Whitefish Mountain Resort Project Report

Justin Bates, Data Scientist
July 17, 2020

Problem Identification Overview

Context

Whitefish Mountain Resort (WMR), located in northwestern Montana, offers spectacular views of Glacier National Park and Flathead National Forest. Whitefish Mountain Resort, initially opened in 1947 as Big Mountain Resort, with an annual snowfall of 333 inches and 3,000 acres of skier and rider accessible terrain. Annually 350,000 skiers and snowboarders ride Whitefish Mountain Resort's 105 named trails, vast bowl, and tree skiing.

WMR has recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This extra chair increases their operating costs by $1,540,000.00 this season. With the new lift, Whitefish Mountain management needs to know how much they can raise the adult weekday and weekend lift tickets to maintain their profit margin of 9.2% this season.

Problem

How much will Whitefish Mountain Resort need to increase their weekday and weekend adult lift tickets to maintain a minimum profit margin of 9.2% for its investors during the upcoming season, while covering additional operating costs of $1,540,000.00 for the new chair lifts?

To generate a solution to the problem, we looked at the current adult weekday and weekend lift ticket prices and other relevant resort data from across the country. The goal of this project was to develop models that predict the expected adult weekday and weekend prices to increase revenue in the upcoming season to account for the increased operating cost of $1,540,000.00 for a new chair lift.

Another avenue we looked down was developing a model that predicts the projected days open for the upcoming season to see if there could be more days then expected.

Problem Solution

By keeping the weekday ticket price at $81.00 but raising the weekend to $86.00 from $81.00, WMR would just miss the shareholder profit goal of 9.2% with a 7.2% rather than without changing the price with a 3.77%. This is calculated with a scenario where 60% of skiers are weekend riders and 40% are weekday rider. WMR has predicted that there will be 123 days (~88 weekdays and ~35 weekend days) in the next ski season with 350,000 skiers passing through. Again, If 60% of those skiers come on the weekend there will be a predicted increase in revenue of $1,050,000.00, which missed the CAPEX of $1,540,000.00. However, with our model that predicts 130 days in the next ski season, WMR would meet their margins and more.

Our linear regression model predicts the season will open for 130 days, an increase of 5% for this upcoming season. This extra week of skiing will bring in an additional 17,500 guests. With 40% of riders purchasing an adult weekday ticket price of $81.00, WMR's revenue would go up by $567,000.00. Plus, With 60% of riders purchasing an adult weekend ticket price of $86.00, WMR's revenue would go up to by $903,000.00. The total revenue for the extra week of skiing would be $1,470,000.00. We accounted for an increase in OPEX of 5%, although it could be more or it could be less. More information would be needed to properly calculate margins.

We generated a second scenario where the weekday skiers were 25% and weekend skiers were 75%. This scenario 2 and our original scenario, scenario 1, can be found in the financial break down below for more in-depth understanding.

Financial Break Down

We must note, with ***Table 1*** below, that the financial scenarios make assumptions for the relevant case study. We assume that the last season's customer tickets sold will stay at 350,000 for this season. We created two scenarios. The first scenario assumes that the number of weekday skiers is 40%, and weekend skiers are 60%; the second scenario assumes that the number of weekday skiers is 25%, and weekend skiers are 75%. We must also assume only the net profit margin because there is not enough information provided by the case study to calculate gross or operating profits. We did not add OPEX cost to run the new lift, but as we add an extra week to the season from the season prediction model, we add a week's worth of OPEX, this ends up being a 5% increase.

