Big Mountain Resort Data Science

Introduction

- What opportunities exist for Big Mountain to recoup the increased operational cost of $1.54 million over the next year through modification of pricing or optimization of importance of their facilities?
- Big Mountain Resort currently charges a premium weekend pass of $81

Objective

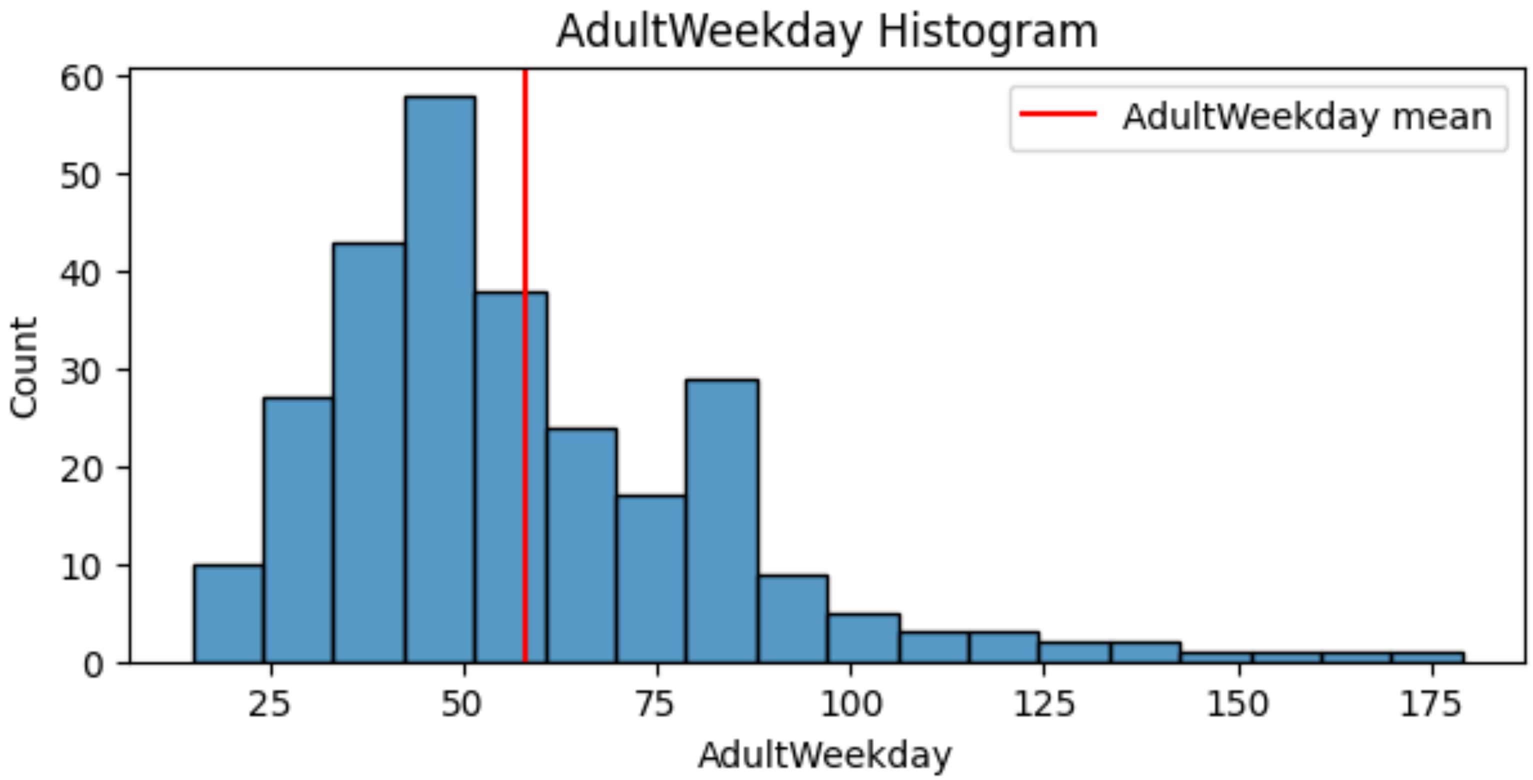
- Create a machine learning model to predict the price Big Mountain should charge depending on their place in the market

Data

- 26 descriptive features of 328 unique ski resorts
  - Name, region, state, summit elevation, vertical drop, trams, fast sixes, longest run etc.
- Target features: AdultWeekday, AdultWeekend

