

Guided Capstone Project Report

Tak Hin Tsang

Problem Statement

Big Mountain Resort in Montana recently installed an additional chairlift to improve visitor distribution across the mountain. While this enhancement increases operational efficiency and customer satisfaction, it also raised the resort's operating costs by \$1,540,000 for the current season. To offset these costs, resort leadership aims to reassess ticket pricing strategies, aligning prices with facility utilization while maintaining or increasing profit margins over the next two years. The core challenge is to quantify the value of Big Mountain's features relative to the broader market and recommend data-driven pricing that reflects both the resort's offerings and market expectations.

Data Wrangling & Preparation

The initial dataset included ~300 North American ski resorts with attributes covering size (skiable terrain), vertical drop, lift infrastructure, runs, snowmaking, region/state, and observed pricing. To ensure a modeling-ready table, I:

- Dropped columns with heavy missingness and excluded rows lacking essential price fields, yielding 277 resorts.
- Corrected evident outliers (e.g., terrain, vertical, operational days) and standardized categorical labels.
- Median-imputed remaining numeric gaps to avoid biased deletions.

This produced a clean, analysis-grade dataset combining the core physical plant (terrain, vertical, chairs, runs) with contextual variables (state, weekday/weekend). Regional and weekend/weekday structure in prices is visible in the state × ticket-type distribution (Fig. 1).

