BIG MOUNTAIN RESORT
FINAL REPORT

1. Problem Identification Overview

Context - Big Mountain Resort (BMR) is a ski resort in northwest Montana. They recently installed a new chair lift to help increase the distribution of visitors across the mountain. This increased their operating costs by $1,540,000. Every year about 350,000 people ski or snowboard at Big Mountain. The business profit margin is 9.2% and the investors would like to keep it there. The business wants recommendations on recouping the increased operating costs from the new chair this season.

In summary, the problem we are trying to address is: **“How can we increase this year’s profit in at least $1.540.000 to recover the increased operating costs resulting from the investment in a new chair lift?”**

This problem can be addressed from different perspectives. I took the route of **analyzing what can be the price change in ticket price**.

Our available data is a dataset with information about BMR and 302 other ski resorts across the nation. The characteristics we have from these ski resorts can be found in annex 1. Unfortunately, we have not received financial data which could help us in our analysis.

2. What deliverables are generated?

I generate 3 deliverables:

- The source code for the project with the models created to analyze the problem
- This document outlining the process
- A PowerPoint presentation with the suggestions to management.

3. Data Preprocessing steps

Data cleaning:

- Several features had missing values (see annex 2). I divided them between a) those that had “nan” values when they should have had the number 0 instead and b) those that truly had missing values.
- For the first case I changed “nan” for 0 and for the second case I filled the missing values with the mean of all the existing data. (Note: a better solution for group “B” would be to fill with the mean of resorts that share same traits).
- Region and State ended up being very similar features so I dropped down “Region” and kept “State”

Exploratory Data Analysis:

- EDA showed that there were several outliers. Here our study created two paths:
    - Creating a model without the outliers (176 rows and 26 columns)
    - Creating a model with the outliers (330 rows and 26 columns)
- Creation of clusters – I used K-Means unsupervised learning clustering technique to create 3 clusters.
- Annex 3 provides figures that explain the EDA process.

Pre-processing:

- I finally proceeded to create dummy variables for the “state” feature
- With our final dataset created I:

- Standardized the data using StandardScaler from sklearn
- Separated our data into the training and testing sets (75/25 respectively)

4. Model Description

The model I used for this analysis is a simple **linear regression**

5. Model Performance

I evaluated the model through explained variance and through mean absolute error

The two paths we took (removing outliers and keeping them) showed a relevant different in the results for the explained variance and MAE (mean absolute error). The following tables show this difference:

Keeping Outliers

Model	Explained Variance	Mean Absolute Error	Features Dropped
Model 1.	0.94	4.9	-
Model 2.	0.92	5.5	'state'
Model 3.	0.93	5.33	'state','summit_elev'

Removing Outliers

Model	Explained Variance	Mean Absolute Error	Features Dropped
Model 1.	error	error	-
Model 2.	0.53	7.37	'state'
Model 3.	0.53	7.34	'state','summit_elev'

Keeping the outliers seemed to offer us better results. This can be due to the fact that removing outliers resulted in some features disappearing from the analysis (like eight and six people chairs), that truly offer relevant information, like the boxplots in annex 3 shows, it is possible that these outliers are not result of a faulty data collection but to real characteristics, thus, we should keep them. Note: there is a case where this probably not true, it is in the “yearsOpen” variable in which there is an observation that shows a ski resort open for more than 3000 years!

As it can be seen, for each case we analyzed three different models:

- Keeping all of the data (including the dummy variables for state)
- Removing state
- Removing state and summit elevation

Results suggest that keeping the outliers and keeping all the features (model 1) offers us the best explained variance and MAE. This is followed closely by model 3. For the final step of the study I chose both of these models to fit the data from our ski resort.

6. Model Findings

I used the data we have for each feature for our ski resort and we fitted it in our model to determine the “Adult Weekend Price”. The current price is set at $81 what our models predict the price should be is:

- Model 1: $84.83
- Model 3: $89.04

Annex 4 provides several scatterplots that show how these new prices would fit in comparison to other resorts.

