BOAK

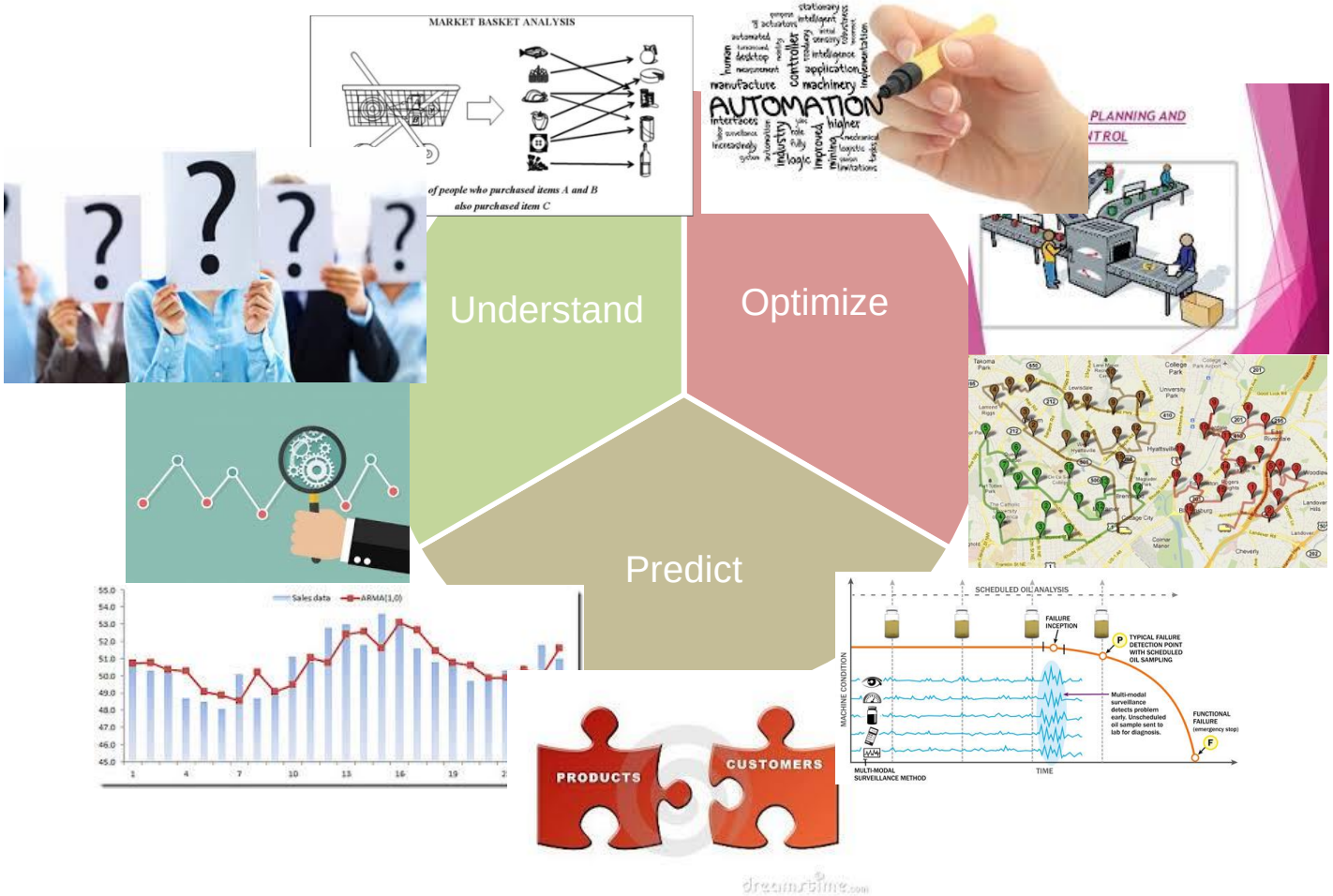
BUSINESS OBJECTS ARBEITSKREIS | 2017

Data Science und Machine Learning in der Praxis

Üetliberg, September 12th, 2017

www.boak.ch

BUSINESS PROCESSES & DATA ANALYTICS

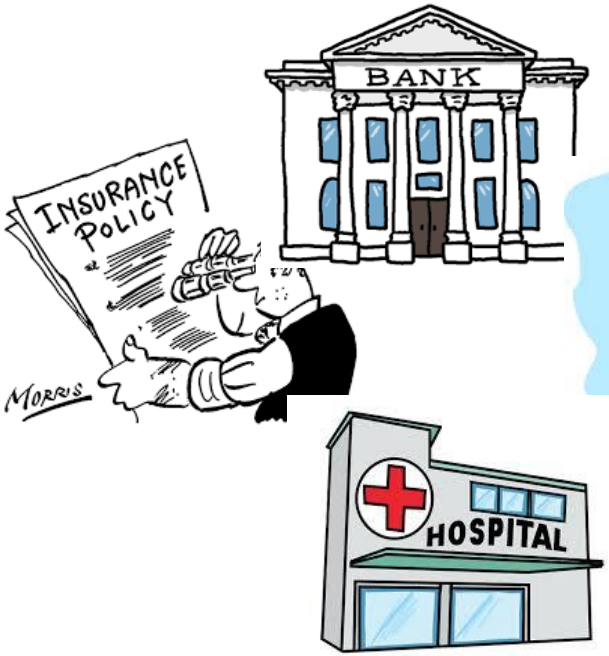


BUSINESS PROCESSES & DATA ANALYTICS

- Value Chain Model

Inbound Logistics → Operations → Outbound Logistics → Marketing & Sales → Service

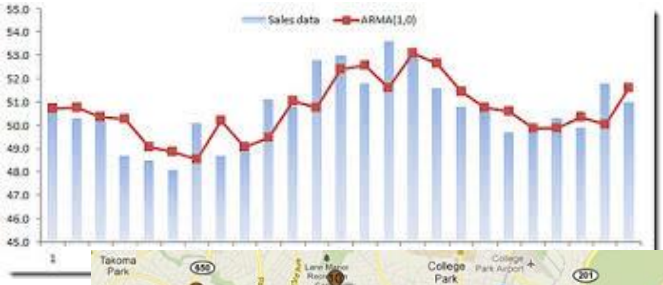
Procurement, HR Management, Infrastructure, Technological Development

