Market Sentiment Analysis

Prepared for the purpose of the Marketing Analytics subject

Contributors:

Hendro Lim, Jose Nelson Zepeda and Yiqing Hu

BUSINESS CASE: AGGREGATE VIEW

zalando

Company Profile:

- Zalando is an e-commerce company based in Berlin

- The company was established in 2008 in Germany by Robert Gentz and David Schneider

- Initially focusing on the sale of footwear but has since included other area of fashion and sports

Competition:

- Based on **datafox.com**, the two main competitors for Zalando are: **Yoox** and **Asos**

- Currently Zalando operates in **15 different countries**; Asos operates in 7 countries, while Yoox in 6 countries

ZALANDO STATISTICS:

22,639 Traffic Rank

64.30% Bounce Rate

3.0 B 2015 Revenue

Comparison Metrics

	Traffic Rank*	Bounce Rate*	2015 Revenue
Asos	334	25.60%	1.7 B
Yoox	2,524	30.40%	1.7 B

Zalando has the highest revenue due to its **larger operational scope**; Lower traffic rank and bounce rate indicate **potential improvement opportunities**

* Traffic rank and bounce rate figure are obtained from **Alexa** sites comparison

BUSINESS CASE: BACKGROUND AND OBJECTIVE

Project Background:

- Given the aggregate revenue size there are still opportunities to improve customers' satisfaction level

- Italy, as one of its biggest customer base*, becomes the main focus for achieving Q4 FY17 Financial Forecast

Scope:

- **Italian market** as one of its biggest customer base

Objective:

- **Meet the Q4 FY17 Financial Forecast** by understanding **Italian market customers' voice** through market sentiment analysis

HELSINKI
Technology hub

DUBLIN
Technology hub

GRYFINO (near Szczecin)
Fulfillment center

MOISSY-CRAMAYEL (near Paris)
Fulfillment center

STRADELLA
Fulfillment center

zalando

* Assumption of Italy as one of the biggest customer base country is based on the fact that Zalando built its fulfilment center in Italy

Market sentiment analysis is critical to perform at this moment in order to be able to understand the **Italian market customers' voice**

