



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**ANÁLISIS DE LOS MOVIMIENTOS EN LAS PARTIDAS DE AJEDREZ DE JUGADORES
PERTENECIENTES A UN CLUB DE AJEDRECISTAS EN GUATEMALA Y SU IMPACTO EN
LOS RESULTADOS OBTENIDOS UTILIZANDO LOS CONCEPTOS DE CIENCIA DE DATOS**

Javier Alejandro Golón López

Asesorado por el Ing. Miguel Ángel Cancinos Rendon

Guatemala, junio de 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**ANÁLISIS DE LOS MOVIMIENTOS EN LAS PARTIDAS DE AJEDREZ DE JUGADORES
PERTENECIENTES A UN CLUB DE AJEDRECISTAS EN GUATEMALA Y SU IMPACTO EN
LOS RESULTADOS OBTENIDOS UTILIZANDO LOS CONCEPTOS DE CIENCIA DE DATOS**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

JAVIER ALEJANDRO GOLÓN LÓPEZ

ASESORADO POR EL ING. MIGUEL ÁNGEL CANCINOS RENDON

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, JUNIO DE 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANA	Ing. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Armando Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIA	Inga. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Ing. Aurelia Anabela Cordova Estrada
EXAMINADOR(A)	Ing. Cesar Augusto Fernández Cáceres
EXAMINADOR(A)	Ing. Cesar Rolando Batz Saquimux
EXAMINADOR(A)	Ing. Álvaro Giovanni Longo Morales
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**ANÁLISIS DE LOS MOVIMIENTOS EN LAS PARTIDAS DE AJEDREZ DE
JUGADORES PERTENECIENTES A UN CLUB DE AJEDRECISTAS EN GUATEMALA Y SU
IMPACTO EN LOS RESULTADOS OBTENIDOS UTILIZANDO LOS CONCEPTOS DE
CIENCIA DE DATOS**

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería Ciencias y Sistemas, con fecha 23 de agosto de 2021.

Javier Alejandro Golon Lopez

A handwritten signature in black ink, consisting of a stylized 'J' followed by 'A', 'G', 'L', and 'L' in a cursive script, with a horizontal line extending to the right.

Guatemala, 11 de mayo de 2023

Ingeniero
Carlos Alfredo Azurdía
Coordinador de Privados y Trabajos de Tesis
Escuela de Ingeniería en Ciencias y Sistemas
Facultad de Ingeniería - USAC

Respetable Ingeniero Azurdia:

Por este medio hago de su conocimiento que en mi rol de asesor del trabajo de investigación realizado por el estudiante **Javier Alejandro Golón López** con carné **201700473** y CUI **3004 46578 0101** titulado **“Análisis de los movimientos en las partidas de ajedrez de jugadores pertenecientes a un club de ajedrecistas en Guatemala y su impacto en los resultados utilizando los conceptos de ciencia de datos”**, lo he revisado y luego de corroborar que el mismo se encuentra concluido y que cumple con los objetivos propuestos en el respectivo protocolo, procedo a la aprobación respectiva.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,


MIGUEL ANGEL CANCINOS RENDON
INGENIERO EN CIENCIAS Y SISTEMAS
COLEGIADO No. 11572
Ing. Miguel Angel Cancinos Rendon
Colegiado No. 11572



Universidad San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala 16 de mayo de 2023

Ingeniero
Carlos Gustavo Alonzo
Director de la Escuela de Ingeniería
En Ciencias y Sistemas

Respetable Ingeniero Alonzo:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante **JAVIER ALEJANDRO GOLÓN LÓPEZ** con carné **201700473** y CUI **3004 46578 0101** titulado **“ANÁLISIS DE LOS MOVIMIENTOS EN LAS PARTIDAS DE AJEDREZ DE JUGADORES PERTENECIENTES A UN CLUB DE AJEDRECISTAS EN GUATEMALA Y SU IMPACTO EN LOS RESULTADOS OBTENIDOS UTILIZANDO LOS CONCEPTOS DE CIENCIA DE DATOS”**, y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo aprobado.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,



Ing. Carlos Alfredo Azurdia
Coordinador de Privados
y Revisión de Trabajos de Graduación



FACULTAD DE INGENIERÍA

SIST.LNG.DIRECTOR.3.EICCSS.2023

El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador de área y la aprobación del área de lingüística del trabajo de graduación titulado: **ANÁLISIS DE LOS MOVIMIENTOS EN LAS PARTIDAS DE AJEDREZ DE JUGADORES PERTENECIENTES A UN CLUB DE AJEDRECISTAS EN GUATEMALA Y SU IMPACTO EN LOS RESULTADOS OBTENIDOS UTILIZANDO LOS CONCEPTOS DE CIENCIA DE DATOS**, presentado por: **Javier Alejandro Golon Lopez**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería.

“ID Y ENSEÑAD A TODOS”



Ingeniero Carlos Gustavo Alonzo
DIRECTOR
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, junio de 2023



USAC
TRICENTENARIA
Universidad de San Carlos de Guatemala

Decanato
Facultad e Ingeniería

24189101- 24189102

LNG.DECANATO.OIE.11.2023

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería En Ciencias Y Sistemas, al Trabajo de Graduación titulado: **ANÁLISIS DE LOS MOVIMIENTOS EN LAS PARTIDAS DE AJEDREZ DE JUGADORES PERTENECIENTES A UN CLUB DE AJEDRECISTAS EN GUATEMALA Y SU IMPACTO EN LOS RESULTADOS OBTENIDOS UTILIZANDO LOS CONCEPTOS DE CIENCIA DE DATOS.**, presentado por: **Javier Alejandro Golon Lopez** después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

Firmado digitalmente por AURELIA ANABELA CORDOVÁ ESTRADA Fecha: 02/06/2023 08:55:26 p.m. Razón: Orden de impresión Ubicación: Facultad de Ingeniería, USAC.

Inga. Aurelia Anabela Córdova Estrada
Decana



Guatemala, junio de 2023

Para verificar validez de documento ingrese a <https://www.ingenieria.usac.edu.gt/firma-electronica/consultar-documento>

Tipo de documento: Correlativo para orden de impresión Año: 2023 Correlativo: 11 CUI: 3004465780101

Escuelas: Ingeniería Civil, Ingeniería Mecánica Industrial, Ingeniería Química, Ingeniería Mecánica Eléctrica, - Escuela de Ciencias, Regional de Ingeniería Sanitaria y Recursos Hidráulicos (ERIS). Postgrado Maestría en Sistemas Mención Ingeniería Vial. Carreras: Ingeniería Mecánica, Ingeniería Electrónica, Ingeniería en Ciencias y Sistemas. Licenciatura en Matemática. Licenciatura en Física. Centro de Estudios Superiores de Energía y Minas (CESEM). Guatemala, Ciudad

ACTO QUE DEDICO A:

- Dios** Por ser una importante influencia en mi carrera, entre otras cosas.
- Mis Padres** Edgar Golón y Rosa María López. Su amor será siempre mi inspiración.
- Mi Hermanos** Eduardo y Julio Golón. Por brindarme su apoyo y consejos durante toda la carrera.
- Mis Abuelos** Nicolasa Cuyan (q.e.p.d) y Ponciano López (q.e.p.d), por su amor y consejos para guiarme a alcanzar mis metas.
- Mis tías** Carolina y Angélica López por brindarme su apoyo y cariño.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ser una importante influencia en mi carrera, entre otras cosas.
Facultad de Ingeniería	Por ser una importante influencia en mi carrera, entre otras cosas.
Mis amigos de la Facultad	José Veliz, Marco López, Andhy Solís, José Alonzo, Ángel Aquino, Diego Martínez, Davis Edward, Jonathan Orantes, Pablo Quevedo, Diego Juárez y Herbert Reyes por apoyarme a lo largo de la carrera.
Mi asesor	Ing. Miguel Ángel Cancinos, por su dirección, orientación y apoyo en la realización de mi trabajo.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	V
LISTA DE SÍMBOLOS	VII
GLOSARIO	IX
RESUMEN.....	XV
OBJETIVOS.....	XVII
INTRODUCCIÓN	XIX
1. EL AJEDREZ.....	1
1.1. Elementos del Ajedrez	1
1.1.1. Piezas.....	2
1.1.2. El tablero.....	3
1.1.2.1. Configuración inicial.....	4
1.2. Reglas básicas de competición.....	5
1.2.1. Normas	6
1.2.2. Ritmos competitivos.....	7
1.3. Sistema de notación de partidas.....	8
1.3.1. Identificación del tablero y las piezas	8
1.3.2. Movimientos.....	10
1.3.2.1. Jugadas especiales	10
1.3.3. Sistema de notación para computadoras	11
1.3.3.1. Formato PGN.....	12
1.4. Sistemas computarizados para análisis de ajedrez	14
1.4.1. Motor Stockfish	15
1.5. El ajedrez en Guatemala.....	16
1.5.1. Competencias nacionales.....	17

1.5.2.	Modalidad Actual de competencias	18
2.	CIENCIA DE DATOS	19
2.1.	Métodos y técnicas	20
2.1.1.	Técnicas algebraicas y estadísticas	20
2.2.	Análisis de datos.....	20
2.2.1.	Tipos de análisis de datos	21
2.2.1.1.	Análisis cuantitativo	22
2.2.1.2.	Análisis cualitativo	22
2.3.	Procesos y técnicas de integración de datos.....	23
2.3.1.	Consolidación de datos	24
2.3.2.	Extraer transformar y cargar (ETL).....	25
2.3.3.	Uso del lenguaje Python.....	27
2.4.	Importancia de la visualización de datos	28
2.4.1.	Tipos de visualizaciones.....	29
2.4.1.1.	Gráficos más comunes.....	30
3.	RECOPIACIÓN Y PREPARACIÓN DE DATOS	31
3.1.	Selección de partidas.	31
3.2.	Conversión de PGN a JSON	32
3.2.1.	Uso del intérprete de Python	32
3.2.2.	Librería Chess.pgn	32
3.2.3.	Análisis y creación del archivo JSON	33
4.	TRANSFORMACIÓN Y LIMPIEZA DEL ARCHIVO JSON.....	35
4.1.	Conexión a base de datos mongo	35
4.2.	Estructura del <i>script</i> de carga y transformación.....	35
4.2.1.	Inserción del archivo JSON a la base de datos	36
4.2.2.	Eliminación de duplicados	37

4.2.3.	Eliminación de campos	37
4.2.4.	Actualización de valores	38
4.3.	Obtención del archivo JSON procesado.	39
5.	ANÁLISIS Y REPRESENTACIÓN DE RESULTADOS	41
5.1.	Preprocesamiento de datos	41
5.1.1.	Manipulación de datos con Python	42
5.1.1.1.	Patrones de jaque mate más comunes	42
5.1.1.2.	Porcentaje de éxito por apertura.....	44
5.1.1.3.	Patrones de enroque más utilizados....	47
5.1.1.4.	Aperturas con mayor porcentaje de éxito.....	49
5.1.2.	Manipulación de datos con Power Bi.....	51
5.1.2.1.	Análisis de porcentaje de victorias por color	51
5.1.2.2.	Análisis de captura de piezas	53
5.2.	Identificación de las piezas involucradas en jaque mate más habituales.....	55
5.3.	Evaluación de la efectividad de las aperturas en el ajedrez mediante análisis de porcentaje de éxito	57
5.4.	Evaluación de las estrategias de enroque más utilizadas en el ajedrez	60
5.5.	Análisis de las piezas más agresivas en ajedrez	62
5.6.	Análisis de la distribución de victorias por color	64
5.7.	Análisis de aperturas más frecuentes.	66
5.8.	Optimización de jugadas de ajedrez con el motor Stockfish....	71
5.8.1.	Obtención de Mejor Movimiento a la primera apertura más frecuente.....	72

5.8.2.	Obtención de mejor movimiento a la segunda apertura más frecuente	72
5.8.3.	Obtención de mejor movimiento a la tercera apertura más frecuente.	73
5.9.	Optimización de la toma de decisiones en ajedrez mediante el análisis computacional de áreas críticas.....	74
CONCLUSIONES.....		81
RECOMENDACIONES		83
REFERENCIAS		85
APÉNDICES.....		89

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Numeración del tablero de ajedrez.....	4
2.	Distribución de las piezas en el tablero	5
3.	Identificación numérica del tablero	9
4.	Ejemplo partida de ajedrez en notación PGN.....	14
5.	Técnicas fundamentales de integración de datos	24
6.	Representación general del flujo de los procesos de ETL	27
7.	Página web Chess24.com.....	31
8.	Formato final de una partida en formato JSON	33
9.	Carga el archivo JSON de las partidas	36
10.	Método para la inserción del archivo JSON	36
11.	Ejemplo de archivo JSON final.....	40
12.	<i>Script</i> en Python para detectar patrones de jaque mate más comunes.....	44
13.	<i>Script</i> en Python para detectar porcentaje de éxito en apertura, sección 1	46
14.	<i>Script</i> en Python para detectar porcentaje de éxito en apertura, sección 2	47
15.	<i>Script</i> en Python para detectar patrones de enroques más utilizados....	48
16.	<i>Script</i> en Python para detectar aperturas más frecuentes, sección 1. ...	50
17.	<i>Script</i> en Python para detectar aperturas más frecuentes, sección 2. ...	50
18.	Paso para agregar condiciones a nueva columna en Power Bi	53
19.	Gráfico Análisis piezas involucradas en jaque mate	56
20.	Gráfico Análisis porcentaje de victorias por apertura	58

21.	Gráfico análisis cantidad de Enroques por tipo	60
22.	Gráfico análisis cantidad de capturas por piezas	63
23.	Gráfico análisis porcentaje de victorias por color	65
24.	Primera apertura más frecuente	67
25.	Segunda apertura más frecuente.....	68
26.	Tercera apertura más frecuente.....	70
27.	Análisis áreas atacables y débiles 1.	76
28.	Análisis áreas atacables y débiles 2	78
29.	Análisis áreas atacables y débiles 3.	79

TABLAS

I.	Etiquetas Formato PGN.....	13
II.	Descripción de los componentes del proceso ETL.	26
III.	Descripción de resultados de las partidas	39
IV.	Representación remplazos de piezas	55

LISTA DE SÍMBOLOS

Símbolo	Significado
json	Extensión de un archivo en formato JSON.
csv	Extensión de archivo con valores separados por comas.
Pgn	Formato de archivo utilizado para almacenar partidas de ajedrez en texto plano.
SVG	Formato de imagen vectorial escalable utilizado en la <i>web</i> .

GLOSARIO

<i>Big data</i>	Término que se refiere a la gestión y análisis de grandes conjuntos de datos.
<i>BlackElo</i>	Puntuación de Elo de un jugador de ajedrez con las piezas negras.
<i>Blitz</i>	Partida de ajedrez en la que cada jugador dispone de un tiempo muy limitado para realizar todas sus jugadas.
C	Lenguaje de programación de bajo nivel ampliamente utilizado para desarrollar aplicaciones de sistema y <i>software</i> de base.
Chess24	Plataforma de ajedrez en línea que ofrece partidas, lecciones y otros recursos relacionados con el ajedrez.
CUDA	Plataforma de computación paralela desarrollada por Nvidia que permite el procesamiento en paralelo en las GPU.

<i>Data Science</i>	Campo interdisciplinario que involucra métodos estadísticos, algoritmos de aprendizaje automático y herramientas de visualización de datos para extraer conocimientos y entendimiento de datos estructurados y no estructurados.
<i>Data Warehouse</i>	Sistema utilizado para almacenar y gestionar grandes conjuntos de datos estructurados para el análisis y la toma de decisiones empresariales.
EPD	Formato de archivo utilizado para almacenar posiciones de ajedrez.
ETL	Proceso de extracción, transformación y carga de datos desde múltiples fuentes a un sistema de destino.
FEN	Formato utilizado para describir una posición de ajedrez en términos de las piezas en el tablero y el estado del juego.
FIDE	Federación Internacional de Ajedrez, organismo rector del ajedrez mundial.
Fichero	Término que se utiliza en programación para referirse a un archivo que se utiliza para almacenar información.

Fortran	Lenguaje de programación de alto nivel utilizado principalmente para aplicaciones científicas y de ingeniería.
<i>Framework</i>	Estructura de trabajo o conjunto de herramientas que se utiliza para desarrollar aplicaciones de <i>software</i> .
GPU	Unidad de procesamiento gráfico, componente de hardware utilizado para acelerar el procesamiento gráfico en ordenadores.
IBM	Empresa multinacional estadounidense de tecnología y consultoría.
Java	Lenguaje de programación orientado a objetos utilizado para desarrollar aplicaciones de <i>software</i> .
LAPACK	Biblioteca de subrutinas de álgebra lineal utilizada en aplicaciones de procesamiento numérico intensivo.
<i>Machine Learning</i>	Rama de la inteligencia artificial que se centra en el desarrollo de algoritmos que permiten a los sistemas aprender y mejorar a partir de los datos.
Mongo Client	Biblioteca de Python para interactuar con bases de datos MongoDB.
MongoDB	Sistema de base de datos NoSQL orientado a documentos.

NEA	Siglas de no ejecución de adversidad, término utilizado en ajedrez para describir una partida en la que ambos jugadores juegan de manera pasiva.
Panda	Biblioteca de Python utilizada para el análisis y la manipulación de datos.
Power Bi	Plataforma de análisis de datos y generación de informes de Microsoft.
Power Query	Herramienta de Excel que permite la transformación y la combinación de datos de diferentes orígenes.
Python	Lenguaje de programación interpretado de alto nivel utilizado en una amplia variedad de aplicaciones de <i>software</i> .
Pymongo	Biblioteca de Python para interactuar con bases de datos MongoDB.
SARS-Cov2	Virus responsable de la pandemia de COVID-19 que comenzó en 2019.
SciPy	Biblioteca de Python utilizada para el procesamiento de señales, la optimización, el álgebra lineal y otros cálculos científicos.
Script	Secuencia de comandos o código utilizado para realizar una tarea automatizada.

TensorFlow	Plataforma de aprendizaje automático de código abierto desarrollada por Google.
WhiteElo	Puntuación de Elo de un jugador de ajedrez con las piezas blancas.
Windows	Sistema operativo desarrollado por Microsoft.

RESUMEN

El presente trabajo aborda el análisis de partidas de ajedrez utilizando técnicas de ciencia de datos. Se explica la estructura del ajedrez, sus reglas básicas y su sistema de notación de partidas. Se detalla el proceso de recopilación y preparación de los datos, así como su transformación y limpieza.

Además, se presentan los resultados del análisis de los datos, entre ellos, los patrones de jaque mate más comunes, el porcentaje de éxito de las aperturas y las estrategias de enroque más utilizadas. También se analizan las piezas más agresivas y la distribución de victorias por color. Se emplean herramientas como Python y Power BI para la manipulación y visualización de datos. Este trabajo contribuye a la comprensión del y análisis de ajedrez la cual puede ser útil para entrenadores y jugadores que busquen mejorar su desempeño en el juego.

