Guided Capstone Project Report

1.Problem statement

Big Mountain Resort has recently installed an additional chair lift that increases their operating costs by $1,540,000 this season. Basing their pricing on just the market average does not provide the business with a good sense of how important some facilities are compared to others. This hampers investment strategy. So, we need to find out how to select a better value for their ticket price.

Recommend strategy in two weeks for Big Mountain resort based on scientific annual revenue forecast over the next year, realize to recoup the increased operational cost of $1.54MM for installing new chair list this season, and keep the profit margins at 9.2% after executing the strategy.

2.Data Wrangling

We dealt with missing values and checked the mistakes in the data.

The original number of rows is 277 and a total of 27 columns. I found that Big Mountain Resort was present, which doesn't appear to have any missing values. Fast Eight has the most missing values, at just over 50%.

There is wrong record of Silverton Mountain'SkiableTerrain\_ac that is 26819.0. But the value I've just looked up is 1819. So I updated 26819.0 with 1819.

We removed rows from ski\_data where both price values are missing. And I added 'state\_population','state\_area\_sq\_miles' from the Internet to the data. I confirm that Weekend prices being higher than weekday prices seem restricted to sub $100 resorts. Finally, there are 277 rows left in the data.

3.Exploratory data Analysis

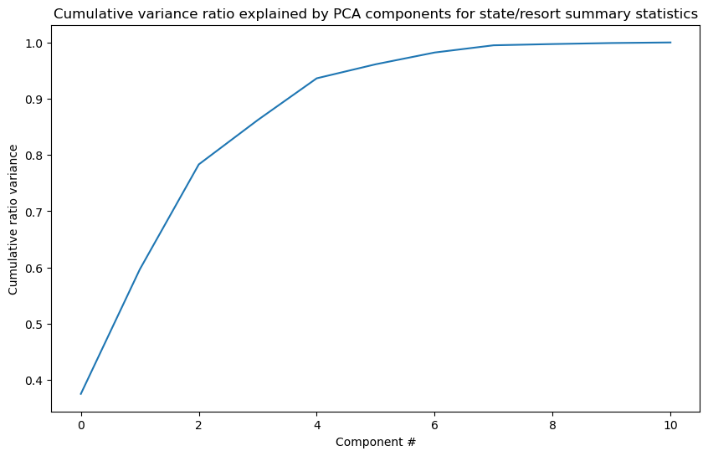
There are states that host many resorts, but other states host a larger total skiing area. The states with the most total days skiing per season are not necessarily those with the most resorts.

A graph with blue squares

Description automatically generatedA graph with blue squares

Description automatically generated

We adopted principle components analysis (PCA) to reduce dimensionality. The first two components seem to account for over 75% of the variance, and the first four for over 95%.

