TRIPLE A DATA QUALITY

DQ Engine as copilot for the pension transition

TRIPLE A
RISK FINANCE

ACTUARIAATCONGRES

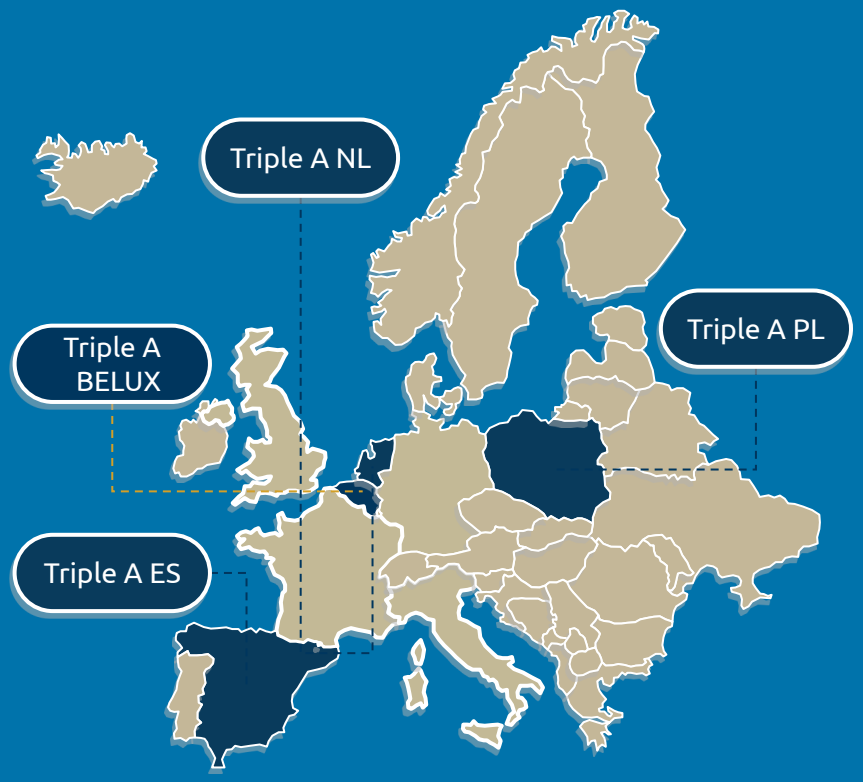
VSAE

INTRODUCTION

Hen Veerman
Partner & Pensions Lead
Triple A - Risk Finance

Hen Veerman
Practice Lead Data & Analytics
Triple A - Risk Finance

TRIPLE A
RISK FINANCE



Triple A Offices

150+ consultants

Serving the vast majority of European top 20 insurers

Part of Redmore, part of House of HR (57k people at work, EUR 2+ Bn turnover)

VOETTEKST NOG TOEVOEGEN

TRIPLE A RISK FINANCE

100+
Clients

170
Consultants (mostly in the Netherlands)

4
Offices (Amsterdam, Barcelona, Brussels, Warsaw)

- Insurance
- Pensions
- Banking
- Risk & Strategy Consulting
- Data & Analytics
- Actuarial Technology

Triple A Risk Finance

is an independent and innovative consultancy company specialised in actuarial, risk management and data science services.

We are open and informal, and at the same time secure and focused on results.

Our clients include insurers, banks, pension funds, pension providers and (large) companies.

Services

Consulting
Modelling
Certification
Project Management
Operational Support
Outsourcing
Key Functions
Training

OUTLINE

1. Playing field: what do we encounter
2. Data Quality Engine: copilot for the transition
3. Future: how to stay clean & what can we expect

PLAYING FIELD

What do we encounter

TRIPLE A
RISK FINANCE

TRANSITION

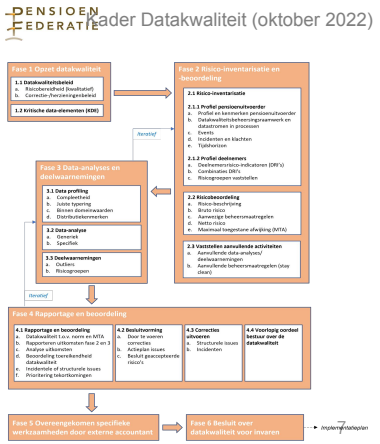
- Transition from a defined benefit pension to a defined contribution pension without consent
- Only applicable for pension funds
- Employer after consultation with unions or works council decides (and not pension fund!)
- Pension fund needs to test whether a balanced and fair transition is possible and execute
- Supervisors need to consent (DNB : policies and data quality and AFM : communication)

