**1 Data Description**

1.1 Missing Data

As shown in figure 1.1 some of our variables had a lot of missing data. For example, the viable fastEight was missing half of its entries. So that individual data point was dropped. Also, the weekend and weekday ticket price variables were missing about 15% of the total data. Resorts without ticket information were dropped because they did not have the target metric. Fortunately, Big Mountain Resort is not missing any information itself.

1.2 Geography

The geography of the data is spread over 35 states. As figure 1.2 shows the location of the resorts is concentrated around a select number of states. The pricing of tickets by state has a closer range among states than the pricing around states as seen in figure 1.3.

1.3 Variables

The distribution of variables was looked at to see if there were any potential issues with them. For the most part the variables were acceptable for use in data analysis. Some variables were skewed which means they may not be useful to incorporate in modeling because they will be unreliable. See figure 1.4.

1.4 Population

The population and area of each state were added. The addition of these variables will potentially be used in further analysis.

1.5 Summary

The starting number of rows and columns were 330 and 27, respectively. Big Mountain Resort was present in the dataset and had no missing values. The column fasteight was dropped and the rows with no ticket prices were. The cleaned data has 25 columns and 277 rows.

**2 Exploratory data Analysis**

The exploratory data analysis began by looking at individual attributes of ski resorts at the state level. These attributes were total area and population, resorts per state, skiable area, night skiing offered, and days open over the season. To dive deeper the relationship between population density and ski resort density were looked at. The visualization showed some consistency between population and presence of ski resorts. When looking at correlations between variables, there were strong positive relationships between vertical drop, number of fast quad lifts, and number of runs (see figure 2).

**3 Modeling**

3.1 Preliminary Modeling

While preprocessing the data the average was tested as a means of deciding ticket prices. Upon investigation it was found using the average was inferior to modeling and had large variation. A linear model was able to explain a large variation in ticket prices but was not as good an estimator as random forest regression. The difference was less variation for estimation than the linear model. With the forest regression model median values and scaling features were used. It was found that median values helped whereas scaling the features did not. Cross validating the models by absolute error was the method that determined the forest regression was superior.

3.2 Conclusions

Big Mountain currently charges $81.00 whereas the pricing model predicted that the price should be $95.87. Relative to other ski resorts Big Mountain is has a larger vertical drop, makes more snow, has more chair lifts, has the more fast quad lifts, has more runs, ranks at the top for longest run, and has a more skiable acres. With that said, a price above most ski resorts is justified and following the model’s prediction is appropriate. If Big Mountain were to entertain another ski lift for 0.2 miles, it would justify a price increase of $1.99 with a gross revenue increase of $3,474.638. However, adding snow making on this new section would not be worth the investment. Another option is trail closure which closing one run does not matter. However, closing 2 or three would reduce support for ticket price, and closing 4 or 5 does not lead to more revenue losses however closing 6 or more would result in large price drops.

3.3 Future Considerations

There were many absent values for many variables. Missing values means part of the story is missing which could change the conclusions from this analysis. Other cost information that would be helpful would be the cost of opening new runs and snow making. This price disparity can be explained by how much higher Big Mountain ranks than other resorts for relevant factors for predicting ticket price. This business can use this model as a tool to inform the decision-making process as to what the premium pricing should be.

