Executive Summary: Big Mountain Resort DSM Review & Recommendations
Prepared by: Jake Charney, Junior Data Scientist
11/22/2023

Overview:

The goal of this project was to analyze and model ticket prices for ski resorts, with a specific focus on our client Big Mountain Resort. The analysis involved exploring the dataset, cleaning the data, performing exploratory data analysis (EDA), and building a predictive model for ticket prices. The key findings and recommendations are summarized below.

Data Collection and Cleaning:

The initial dataset comprised 330 rows and 27 columns, with no missing values for Big Mountain Resort. However, missing values were identified in crucial columns such as 'AdultWeekend' and 'AdultWeekday.' The 'fastEight' column, with 50% missing values, was deemed unnecessary and removed. Resort names were analyzed, uncovering a duplication issue for Crystal Mountain, which was resolved. I also determined that there were discrepancies in resort names and identified a non-one-to-one relationship between 'Region' and 'State.' I explored ticket prices across states and identified variations, especially in weekend prices. Further data cleaning involved removing rows with missing prices and addressing missing numerical values using imputation. I then derived state-wide summary statistics for relevant features and created a merged dataset.

	yearsOpen	**SkiableTerrain_ac**	**Runs**	**LongestRun_mi**	**averageSnowfall**	**SnowMaking_ac**	**projectedDaysOpen**	**TerrainParks**	**daysOpenLastYear**	**AdultWeekend**	**AdultWeekday**	**NightSkiing_ac**	**fastEight**
Missing value count	1	3	4	5	14	46	47	51	51	51	54	143	166
Percent missing	0.3030	0.9090	1.212	1.515	4.242	13.939	14.242	15.454	15.454	15.454	16.363	43.333	50.303

Figure 1: Table of missing value counts and ratio of missing values compared to the total. Columns with no missing values were removed for clarity.

Data Exploration and Analysis:

Exploratory Data Analysis (EDA) revealed insights into ticket price distribution across states, highlighting variations and trends. The relationship between ticket prices and state-specific features, such as skiable terrain area, was visualized. Boxplots and scatterplots provided a comprehensive understanding of the dataset. At the resort level I explored correlations between various resort features and ticket prices. Furthermore, I applied principal component analysis (PCA) to identify key features contributing to variance in ticket prices.

