

7506 Argerich Entrega - TP1 2023 - 1C

Pedro Gallino

TOTAL POINTS

19.5 / 23

QUESTION 1

1 Ejercicios Pandas 8 / 8

1 ☐

✓ - 0 pts Correcto

2 ☐ - 1

✓ - 0 pts Correcto

2 ☐ - 2

✓ - 0 pts Correcto

3 ☐

✓ - 0 pts Correcto

1 (calificado como bien en la preentrega)

2 Bien. Dos observaciones, si usabas en el final un count y un size o algo similar, sacabas las cantidades sin tener que hacer 2 veces un "mismo loc". Y trata de ponerle nombres entendibles a las cosas, esta bien el ej pero me costo por momentos entender que pasaba

3 Bien!

4 Bien

QUESTION 2

Visus 7 pts

2.1 Histograma 1 / 1

✓ - 0 pts Correct

2.2 Heatmap 1 / 1

✓ - 0 pts Correct

2.3 Violin/Box 1 / 1

✓ - 0 pts Correct

2.4 Visu original 1 / 4

✓ - 1 pts Falta una vuelta creativa que la convierta en una visu realmente original.

✓ - 2 pts Falta bastante trabajo

5 Barplot, bien manejada la metáfora, aun así es muy simple, falta trabajo.

QUESTION 3

3 Ejercicios Spark 7.5 / 8

1 ☐

✓ - 0.5 pts Resuelve pero con errores / de forma ineficiente

2 ☐ - 1

✓ - 0 pts Correcto

2 ☐ - 2

✓ - 0 pts Correcto

3 ☐

✓ - 0 pts Correcto

6 Se debía filtrar las acciones block y reblock. El uso de takeOrdered es ineficiente en este

ejercicio, ya que supone ordenamiento. Para hallar aquel usuario con mayor cantidad de bloqueos, se debía utilizar el reduce.

7 No era necesario ordenar el rdd, podrías haber obtenido el resultado con un reduce.

8 Muy bien resuelto!

Hubiera estado bueno que hagas alguna aclaración sobre por qué lo corres 3 veces en algún lado.

Ojo con el uso del cache en algunos casos, acordate que vale la pena usarlo si vamos a usar ese rdd por lo menos, dos veces.

9 Excelente

1 Ej 2 (★)

Qué porcentaje de las versiones actuales son páginas que se editaron una sola vez

[Link a colab Ej-2](#)

Ej 27 (★★)**4**

¿Cuál es el mínimo que ha durado desde su registro un usuario bloqueado en la plataforma?

[Link a colab Ej-27](#)

3**Ej 54 (★★)**

¿Cuál es la acción más realizada por usuarios que no están registrados?

[Link a colab Ej-54](#)

2 Ej 51 (★★★★)

Si la experiencia de un usuario es la cantidad de logs en los que participó, queremos saber que tanto nos sirve para predecir el futuro vandalismo: ¿Cuál es la probabilidad de que un usuario sea bloqueado según experiencias: ¡10, 10-40, 40-100, ¿100? Tener en cuenta que esta experiencia debe ser PREVIA al bloqueo del usuario.

[Link a colab Ej-51](#)