OBJETIVOS

General

Analizar los movimientos en las partidas de ajedrez de jugadores pertenecientes a un club de ajedrecistas en Guatemala y su impacto en los resultados obtenidos utilizando los conceptos de ciencia de datos.

Específicos

1. Recolectar la información de la base de datos chess.com que permita obtener un panorama completo y preciso del comportamiento de los ajedrecistas guatemaltecos.
2. Procesar los datos y estructurar un formato que permita producir información significativa de los movimientos de las partidas de ajedrez.
3. Prevenir y corregir los errores de los datos que puedan afectar el análisis final de las partidas de ajedrez.
4. Relacionar los resultados finales de las partidas de ajedrez con los movimientos obtenidos de cada partida.
5. Brindar detalles e información que ayuden a tomar mejores decisiones a los ajedrecistas en futuras competiciones nacionales.

INTRODUCCIÓN

El ajedrez es un juego milenario que ha sido estudiado desde diferentes enfoques, incluyendo la teoría de juegos, la inteligencia artificial y la psicología. En la actualidad, la ciencia de datos se ha convertido en una herramienta fundamental para analizar y entender el comportamiento de los jugadores y las estrategias utilizadas en este juego.

El propósito de esta investigación es analizar las partidas de ajedrez jugadas en Guatemala y utilizar técnicas de ciencia de datos para identificar patrones y tendencias en el juego. Para ello, se recopilieron partidas jugadas en competencias nacionales y se convirtieron a un formato adecuado para el análisis de datos. A continuación, se aplicaron técnicas de integración, transformación y limpieza de datos para obtener un conjunto de datos consistente y completo.

El análisis de los datos se llevó a cabo utilizando Python y Power Bi, aplicando técnicas de visualización y estadística para identificar patrones en el juego. Se evaluaron aspectos como la efectividad de las aperturas, los patrones de enroque más utilizados y la distribución de victorias por color, entre otros.

El resultado de esta investigación permitirá tener una visión más clara del ajedrez en Guatemala y de las estrategias utilizadas por los jugadores. Asimismo, se espera que los resultados puedan ser utilizados para mejorar la enseñanza y el entrenamiento de este juego en el país.

1. EL AJEDREZ

El ajedrez es uno de los juegos de estrategias más conocidos alrededor del mundo, la trascendencia del propio juego ha llegado a provocar que el Comité olímpico (COI) lo considere como un deporte el cual incluso es regulado de manera global por la Federación Internacional de Ajedrez (FIDE). Su denominación como deporte surge de la competitividad y la intensidad con la que dos oponentes se enfrentan en un periodo de tiempo tan continuo. Se requiere de una alta concentración, tal vez de las más exigentes entre otros deportes, lo cual implica un desgaste psicológico y físico muy demandante.

El origen del ajedrez ha sido un tema de debate por mucho tiempo, con el transcurrir de los años, y de investigaciones de diferentes historiadores, se han determinado numerosas teorías, sin embargo, la más acertada es en la que se argumenta que proviene de la india cerca del siglo VI d.C. Inicialmente no era tomado como un pasatiempo sino como una forma de representar una batalla y de esa manera poder idear estrategias que luego eran llevadas a cabo por los soldados en el campo de guerra.

1.1. Elementos del Ajedrez

Para poder llevar a cabo una partida de ajedrez únicamente son necesarios dos elementos; un tablero y las diferentes piezas. Cada jugador debe tener conocimiento de las reglas y de un espacio físico en donde desarrollar el enfrentamiento.

De manera competitiva se suele emplear como elemento opcional un reloj el cual se encarga de contabilizar el tiempo que cada jugador ha invertido en pensar la estrategia de sus jugadas en una partida con el objetivo de evitar que la partida se prolongue indefinidamente.

1.1.1. Piezas

Para poder diferenciarse dentro del tablero de juego, cada jugador recibe piezas de diferente color, uno controla las de color claro y otro las de color oscuro, normalmente se le llaman las blancas y las negras. Al comienzo de la partida cada uno recibe un total de 16 piezas que se distribuyen en: 8 peones, 2 caballos, 2 alfiles, 2 torres, 1 dama y 1 rey. Todo el juego gira alrededor del comportamiento de las 32 piezas en el tablero. Cada pieza tiene un comportamiento diferente el cual se ve reflejado en su forma de desplazarse por el tablero.

- El rey se puede mover en cualquier dirección (vertical, horizontal y en diagonal) avanzando o retrocediendo una única casilla (excepto en el enroque, en el cual se puede llegar a mover de 2 a 3 casillas).
- La dama o reina al igual que el rey se puede mover en cualquier dirección (vertical, horizontal y en diagonal) con la diferencia que puede avanzar o retroceder el número de casillas que se desee, hasta topar con el borde del tablero u otra pieza.
- El alfil únicamente se puede mover en dirección diagonal, las casillas que se deseen, hasta topar con otra pieza o el borde del tablero.
- La torre se puede mover únicamente en direcciones verticales y horizontales, hasta topar con otra pieza o el borde del tablero.

- El caballo se caracteriza por realizar su movimiento en forma de 'L', siendo la única pieza que es capaz de saltar por encima de otras piezas. De forma más técnica su movimiento está definido por un avance de dos casillas en vertical y uno en horizontal.
- El peón es la pieza que más representación en cantidad tiene sobre el tablero, su movimiento se representa únicamente como un avance de una o dos casillas en vertical, después de moverse por primera vez sólo puede avanzar una casilla, no tiene permitido retroceder en el tablero y la forma de captura de piezas contrarias se realiza de forma diagonal cuando tiene una pieza delante de él (excepto en la jugada del peón al paso).

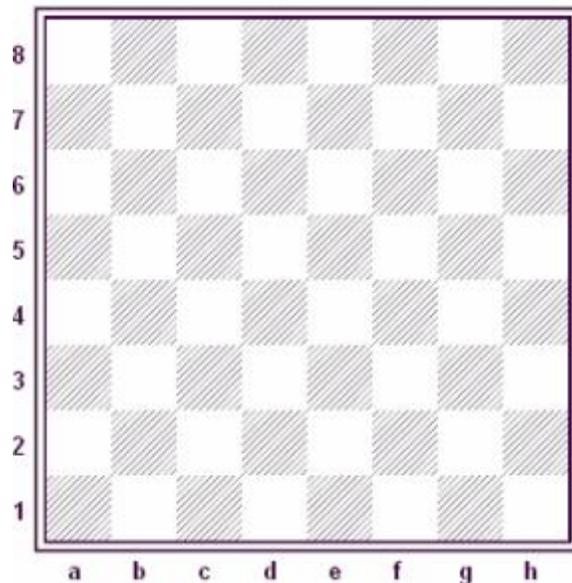
1.1.2. El tablero

El tablero se compone de ocho filas y ocho columnas las cuales forman 64 casillas de igual tamaño. Las casillas van alternando su color entre claro y oscuro.

Los elementos básicos del tablero son:

- Fila. Son las ocho hileras horizontales del tablero. Se nombran con números del 1 al 8, comenzando con la primera fila del lado en donde se encuentran las piezas de color claro o blancas.
- Columna. Es cada una de las ocho hileras verticales del tablero. Se nombran con letras minúsculas de la a, a la h, comenzando desde la primera columna izquierda de lado en donde se posicionan las piezas de color claro.
- Diagonal. Es una línea recta que forman las casillas del mismo color de borde a borde en ángulo.
- Centro. Son las cuatro casillas situadas en medio del tablero.
- Esquinas. Cada una de las cuatro casillas ubicadas los vértices del tablero.

Figura 1. **Numeración del tablero de ajedrez**



Fuente: Aula de Ajedrez (2012). *El tablero de ajedrez*. Consultado el 13 de abril de 2023.
Recuperado de https://www.chess-poster.com/english/learn_chess/notation/images/coordinates_1.gif.

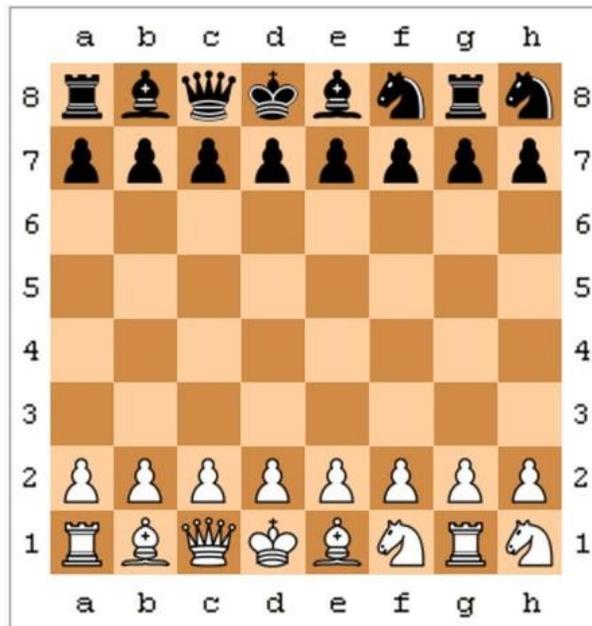
1.1.2.1. Configuración inicial

Cada jugador se debe posicionar de frente con el ajedrecista contrincante, colocando el tablero de tal forma que cada jugador tenga una casilla blanca a su derecha.

Las piezas se deben colocar empezando con la primera fila, en donde en los extremos se colocan las torres, caballos y el alfil, dejando para las casillas de en medio el espacio para la reina, que se coloca en la casilla de su respectivo color, y el rey en la casilla restante. En la segunda fila se colocan los 8 peones

formando una hilera. La disposición final queda como se indica en la siguiente imagen.

Figura 2. **Distribución de las piezas en el tablero**



Fuente: Xataka (2021). *Si no te gusta el ajedrez tradicional, aquí tienes 17 "forks" del juego que conquista a medio mundo*. Consultado el 13 de abril de 2023. Recuperado de <https://www.xataka.com/especiales/no-te-gusta-ajedrez-tradicional-aqui-tienes-forks-juego-que-conquista-a-medio-mundo>.

1.2. Reglas básicas de competición

Explicar todos los movimientos posibles dentro de una partida de ajedrez y las reglas que deben de llevar cada uno de ellos, sobrepasa el objetivo general de este tema de investigación, por lo que únicamente se van a resaltar las reglas que se deben de cumplir en competiciones oficiales que van más allá del desarrollo de una partida, así como un resumen de las normas básicas de juego.

De forma resumida una partida de ajedrez se juega entre dos adversarios que mueven alternativamente sobre el tablero de ajedrez. El jugador que posee las piezas de color claro es el que realiza el primer movimiento, posteriormente el jugador con piezas negras contesta con un movimiento y así sucesivamente transcurre el juego.

El objetivo de cada jugador es colocar al rey del contrincante en una posición en donde se le disminuya la cantidad de movimientos legales que este puede realizar, a esta jugada se le denomina jaque y jaque mate cuando el contrincante ya no tiene casillas a donde mover su rey bajo movimientos legales. El adversario cuyo rey haya recibido el jaque mate, pierda la partida, si en dado caso ninguno de los dos jugadores logro dar un mate la partida queda empate o en tablas, como se le denomina comúnmente.

1.2.1. Normas

Conociendo un poco del funcionamiento del ajedrez, una partida competitiva se juega con las mismas reglas que una informal. Sin embargo, en competiciones se agregan normas las cuales evitan tener inconvenientes con los participantes.

Las normas que son obligatorias en una competición oficial son las siguientes:

- **Pieza tocada, pieza movida:** si se toca una pieza es obligatorio hacer el movimiento con esa pieza. Esto pretende que el jugador tenga que analizar el entorno antes de tomar una decisión. En el caso que se toque una pieza del oponente, el contrario tiene la oportunidad de capturar una pieza nuestra.

- Hablar durante las partidas: está completamente prohibido mantener una conversación con el rival durante el desarrollo de una partida de ajedrez. Tampoco es permitido hablar con un espectador o con cualquier persona que intente interrumpir el juego y la concentración de los jugadores. La única excepción de comunicación son las acciones de proponer tablas, para solicitar un reloj y hablar con el árbitro para indicar el resultado final de la partida o algún problema.
- Uso de dispositivos móviles: llevar a una partida de ajedrez oficial un dispositivo móvil es sancionado con la pérdida de la partida de manera automática, sin embargo, hoy en día, algunos torneos permiten llevar un dispositivo siempre y cuando este apagado.
- Uso del reloj de ajedrez: la regla más importante es presionar el reloj luego de cada movimiento y hacerlo con la misma mano que se utilizó para mover la pieza.

1.2.2. Ritmos competitivos

El tiempo de duración de una partida de ajedrez amistosa puede llegar a durar horas o días, cuando hablamos de competiciones oficial, es necesario contar con una estructura que limite el tiempo disponible para cada jugador.

De manera oficial se cuentan con los siguientes modos de controlar el tiempo en una partida.

- *Blitz* o relámpago: cada jugador dispone de un máximo de 15 minutos para tratar de obtener la victoria. En este tipo de modalidad la estrategia suele dejarse a lado, la presión del tiempo afecta las decisiones que los jugadores deben tomar.

- Partida rápida: en este tipo de modalidad se suele otorgar entre 15 y 60 minutos de tiempo para cada jugador. Es muy utilizado en competiciones oficial para determinar el ganador luego de un empate.
- Clásico: el tiempo para cada jugador suele ser mayor a los 60 minutos. Es el formato utilizado en las Olimpiadas de ajedrez, donde se le agregan 30 segundos de incremento por jugada.

1.3. Sistema de notación de partidas

Las competencias oficiales de ajedrez requieren que los jugadores escriban sus movimientos utilizando una notación específica, la razón de plasmar cada jugada es un simple requisito para poder participar en los torneos y poder así tener un acta en donde el árbitro pueda ver la progresión del juego y tener una base sólida por si se presentan problemas.

Actualmente la tarea de apuntar cada jugada está cobrando mucho auge, ya que a partir de esos datos se pueden generar registros que permitan analizar los momentos críticos de las partidas jugadas con el fin de mejorar el rendimiento del ajedrecista, precisamente lo que esta investigación busca apoyar.

1.3.1. Identificación del tablero y las piezas

La notación que se utiliza se denomina notación estándar algebraica (NEA), tiene como base el sistema de coordenadas que posee el tablero, las filas siguen una numeración de 1 a 8 y las columnas de la letra a, a la h, con esto tenemos todo lo necesario para dar a las 64 casillas del tablero un identificador único compuesto por una letra y un número, las cuales se obtiene de la intersección de la fila y columna.

Figura 3. **Identificación numérica del tablero**

El diagrama muestra un tablero de ajedrez de 8x8 casillas. Las casillas están etiquetadas con letras minúsculas (a-h) y números (1-8) que indican su posición en el sistema algebraico. Las casillas de color claro (blanco) tienen la letra y el número en verde, mientras que las casillas de color oscuro (negro) tienen la letra y el número en rojo. Las etiquetas de las columnas (a-h) están en la parte inferior, y las etiquetas de las filas (1-8) están a la derecha.

a8	b8	c8	d8	e8	f8	g8	h8	8
a7	b7	c7	d7	e7	f7	g7	h7	7
a6	b6	c6	d6	e6	f6	g6	h6	6
a5	b5	c5	d5	e5	f5	g5	h5	5
a4	b4	c4	d4	e4	f4	g4	h4	4
a3	b3	c3	d3	e3	f3	g3	h3	3
a2	b2	c2	d2	e2	f2	g2	h2	2
a1	b1	c1	d1	e1	f1	g1	h1	1
a	b	c	d	e	f	g	h	

Fuente: Chess Campus (2021). *Como se anotan las jugadas en el sistema algebraico.*

Consultado el 13 de abril de 2023. Recuperado de <https://chesscampus.com/curso/ajedrez-ninos/como-se-anotan-las-jugadas-en-el-sistema-algebraico/>.

Ya que se tiene identificadas las posiciones del tablero de juego, es necesario tener una abreviatura para cada pieza de juego. Se utiliza la primera letra del nombre de la pieza en mayúscula. En caso de hablar otro idioma, se utiliza de igual forma la primera letra, pero de la palabra en idioma en el que se está contextualizando.

En caso de utilizar el idioma español, se utilizan las siguientes abreviaciones.

- Rey. R
- Dama. D
- Alfil. A

- Caballo. C
- Torre. T
- Peón: este es un caso especial el cual no lleva notación, únicamente se indica usando el nombre de la casilla.

1.3.2. Movimientos

Teniendo la base de la notación para el tablero y las piezas, para indicar los movimientos basta hacer la unión y tener una cuenta excepciones para jugadas especiales.

Los movimientos se representan en parejas, mostrando como primero el turno de las piezas claras y luego el de las oscuras; Por ejemplo, e4 d5 expresa que el jugador primer jugador movió su peón a la casilla e4 y su oponente respondió desplazando su peón a e5.

1.3.2.1. Jugadas especiales

El ajedrez tiene movimientos especiales que se suelen utilizar para ampliar la estrategia de cada jugador y se pueden ejecutar en un único turno, cada jugada tiene notación especial.

- Capturas. Se presenta cuando se captura una pieza enemiga, su forma de representación es un x entre la pieza movida y la casilla en la que ubicaba la pieza capturada. Un caso especial es la captura que se realiza entre peones, para este caso es necesario indicar de que columna proviene el peón de captura.

- Jaque y Jaque mate. Cuando el rey adversario esta ante una amenaza de captura (jaque), se agrega el signo + al final de la notación. Jaque Mate se denota con ++ o #, seguido por el resultado del juego (1-0 si el jugador de piezas claras ha ganado, o 0-1 si el oscuro ha salido triunfante). El movimiento.
- Enroque. Para este caso tenemos dos opciones, enroque en el flanco de rey, el cual se denomina como 0-0 y enroque de Dama como 0-0-0, también los movimientos son conocidos como enroque largo y enroque corto, uno comienza en la columna h y el otro en la a.
- Coronación del peón. Esta situación se da cuando un peón logra alcanzar el final del tablero, la forma de indicarlo es colocando el símbolo =, seguido de la letra en mayúscula de la abreviación de la pieza a la cual se elige transformar.

Por último, se tiene una notación en caso de que exista ambigüedad en el tablero, esta situación se suele dar cuando existen dos piezas del mismo tipo en la misma fila y ambas tiene movimientos posibles en común, en este caso se agrega un identificador de inicio entre la pieza y la casilla.

1.3.3. Sistema de notación para computadoras

Con el tiempo se han desarrollado varios modos de registrar los movimientos realizados en una partida de ajedrez, existen formatos para notación de humanos y para sistemas informáticos. Nuestro tema de interés se centra en los formatos informáticos, en donde encontramos tres notaciones.

La notación más común en el ámbito informático es; de Forsyth-Edwards (FEN), Descripción de posición extendida (EPD), Steno-chess y la notación portátil de partida (PGN), esta última es la más utilizada por los programas y en la cual nos enfocaremos.

1.3.3.1. Formato PGN

La notación portátil de partida (PGN) es la más utilizada por los programas informáticos de ajedrez para almacenar y gestionar partidas de ajedrez. Se destaca por ser fácil de acceder y de ser manipulable.

