



Universidad Católica Boliviana “San Pablo”

Facultad de Ingeniería

Departamento de Ingeniería de Sistemas

La Paz - Bolivia

PROYECTO DE FASE 1 PRESENTADO AL DEPARTAMENTO DE
INGENIERÍA DE SISTEMAS PARA LA OBTENCIÓN DEL GRADO
ACADÉMICO DE MAGISTER EN CIENCIA DE DATOS

“ANÁLISIS DE RENDIMIENTO Y DETECCIÓN DE
PATRONES ORIENTADO A LA ANALÍTICA DE DATOS
EN EL FÚTBOL CON ALGORITMOS DE REGRESIÓN”

Realizado por:

José Renán Estenssoro Mercado

AÑO 2023

1.	<i>Introducción.....</i>	<i>5</i>
2.	<i>Marco referencial.....</i>	<i>5</i>
2.1	<i>Introducción al tema de investigación</i>	<i>5</i>
2.2	<i>Antecedentes</i>	<i>6</i>
2.3	<i>Descripción del objeto de estudio</i>	<i>7</i>
2.4	<i>Identificación y formulación del problema de investigación</i>	<i>7</i>
2.4.1	<i>Planteamiento problema</i>	<i>7</i>
2.5	<i>Objetivos</i>	<i>8</i>
2.5.1	<i>General.....</i>	<i>8</i>
2.5.2	<i>Específicos</i>	<i>8</i>
2.6	<i>Límites y alcances</i>	<i>8</i>
2.6.1	<i>Limites</i>	<i>8</i>
2.6.2	<i>Alcances.....</i>	<i>9</i>
2.7	<i>Justificación</i>	<i>9</i>
2.7.1	<i>Justificación teórica.....</i>	<i>9</i>
2.7.2	<i>Justificación practica.....</i>	<i>9</i>
2.8	<i>Modelos y herramientas a utilizar</i>	<i>9</i>
2.8.1	<i>Modelos.....</i>	<i>9</i>
2.8.2	<i>Herramientas.....</i>	<i>9</i>
3	<i>Marco teórico</i>	<i>10</i>
3.1	<i>Futbol</i>	<i>10</i>
3.2	<i>FIFA</i>	<i>10</i>
3.3	<i>Premier league.....</i>	<i>11</i>
3.3.1	<i>Los Clubes miembros de la premier league</i>	<i>11</i>
3.4	<i>Análisis de datos</i>	<i>12</i>
3.5	<i>Análisis de datos en el futbol.....</i>	<i>12</i>
3.6	<i>Análisis exploratorio de datos</i>	<i>13</i>
3.7	<i>Regresión.....</i>	<i>13</i>
4	<i>Marco Practico.....</i>	<i>14</i>
4.1	<i>Notebook y las librerías a utilizar</i>	<i>14</i>

4.2	<i>Obtención de los datos</i>	15
4.3	<i>Entendimiento de los datos</i>	16
4.4	<i>Preprocesamiento de datos</i>	17
4.5	<i>Análisis de los datos</i>	21
4.5.1	<i>Análisis de rendimiento</i>	21
4.6	<i>Regresión de los datos</i>	28
4.6.1	<i>Regresión Lineal</i>	28
4.6.2	<i>Suport Vector Machine</i>	31
5	<i>Conclusiones y Recomendaciones</i>	33
5.1	<i>Conclusiones</i>	34
5.2	<i>Recomendaciones</i>	35
6	<i>Bibliografía</i>	35

Ilustración 1	Librerías	15
Ilustración 2	importar dataset.....	16
Ilustración 3	Datos importados.....	16
Ilustración 4	Columnas dataset.....	17
Ilustración 5	filas duplicadas	17
Ilustración 6	valores nulos (parte1)	Ilustración 7 valore nulos (parte 2)..... 18
Ilustración 8	reemplazar valores nulos.....	
Ilustración 9	convertir a int	18
Ilustración 10	tabla 1	19
Ilustración 11	agregar empates.....	19
Ilustración 12	Agregar puntos	19
Ilustración 13	tabla 2	20
Ilustración 14	set index	20
Ilustración 15	separar dataset.....	20
Ilustración 16	ordenar dataset.....	21
Ilustración 17	tabla de campeones	21
Ilustración 18	Top4.....	22
Ilustración 19	victorias empates y derrotas	22
Ilustración 20	goles y goles concedidos.....	23

Ilustración 21 goles vs puntos	23
Ilustración 22 goles vs puntos man city	24
Ilustración 23 matriz correlación 1	25
Ilustración 24 Intentos de gol	26
Ilustración 25 intentos de gol con dirección	26
Ilustración 26 goles recibidos por temporada	27
Ilustración 27 porterías en 0	28
Ilustración 28 regresión lineal	29
Ilustración 29 valores reales vs parecidos	30
Ilustración 30 matriz de confusión.....	Error! Bookmark not defined.
Ilustración 31 reporte F1	Error! Bookmark not defined.

1. Introducción

El fútbol es la actividad más popular y rentable del mundo. Millones de personas siguen diariamente los partidos de las ligas nacionales y campeonatos internacionales e invierten no solo su tiempo en los mismos, sino también sus recursos al comprar entradas para los estadios, pagar los servicios de cable de televisión y comprar souvenirs de su equipo favorito.

El análisis de datos se ha convertido en una herramienta cada vez más importante en el mundo del fútbol, ya que los equipos y las organizaciones buscan obtener una ventaja competitiva mediante su uso para tomar decisiones informadas. El análisis de datos en el fútbol puede ayudar a rastrear y analizar el rendimiento de los jugadores, las estadísticas del juego y las tácticas de los equipos, y además puede proporcionar información valiosa sobre el desarrollo del juego. Este proyecto explorará el uso del análisis de datos en el fútbol y evaluará la eficacia de esta técnica para predecir y mejorar el rendimiento de los equipos y de sus jugadores. Al recopilar y analizar datos sobre una variedad de métricas, como son las estadísticas del desempeño de los jugadores y los resultados de los equipos, se ofrecerá información valiosa y estratégica sobre los posibles beneficios y desventajas del uso de análisis de datos en el mundo del fútbol.

2. Marco referencial

2.1 Introducción al tema de investigación

El uso de datos en muchos campos permite tomar decisiones informadas y predicciones basadas en evidencia. Los datos proporcionan una fuente de información fiable y objetiva que se puede utilizar para respaldar argumentos, validar hipótesis, realizar proyecciones y evaluar la eficacia de diferentes enfoques. Además, los datos se pueden usar para identificar tendencias, patrones y relaciones que pueden no ser evidentes de inmediato, lo que brinda información valiosa que puede ayudar a comprender sistemas complejos y hacer predicciones más precisas.

Adicionalmente, el manejo y el uso de los datos en deportistas profesionales puede ser de gran ayuda para mejorar su rendimiento, evitar posibles lesiones y sacar cierta ventaja deportiva al extraer y analizar datos sobre los pases que estos llegan a completar, los disparos acertados y fallidos, los centros que realizaron, las distancias que recorren por

partido, sus regates, etc. Con esta información los técnicos y sus asistentes pueden comprender mejor las fortalezas y debilidades de sus jugadores para, luego del análisis, optimizar sus potencialidades y, al mismo tiempo, tener una visión más fidedigna del equipo rival.

2.2 Antecedentes

En el año 2003 se publicó el libro “Moneyball: el arte de ganar en un juego injusto” escrito por Michael Lewis. Muchos especialistas en el mundo del deporte dicen que este libro es la razón por la que se comenzó a utilizar el análisis de datos en el deporte (Lewis, 2004).

El libro trata de Billy Beane, gerente general de los “Oakland Athletics”, un equipo de beisbol con base en los Estados Unidos. En el libro Lewis menciona como Beane comenzó a utilizar los datos y la teoría del juego para identificar a jugadores subvalorados con altas tasas de éxito. Beane construye su equipo aplicando la teoría del juego y obtiene buenos resultados. Además lo logra armar su equipo con mucho menos presupuesto que otros. Cuando los gerentes y entrenadores se dan cuenta del éxito que alcanza, se oponen al uso de estos métodos en el deporte (Lewis, 2004).

En el mundo existen muchas empresas que se dedican a la extracción y análisis de los datos de los partidos de futbol. Por ejemplo, “Opta Sports” es una empresa inglesa que genera datos para 30 deportes diferentes en 70 países. “Opta Sports” extrae los datos de los partidos mediante varios softwares y equipos de personas que se dedican a analizar los partidos fecha a fecha. Una vez recolectados, los datos se ponen a disposición de sus socios para que realicen sus respectivas inversiones y actividades (opta, 2023).

La utilización de los datos de “Opta Sports” puede variar según las necesidades de cada socio. Muchos de sus clientes son equipos de futbol profesional que utilizan sus datos para el reclutamiento de jugadores y detección de patrones de juego en los rivales, entre otros tipos de análisis (opta, 2023).

“Opta Sports” también ofrece sus servicios a casas de apuestas que utilizan sus servicios para realizar predicciones en los resultados de los partidos (opta, 2023).

