Big Mountain Resort Project

Author | Rand Sobczak

TABLE OF CONTENTS

01

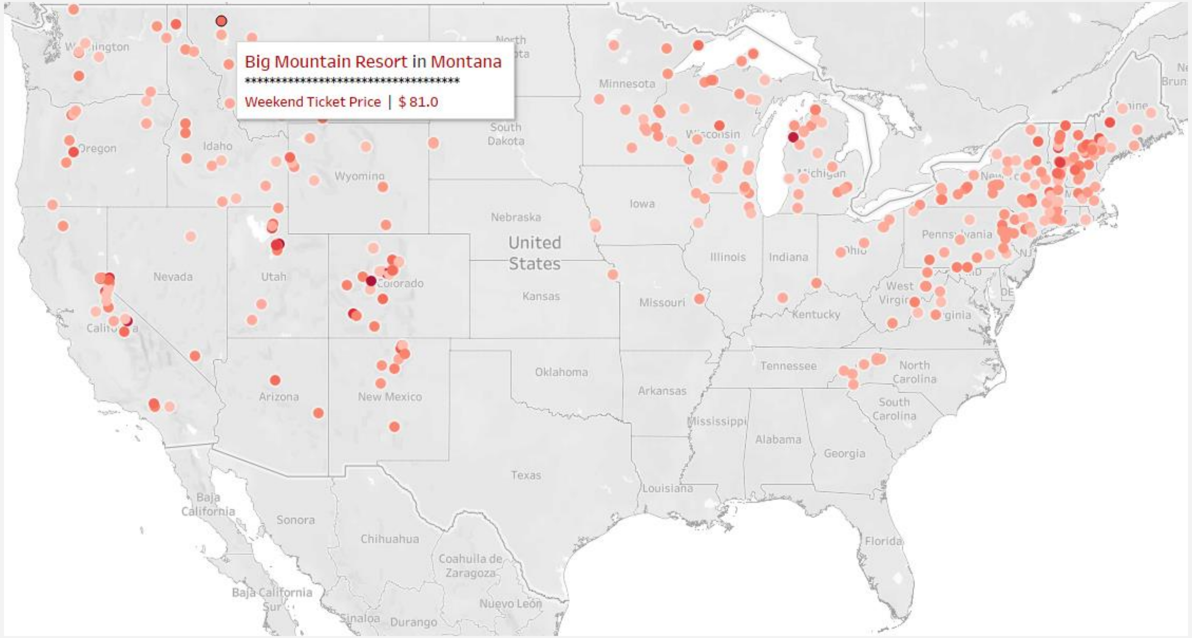
PROBLEM IDENTIFICATION

PROBLEM IDENTIFICATION

INCREASE PROFITS

Big Mountain Resort (BMR) is a ski resort in Montana. **BMR recently installed a new chair lift which increased OPEX by ~$1.54M per annum.**

The problem we are addressing is, **how can we increase profits**, starting this year, **by a minimum of +$1.54M per annum to offset the costs of the new chair lift?**



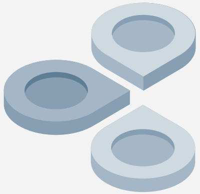
** Three (3) resorts in Alaska not shown on the map.*

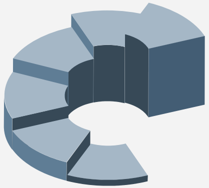
02
GENERAL DELIVERABLES

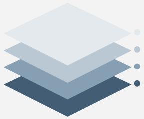
GENERAL DELIVERABLES

THREE (3) DELIVERABLES GENERATED

1. The **source code** for the modeling developed to analyze the aforementioned problem. ( )
2. A **report** found here. ( )
3. This **presentation outlining our advice to mgmt to potentially address the problem.**







GENERAL DELIVERABLES

SOURCE CODE DEFINED

Name	The name of the resort.
Region	The region within the United States where the resort is located.
state	The state within the United States where the resort is located.
summit_elev	The elevation in feet of the summit mountain at the resort.
vertical_drop	The 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 operational trams.
fastEight	The number of fast eight chairs.
fastSixes	The number of fast six chairs.
fastQuads	The number of fast four chairs.
quad	The number of regular speed four person chairs.
triple	The number of regular speed three person chairs.
double	The number of regular speed two person chairs.
surface	The number of regular speed single person chairs.
total_chairs	The number of chairlifts at the resort
Runs	The total number of runs at the resort.
TerrainParks	The total number of terrain parks at the resort.
LongestRun_mi	The length of the longest run at the resort in miles.
SkiableTerrain_ac	The total amount of skiable acres at the resort.
Snow Making_ac	The total number of acres covered by snow making machines.
daysOpenLastYear	The total number of days open last year.
yearsOpen	The total number of years the resort has been open.
averageSnowfall	Average annual snow fall at the resort in inches.
AdultWeekday	Cost of an adult weekday chairlift ticket
AdultWeekend	Cost of an adult weekend chairlift ticket
projectedDaysOpen	The projected number of days open in the upcoming season.
NightSkiing_ac	Total number of acres covered by light for night skiing.

03

DATA PRE-PROCESSING STEPS

DATA PRE-PROCESSING

DATA CLEANING

1. I was provided **a single CSV file** by the Database Manager.
2. I amended SkiableTerrain_ac for a resort* as the data was noticeably off kilter with the dataset & did not comply with the data on their website (🌐); I used the website's data.
3. The **fastEight column was removed as 50%** of the resorts **had no values.**

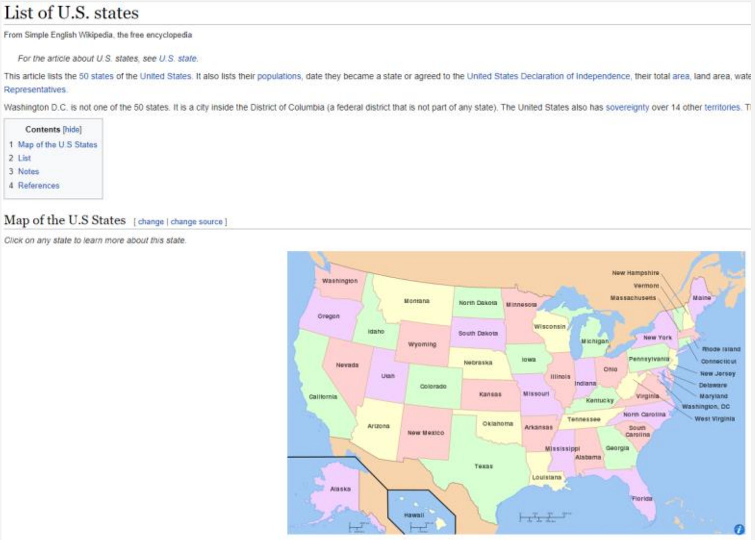
** The resort is Silver Mountain, based in Colorado*

	Name	Region	state	summit_elev	vertical_drop	base_elev	trams	fastEight
0	Alyeska Resort	Alaska	Alaska	3939	2500	250	1	0.0
1	Eaglecrest Ski Area	Alaska	Alaska	2600	1540	1200	0	0.0
2	Hilltop Ski Area	Alaska	Alaska	2090	294	1796	0	0.0
3	Arizona Snowbowl	Arizona	Arizona	11500	2300	9200	0	0.0
4	Sunrise Park Resort	Arizona	Arizona	11100	1800	9200	0	NaN

	count	%
fastEight	166	50.303030
NightSkiing_ac	143	43.333333
AdultWeekday	54	16.363636
AdultWeekend	51	15.454545
daysOpenLastYear	51	15.454545
TerrainParks	51	15.454545
projectedDaysOpen	47	14.242424
Snow Making_ac	46	13.939394
averageSnowfall	14	4.242424
LongestRun_mi	5	1.515152
Runs	4	1.212121
SkiableTerrain_ac	3	0.909091
yearsOpen	1	0.303030
total_chairs	0	0.000000
Name	0	0.000000
Region	0	0.000000
double	0	0.000000
triple	0	0.000000
quad	0	0.000000
fastQuads	0	0.000000
fastSixes	0	0.000000
trams	0	0.000000
base_elev	0	0.000000
vertical_drop	0	0.000000
summit_elev	0	0.000000
state	0	0.000000
surface	0	0.000000

```
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Name               330 non-null    object
 1   Region             330 non-null    object
 2   state              330 non-null    object
 3   summit_elev        330 non-null    int64
 4   vertical_drop      330 non-null    int64
 5   base_elev          330 non-null    int64
 6   trams              330 non-null    int64
 7   fastEight          164 non-null    float64
 8   fastSixes          330 non-null    int64
 9   fastQuads          330 non-null    int64
 10  quad               330 non-null    int64
 11  triple             330 non-null    int64
 12  double             330 non-null    int64
 13  surface            330 non-null    int64
 14  total_chairs       330 non-null    int64
 15  Runs               326 non-null    float64
 16  TerrainParks       279 non-null    float64
 17  LongestRun_mi      325 non-null    float64
 18  SkiableTerrain_ac  327 non-null    float64
 19  Snow Making_ac     284 non-null    float64
 20  daysOpenLastYear   279 non-null    float64
 21  yearsOpen          329 non-null    float64
 22  averageSnowfall    316 non-null    float64
 23  AdultWeekday       276 non-null    float64
 24  AdultWeekend       279 non-null    float64
 25  projectedDaysOpen  283 non-null    float64
 26  NightSkiing_ac     187 non-null    float64
dtypes: float64(13), int64(11), object(3)
```

DATA PRE-PROCESSING


List of U.S. states

From Simple English Wikipedia, the free encyclopedia

For the article about U.S. states, see U.S. state.

