1. **Problem Statement**

The goal of this project is to develop a machine learning model to accurately predict the optimal pricing for Big Mountain ski resort, aiming to maximize company profitability and meet customer demand effectively. Our input was a CSV file with the raw data about different ski resorts in the USA. After exploring this raw data, we cleaned it by removing missing values and outliers. At this very first step, we already see that our target resort’s price has a mean of approximately $51.91 per ticket.

1. **Data Cleaning & Preprocessing**

-Removed missing values and outliers

-Target variable: AdultWeekend

-Dropped non-numeric or irrelevant features like Name, State, and Region.

A graph of a number of tickets

Description automatically generated

**Figure 1. Average Adult Weekend Ticket Price by State**

1. **Exploratory Data Analysis (EDA)**

-Montana is one of the largest states in skiable area and vertical drop.

-PCA was used to reduce dimensionality due to weak feature correlations.

-Key predictors:

1.) vertical\_drop, fastQuads, Runs, Snow Making\_ac

2.) Engineered features: total\_chairs\_runs\_ratio, fastQuads\_skiable\_ratio, etc.

|  |  |
| --- | --- |
| A group of blue dots  Description automatically generated | A group of blue dots  Description automatically generated |

**Figure 2.1 & 2.2. Relationship between numeric features and price**

A screenshot of a computer screen

Description automatically generated

**Figure 3. Correlation heatmap of numerical features**

1. **Model Preprocessing and modeling**
2. We split the dataset 70/30 for training and testing. The mean of the training target (AdultWeekend) is 63.81

**Table 1. Model Performance Comparison (Train/Test Split)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **R² (Train)** | **R² (Test)** | **MAE**  **(Train)** | **MAE**  **(Test)** | **MSE (Train)** | **MSE (Test)** |
| **Dummy Regressor** | **0** | -0.003 | **17.92** | **19.14** | **614.13** | **518.44** |
| **Linear Regression (Median Imputer + Scaler)** | **0.82** | **0.72** | **8.55** | **9.41** | **111.90** | **161.73** |
| **Linear Regression (Mean Imputer + Scaler)** | **0.82** | **0.72** | **8.54** | **9.42** | **112.38** | **164.39** |
| **Linear Regression (Median Imputer + Scaler, k=8)** | 0.76 | 0.60 | **9.48** | **11.79** | **145.90** | **233.71** |
| **Linear Regression (Median Imputer + Scaler, k=10)** | **0.76** | **0.62** | **9.50** | **11.20** | **142.79** | **216.79** |
| **Linear Regression (Median Imputer + Scaler, k=15)** | **0.79** | **0.64** | **9.21** | **10.49** | **127.49** | **210.04** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **CV\_scores (mean)** | **Standard deviation** | **Estimated 95% Confidence Interval** |
| **Random Forest (Median Imputer)** | 0.69 | 0.07 | [0.56, 0.84] |
| **Random Forest (GridSearchCV)** | 0.71 | 0.06 | [0.58, 0.84] |
| **Linear Regression (Median Imputer + Scaler, k = 15)** | 0.63 | 0.09 | [0.44, 0.82] |

**Table 2. Cross-Validation Metrics Summary for Best Models**

|  |  |  |
| --- | --- | --- |
| **Model** | **MAE** | |
| Linear Regression | 10.49 ± 1.62 | |
| Random Forest | | 9.63 ± 1.35 |

**Table 3. Final model comparison**

From the table, we can clearly see that Random Forest clearly outperformed the linear model in terms of error and generalization. After thorough exploration and experimentation, the Random Forest Regressor was selected as the final model due to its stronger performance, stability, and ability to model non-linear patterns in the data.

A graph with blue and white text

Description automatically generated

**Figure 4. Best random forest regressor feature importances**

1. **Conclusion**

The modeling results reveal that Big Mountain’s current AdultWeekend ticket price of $81 is substantially below the predicted optimal price ($95.87 ± $10.39 MAE), indicating room to raise rates without losing competitiveness. Scenario analyses further demonstrate that investments in additional lifts, increased vertical drop, and expanded snowmaking capacity could support price increases, driving meaningful revenue gains.