## Problem statement

How can the Big Mountain Ski Resource make informed decisions on a better ticket pricing strategy for next opening winter season 2023 to cover the increasing operation cost from the recent facilities upgrade ,reduce cost by 10%  and yet still make profit by analysing the other resorts information ?

## Data Wrangling

The ski resort data was loaded from a CSV file and examined using the info() method, revealing that there are 330 rows and 27 columns. Some columns have missing values. To handle this, the fastEight, TerrainParks, and Snow Making\_ac columns, which had excessive missing values, were dropped. A spot check was done to ensure the data's accuracy and was cross-referenced with an external source. Any discrepancies found were corrected. Duplicates and missing values in the remaining rows were then addressed. Missing values in the Runs and LongestRun\_mi columns were filled with median values. Rows with incomplete price data were removed. Population data for the US was extracted from a Wikipedia table, underwent data cleaning, and was merged with the ski resort data. This combined dataset was then utilized for further analysis and saved as "state\_summary.csv" in a subsequent notebook.

Upon analysing pricing features, it was noted that the difference between Adult Weekday and AdultWeekend prices wasn't significant. As the Adult Weekday column had more missing values, the decision was made to retain only the Adult Weekend column.

A graph with blue dots

Description automatically generated

Figure WeekendPrice vs WeekdayPrice (there is a clear line where weekend and weekday prices are equal.)

The final dataset consisted of 277 rows and 25 columns and was saved as "ski\_data\_cleaned.csv" file, although some missing value issues still remained.

## Exploratory Data Analysis

The exploratory data analysis involved a cleaned dataset containing ticket prices, and additional summary of US state population and size data. The data is intended to predict adult weekend ticket prices for ski resorts. The dataset contained several important features, including numerical data (vertical\_drop, fastQuads, Runs, Snow Making\_ac, total\_chairs, SkiableTerrain\_ac, daysOpenLastYear, TerrainParks, NightSkiing\_ac, resorts\_per\_100kcapita, resorts\_per\_100ksq\_mile, resort\_skiable\_area\_ac\_state\_ratio, resort\_days\_open\_state\_ratio, resort\_terrain\_park\_state\_ratio, resort\_night\_skiing\_state\_ratio, total\_chairs\_runs\_ratio, total\_chairs\_skiable\_ratio, fastQuads\_runs\_ratio, and fastQuads\_skiable\_ratio) and categorical data (state).

Upon examining scatterplots and a correlation heatmap for other numerical features against ticket price, the relationship between a resort's share of the supply for a given state and ticket price may also be more complex than expected, as some resorts may be able to charge higher prices despite serving fewer visitors.

A screenshot of a computer generated image

Description automatically generated

Figure scatterplots and a correlation heatmap

A group of blue dots

Description automatically generated

Figure Scatterplots of numeric features against ticket price (a strong positive correlation with vertical\_drop. fastQuads)

Further exploration was undertaken to understand correlations with additional numerical features that relate to a resort's transportation capabilities, such as the ratio of chairs to runs and the ratio of chairs to skiable terrain. Additionally, features were engineered to capture a resort's share of the supply for a given state, such as the ratio of resort skiable area to total state skiable area and the ratio of resort days open to total state days open. These features were deemed potentially useful in subsequent modelling.

## Model Pre-processing with Feature Engineering

Feature engineering plays an important role in the modelling process. It was noted that the difference between AdultWeekday and AdultWeekend prices wasn't significant. As the AdultWeekday column had more missing values, the decision was made to retain only the AdultWeekend column. The final dataset consisted of 277 rows and 25 columns and was saved as "ski\_data\_cleaned.csv" file, although some missing value issues still remained.

## Algorithms Used to Build the Model with Evaluation Metric

The process outlined building a machine-learning model to predict ski resort ticket prices. The average price was initially taken as a baseline idea of performance. Two models, a linear regression model, and a random forest regression model, were trained and evaluated using cross-validation. A linear model was built and four strong features were identified: fastQuads, Runs, Snow Making\_ac, and vertical\_drop. Cross-validation was used to estimate the performance of the model, and it was found that the mean absolute error was around 10.4.

The random forest regressor was also chosen as the model going forward. The estimated performance of the random forest regressor via cross-validation was a mean absolute error of approximately 9.5.

## Winning Model and Scenario Modelling

The random forest regressor was the best performing model, with a cross-validated mean absolute error of approximately 9.5. This model was then utilized to make predictions on the test data with a mean absolute error of 10.1.

A graph with a line

Description automatically generated

Figure The random forest model performace with cross-validation (This shows that you seem to have plenty of data. There's an initial rapid improvement in model scores as one would expect, but it's essentially levelled off by around a sample size of 40-50.)

## Pricing Recommendation

Based on the analysis, the recommended adult weekend ticket price for Big Mountain Resort is $95.

## Conclusion and Future Scope of Work

In conclusion, based on our analysis, we recommend that Big Mountain Resort set their adult weekend ticket price at $95. Our analysis shows that the random forest regressor model provides the best prediction results. However, further improvements can be made by collecting more data on the resorts and their features. Future work could include expanding the dataset to include additional features, such as customer satisfaction ratings, amenities, and customer demographics. Additionally, further exploration of the relationship between a resort's share of the supply for a given state and ticket price may also be useful.