**Title: Guided Capstone Project Report**

**Report Submitted by: Priyanshi Kathuria**

**Date: 29 March 2024**

**Introduction & Problem Statement:**

My client is Big Mountain Resort, a ski resort located in Montana. Big Mountain Resort offers spectacular views of Glacier National Park and Flathead National Forest, with access to 105 trails. Every year about 350,000 people ski or snowboard at Big Mountain. This mountain can accommodate skiers and riders of all levels and abilities.

These are serviced by 11 lifts, 2 T-bars, and 1 magic carpet for novice skiers. The longest run is named Hellfire and is 3.3 miles in length. The base elevation is 4,464 ft, and the summit is 6,817 ft with a vertical drop of 2,353 ft. Big Mountain Resort has recently installed an additional chair lift to help increase the distribution of visitors across the mountain.

This additional chair increases their operating costs by $1,540,000 this season. The resort's pricing strategy has been to charge a premium above the average price of resorts in its market segment.

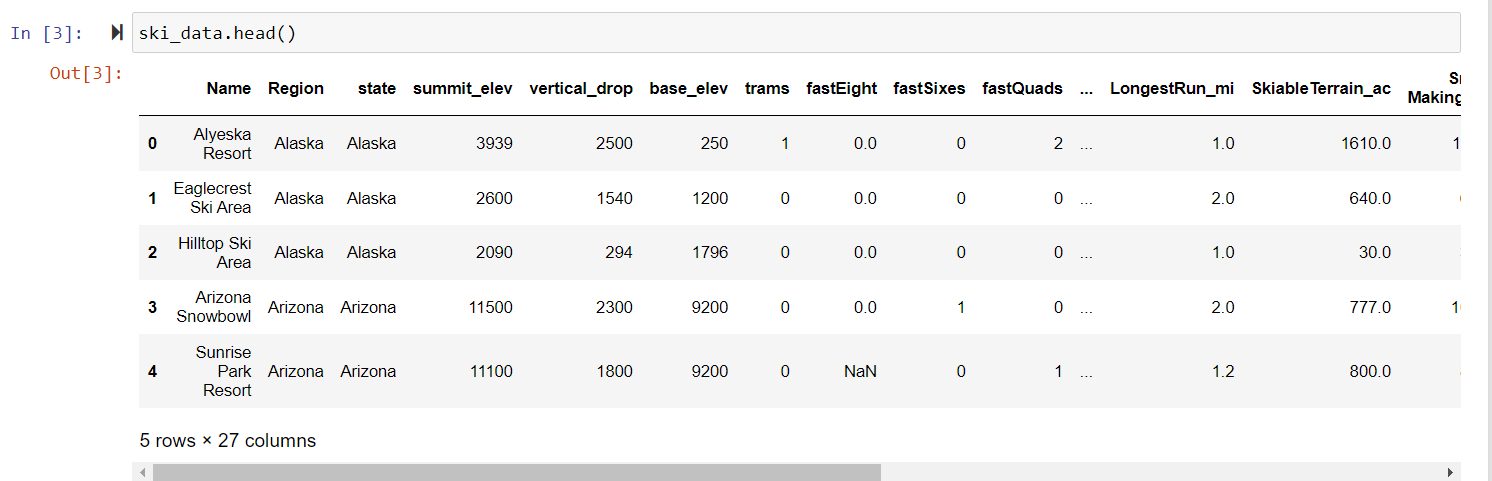
They know there are limitations to this approach. There's a suspicion that Big Mountain is not capitalizing on its facilities as much as it could. Basing their pricing on just the market average does not provide the business with a good sense of how important some facilities are compared to others. This hampers investment strategy.

The business wants some guidance on how to select a better value for their ticket price. They are also considering a number of changes that they hope will either cut costs without undermining the ticket price or will support an even higher ticket price.

In this report, I'll share the insights gained and offer recommendations based on the analysis conducted as part of our guided capstone project.

