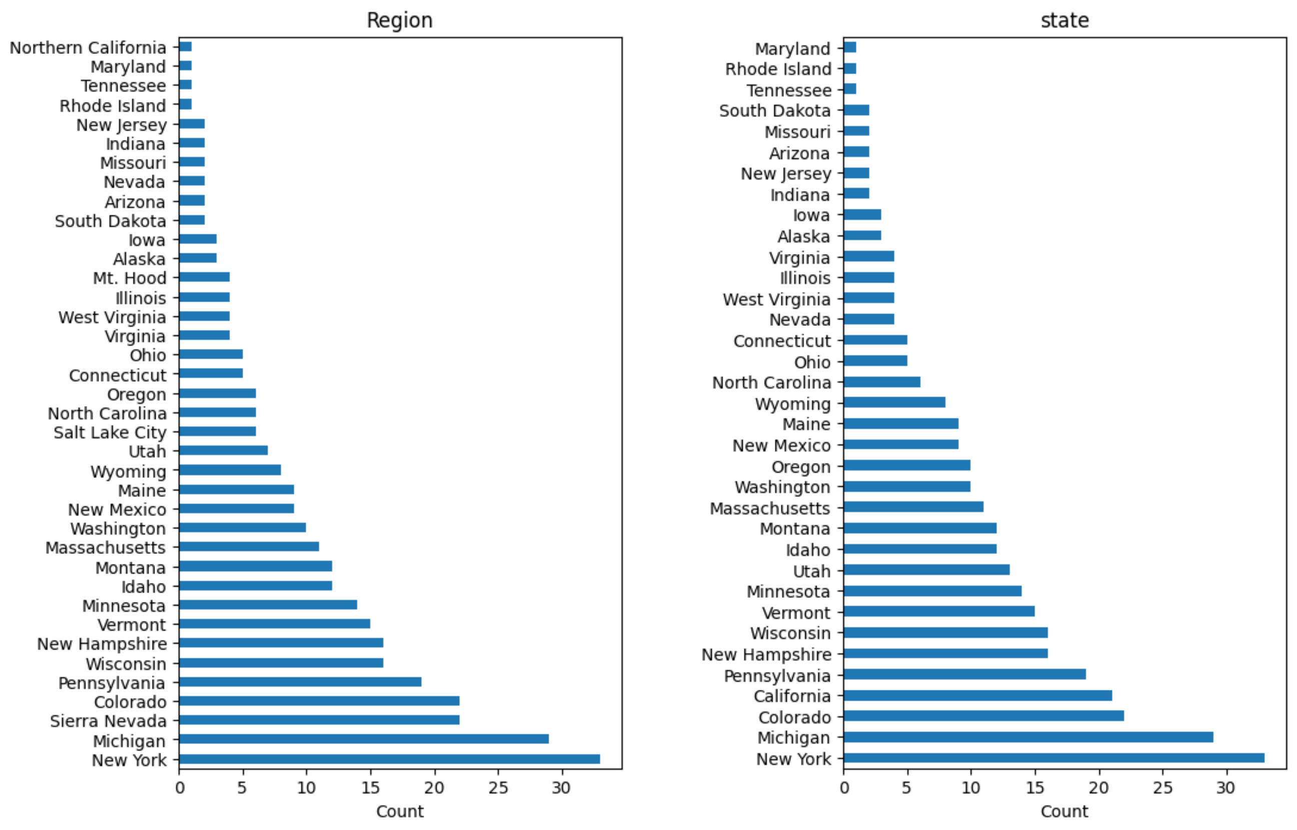
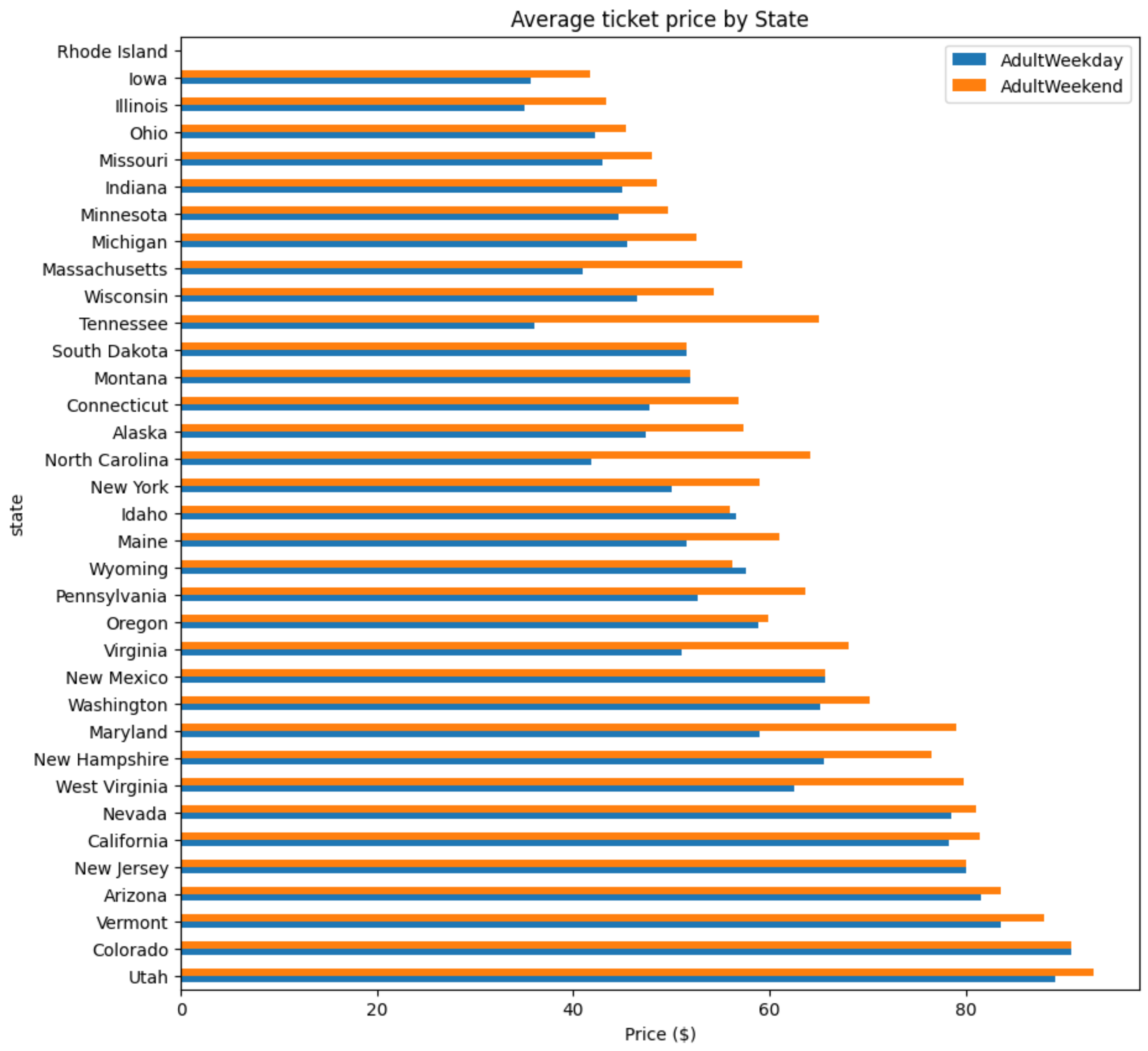
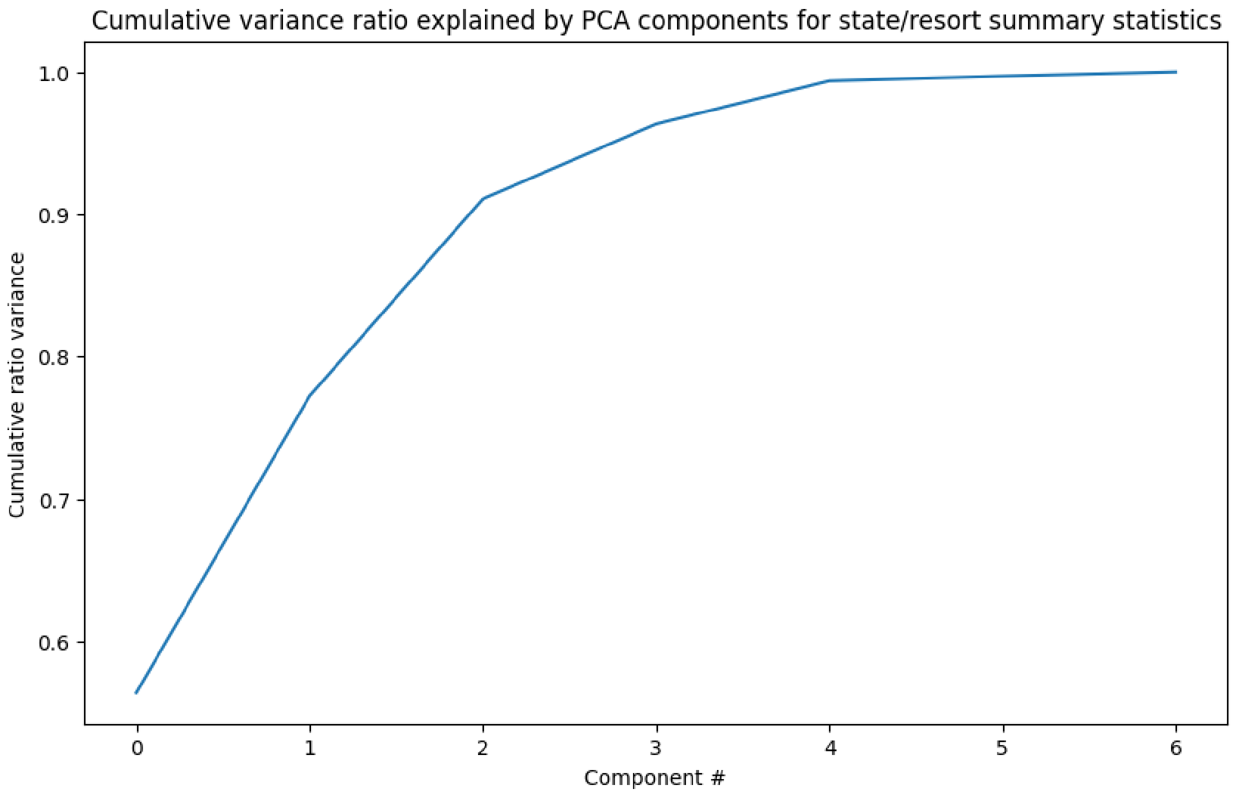
Big Mountain resort recently installed an additional chair lift to increase the distribution of visitors, with the ski resort increasing operating costs by $1.54 million, our data science team was tasked to find and implement a strategy on how to better value our ticket prices and facilities in order to maintain or increase revenue while keeping price margins within 10-20%.