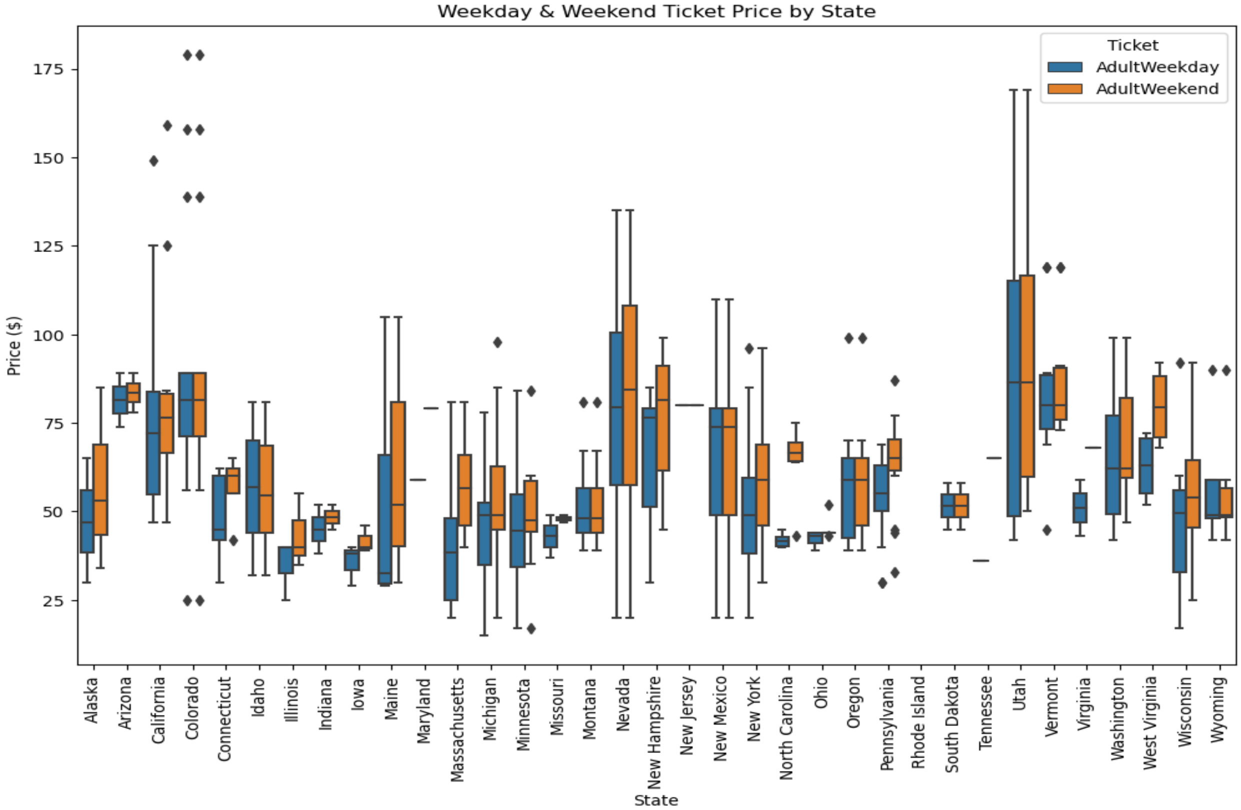

Figure 2: Box plot created using seaborn of 'AdultWeekday' & 'AdultWeekend' prices by state. The boxes represent the IQR, spanning from the 25th to the 75th percentile of ticket prices in each state. The height of the box indicates the spread of prices within this range. The Lines extending from the box indicate the range of prices outside the IQR. Outliers beyond this range are represented as individual points.

