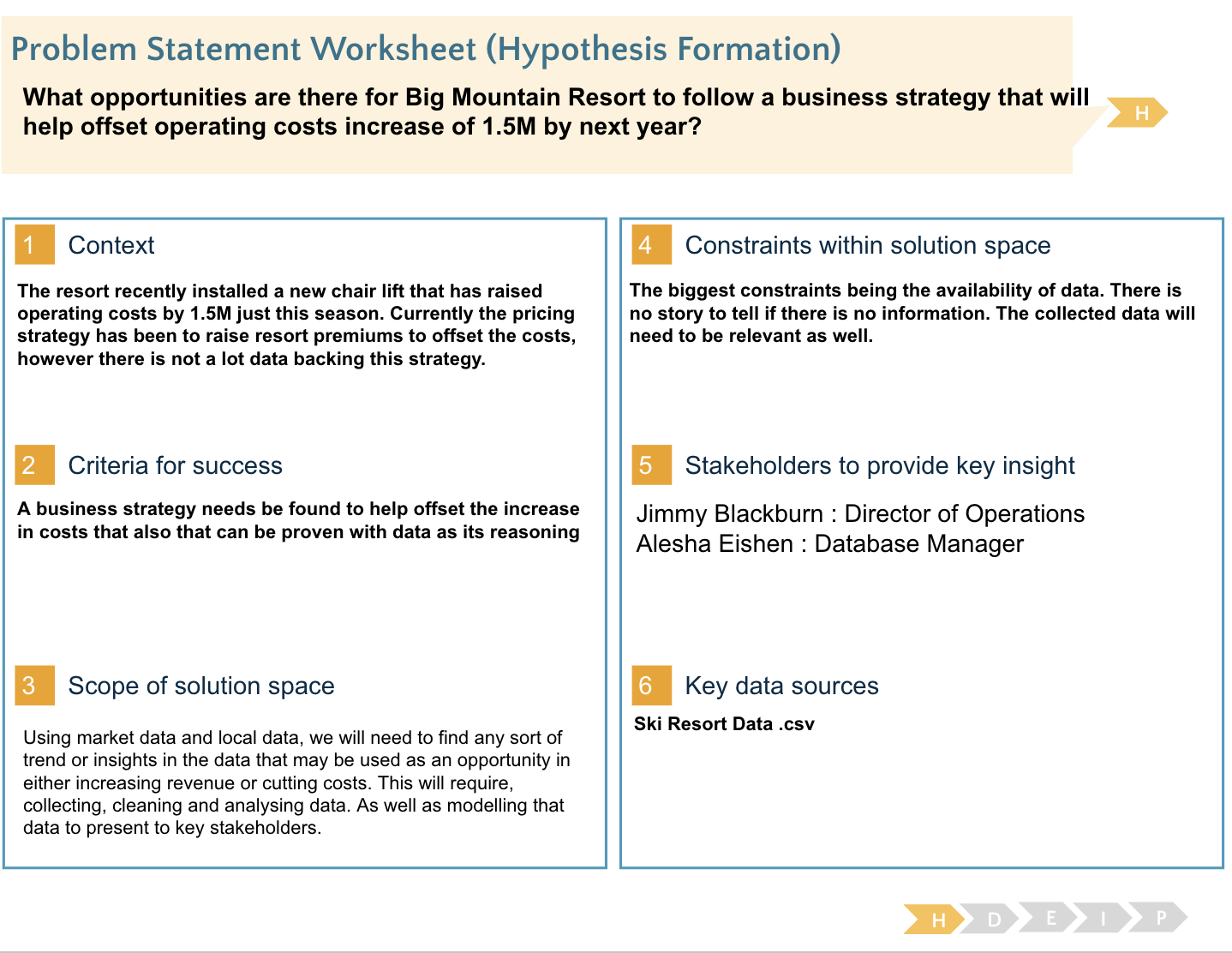
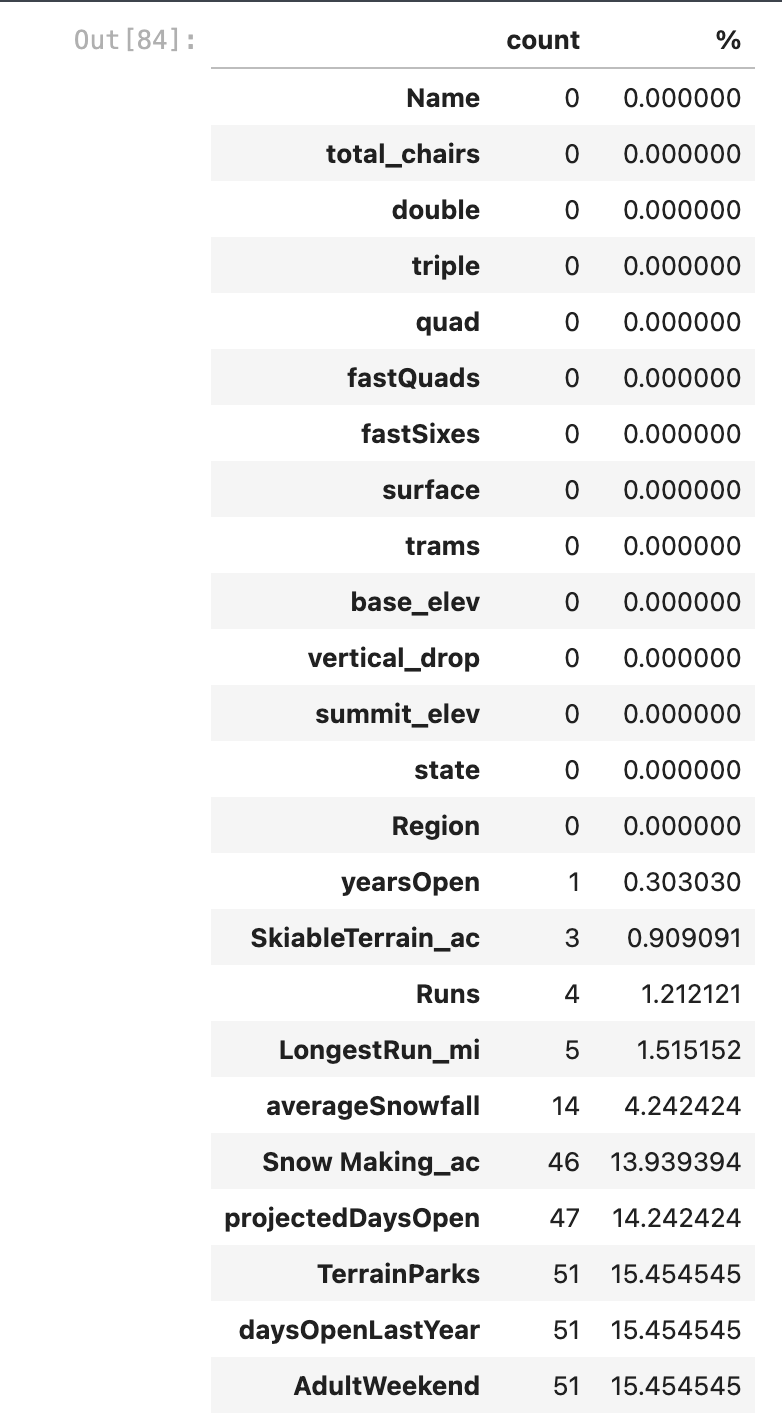
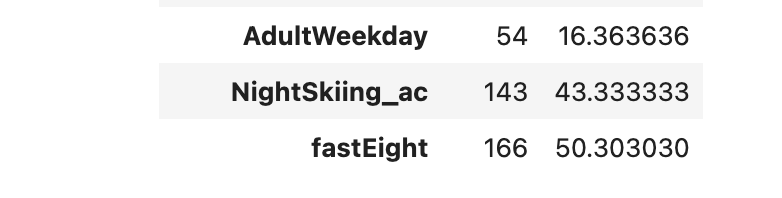
What opportunities are there for Big Mountain Resort to follow a business strategy that will

