Big Mountain Resort Data Analysis Documentation

James Lo
July 7, 2023

Problem Statement
How can Big Mountain Resort overcome this season's ~$1.5 million increase in operating costs by the end of this season by emphasizing value and adjusting ticket prices accordingly, while at the minimum maintaining profits?

Context
In order to increase the distribution of visitors, Big Mountain Resort (BMR) has recently installed an additional lift that increases this season's operating costs by ~$1.5 million. To maintain or possibly increase profitability, BMR is looking to implement data-driven, value-based ticket pricing in addition to implementing data-driven cost cutting measures that will increase profitability.

Criteria for Success: Establishing a value-based ticket price based on market competition and BMR's facilities.

Scope of Solution Space
- Determine the effect of the many variables including (but not limited to) number of chairs, elevation, vertical drop, region, longest runs, etc. on pricing (if any).
- Thoroughly examine if region plays a role in pricing.
- Determine an appropriate value-based ticket price based on data.

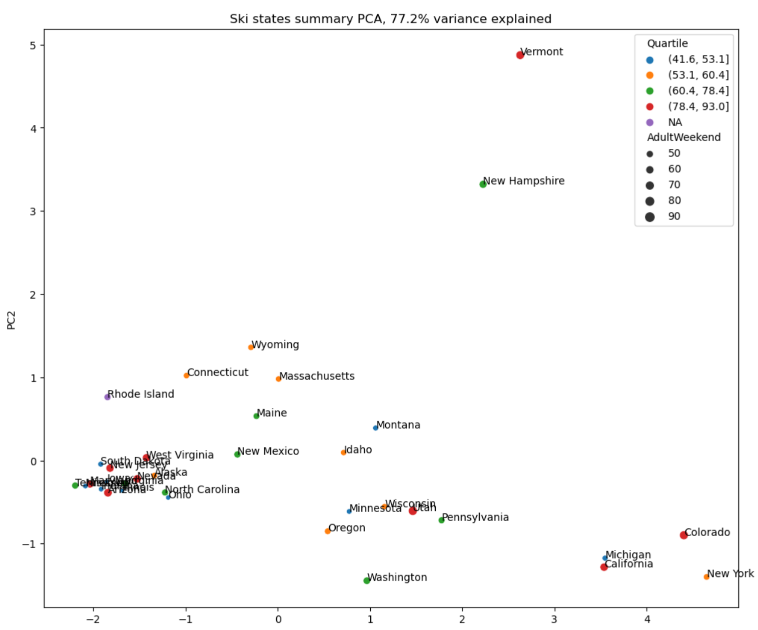
Constraints Within Solution Space
- Data sources are restricted to the provided CSV file.
- In-house metrics and financial data from BMR are not available.

Stakeholders to Provide Key Insights
- Jimmy Blackburn (Director of Operations)
- Alesha Eisen (Database Manager)

Data Source: CSV file of US ski resort data provided by Alesha Eisen (Database Manager)

Data Wrangling
Originally, the dataset contained 27 columns and 330 rows. An initial passthrough of the data resulted in 2 columns and 51 rows being removed. Irregularities in certain rows with SkiableTerrain_ac, Snow_making_ac, and yearsOpen were corrected Target feature was identified as AdultWeekend.