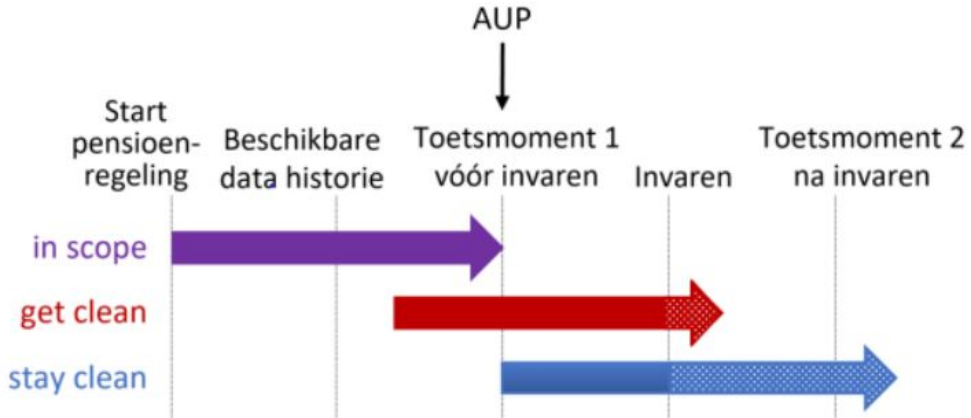
DATA QUALITY TESTING

- Framework for data quality testing (sector approach)

- Pension entitlements at moment of truth are:
  - Correct, complete, demonstrable and reproducible

- Data quality testing approach consists of 6 phases
  - 1. Set-up, data quality policy and definition of CDE
  - 2. Risk assessment
  - 3. Data analysis
  - 4. Reporting, judgement by board
  - 5. Auditor's review (AUP)
  - 6. Decision (go / no go)

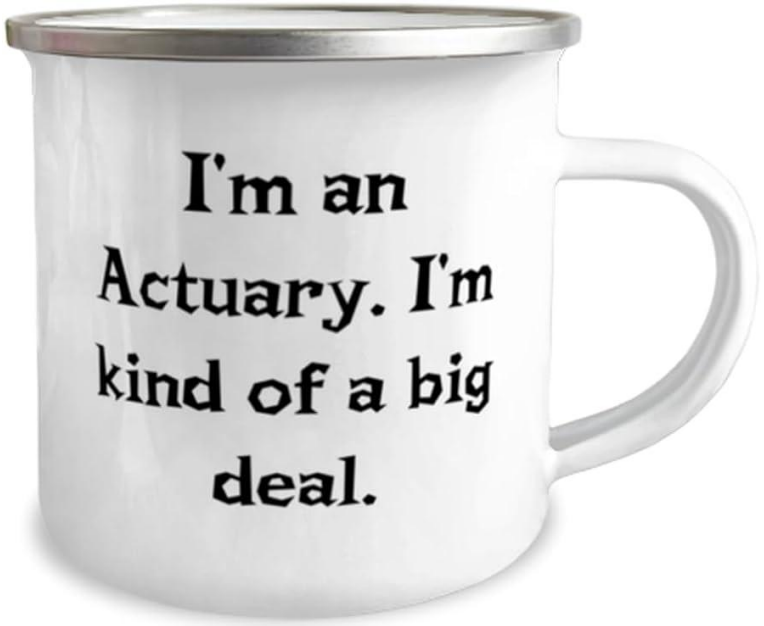




SOME REMARKS

- Data quality is ‘THE’ domain for the actuarial function!

- Issues with pension administration
  - Difference between historic and current requirements
  - Administration of ‘facts’ or administration of ‘positions’?
  - Availability of historic data
  - Delivery of employee service data
  - Specific issues industry wide pension funds



- Assessment and estimate whether impact of missing historic data is within risk appetite