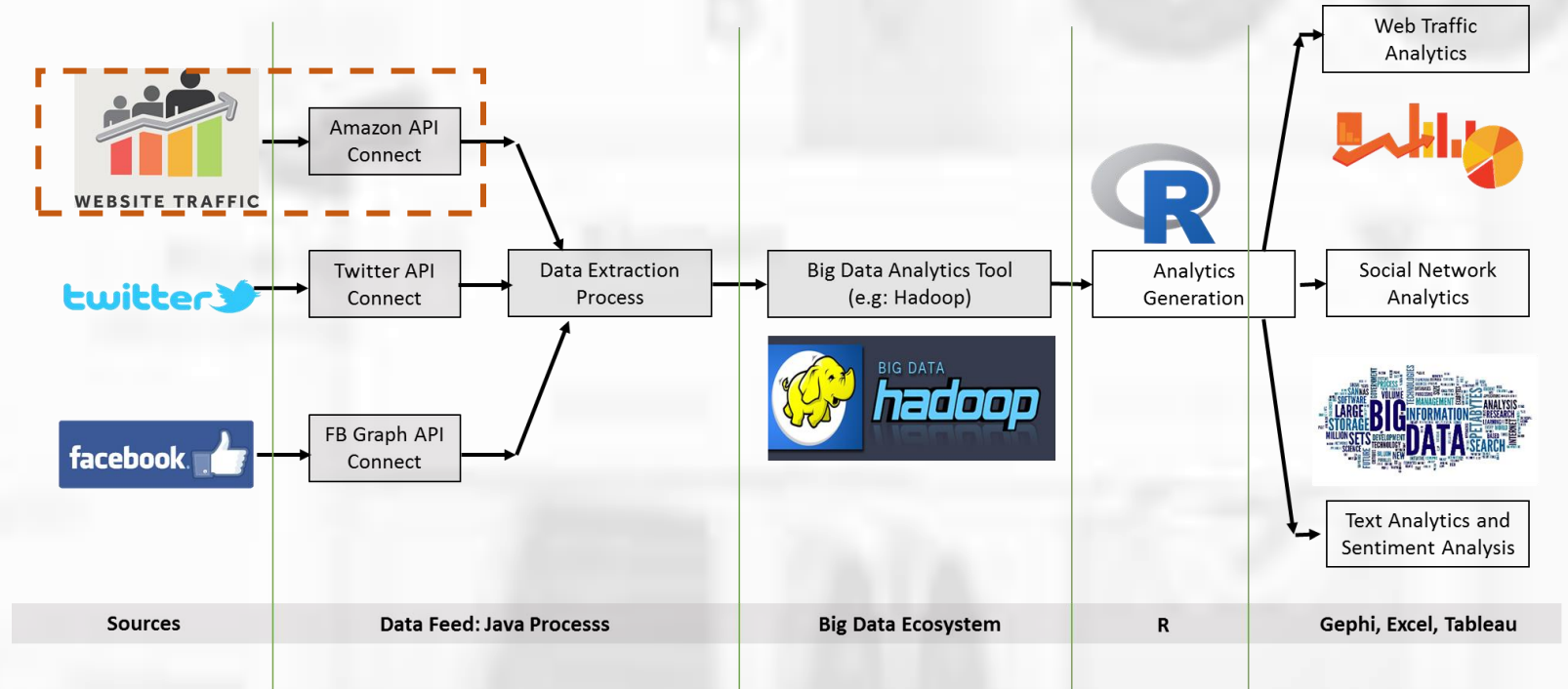# High level architecture for data analysis:
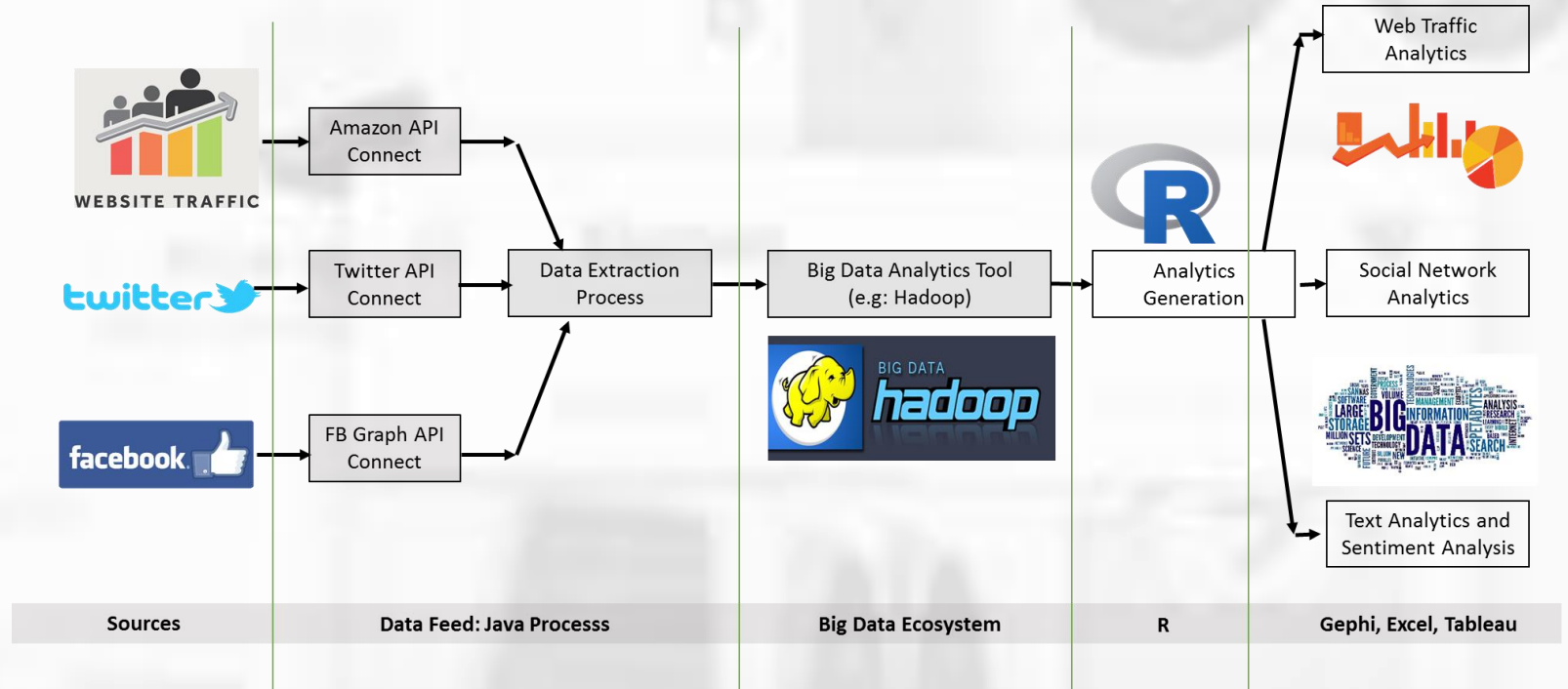
Note:

In our analysis we collect all the data separately through each of their APIs (Web traffic through **Alexa**, Twitter through **Socioviz** and Facebook through **Netvizz**). In the production environment, we recommend to use more advanced system, such as **Hadoop**, in order to get integrated figure

1

ALEXA – Web Traffic and Demographic Analysis

Leveraging **Alexa site comparison tools**, we able to gain valuable insights on our Italian customers, which may then be leveraged to create relevant strategy

High level architecture for data analysis:

| Sources | Data Feed: Java Processs | Big Data Ecosystem | R | Gephi, Excel, Tableau |

Note:

In our analysis we collect all the data separately through each of their APIs (Web traffic through **Alexa**, Twitter through **Socioviz** and Facebook through **Netvizz**). In the production environment, we recommend to use more advanced system, such as **Hadoop**, in order to get integrated figure

ZALANDO.IT STATISTICS:

