Krystal Osei

Guided Capstone Project Report : Blue Mountain Resort

I. **Problem Identification:**

This project is focusing on the Blue Mountain Resort, located in Montana, NY. The resort offers spectacular views of Glacier National Park and National Forest giving their customers access to 105 trails. Every year, the ski resort receives about 350,000 customers, and one great feature of this resort is their mission to accommodate people of all levels and abilities. Recently, the Blue Mountain purchased an additional chair lift to help increase the distribution of visitors across the mountain.The amount needed to install the chair increased their operational costs to $1,540,000.00. The resort decided to brainstorm a price strategy where they will charge customers a higher premium price to cover their operational costs. However, it was difficult for them to determine the price because the company was not fully aware of how to maximize their facilities. After hearing these challenges, the data science team decided to take on this project. Through wrangling, analysis, and modeling we will compare the ticket prices at Blue Mountain Resort to other resorts in the United States, examine the performance of their facilities and features, and help Blue Mountain Resort determine a reasonable price for their customers.

II. **Data Wrangling**

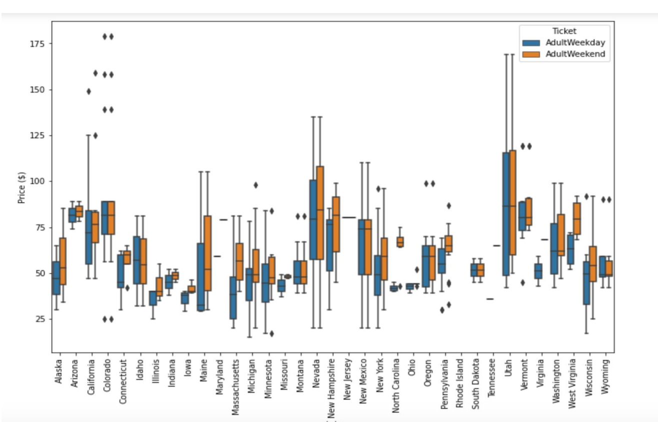
The first step taken in figuring out the right price for customers to attend the Blue Mountain required us to wraggle the data. First, we loaded the ski resort data which contains information of all ski resorts across the United States. After importing the data, the next step taken was to display the data frame and take note of the organization, information, columns, rows and applying panda methods to see the first few rows of the data frame. This is the data frame that displayed after using the head method:

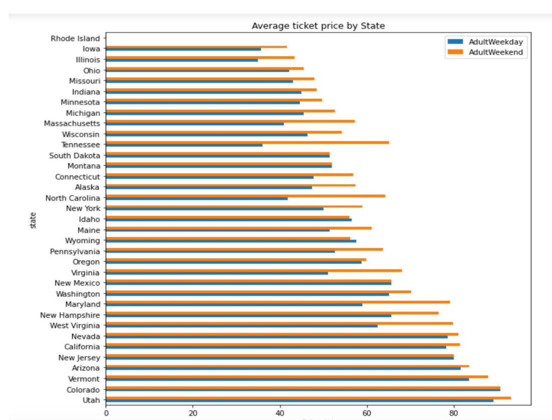
Name	Region	state	summit_elev	vertical_drop	base_elev	trams	fastEight	fastSixes	fastQuads	...	LongestRun_mi
Alyeska Resort	Alaska	Alaska	3939	2500	250	1	0.0	0	2	...	1.0
Eaglecrest Ski Area	Alaska	Alaska	2600	1540	1200	0	0.0	0	0	...	2.0
Hilltop Ski Area	Alaska	Alaska	2090	294	1796	0	0.0	0	0	...	1.0
Arizona Snowbowl	Arizona	Arizona	11500	2300	9200	0	0.0	1	0	...	2.0
Sunrise Park Resort	Arizona	Arizona	11100	1800	9200	0	NaN	0	1	...	1.2