- Buffer in operational reserve for corrections

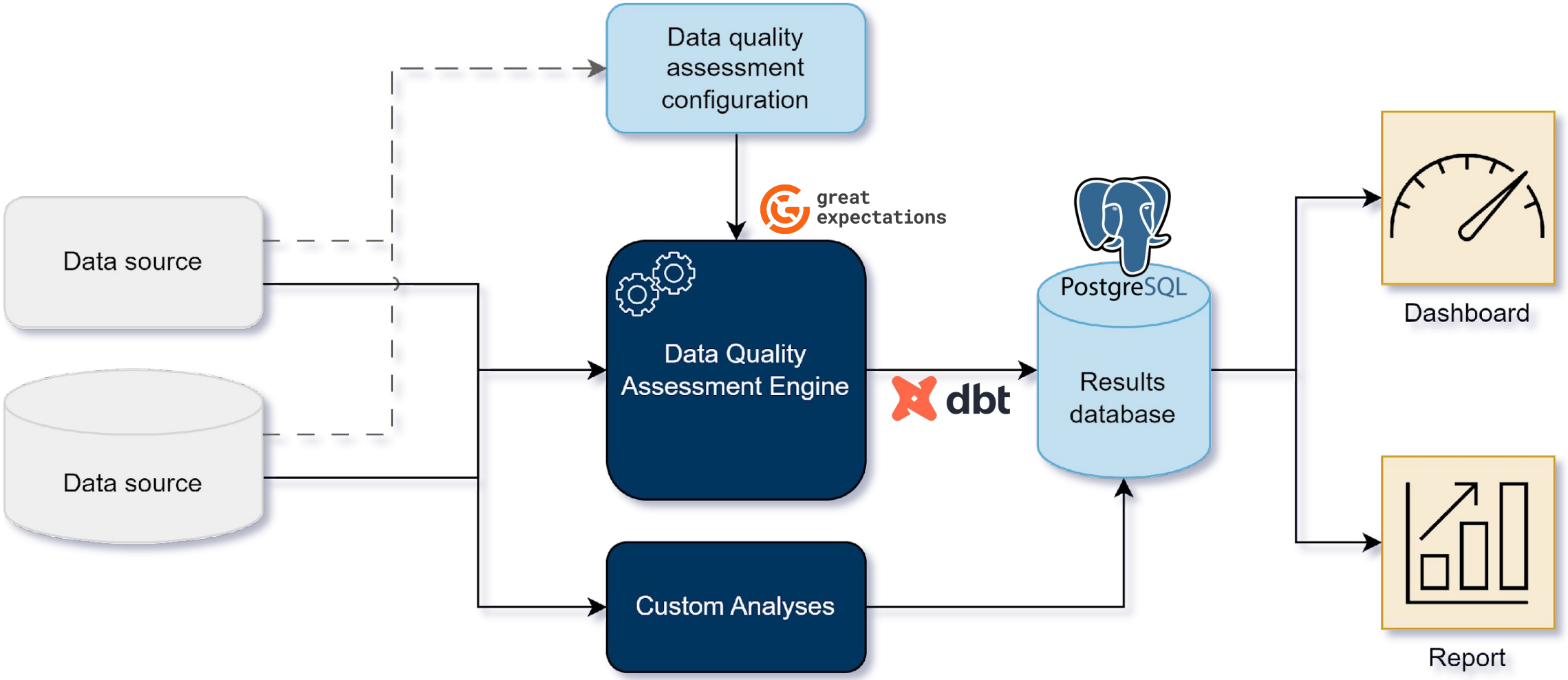
DATA QUALITY ENGINE

Copilot for the transition

TRIPLE A
RISK FINANCE

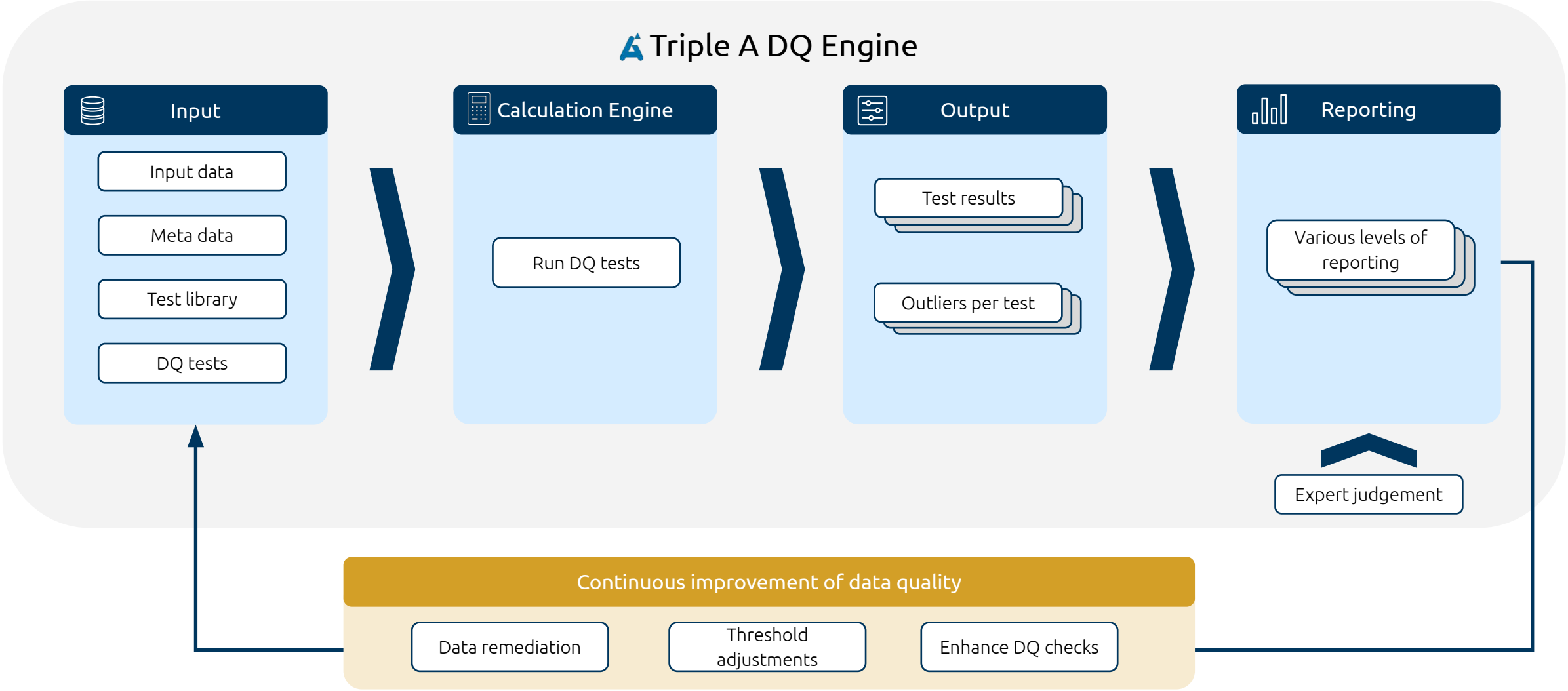
TRIPLE A DATA QUALITY ENGINE

Architecture



TRIPLE A DATA QUALITY PROCESS

From data input to reporting



Data Quality Engine

Configuration

Run

Reporting

About

TRIPLE A RISK FINANCE

Configuration

1 Test configuration

Upload configuration file | Manual test configuration

2 Data selection

Upload data file | Upload data map

3 CDE specification

Upload CDE file | Manual CDE specification

4 Primary key specification

Upload primary key file | Manual primary key specification

5 Test-CDE mapping

Upload test-CDE mapping file | Manual test-CDE mapping

Test configuration

Test library	Threshold lowerbound	Threshold upperbound	Reporting dimension
Determine number of missing attribute values	0.95	0.99	Completeness
Determine number of duplicate records based on primary key	0.99	1	Uniqueness
Assess whether attribute has the expected properties	n.a.	n.a.	Validity
Determine number of attribute values not within domain	0.92	0.97	Validity
Compute summary statistics	n.a.	n.a.	Accuracy
Determine number of the IQR outliers	0.9	0.95	Accuracy
Assess logical relationship between two or more attributes within dataset	0.9	0.95	Consistency

Save | Download as CSV file | Back to main screen

- The test configuration lists all types of tests that can be performed.
- Default threshold levels and reporting dimensions can be customized.
- In case of manual configuration, the file can be saved and downloaded as a csv file. When performing a new data quality analysis based on the same test definitions, the file can be uploaded again.

Data Quality Engine

Configuration

Run

Reporting

About

TRIPLE A RISK FINANCE

Configuration

1 Test configuration

Upload configuration file | Manual test configuration

2 Data selection

Upload data file | Upload data map

3 CDE specification