Otro tipo de cliente son las empresas televisivas que utilizan sus datos para informar sobre las estadísticas de los partidos, antecedentes de los mismos, desempeño de los jugadores,

y posibles resultados, entre otros datos de interés, lo que enriquece las transmisiones y hace que los televidentes se adentren más en el juego (opta, 2023).

2.3 Descripción del objeto de estudio

El Análisis de datos en el fútbol puede ayudar mucho a los equipos profesionales a mejorar sus resultados en los partidos de la liga. Existe una gran cantidad de datos disponibles sobre el rendimiento de los equipos y sus jugadores en los partidos. Analizar los datos puede proporcionar información valiosa sobre cómo mejorar el rendimiento.

Por ejemplo, el análisis de datos puede ayudar a los entrenadores a identificar patrones en el rendimiento de su equipo como del equipo rival. Y con esto puede tomar decisiones antes, durante y después del juego. De igual manera puede ayudar a estructurar los entrenamientos según la información recolectada de cada jugador.

2.4 Identificación y formulación del problema de investigación

2.4.1 Planteamiento problema

El análisis de datos en el deporte ha ganado importancia debido a la inversión creciente de los equipos en la gestión de datos para mejorar sus resultados. Los datos se utilizan para identificar jugadores talentosos, analizar tácticas rivales y generar modelos predictivos para tomar decisiones sobre el rendimiento de los jugadores.

Los equipos deportivos también aprovechan los datos para optimizar el rendimiento de los jugadores, identificar áreas de mejora, planificar la preparación física, determinar cambios oportunos durante un partido, organizar el equipo y aplicar tácticas específicas. Además, el análisis de datos ayuda a identificar patrones y tendencias que pueden mejorar el rendimiento general del equipo.

Sin embargo, a pesar del reconocimiento de la importancia del análisis de datos en el deporte, existen desafíos que deben superarse para un aprovechamiento efectivo. Estos desafíos son:

- ⇒ La interpretación de los datos: La correcta interpretación de los datos y la extracción de conocimientos relevantes para la toma de decisiones estratégicas es primordial. Es necesario contar con personal capacitado en el análisis de datos y en su comprensión profunda. Esto quiere decir que deben realizarse capacitaciones

permanentes para que los operadores puedan aplicar los resultados obtenidos de manera práctica y efectiva.

- ⇒ Identificación de patrones y tendencias: El análisis de datos busca identificar patrones y tendencias significativas que puedan contribuir a mejorar el rendimiento del equipo. Sin embargo, es necesario contar con algoritmos y modelos de análisis apropiados, adaptados a las características específicas del deporte y que consideren los factores que influyen en el rendimiento de los jugadores.

2.5 Objetivos

2.5.1 General

Analizar y predecir el rendimiento de los equipos de futbol en la liga inglesa utilizando técnicas de análisis exploratorio de datos y algoritmos de regresión para identificar patrones de rendimiento y hacer predicciones.

2.5.2 Específicos

- Realizar el “análisis exploratorio de datos” para tener un mejor entendimiento de los datos.
- Identificar cuáles son los equipos que mejor rendimiento tienen en los últimos años para poder identificar cuáles son los patrones que los diferencian y potencian.
- Identificar que similitudes tienen los equipos con mejores desempeños para poder identificar los patrones de rendimiento.
- Utilizar algoritmos de regresión para intentar predecir los resultados de las siguientes temporadas.

2.6 Límites y alcances

2.6.1 Limites

- No habrá implementación del proyecto, solo se realizará un prototipo.
- No se diseñará ningún nuevo algoritmo, todos los algoritmos a implementar ya existen.
- Solo se realizará el análisis de los resultados regresión lineal
- Solo se realizará el análisis de los resultados de regresión logística

2.6.2 Alcances

- Se obtendrán los datos de fuentes oficiales
- Se realizará el análisis de una sola liga
- Las ligas y partidos a analizar serán a partir del año 2008.

2.7 Justificación

2.7.1 Justificación teórica

El desarrollo de los algoritmos y el análisis de los resultados de Ciencia de Datos en la Universidad Católica Boliviana. Se pondrá en práctica conocimientos en las áreas como la programación en Python, desarrollo de algoritmos, inteligencia artificial, análisis de datos, extracción de los datos entre otros conceptos importantes.

2.7.2 Justificación practica

La detección de los patrones y el análisis de rendimiento, permitirá que los equipos de futbol como los entrenadores puedan implementar nuevas tácticas en los entrenamientos para mejorar los resultados en el campo de juego. También permitirá que los equipos puedan reclutar más fácilmente jugadores con un amplio potencial e identificar posibles defectos en los armados de los equipos.

2.8 Modelos y herramientas a utilizar

2.8.1 Modelos

El desarrollo del proyecto será implementado con la metodología Scrum que es sumamente ágil y de fácil implementación

Scrum encaja perfectamente en las necesidades del proyecto al ser un marco de trabajo donde se aplican buenas prácticas. Con esta metodología se realizan entregas parciales de los trabajos finales lo que permite ser un marco ideal para proyectos donde los plazos son cortos.

2.8.2 Herramientas

Las herramientas a utilizar son las siguientes:

- Python

Para el desarrollo del proyecto se utilizará el lenguaje de programación Python con la versión 3.8.8 este lenguaje es el estándar para todo lo que es la ciencia de datos.

- Jupyter Notebook

Para la ejecución y visualización de los algoritmos en Python se utilizará la herramienta de Jupyter Notebook.

- Kaggle

Todos los dato se obtendrán de la página open Source de Kaggle

- MacOS Ventura

Para ejecutar programas como “Jupyter Notebook” entre otros se utilizará el sistema operativo de Apple MacOS Ventura

- Microsoft Office Word

Microsoft Office Word es la herramienta por excelencia para la redacción de trabajos.

3 Marco teórico

3.1 Futbol

El Futbol es un deporte en el que dos equipos rivales, de 11 jugadores cada uno, se enfrentan en un partido de 90 minutos de duración. Cada equipo dispone de 10 jugadores de campo y de un portero, el objetivo es tratar de que la pelota ingrese en el arco del equipo rival más veces de las que entra en la propia portería, respetando un reglamento ya establecido (FIFA, 2023).

3.2 FIFA

La FIFA es la máxima entidad rectora del fútbol a nivel internacional. Existe para gobernar el fútbol y reglamentar el juego y las competencias en todo el mundo. Desde 2016, la organización ha evolucionado rápidamente hasta convertirse en un organismo que puede servir de manera más efectiva a la evolución del juego en beneficio de todo el mundo (FIFA, 2023).

La FIFA está modernizando el fútbol para que sea global, accesible e inclusivo en todos los aspectos. No solo en uno o dos continentes, sino en todo el planeta (FIFA, 2023).

3.3 Premier league

La Premier League es el organismo organizador de la Liga de fútbol inglesa, y se responsabiliza de la competición, su Libro de Reglas y la transmisión centralizada y otros derechos comerciales. (Premier League, 2023)

Sin embargo, La Premier League no opera de forma aislada. Esta institución trabaja de manera proactiva y constructiva con los clubes miembros y otras autoridades del fútbol (Premier League, 2023).

3.3.1 Los Clubes miembros de la premier league

La Premier League es una empresa privada cuya propiedad pertenece a sus 20 clubes miembros. Cada club perteneciente a la Premier League, es independiente y trabaja dentro de las reglas del fútbol, según lo definido por la Premier League, la FA, la UEFA y la FIFA, además de estar sujeto a las leyes inglesa y europea (Premier League, 2023).

Cada uno de los 20 clubes es accionista de la Premier League. La consulta y el consenso es la base de su normativa y método de operación. Para la toma de cualquier decisión, los clubes deben solicitar una reunión con los demás clubes ya que las reuniones de accionistas son el último foro de toma de decisiones y se llevan a cabo a intervalos regulares durante el transcurso de la temporada (Premier League, 2023).

La Asamblea de Socios de la Premier League tiene lugar al final de cada temporada. Es el momento en el que los clubes relegados transfieren sus acciones a los clubes promovidos a la Premier League desde el Football League Championship (Premier League, 2023).

Los clubes tienen la oportunidad de proponer nuevas reglas o enmiendas en la junta de accionistas. Cada Club Miembro tiene derecho a un voto y todos los cambios en las reglas y los principales contratos comerciales requieren el apoyo de al menos dos tercios de los votos, o 14 clubes, para ser acordados (Premier League, 2023).

Cualquier incumplimiento grave del Libro de Reglas se resuelve en un tribunal independiente integrado por tres personas que se reúne para escuchar el caso, determinar las responsabilidades y establecer, si fuese el caso, una sanción, que puede ir desde

multas hasta deducciones de puntos y, en casos extremos, la expulsión de la competencia (esto nunca ha sucedido en la historia de la Premier League) (Premier League, 2023).

En la Junta General de la Premier League en junio de 2022, los clubes acordaron una Carta de Propietarios. Los Propietarios y Directores de los Clubes Miembros son custodios de esas organizaciones y mantendrán el espíritu de los compromisos de la Carta (Premier League, 2023).