help offset operating costs increase of 1.5M by next year?

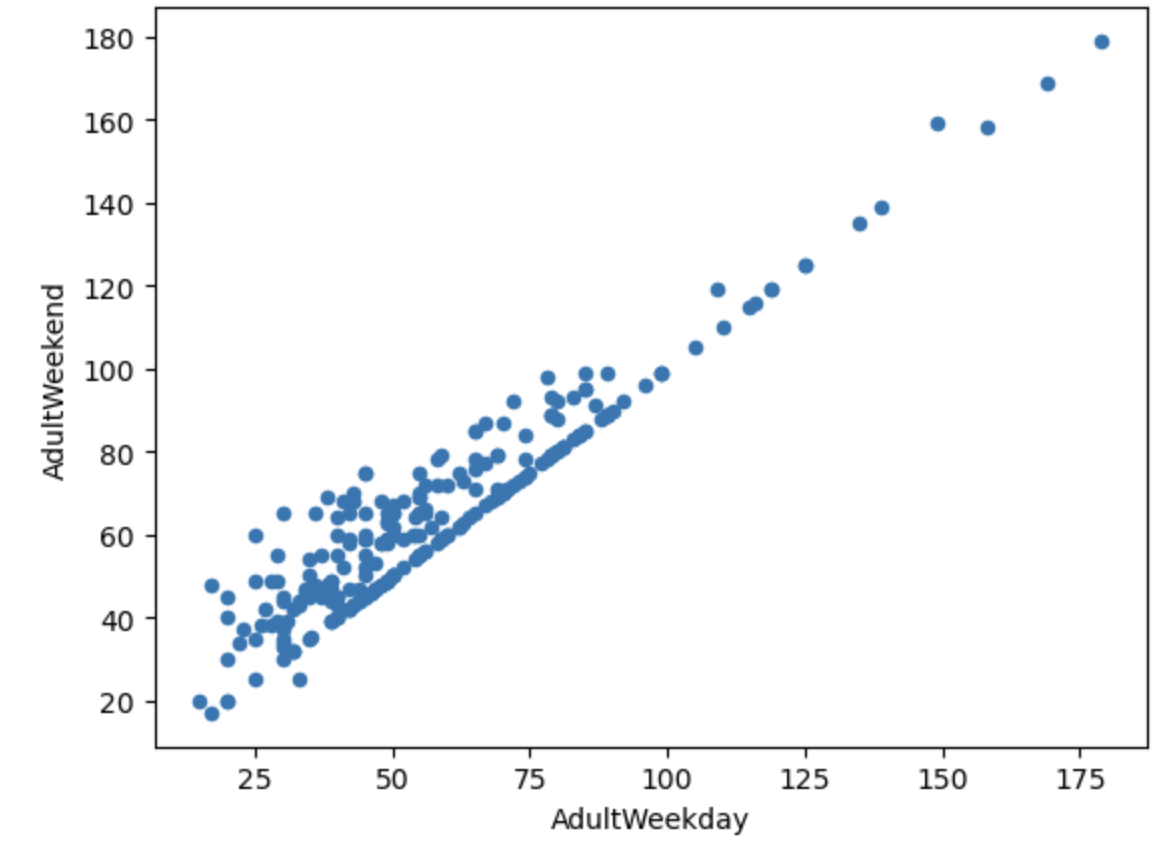


Originally we had 330 rows and 27 columns in a dataframe. We then transposed that data frame so we could then find missing values as well as make it easier to find Big Mountain Resort which we found had no missing values. Through this process we found the rows with missing data in columns and either adjusted the data or dropped it.

