Optimizing Pricing Strategy at Big Mountain Resort

1.   Introduction

As a leader in the ski industry, it is important for Big Mountain Resort to continuously assess and optimize its pricing strategies to maximize revenue and remain competitive. Our team has conducted an analysis of the resort's current pricing approach and identified opportunities for improvement. Our analysis has revealed that the resort's current pricing strategy, which is based on the market average, may not fully capitalize on the value of its facilities. Additionally, this approach does not provide a clear understanding of the relative importance of different facilities to visitors. This hinders the resort's investment strategy and may not be the most effective way to increase revenue.

To address these issues, our team developed a data-driven pricing model that considers the various facilities and properties offered by the resort. We compared the performance of two different models, a linear model and a random forest regressor, and selected the random forest regressor because it had a slightly better performance and was easier to interpret. Our goal is to increase profitability through the implementation of a pricing model that optimizes the value of the resort's facilities. We plan to achieve this by increasing ticket prices by 5%, while also considering the potential impact on demand and ensuring that the price increase is feasible in the current market. The success of these pricing strategies will be measured based on their ability to optimize the value of the resort's facilities, maximize revenue, and achieve cost savings without undermining the current ticket price.

2.   Data

Big Mountain Resort is concerned that it may not be fully optimizing its financial performance relative to its market position and lacks a clear understanding of which facilities drive the most value for visitors. To address these concerns, we are implementing a project to build a predictive model for adult weekend ticket prices based on the various facilities and properties offered by resorts. This model will be used to inform Big Mountain's pricing and investment decisions moving forward. Our objective during the data preparation phase was to efficiently collect, organize, and define the necessary data.

```
ski_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 330 entries, 0 to 329
Data columns (total 27 columns):
 #   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)
memory usage: 69.7+ KB
```

Ski Data

Note: 'AdultWeekday' is the price of an adult weekday ticket. 'AdultWeekend' is the price of an adult weekend ticket. The other columns are potential features.

```
ski_data[ski_data.Name == 'Big Mountain Resort'].T
```

|	151
Name	Big Mountain Resort
Region	Montana
state	Montana
summit_elev	6817
vertical_drop	2353
base_elev	4464
trams	0
fastEight	0.0
fastSixes	0
fastQuads	3
quad	2
triple	6
double	0
surface	3
total_chairs	14
Runs	105.0
TerrainParks	4.0
LongestRun_mi	3.3
SkiableTerrain_ac	3000.0
Snow Making_ac	600.0
daysOpenLastYear	123.0
yearsOpen	72.0
averageSnowfall	333.0
AdultWeekday	81.0
AdultWeekend	81.0
projectedDaysOpen	123.0
NightSkiing_ac	600.0

Big Mountain Resort Data

Next, we looked into possible relationships between region, state, and ticket prices.
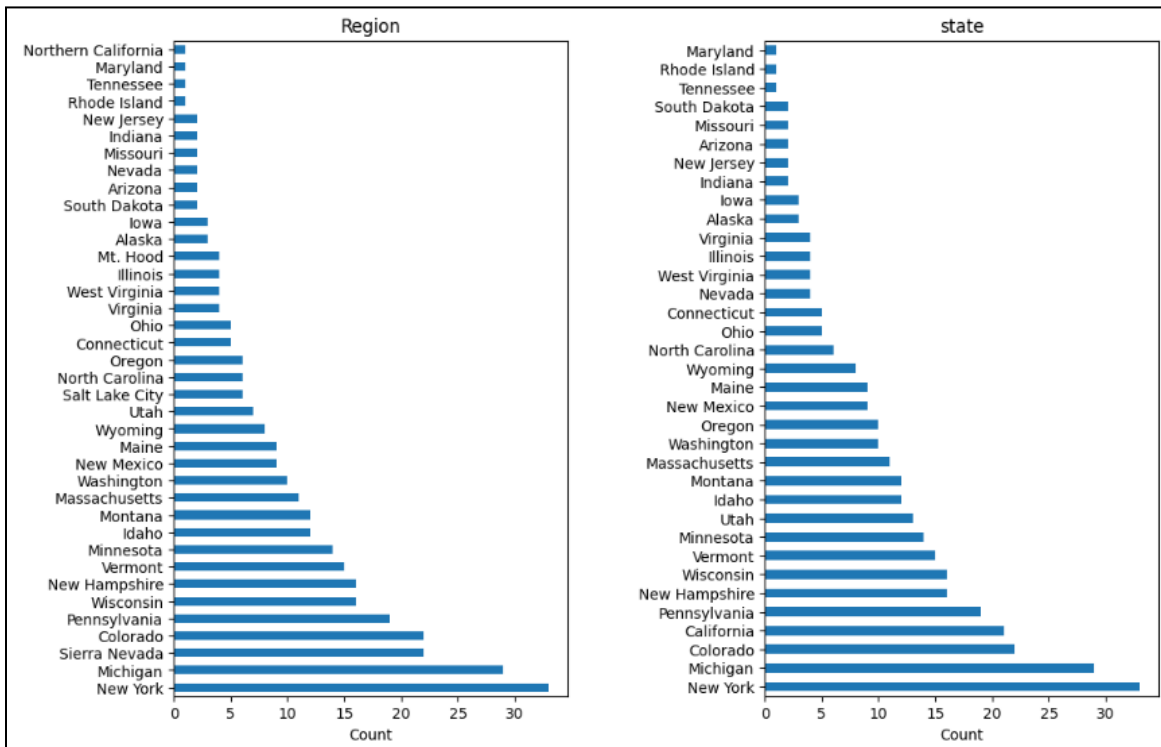
Distribution Of Resorts By Region And State

We analyzed the distribution of resorts across different states and found that New York has the most resorts, while Montana ranks 13th. Before examining ticket prices, we had questions about whether the popularity of resorts in New York is due to its proximity to a larger population. Additionally, we had questions about whether the state in which a resort is located can be a useful predictor of ticket prices. We also needed to determine whether weekend or weekday ticket prices should be the primary focus of our analysis. To compare prices, we created a box plot and looked for relationships between states and ticket prices.

Distribution Of Resorts By Region And State
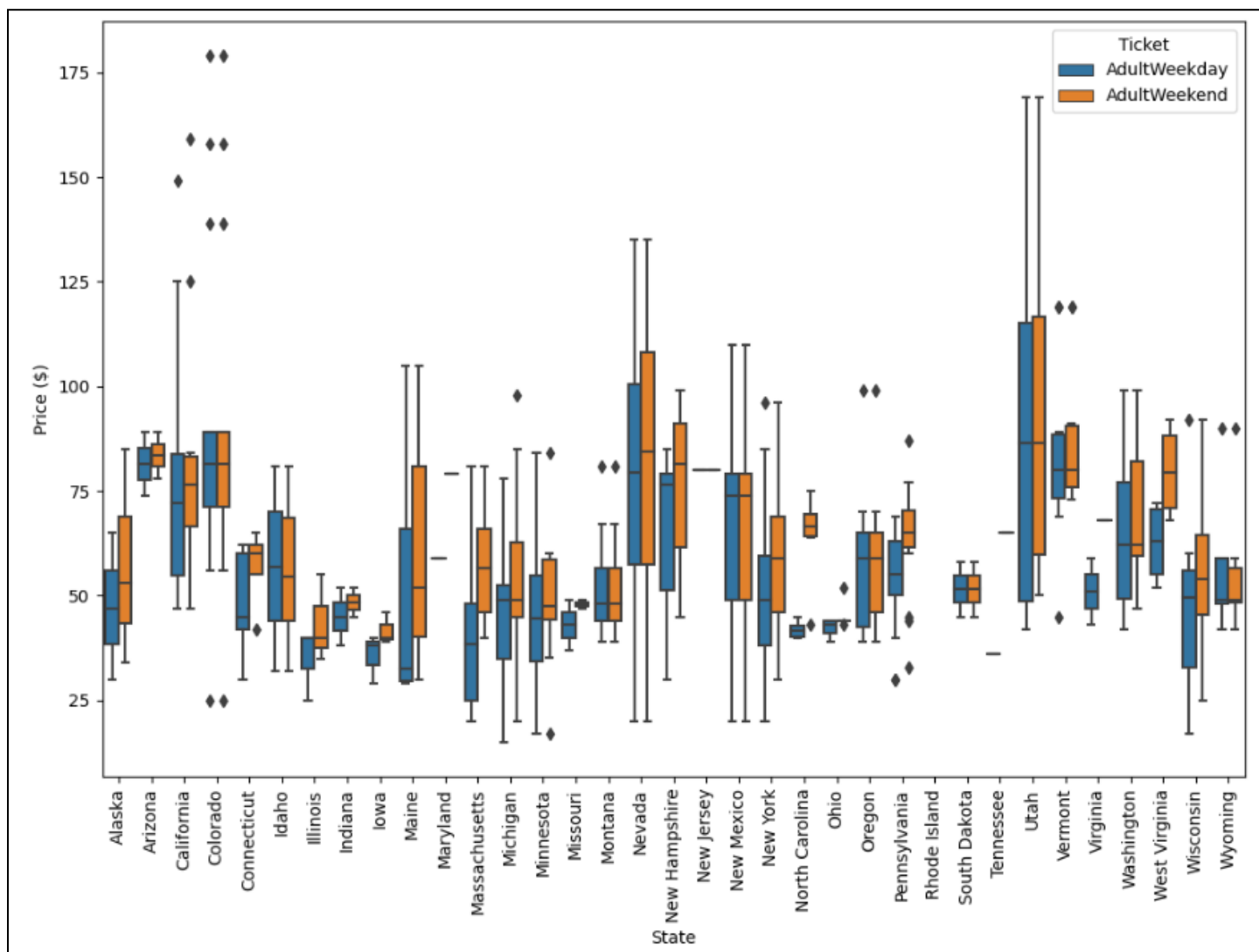
Our analysis showed that, aside from relatively high ticket prices in California, Colorado, and Utah, most prices fall within a broad range from around $25 to over $100. Some states showed more variability in prices than others, with Montana and South Dakota displaying relatively small ranges and similar weekend and weekday ticket prices, while Nevada and Utah had the largest price ranges. Given the lack of a strong relationship between state and ticket prices, we were unsure whether to retain or disregard state information in our analysis.

