Business Analysis for Big Mountain Resort

Soyoung An

Introduction

A ski resort located in Montana, Big Mountain, offers spectacular views of Glacier National Park and Flathead National Forest. About 350,000 people come to ski or snowboard every year. The ski resort has access to 105 trails, 3.3 miles longest run, 4,464 ft base elevation, and 6,817 ft on the summit. It provides 11 lifts, 2 T-bars and 1 magic carpet for novice skiers. Recently, Big Mountain resort invested an additional chair lift, which increases its operation cost by **$1.54 M** this season. The resort plans to charge a premium above the average price of resorts in its market segment. However, they are unsure if this will be beneficial for the business. Besides that, the resort may not capitalize on its facilities as much as it could, so it changes in business strategy to improve its profit margin. Using a collected dataset from other ski resorts in the United States market share, it is aimed to identify the optimal ticket prices and opportunities to increase an additional $1.7 M revenue next year.

Data Wrangling

The dataset has many resorts in different state and regions of the United States, including Big Mountain Resort in Montana. The data provides the facilities' information and services with their ticket prices: adult weekday and adult weekend prices. Big Mountain resort doesn't appear to have any missing values. However, the other resorts have some missing features. Columns with a large amount of missing values and incorrect data were dropped. Resorts with more than 1000 years open period and missing price information was dropped for analysis. Both ticket prices are similar to each other and same in Montana, so the AdultWeekend feature is kept because the weekend prices have the least missing value from the two prices. Besides, suspicious data such as 26819 acres in skiable terrain in Silverton Mountain Resort had been replaced with an accurate value (1819 acres) which is found from the trustable source. There are 277 rows and 25 columns left from 330 rows and 27 columns initially. Also, the U.S. population and area information by state was scrapped from Wikipedia website. This scrapped state information is combined with ski resort data and modified the resorts information by state. The state summary dataset consists of 35 rows and 8 columns (features).

Exploratory Data Analysis

The categorical features are Name, Region, and state. The rest are numerical features: summit_elev, vertical_drop, base_elev, total_chairs, runs, and more. Grouping numerical characteristics for each state corresponding to its state population and space distance is the way to discover and detect any new patterns. After analyzing all of these traits and noticing different trends, principal component analysis (PCA) is used to reduce the dataset's dimensionality into smaller sets that can be easily visualize and analyzed. Once the six numeric features are scaled and fit the PCA transformation, the first two components (resorts_per_state, state_total_skiable_area_ac) account for over 75% while the first four accounts for over 95%. I applied the average ticket price to the scatter plot from the first two PCA elements, but the

pattern was not clear. I created a heatmap and scatter plots to see the relationship and correlation amongst the features.
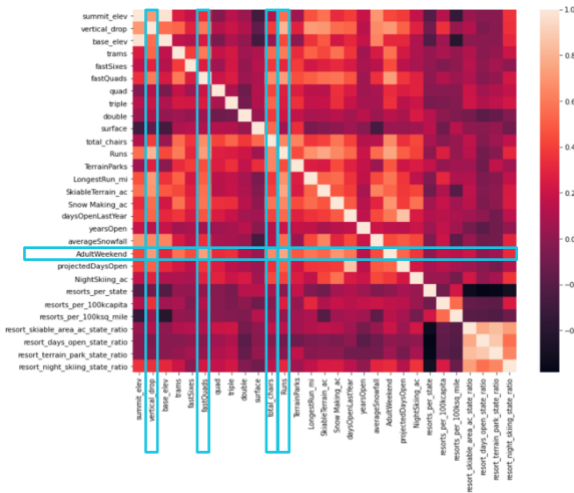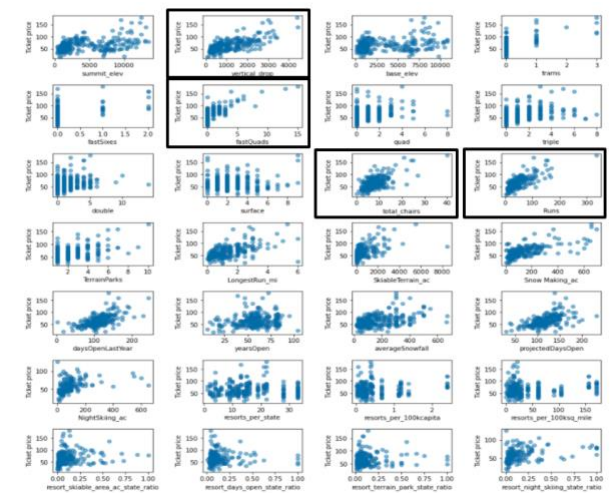

Figure 1: Heatmap   Figure 2: Scatter plot

Figure 1 and 2: Both showing the relationship between resorts ticket price vs. other features

Based on the heatmap and scatterplot from Figure 1 and 2, there are positive correlations between the ticket price with vertical_drop, fastQuads, total_chairs, and runs for each state.

Preprocessing and Training

In the preprocessing process, a baseline model is an essential solution to see how good the mean value is, which is also a metric compared to other models. I split the dataset into a 70 train/30 test set that only contains numerical data. I use three methods: R-squared (R2), mean absolute error (MAE), and mean squared error (MSE). On the test set, the R2 is -0.00312, which is also lower than on the training set. The MAE on the test set is worse than the training set, but the MSE did better on the test set. Therefore, I check the two models: the linear regression model and the random forest mode.

For the initial modes, I impute missing features with the median for one model and with the mean for the other model on both the train and test sets. Then, I scale the data to have a consistent measurement. I train the linear regression model on the training set and make a prediction. The R-squared performance is over 80% on both models on the train set and over 70% on the test set models. The MAE and MSE train test scores are similar, whether using median or mean to replace missing data. The model may be overfitting. Therefore, I use a cross-validation pipeline to find the best k, which divides the training set into k folds and trains the model on k-1 folds. The best k is 5. Vertical_drop has the most impact, and other essential features are Snow_Making_ac, total_chairs, fastQuads, and Runs. With a random forest model, I impute the missing value with mean and median without feature scaling and split the train set into 5-fold cross-validations. On this model, the four dominant features are fastQuads, Runs, Snow_Making_ac, and vertical_drop. The mean absolute error using cross-validation is 11.79 and 9.537 for linear regression and random forest, respectively. With a smaller error to almost $1 different, the random forest model performs better.

Modeling

The model suggests $95.87 for a Big Mountain ticket price, with mean absolute error $10.39. This is higher than its current price of $81.00 because the model assumes that other resorts set their prices on a market-share basis. Big Mountain can either cut costs or increase revenue and evaluate the estimated result for each scenario. The first scenario is close down to up 10 of the least used runs. Statistically, closing one run makes no difference, but closing from 2-3 runs reduces ticket prices and decrease revenue. Closing 4 or 5 runs have the same result as closing 3 runs, but closing 6 runs or more lead to a large revenue drop. The second scenario is increasing the vertical drop by 150 feet and installing an additional chair lift, while the third scenario is adding 2 acres of snow making area in addition to the second scenario. In both scenarios, the business can increase **$2** towards the ticket price and this will lead to the revenue increase up to **$3,474,638.00** over the season. The last scenariois increasing the longest run by 0.3 miles and adding 4 acres of snowmaking capacity, which shows no revenue difference.