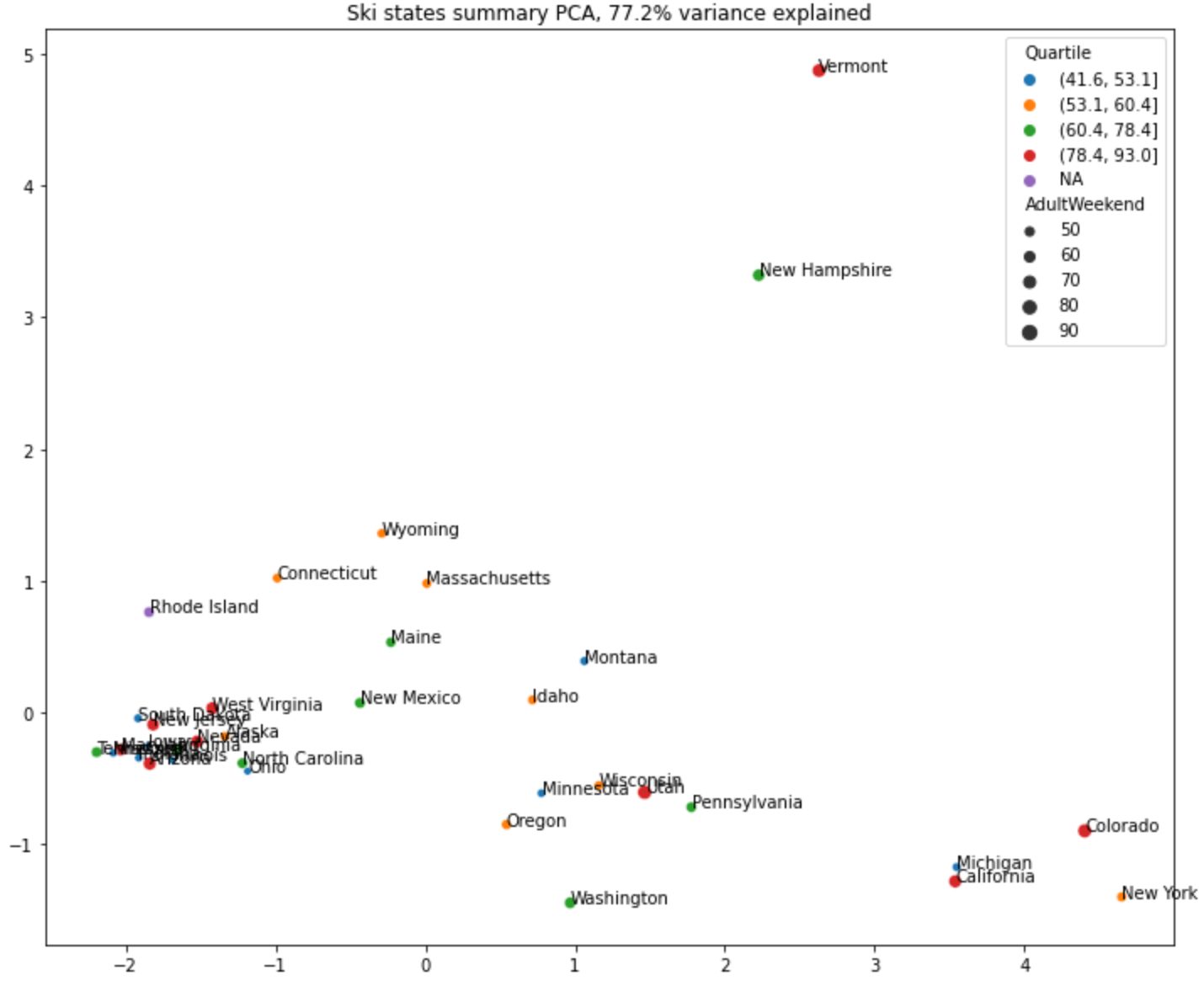
**Analysis of the Ticket Pricing Strategy of Ski Resorts in The United States**

