



# Credit scoring and Fraud detection in retail

## The story of 10 years of risk analytics at Unigro

Geert Verstraeten  
Python Predictions

@pythongeert

# Unigro



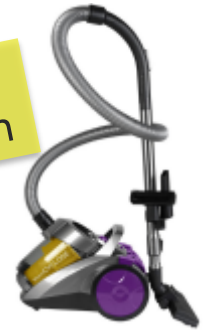
8 EUR x  
15 month

furniture



4 EUR x  
2 month

linen



6 EUR x  
13 month

home  
appliances



40 EUR x  
20 month

hifi & multimedia



7 EUR x  
5 month

beauty



9 EUR x  
18 month

leisure

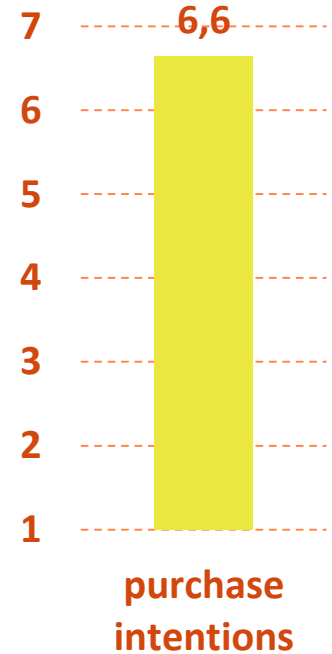
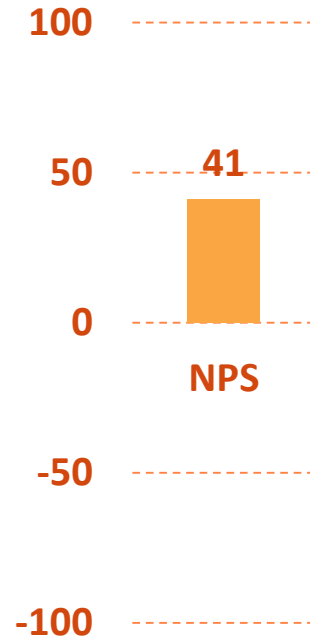
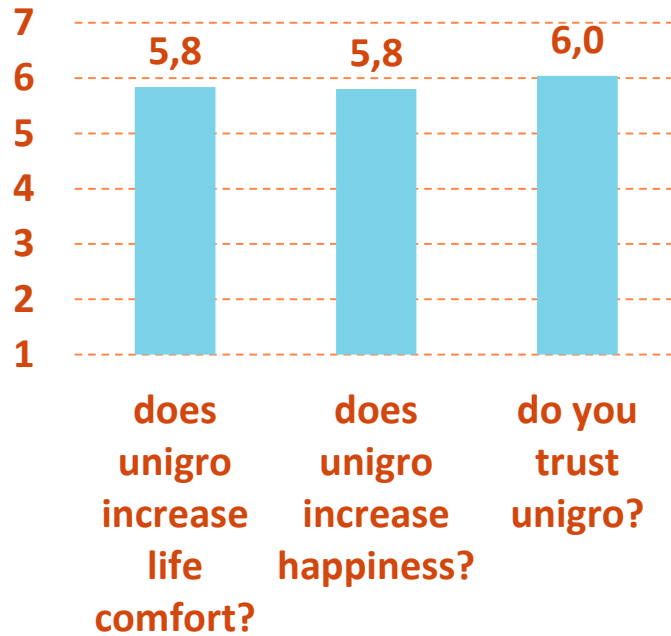


# Unigro - Mission

The brand contributes to making  
the **lives of its customers more comfortable**  
by facilitating access to  
a large number of products and services,  
offering **purchases on credit, granted responsibly**




# Unigro – Mission Execution



# Unigro - Facts

- Since 1948

• Structure:  <  < *otto group*

- Figures:

- 220 employees

- 8000 products

- 205 000 active clients

- 300 000 orders / year

- 450 000 articles sold / year

- 42 Mio EUR revenue / year

- 40% of revenues online





## Challenge

How to grow without accepting more risk?  
Marketing vs Risk = two sides of the coin

## Solution

Credit scoring!



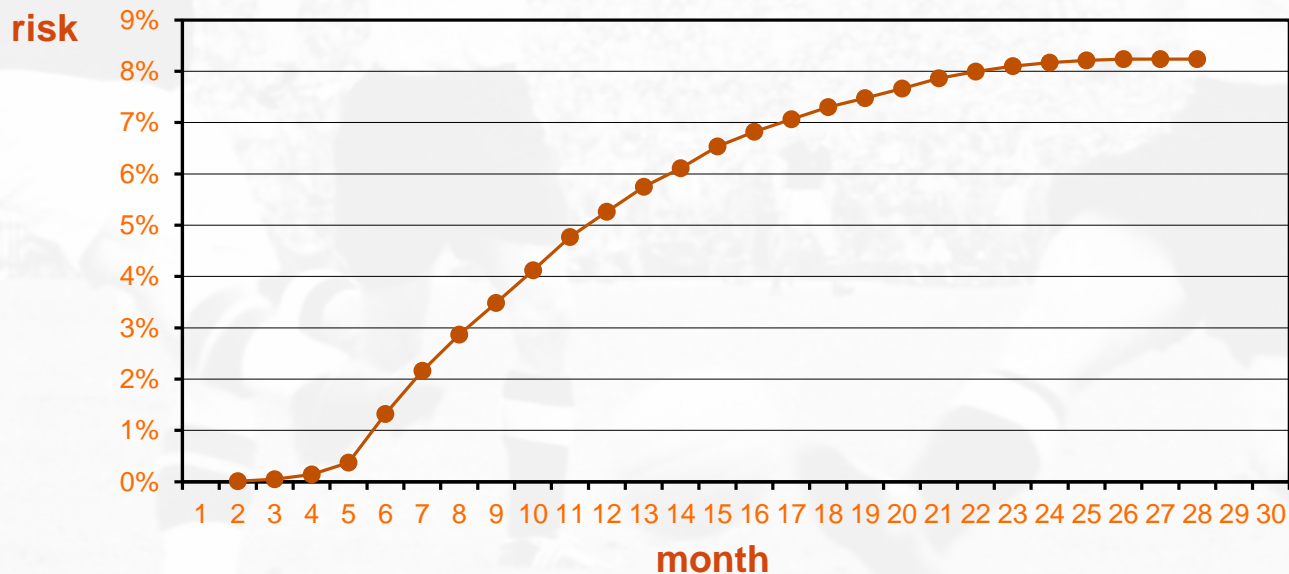
*“Batistuta gets most of  
his goals with the ball”  
Ian St John*

