



## Data Scientist for 1h

Silvia Quarteroni, NLP Expert

Jérôme Berthier, Head of BI & Big Data

Geneva, 22.04.2015

# OUTLINE

1. ELCA
2. From myth to reality
3. Some proofs 😊
4. Conclusion

Visit us at our booth no. A23

# ELCA

## Presentation

ELCA is one of the largest information technology companies in Switzerland, and a leader in the fields of **software development, systems integration, business consulting and applications management.**

- **Founded:** 1968 as electro-calcul for the software navigation of the Grand Dixence Dam
- **Employees:** Over 700 employees, mostly highly skilled (IT) professionals with university degrees
- **Offices:** Lausanne (headquarters), Zurich, Geneva, Bern, Madrid, Paris, Ho Chi Minh City (Vietnam)
- **Turnover:** More than CHF 100 million in 2014
- **Quality Standards:** ISO 9001 (since 1993), ISO 14001 (since 2011), CMMI level 3 (since 2007)
- **Project Experience:** > 1000 customer projects in ten years
- **Awards:**



---

# ELCA

## Business intelligence and Big Data

More than **50 collaborators** specialized in BI (ZH-BE-LS-GE)

- Elicitation of needs, support, training...
- Extraction, transformation, consolidation or migration of Data
- Modeling and implementing of Databases
- Implementing of standard reporting, ad-hoc reporting and dashboard
- Handling of unstructured data
- Social Media and/or Web data extraction
- Advanced analysis (Text processing, Predictive analysis, graph analysis, GIS, Search Engine...)
- Implementation of Big Data Projects

---

## FROM MYTH TO REALITY

### Urban legends

Big data = Volume + Variety + Velocity

Need petabytes of data to learn anything

Need «the cloud» to do anything «big» with your data

All data becomes public and accessible

Statistical models are unreliable/unfit for production

Big data is an automatic solution generator

## MYTH

---

## REALITY

Big data = Volume and/or Variety and/or Velocity

See upcoming examples

Big data without internet connection

Access to data remains limited

Statistical models are fast and have predictable accuracy

Smoke in, smoke out -> you still need to know your business

# THE BIG DATA «STACK»

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

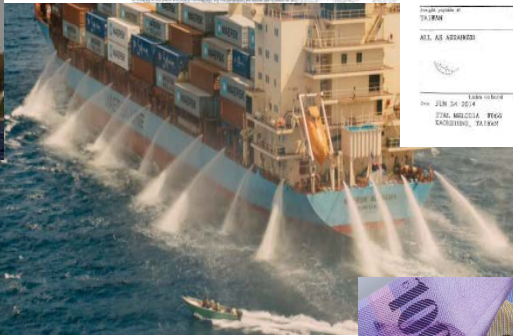
Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# ENTITY SEARCH

Reduce manual checking of invoices, contracts, bills, certificates



**ORIGINAL BILL OF LADING**

TO THE ORDER OF: HANSHIN OCEAN CONTAINER LTD. HANSHIN BAYVIEW QUAY AREA, LEBANON

CONSIGNEE: TO THE ORDER OF: HANSHIN OCEAN CONTAINER LTD. HANSHIN BAYVIEW QUAY AREA, LEBANON

SHIPPER: HANSHIN OCEAN CONTAINER LTD. HANSHIN BAYVIEW QUAY AREA, LEBANON

DATE OF ISSUE: JUN 24 2014

ISSUE AT: LEBANON

SHIP: WAN HAI

PORT OF ORIGIN: SINGAPORE, SINGAPORE

PORT OF DESTINATION: HALIM, MALAYSIA

DESCRIPTION OF GOODS: 1000 BAGS OF SUGAR

QUANTITY: 1000 BAGS

WEIGHT: 100000 KG

VOLUME: 10000 CBM

MARKS AND NUMBERS: AS ATTACHED C/B

**WAN HAI**

**BILL OF LADING**

BL No. 0019481443

S/O No. 0024

**ORIGINAL**

SHIPPER: HANSHIN OCEAN CONTAINER LTD. HANSHIN BAYVIEW QUAY AREA, LEBANON

CONSIGNEE: TO THE ORDER OF: HANSHIN OCEAN CONTAINER LTD. HANSHIN BAYVIEW QUAY AREA, LEBANON

SHIP: WAN HAI

PORT OF ORIGIN: SINGAPORE, SINGAPORE

PORT OF DESTINATION: HALIM, MALAYSIA

DESCRIPTION OF GOODS: 1000 BAGS OF SUGAR

QUANTITY: 1000 BAGS

WEIGHT: 100000 KG

VOLUME: 10000 CBM

MARKS AND NUMBERS: AS ATTACHED C/B

**CERTIFICATE OF ORIGIN FORM B**

Issued in Vietnam

1. Goods consigned from (Exporter's business name, address, country): HANSHIN OCEAN CONTAINER LTD. HANSHIN BAYVIEW QUAY AREA, LEBANON

2. Goods consigned to (Consignee's name, address, country): TO THE ORDER OF: HANSHIN OCEAN CONTAINER LTD. HANSHIN BAYVIEW QUAY AREA, LEBANON

3. Nature of transport and route (as known):

4. Marks, numbers and kind of packages; description of goods: 1000 BAGS OF SUGAR

5. Gross weight or other quantity: 100000 KG

6. Number and date of invoice: JUN 24 2014

7. Country of origin: VIETNAM

8. Declaration by the exporter: I hereby declare that the goods are produced in Vietnam and that they comply with the origin requirements specified for goods consigned to the above-mentioned consignee.

\* No contractual document

# ENTITY SEARCH

Business

<b>Banking</b>	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools



# ENTITY SEARCH

Business

<b>Banking</b>	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	<b>Named entity recognition</b>	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# ENTITY SEARCH

Business

<b>Banking</b>	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	<b>Named entity recognition</b>	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

<b>Search engines</b>	ML/Statistical packages	Big data algorithms	
<b>NLP toolkits</b>	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# ENTITY SEARCH

Business

<b>Banking</b>	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	<b>Named entity recognition</b>	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

<b>Search engines</b>	ML/Statistical packages	Big data algorithms	
<b>NLP toolkits</b>	Graph analysis	Web crawling	