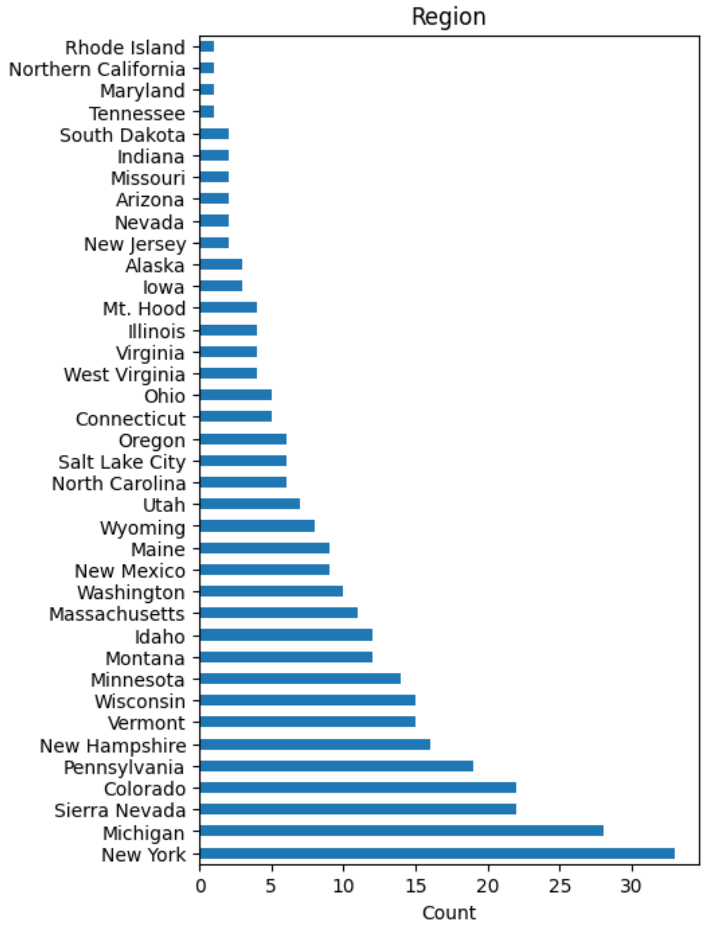
Exploratory Data Analysis

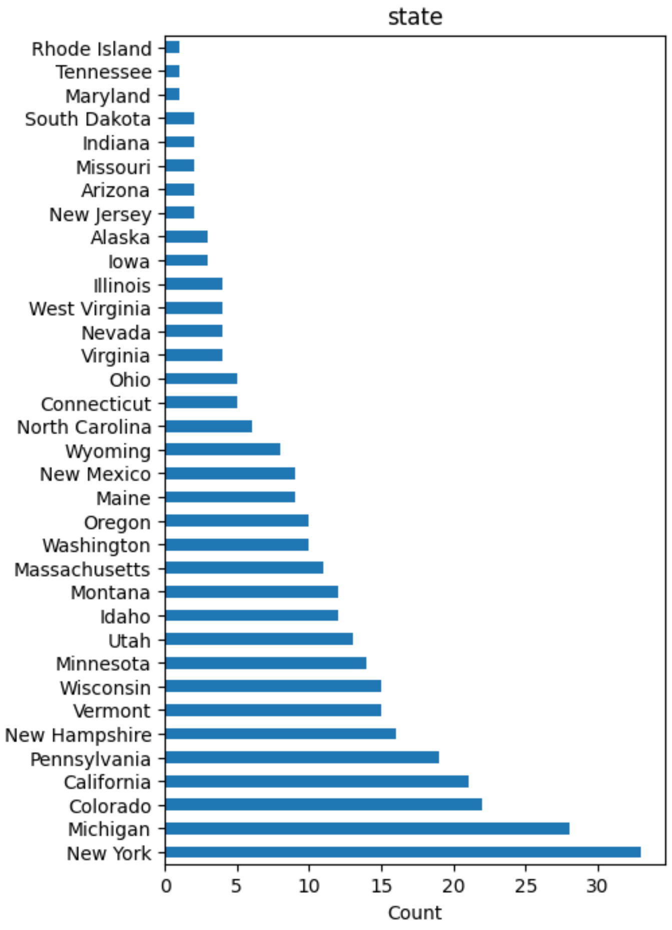
- AdultWeekend is right- skewed



