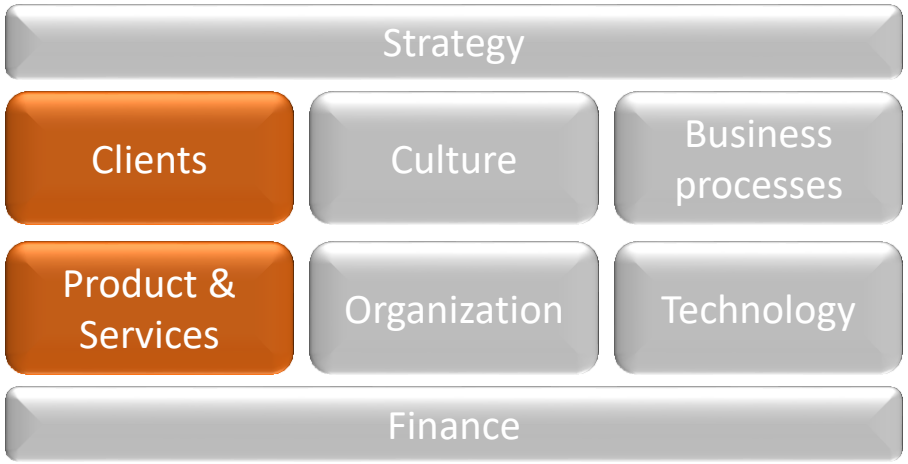
Business Analytics

UNIVERSITAT DE BARCELONA
Facultat d'Economia i Empresa

What are we going to do?

- Presentation.
- What is Business Analytics?
- Impact of Business Analytics in a Company. General approach.
  - Objectives: We will dedicate this sessions to analyze what is the impact of gathering, managing and analyzing data in the different areas of a Company
    - Strategy.
    - Technology.
    - Business Processes.
      - Internet of Thinks.
    - Clients, Products & Services. Marketing.
      - Social Media.

Clients, Product & Services. Marketing



What is Social Media Data

What is Social Media Data.

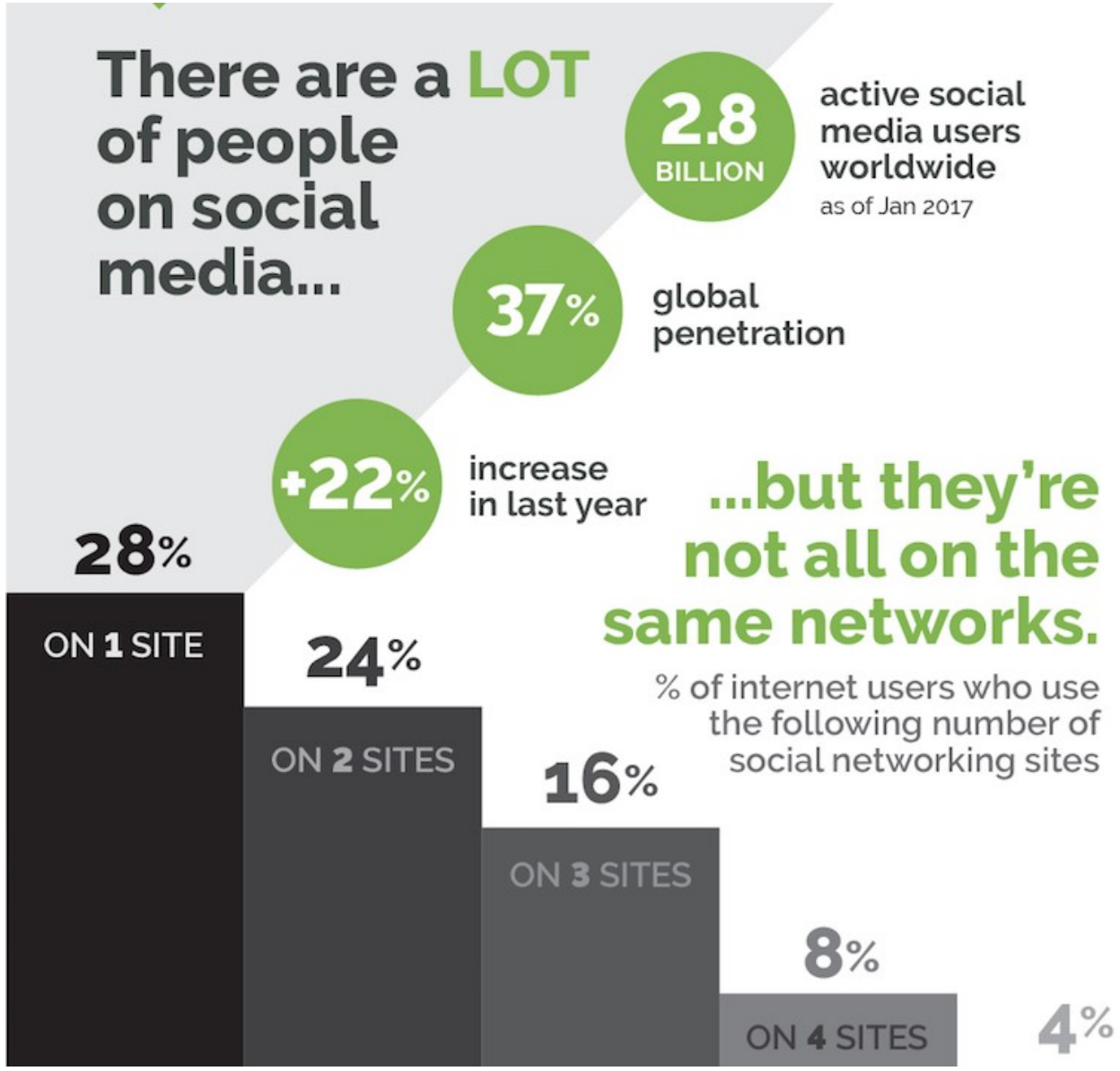
- Within online generated data, only a portion of it is considered truly social media data. There is always a grey area of debate on what social media is and what it is not. But to make it simple, we can define social media data as

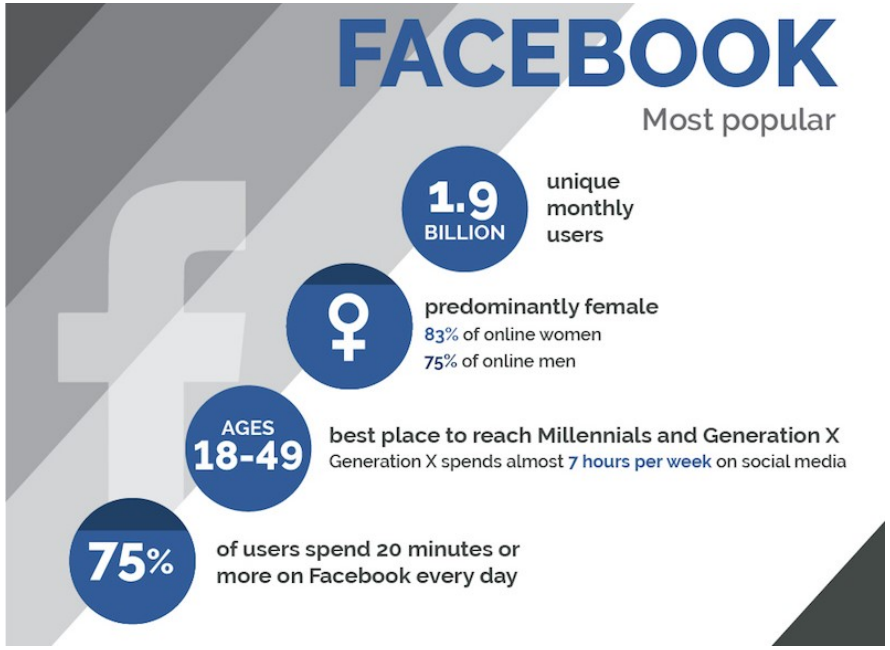
  Data generated within a self-named social media platform.

