Task1

Define which DQ measures should be checked on different layers

|  |  |  |  |
| --- | --- | --- | --- |
| DQ | LZ | RAW | CLEANSED |
| Completeness | YES | YES | Yes |
| Uniqueness | YES | YES | Yes |
| Timeliness | YES | YES | Yes |
| Validity | No | No | Yes |
| Accuracy | No | No | Yes |
| Consistency | No | No | Yes |

In the initial stages of data processing, particularly in the landing and raw layers, emphasis is placed on dimensions of data quality such as completeness, uniqueness, and timeliness. These layers serve as the entry points for data into the pipeline, where the primary goal is to ensure that the incoming data is accurately captured, and promptly available for next processing stages.

Other dimensions of data quality such as validity, accuracy, and consistency will be performed in the cleansed layer, where the data is subjected to rigorous validation against predefined rules, standards, and business requirements.