Exploratory Data Analysis

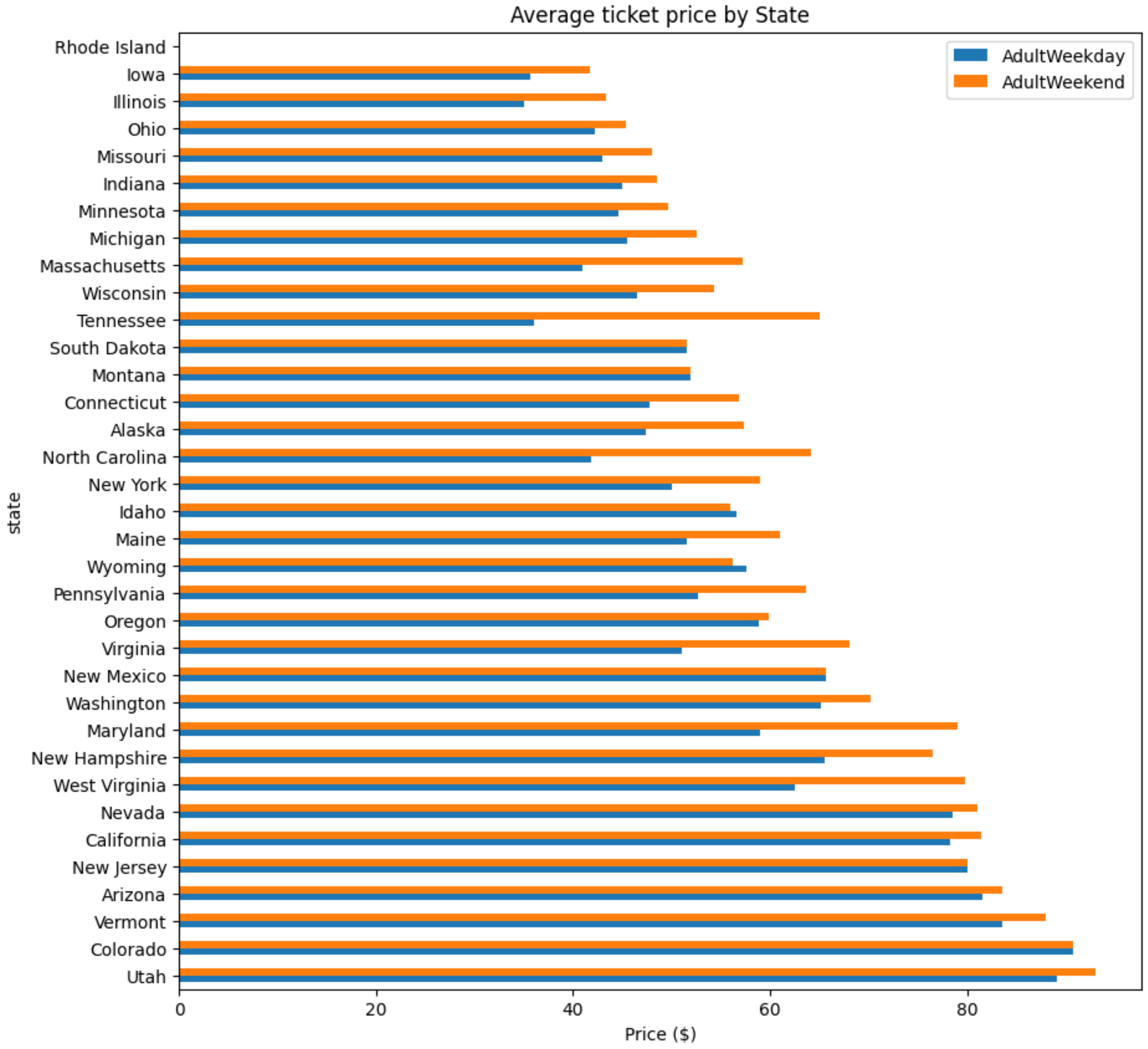
- New York and Michigan are the states and regions with the most ski resorts





