Big Mountain Resort Project Report

Introduction

Big Mountain Resort, a ski resort in Montana, known for offering a spectacular views of Glacier National Park and Flathead National Forest, with access to 105 trails, recently installed an additional chair lift to help increase the distribution of visitors across the mountain. However, this additional chair is estimated to increase the operating costs by $1,540,000 this season. Management has expressed a desire to build a new pricing strategy to increase the revenue for the resort and select a better value for their ticket price.

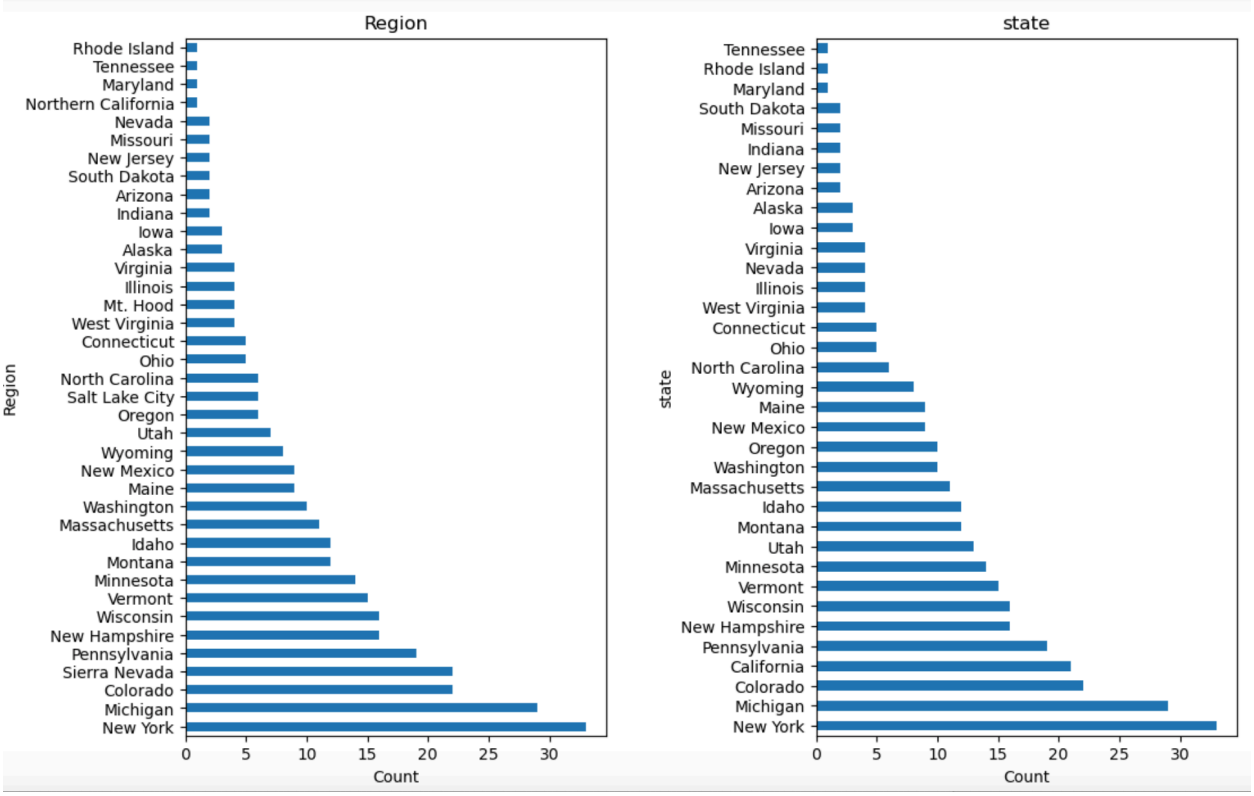
Problem Identification

What opportunities exist for Big Mountain Resort to effectively develop and implement a new pricing strategy that can maximize capitalization and increase revenue to offset their recent operating cost by $1,540,000 this season.