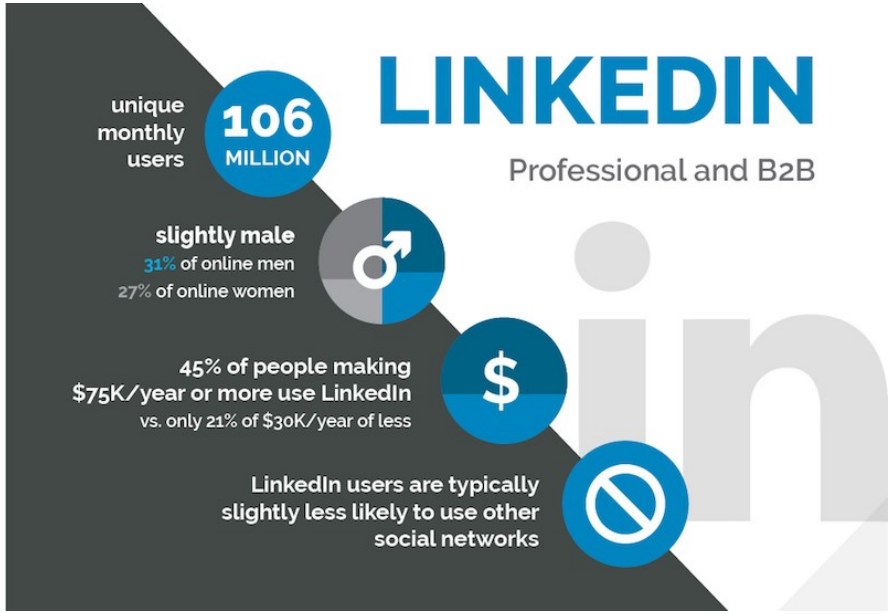
What is Social Media Data.

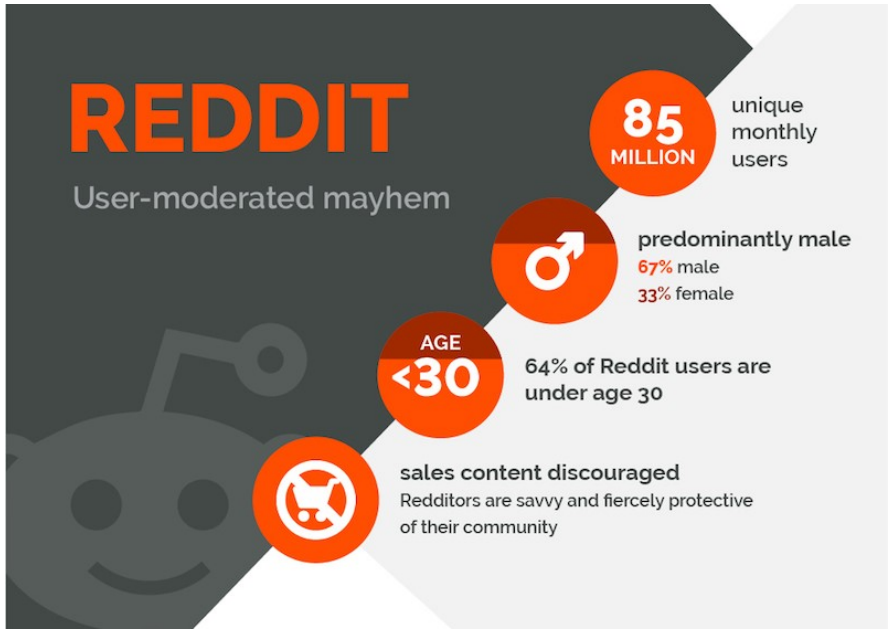
- So we rely on how the platforms present themselves and on their level of publicly available data. In such case, **excluded from our list of data sources are web sites, blogs, mobile apps that are not social platforms, streaming services not social by design, and so forth**. They are sources that do not name themselves as social media technologies. These sources are important for a digital marketing strategy, and we will look at integrated analytics as well, but focus mainly on social media data sources and metrics.
- **We include well-known platforms such as Facebook, Twitter, LinkedIn, Instagram, Google+, YouTube, Pinterest, Snapchat, and others**. The main focus is not to tackle every single specific point of every network, but instead to develop your knowledge to a level where you can tackle any of these networks as targets of your marketing strategies, and be ready for cross-network strategic planning and analysis.

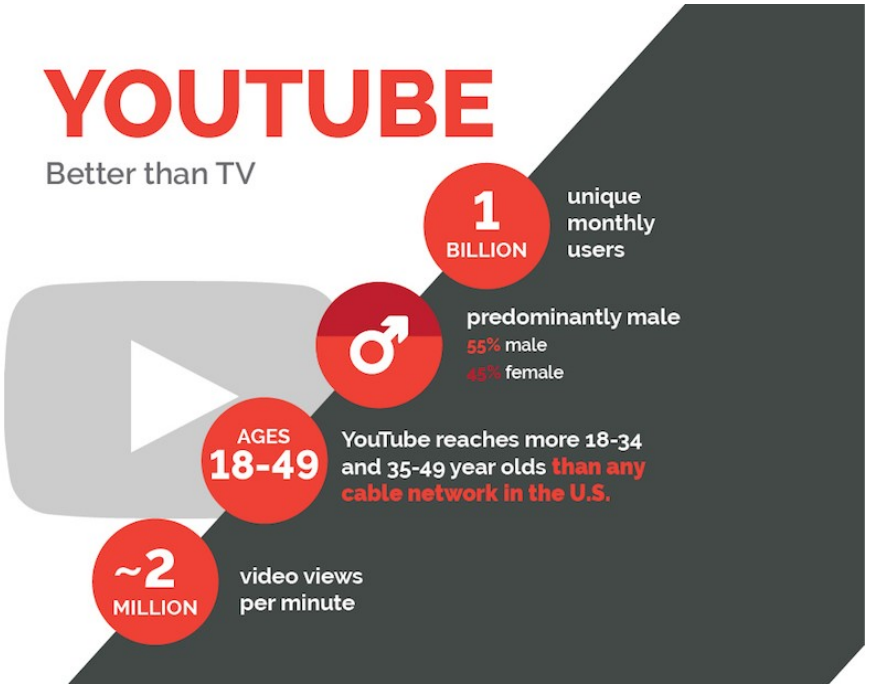
What is Social Media Data.

