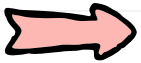
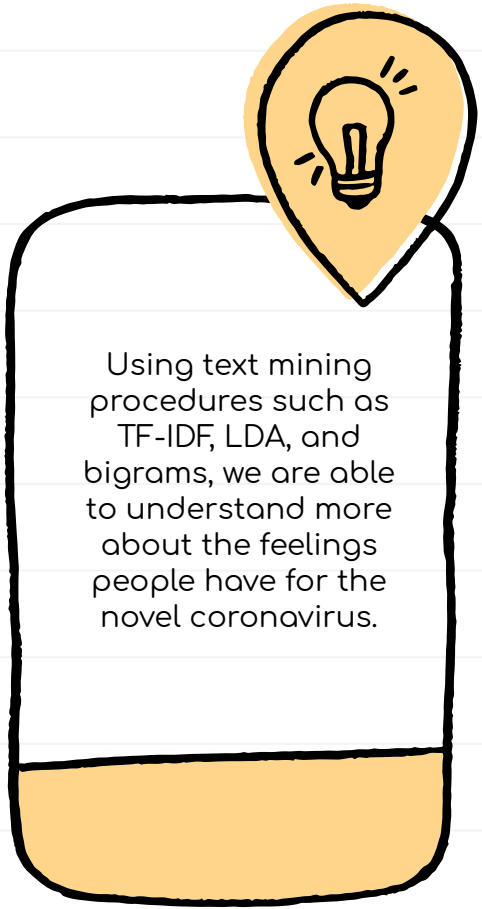
# Coronavirus Text analysis

Cindy | Isabel | Gustavo

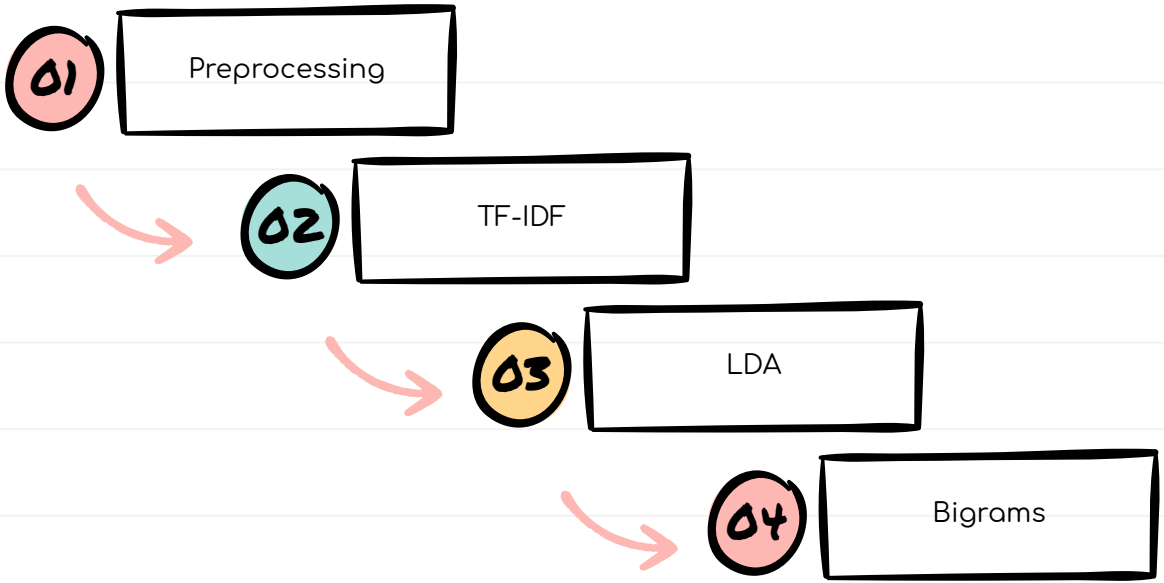
# OVERVIEW


Lots of “dirty” COVID-19 data exists, but without proper procedures, robust and accurate text analysis cannot exist.

