**Big Mountain Resort**

Our client, Big Mountain Resort, has asked us how Big Mountain Resort can select a better value for their ticket price by Nov 1, 2024, while also considering changes that will cut costs without undermining ticket price by 10% or supporting an even higher ticket price.

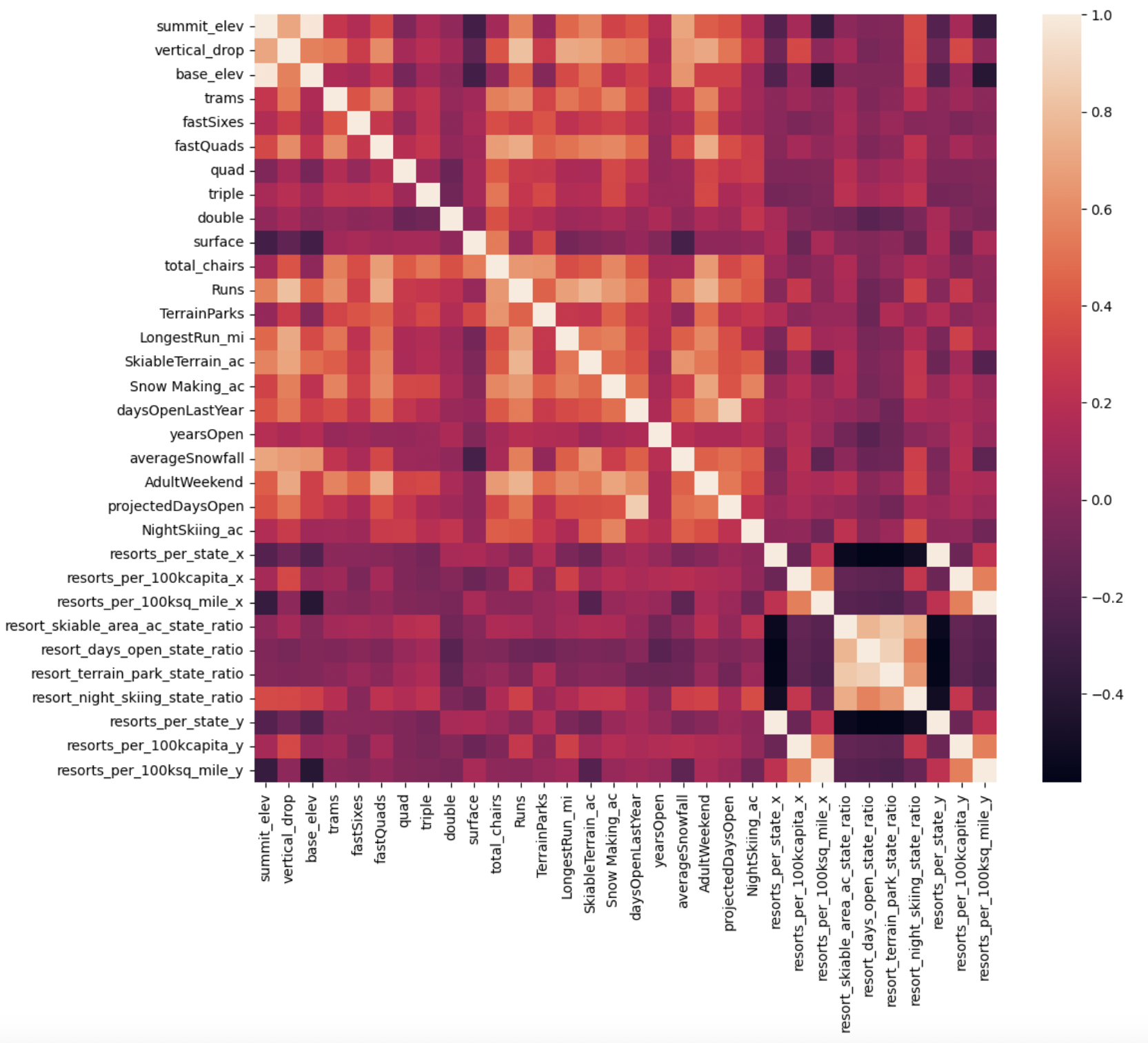
The initial process for Data Wrangling involves setting up the environment by importing the required library and loading the data. The original number of rows in the dataset is 330, and the Big Mountain Resort is identified by exploring the loaded data. Once the data is loaded and verified, we check for the number of missing values per column. One column that was removed is the 'fastEight' column due to half of its values being missing, and the other half being zero, rendering this column essentially devoid of information. The second column that needed removal in the last step was 'AdultWeekday' because 'AdultWeekend' prices have the least missing values. Some data rows were eliminated because the resort lacked any price data. Additionally, there were issues with column names that needed adjustment, addressed by renaming after modifying the content of the columns.

