Big Mountain Resort Project Report

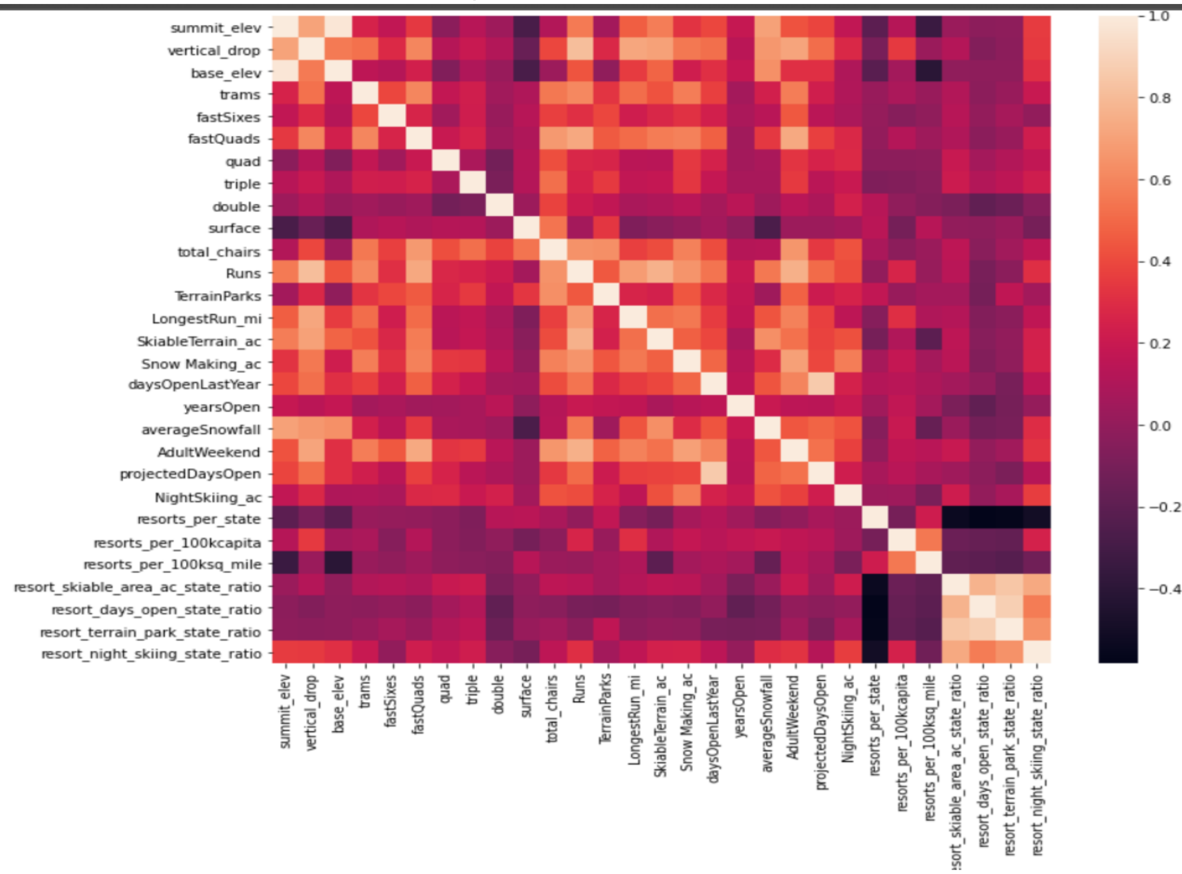
The problem we are facing is that Big Mountain Resort bases their pricing just on the market average but that price does not reflect the value of the facilities they offer. They also have an increase of 1.54 million dollars in operating costs. We want to figure out what facilities we can cut down on while still maintaining their ticket value. Also, Is a higher ticket price while cutting facilities feasible? Based on this problem, our project is trying to build a predictive model for ticket prices based on a number of facilities or properties at the resorts. We can then use this model for Big Mountain's pricing and future plans. The following is the *statistical summary of the numerical columns.*