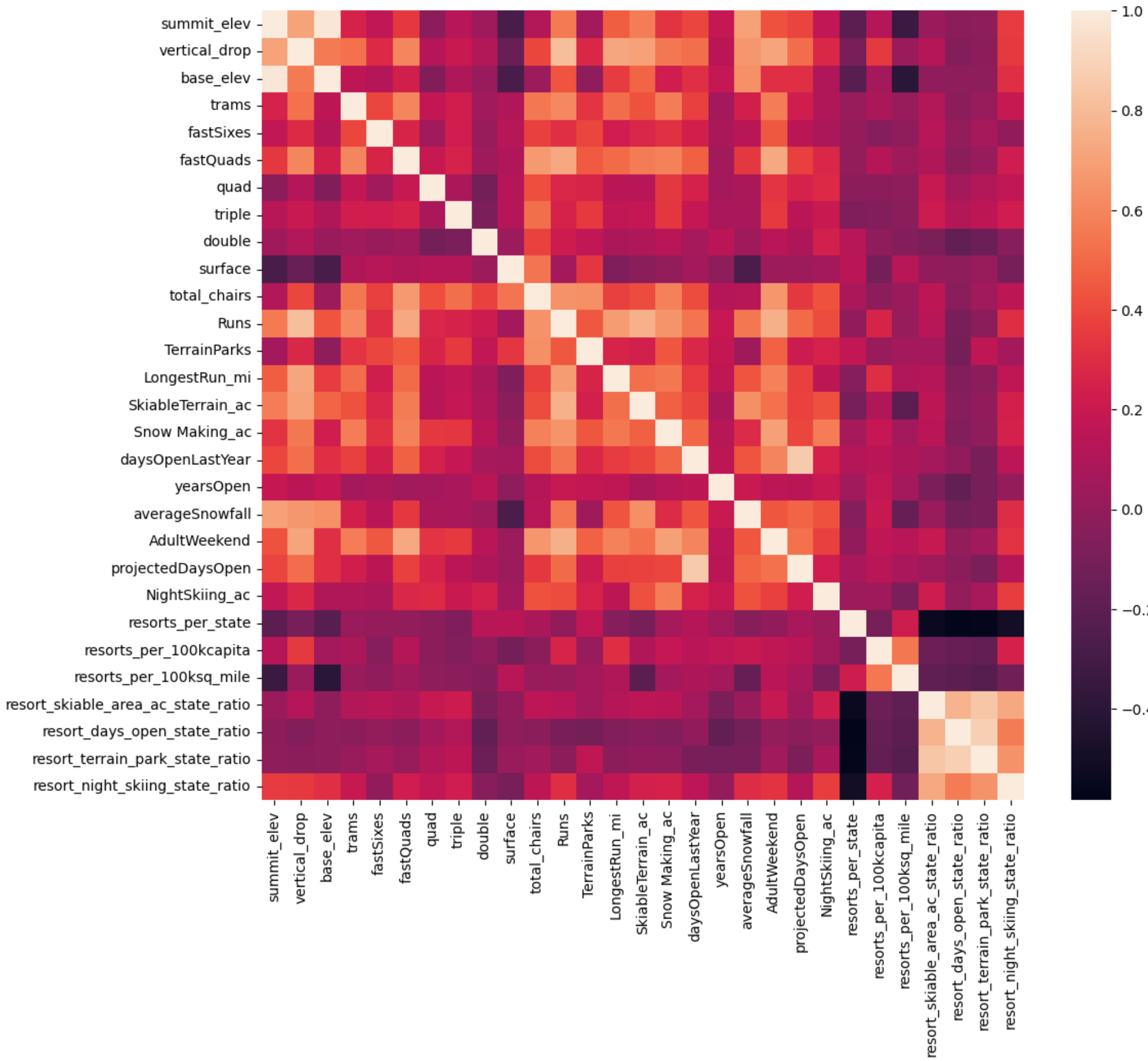
Exploratory Data Analysis
To determine whether state affected ticket price, PCA was conducted. A scatterplot was generated with price:



The lack of any obvious, distinct pattern suggested a lack of relationship between state and ticket price. Nonetheless, the following state resort competition features were derived and added to the cleaned dataset:

- ratio of resort skiable area to total state skiable area
- ratio of resort days open to total state days open
- ratio of resort terrain park count to total state terrain park count
- ratio of resort night skiing area to total state night skiing area

