**Big Mountain Resort Recommendation**

# Problem Statement

* For the upcoming ski season, what is the most cost efficient ticket price Big Mountain Resort can charge its expected 350,000 customers, with the additional cost of the chair lift ($1.54m) in mind?

# Data wrangling major takeaways

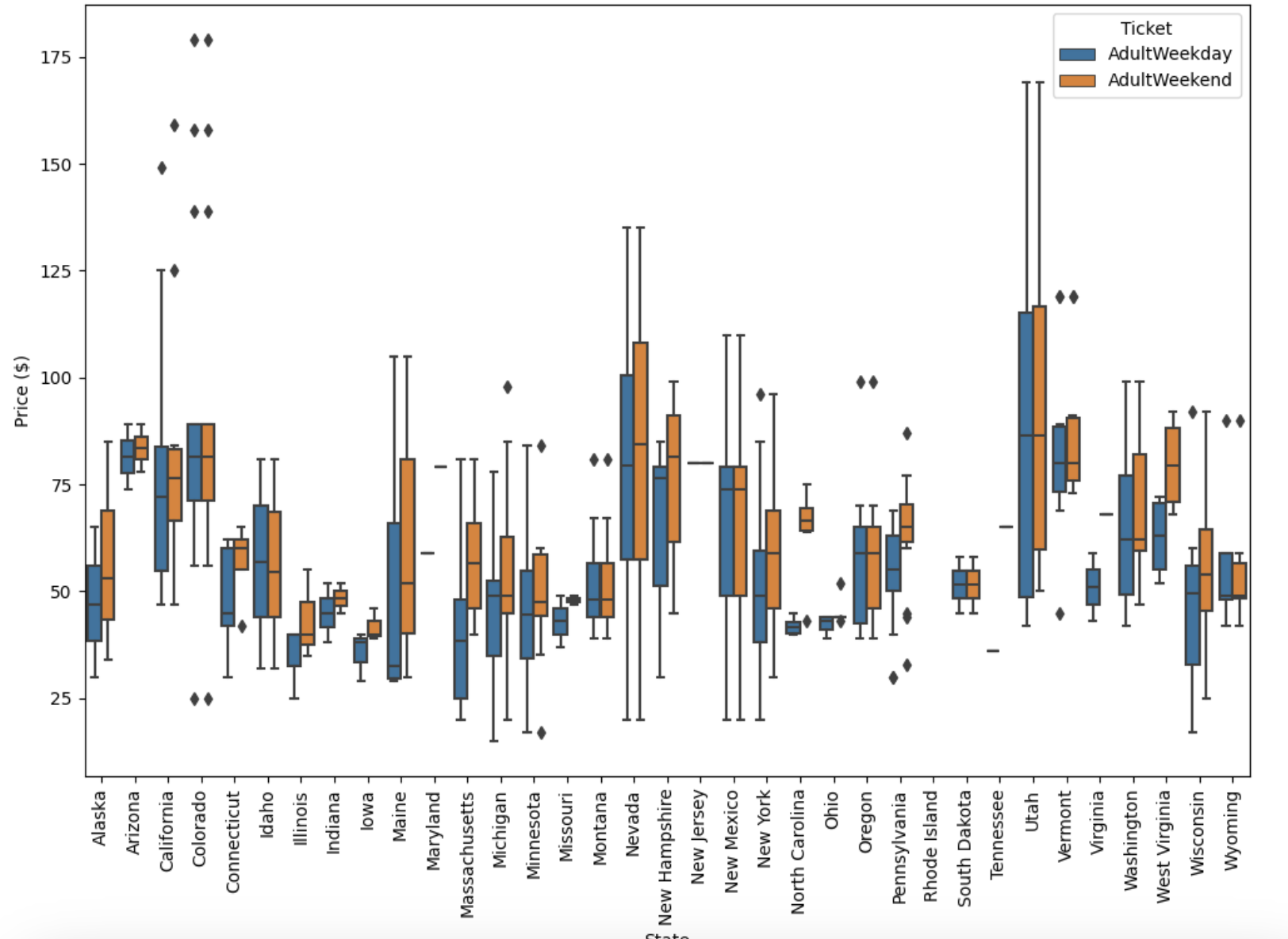
* Proportioning the total amount of N/As within each field of the base ski dataset was the first priority. Fast Eight and Night Skiing showed a high percentage of N/As and posed a large problem to the integrity of those fields.

