Guided Capstone Project Report

The problem statement:

How can Big Mountain Resort select a better ticket price or cut costs to increase their revenue by $1,540,000 for their additional chair lift cost this season?

Context:

The Big Mountain Resort has recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This additional chair increases their operating costs by $1,540.000 this season. The business wants some guidance on how to select a better value for their ticket price. They hope will either cut costs without changing the ticket price or will support an even higher ticket price.

Criteria for success:

Revenue increases by not less than $1,540,000 this season through higher ticket price or cutting costs.

Scope of solution space:

The resort's pricing strategy has been to charge a premium above the average price of resorts in its market segment. There are limitations to this approach. To compare the facilities and ticket prices with 330 ski resorts in the US to evaluate the operating costs and select the good price which can provide the business with a good sense of how important some facilities are compared to others.

Constraints with solution space:

1. Gaining access to the proper data sources can be difficult.
2. There are too many influencing factors of the ticket price such as location, the heights of the resort base and summit mountain, kinds and number of lifts and trams, kinds, number, lengths, and areas of runs and parks, open schedules for last season, number of years the resort has been opened, average annual snowfall, projected days, night skiing.

Stakeholders to provide key insight:

Jimmy Blackbum – the director of Operations
Alesha Eisen – the Database Manager

Key data source:

A csv file including 330 resorts in the US with the facilities information provided by the database manager.

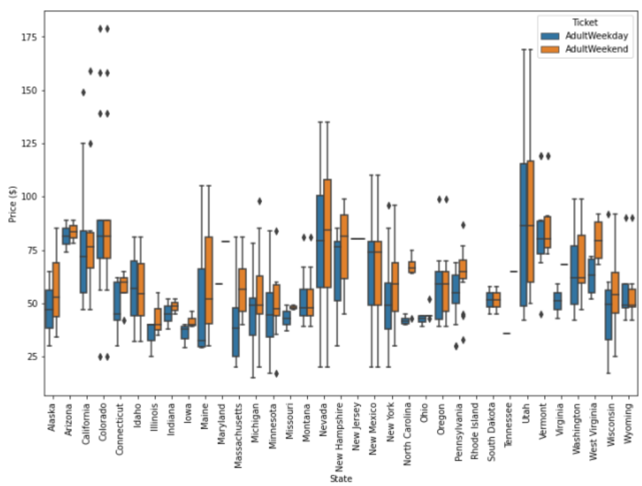
Data Wrangling

Purpose:
To check out and drop the invalid element from data.
To select useful information from the original data for the following analysis.

To find out and fix any other issues from original data.
To combine and sort data in groups according to the requirements of the following analysis.

The processes and findings from this step:

- There are 330 original rows and 27 columns in the data of ski_resort_data.csv.
- To check out the information of Big Mountain Resort Montana from data, and it is shown there is no missing values for our target resort.
- To calculate number of missing values by column: for the key information about prices of weekday and weekend, there are over 82% resorts have no missing ticket price, 3% are missing one value, and 14% are missing both, we drop the records with no price information.
- To check if the names of resorts listed are unique, and to fix the issue by adding region or state.
- To compare the distributions of resorts by region and state.
- To check out the distribution of ticket price by state, and the distribution of weekday and weekend price by state.



- To check out and drop the invalid values by finding out the abnormal distributions of feature values, for instance, the abnormal values of Skiable Terrian area, Snow Marking area, open years, and etc.
- To derive state-wide summary statistics for our market segment, also to check the missing resorts and to add them.
-
- To check more details for our target state of Montana, for instance, the relationship between weekday and weekend prices, and number of missing values by resorts.
-
- So, we finally get cleaned ski data and summary data grouped by state.

Exploratory Data Analysis

1. The main purpose of the step is to check and determine the exit and derived features may affect the ticket price most.
2. Checked out the features of states to see if there are some relationships with ticket price:

a. From data wrangling, we decided to use adult weekend ticket price as our aim ticket price to use. (Details see the analysis for data wrangling of ticket prices)

b. The aim of this step is to find out the features may affect the ticket prices, so we can use them in subsequent modeling.

c. Looked at the top states by order of the features may affect the ticket price such as total state area, total state population, number of resorts per state, total skiable area, total night skiing area, total days open, etc.

d. Roughly looking these features, we found the feature total state area, total population, and number of resorts per state are not indecently affect the ticket price, and the resort density ratio suggested more clear relationship with ticket price, so we used number of resorts per 100k population and number of resorts per 100k square miles instead of absolute state population and state area.

