SUPER X
PROJECT RESULTS

Course:

Data Warehousing (581091)

HWR Supervisor:

Prof. Dr. Roland Müller

Presenters:

Nitish Acharya (1830596)

Vahe Shelunts (1850563)

Quynh Tran (569677)

Agenda

1. Requirement Analysis
2. Analysis of Data Sources
3. Multidimensional Design and Implementation
4. Dashboard
5. Process Intelligence
6. Business Recommendations

Requirement Analysis

Monthly **order** per retailer
Monthly **revenue** per retailer
Sold quantity of **OEMs and own products** per month
The average **deviation** of **forecast and actual order** per retailer
Top retailer per region
Top sold **products**
Average **price** of the sold **product** per quarter/month and per location
Average **shipping period** per **retailer**
Average **shipping period** per **order**
Cancellation % of **order** per retailer

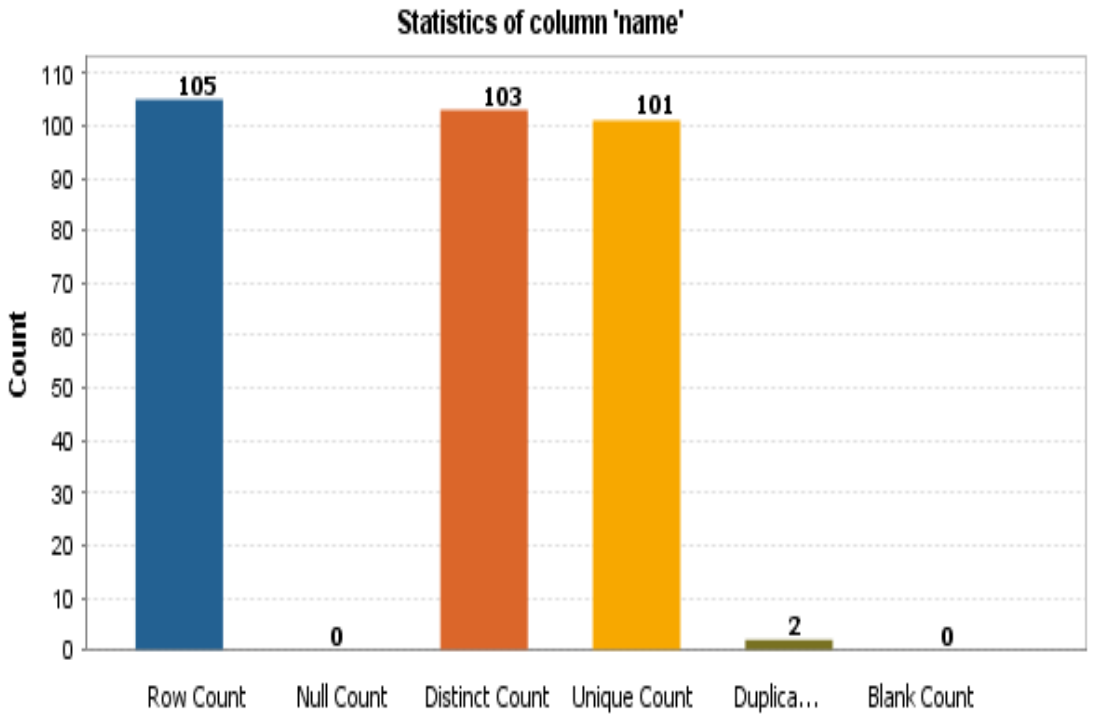
KPIs:

Sales (Revenue and Quantity)

Analysis of Data Sources

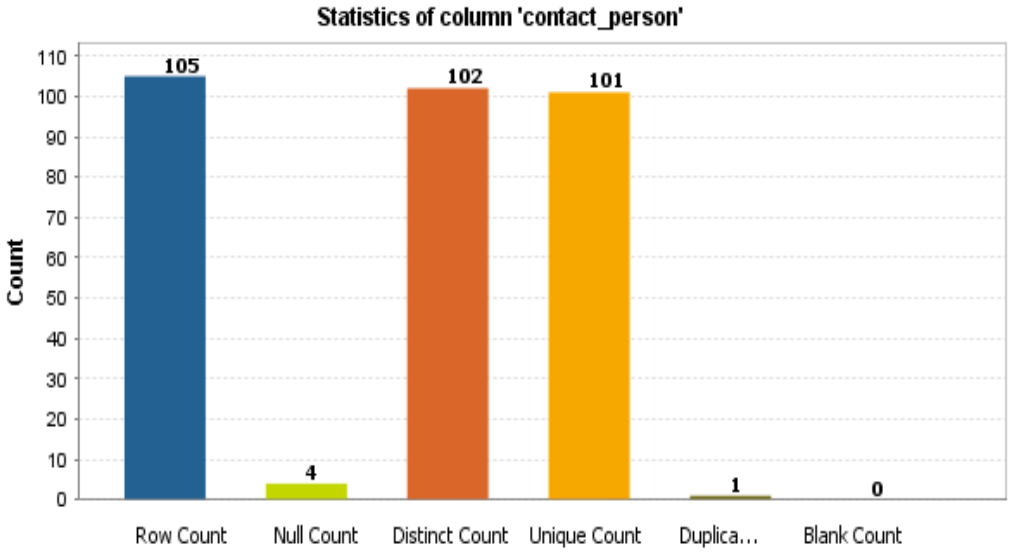
Table	Rows	Keys	indexes
orders	9111	1	3
order_items	104274	1	3
forecasts	121536	1	3
retailers	105	1	1
materials	31	1	1

Data Quality: Duplicates



id	name	category	description	address
34	Rath, Schuppe and Runte	e-shop	Optimized real-time software	8577 Carrie Trail, 82023-8823 Daleberg, Ohio, U.S.A.
98	Charles et Lucas	offline	Open-source 5th generation interface	97 Place de la Victoire, 16963 Beauvais, Languedoc-Roussillon, France
103	Charles et Lucas	offline	Configurable systematic pricing structure	97 Place de la Victoire, 16963 Beau%$is, Languedoc-*oussillon, France
104	Rath, Schuppe and Runte	e-shop	Reverse-engineered leading edge infrastructure	8577 Carrie Trai*, 8202%-8823 Dale$erg, Ohio, U.S.A.

