Guided Capstone Project Report

Client: Big Mountain Resort, Montana.

Subject: Big Mountain is not capitalizing on its facilities as much as it could. The business wants 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.

Date: 12/2/2020

Data Wrangling

Ski Resort data has 330 entries(resorts) and 27 Data columns. AdultWeekday and AdultWeekend are the 2 columns which are the price of the tickets and the other columns are features of the Resort. fastEight is one of the columns in which half of the values are missing and others have zero value so decided to drop the column. The data is verified if each row is unique and made sure there are no duplicate values. Dropped the records(rows) which have missing ticket prices. The new ski_data after data wrangling has 25 columns and 277 entries(rows) after organizing the data. Merged Ski Resort data with state population data to compare number resorts per state with population.

Exploratory data analysis

Our home state, Montana, is the third largest state and is less densely populated.
New York comes top in the number of resorts in our market.
New York state may have the most resorts, but not in the top 5 most skiable area . Montana makes it into the top five. New York dominates the area of skiing available at night.
The total days open seem to bear some resemblance to the number of resorts. Montana comes in 4th in Resorts per capita. Data is scaled and mean is calculated. The data has all numerical features except only one categorical feature state which is converted into Index before starting off with data exploration. There was no pattern found for the relationship between Name, Region, state and Ticket price.
A strong positive correlation was found with vertical_drop. fastQuads seems very useful. Runs and total_chairs appear quite similar and also useful. resorts_per_100kcapita shows something interesting that you don't see from just a headline correlation figure.
Total_chairs, Runs, SkiableTerrain_ac and fastQuads are the features chosen for modelling since they effect the price of the ticket. total_chairs_runs_ratio, total_chairs_skiable_ratio, fastQuads_runs_ratio and fastQuads_skiable_ratio are calculated to plot scatter plots.

Processing and training

Mean absolute error tells you that, on average, you might expect to be off by around $19 if you guessed ticket price based on an average of known values. The missing values to be imputed by median. Using this model, then, on average you'd expect to estimate a ticket price within $9 or so of the real price.
A model that can work very well in a lot of cases is the random forest model which i will be using going forward. Reason: Time to stop the bad practice of repeatedly checking performance on the test split. Instead, go straight from defining the pipeline to assessing performance using cross-validation. cross_validate will perform the fitting as part of the process. This uses the default settings for the random forest so you'll then proceed to investigate some different hyperparameters.
The random forest model 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.
Plenty of data is already available. There's an initial rapid improvement in model scores as one would expect, but it's essentially levelled off by around a sample size of 40-50. There is no need to collect more data.

Modeling

Features that came up as important in the modeling included vertical_drop, Snow Making_ac, total_chairs, fastQuads, Runs, LongestRun_mi, trams and SkiableTerrain_ac. Big Mountain is doing well for all the features, but there are no trams. The vast majority of resorts have no trams similar to Big Mountain. so increase in ticket price shouldn't be a problem.
Big Mountain resort currently charges $81.00. Big Mountain Resort modelled price is $104.08 and expected mean absolute error of $10.66. This suggests that there is room for an increase.
Closing one run 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.
Big Mountain by adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift increases support for ticket price by $3.50.Over the season, this could be expected to amount to $612500.
By adding 2 acres of snow making, increasing the longest run by 0.2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability makes no difference for the ticket price.

Big Mountain is fairly high on most of the league charts of facilities offered except the trams, so the modeled price is much higher than the current price. This mismatch will definitely surprise the business executives. They can think about increasing the ticket prices.