Untitled

November 19, 2025

1 Guided Capstone Project Report

1.1 Problem Statement

Big Mountain Resort aims to determine an optimal lift-ticket price that reflects its improved offerings while remaining competitive in the ski resort market. With the installation of a new chairlift that increases operating costs, the resort needs to understand the following:

- How its current pricing compares with similar resorts
- What factors drive ticket prices across the industry, and
- What a fair, data-driven price should be after improvements.

1.2 Data Wrangling

The dataset required substantial preparation before analysis and modeling:

- Missing values were identified and handled through imputation or removal based on their impact.
- Incorrect or extreme outliers (e.g. unusually large skiable acreage) were flagged and validated.
- Data types were fixed, including converting numerical-looking strings to numeric formats.
- New variables such as lift density, terrain size categories, and average facility indices were created to support deeper modeling.

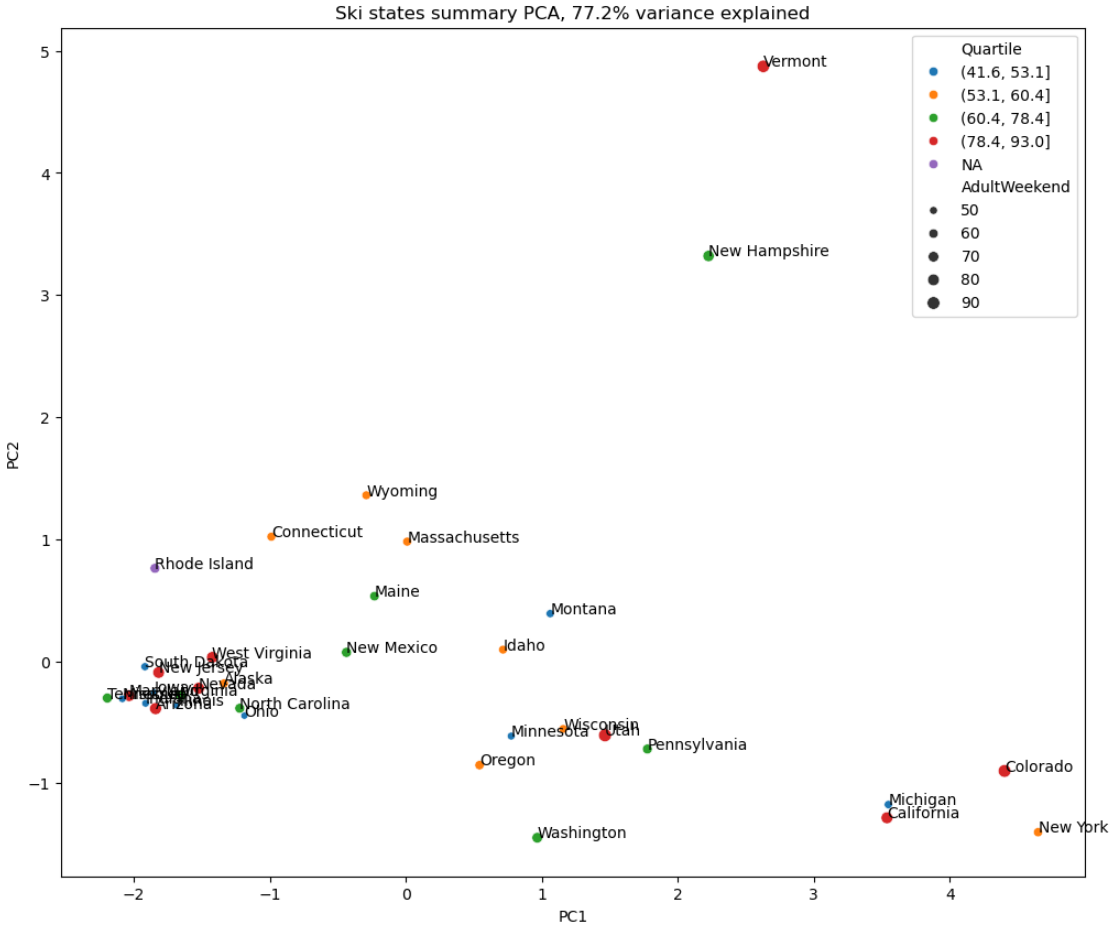
1.3 Exploratory Data Analysis (EDA)

EDA focused on understanding pricing patterns and the factors that differentiate low, mid and high-priced resorts.

