Data Scientist for 1h

Silvia Quarteroni, NLP Expert

Jérôme Berthier, Head of BI & Big Data

Geneva, 22.04.2015

ELCA We make it work.

eCom GENÈVE | SALON DU E-COMMERCE, DU WEB, DU MOBILE & DU E-MARKETING

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

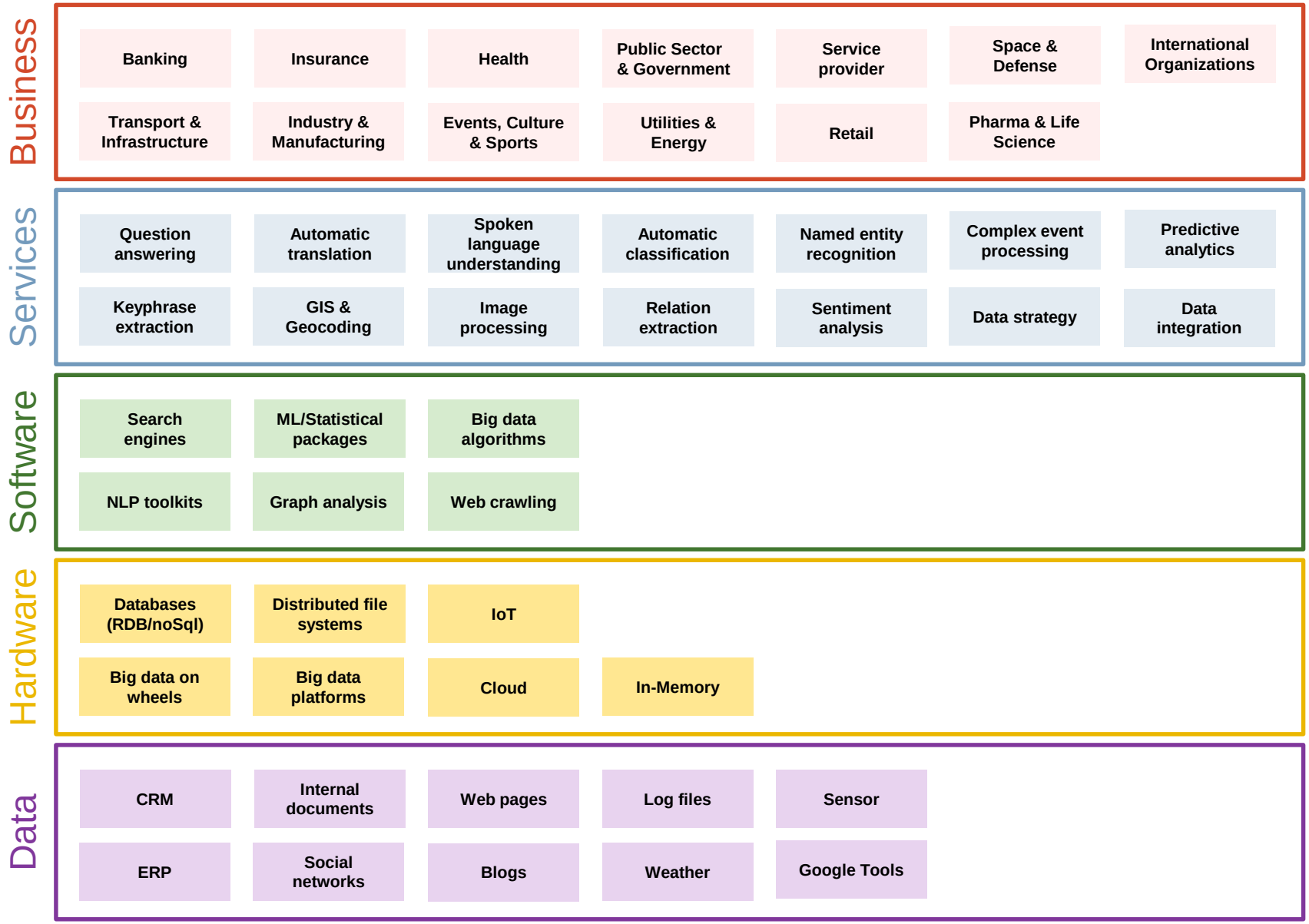
Big data without internet connection

Access to data remains limited

Statistical models are fast and have predictable accuracy

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

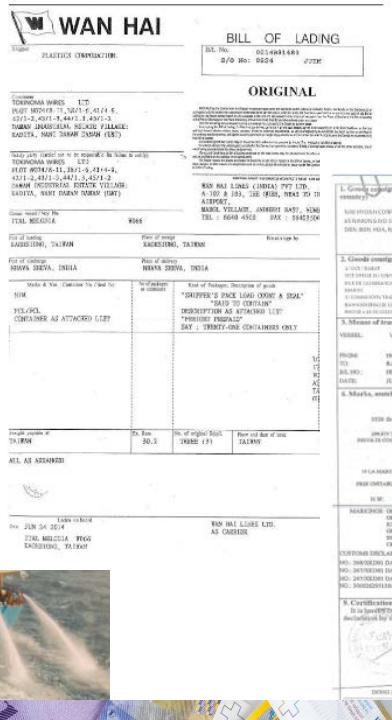
THE BIG DATA «STACK»



