



Data Quality Metrics Definition for Aegis & Jarvis

Project	Table	Describe	Metric	
Jarvis	<ul style="list-style-type: none">y4a_cdm.y4a_dwa_amz_ads_sps_src_trm_v3	Get Sponsored Products data (latest version) for table y4a_analyst.tb_amz_ads_searchterm	<ul style="list-style-type: none">CompletenessAccuracyTimeliness	<ul style="list-style-type: none">Kiểm tra xem run_date đã latest hay chưa? => Nếu chưa trigger.impression>click => nếu sai thì triggerclick,cost, impression, keywordbid,purchases14d, sales14d,unitss oldclicks14d >=0 => nếu sai thì trigger
Aegis	<ul style="list-style-type: none">y4a_cdm.y4a_dwa_amz_ads_sps_tgt_v3	Get Sponsored Products data (latest version) for table y4a_analyst.tb_amz_target_kw_ads	<ul style="list-style-type: none">CompletenessAccuracyTimeliness	<ul style="list-style-type: none">Kiểm tra xem run_date đã latest hay chưa? => Nếu chưa trigger.impression>click => nếu sai thì triggerclick,cost, impression, keywordbid,purchases14d, sales14d,unitss oldclicks14d >=0 => nếu sai thì trigger

Aegis, Jarvis	<ul style="list-style-type: none"> y4a_cdm.y4a_d wa_amz_ads_sp s_adv_prd_v3 	Get Sponsored Products data (latest version) for table y4a_analyst.tb_y4a_amz_ads_perf_by_sku	<ul style="list-style-type: none"> Completeness Accuracy Timeliness 	<ul style="list-style-type: none"> Kiểm tra xem run_date đã latest hay chưa? => Nếu chưa trigger. impression>click => nếu sai thì trigger click,cost, impression, keywordbid,purchases14d, sales14d,unitss oldclicks14d >=0 => nếu sai thì trigger
Jarvis	<ul style="list-style-type: none"> y4a_cdm.y4a_d wa_amz_ads_spb_src_trm_v3 	Get Sponsored Brand data (latest version) for table y4a_analyst.tb_amz_ads_searchterm	<ul style="list-style-type: none"> Completeness Consistency Accuracy Uniqueness Timeliness Anomaly Detection 	<ul style="list-style-type: none"> Kiểm tra xem run_date đã latest hay chưa? => Nếu chưa trigger. impression>click => nếu sai thì trigger click,cost, impression, keywordbid,purchases14d, sales14d,unitss oldclicks14d >=0 => nếu sai thì trigger
Aegis	<ul style="list-style-type: none"> y4a_cdm.y4a_d wa_amz_ads_spb_tgt_v3 	Get Sponsored Brand data (latest version) for table y4a_analyst.tb_amz_target_kw_ads	<ul style="list-style-type: none"> Completeness Consistency Accuracy Uniqueness Timeliness Anomaly Detection 	<ul style="list-style-type: none"> Kiểm tra xem run_date đã latest hay chưa? => Nếu chưa trigger. impression>click => nếu sai thì trigger click,cost, impression, keywordbid,purchases14d, sales14d,unitss oldclicks14d >=0 => nếu sai thì trigger

