DATA QUALITY: THE KEY TO UNLOCKING THE VALUE OF YOUR DATA

ESTHER MUNYI

Data Enthusiast
Chief Data and Analytics Officer
Sasfin Bank

Data Quality is often defined as the “**fitness of use.**

- Data fit for intended use
- Data reflects real value
- Data meets reasonable standards

$12.9 Million

Is lost every year due to poor data quality

based on Gartner’s survey(2020), and this situation is likely to worsen as business operations and data ecosystems become increasingly complex

WHY IS DATA QUALITY IMPORTANT?

Make Better Decisions

Organizations can be more confident in the decisions they make with reliable data.

Comply with Regulatory Requirements

Organizations can avoid costly fines and penalties for non compliance due to poor data quality.

Improve Efficiency

Accurate data can help optimize operational costs by identifying business processes that are time consuming, resource intensive or inefficient.

Improve Customer Service

Accurate and complete data of customers can lead to better customer service and customer satisfaction.

Drive Growth and Profitability

Good data increases the chances of success on opportunities.

Better Performance in AI Models

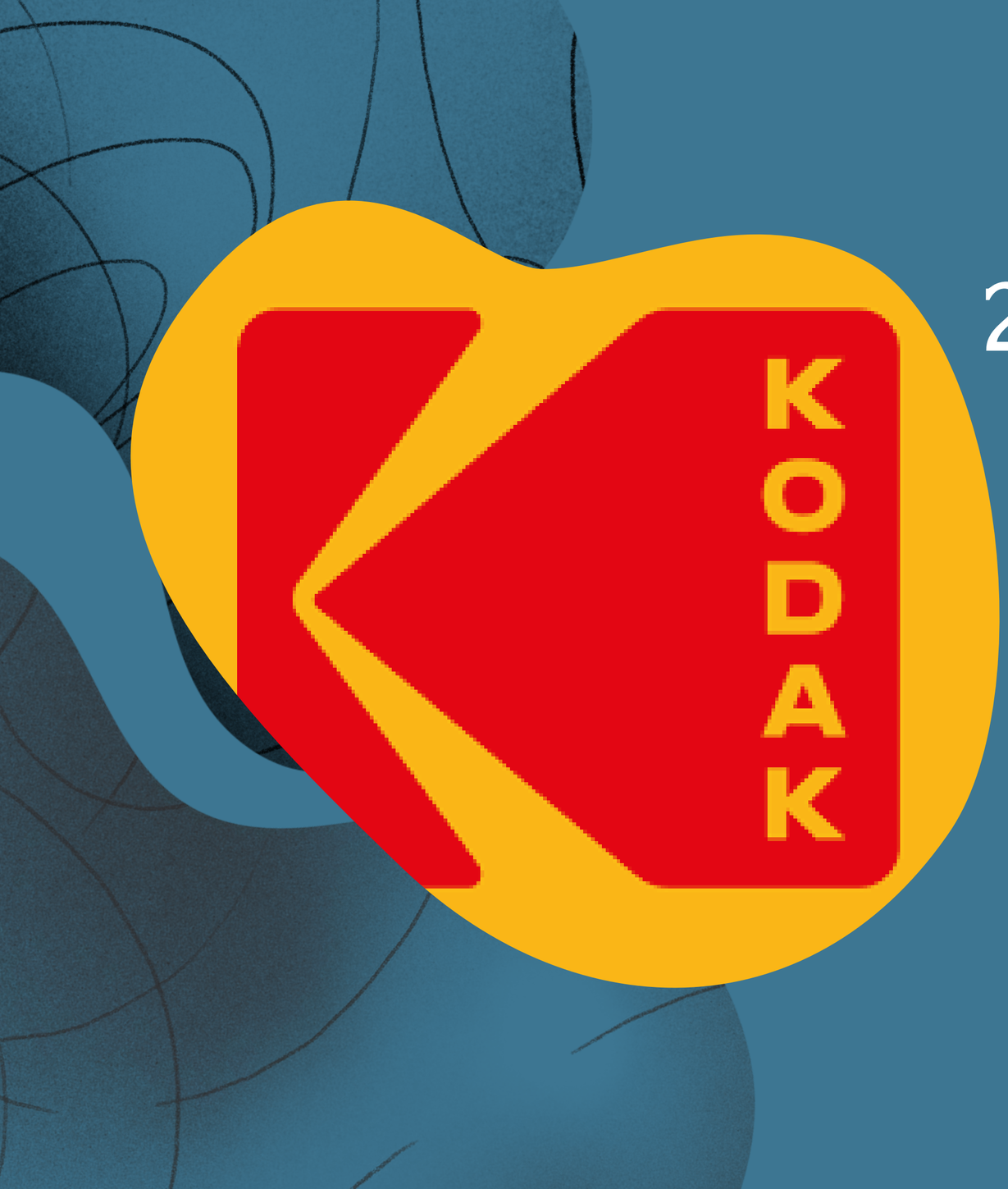
High-quality data enables models to make better predictions and produce more reliable outcomes.