*“The ball is an essential  
part of the game”  
Johan Cruijff*



# History

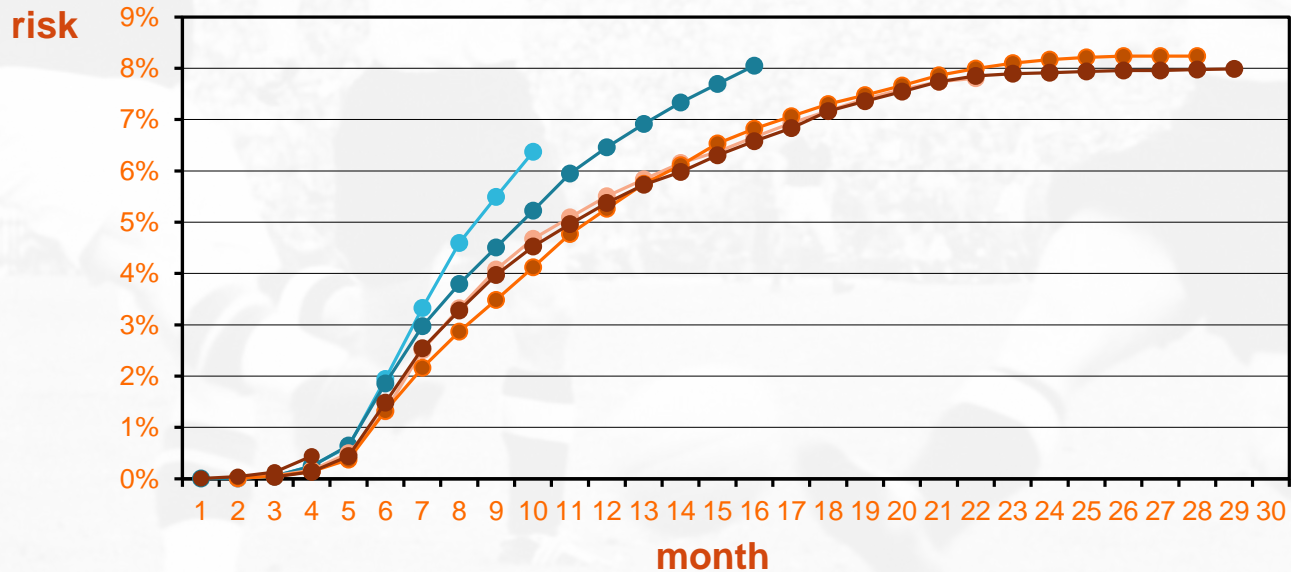
- Existing score performed by Cofidis (groupe 3SI)





# History

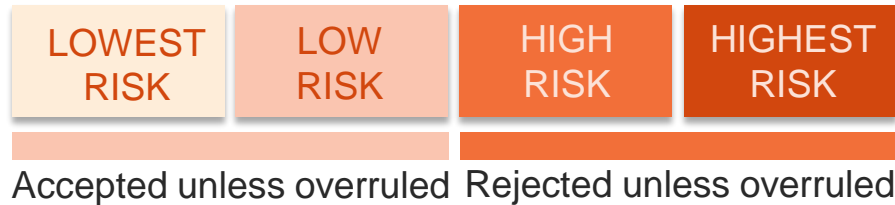
- Existing score performed by Cofidis (groupe 3SI)



# Credit Scoring at Unigro

## Decisions

- Scoring orders, not clients
- Third reminder
- Split modeling
- Interaction with manual process:



# Credit Scoring at Unigro

## Predictors

- Socio-demographic
- Occupation
- Financial
- Relationship with Unigro
- Default history
- Order info



1970

---

TELSTAR





2004 2005 2006 2007 2008

# TELSTAR

2009 2010 2011 2012 2013

- Student project Ghent University
  - Understand Unigro (goal, business processes, data)



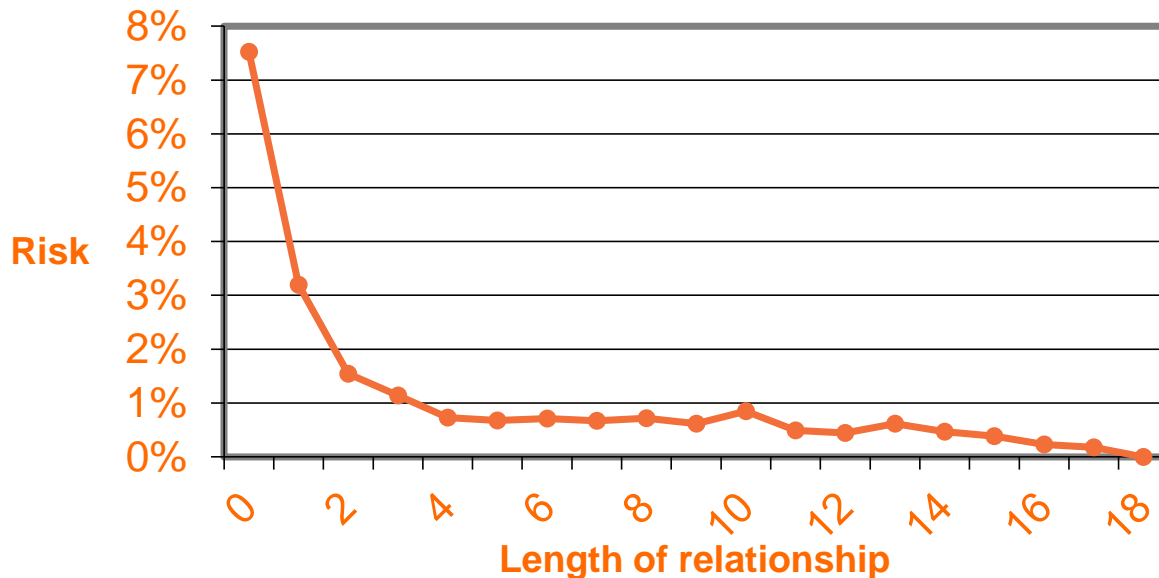


2004 2005 2006 2007 2008

# TELSTAR

2009 2010 2011 2012 2013

- Data understanding





2004 2005 2006 2007 2008

# TELSTAR

2009 2010 2011 2012 2013

- Student project Ghent University
  - Understand Unigro (goal, business processes, data)
  - Construct Basetable (150 variables)
  - Model Building & Validation
    - Logistic regression without feature selection
  - Calculation of quality, parameters & cutoffs



