**Guided Capstone Project Report**

**Problem statement**

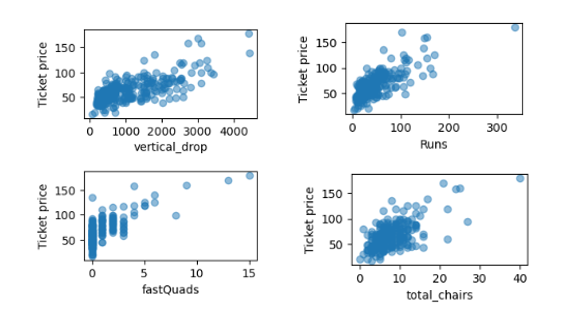
The Big mountain resort, located in Montana has the additional $1.54 million operation cost on adding extra chairs. The resort plan to implement a data-driven business strategy on optimizing ticket price to cover this $1.54 million cost by the end of this season.

**Data Wrangling**

The raw ski data contains 330 resorts including the Big Mountain with 27 features of each resort.  In this step, some fundamental issues were addressed, for example dropping column fastEight because of missing over 50% values, review each feature distribution and corrected the information input such as skiable Terrain\_ac for Silverton Mountain. The distribution of ticket price by state was explored. AdultWeekend ticket price was selected as target value based on observation of equivalence of both tickets and less missing values compared to AdultWeekday.

**Exploratory data Analysis**

Is there any pattern of ticket price distribution by state? How the correlation between the ticket price AdultWeekend with each feature? The contribution of each features to the ticket price was analyzed by principle components analysis (PCA).The analysis showed no clear pattern was found between states and ticket price. Therefore building a price model should consider all states equally. Feature engineering a resort's share of the supply for a given state was investigated by melting the two data frames. Features such as vertical\_drop, fastQuads, Runs and total\_chairs have strong positive correlation with ticket price.



**Model Preprocessing with feature engineering**

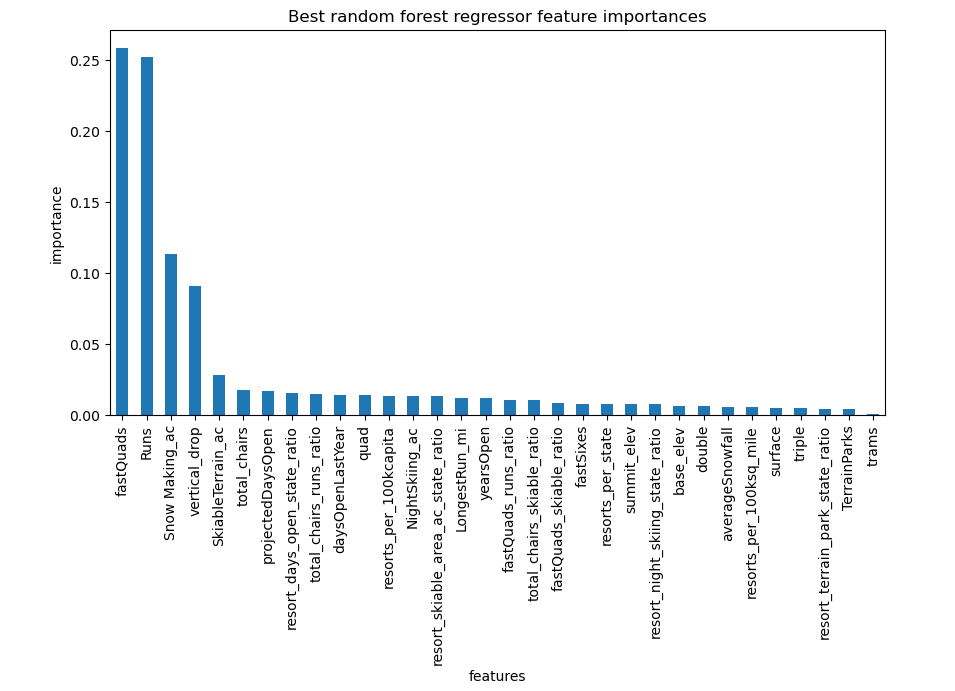
The study target Big Mountain Resort, AdultWeekend were extracted from ski\_data; The numeric features were extracted for modeling by drop state, region, and Name from data frame; the data set was splitted into 70% as train and 30% as test to evaluate the performance of the model; a baseline mode was build by using the average price as predictor. A subset of features instead of whole features should be used in the final modeling to avoid the over-fit.

**Algorithms used to build the model with evaluation metric**

Three metrics can evaluate a model performance: coefficient of determination R2, mean absolute error MAE and mean squared error MSE. R2 explains the variance of the modeling, R2 =1 suggests a model perfectly predicts the observed values. If models do worse than predicting the mean with produce a negative R2.  MAE and MSE summarize the difference between predicted and actual values. The performance of each model is evaluate via cross-validation.

**Winning model and scenario modeling**

Two modelling were explored: linear regression and random forest. In linear regression model,The biggest positive features associated with ticket price are vertical\_drop, Snow Making\_ac,total\_chairs, fastQuads and Runs. trams and SkiableTerrain\_ac are the negative features. In random forest model, the best features of importance are fastQuads,Runs, Snow Making\_ac and vertical\_drop as show below.



The random forest model was selected for further business modeling because of its lower , cross validation mean absolute error.

Among the four business options. Two options (as below) through the modeling can be used based on the assumption that each visitor buys 5 day tickets with a total 350000 visitors in a season. Both options can increase each ticket price by $1.99, and over the season can increase by $3.47M per each option. Either of the options will cover the cost of adding extra chairs.

1. Adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift.

2. Adding a run, increasing the vertical drop by 150 feet, installing an additional chair lift, and adding 2 acres of snow making.

**Pricing recommendation**

The current ticket price for Big mountain is $81, the modeling suggests the highest ticket price can increase to $95.87. If implant either business option mentioned above, only increasing the ticket price by $1.99 can cover the extra chair cost $1.54M over the end of season.

**Conclusion**

The data analysis suggests the extract cost of $1.54 can be covered at the end of the season by increasing ticket price by only $1.99 under the two business options.

**Future scope of work**

Facility cost data such as snow making, operating runs should be provided in order to accurately project the modeling scenarios. Besides adjusting ticket prices, cutting facility cost is another option, with the cost data, we can generate modeling for the ideal amount of cost reduction. Ticket information of Big Mountain such as the average tickets per custom buy and the total amount of ticket sold in the past seasons will be helpful in modeling. Because of the assumption in the current model that each custom buys 5 tickets, what if this assumption is not correct?