This article lists the 50 states of the United States. It also lists their populations, date they became a state or agreed to the United States Declaration of Independence, their total area, land area, wate
Representatives.

Washington D.C. is not one of the 50 states. It is a city inside the District of Columbia (a federal district that is not part of any state). The United States also has sovereignty over 14 other territories. T

Contents [hide]
1 Map of the U.S States
2 List
3 Notes
4 References

Map of the U.S States [change | change source]

Click on any state to learn more about this state.


```
0     82.317073
2     14.329268
1      3.353659
dtype: float64
```

	Name	Region	state	yearsOpen
34	Howelsen Hill	Colorado	Colorado	104.0
115	Pine Knob Ski Resort	Michigan	Michigan	2019.0

DATA CLEANING (cont.)

4. Our goal is to predict prices & as such, **I removed ~14% of the resorts as their price data* was incomplete**; thus not useful.
5. Wikipedia ( ) was used to directly draw **population & state size data to establish per capita & acre metrics.**
6. For one of the resort's columns, **data pertaining to the total years open was 2019**. **This was removed as that data is unlikely to be true.****

** Price data = AdultWeekday & AdultWeekend*

*** The column's name is yearsOpen; it was probably open in 2019 as ski resorts haven't been open for 2019 years in the USA*

DATA PRE-PROCESSING

EXPLORATORY DATA ANALYSIS

Two datasets were pulled from the previous step; namely:

1. **ski_resort_data**:
   - provided by the Database Manager on information which was cleaned down to 277 resorts across 34 states
2. **state_summary**:
   - an assembled dataset from Wikipedia which provided high level population / state size information

Name	Region	state	summit_elev	vertical_drop	base_elev	trams	fastEight	fastSixes	fastQuads	...
Alyeska Resort	Alaska	Alaska	3939	2500	250	1	0.0	0	2	...
Eaglecrest Ski Area	Alaska	Alaska	2600	1540	1200	0	0.0	0	0	...
Hilltop Ski Area	Alaska	Alaska	2090	294	1796	0	0.0	0	0	...
Arizona Snowbowl	Arizona	Arizona	11500	2300	9200	0	0.0	1	0	...
Sunrise Park Resort	Arizona	Arizona	11100	1800	9200	0	NaN	0	1	...

DATA PRE-PROCESSING

EXPLORATORY DATA ANALYSIS (cont.)

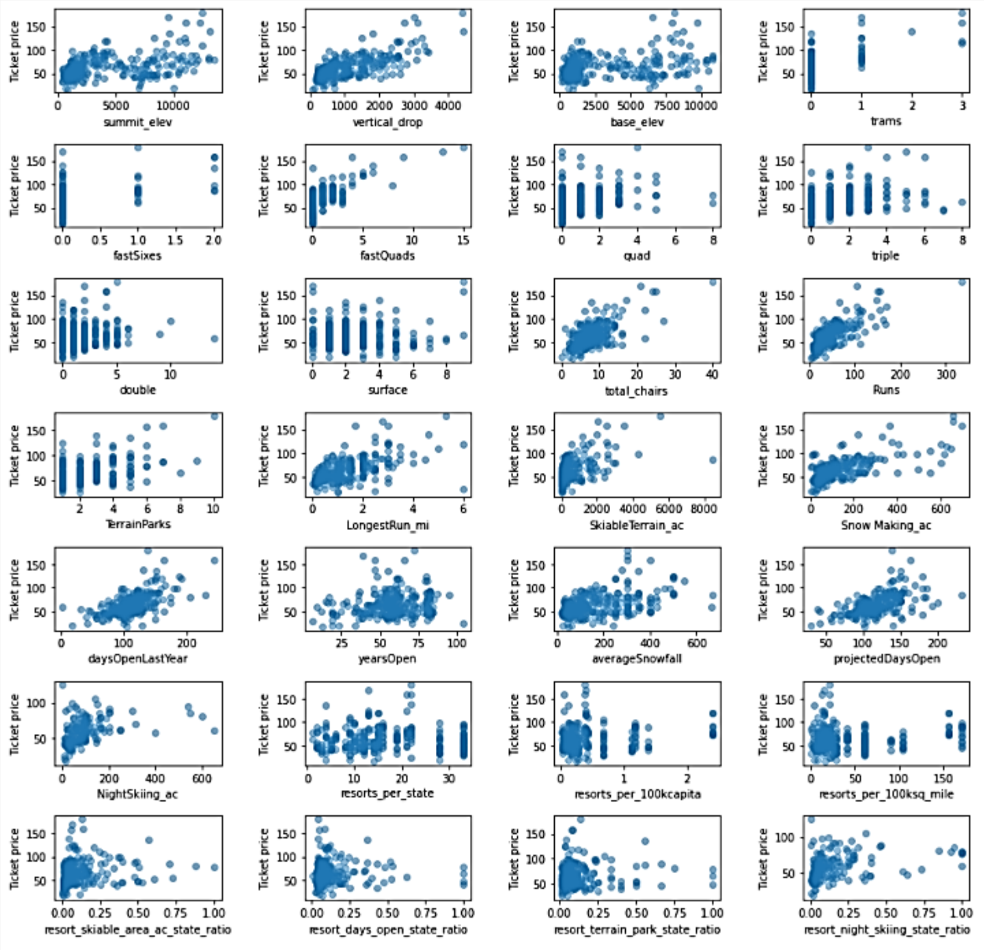
The **categorical features** are:

- The resort names
- Respective region (which did not appear useful)
- The state in which the resort is located in

Numerical features inside the 2 datasets **are plentiful** & additional variables (x7) were added. **Initial correlations with Adult Weekend prices were drawn** (next page).

DATA
PRE-PROCESSING

EXPLORATORY DATA ANALYSIS (cont.)



DATA PRE-PROCESSING

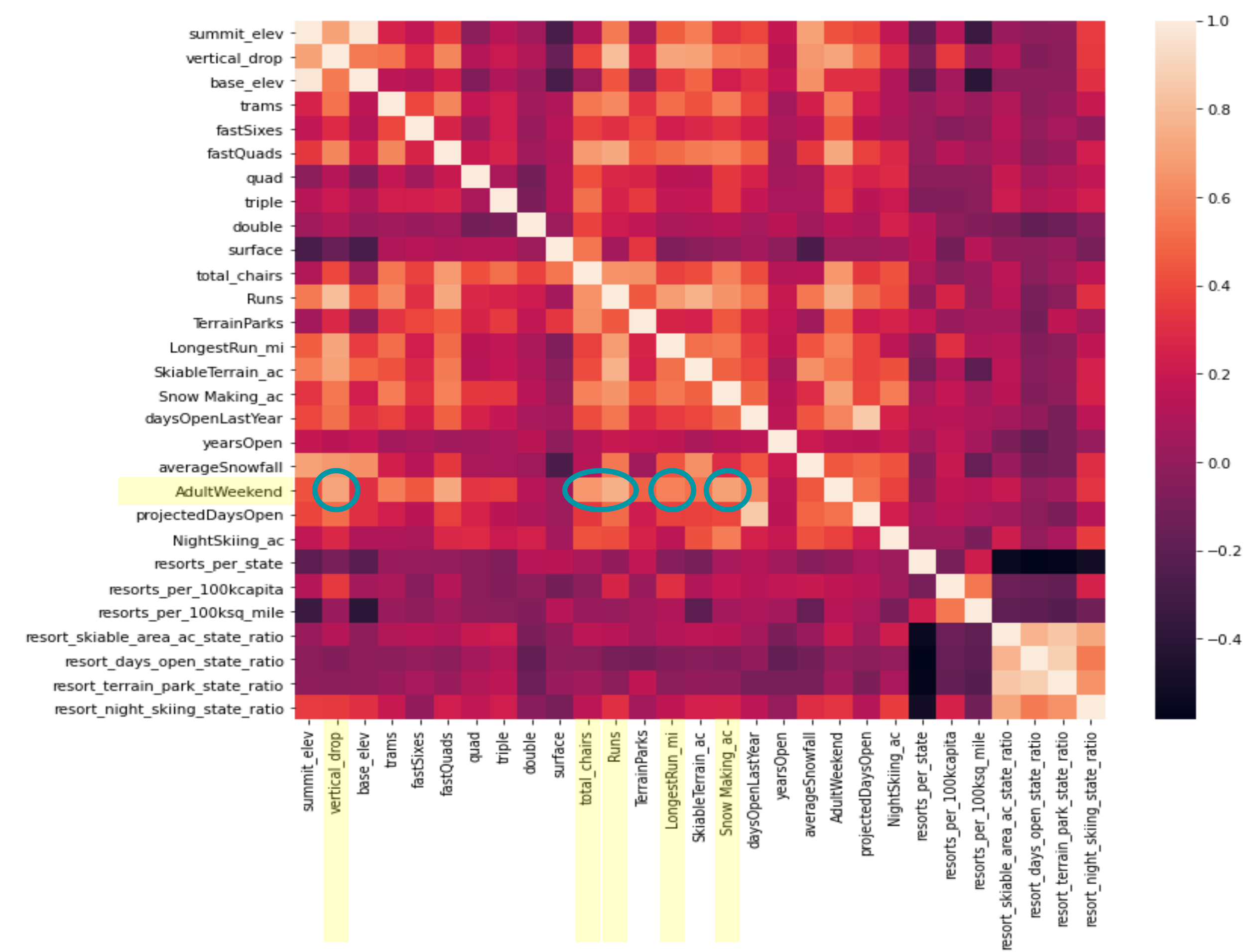
EXPLORATORY DATA ANALYSIS (cont.)

