Guided Capstone Project Report

# Problem Statement

Big Mountain Resort (BMR), a ski resort located in Montana, wants to be more efficient and profitable. With the recent addition of a new chair lift aiming to facilitate the mobility of its 350,000 annual visitors across the property, increased operational costs for the new lift are forecasted to be an estimated $1.54M for the upcoming season.

To offset this substantial expenditure and to potentially increase the resort's revenue, BMR is contemplating revisiting its current pricing strategy. There is a growing suspicion that the resort might not be fully capitalizing on its various facilities, and an adjustment in the ticket pricing could potentially uncover untapped revenue streams.

At the same time, BMR is keen on exploring other avenues for cost savings. The dual strategy involves a careful analysis of the current facilities, amenities, and services, and evaluating their performance and contribution to the resort's revenue.

# Data Wrangling

Data wrangling is a process by which you begin the process of creating tidy data for modeling. This step focuses on collecting and organizing data.

* **Analyzing and Visualizing Data** 
  + **Trends** Utilizing statistical methods and visualizations to discern patterns and anomalies, facilitating informed decision-making to achieve tidy data.
* **Categorical and Numerical Data Examination**
  + Detailed scrutiny of categorical and numerical data, wherein their uniqueness, missing values, and distributions were meticulously explored, laying the groundwork for meaningful insights.
* **Data Cleaning and Optimization**
  + Engaging in the strategic elimination of unnecessary rows and columns, streamlining the dataset for optimal utilization in the forthcoming phases.
* **Feature Engineering and Aggregation**
  + Creating aggregations based on the available data and appending them as new columns, thereby enhancing the dataset's comprehensiveness and richness.
* **Integration of Additional Data**
  + Incorporating population data to bolster the existing dataset, ensuring a more rounded analysis in the subsequent stages.

The chart below shows the average weekday and weekend ticket price for each state.

A graph of different colored and black lines

Description automatically generated with medium confidence

This chart shows that, barring a few high-priced outliers, the majority of ticket prices hover in the 25-100 range. Moreover, it indicates a general trend of states maintaining parity between their weekday and weekend rates, fostering a uniform pricing strategy.

This exploration brought forth two significant queries:

1. How should the resort approach the disparity between the two ticket prices?
2. What is the potential impact of including state data as a feature in the pricing model?

# Exploratory Data Analysis

Exploratory data analysis is a more comprehensive analysis of the data on hand.

* Data exploration
  + examining statistical summaries of different states and analyzing resort density
  + to visualizing high dimensional data, employing techniques like PCA transformation
* Feature Engineering with numeric data
  + Creating correlation heatmaps and scatter plots to understand the potential relationships between different features and ticket prices

A red and pink squares with black text

Description automatically generated

The heatmeap illustrates a high correlation between our target “AdultWeekend” price and other features such as “vertical\_drop” and “Runs”.

# Pre-Processing and Training Data

During this phase, the focus was to initiate the development of predictive models.

* Establishing a Baseline using Mean: Here, the goal was to gauge the performance of our predictive models against a baseline model that simply predicts the mean. Various metrics such as R-squared, Mean Absolute Error, and Mean Squared Error were employed to understand and quantify the performance.
* Model Development
  + Data Treatment: Imputing missing values using strategies such as median and mean imputations.
  + Data Processing: scaling of features and splitting the data into training and testing sets.
  + ModelTrainingandTesting: Building and testing models, followed by an assessment of their performance.
* Model Types and Techniques
  + Linear Model: Implemented through pipeline configurations, utilizing cross-validation techniques and GridSearchCV for hyperparameter tuning and optimization.
  + Random Forest Model: Followed a similar process as the linear model but employed a different algorithmic approach.
* Final Model Selection and Saving
  + The best-performing model, the Random Forest model, was selected and saved for future use.

# Modelling

In the modeling phase, the selected model is used to make precise predictions.

* Calculating Expected Ticket Price
  + The model is utilized to calculate an expected ticket price for Big Mountain Resort, serving as a point for further analysis.
  + The model projects a ticket price of $95.87 with a standard error of $10.39. Even considering this variation, there's evident room for a price hike, chiefly because BMR boasts several top-tier amenities sought after in the ski resort industry
* Modeling Scenarios
  + Here, different options to alter the resort are modeled to understand different potential outcomes.Top of Form
* Recommendations
  + From the scenarios assessed, Scenario #2, which suggests adding a run, augmenting the vertical drop by 150 feet, and introducing a new chair lift, stands out. Given that vertical drop, chair quantity, run count, and run length are pivotal price determinants, this scenario could justify a ticket price boost of $1.99. This increment, when projected over the entire season, can yield an additional revenue of approximately $3.47 million. Significantly, the revenue from this price surge alone could offset the operational expenses of the new chair lift.
  + scenario 1 shows promise but requires further investigation; the modeling indicates closing up to 5 runs would result in a loss of less than 0.75/ticket. Further analysis could potentially involve looking into more granular data to understand which specific runs are least used and evaluating the cost-benefit analysis of closing them. A graph with blue lines and numbers

    Description automatically generated

# Conclusion

The model suggests that Big Mountain Resorts can most certainly raise its ticket prices with or without adding features. If it chooses to add features, like in scenario 2, another price increase is justified. More detailed analysis will reveal the true cost savings associated with closing runs.