Big Mountain Resort Final Report

Problem Statement:

How to increase profit for Big Mountain Resort by $1,540,000 by either increasing the ticket price (with or without changing available facilities), or cutting down costs without undermining ticket price for the coming year

Context: Big Mountain Resort (BMR) 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 this season. Now, to maintain profits, Big Mountain Resort wants to review its ticket pricing strategy by using a data-driven approach

Data: Due to limited access to the resort's database, the only date we had was an excel file with information about different resorts across the US with ticket prices and other features.

Data Wrangling:

During data wrangling we accomplished two things:

1. We explored ski resort data provided by the db manager. After basic exploration, we did basic cleaning involving either fixing values or dropping columns or rows. After the wrangling, the shape of ski resort data changed from 330,27 to 277, 25. The figure below shows distribution of different features and values post wrangling.

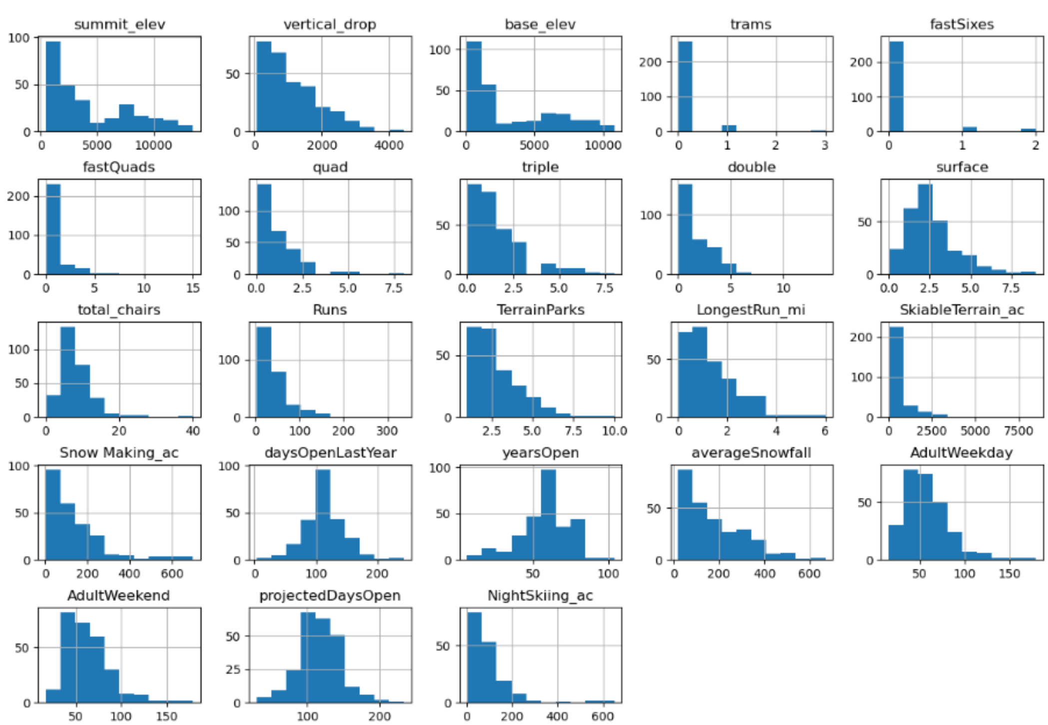

Figure 1: The distribution of features after basic cleanup. Most features have even distributions except few such as fastQuads, fastSixes and trams

2. We also built a state-wide table having information such as population, number of resorts, population summary and other information by using data from ski data provided to us and public data

EDA:

During the EDA stage we analyzed a state-wide picture for the market and added new features such as resorts_per_100K capita and resorts_per_100k sq mile. We also did PCA concluding that state-based analysis is not needed and thus all the resorts should be treated equally irrespective of their location.