Aegis, Jarvis	<ul style="list-style-type: none"> y4a_cdm.y4a_d wa_amz_ads_pr ofile_info 	Get Account Info data for table <ul style="list-style-type: none"> y4a_analyst.t b_y4a_amz_ads _perf_by_sku y4a_analyst.t b_amz_target_ kw_ads y4a_analyst.t b_amz_ads_sea rchterm 	<ul style="list-style-type: none"> Completeness Consistency Accuracy Uniqueness Timeliness Anomaly Detection 	
Aegis, Jarvis	<ul style="list-style-type: none"> y4a_cdm.full_ metrics_dail y 	Get Performance Metrics data	<ul style="list-style-type: none"> Completeness Consistency Accuracy Uniqueness Timeliness Anomaly Detection 	
Aegis, Jarvis	<ul style="list-style-type: none"> y4a_cdm.y4a_d wa_amz_avc_sl s_rpt 	Get Tracking Traffic by country, date, ASIN	<ul style="list-style-type: none"> Completeness Consistency Accuracy Uniqueness Timeliness Anomaly Detection 	
Aegis, Jarvis	<ul style="list-style-type: none"> y4a_cdm.y4a_d wa_mkt_kee_in f_his 	Get Historical Information data, Market Performance	<ul style="list-style-type: none"> Completeness Consistency Accuracy Uniqueness Timeliness Anomaly Detection 	
Aegis, Jarvis	<ul style="list-style-type: none"> y4a_cdm.y4a_a pi_pims_v4 	Get PIMS Information data	<ul style="list-style-type: none"> Completeness Consistency Accuracy Uniqueness Timeliness Anomaly Detection 	
Aegis, Jarvis	<ul style="list-style-type: none"> y4a_cdm.ALL_D WB_EXR 	Get Exchange Rates between USD and Other Currencies	<ul style="list-style-type: none"> Completeness 	
Aegis, Jarvis	<ul style="list-style-type: none"> y4a_cdm.y4a_d wb_amz_ads_pr 	Get Amazon Advertising	<ul style="list-style-type: none"> Completeness Consistency 	

	f	campaigns, capturing Key Metrics	<ul style="list-style-type: none"> • Accuracy • Uniqueness • Timeliness • Anomaly Detection 	
--	---	--	---	--

We have tables for Data Quality Alert

- Sponsor Product: y4a_cdm.y4a_dwa_amz_ads_sps_src_trm_v3
y4a_cdm.y4a_dwa_amz_ads_sps_tgt_v3
y4a_cdm.y4a_dwa_amz_ads_sps_adv_prd_v3
- Sponsor Brand: y4a_cdm.y4a_dwa_amz_ads_spb_src_trm_v3
y4a_cdm.y4a_dwa_amz_ads_spb_tgt_v3
- Other:

y4a_cdm.y4a_dwa_amz_ads_profile_info
y4a_cdm.full_metrics_daily
y4a_cdm.y4a_dwa_amz_avc_sls_rpt
y4a_cdm.y4a_dwa_mkt_kee_inf_his
y4a_cdm.y4a_api_pims_v4
y4a_cdm.ALL_DWB_EXR
y4a_cdm.y4a_dwb_amz_ads_prf

Ensuring data quality is essential for Aegis & Jarvis to maintain accurate, reliable, and actionable insights. Below are key data quality metrics along with their definitions, formulas, and examples.

1. Completeness [↗](#)

Definition: [↗](#)

Completeness measures the proportion of non-missing values in a dataset. It ensures that essential fields contain valid data without **NULL** or blank values.

Formula: [↗](#)

$$\text{Completeness} = 1 - \frac{\text{Number of NULL values}}{\text{Total records}}$$

Example: [↗](#)

If a dataset contains 10,000 records and 500 of them have missing values in the **clicks** column, completeness is:

$$1 - \frac{500}{10,000} = 95\%$$

Alert Criteria: [↗](#)

- If completeness falls below 98%, trigger a warning.
 - If completeness falls below 95%, trigger a critical alert.
-

2. Consistency [↗](#)

Definition: [↗](#)

Consistency ensures that data values conform to expected rules and formats. It verifies internal logic within the dataset.

Examples of Consistency Checks: [↗](#)

- `cost` and `sales` should not be negative.
- `clicks` should not be greater than `impressions`.
- $ACOS = cost / sales$ should not be negative.
- `currency_code` should correctly match `country_code`.

Alert Criteria: [↗](#)

- If inconsistencies exceed 1% of total records, trigger an alert.
-

3. Accuracy [↗](#)

Definition: [↗](#)

Accuracy measures how well data reflects real-world values. It ensures that information is correct and matches expected references.

Examples of Accuracy Checks: [↗](#)

- `campaignid` and `adgroupid` should exist in the system.
- `conversion_rate` should not exceed logical thresholds.
- `sales` should align with historical data patterns.

