**Guided Capstone Project Report [Emine Erdogan]**

Big Mountain Resort in Montana offers great skiing and outdoor activities with 11 lifts and a 3.3-mile-long run. The resort hosts 350,000 visitors annually for an average stay of 5 days at $81 per day. The resort is near natural wonders like Glacier National Park and the towns of Whitefish, Columbia Falls, and Kalispell.

The purpose of this data science project is to come up with a pricing model for ski resort tickets. Big Mountain Resort aims to maximize its returns, relative to its position in the market and this project will build a predictive model for ticket prices based on several facilities. They recently installed a chairlift, increasing operating costs by $1.54 million. The model is going to be used to provide guidance for Big Mountain's pricing and future facility investment plans.

**Data wrangling section**: US population data from Wikipedia was added to the dataset. Different resort prices from national market have been compared and results revealed that ticket prices lie in a broad band from $40 to $179 in the national market, averaging around $64. Some states show more variability than others. Montana and South Dakota, for example, both show fairly small variability as well as matching weekend and weekday ticket prices. Nevada and Utah, on the other hand, show the most range in prices. Some States, notably North Carolina and Virginia, have weekend prices far higher than weekday prices.

In addition to that, it is revealed that the weekend prices have the less missing values than the weekday prices, so the weekday price column was dropped from the dataset and the weekend price column as the target variable was kept. The weekend prices column contained some missing values that led to several rows being dropped from the dataset completely. The original number of columns/rows in the data were 27/330 and the final number of columns/rows are now 25/277.

**Preprocessing:** we identified that there are 8 features that will have significant impact on resorts’ returns. Features that came up as important in the modeling (not just our final, random forest model) included:

* vertical\_drop (vertical change in elevation from summit to the base in feet)
* Snow Making\_ac (total area covered by snow making machines in acres)
* total\_chairs (sum of all chairlifts at the resort)
* fastQuads (the number of fast four person chairlifts)
* Runs (count of the number of runs on the resort)
* LongestRun\_mi (length of the longest run in the resort in miles)
* Trams (the number of trams)
* SkiableTerrain\_ac (total skiable area in square acres)

**Modeling**: Linear regression and random forest models has been used to predict the adult ticket price for Big Mountain ski resort. To ensure that the features are on a consistent scale, the StandardScaler is used to scale each feature to zero mean and unit variance. The linear regression model has an R2 value of 0.81, indicating good model accuracy. However, the much lower value for the test set suggests overfitting to some extent.

Both models show good performance in predicting ticket prices for resorts in the USA, but the random forest model has slightly better performance than linear model, because it has a lower cross-validation mean absolute error by almost $1, and it also exhibits less variability. The model's performance is much better than guessing using the average, with an estimate of ticket prices within $9 of the real price on average.

According to the results produced by Random Forest model produced, dominant top four features are fastQuads, Runs, Snow Making\_ac, vertical\_drop. More specifically, the results suggest that vertical drop is the biggest positive feature. This makes intuitive sense and is consistent with what we found during the EDA work. Also, the area covered by snow making equipment is a strong positive as well. People like guaranteed skiing! The skiable terrain area is negatively associated with ticket price! It could be an effect whereby larger resorts can host more visitors at any one time and so can charge less per ticket. As has been mentioned previously, the data are missing information about visitor numbers. Bear in mind, the coefficient for skiable terrain is negative for this model. For example, if you kept the total number of chairs and the number of fast four chairs (fastQuads) constant, but increased the skiable terrain extent, you might imagine the resort is worse off because the chairlift capacity is stretched thinner.

The report suggests that further model selection is necessary to explore possible hyperparameters and remove the least useful features. Building a simpler model with fewer features can have the advantage of being easier to sell and explain to stakeholders. Although the linear regression and random forest models were both viable, further analysis was required to choose the final model. Cross-validation showed that both models had similar performance, with the linear regression model having a mean absolute error of $9.28 and the random forest model having a mean absolute error of $9.32.

**Possible scenarios to cut the costs and increase the price**

We then take the model for ski resort ticket price and leverage it to gain some insights into what price Big Mountain's facilities might actually support as well as explore the sensitivity of changes to various resort parameters. Note that this relies on the implicit assumption that all other resorts are largely setting prices based on how much people value certain facilities. Essentially this assumes prices are set by a free market. The expected number of visitors over the season is 350,000 and, on average, visitors ski for five days. Assume the provided data includes the additional lift that Big Mountain recently installed.

Big Mountain Resort has been reviewing potential scenarios for either cutting costs or increasing revenue (from ticket prices). The results are shortlisted in the following:

1- 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.

2- In this scenario, if Big Mountain add a run, increasing the vertical drop by 150 feet, and installing an additional chair lift it increases support for ticket price by $1.99, over the season, this could be expected to amount to $3,474,638

### 3- if Big Mountain add 2 acres of snow making it increases support for ticket price by $1.99

Over the season, this could be expected to amount to $3,474,638. Such a small increase in the snow making area makes no difference!

### 4- Increasing the longest run by .2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability. No difference whatsoever.

Chart, histogram

Description automatically generated

Chart

Description automatically generated

A picture containing text, antenna

Description automatically generated

Chart, histogram

Description automatically generated