Frankfurter Allgemeine

Identify and maintain subscribers likely to leave

The F.A.Z. Churn Prevention Model

WAN-IFRA Data Science Day 2022 – Paris, October 21, 2022

Agenda

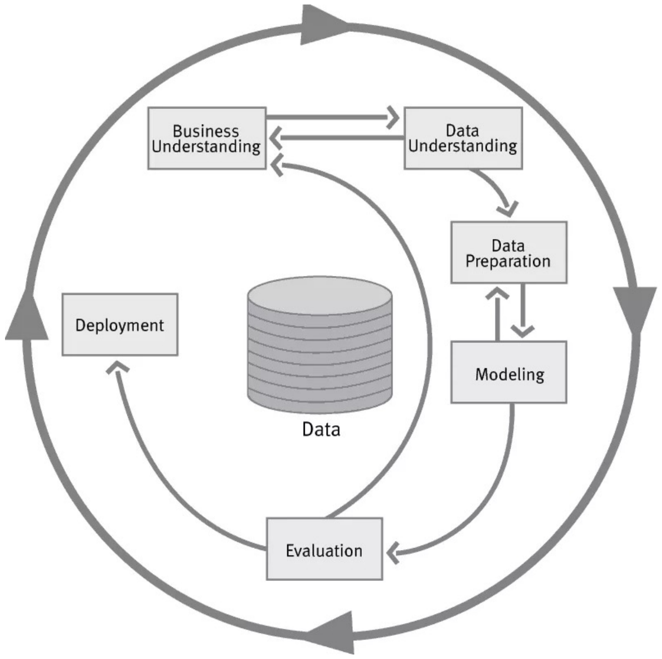
1. Churn Prediction
   - General model overview
   - Technical setting
   - Output files
   - Variable overview
2. Churn Prevention
   - Model validation
   - Market test results

1. Churn Prediction

The churn prediction model identifies customers at high churn risk.

General model overview

- Churn scores are calculated for **each print** and **digital newspaper customer**
- Fully-automatized **calculation** incl.
  - raw data integration, transformation and aggregation
  - identification of **significant parameter** subset and interpretation help
  - selection of best suited data transformation steps
  - selection of **best** suited **model**
  - control instances and log file documentation
  - integration into **marketing systems**



https://datasolut.com/wp-content/uploads/2019/11/CRISP-DM.png.webp

How are the churn scores calculated?

