Guided Capstone Project

Big mountain Resort Ticket Price Prediction

Li Li

December 2021

Springboard

Problem Statement and Objective

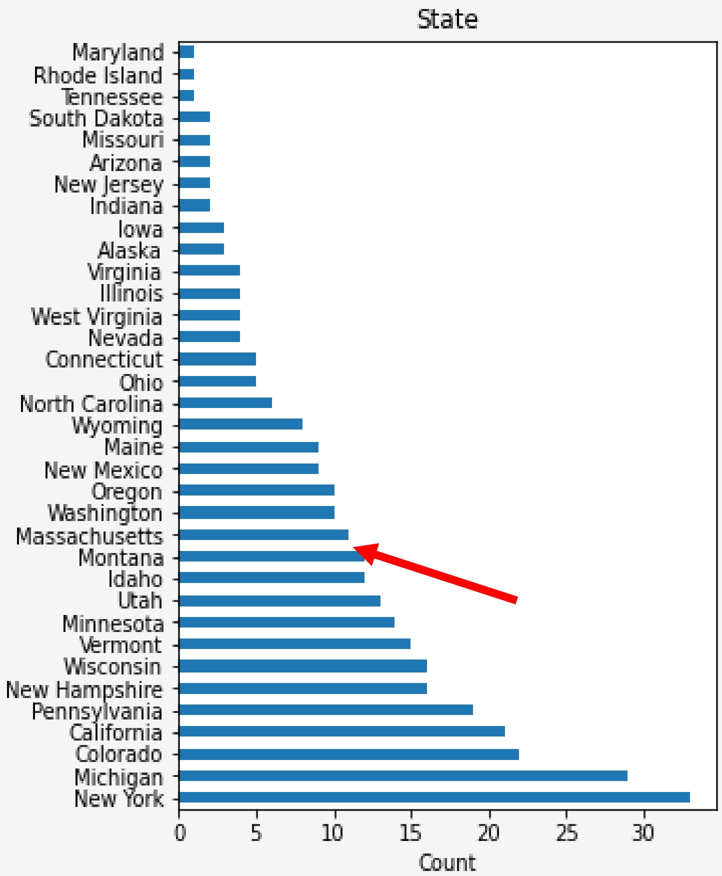
➢ Problem Statement:

- Big Mountain Resort, a ski resort located in Montana, has 105 trails, 11 lifts, 2T-bars and 1 magic carpet for skiers of all levels.
- It has about 350,000 visitors per year.
- An additional lift chair has been installed with operating cost of 1.5M this season.

➢ Objective: The business wants to have a guidance to select a better pricing strategy to increase revenue by:

- identifying which facilities more matters with higher capitalization rate and increasing the ticket price?
- or by reducing the operational cost while maintaining the ticket price?

