Big Mountain Report

Problem Statement:

Big Mountain Resort, a ski resort located in Montana. It has recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This additional chair increases their operating costs by $1,540,000. And currently its pricing on market average does not provide the business with a good sense of how important some facilities are compared to others. So, we need to decide how to make up for the additional $1,540,000 cost for this season by a) adjusting the ticket price b) cut down the unnecessary cost. Currently its ticket prices appear to be lower than our model generated price. Stakeholders are seeking our analyst and make decision based on current price probability and compare to all states.

Data Wrangling:

We have performed the data wrangling by calculate the missing value

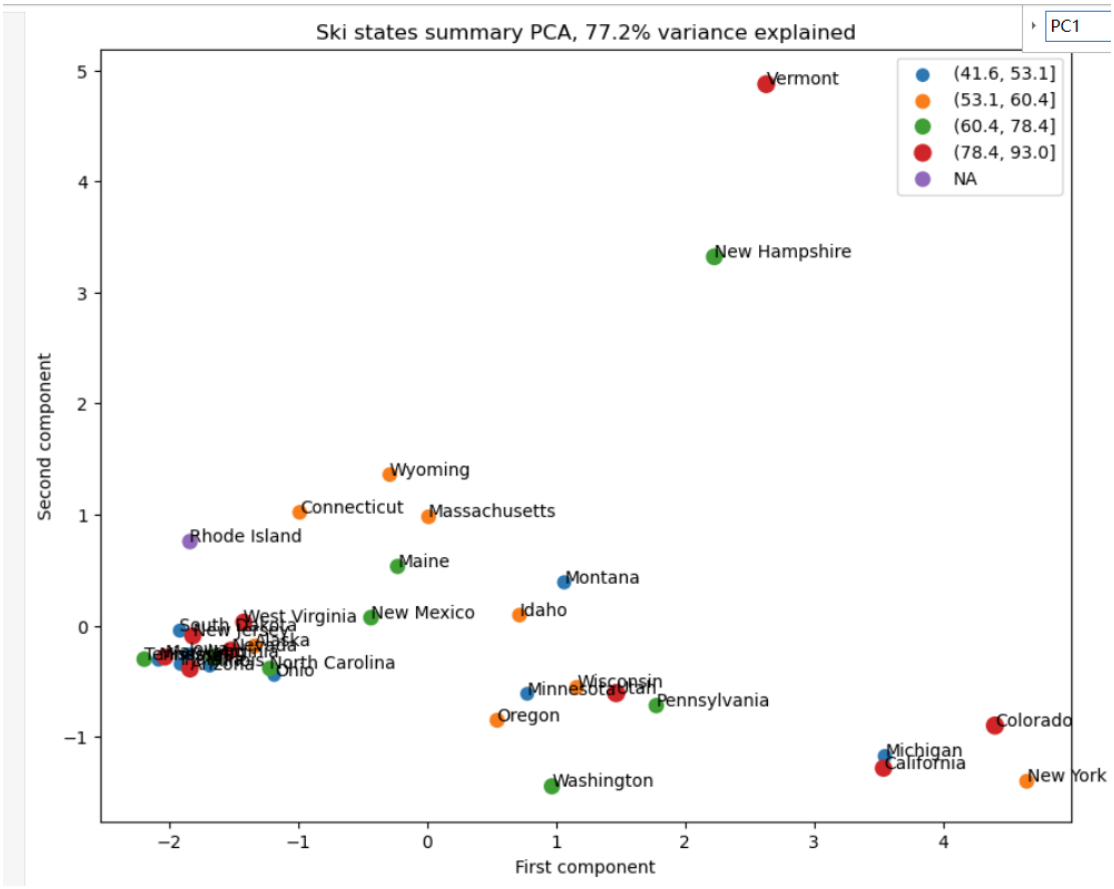
```
summit_elev                          4074.554404
vertical_drop                        1043.196891
base_elev                            3020.512953
trams                                   0.103627
fastSixes                               0.072539
fastQuads                               0.673575
quad                                    1.010363
triple                                  1.440415
double                                  1.813472
surface                                 2.497409
total_chairs                            7.611399
Runs                                   41.188482
TerrainParks                            2.434783
LongestRun_mi                           1.293122
SkiableTerrain_ac                     448.785340
Snow Making_ac                        129.601190
daysOpenLastYear                      110.100629
yearsOpen                              56.559585
averageSnowfall                       162.310160
projectedDaysOpen                     115.920245
NightSkiing_ac                         86.384615
resorts_per_state                      16.264249
resorts_per_100kcapita                  0.424802
resorts_per_100ksq_mile                40.957785
resort_skiable_area_ac_state_ratio      0.097205
resort_days_open_state_ratio            0.126014
resort_terrain_park_state_ratio         0.116022
resort_night_skiing_state_ratio         0.155024
total_chairs_runs_ratio                 0.271441
total_chairs_skiable_ratio              0.070483
fastQuads_runs_ratio                    0.010401
fastQuads_skiable_ratio                 0.001633
dtype: float64
```

