

Group Member:

Yuwei Chen, Hui Lyu, Mingwei Gao

Dataset Chosen:

US Farmers Markets

Status Update:

Based on the initial inspection for the dataset, it is obvious that there are pretty much null values. The format of dates is not uniform. For the zip codes, they should be five-digit but some of them only have four digits. Fortunately, there is no duplicate FMID.

Our goals are:

1. Clean the dataset, as much as we could. Including uniforming the format of dates, validating the latitude and longitude figures for the location of the farm, using latitude and longitude figures to fill the vacant zip codes, etc.
2. Creating SQL queries to show those records violating integrity checks (ICs).
3. Creating SQL queries to return results such as how many farms in specific state, what payment types they accept and what products they produce.