

# Informe técnico NBA 2003 a 2013



DATAMORF

## Introducción



El presente informe técnico tiene como objetivo analizar el desempeño y las tendencias clave de la NBA durante el período comprendido entre los años 2003 y 2013. Este análisis abarca aspectos fundamentales del juego, incluyendo estadísticas de jugadores, rendimiento de los equipos, y patrones estratégicos que marcaron una década significativa en la historia de la liga.

Durante este lapso, la NBA experimentó importantes transformaciones, tanto a nivel competitivo como estructural, que influyeron directamente en el desarrollo del deporte. La llegada de jugadores de renombre como LeBron James, Dwyane Wade y Carmelo Anthony en el Draft de 2003, así como la consolidación de leyendas como Kobe Bryant y Tim Duncan, redefinieron los estándares de excelencia en el baloncesto profesional.

El análisis realizado utiliza datos históricos y métricas avanzadas para identificar las tendencias de rendimiento, evaluar la influencia de los jugadores clave y examinar el impacto de estrategias ofensivas y defensivas a lo largo de la década. Asimismo, se exploran dinámicas relevantes como la evolución del juego perimetral, el rol del tiro de tres puntos, y los cambios en los patrones de victorias y derrotas.

# Documentación del Proceso de ETL y Análisis de Datos



## Herramientas y Tecnologías Utilizadas

1. Entorno de Desarrollo: Visual Studio Code
2. Librerías Python Implementadas:
  - kagglehub: Importada para descargar y gestionar los datos desde la plataforma Kaggle.
  - pandas: Utilizada para la manipulación y análisis de datos estructurados.
  - numpy: Empleada para cálculos numéricos y manejo eficiente de arrays.
  - seaborn y matplotlib: Librerías gráficas para la visualización de datos.
  - pyodbc: Implementada para la conexión y manipulación de bases de datos en SQL Server.

## Descripción del Proceso de ETL (Extract, Transform, Load)

1. Extracción de Datos:
  - Descargamos los conjuntos de datos directamente desde Kaggle mediante la librería kagglehub.
  - Definimos una ruta específica para la carpeta que contiene los archivos CSV descargados.
  - Creamos una lista de archivos relevantes que serán utilizados en el análisis, filtrando solo los necesarios.
2. Transformación de Datos:
  - Iteramos sobre los archivos seleccionados dentro de la carpeta especificada.
  - Limpieza de datos:
    - Eliminación de registros duplicados.
    - Manejo de valores nulos o vacíos.
  - Unificamos tablas mediante operaciones de merge para obtener una visión consolidada de la información.
  - Creamos nuevas variables que sintetizan la información clave, permitiendo una exportación precisa a archivos CSV.
  - Optimizamos el proceso mediante parámetros definidos para reducir el consumo de recursos y mejorar el rendimiento.
3. Carga de Datos:
  - Exportamos los datos transformados a archivos CSV, listos para ser utilizados en análisis y visualización.
  - Conectamos la base de datos con SQL Server utilizando pyodbc para un manejo eficiente y la posibilidad de integración con futuras fuentes de datos.

## Exploratory Data Analysis (EDA)

Durante esta etapa, realizamos un análisis exploratorio exhaustivo para identificar patrones, tendencias y relaciones entre los datos.

### Pasos Clave del EDA:

- Análisis gráfico mediante seaborn y matplotlib para visualizar:
- Distribuciones de datos.

- Comparaciones entre equipos locales y visitantes.
- Correlaciones entre variables clave.
- Identificación de datos relevantes para el diseño del dashboard en Power BI.

### Consolidación y Preparación de Datos

- Se realizó un "merge" para integrar información sobre equipos locales y visitantes, permitiendo un análisis comparativo detallado.
- Los datos consolidados están optimizados para responder a preguntas específicas y generar insights accionables.

## CONEXIÓN SQL SERVER Y PYTHON



### Conexión con SQL Server

La conexión a SQL Server mediante pyodbc permitió:

- Manipular y consultar los datos de manera eficiente.
- Establecer una infraestructura que facilita la integración de nuevas fuentes de datos en el futuro.

### Uso en Visualización y Dashboard

Los datos procesados fueron integrados en Power BI para crear un dashboard interactivo y fácil de interpretar. Este dashboard permite:

- Analizar información clave sobre equipos, desempeños y tendencias.
- Generar insights específicos que apoyan la toma de decisiones basada en datos.

### Conclusión

El proceso ETL implementado, combinado con un análisis exploratorio detallado y la integración en Power BI, proporcionó una base sólida para el análisis de datos y la generación de insights relevantes. Además, la conexión con SQL Server asegura escalabilidad y flexibilidad para futuros proyectos.