Upload CDE file | Manual CDE specification

4 Primary key specification

Upload primary key file | Manual primary key specification

5 Test-CDE mapping

Upload test-CDE mapping file | Manual test-CDE mapping

CDE specification

+ Add row | - Remove row

CDE ID	Dataset	Name of column	Type	Missing allowed	CDE	Field length	Decimal length (in case of numeric)	Domain values	Notes
1	Customer_loans	Name	Str	☐	☑	50		[ ]	
2	Customer_loans	Date of birth	Str	☐	☑	10		[01-01-1900:01-01-2006]	
3	Customer_loans	Address	Str	☑	☑	100		[ ]	
4	Customer_loans	Income (monthly)	Num	☐	☑	7	2	[0:100000]	
5	Customer_loans	Employment status	Str	☐	☑	20		[P, F, parttime, fulltime]	
6	Customer_loans	Assets	Num	☐	☑	7	2	[0:100000]	
7	Customer_loans	Loan ID	Str	☐	☑	8		[CCNNNNNN]	
8	Customer_loans	Loan amount	Num	☐	☑	7	2	[0:100000]	
9	Customer_loans	Interest rate	Num	☐	☑	6	4	[-100:100]	
10	Customer_loans	Loan term (months)	Num	☐	☑	3	0	[0:360]	

Save | Download as CSV file | Back to main screen

- The critical data elements are based on the columns that need to be tested on data quality.
- For each CDE, several properties must be specified.
- As with the test configuration: after inserting all CDE's by hand, this table can be saved and downloaded as a csv file to be uploaded again for future data quality analyses.

Data Quality Engine

Configuration

Run

Reporting

About

TRIPLE A RISK FINANCE

Configuration

1 Test configuration

Upload configuration file | Manual test configuration

2 Data selection

Upload data file | Upload data map

3 CDE specification

Upload CDE file | Manual CDE specification

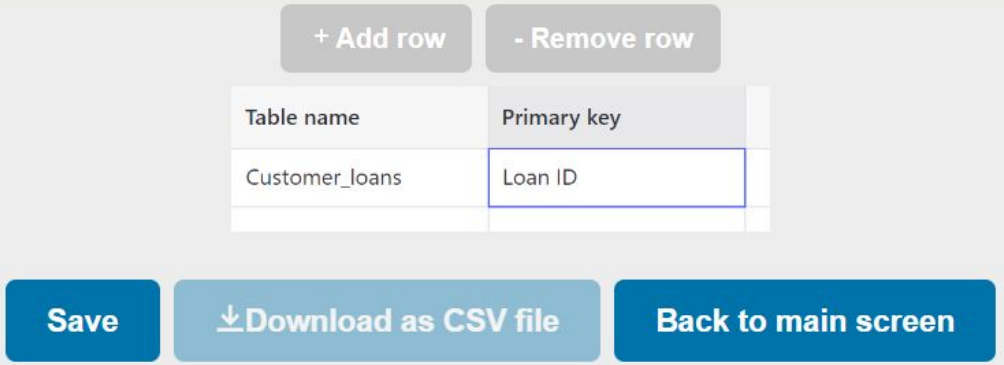
4 Primary key specification

Upload primary key file | Manual primary key specification

5 Test-CDE mapping

Upload test-CDE mapping file | Manual test-CDE mapping

Primary key specification



- The primary keys can again be uploaded or manually specified.
- Based on the primary key(s) of a data set, which can consist of one or multiple CDEs, a test for uniqueness is conducted.