3. Visualized high dimensional data:

    a. scale the data: a data frame combined a scaled ndarray with index coordinated to state labels and column names.

    b. fit the PCA transformation: we found the first two components account for 77.2% variance.

    c. applies the transformation to the data to create the derived features: by adding the state average price to first two components scatter plot, to check how the first two components affect the price.

    d. uses the derived features to look for patterns in the data and explore the coefficients: We checked the high values for both first two components, to see which features are significantly affect the prices.

4. Merged the state summary features into the ski resort data, by adding "state resort competition" features:

    a. ratio of resort skiable area to total state skiable area

    b. ratio of resort days open to total state days open

    c. ratio of resort terrain park counts to total state terrain park count.

    d. ratio of resort night skiing area to total state night skiing area

5. Used feature correlation heatmap to check out the relationship between features and the Adults weekend price which our interest of ticket price.

    a. The state resort competition features are negatively correlated with number of resorts per state.

    b. We found the adults weekend price is more affected by fast quads, runs, snow making area, total chairs, and the ratio of resort night skiing area to total state night skiing are

6. Used the scattering plots of each feature against adults weekend price to check the coordinate correlation with ticket price further:

    a. There's a strong positive correlation with vertical drop. fastQuads seems very useful. Runs and total chairs appear quite similar and useful.

    b. There's some interesting points with the plot of resorts per 100k partials: The lower ticket price when fewer resorts serve a population may similarly be because it's a less popular state for skiing. The high price for some resorts when resorts are rare (relative to the population size) may indicate areas where a small number of resorts can benefit from a monopoly effect.

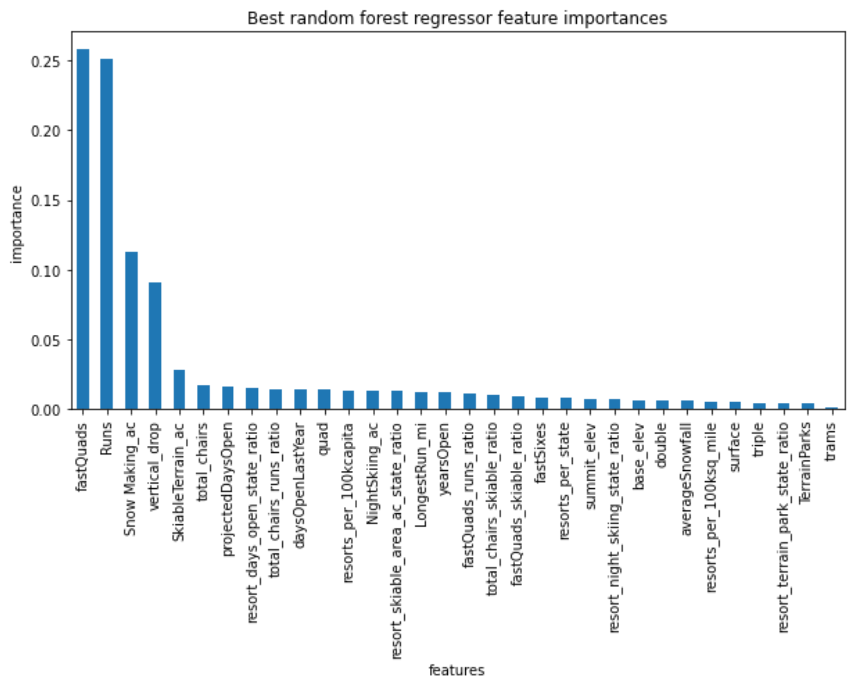
    c. We checked out the derived features may affect the ticket price: 1. the ratio of total chairs to runs and the ratio of total chairs to skiable area: we found less chairs may get higher ticket price but lower visitors (we lacked the number of visitors per year); 2. the ratio of fast quads to total runs and the ratio of fast quads to total skiable area: we found no fast quads may limit the ticket price, but wide skiable area with small number of fast quads may be beneficial.