El formato para almenar la información se rige por etiquetas las cuales se presentan en el siguiente orden.

Tabla I. **Etiquetas formato PGN**

Etiqueta	traducción	Descripción
<i>Event</i>	Evento	Nombre del torneo o competencia
<i>Site</i>	Lugar	El lugar donde se llevó a cabo el evento, Esto debe ser en formato Ciudad, Región PAÍS, donde PAÍS es el código de este en tres letras de acuerdo con el código del Comité Olímpico Internacional
<i>Date</i>	Fecha	La fecha de inicio de la partida. en formato AAAA.MM.DD. Cuando se desconocen los valores se utilizan ??.
<i>Round</i>	Ronda	Numero de ronda del torneo o competencia
<i>White</i>	Blancas	El jugador de las piezas blancas, en formato apellido, nombre.
<i>Black</i>	Negras	El jugador de las negras en el mismo formato.
<i>Result</i>	Resultado	El resultado del juego. Sólo puede tener cuatro posibles valores: 1-0 (las blancas ganaron), 0-1 (Las negras ganaron), 1/2-1/2 (Tablas), o * (cuando el juego está actualmente en disputa, un jugador fallece durante la partida o la partida ha sido postergada).
<i>WhiteCountry</i>	País blancas	Abreviación del país de origen del jugador con piezas blancas
<i>BlackCountry</i>	País negras	Abreviación del país de origen del jugador con piezas Negras

Fuente: elaboración propia, realizado con Microsoft Word.

Para indicar todos los movimientos que surgieron en una partida, el texto de las jugadas se describe bajo el formato (NEA).

Figura 4. Ejemplo partida de ajedrez en notación PGN

```
[Event "Club Caissa Guatemala"]
[Site "chess24.com"]
[Date "2020.01.08"]
[Round "1"]
[White "Mareco, Sandro"]
[Black "Juarez Flores, Gustavo E."]
[Result "1-0"]
[Board "1"]
[WhiteElo "2619"]
[WhiteTitle "GM"]
[WhiteCountry "ARG"]
[WhiteFideId "112275"]
[WhiteEloChange "1"]
[BlackElo "2065"]
[BlackTitle "FM"]
[BlackCountry "GUA"]
[BlackFideId "7300034"]
[BlackEloChange "-2"]

1. d4 {[%clk 1:30:56]} Nf6 {[%clk 1:30:41]} 2. c4 {[%clk 1:31:23]} g6 {[%clk
1:30:30]} 3. Nc3 {[%clk 1:31:48]} Bg7 {[%clk 1:30:50]} 4. e4 {[%clk 1:32:14]} d6
{[%clk 1:31:12]} 5. Nf3 {[%clk 1:32:00]} O-O {[%clk 1:31:31]} 6. Be2 {[%clk
1:32:26]} e5 {[%clk 1:30:29]} 7. O-O {[%clk 1:32:50]} Nc6 {[%clk 1:30:21]} 8. d5
{[%clk 1:33:12]} Ne7 {[%clk 1:30:43]} 9. b4 {[%clk 1:33:37]} a5 {[%clk 1:30:45]}
10. Ba3 {[%clk 1:32:18]} axb4 {[%clk 1:29:58]} 11. Bxb4 {[%clk 1:32:42]} Nh5
```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.

1.4. Sistemas computarizados para análisis de ajedrez

Los motores de ajedrez son programas informáticos que utilizan algoritmos de búsqueda y evaluación de posiciones para analizar partidas de ajedrez y posiciones individuales. Estos programas pueden proporcionar a los jugadores de ajedrez una valiosa información sobre el juego, como la mejor jugada posible en una posición dada, las fortalezas y debilidades de una posición, y posibles líneas de juego para ambos jugadores. Los motores de ajedrez también se utilizan en el análisis de partidas históricas y en la preparación de

aperturas y defensas específicas. En este sentido, los motores de ajedrez se han convertido en una herramienta fundamental para el análisis y mejora del juego de ajedrez, tanto para jugadores profesionales como aficionados.

1.4.1. Motor Stockfish

Stockfish es un motor de ajedrez de código abierto y uno de los programas de ajedrez más fuertes del mundo. El motor de ajedrez utiliza algoritmos de búsqueda y evaluación de posiciones para analizar las posiciones de ajedrez y tomar decisiones sobre cuál es el mejor movimiento para realizar.

Stockfish funciona mediante el uso de un árbol de búsqueda. El árbol de búsqueda comienza en la posición actual del tablero y se ramifica en todas las posibles jugadas que puede realizar el motor. Luego, para cada una de estas posibles jugadas, el motor realiza un análisis de búsqueda en profundidad, evaluando las posibles jugadas subsiguientes y seleccionando la mejor jugada posible. Este proceso se repite varias veces hasta que se alcanza la profundidad de búsqueda deseada o se alcanza un tiempo límite de cálculo.

Stockfish utiliza un conjunto de evaluaciones de posición para determinar la fuerza de cada posición. Estas evaluaciones tienen en cuenta diversos factores, como la materialidad, la estructura de peones, la movilidad de las piezas y la seguridad del rey. Utilizando estas evaluaciones, el motor de ajedrez puede determinar cuál es la mejor jugada posible en una posición dada.

El motor Stockfish se utiliza principalmente para analizar posiciones de ajedrez y partidas, para entrenar y mejorar el juego de los jugadores y para competir en torneos de ajedrez en línea y en computadora. Además, Stockfish también se utiliza como base para desarrollar otros programas de ajedrez y como

una herramienta para la investigación en inteligencia artificial y aprendizaje automático.

1.5. El ajedrez en Guatemala

Se especula que, en la época colonial, algunas familias practicaban el ajedrez de manera amistosa, siendo una costumbre que los españoles trajeron consigo al momento de la Conquista Española de Centro América en el Siglo XVI.

Durante años el ajedrez se había visto en el país como un juego de mesa más, pero con el nombramiento del General Miguel García Granados como presidente facto de Guatemala (1871- 1873), el ajedrez tuvo un despegue de popularidad ya que el mismo presidente se había declarado gran aficionado al ajedrez, incluso participaba de partidas con apuestas de dinero ya que en ese entonces atravesaba una difícil situación económica y su habilidad como ajedrecista lo hacía destacar.

Posteriormente en el año 1923 se celebró el primer campeonato nacional de ajedrez, el cual fue organizado por la institución Sociedad de Auxilios Mutuos del Comercio de Guatemala, cabe mencionar que al inicio eran eventos en donde únicamente participaban las clases altas de la sociedad guatemalteca.

Guatemala aun no contaba con entidades que rigieran el ajedrez a nivel nacional, el 25 de abril de 1927 se fundó el Centro de Ajedrecistas de Guatemala. Bajo esta entidad se realizaron múltiples campeonatos, hasta que en el año de 1939 se logró la afiliación con la FIDE (Federación Internacional de Ajedrez), abriendo la oportunidad para que Guatemala pudiese participar en eventos internacionales.

Sin embargo, debido a los conflictos generados durante la Segunda Guerra Mundial, Guatemala perdió el contacto con la FIDE, teniendo como consecuencia la pérdida de participación en eventos internacionales.

Habiéndose resuelto los conflictos generados por la Segunda Guerra Mundial, el 2 de mayo de 1996 el Centro de Ajedrecistas de Guatemala se transformó en lo que se conoce actualmente como la Federación Nacional de Ajedrez. La relación con la FIDE se fortaleció y desde ese entonces Guatemala ha estado en constante participación en eventos internacionales.

1.5.1. Competencias nacionales

El primer campeonato jugado en la ciudad de Guatemala se llevó a cabo en el año de 1946, con la participación de El Salvador, Nicaragua, Costa Rica y Guatemala, tuvo tanto éxito que se decidió establecer como un evento anual.

Viendo la expectativa que generó el campeonato realizado por Guatemala, años más tarde de celebrarse la primera edición, se decidió expandir la cantidad de invitados y se incluyeron múltiples países pertenecientes al Caribe y Centro América.

Cada año el público interesado en el ajedrez fue creciendo, y con ello la participación de guatemaltecos en competencias internacionales de gran renombre, Guatemala fue ganadora de varios campeonatos en países como México, Colombia, Filipinas y Cuba. Incluso se tuvo participación en múltiples Olimpiadas de Ajedrez.

Actualmente en Guatemala el ajedrez se practica desde temprana edad, colegios y universidades implementan clases de ajedrez dentro de su malla curricular. El ajedrez tiene cobertura nacional y a la fecha se cuentan con 22 Asociaciones Deportivas Departamentales.

1.5.2. Modalidad Actual de competencias

La actual pandemia por el coronavirus SARS-Cov2 ha provocado que el aplazamiento de múltiples campeonatos, incluso las partidas entre amigos del colegio o universidad se han visto canceladas debido a la necesaria acción del distanciamiento social. Pero sucedió algo inesperado y es que la pasión por el ajedrez se ha disparado y es gracias a su traslado al internet. La FIDE viendo esto como una oportunidad de oro y ahora organiza un promedio de 80 torneos diarios.

La FIDE anuncio que las plataformas y clubes de ajedrez online informan que han tenido un crecimiento entre el 30 % y el 40 %. Millones de usuarios ahora utilizan plataformas *web* para practicas el ajedrez, ahora es más fácil organizar una competencia con personas de diferentes países.

El trasladar la interacción al *internet* brinda la posibilidad de generar millones de datos, datos que antes no eran sencillos de obtener o no había una cantidad significativa de ellos. Un desarrollo de análisis de datos sobre el ajedrez ahora es más sencillo debido se tiene toda una base de datos de partidas a la cuales podemos acceder.

2. CIENCIA DE DATOS

El aumento y la disponibilidad de datos generados por usuarios, servicios, dispositivos y cualquier actividad que se realice de forma masiva, han influenciado el nacimiento de métodos científicos que estén orientados a explotar datos para extraer conocimiento

Según cálculos del IBM, a diario se generan más de 43 millones de *terabytes* datos y se anticipa un crecimiento exponencial durante la próxima década.

La ciencia de datos es hoy en día la herramienta fundamental para la explotación de datos y la generación de conocimiento. Entre los objetivos que persigue se encuentra la búsqueda de modelos que describan patrones de y comportamientos a partir de datos con el fin de tomar decisiones o hacer predicciones. (García, et al, 2018, p. 5)

La ciencia de datos combina técnicas, matemáticas, estadísticas, de reconocimiento de patrones y modelos de visualización que trabajan en conjunto para brindar métodos de computación e infraestructura que logren capturar, almacenar y procesar datos con el objetivo de encontrar la forma de sacar ventaja de los datos.

2.1. Métodos y técnicas

Los modos en los cuales podemos llevar a cabo un análisis de grandes volúmenes de información son muy elevados, hay un método o técnica que se especializa para cada situación, esto no quiere decir que obligatoriamente debamos de usar una técnica en específico para la situación que queramos analizar, cada técnica puede ser empleada en cualquier situación, su efectividad es la que determinara si es la adecuada o no.

2.1.1. Técnicas algebraicas y estadísticas

La estadística es la ciencia que puede dar respuesta a ambas necesidades: por una parte, la estadística descriptiva elabora técnicas para la representación y reducción de datos y, por otra, la estadística inductiva estudia la probabilidad de acierto de cada una de las opciones posibles. (Fernández Fernández, Cordero Sánchez, & Córdoba Largo, 2002, p. 18)

Los modelos para las técnicas algebraicas y estadísticas se forman mediante la utilización de funciones como la media aritmética, desviación estándar, correlación, regresión, entre otros.

2.2. Análisis de datos

Corresponde a una etapa de la ciencia de datos en donde se realiza el proceso de recopilación, limpieza y transformación de datos identificando al final la relación que se tiene y los impactos que generan.

El análisis de datos integra distintas operaciones en la que el investigador o analista somete ciertos datos, bien sea de orden cuantitativo o cualitativo, a una serie de análisis, lecturas e interpretaciones, según sea el enfoque de su investigación o requerimiento informativo. (Peña, 2017, p. 30)

2.2.1. Tipos de análisis de datos

Dependiendo del tipo de información a procesar existen dos técnicas las cuales podemos aplicar y que proveen técnicas para generar los modelos de análisis de datos, estas técnicas son: análisis cuantitativo y análisis cualitativo. Ambas técnicas aplicadas de manera correcta deben ser capaces de proveer un alto nivel de confiabilidad.

Ambos enfoques (cuantitativo y cualitativo) utilizan fases similares para su desarrollo.

- Llevar a cabo observación y evaluación de los fenómenos.
- Establecer las ideas que surgen de la observación y evaluación de los fenómenos.
- Fundamentar las ideas y las suposiciones.
- Revisar las suposiciones o ideas sobre la base del análisis.
- Esclarecer las ideas generadas mediante la proposición de nuevas evaluaciones y observaciones.

2.2.1.1. Análisis cuantitativo

Considerada como una técnica de análisis objetiva, ya que se centra en el manejo de los datos numéricos que caracterizan el proceso o fenómeno a analizar, lo cual es muy útil en la aplicación de las ciencias exactas y las ciencias naturales, en razón a la certeza de las conclusiones y leyes que se pueden obtener al finalizar el proceso. (Peña, 2017, p. 32)

Este tipo de análisis se emplea generalmente para crear modelos predictivos, normalmente aplicados a la inversión de bolsa, publicidad, crecimiento poblacional, gestión de riesgos y la salud.

2.2.1.2. Análisis cualitativo

Tiene como característica particular la relación de las variables del modelo, con criterios preestablecidos del analista, que depende en general de su experiencia en el campo de estudio, condición que tiene desventaja, que la solución propuesta no tiene un alcance universal. (Peña, 2017, p. 32)

La información que se suele analizar proviene de contenido distribuido a gran escala como, periódicos, revistas, radio, televisión y la más común, el internet.

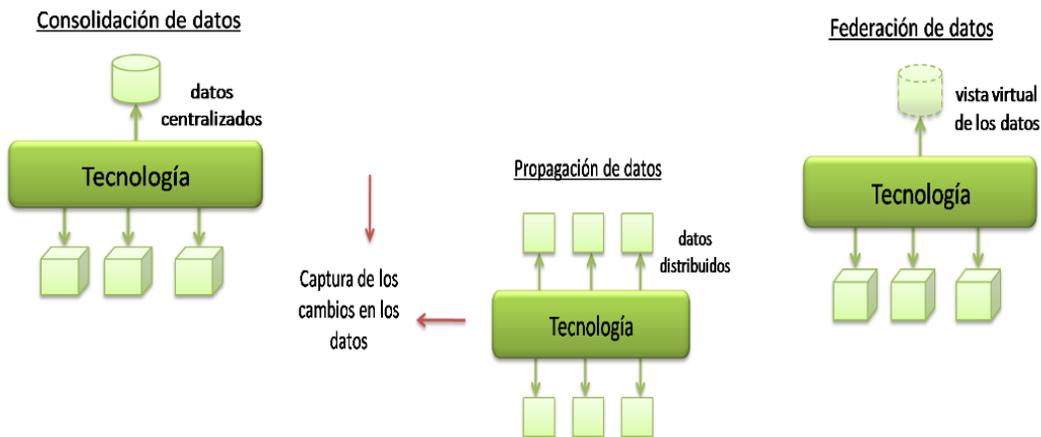
2.3. Procesos y técnicas de integración de datos

La información está distribuida en cientos de bases de datos y archivos y al proceso de unificar la información se le conoce como integración de datos. “Puede ser una vista física de datos que han sido recuperados de diferentes fuentes de información y consolidados en un repositorio, o puede ser una vista virtual que es construida dinámicamente, a la vez que se accede a los datos” (Oliva Alfonso, Pineda Alfonso, Kindelán Castro, & Carralero Iznaga, 2011, p. 6).

La integración de datos se puede realizar con diferentes técnicas o estrategias según sean los datos que se requieren centrar y de la ubicación de origen de cada uno de ellos.

Las principales técnicas utilizadas para la centralización de datos son tres: Consolidación de datos, Federación de datos y Propagación.

Figura 5. **Técnicas fundamentales de integración de datos**



Fuente: Revista Cubana de Ingeniería (2012). Propuesta de herramientas para la integración de datos. Consulta: 30 de septiembre de 2021. Recuperado de <https://rci.cujae.edu.cu/index.php/rci/article/view/65/pdf>.

2.3.1. **Consolidación de datos**

Es el proceso de capturar y combinar la información de diferentes fuentes de datos y crear un único repositorio centralizado. El fin de un almacén de datos central es poder generar reportes, analizar datos, y servir como origen de datos para aplicaciones.

El uso de esta técnica implica que exista un retraso, o latencia, en cuanto a la actualización de la información, pues como el proceso de obtención y almacenamiento ocurre periódicamente, puede que cierta información no esté del todo actualizada en un momento determinado. (Oliva Alfonso, Pineda Alfonso, Kindelán Castro, & Carralero Iznaga, 2011, p. 7)

2.3.2. Extraer transformar y cargar (ETL)

Es la denominación que se le da al proceso de extraer, transformar, y cargar (*Extract, Transform, Load*) datos a partir varias fuentes, para almacenarlo en un repositorio central para su posterior uso. En cuanto a las tareas de un ETL, (Bustamante Martínez, Galvis Lista, & Gómez Flórez, 2013) define que consiste en diseñar el flujo de trabajo que extraiga los datos desde las fuentes, repare sus inconsistencias, los transforme en un formato deseado, y, finalmente, los inserte en la bodega de datos.

El proceso de construcción de un ETL se puede dividir en tres componentes: extracción, transformación y carga. La siguiente tabla describe cada componente, las operaciones realizadas y los resultados esperados.

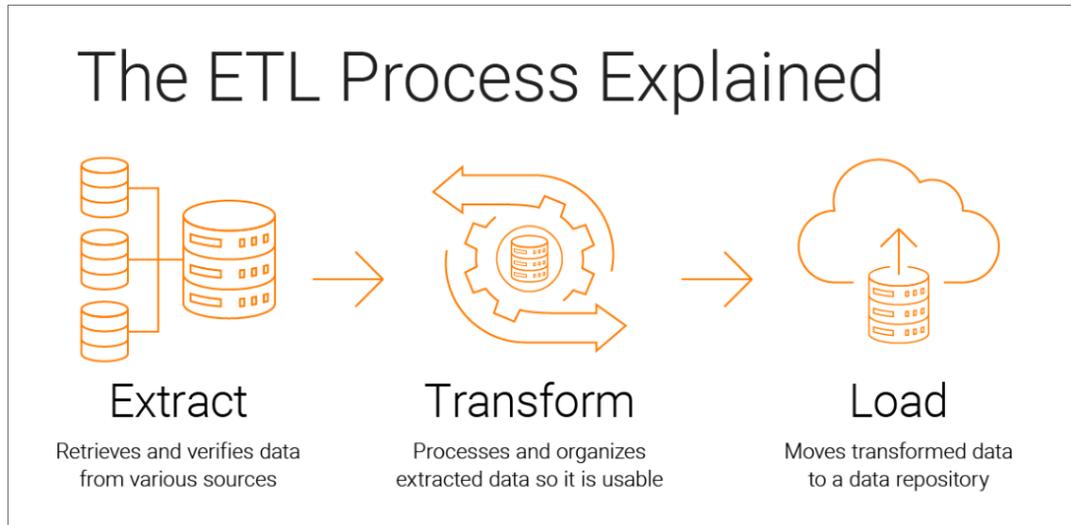
Tabla II. **Descripción de los componentes del proceso ETL**

Componente	Elementos Objetivo(entrada)	Operaciones realizadas (proceso)	Resultado de la tarea (salida)
Extracción	Fuente de datos, sistemas, transacciones, hojas de cálculo, archivos de texto.	Selección	Datos cargados en memoria
Transformación	Datos cargados en memoria	Limpieza, transformación, personalización, realización de cálculos y aplicación de funciones de agregación.	Datos formateados, estructurados, y resumidos de acuerdo con las necesidades (aún en memoria)
Carga	Datos estructurados, formateados y resumidos (aún en memoria)	inserción	Datos formateados, estructurados y resumidos con persistencia en los <i>data warehouse</i>

Fuente: elaboración propia, realizado con Microsoft Word.