Data Quality Engine

- Configuration
- Run
- Reporting
- About

TRIPLE A RISK FINANCE

Configuration

1 Test configuration

Upload configuration file | Manual test configuration

2 Data selection

Upload data file | Upload data map

3 CDE specification

Upload CDE file | Manual CDE specification

4 Primary key specification

Upload primary key file | Manual primary key specification

5 Test-CDE mapping

Upload test-CDE mapping file | Manual test-CDE mapping

Test-CDE mapping

+ Add row | - Remove row

Check ID	Dataset	Name of column	Test name	Reporting dimension	Test ID	Condition	Threshold lowerbound	Threshold upperbound	Reporting dimension override
1	Customer_loans	Name	Determine number of mi	Completeness	1		0.98	1	
2	Customer_loans	Date of birth	Determine number of mi	Completeness	1		0.95	0.99	
3	Customer_loans	Address	Determine number of mi	Completeness	1		0.95	0.99	
4	Customer_loans	Income (monthly)	Determine number of mi	Completeness	1	Group by gender	0.95	0.99	
5	Customer_loans	Employment status	Determine number of mi	Completeness	1		0.95	0.99	
6	Customer_loans	Loan ID	Determine number of mi	Completeness	1		0.98	1	
7	Customer_loans	Loan amount	Determine number of mi	Completeness	1		0.95	0.99	
8	Customer_loans	Interest rate	Determine number of mi	Completeness	1		0.95	0.99	
9	Customer_loans	Loan term (months)	Determine number of mi	Completeness	1		0.95	0.99	

Save | Download as CSV file | Back to main screen

- From the test definitions defined in step 1, the tests that need to be performed can be selected.
- For each selected test, the CDE's can be chosen to which the test should be applied.
- For each combination of tests and CDEs, optional or mandatory test conditions can be specified.
- It is possible to adjust the threshold levels and reporting dimension when deviation from the default values is desired.

Data Quality Engine

Configuration

Run

Reporting

About

TRIPLE A RISK FINANCE

Run

Data Quality Engine

Configuration

Run

Reporting

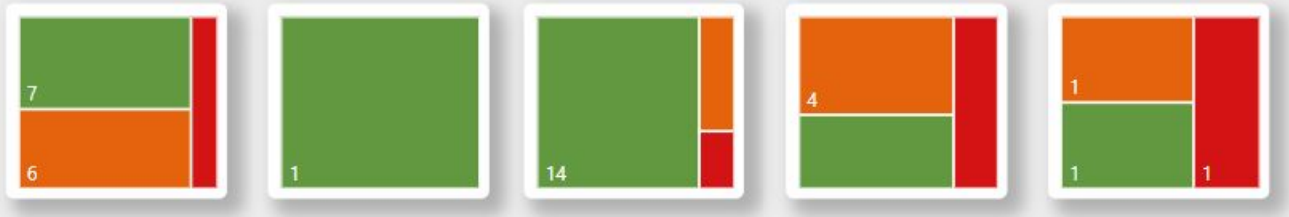
About

TRIPLE A RISK FINANCE

Reporting

RAG FOR EACH DQ DIMENSION

STATUS



CDE	Completeness	Uniqueness	Validity	Accuracy	Consistency
Purpose of the loan					
Address					
Name					
Employment status					
Loan ID					
Interest rate					
Default2					
Default					
Assets					
Loan term (months)					
Number of days past due					
Loan amount					
Date of birth					
LTV					
Credit score					
Income (monthly)					
Monthly payment amount					

TRIPLE A
RISK FINANCE

PERCENTAGES AND RAG PER TEST - COMPLETENESS

CDE	Determine records with non-missing values	Threshold lowerbound	Threshold upperbound	Number of records checked	Determine records with non-missing values using dependencies on other fields in the same data set	Threshold lowerbound	Threshold upperbound	Number of records checked
Address	98%	95%	99%	4304				
Credit score	✓ 99%	95%	99%	4304				
Date of birth	✓ 100%	95%	99%	4304				
Default2	✗ 92%	95%	99%	4304				
Employment status	✓ 99%	95%	99%	4304				
Income (monthly)	96%	95%	99%	4304				
Interest rate	✓ 99%	95%	99%	4304				
Loan ID	✓ 100%	95%	99%	4304				
Name	✓ 100%	95%	99%	4304				
Default	✗ 89%	95%	99%	4304	✗ 89%	95%	99%	2754
Loan term (months)	97%	95%	99%	4304	95%	95%	99%	2754
LTV					95%	95%	99%	2754
Number of days past due					96%	95%	99%	87
Monthly payment amount	98%	95%	99%	4304	97%	95%	99%	2754
Loan amount	✓ 99%	95%	99%	4304	✓ 99%	95%	99%	2754

