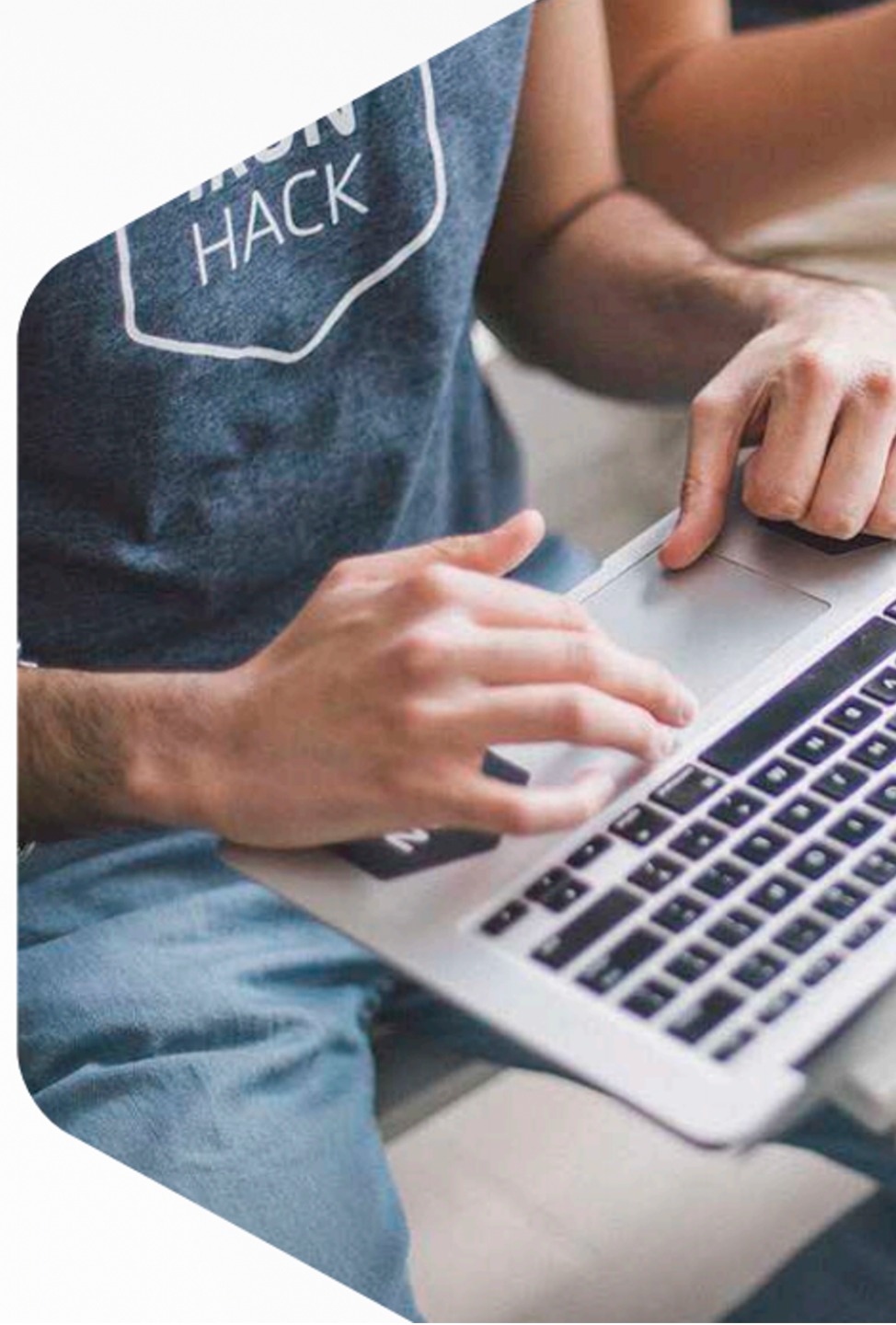# FINAL PROJECT – PT Data Analytics

Madrid 0419

April 2020
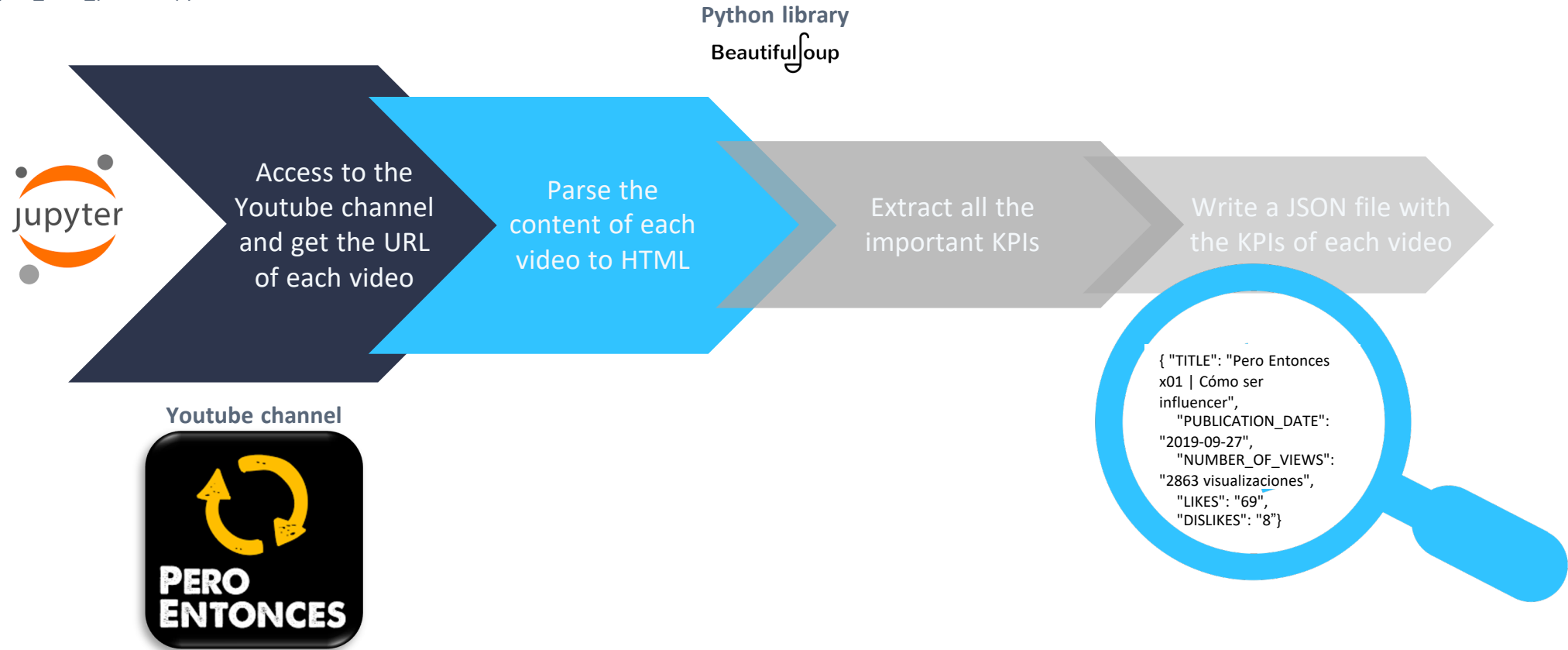