and replace with predicted value using our generated model and based on the all-state information relating to all features. We have tried mean and median, seems like the predict value is not that much different in our case, but for general testing case, we still prefer use median in case some outlier exist. We used state information data to add in our case to calculate the resorts_per_100kcapita and resorts_per_100ksq_mile and then use our original data based on state and hotel information to get all ratio value for comparison.

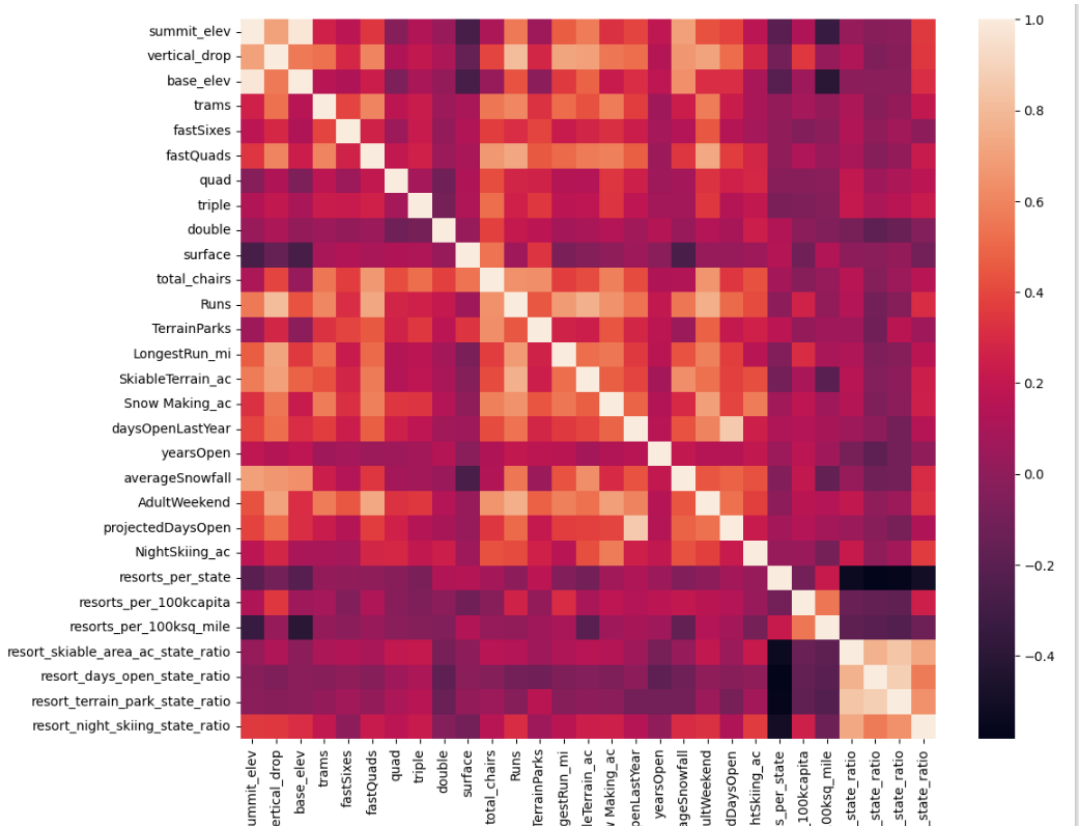
Exploratory Data Analysis (EDA):