Another issue was the presence of two resorts with the same name, Crystal Mountain. By examining location information, such as the state and region of the resort, it was confirmed that they were not duplicates. The corrective action was to treat them as unique records, handling them as separate resorts rather than removing them as duplicates. I have completed the project and achieved the desired ticket price for the weekend, which is higher than the weekday. Prices appear to be restricted to resorts with prices below $100. The remaining dataset consists of 277 rows.

The primary stage of Exploratory Data Analysis (EDA) involves focusing on parameters such as the total state area, population, skiable area, days open, night skiing area, and resorts per state. Resort density was also analyzed as part of the EDA process. Moving forward in the EDA, the next step was to visualize high-dimensional data. The first task in this phase was to scale the data, followed by verification of the scaling. Subsequently, Principal Components Analysis (PCA) was performed, and the final step involved identifying correlations, with a specific focus on the target feature.

A robust positive correlation was observed between 'vertical\_drop' and ticket price in numerical and categorical features. The target feature, AdultWeekend ticket price, showed reasonable correlations, particularly with 'Runs' and 'Snow Making\_ac.' The most correlated feature with ticket price was 'resort\_night\_skiing\_state\_ratio.' Patterns also emerged, indicating a connection between 'Runs' and 'total\_Chairs' with ticket prices, suggesting that more runs require more chairs for transportation.

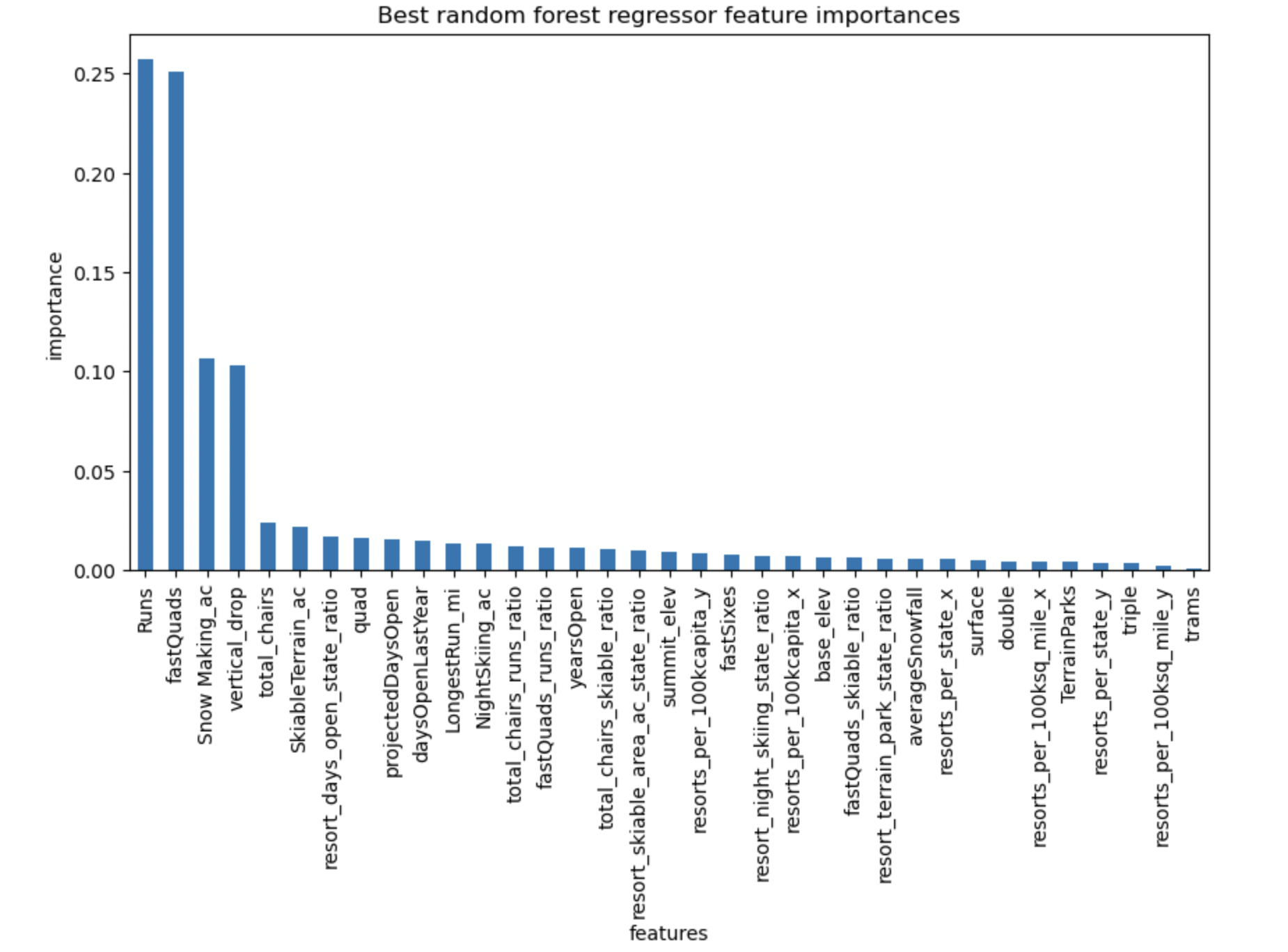
A key observation regarding state and prices is the absence of an obvious pattern. Consequently, the decision was made to treat all states equally and work towards building a pricing model that considers all states together without treating any state as particularly special. However, a cautious approach is required when dealing with the negative sign in the data. For instance, features like 'resorts\_per\_100Kcapita' and 'resorts\_per\_100ksq\_mile' exhibit a large negative coefficient, which results in a large positive PCA score when multiplied by a large negative feature.