The rest of the wrangling process included cleaning up the numeric features, such as 'vertical_drop', 'fastEight', and 'Runs'.
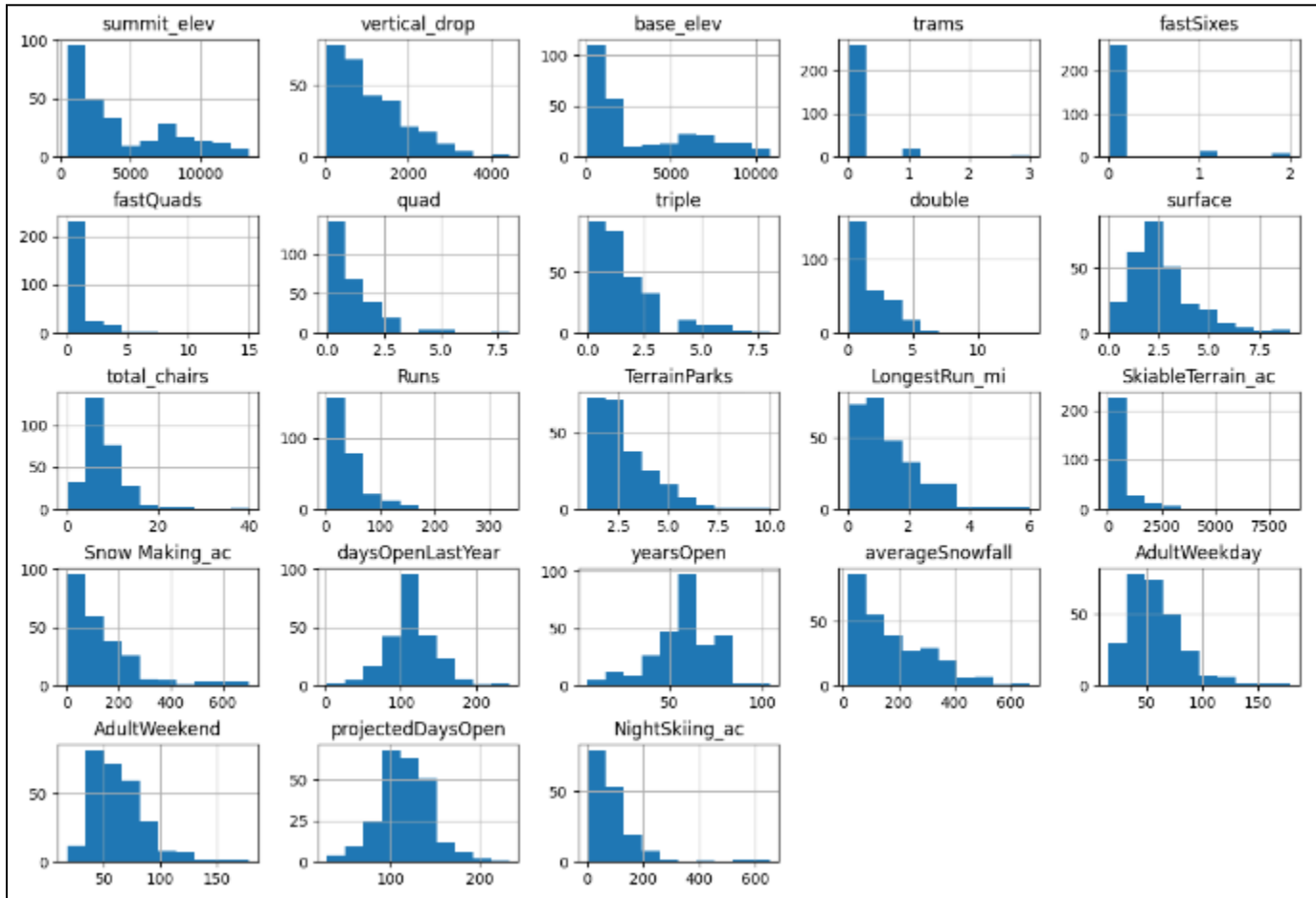
Distributions of features
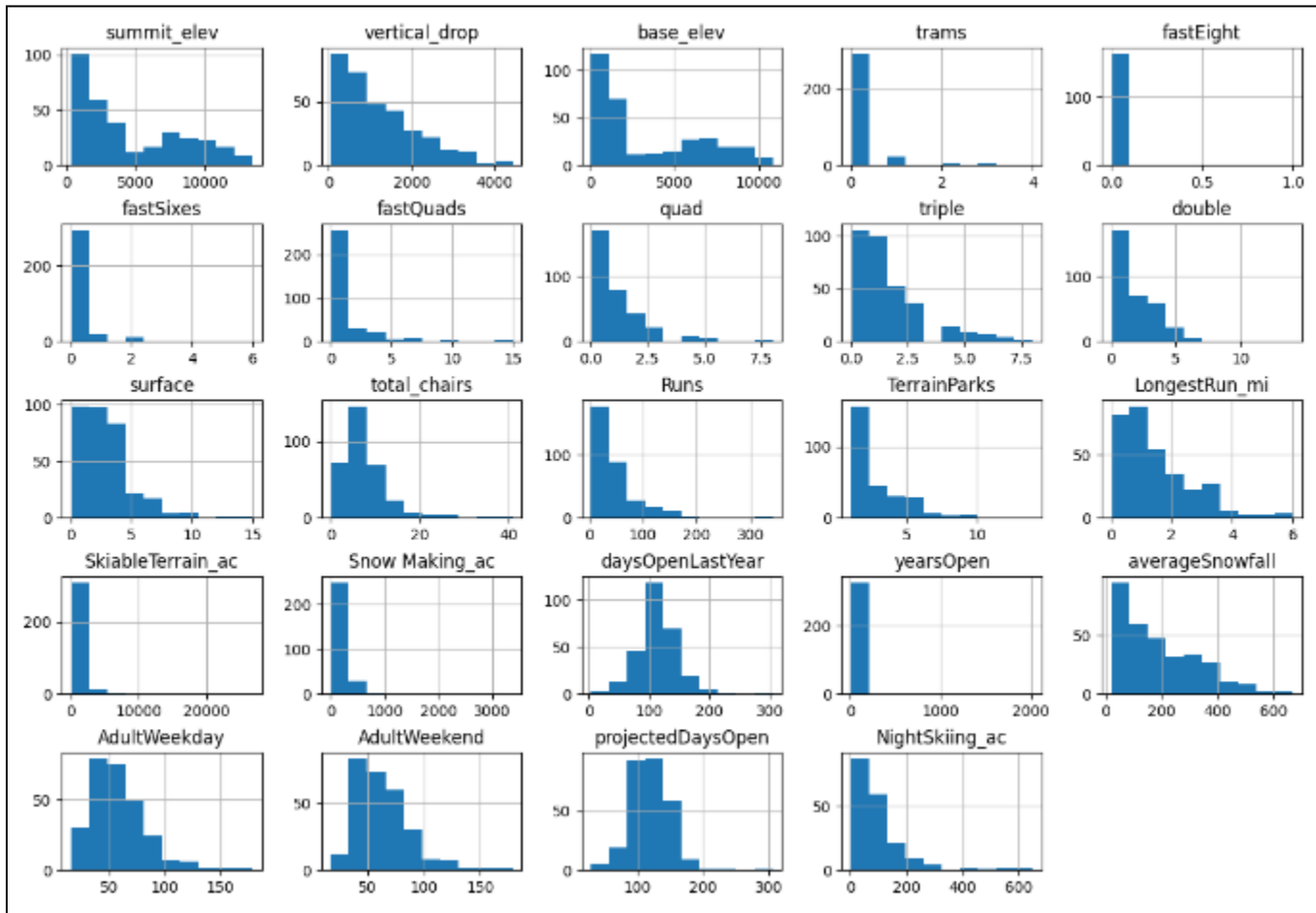
We removed the 'fastEight' column in its entirety because half of the values are missing and all the remaining values are zero. We also removed features related to chairlifts and altitudes, as they are not relevant to our analysis. Likewise, we then removed rows with no price data, which is our target feature.
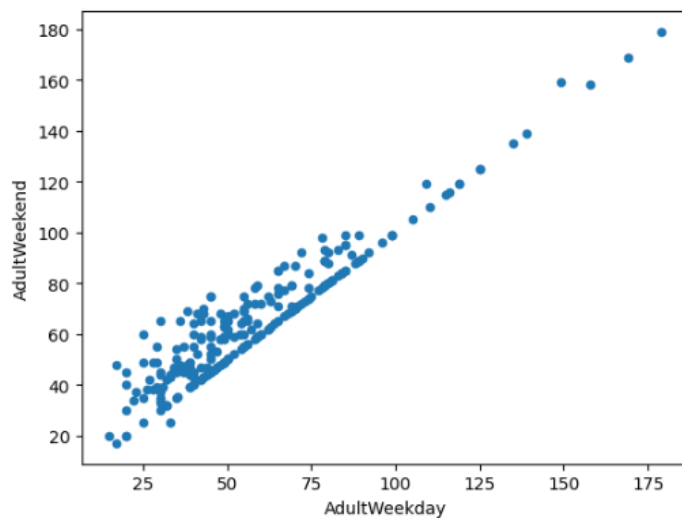
Missing Price Data

```
missing_price = ski_data[['AdultWeekend', 'AdultWeekday']].isnull().sum(axis=1)
missing_price.value_counts()/len(missing_price) * 100

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

In addition, about 14% of the rows had no price data, so those rows were dropped.

Updated distribution of features

Finally, we examined options for the target feature when modeling ticket prices. We considered using adult weekend and weekday prices as the target and created a scatter plot to compare them.

Adult Ticket Prices Comparison

There is a clear line where weekend and weekday prices are equal. Weekend prices being higher than weekday prices seem restricted to sub $100 resorts. The box plot from earlier showed that the distribution for weekday and weekend prices in Montana seemed equal.

Adult Ticket Prices Comparison

	AdultWeekend	AdultWeekday
141	42.0	42.0
142	63.0	63.0
143	49.0	49.0
144	48.0	48.0
145	46.0	46.0
146	39.0	39.0
147	50.0	50.0
148	67.0	67.0
149	47.0	47.0
150	39.0	39.0
151	81.0	81.0

Weekend prices have the least missing values of the two, so drop the weekday prices and then keep just the rows that have weekend price.

Missing Ticket Prices

