Overview of the problem and context:

Big Mountain Resort is a ski resort located in Montana which offers high-quality facilities for its approximately 350,000 visitors per year. Its base elevation is 4,464 ft, and the summit is 6,817 ft with a vertical drop of 2,353ft. With the recent capital investment made to its facilities, resulting in additional operating costs ($1,540,000), the executives wanted to look into generating more revenue. Therefore, the resort contracted the data science team to evaluate the efficiency of the current pricing strategy as the business' executives suspect it is not reflecting Big Mountain's well-equipped facilities. To solve this, the data science team focused on creating a benchmark with the resort's ticket price data and its most influential features to use an ML model that predicts a price that supports revenue increase based on this data.

Data Wrangling:
In the project's first stage, I applied cleaning and organizing techniques to make the raw data readable and easy to understand. I also highlighted critical variables to help find a suitable predictive model for Big Mountain Resort ticket price. Technical details of this process are.
I will cover it as an overview of the conclusions made rather than the technical structure followed. Feel free to refer to the Jupyter Notebook for technical details [here](#).

I performed a series of steps writing code in Python to tidy up the data. Throughout the process, I discovered that 16% of the resorts were missing at least one value on the ticket price feature. Also, I investigated the relationship between State and Region, imputed values on the terrain acres of a resort to prevent outliers that would disrupt the metrics, and grouped the data by state. I then continued developing a dataset that will raise questions to tackle the current inefficiencies in pricing, if any. [The boxplot_1](#) shows the distribution of tickets price by State, which encouraged me to retain all state information with available prices and drop the others. Finally, I defined the ticket price as the target feature.

After transforming and tidying up the data, I created some visualization. The [histogram_1](#) shows the feature distributions. It shows some skewed distributions to keep an eye on. By the end of this stage, I went web harvesting; I extracted a table with the US Population by state and merged it with the original data to generate a state-wide summary of statistics for Big Mountain's market target. This step will hint at the core solution of the business problem, a benchmark, as the analysis gets deeper into the process. This reference point will significantly help create a customized model to restructure the pricing strategy.

Exploratory Data Analysis
With a cleaner dataset, I explored the data in different sections with a newly state-wide merged table that provided relevant insights into both numerical and categorical variables that are useful to predict a price that reflects the added value of the resort facilities. By performing statistical investigation through formulas and visualizations, I noticed that night skiing, the vertical drop, the number of runs, fast quads, and snow-making acres positively correlate to the ticket price. The later discovery will play a relevant role in the modeling stage as we directly tackle the intention of generating more revenue by assessing the features that positively influence the

price. Therefore, having these variables identified, it will be easier to customize the predictive model for business needs to highlight Big Mountain's position against competitors.

In the heatmap, I visualized the correlation of the features showing a positive correlation with ticket price on the features mentioned above. It is critical to identify what moves the price and direction so that the ML model can do its job correctly and support Big Mountain Resort's decision-making process for its mission to maximize profit. As you will see, the tailored technical process behind the scenes ensures considering every price-influential aspect to exceed the business expectations.

Preprocessing and training:

I will not touch on complex technical functions utilized to build the machine learning model at this stage; you can review it here. Instead, I will summarize the relevant points and the way it all collaborates to create a successful model.

I decided to run two regression models and test them in the train/test split. This concept is a way to hold back some of the data to avoid biased models. I split the resort data into two partitions (70% - 30%). I decided to test the performance of a Linear Regression Model and a Random Forest Model. The first one had a formidable performance when assessed on the test set, though it presented a more considerable variability in its execution. The latter showed consistent performance with the cross-validation results and a lower mean absolute error. In other words, I selected the Random Forest Model because it proved to be a better fit for the objective of the business more accurate: to increase revenue by restructuring the pricing strategy to reflect the resort's high-quality facilities.

Also, I created a barplot_2 of the Random Forest's feature importances, highlighting what the EDA stage did. It is a critical verification of the process's success as it's reassuring that the previous analysis pointed in the right direction.

Modeling & Conclusions:

Having picked a model, I deployed it to the data to predict the best price for a one-type ticket to satisfy the business needs. I also evaluated the given scenarios through the model gaining relevant insights explained below. To review the programming techniques, please refer to this notebook.