Summary

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.
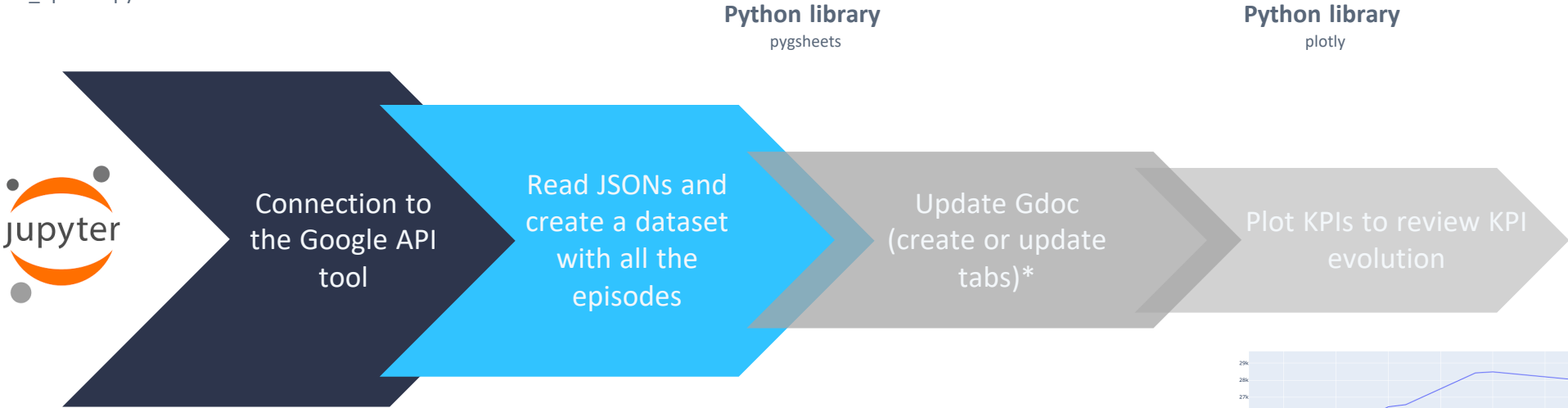- Web scraping -- Youtube.
- Stackoverflow.