Let’s go to the question at hand and how this new information can help us cover the cost incurred with the new chairs.

Unfortunately, we have limited data but a quick financial analysis offers the following insights.

	Current Scenario	Scenario 1	Scenario 2	Scenario 3
Number of Skiers	350,000	350,000	350,000	350,000
Ticket Price	$ 81.00	$ 84.93	$ 89.04	$ 86.33
Approx. Revenues	$ 28,350,000	$ 29,725,500	$ 31,164,000	$ 30,216,117
Approx. Costs	$ 25,741,800	$ 26,990,754	$ 28,296,912	$ 27,435,800
Profit Margin	9.20%	9.20%	9.20%	9.20%
Cost difference		$ 1,248,954	$ 2,555,112	$ 1,694,000
CAPEX		$ 1,540,000	$ 1,540,000	$ 1,540,000
Approx. OPEX (10%)		$ 154,000	$ 154,000	$ 154,000
Difference		$ (445,046)	$ 861,112	$ -
New Profit Margin		7.70%	11.96%	9.20%

Knowing we want to keep the same profit margin for investors, we have limited options to play with*. One of the options is ticket price. The three proposed scenarios play with the ticket price. If we were to follow the suggestions of model 1 ($84.93), the profit margin would fall to 7.70%, if we followed model 3 ($89.04), the profit margin would increase to 11.96%. If we wanted to keep the same profit margin, we would need to put a price of around $86.33.

*Note: This is simplified model makes several assumptions that are relevant for our case study. I list them here:

- We assume that we are talking about Net profit margin not Gross or Operational profit margins.
- We expect that the number of skiers is kept the same as last year (350.000)
- No other income a part from ticket prices has been considered (like food sales)
- Ticket price increased was considered for the whole set of skiers while the analysis was only done for weekend ticket price. Initial figures show that BRM should consider reducing the weekly price.
- We assume that there will be ongoing costs of running the new chair, this could mean an increase in salaries, in utilities, in depreciation, and maybe other costs. We have made an approximation of 10% cost over the CAPEX.

Recommendation to management: Increase the price of the ticket to $86

Next Steps

Further analysis can be carried out. I suggest some follow up points divided by org. departments:

- Data Science department: Review more features e.g. weekday price or projected open days
- Marketing Department: Demand elasticity, advertisement expenses to promote new chair.
- Financial Department: Detailed financial analysis, estimation of increase in OPEX
- Operations Department: Queuing analysis, resource requirements

