Springboard

6.7 Guided Capstone Report - Big Mountain Resort

Introduction

The task in the first Springboard capstone was to identify ways for Big Mountain Resort (called Whitefish Mountain Resort in the data) to increase revenue and maintain profit margins even with increased costs. The data was provided in csv form. A linear model was fit using least squares regression.

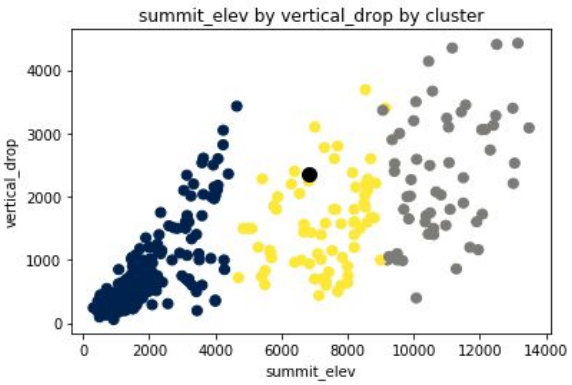
Findings

The key finding from the guidance was the predicted ticket price for the resort. After finding the resort by matching the data from the description to the row (and checking Springboard comments), I found the expected price was $88.51. This was an increase from $81. The top features were:

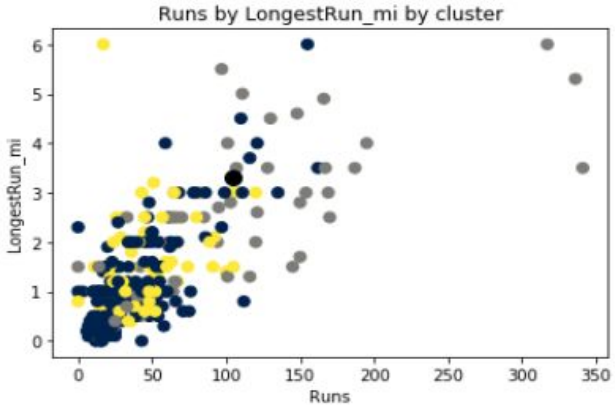
1. Adult weekday fare
2. Vertical drop
3. Number of runs
4. Number of quad chairs
5. Average annual snowfall

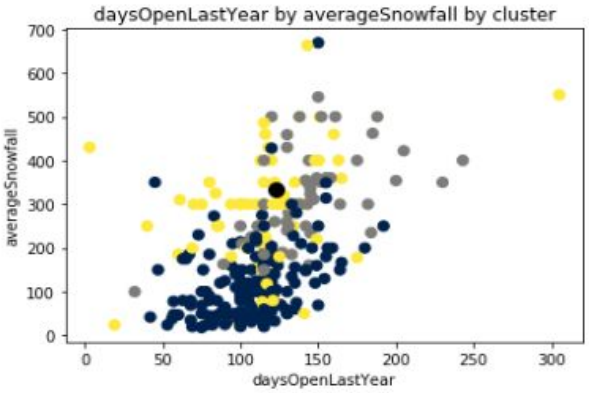
It is interesting to see vertical drop included, since summit_elevation and base_elevation were excluded. However, this could be somewhat in the resort's control in how they build their runs.

I generated plots for the top 5 features and colored by cluster, as in the example. The first was the example of summit elevation and vertical drop. Here (**right**) we see that the target resort lies somewhere in the middle to top third of the resorts.



I then plotted the number of runs by longest runs (**below, left**) and again saw the target resort (black spot) was in the upper third of the resorts. It's placement among its competitors suggests it could charge similar prices. The final plot was to look at the fifth top variable - average snowfall, with a potential target variable of the number of days open in the last year. This plot (**below,right**) was less illuminating, but showed again that the target report was clustered near the top third.





Challenges

My difficulties with this dataset were centered on outliers and scaling data. I found when I initially went through and followed the instructions that I ended up with a small dataframe of ~170 rows. I was surprised by the variance of 0.66 for an initial project and that my data's top features did not line up with the guidance in the notebook.

I first investigated how I preprocessed my data. I changed how I replaced NaN - all zeros, all with mean, and 'adjusted' means filled when the data followed a normal distribution and 0 otherwise. This had some effect but my variance remained in the 0.6 - 0.75 range.

I then found a mistake where I had not passed X_scaled to my linear fit, but rather X. This explained the strange value for the intercept (which was nowhere near the mean of the data). However, when I passed X_scaled to the fit for the first model, my error was huge, on the order of 10^22. My hypothesis is an error of division of something very close to 0 when creating the scaling Z score. This issue only appeared in model 1, after the state dummy variables were made.