**REPORT SUMMARY:**  **BIG MOUNTAIN SKI RESORT TICKET PRICE RECOMMENDATION**

**Le-Thuy Tran**

1. **Problem Statement**

Big Mountain Resort, a ski resort located in Montana. Big Mountain Resort offers spectacular views of Glacier National Park and Flathead National Forest, with access to 105 trails. Every year about 350,000 people ski or snowboard at Big Mountain. This mountain can serve skiers and riders of all levels and abilities with 11 lifts, 2 T-bars, and 1 magic carpet for novice skiers. Hellfire, 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. Big Mountain Resort has recently installed an additional chair lift to help increase the distribution of visitors across the mountain with an operating cost increase of $1,540,000 this season. The dataset contains information from 330 resorts in the US is available to answer the following questions from the Big Mountain Ski Resort’s management teams:

* Has Big Mountain Ski Resort maximized its returns, relative to its position in the market?
* What facilities or properties matter most to the visitors or may be used to boost the ticket price?
* What model may be used to provide guidance for Big Mountain’s pricing and future facility investment plans?
* What is the increased ticket price that Big Mountain supports with the additional installed chair lift ?

1. **Data**

A single CSV file from the database manager that contains information from 330 resorts in the US including Big Mountain Resort that can be considered part of the same market share. Each row of the CSV file contains the information of a distinct resort: name, region, state, elevation of the summit mountain (summit\_elev), vertical change in elevation from the summit to the base (vertical\_drop), elevation at base (base\_elev), number of trams (trams), number of fast eight person chairs (fastEight), number of fast six person chairs (fastSixes), number of fast four person chairs (fastQuads), regular speed four person chairlifts (quad), regular speed three person chairlifts (triple), regular speed two person chairlifts (double), regular speed single person chairlifts (single), sum of all chairlifts (total\_chairs), number of runs (Runs), number of terrain parks (TerrainParks), length of longest run (LongestRun\_mi), total skiable area (SkiableTerrain\_ac), total area covered by snow making machines (Snow Making\_ac), total number of days open last year (daysOpenLastYear), number of years the resort has been open (yearsOpen), average annual snowfall (averageSnowfall), cost of an adult weekday chairlift ticket (AdultWeekday), cost of an adult weekend chairlift ticket (AdultWeekend), projected days open in the upcoming season (projectedDaysOpen), total skiable area covered in lights for night skiing (NightSkiing\_ac). There are missing data in the following columns: fastEight, NightSkiing\_ac, AdultWeekday, AdultWeekend, daysOpenLastYear, TerrainParks, projectedDaysOpen, Snow Making\_ac, averageSnowfall, LongestRun\_mi, Runs, SkiableTerrain\_ac, yearsOpen. All the columns are numerical data except for the columns name, state, and region.

We have the following possible cause for concern after reviewing the distributions of these feature values:

* SkiableTerrain\_ac because values are clustered down the low end,
* Snow Making\_ac for the same reason,
* fastEight because all but one value is 0 so it has very little variance, and half the values are missing,
* fastSixes raises an amber flag; it has more variability, but still mostly 0,
* trams also may get an amber flag for the same reason,
* yearsOpen because most values are low but it has a maximum of 2019, which strongly suggests someone recorded the calendar year rather than the number of years.

We have taken the following actions to clean the data:

* Replace the value of SkiableTerrain\_ac of 1819.00 for the row with high value of this column 26819.00
* Drop the rows with a high value of snowMaking\_ac.
* Drop the fastEight column entirely because half the values are missing and all but the others are the value zero.
* Drop a row with the wrong value of yearsOpen
* Remove rows from data where both AdultWeekend and AdultWeekday are missing.

The final dataset has 277 rows and 25 columns after dropping rows, dropping and adding columns.

