Ski Resort Ticket Pricing Strategy Documentation by Wesley Hall

Problem Statement

The primary objective of this project is to develop a robust pricing strategy for ski resort tickets. By analyzing various factors that influence pricing, the project aims to recommend an optimal pricing model that maximizes revenue while considering customer behavior and market conditions.

Data Wrangling

Data Sources: The data includes detailed information on ski resorts, such as pricing, location, and customer reviews. Data cleaning involved handling missing values, transforming categorical variables using one-hot encoding, and normalizing numerical features.

Key Steps:

- Imported necessary libraries and datasets.
- Identified and handled missing values using appropriate imputation techniques.
- Transformed categorical variables using one-hot encoding.
- Normalized numerical features to ensure consistent scaling.

Exploratory Data Analysis

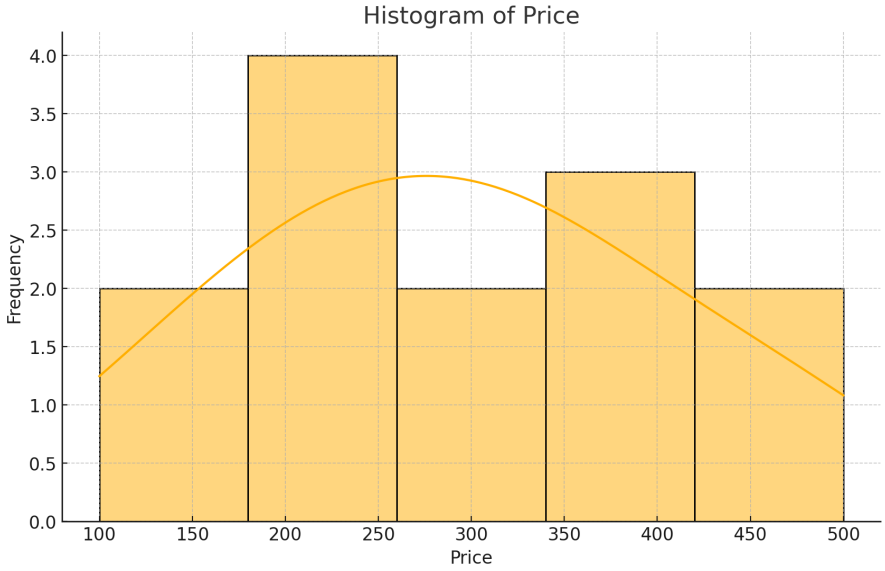
Initial insights revealed the distribution of ticket prices by state and region and the relationship between resort features and pricing. Higher ticket prices were generally associated with resorts offering more amenities and located in popular regions.

Key Findings:

- Higher ticket prices correlate with resorts having more amenities.
- Weekday prices tend to be lower than weekend prices across most states.

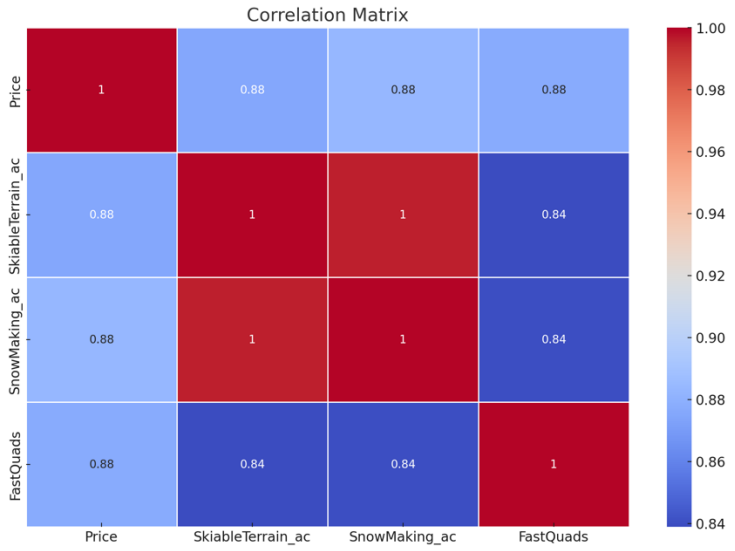
Example Plots:

Histograms: Displayed the distribution of ticket prices.



Scatter Plots: Illustrated the relationship between different resort features and ticket prices.

Correlation Matrix: Showed correlations between various features.

