Guided Capstone Project Report

Problem Statement

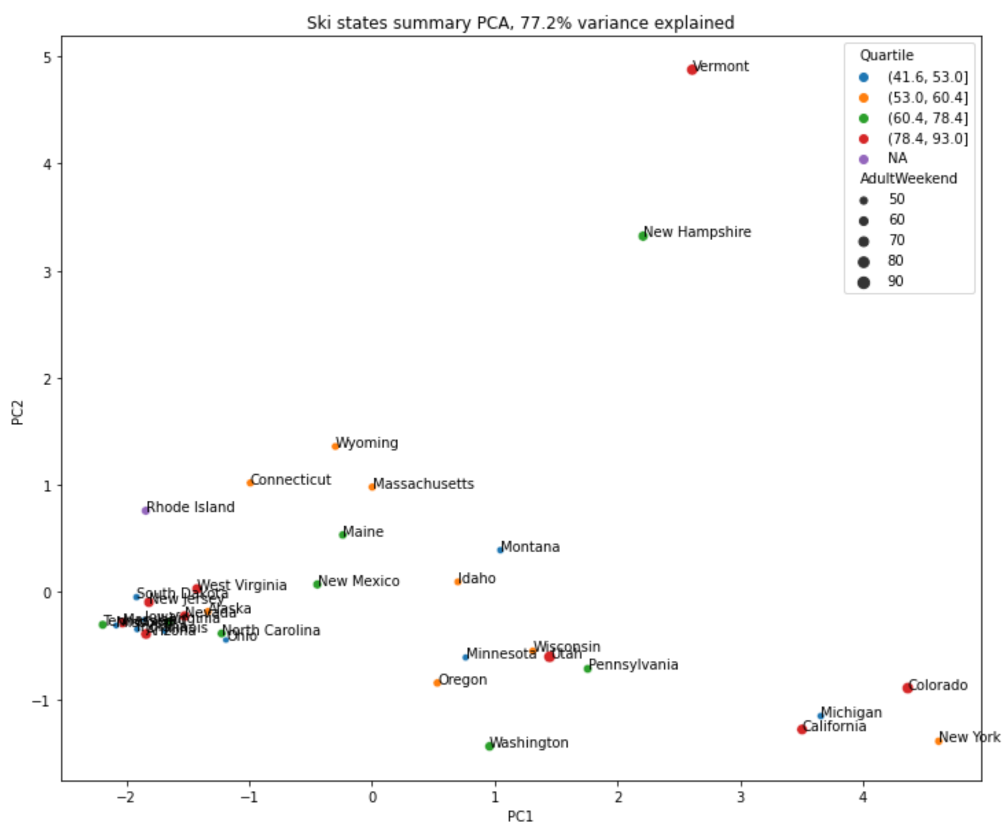
The problem statement was created with the information given in mind. Big Mountain Resort had added a new chair, which increased operating costs by $1.5M. Due to this increase in operating costs, and with the rudimentary method Big Mountain Resort had been deciding their ticket prices, they decided to use a more scientific method to determine prices, hopefully covering at least the cost of the new chair.

Data Wrangling

The only data provided by Big Mountain Resort was a .csv file with data on many resorts in the United States. The most important columns of data in this .csv file are the 'AdultWeekday' and 'AdultWeekend' columns because these are the ticket prices for the different resorts. We dropped on resort information for resorts that did not have 'AdultWeekday' prices because most resorts that posted a price posted it in 'AdultWeekend'. We also merged this .csv file with census data for the state in which each resort resides.

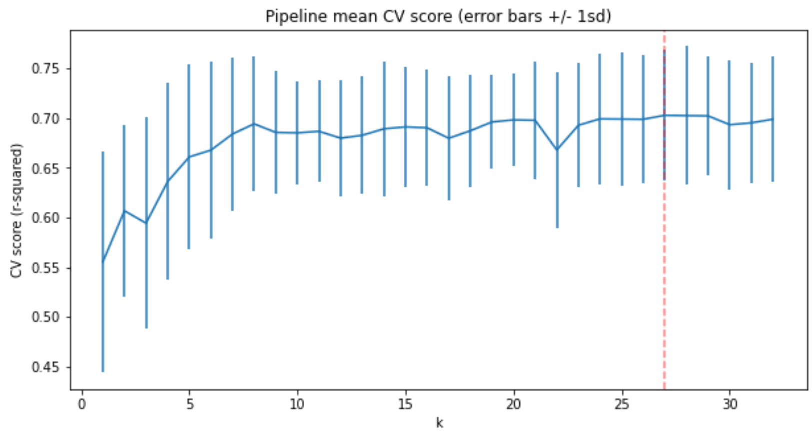
Exploratory Data Analysis

Our exploratory data analysis included first looking at resort statistics by state (total resorts, total skiable area, total days open, etc.). We then defined some new columns, such as the resort density (given by number of resorts divided by state area). We then standardized the data, since we have very different units of measurements in our data. After this we calculated the PCA of the data on ticket price. Below is a useful plot of the first two components of PCA on different state's average ticket prices. We also looked at a correlation heatmap for all of the variables.



Model Preprocessing and Winning Model

Before attempting any models, we first imputed missing values with the median of that column's value. We chose the median because many of our distributions are skewed. We also standardized the data to mean zero and variance one. We first tried a linear model and got a MAE of about $10. We then did a cross validation grid search to find the k-best parameters to use for fitting. Below is a line chart of the k-best parameters. We then did similar cross validation for a random forest linear regression model, and got a MAE of about $9, so we decided to use the random forest model as our final model.



Pricing Recommendation and Conclusion

We recommend that if Big Mountain Resort does not change any of their current facilities, they should raise their ticket price to $92.39. This is an increase of $11.39 from the current ticket price of $81, which translates to $19.9M in revenue per year, assuming the same number of skiers visit. When looking at Big Mountain Resort as compared to other resorts, they are on the high end of most metrics, and Big Mountain Resort has known it has needed an increase in ticket prices for a long time. If Big Mountain Resort gets one more run that increases vertical drop by 150 feet, while only adding one chair, they could increase their price by $16.28, which is $28.4M more in revenue per year. At the very least, Big Mountain Resort has room to increase ticket sales to match the industry average for a resort of their kind.

Future Work

Currently, we have done a small amount of pre-testing different parameters, but it would be best for a business analyst to be able to do this instead. In the future, we can make the code that predicts the new price increase given changes to parameters readily available for analysts to use. In addition, we could possibly implement the new operating costs for adding things such as chairs or more snow making acres into the function, so that we could get the new profit, rather than the new revenue, which would be more useful for making business decisions.