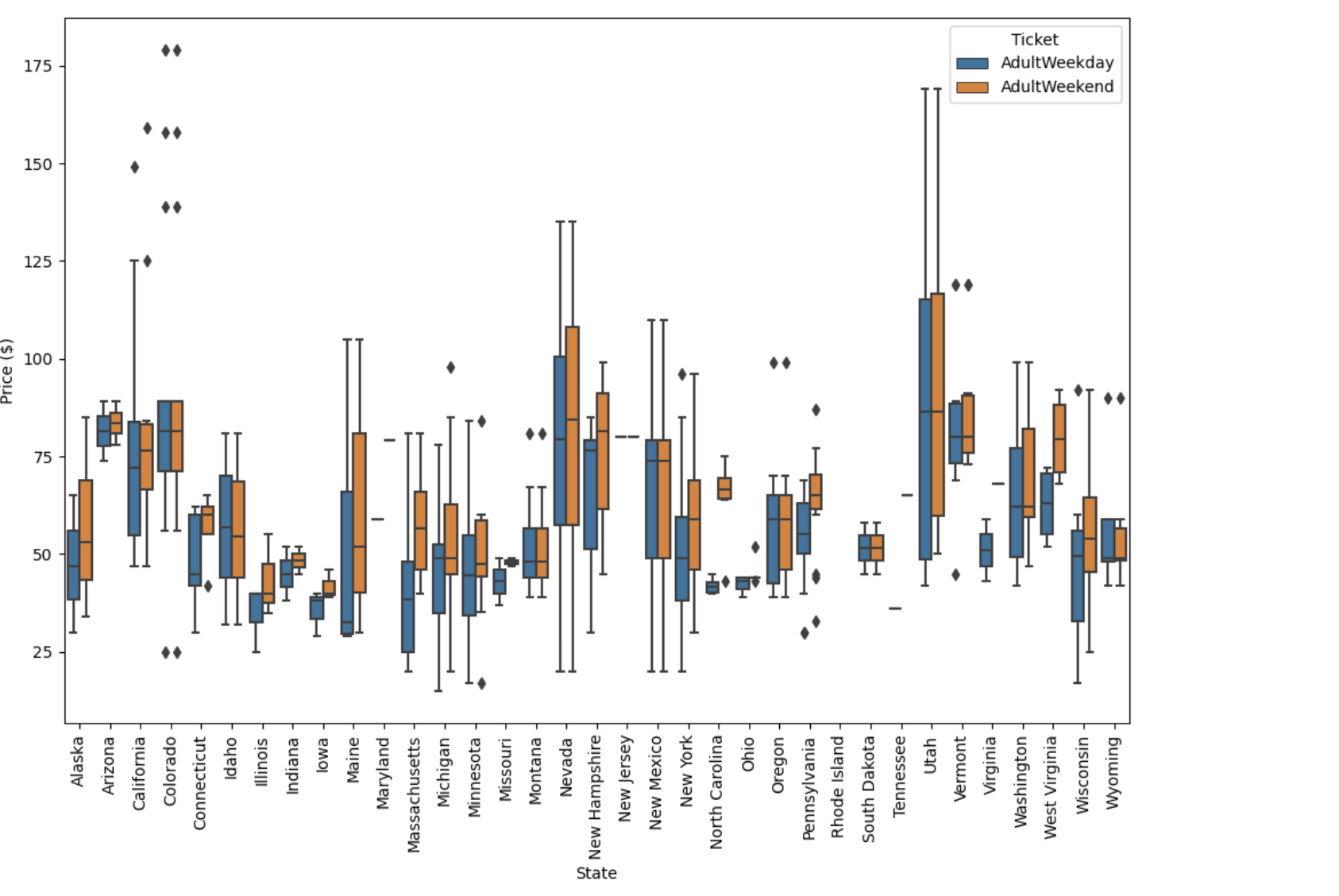
**Guided Capstone Project Report**

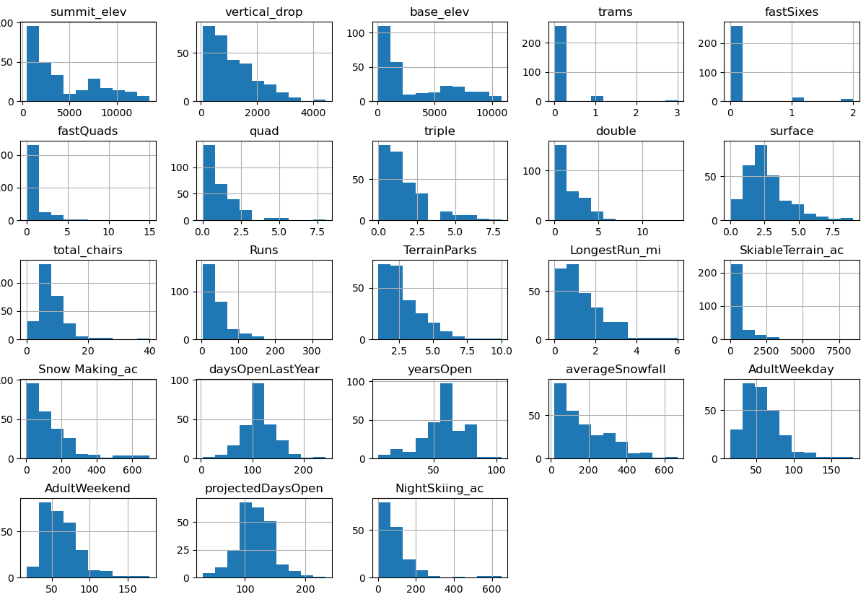
Big Mountain Ski Resort (BMSR) is a ski resort in Montana that has tasked and assigned the data science team to identify opportunities to increase their ticket prices to obtain the best value from the resort’s features, increase resort revenues, and potentially reduce operating costs by the end of next ski season. At the start of this assignment, BMSR reported that their operating costs amount to over $1.5 million after installing an additional chair lift to accommodate for potentially increased number of visitors, in addition to their 11 lifts, 2-Tbars, and 1 magic carpet lifts. From a high level, the data analysis observed at the conclusion of this assignment showed that increasing BMSR’s vertical drop feature by 150 ft, adding a chair and run would justify a ticket pricing increase, while increasing their revenue by over $1 million. In addition, eliminating up to five runs would not only save on capital expenditure on operating costs, but also provide the least impact to ticket pricing and revenues.