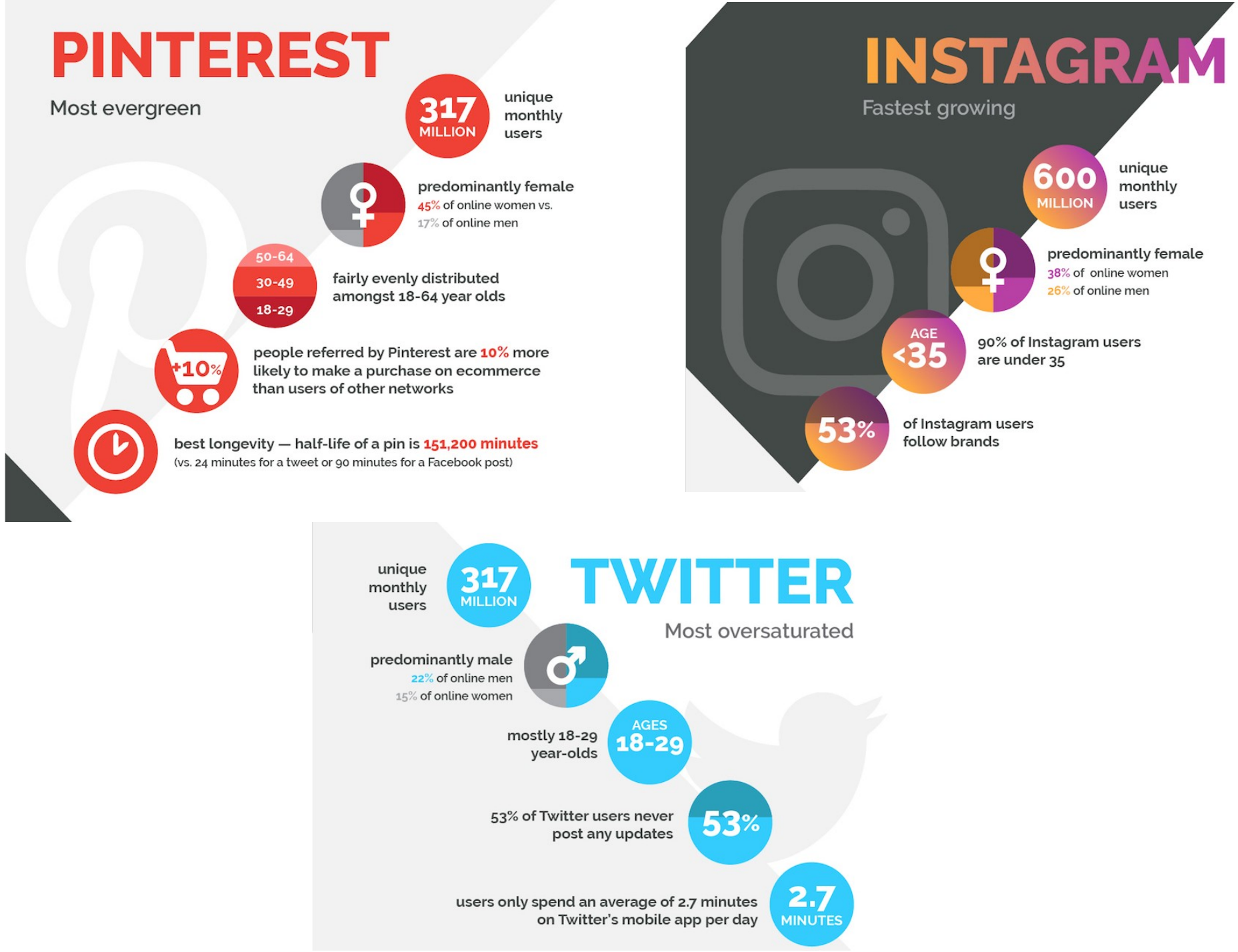



FACEBOOK
Most popular
1.9 BILLION
unique monthly users
predominantly female
83% of online women
75% of online men
AGES 18-49
best place to reach Millennials and Generation X
Generation X spends almost 7 hours per week on social media
75%
of users spend 20 minutes or more on Facebook every day


LINKEDIN
Professional and B2B
unique monthly users
106 MILLION
slightly male
31% of online men
27% of online women
45% of people making $75K/year or more use LinkedIn
vs. only 21% of $30K/year of less
LinkedIn users are typically slightly less likely to use other social networks


REDDIT
User-moderated mayhem
85 MILLION
unique monthly users
predominantly male
67% male
33% female
AGE <30
64% of Reddit users are under age 30
sales content discouraged
Redditors are savvy and fiercely protective of their community


YOUTUBE
Better than TV
1 BILLION
unique monthly users
predominantly male
55% male
45% female
AGES 18-49
YouTube reaches more 18-34 and 35-49 year olds than any cable network in the U.S.
~2 MILLION
video views per minute


PINTEREST
Most evergreen
317 MILLION
unique monthly users
predominantly female
45% of online women vs.
17% of online men
50-64
30-49
18-29
fairly evenly distributed amongst 18-64 year olds
+10%
people referred by Pinterest are 10% more likely to make a purchase on ecommerce than users of other networks
best longevity — half-life of a pin is 151,200 minutes
(vs. 24 minutes for a tweet or 90 minutes for a Facebook post)
INSTAGRAM
Fastest growing
600 MILLION
unique monthly users
predominantly female
38% of online women
26% of online men
AGE <35
90% of Instagram users are under 35
53%
of Instagram users follow brands
TWITTER
Most oversaturated
317 MILLION
unique monthly users
predominantly male
22% of online men
15% of online women
AGES 18-29
mostly 18-29 year-olds
53%
53% of Twitter users never post any updates
2.7 MINUTES
users only spend an average of 2.7 minutes on Twitter's mobile app per day

What is Social Media Data.

- But don’t forget that in general we will have to manage:
  - **Offline originated data**. Data that has been generated with no connection to the Internet, and then registered into a system, which may be accessed via the Internet later. Examples include physical retail, printed press, live events, telephone marketing and customer support, traditional television audience measurements, and so forth.
  - **Online originated data**. Created from systems connected to the Internet. Examples include web sites, e-commerce, media streaming, e-mail, mobile applications, **social media**, online devices, and so forth.

Types of Analytics in Social Media

Types of Analytics.

- To make it simple, let’s divide the analytics types into the following categories.
  - Analytics
  - Listening
  - Advertising analytics
  - CMS analytics
  - CRM Analytics

Types of Analytics.

- Analytics or Channel Analytics
  - When the source of the data is exclusively the social channels that we add into an analytics tool, that data falls into the analytics category.
  - Data sources that are included here are the content that the channel publishes, interactions related to the content published, number of followers, and some information on these followers. The social media networks offer many different connection points for their data. Some tools only connect to one or a few of these connection points.
  - Most social media networks offer some kind of analytics directly to their users. Usually, they are not complete or are not easy to work with on a professional level
- Social Media Listening
  - Keyword and Mention-Based Analysis Social media listening received this name because it relates to the analyst being able to “hear” what the market is saying about the brand via social media channels. Many marketers also refer to it as social media monitoring
  - Most listening tools add a few processes to enrich the data with more details.
    - Demographics: gender, age, location
    - Interests of people mentioning
    - Sentiment of mentions: positive, neutral, or negative
    - Influencer factors, such as the number of followers of the people making mentions, or how relevant they are to the brand based on what they talk about or the number of people interacting with their content

Types of Analytics.

