**Guided Capstone Project Report**

**Notebook 02 - Data Wrangling**

We first imported the relevant .csv file with the original messy data. We found that several columns had around 15% missing data, and the fastEight column had 50% missing so we dropped that column from the dataframe. More data analysis found other discrepancies such as duplicated rows, and 14% of rows had missing values for both price columns. One resort had an incorrect value for skiable\_area. After dealing with all missing or incorrect values we saved the cleaned data to a new .csv file.

**Notebook 03 - Exploratory Data Analysis**

After cleaning up the state\_summary dataframe, we graphed several columns to see distribution. We found that Vermont state in particular scored high in resorts per capita (2.4/100k) and in resorts per area. We plotted the cumulative variance ratio explained by PCA components for state/resort summary statistics and found that the first four components accounted for over 95% of the variance. With PCA we found that resorts\_per\_100kcapita and resorts\_per\_100ksq\_mile are quite important columns when it comes to determining price, and we also found that runs and total\_chairs columns are also highly correlated with price.

**Notebook 04 - Pre-Processing and Training Data**

First we split our data into train and test data. First we just found the mean of train data for a starting point. Then we determined that R-Squared values as well as Mean Absolute Error and Mean Squared Error. Using the average value as prediction gave us an R-squared of 0.0 on train data but -0.000719 on test data. Next we imputed values using scikit-learn and median values. This gave us a better r2. Using this model we can expect to estimate ticket prices within $8-12 of real price. We then built a pipeline for Imputing, Scaling and Regression. Then we added KBest to the pipeline which estimated ticket prices to $9-13 of real price. So we tried a Random Forest pipeline which named Runs, vertical\_drop, and fast\_quads as three most important features, and allowed us to drop our estimation of price to within $9.

**Notebook 05 - Modeling**

In this notebook, we now take our model for ski resort ticket price and leverage it to gain some insights into what price Big Mountain's facilities might actually support as well as explore the sensitivity of changes to various resort parameters. We used our best model from the previous notebook to calculate a Big Mountain Resort price of $92.20 (actual is $81.00). We also found that the more runs are closed the bigger the impact on ticket price (with 6 or more closed making a big difference). We calculated the expected ticket price increase of several scenarios and found that increasing Runs by 1, Vertical Drop by 150ft, and Total Chairs by 1 would allow an increase in ticket price by $1.85, generating $3240741 over the season. Adding additional snow making areas makes no difference to estimated season revenue, and neither does increasing the longest run. This is our conclusion for how Big Mountain Resort can increase ticket price and increase revenue with the least amount of new features added.