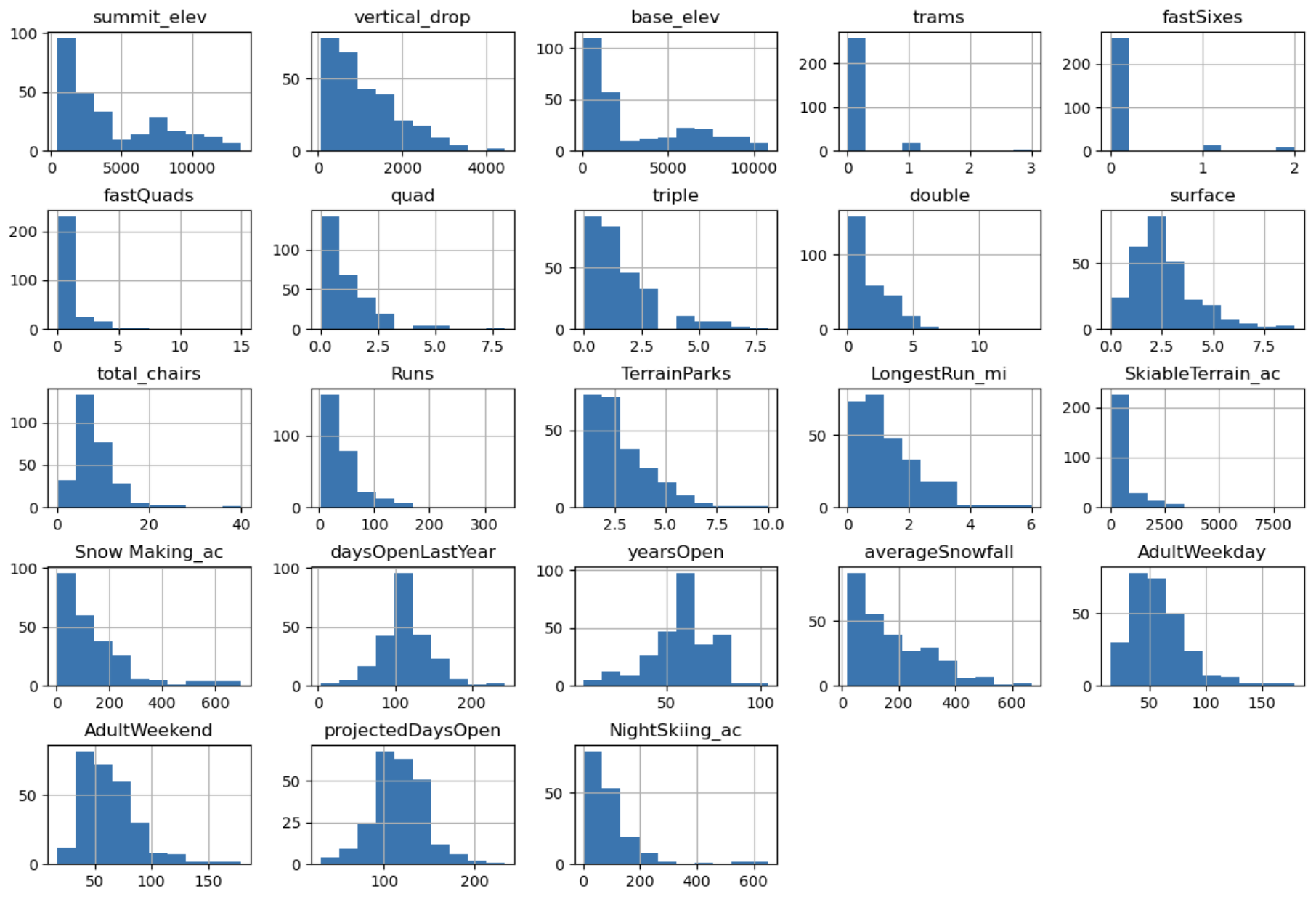




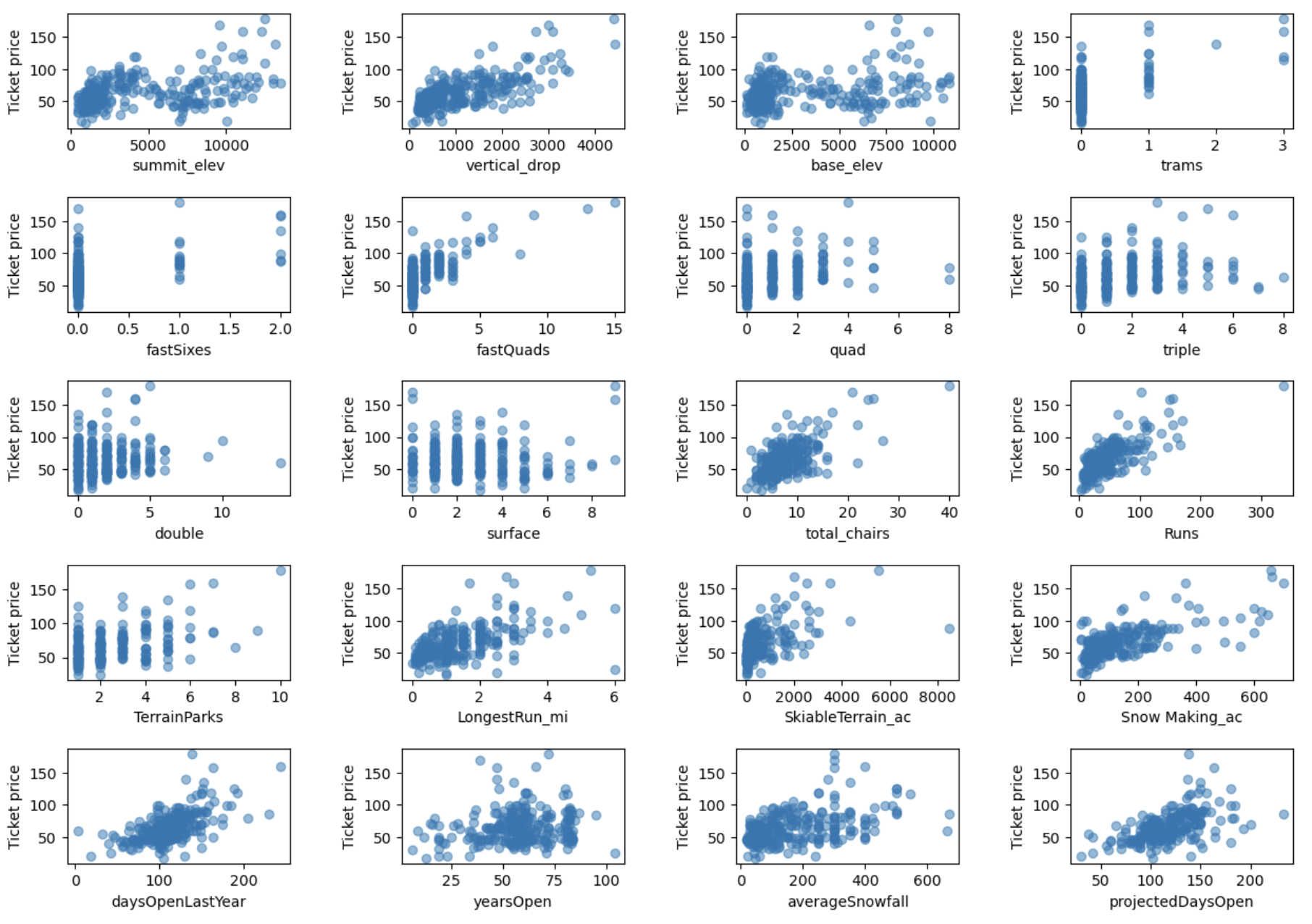
From there we then started to look at what columns we really needed. The data had ticket prices in two different columns, 'AdultWeekend' and 'AdultWeekday'. We plotted both of them to determine they both follow a linear trend line so we can use either and we chose to drop the Weekday price column since it had more missing values, making 'AdultWeekend' price our target data we were searching for.

