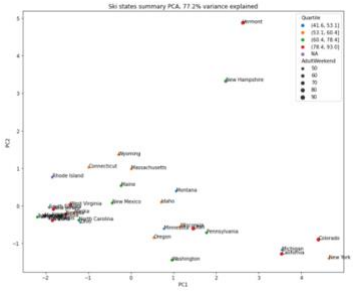
Guided Capstone Project Report

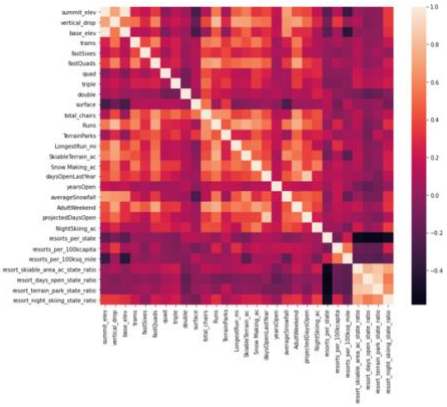
Big Mountain Resort in Montana is a recreational winter resort containing access to 105 trails and serves about 350,000 skiers and snowboarders each winter season. The resort has recently installed an additional chair lift to increase distribution of visitors, at an increase to operational costs of $1,540,000. The resort's pricing strategy has been to charge a premium above the average price of resorts in its market segment, however, there is a concern that marking the pricing strategy to the average is not allowing the resort to capitalize on its facilities as much as it could. Thus, the data evaluation of 330 market-participant resorts and subsequent modeling was undertaken to determine a statistically substantiated and advantageous ticket price change as well as the associated revenue change.

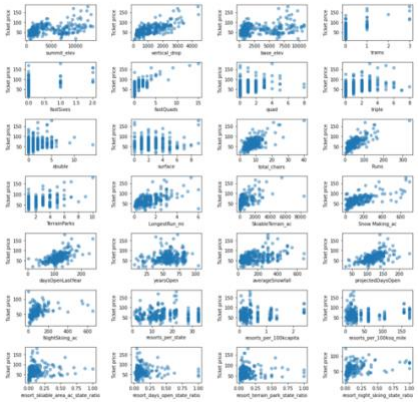
The data available for this analysis contains information from 330 resorts from around the United States in the same market segment as Big Mountain Resort. The original dataset contains information about mountain characteristics, resort characteristics, and pricing information. This dataset was transformed during the data wrangling phase of the project to include the highest fidelity data, according to number of missing values, resulting in approximately 277 rows of data (Fig. 1).

```
Int64Index: 277 entries, 0 to 329
Data columns (total 25 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Name               277 non-null    object
 1   Region             277 non-null    object
 2   state              277 non-null    object
 3   summit_elev        277 non-null    int64
 4   vertical_drop      277 non-null    int64
 5   base_elev          277 non-null    int64
 6   trams              277 non-null    int64
 7   fastSixes          277 non-null    int64
 8   fastQuads          277 non-null    int64
 9   quad               277 non-null    int64
 10  triple             277 non-null    int64
 11  double             277 non-null    int64
 12  surface            277 non-null    int64
 13  total_chairs       277 non-null    int64
 14  Runs               274 non-null    float64
 15  TerrainParks       233 non-null    float64
 16  LongestRun_mi      272 non-null    float64
 17  SkiableTerrain_ac  275 non-null    float64
 18  Snow Making_ac     240 non-null    float64
 19  daysOpenLastYear   233 non-null    float64
 20  yearsOpen          277 non-null    float64
 21  averageSnowfall    268 non-null    float64
 22  AdultWeekend       277 non-null    float64
 23  projectedDaysOpen  236 non-null    float64
 24  NightSkiing_ac     163 non-null    float64
dtypes: float64(11), int64(11), object(3)
```

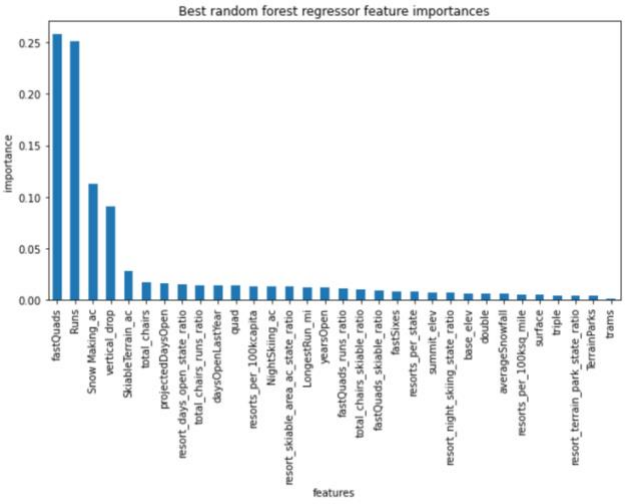


During the Exploratory Data Analysis phase of the analysis, we aggregated the data by state to observe state and regional trends, ultimately deciding whether to treat states/regions independently or not. From the principal component analysis (PCA), the first two components(resorts_per_state and state_total_skiable_area_ac) accounted for approximately 77% of the variance, and resulted in no clear price quartile trends by state (Fig. 2). Thus, we have elected to treat all states equally in the dataset and subsequent price modeling. Additionally, a feature correlation heatmap (Fig. 3) and 'ticket price vs variable' scatter plots (Fig. 4) were created to identify possible variable correlations. The variables with correlation to ticket price, include: fastQuads, vertical_drop, total_chairs, Runs, and Snow Making_ac.





We started the pre-processing and training effort by parsing the data into train and test data sets and built linear models, imputing missing values using both the median and mean to fill non-existant numbers. Since we suspected the model was over-fitting, we decided to refine the model using 'SelectKBest', cross-validation, and optimizing hyperparameters of the estimator with GridSearchCV. We then tried a random forest regressor; it was clear that 69 trees, 'median' imputation strategy, and no scaling features were the best parameters. The best random forest regressor feature importances were 'fastQuads' and 'Runs', followed by 'Snow Making_ac' and 'vertical_drop'. The RF regression model has a lower CV MAE by almost $1, exhibits less variablity, and is verified by the performance on the test set.



The current ticket price at Big Mountain resort is $81. Since we are considering a ticket price change, we decided to refit the model on all of the available data excluding Big Mountain resort, effectively removing the bias of its current ticket price in the model. After running the model, the Big Mountain Resort price result was $95.87, which is significantly different from the current price of $81. The validity of this price is based on the assumption that other resorts are accurately pricing their tickets to what the market supports, and Big Mountain Resort sits high on many of the league charts.

From the scenario modeling effort, we have identified two scenarios that resulted in influencing the support for ticket prices. Closing a number of chair lifts had a negative impact on the ticket depending on the number of lifts that were closed but could provide cost-savings in operational costs. Additionally, the scenario adding a run, increasing the vertical drop, and adding a chair lift showed the combination of these parameters showed support for a $1.99 increase in ticket price which would result in an additional $3,474,638 in revenue. This additional revenue would account for nearly double the cost of the operating costs of the additional chair lift described in the proposition from the business. That being said, there are a number of additional cost factors associated with adding an additional run or extending the vertical drop. These would likely be one-time costs, as opposed to the recurring operating costs of the new lift but a consideration, nonetheless. Thus, increasing the ticket price $1.99 will cover the costs of the new lift with increased revenue and meet the objectives of the study.