Capstone 1

Big Mountain resort case study

User

2021

# Problem identification

Big Mountain Resort is a ski resort located in Montana that has access to 105 trails, serviced by 11 lifts, 2 T-bars, and 1 magic carpet for novice skiers. The longest run is 3.3 miles in length. The base elevation is 4,464 ft, and the summit is 6,817 ft with a vertical drop of 2,353 ft.

In addition, an additional chair lift has been installed recently to help increase the distribution of visitors across the mountain. This additional chair increases their operating costs by $1,540,000 this season.

The pricing strategy has been to charge a premium above the average price of resorts in its market segment. However, this comes with certain limitations, especially since it is not obvious how important some facilities are compared to others and whether it can be further capitalized. Therefore, two important questions arise from the business side:

1. How can Big Mountain resort select a better value for their ticket price?
2. Would any of the planned changes cut costs without undermining the ticket price or support an even higher ticket price?

Current ticket price Big Mountain resort is charging for Adults is $81 for both weekdays and weekends, no difference. After further analysis proved that adult weekday price column was missing values for many resorts, it was decided to focus only on adult weekend ticket price column.

Besides that, the business has shortlisted some options for further cutting costs or increasing revenue:

1. Scenario 1: Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
2. Scenario 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, without additional snow making coverage
3. Scenario 3: Same as number 2, but adding 2 acres of snow making cover
4. Scenario 4: Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

The expected number of visitors over the season is 350,000 and, on average, visitors ski for five days.

# Summary of the project data

For the purpose of this project, there was a CSV file created that contains information from 330 resorts in the US that can be considered part of the same market share, including Big Mountain Resort data as well. The data included information about the features available in the resorts (e.g. night skiing area, number of chairs) as well as tickets price during weekdays and weekends.

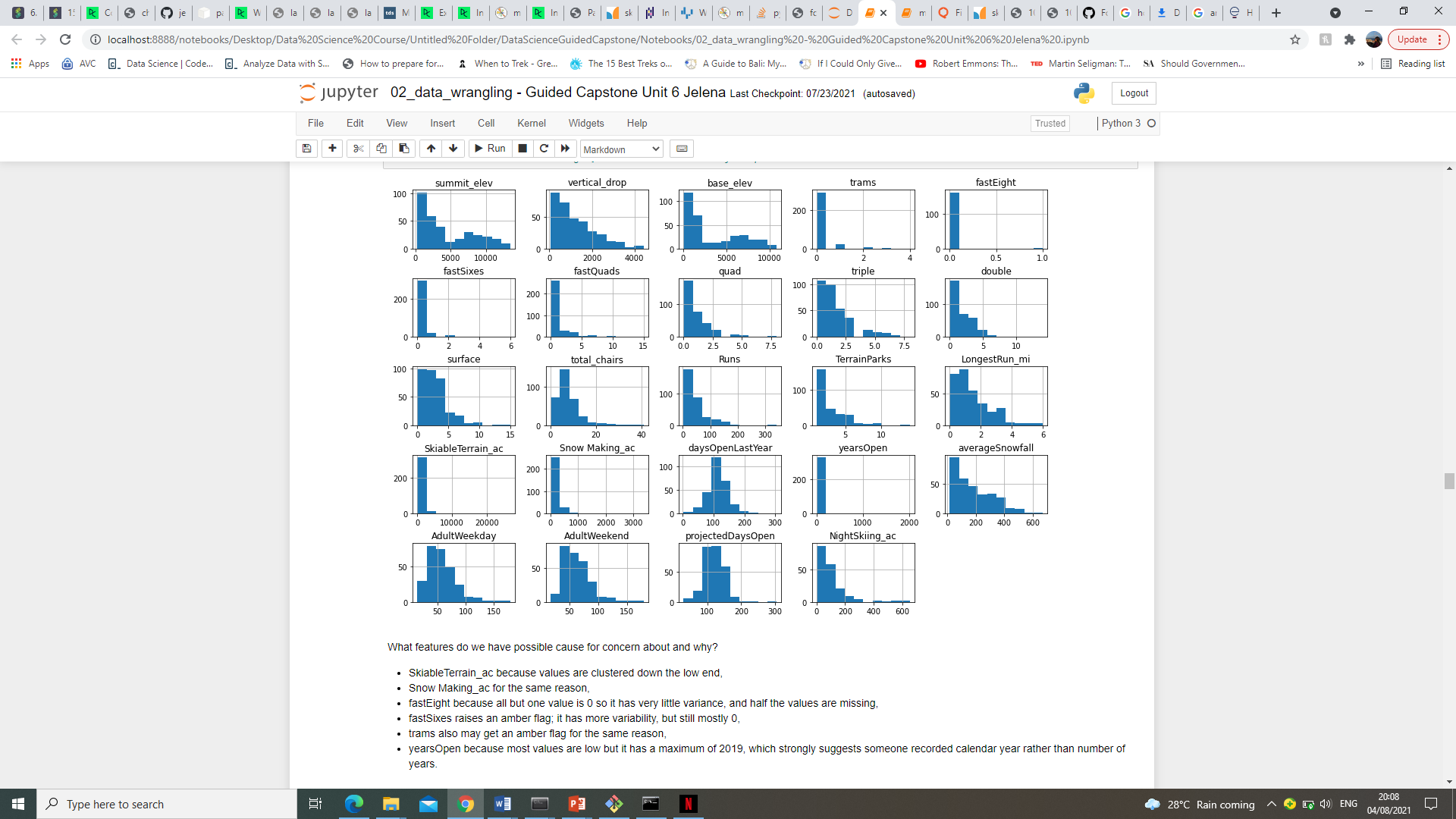
In order to get better understanding of each state’s population and skiable area, we introduced additional dataset obtained from Wikipedia. This table of data is useful because it allows to easily pull and incorporate an external data set. It also allows to proceed with an analysis that includes state sizes and populations that can be very useful for the model and can help better understand density of the resorts in each state as well as total availability of skiable area or for example, area intended for night skiing.

