SPRINGBOARD GUIDED CAPSTONE PROJECT

Big Mountain Resort

Presented by Andrew Vo

Problem Identification

A BIG SKI PROBLEM

Big Mountain Resort plans on adding an additional chair lift and wants insights into its pricing and operational strategy in order to maintain profit margins in light of the additional operational costs.

We leverage data from over 300+ ski resorts to arrive at a pricing model that can provide insights for the executive team.

Recommendations and Key Findings

INCREASE TICKET PRICE

We recommend Big Mountain increase ticket prices to about $94, representing an 18% premium over current price.

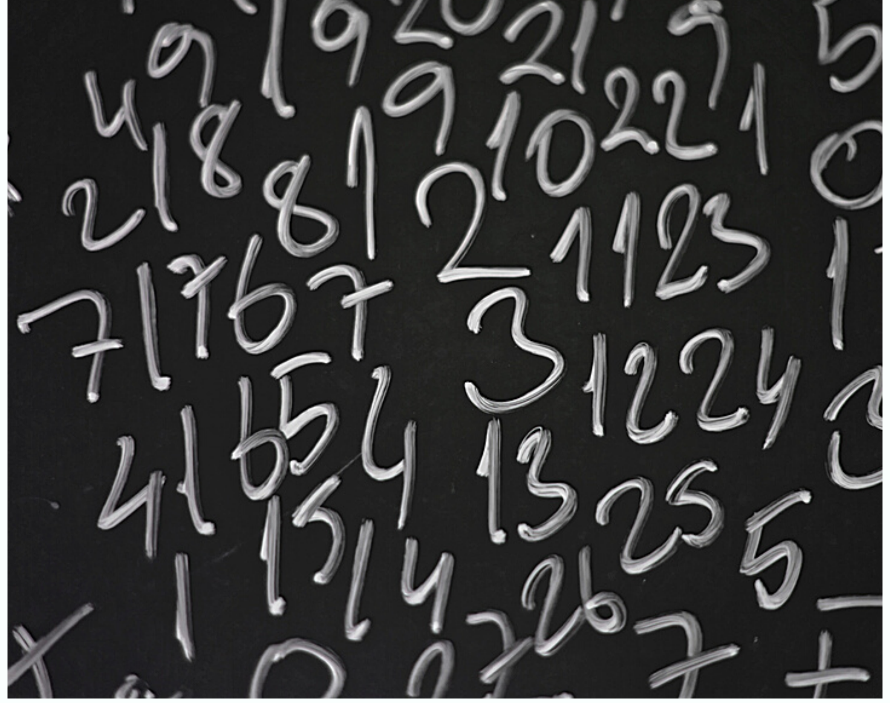
CLOSE ONE RUN

We recommend Big Mountain close one run, which will reduce operational expense with minimal revenue impact.

ADD. PRICE INCREASE

With the additional chair lift, we recommend Big Mountain increase ticket price another $2 to $96 to be in line with market rates.

Our Data Science Process



DATA WRANGLING

Reviewing and cleaning up the data from 300+ ski resorts.

EXPLORATORY DATA ANALYSIS

Reviewing correlations and identifying principal components in the dataset.

TRAINING AND MODELING

Training, testing and reviewing a linear model and a Random Forest.

