Problem Statement for Big Mountain Resort's Lift Pricing Strategy

Context.—Big Mountain Resort (BMR) is a ski resort in Montana. BMR has some strengths, since they can offer guests excellent views of Glacier National Park and Flathead National Forest and have a 3.3 mile run called Hellfire. Every year 350,000 skiers and snowboarders use their slopes. They have 11 aerial lifts, 2 t-bar lifts, and one magic carpet. They recently added another aerial lift to their facilities "to help increase the distribution of visitors across the mountain". The new lift increased their operational cost for the season $1,540,000.

BMR's current strategy is to set prices relative to the market average with a premium on account of their superior facilities. They want a new pricing strategy that will help them cover the costs from the new ski lift.

Criteria for Success.—BMR will implement a new lift ticket pricing strategy within two-weeks to help increase revenue to the cover the increase in operational costs.

Scope of the Solution Space.—BMR is aware that its current lift ticket pricing strategy is deficient. Therefore, the solution should be in the form of a new ticket pricing strategy.

One method for determining the best new strategy could be to use (internal) historical figures and compare the predicted outcomes of different pricing strategies. Unfortunately, the provisionally available data does not appear to support this approach, since its figures are on the facilities of other resorts. It might be more fruitful to consider BMR basing its prices on its *nearest competitors* rather than the market average, which would give them a better estimate of their worth. The available data supports this method by providing insights into the pricing and facilities of its competitors.

Constrains within the Solution Space.—In addition to a new pricing strategy, BMR is taking other cost cutting measures that should either improve the quality of their facilities or diminish their quality not at all. The new strategy should be harmonious with or complement those other measures. Moreover, BMR suspects that it is undervaluing its facilities. Therefore, BMR wants the new pricing strategy to either leave the new lift ticket prices where they are at or increase them.

Additionally, BMR has reason to believe that some of its facilities are more valuable than others and needs a more precise estimate of their values. An ideal solution should offer insights into those features and leaver them in a more sophisticated way. (A potentially desirable new strategy might be one that charges a premium for its more valuable lifts.)

A new pricing strategy has the potential to influence turn-out, which factors into revenue. An ideal solution should take into account the potential for it to influence customer turn-out. It would undermine the purpose of the new lift to push customers away, since the lift is intended to increase an already massive number of skiers. But if the suspicion that BMR is undervaluing its facilities is correct, that would mean that an increase in prices would not result in a decrease in turn-out, let alone negate the increase in prices.

Finally, it might not be necessary to increase revenue to something proportional to the new cost. It is important not to set the target increase in revenue to something too high. We need to find an *adequate* target to increase revenue by to take full advantage of the new lift.

Stakeholders to Provide Key Insight.—The Data team should consider the interests of the following:

- BMR owners are invested in the success of the resort.
- BMR upper management has the authority to implement the new strategy and depends on its success.
- BMR facility maintenance are discovering and implementing other cost cutting measures to go along with the new pricing strategy.
- The BMR database manager has access to crucial data on the activity of the resort.

What Key Data Sources are Required?—The Data team should consider the following data sources:

- The CSV file provided by the BMR database manager with information on other resorts.

How many rows and columns did we start with? What rows and columns did we remove? Why did we remove them? How many did we finish with?

Was Big Mountain Resort present in the data?

Our project is to find a pricing model for ski resort tickets in our market segment. The first step in our search is data wrangling and that is what we did in this notebook. We collected, explored, and cleaned our data. It did not take us long to collect the data and we spent the majority of our time exploring and cleaning.

The raw ski resort DataFrame had 330 rows and 26 columns. The observation for Big Mountain Resort was on row 151 and had no missing values. Next we divided our exploration into two branches. In the first branch, we explored the categorical features of the ski resorts DataFrame. The first branch ended up being inconclusive, although we constructed some potentially useful plots.

In the second branch, we explored its numerical features. We examined each column individually for suspicious values. The reported skiable acres for Silverton Mountain was extravagantly high and we found a more accurate number through research. Heavenly Mountains Resort had a suspiciously high number of snow making area. We noticed that it had missing values in both price columns, so we dropped that observation. We dropped the column for fast eight person chair-lifts, since only one resort had any and it meant very little. One observation said a resort has been open for thousands of years. We dropped that row. The second branch ends here.

We dropped all rows that were missing values in both price columns. We dropped the column for weekday prices because it had 7 missing values while the weekend prices had 4 missing values. We considered the weekend prices column our target feature. Then we dropped all of the rows that were missing a value in the weekend prices row. We ended this notebook with 277 rows and 25 columns.

Our project is to find a pricing model for ski resort tickets in our market segment. The first step in our search is data wrangling and that is what we did in this notebook. We collected, explored, and cleaned our data. It did not take us long to collect the data and we spent the majority of our time exploring and cleaning. After we loaded the ski resort data from the CVS file into a DataFrame, we identified two columns as candidates for our target feature: 'AdultWeekday' and AdultWeekend' ski-lift ticket prices.

