UDACITY

< Return to Classroom

Wrangle and Analyze Data

REVIEW

HISTORY

Meets Specifications

Hi,
This is a very good submission. Well done. The code is well written and easy to follow. Some additional comments could be inserted in some sections but overall I find it appealing. Please keep up the good work.

Additional Resources

- I'd like to invite you to look at this link on this article on data wrangling

Code Functionality and Readability

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

Great job, the code runs with no issues

Suggestion

- In case you get errors or you feel the need to debug your code, please take a look at this post that goes into it debugging tips.

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.

The notebook is appropriately formatted. Markdown cells are well placed and there are some intuitive code comments. This makes your work easy to follow.

Suggestion

- A table of content can also be added in your files. This will make navigation even easier. Take a look at this link to have an idea on how to do this

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.

Each piece of data is successfully gathered and separated into pandas dataframes. The data files are in 3 different file formats. Well done.

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.

Good job here. You have effectively performed visual and programmatic assessments.
Note that this is one of the most important parts of the data wrangling step. This gives you some insight on the data you are about to work on.

Tip

- Some more functions that could be used for assessment are:

- unique() The `unique()` function is used to get unique values of Series object.
- isna() This indicates whether values are missing
- info() This method prints information about a DataFrame including the index dtype and columns, non-null values and memory usage.

- columns This produces a list of all the columns in the dataframe
  There are many more. Please feel free to explore when you can.

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.

Well done in identifying at least 8 quality issues and 2 tidiness issues

Cleaning Data

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

The different steps in cleaning your project have been clearly defined.

Suggestion

- To make this stand out a little more, you could perform a test before cleaning in order to show that the results are different after cleaning.

Tip

- Take a look at the testing assert statement It helps ease testing by introducing python familiar asserts.

Note

- Fully cleaning the entire dataset requires exceptional effort so only a subset of its issues should be cleaned. The ones that are cleaned need to be cleaned correctly.

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.

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 the necessary files are present in the submission.

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

START