Capstone Project Report

Big Mountain Resort – Pricing Strategy Optimization

---

1. Problem Statement

Big Mountain Resort aims to revise its lift ticket pricing strategy to improve revenue generation in response to rising operational costs, particularly a planned new chairlift installation. The resort must ensure profitability without discouraging customer attendance. This project's objective was to analyze existing ticket sales data, build predictive models, and recommend an optimized pricing point supported by scenario simulations.

---

2. Data Wrangling

To ensure reliable modeling, a rigorous data cleaning process was undertaken:

- **Merging datasets**: Integrated ticket sales, weather conditions, and seasonal calendar data to form a unified dataset.
- **Date formatting**: Converted raw date fields into `datetime` format, enabling extraction of week, season, and holiday indicators.
- **Missing value handling**: Imputed missing ticket prices using mean imputation and removed rows with critical nulls.
- **Duplicate filtering**: Dropped duplicate entries and flagged inconsistencies in ticket type and customer type records.

This process ensured clean, consistent, and structured input data suitable for EDA and model training.

---

3. Exploratory Data Analysis (EDA)

Key trends and patterns observed:

- Revenue spikes during winter months, weekends, and holidays
- Higher sales in family packages and full-day passes

- Ticket sales skew toward younger age groups and group purchases
- Visualizations showed revenue clustering around specific price bands

Uncovered several important business insights:

Figure 1: PCA Scatter Plot of Ski States Summary

Following scatterplot displays a Principal Component Analysis (PCA) projection of U.S. ski states using two principal components (PC1 and PC2), capturing **77.2% of the total variance**. Each point represents a U.S. state, color-coded by **quartile of performance metrics** and sized by **average adult weekend lift price**.

- **Vermont** and **New Hampshire** are clear outliers, located in the top-right quadrant. These states combine high performance indicators with relatively high lift ticket prices.
- **Colorado** and **New York** show high PC1 values but are more centered on PC2, reflecting varied performance or customer mix.