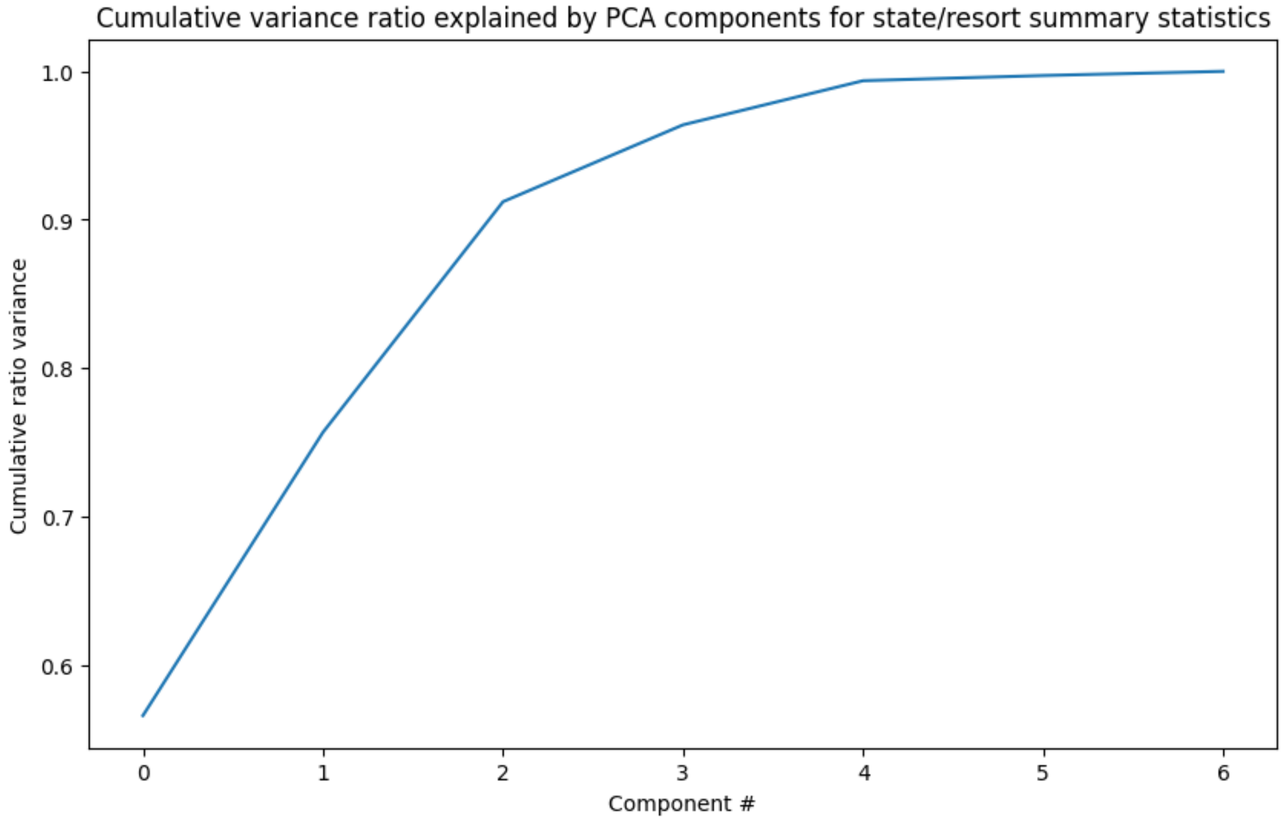
Exploratory Data Analysis

- Utah, Colorado and Vermont have the highest average ticket prices



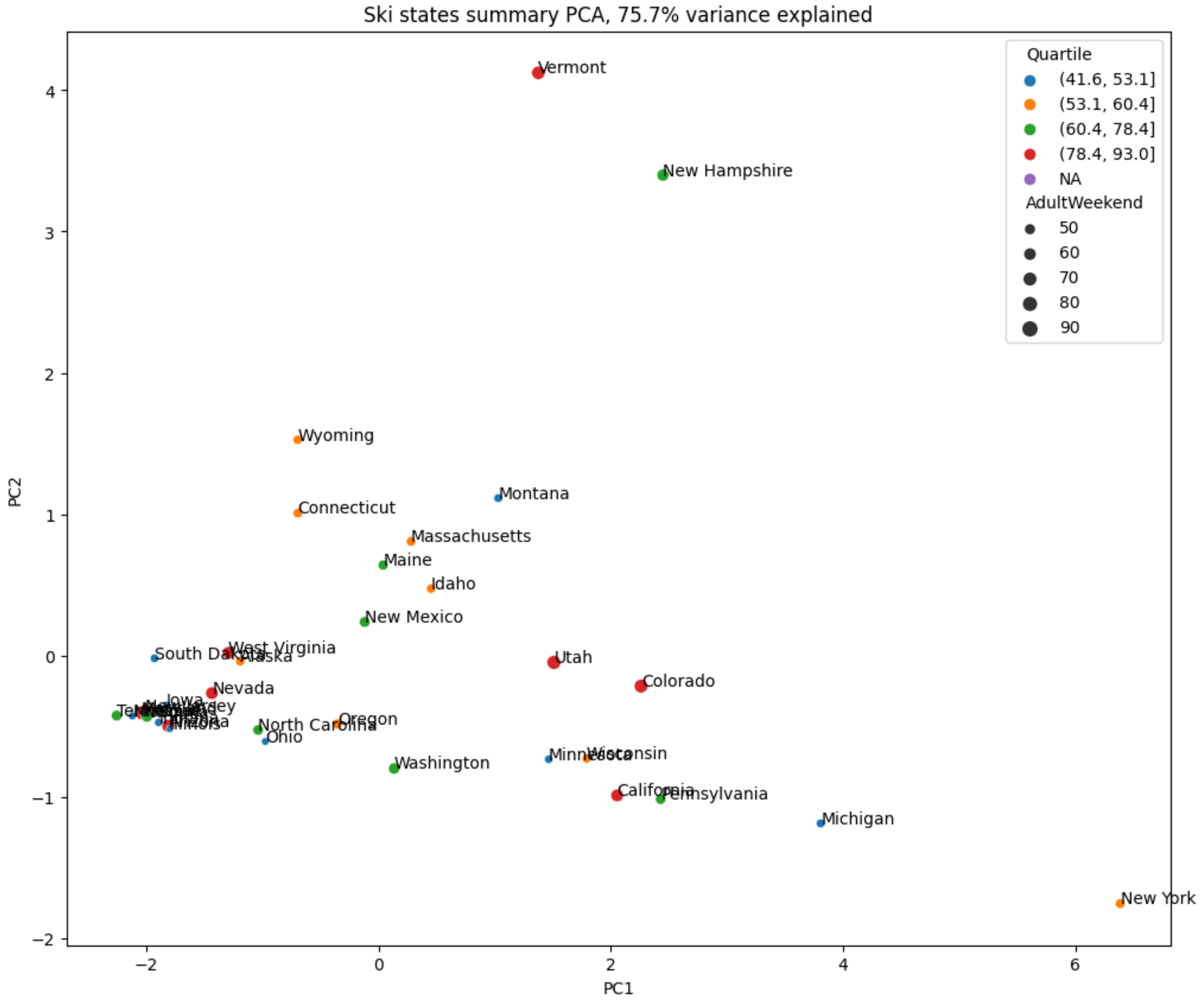
Exploratory Data Analysis

- First two components account for over 75% of the variance
- First four components account for over 95% of the variance



