Guided Capstone Project Report-Big Mountain Resort

Introduction

Big Mountain Resort, is a ski- resort with around 3,50,000 customers a year and management is committed to provide customers better experience. The company has recently added an additional chair lift to improve distribution of visitors which resulted in additional operating expenses of $1,540,000 and management wants to recover the cost along with generating additional revenue to justify the investment strategy.

Problem Statement

How can Data Science be leveraged at Big Mountain Resort to implement data-driven business strategy, and henceforth realize higher ticket price by showcasing better facilities & implement cost cutting measures, to generate additional profit higher than $1,540,000 in one year to recover additional operational costs and investments.

Scope

Only data available for analysis is coming from SQL databases in a single CSV file and consists of key features of different ski resorts. It will be used to identify important attributes and generate insights about how the ticket prices can be adjusted to impact additional revenues

Constraints within solution space

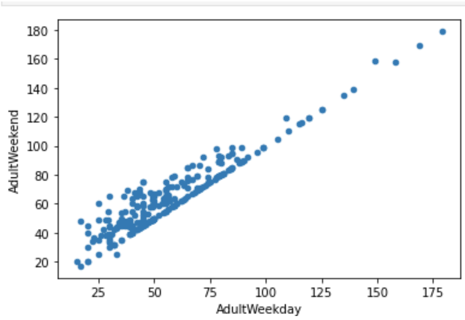
- Only one csv file available for data analysis and modeling
- The data does not provide information about total visitors to the resorts in a year
- It does not provide any measurable metrics about the visitors during weekdays and weekends to see the impact of pricing

Dataset

Original dataset consists of 330 rows and 27 features. Out of 27 features 3 were categorical and remaining 24 were numeric.

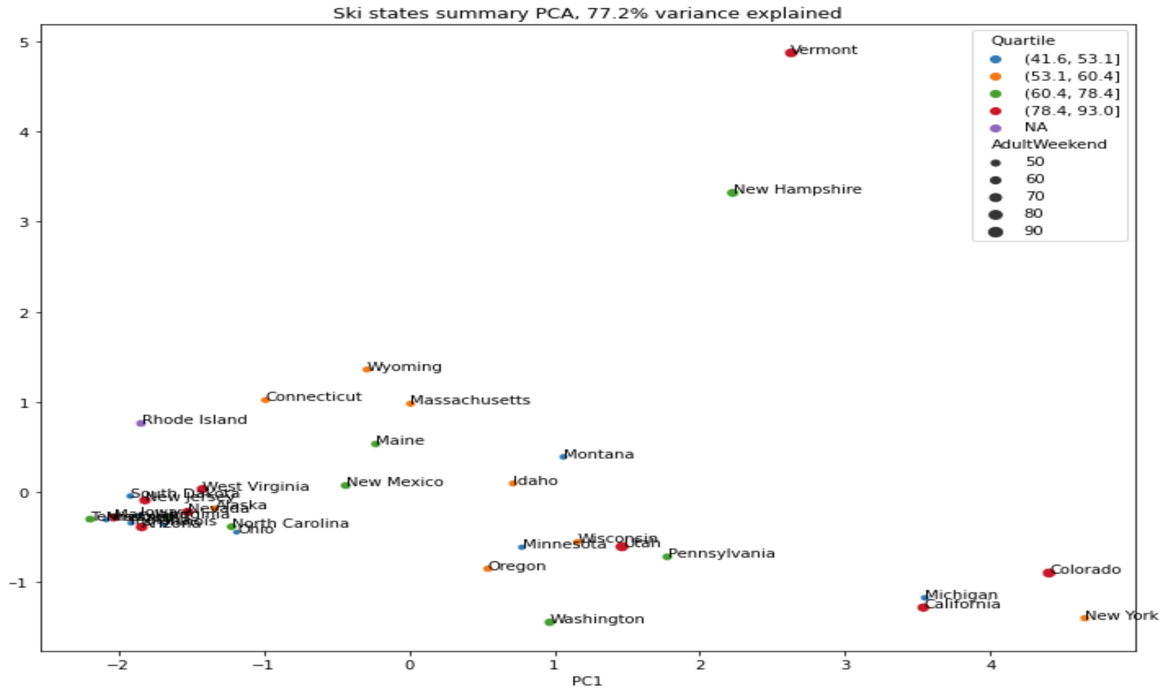
Data Wrangling

During this process, data is checked for missing values, type of data and data discrepancies. The fast Eight feature has 50.30% missing values and that column is dropped. Additionally the resorts without any price information were also dropped. Further, Adult Weekend had more missing values and showed linear relationship with Adult Weekday prices, it is decided to keep Adult Weekend Price as our target variable and AdultWeekday column is dropped