1 Ejercicios Pandas 8 / 8

1 ☐

✓ - 0 pts *Correcto*

2 ☐ - 1

✓ - 0 pts *Correcto*

2 ☐ - 2

✓ - 0 pts *Correcto*

3 ☐

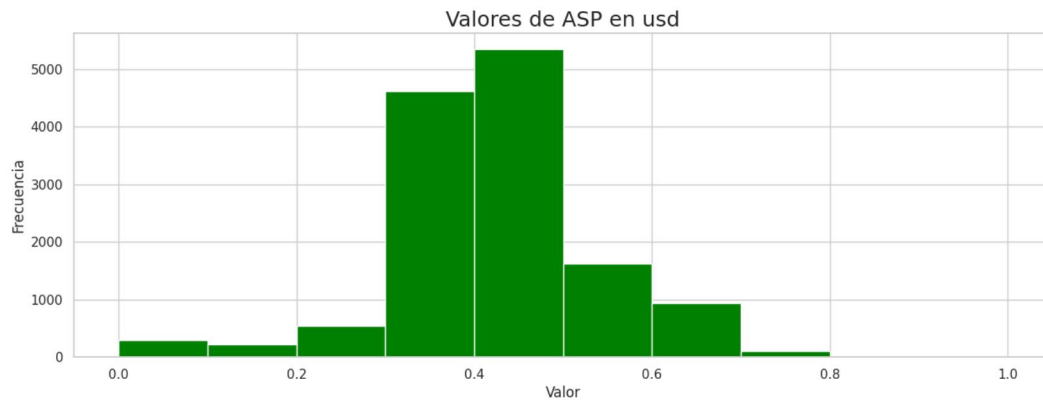
✓ - 0 pts *Correcto*

1 (calificado como bien en la preentrega)

2 Bien. Dos observaciones, si usabas en el final un count y un size o algo similar, sacabas las cantidades sin tener que hacer 2 veces un "mismo loc". Y trata de ponerle nombres entendibles a las cosas, esta bien el ej pero me costo por momentos entender que pasaba

