UDACITY

‹ Return to "Data Analyst Nanodegree" in the classroom

DISCUSS ON STUDENT HUB

Wrangle and Analyze Data

REVIEW

HISTORY

Requires Changes

2 SPECIFICATIONS REQUIRE CHANGES

Hi there,

Awesome work on the project. All the segments are working fine. Just one issue is there that `act_report.pdf` file is missing.

Please include the file in the next submission with required details and I am confident the project will pass.

Code Functionality and Readability

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

All the code is present in the `wrangle_act.ipynb` notebook and run without errors. Good work on checking that every cell works correctly.

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.

It's great to see that you have organized the notebook in the 4 distinct sections of GATHER / ASSESS / CLEAN and ANALYZE. The notebook is interspersed with code and markdown text. This helps anyone in following

along the work and can also understand the process flow that you have taken.
Nice job.

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 all the 3 sources and is saved to file locally.

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.**

Both visual and programmatic assessments are done in the notebook.
Nice job on using the functions like `info(), describe(), value_counts(), sum() and duplicated()` to explore more about 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.

Issues have been found and classified into Quality / Tidiness issues.

All the issues are classified correctly. Awesome.

Cleaning Data

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

The DEFINE / CODE / TEST steps are clearly documented in the cleaning section. This helps us a lot in identifying each issue and how it's cleaned and tested.

Suggestions

You can make them more prominent by using markdown headings to identify these sections.

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.

Copies of the dataset are made prior to cleaning. That's an important step as we may need to refer to original dataset later on.

All the issues identified are cleaned.

Suggestions

For the ratings numerator, don't directly consider `rating > 15` as wrong. if some rating feels wrong, then it should be matched with the text and if they don't match then they can be 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 gathered/cleaned data is 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.

The master dataset is analyzed using the pandas and insights and visualizations are given.
Good job.

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.

Good work on creating the report for the wrangling efforts. It's clear and concise and reflects the wrangling process taken for the data set.

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.

The `act_report.pdf` file is missing. Please include this report in the next submission.

This document should not be coding but a report.
Make sure this report contains

1. At least 3 insights
2. At least 1 visualization
3. The Visualisations are explained
4. The report length is at least 250 words.

Suggestion:
Try to consider this report as a blog post. You can add images of dogs, tweets, ratings to make it a fun read. Anything to get the reader engaged.

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.

Please include the `act_report.pdf` file in the next submission.

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