Data Science Career Track

**Guided Capstone**

Building a Predictive Model for Ticket Prices

horizontal line

# Objective:

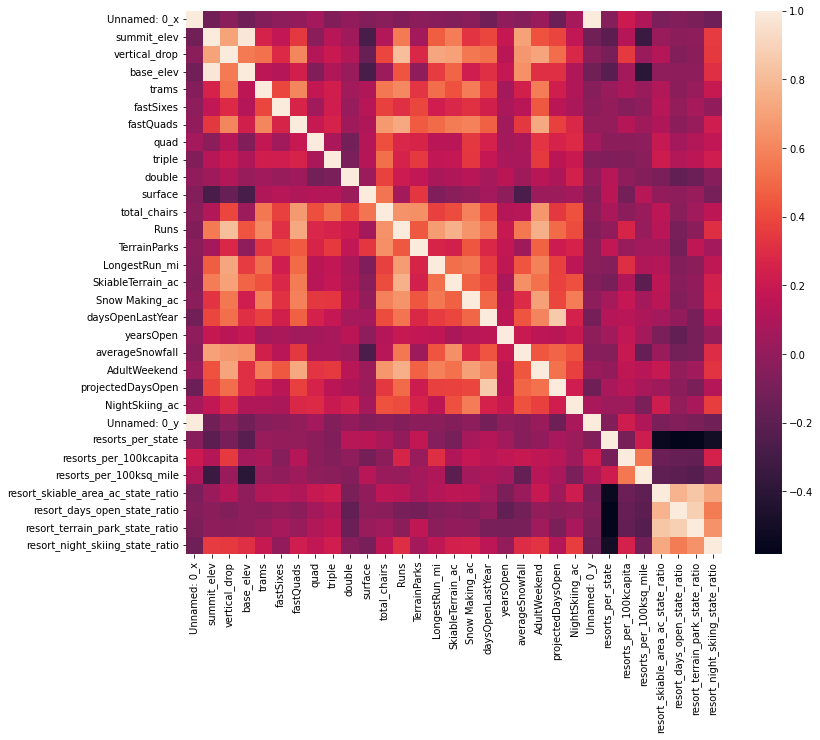
The purpose of this data science project is to come up with a pricing model for ski resort tickets in our market segment. Big Mountain, a popular ski resort, suspects it may not be maximizing its returns, relative to its position in the market. It also does not have a strong sense of what facilities 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 several facilities, or properties, boasted by resorts*.* This model will be used to provide guidance for Big Mountain's pricing and future facility investment plans.

## Exploratory Data Analysis:

This analysis started with a main dataset of 27 columns and 330 rows. This included details about the resort, some key amenities provided by the resort and its weekend Adult prices and Weekday adult prices. The second dataset was the number of resorts in each state. After a detailed data cleaning, the main dataset was used to understand the density of resorts and the state with the large number of resorts. On plotting the resort data state wise, we observe that New York has the greatest number of resorts.

A correlation heatmap of the feature’s reveals that summit and base elevation are quite highly correlated. We also notice that when resorts are more densely located with population, more night skiing is provided. The ratio features are negatively correlated with the number of resorts in each state which means that with an increase in the number of resorts in a state, the share of all the other state features will drop for each. The target feature, AdultWeekend ticket price correlates with fastQuads, Runs and Snow Making\_ac. Visitors value more guaranteed snow, which would cost in terms of snow making equipment, which would drive prices and costs up. Of the new features, resort\_night\_skiing\_state\_ratio seems the most correlated with ticket price which could mean that seizing a greater share of night skiing capacity is positive for the price a resort can charge.

Runs, total\_chairs are well correlated with ticket price. The more runs you have, the more chairs you'd need to ferry people to them. They may count for more than the total skiable terrain area. The total skiable terrain area is not as useful as the area with snow making. People seem to put more value in guaranteed snow cover rather than more variable terrain area. Vertical drop seems to be a selling point that raises ticket prices as well.