Data Information

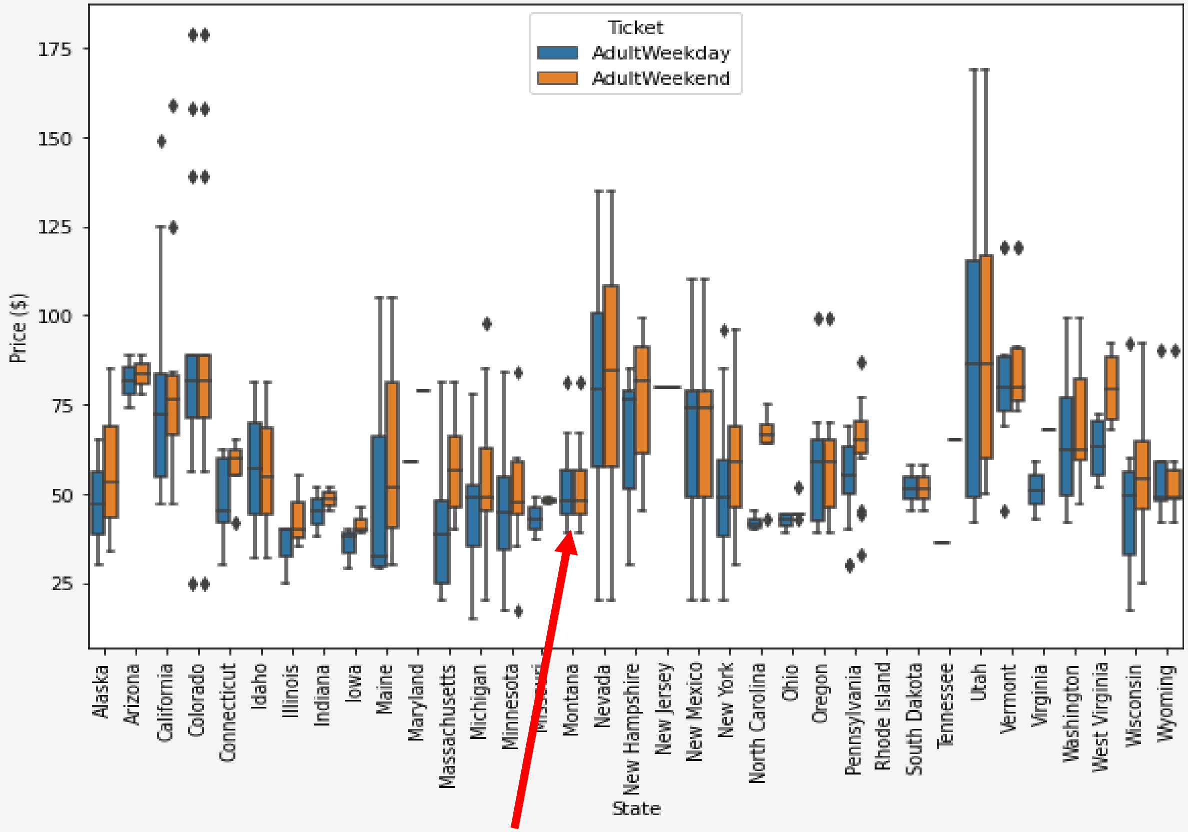


- ➢ Raw Data (330 rows and 27 columns)
- All resorts in 35 states belong to the same market
- Big Mountain resort is 1 of 12 in Montana

- ➢ Data Preparation (cleaning and transformation)
- Add statistic summary columns
- Imputing the missing data
- Explore the relationship between features
- Correct the errors
- Drop rows with no price data

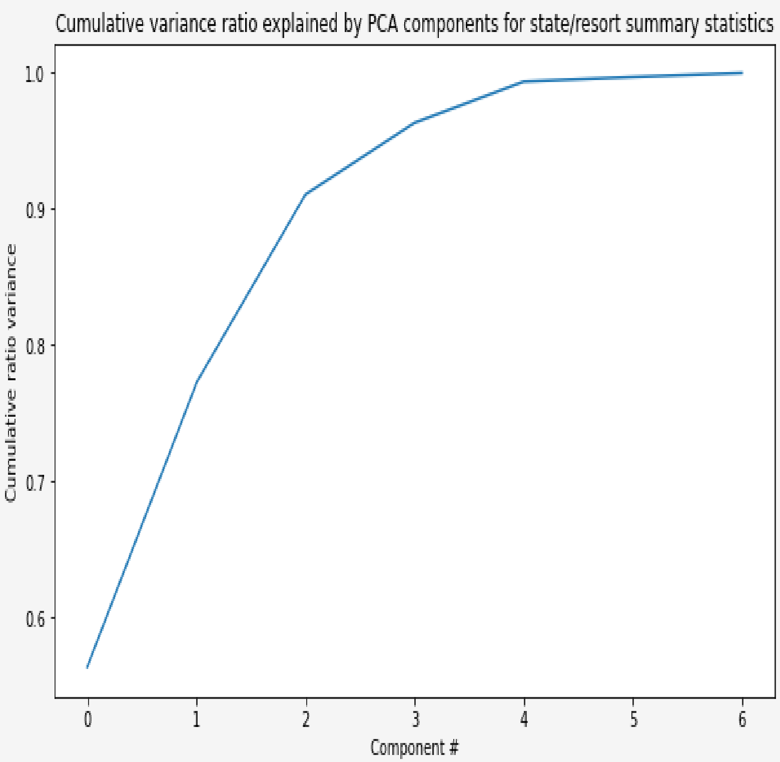
- ➢ Final Data (276 row and 36 columns)

