Wrangle Report

Introduction

In this project I will put what I have learned from Udacity Data Wrangling lesson in Data Analyst Nanodegree program. First the datasets are wrangled from the tweets of the Twitter user @dog_rates or known as WeRateDogs. This Twitter user is known to rate people's dogs usually out of 10 and add a humorous comment of his own about people's dogs.

Project phases

- Gathering the data
- Assessing the data
- Cleaning the data

First Gathering the data phase:

The required data consists of three datasets

1- **Twitter archive enhanced**: which is CSV file provided by Udacity support team and downloaded from the classroom manually.
2- **Image predictions**: which is TSV file provided by Udacity support team and downloaded programmatically by using Requests library in python that request the file via URL then saved to my computer. The file contains predictions of what animal breed in the image.
3- **Tweet JSON file**: the file can by gathered using Twitter API to gather more information from WeRateDogs tweet using the IDs of his tweets to get Favorite counts and retweets counts using Tweepy library in python then store each tweet in JSON file unfortunately I couldn't get the twitter API so I used the txt file provided by the Udacity support team.

Second Assessing the data phase:

After gathering all the necessary datasets, I begin assessing the data visually and programmatically

1- Visually assessing the data by printing the datasets then checking the data for any quality issues and tidiness issues.
2- Programmatically assessing the data by using python code to assess with various methods like info and describe.

Then documented all the quality issues and tidiness issues.

Cleaning the data phase:

This is the last phase of data wrangling in it you clean the quality issues and tidiness issues found from the previous phase cleaning are done programmatically.

Conclusion:

Data wrangling is one of the most important phases of the data analysis process and it's a most have skill for any successful data analyst without it the data analysis could lead to false outcomes.