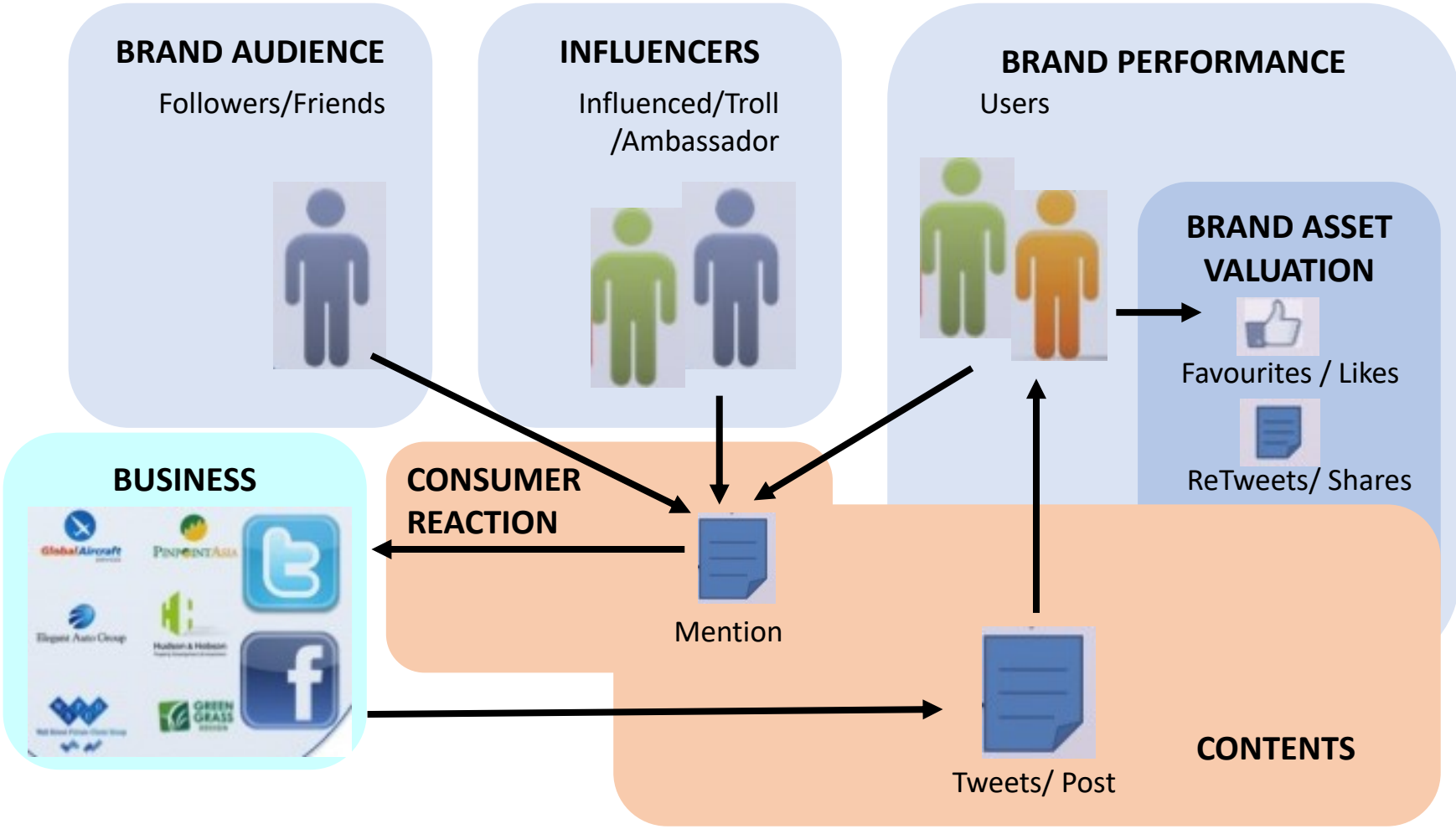
- Advertising Analytics
  - Focus on Conversions and ROI of Paid Social Media Campaigns. The information that it delivers is focused on showing to the marketer what is working best and why, based on results from the direct investment on specific content
  - It depends on the objective of an ad or paid content Social media networks offer different kinds of conversions for any paid content
    - following a profile or liking a business page
    - clicking a link in a post
    - signing up or downloading an app
- CMS Analytics
  - Another big aspect of social media is the management of content in a professional level. When we have a brand identity to maintain, many different social media channels to publish into, and an ongoing significant amount of very specific content that we wish to publish, we need tools to help us get it done
  - So we can eventually find a metric such as “engagement per team member” to see who in our team is creating the most engaging content.

Types of Analytics.

- CRM Analytics
  - Usually, when we think of CRM, or customer relationship management, we don’t immediately think of social media. This may happen because of how hard it is to maintain a long conversation through a social media channel, and how easy it is to lose track of the conversations we had. Usually, conversations related to customer support and sales that start on social media quickly move to other channels in order to be continued.
  - The metrics here vary by task, ranging from the number of questions answered by customer support and how fast the team can answer them, to the lead score in sales, where a mix of metrics are used to determine if one person is a potential client or not.

Elements of Analytics in Social Media

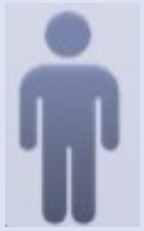
Elements of Social Media Analytics



USERS

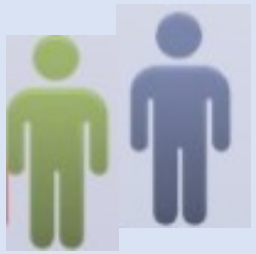
BRAND AUDIENCE

Followers/Friends



INFLUENCERS

Influenced/Troll /Ambassador



Users

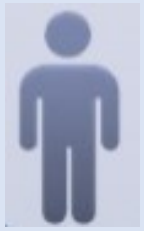
Depending on your affiliation or activity we have the following ROLES:

- USERS
  - Any ACCOUNT of a Social Media PLATFORM.
- FOLLOWERS
  - USERS who follow a certain PROFILE.
- AUTHORS
  - USER who posts a MESSAGE.

BRAND AUDIENCE

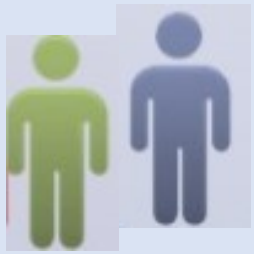
BRAND AUDIENCE

Followers/Friends



INFLUENCERS

Influenced/Troll /Ambassador



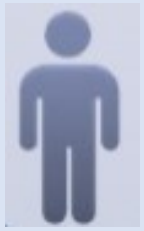
Users

- BRAND AUDIENCE. Analysis of the FOLLOWERS of a COMPETITOR'S PROFILES.
  - Scope: Focuses on analysing different aspects of the FOLLOWERS of a COMPETITOR'S PROFILES:
    - PROFILE OF A COMPETITOR'S FOLLOWERS
    - COMPARISON OF FOLLOWERS AMONG COMPETITORS
    - FOLLOWER MESSAGES
- UNIQUE FOLLOWERS Of all the FOLLOWERS of a COMPETITOR'S PROFILES are counted without repeats.
- PROFILE SPECIFICITY Number of FOLLOWERS who only follow the PROFILE of a COMPETITOR.

INFLUENCERS

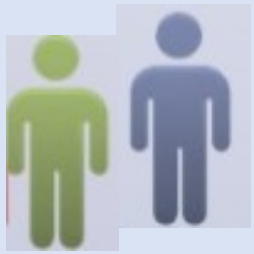
BRAND AUDIENCE

Followers/Friends



INFLUENCERS

Influenced/Troll /Ambassador



Users

- INFLUENCERS Analysis of AUTHORS who make DIRECT or INDIRECT MENTIONS of a COMPETITOR.
  - Scope: Focuses on analysing different aspects of AUTHORS who make MENTIONS of a COMPETITOR.
    - INFLUENCERS BY SCOPE OR ACTIVITY
    - RELEVANCE ANALYSIS
    - SENTIMENT ANALYSIS
    - KLOUT INDEX
    - AMBASSADORS, TROLLS

USERS


Jordi Évole
@jordievole
Uno de Salvados @salvadostv
Barcelona · elterrat.com
6.664
TWEETS
531
SIGUIENDO
886.199
SEGUIDORES
Seguir
Seguidores Todos / Tú conoces
Cristian Pérez @Crishunder
Seguir
USER
FOLLOWER
AUTHOR
TWEET
Tweets
Jordi Évole @jordievole 13 h
Arrivals/Llegadas/Chegadas pic.twitter.com/W0bRKCqNf0
Ver foto
Jordi Évole @jordievole 13 h
Jajajajaja!!! "@loulancho: #habemusPampa"
Abrir