Transforming multiple data sources to one churn score table

Customer data

Customer service contact data

Payment data

Customer benefits portal data

Web analytics data

integration, transformation and aggregation

parameter estimation with best suited model

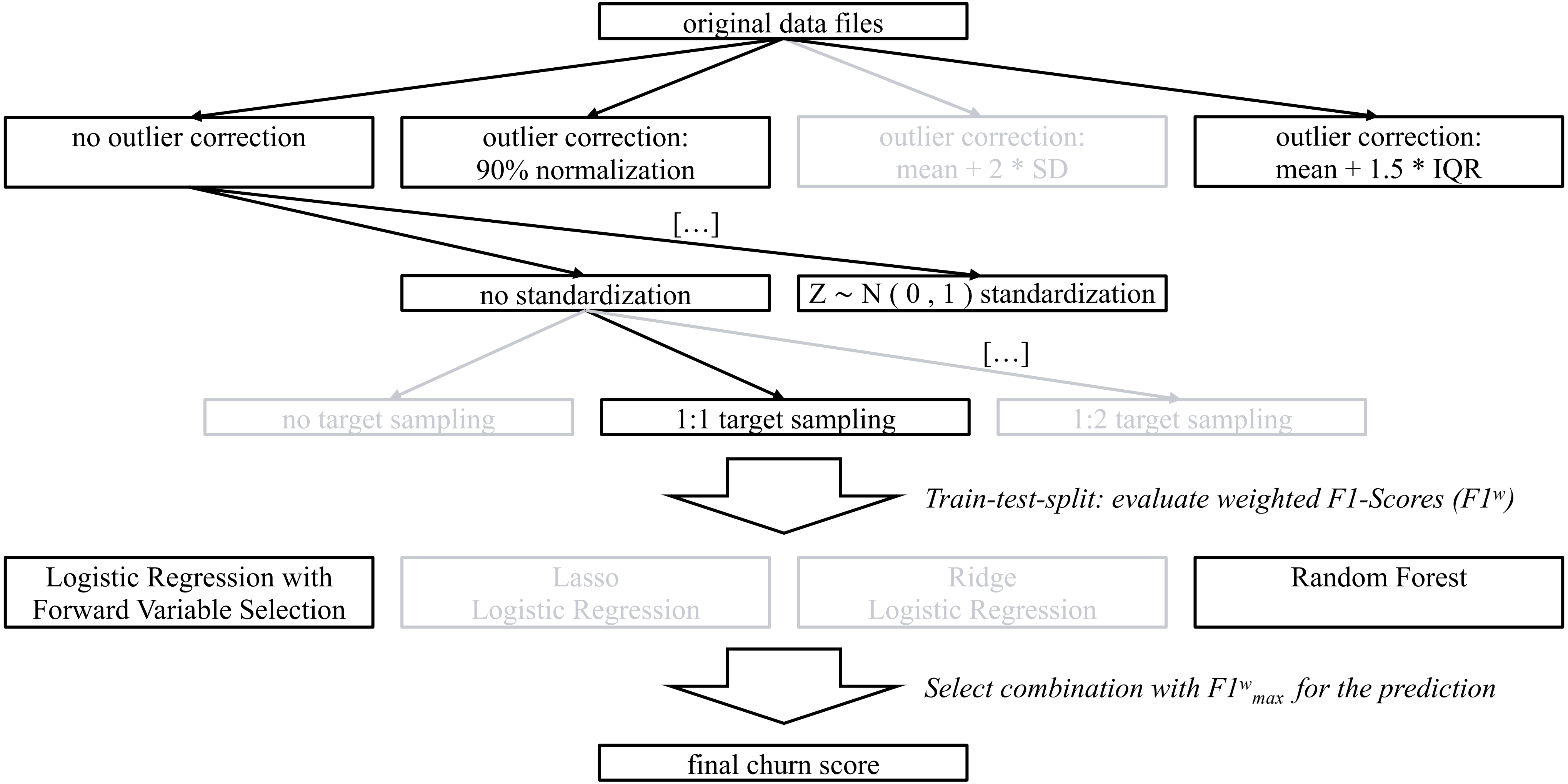
identification and interpretation of significant variables

customer-individual score calculation

CSV file with churn scores

Customer ID	Order ID	Churn Score
56xxxxxxxx	19xxxxx	91.5%
59xxxxxxxx	16xxxxx	69.2%
56xxxxxxxx	18xxxxx	10.9%

How does the algorithm identify the best suited model and data?



Parameter interpretation and calculation log files allow to supervise the automatized calculation output.

Output and log files

– Output files include model summaries and parameter interpretation assistance

	variable interpretation
age	Every additional year increases the probability to churn by (c.p.) *(percentage)* % on average.
complaint_phone	A by 1% increased share of phone call requests of the total number of customer service contacts decreases the probability to churn by *(percentage)* % on average.
marketing_channel	Customers with a direct website order have a (c.p.) (percentage) lower churn probability than customers with any other marketing channel order.
optin_phone	Customers with a phone opt-in consent have a (c.p.) (percentage) higher churn probability than customers without the phone opt-in.
payment_method	Customers who pay by credit card have a (c.p.) (percentage) lower churn probability than customers with other payment methods.

– Log files track warnings in the calculation progress to ensure control

	warnings
Unbalanced_Factors	Strongly unbalanced factor variable(s) *(variable names)*. The variable(s) will be removed from the dataset.
Correlation	Highly correlated variable(s) *(variable names)*. The variable(s) *(variable names)* will be removed from the dataset.
Warnings_AIC.Base	
Warnings_AIC.Base (1:1 sample)	glm.fit: fitted probabilities numerically 0 or 1 occurred
Warnings_VIF.Base	Multicollinearity problem with variable(s) *(variable names)*. The variable(s) will be removed for parameter estimation.