Traffic Rank

Bounce Rate

134
23.60%

As the comparison, In Italy, Asos ranked **370** and Yoox ranked **416** on the traffic. For the bounce rate, Asos has **25.60%**, while Yoox has **30.40%**
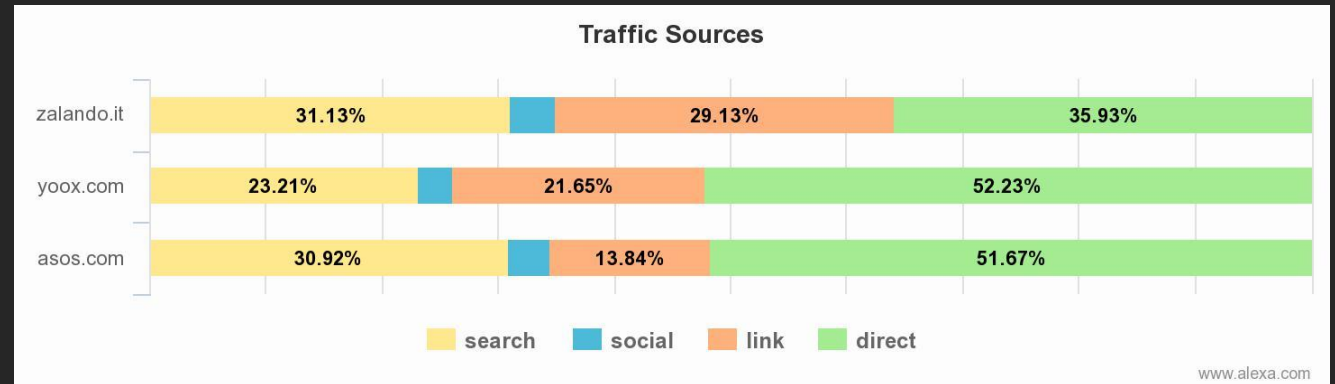
About Italian Market

- In Italy, Zalando operates under **Zalando.it** domain, which has its own unique content

- The system structure is totally different with its two main competitors, which have general content and do adjustment only on the language (**Asos.com/it** and **Yoox.com/it**)

- Based on Alexa traffic rank, **Zalando.it** has the **best traffic** and **bounce rate** compare with its competitors

ZALANDO ITALIA INSIGHTS:

ENGAGEMENT METRICS AND TRAFFIC SOURCES

8.71 Pageviews per user

Time on site **8.8** minutes

Traffic Sources

	search	social	link	direct
zalando.it	31.13%		29.13%	35.93%
yoox.com	23.21%		21.65%	52.23%
asos.com	30.92%		13.84%	51.67%

