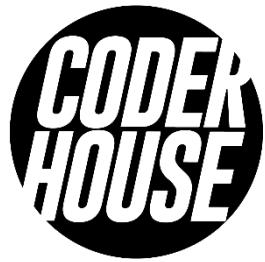




PROYECTO DATA ANALYTICS



Autor: Joaquín Perusia

Institución: Coderhouse

Fecha entrega: 11/2/2025



## Índice

Glosario palabras clave	5
Introducción	6
Temática del proyecto	6
Objetivo	6
Hipótesis	7
Alcance	7
Nivel de aplicación del análisis	8
Diagramas	8
Diagrama Entidad Relación	9
Diagrama de relaciones	9
Listado de tablas	10
Tabla Equipos	10
Tabla Estadios	10
Tabla Partidos	10
Tabla Estadísticas	11
Modelo relacional en Power BI	11
Modificaciones en tablas	12
Segmentaciones elegidas	12
Medidas calculadas	13
Visualización de los datos	14



Página 'HOME'	14
Página 'TEMPORADAS'	14
Página 'ESTADIOS'	15
Página 'EQUIPOS'	16
Página 'COMPARACIÓN'	17



## Glosario palabras clave

**Asistencias** → Pase que conduce directamente a una anotación.

**Conferencia** → Sistema geográfico en que se divide la NBA. Dos conferencias, Este y Oeste, formadas por 15 equipos cada una.

**Foreign key (FK)** → clave foránea. Columna o conjunto de columnas que se usa para vincular dos tablas en una base de datos.

**Localía** → Ventaja de jugar en el propio estadio.

**Modalidad burbuja** → Medida especial durante la temporada 2019 por la aparición del COVID-19, en la que todos los equipos jugaban en un único estadio sin público.

**Modelo predictivo** → Algoritmo o método utilizado para hacer predicciones basadas en datos históricos.

**Porcentaje de tiros de campo** → Relación entre tiros anotados y tiros intentados.

**Porcentaje de tiros libres** → Relación entre tiros libres convertidos y tiros intentados desde la línea de tiros libres.

**Porcentaje de triples** → Relación entre triples anotados y triples intentados.

**Primary key (PK)** → clave primaria de las tablas de base de datos.

**Rebotes** → Captura del balón después de un intento fallido de tiro.



## Introducción

Cuando alguien quiere expresar comodidad y seguridad, suele decir: “*me siento como en casa*”. En el deporte, esta sensación puede ser un factor determinante en el rendimiento de los equipos. Jugar en casa no solo significa competir en un entorno conocido, sino también contar con el apoyo del público, evitar largos viajes y estar habituado a las condiciones del estadio. Pero, ¿realmente esta ventaja se traduce en un mayor porcentaje de victorias?

El análisis de datos en el ámbito deportivo permite obtener información valiosa sobre el rendimiento de los equipos y los factores que influyen en los resultados de los partidos. En este proyecto, se busca predecir la probabilidad de que un equipo local gane un partido en función de sus estadísticas. Para ello, se analizarán métricas clave del rendimiento. Además, se estudiará si la ventaja de jugar en casa tiene un impacto significativo en los resultados.

## Temática del proyecto

El análisis del rendimiento en los deportes ha evolucionado con el uso de la analítica de datos, proporcionando información clave sobre el desempeño de los equipos y factores que influyen en los resultados. En este proyecto, se estudiará la influencia de la localía en los partidos de la NBA, explorando si los equipos que juegan en casa tienen una ventaja estadística sobre los visitantes.

## Objetivo

El objetivo de este análisis es evaluar el impacto de la localía en los resultados de los partidos de la NBA y desarrollar un modelo predictivo que estime la probabilidad de victoria del equipo local en función de sus estadísticas de rendimiento.

