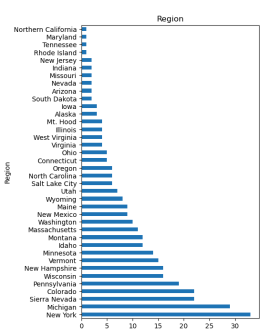
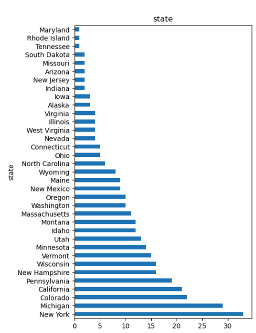
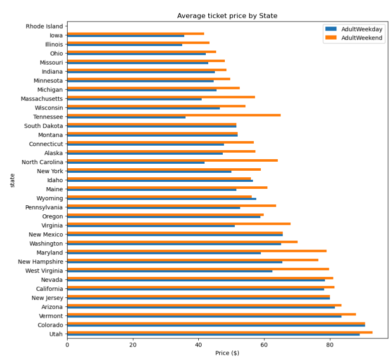
The problem for Blue Mountain was how to find an optimum pricing point for the resort based on the services they offered from studying the competitors data. This is because Blue Mountain just increased their operating cost by $1.54 Mil, by installing a chair lift. They are known for charging a premium price for their services and are not willing to cut it or increase it but instead looking for an alternative way to create a better value for their ticker to result in more customers to increase revenue. The best way to move forward for this business is to find a way to increase their revenue so that it can outway their expenses which will result in a net profit. The way to do this was by making machine learning models to construct an ideal price to increase revenue. The problems that are hindering their success right now is not having enough traction for their underutilized services and not knowing if the customers are not interested in it or see the point of their service.

The ski_data was a csv file that had the data for 330 resorts in America. It contained 330 rows and 27 columns. While investigating the data seemed like the state names and regions were very similar after finding the difference. The best decision was to use the state column instead since a few states came across different regions. After moving forward to see the competitors for Big Mountain a bar graph of the state of Montana was created to see its rank and how it compares to its competitors. It placed 13th on the highest count from all the states and regions it was compared with. To find a more realistic market price the best choice was to compare it with other resorts within the state.



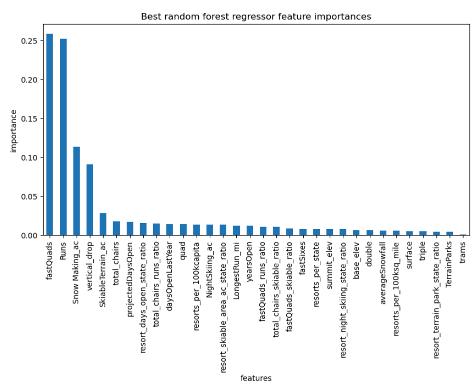




The two columns that were used were Adult Weekday prices and the Adult weekend Prices. When trying to compare these two columns, 3.34% were missing one column and 14.24% were missing both. This resulted in us dropping 14.24%. In the process of cleaning the data we dropped the fastEight column since only one resort had it and it was not needed to compare resorts with, we dropped rows that did not contain information on the weekday and weekend prices.Also added a DataFrame that contained the population of states which gave us new reports i.e resorts per state, state population, state skiable area, etc.

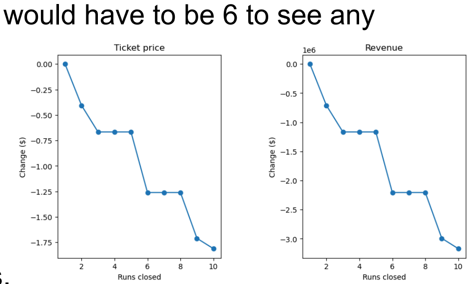
The data that was presented was regarding ski data in relation to their facility and state data relating to ski resorts within the state.The two(State_summary and Ski_data ) were merged later on. We also added a two columns resorts_per_100kcapita and resorts_per_100ksq_mile and After multiplying state_population and state_area_sq_miles because by 100,000 we added the result to two new columns: resorts_per_100kcapita and resorts_per_100ksq_mile. THis allowed us to drop state_population and state_area_sq_miles because we put the data to better us to present density of resorts relative to the state population and size. This all leads us to see the relationship between ticket prices and other features. From the data display it seemed like the more 'total_chairs_runs_ratio', 'total_chairs_skiable_ratio' it would relate to price increasing.

Using the data we constructed a Train/Test Split making 70% of it train and 30% of it test. After making sure that we were only using numbers we tested to see the train mean and a dummy regression mean to see their similarity, and they both resulted in with the same outcome which indicated the right track. Later we calculated the MAE and MSE. The MAE was done to tell us around what number we would off if guessed ticket price was based on an average of know values and the MSE was done to compare the actual and predicted values. A key value when calculating the sklearn metrics was to make sure values are in order so it does not print an error. I later used the Linear model features. fill in method to replace the 0s with the median and against with the mean, then made a prediction using the model from both train and test. The results were not too different from the median and MAE and could notice a difference in the MSE. The next step was to create a pipeline and noticed a difference in Median and the MAE by a little bit. We next repeated the process again with the SelectKBest way which picks the best K feature, which increased the MAE. In the cross-validation process we saw that the best good value for k was 8. It showed that until k there was an increase and after k there was a decrease. After that the Random Forest Regressor from sklean and after making the strategy the median.  After comparing the Linear model and the Random Forest the conclusion came to that the random fores has a lower cross validation absolute error by almost $1, and we also saw by a learning curve that we have plenty of data.





Big Mountain currently charges $81. A suggested price from the modeling is to add $1.99 to the current price to make it $82.99. By making this change your total revenue would end up $3,474,628 and would allow you to add a run, increase the vertical drop to 150, install a chair lift, and add 2 acres of snow making. Although there is no data on just the new chair increasing revenue without adding or subtracting from the original cost. We do have data that shows closing 1 run would not increase revenue too much. You would have to close 2-3 to allow you to reduce tickets and increase revenue and if you want to close any more than 3 it would have to be 6 to see any




difference in increase of revenue and space to reduce tickets.