La necesidad de integrar datos de manera sencilla creció con rapidez, (Solutions, 2021) explica que por esa razón el proceso ETL se convirtió en el método estándar para extraer datos de múltiples orígenes y transformarlos para luego insertarlos en una fuente destino.

Figura 6. **Representación general del flujo de los procesos de ETL**



Fuente: informatica (2023). *What is ETL (extract transform load)?* Consultado el 13 de abril de 2023. Recuperado de <https://www.informatica.com/resources/articles/what-is-etl.html>.

2.3.3. **Uso del lenguaje Python**

Cada vez se utiliza más por las empresas y programadores que trabajan con datos (*Business Intelligence*, Integración de datos, *Data Science*, *Machine Learning*, *Big Data*...). El motivo de que cada vez cobra más importancia en su uso es en la gran cantidad de librerías existentes para realizar prácticamente todo y más aún si el objetivo es trabajar y gestionar datos, también por lo optimizado que está Python respecto a JAVA. (Francisco Rodríguez, 2020, p. 1)

Las librerías que ofrece Python son versátiles, se tiene a disposición librerías que pueden ser usadas para cualquier aplicación y en cualquier momento. Desde funciones estadísticas hasta aplicaciones para *Deep learning*, Python ofrece librerías programadas que solo esperan ser implementadas para un uso en específico.

Ofrece un alto rendimiento. A diferencia de otros lenguajes como Java, el cual se tiene que ejecutar sobre una Java máquina virtual, una capa extra que agrega latencia.

Python utiliza librerías de alto rendimiento construidas en otros lenguajes. Por ejemplo, SciPy que usa LAPACK, el cual es una librería de Fortran para aplicaciones de algebra lineal. Tensorflow utiliza CUDA, el cual hace uso de una librería en C para el procesamiento del GPU en paralelo. (Mukhopadhyay, 2018, p. 2)

2.4. Importancia de la visualización de datos

Encontrar una representación adecuada para un conjunto de datos en específico es crucial para que el resultado del proceso de visualización será satisfactorio. Brindar las interacciones apropiadas proveerá de información más relevante al usuario.

El objetivo de una visualización es encontrar una metáfora visual que permita entender y percibir en forma efectiva un conjunto de datos. Una visualización debe proveer también un conjunto de interacciones a partir de las cuales el usuario explorará el conjunto de datos con una mínima carga cognitiva. (Castro, Larrea, Urribarri, Ganuza, & Escarza, 2018, p. 361)

Es más factible recordar imágenes con muchos patrones que datos sin procesar. Los gráficos son una forma simple de expresión que tiene más impacto a largo plazo ya que no requiere de un gran esfuerzo mental para analizarlo. La visualización de datos es una forma muy eficaz de transmitir información.

La visualización de datos no únicamente se basa en la elección de un gráfico adecuado, también se requiere de una forma de clasificar, comprender y explicar datos para que tengan sentido para las personas que van a tomar decisiones a partir de los datos presentados.

2.4.1. Tipos de visualizaciones

Para la presentación de los datos y un correcto análisis de estos se suelen utilizar dos tipos de gráficos, de representación y exploratorios. Las diferencias entre cada uno recaen sobre el uso y forma que se les da. Los gráficos de presentación son generalmente estadísticos, deben incluir definiciones y explicaciones de las variables involucradas. Los gráficos exploratorios se utilizan para visualizar resultados de forma rápida e informativa, la precisión no es su objetivo principal. La descripción de las variables involucradas es innecesaria en los gráficos exploratorios.

Los grupos en los cuales se dividen los tipos de visualización son los siguientes:

- Cambios en el tiempo: se utilizan cuando los datos utilizados para la investigación tienen un elemento del tiempo involucrado. Los cambios durante un periodo de tiempo específico.
- Identificación de relaciones: se utiliza para identificar la relación entre diferentes puntos de datos, su uso es valioso para la visualización de datos.
- Análisis de red: se utiliza para identificación de grupos mediante la relación indicada con líneas que conectan puntos.

- Determinación de la frecuencia: se utiliza para el análisis de la información en un periodo de tiempo, se suele relacionar con datos que presentan frecuencia en eventos.

2.4.1.1. Gráficos más comunes

Transformar la información en conocimiento requiere de una base sólida de información y sobre todo de una correcta utilización de los componentes visuales. Mostrar los datos de una analítica requiere de la elección correcta de tipos de gráficos. Las representaciones más utilizadas son las siguientes:

- Tablas. Son utilizadas para mostrar números de manera contundente, es una visualización que está diseñada para ofrecer una cantidad amplia de información en un solo componente.
- Gráfica de barra. Se utiliza para hacer énfasis en la comparación entre elementos dentro de un periodo de tiempo en específico.
- Gráfico Circular. Se utiliza para definir las proporciones y porcentajes de un conjunto de datos
- Gráfico de línea. Su principal función es mostrar tendencias de un conjunto de datos durante un periodo continuo de tiempo.
- Gráfico de dispersión. son útiles para mostrar la relación entre diferentes puntos de datos y mostrar cuanto afecta una variable a otra.
- Gráfico de barras. Se utiliza para resumir un conjunto de datos por categorías.

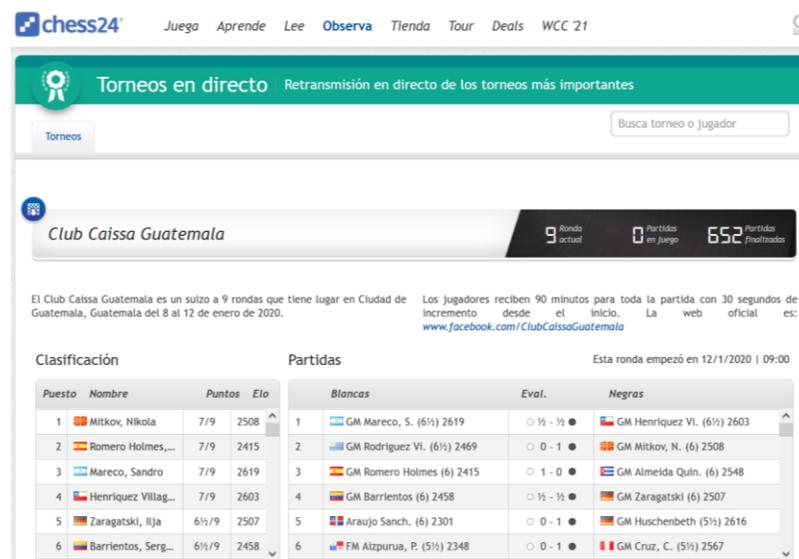
3. RECOPIACIÓN Y PREPARACIÓN DE DATOS

En esta sección se describe el proceso necesario para la obtención de los datos de las partidas de ajedrez y las transformaciones necesarias para tener el fichero con formato adecuado para la inserción de los datos al *Data Warehouse*.

3.1. Selección de partidas

La captura de los datos se hará desde la página de Chess24.com. Las partidas a analizar corresponden al torneo realizado por el club Caissa Guatemala, llevado a cabo del 18 al 12 de enero del 2020.

Figura 7. **Página web Chess24.com**



The screenshot shows the Chess24.com website interface. At the top, there is a navigation bar with links for 'Juega', 'Aprende', 'Lee', 'Observa', 'Tienda', 'Tour', 'Deals', and 'WCC 21'. Below this is a green banner for 'Torneos en directo' (Live Tournaments) with a search bar. The main content area features a header for 'Club Caissa Guatemala' with statistics: '9 Ronda actual', '0 Partidas en juego', and '652 Partidas finalizadas'. Below the header, there is a brief description of the tournament and a link to the club's Facebook page. The page is divided into two main sections: 'Clasificación' (Classification) and 'Partidas' (Matches). The 'Clasificación' table shows the ranking of players based on points and rating. The 'Partidas' table shows the results of individual matches, including the player names, ratings, and scores.

Puesto	Nombre	Puntos	Elo
1	Mitkov, Nikola	7/9	2508
2	Romero Holmes,...	7/9	2415
3	Mareco, Sandro	7/9	2619
4	Henriquez Villag...	7/9	2603
5	Zaragatski, Ilya	6½/9	2507
6	Barrientos, Serg...	6½/9	2458

Biancas	Eval.	Negras
GM Mareco, S. (6½) 2619	○ ½ - ½ ●	GM Henriquez Vl. (6½) 2603
GM Rodriguez Vl. (6½) 2469	○ 0 - 1 ●	GM Mitkov, N. (6) 2508
GM Romero Holmes (6) 2415	○ 1 - 0 ●	GM Almeida Quin. (6) 2548
GM Barrientos (6) 2458	○ ½ - ½ ●	GM Zaragatski (6) 2507
Araujo Sanch. (6) 2301	○ 0 - 1 ●	GM Huschenbeth (5½) 2616
FM Alzpurua, P. (5½) 2348	○ 0 - 1 ●	GM Cruz, C. (5½) 2567

Fuente: chess.com (2019) Club Caissa Guatemala. Consultado el 19 de octubre de 2022.

Recuperado de <https://chess24.com/en/embed-tournament/irt-open-2019>.

Se descargará un único archivo con formato PGN, que pone a disposición la página, el cual contiene un total de 652 partidas, cada una con la misma notación.

3.2. Conversión de PGN a JSON

Para insertar y procesar la información es necesario tener cada partida en formato JSON, por lo que se hará uso de un *script* que permita la obtención de un único archivo en formato JSON con todas las partidas.

Para realizar el *script* fue necesario hacer uso el lenguaje Python y de la librería Chess.pgn

3.2.1. Uso del intérprete de Python

Para el desarrollo del *script* es necesario descargar a instalar el lenguaje de programación interpretado Python en su versión 3.9.2. La descarga se puede realizar gratuitamente desde su página *web* oficial.

3.2.2. Librería Chess.pgn

Chess.pgn es una librería escrita para Python para el validación y generación de movimientos de ajedrez que soporta la lectura de archivos en formato PGN.

Utilizando la librería es posible realizar un análisis sintáctico del archivo de partidas y a partir de este formar la estructura JSON que se requiere para el análisis posterior.

3.2.3. Análisis y creación del archivo JSON

El siguiente proceso consta de construir el archivo JSON que contendrá los datos de todas las partidas a analizar, los pasos necesarios pueden verse en el apéndice 1, en el cual mediante un *script* escrito en lenguaje de programación Python y con la implementación de la librería Chess.pgn se realiza la extracción de los atributos y la creación de nuevos campos que serán insertados al JSON final.

Figura 8. Formato final de una partida en formato JSON

```
{
  "Event": "?",
  "Site": "Chess.com",
  "Date": "2021.07.28",
  "Round": "?",
  "White": "Willygt",
  "Black": "BladeRunner01",
  "Result": "1-0",
  "Moves": "1. d4 d5 2. c4 c6 3. Nf3 Nf6 4. Bg5 e6 5. e3 Be7 6.
  "Open": "d2d4",
  "EndMotive": "Abandono"
},
```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.

El archivo final está compuesto por un arreglo de partidas con el formato expuesto en la figura 8.

4. TRANSFORMACIÓN Y LIMPIEZA DEL ARCHIVO JSON

El archivo JSON obtenido del proceso anterior será cargado a una colección local de MongoDB esto con el objetivo facilitar la corrección y eliminación de los registros que no aporten un valor significativo al análisis final.

El proceso de transformación y limpieza de datos se realiza mediante la ejecución de un *script* en Python el cual contiene la conexión a la base de datos mongo y las consultas a la colección para eliminar, actualizar y modificar campos de los registros de cada partida.

4.1. Conexión a base de datos mongo

Los registros de las partidas serán almacenados y procesados en una instancia local de MongoDB. La instalación se hará mediante la configuración del asistente de instalación de Mongo DB en su versión *Community* 5.0 para el sistema operativo Windows.

Para realizar la comunicación del *script* hacia el almacén de datos se utiliza la librería Pymongo la cual es de nativa de Python para la conexión con bases de datos MongoDB. El proceso de su importación y funcionamiento se explica a detalle en el apéndice 2.

4.2. Estructura del *script* de carga y transformación

Mediante la implementación de funciones nativas del lenguaje Python y en conjunto con la librería de Pymongo, se realiza la lectura de las partidas en

formato JSON para insertarlas a la colección en la base de datos y luego realizar el análisis de los atributos que se requieren para la visualización de resultados final.

Para la carga de las partidas se hará uso de la librería json, nativa de Python para trabajar con datos en formato JSON.

Figura 9. **Carga el archivo JSON de las partidas**

```
with open('PGNJSON.json') as f:  
    file_data = json.load(f)
```

Fuente: elaboración propia, realizado con Visual Studio 1.77.0.

4.2.1. **Inserción del archivo JSON a la base de datos**

La razón por la cual hemos transformado las partidas a formato JSON es porque MongoDB trabaja con documentos los cuales son almacenados con el formato de clave-valor. Desde el *script* de Python la importación se realiza mediante la inserción de los datos por medio del método *insert_many* el cual inserta cada partida en la colección especificada.

Figura 10. **Método para la inserción del archivo JSON**

```
collection.insert_many(file_data)
```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.

Con la información ya carga en la base de datos se procede a realizar el procesamiento de los datos. En las siguientes secciones se decide que campos utilizar y cuales formaran parte de la limpieza de la información.

4.2.2. Eliminación de duplicados

Al momento de descargar las partidas en formato PGN de la página Chess21.com es posible que algunos registros de partidas estos duplicados, por lo cual debemos eliminar ese factor del análisis.

Una partida de ajedrez puede ser jugada múltiples veces por los mismos jugadores e incluso con el mismo color de piezas, sin embargo, los movimientos que surjan en cada partida son diferentes. El del campo que contiene los movimientos de cada partida en el archivo tiene como nombre *Moves*. El campo debe ser único, el procedimiento para verificarlo se encuentra detallado en el apéndice 3.

4.2.3. Eliminación de campos

Luego de eliminar los registros duplicados se realiza la toma de decisión de cuáles son los campos que no aportan un valor a la visualización de resultados finales.

Los atributos que contienen los registros insertados hasta al momento se pueden visualizar en la figura 5.

Los campos que son irrelevantes para el desarrollo y el motivo son los siguientes:

- *BlackElo*. La habilidad teórica del jugador con piezas negras es intrascendente para el objetivo del análisis. No es necesario conocer el *ranking* de cada jugador ya que se trata del análisis de un torneo en específico.
- *Date*. La fecha de la partida es despreciable debido a que la variable del tiempo no desempeña un papel relevante en el análisis.
- *Round*. El número de ronda del evento no proporciona información significativa ya que la observación de resultados se realiza al torneo en general.
- *WhiteElo*. La habilidad teórica del jugador con piezas blancas al igual que *BlackElo* es irrelevante.

El método para eliminar los campos descritos anteriormente se detalla en el apéndice 4.

4.2.4. Actualización de valores

El campo *Result* describe el ganador de la partida con formato de puntuación, para la realización de un análisis más sencillo se modificará la nomenclatura de las victorias.

El formato PGN descargado de la página Chess24.com describe los resultados de las partidas de la siguiente forma.

Tabla III. **Descripción de resultados de las partidas**

Nomenclatura	Descripción
1-0	Victoria para el jugador de piezas blancas
0-1	Victoria para el jugador de piezas negras
1/2-1/2	Empate o tablas

Fuente: elaboración propia, realizado con Microsoft Word.

La modificación de las nomenclaturas se realiza colocando la palabra *White* cuando el resultado sea 1-0, *Black*, cuando el resultado sea 0-1 y *Draw* cuando el resultado sea de tablas (1/2-1/2). El comando para modificación de los valores del campo *Result* se describe a detalle en el apéndice 5.

4.3. Obtención del archivo JSON procesado.

Concluida la transformación y limpieza de los datos, los registros almacenados en la base de datos ya contienen únicamente los campos necesarios que aportan un valor significativo al análisis final del torneo de ajedrez.

En este paso se realiza la extracción, a un archivo JSON, la información obtenida luego del procesamiento y transformación de los datos, el cual será importado en Power Bi para realizar la creación de los gráficos.

El procedimiento para la obtención del fichero se encuentra dentro de las instrucciones del *script* de Python utilizado anteriormente para la inserción de las partidas a MongoDB y para todo el procesamiento de datos. Su funcionamiento y explicación se encuentran a detallados en el apéndice 3.

Figura 11. Ejemplo de archivo JSON final

```
{
  "_id": {
    "$oid": "642fb590a77794c386a48054"
  },
  "Event": "?",
  "Site": "Chess.com",
  "White": "Francis210",
  "Black": "MasterPadawan1",
  "Result": "White",
  "Moves": "1. e4 g6 2. d4 Bg7 3. Nf3 d6 4. Nc3 Nf6 5. Bc4 O-O 6. O-O Nxe4 7.",
  "Open": "e2e4",
  "EndMotive": "Abandono"
},
```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.

5. ANÁLISIS Y REPRESENTACIÓN DE RESULTADOS

El preprocesamiento de datos es una etapa crítica en cualquier análisis de datos y, en el caso del análisis de partidas de ajedrez, no es diferente. En esta etapa, se realizarán una serie de pasos para preparar los datos para el análisis posterior.

Una vez completado el preprocesamiento de los datos, se procederá al análisis en sí. En este caso, se emplearán técnicas de estadística descriptiva para identificar patrones y tendencias en las partidas de ajedrez. Algunos ejemplos de estas técnicas incluyen la identificación de valores atípicos, la media, la mediana y la moda, la desviación estándar, entre otras.

Para presentar los resultados obtenidos, se utilizarán una variedad de gráficos y visualizaciones que permitirán una mejor comprensión de los patrones y tendencias identificados.

