Big Mountain Resort (BM) Report

- Suggestions for Revenue Increase -

EC, Data Science Team

1. Problem Identification

- **Problem Statement:** How can BM increase revenue for the resort by at least $1,540,000 by the next season by (1) cutting costs down without undermining the ticket price or (2) setting a higher ticket price via data-driven standards?
- **Context**: BM has recently installed an additional chair lift. This additional chair increases their operating costs by $1,540,000 this season. Our goal is to find some guidance on selecting a better value for their ticket price based on data and recommend several ideas to increase revenue.
- **Criteria for Success**: New pricing strategy and a clear set of actions to reduce operating costs.
- **Scope Of Solution Space**: Ski resorts located in Montana and Washington where can be the competitors for the BM, and Ski resorts with similar specifications to the BM.
- **Constraints Within Solution Space**: Insufficient data regarding operating expenses and revenue sources.
- **Stakeholders To Provide Key Insights**: Jimmy Blackburn (Director of Operations), Alesha Eisen (the Database Manager), and any related employees in the management team or data science Team
- **Date Sources**: Single CSV file containing information from 330 resorts in the U.S.

2. Data Wrangling

Data wrangling was conducted through examination, cleaning, and exploring additional data.

- **Dropped Data**: The 'fastEight''s: half the values are missing and all but the others are the value zero. There is essentially no information in this column. And there is no significant difference between weekend and weekday prices. And weekend prices have lesser missing values than weekday prices. So drop the weekday prices and then keep just the rows that have weekend prices. And there were a number of missing values that led to several rows being dropped completely.
- **Error corrected Data**: Some suspicious data were corrected based on research; Silverton Mountain's skiable terrain area, and Heavenly Mountain Resort's snowmaking area.
- **Augmented Data**: US state population and size data after cleaning.

3. Exploratory Data Analysis (EDA)

Through EDA, several features were selected for modeling

- **Features:** selected features describe the various attributes of each resort and state
  - **Numerical features**: 21 features from original data, and 7 features from US population data
  - **Categorical features**: 'Name', 'Region', and 'state'.

- **Target feature**: ‘AdultWeekend’
  - **Selected features for subsequent modeling**: ‘vertical_drop’, ‘fastQuands’, ‘Runs’, and ‘total_chairs’.
- **How to handle the state labels:** There was no clear pattern suggestive of a relationship between state and ticket price. Due to this, state labels will be treated equally in building a pricing model that considers all states together.
- **PCA (Principal component analysis) Findings** [Fig.1]
  - States, outliers in the dataset: Vermont, New Hampshire, New York, and California. All had either a huge number of resorts per capita or per square mile relative to the remaining states being examined.
  - However those outliers don't appear to weigh the ticket price at all, so this isn't useful for the purposes of modeling ticket price in later steps.
- **Feature correlation heatmap** [Fig.2]
  - Features that were more positively and negatively associated with the ticket price:
    - primary features: FastQuads, Runs, Vertical Drop, Total Chairs
    - secondary features: Longest Run, Skiable Acres, Snowmaking Acres, and Night Skiing.

4. Model Preprocessing with feature engineering

- **Baseline idea of performance:** The mean of the target train set, y_train, is approx. 63.81. And the mean of the dummy regressor on the training data is approx. 63.81. And the results whos that both values are the almost same. Therefore it can tell that the best guess could be the average price.
- **Linear Model:** LinearRegression fits a linear model with coefficients to minimize the residual sum of squares between the observed targets in the dataset and the targets predicted by the linear approximation. The linear model was built by imputing missing values, scaling the features, and training a model. Imputing the missing values in the data set using either the mean or the median did not seem to matter or significantly change the results. Both results show that The estimated ticket price will be within 9 dollars or so of the actual price.
- **Cross-Validation:** These results highlight that assessing model performance is open to variability. The result could be different depending on the quirks of which points are in which fold. But the result shows that the mean of the test score is approx. 0.68, and the standard deviation is approx. 0.05. And using GridSearchCV, the best parameter for the k in the pipeline was derived as '8'. So the vertical drop is your most significant positive feature (10.767857). And the coefficient for six features is positive for this model.
- **Random Forest:** The key processing steps are conducted by defining the pipeline Fit and assessing performance using cross-validation. Hyperparameter search using GridSearchCV As a result, fastQuads, Runs, Snow Making_ac, and vertical_drop are the dominant top four features.

