Guided Capstone Report

Big Mountain is one of many ski resorts in the nation, and is based out of Montana. They just put in a new ski lift that will help get customers up the mountain faster and more efficiently. This however comes with the increased operational cost by $1.5 million. Although fine with the increased cost, ownership would like to be able to maintain the 9.2% profit margin that they achieved last year. We a file with public information on many other ski resorts that includes information such as ticket prices for both weekdays and weekends, how large of an area is skiable terrain, and much more.

Having all of the data is one thing but figuring out what to do with it is another. The first step was to clean the data frame so we could get a clear understanding of what is going on with the other firms. The data frame had a lot of missing values that needed to be accounted for before we could get insight from the data frame. The way that we did that was by isolating the columns that had the null values and column by column we had to decided wither putting zero in place of the null value or the mean of values in that column in place of the null value.

Once we had the data frame with no missing values, we could start to look at what we would want to isolate as a possible solution. The way we did that was by trying to find columns in the data frame that Big Mountain has control over, such as weekday ticket prices, weekend ticket prices, and number of quad ski lifts. The reason we want to focus in on columns such as these is because other columns such as the state the resort is in as well as the summit height is something that can not be changed. However when building the predictive model these factors are still important because there is a correlation between the ticket price and the state that the resort is in. If we simply omitted those then the model would not have all the information available when it comes down to a prediction.

When it came time to build the model we needed to feed the model the data frame of the known values, without the outliers that skew the statistics. From there we end up with a map that shows how each of the columns are correlated to each other. From there we select the columns that we want to find the relationships between and we are able to see how the remaining columns effect what we want to control for.