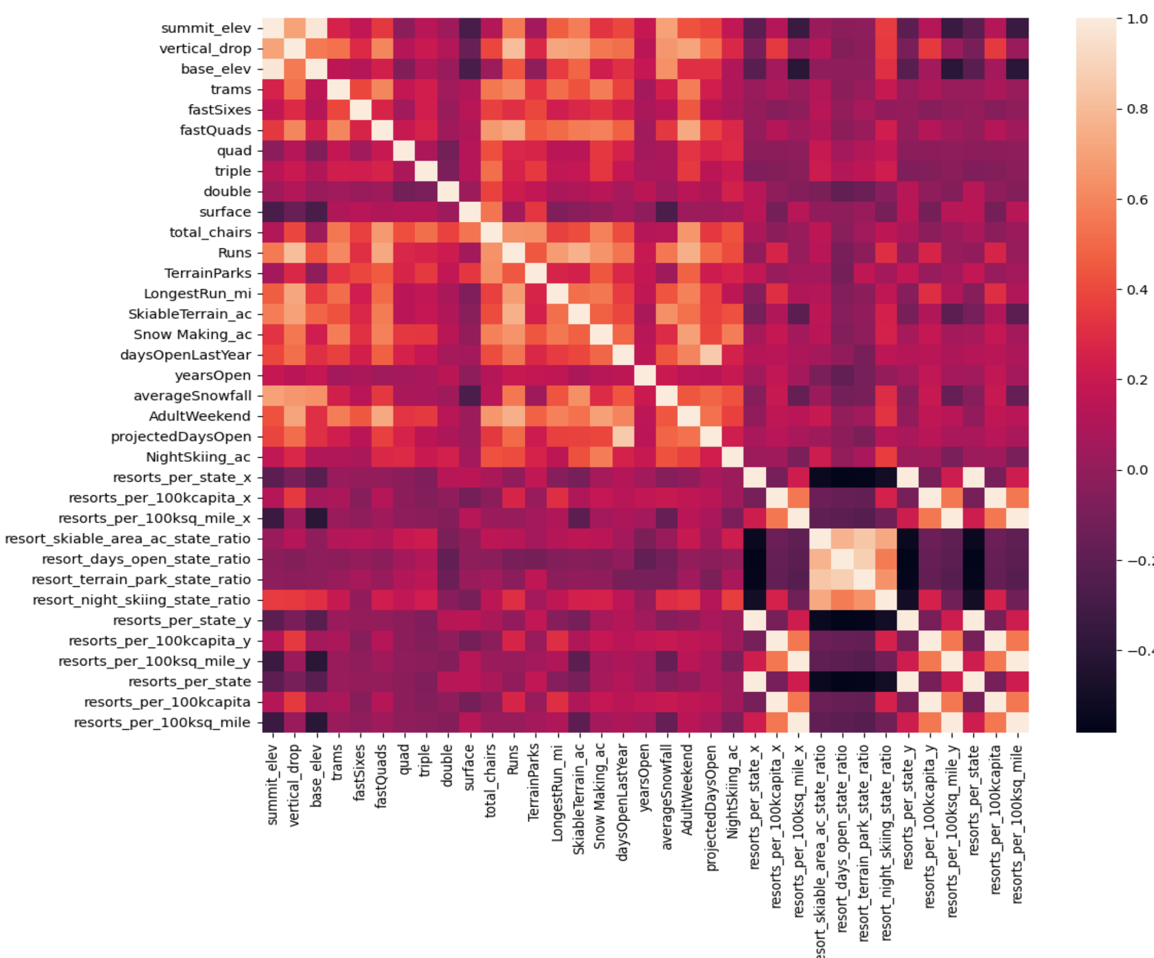

Figure 3: Heatmap displaying a comprehensive overview of the relationships between various features of ski resorts, facilitating insights into potential correlations and patterns. Each row and column represent specific features, and the color intensity reflects the strength and direction of their correlation. The color spectrum, ranging from cool to warm tones, signifies the correlation strength. Cool tones (e.g., orange) indicate a negative correlation, while warm tones (e.g., violet) suggest a positive correlation.

Preprocessing and Predictive Modeling:

Machine learning models, including linear regression and random forest, were trained to predict ticket prices. I trained the linear regression model using a test and training group and evaluated its performance using the coefficient of determination (R2), mean absolute error (MAE), and mean standard error (MSE). I utilized feature selection techniques to identify key features impacting ticket prices and identified crucial factors influencing ticket prices, including vertical drop, snowmaking acreage, and fastQuads. I then sought to improve predictive accuracy by developing a Random forest model (RFM). The RFM was then used to confirm dominant features affecting ticket prices and flagged the same features as the linear regression model. However, the random forest model exhibited better performance in cross-validation and on the test set. The RFM was then utilized to model scenarios based on client propositions, such as closing runs or adding facilities and the revenue impact of each scenario was evaluated, providing insights for business decisions.