Data Wrangling

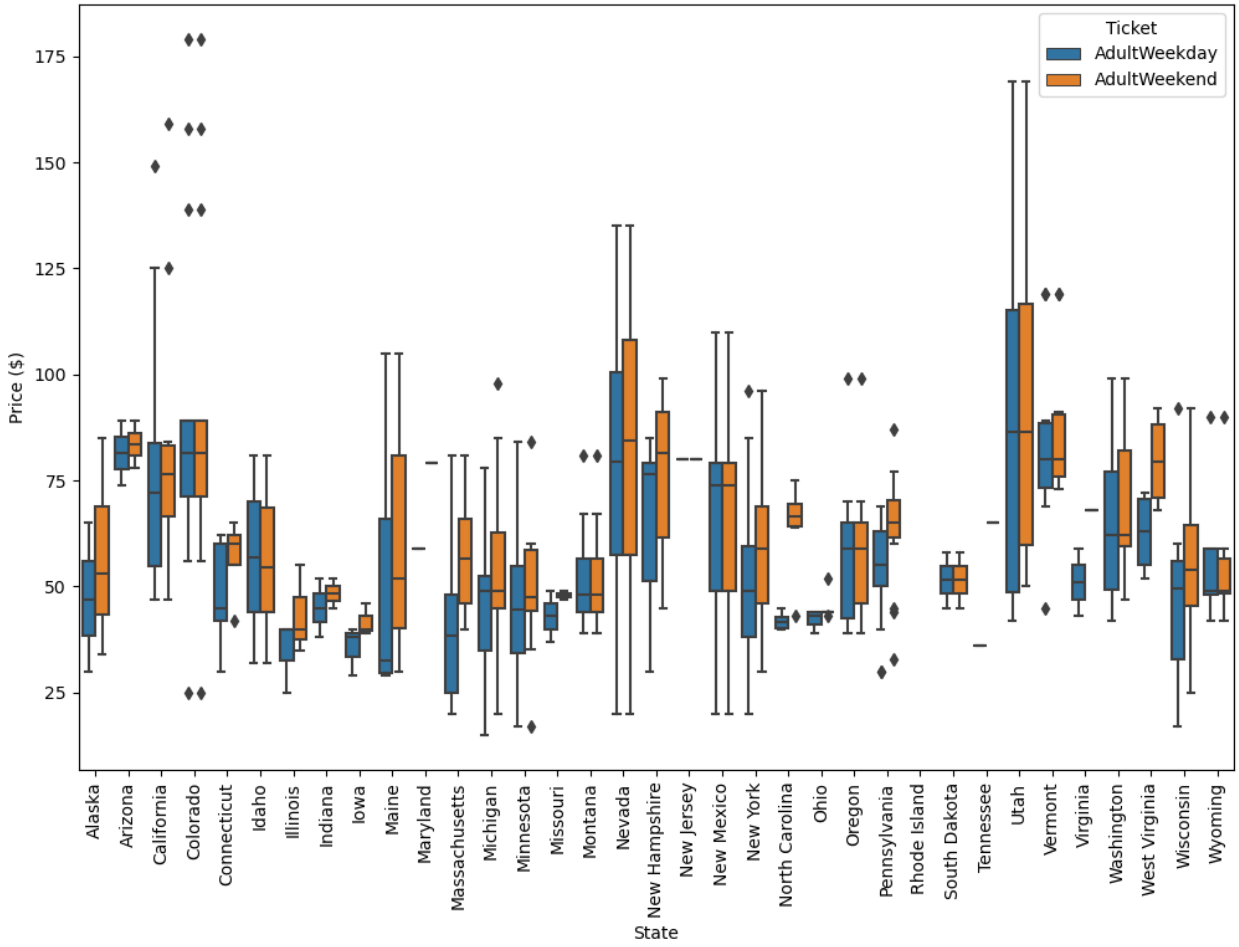
In this step, we performed a series of processes to explore, transform, and validate raw dataset retrieved into a high-quality and reliable data.

During our initial step, we compare the distribution of resorts by region and state and even though not much detail on Montana state (other than it lies at no. 13th spot of number of resorts), New York tops the state with >30 resorts in a state.

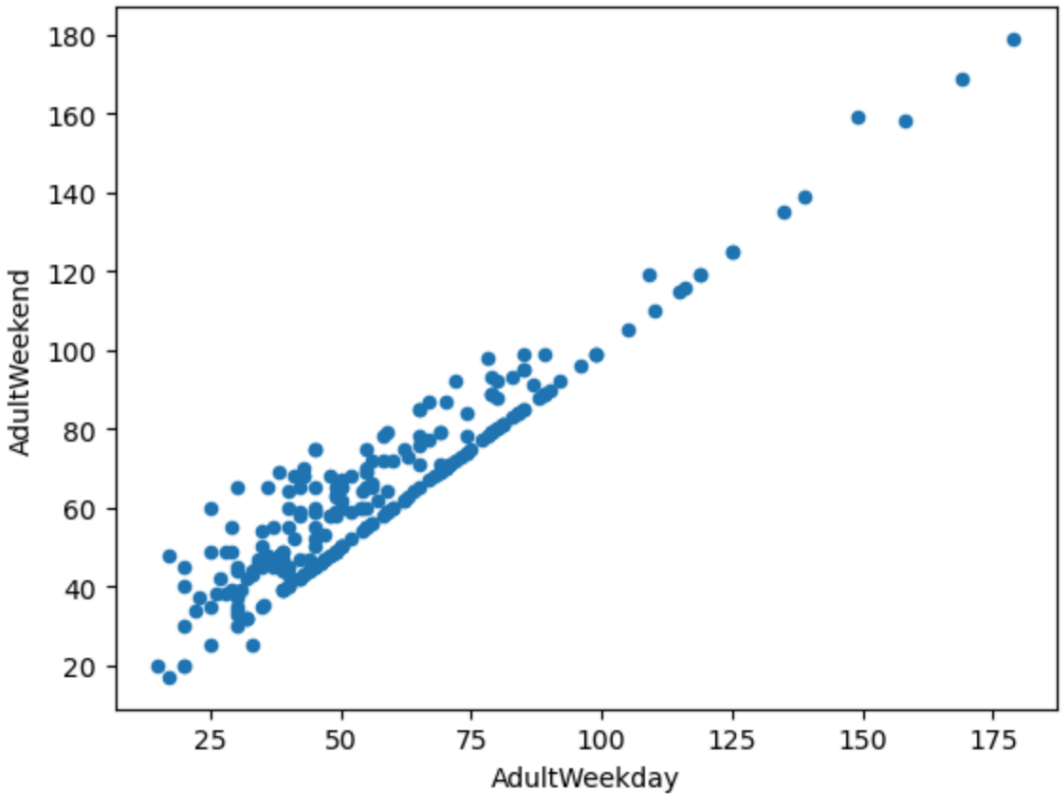


We also compare each state based on the two prices, AdultWeekEnd and AdultWeekday and the takeaways are as follows:

- Most state prices including Montana lie between $25-$100, except California, Colorado and Utah, which has relatively expensive ticket prices.
- Montana and South Dakota are the only two states with fairly small variability and have matching weekend and weekday ticket prices. Some states show more variability than others specially North Carolina and Virginia, that have weekend prices far higher than weekday prices.



