

Text Mining in Social Media for Market Analysis of Products.

Evandro Botti de Cerqueira
Federal University of Bahia - UFBA

Abstract: The fast progress of digital technologies has transformed the world in many ways over the last decades, in particular with new media such as social media. The main objective of this work was to understand the main characteristics, motivations, habits, interests and values of a certain target audience through their text messages published on twitter, in order to propose a psychological and demographic model for use in marketing campaigns.

Related publications:

- FEINERER, I, HORNIK, K, MEYER, D (2008). Text Mining Infrastructure in R. Journal of Statistical Software.
- KING, M. (Ed.). Music Marketing: Press, Promotion, Distribution, and Retail. Boston: Berklee Press, 2009



evandrobotti@gmail.com



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Problem: Understanding the characteristics, preferences, habits, interests and values of a target audience through their text messages.

Dataset: 25.054 tweets written from 07/09/2007 to 09/22/2022, using the *hashtag* : umbriajazz

A	B	C	
	Date	User	Tweet
0	2022-09-22 14:47:00+00:00	UmbriaJazz_OFC	@mariobiondi https://t.co/XL6P7pyPTM
1	2022-09-20 10:59:04+00:00	mbriotti	https://t.co/sRmZ6NY3aV
2	2022-09-19 23:27:59+00:00	GaffeBrio	@UmbriaJazz_OFC Come vedete, a Terni si fa sempre il Sold out
25051	2008-07-09 13:44:50+00:00	morello	@ninna_r che meraviglia - grazie. Speriamo che regga cos'! Qui a Perugia d'altraparte inizia UmbriaJazz. E noi smettiamo di dormire.
25052	2008-06-13 13:55:14+00:00	Googlisti	grazie a @abidun trovano il programma di UmbriaJazz 2008 e un pensiero nostalgico corre agli anni di università ...che bei ricordi!!
25053	2007-07-12 16:24:50+00:00	clarissagigante	si gode umbriajazz 2007
25054	2007-07-09 19:24:14+00:00	rainbowbreeze	termina un post per onemobile sull'umbriajazz e le tecnologie mobili. e' soddisfatto delle sue elucubrazioni, aspettando la pubblicazione..

Previous Works:

- FEINERER, I, HORNIK, K, MEYER, D (2008). Text Mining Infrastructure in R. Journal of Statistical Software.
- KING, M. (Ed.). Music Marketing: Press, Promotion, Distribution, and Retail. Boston: Berklee Press, 2009



evandrobotti@gmail.com



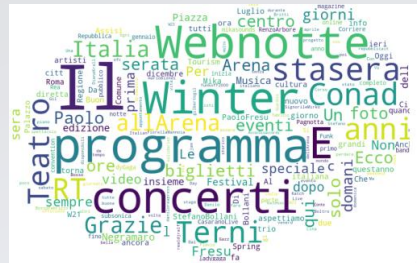
4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Methods: Analysis of users, analysis of publication time, analysis of the most frequent terms, analysis of n-grams, sentiment analysis, topic modeling analysis.

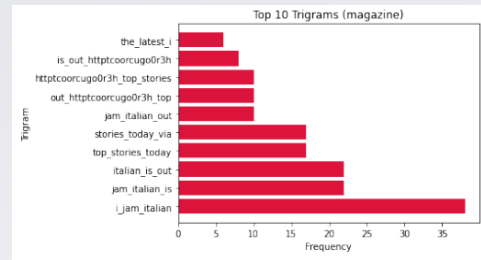
Tools: Python, Snsrape , Pandas, Langdetect , Matplotlib, Nltk, Numpy, Pycountry .