IMPACT OF BAD DATA

1. WRONG DECISIONS

2008 World Financial Meltdown

Financial institutions used bad data to assess the risk of subprime mortgages, which led to the creation of complex financial products that were too risky. When these mortgages defaulted, key financial institutions collapsed, leading to widespread evictions, foreclosures, and job losses globally.

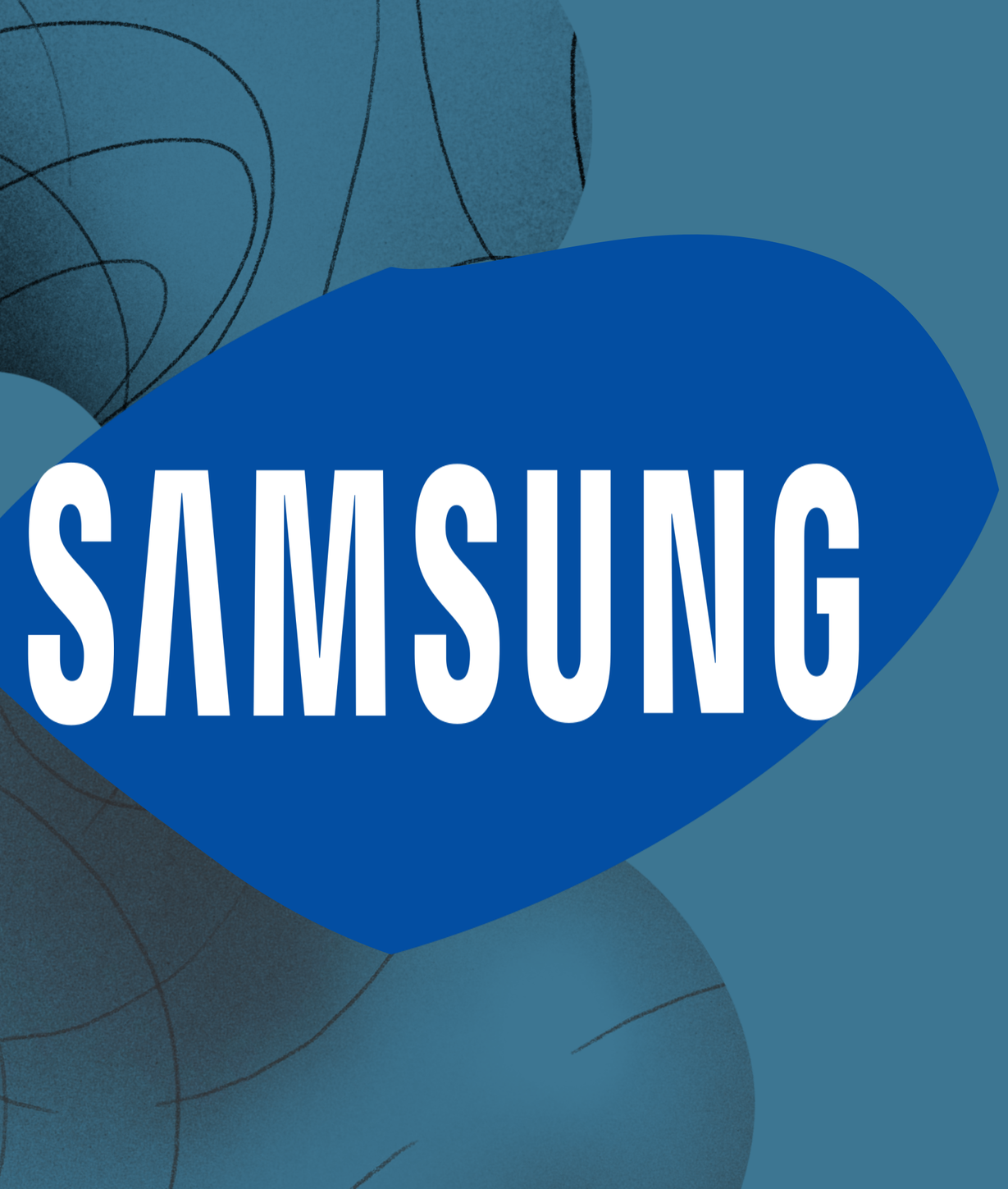


2. MISSED OPPORTUNITY

Kodak’s Missed Opportunity

Kodak had 10 years to prepare, but instead chose to focus on its existing film business. As a result, Kodak was caught off guard by the rapid adoption of digital photography and eventually went bankrupt.

Its important to take steps even if it seems there is plenty of time.



3. INCREASED COSTS

Samsung: Data Entry Error Cost $105 Billion

In 2018, a Samsung Securities employee mistakenly issued 1,000 shares to workers instead of 1,000 won per share in dividends, costing the company $300 million. The error was fixed within 37 minutes, but the company still had to pay dividends worth 1,000 times the value of each share to 2,018 employees.

NEWS & COMMENTARY

The Computer Got it Wrong: Why We're Taking the Detroit Police to Court Over a Faulty Face Recognition 'Match'

Where to Start:

