



Customer Matching Deduplication

—

AI deduplication and Matching



Challenges

- Create a master table: detect duplicate records in a dataset.
- Match two different tables
- Match a messy table against a clean one



Solution

Highly accurate and fast record deduplication and matching based on:

- Natural language processing
- User input: are records the same or not?

1

Train model

Users can provide input to the model through an app embedded in the report

2

Cluster records

The model returns the record with a cluster ID for related records in real-time

3

Gain insights

Improve data quality or perform business logic based on the origin of the dataset



Benefits

- Improve analytics by adding more value to the data
- Enrich and improve data quality by replacing less valuable data or adding missing data.
- Because of the user input, the user can safeguard the correctness of the matching.

Data

- Product and customer data
- Metadata: products and customers
- ...

Sources:



Integration



Census Data			
First Name	Last Name	Phone	Zip
Matt	Michelson	555-5555	12345
Jane	Jones	555-1111	12345
Joe	Smith	555-0011	12345

A.I. Researchers			
First Name	Last Name	Phone	Zip
Matthew	Michelson	555-5555	12345
Jim	Jones	555-1111	12345
Joe	Smeth	555-0011	12345


Arrows labeled 'match' connect the rows between the two tables: Matt to Matthew, Jane to Jim, and Joe to Joe.

Multinational bottling company

Industry Beverage

Coca-Cola European Partners is a British company formed as a result of the combination of the three main bottling companies for The Coca-Cola Company in Western Europe

 42 bottling plants

 + 33200 employees

 Serverd in +200 Countries



Challenges

- Gather sales figures from beverage wholesalers to extract **market insights**
- Beverage wholesalers use **different product descriptions** than coca-cola
- Product descriptions may contain **spelling mistakes**



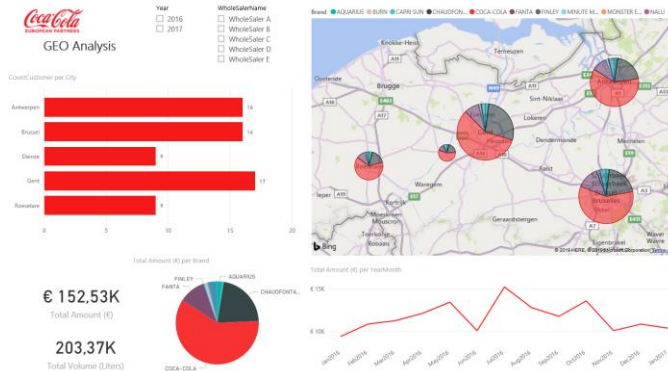
Solution

- Build a central **data platform** where beverage wholesalers can upload sales figures
- **Link** messy product descriptions of wholesalers to clean product descriptions of CCEP
- **Visualise** market insights



Benefits

- **Close to 100% mapping** of product descriptions
- **From 5 FTE** manual mappers **to 1** controller FTE



% Products Mapped over all wholesalers