A continuación damos una descripción del código que se ejecuto en Python para la conexión con SQL server:

```
import pyodbc # Importamos la librería
```

```
# Parámetros de conexión
```

```
server = "DESKTOP-UTQDVSQ\SQLEXPRESS" # Nombre del servidor

database = "DATOSFINALESNBA2023"      # Nombre de la base de datos

# Lista de tablas a consultar

tablas = ["game_detail_table", "games_table", "inactive_players_table", "teams_table"]

# Establecemos la conexión

try:

    # Intentamos conectar

    conexion = pyodbc.connect(

        f"DRIVER={{SQL Server}};SERVER={server};DATABASE={database};Trusted_Connection=yes;"

    )

    print("Conexión exitosa")

# Cursor para ejecutar consultas

cursor = conexion.cursor()

# Iterar sobre las tablas

for tabla in tablas:

    print(f"\nConsultando datos de la tabla: {tabla}")

    query = f"SELECT TOP 10 * FROM dbo.{tabla}" # Consulta para cada tabla

    cursor.execute(query)

# Imprimir resultados

for row in cursor:

    print(row)
```

```
except pyodbc.Error as e:
```

```
    # Manejo de errores
```

```
    print("Error en la conexión:", e)
```

```
#finally:
```

```
    # Cerrar conexión
```

```
    #if 'conexion' in locals():
```

```
        #conexion.close()
```

```
    # print("\nConexión cerrada")
```

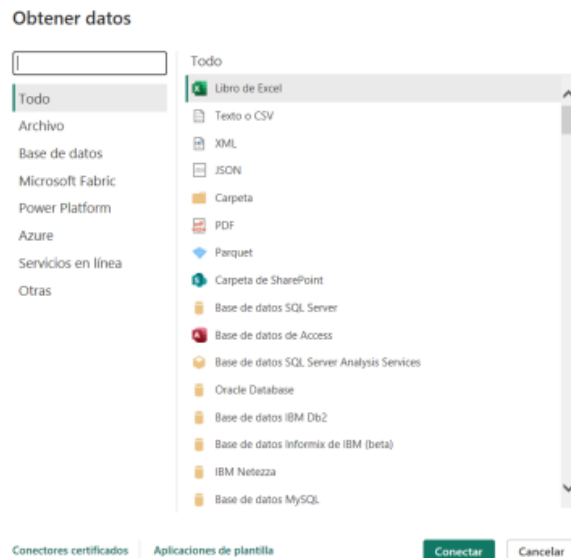
Con estos pasos, se puede conectar Python con SQL y realizar consultas sobre tablas ya cargadas. Este enfoque simplifica el trabajo con datos relacionales y facilita el análisis posterior.

# Dashboard power bi



## Documentacion Power BI

### Obtención de los orígenes de datos .



### Conexión Base de Datos.

## Base de datos SQL Server

Servidor ⓘ



Base de datos (opcional)

