No duplicate row is found in OEM Emergency Notifications and NYC coronavirus data.

OEM Emergency Notifications dataset has 16827 entries.

NYC coronavirus dataset has 54 entries.

1728 rows in OEM Emergency Notifications dataset notification title is blank

One row in the OEM Emergency Notifications dataset date_and_time might be wrong.

steps

First, use pandas read .csv format OEM Emergency Notifications dataset.

Then get a high-level overview of the data types and the number of null values. There are 16827 entries in the OEM Emergency Notifications dataset. And 523 of them have a null value in the email_body column. All of the data types are object.

Because email_body is about notification detail and has some of them are empty, so we drop this column.

Check duplicated entries and keep the first one. And did not find duplicated entry.

Convert the data type of notificationtype from object to category.

Convert the data type of date_and_time from object to datetime.

Because we do not plan to use time, so we only keep date and delete time in date_and_time. Rename date_and_time as Date.

Count the number of occurrences of different notification types. Found there has no wrong value, and the least type has occurred 16 times.

Count the number of occurrences of different notification title. [blank] occurs 1728 times. Convert [blank] to np.NaN

Check date range and find min value is 1900-01-01, which seems like a wrong value.

Use pandas read .csv format NYC coronavirus dataset.

Drop HOSPITALIZED_CASE_COUNT and DEATH_COUNT

Rename DATE_OF_INSERT as Date and rename NEW_COVID_CASE_COUNT as Newcase.

Keep date and delete time in Date.

Have an overview of this dataset. There are 54 entries and have no null value.

Check the range of date, min is 2020-03-03, max is 2020-04-25.

To do integration, we select the entries after 2020-03-02 and join these two datasets together.