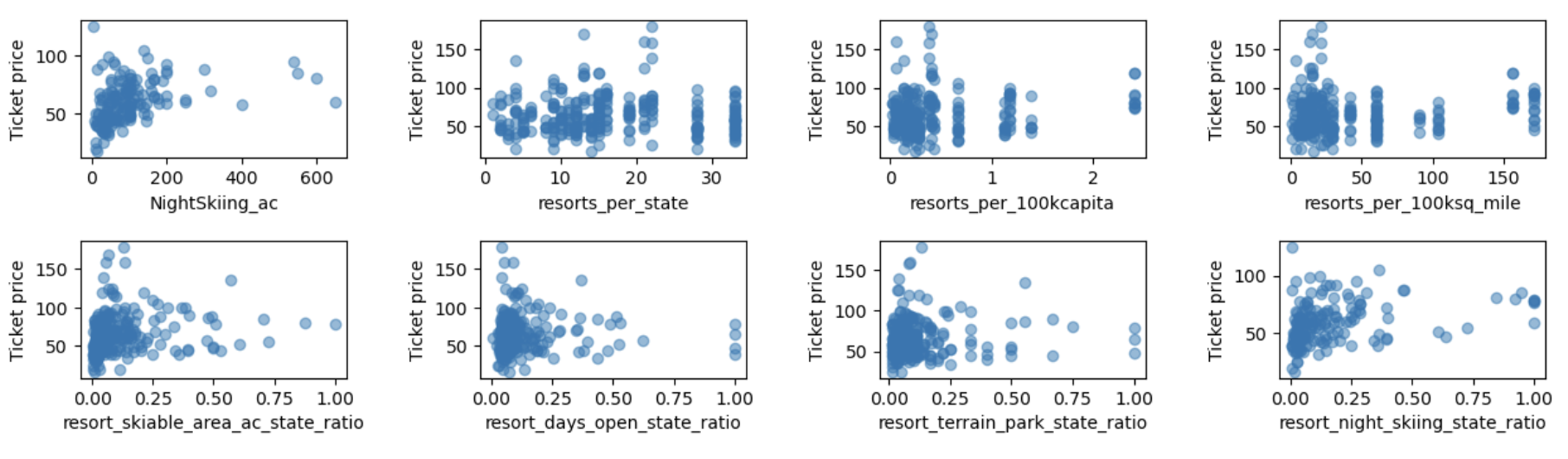


Some rows also got dropped due to too many missing values or due to incorrect or typo values. We found this out by plotting histograms and seeing the distribution of data for each column and found the values skewing the data due to a typo or too many missing values.



At first the data we began with was seven columns and thirty five rows consisting of mainly float value types. We then just changed the index and scaled the values in that new data frame. We plotted a cumulative variance ratio. The mean average ticket price on weekends was also made into a new dataframe and it was plotted to see the distribution of those values. We used quartile functions and PCA values to help perform some more analysis by plotting them using the 2 PC values and the quartiles for hues. We also plotted scatterplots of all the columns in ski\_data. Based on our initial analysis, states with more skiable areas and larger populations seem to be related to higher ticket prices like Colorado and California. However through more in depth modeling we noticed price was correlated with the number of runs and average snowfall. This gives the idea that customers are willing to pay more for tickets at resorts where there is more snow and accessibility to the mountain. Number of visitors per year is something that may be needed for further analysis because some resorts benefit from having more lifts because they serve more people but this may reduce ticket price to an extent. The same is true for resorts with smaller numbers of lifts since they attract less customers they may charge more however this to an extent can reduce demand and therefore ticket price. Night skiing ratio also seems indicative of how much a resort can charge as this seems related to driving up ticket prices. States similar to Montana where our resort is should be taken more into account in future modeling such as ones with similar skiable area and population. That way we can see more closely which features are more important for driving up ticket prices.





In order to begin implementing our model we had to separate our data in two types, train and test data. We did it with a 70/30 split. From there we double checked all the value types of our new train and test sets. At first we tested using the average prices as a baseline however that did not yield strong results at all. We built a linear model initially that we ran multiple tests on, through this process we made modifications to the model. In order to modify the K value to yield better model results we used cross validation. We found that a K value of 8 was best for the model and through that we found key features such as vertical\_drop, Snow Making\_ac, total\_chairs, fastQuads and Runs. These features were what we expected from the previous EDA. We repeated the process with a Random forest regression model however we did not test K values we created a pipeline and immediately used cross validation to find the best parameters. These parameters included using median value for imputations and non using scalers, the models best features were fastQuads, Runs, Snow\_Making\_ac, and Vertical\_drop. After testing both models the Random Forest Regression was the model to be used going forward. We determined this using mean absolute error and the RF model had a slightly smaller value, meaning that it exhibits less variability in determining price for tickets.



The scenarios we were given were as follows:

* Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
* Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage
* Same as number 2, but adding 2 acres of snow making cover
* Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

Currently Big Mountain Resort has a ticket price of

81,our modeled price for Big Mountain Resorts is 95.87. For Senior leadership there lies opportunity for Big Mountain Resort to increase this price even more as our data is based on the facilities of other resorts and their prices. This means if we assume other resorts are pricing tickets correctly we are undervaluing our resort ticket prices. In order increase ticket prices the resort could try scenario 1, close some runs and hedge this decrease in ticket price with scenario 2, build the extra lift and increase the ticket price. If 10 runs are closed the model dictates decreasing ticket price by 1.75 but if 5 runs are closed its only decrease of

0.75. So in theory you would raise the ticket price by $1.25. I’d suggest closing the runs slowly first without immediately building the new lift to see how much it affects the number of visitors per year. Then once that price is stable, build the new lift and increase the prices to match accordingly.

