**Nextline data scientist challenge**

**Descripción del problema a resolver**

* El Blog de Nextline es administrado y escrito por el departamento de comunicación, desde hace ya varios años (<https://blog.nextline.mx> ), este cuenta con notas de diferentes temas, las cuales por esto, tienen distintas clasificaciones, para conocer mejor los alcances y resultados del Blog, así como ver que temas son de mayor interés, se requiere hacer un listado de las notas con su categoría y el número de comentarios.

**Rúbrica y entregables**

* Programa de python que obtenga y procese información de la página web del blog de nextline y genere csv con dicha información.
  + Si solo se saca información de la primera página (20 pt)
  + Si la información está casi completa pero faltan algunos datos (30 pt)
  + Si la información se obtiene de todas las páginas (40 pt)
* El programa debe de generar un documento .csv con el listado de las publicaciones que contenga lo siguiente (ver archivo adjunto “**example-list.csv**”):
  + Título (5 pt)
  + Categoría (5 pt)
  + Fecha (5 pt)
  + Número de comentarios (5 pt)
  + Último comentario (20 pt)
* El programa debe de generar un documento .csv con las categorías de las publicación que contenga lo siguiente (ver archivo adjunto “**example-categorization.csv**”):
  + Contenga todas las categorías (5 pt)
  + Total de publicaciones por categoria (20 pt)
  + Título de las publicaciones en su columnas correspondientes (10 pt)

**Nota importante**

La evaluación del ejercicio se llevará a partir de una serie de rúbricas, cada elemento a evaluar contará con una puntuación (Ver sección anterior). Así, los puntos totales corresponderán de acuerdo a los requisitos que hayas cumplido.

El máximo de puntos que se pueden obtener son 115 puntos.

**Tips**

* Recomendamos utilizar la librería lxml para el procesamiento de html, pero es totalmente a su criterio.
* La paginación del blog solo cambia un número de la url:
  + [https://blog.nextline.mx/page/1](https://blog.nextline.mx/page/2/)
  + [https://blog.nextline.mx/page/2](https://blog.nextline.mx/page/2/)
  + [https://blog.nextline.mx/page/3](https://blog.nextline.mx/page/2/)
  + …………..
* Los comentarios se obtienen accediendo al post (Ejemplo: <https://blog.nextline.mx/katherine-martinez-la-nextliner-deja-intentar/#li-comment-39>)
* Revisa los archivos adjuntos para entender los entregables:
  + **example-categorization.csv**
  + **example-list.csv**
* Url del blog : <https://blog.nextline.mx>