www.alexa.com

Zalando.it traffic sources is contributed mostly through **direct access** to the website and **search** the website through search engine
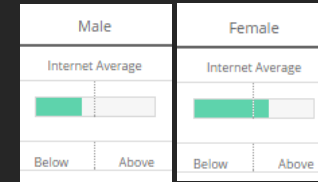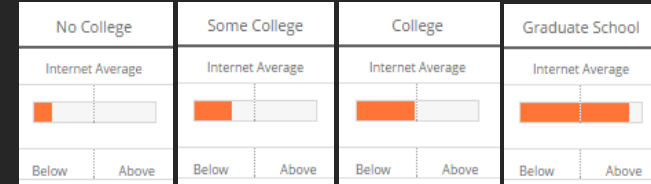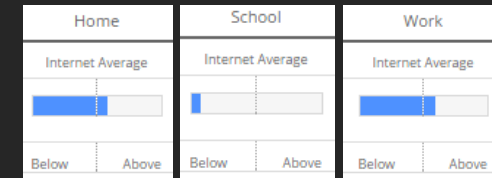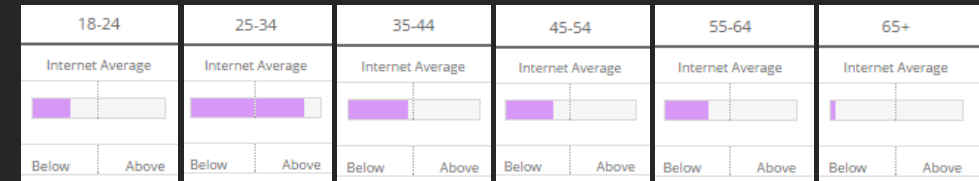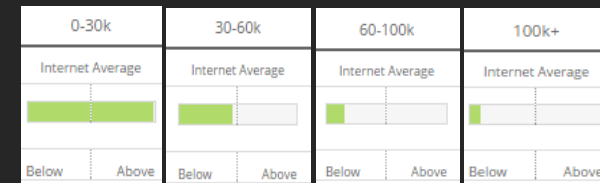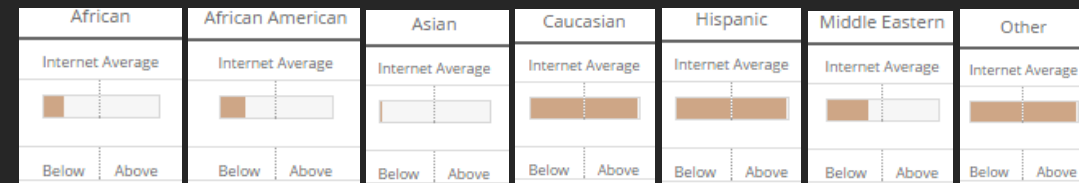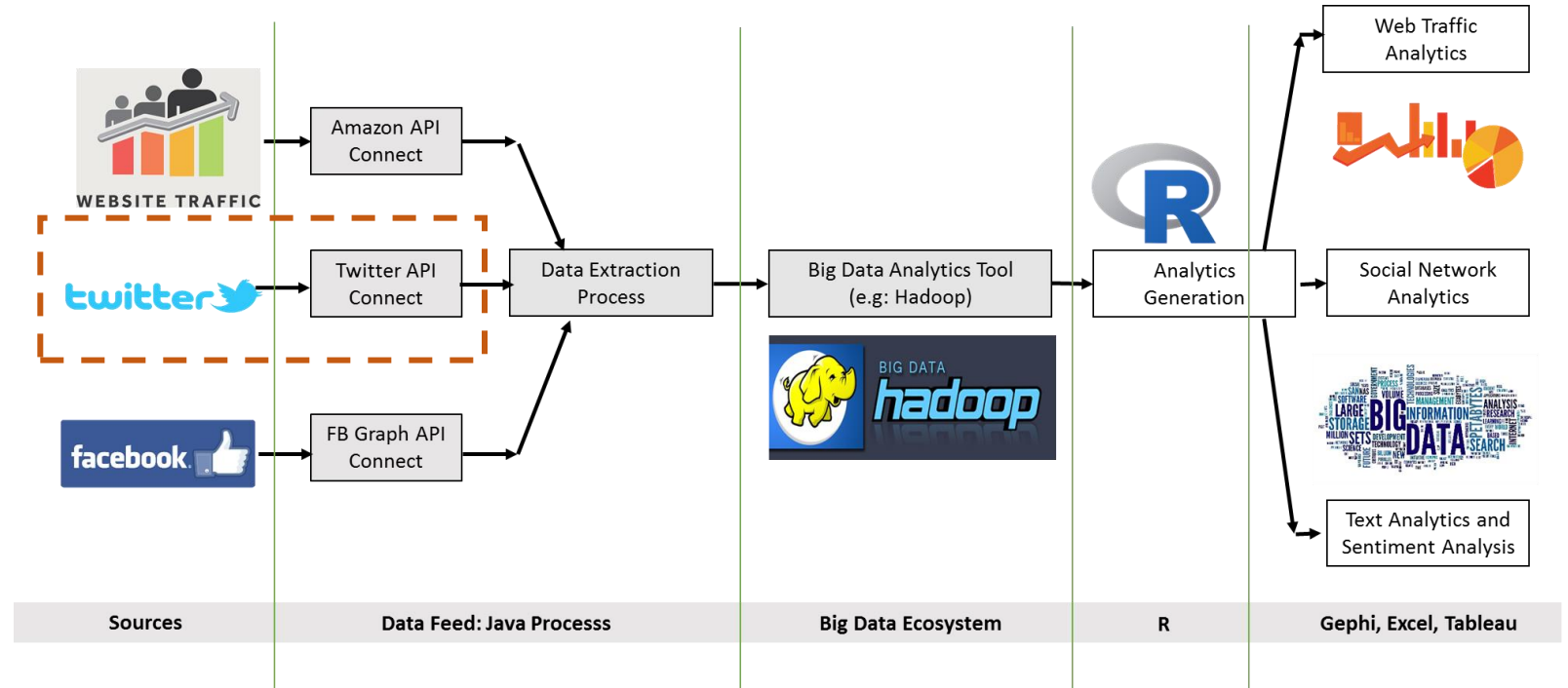
2

SOCIOVIZ.NET – Twitter Social Network Analysis

Through **socioviz.net**, we able to analyse mentions and hashtag to identify key influencers, opinions and contents, which will be projected in a network graph
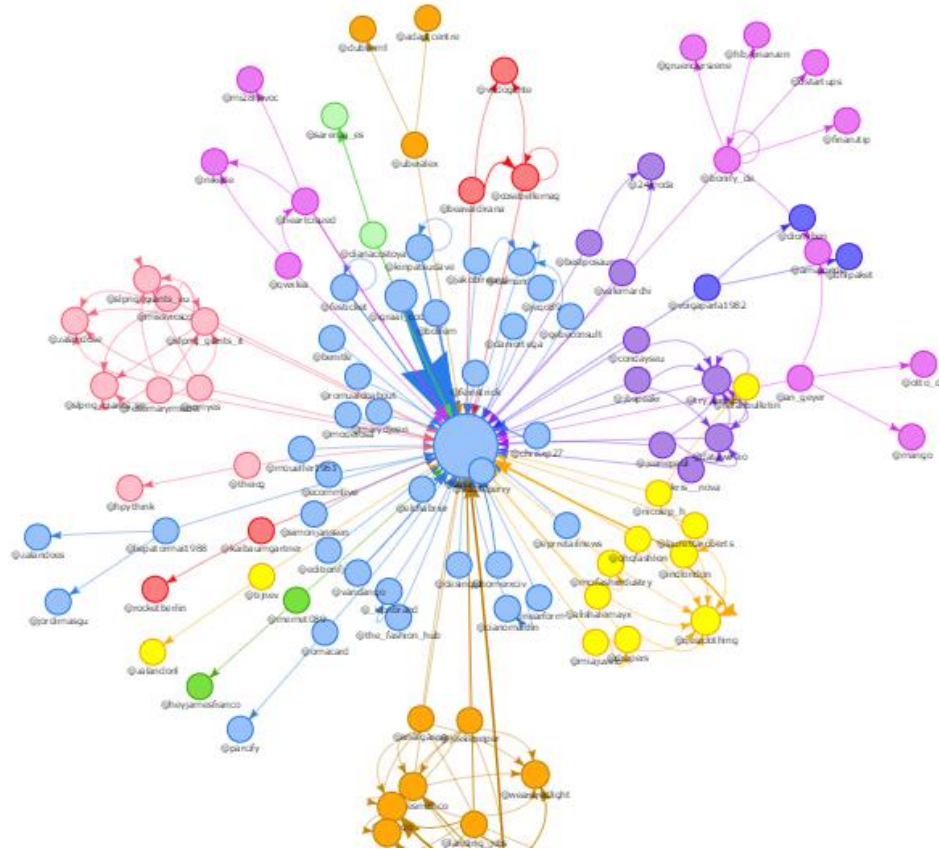
High level architecture for data analysis:

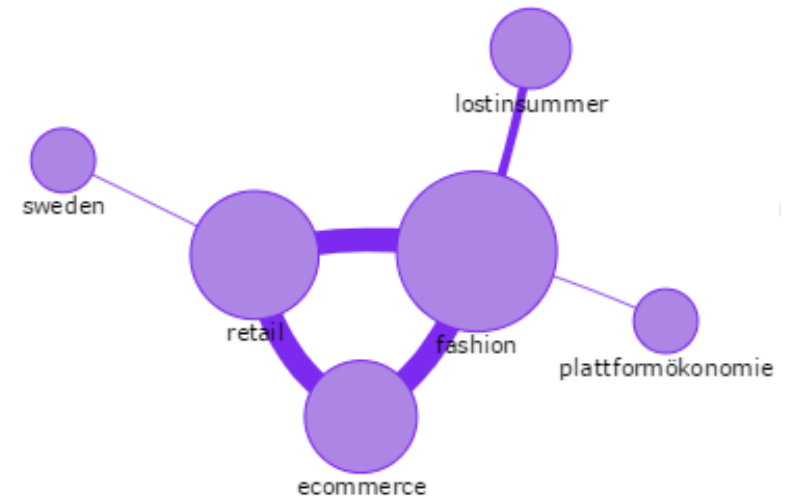| Sources	Data Feed: Java Processs	Big Data Ecosystem	R	Gephi, Excel, Tableau

Note:

In our analysis we collect all the data separately through each of their APIs (Web traffic through **Alexa**, Twitter through **Socioviz** and Facebook through **Netvizz**). In the production environment, we recommend to use more advanced system, such as **Hadoop**, in order to get integrated figure

Mentions Analysis:
Identify Key Influencers

# Hashtag Analysis:
Identify Customers' Opinions

Zalando's graphical representations above are obtained through socioviz feature. Based on that, we understand that there is no key influencers (customers is **fragmented**) and the customers perceived Zalando as a **retail**, **e-commerce** and **fashion** brand

