

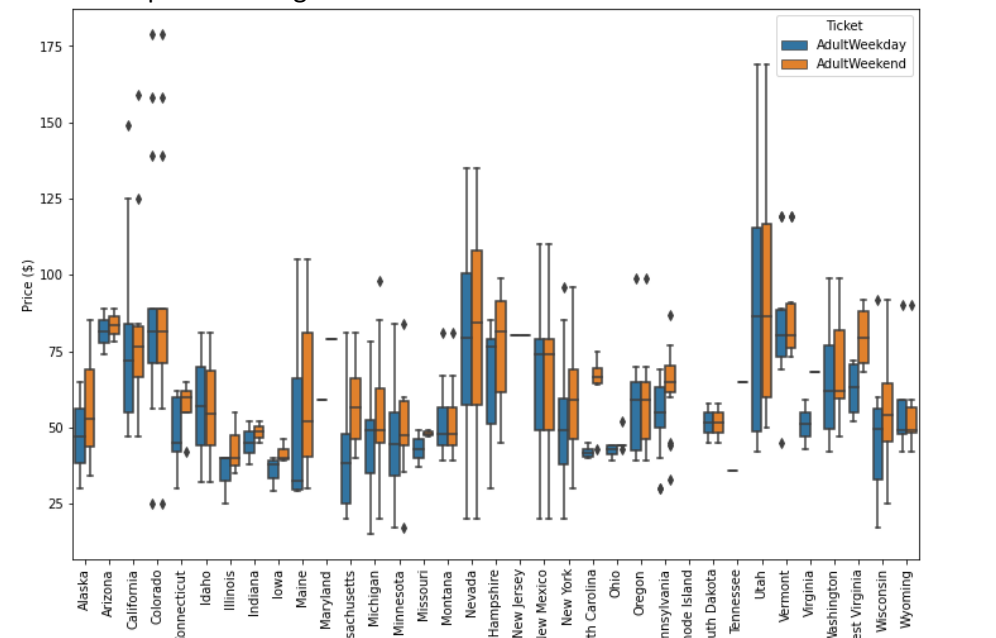
Documentation

Problem Statement

The first task was to identify the problem. It was most important to fill out the problem statement at the top, so that it would be clear on what steps I would need to take. This problem statement was to find out the price of a ticket using data we have from other resorts, so that the additional \$1.5 million operational cost can be better invested by the end of the fiscal year 2022. It was also important to fill out the problem statement worksheet, and its categories: Context, Criteria for success, scope of solution space, constraints, stakeholders to provide key insight, and the key data sources. This created a good framework to go about the problem.

Data Wrangling

Data wrangling was the first assignment. This means cleaning up the data. The first task that I had to do was import the necessary libraries and the csv data. I made sure that I had all the data that I would need. This meant that there should be no null or missing values. Exploring null values, I found some columns that no data, particularly fastEight. Between weekend and weekday missing values, weekday was missing more. Within categorical values, I found two resorts with the same name, but with different locations. This boxplot was created to show the differences between average adult weekday and adult weekend prices among the states.



There were a few questions that had to be answered after this boxplot was created. These were:

What do you do about the two types of ticket price?

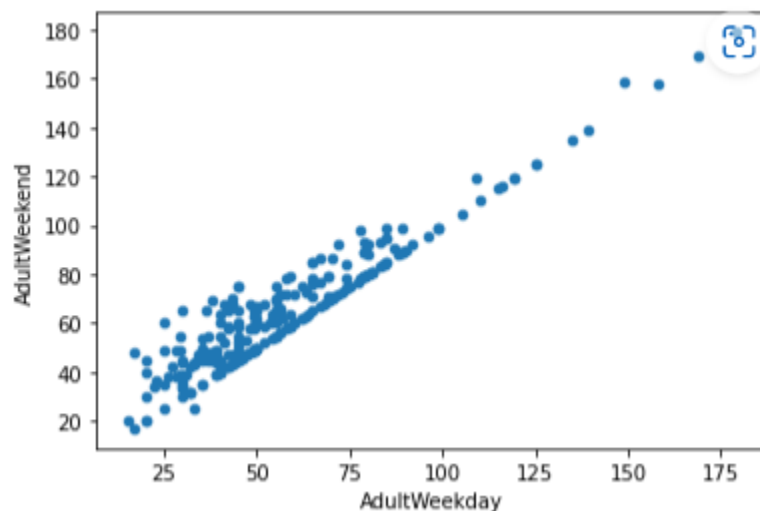
What do you do about the state information?