The beginning of the data science assignment kicked off with wrangling data that was extracted from BMSR’s data base in the form of a CSV file, which contained resort feature detail on over 300 resorts across America. The data provided by BMSR was cleaned and evaluated by identifying the data types available, the categorical value relationships, identifying any outliers, missing, or duplicate values, and assessing which features that could be eliminated from the data frame. After establishing the dataframe, the following data and features are shown in Figure 1. In addition, since BMSR’s pricing strategy includes charging a premium over the mean ticket prices of other resorts, the mean ticket price was compared with other average ticket prices by states during the data wrangling step. Refer to Figure 2. Furthermore, the data wrangling step also included initial visualization of distribution of the feature data to see any initial correlations that we could potentially explore during the next step, Exploratory Data Analysis (EDA). Refer to Figure 3 for initial visual models.

**Figure 1: Cleaned Data during Data Wrangling** 

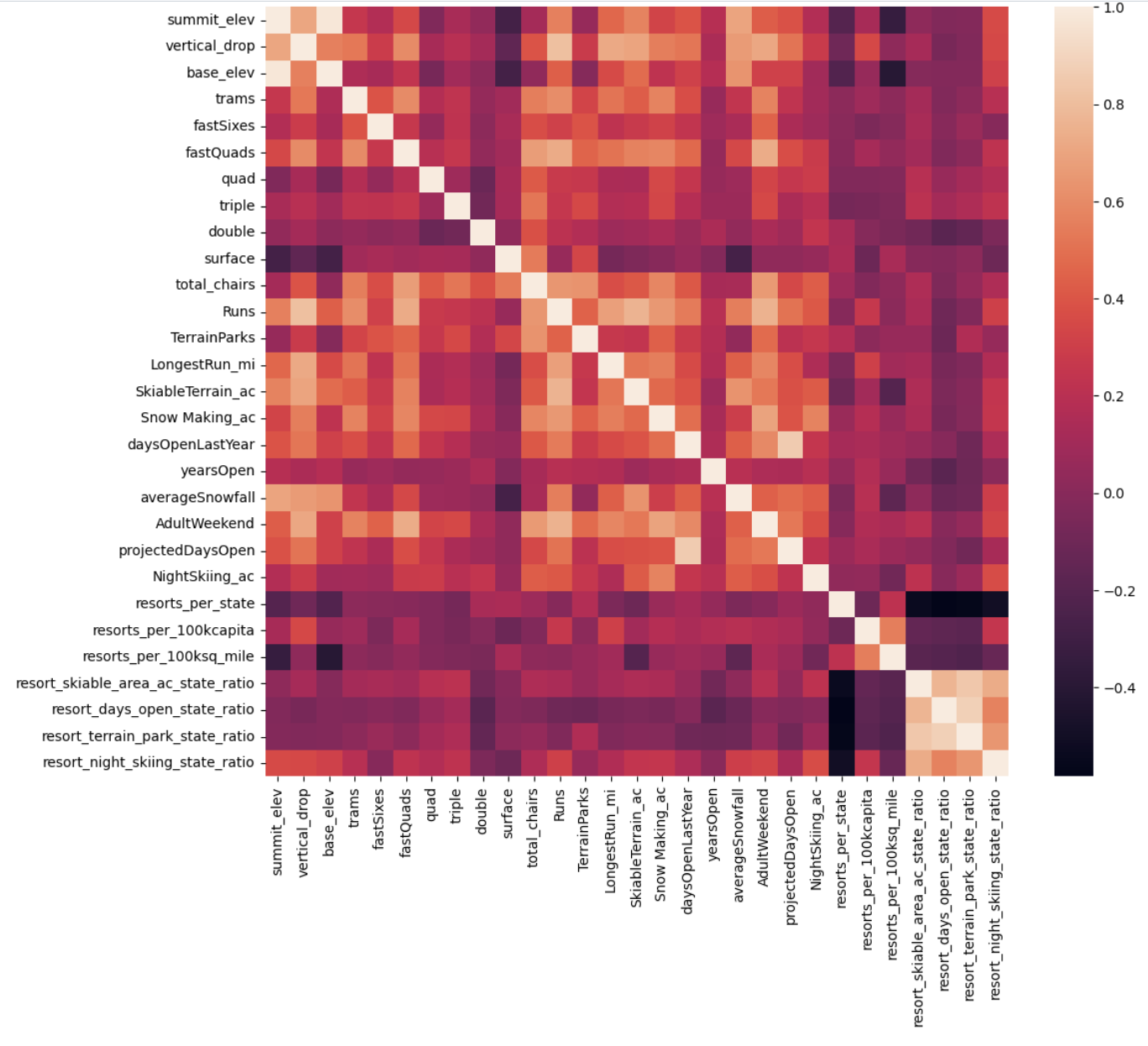
**Figure 2: Mean Resort Ticket prices by state**