USERS


facebook
Jordi Évole
M'agrada
Missatge
Personatge públic
http://www.salvados.lasexta.com/
413 mil
Fotos
Vídeos
Jordi Évole
dimarts
AUTHOR
Jajajajajaja!!!
"@la_podadora: OLA K AC MILAN"
POST
M'agrada · Comenta-ho · @jordievole on Twitter
A 2.617 persones els agrada.
Mostra els comentaris anteriors
2 de 86
USER
FRIEND

USERS

Depending on the scope and tone of their activity, we have the following ROLES:

- INFLUENCER
  - AUTHOR who mentions a KEYWORD or mentions a PROFILE.
  - They must have a large number of FOLLOWERS to be considered INFLUENCERS.
- AMBASSADOR
  - INFLUENCER where most of their MENTIONS are in POSITIVE TONE.
- TROLLS
  - INFLUENCER where most of their MENTIONS are in NEGATIVE TONE

USERS


AMBASSADOR
NEUTRAL TONE
FOLLOWER
INFLUENCER
DIRECT MENTION
INDIRECT MENTION
somos.som producciones audiovisuales
11.000
GRACIAS
Julia Otero
Missatge
11.247
Julia Otero ha compartit la foto de D ías Baixas
dimarts
El otro lado de la redacción provisional de JELO en Pontevedra. Simulando que trabajamos.
Mostra la traducció
Continúa la programación de Onda Cero en Pontevedra con motivo del 25 aniversario de la Denominación de Origen D.O. Rías Baixas hoy desde Pontevedra. Ahora con Julia en la Onda, su equipo ultima detalles en la redacción improvisada en el Pazo de Mugartegui
Més informació · Suggereix una edició
Fotos
M'agrada
Autógrafo Julia
Videografia

BUSINESS

- PLATFORM
  - Definition: Social Media application. Facebook, Twitter, etc.
- COMPETITOR
  - Definition: Logical grouping of PROFILES, usually companies, that uses PLATFORMS for their Social Media strategy.
- PROFILE
  - Definition: Channel created within a PLATFORM to execute a part of the Social Media strategy.

BUSINESS


twitter
PLATFORM
PROFILE
DIRECT MENTION
INDIRECT MENTION
KEYWORD
POST
La Chocita del Loro
@chocitadelloro
Seguir
Oferta loca de la @chocitadelloro de #GranVia es.letsbonus.com/espectaculos/m ... Si haces RT y eres follower sorteo de 2 entradas a elegir
Responder
Retwittear
Favorito
Más
275
RETWEETS
6
FAVORITOS
07:37 - 14 dic 2012

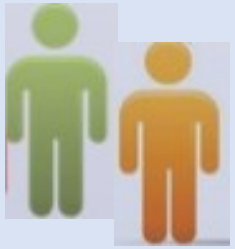
BUSINESS


facebook
PLATFORM
PROFILE
La Chocita del Loro
Agrada a 3.111 · 73 en parlen · 1.728 hi han estat
M'agrada
Missatge
Arts i entreteniment · Afegeix una categoria
Gran Via 70, Hermosilla 77 y Doctor Urquiola 23,...
91.524.25.03 / 91.435.30.46 / 91.466.75.04
3.111
Compra tus entradas
Tickets
Més informació · Suggereix una edició
DIRECT MENTION
POST
Carlos Ramos ▶ Corta el Cable Rojo
4 febrer
Hoy estamos más compenetrados que nunca. Ven a vernos a La La Chocita del Loro de Gran Vía a las 21:30.
Mostra la traducció
baños-2.JPG
db.tt
Shared with Dropbox
M'agrada · Comenta-ho · Comparteix
5
KEYWORD
INDIRECT MENTION

BRAND PERFOMANCE

BRAND PERFORMANCE

Users

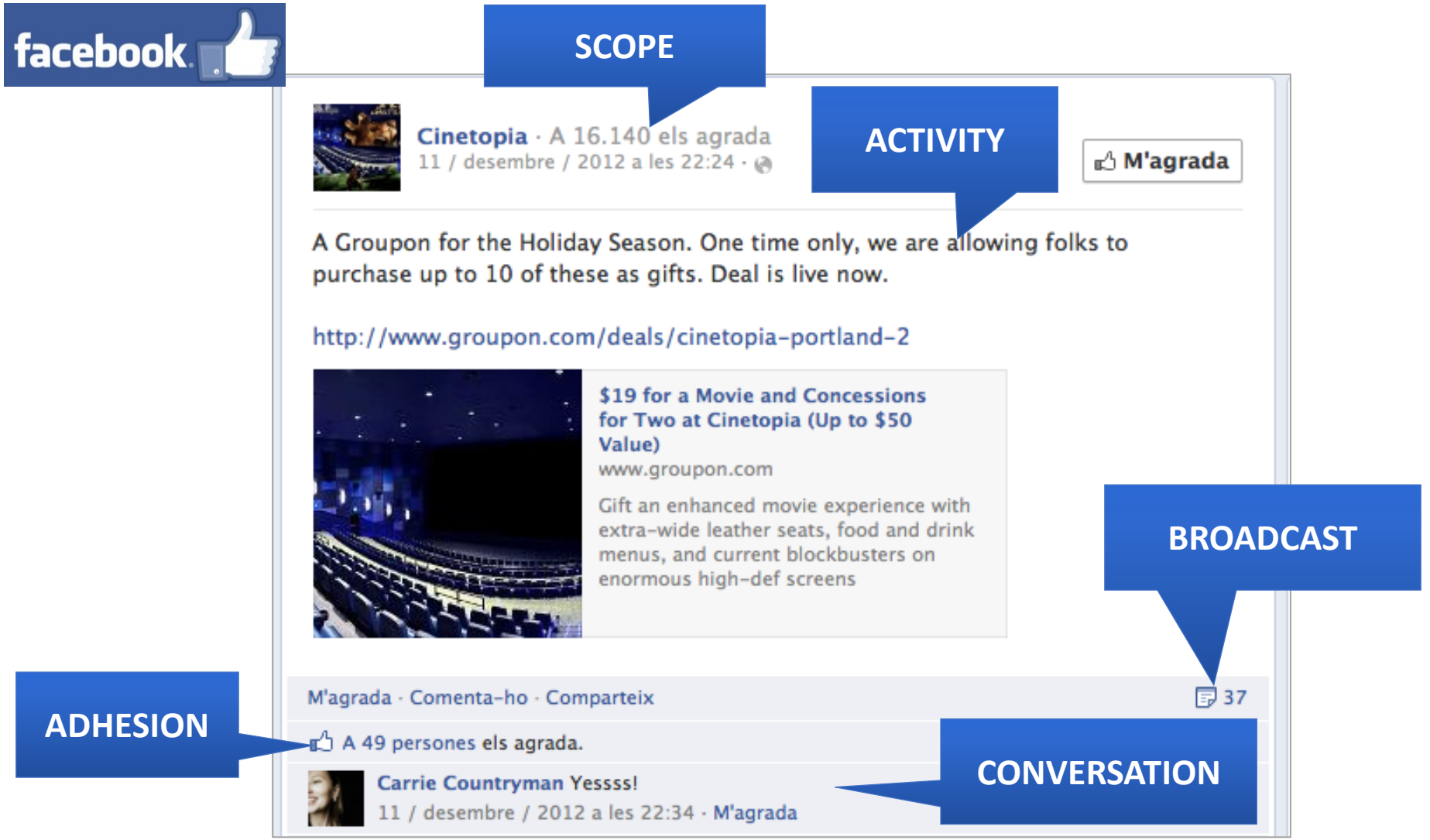


Comparison of the production impact of each COMPETITOR's PROFILES.