It was important to pull out any important information from the missing weekday and weekend prices before dropping it. Some key pieces of data were: fastEight had half the values missing. The remaining values were 0, except for one. yearsOpen had a maximum of 2019, which probably means that someone

recorded the calendar year, instead of the number of years open. SkiableTerrain_ac and Snow Making_ac had values clustered on the lower end. I verified and then replaced outliers in the data. This was only done on resorts that would not be dropped from missing price data.

I had to pull in state population and state area (square miles) data. This was done so that I could have a real-world guideline to compare to Montana, since that is where the resort is located.

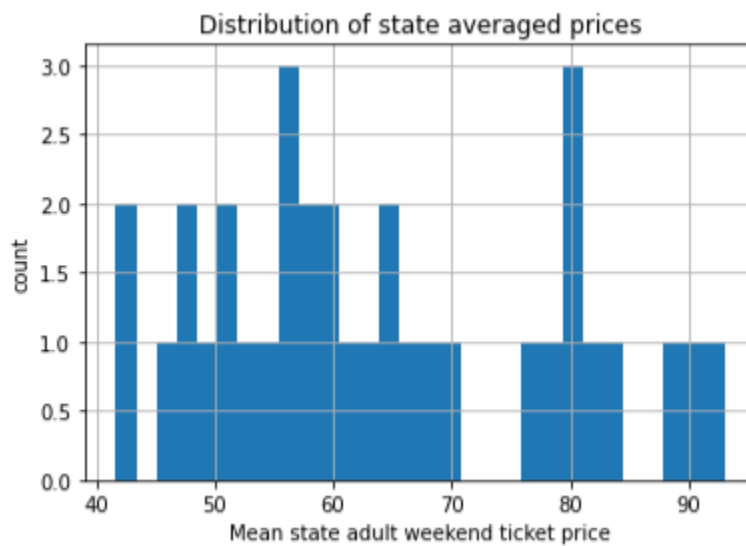
In Montana, weekday and weekend prices were equal. So, the weekday prices column was dropped until it was equal with the number of weekend prices. Finally, the two dataframes(ski data cleaned and state summary) were saved as CSV files in my folder. This scatter plot helped get a ballpark estimate on what the price of the ticket would be for the resort, and showed the relationship between weekend and weekday prices.



Exploratory Data Analysis

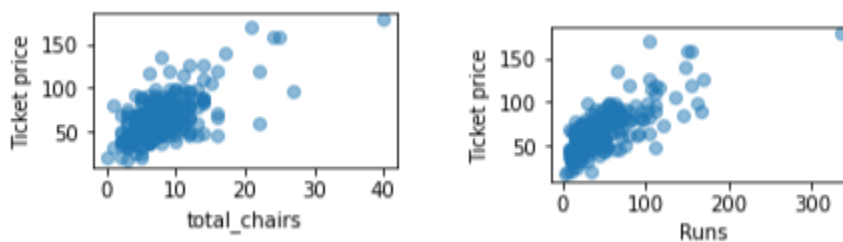
The next step was exploratory data analysis. Comparing the correlations between different states and other factors was the very first thing that had to be done. The first task that was done was to compare the States by total area, population, and resorts per state along with other skiable data. A few findings were that the total number of days open favored the states with the highest number of resorts. The total night skiing favors those in the northern most states. This could be due to the shorter days and the need to extend skiing days. A view of the number of resorts per 100k population and number of resorts per 100k square miles removes the larger states. Then, the data had to be scaled, which normalized the data.