More than 50 variables have been included in the churn model.

Variable overview including significant factors, F.A.Z. print, 2021

customer data

- age
- gender
- country
- region
- **student (**)**

- **current subscription duration (**)**
- **subscr. duration since first contact (**)**
- **# current subscriptions (**)**
- **# subscriptions in customer history (**)**
- # free samples in customer history
- **# trial subs in customer history (**)**
- **conversion from free sample / trial sub (**)**
- cross-usage (F.A.Z. vs. F.A.S. print and digital)

- **opt-in consent (*)**
- **marketing channel (*-**)**
- incentive received

payment data

- **price (**)**
- **payment frequency (**)**
- **payment method (**)**
- *payment credibility*

customer benefits portal

- # lottery participations
- # lottery wins

customer service contact data

- **# of contacts/complaints (**)**
- **complaint type (*-**)**
- **complaint channel (*)**

(**) p < 0.01; (*) p < 0.05

2. Churn Prevention

The model validation provided very good results for both churn- and non-churn customers on test data.

Accuracy, precision and recall results, 2021

		Actual target	
		no churn	churn
Predicted target	no churn	**99.6%**	12.3%
	churn	0.4%	**87.7%**

Accuracy: 97.3%

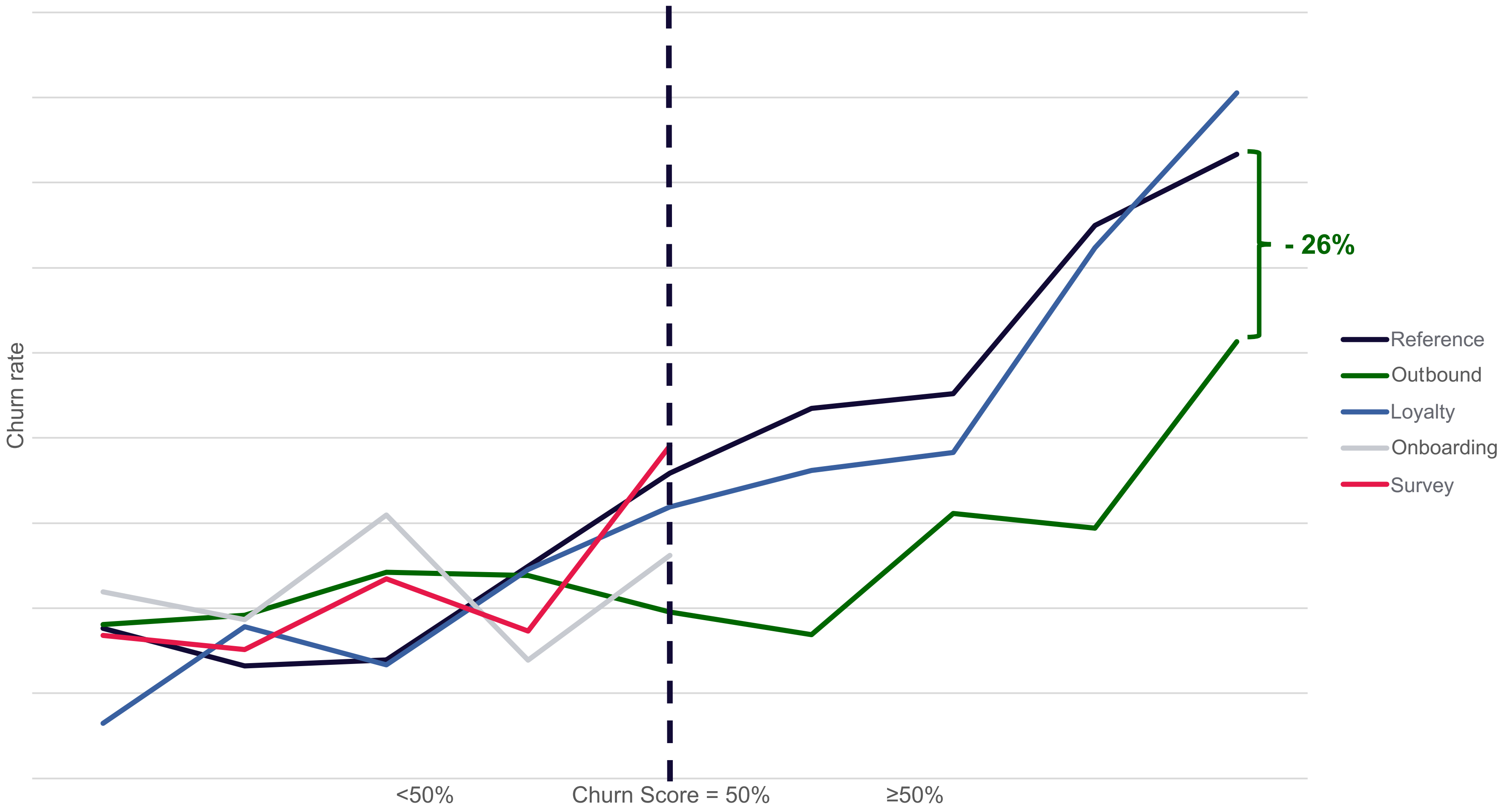
Precision: 97.9% (churn), 97.1% (no churn)

