Guided Capstone Project Report

Introduction

Big Mountain Resort is a ski resort in Montana. It offers stunning views of Glacier National Park and the Flathead National Forest and access to 105 trails. It attracts approximately 350,000 visitors of all skill levels each year who enjoy skiing or snowboarding.

While Big Mountain Resort continues to invest and maintain a high level of service for its guests, management believes that the resort is not maximizing its profits sufficiently, given its position in the market sector. It also does not have a clear idea of the facilities that matter most to visitors, including those they would be more likely to pay more.

Recognizing the need for change, management is actively seeking advice on how to best price its tickets. It's clear that a new pricing strategy is necessary to maximize profits and better serve their guests.

This project aims to revolutionize the resort's pricing strategy by developing a model capable of predicting ticket prices from a number of data available in ski resorts. The potential for this project to significantly impact the resort's profitability is both exciting and intriguing.

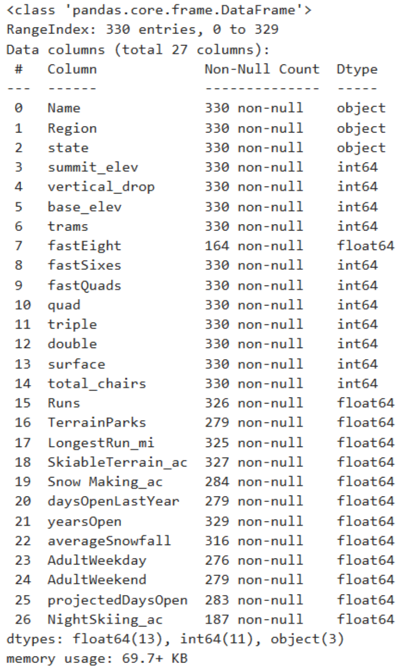
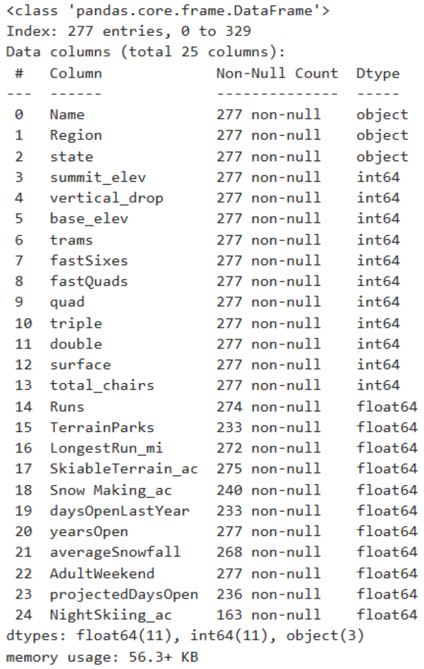
Problem statement

Big Mountain Resort would like to define a new pricing strategy based on data from other ski resorts nationwide. What pricing model can be specified to determine a competitive price for guests accurately?

Data Wrangling

Data wrangling, the process of preparing data for analysis, involves, among other things, importing, loading, exploring, cleaning (deleting or removing), updating, and re-exploring the raw data.

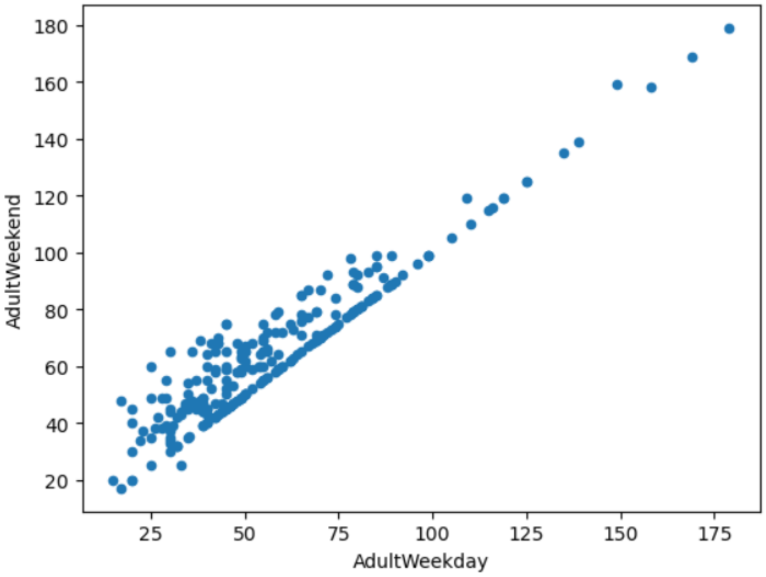
The first steps of auditing the raw data noted a total of 27 columns and 330 rows. These include, among other things, the total vertical drop, the number of chairlifts, the weekday price, the weekend price, and the total number of runs for each resort.

