# Project Report - Guided Capstone - Step 6

1. Problem Statement

Enhance Big Mountain Resort's revenue through a data-driven ticket pricing strategy, aiming to increase the profit margin by 20% by the end of the next financial year. Specifically, analyze correlations between resort features, projected operating days, and geographic locations with AdultWeekday and AdultWeekend ticket prices. Develop actionable pricing recommendations to attract visitors and maintain competitiveness.

Data Wrangling
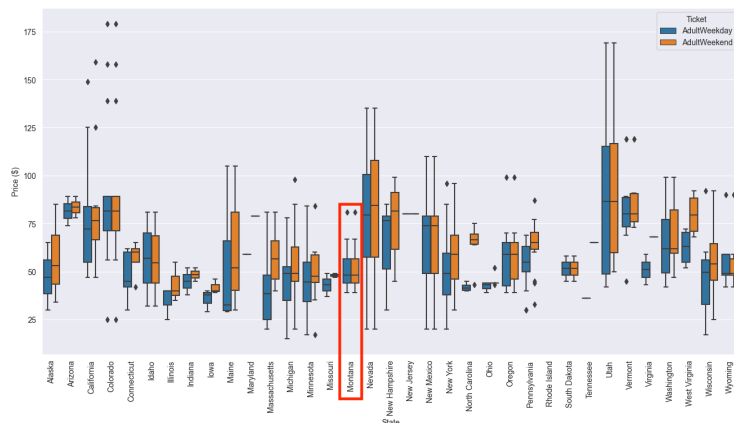
Montana state showed 1:1 ratio between AdultWeekday and AdultWeekend ticket prices. Many states with ticket prices ranged between 25-100, Montana showed little price variation (around 15 IQR) while some states like Nevada & Utah showed big price range (around 50 IQR), some States like California & Utah showed much higher price on weekends.

Exploratory data Analysis

Features where there are possible correlations with "AdultWeekend" ticket price:
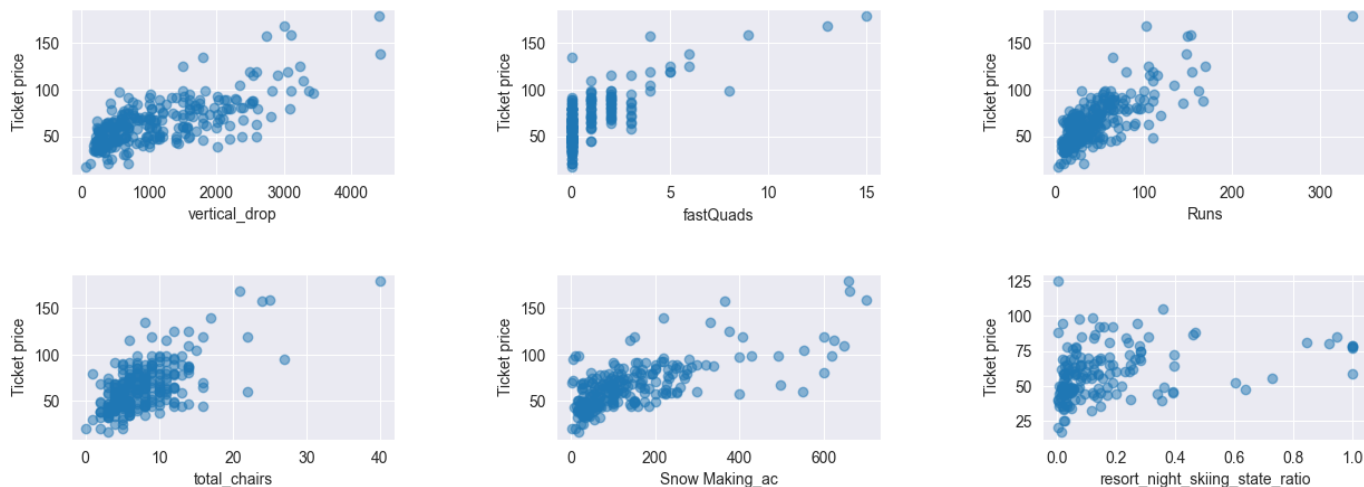
Model Preprocessing with feature engineering
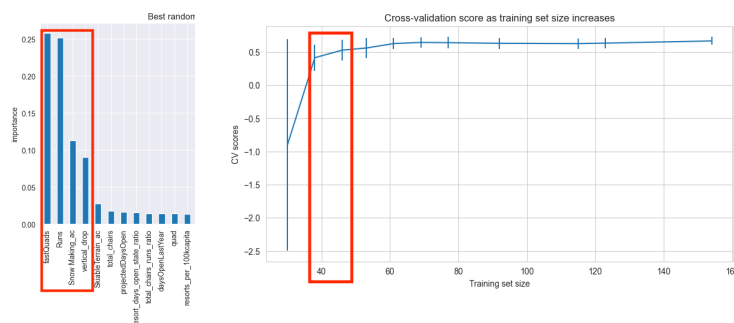
Top-4 Features that would influence the "AdultWeekend" ticket price, based on Random Forest Model with Cross Validation (cv=5), are "fastQuads", "Runs", "Snow Making_ac", and "vertical_drop". Recommended sample size: 40 - 50.

Algorithms used to build the model with evaluation metric

The first model built was linear model, using "sk_learn.pipeline.make_pipeline" method, with parameters of "SingleImputer(strategy='median'), StandardScaler(), LinearRegression()". This model seemed to be off about 9 USD (which is already much better than the 19 USD "mean absolute error" of just using the average).