			Scenario 1 40% weekday : 60% weekend	Scenario 2 25% weekday : 75% weekend
Operating Revenue		Normal	Changing Price & Predicted Season	Changing Price & Predicted Season
	Total Operating Revenue	$ 28,350,000.00	$ 30,870,000.00	$ 31,145,625.00
Operating Expenses				
	Total Operating Expenses	$ 25,741,800.00	$ 27,746,780.49	$ 27,746,780.49
Net Profit				
	Total Net Profit	$ 2,608,200.00	$ 3,123,219.51	$ 3,398,844.51
	Net Profit Margin	9.20%	10.12%	10.91%

Table 1: The table displays the projected income statement of 2 scenarios. In the first scenario, 40% of visitors buy weekdays, and 60% buy weekend tickets. While in the second, 25% of visitors buy weekday and 75% of visitors buy weekend tickets. Located in the link below is the income statement breakdown:
https://docs.google.com/spreadsheets/d/1cxcq2yBNPQKfL-DYHagFRaa9jkbzBw7_q0WcD3u1rbk/edit?usp=sharing

Data

A CSV file that contains 27 features from 330 unique resorts across America, including Whitefish Mountain Resort. Alesha Eisen, the Database Manager at WMR, provided this CSV file.

Models

We generated three linear regression models to predict the adult weekend ticket price, adult weekday ticket price, and the projected number of days the resort will be open next season. We also generated 3 random forest regression models to predict the adult weekend ticket price. We didn't bother with the rest because after testing them with weekend prices the random forest regression models did poorly. We believe that the data was too noisy to get an accurate model, even with removing the outliers from data.

Deliverables

- Jupyter notebooks outlining the data science method used to produce models from data set to solve the defined problem best
- Project Report
- Executive Summary Presentation

Data Preprocessing: Notable Steps

We started the preprocessing steps by replacing any Nan values found in the data set with median, mean, and zeros, respectively. Next, we developed clusters as a new feature using the other numeric features and K-means. We fit the K-means algorithm with a k parameter set to the optimal number, 3. The clusters seem to be created based on the summit elevations, and similar features which are highly correlated, as shown in ***Figure 1***.

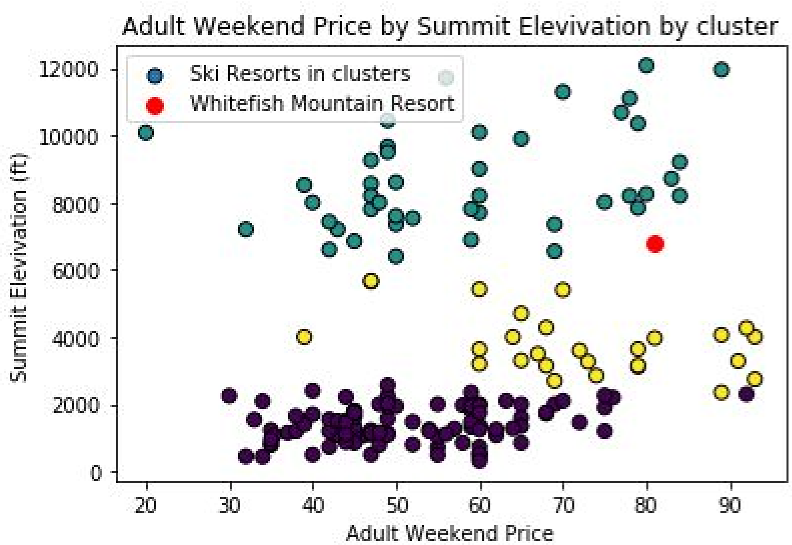

Figure 1: The purple cluster is the resorts at 3000 feet or lower (< 3000). The yellow cluster is the resorts between 3000 and 6000 feet (>3000 and <6000). The green cluster is the resorts higher than 6000 feet (>6000). Pinned with the big red dot is the Whitefish Mountain Resort, and it sits at the bottom end of the green cluster.

We created dummy features for the categorical variable state. Next, we standardized the magnitude of numerical features. Finally, we used the sklearn model selection import `train_test_split` and created a 75/25 split with the model response variables.