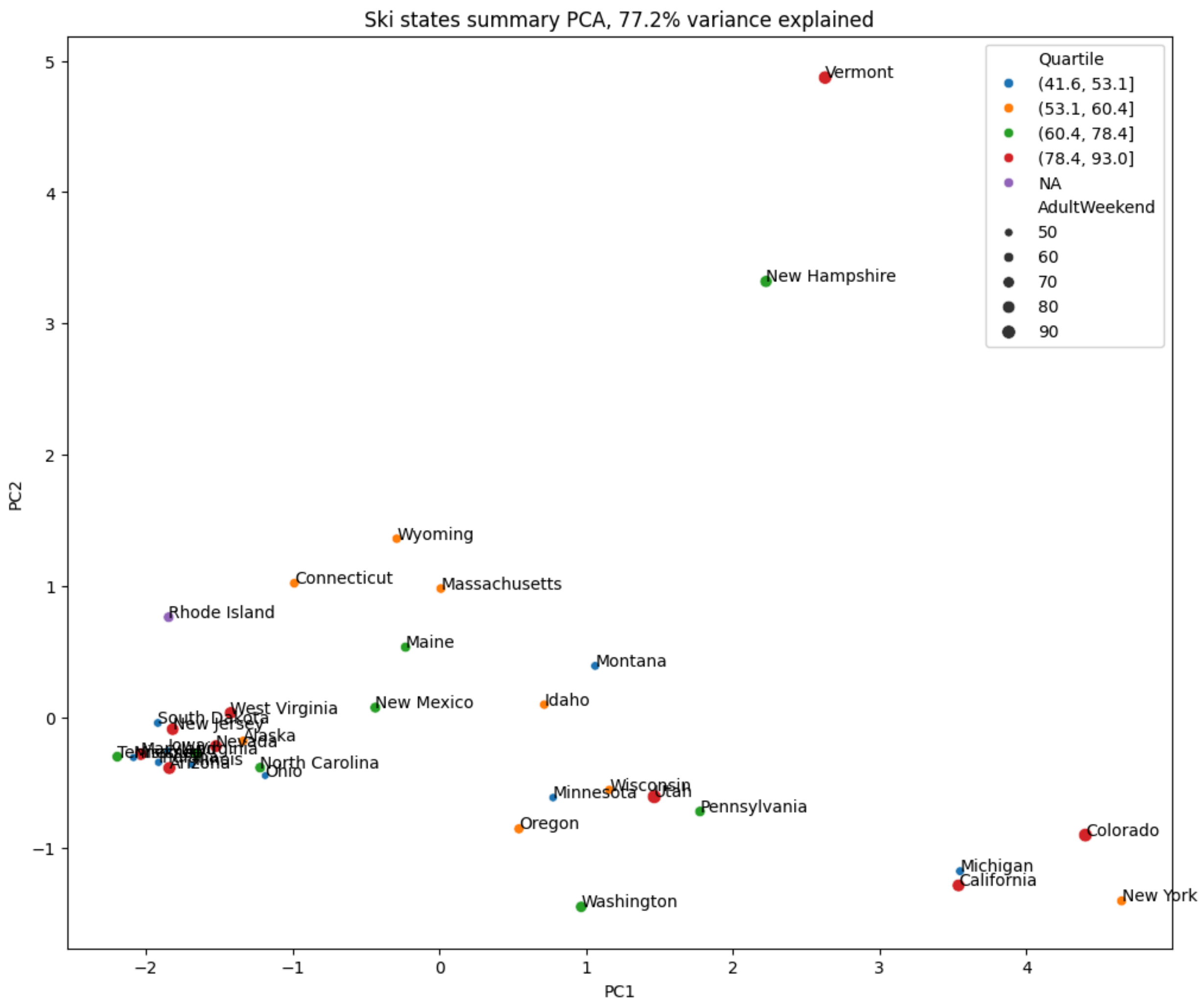
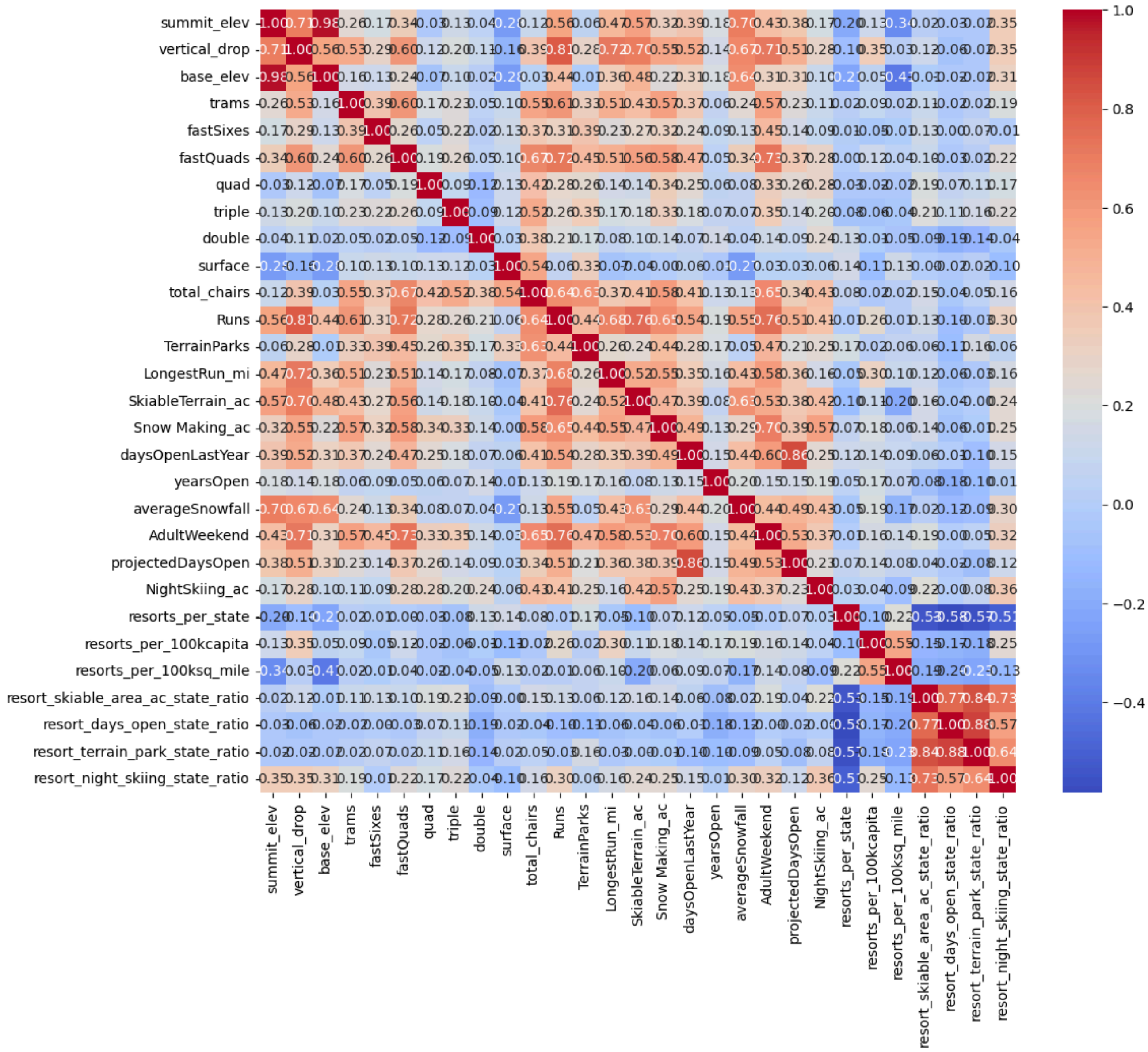


