# What changes can Big Mountain pursue to increase revenue and successfully offset the perceived $1.54 million rise in operating costs for this season?

Our objective is to build a model that will help Big Mountain to initiate changes that would support a higher ticket value or cut operational cost to compensate for the $1.54M additional operational cost.

After initial assessment of the dataset quite a few columns with missing values were discovered. Column with mostly missing values were dropped. Some suspicious values were taken care of by doing some research on the internet.
A state-wide summary was created using groupby behavior named aggregation. Also, population data and area data are collected from Wikipedia and merged with the state-wide dataset.

For our target feature when building our model will be the features that relates with ticket prices namely AdultWeekday and AdultWeekend.

The scatter plot of AdultWeekday vs. AdultWeekend shows Weekend prices being higher than weekday prices seem restricted to sub $100 resorts as shown by Fig: 1.

The two columns AdultWeekday and AdultWeekend are same for our resort and all the other resorts in Montana state. As AdultWeekend price has least missing values, AdultWeekday was dropped. We chose AdultWeekend to be our target feature. Therefore, rows with both AdultWeekday and AdultWeekend missing are dropped.



Fig: 1. Scatter plot of AdultWeekend vs.

Analyzing the state-wide summary combined with average ticket price does not reveal any useful insight to treat any state differently. Principal Component Analysis was used to achieve this. We can observe a 77.2% variance based on PC1 and PC2 that is shown in Fig: 2, no clear pattern is visible here. This supports building a generalized model that doesn't weigh state labels disproportionately but allows room for potential state-derived features.



