**Project Overview:**

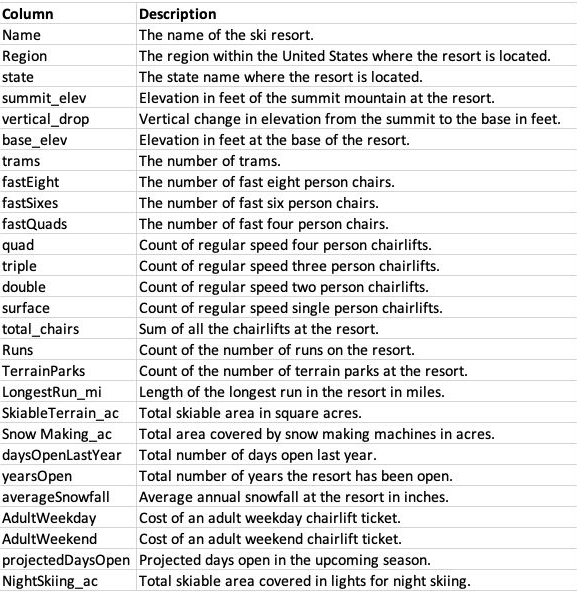
Your client is Big Mountain Resort, a ski resort located in Montana. Big Mountain Resort offers spectacular views of Glacier National Park and Flathead National Forest, with access to 105 trails. Every year about 350,000 people ski or snowboard at Big Mountain. This mountain can accommodate skiers and riders of all levels and abilities.

The resort's pricing strategy has been to charge a premium above the average price of resorts in its market segment. They know there are limitations to this approach. There's a suspicion that Big Mountain is not capitalizing on its facilities as much as it could. Basing their pricing on just the market average does not provide the business with a good sense of how important some facilities are compared to others. This hampers investment strategy. You are part of a new data science team brought in to implement a more data-driven business strategy. The business wants some guidance on how to select a better value for their ticket price. They are also considering a number of changes that they hope will either cut costs without undermining the ticket price or will support an even higher ticket price.

**Problem Statement:**

How can Big Mountain Resort use market data from resorts around the country to derive the best value for their ticket price and determine the financial impact of facility changes for the upcoming season?

**Market Data:**



**Goal:**

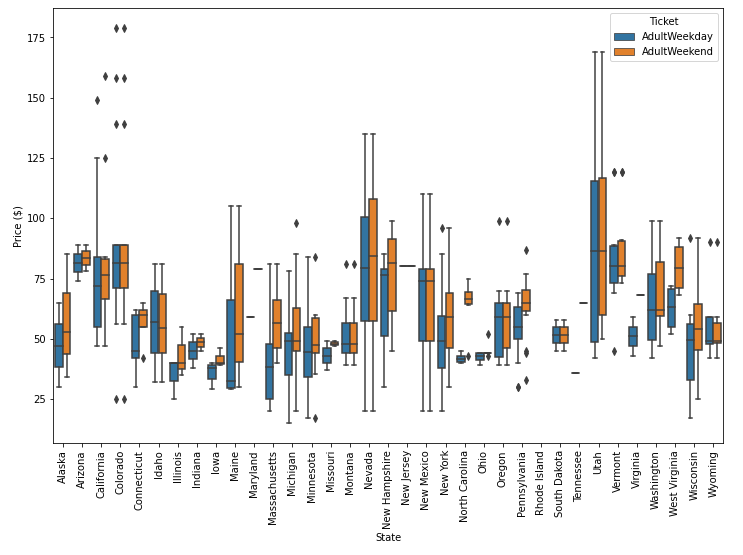
The purpose of this data science project is to come up with a pricing model for ski resort tickets in our market segment. Big Mountain suspects it may not be maximizing its returns, relative to its position in the market. It also does not have a strong sense of what facilities matter most to visitors, particularly which ones they're most likely to pay more for. This project aims to build a predictive model for ticket price based on a number of facilities, or properties, boasted by resorts (*at the resorts).* This model will be used to provide guidance for Big Mountain's pricing and future facility investment plans.