Recall: 87.7% (churn), 99.6% (no churn)

Especially outbound telephony helped to prevent churn of customers at high risk.

F.A.Z. print: Churn rate comparison of several measures

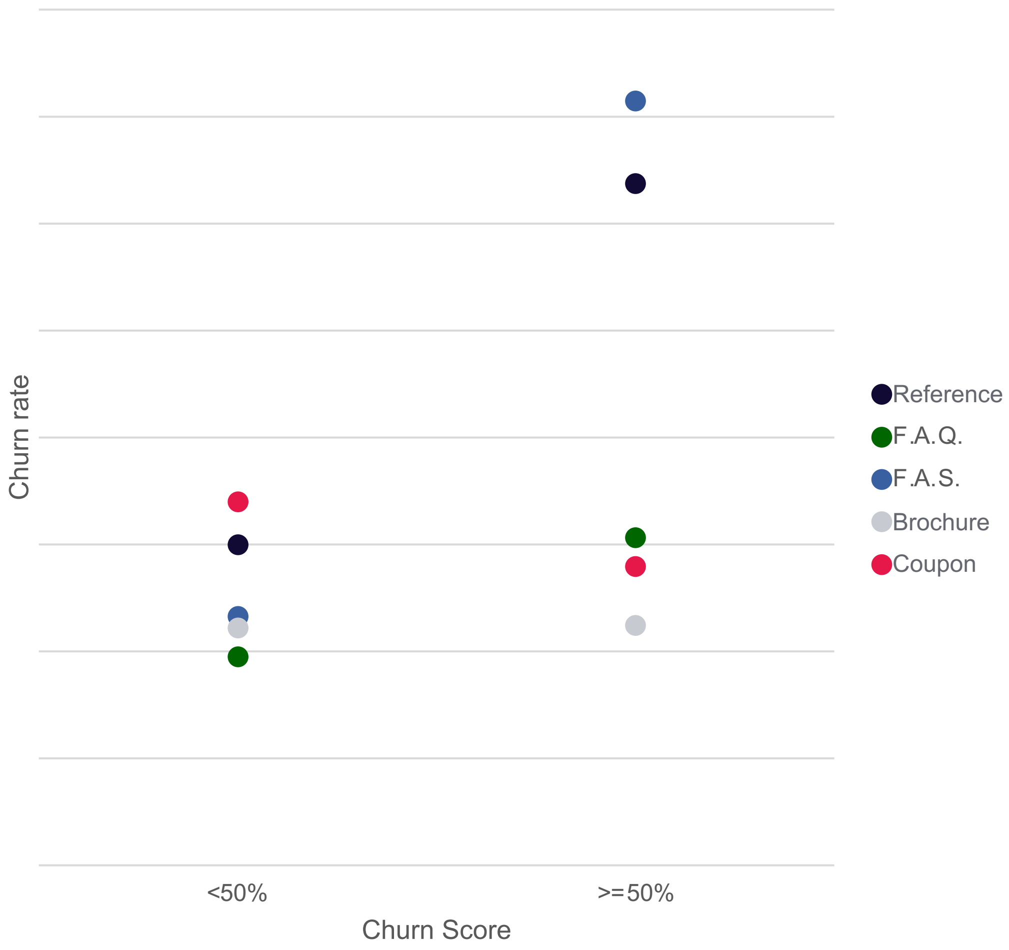
Print



The editorial department brochure and the free annual subscription of „Quarterly“ significantly lowered customer churn.

F.A.Z. print: comparison of incentives by churn rate

Print



Contact

Frankfurter Allgemeine Zeitung GmbH

Fabian Wörz
Senior Data Scientist
Product + Sales
Phone +49 69 7591 1830
E-Mail f.woerz@faz.de
www.faz.net

Freiheit beginnt im Kopf.

Thank you very much.

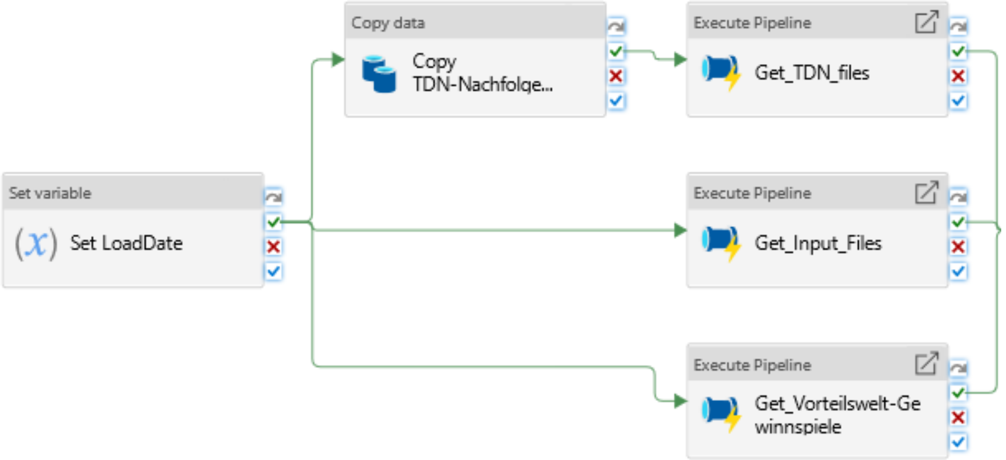
Appendix

1. Churn Prediction

Churn scores are integrated into the Salesforce Marketing Cloud as an underlying factor for marketing campaigns.

Data pipeline

raw data and parameter import



More than 50 variables have been included in the churn model.

Variable overview including significant factors, F.A.Z. digital, 2021

customer data

- **age (**)**
- gender
- country
- **region (*)**
- student

- **current subscription duration (**)**
- **subscr. duration since first contact (*)**
- # current subscriptions
- # subscriptions in customer history
- # free samples in customer history
- **# trial subs in customer history (**)**
- conversion from free sample / trial sub
- cross-usage (F.A.Z. vs. F.A.S. print and digital)

- **opt-in consent (*)**
- **marketing channel (*)**
- incentive received

payment data

- price
- payment frequency
- **payment method (*)**
- *payment credibility*

customer benefits portal

- # lottery participations
- # lottery wins

customer service contact data

- # of contacts/complaints
- **complaint type (**)**
- complaint channel

web analytics data

- Ø page views per day
    - Δ last two weeks
    - Δ last four weeks
- **Ø days with ≥1 visit(s) (**)**
    - **Δ last two weeks (*)**
    - Δ last four weeks
- Ø minutes per visit
    - Δ last two weeks
    - Δ last four weeks