SkiableTerrain_ac	Snow Making_ac	daysOpenLastYear	yearsOpen	averageSnowfall	AdultWeekday	AdultWeekend	projectedDaysOpen
1610.0	113.0	150.0	60.0	669.0	65.0	85.0	150.0
640.0	60.0	45.0	44.0	350.0	47.0	53.0	90.0
30.0	30.0	150.0	36.0	69.0	30.0	34.0	152.0
777.0	104.0	122.0	81.0	260.0	89.0	89.0	122.0
800.0	80.0	115.0	49.0	250.0	74.0	78.0	104.0

After looking at the dataframe, we will first explore at the price tickets, then the features, and then we will look into the resorts. As we exam and look at the visuals we will clean the data and take note of important observations that will help us in the long run to determine a higher premium price ticket.

Price Tickets: There are two important observations taken into account when looking at the prices at the resorts. First, there are two columns displaying the price tickets, *AdultWeekend* and *AdultWeekday* and second, the price column about 14% of the rows are missing price data which we will drop those rows since determining a grice is our target. and if we take visual look at the price distribution of the prices across all the states,





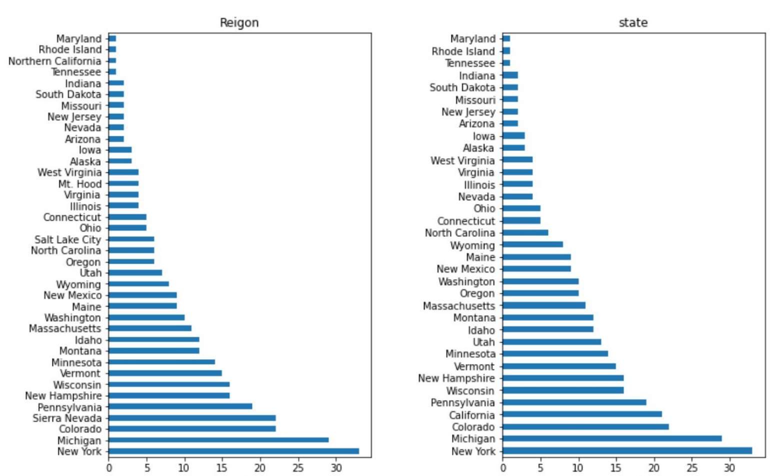
We can take note that most of the ticket prices across the states are between $25 - $100. In the box plot and in more of the populated states such as California there is more variability in their prices, but in Montana, we do not see much variability meaning leaving weekend and weekday prices are around the same price. After seeing the variability within the other states it is expected for Blue Mountain to increase their ticket price taking into account the states. The *AdultWeekend* column is missing 15.45% of values and *AdultWeekday* is missing 16.36% of its values so we will have to model the *AdultWeekend* since it's missing the least amount of data.

Features Category:

Our goal is to get the Blue Mountain to maximize on its features so when we take a look at the features there are a lot of features in regards to chairlifts. We should also focus on features relating to the altitude which are, Terrainarks, Skiable Terrain_ac, days open last year, and nightskiing_ac. When taking a look at the years open there was one resort that opened in 2019 and there was one resort that was open for just six years, so we will drop those two rows since we want to work with data that gives us a lot of data to work with. Thefasteight column was dropped since 52% of the values are missing values.

State & Resort

Taking a look at the geography and the population of the states is a great use of information in helping to see if these factors have an effect on the determining a ticket price. Below we see a graph showing that



