



Il ruolo dei big Data nella scuola e nella didattica

**Francesco M. Sacco,
Ferdinando Pennarola**

26 Novembre 2018

Siamo Un Data Cluster....

L'effetto combinato della pervasività degli accessi alla rete Internet e dell'utilizzo di device mobili, ampiamente penetrati nella popolazione a tutti i livelli di età di e dimensioni sociali, *alimenta un database di dimensioni inaudite, che cresce ogni giorno in modalità additiva*

Siamo Un Data Cluster....

Le persone, gli studenti, i docenti, gli staff amministrativi (tutti i sotto-insiemi di una unità scolastica e/o didattica) lasciano sempre più *tracce* circa il loro comportamento giornaliero, per almeno due ragioni:

- Possiedono e usano uno smartphone sempre connesso alla rete
 - Utilizzano la rete per finalità inerenti alla didattica

Dati Digitali

I dati digitali valgono «doppio»

Ci sono sempre due dimensioni di un dato in formato digitale:

Il **dato** in sé

Il **metadato** (dove, quando, da chi, da quale interfaccia tecnica è stato originato)

La parte preziosa, e spesso nascosta, dei «big data» è quella sommersa e non visibile



**The
Economist**

MAY 5TH-12TH 2017

Crunch time in France

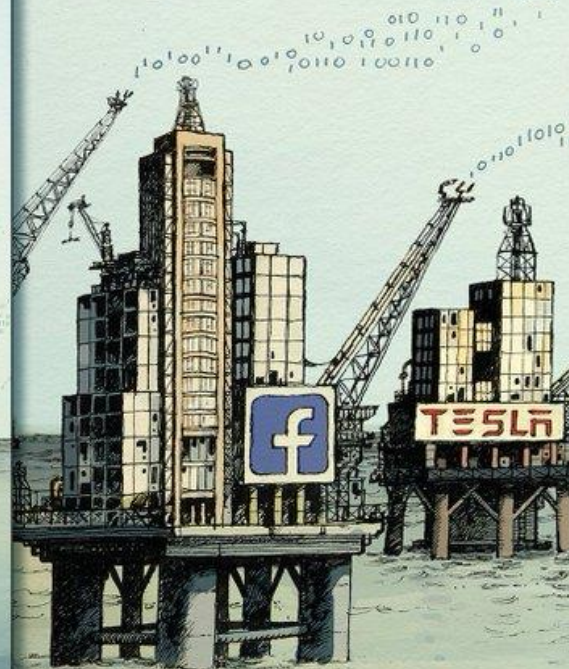
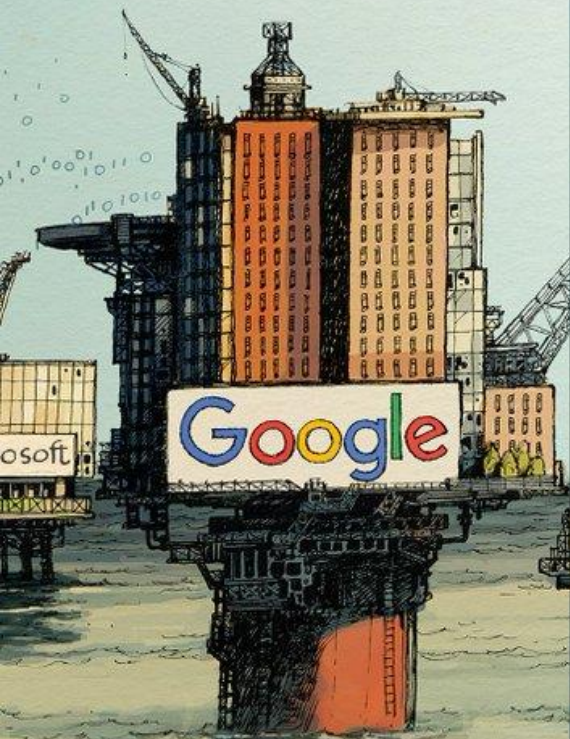
Ten years on: banking after the crisis

South Korea's unfinished revolution

Biology, but without the cells

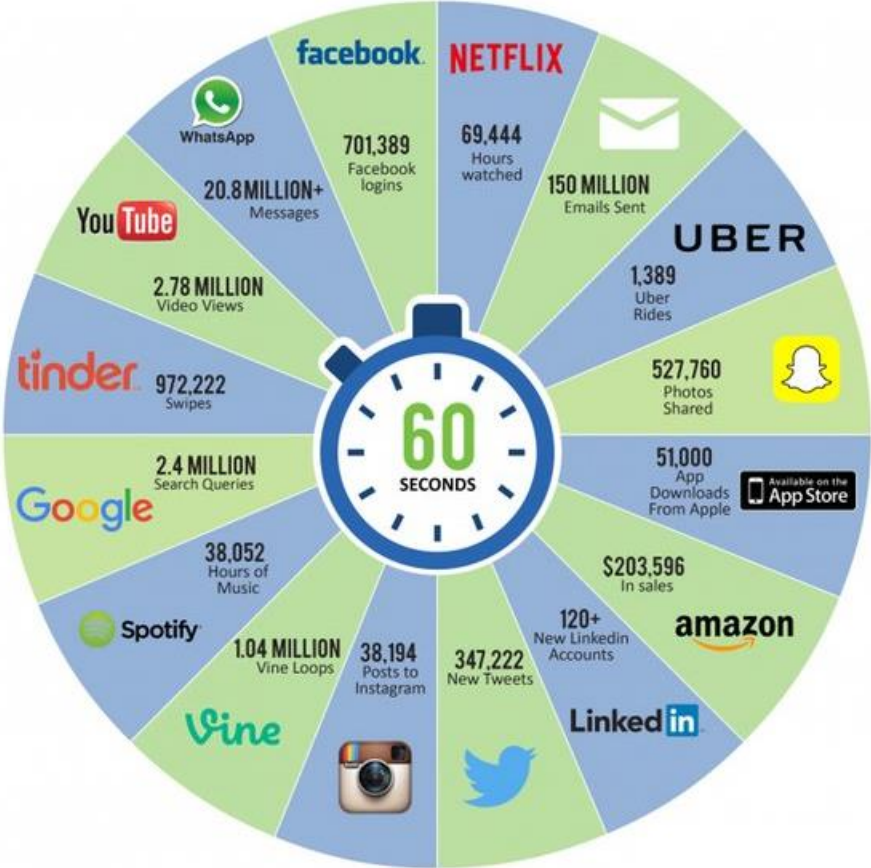
The world's most valuable resource

**Data and the new rules
of competition**



Big data...

Un minuto di Internet nel 2016...



...e nel 2017



Big data...

...e nel 2018





Cosa si può fare...

CORRIERE DELLA SERA
PRESENTA



iHAPPY 2014

di
Andrea Ceron
Luigi Curini
Stefano M. Iacusi



ECCO COME È ANDATA IN DETTAGLIO QUESTO ANNO DI FELICITÀ

VALORE MEDIO
INDICE IHAPPY

58,6

NEL 2013: 60,3%

NUMERO DI GIORNI
SOPRA AL 50%

310

PARI A CIRCA L'85%

VALORE MEDIANO
INDICE IHAPPY

59,9

NEL 2013: 62,6%

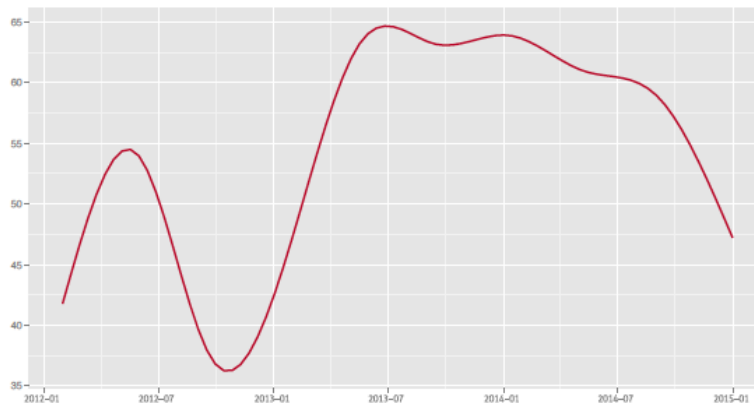
NUMERO DI GIORNI
SOTTO AL 50%

55

PARI A CIRCA IL 15%

2012-14 ANDAMENTO DI IHAPPY IN ITALIA

% di target Italia



I CINQUE
GIORNI
PIÙ FELICI

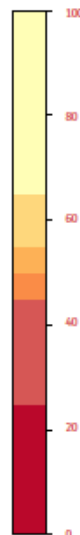
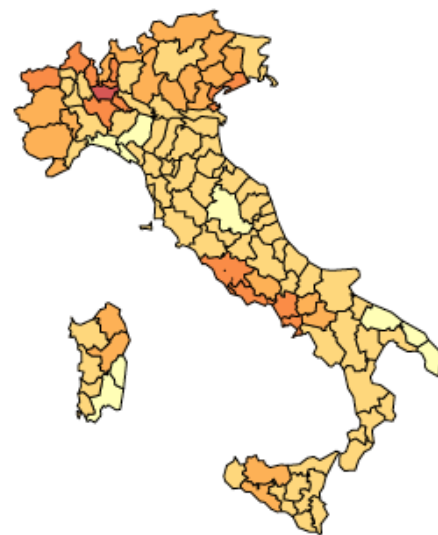
indice iHappy