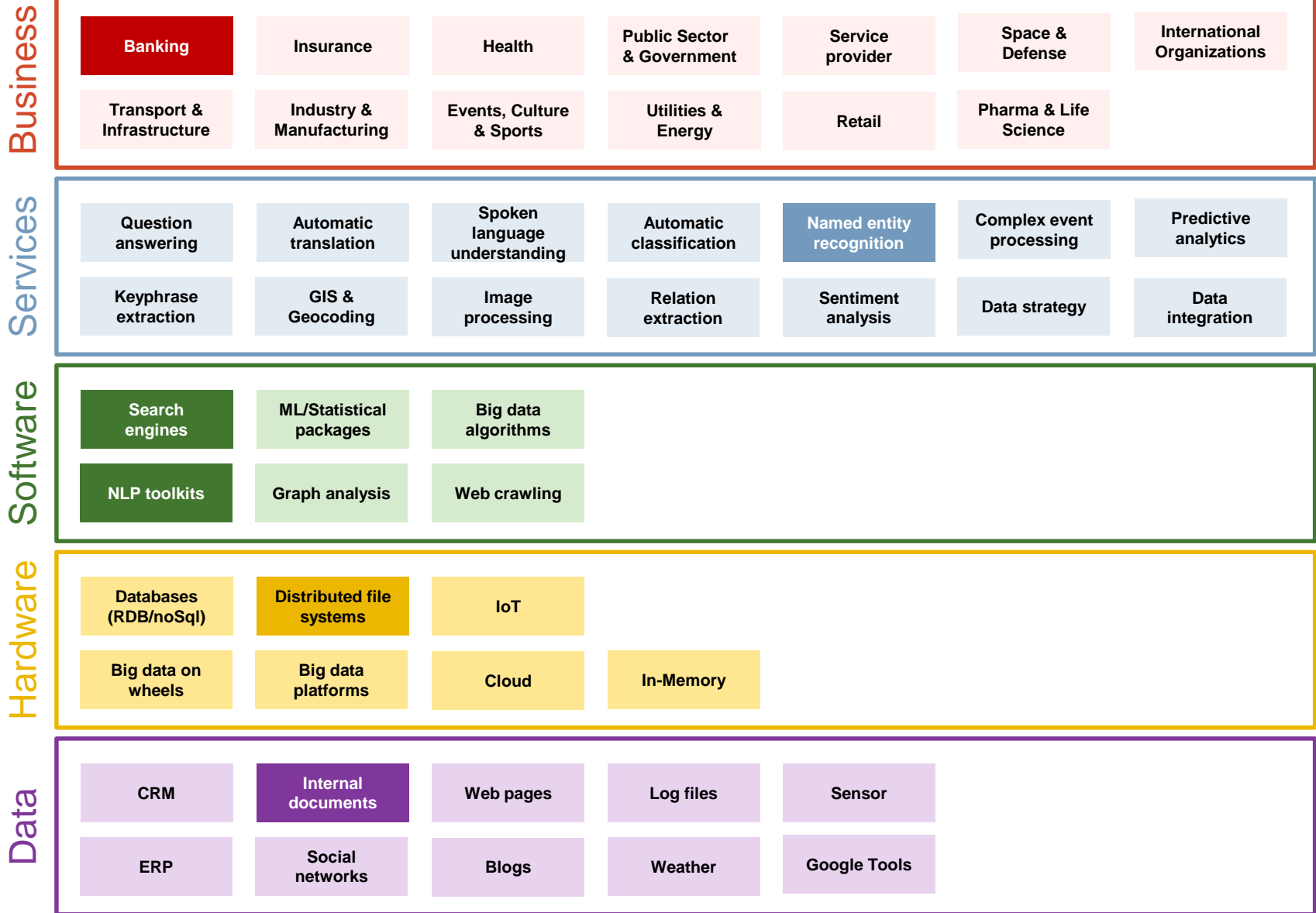
Hardware

Databases (RDB/noSql)	<b>Distributed file systems</b>	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# ENTITY SEARCH





# ENTITY SEARCH

From OCR to conformity checks

The image shows a document with an entity search overlay. The document text includes:

**TUKITNACHRA STEEL & COMMODITIES TRADING LTD.**  
The Financial Services Center, Stone Ground Kingstown VC65398  
St Vincent and The Grenadines

To: **ATG NOUR STEEL**  
529 Silver Road, Me...  
Place, behind tati m...  
Sector-2 Rodhini, **India**

Shipping Information Fo...

Quantity	
132 tons	Roll... As... No...

►Country of Origin :  
►Port of loading : **India**  
Place of Delivery : **India**  
Terms of delivery : **India**

Drawn under the doc...  
Delhi 110 085, **India**

Bank account : **Lukinos Bank SA, Toronto**  
IBAN : CA15 2315 1505 0558

The overlay features the text "ENTITY SEARCH" at the top left, a large "ENTITY SEARCH" in the center, and a search bar at the bottom with the text "arion|".

# ENTITY SEARCH

From OCR to conformity checks

**LOCATIONS**

- Corinth 1
- FOB AGHII THEODORI 1
- Greece 1
- TURKEY 1

**ORGANIZATIONS**

- IZMIR REFINERIES S.A. 1

**DOCUMENTS**

Filters: arion

**Certificate of origin 2.txt**  
Created by Hugo on 2015-04-17 in /Hugo\_4299/out/ [EXPORT](#)

(J)OR ATATURC  
01L IZfVIR REFINERIES S.A. AGml™EODOM™™™H  
CERTIFICATE OF ORIGIN N° ..335.....  
We, the undersigned, mor ataturc oil izmir refineries s .a. , Marousi, Greece  
hereby certify that the merchandise described shipped on-ARION...-  
Date,.....u/07/2014.....consigned to .....  
is of the quantity cited and has been produced from the Company's Refinery at Aghii  
Theodori, Corinth Greece.  
Description  
t ir .sn î nppM s ut. phi jr max  
Quantity  
20000.000 METRIC TONS DJ VAC-  
19973.735 METRIC TONS IN AIR  
23940711 LITRES AT 15 DEC C  
-M3.7 15LC-  
AT 60° F  
LOADING PORT: FOB AGHII THEODORI  
DESTINATION : TURKEY  
For  
MOR ATATURC  
OIL IZMIR REFINERIES S.A.

First Previous 1 Next Last

**VESSELS**

- ARION 1

**FILTER BY AUTHOR**

- Hugo 1

# ENTITY SEARCH

From OCR to conformity checks

The screenshot displays a web application interface for entity search. It is divided into three main vertical panels:

- LOCATIONS (Left Panel):** A teal sidebar containing a list of locations with counts in circles. Locations include US (2), Cl F Tamato Port (1), Delhi (1), GREAT BRITAIN (1), Grenadines (1), HAIFA (1), ISRAEL (1), ITALY (1), India (1), Kingstown (1), Madagascar (1), and Nemrut Bav (1).
- ORGANIZATIONS (Left Panel):** An orange sidebar containing a list of organizations with counts in circles. Organizations include ATG NOUR STEEL LTD (1), Blabla Bank Ltd (1), COMMODITIES TRADING LTD (1), DUMONT PETROLEUM LLP (1), Financial Services Center (1), KORIMOV ENERGY GROUP SA (1), Lukinos Bank SA (1), STEEL LTD (1), and Total (1).
- 2 DOCUMENTS (Center Panel):** A document viewer showing two invoices. The first is "Invoice 1.txt" created by Jonathan on 2015-04-17, with an "EXPORT" button. The text includes "TUKITNACHRA STEEL & COMMODITIES TRADING LTD.", "The Financial Services Center, Stone Ground Kingstown VC65398", and "St Vincent and The Grenadines". The second is "Invoice 2.txt" created by Zoe on 2015-04-17, also with an "EXPORT" button. The text includes "oriinov Energy" and "KORIMOV ENERGY GROUP SA". A pagination bar at the bottom shows "First", "Previous", "1", "Next", and "Last".
- VESSELS (Right Panel):** A purple sidebar showing "MT AZOV MARINER" with a count of 1.
- FILTER BY AUTHOR (Right Panel):** A red sidebar showing authors "Jonathan" (1) and "Zoe" (1).

