Anthony Szabo
Guided Capstone Project Report

Throughout my time working with the ski resort data, the goal was always to see what changes could be made to ticket prices in the next 12 months to see a 10% increase in profit. Big Mountain Resort is a very popular resort in Montana that offers as much as any resort in the country. They initially charged $81 on weekdays and weekends to adults, and I think they can raise that price, since they offer a lot when compared to other ski resorts in the country.

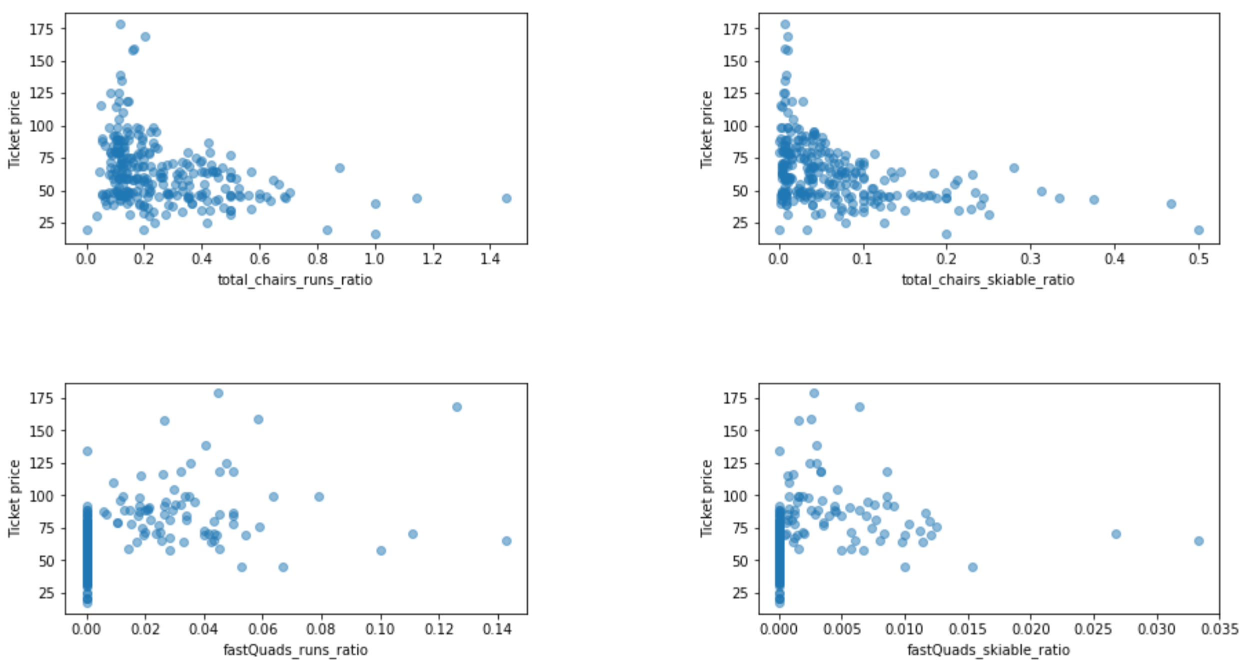
At first I was given a CSV file of all ski resort data of 329 of the most popular ski resorts in the country. Right away I was able to find that there were a lot of categories that many resorts had missing data in, for example fast eights or night skiing acreage. It is also noteworthy that we were missing at least 1 of the two values for ticket price in just over 17% of resorts. All of those instances were dropped because there's not a whole lot we can do without the ticket price. During this phase, we were able to get a good look at where all of this information was coming from geographically. The regions with the most resorts were New York, Michigan, Sierra Nevada, and Colorado, while the regions with the highest ticket price were Utah, Colorado, Vermont, and Arizona. It is interesting to see Colorado on the top of both of those lists, and it speaks to the skiing market in that region. There were a couple spots where I manually added in data, such as for snowmaking acreage in Heavenly Mountain Resort, as well as spots where I amended incorrect data, such as one resort being reported as being open for 2019 years. Originally, there were 330 rows and 27 columns of data, and Big Mountain Resort was in there. Some rows that got removed were rows without pricing information, and we had to remove the Fast Eights column because more than half of the resorts had no data for that. In the end, there were 277 rows and 25 columns of data that I brought over to EDA.

This data can be displayed and interpretted in so many different ways. This data featured statistics on each of the parks including how many of each type of run the certain park had, how many chairs they had, how many days they were open for, amongst more. There was not a direct correlation between ticket price and state unfortunately. Instead, we had to look into which states had a lot of runs, or how many chairs they had, or even how much space skiing took up in the certain state. The first thing I noticed was larger western states such as California, Colorado, and Montana, as well as north-eastern states such as Vermont, New York and New Hampshire, had the biggest hands in the skiing market. Additionally, it is noteworthy that the each state that makes up New England with the exclusion of Maine (Vermont, New Hampshire, Rhode Island, Massachusetts, and Connecticut) make up the top 5 list for resorts per 100k sq miles. I think that speaks to the culture in that area where those states are able to keep so many resorts open for the small amount of total land that the state actually takes up. Those states' residents are such a big skiing market that they all go and drive to those mountains and resorts on the weekend or even as a day trip. It is fair to assume the opposite for the larger western states (California, Colorado, Montana). Those states are likely places where people take trips to for longer periods of time. We also got some interesting insight [1] into total chairs to runs, total chairs to skiable terrain, fast quads to runs, and fast quads to skiable terrain and how they relate to ticket price. Not only does there seem to be no correlation, in the case of total chairs to skiable terrain, there seems to be a negative correlation. It is good to keep in mind that those ratios specifically don’t contribute to ticket price exactly.

