Rachel Chiang

Big Mountain Resort is a ski resort in Montana that entertains and accommodates 350,000 skiers and snowboarders every year. Recently, the resort installed a new chairlift, which increased operating costs by $1,540,000. The resort wants to know how to select a better price for their ticket and is considering changing their features to reduce operating costs or support a higher ticket price. We would like to investigate how Big Mountain Resort can capitalize on its sport-accommodating facilities to increase its ticket price to at least cover the $1.54M increase in operating costs within the next season. Specifically, we will examine the relationship between resort features and ticket prices.

Ski resort data was received from the database manager, and to begin with, the data was assessed and adjusted for quality. Roughly half of the features were reported with missing values, and there were a few data points that were suspicious, but only Silverton Mountain resort's skiable terrain was corrected after finding the right value advertised on its website. We dropped two resorts, Heavenly Mountain Resort for its lack of price data and Pine Knob Ski Resort for its implausible years open value. A feature that we chose to ignore was the number of fast eights at each resort because only one resort reported having one. At this stage, we wanted to consider whether the state was important to the price, so a few statewide statistics and generic information were collected and cleaned. Our target feature was the ticket price, but the data we received included weekend and weekday adult prices, and inputs for the two were inconsistent. In the end, we decided to proceed with the weekend adult ticket prices as our target feature.

We examined the state-wide statistics for the resorts and calculated ratios for specific features and for resort density for each state because the absolute size or population of a state may not be as relevant. Simply looking at the distributions and rankings, or patterns within these, would not help understand the relationships between the features alone, so we used principal components analysis (PCA) to try to disentangle the relationships between the features. We generated a scatterplot for the first two PCA components to see if there is a pattern with the prices and states, but there did not seem to be an obvious pattern. We plotted a correlation heatmap for each of the features (see Appendix, Graph 1). Looking specifically at the weekend adult ticket price, the vertical drop, number of fast quads, total number of chairs, number of runs, and snowmaking cover seemed to have high correlation. To focus on the relationships between ticket price with the other variables, we plotted scatterplots for each (see Appendix, Graph 2). There seemed to indeed be good positive correlations with vertical drop, number of fast quads, total number of chairs, and runs, as well as some for snowmaking cover.

We used two models and compared their performances: a linear model and a random forest model. We partitioned our data with a 70/30 train/test split. For the linear model, we imputed missing values with median instead of mean, scaled the data with a `StandardScaler`, selected k best features to subset on using `f_regression`, and used linear regression. We assessed performance and searched for the best parameters with cross-validation and `GridSearchCV`. In the end, we used $k = 8$, and the model found the following coefficients as most important: vertical drop, snowmaking area, total number of chairlifts, and number of fast quads. For the random forest model, we used 69 trees, imputed missing values with the median, did not scale

the data, and used random forest regression. The model found that the most important features were the number of fast quads, the number of runs, snowmaking coverage, and vertical drop.
To score the two models, we used the mean absolute errors (MAE), where $MAE = \frac{1}{n}\sum_i^n |y_i - \hat{y}|$.

	Linear Model	Random Forest Model
Mean of MAE	10.499032338015294	9.644639167595688
Standard deviation of MAE	1.6220608976799664	1.3528565172191818
MAE	11.793465668669324	9.537730050637332

The random forest model's performance seemed to be better, as it is $1 closer when estimating the actual value of a ticket price, and it exhibits less variability. Thus, the random forest model was used for further modeling.

