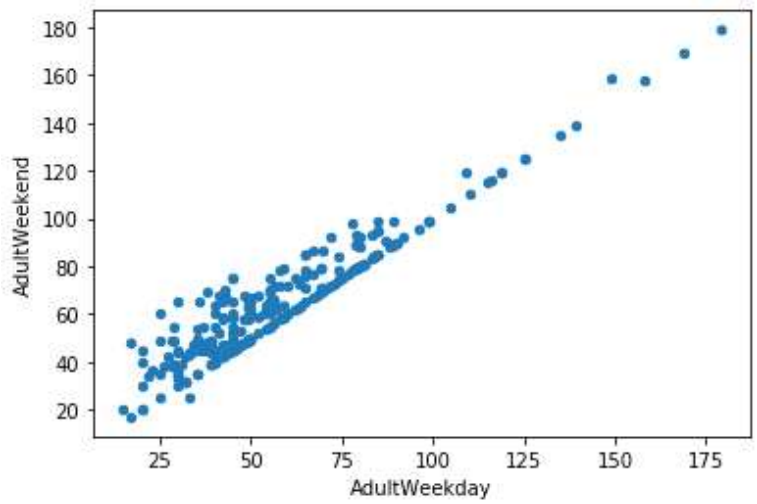
I. Introduction

This project is about 'Big Mountain Ski resort' located in Montana. The business wants some guidence on how to select a better value for their ticket price. Management is also considering a number of changes that they hope either cut costs without undermining the ticket price or which support even higher ticket price. Big mountain resort also wants to capitalize on its facilities as much as it could.
The primary stakeholder is Director of Operations Jimmy Blackburn. Another point of contact provided is database manager Alesha Eisen

II. Data Wrangling

In this step I have cleaned data, dropped rows with no price data, dropped column such as "fastEight" which contains too many missing values. Removed non state values such as Massachusetts, Pennsylvania, Rhode Island and Virginia. Converted dataset into desired format. There are two price values weekday and weekend. To decide what will be the target for modeling the price used a scatter plot which describes the relation between weekday and weekend price as follow



there is a clear line where weekday and weekend prices are equal, and there are more missing prices in weekday than weekend. So dropped the weekday prices and kept only weekend prices. And target would be the pricing model for weekend prices.

