Wrangle Report

Introduction :

This Wrangle and Analyze data project was the 4th project of Udacity's Data Analyst Nanodegree program.It involved wrangling and analyzing the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

Gathering Data :

This project involved gathering data from 3 sources :
1.Csv file - **Enhanced Twitter Archive**
2.Tsv file - **Image Predictions File**
3.Api data - **Additional Data via the Twitter API**

Assessing Data :

After gathering data, the assessment on the data was done using following methods :

- head
- Info
- Describe
- value_counts()

Tidiness issue that were solved :

- Combining data from all 3 sources into 1 final table
- Combining 4 different variables(columns) into 1 named dog_type

Quality issues were cleaned :

- **Removing for null values**
- **Removing retweets**
- **Removing tweet replies**
- **Removing unwanted columns from the table**

- **Checking for duplicates**
- **Converting tweet_id into string format**
- **Converting timestamp into Datetime format**
- **Converting rating_numerator and rating_denominator into float**
- **Extracting the device type from source column**
- **Checking the numerator and denominator column**
- **Removing inaccurate ratings**
- **Checking the text column for more than 1 mention of dogs**
- **Reducing the doggo , floofer , pupper and puppo columns into one column name dog_type**
- **Issues with the name**
- **Standardizing the rating system**

Visualizations :

- **Checking the source of all the tweets**
- **Checking the relation between retweets and favourite count**
- **Plotting a line chart for rating over the time**
- **Checking different types of dog types**
- **Checking Top 10 breeds of dog**

Conclusion :

I collected data from 3 different source , analysed the data and then cleaned the data. This project gave me insights about how to gather data from multiple source and how to improve the tidiness of the data.