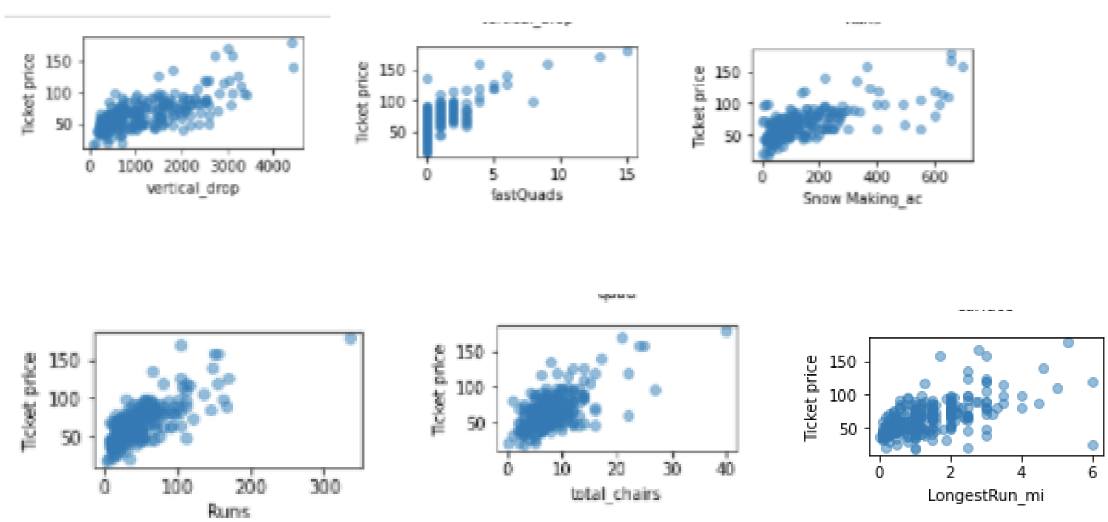


During the process we also added state population, area data from wikipedia[1] and made a state_summary dataframe. The dataset after wrangling consisted of 277 rows and 25 columns.

Exploratory Data Analysis
PCA component analysis is performed on state summary data and PCA1 & PCA2 alone explained over 75% of the variation. Average ticket prices for states were explored with PCA1 and PCA 2 components and it was found that state specific features are not influencing the ticket prices in a meaningful way.

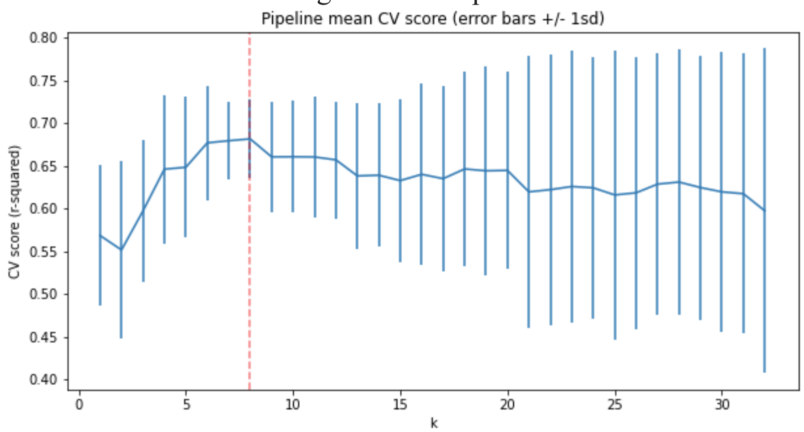


State summary data frame is merged with ski data dataframe and added 10 more features showing resort features relative to their respective states. Scatter plots of features relative to price indicated very strong positive correlation with vertical drop (0.71), fastQuad(0.73), Run (0.76),total chair (0.65), LongestRun Mile (0.58) and Snowmaking(0.70)

