BIG MOUNTAIN PRICING STRATEGY

Building a predictive model
for ticket prices

TABLE OF CONTENTS

PROBLEM IDENTIFICATION

Big Mountain recently installed an additional chair lift which increases their operating costs by $1,540,000 this season. The company is looking for guidance to assess how important some facilities are compared to others, to select a better value for their ticket price. Big Mountain suspects it may not be maximizing its returns, relative to its position in the market. It also does not have a strong sense of what facilities matter most to visitors, particularly which ones they're most likely to pay more for.

The purpose of this initiative is to build a predictive model for ticket prices based on a few facilities, or properties at the resorts. This model will be used to provide guidance for Big Mountain's pricing and future facility investment plans.

RECOMMENDATION AND KEY FINDINGS

DATA WRANGLING

We have a data set with 330 rows and 27 columns, but the focus has been placed to: **TerrainPar**, **SkiableTerrain**, **daysOpenLast** and **NightSkiing_ac**.

According to the data available, the Weekend price (**AdultWeekend** column in our data) is the most suitable feature to predict ticket price.

EXPLORATORY DATA ANALYSIS

We couldn’t find an obvious pattern for each state. So, we decided to treat all states equally.

Useful features: **numbers of various chairs**, and the **number of runs**, but we don't have the ratio of chairs to runs which could inform us how easily, and so quickly, people could get to their next ski slope!

We have seen an exclusive versus mass market resort effect.

Something very useful that's missing from the data is the number of visitors per year.

MODELING RESULTS & ANALYSIS (1/4)

INITIAL NOT-EVEN-A-MODEL

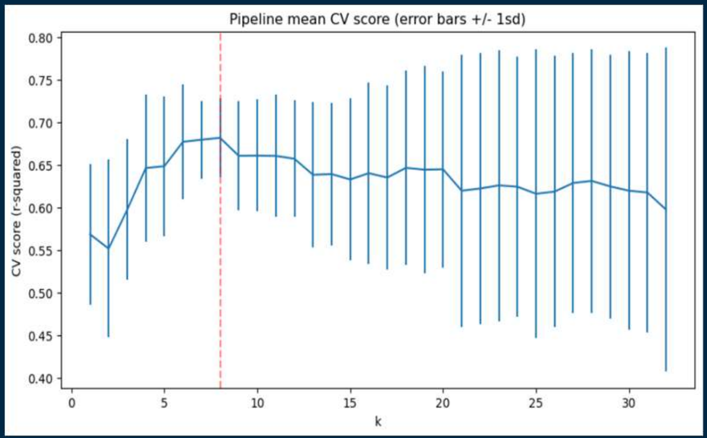
A good place to start is to see how good the mean is as a predictor. In other words, what if we simply say our best guess is the average price?

The **Average Price** is: **$63.81** and the **Mean Absolute Error** (MAE) is: **19.14**

The Mean Absolute Error is arguably the most intuitive of all the metrics, this essentially tells us that, on average, we might expect to be off by around **$19** if we guessed ticket price based on an average of known values.

MODELING RESULTS & ANALYSIS (2/4)

Linear Regression Model



We used cross-validation for multiple values of k (number of features) to pick the value of k that gives the best performance.

The above suggests a good value for k is 8. There was an initial rapid increase with k, followed by a slow decline. Also noticeable is the variance of the results greatly increase above k=8.

After imputed missing values and scaling the data we trained a Linear Regression (LR) model,

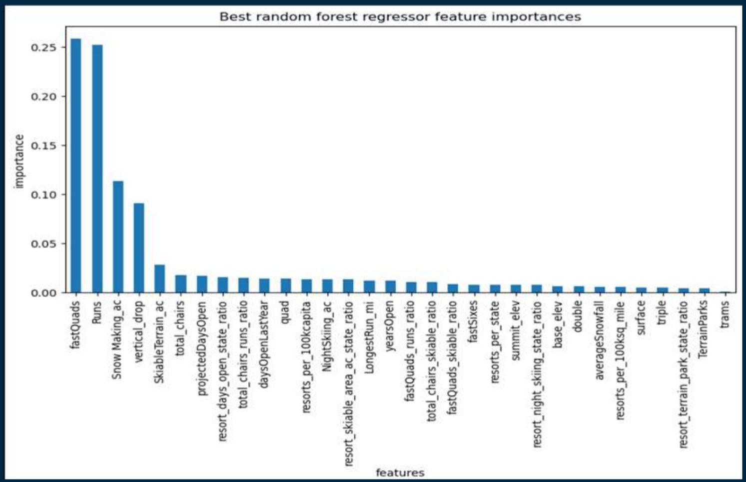
The **R_Squared Score** was: **0.79** on the training set and **0.64** on the test set, meaning the features included in the model explains **79%** of the variability of the ticket price which is not a bad score for our predictive model. The more this value is close to **100%** the better our model.