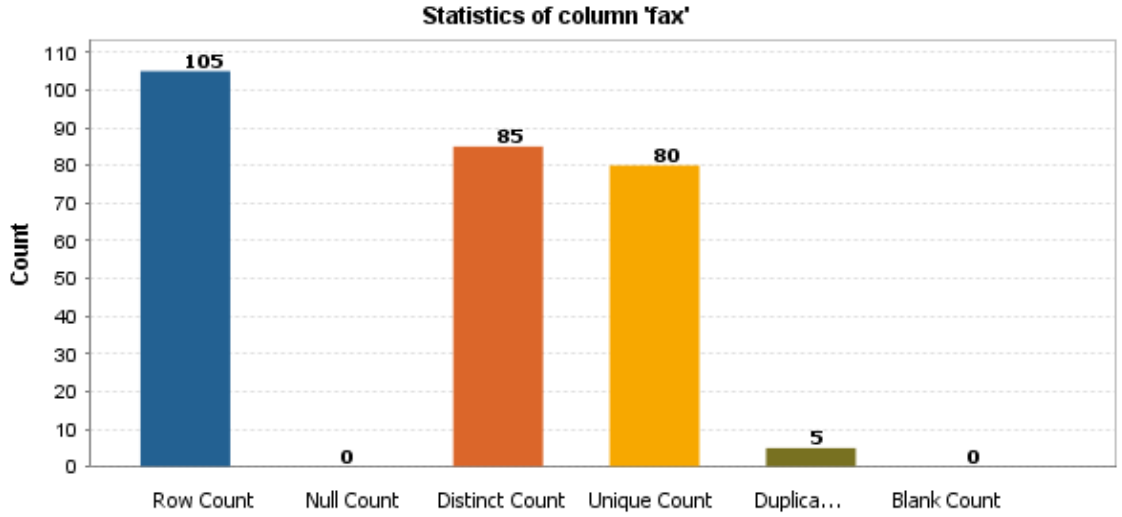
Data Quality : Missing values



Column: order_items.currency

Simple Statistics

Beschriftung	Count	%
Row Count	104274	100.00%
Null Count	4229	4.06%
Distinct Count	6	5.754E-3%
Unique Count	0	0.00%
Duplicate Count	6	5.754E-3%
Blank Count	0	0.00%



(999) 999-999

Data Quality : Invalid values

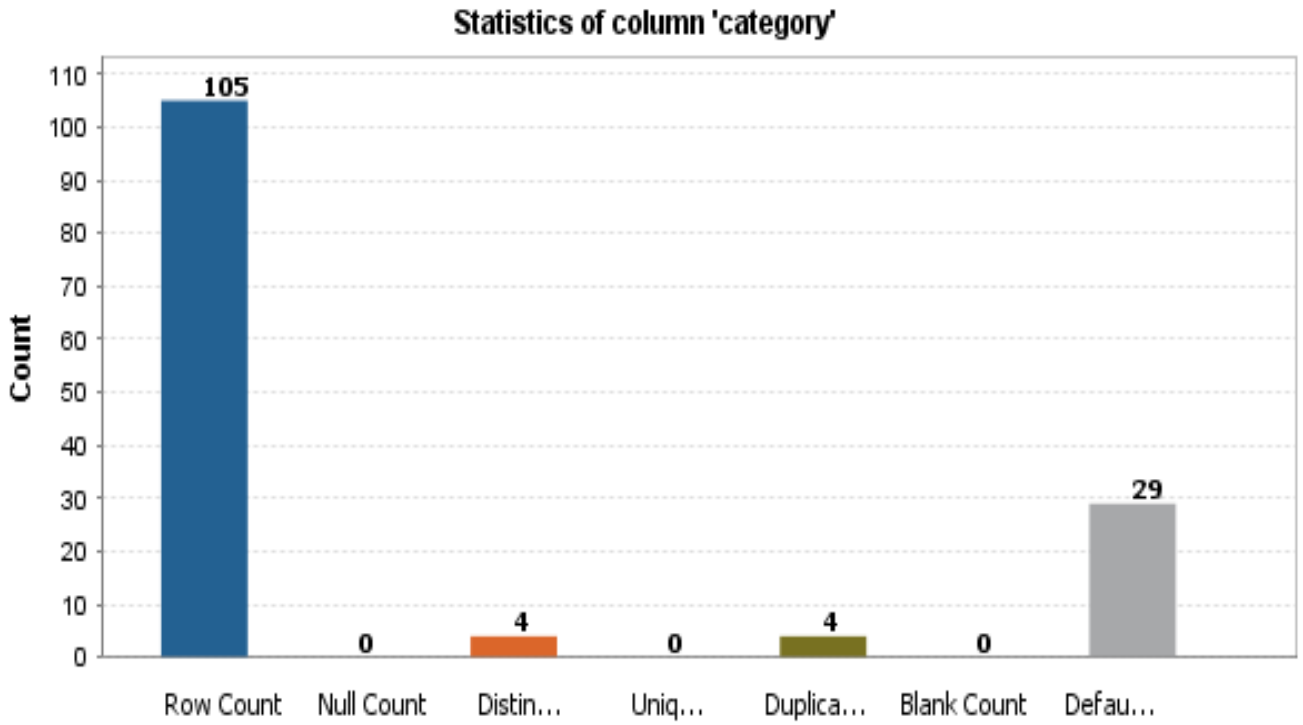
email
mme_baro _romane@gonzalez.com
eva.brin @vriesbrouwer.org
berge_beth@f hey.org
ebony.schaefer@harveytremblay com
prof_juliette_colin@ irard.fr
pri to.lucia.almonte@vega.es
jewel.osinski@rayno kirlin.info
elna_stehr@stantonn colas.biz
jermaine.rutherford@wit ing.ca
mazza_adriano@milan .org
mathis.m. eclerc@robert.org
niak.oleg.le@decfalkows i.pl
kobe. oodwin@towne.co.uk
urbansky_la ra@hommel.com
mclau hlin_derick@ullrichwest.com
wiza.darian@s orer.name
raik.birkemeyer@g eithanner.org
bradtke.sydnie@bergstr m.biz
olivier.alexis@richa d.net