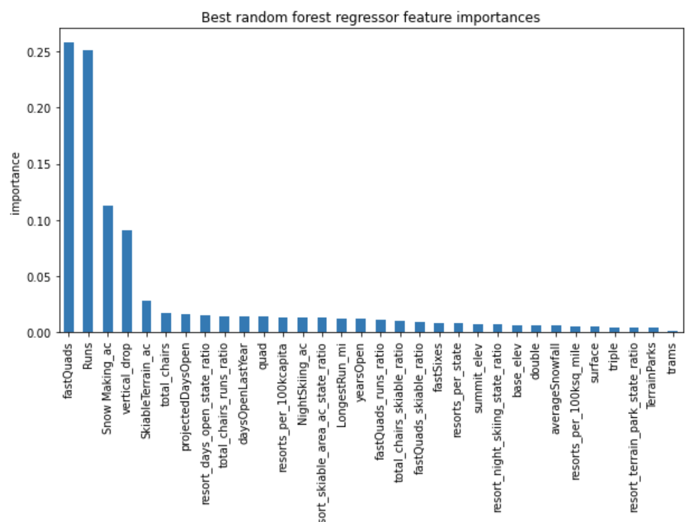


Preprocessing & Training Data

Dummy Regressor (mean),Linear Regression & Random Forest Regressors are evaluated for training and testing data. On the test data,Random Forest Regressor with ("selectkbest__k" and "simpleimputer__strategy': 'median") performed the best with lowest mean absolute error value of 9.53, compared to 11.79 for Linear Regression & 19.13 for Dummy Regressor. The CV scores plot below indicates that k=8 features give the best performance.



Top 5 features based on feature importance for Random Forest Regressor models are fastQuad, Runs,Snow Making_ac ,Vertical Drop,Skiable Terrain_ac, & total chairs



Modeling

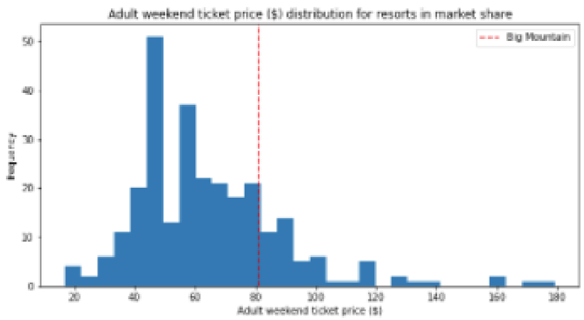
Big Mountain Resort is evaluated for important features in the market context. Big Mountain Resort ticket price is on the higher end of the price spectrum at $81, but the suggested price based on the model is $95.87. A ticket price of $95.87 is sufficient to generate additional revenue of $26,022,500