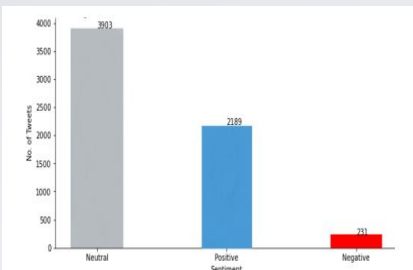
1- Word cloud of the Most Frequent Terms in English



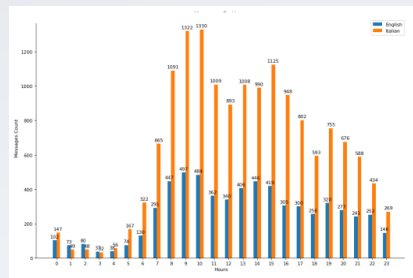
2- Wordcloud of the Most Frequent Terms in Italian



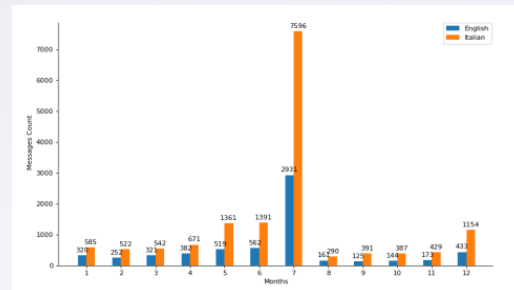
3 - Histogram of the Top 10 Trigrams



4- Sentiment Analysis Bar Plot



5 - Tweets per hour histogram.



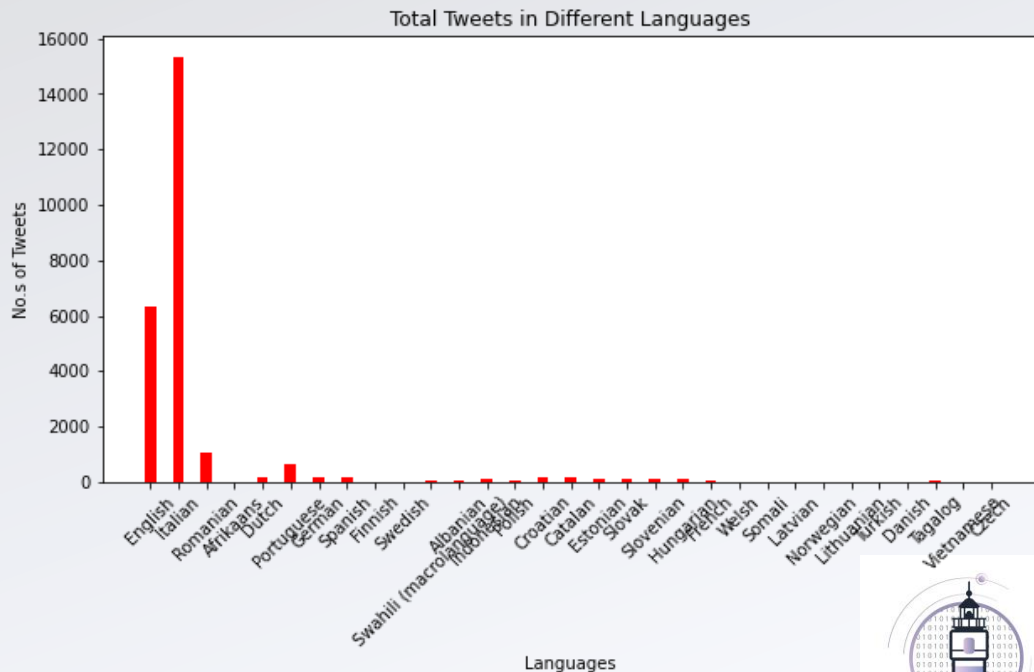
6 - Tweets per Month histogram.



Conclusions:

Main characteristics of a Perfect Client

Questions	Answers
Main countries ?	Italia, USA, Romania, Holanda
Main Languages ?	Italian, English, Romanian, Dutch
Main Blogs ?	Webnotte, CasaranoLive
Main Professions ?	Journalist, Musician, Digital Influencer
Main Radios ?	Montecarlo, Rai3, Jazzbittes
Main Magazine ?	I JAM
Main Journal ?	La Repubblica
Main Artists ?	Lady Gaga, Paolo Fresu, R.Glasper
Main Hobbies ?	Fotografia, Tourism
Main Festivals ?	Montreux Jazz e Umbria Jazz Winter
Main Feelings on Umbria Jazz?	Positives



evandrobotti@gmail.com