Para ello, se analizarán variables clave como puntos anotados, porcentajes de tiros de campo, triples y tiros libres, asistencias y rebotes de cada equipo. A partir de estos datos, se buscará identificar patrones y tendencias que permitan determinar si jugar en casa representa una ventaja significativa para los equipos y qué factores influyen en el desenlace de los encuentros.



Además, el análisis permitirá responder preguntas como:

- ¿Cuánto influye la localía en la probabilidad de ganar un partido?
- ¿Existen métricas específicas que tengan un mayor peso en la victoria del equipo local?
- ¿Ha cambiado la ventaja de la localía a lo largo de las temporadas?

## Hipótesis

La hipótesis central de este proyecto es que los equipos que juegan en condición de local tienen una mayor probabilidad de ganar debido a diversos factores, como el apoyo del público, la familiaridad con la cancha y la reducción del desgaste físico causado por los viajes.

Para probar esta hipótesis, se aplicarán distintos métodos de análisis de datos. Se realizará una exploración descriptiva de los datos para identificar tendencias en las tasas de victorias locales a lo largo de varias temporadas.

Además, se analizará si esta ventaja ha cambiado con el tiempo, tomando como referencia las temporadas 2017, 2018, 2020, 2021 (ignorando la 2019 debido a que se jugó en una modalidad de burbuja) y si existen otros factores que puedan influir en los resultados, como el rendimiento del equipo visitante o la calidad del plantel del equipo local.

## Alcance

El proyecto tiene como objetivo analizar la probabilidad de victoria del equipo local en los partidos de la NBA. Este análisis está dirigido a un público diverso que incluye a analistas deportivos, equipos técnicos, periodistas y fanáticos del deporte.

Para los analistas deportivos y estadísticos, el proyecto busca proporcionar información valiosa sobre las tendencias de victorias locales, ayudando a identificar patrones de rendimiento y mejorar pronósticos en base a datos históricos. Además, los equipos técnicos y entrenadores podrán utilizar estos datos para ajustar tácticas y estrategias,



aprovechando las métricas de desempeño para tomar decisiones más informadas tanto en partidos locales como de visitante.

Por otro lado, el análisis también apunta a periodistas y medios deportivos, quienes podrán enriquecer sus coberturas con información fundamentada sobre el impacto de la localía en los resultados de los partidos. Finalmente, se espera que el proyecto despierte el interés de los aficionados al deporte, permitiéndoles comprender mejor los factores que influyen en las victorias de sus equipos favoritos.

## Nivel de aplicación del análisis

Este proyecto se sitúa en un nivel táctico, ya que proporciona información clave para la toma de decisiones a corto y mediano plazo. El análisis de las estadísticas de rendimiento de los equipos permite ajustar alineaciones y estrategias de juego según las tendencias observadas en los partidos locales y de visitante.

El análisis no solo ofrece una comprensión más profunda del juego, sino que también se convierte en una herramienta para entrenadores y analistas, pudiendo planificar partidos de manera más eficiente.

## Diagramas

Se presentan los diagramas del proyecto en los que se establecen las relaciones correspondientes para el entendimiento del objeto de estudio.

