

Análisis Exploratorio de Datos en Ajedrez



Autor: Jesús González Suárez

Fecha:

30 de julio de 2024

[The Bridge Data Science Bootcamp]

Índice

Índice	1
Índice de figuras	1
Índice de cuadros	1
1 Introducción	2
2 Descripción del Dataset	2
3 Hipótesis Formuladas	3
4 Análisis y Resultados	4
4.1 Hipótesis 1: Los jugadores con una calificación más alta tienden a ganar más partidas	4
4.2 Hipótesis 2: El resultado de la partida depende del tipo de apertura utilizado, es decir ciertas aperturas específicas tienen una mayor tasa de victorias para las blancas o las negras	5
4.3 Hipótesis 3: Las partidas con un mayor número de turnos tienden a terminar en tablas	7
5 Conclusiones	8
5.1 Hipótesis 1: Los jugadores con una calificación más alta tienden a ganar más partidas	8
5.2 Hipótesis 2: El resultado de la partida depende del tipo de apertura utilizado, es decir ciertas aperturas específicas tienen una mayor tasa de victorias para las blancas o las negras	8
5.3 Hipótesis 3: Las partidas con un mayor número de turnos tienden a terminar en tablas	8
5.4 Conclusión final	8

Índice de figuras

1	Calificaciones promedio de los ganadores y perdedores	4
2	Distribución de turnos por estado de victoria	5
3	Resultados de partida por apertura	6
4	Tasas de victoria para las 10 aperturas más comunes	7

Índice de cuadros

1	Tabla de ganadores cuando ganan Blancas	4
2	Tabla de ganadores cuando ganan Negras	4
3	Comparación de métodos para ganadores con piezas blancas y negras . .	4
4	Resultados del Test de Chi-Cuadrado	5
5	Resultados del Análisis de Varianza (ANOVA)	7

Análisis Exploratorio de Datos en Juegos de Ajedrez

Autor: Jesús González Suárez

30 de julio de 2024

Resumen

Este documento presenta un análisis exploratorio de datos (EDA) de un conjunto de juegos de ajedrez obtenidos de Kaggle. Se formularon y probaron varias hipótesis utilizando pruebas estadísticas y visualizaciones. Los resultados de las pruebas se discuten y se proporcionan conclusiones basadas en los datos.

1. Introducción

El objetivo de este análisis es explorar un conjunto de datos de juegos de ajedrez y formular varias hipótesis para entender mejor los patrones y tendencias en los datos. El conjunto de datos utilizado en este análisis fue obtenido de Kaggle y contiene información detallada sobre más de 20,000 juegos de ajedrez.

2. Descripción del Dataset

El conjunto de datos contiene las siguientes columnas:

- **Game ID:** Identificador único del juego.
- **Rated (T/F):** Indica si el juego está clasificado.
- **Start Time:** Hora de inicio del juego.
- **End Time:** Hora de finalización del juego.
- **Number of Turns:** Número de turnos en el juego.
- **Game Status:** Estado del juego (jaque mate, resignación, tiempo agotado, tablas).
- **Winner:** Ganador del juego (blancas, negras, tablas).
- **Time Increment:** Incremento de tiempo por movimiento.
- **White Player ID:** Identificador del jugador de blancas.
- **White Player Rating:** Calificación del jugador de blancas.
- **Black Player ID:** Identificador del jugador de negras.
- **Black Player Rating:** Calificación del jugador de negras.

- **All Moves:** Todos los movimientos en notación estándar de ajedrez.
- **Opening Eco:** Código estandarizado de la apertura.
- **Opening Name:** Nombre de la apertura.
- **Opening Ply:** Número de movimientos en la fase de apertura.

3. Hipótesis Formuladas

Se formularon las siguientes hipótesis para el análisis:

1. Los jugadores con una calificación más alta tienden a ganar más partidas.
2. Ciertas aperturas específicas tienen una mayor tasa de victorias para las blancas o las negras.
3. Las partidas con un mayor número de turnos tienden a terminar en tablas.

4. Análisis y Resultados

4.1. Hipótesis 1: Los jugadores con una calificación más alta tienden a ganar más partidas

Se realizaron pruebas t de Student y U de Mann-Whitney para comparar la relación entre el rating y la cantidad de partidas ganadas. Los resultados indican que hay una relación entre el rating y la cantidad de partidas ganadas al obtener $p\text{-value} < 0,05$.