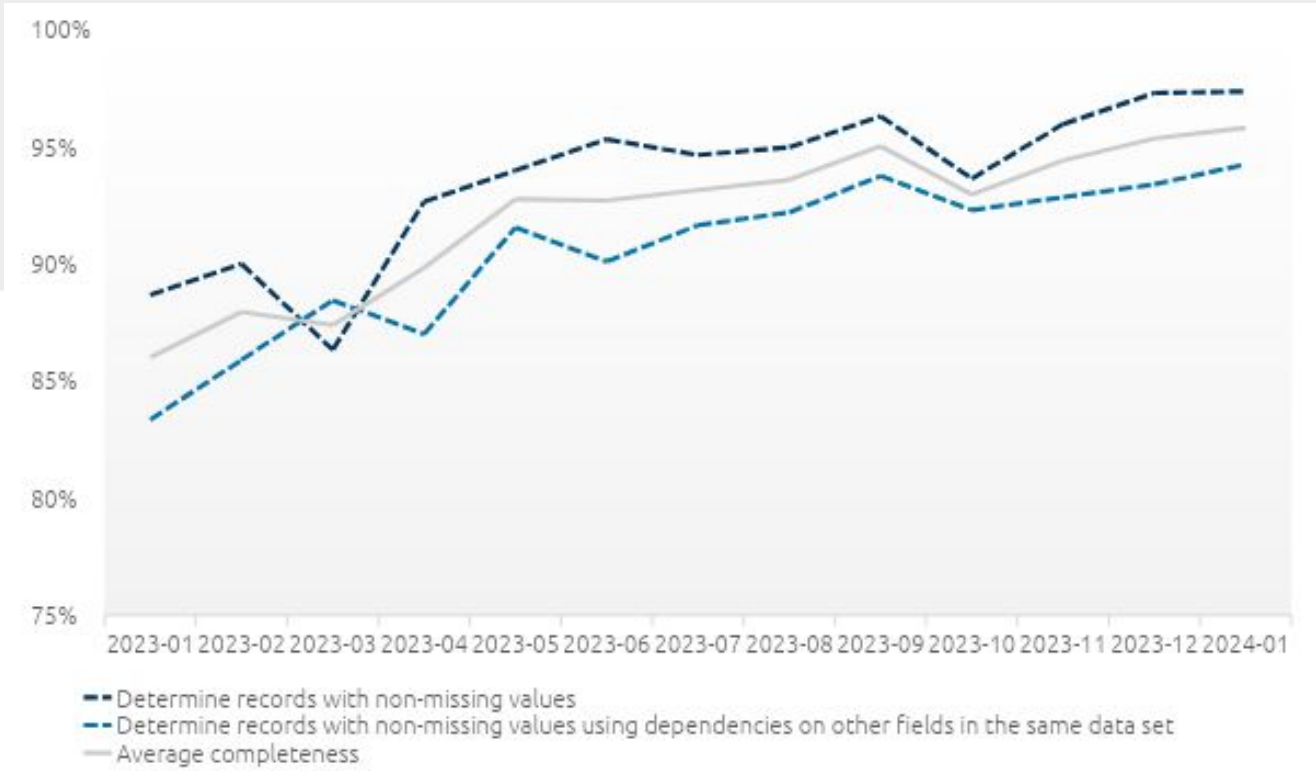
TRIPLE A RISK FINANCE

OUTLIERS TABLE - COMPLETENESS

Test ID	Table	Loan ID (primary key)	Date of birth	Loan amount	Loan term	Monthly payment amount
16	Customer_loans	133709	13-4-1957		12	
16	Customer_loans	4678896	30-3-1944		2	
16	Customer_loans	9008779	24-12-1959			300
16	Customer_loans	23456764	7-7-1939			800
16	Customer_loans	628463	8-10-1959		1	200
16	Customer_loans	8595749	9-1-1959		10	
16	Customer_loans	...	...		...	...

TRIPLE A
RISK FINANCE

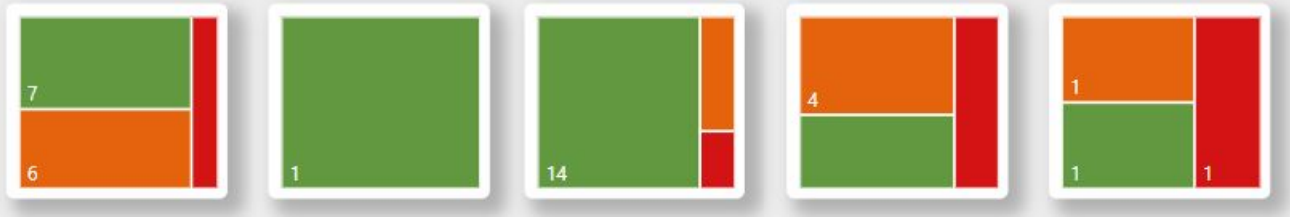
AVERAGES OVER TIME - COMPLETENESS



TRIPLE A
RISK FINANCE

RAG FOR EACH DQ DIMENSION

STATUS



CDE	Completeness	Uniqueness	Validity	Accuracy	Consistency
Purpose of the loan					
Address					
Name					
Employment status					
Loan ID					
Interest rate					
Default2					
Default					
Assets					
Loan term (months)					
Number of days past due					
Loan amount					
Date of birth					
LTV					
Credit score					
Income (monthly)					
Monthly payment amount					