**Figure 3: Initial Distribution by Features**



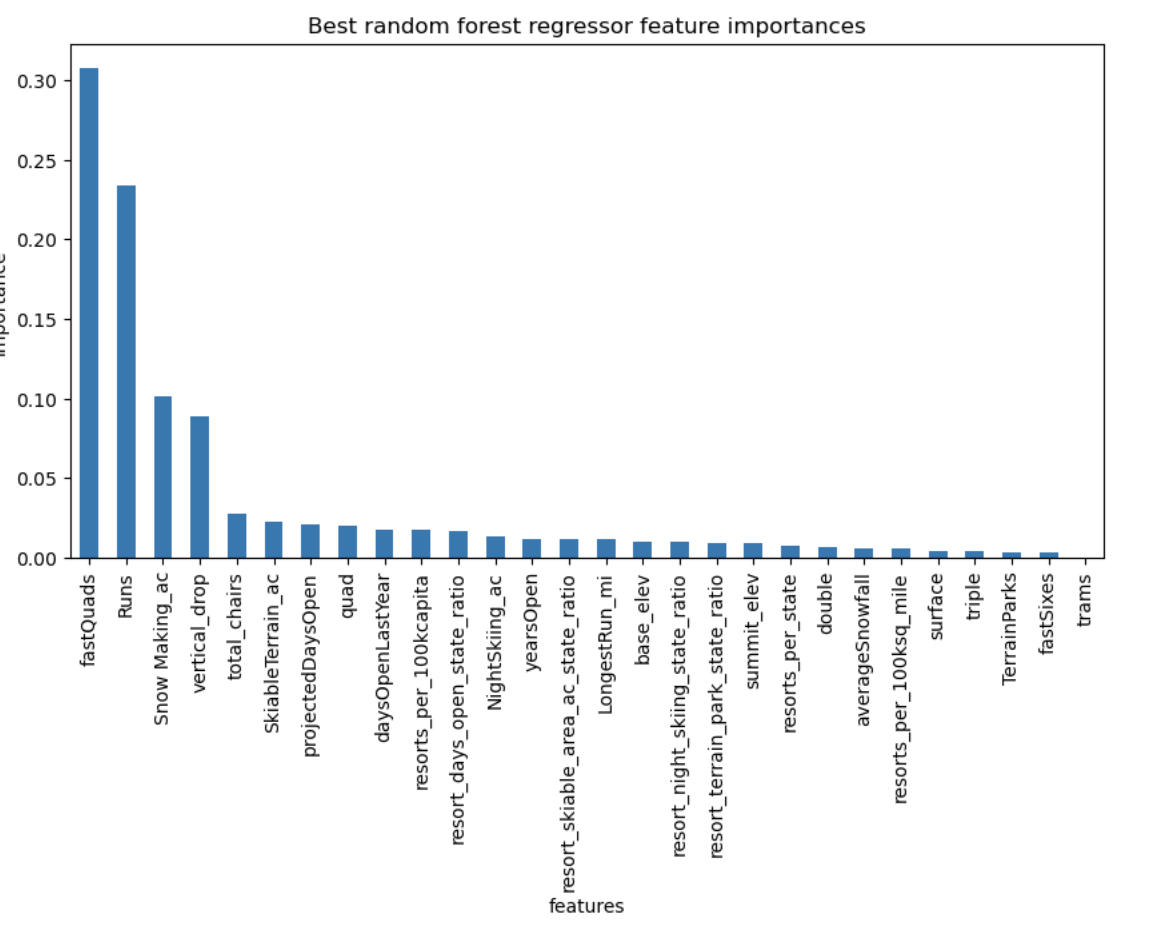
The next step in the data science assignment is Exploratory Data Analysis (EDA), in which hypotheses, observations, and relationships were identified. A primary measure during this step was to scale and perform primary component analysis (PCA). The dataframe was first scaled to narrow the dataframe and the features down to numerical values, and also since the features are heterogenous. After scaling, PCA was then performed as a measure to determine linear relationships and combinations with the original features that may be uncorrelated with one another and ordering them by amount of variance. The purpose of PCA was to ultimately to find correlation with the features and average ticket prices. Thus, a heat map was insituted to initially visualize the features with the strongest correlation to price increase. -1 being negatively correlated, 0 being not correlated, and 1 being strongly correlated. (refer to **Figure 4**). The results from the variance we observed from PCA and the heatmap allowed for initially narrowing the most important features correlated with resort ticket price increases: vertical drop, fast Quads, Runs and total chairs. These features were noted and further evaluated in the next step of pre-processing the data for generating models to train and test for most accurate predictions.