3.4 Análisis de datos

La analítica de datos es convertir los datos en información para que las empresas, negocios u otros puedan adquirir información práctica de sus negocios. La analítica de datos incluye una variedad de procesos, herramientas y tecnologías para poder encontrar ciertas tendencias que los datos por sí solos no nos muestran y también nos pueden ayudar a resolver los problemas. (Amazon AWS, 2022)

El análisis de datos ayuda a las empresas a obtener una mayor visibilidad y un conocimiento más profundo de sus procesos y servicios. Les proporciona información detallada sobre la experiencia del cliente y sus problemas. Al cambiar el paradigma más allá de los datos para conectar los conocimientos con la acción, las empresas pueden crear experiencias personalizadas para los clientes y productos digitales relacionados, optimizar las operaciones y aumentar la productividad de los empleados. (Amazon AWS, 2022)

3.5 Análisis de datos en el fútbol

El análisis de datos ayuda a formar un equipo con cualidades específicas. En otras palabras, los datos disponibles permiten buscar jugadores que cumplan con un perfil específico, incluso antes de mandar a los cazatalentos a observar nuevos jugadores (Altim Analytics, 2020).

En cada partido y entrenamiento se generan una gran cantidad de datos que permite a los jugadores, clubes, entrenadores obtener información que ayuda para la creación de nuevas estrategias que ayudan a evitar los errores en los partidos (Altim Analytics, 2020).

Todos los movimientos que hacen los jugadores pueden generar datos y estos se los recoger, transformar y analizar, para que los directivos, entrenadores y jugadores puedan basar sus decisiones en información contrastada y correcta. (Altim Analytics, 2020)

Hace un par de años atrás los analistas deportivos recolectaban la información de los partidos de manera manual, esto quiere decir que tenían que ver las repeticiones de los partidos y anotar los datos en plantillas de Excel. (altim analytics, 2020)

En la actualidad, se continúan grabando los partidos para extraer los datos que son analizados mediante un software especializado. La meta es detectar patrones, capturar, almacenar y organizar la información no estructurada y de esta manera obtener los datos sobre los pases, disparos, atajadas, etc.. También se pueden obtener datos individuales de los jugadores, como los partidos en los que juega como titular, la distancia que recorre, su velocidad, los minutos que juega, los goles que anota, etc (altim analytics, 2020). El uso de un software especializado y de los de algoritmos permite obtener la información limpia para realizar informes claros y precisos (altim analytics, 2020).

3.6 Análisis exploratorio de datos

El análisis exploratorio de datos o EDA, por sus siglas en inglés, es utilizado por científicos de datos y analistas para analizar e investigar grandes cantidades de datos y resumir sus principales características, utilizando métodos de visualización de datos como las gráficas. Realizar el EDA permite manipular y modificar los datos a gusto para obtener las respuestas que el analista necesita. También permite descubrir patrones, detectar anomalías o probar hipótesis.

El EDA ayuda a identificar mucho mejor las variable de los datos y las relaciones entre ellas, también ayuda a determinar y elegir las técnicas estadísticas a utilizar para realizar el análisis estadístico.

3.7 Regresión

La dependencia estadística se refiere a la relación entre variables en la que conocer los valores de la variable independiente no permite determinar con exactitud el valor de la variable dependiente, pero sí su comportamiento global. Este análisis admite dos enfoques: la teoría de correlación para medir el grado de dependencia entre variables y la regresión para determinar la estructura de dependencia. La finalidad de la regresión es asignar valores a la variable Y en función de los valores de la variable X. En el caso bidimensional, la regresión de Y sobre X y la regresión de X sobre Y explican cada variable para cada valor de la otra, pero no necesariamente coinciden (Universidad de Valencia(CEACES), 2020).

La dependencia estadística se refiere a la relación entre variables en la que conocer los valores de una variable independiente no permite determinar con exactitud el valor de la variable dependiente, aunque sí se puede llegar a establecer un comportamiento general. El análisis de la dependencia estadística admite dos enfoques: la correlación, que estudia el grado de dependencia entre variables, y la regresión, que determina la estructura de dependencia para poder asignar valores a la variable dependiente en función de los valores de las variables independientes. En el caso bidimensional, se pueden establecer dos tipos de regresión, una de Y sobre X y otra de X sobre Y, cuyas funciones no tienen por qué coincidir. En conjunto, el análisis de la dependencia estadística resulta útil en campos como la investigación de mercados, la biología, la psicología y la sociología, entre otros. (Universidad de Valencia(CEACES), 2020)

4 Marco Practico

4.1 Notebook y las librerías a utilizar

En el proyecto se utilizó el lenguaje de programación Python con “Jupyter Notebook” como interprete. Python es uno de los lenguajes de programación más usado para el análisis de datos y la inteligencia artificial. Para usar el Notebook de Jupyter no es necesario estar conectado a internet.

La librerías utilizadas se las puede observar en la Ilustración 1 . Donde se muestran todas la librerías utilizadas en el proyecto. Las librerías en el proyectos son pandas, utilizada para manejar Dataframes en Python, Matplotlib, utilizada para realizar gráficas, Numpy, utilizada para realizar operaciones matemáticas y finalmente Sklearn es la librería de inteligencia artificial por defecto de Python.

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
import numpy as np
from scipy.stats import norm
from sklearn.preprocessing import StandardScaler
from sklearn.feature_selection import VarianceThreshold
from sklearn import preprocessing
from scipy import stats
from pandas import Series, DataFrame
from pandas.plotting import autocorrelation_plot
#from pandas.tools.plotting import scatter_matrix
from pylab import rcParams
from matplotlib import collections as collections
from matplotlib.patches import Rectangle
import seaborn as sns
from itertools import cycle

import warnings
warnings.filterwarnings('ignore')
%matplotlib inline

rcParams['figure.figsize'] = 5,4
sb.set_style('whitegrid')
from numpy import median
from numpy import mean

```

Ilustración 1 Librerías

4.2 Obtención de los datos

Existen muchas maneras de obtener los datos para este tipo de proyectos, pero para agilizar la realización de este proyecto los datos fueron obtenidos de la página Kaggle.com, esta página de software libre contiene un gran cantidad y variedad de datos que contiene y su fácil entendimiento.

Al realizar la búsqueda de datos se encontró un dataset con los resultados y las estadísticas de los equipos de la premier league inglesa en el siguiente enlace <https://www.kaggle.com/datasets/zaeemnalla/premier-league?select=stats.csv>. Estos datos son desde la temporada 2006/2007 hasta la temporada 2017/2018. Lo que significa que son 11 años de datos con los resultados de 4560 partidos de futbol una cantidad suficiente para ser analizada y trabajada.

Estos datos encontrados estaban en dos archivos .csv con los nombres “results.csv” y “stats.csv”. para el presente proyecto se utilizará únicamente el archivo stats.csv.

4.3 Entendimiento de los datos

Una vez obtenidos los datos se los almacena en la misma carpeta donde se creo el Notebook. Después se tiene que importar con el método “read_csv” de Pandas, una vez que los datos están importados se muestra un resumen de los datos con el método head() como se puede apreciar en la Ilustración 2 importar dataset.

```
#importar dataset
stats = pd.read_csv('resultados/stats.csv')
stats.head()
```

Ilustración 2 importar dataset

El resultado obtenido se lo muestra en la Ilustración 3 Datos importados donde se puede apreciar el nombre de los equipos y varias columnas, como el número de victorias por equipo los goles, tarjetas amarillas etc.

team	wins	losses	goals	total_yel_card	total_red_card	total_scoring_att	ontarget_scoring_att	hit_woodwork	att_hd_goal
Manchester United	28.0	5.0	83.0	60.0	1.0	698.0	256.0	21.0	12.0
Chelsea	24.0	3.0	64.0	62.0	4.0	636.0	216.0	14.0	16.0
Liverpool	20.0	10.0	57.0	44.0	0.0	668.0	214.0	15.0	8.0
Arsenal	19.0	8.0	63.0	59.0	3.0	638.0	226.0	19.0	10.0
Tottenham Hotspur	17.0	12.0	57.0	48.0	3.0	520.0	184.0	6.0	5.0

Ilustración 3 Datos importados

El dataset tienen las siguientes columnas:

Columna	significado	Columna	significado	Columna	significado
teams	equipo analizado.	goal_fastbreak	goles al inciar un partido	total_pass	cantidad total de pases hechos
wins	victorias	total_offside	fuera de lugar	total_through_ball	pases en profundidad
losses	derrotas	clean_sheet	cantidad de partidos sin recibir goles	total_long_balls	pases largos
goals	goles	goals_conceded	goles recibidos	backward_pass	pases hacia atras
total_yel_card	tarjetas amarillas	saves	salvadas por el portero	total_cross	centros realizados
total_red_card	tarjetas rojas	outfielder_block	remates bloqueados	corner_taken	corner pateados
total_scoring_att	remates al arco	interception	pases interceptados	touches	toques al balon
ontarget_scoring_att	remates al arco con direccion a porteria	total_tackle	cantidad de tacleadas totales	big_chance_missed	grandes chances creadas.
hit_woodwork.	remates a los postes	last_man_tackle	tacleadas de ultimo hombre	clearance_off_line	depejes en la linea
att_hd_goal	remates de cabeza	total_clearance	despejes	dispossessed	perdidas de balon
att_pen_goal	remates de penal	head_clearance	despejes de cabeza	penalty_save	penales salvados
att_freekick_goal	remates de tiro libre	own_goals	goles en propia puerta	total_high_claim	salidas del arquero
att_ibox_goal	remates dentro del area del portero	penalty_conceded	penales concedidos	punches	depeje con los puños del arquero
att_obox_goal	remates fuera del area del portero	pen_goals_conceded	goles de penales concedidos.	season	temporada jugada

Ilustración 4 Columnas dataset

4.4 Preprocesamiento de datos

En el punto 4.3 Entendimiento de los datos se pudo apreciar de manera visual que no existen columnas repetidas, pero debido a la gran cantidad de datos que el dataset tiene, no se puede apreciar de esta manera si existen filas duplicadas por lo que se utilizó el método “duplicated()” como se puede apreciar en la Ilustración 5 filas duplicadas.