Model Descriptions

Model A is a linear regression model with a model response variable of `AdultWeekend` and performed the best when using all the features except base elevation, summit elevation, and the clusters made in the preprocessing step. We looked into using a random forest regression model, but because of the dataset's noise caused by user input error, we had higher accounts of error in the model evaluation than the linear model, Model A, we selected. Model B is also a linear regression model with the model response variable of `AdultWeekday` that performed the best when using all the features except base elevation, summit elevation, and the clusters made in the preprocessing step. We found better results when drop those 3 features with this model because they are highly correlated and generated noise within the model. Model C is a linear regression model with a model response variable of `projectedDaysOpen` that performed the best when using all the features except base elevation and states.

Model Performances

Model A produced a higher explained variance than the other two draft models, showing that 94.2% of the variance in the `AdultWeekend` model response variable (y) is explainable by a linear relationship between the input features (X) and the output prediction `AdultWeekend` (y). The mean absolute error for model 1 is lower than the other draft models at 4.891.

Model B produced a higher explained variance than the other two draft models, showing that 93.1% of the variance in the `AdultWeekday` model response variable (y) is explainable by a linear relationship between the input features (X) and the output prediction `AdultWeekday` (y). The mean absolute error for model 2 is lower than the other draft models at 5.343.

Model C produced a higher explained variance than the other two draft models, showing that -4.1% of the variance in the `projectedDaysOpen` model response variable (y) is explainable by a linear relationship between the input features (X) and the output prediction `projectedDaysOpen` (y). The mean absolute error for model 2 is lower than the other draft models at 13.738. *This model was not very accurate and the evaluation was terrible but we still decided to trust it for the sake of this case study.*

Model As' Evaluation

Model	Explained Variance	Mean Absolute Error	Features Dropped
Linear Regressor			
Model 1.	0.941	4.891	'summit_elev','base_elev', 'clusters'
Model 2.	0.925	5.404	'state'
Model 3.	0.931	5.211	'state', 'base_elev', 'summit_elev'
Random Forest Regressor			
Model 4.	0.906	5.808	'summit_elev','base_elev', 'clusters'
Model 5.	0.910	5.749	'state'
Model 6.	0.909	5.772	'state', 'base_elev', 'summit_elev'

Model Bs' Evaluation

Model	Explained Variance	Mean Absolute Error	Features Dropped
Linear Regressor			
Model 1.	0.931	5.343	'summit_elev','base_elev', 'clusters'
Model 2.	0.914	5.745	'state'
Model 3.	0.918	5.616	'state', 'base_elev', 'summit_elev'

Model Cs' Evaluation

Model	Explained Variance	Mean Absolute Error	Features Dropped
Linear Regressor			
Model 1.	-0.227	15.448	'summit_elev','base_elev', 'clusters'
Model 2.	-0.041	13.738	'state'
Model 3.	-0.056	13.775	'state', 'base_elev', 'summit_elev'

Model Results & Insights

The models, along with some frequency and scatter plots, give us a ton of insights on how to maintain a minimum profit margin of 9.2%. Many things go into a customer's mind when choosing a ski resort, including location, commute convenience, quality snowfall, and skiable terrain. ***Figure 2*** shows a heat map of the US states statistics for some resort attributes. We will dive more in-depth in the Results and Insights subsections.