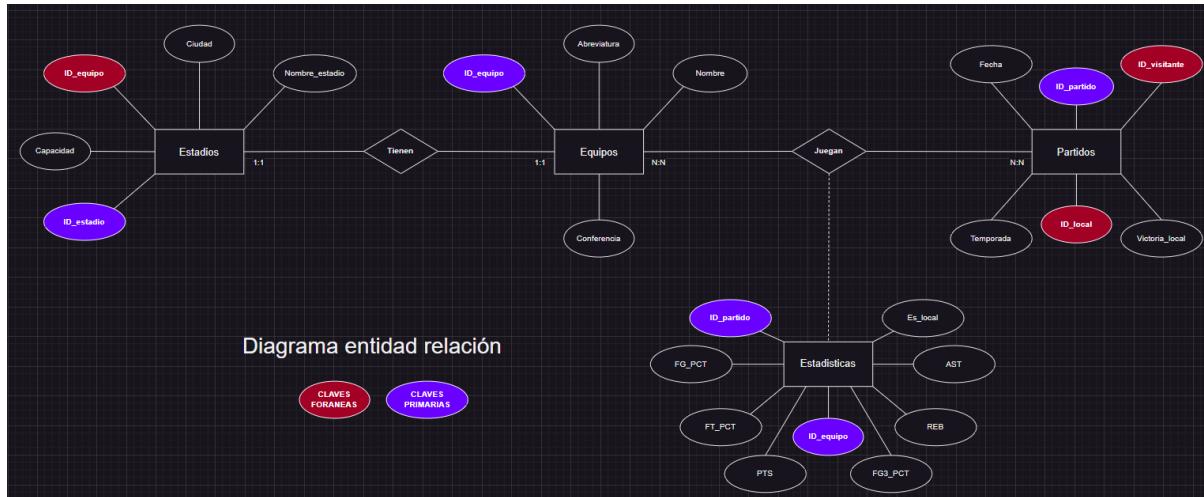
Los diagramas se podrán visualizar de mejor manera en:



