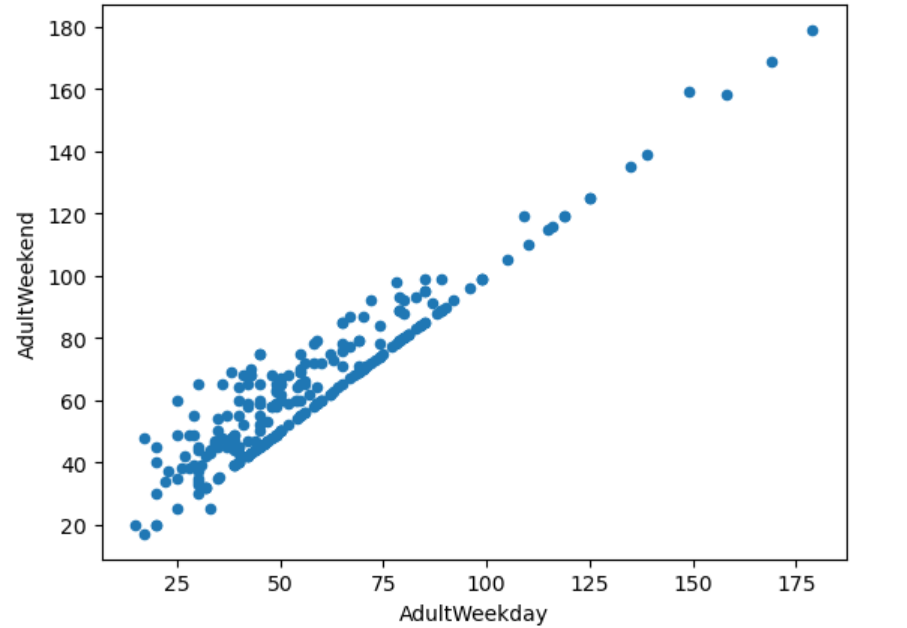
Big Mountain Resort, a ski resort in Montana, has recently installed another chair lift to better distribute the visitors around the mountain. With the increase of another chair, the operating costs also increased by $1,540,000. Big Mountain resort would like to have a better ticket price or cut costs that can accommodate the new operating costs and does not undermine the current ticket price or can support a higher ticket price by the end of this season.

**Data Wrangling**

This is when we would be collecting, organizing, and cleaning the data. We load in the data set that was sent via CSV. We check to see if any data is missing from the columns and we see that there are about 13 columns with missing values, but only one column that had over 50% of values missing, fastEight. Given that this column has so many missing values, this column was dropped from our data.

We also want to check the differences between ticket prices from weekday to weekend, we can see that there is a fine line where the weekend and weekday prices are equal, and it appears that the resorts that have greater Weekend prices are those that have less than $100 tickets already.

. 

We then check to see if Weekday prices or Weekend prices have more missing values, turns out Weekday does, and we drop that column. There was an issue, where a row seemed to be open for 2,019 years, so any row that claimed to be open for over 1000 years was dropped. With our data clean, organized and saved into a new file, we move on to our next step.

**Exploratory Data**

For this section, we start by using summary statistics to see which states were the largest in area, population, skiable area, night-skiing area, resorts per state and how many days the resort was open. New York was found to have the most resorts, but did not place in the top 5 when it came to skiable areas. This revelation led into the next section of resort density. Resort density consisted of two measurements, a ratio of resorts and the population and the ratio of resorts to the state area in square miles.