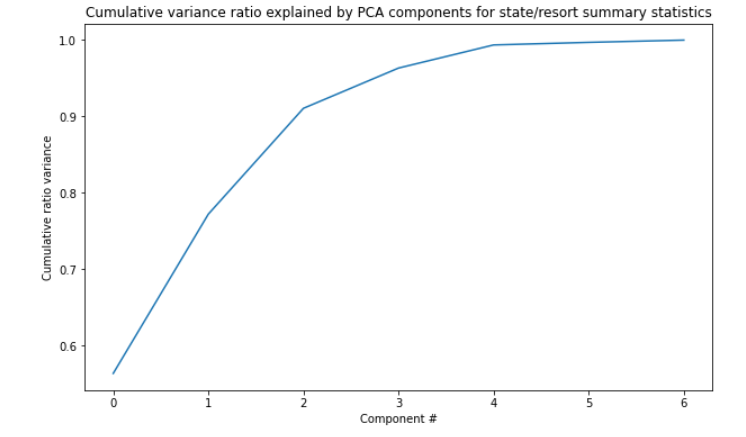
1. **EDA and Statistical Analysis**

The following remarks resulted from our exploratory data analysis:

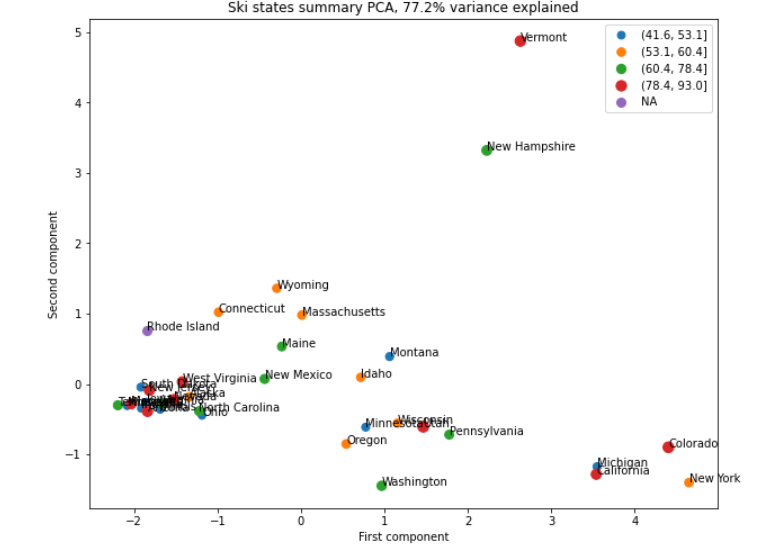
* Montana (the home state of Big Mountain Ski Resort) comes in at third largest area but does not figure in the most populous states.
* There are states that host many resorts, but other states host a larger total skiing area. The states with the most total days skiing per season are not necessarily those with the most resorts.
* New York State boasts an especially large night skiing area. New York had the most resorts but wasn't in the top five largest states, so the reason for it having the most resorts can't be simply having lots of space for them.

The plot of cumulative variance ratio explained by PCS components for state/resort summary statistics in **Figure 1**. As seen in **Figure 1**, the first two components seem to account for 75% of the variance, and the first four for over 95%.

In **Figure 2**, we plot the first two derived features (the first two principal components) and label each point with the name of the state. In this figure, the price ticket, AdultWeekend, is specified by the pointsize and the quartile is specified by hue. There isn't an obvious pattern. The red points representing the upper quartile of price can be seen to the left, the right, and up top. There's also a spread of the other quartiles as well. In this representation of the ski summaries for each state, which accounts for some 77% of the variance, we simply do not see a pattern with price. So all states should be treated equally when building the pricing model.