Using the Feature Correlation Heatmap (next page) **I identified numerical features that have the strongest correlations to AdultWeekend prices** (easier to read than previous page) **& may provide valuable insights to achieve our goal of higher prices**. The **below are the strongest positive correlations** & the possible justification for the relationships:

- **Vertical Drop**
  - *Desire for speed*
- **Total Chairs**
  - *Limited wait times*
- **Runs**
  - *Value on different run options*
- **Longest Run**
  - *Desire for being in motion longer*
- **Snow Making acres**
  - *Some guarantee of snow*

DATA PRE-PROCESSING

EXPLORATORY DATA ANALYSIS (Feature Correlation Heatmap)



DATA PRE-PROCESSING

EXPLORATORY DATA ANALYSIS (cont.)

Issues of the data & it's visualization that I addressed modeling include but are not limited to:

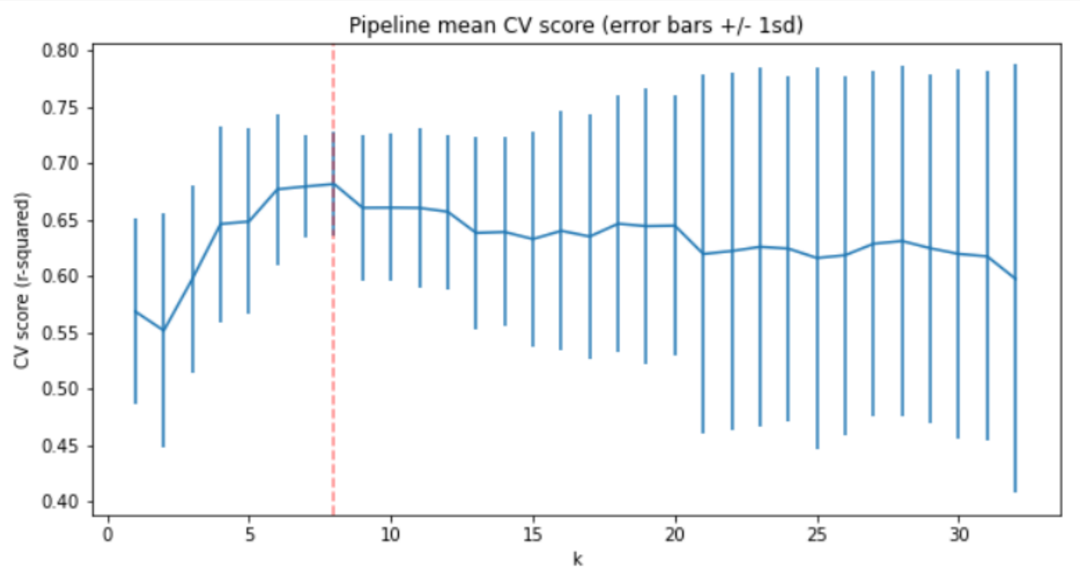
- I **filled in missing numbers with averages** of others to address missing data which may throw off scatter plots & quartile numbers
- I **utilized seaborn rather than** relying on **matplotlib** for scatter plots which may not be the ideal display candidate
- I **scaled the numbers** when required

DATA PRE-PROCESSING

PRE-PROCESSING

1. The **first** pre-processing step was to **split the data** into a training and testing splits, 70% / 30% respectively, & removing any object types so that the dataset is only composed of numeric values.
2. **With the help of a standard mean & the DummyRegressor functions**, I then **determined the "Best Guess"** number which was determined **to be 63.81.**
3. **Using the mean absolute error**, I determined that on average **the metric may be off by ~$19.**
4. I then **scaled the data** to ensure that the data can work better together.
5. Once scaled, the **train & test R squared became ~81% & 72%**. The **mean absolute errors dropped down to ~$9**; much better than the $19 unscaled guess.
6. I also replaced any missing data but the results weren't much different.

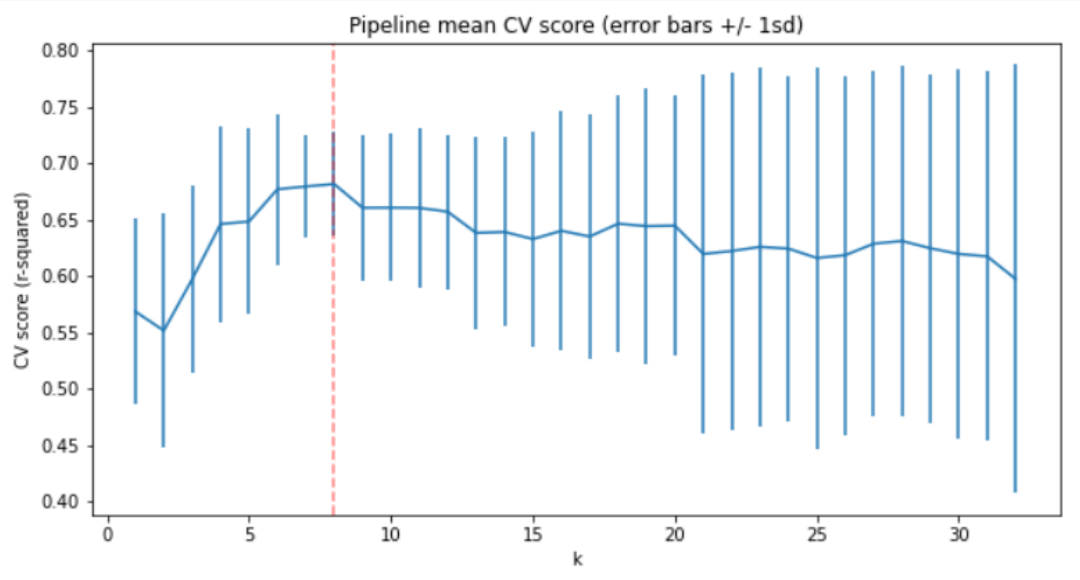
DATA PRE-PROCESSING (cont.)




```
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
dtype: float64
```

7. Next, I ran the training & testing splits through the **Pipeline function**. With **cross-validation**, I determined that the **best k value is 8.**
8. Using the linear coefficient numbers for each item versus AdultWeekend price, the **most positively correlated item was the vertical_drop**

DATA PRE-PROCESSING (cont.)



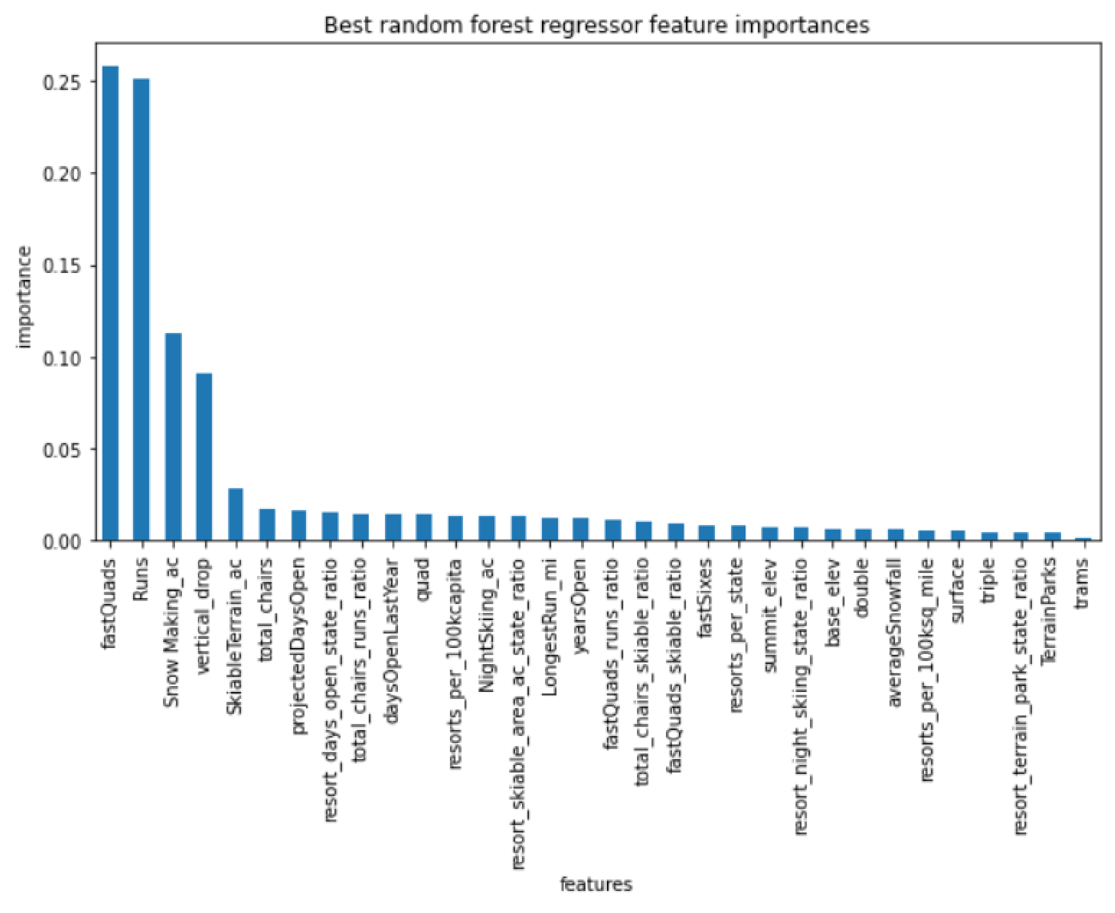

```
vertical_drop          10.767857
Snow Making_ac          6.290074
total_chairs            5.794156
fastQuads               5.745626
Runs                    5.370555
LongestRun_mi           0.181814
trams                  -4.142024
SkiableTerrain_ac      -5.249780
dtype: float64
```