- SCOPE
  - Definition: Number of FOLLOWERS of a PROFILE.
  - When a PROFILE publishes a MESSAGE it will reach at least as many FOLLOWERS as that PROFILE has. This will be the MESSAGE SCOPE.
- ACTIVITY
  - Definition: number of MESSAGES published by each PROFILE.
- BROADCAST
  - Definition: Number of times a MESSAGE from a PROFILE is broadcasted by USERS.
- ADHESION
  - Definition: Number of times a MESSAGE is liked by any USER.
- CONVERSATION
  - Definition: Number of REPLIES on the MESSAGES posted by a PROFILE.

BRAND PERFOMANCE

BRAND PERFOMANCE



BRAND PERFOMANCE

- ENGAGEMENT: Direct interaction of the community with our strategy.
  - What percentage of users actually interact:

$$\frac{\text{ADHESION + CONVERSATION + BROADCAST}}{\text{SCOPE}} \text{ X 100}$$

- ROI: Return on investment in Social Media.
  - Relationship of Social Media INDICATORS with more traditional MARKETING.
    - Leads = SCOPE. CPL (Cost per Lead): cost of acquiring a contact.
    - Impressions = SCOPE x ACTIVITY. CPM (Cost per thousand): cost of producing one thousand impacts on the target audience.
    - Interaction = BROADCAST, ADHESION, CONVERSATION CPI (Cost of Interaction) Value of interactions with consumers.
    - Sponsorship = MENTIONS from third parties. The cost is variable depending on the relevance of the institution or person.

BRAND ASSET VALUATION

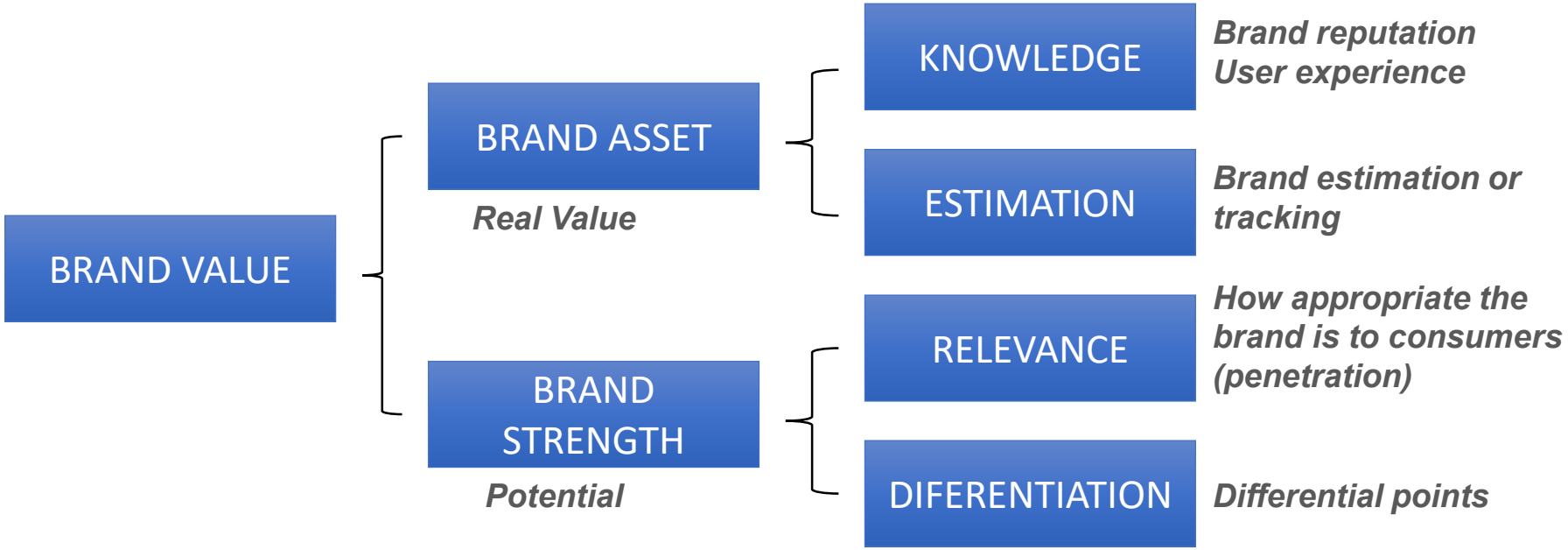
BRAND ASSET VALUATION: Analysis that measures the value of each COMPETITOR with respect to the others.

BRAND ASSET VALUATION

Favourites / Likes

ReTweets/ Shares

Reply



CONSUMER REACTION

CONSUMER REACTION

Mention

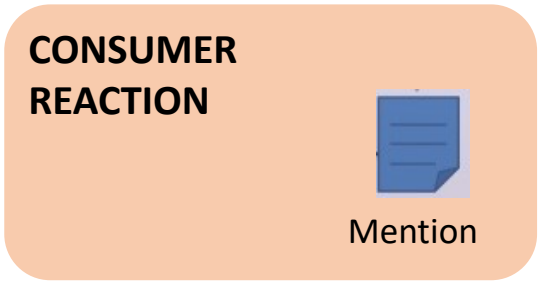
CONSUMER REACTIONS Analysis of activity and MENTIONS made by USERS.

- Scope: Analyzes messages made by USERS where comments are made:
  - DIRECTLY a PROFILE
  - INDIRECTLY a KEYWORD

- MENTION
  - Is a MESSAGE where its AUTHOR is any USER of the PLATFORM, and in the MESSAGE a COMPETITOR'S PROFILE or a COMPETITOR'S KEYWORD is directly mentioned or quoted.

- DIRECT MENTION.
  - MENTION where you explicitly quote the PROFILE or write on their wall to make it clear that you want to refer to the PROFILE and that the PROFILE receives the MESSAGE.

CONSUMER REACTION

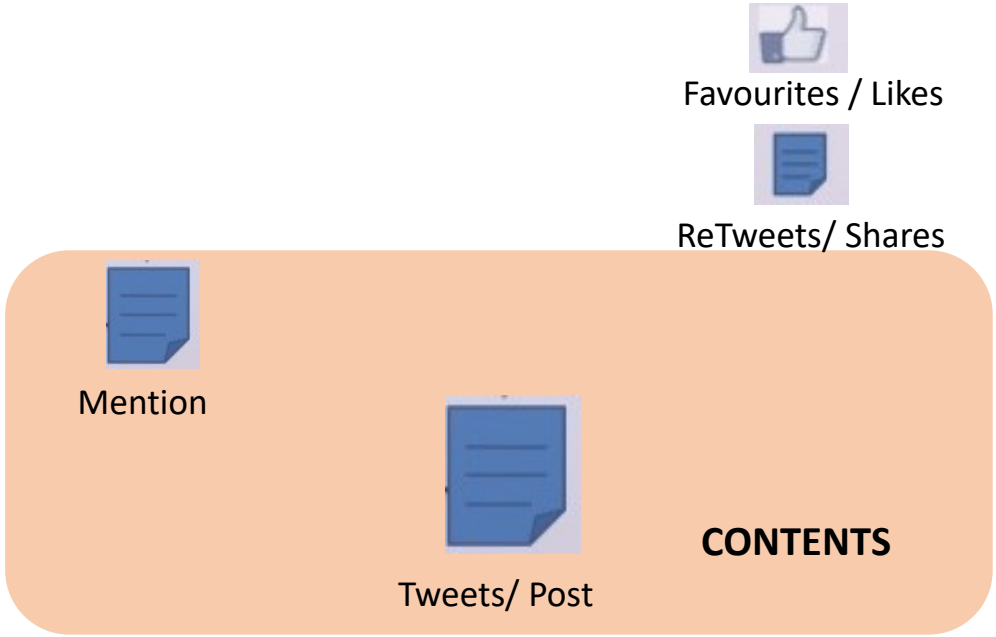


- INDIRECT MENTION.
  - MENTION where a KEYWORD is explicitly quoted that is followed by a COMPETITOR. The AUTHOR of the MESSAGE does not seek to refer directly to the PROFILE and the PROFILE receives the MESSAGE.