▼ **Pattern Frequency**

Wert	Count	%
() -	20	19.05%
---	13	12.38%
Empty field	12	11.43%
	9	8.57%
-- a	9	8.57%
+	7	6.67%
+--	7	6.67%
--	6	5.71%
()	6	5.71%
.. a	5	4.76%

Phone number pattern frequency

Data Quality : Inaccuracy



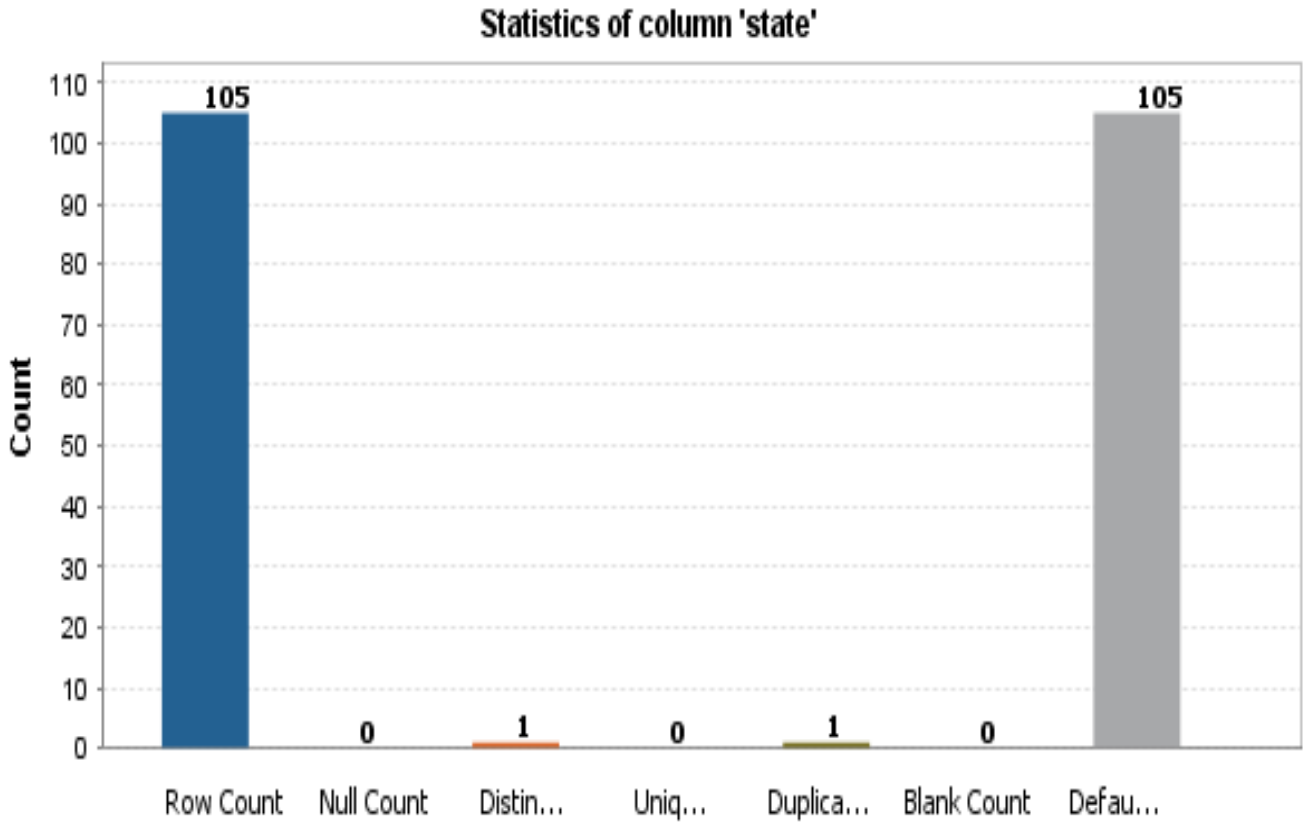
Online

Offline

e-shop

Offline retailer

Data Quality : Irrelevant

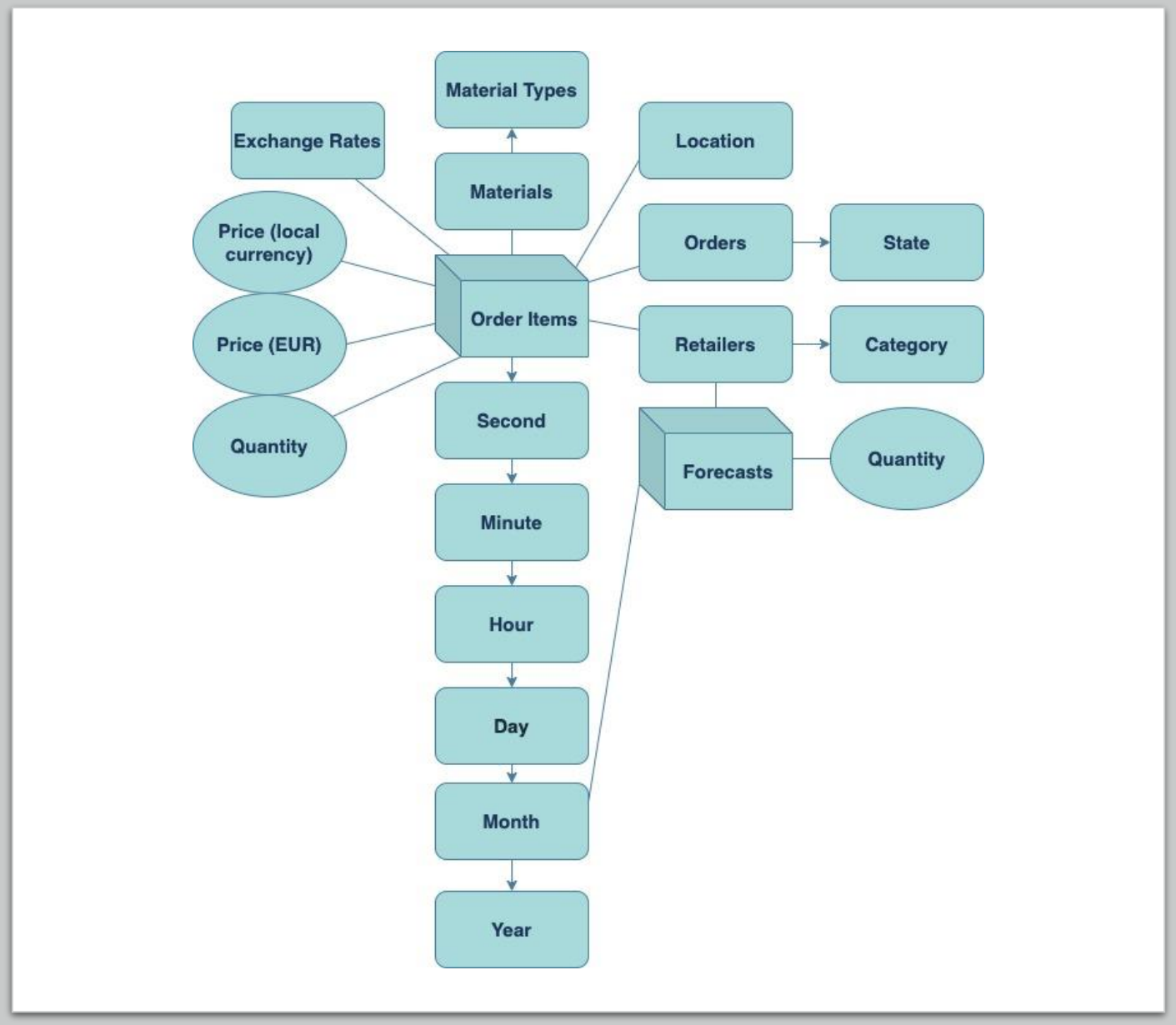


Single value

ACTIVE

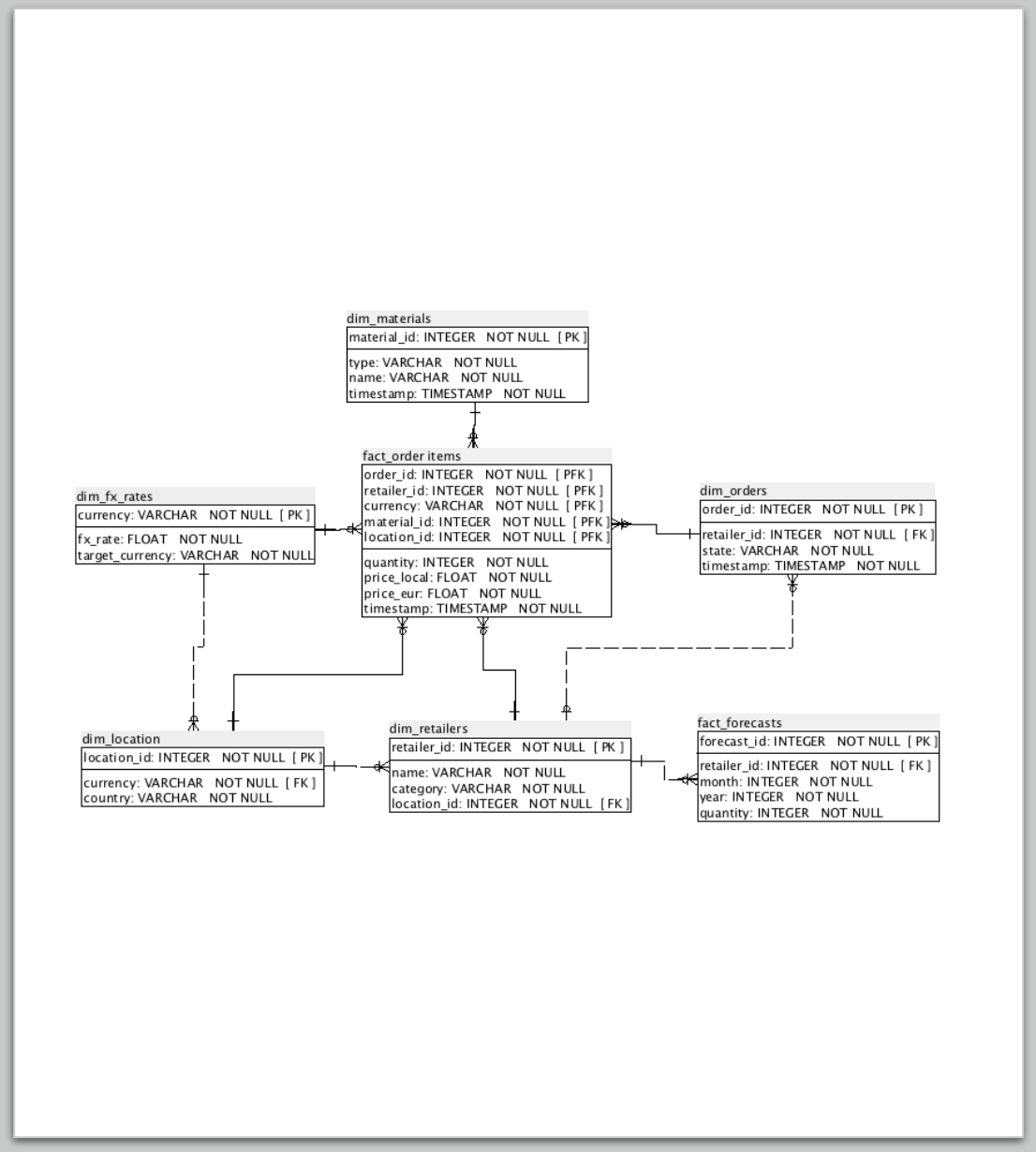
Multidimensional Implementation (Conceptual design)