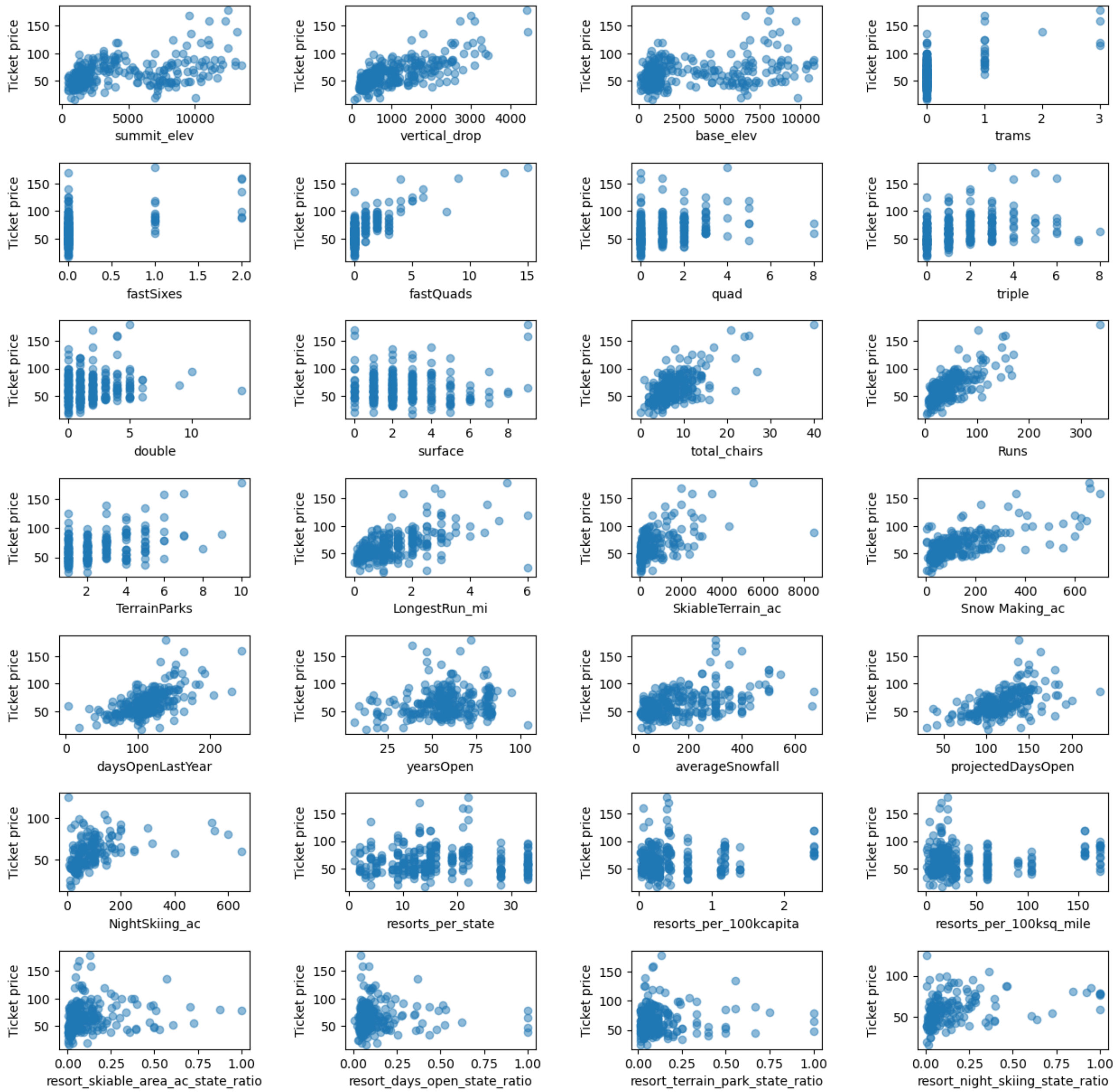
Due to the revelations from PCA, subsequent modeling and feature selection would be focused on the resort level. The following correlation heatmap was generated to gain a high level view of the relationships amongst the features:



The heatmap shows that the following features are strongly correlated to price:

- Vertical Drop
- Number of Fast Quads
- Number of Runs
- Area Covered by Snow Makers
- Total Number of Chairs
- The ratio of the resort's night skiing area to the total state's night skiing area (i.e. the resort's state-wide night skiing share)

To further examine the relationships among price and the other numeric features, a series of scatterplots were generated and analyzed.