**Data Wrangling:**

**Our data:**



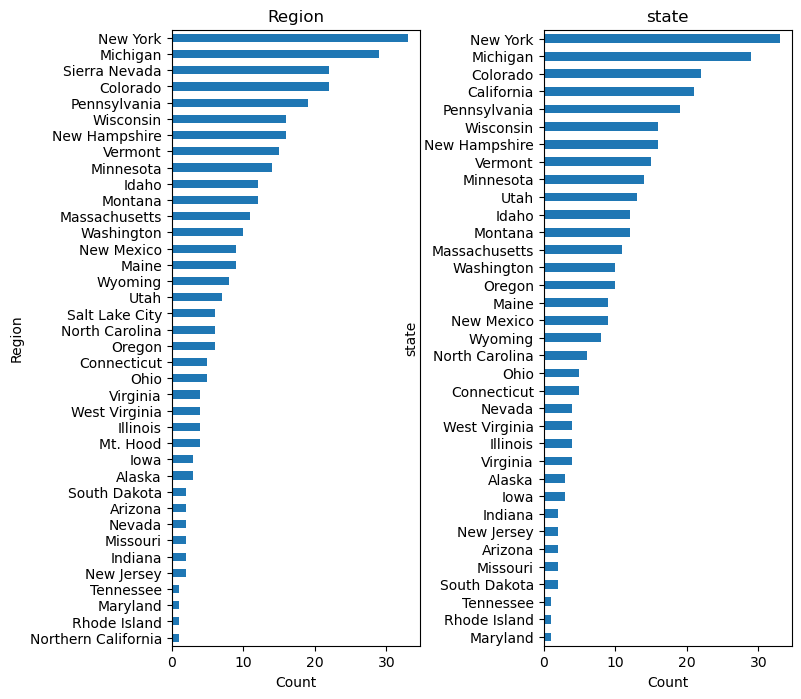
1. While exploring Silverton Mountain information, we see that its skiable terrane was messed up. The value popping up was 26819, but the original value was 1819. We replaced it.

2. We Dropped the 'fastEight' column from ski\_data because half the values are missing and all but the others are the value zero. There is essentially no information in this column.

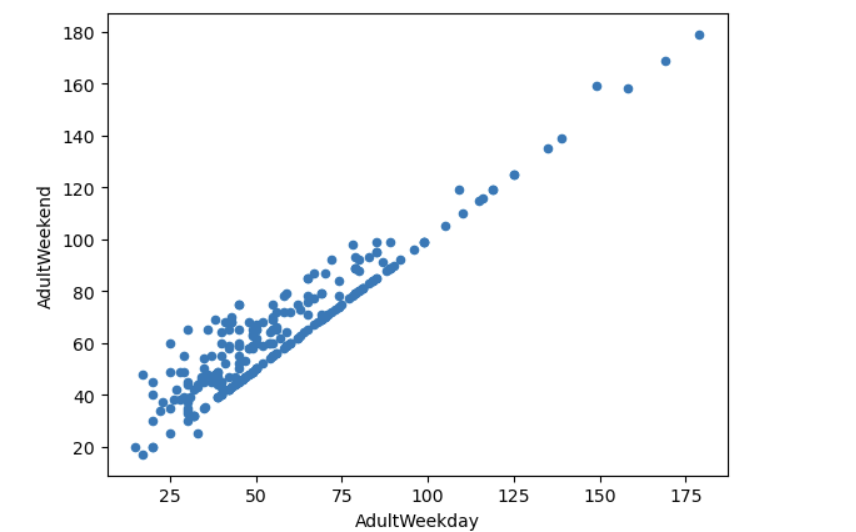
3. We have 4 states missing in our data which we merged later from the Wikipedia data - which were that \_contain\_ 'Massachusetts', 'Pennsylvania', or 'Virginia'

4. We calculated the missing values and % of that for our final data.

**We explored a Distribution Of Resorts By Region And State which landed in 13th place in the USA.**

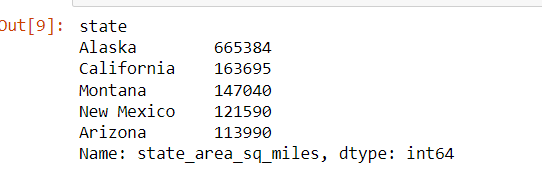


We further studied a scatter plot (x='AdultWeekday', y='AdultWeekend', kind='scatter'). There is a clear line where weekend and weekday prices are equal. Weekend prices being higher than weekday prices seem restricted to sub $100 resorts. We also observed that the distribution for weekday and weekend prices in Montana seemed equal.

