



PROYECTO FINAL

Bootcamp Data Analytics

*Estrategia basada en datos
para la adquisición de derechos
televisivos de partidos de
Básquet Profesional*



Adquisición de derechos televisivos de partidos de Básquet profesional

Autores: Fernando Colazo – Data Engineer

Laura Constanzo – Data Analyst/Team Lead

Lucas Gebhardt – Data Analyst/Data Viz

Luis Velázquez – Data Base Administrator

Mauro Díaz Ruffino – Data Engineer

Cohorte: DA-FT01

Fecha de entrega: 31/05/2024

Institución: 3LFM Consulting

1.- Introducción

Este proyecto se enfoca en el análisis de datos deportivos de la NBA para ayudar a una cadena televisiva a decidir de qué partidos de básquet profesional comprar los derechos para transmitir y obtener los mayores beneficios económicos a través de altos niveles de audiencia y publicidad.

2.- Industria y Contexto

El proyecto apunta al análisis de datos deportivos de la NBA, incluyendo datos históricos desde 1946 hasta 2023, para que una cadena televisiva pueda decidir qué partidos comprar para su transmisión, maximizando la audiencia y los ingresos publicitarios.

3.- Descripción del Problema

Somos la consultora 3LFM Consulten y una cadena televisiva nos indica que quiere competir contra ESPN, TNT y NBA TV por la televisación de partidos. Necesita obtener insights valiosos para decidir qué partidos adquirir para transmitir y así obtener la mayor rentabilidad. La audiencia de los partidos se ve influenciada por la presencia de jugadores populares y equipos con altos desempeños, tanto en la NBA como en la Liga Universitaria.

4.-Objetivos

4.1.- Objetivo Principal

Ayudar a la cadena televisiva a decidir qué partidos adquirir para maximizar la rentabilidad respondiendo a preguntas clave sobre equipos universitarios y de la NBA, jugadores destacados, y estadísticas relevantes.

4.2.- Objetivos Específicos

- Crear una Base de Datos que contenga toda la información relevante sobre partidos, equipos y jugadores de la NBA y Liga Universitaria.
- Desarrollar analíticas claves a través del uso de herramientas de BI.
- Construir un proyecto que sea escalable en el tiempo, pudiendo automatizar la carga de datos a futuro y permitiendo sumar nuevas Features al desarrollo.

4.3.- Objetivos del Equipo

- Afianzar los conocimientos aprendidas durante el Bootcamp de Henry.
- Fortalecer habilidades blandas y de comunicación. Trabajar de forma conjunta y colaborativa simulando un entorno laboral.

5.- Resultados Esperados e Impacto

Se espera que los resultados ayuden al cliente a tomar decisiones basadas en datos que representen un beneficio económico sostenido. El análisis puede actualizarse con cada temporada de la NBA, anticipándose a futuras campañas de televisación.

6.- Fuentes de Datos

6.1.- Dataset Principal

- *NBA Database*: Disponible en Kaggle, incluye estadísticas de la NBA desde 1946 hasta 2023 en formato .csv con 2.31 GB de datos, 16 tablas y 341 columnas.

El dataset incluye datos de más de 4,800 jugadores, 65,000 partidos, estadísticas individuales y de equipo, historia de los drafts, capacidades de los estadios, y audiencias.

6.2.- Datasets Auxiliares

- *NBA Player Salaries (2022-23 Season)*: Disponible en Kaggle, incluye información sobre salaries y estadísticas de juegos de todos los jugadores de la NBA durante la última temporada de la NBA (22/23).

- *NBA: Instagram followings 2024, by team*: Dataset creado por Statista Research Department, nos muestra información sobre seguidores en redes sociales de jugadores y equipos de la NBA.

7.- Plan de Análisis

El trabajo se ejecuta en el lapso de dos semanas

7.1- Requerimientos del Caso

- Recopilar datos relevantes sobre equipos universitarios y de la NBA, historial de drafts, salaries de jugadores, resultados de partidos y estadísticas.