This report summarizes the data, methodology and results of modeling the ticket pricing strategy of ski resorts in The United States. This analysis was put together by request of Big Mountain Resort management, and in addition to constructing the model, we have also analyzed how a few different scenarios would impact Big Mountain Resort’s ticket price. Our results indicate that Big Mountain Resort is currently undercharging for their tickets based on the level of service and quality of facilities that they provide. The most important factors that we determined to be positively related to ticket price were the number of fast quads, total ski runs, total snowmaking coverage, and the vertical drop of the ski resort’s runs. More details about these factors can be found in the results section of this report.

**Data Cleaning and Exploration**

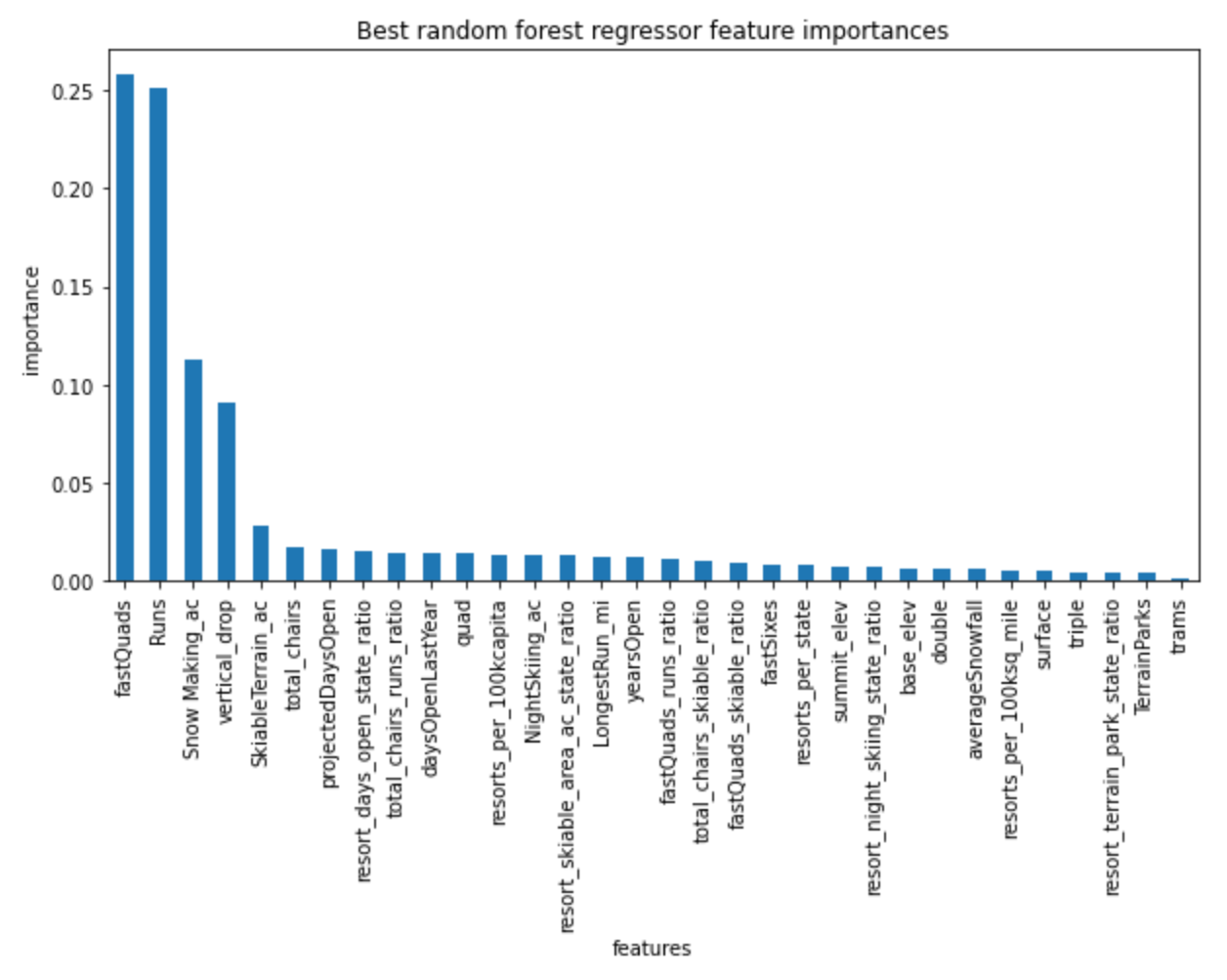
The original dataset was provided by Alesha Eisen, the database manager, and included 330 rows representing individual ski resorts, along with 27 columns, representing individual data points for those ski resorts. The dataset included both weekday and weekend ticket price columns, but only weekend ticket price was used in the model due to having fewer missing values. After making other adjustments to remove missing and inconsistent values, the dataset contained 25 columns of data for 277 ski resorts.

