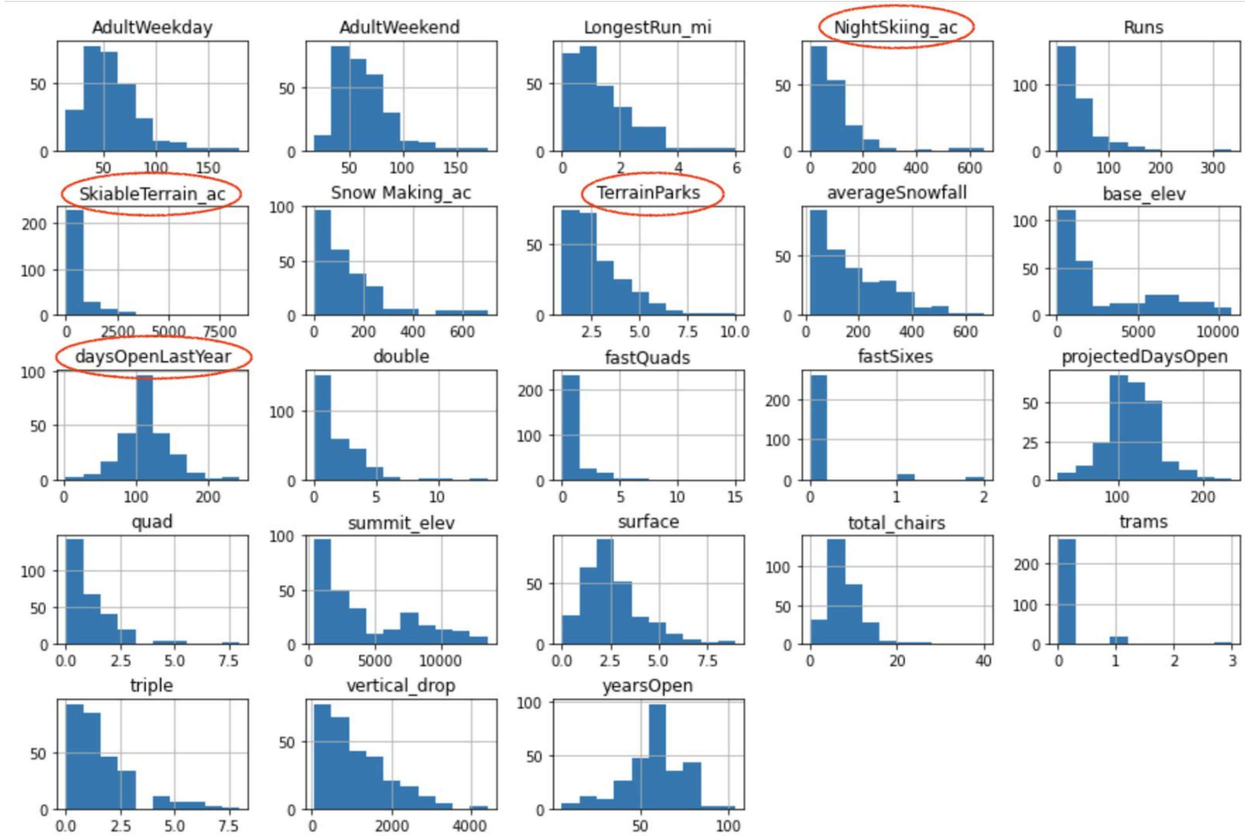
Guided Capstone Project Report

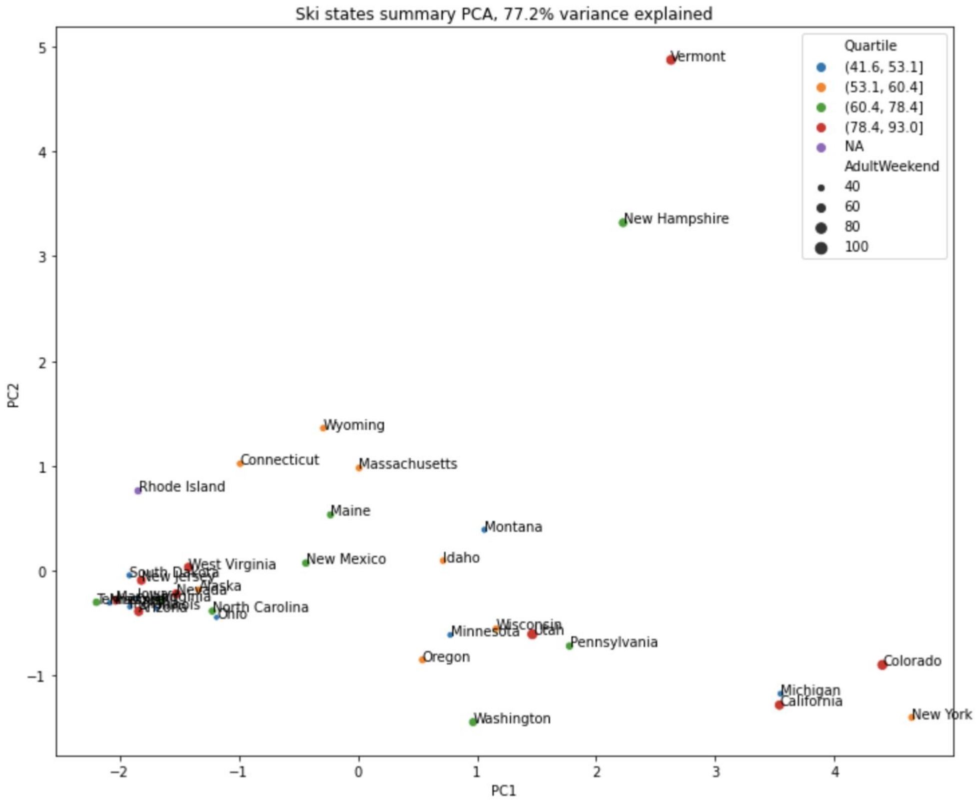
Recommendations

After a new chair lift installation, the added operational cost needs to be covered either by cutting unnecessary costs or increasing the ticket prices. Big Mountain resort tickets are already selling more than the average market ticket prices, therefore we need to find some features that are positively correlated with ticket prices and see how to predict the new prices based on these features.

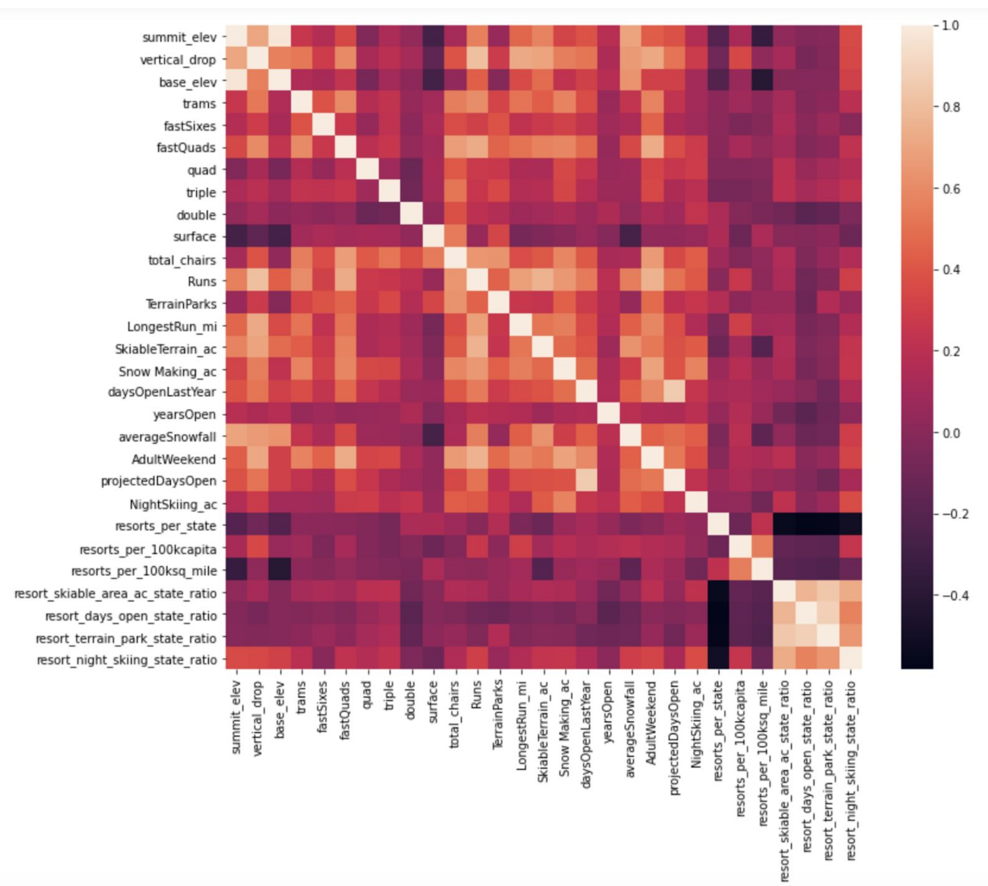
The problem solution started with data wrangling to make sure all the features and required target values are available in the correct format, Big Mountain resort has been checked and it was available in the dataset without any missing values, all the numerical and categorical fields have been searched for missing values either to input or remove. Then decided to keep weekend prices as weekday prices had more missing values. Finally, added two new columns as population and area for each state concatenating with five other columns as resorts per state, state total skiable area, state total days open, state total terrain parks, and state total night-skiing area, as the state summary data frame. The result of all these data wranglings is available in the below distribution plot:



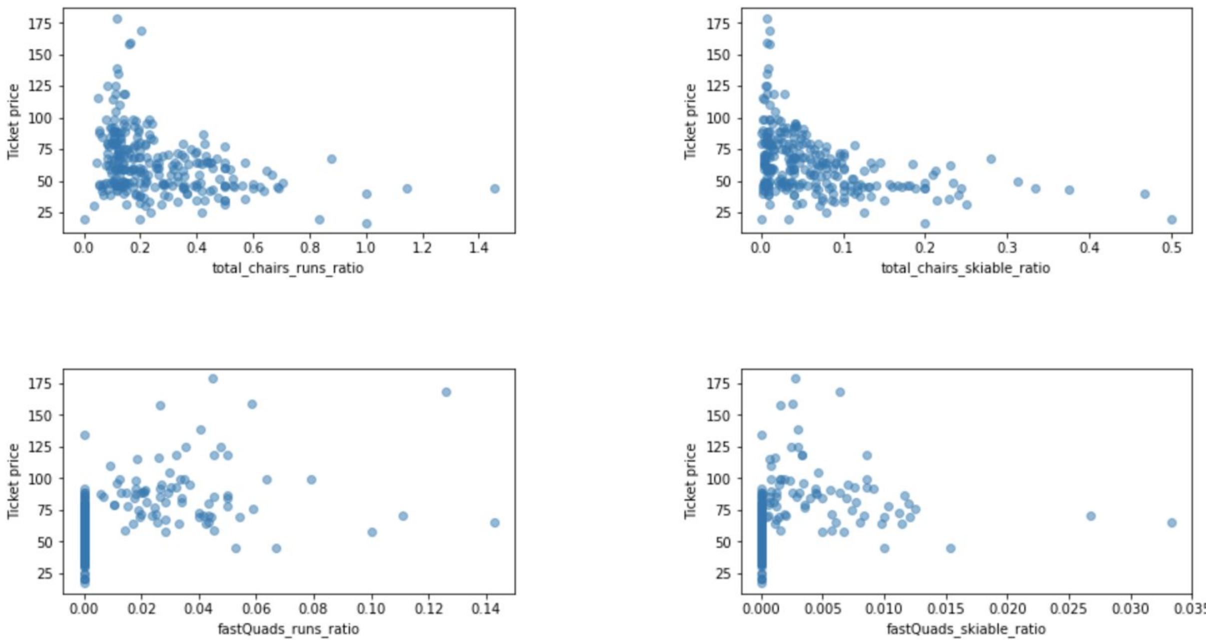
Te visualization and exploratory data analysis show the correlation between some features and the ticket price. Here are the plots:



Ski states summary PCA, 77.2% variance explained
Quartile
(41.6, 53.1]
(53.1, 60.4]
(60.4, 78.4]
(78.4, 93.0]
NA
AdultWeekend
40
60
80
100
Vermont
New Hampshire
Wyoming
Connecticut
Massachusetts
Rhode Island
Maine
Montana
New Mexico
Idaho
West Virginia
North Carolina
Ohio
Minnesota
Wisconsin
Utah
Pennsylvania
Oregon
Colorado
Michigan
California
New York
Washington
PC1
PC2



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



Now it's time to create a model to predict the ticket price, after partitioning the dataset to test and train the dataset, and impute the missing values with mean the R^2 is equal to (0.8170, 0.7163) and mean absolute error is (8.5368, 9.4163) and mean squared error is (112.3769, 164.3926).

Using the median value to impute the missing values and train the model will result R^2 equal to (0.8177, 0.7209) and mean absolute error equal to (8.5478, 9.4070) and mean squared error equal to (111.8958, 161.7315).

After a deep investigation based on linear regression the coefficient for the first eight effective parameters are as below:

vertical_drop	10.767857
Snow Making_ac	6.290074
total_chairs	5.794156
fastQuads	5.745626
Runs	5.370555
LongestRun_mi	0.181814
trams	-4.142024
SkiableTerrain_ac	-5.249780

Those features colored in green are also the features selected based on the random forest model. And final results about the Big Mountain Resort are as below:

1. Big Mountain has the most expensive ticket in Montana but somewhere in the middle within all resorts.
2. Big Mountain is doing well for the vertical drop, but there are still quite a few resorts with a greater drop.
3. Big Mountain is very high up the league table of snowmaking area.
4. Big Mountain has amongst the highest number of total chairs, resorts with more appear to be outliers.

5. Most resorts have no fast quads. Big Mountain has 3, which puts it high up that league table. There are some values much higher, but they are rare.
6. Big Mountain compares well for the number of runs. There are some resorts with more, but not many.
7. Big Mountain has one of the longest runs. Although it is just over half the length of the longest, the longer ones are rare.
8. The vast majority of resorts, such as Big Mountain, have no trams.
9. Big Mountain is amongst the resorts with the largest amount of skiable terrain.

There are four different business scenarios which have been explained below:

1. Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
2. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snowmaking coverage
3. Same as number 2, but adding 2 acres of snowmaking cover
4. Increase the longest run by 0.2 miles to boast 3.5 miles length, requiring additional snowmaking coverage of 4 acres

The random forest model feedback about each scenario is as below:

1. The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in the ticket price. Increasing the closures down to 6 or more leads to a large drop.
2. Increasing the vertical drop, support for ticket price by $1.99. Over the season, this could be expected to amount to $3,474,638
3. Increasing the vertical drop and 2 acres of snowmaking makes no difference and will result in exactly the same as the previous scenario.
4. The last scenario also makes no difference.