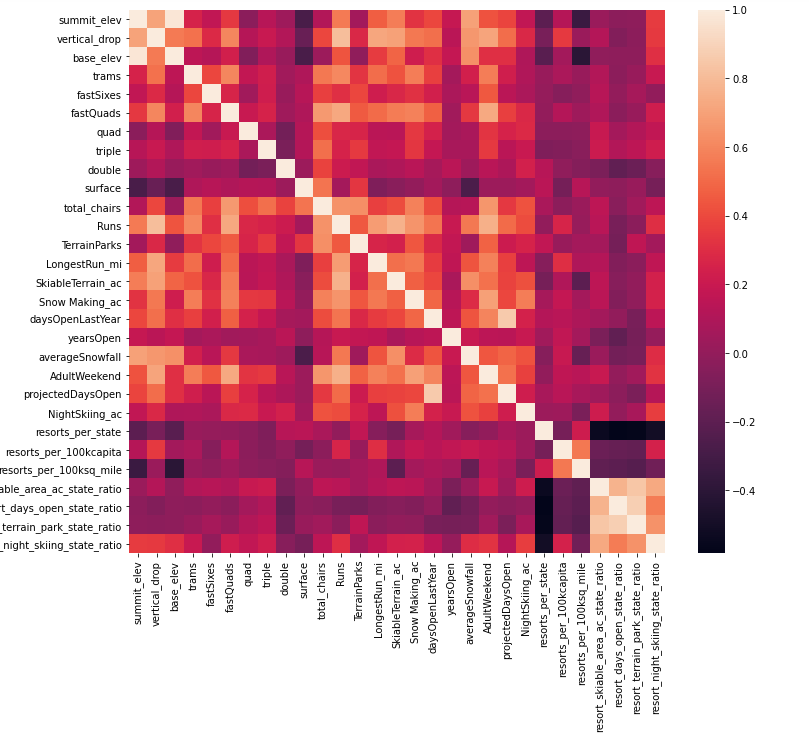


**Figure 1**. Cumulative variance ratio explained by PCS components for state/resort summary statistics



**Figure 2.** The plot of the first two derived features (the first two principal components) are labeled with the names of the states. In this figure, the price ticket, AdultWeekend, is specified by the pointsize and the quartile is specified by hue

In order to gain a level view of relationships amongst the features, we plot in the heatmap of correlations in ski data in **Figure 3.**



**Figure 3**. Feature correlation heatmap

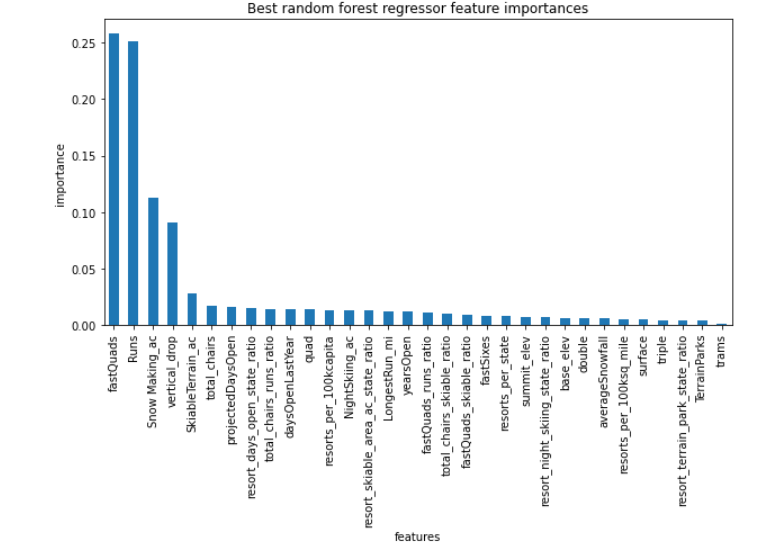
As shown in **Figure 3**, fastQuads, Runs, total chairs, vertical\_drop, LongestRun\_mi, trams, SkiableTerrain\_ac, and SnowMaking\_ac seem quite well correlated with the ticket price AdultWeekend.

1. **Machine Learning**

The following steps have been taken to perform machine learning on the ski data:

* Imputes missing values with strategy of ‘median’
* Scales the data with standard scaler
* Select the k best features
* Train and test model using either Linear Regression or Random Forest model
* Use 5-fold cross-validation for estimating model performance