3

NETVIZZ –
Facebook Text and Sentiment Analysis

In order to do text and sentiment analysis, we utilize **Netvizz** to collect all the unstructured data from Facebook comments, likes, shares, posts and fans
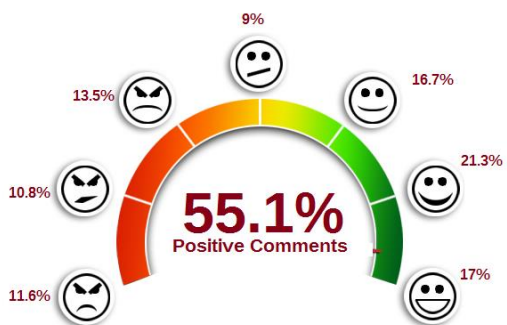
High level architecture for data analysis:
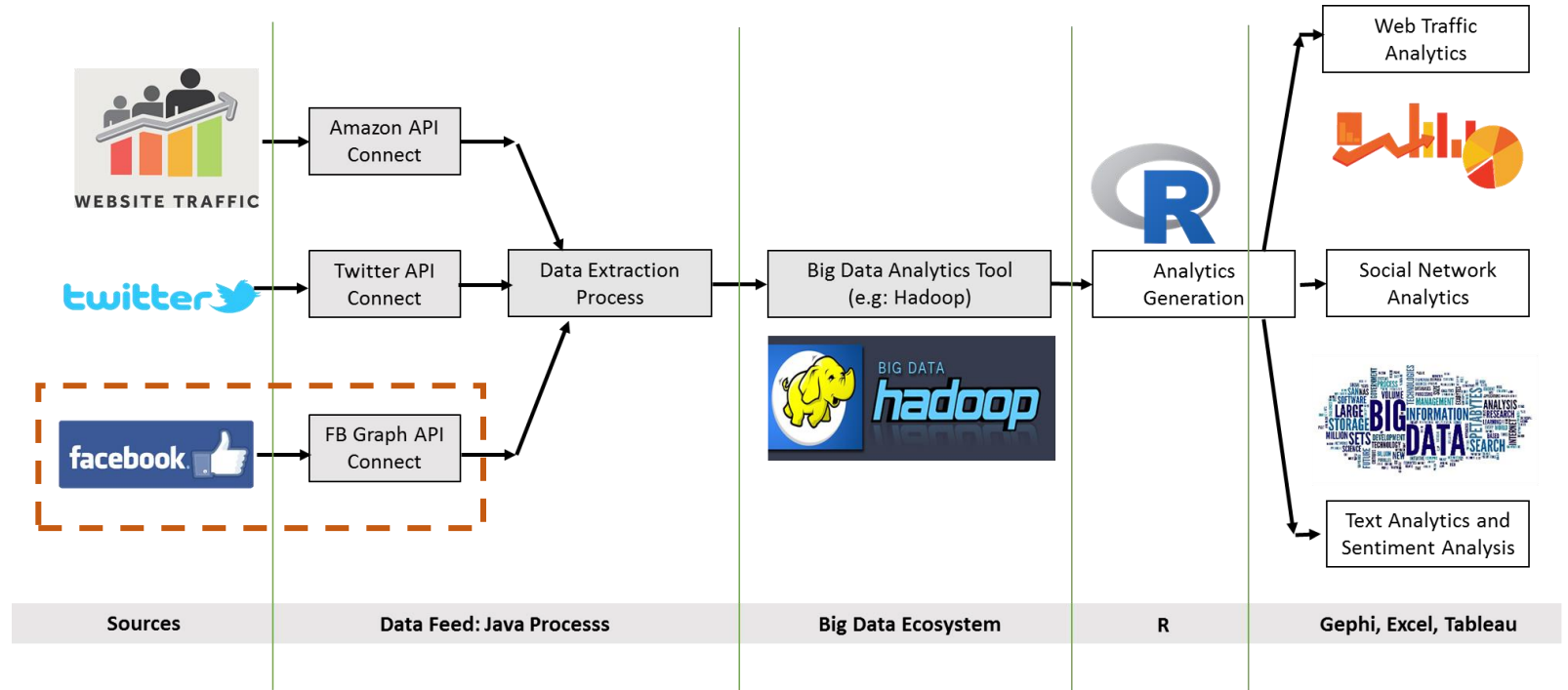
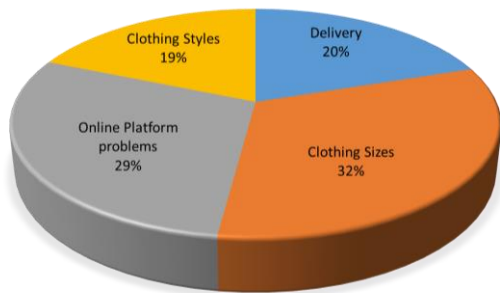| Sources | Data Feed: Java Processs | Big Data Ecosystem | R | Gephi, Excel, Tableau |

Note:

In our analysis we collect all the data separately through each of their APIs (Web traffic through **Alexa**, Twitter through **Socioviz** and Facebook through **Netvizz**). In the production environment, we recommend to use more advanced system, such as **Hadoop**, in order to get integrated figure

Sentiment Analysis

WordCloud – Customers

WordCloud – Zalando

55.1% Positive Comments

9% 16.7% 13.5% 21.3% 10.8% 17% 11.6%

The **44.9%** Negative Comments Distribution:
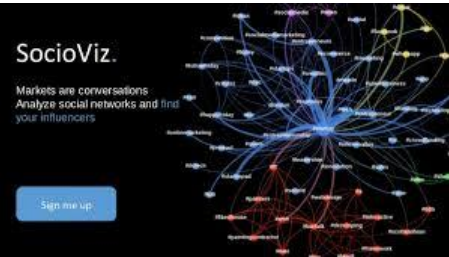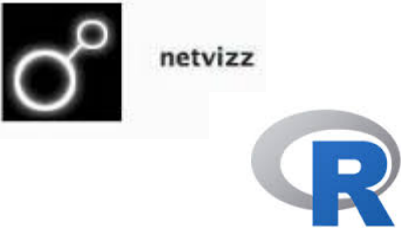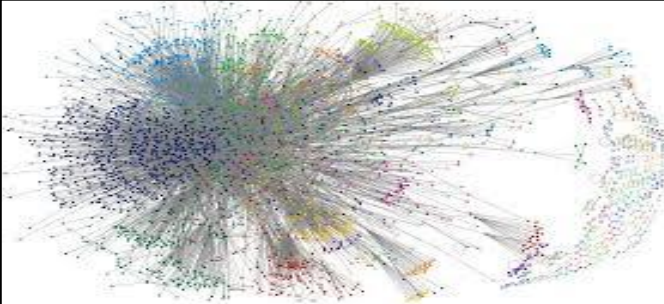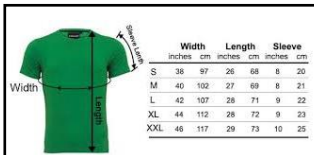
Delivery 20%
Clothing Styles 19%
Clothing Sizes 32%
Online Platform problems 29%

grazie sempre piace molto scarpe cosaÃ prezzo pagamento bene bello pagare sempre

We utilize R to analyse data obtained from Netvizz. Overall the **Sentiment Analysis** and the Natural Language Processing (NLP) through **WordCloud – Customers** analysis showed positive response; The **WordCloud – Zalando** also performed to help us checked whether our interaction with customers aligned with the way we want to be perceived

Eventually, all the insights will help the business in **shaping the strategy** as follow..