4. REPUTATIONAL DAMAGE

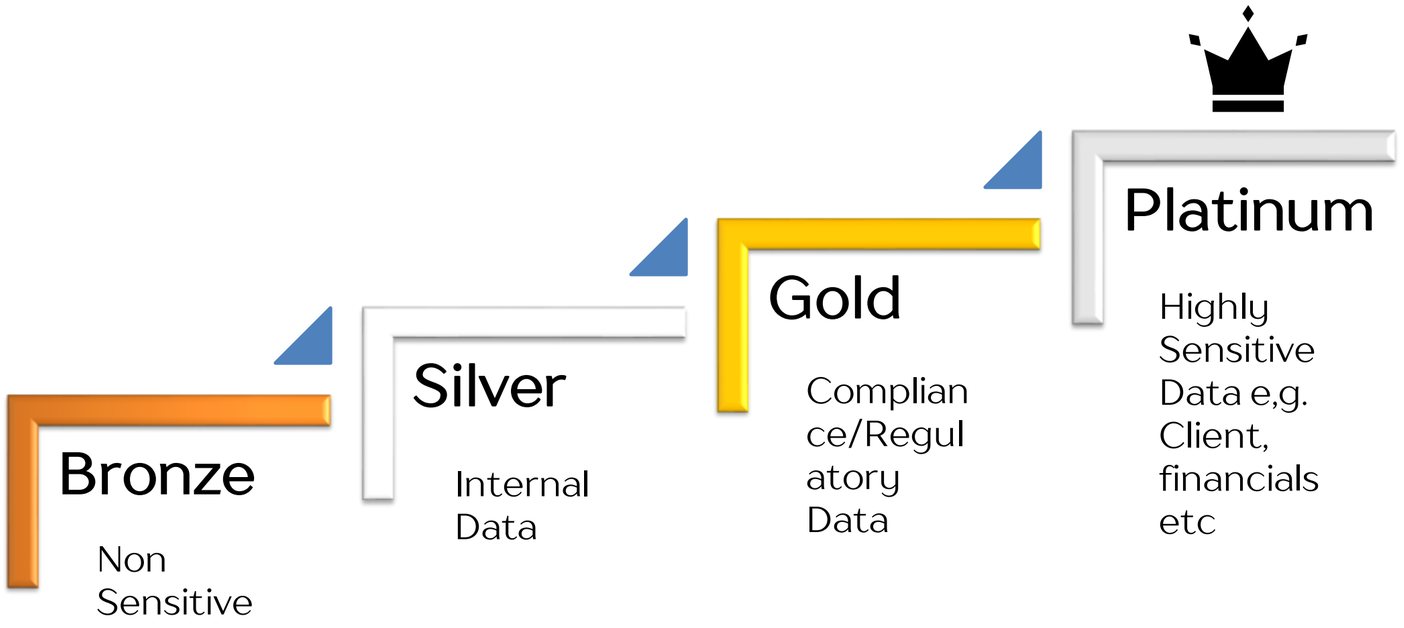
Detroit Police Wrongful Arrests

Detroit's police chief admitted that facial recognition technology used by the department misidentifies suspects about **96 percent** of the time. It is well documented that face recognition technology is deeply flawed.

The technology has a disturbing record of racial bias against people of color and other marginalized groups.

WHERE TO START?

1. Identify & focus on what's important



2. Define the data quality rules

Business Requirements

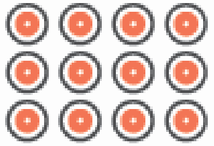
Industry Standards

Regulatory, Statutory, Compliance Requirements

6 dimensions of data quality

Accuracy
The degree to which the data correctly represents the entity or attribute being described.

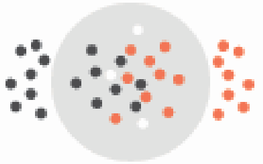
Completeness
The percentage of missing data from a given data set.



Consistency
The absence of difference or contradiction in data irrespective of the data's source.

Validity
Invalid data affects the accuracy and completeness of a data set.

Integrity
The validity of relationships across various data entities.



Uniqueness
Ensures duplicate or overlapping data is identified and marked.

