SANREMO21 ANALYSIS

Presentation @UniPI

25/03/2021

REPLY
TARGET

MEETING AGENDA

1 Team presentation

2 Introduction

3 Our Analysis

- Source
- Crawler method
- Storage
- Data Prep
- Analysis
- Demo

4 Q&A

TEAM PRESENTATION

MUSIC ANALYTICS GURU ☺

Tommaso Furlan
Senior Consultant
Data Science & Business Informatics @**UniPI**

Giulia Maggi
Consultant
Statistics & Data Science @**Milano-Bicocca**

Giovanni Valentini
Consultant
Statistics @**Milano-Bicocca**

Sara Pisaniello
Consultant
Economics Science & Statistics @**Milano-Bicocca**

Vincenzo Cimino
Consultant
Informatics @**UniPA**

…and many other gurus!

SANREMO
71°
FESTIVAL
DELLA
CANZONE
ITALIANA
2021

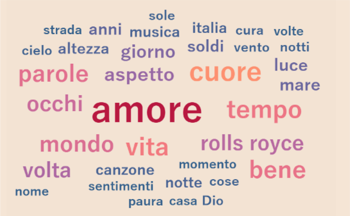
INTRO

OUR ANALYSIS

INTRODUCTION

Social Monitoring

Tweets Sentiment Analysis



NLP Text Analysis

Winner Prediction Analysis

Tableau Public for Sanremo Analysis by Target Reply: https://tabsoft.co/2OerpBm

SOCIAL MONITORING

SOCIAL MONITORING

Goal

Analyze the evolution of the instagram followers before - during - after Sanremo2021

Dataset

- Number of followers for each singer
- Post-like number for each singer

Ingestion:

- Scrapping of IG Profile directly from HTML pages (no API)
- Frequency: 2 times a day

Data Prep / Pre-Processing:

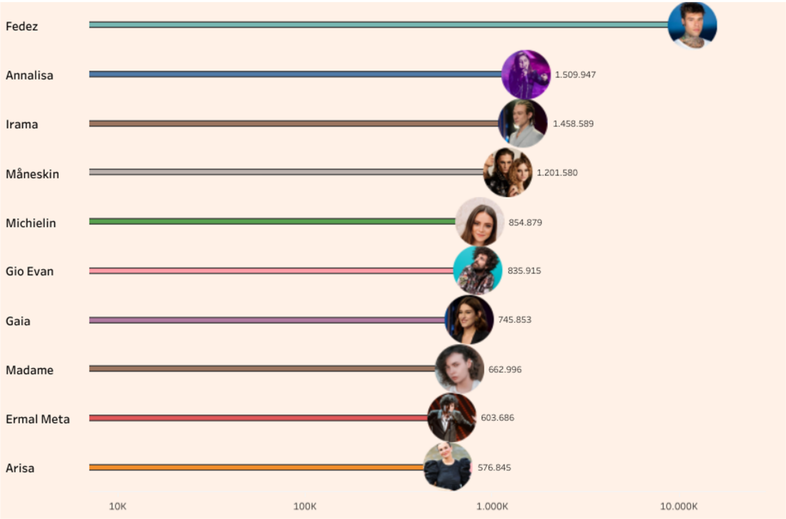
- Aggregation

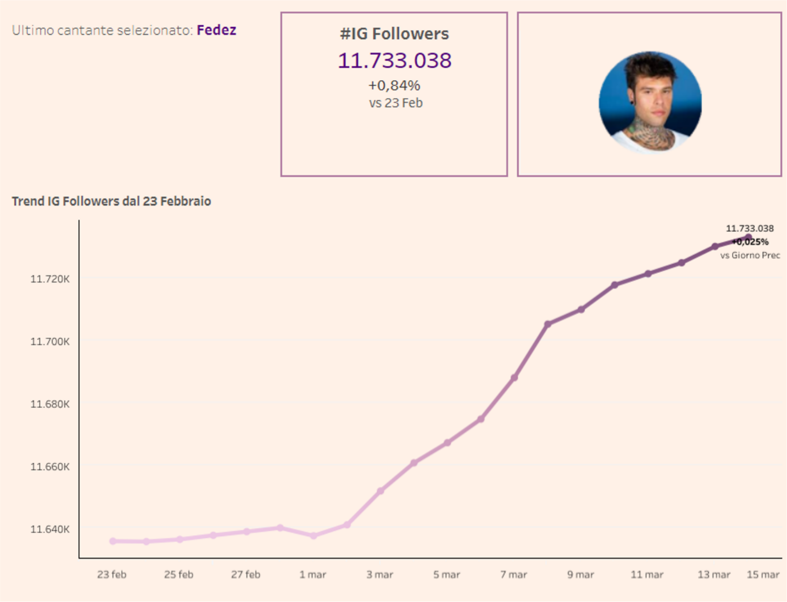
Analysis:

- IG followers evolution
- Social Activity

SOCIAL MONITORING

IG Followers





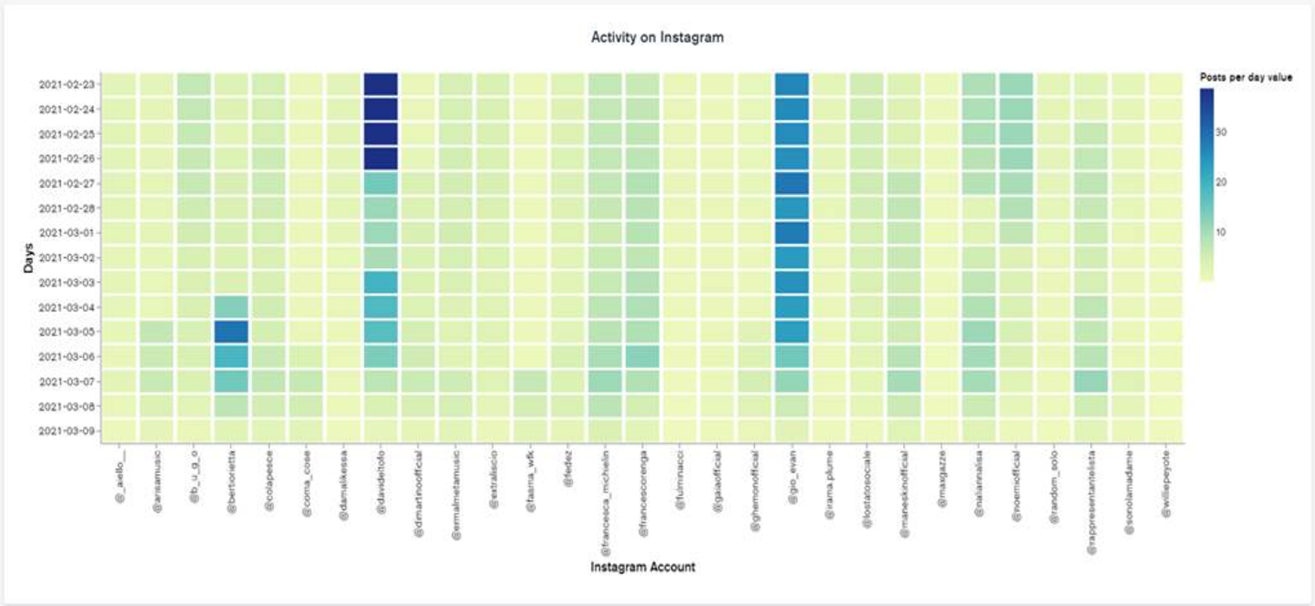
Ultimo cantante selezionato: **Fedez**

#IG Followers
11.733.038
+0,84%
vs 23 Feb

Trend IG Followers dal 23 Febbraio

SOCIAL MONITORING

Social Activity



NLP TEXT ANALYSIS

NLP TEXT ANALYSIS

Goal

Analyze the lyrics of the songs in the competition from multiple points of view and to try to identify the candidate winner.

Dataset

- Sanremo21 song lyrics
- 2010-2020 winning song lyrics
- Entire discography of Sanremo21 singers

Ingestion:

- API (genius.com)

Data Prep / Pre-Processing (R):