# Data wrangling

There were 330 rows and 27 columns originally, including the data on Big Mountain Resort. Column "fastEight" has been removed as the data wasn't relevant for the analysis (all the values were 0 or missing, except one values which was 1). Also, column "AdultWeekday" was excluded since it has more missing values than column "AdultWeekend", so it is more valuable to observe weekend prices. In terms of rows, all the rows that didn't have a value inserted for "AdultWeekend" have been removed since price is the main value required for our analysis.

There are now 277 rows in the dataset after the cleaning. Some other issues were identified that required checking the data online, e.g. there was wrong value inserted for column "SkiableTerrain\_ac" for one row (row 39) (actual value is 1819, inserted value was 26819). Another issue was relevant for column "yearsOpen" where in one case (row 115) it was inserted 2019, instead of actual count of years open, so this observation is excluded from analysis since it also didn't have price.

During this process, we identified how values for different features are distributed and whether there are some major discrepancies, as you can see per the graph 1.



Graph 1. Distribution of values for all the features included in the dataset.

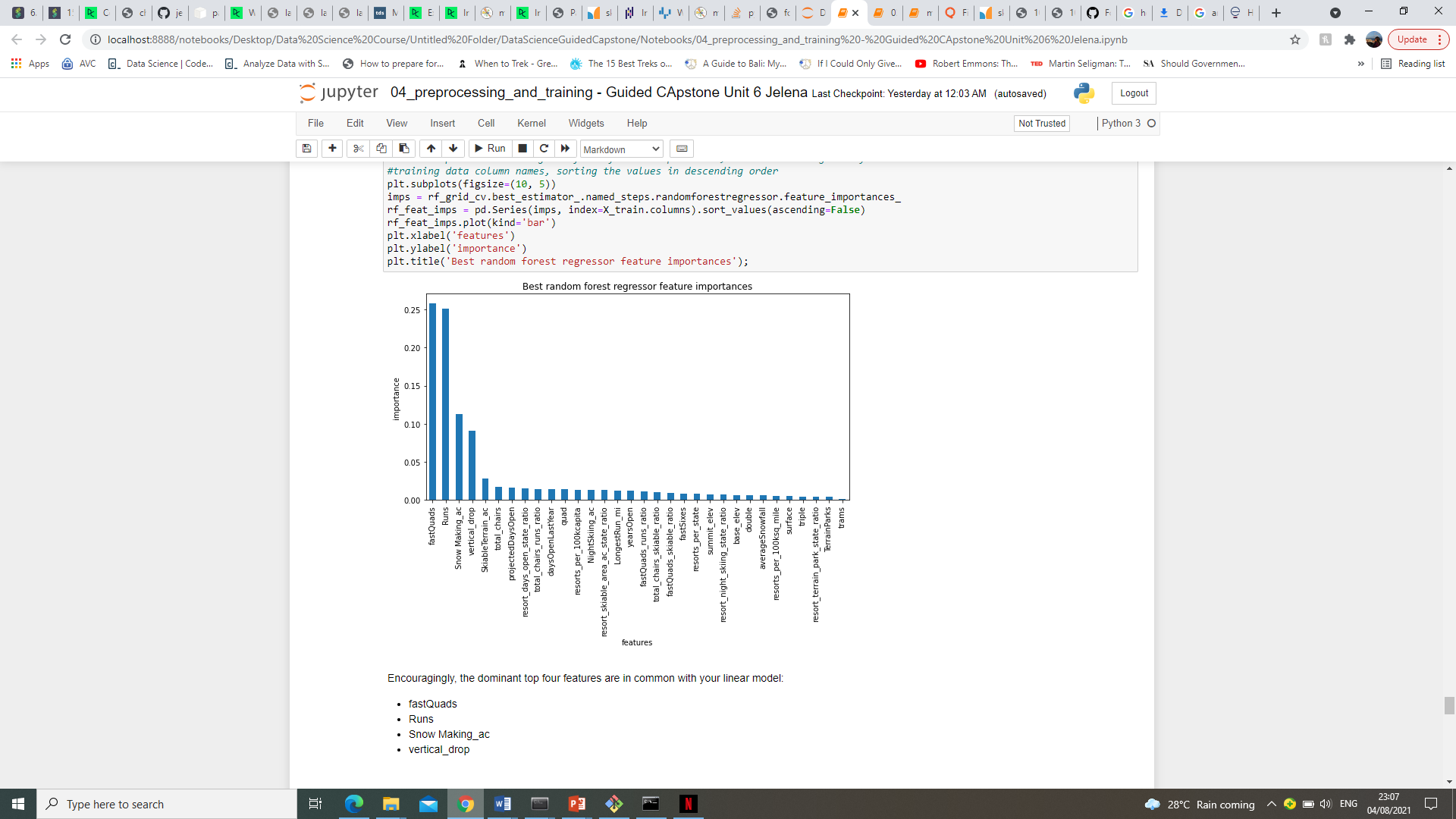
# Data story

While checking the states and ticket price relationship, the scatterplot showed that there is no obvious pattern that can be further investigated. Namely, the red points representing the upper quartile of price can be seen all over the place (left, right and up top). This accounts for some 77% of the variance. Therefore, from the available data, we cannot clearly highlight why some states have higher prices compared to others. This is why we treated all states equally and built a model that includes all states with new features created in this analysis (e.g. resort's night skiing to total state night skiing area or resorts\_per\_100kcapita).