A Random Forest Regressor was also tried, using "RandomForestRegressor(random_state=47)", combined with "GridSearchCV" to test the pipeline using "mean/median" and "StandardScaler/None". Using "best_params_", it appeared the best parameters are:

randomforestregressor__n_estimators=69, imputer=median, and no standard scaler. Using "best_estimator_.named_steps.randomforestregressor.feature_importances_" method, we were able to identify the most important features, such as: fastQuads, Runs, Snow Making_ac, and vertical_drop.

Winning model and scenario modeling

After using "cross_validate" methods to compare the performance of Linear Regression vs Random Forest, it turned out that **Random Forest was chosen** because of **lowest Mean Absolute Error (9.5 USD)**, it also exhibits less variability.

Pricing Recommendation & Conclusion

Big Mountain Resort currently charges 81.00 USD for AdultWeekend tickets. Model's suggested ticket price came to 95.87 USD with 10.39 USD Mean Absolute Error (which still shows room for a ticket price increase). The model's suggested ticket price was based on the facilities offered compared to the marketplace.
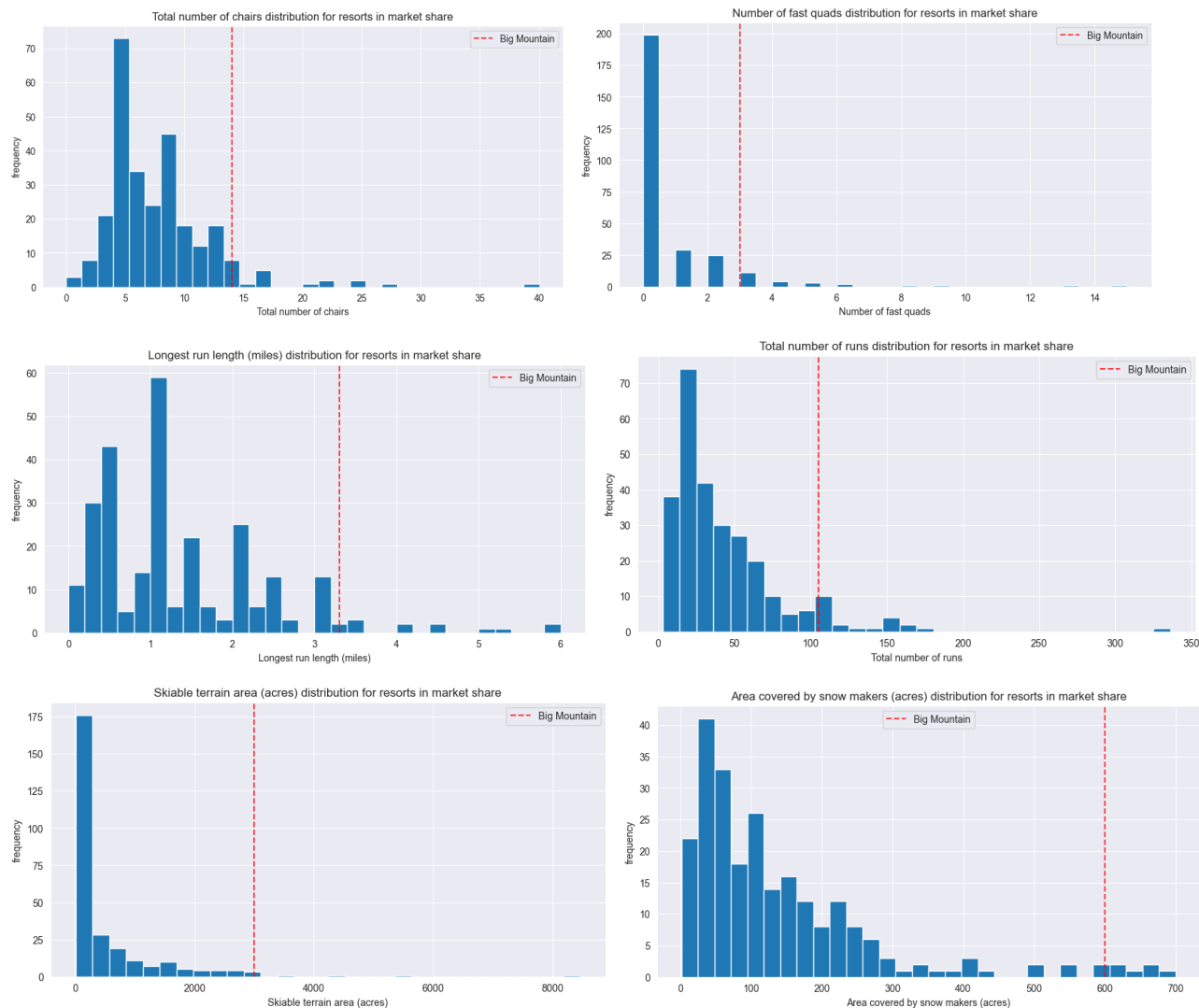
Compared to all resorts in US, Big Mountain Resorts seemed to have competitive advantage in #1 Areas covered by snow makers (acres), #2 Total number of chairs, #3 Number of Fast Quads, #4 Total number of runs, #5 Longest run length (miles), and #6 Skiable terrain area (acres).

Future scope of work

There are additional data points that can further improve the model, such as incorporating operating cost data, number of visitors across the US, and any data that could indicate customer's willingness in paying premium price for more features.

Another suggestion is to include cross-functional business experts to test the model under different scenarios. To make the model accessible, we could deploy the model in API where user can easily call them with the features that they want to test.