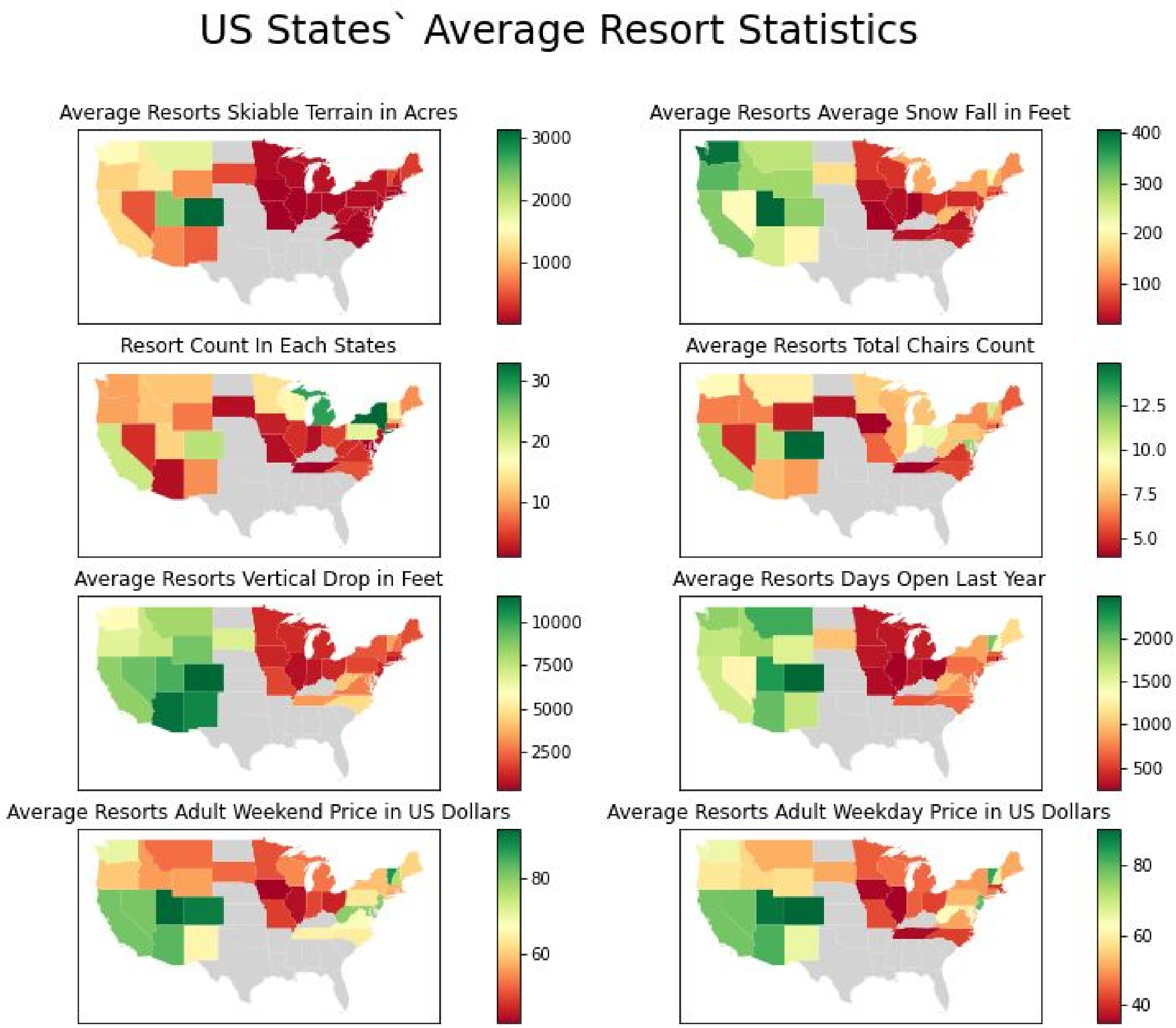

Figure 2: The heat map of the US states statistics on relevant ski resort data. The red is lowest, moving through yellow to green representing the highest value. Grey states represent states that don't have any data available for in the used dataset.

As you can see, the northeastern states like New York, Michigan, and Pennsylvania have more options for a more modest experience. The northeast tends to have on average tickets that are much cheaper, the mountain is much smaller, little snowfall on average, and a fraction of the number of runs. The northeast has a lot more options in the area, meaning the commute would be short, which is excellent for day trips.

On the other hand, the northwestern states like Colorado, California, Utah, Idaho, Montana, Washington have slightly fewer spectacle facilities. The resorts showcase countless scenic runs with an almost endless skiable territory. The average snowfall is hard to beat with the northwestern mountain ranges. This ends up creating a much longer ski season for the region.