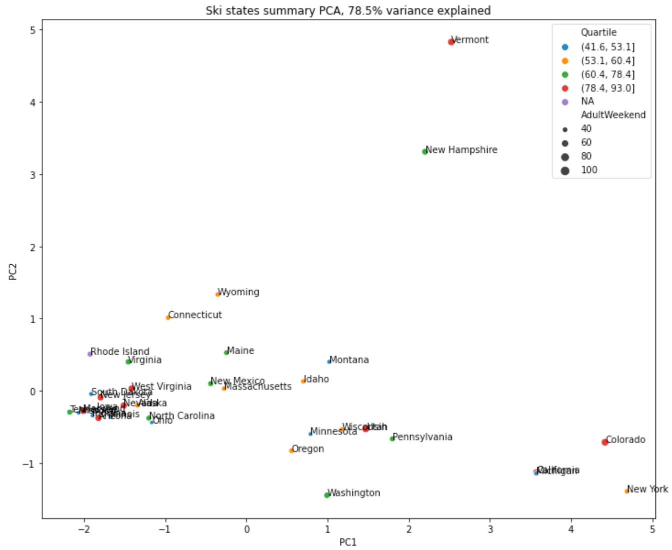
III. <u>Exploratory Data Analysis</u>

During this section of the project, our approach is to explore the data after cleaning the ski resort data. After data wrangling in the previous section, we came to the conclusion that the focus of the project will not be predicting a price but specifically, the adult weekend ticket price at Blue Mountain. This section is focused on examining the state summary statistics and finding correlation within the states and taking note if that may have any effect on how or why state prices are like that. The state summary includes information on the *Total state area, total state population, resorts per state, Total skiable area, Total night skiing data, and Total days open*. Montana is the third largest state area, it is not amongst the top five in state population. New York was number five in state population and number one in having the most resorts, but it was not in the top five when it came to . However, New York came first in most of the state summary statistics which intrigues the question of looking into the ratio of the resorts given the population and the area of the state.

We first dropped the absolute population and the state size columns of the state summary statistics so we can focus on only the competitive landscape of each state. After calculating the ratio we notice that Montant is in the top five as well as Vermont. At this point we can see that when we try to findcorrelation with the states given the area or the resorts some have a correlation while others have a correlation with other features. Our goal is to find a consistent correlation so that we can explore or data and that will aid us in determining the price. In order to find this correlation, we will first follow three steps.

1. Scale data
2. Fit PCA (principal components analysis) using the scaled data
3. Apply the transformation to the data to create derived features

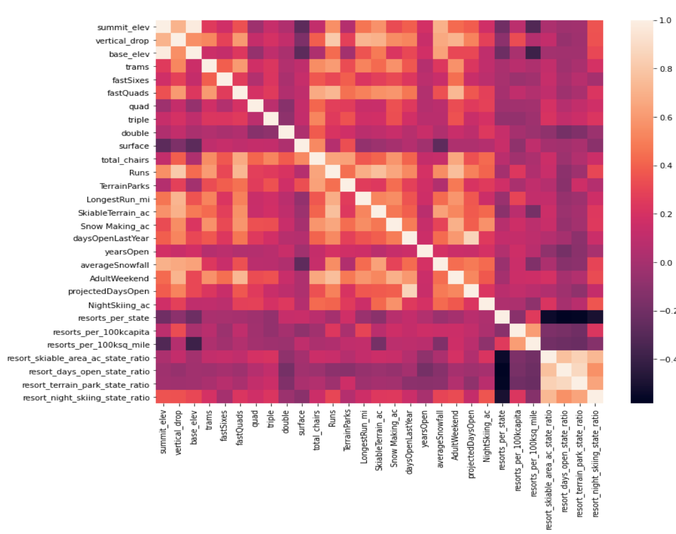
The technique of PCA will find linear combinations of the original state summary features that are uncorrelated and then order them by the amount of variance.