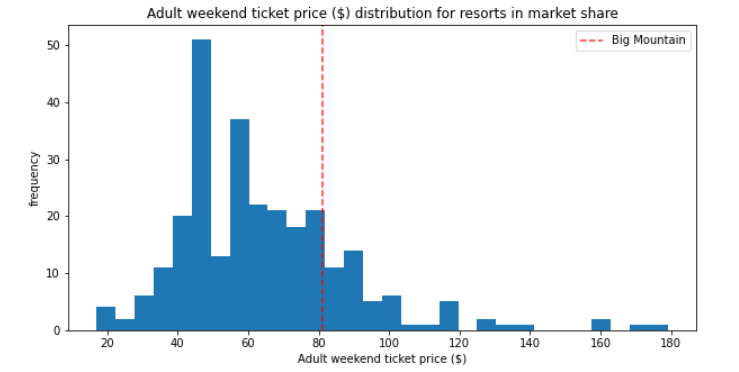
The assessment of the result model has over 60% of the variance and cross-validation shows that assessing model performance is open to variability. The Random Forest model performs better, has a lower cross-validation mean absolute error, and exhibits less variability than a Linear Regression Model. For the Random Forest Regressor, we plot the features importances as shown in **Figure 4**. Features that came up as important in the modeling included: vertical\_drop, Snow Making\_ac, total\_chairs, fastQuads, Runs. LongestRun\_mi, trams, and SkiableTerrain\_ac,

Figure 

**Figure 4.** The best random forest regressor features importances.

Current ticket price at Big Mountain Ski resort is $81. The AdultWeekend ticket price amongst resorts is shown in **Figure 5**.

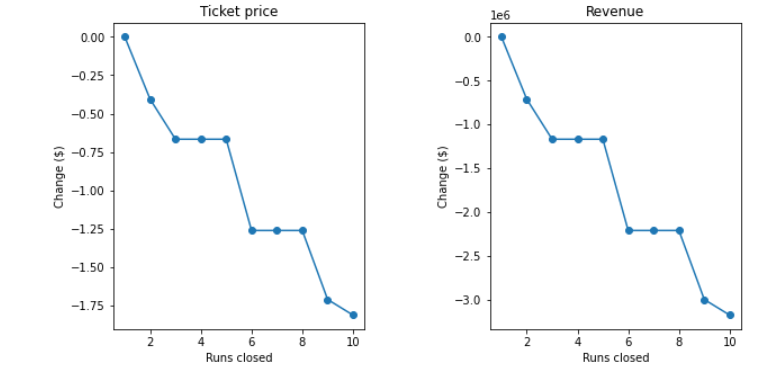
Using the Random Forest Regressor model to predict the ticket price for Big Mountain Ski resort, the obtained ticket price is $95.87.



**Figure 5**. The AdultWeekend ticket price distribution for resorts in market share

If Big Mountain ski resort is going to install an additional chair lift in addition to a run and to increase the vertical drop by 150ft, it supports the increase of ticket price by $1.99. Over the season, the revenue can amount to $3474638 to cover the additional operating cost of the new chair lift if the average number of visitors this season is 350000.

We plot the change in supported ticket price versus the number of runs closed as in **Figure 6**. As shown in **Figure 6**, closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. 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.



**Figure 6**. The change in ticket price and revenue versus the number of runs closed.

The model also suggests that Big Mountain can close down up to 5 runs with very little change in the price support

1. **Conclusions and Recommendations**

Big Mountain ski resort currently charges $81 per adult ticket. A Random Forest Regressor model has been built and trained with data from all other resorts to predict the price for Big Mountain ski resort. The features that are most important to the model include fastQuads, Runs, Snow Making\_ac, vertical\_drop, and SkiableTerrain\_ac. The predicted ticket price for Big Mountain using our developed model is $95.87. The result suggests that Big Mountain resort might be undercharging. In comparision with other resorts, the facilities of Big Mountain ski has high numbers of fastQuads, Runs, SnowMaking\_ac, vertical\_drop, and SkiableTerrain\_ac. This is the reason to justify why Big Mountain ski resort can support such a high ticket price. If Big Mountain ski resort is going to install an additional chair lift in addition to a run and increase of vertical drop by 150ft, it supports the increase of ticket price by $1.99. Over the season, the revenue can amount to $3474638 to cover the additional operating cost of the new chair lift if the average number of visitors this season is 350000. The model also suggests that Big Mountain can close down up to 5 runs with very little change in the price support