**Wrangle and Analyze Data**

**Meets Specifications**

**CONGRATULATIONS !!!!**

**Code Functionality and Readability**

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

Code submitted reflects work done in the submitted document.

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 organization on the notebook. Very easy to read through.

**Additional Recommendations**

* I recommend reviewing python standards for functional programming as this will be very helpful for your future.  
  [Functional Programming HOWTO](https://docs.python.org/3/howto/functional.html)
* There are several different ways to annotate your code.  
  [How To Write Comments in Python](https://www.digitalocean.com/community/tutorials/how-to-write-comments-in-python-3)
* Finally, review the [PEP 8 -- Style Guide for Python Code](https://www.python.org/dev/peps/pep-0008/)

**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.

All elements imported correctly!

**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.

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.

At least 8 Data Quality issues inspected and written about.  
At least 2 Tidiness issues inspected and written about.

**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.

I cannot overstate how important it is to create copies. Great work here.

**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.

**Advanced Visualization Packages**  
I would recommend playing around with the following packages.

1. seaborn  
   a. Install in pip pip install seaborn  
   b. <https://seaborn.pydata.org/examples/index.html>
2. plotly  
   a. Install in pip pip install plotly  
   b. <https://plot.ly/python/>

**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.

**Great** explorations into the data. Start thinking about external variables and possible expansions to the investigation.

**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.