NYPL Menus Data Cleaning by Team Data Clorox

1. OpenRefine Cleaning

* MenuItem.csv [Pending]
* Was unable to perform transform column “created\_at”, “updated\_at” to date due to large number of records
* MenuPage.csv
* Transformed id, menu\_id, page\_number, image\_id, full\_height, full\_width to number
* Menu.csv
* Clustering on Column “event” for grouping similar texts together
* Clustering on Column “name” for grouping similar texts together
* Common Transform Titlecase on Column “event”
* Clustering on Column “sponsor” for grouping similar texts together
* Common Transform Titlecase on Column “sponsor”
* Common Transform Titlecase on Column “venue”
* Clustering on Column “venue” for grouping similar texts together
* Common Transform Titlecase on Column “venue”
* Clustering on Column “place” for grouping similar texts together
* Common Transform Titlecase on Column “place”
* Clustering on Column “physical\_description” for grouping similar texts together
* Clustering on Column “occasion” for grouping similar texts together
* Removed Column “location\_type” due to all blank entries
* Removed Column “keywords” due to all blank entries
* Removed Column “language” due to all blank entries
* Clustering on Column “location” for grouping similar texts together
* Transformed page\_count, dish\_count, id columns to numbers
* Trnasformed date column to date
* Dish.csv
* Common Transform Titlecase on Column “name”
* Removed Column “description” due to all blank entries
* Transformed columns menus\_appeared, times\_appeared, lowest\_price, highest\_price, id to number
* Was unable to perform clustering on column “name” due to large number of records