2a.



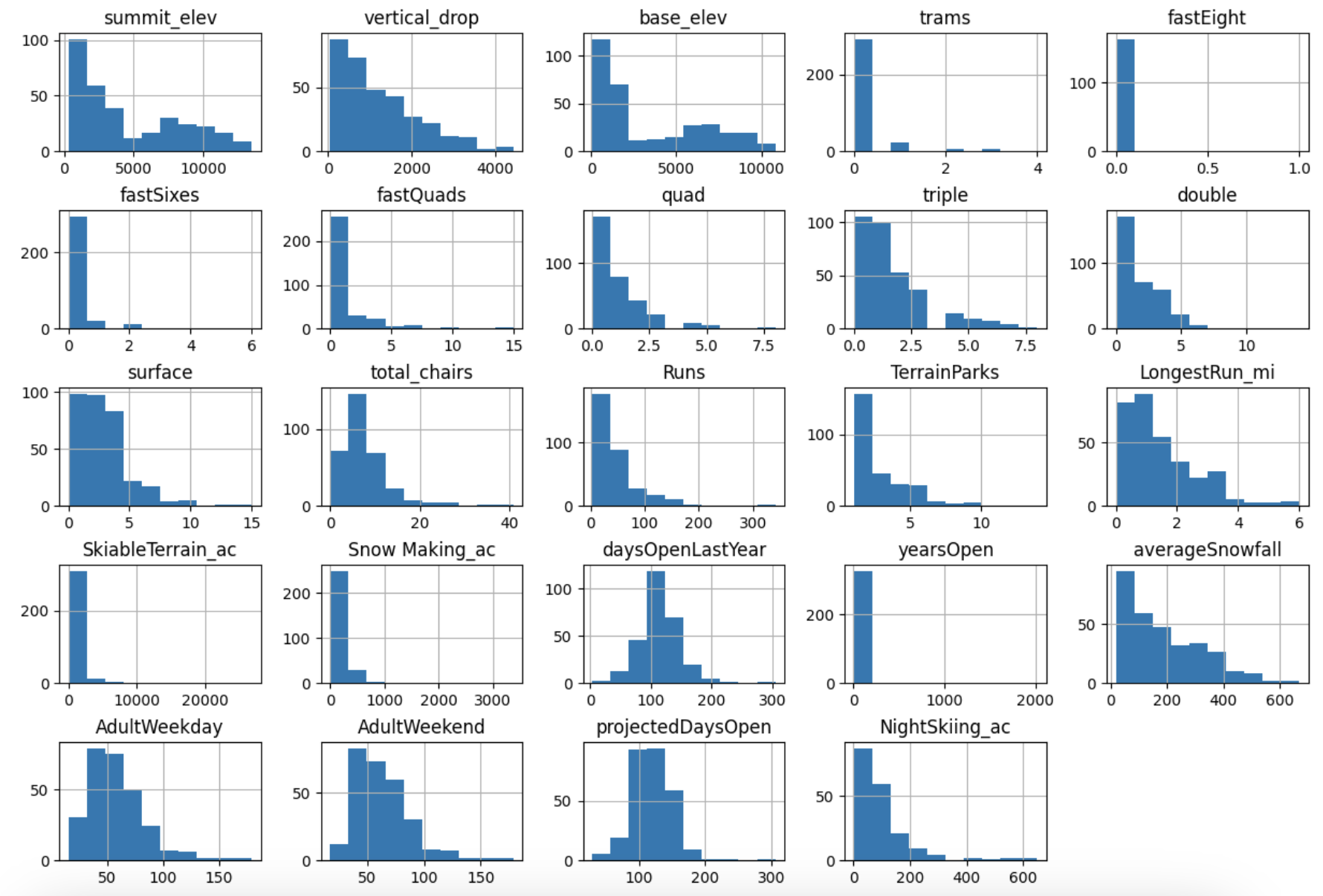
* After, understanding the shape of the data for each field and how they skew was beneficial for unearthing any trends or further areas of interest. Overall, average ticket price by state (2b) displayed the highest range of pricing from California, Utah, and Vermont. Less variability was seen in Montana and South Dakota.

2b.