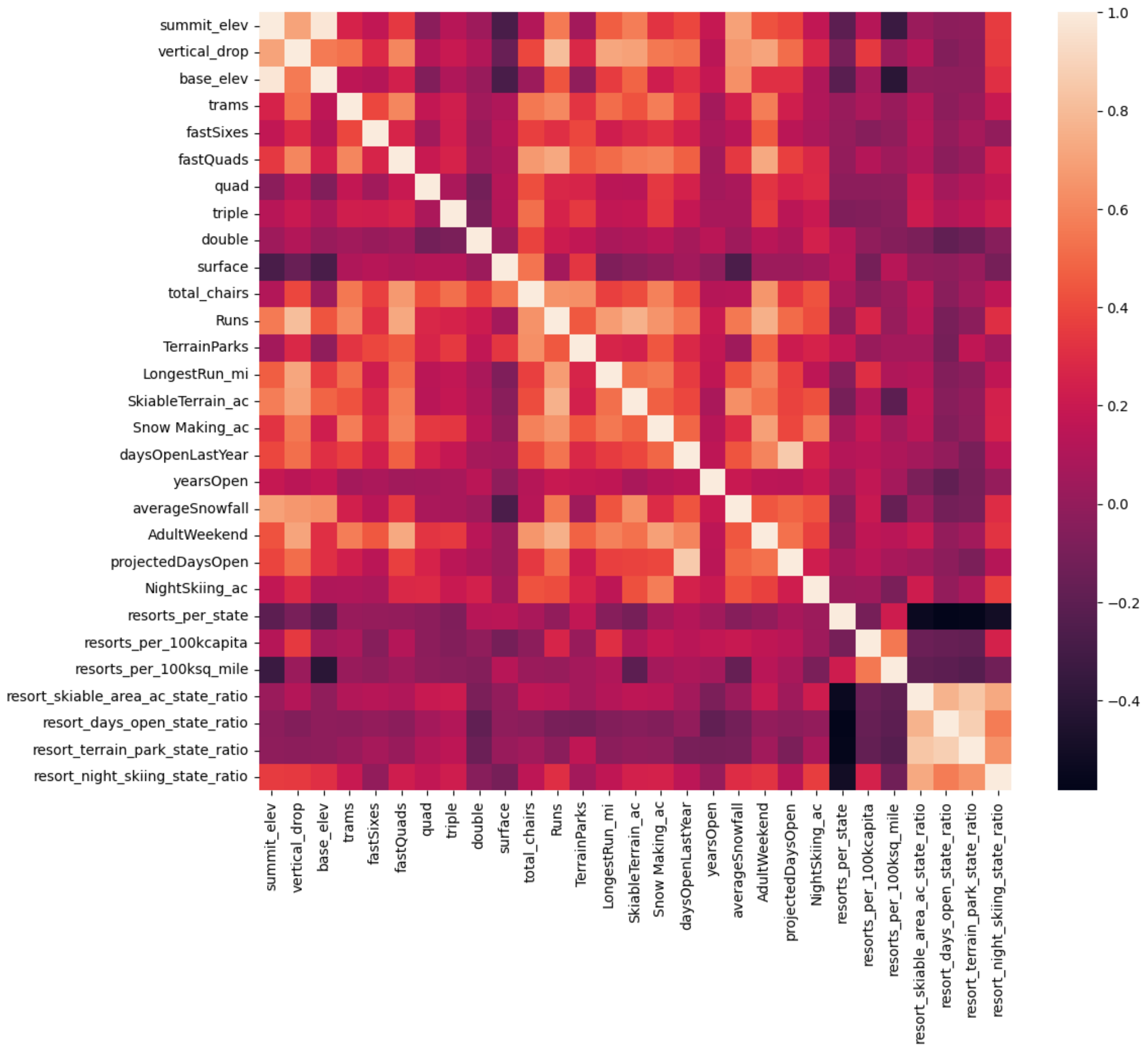
For the data wrangling portion, a dataset of different ski resorts and their prices from across the US was examined. The dataset was initially checked for missing values, and duplicates in resort names were identified and removed. The data was then split by state, allowing for an examination of ticket price distributions within each state. Additionally, the distribution of other feature values was explored. Rows without price data were dropped from the analysis. A comparison of weekend and weekday prices revealed that weekend prices were preferred as target features for modeling purposes. Overall, these steps were taken to thoroughly analyze the dataset and make informed decisions for further modeling and analysis.



