Matt Snyder

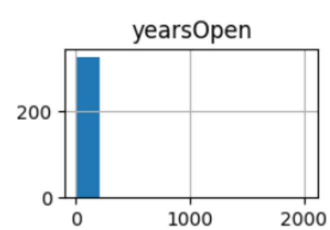
Springboard Data Science Career Track

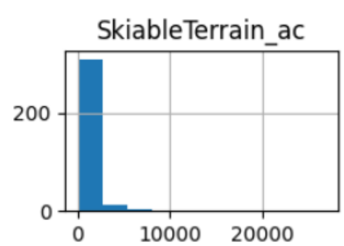
11/7/2025

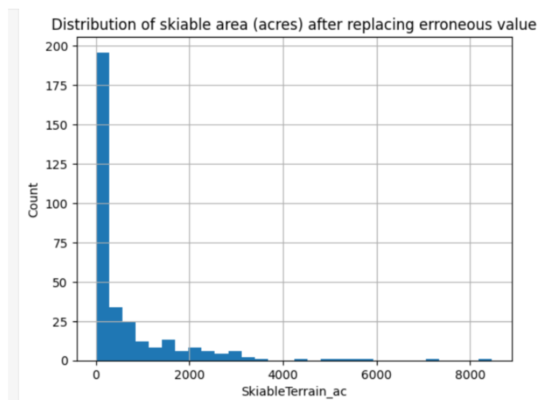
Guided Capstone Project Report – Big Mountain Resort

Big Mountain Resort is looking to increase their top-line either by reducing operation costs or capitalizing on undervalued facilities. They have also been looking to recoup their $1.5 million (M) that was spent on additional chair lifts. We will be utilizing various modelling techniques to determine the relations and significance of different amenities the resort owns and we will look to see if there's any strong correlation between amenities and features in regard to ticket pricing. Specifically, adult weekend tickets (ticket price on weekends) has been selected to be our target variable because it reflects the peak demand period and is most representative and comparable measure to a resort's pricing strategy.

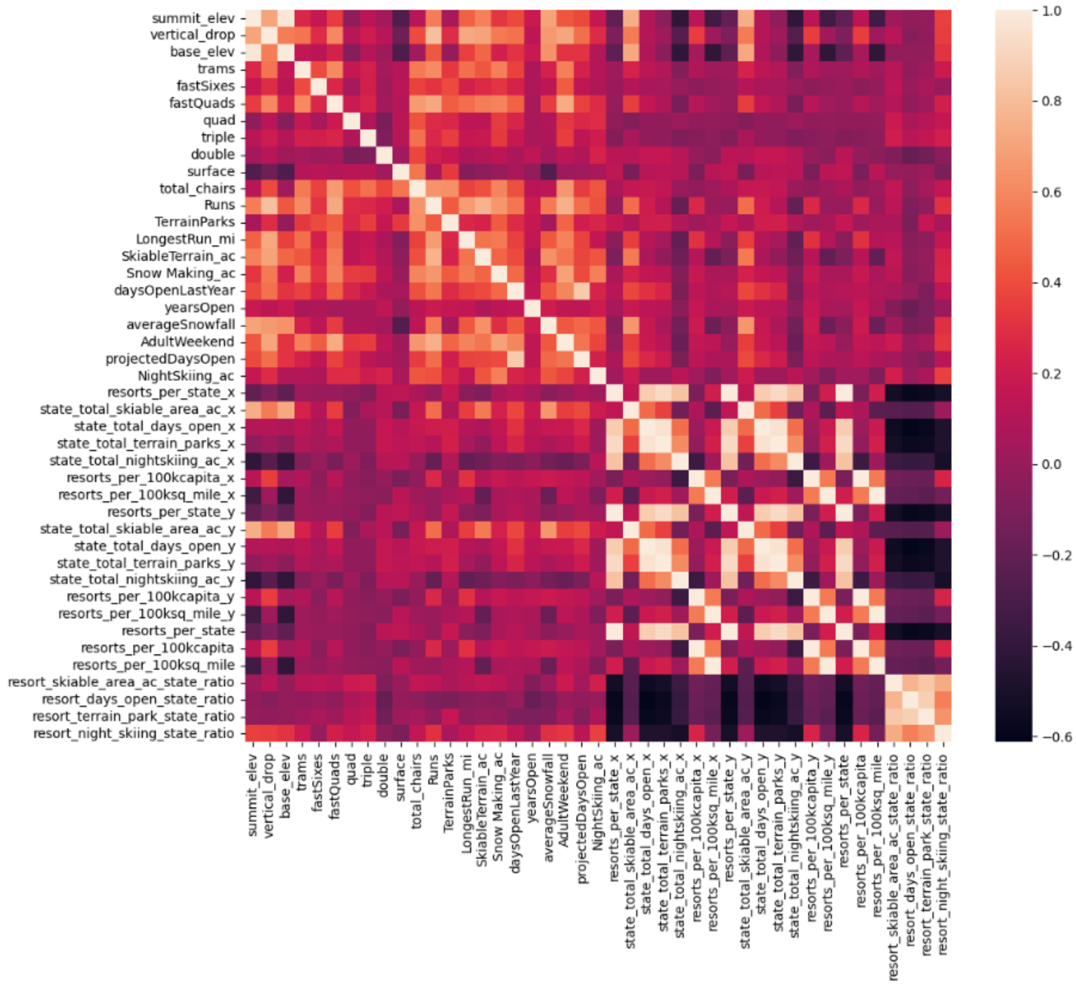
We will be using the Ski Resorts dataset which contains numerical and categorical information for over 300 U.S. ski resorts, covering features such as vertical drops, skiable acreage, total chairs, fast quads. The ski resorts are categories by state and broader geographic region. Amenities will also be compared, such as terrain parks, night skiing and snow making capacity.



