



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

[DISCUSS ON STUDENT HUB](#)

Wrangle and Analyze Data

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

Congratulations 🎓

You've made it!! 🎉

Wrangle and Analyze Data is a project that requires a lot of time and effort but you have been up to the task and you did everything very well! Data science is one of those fields in which you never stop learning and it can sometimes be frustrating to spend hours and hours re-examining problems, but I assure you that once you learn the basics well it's easier to learn more complex concepts and this it is one of those jobs that gives enormous satisfaction! I'm sure you'll do great things! Never stop believing in yourself and in your skills!

I wish you all the best with your Nanodegree and your career!



...some resources that might inspire your data journey and help you to improve your fantastic skills 🚀 :

- Regular articles that explore various topics and tell stories through visualization: <https://flowingdata.com/>
- Amazing website with full interactive data visualizations: <https://pudding.cool/>

Code Functionality and Readability

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

Great job checking that every code cell runs properly and fixing the previous code errors, now all the code is present in the `wrangle_act.ipynb` notebook and run without errors. Awesome! 💪

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.

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.

Data is successfully gathered from three different sources and in different formats, plus It's great that you have applied the suggestion from the previous review, hiding your API credentials! 👍

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.

A very nice work here! Both visual and programmatic assessment are done properly in the notebook! Good job applying almost all suggestions from the previous review in order to explore more about each piece of data gathered using different useful pandas functions and external applications! 👍

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.

Great job here, now both data quality and data tidiness issues are properly detected, plus each of them are documented in a very clear and detailed manner, well done! 💪

Cleaning Data

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

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.

Well done using the `copy()` function, making a copy of the datasets, before cleaning the data! 👍

```
[ ] 1 # Take a copy of twitter_archive on which the cleaning tasks will be performed
    2 archive_clean = twitter_archive.copy()
    3
    4 status_df = status_df_original.copy()
    5 image_prediction = image_prediction_original.copy()
```

Cleaning issues

Excellent job applying almost all suggestions included in the previous review, indeed now almost all the issues detected are correctly cleaned, no duplicates are still included and the multiple dog stages are properly handled. Being able to interpret well and apply the suggestions of others allows you to always improve your skills by working in a team, it is a skill that not everyone has because you have to be able to get in the position to learn and recognize that the other has more experience than us. Remember that this is one of those fields where you never stop learning and improving and this is one of those skills that really

make a difference in the world of work! Keep it up! 💪

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.

Great job fixing the code error, now the master dataset are correctly saved to a CSV file!

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.

Amazing job fixing all the issues and updating properly each insight and visualization produced. Now this section meets all the specifications!

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.

You have made a very nice wrangle_report document, concise, clear and interesting to read. Well done!

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.

It's great that you have not only fix the issues and updating the insights but you have also enhanced the entire document including numerous references such as images, personal thoughts and experiences, all these details help to increase the engagement with the reader. Keep it up!

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](#)