**Guided Capstone Project Report**

Cyndie Jean

University of Maryland Global Campus

**Guided Capstone Project Report**

`Big Mountain Resort is a ski resort in Montana, looking at different ways to increase its revenue. They decided to add chair lift, which would see an increase in operating costs by 1.5 million dollars. To deal with this additional cost, they were trying to decide what ways they could make up for the increase in operating costs. The goal of the analysis and modeling was to figure out whether Big Mountain Resort should increase ticket prices or cut down costs so that profit would increase by the end of next season. The dataset used was provided by the Database Manager, Alesha Eisen. The dataset included data not only from Big Mountain Resort but data from numerous ski resorts across the United States including some in Montana. To work with the data, there was some data removed due to inaccuracy, multiple missing values, or not being useful for the problem being solved. After the data was cleaned the dataset was now useable to answer the hypothesis.

The next step after cleaning and identifying the problem was to do an exploratory data analysis (EDA) to see if the dataset had any pattern that should be noted. Figure 1 captured a principal component analysis (PCA) summary between states and the variance of their adult weekend ticket prices. Since the scatterplot was scattered in the majority of one region, this meant it was safe to treat all the states equally, which meant that each resort could be treated the same. Another scatterplot was created to see the variable's relationship with adult weekend ticket prices and to see if any patterns stood out. One of the variables that seemed to have a strong relationship with the adult weekend price was the variable longestrun\_mi, which was one variable that was used later for modeling.

Before a model can be done, the data set needs to be split into a testing and training dataset. This is so that the model performance can be tested to make sure that it is accurately doing what it is designed to do. So, after the Big Mountain Resort was removed the data was split into a 70/30 split. 70 percent of the data was used to train the dataset and the rest was used to test the dataset. Then the mean, standard error, R squared, mean absolute error, and mean squared error were checked and compared to make sure that the test data was correct compared to the training data. The next step was to scale the data, after filling missing values with the median value of the respected column. The scaled data was compared to the training/testing data to make sure it was safe to use. There were various pipelines created to find the correct model. The pipeline that did not overfit the data was used. In this pipeline, all missing values were filled with the median value, scaled the data to zero, trained a linear regression model, the best k feature using f regression where k was set to equal 15. The cross-validation was used to assess performance. In figure 2, found that the top 4 features related to adult weekend ticket prices were: fastQuads, Runs, Snow Making\_ac, and vertical\_drop.

The final stage looked at Big Mountain Resort compared to all other resorts and then created a model based on the data given. Big Mountain seems to be around the upper middle comparing the variables vertical drop, total number of chairs, and fast quads distribution for resorts. For the variable snow-making area, Big Mountain was on the higher end of the distribution. The final thing was creating 4 different models. The first looked at how ticket price and revenue would be affected if a certain amount of runs closed which saw a decrease in both ticket price and revenue (Figure 3). The next model looked at how the ticket and revenue would be affected if Big Mountain added a run, increased the vertical drop by 150 feet, and installed one more chair lift. This model supported a price increase of $8.61 and expected the revenue to be around 15 million dollars. The third model looked at if a run was added, the vertical drop increased by 150 feet, one more chair lift and the snow-making acres were increased by 2 acres. This supported a ticket increase of $9.90 and expected revenue of about 17 million dollars. The final model looked to see what would happen if the longest run was increased by .2 miles and added 4 acres of snowmaking. This model saw no difference. After looking at the models produced, the suggestion would be to increase the ticket price by $9.90 to get the biggest revenue possible given the dataset and modeling process.

Figures

**Figure 1**

*Ski States Summary PCA*

*A graph of states with numbers and names

Description automatically generated with medium confidence*

**Figure 2**

*Best Random Forest Regressor Feature Importances*A graph with blue and white text

Description automatically generated

**Figure 3**

*Runs closed v. Ticket Price and Revenue*A graph of a price change

Description automatically generated with medium confidence