]:

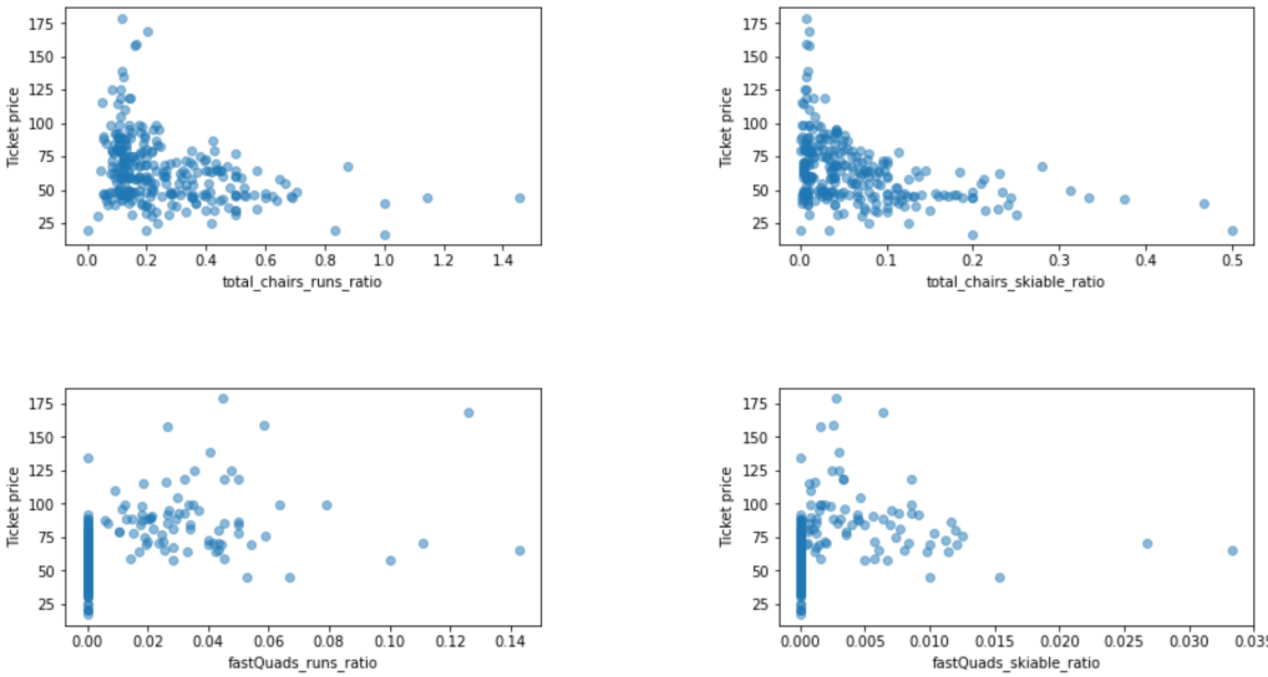
	count	mean	std	min	25%	50%	75%	max
summit_elev	330.0	4591.818182	3735.535934	315.0	1403.75	3127.5	7806.00	13487.0
vertical_drop	330.0	1215.427273	947.864557	60.0	461.25	964.5	1800.00	4425.0
base_elev	330.0	3374.000000	3117.121621	70.0	869.00	1561.5	6325.25	10800.0
trams	330.0	0.172727	0.559946	0.0	0.00	0.0	0.00	4.0
fastEight	164.0	0.006098	0.078087	0.0	0.00	0.0	0.00	1.0
fastSixes	330.0	0.184848	0.651685	0.0	0.00	0.0	0.00	6.0
fastQuads	330.0	1.018182	2.198294	0.0	0.00	0.0	1.00	15.0
quad	330.0	0.933333	1.312245	0.0	0.00	0.0	1.00	8.0
triple	330.0	1.500000	1.619130	0.0	0.00	1.0	2.00	8.0
double	330.0	1.833333	1.815028	0.0	1.00	1.0	3.00	14.0
surface	330.0	2.621212	2.059636	0.0	1.00	2.0	3.00	15.0
total_chairs	330.0	8.266667	5.798683	0.0	5.00	7.0	10.00	41.0
Runs	326.0	48.214724	46.364077	3.0	19.00	33.0	60.00	341.0
TerrainParks	279.0	2.820789	2.008113	1.0	1.00	2.0	4.00	14.0
LongestRun_mi	325.0	1.433231	1.156171	0.0	0.50	1.0	2.00	6.0
SkiableTerrain_ac	327.0	739.801223	1816.167441	8.0	85.00	200.0	690.00	26819.0
Snow Making_ac	284.0	174.873239	261.336125	2.0	50.00	100.0	200.50	3379.0
daysOpenLastYear	279.0	115.103943	35.063251	3.0	97.00	114.0	135.00	305.0
yearsOpen	329.0	63.656535	109.429928	6.0	50.00	58.0	69.00	2019.0
averageSnowfall	316.0	185.316456	136.356842	18.0	69.00	150.0	300.00	669.0
AdultWeekday	276.0	57.916957	26.140126	15.0	40.00	50.0	71.00	179.0
AdultWeekend	279.0	64.166810	24.554584	17.0	47.00	60.0	77.50	179.0
projectedDaysOpen	283.0	120.053004	31.045963	30.0	100.00	120.0	139.50	305.0
NightSkiing_ac	187.0	100.395722	105.169620	2.0	40.00	72.0	114.00	650.0

