Data Quality Monitoring
Fact Sheets

DATAGAPS
THE GREAT DATA COMPANY

" Defines data rules using an easy to use web interface and share the results with your business users. "

Data is entering the enterprise more rapidly and from more places than ever before in today's environments. Whether the data is manually entered as in HR or ERP systems, machine-generated coming from IoT or Log files, the need has never been higher for data validations and scoring. Companies are faced with fines, regulatory pressures, loss of credibility and many other issues if they report off of data that has even minor data quality issues.

Key Features

- Table Level
- Model Level
- Enterprise Level

- Domain level
- Rules Based
- AI-Driven Predictions

- Data Quality Score Monitoring
- Web-Based Implementation
- Cloud or on Premise Deployment

Data quality can be defined as the **accuracy** and **reliability** of the data collected and the conclusion drawn from that data. Data quality is always a concern and can be analyzed in many ways.

Data quality scoring(DQS) is a metric-based way that helps companies assess the quality of their data and the confidence in the data's accuracy.
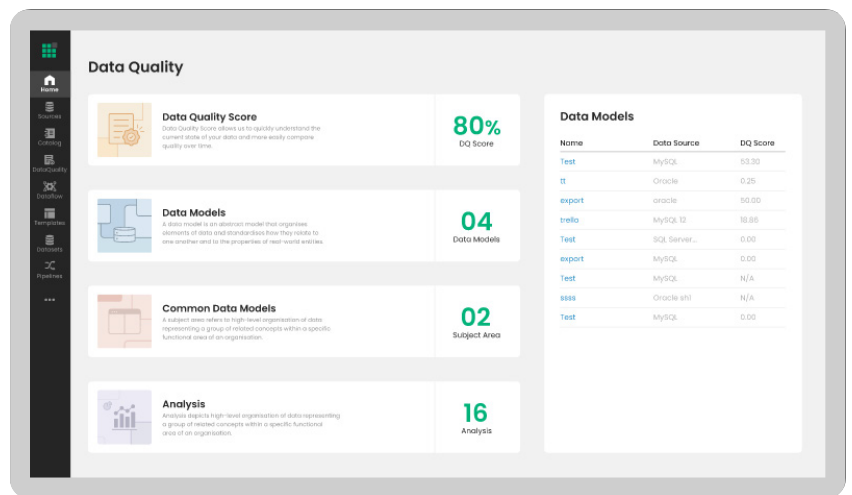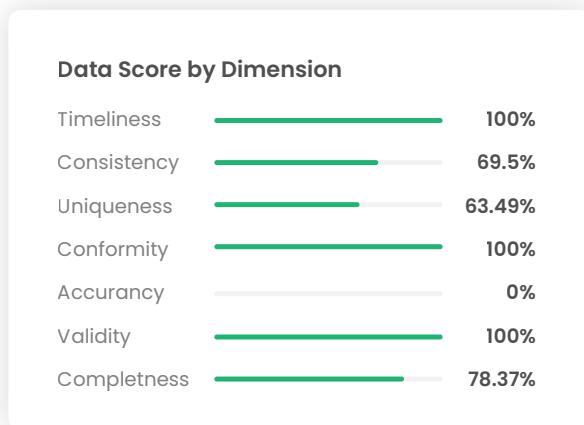

This metric can help companies better understand their data, pinpoint where errors are, and predict whether data accuracy is improving or falling. Data Quality Monitoring is a Datagaps approach to deriving your data landscape's scoring. It can be implemented to score DQ at any organization level and include as many data assets as you might like to track. DQS can be sliced in several distinct ways. One, along the following dimensions that roll up through your data hierarchy.

The other is the data hierarchy itself, i.e., table, data element, model, Datawarehouse, enterprise. The third is the user-defined subject area so that you can score data elements according to function such as sales, finance, HR, or any other subject area that you like.

The flexibility that you get with our approach is the ability to view your DQS at any level, even user-defined.

Data Score by Dimension

Dimension	Score
Timeliness	100%
Consistency	69.5%
Uniqueness	63.49%
Conformity	100%
Accurancy	0%
Validity	100%
Completness	78.37%

However, there is another aspect of our Data Quality Monitoring solution. Data Quality Monitoring Boundaries. Most numeric data flowing into your enterprise follows a pattern that can be learned over time. We use AI-based algorithms to learn what these patterns are to predict the boundaries of the ingested data over time. The longer the process runs, the better it gets at the predictions and the more accurate it becomes. After a short period of time, you can run the new data that you ingest through the process and find anomalies in the data as it arrives. It can notify you of these anomalies so that they can be remedied and thus keep your DQS in line with your expectations.

Confirmed_Cases (Country: INDIA)
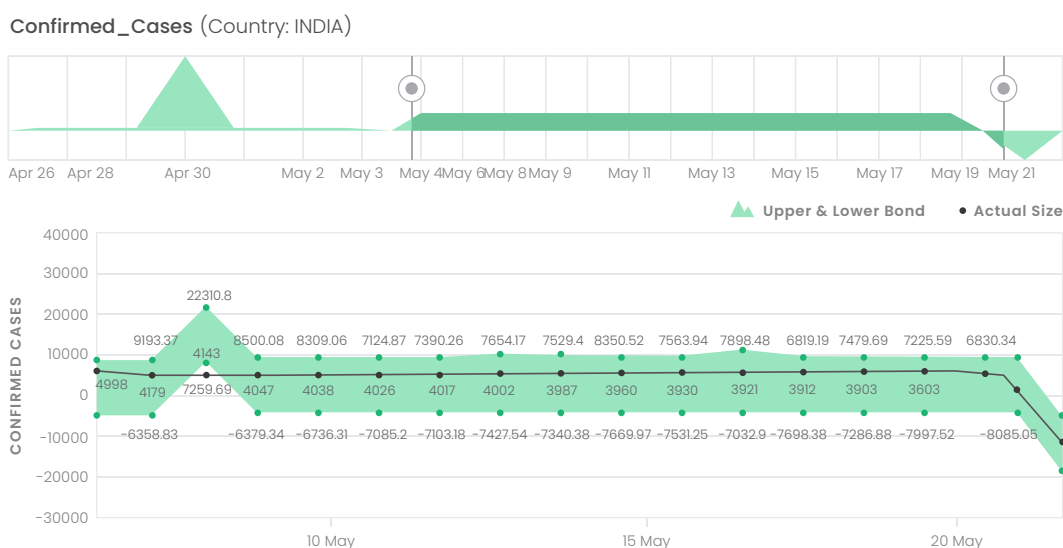
As you can see, the predicted boundaries are illustrated by the green graphics and the new data element is the line chart. You can see on May 8 th the data element fell outside of the normal boundaries, so it was flagged and the appropriate person was notified. With this example, we show just one data element that is being predicted. However, there is no limit to the number of numeric elements that can be predicted and tracked.

In Conclusion, The combination of DQS and Data Monitoring allows organizations to validate their essential data assets as these elements are ingested and stored. The high-level view shows an overall score and if it deteriorates, the granular views quickly point to the data attributes causing the degradation. Once that is identified, corrective action can take place to remedy the anomalies.