

## 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: lunes 06 de marzo a las 11:55 a.m.
- 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 futbol de la liga europea (entre las temporadas de 2,008 a 2,016 es decir 8 años), así como información de los jugadores y sus características.

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 según este modelo, ¿a qué equipo le apostaría usted? (debe de dar fundamentos basados en los datos)**

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

### Etapas

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 deberá:

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

## **Etapla 2**

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

Para esto deberá obtener lo siguiente:

### **Según estadísticas:**

1. La cantidad de juegos jugados en cada temporada por cada equipo, de cada liga (tome en cuenta que cada equipo puede jugar como visitante o como anfitrión.
2. ¿Quién es el mejor equipo de todas las ligas y de todas las temporadas según las estadísticas?

*Hint:* Obtenga la cantidad de goles a favor, goles en contra y la diferencia entre las dos anteriores, esto por cada temporada y por cada equipo de cada liga.

Utilizando este mismo query, obtenga el ranking de los equipos por temporada y por liga, ordenados por ese ranking de manera descendente por diferencia (utilice la función Rank () over patition), para obtener el equipo ganador (la respuesta es obvia).

### **Según apuestas:**

3. Realice un promedio de las probabilidades de todas las casas de apuesta por temporada, liga y equipo (elimine aquellos equipos que no tienen estadísticas en ninguna casa de apuesta (casas de apuesta como B36, IW, LB, PSH, etc).
4. ¿Quién es el mejor equipo de todas las ligas y de todas las temporadas según las apuestas? Observe la información dada en la siguientes páginas para poder interpretar las cuotas promedio para obtener las probabilidades:
  - <https://www.apuestas-deportivas.es/calculadora-probabilidades-apuestas-deportivas/>
  - <https://www.pasionamarilla.com/ud-laspalmas-noticias/como-se-calculan-las-probabilidades-en-las-apuestas-deportivas/>

*Hint:* Obtenga el promedio de todas las casas de apuesta por partido, y luego obtenga el promedio de estas por temporada, liga y equipo.

### Otros:

5. ¿Quiénes son los jugadores de cada liga y cada temporada que tienen los mejores atributos – características de juego? ¿De acuerdo a este inciso y comparándolo con el inciso 2 y 4 anteriores, alguno de los jugadores más valiosos se encuentra dentro del mejor equipo?
6. ¿Cuál es el equipo cuyos jugadores han incurrido en menos faltas?
7. ¿Cuáles son las características/atributos de los equipos que han sido los líderes de sus ligas en las distintas temporadas? ¿Sus comportamientos son similares?
8. ¿Quiénes son los 3 países líderes según apuestas y según estadísticas?

### Etapas 3

A continuación, debe plantear sus propias preguntas que le permitan justificar la decisión que tomará acerca de en qué equipo invertirá. Todas sus conclusiones deben estar basadas en el resultado de consultas SQL. Por ejemplo (sugerencias):

- Podría plantearse apostar en el equipo que sea más consistente en la cantidad de partidos que gana por temporada
- Podría plantearse apostar en el equipo que haya mejorado en las últimas tres temporadas
- Podría plantearse invertir en el equipo tienen características de ser más que son más creativos y que tienen una alta presión sobre la defensa, o a lo mejor los que tienen un juego rápido y un juego más agresivo.
- 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:
  - a. Diagrama Entidad/Relación de la base de datos construida
  - b. Script desarrollado para procesar archivos CSV y alimentar base de datos
  - c. Documento PDF con las preguntas, queries y resultados obtenidos; que incluya la respuesta a la pregunta de negocio planteada y su justificación

### **Evaluación:**

1. Diseño y construcción de base de datos: 10 puntos
2. Diseño y construcción de script para procesar archivos CSV y alimentar base de datos: 10 puntos
3. Diseño de queries iniciales e interacción con base de datos: 25 puntos
4. Diseño de preguntas propias y queries para responderlas: 40 puntos
5. Análisis de resultados y presentación de solución a pregunta de negocio: 15 puntos

**Total: 100 puntos**

### **Puntos extras:**

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