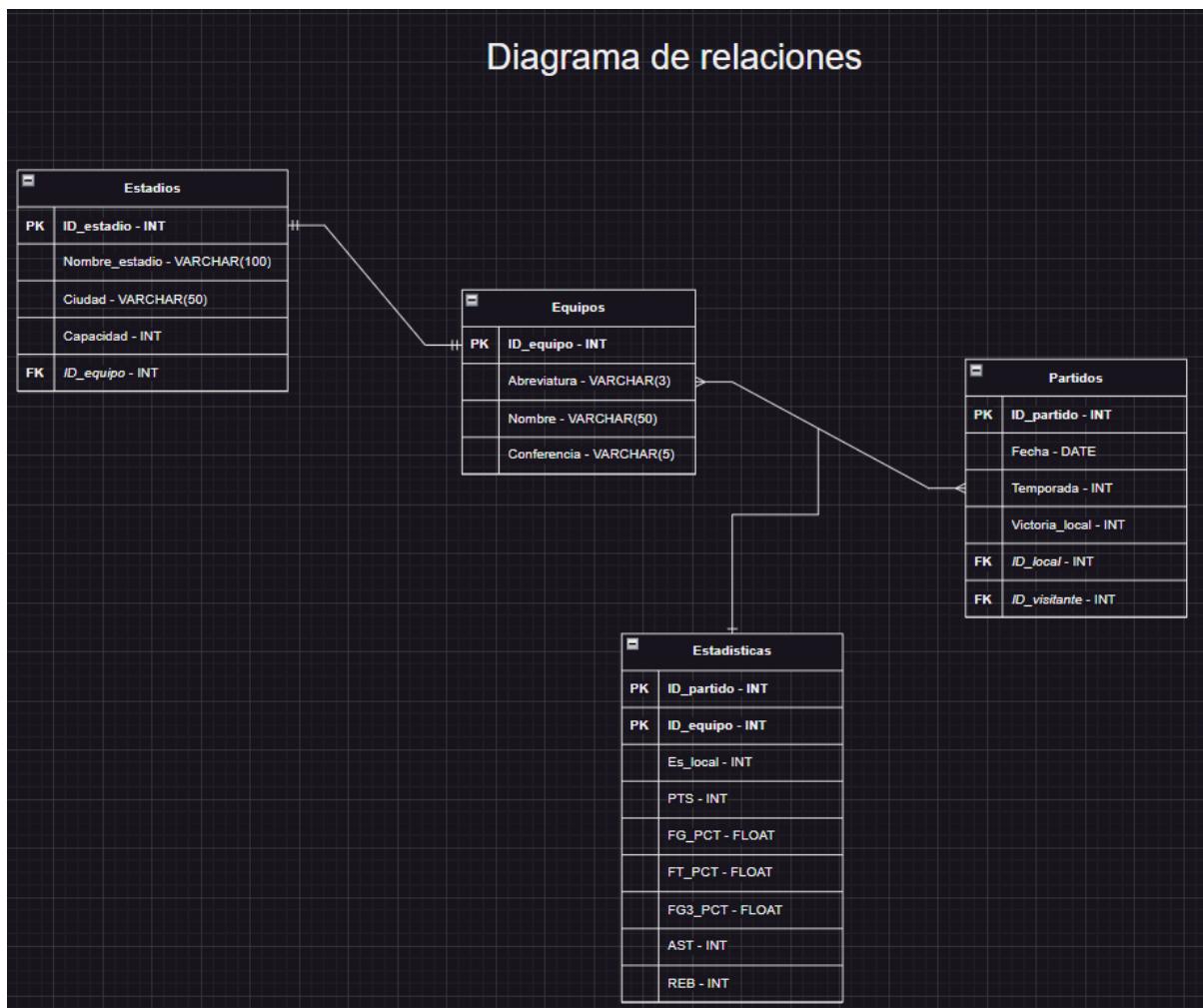
DER-NBA.drawio



## Diagrama Entidad Relación



## Diagrama de relaciones





## Listado de tablas

En las siguientes páginas presentamos las tablas del proyecto final, con sus Columnas, Descripción, Tipo de Dato, Primary key y Foreign key.

### Tabla Equipos

Columna	Descripción	Tipo de dato	Es PK	Es FK
ID_equipo	Identificador numérico de cada equipo	INT	SI	NO
Abreviatura	Abreviación (3 letras) del nombre del equipo	VARCHAR(3)	NO	NO
Nombre	Nombre del equipo	VARCHAR(50)	NO	NO
Conferencia	En que división (Este u Oeste) se encuentra el equipo	VARCHAR (5)	NO	NO

Esta tabla almacena información sobre los equipos de la NBA, incluyendo su nombre, abreviatura y conferencia (Este u Oeste). Es fundamental para identificar a los equipos en los partidos y relacionarlos con sus estadísticas y estadios. La clave primaria (PK) es el identificador numérico de cada equipo.

### Tabla Estadios

Columna	Descripción	Tipo de dato	Es PK	Es FK
ID_estadio	Identificador numérico del estadio del equipo	INT	SI	NO
Capacidad	Cantidad de personas que entran en el estadio	INT	NO	NO
Ciudad	Ciudad en la que se encuentra el estadio	VARCHAR(50)	NO	NO
Nombre	Nombre del estadio	VARCHAR(100)	NO	NO
ID_equipo	Identificador numérico del equipo	INT	NO	SI

Incluye información sobre los estadios de los equipos, como nombre, ciudad y capacidad. Se utiliza para analizar la influencia de jugar en casa y su relación con las victorias locales. La clave primaria es el identificador numérico del estadio. La clave foránea es el identificador numérico de los equipos, para así poder establecer la conexión apropiada.

### Tabla Partidos

Columna	Descripción	Tipo de dato	Es PK	Es FK
ID_partido	Identificador numérico del partido	INT	SI	NO
Fecha	Fecha en que se disputó el partido	DATE	NO	NO
Temporada	Temporada en que se disputó el partido	INT	NO	NO
Victoria_local	Columna que indica si el equipo local ganó el partido	INT	NO	NO
ID_local	Identificador numérico del equipo local	INT	NO	SI
ID_visitante	Identificador numérico del equipo visitante	INT	NO	SI

Contiene datos de cada partido jugado, como la fecha, temporada, equipos participantes (local y visitante) y el resultado del equipo local. Es esencial para analizar el



rendimiento de los equipos a lo largo del tiempo. La clave primaria es el identificador numérico de cada partido. Las claves foráneas son los identificadores numéricos de los equipos que disputaron ese partido, uno para el local y uno para el visitante.