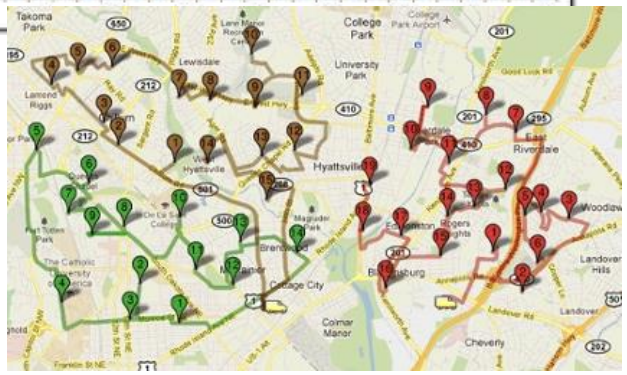


BUSINESS PROCESSES & DATA ANALYTICS

- Value Chain Model

- Forecasting of sales
- Efficient fleet rooting
- …





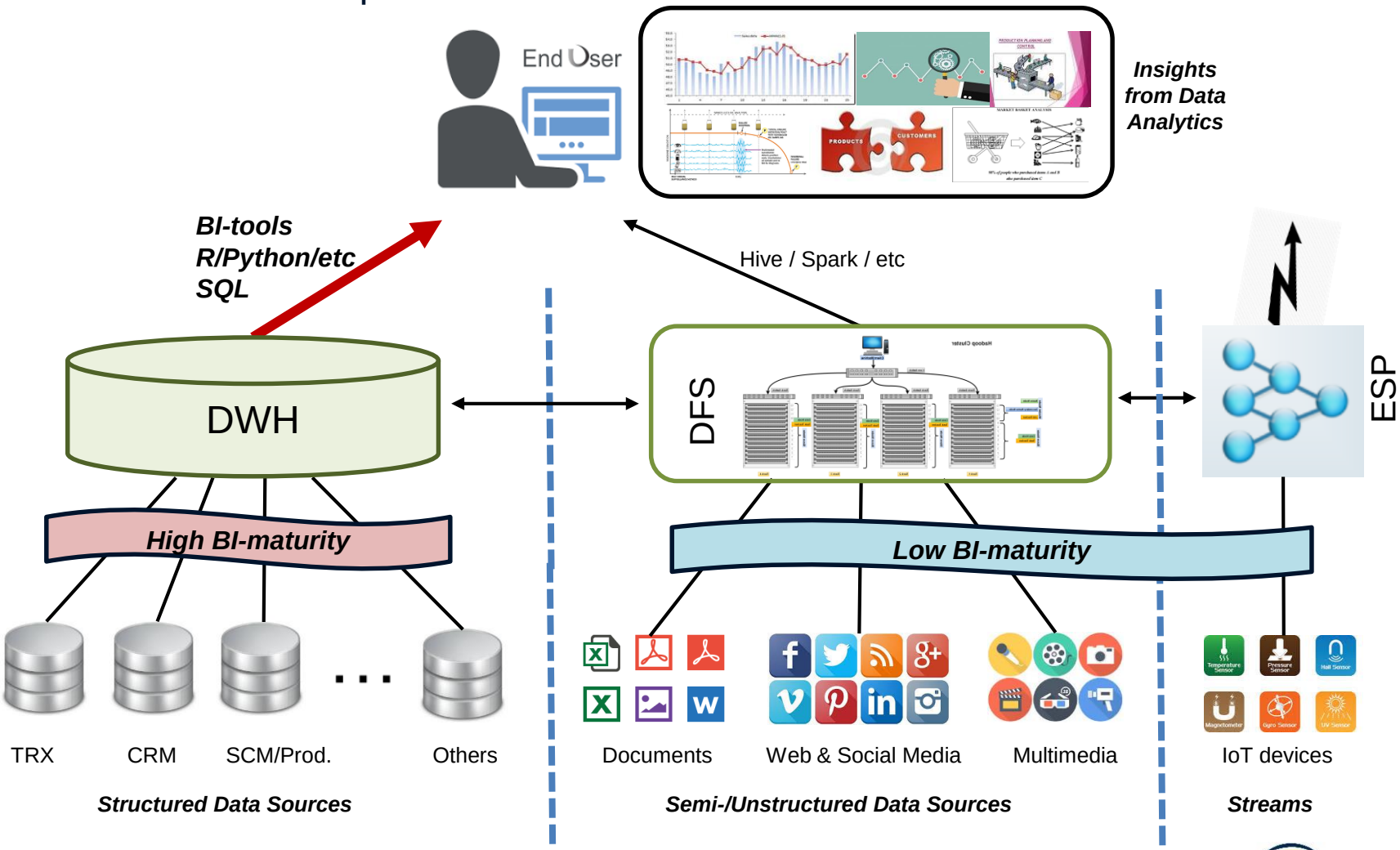
BUSINESS PROCESSES & DATA ANALYTICS

- Value Chain Model

- Customer understanding
- 1:1 marketing
- …

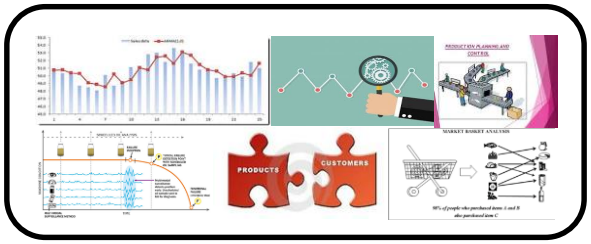
DATA SCIENCE & ARCHITECTURE

- Swiss mid-sized companies: current status



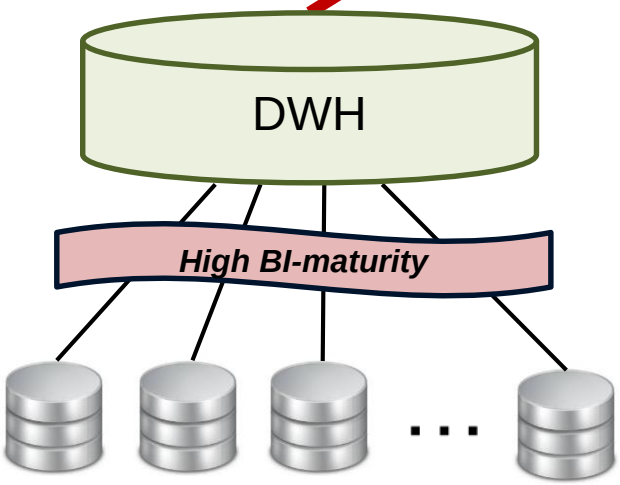
DATA SCIENCE & ARCHITECTURE

- Swiss mid-sized companies: current status



Insights from Data Analytics

BI-tools
R/Python/etc
SQL



Structured Data Sources

Infrastructure

Automation

IBM sas rapidminer KNIME alteryx Microsoft SAP Angoss BOAK

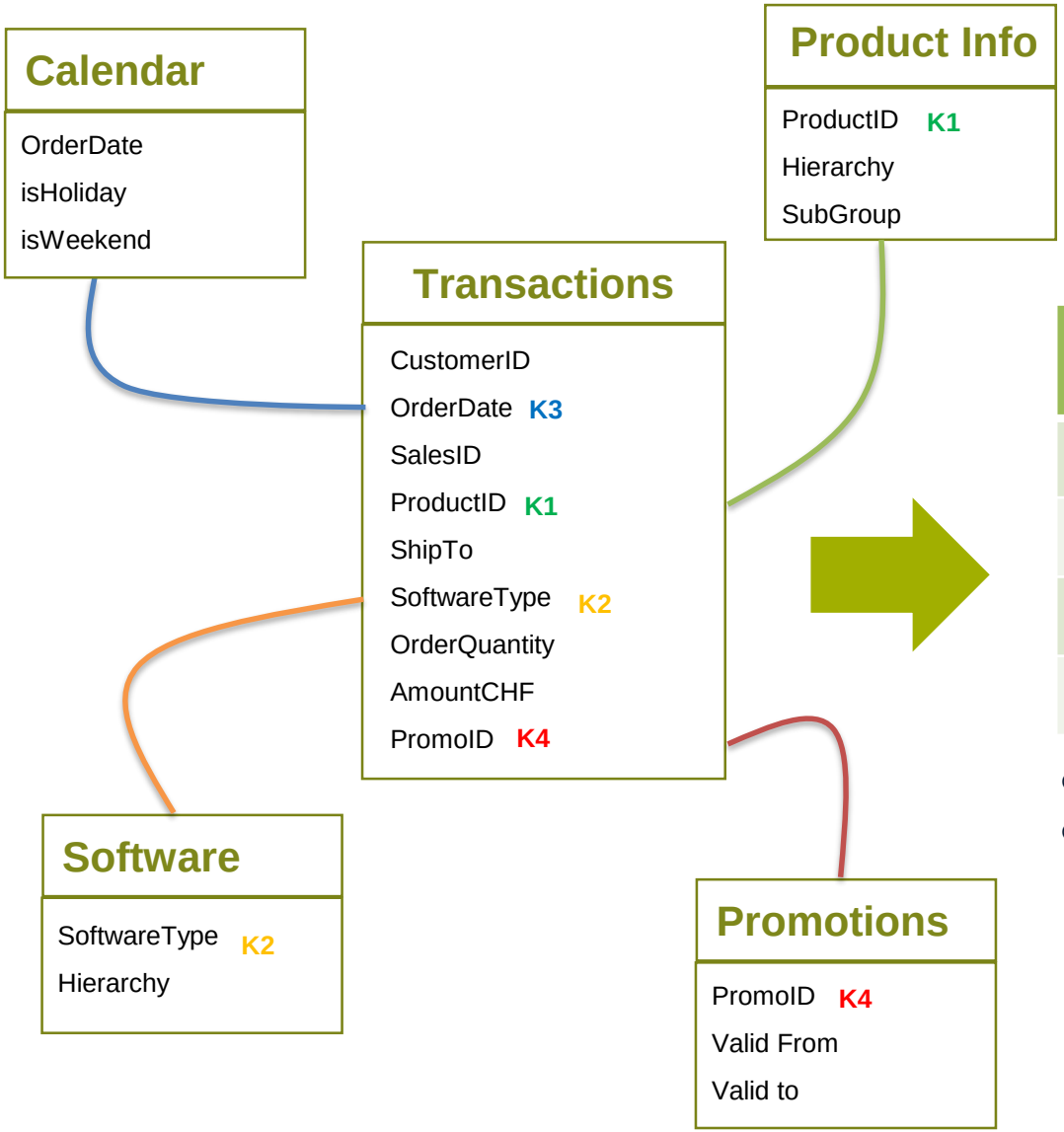
OUTLOOK

- Customer analytics
  - Better understand your customers
- Predictive analytics
  - Run predictions of quantities
- Geo-analytics
  - Leverage geodata
- Others
  - Data analytics projects @IT-Logix

CUSTOMER ANALYTICS

BOAK
BUSINESS OBJECTS ARBEITSKREIS | 2017

CUSTOMER ANALYTICS



Customer ID	feat1	…	featN
5D846CEF…			
70CEA195…			
….			
826E1AE7…			

- Table / One row for each customer
- More dimensions for each customer

CUSTOMER ANALYTICS

- Extraction of customer features
  - Table / One row for each customer
  - More dimensions for each customer

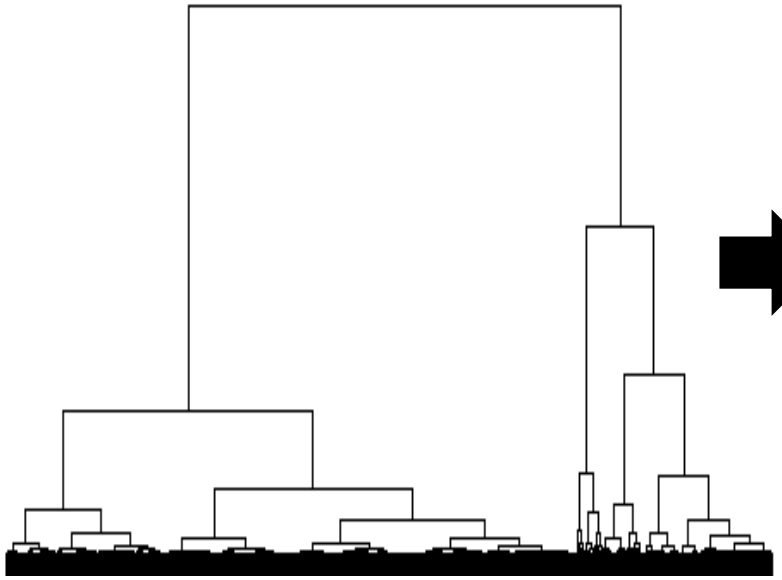
Customer ID	Feat-1	…			Feat-N
5D846CEF…	Transactional Features	Product Preference Features	Channel Preference Features	Calendar Preference Features	Marketing Preference Features
70CEA195…					
….					
826E1AE7…					

