

Local Food Directories: National Farmers Market Directory

Dataset Background & Structure

The Farmers Market Directory lists markets that feature two or more farm vendors selling agricultural products directly to customers at a common, recurrent physical location. Maintained by the Agricultural Marketing Service. The dataset was downloaded as a CSV, it contains 56 columns and 8739 rows for which 8173 appear to be unique. It contains the market name, website, address, social media information, products, season times, and updated time for the entry.

Quality Issues

Some of the quality issues that we can see on a quick analysis are:

- Repeated entries (Markets with similar names)
- Missing social media
- Missing addresses
- Other missing values
- Different date formats
- No apparent data entry rules (constraints)

Use Case and Data Cleaning Goals

The dataset provides information to the consumer to look for a nearby farmers market, the website <https://www.ams.usda.gov> provides good filtering options, which can help you find what you need. However, due to the dataset's poor quality it might make you ask your neighbor or cousin because that might be easier.

We'll have to correct the missing values, format dates correctly, cluster repeated entries based on name and addresses and possibly setup data entry rules to avoid repetitions in the future. Once the cleaning process is complete, a possible use for the dataset would be to create an application that uses the location of the user and provides the nearest farmers market based on the coordinates available in the dataset.

As it stands the dataset is actually fit for use, and it contains a lot of useful information. I actually used the website to look for a farmers market near where I live and provided me with a few locations. Unfortunately for some other users, they might get duplicates, missing information, or not all of the farmers markets available and for that reason we need to clean the dataset.