Dataset:

The dataset was originally used by the authors of *The Min-Max Test: An Objective Method for Discriminating Mass Spectra* (The Min-Max Test: An Objective Method for Discriminating Mass Spectra | Analytical Chemistry) to classify mass spectra using the Min-Max test. It has been publicly released on the NIST Public Data Repository website (*PDR: Supplemental Data and Source Code for Min-Max Test Research*). The dataset contains mass spectra for various compounds, providing detailed information about their molecular structure and composition

This dataset is comprised of four different files: CM1, CM2, CM3, and Isomers. Each of these files is dedicated to different compound categories:

- CM1: Fentanyl 29 compound samples
- CM2: Synthetic Cathinones 59 compound samples
- CM3: Cannabinoids 49 compound samples

All CM files mainly contain three columns:

- #Point: A unique identifier in descending order.
- X (Thompsons): mass-to-charge ratio of compounds, measured in Daltons
- Y (Counts): Relative intensity or abundance of ions.

At the top of these columns, it provides notes indicating the formula applied to generate these data. Both the X (Thompsons) and Y (Counts) columns contain continuous float values. It is to be noted that, In CM1, CM2, and CM3, there are 10 mass spectra samples of each compound.

When combined into a single CSV for all compounds:

- CM1 contains 92,310 Rows, 3 Columns.
- CM2 contains 53,210 Rows, 3 Columns.
- CM3 contains 301,237 Rows, 3 Columns.

If the data from CM1, CM2, and CM3 are all combined, then the resulting dataset contains 446,766 rows and 3 columns.

The Isomers file, as the name suggests, deals with isomers and contains 9 samples. Like the CM files, it contains the same three columns (#Point, X (Thompsons), and Y (Counts)), with continuous float values for X and Y column. When combining all CSV files in this category, the dataset contains 49,189 rows and 3 columns.

Moreover, another categorical data was given in a file called Codename_withformulas. It had the compound name, its molecular formula, and how many of those compounds' samples are there in CM1, CM2, and CM3.