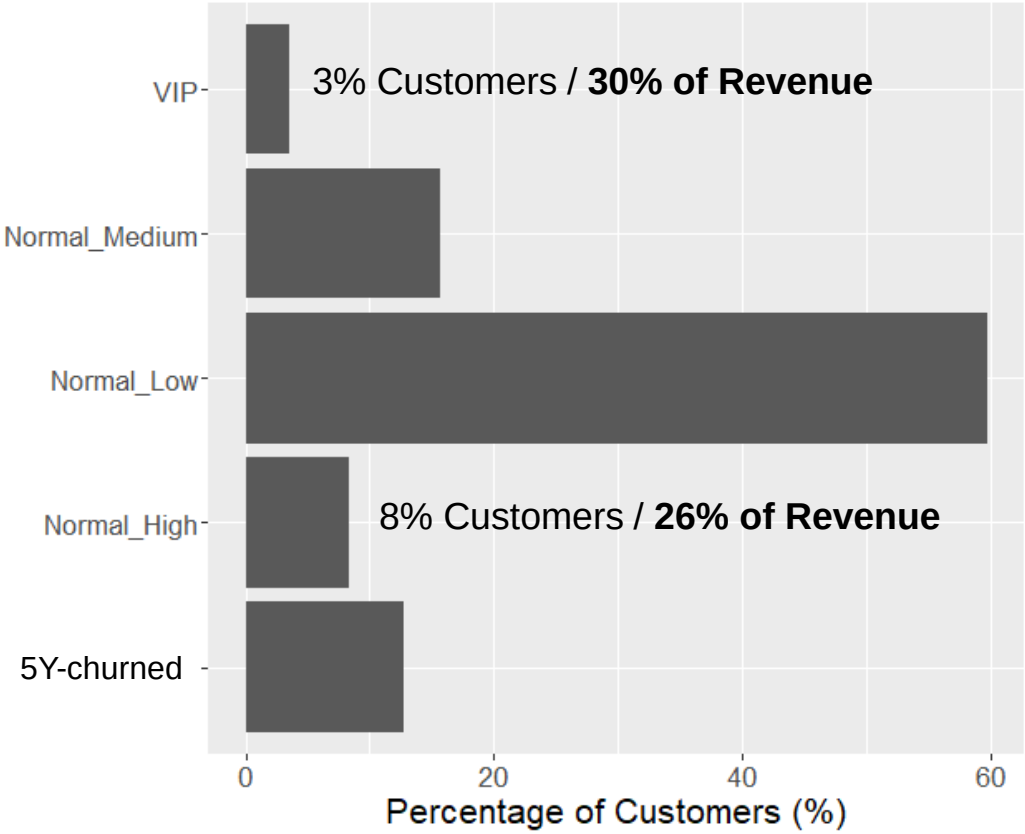
Transactional Behavior	Product Preference	Channel Preference	Time Preference	Marketing Preference

CUSTOMER FEATURES

Transactional Features

- Amount Class
  - Clustering based on the “Purchase.Amount.Score”
  - Purchase.Amount.Score = **4 x** [1Y]$_{CHF}$ + **2 x** [2Y]$_{CHF}$ + **1 x** [3Y-5Y]$_{CHF}$

