1 Data wrangling

       First, loading original data, exploring data, we found out there are 330 entries and 27 columns in the data frame, and our resort 'Big Mountain Resort' is in the data frame columns. started to clean the data, we saw 3% of resorts are missing one value, and 14% are missing both.

       we dropped these 14% data. Because the price is our target, these rows are of no use. We also corrected some wrong outlier data. Must drop the whole row of 'Heavenly Mountain Resort'. Because there is no ticket pricing information. Then we dropped the 'fast Eight 'column in its entirety, cause half the values are missing and all but the others are the value zero. We don't get valuable information from this column.

      While we were targeting price, Weekend prices have the least missing values of the two, so we dropped the weekday prices and then kept just the rows that have weekend prices. There are still some missing values, but right now we just leave them as it. Finally, we saved our data.

Chart, scatter chart

Description automatically generated

2 Exploratory Data Analysis

      Montana was in the top five for size and makes it into the top five for the most skiing area but doesn't figure in the most populous states. It is less densely populated. New York comes top in the number of resorts in our market. But they don't account for the most skiing area. In fact, New York doesn't even make it into the top five of skiable areas. The top five of the area of skiing available at night are more northerly states including New York.

      If we increase the number of resorts in a state, the share of all the other state features will drop for each. And there is some positive correlation between the ratio of night skiing area with the number of resorts per capita. It means that when resorts are more densely located with population, more night skiing is provided.

      Our target feature is 'Adult Weekend ticket price', we see quite a few reasonable correlations. fast Quads stands out, along with Runs and Snow Making\_ac. Visitors would seem to 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.

      As well as Runs, total\_chairs are quite well correlated with ticket price. This is plausible; the more runs you have, the more chairs you'd need to ferry people to them! Interestingly, they may count for more than the total skiable terrain area. For sure, people seem to put more value in guaranteed snow cover rather than more variable terrain area. The vertical drop seems to be a selling point that raises ticket prices as well.

Graphical user interface

Description automatically generated with medium confidence

4 Pre-Processing and Training Data

      Started by Guessing A good place to start is the average price. We then built a machine learning model.

      We can confidently present our results to business: We have enough data to say the random forest model has a lower cross\_validation mean absolute error by almost $1. It also exhibits less variability.

A picture containing timeline

Description automatically generated

5 Modeling

     Big Mountain currently charges $81. our model price is $95.87. There is room for an increase even with the expected mean absolute error of $10.39.

    The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. If Big Mountain closes 3 runs, it seems they may as well close 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop.

    This scenario increases support for ticket price by $1.99. Over the season, this could be expected to amount to $3474638

    In this scenario, Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift.

    This increases the longest run by .2 miles and guarantees its snow coverage by adding 4 acres of snow making capability.

    Although the longest run feature was used in the linear model, the random forest model because of its better only has the longest runway down in the feature importance list. It made no differences.

Chart, line chart

Description automatically generated