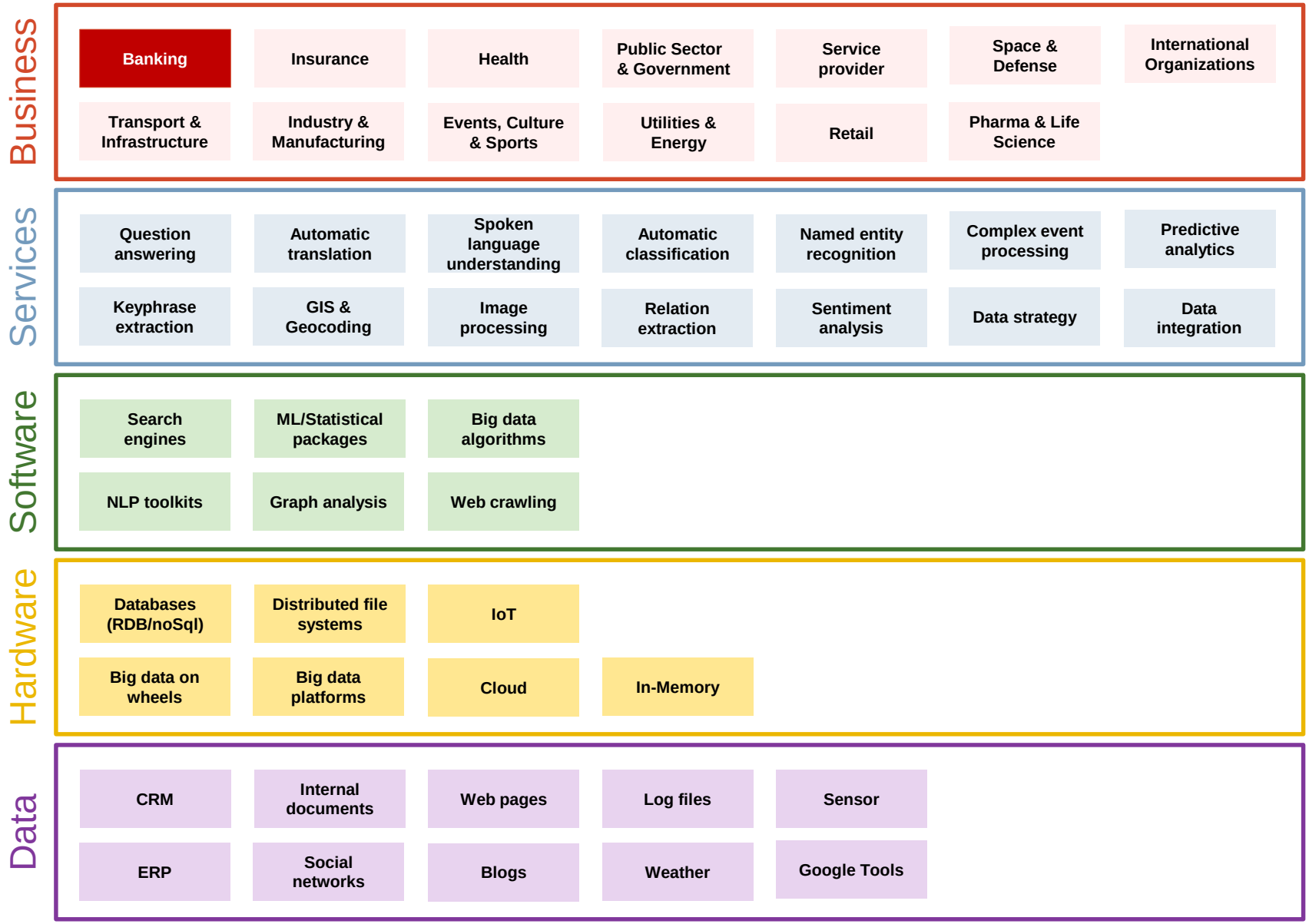
ENTITY SEARCH

Reduce manual checking of invoices, contracts, bills, certificates

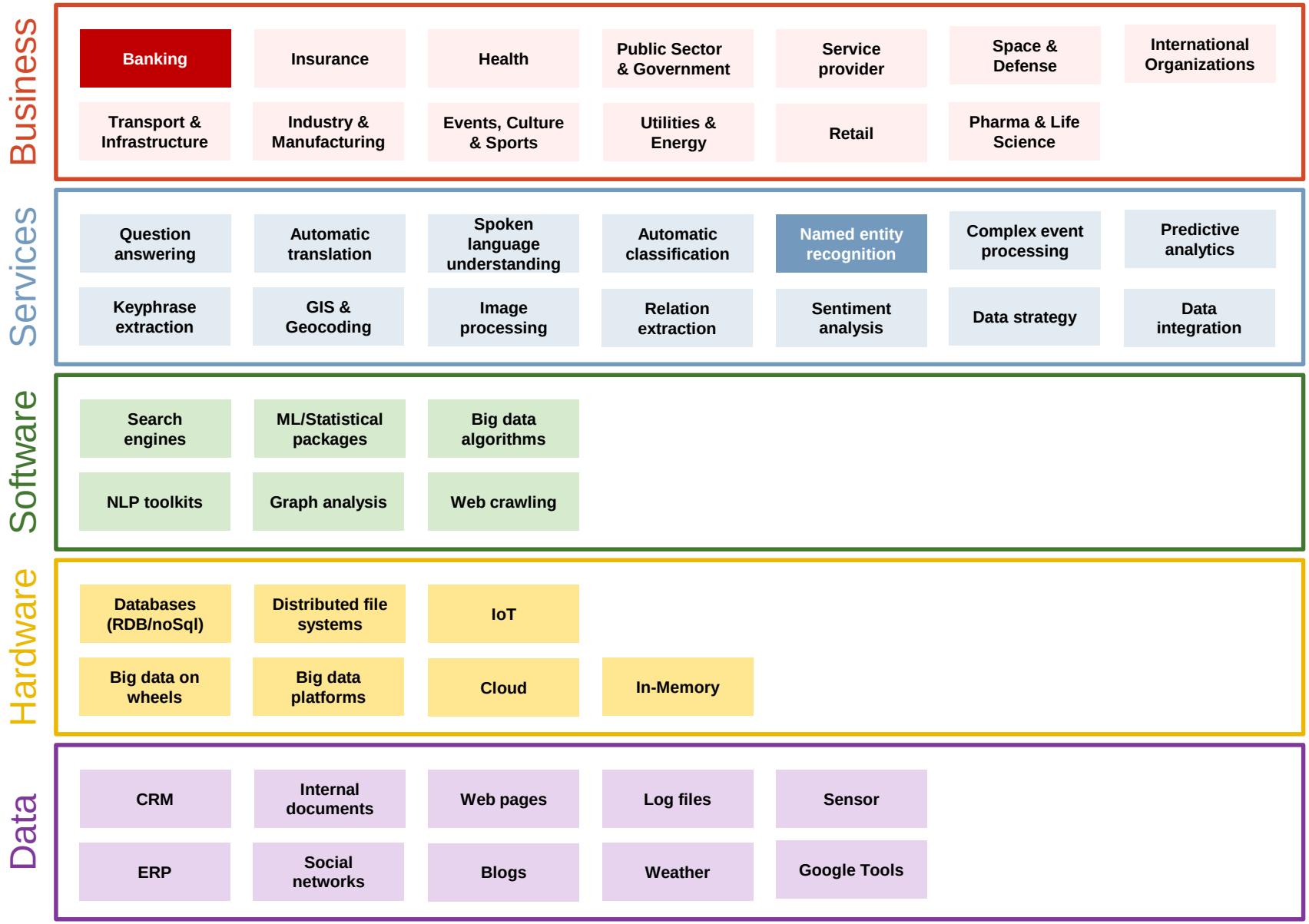
* No contractual document



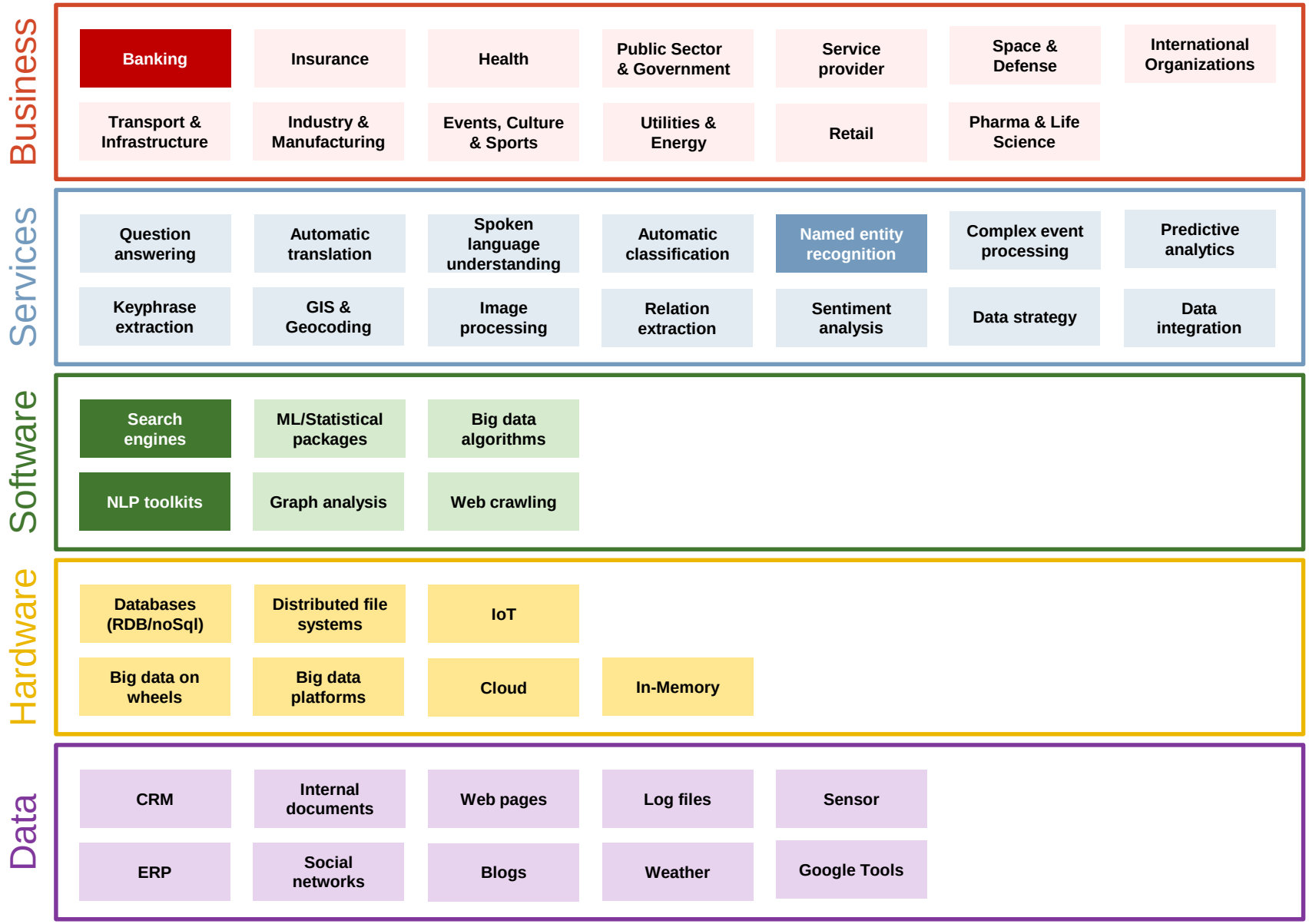
ENTITY SEARCH



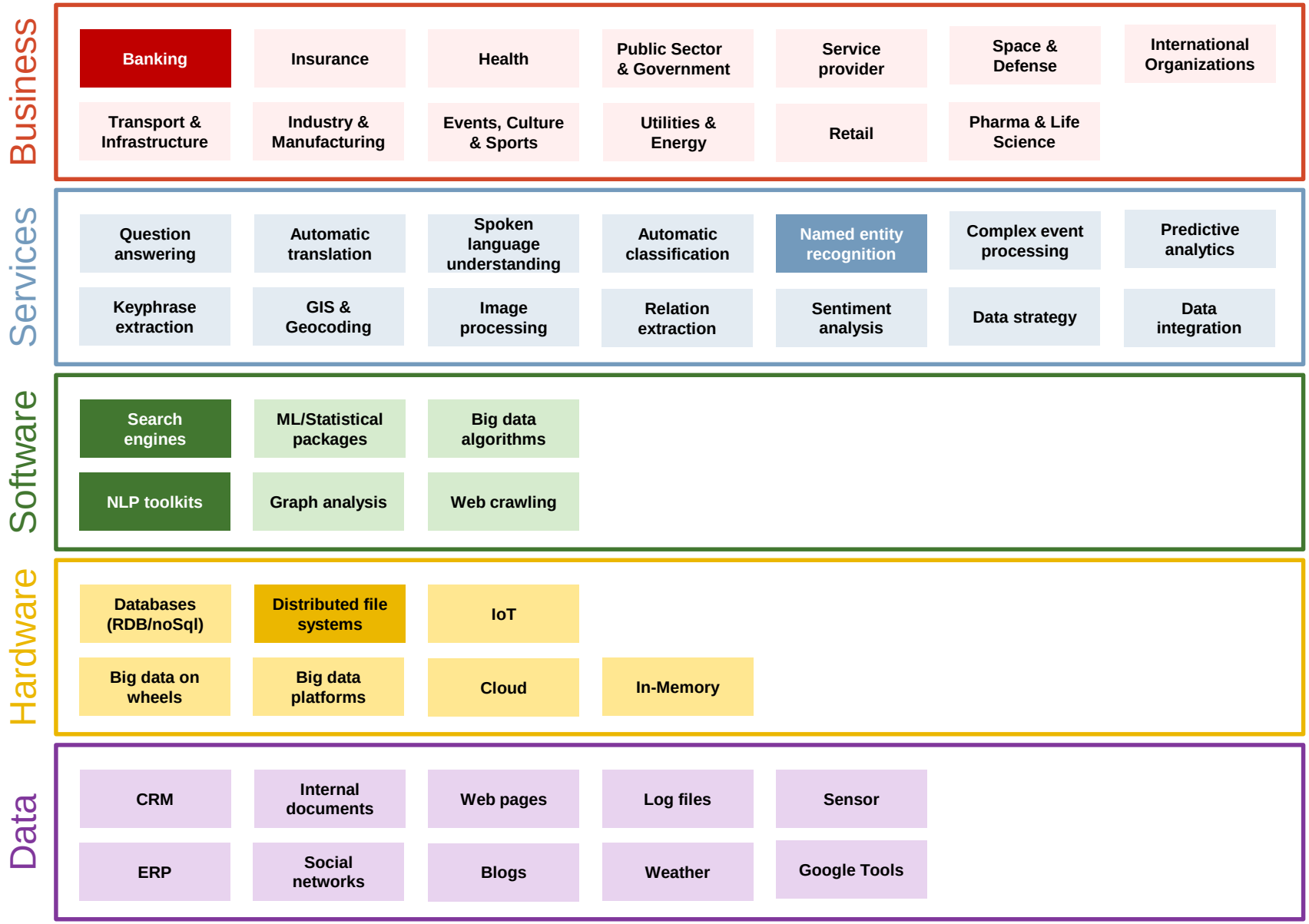
ENTITY SEARCH



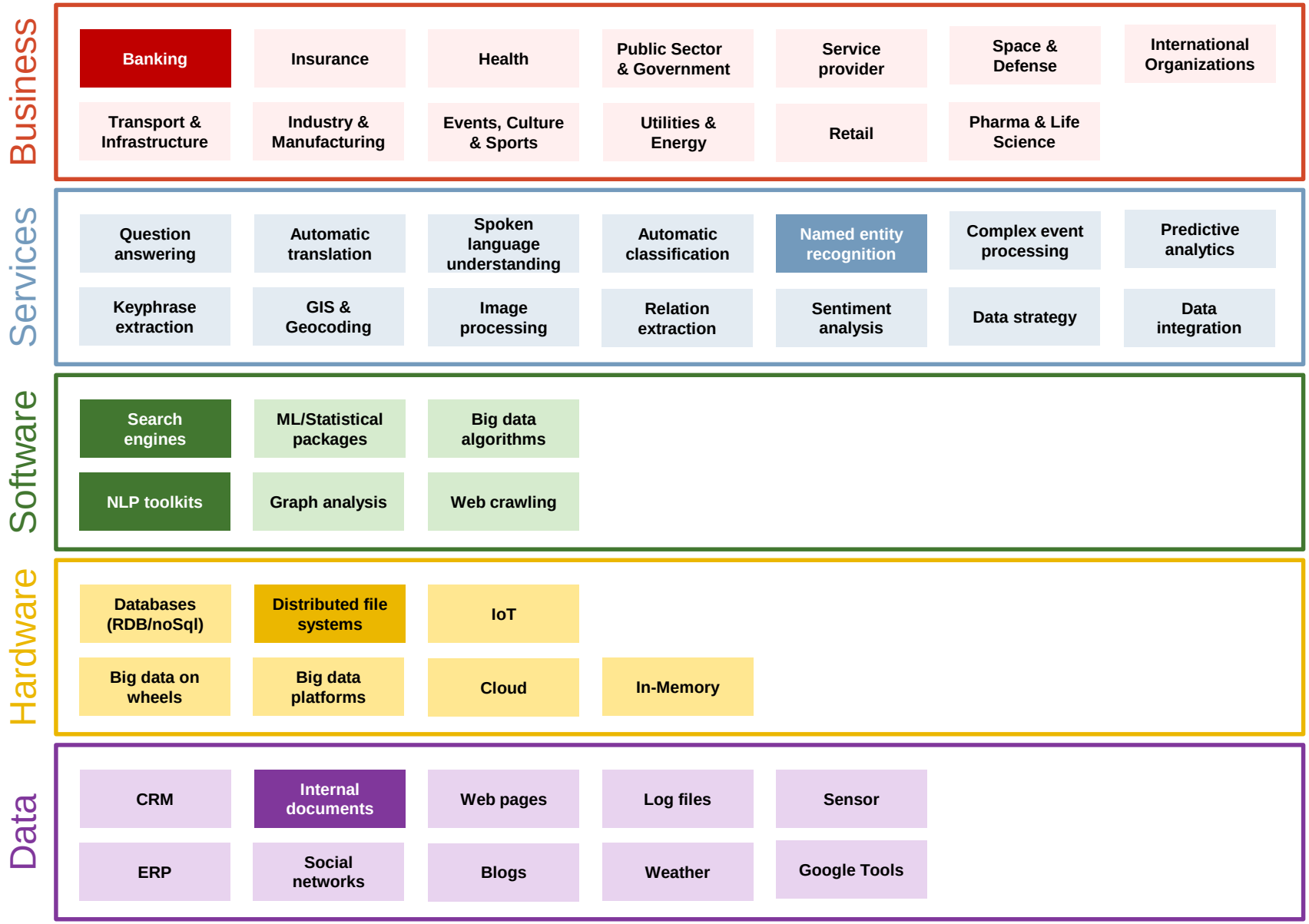
ENTITY SEARCH



ENTITY SEARCH

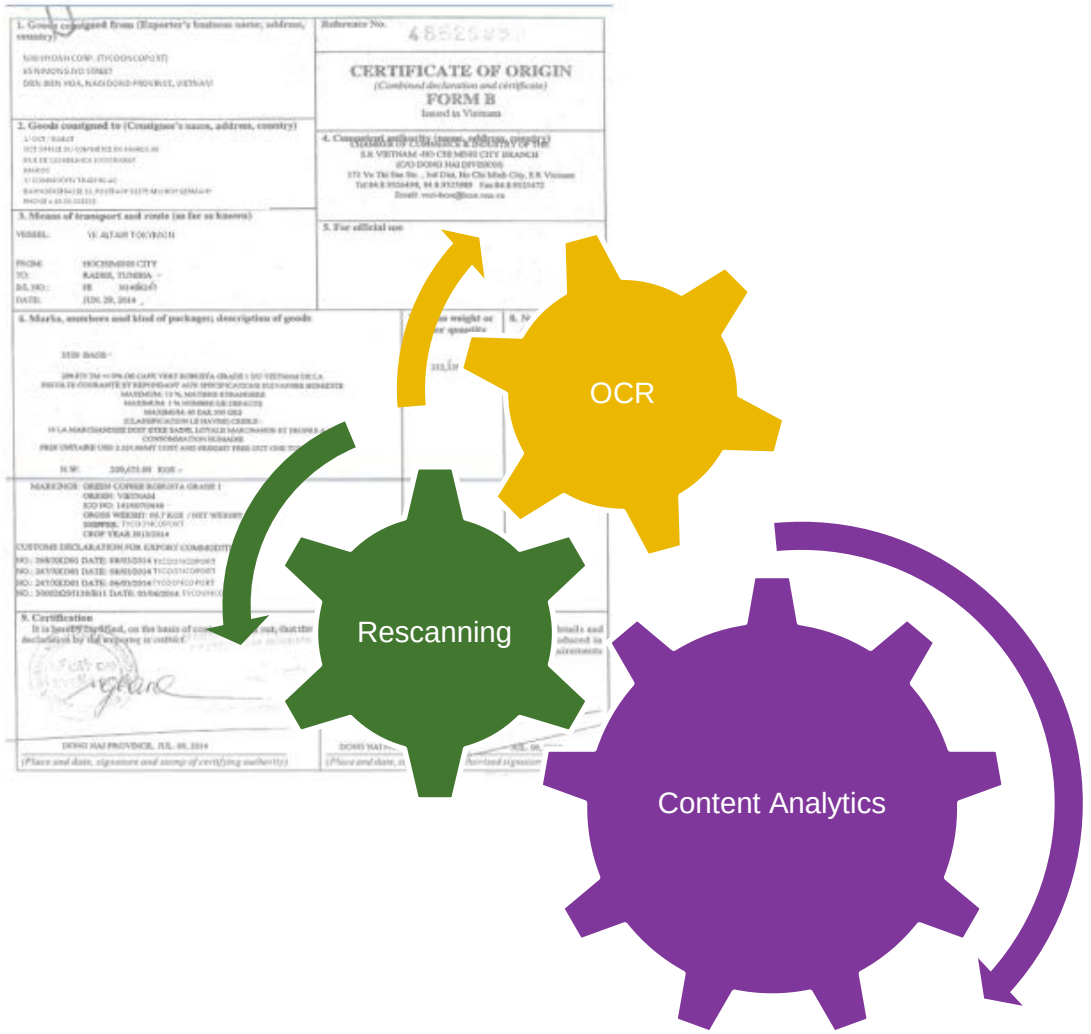


ENTITY SEARCH



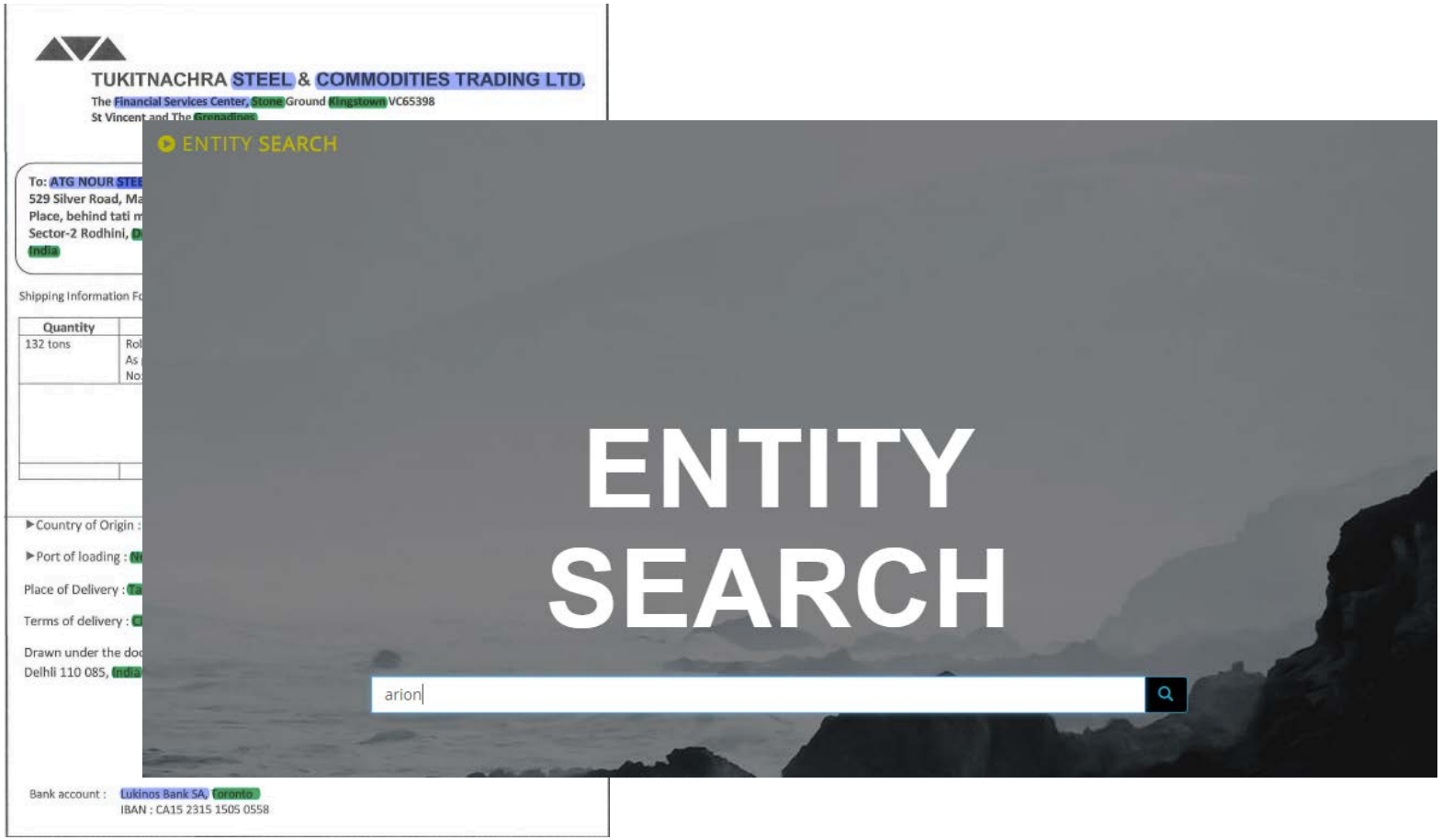
