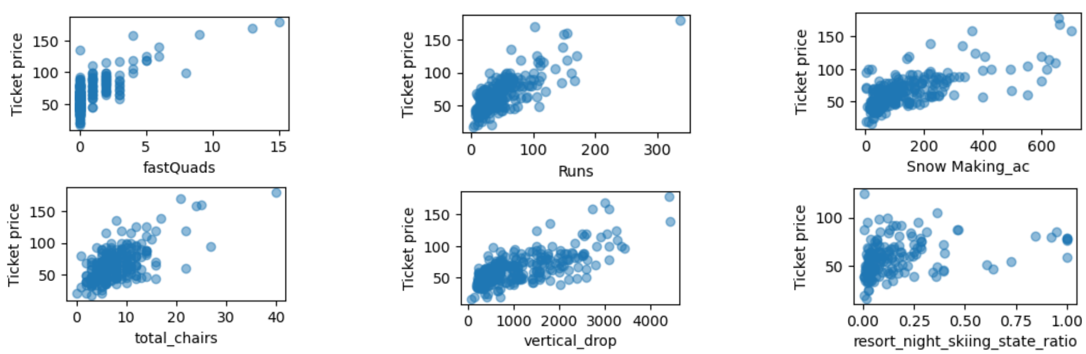
Big Mountain Resort – Pricing Strategy Data Science Project

Big Mountain Resort wanted to determine the right ticket price that is aligned with its facilities and covers its operating costs. To achieve this goal, we constructed a model to determine a suitable ticket price based on the resort's amenities.

The dataset utilized in the model includes the names and locations of 330 ski resorts in the United States, along with details of their offerings to customers. We initiated the data wrangling process, which involves converting raw data into a usable format. During this phase, we examined missing and duplicated values to ensure the data's reliability. We then proceeded to analyze the data distribution and removed rows and columns that were incomplete or irrelevant for our analysis. Additionally, we gathered data about the states' population and area from the internet, this helped us to build state competition features. When selecting the target variable, we had to choose between weekday and weekend ticket prices. We opted for the weekend prices due to their lower frequency of missing values and similar distribution to the weekday prices.

An exploratory data analysis was conducted to investigate whether the state impacts the ticket price. To assess this, we performed Principal Component Analysis (PCA) and determined that all states could be treated equally since they do not significantly influence the ticket price. Furthermore, we examined relationships between features using a correlation matrix and scatterplots. Our analysis revealed several noteworthy relationships between features, particularly with our target variable. Features such as fast quads, runs, snowmaking area, total chairs, vertical drop, and resort night skiing state ratio demonstrated correlations with the ticket price.

