Guided Capstone Project Report

Summary of the findings:

From the data we can see that Montana has the 4th largest skiable area and has the 4th most ski resorts per 100k population. Vermont stands out as the state with highest density per capita and New Hampshire has the highest number of resorts per 100ksq miles.

We calculated the resort ratios compared to the state total for variables such as; resort skiable area, resort days open etc. After visualizing the data and creating the correlation heatmap, we can see that **vertical drop, number of trams, fastQuads, runs, total chairs, snowmaking ac can be important variables that positively effects price**. Resorts per 100k capita seems interesting since if the value is low there can be quite variability in ticket price and price can drop a little before climbing up as the number for resorts per capita increases. The high ticket price when the resort numbers are low can indicate a monopoly.

We asked the question: ‘ How good the mean is for predicting the ticket price? ’ From here, by using dummy regressor, the MAE, MSE and R-squared is calculated for these predictions.

MAE : 17.9, 19.13

MSE :614, 581 – Numbers are for train and tests sets respectively for all stages.

When we use median to fill missing values and fit a linear regression model. The error rates change as

R2 = 0.81 , 0.72

MAE : 8.5 , 9.4

MSE : 111.9 , 111.9 ( When we use mean to fill missing values nothing changes much )

After this step we use pipeline from sklearn to create different models. We use SelectKbest to select number of important variables. We try model with 15 variables and to check the performance using cross- validation.

We used another algorithm : Random Forest. After fitting the model and searching for the most important features, we see that fastQuads, Runs, snow making ac and vertical drop are our most important variables. This aligns with our linear model. (And also became our main model)

Next we calculate the price for Big mountain resort by using the random forest model and we see that the predicted price is $95 (actual is $81) which means that there is room for increase.

In our random forest model, most important features were;

• vertical\_drop

• Snow Making\_ac

• total\_chairs

• fastQuads

• Runs

• Longest Run\_mi

• trams

• SkiableTerrain\_ac

We see that Big Mountain resort lies on the higher side of the graph for each of these features.

A picture containing chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart

Description automatically generated

After this , we tried to answer some of the questions that business wants to understand.

1. Close up to 10 of the least used runs.

By using the function predict increase, (which adjusts the values of the features by the value of delta) we predict the changes for each number of runs that are closed. Here we find that closing one run makes no difference where closing 2 and 3 reduces support for ticket price. If the resort closes 3 runs it can close 4 and 5 which will not create any revenue impact.

Chart, line chart

Description automatically generated

1. In this scenario, 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 $8.61

Over the season $15,065,471

1. This scenario repeats previous but adds 2 acres of snow making.

Support for ticket price by $9.90.

Over the season $17,322,717

1. This scenario calls for increasing the longest run by 0.2 miles and adds 4 acres of snow making capability.

No difference for this scenario