WeRateDogs Report

The WeRateDogs Twitter datasets had many ways to wrangle the data through three datasets differently.

Firstly, start with gathering the data with different ways by uploading data or importing it into the notebook by Pandas functions.

secondly, I started to look into the data with two different views,

- one is visually by looking into the structure or the information or description of the datasets to find out any anomalies in it then take an action according to cleaning the anomalies.

- Then I started to look programmatically through the datasets by looking into the null or N/A values, looking into the duplicated data or unlogical data or not accurate numbers that might be too far away from the realistic numbers we are assessing.

Thirdly, starting to clean the data and choose what is the action should be done whether it removing this data or changing or replacing it with appropriate values, beside removing the redundant data which we will not use for our analysis or visualizations as this take a storage for dataset and consume a space of memory, and clean step is cascaded into three steps as follow:

-By defining which columns, rows or values will be defined as not good data and which are food.

-Then I take the steps for coding the actions I defined to clean those bad data and shape the dataset to be analyzed properly.

-After that started to test is the data is more readable or insightful to take it to the analysis and visualizations steps to get the insights out of it or not.

Additionally, before starting the analysis theses datasets have to be consolidated in one data frame as one dataset to get the best of analysis from it and to reduce headache on transforming or moving from one dataset to another.

Finally, data was ready to start scraping for insights through visualizations, and look for distributions of the categories in dataset.