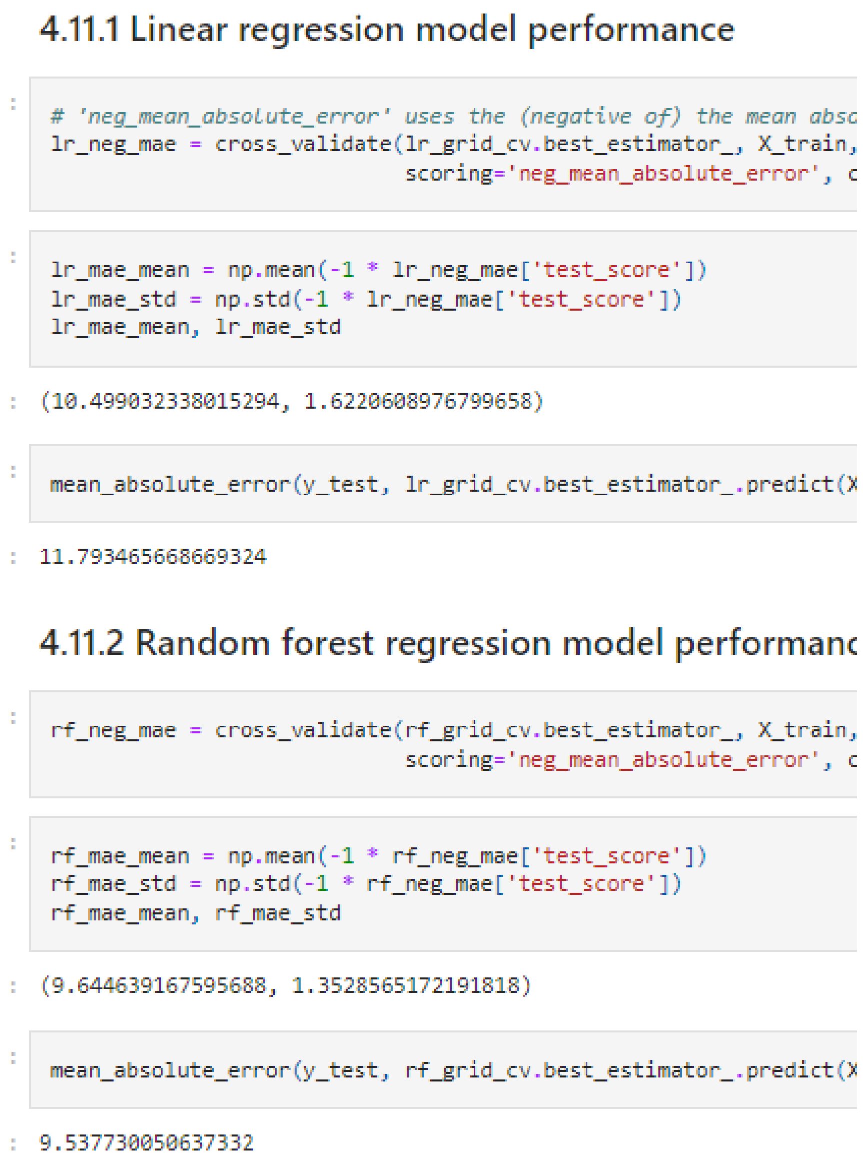


Next, is the EDA section. The data was explored in more detail in this section. The top states by order of each of the summary statistics were examined, including total state area, total state population, resorts per state, total skiable area, total night skiing area, and total days open. Resort density was calculated and the top states by resort density were found. Finally, the data was visualized by scaling the data and calculating the PCA transformation. It was found that the first two transformations accounted for 75% of the variance, and the first four accounted for 95%. Average ticket price was then added to the scatter plot and a conclusion was drawn on how to handle the state label. The ski resort numeric data was also examined, including feature engineering, feature correlation heatmap, and scatterplots of numeric features against ticket price. The heatmap showed a lot of important information: summit and base elevation are highly correlated, our target feature of the weekend prices are correlated to runs, guaranteed snow, and fast quads, and more variable terrain area is not as highly correlated with the ticket prices.