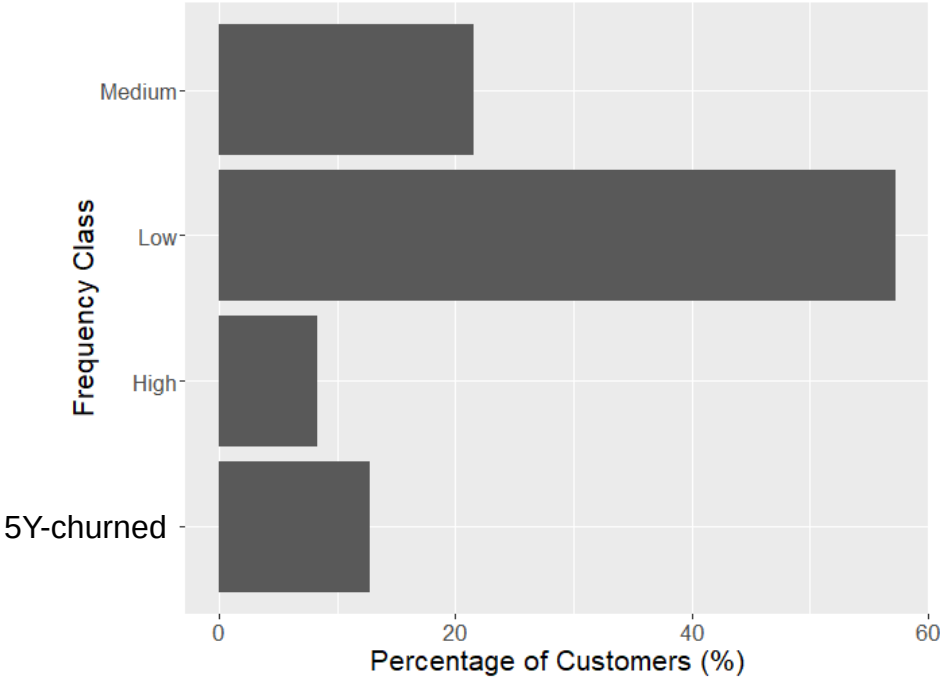


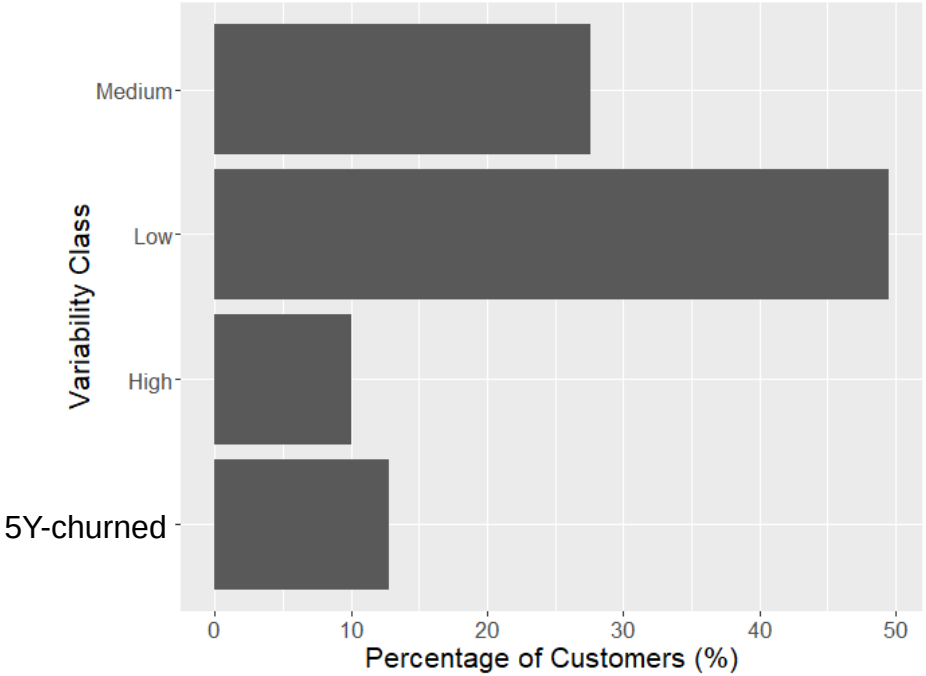


CUSTOMER FEATURES

Transactional Features

- Frequency & Variability Class

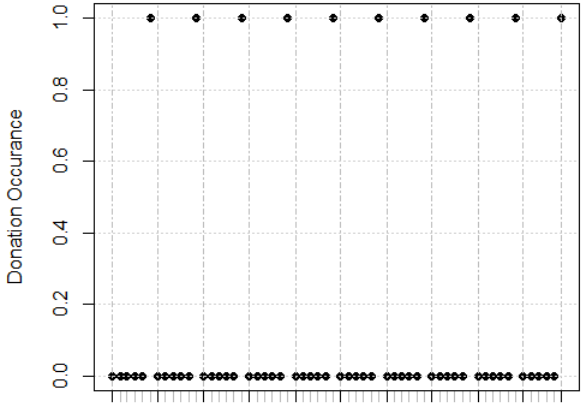




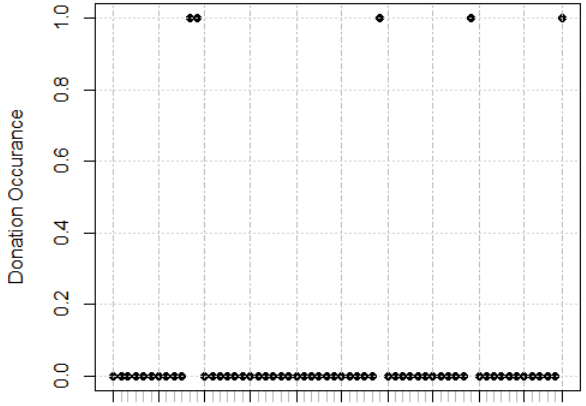
CUSTOMER FEATURES

Transactional Features

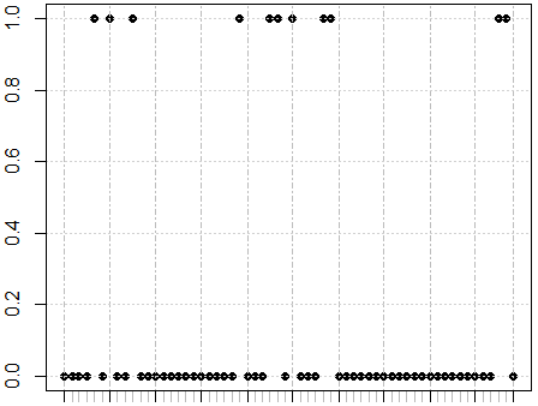
- Periodicity Class



Periodicity Class	Customer Base (%)
Strong	0.5%



Periodicity Class	Customer Base (%)
Weak	5%

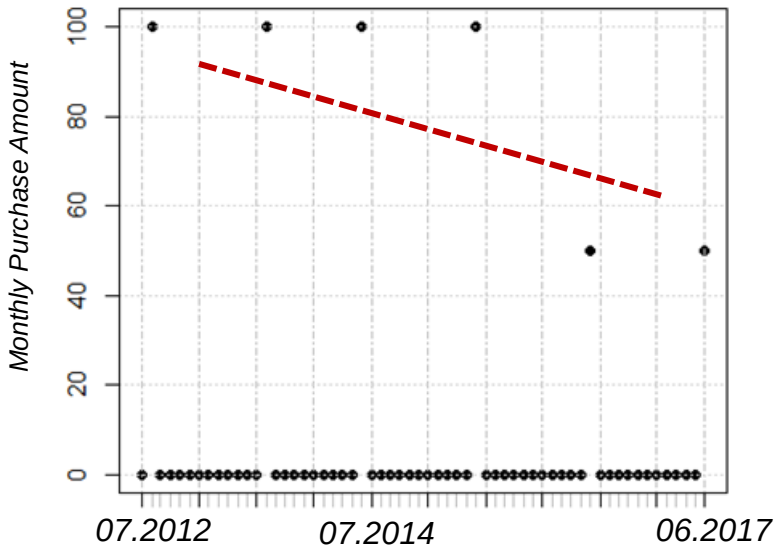


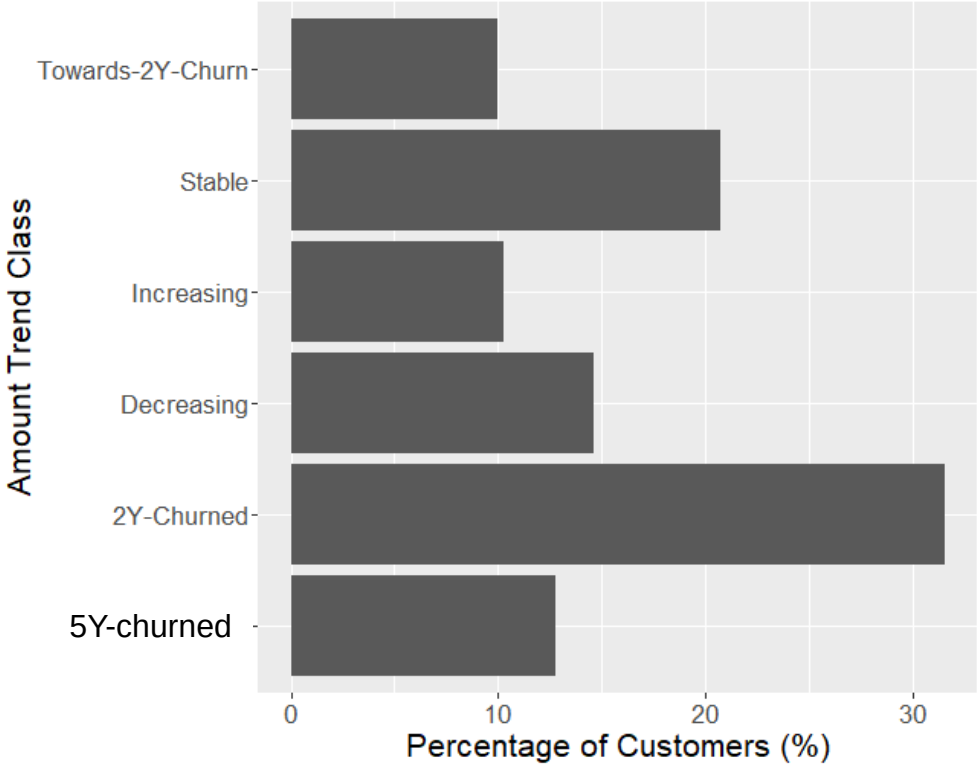
Periodicity Class	Customer Base (%)
Aperiodic	82%

CUSTOMER FEATURES

Transactional Features

- Annual Amount Trend Class





CHANNEL PREFERENCE FEATURES

Channel Preference Features

- Based on channel category

Frequency of Mobile
Frequency of Download
Frequency of Online
Frequency of Diverse
Frequency of FALSE

- Example:

Customer ID	Channel.Mobile	Channel.Online	Channel.*
07E58F094891FF5….	0.25	0.75	0

25% of times via Mobile Channel

75% of times via Online Channel

CUSTOMER ANALYTICS

Customer Behavior Table

Customer ID	Feat-1	…			Feat-N
5D846CEF…	Transactional Features	Product Preference Features	Channel Preference Features	Calendar Preference Features	Marketing Preference Features
70CEA195…					
….					
826E1AE7…					

- Additional customer dimensions
- Contains DA in an “easy-to-use” form
- Basis for personalized marketing

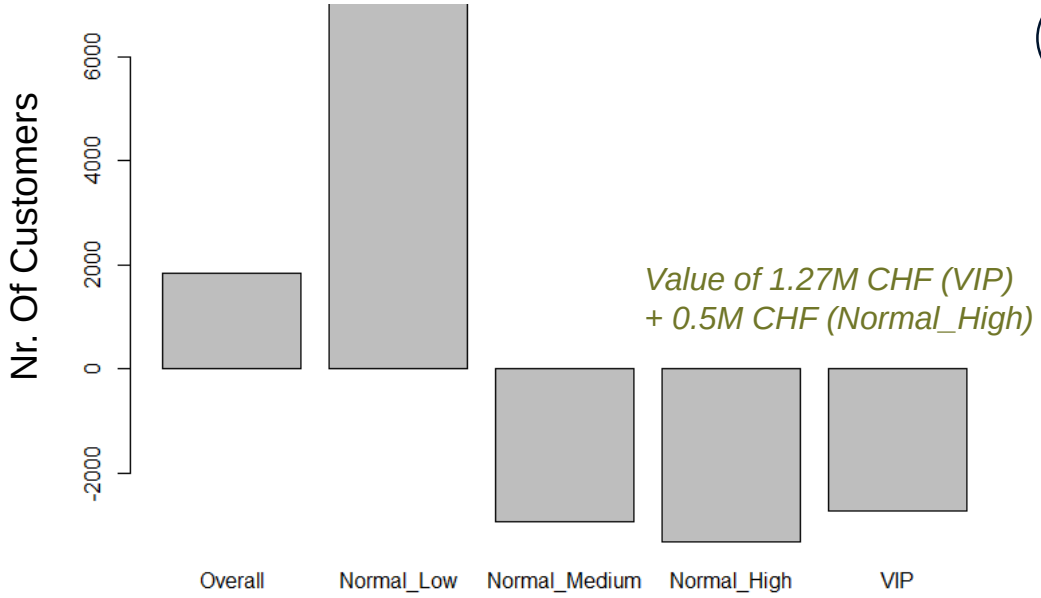
THE AFTER-"MATH"

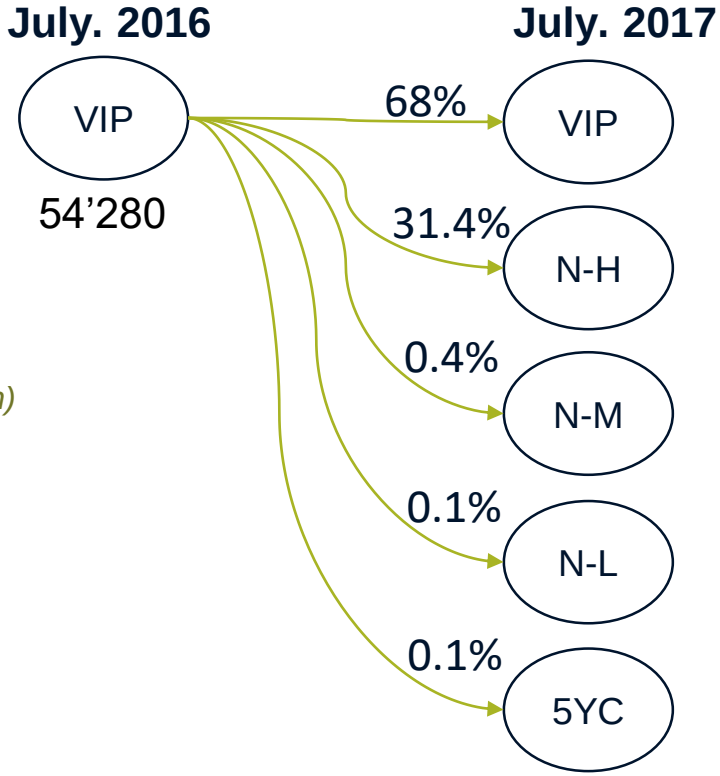
- Immediate Value: Retention, Re-activation, Acquisition

Retention

- "VIP" & "Normal_High": **11%** of the customer base
- They make **~56%** of the total revenue

Difference between July. 2017 and July. 2016





THE AFTER-”MATH”

- A/B Testing

Empfänger:
VIP, Normal High

Testszenario wie folgt:	**Testcases**	**Retention Cycle:**
	A) 5’000 Adressen	**kein Rabatt**
	B) 5’000 Adressen	**CHF 20.- Rabatt auf alle DPP’s**
	C) 5’000 Adressen	**20% Rabatt auf alle DPP’s**

Gültigkeit der Aktion: ca 4-6 Wochen. 14 Tage vor Ablauf der Gültigkeit wird nochmals ein Reminder an alle Nicht-Käufer versendet.

- Choice of customers based on:
  - Amount.Trend = “Decreasing” or “Towards-Churn”

THE AFTER-"MATH"

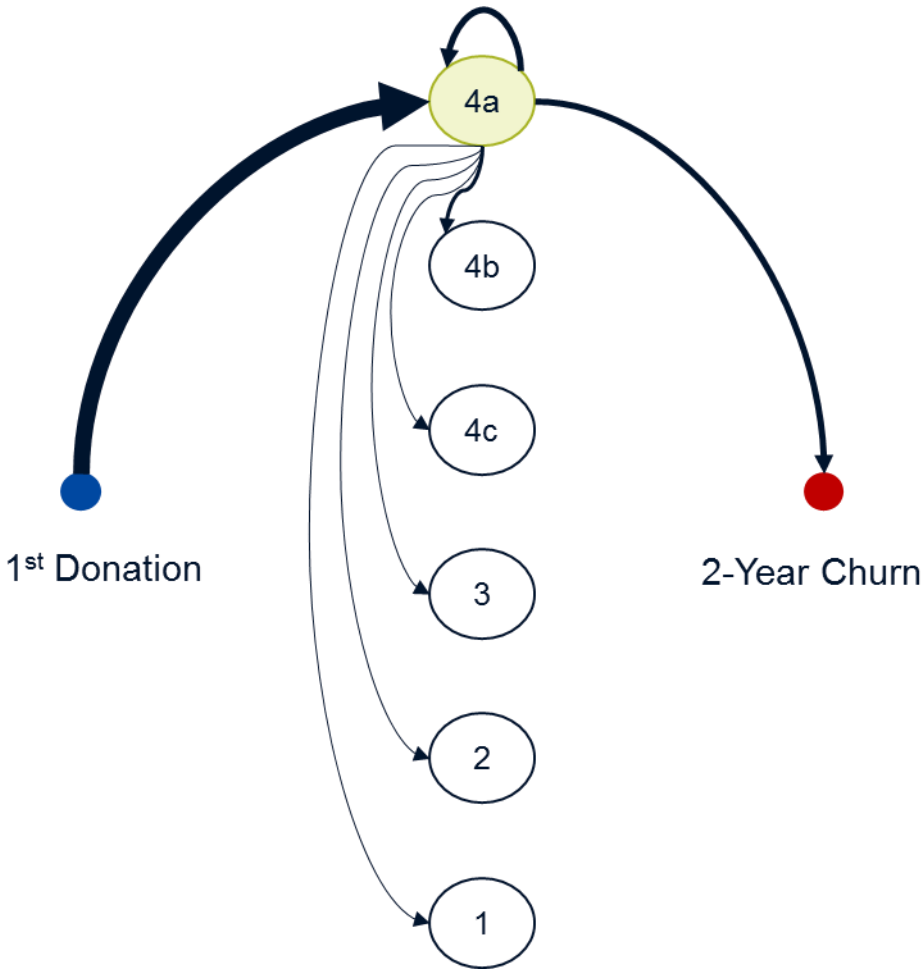
- Flexible grouping of donors

ID	Features	Population (%)	Revenue (%)
4a	Amount_Group= Normal_low Nr. Of Reactions 5Y = 1	30%	5.5%
4b	Amount_Group= Normal_low Nr. Of Reactions 5Y > 1 Frequency = Low Periodicity = Aperiodic	33%	14%
4c	Amount_Group= Normal_Low Nr. Of Reactions 5Y > 1 Frequency = Medium, High | Periodicity = weak, strong	22%	14%
3	Amount_Group = Normal_Medium	13%	28%
2	Amount_Group = Normal_High	2%	10%
1	Amount.Class = Special	1%	30%

