For this data wrangling notebook the data was pulled from an .csv file supplied by the Big Mountain Resort called “ski\_resort\_data.csv.” The Pandas library DataFrame was used to store the data. I used Matplotlib and seaborn to create graphs and chart visualizations for the data.

The first thing looked at was in the data was to analyze where the number and ratio of missing values in the data. This gave me an idea of how “complete” the data was. After that I looked at the categorical features to determine which features aligned most with the problem I am trying to figure out. I also took this time to check the uniqueness of the data to determine if there were any duplicates or incomplete data. I learned there were no duplicates but there were two different resorts with the same name in different regions. From there I analyzed states and regions to determine the differences and what distinctions were made in categorizing the data. I learned that the most resorts were in New York and somehow needed to determine how that knowledge impacted our resort in Montana.

I also discovered that the data does have some incorrect values when looking at outliers. The resort Silverton Mountain had a skiable area listed greatly higher than the rest of the data. After looking into that outlier, I discovered the number was indeed wrong and corrected it in the dataset. After that I looked a the distribution of feature values. Something that was noticed is how skiable terrain is clustered at the lower end for most resorts as well as snow making. Features such as fast eight had too much missing data or not enough resorts had them for the to be a determining factor. The key take away is that among most resort features the data is clustered in tight regions., excluding some outliers.

After analyzing the distributions and how they varied I came back to the primary task of the analysis. How does this all relate to ticket prices? Ticket prices are the key feature I hope to be able to model from this data. What I can tell so far is under the $100 mark adult weekend tickets tended to be higher than adult weekday prices. This will be the feature that I need to find out what most correlates the price with moving forward in my analysis.