Analytics Tools	Analysis Performed	Insights	Managerial Implication
amazon alexa	• Web Traffic • Demographic Figure	• Well-balanced traffic source • Educated female, age 25-34 with income 0-30K; Search at home and work; Hispanic and Caucasian ethnic	• **Special promotion** for products related with educated female (worker) • Selection of Zalando **brand ambassador** (Hispanic/ Caucasian)
SocioViz. Markets are conversations Analyze social networks and find your influencers Sign me up	• Key Influencers • Customers' Opinion	• Fragmented customers – no key influencers • Zalando perceived as retail, e-commerce and fashion brand	• **Select key influencers** (e.g: brand ambassador) to increase awareness • **Periodic image review** on what Zalando wants to be perceived in the market by its customers
netvizz R	• Sentiment Analysis • WordCloud – Customers • WordCloud – Zalando	• Overall, customers' satisfied • Top 3 improvement opportunities: clothing sizes, platform problem and delivery	• Improve **clothing size details** • Prioritizing **improvement on the platform issues** • **Delivery performance review** for Zalando's vendor

Further details on the **promotions or actions** taken and **potential values creation**

Managerial Implication	Promotions/ Actions		Potential Values
• **Special promotion** for products related with educated female (worker) • Selection of Zalando **brand ambassador** (Hispanic/ Caucasian)	**Promotion for female worker**		• Increase **awareness** • Increase **sales** through promotion and better service on the clothes size, platform and delivery • Relevant **brand image creation**
• **Select key influencers** (e.g: brand ambassador) to increase awareness • **Periodic image review** on what Zalando wants to be perceived in the market by its customers	Brand ambassador that creates brand image: *YOUTH AND STRONG*	**Performing hashtag/ sentiment analysis to see whether the customers perceived us the way we want to be perceived**	
• Improve **clothes size details** • Prioritizing **improvement on the platform issues** • **Delivery performance review** for Zalando's vendor	**Share details on clothes size**	**Top 3 Platform Problems:** 1. Difficulties in changing password 2. Different experience between mobile apps and website 3. Paypal payment related issues	**Review delivery performance**

CONCLUSION

Input Data*	Source Data	Data Avail.	Variables and Characteristics	Data Processing Methods**	Managerial Implication	Potential Value
Web Traffic	Amazon API	Alexa	**Structured:** Bounce rate, Pageview per customer, Time on site, Traffic source, Demographic	Online processing	• Promotion for relevant product • Selection of Zalando's brand ambassador	• Increase awareness • Increase sales • Relevant brand image creation
Demographic						
Mentions	Twitter API	Socioviz	**Unstructured:** Key influencers and customers' opinion	Online processing, API and NLP	• Select key influencers • Periodic image review	
Hashtag						
Sentiment Analysis	Facebook API	Netvizz	**Unstructured:** Perception towards Zalando	Online processing, API and NLP	• Improve clothing size details • Platform issues resolution • Delivery performance review	
WordCloud – Customers						
WordCloud - Zalando						

* For more details on the input data, please refer to appendix

** In our analysis, we extract the data by leveraging free tools that connect directly to each social media APIs. For production environment, we recommend to use advanced big data ecosystem to get collective insights, such as Hadoop

Research/Academic Documents

- Paul M. Leonardi,Marleen Huysman,Charles Steinfield. *Enterprise Social Media: Definition, History, and Prospects for the Study of Social Technologies in Organizations*; http://onlinelibrary.wiley.com/doi/10.1111/jcc4.12029/full

- Kurniawati, Graeme Shanks. *The Business Impact Of Social Media Analytics*; http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1271&context=ecis2013_cr

- Gunther Eysenbach. *Can tweets predict citations?;* http://www.jmir.org/2011/4/e123/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+JMedInternetRes+(Journal+of+Medical+Internet+Research+(atom))

Blogs and other websites

- **How Starbucks And Other Companies Use Complex Math Algorithms To Read Your Feelings Online** *http://www.businessinsider.com/twitter-facebook-monitoring-2012-11?IR=T*

- **PlayStation VR Customer Sentiment Analysis** *https://www.strategyanalytics.com/strategy-analytics/news/strategy-analytics-press-releases/strategy-analytics-press-release/2017/03/20/strategy-analytics-playstation-vr-customer-sentiment-analysis-shows-device-delights-consumers#.WS14Z2jyjIU*

- **Why Retailers Need Sentiment Analysis To Connect With Customers** http://www.clarabridge.com/blog/the-heart-of-the-matter-why-retailers-need-sentiment-analysis-to-connect-with-customers/

- **Customer Sentiment Analysis: A crucial need in E-Commerce Data Initiatives** https://blog.nabler.com/customer-sentiment-analysis-a-crucial-need-in-e-commerce-data-initiatives/