5. Winning model and scenario modeling

- **Final Model Selection** (**L: Linear regression model, R: Random forest regression**): The random forest model has a lower cross-validation MAE by almost $1 (L: approx. 10.50, R: approx. 9.64). It also exhibits less variability (L: approx. 1.62, R: approx. 1.35). Therefore the model of Random forest will be used to go forward.

- **Scenarios**: (1) Closing up to 10 of the least used runs, (2) adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift, (3) previous scenario + adding 2 acres of snowmaking, and (4) Increasing the longest run by .2 miles and guaranteeing its snow coverage by adding 4 acres of snowmaking capability. The first and second are worth considering.
    - [Scenario 1] **Permanently closing down up to 10 of the least used runs** [Fig.3]: This suggestion will cause the price to go down. And the annual revenue will be decreased by up to about $3,170k. But if the closing reduces annual operating expense more than that, it is worth considering closing the runs. Furthermore, in the case of the closing only one run, the model says closing one run makes no difference. Therefore, to close only one run is worth trying.
    - [Scenario 2] **Adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift:** This suggestion increases support for ticket price by $1.99. Therefore, over the season, this could raise the revenue by about $3,474k.

6. Pricing recommendation

The Big Mountain Resort's ticket price of $81.00 is higher than other resorts. However, our resort's modeled price is $95.87. Therefore, even with the expected mean absolute error of $10.39, there is room for an increase. Furthermore, if the price ticket is increased to $95.87, it leads to an annual revenue increase of $2,602,250, which is more than the increased operating expenses of $1,540,000.

7. Conclusion

As suggested by the model, initially increase the ticket price from $81 to $85.34/day/customer and monitor the daily increase of revenue and validate with recent sale information. If continues, increase the price to a maximum of $106.40. And the ticket price could be higher if more ski facilities are enhanced. But it does not necessarily mean that revenue is also increased. To verify the revenue change, we need to see the data about expense and investment as well as ticket price.

