



### **LIMITACIONES DE DATOS ACTUALES**

- Inconsistencia de datos:
  - Existe información que no concuerda con la proporcionada en otros lugares, el ejemplo más claro de esto es la selección de Rusia 2014, se encuentra dentro de los participantes, pero no se encuentra la plantilla de Rusia en 2014.
- Codificación de datos problemática
  - Existen muchos nombres de equipo y de jugadores que poseen acentos, los cuales generan conflictos al tratar de hacer búsquedas o normalizar los datos.
- Sin estandarización de los datos
  - Algunos nombres, no solo dificultan la normalización por acentos y caracteres especiales, algunos poseen nombres en diferente orden, un ejemplo claro de esto son los nombres de Japón, cuyo orden suele variar.
- Datos incompletos o sin relación
  - No se puede realizar la relación entre clubes y tarjetas, debido a que la información de las tarjetas se proporcionó como un consolidado por equipos en los mundiales, y no por jugador, quien posee la información del club, por lo que no se puede relacionar una cantidad de tarjetas a un club.

### **PROPUESTAS DE MEJORA A FUTURO**

- Estandarización de datos
  - Seleccionar un medio del cual obtener la información relacionada a los jugadores de esta manera evitar diferentes nombres para los mismos jugadores, ya sea por acentos o por la visualización de su nombre.
- Detallar información



- No proporcionar consolidados por mundiales o de datos generales por selección, dado que teniendo un detalle se puede obtener un consolidado, pero del consolidado no se pueden realizar análisis más concretos.
- Verificar la integridad de datos
  - Para solucionar los problemas como el equipo de Rusia 2014, o bien, solucionarlo utilizando el primer punto (Una fuente única que facilite toda la información de manera estandarizada).
- Aumento en la cantidad de datos
  - Dado que no se posee información detallada sobre todos los mundiales, sino, sólo de los últimos 3, los datos estadísticos calculados pueden no llegar a ser del todo confiables, esto debido a la poca cantidad de información que se posee o que se cargó en el sistema, por lo que tener un registro más amplio enriquecería el cálculo y las estadísticas calculadas.