	
Before	After

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 330 entries, 0 to 329
Data columns (total 27 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Name               330 non-null    object 
 1   Region             330 non-null    object 
 2   state              330 non-null    object 
 3   summit_elev        330 non-null    int64  
 4   vertical_drop      330 non-null    int64  
 5   base_elev          330 non-null    int64  
 6   trams              330 non-null    int64  
 7   fastEight          164 non-null    float64
 8   fastSixes          330 non-null    int64  
 9   fastQuads          330 non-null    int64  
 10  quad               330 non-null    int64  
 11  triple             330 non-null    int64  
 12  double             330 non-null    int64  
 13  surface            330 non-null    int64  
 14  total_chairs       330 non-null    int64  
 15  Runs               326 non-null    float64
 16  TerrainParks       279 non-null    float64
 17  LongestRun_mi      325 non-null    float64
 18  SkiableTerrain_ac  327 non-null    float64
 19  Snow Making_ac     284 non-null    float64
 20  daysOpenLastYear   279 non-null    float64
 21  yearsOpen          329 non-null    float64
 22  averageSnowfall    316 non-null    float64
 23  AdultWeekday       276 non-null    float64
 24  AdultWeekend       279 non-null    float64
 25  projectedDaysOpen  283 non-null    float64
 26  NightSkiing_ac     187 non-null    float64
dtypes: float64(13), int64(11), object(3)
memory usage: 69.7+ KB
```

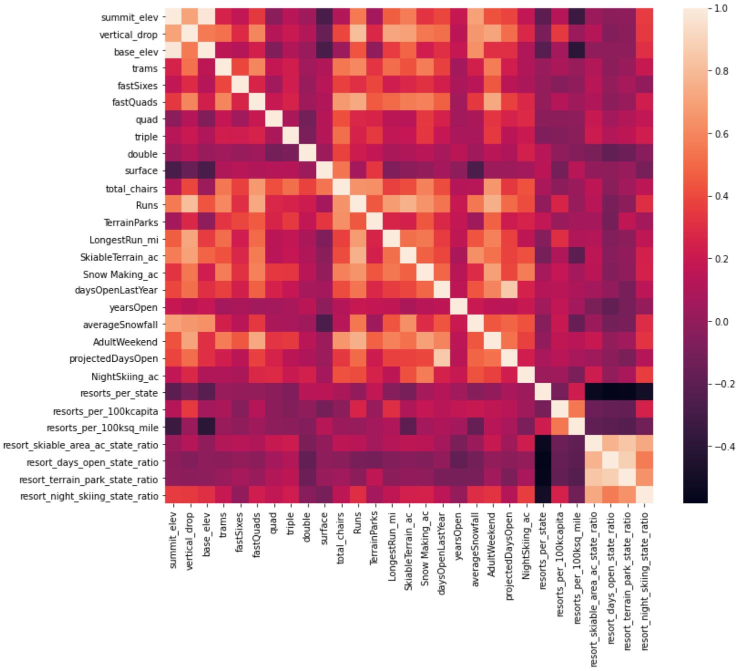
```
<class 'pandas.core.frame.DataFrame'>
Index: 277 entries, 0 to 329
Data columns (total 25 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Name               277 non-null    object 
 1   Region             277 non-null    object 
 2   state              277 non-null    object 
 3   summit_elev        277 non-null    int64  
 4   vertical_drop      277 non-null    int64  
 5   base_elev          277 non-null    int64  
 6   trams              277 non-null    int64  
 7   fastSixes          277 non-null    int64  
 8   fastQuads          277 non-null    int64  
 9   quad               277 non-null    int64  