ENTITY SEARCH

From OCR to conformity checks



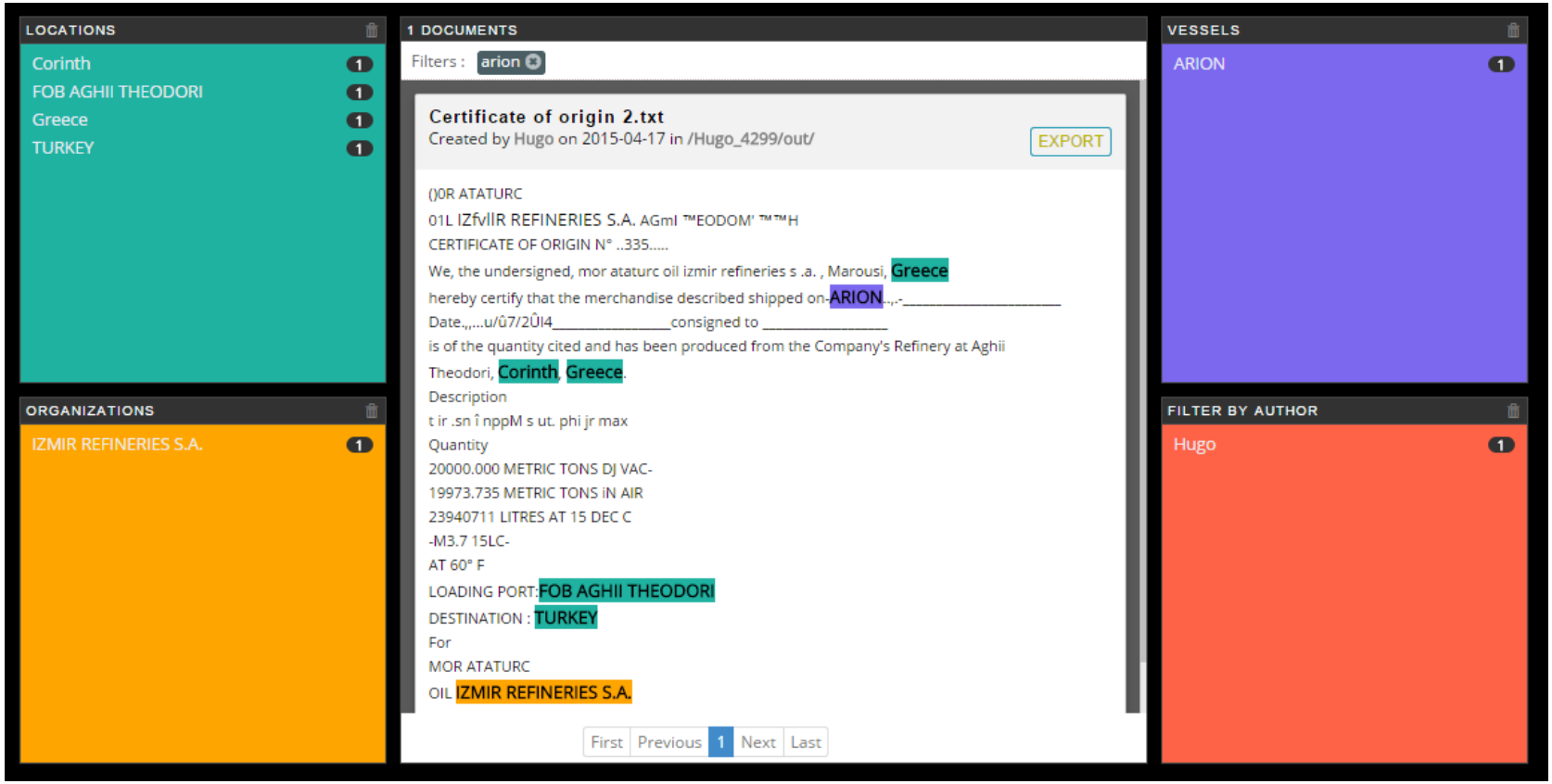
ENTITY SEARCH

From OCR to conformity checks



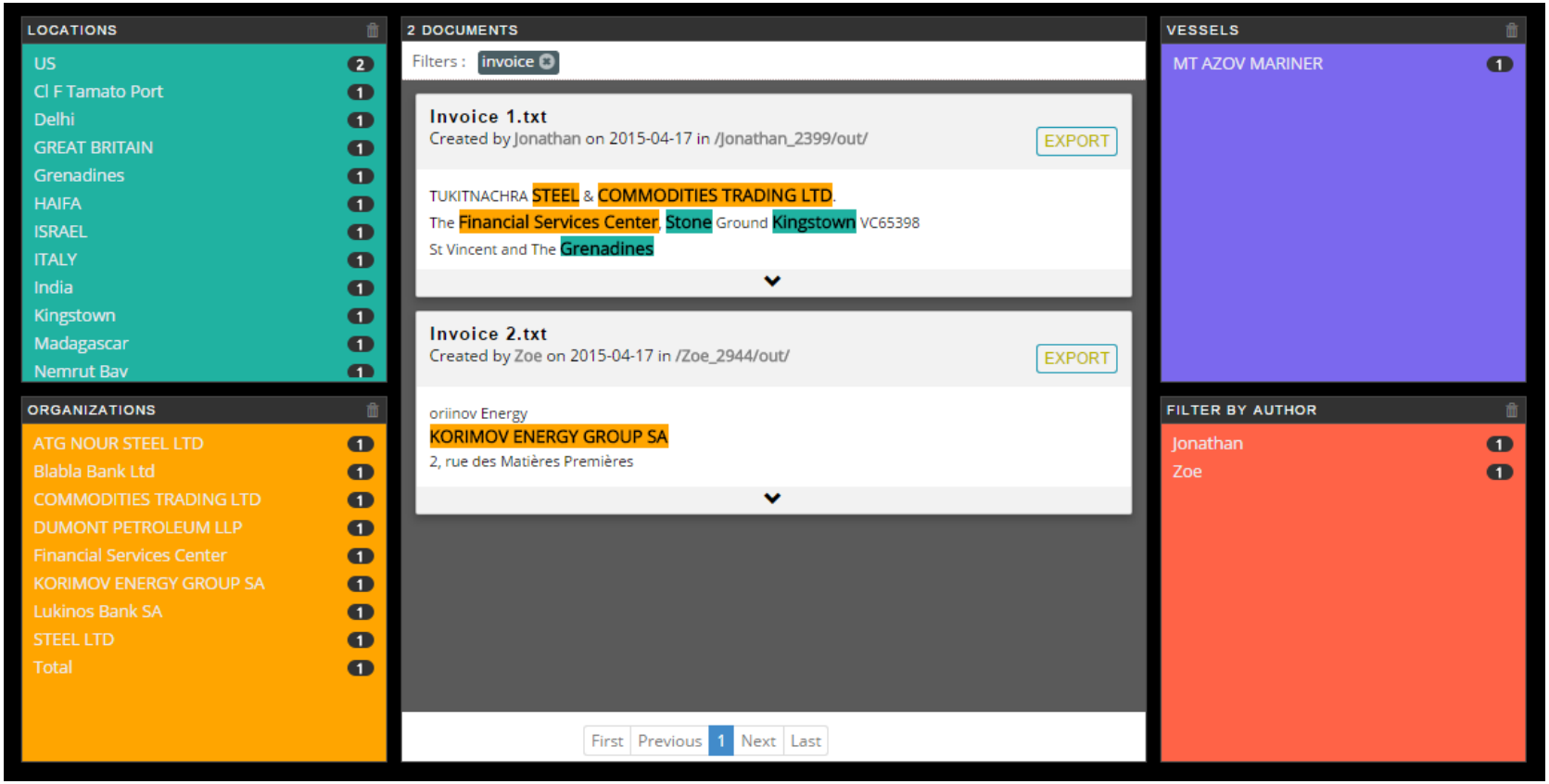
ENTITY SEARCH

From OCR to conformity checks



ENTITY SEARCH

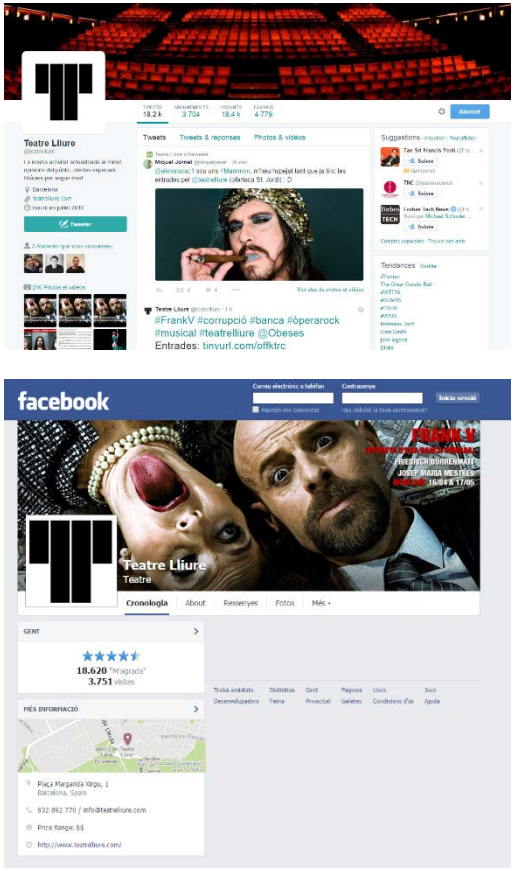
From OCR to conformity checks



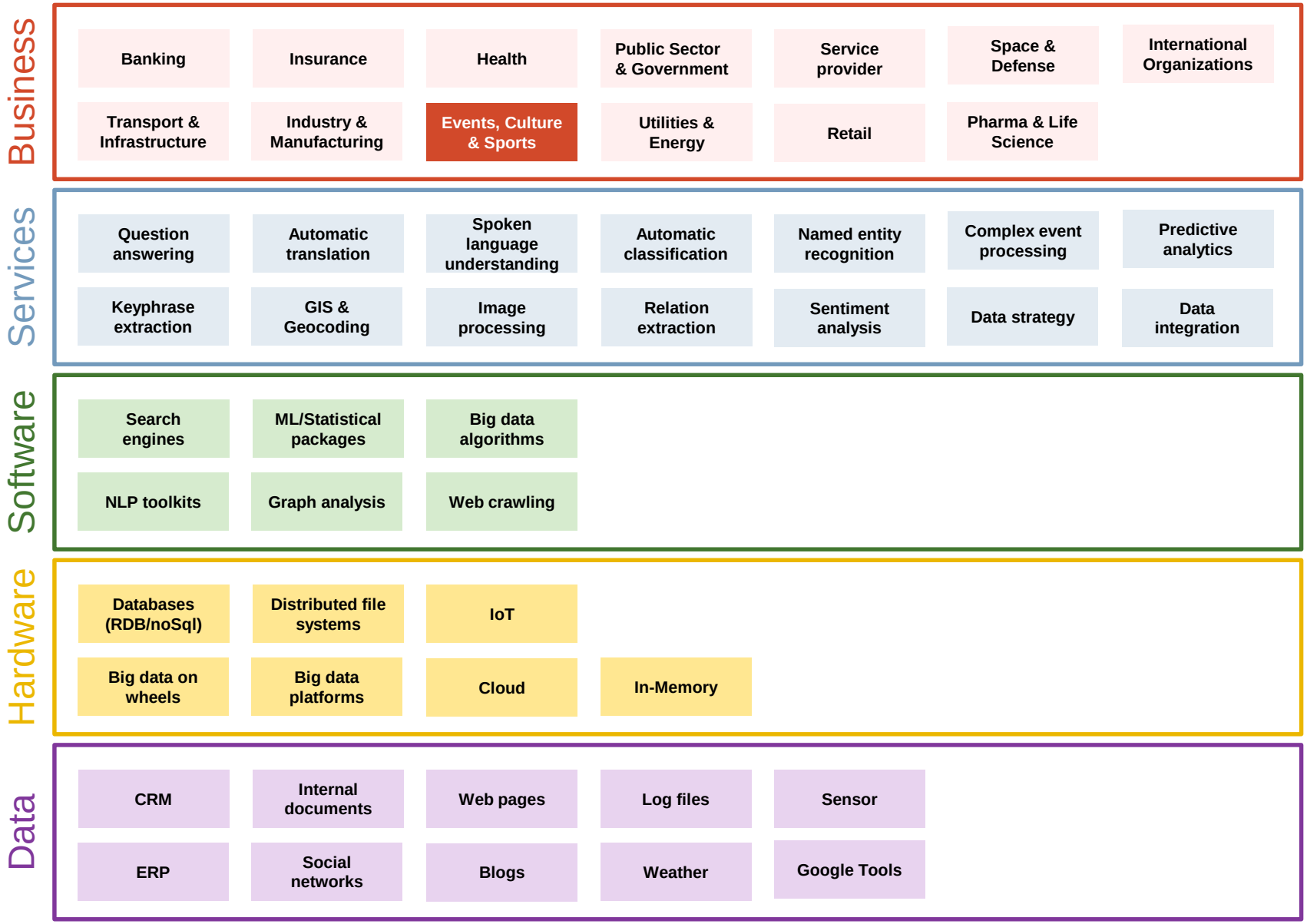
TWITTER FOR SALES

How well do we know our customers on social networks?

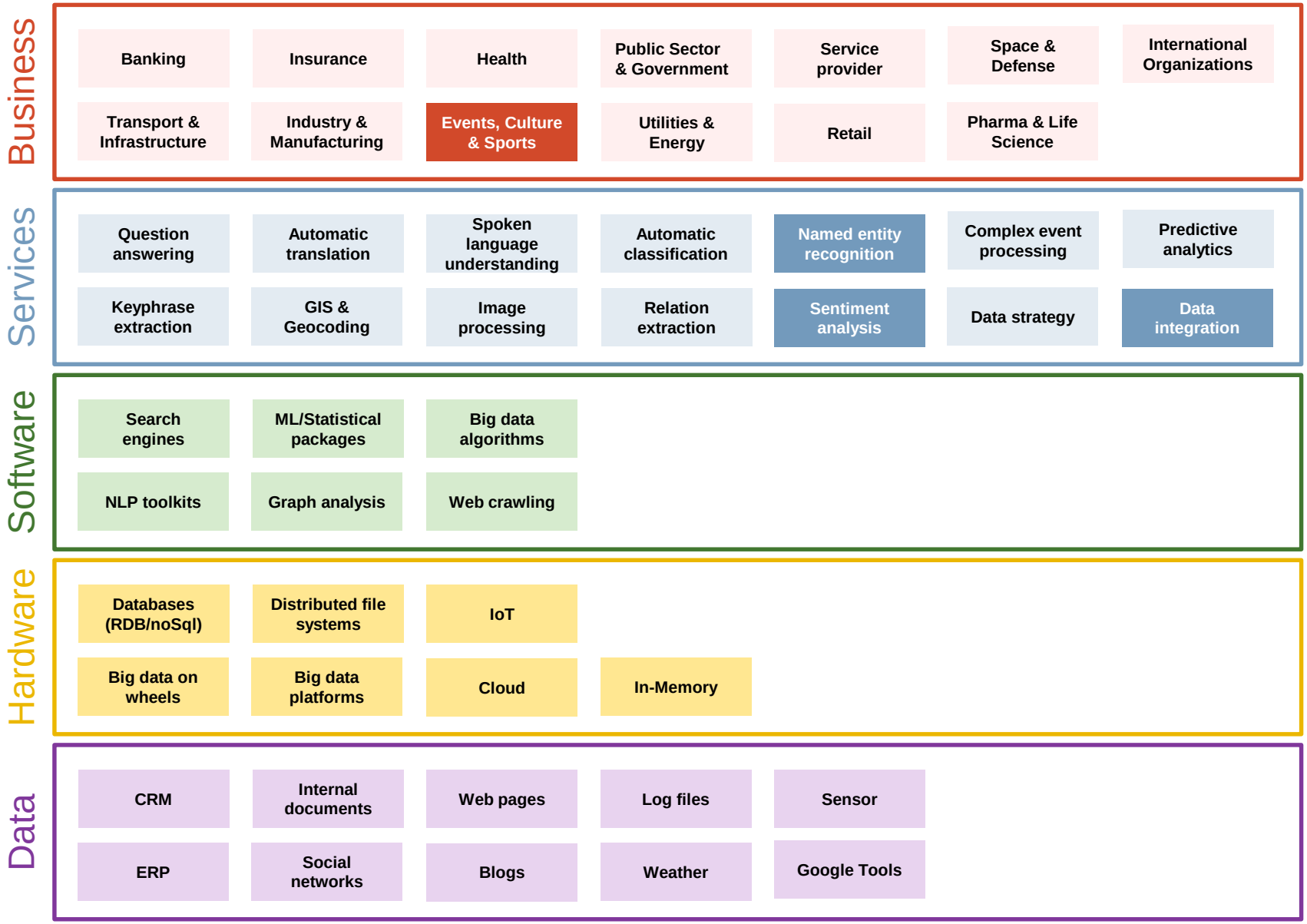
Solution ?