Target Feature

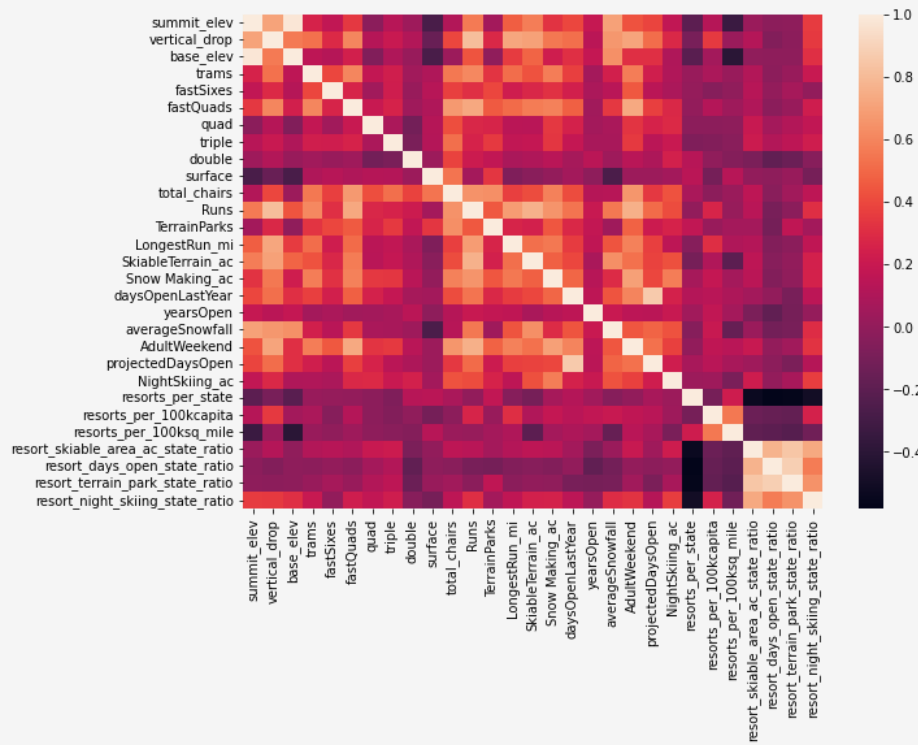


- 2 kinds of ticket prices: Weekend & Weekday and missing about 15~16% in data.
- The distribution of two prices appear equal in Montana.
- Weekend prices have the less missing data and been used as the target feature for the model.

Key Features



- The correlation heatmap below shows how the features correlated to ticket prices for each state:
- Identify key features: vertical drop and the area covered by snow making equipment, runs, total chairs, snow Making_ac, and fastQuads.



- Implement PCA to perform dimension reduction.
- The first 2 principal components of 7 account for 77% of total variance.

Linear Regression Model and Parameters Optimization

- All features (32 columns)

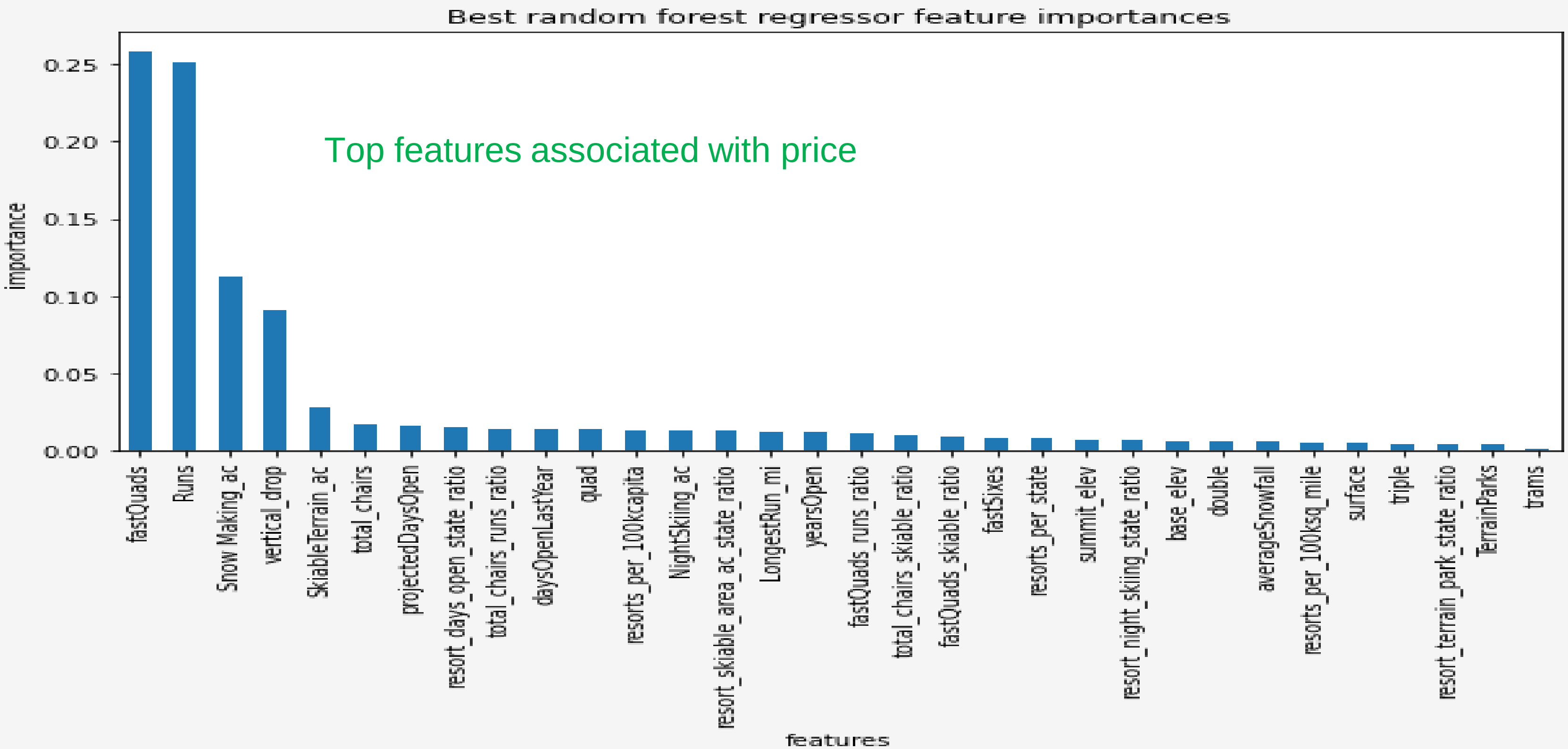
- Assess the model using cross validation:
- R2: 0.66
- Std of scores: 0.064
- mae of the precited: $11.79