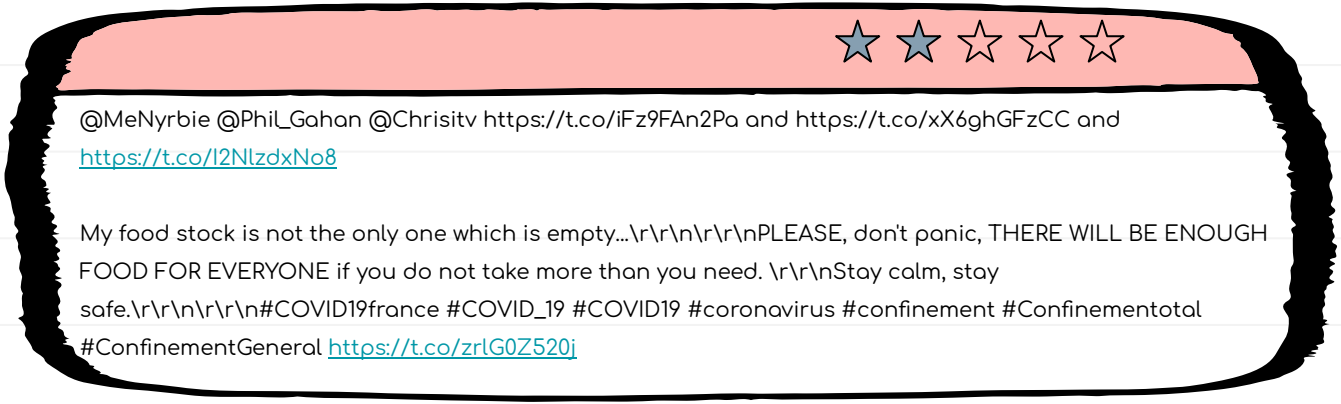



Using text mining procedures such as TF-IDF, LDA, and bigrams, we are able to understand more about the feelings people have for the novel coronavirus.

# Workflow Overview

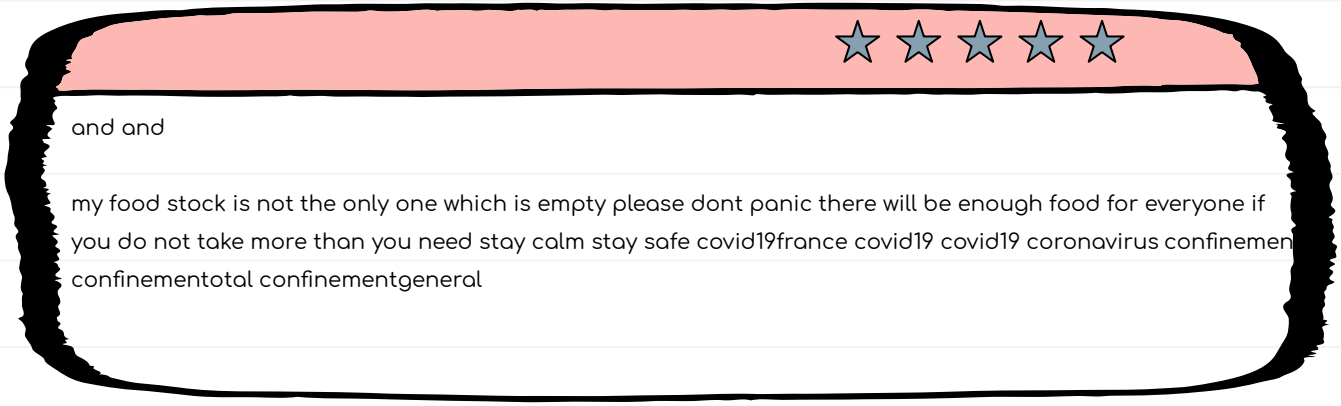


# Preprocessing



@MeNyrbie @Phil_Gahan @Chrisitv https://t.co/iFz9FAn2Pa and https://t.co/xX6ghGFzCC and https://t.co/I2NlzdxNo8

My food stock is not the only one which is empty...\r\r\n\r\r\nPLEASE, don't panic, THERE WILL BE ENOUGH FOOD FOR EVERYONE if you do not take more than you need. \r\r\nStay calm, stay safe.\r\r\n\r\r\n#COVID19france #COVID_19 #COVID19 #coronavirus #confinement #Confinementotal #ConfinementGeneral https://t.co/zrlG0Z520j



and and

my food stock is not the only one which is empty please dont panic there will be enough food for everyone if you do not take more than you need stay calm stay safe covid19france covid19 covid19 coronavirus confinemen confinementotal confinementgeneral