Annex 1 – Table with definitions of available features

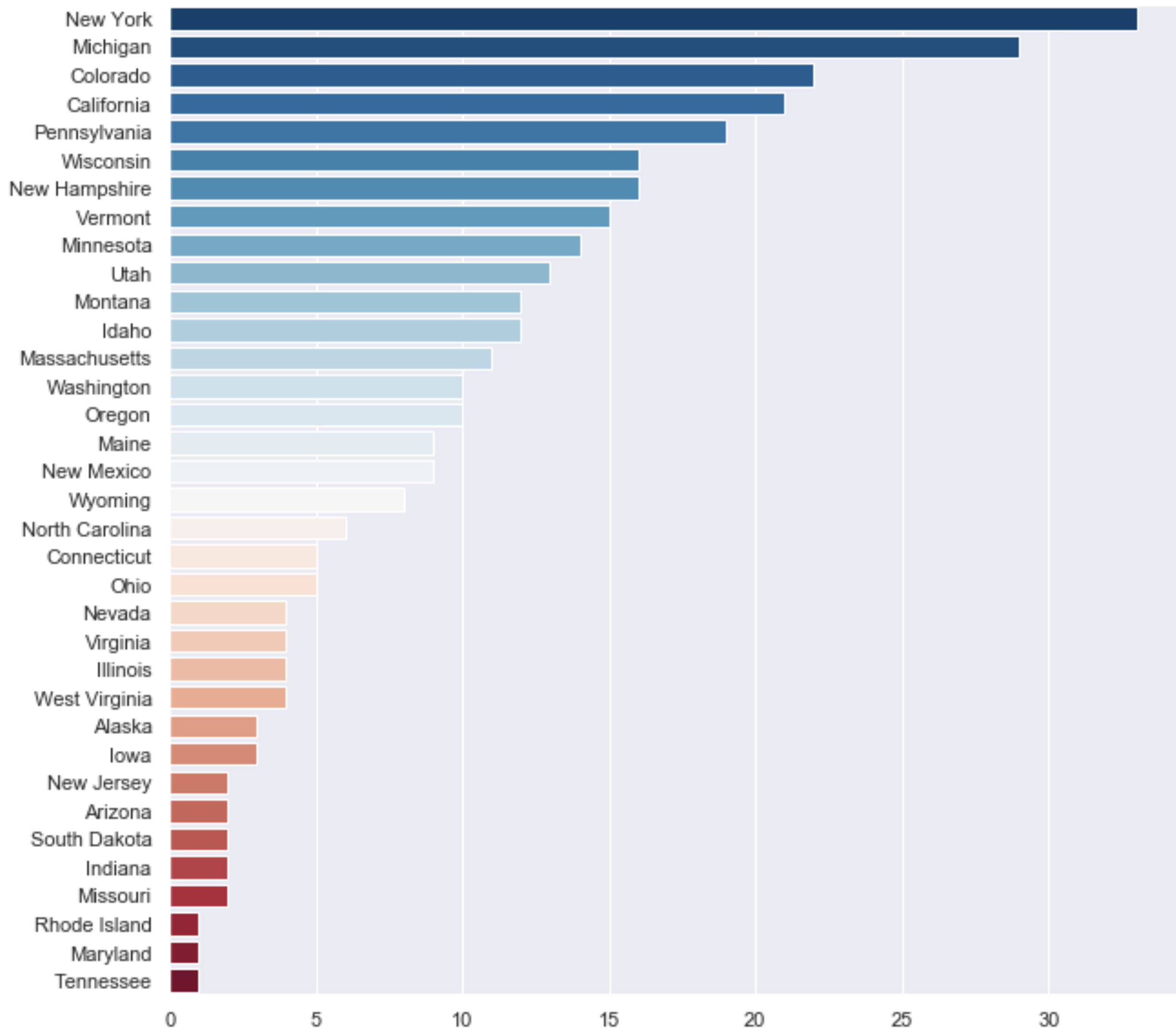
Column	Description
Name	The name of the ski resort.
Region	The region within the United States where the resort is located.
state	The state name where the resort is located.
summit_elev	Elevation in feet of the summit mountain at the resort.
vertical_drop	Vertical change in elevation from the summit to the base in feet.
base_elev	Elevation in feet at the base of the resort.
trams	The number of trams.
fastEight	The number of fast eight person chairs.
fastSixes	The number of fast six person chairs.
fastQuads	The number of fast four person chairs.
quad	Count of regular speed four person chairlifts.
triple	Count of regular speed three person chairlifts.
double	Count of regular speed two person chairlifts.
surface	Count of regular speed single person chairlifts.
total_chairs	Sum of all the chairlifts at the resort.
Runs	Count of the number of runs on the resort.
TerrainParks	Count of the number of terrain parks at the resort.
LongestRun_mi	Length of the longest run in the resort in miles.
SkiableTerrain_ac	Total skiable area in square acres.
Snow Making_ac	Total area covered by snow making machines in acres.
daysOpenLastYear	Total number of days open last year.
yearsOpen	Total number of years the resort has been open.
averageSnowfall	Average annual snowfall at the resort in inches.
AdultWeekday	Cost of an adult weekday chairlift ticket.
AdultWeekend	Cost of an adult weekend chairlift ticket.
projectedDaysOpen	Projected days open in the upcoming season.
NightSkiing_ac	Total skiable area covered in lights for night skiing.