Alert Criteria: [↗](#)

- If data accuracy deviations exceed 2% from expected values, trigger an alert.
-

4. Uniqueness [↗](#)

Definition: [↗](#)

Uniqueness ensures that there are no duplicate records in the dataset.

Formula: [↗](#)

$$\text{Uniqueness} = 1 - \frac{\text{Number of duplicate records}}{\text{Total records}}$$

Example: [↗](#)

If a dataset contains 10,000 records and 200 records are duplicates, uniqueness is:

$$1 - \frac{200}{10,000} = 98\%$$

- If uniqueness falls below 99%, trigger a warning.
- If uniqueness falls below 97%, trigger a critical alert.

5. Timeliness [↗](#)

Definition: [↗](#)

Timeliness ensures that the dataset contains the most recent and up-to-date records.

Formula: [↗](#)

$$\text{Timeliness} = 1 - \frac{\text{Current date} - \text{Latest data date}}{\text{Threshold limit}}$$

Example: [↗](#)

If today's date is 2025-03-17 and the most recent data entry is 2025-03-15, assuming a threshold of 2 days:

$$1 - \frac{17-15}{2} = 50\%$$

Alert Criteria: [↗](#)

- If data is older than the threshold (e.g., 2 days), trigger an alert.

6. Anomaly Detection [↗](#)

Definition: [↗](#)

Anomaly detection identifies unexpected fluctuations in key metrics compared to historical trends.

Examples of Anomalies: [↗](#)

- clicks or sales change by more than 50% compared to the 7-day average.
- CTR = clicks / impressions drops significantly.
- ROAS = sales / cost is abnormally low.

Formula for Anomaly Detection: [↗](#)

$$\text{Anomaly Score} = \frac{\text{Current Value} - \text{Average of last 7 days}}{\text{Standard Deviation}}$$

Alert Criteria: [↗](#)

- If deviation exceeds ± 2 standard deviations, trigger a warning.
- If deviation exceeds ± 3 standard deviations, trigger a critical alert.

Implementation Strategy for Alerts [↗](#)

To ensure real-time monitoring, alerts can be implemented using SQL queries in PostgreSQL, Airflow DAGs, or Power BI dashboards.

Example SQL Query for Missing Data Alert: [↗](#)

```
1 SELECT COUNT(*) AS missing_values
2 FROM getDataRaw
3 WHERE clicks IS NULL OR cost IS NULL OR sales14d IS NULL;
```

Example SQL Query for Anomalies:

```
1 WITH avg_last7 AS (  
2     SELECT searchterm, AVG(clicks) AS avg_clicks, AVG(sales14d) AS avg_sales  
3     FROM getDataRaw  
4     WHERE report_date >= CURRENT_DATE - INTERVAL '7 days'  
5     GROUP BY searchterm  
6 )  
7 SELECT g.*,  
8     (g.clicks - a.avg_clicks) / NULLIF(a.avg_clicks, 0) AS click_change_ratio,  
9     (g.sales14d - a.avg_sales) / NULLIF(a.avg_sales, 0) AS sales_change_ratio  
10 FROM getDataRaw g  
11 LEFT JOIN avg_last7 a ON g.searchterm = a.searchterm  
12 WHERE g.report_date = CURRENT_DATE  
13 AND (  
14     ABS((g.clicks - a.avg_clicks) / NULLIF(a.avg_clicks, 0)) > 0.5  
15     OR ABS((g.sales14d - a.avg_sales) / NULLIF(a.avg_sales, 0)) > 0.5  
16 );  
17
```

Conclusion

The defined metrics ensure high data quality for Aegis & Jarvis by:

- Ensuring **completeness** and **uniqueness** to prevent missing or duplicate records.
- Maintaining **consistency** and **accuracy** for logical correctness.
- Monitoring **timeliness** for up-to-date information.
- Detecting **anomalies** to identify unusual patterns.

By implementing automated alerts, the system can proactively detect and resolve data quality issues before they impact decision-making.