The random forest model suggested that Big Mountain Resort should price their tickets at $95.87. The resort tends to place highly in almost all features, including in those that this model considers as most important (fast quads, runs, snowmaking cover, and vertical drop). Note that to cover the $1.54M increase in operating costs, the resort must increase ticket prices by at least $0.88 each.
Big Mountain Resort proposed four scenarios to cut costs or to increase revenue:

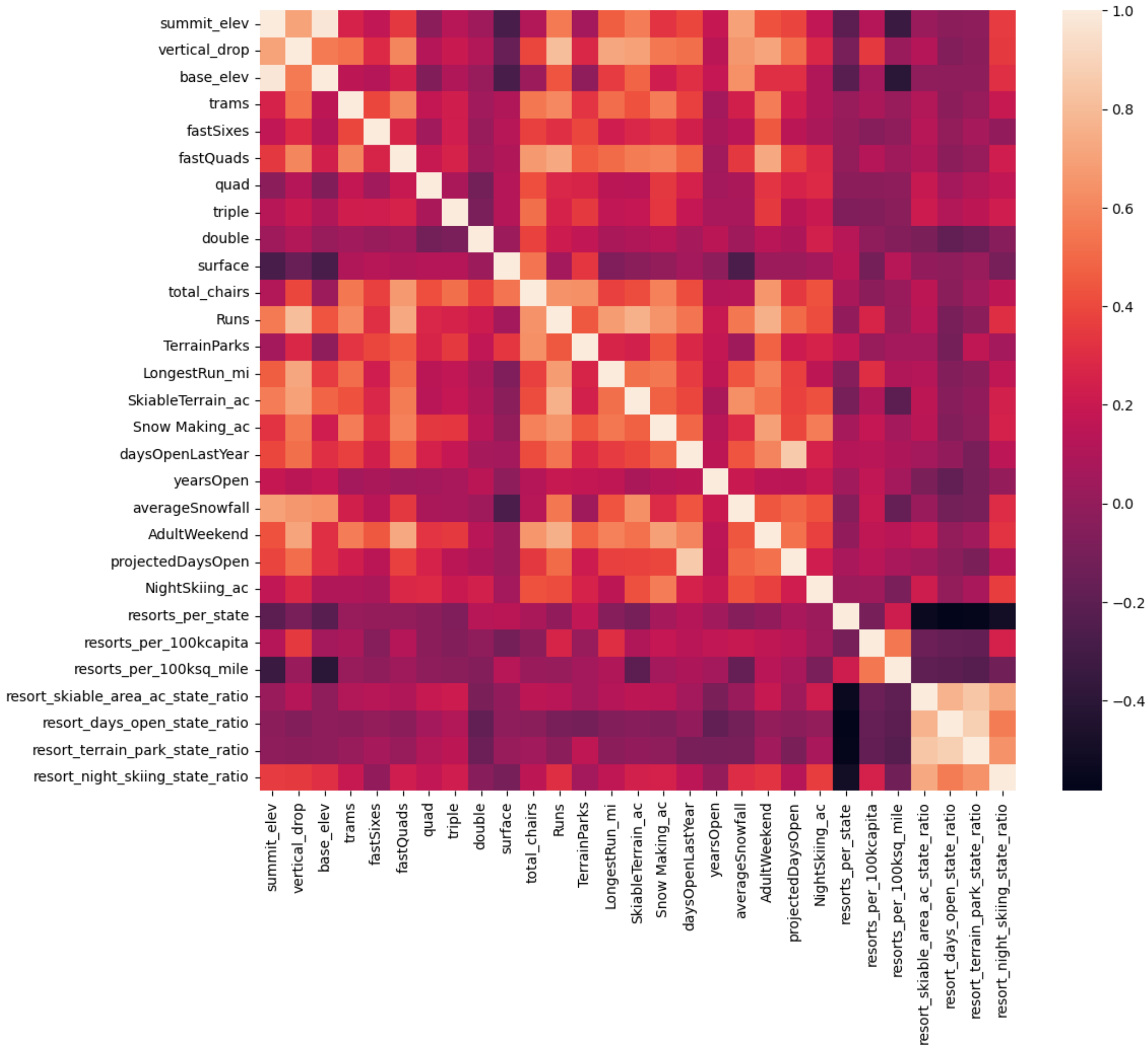
1. Permanently close 1-10 of the least used runs.
2. Increase the vertical drop by adding a run to a point 150 feet further down and install another chairlift only.
3. Do option 2 but also add 2 acres of snowmaking cover.
4. Increase the longest run by 0.2 miles and add 4 acres of snowmaking cover.