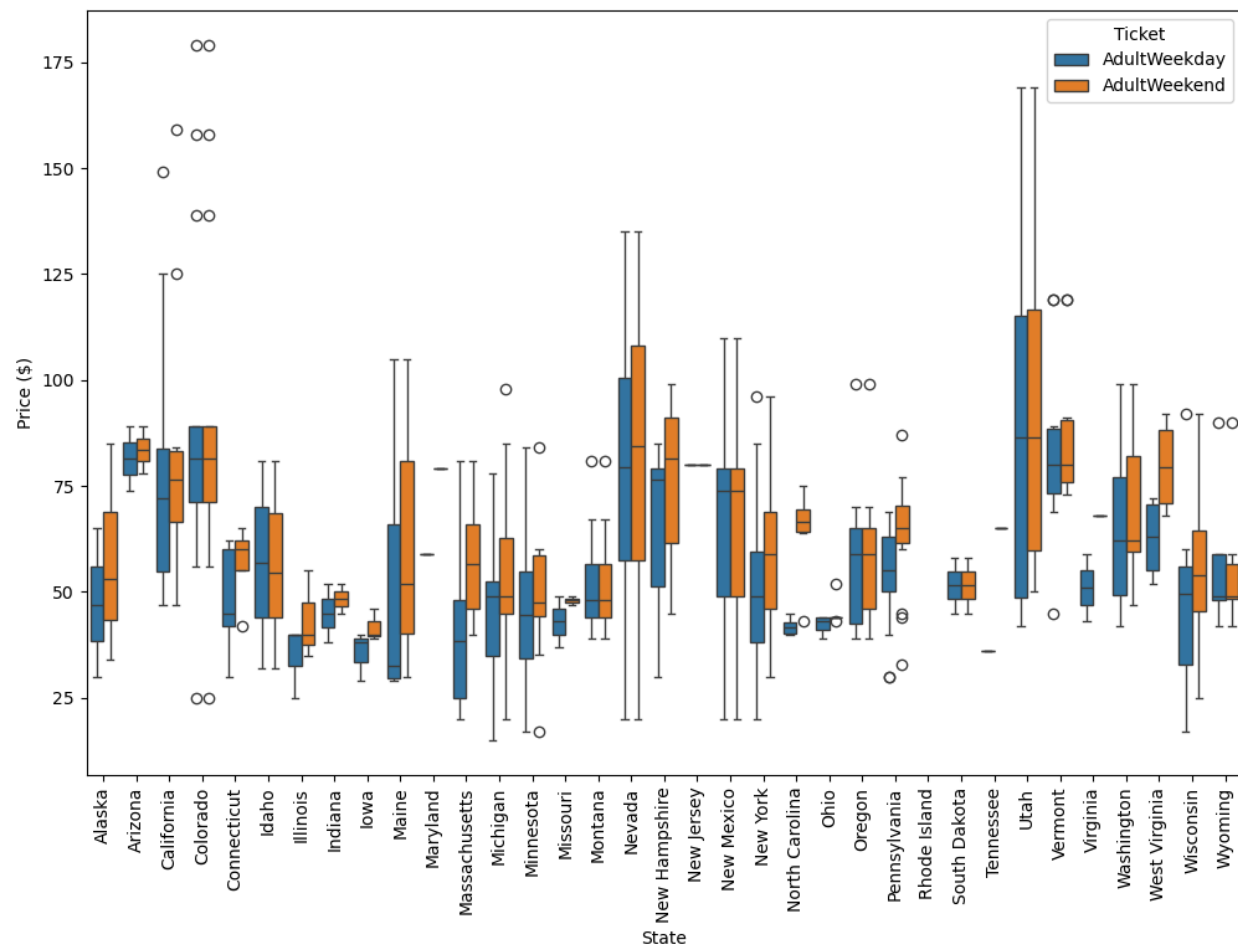


Figure 1. Boxplot of ticket prices by state with weekday/weekend split - *02_data_wrangling.ipynb*

Exploratory Data Analysis

Two consistent patterns emerge. First, weekend prices exceed weekday prices across states; second, Western states trend higher overall (Fig. 1). Correlation structure reinforces economic intuition: skiable terrain, vertical drop, snowmaking acreage, and total chairs are the strongest price-related numeric features (Fig. 2). These patterns guided feature selection and interactions for modeling, focusing on signals of scale, challenge, and capacity.

