ABSTRACT
This project aims to perform an in-depth analysis on the datasets given. We aim to understand our dataset before any further action can be done, for example, any algorithms formulated. In some instances, Data science is not really needed to solve all 5 challenges we encounter. As such, it is essential to analyze the problem statement and the data before we use personnel and finances to perform a job that could have been done efficiently and using less finances. Data wrangling needs to be done in instances where data science is used to ensure the data is up to standard before analysis even begins. The project aims to expound on data wrangling.

CONTEXT
Dog rating is a big idea on social media apps that generally is done for fun. The dog rating idea is usually supposed to be on a scale of 1-10, but unfortunately, a significant number of people rate the pictures over 10. The data is also from different sources. As such, our data scale needs to be assembled, and since it is majorly littered with unreasonable ratings, we need to sort it out.

The attached link entails a Git repository with the Data Report and Analysis.
https://github.com/Kathi3/fictional-palm-tree

CRISP-DM
In this Wrangle_Act project, the CRISP-DM model is used, which is a leading data mining methodology. CRISP-DM has provided the guidelines that have enabled the execution of this project in an organized and transparent manner. This model groups all tasks into six phases, as listed below.
- Business understanding
- Data understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

3.1. BUSINESS UNDERSTANDING

The main task is to use the data provided to answer the questions we set out to understand and visualize the data to make it easier for non-data scientists to understand and follow our work. Our resources are as seen below.

- Personnel- an aspiring data scientist
- Data - Provided Dataset (tmdb Movies)
- Software- Jupyter Notebook

DATA EXPLORATION The idea of this section is to create visualizations that can actually prove or disprove the objectives we set out to achieve to help us make more informed decisions. The idea is also to have diagrams that are easy to follow for people who may not be in the DS field.

CONCLUSIONS & RECOMMENDATION

From the descriptive analysis of the data, we can see that we have achieved what our objectives were.

1. To Load the Data and Access it using 3 methods - Able to load the data in 3 methods. Through the csv, tsv, and the Twitter API.

2. Combined the files to make 1 data frame.

3. Tidied up the Data to get a cleaner Dataframe - Through removing duplicates, removing the tenses in tweets

4. Detected and corrected 8 quality issues.

5. Visualized the data

Recommendations -
1. The Twitter API process is a great learning process. Long but it helps unleash an avenue to analyze data on twitter and build more projects..
2. The guideline was helpful but didn't necessarily reflect the actual work as some of the codes didn't work. More needs to be done to help in the process. The cleaning process is tedious. Allow yourself to pace it at least, else it feels overwhelming.