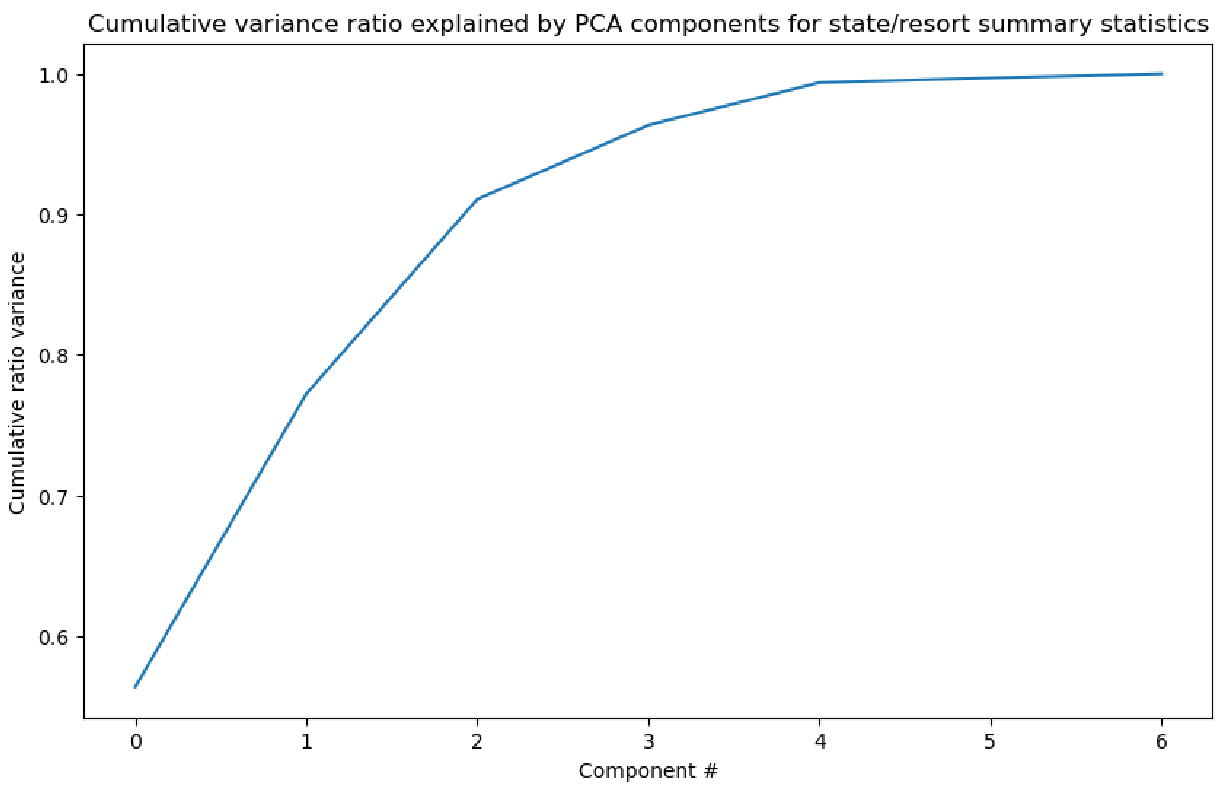

Figure 2: PCA analysis shows 2 features accounting for 75% in variability

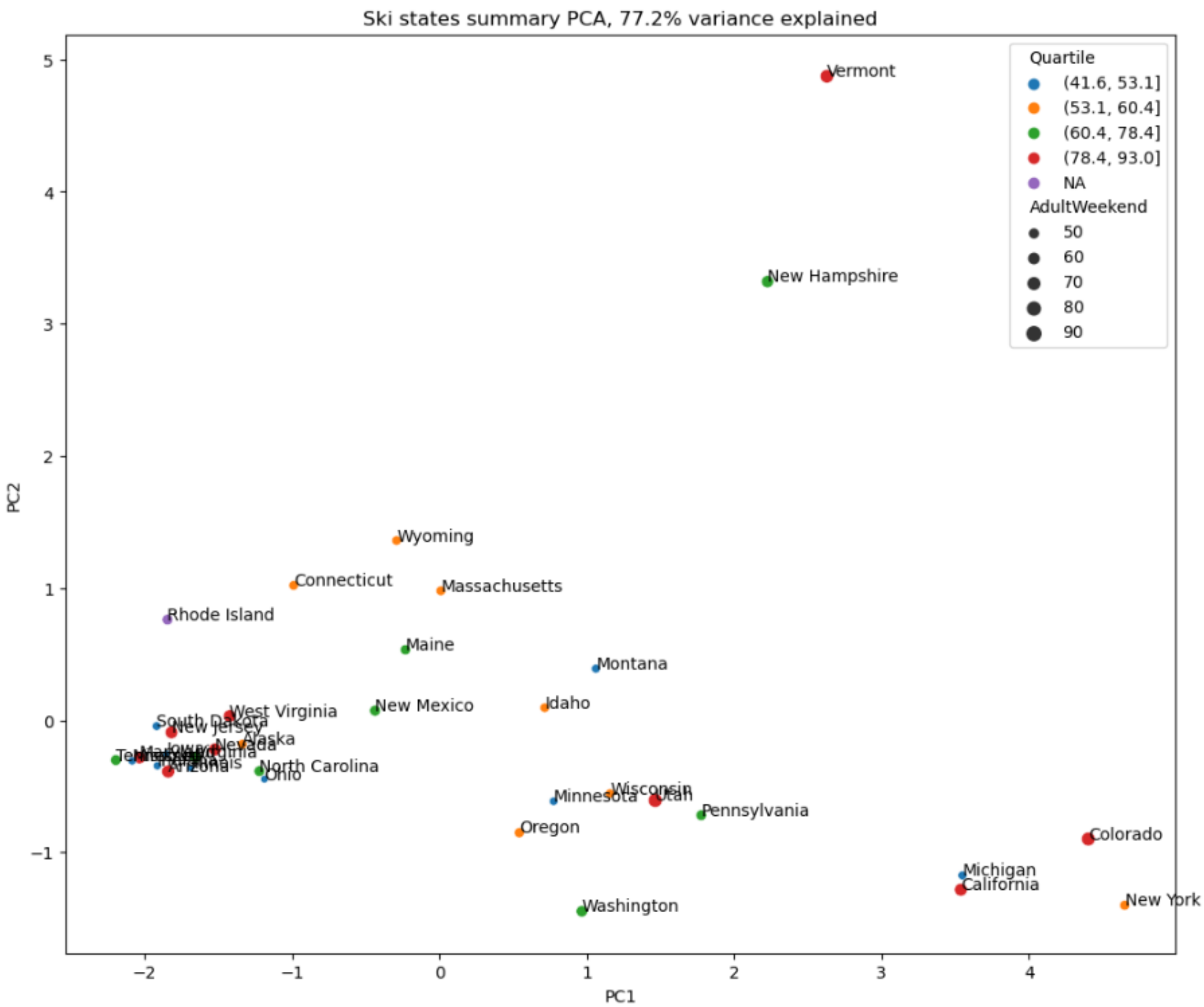

Figure 3: Price distribution for two dominant components show no clear pattern highlighting state labels can be ignore for price modeling

Lastly after doing correlation analysis between ticket price and other features and then plotting scatter plots and heatmap, we concluded that the most important features to consider for modeling the ticket prices would be vertical_drop, fastQuads, runs, total_chairs, snow making_ac, LongestRun_mi and skiableTerrain. The figure below show scatter plots between ticket price and individual features

