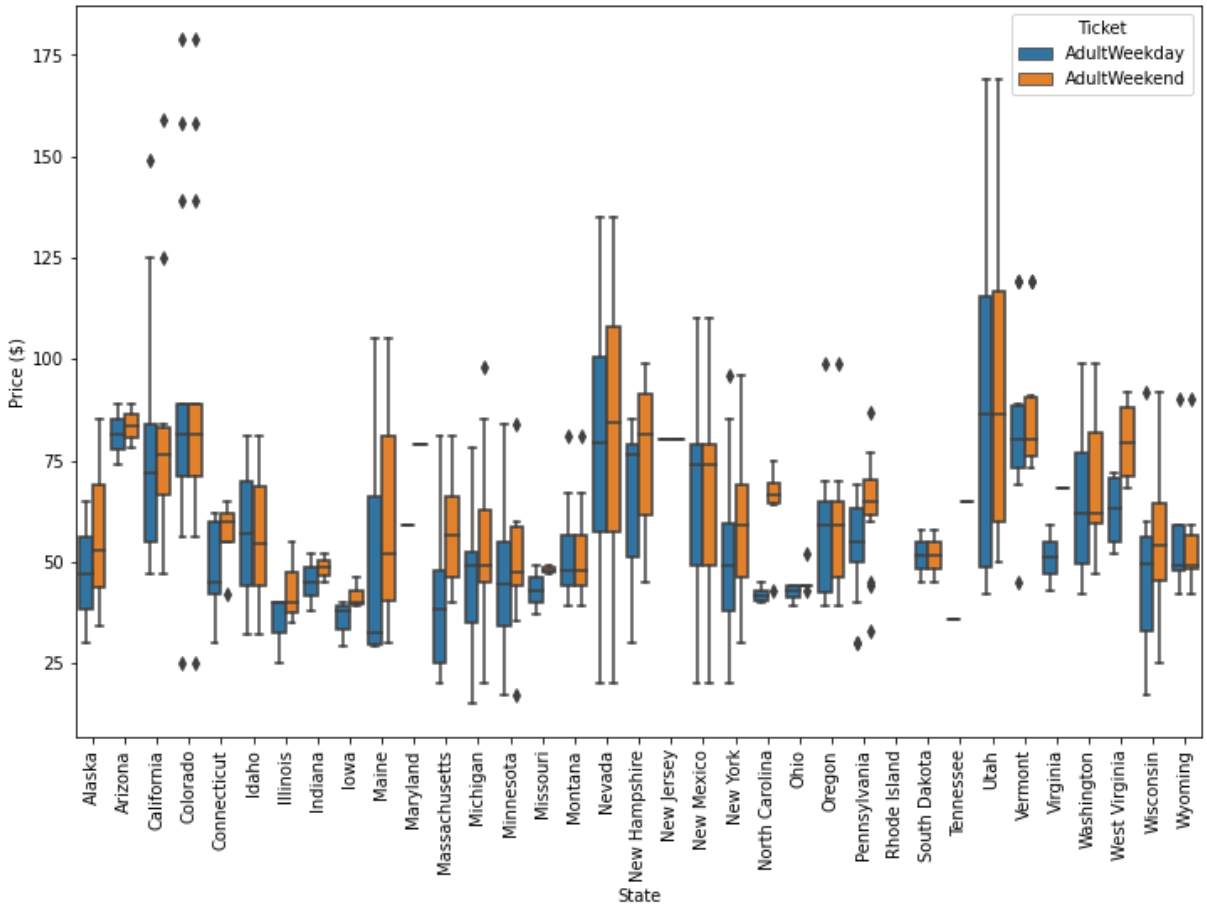
Guided Capstone Project Report – Ron Hankey

Problem Statement: The goal is to determine an optimal ticket pricing strategy for Big Mountain Resort. Project success can be measured by evaluating the impact of the recommended pricing strategy, along with cost-cutting measures, on profitability. This strategy aligns with the resort's goal of implementing a more data-driven business approach and needs to be implemented within the next year.

