(2) Jueves

- Python
> Anaconda > Jupyter
> Importar bibliotecas
> Conexión con SQL
> Query donde se hace JOIN STATS con Goles_casa y Goles_vis
> Eliminar columnas duplicadas (team y season)
> Ver la correlación de las variables con el target "win"
> Nuevo DF con variables ~~correlacionadas~~ con mayor correlación (goles_casa, goles_visita, total_pacs, corner_taken, touches, season).
> ~~Ver nulos, duplicados, etc~~
> Limpieza datos, mínimos, máximos, nulos, outliers
> Revisar la distribución de los datos

> Train y Test del Modelo Lineal
> DF 2 donde este de la primera a la penúltima temporada

1° Tratar
> Dropear target "wins"
> Dividir categoricos y numericos
> Normalizar numericos
> Estandarizar numericos
> Tratar categoricos "season" con OneHot Encoding
> Concatenar con X_norm o X_standarized con Encoding

2° Test
> ~~Split~~ Split train test
> Crear Modelo
> Entrenar modelo
> Predecir data
> R^2
~~X~~ normalizadas (X_norm) estandarizadas (X_standar..)
$R^2 = .75$ $R^2 = .73$

3° Cross-validation
> Elegir modelo → X_norm
> Hacer validación del modelo con los datos de la ~~X3~~ última temporada. (Cross-Validation).
Comparar resultados de Predicción de "wins" con ~~el~~ las verdaderas "wins"

Norma