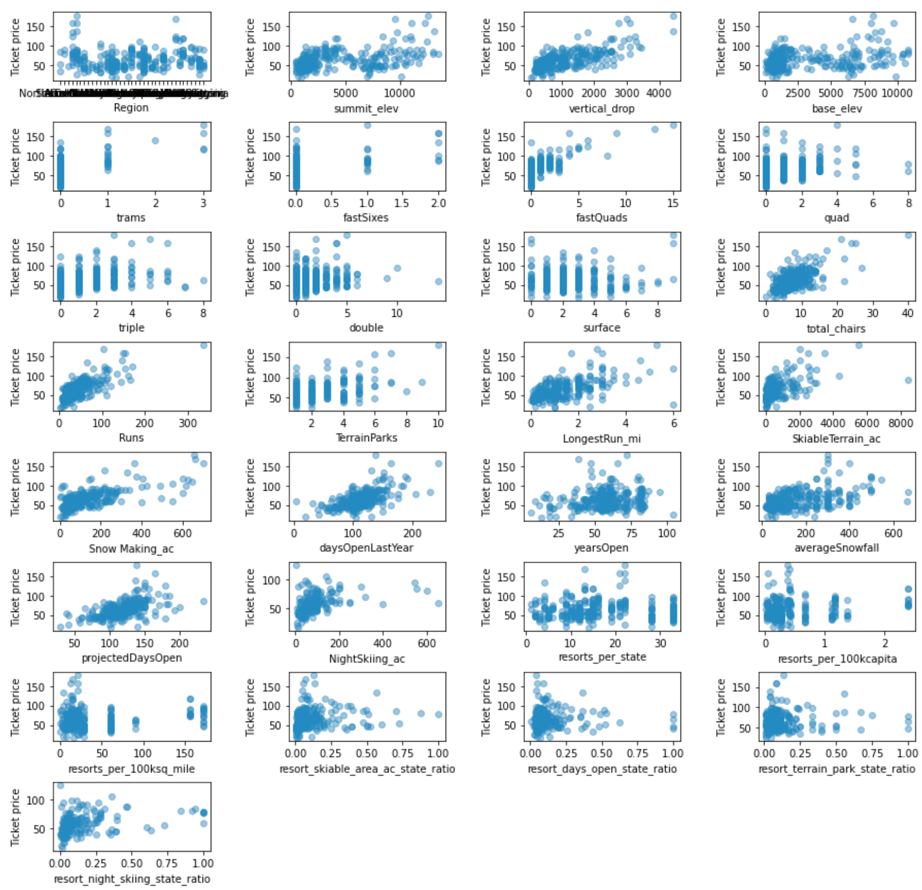
The image above gives us a good visual of the state summary statistics along with the average prices. We can see that Vermont and New Hampshi

After merging our state summary data with our ski resort data we can look into these features and gain a view of the relationships between these features below and how the *Adult Ticket Price* changes based on these ratios.

- 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



When we graph these features we get the graph below which gives information on the relation of the features and how it affects the Adult Ticket Price (see scatterplots below). First, the summit and base elevation have a high correlation to one another, as opposed to the number of resorts in each state vs have a low correlation. Looking further into the Adult Ticket Price, the fastQuads stands out, Runs, and Snow Making_ac. The more snow and runs will require more equipment which will definitely have an effect on the Adult Ticket Price.



Throughout this section we saw some correlations we tried grouping the data, but noticed that when we tried to group the state it did not make a difference or aid in influencing the *Adult Ticket Price* You didn't see any and decided to treat all states equally; the state label didn't seem to be particularly useful.

IV. Pre-Processing and Training Data Development

In the previous sections, we identified the root problem, cleaned the data, dropped rows and columns that did not give much insight, and we explored the data future and looked at how the feature may affect the ticket price we then decided to. We thought grouping the states may have an influence on determining the price, but after analysis we decided to treat all states equally; the state label didn't seem to be particularly useful. Now with exploring we are now at the 4 Data science steps and we are going to build our model.

We will then dive into doing a train/test technique to create a model that helps us not only for the data we have now but the data we will have in the future. When cleaning or exploring the ski resort data we have used the data as one data, but in order to build a model and if we train all of our data we will not have data left over to test and apply our model to see if it works. So the first step we will take is to partition the data in a 70/30 test split so we will have data to train 70% of the data and build our model and then we will test our model on the 30% of data, so in the future the company can continue to use this model when they accumulate new data about their features and other resort.