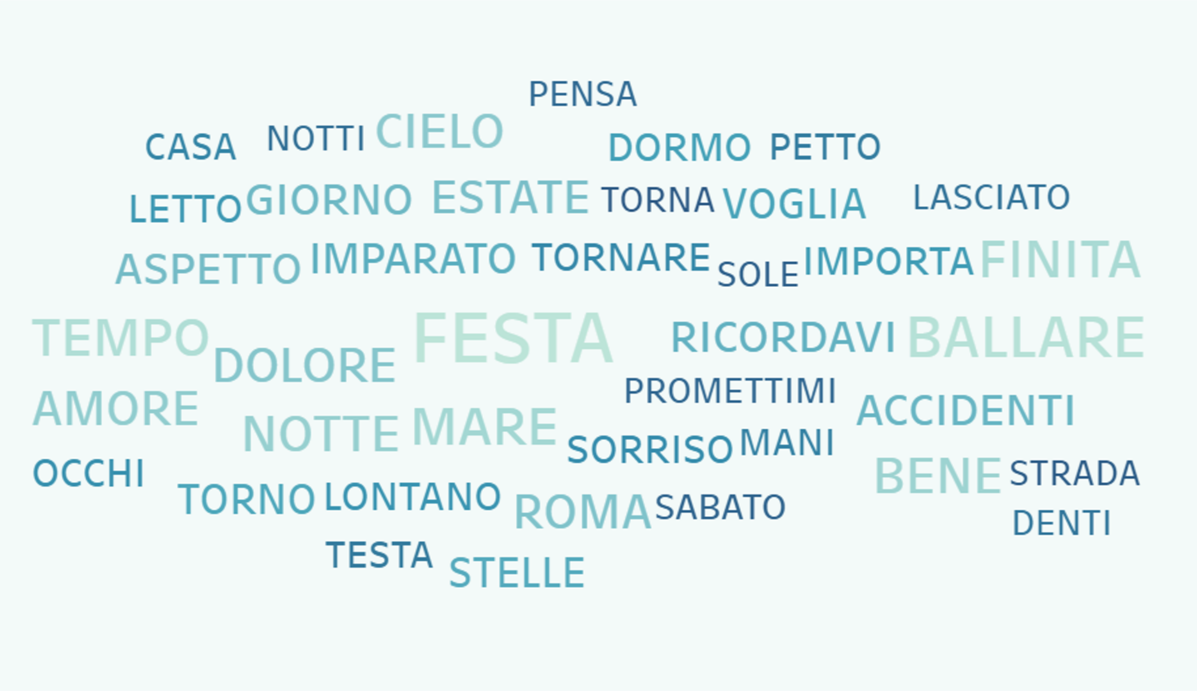
- Python library: «lyricgenius»
- Tokenization
- Remove stop-words
- Stemming (next step)

Analysis:

- Wordcloud
- Frequency word-singer
- Correlations
- Sentiment Analysis

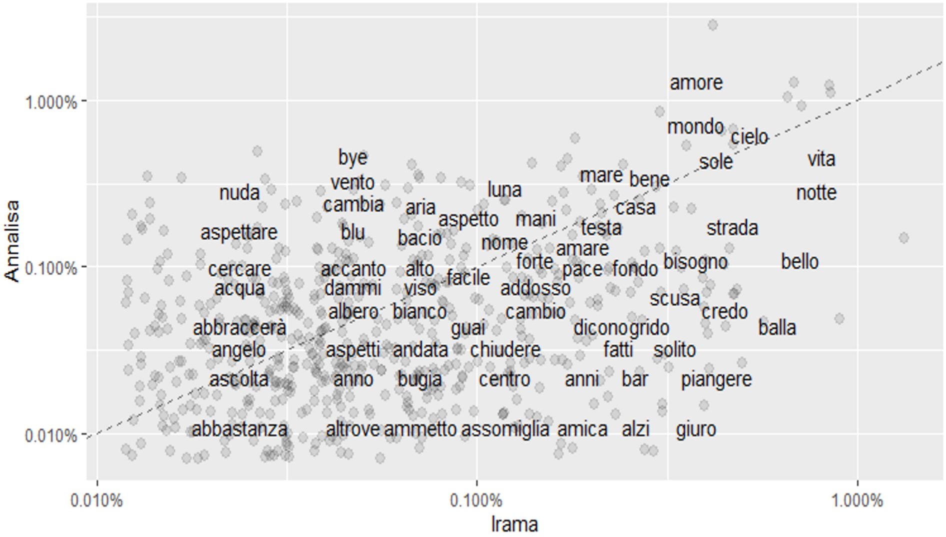
NLP TEXT ANALYSIS

Wordcloud



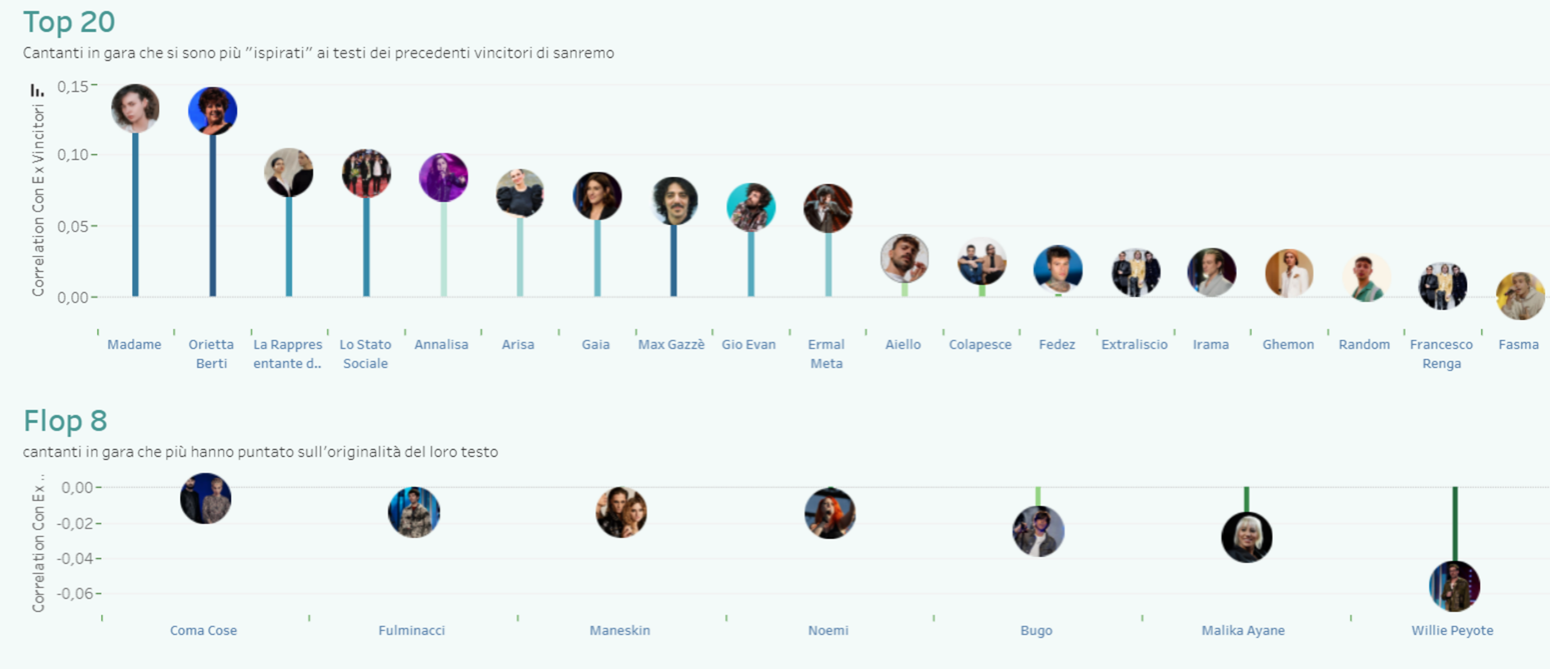
NLP TEXT ANALYSIS

Frequency word-singer



NLP TEXT ANALYSIS

Correlation (previous winners)



Top 20

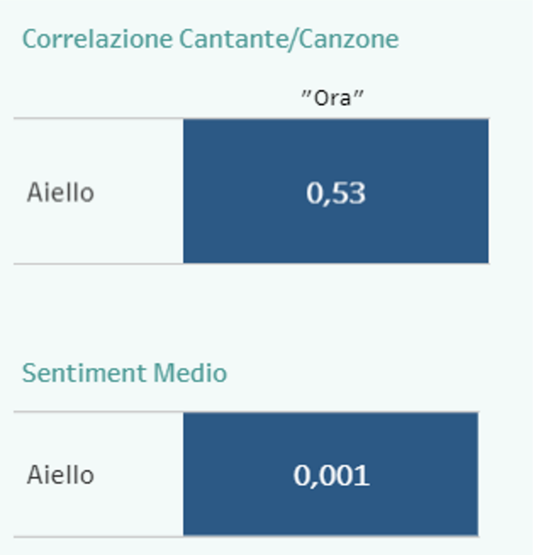
Cantanti in gara che si sono più "ispirati" ai testi dei precedenti vincitori di sanremo

Flop 8

cantanti in gara che più hanno puntato sull'originalità del loro testo

NLP TEXT ANALYSIS

Correlation (singer discography)