- 5 dimensions
- 2 fact tables
- Hierarchies in dimensions

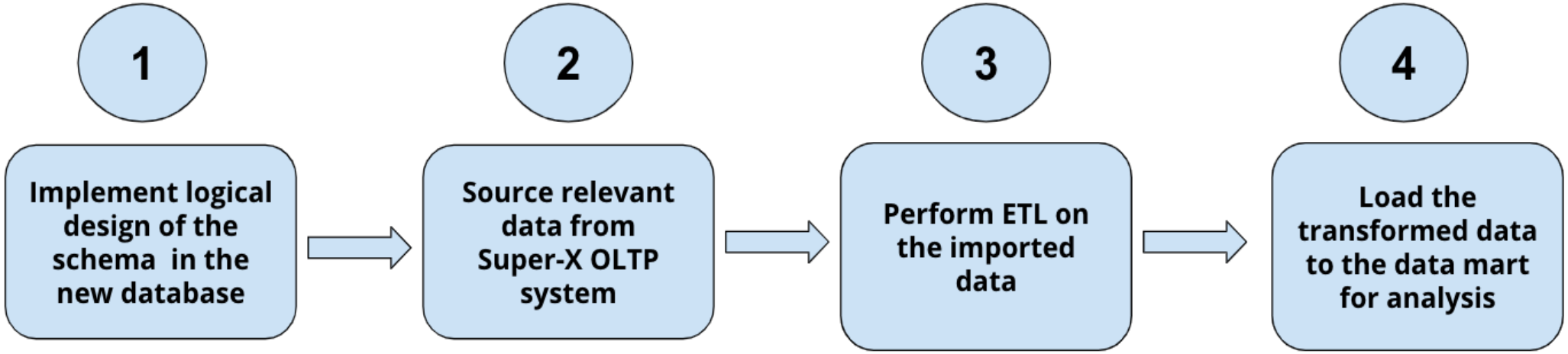


Multidimensional Implementation (Logical design)

- New dimensions (Exchange rates, Location)
- Fact table with primary foreign keys from dimensions
- Price (EUR), quantity – facts in Sales table
- Quantity – fact in Forecasts table

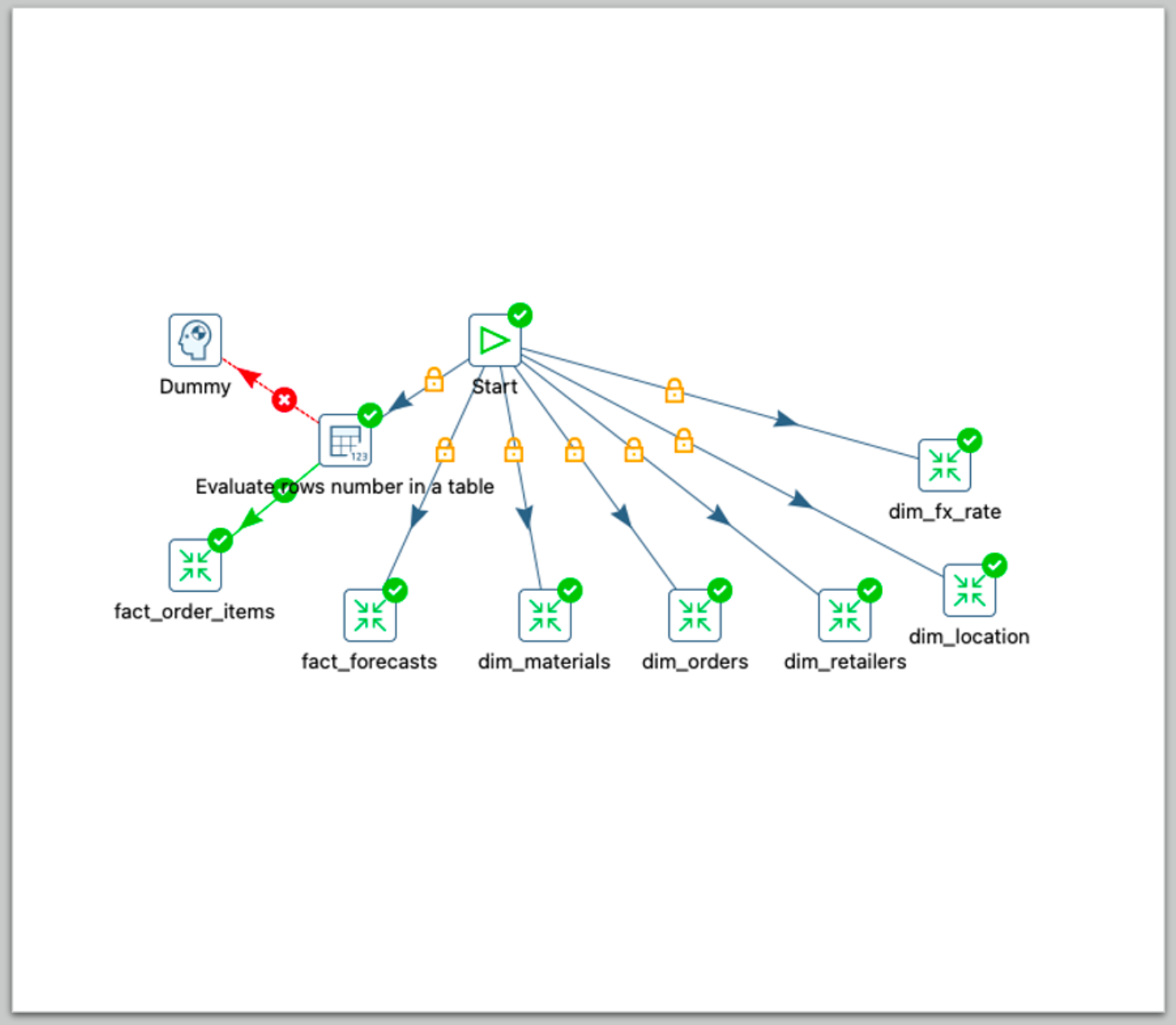


Multidimensional Implementation steps



ETL Overview

1. Exchange rates dimension
2. Location dimension
3. Retailers dimension
4. Materials dimension
5. Orders dimension
6. Forecasts fact table
7. Order items fact table