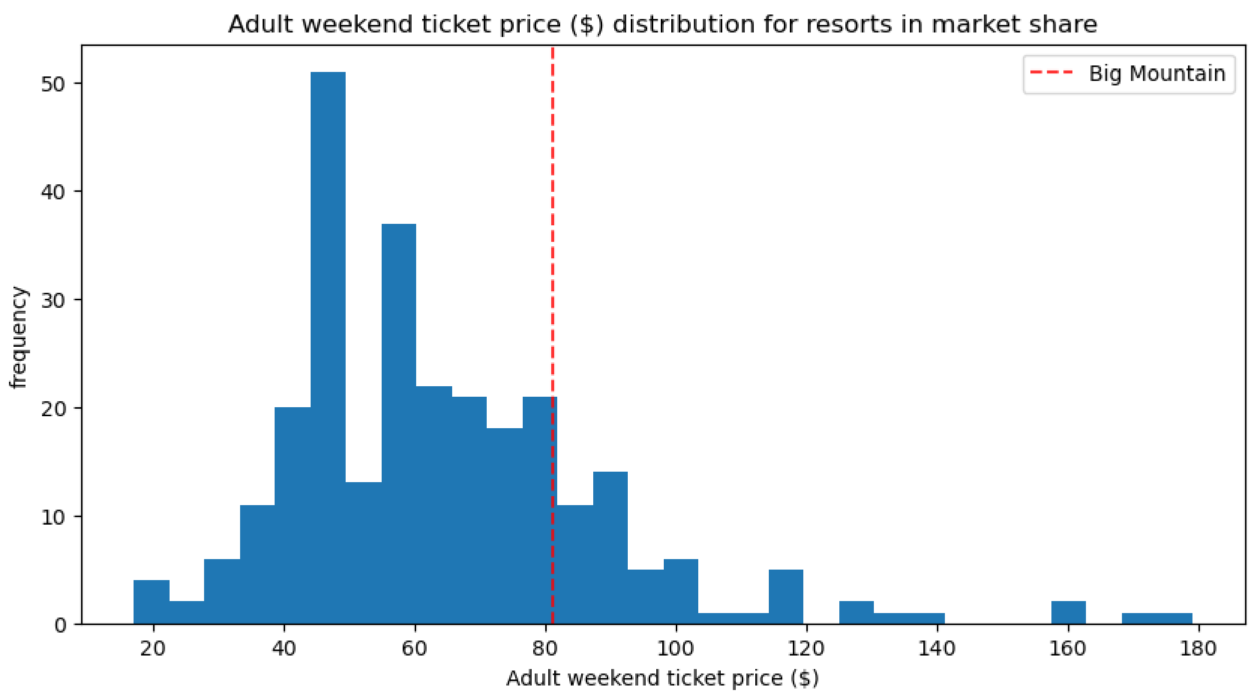

Figure 4: Histogram created with seaborn illustrates the frequency distribution of adult weekend ticket prices across various ski resorts. Each bar represents a price range, and its height corresponds to the number of resorts falling within that range. The x-axis denotes the price intervals, while the y-axis indicates the count of resorts. The red dashed line is positioned at the ticket price of Big Mountain Ski Resort. This line serves as a reference point, allowing for an immediate comparison of Big Mountain's ticket price to the overall distribution.

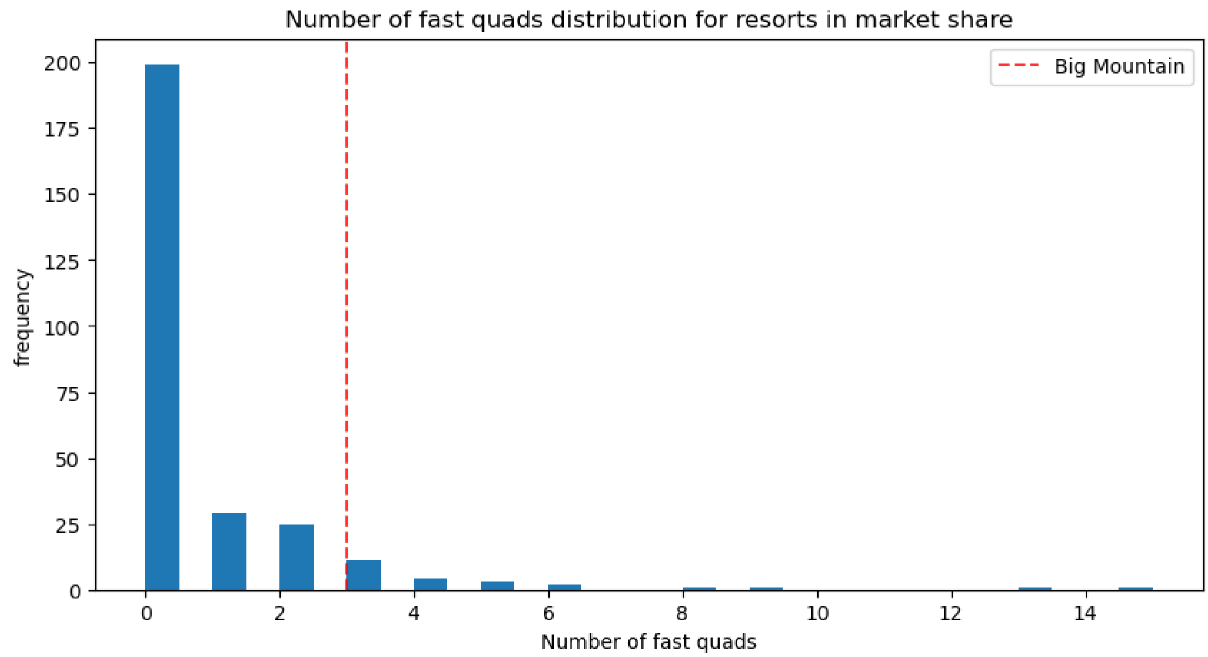

Figure 5: Histogram created with seaborn illustrates the frequency distribution of the number of fast quad chair lifts across various ski resorts. Each bar represents a price range, and its height corresponds to the number of resorts falling within that range. The x-axis denotes the number of fast quad chair lifts, while the y-axis indicates the count of resorts. The red dashed line is positioned at the amount of fast quad chair lifts operated by Big Mountain Ski Resort.