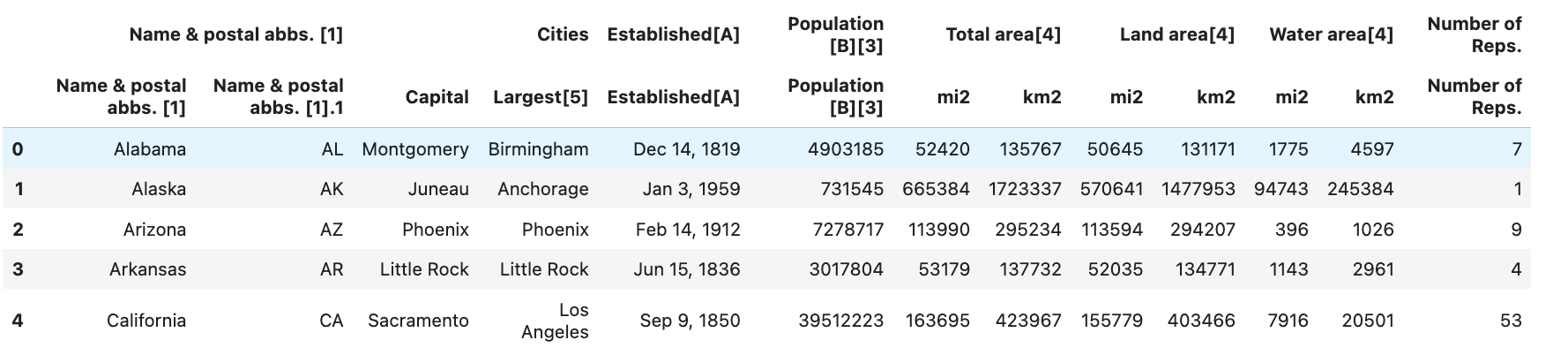
* When looking at features and their distributions (2c), most have a right tail distribution, aside from Skiable terrain, which is mainly concentrated to the lower amount of acres.

2c.



* Finally, is a look at the raw data (2d) being appended to strengthen the ski dataset’s density measures. It contains various qualitative and quantitative columns that will be valuable moving forward.

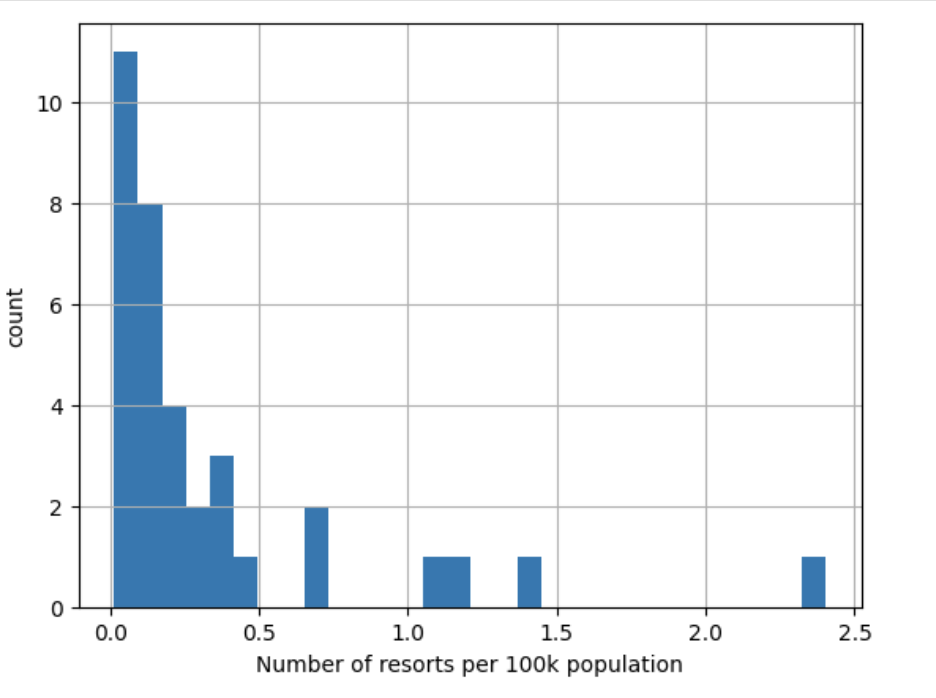
2d.