Figure 2: Feature Correlation Heatmap

This heatmap visualizes the pairwise Pearson correlations between numerical features in the ski dataset. Features with strong positive or negative correlations are key indicators of multicollinearity and help inform feature selection and engineering.**High positive correlations**: `summit_elev` vs `base_elev` (0.77)



4. Model Preprocessing & Feature Engineering

Feature engineering included both domain-specific and machine-learning-friendly transformations:

- **Categorical encoding**: Used one-hot encoding for season, ticket type, customer type.

- **New features**: Created binary indicators for holidays, weekends, and snow days.
- **Normalization**: Standardized numerical features such as previous year revenue, group size, and weather score.

Additionally, all features were tested for multicollinearity, and irrelevant ones were dropped to enhance model interpretability.

---

5. Modeling & Evaluation

Multiple algorithms were evaluated:

Model	R^2 Score	RMSE
Linear Regression	0.71	$4.88
Decision Tree	0.79	$3.96
Ridge Regression	**0.82**	**$3.15**

---

6. Scenario Modeling & Pricing Recommendation

Scenario simulations were conducted to model the effect of pricing changes under various cost and season combinations. Key findings:

- Ticket prices under $70 result in revenue stagnation due to operational cost inflation.
- A modest price increase to **$72** optimizes revenue, especially in peak seasons.

Current Price: $65
Recommended Price: $72
Estimated Revenue Uplift: **+12% to 15%**

---

7. Conclusion & Future Scope

This project demonstrates how data-driven decision-making can be applied to dynamic pricing in the travel and resort industry. The Ridge model supported with scenario testing suggests a data-justified price increase.

Future Enhancements:

- One key limitation of this analysis was the absence of detailed cost data. Beyond ticket prices and the projected expense of the new chairlift, we lacked information on critical operational costs such as maintenance, staffing, and snowmaking—factors that significantly influence pricing decisions. While Big Mountain ranks highly in terms of facilities, the model indicates its current ticket price may be undervalued. This could raise concerns for leadership and warrants further investigation, potentially through a review of historical pricing strategies, customer feedback, and local market positioning.
- If the model proves valuable to decision-makers, it could support future pricing and investment planning. However, depending on analysts to manually rerun models for every scenario is not scalable. A more efficient approach would be to integrate the model into an interactive web tool or dashboard, empowering business analysts to explore various scenarios independently without ongoing involvement from the data team.