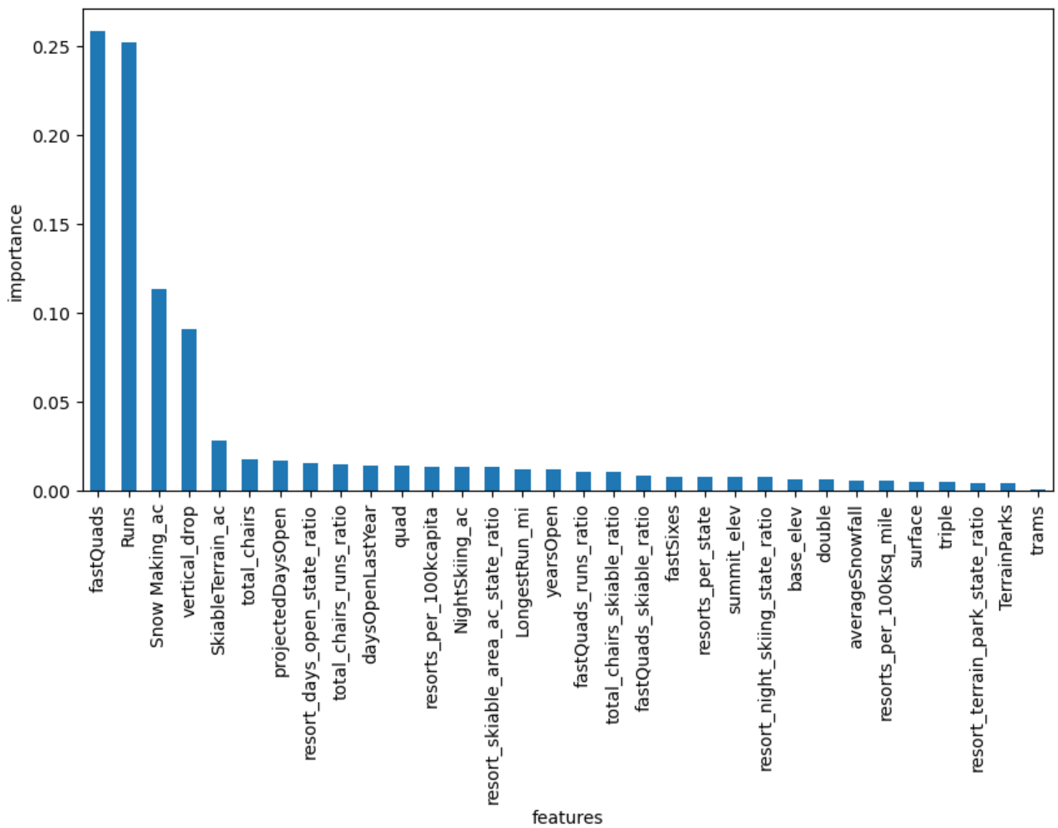
Scatterplots of the features more related to the Ticket Price



With our features prepared, we commenced by dividing the data into training and testing sets. Subsequently, we were poised to construct and evaluate various models. Initially, we employed the mean of the known values as a predictor, which exhibited poor performance (with an expected deviation of approximately $19). In response, we explored alternative models: Linear Regression and Random Forest Regressor. However, before training the models with the data, preprocessing steps were essential. To streamline this process, we utilized pipelines, comprising an imputation of missing values, data scaling, and an advantageous scikit-learn algorithm called SelectKBest. This algorithm selects the most pertinent features for the model, thereby mitigating the risk of overfitting resulting from collinearity among variables. Additionally, Grid Search proved invaluable in determining the optimal parameters for our preprocessing steps and model training. To evaluate our models' performance, we employed both the test split and cross-validation technique, which aids in enhancing the model's performance on unseen data. The evaluation metrics utilized included R^2 scores, Mean Squared Error (MSE), and Mean Absolute Error (MAE). Ultimately, the Random Forest

Regressor outperformed the linear model, yielding a lower MAE of $9.54 compared to $11.79 for the linear model. Then, we designated the Random Forest Regressor as our chosen model for price determination. Scikit -learn provides a valuable tool for assessing feature importance, revealing similar features to those identified as crucial in our Exploratory Data Analysis.

Best Random Forest Regressor Feature Importances



Once we finalized our selected model with its tuned parameters, we proceeded to train it using the entire dataset, excluding the data for Big Mountain Resort. The model suggested a price of $95.87, with a mean absolute error (MAE) of $10.39, despite this, there is still room for a potential increase in price. Given that Big Mountain Resort boasts several significant facilities, an increase in price can be justified.

Furthermore, our analysis revealed some interesting insights regarding the impact of various facility enhancements on ticket price. For instance, small changes such as adding 4 acres of snowmaking or increasing the longest run by 0.2 miles did not significantly influence support for a price increase. However, closing runs had a discernible effect on ticket price support, with the closure of multiple runs resulting in incremental reductions in support.

Moving forward, it would be beneficial to obtain data on daily visitor numbers, additional operating costs, and competitor prices. If deemed useful by the leadership, the model could be integrated into an application, allowing business stakeholders to utilize it at their convenience and explore different pricing scenarios. This comprehensive approach to data analysis and pricing strategy provides valuable insights for decision-making and future planning.