# Exploratory data analysis major takeaways

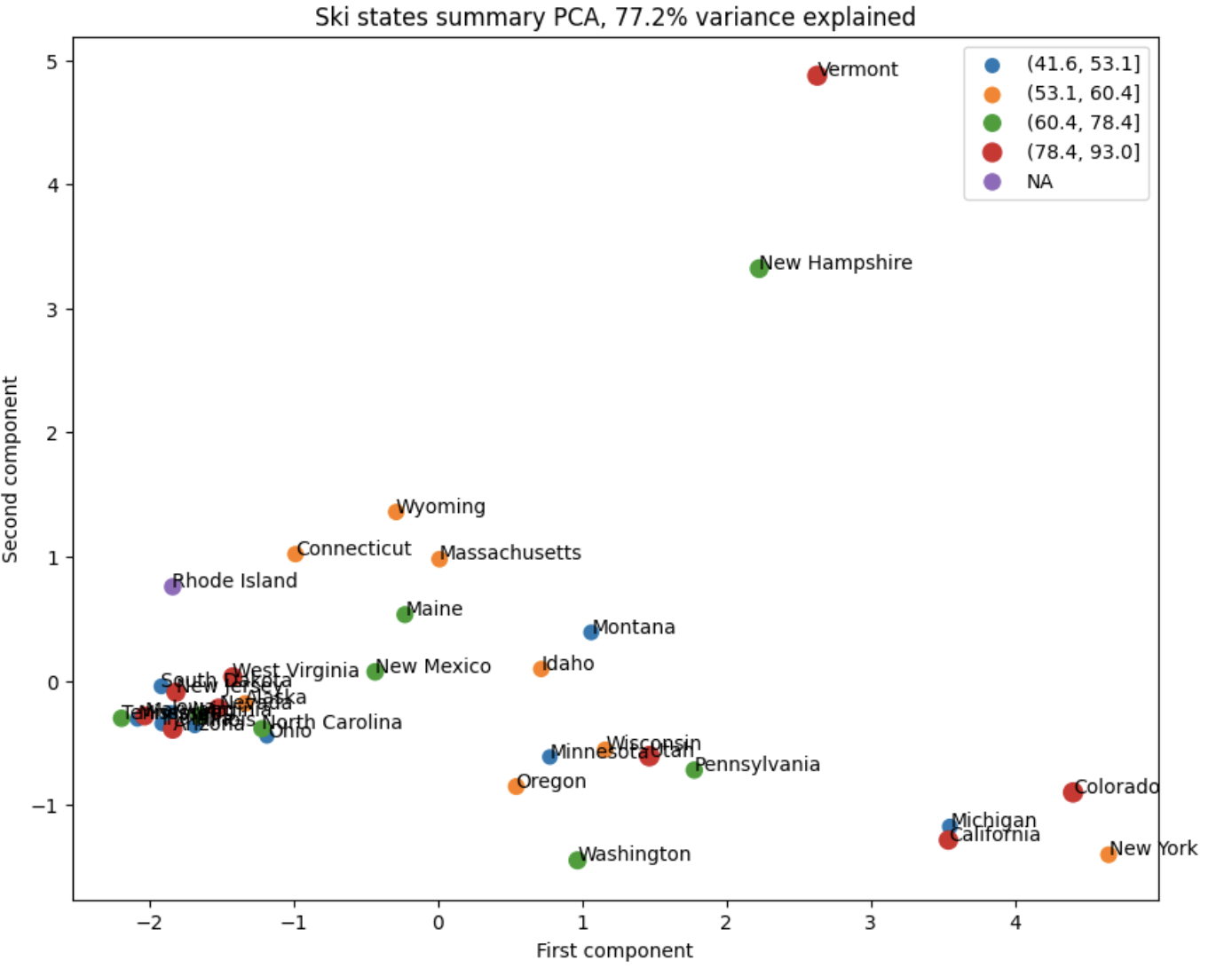
* A natural transition from data wrangling to the EDA phase illuminates the benefits of merging the ski dataset with the state data. The figure below (3a) shows the relationship between resorts and population, as a distribution; explaining that there are not many resorts that are within range of a 100k population, stated by the right tail distribution (with a large amount of values close to 0.0). As a result, this finding is an example of one of the many checks on the dataset to contextualize how the data is shaped.

3a.