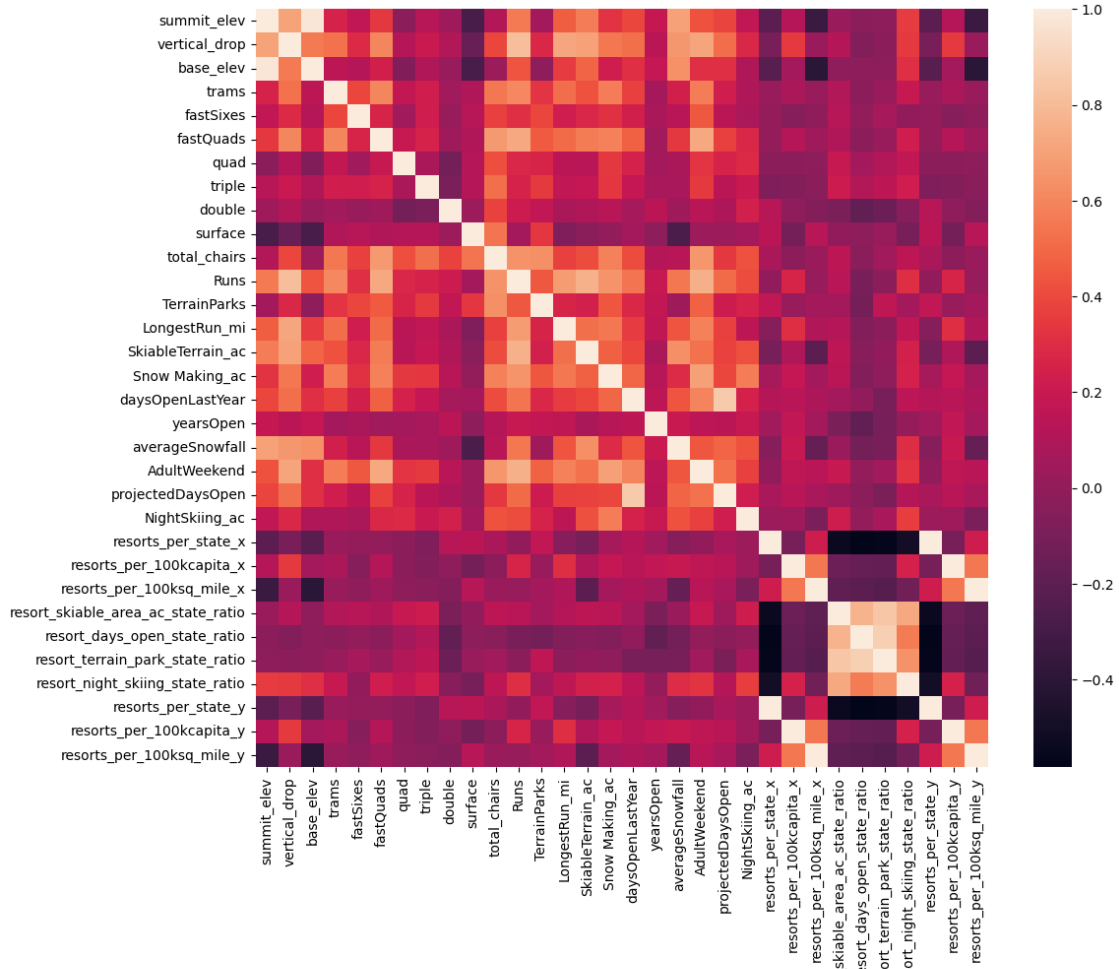
1.4 Key insights:

- Resorts with larger skiable terrain, more chairlifts, and higher facility scores generally charge more.
- Big Mountain already outperforms many competitors in facilities and terrain.

