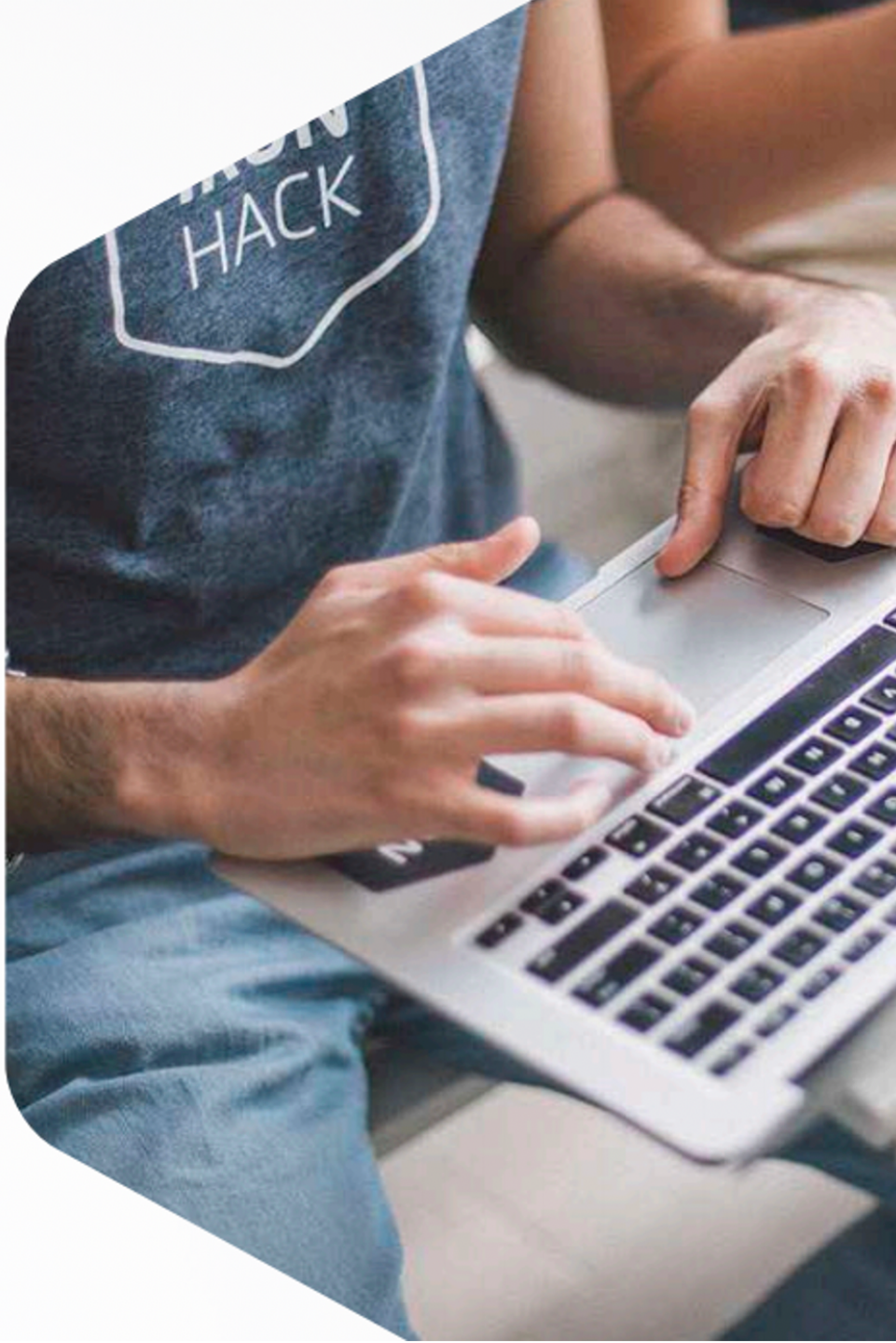
FINAL PROJECT – PT Data Analytics
Madrid 0419

April 2020



Summary

Objective of the project

- Find the best title for a Podcast to increase views and likes in Youtube

Skills, tools and technologies used

- Python & Jupyter Notebook.
- Statistical analysis.
- Data visualization -- via Plotly.
- Google: API & Google Sheets.
- Deep Neural Network.
- Word Embedding.
- JSON and csv output files.
- Web scraping -- Youtube.
- Stackoverflow.

