- Hey there! ✋

  How are you doing today?

- I will be helping you to assess the quality of your dataset to gather high quality requirements for your system.

- Having trust in your data is crucial because it impacts the effectiveness of your decisions and any requirements driven from this data.

- Take a couple of minutes to get familiar with the data. Understanding your data makes a big difference. 🔍

- Do you think this dataset is relevant to what you are looking for? In other words, will it help you in understanding your users and system requirements more?

- What was the period of time covered in this dataset? Do you think the data could be outdated and changes might have happened since then?

- How many records under each year? Look for any change in behavior over the years. Such factors would affect the importance of some records over the others.

- Do not overlook duplication in the data. Multiple variations of the same record may provide skewed analytical results and incorrect insights.

- How about missing values, did you notice any critical ones? Ignoring them might affect the accuracy of other related fields.

- Do you know that 41% of companies in the US cite inconsistent data across technologies as their biggest challenge because it costs them alot! 💡

- Data consistency and integrity is very important to trust your decisions. An example of inconsistency in data is to have contradicting values or same values in different formats.

- Can you give me one case where the data is inconsistent in this dataset?

- Also, make sure your dataset is not over-representing a specific age group, race, or gender for example. This is called data bias and it can wreck your machine learning models if go undetected.

- Share with me any bias cases you find in the data (if any)?

- Take more time investigating the quality of the dataset in hand, and let me know when you are done.

- Great job 👏

  Early detection of data quality problems will save you time and money. IBM estimates that bad data costs the U.S. economy $3.1 trillion per year to make corrective actions.