We calculated the Total state area, Total state population, Resorts per state, Total skiable area, Total days open to gain some perspective for all the state.

To reduce dimensionality and better visualize the dataset, we applied PCA, for first 2 compoennts PC1 and PC2 explained 77 %of the variance.

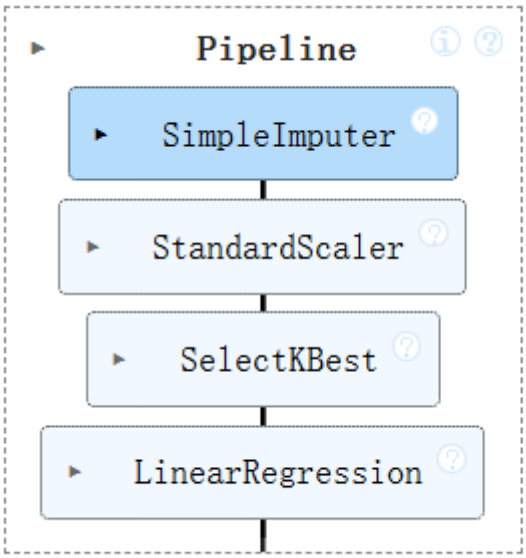


We also used correlation heatmap to gain a high-level view of relationships amongst the features.

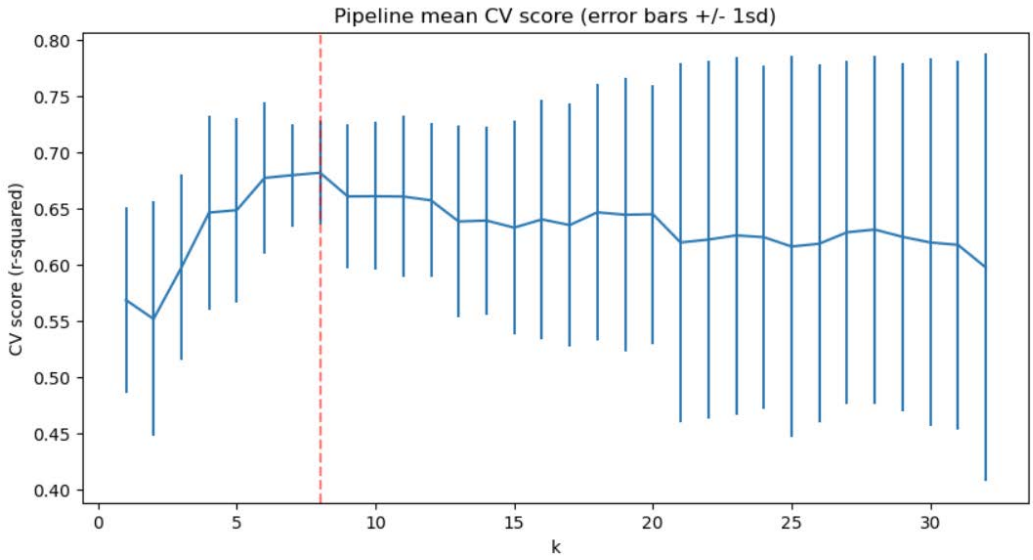


Model Preprocessing with feature engineering:

We used 70/30 train/test split to evaluate out-of-sample performance We used Baseline Model. with sklearn's DummyRegressor to generate the average price.



