

Data Wrangling

Trends in Restaurant Inspections in New York City Capstone Project

Vicente López-Díaz

From the data available in [Open Data](#) from the City of New York, with information of violations reported during inspection, for each restaurant inspected by NYC Department of Health and Mental Hygiene (DOHMH). I wrangled the data to calculate the change on violations reported in an inspection. The steps I took are:

1. Separate date into year, month and day columns. Keep only data from 2014, 2015, 2016 and 2017.
2. Keep only 'Inspection Types' with 'Cycle Inspections / Initial Inspection' and 'Cycle Inspections / Re-inspection'.
3. Make violation groups from two digits in 'Violation Code'.
4. Separate tables Initial Inspection and Re-inspection.
5. Include a violation reported variable, to get the number of violations in each group.
6. Pivot tables to get violation groups in columns and inspection in columns. Missing values correspond to zero violations reported in that inspection. For both Initial and Re-inspections.
7. Join the Initial and Re-inspection in one table to put initial and its re-inspection in one row.