7. Next, I ran the training & testing splits through the Pipeline function. With cross-validation, I determined that the best k value is 8.
8. Using the linear coefficient numbers for each item versus AdultWeekend price, the most positively correlated item was the vertical_drop & **the most negatively correlated item was SkiableTerrain_ac**
   - Vertical Drop we touched on earlier (speed), but **Skiable Terrain** (acres) **brings thoughts about** a potential "**exclusivity**" consideration
     - *Touch on this later*

DATA PRE-PROCESSING (cont.)

9. I then **moved to a Random Forest Generator** which **marginally improved our cross-validation results** & **showed that the top four dominate features were similar to those in the linear model.**
10. I **chose the Random Forest Model for my modeling, as it**:
    - Has a **lower cross-validation mean absolute error** ~$1
    - Exhibits **less variability.**

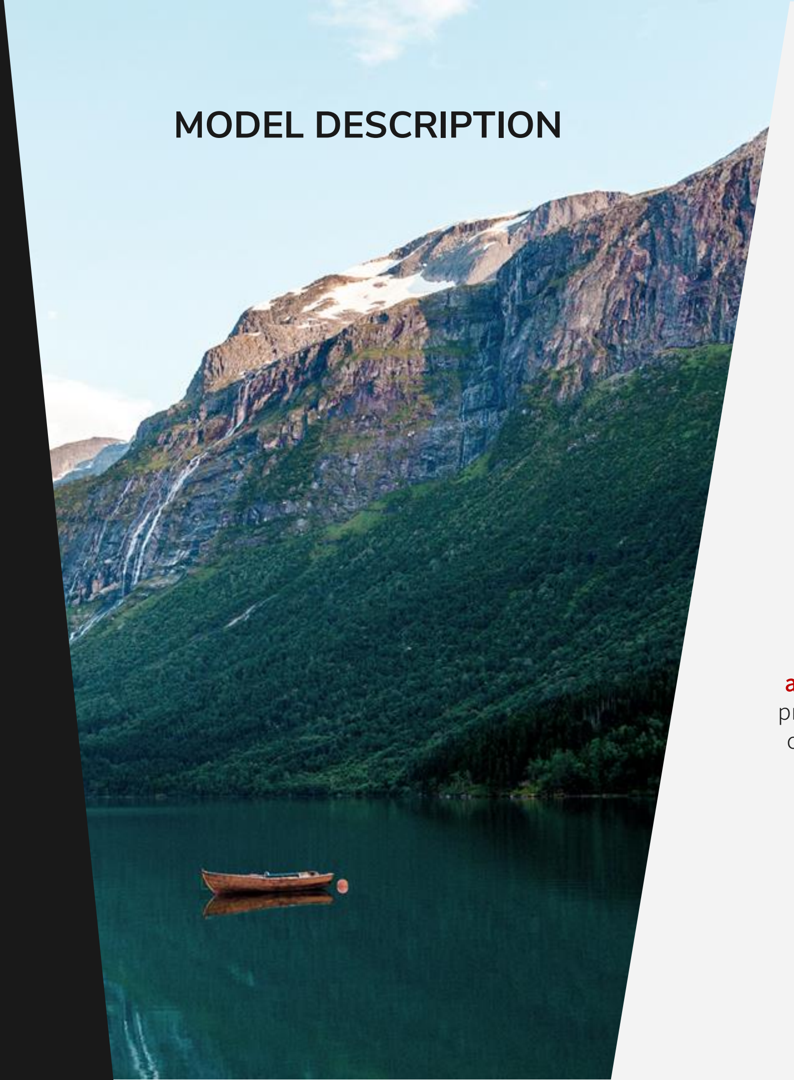
To note, the verifying performance on the test set produced performance consistent with the cross-validation results.



04

MODEL DESCRIPTION

MODEL DESCRIPTION



RANDOM FORESTS

As mentioned, **the model I used was the Random Forest Model with the goal of determining what price BMR's facilities support** as well as sensitivity of changes to various resort parameters.

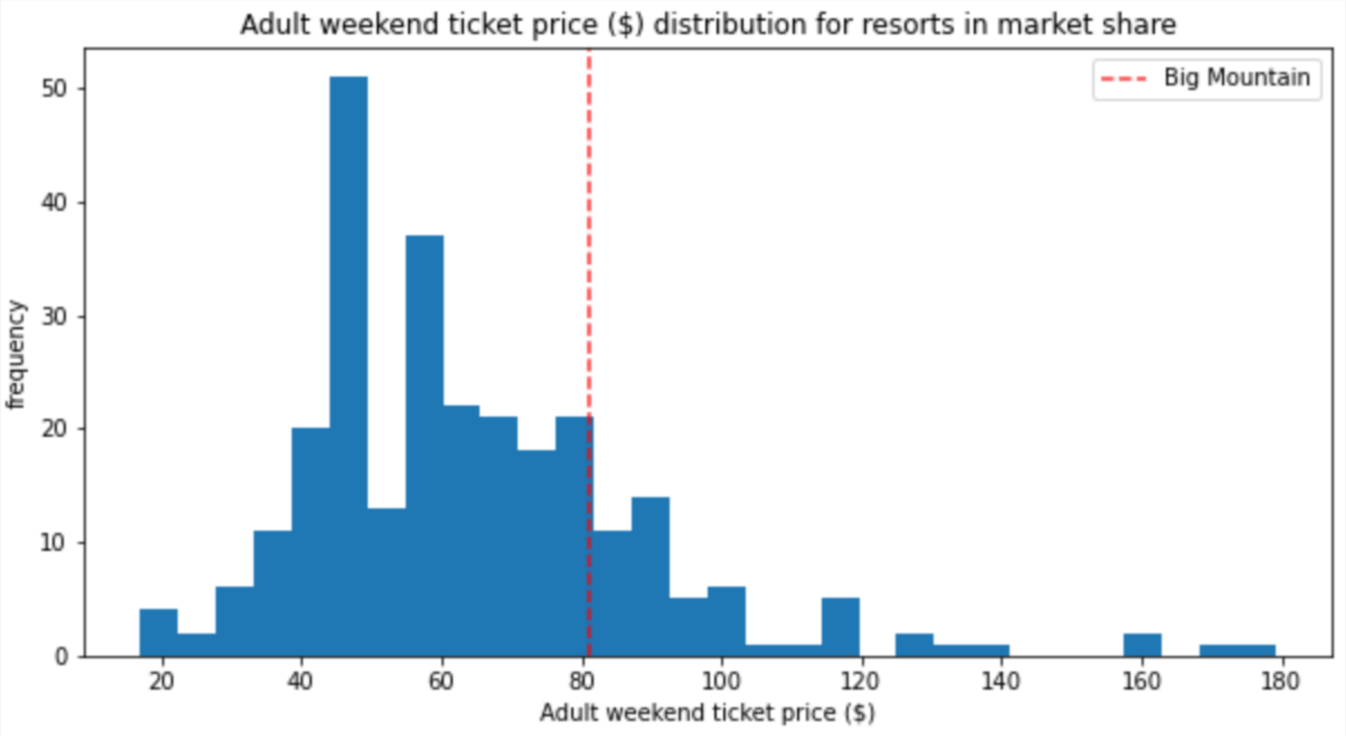
ASSUMPTION

This model relies on the implicit assumption that all resorts are setting prices based on how the market values certain facilities; essentially **prices are set by the free market.**