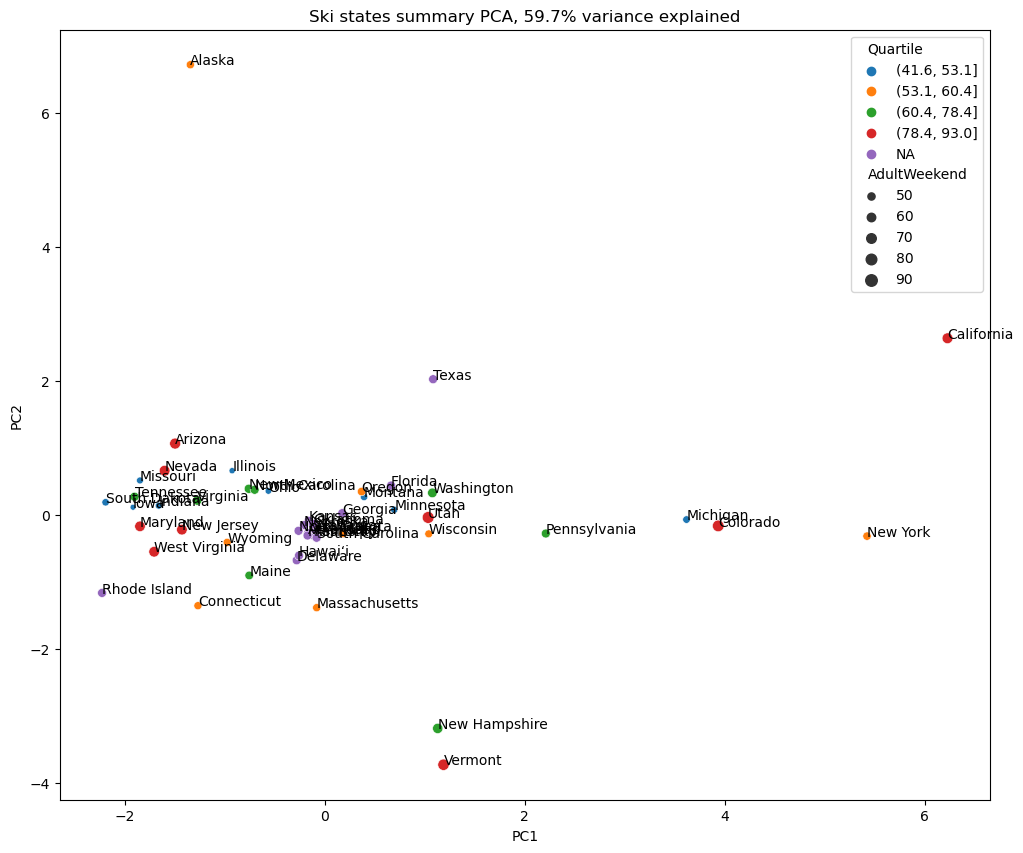


We can see distribution of state averaged prices.

A graph of a number of blue squares

Description automatically generated

We can see the same distribution of states as before, but with additional information about the average price. We haven't seen any clear grouping yet.



It looks like resorts\_per\_100kcapita and resorts\_per\_100ksq\_mile might count for quite a lot, in a positive sense.

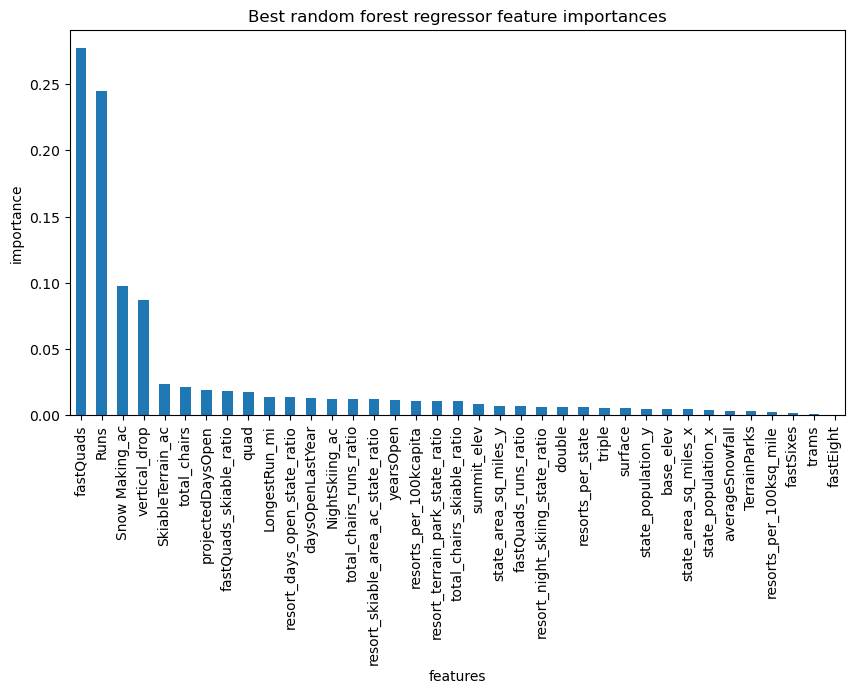
Based on feature correlation heatmap, visitors would seem to value more guaranteed snow, which would cost in terms of snow making equipment, which would drive prices and costs up. Of the new features, resort\_night\_skiing\_state\_ratio seems the most correlated with ticket price. If this is true, then perhaps seizing a greater share of night skiing capacity is positive for the price a resort can charge. As well as Runs, total\_chairs is quite well correlated with ticket price.