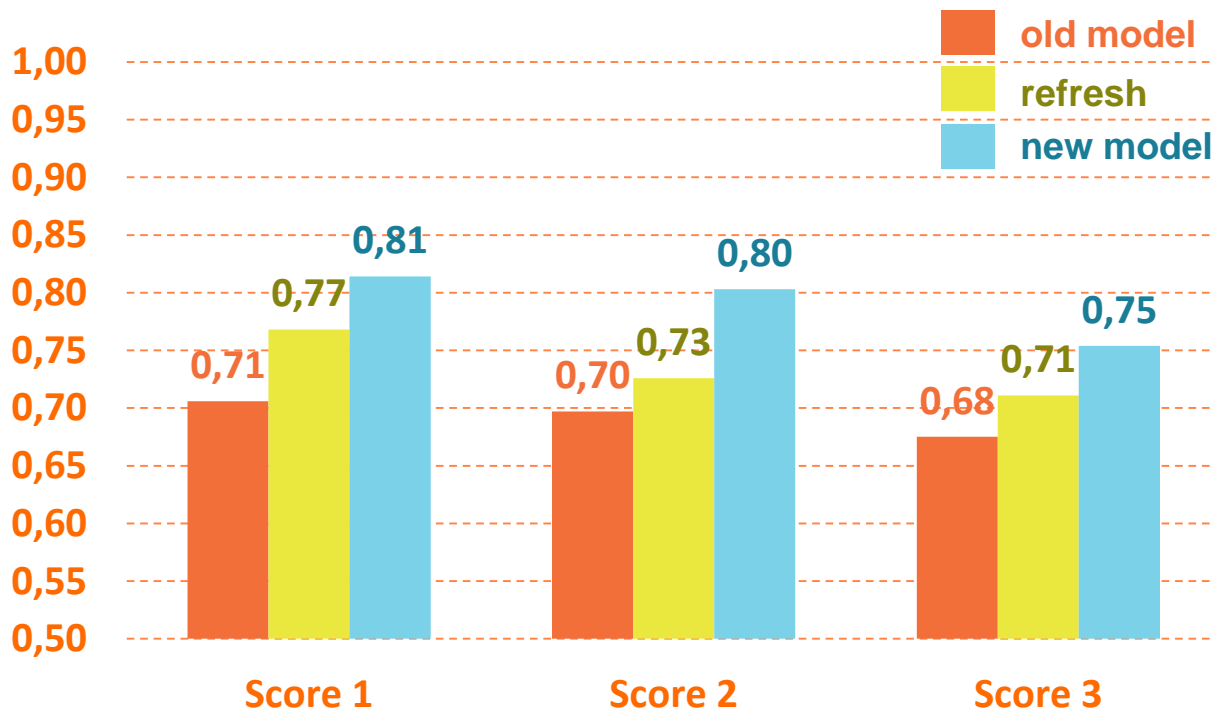


2004 2005 2006 2007 2008

# TELSTAR

2009 2010 2011 2012 2013

AUC  
(predictive  
Performance)





