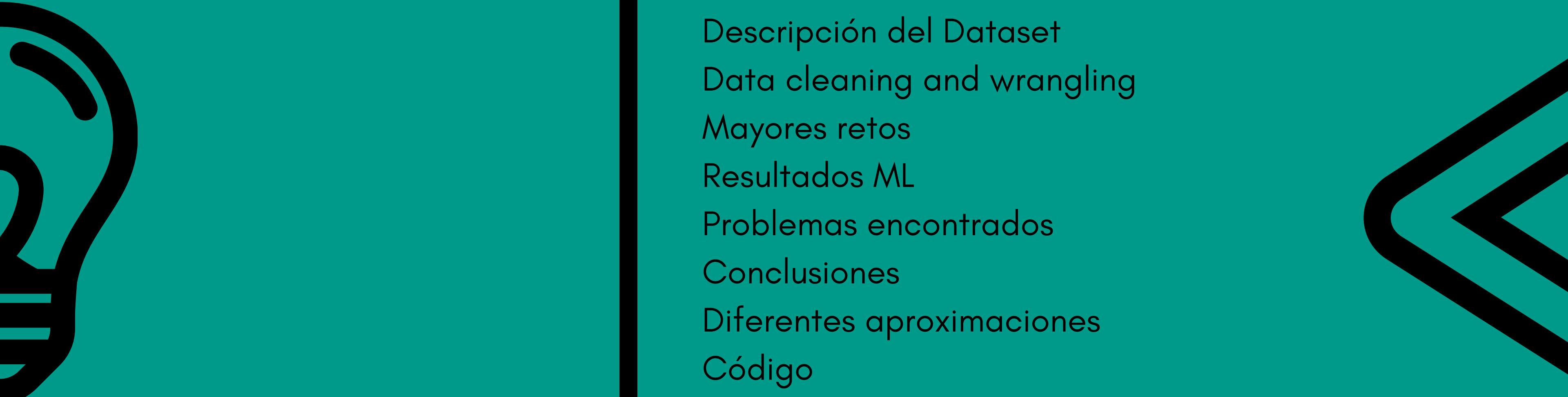


Sistema de Recomendación: Codewars

Alberto García Cobo

CONTENIDO



Descripción del Dataset
Data cleaning and wrangling
Mayores retos
Resultados ML
Problemas encontrados
Conclusiones
Diferentes aproximaciones
Código

