**Pruebas IT (NODEL)**

**Retos:**

1: Dado el sheet [click](https://docs.google.com/spreadsheets/d/1DRD97TAw2WIuTCG0Nh6BW-aVvDKAgY1wvJb38V-3vU8/edit#gid=0) , dirigirse al tab reto1 donde se encuentra una tabla con los autores que han sido clasificados en base a sentimiento, country y theme. Se necesita que se realice una pivot table en la cual las columnas base sean author y sentimiento y las columnas variables sean country y thema.

[Aqui](https://docs.google.com/spreadsheets/d/1DRD97TAw2WIuTCG0Nh6BW-aVvDKAgY1wvJb38V-3vU8/edit#gid=1217311985) un ejemplo del formato de salida del doc. Las columnas de country y tema serán variables dependiendo del total de countries y temas que exista en el sheet de base, ejemplo si en la base existen 12 countries , esta sección en el sheet de salida deberá tener 12 columnas , del mismo modo con temas.

El sheet de salida tiene que estar en el mismo formato que el del ejemplo , incluido las columnas de cabecera del ejemplo además que el tab deberá ser creado desde 0, no se admite que exista un template fijo al momento de crearlo

Este reto debe ser realizado en python en conjunto con el uso del api de google sheet y el test del reto será modificando la data de la base para comprobar su dinamismo.

2.Extraer de Instagram TODOS los comentarios del siguiente post:

<https://www.instagram.com/p/B166OkVBPJR/>

Este post tiene alrededor de 1480 comentarios y se deberá extraer también los comentarios de los comentarios.

El dataframe de salida deberá contener los siguientes campos:

Post,Caption,Date,likesComment,IdFatherComment,IdChildComment,Username

IdFatherComment es el id de un comentario directo al posts y IdChildComment es el id de un comentario hecho sobre otro comentario

3. Algunas veces la data requerida a extraer no está disponible en alguna api o link oficial de descarga, por eso en algunas ocasiones acudimos al Scraping de páginas web para lograr este reto. Una manera de realizar esto es usando el api de SELENIUM para python , la cual nos permite simular una navegación web , interactuar con la página y bajar la data requerida.

Esta página [link](http://www.pbclibrary.org/raton/mousercise.htm) está orientada a ambientar a las personas con el uso del mouse en sus computadoras. El desafío en este reto es mediante el uso de python y SELENIUM, crear un script que permita avanzar en los diferentes niveles de la página mencionada interactuando con los elementos y siguiendo las órdenes que se muestran. La página tiene 41 niveles , y cada una de ellas equivaldría a 1 punto en la evaluación (el tema vale 35 puntos , si se logra completar los 41 niveles se obtendrá 6 puntos adicionales en la evaluación)

Tip. Cada página tiene una orden o petición específica que se debe hacer, se recomienda crear un diccionario o sistema de reglas dependiendo de la página en donde se encuentra para setear la orden a ejecutar. Ejemplo. Pag 22, clickear todos los links y continuar, Pag 29 doble click en todas las imágenes y continuar

**Entregables**:

Link de GitHub al repositorio:

[stevenaraujo98/EjercicioNodel (github.com)](https://github.com/stevenaraujo98/EjercicioNodel)

Reto 1: Link del sheet con el tab de salida requerido. Código del script en github

<https://docs.google.com/spreadsheets/d/19QrSGevwUbouu9jSvXotbbeIkZyA8r50lqn9UDPs77o/edit?usp=sharing>

Reto 2: csv con el dataframe requerido. Código del script en github

CSV también en el repositorio.

Reto 3: . Código del script en github

**¿Cuánto vale cada tema?**

Reto1 =30 Puntos

Reto2 =35 Puntos

Reto3 =35 Puntos