- product usage (F.A.Z. vs. F.A.S.)
- multimedia usage
- max. # visits per day

- channel type
- country of access
- **device (**)**
- operating system

(**) $p < 0.01$; (*) $p < 0.05$

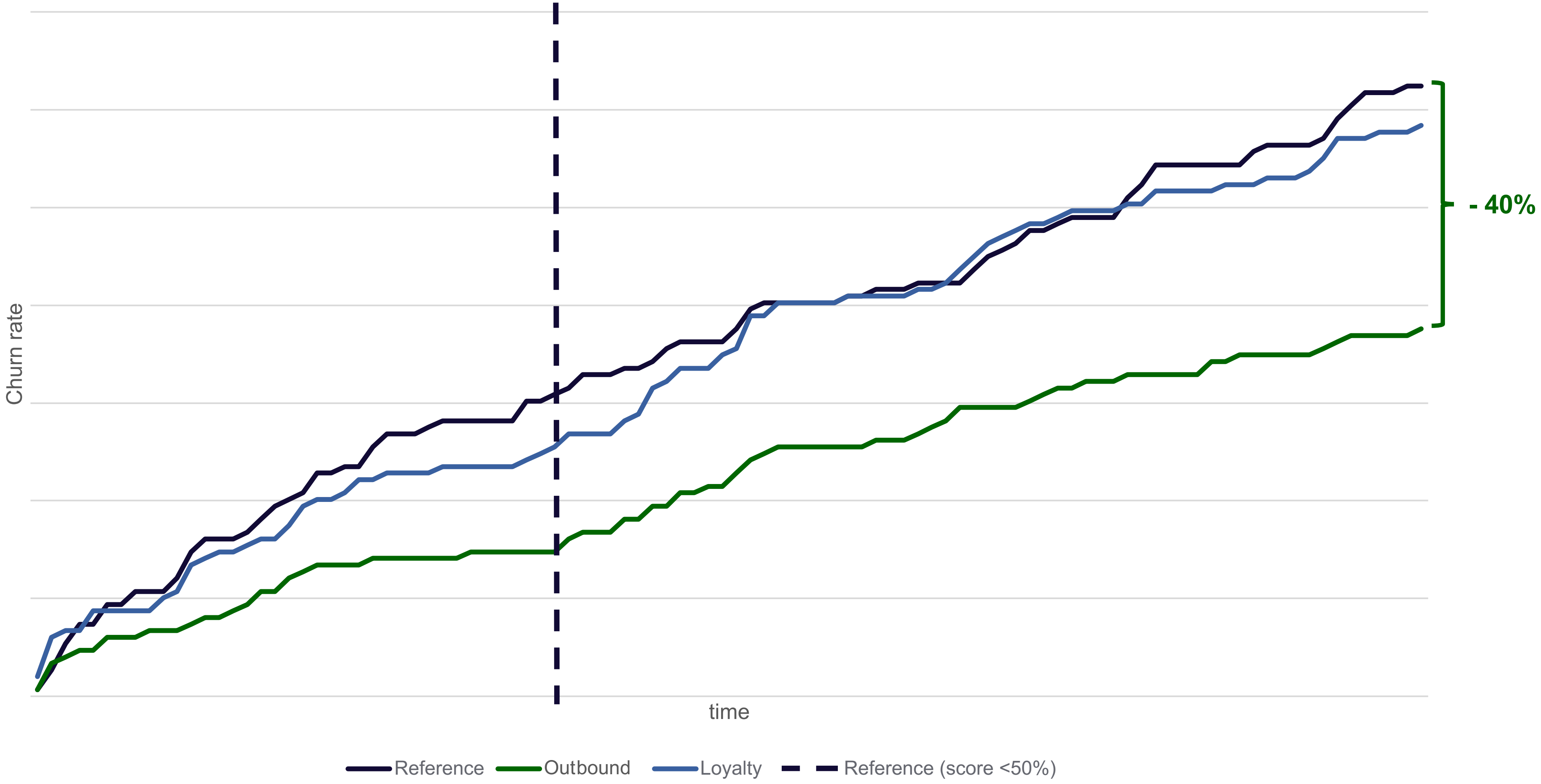
2. Churn Prevention

Especially outbound telephony helped to prevent churn of customers at high risk.

