



EDA: Ganar a la Casa de Apuestas



Análisis Exploratorio - Tenis ATP 2020-2024



¿Existen factores que las casas de apuestas subestiman?



1. Datos e Integración

FUENTES

Partidos ATP con Datos de las apuestas
(Tennis-Data.co.uk)

Características de los jugadores (Kaggle)

PROCESO DE INTEGRACIÓN

1. UNIR 5 AÑOS

```
df_partidos = pd.concat([2020...2024])
```

2. IDENTIFICAR JUGADORES

apellido + inicial + control de edad

3. UNIR POR player_id

clave real, no posición de fila

11.794 partidos · 510 de 517 jugadores identificados

2. Variables

Tras la integración, seleccionamos solo variables relevantes para el caso de negocio

1. Variables de las casas de apuestas

Variable Qué es Para qué la necesito
--

Date Fecha del partido Análisis temporal: ¿ha cambiado algo en 5 años?
--

Series Tipo Torneo (ATP250, 500, 1000, GS) Identificar Grand Slams
--

Surface Hierba, tierra o pista dura ¿la altura importa más en hierba?

Best of Partido a 3 o 5 sets ¿los veteranos aguantan mejor partidos largos?

Round Ronda del torneo Ver si en finales pasa algo diferente
--

AvgW, AvgL Cuotas medias Se usan las medias, no las mejores: más conservador
--

Tenemos datos de varias casas de apuestas, pero nos quedamos con las medias

2. Variables de los jugadores (Kaggle)

Variable Tipo Uso

Winner, Loser Texto Identificar jugadores

WRank, LRank Numérica Rankings

Winner_Height, Loser_Height Numérica Análisis altura
--

Winner_Hand, Loser_Hand Categórica Zurdos vs diestros

Winner_Age, Loser_Age Numérica Factor experiencia

3. VARIABLES CREADAS

Height_Diff

Winner_Height - Loser_Height

¿Ganó el más alto?

Age_Diff

Winner_Age - Loser_Age

¿Ganó el veterano?

Puntos_Diff

WPts - LPts

¿Hubo sorpresa?

Winner_Was_Favorite

AvgW < AvgL

¿Ganó el favorito?

Winner_Is_Lefty

Winner_Hand == 'L'

¿Es zurdo el ganador?

3. Limpieza de Datos

1. Verificación inicial

✓ 0 duplicados

! 15 partidos sin cuota media

2. Tratamiento de nulos

✓ 77 walkovers y adjudicados fuera: la casa los anula

✓ Masters Cup: Best of = 3

3. Los «outliers» eran un error de cruce

! Había 434 registros con 45-77 años

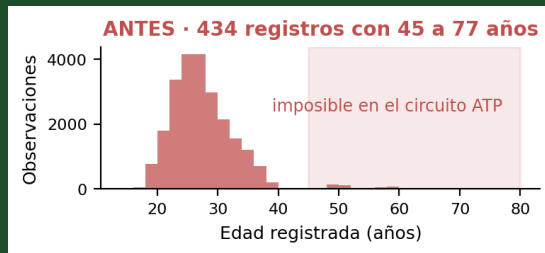
✓ No eran outliers: el cruce asignaba padres por hijos

El filtro IQR anterior borraba 466 partidos, 432 de ellos (93 %) por este error

11.717

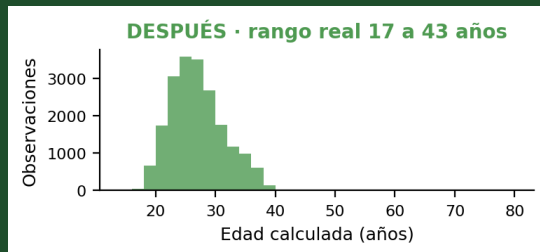
partidos analizados

ANTES



Edad registrada: hasta 77 años

DESPUÉS



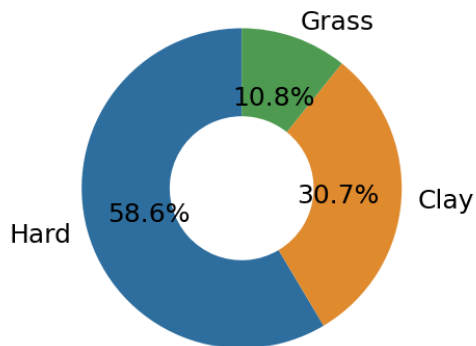
Edad calculada: 16,7 a 42,5 años

4. Análisis Univariante

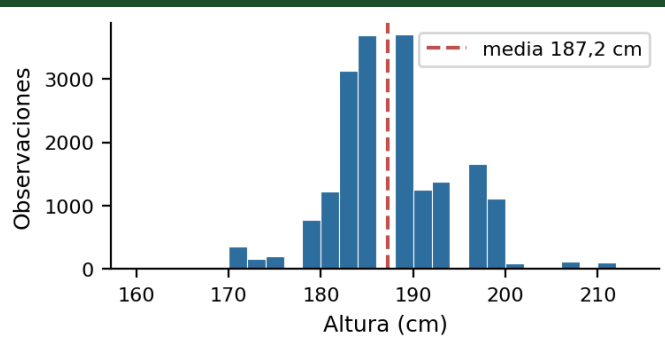
Conociendo la distribución de nuestras variables clave

Partidos por Superficie

Partidos por superficie



Mayoría en pista dura (58,6 %). Hierba solo 10,8 %: la temporada de hierba dura unas semanas.



