Brad Johnson

Guided Capstone Step 6

2/9/2023

Big Mountain Resort, a ski resort located in Montana has reached out for assistance in regard to increasing revenue. Reading through the provided information, it was clear to me that they were not capitalizing on their strengths and facilities enough and needed direction on how and why they should increase their ticket pricing to achieve the sought gain in revenue. A data set was provided that included information on all 330 resorts in the same market share. This included information on summit elevation, vertical drop, runs, ticket prices, and several more. This is what I worked with to create a plan for the resort.

Upon importing, the data set started with 330 rows and 27 columns. The rows represent the resorts while the columns represent various features and weekday/weekend prices. I began by calling the desired resort, Big Mountain Resort, to see if the data file existed and fortunately it did result in a report without any missing values. I inspected the info and explored the differences between region and state while plotting a bar graph to illustrate the number of resorts in both each region and state. I then proceeded to graph the average ticket prices by state to compare how individual states set their ticket pricing in not only total but variation from weekday to weekend. While studying the distribution of features and seeking out any outliers or interesting values, I discovered that the skiable terrain for Silverton Mountain, Colorado had an incorrect data point and was an obvious outlier. This was easily modified and adjusted for by using the .loc accessor. The fastEight column also was missing half of its values and nearly all the others were at 0. This column was deemed useless and was dropped from the dataset. In yearsOpen, a value of 2019 was discovered. It is hypothesized that this might mean the resort has yet to have opened. Either way, the row was determined to be impractical to use thus was also dropped from the dataset. Isolating features to those that would be interesting to our concern, I was left with TerrainParks, SkiableTerrain, daysOpenLast year, and NightSkiing. These were summed using aggregations. Focusing on price, 14% of the rows were found to have no price data at all. These rows were of no use to our concern and were removed from the data set. Another error was discovered while diving into population and area data for US states. A few states were listed as commonwealths and not states which created confusion. By removing the brackets and contents, I was able to safely include these states back into the dataset. From there, one more column was dropped. In Montana, weekday and weekend prices were identical thus I only need to use whichever row had the least amount of missing data, which was weekend prices. The Weekday price column was then dropped. This left me with 277 rows and 25 columns for the final version of the dataset.

Once providing myself with a much cleaner and efficient version, I began exploring this data set by ordering states based on their statistics. This revealed New York as a front runner for most categories as it boasts both a high population and a high number of resorts. This information presented some potentially relevant features for each state. In order to dive deeper, I used principle components analysis. Using PCA, I created a scatterplot to get a better look at the relationship between state and ticket pricing, but unfortunately was unable to notice any obvious pattern. This leads me to believe it makes more sense to treat all states equally, and to build a price model around that idea. From there I began to study the relationship between features and ticket prices as this would reveal which features could help the most in our business strategy. This data showed the following features to be of high correlation: vertical\_drop, fastQuads, Runs, and total\_chairs. The relationships of these features are hard to gauge at this point since we do not have data on the number of visitors per year, however it is does appear that having no fast quads may limit ticket sales.

Next, I began preprocessing by testing the mean of the set to see if that would be a fair predictor of ticket pricing. Using mean absolute error, it was determined that based on an average of known values, you would expect to be off by around $19 which is obviously not what I was looking for. From there I created a linear model using a pipeline. This dropped our mean absolute error from 19 to about 9 which is significantly better. Cross validating allowed me to both obtain an estimate of uncertainty in the performance estimate and discover which features saw the most positive return. These features include vertical drop, snow making, runs, and fast quads, all of which were previously seen in the EDA work earlier. A random forest regressor was tried using the median as the input, as that was determined to be the best parameter. This marginally improved on the CV results and suggested the same four features. I chose to continue going forward with the random forest model because it had a lower mean absolute error and showed performance consistent with the CV results when used on the test set. This model was saved for future use.

Using the newly designed model, it is suggested that the price should be as high as $95.87. Currently Big Mountain Resort is selling tickets at a price of $81 per day. I strongly believe the pricing should be increased to at least $90 as it was explicitly shown through generated charts that Big Mountain Resort is in the upper echelon of nearly all features that were deemed valuable. This price increase, especially knowing that many visitors typically buy five-day tickets, would increase revenue tremendously. Over the course of the season, a price raise of only $9 would increase revenue by a staggering $15 million. This will easily offset the operating costs of the new chairlift ($1.5M) and provide much improved financial stability. Running tests on the listed scenarios, only scenario two has any amount of significant expected price change. By adding the run, increasing the vertical drop, and installing an additional lift, it is predicted to support a price increase of nearly another $9. This would provide an additional $15 million in revenue which would offset the price for the operating costs on this new chair lift as well. This seems like a strong option with incredibly positive upside if the management is interested in pursuing.

Overall, I feel like this project was a success and that Big Mountain Resort will be very happy with my findings. Despite the success, I do believe that receiving more information on cost and resort expenses would greatly aid in a more complete understanding of the business issue. How much do you charge for rentals, if any? Is there lodging available? If so, does that factor into resort revenue? Parking fees, concessions, entertainment, etc. all would be appreciated information. Ultimately the model worked well and illustrated that the high ranking of Big Mountain Resort in the features we deemed valuable was proof that an increase in ticket prices is justifiable and encouraged. Going forward, I could save and package this model and distribute it to their analyst team to use for future concerns. I could provide some insight on how to operate and then allow them to play with different scenarios like the four I covered in this project. This will ensure success not only in the short term, but hopefully in the long term as well.