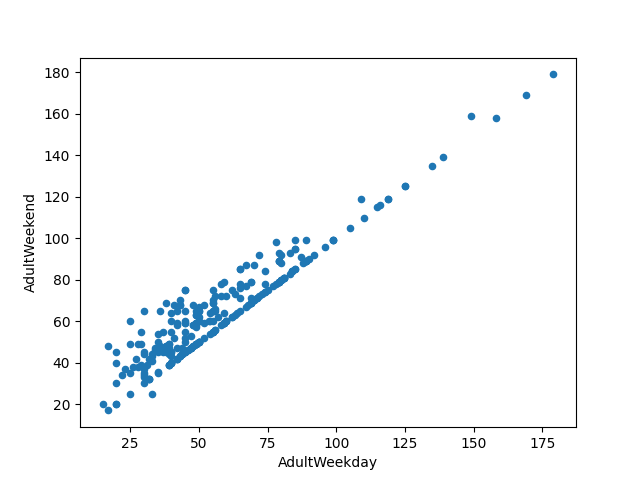
**BIG MOUNTAIN RESORT PROBLEM**

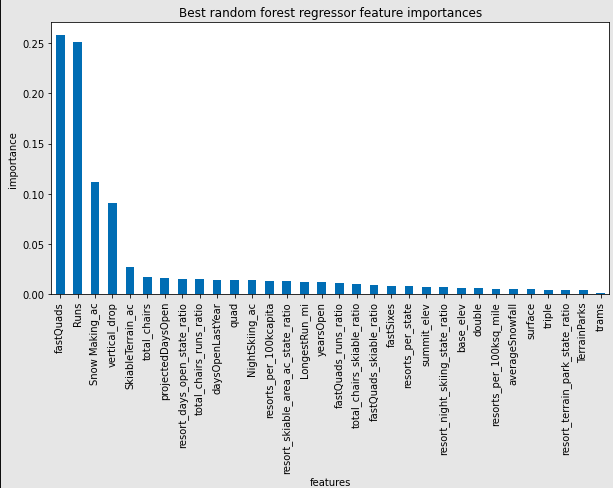
In this project, we began with the problem statement, “ What opportunities exist for Big Mountain Resort to implement a more data-driven business strategy to increase the revenue for recouping the increased operational cost of $1.54 M for installing new chair list this season, while keeping the profit margins at 9.2% and give an insight on annual revenue for the season over the next year?”

We began by wrangling the data which involved focusing on collecting the data, organizing it, and making sure it’s well defined. We began by loading the sky-resort data and exploring the data for the resort of interest – Big Mountain Resort. The Big Mountain Resort data had no missing values. We analyzed the data in a number of ways.

In one such analysis, we checked the average prices of AdultWeekday and AdultWeekend ticket (adjacent figure). This showed how the average ticket price varied from state to state.

We observed, in the model ticket price target plot, there is a clear line where weekend and weekday prices are equal (figure below). Weekend prices being higher than weekday prices seemed restricted to sub $100 resorts. Finally, we saved cleaned and summary data to our data directory for further analysis in the future.

We further performed exploratory data analysis. We used seaborn heatmap to gain a high level view of relationships amongst the features of the dataframe and found summit and base elevation were quite highly correlated. Also, when resorts were more densely located with population, more night skiing was provided. On AdultWeekend ticket price, we saw quite a few reasonable correlations. fastQuads stood out, along with Runs and Snow Making\_ac. It implied visitors would seem to value more guaranteed snow, which would cost in terms of snow making equipment, which would drive prices and costs up. We created a series of scatterplots to understand how ticket price varied with other numeric features. In the scatterplots, we saw, a strong positive correlation with vertical\_drop. fastQuads seemed very useful. Runs and total\_chairs appeared quite similar and also useful.

While pre-possessing and training the data, we constructed a pipeline that imputed missing values, scaled the data, and selected the k best features, trained a linear regression model. We also used the cross-validation technique to estimate the model performance. Using random forest model, we found the dominant top four features are in common with the linear model: fastQuads, Runs, Snow Making\_ac, vertical\_drop.

We, then, took our model for ski resort ticket price and leveraged it to gain some insights into what price Big Mountain's facilities might actually support as well as explore the sensitivity of changes to various resort parameters. This relied on the implicit assumption that all other resorts were largely setting prices based on how much people valued certain facilities. Essentially this assumed prices were set by a free market. We used our model to gain insight into what Big Mountain's ideal ticket price could/should be, and how that might change under various scenarios.

We refitted the model on all available data except that of Big Mountain. That was chosen so because we wanted to train a model to predict Big Mountain's ticket price based on data from all the other resorts! We didn't want Big Mountain's current price to bias this prediction. We explored four following modeling scenarios : a) Close up to 10 of the least used runs. The number of runs is the only parameter varying : 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. b) Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift: This scenario increases support for ticket price by $1.99. Over the season, this could be expected to amount to $3474638. c) Repeating the previous one but adding 2 acres of snow making : Such a small increase in the snow making area made no difference ! d) Increasing the longest run by 0.2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability: No difference !