 10  triple             277 non-null    int64  
 11  double             277 non-null    int64  
 12  surface            277 non-null    int64  
 13  total_chairs       277 non-null    int64  
 14  Runs               274 non-null    float64
 15  TerrainParks       233 non-null    float64
 16  LongestRun_mi      272 non-null    float64
 17  SkiableTerrain_ac  275 non-null    float64
 18  Snow Making_ac     240 non-null    float64
 19  daysOpenLastYear   233 non-null    float64
 20  yearsOpen          277 non-null    float64
 21  averageSnowfall    268 non-null    float64
 22  AdultWeekend       277 non-null    float64
 23  projectedDaysOpen  236 non-null    float64
 24  NightSkiing_ac     163 non-null    float64
dtypes: float64(11), int64(11), object(3)
memory usage: 56.3+ KB
```

At the end of these data inspections, two columns ("fastEight" and "AdultWeekday") were found to have several missing values and were removed. Additionally, many missing values had to be deleted, and a few other columns had to be removed. At the end of this process, 277 of the original 330 rows remained, and the target column for pricing was "AdultWeekday."



Exploratory data Analysis

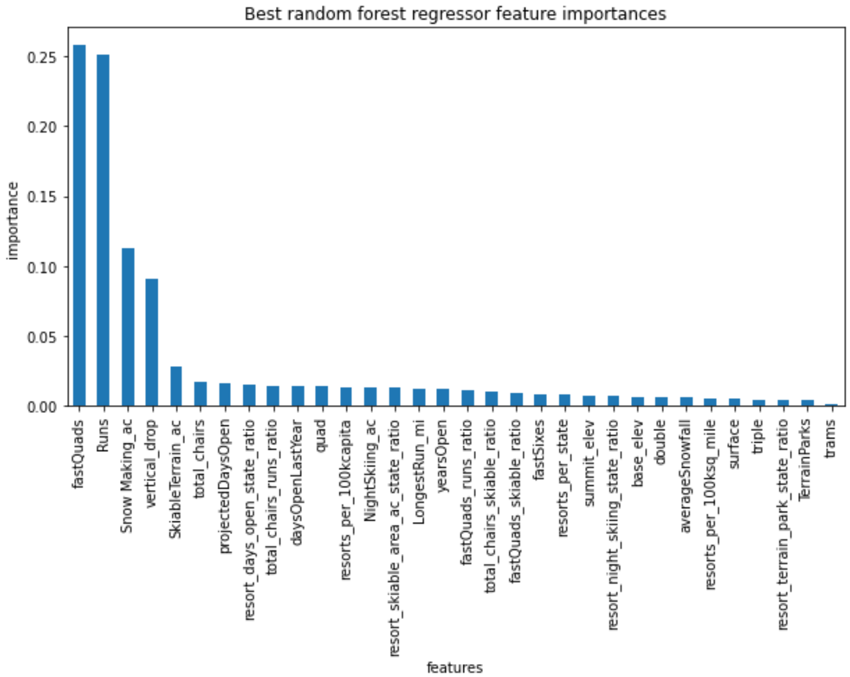
This step aims to find trends, patterns, and actionable insights in the data. With this in mind, the first relationship explored between total resorts per capita and total resorts per area did not reveal any exciting trends for our target resort, Big Mountain Resort. The second relationship explored required a Principal Cumulative Analysis (PCA). It showed that the first two components account for 75% of the variance, and the first four account for 95%. To better understand the relationship between price and components, one created a heat map to visualize the relationship between each feature.



The most significant correlation was found with fastQuads, Runs, and Snow Making_ac. So now, we can use these features to create a model that can determine a new ticket price based on the data.

Model Preprocessing with feature engineering