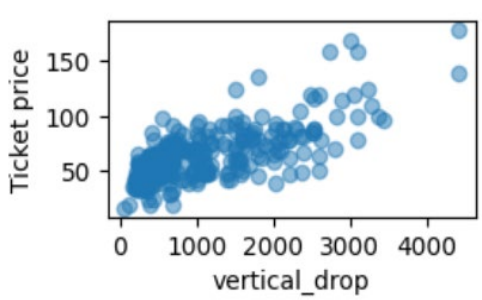


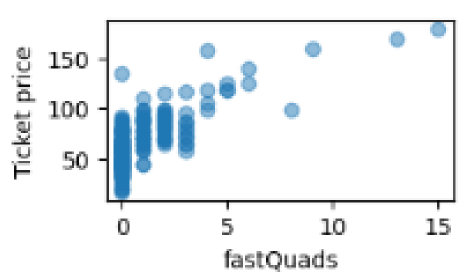


Cleaning the data included filling in missing values using median imputation for continuous (numerical) variables. We used median imputation, over say the mean, because the median won't be distorted by extreme values. We also found outliers in our data set, by creating a histogram for all every metrics, we found several skewed datasets. Namely, 'yearsOpen' had an outlier of over 2000+, which tells us that a calendar year was put in rather than the actual years open. Moreover, we found that our 'SkiableTerrain_ac' (acreage) looked to also have an outlier after plotting our histogram. To find the resort, we located any resorts with 'SkiableTerrain_ac' greater than 1000 acres and found that our problem lied with Silverton Mountain in Colorado. After further investigation using its own website, we found that the actual skiable acreage was closer to 1819, which allowed us to clear this error and have a more accurate graph, which can be seen in the plot below.