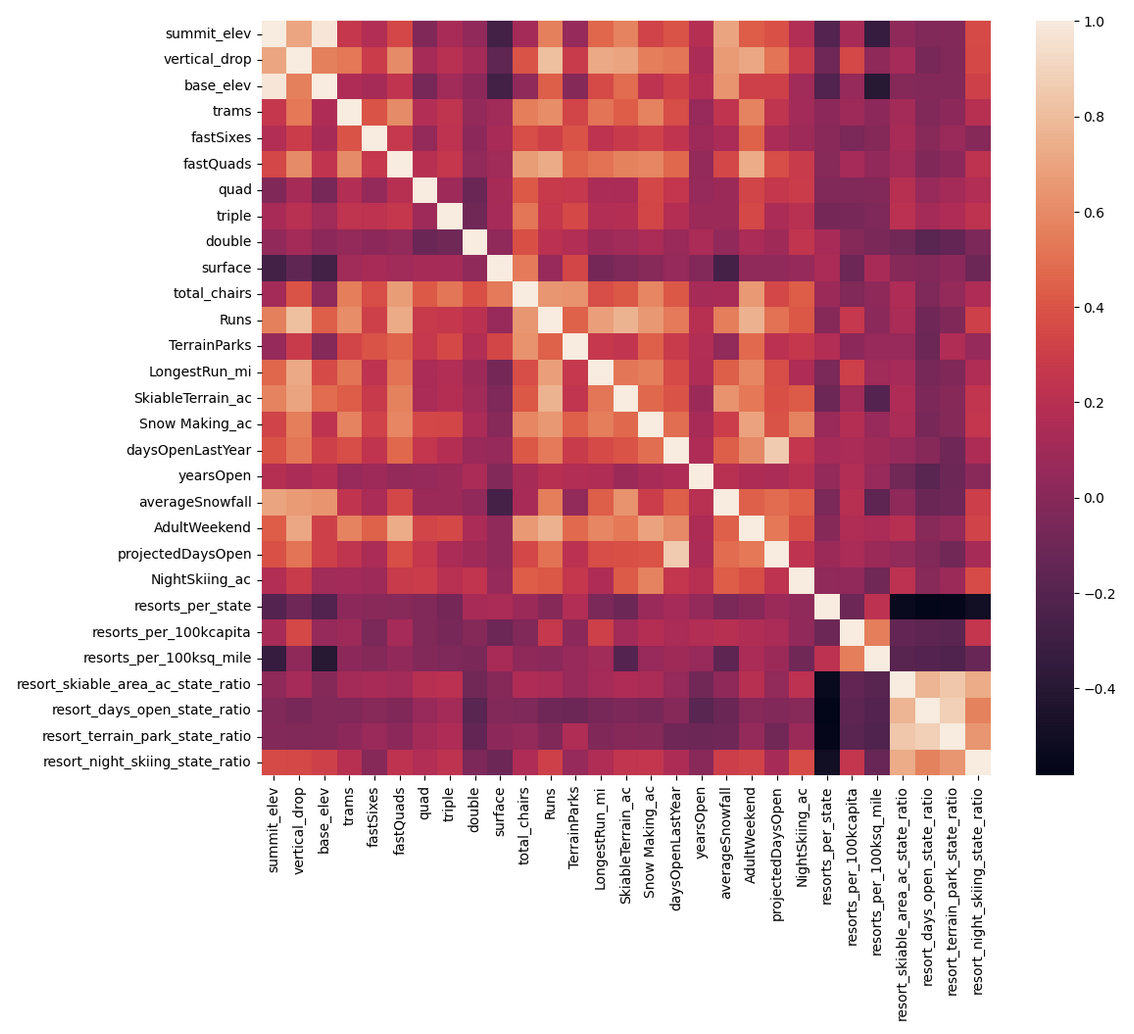
* The next step determined what features have the most influence on the dataset by taking a principal component analysis (3b) of the merged dataset, which is indexed by state. The PCA, which reduces the dimensionality of this large dataset and shows how the top two components fare for each state and layers in the quartiles of adult weekend prices.

3b.



* Correlation between fields is also measured using a correlation heatmap (3c). This provides a high level view of the relationships between all fields. Our price field stands out for obvious reasons, but the consensus top fields that show strong correlation include: runs, snow making acres, and fastquads.