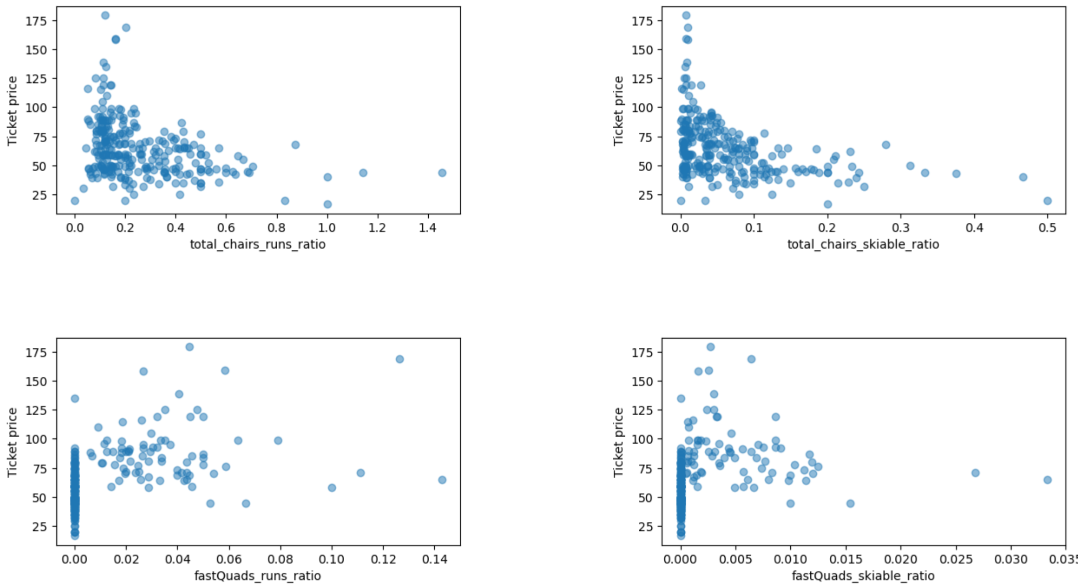


While the scatterplots supported the correlations the heatmap presented, a revelation was discovered in the price vs resorts_per_100k_capita. While the scatterplot reveals high variability in ticket price when there are fewer resorts per 100k capita, there is a slightly increasing price trend as the resorts per 100k capita increases.

To analyze how easily a resort can transport people around, the following features were derived and added to the dataset:

- total_chairs_runs_ratio
- Total_chairs_skiable_ratio
- fastQuads_runs_ratio
- fastQuads_skiable_ratio

These features were used to generate scatterplots against price.



The scatterplots revealed that resorts with a high number of chairs relative to the number of runs charged lower ticket prices. Also, with the exception of 1 resort, resorts with no fast quads seem to have a price cap at approximately under $95.

Model Preprocessing with feature engineering
Moving forward, Big Mountain was temporarily removed from the dataset in order to process the dataset for modeling.

A 70/30 train/test split was utilized.

First, a dummy regressor utilizing the mean was fit on the training data for a baseline measure of performance with the following results:

- R-squared: -0.0031235200417913944
- Mean Absolute Error: 19.136142081278486
- Mean Squared Error: 581.4365441953483

Linear and random forest regression were used for modeling. Implementation followed the following steps:

1. Impute missing values.
2. Scale the features.
3. Train the model.
4. Calculate model performance.

Both mean and median imputation were analyzed. For linear regression different imputation methods had negligible impact in scoring. On the other hand, median value imputation was superior for random forest regression.

Algorithms used to build the model with evaluation metric
Two final models were built for comparison. The first model utilized linear regression with cross validation. The second model utilized random forest regression with cross validation. Both models used median value imputation.

Mean absolute error was utilized as the evaluation metric.

Winning model and scenario modeling
Random forest regression was the winning model with the results as follows:

Method	Training MAE	Training σ	Test MAE
Linear Regression	10.50	1.62	11.79
Random Forest Regression	9.64	1.35	9.54

The following 4 shortlisted scenarios were examined with the random forest regression model::

1. Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
2. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage
3. Same as number 2, but adding 2 acres of snow making cover
4. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

When calculating revenue changes, the expected number of visitors was 350,000 and each visitor would ski on average for 5 days (a total of 1.75 million tickets sold).

The model suggests that scenarios 3 and 4 do not warrant a price increase.

In scenario 1, closing one run does not suggest a change in ticket prices and should be done immediately. The financial feasibility of closing more runs should be examined with the following considerations on revenue:

Closing X Number of Runs	Revenue Change (in millions of $)
2	-0.71
5	-1.17
8	-2.21
9	-2.99
10	-3.17

Scenario 2 increases support for ticket price by $1.99 or over the season $3.47 million.

Pricing recommendation
Currently, Big Mountain charges $81.00 while the model suggests that Big Mountain should be charging $95.87 (with an expected mean absolute error of $10.39). At the minimum without implementing any of the above suggested scenarios, Big Mountain should increase ticket price to $85.87 (a $4.87 increase or 6% increase).

Conclusion
To maintain profits, an $0.87 or 87 cent increase in ticket price will cover the approximately $1.5 million increase in operating costs. Even with this nominal increase in ticket prices, Big Mountain is undercharging and should increase ticket prices to $85.87.

An examination of the proposed scenarios suggests that scenarios 1 and 2 should be examined in more detail..

Future scope of work
Further examination of scenarios 1 and 2 need to be conducted with operating cost and installation/shutdown costs.

While there was a lack of relationship between ticket price and state, ticket prices could be analyzed by location radius.

In addition, visitor numbers, revenue information, and operating cost data could be used to build a more robust model.