The initial mean was identified as the “best guess” of the price, and $83.81 was calculated as a baseline to use for comparing other prices. With a mean absolute error of about $19, the mean is not a better solution, as it is much too high. A regression was therefore performed using the median between the results. This time, the mean absolute error was only off by $9. A data pipeline was therefore created for both linear and random regression to efficiently produce identical results and thus facilitate comparisons. However, only the latter one performed better. The subsequent regression using the random forest model clearly shows that imputing the median value allows for estimating the average price of the four components previously identified. Furthermore, the analysis revealed that the vertical drop also plays an important role in determining the ticket price. Once all the components are included in the random forest model, the mean absolute error is only about $1, which is acceptable variability.

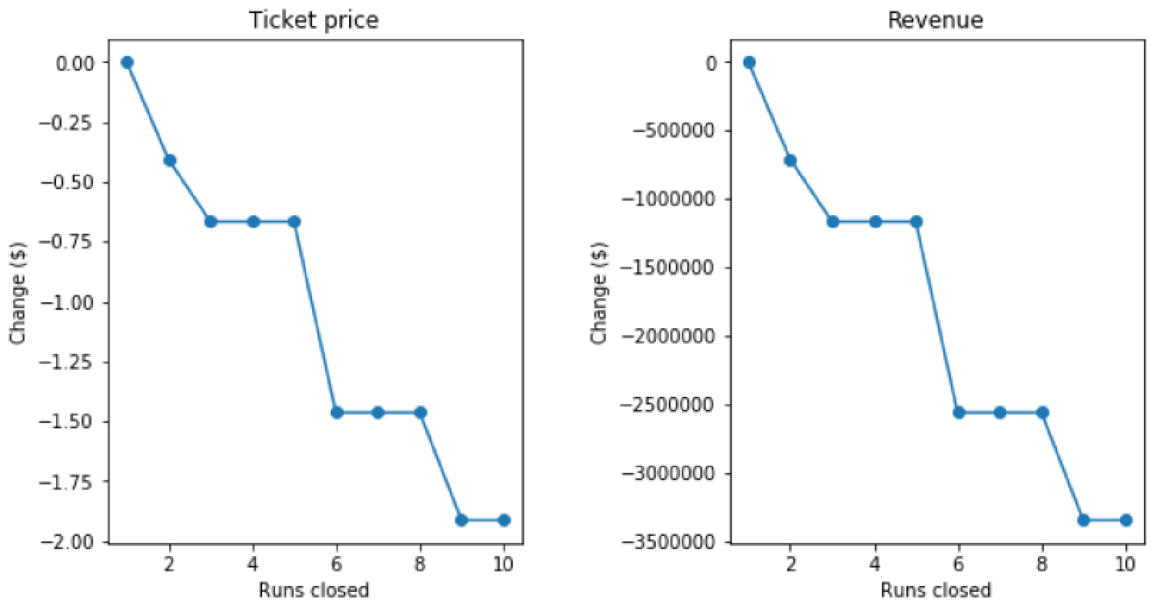


The model was created using the principal components and random forest regression method. Based on the data, it can predict the ticket price. By adding more relevant components, the model can be made as accurate as possible. While Big Mountain Resort's current fare is $81.00, the modeled fare is $95.87, suggesting that there is room for a substantial price increase.

Conclusion

Finally, we see that the resort can do much better in pricing. Seven of eight characteristics favor a ticket price of at least $10. In addition, there are margins for savings, particularly in closing unprofitable runs. The model predicts that keeping up to five runs closed without significantly impacting revenue is possible.

Big Mountain Resort has the opportunity to enhance its revenues while delivering quality services to its guests. We hope these results will enable resort management to make informed decisions that ensure the sustainability of its operations.



Future scope of work

From a perspective, it would be interesting to analyze how the closure of certain runs and their corollary, as well as the reduction of operating costs, can contribute to improving ticket prices. Another promising option is increasing the length of specific runs using some investments in chairlifts, the effects of which must be anticipated.