**Pre-processing and Training:**

We calculated the mean as a predictor of average price using the mean function and dummy regressor which matched. We also calculated mean squared error and mean absolute error. We use the R squared method to determine the metrics leading us to conclude that if we use the average value as the prediction, we get an R2 of zero on our training set. Then we replace missing values with the median for each feature, scale the data to zero mean and unit variance and train a linear regression model. We also perform the above functions by using means to impute the missing values. Further, using pipelines to build a linear regression model, we demonstrate that the above steps can be achieved with just two functions, fit() and predict(). As we suspected overfitting, we changed the model by selecting ‘k’ best features. We use cross-validation for multiple values of k and use cross-validation to pick the value of k that gives the best performance. We zero in on 8 best features some of which are vertical drop, snow making equipment are positively impacting ticket price. Skiable area is negatively associated with ticket price, which seemed odd. Then we use RandomForest to build a model that concurs the dominant top four features that are in common with the linear model are fastQuads, Runs, Snow Making\_ac and vertical\_drop Finally we select the best model. The random forest model has a lower cross-validation mean absolute error by almost $1. It also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

**Modeling:**

We now take our model for ski resort ticket price and leverage it to gain some insights into what price Big Mountain's facilities might support as well as explore the sensitivity of changes to various resort parameters. Note that this relies on the implicit assumption that all other resorts are largely setting prices based on how much people value certain facilities. Essentially this assumes prices are set by a free market. Big Mountain Resort modeled price is 95.26, 𝑎𝑐𝑡𝑢𝑎𝑙 𝑝𝑟𝑖𝑐𝑒 𝑖𝑠 81.00. Even with the expected mean absolute error of 10.37, this suggests there is room for an increase. Features that came up as important in the modeling included: 𝑣𝑒𝑟𝑡𝑖𝑐𝑎𝑙𝑑𝑟𝑜𝑝, 𝑆𝑛𝑜𝑤𝑀𝑎𝑘𝑖𝑛𝑔𝑎𝑐, 𝑡𝑜𝑡𝑎𝑙𝑐ℎ𝑎𝑖𝑟𝑠, 𝑓𝑎𝑠𝑡𝑄𝑢𝑎𝑑𝑠, 𝑅𝑢𝑛𝑠, 𝐿𝑜𝑛𝑔𝑒𝑠𝑡𝑅𝑢𝑛𝑚𝑖, 𝑡𝑟𝑎𝑚𝑠, 𝑆𝑘𝑖𝑎𝑏𝑙𝑒𝑇𝑒𝑟𝑟𝑎𝑖𝑛𝑎𝑐. Big Mountain has amongst the highest number of total chairs and has a pretty good vertical drop. Most resorts have no fast quads, Big Mountain has 3. It is high up the league table of snowmaking acres. It also compares well for the number of runs. Most resorts, such as Big Mountain, have no trams. It is amongst the resorts with the largest amount of skiable terrain. We test four different scenarios that help Big Mountain cut its costs or increase revenue. Scenario one concludes that closing 6 or more least used runs leads to a large drop in revenue. However, it can safely close up to 4 or 5 and there is no higher loss when compared to closing down 3. Second scenario tests the adding of a run and increasing vertical drop by 150 ft and installing an additional chairlift. This could increase ticket price by 9.46 and over the season, this could be expected to amount to $16561594. In addition to the changes in scenario two, scenario three adds 2 acres of snow- making, which translates to an increase in ticket price by 10.23 and over the season additional revenue of $1790579. Scenario 4 calls for increasing the longest run by .2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability, which created no difference whatsoever.

**Conclusion:**

Scenario 3 tested above gives a good increase in the ticket price. However, there are limitations to this as we are not aware of the cost of adding 2 acres of snow making, a run and increasing the vertical drop by 150 feet. The predicted price is higher than its current price as it is yet to capitalize on its features. A thorough presentation and documentation of how the features impact pricing should be shared with the management so future capital investments can be made in accordance with the same.