```
ski_data[['AdultWeekend', 'AdultWeekday']].isnull().sum()

AdultWeekend    4
AdultWeekday    7
dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 277 entries, 0 to 329
Data columns (total 25 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Name               277 non-null    object
 1   Region             277 non-null    object
 2   state              277 non-null    object
 3   summit_elev        277 non-null    int64
 4   vertical_drop      277 non-null    int64
 5   base_elev          277 non-null    int64
 6   trams              277 non-null    int64
 7   fastSixes          277 non-null    int64
 8   fastQuads          277 non-null    int64
 9   quad               277 non-null    int64
 10  triple             277 non-null    int64
 11  double             277 non-null    int64
 12  surface            277 non-null    int64
 13  total_chairs       277 non-null    int64
 14  Runs               274 non-null    float64
 15  TerrainParks       233 non-null    float64
 16  LongestRun_mi      272 non-null    float64
 17  SkiableTerrain_ac  275 non-null    float64
 18  Snow Making_ac     240 non-null    float64
 19  daysOpenLastYear   233 non-null    float64
 20  yearsOpen          277 non-null    float64
 21  averageSnowfall    268 non-null    float64
 22  AdultWeekend       277 non-null    float64
 23  projectedDaysOpen  236 non-null    float64
 24  NightSkiing_ac     163 non-null    float64
dtypes: float64(11), int64(11), object(3)
memory usage: 56.3+ KB
```

After data file was cleaned

3. Analysis

Now that we knew that weekend prices were going to be our target feature for price modeling, it was time to revisit the State dataset.

Total state area	*Total state population*

```state```
```Alaska   665384```
```California   163695```
```Montana   147040```
```New Mexico   121590```
```Arizona   113990```
```Name: state_area_sq_miles, dtype: int64```
	```state```
	```California 39512223```
	```New York 19453561```
	```Pennsylvania 12801989```
	```Illinois 12671821```
	```Ohio 11689100```
	```Name: state_population, dtype: int64```

Resorts per state	*Total skiable area*

```state```
```New York   33```
```Michigan   28```
```Colorado   22```
```California   21```
```Pennsylvania   19```
```Name: resorts_per_state, dtype: int64```
	```state```
	```Colorado 43682.0```
	```Utah 30508.0```
	```California 25948.0```
	```Montana 21410.0```
	```Idaho 16396.0```
	```Name: state_total_skiable_area_ac, dtype: float64```

*Total night skiing area*

```
state
New York          2836.0
Washington        1997.0
Michigan          1946.0
Pennsylvania      1528.0
Oregon            1127.0
Name: state_total_nightskiing_ac, dtype: float64
```

Total days open

```
state
Colorado          3258.0
California         2738.0
Michigan          2389.0
New York          2384.0
New Hampshire     1847.0
Name: state_total_days_open, dtype: float64
```

Montana is the third-largest state in terms of total area, but it is not among the most populous states. However, despite having fewer resorts than states like New York, Montana ranks among the top five states for total skiable area. There are large states that are not necessarily the most populous and states that have many resorts, but others have a larger total skiing area. The states with the most total days of skiing per season are not necessarily those with the most resorts. Therefore, instead of solely considering the absolute size or population of a state, we were more interested in the ratio of resorts serving a particular population or area.

Resort Density
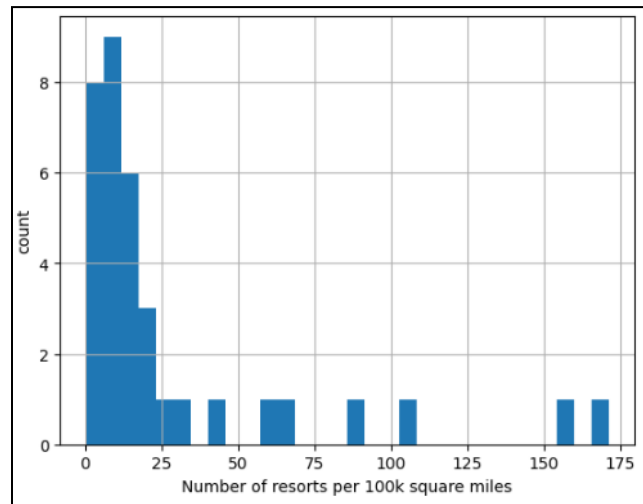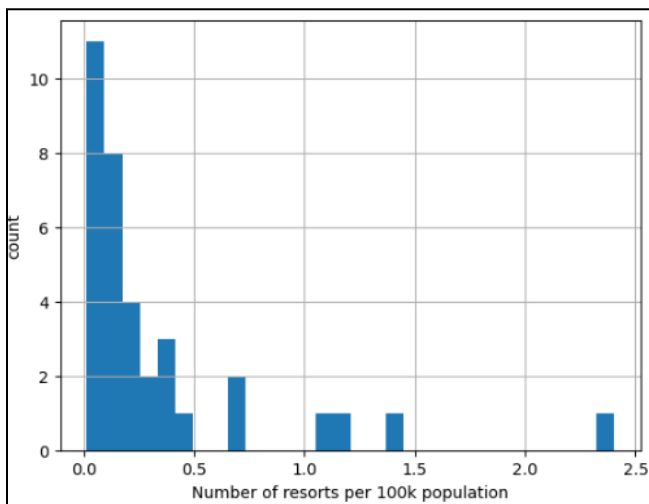
Top states by resort density

```
state
Vermont           2.403889
Wyoming           1.382268
New Hampshire     1.176721
Montana           1.122778
Idaho             0.671492
Name: resorts_per_100kcapita, dtype: float64
```