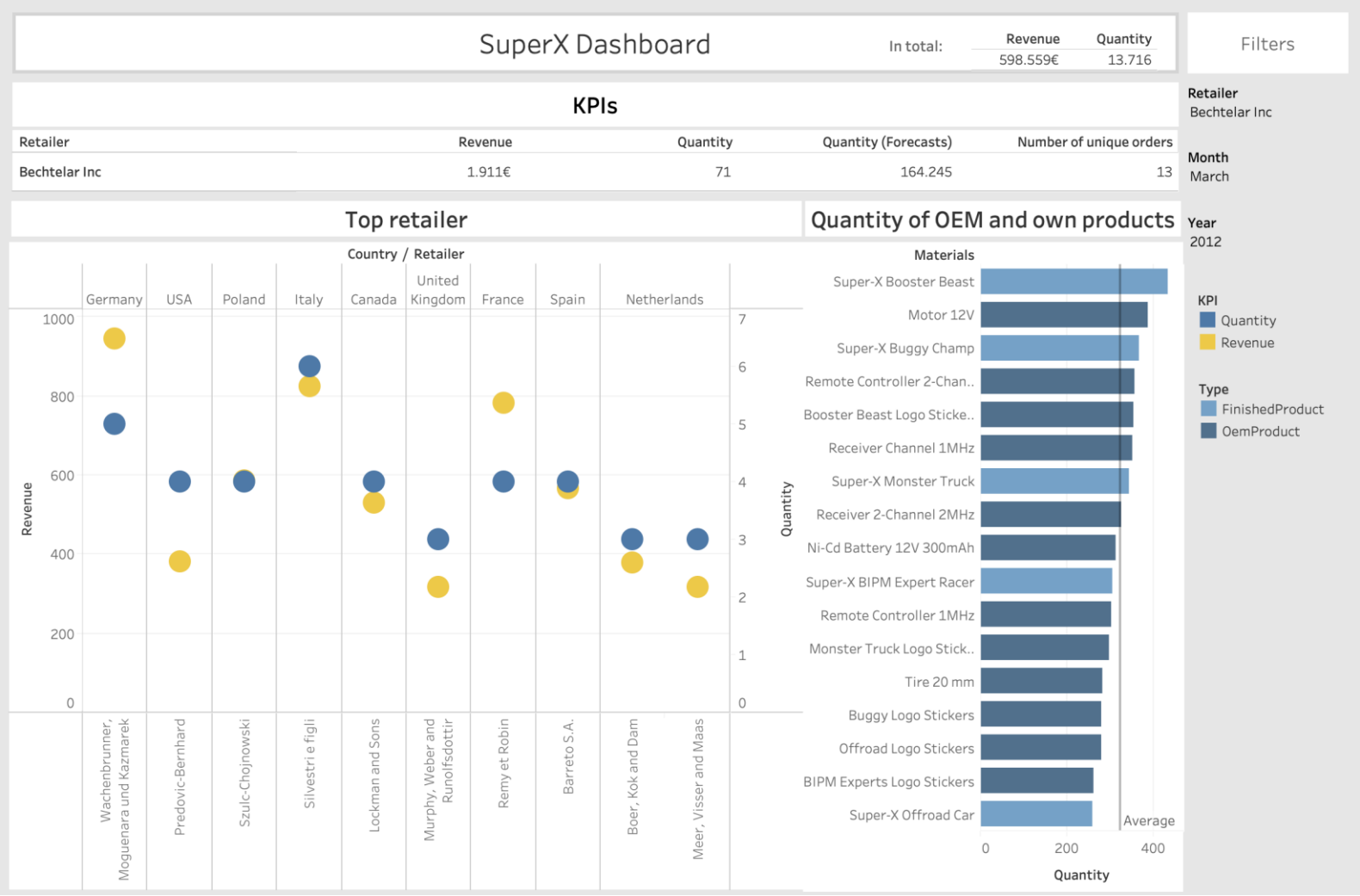
Exchange rates dimension

- Target currency – EUR
- Exchange rates defined for all currencies
- Aggregated average exchange rate across years

Year	USD/EUR	CAD/EUR	GBP/EUR	PLN/EUR
2012	0,78	0.7	1,23	0,24
2013	0,75	0,73	1,18	0,24
2014	0,75	0,68	1,20	0,24
2015	0,90	0,71	1,30	0,24
2016	0,90	0,68	1,22	0,23
2017	0,89	0,68	1,14	0,23
Average	**0,83**	**0,70**	**1,21**	**0,24**
Standard Deviation	**0,08**	**0,02**	**0,05**	**0,00**
% Standard Deviation	**9%**	**3%**	**4%**	**2%**