3 Bien!

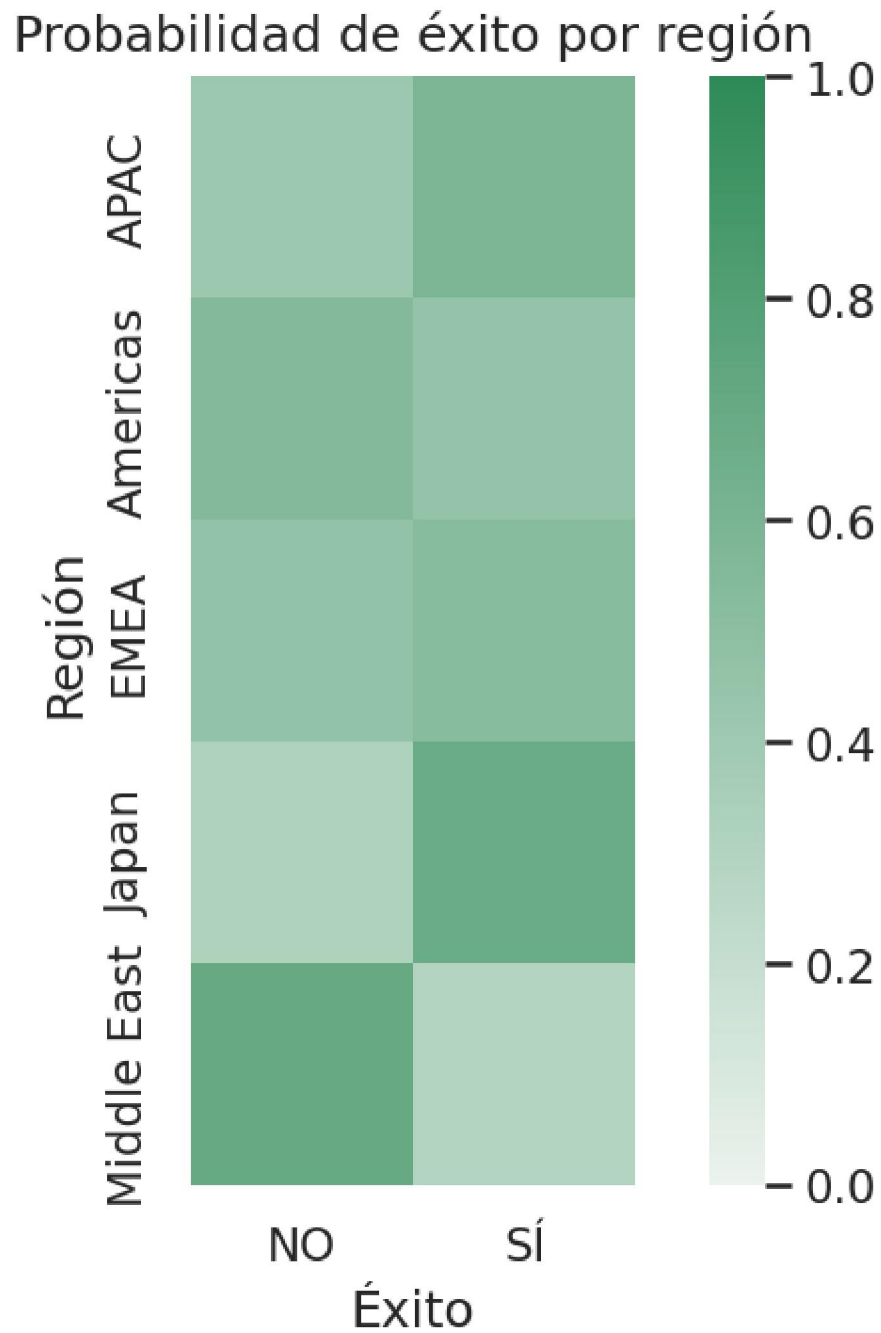
4 Bien



<https://colab.research.google.com/drive/1sOYxSEFZBU8lU4x1gwRnVenGlcFqVJP?usp=sharing>

2.1 Histograma 1 / 1

✓ - 0 pts Correct



https://colab.research.google.com/drive/1jjZn0ExCRcNM0V0y5TlVD_Jqnvwputf3?usp=sharing

2.2 Heatmap 1 / 1

✓ - 0 pts Correct



<https://colab.research.google.com/drive/1yZNEoqlgnNt8iArq2cs7GhmWAYXOkjm8?usp=sharing>

2.3 Violin/Box 1 / 1

✓ - 0 pts Correct



<https://www.vinetur.com/2015060119670/la-10-marcas-que-venden-mas-vino-en-el-mundo.html>

2.4 Visu original 1 / 4

✓ - 1 pts *Falta una vuelta creativa que la convierta en una visu realmente original.*

✓ - 2 pts *Falta bastante trabajo*

5 Barplot, bien manejada la metáfora, aun así es muy simple, falta trabajo.

Ej 34 (★)

6

¿Quién es el usuario que más ha bloqueado a otros?
Link a colab Ej-34

Ej 28 (★★)

9

El porcentaje de contenidos que están publicados cuya última edición no tiene comentario para los usuarios que realizaron 1, ¿10 y ¿100 de las últimas ediciones
Link a colab Ej-28

7

Ej 39 (★★)

¿Cuál es la acción más realizada por usuarios que no están registrados?
Link a colab Ej-39

8

Ej 4 (★★★★)

Considerando el pagelink_sample.csv representar como un grafo en Spark los contenidos de wikipedia (considerando los contenidos de wikipedia)
Link a colab Ej - 4

3 Ejercicios Spark 7.5 / 8

1 □

✓ - 0.5 pts *Resuelve pero con errores / de forma ineficiente*

2 □ - 1

✓ - 0 pts *Correcto*

2 □ - 2

✓ - 0 pts *Correcto*

3 □

✓ - 0 pts *Correcto*

6 Se debía filtrar las acciones block y reblock.

El uso de takeOrdered es ineficiente en este ejercicio, ya que supone ordenamiento. Para hallar aquel usuario con mayor cantidad de bloqueos, se debía utilizar el reduce.

7 No era necesario ordenar el rdd, podrías haber obtenido el resultado con un reduce.

8 Muy bien resuelto!

Hubiera estado bueno que hagas alguna aclaración sobre por qué lo corres 3 veces en algún lado.

Ojo con el uso del cache en algunos casos, acordate que vale la pena usarlo si vamos a usar ese rdd por lo menos, dos veces.

9 Excelente