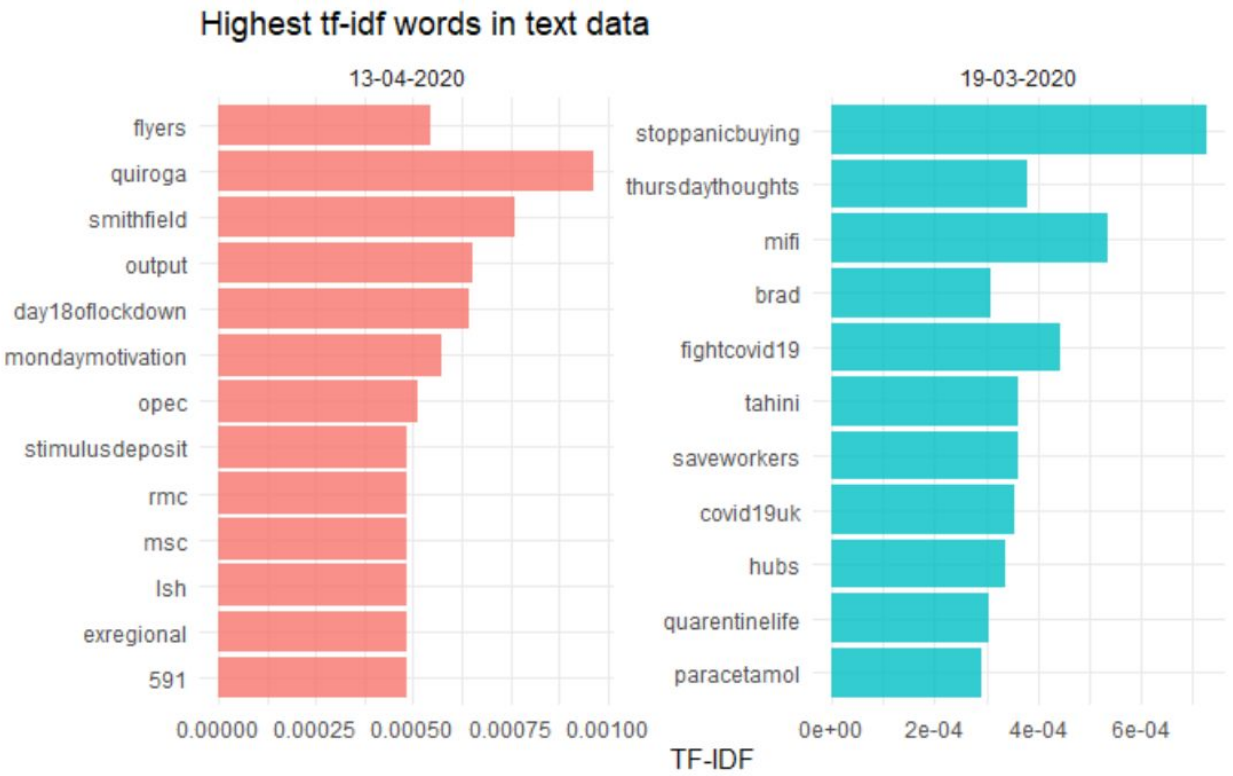
PUNCTUATION

ESCAPE CHARACTERS

LINKS

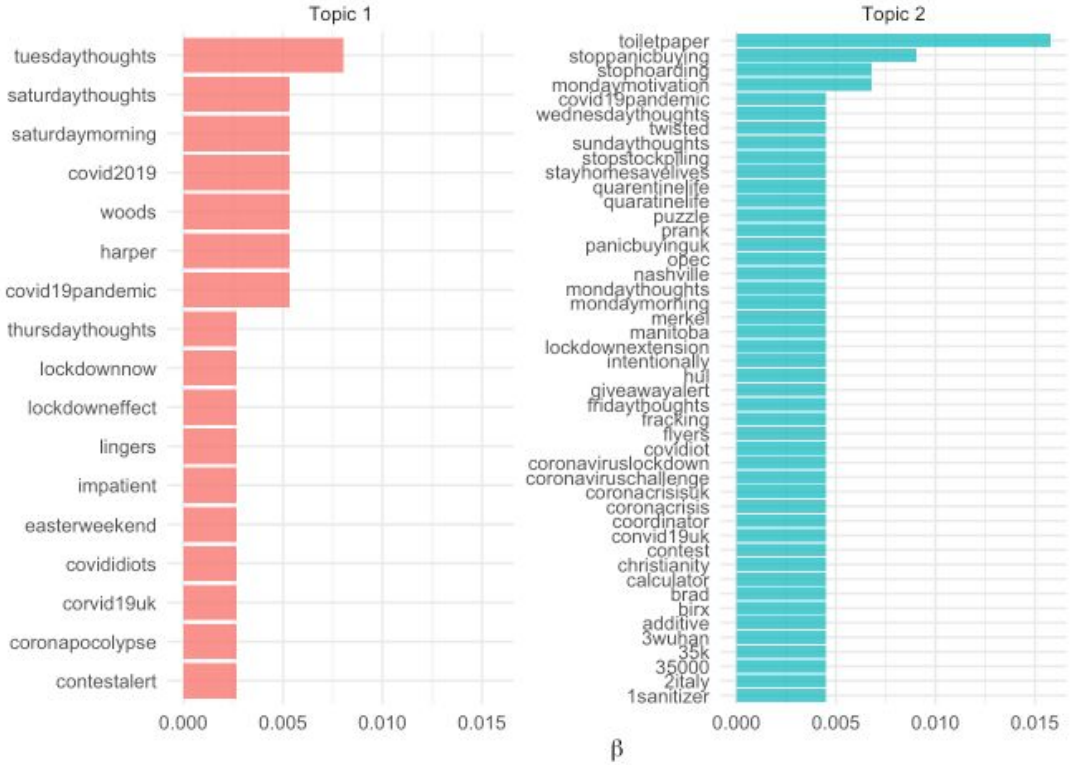
CAPITALIZATION

# Findings: TF-IDF




Hashtag words that trended on March 19th and April 13th, 2020

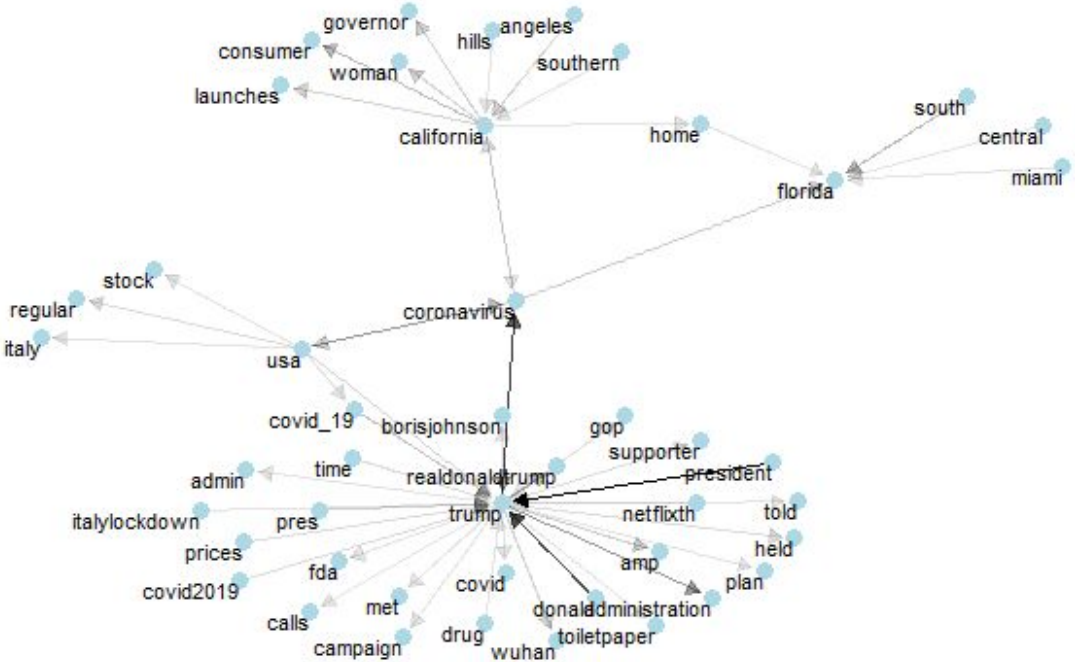