5.1. Preprocesamiento de datos

El archivo JSON obtenido del proceso de ETL en el capítulo anterior no basta para obtener análisis detallados de las partidas de ajedrez, En este capítulo se describirá en detalle todo el proceso necesario para preparar los datos para su representación posterior.

Se presentarán las técnicas más comunes utilizadas en el preprocesamiento de datos, así como las consideraciones y decisiones que se tomarán durante esta fase. También se describirán las herramientas y técnicas utilizadas para realizar el preprocesamiento de datos de manera eficiente.

5.1.1. Manipulación de datos con Python

Los análisis que se describirán a continuación requieren de un tratamiento previo de los datos más complejo para poder obtener un formato que permitan una representación más legible de los análisis que se realizaran.

- Análisis de patrones de jaque mate más comunes
- Análisis de porcentaje de éxito por apertura
- Análisis de patrones de enroque más utilizados
- Análisis aperturas más frecuentes

Cada uno de los análisis requieren una serie de tratamientos por separado que se describirá a continuación:

5.1.1.1. Patrones de jaque mate más comunes

Los patrones de jaque mate más comunes son las jugadas finales que resultan en la captura del rey del oponente y que se producen con mayor frecuencia en las partidas de ajedrez. Estos patrones pueden variar desde jaque mate con dama hasta jaque mate con alfil, pasando por jaque mate con caballo, entre otros.

Para poder extraer los patrones del archivo JSON generado en el proceso de ETL, se creó un *script* con los siguientes pasos para el preprocesamiento utilizando Python y las librerías Json y Panda.

- Cargar el archivo JSON a una estructura de tabla
- Inicializar la estructura que almacenara el conteo y descripción de los patrones encontrados
- Iterar cada partida
- Dividir los movimientos en una lista
- Iterar cada movimiento
- Verificar que el movimiento representa una jugada de jaque mate
- Extraer el movimiento
- Agregar el patrón a la estructura que almacenara el conteo
- Por último, exportar la estructura a un archivo CSV

Figura 12. **Script en Python para detectar patrones de jaque mate más comunes**

```
1 import json
2 import pandas as pd
3 import re
4 # Cargar el archivo JSON a un DATAFRAME
5 with open('../Mongo ETL/chess_records_etl_output.json', 'r') as f:
6     data = json.load(f)
7 df = pd.json_normalize(data)
8 #Inicializar diccionario para almacenar el conteo de los patrones de jaque mate
9 checkmate_counts = {}
10 # Iterar cada partida
11 for game in df['Moves']:
12     # Dividir movimientos en una lista
13     moves = game.split()
14     # Iterar cada movimiento
15     for i in range(len(moves)):
16         # Verificar que el movimiento es un jaque mate
17         if re.match(r'[KQRBN]\w*#[+#]?', moves[i]):
18             # Extraer el patron del movimiento
19             pattern = re.search(r'[KQRBN]\w*#[+#]?', moves[i]).group()
20             # Agregar el patron al diccionario e incrementar su aparicion
21             checkmate_counts[pattern] = checkmate_counts.get(pattern, 0) + 1
22 # crear un DataFrame del diccionario de jaques mates
23 checkmate_df = pd.DataFrame(list(checkmate_counts.items()), columns=['Checkmate Pattern', 'Count'])
24 # Ordenar DataFrame por apariciones en orden descendente
25 checkmate_df = checkmate_df.sort_values(['Count', ascending=False])
26 # Exportar DataFrame como archivo CSV
27 checkmate_df.to_csv('checkmate_patterns.csv', index=False)
```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.0.

El archivo CSV, que genera el *script* en Python, se utilizará posteriormente para representar los datos mediante un gráfico en Power Bi.

5.1.1.2. Porcentaje de éxito por apertura

El análisis del porcentaje de éxito por apertura es una técnica común utilizada en la estrategia del ajedrez para evaluar la efectividad de diferentes aperturas. Esta técnica implica analizar las partidas de ajedrez y registrar el número de veces que se ha utilizado una apertura específica, así como el porcentaje de veces que se ha ganado o perdido al usar dicha apertura.

Para obtener los porcentajes de éxito por apertura del JSON generado en el proceso de ETL, se creó un *script* con los siguientes pasos para el preprocesamiento utilizando Python y las librerías Json.

- Cargar el Archivo JSON Python
- Crear la estructura que almacenara los resultados de movimientos de las aperturas y el conteo de victorias
- Iterar cada partida del archivo
- Dividir los movimientos de cada partida en una lista
- Obtener los primeros 3 movimientos de la lista
- Agregar a la estructura de movimientos la apertura si no existe
- Actualizar el contador de victorias basado en el resultado de la partida
- calcular el porcentaje de victorias para cada apertura almacenada en la lista de resultado
- Exportar los resultados en un archivo CSV.

Figura 13. **Script en Python para detectar porcentaje de éxito en apertura, sección 1**

```
1 import json
2 import csv
3 # Cargar el archivo JSON
4 with open('../..//Mongo ETL/chess_records_etl_output.json') as f:
5     data = json.load(f)
6 # Crear un diccionario para almacenar los movimientos de las aperturas y el conteo de victorias
7 openings = {}
8 # Iterar cada partida en los datos
9 for game in data:
10     moves = game['Moves'].split() # Dividir los movimientos en una lista
11     opening = ' '.join(moves[:3]) # Obtener los primeros 3 movimientos
12     result = game['Result']
13     # Agregar el movimiento si no esta creado e inicializar los valores
14     if opening not in openings:
15         openings[opening] = {'wins': 0, 'losses': 0, 'draws': 0}
16     # Actualizar el contador de victorias basado en el resultado de la partida
17     if result == 'White':
18         openings[opening]['wins'] += 1
19     elif result == '0-1':
20         openings[opening]['Black'] += 1
21     else:
22         openings[opening]['draws'] += 1
```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.

Figura 14. **Script en Python para detectar porcentaje de éxito en apertura, sección 2**

```
23 # Calcular el porcentaje de victorias para cada apertura almacenada en la lista de resultados
24 results = []
25 for opening, counts in openings.items():
26     total_games = counts['wins'] + counts['losses'] + counts['draws']
27     win_percent = round(counts['wins'] / total_games * 100, 2)
28     results.append({'opening': opening, 'win_percent': win_percent, 'Total_Games' : total_games})
29 # Ordenar los resultados por porcentaje de victorias en orden descendente
30 results.sort(key=lambda x: x['win_percent'], reverse=True)
31 # Imprimir los resultados en consola
32 for result in results:
33     print(f"{result['opening']}: {result['win_percent']}% {result['Total_Games']}")
34 # Guardar los resultados en un archivo CSV
35 with open('opening_success.csv', mode='w', newline='') as f:
36     fieldnames = ['opening', 'win_percent', 'games']
37     writer = csv.DictWriter(f, fieldnames=fieldnames)
38     writer.writeheader()
39     for result in results:
40         writer.writerow({'opening': result['opening'],
41                         'win_percent': result['win_percent'],
42                         'games': result['Total_Games']})
43
```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.

El archivo CSV, que genera el *script* en Python, se utilizará posteriormente para representar los datos mediante un gráfico en Power Bi.

5.1.1.3. Patrones de enroque más utilizados

Los patrones de enroque son movimientos estratégicos utilizados en el juego de ajedrez para cambiar la posición de las piezas y fortalecer la defensa del rey. Los enroques corto y largo son los patrones de enroque más comunes y utilizados en el juego de ajedrez. El enroque corto se realiza moviendo el rey dos casillas hacia la torre en el flanco del rey y la torre se mueve hacia la casilla contigua al rey en el lado opuesto. Mientras que el enroque largo se realiza moviendo el rey dos casillas hacia la torre en el flanco de la reina y la torre se mueve hacia la casilla contigua al rey en el lado opuesto.

Para poder obtener los patrones de enroque más utilizados del JSON generado en el proceso de ETL, se creó un *script* con los siguientes pasos para el preprocesamiento utilizando Python y las librerías Json.

- Cargar el archivo JSON a Python
- Inicializar la estructura que almacenara los resultados
- Definir la expresión regular que coincida con los patrones de enroque
- Contar el número de veces que el patrón es utilizado
- Guardar el conteo de enroques en un archivo CSV

Figura 15. **Script en Python para detectar patrones de enroques más utilizados**

```
1 import json
2 import re
3 from collections import defaultdict
4 import csv
5 # Cargar el archivo JSON
6 with open('.././Mongo ETL/chess_records_etl_output.json') as f:
7     data = json.load(f)
8 # Definir la expresion regular que coincida con los patrones
9 castling_regex = re.compile(r'(O-O(-O)?|O-O(-O)?)')
10 # Contar el numero de veces que el patron es utilizado
11 castling_counts = defaultdict(int)
12 for game in data:
13     moves = game['Moves']
14     castlings = castling_regex.findall(moves)
15     for castling in castlings:
16         if castling[0] == 'O-O':
17             castling_counts['kingside'] += 1
18         elif castling[0] == 'O-O-O':
19             castling_counts['queenside'] += 1
20         else:
21             castling_counts['other'] += 1
22
23 # Guardar el conteo de enroques en un archivo CSV
24 with open('castling_counts.csv', 'w', newline='') as f:
25     writer = csv.writer(f)
26     writer.writerow(['Castling Type', 'Count'])
27     for castling_type, count in castling_counts.items():
28         writer.writerow([castling_type, count])
29
```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.

El archivo CSV, que genera el *script* en Python, se utilizará posteriormente para representar los datos mediante un gráfico en Power Bi.

5.1.1.4. Aperturas con mayor porcentaje de éxito

En el ajedrez, las aperturas son los movimientos iniciales que se realizan en la partida para desarrollar las piezas y establecer una posición sólida en el tablero. Hay muchas aperturas diferentes en el ajedrez, pero algunas son más frecuentes y populares que otras.

Para poder obtener los patrones de enroque más utilizados del JSON generado en el proceso de ETL, se creó un *script* con los siguientes pasos para el preprocesamiento utilizando Python y las librerías JSON y Chess.pgn.

- Cargar el archivo JSON a Python
- Inicializar la estructura que almacenara las aperturas
- Iterar cada partida
- Con la librería Chess.pgn crear un tipo de objeto que mapea la apertura en un tablero digital
- Extraer los movimientos de la partida y remover el indicar de numero de movimiento
- Extraer la posición del movimiento en el tablero y almacenarla en la estructura de partida
- Incrementar el contador que almacena la frecuencia con la que la apertura se utilizar
- Ordenar los movimientos más frecuentes en orden descendente
- Iterar las aperturas obtenidas
- Obtener el tablero del arreglo de movimientos de la apertura
- Convertir el tablero en notación SVG

- Crear un archivo SVG para cada apertura dentro de las 10 más frecuentes

Figura 16. **Script en Python para detectar aperturas más frecuentes, sección 1**

```

1  import json
2  import chess
3  import chess.svg
4  # Cargar el archivo JSON
5  with open('../Mongo ETL/chess_records_etl_output.json', 'r') as f:
6      data = json.load(f)
7  #Inicializar un diccionario que almacenara la frecuencia de cada apertura
8  opening_moves = {}
9  # Iterar cada partida
10 for game in data:
11     # Crear un nuevo objeto de tipo tablero de ajedrez
12     board = chess.Board()
13     # Extraer los movimientos de cada juego y remover el indicador de numero de movimiento
14     moves = [move for move in game['Moves'].split() if '.' not in move]
15     # Iterar cada movimiento de inicio de la partida y agregarla al tablero
16     for move in moves[:10]:
17         board.push_san(move)
18     # Extraer la posicion actual del tablero
19     fen = board.fen().split()[0]
20     # Incrementar el contador de la actual posicion en el tablero
21     opening_moves[fen] = opening_moves.get(fen, 0) + 1

```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.

Figura 17. **Script en Python para detectar aperturas más frecuentes, sección 2**

```

22 # Ordenar los movimientos mas frecuentes en orden descendente
23 for i, (fen, count) in enumerate(sorted(opening_moves.items(), key=lambda x: x[1], reverse=True)[:10]):
24     print(f'{fen}: {count}')
25     #Obtener el tablero del arreglo de movimientos
26     board = chess.Board(fen)
27     # Convertir el tablero en notacion SVG
28     svg_board = chess.svg.board(board=board)
29     file_name = 'Board_opening_' + str(i) + '.svg'
30     # Crear archivo SVG de la apertura
31     with open(file_name, 'w') as f:
32         f.write(svg_board)

```

Fuente: elaboración propia, realizado con Visual Studio Code 1.77.0.

El *script* en Python genera una imagen en SVG del tablero por cada apertura en el 10 más frecuentes.

5.1.2. Manipulación de datos con Power Bi

Mediante la herramienta de Power Query de Power Bi se pueden realizar transformaciones y cálculos personalizados que permiten a los usuarios analizar y comprender mejor los datos.

Los análisis que se describirán a continuación requieren de transformaciones previas para una representación más legible de los análisis que se realizaran.

- Análisis de porcentaje de victorias por color
- Análisis de captura de piezas

Cada uno de los análisis requieren una serie de transformaciones por separado que se describirá a continuación:

5.1.2.1. Análisis de porcentaje de victorias por color

El análisis de porcentaje de victorias por color en el ajedrez es un enfoque estadístico que evalúa el rendimiento de los jugadores en función del color de sus piezas. Debido a que el jugador que tiene el primer movimiento en una partida de ajedrez tiene una pequeña ventaja, es importante considerar el porcentaje de victorias por color para determinar si un jugador es más fuerte con las piezas blancas o negras.

Para realizar el análisis utilizando Power BI se deben seguir los siguientes pasos:

- Carga el archivo JSON de datos en Power BI. Ir a la pestaña Inicio y seleccionar Obtener datos. En la ventana emergente, elegir Archivos y seleccionar el archivo JSON de datos en formato PGN.
- En la siguiente ventana, elegir la opción JSON como formato de archivo y Cargar para importar los datos en Power BI.
- Una vez que los datos estén cargados, Power BI te abrirá la vista de Editor de consultas. Aquí, se realizarán las transformaciones necesarias en los datos antes de crear la visualización.
- En el editor de consultas, en la pestaña agregar columna, seleccionar columna condicional.
- Crear una condición donde basado el resultado de la partida crear una nueva columna con el color ganador.
- Cerrar y aplicar los cambios.

Figura 18. **Paso para agregar condiciones a nueva columna en Power BI**

Agregar una columna condicional

Agregue una columna condicional que se calcula a partir de las otras columnas o valores.

Nuevo nombre de columna
Winner

	Nombre de columna	Operador	Valor	Salida
Si	Result	es igual a	1-0	Enton... White
O si	Result	es igual a	0-1	Enton... Black

Agregar cláusula

De lo contrario
Draw

Aceptar Cancelar

Fuente: elaboración propia, realizado con Microsoft Power BI.

Terminada la transformación, esto nos genera una nueva tabla con la columna que será utilizada para representar los resultados.

5.1.2.2. Análisis de captura de piezas

El análisis de captura de piezas es una técnica comúnmente utilizada en el ajedrez para evaluar la fuerza relativa de las piezas en el tablero.

Para realizar el análisis utilizando Power BI se deben seguir los siguientes pasos:

- Carga el archivo JSON de datos en Power BI. Ir a la pestaña Inicio y seleccionar Obtener datos. En la ventana emergente, elegir Archivos y seleccionar el archivo JSON de datos en formato PGN.

- En la siguiente ventana, elegir la opción JSON como formato de archivo y Cargar para importar los datos en Power BI.
- Una vez que los datos estén cargados, Power BI te abrirá la vista de Editor de consultas. Aquí, se realizarán las transformaciones necesarias en los datos antes de crear la visualización.
- En la pestaña Transformar seleccionar la opción dividir columna por delimitador, y elegir las opciones para dividir en cada aparición del delimitador, dividir en filas para el delimitador espacio.
- Agregar una columna condicional para filtrar los movimientos que no contienen el valor x (indicador de captura) y colocarlos como nulos.
- Filtrar filas con el valor nulo.
- En la pestaña agregar columna, seleccionar la opción extraer texto antes del delimitador utilizando el carácter x como delimitador.
- Agregar una nueva columna extrayendo el primer carácter de cada fila, el cual indica la pieza.
- Utilizar la función de reemplazar para cambiar los valores indicados en la tabla II.
- Clic derecho sobre la columna con los reemplazos realizados y seleccionar la opción agrupar por.
- Agrupar por la nueva columna utilizando la operación de recuento de filas.
- Guardar los cambios y cerrar el editor.

Tabla IV. **Representación remplazos de piezas**

Valor que buscar	Remplazar con
K	<i>King</i>
Q	<i>Queen</i>
R	<i>Rook</i>
B	<i>Bishoop</i>
N	<i>Knight</i>
a, b,c,d,e,f,g,h	<i>Pawn</i>

Fuente: elaboración propia, realizado con Microsoft Word.

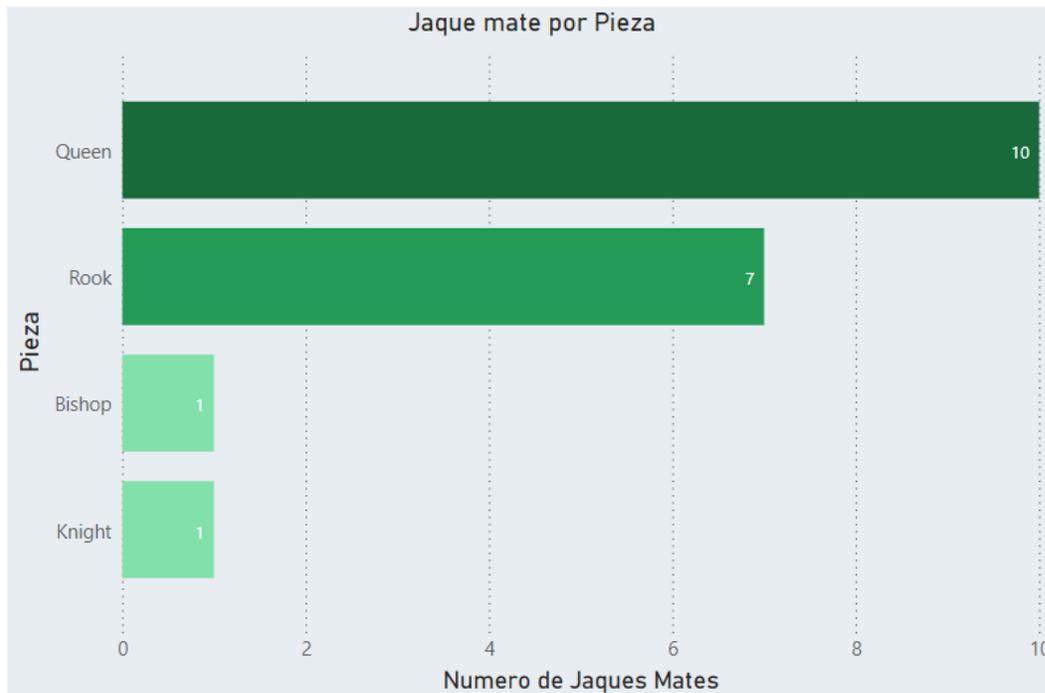
Terminada la transformación, esto nos genera una nueva tabla con la columna que será utilizada para representar los resultados.