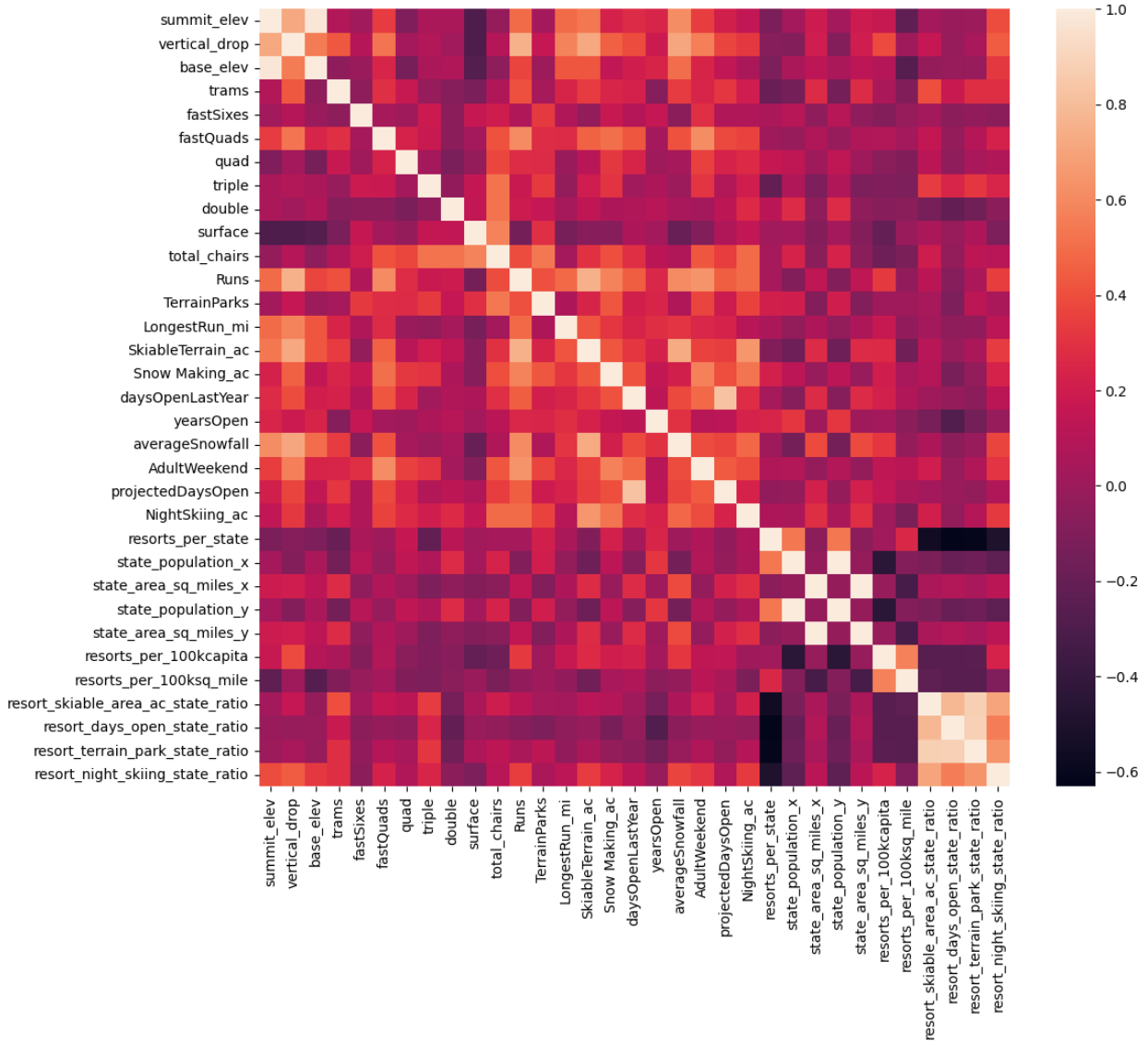
To further verify it, we explored **the relationship between "AdultWeekday" and "AdultWeekend" ticket prices in Montana state** and it confirms there is no difference in **"AdultWeekday" and "AdultWeekend ticket prices.**



