# **Weekly Data Validation Framework – Genome DQ Monitoring**

### **Purpose**

To ensure reliable and high-quality data for marketing campaigns, the A&D Data Strategy team has implemented a standardized validation process for the ET Digital Direct Genome platform. This process revolves around weekly checks that evaluate data integrity across multiple dimensions using the ICEDQ platform. The objective is to detect anomalies early, involve the appropriate stakeholders for resolution, and formally sign off validated data before it is used downstream.

### **Key Reports Monitored**

Every Sunday night at 11 PM, four critical ICEDQ-based reports are generated and reviewed. Each focuses on a specific aspect of data quality:

1. **DQ Metric Report**
   * **Platform**: ICEDQ
   * **Purpose**: Tracks 127 essential metrics, each with a defined threshold to flag abnormal fluctuations.
   * **Focus**: Ensures key KPIs and business-critical attributes stay within expected variability.
2. **Table Count Report**
   * **Platform**: ICEDQ
   * **Purpose**: Compares row counts between the Enterprise Data Warehouse (EDW) and Snowflake to verify data synchronization.
   * **Focus**: Detects missing or extra records that may indicate upstream ingestion or transformation issues.
3. **Null Check Report**
   * **Platform**: ICEDQ
   * **Purpose**: Identifies sudden changes in the proportion of null values at the column level.
   * **Focus**: Highlights potential pipeline breaks, schema changes, or ingestion errors.
4. **Volume Trend Report**
   * **Platform**: ICEDQ
   * **Purpose**: Monitors table-level volume trends using statistical deviations.
   * **Focus**: Flags sharp increases or decreases in data volume that deviate from historical patterns.

### **End-to-End Workflow**

The data validation process flows as follows:

1. **Genome Weekly Jobs Complete** →
2. **ICEDQ Reports Generated (DQ Metric, Table Count, Null Check, Volume Trend)** →
3. **Data Strategy Team Reviews Reports** →
4. **SME Consultation for Anomalies** →
5. **Sign-Off Provided** →
6. **Campaign Team Signs Off** →
7. **Campaigns Deployed**

This layered structure ensures issues are caught before any marketing action is taken.

### **Validation Partner: iceDQ**

ICEDQ, an external validation platform, powers all the automated checks described. It compares current data snapshots with historical benchmarks and surfaces deviations. The Data Strategy team owns the review and escalation process, coordinating with Data SMEs when thresholds are breached.

### **Step-by-Step Weekly Validation Process**

1. **Post-Job Trigger**
   * All Genome-related weekly jobs must complete successfully across EDW and AWS Snowflake.
   * ICEDQ reports are then auto-generated and published in the SharePoint folder for the 2025 campaign.
2. **Anomaly Review**
   * The team analyzes failed DQ checks from the DQ Metric report and follows these steps:  
     + **a. Attribute Analysis**: Confirm if the flagged metric deviated due to genuine data change (e.g., market volatility) or system issues.
     + **b. Trend Evaluation**: Compare the metric trend over time in EDW and Snowflake to validate whether the deviation is unusual or explainable.
     + **c. Release Impact Review**: Check if recent system releases or schema changes caused unintended metric shifts.
     + **d. SME Confirmation**: If an issue is unclear, escalate to the responsible SME for context and validation.
     + **e. Final Sign-Off**: Once diagnosed, document findings and provide formal sign-off.

### **Conclusion**

This process ensures that every week, before marketing campaigns go live, data quality is comprehensively validated across critical dimensions—accuracy, completeness, and consistency. With automated monitoring via ICEDQ and human-in-the-loop review from the Data Strategy team and SMEs, this approach minimizes data risk and supports data-driven campaign execution with confidence.