5.2. Identificación de las piezas involucradas en jaque mate más habituales

Identificar las piezas que suelen estar involucradas en los patrones de jaque mate más habituales es una parte fundamental del estudio del ajedrez. Al conocer cuáles son las piezas más importantes en estas combinaciones tácticas, los jugadores pueden prepararse mejor para su uso en partidas reales. Por ejemplo, en muchas situaciones de jaque mate, el caballo y el alfil pueden ser piezas cruciales para cortar las líneas de escape del rey enemigo.

El siguiente gráfico se ha obtenido mediante el preprocesamiento mencionado anteriormente.

Figura 19. **Gráfico Análisis piezas involucradas en jaque mate**



Fuente: elaboración propia, realizado con Microsoft Power Bi.

En general, se identificaron tres piezas principales que suelen estar involucradas en los patrones de jaque mate más habituales analizados: la dama, la torre y el alfil.

La dama es una pieza muy poderosa que puede moverse en cualquier dirección y es capaz de controlar varias casillas a la vez. Una de las formas más comunes en que la dama se utiliza en los patrones de jaque mate es como parte de una combinación con otras piezas, como el alfil o el caballo. La dama puede ser utilizada para controlar las diagonales y las filas del tablero, mientras que el alfil y el caballo pueden ser utilizados para cortar las líneas de escape del rey enemigo.

La torre es una pieza que a menudo se utiliza en combinación con otras para llevar a cabo jaques mate efectivos, la torre se utiliza para controlar las columnas que conducen al rey enemigo, impidiendo que éste escape y facilitando su captura.

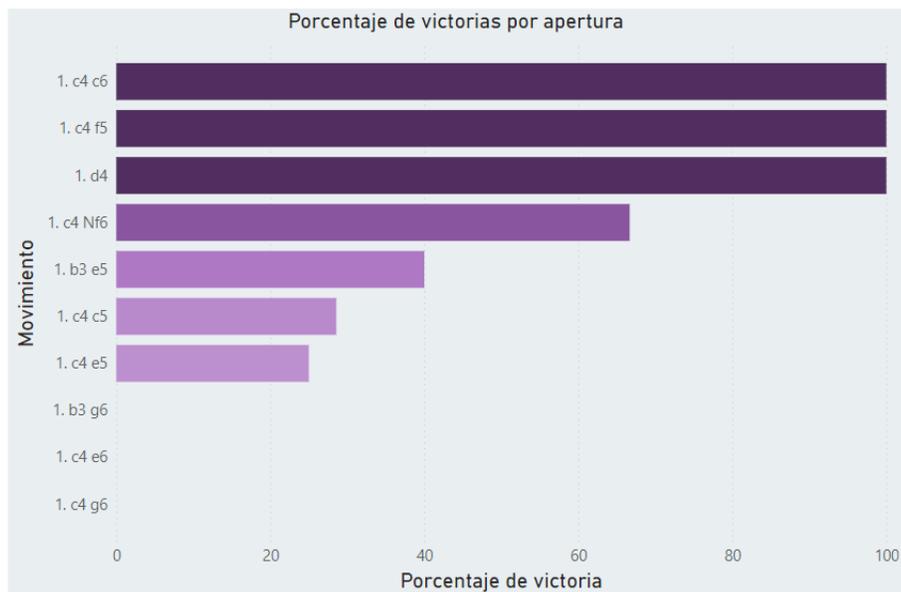
El alfil, por último, es una pieza que puede controlar diagonales enteras y es especialmente efectiva en combinación con otras piezas.

5.3. Evaluación de la efectividad de las aperturas en el ajedrez mediante análisis de porcentaje de éxito

El análisis de porcentaje de éxito por apertura puede ser una herramienta muy útil para los jugadores de ajedrez que buscan mejorar su juego. Al evaluar la efectividad de las aperturas, los jugadores pueden tomar decisiones informadas sobre qué aperturas utilizar en sus partidas, y adaptar su estrategia según los resultados de cada una de ellas. Es importante tener en cuenta que el análisis de porcentaje de éxito no es una ciencia exacta y los resultados pueden variar dependiendo del nivel de los jugadores y las circunstancias específicas de cada partida.

La siguiente gráfica toma en consideración la cantidad de partidas analizadas el archivo JSON y obtiene el porcentaje de victoria de cada una de ellas.

Figura 20. **Gráfico Análisis porcentaje de victorias por apertura**



Fuente: elaboración propia, realizado con Microsoft Power Bi.

Es importante tener en cuenta que el análisis de porcentaje de éxito por apertura no es una herramienta mágica que garantice la victoria en todas las partidas. Los jugadores deben utilizar el análisis de manera inteligente y estratégica, teniendo en cuenta sus propias fortalezas y debilidades, y adaptando su juego según las circunstancias específicas de cada partida.

Al analizar los resultados del top 10 de aperturas, se puede observar que la apertura más efectiva es c4 c6 (también conocida como apertura inglesa), con una tasa de éxito del 100 %. Es una apertura abierta que se caracteriza por el avance del peón de c4 en el primer movimiento de las blancas, y la respuesta del peón de c6 por parte de las negras. Esta apertura suele conducir a posiciones cerradas y estratégicas, en las que ambas partes intentan controlar el centro del tablero y desarrollar sus piezas de manera armoniosa.

La segunda apertura más efectiva, que también cuenta con un porcentaje del 100 % es la que cuenta con los movimientos c4 f5 (también conocida como Apertura Bird). Es una variante de la Apertura Inglesa, en la que las negras juegan un avance temprano del peón de f5, buscando desequilibrar la posición y presionar a las blancas desde el principio. Esta apertura puede llevar a posiciones abiertas y tácticas, en las que ambas partes deben estar alerta para evitar posibles trampas y amenazas tácticas.

Por su parte la tercera apertura, con una eficacia completa, está representada por el movimiento d4 (también conocida como Apertura de Peón de Dama). Es una apertura abierta en la que las blancas avanzan el peón de d4 dos casillas, controlando el centro del tablero y preparándose para desarrollar sus piezas. La Apertura de Peón de Dama puede llevar a diversas variantes, incluyendo la Defensa India de Rey, la Defensa Nimzo-India, la Defensa Grunfeld, entre otras. Esta apertura es muy popular y se ha utilizado en partidas de ajedrez de alto nivel durante décadas.

También es interesante notar que algunas aperturas que suelen ser muy populares entre los jugadores, como c4 nf6 (también conocida como Defensa Anglo-India) tienen una tasa de éxito relativamente baja.

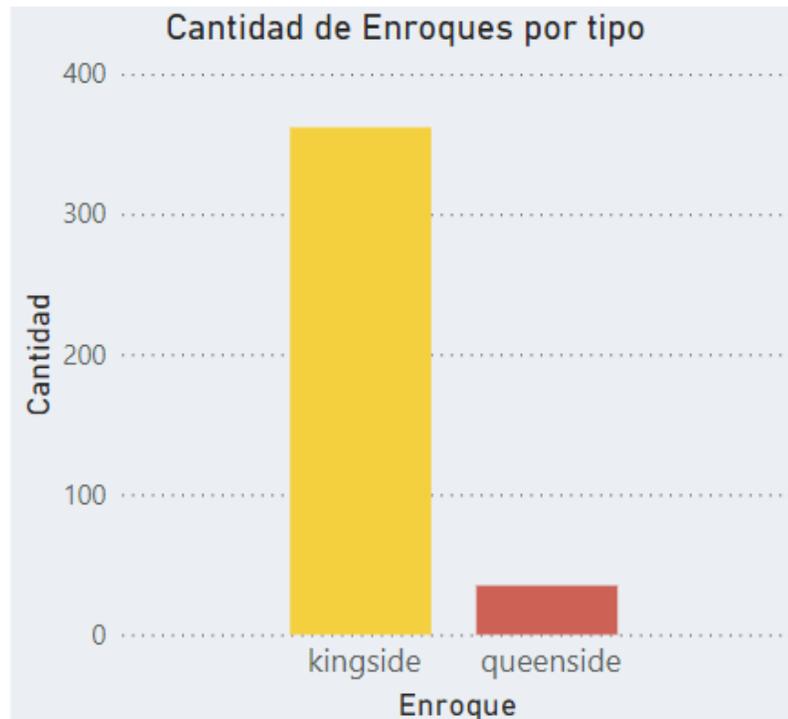
La apertura Defensa Anglo-India es una apertura abierta en la que las blancas avanzan el peón de c4, controlando el centro y preparándose para desarrollar sus piezas, y las negras responden con el caballo en f6, apuntando a e4 y controlando el centro de manera indirecta. Esta apertura puede llevar a posiciones cerradas y estratégicas, en las que ambas partes luchan por el control del centro del tablero y buscan desarrollar sus piezas de manera armoniosa. La Defensa Anglo-India es una de las respuestas más populares a la Apertura Inglesa, y ha sido utilizada en partidas de ajedrez de alto nivel durante décadas.

Esto demuestra la importancia de no confiar ciegamente en aperturas populares, y en cambio, utilizar el análisis de porcentaje de éxito para evaluar la efectividad de cada apertura de manera individual.

5.4. Evaluación de las estrategias de enroque más utilizadas en el ajedrez

El enroque es una de las jugadas más importantes en el ajedrez, y puede ser fundamental para el éxito de un jugador. A través del enroque, el rey se mueve hacia una posición más segura y se protege detrás de una línea de peones, lo que lo hace menos vulnerable a los ataques del oponente. Sin embargo, aunque el enroque puede ser una jugada muy efectiva, es importante recordar que no es la única estrategia disponible, y su éxito puede depender en gran medida de la situación específica del juego.

Figura 21. **Gráfico análisis cantidad de Enroques por tipo**



Fuente: elaboración propia, realizado con Microsoft Power Bi.

De acuerdo con la gráfica anterior el enroque del rey parece ser la estrategia de enroque más utilizada en el ajedrez en comparación con el enroque de la reina. Esto se puede ver claramente en el gráfico proporcionado, que muestra que el enroque del rey se utiliza con mayor frecuencia que el enroque de la reina.

Si bien el enroque del rey puede ser una jugada efectiva en muchas situaciones, es importante tener en cuenta que el enroque de la reina también puede ser una estrategia muy efectiva en ciertas situaciones. Por ejemplo, en algunas aperturas de ajedrez, el enroque de la reina puede ser preferible debido a la posición de las piezas del oponente. Por lo tanto, aunque el enroque del rey

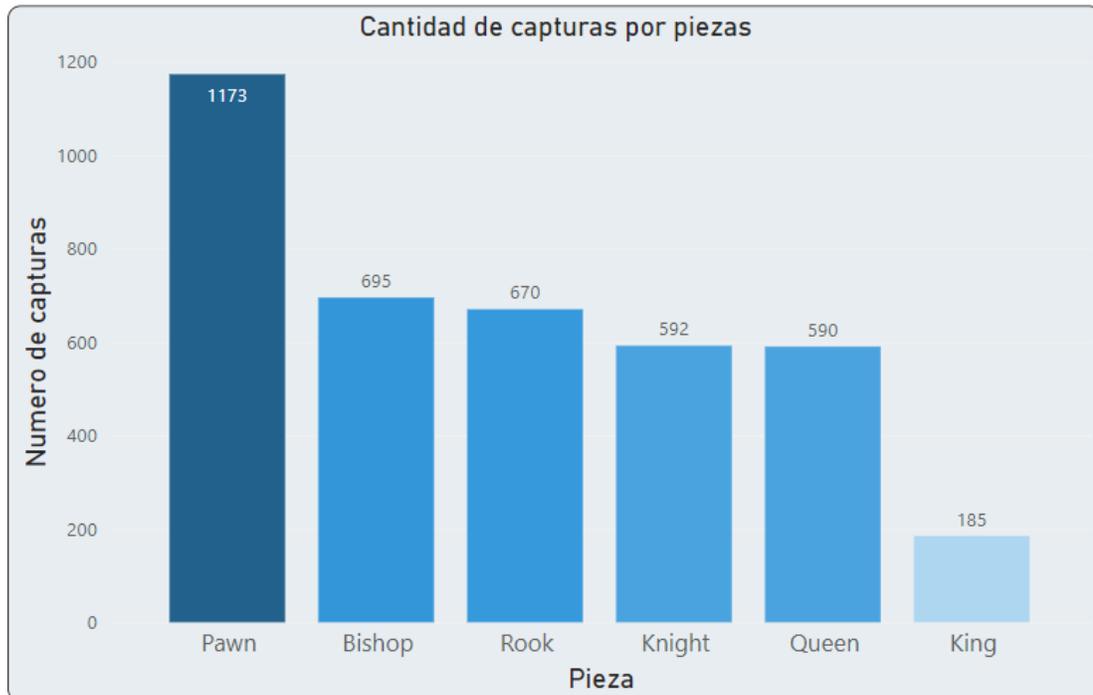
es una jugada común en el ajedrez, es importante no descartar otras estrategias de enroque, como el enroque de la reina.

Cada situación es única y puede requerir un enfoque diferente, por lo que es esencial que los jugadores comprendan las ventajas y desventajas de cada estrategia de enroque y puedan tomar decisiones informadas sobre qué estrategia utilizar en una situación determinada.

5.5. Análisis de las piezas más agresivas en ajedrez

Una de las habilidades más importantes en el ajedrez es la capacidad de capturar las piezas del oponente. Por lo tanto, el análisis de las piezas más agresivas en el ajedrez es una tarea crucial para cualquier jugador que quiera mejorar su juego.

Figura 22. **Gráfico análisis cantidad de capturas por piezas**



Fuente: elaboración propia, realizado con Microsoft Power Bi.

En este caso, se puede observar que el peón es la pieza que más captura en el juego, seguido por el alfil, la torre, el caballo, la reina y el rey, respectivamente.

El hecho de que el peón sea la pieza que más captura puede ser sorprendente para algunos jugadores, ya que los peones suelen considerarse piezas más defensivas en el juego. Sin embargo, los peones tienen un papel importante en el juego y pueden ser muy efectivos para controlar el centro del tablero y atacar las piezas del oponente. Además, los peones son las únicas piezas que pueden capturar de manera diagonal, lo que les da una mayor flexibilidad en el tablero.

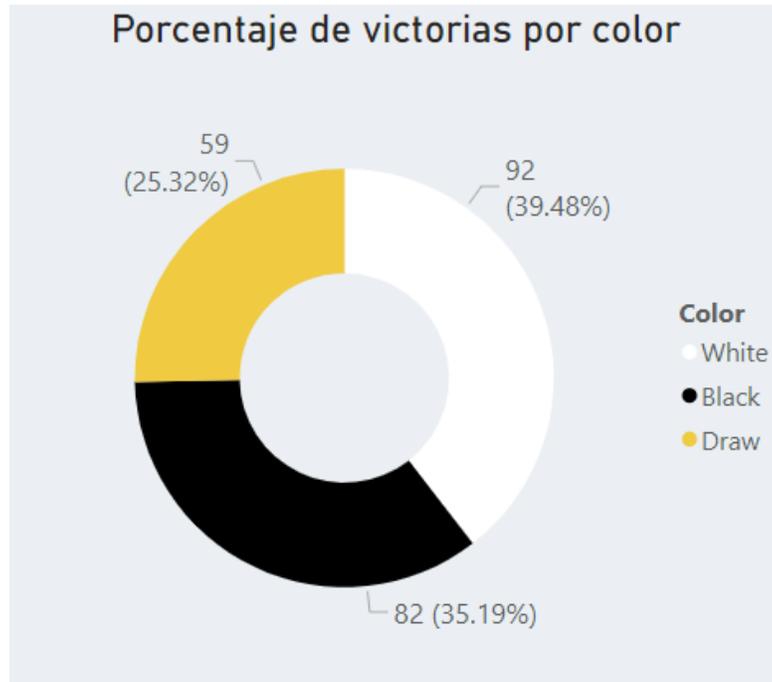
El alfil y la torre son piezas muy poderosas y su posición en la lista de piezas que más capturan en el juego no es sorprendente. Ambos tienen un alcance significativo en el tablero y pueden ser muy efectivos en la captura de piezas. El caballo, por otro lado, tiene un movimiento único y puede saltar sobre otras piezas, lo que lo hace muy efectivo para atacar las piezas del oponente. Sin embargo, su alcance en el tablero es más limitado en comparación con el alfil y la torre.

Finalmente, la reina y el rey son las piezas que menos capturan en el juego. La razón por la que la reina captura menos piezas que el resto de las piezas en la lista puede ser porque se utiliza más para proteger al rey y atacar al oponente de forma indirecta. El rey, por otro lado, es una pieza muy importante que debe ser protegida en todo momento y, por lo tanto, no se utiliza con frecuencia para capturar otras piezas.

5.6. Análisis de la distribución de victorias por color

El análisis de la distribución de victorias por color en el ajedrez es una herramienta importante para comprender la influencia del color en los resultados de las partidas. Se trata de un enfoque que permite evaluar la proporción de victorias obtenidas por cada color y determinar si hay alguna correlación entre la elección del color y la probabilidad de ganar un juego.

Figura 23. **Gráfico análisis porcentaje de victorias por color**



Fuente: elaboración propia, realizado con Microsoft Power Bi.

En el gráfico se puede observar que las blancas ganan una proporción ligeramente mayor de partidas que las negras, con un porcentaje del 39.8 % para las blancas y del 35.19 % para las negras. Este resultado sugiere que jugar con las piezas blancas podría ofrecer una pequeña ventaja en términos de la probabilidad de ganar una partida.

Sin embargo, también es importante tener en cuenta que el porcentaje de empates es relativamente alto, con un 25.32 % de las partidas terminando en empate. Esto sugiere que el color de las piezas puede no ser un factor determinante en el resultado de la partida, y que otros factores, como la habilidad de los jugadores y la estrategia utilizada, pueden ser más importantes.

Es posible que las blancas tengan una pequeña ventaja debido a la iniciativa que obtienen en la apertura, lo que les permite establecer el tono de la partida y dictar el ritmo del juego. Sin embargo, esta ventaja puede ser contrarrestada por una buena preparación por parte de los jugadores negros y una sólida defensa.