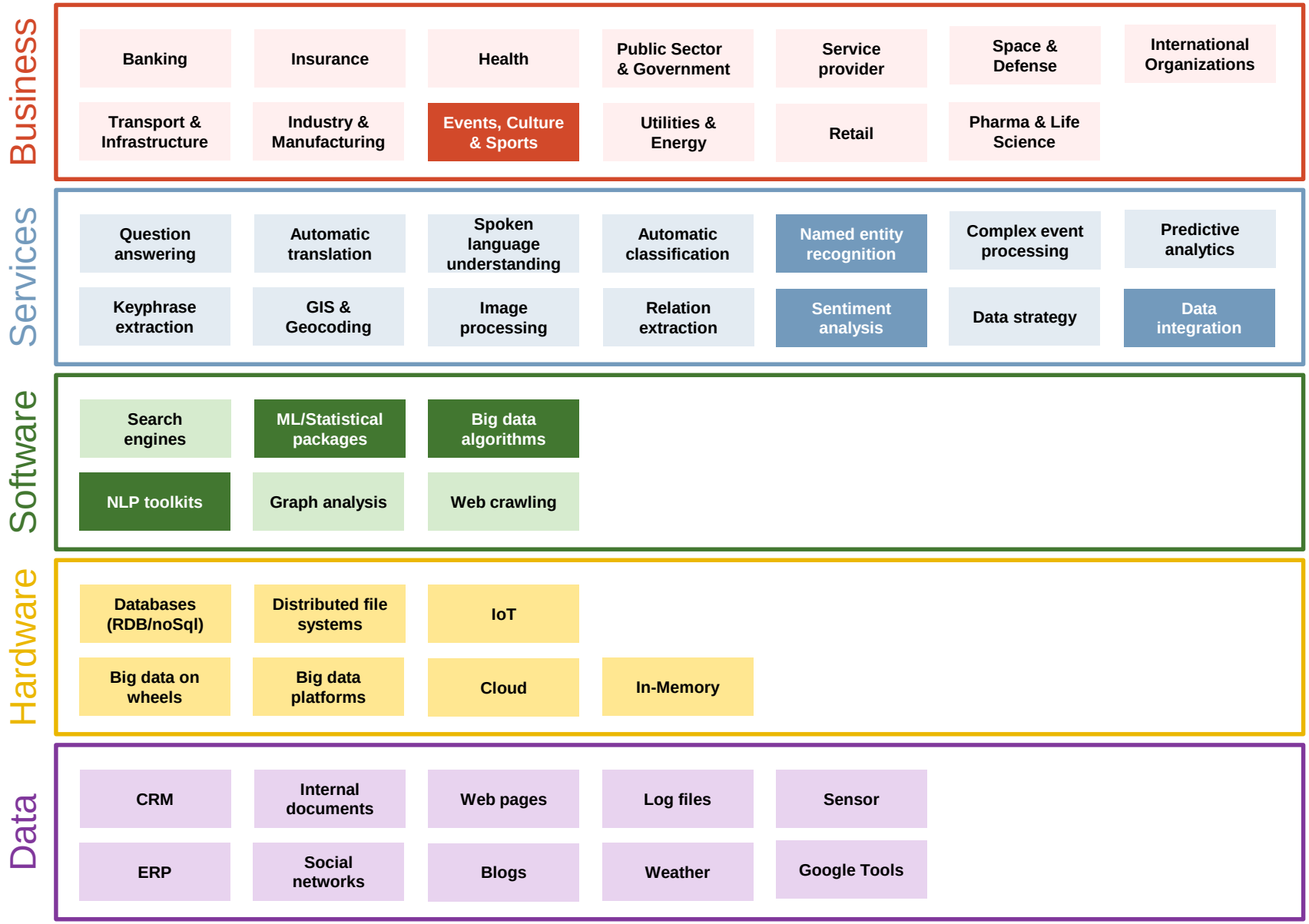
TWITTER FOR SALES



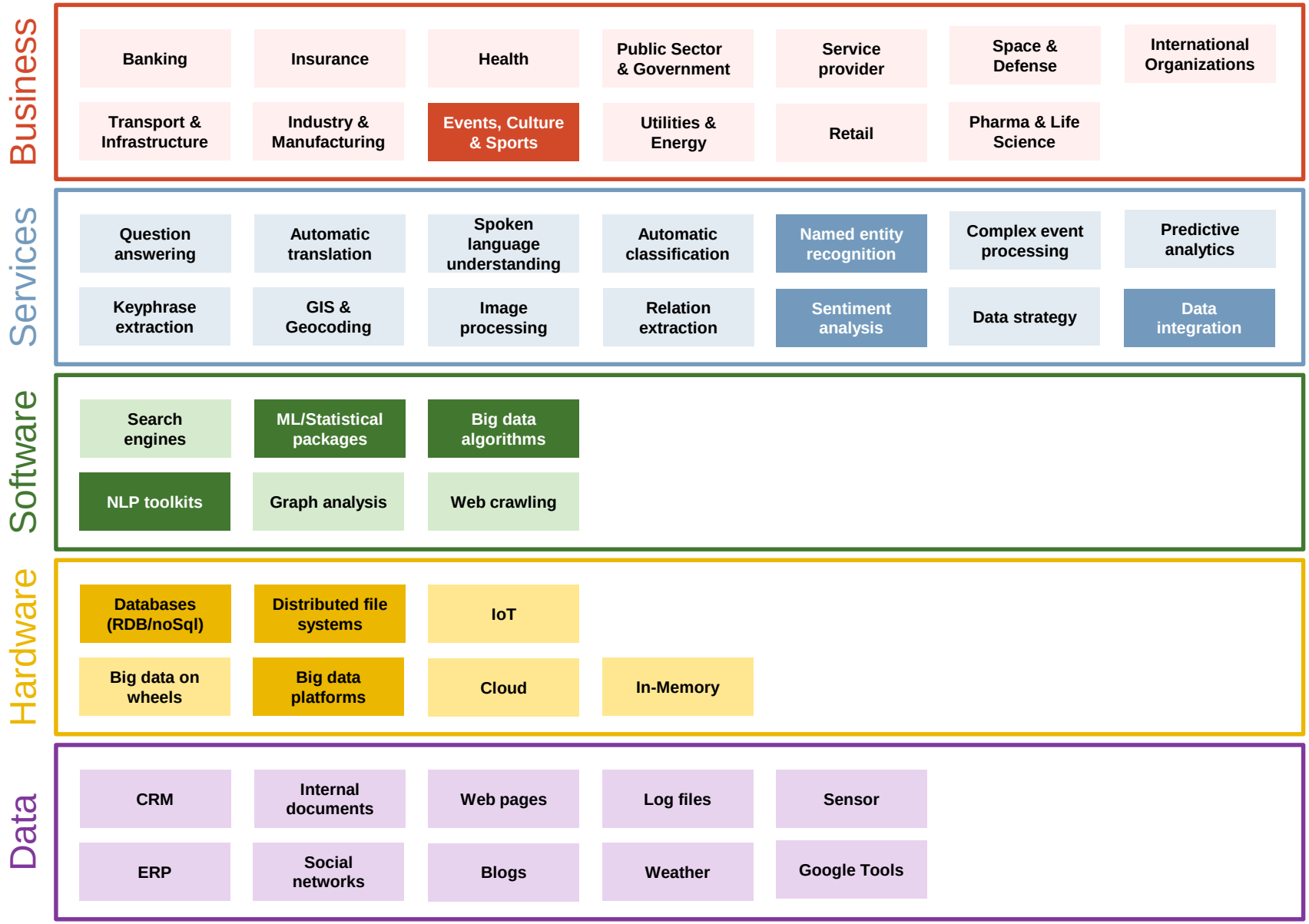
TWITTER FOR SALES



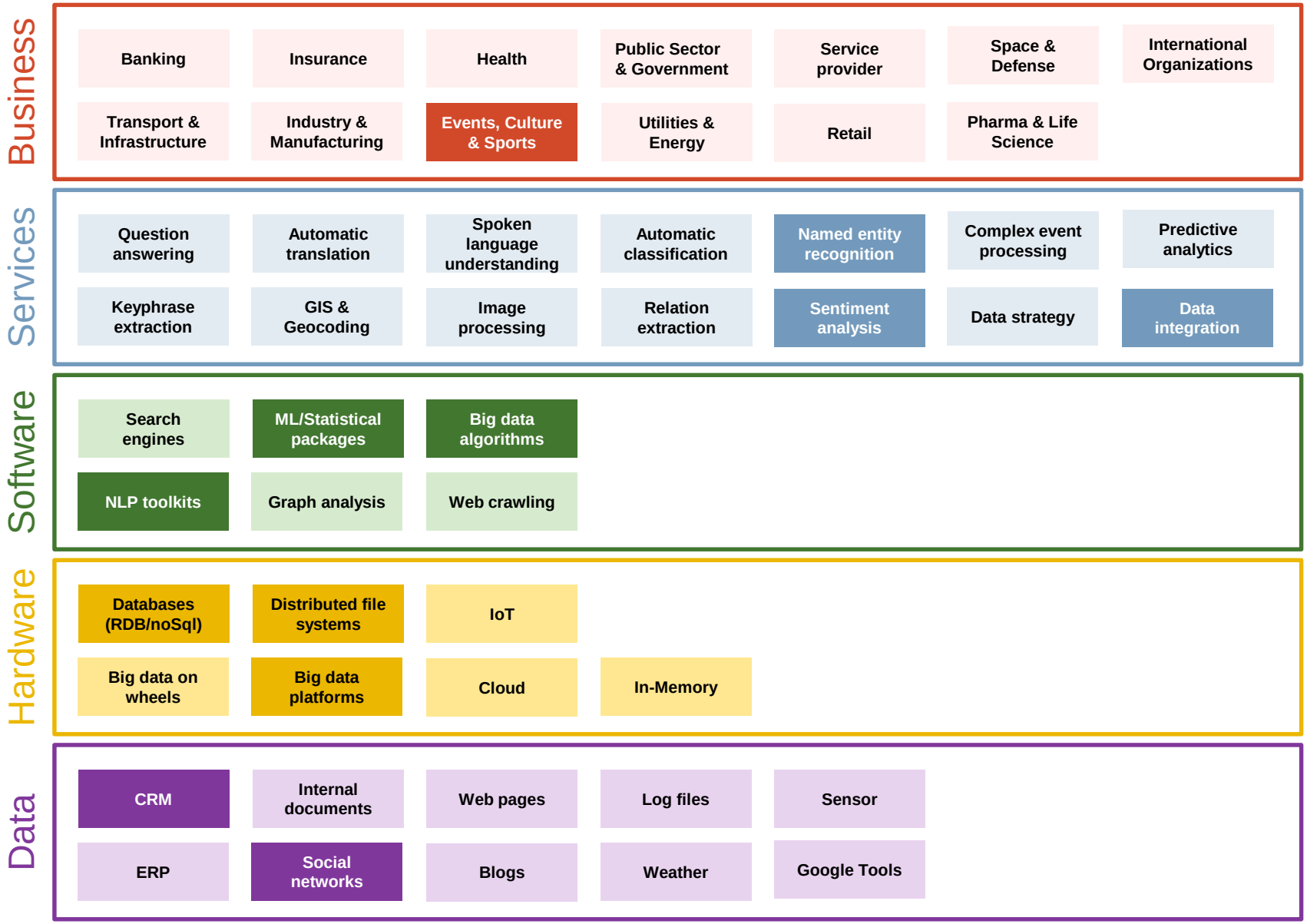
TWITTER FOR SALES



TWITTER FOR SALES



TWITTER FOR SALES



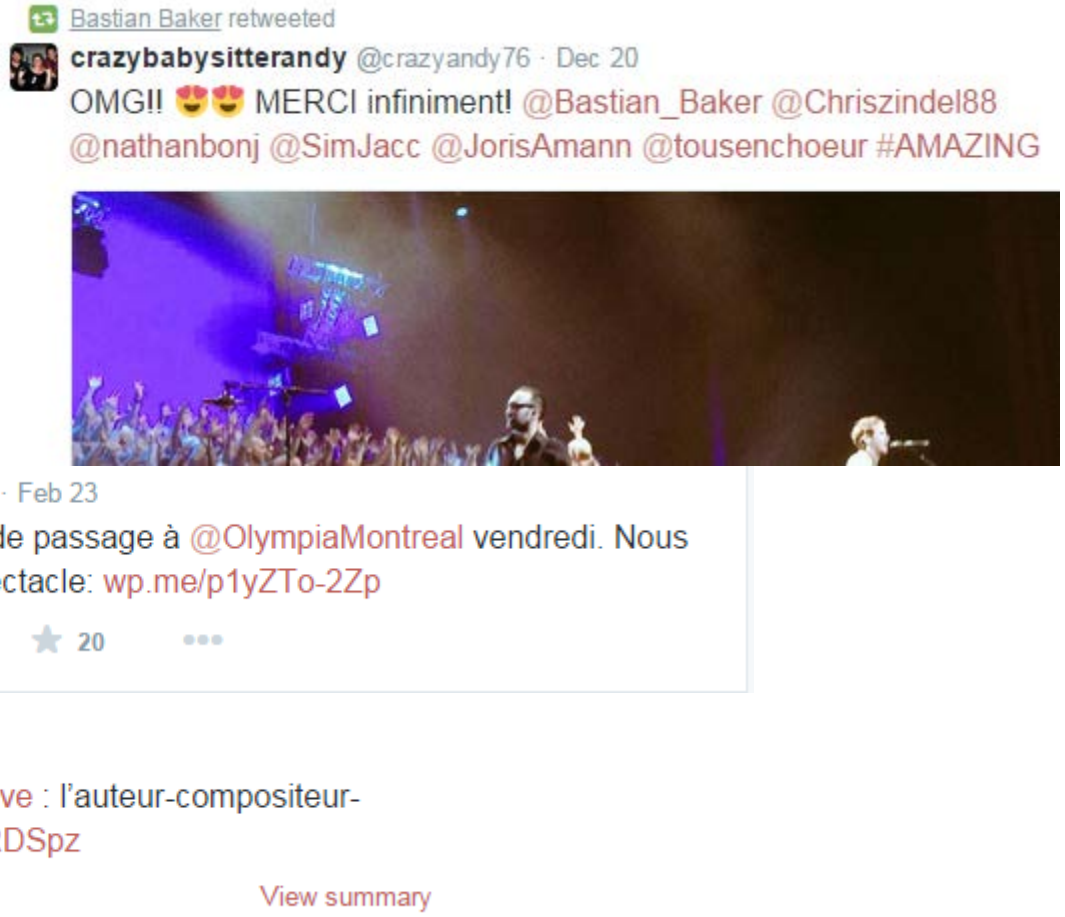
TWITTER FOR SALES

NLP and social networks

Sentiment analysis

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

Identification of **dialog act**



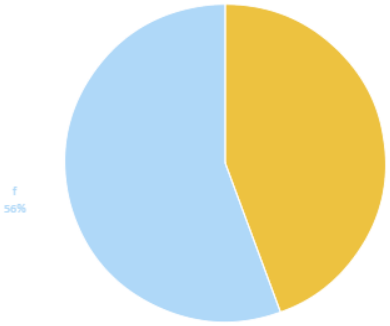
TWITTER FOR SALES

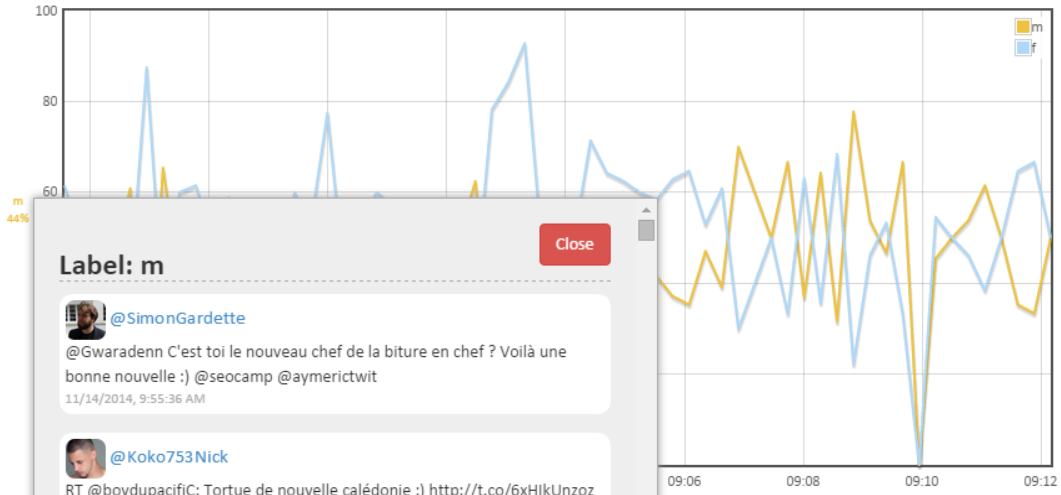
Automatic prediction of age and gender on Twitter