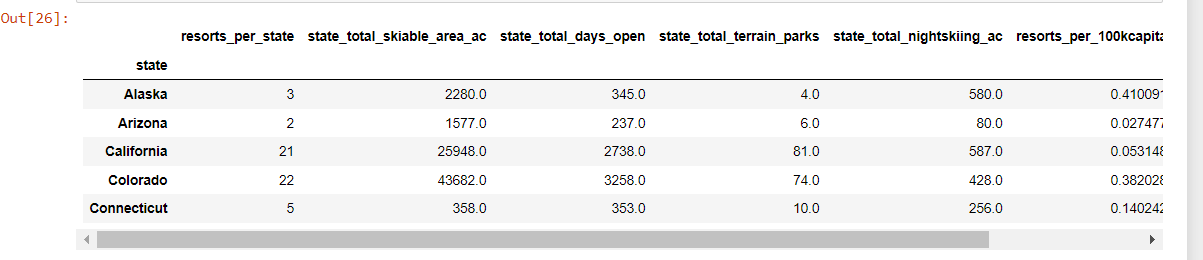


**Exploratory Data Analysis**

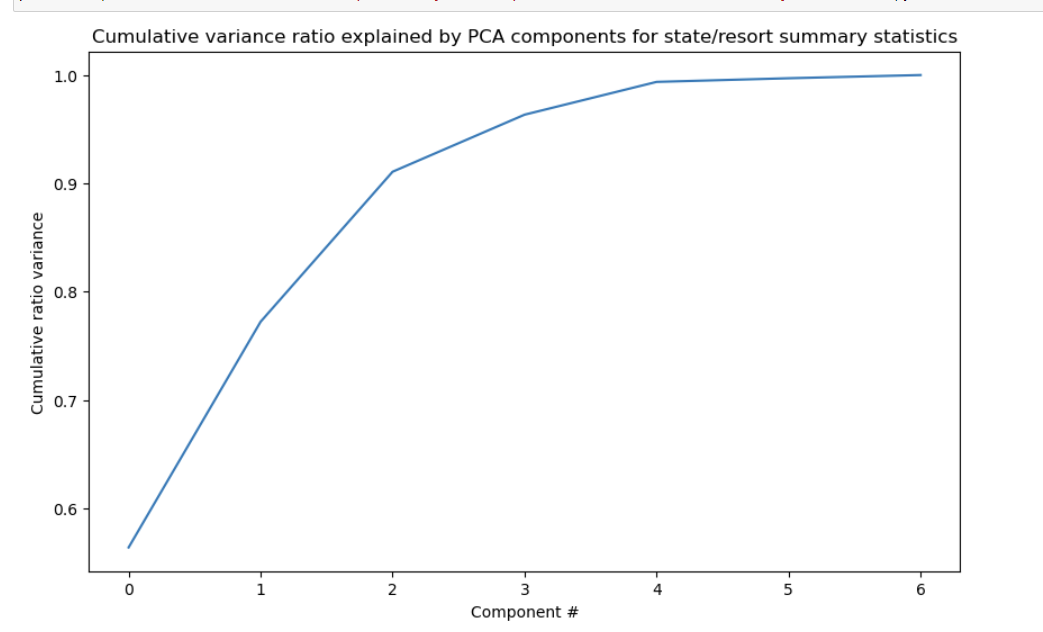
1. We started by noticing the top 5 states in terms of skiing state area: We are in top 3



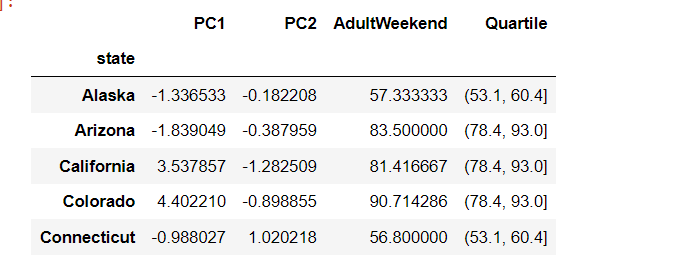
1. We obsered that, in terms of resorts density, Montana did not appear in the top 5.
2. In order to make our calculations easy, we made state name - INDEX so we can easily study the rest of the numerical features.



1. We performed PCA analysis - The first two components seem to account for over 75% of the variance, and the first four for over 95%.



