**METODOLOGIA ESTRUCTURA DE CARPETAS PROYECTO DE CIENCIA DE DATOS:**

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamente

**METODOLOGIA NOMBRE DE ARCHIVOS:**

Independientemente de la fase del proyecto en la que nos encontremos es necesario nombrar a los archivos con una estructura común. A continuación, se propone un ejemplo:

**XX\_fase\_NombrePrograma\_fecha.py**

**XX** siempre es el número de programa para ejecutar (01,02...).

A continuación, la **fase** es en la que estemos del proceso: CD para comprensión de datos, PD para preparación, MOD para modelado...

**NombrePrograma**: algo identificativo y corto que defina lo que realiza el script.

**Fecha**: fecha con formato añomesdia

Un ejemplo en nuestro caso podría ser el siguiente:

01\_PD\_Preparacion\_reviews\_20230110.py

**METODOLOGIA NOMBRE DE LAS TABLAS:**

* RAW: Cargamos la tabla entera en el formato que R/Python determine.
* SRC: Hacemos cambios y modificaciones para dejar la tabla más o menos depurada. Cambio de formatos al correcto, alguna regla lógica para depurar info, si es necesario cruzar con alguna...
* TMP: Tablas temporales
* VAL: Tabla Validada. Está ya fetén y super bien auditada
* MOD: Tabla con las variables que se vayan a usar en el modelado.
* VUN (Visión única de negocio): Si tenemos que crear una tabla que integre la VAL, MOD y los resultados del modelo

RAW > SRC > TMP > VAL > MOD > VUN

La nomenclatura en el caso de las tablas es la siguiente:

**tipoTabla\_Identificador\_fecha.tipoDeArchivo**

Un ejemplo en nuestro caso podría ser el siguiente:

RAW\_interacciones\_20230110.csv -> SRC\_interacciones\_procesadas\_20230210.pickle