GENNAIO	OTTOBRE	SETTEMBRE	GIUGNO	MARZO
1	16	28	1	28
75,1%	74,2%	74,1%	73,6%	73,0%
Nuovo anno; Fiat compra tutta Chrysler; De Blasio giura come sindaco di NYC; salvi due bambini dopo una notte nei boschi su Monte Livata	Presentazione manovra finanziaria: meno tasse sul lavoro, 80 euro confermati, Tfr in busta paga; Summit asiatico ASEM a Milano	Due foto: a) Abbraccio tra i due Papi; b) i Clinton coccolano la piccola Charlotte. Inizia la rivoluzione degli ombrelli ad Hong Kong	Presentata la lista degli azzurri per i Mondiali brasiliani	Obama dal Papa a Roma elogia Italia su riforme; si discute sulla riforma del Senato; Putin chiama Obama per trattare su Ucraina

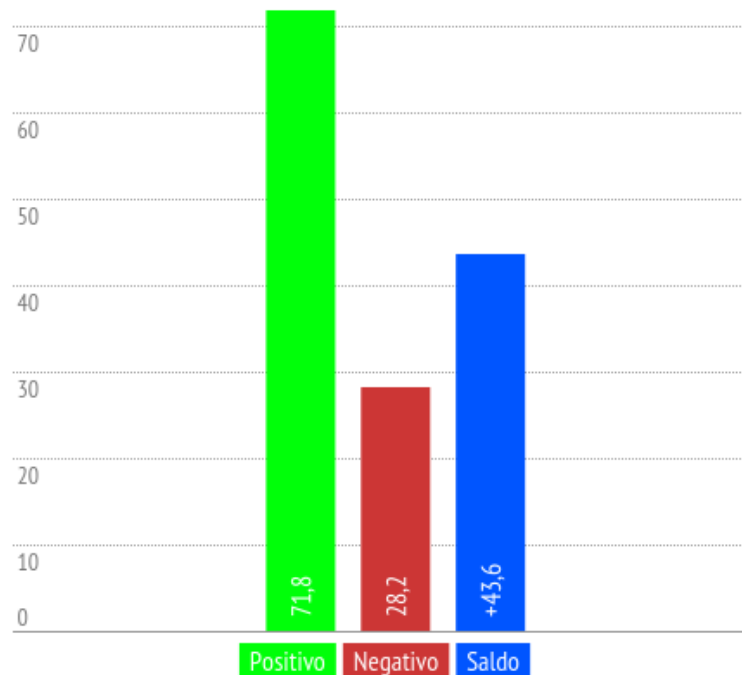
I CINQUE
GIORNI
PIÙ TRISTI

indice iHappy

OTTOBRE	OTTOBRE	DICEMBRE	AGOSTO	OTTOBRE
22	31	12	23	12
33,1%	33,9%	35,0%	36,0%	36,6%
Attacco al parlamento canadese in diretta TV	Tutti assolti per la morte di Stefano Cucchi	Morte di Virna Lisi	I russi oltre il confine ucraino; tragedia dei migranti in mare: 200 morti dalla Libia	Il giorno dopo l'alluvione di Genova: polemiche; paura per il contagio dell'Ebola

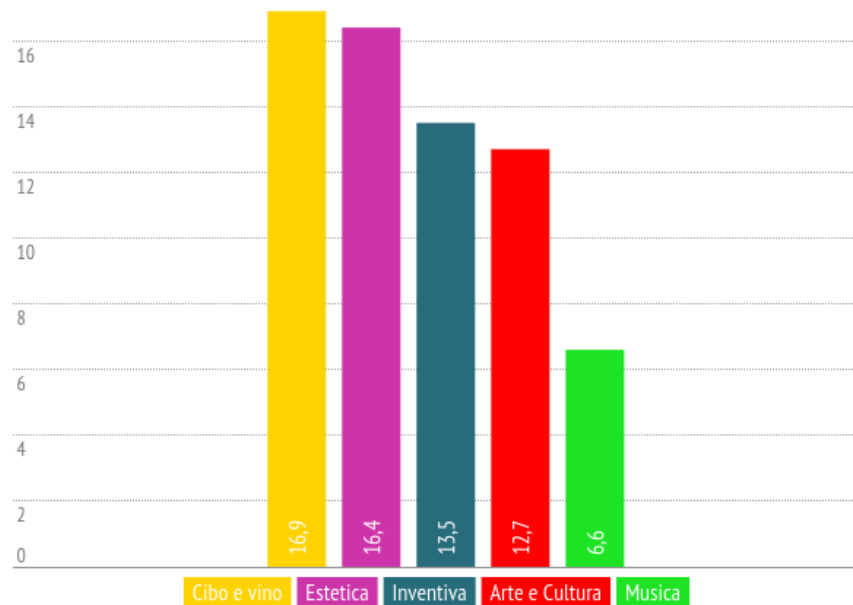
FELICITÀ MEDIA ANNUALE
DELLE SINGOLE
PROVINCE ITALIANE

Il sentiment verso l'Italia

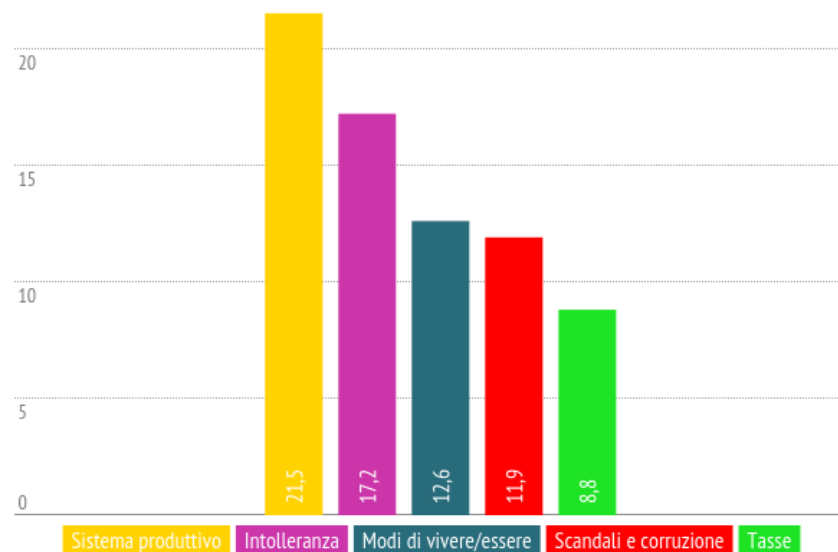


Nota: Al netto dei commenti neutri

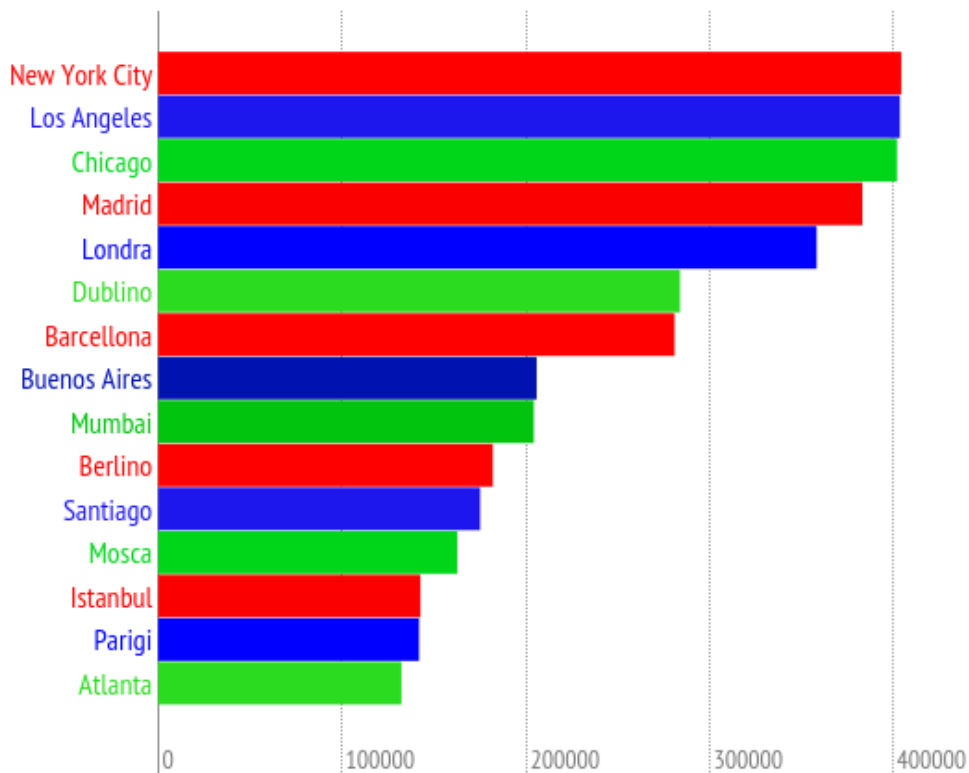
I cinque principali aspetti positivi dell'Italia