The raw ski resort DataFrame had 330 rows and 26 columns. We made a preliminary observation of the number of missing values in each row. We found an entry for Big Mountain Resort with no missing values. From there, we divided our exploration into four branches. The first two are the longest. In the first branch, we explore the categorical features of the ski resorts DataFrame. In the second branch, we explore its numerical features. In the third branch, we construct a new DataFrame for state-wide summary statistics for our market segment. In the final branch, we narrow our target feature to adult weekend ski-lift ticket prices. We end this notebook with clean ski resorts DataFrame with 277 rows and 25 columns. Each branch of our exploration is, perhaps, best characterized as the process by which we arrived at that DataFrame.

We began our exploration of the categorical features of the ski resort DataFrame by selecting only columns with the object data type, which gave us three columns: 'name', 'Region', and 'state'. We wanted to make sure that every row represents a unique resort, so we counted the number of distinct resort names in the 'name' column and sorted the count in descending order. If the largest value in the list was one, we could be sure that every row contained a unique observation. However, the name 'Crystal Mountain' appeared twice. It was possible that there were either two observations for the same resort, which would be a problem, or two resorts named 'Crystal Mountain'. We decided to count the number of unique name-state and name-Region combinations. Since in both counts every combination occurred only once, we could be sure that there were two resorts named 'Crystal Mountain'. We double checked our suspicion by displaying all the two rows with 'Crystal Mountain' in the 'name' columns and we saw that one was in Minnesota and one was in Washington. We have not yet dropped any observations or columns.

At that point, every row we had seen had the same value in the 'state' column as the 'Region' column, so we checked to see if the two columns differed at all. We counted 33 rows where the state and Region were different and 297 rows where they were the same. We filtered all of the Regions that were not states: Northern California, Sierra Nevada (California), Sierra Nevada (Nevada), Salt Lake City, and Mt. Hood. There were 38 distinct regions and 35 states with ski resorts. Next we plotted the number of ski resorts in a state or region on two horizontal barcharts. We ended by speculating on the viability of a Montana-specific pricing model.

We finish our exploration of the categorical features of the ski resort DataFrame by forgetting about region for a moment to focus solely on states. In particular, we wanted summary statistics for the price of ski-lift tickets for each state on weekends and weekdays. By doing this, we hoped to represent pricing data about the ski resort market in Montana. We constructed a horizontal bar chart to compare the average price on weekends and weekdays for tickets in each state. To extract more information, we constructed a boxplot the represented the distribution of the two kinds of price in each state. We speculated that it might be useful to consider the difference in prices between weekdays and weekends. With that, we departed from our exploration of categorical features to explore numerical features. We did not drop any rows or columns in our exploration of the categorical features.

We started our exploration of the numerical features of our ski resort DataFrame by describing the data in the numerical columns and representing the distribution of values in every columns with histograms. However, we did make a brief detour into the proportion of rows that were missing values for any or both sorts of price. Now we started to move through each numerical column one by one in search of suspicious values. The first column we examined was the skiable area in acers column. The value for Silverton Mountain is extravagantly high. Internet research revealed that the value represents the non-lift accessible skiable area. In terms of lift accessible area, Silverton Mountain was in middle of the pack, so we replaced its observation in that column with that value. After we replaced that value, we made a new histogram of to represent skiable area.

We considered ourselves done with the skiable area columns and shifted our attention to the snow making area. This time Heavenly Mountains Resort has a value that was suspiciously high. However, before we began our search from a more realistic estimate, we noticed that Heavenly Mountain Resort is missing values for weekend and weekday prices, so we dropped that observation. We felt justified in doing so because rows without any price data are irrelevant to our project. We dropped the column for fast eight person chair-lifts, since only one resort had any and it meant very little. We found an observation that claimed a resort was open for thousands of years. We created a histogram of years open that excluded that row which looked reasonable. We speculated that the value in the years open column was a misplaced date for the year the resort opened. However, no resort was younger than 6 years old and we could not be sure from the data that the resort was ever open in the first place so we dropped that row. We decided to skip the columns for fast six person ski lifts and trams.

For our third branch, we constructed a new DataFrame what contained state-wide skiing summary statistics. We digressed for a moment to drop all rows that have no price data and we summited all of the numeric columns for a second review with histograms. This time we felt satisfied. Now we wanted to add rows for state population and area (square miles) to the new DataFrame. We imported a table from Wikipedia with that data into a temporary DataFrame which we merge with the state-wide skiing summary statistics DataFrame.

We wanted to figure out which of the two price columns was a better target feature. We used the two columns to construct a scatter plot. We made a little table with all of the known prices in Montana on weekdays and weekends. We counted the number of missing values in the two columns and found that the weekdays columns was missing 7 values while the weekends column was missing 4 values. We decided to make the weekend prices our target feature. So we dropped the weekday prices column and all of the rows in the weekends data set with missing values. We conclude the notebook by saving the two DataFrames we made to our computer.