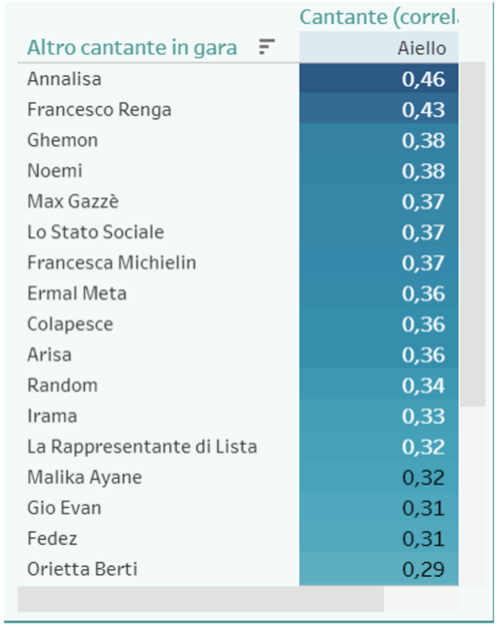
Correlazione Cantante/Canzone

	"Ora"
Aiello	0,53

Sentiment Medio

Aiello	0,001

Correlation (other partecipants)



	Cantante (correl
Altro cantante in gara	Aiello
Annalisa	0,46
Francesco Renga	0,43
Ghemon	0,38
Noemi	0,38
Max Gazzè	0,37
Lo Stato Sociale	0,37
Francesca Michielin	0,37
Ermal Meta	0,36
Colapesce	0,36
Arisa	0,36
Random	0,34
Irama	0,33
La Rappresentante di Lista	0,32
Malika Ayane	0,32
Gio Evan	0,31
Fedez	0,31
Orietta Berti	0,29

TWEETS SENTIMENT ANALYSIS

TWEETS SENTIMENT ANALYSIS

Goal

Analyzing the language of tweets over the 5 evenings of Sanremo 2021.
Hashtag: #Sanremo2021 - #Sanremo21 - #Sanremo

Ingestion

- Stream real-time on official Twitter API

Analysis (for every day):

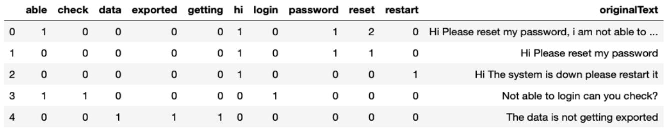
- Wordcloud
- Sentiment Analysis
- Graduation Index

Data Prep / Pre-Processing (R):

- Add REFERENCE_DAY column
- Remove stop-words/numbers/url/etc.
- Remove 'Sanremo-words' (es: festival)

Models:

- DocumentTermMatrix

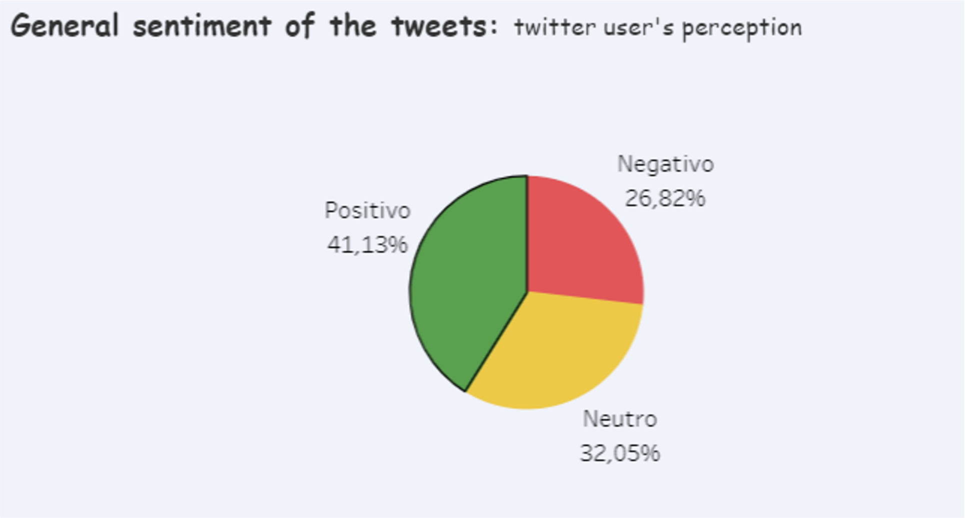


	able	check	data	exported	getting	hi	login	password	reset	restart	originalText
0	1	0	0	0	0	1	0	1	2	0	Hi Please reset my password, i am not able to ...
1	0	0	0	0	0	1	0	1	1	0	Hi Please reset my password
2	0	0	0	0	0	1	0	0	0	1	Hi The system is down please restart it
3	1	1	0	0	0	0	1	0	0	0	Not able to login can you check?
4	0	0	1	1	1	0	0	0	0	0	The data is not getting exported