```
state
New Hampshire     171.141299
Vermont           155.990017
Massachusetts     104.225886
Connecticut        90.203861
Rhode Island       64.724919
Name: resorts_per_100ksq_mile, dtype: float64
```

At this point, we have constructed several potentially useful and business-relevant features based on summary statistics for each of the states that we are interested in. We also explored many of these features and found various trends. Some states are higher in some features but not in others, and some features are more correlated with each other than others.

To better understand these trends, we decided to use a technique that identifies linear combinations of the original features that are uncorrelated with each other and orders them by the amount of variance they explain. The chart below visualizes the distribution of states and their average ticket price. Despite accounting for 77% of the variance, we did not see a clear pattern or grouping between states and ticket price.

Ski States Summary – Principal Component Analysis

Ski states summary PCA, 77.2% variance explained
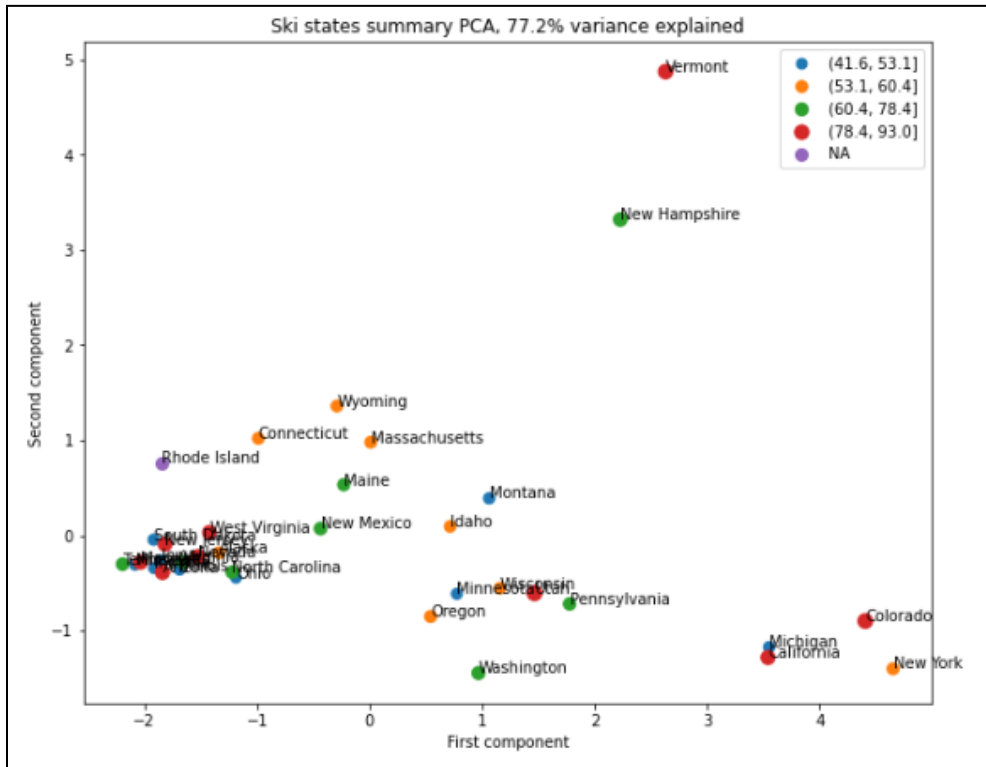
We then merged the two datasets (ski_data and state_summary) and engineered some intuitive features.

Ski Data

|	0	1	2	3	4
Name	Alyeska Resort	Eaglecrest Ski Area	Hilltop Ski Area	Arizona Snowbowl	Sunrise Park Resort
Region	Alaska	Alaska	Alaska	Arizona	Arizona
state	Alaska	Alaska	Alaska	Arizona	Arizona
summit_elev	3939	2600	2090	11500	11100
vertical_drop	2500	1540	294	2300	1800
base_elev	250	1200	1796	9200	9200
trams	1	0	0	0	0
fastSixes	0	0	0	1	0
fastQuads	2	0	0	0	1
quad	2	0	0	2	2
triple	0	0	1	2	3
double	0	4	0	1	1
surface	2	0	2	2	0
total_chairs	7	4	3	8	7
Runs	76.0	36.0	13.0	55.0	65.0
TerrainParks	2.0	1.0	1.0	4.0	2.0
LongestRun_mi	1.0	2.0	1.0	2.0	1.2
SkiableTerrain_ac	1610.0	640.0	30.0	777.0	800.0
Snow Making_ac	113.0	60.0	30.0	104.0	80.0
daysOpenLastYear	150.0	45.0	150.0	122.0	115.0
yearsOpen	60.0	44.0	36.0	81.0	49.0
averageSnowfall	669.0	350.0	69.0	260.0	250.0
AdultWeekend	85.0	53.0	34.0	89.0	78.0
projectedDaysOpen	150.0	90.0	152.0	122.0	104.0
NightSkiing_ac	550.0	NaN	30.0	NaN	80.0
resorts_per_state	3	3	3	2	2
state_total_skiable_area_ac	2280.0	2280.0	2280.0	1577.0	1577.0
state_total_days_open	345.0	345.0	345.0	237.0	237.0
state_total_terrain_parks	4.0	4.0	4.0	6.0	6.0
state_total_nightskiing_ac	580.0	580.0	580.0	80.0	80.0
resorts_per_100kcapita	0.410091	0.410091	0.410091	0.027477	0.027477
resorts_per_100ksq_mile	0.450867	0.450867	0.450867	1.75454	1.75454

After merging the state summary features into the ski resort data, we then added the following features and dropped the state column:
- ratio of resort skiable area to total state skiable area
- ratio of resort days open to total state days open

- ratio of resort terrain park count to total state terrain park count
- ratio of resort night skiing area to total state night skiing area.

We were then able to produce a feature collection heatmap to gain a high live view of relationships amongst the features.
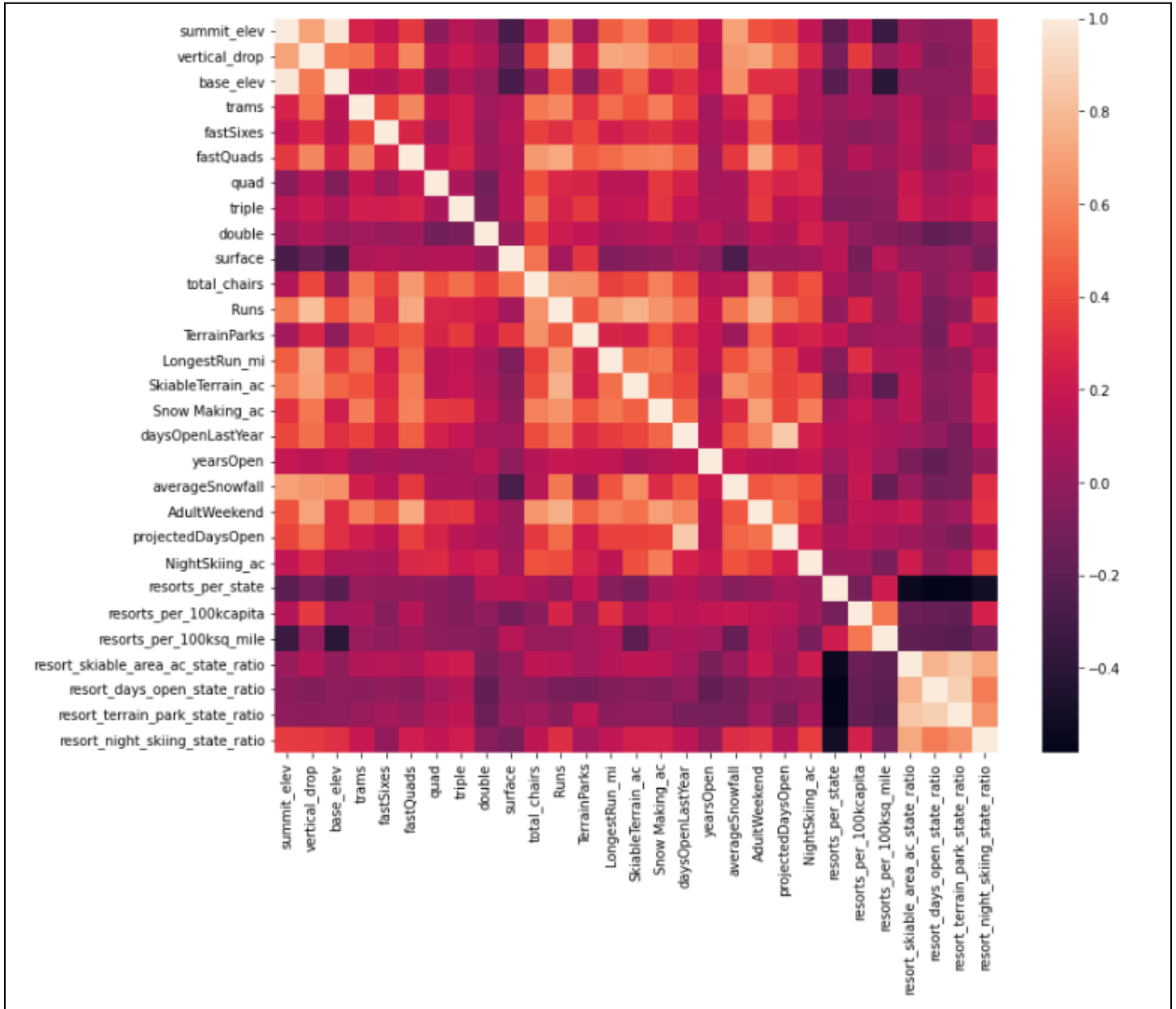
Our target feature, 'AdultWeekend' ticket price, has several reasonable correlations. 'fastQuads', 'Runs', and 'Snow Making_ac' stand out as correlated features. It appears that visitors value resorts with more guaranteed snow, which likely requires more snow making equipment and drives up prices and costs. Among the new features, 'resort_night_skiing_state_ratio' appears to be the most correlated with ticket price. It could be that having a larger share of night skiing capacity allows a resort to charge higher prices. 'Runs' and 'total_chairs' are also fairly well correlated with ticket price. It appears that the more runs a resort has, the more chairs it would need to transport people to them. Additionally, people appear to value guaranteed snow cover over a larger, potentially more variable terrain area. The 'vertical drop' feature also seems to be a selling point that increases ticket prices.

We then created a series of scatter plots to really dive into how ticket prices varies with other numeric features.
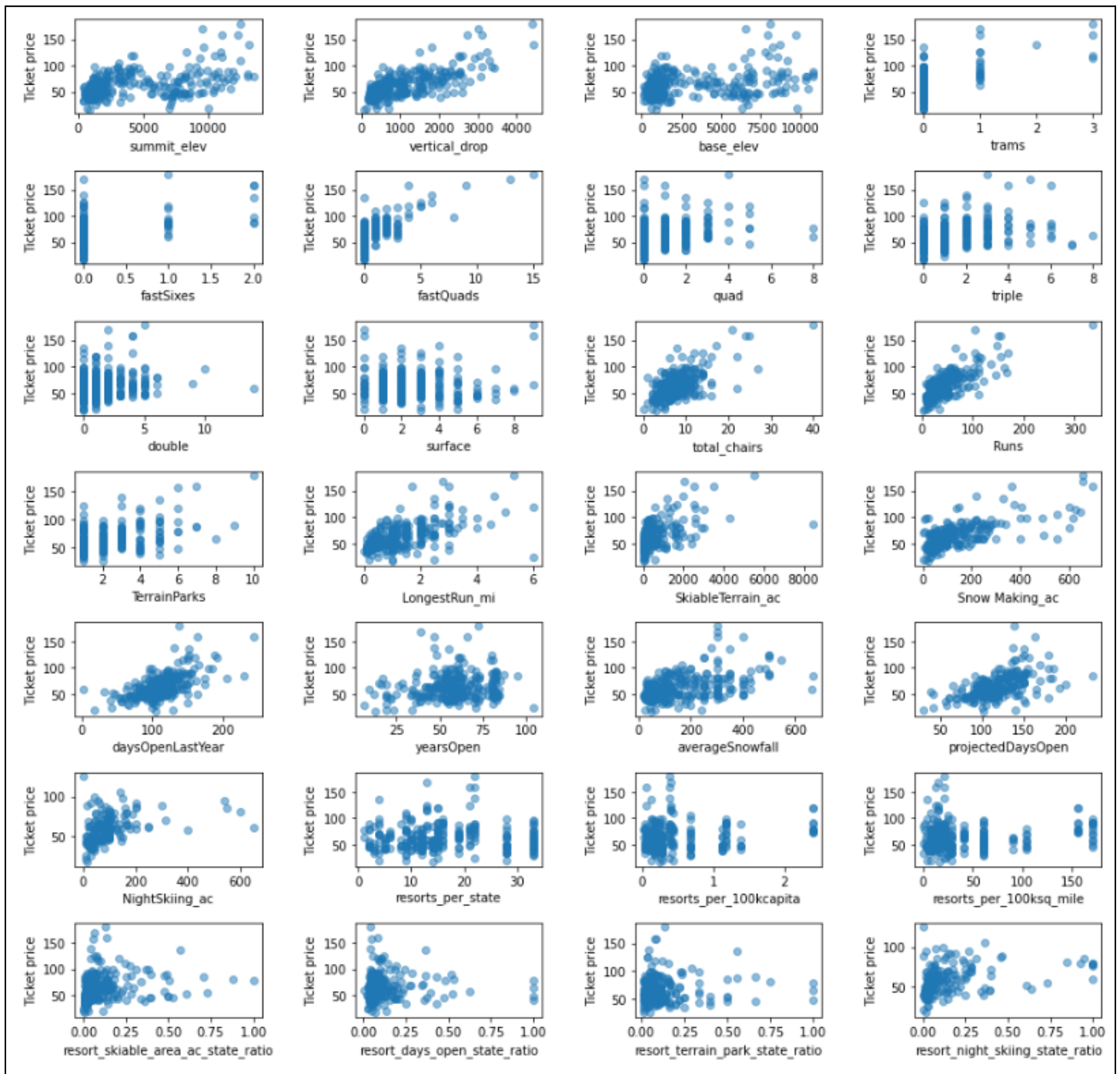
There is a strong positive correlation with 'vertical_drop'. 'fastQuads' appears to be a very useful feature. 'Runs' and 'total_chairs' seem similar and also useful. 'resorts_per_100kcapita' shows some interesting trends that are not apparent from just the headline correlation figure. When the value is low, there is significant variability in ticket price, although it is capable of reaching high values. Ticket price may initially drop before increasing as the number of resorts per capita increases. This could be because ticket price may rise with the number of resorts serving a population, indicating a popular area for skiing with high demand. On the other hand, the lower ticket price when fewer resorts serve a population may be because it is a less popular state for skiing. The high price for some resorts when resorts are scarce

(relative to the population size) may be due to a monopoly effect in those areas. While the picture is not clear, we have identified some interesting trends.

Finally, we can consider additional features that may be useful in terms of how easily a resort can transport people around. While we have the numbers of various chairs and runs, we do not have the ratio of chairs to runs. It seems logical that this ratio would indicate how easily and quickly people can reach their next ski slope. We can create these features now.
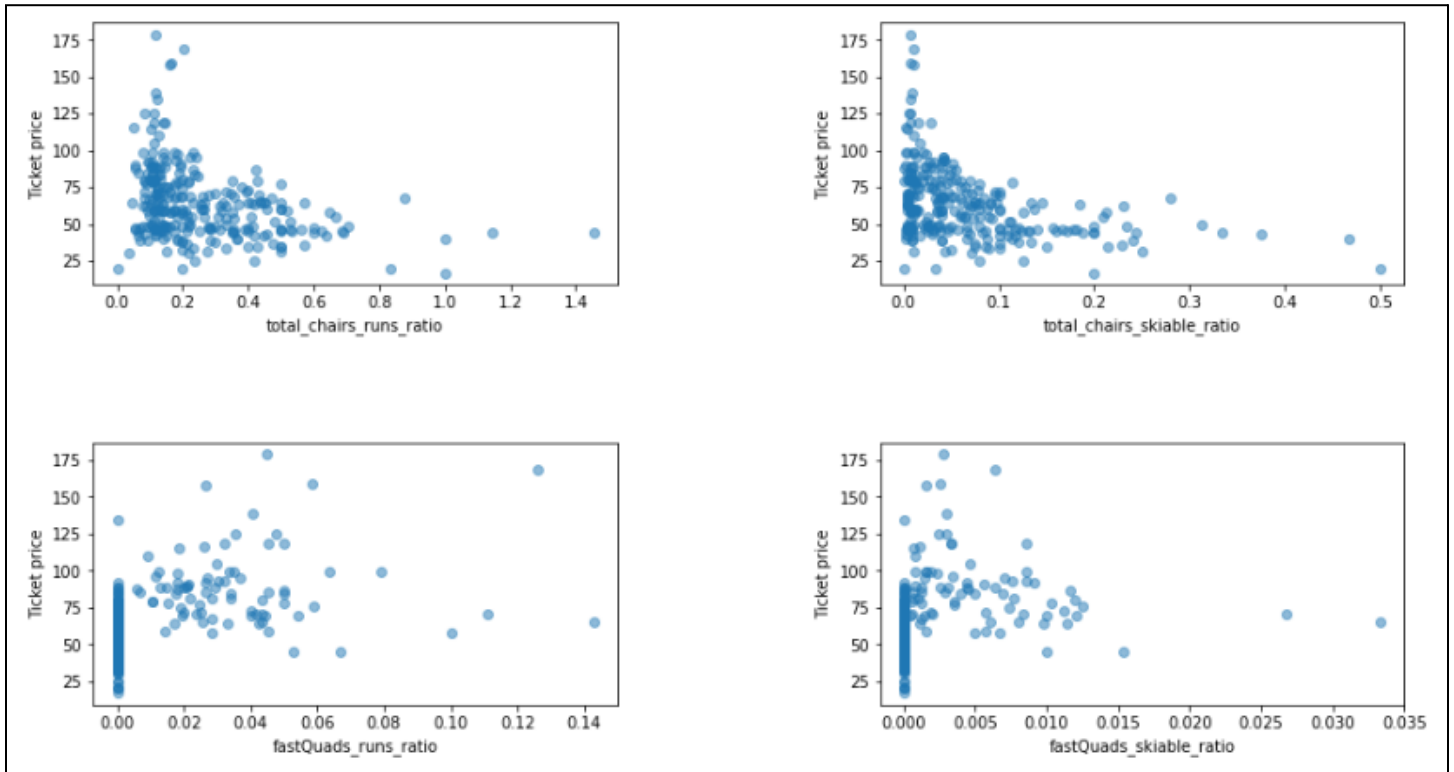
Initially, these relationships may seem counterintuitive. It appears that the more chairs a resort has to transport people, relative to the number of runs, ticket price decreases rapidly and remains low. This may be due to an exclusive versus mass market resort effect; if a resort has fewer chairs, it can charge more for tickets, but it will also be able to serve fewer visitors. The price per visitor is high, but the number of visitors may be low. One important piece of information that is missing from the data is the number of visitors per year.

Additionally, it appears that having no fast quads may limit ticket price, but if a resort has a wide coverage area, then a few fast quads may be beneficial for ticket price.

Updated ski data file