TRIPLE A
RISK FINANCE

PERCENTAGES AND RAG PER TEST - ACCURACY

CDE	Mean	25th percentile	75th percentile	Min	Max	Inter Quartile Range	IQR outliers	Threshold lowerbound	Threshold upperbound	Number of records checked
Date of birth	18-3-1968	16-11-1949	21-2-1983	2-5-1921	11-9-9999	12150	✓ 97%	90%	95%	4304
Income (monthly)	3500	1873	4780	1200	39000	2907	✗ 89%	90%	95%	4304
Assets	750000	230000	1700000	0	3600000	1470000	92%	90%	95%	4304
Loan amount	600000	325500	1200000	0	2750000	874500	92%	90%	95%	4304
Interest rate	6	2,3	12,7	0	15.8	10	✓ 99%	90%	95%	4304
Loan term (months)	125	70	268	12	420	198	90%	90%	95%	4304
Monthly payment amount	780	240	3670	0	12000	3430	✗ 87%	90%	95%	4304
Credit score	7	3	8	0	10	5	✓ 96%	90%	95%	4304
LTV	70	45	80	20	100	35	✓ 96%	90%	95%	4304
Number of days past due	60	7	120	0	360	113	94%	90%	95%	4304

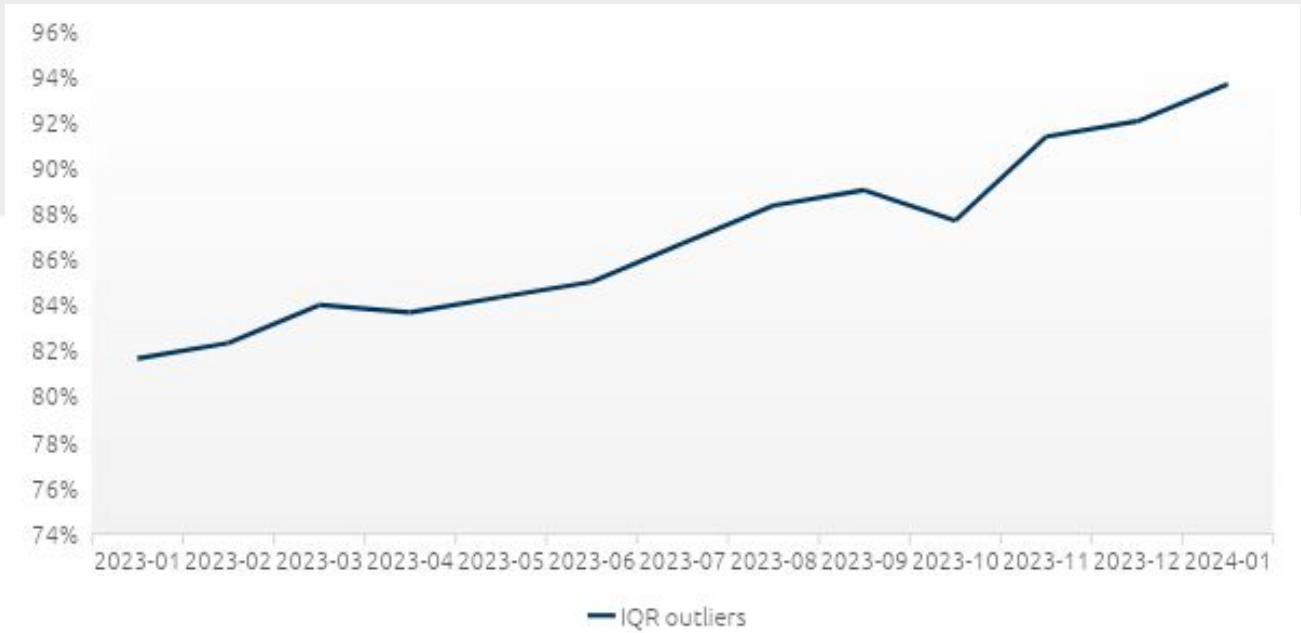
TRIPLE A
RISK FINANCE

OUTLIERS TABLE - ACCURACY

Test ID	Table	Loan ID (primary key)	IQR lower bound	IQR upper bound	Income
71	Customer_loans	2435553	143	6631	6906
71	Customer_loans	2488001	143	6631	12773
71	Customer_loans	55211845	143	6631	125
71	Customer_loans	78782110	143	6631	19550
71	Customer_loans	...	...	...	...

TRIPLE A
RISK FINANCE

AVERAGES OVER TIME - ACCURACY



TRIPLE A
RISK FINANCE

FUTURE

How to stay clean & what can we expect

TRIPLE A
RISK FINANCE

HOW TO STAY CLEAN

- Data Governance: have clear definitions of data policies and owners, who is responsible for the quality of which data domain
- Data Catalog: have all data definitions and mappings documented in a centralized tool or application (e.g. Collibra or Purview). This will improve data management and data trust
- Use the CDE values from the get clean phase as constraints on the input of the pension front-end systems. The transition is a great opportunity to improve the data ecosystem significantly
- We suggest to expand the existing Actuarial Function, Risk Management Function and Internal Audit Function with a new to be formed Data (Officer) Function

WHAT CAN WE EXPECT

1. Personalized AI Customer Service

“The new system will leave a lot of customers with specific questions that have never been asked before “

Through Natural Language Processing (NLP) AI can:

- Answer new and specific client questions
- Offer preliminary advice on retirement planning
- Help providing financial insights or even decisions
- Create insights based on specific client needs, goals and circumstances

Personalized AI pension chatbots can be trained in a secure environment on client portfolios without the intervention and costly hours of a person

WHAT CAN WE EXPECT

2. AI Risk Management, Predictive Analysis and Fraud Detection

Generative AI can be used to identify and predict risk in the (new) pension system
Recognizing patterns in the field of market trends, returns, demographic shifts etc

In a new system new "opportunities" for fraud arise!
Therefor AI will play a big role in recognizing not earlier seen anomalies and patterns

FRAUD DETECTION

WHAT CAN WE EXPECT

3. Personalized Interactive dashboards supported by AI

Portfolio insights are easiest gained by means of visuals such as images and graphs

Online dashboards in client portals are the perfect means to this end.

