



---

## **Proyecto No. 1** **Desarrollo y consulta de Base de Datos** **(Parejas)**

### **Modalidad y fecha de entrega**

- El proyecto se hará en parejas y debe de ser enviado antes de la fecha límite de entrega: martes 23 de agosto a la hora de la clase
- No se permitirá la entrega o envío de proyecto más allá de la fecha límite

### **Descripción general del proyecto**

El proyecto contempla el uso de tecnologías de bases de datos para la creación y carga de modelos de datos, con el objetivo de utilizar lenguaje SQL para investigación, desarrollo y presentación de resultados sobre preguntas de negocio para apoyo de toma de decisiones.

El conjunto de datos a utilizar serán los archivos en formato CSV que pueden descargarse en Canvas. Este conjunto de datos incluye todos los juegos de la NBA registrados de 1946 a mayo de 2021, así como información de los jugadores históricos y activos actualmente. Para fines del proyecto utilizaremos los datos de la temporada 2015/2016 a la temporada 2020/2021 (es decir, cinco años).

El objetivo general del proyecto es investigar los datos presentados para responder a la siguiente pregunta: **Basado en el desempeño de los equipos y jugadores de los últimos cinco años, ¿en qué equipo invertiría usted para la temporada 2022/2023?**

Para responder fundamentadamente a esta pregunta usted deberá analizar y entender el modelo de datos presentado, definir métricas podrían servir de base para justificar una decisión como esta y desarrollar los queries necesarios para calcular dichas métricas.

### **Etapas 1**

En la etapa 1 se encargará de procesar los archivos CSV proporcionados y levantarlos en una base de datos PostgreSQL donde pueda ejecutar sus queries para análisis. Para esto se recomienda:

1. Descargar los archivos CSV y verificar qué información se incluye en cada uno
2. Crear la base de datos y las tablas en donde almacenará los datos
3. Desarrollar un script de Python (o lenguaje a su elección) que pueda leer la información de los archivos CSV, conectarse a su base de datos y alimentar la información en las tablas creadas

### **Etapas 2**

En la etapa 2 usted ejecutará algunos queries que le permitan familiarizarse con el modelo de datos presentado.

Las preguntas a responder son:

1. ¿Quién es el jugador activo más alto? ¿Y el más bajo?

2. ¿Cuál fue el promedio de puntos anotados y recibidos por cada equipo en cada una de las temporadas relevantes?
3. Muestre el top 5 de árbitros en cuyos juegos pitados el equipo visitante pierde.
4. ¿Qué equipo maneja los salarios más altos actualmente?
5. ¿Cuál fue la temporada con más partidos en la historia de la NBA? ¿Cuál fue la temporada que más se ha prolongado en fechas?
6. ¿Cuál es el equipo que más diferencia de puntos a favor obtuvo en promedio por partido para la temporada 2017? ¿Y para la temporada 2018?
7. ¿Quién es el jugador más valioso del draft del 2018 hoy en día<sup>1</sup>?
8. Calcule el top 5 de los estados que más salarios pagaron durante las temporadas 2020/2021 y 2021/2022.

### Etapas 3

A continuación debe plantear sus propias preguntas que le permitan justificar la decisión que tomará acerca de en qué equipo invertirá para la temporada 2022/2023. Todas sus conclusiones deben estar basadas en el resultado de consultas SQL. Por ejemplo:

- Podría plantearse invertir en el equipo que sea más consistente en la cantidad de partidos que gana por temporada
- Podría plantearse invertir en el equipo que haya mejorado en las últimas cinco temporadas
- Podría plantearse invertir en el equipo que los jugadores con mejores promedios de puntos (PTS), asistencias (AST), rebotes (REB) y apariciones en *ALL STAR games*<sup>2</sup>.
- Etc.

### Requerimientos mínimos a completar:

- Se debe presentar el resultado de al menos 15 queries en todo el proyecto
- Deben presentarse al menos tres queries *diferentes* con agrupaciones (GROUP BY)
- Deben presentarse al menos tres queries *diferentes* con JOINS entre dos o más tablas
- Debe presentarse al menos una consulta que haga uso de subqueries

### Especificación de tecnología:

- Sistema gestor de base de datos: PostgreSQL
- Interfaz de interacción con base de datos: a discreción

### Temas a reforzar:

- Lenguaje SQL: DDL / DML
- PostgreSQL
- Consultas SQL hacia lógica de negocio

### Documentos a entregar:

- Archivo comprimido con:

---

<sup>1</sup> Puede investigar acerca del “draft” de la NBA aquí: [https://es.wikipedia.org/wiki/Draft\\_de\\_la\\_NBA](https://es.wikipedia.org/wiki/Draft_de_la_NBA)

<sup>2</sup> Puede investigar sobre otras estadísticas de Basketball aquí: [https://en.wikipedia.org/wiki/Basketball\\_statistics](https://en.wikipedia.org/wiki/Basketball_statistics)



- 
- Diagrama Entidad/Relación de la base de datos construida
  - Script desarrollado para procesar archivos CSV y alimentar base de datos
  - Documento PDF con las preguntas, queries y resultados obtenidos; que incluya la respuesta a la pregunta de negocio planteada y su justificación

**Evaluación:**

- Diseño y construcción de base de datos: 10 puntos
- Diseño y construcción de script para procesar archivos CSV y alimentar base de datos: 10 puntos
- Diseño de queries iniciales e interacción con base de datos: 25 puntos
- Diseño de preguntas propias y queries para responderlas: 40 puntos
- Análisis de resultados y presentación de solución a pregunta de negocio: 15 puntos (la presentación se realizará el día martes 23 de agosto durante la clase)

**Total: 100 puntos**

**Puntos extras:**

- Creatividad para presentación de análisis
- Creatividad para presentación de resultados de queries de forma gráfica