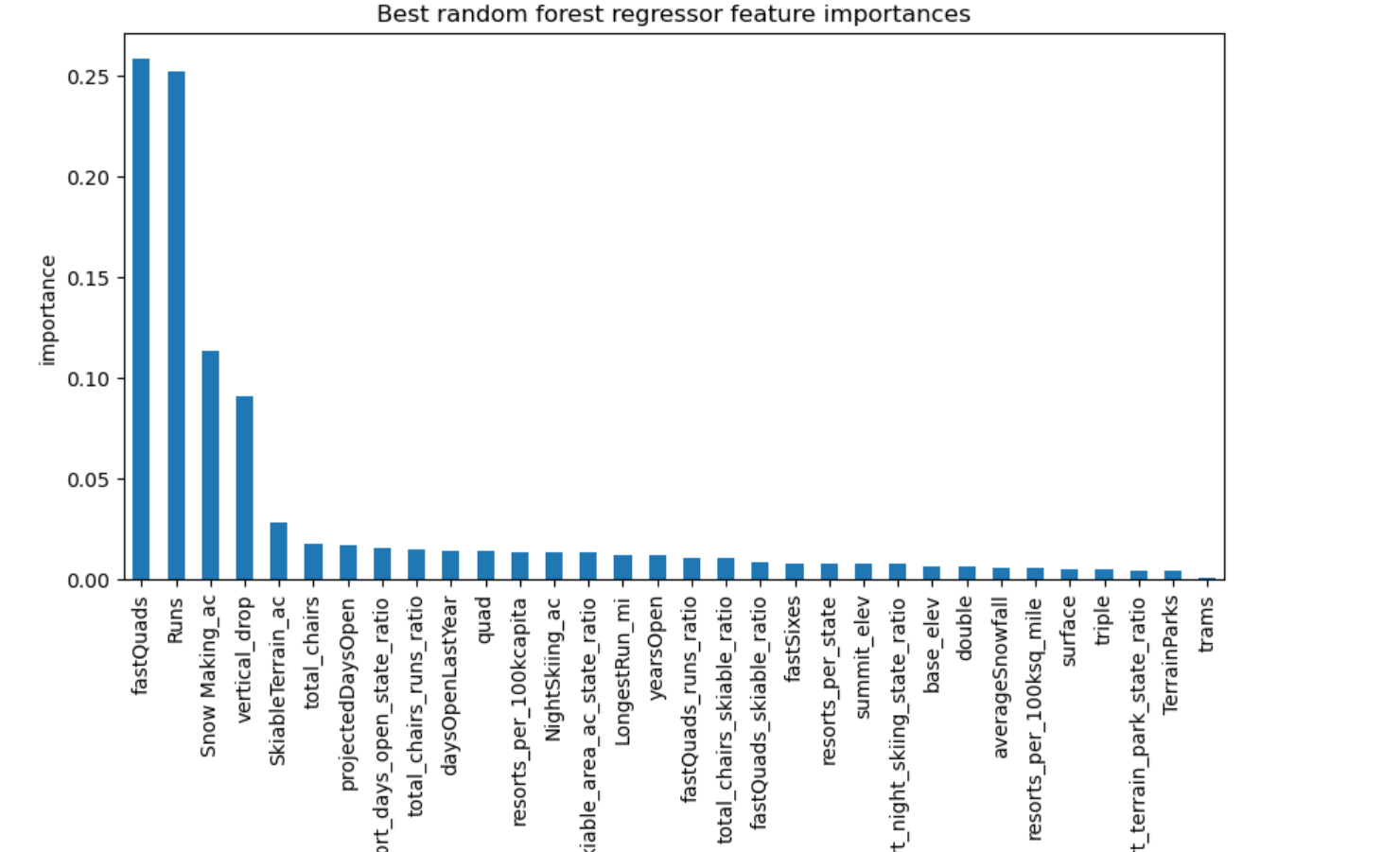
To get an idea of the web of connections for the features, we scaled our data, verified the scaled data by summary statistics and fit our scaled data to a PCA(principle components analysis ) transformation, By doing this we found that even though certain states had a higher PCA, there is not a reason to treat any state differently than the rest and we should continue forward with a model that considers all states equally.

There were also some correlations found between ticket price and vertical drop, runs, total chairs, snow making acres and fast quads.