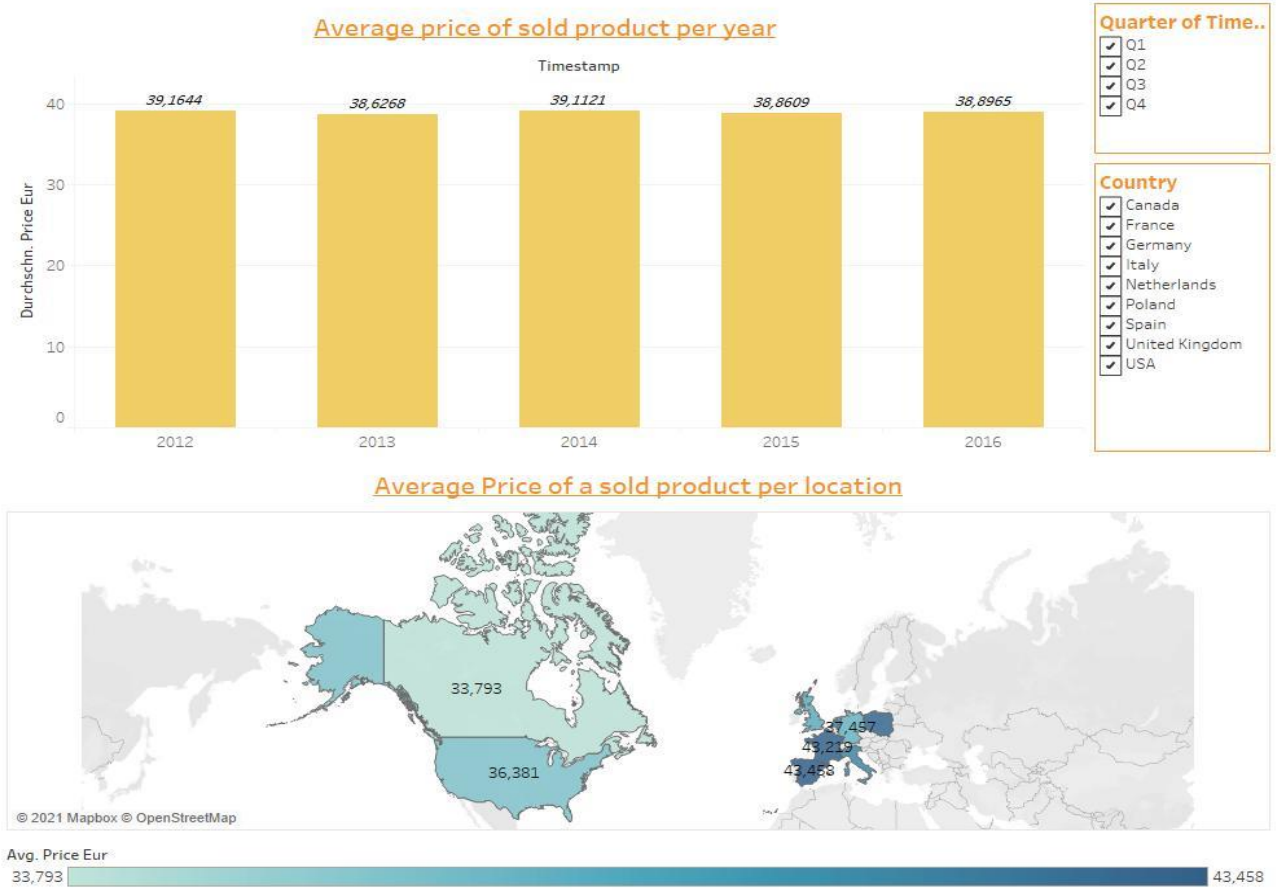
Source: https://www.ofx.com/en-au/forex-news/historical-exchange-rates

Tableau Dashboard



- Monthly order per retailer
- Monthly revenue per retailer
- Sold quantity of OEMs and own products
- Top retailer per region

