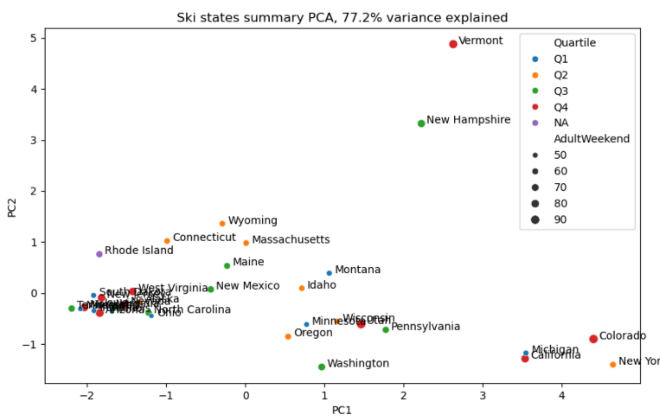
Guided Capstone Project Report

Problem statement: Big Mountain Resort in Montana boasts spectacular National Park views, 105 trails, and approximately 350,000 annual skiers or snowboarders. A recent chairlift addition, intended to improve visitor distribution across the mountain, has increased operating costs by $1.54 million this season. Recognizing limitations in their current premium-above-average pricing approach, the resort acknowledges the need to capitalize on its unique facilities and optimize investment strategies based on facility importance. To address this, Big Mountain desires to develop a new, competitive, data-driven ticket pricing model to offset the increased operating cost reflecting the true value and explore opportunities for cost-saving measures or even higher-priced value propositions.

Data Wrangling: Data provided by Alesha Eisen, the database manager, included information on 330 ski resorts across the USA. After cleaning the dataset by removing resorts with missing ticket prices and unhelpful features, it was refined to 277 observations with 25 initial features. Additional data from the internet was incorporated to create features based on density relative to area and population, resulting in a total of 36 features.

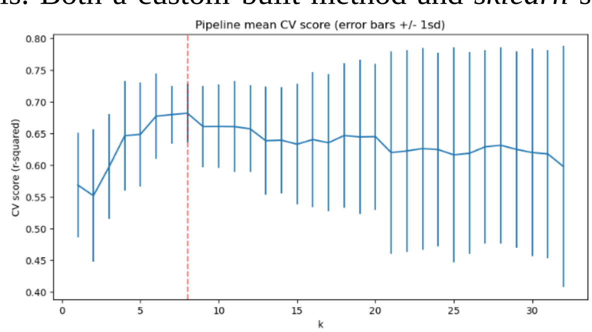


Exploratory data Analysis: The **Principal Component Analysis (PCA)** revealed no discernible pattern or clustering of states within different ticket price quartiles. This finding indicated that states could be treated uniformly, obviating the need for state-specific analysis in subsequent steps.

Model Preprocessing with feature engineering: Subsequent analysis identified features strongly correlated with ticket prices using heatmaps, correlation tables, and scatter plots. Features like '*Runs*', '*fastQuads*', '*vertical_drop*', and '*Snow Making_ac*' showed the highest correlations, making them key factors for potential price adjustments. This was further supported by the **Random Forest Regression Model**'s feature importance analysis. The data was then split into a Train set (70%) and a Test set (30%) to ensure the model remains uninfluenced by the test data, allowing for an accurate evaluation of the model's expected performance on future data.

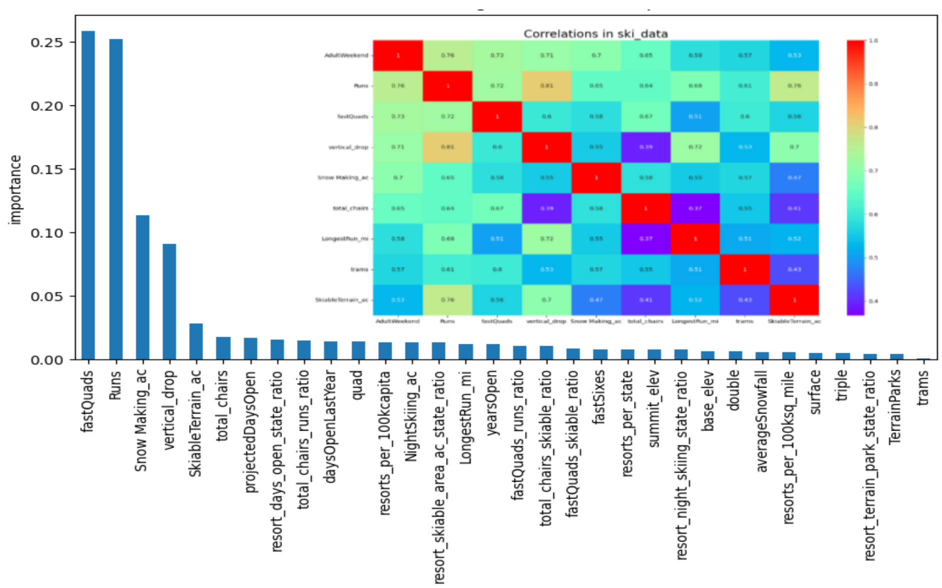
Algorithms used to build the model with evaluation metric:

i. A **baseline model** considered the average ticket prices as the predicted price for Big Mountain, serving as a comparison for later models. Both a custom-built method and *sklearn's DummyRegressor* were used, revealing an average prediction error of approximately $19.