The **Mean Absolute Error** (MAE) was: **10.49** and the **Standard Deviation** was **1.35**.

So, we might expect to be off by around **$10** if we guessed ticket price based on our **Linear Regression** Model.

MODELING RESULTS & ANALYSIS (3/4)

Random Forest Model



Encouragingly, the dominant top four features are in common with our linear model:

fastQuads
Runs
Snow Making_ac
vertical_drop

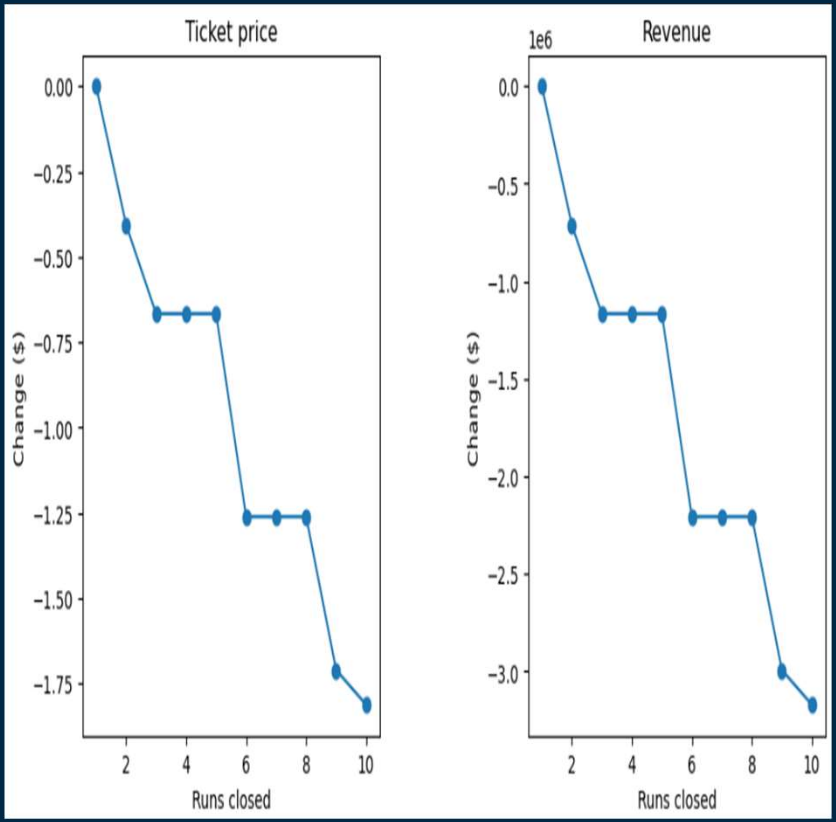
We also trained a Random Forest (RF) model,

The **R_Squared Score** was: **0.70** and the standard deviation was **0.07**, meaning the features included in the model explains **70%** of the variability of the ticket price which is not a bad score for our predictive model. The more this value is close to **100%** the better our model.

The **Mean Absolute Error** (MAE) was: **9.64** and the **Standard Deviation** was **1.35**.

So, we might expect to be off by around **$9.64** if we guessed ticket price based on our **Random Forest** Model.

MODELING RESULTS & ANALYSIS (4/4)



The model says closing one run makes no difference. The business might test, and progress, with run closures by closing one run, then 2 and 3 successively, then closing 3 runs, then closing 4 or 5 runs, then closing 6 or more and assess the impact on ticket price at each stage of run closures.

Currently Big Mountain is charging 81 USD for AdultWeekend. The model suggests a ticket price that could be supported in the marketplace by Big Mountain's facilities to be 95.87 USD. Even with the expected mean absolute error of 9.64 USD, this suggests there is room for an increase.

Big Mountain has amongst the highest number of total chairs, resorts with more appear to be outliers.

The scenario where Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift is the one I would recommend for further consideration because this scenario increases support for ticket price by 8.61 USD over the season. This could be expected to amount to 15,065,471 USD.

SUMMARY & CONCLUSION

Random Forest vs. Linear Regression

The Random Forest model is our choice for predicting the ticket price. It has a lower cross-validation mean absolute error by almost $2. It also exhibits less variability.

Data Availability

In this use case we already have plenty of data

Operating costs and missing data

Besides the additional operating cost of the new chair lift, operating cost on vertical drop, snow making area, fast quads, runs, trams and skiable terrain area may also be useful. The data is missing information about visitor numbers.

Pricing Strategies

It's reasonable to expect that some resorts will be "overpriced" and some "underpriced." Or if resorts are pretty good at pricing strategies, it could be that our model is simply lacking some key data. Certainly, knowing more about operating costs, of key important features in the model would help.