```
ski_data.head().T
```

	0	1	2	3	4
Name	Alyeska Resort	Eaglecrest Ski Area	Hilltop Ski Area	Arizona Snowbowl	Sunrise Park Resort
Region	Alaska	Alaska	Alaska	Arizona	Arizona
state	Alaska	Alaska	Alaska	Arizona	Arizona
summit_elev	3939	2600	2090	11500	11100
vertical_drop	2500	1540	294	2300	1800
base_elev	250	1200	1796	9200	9200
trams	1	0	0	0	0
fastSixes	0	0	0	1	0
fastQuads	2	0	0	0	1
quad	2	0	0	2	2
triple	0	0	1	2	3
double	0	4	0	1	1
surface	2	0	2	2	0
total_chairs	7	4	3	8	7
Runs	76.0	36.0	13.0	55.0	65.0
TerrainParks	2.0	1.0	1.0	4.0	2.0
LongestRun_mi	1.0	2.0	1.0	2.0	1.2
SkiableTerrain_ac	1610.0	640.0	30.0	777.0	800.0
Snow Making_ac	113.0	60.0	30.0	104.0	80.0
daysOpenLastYear	150.0	45.0	150.0	122.0	115.0
yearsOpen	60.0	44.0	36.0	81.0	49.0
averageSnowfall	669.0	350.0	69.0	260.0	250.0
AdultWeekend	85.0	53.0	34.0	89.0	78.0
projectedDaysOpen	150.0	90.0	152.0	122.0	104.0
NightSkiing_ac	550.0	NaN	30.0	NaN	80.0
resorts_per_state	3	3	3	2	2
resorts_per_100kcapita	0.410091	0.410091	0.410091	0.027477	0.027477
resorts_per_100ksq_mile	0.450867	0.450867	0.450867	1.75454	1.75454
resort_skiable_area_ac_state_ratio	0.70614	0.280702	0.013158	0.492708	0.507292
resort_days_open_state_ratio	0.434783	0.130435	0.434783	0.514768	0.485232
resort_terrain_park_state_ratio	0.5	0.25	0.25	0.666667	0.333333
resort_night_skiing_state_ratio	0.948276	NaN	0.051724	NaN	1.0
total_chairs_runs_ratio	0.092105	0.111111	0.230769	0.145455	0.107692
total_chairs_skiable_ratio	0.004348	0.00625	0.1	0.010296	0.00875
fastQuads_runs_ratio	0.026316	0.0	0.0	0.0	0.015385
fastQuads_skiable_ratio	0.001242	0.0	0.0	0.0	0.00125

4. Pre-Processing and Training Data

Next, we partitioned the data into training and testing split, with 70% of the data being used for the training split. We tested out our data using two models, linear regression and random forest regression model.

Linear Regression Model