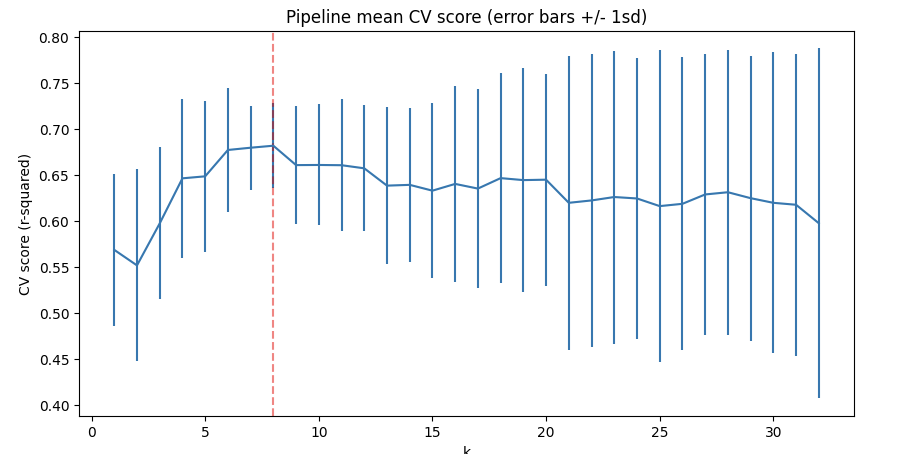
3c.



# Preprocessing and training data

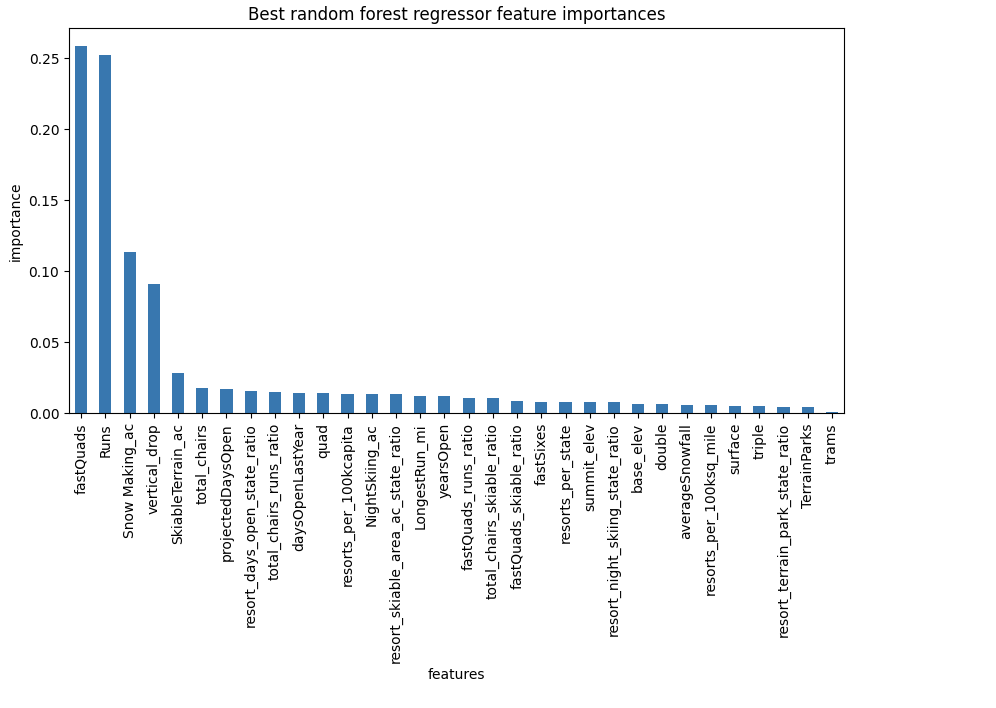
* The refined dataset required a data split into training and test groupings for modeling purposes. Training set is weighted at 70% while the test data takes the remaining 30%, and the x and y are the rest of the data and the adult weekend price field, respectively.
* Imputations were performed to account for the other missing values in the dataset.
* From the initial linear model using a pipeline (4a), an expedited form of scaling features that accepts inputs to fit, predict, and conducting a regression in a simple format, the right amount of features are selected

4a.



* Finally, more parameters were inputted such as random forest regressors, and cross validation was assessed to improve results and finalize the model (4b). The top features are shown below.

