

Plan de Trabajo del Proyecto Final

Bases de datos no relacionales - 2025

1. Identificación del grupo

- **Número del grupo:** 18
- **Integrantes (nombre completo y número de cédula):** Maximiliano Martín Lucero Bacardat - 5.248.214-3

2. Familia de proyecto elegida

Seleccionar una de las siguientes familias:

- **Familia 1:** Migración entre modelos
- **Familia 2:** Evaluación de performance
- **Familia 3:** Exploración en profundidad de un motor
- **Familia 4:** Coast to coast (de la obtención al análisis, posible solución poliglota)
- **Familia 5:** Modelado cruzado o variantes
- **Familia 6:** Relevamiento y comparación de bases analíticas

3. Objetivos del proyecto

El presente proyecto tiene como propósito principal estudiar los factores que influyen en la elección del **Jugador Más Valioso** (MVP) de la NBA, a partir de datos históricos de desempeño individual y colectivo. Se busca, además, aplicar un enfoque predictivo sobre la base de los patrones observados en temporadas anteriores.

Para alcanzar este objetivo general, se plantean los siguientes objetivos específicos:

- Analizar las estadísticas históricas de los jugadores que han ganado el premio MVP.
- Evaluar el desempeño de estos jugadores y sus equipos en sus respectivas temporadas MVP.
- Comparar estos desempeños con los de otros jugadores destacados que no ganaron el premio.
- Modelar las relaciones entre jugadores, estadísticas, equipos y premios mediante una base de datos de grafos.
- (Opcional) Utilizar esta base para estimar candidatos al MVP de la temporada actual, en base a variables históricas relevantes.

4. Descripción general del proyecto

Se propone diseñar una solución que recorra todas las etapas del ciclo de vida de los datos —recolección, análisis, modelado, transformación, persistencia y explotación— para estudiar el premio MVP en la NBA.

El proyecto consistirá en construir un modelo basado en grafos que represente las relaciones entre jugadores, temporadas, equipos, métricas estadísticas y premios. A través de esta representación, se buscará entender qué características comunes tienen los jugadores que han ganado el premio MVP, y qué factores tienden a influir con mayor peso en su elección.

Entre las variables a considerar se encuentran estadísticas individuales (puntos, asistencias, rebotes, eficiencia, etc.), desempeño colectivo (porcentaje de victorias del equipo), logros destacados y la obtención de premios.

La hipótesis es que ciertos indicadores estadísticos, combinados con el rendimiento del equipo, permiten identificar patrones en la elección del MVP. En caso de que los resultados lo permitan, se considerará la posibilidad de entrenar un modelo de predicción complementario que estime la probabilidad de que un jugador actual obtenga el premio.

5. Actividades previstas y cronograma tentativo

Actividad	Breve descripción	Semana(s) prevista(s)	Duración (hs)
Recolección de datos	Extracción de datos históricos de jugadores, estadísticas, equipos y premios.	Semana 1	6
Limpieza de datos	Unificación de formatos, eliminación de inconsistencias, imputación si necesario.	Semana 1	6
Diagramar el grafo	Identificación de nodos, relaciones y propiedades clave a modelar.	Semana 2-3	10
Construcción del grafo	Construcción e implementación del modelo en Neo4j.	Semana 3-4	10
Consultas e insights	Desarrollo de consultas para responder preguntas sobre los MVP.	Semana 4-5	10
(Opcional) Modelo predictivo	(Si se considera viable) Entrenamiento y evaluación de un modelo de clasificación.	Semana 5-6	8
Informe final	Redacción del informe final y elaboración de visualizaciones.	Semana 6	10

6. Base teórica o artículos relacionados

¿El proyecto se inspira en algún artículo académico o proyecto existente?

- Sí — Indicar referencias completas:

- ...

- No

7. Datos y entornos

- **Juegos de datos a utilizar** (a definir su disponibilidad):

Los datos se obtendrán de fuentes públicas y gratuitas, entre ellas:

- *NBA Stats API*: Proporciona estadísticas completas de jugadores, equipos, partidos, temporadas, galardones, rankings, etc.
 - [Basketball Reference](#): Sitio con estadísticas históricas extensas (desde 1946) y premios individuales (como MVPs, All-NBA, etc.).
 - [balldontlie.io](#): API gratuita con acceso libre, cubre estadísticas de partidos y jugadores desde la temporada 1979–1980.
 - [Kaggle Datasets](#): Plataforma con múltiples datasets relevantes, incluyendo historiales de partidos, estadísticas avanzadas, y datos anotados sobre premios MVP.
 - [FiveThirtyEight GitHub](#): Incluye datasets de predicción y rendimiento.
- **Entornos de prueba y herramientas:**
 - Python (recolección y análisis de datos)
 - Neo4j (modelado y consultas de grafos)
 - Google Colab y Jupyter Notebooks (desarrollo)
 - (Opcional) Scikit-learn o XGBoost si se implementa una etapa predictiva

8. Otras observaciones (opcional)

El premio MVP, si bien se basa en criterios de rendimiento, también puede estar influenciado por aspectos subjetivos como la narrativa o el impacto mediático, los cuales no se reflejan completamente en los datos numéricos. El modelo propuesto se enfocará en identificar patrones objetivos a partir de datos históricos que permitan entender qué variables tienen mayor correlación con la obtención del premio.