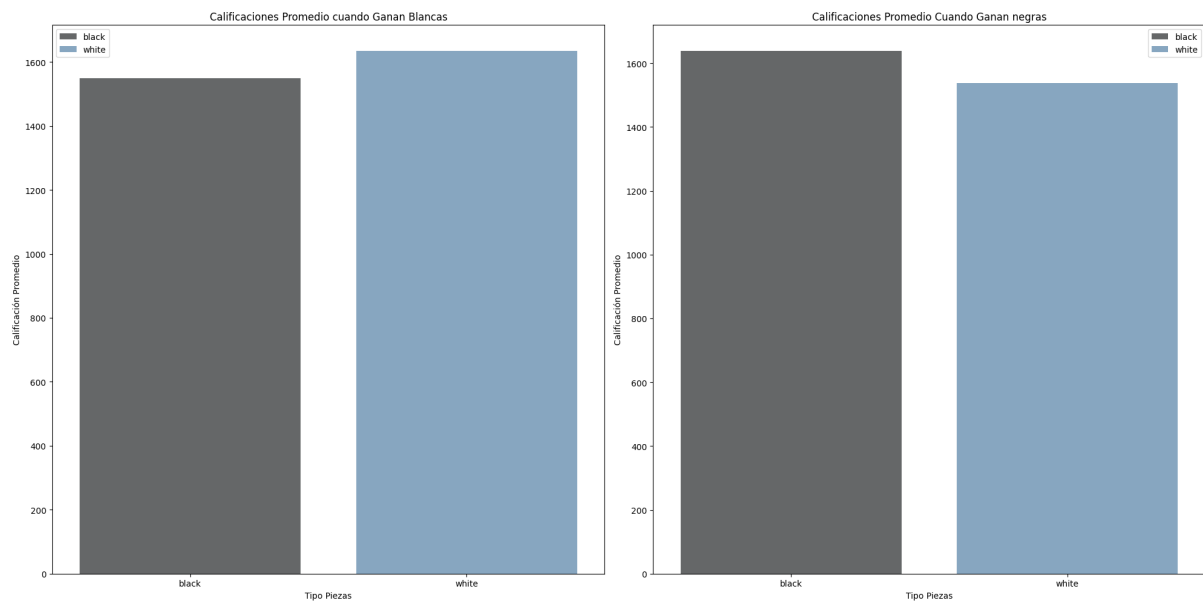


Figura 1: Calificaciones promedio de los ganadores y perdedores

winner	white_rating
black	1549.161117
white	1633.757659

Cuadro 1: Tabla de ganadores cuando ganan Blancas

winner	black_rating
black	1637.666891
white	1538.560151

Cuadro 2: Tabla de ganadores cuando ganan Negras

	Ganan Blancas		Ganan Negras	
metdod	u-stat	p-valor	u-stat	p-valor
t-student	20.23	4.11E-90	23.71	1.69E-122
Mann-Whitney	50495725.5	1.92E-76	51737988.5	4.98E-106

Cuadro 3: Comparación de métodos para ganadores con piezas blancas y negras

4.2. Hipótesis 2: El resultado de la partida depende del tipo de apertura utilizado, es decir ciertas aperturas específicas tienen una mayor tasa de victorias para las blancas o las negras

Al comparar 2 Variables Categoricalas, se realizó una prueba de chi-cuadrado para comparar las tasas de victorias entre diferentes aperturas. El test chi-2 tiene como hipótesis nula (o de partida) la independencia de las variables, por tanto como hemos obtenido un $p\text{-value} < 0,05$. podemos rechazar la hipótesis de partida con seguridad y pensar que existe una relación estadísticamente significativa, por lo que podemos decir que ciertas aperturas favorecen a un jugador sobre el otro.

Por otro lado se hizo una agrupacion por tipo de apertura y ganador para ver la distribucion y ver como se relacionaba el tipo de apertura con las victorias tanto de piezas blancas como negras, obteniendo un mapa de calor y un $p\text{-value} < 0,05$, por lo que tambien podemos decir que determinadas aperturas estan relacionadas con victorias de las peizas blancas o piezas negras