Annex 2 – Percentage of missing values per feature

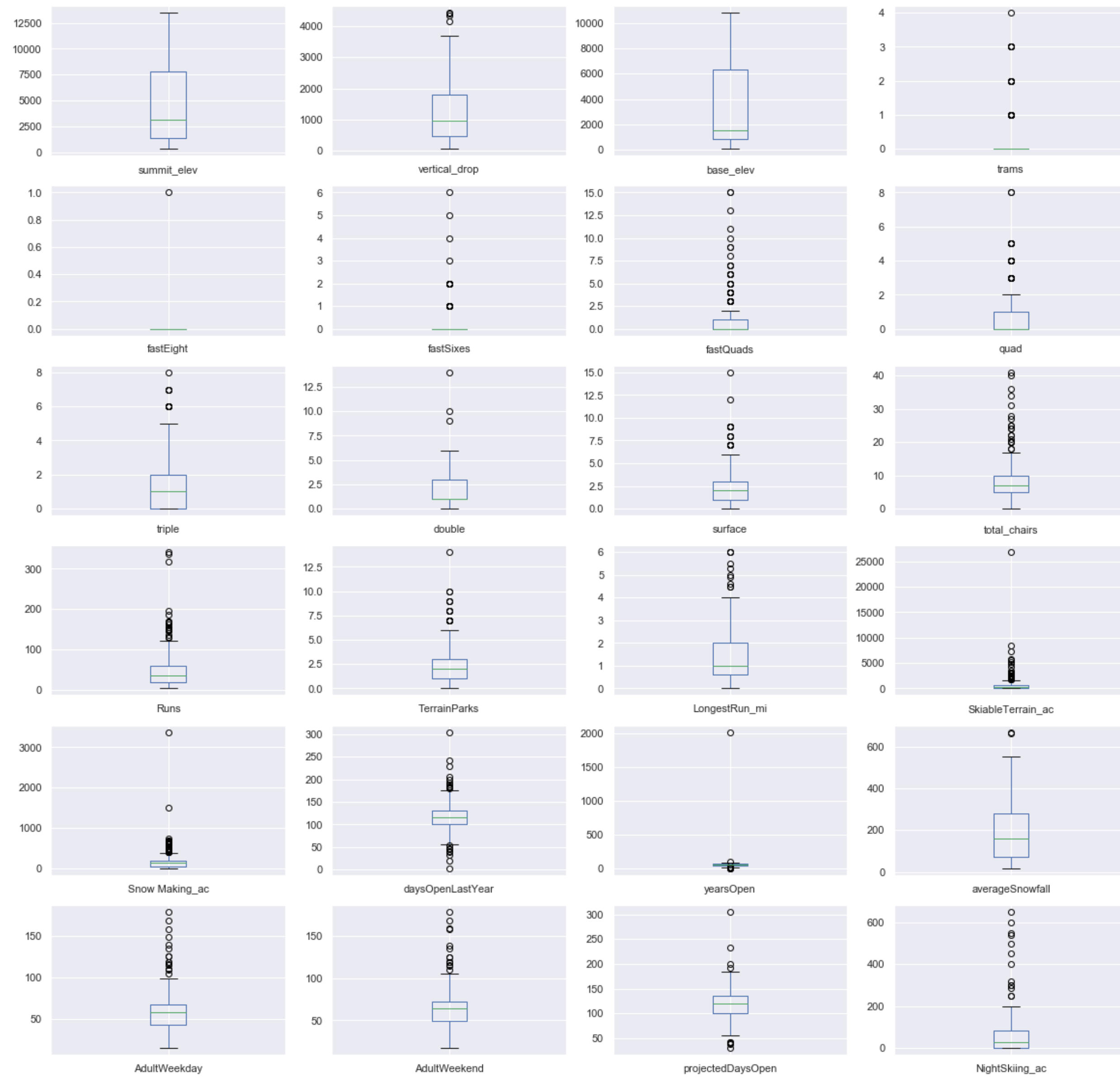
	percent
fastEight	0.503030
NightSkiing_ac	0.433333
AdultWeekday	0.163636
AdultWeekend	0.154545
daysOpenLastYear	0.154545
TerrainParks	0.154545
projectedDaysOpen	0.142424
Snow Making_ac	0.139394
averageSnowfall	0.042424
LongestRun_mi	0.015152
Runs	0.012121
SkiableTerrain_ac	0.009091
yearsOpen	0.003030

