Machine Learning Engineering Bootcamp

Capstone: Data Wrangling

Step 5

Learning Objective

- Clean the dataset and address issues such as missing values and outliers.
- Apply data wrangling techniques, including standard datasets and sources.

Criteria	Meets Expectations
Time Estimates	15 - 20 Hours
Completion	❏ Data and code are uploaded to Github (where applicable).
Process and understanding	❏ The submission shows that the student understands how to **wrangle** and **clean data**. ❏ The submission demonstrates that the student made thoughtful decisions about how to work with missing values and outliers as needed.
Presentation	❏ Well-documented Github repository and code. The Jupyter notebooks for the code provide step-by-step documentation that's easy to follow.

Excellence:
1. *The student collected data from multiple disparate sources (APIs, CSVs), and extracted and merged the relevant data, enhancing the quality of the results.*
2. *The student explored the data systematically through visualizations that guided their next steps.*

For reference, review how this interim step fits into the overall Capstone Project Rubric.