- Realizar análisis profundos utilizando técnicas avanzadas.

- Presentar resultados de manera comprensible mediante visualizaciones claras y atractivas.

7.2- Estrategia de Análisis

- Recopilación de Datos: Obtención y limpieza de datos.

- Análisis Descriptivo: Análisis de la distribución de selecciones en el draft, salaries, asistencia a partidos, etc.

- Análisis Comparativo: Comparación de rendimiento y estadísticas entre diferentes grupos.

- Presentación de Resultados: Creación de informes y visualizaciones interactivas.

8.- Fases y Tareas del Proyecto

8.1.- Recopilación y Procesamiento de Datos:

Obtención y limpieza de datos.

8.2.- Análisis Exploratorio de Datos:

Identificación de patrones y tendencias.

8.3.- Desarrollo de Métricas y Estadísticas Personalizadas:

Creación de métricas específicas.

8.4.- Análisis Comparativo:

Comparación entre diferentes grupos.

8.5.- Presentación de Resultados:

Informes detallados y visualizaciones.

9.- Herramientas y Tecnologías

- Python
- SQL Server
- Power BI
- Pandas
- Numpy
- GitHub
- Kaggle
- Canvas

10.- Pipeline de Datos a Implementar

10.1.- Ingesta de Datos

Fuente de Datos

Descarga y almacenamiento del dataset desde Kaggle.

Configuración de scripts para la actualización periódica de los datos si el dataset se actualiza con frecuencia.

- Tecnologías:

Python: Pandas para la manipulación de datos.

GitHub: Para el almacenamiento y versionado del código.

10.2.- Almacenamiento de Datos

Base de Datos

Configuración de una base de datos en SQL Server para almacenar los datos de manera estructurada.

Esquema de la Base de Datos:

Tablas:

- [Proveniencia]: Contiene información sobre la institución de donde vienen los jugadores.
- [Player]: Contiene información básica y biográfica de los jugadores de la NBA.
- [Draft]: Información del historial de los jugadores seleccionados en el draft de la NBA.
- [Teams]: Información sobre los equipos de la NBA.
- [Location]: Información sobre las localizaciones (ciudades y estados).
- [State]: Tabla con los estados y su respectivo identificador.
- [City]: Tabla con las ciudades y su respectivo identificador.
- [Arena]: Información sobre los estadios ("arenas") donde se juegan los partidos.
- [Games]: Información sobre los partidos jugados.
- [Game_Stats]: Estadísticas detalladas de cada partido.
- [Season]: Información de la temporada.
- [Team_historty]: Información histórica de los equipos.
- ETL (Extract, Transform, Load)

Extracción: Importación de datos desde archivos CSV de Kaggle.

Transformación: Limpieza de datos, tratamiento de valores nulos, normalización y creación de nuevas columnas calculadas.

Carga: Inserción de los datos limpios y transformados en las tablas correspondientes de la base de datos.

10.3.- Procesamiento de Datos

Limpieza y Transformación

Uso de Pandas para realizar operaciones de limpieza y transformación.

Tratamiento de valores atípicos y datos faltantes.

Automatización: Implementación de scripts de ETL para la actualización automatizada de datos.

Programación de tareas para la ejecución periódica de pipelines.

10.4.- Análisis Exploratorio de Datos (EDA)

Exploración Inicial:

- Análisis descriptivo utilizando Python (Pandas, Matplotlib, Seaborn).
- Visualización de estadísticas clave y distribución de datos.

Visualizaciones:

- Creación de gráficos de dispersión, histogramas, gráficos de barras y mapas de calor para identificar patrones y tendencias.

10.5.- Visualización y Reporte

Dashboards Interactivos

Creación de dashboards en Power BI para visualizar los resultados del análisis.

Implementación de segmentadores, gráficos interactivos y reportes dinámicos.

Documentación y Presentación

Documentación detallada del proceso de análisis y resultados obtenidos.

Presentación de hallazgos clave utilizando técnicas de storytelling y visualización atractiva.