As stated before, our target feature for modelling is AdultWeekend ticket price. While checking the correlation with other features, it was noticed that some have higher correlation with ticket price than others. The most important being fastQuads and Runs, while also one of the most interesting is snow\_making\_ac as visitors seem to value more guaranteed snow.

Further investigation on these features (ratio of total chairs and runs or ratio of fast quads and skiable area) helps to see these relationships a bit better. One important feature missing in the dataset is data on visitors per year, this would help better understand some relationships (e.g. if a resort does not have many chairs, it can charge more for a ticket, but implicitly have less visitors).

The importance of the features to ticket price (derived through random forest regressor) can be observed in graph 2.



Graph 2. Importance of different features for ticket price

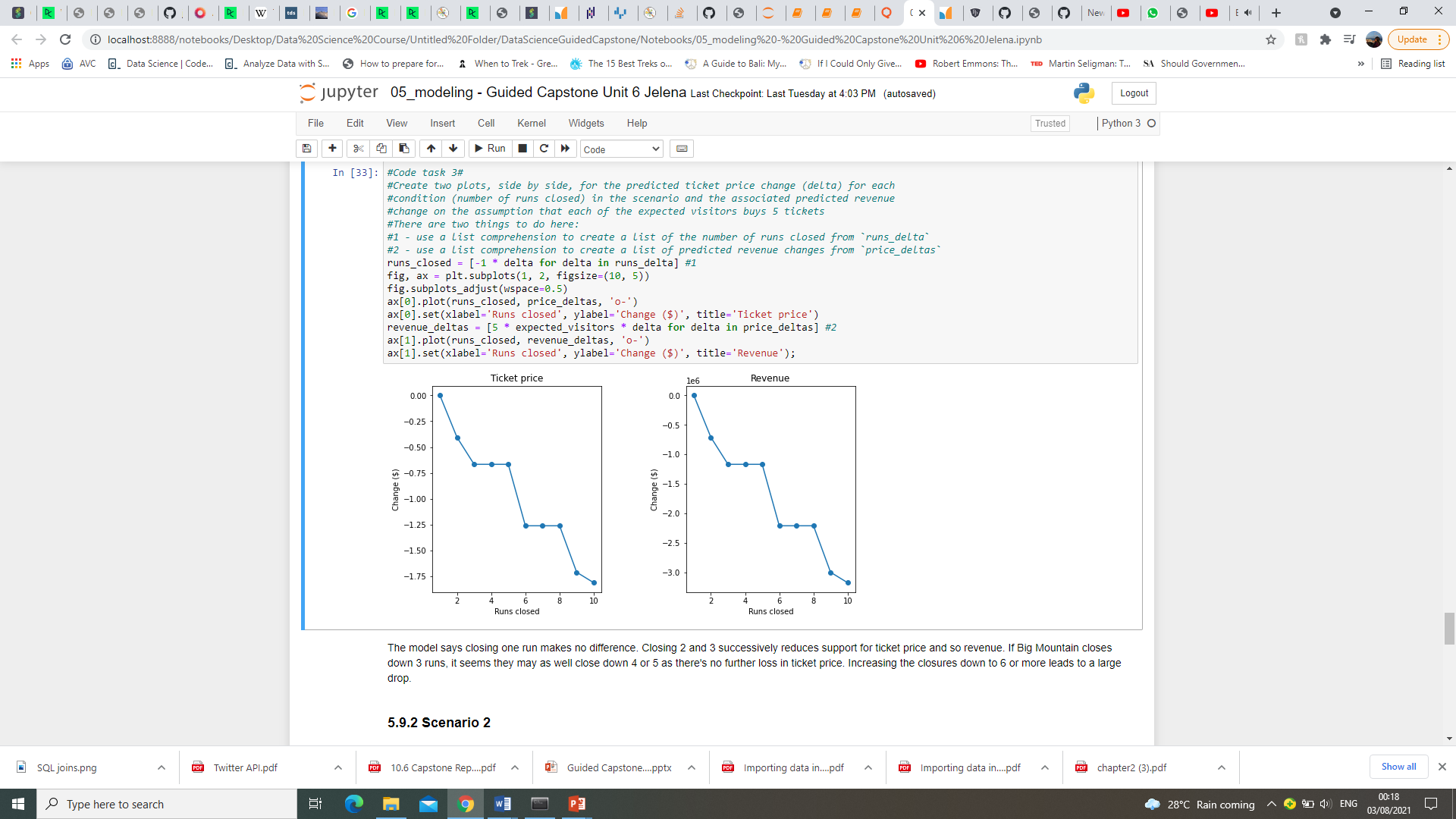
As the very first step after cleaning of the data is done, I have tested whether average price would be good indicator, the strategy Big Mountain resort is currently having. R squared (which shows how one variable is explained by another) for the training set was 0, whereas it was negative for the test set confirming that it is not a good approach to use average price. Also, mean absolute error showed that on average, the ticket price would be off by around 19 dollars if guessed based on an average of known values. Therefore, it is clear that the company should adopt different pricing strategy.

