

METODOLOGIA ESTRUCTURA DE CARPETAS PROYECTO DE CIENCIA DE DATOS:

- 📁 0 Datos
- 📁 1 Comprension de negocio
- 📁 2 Comprension de datos
- 📁 3 Preparacion de datos
- 📁 4 Modelado
- 📁 5 Evaluacion
- 📁 6 Implementacion

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 año-mes-día

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 fétén 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