Welcome to the Classification Machine

Analytics result

Genders





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 @boydupacifiC: Tortue de nouvelle calédonie :) http://t.co/6xHIkUnzoz

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

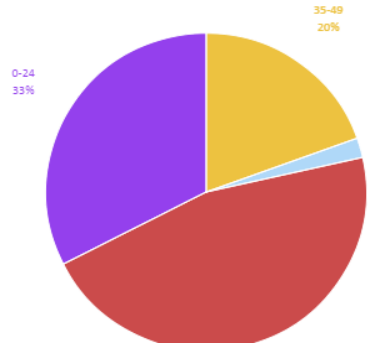
@_not_101

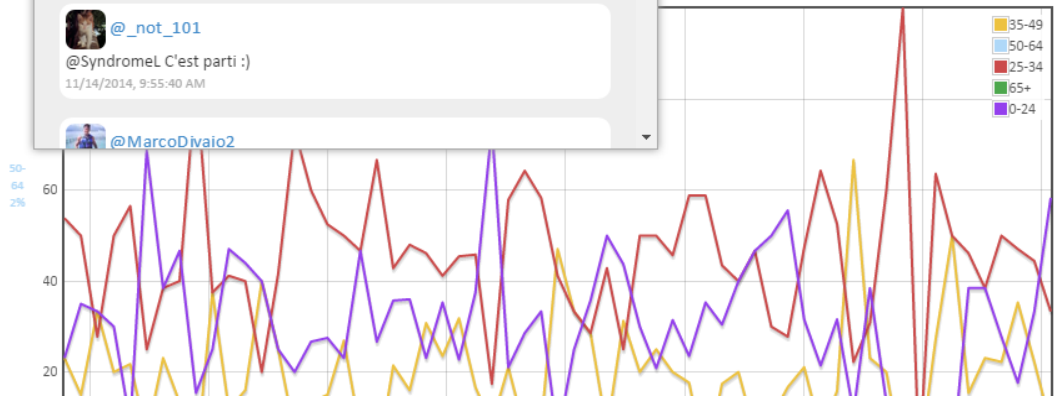
@SyndromeL C'est parti :)

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

@MarcoDivaio2

Ages

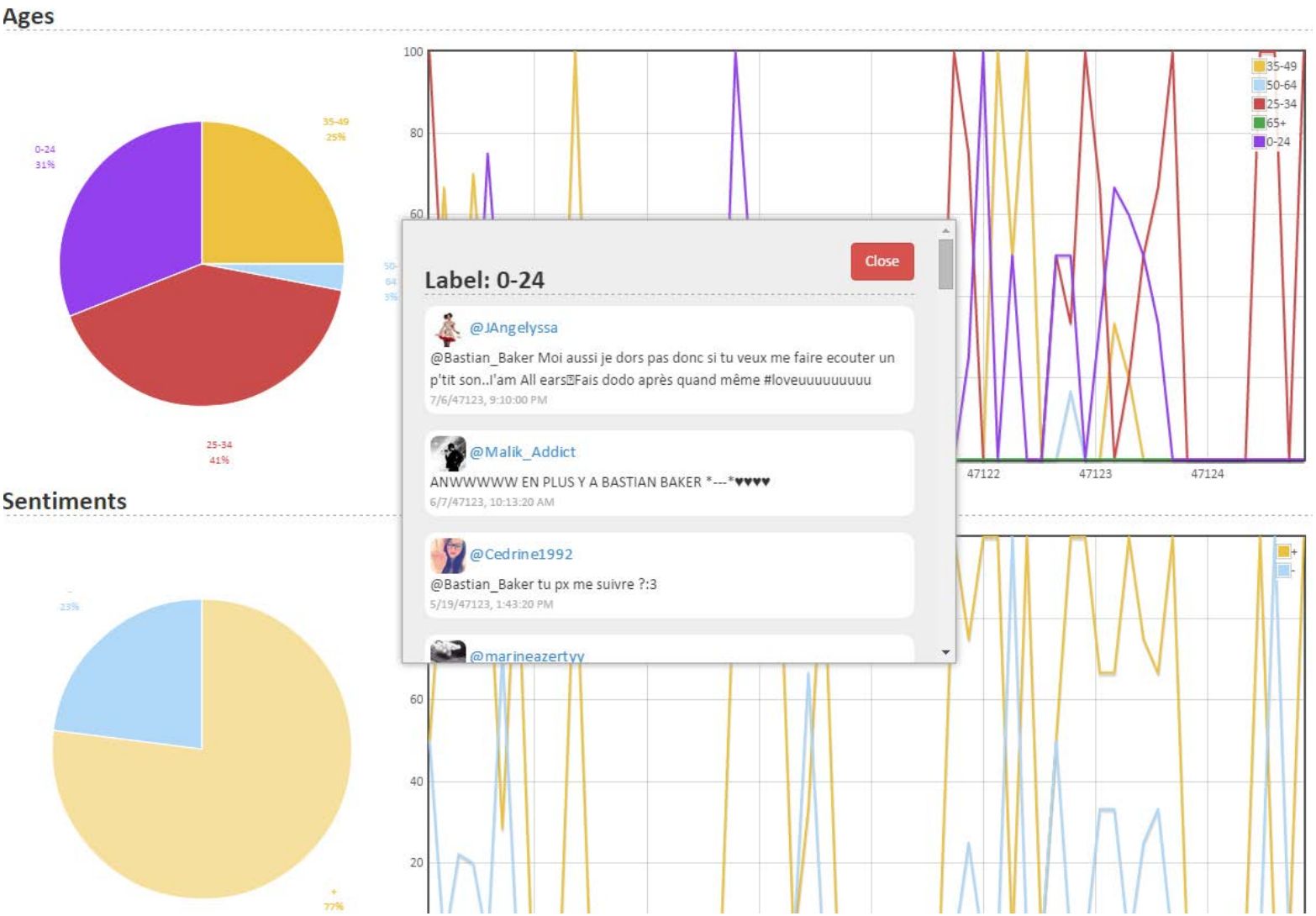




Filters

TWITTER FOR SALES

Age and sentiment of tweets about Bastian Baker



TWITTER FOR SALES

User profiling on Twitter: “Bastian Baker”

Most active users

Show 10 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%)
hxranlik	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

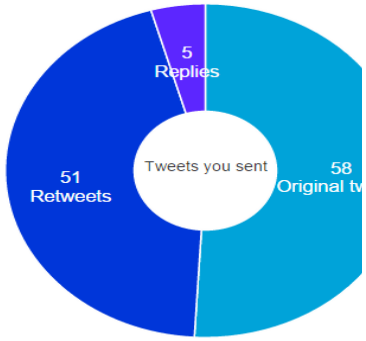
TWITTER FOR SALES

The Secutix monthly report

Your social activity

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

74 retweets (+722%) have been create
that generated **53 900 impressions (+1**



Your top 10 ambassadors

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

new — 27 063 followers ● 3 relevant tweets
Ànima és el magazin d'actualitat escènica del @canal_33. Presentat i dirigit per @tonipunti, s'emet els dilluns a les 22:45.

+1 — 3 107 followers ● 8 relevant tweets
Journalist. Social Media Strategist. Theater

new — 16 814 followers ● 2 relevant tweets
HOMBRE LOBO

new — 9 977 followers ● 2 relevant tweets
Festival de Tardor de Catalunya · Girona Sa

new — 1 251 followers ● 5 relevant tweets
«Just be you, but learn from other people» London??-

new — 5 093 followers ● 2 relevant tweets
?'If you can imagine, you can get'? Director @BancoSabadell

new — 20 361 followers ● 1 relevant tweets
Departament de Cultura. Generalitat de Ca Government of Catalonia.

new — 746 followers ● 5 relevant tweets
teatre i més encara

-5 — 15 613 followers ● 1 relevant tweets
El digital de cultura. Busquem el que és únic entre el que ens és comú.

new — 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://t.co/P7k6uqcnsR

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 guia defini

The 3 most retweeted tweets about you

Ànima és el magazin d'actualitat escènica del @canal_33. Presentat i dirigit per @tonipunti, 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 #Ànima33, coneixem a fons a l'actor Pol López.

Ànima és el magazin d'actualitat escènica del @canal_33. Presentat i dirigit per @tonipunti, s'emet els dilluns a les 22:45.

28 sept. 2014 ↻ 8
Volem seguir dient #ensanima amb @teatrenacional, @teatrelliure, @mercatflors i ara que també comencen @temporadaalta o @bcnjazzfest.

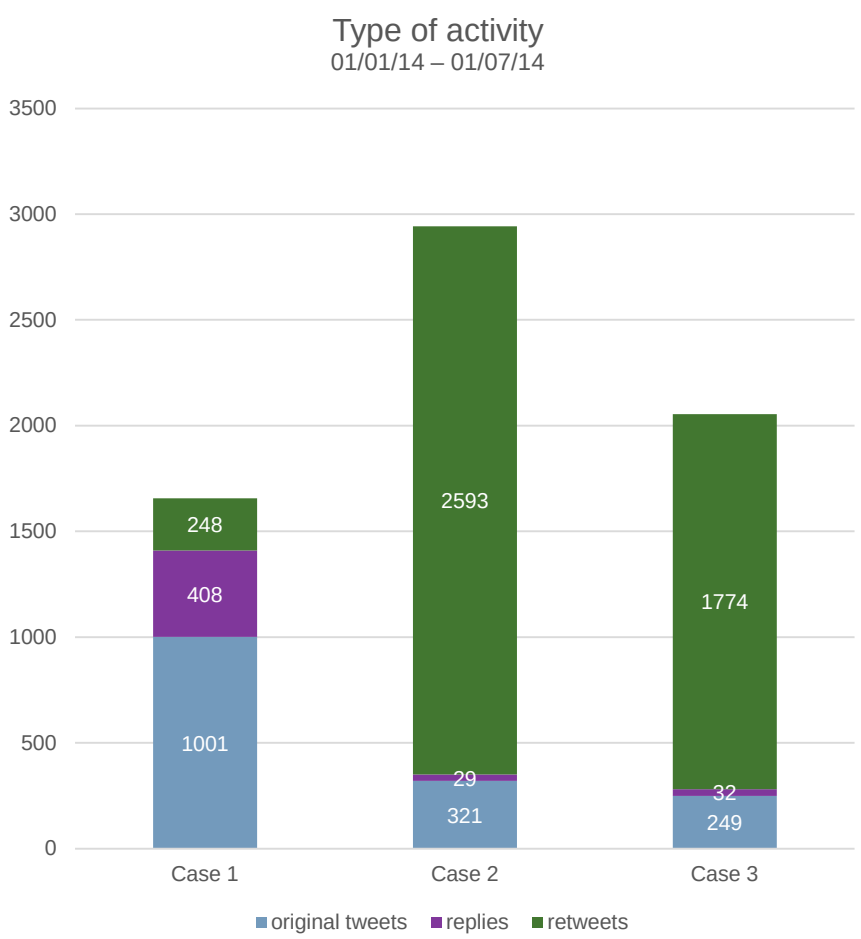
@MiriamIsclaNews

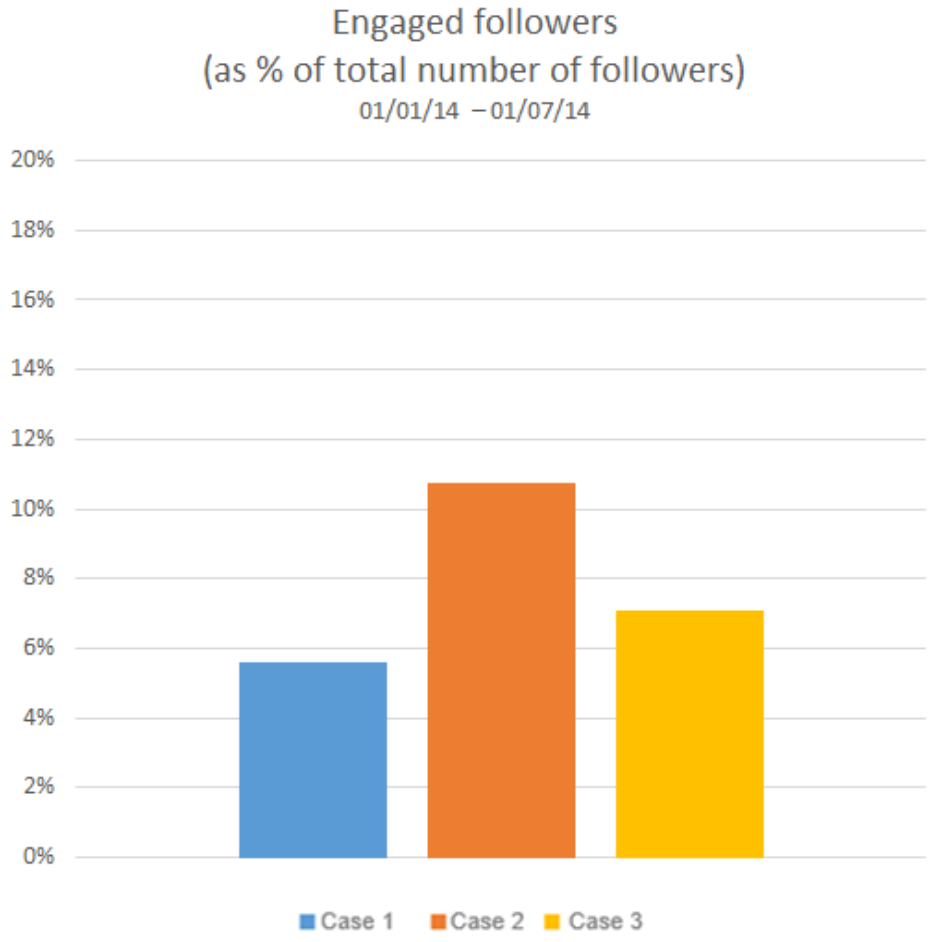
2 sept. 2014 ↻ 6
"Per fi!!!" ???? #annalizaran #salaannalizaran @teatrelliure http://t.co/T12AlcxmMd

TWITTER FOR SALES

What's better, to tweet or to retweet?