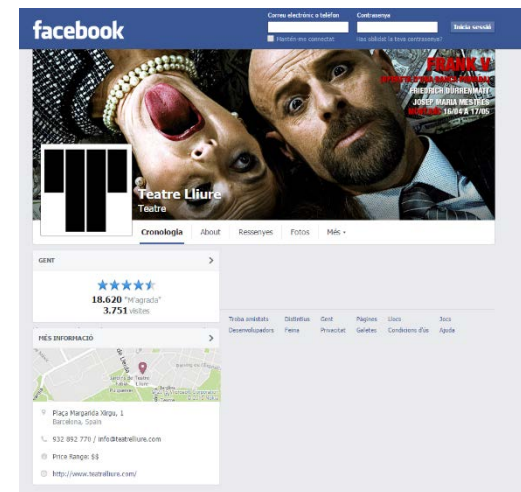
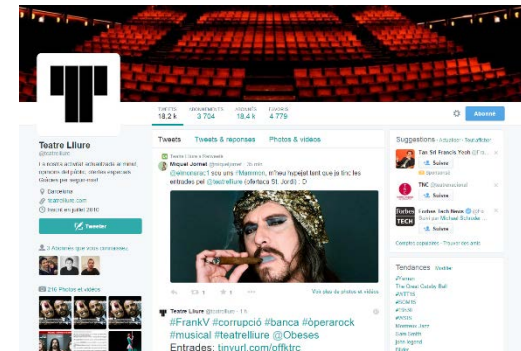


# TWITTER FOR SALES

How well do we know our customers on social networks?



Solution ?



# TWITTER FOR SALES

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# TWITTER FOR SALES

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# TWITTER FOR SALES

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# TWITTER FOR SALES

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# TWITTER FOR SALES

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# TWITTER FOR SALES

NLP and social networks

## Sentiment analysis

Automatic prediction of user **gender** and **age** via tweets

Identification of **dialog act**

 Bastian Baker retweeted  
 **crazybabysitterandy** @crazyandy76 · Dec 20  
OMG!! 😍😍 MERCI infiniment! @Bastian\_Baker @Chriszindel88  
@nathanbonj @SimJacc @JorisAmann @tousenchoeur #AMAZING



 Bastian Baker retweeted  
 **TempoMag** @Tempo\_Mag · Feb 23  
.[@Bastian\\_Baker](#) était de passage à [@OlympiaMontreal](#) vendredi. Nous avons assisté à son spectacle: [wp.me/p1yZTo-2Zp](http://wp.me/p1yZTo-2Zp)

  12   20 

 Bastian Baker retweeted  
 **Musimax** @Musimax · Feb 20  
La dernière belle découverte de [@CheliSauve](#) : l'auteur-compositeur-interprète suisse [@Bastian\\_Baker](#) [bit.ly/1zRDSpz](http://bit.ly/1zRDSpz)

  11   18 

[View summary](#)

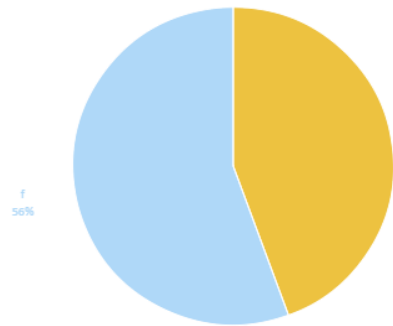
# TWITTER FOR SALES

## Automatic prediction of age and gender on Twitter

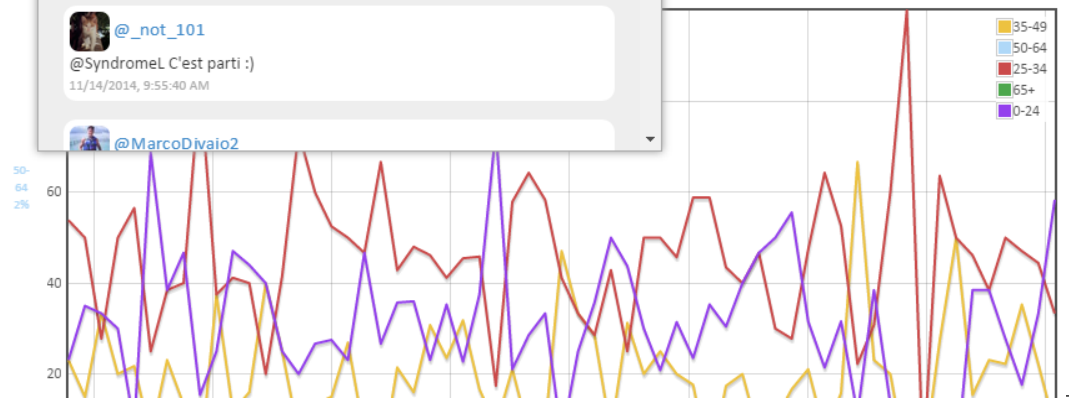
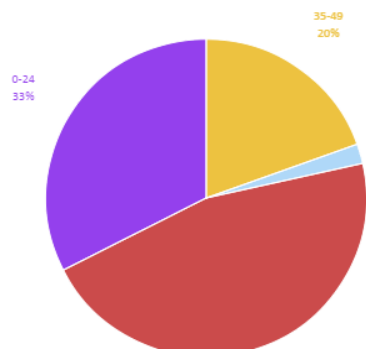
### Welcome to the Classification Machine

Analytics result

#### Genders



#### Ages



Label: m

Close



@SimonGardette

@Gwaradenn C'est toi le nouveau chef de la biture en chef ? Voilà une bonne nouvelle :) @seocamp @aymerictwit

11/14/2014, 9:55:36 AM



@Koko753Nick

RT @boydupacifi: Tortue de nouvelle calédonie :) <http://t.co/6xHikUnzoz>

11/14/2014, 9:55:38 AM



@\_not\_101

@SyndromeL C'est parti :)