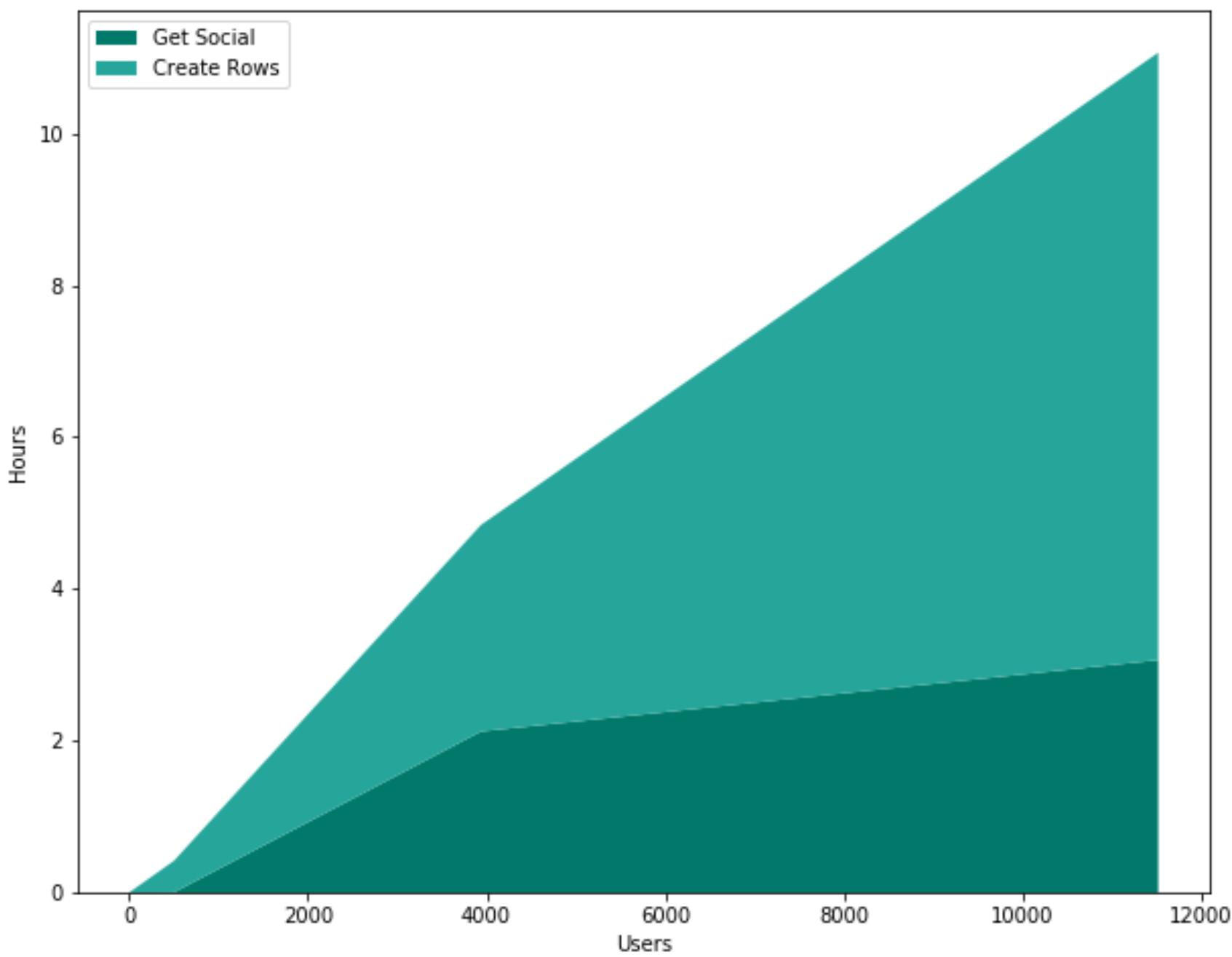
**Usuarios de
Codewars**

**Web Scraping
+ API requests**

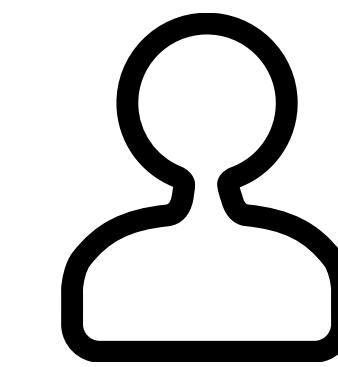
**11500 Registros
3 Iteraciones**

Descripción del Dataset

Coste en Tiempo en cada diferente iteración



Descripción del Dataset



0.0+0.4 horas
2.1+2.7 horas
3.0+8.0 horas

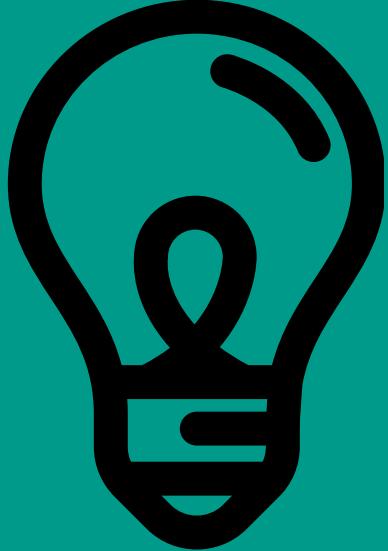
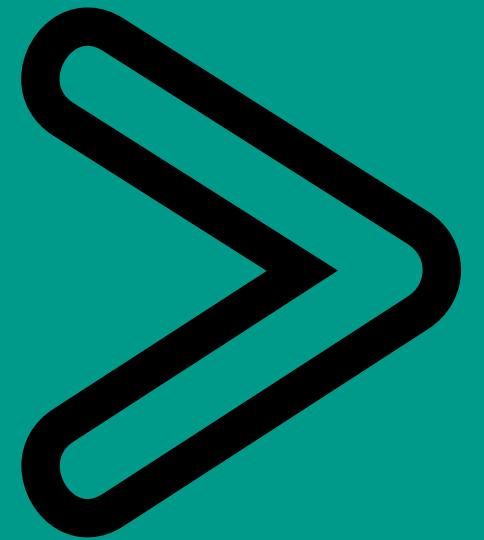
500
3939
11500