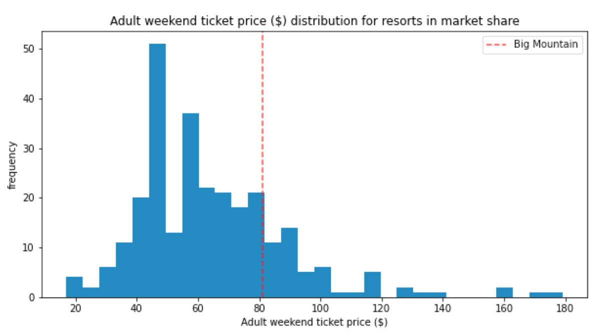
1. First we calculated the mean which will be the average price and used this to guess the price ticket.
2. Then we build the model based on mean absolute error: $MAE = \frac{1}{n}\sum_{i}^{n}|y_i - \hat{y}|$
3. We calculated the mean absolute area and when we applied the model it stated that we will be \$19.00 off if we guessed the tickets based on the average.

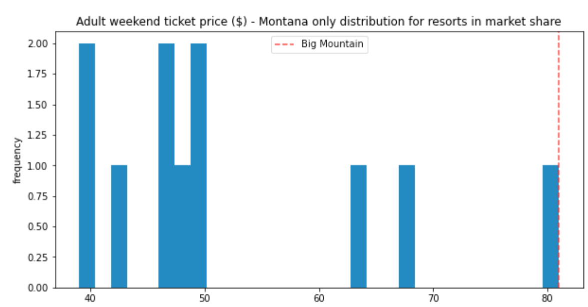
V. Modeling Checkpoint

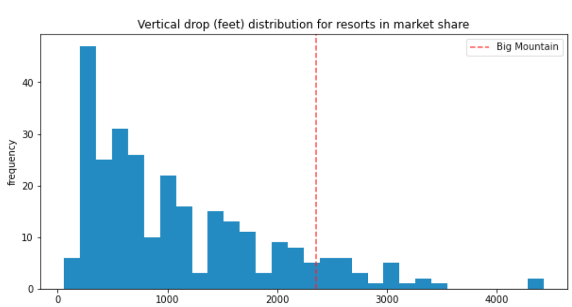
In this section, we will take our model we built in section four, for ski resort ticket price and use it to gain some insights into what price Big Mountain resort might determine as the *AdultWeekend* ticket price. As we use the model, the price will be predicted on the fact that Blue Mountain set their prices within a free market so prices will depend on how customers value their features and services provided at the resort.

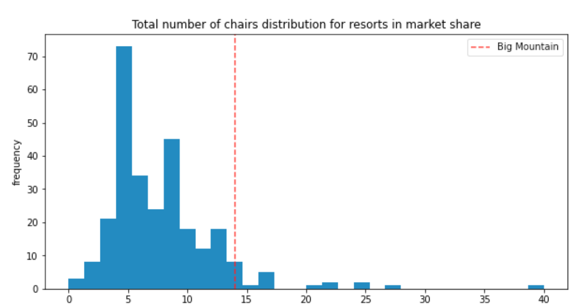
First, we refit the model on all ski resort data excluding the Big Mountain Resort. Then there was an expected mean absolute error of 10.42. Then we calculated the expected ticket price at Big Mountain using the model and the modelled price is $104.03 while the actual price is $81.00. This suggests Blue Mountain is undercharging their customers which is hurting their revenue and investments. The question of why Blue Mountain is not charging their customers is what we can look into further, but taking note of how the Blue Mountain resort provides great features and facilities may encourage the CEO to see the resorts worth and increase their prices considering 350,000 people attend Blue Mountain yearly..