05

MODEL FINDINGS

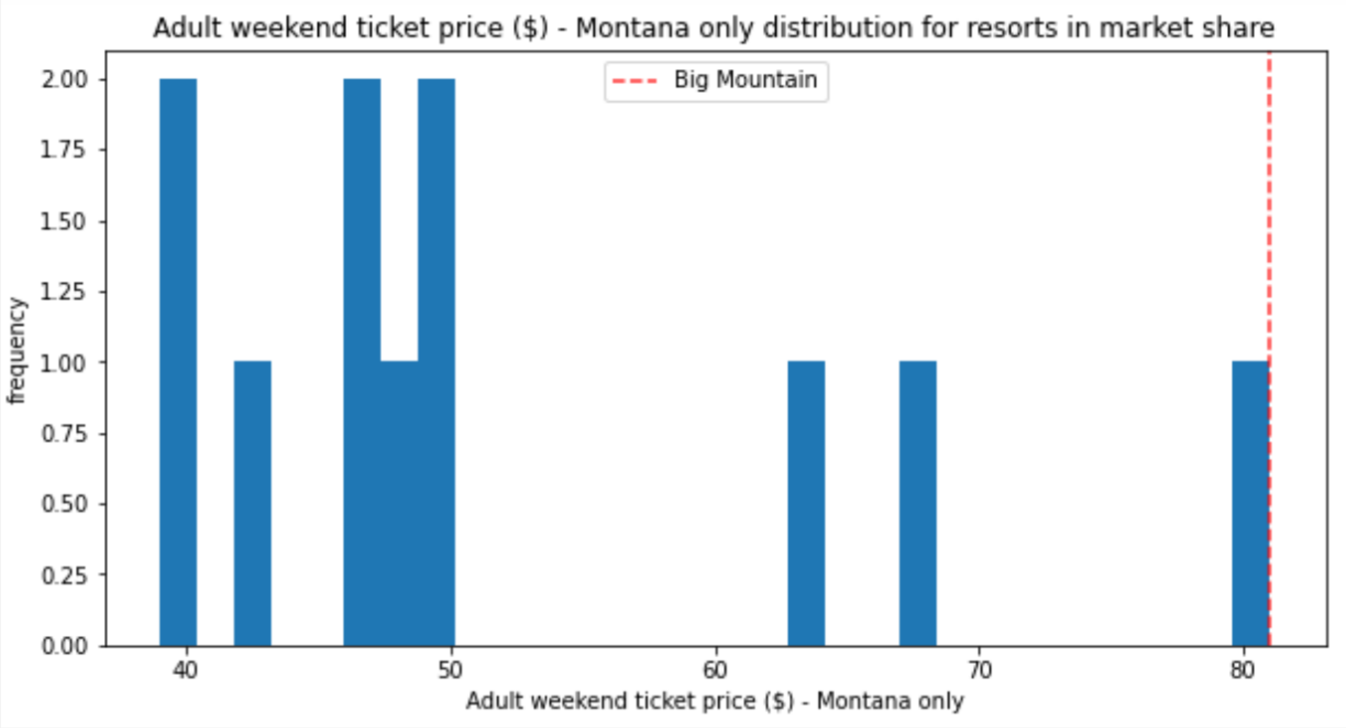
MODEL FINDINGS



As the business sought to review potential scenarios for either cutting costs or increasing revenue from ticket prices, **I explored where BMR stands against BMR's competitors nationally.**

To begin, **I find that BMR's price is slightly above the mean of the country**

MODEL FINDINGS (cont.)

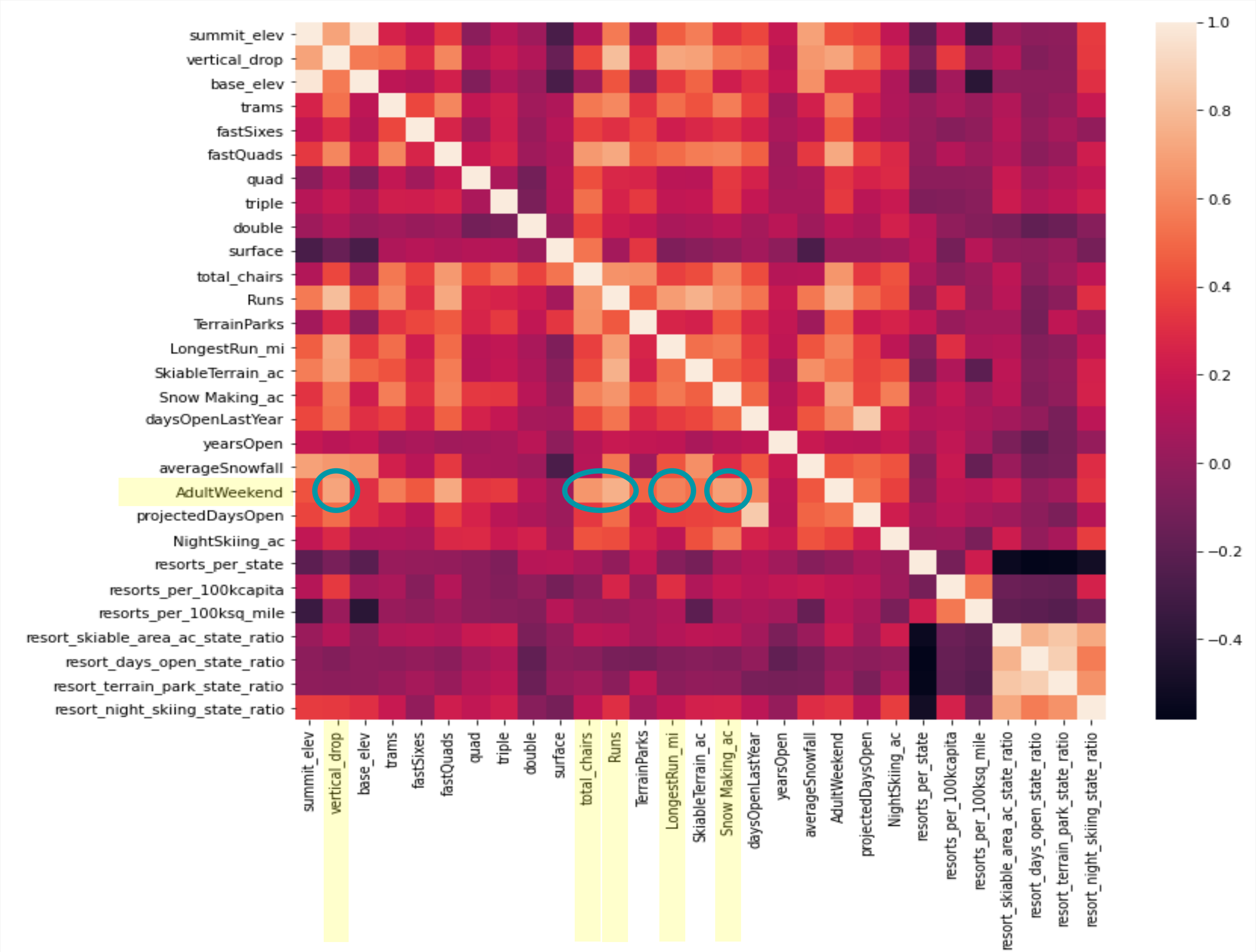


As the business sought to review potential scenarios for either cutting costs or increasing revenue from ticket prices, I explored where BMR stands against BMR's competitors nationally

To begin, I find that BMR's price is only slightly above the mean for the country **but the most expensive in Montana**. No actionable requirement, it's just something to note.

MODEL FINDINGS (cont.)

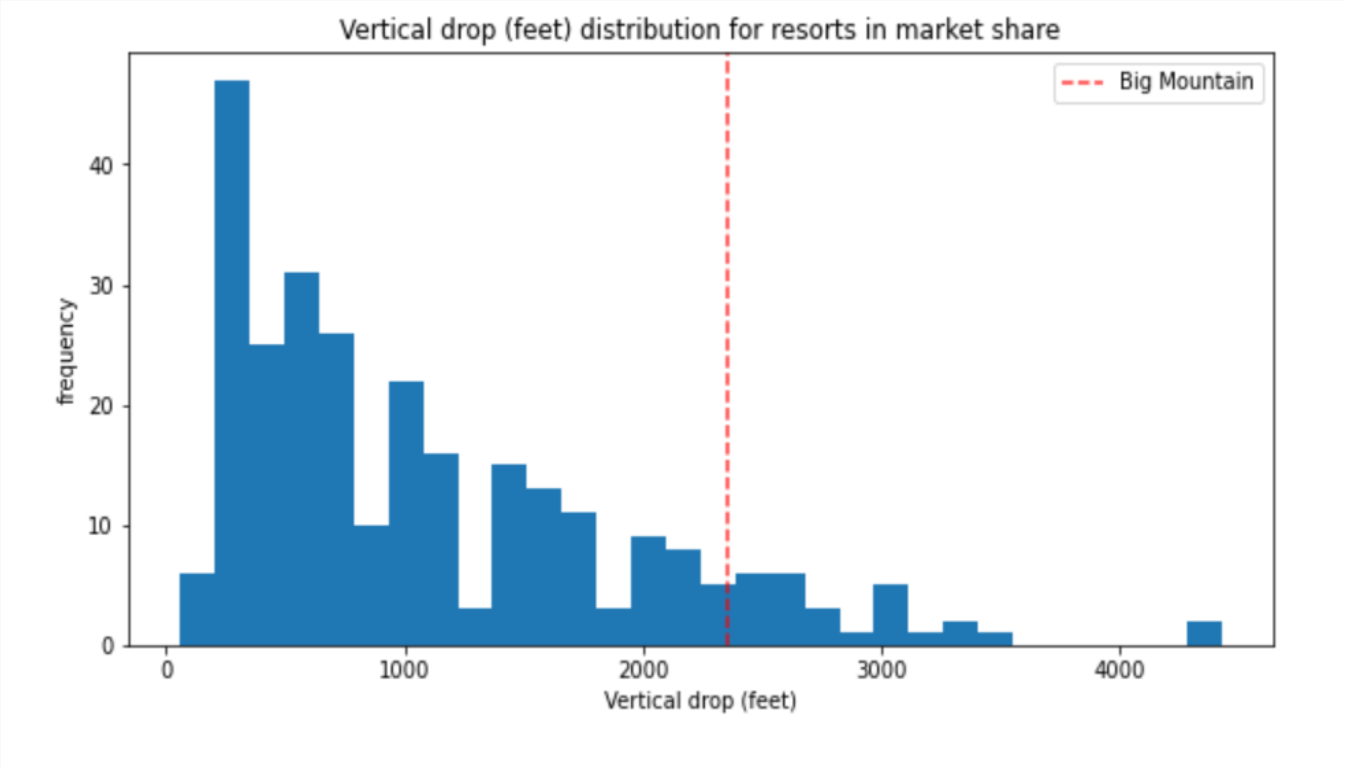
Heading back to the **Feature Correlation Heatmap**, I start by **investigating where BMR stands against competitors with the most positively correlated features for ticket prices.**



MODEL FINDINGS (cont.)