ID	Features	Population	Revenue (%)
3p	Periodicity = Strong,Weak Frequency = Low	2.6%	5.4%

ID	Features	Population	Revenue (%)
3f	Frequency = Medium, High	7.2%	15 %

- Monitoring of groups



THE AFTER-"MATH"

Optimize Marketing resources

Group	Donors (%) / Revenue (%)	Marketing Spending	PROPOSED ACTIONS
4a	30% / 5.5%	19%	• Less actions, but in proper timing • Mild Retention Process for new donors • Soft Retention Process for 2Y-churned donors
4b	33% / 14%	39%	• Less actions, but in proper timing • Mild Retention Process for new donors • Soft Retention Process for 2Y-churned donors
4c	22% / 14%	25%	• No changes • Full Retention Process
3	13% / 28%	14%	• Increase Marketing touchpoints • Full Retention Process • Fast Churn Monitoring
2	2% / 10%	2%	• Increase Marketing touchpoints • Full Retention Process • Fast Churn Monitoring / Personalized contact
1	1% / 30%	1%	• Increase Marketing touchpoints • Full Retention Process • Fast Churn Monitoring / Personalized contact

Predictions and notifications

Prediction Status: "Delayed"

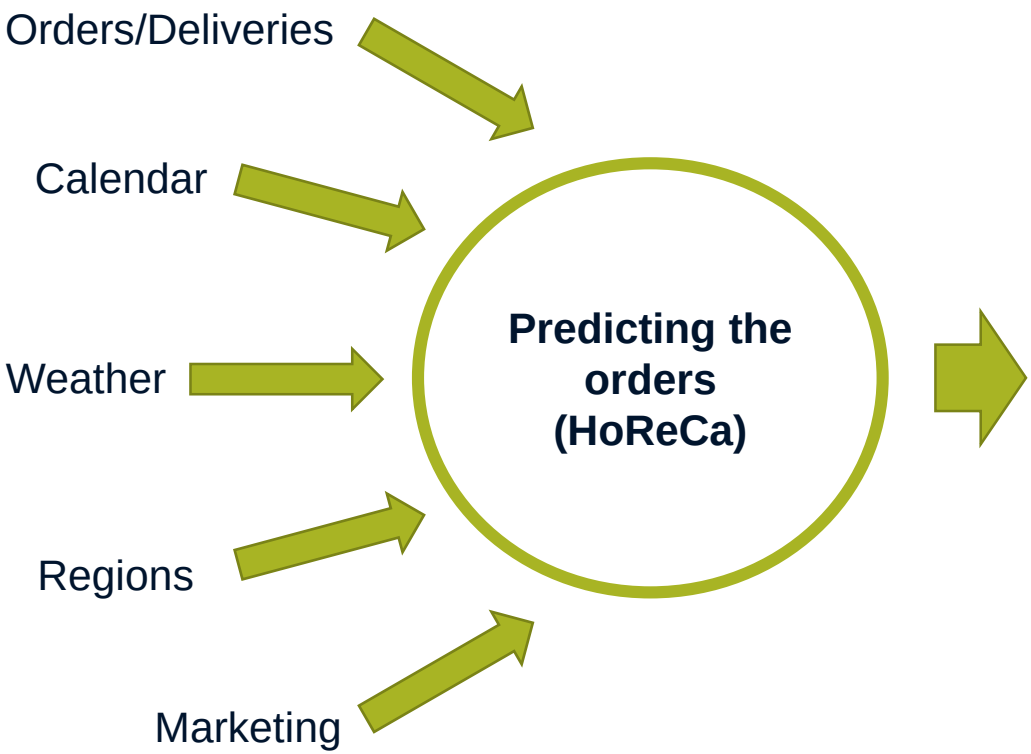
- Avoid churn
- Reduce forgotten donations

PREDICTIVE ANALYTICS

BOAK
BUSINESS OBJECTS ARBEITSKREIS | 2017

MOTIVATION: BUSINESS PERSPECTIVE

- Retailer of drinking products in **Ho**tels-**Re**staurants-**Ca**tering (HoReCa)
- Logistics Department



Logistic Processes

- ❑ Prevent express deliveries
- ❑ Avoid “out-of-stock” situations
- ❑ Group delivery routes
- ❑ Recommend assortment within the customer (inter)
- ❑ Improve assortment across customers (intra)
- ❑ Handle customer stock
- ❑ Identify non-engaged customers and develop strategies

DESCRIPTION OF ANALYZED CUSTOMERS

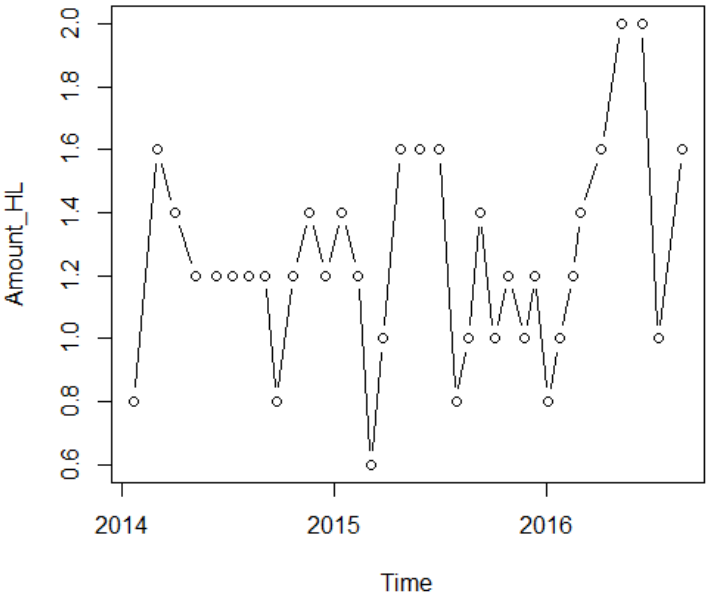
- What type of analysis did we perform?

We analyzed
45’360 time-series

Example:
Customer ID =
77 294 025

- 43 deliveries in total
- 29 different products

Material	Nr. Of Deliveries
10041	35
10099	32
10152	29
10379	25
10476	24
10514	23
10601	19
10409	14
10153	12
10975	12
10606	10
10448	9
11542	8
11543	7
…	…



- Deliveries start from 2014-01-22
- Amount_HL in [0.6, 2] HL
- Time interval between consecutive orders in between 14 to 42 days

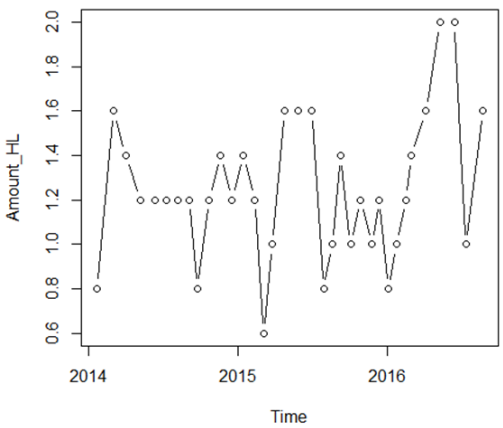
RESULTS

- Engaged and non-engaged customers

  - **Gr.1: 972 out of 1579 Customers (62%)**

    - Patterns exist for at least 1 product of these customers

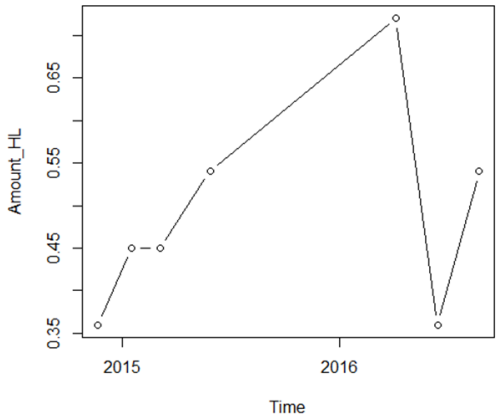
    ✓ ***88%*** of the total deliveries
    ✓ ***83%*** of the total express-deliveries



  - **Gr.2: 607 out of 1579 Customers (38%)**

    - Patterns do not exist (at least based on the history of deliveries)

    - Example: Between all consecutive orders
      Min Day Difference = 49
      Max Day Difference = 313