```
duplicados = stats.duplicated().any()
duplicados
```

Ilustración 5 filas duplicadas

La respuesta al código previo es de un dato “booleano” de valor falso, lo que nos indica que en el data set no existen filas duplicadas. Después se requirió ver si existen valores nulos en el dataset para eso se utilizó la función “isna()” y se hizo la suma de los resultados

con la función “.sum()” el resultado se lo puede apreciar en la Ilustración 6 valores nulos (parte1) Ilustración 7 valore nulos (parte 2)

team	0	last_man_tackle	0
wins	0	total_clearance	0
losses	0	head_clearance	20
goals	0	own_goals	0
total_yel_card	0	penalty_conceded	0
total_red_card	0	pen_goals_conceded	0
total_scoring_att	0	total_pass	0
ontarget_scoring_att	0	total_through_ball	20
hit_woodwork	0	total_long_balls	0
att_hd_goal	0	backward_pass	80
att_pen_goal	0	total_cross	0
att_freekick_goal	0	corner_taken	0
att_ibox_goal	0	touches	0
att_obox_goal	0	big_chance_missed	80
goal_fastbreak	0	clearance_off_line	0
total_offside	0	dispossessed	20
clean_sheet	0	penalty_save	0
goals_conceded	0	total_high_claim	0
saves	20	punches	0
outfielder_block	0	season	0
interception	0	dtype: int64	
total_tackle	0		

Ilustración 6 valores nulos (parte1)

Ilustración 7 valore nulos (parte 2)

En las dos imágenes se puede observar que hay 6 columnas que tienen valores nulos y cuantos valores nulos tienen en total cada columna. Para realizar el tratado de estos valores se decidió darles 0 como valor por defecto. Para realizar esto se utilizó el método “.replace()” de pandas como se puede observar en la Ilustración 8 reemplazar valores nulos.

```
stats.replace([np.inf, -np.inf, np.nan], 0, inplace=True)
```

Ilustración 8 reemplazar valores nulos

Finalmente, como se pudo ver en la Ilustración 3 Datos importados los datos en el dataset son datos tipo “float”. Para este trabajo se necesita que estén en formato tipo “int”. Para eso se seleccionan todos los valores numéricos y se los convierte en tipo “int” con el función “astype()” y a la función le pasamos el valor a convertir.

```
numeric_cols = stats.select_dtypes(include=[np.number]).columns
stats[numeric_cols] = stats[numeric_cols].astype(int)
stats.head()
```

Ilustración 9 convertir a int

Y este es la Ilustración 10 tabla 1 se puede apreciar el resultado final.

	team	wins	losses	goals	total_yel_card	total_red_card	total_scoring_att	ontarget_scoring_att	hit_woodwork	att_hd_1
0	Manchester United	28	5	83	60	1	698	256	21	
1	Chelsea	24	3	64	62	4	636	216	14	
2	Liverpool	20	10	57	44	0	668	214	15	
3	Arsenal	19	8	63	59	3	638	226	19	
4	Tottenham Hotspur	17	12	57	48	3	520	184	6	

Ilustración 10 tabla 1

Pero en la Ilustración 10 tabla 1 se puede apreciar que faltan dos datos importantes que no pueden faltar en nuestro análisis de datos y estos son los empates obtenidos y los puntos ganados en cada temporada. Estos valores se los puede obtener de una manera muy rápida y simple. cada temporada de la liga inglesa tiene 38 jornadas en total por los que si se le resta las victorias y las derrotas a cada equipo se puede calcular el número de empates.

```
stats["draws"] = 38 - stats["wins"] - stats["losses"]
stats.insert(2, "draws", stats.pop("draws"))
stats.head()
```

Ilustración 11 agregar empates

Y para calcular los puntos simplemente se multiplica las victorias por 3 y se le suma 1 a los empates y el resultado total serán los puntos obtenidos.

```
stats["points"] = stats["wins"]*3 + stats["draws"]
stats.insert(4, "points", stats.pop("points"))
stats.head()
```

Ilustración 12 Agregar puntos

Después de agregar los empates y los puntos obtenidos, durante cada temporada, a la tabla se lo puede apreciar en la Ilustración 13 tabla 2.

	team	wins	draws	losses	goals	total_yel_card	total_red_card	total_scoring_att	ontarget_scoring_att	hit_woodwork
0	Manchester United	28	5	5	83	60	1	698	256	21
1	Chelsea	24	11	3	64	62	4	636	216	14
2	Liverpool	20	8	10	57	44	0	668	214	15
3	Arsenal	19	11	8	63	59	3	638	226	19
4	Tottenham Hotspur	17	9	12	57	48	3	520	184	6

Ilustración 13 tabla 2

se cambia el índice numérico de la tabla para que la columna “team” sea el nuevo index, esto se realiza con el método “set_index()”.

```
stats = stats.set_index('team', drop=True)
stats.head()
```

Ilustración 14 set índice

Una vez realizado todo el preprocesamiento de datos se obtiene una tabla con 43 columnas 240 filas. Pero antes de pasar al análisis de datos se tienen que partir el dataset por temporadas, debido a que partirlo el dataset por temporada permite realizar el análisis de los datos temporada a temporada y de esta manera los datos no se mezclan.

```
# Create a list of unique seasons in the dataset
seasons = stats['season'].unique()
print(seasons)

# Create a dictionary to store each season's data as a separate dataframe
season_data = {}

# Loop through each season, filter the original dataframe by that season,
#and store the resulting dataframe in the dictionary
for season in seasons:
    season_data[season] = stats[stats['season'] == season]
```

Ilustración 15 separar dataset

Una vez que se tiene el dataset separado, se debe ordenar los datos y el orden de los equipos que vienen desordenados. Los datos deben colocarse en un orden específico, el

equipo con mayor cantidad de puntos tiene que ir primero y en forma descendente hasta llegar al equipo con la menor cantidad de puntos.