Heading back to the Feature Correlation Heatmap, I start by investigating where BMR stands against competitors with the most positively correlated features for ticket prices:

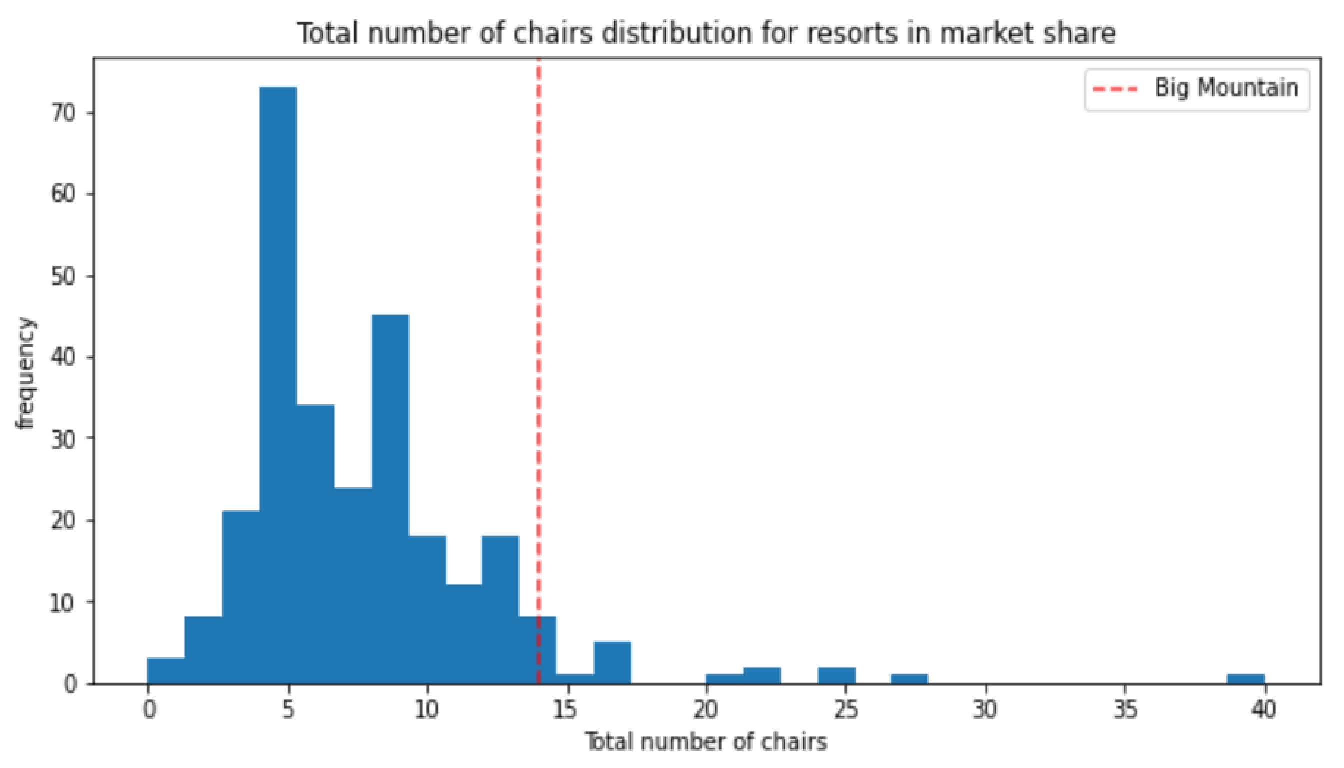
- **Vertical Drop**
  - *Doing well albeit a notable amount have steeper drops*



MODEL FINDINGS (cont.)

Heading back to the Feature Correlation Heatmap, I start by investigating where BMR stands against competitors with the most positively correlated features for ticket prices:

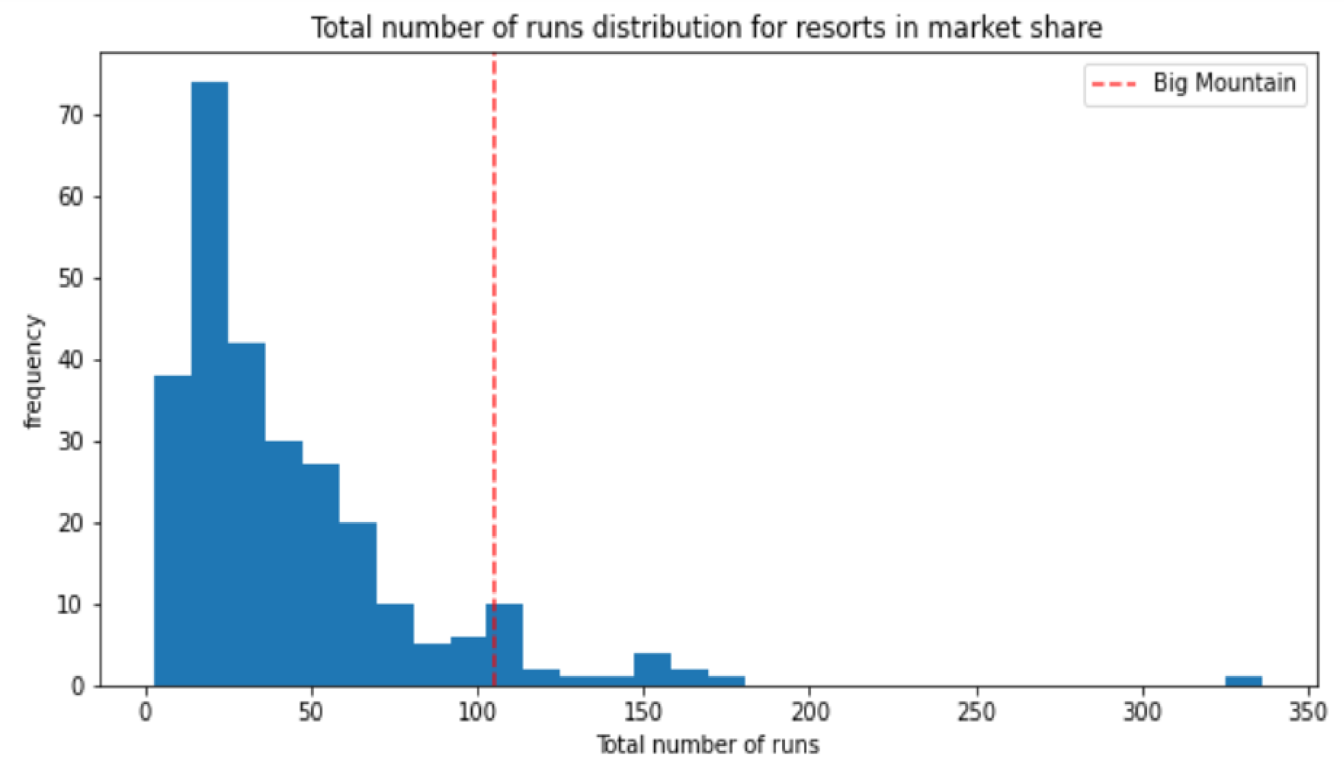
- Vertical Drop
- **Total # of Chairs**
  - *Currently amongst the highest; others may be outliers*



MODEL FINDINGS (cont.)