11/14/2014, 9:55:40 AM



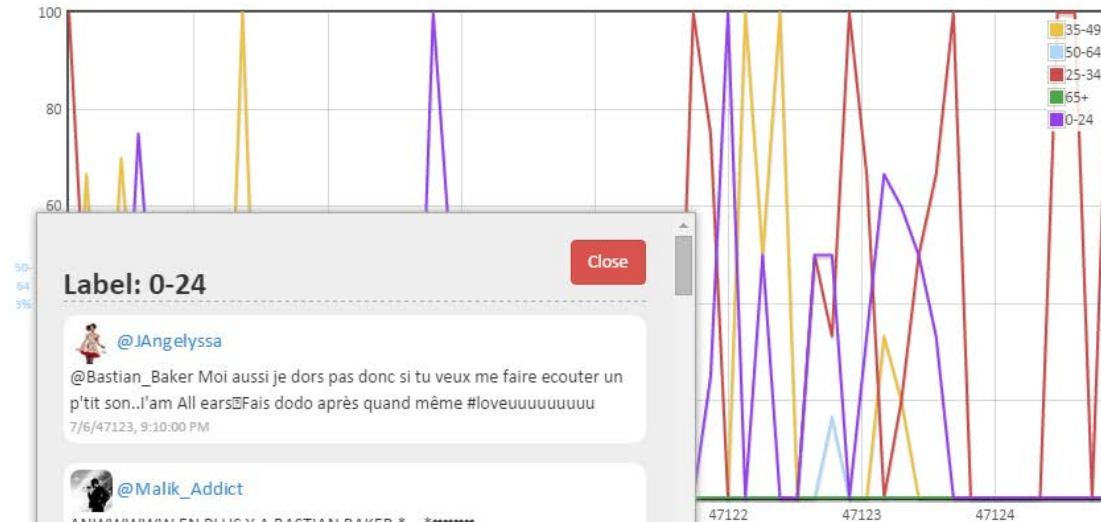
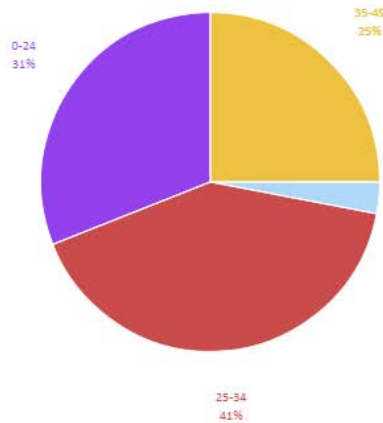
@MarcoDivaio2



# TWITTER FOR SALES

## Age and sentiment of tweets about Bastian Baker

### Ages



Label: 0-24

Close

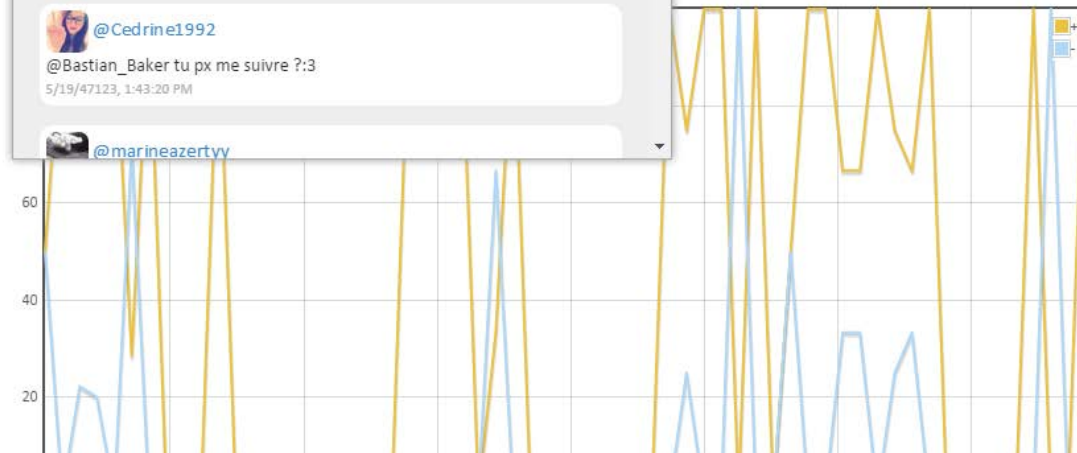
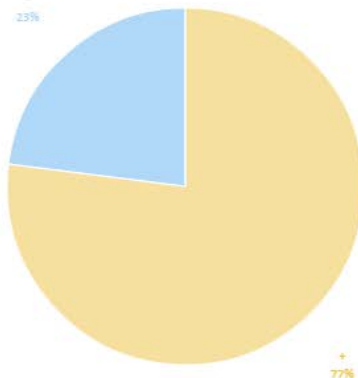
@JAngelyssa  
@Bastian\_Baker Moi aussi je dors pas donc si tu veux me faire ecouter un p'tit son..l'am All ears☺Fais dodo après quand même #loveuuuuuuuuu  
7/6/47123, 9:10:00 PM

@Malik\_Addict  
ANWWWWW EN PLUS Y A BASTIAN BAKER \*---\*♥♥♥♥  
6/7/47123, 10:13:20 AM

@Cedrine1992  
@Bastian\_Baker tu px me suivre ?:3  
5/19/47123, 1:43:20 PM

@marineazertyv

### Sentiments



# TWITTER FOR SALES

## User profiling on Twitter: “Bastian Baker”

### Most active users

Show  entries

Search: |

Username	Tweets	Sentiment	Age	Gender	Dialog Act
Brigitte_Clerc	10	+ (75.7%)	25-34 (42.4%)	f (51.6%)	com (70.9%)
JAngelyssa	6	+ (79.1%)	25-34 (79.6%)	f (88.2%)	com (74.6%)
sophieguntern	4	+ (67.1%)	25-34 (47.5%)	m (58.6%)	com (39.1%)
crazyandy76	3	+ (86.3%)	25-34 (35.6%)	f (58.4%)	mis (62.1%)
egnalos98	3	+ (65.6%)	25-34 (37.0%)	f (64.4%)	mis (93.0%)
Nora_Elidrissi	2	+ (85.3%)	35-49 (39.0%)	m (79.1%)	mis (99.0%)
Geeti_801	2	+ (79.4%)	0-24 (42.4%)	f (82.6%)	com (51.7%)
hxrانlik	2	+ (74.5%)	0-24 (46.4%)	f (51.1%)	com (50.2%)
ihavelrwin	2	- (59.8%)	0-24 (69.9%)	m (53.0%)	com (35.5%)
SabrinaB0603	2	+ (83.5%)	25-34 (37.7%)	m (80.3%)	mis (98.6%)

