My wrangle Report

I took a systematic approach to how i organised my wrangle act file with sections in markdown cells so i will try to do the same here!

Importing important libraries:

I imported the essential libs like pandas for the assessing and wrangling And requests for gathering from the twitter api, but afterwards i started to add the important libs that had a use in a specific cell or IS the main lib in some of the sections like matplotlib in the last section

1st Section: Gathering

In the 1st section i started gathering the data and followed the rubric thoroughly so got the data from 3 different sources which were
1.twitter_archive_enhanced.csv
2.image_predictions.tsv
3.tweet_json.txt
So after i have gathered all of the data i got into the 2nd section

2nd Section: Assessing

So this section was the largest as i had been working on finding quality and tidiness issues in every file so i used all the basic functions with goals in mind i found inaccuracies in the names of the dogs i found alot of strange things like source column having htm tags so it was an adventure to say the least and i had to look for code in decumntaries and used substack with ai getting the code explained so i could make mine also i made a very good markdown system that has 2 main headings

which were the visual and programmatic assessment and each on of them had 3 subheadings which were the names of the files to be explored:1.twitter_archive_enhanced.csv
2.image_predictions.tsv
3.tweet_json.txt

The documentation:-

After i had found all i could find i documented every little issue and made sure to fit the rubric criteria, i also tried to improve the project by finding more than expected as suggested

Before the cleaning section i read from the content that i should copy the df of the files in other df variables that should be cleaned afterwards so i did that

3rd Section: Cleaning

I started using the rubric as the standard again and used a framework of the define code test and worked on it so i took the define from the documentation and modified it to the need and made the code to fix the issue the copied the code from the assess section sometimes to test and other times i just tested using other code made

Afterwards i merged the dfs using the merge function and made it look clean by dropping useless columns

Note: as per the reviewer's suggestions i changed the layout of the define code test format to look more clean and that was amazing

I apolagize and admit that it was unnecessary to replace nan in the name column and so i returned it back to how it should be

Also thanks to the reviewers suggestions i realized that all the inaccuracies in the name column were in lower case so i don't need to make list rather i will filter them out using the functions like islower

4th Section: Storing

I stored the data and saved it in a master df in csv
Note: i made indexing to false as per the suggestions

5th Section: Analyzing and visualizing data

I tried to ask interesting questions that had interesting insights as well
So i did that and was shocked by some of the results!!
Note: i used a trendline as per the suggestions which further enhanced my previous insight.

6th Section: Reporting

And this is this is this section! I hope that my project was fun to you as it was fun for me!!