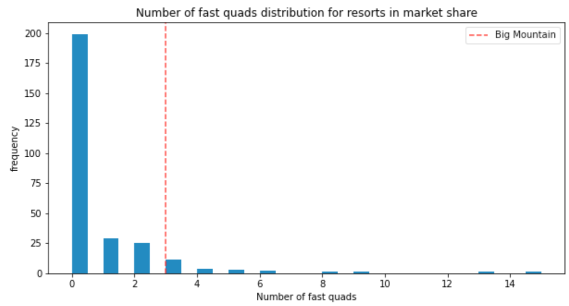
We then took a look at how Blue Mountain is doing in the market context and visually seeing how Blue Mountain is doing compared to the other resorts in within these eight features *vertical_drop, Snow Making_ac, total_chairs, fastQuads, Runs, LongestRun_mi, trams, and SkiableTerrain_ac.* Looking at these features we may be able to get a sense of how the features impact the ticket price that may influence the approach to increase the *AdultTicket* prices at Blue Mountain. The graphs below give us valuable information about the features at Blue Mountain that came up when modeling the data.

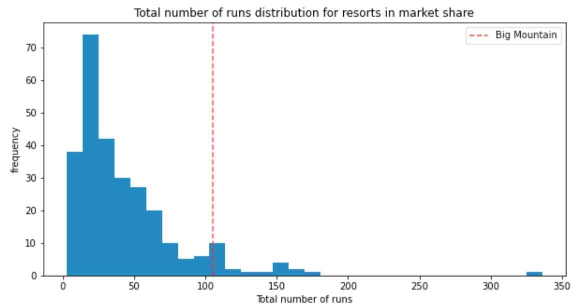


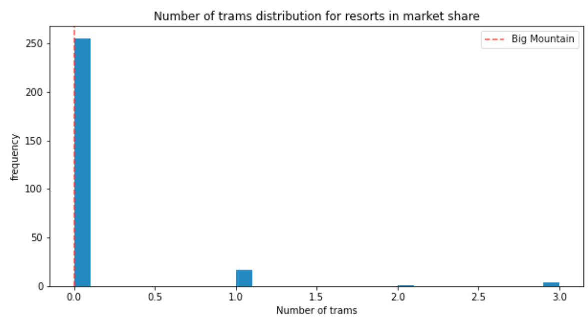


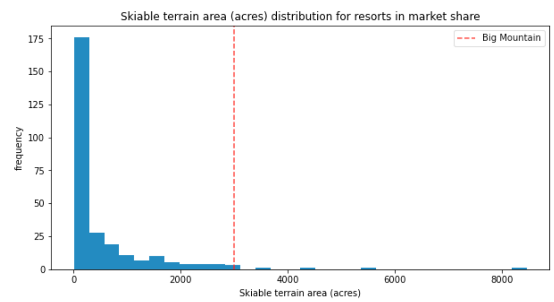












From the graphs, Blue Mountain stand out in these seven ways:

1. Big Mountain contains has vertical but it is not as high as the other resorts
2. The Big Mountain resort is top in making snow making areas.
3. Big Mountain contains a high number in total of chairs.
4. Big Mountain also has three fast quads which is more than most of the other resorts.
5. The vast majority of resorts, such as Big Mountain, have no trams.
6. Big Mountain has one of the longest runs.
7. Big Mountain is amongst the resorts with the largest amount of skiable terrain.

From the graphs above, we can take note of how the Blue Mountain stands out amongst the other resorts. We can now use our model in order to see if using these features where Blue Mountain has an impact on the *Adult Ticket* price.

Below are four possible scenarios Blue Mountain can take to increase their ticket prices keeping in mind the features:

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,
3. Add 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.

However, after using our model on each scenario, the *AdultWeekend Ticket* price increased only by $1.36 which is not much of an increase. It will be best if Blue Mountain increased their ticket price from $81.00 to $104.00 given how they stand out with the features and facilities they offer and not based their price increase on adding or increasing their features.