Guided Capstone Project Report: Big Mountain Resort

Big Mountain Resort in Montana offers stunning views, 105 trails, and facilities for all skill levels, including 11 lifts, 2 T-bars, and a magic carpet. The resort boasts a 3.3-mile run, a base elevation of 4,464 ft, summit of 6,817 ft, and a 2,353 ft vertical drop. Despite a recent $1,540,000 chair lift addition, the pricing strategy may be holding back its potential. The resort seeks a data-driven approach to optimize ticket pricing and consider cost-effective changes while staying competitive.

To analyze and make informed decisions, the raw data from ski_resort_data.csv was studied. The dataset comprised 27 columns with 330 entries, but it required cleaning due to a considerable number of missing values and inconsistencies. A thorough exploration of the data was conducted for all states, revealing that the 'fastEight' column had the highest proportion of missing values, just over 50%. Unfortunately, there was also a notable absence of data for the desired target variable, ticket price, with 15-16% of values missing. Adult Weekday had slightly more missing records compared to Adult Weekend. Given our focus on Big Mountain, outliers and missing values in relevant columns were rectified. The final dataset for predictive modeling consists of 25 columns and 277 rows. The data was then prepared for further investigation."

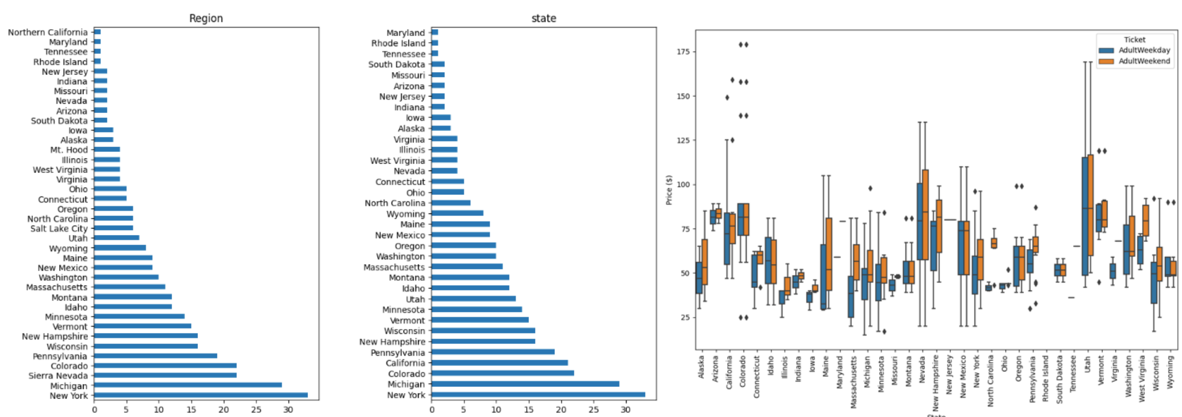

Fig: Distribution of resorts(left), and price(right)

The cleaned data underwent further detailed analysis, with a primary focus on price. Exploratory data analysis was conducted using techniques such as pair plots, box plots, and scatter plots. Additionally, feature extraction was performed. Principal Component Analysis (PCA) was implemented, as it aids in reducing the number of features in a dataset while preserving as much of the original information as possible. PCA enables the visualization of high-dimensional data in lower dimensions, and it can be used to extract the most important features or components of a dataset. The refined data with features for further analysis has been saved.

For the next step, the refined data was utilized for preprocessing and training. The data underwent preprocessing and was split in a ratio of 70% for training and 30% for testing. By splitting the data into training and testing sets, we ensure that the model doesn't learn from the test split. The target column was 'Adult Weekend'. During data processing, various performance metrics from the Sklearn metrics module were demonstrated, such as R-squared (R2), Mean Absolute Error (MAE), and Mean Squared Error (MSE). For further processing, the focus shifted to imputing missing values using the median, which is a robust measure of central tendency based on the skewed distributions observed during exploratory data analysis (EDA). To maintain consistent scaling, the 'Standard Scaler' was employed to ensure zero mean and unit variance. This scaler was fitted on the training data and then applied the transformation to both the training and test data. The 'Standard Scaler' is a preprocessing step in machine learning that standardizes features by removing the mean and scaling to unit variance. This aids in making the features more comparable.