**Preprocessing and training**

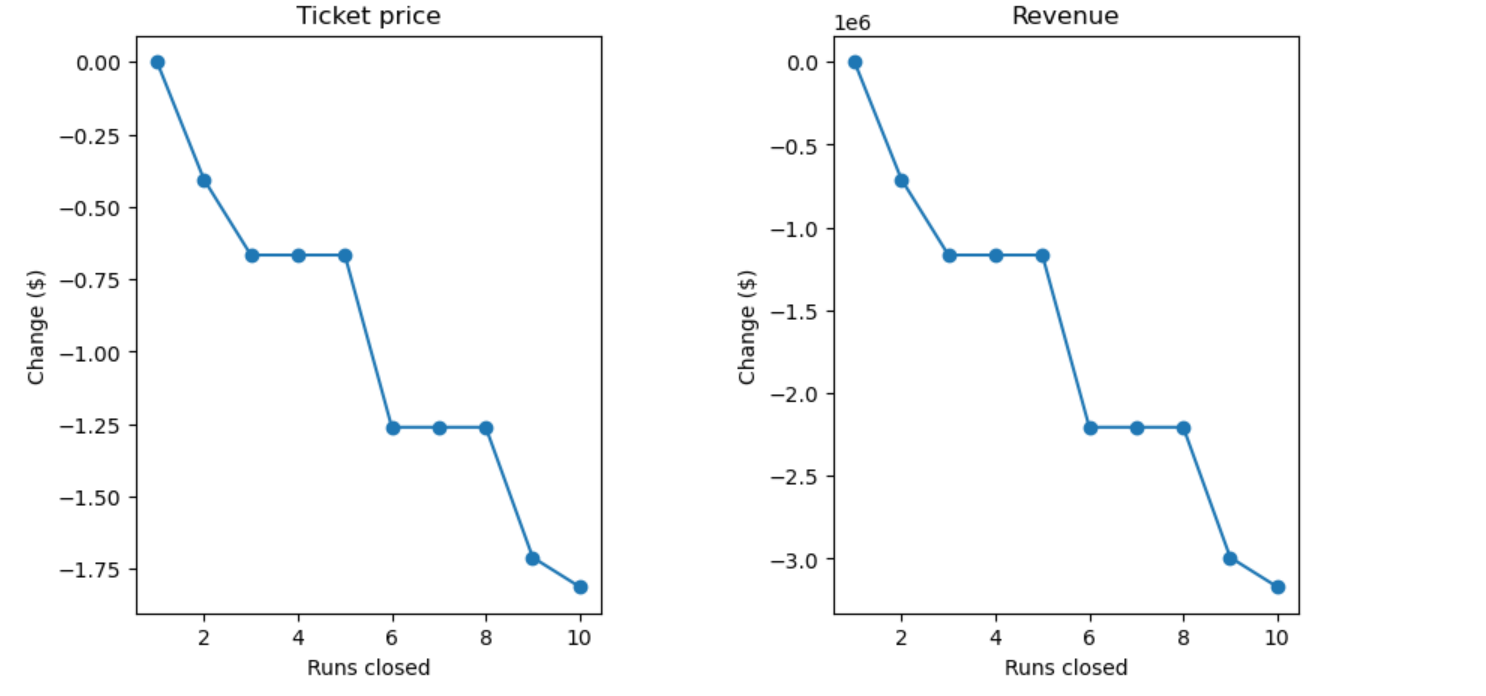
For this process, we separate Big Mountain data from the original dataset, then we use our updated ski\_data for a train/test split. Splitting that data in 70/30 partitions to train and assess which model works best. For our models we need to know the metrics R squared, mean absolute error, and mean squared error. For our linear model, we imputed our missing values once with mean and then with median, both produced similar results, which may mean our model is being overfitted. We use GridSearchCV that takes a pipeline object and takes any possible combinations to find the best option, in this case the number of best features, k, is 8. Then we find our best feature which was vertical drop followed by snow\_making\_ac.

Then to compare, we use Random Forest Model, which showed the best features being fastQuads, Runs, Snow Making\_ac, and vertical\_drop. After calculating absolute mean and the result being about $1 less than our Linear model, we continue the next step with the Random Forest Model



**Modeling**

In the final section, we use our model to gain some insights for Big Mountain resort facilities and if the resort’s ticket price should be updated or if certain cuts need to be made. To start, we calculate the expected modelled price for Big Mountain resort based on the best features, which is approximately $14 more than what they are currently offering, $81, with a $10 absolute error. When comparing Big Mountain resorts features to others, usually Big mountain is usually above the average.

Next we go through some model scenarios, which the business has shortlisted. The first scenario is to close the 10 least used runs.

We can see that dropping 2-3 runs, there is a reduction in ticket price/revenue and if the business wants to drop 3, then they should just drop 4 or 5 since there is no financial loss at that point. Dropping 6 or more we will see a big drop in price.

The second, third and fourth scenario all garner the same result, we get ticket price support by $1.99 for adding a run, chair lift, snow making acres, and extending vertical drop. At the end of the season we should be making $3474638.