Big Mountain Resort currently prices its ticket for $81, which positions it at the high end among Montana's ski resorts. However, the model suggests that the ticket price could be around $5 to $14 higher than the current price - $95.87. The predictive model takes facilities' influence on price around the other resorts in the same market share in the US. Based on the leadership concerns about the current pricing strategy not highlighting the superior Big Mountain's facilities. I'd suggest adjusting the price as the model was specifically designed to predict a price that reflects the resort's competitive advantage - its facilities - that is being overlooked.

Regarding the operational cost increase due to the newly acquired chair lift, increasing the price, as shown in scenarios 2 and 3, might cover the investment based on each visitor buying five tickets. Therefore, the revenue will increase by about $3,474,638. Also, closing runs will impact ticket prices; thus, profit will feel the effect. However, the change between closing 4 to 6 runs is the most significant, going from a 0.75 difference in the ticket price to 1.25. According to the model behavior, there could be improvements to the suggested scenarios from Big Mountain leadership. These improvements might focus on increasing dominant features by more significant amounts.

The data had limitations on total visitor numbers, and other price data, such as ski clothing rentals or purchases, ski accessories, and gear prices, among others, would have been valuable. In addition, the price of rooms for overnight visitors would have also contributed to the predictive model. Furthermore, more profound insight into Big Mountain's cost structure could've helped to customize the model to its particular costs' needs and predict a price that would sustain profit based on those needs. Nevertheless, the ML model went through a tailored-made process to ensure its success by taking the most relevant pricing-drive features and developing a complex series of integrated formulas to return an efficient pricing strategy. As noted in the first paragraph of the conclusion, its predicted price reflects Big Mountain facility capacity among the 330 resorts exhibited in the dataset.
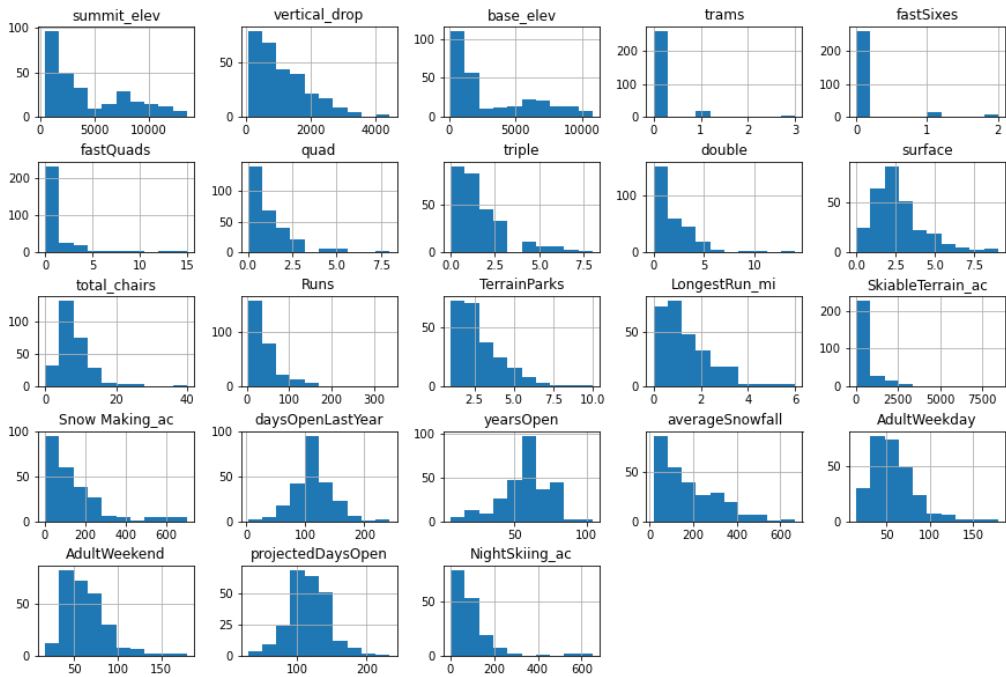
As the resort management suspected, the price does not reflect Big Mountain's top facilities as it is positioned high in the ranking on some of the most influential ticket price features.