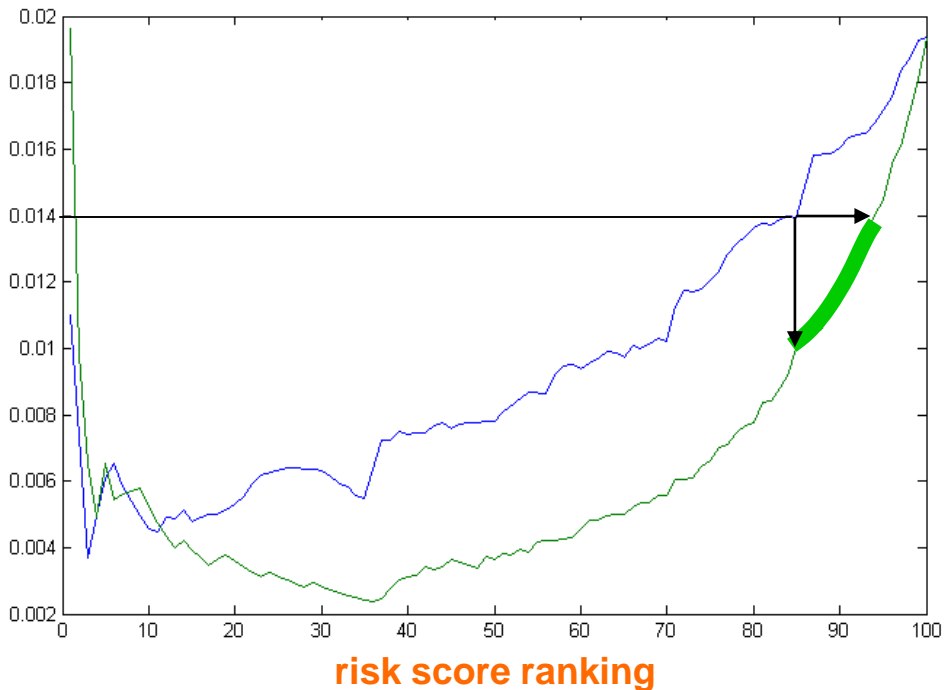


2004 2005 2006 2007 2008

# TELSTAR

2009 2010 2011 2012 2013

**cumulative  
risk**



**old model**  
**new model**



1978

---

TANGO



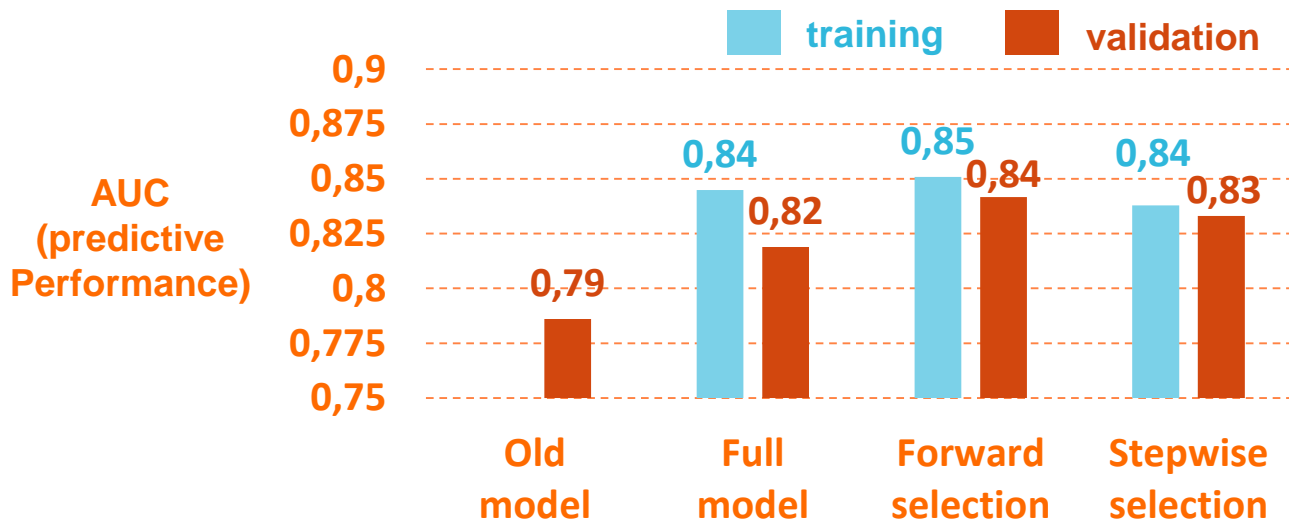


2004 2005 2006 2007 2008

# TANGO

2009 2010 2011 2012 2013

- Variable selection





# TANGO



- Sign violations: parameters

Variable	Model 1	Model 2	Model 3
Reminder Type 1	0,02	0,02	0,06
Reminder Type 2	-0,09	0,01	-0,08
Reminder Type 3	0,54	0,00	0,56



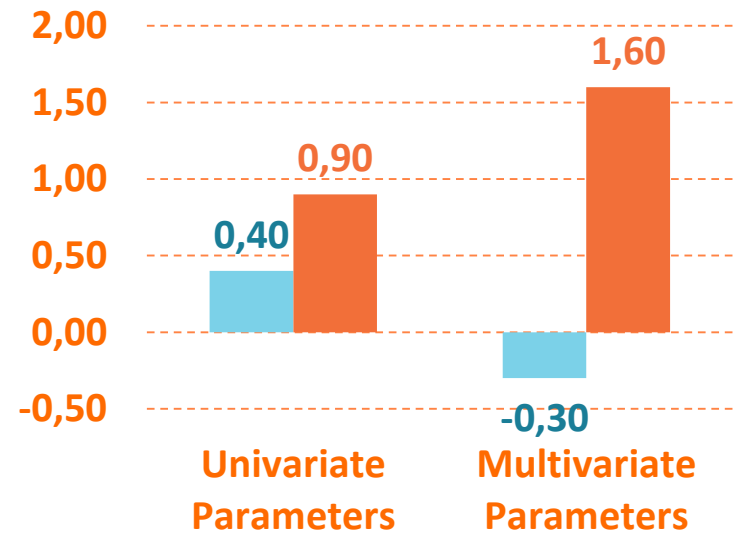


# TANGO



- Sign violations (example)

