**Overview and initial assessment of the dataset (narrative and supplemental information). You should describe the structure and content of the dataset and quality issues that are apparent from an initial inspection. You should also describe a (hypothetical or real) use case of the dataset and derive from it some data cleaning goals that can achieve the desired fitness for use.**

**In addition: Are their use cases for which the dataset is already clean enough? Others for which it well never be good enough? You can speculate a bit here – but the rest of the project should focus on a “middle of the road” use case that requires a practically feasible amount of data cleaning.**

This dataset is from the New York Public Library (menus.nypl.org). It consists of over 45,000 menus and includes information on each menu such as what establishment the menu came from and what individual items are found on that menu. The dataset is split into several CSVs. Dish.csv contains information of individual items on a given menu such as name of the dish, description of the dish, menus it appeared on, prices it sold for and when the item appeared on any menu. Menu.csv contains information on each menu as a whole, e.g., name of the establishment, location of the establishment, language in which menu is written, currency accepted, and overall page count and dish count. MenuItem.csv contains additional information about when an item was added to the dataset and when it was last updated. PageItem.csv also contains additional information about menu items with a page number and location of each item within the dataset.

On a quick inspection, there are several common data cleaning problems that are evident. For instance, there are formatting inconsistencies between the names of the establishments and their location names (NEW YORK vs. NEW YORK, [NY] or PACIFIC MAIL STEAMSHIP COMPANY vs. PACIFIC MAIL STEAMSHIP CO.).

A simple use case for this dataset would be to map the distribution (location) of restaurants included in the dataset. This would require data cleaning to ensure consistent restaurant and location names. One can also extend this use case to include time (i.e., mapping the distribution of restaurants geographically over time). This would add additional constraints, such as searching for missing or nonsensical dates.

Another use case would be to track the most popular dishes over time. This would require using multiple CSVs (i.e., Dish.csv and Menu.csv) and would require similar data consistency as the above use cases.