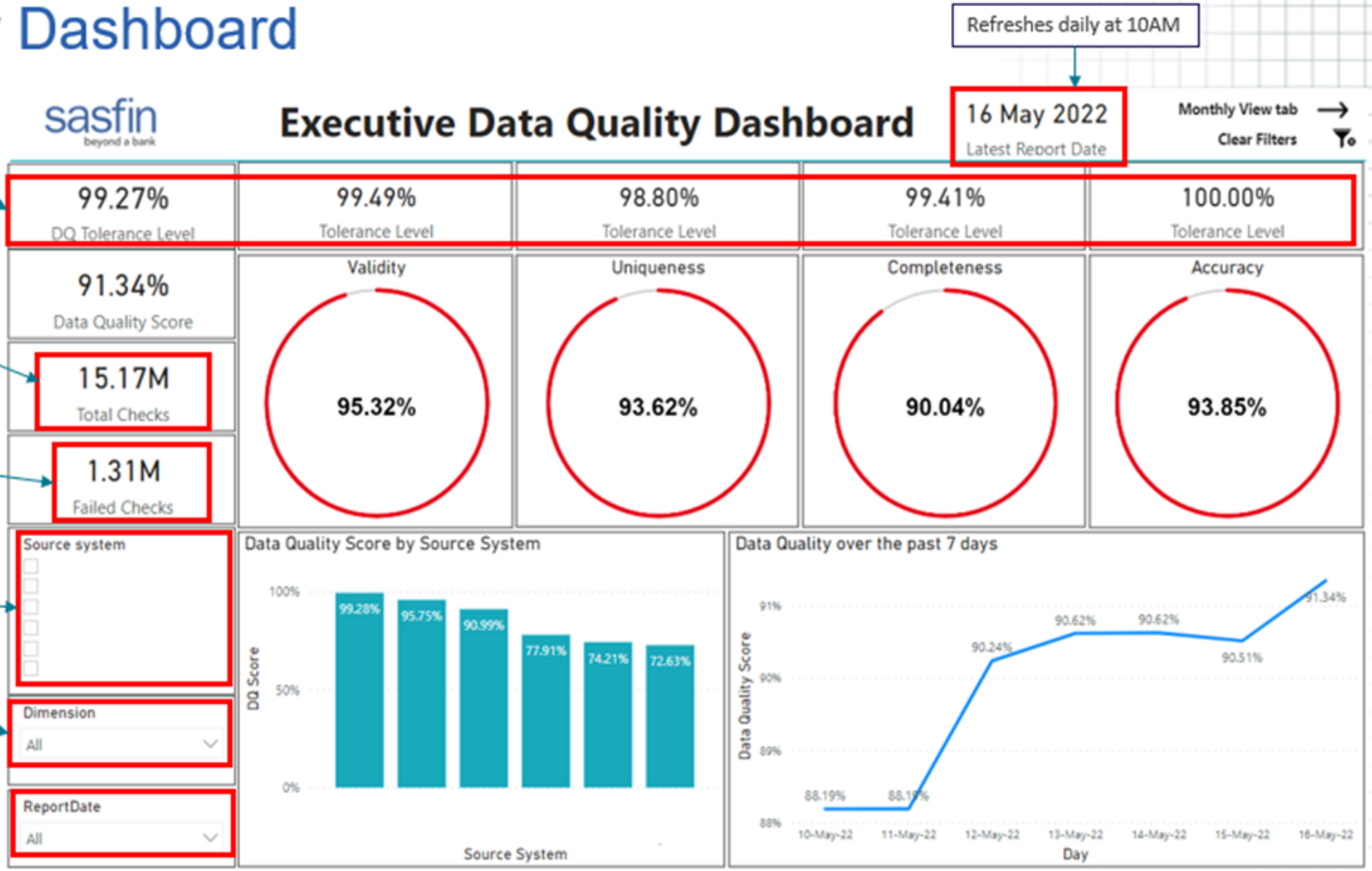
DESIGN: LINDA KOURY; ICONS: PRESSUREUA/GETTY IMAGES

3. Determine your tolerance levels

80:20 Rule

1:10:100 Rule

Data Quality Dashboard



DQ Tolerances: range of what is acceptable quality

Total Checks: Total number of all check as documented in the DQ Rules. Not the number of records checked

Total Checks: Total number failed check. Not the number of failed records.

Source System: Options to view DQ Stats for the Source System(s) selected

Dimensions: Options to view DQ Stats for the Dimension(s) of interest

Dimensions: Options to view DQ Stats for Report dates in the past or present

sasfin

beyond a bank

WHO'S RESPONSIBLE?

THABO BESTER PRISON ESCAPE

Who was responsible?

correctional services

Department:
Correctional Services
REPUBLIC OF SOUTH AFRICA

EVERYONE IS RESPONSIBLE!

Data Stewards Forum

- *Review of data quality exceptions reports
- Performance of deep dives into data quality defects
- Report on remediation efforts and initiatives undertaken by business
- Provide guidance on business context and reporting to Business Owners and respective committees

Frequency: Monthly

Business Owners and Management

- Review of data quality remediation and improvements
- Provide support to Data Stewards in the respective business teams in executing their data remediation initiatives
- Monitor performance of data quality metrics against set targets, tolerances and thresholds
- Deliberate on high-risk data quality issues that are potential limitations
- Assess impact of identified data quality issues on business strategy

Frequency: Monthly, Bi-Monthly

Data Governance Committee

- Review of data quality exceptions per line of business
- Review of commentary provided by lines of business (input from DS Forum and Business Owners)
- Provide oversight on bank-wide data quality performance
- Report high risk and impact data quality issues to the Board and Executive Committee/s