- Optimize k value as 8 and determine the significant features contributing to ticket price:
- Positively associated with prices:
- Vertical drops: 10.77
- Snow_Making_ac: 6.29
- fastQuads: 5.75
- Runs: 5.37
- LongestRun_mi; 0.18
- Trams: -4.14
- Skiable_Terrain_ac: -5.25

Note: First five features are positively associated with price, the last two negatively.

Random Forest Model

- Assess the model using cross validation:
- R2: 0.70
- Std of scores: 0.064
- MAE: $10.4 (lower than $11.79 of linear regression model)
- The predicted price for Big Mountain: $95 compared with the original price $81



Data Quality Assessment

- Plenty of data in the training
- Based on model cross-validation scores, an initial rapid improvement is observed at 40~50% of training data size

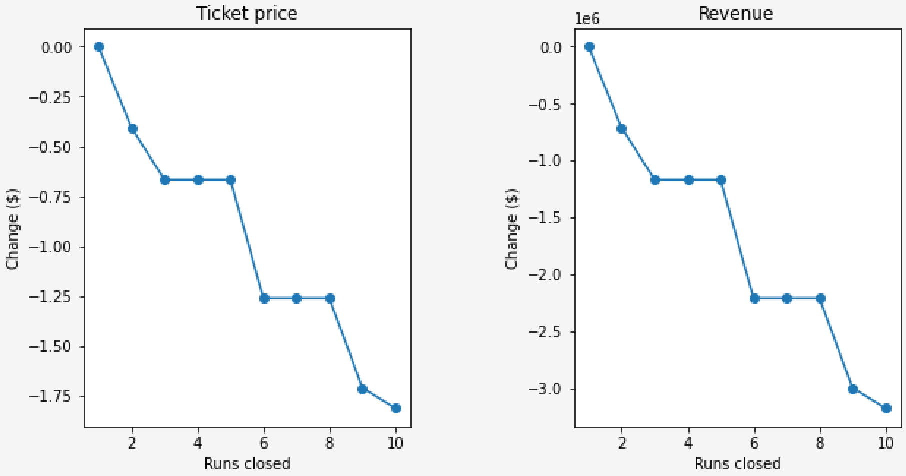
Scenarios Tests based on Various Features Selection

- ➢ #1: close one run is ok
- with no affect on price or revenue (as shown in the plot).

- ➢ #2: adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift.
- ticket price would be increased by $1.99 and revenue by $3,474,638.

- ➢ #3: repeating scenario#2 but adding 2 aces of snow making
- make no more difference on ticket price and revenue

- ➢ #4: increasing the longest run by 0.2 mile and adding 4 acres of snow making capability
- Make no difference on the model result



Conclusion//further work

- ➢ The model predicts that the ticket price for Big Mountain resort could be raised to $95 compared with the current $81.

- ➢ Based on scenario test, it suggests that Big Mountain could expect to have a revenue increase of $3,474,638 (if the ticket price is raised by 1.99 USD, and expected visitors are 350,000 people over the season, and on average, visitors ski for 5 days.)

- ➢ There are a good number of features included in the model. The model explains about 62%~77% of variance with all the features. It indicates that some other features influencing the price might be missing in this dataset.

- ➢ Additional data may need be sourced and considered, such as:

- Visitors data per year (volume of weekend vs. weekdays)
- Customer data (demographics, new vs. regular)
- Operational costs for facilities

- ➢ This random forest model is providing at best a baseline starting point. More combinations of features can be tested to refine the model and then to make more accurate prediction.