**Figure 4: Resort Feature Heatmap**



The pre-processing step and model generation step kicked off by ensuring all data had to be numeric, as any categorical data such as state and region were not needed and thus eliminated. In addition, any null values were imputed with either median and mean values for the respective features. After pre-processing, the dataframes had undergone predictive model generation, training, and testing through linear regression models. The linear regression models were constructed using pipelines that fit the training data to make the best predictions based on features. The accuracy of the models were evaluated using R squared, mean absolute error, and mean squared error calculations. In addition, the accuracy of the linear regression model from the fitted training model was also further evaluated by experimenting with different hyperparamaters. However, experimenting with different hyperparamters yielded no improved accuracy results. As a comparability method, a Random Forest Regressor model was also generated with a pipeline derived from the developed training model to see if the model could make more accurate predictions. The Random Forest Regressor model was also experimented with different hyperparamters, and as a result, the final input of hyperparamaters had yielded marginally more accurate predictions from the R squared, mean absolute error, and mean squared error calculations. The Random Forest Regressor model also ranked the most impactful features (columns) from the data frame, and the most impactful features correlated positively with price increases were fastQuads, Runs, Snow Making, vertical drop, skiable terrain, and total chairs. Refer to **Figure 5**.

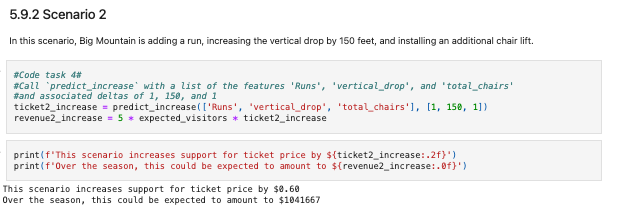
**Figure 5: Best Random Forest Regressor Features (Ranked)**



The results of the Random Forest Regressor Model provided a more narrow insight of the most impactful resort features that had the most influence to justify a resort ticket price increase. Thus, the Random Forest Regressor Model was chosen for final step for generating models that would predict the price of BMSR’s ticket price increases based on adjustments to features. The final model had shown that BMSR’s current ticket price of $81 could yield a ticket increase of up to $99, without significant impact to operating costs nor costs of investments. In addition to the final model that was generated, a custom price increase prediction function was generated that would allow BMSR to experiment with adding or removing feature values to predict both ticket prices and revenues. Refer to **Figure 6**.

In conclusion, it is recommended for BMSR to at least add 1 run, increase the vertical drop, and add a chair lift. In addition, the prediction model function also allows BMSR to add 300 ft in vertical drops and add 2 additional chairs and 2 additional runs, the prediction model would suggest that BMR could support an increase in ticket price by over five dollars with an additional revenue of 9.5 million dollars. Which covers the baseline cost of $1.5 million and generates significant more revenue. This is justified, as the final model showed that BMSR’s current ticket price of $81 could yield a ticket increase of up to $99, without significant impact to operating costs nor costs of investments.

**Figure 6: Price prediction function**

****