10.6.- Despliegue y Mantenimiento

Publicación de Informes

Uso de Power BI Service para compartir dashboards con stakeholders.

Configuración de alertas y actualizaciones automáticas de datos.

Mantenimiento

Monitoreo del rendimiento del pipeline y ajustes según sea necesario.

Documentación y actualización continua del código y scripts utilizados.

10.7- Herramientas y Tecnologías

- Python: con la implementación de librerías como Pandas, Numpy, Matplotlib, Seaborn, Watchdog, SQL Alchemy.
- SQL Server: Para almacenamiento y consultas eficientes.
- Power BI: Para visualización y creación de dashboards.
- Watchdog: Para la automatización de tareas de ETL.

Con este pipeline proporcionamos un flujo estructurado y sistemático para manejar, procesar y analizar el dataset, asegurando que los datos sean precisos, actualizados y presentados de manera efectiva para la toma de decisiones.

11.- Creación de Dashboard

11.1.- Fuente de Datos

Conexión del Power BI a la fuente de datos: base de datos en SQL.

11.2.- Tecnologías

Power BI: Para crear el tablero interactivo.

SQL Server: Fuente de datos.

11.3.- ETL (Extract, Transform, Load)

Extracción: Importación de datos desde SQL.

Transformación: Limpieza de datos, tratamiento de valores nulos, normalización y creación de nuevas columnas calculadas, como así también de tablas como la de calendario.

Carga: Inserción de los datos limpios y transformados en las tablas correspondientes.

11.4.- Análisis Exploratorio de Datos (EDA)

Exploración Inicial

Análisis descriptivo utilizando Python (Pandas, Matplotlib, Seaborn).

Visualización de estadísticas clave y distribución de datos.

11.5.- Visualizaciones del EDA

Creación de gráficos de dispersión, histogramas, gráficos de barras y mapas de calor para identificar patrones y tendencias.

11.6.- Dashboards Interactivos

Creación de dashboards en Power BI para visualizar los resultados del análisis.

Implementación de segmentadores, gráficos interactivos y reportes dinámicos.

11.7. Documentación y Presentación

Documentación detallada del proceso de análisis y resultados obtenidos.

Presentación de hallazgos clave utilizando técnicas de storytelling y visualización atractiva.

Con este texto explicamos el proceso completo para la creación del dashboard interactivo y todos sus pasos previos necesarios. Para poder llegar así a las conclusiones que se comunicaran ante nuestros stakeholders.

12.- Conclusiones

12.1.- Recomendaciones Estratégicas

Luego de un análisis exhaustivo de nuestra base de datos y una comprensión profunda del contexto del negocio de nuestro cliente, definimos un índice que relaciona tres KPIs:

Desempeño de Equipo como indicador del éxito y la competitividad de un equipo durante las temporadas. Se calcula como: $(\text{Victorias} - \text{Derrotas}) + (\text{Puntos anotados} - \text{Puntos permitidos o en contra}) / \text{Juegos jugados}$. Considera tanto el éxito en términos de victorias y derrotas como la eficacia relativa del equipo tanto en ataque como en defensa. Esta combinación ofrece una visión general del rendimiento general del equipo durante la temporada, es decir los valores altos muestran a los equipos que mejor les va.

Redes Sociales como un indicador de popularidad del equipo a través de la cantidad de seguidores. Mide el interés y el atractivo del equipo para los aficionados. Está normalizado basado en el número de seguidores en una escala de 0 a 5.

Finanzas del equipo como un indicador que evalúa la capacidad para atraer y mantener jugadores estrella en función de los salarios que puede pagar, indicando también la presencia de jugadores de alto valor. Es la suma total de los salarios de los jugadores del equipo normalizada a una escala de 0 a 5.