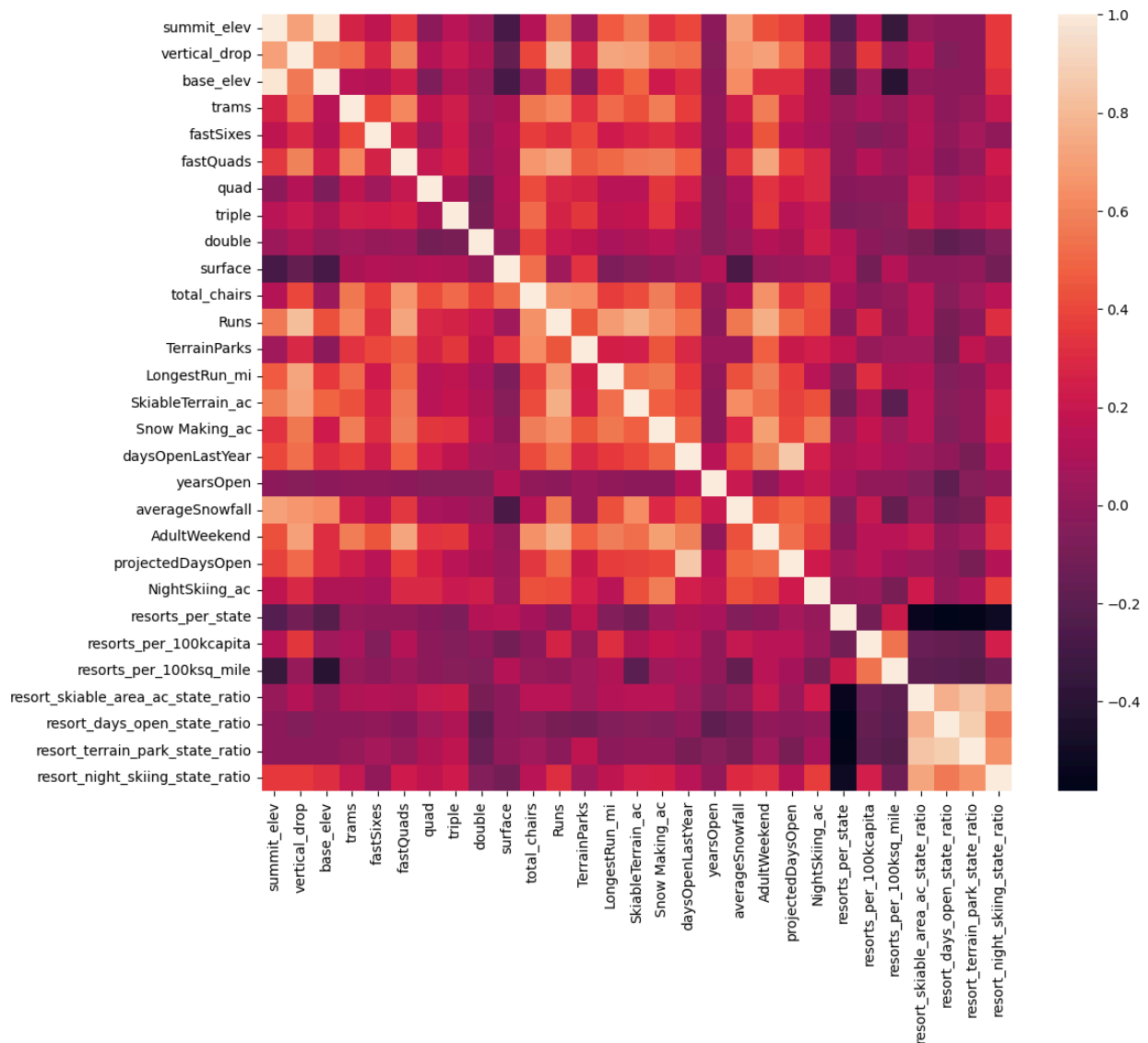


Figure 2. Correlation heatmap - 03_exploratory_data_analysis.ipynb

Model Preprocessing & Feature Engineering

Categorical variables (e.g., state) were one-hot encoded. Numeric features were kept in their native, interpretable units, and we added interaction terms that proxy congestion and capacity (e.g., lifts, terrain). Data were split 80/20 (train/test) with 5-fold cross-validation on train to set hyperparameters and estimate generalization. I used a stratified train/validation split with cross-validation to control variance. This setup keeps the model explanatory enough for pricing governance while capturing key nonlinearities.

Algorithms & Evaluation

I compared linear/regularized baselines with tree-based methods. Random Forest Regression performed best, capturing interactions and modest nonlinearities while remaining stable. Evaluation on hold-out data produced $R^2 \approx 0.89$ and MAE $\approx \$6.37$, which is accurate enough to support price-band decisions at typical weekend price levels. The model's feature importances mirror EDA: skiable terrain, vertical, snowmaking, and total chairs dominate (Fig. 3), reinforcing face validity.

