U D A C I T Y

< Return to "Data Analyst Nanodegree" in the classroom

Wrangle and Analyze Data

REVIEW
HISTORY

Requires Changes

2 SPECIFICATIONS REQUIRE CHANGES

Student, good submission. Good job on wrangling, cleaning and visualizing your data! You just have a few more points to finish and you should be able to pass this project.

Code Functionality and Readability

All project code is contained in a Jupyter Notebook named wrangle_act.ipynb and runs without errors.
Nicely done. Files submitted correctly, and your notebook doesn't show any execution errors.

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.
Student, your notebook as an easy to follow structure. You use the **gather assess clean** process quite well. Good work. This is important for when readers look at your work (or even yourself). It helps them to follow along.

TIP

One tip as well would be to include these steps in a table of contents at the beginning of your notebook. See this link for how to do this.

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.

You successfully gathered data from three different sources: local file (xls), URL, and programmatically extraction of JSON using an API. You have also stored correctly the gathered data in a format according the project instructions.

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.**

Student, good job on assessing your dataset. You have used `head` , `tail` , `info` , `value counts` quite well. Another useful function is `describe` .

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.

Student good job so far. Please note however that for the dog stage cleaning issue. Instead of having None as a value you should only have dog stages. Of course there are some that have multiple in one row, so for these you can use the value as 'mutiple'. Either the output can be multiple, or the output can have two dogs as one value.

```
df_1_clean.stage.value_counts()
```

```
pupper          224
doggo            75
puppo            24
multiple         12
floofer           9



Name: stage, dtype: int64
```

[54]: `df_1_clean.stage.value_counts()`

```
[54]:  pupper              224
       doggo                75
       puppo                24
       doggo,pupper         10
       floofer               9
       doggo,puppo           1
       doggo,floofer         1
       Name: stage, dtype: int64
```

Cleaning Data

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

The **define code test** steps have been used successfully.

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.

✅ Archive dataframe was copied before being cleaned.

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.

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 detailing insights about the data. Nice job with the visualization.

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.

✅ Act report completed successfully. Good work.

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

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.

Student, the rubric states that at least one visualization should be included in this report. Good details included, but please include visualizations. This report should be like a magazine or blog post, so feel free to include

pictures of tweets, dogs, or whatever else will make it fun and exciting.

Project Files

The following files (with identical filenames) are included:

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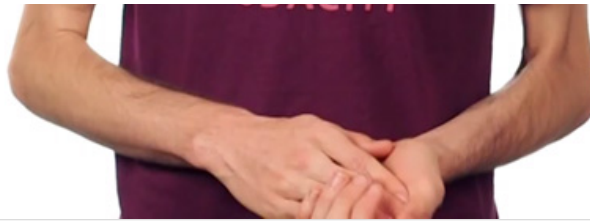- **act_report.pdf or act_report.html**

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

✅ All files included.

☑ RESUBMIT

⬇ DOWNLOAD PROJECT

Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

⊙ Watch Video (3:01)

RETURN TO PATH