I cinque principali aspetti negativi dell'Italia

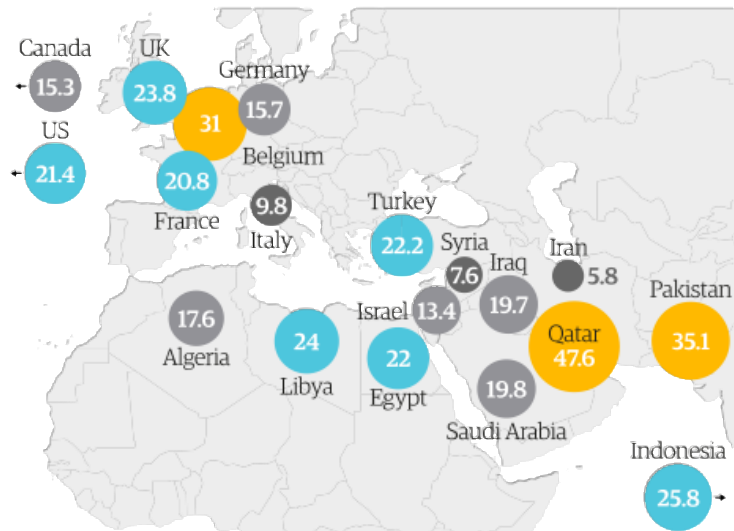


Le 15 città in cui si parla di più di Italia



Arabic language social media support for Isis

Percentage of posts expressing positive sentiment towards Isis in a study of Arabic language social media

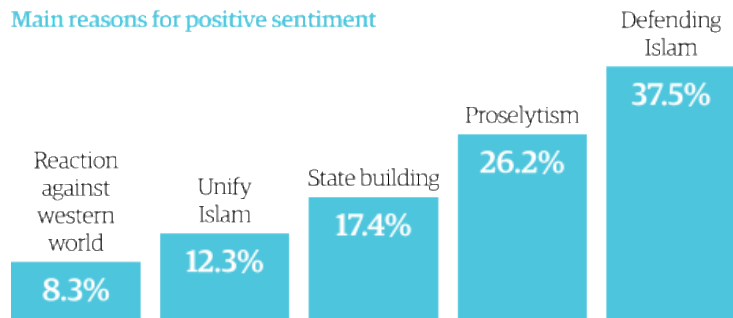


GUARDIAN GRAPHIC

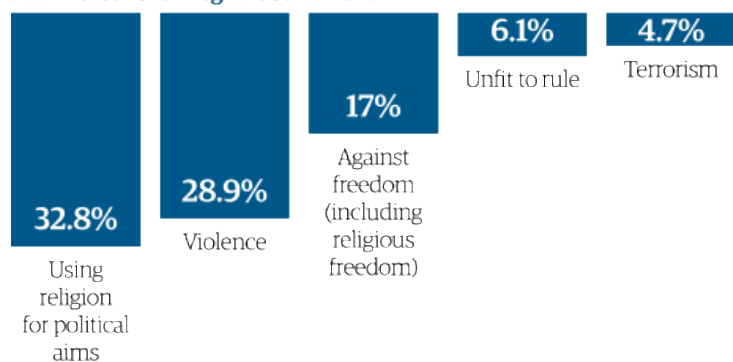
SOURCE: VOICES, SOCIAL MEDIA

Isis in Arabic language social media

Main reasons for positive sentiment



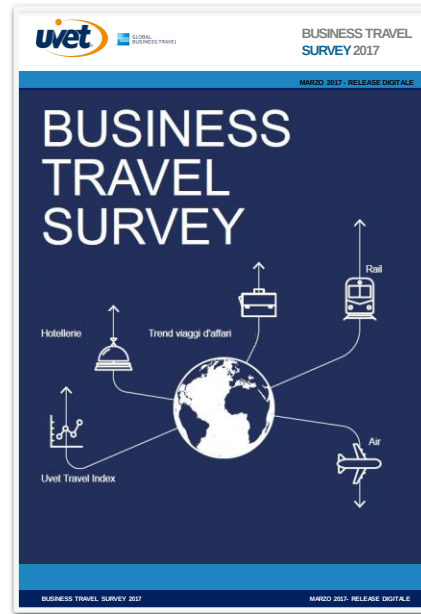
Main reasons for negative sentiment



GUARDIAN GRAPHIC

SOURCE: VOICES, SOCIAL MEDIA

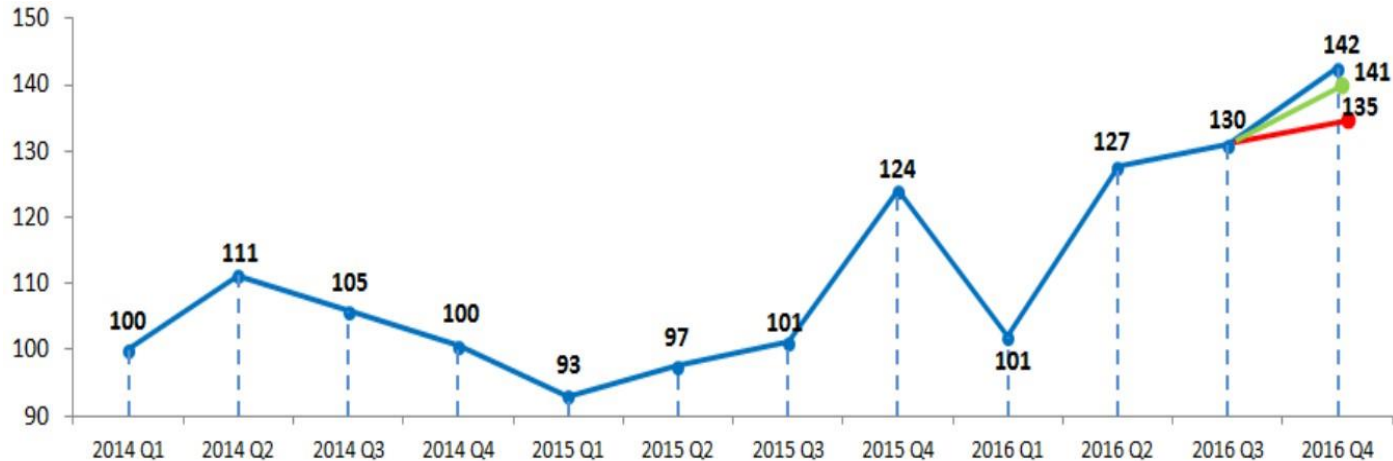
Travel Volumes, Fares and the Economy



Travel from Italy only: does Economic growth drive Business Travel?

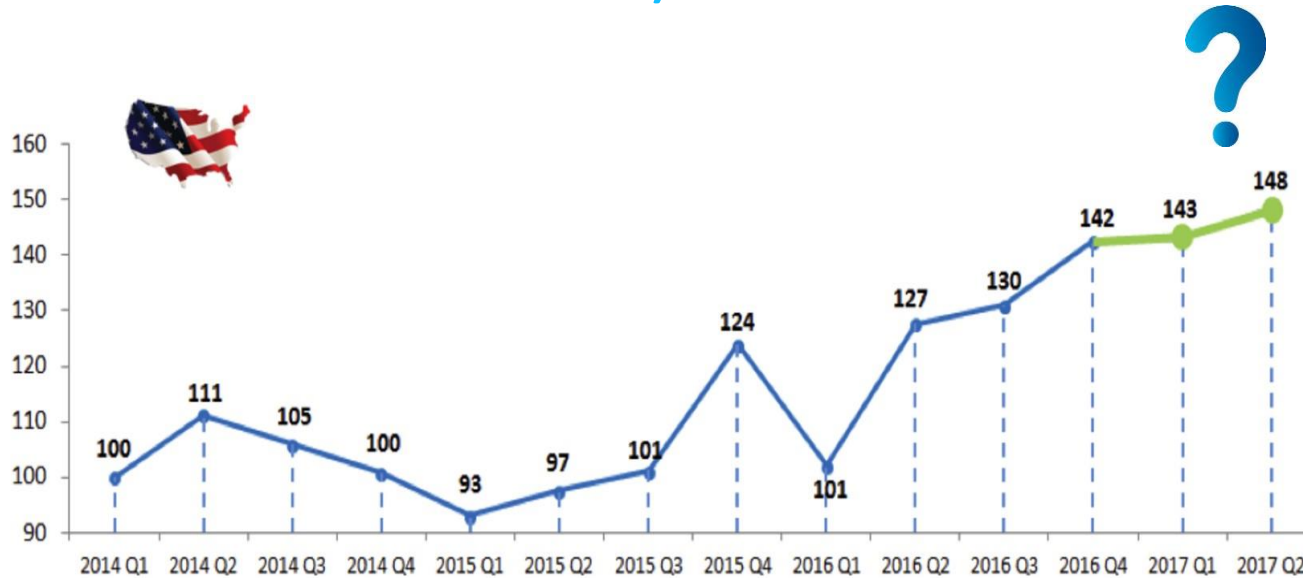
Destination	Lag (number of quarters)	Spearman Correlation
United States	3	0.93**
Japan	1	0.92**
Poland	2	0.92**
Chile	1	0.85**
Netherlands	1	0.72**
Israel	3	0.71**
Portugal	2	0.71**
Germany	3	0.67**
Denmark	1	0.63**