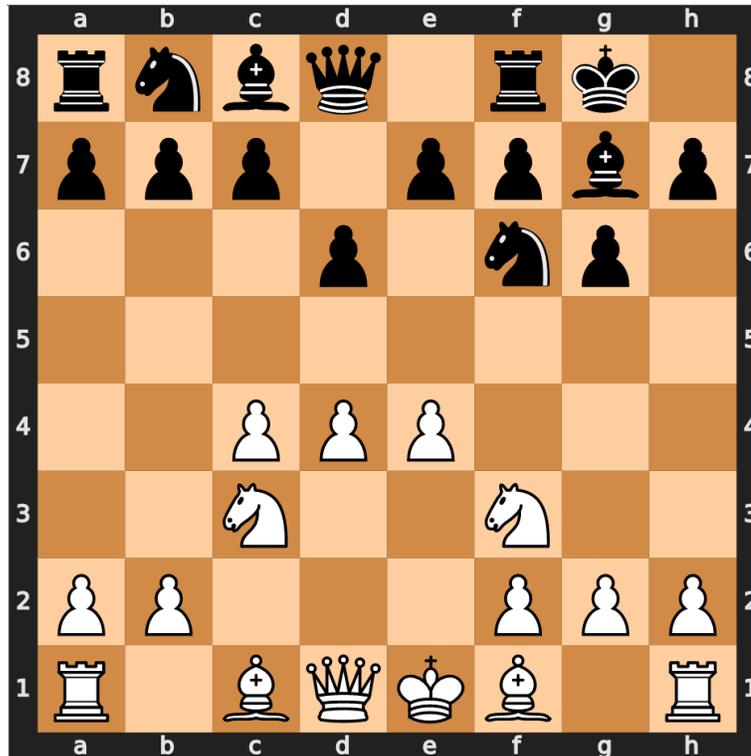
5.7. Análisis de aperturas más frecuentes

El análisis de las aperturas más frecuentes en el ajedrez puede ser de gran ayuda para mejorar el juego y preparar estrategias más efectivas. Identificar las aperturas que se utilizan con mayor frecuencia en las partidas de ajedrez permite a los jugadores conocer las tendencias actuales y adaptar su juego a las mismas. Además, al conocer las aperturas más utilizadas, se pueden diseñar contra estrategias para contrarrestarlas y así obtener ventaja en la partida.

Entender las aperturas más populares y su frecuencia de uso puede proporcionar una mayor comprensión del juego y permitir un seguimiento más activo de las partidas. Además, conocer las aperturas más populares puede ser de gran ayuda para seguir y analizar partidas de jugadores profesionales, ya que muchos de ellos tienen patrones de juego bien definidos en sus aperturas favoritas.

Mediante las imágenes en formato SVG generadas por el *script* de análisis de partidas más frecuentes, se obtuvieron las siguientes tres aperturas más frecuentes:

Figura 24. Primera apertura más frecuente



Fuente: elaboración propia, realizado con Python.

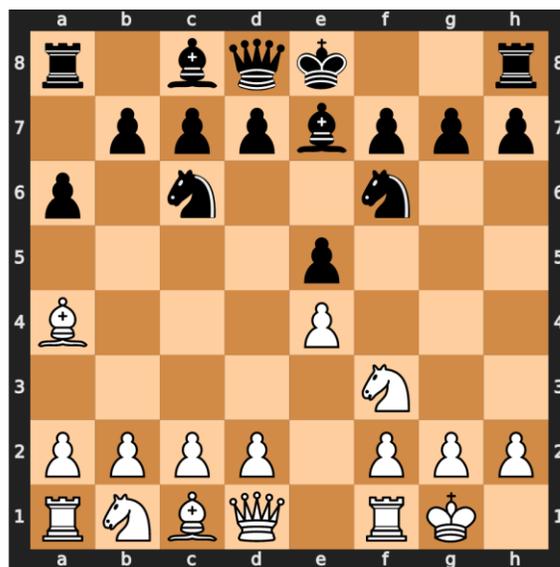
La figura 24 representa la primera partida más jugada con un número de frecuencia de tres veces. No se puede decir que esta posición represente una apertura específica, ya que es el punto de partida común para todas las aperturas en el ajedrez.

En esta posición, ambos bandos han desarrollado sus caballos hacia el centro del tablero y han avanzado sus peones centrales. El alfil de casillas blancas está en una buena posición, ya que controla la diagonal blanca que apunta hacia el flanco de rey negro. El alfil de casillas negras también está listo para entrar en juego, aunque su camino está bloqueado momentáneamente por

el peón de c7. El rey blanco está seguro en el flanco de rey, mientras que el rey negro se encuentra en una posición similar en el flanco de dama.

En esta posición, hay varias posibles jugadas que ambos jugadores podrían considerar. Algunas opciones populares para las blancas incluyen avanzar el peón de d4 para controlar más espacio en el centro del tablero, mover el caballo de c3 para liberar al alfil de casillas blancas y/o avanzar el peón de g2 para preparar el enroque largo. Por su parte, las negras podrían responder con la defensa siciliana (2...d6) o la defensa francesa (2...e6), entre otras opciones.

Figura 25. **Segunda apertura más frecuente**



Fuente: elaboración propia, realizado con Python.

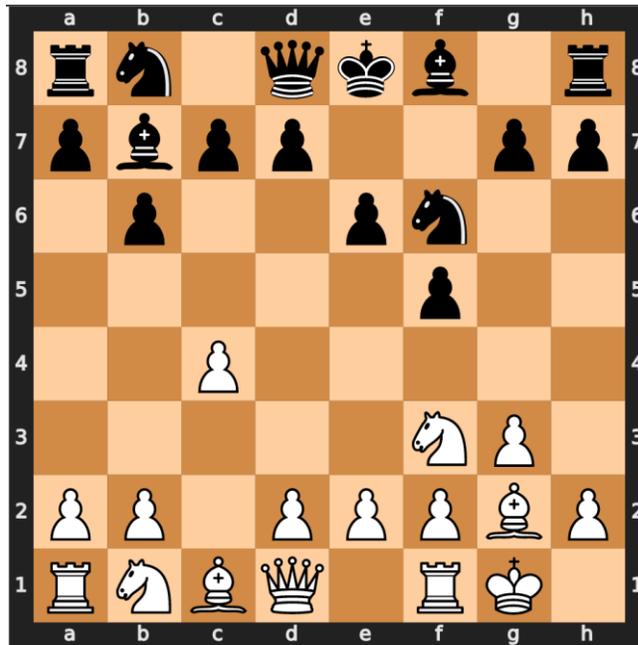
La figura 24 representa la segunda partida más jugada con un número de frecuencia de tres veces, representa la segunda apertura con mayor frecuencia de las partidas analizadas, constituye la posición inicial de la Apertura Española, también conocida como Ruy López.

En esta posición, el alfil de casillas blancas ha avanzado hasta la casilla b5, atacando al caballo negro en c6. Este es el movimiento característico de la Apertura Española. El objetivo de las blancas es presionar al negro en el centro del tablero y controlar el espacio en el flanco de rey.

En cuanto a la estructura de peones, las blancas tienen una ventaja espacial gracias a su control sobre las casillas centrales. Ambos bandos han avanzado sus peones centrales, pero las blancas tienen un peón adicional en el centro del tablero (e4). El alfil de casillas blancas también está en una buena posición, ya que controla la diagonal blanca que apunta hacia el flanco de rey negro.

Por su parte, las negras también han desarrollado sus piezas hacia el centro del tablero, aunque el caballo en c6 está atacado por el alfil blanco. Las negras podrían considerar mover su caballo a d4 para controlar más espacio central, o avanzar su peón de e6 para liberar su alfil de casillas negras.

Figura 26. Tercera apertura más frecuente



Fuente: elaboración propia, realizado con Python.

Representa la posición inicial de la Defensa Siciliana, una de las aperturas más populares en el ajedrez.

En esta posición, las negras han avanzado su peón de c7 a c5 para controlar el centro del tablero. El objetivo de esta jugada es desafiar el control blanco sobre el centro y liberar el camino para el desarrollo del caballo en c6. El peón de e7 está preparado para avanzar a e6, lo que liberará al alfil de casillas negras y le dará más espacio para moverse. El caballo de f6 también está desarrollado, controlando la casilla central d5.

Por su parte, las blancas han avanzado su peón de d2 a d4 para controlar el centro del tablero. El caballo de g1 está preparado para moverse a f3, desde donde puede controlar la casilla central d4 y apuntar al flanco de rey negro. El alfil de casillas blancas también está en una buena posición, controlando la diagonal blanca que apunta hacia el flanco de dama negro.

5.8. Optimización de jugadas de ajedrez con el motor Stockfish

La estrategia en el ajedrez es un tema fascinante que ha sido objeto de estudio y análisis desde hace décadas. En particular, la evaluación de posiciones de ajedrez y la determinación de los mejores movimientos posibles para cada jugador son áreas fundamentales para el desarrollo de la estrategia. En la actualidad, se dispone de herramientas muy potentes que permiten analizar posiciones y determinar las mejores jugadas en cuestión de segundos. Entre estas herramientas, se encuentra el motor de ajedrez Stockfish, que es uno de los más utilizados y respetados en el mundo del ajedrez.

En este trabajo, se ha desarrollado un análisis haciendo uso del motor Stockfish para la evaluación de posiciones de ajedrez. Para ello, se ha utilizado el módulo Chess.engine de Python, que proporciona una interfaz para interactuar con el motor de ajedrez. En concreto, se ha implementado un *script* que carga una posición de ajedrez, solicita al motor Stockfish que encuentre el mejor movimiento para esa posición y evalúa la puntuación del movimiento recomendado. Además, se han realizado pruebas con diferentes posiciones de ajedrez para evaluar el rendimiento del motor en diferentes situaciones. Los resultados obtenidos muestran la efectividad y potencia del motor Stockfish en la evaluación de posiciones de ajedrez y su utilidad para la mejora de la estrategia en el ajedrez.

Las posiciones de ajedrez a evaluar son las obtenidas del resultado del capítulo: Análisis de aperturas más frecuentes.

5.8.1. Obtención de Mejor Movimiento a la primera apertura más frecuente

La posición de ajedrez dada en formato FEN, *rnbq1rk1/ppp1ppbp/3p1np1/8/2PPP3/2N2N2/PP3PPP/R1BQKB1R*, fue evaluada mediante el uso del motor Stockfish, y se encontró que el mejor movimiento posible es c1g5. Este movimiento involucra el movimiento del alfil blanco de su posición original en c1 a la casilla g5, donde pone en jaque al rey negro y amenaza capturar la dama negra en d8.

La evaluación de la posición por parte del motor Stockfish también arrojó una puntuación de -13 para el jugador blanco, lo que sugiere que la posición es ventajosa para el jugador negro. La puntuación se calcula utilizando la función *PovScore*, que proporciona una evaluación de la posición desde el punto de vista del jugador blanco. En este caso, el valor negativo de la puntuación indica que el jugador blanco está en una posición desventajosa.

5.8.2. Obtención de mejor movimiento a la segunda apertura más frecuente

La posición de ajedrez dada, *r1bqk2r/1pppbppp/p1n2n2/4p3/B3P3/5N2/PPPP1PPP/RNBQ1RK1*, fue evaluada mediante el uso del motor Stockfish, y se encontró que el mejor movimiento posible es f1e1. Este movimiento involucra el movimiento del rey blanco de su posición original en f1 a la casilla e1, donde el rey blanco se coloca en una posición más segura y se prepara para la continuación del juego.

La evaluación de la posición por parte del motor Stockfish también arrojó una puntuación de +98 para el jugador blanco, lo que sugiere que la posición es ventajosa para el jugador blanco. La puntuación se calcula utilizando la función *PovScore*, que proporciona una evaluación de la posición desde el punto de vista del jugador blanco. En este caso, el valor positivo de la puntuación indica que el jugador blanco está en una posición ventajosa.

5.8.3. Obtención de mejor movimiento a la tercera apertura más frecuente

La posición de ajedrez dada, rn1qkb1r/pbpp2pp/1p2pn2/5p2/2P5/5NP1/PP1PPPBP/RNBQ1RK1, fue evaluada mediante el uso del motor Stockfish, y se encontró que el mejor movimiento posible es d2d4. Este movimiento implica el avance de peón blanco de su posición original en d2 a la casilla d4, lo que permite al jugador blanco tomar más espacio en el centro del tablero y potencialmente amenazar la posición del caballo negro en c5.

La evaluación de la posición por parte del motor Stockfish también arrojó una puntuación de +155 para el jugador blanco, lo que sugiere que la posición es muy ventajosa para el jugador blanco. La puntuación se calcula utilizando la función *PovScore*, que proporciona una evaluación de la posición desde el punto de vista del jugador blanco. En este caso, el valor positivo de la puntuación indica que el jugador blanco está en una posición muy favorable.

5.9. Optimización de la toma de decisiones en ajedrez mediante el análisis computacional de áreas críticas

La toma de decisiones en el ajedrez es un proceso complejo que requiere la evaluación de múltiples factores y variables en tiempo real. Los jugadores de ajedrez tienen que considerar no solo las posibles jugadas disponibles, sino también las posibles respuestas de su oponente, las implicaciones a largo plazo de cada movimiento y la posición general del tablero. La habilidad de tomar decisiones rápidas y precisas en el ajedrez es fundamental para cualquier jugador, ya sea un principiante o un gran maestro.

En los últimos años, el análisis computacional de posiciones críticas se ha convertido en una herramienta valiosa para ayudar a los jugadores de ajedrez a optimizar su toma de decisiones. Los programas de ajedrez y los motores de juego pueden analizar rápidamente miles de posibles movimientos y evaluar su efectividad en función de una serie de criterios, como el valor material, la estructura de peones y la seguridad del rey. Este análisis permite a los jugadores ver las posibles consecuencias de cada movimiento antes de hacerlo, lo que les ayuda a tomar decisiones más informadas.

En el capítulo: análisis de jugadas más frecuente, Se obtienen las tres jugadas más frecuentes, Para este análisis se utilizaron los resultados de las posiciones de las jugadas en formato FEN. Debido a que, para obtener un mejor análisis, se requieren capacidad de procesamiento muy altas, para ciertas posiciones no se obtuvieron áreas a evaluar. Los resultados de a continuación son únicamente de posiciones donde la capacidad computación logro realizar análisis.

Los resultados obtenidos para la posición r1bqk2r/1pppbppp/p1n2n2/4p3/B3P3/5N2/PPPP1PPP/RNBQ1RK1 fueron los siguientes:

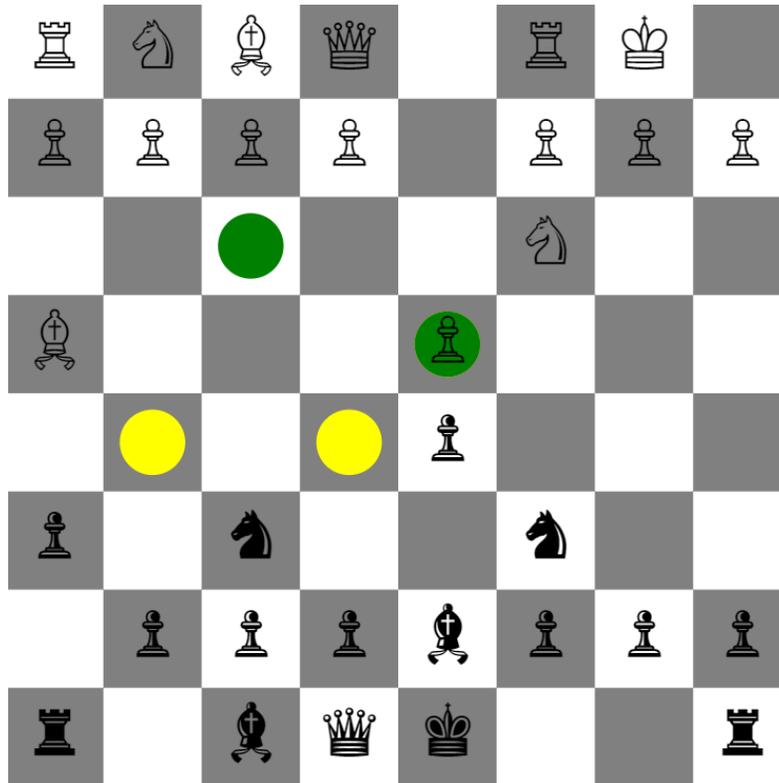
Áreas atacables por las blancas: set() - Esto indica que las blancas no tienen áreas a las que puedan atacar en este momento.

Áreas débiles de las blancas: {25, 27, 36} - Esta salida indica las áreas de la posición de las blancas que podrían ser vulnerables a futuros ataques de las negras. En este caso, las áreas débiles son la casilla 25 (donde hay un peón aislado), la casilla 27 (donde no hay peón para defender al caballo) y la casilla 36 (donde se encuentra un peón aislado y avanzado).

Áreas atacables por las negras: {42, 36} - Este resultado indica las áreas de la posición de las blancas que las negras pueden atacar en este momento. En este caso, las áreas atacables son la casilla 42 (donde se encuentra un peón aislado y avanzado) y la casilla 36 (donde no hay peón para defender al caballo).

Áreas débiles de las negras: set() - Este resultado indica que las negras no tienen áreas débiles en este momento.

Figura 27. **Análisis áreas atacables y débiles 1**



Fuente: elaboración propia, realizado con Python.

Los resultados obtenidos para la posición `r1bqk2r/1pppbppp/p1n2n2/4p3/B3P3/5N2/PPPP1PPP/RNBQ1RK1` fueron los siguientes:

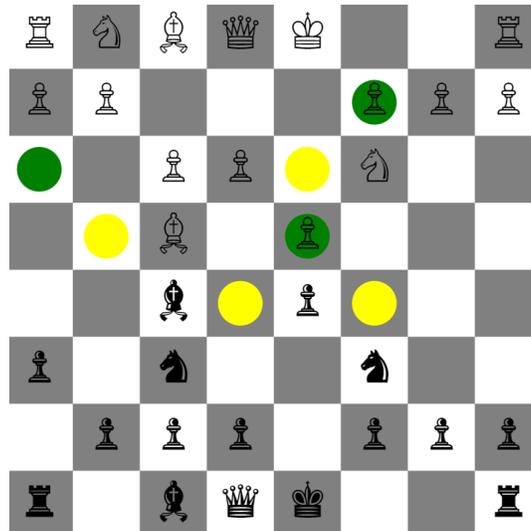
Áreas atacables por las blancas: Este resultado indica que las blancas no tienen áreas a las que puedan atacar en este momento.

Áreas débiles de las blancas: Casillas 33, 27, 44 y 29. Esta salida indica las áreas de la posición de las blancas que podrían ser vulnerables a futuros ataques de las negras. En este caso, las áreas débiles son la casilla 33 (donde está el peón aislado), la casilla 27 (que queda sin control cuando el alfil mueve), la casilla 44 (donde se encuentra un peón aislado y avanzado) y la casilla 29 (que también queda sin control cuando el alfil mueve).

Áreas atacables por las negras: casillas 40, 36 y 53 - Este resultado indica las áreas de la posición de las blancas que las negras pueden atacar en este momento. En este caso, las áreas atacables son la casilla 40 (donde se encuentra el peón avanzado de las blancas), la casilla 36 (que queda sin control cuando el alfil mueve) y la casilla 53 (donde se encuentra el caballo de las blancas).

Áreas débiles de las negras: Este resultado indica que las negras no tienen áreas débiles en este momento.

Figura 28. **Análisis áreas atacables y débiles 2**



Fuente: elaboración propia, realizado con Python.

