**Problem Statement:**

Big Mountain Resort, located in Montana, is exploring opportunities to create a new competitive ticket price to maximize revenue on its facilities by the end of the fiscal year. It also desires to find cost cutting measure to offset the 1,540,000 operating costs from the installation of the new lift chair.

**Context:**

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. The business has recently installed an additional chair lift to increase visitors’ distribution. This addition increases their operating costs by $1,540,000 this season. The business wants some guidance on how to select fair ticket price to offset this operating cost and to maximize capitalization on their facilities investment.

**METHODOLOGY:**

**Data Wrangling:** Here we explore and clean our raw ski data. While exploring our data, we find that New York has the most resorts, while our own resort come at 13th place (Figure 1)

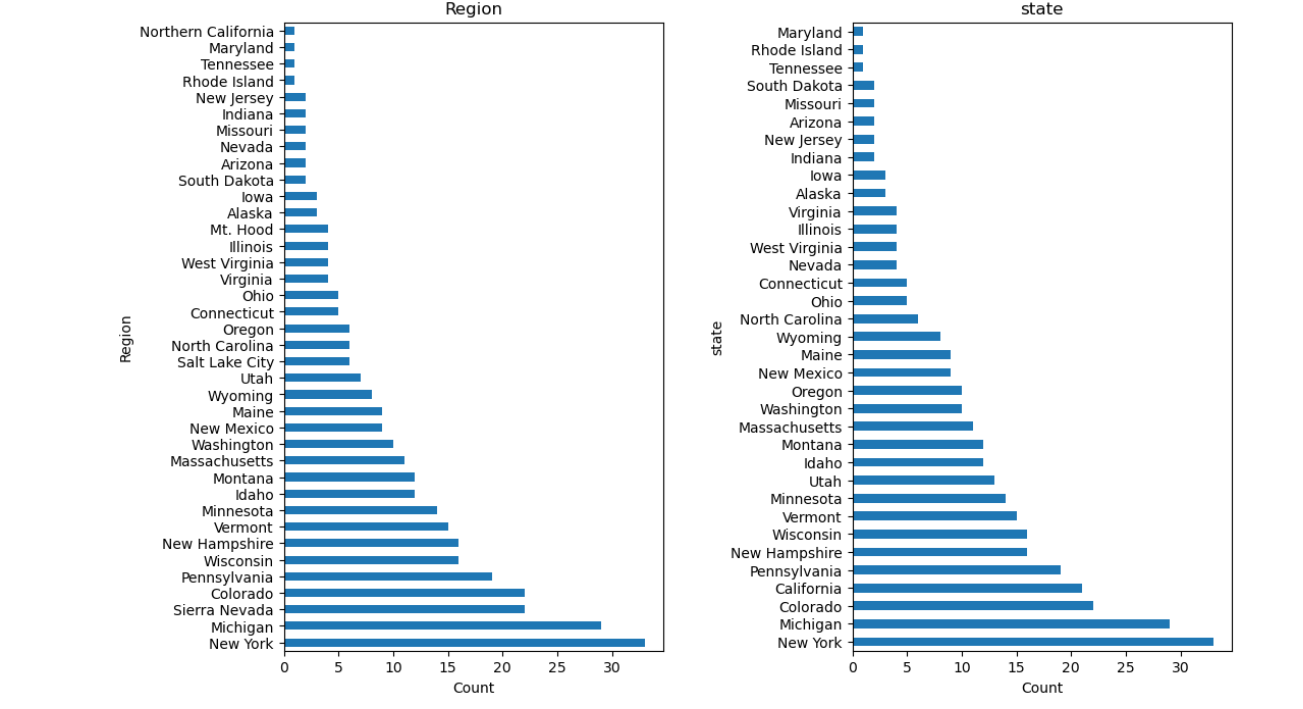


Figure 1

At this stage we also need to define our target features. Since adult weekday ticket price have more missing values in the dataset, we decide to use adult weekend ticket price as our target feature. We dropped rows that have no price data (about 14%) We also add features that captures the state-wide market size by summing TerrainParks, SkiableTerrain\_ac, daysOpenLastYear, NightSkiing\_ac. At the end of this state we have cleaned ski data set consist of 277 rows and 25 columns in addition to state summary table.

