**Dataset Title**: Pesticide Data Program 2021-USDA

**Summary:** The Pesticide Data Program (PDP) is a national pesticide residue monitoring program and produces the most comprehensive pesticide residue database in the U.S. The Monitoring Programs Division administers PDP activities, including the sampling, testing, and reporting of pesticide residues on agricultural commodities in the U.S. food supply, with an emphasis on those commodities highly consumed by infants and children.

See link below for the dataset which is not provided on GitHub due to size.

[**https://www.ams.usda.gov/datasets/pdp**](https://www.ams.usda.gov/datasets/pdp)

**Tech Stack:** SQL, Pandas

**Tool:** MySQL Workbench, Jupyter Notebook

**Objective(s):**

* Explore the dataset using SQL.
* Explore cleaning the dataset using SQL.
* Create queries below to analyse the dataset to generate insights using SQL
* Use Pandas as a control feature for the SQL analysis.

**Tasks:**

* Load the dataset tables into MySQL.
* Create a data model and normalize the database.
* Execute the following queries:
  + How many samples were tested?
  + How many distinct samples were tested?
  + Average number of samples tested per location
* Use Pandas to run a similar analysis of each query.
* Execute the query in a Jupyter notebook and save in a CSV file for further use.