The predictive model is designed to run different scenarios with relative ease. It can be made available in a user-friendly platform that takes the input per Scenario and returns results in a non-technical context. It would be helpful to briefly train the business analyst on the model and how it's built but focus on running different scenarios.
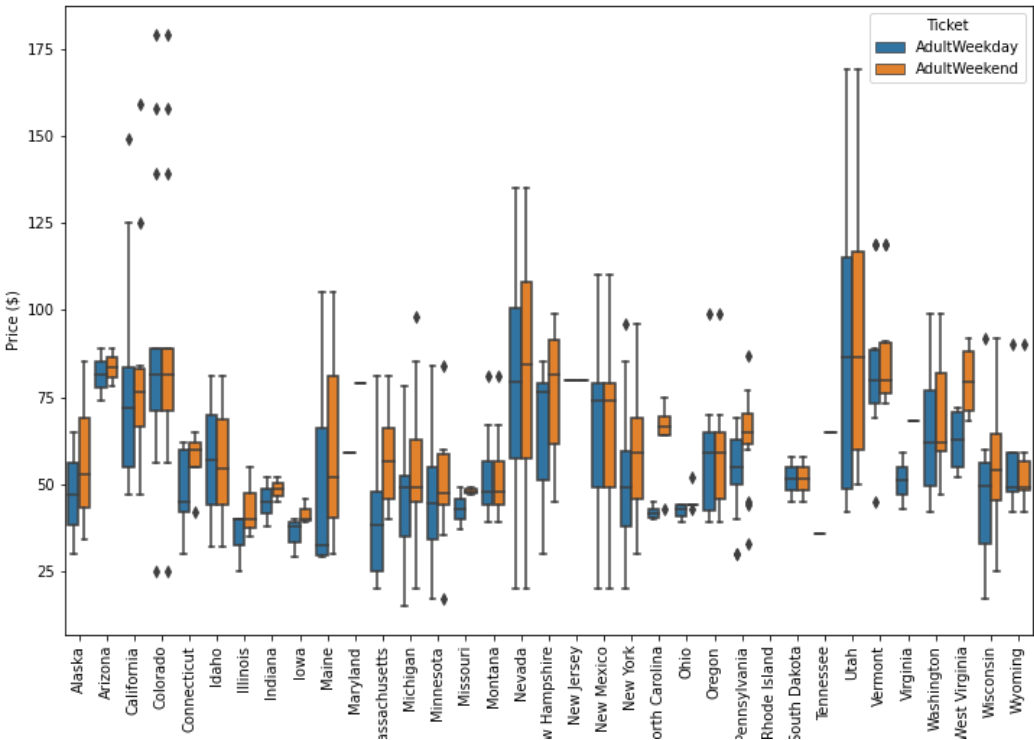
Also, on the last page of this document, you can review where Big Mountain stands for the top pricing-drive features.
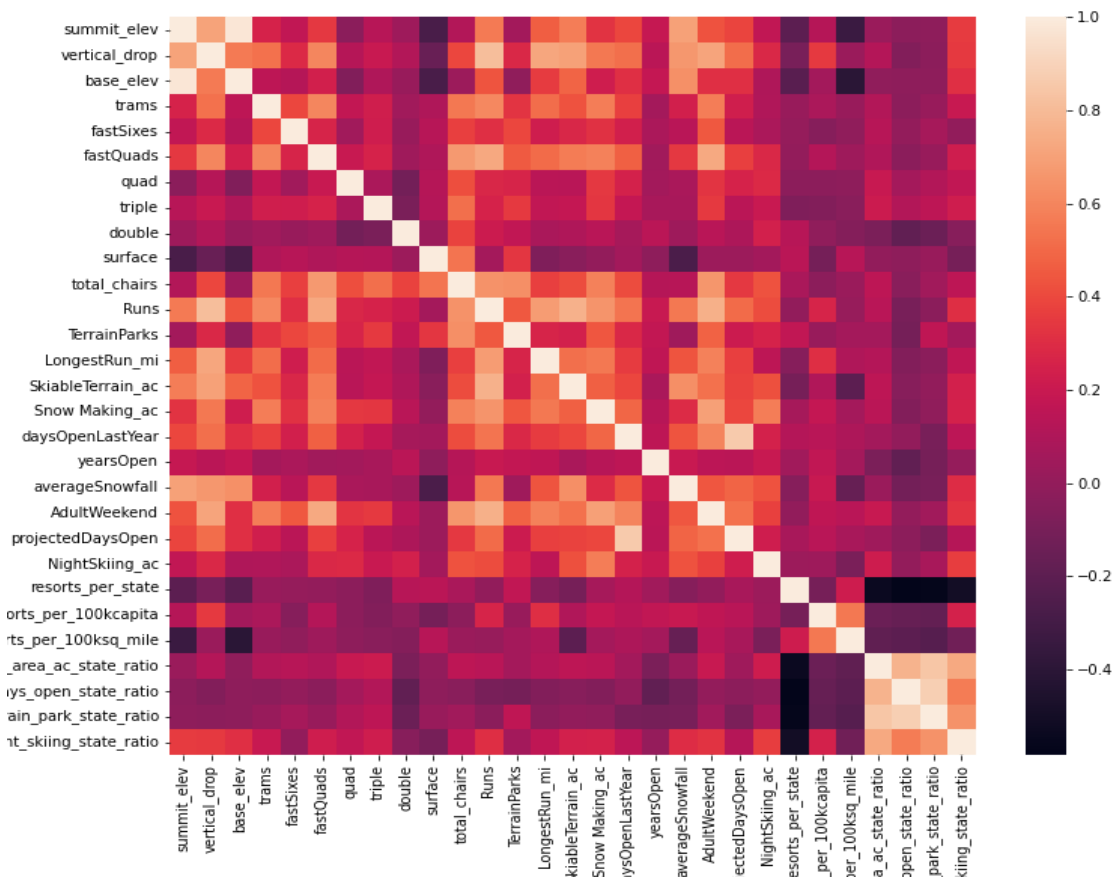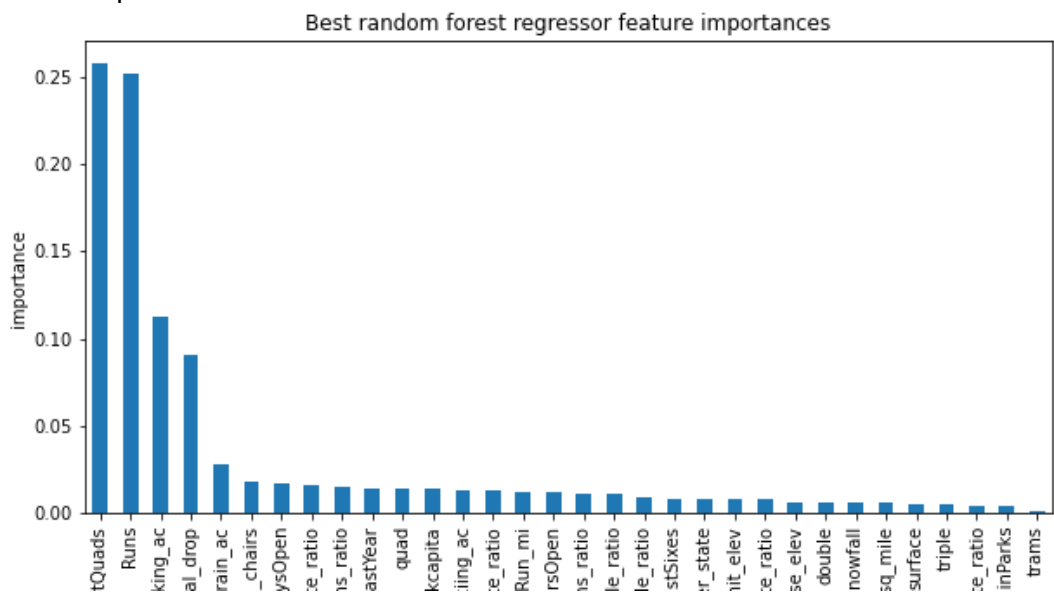
Figures:

Histogram_1

Box_plot1

Heatmap

Barplot

Best random forest regressor feature importances

Big Mountain Position in Top Features:

Adult weekend ticket price ($) - Montana only distribution for resorts in market share

Adult weekend ticket price ($) distribution for resorts in market share

Vertical drop (feet) distribution for resorts in market share

Area covered by snow makers (acres) distribution for resorts in market share

Total number of chairs distribution for resorts in market share

Number of fast quads distribution for resorts in market share

Total number of runs distribution for resorts in market share

Longest run length (miles) distribution for resorts in market share