We modeled the four options using the random forest model and would recommend further consideration to be given to options 1 and 2. With option 1, permanently closing just one of the runs may decrease operating price without negatively affecting the ticket price, so it would likely be beneficial overall. If additional runs are closed, the resort can also consider closing them in groups (see Appendix, Graph 3). With option 2, the model predicted an increase in ticket price of $1.99, and given the 350,000 expected visitors in the next season, this could provide a total of $3,474,638 increase in revenue. However, it is important to think about how much these changes cost.

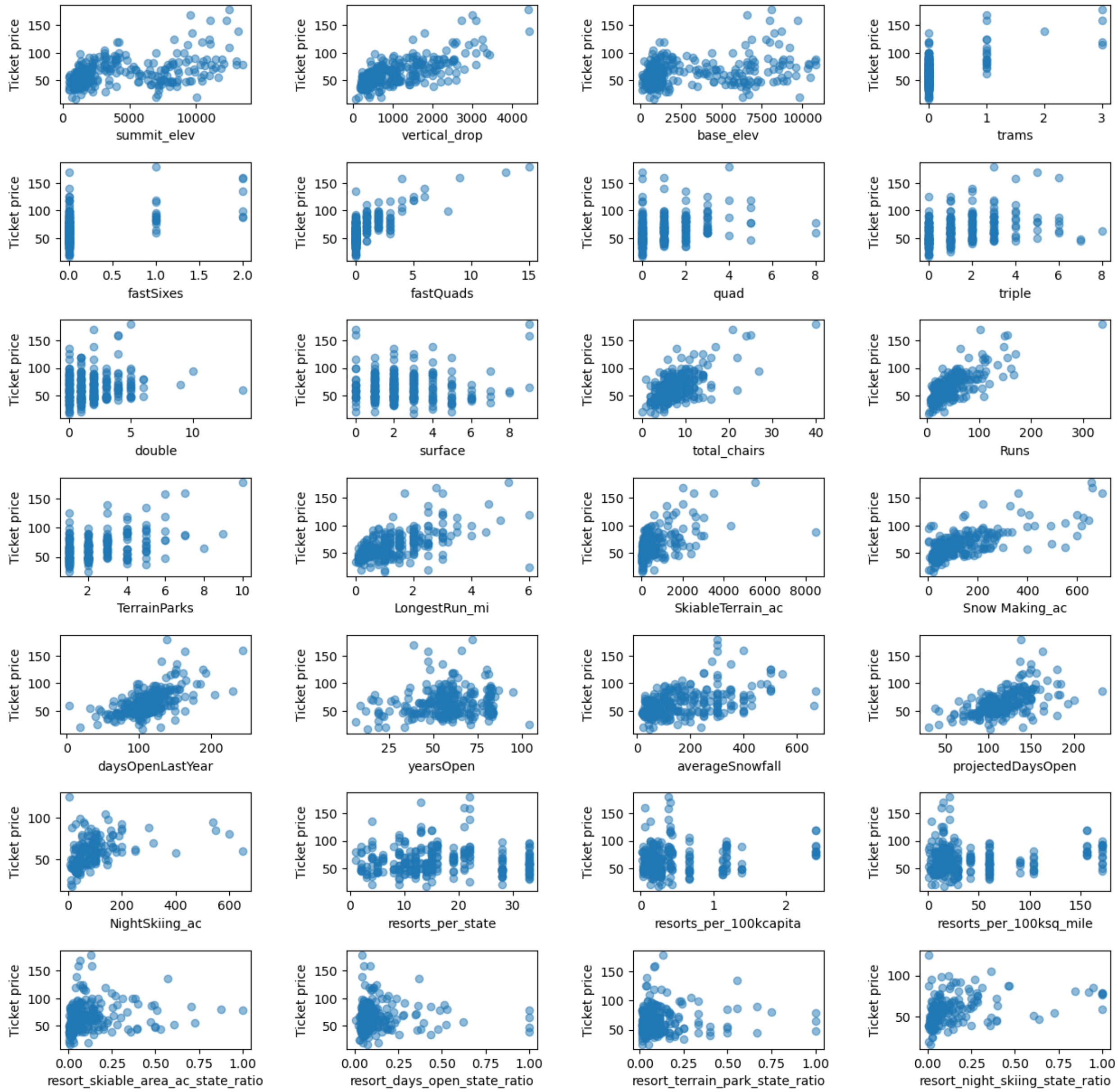
Assuming that the resorts in our data were not overpricing or underpricing their tickets and the resort is in a free market, Big Mountain Resort could consider simply adjusting the ticket price to match its already-existing features to meet the increase in operating costs. Otherwise, or additionally, the resort could consider closing one of the least used runs, or adding a new run and chairlift to increase the vertical drop for potentially more revenue.