Annex 3 – EDA Figures

Number of Ski Resorts per State

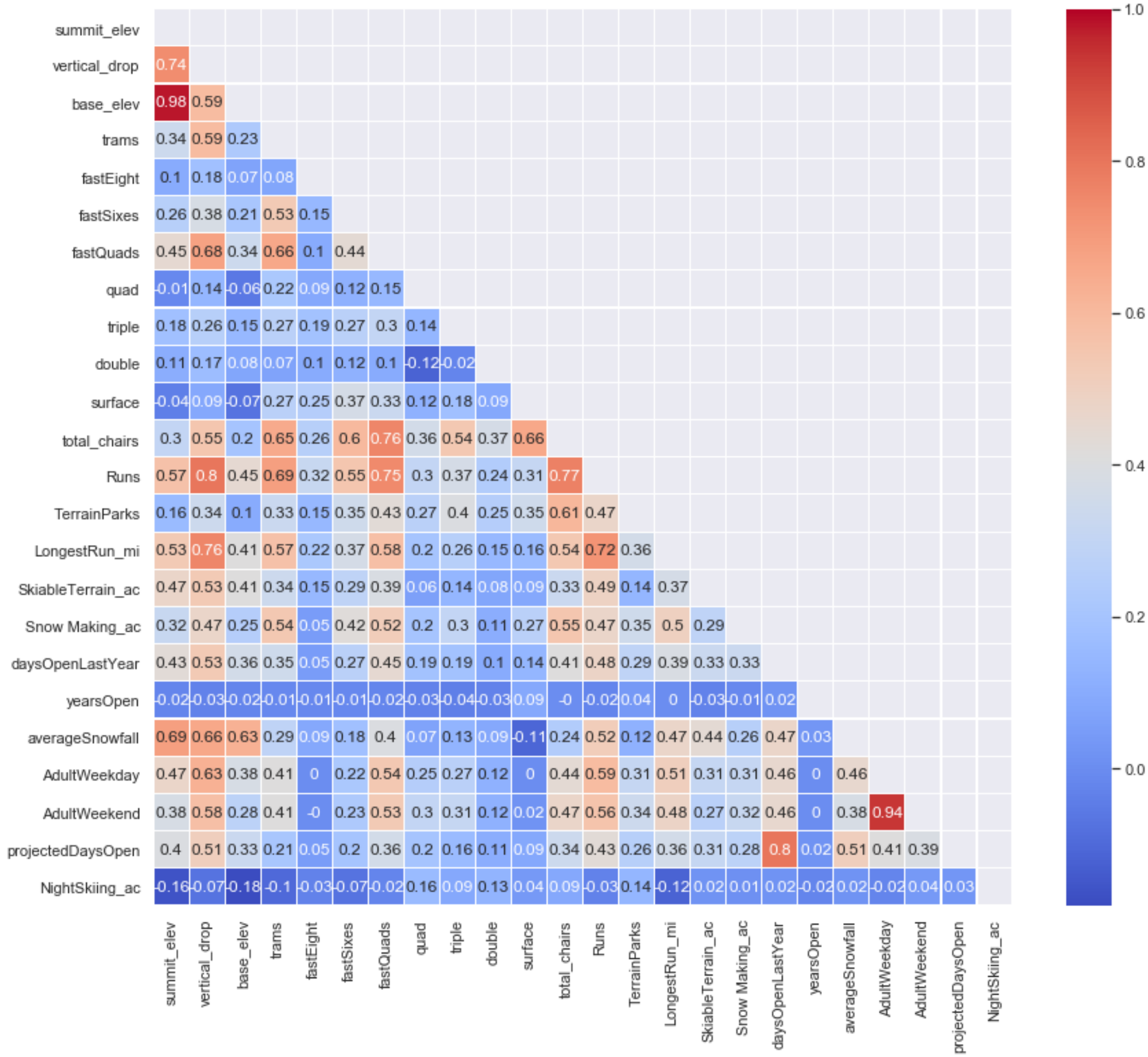