Showing 1 to 10 of 68 entries

Previous

1



2

3

4

5

### Brigitte\_Clerc

-  [@Brigitte\\_Clerc](#)  
@Bastian\_Baker ah le hand claping ça je sais faire...t as vu à Montréal ?! ;-)  
8/11/47123, 8:33:20 AM
-  [@Brigitte\\_Clerc](#)

# TWITTER FOR SALES

## The Secutix monthly report

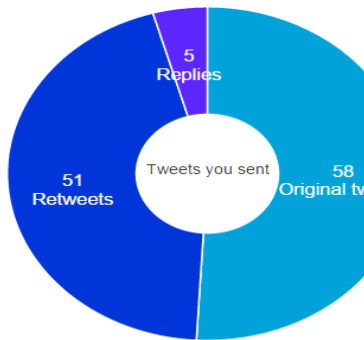
### Your social activity



In **sept. 2014**, you published **114 tweets** that generated **1 749 045 impressions** (



**74 retweets (+722%)** have been created that generated **53 900 impressions (+14**



### Your top 10 ambassadors

Scoring is done using a non-linear formula taking into account the number of followers and the number of tweets mentioning you

- new **Anima 33**  
27 063 followers ● 3 relevant tweets  
Anima és el magazín d'actualitat escènica del @canal\_33. Presentat i dirigit per @tonipuntí, s'emet els dilluns a les 22:45.
- +1 **Ataques i Somriures**  
3 107 followers ● 8 relevant tweets  
Journalist. Social Media Strategist. Theater
- new **HOMBRE LOBO**  
16 814 followers ● 2 relevant tweets
- new **Festival de Tardor de Catalunya - Girona Sa**  
9 977 followers ● 2 relevant tweets
- new **Just be you, but learn from other people! London??**  
1 251 followers ● 5 relevant tweets
- new **Banco Sabadell**  
5 093 followers ● 2 relevant tweets  
?If you can imagine, you can get? Director @BancoSabadell
- new **Cultura**  
20 361 followers ● 1 relevant tweets  
Departament de Cultura. Generalitat de Ca Government of Catalonia.
- new **Teatre i més encara**  
746 followers ● 5 relevant tweets
- 5 **El digital de cultura. Busquem el que és únic entre el que ens és comú.**  
15 613 followers ● 1 relevant tweets
- new **Teatre Nacional de Catalunya**  
15 498 followers ● 1 relevant tweets  
El teatre de tothom, clàssic i contemporani

### The 5 most important followers you gained

- Estudiante y narrador.No espero nada de los demás, es conmigo mismo con quien soy exigente. http://Lco/P7k6uqnsR**
- Twitter oficial de Il Bello Canto.Formación compuesta por cuatro jóvenes cantantes líricos con apasionantes versiones y temas propios que sin duda te encantarán**
- Agencia #eve para superar #events #pla #market**
- La guía defini**

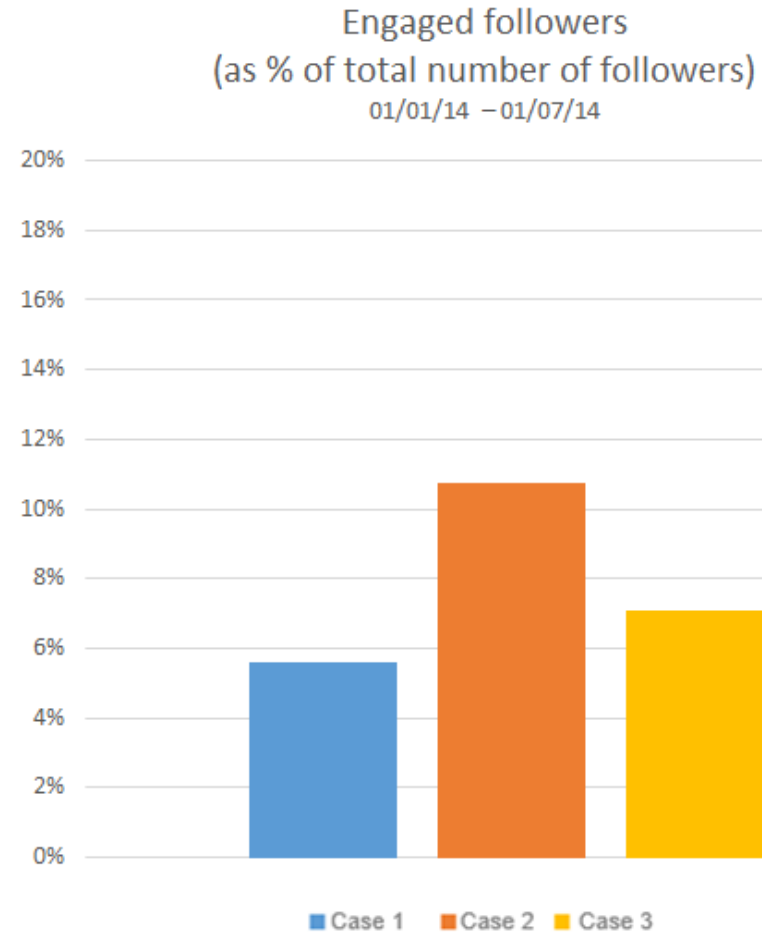
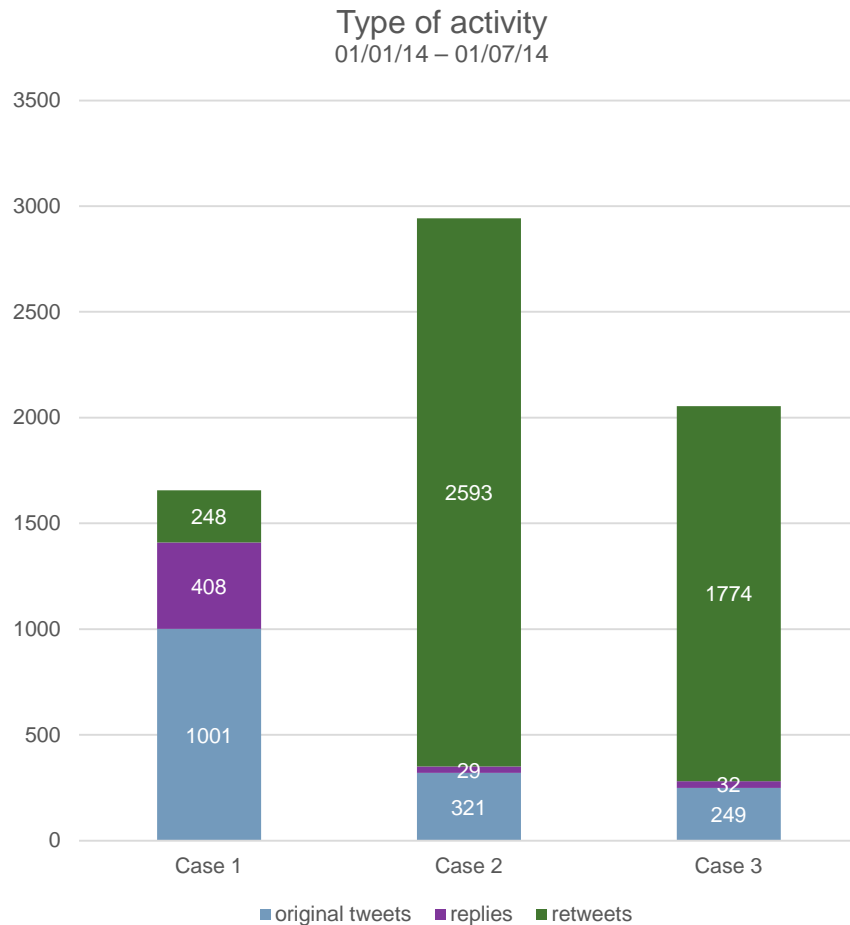
### The 3 most retweeted tweets about you