Data Wrangling

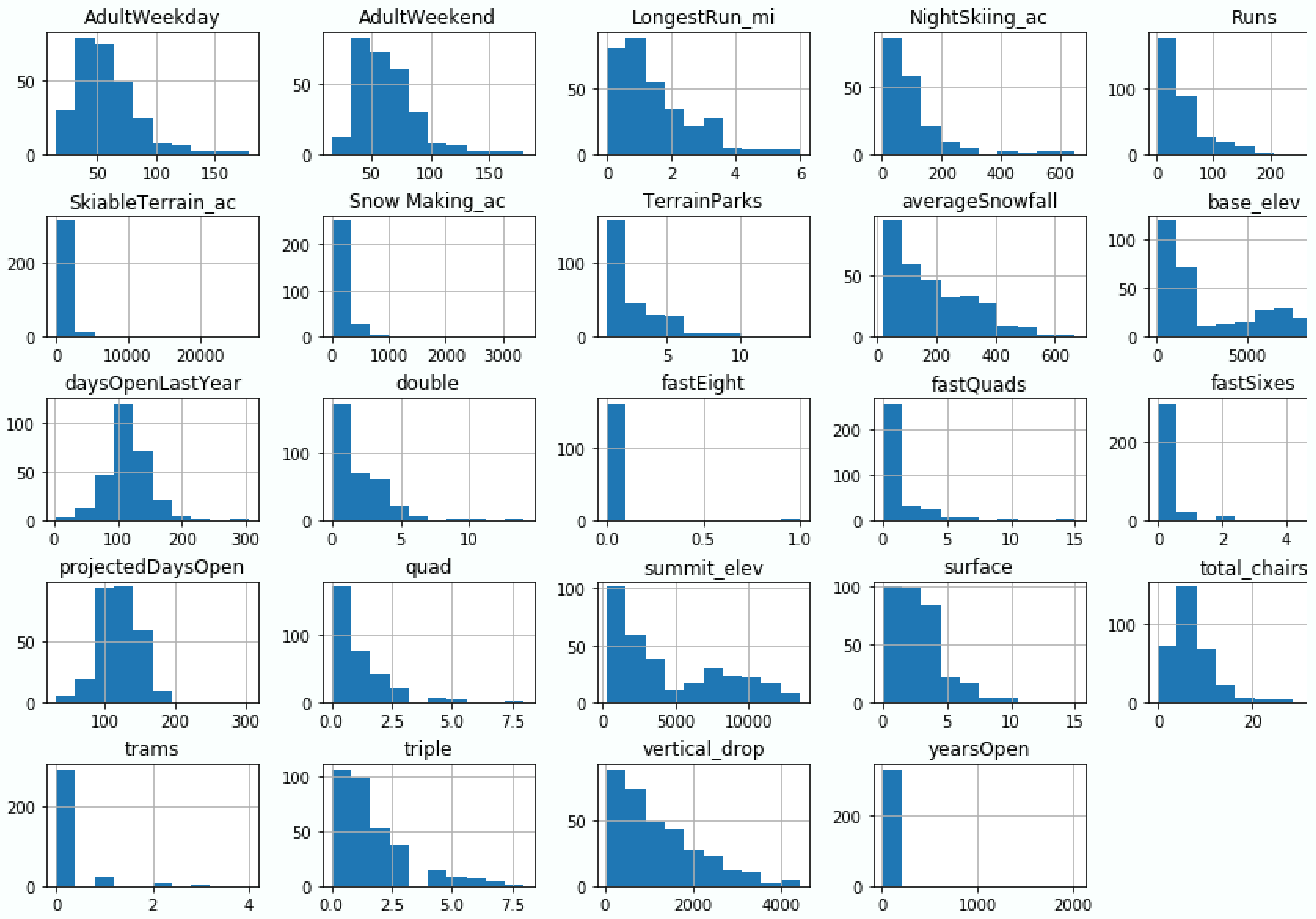
Principal anomalies in the dataset include:

1. Missing values like number of Fast Eight lifts.
2. Over 14% of observations no ticket prices.
3. Several outliers in Skiable Area and Years Opened

