UDACITY

‹ Return to "Data Analyst Nanodegree" in the classroom

DISCUSS ON STUDENT HUB

Wrangle and Analyze Data

REVIEW

HISTORY

Meets Specifications

Good work! Thank you for addressing the **last reviewer's** comments.
You have provided a wonderful project :)
However, to improve your skill and knowledge for your journey of being an aUdacious data analyst, here I provide you with some comments 😉

Code Functionality and Readability

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

No error found, well done!

- If you want a very succinct cheat sheet for data wrangling using python, I think this will be very helpful for you :)

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.

I love how you have structured your project and commented on all complex code structures. This is a very good practice. As a reviewer myself, I found it very helpful to understand the code and how the code produces a correct/wrong result. In a workplace, such clear structure and a well-documented code will be very helpful for colleagues that might be continuing your work or learning from your work.

Keep doing the good practice!

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 have included the three data sources correctly with the correct methods :)

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.**

You have provided your assessments here :)

- But, it is better if want to separate the visual and the programmatic assessment in two separate headers. This way you show that you have conducted both assessments explicitly.
- Additionally, you can also assess the visually using other software such as Excel and Google Sheets as Jupyter Notebook sometimes deprecating the data that are being displayed.
- By the way, good work on using pandas functions for the programmatically assessment :)

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.

All issues have been mentioned correctly.

I appreciate that you have shown a further effort on identifying the issues more than required. I think it is one of the traits that a data analyst should have, curiosity 👏🏻

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.

You have copied the original dfs before you clean it :)
This is a good practice.
For further information why is it so important, please read this.

You also have cleaned all mentioned issues and merge the tables into a master table :)

Suggestion
For the dog stages, if you look closely to the individual text of multiple dog stage, there are some row that actually has one, but wrongly extracted as two stages.

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.

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review