1. We saw a range in average ticket price histogram, but it may be hard to pick out differences if using the value for point size. We added another column where I separated these prices into quartiles

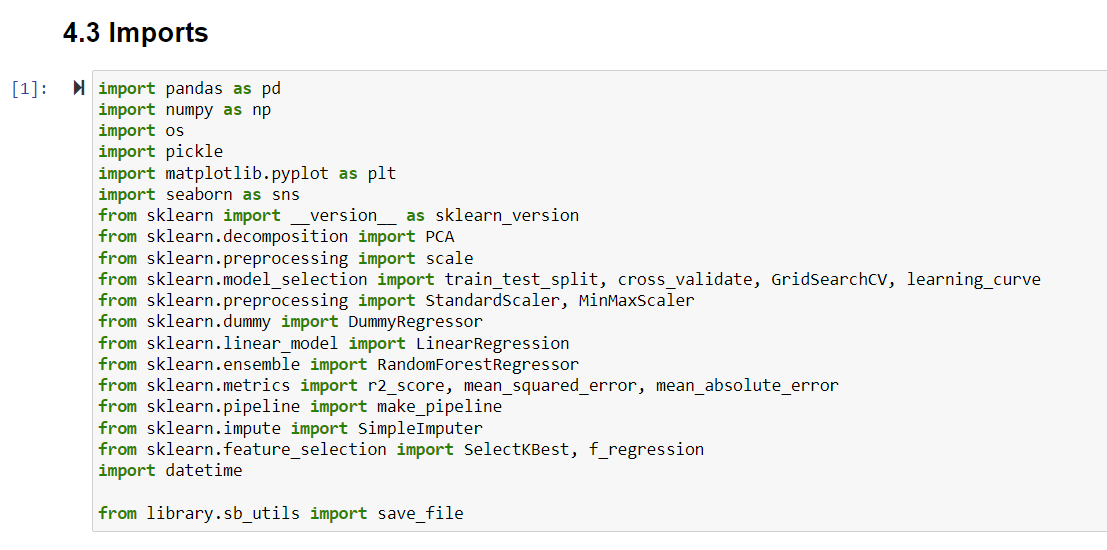


I concluded this sheet with the following observations:

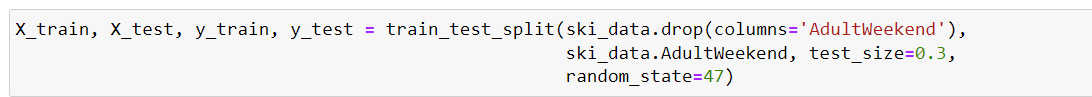
* In the exploratory data analysis (EDA) of the ski\_data dataframe, several numerical and categorical features were identified. Numerical features typically included attributes such as ticket prices ('AdultWeekend', 'AdultWeekday', etc.), various resort characteristics (e.g., 'vertical\_drop', 'base\_elev', 'fastQuads'), and skiable terrain area ('TerrainParks', 'Runs', 'total\_chairs', etc.). Categorical features comprised attributes such as 'Name', 'Region', and 'state', which denote the name of the resort, the region it belongs to, and the state where it's located, respectively.
* We observed a potential relationship between states and ticket prices, suggesting that prices might vary based on the state where a resort is located. This led us to consider including the 'state' feature in subsequent modeling to capture this geographical influence.
* When selecting features for modeling, we need to be cautious of multicollinearity and spurious correlations between features. Additionally, we must carefully handle the 'state' labels to avoid bias or overfitting. Encoding states as dummy variables or grouping them into broader regions could be helpful.
* For modeling, 'AdultWeekend' ticket prices seem like a reasonable target variable. We'll need to check its distribution and possibly transform it to meet model assumptions. Handling 'state' labels effectively is crucial for capturing geographical influences on ticket prices while avoiding model complexity.

**Model Preprocessing with feature engineering & Algorithms used to build the model with evaluation metrics.**

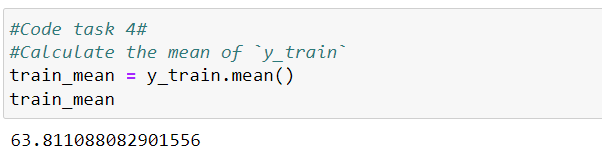
Imported the following libraries:



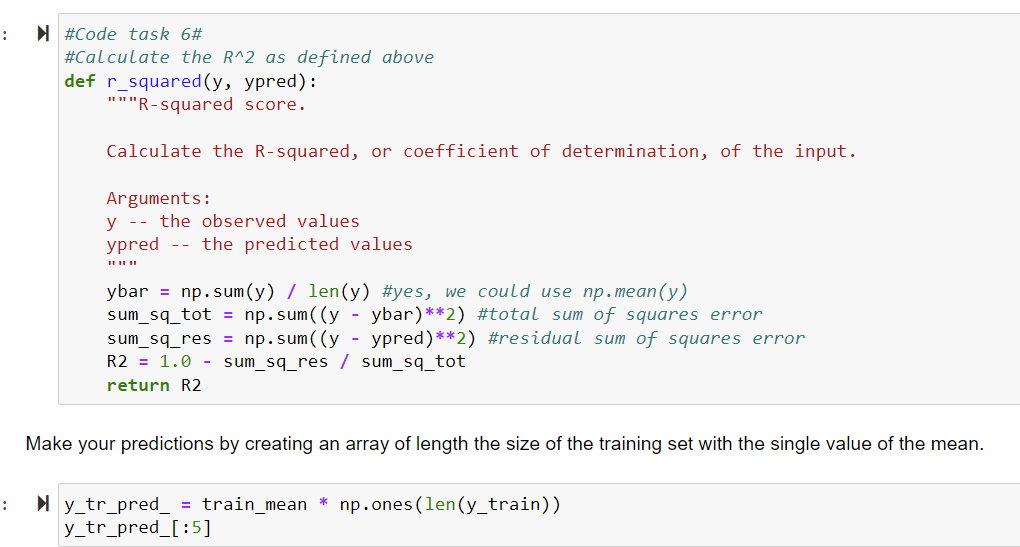
We split our data into sizes 70/30 train/test split.



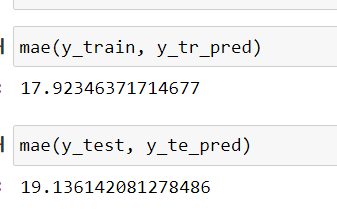
Before even starting with learning a machine learning model, we started by considering how useful the mean value is as a predictor.



We calculated the R2 score and did a comparison with dummy regressor.



We further calculated the Mean Absolute Error and it essentially tells you that, on average, you might expect to be off by around ~$19 if you guessed the ticket price based on an average of known values.



Before starting to work on our model:

We Impute missing values with median

Apply the imputation to both train and test splits

Scale the data

Train the model on the train split

Assess model performance

Using this model, then, on average we expected to estimate a ticket price within ~$9 or so of the real price. This is much, much better than the ~$19 from just guessing using the average.