The ticket prices were looked at next. First, the average ticket prices for weekend per state was calculated, with the data easily viewed in a histogram.



Preprocessing

PCA was used to provide summary statistics for the state/resort differentiation. From this data, we could determine the average ticket price by state. This data was easily viewed through a scatter plot. Rhode Island was the data that was not easily carried over. I wanted to keep the important data in a row before dropping anything with missing prices. However, since Rhode Island only had one resort, and it had a missing price, it stayed. A correlation heat map was generated to look for correlations in the column variables. It was determined that summit and base elevation were highly correlated along with night skiing and number of resorts per capita. Looking at "AdultWeekend", there are quite a few correlations including vertical drop, snow making, and runs. Scatter plots were generated so that the correlation of ticket prices and other variables could be visualized. The correlation between ticket prices and chairs /runs and having a quad lift could benefit the ticket price as shown below.

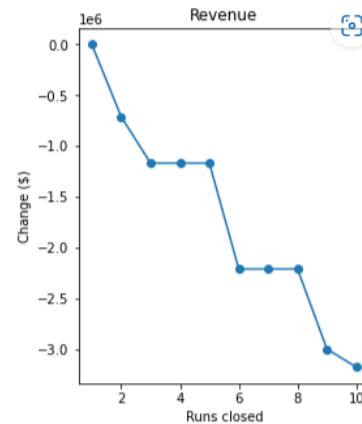


Modeling & Results

The current price of a one-day adult ticket is 81 dollars. Assuming average visitors ski for 5 days, and 350,000 visitors come in the season, the model predicts that there is a 3.47 million dollar increase in revenue from raising the ticket price by \$1.99 and adding a run 150 ft lower down and a chair lift to support it. The model, however, does not account for the additional capital expenditure and on-going operation since this data has not been provided. It is assumed that the additional chair lift would increase the operating cost by \$1.54 million.

Furthermore, closing one of the least used runs is expected to have no impact on existing revenue. Closing more runs would have a massive impact.

The price predicting model suggests the potential for raising the price above \$83, but this would mean making additional enhancements to the facility to support the price increase. The resort is already positioned in the premium segment of the market and is priced competitively compared to other resorts offering similar amenities. Since Big Mountain has a geographical advantage to compete in the premium market, upgrading the facility to charge a higher price while attracting more visitors would increase the resort's competitive advantage in a pro market cycle.