Exploratory Data Analysis

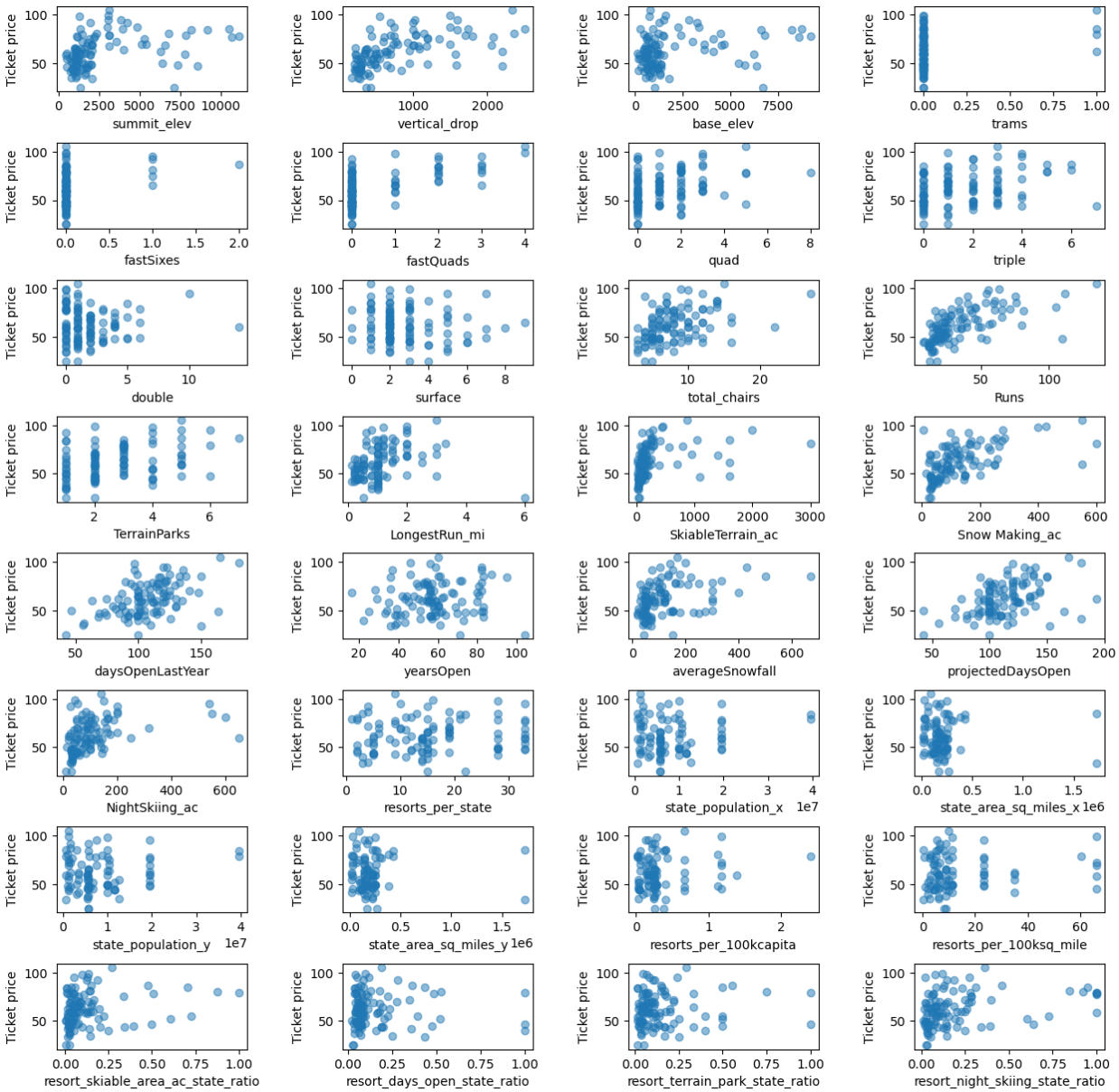
At first we performed **Principal Component Analysis (PCA)** to reduce the scope of the feature space and identify the principal components. Then, performed feature engineering which help us guide on how (or whether) to use the state labels in the data. Then, we create **Feature correlation heatmap** to gain a high level view of relationships amongst the features. An interesting observations are as follows:

- Resorts situated with densely located population provides night skiing too, as there is some positive correlation between the ratio of night skiing area with the number of resorts per capita
- A night skiing capacity is positively related to the price a resort can charge.
- Runs, total_chairs is positive correlated with ticket price too. High the total no. of chairs, higher the ticket price.



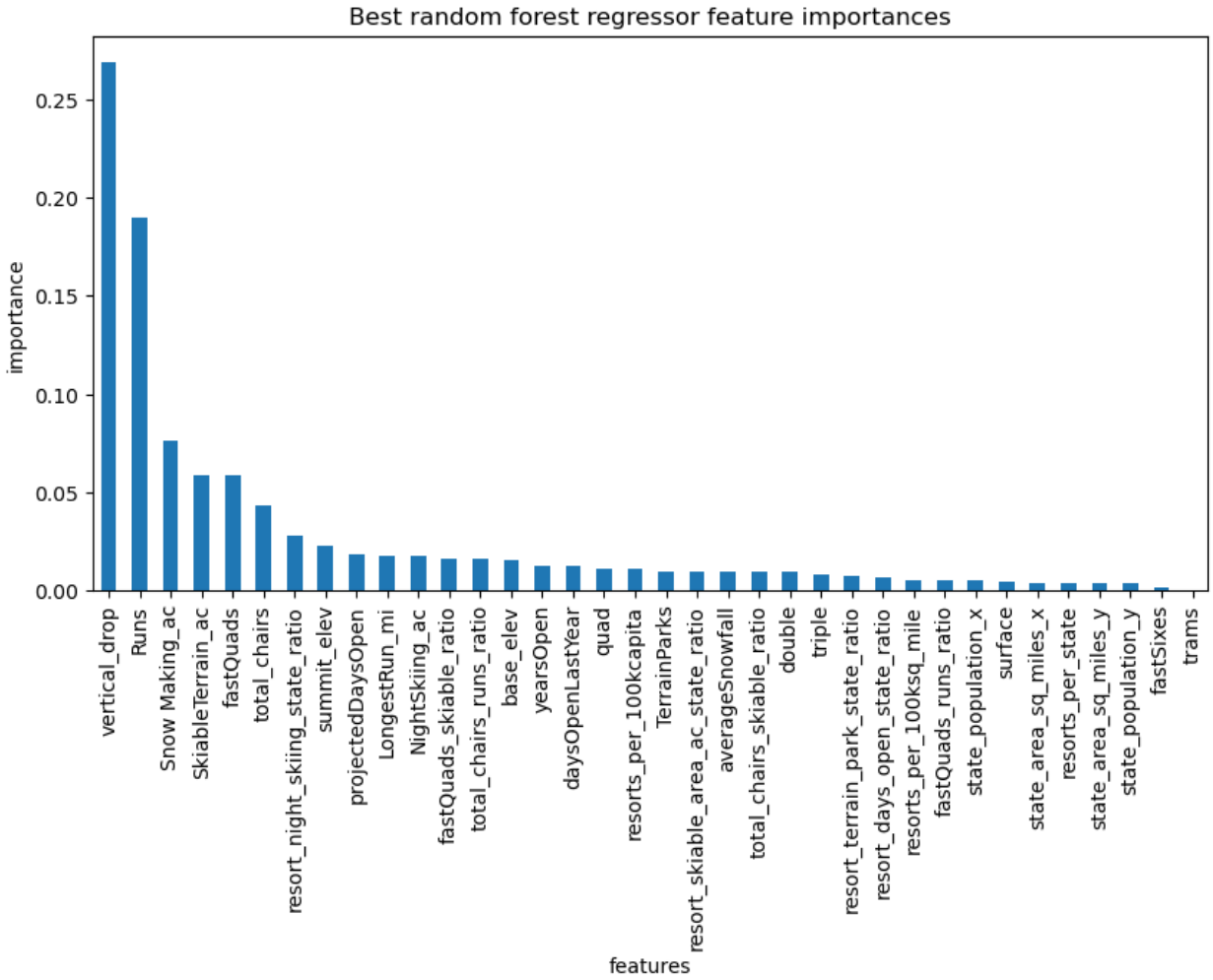
Another insights comes when we compare the features with ticket prices which are as follows:

- There's a strong positive correlation between vertical_drop and ticket prices.
- fastQuads seems very useful.
- Runs and total_chairs appear quite similar and also useful.
- Under resorts_per_100kcapita, when the value is low, there is quite a variability in ticket price, which is capable of going quite high.
- Ticket price first drop a little before climbing upwards as the number of resorts per capita increases.



Pre-processing and Training data

To further improve our data quality and make the data useful for machine modeling, we further explored the data by first Impute missing values and then select/finalize the best model (Linear Regression model or Random forest regression model) and calculate the mean absolute error using cross-validation. By using Random forest regression model, we are able to identify the dominant top features that are in common with linear model, and those are vertical_drop, Runs, Snow Making_ac, SkiableTerrain_ac, fastQuads.



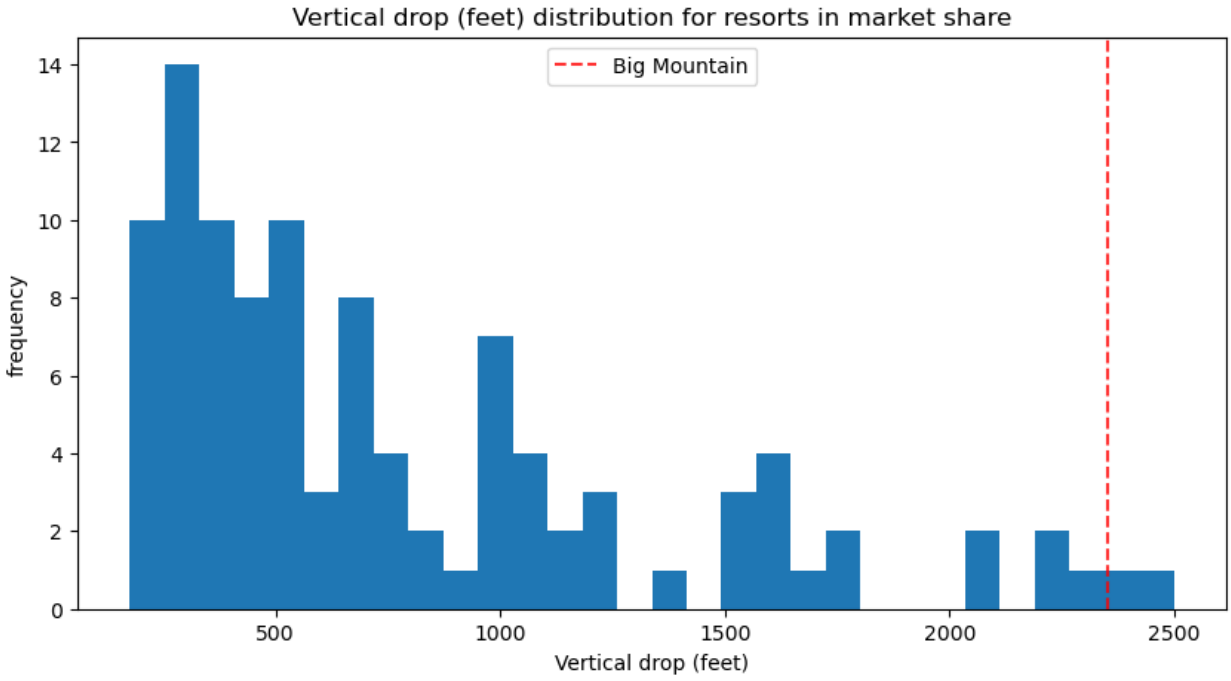
Later, we also perform the data quality assessment where we assess the cross validation score.