**MODELS:**

1. **Linear Regression Model**

We created a pipeline that

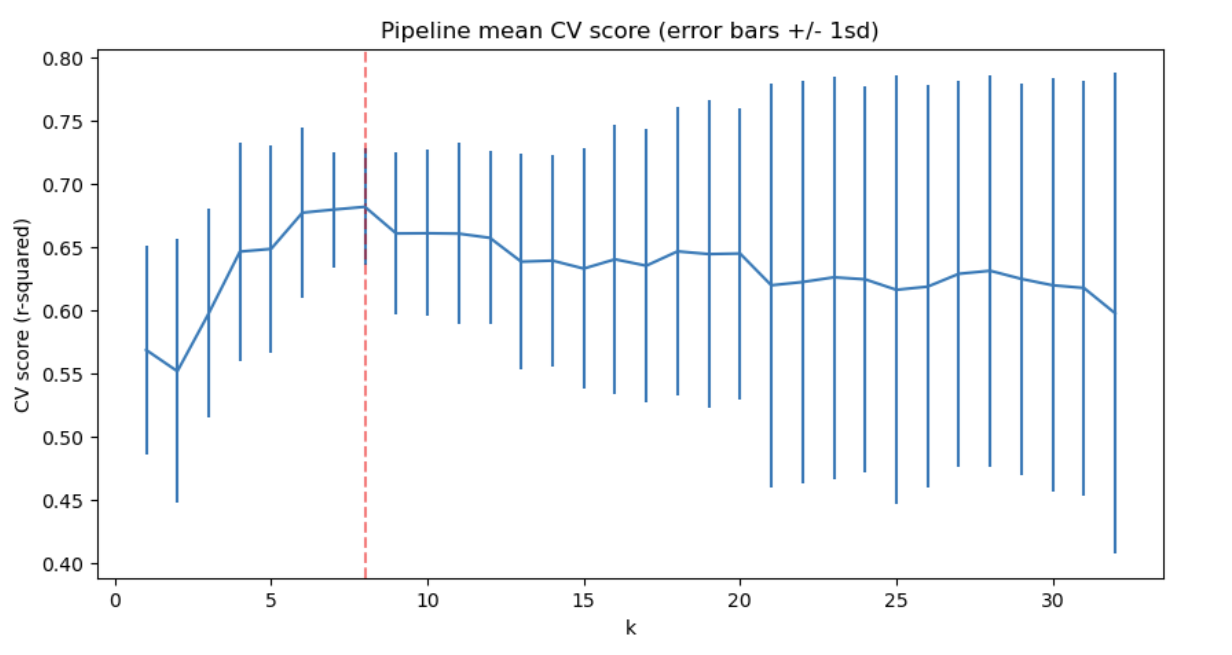
\* imputes missing values

\* scales the data

\* selects the k best features

\* trains a linear regression model

\* a technique (cross-validation) for estimating model performance

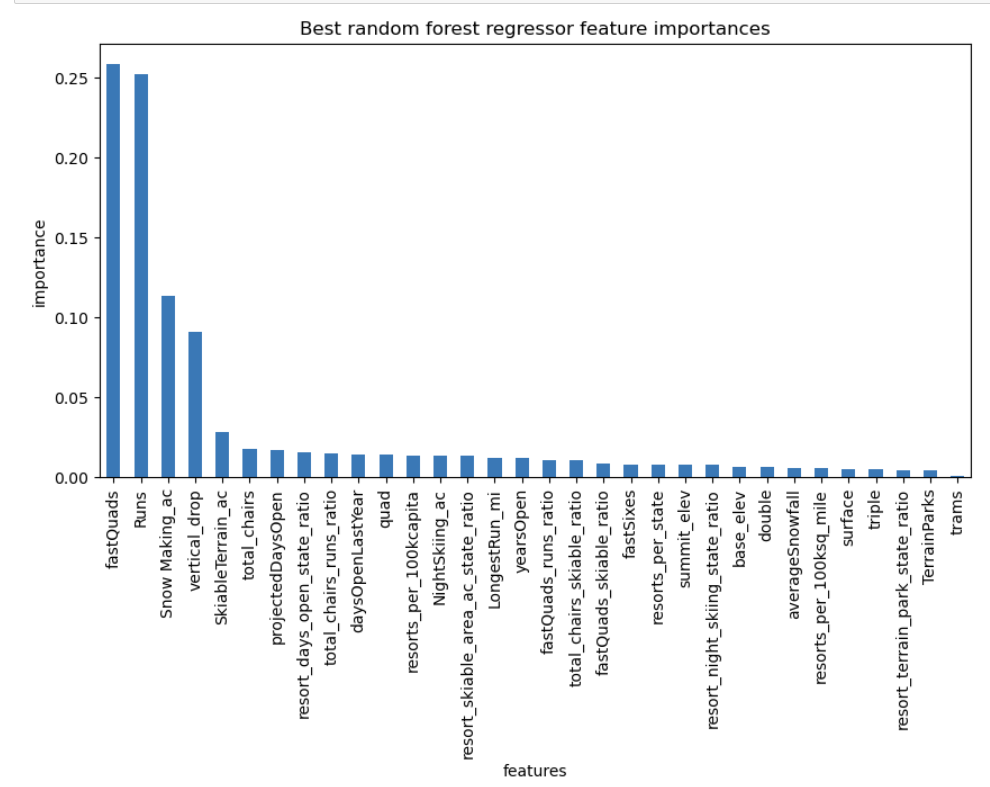


The above suggests a good value for k is 8. There was an initial rapid increase with k, followed by a slow decline. Also noticeable is the variance of the results greatly increases above k=8.