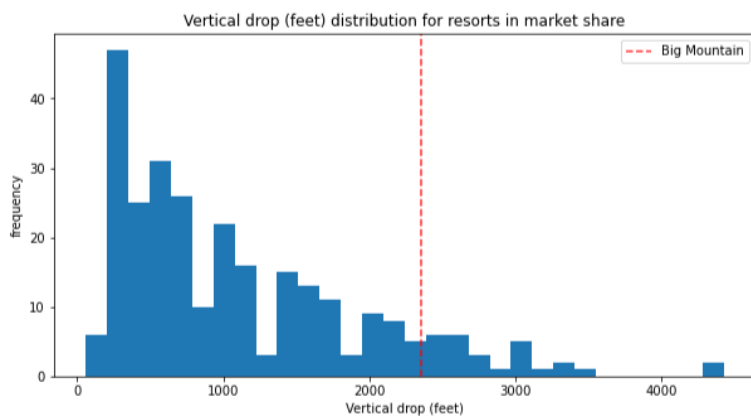


Future Research

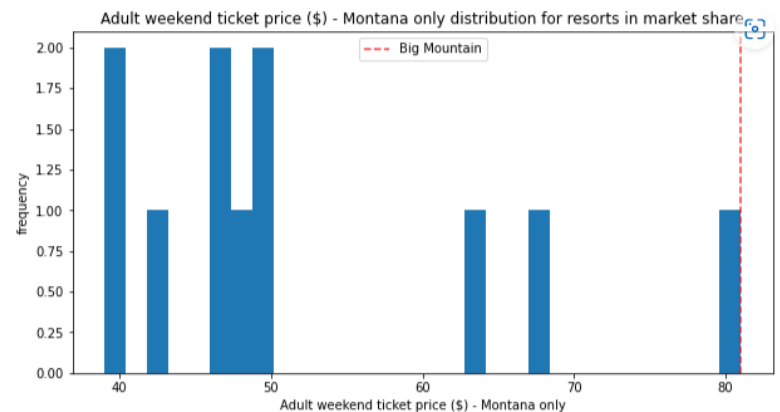
Some of the limitations was the lack of some data, including operating cost data and any additional capital expenditure and operation. It also does not include visitor volumes across the U.S.

So, considering this, improving the model can include: (1) operating cost data and (2) visitor volume across the U.S. (3) Surveying the market for willingness to pay a premium by amenities and features should be considered as well.

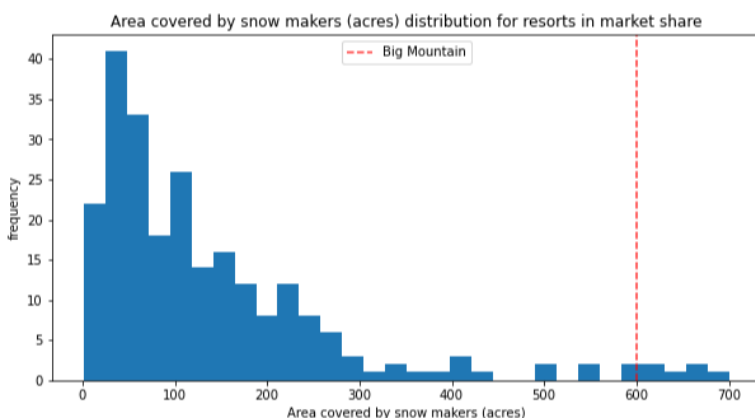
Appendix



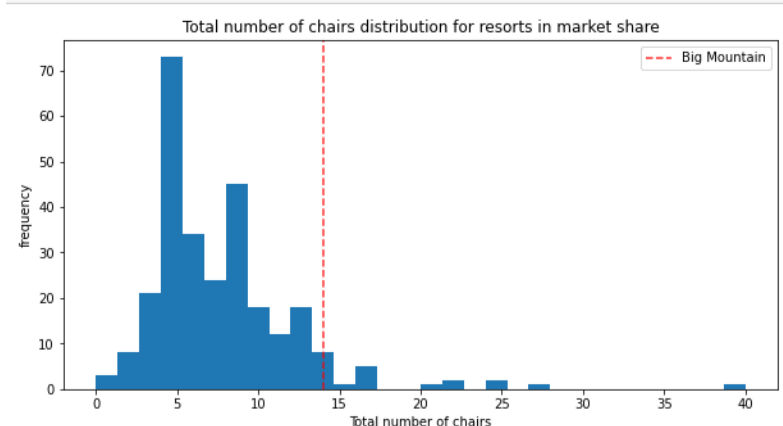
Big Mountain has a good vertical drop.



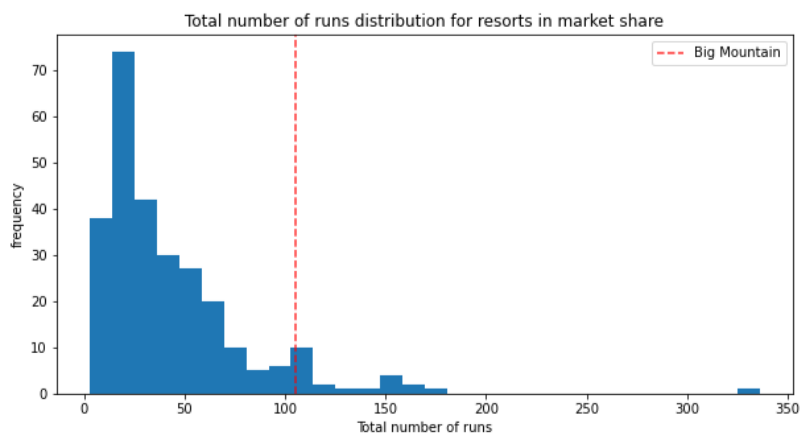
Big Mountain is a priced competitively.