- Wordcloud package for R (wordcloud analysis)
- Textwiller package for R (sentiment analysis)

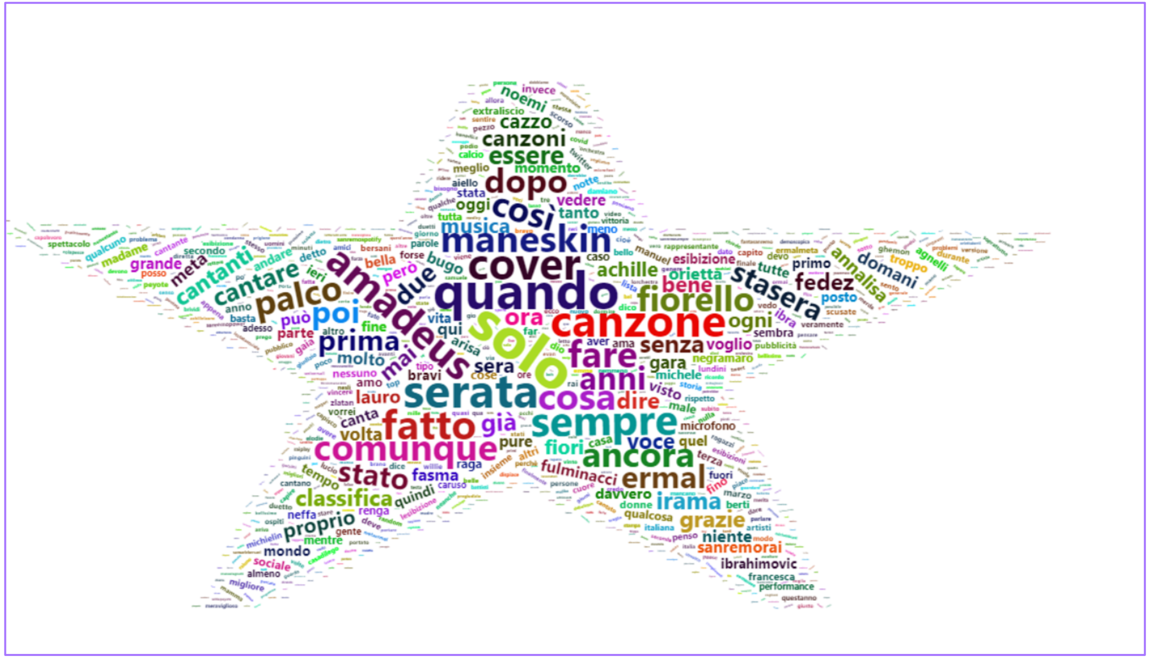
TWEETS SENTIMENT ANALYSIS

Sentiment Analysis



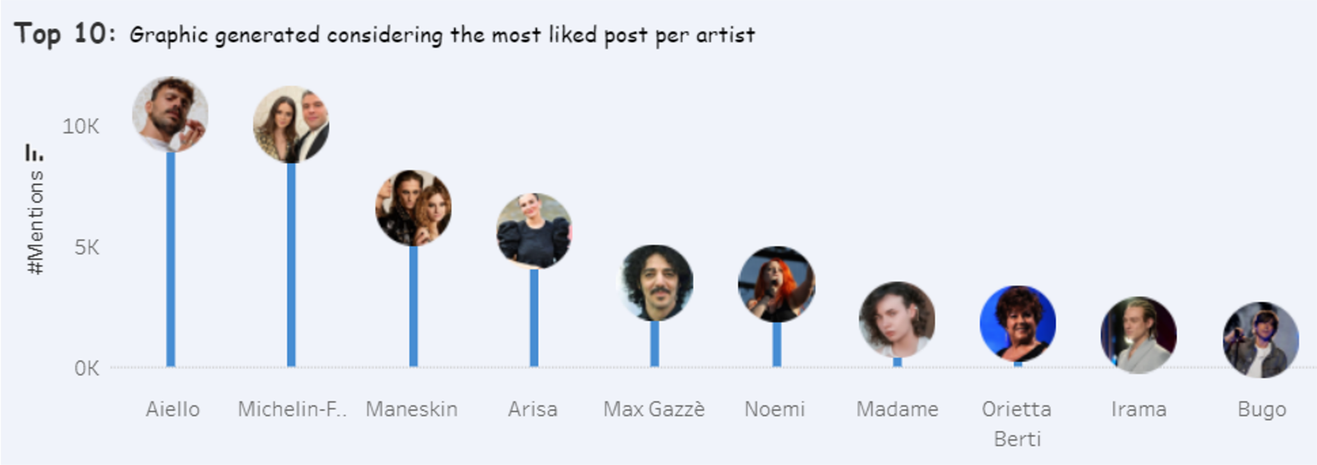
TWEETS SENTIMENT ANALYSIS

Wordcloud Analysis



TWEETS SENTIMENT ANALYSIS

Graduation Index



WINNER PREDICTION ANALYSIS

WINNER PREDICTION ANALYSIS

Goal

Predict the winner based on the characteristics of the songs

Dataset

- All songs of the 2010-2020 editions
- API Spotify *(https://developer.spotify.com/documentation/web-api/)*