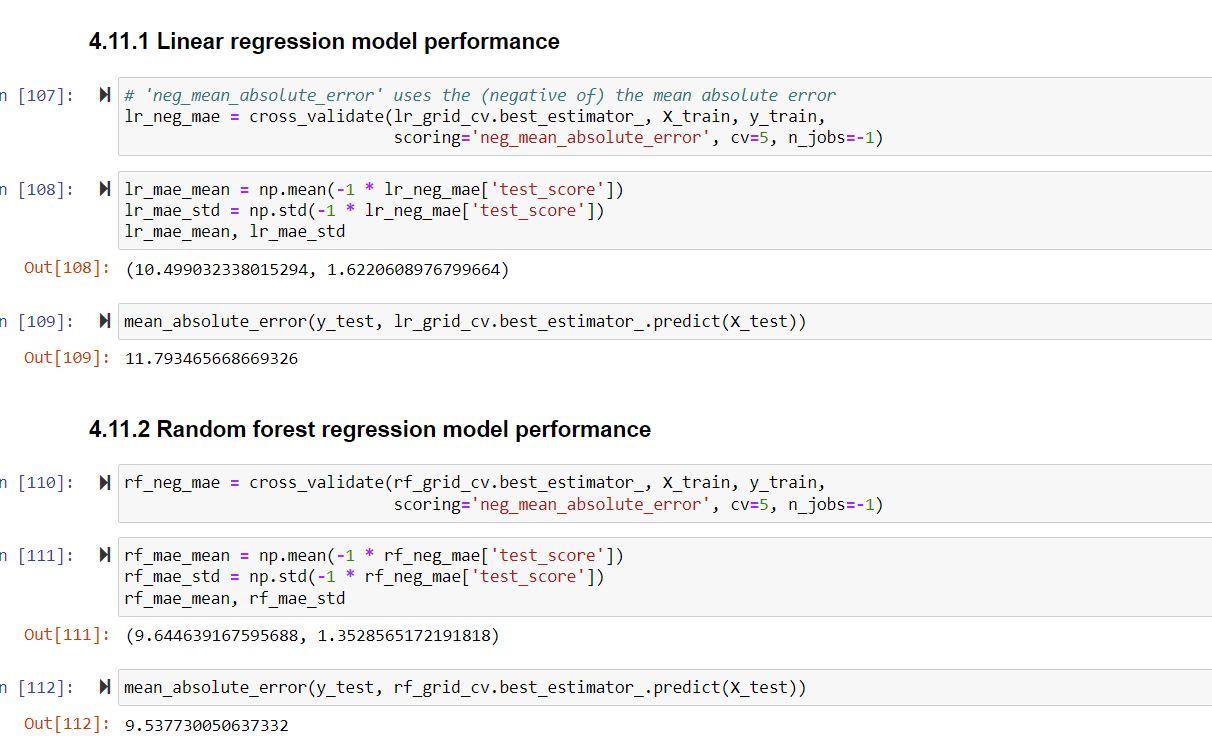
Our final findings:

These results suggest that vertical drop is your biggest positive feature. This makes intuitive sense and is consistent with what you saw during the EDA work. Also, you see the area covered by snow making equipment is a strong positive as well. People like guaranteed skiing! The skiable terrain area is negatively associated with ticket price! This seems odd. People will pay less for larger resorts. There could be all manner of reasons for this. It could be an effect whereby larger resorts can host more visitors at any one time and so can charge less per ticket.

1. **Random Forest Model**

****

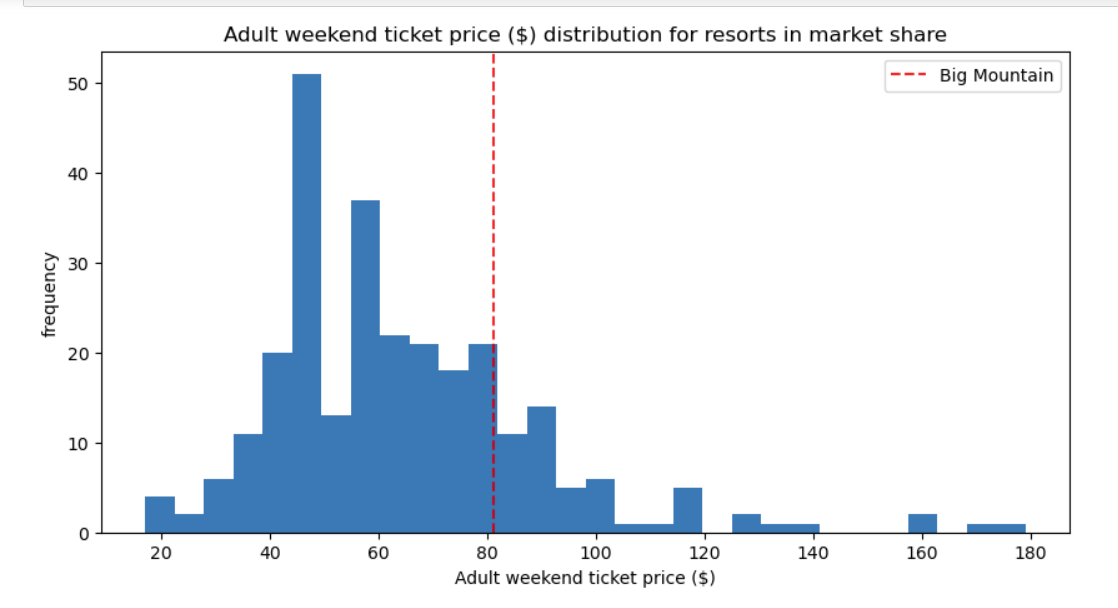
**Winning model and scenario modeling**

****

Conclusion:

