



[◀ Return to "Data Analyst Nanodegree" in the classroom](#)

[DISCUSS ON STUDENT HUB](#)

Wrangle and Analyze Data

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

Congratulations!

EXCELLENT JOB ON THIS PROJECT! YOU HAVE MET ALL THE REQUIREMENTS OF THE RUBRIC.

It's a **dog eat dog** world out there, but take a moment to **paws** and congratulate yourself on **woofing** down this data challenge and passing all the requirements of our rubric! This was no small feat. You went through multiple ways of gathering, accessing, and cleaning data, and presented your results in concise, meaningful reports. Keep up the fantastic work as we go deeper into the modeling and analysis side of the data! It may get **ruff** so don't be afraid to ask for help!

Before you move on to your next lessons, take pride in the effort you've put into this project. I hope you found this exercise both challenging and rewarding. Keep up the exceptional work and effort here, and I look forward to seeing you rock those future submissions!

13/10

Code Functionality and Readability

All project code is contained in a Jupyter Notebook named `wrangle_act.ipynb` and runs without errors.

Your code is correctly contained in a Jupyter Notebook named `wrangle_act.ipynb`. Always a good start!

The Jupyter Notebook has an intuitive, easy-to-follow logical structure. The code uses comments effectively and is interspersed with Jupyter Notebook Markdown cells. The steps of the data wrangling process (i.e. gather, assess, and clean) are clearly identified with comments or Markdown cells, as well.

Nice work here in clearly identifying the sections between Gather, Assess, and Clean to provide the reader with a logical path of your work. Your code is well commented with appropriate print outs and markdowns to reflect each step of your work. This attention to detail really goes a long way to showcase your understanding of the material and demonstrates a skill set that is a key element of finding success in data science. Right on!

Gathering Data

Data is successfully gathered:

- From at least the three (3) different sources on the Project Details page.
- In at least the three (3) different file formats on the Project Details page.

Each piece of data is imported into a separate pandas DataFrame at first.

Excellent job here pulling data from the `twitter-archive-enhanced.csv` file, the neural network (tsv) pulled through the `requests` library, and leveraging Twitter's API using the `tweepy` library to explore the WeRateDogs Twitter archive! You have successfully gathered data from three different sources and in three different file formats on the Project Details page, a useful skill for any data guru!

Assessing Data

Two types of assessment are used:

- Visual assessment: each piece of gathered data is displayed in the Jupyter Notebook for visual assessment purposes. Once displayed, data can additionally be assessed in an external application (e.g. Excel, text editor).
- Programmatic assessment: pandas' functions and/or methods are used to assess the data.

Great job showcasing the data gathered in your notebook. I love that you used `DataFrame.info()` to get a concise summary of each `DataFrame`, as well as leveraging `value_counts()`. You could have also added to your work with `DataFrame.describe()`, to generate descriptive statistics that summarize the shape of the data distribution.

One tool I often love in my data exploration is `seaborn`. It makes finding data distributions a breeze and serves them up in perfect visualizations with a minimal amount of coding. You can find more information on

how to leverage seaborn in your analysis here:

<https://seaborn.pydata.org/tutorial/distributions.html>

At least eight (8) data quality issues and two (2) tidiness issues are detected, and include the issues to clean to satisfy the Project Motivation. Each issue is documented in one to a few sentences each.

Looks great! Excellent work here in finding data quality and tidiness issues from the tweets. You structured your issues well and it was easy to follow how you addressed these issues in the next section. While your list was not comprehensive (finding every potential issue with the data would require quite a bit of effort), and there were a number of other areas that could have been cleaned, you have nonetheless satisfied the requirements in the rubric. Woo!

Cleaning Data

The define, code, and test steps of the cleaning process are clearly documented.

I really appreciate your documentation of each of the cleaning process, with markdowns for **Define**, **Code** and **Test** steps well labeled. Terrific work.

Copies of the original pieces of data are made prior to cleaning.

All issues identified in the assess phase are successfully cleaned (if possible) using Python and pandas, and include the cleaning tasks required to satisfy the Project Motivation.

A tidy master dataset (or datasets, if appropriate) with all pieces of gathered data is created.

Nice work leveraging `df.copy()` here prior to cleaning up the data! We definitely want to retain the original data in case mistakes were made along the way. You've also successfully managed to clean up the issues first identified in the prior assessment section. I'm sure you recognize that this took a non-trivial amount of effort and represents a key component of our wrangling process.

Storing and Acting on Wrangled Data

Students will save their gathered, assessed, and cleaned master dataset(s) to a CSV file or a SQLite database.

Well done here in using `.to_csv()` to store you cleaned data to CSV.

The master dataset is analyzed using pandas or SQL in the Jupyter Notebook and at least three (3) separate insights are produced.

At least one (1) labeled visualization is produced in the Jupyter Notebook using Python's plotting libraries or in Tableau.

Students must make it clear in their wrangling work that they assessed and cleaned (if necessary) the data upon which the analyses and visualizations are based.

Good job here in coming up with interesting insights from your cleaned data results. Your visualizations were spot on and demonstrated your ability to translate cleaned data to actionable output. Terrific!

Report

The student's wrangling efforts are briefly described. This document (wrangle_report.pdf or wrangle_report.html) is concise and approximately 300-600 words in length.

Terrific job with your report here! Your report is well formatted and easy to read, with the Gather, Assess, and Clean sections clearly marked. I appreciated the detail with which you elaborated on some of the work you've done in this project, crafting a summary of your work so that the audience can get an overview of the processes used in your analysis.

The three (3) or more insights the student found are communicated. At least one (1) visualization is included.

This document (act_report.pdf or act_report.html) is at least 250 words in length.

As with your `wrangle_report.pdf`, you crafted a structured report to showcase the results of your work. You gave some important insights here in your `act_report.pdf`, and I appreciate the level of specificity you went through to detail these outputs. Your word count was a little light here, but I let it slide since your insights were clear and straight forward.

Project Files

The following files (with identical filenames) are included:

- wrangle_act.ipynb
- wrangle_report.pdf or wrangle_report.html
- act_report.pdf or act_report.html

All dataset files are included, including the stored master dataset(s), with filenames and extensions as specified on the Project Submission page.

 [DOWNLOAD PROJECT](#)

[RETURN TO PATH](#)
