New York Inspection CSV

The New York csv originally came with 399918 rows with 18 columns. The values in the zip code column containing a decimal place and the address being separated between building number and street. ‘DBA’ represents the name of the establishment, ‘CAMIS’ being the code assigned to the establishment and ‘RECORD DATE’ being the same throughout the data frame.

* First started by making a copy of the original data frame so that the original data frame is not manipulate it in any way.
* Next, dropped any duplicates and filled in any missing value under the ***zipcode*** column with 0 and defined all the values within that column as integers.
* Then, simplified the ***building*** and ***street*** columns into a new ***address*** column by combing the 2 columns.
* After organizing the columns, a .value\_counts() was performed for the **RECORD DATE** column and got only 08/28/2017. I unselected the RECORD DATE column because it only describes the date the inspection results recorded into the file which is not needed.

San Francisco Inspection CSV

The San Francisco csv had 53973 rows with 23 columns. The latitude, longitude, location columns were filled with null values and the **inspection\_id** was a combination of the building id and the inspection date. The **violation\_id** was also a combination of the inspection date along with the violation id. The **inspection\_date** included the timeframe, but all its values were at 0.

* First, dropped all duplicate rows and got rid of unnecessary columns.
* The phone numbers within the column contained decimals which signified that the phone numbers were floats, so I turned those into strings to reveal the whole phone number. I later removed the decimals from the values.
* Converted datetime from the inspection date column to just show the date.
* Removed “business\_” from the **address**, **city**, **state**, **postal\_code** and **phone\_number** columns to simplify and shorten the length of the data frame.
* Looking at the **violation\_id** column, the actual id were the last 6 numbers for the element, so I cut the string to where the last 6 numbers were present, removing the **inspection\_id** values from it.