- Anima 33**  
Anima és el magazín d'actualitat escènica del @canal\_33. Presentat i dirigit per @tonipuntí, s'emet els dilluns a les 22:45.  
20 sept. 2014 9  
És jove, atípic, té talent i és una de les apostes de futur del @teatrelliure. Dilluns, a #Anima33, coneixem a fons a l'actor Pol López.
- Anima 33**  
Anima és el magazín d'actualitat escènica del @canal\_33. Presentat i dirigit per @tonipuntí, s'emet els dilluns a les 22:45.  
28 sept. 2014 8  
Volem seguir dient #ensanima amb @teatrenacional, @teatrelliure, @mercatfllors i ara que també comencem @temporadaaita o @bcnjazzfest.
- Miriam's News**  
"Per fill!" ???? #annalizaran #salaannalizaran @teatrelliure http://t.co/T12AlcxmMd  
2 sept. 2014 6

# TWITTER FOR SALES

## What's better, to tweet or to retweet?

User engagement vs activity type (tweet, reply, retweet)



# EMARKETING OPTIMIZATION

How to get from Spam to Target ?



VS



**MARKETING  
TEAM**



# EMARKETING OPTIMIZATION

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# EMARKETING OPTIMIZATION

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# EMARKETING OPTIMIZATION

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools



# EMARKETING OPTIMIZATION

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# EMARKETING OPTIMIZATION

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# EMARKETING OPTIMIZATION

Targeted eMarketing for customer segments



Targeted eMarketing for each show



# EMARKETING OPTIMIZATION

Targeted eMarketing for customer segments



Targeted eMarketing for each show



# EMARKETING OPTIMIZATION

## CRM Segmentation

Use customer CRM information (gender, location, purchase history, preferred genres...) to cluster clients using learning algorithms.

Identify «VIP» cluster(s) with highest profitability.

### Cluster: 7

1,663	3.02	20.77	0.90	3.90
Count of contact_dwid	Average of advantages_campaigns_used	Average of advantages_received	Average of isEasyAdopter	Average of number_of_active_seasons
159.62	0.56	0.44	0.25	0.39
Average of days_since_last_perf_purch	Average of fraction_internet	Average of fraction_boxoffice	Average of SeasonTicketSeason1	Average of SeasonTicketSeason2
0.62	0.63	1.99	4.50	5.54
Average of SeasonTicketSeason3	Average of SeasonTicketSeason4	Average of PerformancesSeason1	Average of PerformancesSeason2	Average of PerformancesSeason3
5.96	4.02	9.89	11.08	12.49
Average of PerformancesSeason4	Average of QuantitySeason1	Average of QuantitySeason2	Average of QuantitySeason3	Average of QuantitySeason4
0.67	0.21	0.78	0.56	0.60
Average of barcelona	Average of male	Average of female	Average of fraction_times_early	Average of percentage_in_catalan
0.23	0.63	0.16	0.23	0.68
Average of percentage_in_spanish	Average of percentage_dramas	Average of percentage_comedies	Average of percentage_times_alone	Average of percentage_times_couple

# EMARKETING OPTIMIZATION

## CRM Segmentation

Drill down into the VIP cluster and cluster it further.

### Cluster: 1

600	0.55	0.12	0.17
Count of contact_dwid	Average of fraction_times_early	Average of fraction_times_last_minute	Average of percentage_comedies
0.64	0.01	0.02	3.79
Average of percentage_dramas	Average of percentage_music_dance	Average of percentage_musicals	Average of advantages_campaigns_used
1.00	0.00	0.00	1.00
Average of barcelona	Average of upto50kkm	Average of male	Average of female

### Cluster: 3

246	0.57	0.10	0.17
Count of contact_dwid	Average of fraction_times_early	Average of fraction_times_last_minute	Average of percentage_comedies
0.67	0.01	0.02	3.96
Average of percentage_dramas	Average of percentage_music_dance	Average of percentage_musicals	Average of advantages_campaigns_used
0.00	0.31	0.00	1.00
Average of barcelona	Average of upto50kkm	Average of male	Average of female

### Cluster: 4

587	0.55	0.30	0.13
Count of contact_dwid	Average of fraction_times_early	Average of fraction_times_last_minute	Average of percentage_comedies
0.61	0.03	0.02	2.25
Average of percentage_dramas	Average of percentage_music_dance	Average of percentage_musicals	Average of advantages_campaigns_used
0.87	0.03	0.44	0.55
Average of barcelona	Average of upto50kkm	Average of male	Average of female

### Cluster: 5

229	0.59	0.19	0.16
Count of contact_dwid	Average of fraction_times_early	Average of fraction_times_last_minute	Average of percentage_comedies
0.66	0.01	0.02	1.99
Average of percentage_dramas	Average of percentage_music_dance	Average of percentage_musicals	Average of advantages_campaigns_used
0.00	0.88	0.42	0.55
Average of barcelona	Average of upto50kkm	Average of male	Average of female