And we used Evaluation Metrics, used proportion of variance in the dependent variable (our ticket price) that is predicted by our "model". With Mean Absolute Error and Mean Squared Error to predict the value is off range to better understand the test case accuracy. We Imputing missing feature (predictor) values with median and mean. And then we use cross-validation for multiple values of k and use cross-validation to pick the value of k that gives the best performance.



Lastly, we use Random Forest Model to perform nonlinear test.

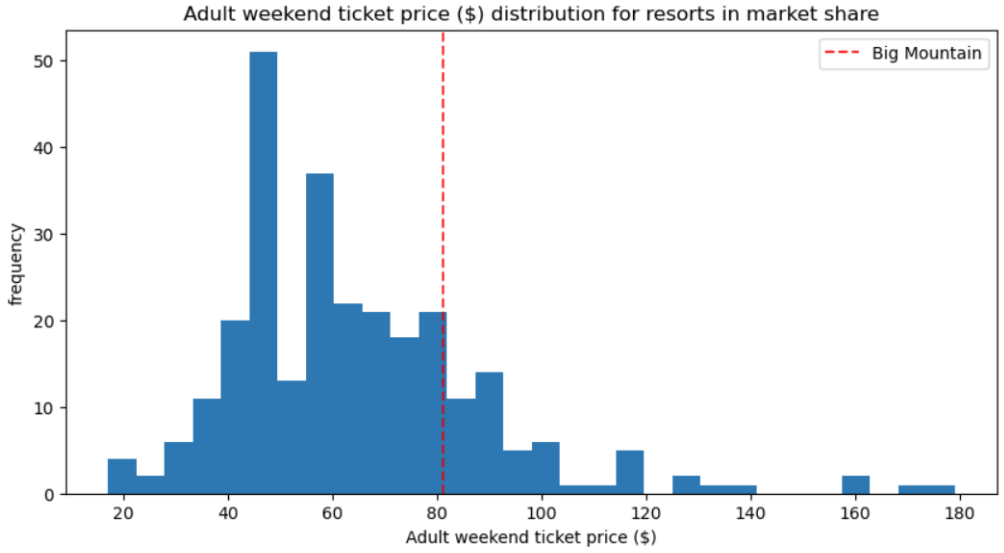
Algorithms used to build the model with evaluation metric:

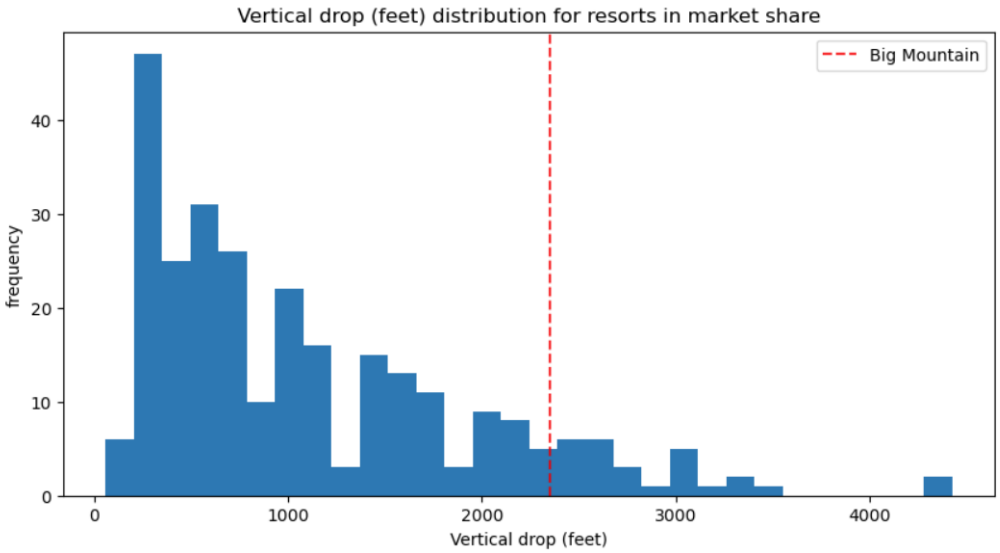
Dummy Regressor to calculate MSE and MAE. Linear regession, Cross-Validation for Hyperparameter (k-fold cross-validation) and Random Forest Regressor