**Columnas Nulas
(Tags)**

**Strings
Representando
Numéricos**

**Incrementar
Score Futuro**

| **Data cleaning and wrangling**



"JAVASCRIPT (5 KYU)"
"3 (1 REPLIES)"
"1 (1 STARTED)"
"SEP 2018"
"3,101"

| **Data cleaning and wrangling**

**Combinar Datos
de la API y Web
Scraping**

**Treemap:
javascript +
json ad hoc**

**Sistema de
Recomendación
Vectorial**

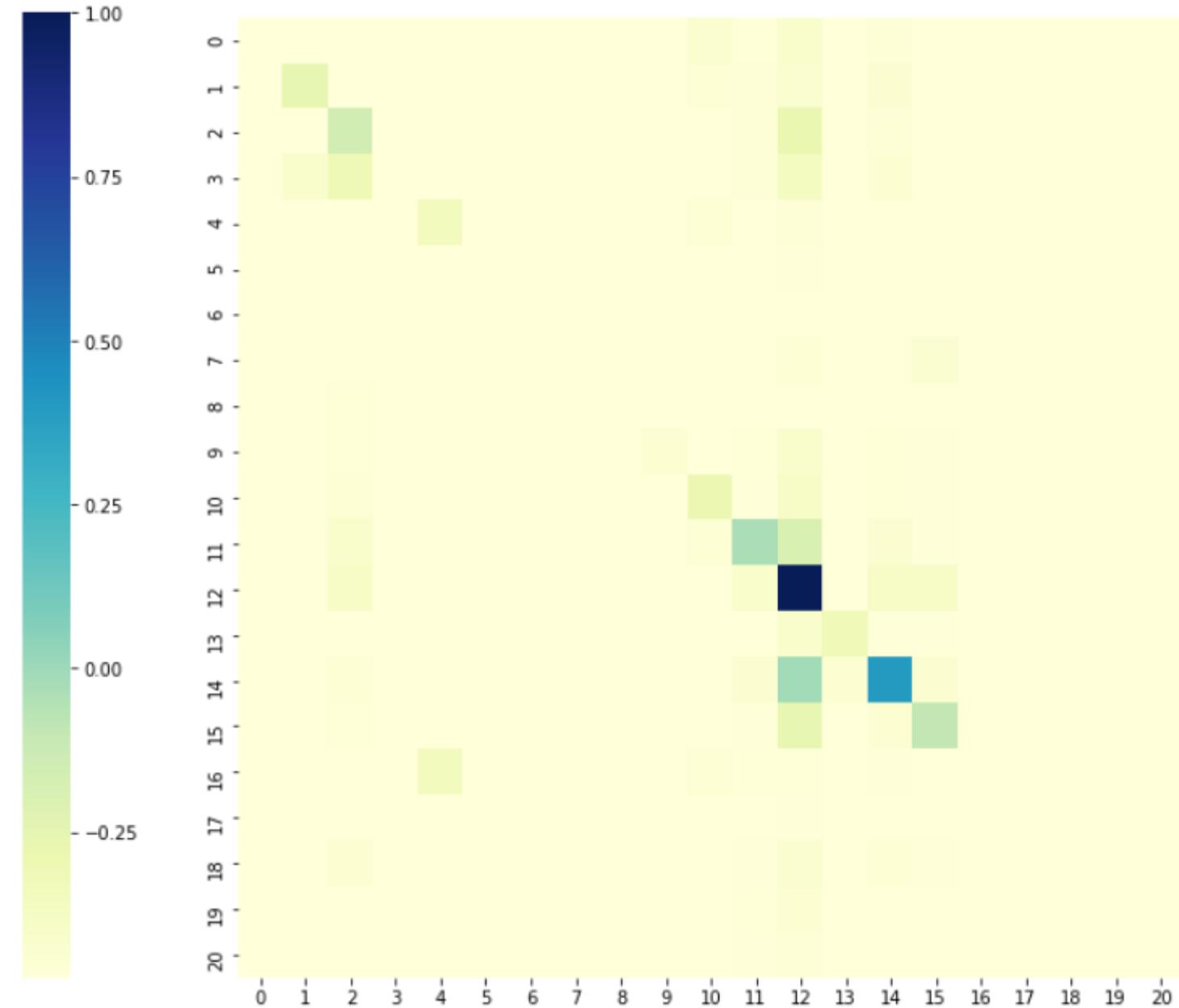
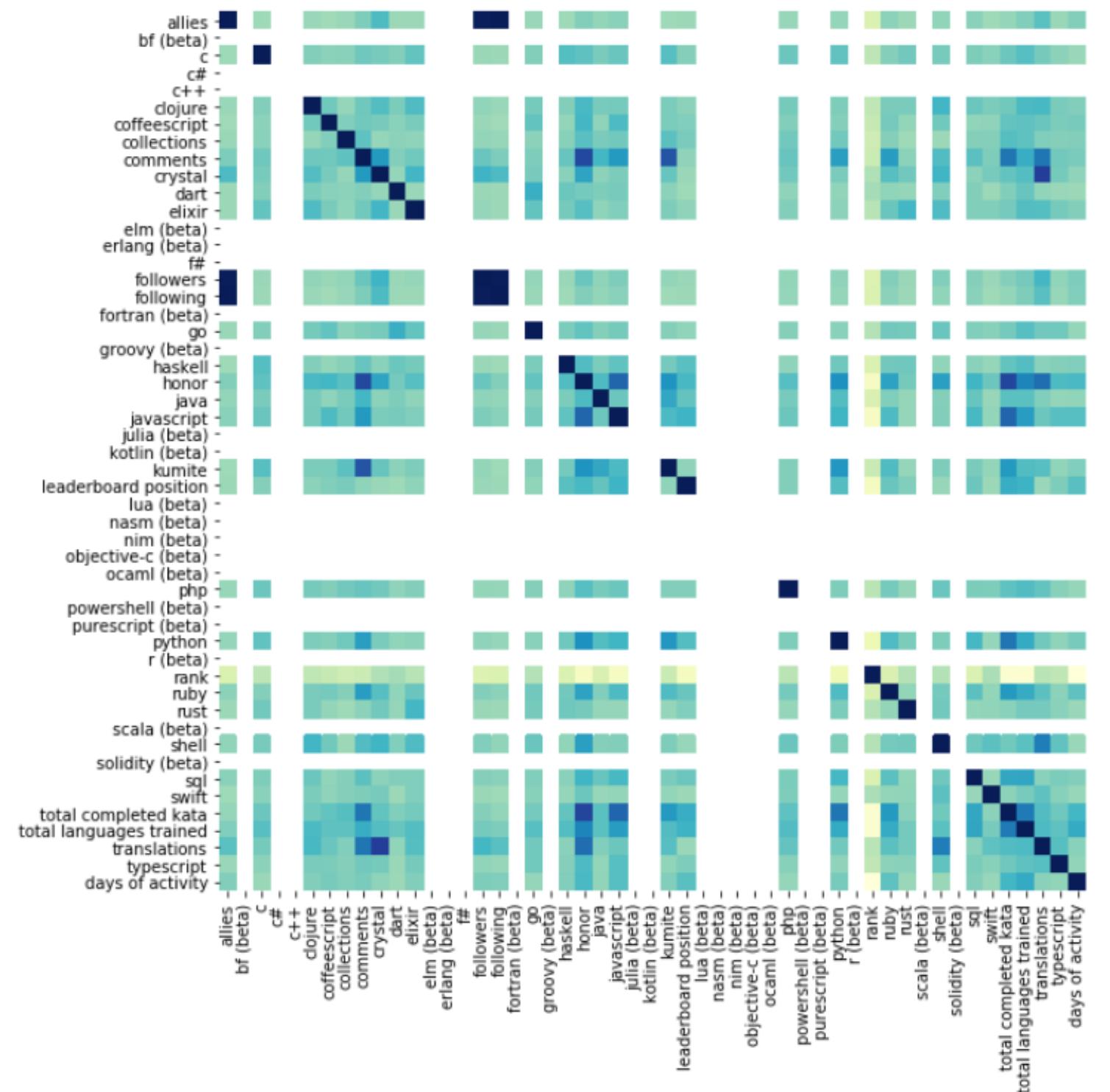
Mayores retos