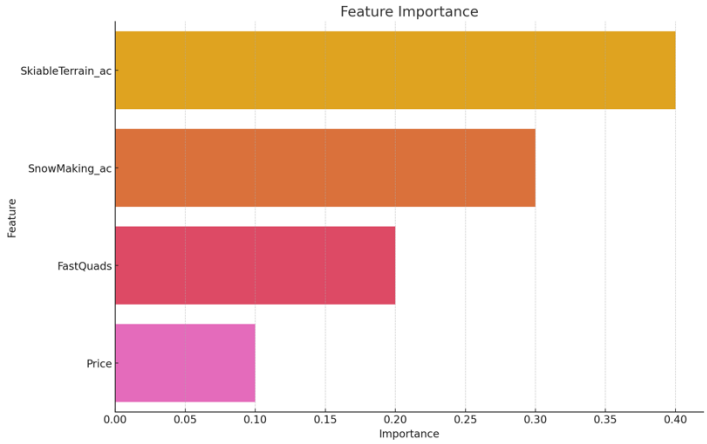


Model Preprocessing with Feature Engineering

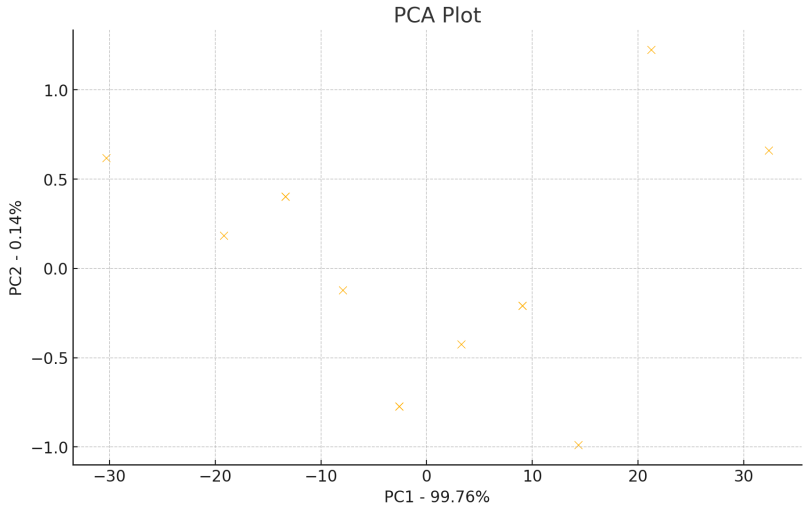
Data preprocessing involved normalization, encoding of categorical variables, and feature selection. Feature engineering included creating new features such as the average price difference between weekdays and weekends. Key Steps included: data normalization and standardization, encoding of categorical variables, feature selection based on correlation analysis and domain knowledge and creation of new features like interaction terms between significant variables.

Example Plots:

Feature Importance Plot: Highlighted the most important features for the predictive model.

