UDACITY

PROJECT

Wrangle and Analyze Data

A part of the Data Analyst Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

Excellent job, I hope you have had lots of fun working on this project!

It's true that data wrangling is seen as a somewhat tedious task, but I hope you have found this project engaging, interesting, and fun. I like the way how you identified and cleverly solved some issues in the dataset. Keep up with the good work, and please continue practicing and learning many other functions and tools to improve and perfect your data wrangling skills. Be Udacious!

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.

Your notebook is clean and structured. Different sections are clearly shown for each one of the steps of the data wrangling process. Also you detail appropriated comments were needed to explain what your code does and how it works.

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.**

You have correctly used functions like info, value_counts, describe, to perform the assessment of data. Also, it's always a good idea to use another application (like Sheets, Excel, etc) to perform a quick visual 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.

You have identified data quality and tidiness issues, and explained the action that should be taken to fix them. You have given
a clear and succinct explanation for each one of them. Just in case keep in mind, that there is no definitive list, there are definitely other different points to consider, but you have identified the most important issues.

One issue you missed:
Check the numerator - in the source data there seems to be fractional numerators. Are those preserved when exporting the data? Take a look at the id = 681340665377193000

You can read this description:

> I've been told there's a slight possibility he's checking his mirror. We'll bump to 9.5/10. Still a menace

As you see, the numerator should be 9.5 - but in the numerator it's extracted as "5" (a wrong regular expression was used to extract it).
You could fix that issue also using a regular expression, to extract "9.5" instead of "5".

Something else: it would be better to unify all different dog stages column (doggo, puppo, etc) into a single column - this is a tidiness issue.

Cleaning Data

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

You have correctly performed the cleaning data step of the data wrangling process. The define-code-test framework has been correctly applied.

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 copied the original data before cleaning (important if at some point you need to trace back on your steps), and also you created the final clean master dataset. Great job!

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.

Master dataset correctly stored into 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.

Good job detailing insights about the data. Great and relevant visualizations, clearly labeled and engaging

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.

Nicely done, it's a great "executive report" style document. To the point, concise, and clear.

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.

This is not only an interesting, but also a very fun report! You don't only "tell" but also "show" the insights found

Notice the html isn't showing properly the plots - that's why it's always better to export as a PDF

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

Student FAQ