Upon completion of the initial review and cleaning of the raw data (ski\_resort\_data.csv), which includes our resort as a data point, the shape of the Dataframe (ski\_data) was reduced from 330 rows x 27 columns to 277 rows x 25 columns. This includes the removal and/or addition of the following rows and columns:

* Column ‘fastEight’
  + Only included data for 164 resorts, 163 of which matched. Therefore it serves little purpose for correlation or comparison
* Column ‘AdultWeekday’
  + Based on the limited differences between this column and “AdultWeekend’ prices, we chose to use only “AdultWeekend
* Rows
  + 47 rows removed due to lacking either ‘AdultWeekend’ or ‘AdultWeekday’ data
  + 4 rows removed due to lacking ‘AdultWeekend’ data only after elimination of ‘AdultWeekday’
  + 1 row removed for having ‘yearsOpen; data as 2019

As part of the review additional information a new Dataframe (states\_summary) including aggregated details about each state was created. Additional information was pulled in from Wikipedia’s US States page to add the following columns to the new Dataframe: ‘state\_population’, ‘state\_area\_sq\_miles’. This Dataframe will be used to determine to what extent different states affect the ticket prices as well as used to group Montana (where Big Mountain is located) into a smaller state subset with similar characteristics for more accurate modeling.

Efforts were made through visualization (boxplots, histograms) to verify values given were within reasonable distributions. One ‘SkiableTerrain\_ac’ was corrected for Silverton Mountain, new value 1819 acres.

At this point we can move forward with exploratory data analysis, though we will need to pay attention to the fact that some columns still have missing data. The target feature will be the ticket price represented by “AdultWeekend’.