Pre-processing and training data

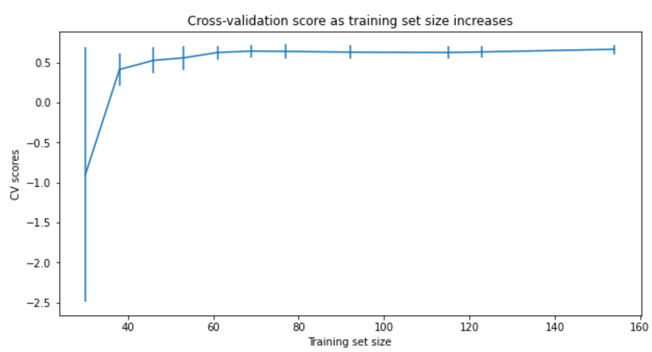
The performance of the model with mean absolute error is 9.53 as which for Linear regression model is 11.79.

we can tell the Random Forest Regression Model performance is better than the linear regression model performance, it has a lower cross-validation mean absolute error by almost 1 dollar and it also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

We decided to use the random forest regression model. Got the Best random forest regressor feature importance shown bellow:



We checked how performance varies with differing data set sizes, and the result tell us a sample size 40-50 is good enough. We do not have to increase our training set.



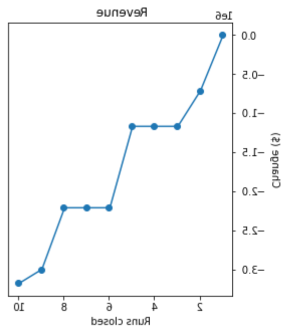
Modeling

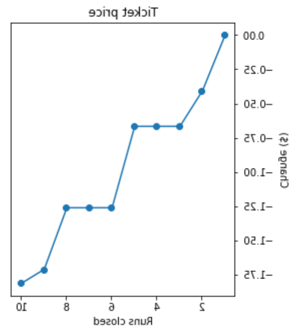
The Current charge of Big Mountain is $81.00 as our model suggest charge is $95.87, even with the expected mean absolute error of $10.39, this suggests there is room for an increase.

We checked all the important features in our modeling for Big Mountain compared which in other resorts in the whole market and the Montana market, found that all features in Big Mountain are at the top level in the market, that makes the prediction price high.

Big Mountain has been reviewing potential scenarios for either cutting costs or increasing revenue. We checked 4 scenarios to see how it works:

1. To close 10 of the least runs: we can close 1 run with no effect to revenue, and 2-5 runs reducing the same revenue as $1.2 million, and 6-8 runs reducing the same revenue as $2.3 million which is too much.





2. To add 1 run, increase the vertical drop by 150 feet, and install 1 chair lift: will increase ticket price by $1.99 and will increase revenue $3474638.
3. To add 2 acres of snow making: no difference.
4. To increase the longes runs by 0.2 miles and add 4 acres of snow making: no difference.

From the 4 scenarios we tested, we can see the two ways cutting cost and increase ticket price can work both ways, we need to check which scenarios can cutting the cost more efficiently and to increase the ticket price. First, we suggest closing 5 of the least runs to cut the cost about $1.2 million, at the same time add vertical drop by 150 feet following the new installed additional chair lift to increase price by $1.99 so can cover the $1,540,000 operating increasing in totally.

So we suggest ticket price is $81.00 + $1.99 = $82.99, and closing 5 of least runs, adding 150 feet to the vertical drop at the same time.

We suggest Big Mountain can consider decreasing the vertical drop, snow maker cover area, number of fast quads, longest run length to cut the cost in the future or can increase the ticket price to match the top features in the market.

The current price strategy has been to charge a premium above the average price of resorts in its market segment, but the most features of the resort are much better than the average level in its market. Considering it's the highest ticket price in Montana market, the higher ticket price may cannot be accept by the local market. So, we may pay more attention to consider how to reduce cost in the further work.