Some options for future work may be further adjustments to the parameters of the model or adding more data that may be useful to know. Throughout the project, a few data points that may have been useful in understanding the relationships of features would include the number of visitors per year, how many tickets were sold, or general operating costs.

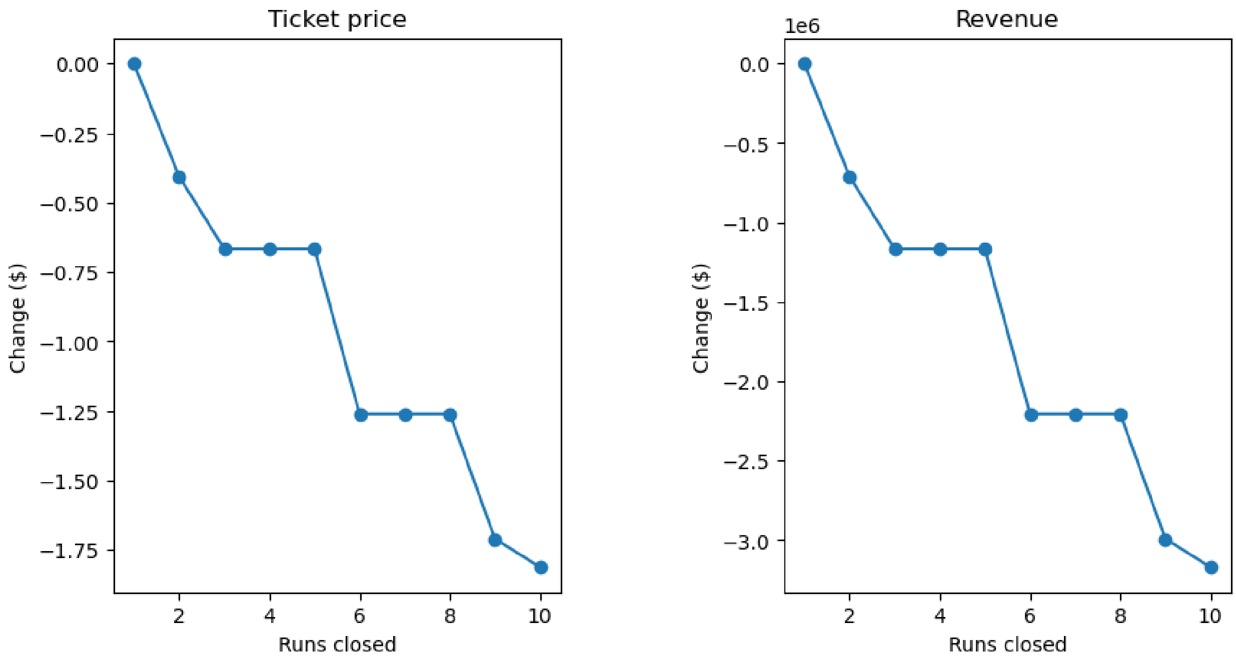
Appendix



Graph 1. Heat map of features of ski data.



Graph 2. Scatterplots of features versus ticket price.



Graph 3. Predicted change in ticket price given scenario 1 using the random forest model.