Wrangle Report
By: Bilal Almajnooni

Introduction

The objective of this project is to work on my data wrangling skills on which I have learned through Data Analysis Nanodegree from Udacity. The dataset in question belongs to WeRateDogs twitter account, from roughly 2015 till 2017.

Data Wrangling

Data wrangling consists of many parts, starting from gathering the data from any source be it programmatically through code or manually through downloading it from the internet, or acquiring it on a USB drive.

After that comes the assessing part, in which you view the data either programmatically using methods such as head(), info(), and describe() from pandas library in a meaningful and impactful way as summary. Or manually by opening the data in an excel sheet or a google spreadsheet and thoroughly examine it.

The third step is cleaning the data. Data cleaning is the part that takes the most time in finishing it, because it involves programming the assessed issues in the data and make it work somehow. The best way of going about in this process is to divide it into three parts: Define, Code, Test. Define part is all about changing our made assessments to well-defined cleaning tasks. Think of it like a pseudocode. The code part is straightforward: convert your well-defined pseudocode into code that works. The last part is about testing your code if it works or not.

After finishing all of that, you must store your master dataset as to maintain your finished work. Then comes the real work, visualizing your findings.

Analyzing and visualizing your master dataset is all about asking the right questions and using the correct way to convey it. One of the questions in which I had was about finding the number of tweets the account had posted. To visualize it, I used groupby() function and supplying it with the weekly count of tweets and plotting it on a line chart.

Conclusion

To conclude, I believe that data wrangling is the most important skill to have as a data analyst. The steps in which you wrangle data are critical and you must be able to apply each one correctly to produce good results.