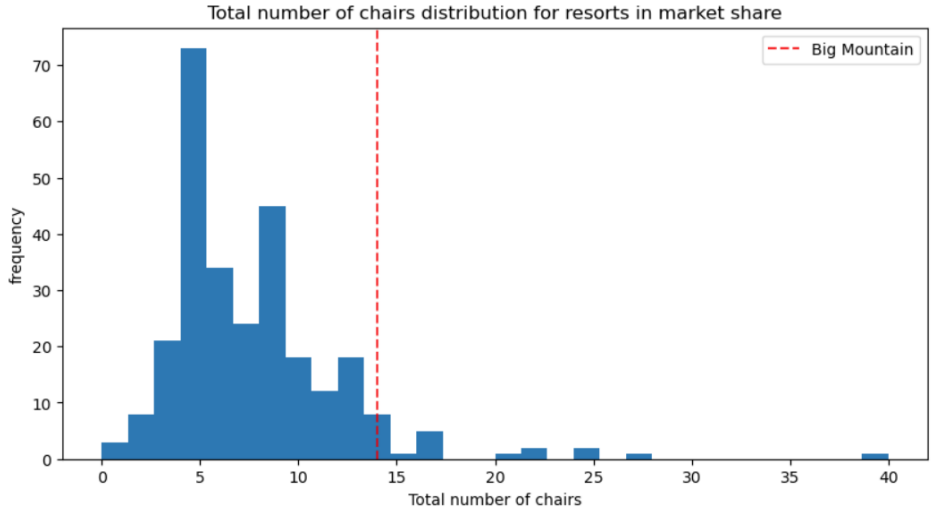
$$MAE = \frac{1}{n}\sum_{i}^{n}|y_i - \hat{y}| \qquad MSE = \frac{1}{n}\sum_{i}^{n}(y_i - \hat{y})^2$$

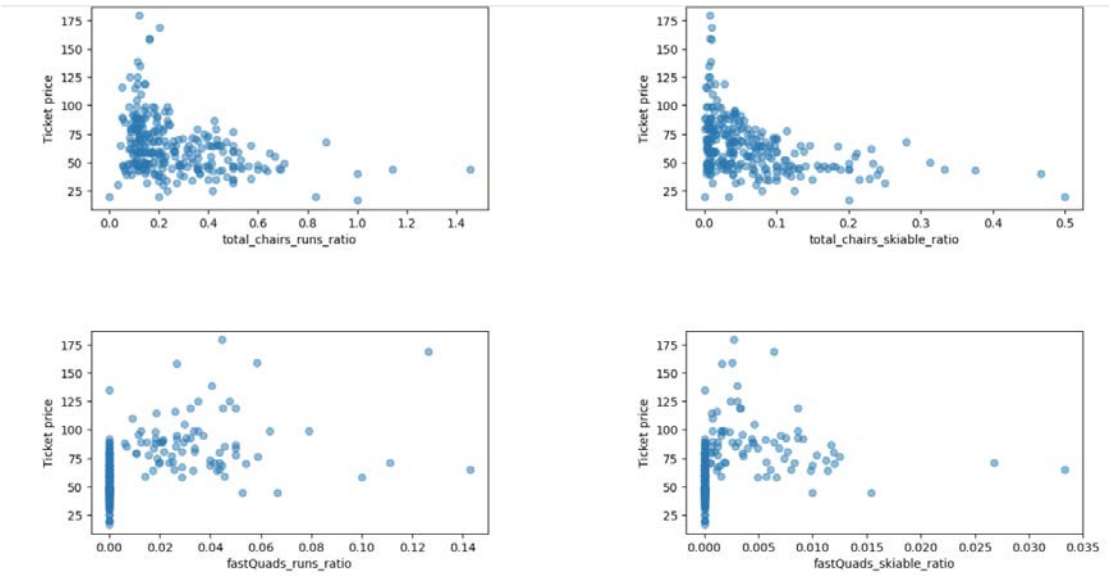
Winning model and scenario modelling:

Random Forest model win since it captures nonlinear relationship which is good, because the ticket price is not always linear regression. Random Forest model also captures all different interactions such as total_ski, vertical drop, chair information and outliers doesn't affect this model as linear model.









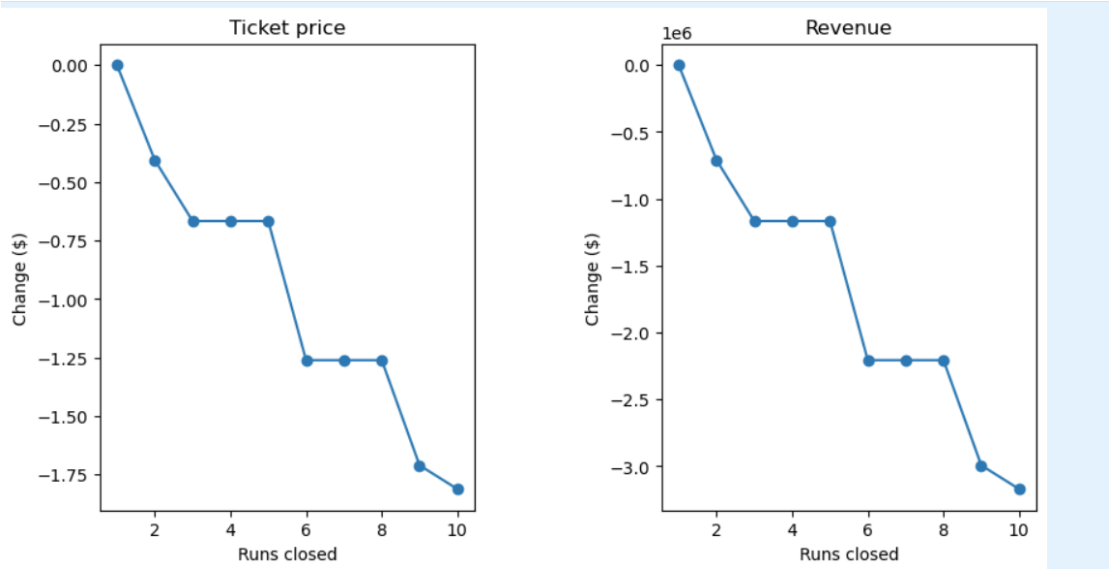
Pricing recommendation:

We recommend price to $4.33 by increasing the vertical drop by 150 feet, and installing an additional chair lift and adding 2 acres of snow making. In this case in one season, it could be

expected to amount to $7583333.

Conclusion:

I will suggest business leadership to conside cut down runs for saving cost or add vertical drop to gain more revenue. Adding new chair reach max profit to 6. for future improvement will suggest use model 2 to add drop and chair to increase price. Cut down runs have negative impact altough can lower cost but unreliable.



Future scope of work:

There is other cost information would be useful: staffing cost, utilities maintenance, guest insurance, equipment replacement, food and energy and rental fees. Without all this info, the result can be easily bias, so for better predication these should also be added into our model. The reason why the big mountain price modeled price is much higher than current price is that we use the data from all state. If we took a look at state Montana, the price is already the highest. So I believe the business executives is being conservative since not many competitors in the same state. If the business leaders felt this model was useful, I will start developed user interface like dashboard to show features of this model for stakeholders to better use it. And scale the model future more with more useful data added in and enrich the features.