```
for season in seasons:
    season_data[season]["position"] = season_data[season]['points'].rank(ascending=False, method='dense')
    season_data[season].insert(0, "position", season_data[season].pop("position"))
    season_data[season]["position"] = season_data[season]["position"].astype(int)
    season_data[season] = season_data[season].sort_values(by=['position'], ascending=True)
```

Ilustración 16 ordenar dataset

4.5 Análisis de los datos

4.5.1 Análisis de rendimiento

Uno de los objetivos más importantes temporada a temporada para los equipos grandes es salir campeón de la liga, pero son solo unos pocos equipos los que pueden salir campeones cada temporada. En la Ilustración 17 tabla de campeones se puede apreciar los campeones desde la temporada 2006/2007 hasta la temporada 2017/2018.

	wins	draws	losses	points	goals	goals_conceded	clean_sheet	season
Manchester United	28	5	5	89	83	27	16	2006-2007
Manchester United	27	6	5	87	80	22	21	2007-2008
Manchester United	28	6	4	90	68	24	24	2008-2009
Chelsea	27	5	6	86	103	32	18	2009-2010
Manchester United	23	11	4	80	78	37	15	2010-2011
Manchester City	28	5	5	89	93	29	17	2011-2012
Manchester United	28	5	5	89	86	43	13	2012-2013
Manchester City	27	5	6	86	102	37	16	2013-2014
Chelsea	26	9	3	87	73	32	17	2014-2015
Leicester City	23	12	3	81	68	36	15	2015-2016
Chelsea	30	3	5	93	85	33	16	2016-2017
Manchester City	32	4	2	100	106	27	18	2017-2018

Ilustración 17 tabla de campeones

En Inglaterra actualmente solo son 6 los equipos considerados grandes, de estos 6 equipos solo tres pudieron salir campeones en los años analizados, estos son el

Manchester United con 6 títulos, Manchester City y Chelsea cada uno con 3 títulos, existe un cuarto campeón que no es considerado un equipo, este es el Leicester City que sólo tiene un solo título.

	1	2	3	4
Arsenal	0	2	5	3
Chelsea	3	3	3	0
Leicester City	1	0	0	0
Liverpool	0	2	1	3
Manchester City	3	3	1	1
Manchester United	6	2	0	1
Tottenham Hotspur	0	1	3	1

Ilustración 18 Top4

En la Ilustración 19 victorias empates y derrotas se puede apreciar que existe un patrón en los equipos campeones, para que un equipo salga campeón de la liga inglesa tendría que obtener de 25 victorias para adelante y 5 o menos derrotas, de los 12 campeones analizados solo son dos equipos que lograron salir campeones con menos de 25 victorias en la temporada, pero a esto reducida cantidad de victorias se compensa con también una muy reducida cantidad de derrotas lo que ayudo a compensar esa falta de puntos.

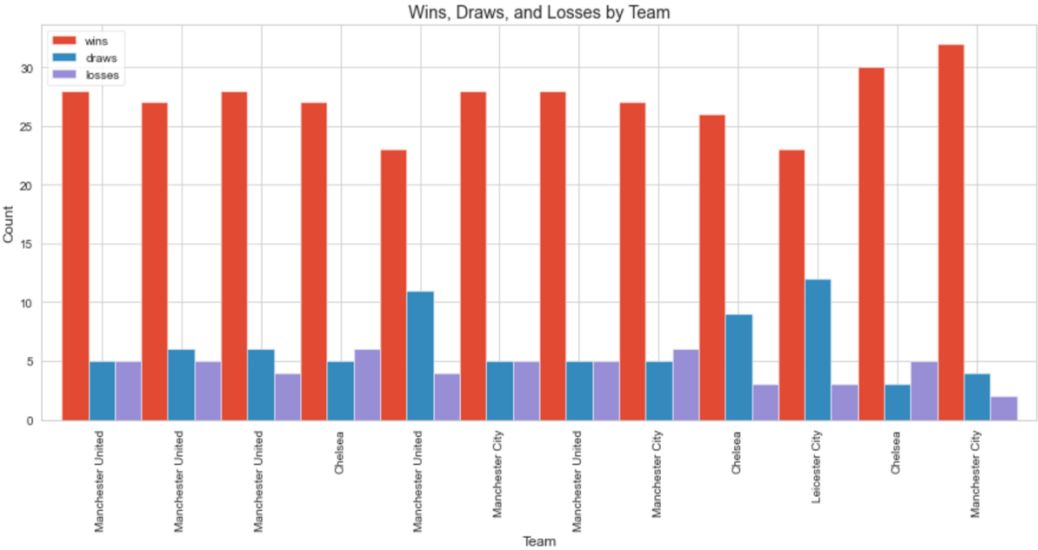


Ilustración 19 victorias empates y derrotas

En la Ilustración 20 goles y goles concedidos se analiza la cantidad de goles que los equipos marcan y conceden. Los equipos tienden a marcar un promedio aproximado de 80 goles temporada y recibir una cantidad menor de 40 goles.

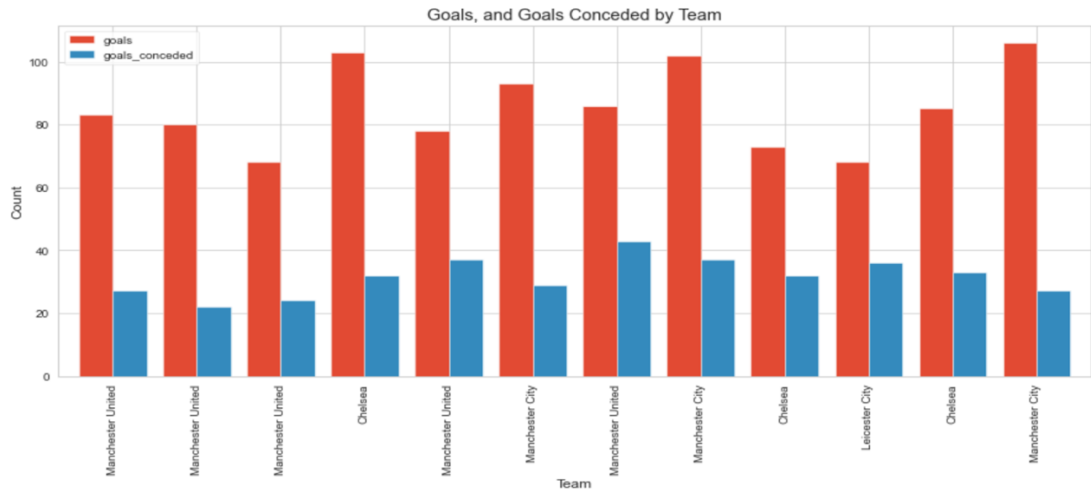


Ilustración 20 goles y goles concedidos

En la Ilustración 21 goles vs puntos se muestra la comparación de los equipos que tienen de la cantidad de goles que anotan comparado con los puntos obtenidos. Se puede apreciar que por lo general que mientras más goles se anoten más puntos se obtienen. En ciertos puntos se puede observar que existen una cierta linealidad en los puntos

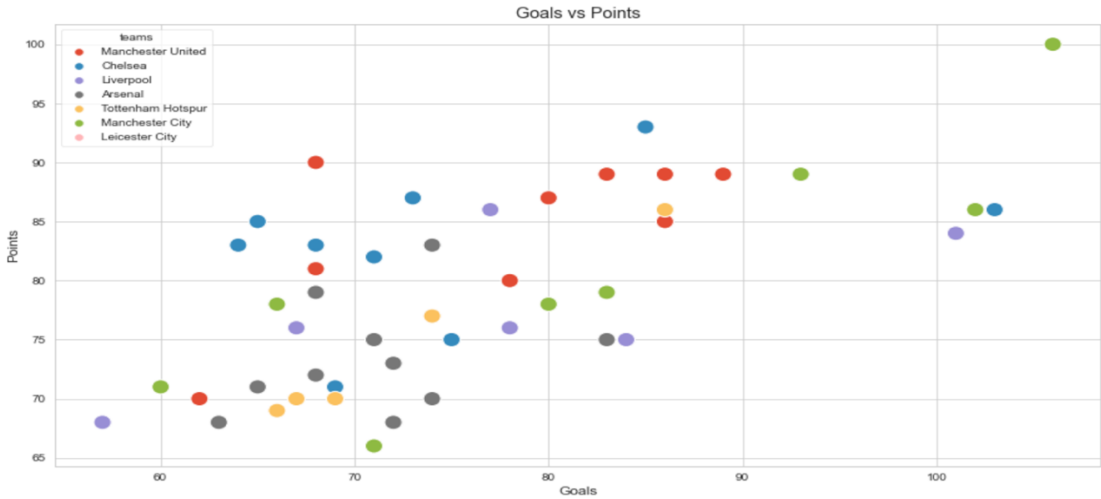


Ilustración 21 goles vs puntos

Para demostrar esta linealidad usaremos de ejemplo como le fue al equipo del Manchester City, para eso podemos ver la Ilustración 22 goles vs puntos man city, donde se puede apreciar que el equipo que mientras más goles anoto más puntos obtuvo.



Ilustración 22 goles vs puntos man city