The prediction for 2016 Q4: USA



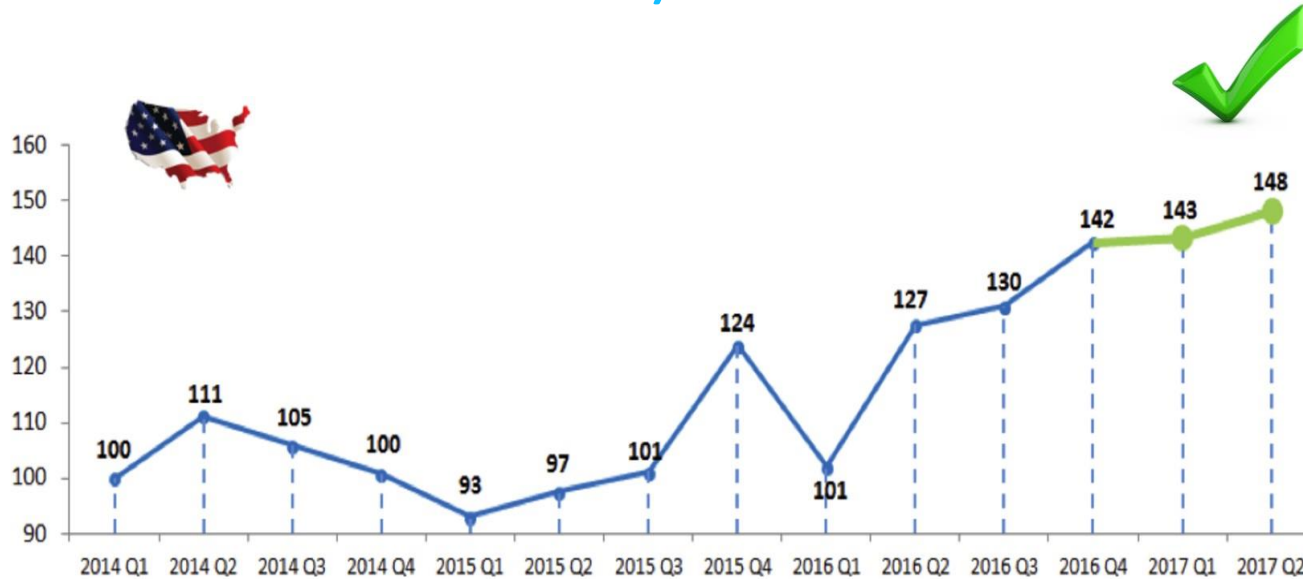
- Quarterly real volume
- Prediction for 2016 Q4 with macroeconomic variables (GDP, etc)
- Prediction for 2016 Q4 without macroeconomic variables

The prediction for 2017 Q1 & Q2 (as of December 2016)



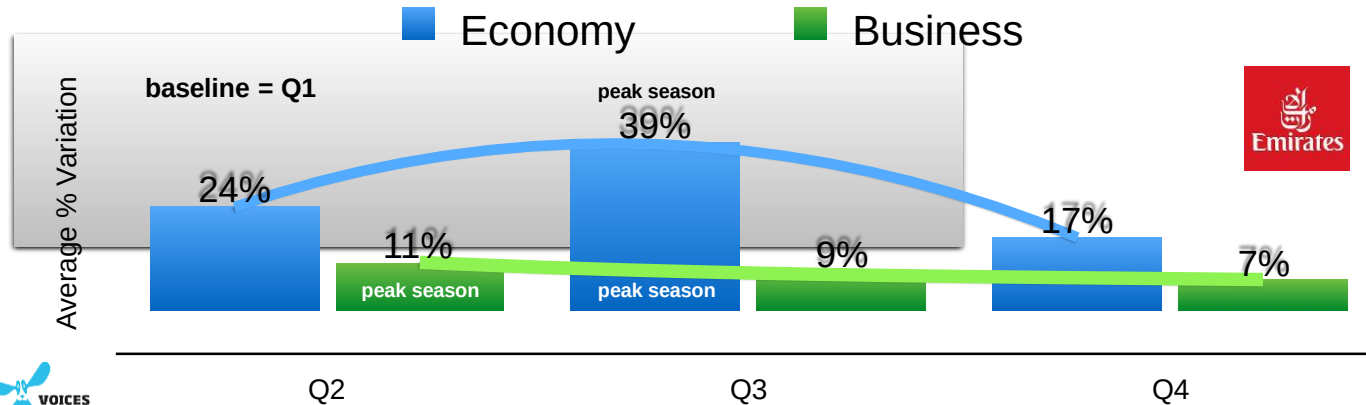
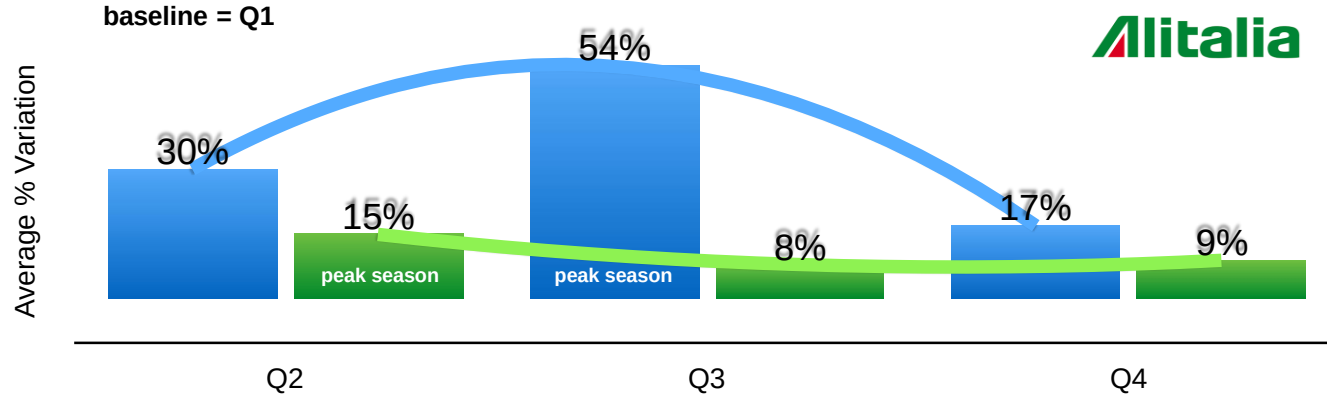
- Quarterly real volume
- Prediction for 2017 Q1 & Q2 with macroeconomic variables

The prediction for 2017 Q1 & Q2 (as of December 2016)

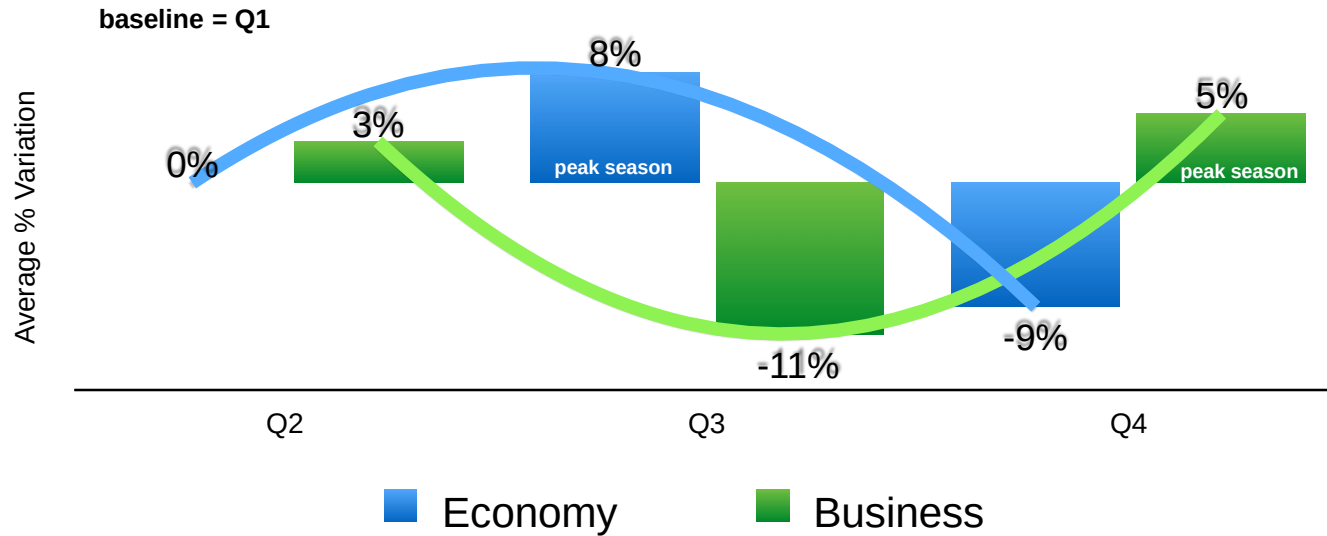


- Quarterly real volume
- Prediction for 2017 Q1 & Q2 with macroeconomic variables