These dashboards will have AI supporting specific client needs.

“Clients will be able to prompt the dashboard regarding specific questions and needs”

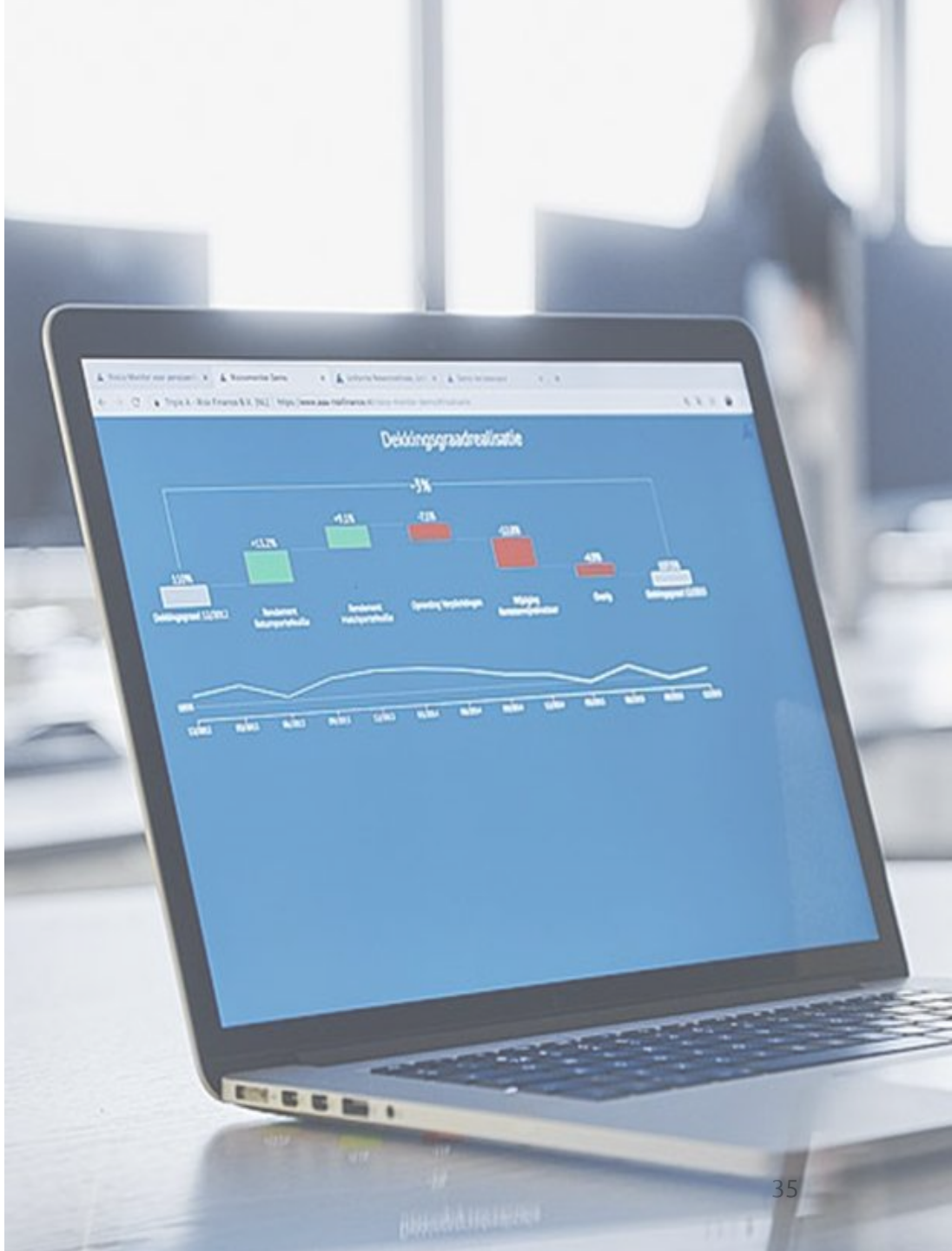
WHAT CAN WE EXPECT

4. Reporting and compliance processes automated by AI

Gen AI can be used to automatically generate and analyze reports
This helps with meeting the reporting requirement of WTP and other regulatory bodies

By automating processes:

- Errors are reduced
- Costly employee hours are saved
- More resources (employee hours) can be used for R&D and other improvements



TRIPLE A
RISK FINANCE

THANK YOU FOR YOUR ATTENTION

Hen Veerman
https://www.linkedin.com/in/hen-veerman-4946b22/

Pieter Stel
https://www.linkedin.com/in/pieterstel/