Fig: 2. Observed variance of 77.2% between PC1 and PC2
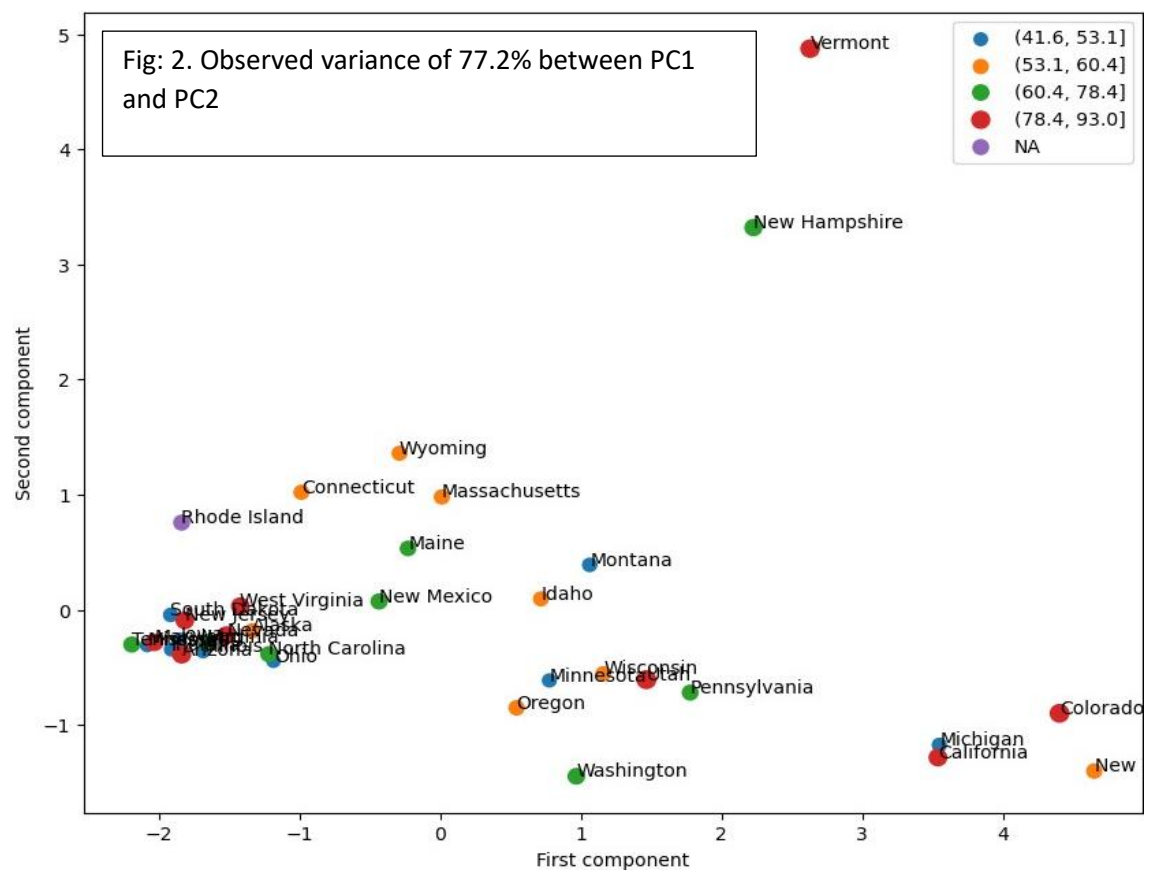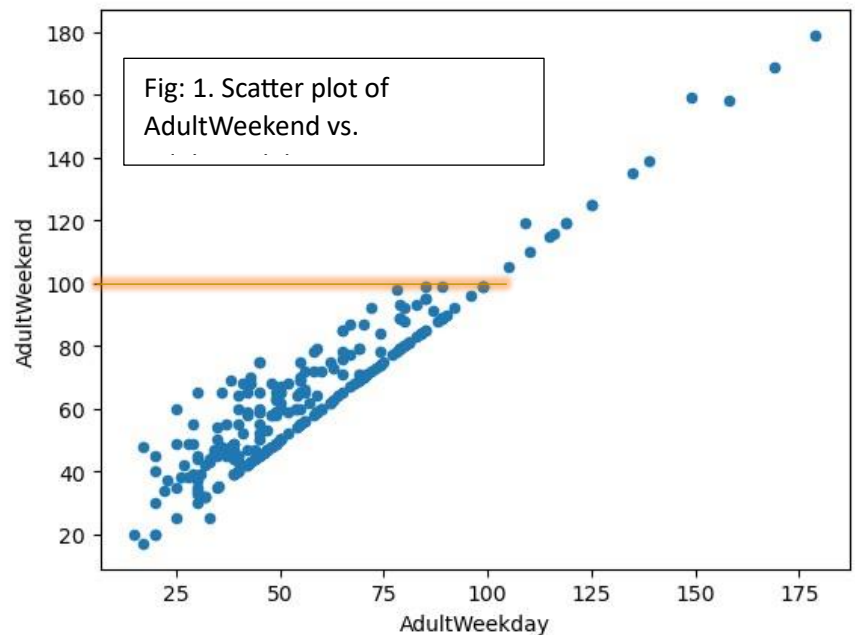
By using correlation heatmap useful insights were drawn out form the numerical features.

If we focus on the correlation (In Fig: 3)between our target feature (AdultWeekend) and other features, we could see that Snow Making_ac, Runs, total_chairs, fastQuads and vertical_drop is pretty good positively correlated and could be our potential predictor features. There were also some positive correlations with DaysOpenLastYear,