Model Results

Model A, with the `AdultWeekend` model response variable, predicts that WMR should be charging $84.90 rather than $81.00 for an adult weekend lift ticket. ***Figure 3*** shows a histogram for the frequency of `AdultWeekend ` along with the WMR actual and predicted `AdultWeekend` ticket cost. Model B, with the `AdultWeekday` model response variable, predicts that WMR should be charging $77.00 rather than $81.00 for an adult weekday lift ticket. ***Figure 4*** shows a histogram for the frequency of `AdultWeekday ` along with the WMR actual and predicted `AdultWeekday` ticket cost.

The range difference on average between resorts for the weekend and weekday price is $6.25. WMR's current range difference between weekend and weekday price is $0.00. Our model predicted that the range difference would be above average at $7.90. We could play with this range on a sliding window without changing the weekday price, but adding the $7.90 to the origins weekend price making the possible value of WMR as high as $88.90. We believe that a small, more subtle, change to $86.00 is valid. The $5 is small enough that most people wouldn't notice the difference nor would it change their ski plans.

Model C, with the `projectedDaysOpen` model response variable, predicted the WMR should be open 130 days rather than 123 days. ***Figure 5*** shows a histogram for the frequency of `projectedDaysOpen ` along with the WMR actual and predicted `projectedDaysOpen.`

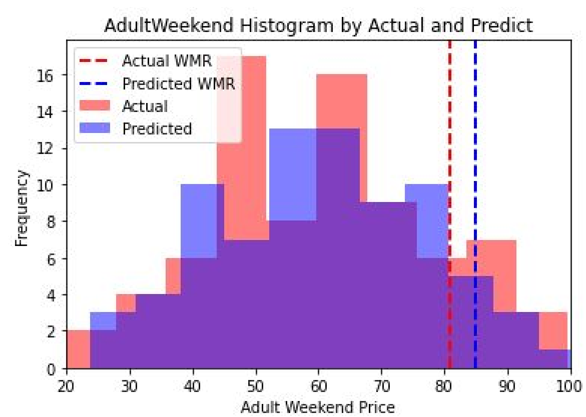

Figure 3: The histogram of the `AdultWeekend` shows the resorts' actual and predicted ticket costs frequency alongside the WMR's actual and predicted ticket costs. The red bins and line represent the actual, while the blue bins and line represent the predicted.

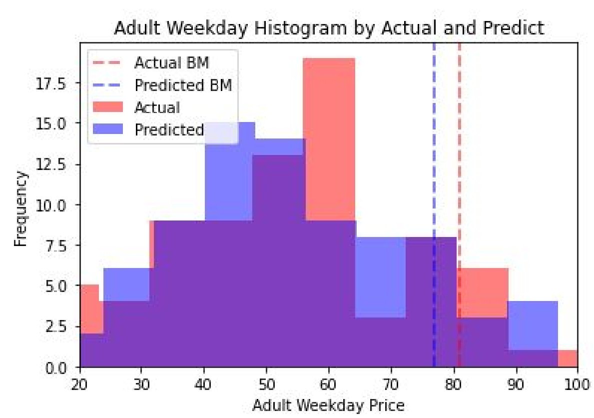

Figure 4: *The histogram of the `AdultWeekday` shows the resorts' actual and predicted ticket costs frequency alongside the WMR's actual and predicted ticket costs. The red bins and line represent the actual, while the blue bins and line represent the predicted.*

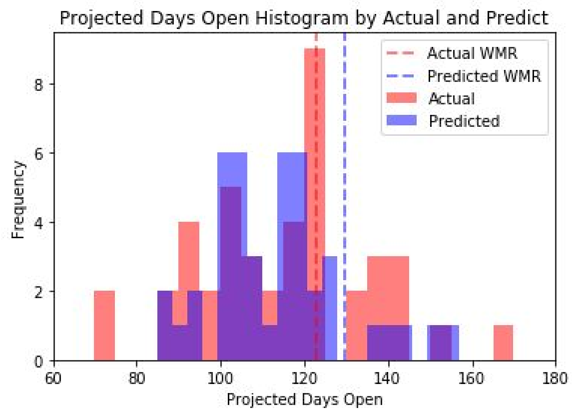

Figure 5: *The histogram of the `projectedDaysOpen` shows the resorts` actual and predicted days open for the next ski season frequency alongside the WMR's actual and predicted days open for the next ski season. The red bins and line represent the actual, while the blue bins and line represent the predicted.*

Model Insights

When we dig deeper and take a look at some of the features alongside the `AdultWeekend` and `AdultWeekday,` we can begin to uncover the true value of WMR. Plus, when we use the prediction for the `projectedDaysOpen` to get a more extended season, we can start to see how we can maintain their profit margin of 9.2% this season.