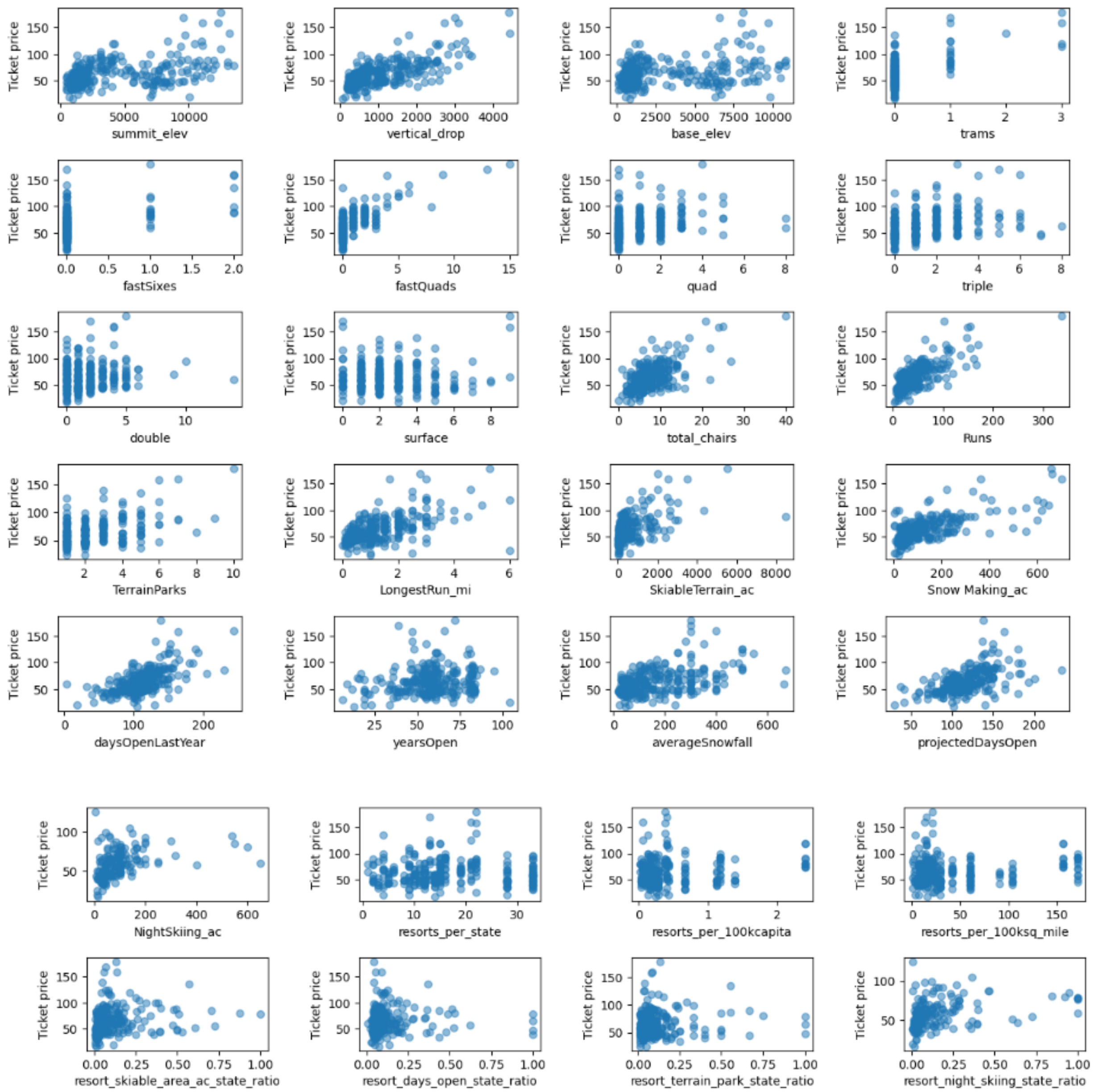

Figure 4: Scatterplots show high correlations of ticket prices with `vertical_drop`, `fastQuads`, `Runs` and `total_chairs`

Preprocessing & Training

After establishing the base model performance by using average value, we developed and analyzed the performance of two models - Linear regression based and Random forest based. To avoid overfitting, we identified 8 features using k-feature functionality.