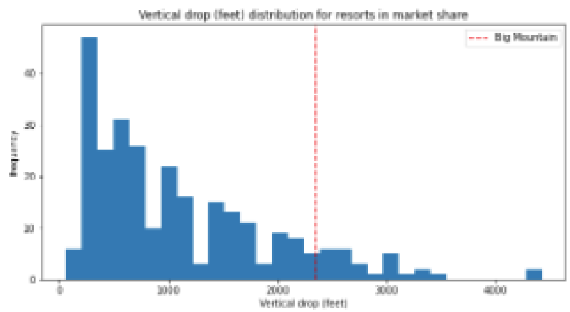
New Price

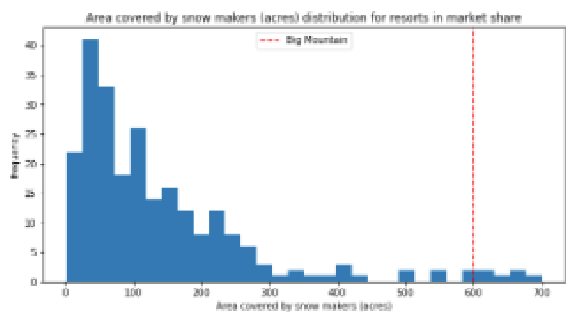
```
print(model.predict(X_bm))
```

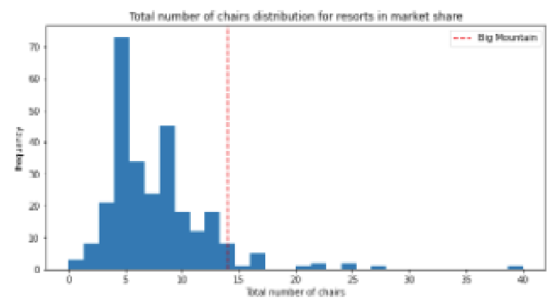
```
[95.86942029]
```

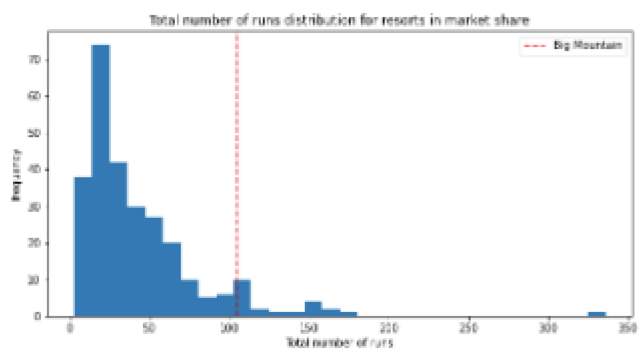
The higher price can be attributed to the features for Big Mountain Resort, which are in the top end of the market place depicted by the red dotted line.





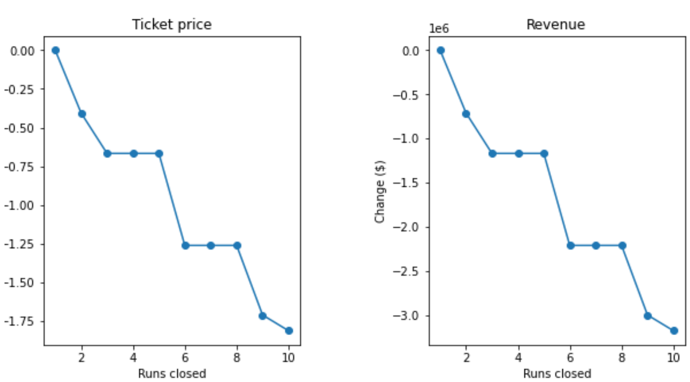






Scenario Analysis Assumptions:
Four different business scenarios assumptions based on 3,50,000 visitors per year & 5 tickets per visitor, were evaluated by plugging the modified values in the model.

Scenario 1. Permanently closing down up to 10 of the least used runs.



- Ticket prices are expected to drop gradually when the first 3 runs are closed, but no major difference in ticket price for the next 2 drops. This scenario leads to reduction in ticket price

Scenario 2. Increase the vertical drop by adding a run to a point 150 feet lower down & installation of an additional chair lift to bring skiers back up

- This scenario supports an increased ticket price of $8.61 and annual increased revenue of $15,065,471

Scenario 3. Increasing the vertical drop by 150 ft by adding a run, along with adding 2 acres of snow making.

- This scenario supports an increased ticket price of $9.90 and annual increased revenue of $17,32,2717

Scenario 4:Increasing the longest run by 0.2 miles by increasing 4 more acres of snow making cover.

- No increase in price

Conclusion

Random Forest Regressor with 8 features provided the best model performance with lowest mean square error. The base model without any modifications to the features supports ticket price of $95.87 with MAE of $10.39, which definitely builds a case for increase in ticket prices. In perspective, $95.87 ticket price is enough to generate additional revenue of $26.02 Million which is much higher than additional operational cost of $15.4 million and easily covers the additional operational cost Scenario 2 and Scenario 3 can be considered for future changes after evaluating capital and operational costs associated with the changes.

Future Work

Data showing the number of visitors for each resort can be collected to have a better grasp over price elasticity of demand. Additionally, the capital and operational costs associated with different features would be extremely useful in generating more useful insights