Este índice nos otorga un ranking de equipo, como puede verse en el dashboard. A partir de este resultado y con el conocimiento adquirido en el desarrollo de este proyecto, *pudimos establecer que nuestra recomendación estratégica sea que la cadena televisiva compre los derechos para transmitir al aire los partidos en donde esté involucrado el equipo Boston Celtics. De esta forma podría obtener los mayores beneficios económicos por los altos niveles de audiencia y publicidad.*

Boston Celtics es un equipo que se encuentra en el TOP 3 del ranking obtenido a partir de nuestro índice de popularidad. Además, como vimos en la hoja de audiencia, es uno de los equipos con mayor convocatoria cuando juega de local. Esto es una muestra clara de que tiene muchos fanáticos que siguen sus campañas por temporada.

Por otro lado, posee un alto porcentaje de victoria cuando juega de visitante, lo que demuestra que sus partidos son interesantes de ver a través de la tv cuando sus aficionados y el público en general no puede concurrir al estadio. También observamos que no todos los partidos son televisados y de esta forma podría haber un mercado para que nuestra cadena televisiva pueda introducirse.

Finalmente, analizando los partidos de la liga universitaria en donde su transmisión permitiría obtener la mayor rentabilidad para nuestro cliente, recomendamos que se intente adquirir los derechos televisivos del equipo de la Universidad de Duke, ya que esta institución viene aportando muchos jugadores a la NBA, tanto en la última temporada y en el histórico estando en el top 5. Esto indica que sus partidos son atractivos por tener la presencia de jugadores de alto nivel que luego desplegarán sus habilidades en la NBA.

Para dar con estas conclusiones se evaluaron los puntos que se detallas a continuación

12.2.-Ranking de Equipos

Se creó un ranking de los equipos en base a un índice compuesto por varios KPIs: desempeño de los equipos, cantidad de seguidores en redes sociales y salarios del plantel.

Este índice refleja no solo el rendimiento deportivo, sino también la popularidad y la estabilidad financiera de los equipos.

12.3.- Recomendaciones de Equipos

- Golden State Warriors: Destacan por su alto rendimiento reciente, gran cantidad de seguidores en redes sociales y buenos salarios, lo que indica una situación financiera sólida.
- Los Angeles Lakers: Con un historial de éxito y jugadores populares, tienen una base de seguidores robusta y convocan grandes audiencias.
- Brooklyn Nets: A pesar de su variabilidad en rendimiento, su reciente reestructuración y adquisición de jugadores estrella han incrementado su popularidad.
- Milwaukee Bucks: Su rendimiento consistente y la presencia de Giannis Antetokounmpo los hace atractivos para las audiencias.
- Boston Celtics: Un equipo con historia y buena presencia en redes sociales, además de un desempeño sólido.

12.4.- Asistencia a Partidos

Se puede observar que equipos con altos índices de popularidad y buenos desempeños tienden a tener una mayor asistencia a los partidos.

Equipos como los Golden State Warriors y Los Angeles Lakers consistentemente llenan sus estadios, reflejando su capacidad para atraer grandes audiencias en transmisiones televisivas.

12.5- Convocatoria de Equipos

Equipos con jugadores estrella tienen una mayor capacidad de convocatoria, atrayendo no solo a fanáticos locales sino también a una audiencia global.

La presencia de jugadores reconocidos internacionalmente incrementa la visibilidad y atractivo de los partidos.

12.6- Métricas Clave

- Desempeño de equipos: Incluye estadísticas de victorias, títulos y rendimiento en playoffs.

- Popularidad en redes sociales: Medida a través de seguidores, interacciones y tendencias.
- Salarios del plantel: Indicador de la inversión y capacidad financiera del equipo.

13.- Impacto del Proyecto

Este proyecto proporciona una base sólida para la toma de decisiones estratégicas, permitiendo a la cadena televisiva posicionarse mejor en el mercado competitivo.

La implementación de este análisis basado en datos concretos y actualizados asegura que la cadena pueda competir eficazmente con grandes nombres como ESPN, TNT y NBA TV. Además, el enfoque escalable y automatizable del proyecto permitirá que la cadena se mantenga actualizada y relevante en futuras temporadas de la NBA.