Modeling

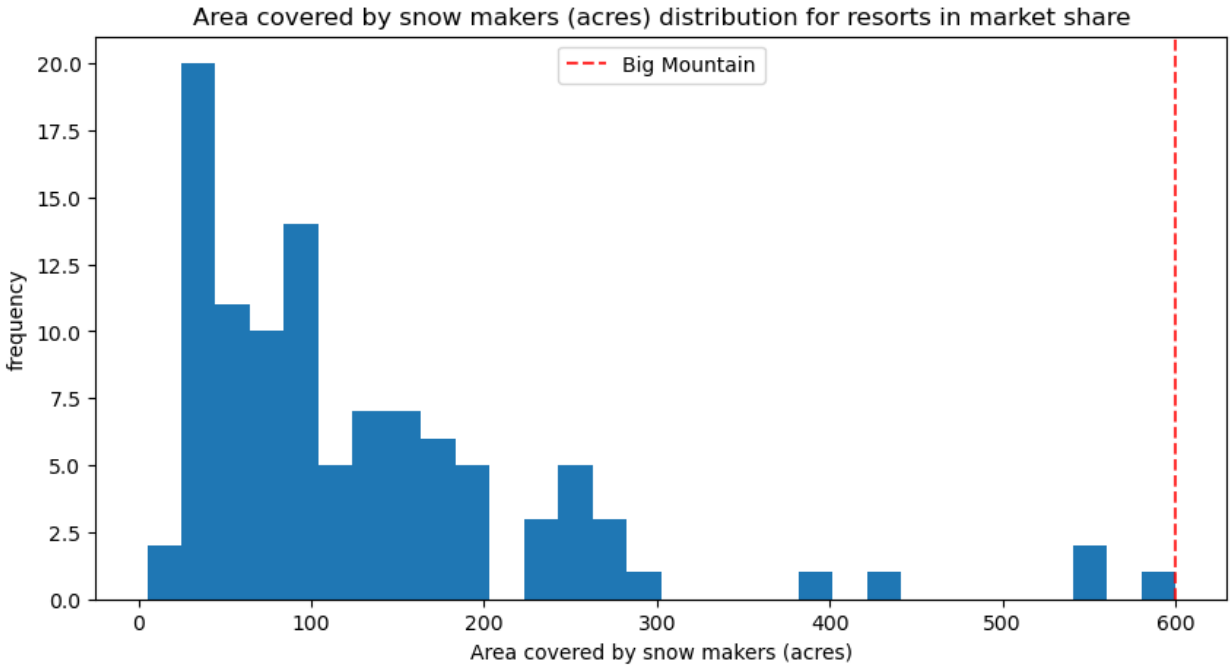
In this process, we graphically compare the distribution of some of the important features such as vertical_drop, Snow Making_ac, total_chairs, fastQuads, Runs, LongestRun_mi, trams and SkiableTerrain_ac, in terms of Big mountain specifically.

Some of the important observations are as follows:

Big Mountain is doing well for vertical drop, but there are quite a few resorts with a greater drop of 2500 feet.

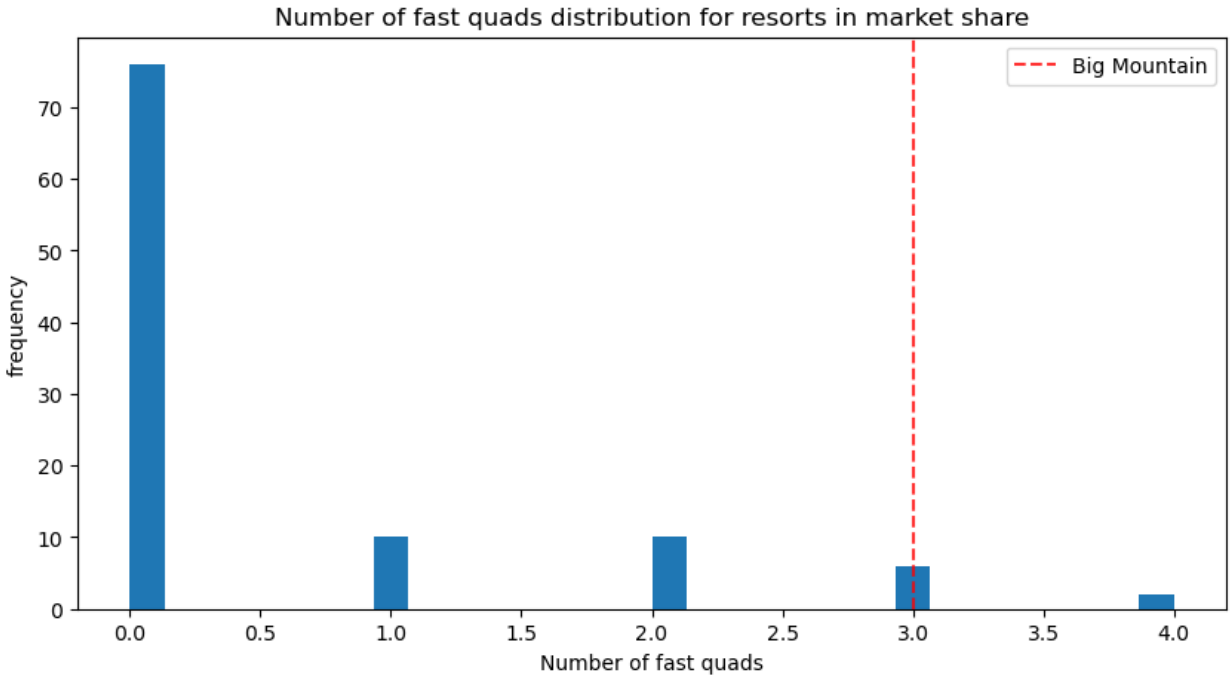


Big Mountain tops the league table of snow making area.