Patterns in Air Fares? An example: Milan to New York

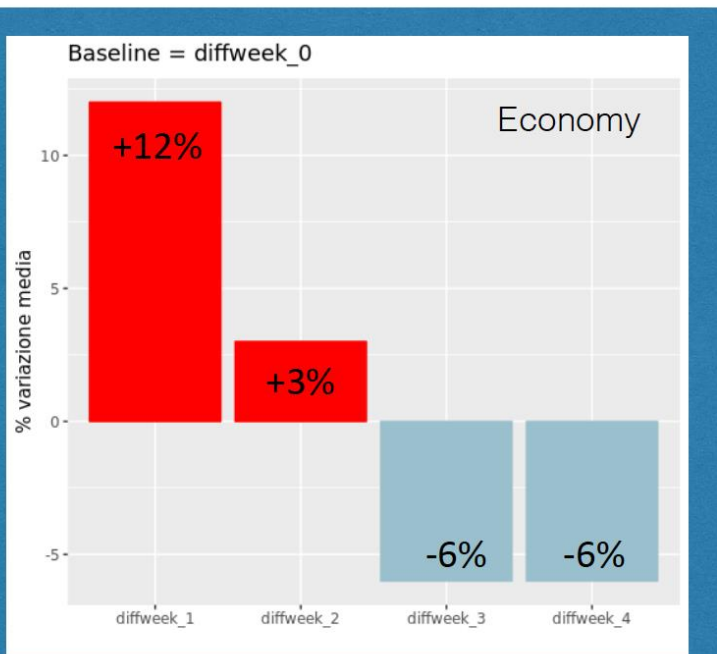


Patterns in Air Fares in Asia? An example: Milan to Shanghai



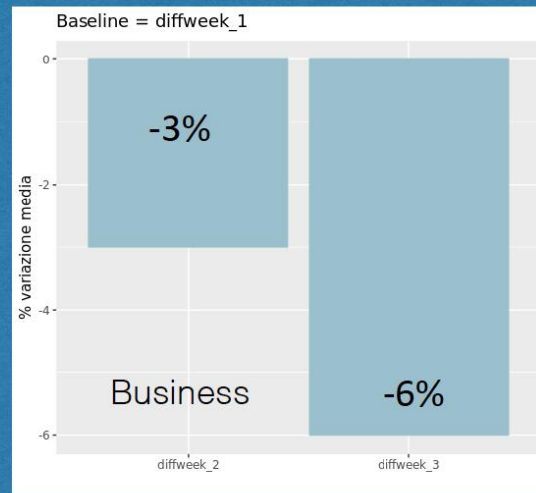
*Lowest fare = Q4, no longer Q1
Business class behaviour also changes*

Advanced booking pays back in Business class, not always in Economy class



Last minute offers exist, otherwise
book at least three weeks in
advance

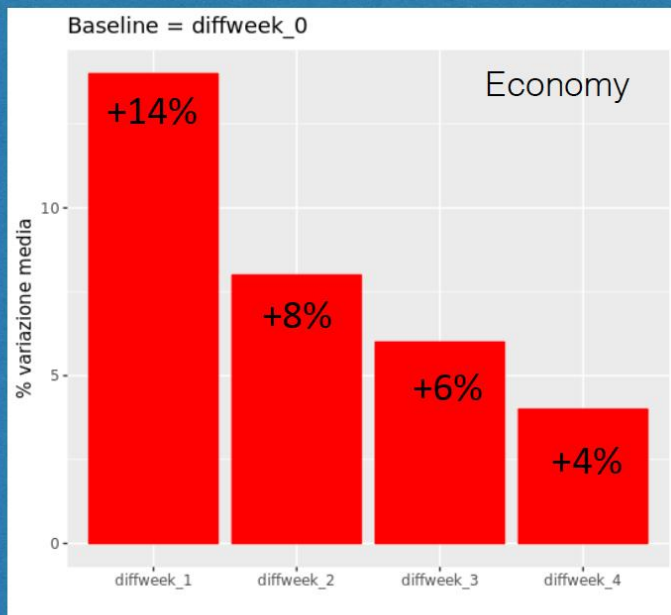
Alitalia



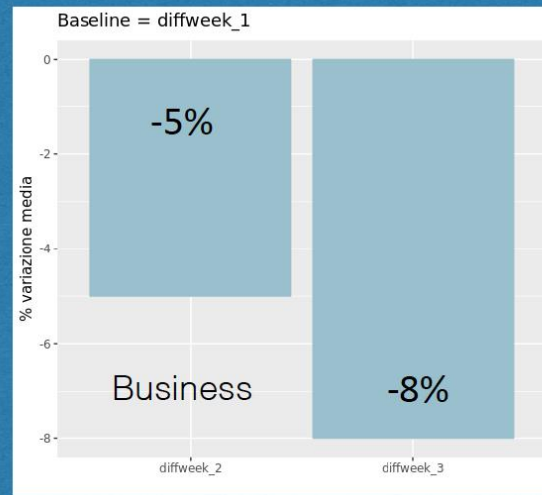
*Business flight always booked in
advance, the sooner the better!*

diffweek: number of weeks in advance, diffweek_0 = same week

Advanced booking pays back in Business class, not always in Economy class



Last minute offers exist, otherwise
book the sooner as possible

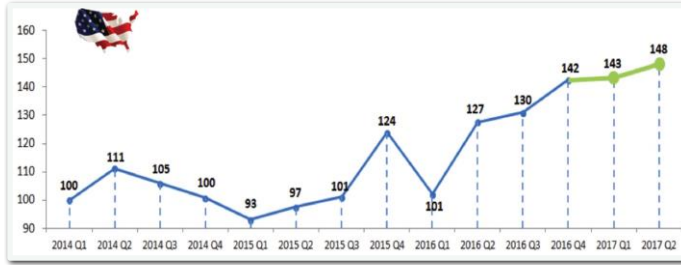


*Business flight always booked in
advance, the sooner the better!*

diffweek: number of weeks in advance, diffweek_0 = same week

Mix all information

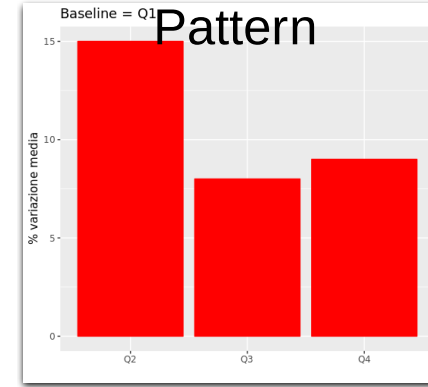
GDP - Flight Prediction



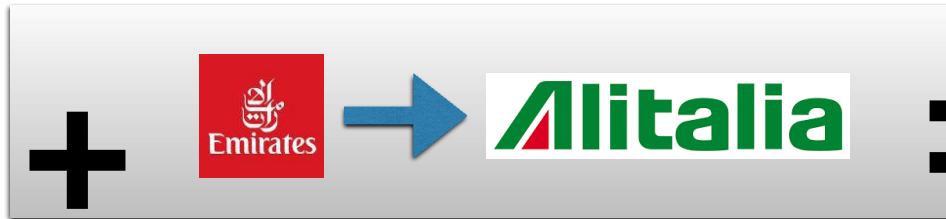
+

Air Fare

Pattern

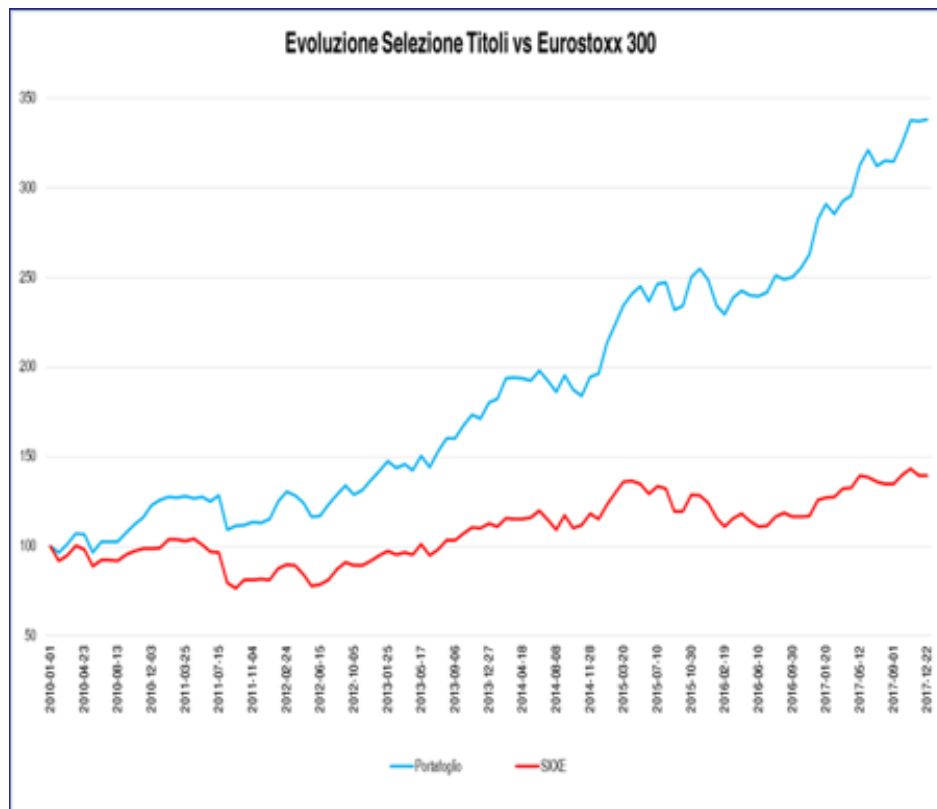


Competitors Analysis



Pre buy seats
& Resell for
better service
(or profit)