Heading back to the Feature Correlation Heatmap, I start by investigating where BMR stands against competitors with the most positively correlated features for ticket prices:

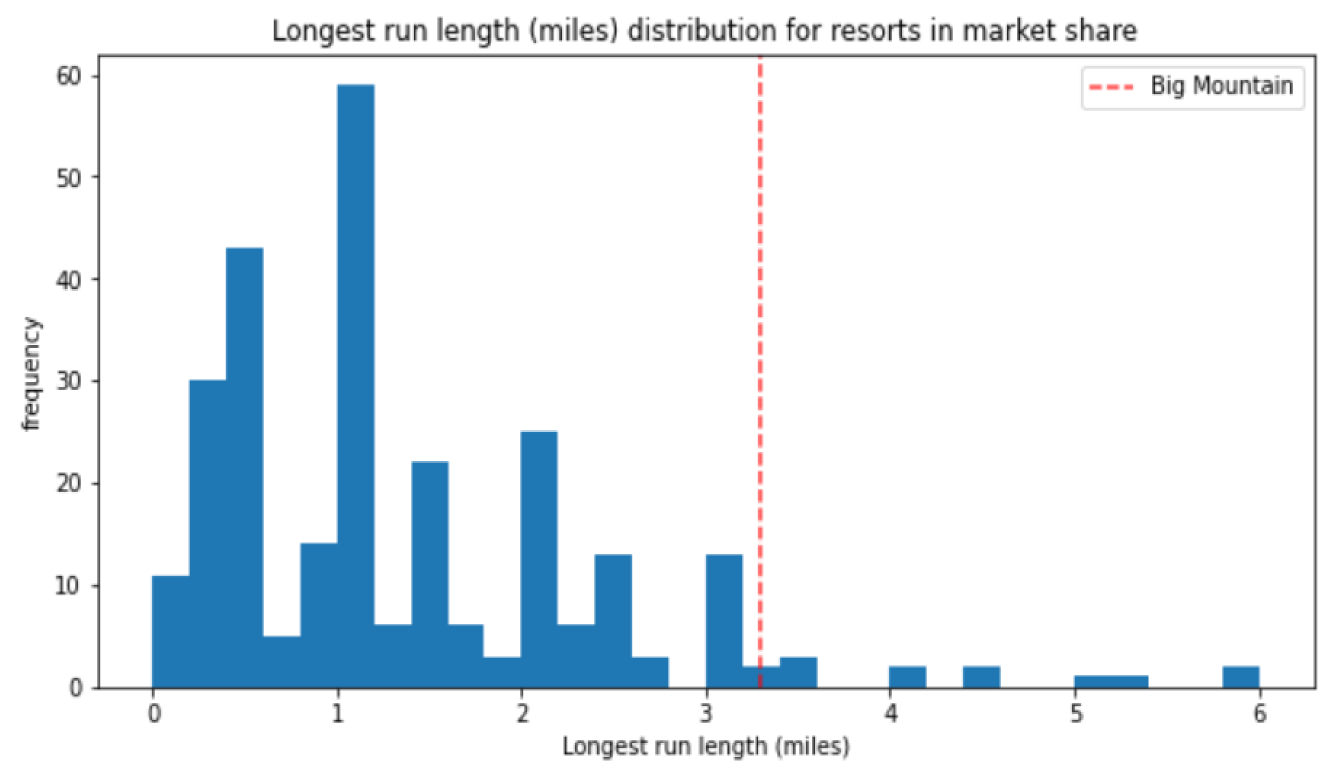
- Vertical Drop
- Total # of Chairs
- **Total # of Runs**
    - *Currently quite competitive albeit some offer more*



MODEL FINDINGS (cont.)

Heading back to the Feature Correlation Heatmap, I start by investigating where BMR stands against competitors with the most positively correlated features for ticket prices:

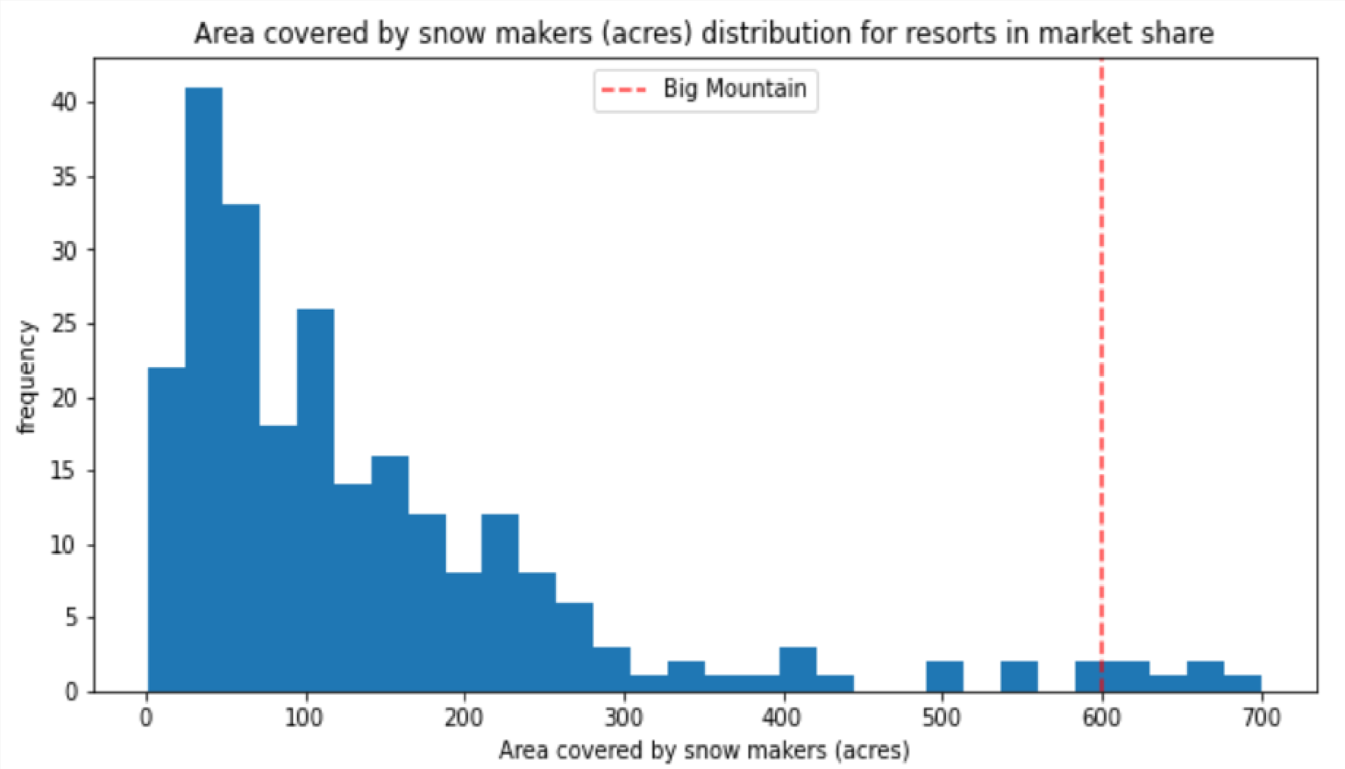
- Vertical Drop
- Total # of Chairs
- Total # of Runs
- **Longest Run**
  - *Again, quite competitive but BMR's longest run is ~approx. half that of the longest in the country*



MODEL FINDINGS (cont.)

Heading back to the Feature Correlation Heatmap, I start by investigating where BMR stands against competitors with the most positively correlated features for ticket prices:

- Vertical Drop
- Total # of Chairs
- Total # of Runs
- Longest Run
- Acres covered by Snow Makers
    - *Currently, very competitive*



COST CUTTING OR REVENUE INCREASING SOLUTIONS MODELLED IN 4 SCENARIOS

SCENARIO 1

Permanently **closing down up to 10 of the least used runs**

SCENARIO 2

Increasing the vertical drop to a point 150 feet lower **but requiring the installation of an additional chair lift** without additional snow making coverage

SCENARIO 3

Same as Scenario 2 albeit **adding 2 acres of snow making cover**

SCENARIO 4

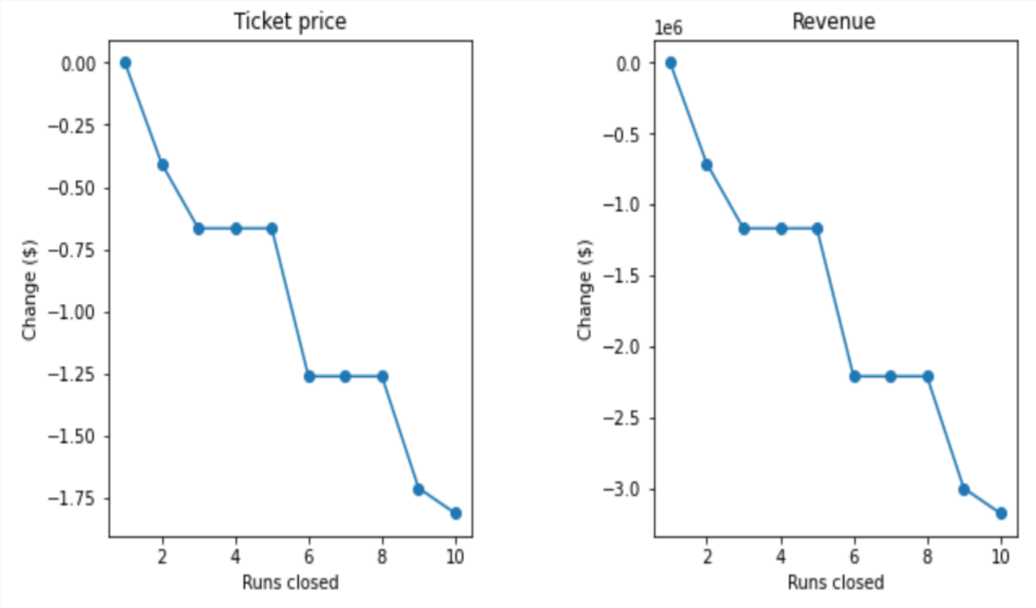
Increasing the longest run by 0.2 miles to 3.5 miles in length **& adding 4 additional acres of snow making cover**

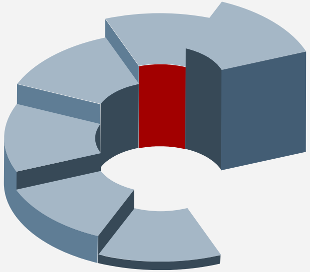
SCENARIO #1

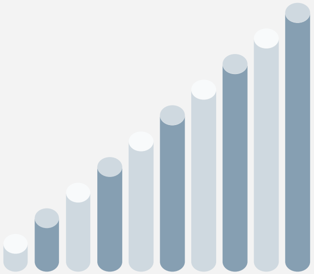
The model shows that **closing 1 run makes no difference to either Ticket Prices or Revenue**.

