Guided Capstone Project Report

Problem Statement:

Big Mountain Resort in Montana is a ski resort with access to 105 trails and over 350,000 skiers visits every year.In addition to 11 lifts the resort has recently installed an additional chair lift to help distribution of visitors resulting in increasing their operating costs by $1,540,000 for the season.The resort wants some guidance on how to select a better value for their ticket price. They are also considering a number of changes that they hope will either cut costs without undermining the ticket price or will support an even higher ticket price

Data Collection :

Ski data Csv file - Data contains information from 330 resorts in the US that can be considered part of the same market share and have the same data columns for Big Mountain Resort.

Population and area data for the US states can be obtained from wikipedia. The data includes state sizes and populations that will be using for our model.

Data Cleaning and Preprocessing:

Original data contains 330 Rows & 27 Columns.Some columns need to be dropped with missing values and outliers needs to be either fixed or removed.In our dataset we dropped column 'fastEight' because it was found that 50.3% of resorts values are missing and remainings are 0.
We picked “adult weekend” ticket price as the target variable based on its relatively low missingness (15%) across all records compared to the adult weekday price (16%).
We also performed PCA analysis plots but are not able to find any pattern of a relationship between state and ticket price.From the correlation heatmap we found some features that are highly correlated with our target variable 'weekend price' are fastQuads,Runs, Snow Making_ac resort_night_skiing_state_ratio, total_chairs and vertical drop.
The final data (ski_data_cleaned.csv) has 277 Rows and 25 columns.

Machine Learning Models:

For predicting best ticket price for Big Mountain Resort we first built a pipeline which involves the following steps:

1. Imputing missing values with median / mean.
2. Scale the data to zero mean and unit variance
3. Select the k best features based on univariate statistical tests.

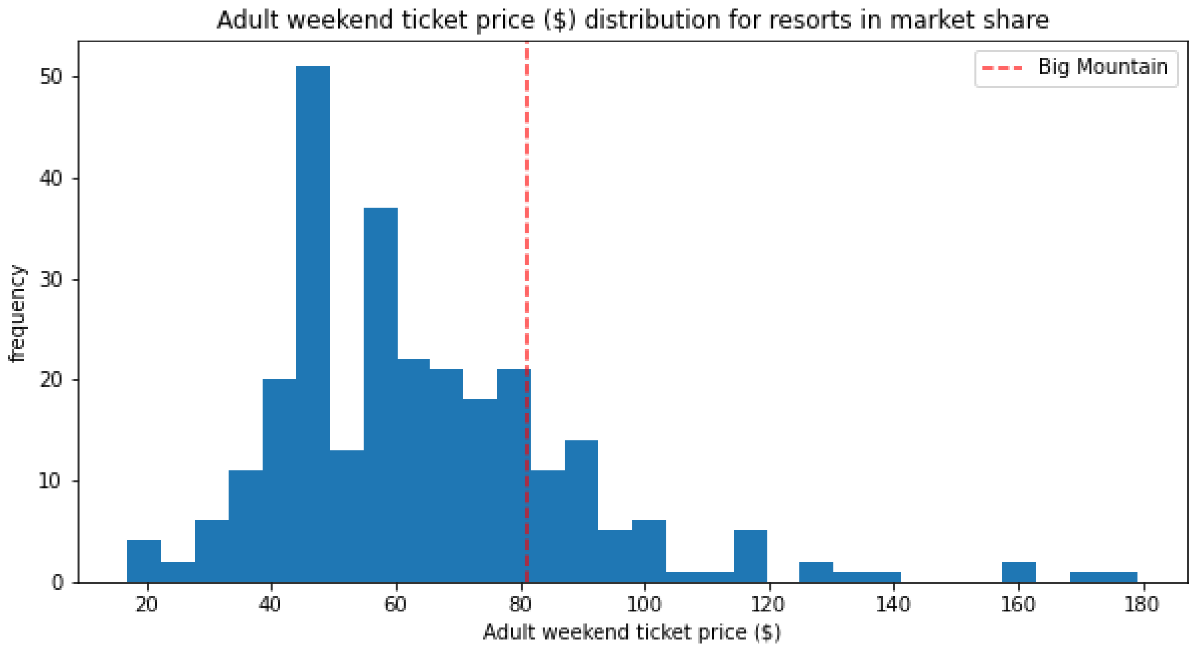
4. Applying the machine learning model (we used Linear Regression and Random Forest).

Below are the results of both the models:
Linear Regression : Cross validation MAE = 10.49, Test MAE = 11.79
Random Forest Regressor : Cross validation MAE = 9.64, Test MAE = 9.53