| Resultados ML

**Correlación
heatmap**

**Supervised
Learning:
score = 0.51**

**Unsupervised
Learning**



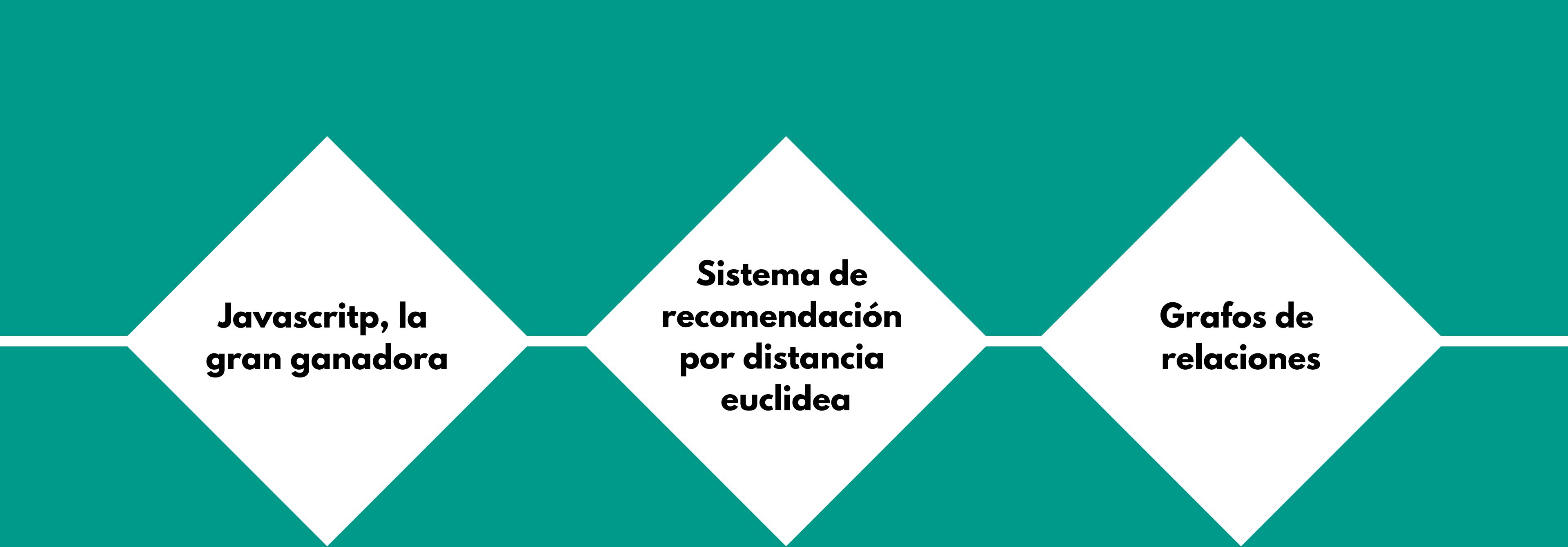
Correlation y confusion heatmap

**Tiempos
ETL**

**Web Scraping
de una Web
Dinámica**

**Creación
de Grafos**

Problemas encontrados



**Javascript, la
gran ganadora**

**Sistema de
recomendación
por distancia
euclídea**