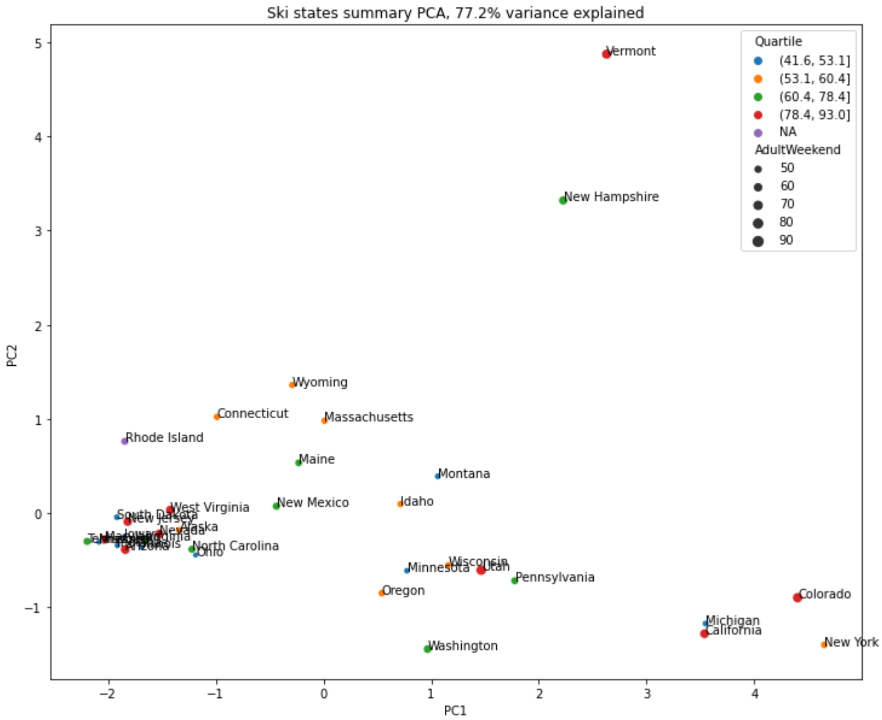
8. Further scope of work

To better verify net income increase, we need to not only revenue but also expenses. Therefore the following data is required;

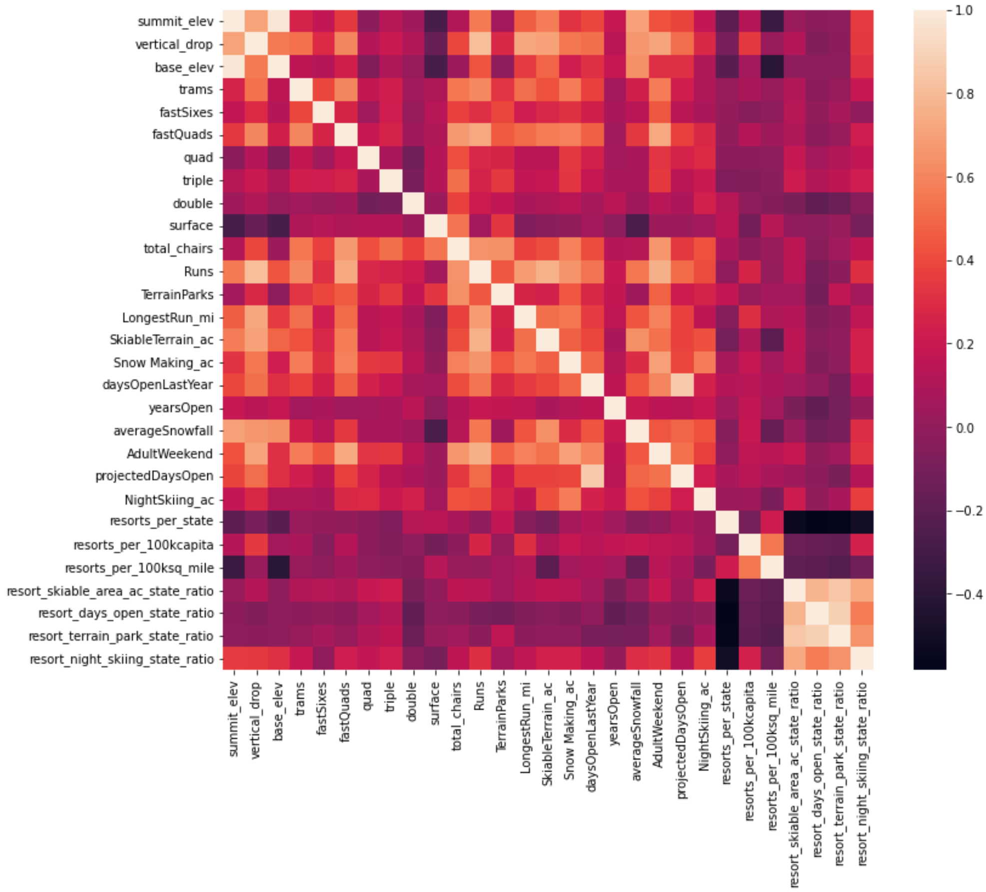
(a) Operating expense: How much cost is expensed for each facility.
(b) Investment expense: How much cost is required to install each facility.

9. Figures

[Fig.1] Ski states summary PCA, 77.2% variance explained



[Fig.2] Feature correlation heatmap



[Fig.3] Ticker price change by runs closed & Revenue change by runs closed.