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



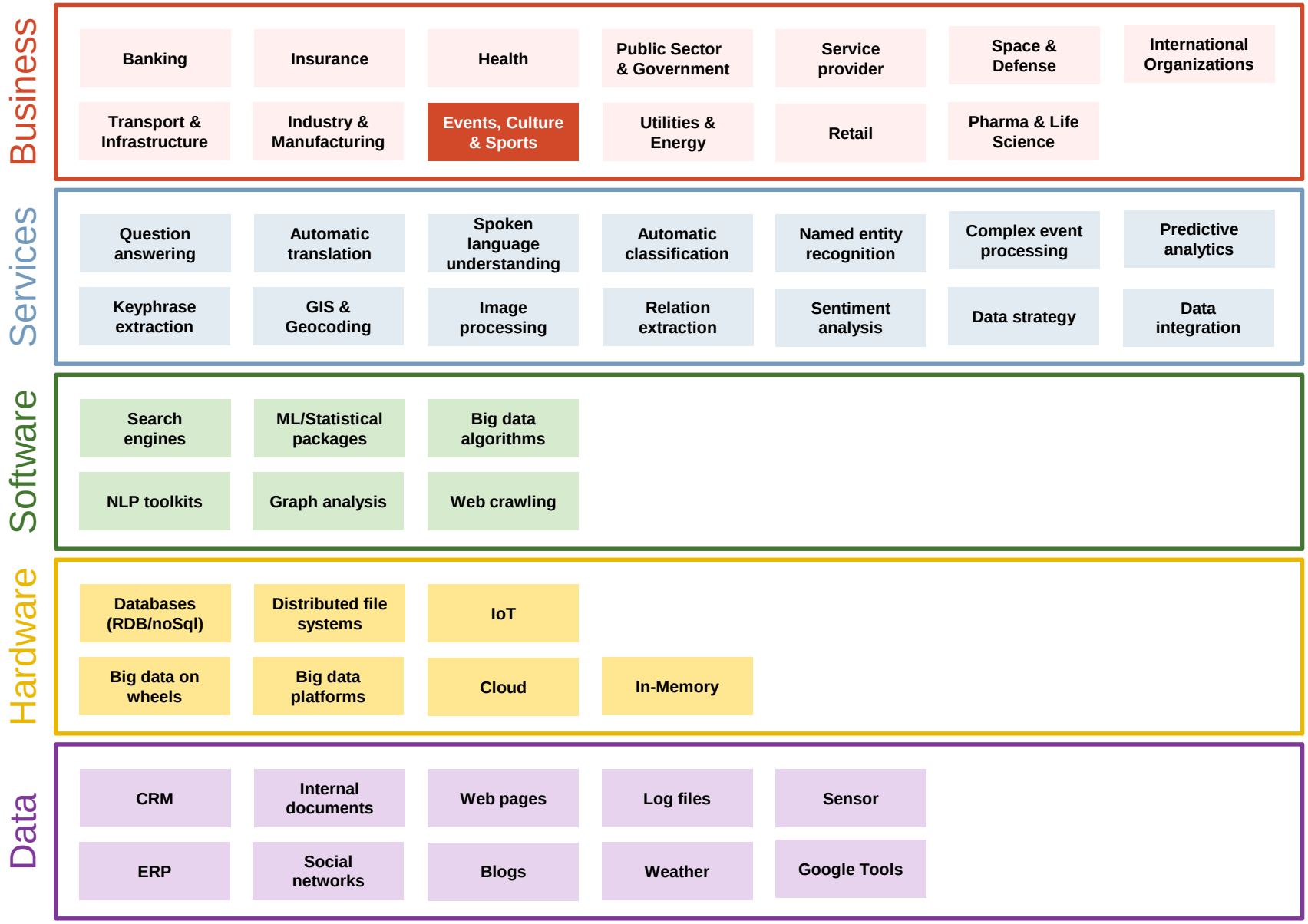


EMARKETING OPTIMIZATION

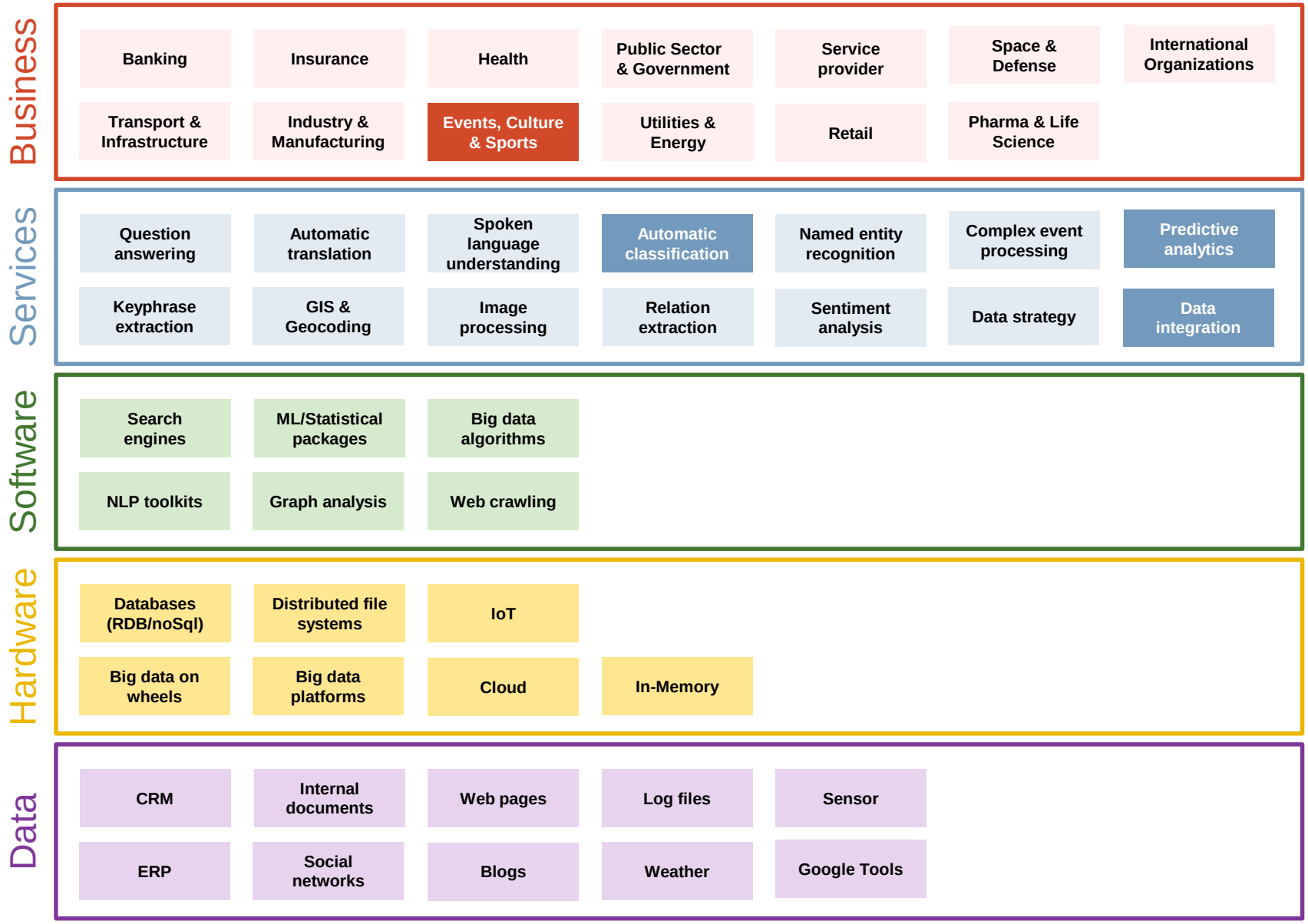
How to get from Spam to Target ?

VS

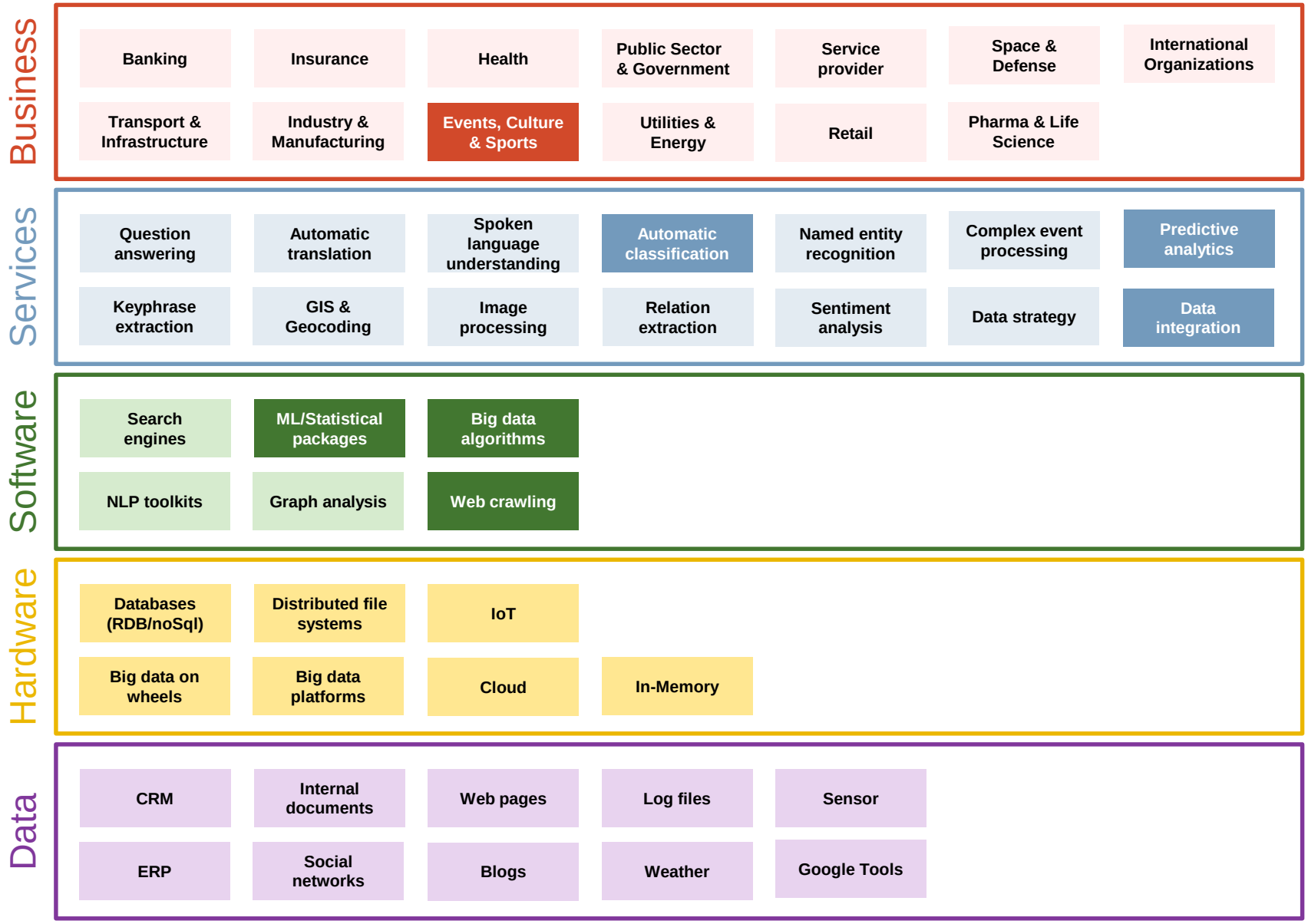
EMARKETING OPTIMIZATION



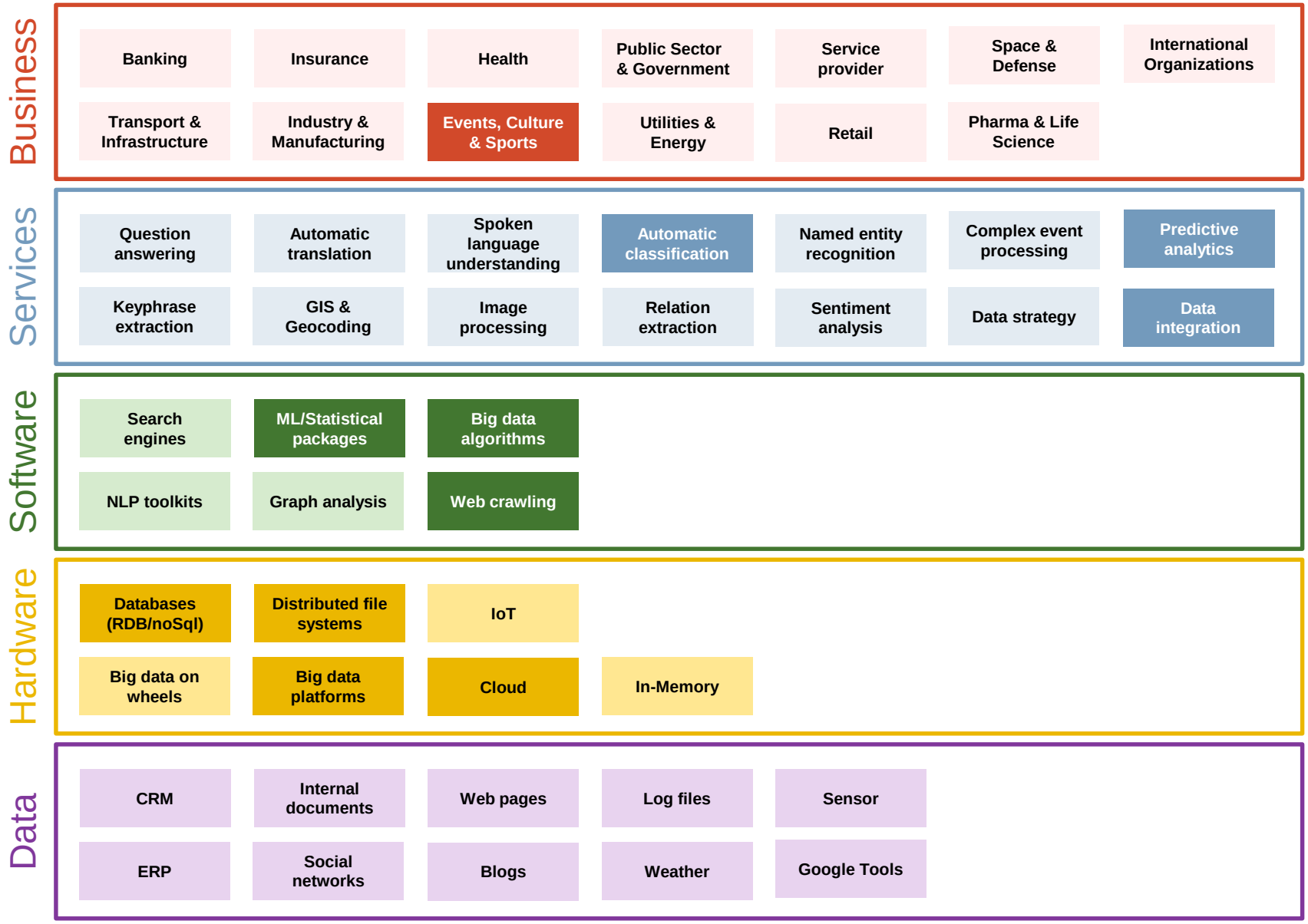
EMARKETING OPTIMIZATION



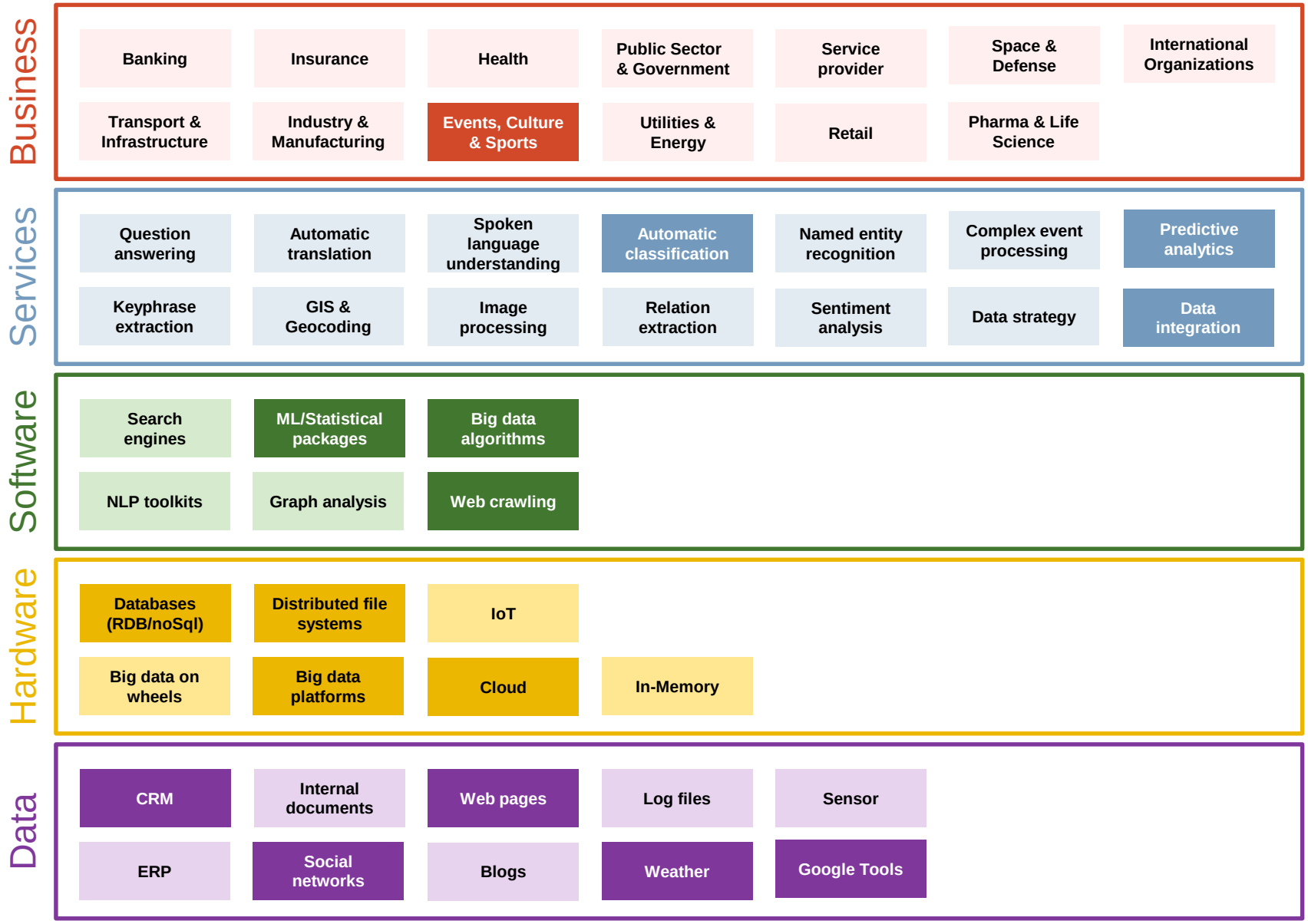
EMARKETING OPTIMIZATION



EMARKETING OPTIMIZATION



EMARKETING OPTIMIZATION



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

EMARKETING OPTIMIZATION

CRM Segmentation

Drill down into the VIP cluster and cluster it further.