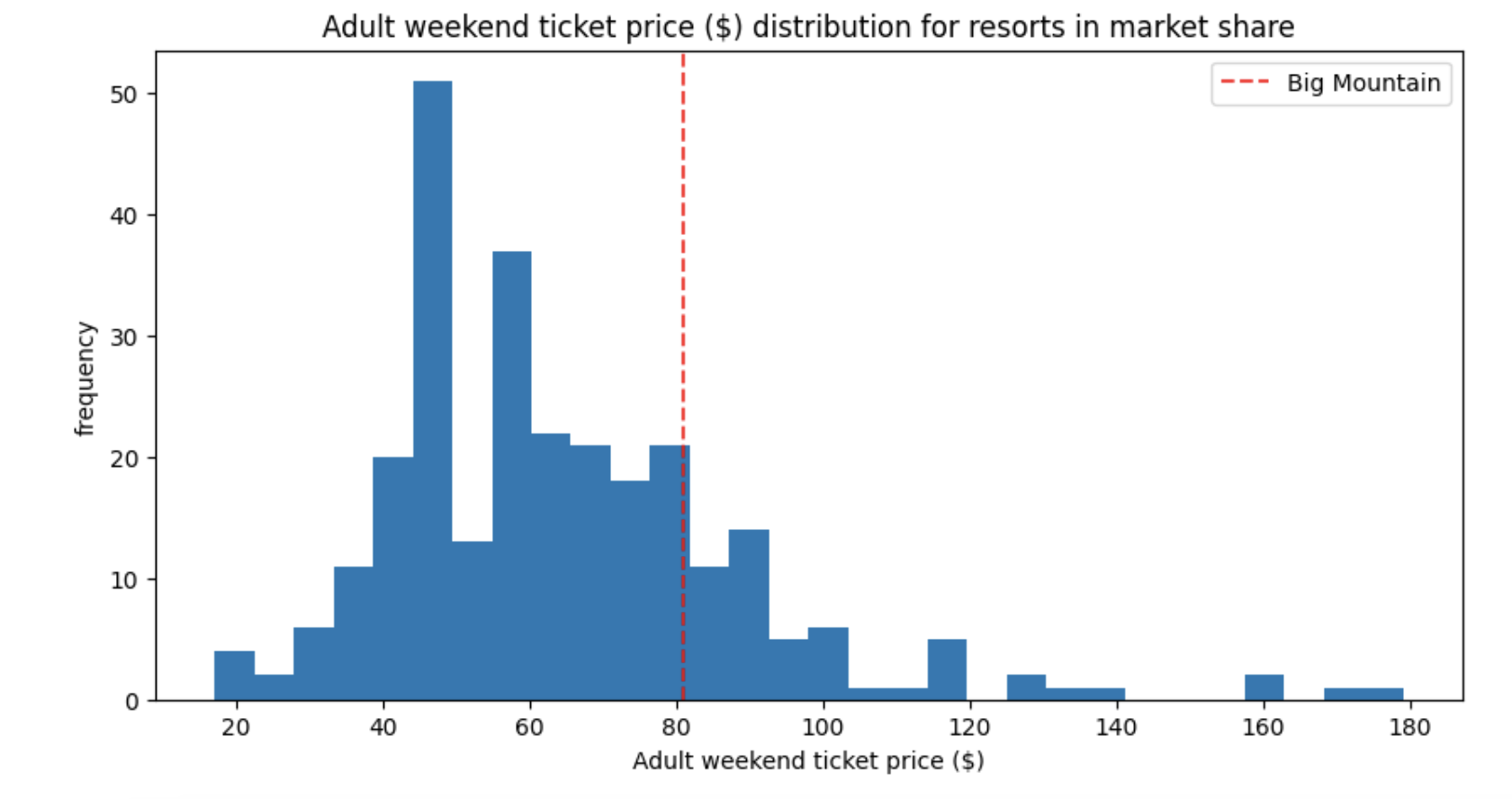
4b.