Modo Conectividad de datos ⓘ

☒ Importar

☐ DirectQuery

▸ Opciones avanzadas








Aceptar

Cancelar

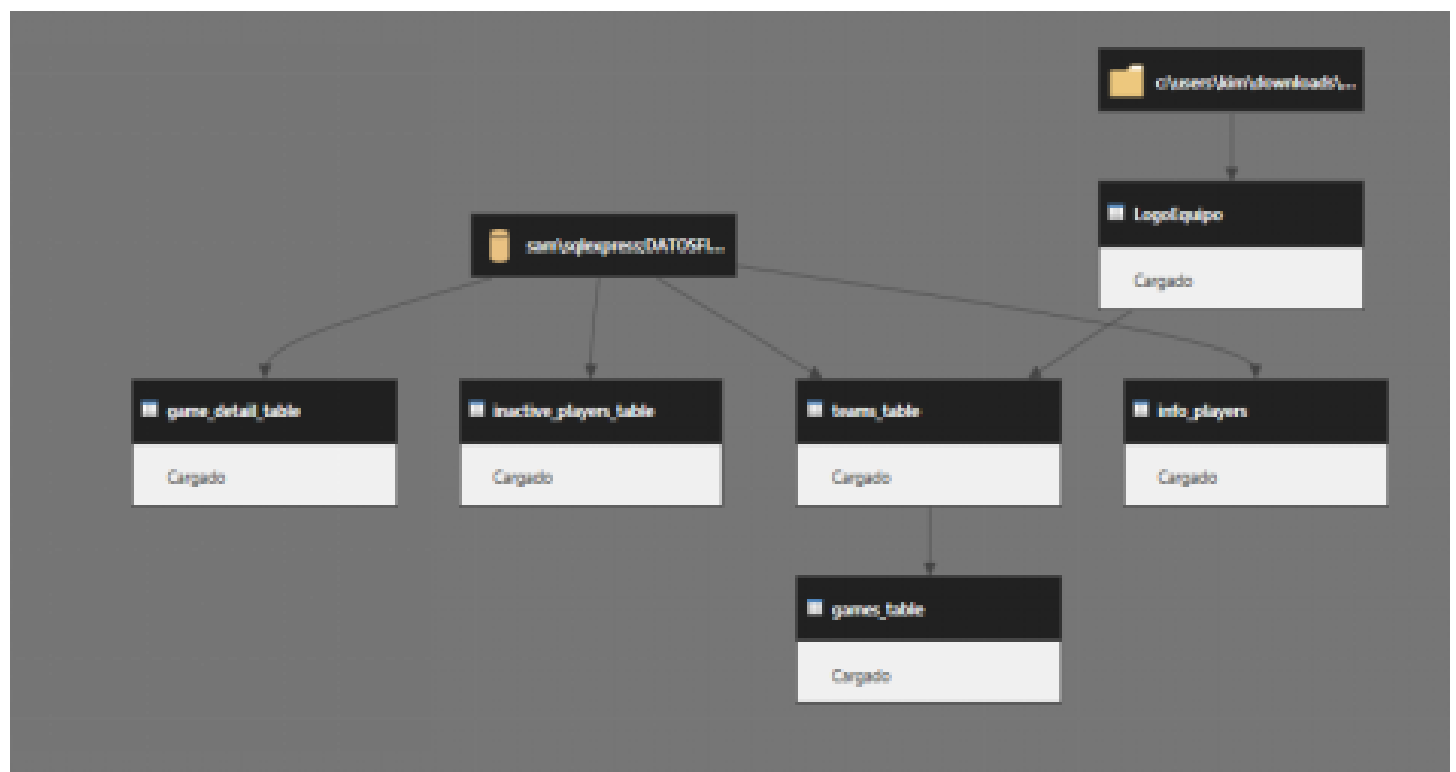
### Relación Tablas SQL



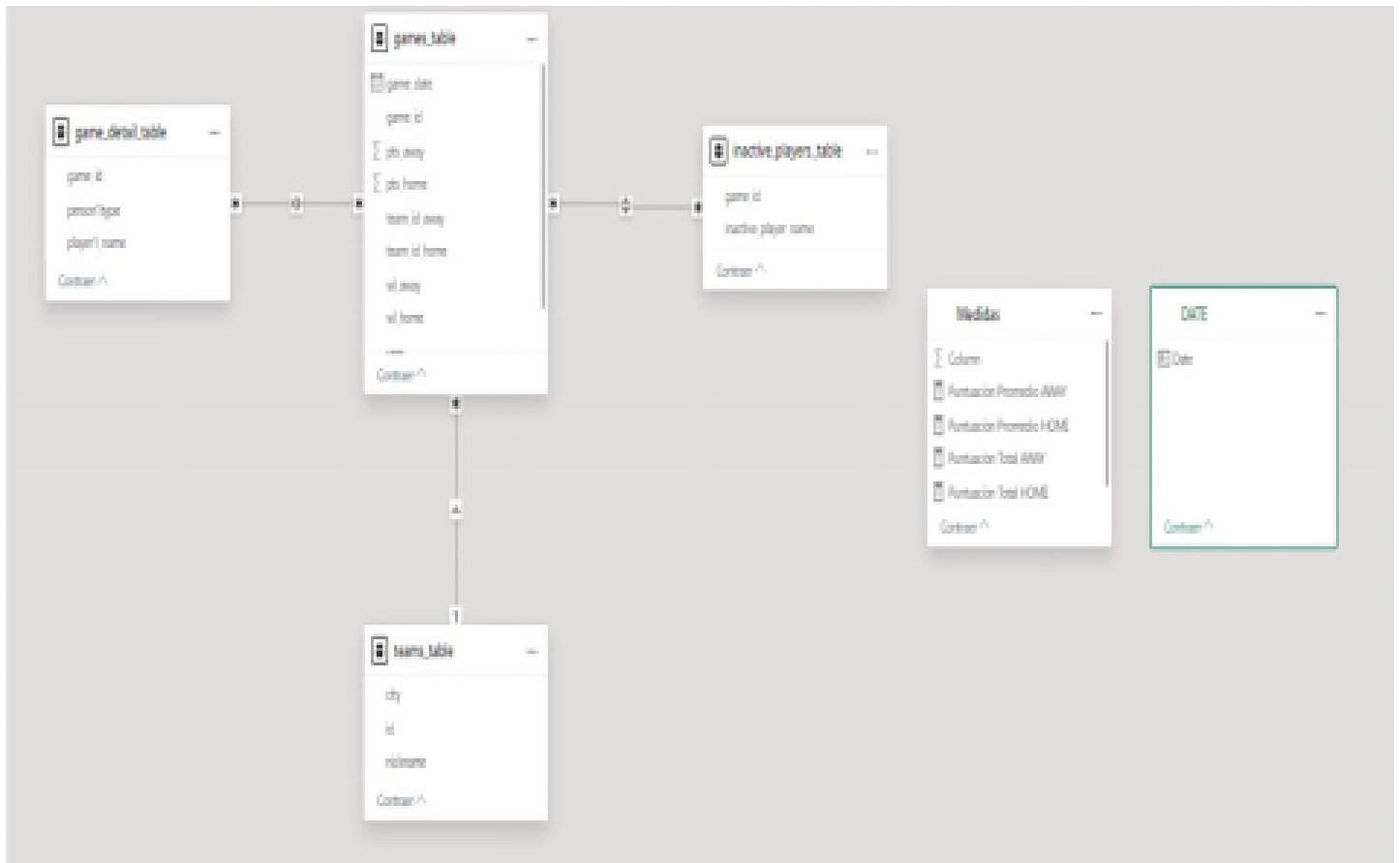
## DATOSFINALESNBA2023 [7]

