Prueba técnica.

Hipótesis:

¿Cuál era su estilo de juego? ¿Al ataque? ¿De posesión, contragolpes, defensivo?

¿Qué jugadores tenía?

PASO 1.

* Los datos que vamos a usar se obtuvieron de la página Kaggle y es el dataset llamado: **Football Data from Transfermarkt.**
* La preparación de los datos se usará para el análisis de por qué el Barcelona de guardiola de la temporada 2008-2009 es considerado por muchos uno de los mejores equipos (sino el mejor) de la historia del futbol, analizaremos sus rivales, su estilo de juego, jugadores, títulos ganados, etc.

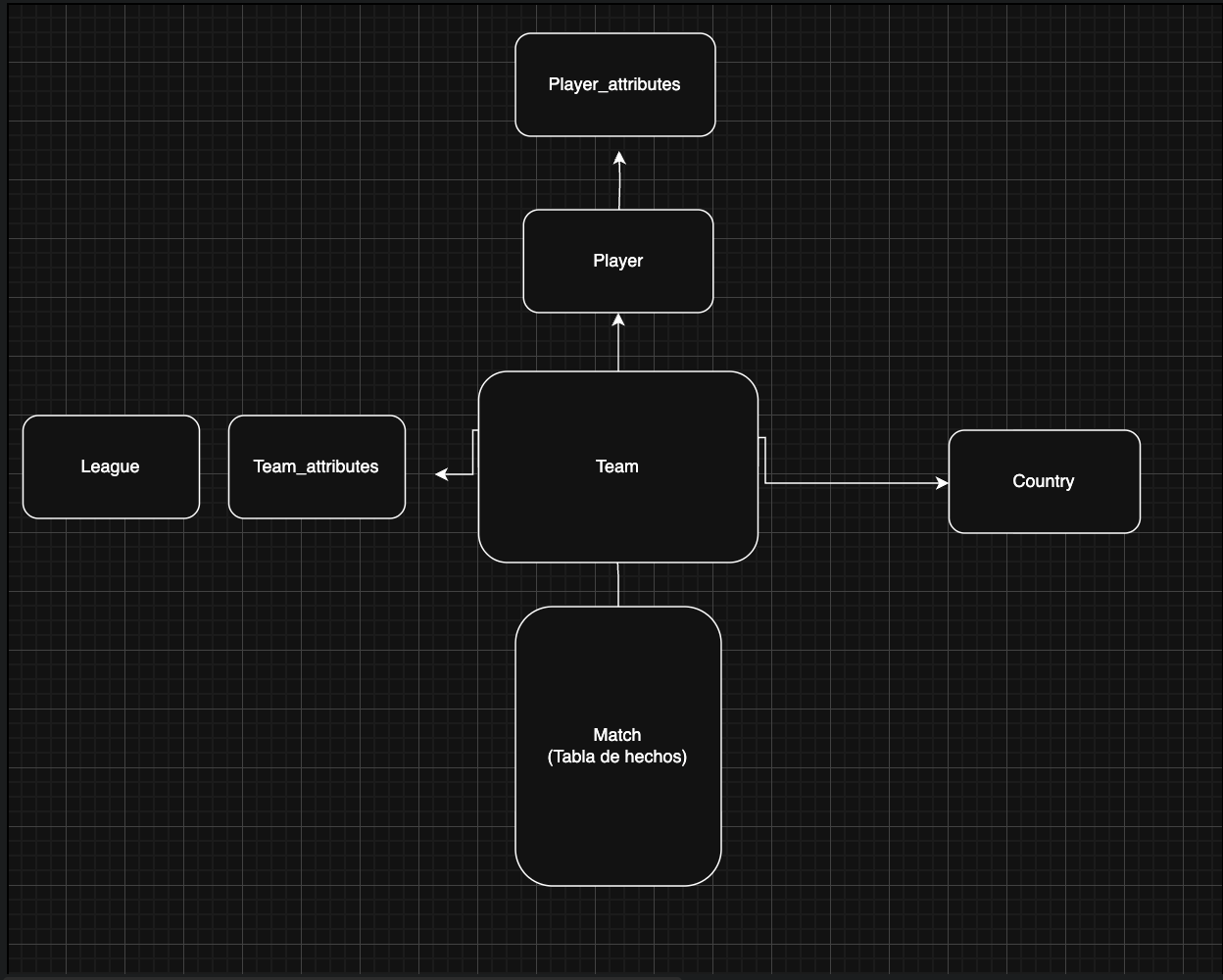
PASO 2. Análisis exploratorio de datos (EDA)

Durante esta fase se realizó una exploración de las tablas disponibles del dataset, con el objetivo de identificar las variables más importantes para el análisis del paso 1. Se revisaron los siguientes aspectos:

* Identificación de las tablas disponibles y los registros de cada una.
* Exploración de las columnas, tipos de datos y rangos de fechas en cada una.
* Análisis preliminar para ver que columnas podrían resolver la problemática planteada en el paso anterior.
* Revisión de valores nulos en las diferentes tablas.

Durante este proceso se detectó que la tabla Players\_Attributes tiene bastantes registros nulos en algunas columnas. Sin embargo, estos registros representan menos del 2% del dataset, por esa misma razón se decidió no hacer ninguna transformación inicial.

**Diagrama:**



PASO 3. Definir el modelo de datos

Se eligió un modelo estrella por su claridad a la hora de realizar análisis más profundos usando consultas más sencillas y eficientes pudiendo juntar varias tablas las cuales nos pueden ayudar al análisis correcto del análisis de rendimiento de equipos y jugadores a lo largo del tiempo.

**Arquitectura del proyecto:**

Para este proyecto, se usó una arquitectura liviana y local enfocada en el análisis exploratorio de los datos. El flujo se compone de los siguientes elementos:

* **Fuentes de datos:** archivo local .sqlite, que corresponde a una base relacional ligera.
* **Lenguaje de programación:** Por su versatilidad y amplio ecosistema en la ciencia de datos.
* **Conexión a base de datos:** se estableció mediante el módulo sqlite3, permitiendo hacer consultas desde Python.
* **Manipulación y análisis de datos:** se utilizó la librería panda, la cual permite hacer análisis, transformaciones, etc, todo en formato dataframes.
* **Entorno de trabajo:** Jupyter notebook, que combina la creación de código, visualización y documentación todo en un mismo espacio.

**Frecuencia de los datos:**

Ya que los datos que se usan para este análisis son de un periodo en específico, no se requiriere una actualización como tal, ya que se trata de un conjunto cerrado de información que no cambia.

Sin embargo, si este proyecto migrará a una comparativa que incluya equipos más actuales, se podría diseñar una arquitectura en la cual los datos se actualicen en un periodo de tiempo establecido:

* Actualización mensual o al finalizar cada jornada.
* Actualización cada vez que el analista desee comparar o analizar un equipo o situación en especifica.

En resumen, para este caso puntual no se necesita una actualización con una frecuencia de tiempo ya que se está determinando un análisis en un periodo de tiempo determinado, ahora bien, si eventualmente se planea usar una frecuencia en producción se deberá terminar un periodo de tiempo semanal o mensual según requerimientos del proyecto.