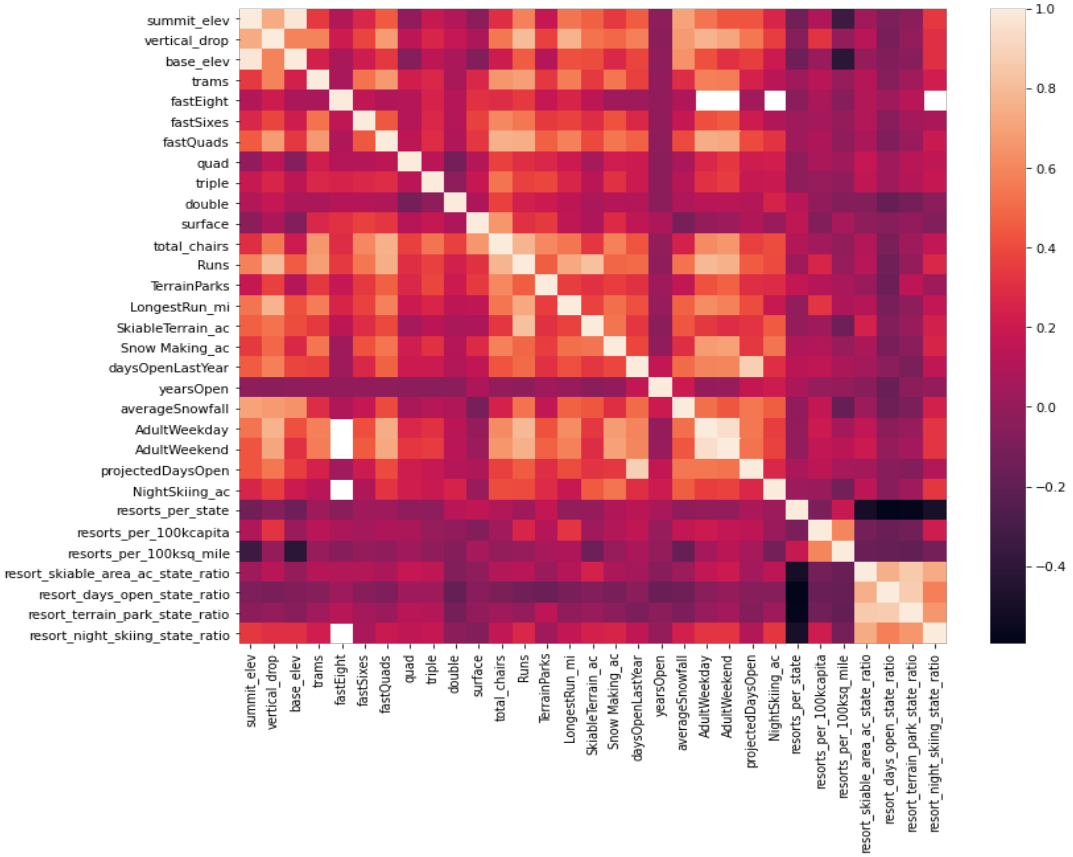
Report

Big Mountain Resort offers spectacular views of Glacier National Park and Flathead National Forest, with access to 105 trails. Every year about 350,000 people ski or snowboard at Big Mountain. Recently the Resort installed an additional chairlift that increases their operating costs by $1,540,000 this season. The Resort's pricing strategy has been to charge a premium above the resorts' average price in its market segment. Basing their pricing on just the market average does not provide the business with a good sense of how important some facilities are compared to others. Big Mountain wants to make operations cut without compromising the quality services. The Resort is looking for a model to predict the ticket price and be more competitive with other alternatives. Can the data science team create a prediction model that helps Big Mountain Resort predict ticket prices contemplating each operation's value and compare themself with the competition?