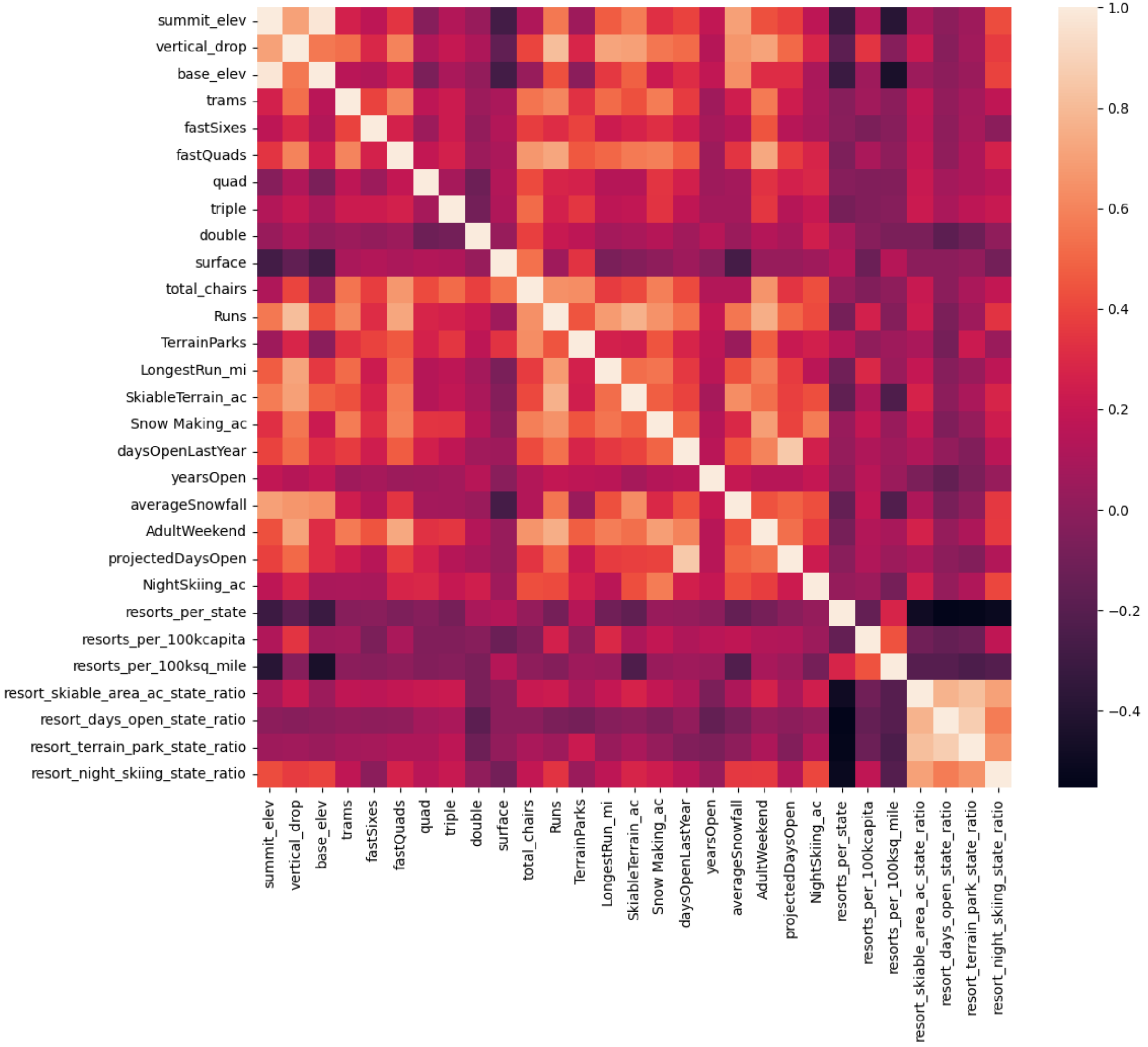
Exploratory Data Analysis

- No clear pattern with distribution of states with the first two components



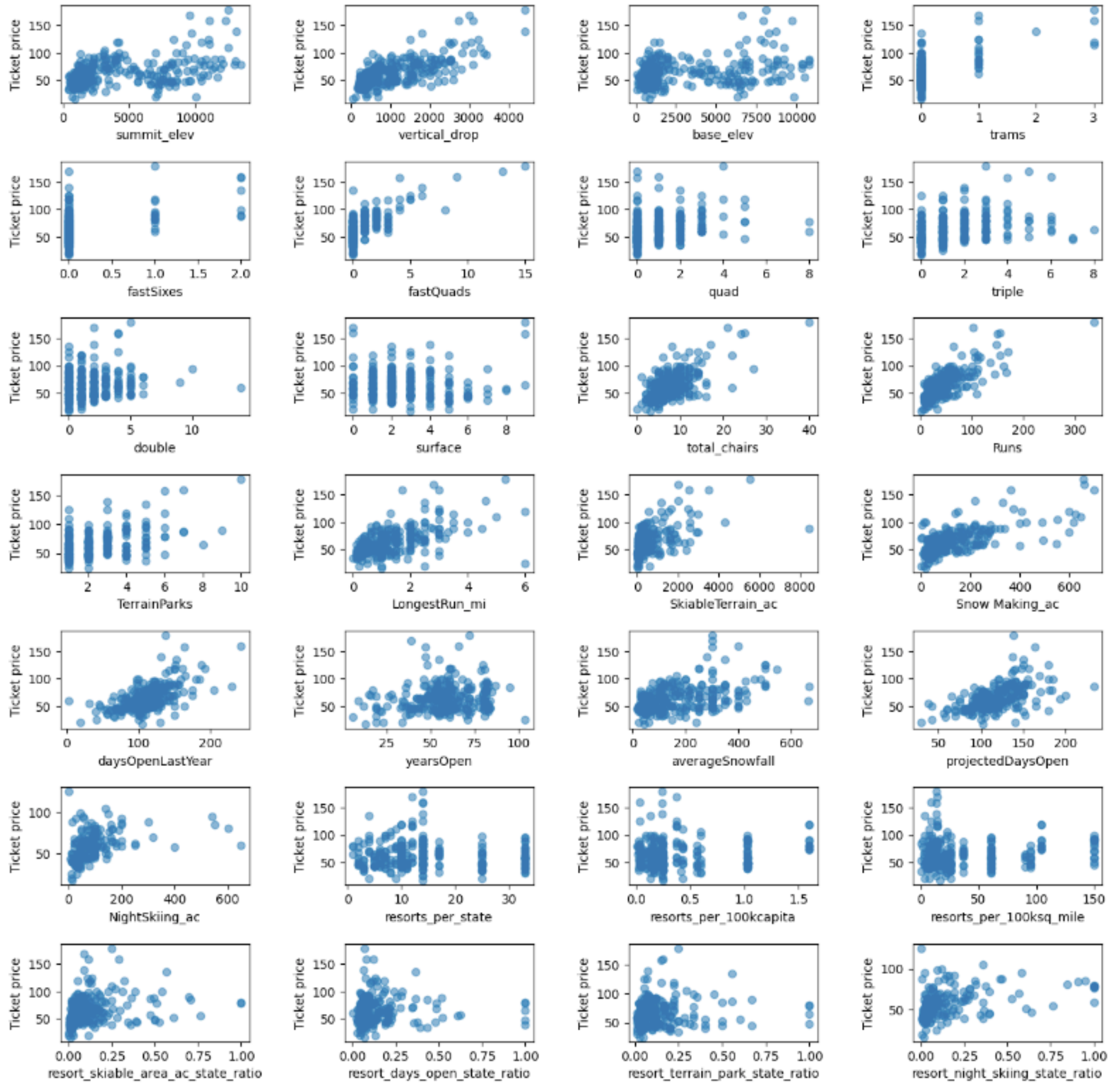
Exploratory Data Analysis

- fastQuads, Runs, tota_chairs, and Snow Making_ac are strongly correlated with AdultWeekend
- Resort_night_skiing_state_ratio correlated with AdultWeekend



Exploratory Data Analysis

- Scatter plots of features vs ticket price matches the correlation matrix



Preprocessing and Training

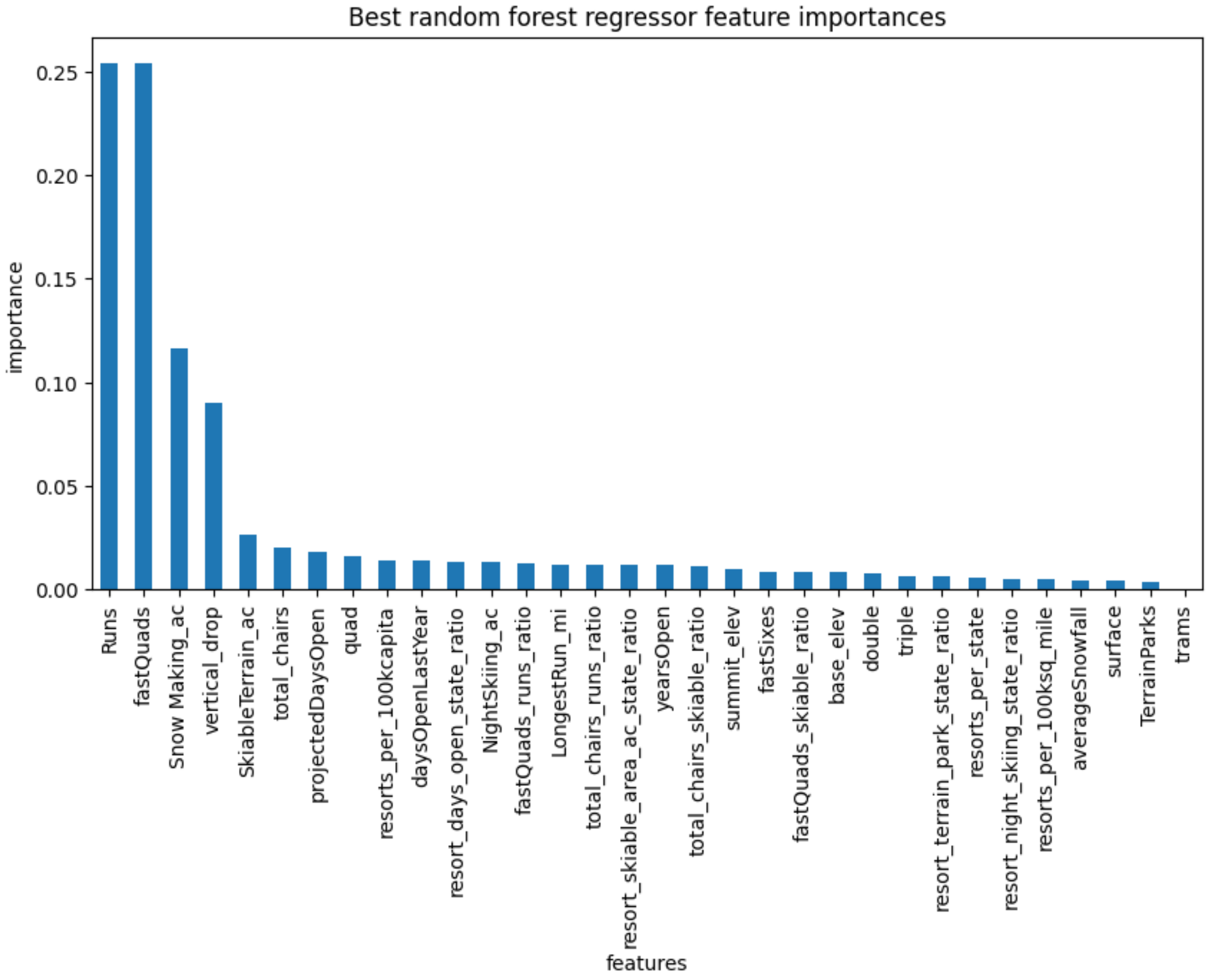
- 70/30 train/test split
- Impute missing data with mean or median
- Scale data around 0 using a standard scaler