The first step is scraping from Youtube the main KPIs* for each podcast of both the child and the parent channels

Scraping_info_from_youtube.ipynb && Scraping_info_from_youtube-others.ipynb

Python library

BeautifulSoup

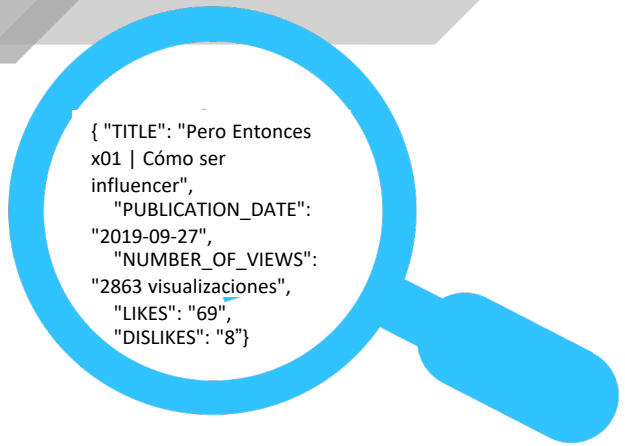
jupyter

Access to the Youtube channel and get the URL of each video

Parse the content of each video to HTML

Extract all the important KPIs

Write a JSON file with the KPIs of each video



Youtube channel

CHILD CHANNEL

PERO ENTONCES

PARENT CHANNEL

ΦΒΛ PODCAST

The objective is to analyse "Pero Entonces" however the volume of data is small, so the parent channel helps us to have a more robust model**

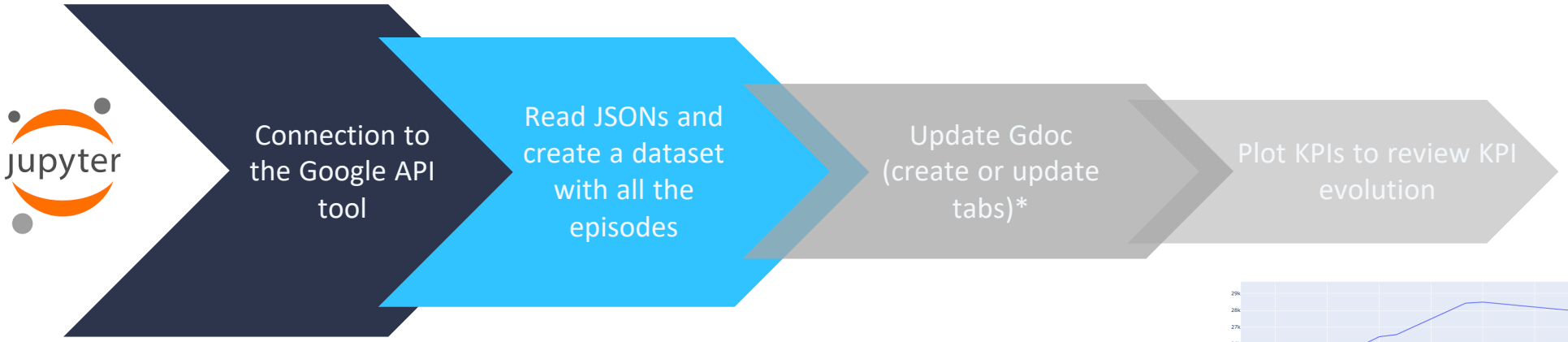
*KPIs: likes, dislikes, date of publication, views; **As we will see In the last slide, the volume of data makes the model more robust, but it still needs more words to make a more accurate assessment

Then, after establishing the connection with the Google API, a summary is written in a GSheet so that the author can monitor each episode evolution