Bad Payment	Reminder Type 1	Reminder Type 2
1	10	6
1	5	2
0	8	3
0	1	1
0	0	0

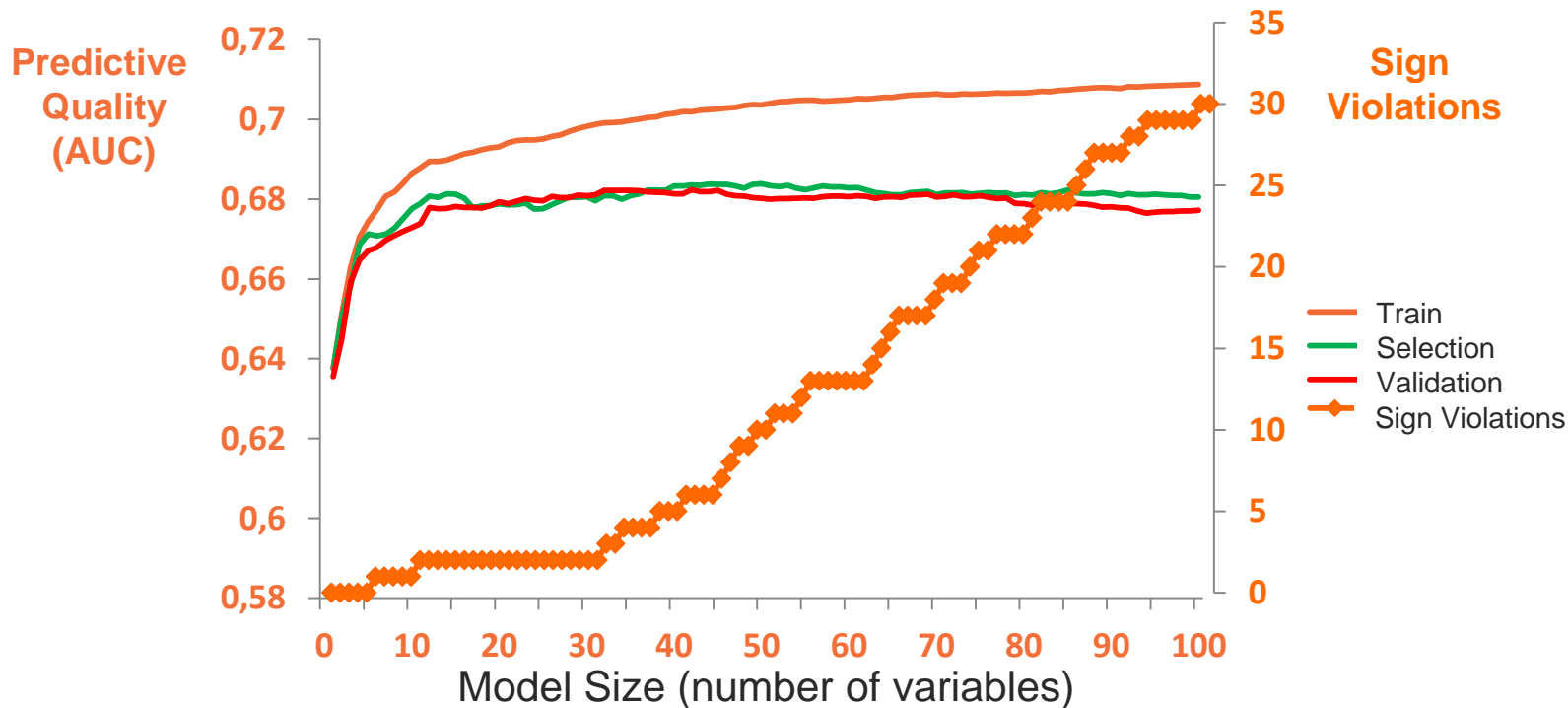




2004 2005 2006 2007 2008

# TANGO

2009 2010 2011 2012 2013

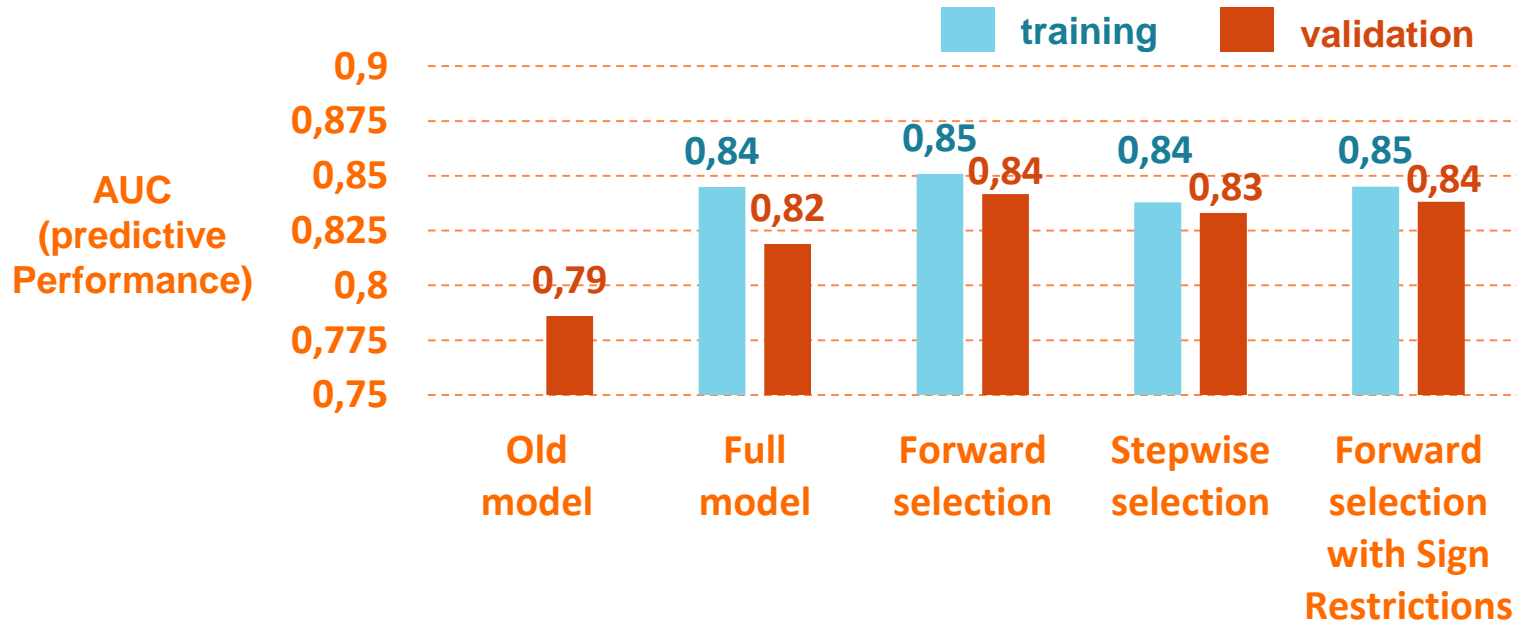




# TANGO



- FSR



# PYTHON PREDICTIONS



- Predictive Analytics in Marketing, Risk & Operations
- Since 2006
- References:





1986

---

AZTECA





# AZTECA



- Evaluation / recalibration
- New models
  - New variables (e.g. Internet orders)
  - Validation by Cofidis
  - Discretisation of variables

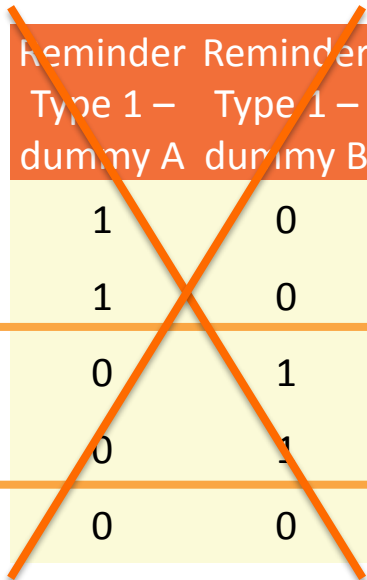




# AZTECA



Bad Payment	Reminder Type 1 - original	Reminder Type 1 - grouped	Reminder Type 1 - dummy A	Reminder Type 1 - dummy B	Reminder Type 1 - incidence
1	10	A	1	0	100%
1	8	A	1	0	100%
1	5	B	0	1	50%
0	1	B	0	1	50%
0	0	C	0	0	0%





# AZTECA



- Evaluation of scores
- New scores
  - New variables (e.g. Internet orders)
  - Validation by Cofidis
  - Discretisation of variables
  - Smoothed weights of evidence
- Monitoring

