**Proyecto 5: Data Wrangling**

Objetivos de aprendizaje y competencias

1. Objetivos de aprendizaje:
   1. Limpiar y estandarizar datos.
   2. Realizar análisis exploratorio y estadístico de datos.
   3. Aplicar uso gráfico de histogramas, barplots, violín plots y ploteos varios.
   4. Trabajar con más de una base de datos.
2. Competencias que se desarrollan:
   1. Analizar un problema para formular una solución.
   2. Idear un algoritmo para resolver un problema.
   3. Trabajar en equipo resolviendo un problema.
   4. Presentar los resultados de un proyecto.

# Descripción

Realice un análisis exploratorio sobre las BBDD de los mundiales de fútbol de la FIFA, limpieza y estandarización de los datos en base a el concepto “tidy data” del libro “The elements of data analytics style” (link : <http://worldpece.org/sites/default/files/datastyle.pdf>), En base a lo anterior deberá entregar claramente conclusiones basadas en las características analizadas.

# Antecedentes

Para el desarrollo de esta actividad se utilizarán las siguientes tablas las cuales se encuentran adjuntas a este archivo:

* *WorldCupMatches.csv*
* *WorldCupPlayers.csv*
* *WolrdCups.csv*

# Instrucciones

En primera instancia se les solicitara resolver el siguiente problema:

Realizar un análisis estadístico en profundidad con indicadores creados por ustedes, en donde se exponga todo el historial de los mundiales a través de las 3 BBDD, por lo tanto debe:

1. Realizar un análisis exploratorio de los datos identificando la presencia de datos nulos y establezca una normativa para tratar dichos datos, identificar variables categóricas y cuantitativas.
2. Identifique y establezca un protocolo para tratar outliers.
3. Obtenga visualizaciones pertinentes a su análisis sobre los mundiales.
4. Hacer un análisis de la concurrencia de público para los años donde se disputó el mundial FIFA, repita lo anterior según los distintos horarios en los que se disputan los encuentros. Indique los partidos con mayor concurrencia de público.
5. Visualice la cantidad de goles anotados por cada uno de los eventos mundialistas. Luego, muestre la distribución de los goles de local y visita. Comente sus resultados.
6. Visualizar los resultados de los partidos por países (ganados, empatados, perdidos).
7. Realice tanto *Cross - Table* como *Pivot - Table* para mejorar su análisis en base a elementos como por ejemplo, año y ciudad del evento mundialista.
8. Realice un análisis de los jugadores con las mejores estadísticas según los mundiales disputados (participaciones, rendimiento, etc).
9. Evalúe dependencia lineal entre las variables de cada BBDD y obtenga una relación para determinar el ganador de cada partido. Relacione el máximo goleador del mundial con el ganador del torneo.
10. Realice un reporte en donde clarifique sus conclusiones por mundial y de manera clara se de cuenta una descripción estadística específica y a su vez muestre la metodología expuesta para determinar dichas observaciones.

La entrega del trabajo será en un documento de google que contenga el reporte nombrado en el punto ocho. Además, deberá presentar su análisis y resultados por medio de una presentación de no más de 10 slides o diapositivas (formato pdf).