Boxplots before removing outliers

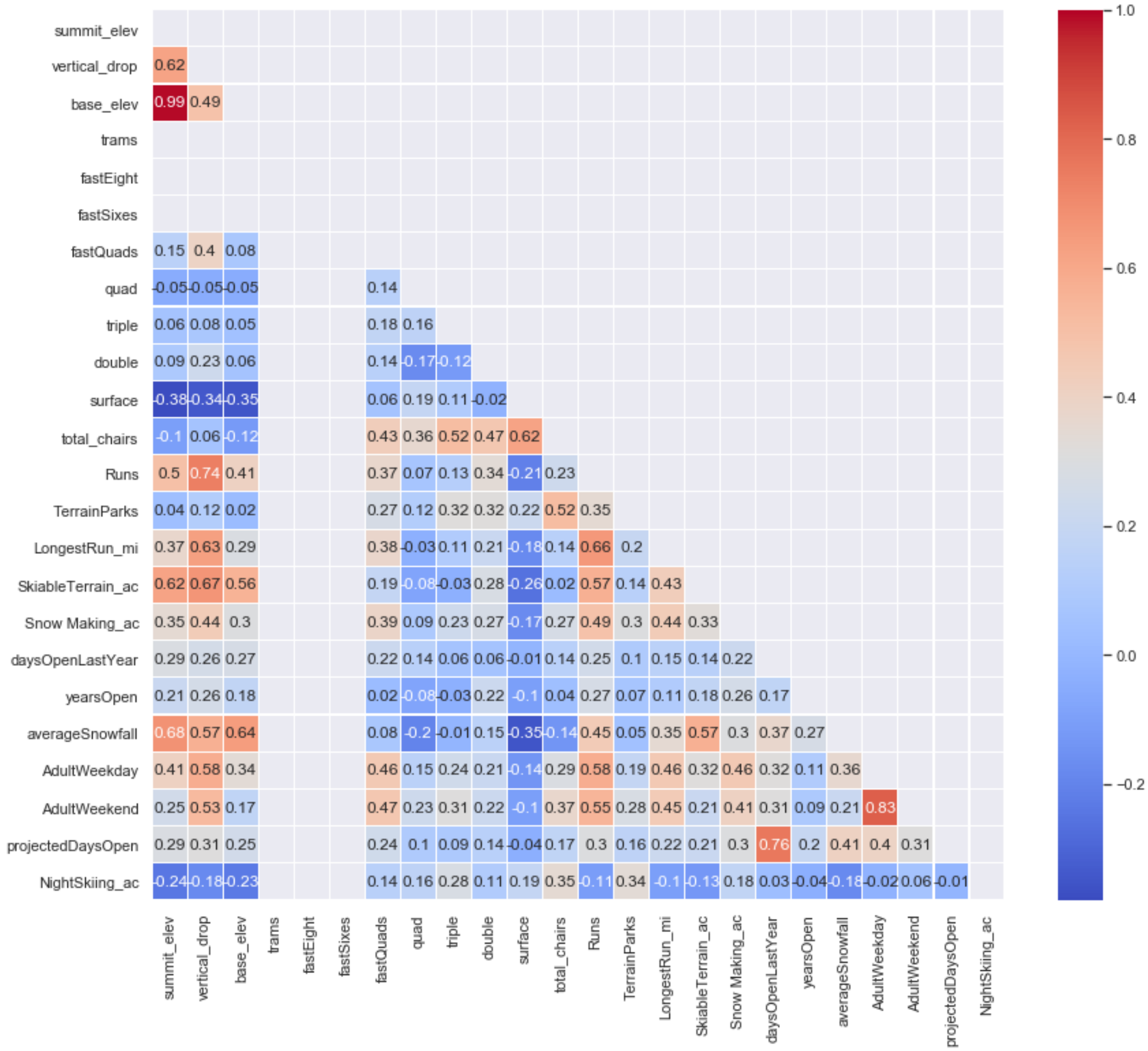


Boxplots after removing outliers (> 1.5 * IRQ <)