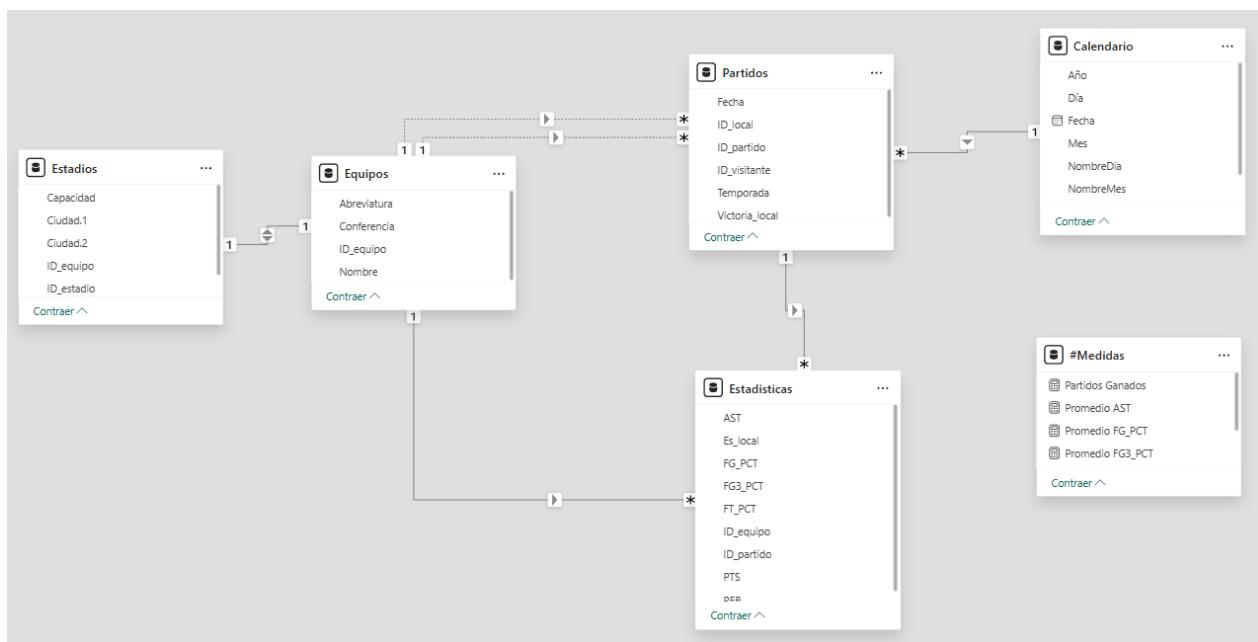
## Tabla Estadísticas

Columna	Descripción	Tipo de dato	Es PK	Es FK
ID_partido	Identificador numérico del partido	INT	SI	NO
ID_equipo	Identificador numérico del equipo	INT	SI	NO
Es_local	Columna que indica si el equipo era local o visitante	INT	NO	NO
PTS	Puntos anotados por el equipo en un determinado partido	INT	NO	NO
FG_PCT	Porcentaje de tiros de campo del equipo en un determinado partido	FLOAT	NO	NO
FT_PCT	Porcentaje de tiros libres del equipo en un determinado partido	FLOAT	NO	NO
FG3_PCT	Porcentaje de tiros de 3 puntos del equipo en un determinado partido	FLOAT	NO	NO
AST	Asistencia totales del equipo en un determinado partido	INT	NO	NO
REB	Rebotes totales del equipo en un determinado partido	INT	NO	NO

Almacena las estadísticas individuales de cada equipo en cada partido, como puntos, asistencias, rebotes y porcentajes de tiro. Permite evaluar el desempeño de los equipos y comparar su rendimiento en partidos locales y de visitante. La clave primaria es una compuesta, formada por el identificador numérico del partido y el del equipo.