La cantidad de victorias que un equipo tiene en una temporada es correlativa a la cantidad de goles que el equipo anota, mientras más goles anota un equipo más victorias puede obtener. En la Ilustración 23 matriz correlación 1 se puede apreciar que la mayor correlación existente es entre los goles anotados y las victorias con un valor de 0.55 lo que nos indica una fuerte correlación de los dos. Otra relación fuerte existente es la cantidad de goles concedidos con los empates y las derrotas teniendo una correlación de 0.26 y 0.15.

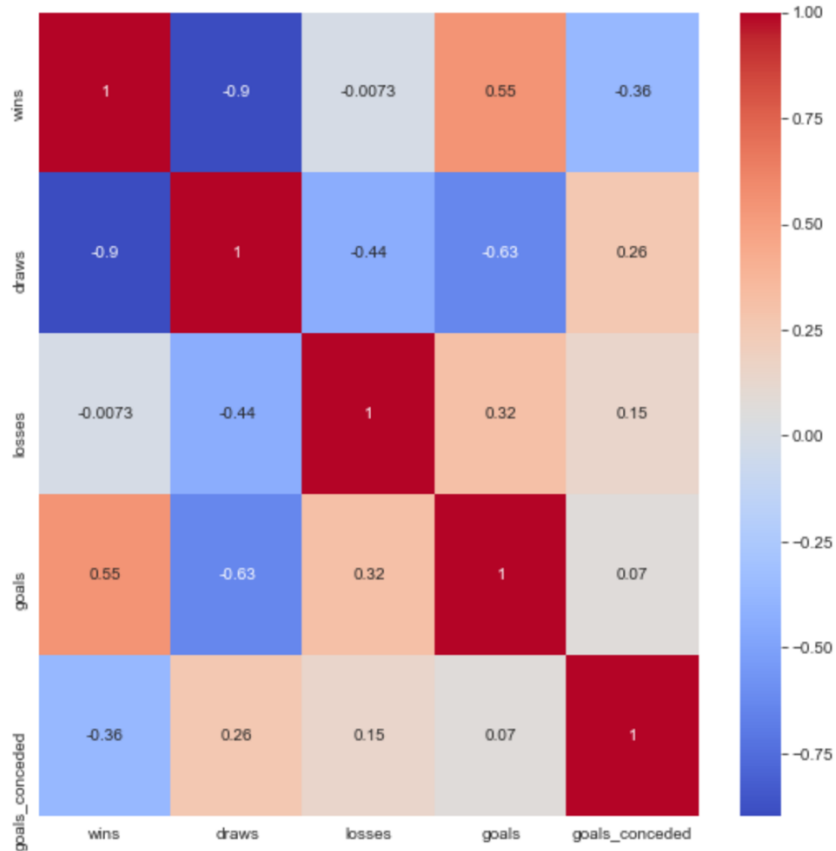


Ilustración 23 matriz correlación 1

Debido a que anotar goles es fundamental para la obtención de victorias, la creación de oportunidades de gol es muy importante. En la Ilustración 24 Intentos de gol se pueden apreciar dos graficas que muestran la cantidad de intentos de goles realizados por los equipos en cada temporada. También se puede apreciar la posición en la que salió y la cantidad de goles anotados esa temporada.

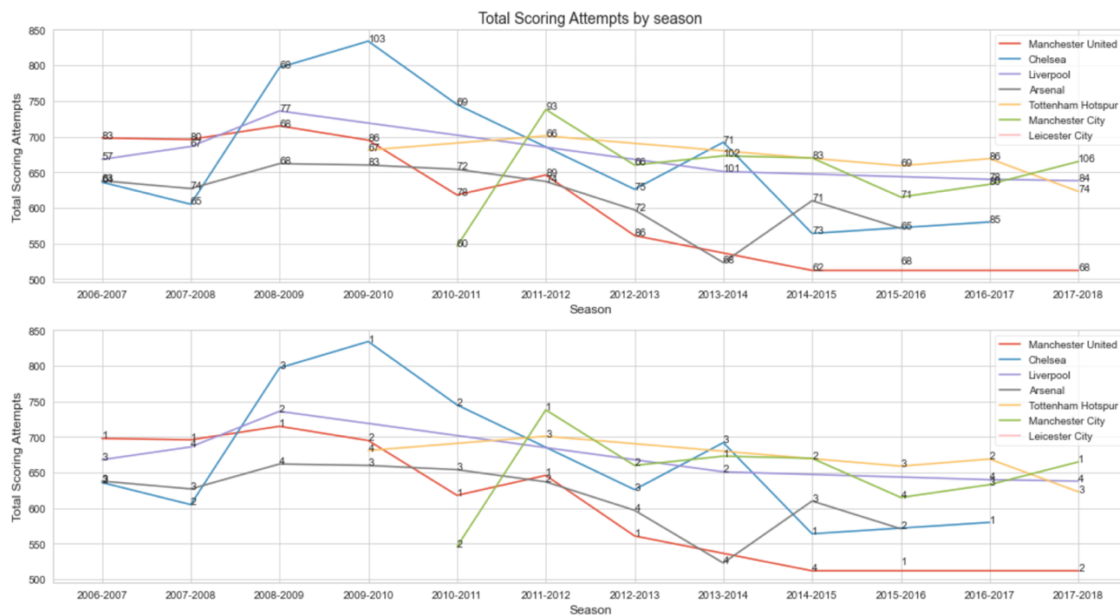


Ilustración 24 Intentos de gol

En la Ilustración 25 intentos de gol con dirección también se puede apreciar la cantidad de oportunidades de gol que los equipos obtuvieron pero en vez se muestra la cantidad de oportunidades reales que tuvieron.

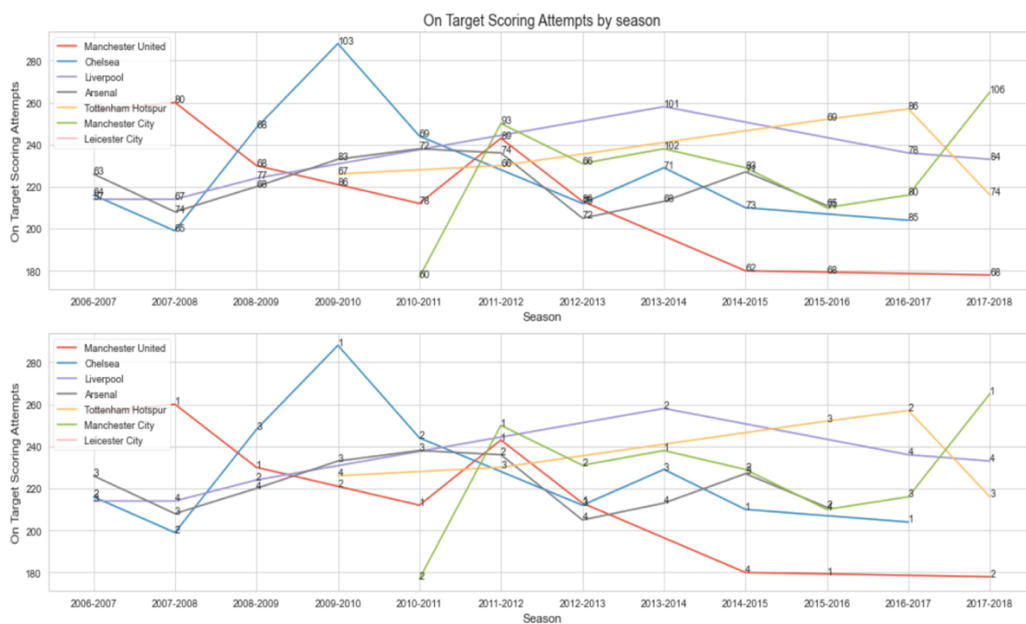


Ilustración 25 intentos de gol con dirección

Pero el solo convertir goles no es lo más importante en un equipo de alto rendimiento, también lo es el recibir la menor cantidad de goles posibles. En realidad, el objetivo es tratar de mantener el arco en 0. En la Ilustración 26 goles recibidos por temporada el equipo que reciba la menor cantidad de goles en una temporada no siempre va a ser el equipo con mayor cantidad de goles. Se puede apreciar que existe cierta correlación entre la menor cantidad de goles recibidos y las primeras posiciones en la tabla. Y también es importante señalar que si se recibe una menor cantidad de goles, se estará en las posiciones de arriba.

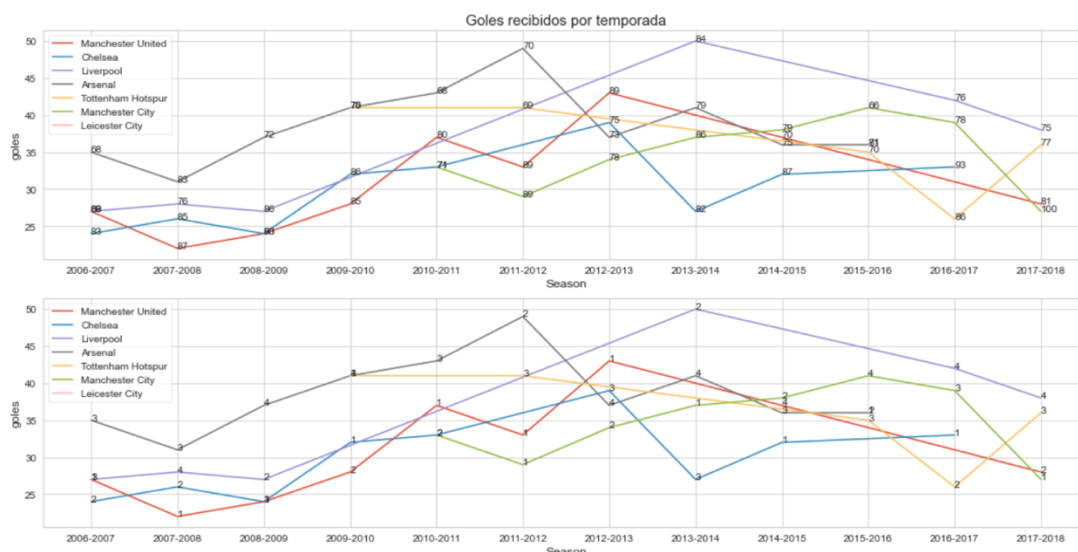


Ilustración 26 goles recibidos por temporada

En la Ilustración 27 porterías en 0 se tiene un gráfico que muestra la cantidad de veces que los equipos terminaron un partido sin recibir goles. La mayoría de los equipos que salen campeones tienden a salir de un partido sin recibir goles con un promedio de 16 veces por temporada. Si bien esto no es un excluyente, puede ampliar mucho las posibilidades de salir primero en la tabla de posiciones.

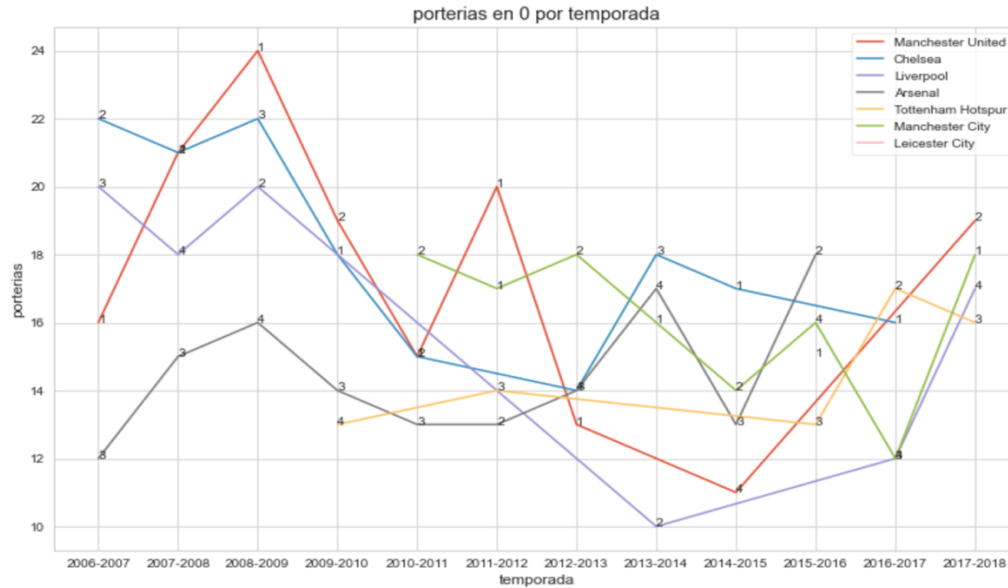


Ilustración 27 porterías en 0

4.6 Análisis de los datos

4.6.1 Regresión Lineal

Para realizar la regresión se estableció a los puntos como la variable dependiente, y como las variable independientes se estableció a las siguientes variables, 'goals', 'total_scoring_att', 'ontarget_scoring_att', 'hit_woodwork', 'att_hd_goal', 'att_pen_goal', 'att_freekick_goal', 'att_ibox_goal', 'att_obox_goal', 'clean_sheet', 'goals_conceded', 'saves', 'own_goals', 'penalty_conceded' y finalmente 'pen_goals_conceded'. Se decide utilizar estas columnas debido a su alta influencia en el la posibilidad de obtener victorias, como se pudo apreciar en los puntos previos. Se dividió el dataset en dos: En train y en test. El resultado final se lo puede apreciar en la **Error! Reference source not found..**

	Coeficientes	Intercepto	Error cuadrático medio (MSE)	Coeficiente de determinación (R2)
0	1.606628	40.232229	15.56388	0.911471
1	-0.011150	40.232229	15.56388	0.911471
2	0.032125	40.232229	15.56388	0.911471
3	-0.054686	40.232229	15.56388	0.911471
4	0.073371	40.232229	15.56388	0.911471
5	-0.227039	40.232229	15.56388	0.911471
6	0.136203	40.232229	15.56388	0.911471
7	-1.006872	40.232229	15.56388	0.911471
8	-0.797697	40.232229	15.56388	0.911471
9	0.504557	40.232229	15.56388	0.911471
10	-0.499423	40.232229	15.56388	0.911471