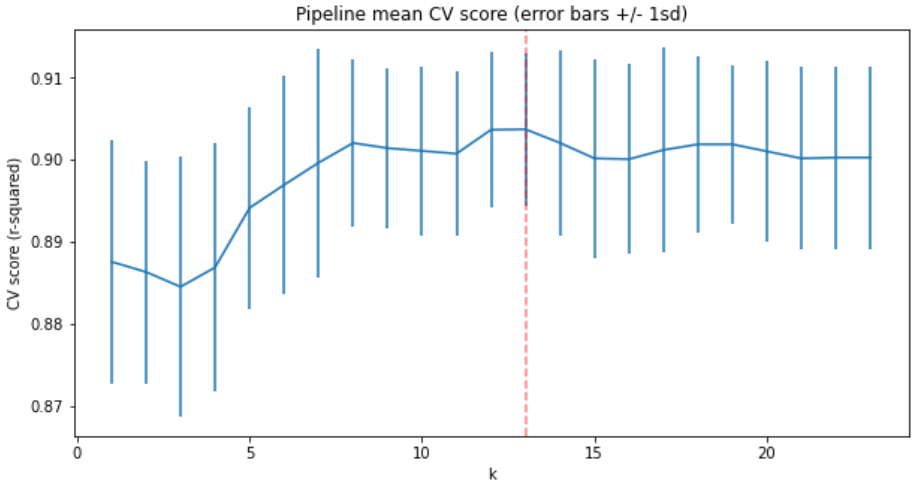
Since there are historical data, we studied the data and created a dataset with the feature to answer the principal question. The Resort provided the ski_data, and we explored the quality of historical data. We preprocessed the raw data using Python tools like pandas, seaborn, os, and matplotlibrary to remove the missing values and remove features that are not essential for creating the prediction model. The prediction model's essential features are Adultweekday and AdultWeekend; they contain the tickets prices by week and weekend. Other crucial elements are the Resort by region and state, each Resort area, the total night skiing, the amount of day open last year, etc. The mention feature will help to create an accurate prediction model and the ability to project different scenarios.

The analysis indicated that Montana was one of the more prominent Resorts but didn't figure in the most populous states. The study did not reveal any justification for treating all states equally in the early exploratory data and building a pricing model that considers all states together. Without treating anyone particularly, we created a heatmap to see the correlation between features.

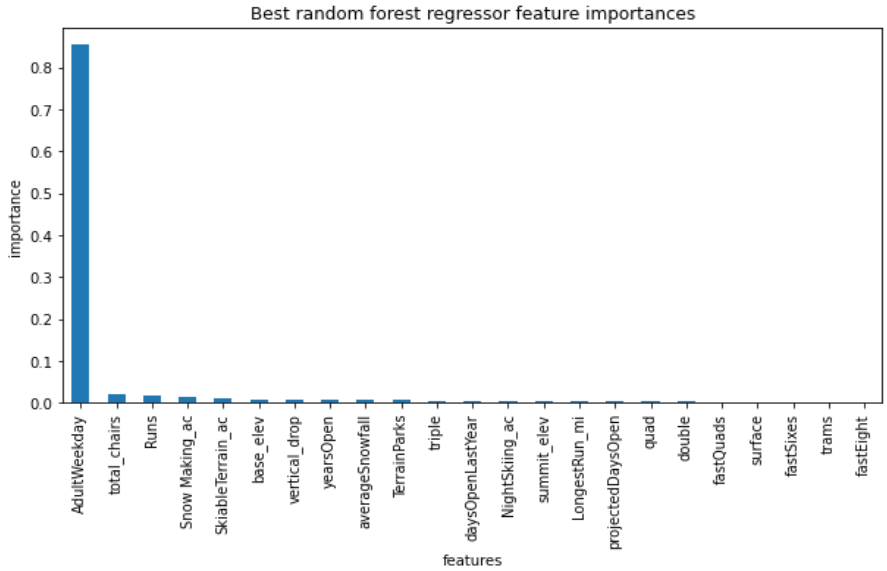


The heatmap suggest features ‘summit’ and ‘base elevation’ are positively correlated and negatively correlated with the number of resorts in each state. If any incrementation on Resort's number in a state, the share of all the other state features will drop for each. An interesting observation in this region of the heatmap is that there is some positive correlation between the ratio of night skiing area with the number of resorts per capita. In other words, it seems that when resorts are more densely located with population, more night skiing is provided.

In the preprocessing and training analysis phase we partitioned the data to a train and test split by 80% and 20% to test some model performance, compare and then choose the best model. We used the coefficient of determination to measure the proportion of variance in the dependent variable (ticket price) predicted by the "model". After running some models, the best model's mean absolute error indicates that it is expected to be off by around $9 if you guessed ticket price based on an average of known values. We used Cross-validation technique to test model performance. The cross-validation analysis helped to pick the value of k that gives the best performance.



Essential steps, like defining the pipeline and experiment with hyperparameter search using GridSearchCV are important for the creation of an accurate prediction model. After using these tools, the analysis indicates AdultWeekday, total_chairs, Runs, and Snow Making_ac is the best random forest regressor.



The random forest model has lower cross-validation mean absolute error by almost $1. It also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

The best prediction model indicates the Big Mountain Resort displayed the best price at $87.48, and the actual cost is $81.00. Even with the expected mean absolute error of $6.76, this suggests there is room for an increase. The model suggested an incrementation because Big Mountain Resort is very competitive in some areas that justified the increasing price.