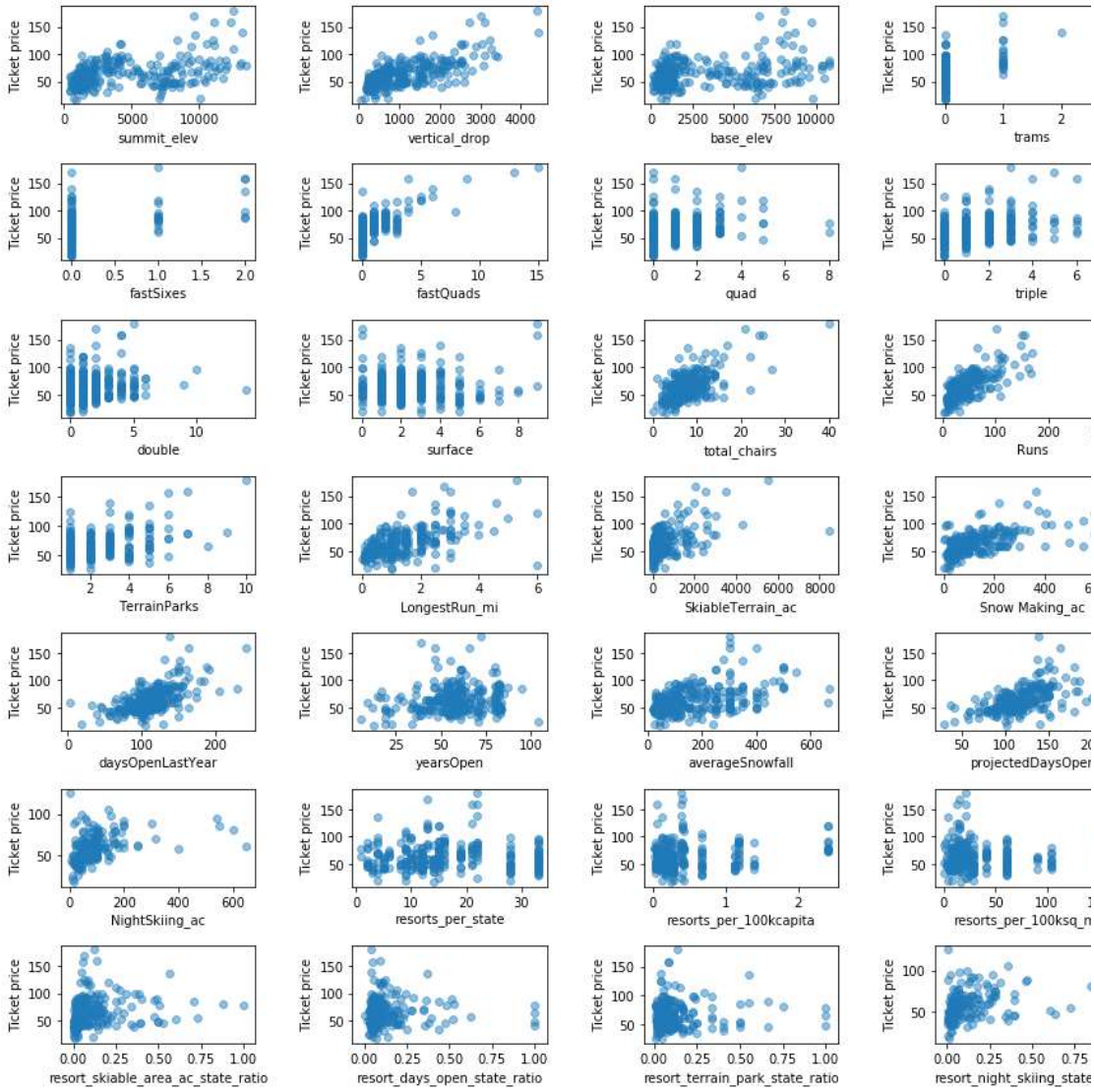
III. Exploratory Data Anaysis

This step is used to find patterns/trends in given data. To check the state wide picture of market I explored top states by order of each of summary statistics. In total state area Montana comes first. California has higest state population, New York comes first in the number of resorts, Colorado has highest skiable area.

There is no clear statewise pricing trends found in the data so decided to consider all states

equally. For further exploration I considered only numeric value features.

To test ticket price against all numeric features used scatterplot as follows.