11	0.001149	40.232229	15.56388	0.911471
12	-0.038260	40.232229	15.56388	0.911471
13	0.183566	40.232229	15.56388	0.911471
14	-0.048574	40.232229	15.56388	0.911471

Ilustración 28 regresión lineal

Coeficientes: Los coeficientes representan el cambio esperado en la variable objetivo (puntos) por cada unidad de cambio en las características correspondientes. En este caso, cada coeficiente se refiere a una característica específica. Por ejemplo, el coeficiente asociado a la característica "goals" es 1.6066, lo que indica que se espera un aumento de aproximadamente 1.6066 puntos por cada gol anotado. De manera similar, los otros coeficientes representan el impacto de cada característica en los puntos.

Intercepto: El intercepto es el valor esperado de la variable objetivo cuando todas las características son igual a cero. En este caso, el intercepto es de 40.2322, lo que significa que si todas las características son cero, se esperaría que el equipo tenga alrededor de 40.2322 puntos.

Error cuadrático medio (MSE): El MSE es una medida del error promedio cuadrático entre las predicciones del modelo y los valores reales. En este caso, el MSE es de 15.5639, lo

que indica que, en promedio, las predicciones del modelo tienen un error de aproximadamente 15.5639 puntos con respecto a los valores reales.

Coefficiente de determinación (R^2): El coeficiente de determinación, también conocido como R^2 , es una medida de la proporción de la varianza total de la variable objetivo que puede ser explicada por el modelo. En este caso, el R^2 es de 0.9115, lo que significa que aproximadamente el 91.15% de la variabilidad en los puntos puede ser explicada por las características utilizadas en el modelo.

En resumen, el modelo de regresión lineal muestra que las características tienen diferentes impactos en la cantidad de puntos que un equipo puede obtener. Algunas características, como los goles anotados, tienen un impacto positivo significativo en los puntos, mientras que otras características, como los goles en contra, tienen un impacto negativo. El modelo tiene un buen ajuste, ya que el coeficiente de determinación (R^2) es alto (0.9115), lo que indica que las características seleccionadas explican una gran proporción de la variabilidad en los puntos. Sin embargo, el error cuadrático medio (MSE) muestra que las predicciones del modelo tienen un error promedio de alrededor de 15.5639 puntos.

Y finalmente se obtuvo un gráfico comparativo de los valores reales con los valores precedidos.



Ilustración 29 valores reales vs parecidos

4.6.1.1 Ejecución de los resultados

Con los datos que se muestran en la Ilustración 30 datos de prueba se realiza la ejecución y el análisis de rendimiento de los equipos. Para poder realizar este análisis los datos previos se los tiene que correr bajo el análisis de la regresión lineal realizada.

	goals	total_scoring_att	ontarget_scoring_att	hit_woodwork	att_hd_goal	att_pen_goal	att_freekick_goal	att_ibox_goal	att_obox_goal	clean_sheet
team										
Manchester United	83	698	256	21	12	5	1	72	11	16
Chelsea	64	636	216	14	16	3	6	41	23	22
Liverpool	57	668	214	15	8	6	1	46	11	20
Arsenal	63	638	226	19	10	10	3	53	10	12
Tottenham Hotspur	57	520	184	6	5	6	2	44	13	6
Bolton Wanderers	47	404	120	7	10	6	0	38	9	12
Reading	52	419	132	8	15	3	0	42	10	13
Blackburn Rovers	52	478	155	5	12	5	3	43	9	8
Everton	52	465	153	9	9	8	2	44	8	14
Portsmouth	45	525	186	9	13	2	0	32	13	12

Ilustración 30 datos de prueba

Una vez realizada la regresión se obtuvo los observados en la Ilustración 31 resultados reg lineal en estos datos se puede observar que el equipo con mayor puntuación es el Manchester United con 87 puntos seguido del Chelsea con 83 puntos y el último equipo siendo el Bolton Wander con 50 puntos.