Clusters 4 and 5: good targets for season tickets offers

Clusters 1 and 3: good targets for promotional codes

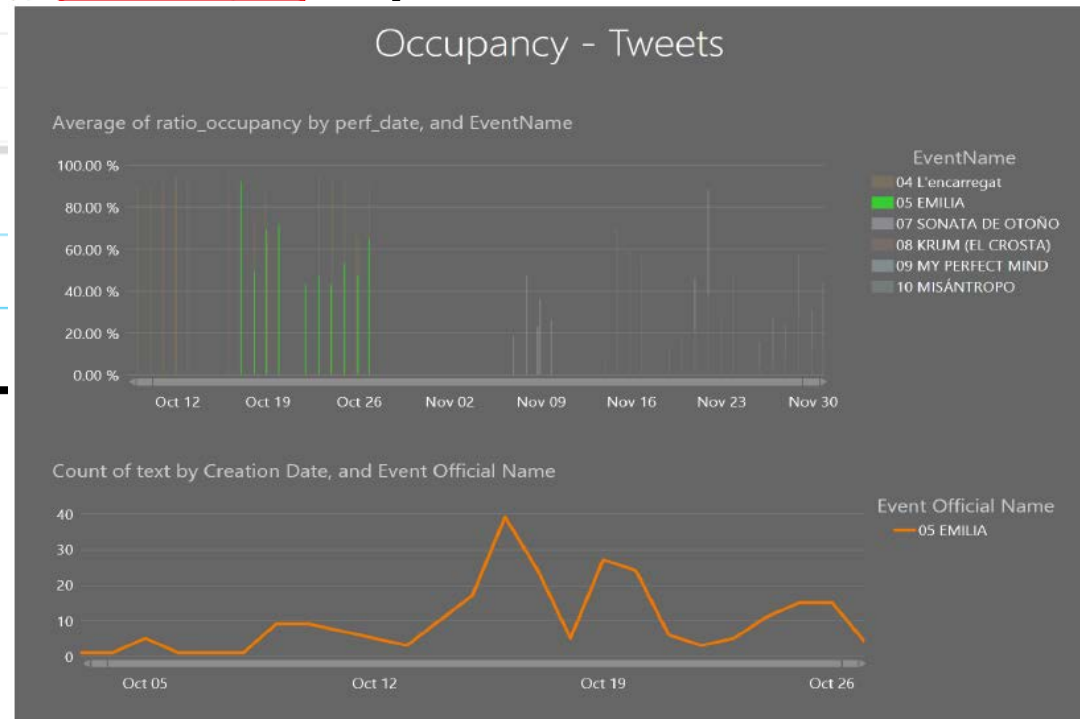
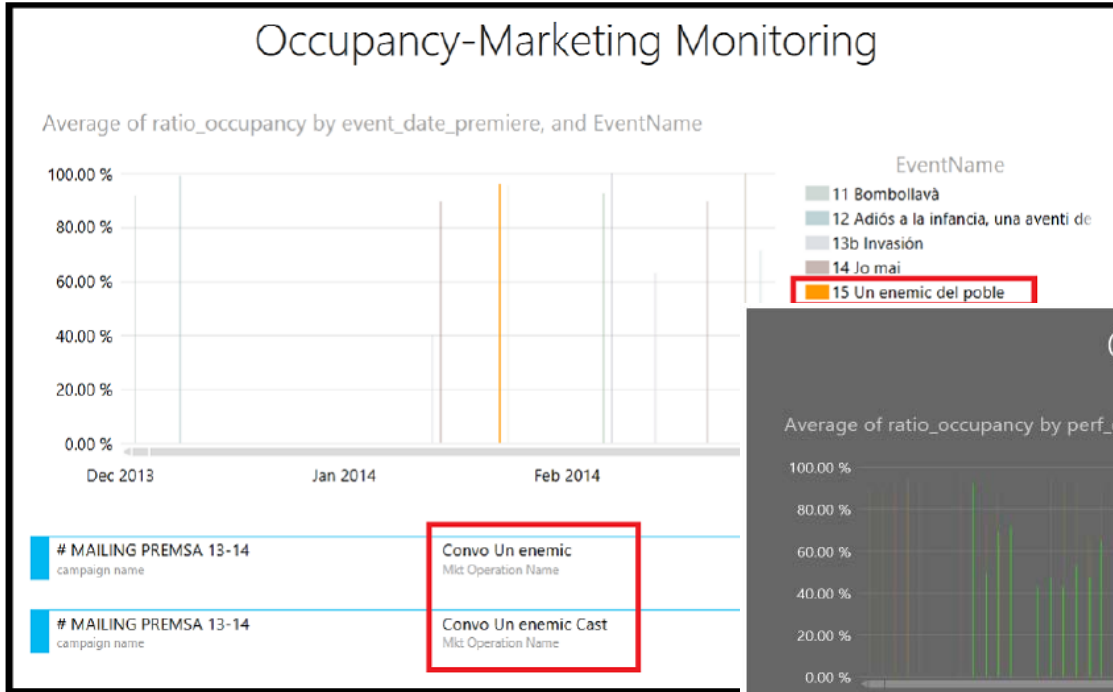
Clusters 1, 4 and 5: good targets for last-minute offers

Cluster 4: good target for afternoon shows

# EMARKETING OPTIMIZATION

## OCCUPANCY ANALYSIS

A marketing tool for monitoring the occupancy of shows

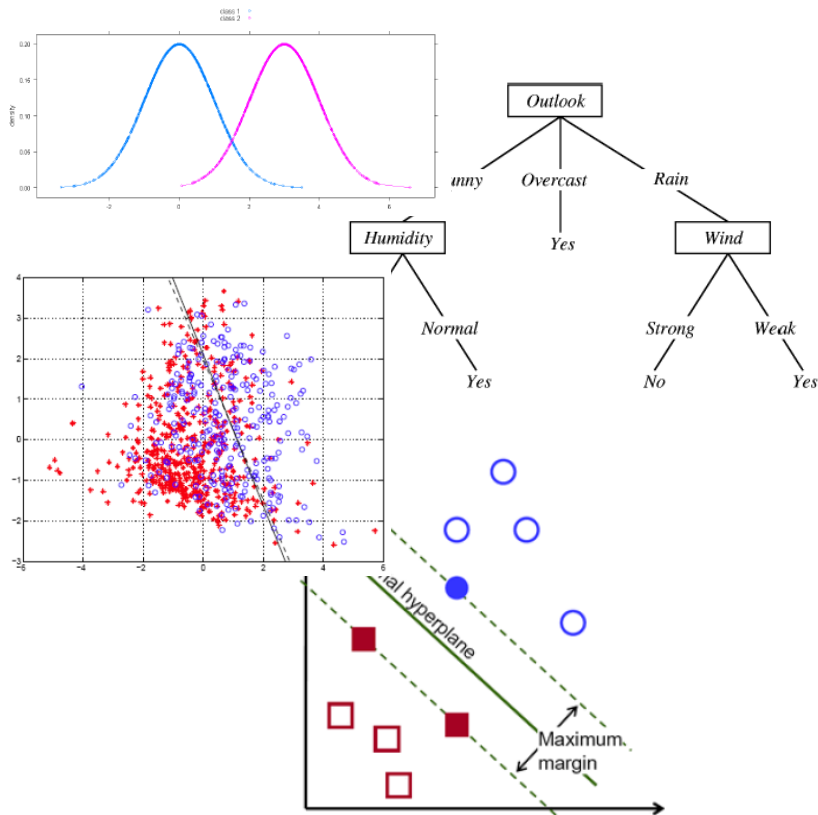




# EMARKETING OPTIMIZATION

## OCCUPANCY ANALYSIS

Statistical analysis: predict the occupancy of future shows based on show type, time, weather information, .... and devise a targeted eMarketing strategy for each type of show (drama, musical...) or even for individual shows.