## Modelo relacional en Power BI

Se cargaron en Power BI los datos previamente mostrados, realizando algunos cambios fundamentales en el diagrama entidad-relación, viéndose afectadas las relaciones entre las tablas de la siguiente manera:





La relación que antes se daba entre ‘Equipos’ y ‘Partidos’ de muchos a muchos (N:M) dando como resultado una tabla derivada ‘Estadísticas’ se modificó. En el nuevo diagrama presentado las relaciones son:

‘Equipos’ → (1:\*) → ‘Partidos’

‘Equipos’ → (1:\*) → ‘Partidos’

Esta relación se da dos veces, parte de ID\_equipo (Equipos), una es para el ID\_local (Partidos) y la otra para el ID\_visitante (Partidos). Una de las dos relaciones se mantiene inactiva para no generar errores de ambigüedad a la hora de utilizar visualizaciones de Power BI.

‘Partidos’ → (1:\*) → ‘Estadísticas’

‘Equipos’ → (1:\*) → ‘Estadísticas’

Se agregó una tabla nueva a la relación, ‘Calendario’, que sirve para delimitadores de fecha. Se conecta con ‘Partidos’:

‘Calendario’ → (1:\*) → ‘Partidos’

## Modificaciones en tablas

Las tablas mantuvieron su forma original, con la excepción de ‘Estadios’, la cual pasó de tener la columna ‘Ciudad’ a ‘Ciudad.1’ y ‘Ciudad.2’. Esta modificación se realizó a partir de la utilización de un mapa coroplético, separando la columna en dos, conservado en ‘Ciudad.1’ sólo el nombre de la ciudad para así ver un correcto funcionamiento de la herramienta de visualización.

## Segmentaciones elegidas

Se utilizaron filtros de distintos tipos (menús desplegables, listas verticales, mosaicos) para dividir los datos de manera correcta. Las segmentaciones usadas durante el trabajo fueron:

- Selección de Equipos.
- Selección de Temporadas.



- Selección de Conferencias.

## Medidas calculadas

Se creó una tabla '#Medidas' que contiene medidas calculadas basadas en datos de las tablas originales. Las medidas creadas son:

1. Partidos ganados
2. Promedio AST
3. Promedio FG\_PCT
4. Promedio FG3\_PCT
5. Promedio FT\_PCT
6. Promedio PTS
7. Promedio REB

La medida del punto 1 (Partidos ganados) se creó con el fin de hacer la comparación entre los distintos equipos y poder visualizar la cantidad de partidos que el equipo ganó en una cierta temporada. El código utilizado fue el siguiente:

```
Partidos_Ganados =  
VAR LocalWins =  
  
    CALCULATE(  
        COUNTROWS(Partidos),  
        Partidos[Victoria_local] = 1,  
        Partidos[ID_local] = SELECTEDVALUE(Estadisticas[ID_equipo])  
    )  
VAR VisitanteWins =  
    CALCULATE(  
        COUNTROWS(Partidos),  
        Partidos[Victoria_local] = 0,  
        Partidos[ID_visitante] = SELECTEDVALUE(Estadisticas[ID_equipo])  
    )  
RETURN  
    LocalWins + VisitanteWins
```

Las medidas del punto 2, 3, 4, 5, 6, 7 se crearon para poder conocer el promedio de cada métrica de un equipo en determinada temporada y así poder hacer las comparativas correspondientes. Los códigos utilizados fueron los siguientes:

2. Promedio AST = `AVERAGE(Estadisticas[AST])`
3. Promedio FG\_PCT = `AVERAGE(Estadisticas[FG_PCT])`
4. Promedio FG3\_PCT = `AVERAGE(Estadisticas[FG3_PCT])`
5. Promedio FT\_PCT = `AVERAGE(Estadisticas[FT_PCT])`
6. Promedio PTS = `AVERAGE(Estadisticas[PTS])`
7. Promedio REB = `AVERAGE(Estadisticas[REB])`

## Visualización de los datos

El diseño del tablero partió de un mockup generado en Canva que luego fue llevado a Power BI. A continuación, se van a mostrar las páginas creadas.