# Modeling

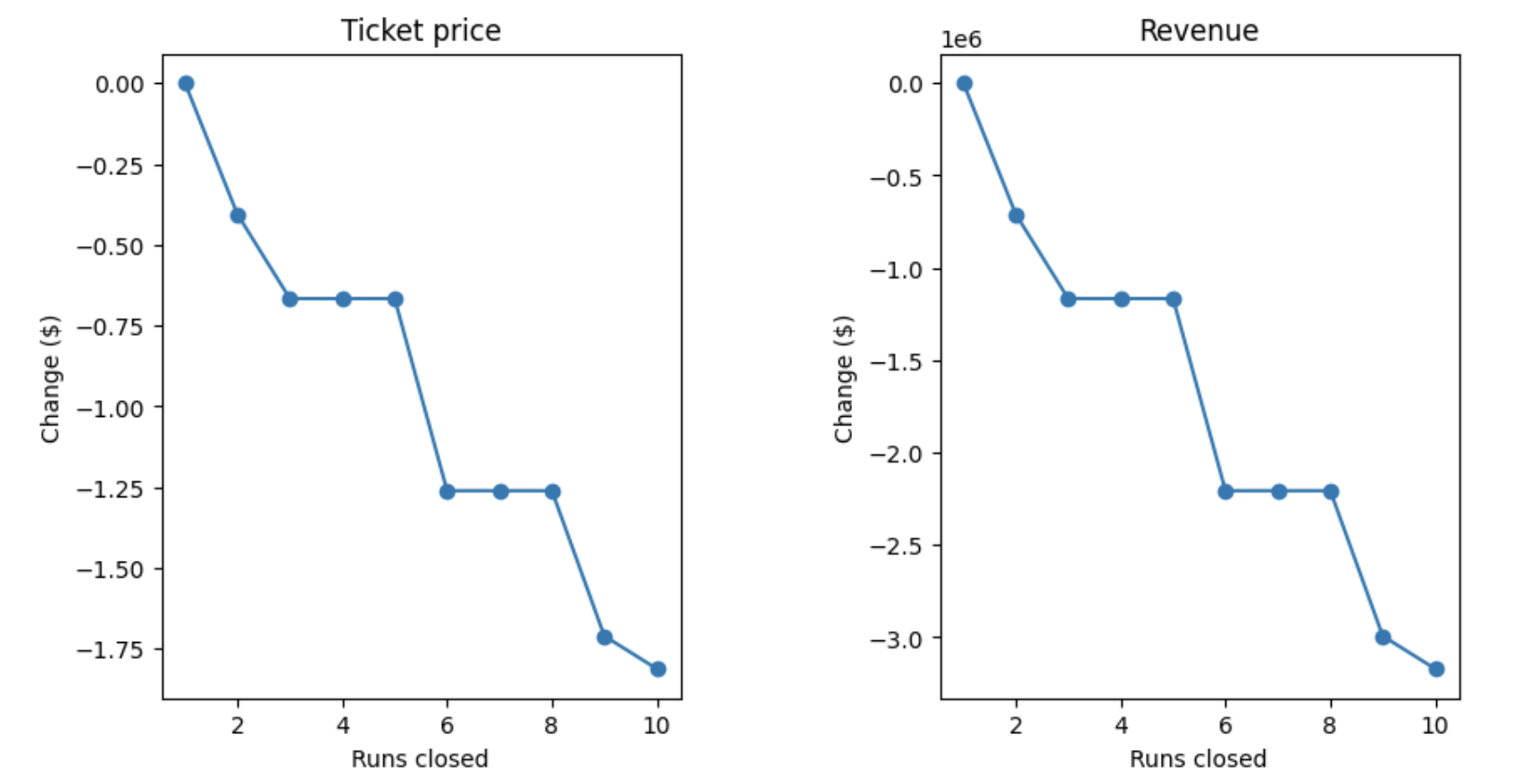
* Several tests are performed to ensure the veracity of the model. The difference between actual price and modeled price is determined to be around $14.
* Understanding Big Mountain Resort in relation to competitors was the next step in the progression

5a.



* Scenarios that impact ticket price, both increase and decrease, are mocked. One of which shows the relationship of number of runs closed and ticket price and number of runs closed and revenue. Ultimately, there are certain intervals where increasing the number of closed runs does not impact ticket price or revenue.

5b.



# Recommendation

* With operating costs of a new chair lift at $1.54m, and it being factored into our estimation for increases in features, Big Mountain Resort should be aggressive in their stance on pricing given their position in the market and the potential to expand the business. The features that are most relevant to this recommendation are runs, vertical drop, total chairs, and snow making acres. Increasing these is paramount to the increase in ticket price and revenue. This move would output more than $15m to the business, lifting overall ticket price by $9 and would more than offset the cost of the new lift.