**README**

En este documento, se detallará la estructura seguida en el proyecto, explicando el contenido de los distintos ficheros y directorios.

**Estructura del Proyecto**

* **CSV.** Directorio que contiene los CSVs utilizados en el proyecto
  + **Originales**: CSV originales usados para la generación de CSVs con una mejor estructura con los scripts de Python
  + **Generados**: CSV generados por los scripts de Python, los usados posteriormente en R.

**Nota:** En ningún momento se realizaron modificaciones sobre los CSV originales, solo se reestructuró la información contenida en ellos en otro CSV generado en Python para facilitar la carga en el datawarehouse en R

* **Scripts Pyhton de Limpieza.** Directorio con los scripts usados para generar los CSV que se utilizarán en R.
* **Carga DWH.** Directorio con el código R usado para cargar el datawarehouse, hay dos versiones una usada para la 1º iteración y otra usada en el resto de las iteraciones (se mejoró el proceso de carga).
* **Modelos.** Carpeta en la que se encuentran todos los modelos probados durante las distintas iteraciones, dichos scripts poseen alguna explicación de lo realizado en cada iteración. Los modelos definitivos se encuentran en la carpeta Iteración Final, donde el modelo principal corresponde al fichero **modeloDWH\_it4\_2007.R**
* **Sql.** Directorio con la de la base de datos con datos y sin datos. (La versión definitiva del datawarehouse). La base de datos correspondiente a la versión definitiva del Datawarehouse es **min\_v4\_Definitiva.sql**.

**Nota:** Ejecutar directamente los scripts de Python sobre los CSV no generarían unos CSV 100% válidos para ser usados en R, ya que además de por los scripts han pasado por algunas mejoras manuales, estos CSV perfectos están dentro de la carpeta con el Proyecto R.

El motivo de estas mejoras manuales es la simpleza que ofrecen los editores actuales (ejemplo: cambiar todos las ',' de un fichero por un '.', esta operación se puede hacer fácilmente con un editor, implementarla en algún lenguaje seria también relativamente sencillo, pero suponía una pérdida de tiempo para el desarrollo del proyecto.