- FEELING
  - Depending on the content of the MESSAGE, a MENTION has a SENTIMENT which can be POSITIVE, NEUTRAL or NEGATIVE.

- Categories:
  - POSITIVE: When the MENTION speaks positively about the PROFILE it mentions.
  - NEUTRAL: When the MENTION simply comments without wanting to influence positively or negatively the PROFILE it mentions.
  - NEGATIVE: When the MENTION speaks negatively about the PROFILE it mentions.

CONTENTS



CONTENT Analysis of the entire CONTENT of MESSAGES that have INDIRECT MENTIONS (that a KEYWORD of a COMPETITOR has been referred to in the MESSAGE).

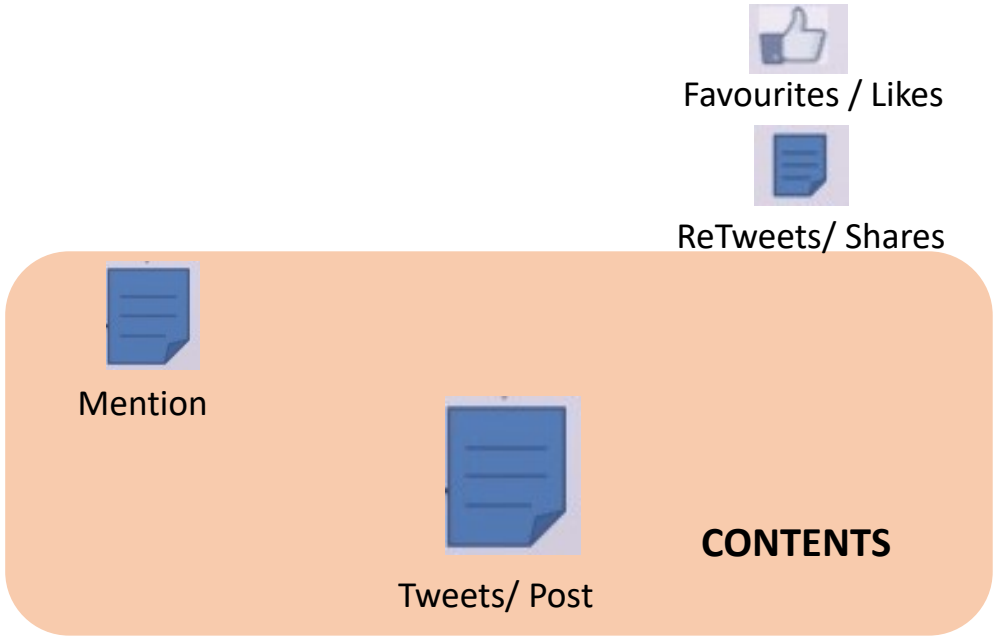
- Scope: Focuses on analyzing the CONTENT of MESSAGES with INDIRECT MENTIONS to detect the words used in them, which we will call TOPICS.

- TOPIC.
  - Simple or compound words that appear in the CONTENT of MESSAGES that have INDIRECT MENTIONS of KEYWORDS followed by COMPETITORS.

- TRENDING TOPIC
  - TOPIC most repeated according to temporality and location.

- CATEGORIES
  - CATEGORIES are used to group TOPICS.

CONTENTS



- LOCALIZATION Analysis of the location of LOCALIZED MESSAGES.
  - Scope: Only the TWITTER PLATFORM is analyzed.
  - It focuses on performing a geospatial analysis on a MAP of TWITTER MESSAGES that have their sending position reported.

- LOCALIZED MESSAGE
  - MESSAGE that has shared its location at the time of posting.
- COUNTRY
  - COUNTRY from where a MESSAGE is posted.
- POPULATION
  - POPULATION from where a MESSAGE is posted.
- LANGUAGE
  - LANGUAGE in which a MESSAGE is written.

Examples and applications

Examples and applications.

- The social big data analysis can be applied to social media data sources for discovering relevant knowledge that can be used to improve the decision making of individual users and companies.
- This field includes methodologies that can be applied to different areas such as e-commerce, marketing, security, and healthcare; more recent methodologies have been applied to treat social big data.
- We will mention a short descriptions of some applications of these methodologies in domains that intensively use social big data sources for business analytics.
  - Stock Market.
  - Marketing
  - Crime Analysis
  - Epidemic Intelligence

Stock Market: Social Media Analysis applied to Stock Market

- The stock market is influenced by several factors, such as macroeconomics, regulatory, purely speculative ones, and many others. However, one of the most relevant and meaningful is the general opinion and the overall investors' sentiment, i.e., what investors think about a certain firm and, as a consequence, about the relative stock.
- This investors' sentiment is here proxied by the Twitter content, and the study sums up to the recent outbreak of works that exploit sentiment analysis and Twitter data for stock market predictions. The sample analysed concerns three major technology companies over a two-months period, on a minute basis. Using microblogging activities and a scoring algorithm for each tweet, it was possible to formulate interesting forecasting models identifying a new set of variables and indicators of the stock market future movements.
- A selection model has been used to implement the study, and the evidences found were encouraging, since it has been possible to draw the conclusion that this new source of data may increase the explanatory power of financial forecasting models. More in detail:
  - It looks like that the average sentiment associated to any tweet is not so relevant as expected in prediction terms
  - While the posting volume has a greater forecasting power and it could be used to augment the models.
- Further implementations could be studied in the next future, such as considering longer timeframes, different companies and sectors, or analyzing special situations such as IPO and company's announcements (dividends, etc.). Of particular interest would also be studying the structure of the network who is talking about a certain stock or firm and assess how this affect the company's evaluation on the stock market.

- Working Paper: "Can Twitter proxy de investors sentiment"
- Working Paper: "Twitter mood predicts the stock market"

Marketing.

- Big social media analytics and cloud computing offer a unique opportunity for businesses to obtain opinions from a vast number of customers, improving traditional strategies.
- Social network analysis extracts user intelligence and can provide firms with the opportunity for generating more targeted advertising and marketing campaigns.
  - A case study using Facebook to determine users perceptions regarding Facebook ads found that most of the participants perceived the ads on Facebook as annoying or not helpful for their purchase decisions.
  - However another study show how ads placed on users social streams that have been generated by the Facebook tools and applications can increase the number of visitors and the profit and ROI of a Web-based platform.

Marketing.