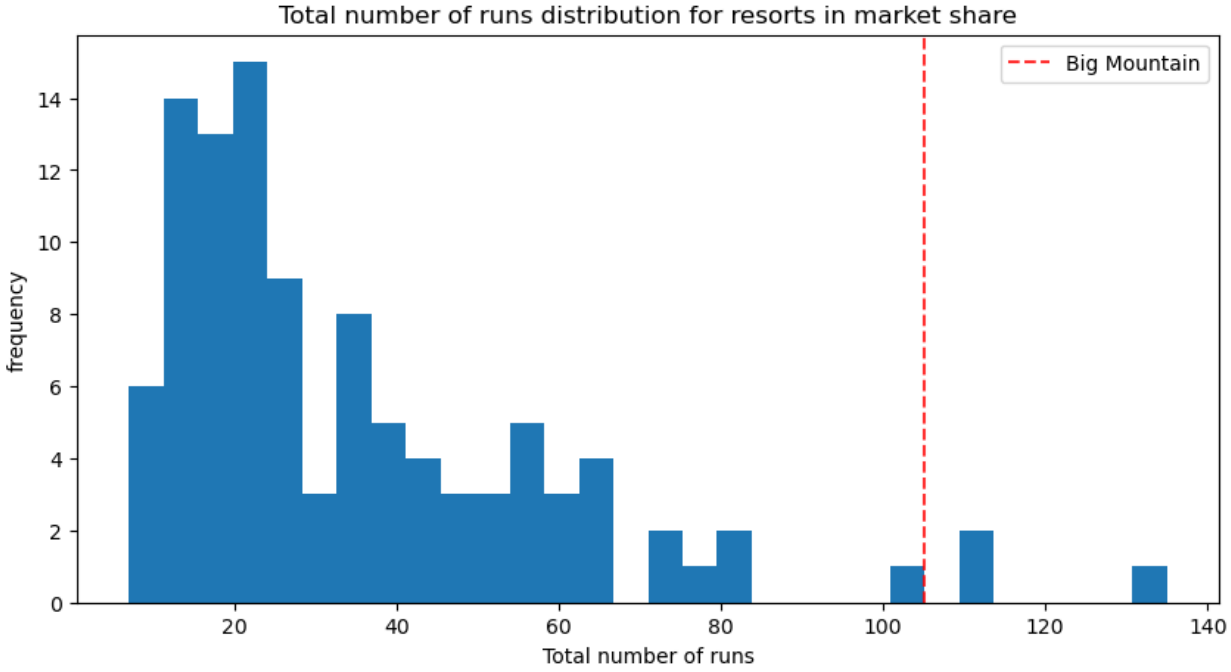


Big Mountain has amongst the highest number of total chairs, resorts with more appear to be outliers.

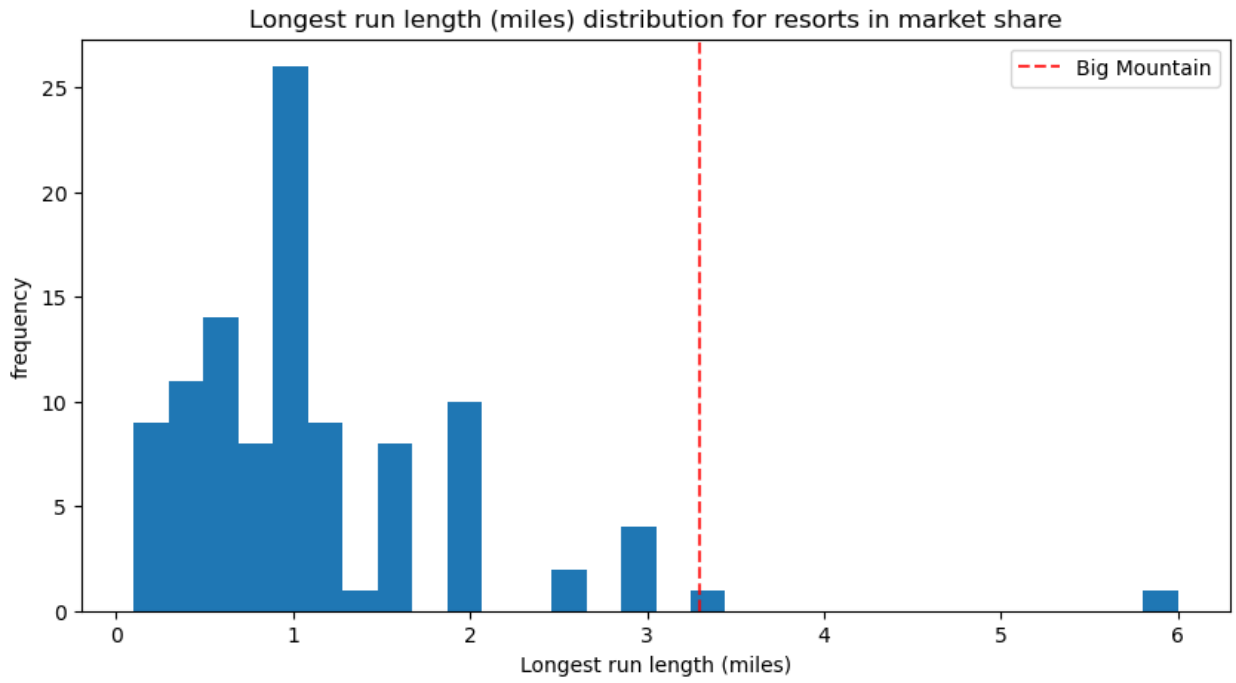
Most resorts have no fast quads. Big Mountain has 3, which is quite high, though there are some values much higher, but they are rare.



In terms of Runs, Big Mountain compares well for the number of runs. However, there are some resorts with more, but not many.