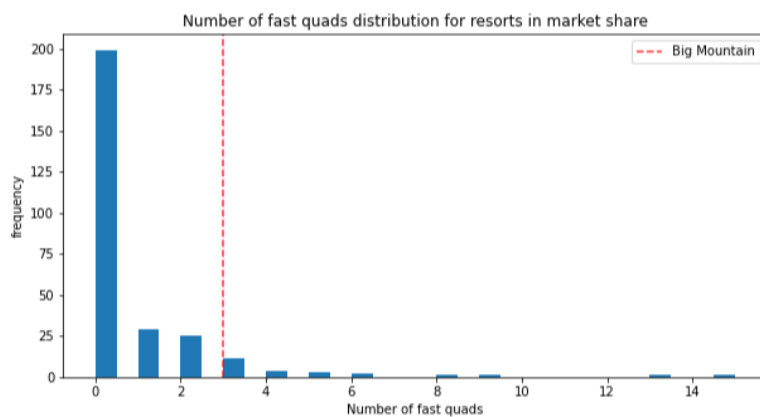
Big Mountain is high in terms of snow making area.



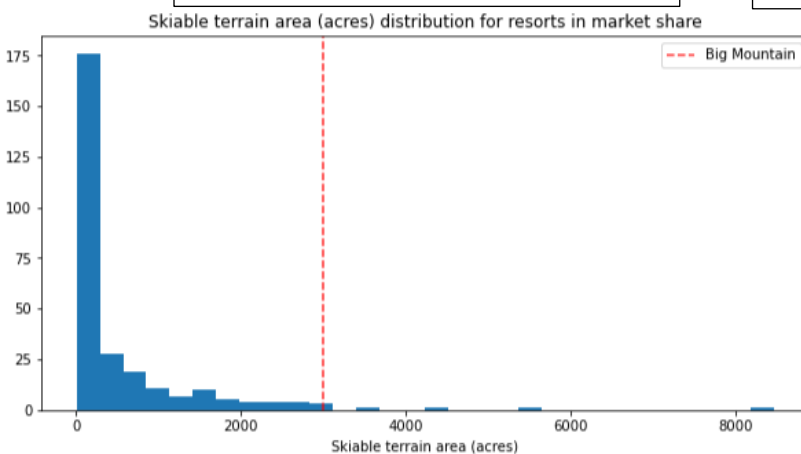
Big Mountain has among the highest number of total chairs.



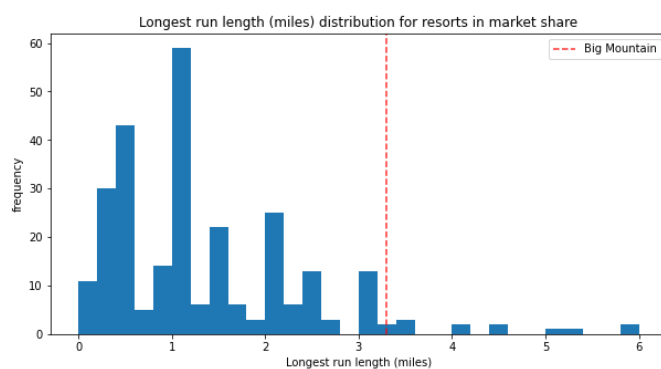
Big Mountain compares well with runs.



Big Mountain has 3 fast quads, which puts it higher than most resorts.



Big Mountain is among the resorts with the largest amount of skiable area.



Big Mountain has one of the longest runs.