RESULTS

- Engaged customers (Gr.1): Main Products - Concept

Main Products: at least one of them is ordered almost always

Secondary Products: ordered infrequently

RESULTS

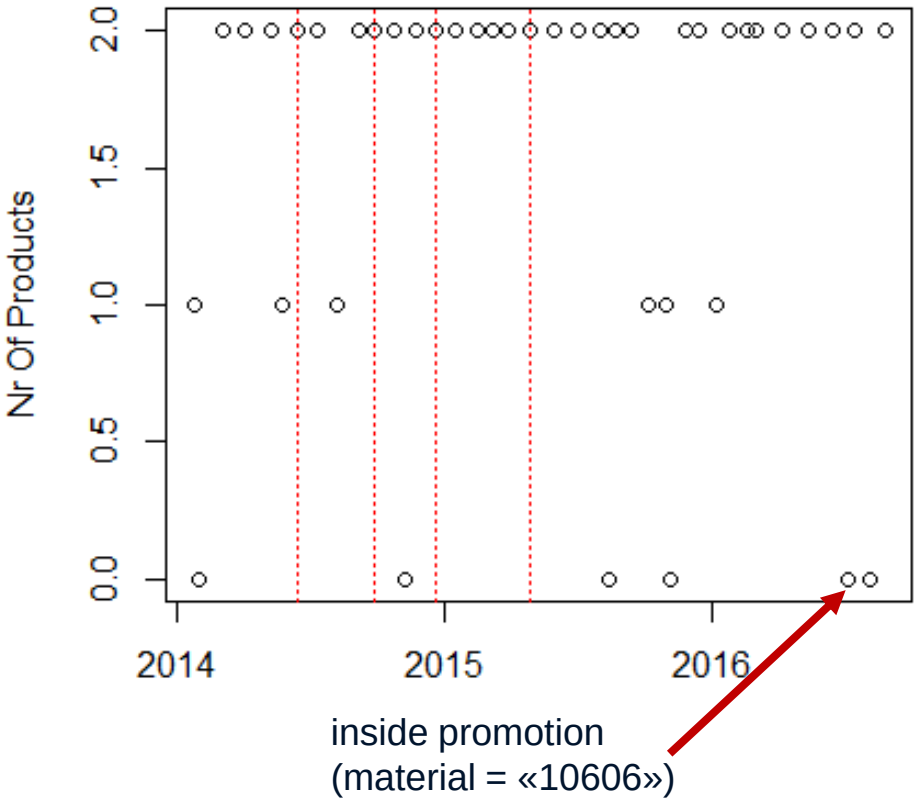
- Engaged customers (Gr.1): Main Products - Example

Example: Customer ID = 77 294 025

«main» products

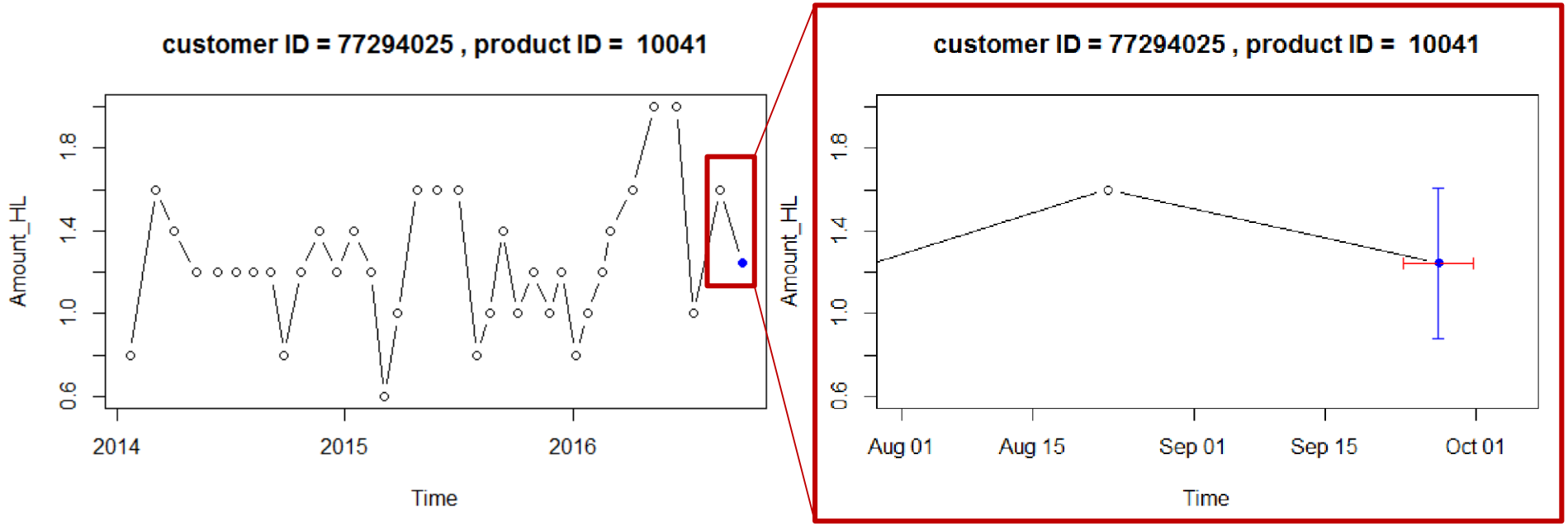
Material	Nr. Of Orders
10041	35
10099	32

- ✓ 37 out of 43 times (86%) at least one of the 2 «main» products was delivered
- ✓ 297 out of 307 product-deliveries (97%) happened in these 37 deliveries
- ✓ 4 out of 4 express-deliveries (100%) happened in these 37 deliveries (red dashed lines)



RESULTS

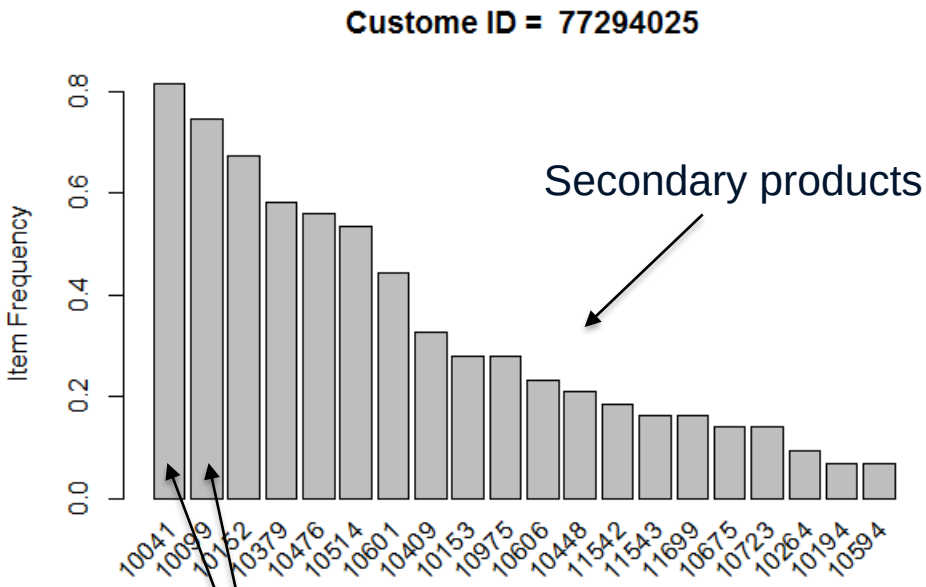
- Engaged customers (Gr.1): Main Products - Predictions



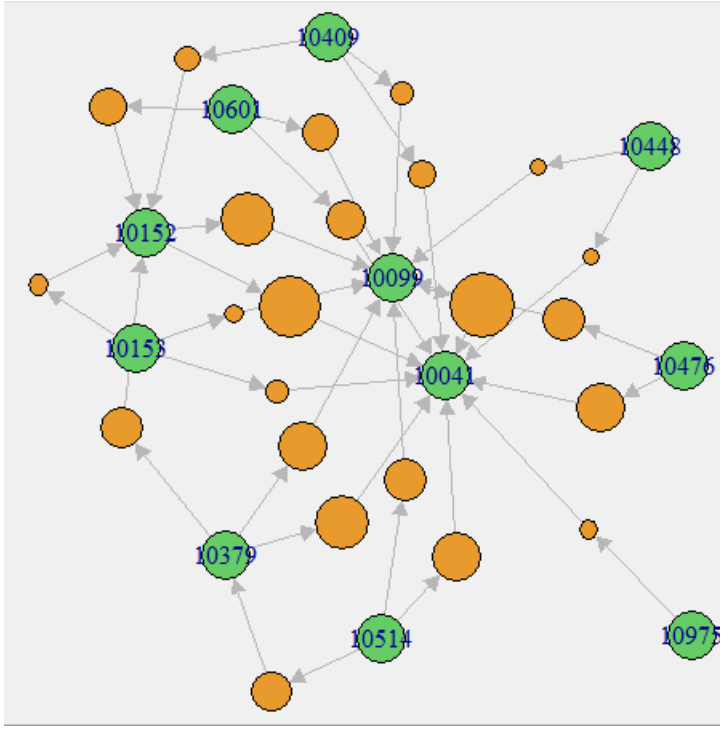
- For every prediction 2 confidence intervals are computed:
  1. Time of delivery (red line)
  2. Amount of product delivered (blue line)

RESULTS

- Engaged customers (Gr.1): Secondary products



Association Graph of products



Can be used for product recommendations

RESULTS

- Engaged customers (Gr.1): Overall Strategy & Benefits

  - ✓ For 62% of the customers (972 out of 1579) we can make reliable predictions
  - ✓ 88% of the all deliveries , and 83% of all the express deliveries are made to the Gr.1 customers
  - ✓ For the Gr.1 customers, deliveries can be predicted based on the «main» product(s)
    - ↳ Majority of total deliveries can be predicted
    - ↳ Majority of express deliveries can be avoided
    - ↳ Change the order-behavior via recommendations for known «secondary» products (reduce single orders of «secondary» products)

GEO-ANALYTICS

BOAK
BUSINESS OBJECTS ARBEITSKREIS | 2017

GEO-DATA IN SWITZERLAND

- BFS as a source of information

Regional Level

State (Kanton)

District (Bezirk)

Community (Gemeinde)

Hectare (100m x 100m)

Features / Attributes

- Age groups
- Housing Information
- Country of origin
- Marital status
- Language
- Work sectors
- Family and kids
- Religion
- Political opinions
- Commuting
- Urbanization
- Educational level
- Income
- etc