**Figures**

Figure 1. Missing Data

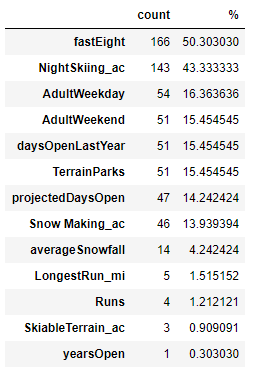


Figure 1. Geography of the data

Chart, bar chart, histogram

Description automatically generated

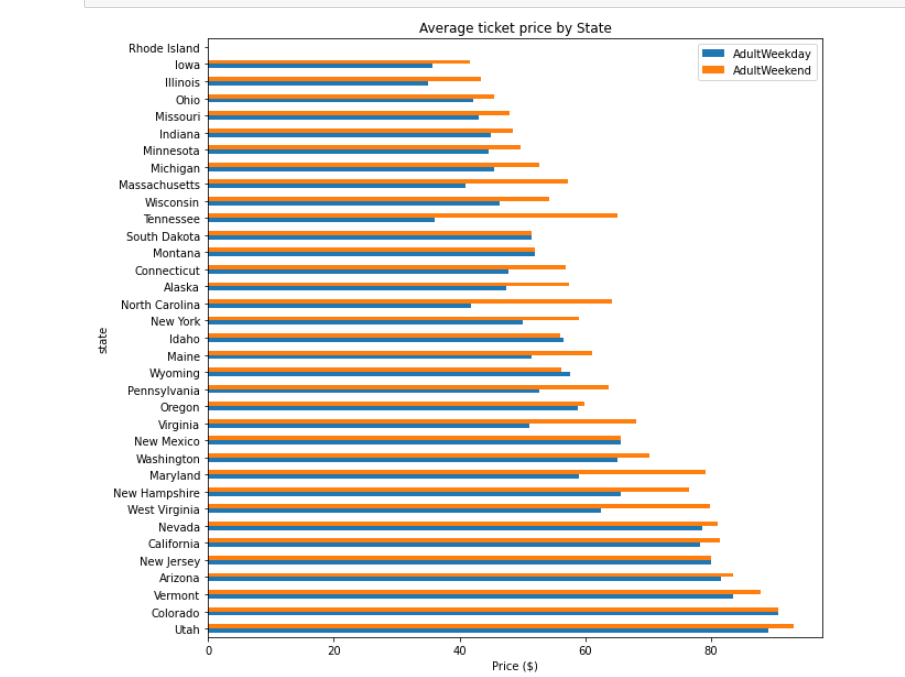
Figure 1.3 Price by state

Figure 1.4 Distribution of Variables

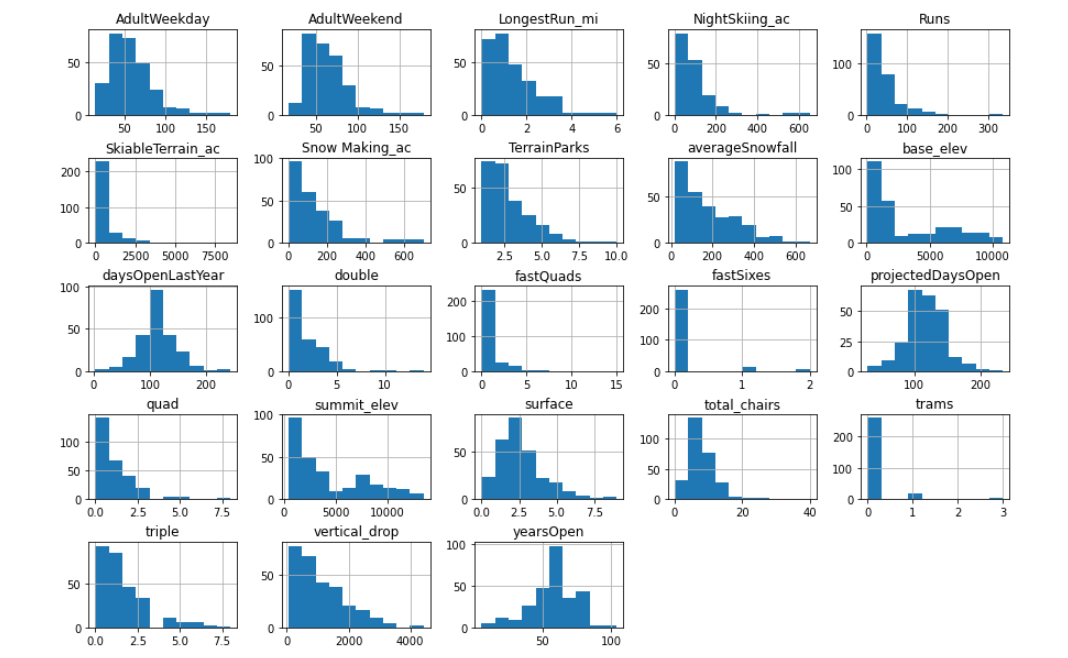


Figure 2