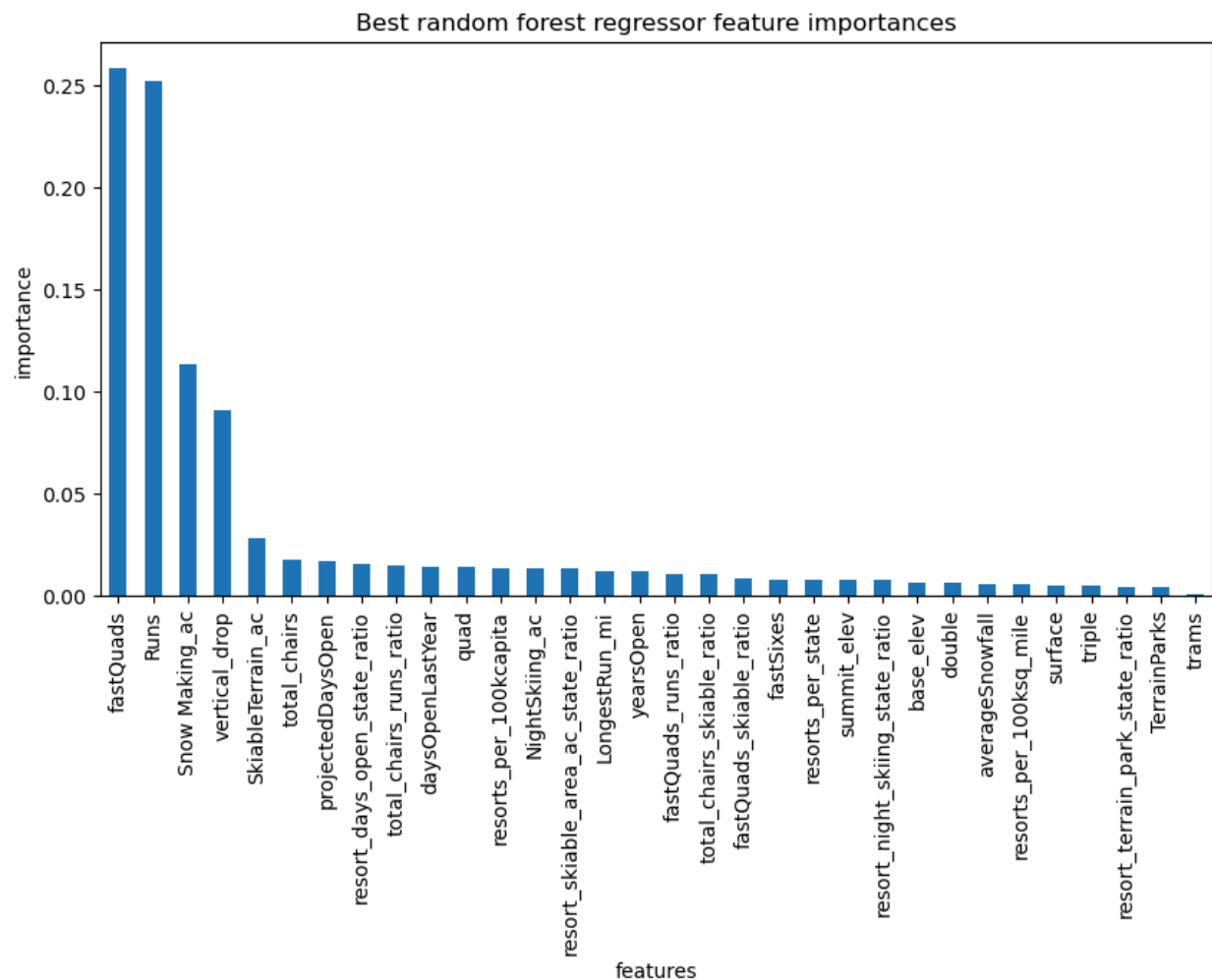


Figure 3. Random Forest feature importances - *04_preprocessing_and_training.ipynb*

Winning Model & Scenario Insights

The Random Forest indicates Big Mountain's feature mix supports a higher adult weekend price than current levels. To make the implications concrete for operators, I used the model to assess a capacity-reduction scenario (runs closed) and computed the associated predicted price and revenue deltas. The side-by-side view quantifies how operational changes translate into pricing room and revenue sensitivity—useful for both proactive planning and day-of adjustments (Fig. 4).

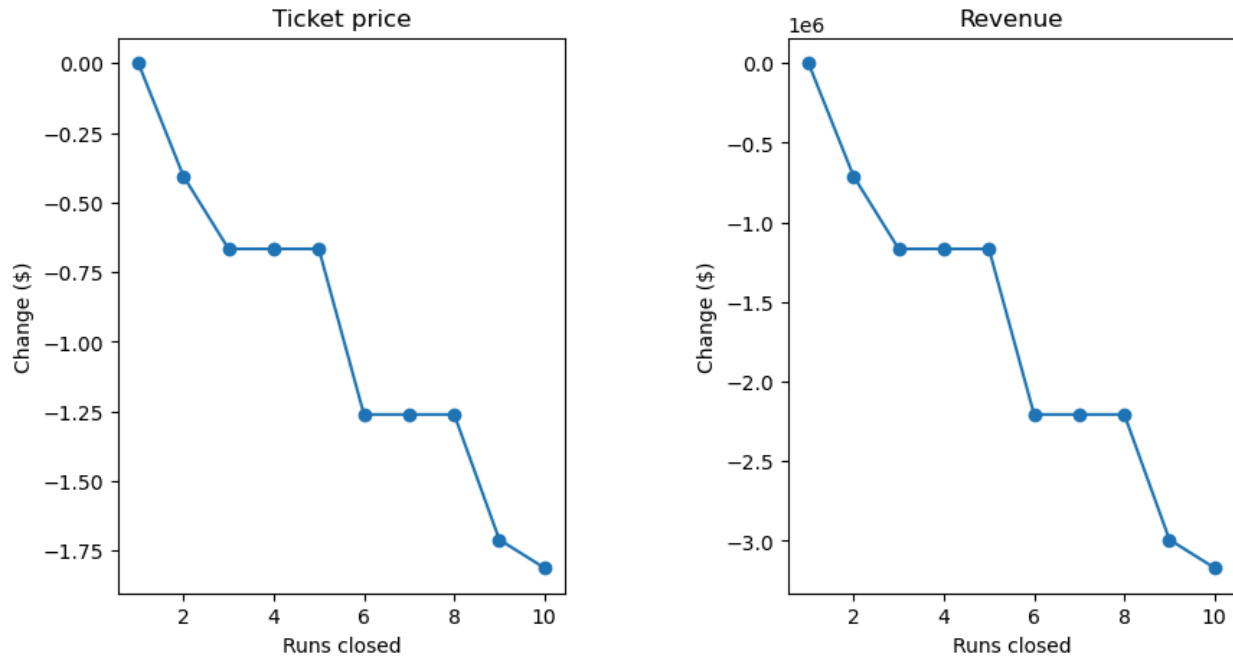


Figure 4. Runs-closed scenario — price and revenue deltas - *05_modeling.ipynb*

In addition to scenarios, market positioning matters. A distributional comparison of adult weekend prices across all resorts, with Big Mountain marked, shows that the current \$81 weekend price sits below comparable peers—consistent with the model's signal that the resort is underpriced (Fig. 5).