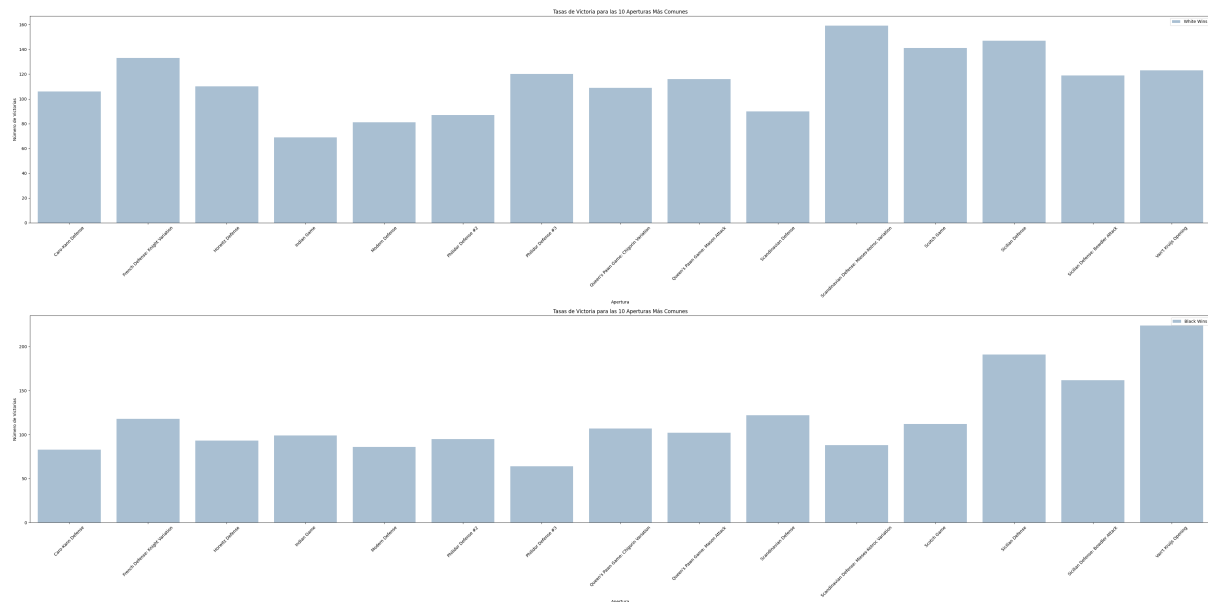


Figura 2: Distribución de turnos por estado de victoria

Chi-Cuadrado	
Chi-Cuadrado	110.15
P-Value	1.07E-11
Grados de Libertad	28

Cuadro 4: Resultados del Test de Chi-Cuadrado

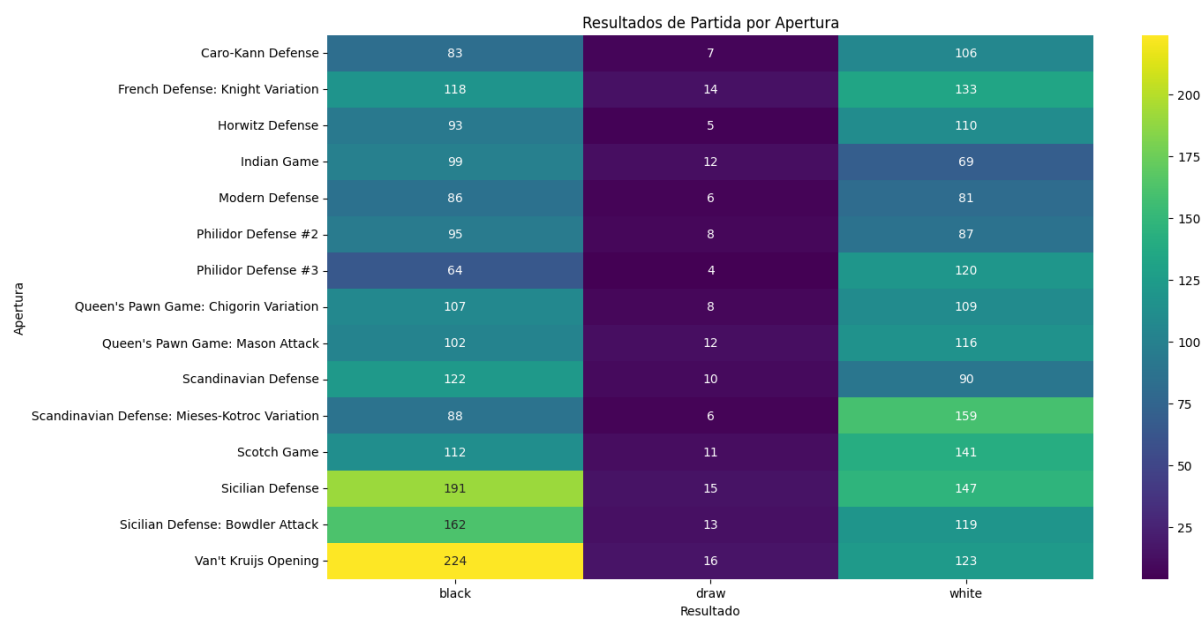


Figura 3: Resultados de partida por apertura

4.3. Hipótesis 3: Las partidas con un mayor número de turnos tienden a terminar en tablas

Se utilizó ANOVA para comparar las medias de tres o más grupos para ver si al menos uno de los grupos difiere significativamente de los demás. En este caso, estamos comparando el número de turnos entre diferentes tipos de resultados. En este caso, el valor p es $2.3e-28$, que es $p\text{-value} < 0,05$. Por lo tanto, rechazamos la hipótesis nula y concluimos que hay una diferencia significativa en el número de turnos entre los diferentes estados de victoria. Esto respalda la hipótesis de que las partidas con un mayor número de turnos tienden a terminar en tablas.