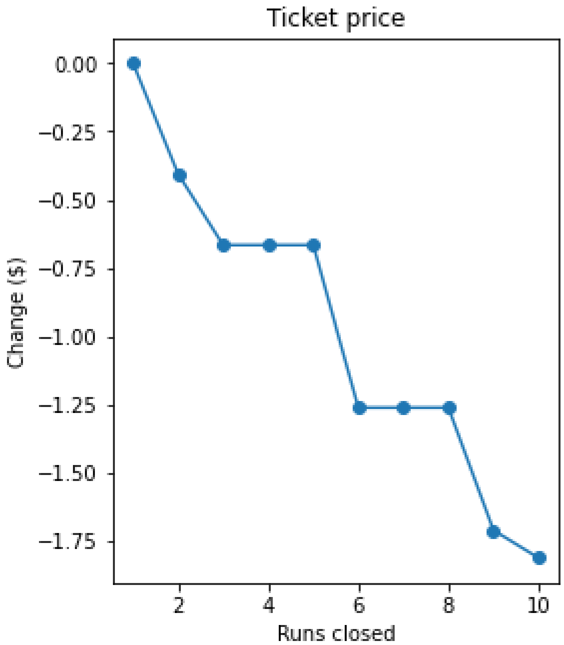
On preprocessing and training, we began with testing our data with some useful error metrics such as mean absolute error, mean squared error, and r squared. We then fit the data with pipeline and linear regression models, calculating those same error metrics, with the linear regression model showing similar results and the pipeline model showing results exactly the same as if we filled the missing data with the median. Next, I used GridSearchCV to find the best k parameter for the pipeline model to fully optimize its r squared value. Finally I used a random forest model to calculate the same error metrics as previous. In the end, the first pipeline model was the best one I created and was the model I took over to modeling.

Big Mountain currently charges $81 per ticket, while our modeling recommends a price of $95.87. Taking into account expected average error of 10.39, that would still leave us at $85.48. Immediately, I would suggest a change of ticket price from $81 to $85 to take into account the current market in respect to what Big Mountain is offering, with consideration for another price increase in the future. I would approach business leadership with the model showing where we are in term of ticket price across all resorts in the US, the model showing how many chairs we run compared to other resorts, the model showing the number of runs we have compared to other resorts, the model showing how many skiable acres we have compared to other resorts, and the model showing our vertical drop compared to other resorts. The set of charts shows us seemingly around the 70th percentile of ticket prices, while the other charts show us in at least the 85th to 90th percentiles. Although it seems like scenario 3 is wasting money adding more acres of snow making for no extra revenue, I would argue that a new run 150 feet lower down would not bring in as many new skiers as a totally new run would. I would recommend scenario 3 for further consideration given the cost of installation for the new chair and snowmakers, plus the cost of maintenance for the snowmakers. I would also note revenue costs in the change in profit vs runs closed chart [2] to management in case they were debating closing a run in the future.

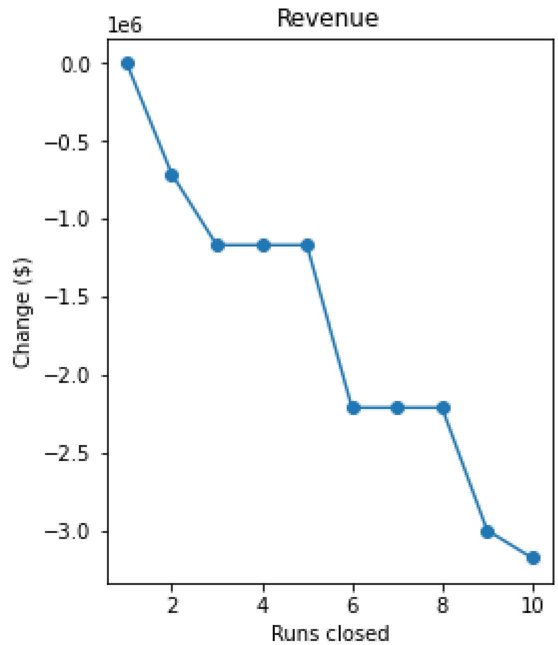
[1]



[2]



Ticket price
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
Change ($)
2
4
6
8
10
Runs closed



Revenue
1e6
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
Change ($)
2
4
6
8
10
Runs closed