Gsheet_update.ipynb

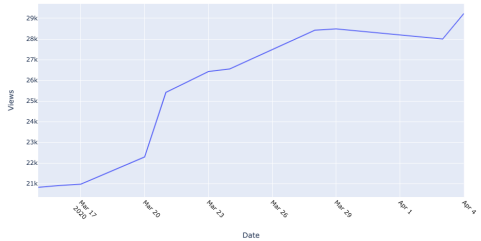
Python library
pygsheets

Python library
plotly

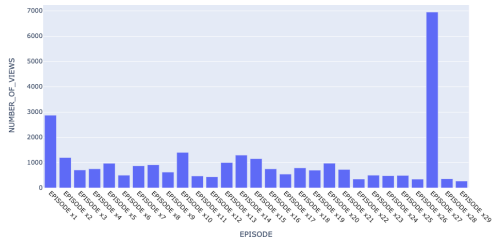


Google APIs

Daily views evolution



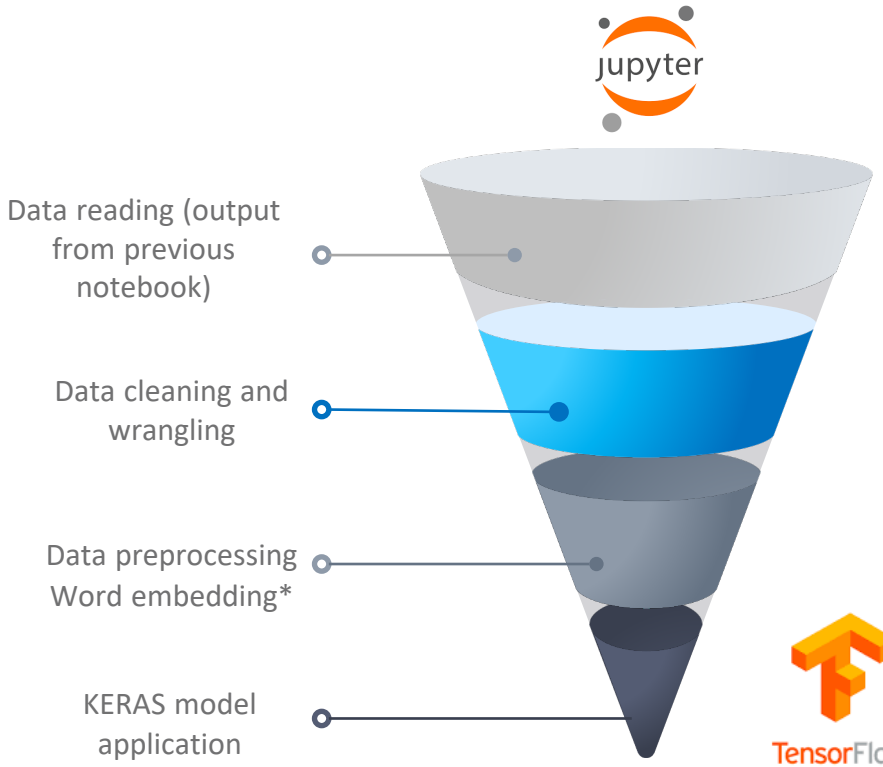
Last day situation

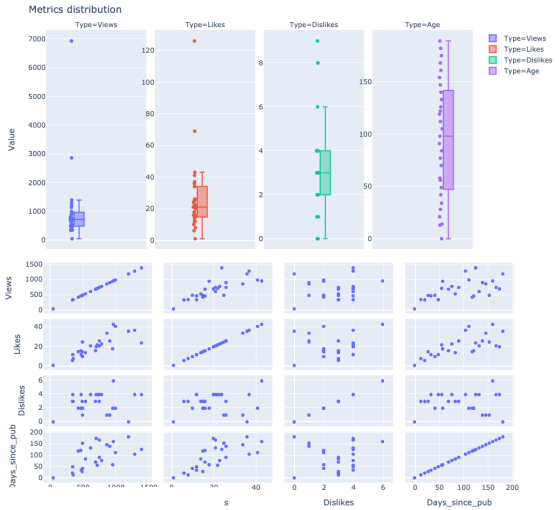


*This modification of the Gsheets is automatically done from the Jupyter Notebook; Also that Gsheets also serves as a database of all the data that is read and written everyday

Finally, a Deep learning model helps us to analyse each title to undertand which words drive the autor to have more views or likes

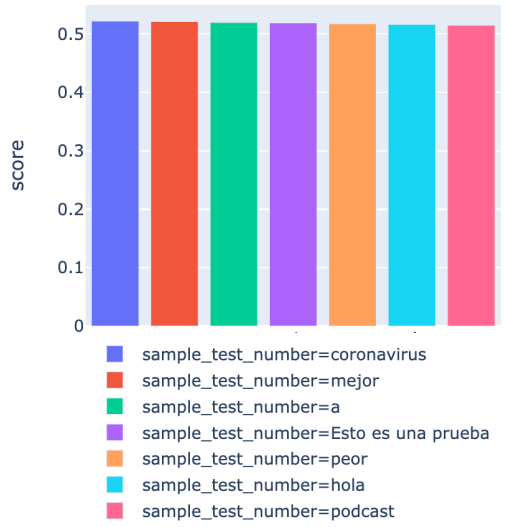
Likes_Views_Estimator.ipynb





Impact of the new title

Score by test sample



*The word embedding is included in the KERAS libraries and transform the text into numerical vectors to feed the model

Next steps….

The dictionary of words is still small we would need a much greater different range of words

- Construct a more robust model with a much wider dictionary of words
- Estimate the likes/views over time

Thank you for your time

April 2020