Make profits on machine learning



Rend. composti Gen10 - Dic17

Portafoglio	Alpha	SXXE
238.1%	198.8%	39.3%

Rendimenti medi annui

Portafoglio	Alpha	SXXE
17.1%	12.2%	4.8%

Alpha

	Annuale	Mensile
Min	+6.2%	-5.6%
Media	+12.2%	+0.8%
Max	+26.9%	+6.0%

Numero medio di titoli selezionati
40

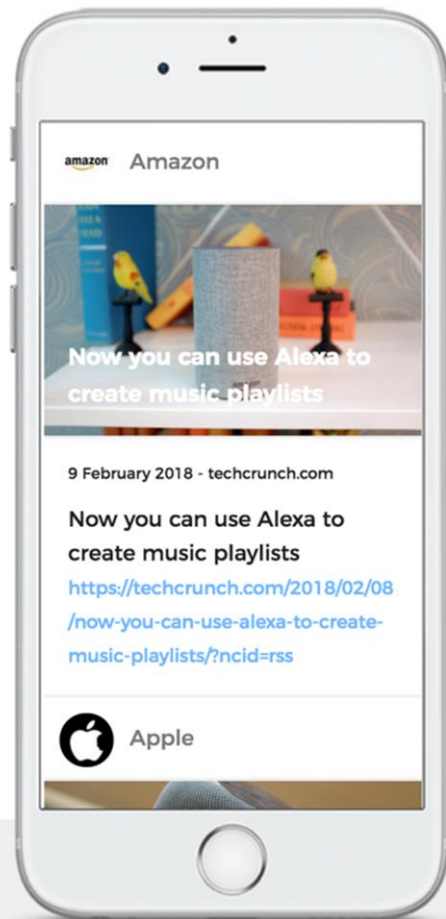
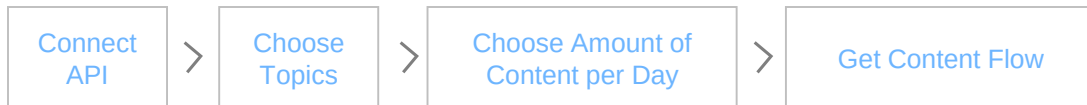
Alla fine di ogni mese mediamente vengono effettuate circa 40 operazioni, per un turnover di portafoglio annuo pari a circa 6x.
In ipotesi di 3bps di commissione di negoziazione si stima un costo complessivo pari a circa 18bps annui.

Content Feed

ContentFeed aggrega in automatico contenuti relativi a qualunque interesse e li fornisce all'utente attraverso un feed di *news cards*.

Ogni *news card* è visualizzata in un format standard, chiaro e leggibile. Dalla card è possibile poi accedere al sito originale che ha pubblicato il pezzo.

Il feed viene fornito attraverso una API, attraverso la quale è possibile specificare gli interessi da trattare, l'ammontare ed anche gli eventuali “pesi” relativi.



The Washington Post

Monkey Cage

How pollsters could use social media data to improve election forecasts

By **Andrea Ceron, Luigi Curini and Stefano Iacus** December 21 at 9:00 AM

Donald Trump's Nov. 8 victory surprised almost everyone. But if pollsters had looked at Twitter, they might have recognized that the race was close — or so we learned in our recent research.

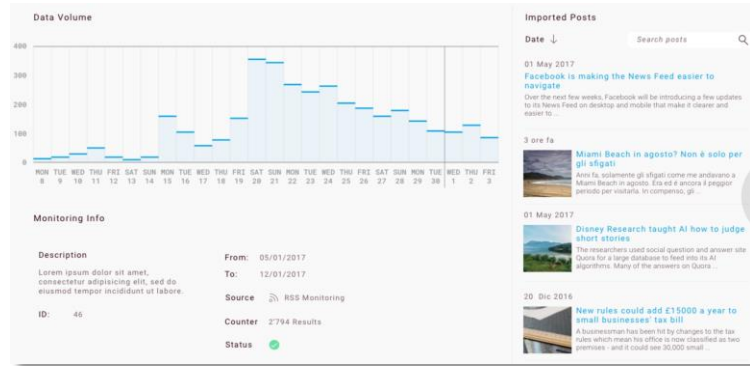
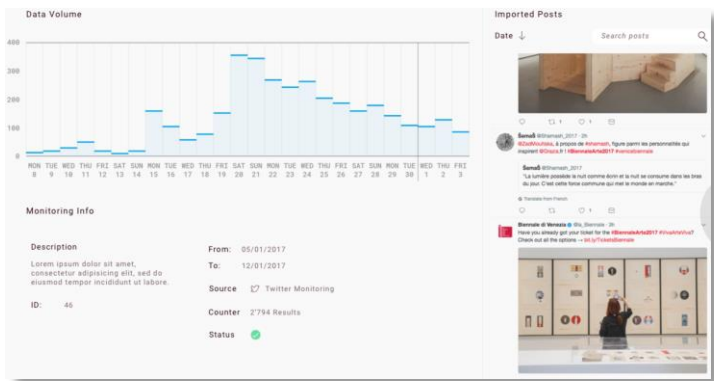
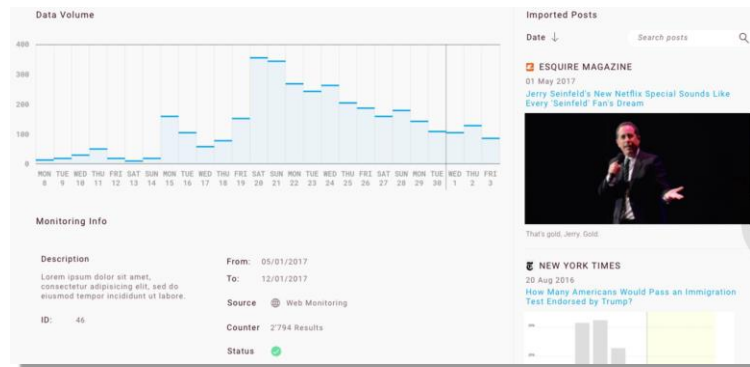
Even when polls were showing a big lead for Hillary Clinton, real-time analysis of social media was showing it getting tighter — even if you subtract the propaganda bots used by Trump staff, estimated to make up about 20 to 30 percent of the total.



Cos'altro si può fare...

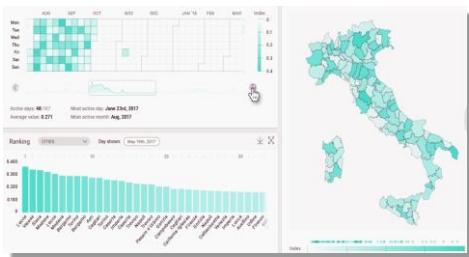
Wireframe of future VOICES ANALYTICS® (Prototype version, may change)

MONITORING MULTIPLE SOURCES

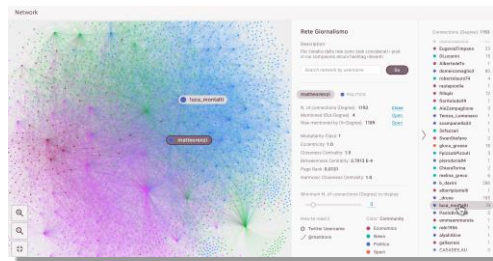


Wireframe of future VOICES ANALYTICS® (Prototype version, may change)

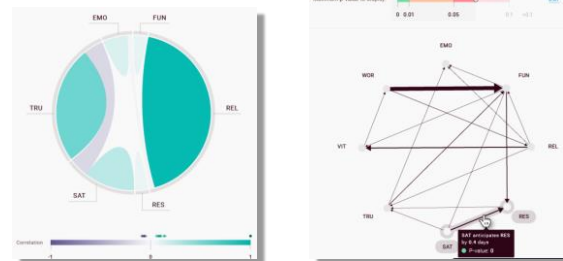
Geolocalized Social Media Index



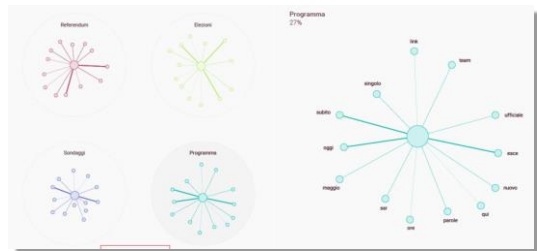
Interactive Social Network Analysis



Lead-Lag & Correlation Analysis



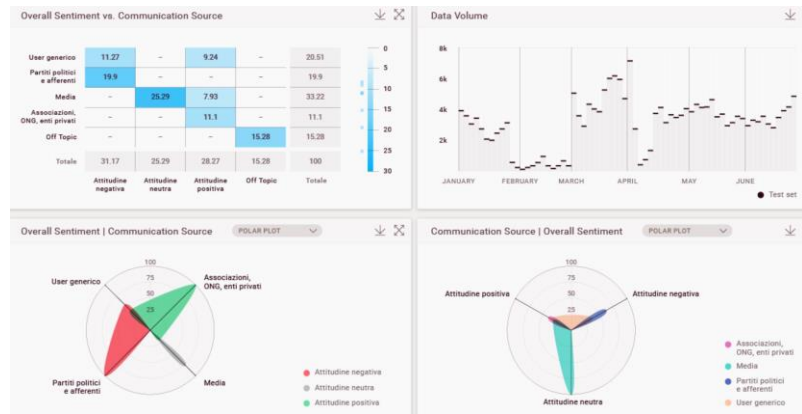
Topic Discovery



Social NPS Net Promoter Score



Cross Tabs



And much more...

