Cristyn Filla

Data and Metadata Profile

Data Curation I

Jan. 27th 2023

My selected data set for this quarter’s term project is route finder data set for Yosemite National Park from the Mountain Project (a link to this dataset can be found below). This data set details the various known and tested rock-climbing routes found in Yosemite National Park, including both bouldering and traditional (trad) climbing routes. There are 574 routes (also referred to as problems) included in this data set. The data is hosted by the Mountain Project which crowd sources data from climbers who have climbed the routes. Climbers input data on a route page that presents the inputter with a number of fillable fields, some of which require selection from a list of inputs, others allow for free text data to be input. The data on the Mountain Project site is managed and organized by regional administrators who are volunteers for the site. From there the data can then be pulled into a csv in order to see data on multiple routes that meet specified qualifications at once. In order to access the dataset for this project the Mountain Project Route Finder was used to create a file that contains information for routes within Yosemite National Park.

As a crowd-sourced platform Mountain Project does not put any restrictions on data usage on their site. Those most likely to benefit from this dataset in particular are climbers, those both near the Yosemite Valley area as well as nationally and globally who have an interest in climbing at the park; the National Park Service, who could find the knowledge of what routes people are climbing and sharing valuable to park upkeep and understanding park traffic data; and various outdoor/adventure businesses, who might use the data to plan climbing trips or target sales of specific climbing gear needs in the region.

When any of these stakeholders navigate to this dataset for use, they will access it through a single CSV file that contains sixteen pieces of metadata. This metadata is route [name], location, URL, average stars, your stars, route type, rating, pitches, length, area latitude, and area longitude. These metadata can be found in the header row of the CSV. It is important to note that. the route page (which, as previously stated, informs the dataset) is included in the dataset via linked data in the form of a URL and that a greater amount of data and metadata can be found for each route, particularly descriptive data, on the route page, thus creating a complex web of data and metadata.

Overall, the metadata for this dataset can best be described as adequate. The main intended user of the data is a climber who is looking for climbing routes in Yosemite National Park. A climber who intends to climb at Yosemite can be assumed to have a strong background in route reading and rating standards, as well as being intimately familiar with their own skills and limits when climbing rock faces outdoors. If the user has this assumed knowledge when approaching the dataset, then the data could prove very helpful in selecting and locating routes to climb. If, however, a non-climber or novice is using this dataset (which should not be ruled out) then the data lacks a level of detail that would be desired. For example, knowledge of route features, such as if the route contains an overhang or mantle, or if the route is particularly “pinchy” or “reachy”, or if it is a slab or at an incline, would be extremely valuable to this type of user. Similarly, the length of a route, both vertically or horizontally, could be of great benefit to a National Park employee looking to better understand the climbing routes being circulated among the climbing community. Some of this missing metadata in the dataset can be found by navigating to the route page through the URL. This does bring into question whether the metadata and data on the route page should be considered in this evaluation and term project. For the moment, I will proceed with treating the route page as outside of the dataset given that it is not feasible at this time to find a method of effectively pulling metadata and data from the route pages into the dataset.

While there are areas for enrichment in the metadata, the method in the repository environment (Mountain Project) to find the dataset is very straightforward. A convenient, route finder tool on the main page of the website can be used to locate the dataset of all routes within Yosemite National Park.

While the Mountain Project has previously been referenced in articles about climbing culture and in pop-culture publications on the subject of climbing/bouldering, there have been no publications referencing or using the data in this dataset to the best of my knowledge. Searches across the largest academic journal repositories, such as JSTOR, Academic Search Complete, and Google Scholar, using key terms such as “Mountain Project, climb\*, boulder\*, and Yosemite turn up no results for the use of this data in the scholarly articles stored by these repositories.

This data set has the potential to make climbing in Yosemite more accessible by pulling together collective knowledge and I am excited to work on data curation for this dataset to add to the usefulness of the data.

Link to the dataset location online:

<https://www.mountainproject.com/route-finder?selectedIds=105833381&type=boulder&diffMinrock=1800&diffMinboulder=20000&diffMinaid=70000&diffMinice=30000&diffMinmixed=50000&diffMaxrock=5500&diffMaxboulder=21700&diffMaxaid=75260&diffMaxice=38500&diffMaxmixed=60000&is_trad_climb=1&is_sport_climb=1&is_top_rope=1&stars=0&pitches=0&sort1=popularity+desc&sort2=rating>