Summary Statement for Data Wrangling – Big Mountain Resort

The first step in wrangling the data was to read in the CSV file into Jupyter Notebook. Then, using indexing, it was confirmed that Big Mountain Resort was in the data. For the next steps, the number of null values in each column were counted. The column fastEight had the most null values at 166. Categorical columns were checked for uniqueness. It turns out after some string manipulation that resort names are unique by state. There are two Crystal Mountain Resorts, but they are located in different states. This shows that name and state can be an index together. Given the dataset only contains 330 resorts, it was chosen that Big Mountain Resort should be compared nationally rather than statewide. Next, outliers were removed or adjusted so that Big Mountain Resort’s mean data isn’t skewed.

The first outlier noted was Silverton Mountain, with a skiable terrain area of 26,819 square acres. In actuality, the value found online is 1819 square acres. This was a data correction step, so things such as the website URL, date accessed, and why the value was suspicious in the first place should be noted in the documentation and raised as a question to the client. For this project, we changed the value to 1819 and moved on.

Heavenly Mountain Resort has a suspicious value for Snow Making\_ac, but there’s no ticket pricing for the resort at all. So, the row was removed. The column fastEight is mostly just null values and zeroes, so it was dropped.

\*\*I want to note at this point that Jupyter Notebook will not connect to the kernel at all and troubleshooting failed, so the rest of this summary will be wrote without having access to running code again.\*\*

Another resort was removed for being open for 2019 years. It was likely that the resort was opened in 2019 instead of being open for 2019 years. After this, it was time to remove rows that are missing all price data. That accounted for 14% of the data in the CSV file. A data set was imported from Wikipedia and transformed to merge onto the CSV file. It added population and state area as columns to the data. Then, ticket pricing was the main issue left to decide.

Weekday tickets for adults were compared to adult weekend tickets for resorts in Montana. I feel that all states’ pricing data should have been involved since Big Mountain Resort attracts “visitors of all skill levels”. This means people may travel a long distance to ski Big Mountain Resort. However, Montana allows a quick choice in which price to choose since its weekday and weekend prices are the same for a resort. Weekends have more pricing data available, so weekday prices are dropped. At this point, the data is clean.

NightSkiing\_ac has 187 values, so this means either the resort has no night skiing or didn’t report it. I see that Big Mountain has some night skiing, but only about 20% of its areas are skiable at night. This data still supports my hypothesis, as Big Mountain can expand on its night skiing and have room to raise ticket prices without being the most expensive.