

< Return to Classroom

Wrangle and Analyze Data

REVIEW

HISTORY

Meets Specifications

Congratulations on completing the project!! You've worked on the comments from earlier review... well done I think the earlier reviewer stated the points just for having a ease of reading...

Distinction between data quality and data tidyness

QUALITY
Completeness. i.e: Do we have all of the records that we should? Do we have missing records or not? Are there specific rows, columns, or cells missing?
Validity. i.e: Do they conform with real-world constraints? (e.g. negative retweets, favorites, etc.)
Accuracy issues. i.e: Incorrect data like denominator ratings greater than 10 for example
Consistency. i.e: Make sure the columns all have a standard format. in your case details about dog tweets and metrics
TIDINESS
Making sure every column is a variable
Making sure every row is an observation
Making sure every cell is a single value

Code Functionality and Readability

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

Gathering Data

✓ Data is successfully gathered:

- From at least the three (3) different sources on the Step 1: Gathering Data page.
- In at least the three (3) different file formats on the Step 1: Gathering Data page.

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

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.

✓ 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.

Cleaning Data

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

Good work following code -> test strategy

✓ 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.

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.

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.

✓ 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.

Excellent work sharing the insights and visualisations

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.

Good work sharing the pdf versions of act_report.pdf and wrangle_report.pdf

📄 DOWNLOAD PROJECT

Rate this review

RETURN TO PATH

START