The first step is scraping from Youtube the main KPIs for each podcast, such as likes, unlikes, views or days since publication

Scraping_info_from_youtube.ipynb

Python library

Beautiful Soup

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

PERO ENTONCES

{ "TITLE": "Pero Entonces x01 | Cómo ser influencer",
   "PUBLICATION_DATE": "2019-09-27",
   "NUMBER_OF_VIEWS": "2863 visualizaciones",
   "LIKES": "69",
   "DISLIKES": "8"}

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
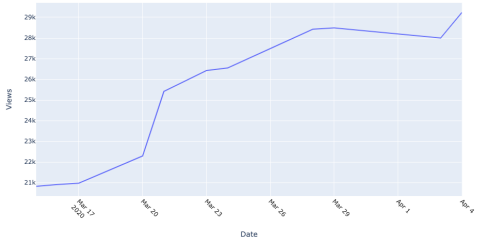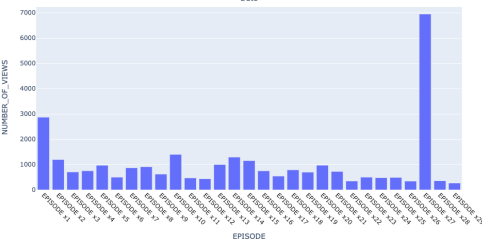
Gsheet_update.ipynb

Python library
pygsheets

Python library
plotly

Connection to the Google API tool

Read JSONs and create a dataset with all the episodes

Update Gdoc (create or update tabs)*

Plot KPIs to review KPI evolution
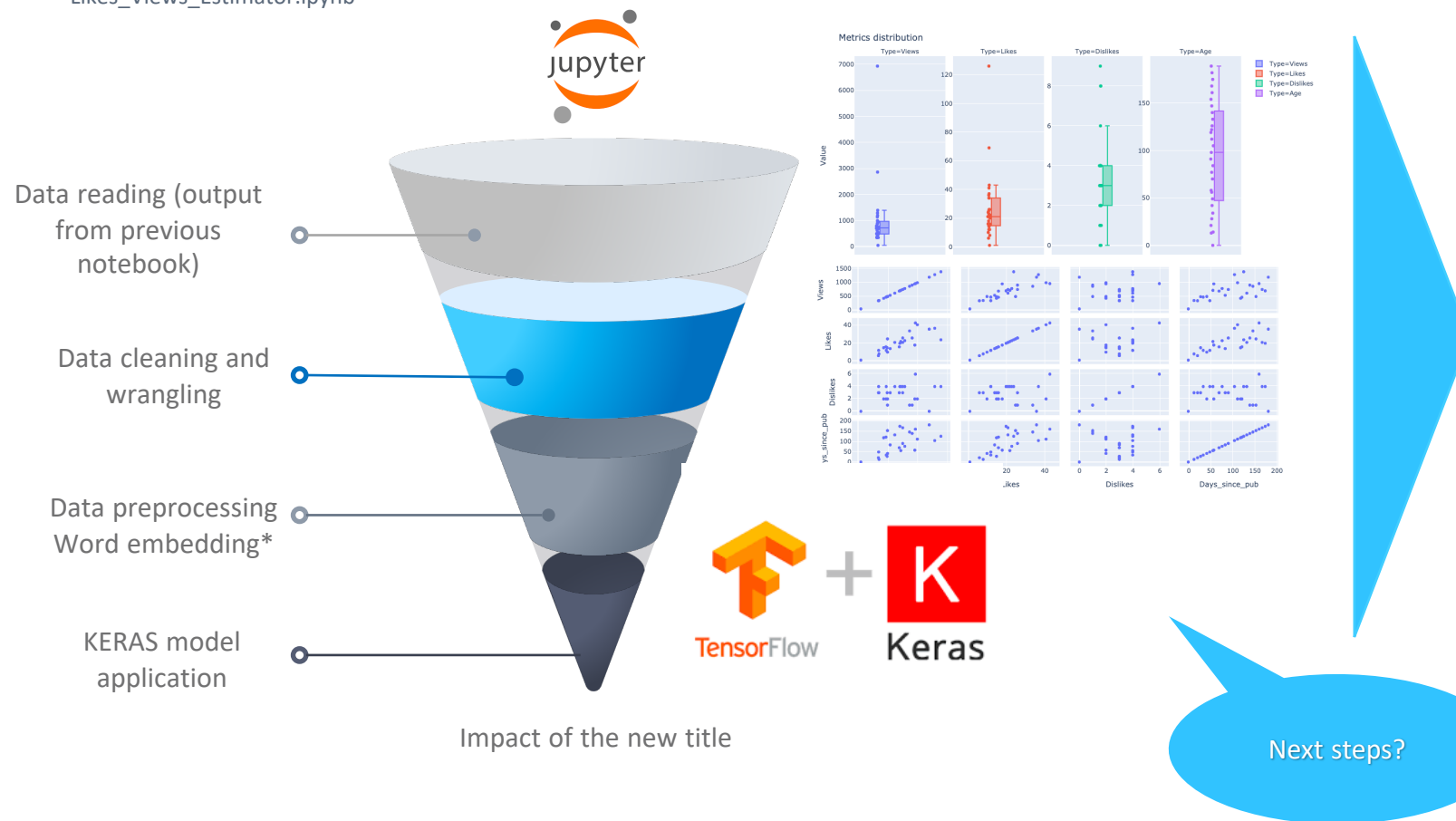
Daily views evolution

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

Data reading (output from previous notebook)

Data cleaning and wrangling

Data preprocessing Word embedding*

KERAS model application

Impact of the new title

Next steps?

- Include other podcasts to have a more robust model

- Estimate the likes/views over time

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

Thank you for your time

April 2020