**Grafos de
relaciones**

| Conclusiones



| d3 Treemap

Buscamos expertos en: python, java y javascript



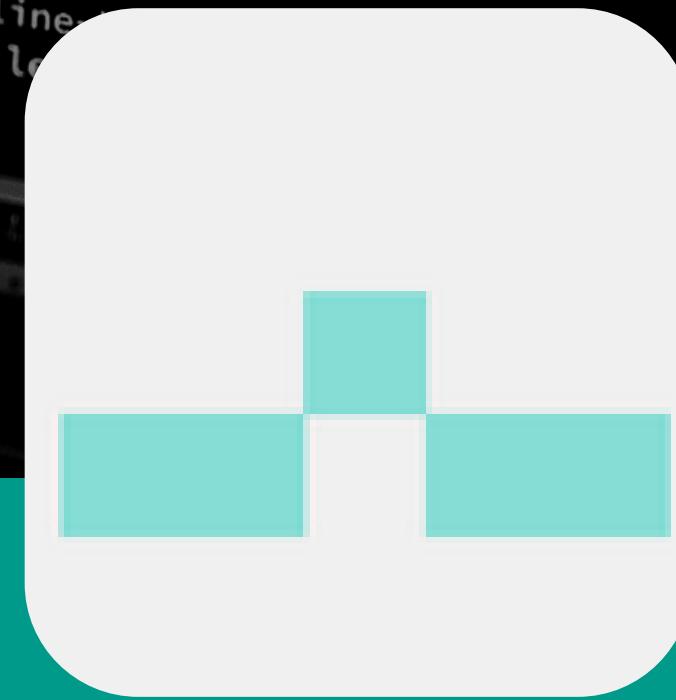
Blind4Basics

<https://github.com/Blind4Basics>



Bubbler

<https://github.com/Bubbler-4>

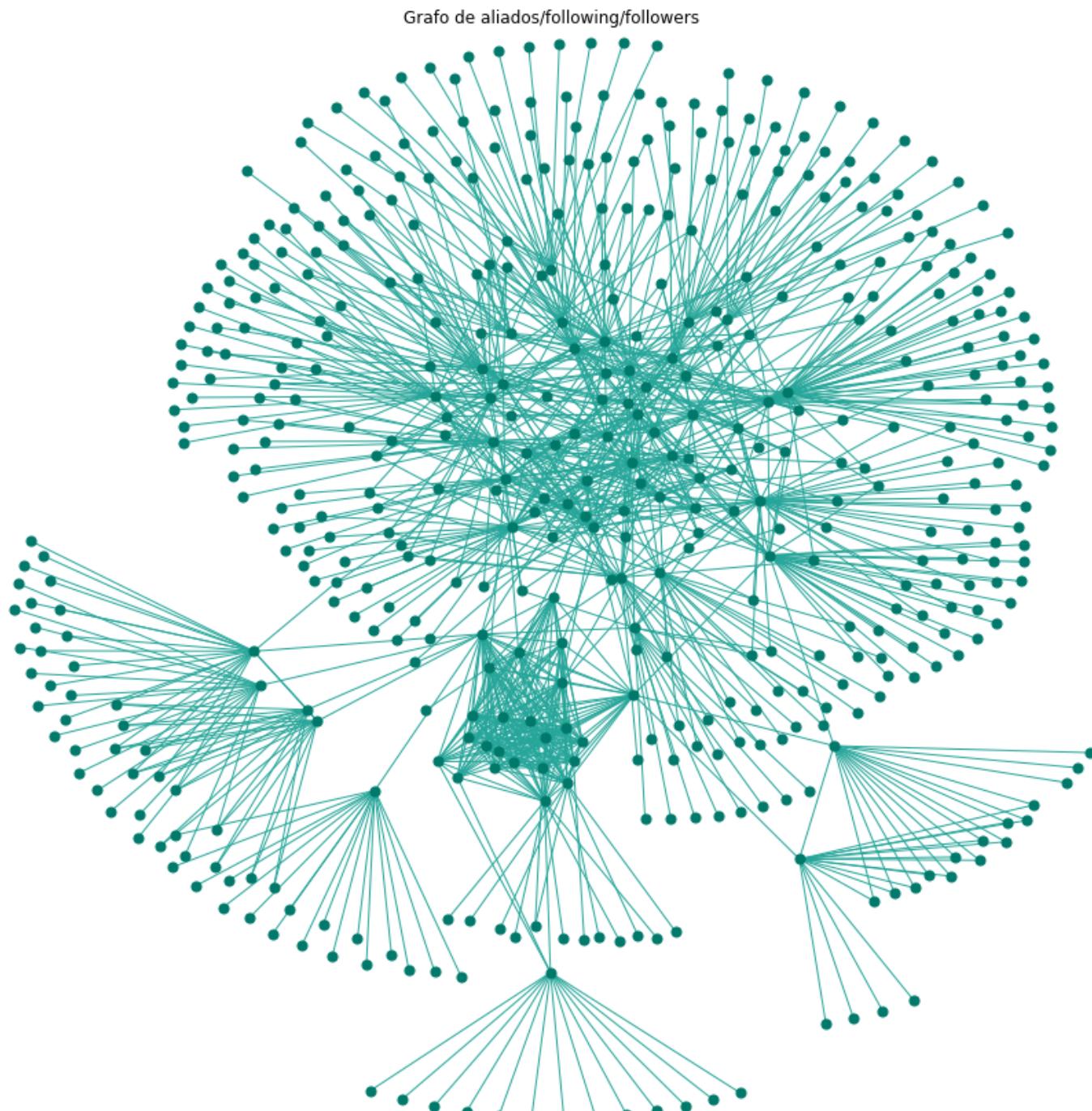


Lyapunov

<https://github.com/Lyapunov>

| **Sistema de recomendación**

□