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

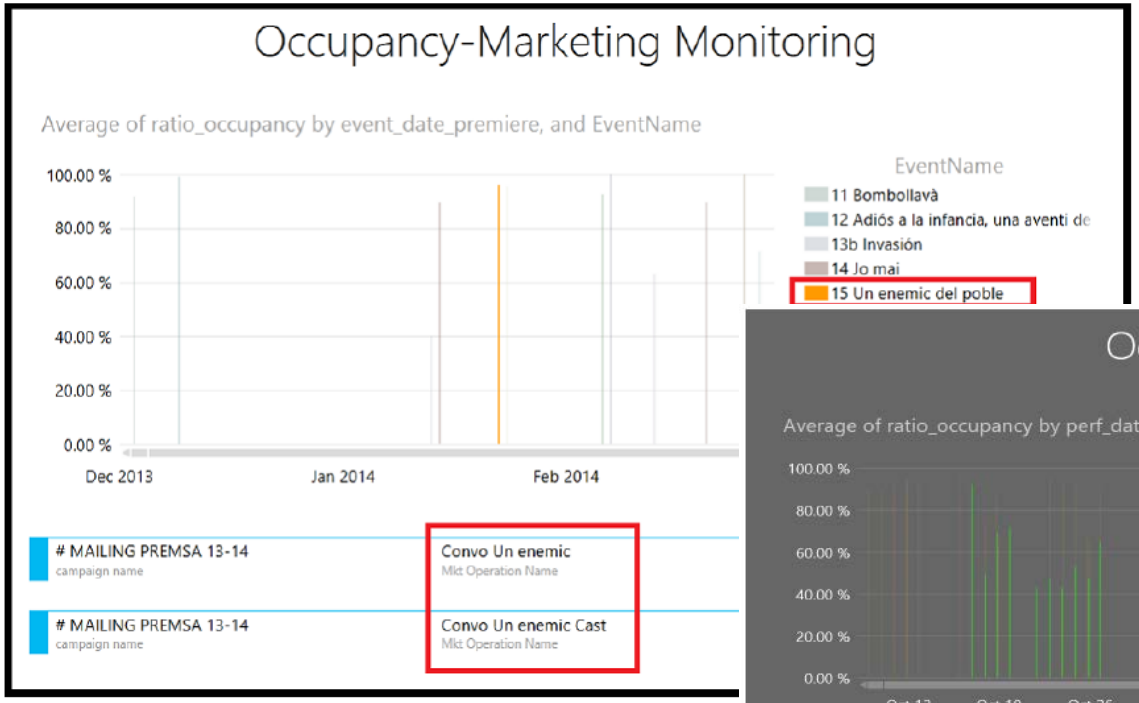
Cluster 4: good target for afternoon shows

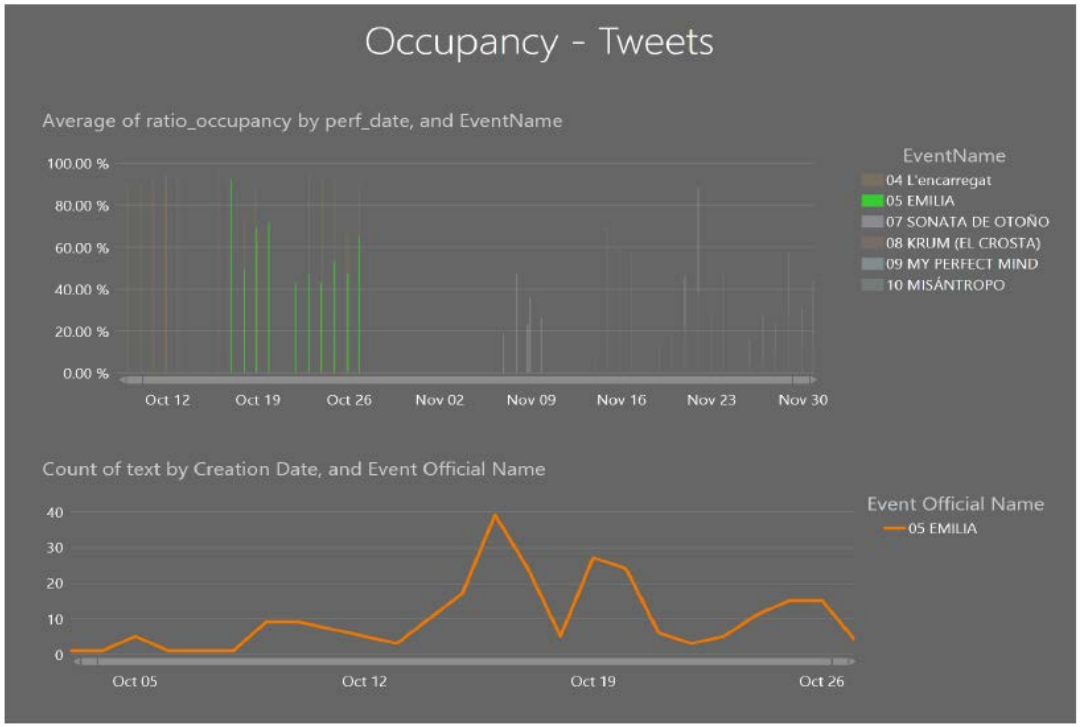
...

EMARKETING OPTIMIZATION

OCCUPANCY ANALYSIS

A marketing tool for monitoring the occupancy of shows

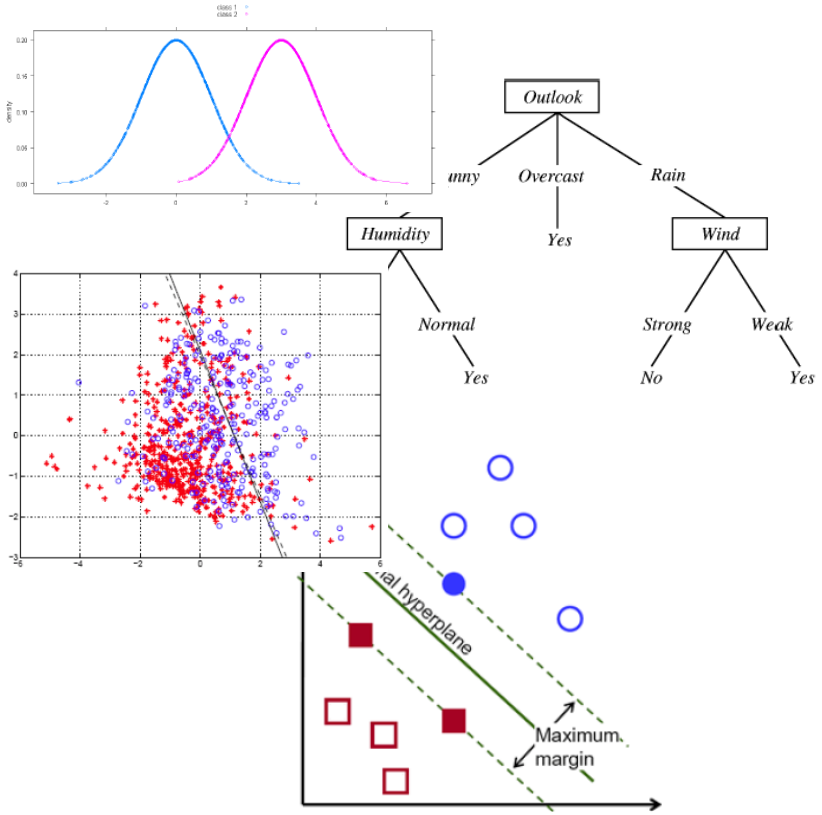




EMARKETING OPTIMIZATION

OCCUPANCY ANALYSIS

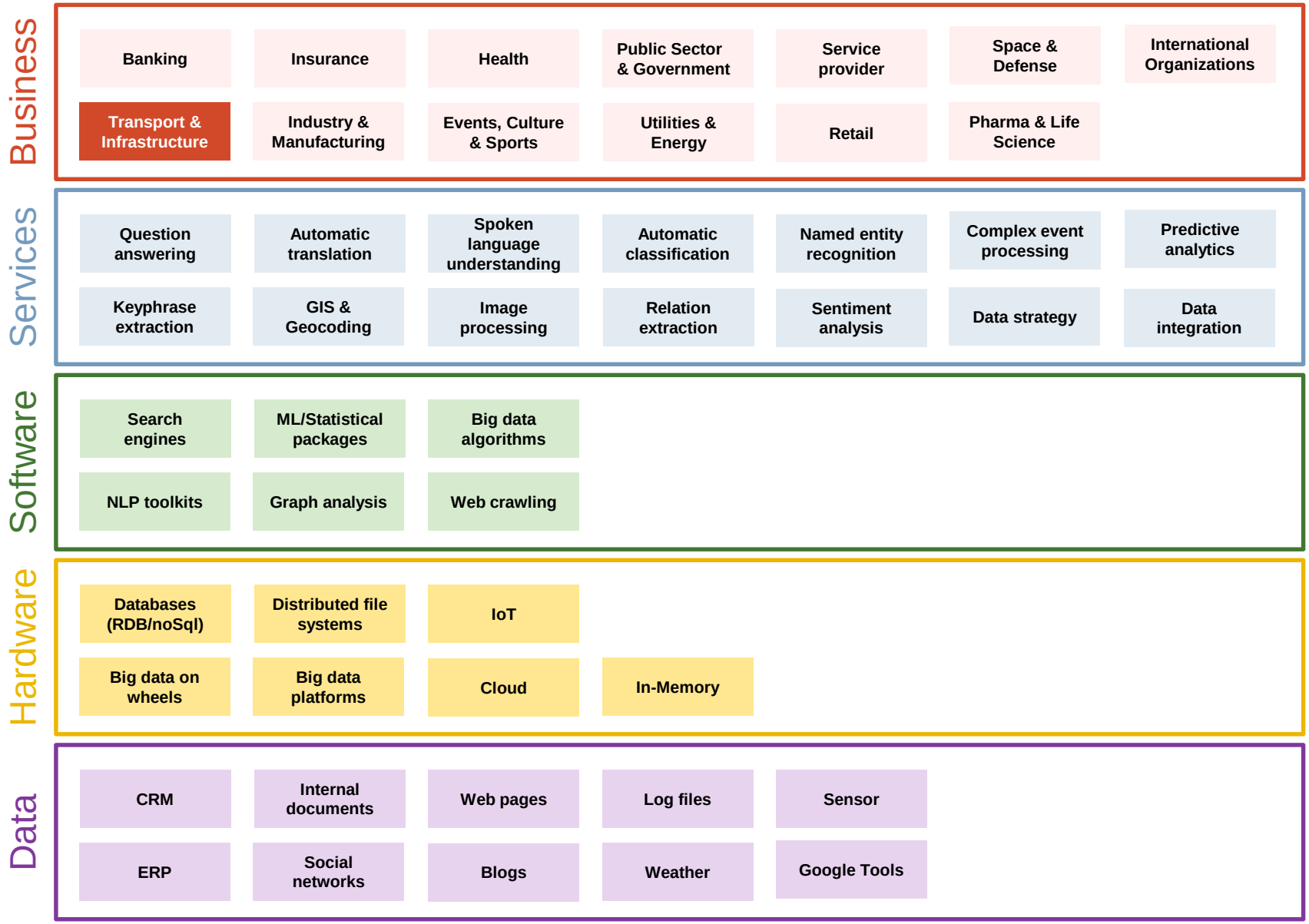
Statistical analysis: predict the occupancy of future shows based on show type, time, weather information, …. and devise a targeted eMarketing strategy for each type of show (drama, musical…) or even for individual shows.