LongestRun_mi, TerrainParks, trams which also could be considered when building ticket pricing model.
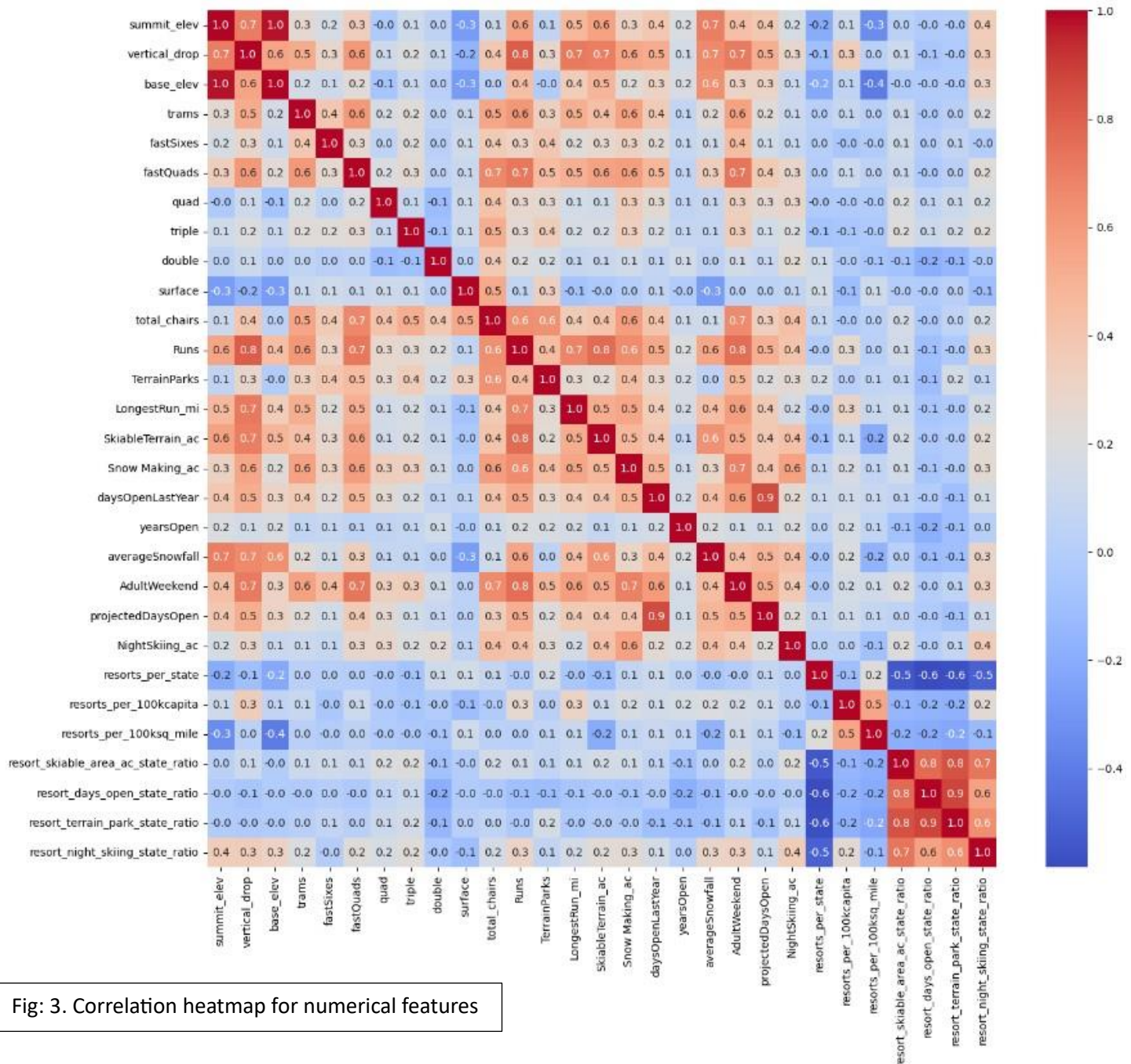


Fig: 3. Correlation heatmap for numerical features

Linear Regression and Random Forest Regression, were built using scikit-learn. Both models were implemented within a pipeline to streamline preprocessing and training. SimpleImputer was used to handle missing values, filling in missing data with the mean or median as needed to enhance model performance. For Linear Regression, using the mean over the median showed no significant difference, while for the Random Forest model, median imputation provided better results. Feature selection was performed using SelectKBest. To optimize the number of selected features, GridSearchCV was employed, testing multiple values of k to find the best configuration. To further validate model performance and minimize overfitting, cross-validation was applied throughout the pipeline.

