**RELACIÓN DE FUNCIONES PROPUESTAS PARA LA REALIZACIÓN DE PROYECTO DEL 1ER. CUATRIMESTRE CURSO 2020-21 Grupos IS4-1 y TI3-2**

**Estructura general del proyecto**

1. Para el “dataset” (fichero) que se elija el alumno debe fabricar un enunciado similar al que se facilita adaptado mismo que debe incluir en una carpeta “doc”.
2. El “dataset” elegido debe incluirlo en una carpeta “data”
3. En una carpeta de archivo fuentes debe incluir en un módulo Python las funciones de los bloques del 1 al 6, mientras que el test que se propone en el bloque 7 debe de incluirse en otro módulo Python que se denomina como el anterior con el prefijo “test\_”.

Para las **condiciones** que se citan en los respectivos ejercicios se facilitan las siguientes ideas:

* Que un campo o varios tomen unos valores concretos.
* Que un campo sea mayor, menor que un valor concreto o comprendido entre otros dos valores
* Que un determinado campo esté contenido en un conjunto de valores
* Que una determinada propiedad de tipo string comience por alguna letra o secuencia de letras
* Cualquier combinación de las cuatro anteriores.

SAMUEL BUZON GIL TRABAJO SOBRE YOUTUBE\_DATA

1. **Bloque 1: Obligatorio**

Lectura del fichero y carga y devolución de una lista con todos los registros leídos.

1. **Bloque 2: A elegir dos** 
   1. Obtener un conjunto con alguno de los campos. (B)
      1. Obtendremos por ejemplo un conjunto de videos con por su título.
   2. Obtener una lista con dos o tres de los campos de cada registro en los que otro campo cumple determinada condición. (D)
      1. Obtendremos por ejemplo el título nº likes y nº dislikes.
2. **Bloque 3:** A elegir dos (*en la condición de este bloque alguno o algunos de los campos deben tener un valor por defecto*)
   1. Calcular la suma de alguna propiedad numérica de los registros que cumplan determinada condición. (A)
      1. Calcularemos la diferencia de likes - dislikes de aquel/aquellos que tenga un determinado título o url.
   2. Calcular la media ponderada de una propiedad numérica respecto de otra de los registros que cumplan determinada condición. (C)
      1. Calcularemos la media de likes que tienen los videos con x título.
3. **Bloque 4: A elegir uno**
   1. Obtener el registro (o algunos campos del registro) que contiene el valor máximo o mínimo de un campo determinado, de los registros que cumplen determinada condición. (B)
      1. Obtendremos el valor máximo de likes o dislikes que un video tiene.
4. **Bloque 5. A elegir uno**
   1. Obtener una lista de registros (o algunos campos del registro) ordenada con los n registros con mayor (o menor) valor en un campo determinado. Donde “n” es un parámetro que debe recibir la función. (A)
      1. Obtendremos una lista de registros ordenadas por los likes o dislikes de tamaño n.
5. **Bloque 6. A elegir dos**
   1. Obtener un diccionario que permita agrupar, los registros que cumplen determinada condición, por algún campo (clave). A cada clave se le hará corresponder una lista con los registros que contienen esa clave. (A)
      1. Agrupar registros mediante su url como clave.
   2. Obtener un diccionario que permita agrupar, los registros que cumplen determinada condición, por algún campo (clave) y que haga corresponder a cada clave una lista con los registros que cumplen determinada condición. (C)
      1. Agrupar registros mediante su url como clave y poner una condición en la que el título sea x.
6. **Bloque 7: Obligatorio**

Un fichero de “test” que pruebe las funciones anteriores con la siguiente estructura:

* + 1. Debe empezar visualizando:
* El número total de registro leídos
* Los 3 primeros registros leídos
* Los 3 últimos registros leídos
  + 1. Probar cada una de las funciones elegidas de los bloques 2 al bloque 6 con, al menos, dos juegos de parámetros distintos, si la función lo permite.