While exploring the relationships within the data, we determined that the state a resort was located in did not have a clear impact on it’s ticket price. For this reason, the resort state was not included in the modelling process. We did, however, add additional columns that were rough measures of how much a resort owned the total market share in the state, to try to see if in-state competition played a role in determining ticket price. Additionally, we added columns that were measures of the ratio between the number of ski lifts and the number of total runs/skiable terrain, to try to get a sense of how fast a customer could navigate the resort. The final dataset that was used for modelling included 36 columns and 277 rows. 

Using a technique known as principal component analysis, the scatterplot shown above is a good indicator that there is not a clear relationship between state and price. States with low, medium and high average ticket prices are seen scattered across the entire plot.

**Modeling Methodology**

Two primary models were constructed and tested during our analysis. The first model was a linear regression model, and the second was a random forest model. Both models used five-fold cross-validation, which is a strategy that seeks to avoid overfitting to the dataset, and helps the model generalize better to future data. The linear regression model had a mean absolute error of 10.5 across the cross-validation dataset, while the random forest model had a mean absolute error of 9.6. This means that on average the predicted ticket price output by the random forest model was almost a whole dollar closer to the actual ticket price than the linear regression’s prediction. While the linear regression model was simpler, the considerable performance advantage of the random forest model made it the best choice for further exploration. In the chart below you can see that the random forest places a large amount of emphasis on four key variables: the number of fast quads, the number of total runs, the amount of snowmaking coverage, and the vertical drop.



**Results**

Based on the random forest model, Big Mountain Resort should be charging $95.87 for weekend tickets, compared to the current price of $81.00. This number is explained by Big Mountain Resort’s high number of fast quads, total runs, snowmaking area, and vertical drop, which the model identified as good indicators of high ticket price. In addition to modeling the current recommended ticket price, we also modeled the following scenarios and the resulting ticket price increase or decrease that could be justified based on each scenario:

1. **Close up to 10 runs**
   1. Closing runs would result in an expected decrease in ticket price from $0.00 (0 runs) to $1.75 (10 runs)
2. **Add a run, add 150 feet to vertical drop, add a chair lift**
   1. This scenario could justify a ticket price increase of $1.99
   2. This would result in around $3.5M in increased revenue annually, based on 350,000 visitors who each buy 5 tickets.
3. **Repeat Scenario 2, but also add 2 additional acres of snowmaking**
   1. The additional 2 acres of snowmaking does not justify any further increase in price
4. **Increase the longest run by .2 miles and add 4 acres of snowmaking**
   1. According to the model, this does not justify any increase in ticket price
5. **Add an additional chair lift**
   1. The current proposal of adding an additional chair lift with an annual operating cost of $1.5M is not cost effective as it would only justify a ticket price increase of around $0.30, or $500K annually.