- ☐  game\_detail\_table
- ☐  games\_table
- ☐  inactive\_players\_table
- ☐  info\_players
- ☐  sysdiagrams
- ☐  teams\_table
- ☐  fn\_diagramobjects

### Dependencias de Consulta Inicial



## Diagrama Inicial



## Transformación de datos

- Corrección de errores de formato fecha, enteros, decimales, texto y numero
- Eliminación de datos Inconsistentes
- Reemplazar valores nulos o erróneos
- Eliminar duplicados
- Corrección de errores de formato

## Cambio de Estructura de Datos

- Ajustar forma en que los datos están organizados para que sean más útiles y eficientes.
- Conversión de columnas en filas.
- Convertir filas en columnas.
- Combinar datos de varias tablas (mediante uniones o fusiones).



## **Filtro de Datos**

- Excluir datos innecesarios.
- Filtrar filas por rango de fechas.
- Se excluyen registros que no cumplen con lo que se requiere.

## **Alimentación de tablas**

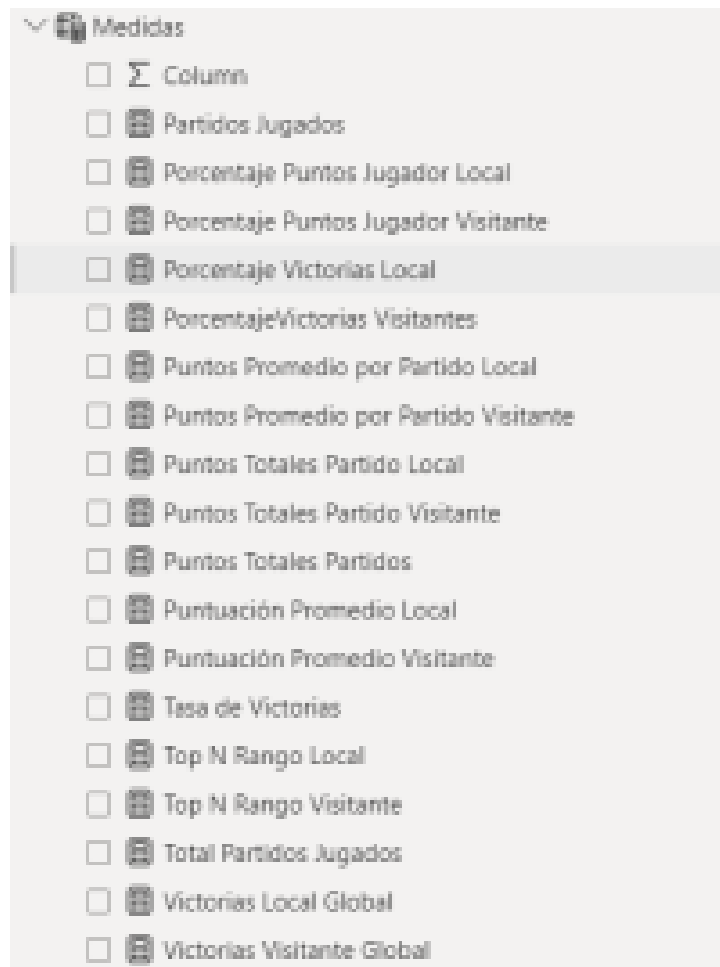
- Creación de nuevas columnas o métricas para agregar valor.
- Creación columnas calculadas.
- Aplicar funciones personalizadas