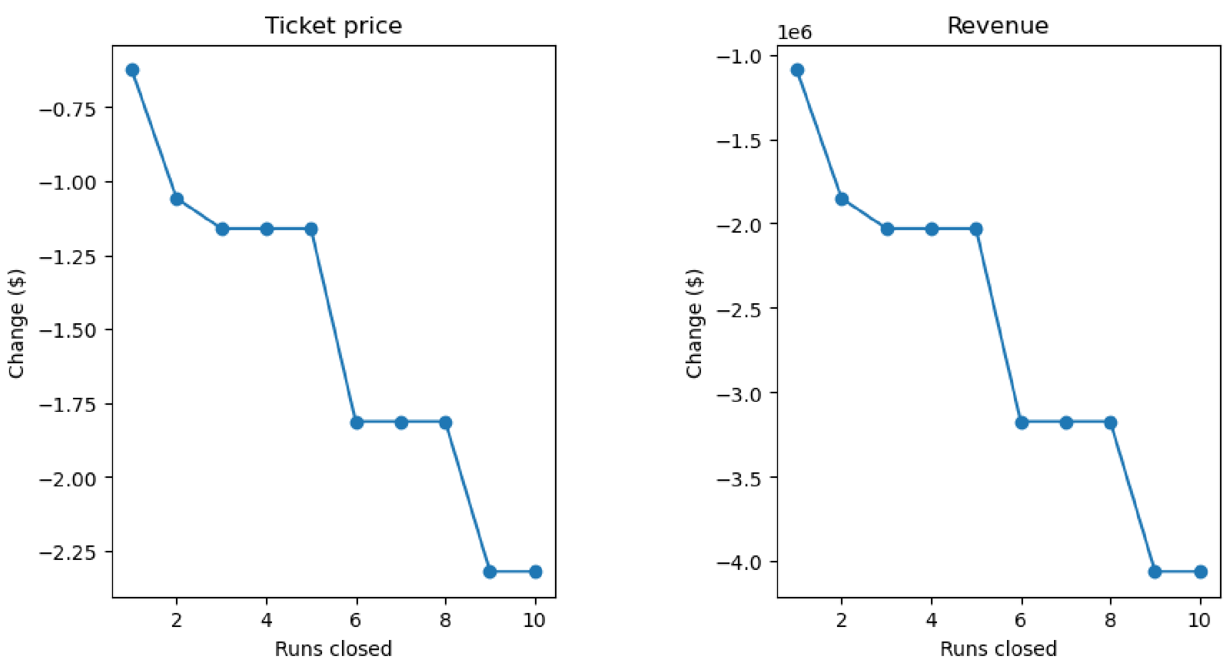

Figure 6: This figure consists of two interconnected line plots, each sharing the same x-axis, representing the amount of runs closed. The left plot illustrates the change in ticket price, while the right plot depicts the corresponding change in revenue. Both the y-axes are scaled to facilitate a direct comparison between the two variables. By examining both line plots simultaneously, viewers can discern the correlation between closing runs, the subsequent change in ticket price, and the corresponding revenue impact. This visual representation aids in making informed decisions by weighing the trade-offs between adjusting ticket prices and the overall financial outcome for the ski resort.

Recommendations:

The recommendations for the client based on my analysis of the data are as follows:

1. Utilize the predictive models, especially the random forest model, to inform pricing decisions to gain an optimized pricing strategy. I recommend exploring tiered pricing models based on facilities, especially focusing fast quads based on Big Mountain's league rank.
2. Invest in snow making technology and increased vertical drops. Given the positive correlation between snowmaking acreage and ticket prices, consider investing in snowmaking technology to enhance skiing conditions and potentially justify higher ticket prices. Also, as vertical drop was identified as a significant positive feature, consider ways to enhance or market the resort's vertical drop to attract visitors
3. Engage in a marketing focus to highlight key facilities that customers value like the ones mentioned above. Emphasize the unique aspects that contribute positively to ticket prices.

Limitations and Next Steps:

While this model provided key insights, it was limited due to a lack of data surrounding key factors that might influence ticket pricing. I suggest acquiring data on visitor numbers, operational costs, rental equipment, and revenue from rentals for more accurate modeling. Moving forward I propose making the model available for business analysts, allowing them to test new scenarios without direct involvement.

Conclusion:

The analysis provided valuable insights into factors influencing ticket prices, enabling informed decision-making for Big Mountain Resort. Further enhancements and exploration of scenarios are recommended for continued business optimization. Regular updates to the models and continuous monitoring of market trends will ensure the resort remains competitive and aligned with customer expectations.