Song features:

- Danceability
- Power
- Rhythm
- Etc.

Data Prep:

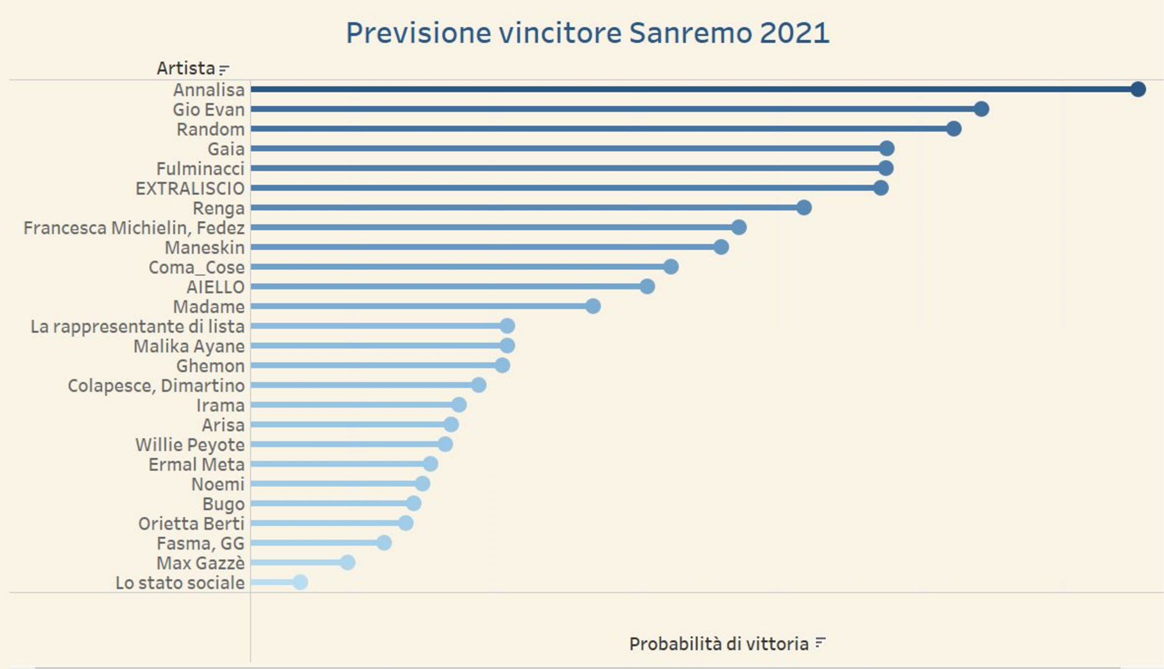
- Add WINNER column
- Removed columns with outliers
- No balancing techniques (test/training set)

Algorithms:

- Logistic Regression
- XGBOOST Model

WINNER PREDICTION ANALYSIS

Winner Analysis



Q&A

Q&A

We remain available to resolve any further doubts!

Our contacts:

- t.furlan@reply.it
- g.valentini@reply.it
- v.cimino@reply.it
- g.maggi@reply.it
- s.pisaniello@reply.it

‘PERCHÈ SANREMO È SANREMO…’

THANK YOU

www.reply.com

REPLY
TARGET