Let's start by looking at WMR features in comparison to the other resorts. We will keep them in the clusters to point out groups by sizes of each resort as explained in ***Figure 1,*** before we look at the insights the features give us. We want to note the outliers have not been removed. We can see that they are in the purple cluster which is a smaller resort cluster. Plus, the ticket price means that were used to replace NaN values cause a bigger issue in the dataset itself. After further research of resorts website ticket prices, most of the NaN valued tickets we greater than $100.00, and in some cases greater than $200.00. We believe that the data was let out when entered by users because they were so high or the resort tickets and always going up each season. So, the mean weekday price ended up being $57.92, and $64.17 for the weekend tickets. The mean can be easily identified by the out of place vertical dot stack about the mean price values.

With each feature we review, the scale of the y-axis begins to zoom in while the x-axis stays constant with the price. The first feature we will compare against price is the skiable terrain in acres as shown in ***Figure 6***. Here we see that WMR is in the top ten percentile for skiable terrain. The second feature we will compare against price is the snowmaking in acres, as shown in ***Figure 7***. Here we see that WMR is in the top ten percentile again for snowmaking. The next feature we will compare against price is the total run count, as shown in ***Figure 8***. Here we see that WMR is in the top ten percentile again for the total run count. The final feature we will compare against price is the total chair count, as shown in ***Figure 9***. Here we see that WMR is in the top ten percentile again for the total chair count.

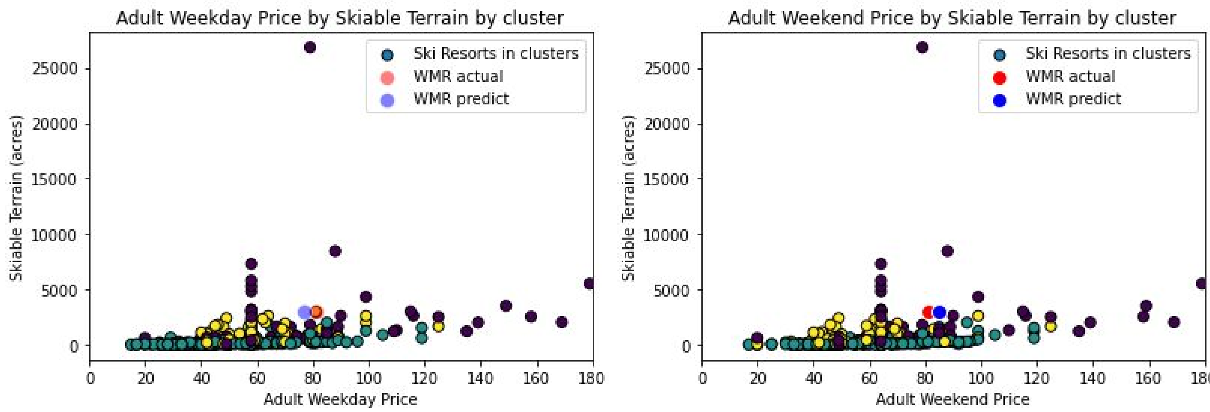

Figure 6: The scatter plot to the left displays weekday price vs. skiable terrain in acres; the scatter plot to the right displays weekend price vs. skiable terrain in acres, both with clusters. The purple cluster is the resorts at 3000 feet or lower (< 3000). The yellow cluster is the resorts between 3000 and 6000 feet (>3000 and <6000). The green cluster is the resorts higher than 6000 feet (>6000). Pinned with the big red dot is WMR as their actual listed price and anchored with the big blue dot is WMR as their predicted price, both vs. the skiable terrain in acres.