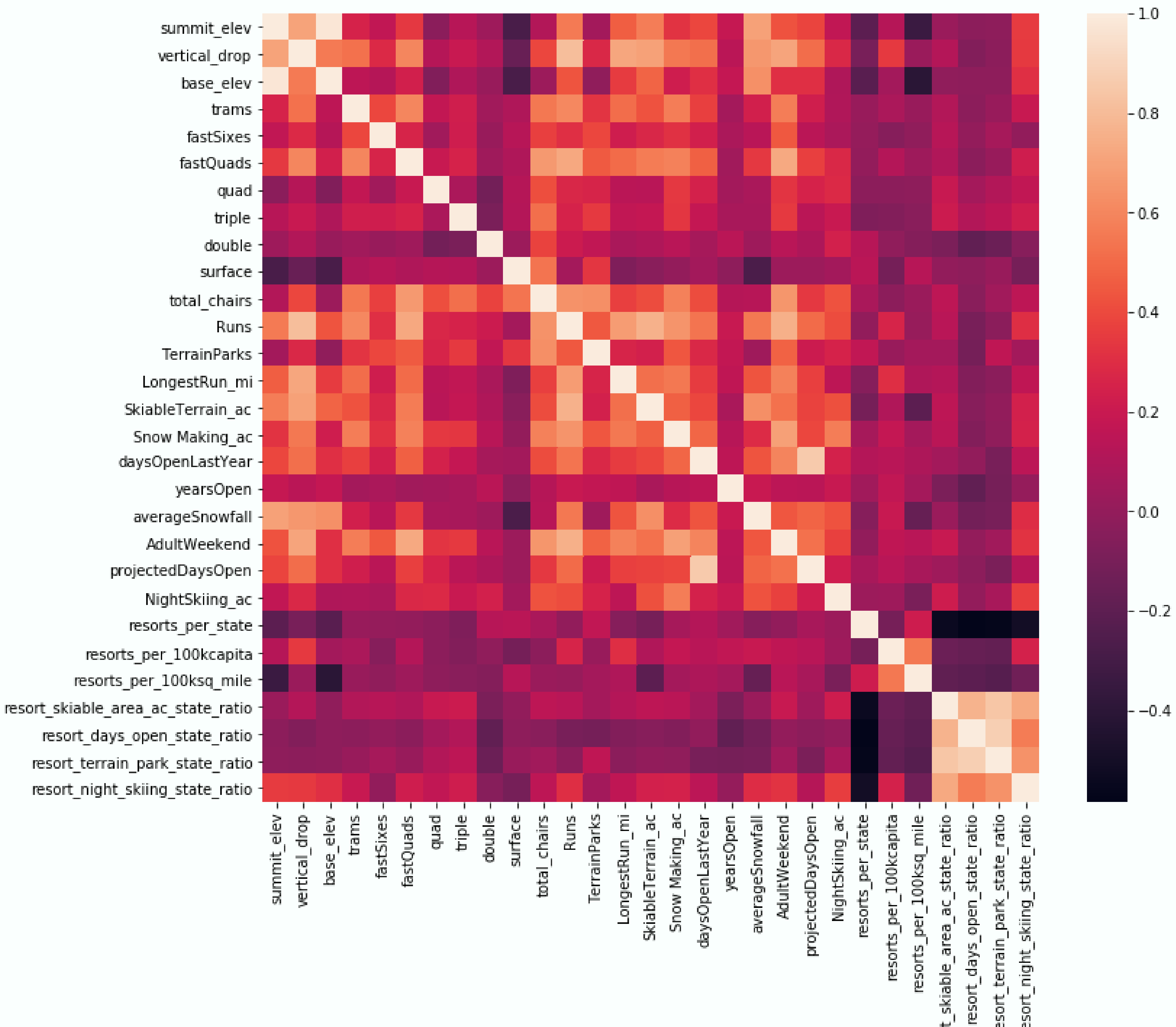
We remove these bad data points in order to prevent our model from being poorly trained and hence the prediction being less meaningful.

We also checked that each observation was unique to prevent multi-collinearity.

Overall, the data is not as exhaustive as we wanted, but we did clean it up to a state were it was usable for our predictive analysis.



Exploratory Data Analysis



In this step, we leverage feature engineering using Principle Component Analysis to find novel features within the data for training.