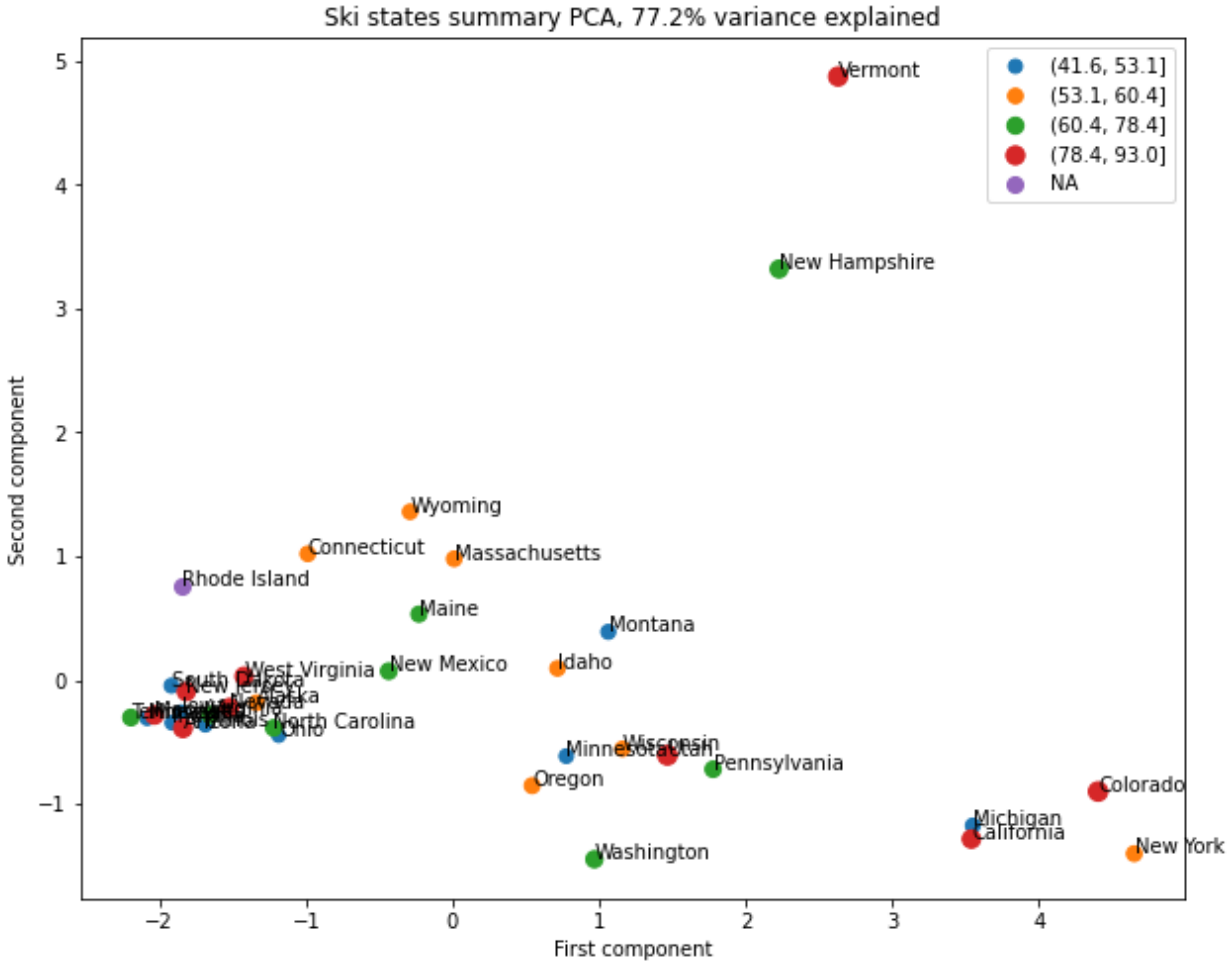
The data for the analysis was supplied in a spreadsheet in CSV format that was then imported into the Python notebook. The data was examined for missing values and fastEight had the most missing values, a little over 50% missing. The desired target value, the ticket price, also was missing 15-16% of values. In the categorical values we observed there were no missing values.

A boxplot shows that most prices appear to lie in a range roughly from $25 to over $100. Some states Price variability is considerable in some states such as Nevada and Utah, while Montana and South Dakota both have fairly small variability. We also note the weekend and weekday ticket prices while Montana and South Dakota are the same.



In the exploratory data analysis phase, in order to unravel the intricate network of relationships, principle components analysis (PCA) offers a solution. This method identifies linear combinations of the initial features that lack correlation and arranges them based on the variance they elucidate. Utilizing these resultant features facilitates visualizing data in reduced

dimensions (e.g., from 7 to 2), providing insight into the explained variance. Additionally, it enables examination of the original features' influence on these derived components.