The first important finding that we made was that there are two kinds of ticket prices, one for the weekend and one for the weekdays. We then looked for missing values in our data and found that the 'fastEight' column has an alarming 50.3% of its data missing. What was even more relevant to us was that 15-16% of values in 'AdultWeekday' and 'AdultWeekend' prices were missing. We decided to drop the ‘fastEight’ and the ‘AdultWeekday’ columns.

We identified our target column as 'AdultWeekday' which is 'Price'. Then our features were engineered by the following columns: TerrainParks, SkiableTerrain_ac, daysOpenLastYear and NightSkiing_ac and they were added as new columns. We then used a heatmap to look at the correlation between our variables. We found some interesting correlation between features of the resort and ticket prices but we did more digging by putting the numerical values in a scatterplot and checking the correlation between each variable and the ticket price. We found that vertical_drop has a very strong positive correlation with ticket price. fastQuadsm total_chairs, and runs, also seem to be useful in there correlation with ticket price.

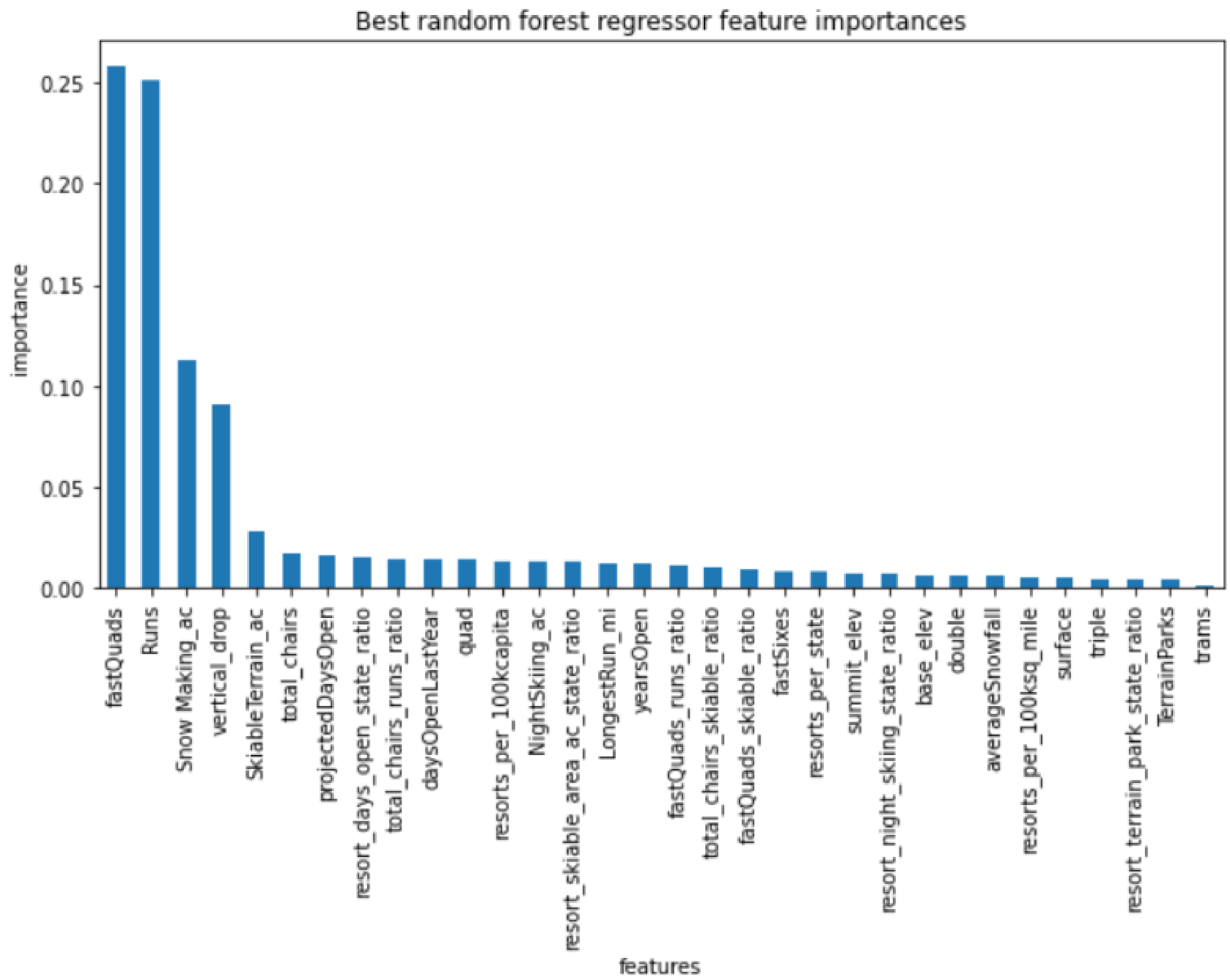


Looking at these results made us realize that we might also want to calculate the following ratios: Total_chairs_runs_ratio, Total_chairs_skibale_ratio, fastQuads_runs_ration, fastQuads_skiable_ration. When putting these columns in a scatter plot we found that the more chairs a resort has relative to the number of runs, the ticket prices plummet. This may be because if there are less chairs then the tickets are more valuable but there will not be a lot of customers. On the other hand, more chairs will make the tickets less valuable and therefore cheaper but you will be able to take in more customers since there are plenty of spots for everyone. We realize that we are missing information that could be very useful: number of visitors per year. Lastly, this scatterplot also showed us that a small number of fast quads is beneficial to the ticket price.



We created pipes for both linear regression and random regressor and found that the random regressor performed better. The best_params was applied and it gave us only 8 features. Then, the parameters were reset and then iterated again. Our final MAE is 9.54 and the mean is 63.81.

Our suggested improvement is that there is room to increase the ticket price by about $5. If you factor that amount with 350K guests who stay 5 days on average, this would add a total of $7.5M to our sum. We came to this conclusion because our modeled price is $95.87 while our actual price is $81. There is also not a huge difference in price decrease when you close up to 6 of the least used runs so closing a couple is acceptable.



Encouragingly, the dominant top four features are in common with your linear model: