Guided Capstone Project Report

Problem statement.—BMR has excellent facilities. Their Montana slopes offer views of Glacier National Park and Flathead National Forest. Their longest run is 3.3 miles and they have 11 aerial lifts, two t-bars, and one magic carpet. Last season they attracted 350,000 skiers and snowboarders.

BMR recently installed new chair lift that increases this season's operational cost by $1,540,000. BMR's business leadership wants a way to cover this additional cost. Aware that their current pricing strategy – charging a premium on the market average – is likely below the market supported price, BMR wants to use industry data to find a pricing model for ski resort tickets in our market segment. BMR is also looking to take other cost cutting measures to improve their facilities.

Data wrangling.—Our data set had rows for 329 resorts across the US including BMR. The data's 27 columns show us the state and region of each resort, its lift facilities, skiing facilities, and weekend and weekday adult ticket prices. After cleaning the data, we decided to make weekend prices our target feature since it had fewer missing values. Finally, we imported and merged state area and population columns to our data set to calculate states summary statistics. Our resulting data set had 277 rows and 25 columns.

Exploratory data analysis.—For our EDA, we tried to segment the market by state. In our univariate exploratory analysis, the data seemed to suggest New England states might belong to a different market segment.

In our multivariate analysis, we performed a principal component analysis to find market segments. However, the results of our PCA revealed no clustering, which led us to give up on segmenting the market by state.[2] For our multivariate graphical analysis, we created a feature correlation heatmap.[3] The revealed correlations between ticket price and the ratio of a resort's night skiing area to the state's total night skiing area, the number of chair, and total runs.

Model preprocessing with feature engineering.—To estimate the supported market price, we constructed two models: a linear regression and random forest regression. The winning model was the random forest search and we used it find the most important features: fast quads, runs, snow making area, and vertical drop.[4] We suggest using these features when considering which facilities to improve for the season.

Algorithms used to build the model with evaluation metrics.—To decide on our model, we preformed similar process on a linear regression and a random forest regression. First, we split the data into train/test sets. Second, we made data pipelines that would impute the median, standardize the data, and preform a regression. (For the linear regression, we added a step. We use a cross-validation grid search to find the best number of parameters.) Then, we fitted the

pipelines to the training data. Next, to evaluate the models, we used cross-validation to estimate the coefficient of determination, mean squared error, and the mean absolute error. Finally, we selected the winning model and graphed the cross-validation scores as the data used to fit the model increased to determine whether we had enough data to produce an accurate model.[5]

Winning model and scenario modeling.—The winning model was the random forest regression because its mean absolute error was $1 less than the linear regression. Our model help us evaluate the impact of the following four scenarios: BMR (1) closes up to 10 runs, (2) extends its vertical drop 150ft below its base elevation and installs a new chair lift, (3) adds to the second Scenario 2 acers of snow makers, or (4) extends its longest run by 0.2 miles.

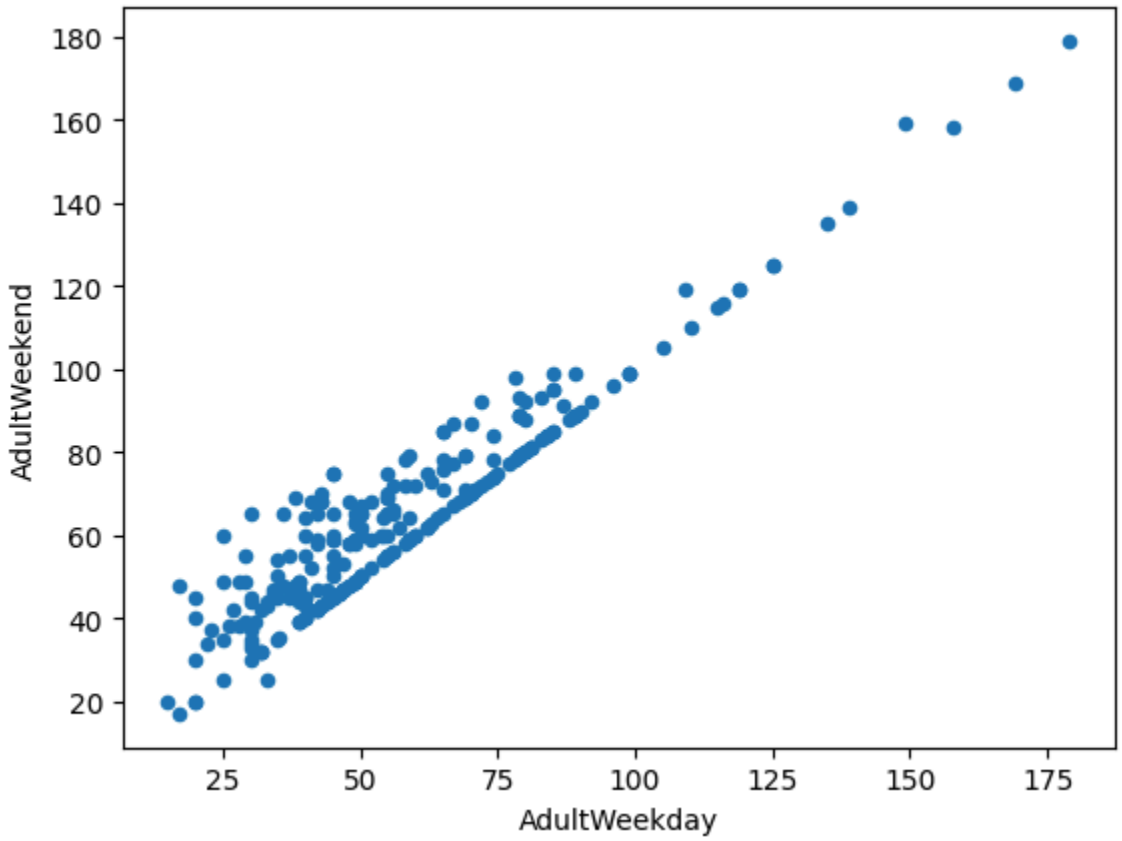
Pricing recommendation.—The current adult weekend ticket price is $81. The estimated market supported price is $96.84 with an mean absolute error of $10.35. We recommend raising ticket general prices between $5.49 and $15.84. Furthermore, we recommend closing one run, since it does not lower the market supported price, and implementing Scenario 2, since it raises the market supported price by $1.99.

Conclusion.—To concluded, we were tasked with finding a new pricing strategy using data on the skiing industry. To accomplish our task, we tried to use the tools of data analysis to segment the market, and tried to select an appropriate model to estimate the market supported price. Ultimately, we discovered that BMR's business leadership was correct that it could afford to raise its ticket prices.

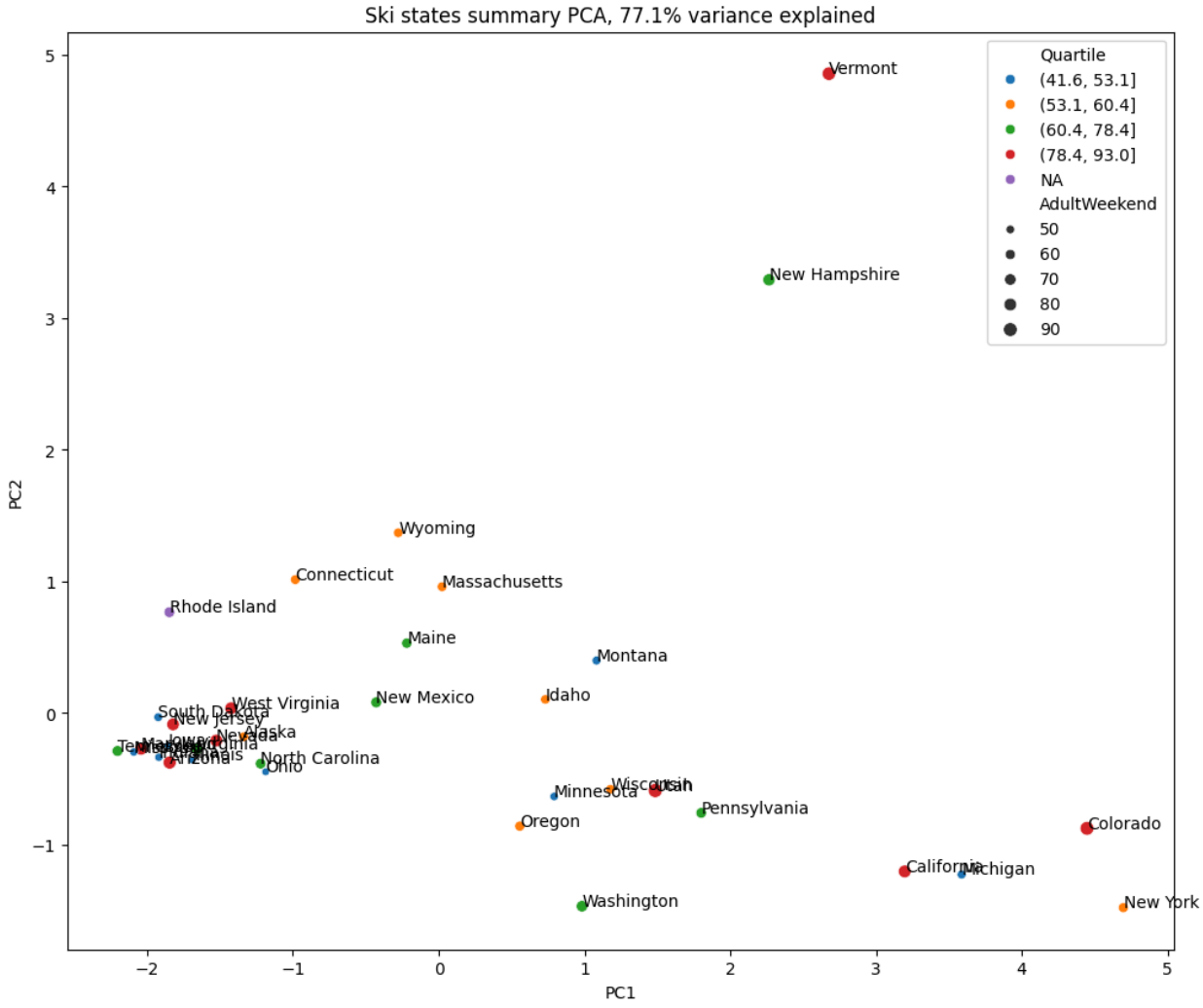
Future scope of work.—It is quite noteworthy that our initial suspicion that BMR was undervaluing its facilities was grounded in the quality of its location. However, our data does not account for natural beauty. It may be possible that the current model undervalues BMR's facilities by not accounting for its real estate.

Figures

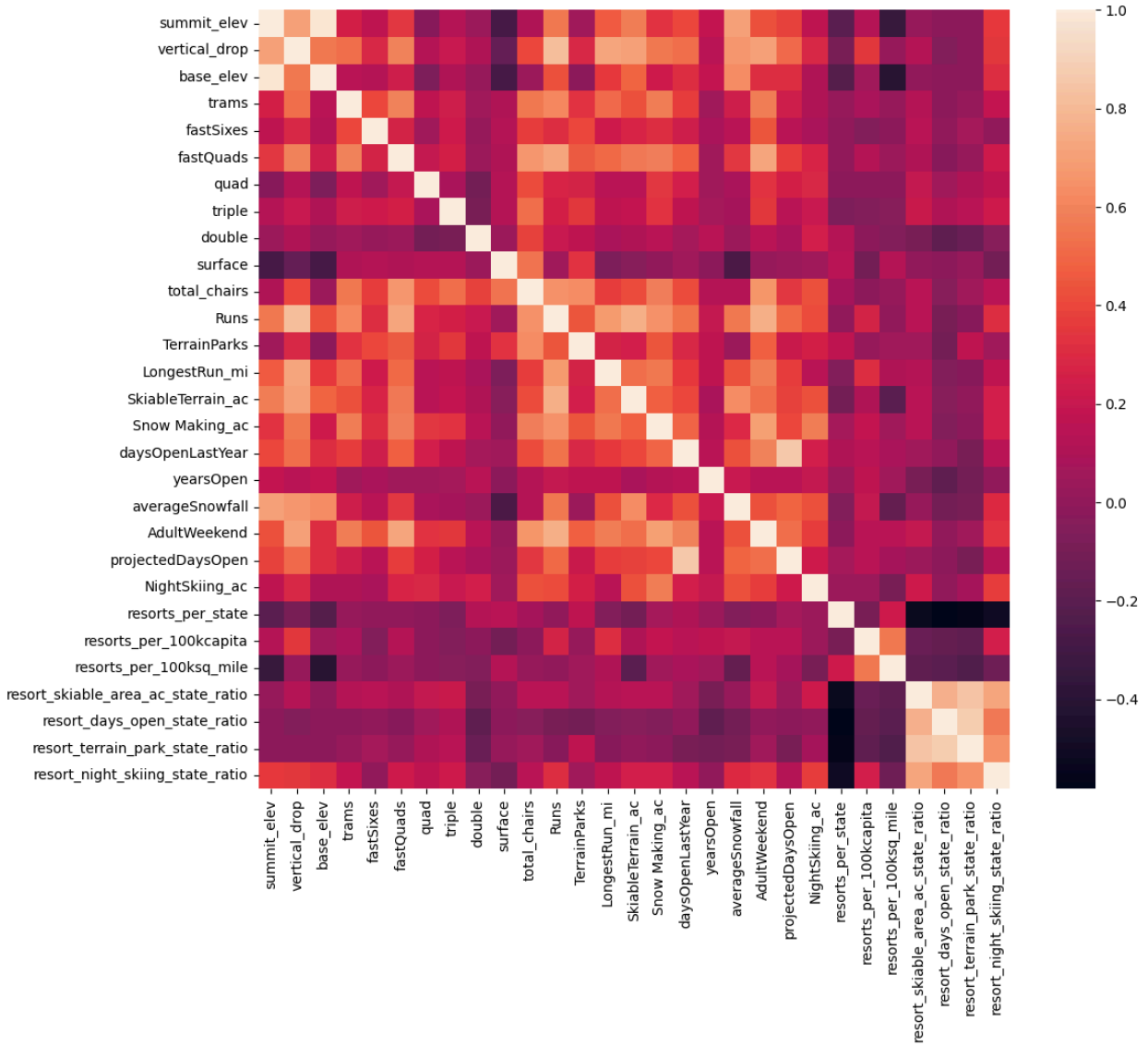
[1] “Target feature scatterplot”



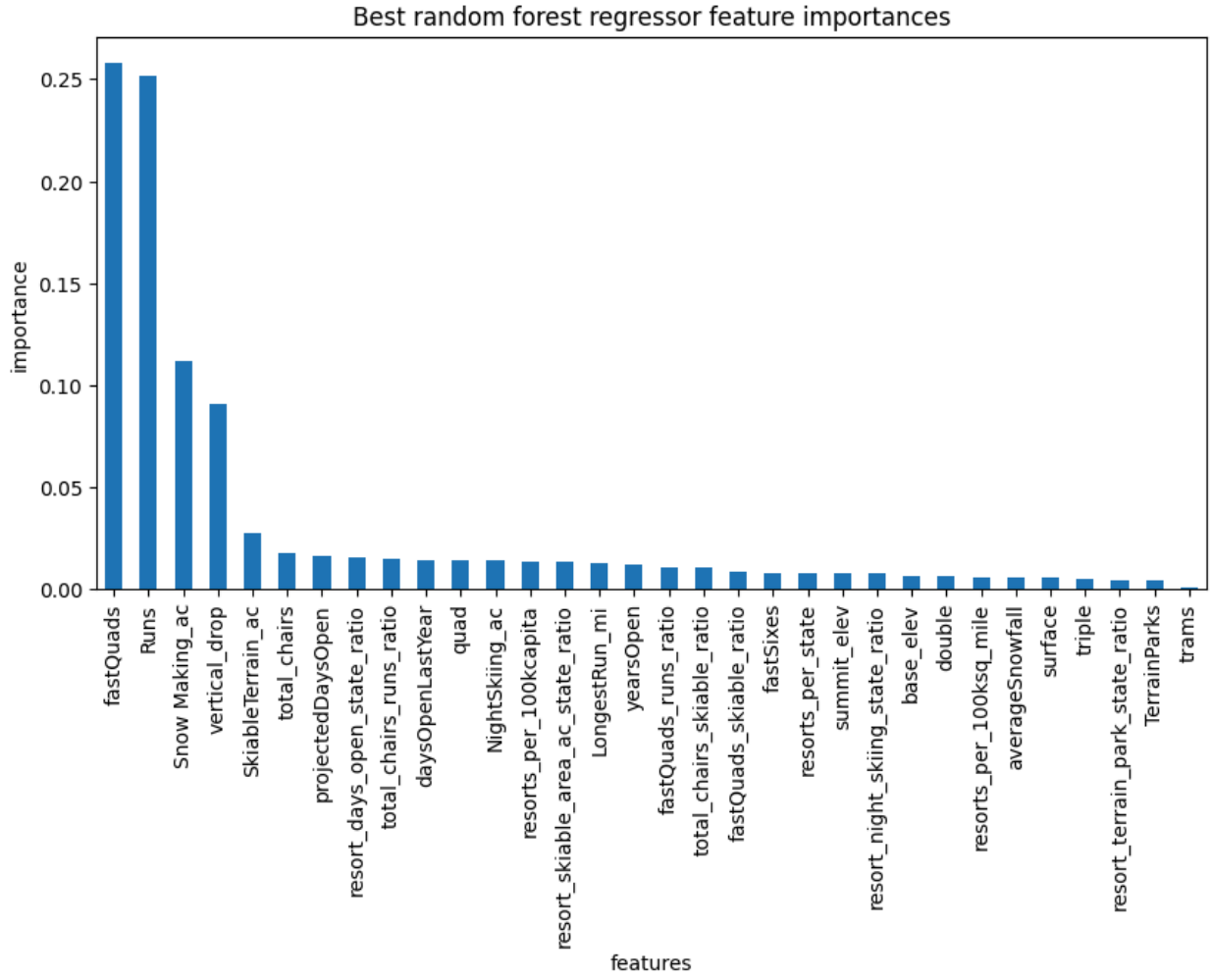
[2]


Ski states summary PCA, 77.1% variance explained
PC1
PC2
Quartile
(41.6, 53.1]
(53.1, 60.4]
(60.4, 78.4]
(78.4, 93.0]
NA
AdultWeekend
50
60
70
80
90
Vermont
New Hampshire
Wyoming
Connecticut
Massachusetts
Rhode Island
Maine
Montana
Idaho
New Mexico
West Virginia
New Jersey
North Carolina
Ohio
Minnesota
Pennsylvania
Oregon
Colorado
California
Michigan
Washington
New York

[3]


summit_elev
vertical_drop
base_elev
trams
fastSixes
fastQuads
quad
triple
double
surface
total_chairs
Runs
TerrainParks
LongestRun_mi
SkiableTerrain_ac
Snow Making_ac
daysOpenLastYear
yearsOpen
averageSnowfall
AdultWeekend
projectedDaysOpen
NightSkiing_ac
resorts_per_state
resorts_per_100kcapita
resorts_per_100ksq_mile
resort_skiable_area_ac_state_ratio
resort_days_open_state_ratio
resort_terrain_park_state_ratio
resort_night_skiing_state_ratio
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

[4]


Best random forest regressor feature importances
importance
0.25
0.20
0.15
0.10
0.05
0.00
fastQuads
Runs
Snow Making_ac
vertical_drop
SkiableTerrain_ac
total_chairs
projectedDaysOpen
resort_days_open_state_ratio
total_chairs_runs_ratio
daysOpenLastYear
quad
NightSkiing_ac
resorts_per_100kcapita
resort_skiable_area_ac_state_ratio
LongestRun_mi
yearsOpen
fastQuads_runs_ratio
total_chairs_skiable_ratio
fastQuads_skiable_ratio
fastSixes
resorts_per_state
summit_elev
resort_night_skiing_state_ratio
base_elev
double
resorts_per_100ksq_mile
averageSnowfall
surface
triple
resort_terrain_park_state_ratio
TerrainParks
trams
features

[5]


Cross-validation score as training set size increases
CV scores
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
40
60
80
100
120
140
160
Training set size