F.A.Z. print (churn probability ≥ 50%): churn over time

Print

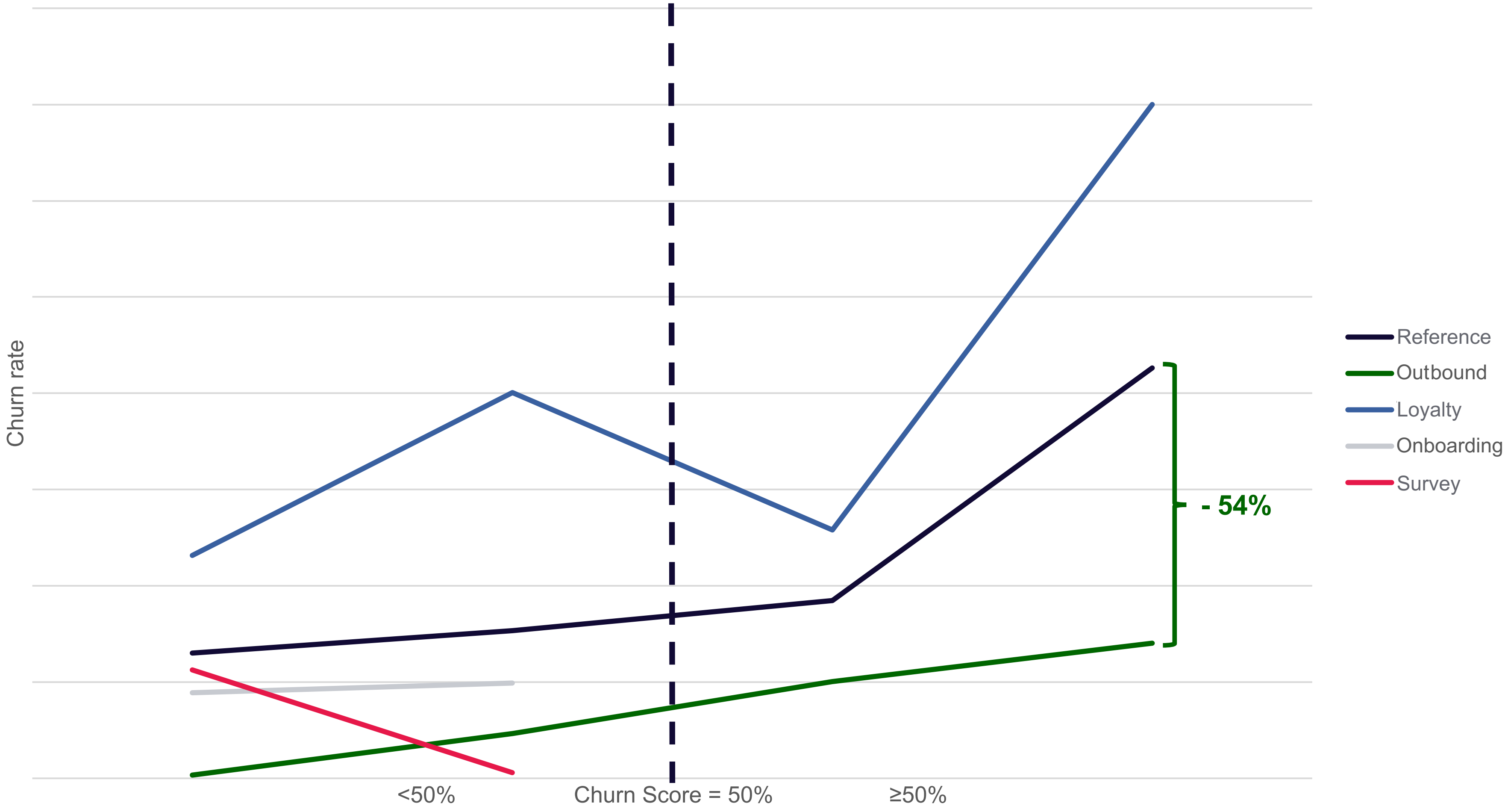
≥ 50%



Especially outbound telephony helped to prevent churn of customers at high risk.