Reputation Opinion Analysis

It reports the sentiment and describes the main opinions around a brand/product/topic.

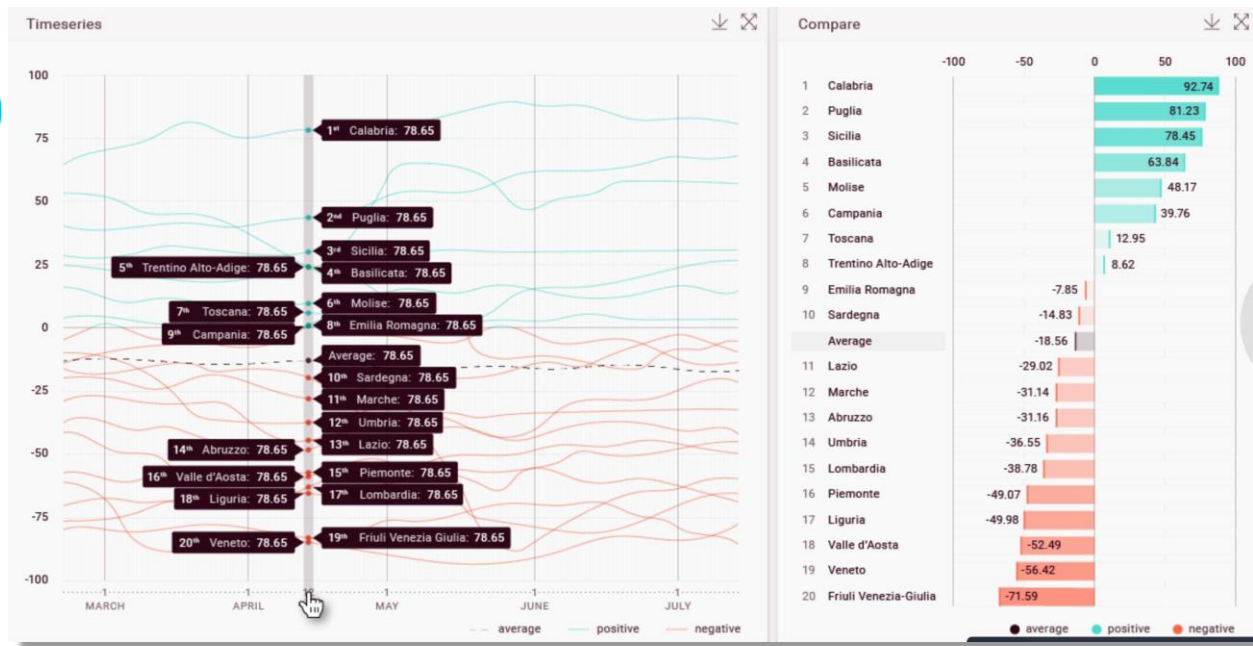
It also provides the reasons for expressing a positive or negative sentiment.



Social NPS (Net Promoter Score)

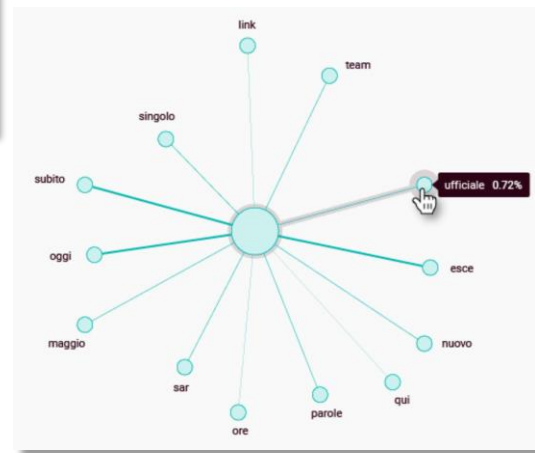
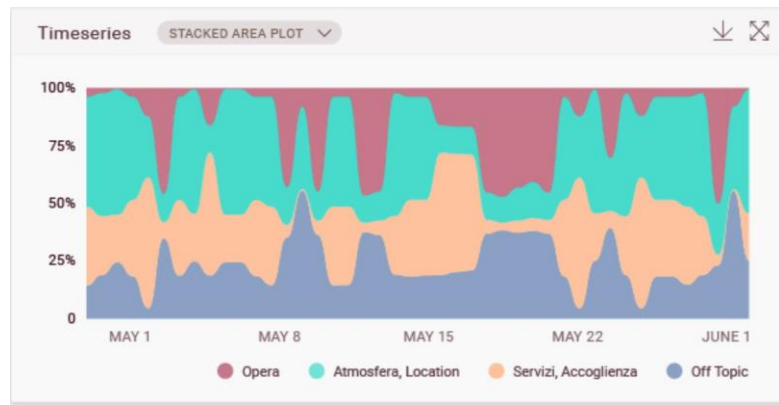
It performs a comparison across the brands active in a peculiar market, showing the net rate between promoters and detractors.

This gives an idea about of the online “word of mouth” around each brand.



Market Trend Discovery

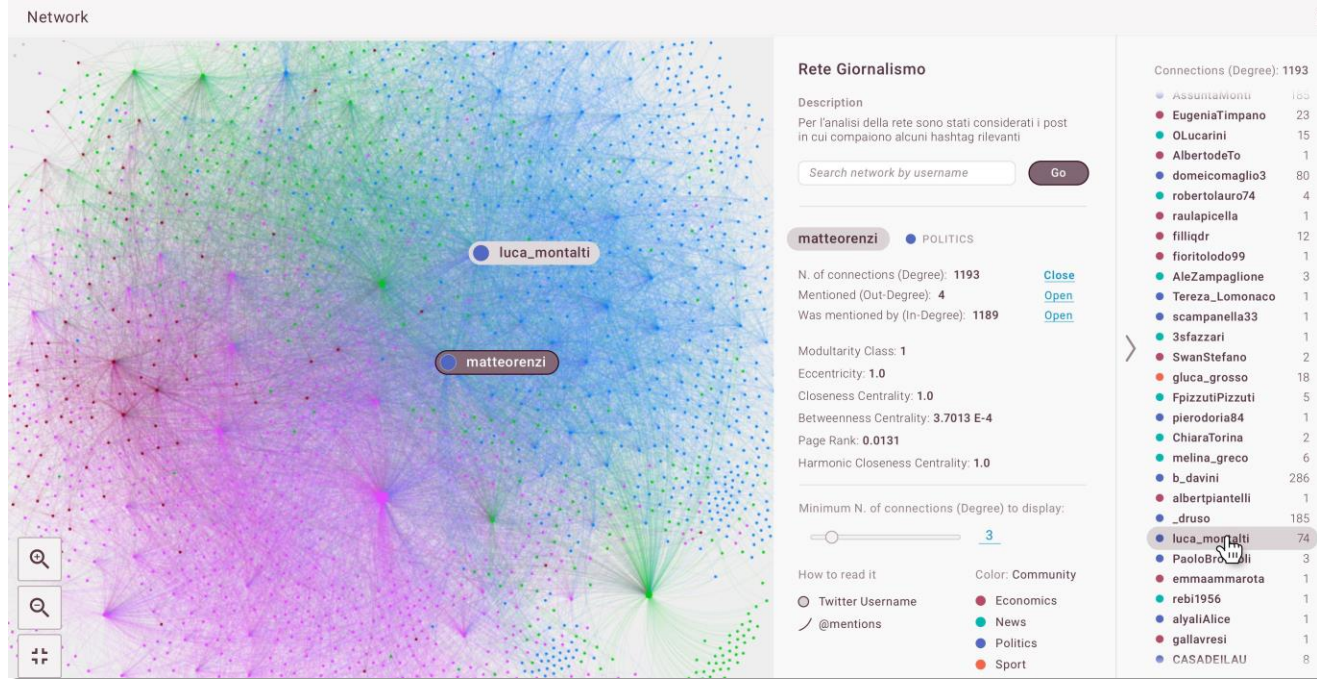
It investigates the words and the concepts associated to a product/market/topic and allows to identify unsatisfied needs of the consumers or new trends in consumption and opinions combining sentiment with topic discovery models.



Influencer Network

It applies social network analysis to identify the influencers and the “hub” active on conversations around a brand/topic.

This allows to select users that must be engaged for marketing activities and to spread communication.



Social Index (geo-located)

This promptly builds indexes of mood/views/attitudes showing the variation across time and space.