Frequency: Monthly

Board and Executive Committees

- Review of the bank's overall data quality performance
- Raise concerns and risks that data quality performance may have on the bank's overall strategy
- Support data quality initiatives undertaken by business

Frequency: Various

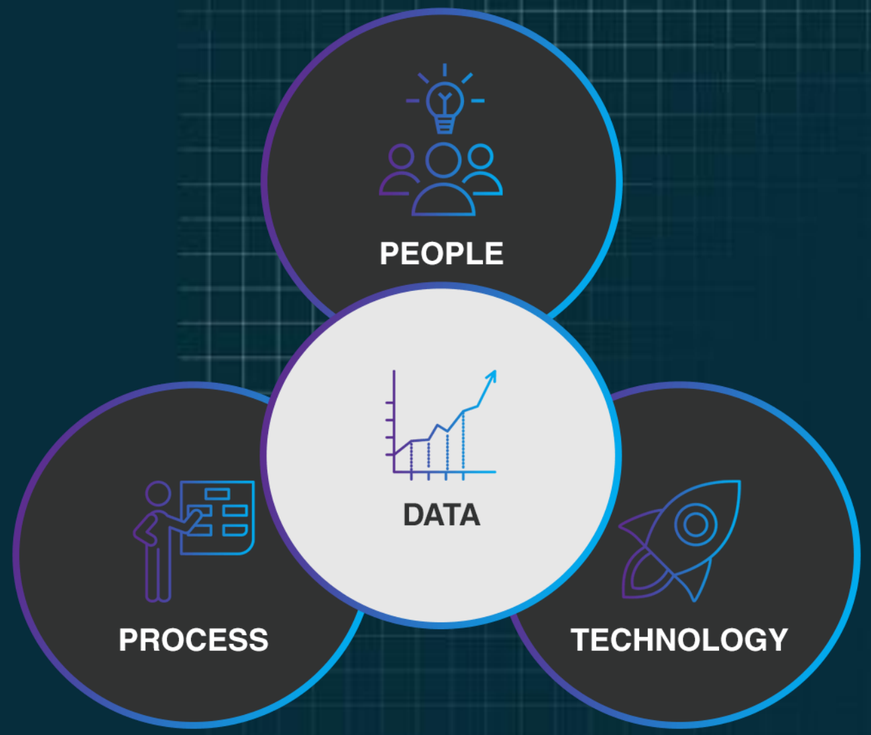
TAKE ACTION

Track Remediation Targets

Perform Root Cause Analysis

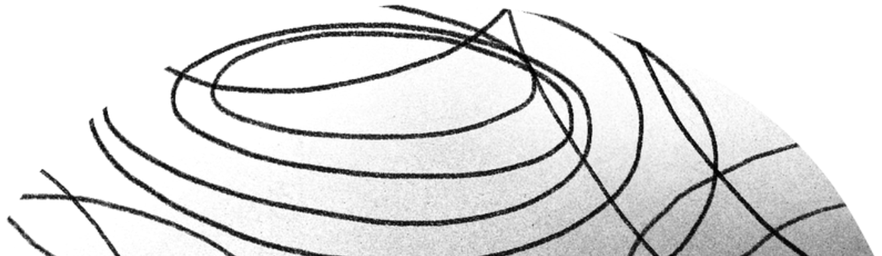
Establish Proactive Incident Prevention

Train employees on data quality best practices



CHALLENGES

- **Lack of resources:** Data quality initiatives can be expensive and time-consuming.
- **Lack of buy-in and commitment**: Data quality initiatives often require changes to business processes and culture. Many organizations are not committed to improving their data quality.
- **Lack of data governance**: Data quality initiatives are more effective when there is a clear governance framework in place.
- **Lack of expertise.** Many organizations do not have the expertise to implement and manage data quality initiatives.
- **Data silos**: Data silos occur when data is stored in different systems and is not easily accessible. This can make it difficult to get a complete picture of the data and can lead to errors.



THANK YOU!

If you have more questions:

Email: esthermunyi@gmail.com

LinkedIn: https://www.linkedin.com/in/esthermunyi/