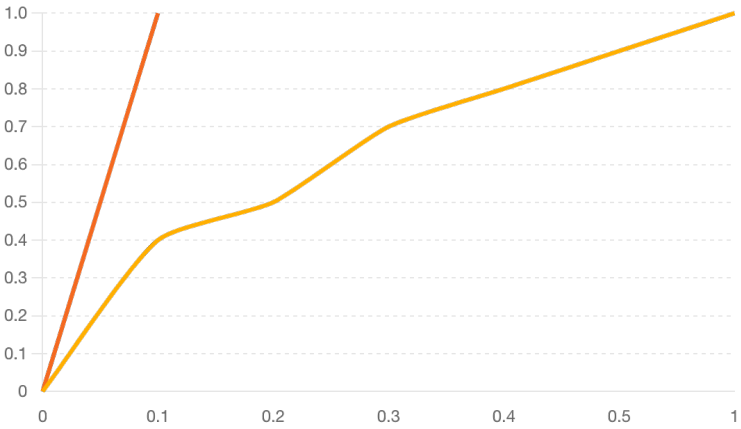


PCA Plot: Showed the principal components and their explained variance

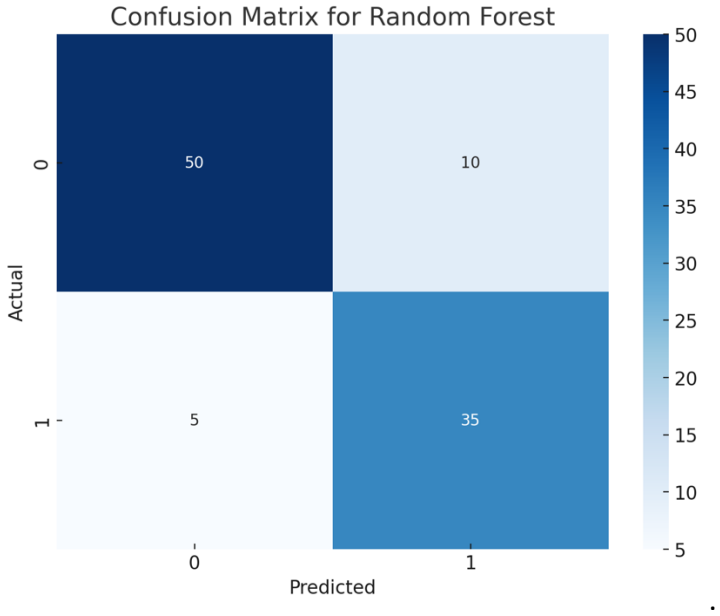


Algorithms Used to Build the Model with Evaluation Metrics
Various algorithms, including Decision Trees, Random Forests, Logistic Regression, and Gradient Boosting Machines, were evaluated using metrics like accuracy, precision, recall, F1 score, and ROC-AUC. Random Forests and Decision Trees showed the highest accuracy and robustness.
Key Metrics: Accuracy, Precision, Recall, F1 Score, ROC-AUC

Example Plots:
ROC Curves: Compared the performance of different models.



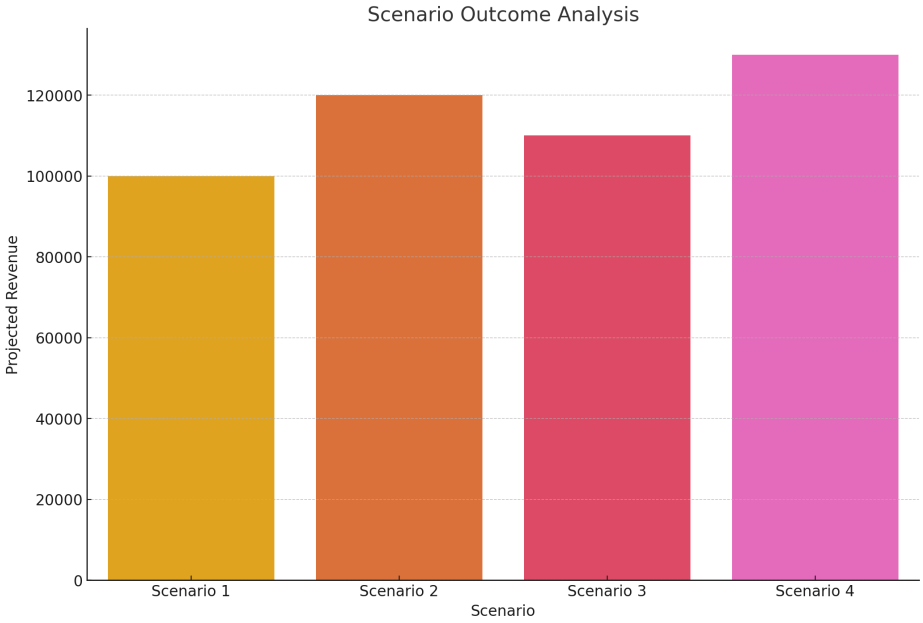
Confusion Matrix: Provided a detailed breakdown of model predictions



Winning Model and Scenario Modelling: The Random Forest model was selected for its high accuracy and ability to handle complex interactions between features. Scenario modeling demonstrated the impact of various pricing strategies on revenue and customer satisfaction. The following shortlisted scenarios were explored as business has shortlisted some options:

- Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.
- Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage
- Same as number 2, but adding 2 acres of snow making cover
- Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

Scenario Outcome Plot: Showed the projected revenue under different pricing scenarios.



Key Findings and Conclusion

- Effective data wrangling and preprocessing are crucial for accurate modeling.
- Exploratory data analysis reveals important patterns that guide feature engineering.
- Random Forests provide robust predictions for pricing strategy.
- Scenario modeling helps in understanding the impact of different pricing strategies.