**Exploratory Data Analysis:** in this stage, we explore our data to get a state-wide picture for our market for:

total state area. (our own resort Montana come at 3rd place), Total state population, Resorts per state (New York comes 1st place), Total skiable area, Total night skiing area, Total days open etc.there is a challenge here since we found no specific pattern suggesting a relationship between state and ticket price. We found various trends. Some states are higher in certain feature, but not in others. Some features will also be more correlated with one another than others. One way to solve this is to use PCA technique to reduce the number of variables of a data set, while preserving as much information as possible. The first two components of PCA account for over 75% of the variance, and the first four for over 95%. We used PCA with 2 components to identify any trend in price (Figure 2)

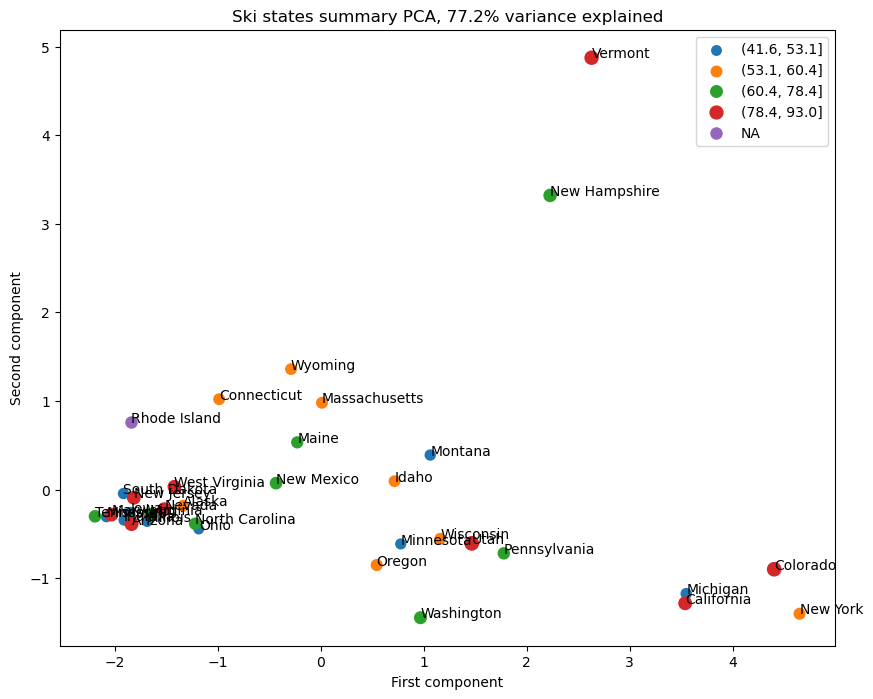


Figure 2

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.

#### We then use some Feature engineering to explore the resort-level data in more detail. We merge ski data w state summary and add "state resort competition" features (ratios)

#### A great way to gain a high level view of relationships amongst these features is Feature correlation heatmap (Figure 3)

#### 

Figure 3

#### These features are well correlated with ticket price: fastQuads, Runs, Snow Making\_ac, total\_chairs  The vertical drop seems to be a selling point that raises ticket prices as well. Of the new features, resort\_night\_skiing\_state\_ratio seems the most correlated with ticket price

**3. Preprocessing and Training:** in this stage we split our dataset into 70/30 train/test split, then develop our initial linear model by imputing missing values with either median or mean (the results are not much difference). The steps in this process are mpute missing values, scale the features, train a model (which can b automated w sklearn pipeline) Model performance is evaluated using Sklearn metrics such as r2 score and MAE

### We enhanced our linear model using sklearn function such as selectKBest to select the k best features and assessing performance using cross-validation. We want to use cross-validation for multiple values of k and use cross-validation to pick the value of k that gives the best performance; this hyperparameter search can b accomplished using sklearn function GridSearchCV. These results suggest that **vertical drop** is our biggest positive feature.

Our second model is Random Forest which work very well in a lot of cases. Encouragingly, the dominant top four features are in common with our linear model: fastQuads, runs, Snow Making\_ac, vertical\_drop

### We evaluate model performance for both Linear regression and Random Forest. Our winning model is Random Forest since it has lower MAE from cross-validation than our linear regression Model and its performance on the test set was consistent with that.

**Modeling:** in this stage we explore different scenarios to find optimal ticket price.Our model came up with predicted price of 95.87 and MAE of 10.39 suggesting there is room for an increase. We model various scenarios to find the best recommendation to suggest to the leadership. Our winning model Random Forest was chosen because of its superior reliability.

**Pricing Recommendation:** from our model scenarios, we recommend Big Mountain Resort to consider a.) increasing ticket price $1.99 coupled with strategic facility enhancements such as increasing vertical drop. This will bring additional 1.94 million revenue increase, enough to cover operating cost of chair lift additionand b.) closing one run to cut cost without affecting ticket price/revenue.

**Conclusion:** we have transformed our raw data into valuable insight to guide Big Mountain Resort leadership in finding optimal ticket price to maximize revenue while stay competitive.

**Future scope of work:** we will continue to optimize and improved our algorithm to enhanced our pricing strategy. Currently we have limited data. In the future we could integrate more data sources such as operating cost, competitor price etc. As for cutting cost by closing unused run, we can do it gradually. We already see closing one run will not affect ticket/revenue. Additional closure should b done in steps couple with monitoring and getting feedback. Lastly, we should continue to monitor and evaluate our pricing model to stay ahead in ski resort industry.