| Grafos de networkx

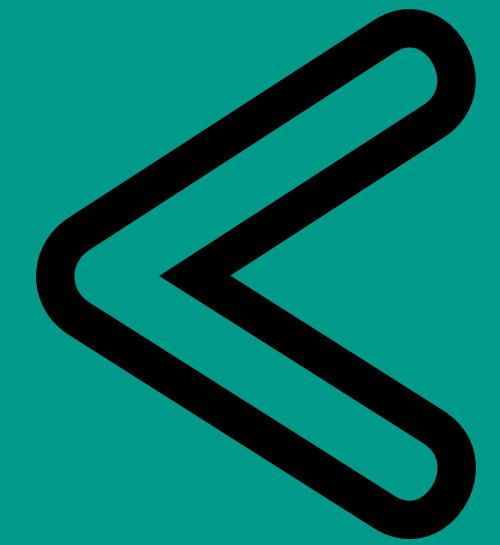
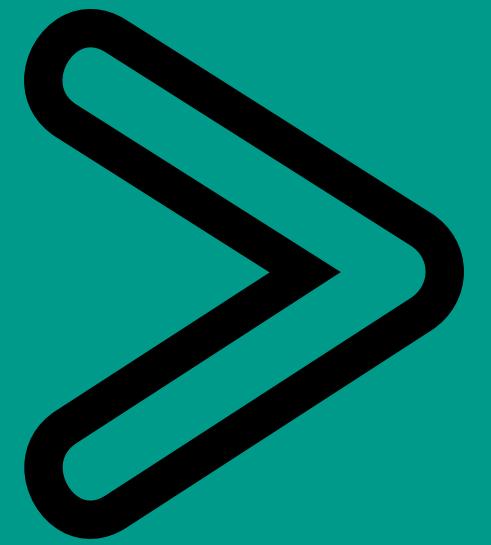
```
graph TD; A((Sistema de Archivos)) --- B((Variables Globales)); B --- C((Sistema de Recomendaciones))
```

**Sistema de
Archivos**

**Variables
Globales**

**Sistema de
Recomendaciones**

Diferentes aproximaciones



import pandas as pd

DF.SHAPE

DF.DTYPES

| Código