# Problem Identification Overview

How can Big Mountain Resort increase revenue for all 330 resorts to recoup their increased operating cost of $1,540 by the end of the year in order to maintain their current business profit margin of 9.2%?

Big Mountain Resort offers spectacular views and while skiing and snowboarding. The recently installed an additional chair lift to help increase the distribution of visitors, which increased operating costs this season. The business expects recommendations on recouping the increased operating costs in order to maintain profit margin of 9.2% and a projection of the this year’s annual revenue based on the recommendations.

**Purpose and Motivation:**

The purpose of this data science project is to come up with a pricing model for ski resort tickets in our market segment. Big Mountain suspects it may not be maximizing its returns, relative to its position in the market. It also does not have a strong sense of what facilities matter most to visitors, particularly which ones they're most likely to pay more for. This project aims to build a predictive model for ticket price based on a number of facilities, or properties, boasted by resorts (at the resorts). This model will be used to provide guidance for Big Mountain's pricing and future facility investment plans.

**Data Acquisition:**

We were supplied with a raw CSV data file by the database manager, which included data (ski resorts’ prices, resort layout, and equipment available) from 330 resorts in the US that can be considered part of the same market share.

**Data cleaning:**

The weekend and weekday prices were pretty equal and since the AdultWeekend column had fewer missing values. We decided to drop the AdultWeekday column and the target feature for our predictive model is the ‘AdultWeekend’ variable.

Chart, scatter chart

Description automatically generated

14% of our data are missing both AdultWeekday and AdultWeekend prices so these records without price information were dropped from the dataset.

Half the values from the The fastEight column were missing and practically all of the other values are zero. There is essentially no information in this column to work with so it was dropped from the dataset.

We also obtained some additional US state population and size data with which to augment the dataset, which also required some cleaning.

In the end we were left with 277 observations out of the original 330.

**Exploratory Data Analysis:**

Since our interest lies in ticket prices here is an image of the ticket prices for resorts in each state.

Chart, histogram

Description automatically generated

We explored the state summary data in great detail and created several new features from the state data: resorts\_per\_100kcapita and resorts\_per\_100ksq\_mile to analyze. Unforunately, there wasn't an obvious pattern between state and the ticket price. The state labels didn't seem to be particularly useful so we will treat all states equally and work towards building a pricing model that considers all states together.

We merged our state summary features into the ski resort data, add "state resort competition" features in order to put each resort within the context of its state:

* ratio of resort skiable area to total state skiable area
* ratio of resort days open to total state days open
* ratio of resort terrain park count to total state terrain park count
* ratio of resort night skiing area to total state night skiing area

We created a series of scatterplots to really dive into how ticket price varies with other numeric features. In the scatterplots we see what some of the high correlations were clearly picking up on. There’s a strong positive correlation with vertical\_drop, fastQuads, Runs, total\_chairs and resorts\_per\_100kcapita appears useful.

Below are further features that may be useful in that they relate to how easily a resort can transport people around.

Chart, scatter chart

Description automatically generated

**Modeling:**

We checked to see how good the mean is as a predictor and as expected, using the average value as our prediction gives us and R2 of zero on our training set. We created a Linear Regression model to make predictions on both the train and test splits. We then applied the Random Forest method and we marginally improved upon the default CV results. Both the linear model and the random forest agree that the dominant top four features are: fastQuads, Runs, Snow Making\_ac and vertical\_drop.

**Model Findings:**

When comparing the linear regression model performance with the random forest regression model performance, The random forest model has a lower cross-validation mean absolute error by almost $1, it the model to use going forward.

The predicted AdultWeekend ticket price for Big Mountain is $94.22, which is a suggested increase from the current cost of $81. This prediction is based on the features identified from the other 329 resorts fitting the model to Big Mountain and the validity of our model lies in the assumption that other resorts accurately set their prices according to what the market (the ticket-buying public) supports.

We reviewed several potential scenarios for either cutting costs or increasing revenue (from ticket prices).

1. Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.

Chart, line chart

Description automatically generated

1. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage
2. Same as number 2, but adding 2 acres of snow making cover
3. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

**Conclusion/Recommendation:**

We recommend they build a similar model to predict AdultWeekend Ticket Prices and test scenario 1, where we increased the vertical drop and installed the additional chair lift, without additional snow making coverage. The increased ticket price is only $1.99, which can be easily accepted by the visitors, and it would produced an increase in profit over the season. An increase in the snow making area makes no difference in the model prediction, therefore it would’t make sense to spend additional money increasing snow making area when it’s not producing additional revenue.