

Summary On The Key Processes And Findings

1. Data Wrangling on the Original Ski Data

The original dataset has 330 rows of records and with 27 columns/features. Among 50 states, New York accounts for the majority of resorts (33 out of 330). Big Mountain Resort is in Montana, which comes in at 13th place and has 12 resorts in total. Aside from some relatively expensive ticket prices in California, Colorado, and Utah, most prices appear to lie in a broad band from around 25 to over 100 dollars. Not all states are homogeneous. Some States show more variability than others. For example, Montana and South Dakota, show fairly small variability. Nevada and Utah, on the other hand, show the most range in prices.

There are missing values. Some resorts have more missing than others, and some features of the resorts are missing more recorded information than others. The most important missing worthy of mentioning is in the two target quantities, the adult weekday ticket price and the adult weekend ticket price. They each has about 15-16% of values missing. The missing in adult weekend ticket price (~15.45%) is less serious than in the adult weekday ticket price (~16.36%), and 47 resorts are missing both prices. Just over 82% of resorts have no missing ticket price. Another heavy missing is on the number of fast eight person chairs ("FastEight"), missing more than half (166 out of 330 total).

There are erroneous recordings as well. 1) Heavenly Mountain Resort has an unusually large snowmaking area. Due to its missing both type of price info, it was dropped from further data analysis. 2) Silverton Mountain Resort has a suspicious skiable terrain area. With some research, we were able to update it with a plausible value and kept it in records. 3) Pine Knob Ski Resort has an erroneous value on years of open. Since there is no clear way to update this information and it was dropped from the dataset. 4). Little Switzerland has a missing value on years of open, and was dropped

Table 1. Features/Records Dropped¹

Feature/Record	Reasons		Action Taken
Fast Eight-person chairlift	Missing 166/330		Dropped
adult weekday ticket prices	Missing 54/330	47 overlap missing	Dropped
adult weekend ticket prices	Missing 51/330		Kept ²
Silverton Mountain Resort	Missing Skiable Terrain Area		Imputed
Pine Knob Ski Resort	Error recording in years of open		Dropped
Little Switzerland	Missing years of open		Dropped
Heavenly Mountain Resort	Error recording in snow making area missing both ticket prices		Dropped ¹
51 Resorts	Missing the weekend price		Dropped
Silverton Mountain Resort	Missing Skiable Terrain Area		Kept and imputed

¹. Heavenly Mountain Resort is included in the 51 resorts dropped due to missing ticket price

². After this step, the data size is updated to 277 rows by 25 columns

Table 1 below summarizes missing information in the original dataset and actions we took in data wrangling. i). We dropped feature "Fast Eight-person chairlift" due to its over 50% missing; ii). We dropped the adult weekday ticket price and kept the adult weekend ticket price for two reasons. On one hand, missing is more serious in weekday prices than that in the weekend price. On the other hand, not only all 12 resorts in Montana but also more than half of all resorts (106 in total) have the

two types of ticket prices matching. iii). We dropped two resorts, Pine Knob Ski Resort and Heavenly Mountain Resort, for erroneous recording, iv). We deleted 51 resort records that are missing adult weekend price, including Heavenly Mountain Resort. v). We kept Silverton Mountain Resort, and imputed its error input on skiable terrain area.

We are now left with a dataset of 277 resorts with 25 features/quantities for analysis.

2. Calculate State Summary and Pull in State Population and Sizes

The remaining 277 resorts are found to be scattered in 35 states. On a state level, we then calculated state summary statistics such as

- i) state total numbers of resorts, state total number of days of resort openings in last year, state total skiable area, state total number of terrain parks and state total night skiing area,
- ii) two state resort density measures: the ratio of state total number of resorts per 100k people, and the ratio of state total number of resorts per 100k square miles.

Because of the nature of the resort business and the heterogeneity of demographic features across 50 states of the United States, On these summary statistics, we also pulled in state population and area information from [wikipedia](https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States).