For model selection, a linear regression model was initially used to predict ticket pricing. Despite showing promising results in cross-validation, it fell short of expectations on the test set, prompting us to explore more complex models. The introduction of the random forest regressor, a powerful tool for revealing intricate data patterns, along with enhanced preprocessing techniques including feature scaling and handling missing data, proved to be a game-changer. This model not only outperformed the linear one in cross-validation but also demonstrated exceptional consistency with outstanding results on the test set. Opting for the random forest regressor due to its superior predictive performance and resilience, the ski resort is now empowered to make data-driven decisions, optimizing revenue and ensuring a top-notch guest experience. This strategic choice ensures that our pricing strategy is well-informed and effective. The model identified several important features, including Vertical Drop, Total Runs, Snow Making Acres, Length of Longest Run, Total Lift Chairs, Number of Trams, and Fast Quad Lifts, as key drivers of ticket pricing.

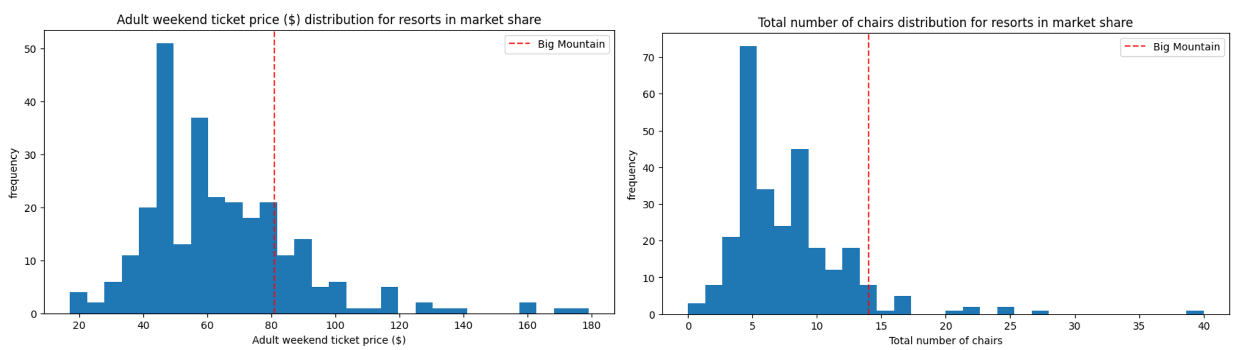

Fig: Histogram distribution of adult weekend price (left) and total number of chairs (right).

For the decisive analysis different scenarios were tested: Scenario 1 analyzed run closures, revealing limited impact on pricing until six or more runs were closed, causing a significant drop. Scenario 2 proposed resort improvements, projecting a $1.99 ticket price increase and potential $3.5M revenue boost. Scenario 3's snow-making expansion showed marginal pricing influence. Scenario 4's run extension and snow-making addition had negligible effects. Scenario 2 stands out, offering the most promise for revenue growth and enhanced offerings. Further testing and customer input are advised for implementation. Other scenarios showed less substantial impact on pricing and revenue.

In line with its premium market position, Big Mountain Resort currently charges $81.00 per ticket, but it has the capacity to support a slightly higher price point, approximately $95.87. With the addition of a new chair lift, the resort will incur an extra cost of $1.85 for each ticket. Scenario 2, which involves expanding the snow-making area, shows promise, projecting a \$2.14 increase in ticket price and potential income growth. Initiating a trial phase for customer feedback to evaluate run closures would be beneficial. To enhance the analysis, additional data on operational expenses such as maintenance, staffing, and marketing is needed. It is crucial to survey business executives for their expectations regarding pricing strategy alignment. If the model proves valuable, integrating it into a user-friendly tool for analysts would be advantageous. Providing training resources would further empower them, reducing the need for constant consultation. The goal is to enable independent pricing evaluations, allowing data scientists to tackle more complex tasks. Thus, creating an accessible, user-friendly model is essential for this transition.