- **Zara Leverages Data Analytics to Understand Consumer Tastes** https://digit.hbs.org/submission/zara-leverages-data-analytics-to-understand-consumer-tastes/

- **How Zalando used Bayesian stattistics to sabe loads of money** http://dataconomy.com/2015/04/how-zalando-used-bayesian-statistics-to-save-loads-of-money/

Netvizz: Netvizz is a tool that extracts data from different sections of the Facebook platform - in particular groups and pages - for research purposes. File outputs can be easily analyzed in standard software.

https://apps.facebook.com/netvizz/?fb_source=search&ref=br_rs
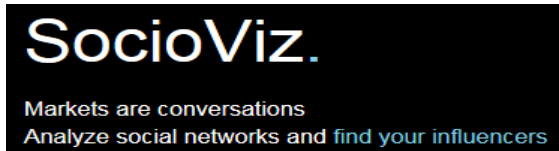
Socioviz: SocioViz is a social media analytics platform powered by Social Network Analysis metrics

http://socioviz.net/SNA/eu/sna/login.jsp

R Language: R is a free software environment for statistical computing and graphics.

https://cloud.r-project.org/

Alexa: is a California-based company that provides commercial web traffic data and analytics. It is a wholly owned subsidiary of Amazon.com.

http://www.alexa.com/

Appendix: Input Data Details

Input Data	Source	Description	Why is it important?	Type	Data Processing Methods	Automatically extracting	Legal Consideration
Web Traffic: Bounce Rate	AMAZON API	It represents the percentage of visitors who enter the site and then leave ("bounce") rather than continuing on to view other pages within the same site.	Bounce rates can be used to help determine the effectiveness or performance of an entry page at generating the interest of visitors.	Structured	Online processing	Yes	The data will be extracted from Internet Alexa API, a Premium Account is needed to start using their API. Alexa Internet, Inc. is a California-based company that provides commercial web traffic data. It is a wholly owned subsidiary of Amazon.com. (Not to be confused with Amazon Alexa Echo Voice)
Web Traffic: Daily Pageviews Per Visitor	AMAZON API	Once a visitor arrives at our website, they will search around on a few more pages. Each individual page a visitor views is tracked as a page view.	We need to understand if the pages are relevant or not, if the links and menu are ok for the user	Structured	Online processing	Yes	
Web Traffic: Daily Time onsite	AMAZON API	An average amount of time all visitors spend on a particular page.	This is a good metric to see if visitors are actually reading our content.	Structured	Online processing	Yes	
Web Traffic: Traffic Sources	AMAZON API	The percentage of all referrals that came from Search engines, social networks, from links on other or direct	It is important to understand the key sources of traffic to improve the actions, generate synergy, etc	Structured	Online processing	Yes	
Web Traffic: Audience Demographics	AMAZON API	A general profile including gender, education, age, income	We will be able to improve our consumer understanding and generate better products, campaigns and user experience	Structured	Online processing	Yes	

Input Data	Source	Description	Why is it important?	Type	Data Processing Methods	Automatically extracting	Legal Consideration
Twitter Mentions	TW API	A Tweet that includes your @handle at any point other than the beginning.	The serious advantage of a mention is that it has the potential to reach Twitter users who may not be following you or aware of your brand's presence on Twitter.	Unstructured	Online processing, API, NLP	Yes	Twitter allows a real time streaming API for R, Python, Java and other languages. We need to create a developers account and get the API and Secret keys.
Twitter Hastags	TW API	A hashtag is simply a way to categorize a tweet's topic(s), which then makes it easier for people to search for other tweets about those topics.	One of the best things about hashtags is that it allows you to create communities of people interested in the same topic by making it easier to find and share related information.	Unstructured	Online processing, API, NLP	Yes	

Appendix: Input Data Details

Input Data	Source	Description	Why is it important?	Type	Data Processing Methods	Automatically extracting	Legal Consideration
Fanpage Comments	FB API	A note, image, and/or reply that a user leaves beneath a Facebook post.	While the number of comments our brand receives is worth paying attention to, it's the tone and content of comments that provide insight into how people feel about our brand, product, and campaign.	Unstructured	Online processing, API, NLP	Yes	We will be able to gather data that a) the logged-in user has access to and b) the logged-in user has given the app access to; we can only retrieve the current user's data (that we have asked for via various permissions scopes), certain pieces of their friend's data and public data.
FB Likes	FB API	Represented by the thumbs up symbol, a Like implies that a Facebook post resonated with a reader and that he or she appreciates the content.	Receiving Likes as a brand is a great way to build brand awareness on Facebook.	Unstructured	Online processing, API	Yes	
FB Shares	FB API	When a user shares your Facebook post with his or her Facebook friends, possibly adding commentary.	A Facebook share provides more exposure for our brand and amplifies our content.	Unstructured	Online processing, API	Yes	
FB Posts	FB API	It is a comment, picture or other media that is posted on a Facebook page or "wall."	Facebook posts are about invoking engagement and action. It is the way to communicate ideas, new products, etc.	Unstructured	Online processing, API, NLP	Yes	
FB Fans	FB API	A Facebook "fan" is a user who "likes" a particular page.	the company can use objective numbers like Facebook followers to measure a campaign's health or success.	Unstructured	Online processing, API	Yes	