There's a strong positive correlation with vertical\_drop. fastQuads seems very useful. Runs and total\_chairs appear quite similar and also useful. Ticket price may drop a little before then climbing upwards as the number of resorts per capita increases. Ticket price could climb with the number of resorts serving a population.

 It seems that the more chairs a resort has to move people around, relative to the number of runs, ticket price rapidly plummets and stays low. It also appears that having no fast quads may limit the ticket price, but if your resort covers a wide area then getting a small number of fast quads may be beneficial to ticket price.

4.Model Preprocessing with feature engineering

It is important to extract Big Mountain Data and separate it from the rest of the data to use later. We train the model on the train split and assess model performance. We can build pipelines to make pre-processing and training data done. We refined the Linear Model and try Random Forest Model. And we make final model selection.



The dominant top four features are in common with your linear model:

* fastQuads
* Runs
* Snow Making\_ac
* vertical\_drop

5.Algorithms used to build the model with evaluation metric

We used 𝑅2 as a common metric to judge the expect performance on a test set. This is a measure of the proportion of variance in the dependent variable (our ticket price) that is predicted by our "model". Another common metric (and an important one internally for optimizing machine learning models) is the mean squared error. This is simply the average of the square of the errors. We used sklearn.metrics that provides many commonly used metrics, include R-squared,mean absoluted error,mean squared error.

6.Winning model and scenario modelling

We built the best linear model and a best random forest model. The random forest model has a lower cross-validation mean absolute error by almost $1. It also exhibits less variability. We chose random forest model at last.

The business has shortlisted some options in the scenarios as follows:

(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 snow making coverage.

(3) Same as number 2, but adding 2 acres of snow making cover.

(4) Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres.

These results suggest that vertical drop is our biggest positive feature, the area covered by snow making equipment is a strong positive as well. The skiable terrain area is negatively associated with ticket price, the data is missing information about visitor numbers. So, we cannot judge the reason based on the data now.

7.Pricing recommendation

Big Mountain currently charges adult ticket at weekend $81/per day. Big Mountain Resort modelled price is $96.39, even with the expected mean absolute error of $10.22, this suggests there is room for an increase.

The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue, so the price of ticket needs to reduce $1.5−$2.38. Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift. This scenario increases support for ticket price by $1.36. Over the season, this could be expected to save $2,386,364.

8.Conclusion

Based on now data, Big Mountain can consider closing one or 2 and 3 run to reduce costs and maintain the ticket price. Big Mountain can also raise the ticket price to $96.39 to increase the revenue.

It is assumed that the additional chair lift would increase the operating cost by $1.54 million. Big Mountain can install an additional chair lift, and charge an adding ticket price by $1.36, finally the investment can gain good profits after deducting costs.

9.Future scope of work

Now work based on limited data: The only price data in our dataset were ticket prices; we didn't know the operating cost of facilities except the additional operating cost of the new chair lift. We can raise the quality of the data by collecting more data about populations and attributes of customers and operating cost of different equipment. So, we can test more combinations of adjustments of price and operation optimization.

Check the reasons why its modeled price was so much higher than its current price, whether there were other factors might affect the price. We can report how key facilities support hight ticket price that will give some insights when the business executives make decision to choose high-end route (better service, better revenue).

Automate this part of the sensitivity analysis including more combinations of parameters in a scenario. So, business analysts can use it to explore new combinations.