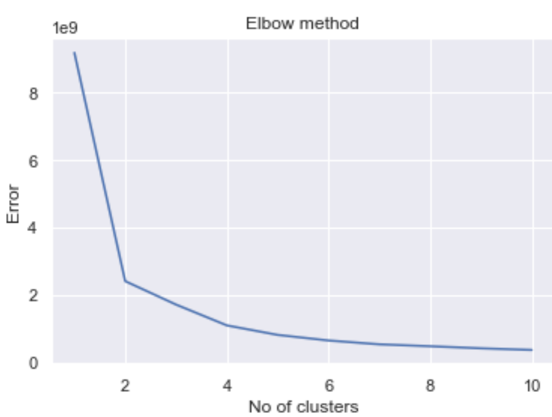
Correlation Matrix for data with Outliers



Correlation Matrix for data without Outliers

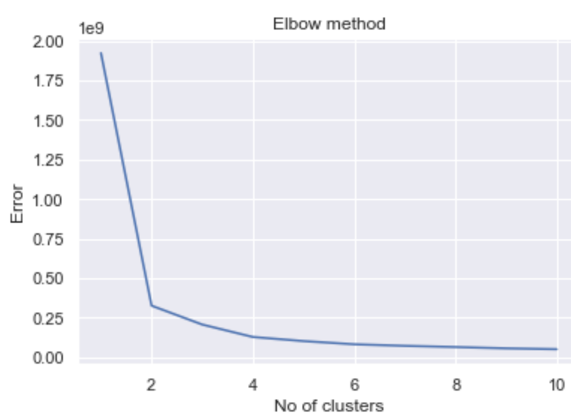


Elbow Chart with outliers

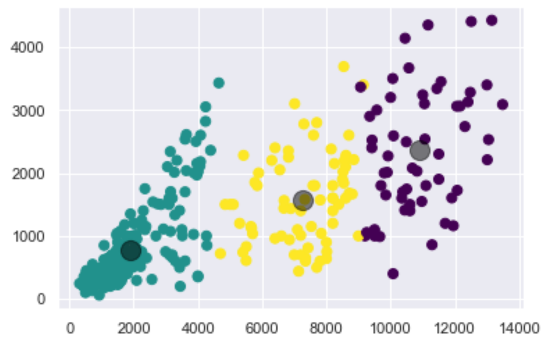


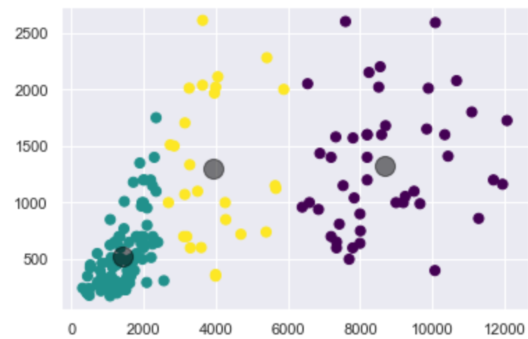
Clusters with outliers

Elbow Chart without outliers



Clusters without outliers

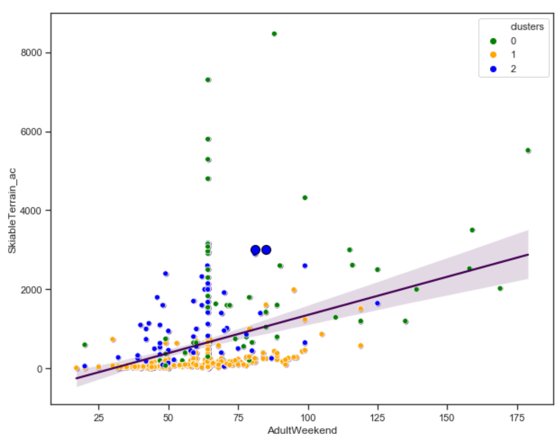




