**Problem Statement**

Big Mountain Resort is a ski resort located in Mountain that offers views of the Glacier National Park and Flathead National Forest and has over 105 trails. Each year about 350,000 people ski or snowboard at the resort. Big Mountain Resort recently installed an additional chair lift to help increase the distribution of visitors across the mountain. The additional increase in the chair increased the operating cost by ~$1.5 million. The business brought in a team of data scientists to implement a data-driven business in order to select a better value for ticket price while also considering a number of changes that could either cut costs without undermining the ticket price or support an even higher ticket price.

**Data Wrangling**

We first obtained a csv file of our ski data and explored our data by identifying the missing data for each ski resort. When looking at our data we identified that fastEight column has the most missing values (50%), ticket price is missing 15-16% of values, and the AdultWeekday column is missing in a few more records than AdultWeekend column. We performed some data deletion for any rows that did not have a price data in order to get our data to be as precise as possible. Any rows without data could potentially skew our results. We performed additional data exploration to determine the following:

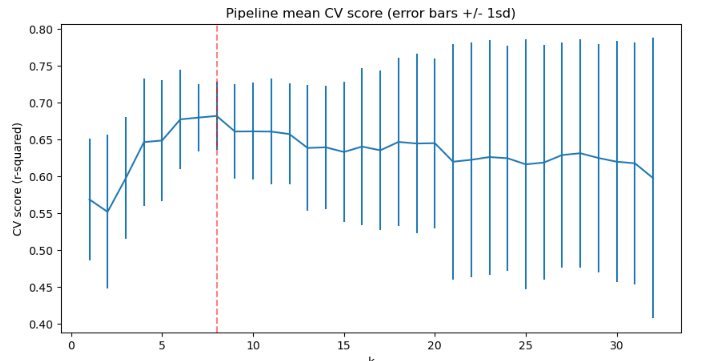
* Distribution of resorts by region and state
* Average ticket price for both adult weekday and adult weekend by state
* The ticket price ranges for each state
* Distribution of feature values (vertical drop, trams, total chairs, snow making area, days opened last year, average snow fall, etc.)

**Exploratory Data Analysis**

After we cleansed our data, we tried to determine if there were any correlation between resorts features (total skiable area, total night skiing, resorts, total days open, population density) in relation with the state. We found potentially useful and business relevant features to derive our statistics with the states we were concerned with. We used the PCA (Principle components analysis) technique to determine if there were any linear combinations from the features with state. We found that there were two (2) components (features) that accounted for 75% of the price. As we increased our components (features), we found that four (4) components (features) that accounted to over 95% of the variance.

**Model Pre-processing and Training**

We initialized our model by imputing the missing values from our ski data with the mean and median values. When using the mean values for our missing data, we assessed how the mean and median values was as a predictor. Using the sklearn's DummyRegresson on our testing and training datasets, assess how well our predicted set values agree with the actual data using metrics (R^2, mean absolute error, mean squared error). Using the metrics mentioned before, we determine whether a mean or a medium will best fit any missing data. We noted that the higher the R^2, the better fit the data. We created a pipeline (SimpleImputer, StandardScaler, and LinearRegression) that would best fit the missing data. We noted that if we used the mean value, we would expect ~$19 difference.

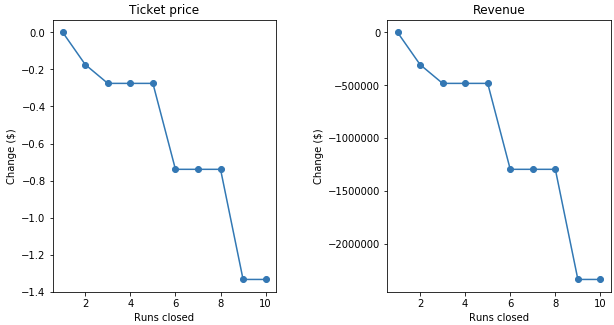
We performed a cross-evaluation technique for estimating model performance with varying levels of k using the GridSearchCV feature in the sklearn package. We noted that the bigger the k (features), the greater the variability in performance in price. 

In our model we determined that the biggest positive feature was 'vertical drop'. We also determine that the skiable area is negatively impacting the ticket price. We created a forest a random forest model which helped determine our top dominant features (fastQuads, Runs, Snow Making\_ac, and vertical\_drop).

**Model**

Based on our selected model above, we fitted Big Mountain resort into our model to determine if the business was accurately reflecting the price based on the features. We identified that Big Mountain Resort modelled price is $95.87 but the actual price was $81 with an expected absolute error of $10.39. This suggest that there is a room for a price increase.

We set out different modeling scenarios to find potential solutions of increasing our revenue or cutting cost. Our expected number of visitors over the season is 350,000 and, on average, visitors ski for five days. Using the above information, I modeled four (4) different scenarios to address the business propositions provided by our client.

* Scenario 1 - Closing up to 10 runs
  + We noticed that if we closed two (2) to three (3) runs, we see that it reduces the support for ticket price and revenue. If we close four (4) to five (5) runs, our analysis demonstrates that there is no difference. If we close six (6) or more runs, we noticed that it leads to a significant revenue decrease.   
      
    
* Scenario 2: Increase vertical drop
  + If we increased the vertical drop by 150 ft, we saw that this supports a price increase of $1.99 and thus over the season, we would expect the revenue to increase by $3,474,638.
* Scenario 3: Increase snow making
  + If we increased the snow making by 2 acres, we saw that this supports a price increase of $1.99 and thus over the season, we would expect the revenue to increase by $3,474,638.
* Scenario 4: Increase longest run by .2 miles and increase 4 acres of snow making
  + If we increased the longest run by run by .2 miles and increase 4 acres of snow making, we saw that there was no difference whatsoever.

We believe that by best scenarios would be to increase the vertical drop by 150ft because it is the most optimal choice without having to increase operating cost or decrease revenue. In scenario 3, we identified that by adding two acres of snow and increase vertical drop would increase revenue by ~$3.5 million dollars over the same option with no additional snow making.