```
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
```

According to the linear regression model, these results suggest that 'vertical drop' is our strongest positive feature. This is consistent with what we observed during the exploratory data analysis phase. The model also showed that 'Snow Making_ac' is a strong positive factor. However, the model indicated that 'skiable terrain area' is negatively associated with ticket prices, which was inconsistent with our earlier observations.

We also analyzed a random forest regression model.

Random Forest Model
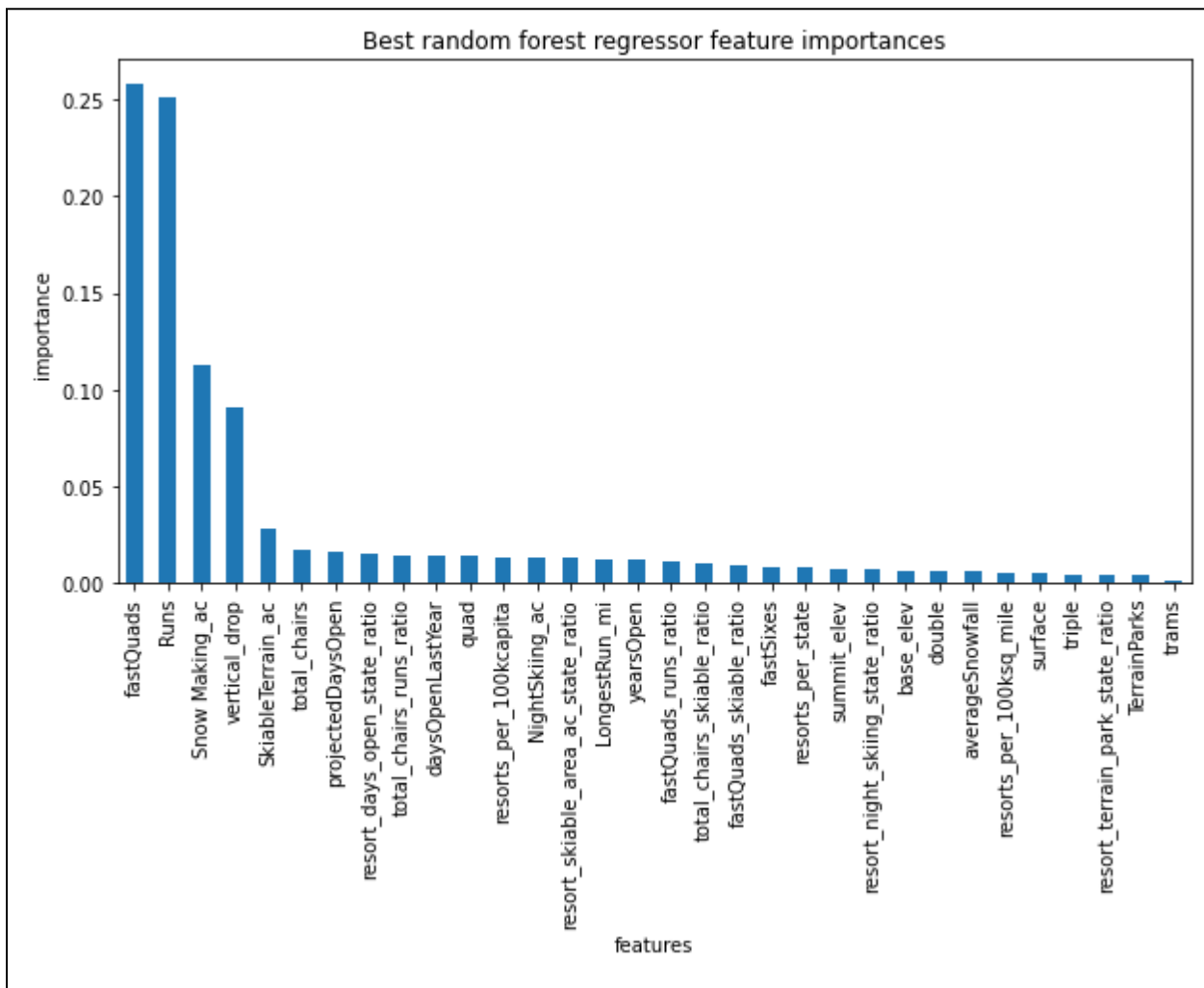
The most dominant features according to this model are fastQuards, Runs, Snow Making_ac, and vertical_drop.

To choose the best model, we calculated the mean absolute error (MAE) and sought the lowest value. In general, a lower MAE indicates that the model is making fewer errors and is more accurate. Between the two models, the random forest model has a lower cross-validation mean absolute error by almost $1 and exhibits less variability. When we tested the model's performance on the test set, the results were consistent with the cross-validation results.

5.   Modeling

After building a model for ski resort ticket prices, we examined how we could use it to gain insights into the prices that Big Mountain's facilities might be able to support. We also explored the sensitivity of changes to various resort parameters.

Using our best model, we calculated the expected Big Mountain ticket price to be $95.81, compared to the actual price of $81.00. While the expected mean absolute error is $10.39, this suggests that there is room for an increase. The validity of our model depends on the assumption that other resorts accurately set their prices based on what the market (the ticket-buying public) supports. The fact that our resort seems to be charging significantly less than what is predicted by the model suggests that it may be undercharging. However, if resorts are generally effective at pricing strategies, our model could be missing some key data, such as operating costs and visitor totals.