GEO-DATA IN SWITZERLAND

- BFS as a source of information

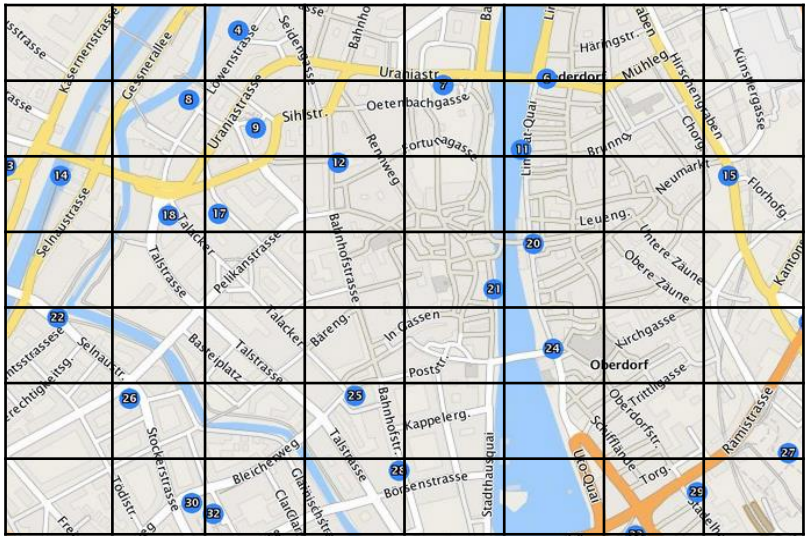
Regional Level

State (Kanton)

District (Bezirk)

Community (Gemeinde)

Hectare (100m x 100m)



Features / Attributes

- Age groups
- Housing Information
- Country of origin
- Marital status
- Language
- Work sectors
- Family and kids
- Religion
- Political opinions
- Commuting
- Urbanization
- Educational level
- Income
- etc

BOAK BUSINESS OBJECTS ARBEITSKREIS | 2017

GEO-DATA IN SWITZERLAND

- BFS as a source of information

Regional Level

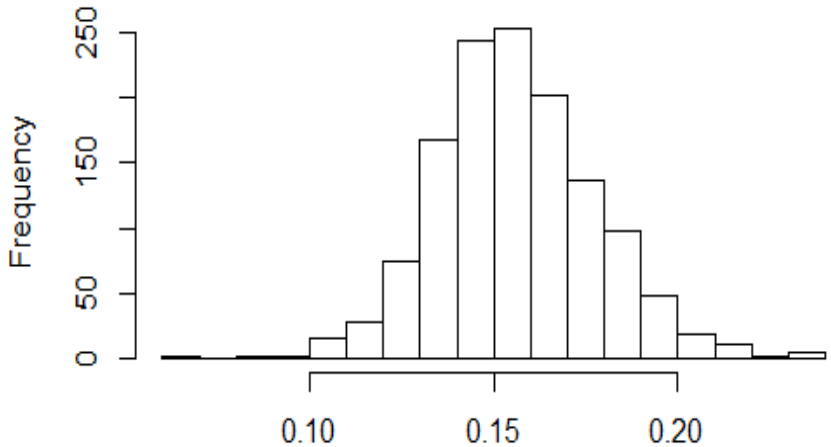
State (Kanton)

District (Bezirk)

Community (Gemeinde)

Hectare (100m x 100m)

Quota of persons of age 0-14 across political communities in CH



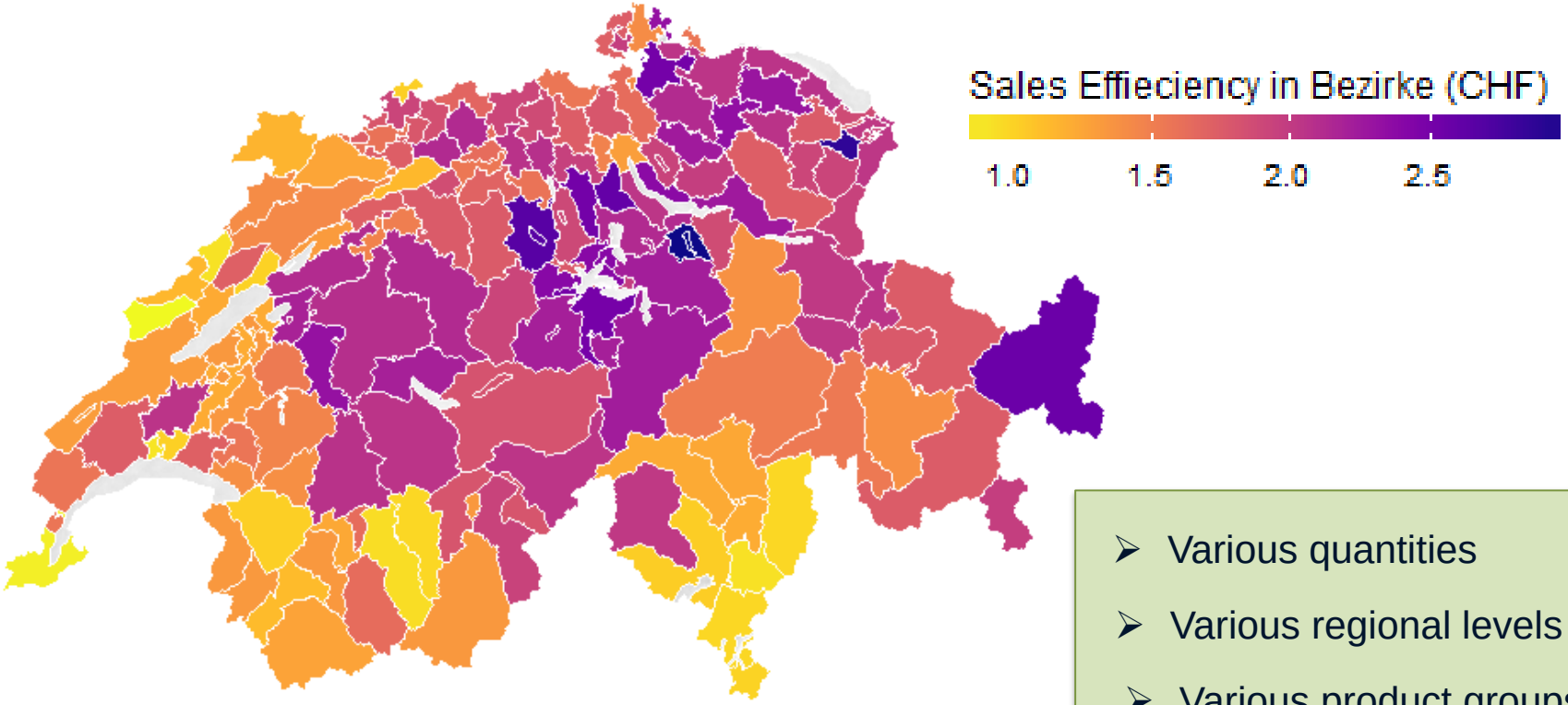
Diversity is the key to informativeness!

Features / Attributes

- **Age groups**
  - **0-4**
  - **5-9**
  - **10-14**
  - 15-19
  - 20-24
  - …
  - 89 and more
- Housing Information
- Country of origin
- Marital status
- Language
- Work sectors
- etc

