**Task 1 report**

***d. Unstructured data can be more difficult to assess than structured data. Just as you did in part b, write 3 - 4 sentences that show how this unstructured data fails to check requirements of good quality data.***

The first two columns (v1 and v2) contain the useful data with labels and messages, while the other columns (Unnamed: 2–4) are mostly empty (99% missing) and lack a consistent format, making them difficult to use. Multiple rows are duplicated, such as lines 95 and 899, which could lead to misleading and inaccurate conclusions due to this noise. Upon closer inspection, some of the text messages appear unfinished, suggesting that the data may be potentially incomplete.Unstructured data quality is less mature and demands more research attention compared to structured data, which has relatively clearer processing frameworks (Mirzaie, Behkamal, and Paydar 10). Jonker and Gomstyn also highlighted that uneven data, like this dataset, can easily lead to errors, illustrating the another challenges of unstructured information (Jonker and Gomstyn)

***b. How would you recognize this poor quality?***

When we first examined the dataset, we observed that many of the key fields were blank: columns such as email (97%missing), phone (97%missing), and LinkedIn (85%missing), which immediately made the table feel incomplete. As we looked closer, we noticed the name column is inconsistent, often containing job titles instead of individuals' names. The name column also contain inaccurate entry like one row listed “Scrum Master” three times. These issues were discussed by (“Guide to Structured vs Unstructured Data”) and (Freeman) when discussing poor data quality: missing values, inconsistent records, and information that can’t really be trusted.

**REFERENCE**

\* Alexandra Jonker, and Alice Gomstyn. “Structured vs. Unstructured Data: What’s the Difference?” IBM Think, 7 Feb. 2025, [Link](https://www.ibm.com/think/topics/structured-vs-unstructured-data#:~:text=DynamoDB%2C%20Hadoop%20%20and%20,43.) Accessed 1 Oct. 2025.

\* Freeman Mark. “Financial Data Quality: Modern Problems and Possibilities.” Gable, 7 Nov. 2024, [Link](https://www.gable.ai/blog/financial-data-quality-management#:~:text=Financial%20institutions%20often%20deal%20with,integration%20issues%20between%20different%20systems.) Accessed 28 Sept. 2025.

\* Guide to Structured vs Unstructured Data: with Real-World Examples.” Domo, 17 June 2025, [Link](https://www.domo.com/learn/article/unstructured-data-vs-strucuctured-data?utm\_source=%20https://www.domo.com/learn/article/unstructured-data-vs-strucuctured-data.) Accessed 1 Oct. 2025.

\* Mirzaie, Mostafa, Behshid Behkamal, and Samad Paydar. “Data Quality: A Systematic Literature Review and Future Research Directions.” Ferdowsi University of Mashhad, 2019. [PDF](https://arxiv.org/pdf/1904.05353)