Los resultados obtenidos para la posición `r1bqk2r/1ppbppp/p1n2n2/4p3/B3P3/5N2/PPPP1PPP/RNBQ1RK1` fueron los siguientes:

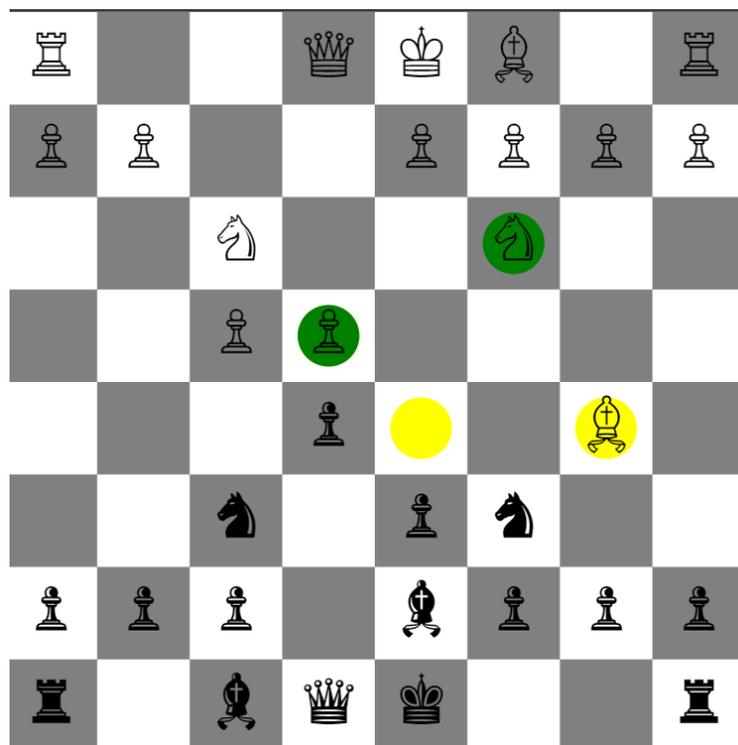
Este resultado indica que las blancas no tienen áreas a las que puedan atacar en este momento.

Áreas débiles de las blancas: casilla 35, que es la casilla donde se encuentra el peón d4 que está clavado por el alfil de g7, y las casillas 28 y 30, que son las casillas que están detrás de los peones c4 y e4 respectivamente y que pueden ser objeto de un ataque de los peones negros.

Áreas atacables por las negras: casillas 35 y 45. La casilla 35 es la casilla donde se encuentra el peón d4 que puede ser atacado por el caballo negro, mientras que la casilla 45 es la casilla donde se encuentra el peón f2 que puede ser atacado por la dama negra si se produce un cambio en la diagonal a8-h1.

Áreas débiles de las negras: Este resultado indica que las negras no tienen áreas débiles en este momento.

Figura 29. **Análisis áreas atacables y débiles 3**



Fuente: elaboración propia, realizado con Python.

CONCLUSIONES

1. A partir del análisis de piezas involucradas en jugadas de jaque mate, se puede concluir que la reina y la torre. La reina es la pieza que más participa en los jaques mates, ya que tiene una gran movilidad y puede atacar en múltiples direcciones. Por otro lado, la torre se posiciona como la segunda pieza que más jaques mates provoca después de la reina, debido a su gran movilidad y su capacidad para controlar columnas y filas completas en el tablero.
2. El análisis de la efectividad de las aperturas en el ajedrez muestra que existen algunas opciones que pueden aumentar las posibilidades de éxito de un jugador. En particular, se ha encontrado que la Apertura Inglesa, la Apertura Bird y la Apertura de peón de dama son las aperturas más efectivas con una tasa de éxito por del 100 % en las tres. Estas conclusiones pueden ser útiles para los jugadores que buscan mejorar su juego y aumentar sus posibilidades de ganar partidas. Cabe recordar que los porcentajes de éxito únicamente aplican para el caso de uso del torneo realizado por el Club Caissa Guatemala.
3. En cuanto a las estrategias de enroque más utilizadas en el ajedrez, se puede concluir que el enroque de rey es, por un margen muy amplio, la estrategia más utilizada.

4. A partir del análisis de las piezas más agresivas, se puede concluir que la pieza que más capturas realiza es el peón, seguida del alfil y la torre. Estas piezas tienen la mayor cantidad de movimientos posibles, lo que les permite moverse con mayor libertad y atacar a las piezas adversarias con mayor facilidad.
5. En cuanto a la distribución de victorias por color, se puede concluir que el color blanco tiene una ligera ventaja sobre el color negro. Esto puede deberse a que el jugador que utiliza las piezas blancas tiene la oportunidad de mover primero, lo que le permite tener una ligera ventaja en la apertura del juego.
6. Los análisis utilizando las herramientas de optimización Stockfish puede ser útil para mejorar la toma de decisiones en el ajedrez y desarrollar habilidades de análisis y estrategia.

RECOMENDACIONES

1. Ampliar la muestra de datos: para obtener conclusiones más precisas y confiables sobre el análisis de ajedrez mediante ciencia de datos, sería recomendable ampliar la muestra de datos analizados. Esto permitiría obtener un conjunto de datos más grande y variado que pueda representar mejor la realidad del juego de ajedrez.
2. Validar las conclusiones con expertos en ajedrez: es importante validar las conclusiones obtenidas a través del análisis de ajedrez con expertos en la materia, como jugadores profesionales, entrenadores o maestros de ajedrez. Esto permitiría comprobar si las conclusiones obtenidas son coherentes con la experiencia práctica y si se pueden aplicar efectivamente en el juego real.
3. Explorar nuevas herramientas y técnicas de análisis de datos: existen diversas herramientas y técnicas de análisis de datos que pueden ser útiles para el análisis de ajedrez, como el aprendizaje automático, el análisis de redes, el procesamiento del lenguaje natural, entre otras. Sería interesante explorar estas opciones para ampliar el análisis y obtener nuevas conclusiones y perspectivas sobre el juego de ajedrez.

4.

REFERENCIAS

1. Alfaro, F. R. (s.f.). *Data Management*. Recuperado de <https://datamanagement.es/2019/10/13/proceso-etl-con-python-desde-cero-y-paso-a-paso/>.
2. Bååth's, R. (10 de junio, 2015). sumsar [Mensaje en un blog]. Recuperado de <http://www.sumsar.net/blog/2015/06/big-data-and-chess/>.
3. BBVA. (7 de enero, 2019). Jubilaciondefuturo [Mensaje en un blog]. Recuperado de <https://www.jubilaciondefuturo.es/es/blog/el-ajedrez-un-deporte-que-estimula-la-mente.html>.
4. Brust, A. V. (5 de agosto, 2021). bitsandbricks.github.io [Mensaje en un blog]. Recuperado de https://bitsandbricks.github.io/ciencia_de_datos_gente_sociable/index.html.
5. Bustamante Martínez, A., Galvis Lista, E. A., & Gómez Flórez, L. C. (2013). *Técnicas de modelado de procesos de ETL: una revisión de alternativas y su aplicación en un proyecto de desarrollo de una solución de BI*. *Red de Revistas Científicas de América Latina, el Caribe, España y Portugal*, 8.

6. Castro, S., Larrea, M., Urribarri, D., Ganuza, L., & Escarza, S. (2018). *Métricas, Técnicas y Semántica para la Visualización de Datos. Bahía Blanca.*
7. Duque Méndez, N. D., Hernández Leal, E. J., Pérez Zapata, Á. M., Arroyave Tabares, A. F., & Espinosa, D. A. (2016). *Modelo para el proceso de extracción, transformación y carga en bodegas de datos. una aplicación con datos ambientales.* Colombia: Universidad Militar Nueva Granada.
8. Echeverria, N. (2018). *Club Deportivo Basilio.* Recuperado de <http://clubdeportivobasilio.com/normas-ajedrez-competicion>.
9. Fernández Fernández, S., Cordero Sánchez, J., & Córdoba Largo, A. (2002). *Estadística Descriptiva.* Madrid: ESIC.
10. García, J., Molina, J., Berlanga, A., Patricio, M., Bustamante, Á., & Padilla, W. (2018). *Ciencia de datos. Técnicas analíticas y aprendizaje estadístico.* Bogotá: Altaria S.L.
11. Giraldo, V. (11 de septiembre, 2020). rockcontent [Mensaje en un blog]. Recuperado de <https://rockcontent.com/es/blog/visualizacion-de-datos/>.
12. Guatemala, F. N. (2018). Manual de normas y procedimientos. Guatemala [Mensaje en un blog]. Recuperado el 20 de septiembre de 2021.

13. Ichess. (16 de julio, 2018). ichess [Mensaje en un blog]. Recuperado de <https://www.ichess.es/blog/notacion-ajedrecistica-guia-principiantes/>.
14. Mukhopadhyay, S. (2018). *Advance Data Analytics Using Python*. Kolkata: Apress. Recuperado el 1 de octubre de 2021.
15. Muñoz, D. (26 de enero, 2018). THE ZUGZWANG BLOG [Mensaje en un blog]. Recuperado de <https://thezugzwangblog.com/ajedrez-deporte/>.
16. Oliva Alfonso, D., Pineda Alfonso, T., Kindelán Castro, D., & Carralero Iznaga, J. (2011). *Propuesta de herramientas para la integración de datos*. *Revista Cubana de Ingeniería*, 13.
17. Peña, S. (2017). *Análisis de datos*. Bogotá: Fondo Editorial Areandino.
18. Pulido, H. A. (2016). *Predicting the Outcome of a Chess Game by Statistical and Machine Learning techniques*. Barcelona.
19. Solutions, A. S. (2021). *Analytics Software & Solutions*. Recuperado de https://www.sas.com/es_ar/insights/data-management/what-is-etl.html.
20. Vega Malagón, G., Ávila Morales, J., Vega Malagón, A. J., Camacho Calderón, N., Becerril Santos, A., & Leo Amador, G. (2017). *Paradigmas en la investigación. enfoque cuantitativo y cualitativo*. Querétaro.

21. Vries, M. D. (2019). *The Use of Data Science for Sports Analysis Purposes. Paises Bajos.*

APÉNDICES

Apéndice 1. ***Script* creación de archivo en formato JSON**

La información descargada de la página de Chess24.com está en formato PGN por lo que es necesario realizar la transformación a un fichero en formato JSON.

La conversión a JSON se realiza mediante la ejecución de un *script* escrito en lenguaje Python. Las siguientes 4 imágenes detallan los pasos realizados para la obtención del nuevo fichero.

Como primer paso se deben importar las librerías necesarias para el análisis y tratamiento de los datos, en este caso se hace la importación de la librería Chess.pgn y la librería nativa json de Python. Aplicando el método de lectura de archivos en Python se carga en memoria las partidas en formato PGN.

Para el funcionamiento del algoritmo ejemplificado en la figura 2 es necesario declarar un arreglo el cual contendrá todas las partidas analizadas en formato diccionario de Python.

El código implementado a detalle para esta sección se puede observar en la siguiente figura.

Continuación del apéndice 1.

```
# Librería para el analizador de partidas en fomato PGN
import chess.pgn
# Librería para el manejo de formato JSON
import json
# Lectura del archivo que contiene todas las partidas
pgn = open("partidas.pgn")

# Declaración de arreglo que contendrá las partidas luego extraer los campos necesarios
GamesList = []
```

El siguiente paso en el *script* es realizar una iteración por cada partida y formar una variable de formato diccionario que será insertada al arreglo de partidas.

Para realizar la extracción de los datos se tiene una estructura de control *While* la cual va realizando un análisis léxico y sintáctico de la partida en secuencia. Dentro de cada iteración se hace uso del método *read_games* de la librería *Chess.pgn* la cual extrae la información en forma de clave-valor que será insertada en una variable temporal en formato diccionario tal y como se puede observar en la figura 2.

Continuación del apéndice 1.

```
stop = False
# Iteración de cada partida de ajedrez que el analizador logro extraer
while stop!=True:
    game = chess.pgn.read_game(pgn)
    if game is None:
        stop = True
    else:
        # Creación del registro en formato JSON con los campos analizados
        struct = {
            "Event": game.headers['Event'],
            "Site": game.headers['Site'],
            "Date": game.headers['Date'],
            "Round": game.headers['Round'],
            "White": game.headers['White'],
            "Black": game.headers['Black'],
            "Result": game.headers['Result'],
            "BlackElo": game.headers['BlackElo'],
            "EventDate": game.headers['EventDate'],
            "PlyCount": game.headers['PlyCount'],
            "TimeControl": game.headers['TimeControl'],
            "WhiteElo": game.headers['WhiteElo'],
            "Moves": str(game.mainline_moves()),
        }
```

La librería únicamente convierte los valores del archivo PGN a clave-valor para que puedan ser consultados de manera sencilla, sin embargo, para la realización de visualizaciones y análisis es necesario conocer el motivo por el cual la partida a finalizado. Mediante la implementación del método *game.end* se puede conocer cuál ha sido el motivo.

Implementando una estructura de control *if*, se crea un nuevo atributo el cual indica la razón de finalización de partida, las razones contempladas son las siguientes: abandono, jaque mate, tiempo, tablas por mutuo acuerdo, tablas por rey ahogado y *default* en caso de no poder determinar la razón.

Luego de completar los datos de la variable temporal se inserta en el arreglo de partidas declarado anteriormente. La implementación a detalle se puede observar en la figura 3.

Continuación del apéndice 1.

```
}
# Obtencion del motivo de finalizacion de partida
if ('abandono' in str(game.end())):
    struct['EndMotive'] = 'Abandono'
elif ('mate' in str(game.end())):
    struct['EndMotive'] = 'Jaque Mate'
elif ('tiempo' in str(game.end())):
    struct['EndMotive'] = 'Tiempo'
elif ('mutuo' in str(game.end())):
    struct['EndMotive'] = 'Tablas por mutuo acuerdo'
elif ('ahogado' in str(game.end())):
    struct['EndMotive'] = 'Tablas por rey ahogado'
else:
    struct['EndMotive'] = 'Indeterminado'

GamesList.append(struct)
```

Por último, con el método *dumps*, de la librería *json* que convierte objetos de Python a *json string*, se genera el arreglo de partidas con el cual se crea un fichero llamado *chess_records_etl_output* el cual contiene todas las partidas en formato *json*.

```
# Transformación del arreglo a formato JSON
jsonoutput = json.dumps(GamesList)
# Metodo crear un archivo en formato JSON del registros de todas las partidas
f = open("PGNJSON.json", "a")
f.write(jsonoutput)
f.close()
```

Fuente: elaboración propia, realizado con Python.

Apéndice 2. **Configuración de la conexión hacia la base de datos**

El segundo *script* que se desarrolla es el de la comunicación con la base de datos para la etapa de procesamiento de los registros de las partidas. En la primera parte del *script* se importa la librería JSON y MongoCLient, la herramienta proveniente de la librería Pymongo que sirve para la conexión con Mongoddb.

Antes de realizar la conexión se configuro la base de datos según la documentación oficial de Mongoddb para la instalación de su versión para Windows 10.

En la continuación del *script* se realiza la conexión a la instancia local de Mongoddb ubicada en el puerto por defecto 27017, la conexión se almacena en la variable *client* con la cual se realiza la autenticación al sistema mediante el usuario administrador y la creación de la colección que contendrá los registros de las partidas.

Para prevenir que la colección tenga registros previos, se elimina la colección con el comando *drop*, con esto se asegura que el registro de datos se hará a un almacén sin datos previos que afecten al análisis posterior.

Continuación del apéndice 2.

```
# Libreria para el manejo de formato JSON
import json
# Libreria para la comunicación con la base de datos local de mongodb
from pymongo import MongoClient

# conexión con la base de datos
client = MongoClient('localhost', 27017)
# autenticación de la base
db = client['admin']
# selección de la nueva colección
collection = db['PGNGames']
# elimina la colección con el objetivo de limpiar los datos
collection.drop()

with open('PGNJSON.json') as f:
    file_data = json.load(f)

collection.insert_many(file_data)
```

Fuente: elaboración propia, realizado con Python.

Apéndice 3. **Eliminación de registros duplicados**

El proceso de eliminación de duplicados se divide en dos secciones. En la primera sección se utiliza el *Framework* de agregación de Pymongo. El operador *\$group* agrupa los registros por su identificador y almacena el id en el campo *duplis* con el operador *\$addToSet*. Con la instrucción *\$sum* se contabiliza el número de duplicados que se presenta en un registro. La siguiente agrupación se realiza con el operador *\$match* el cual filtra aquellos registros que contabilizan 2 o más similitudes.

En la sección 2 se itera sobre el curso que devuelve la función de agregación y se remueve el primer id del arreglo de duplicado, luego se agrega el resto a un arreglo el cual se usará para remover los duplicados.

Por último, se ejecuta la función *delete_many* con los identificadores de los duplicados almacenados en el arreglo *response*. El código completo se puede apreciar en la figura 6.

Continuación del apéndice 3.

```
#Seccion1 . Eliminar registros duplicados (el valor que se toma como duplicado son los movimientos de las partidas)
cursor = collection.aggregate([
    {"$group": {
        "_id": {'moves':"$Moves"},
        'duplis': {'$addToSet':'$id'},
        'count':{'$sum':1}
    }},
    {"$match":{"count":{"$gt":1}}
    ]
})

# Seccion 2. Variable que contiene los registros encontrados duplicados que serán eliminados
response = []
for doc in cursor:
    del doc["duplis"][0]
    for id in doc["duplis"]:
        response.append(id)

# Elimina los registros que coincidan con el id del arreglo
collection.delete_many({"_id": {"$in": response}})
```

Fuente: elaboración propia, realizado con Python.

Apéndice 4. **Comando eliminación campos innecesarios**

Mediante la ejecución del operador `$unset` se elimina de los registros los campos, `BlackElo`, `Date`, `Round` y `WhiteElo`. El comportamiento del operador reemplaza las coincidencias con el valor nulo, en vez de remover del arreglo los elementos esto para mantener el consistente el tamaño y las posiciones del arreglo.

```
# Elimina Los campos (BlackELO, Date, Round y WhiteElo)
collection.update_many({}, {'$unset': {'BlackElo':1, 'Date':1, 'Round':1, 'WhiteElo':1}})
```

Fuente: elaboración propia, realizado con Python.

Apéndice 5. Comandos para cambio de representación de resultado de partidas

Para comprender de una mejor manera los resultados en las visualizaciones se realiza la actualización de los valores del campo *Result*. El análisis de los valores por los cuales van a ser modificados se explica en el capítulo 3.

El operador *set* actualiza cada campo con el cual se obtenga una coincidencia lexicográfica y lo reemplaza con el nuevo valor introducido en el campo *Result*. El procedimiento se realiza para cada variante que presenta el campo como se puede ver a detalle en la figura 8.

```
# Actualización de formato de resultado final de partida
collection.update_many({'Result': "1-0"}, {'$set': {'Result': "White"}})
collection.update_many({'Result': "0-1"}, {'$set': {'Result': "Black"}})
collection.update_many({'Result': "1/2-1/2"}, {'$set': {'Result': "Draw"}})

client.close()
```

Fuente: elaboración propia, realizado con Python.