**Guided Capstone Project Report**

Problem Statement- How can we accurately adjust the price of tickets this season to maximize the profits for Big Mountain Resort accounting for facilities provided by them?

To find the accurate ticket price following step were performed:

1. Data wrangling
2. Exploratory data analysis
3. Preprocessing and training
4. Data modeling

In the first step(data wrangling), missing values were identified(eg. fastEight,NightSkiing\_ac, AdultWeekday, AdultWeekend ), duplicate data were analysed(eg. resort name: Crystal Mountain ) and columns with similar data were investigated(eg. State and Region). We also analysed distribution of tickets based on different features, and identified features of most importance. Some of these features were TerrainParks, SkiableTerrain\_ac, daysOpenLastYear, NightSkiing\_ac, based on initial analysis. Finally, we ignored the records with no price information.

Secondly, in exploratory analysis, we started with a state-wide picture for our market, by including population, area, resorts per state, total skiable area and number of days resorts were open state- wise. PCA transformation was performed to visualise the impact on the ticket price by different features state-wise, and correlation between various features were determined which will have impact on the ticket prices.

Afterwards, we started with splitting the dataset into training and tests. After that we used mean(average) price as predator, and also calculated R-squared, mean absolute error and mean squared error ( later used sklearn metrics to perform these functions). We also imputed the missing features with predictor(mean). Furthermore, we created a pipeline for modelling the data, initially we used a linear regression model. Applied it on training and test data sets to determine the performance. It explains over 80% of the variance on the train set and over 70% on the test set. Estimated increase in price of ticket is $9. After that we used Random Forest Regressor using pipeline. Cross-validation was much better than the Linear regression model(mean: 0.7081612973109424, std: 0.06564715369819316). Using Random Forest Regressor it was found that 9.495505500261919 is the mean absolute error(increase in price). It has a lower cross-validation mean absolute error by almost $1. It also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

Finally, features that came up as important in the modeling completed in the previous stage were vertical\_drop, Snow Making\_ac, total\_chairs, fastQuads. The resort operates within a market where people pay more for certain facilities to determine the approximate price of the ticket that can be set for maximum profit. Analysing the different scenarios, we found that increasing the vertical drop by 150 feet, adding 2 acres of snow making and installing an additional chair lift will support the increase in price by $1.99.

Based on the model, we can increase the cost of a ticket by $1.99. Business can use the finding of this model for increasing the profit by $3474638.