The random forest model has a lower cross-validation mean absolute error by almost ~$1. It also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

**Pricing recommendation**

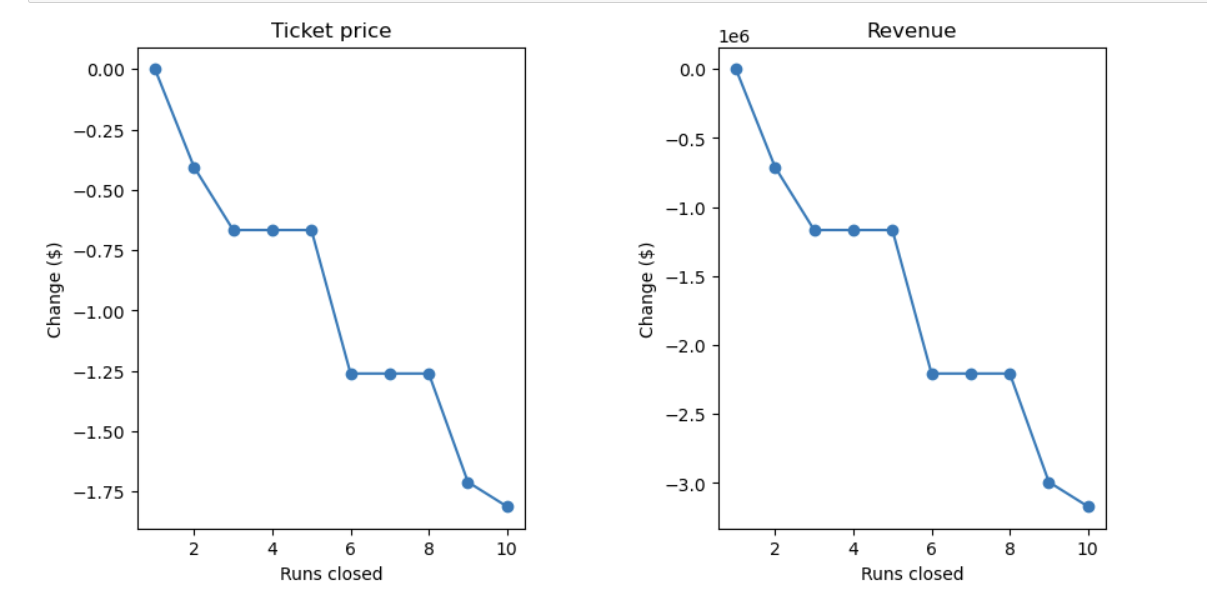
****

At present, Big Mountain Resort charges $81 for a ticket. Our modeling suggests that the market could support a ticket price of around $95. This indicates there's room for the resort to adjust its prices to better match what customers expect.

Impact of Additional Facilities: Adding a new chair lift would significantly boost the predicted ticket price. This suggests that investing in new facilities, like the proposed chair lift, could help Big Mountain justify a higher ticket price and improve its offerings.

Revenue Analysis: We also looked at how different scenarios, such as closing certain runs, could affect revenue. Analyzing these predictions can help Big Mountain make smart decisions about pricing and investments to maximize revenue without sacrificing customer satisfaction.

The model says closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop. Image below:



Recommendations: Based on our analysis, here's what I suggest for Big Mountain Resort:

Adjust Ticket Pricing: Consider raising ticket prices to around $95 to better reflect the value the resort offers.

**Conclusion:**

In summary, our analysis provides valuable insights to help Big Mountain Resort fine-tune its pricing strategy and stay competitive in the ski resort industry. By following these recommendations and staying attuned to market dynamics, the resort can maximize revenue and keep guests happy on the slopes.

**Future scope of work**

Enhance Facilities: Move forward with plans to add a new chair lift to improve the resort's amenities and potentially justify the proposed ticket price increase.

Stay Informed: Keep a close eye on market trends, feedback from guests, and what competitors are doing to ensure decisions about pricing and investments are always well-informed.