## **Integración de Datos**

- Combinación de datos de varias fuentes en el modelo.
- Fusión de datos para combinar tablas con relaciones comunes y agregar filas a las tablas que requieran información específica

## **Creación de Cálculos (medidas)**

- Creación de reglas y cálculos específicos para obtención de Insights

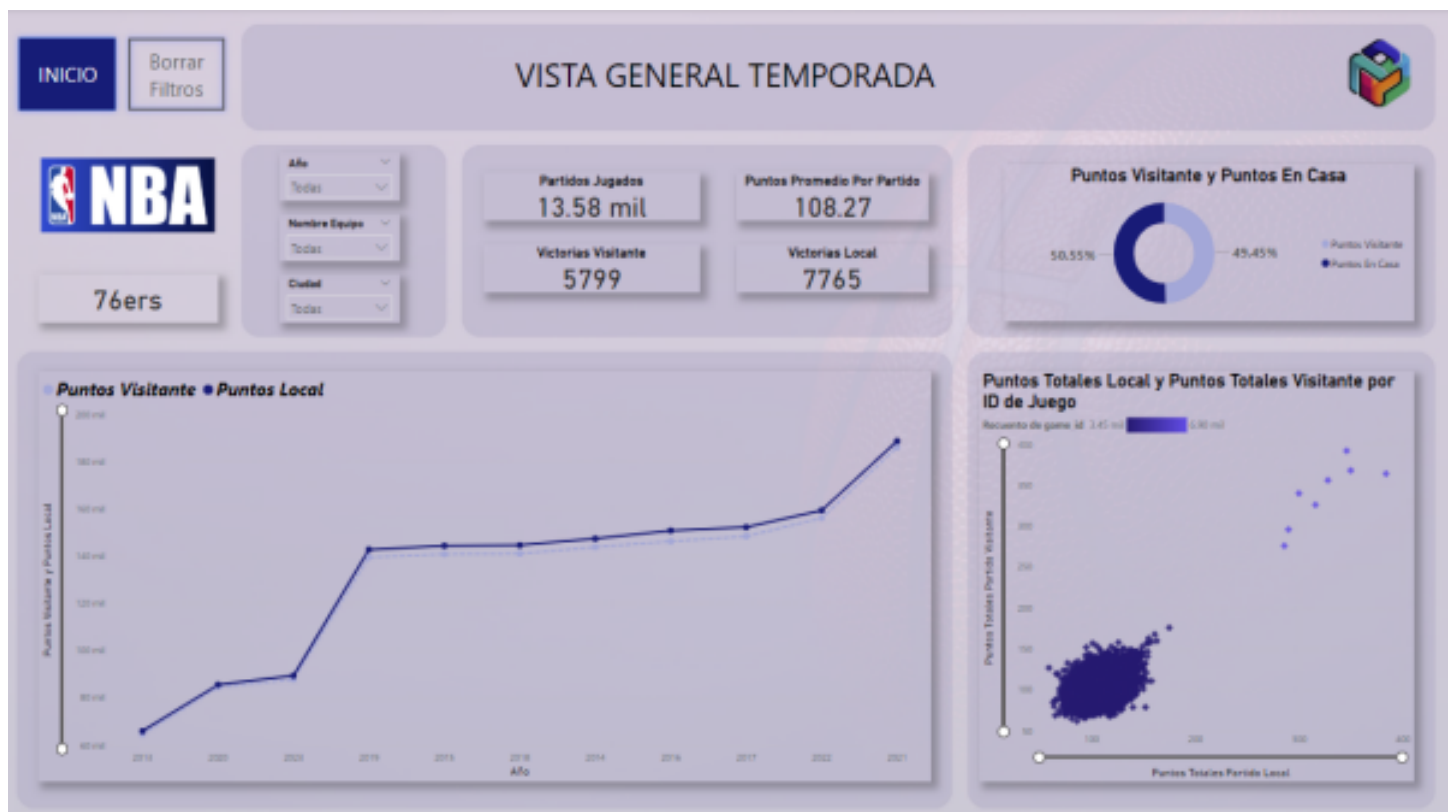


## Relación

- Creación de relaciones entre tablas y medidas.

## Visualización y resultados

- Datos Diagramas y gráficos con insaits necesarios



## Diagrama de Relación Final