### Página 'HOME'

Página de inicio con la que el usuario puede navegar a través de las distintas secciones con los botones mostrados en la pantalla.



### Página 'TEMPORADAS'

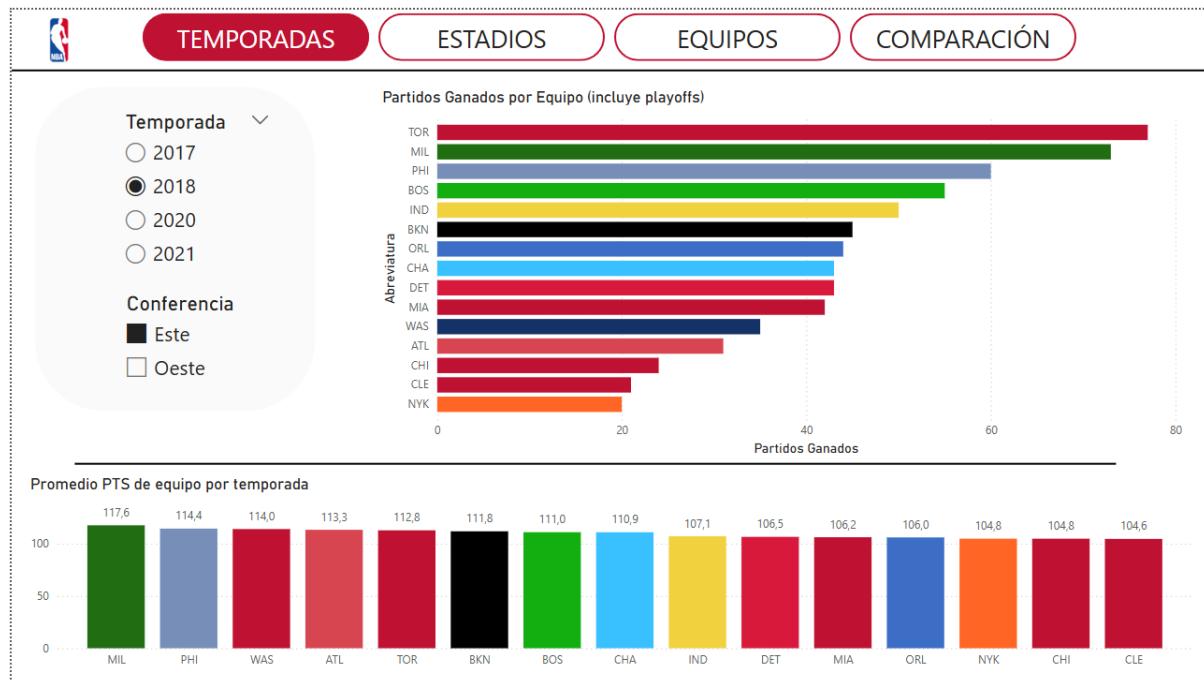
Página que cuenta con dos filtros. Uno de temporada (2017, 2018, 2020, 2021) siendo obligatoria y única una elección. El segundo filtro es de conferencia, pudiendo elegir Este, Oeste o ambas.

Esta página muestra un gráfico de barras horizontales en el que se visualizan las



victorias de los equipos en la temporada seleccionada. Debajo hay un gráfico de barras verticales que muestra la métrica de promedio de puntos de los equipos en la misma temporada.

- Detalle visual, cada equipo está representado con su color de casaca.