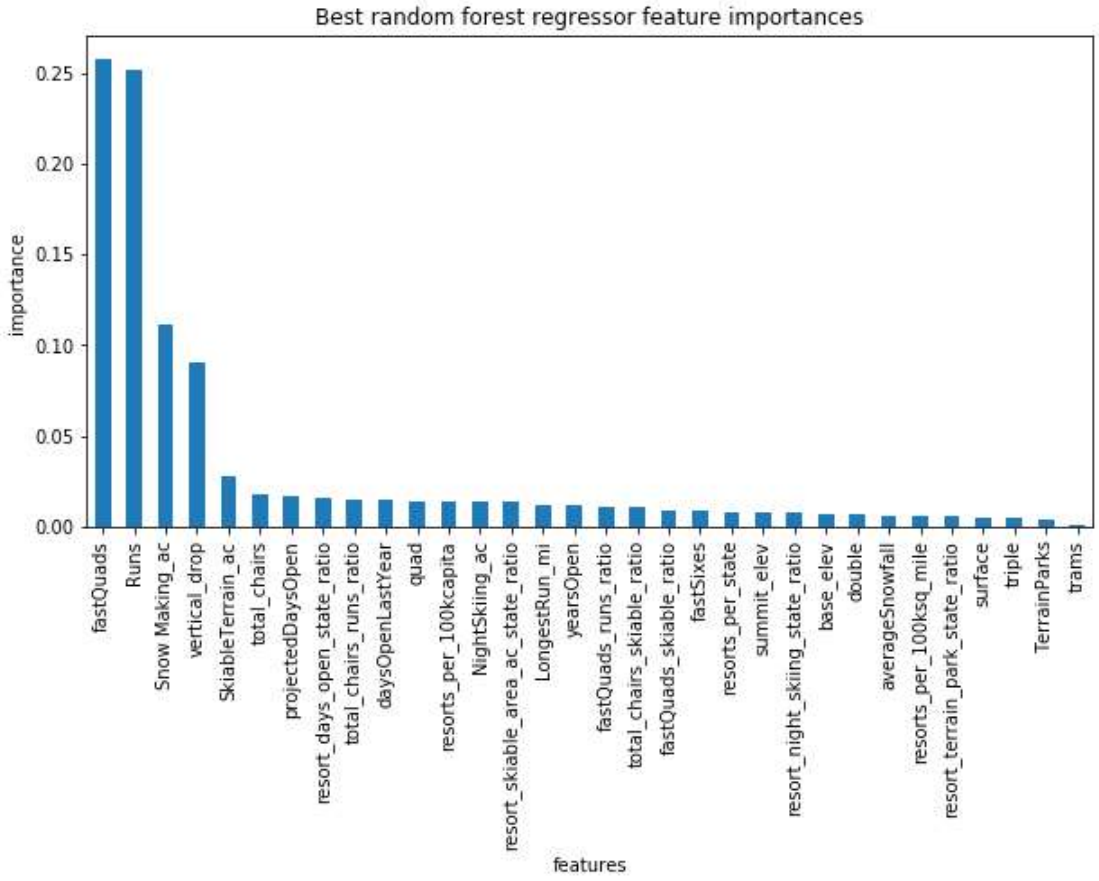
There is a clear positive correlation with some features such as vertical_drop , total_chairs, runs, fastQuads which is usefull.

IV. Pre-Processing and Training Data

In this step I divided the data into two parts to train/test split and used machine learning algorithm to train the data and then test it on test data.

Using the linear regression pipeline using median strategy over the mean is more usefull and shows few features has more positive results for ex. vertical_drop, snow_making_ac, total chairs etc. same as we explored in exploratory data analysis step.

Calculated best random forest regressor feature importances and showed in bar plot as follows



Random forest regressor and linear regressor model has following features in common

- fastQuads
- Runs
- Snow Making_ac
- vertical_drop

The random forest model has a 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.

V. Modelling

We calculated expected Big Mountain Resort ticket price from the model.

Big Mountain Resort modelled price is **$94.22**, actual price is **$81.00**.

Expected mean absolute error is **$10.39** there is room for increase.

Features that came up as important in the modeling

- vertical_drop
- Snow Making_ac
- total_chairs
- fastQuads
- Runs
- LongestRun_mi
- trams
- SkiableTerrain_ac

Big Mountain Resort has been reviewing potential scenarios for either cutting costs or increasing revenue (from ticket prices).

Scenario

1. Permanently closing down 10 least used runs

- $80.25 (Note: Model shows that closing down 10 runs will impact on the revenue significantly)

2. Adding a run, Increase vertical drop by 150ft and install a chair lift

- Increase price by $1.99 (Revenue increase $3.47M / season)

3. Adding a run, Increase vertical drop by 150ft, install a chair lift and add snow making area

- Increase price by $1.99 (Revenue increase $3.47M / season, no difference observed due to snow making area increase)

4. Increase the longest run by .2 miles

- No change to price

VI. Conclusion

Big Mountain Resort currently charges $81.00 ticket price. And our model suggests $94.22 in marketplace. however ehe validity of our model lies in the assumption that other resorts accurately set their prices according to what the market (the ticket-buying public) supports.

The fact that our resort seems to be charging that much less that what's predicted suggests our resort might be undercharging. But if ours is mispricing itself, are others? It's reasonable to expect that some resorts will be "overpriced" and some "underpriced." Or if resorts are pretty good at pricing strategies, it could be that our model is simply lacking some key data? Certainly we know nothing about operating costs, for example, and they would surely help.

Adding a run for 150 ft. and an additional chair supports $1.99 price increase and increase in revenue amount to $3474638.