**Research Problem**

The data science project aims at building a predictive model for the pricing of tickets at Big Mountain, a ski resort. 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 several facilities, or properties, boasted by resorts This model will be used to provide guidance for Big Mountain's pricing and future facility investment plans.

**Main Steps used for the analysis**

The major steps used here includes the following. First step is data preprocessing using different data cleaning methods and finally getting a clean data set. Next step is to do the exploratory data analysis to understand the patterns of data that can be used in the subsequent modeling. Third step is to preprocess and train the data. The fourth step is modeling in which insights are obtained on what Big Mountain's ideal ticket price could/should be, and how that might change under various scenarios.

**Findings and Recommendations**

The average ticket price differs considerably across different states, as shown by the analysis. There were two possible primary target features namely "Weekend" and "Weekday" Ticket Prices. Initial analysis shows that the weekend ticket prices are greater on an average for the states than the weekday ticket prices. The Weekend prices were found to have the least missing values of the two, so the finding was to drop the weekday prices and then keep just the rows that have weekend price. Thus, based on the analysis, the target feature is Weekend ticket price, which will be used for the prediction purpose. To account for the heterogeneity of the features in the ski data set, the data is scaled, which is then used for PCA transformation. The first two components seem to account for over 75% of the variance.In the first two components, there is a spread of states across the first component. Vermont and New Hampshire appear to be off on their own a little in the second dimension, although they're really no more extreme than New York and Colorado area in the first dimension. The next step is to merge the state summary features data into the ski resort data. The heat map for this merged data and scatter plot gives an indication that vertical\_drop. fastQuads,Runs, total\_chairs and resorts\_per\_100kcapita can be very useful in predicting ticket prices in the further analyses.

Comparing the performance of both linear regression model and random forest regressor, it has been found that the random forest model has a lower cross-validation mean absolute error by almost $1 and exhibits less variability. Thus, random forest regressor is used for modeling purposes. The analysis shows that at present Big Mountain charges 81while the modelled price is 94.22. This suggests that there is a scope for increase in the price given the facilities in the Big Mountain resort. The additional operating cost of the new chair lift per ticket can be covered with this rise in prices, according to our finding. The second scenario where Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift is recommended because This scenario increases support for ticket price by $1.94.