ii. A **linear regression model (LRM)** was then developed. Initially, missing features were filled manually using mean and median values, before transitioning to *sklearn*'s *SimpleImputer* for automated imputation. The LRM showed significant improvement with a mean absolute error (MAE) of $9 compared to the baseline's $19. *GridSearchCV*, performing *cross-validation* with five folds, suggested using k=8 features in the LRM. This tool also identified features with the greatest impact on ticket pricing, consistent with earlier findings.

iii. The **Random Forest Regression Model (RFRM)** demonstrated a nearly $1 lower MAE in cross-validation compared to the LRM, with less performance variability. Evaluation on the test set confirmed that the RFRM's performance was consistent with *cross-validation* results.



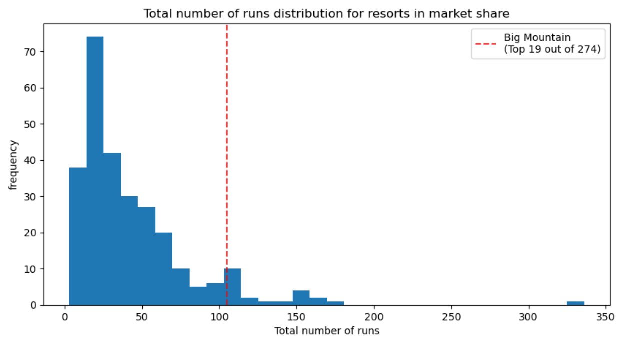
Winning model and scenario modelling: Based on the analysis, the RFRM is the best model for predicting the Big Mountain ticket price. To address the current operating cost increase, four different scenarios within Big Mountain's scope such as permanently closing down up to 10 of the least used runs, increasing the vertical drop by 150 ft by adding a run and a chairlift with and without any addition of snowmaking cover, and just increasing the longest run by 0.2 miles to reach 3.5 miles with an additional 4 acres of snowmaking coverage.

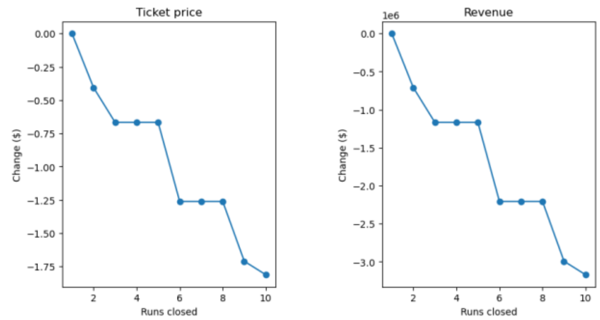
Pricing recommendation:

i) We recommend the **Big Mountain** to increase the ticket prices to $95.87 from current price of $81.00. This will significantly increase their revenue (approx. $ 26.0 M) and is completely supported by the market pressure. However, I would suggest implementing price hikes incrementally to gauge their impact before further adjustments.

ii) The analysis suggests that a $ 1.99 increase in ticket price is feasible, which could generate an additional $ 3.47 M over the season if the resort increases the vertical drop by 150 feet. However, implementing this change necessitates an expansion of the vertical drop, which requires installing a new chair lift with an operating cost of $ 1.54 M for the season. Additionally, an extra run is needed, although the associated operating cost for this has not been provided.

iii) **Big Mountain** ranks second out of eleven resorts in **Montana** and nineteenth out of 277 resorts nationally in terms of the number of runs. This strong positioning allows for potential reductions in the number of runs without impacting its competitive standing locally or nationally.





Conclusion: Our analysis of **Big Mountain**'s features compared to other ski resorts nationally and within **Montana** indicates strong performance. The accompanying table highlights its favorable national ranking and top position within **Montana**, underscoring the rationale for increasing the adult weekend ticket price to better align with competitive benchmarks. Additionally, there are scenarios where the resort could potentially enhance facilities at minimal cost or consider shutting down underutilized facilities.

Features	Rank in USA	Rank in Montana
AdultWeekend	52	1
vertical_drop	29	5
trams	23	1
Runs	19	2
total_chairs	14	1
fastQuads	14	1
LongestRun_mi	12	1
Snow Making_ac	7	1
SkiableTerrain_ac	5	1

Future scope of work: This data provided to us is very useful in determining the possible linkage between the available features within resorts and target price and do not need to collect any more data for the existing features. However, there are a few limitations to consider:

i) *Limited Scope of Data*: The data only considers internal factors while determining the ticket price. It does not consider customers' backgrounds, such as their annual income, the distance they need to travel to reach the resort, and other relevant factors.

ii) *Assumption of Market Pressures*: Our model assumes that all other resorts are following market pressures when determining their ticket prices, similar to our model. This may not be accurate, as there is no mention of data reflecting their profitability.

iii) *Cost Data Omission*: The data does not include information on the features' installation and maintenance costs, which are crucial for making informed decisions on where to invest for profit optimization.