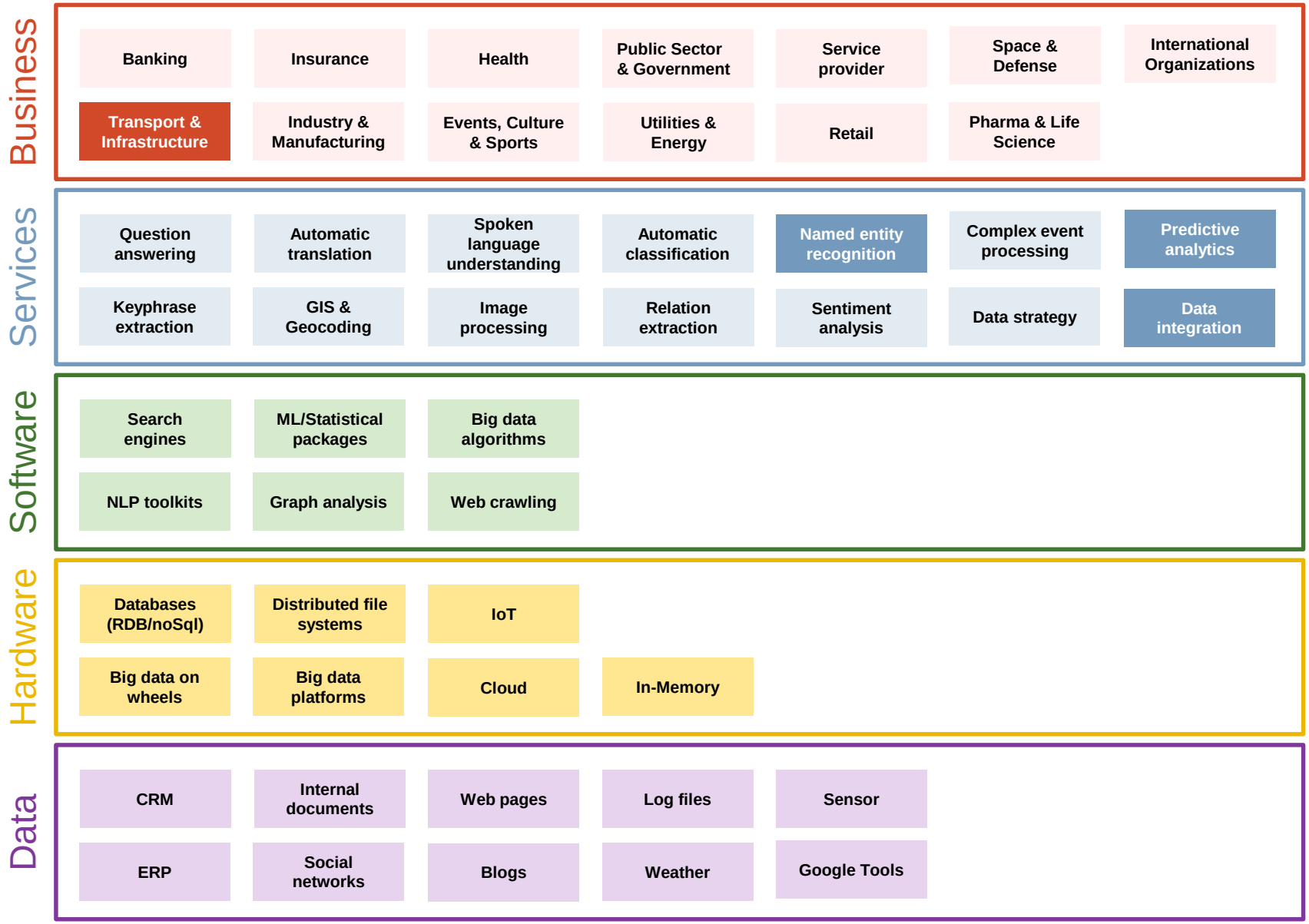
PREDICTIVE MAINTENANCE OF DOORS

Open or Late

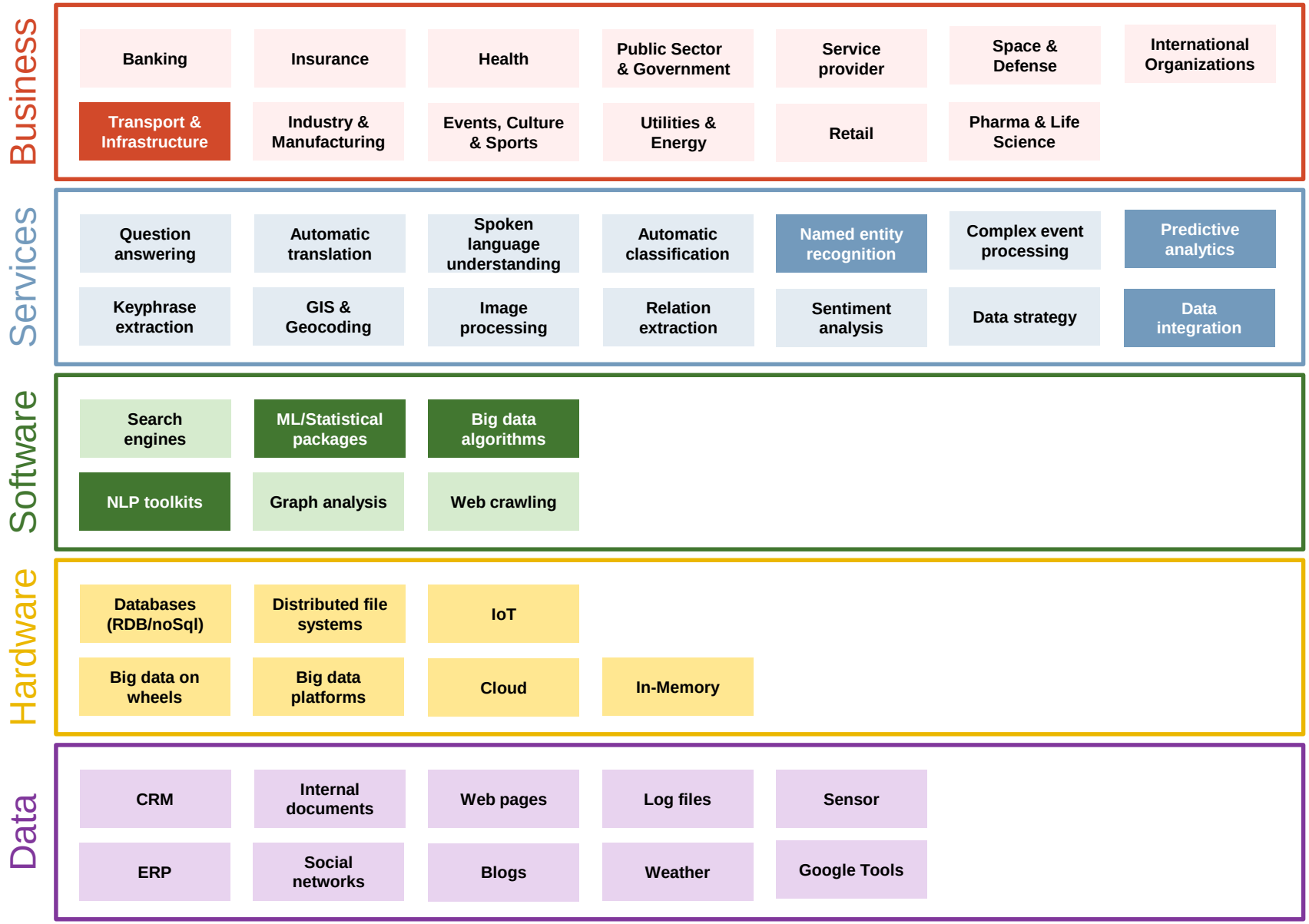
PREDICTIVE MAINTENANCE OF DOORS



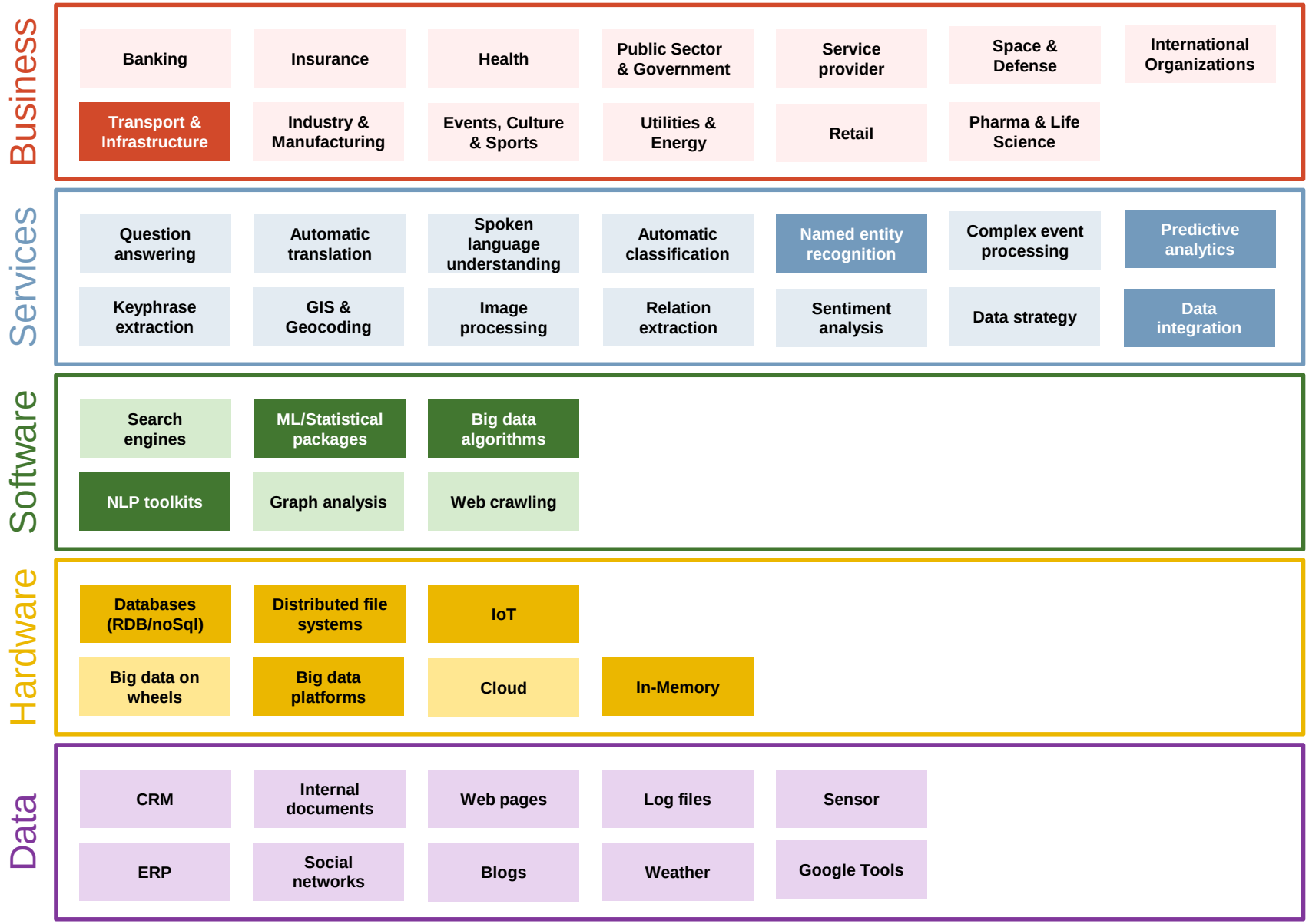
PREDICTIVE MAINTENANCE OF DOORS



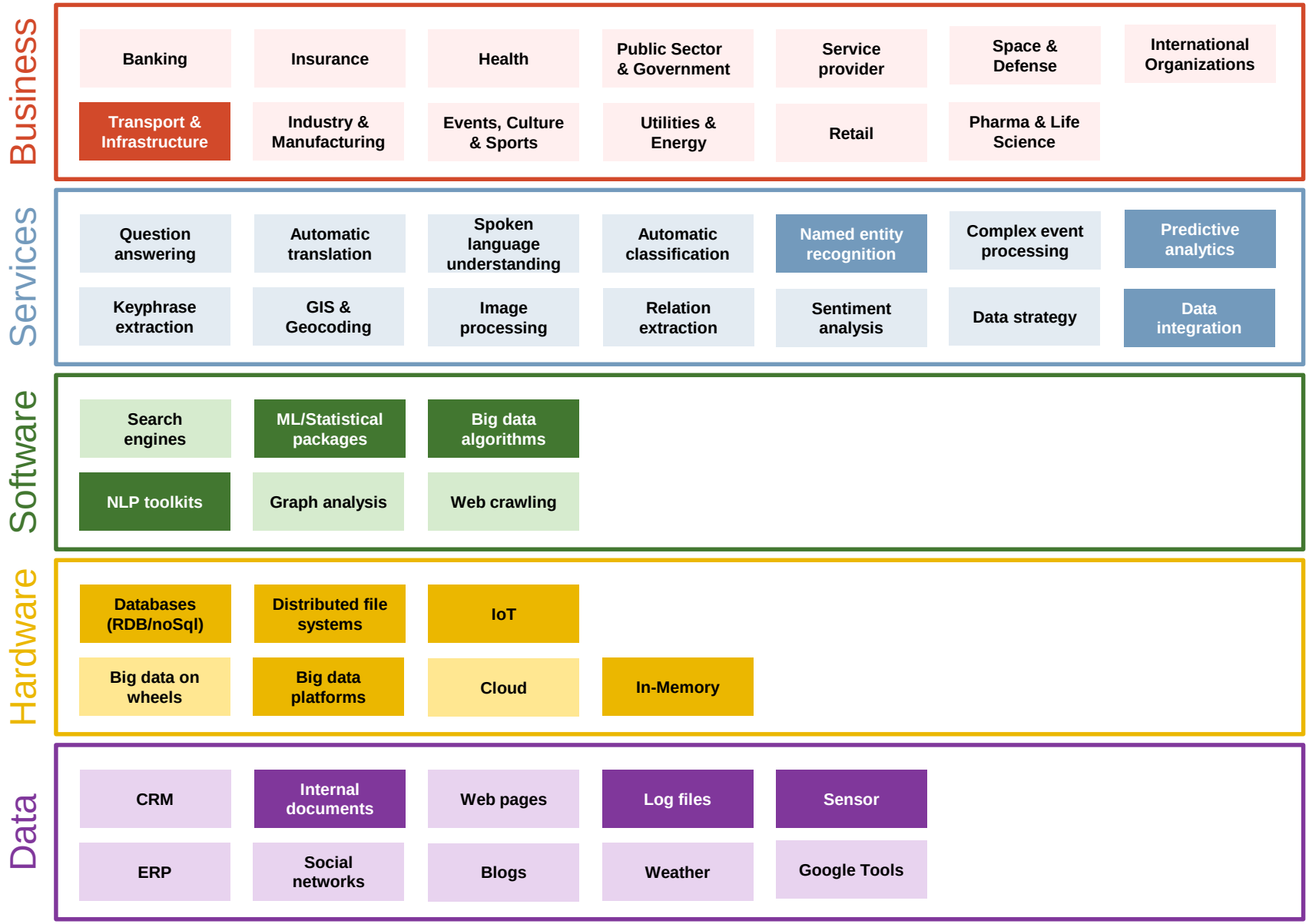
PREDICTIVE MAINTENANCE OF DOORS



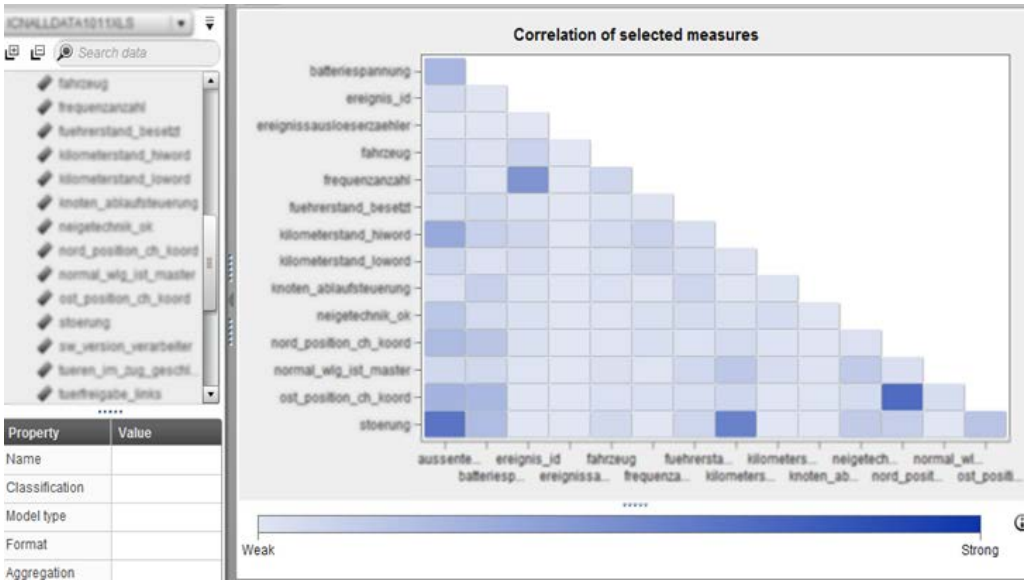
PREDICTIVE MAINTENANCE OF DOORS

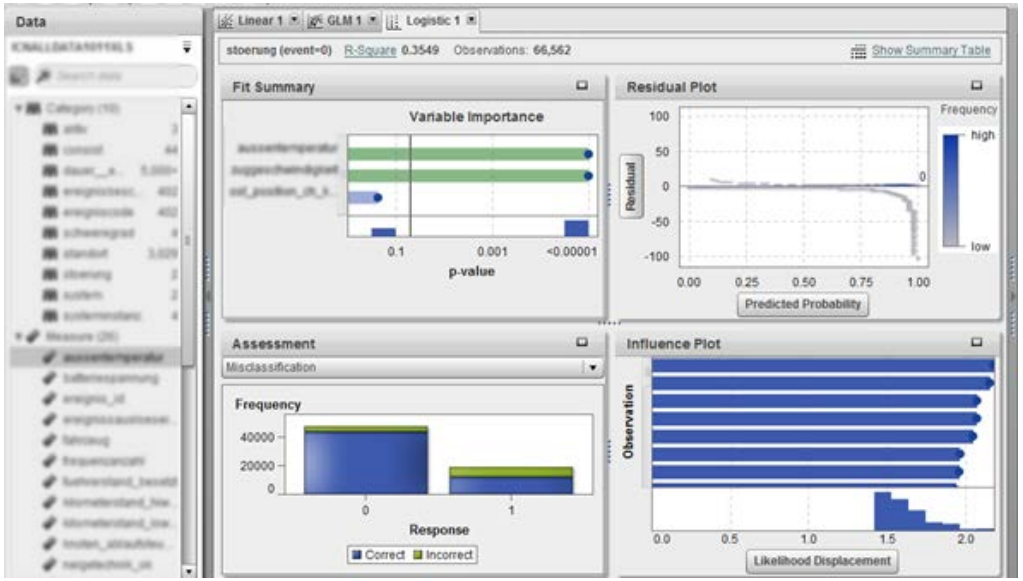


PREDICTIVE MAINTENANCE OF DOORS



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
Tel +41 21 613 21 11

Jérôme Berthier
Head of BI & Big Data
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