Modeling: Linear Regression Model

- Select k best features
- R-squared score on train data: 0.72
- R-squared score on test data: 0.638
- MAE on train data: 9.21
- MAE on test data: 10.49
- CV score mean: 0.633
- CV score standard deviation: 0.095

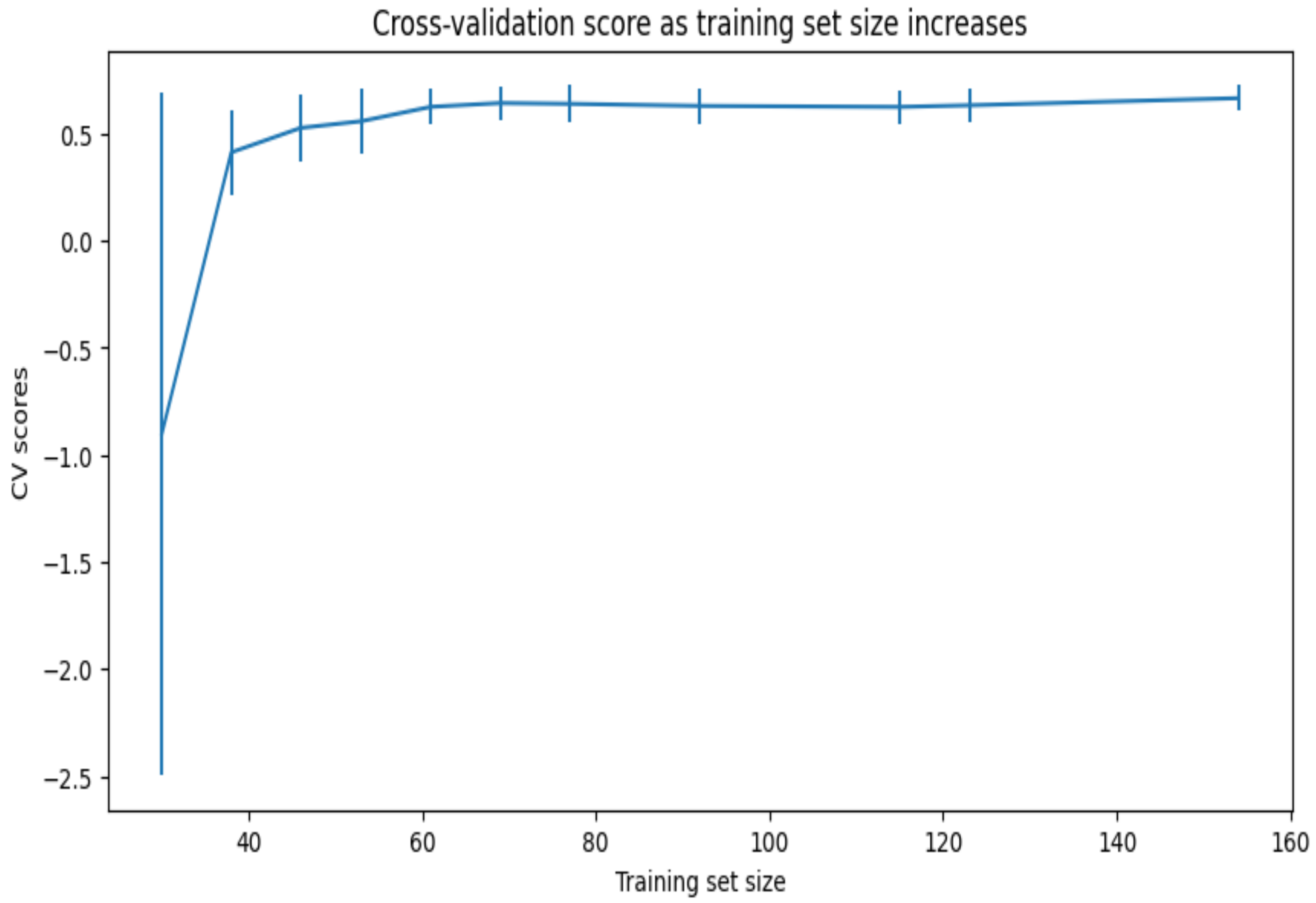
Modeling: Random Forest Regressor

- CV score mean: 0.713
- CV score standard deviation: 0.071
- Runs and fastQuads are most important features
- Snow Making_ac and vertical_drop also important
- Better model than linear regression