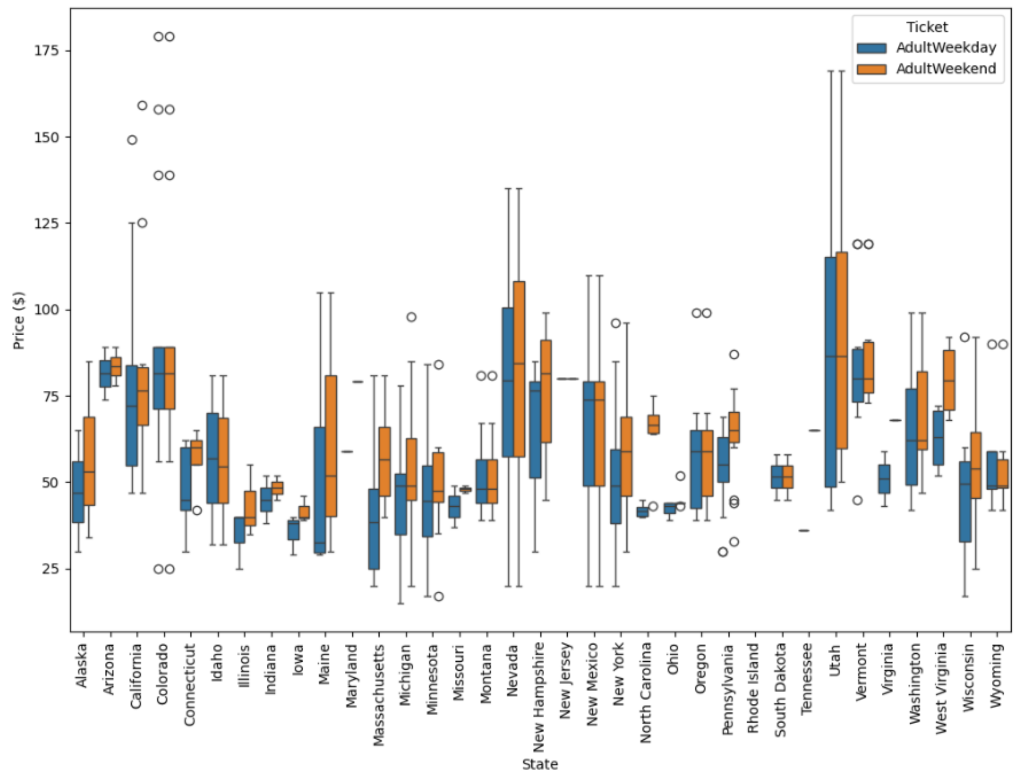


We conducted an exploratory data analysis (EDA) and produced a heat map which showed the correlations between different variables. It's important to note that heatmaps are great for identifying patterns, but high correlations can mask relationships between two variables (the variables could overlap meaning when one rises, the other rises at a near 1:1 ratio), thus usually overfitting and skewing the graph. found that ticket prices generally rise with overall resort size (which usually include more Runs, fastQuads, and TerrainParks); however, resorts with more chairs per fun or per acre often show lower prices, likely reflecting smaller, highly dense areas that have more visitors with regards to size.

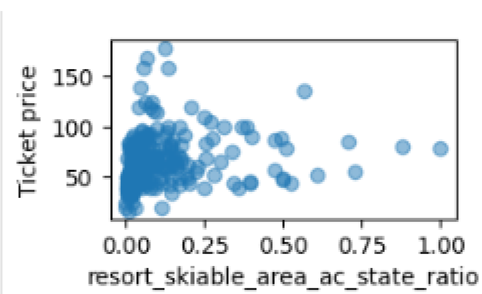




EDA also revealed clear pricing patterns such as large vertical drops and skiable acres which correlated with higher prices. Resorts with more fast quads usually charged more as that also had a direct correlation with higher prices.



There was also some regional variation, as such can be seen with states like Colorado, Utah, and California which had higher ticket prices, while smaller regional markets (Vermont, New Hampshire) trended lower. Upon further analysis, we manufactured a ratio that helped differentiate states, as we still saw smaller states commanding higher prices. Our findings show that while a state might have smaller parks, the bigger parks in the state still priced their tickets higher. See the below figure.



For preprocessing and model training, the data was split into training (80%) and test (20%) sets. Numeric features were scaled and categorical variables encoded via OneHotEncoder. Two regression approaches were tested: a multiple linear regression model to establish a baseline and a random forest regressor to capture nonlinear relationships between resort features and pricing. The random forest outperformed the linear model, achieving an R^2 of approximately 0.76 on test data. Feature importance from the random forest provide insights into which factors affect ticket pricing the most. Hyperparameter tuning was performed using GridSearchCV, optimizing for R^2 and Mean Absolute Error (MAE).

The Random Forest Regressor achieved the best accuracy with an R^2 (train) ~0.91, R^2 (test): ~0.76, and an MAE ~$11.8.

This analysis shows that ski resort pricing is primarily influenced by physical scale (respective to state size) and quality-related features, rather than pure amenities. Larger vertical drops and terrain correlate most strongly with premium pricing. Our recommendation would be to lengthen the vertical drop of the Big Mountain Ski resorts and add in another lift. Our projection shows that this will increase revenue by ~$1.6M per annum, which would cover the $1.5M already spent in operating costs. Additionally, we found that when we were able to valuate key features, we included that into our model and found that Big Mountain resort maybe undercharge $2-$5 per weekend ticket. Our recommendation for this would be to roll this out slowly or possibly only do it during peak hours, while a cost-savings analysis is to be continually conducted to see if there's a better ROI for this pricing strategy.

Limitations we faced are that we had some missing data for features, such as snowmaking. Prices reflected singly-day tickets and may not capture seasonal or dynamic pricing (e.g., cheaper if tickets are bought in bulk or seasonally).