UDACITY

Wrangle and Analyze Data

A part of the Data Analyst Nanodegree Program

PROJECT REVIEW
CODE REVIEW
NOTES

SHARE YOUR ACCOMPLISHMENT! 🐦 📘

Meets Specifications

Code Functionality and Readability

> **All project code is contained in a Jupyter Notebook named wrangle_act.ipynb and runs without errors.**
>
> Well Done for implementing the code that read and clean the data with mo error. All the code blocks perform as expected with no 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.**
>
> The code is well formatted and appropriately commented. That make it easy to follow the analysis steps and identify a specific functional operation.

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.**
>
> Well Done for reading the data correctly. While `.info()` `is an excellent function that provides very useful information, you can also use` .shape``` that return only the dimensions of the data frame.

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.**
>
> The code address several important and meaningful issues that need to be corrected or modified. Each one of this issues is being described and explained.

Cleaning Data

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

For some of the code, you can include more comments that explain what the code performed. When writing a code, the documentation should be such a reader that is not familiar with the programing language will be able to understand what has been performed.

In addition as part of the test or after the test you can show a sample of the field that was cleaned before and after. For example, how many "retweets" that data include before the cleaning?

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.

This is important since cleaning requires several iterations so it is good that you have an option to go back to the data before cleaning it if something went wrong.

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.

Excellent! All the chart include axis labels title and description. That makes it easy for the readers to follow the line of your thoughts and understand the dataset and your analysis.

Report

The student's wrangling efforts are briefly described. This document (wrangle_report.pdf) 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) is at least 250 words in length.

Project Files

The following files (with identical filenames) are included:

- wrangle_act.ipynb
- wrangle_report.pdf
- act_report.pdf

All dataset files are included, including the stored master dataset(s), with filenames and extensions as specified on the Project Submission page.

⬇ DOWNLOAD PROJECT