**Process:**

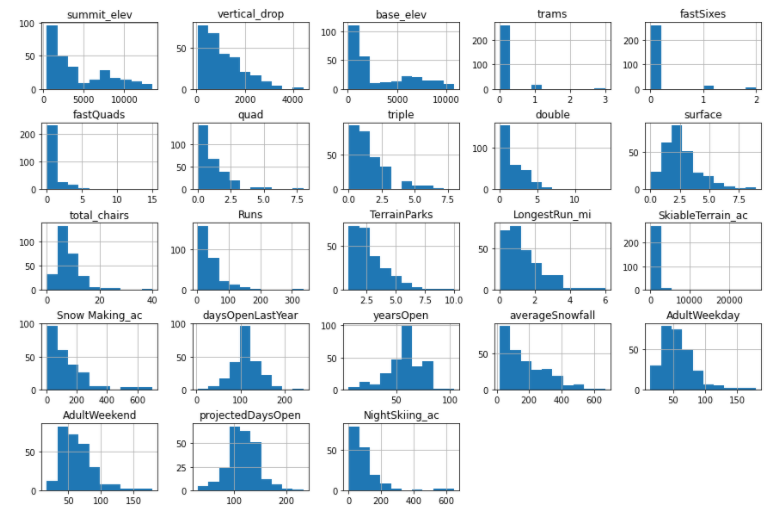
**Data Wrangling:**

The data we started with contained some ticket price values, but with several missing values that led to several rows being dropped completely. We also had two kinds of ticket price. There were also some obvious issues with some of the other features in the data that, for example, led to one column being completely dropped, a data error corrected, and some other rows dropped. We also obtained some additional US state population and size data with which to augment the dataset, which also required some cleaning. Reviewed graphs like the ones below to spot irregularities in numeric values are review state trends.

**Pricing Data per State**



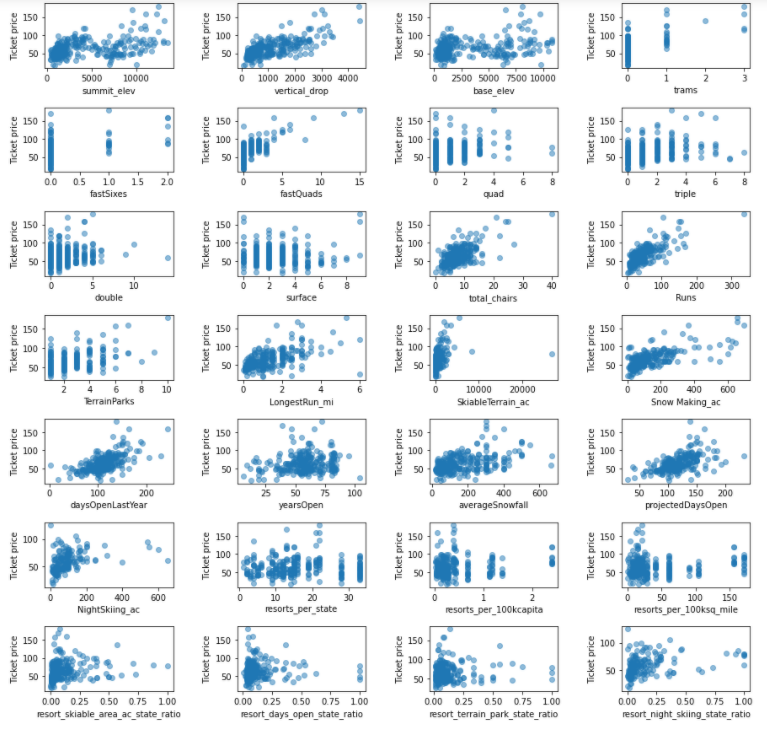
**Feature Distribution**



**Exploratory Data Analysis:**

During EDA, we found a small number of data values that gave clear choices about whether to replace values or drop a whole row. We determined that predicting the adult weekend ticket price was the primary aim. We threw away records with missing price data, but not before making the most of the other available data to look for any patterns between the states. We didn't see any and decided to treat all states equally; the state label didn't seem to be particularly useful. However, we did not have attendance data for each resort so it was hard to tell the different demand levels in different states. We drew a correlation between each feature provided in the data and ticket price as shown below. As far as target features, it looks like vertical drop, runs, total chairs, and resorts per capita show interesting correlations between price.