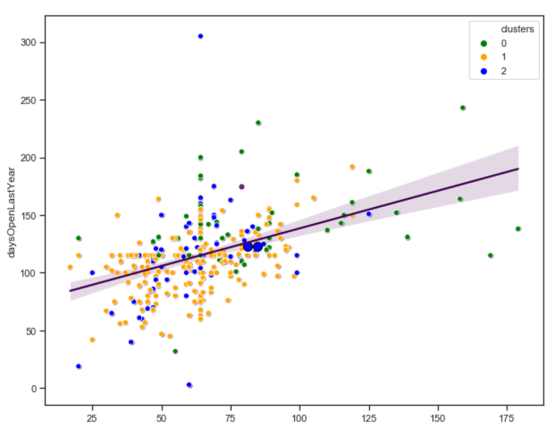
Annex 4 – Comparison BMR with other Resorts on three variables & 2 models

Comparison with other resorts (with states)

Skiable Terrain

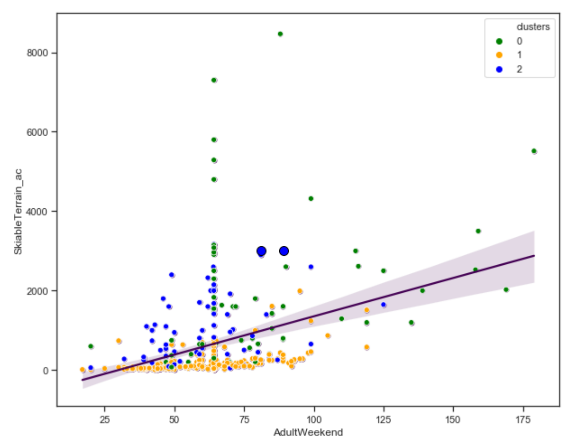


Days Open

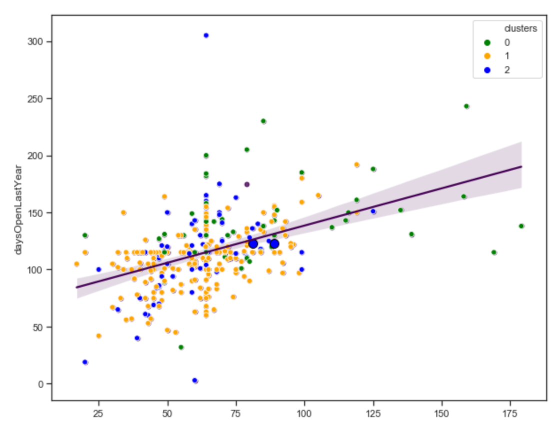


Comparison with other resorts (without states)

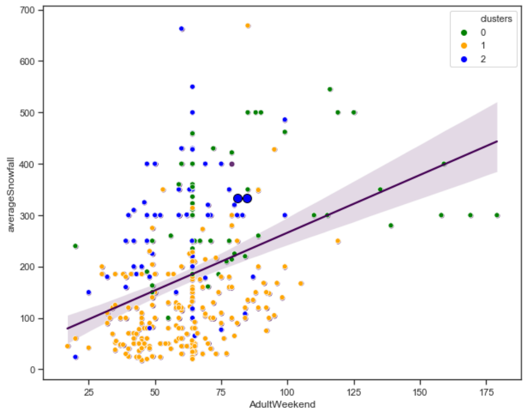
Skiable Terrain



Days Open



Average Snowfall



Average Snowfall