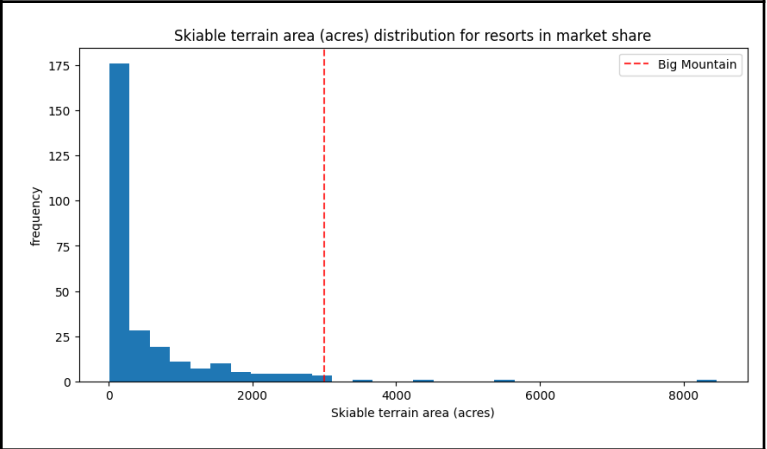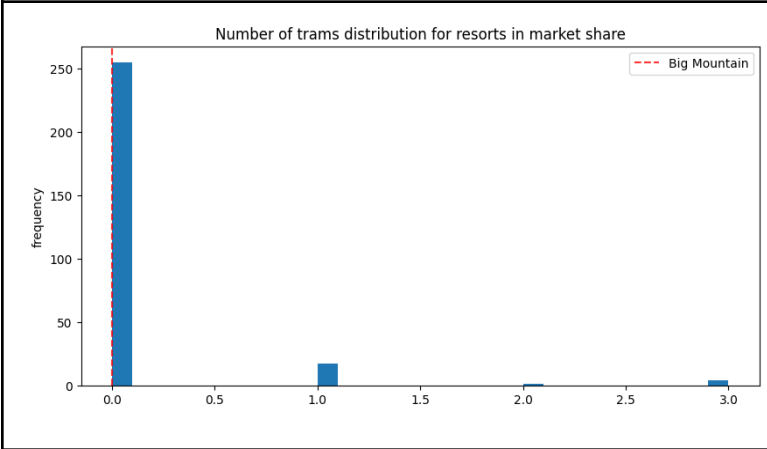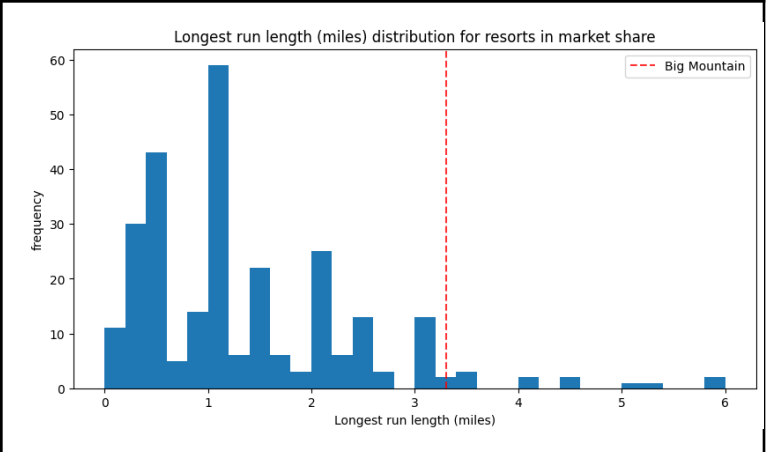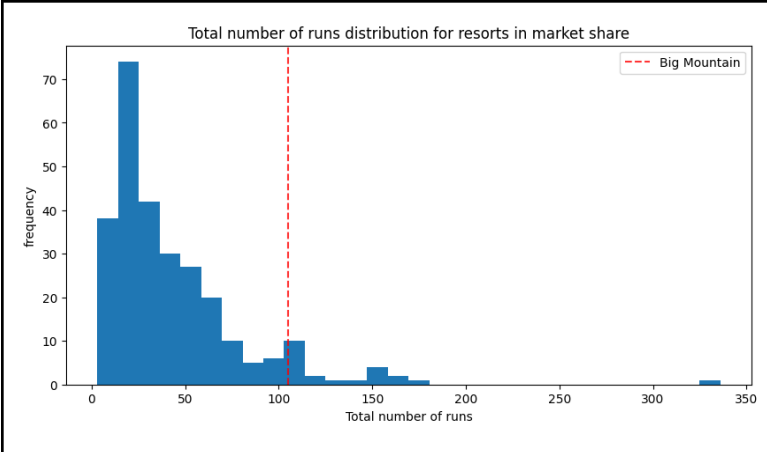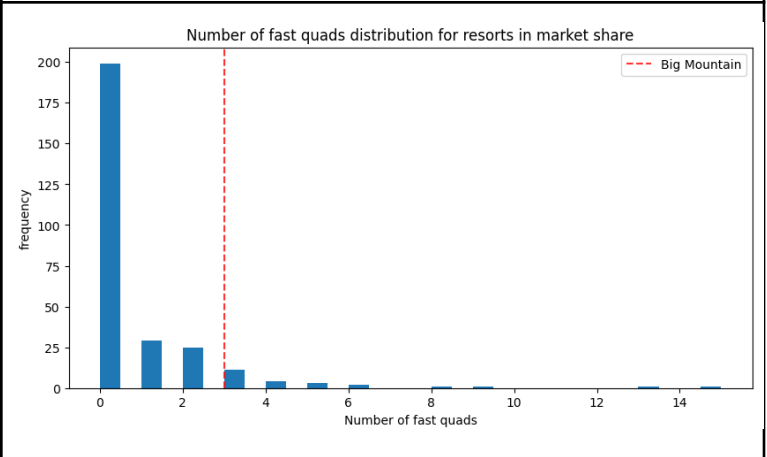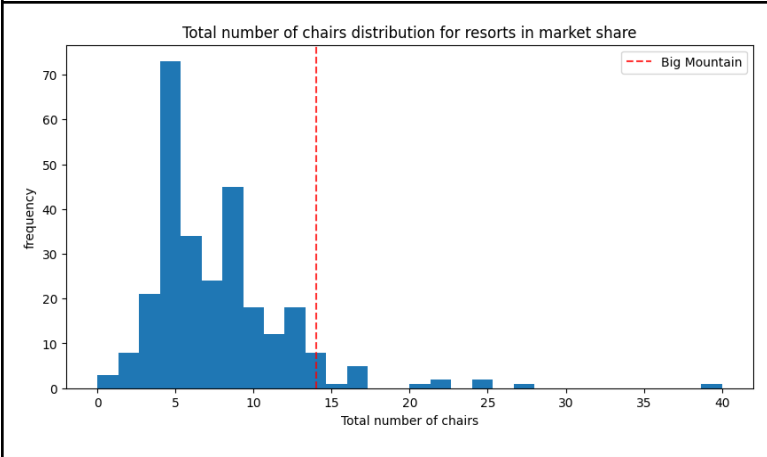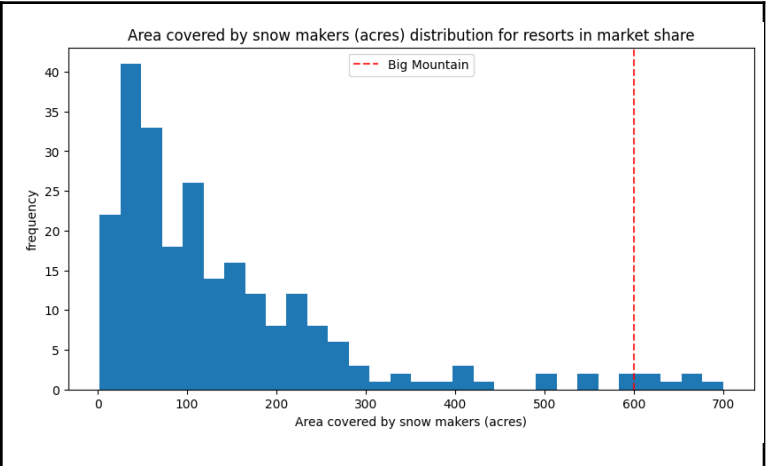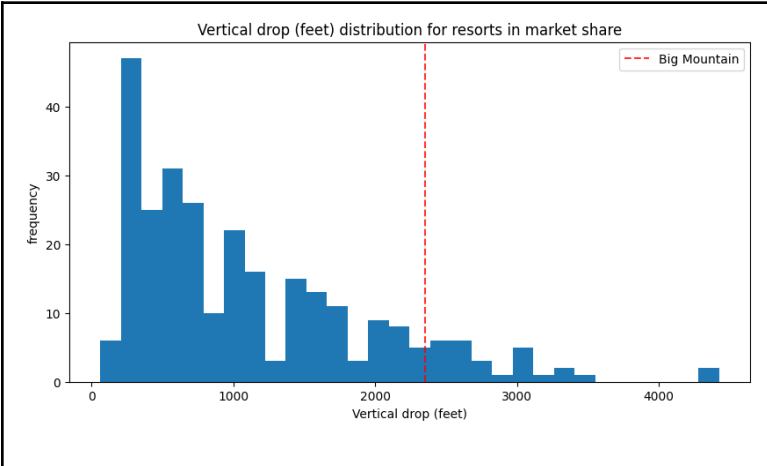
Features that came up as important in the modeling included:
- vertical_drop
- Snow Making_ac
- total_chairs
- fastQuads
- Runs
- LongestRun_mi
- trams
- SkiableTerrain_ac

This is where Big Mountain sits overall among all resorts for price and for just other resorts in Montana.
Ticket price
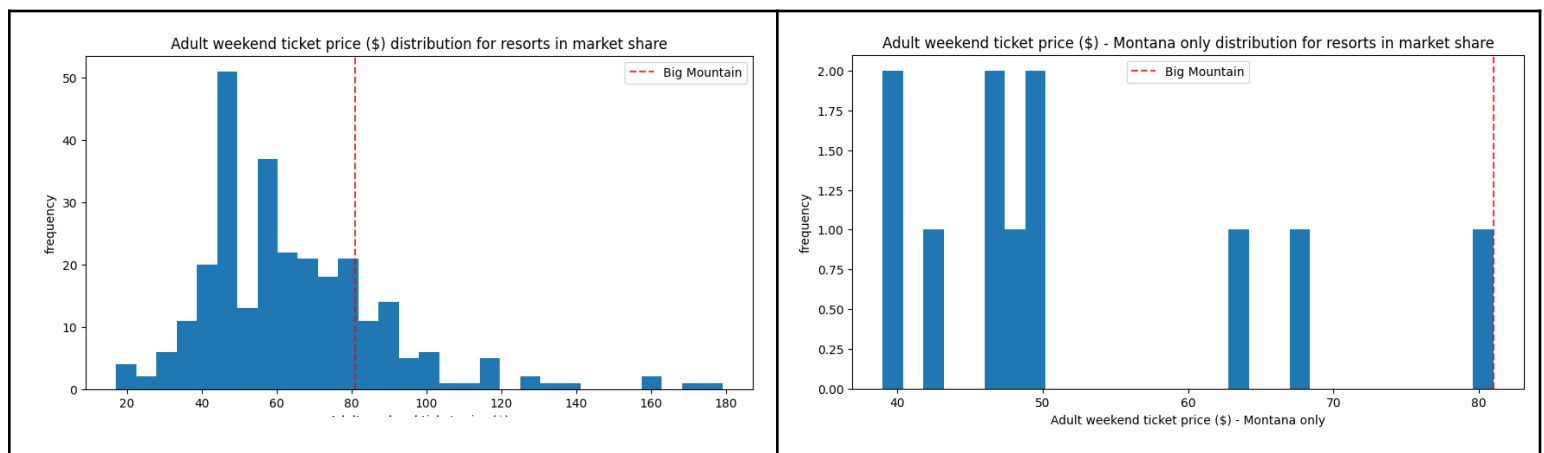
Big Mountain Resort has been considering options for either reducing costs or increasing revenue through ticket prices. The expected number of visitors over the season is 350,000 and, on average, visitors ski for five days. The data also includes information about the additional lift that Big Mountain recently installed.

Scenario 1: Permanently closing up to 10 of the least used runs. This does not impact any other resort statistics.
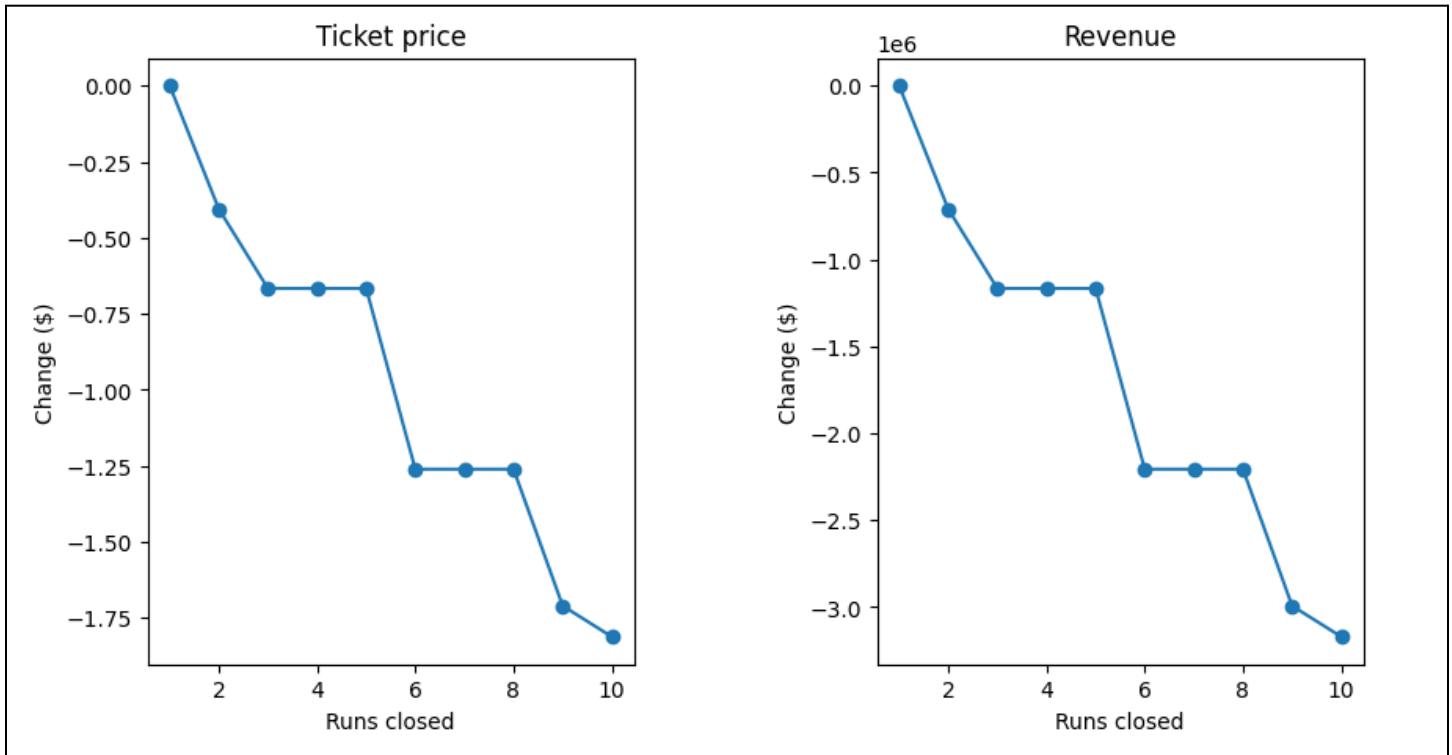
According to the model above, closing one run does not make a difference. However, closing 2 or 3 successively reduces support for ticket price and revenue. If Big Mountain closes 3 runs, it may as well close 4 or 5 as there is no further loss in ticket price. Closing 6 or more runs leads to a significant drop.

Scenario 2: Increase the vertical drop by adding a run to a point 150 feet (45.72 m) lower down, but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage.

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

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

Scenario 3: Same as scenario 2, but adding 2 acres of snow making coverage. Adding 2 acres of snow making has no effect on ticket price or expected revenue, which remains the same as in scenario 2.

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

Scenario 4: Increase the longest run by 0.2 miles (0.32 km) to a length of 3.5 miles (5.63 km), requiring an additional 4 acres of snow making coverage.

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

0.0
```