Testing and comparing two different models – linear model and random forest regressor – showed that the mean estimated validity of random forest model is better by almost 10%. Also, it has lower cross-validation mean absolute error by almost 1 dollar (11,8 for linear model vs. 9,54 for random forest regressor). Verifying performance on the test set produces performance consistent with the cross-validation results. This is why, I have decided to proceed with random forest model in predicting the resort’s price.

As said, Big Mountain resort is currently charging 81 dollars**. The outcome of the modelling is that the resort should charge 95.87 dollars.** Even with the expected mean absolute error of $10.39, this suggests there is room for an increase. Big Mountain resort stands very high up on the board for all the most important features for visitors, which also supports the price increase.

Analysis of the possible scenarios shows that Scenario 2 would be the most profitable, whereas if company is trying to cut costs, it can be done by closing one run (closing it makes no difference to the predicted price).

When we perform Scenario 1 (closing 10 runs), the model says that closing one run makes no difference in price, whereas closing 2 and 3 successively reduces support for ticket price. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop (graph 3). If we want to avoid any price drop, we can only consider closing one run, closing any more runs will lead to price decrease.



Graph 3. Impact on the price from closing runs

In Scenario 2, Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift. Analysis shows that price would increase by 1,99, and over the season this could be expected to amount to 3.474.638 dollars.

In Scenario 3, it is the same condition as in the previous one, but adding 2 acres of snow making. The outcome is the same as in Scenario 2 showing that adding to acres of snow making area makes no difference.

In Scenario 4, the longest run is increased by .2 miles and the snow coverage is guaranteed by adding 4 acres of snow making capability. However, this scenario makes no difference to the price at all. In random forest model, this feature actually was not one of the most important features which proves in this scenario as well.

All in all, if any of the scenarios is to be considered, it can be Scenario 2 which proves to add to the price. Also, closing one run will not affect the price as per the model, so if there are high costs associated with the runs, closing one might lead to lower costs without an effect on price. In addition, the price should be adjusted as the model shows, but analysis of the operating costs should also be done before final decision is made.

# Summary and conclusion

Pricing strategy should definitely be adjusted to reflect the learnings from this model as using market average is proven to be wrong.

Resort should consider implementing changes from Scenario 2 as this would lead to significant additional revenue in the long run.

How can the model be further improved? If there were operating costs included in the data (such as the costs for snow making or all the costs associated with maintenance of runs), it would further support the analysis, as well as adding the number of visitors in other resorts or within different states.

Also, the outcome of the modelling clearly shows that our resort is "undercharging", but it is important to highlight that this model assumes that other resorts accurately set their prices according to what the market supports. But if ours is mispricing itself, are others? However, if others correctly set their prices, it could be that our model is simply lacking some key data, such as data on operating costs.