Since our focus is on Big Mountain Resort, it is crucial to extract data specific to this resort for Pre-processing and training purposes. The first involves considering the mean value as a predictor and establishing a baseline model for subsequent evaluations. To initiate preprocessing, we conduct a Train/Test Split, dividing the data into training and testing sets to independently assess model performance.

Moving forward, we delve into Metrics to evaluate the agreement between different sets of values. One such metric is the coefficient of determination (R-squared), indicating the proportion of variance in the dependent variable (ticket price) predicted by our model. An R-squared of 1 signifies a perfect prediction while using the average value results in an R-squared of 0 on the training set. Generally, performance on the test set is expected to be slightly worse than on the training set.

Transitioning to the Random Forest Model, a robust choice in many cases, we emphasize the need to assess performance through cross-validation rather than repeatedly checking on the test split. The top four dominant features identified by the Random Forest Model are fastQuads, Runs, Snow Making\_ac, and vertical\_drop. A simpler model with fewer features can be advantageous for stakeholder understanding. Comparing the Linear Regression and Random Forest models, we calculate the mean absolute error using cross-validation. The Random Forest Model outperforms the Linear Regression Model, exhibiting lower cross-validation mean absolute error and reduced variability.



Ultimately, performance verification on the test set aligns with cross-validation results, leading us to select the Random Forest Model as the optimal choice.

The subsequent step involves refitting the model using all available data but excluding Big Mountain's data. The objective is to train a model predicting Big Mountain's ticket price based solely on data from other resorts. This approach aims to prevent Big Mountain's current price from biasing the prediction, ensuring the calculation relies solely on its competitors.

Moving forward, we calculate the expected ticket price for Big Mountain Resort based on the model. The modeled price is $97.96, while the actual price stands at $81.00. Despite an expected mean absolute error of $10.36, this suggests there is potential for an increase. The model's validity hinges on the assumption that other resorts accurately set their prices based on market demand. The substantial difference between our resort's actual price and the predicted price implies that Big Mountain Resort might be undercharging. Recognizing that some resorts may be "overpriced" and others "underpriced," we seek a clearer context by evaluating Big Mountain Resort in the market.

Examining contextual information such as vertical drop, Snow Making\_ac, total chairs, fast quads, number of runs, longest run, and Trams provides insights. Big Mountain Resort stands out favorably in terms of vertical drop, snowmaking area, total chairs, and longest runs. The model suggests that closing one run has no impact, but closing 2 and 3 consecutively reduces support for ticket price and revenue. If Big Mountain were to close down 3 runs, it might be as effective as closing down 4 or 5, with no further loss in ticket price. However, increasing closures to 6 or more results in a significant drop.

Conversely, adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift would raise the ticket price by $2.22. Over the season, this price increase would amount to $ 3,888,889. Interestingly, adding 2 acres of snowmaking shows no discernible impact. Considering these factors, it seems reasonable to suggest increasing the ticket price to $97.96, given the current price of $81, indicating room for an increase.

Noteworthy is the fact that Big Mountain has the highest number of total chairs, and resorts with more chairs appear to be outliers. Hence, it's plausible to anticipate that increasing ticket prices would result in a corresponding increase in value. The model indicates a significantly higher predicted price than the current one, suggesting that, when compared to other resorts, Big Mountain has a lot to offer but has priced its services comparatively lower. This incongruity may come as a surprise to business executives who might have anticipated a price increase to position the resort as the most expensive in the market. Despite this, the business can continue utilizing the model with the assistance of data scientists to adapt to changes in data and shifts in the market.

In communicating these findings to business analysts, I would emphasize how they can effectively leverage the model to gain valuable insights. Demonstrating ways to utilize the model for strategic decision-making and highlighting its potential impact on business outcomes could provide them with valuable tools for informed analysis.