Again, this scenario does not make a difference from the current price. Although the 'longest run' feature was used in the linear model, the random forest model ranked 'longest run' low in the feature importance list.

6.   Recommendations

Big Mountain Resort is seeking to optimize the value of its facilities and increase revenue. To address this issue, we recommend that the resort implement our pricing model that consider the specific amenities and facilities offered at the resort. These amenities include, the vertical drop, snow making machines, fast quads, skiable terrains, and runs. To optimize the value of its facilities and maximize revenue, the resort should use our pricing model to track changes in ticket revenue and profitability, determine the impact on ticket demand, and ensure price increases are feasible.

In addition, our analysis has identified several scenarios that have the potential to increase revenue and improve the value of the resort's facilities. Specifically, expanding the skiable terrain, increasing the vertical drop, and adding a new chair lift have the potential to attract more visitors and increase ticket prices. We recommend that the resort consider these scenarios as part of its pricing and investment strategy.

If the business leaders find the model to be useful, it can be utilized in numerous ways. For example, the model can be used to evaluate the potential impact of various pricing scenarios, such as raising ticket prices or expanding facilities. Business analysts can also use the model to explore the sensitivity of different parameters, such as the number of runs or the size of the skiable terrain, to changes in the ticket price.

7.   Conclusion

In conclusion, our analysis has revealed that there is potential for Big Mountain Resort to increase its ticket prices to optimize the value of its facilities and maximize revenue. By tracking changes in ticket revenue and profitability, determining the potential impact on demand for tickets, and ensuring that any price increases are feasible given the resort's market positioning and competition, the resort can make informed decisions on pricing strategy. While there are limitations to our analysis, such as a lack of data on operating costs and visitor totals, our findings suggest that there is room for Big Mountain to adjust its pricing strategy to increase profitability. By carefully considering the potential impact of pricing changes on demand and revenue, Big Mountain can make informed decisions on pricing strategy that optimize the value of its facilities and maximize profitability.