The **impact begins when 2 runs are closed. If 3 runs are closed, BMR could also close 4 or 5** as they have the same impact on Ticket prices & Revenue.

Closing 6 or more is when the model **indicates a substantially negative impact would occur.**







```
ticket2_increase = predict_increase(['Runs', 'vertical_drop',
                                     'total_chairs'],
                                    [1, 150, 1])
revenue2_increase = 5 * expected_visitors * ticket2_increase
```

```
print(f'This scenario increases support for ticket price by ${ticket2_increase:.2f}')
print(f'Over the season, this could be expected to amount to ${revenue2_increase:.0f}')
```

```
This scenario increases support for ticket price by $1.99
Over the season, this could be expected to amount to $3474638
```

SCENARIO #2

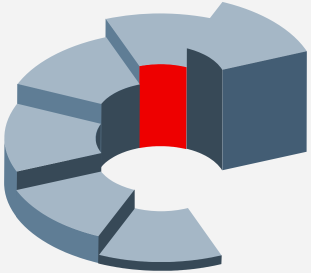
I modeled increasing the vertical drop to a point 150 feet lower but requiring the installation of an additional chair lift without additional snow making coverage.

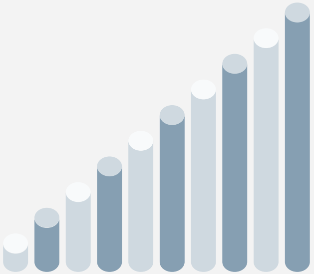
In this scenario, **the model increases support for ticket prices by $1.99** to $82.99 **which could increase revenue by $3,474,638 over the season**.

SCENARIO #3

I modeled **the same as Scenario 2 but added 2 additional acres of snow making**.

The model's **output was identical as Scenario 2** as the model **increases support for ticket prices by $1.99** to $82.99 & **may increase revenue by $3,474,638 over the season**.

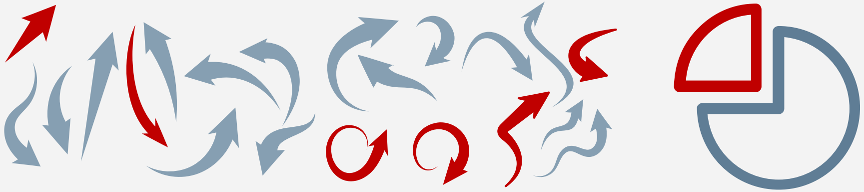




```
ticket3_increase = predict_increase(['Runs', 'vertical_drop',
                                      'total_chairs', 'Snow Making_ac'],
                                     [1, 150, 1, 2])
revenue3_increase = 5 * expected_visitors * ticket3_increase
```

```
print(f'This scenario increases support for ticket price by ${ticket3_increase:.2f}')
print(f'Over the season, this could be expected to amount to ${revenue3_increase:.0f}')
```

```
This scenario increases support for ticket price by $1.99
Over the season, this could be expected to amount to $3474638
```



```
predict_increase(['LongestRun_mi', 'Snow Making_ac'], [0.2, 4])
0.0
```

0.0

SCENARIO #4

I increased the longest run by 0.2 miles & added 4 acres of additional snow making cover.

The output showed no difference from the current. I believe this output is a result of the Random Forest Model placing the longest run low on the importance list.

06

NEXT STEPS

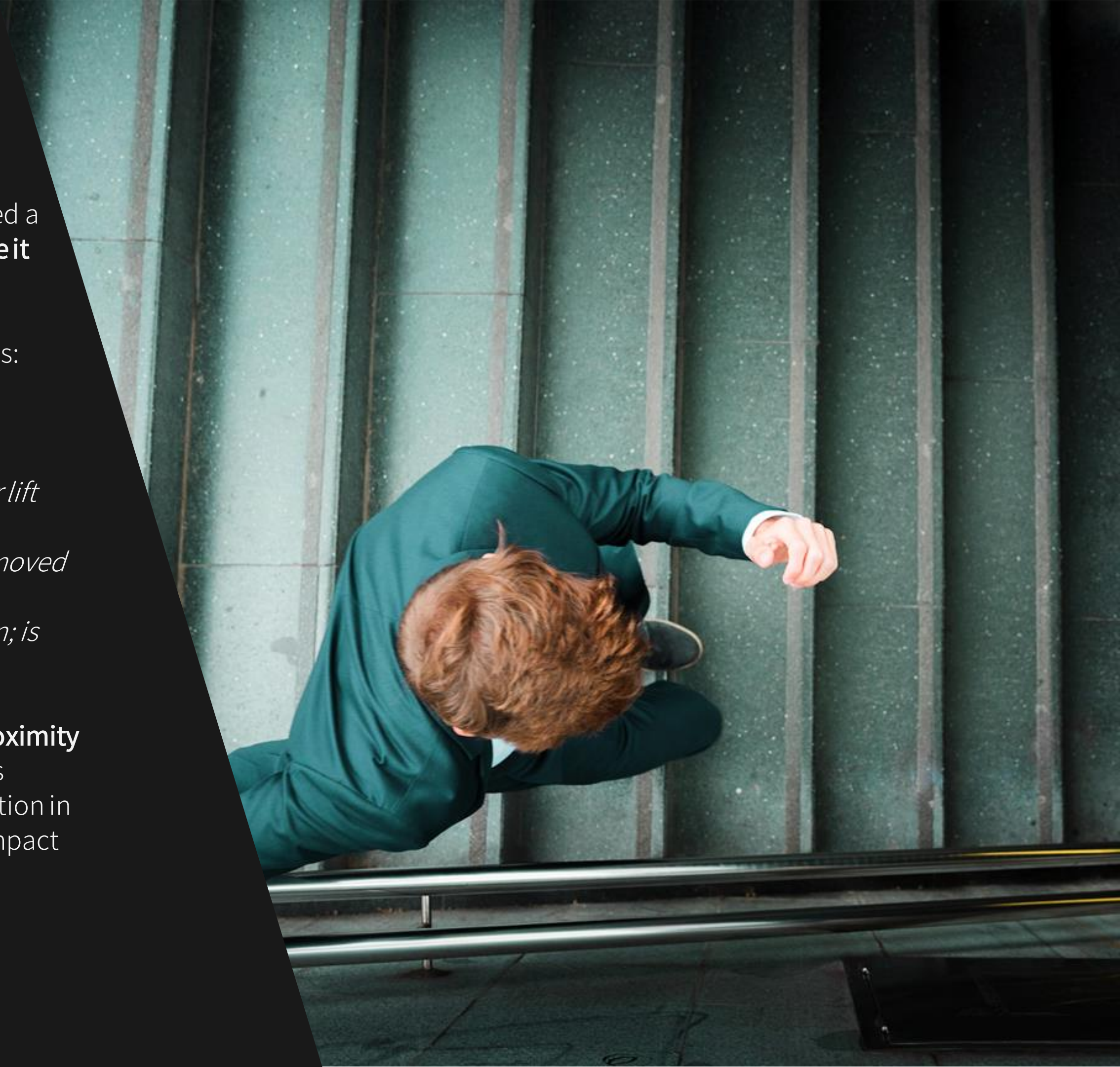
NEXT STEPS | GETTING TO THE ANSWER

<u>I advise Scenario 2 be taken under consideration</u> as it presented a possible solution. **While Scenario 3 presents the same, I believe it may incorporate additional cost(s) above those in Scenario 2.**

Further analysis could be undertaken with various departments:

- **Accounting & Finance**
  - *Further details and or structure of operational expenses*
- **Operations**
  - *The perceived benefit to operations by an additional chair lift*
- **Database**
  - *Review data sourcing methodology; particularly those removed*
- **Marketing**
  - *The most negative correlation to prices was skiable terrain; is exclusivity a possible marketing consideration*

Other data that may paint an even clearer picture could be proximity to the respective airports. This may not be relevant but if BMR's distance to an airport is comparatively higher than the competition in the dataset, this may make it less competitive and negatively impact our models.



THANKS

rand.sobczak@gmail.com
+1 313 447 8634

This is where I give credit to the ones who are part of this project.

- Presentation template by Slidesgo
- Icons by Flaticon
- Infographics by Freepik
- Images created by Freepik - Freepik
- Author introduction slide photo created by Freepik
- Font & Imagescreated by Freepik.com