Tableau Dashboard



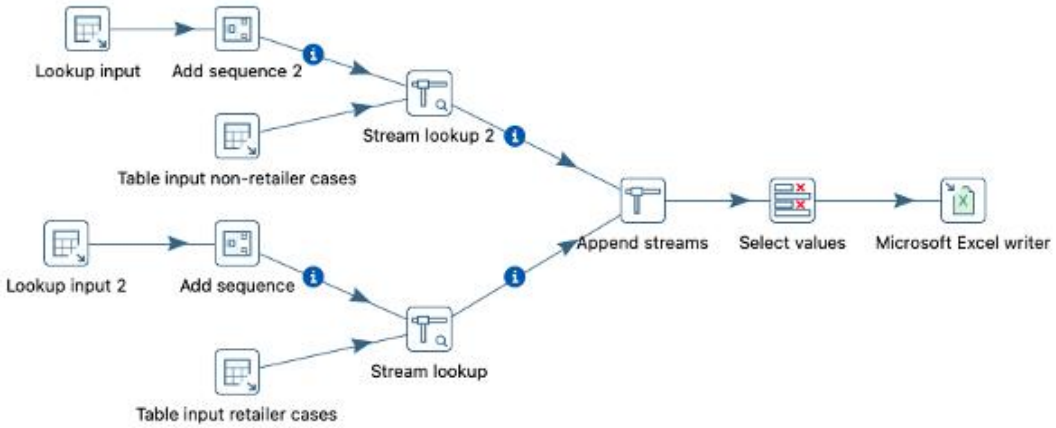
- Average price of sold product

Power BI Dashboard

- Top sold products
- Cancellation % of order per retailer

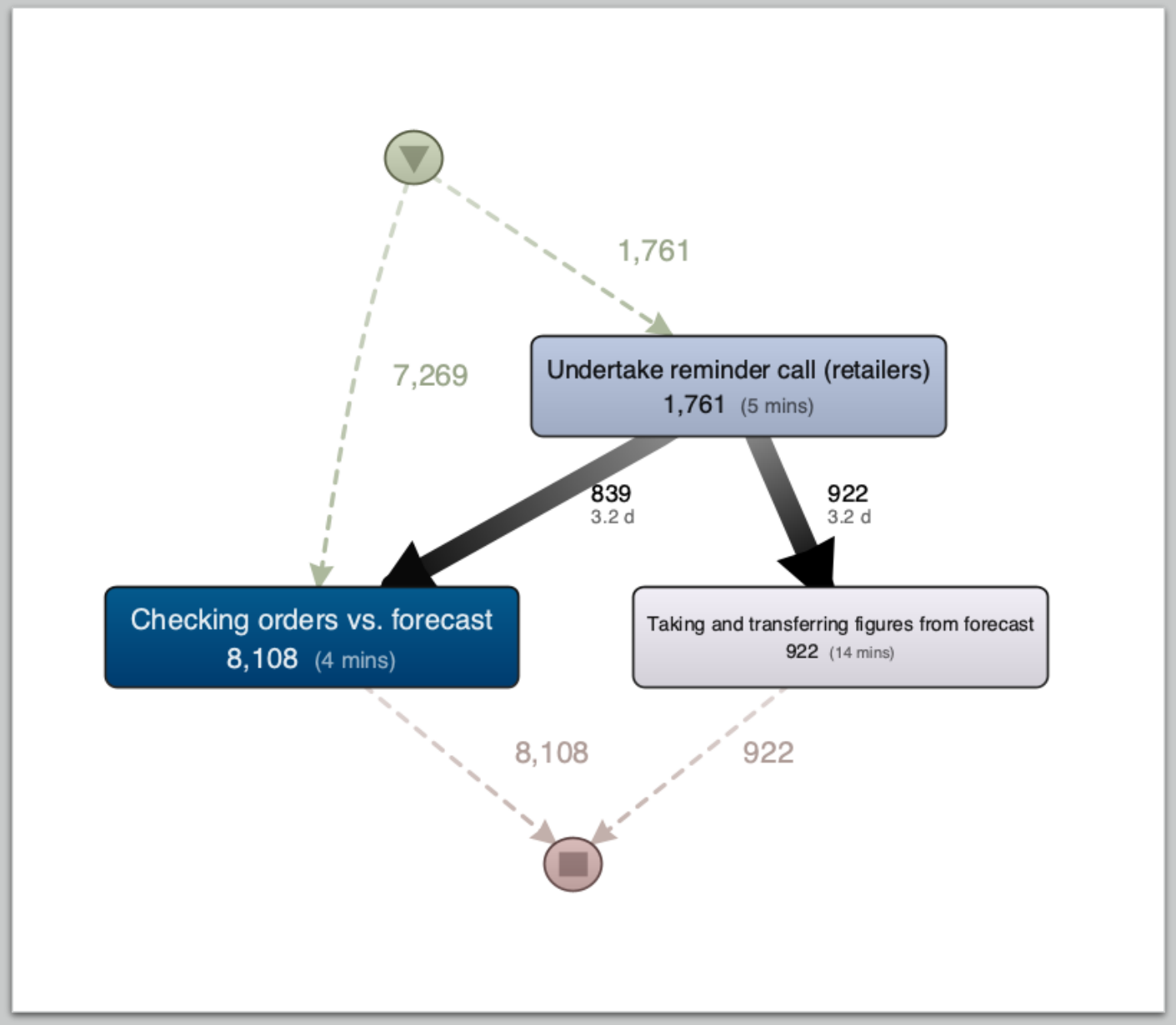
Process Intelligence overview

- Querying sales logs
- Creating Case IDs for Process instances
  - Retailer-order-based cases (on the basis of retailer ID, month and year)
  - Cases not attributable to specific retailers (on the basis of month and year)
- Feeding the output for process mining



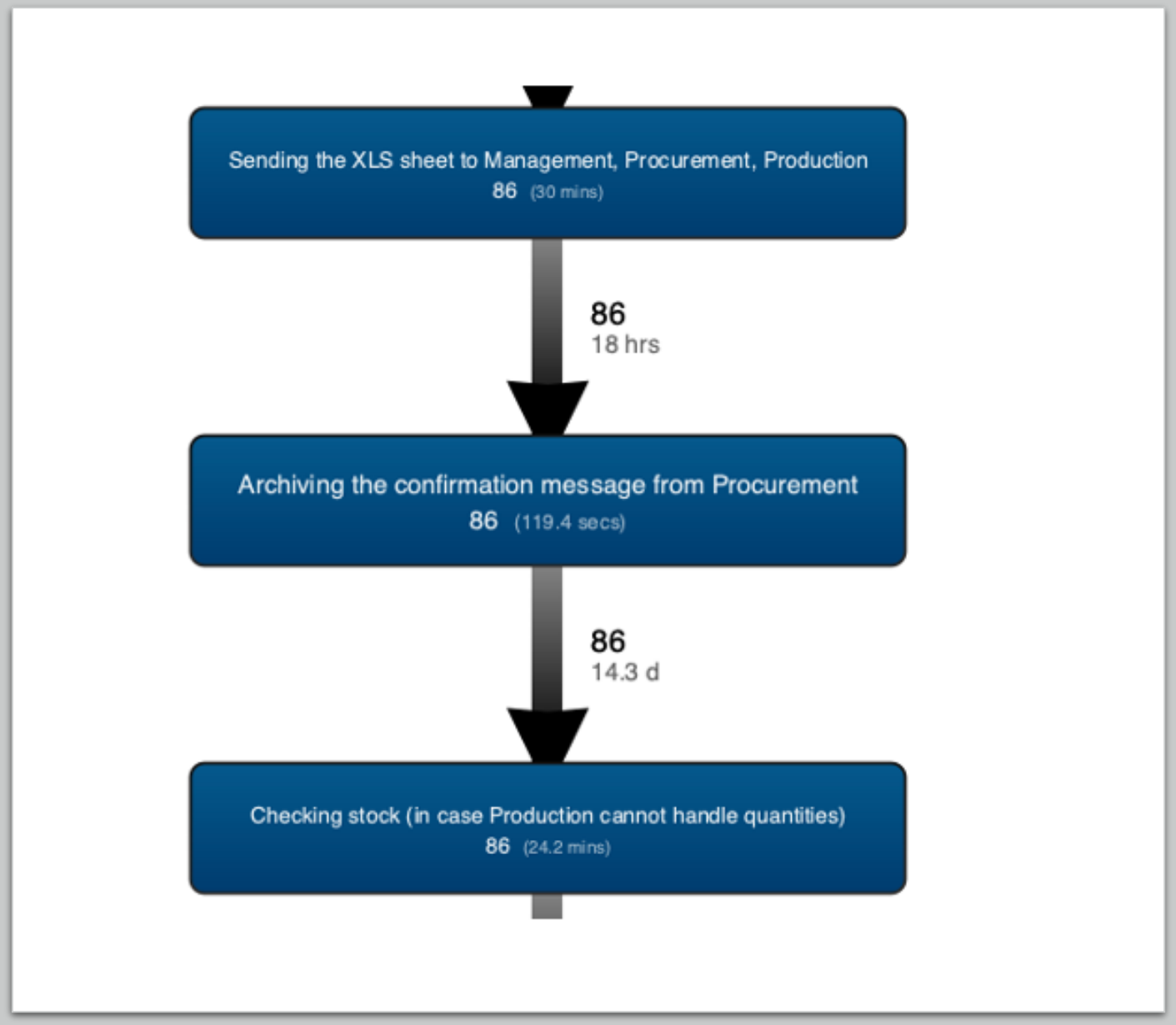
Retailer-order-based cases

- In around 17 % of cases, reminder call is initiated
- Waiting time 3 days (until retailers send their orders)
- Over 50 % of reminder cases are unsuccessful (figures transferred manually from the forecast)



Monthly-based cases

- Media breaks increase processing time
- Redundant and non-value-adding activities
- Potential bottlenecks in the system (takes 14 days for sales to check stock)



Business Recommendations

Identification of root cause for data quality problems

Implementation of Change Data Capture

Implementation of a dashboard

Introduction of an IT system for Sales operational work

Thank you for your attention!