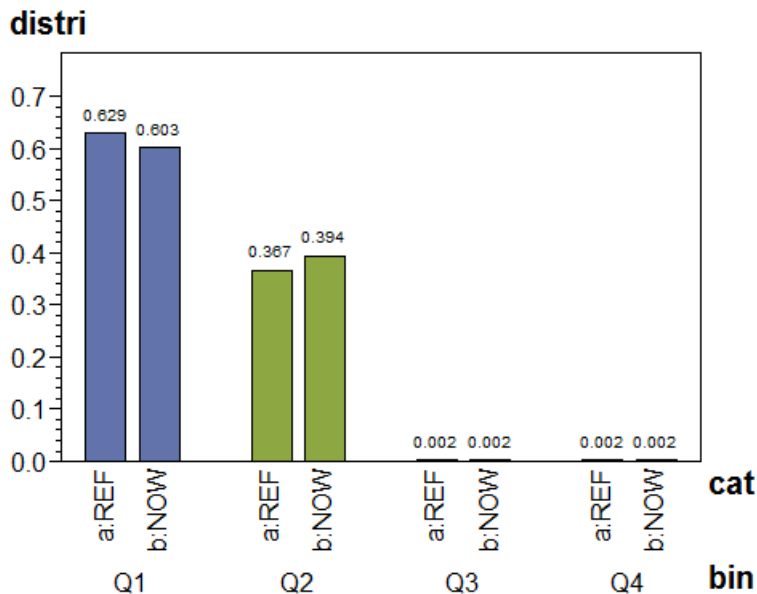




# AZTECA



- Monitoring of recently scored orders: score quadrant

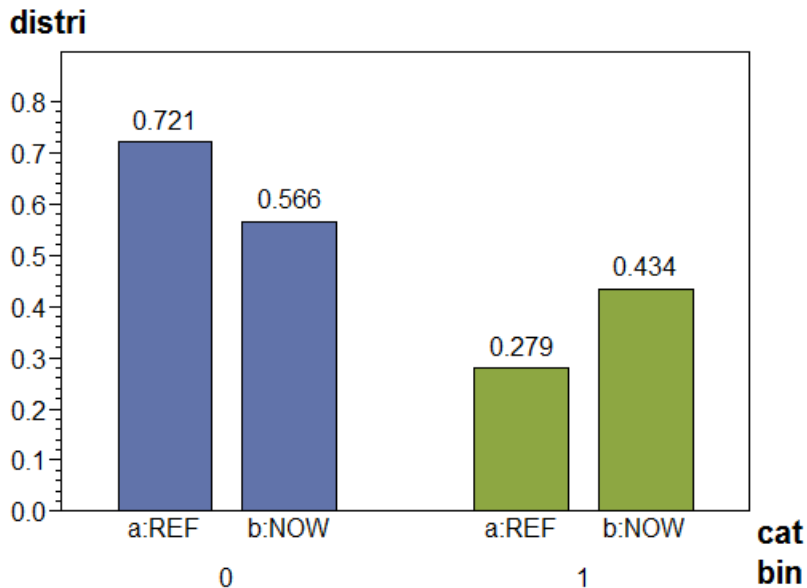




# AZTECA



- Monitoring of recently scored orders: distribution change

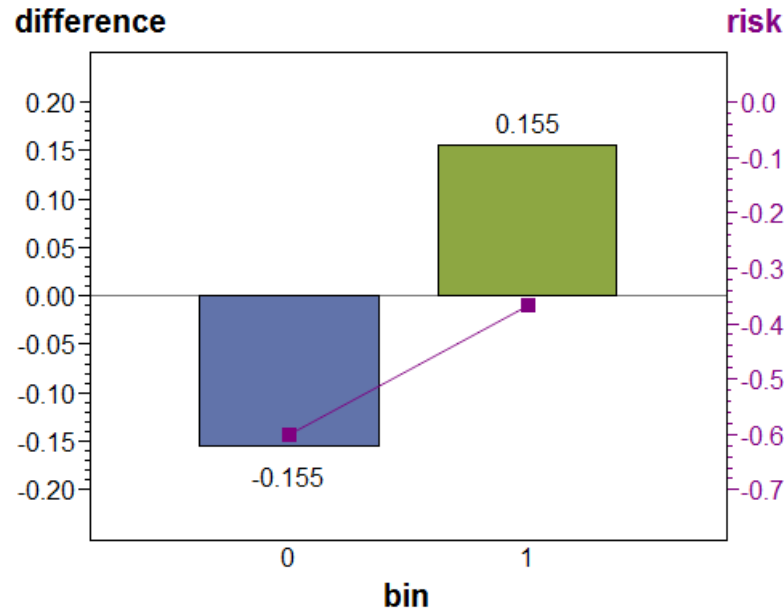




# AZTECA



- Monitoring of recently scored orders: distribution change





# AZTECA



- Monitoring of recently scored orders: estimated impact

Variable	% Change in Risk	Risk Evolution
Score	5.6%	Higher
Predictor 1	-0.3%	Stable
Predictor 2	-15.1%	Lower
Predictor 3	-2.3%	Stable
Predictor 4	1.2%	Stable
...	...	...







# AZTECA



- Monitoring of recently scored orders: log tracking

Step	File	Node	warnings	errors	zero obs	date	time
0	START	Start	0	0	0	25JUN2008	10:48:07
1	PR	<pre>7788 * STEP 0: Extract from server; 7789 * We keep the scores of the correct month and remove ecart direct.; 7790 data min.opvscore2;</pre>					07
2	CA	<pre>7791 set u.opvscore1_juli_okt (where = ((substr(put(datscor,8.),1,6) = "&amp;month") and 7791! (strip(scr_acp) not in ("E","I","N","T")))); 7792 run;</pre>					07
3	NE	<pre>NOTE: There were 20220 observations read from the data set U.OPVSCORE1_JULI_OKT. WHERE (SUBSTR(PUT(datscor, 8.), 1, 6)='201310') and STRIP(scr_acp) not in ('E', 'I', 'N', 'T');</pre>					07
4	CA	<pre>NOTE: The data set MIN.OPVSCORE2 has 20220 observations and 57 variables. NOTE: DATA statement used (Total process time):</pre>					08
5	DIS	<pre>real time          0.11 seconds cpu time           0.07 seconds</pre>					22