# Findings: LDA




Utilizing Latent Dirichlet Allocation to discover topics within the dataset

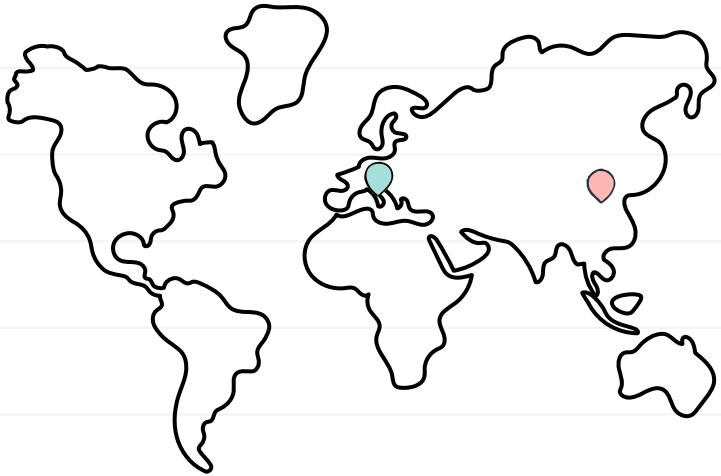
# Findings: Bigrams

- California
- Florida



# Findings: Bigrams

- Italy
- China



# Thanks!

Cindy | Isabel | Gustavo