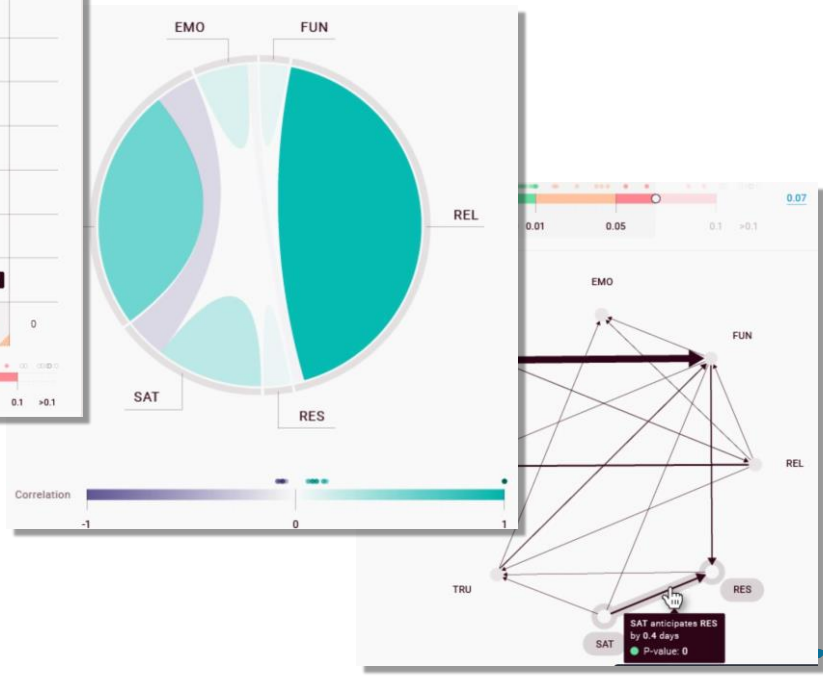
It can be used for content generation purposes or as fire-alarm. The index can integrate sentiment with other sources of data.



Predictive Analytics

It combines sentiment analysis with other sources of data (e.g. google query, websites visits, daily revenues or other own data).

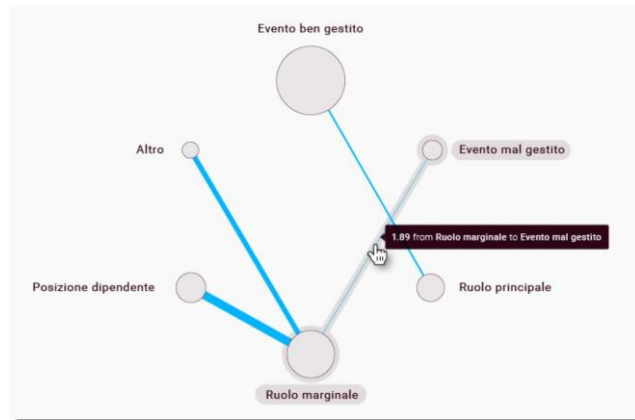
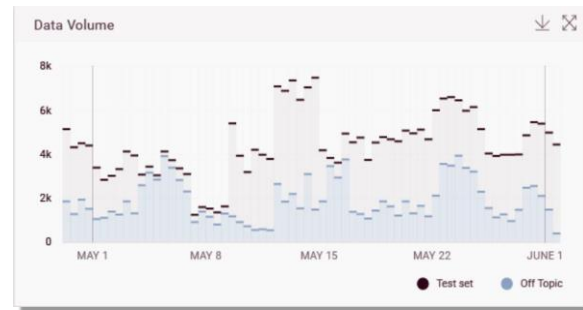
It performs time series analysis to identify breakpoints, trends, predict outcomes and evaluate which time series can anticipate an increase/decrease in the level of another one in order to take action.



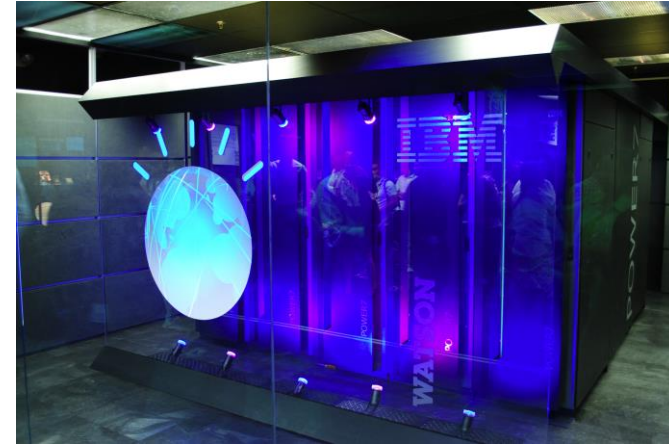
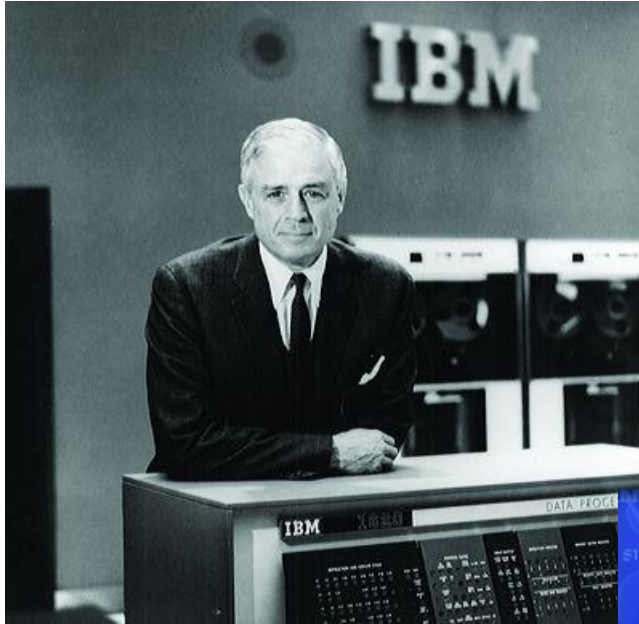
ADV Monitor

It monitors the effectiveness of advertising/communication campaigns by combining reputation analysis with predictive analytics.

This allows to take action adjusting the frame of the campaign and evaluating its effects going beyond the number of “likes”.



Intelligenza artificiale...



Loop.ai, un esempio...

LOOP LEARNING APPLIANCE

domain-specific unsupervised learning
on-premises

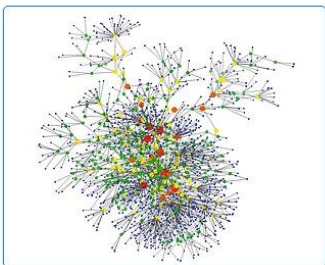


INPUT
UNSTRUCTURED
TRAINING TEXT
(BIG DATA)



**LOOP LEARNING
APPLIANCE**

available also as a
virtual appliance



OUTPUT
LOOP CORTEX
self-learned
knowledge model

LOOP REASONING APPLIANCE

domain-specific unsupervised text understanding
on-premises



LOOP CORTEX
synched from the
Learning Appliance



INPUT
UNSTRUCTURED
TEXT TO ANALYZE
(SMALL DATA)



**LOOP
REASONING
APPLIANCE**

Available also as a
virtual appliance or
embedded



OUTPUT
Structured Output
Discovered concepts in the
INPUT text with weights
and optional sentiment
(JSON)



Cosa si può fare per la scuola e la didattica...

Progettiamo il futuro

Adeguiamo le conoscenze della scuola

Interroghiamo i «big data» per capire che cosa effettivamente accade nei processi didattici giornalieri



Immaginiamo il futuro

Una *convenzione con i network provider* da cui transitano i dati geo localizzati (originati dai plessi scolastici durante gli orari di servizio nei giorni del calendario scolastico)

Accesso del team di ricerca ai «big data» in formato:

- *Anonimo* (non sarebbe possibile altrimenti)
- *Aggregato* (non sono possibili altre modalità)



Interroghiamo i dati

Qual è l'effettivo (e misurato) grado di «**multimedialità**» della generazione di studenti attuali?

Quali sono le **applicazioni** e i servizi più utilizzati tramite smartphone?

Come si **distribuiscono** questi comportamenti d'uso nel corso della giornata e nelle varie classi di vario ordine e grado?



Interroghiamo i dati

Qual è l'uso e l'accesso ad **informazioni** attinenti la didattica svolta nella scuola, in un dato momento?

Quali sono i **network** che si stabiliscono in questa comunità di discenti e docenti e quali sono le dimensioni, i nodi di questi network, le frequenze di accesso nel corso della giornata didattica?



Interroghiamo i dati

Quali sono le **dimensioni** di questo traffico: di dati a cui si è fatto accesso (download) e dati che sono giornalmente creati (upload)?

E' possibile astrarre dai dati quelle informazioni che sono utilizzate per lo **svolgimento delle attività didattiche** e che si presume siano acquisite con gli smartphone personali degli utenti?





Quali implicazioni di policy?

So What?

Una voce per la scuola:

- Indicazioni per gli *investimenti* nelle tecnologie e nei servizi per la didattica, a livello di plesso scolastico, di città, regione o intero Paese
- Indicazione per gli interventi di *formazione* e adeguamento delle competenze
- Indicazione per lo sviluppo di *partnership* della scuola con il mondo del lavoro