GEO-ANALYTICS: RESULTS

- Quantification sales efficiency at the regional level

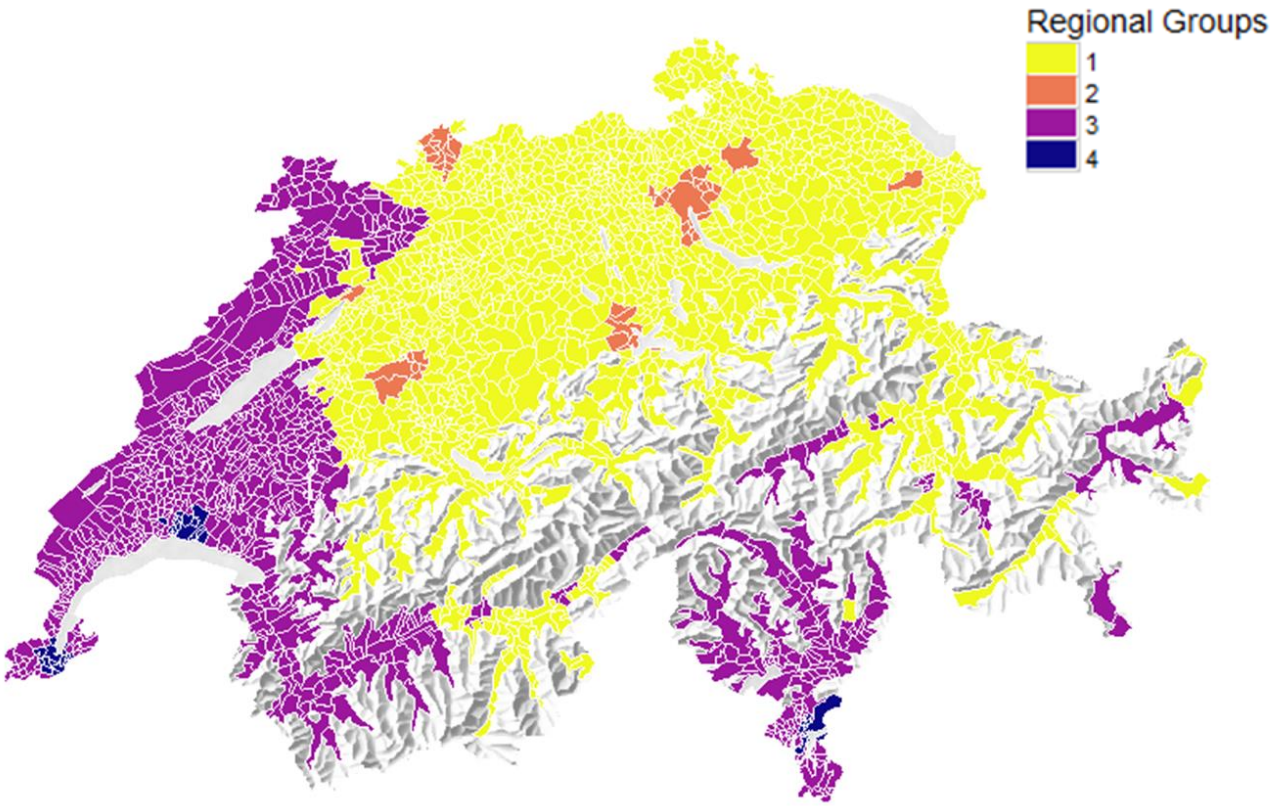


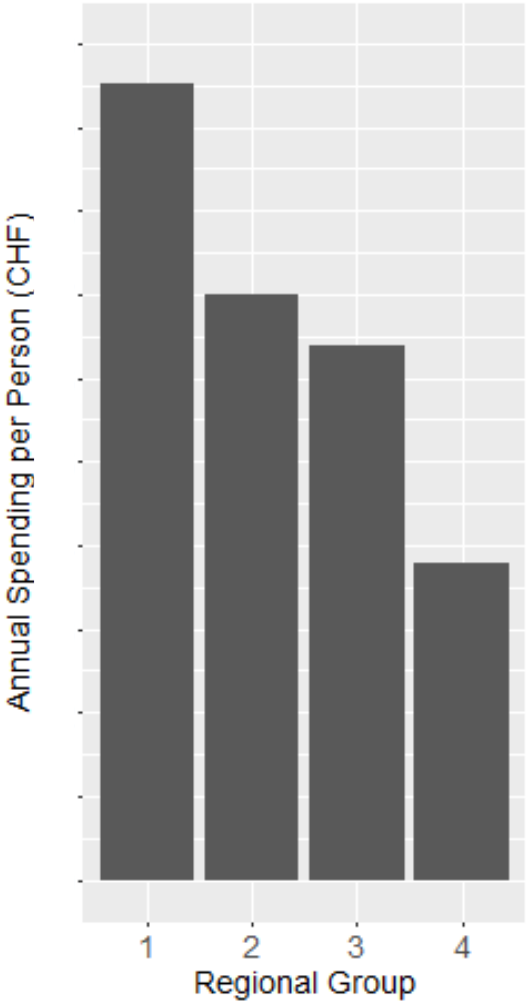
Annual Revenue Efficiency (= Annual Revenue / Population) of a product-group for all Bezirke in 2016

- Various quantities
- Various regional levels
- Various product groups

GEO-ANALYTICS: RESULTS

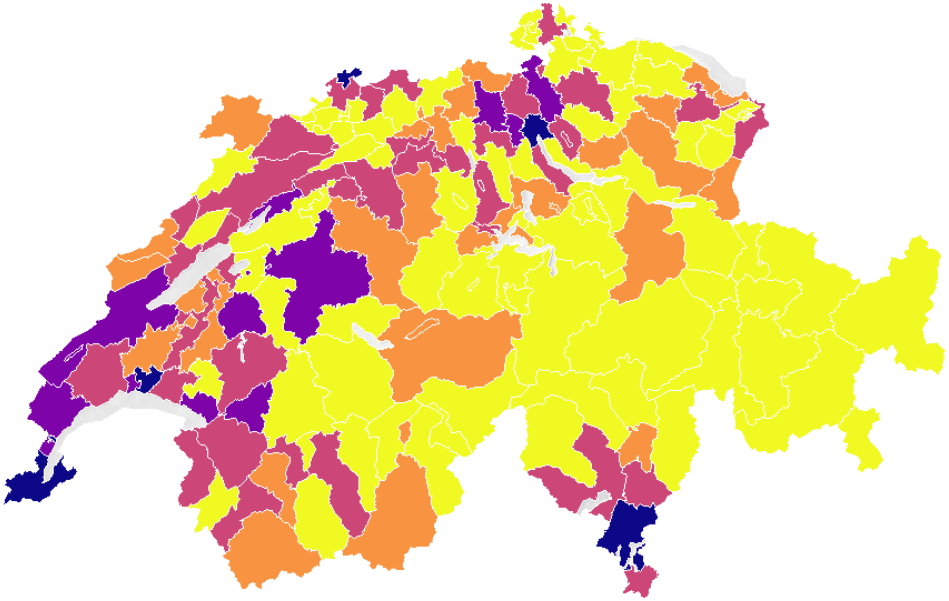
- Regional grouping





GEO-ANALYTICS: RESULTS

- Annual revenue potential



Revenue Potential (CHF)

small

high

Annual revenue potential of a product-group at the Bezirk-level

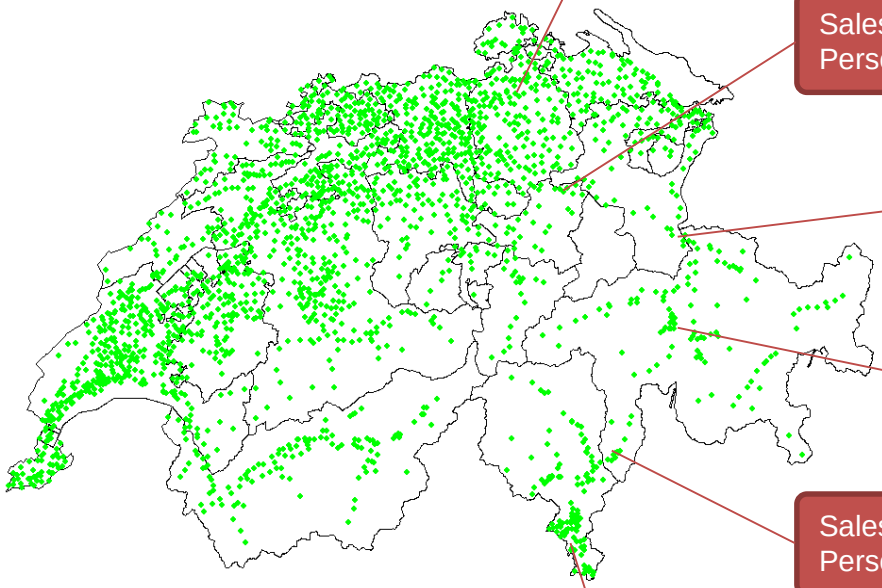
- Which regions should be targeted first?

KANTON_NAME	Revenue Potential (CHF)	Cumulative Percentage (%)
Vaud	5.69*X	14.2
Genève	5.58*X	28.1
Zürich	5.39*X	41.6
Ticino	3.47*X	50.3
…	…	…

Annual revenue potential of a product-group at the Kanton level

GEO-ANALYTICS: RESULTS

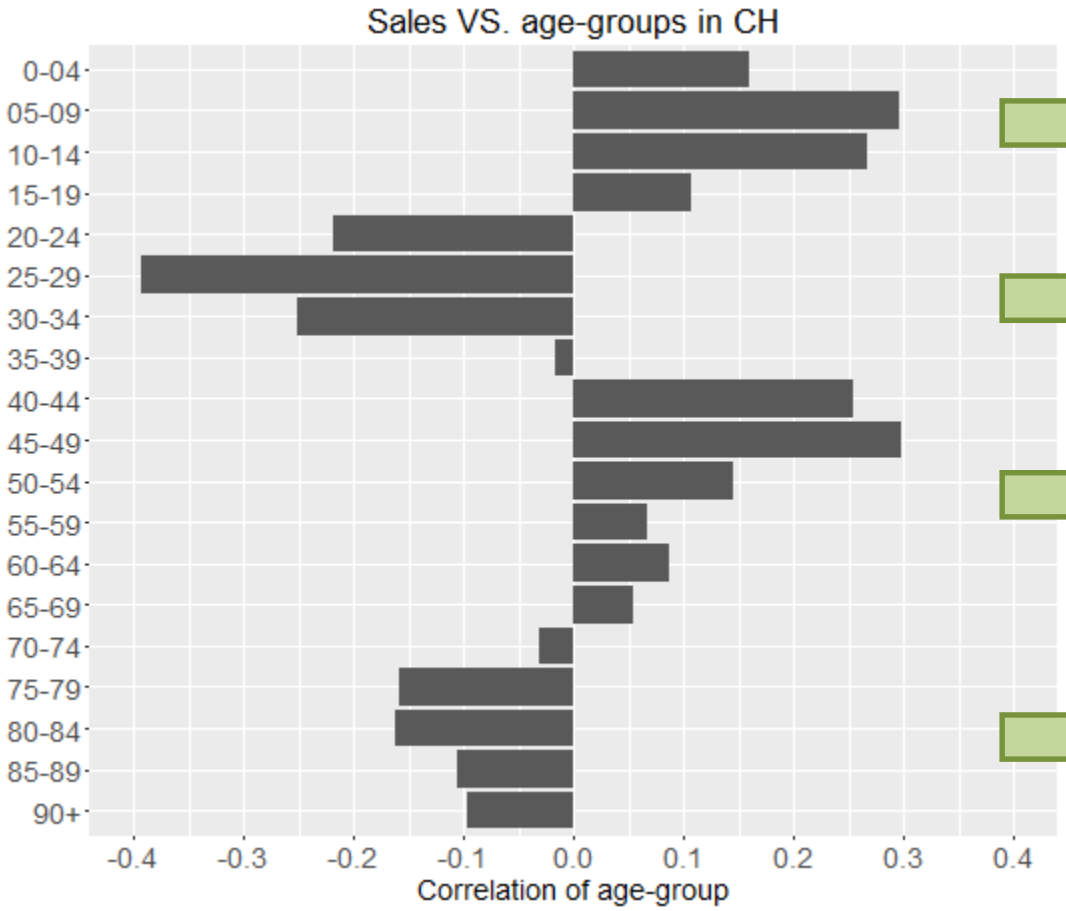
- Analysis for age-groups: exploit regional diversity



Sales Per Person	9.5 CHF	Quota of Age 45-49	10.2%
Sales Per Person	8.1 CHF	Quota of Age 45-49	9.5%
Sales Per Person	7.2 CHF	Quota of Age 45-49	8.1%
Sales Per Person	5.8 CHF	Quota of Age 45-49	6.5%
Sales Per Person	6.7 CHF	Quota of Age 45-49	6.3%
Sales Per Person	5.1 CHF	Quota of Age 45-49	7.2%

GEO-ANALYTICS: RESULTS

- Sales Efficiency for age-groups



Families with kids

Young people

middle-aged people

Old people

OTHERS

BOAK
BUSINESS OBJECTS ARBEITSKREIS | 2017

OTHERS

- Requirements Engineering with Data Science
- Fraud detection (mobilezone)
- Predicting defaulting customers (card center of major swiss bank)
- Optimization of campaigns (Young Swiss)
- Infer customer segments behind products (B2B client)

Wir freuen uns auf angeregte Gespräche mit Ihnen

- Dr. Sotiris Dimopoulos
  Senior Data Science Consultant

Follow us: @itlogixag
http://blog.it-logix.ch