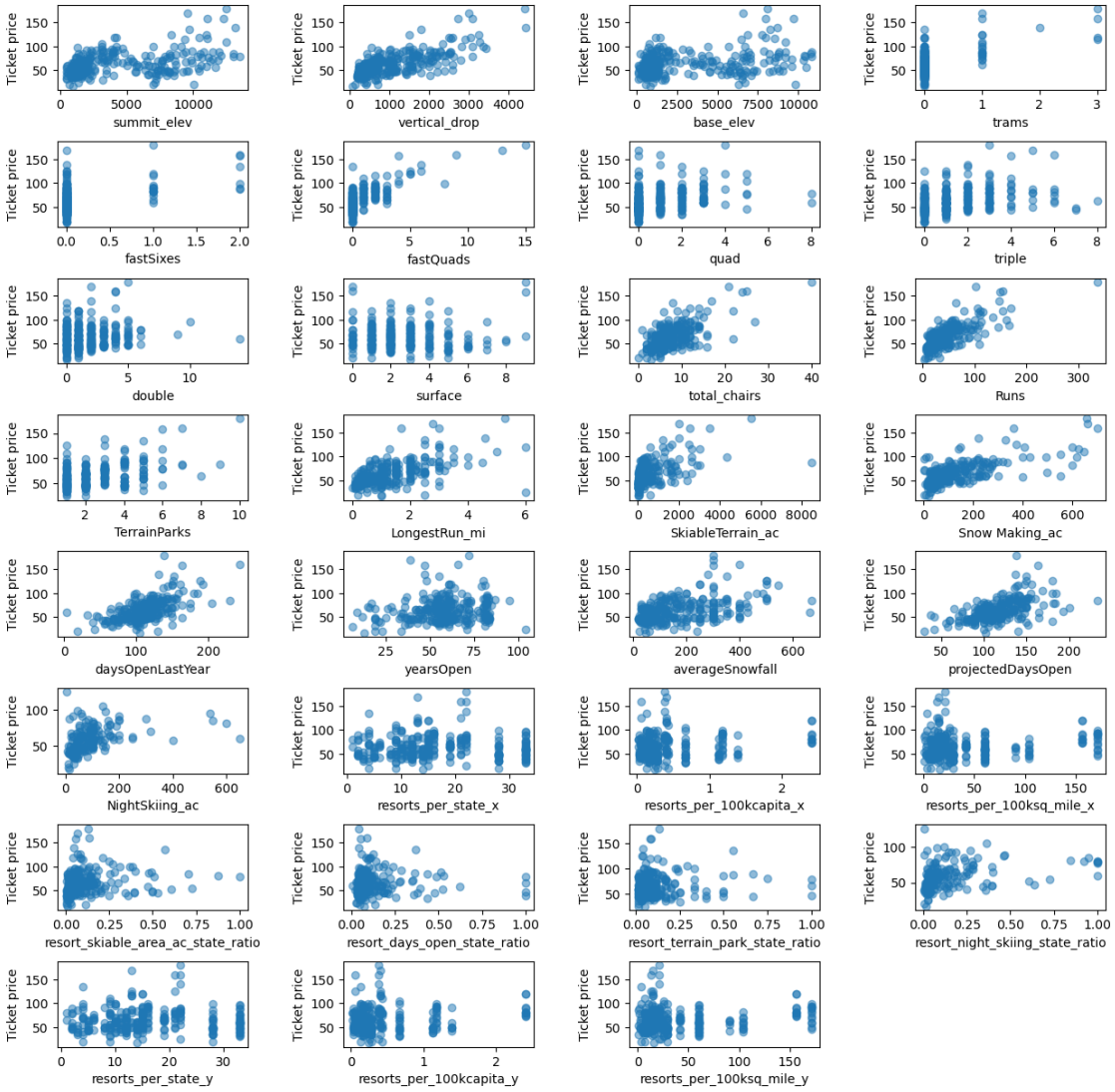
1.5 Distribution of Ticket Prices



1.6 Correlation Heatmap



1.7 Big Mountain vs Comparable Resorts



1.8 Model Preprocessing and Feature Engineering

To prepare for modeling:

- Numerical variables were scaled using StandardScaler.
- Categorical fields (if any existed) were encoded.
- Engineered features included:
  - Terrain per lift ratio
  - Total facility index
  - Operational cost per acre

Cross-validation was used to ensure generalizable estimates.

1.9 Algorithms and Evaluation Metrics

Two main models were tested

1. Linear Regression

- Provided interpretability and directional clarity of features.
- Metrics used: MAE, RMSE, R^2
- Cross-validation showed moderate performance with some underfitting.

2. Random Forest Regressor

- Captured nonlinear relationships better.
- Requiring the tuning of depth, number of trees, and sample splits.
- Outperformed linear models on both validation and test data

1.10 Winning Model and Scenario Modeling

The Radom Forest Regressor was selected as the winning model due to:

- Highest predicitive accuracy
- Consistency between cross-validation and test-set performance
- Scenario Modeling:

Two scenarios were evaluated:

- Base case (current resort features)
- Improved case with the new chairlift and updated facilities The improved scenarion predicted a significantly higher optimal ticket price than what Big Mountain currently charges.

1.11 Pricing Recommendation

Based on the winning model's predicition:

- Big Mountain's current price is below the predicted optimal price relative to its competitors.
- After including the operational cost of the new lift and improved amenities, the model estimate a new optimal ticket price range of approximately (85.51, 106.23)

This recommended price:

- Keeps the resort competitive
- Reflects the increased value offered
- Covers new operational costs while staying aligned with market expectations.

1.12 Conclusion

The analysis shows that Big Mountain is undervalued relative to similar resorts. Data-driven modeling indicates that after upgrades, the resort can safely increase its ticket price without reducing competitiveness. A carefully adjusted pricing strategy will enhance revenue while remaining fair to customers based on the upgraded experience.

1.13 Future Scope of Work

To strengthen future decisions, the following extensions are recommended:

- Gather additional cost data (maintenance, staff, equipment depreciation).
- Integrate customer satisfaction or demand elasticity data.
- Build a time-series model for season-specific pricing.
- Conduct A/B testing for dynamic pricing strategies.
- Expand the dataset to include geographic factors and competitor seasonal