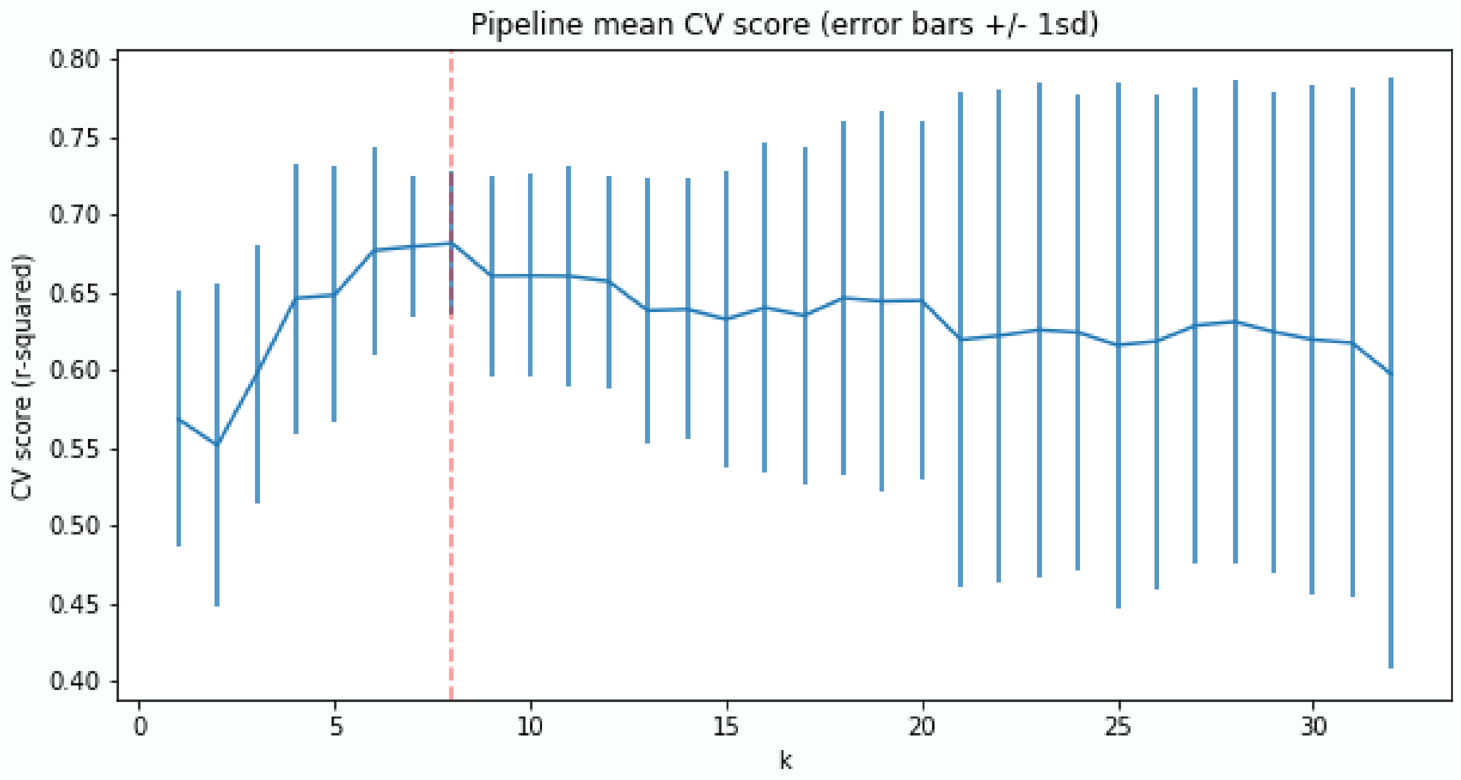
Additionally, we explored the correlations among our features in order to reveal any multi-collinearity in the dataset that may cause over-fitting of the model.

Aside from some higher correlations found in our synthetic features found in the bottom right of the heat map, we're satisfied that many of our features have low correlations.