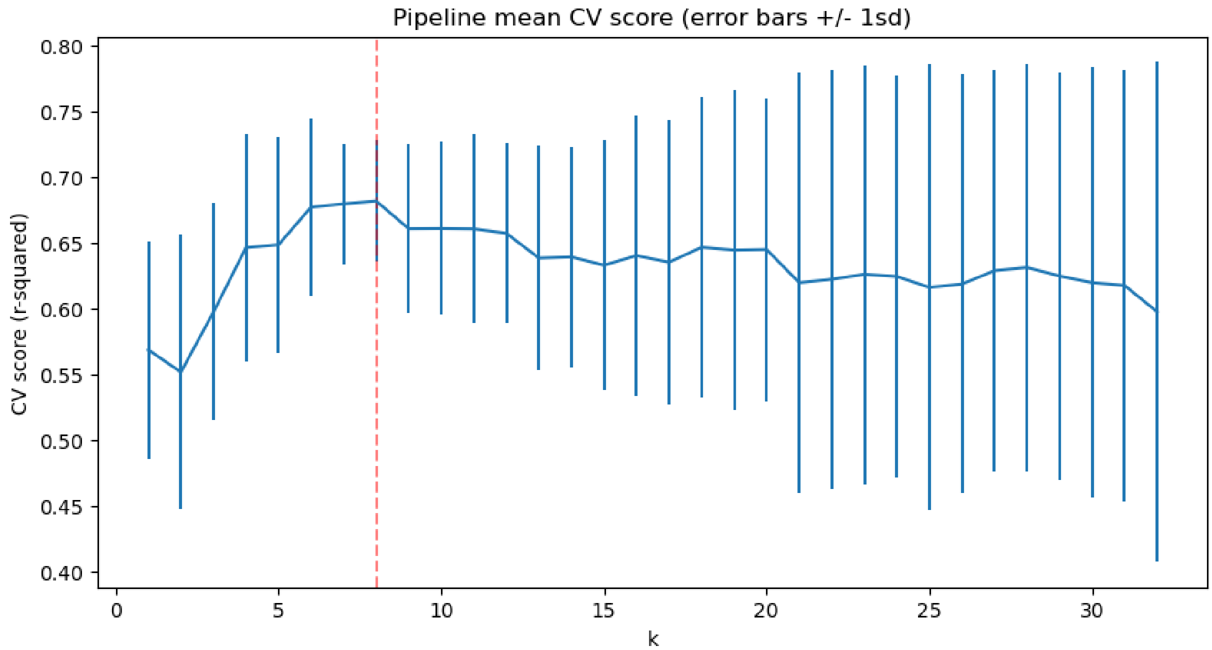

Figure 5: Figure showing 8 is a good value for number of features to be modeled

After measuring performance for each **we finalized the Random forest model for price estimation** as it had the least cross-validation mean absolute error and exhibited least variability based on performance on the test set. Here is a summary of the performance of the two models:

Model	**Performance**
Linear Model	<u>Cross Validation</u> Mean absolute error: 10.499032338015294 Absolute Error Standard deviation: 1.622060897679966 <u>Test Set</u> Mean Absolute Error: 11.793465668669326
Random forest	<u>Cross Validation</u> Mean absolute error: 9.644639167595688 Absolute Error Standard deviation: 1.3528565172191818 <u>Test Set</u> Mean Absolute Error: 9.537730050637332

Scenario Modeling

We modeled four scenarios as described below with the corresponding results:

Scenario	Result
Closing down runs	The model says closing one run makes no difference in price, thus there is an opportunity to save the cost; Closing 2 and 3 runs 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 ticket price. Increasing the closures down to 6 or more leads to a large drop
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 snow making coverage	Adding a vertical drop by 150 feet may allow an increase in price by $1.99 doing, leading to a profit increase by ~$2 MM assuming 350,000 visitors per season
Same as above, but adding 2 acres of snow making cover	This scenario doesn't offer any benefits
Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres	This scenario doesn't offer any benefits

Pricing Recommendation

Based on the modeled price we concluded that Big Mountain should try to raise the prices based on the facilities it provides.

It can aim for $85.48 price ($95.87 - $10.39), which is the lowest price that the model estimated taking into account maximum uncertainty; though there is a chance it may lead to a loss in traffic given its price is highest in Montana. The new price should offset the additional chair cost based on the expected 350,000 visitors this season

Future work

We need to collect data on costs for maintaining additional runs for further evaluating the scenario for closing a run.

Additionally, we need data on customers' willingness to pay to consider any price change. Since Big Mountain's prices are highest in Montana, we should consider whether any price change may lead to a huge drop in customers as customers may already feel they are paying a premium. Big mountain resorts can do a short survey among potential customers to see how customers would perceive any price increase and make decisions accordingly.

Further, for the model to be used without the help of a data scientist, a UI should be built to get recommendations for different scenarios. This UI can then be easily used by both leadership, business analysts, and other stakeholders to check different scenarios using the model