Finally, model selection was based on the Mean Absolute Error (MAE) from cross-validation, with the Random Forest Regression model outperforming Linear Regression.

Using Random Forest Regression Big Mountain Resort modelled price is found $95, actual price is $81.00.
Even with the expected mean absolute error of $10.39, this suggests there is room for an increase.

Now, from the shortlisted suggestions of the business if we consider shutting down runs to minimize cost, we observe that (In Fig: 4) one closed run doesn't have any impact where 2 or 3 successive closed runs reduces support for ticket price and so revenue. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop.
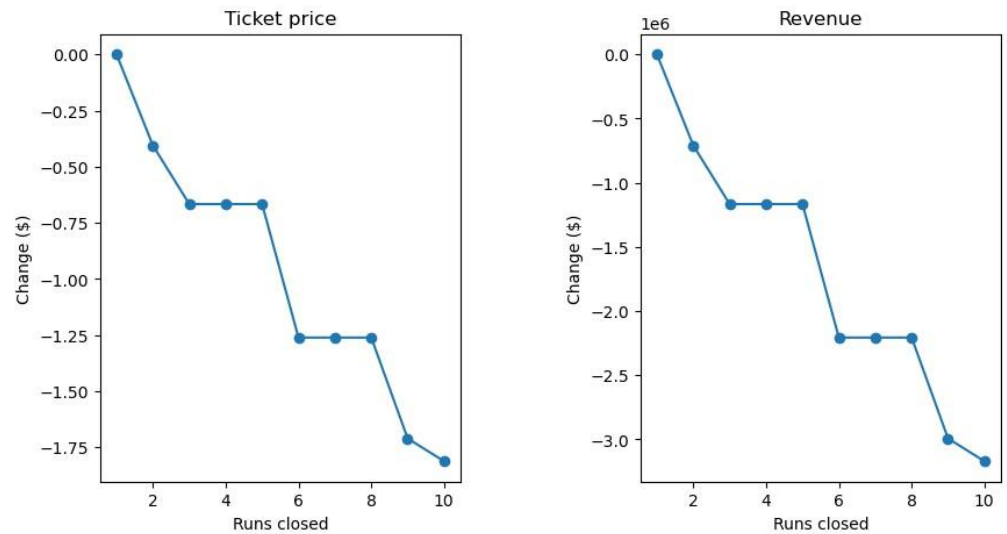


Fig: 4. Impact of run closure on ticket price and revenue

In another scenario to focus on features which contributes most to the ticket price, Big Mountain is adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift. This scenario increases support for ticket price by $8.61. Over the season, this could be expected to amount to $15065471.

In this next scenario, previous one was repeated but with adding 2.0 acres of snow making. This scenario increases support for ticket price by $9.90. Over the season, this could be expected to amount to $17322717. Such a small increase in the snow making area makes no difference!

In another scenario, the longest run is increased by 0.2 miles and guaranteeing its snow coverage by adding 4.0 acres of snow making capability. This one makes no difference whatsoever.

The reason for the modeled price so much higher than its current price is that current price was not based on data driven decision. This could surprise executives if they've historically focused on price competition rather than maximizing revenue through price adjustments.

The model is a valuable tool for exploring pricing adjustments based on amenities changes, helping executives plan for future investments in facilities. Leaders could use the model to estimate revenue impacts of facility changes or price increases. This could inform budgeting and forecasting. The model allows the business to test different pricing and facility configurations to stay competitive while optimizing revenue.

To avoid repetitive requests, the model should be aimed to make accessible to analysts. A user-friendly dashboard with widgets could let analysts adjust features like the number of runs, vertical drop, or chair lifts and see predicted prices and revenue impacts. Tools like Streamlit or Dash could enable this without requiring coding knowledge. If more flexibility is needed, an API could allow analysts to interact with the model directly through tools like Excel or Power BI, making it easy to test scenarios. A short training session for analysts on how to use the model effectively would empower them to leverage it independently.