```
[86.62771628 83.28104604 71.55245326 67.25007676 53.1930921 49.65940066  
57.07710899 50.90999446 60.92403573 56.11197629]
```

Ilustración 31 resultados reg lineal

4.6.2 Suport Vector Machine

Igualmente que en la sección anterior para aplicar el algoritmo SVM se estableció a los puntos como la variable dependiente, y como las variables independientes se estableció a las siguientes variables, 'goals', 'total_scoring_att', 'ontarget_scoring_att', 'hit_woodwork', 'att_hd_goal', 'att_pen_goal', 'att_freekick_goal', 'att_ibox_goal', 'att_obox_goal', 'clean_sheet', 'goals_conceded', 'saves', 'own_goals', 'penalty_conceded' y finalmente 'pen_goals_conceded'.

En la Ilustración 32 reporte SVM se ve parte de el informe de clasificación, informa sobre los coeficientes, los interceptos, el error y otros datos.

```
{'Coeficientes': array([[ -1.16812683e-03, -6.67501043e-04,  1.00125156e-03, ...,
    2.16937839e-03, -1.80225282e-02,  1.66875261e-04],
    [-6.20732464e-04, -7.58673012e-04, -2.06910821e-04, ...,
    1.37940548e-03, -1.15870060e-02,  0.00000000e+00],
    [-3.41938793e-03, -1.65840315e-02, -6.15489827e-03, ...,
    1.19678578e-03, -1.70969396e-04, -1.70969396e-04],
    ...,
    [-1.39389964e-03,  1.10692030e-02,  2.13184651e-03, ...,
    -7.37946868e-04, -5.65759265e-03,  8.19940964e-05],
    [-7.14621533e-03,  9.40291490e-03, -6.58204043e-03, ...,
    -5.64174894e-04, -1.16596145e-02,  5.64174894e-04],
    [-3.51493849e-03, -1.42271320e-02, -1.02100594e-02, ...,
    1.00426814e-03,  1.17164616e-03,  3.34756047e-04]]),
  'Intercepto': array([ 1.12599082,  1.30043451,  7.9748675 , ..., -7.20900295,
    -3.2899859 , 11.78784836]),
  'Error cuadrático medio (MSE)': 11.791666666666671,
  'Coeficiente de determinación (R2)': 0.7000686562862882}
```

Ilustración 32 reporte SVM

En el contexto de una SVM (Support Vector Machine) de clasificación multiclase, los resultados presentados son los coeficientes de los hiperplanos y el término de intercepción para cada una de las clases, así como el error cuadrático medio (MSE) y el coeficiente de determinación (R2) del modelo.

Coeficientes: Los coeficientes representan los pesos asignados a cada una de las características (variables independientes) del modelo para cada clase en el problema de clasificación multiclase. Los coeficientes indican cómo cada característica contribuye a la decisión de clasificación para cada clase. Cada fila en el arreglo de coeficientes corresponde a una clase, y cada columna corresponde a una característica.

Intercepto: El término de intercepción (también conocido como sesgo o bias) es una constante que se suma al resultado de la combinación lineal de las características y los

coeficientes para cada clase. Ayuda a ajustar el hiperplano y desplazarlo en el espacio para que pueda separar correctamente las clases.

Error cuadrático medio (MSE): El error cuadrático medio es una métrica que representa el promedio de los errores al cuadrado entre las predicciones del modelo y los valores reales (etiquetas de clase). Es una medida de la calidad del ajuste del modelo a los datos. Un MSE más bajo indica un mejor ajuste del modelo a los datos de entrenamiento.

Coeficiente de determinación (R2): El coeficiente de determinación R2 es una métrica que indica la proporción de la varianza total de la variable dependiente (en este caso, las etiquetas de clase) que es explicada por el modelo. Un valor de R2 cercano a 1 indica que el modelo explica bien la variabilidad de los datos, mientras que un valor cercano a 0 indica que el modelo no explica bien la variabilidad y puede estar subajustando.

Es importante tener en cuenta que estos resultados pertenecen a una SVM específica con el conjunto de características y datos proporcionados. La interpretación precisa de los resultados puede depender del contexto y la naturaleza del problema específico que estés abordando. Además, es fundamental evaluar el rendimiento del modelo utilizando otras métricas, como la precisión, el recall, la matriz de confusión, etc., para obtener una comprensión completa de la efectividad del modelo en la clasificación multiclase.

4.6.2.1 Ejecución de resultados

Para realizar la ejecución de los resultados se utilizaron los mismos datos del mismo punto de la sección de 4.6.1 Regresión Lineal, en la Ilustración 33 resultados svm se puede observar los resultados que cada equipo tienen. Como se puede apreciar el equipo con mayor puntuación obtenida sería el Manchester United con 89 puntos seguido del Chelsea con 83. En un último lugar tenemos al Blackburn Rovers con 52 puntos.

[89 , 83 , 68 , 68 , 60 , 56 , 56 , 52 , 58 , 58]

Ilustración 33 resultados svm

5 Conclusiones y Recomendaciones

5.1 Conclusiones

- ⇒ **Análisis exploratorio de datos:** El análisis exploratorio de datos nos permitió obtener un mejor entendimiento de los datos disponibles. A través de técnicas como visualizaciones, resúmenes estadísticos y la identificación de patrones, pudimos obtener información relevante sobre las variables y su distribución en los datos. Esto nos proporcionó una base sólida para realizar análisis más profundos y tomar decisiones informadas.
- ⇒ **Identificación de equipos con mejor rendimiento:** Mediante el análisis de los datos históricos, identificamos los equipos que han tenido el mejor rendimiento en los últimos años. Al estudiar los patrones y tendencias presentes en estos equipos, pudimos identificar factores comunes que podrían contribuir a su éxito. Estos factores podrían incluir la calidad de los jugadores, las tácticas implementadas, el estilo de juego o la gestión del equipo.
- ⇒ **Identificación de similitudes entre equipos de alto rendimiento:** Al comparar los equipos con mejor desempeño, encontramos similitudes significativas que podrían explicar su éxito. Estas similitudes pueden incluir la presencia de líderes estratégicos en el equipo, una sólida planificación de la preparación física, una cultura de trabajo en equipo o una mentalidad ganadora. Identificar estas similitudes nos brinda la oportunidad de aprender de los equipos exitosos y aplicar esas lecciones a otros equipos para mejorar su rendimiento.
- ⇒ **Utilización de los algoritmos de regresión y svm para la predicción:** Empleamos algoritmos de regresión y svm para predecir los resultados de las siguientes temporadas. Estos modelos se basaron en datos históricos y variables relevantes para el rendimiento del equipo. Si bien las predicciones pueden contener cierto margen de error, nos brindan una estimación razonable de cómo podrían desempeñarse los equipos en el futuro y permiten tomar decisiones más fundamentadas en la planificación y estrategia.

En conjunto, el análisis exploratorio de datos, la identificación de equipos con mejor rendimiento, el estudio de las similitudes entre ellos y la utilización de algoritmos de regresión y svm nos brindaron una visión más clara y precisa del rendimiento en el deporte. Estas conclusiones pueden ayudar a los equipos deportivos a optimizar su rendimiento, mejorar sus estrategias y tácticas, así como a tomar decisiones más acertadas en relación

con el desarrollo de los jugadores y la planificación de las temporadas futuras. El análisis de datos se consolida como una herramienta valiosa para mejorar los resultados deportivos y maximizar el potencial de los equipos así como el uso eficiente de los recursos.

5.2 Recomendaciones

De acuerdo a las observaciones y al análisis que se realizó durante la formulación de este proyecto, se puede realizar las siguientes recomendaciones:

1. Jupyter Notebook es una herramienta muy fácil e intuitiva de usar, pero al mismo tiempo el usuario depende mucho de las capacidades del equipo, se recomienda utilizar Google Colab para mayor agilidad en los procesos.
2. Se recomienda tener todas las librerías actualizadas
3. Para futuros proyectos se recomienda realizar una recolección de datos más metódica, con una mayor cantidad de precisión.
4. En una futura obtención de datos se recomienda recolectar los datos jugador por jugador en vez de una recolección de datos general.
5. Para futuros proyectos se recomienda utilizar algoritmos mas avanzados.
6. En el análisis exploratorio de datos se recomienda indagar más con los datos y realizar más graficas.
7. Se recomienda capacitar al personal deportivo de los clubes de fútbol, entidades deportivas y organizaciones estatales en el uso de aplicación de las estadísticas y datos en el deporte para la implementación de políticas estatales.

6 Bibliografía

Coleman, J. (2011). *Google academics*. Obtenido de https://www.researchgate.net/profile/Jason_Coleman5/publication/266354609_QR_Codes_What_Are_They_and_Why_Should_You_Care/links/5564780908ae9963a12020fa.pdf

Scrum. (2017). *What Is Scrum?* Recuperado el octubre de 2017, de Scrum.org: <https://www.scrum.org/resources/what-is-scrum>

Android developers. (2020). *Android developers*. Obtenido de <https://developer.android.com/guide/topics/manifest/manifest-intro?hl=es-419&authuser=1>

Facultad de Informática, Universidad de Murcia. (2019). *Visualización de Esquemas en Bases de Datos NoSQL basadas en documentos*. Obtenido de https://biblioteca.sistedes.es/submissions/uploaded-files/JISBD_2017_paper_71.pdf

IEBSchool. (2016). *Firebase, qué es y para qué sirve la plataforma de Google*. Obtenido de <https://www.iebschool.com/blog/firebase-que-es-para-que-sirve-la-plataforma-desarrolladores-google-seo-sem/>

FRANCO, Y. J. (2012). *SISTEMA OPERATIVO ANDROID: CARACTERÍSTICAS Y FUNCIONALIDAD PARA DISPOSITIVOS MÓVILES*. Colombia.

Coplien, J. O. (2019). *A Scrum Book: The Spirit of the Game*. the pragmatic programmers.

Weber, J. (2015). *Practical Google Analytics and Google Tag Manager for Developers*. Apress.

Amason aws. (2022). *¿Qué es el análisis de datos?*

altim analytcs. (2020). *ANÁLISIS DE DATOS APLICADO AL FUTBOL*.

Universidad de Valencia(CEACES). (2020). *Regresion*.

Premier League. (2023). *Premier Legue*. Obtenido de <https://www.premierleague.com>

FIFA. (2023). *FIFA*. Obtenido de <https://www.fifa.com>

Lewis, M. (2004). *Moneyball: El arte de ganar en un juego injusto Libro*. W. W. Norton & Company.

opta. (2023). *opta*. Obtenido de <https://www.statsperform.com/opta/>

kaggle.com. (2018). *resultados premier league*. Obtenido de <https://www.kaggle.com/datasets/zaeemnalla/premier-league?select=stats.csv>