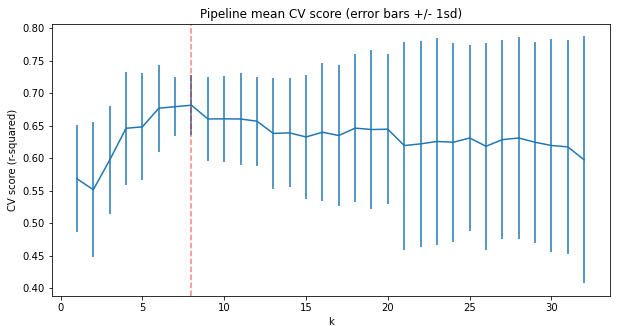
**Features vs Ticket Price**



**Pre-Processing and Training:**

In this step, we trained and tested our model. We ran a linear model and a Random Forest Model and decided that the Random Forest Model had a low mean absolute error and exhibited less variability. We noticed that the dominant features in our model were Vertical Drop, Runs, Snow Making Acreage, and fast Quads and that the optimal number of features to test is 8 as shown by the chart below.

**Pipeline Mean CV Score**



**Modeling:**

We then ran the model on all available data and came up with some results.

**Takeaway:**

**Price**

Currently, Big Mountain resort charges 81 per ticket. According to our modelling, Big Mountain could support a ticket increase anywhere from $4.91 to $15.32 leading to a ticket price of $85.91 to $96.32 respectively. This range is derived from the mean absolute error of the model. These price increases would lead to a revenue increase of $8,592,500 or $26,810,000 respectively.

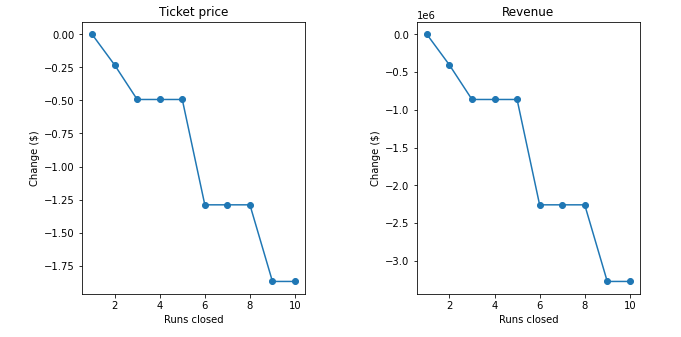
This result should be looked at optimistically and doubtfully! The validity of our model lies in the assumption that other resorts accurately set their prices according to what the market (the ticket-buying public) supports. The fact that our resort seems to be charging that much less than what's predicted suggests our resort might be undercharging. But if ours is mispricing itself, are others? It's reasonable to expect that some resorts will be "overpriced" and some "underpriced." Or if resorts are pretty good at pricing strategies, it could be that our model is simply lacking some key data? Certainly we know nothing about operating costs or total ticket sales for other resorts, for example, and they would surely help. I would not suggest raising the price by $15 initially but I would say that the model implicates that there is room for growth in the price.

The reason for this price optimization is that Big Mountain leads the market in the most important facility metrics that lead to increased ticket price such as Vertical Drop, Snow Making area, Total number of chairs, Fast Quads, Runs, Longest Run, and Skiable Terrain Area. The graphs at the bottom of this document explain this.

**Enhancements**:

Regarding recent renovations, the new chairlift should allow the business to charge $0.29 more per ticket leading to a revenue gain of $507,246 according to the model.

As for future improvements, we would need to know the cost of each enhancements. I would advise against Scenario 4 and Scenario 3 as they do not provide incremental value. I would advise looking into Scenario 2 as that will increase support for a ticket price by $1.99 for a total increase of $3,474,638. If Big Mountain wanted to pursue run closures, I would limit it to one at first as that had little effect on the modeled ticket price. I would not close more than 5 runs as there is a significant drop off in price with that change. I would suggest closing a few at a time and see how it affects crowds and demand. This is show in the chart below:



**Further Works:**

A reason to be skeptical of the model is shown in the plot directly below. Big Mountain leads all resorts in ticket price in Montana. The data we had did not contain attendance data to determine demand levels by state. We also lacked pricing data for upgrades which limited the scope of this analysis. Cost information for each of the scenarios would be helpful in providing additional business insight. Although Big Mountain was a relatively expensive compared to other resorts, the modeled price was so much higher because of Big Mountains superiority in several metrics that proved to be most valuable throughout the market.



**Takeaway Plots:**