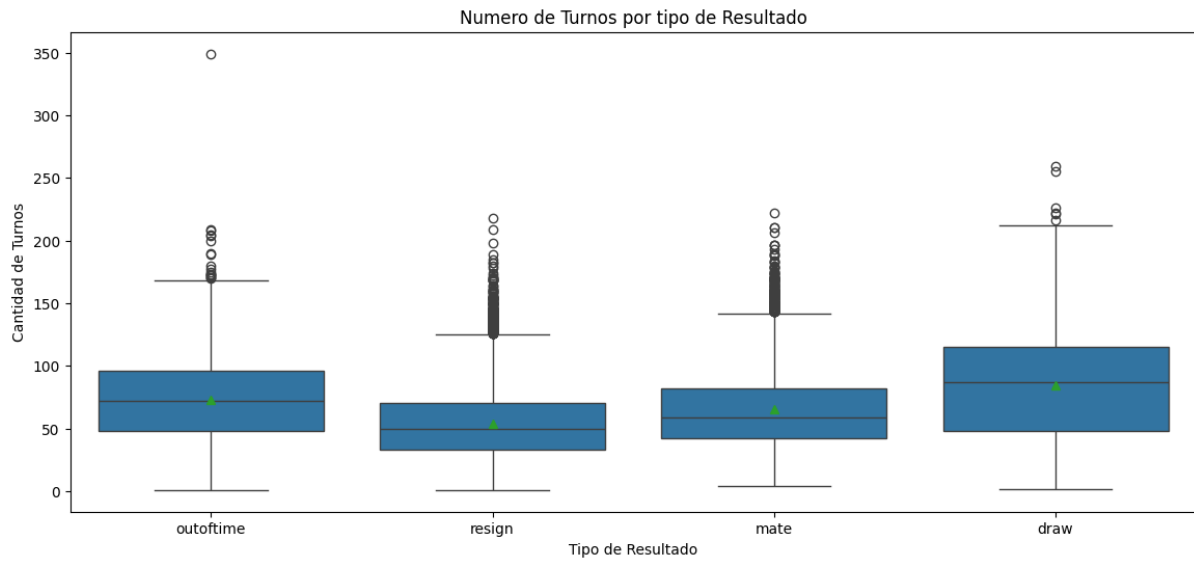


Figura 4: Tasas de victoria para las 10 aperturas más comunes

Anova	
f_val	436.56
p_val	1.37E-274

Cuadro 5: Resultados del Análisis de Varianza (ANOVA)

5. Conclusiones

5.1. Hipótesis 1: Los jugadores con una calificación más alta tienden a ganar más partidas

La calificación de los jugadores es un indicador significativo del éxito en el ajedrez. Los jugadores con calificaciones más altas tienden a ganar más partidas. Esto sugiere que la experiencia y habilidad reflejada en la calificación de un jugador es un factor crucial para el éxito en las partidas de ajedrez.

5.2. Hipótesis 2: El resultado de la partida depende del tipo de apertura utilizado, es decir ciertas aperturas específicas tienen una mayor tasa de victorias para las blancas o las negras

Las aperturas juegan un papel fundamental en el resultado de las partidas. Ciertas aperturas favorecen significativamente a un jugador sobre el otro, y el tipo de apertura utilizada puede determinar en gran medida si la partida será ganada por las blancas o las negras. Esto subraya la importancia de la preparación y el conocimiento de las aperturas en el ajedrez competitivo.

5.3. Hipótesis 3: Las partidas con un mayor número de turnos tienden a terminar en tablas

Las partidas que terminan en tablas tienden a tener un mayor número de turnos en comparación con las que terminan en jaque mate, abandono o tiempo agotado. Esto indica que las partidas más prolongadas y equilibradas, donde ninguno de los jugadores comete errores decisivos, suelen terminar en tablas. La duración de una partida puede, por lo tanto, ser un indicador de un juego bien igualado.

5.4. Conclusión final

El análisis de los datos revela que varios factores influyen de manera significativa en el resultado de una partida de ajedrez. La calificación del jugador, la duración de la partida y las estrategias de apertura son determinantes. Los jugadores de mayor calificación tienen una ventaja inherente, las partidas más largas tienden a ser más equilibradas, y el conocimiento y uso efectivo de las aperturas pueden cambiar el curso de una partida.

En resumen, el éxito en el ajedrez depende de una combinación de habilidad y estrategia. Los jugadores deben trabajar en mejorar su calificación a través de la práctica y el estudio, entender la importancia de la duración y el equilibrio en el juego, y dominar una variedad de aperturas para aumentar sus posibilidades de victoria.

Estas conclusiones proporcionan una visión integral de los aspectos críticos que deben ser considerados por los jugadores que buscan mejorar su rendimiento en el ajedrez.