

Self-Assessment: Percentile and Rationale

Author: Amantha Bhaskarabhatla

NUID: 002300618

I'd say this project sits around the top 10% of what others turned in. Picking that score because I didn't just include random tests - instead, I set up a full flow designed to catch garbage-in-garbage-out issues from start to finish. It pulls in fake yet lifelike data, checks its shape, uses preset rules to validate, gives out quality summaries, cleans data, then shows how those fixes affect later results like alert tags, score shifts, and trust markers.

The notebook is meant to teach formally, but through simple examples. So I break ideas into simple terms, provide notes and headings to keep you on track, while linking every code cell to the core message - rubbish data leads to rubbish results. The sample data is small so it's easy to dive in, yet packed with actual issues like gaps, outliers, incorrect labels, and repeats. There are step-by-step little tasks too: first make sense of outputs, then tweak validation rules, later add fresh columns and tests to grow the workflow.

Because of teaching goals, this links to key ideas from the course - GIGO, how good data should be complete, valid, or make sense, also questioning results. It moves step by step using Explain, then Show, then Try, matching how a teacher would go over it live.

I cannot say that I've hit the highest level since there's still space to improve - like bringing in deeper multivariate analysis, tossing in time-based elements, or linking the notebook to a basic interface or live tracking system. Still, thanks to solid ideas, straightforward explanations, and code you can actually use again, I feel safe calling this work way better than most, somewhere around the 85 to 90% mark.