- A study of microblogging (Twitter) utilization as an eWOM (electronic word-of-mouth) advertising mechanism analyses the range, frequency, timing, and content of tweets in various corporate accounts.
  - The results obtained show that 19% of microblogs mention a brand.
  - Of the branding microblogs, nearly 20% contained some expression of brand sentiments.
  - Therefore, the study conclude that microblogging reports what customers really feel about the brand and its competitors in real time, and it is a potential advantage to explore it as part of companies overall marketing strategies.
- Customers brand perceptions and purchasing decisions are increasingly influenced by social media services, and these offer new opportunities to build brand relationships with potential customers.
  - Another study that uses Twitter data is presented to forecast box-office revenues for movies.
  - The study show how a simple model built from the rate at which tweets are created about particular topics can outperform market based predictors.

Targeted advertising on Facebook

- Recently, social network marketing has gained tremendously in popularity. There are increasing numbers of web-based companies that focus their marketing strategies on social network platforms such as Facebook, Twitter, etc.
- Even if social network marketing is perceived as a huge success, little or nothing is known about how well such social network-based marketing campaigns perform. An study analyses the results of an ad-driven social network-based marketing campaign centred on the social network Facebook. In particular, it demonstrates in the extent to which ads placed on the user's social stream and generated by the Facebook tools and applications can increase the number of visits, profit or ROI of a web-based platform called VirWoX.

- Working Paper: Social stream marketing on Facebook: a case study

Using Twitter to forecast box-office revenues for movies

- The working paper is a study that shows how social media can be utilized to forecast future outcomes.
  - Specifically, using the rate of chatter from almost 3 million tweets from the popular site Twitter, the authors have constructed a linear regression model for predicting box-office revenues of movies in advance of their release.
  - The study shows that the results outperformed in accuracy those of the Hollywood Stock Exchange and that there is a strong correlation between the amount of attention a given topic has (in this case a forthcoming movie) and its ranking in the future.
  - The study also analysed the sentiments present in tweets and demonstrated their efficacy at improving predictions after a movie has released. While this study focuses on the problem of predicting box office revenues of movies for the sake of having a clear metric of comparison with other methods, this method can be extended to a large panoply of topics, ranging from the future rating of products to agenda setting and election outcomes. At a deeper level, this work shows how social media expresses a collective wisdom which, when properly tapped, can yield an extremely powerful and accurate indicator of future outcomes.

  - Working Paper: Predicting the Future With Social Media

Crime Analysis.

- Criminals tend to have repetitive pattern behaviours, and these behaviours are dependent upon situational factors. That is, crime will be concentrated in environments with features that facilitate criminal activities.
- The purpose of crime data analysis is to identify these crime patterns, allowing for detecting and discovering crimes and their relationships with criminals. The knowledge extracted from applying data mining techniques can be very useful in supporting law enforcement agencies. Communication between citizens and government agencies is mostly through telephones, face-to-face meetings, email, and other digital forms. Most of these communications are saved or transformed into written text and then archived in a digital format, which has led to opportunities for automatic text analysis using NLP techniques to improve the effectiveness of law enforcement.
  - Filtering reports and identifying those that are related to the same or similar crimes can provide useful information to analyse crime trends, which allows for apprehending suspects and improving crime prevention.
  - Traditional crime data analysis techniques are typically designed to handle one particular type of dataset and often overlook geospatial distribution. Geographic knowledge discovery can be used to discover patterns of criminal behaviour that may help in detecting where, when, and why particular crimes are likely to occur.
  - Another analytical technique that is now in high use by law enforcement agencies to visually identify where crime tends to be highest is the hotspot mapping. This technique is used to predict where crime may happen, using data from the past to inform future actions. Each crime event is represented as a point, allowing for the geographic distribution analysis of these points.
  - Finally, the use of data mining in fraud detection is very popular, and there are numerous studies on this area. ATM phone scams are one well-known type of fraud.

Identify discussion topics across a city in the United States to predict crimes

- Twitter is used extensively in the United States as well as globally, creating many opportunities to augment decision support systems with Twitter-driven predictive analytics. Twitter is an ideal data source for decision support: its users, who number in the millions, publicly discuss events, emotions, and innumerable other topics; its content is authored and distributed in real time at no charge; and individual messages (also known as tweets) are often tagged with precise spatial and temporal coordinates.
- This study presents research investigating the use of spatiotemporally tagged tweets for crime prediction. We use Twitter-specific linguistic analysis and statistical topic modeling to automatically identify discussion topics across a major city in the United States. We then incorporate these topics into a crime prediction model and show that, for 19 of the 25 crime types we studied, the addition of Twitter data improves crime prediction performance versus a standard approach based on kernel density estimation.

- Working Paper: Predicting crime using Twitter and kernel density estimation

Identification of fraudulent financial statements

- This case explores the effectiveness of Data Mining (DM) classification techniques in detecting firms that issue fraudulent financial statements (FFS) and deals with the identification of factors associated to FFS. In accomplishing the task of management fraud detection, auditors could be facilitated in their work by using Data Mining techniques. This study investigates the usefulness of Decision Trees, Neural Networks and Bayesian Belief Networks in the identification of fraudulent financial statements. The input vector is composed of ratios derived from financial statements. The three models are compared in terms of their performances.

- Working Paper: Data Mining techniques for the detection of fraudulent

Epidemic intelligence

- Epidemic intelligence can be defined as the early identification, assessment, and verification of potential public health risks [106] and the timely dissemination of the appropriate alerts. This discipline includes surveillance techniques for the automated and continuous analysis of unstructured free text or media information available on the Web from social networks, blogs, digital news media, and official sources

Detecting influenza epidemics by analyzing Twitter messages

- Rapid response to a health epidemic is critical to reduce loss of life. Existing methods mostly rely on expensive surveys of hospitals across the country, typically with lag times of one to two weeks for influenza reporting, and even longer for less common diseases. In response, there have been several recently proposed solutions to estimate a population's health from Internet activity, most notably Google's Flu Trends service, which correlates search term frequency with influenza statistics reported by the Centers for Disease Control and Prevention (CDC). In this paper, we analyse messages posted on the micro-blogging site Twitter.com to determine if a similar correlation can be uncovered. We propose several methods to identify influenza-related messages and compare a number of regression models to correlate these messages with CDC statistics.

- Working Paper: Towards detecting influenza epidemics

Assess disease outbreaks from tweets

- Data mining social media has become a valuable resource for infectious disease surveillance. However, there are considerable risks associated with incorrectly predicting an epidemic. The large amount of social media data combined with the small amount of ground truth data and the general dynamics of infectious diseases present unique challenges when evaluating model performance. In this paper, we look at several methods that have been used to assess influenza prevalence using Twitter. We then validate them with tests that are designed to avoid and illustrate issues with the standard k-fold cross validation method. We also find that small modifications to the way that data are partitioned can have major effects on a model's reported performance.

- Working Paper: Validating Models for Disease Detection Using Twitter


BIG DATA is GOING TO IMPACT US SOON !
CALL BRUCE WILLIS !!
DATA
www.nuxeo.com
Soulcié

Key questions.

- Data availability and data sources. What data do you think should be available, which is important to effectively plan on how far you can go with your analytics process.
- Knowledge beyond social media. Do you have a global view of all marketing assets and strategies in place, and even of a marketing roadmap if such knowledge is available? What information you need to punt in place to develop a Social Media Analytics strategy?
- Tools and technology preparation. Do you have all the necessary technology in place, or at least ready to be activated? This includes eventual customization of metrics and dashboards, and integration of data.
- Team preparation. Are you sure that the team is ready to implement the strategy? Education and training can be added at this stage. Plan if necessary.
- Goals and objectives. Have you defined the major goals and all intermediate objectives related to the strategy? What are them?

Thank You

UNIVERSITAT DE BARCELONA

Facultat d'Economia i Empresa