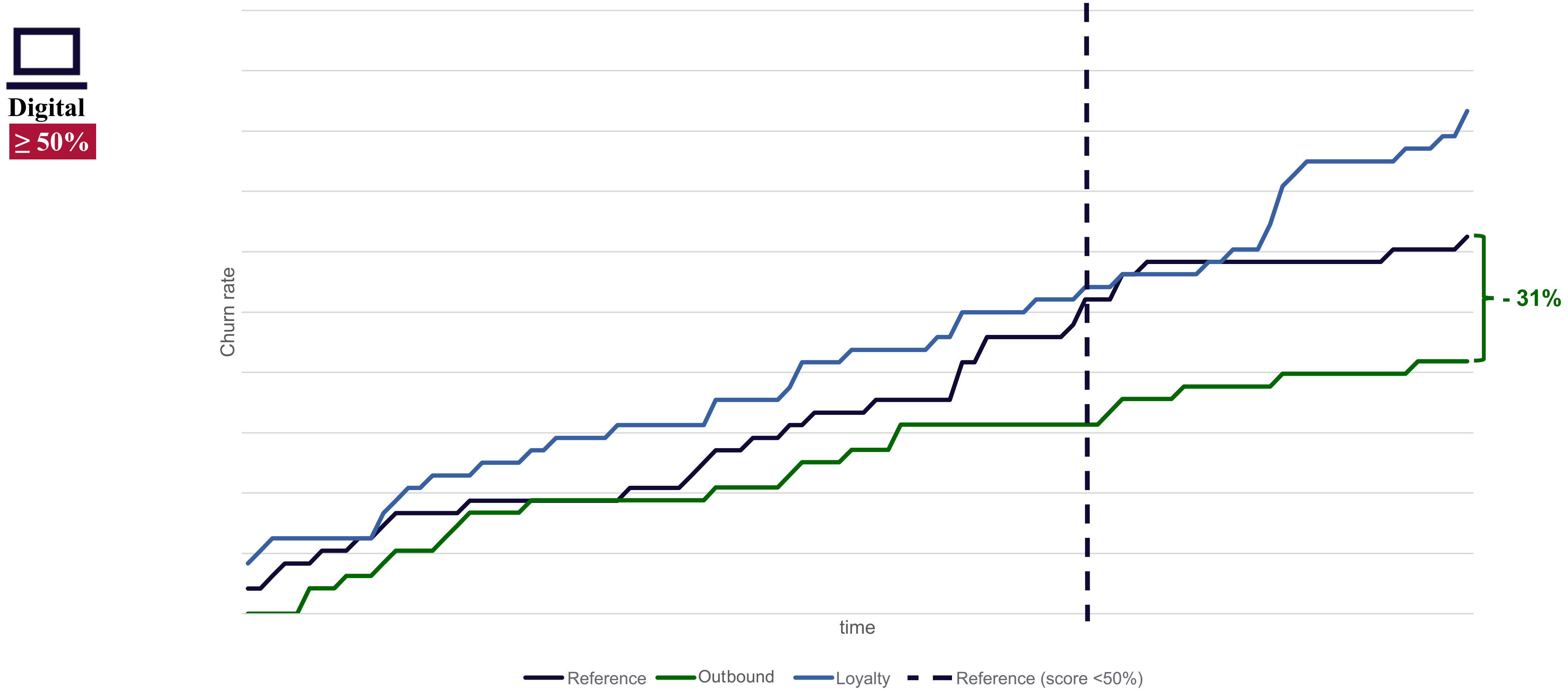
F.A.Z. digital: Churn rate comparison of several measures

Digital



Especially outbound telephony helped to prevent churn of customers at high risk.

F.A.Z. digital (churn probability ≥ 50%): churn over time



The editorial department brochure and the free annual subscription of „Quarterly“ significantly lowered customer churn.

F.A.Z. digital: comparison of incentives by churn rate

Digital