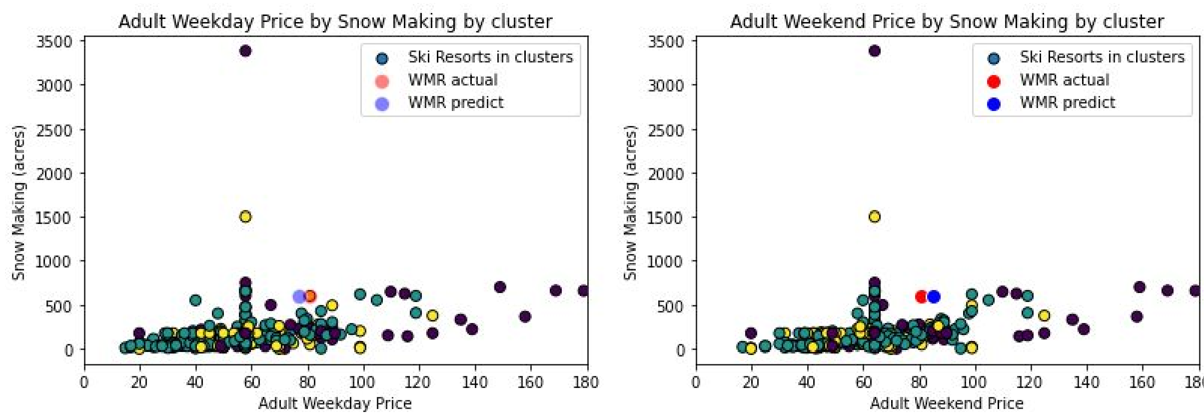

Figure 7: The scatter plot to the left displays weekday price vs. snowmaking in acres; the scatter plot to the right displays weekend price vs. snowmaking in acres, both with clusters. The purple cluster is the resorts at 3000 feet or lower (< 3000). The yellow cluster is the resorts between 3000 and 6000 feet (>3000 and <6000). The green cluster is the resorts higher than 6000 feet (>6000). Pinned with the big red dot is WMR as their actual listed price and anchored with the big blue dot is WMR as their predicted price, both vs. the snowmaking abilities in acres.

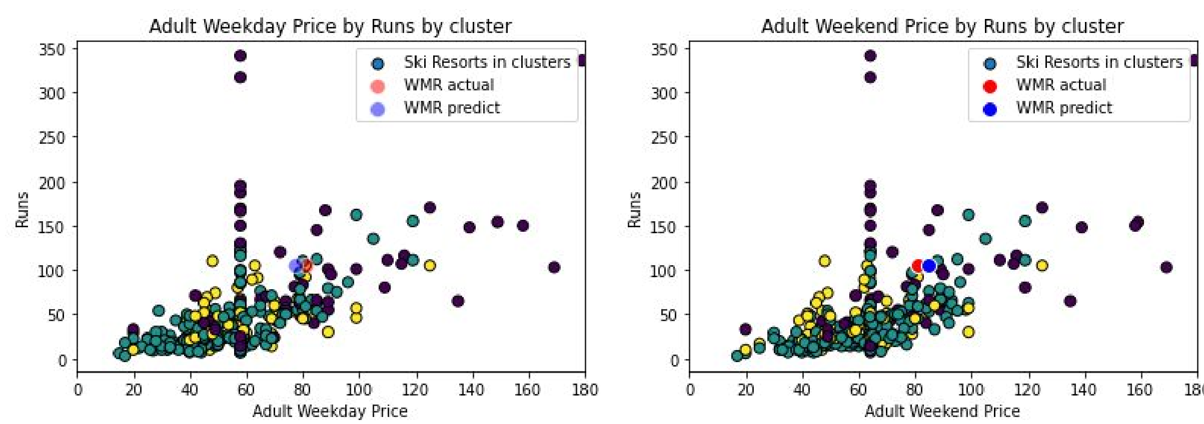

Figure 8: The scatter plot to the left displays weekday price vs. run count; the scatter plot to the right displays weekend price vs run count, both with clusters. The purple cluster is the resorts at 3000 feet or lower (< 3000). The yellow cluster is the resorts between

3000 and 6000 feet (>3000 and <6000). The green cluster is the resorts higher than 6000 feet (>6000). Pinned with the big red dot is WMR as their actual listed price and anchored with the big blue dot is WMR as their predicted price, both vs. the skiable terrain in acres.