In the pre-processing and training portion, the data was split into train and test sets, and various preprocessing techniques were applied to ensure the data was suitable for training the model. Missing values were imputed using the median and mean, and the data was scaled to a standard range. Pipelines were utilized to streamline the preprocessing and training process, combining the steps into a single workflow. The linear model was refined by selecting an optimal set of features using GridSearchCV, and a Random Forest model was also evaluated using cross-validation and GridSearchCV. After extracting and training the data, you find the model of the training data to find the mean and see how well the mean matches actual values, which is the simplest way to predict a value, but does not take into account trends and variance. Then we use the R2 value to find the accuracy of this prediction through the proportion of the variance. The mean of 63.81 and the R2 of -0.003 means that the initial model is a good predictor. However, the MAE being off around $19 means that there are better predictors. With a linear regression model, we first need to fit, train, test, and asses (predict) using the model. The MAE was off by $9, which makes it a better predictor. Next is to cross validate with RF regressor, this is a better model as it tends to overfit easily and is more complex. This was apparent with the fact that it had the lowest absolute error being off of $1.\



4.11.1 Linear regression model performance

```
# 'neg_mean_absolute_error' uses the (negative of) the mean absc
lr_neg_mae = cross_validate(lr_grid_cv.best_estimator_, X_train,
                            scoring='neg_mean_absolute_error', c
```

```
lr_mae_mean = np.mean(-1 * lr_neg_mae['test_score'])
lr_mae_std = np.std(-1 * lr_neg_mae['test_score'])
lr_mae_mean, lr_mae_std
```

```
(10.499032338015294, 1.6220608976799658)
```

```
mean_absolute_error(y_test, lr_grid_cv.best_estimator_.predict(X
```

```
11.793465668669324
```

4.11.2 Random forest regression model performanc

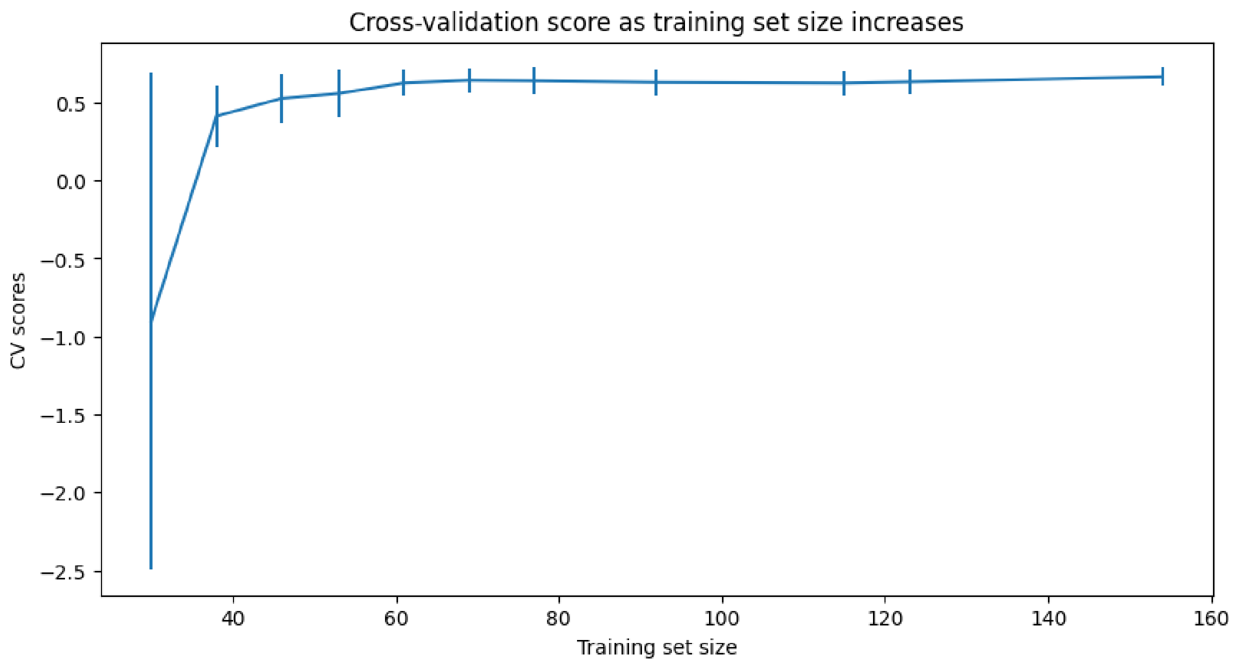
```
rf_neg_mae = cross_validate(rf_grid_cv.best_estimator_, X_train,
                            scoring='neg_mean_absolute_error', c
```

```
rf_mae_mean = np.mean(-1 * rf_neg_mae['test_score'])
rf_mae_std = np.std(-1 * rf_neg_mae['test_score'])
rf_mae_mean, rf_mae_std
```

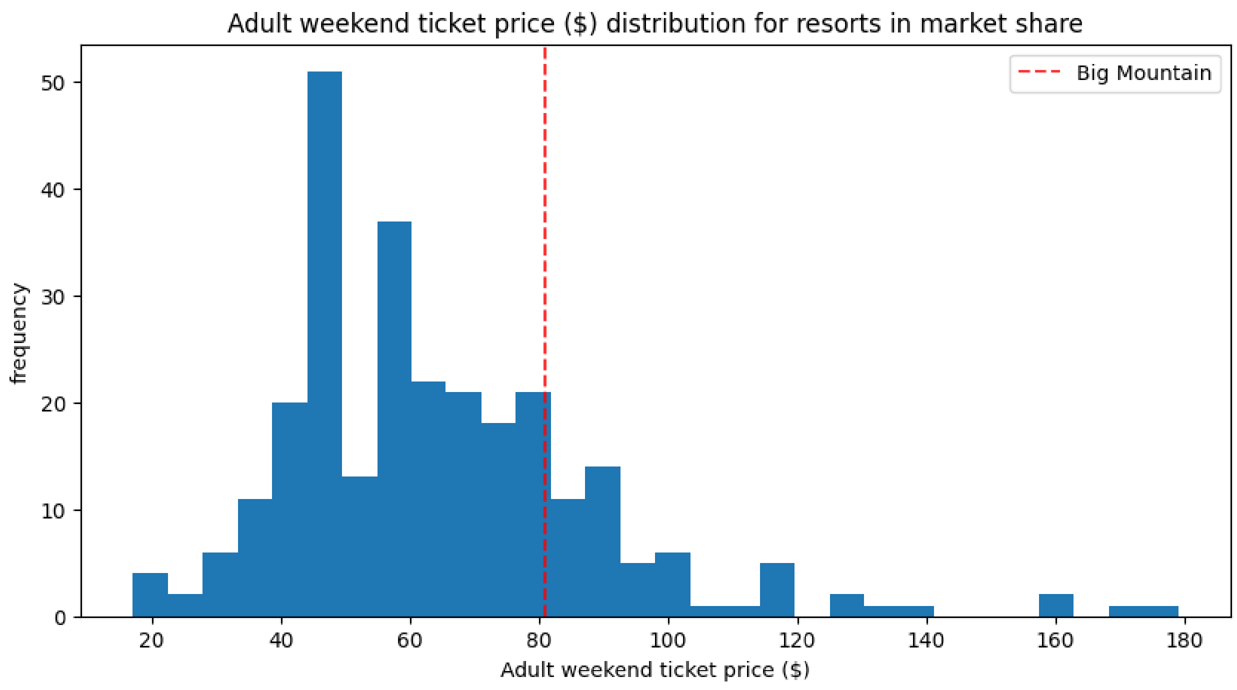
```
(9.644639167595688, 1.3528565172191818)
```

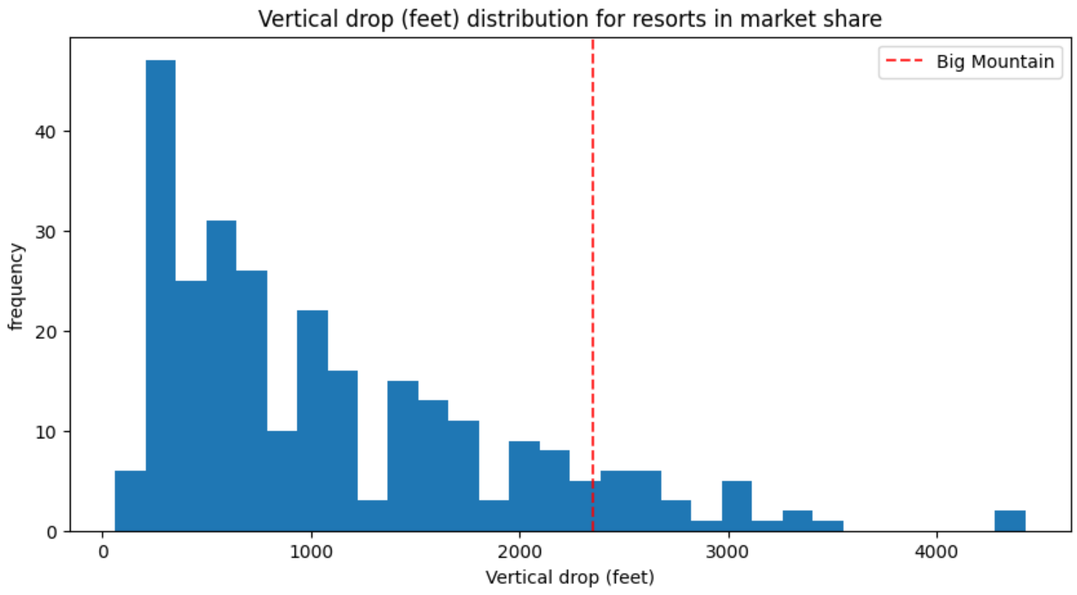
```
mean_absolute_error(y_test, rf_grid_cv.best_estimator_.predict(X
```

```
9.537730050637332
```