## Página 'ESTADIOS'

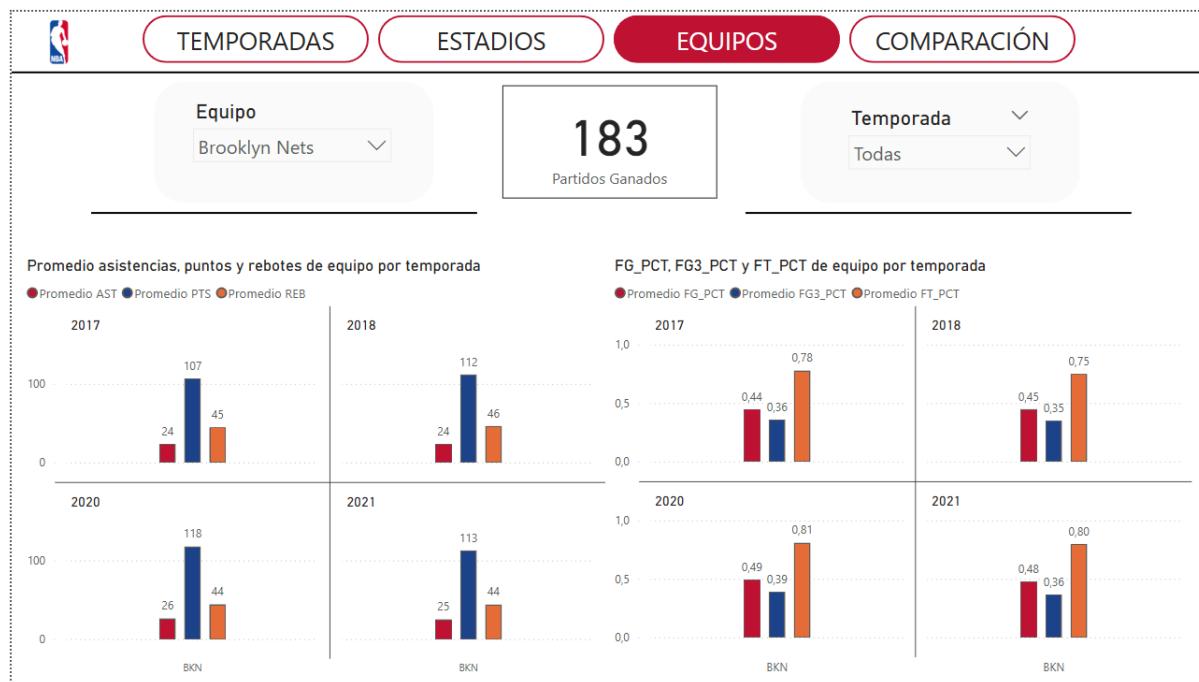
En esta página se muestra un filtro del tipo mosaico en el que se debe seleccionar un equipo para luego mostrar tarjetas con los datos de: conferencia en la que juega, nombre de su estadio y capacidad. A la derecha de las tarjetas se utilizó un mapa coroplético para mostrar la ciudad del equipo elegido previamente.



### Página 'EQUIPOS'

Esta página tiene el fin de poder conocer las estadísticas de un equipo durante las cuatro temporadas contenidas (2017, 2018, 2020, 2021) en el conjunto de datos.

Se muestra una tarjeta con la cantidad de partidos ganados por el equipo y debajo se utilizan dos gráficos de barras apiladas verticales. En el gráfico 1 se muestran los promedios de puntos, asistencias y rebotes del equipo en las distintas temporadas, o solo en una si es que así se selecciona. En el gráfico 2 se muestran los promedios de porcentajes de tiros de campo (FG), de tiros libres (FT) y de tiros de tres (FG3) por temporada.



## Página 'COMPARACIÓN'

En esta última página se selecciona la cantidad de equipos que el usuario deseé y luego una temporada, en base a estos filtros se ajustan los gráficos.

Se encuentra un gráfico de barras horizontales mostrando las victorias de los equipos seleccionados en la temporada.

A la derecha se muestran seis gráficos de áreas en los que se comparan sus métricas (AST, REB, PTS, FG\_PCT, FT\_PCT, FG3\_PCT) en la temporada.