# PREDICTIVE MAINTENANCE OF DOORS

Open or Late



# PREDICTIVE MAINTENANCE OF DOORS

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# PREDICTIVE MAINTENANCE OF DOORS

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms
NLP toolkits	Graph analysis	Web crawling

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# PREDICTIVE MAINTENANCE OF DOORS

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms
NLP toolkits	Graph analysis	Web crawling

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# PREDICTIVE MAINTENANCE OF DOORS

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms	
NLP toolkits	Graph analysis	Web crawling	

Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# PREDICTIVE MAINTENANCE OF DOORS

Business

Banking	Insurance	Health	Public Sector & Government	Service provider	Space & Defense	International Organizations
Transport & Infrastructure	Industry & Manufacturing	Events, Culture & Sports	Utilities & Energy	Retail	Pharma & Life Science	

Services

Question answering	Automatic translation	Spoken language understanding	Automatic classification	Named entity recognition	Complex event processing	Predictive analytics
Keyphrase extraction	GIS & Geocoding	Image processing	Relation extraction	Sentiment analysis	Data strategy	Data integration

Software

Search engines	ML/Statistical packages	Big data algorithms
NLP toolkits	Graph analysis	Web crawling

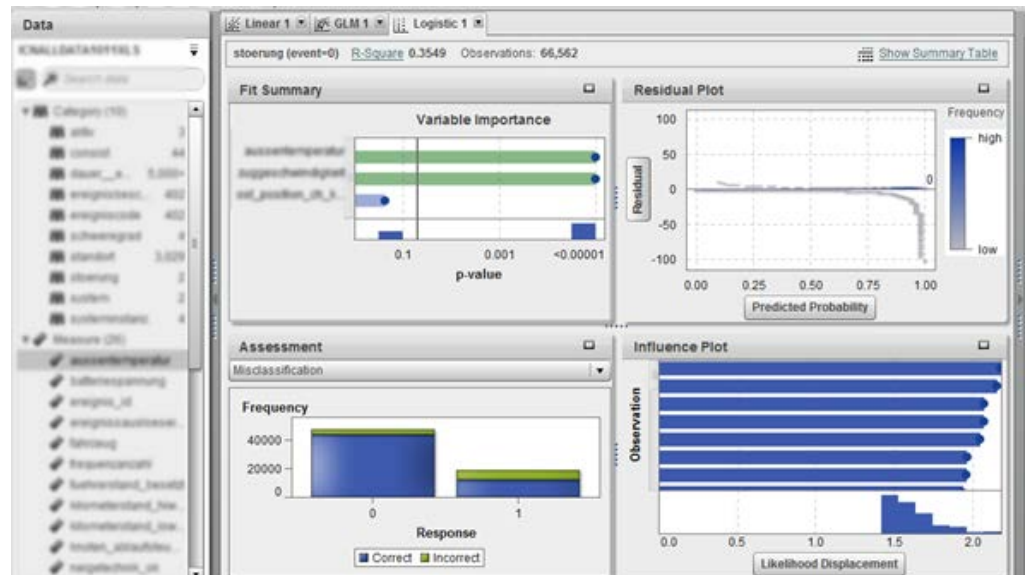
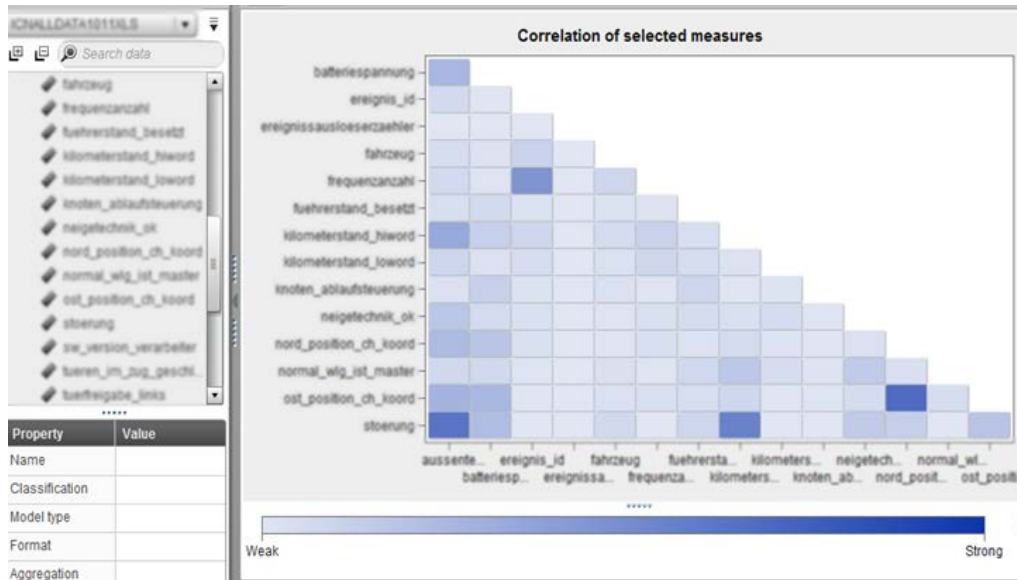
Hardware

Databases (RDB/noSql)	Distributed file systems	IoT	
Big data on wheels	Big data platforms	Cloud	In-Memory

Data

CRM	Internal documents	Web pages	Log files	Sensor
ERP	Social networks	Blogs	Weather	Google Tools

# PREDICTIVE MAINTENANCE OF DOORS



---

## CONCLUSIONS

Making sense of content is not a myth

No massive data → Bring your data and we will provide the rest

No massive investments → Use ELCA big data lab

No massive engagement → Start by POC

Without engaging massive resources,  
you can make Big Data your reality

Visit us at our booth no. A23



# QUESTIONS



# Thank you.

## Contact

Silvia Quarteroni

NLP expert

[silvia.quarteroni@elca.ch](mailto:silvia.quarteroni@elca.ch)

Tel +41 21 613 21 11

Jérôme Berthier

Head of BI & Big Data

[Jerome.Berthier@elca.ch](mailto:Jerome.Berthier@elca.ch)

+41 21 613 21 11



JBerthier

---

ELCA Informatique SA | Lausanne 021 613 21 11 | Genève 022 307 15 11

ELCA Informatik AG | Zürich 044 456 32 11 | Bern 031 556 63 11

[www.elca.ch](http://www.elca.ch)