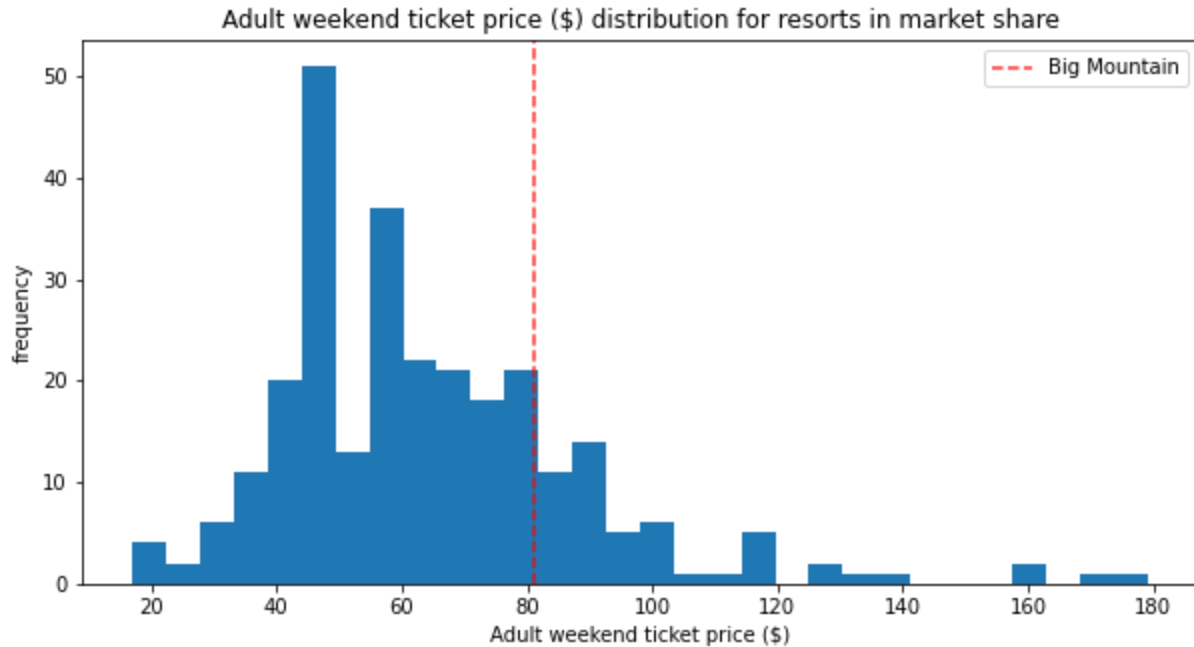


Figure 5. Adult weekend ticket price distribution (all resorts) with Big Mountain marked - *05_modeling.ipynb*

Pricing Recommendation

Adopt an adult weekend base price of ~\$95.87. Maintain the existing weekday differential and layer targeted premiums (e.g., holidays, exceptional conditions) within guest-experience and capacity guardrails. This recommendation is grounded in:

- **Market position:** current weekend price is below comparable resorts (Fig. 5).
- **Drivers of willingness to pay:** terrain, vertical, snowmaking, and chairs align with higher price levels (Figs. 2-3).
- **Operational resilience:** scenario sensitivity quantifies revenue impacts when conditions vary (Fig. 4).

Under typical demand, this policy meaningfully reduces the \$1.54M incremental-cost gap over two seasons without eroding demand. Re-estimate monthly during peak and quarterly otherwise; if forecast/operational inputs are missing, fall back to prior period's setting.

Conclusion

Big Mountain's weekend price under-recognizes the market value of its terrain and lift infrastructure. A move to ~\$95.87, coupled with situational premiums, better aligns price with value and strengthens revenue resilience. The modeling evidence supports both the magnitude and the manageability of this change.

Future Scope

Enhance granularity with dated ticketing/visitation and richer weather feeds; integrate capacity/queue telemetry to refine congestion features; and run controlled A/B tests of weekend/holiday ladders. Over time, automate scenario sweeps in a lightweight dashboard so analysts can recommend daily prices within governance guardrails.