

```
---
title: "Actividad Semanal de la clase 3"
output: html_document
---
```

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```{r}

---
```

Dataset

En esta libreta vamos a trabajar con un conjunto de datos que contiene los datos de la FIFA de la copa del mundo. Disponéis de los archivos ya descargados en la sección de recursos. Utilizaremos estos tres ficheros:

- WorldCupMatches.csv: Datos de los partidos disputados
- WorldCupPlayers.csv: Datos de los jugadores que han jugado en cada partido
- WorldCups.csv: Datos de las copas disputadas y los resultados

Introducción

Mediante esta libreta vamos a explorar los datos almacenados en estos tres ficheros y a responder preguntas haciendo operaciones y transformaciones con pandas. Además de completar las celdas de código para responder a los ejercicios, cuando así se indique se deberá incluir texto en `_markdown_` explicando la información o las conclusiones extraídas.

```
```{r}
library(dplyr)
library(tidyr)
```

```{r}
Cargamos el conjunto de datos WorldCupMatches
df_partidos <- read.csv("WorldCupMatches.csv")
```
```

1. Exploración

Usar las funciones ``head`` y ``summary`` para explorar el dataframe ``df_partidos``. ****Documentar en markdown**** brevemente la información almacenada en cada dataframe (columnas y tipo de datos).

No es necesario hacer una documentación exhaustiva, tan solo comentar aspectos principales como qué tipo de información almacena el DataFrame, cuántas filas y columnas tiene, y comentar las columnas que se consideren más relevantes._

```
```{r}
head(df_partidos)
```

```{r}
summary(df_partidos)
```
```

2. Limpieza de datos

Una de las cosas que llama la atención es la alta presencia de valores perdidos en el DataFrame ``df_partidos``. Cuando trabajamos con datos obtenidos del mundo real siempre nos toparemos con problemas relacionados con la medición, captura o almacenamiento de dicha información.

```
```{r}
is.na(df_partidos)
```
```

Localiza las filas con valores perdido. Analizar y ****documentar en markdown**** a qué se deben

estos valores perdidos. Finalmente usar la función `[drop_na]` (https://tidyr.tidyverse.org/reference/drop_na.html) sobre el `DataFrame` para eliminar los valores perdidos del `DataFrame`, y almacena el resultado en ``df_partidos`` de nuevo.

****Completar****

```
```{r}
head(df_partidos)
```{r}
df_partidos <- drop_na(df_partidos)
```
```

### ## 3. Cargar datos

Crear las variables ``df_jugadores`` y ``df_copas`` que contengan los dataframes correspondientes a la lectura de los csv ``WorldCupPlayers`` y ``WorldCups``.

**\*\*Completar\*\***

# Cargamos el conjunto de datos de\_copas

```
```{r}
df_jugadores <- read.csv("WorldCupPlayers.csv")
```
```{r}
head(df_jugadores)
```
```{r}
summary(df_jugadores)
```
```

```
```{r}
df_copas <- read.csv("WorldCups.csv")
```
```

```
```{r}
head(df_copas)
```{r}
summary(df_copas)
```
```

4. Rango temporal

El dataframe ``df_copas`` contiene datos de todos los mundiales disputados. ¿Cuál es el año del mundial más antiguo disputado? ¿Y el año del mundial más reciente?

```
```{r}
min_Year <- min(df_partidos$Year)
print(min_Year)
```
```

```
```{r}
max_Year <- max(df_partidos$Year)
print(max_Year)
```
```

****Pista:**** recuerda que podemos usar funciones de agregación (``min`` ``mean``, etc.) directamente sobre columnas. Por ejemplo, el siguiente código nos muestra el mayor número de goles marcado en un mundial.

```
GoalsScored
```{r}
head(df_copas)
```
```

```

```{r}
max(df_copas$GoalsScored)
print(max(df_copas$GoalsScored))
```

```{r}
min(df_copas$GoalsScored)
print(min(df_copas$GoalsScored))
```

```

****Opcional:**** Una vez localizados los años de interés (más antiguo y más reciente) visualizar las filas completas correspondientes a cada año usando la función `filter`.

```

```{r}

fila_antiguos <- filter(df_partidos, Year == min_Year)

print(fila_antiguos)
```

```{r}
fila_reciente <- filter(df_partidos, Year == max_Year)

print(fila_reciente)
```

```

...

"El documento proporciona una guía detallada para realizar un análisis de datos de los partidos de la Copa del Mundo utilizando R y los paquetes dplyr y tidyr. Aquí se explica el proceso analítico llevado a cabo:

Carga de Datos:

Se cargan tres conjuntos de datos que contienen información sobre los partidos de la Copa del Mundo (WorldCupMatches.csv), los jugadores (WorldCupPlayers.csv) y las copas disputadas (WorldCups.csv).

Exploración de Datos:

Se utilizan las funciones `head` y `summary` para explorar el contenido del dataframe `df_partidos`, proporcionando una visión general de las primeras filas y un resumen estadístico de las columnas.

Documentación en Markdown sobre la estructura y contenido del dataframe, resaltando aspectos como el número de filas y columnas y la relevancia de ciertas columnas.

Limpieza de Datos:

Identificación de valores perdidos usando `is.na`.

Análisis de la causa de los valores perdidos y documentación en Markdown.

Uso de la función `drop_na` de `tidyr` para eliminar filas con valores perdidos, limpiando el dataframe `df_partidos`.

Cargar Datos de Jugadores y Copas:

Creación de las variables `df_jugadores` y `df_copas` al leer los archivos `WorldCupPlayers.csv` y `WorldCups.csv`.

Exploración de estos dataframes usando `head` y `summary`.

Rango Temporal:

Determinación del año del mundial más antiguo y más reciente usando las funciones `min` y `max` sobre la columna `Year` del dataframe `df_partidos`.

Opcionalmente, se pueden filtrar y visualizar las filas correspondientes a estos años usando la función `filter`."