2014

---

# BRAZUCA

*“Der Ball ist rund”  
Sepp Herrberger*





# BRAZUCA



- New models
  - Brainstorm with internal experts (manual procedure)
  - Clean discretisation → incidence

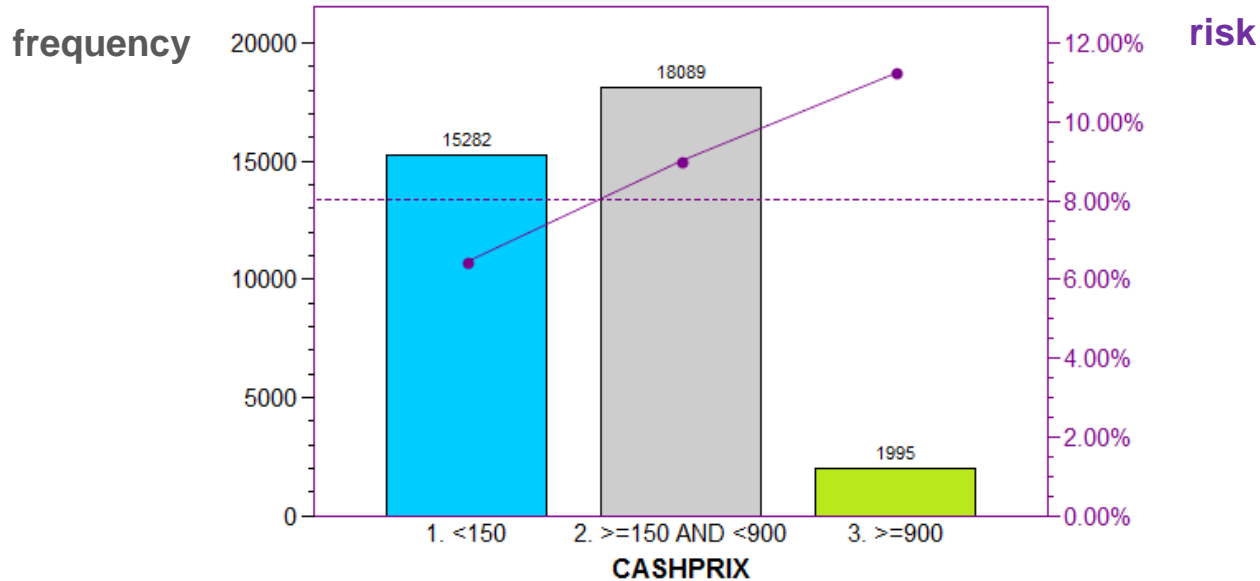




# BRAZUCA



- Clean discretisation: profiling





# BRAZUCA



- New models
  - Brainstorm with internal experts (manual procedure)
  - Clean discretisation → incidence
  - Logistic regression using FSR procedure
  - Credit vs fraud models
  - Impact graphs
- New monitoring

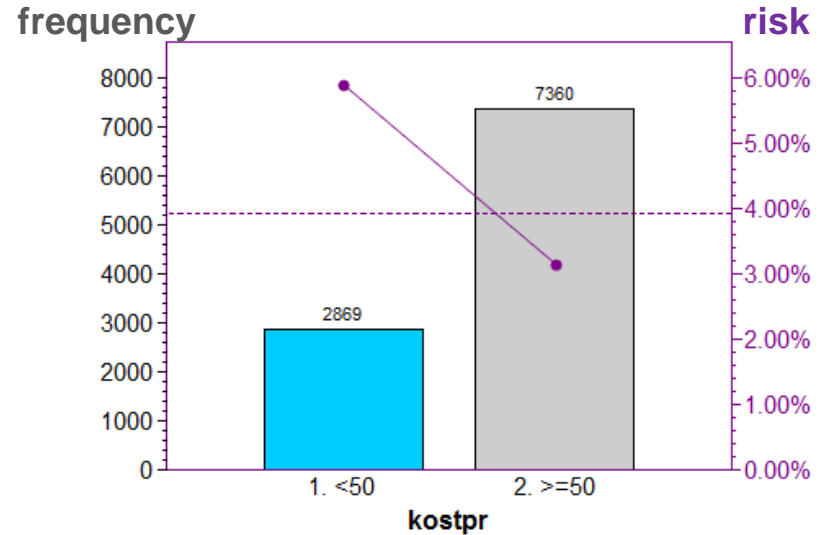
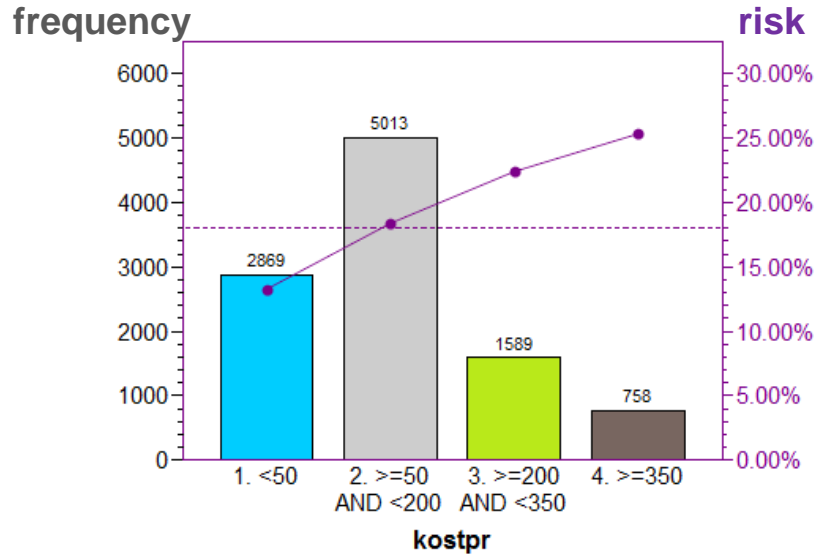