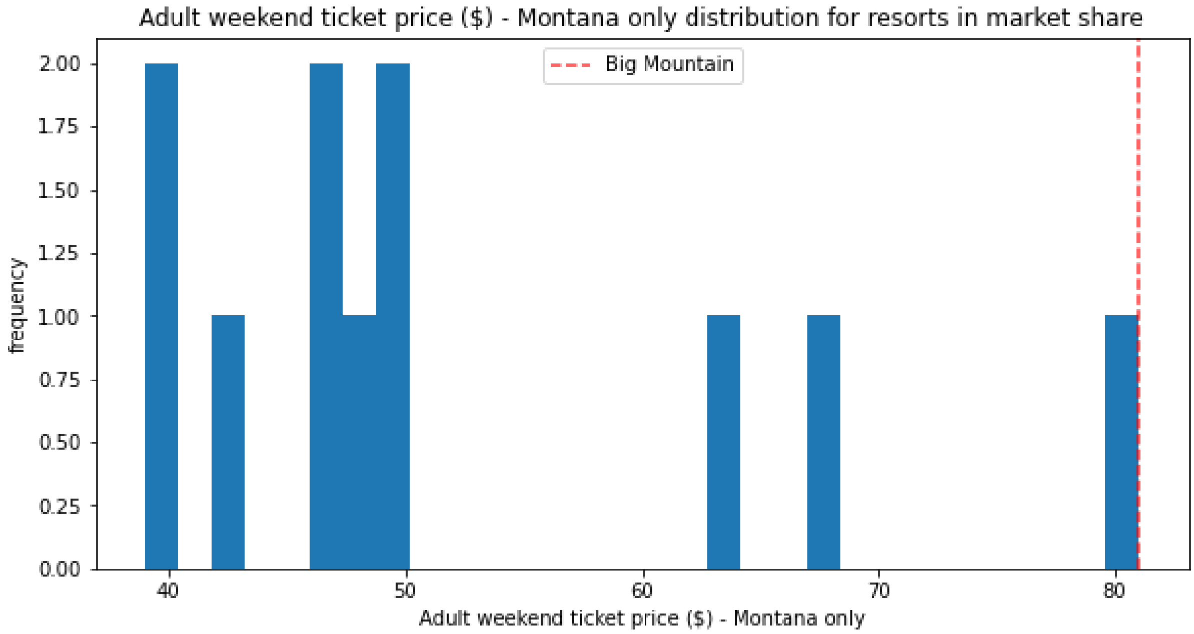
We decided to use the random forest model because it has a lower cross-validation MAE than the linear regression model's cross-validation MAE by almost $1 and exhibits less variability than the linear regression model.The random forest model's performance on the test set produced performance consistent with its cross-validation results.Also we found the top 4 features are fastQuads, Runs, Snow Making_ac, and vertical_drop.

Visualisations :

The below plot shows where Big Mountain sits overall amongst all resorts for price.



The below plot shows where Big Mountain sits overall amongst all resorts for just other resorts in Montana.



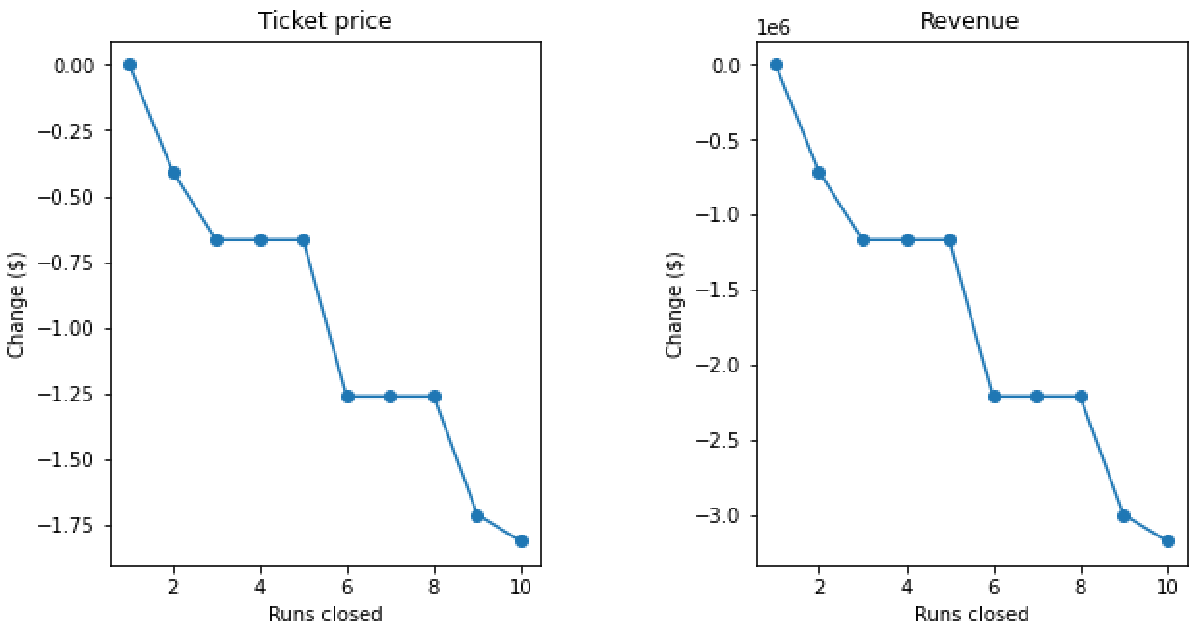
Results:

Our model suggests Big Mountain Resort modelled price is $95.87 with mean absolute error of $10.39.We looked at some histogram comparisons, it seems like Big Mountain is above average in many of the facilities it offers. This could mean that there is room to increase ticket prices.

We have taken into consideration four scenarios to increase the revenue and reduce the cost

1. Scenario 1
   Consider closing up to 10 of the least used runs.Closing 1 runs 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.

   The below plots show how closing runs affect ticket prices and revenue.



2. Scenario 2
   Proposed to add a run, increase the vertical drop by 150 feet and install an additional chair lift will increase ticket price by $1.99 and revenue by $3474638 over the season.
3. Scenario 3
   In addition to scenario 2 add snow making area of 2 acres the results are same as scenario 2.

4. Scenario 4
   Increasing the longest run by .2 miles and guaranteeing its snow coverage by adding more 4 acres of snow making capability will make no difference.

Conclusion:
Pricing strategies from scenarios 1 and 2 are recommended. Since closing 1 run does not influence the ticket price but can reduce the operation cost it is worth considering.Adding a run, increasing the vertical drop by 150 feet and installing an additional chair lift can support the ticket price significantly compared with other strategies. But more detailed information about the operation cost in scenario 2 is required before it is finally adopted.We could perform more analysis on less important features that contribute little to the price so cutting costs on these features may reduce the costs.