UDACITY

Back to Data Analyst Nanodegree

Wrangle and Analyze Data

REVIEW

HISTORY

Meets Specifications

Avid Udacity Student,

Congrats! You passed the project on your first submission. Quite impressive! 🙌 This is a great piece of work and tells a ton about the type of person you are - organized, hardworking, and quality oriented. Going thoroughly through the work, I could see a lot of time and effort invested in the project, and I think this is commendable. I exhort you to keep up this good work as it will make you an outstanding Data Analyst. Keep learning and keep up the good work! 😄

Code Functionality and Readability

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

Awesome, all code is functional and runs free from significant error.

Learning Notes

I am a fan of using shortcuts with Jupyter Notebook. Check out this medium post on **Jupyter Notebook Shortcuts**.

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.

Great job on making the work very clear and easy to follow. Markdown cells are used to clearly indicate different sections of the notebook. Apart from markdown cells, the project also makes use of inline python comments to make the work clearer. Nice!

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.

Good work gathering data from three different sources and converting all of them into python data frames before performing wrangling.

Data was gathered from the following sources:

- Twitter-archive **CSV** file.
- Image prediction **TSV** file.
- **tweet_json.txt** file containing retweet counts and likes.

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.**

Well done! Both visual and programmatic assessments are used in the notebook and the results are well documented.

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.

Good work identifying 13 quality issues and 3 tidiness issues in the dataset!

Cleaning Data

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

The different steps of the cleaning process are clearly documented. We have the `define`, `code` and `test` steps which are clearly stated with some explanations of what process you intend to do at each level. Excellent work!

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.

Good work using pandas.DataFrame.copy function to make a copy of the datasets before cleaning.

```
#creat copies of the dataframes to clean
df_archive_clean = df_archive.copy()
```

```
df_breeds_clean = df_breeds.copy()
df_tweet_clean = df_tweet.copy()
```

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 cleaned master dataset is saved to a csv file. Good work!

```
#save the merged table in CSV file named 'twitt
er_archive_master.csv'
df_main.to_csv('twitter_archive_master.csv')
```

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 work analyzing the cleaned data! The notebook contains four separate insights which are well described in the report and in addition, several visualizations have been produced. 

Learning Notes

- Check out the python visualization documentation for various ways of visualizing data.
- There are several other ways to visualize data including **Box plots, Line graphs, Pie charts. Check out this documentation on Visualization with Seaborn.**

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 write-up (`wrangle_report.pdf`) is very detailed and within the limit required. Awesome job!

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.

Eight interesting insights have been reported in the `act_report.pdf` and they have been analysed in details. All information is well communicated and the write-up is more than 250 words in length. Awesome work!

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 required files are present. The submission contains the `wrangle_act.ipynb` with all the necessary code, the `wrangle_report.pdf` which contains a brief discussion of the wrangling efforts made in the project with the use of 3 steps which are Gathering Data, Assessing Data and Cleaning Data. Finally, we also have the `act_report.pdf` which is mainly a discussions on the insights discovered in the project. Brilliant!

DOWNLOAD PROJECT

RETURN TO PATH