PROJECT SPECIFICATION

Wrangle and Analyze Data

Code Functionality and Readability

CRITERIA	MEETS SPECIFICATIONS
The student's code is functional.	All project code is contained in a Jupyter Notebook named wrangle_act.ipynb and runs without errors.
The student's code is readable, i.e., uses good coding practices.	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.

Gathering Data

CRITERIA	MEETS SPECIFICATIONS

CRITERIA	MEETS SPECIFICATIONS
The student is able to gather data from a variety of sources and file formats.	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.

Assessing Data

CRITERIA	MEETS SPECIFICATIONS
The student is able to assess data visually and programmatically for quality and tidiness.	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.

CRITERIA	MEETS SPECIFICATIONS
The student is able to thoroughly assess a dataset.	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.

Cleaning Data

CRITERIA	MEETS SPECIFICATIONS
The student uses the steps in the data cleaning process to guide their cleaning efforts.	The define, code, and test steps of the cleaning process are clearly documented.
The student is able to thoroughly clean a dataset programmatically.	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.

Storing and Acting on Wrangled Data

CRITERIA	MEETS SPECIFICATIONS
The student is able to store a gathered, assessed, and cleaned dataset.	Students will save their gathered, assessed, and cleaned master dataset(s) to a CSV file or a SQLite database.
The student is able to act on their wrangled data to produce insights (e.g. analyses, visualizations, and/or models).	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

CRITERIA	MEETS SPECIFICATIONS

CRITERIA	MEETS SPECIFICATIONS
The student is able to reflect upon and describe their data wrangling efforts.	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 student is able to describe some insights found in their wrangled dataset.	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

CRITERIA	MEETS SPECIFICATIONS

CRITERIA	MEETS SPECIFICATIONS
Are all required files included in the student's submission?	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.

Suggestions to Make Your Project Stand Out!

- Assess and clean more than the required issues (eight and two, respectively, for quality and tidiness).
- In act_report.pdf (or act_report.html), additional images beyond any visualizations are encouraged to make this report more engaging. Make this report as detailed as you'd like!
- Gather additional data beyond the required pieces.
- Create a model that can make predictions based on your wrangled data.