

INTRODUCTION

This is the wrangling report on how I gathered my three data set, identifying the data quality and tidiness issues and how I was able to clean the messy data up.

SUMMARY

On loading all the three (3) dataset for this project, I checked to confirm the shapes, identified the quality and tidiness issues of the data. Accessing the data manually and programmatically, I discovered

1. So many columns in the data contained missing values
2. Some missing values were misrepresented, with entries as **None** - This issue was resolved using the pandas `replace()` function, to replace all instances of the **None** value with **N/A** in the dataset
3. Tweet id is loaded as integer instead of string – this was also resolved using the `astype()`
4. The timestamp column is a string, instead of a date-time object – I used pandas to `datetime()` function to change the timestamp column to date-time.

I also used the `describe()` function to analyze the dataset. This is where I identify the following issues: o Abnormal rating's numerator and denominator.

CONCLUSION

A significantly large number of data observations were lost during the wrangling process. Interestingly, all the dataset is now clean and devoid of tidiness and quality issues. All pestering issues were promptly addressed, and the project documented