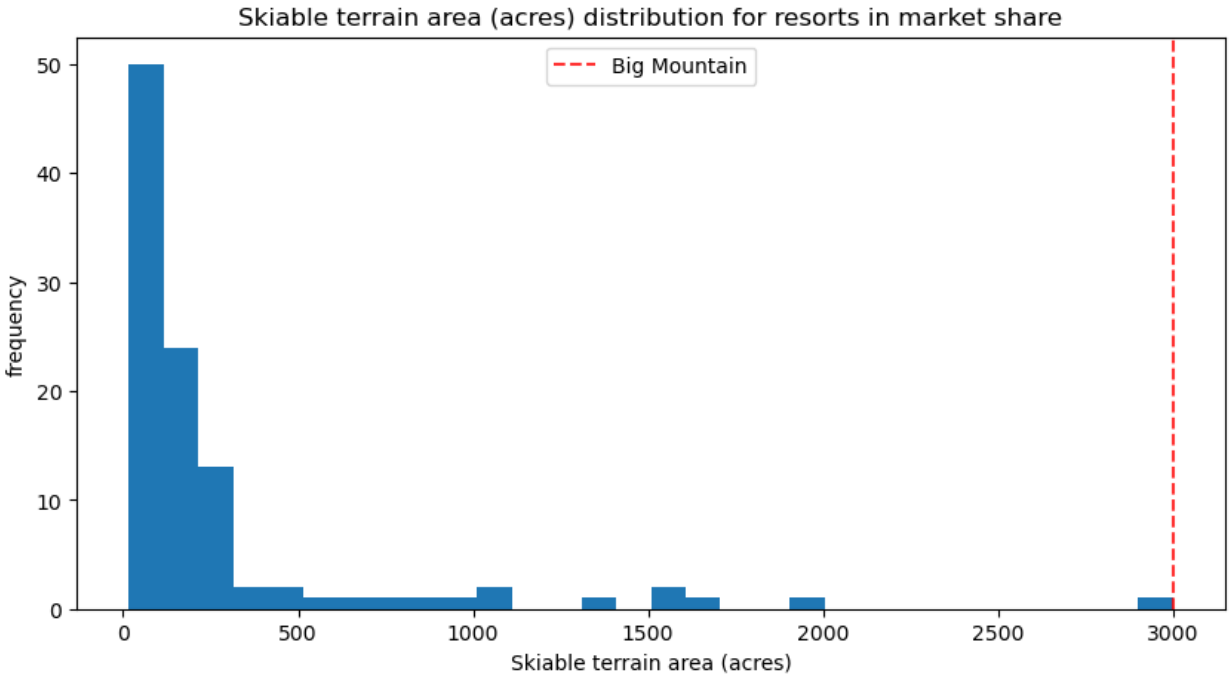


In terms of longest runs, Big Mountain is exceeding. Although it is just over half the length of the longest, the longer ones are rare.

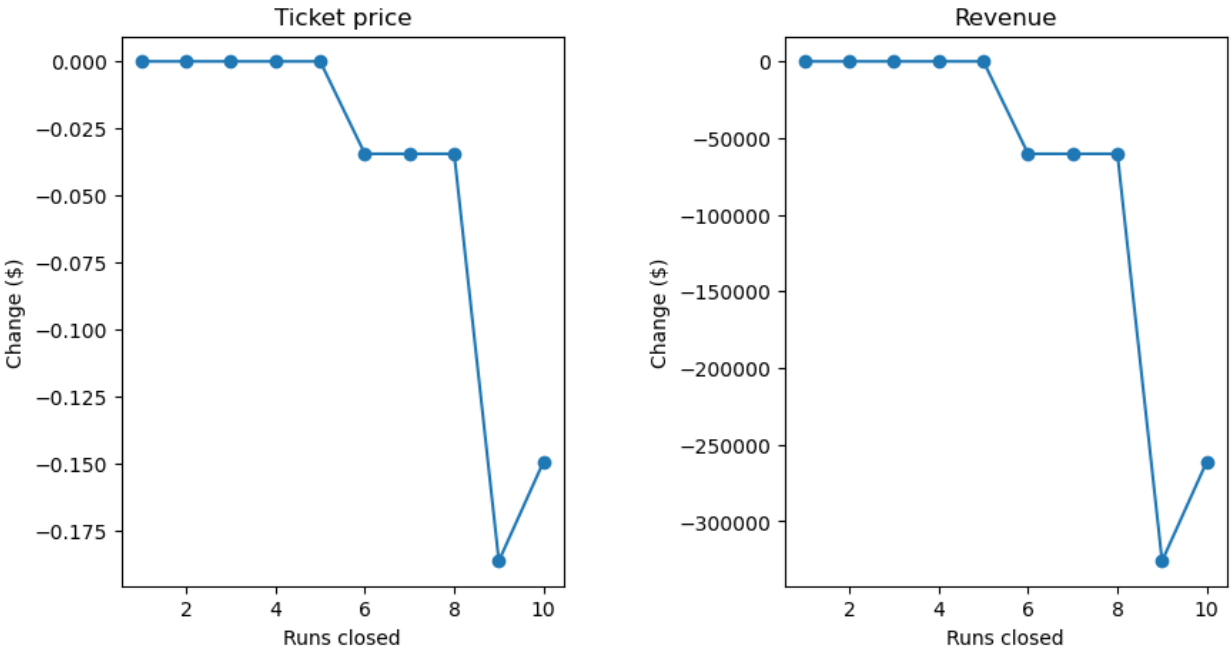


The vast majority of resorts, including Big Mountain, have no trams.

Finally, Big Mountain is amongst the resorts with the largest amount of skiable terrain.



We also analyzed the impact of scenario as what if we close up to 10 of the least used runs, where we created two plots for the predicted ticket price change (delta) and revenue change and the model shows there's no change until 5 close runs but Increasing the closures down to 6 or more leads to a sharp drop.



Recommendations

I'd recommend adding a run, increase the vertical drop by 150 feet, and installing an additional chair lift and still be able to increase our ticket price support by $0.14 which totaled to $247768 over the season.

Other than ticket prices, it would be interesting to learn more about the data associated with day-to-day operational costs, actual number of visitors, that can affect profitability. As we have seen Big Mountain Resort modeled price came as 90.12 dollars with expected mean absolute error of 9.57 dollars, which suggests there is a room for an increase.

Another recommendation is that there has to be a centralized system/dashboard for visualization and further exploration.