**Data Science Career Track**

**Guided Capstone Step 6**

The first step of my work was to identify what Big Mountain Resorts problem was. After completing a problem statement worksheet, I had a better idea of what my goals were. The main goal was to improve the pricing model of the resort based on the data they have collected. To do this, I want to first have an idea of what data the Resort has been collecting over the years, after reading the data, we need to first perform data wrangling, cleaning and exploratory analysis to figure out how to create the model and what type of model is the best model to create with the provided data.

After going through the data, I realize it is possible to create a predictive pricing model for tickets based on the number of features the resort has. However, the data I received clearly needs some cleaning and order before getting started with creating the model. The first step for this was getting rid of rows with a lot of missing or incomplete data. I also noticed there are 2 rows with ticket prices, AdultWeekend and AdultWeekday, I decided to use the row with the least missing values which was AdultWeekend. AdultWeekend will be the price to predict. During this step I also decided to create a DataFrame containing the data of states and its population which will be added later to the ski resort data.

A screenshot of a graph

Description automatically generated

The third step of my work required EDA (exploratory data analysis), in this section I was able to load all the previously cleaned data and visualize it after performing principal component analysis which helped me understand the relationship between features and ticket prices. Here I also performed feature engineering, categorical features were related to state names, resort names, and region, all other data were numerical or other. After exploring the data I chose to use the features that are correlated more to the ticket prices (Vertical\_drop, Total Chairs, Runs, SkiableTerrain\_ac) this four features were found to be highly correlated to our AdultWeekend price.

A close-up of a grid

Description automatically generated

The fourth step I started applying machine learning to the data. I first separate the data of the resort the bosses are interested in which Big Mountain Resort. Then I split the ski\_data into training (70%) and testing (30%) while also dropping non-numeric rows. I explore the mean as a predictor by using a dummy model, linear regression model and random forest regression assessing how accurate its results are by using R2 (coefficient of determination), the MSE (mean squared error), and MAE (mean absolute error). After evaluating all the models I have to decide which one I’m using, the linear model testing found vertical\_drop Snow Making\_ac total\_chairs and fastQuads were the biggest positive features in prediciting AdultWeekend ticket price, the random forest model found fastQuads Runs Snow Making\_ac vertical\_drop were the biggest positive features. In the end I chose the random forest model because it is the one with the lower mean absolute error, this model ensures lower variability and predictions closer to ticket price.

A graph with blue and white text

Description automatically generated

For the fifth and final step we take all of our ski resort data and leverage it in order to finally propose how much can Big Mountain Resort increase its ticket prices. During this process we can model different scenarios in which Big Mountain Resort is able to increase its prices. After successfully modeling the data I conclude the resort can increase the ticket price by $14.43. I also consider the possibility of creating a dashboard that is easy to use for executives in which they can explore on their own different scenarios to increase ticket prices.

A screenshot of a computer

Description automatically generated