Big mountain charges $81 for tickets, but modeling suggests that the tickets should be 95.87. I would explain to leadership that this model is based on other ski places, so our competition is clearly charging more by a lot, so we should increase prices slightly, to make more money but still be cheaper than other competitors.The chair lift and increasing the vertical drop by 150 ft can happen if you increase the price of tickets by 1.99. To determine future scope of work it's important to look at what the data can do. Deficiency of the data is that we only have ticket prices for the weekends instead of weekdays. There are also other reasons for tickets to be differently priced, like discounts or special pricing. Additional operating costs for the chair lifts would be day-to-day costs, construction costs, and maintenance costs. Big Mountain ranks highly on many league charts possibly because the facilities did not have much effect on the price. Instead we would need to model the prices on the features that influence the ticket pricing.These businesses would make use of it by looking at which one of the features effect ticket prices and looking how to adjust various things to make the ticket prices reflect a more efficient business for practice.


Adult weekend ticket price ($) distribution for resorts in market share
Big Mountain
frequency
Adult weekend ticket price ($)
0
10
20
30
40
50
20
40
60
80
100
120
140
160
180


Vertical drop (feet) distribution for resorts in market share
Big Mountain
frequency
Vertical drop (feet)
0
10
20
30
40
0
1000
2000
3000
4000