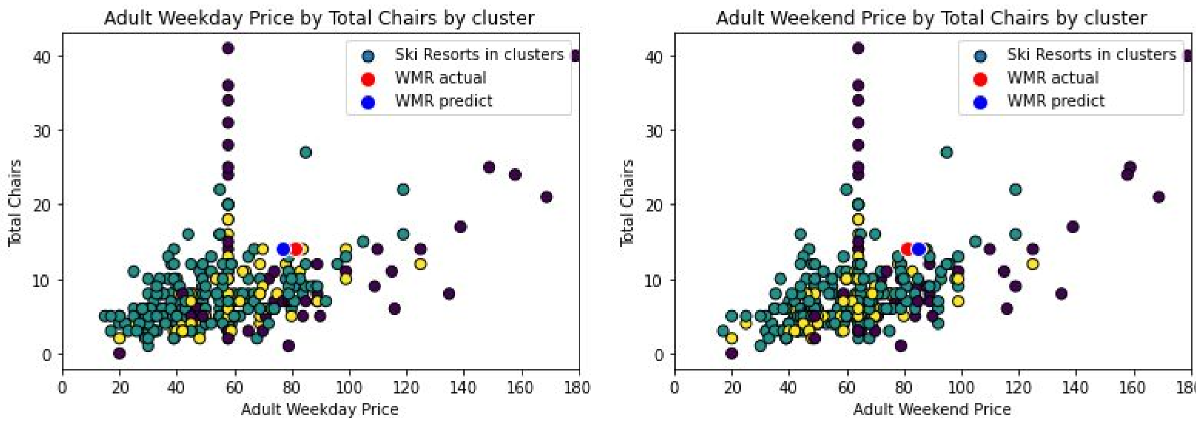

Figure 9: The scatter plot to the left displays weekday price vs total chair count; the scatter plot to the right displays weekend price vs. total chair count, both with clusters. The purple cluster is the resorts at 3000 feet or lower (< 3000). The yellow cluster is the resorts between 3000 and 6000 feet (>3000 and <6000). The green cluster is the resorts higher than 6000 feet (>6000). Pinned with the big red dot is WMR as their actual listed price and anchored with the big blue dot is WMR as their predicted price, both vs the skiable terrain in acres.

Conclusion

To conclude the report, Whitefish Mountain Resort can increase their adult weekend price to $86.00 to grow the shareholder profit margins goals of 10.12%, but to meet this new profit margin with this new price, the resort needs to stay open an extra week as predicted by model C. WMR has the right features to raise the price higher with the current chair. Plus, if they continue to make the resort more accessible they could begin to see more foot traffic to their beautiful resort.

Next Steps

1. Do some more research into gathering more accurate data to reduce the number of initial NaN values.
2. Look into obtaining additional data such as other revenue streams, the number of customers per weekday/weekend, and other expenses(i.e.COGS, advertising, lawsuits) for each resort to improve the models.
3. Try to obtain the data over the last 30 years for each resort to see how the price increase as the resorts expand.
4. Use ensemble learning models to generate more accurate models once the noise is reduced.
5. With COVID-19, how can this affect the analysis?
    a. Many people are working from home and learning from home which means traveling is easier to schedule. This could provide more skiers.
    b. Many people are not financially stable right now and are they going to have the money to go on a ski trip? Will they try and go somewhere where it’s cheaper? Will they try and go on weekdays instead of weekends? So there could also be fewer skiers or fewer skiers on weekends.