Evaluation

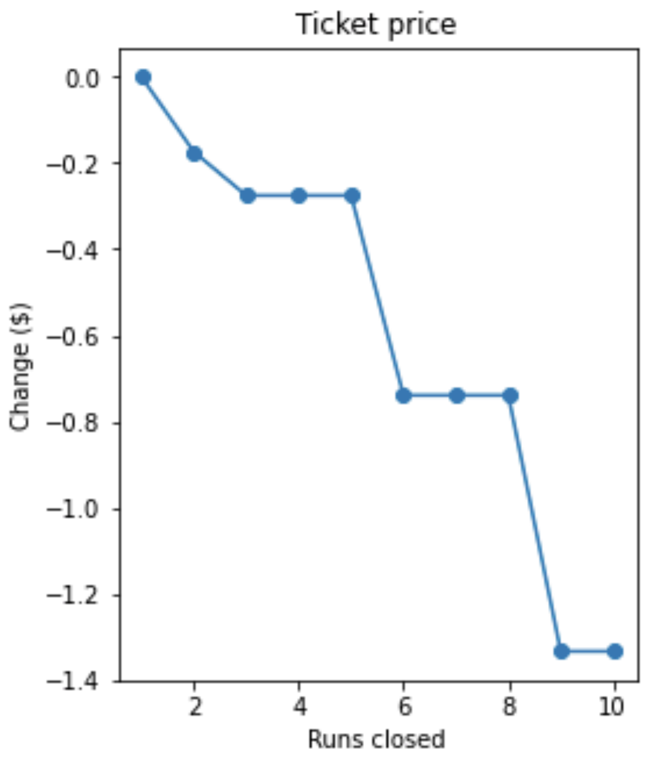
- Cv scores level off by around 40-50 samples indicating there is sufficient data

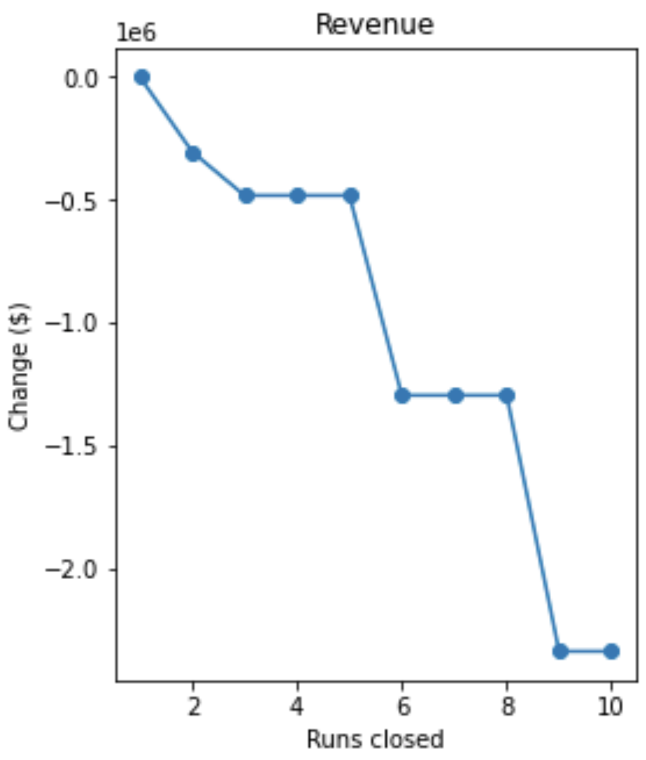


Modeling Results: Scenario 1

- Close up to 10 of the least used runs

- Closing one run makes no difference
- Closing 2 and 3 runs reduces support for ticket price
- Closing 4 or 5 runs produces the same loss in revenue as closing 3 runs
- Closing 6 or more runs leads to large drop in ticket price and revenue





Modeling Results: Scenario 2

- Add a run, increase the vertical drop by 150ft, and install a new chairlift

- Increases support for an increase of ticket price by $161
- Over the season, this is expected to amount to $2,815,217

Modeling Results: Scenario 3

- Add a run, increase the vertical drop by 150ft, install a new chairlift, and add 2 acres of snow making

- This scenario gives the exact same results as scenario 2
- Increases support for an increase of ticket price by $161
- Over the season, this is expected to amount to $2,815,217

Modeling Results: Scenario 4

- Increase the longest run by 0.2 miles and guarantee its snow coverage by adding 4 acres of snow making capability.

- This scenario resulted in making no difference

Conclusion

- Scenario 2 yields the best results
  - It will allow for an increase in price that will cover the seasonal operational cost of adding the new chairlift