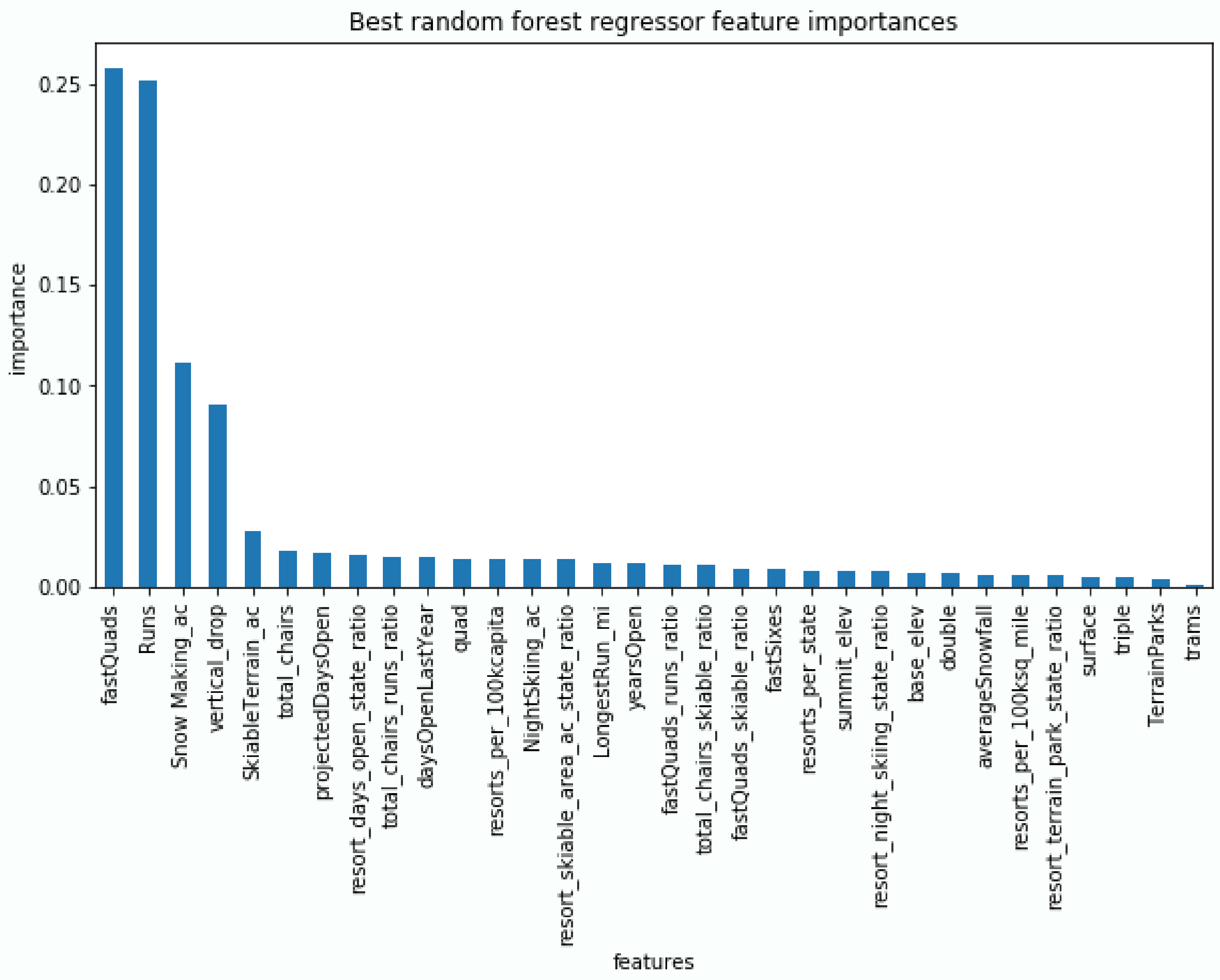
Training and Modeling

This allowed us to quickly evaluate (using error/variance metrics) that the Random Forest Regressor was the most optimal model for our dataset.

Model Prediction: The output of the model (development highlighted herein) recommended a price of $94 (18% premium) and an additional $2 for an additional chair lift + vertical drop increase of 150 ft.



We developed a training and modeling pipeline to quickly evaluate two models (a linear regression and a Random Forest) based on cross-validation datasets.



Pictured above is the sorted importance of features with the top Four being the number of Fast Quad lifts, runs, Snow making area, and vertical drop.

Conclusion

FINAL RECOMMENDATION

Increase ticket price (18-20% premium), close one run.

DISCLOSURE

Does not consider state-specific effects in Montana or operational costs associated with new chair lift.

FURTHER WORK

Adding new features or collecting more data, honing into resorts similar to Big Mountain Resort, adjusting for state-level effects, using more advanced machine learning models like neural networks.