The red points, indicating the upper quartile of price, are visible on the left, right, and top. There's also a dispersion of the other quartiles. In this depiction of ski summaries across states, which captures approximately 77% of the variance, no discernible pattern with price is apparent.

In the Model Preprocessing with feature engineering phase, we chose to use the median for filling missing values because of the skew of many of our predictor feature distributions. We then build a linear regression model and find a simple linear regression model explains over 80% of the variance on the training set and over 70% on the test set. We then set up a pipeline to compare a regression model with a random forest model and performed hyperparameter tuning using the GridSearchCV model. We perform a comparison between the linear regression and random forest models linear model: Mean MAE = 10.5 / Std Dev MAE = 1.62 / MAE = 11.79 random forest: Mean MAE = 9.64 / Std Dev MAE = 1.35 / MAE = 9.54.

The random forest has a lower MAE and lower std deviation so we choose this as our model for production

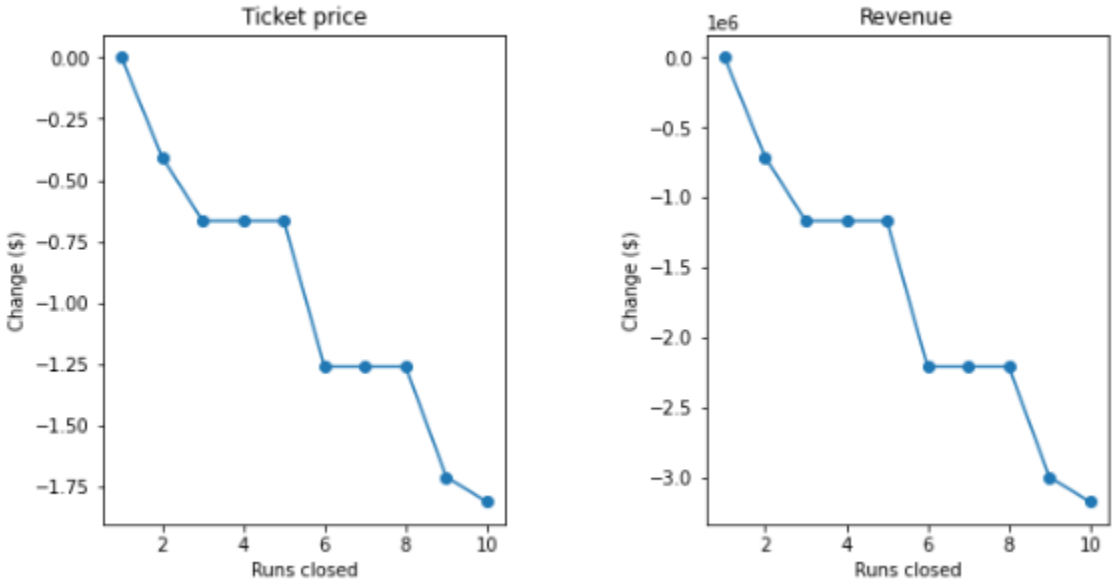
In the final phase we created a model examining ticket prices using these features which were found to be most important in the random forest model:

- vertical_drop
- Snow Making_ac
- total_chairs
- fastQuads
- Runs
- LongestRun_mi
- trams
- SkiableTerrain_ac

From the model we were able to model a price:
Big Mountain Resort modelled price is $95.87, actual price is $81.00.
Even with the expected mean absolute error of $10.39, this suggests there is room for an increase.

Big Mountain Resort has been exploring various scenarios to reduce expenses or boost revenue, particularly through ticket price adjustments. The resort has the freedom to set ticket prices independently of predefined parameters. Nevertheless, it operates within a market where pricing is influenced by the perceived value of different facilities. Understanding how facilities contribute to ticket prices is crucial business intelligence for the resort.

We also used the model to look at how cutting down up to 10 of the least used runs.



The graph above indicates that shutting down a single run has no impact. However, closing two or three runs sequentially decreases support for ticket prices and consequently revenue. If Big Mountain closes three runs, it appears that closing four or five runs would yield no additional loss in ticket prices. Expanding closures to six or more runs results in a significant decrease in revenue.