# BRAZUCA



- Credit vs fraud models



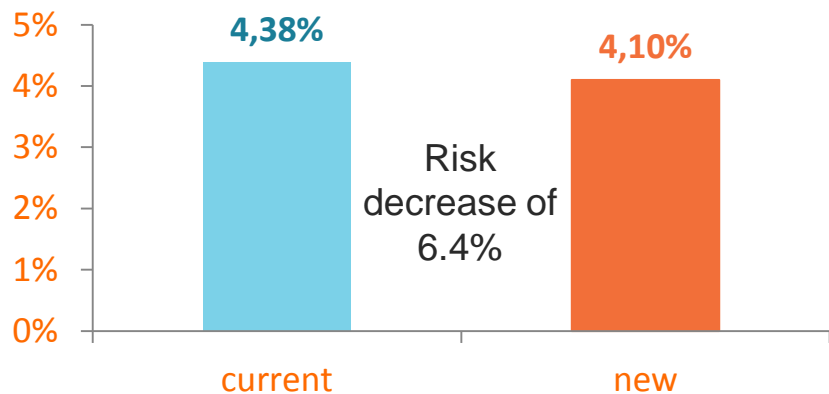


# BRAZUCA

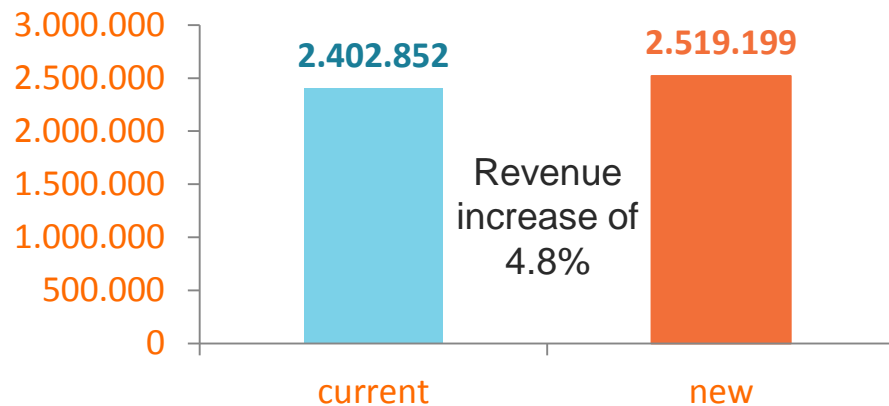


- Impact graphs: amount@risk= 100k

### Risk



### Amount

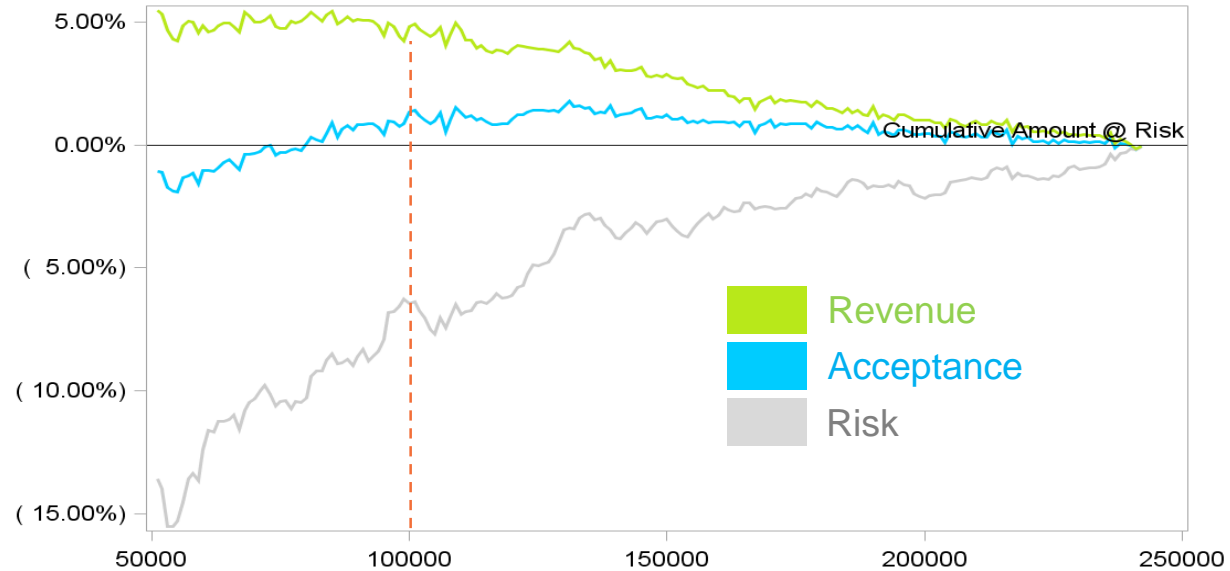




# BRAZUCA



- Impact graph





# Conclusions

- Discretize variables -> incidence
  - Interpretability
- Preference for Logit with FSR
  - Interpretability & performance
- Monitoring
  - Continuous evaluation
- Performance graphs
  - Alignment with business objectives



# Results

Between 2006 and 2012:

- Unigro's revenue increased with 25%
- Risk decreased with 0,9 percentage points
- Revenue on credit orders increased with 38%

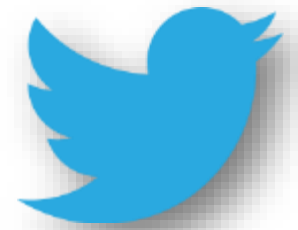


# Lessons: BRAZUCA vs credit scoring

- Performance is key but looks & story will sell
- Elaborate testing is crucial (internal & external)
- Focus on continuous improvements
- At some point we (will) need automation



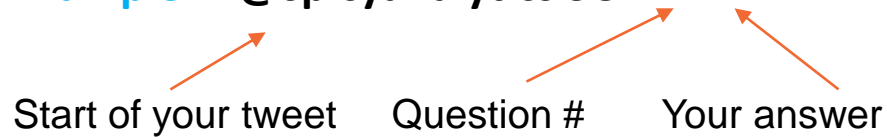
## SAS Forum | Twitter Contest – Tweet to win prizes!



- 3. <Question?>
- A. <Answer 1>
- B. <Answer 2>
- C. <Answer 3>

**Tweet your answer:**

**Example:** @spicyanalytics 3C



**Prizes to win:**

- 1<sup>st</sup> prize: a ticket for Analytics 2015
- 2<sup>nd</sup> prize: a book of Prof Bart Baesens: “Analytics in a big data world”
- 3<sup>rd</sup> to 30<sup>th</sup> prize: chocolates with pepper

**Winners will be contacted post-Forum !**