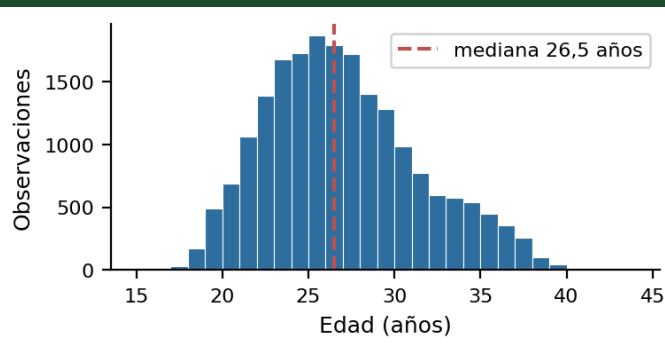
Altura

Media:

187,2 cm

Rango:

163-211



Edad

Pico:

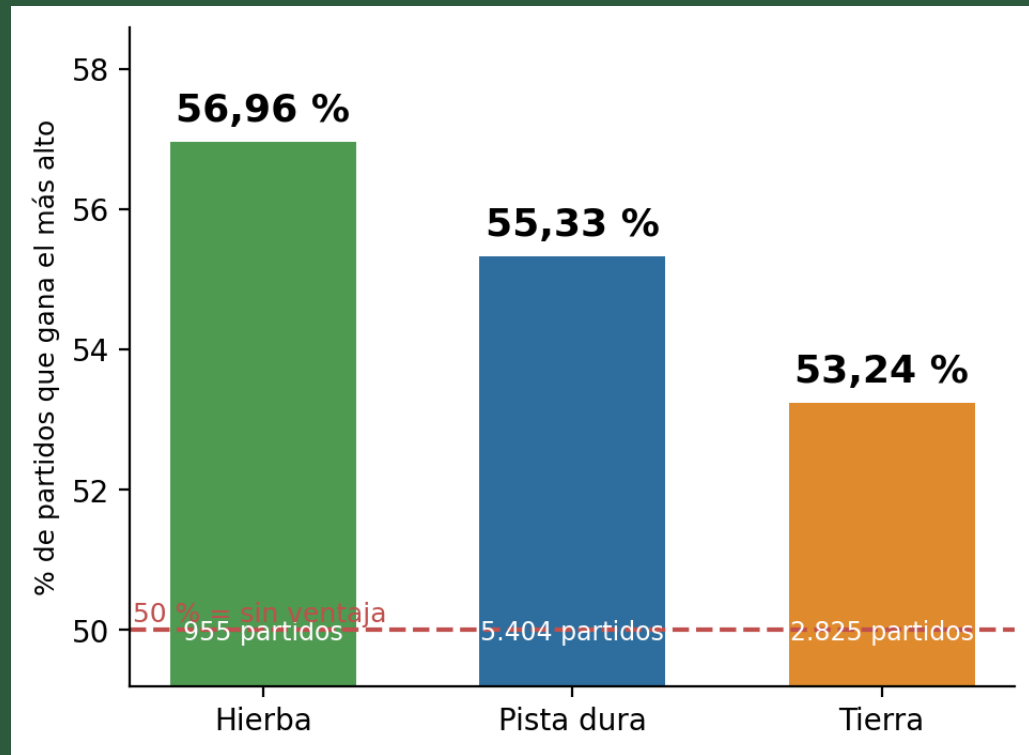
24-27
años

Máximo:

42,5 años

4. ANÁLISIS BIVARIANTE

¿Los altos ganan más en superficies rápidas?



RESULTADO

Hierba: 56,96 %

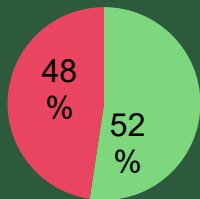
Pista dura: 55,33 %

Tierra: 53,24 %

Sí hay ventaja

El orden coincide con la teoría del saque, y eso es lo que lo hace creíble

¿Los zurdos tienen ventaja oculta?



■ Gana Diestro

■ Gana Zurdo

Resultado: Zurdo vs Diestro

Nadal: La excepción

EL MITO VS LA REALIDAD

Se dice que los zurdos tienen ventaja porque los rivales no están acostumbrados...

47,62 %

Los zurdos ganan MENOS del 50%

¿Por qué el mito? → Nadal

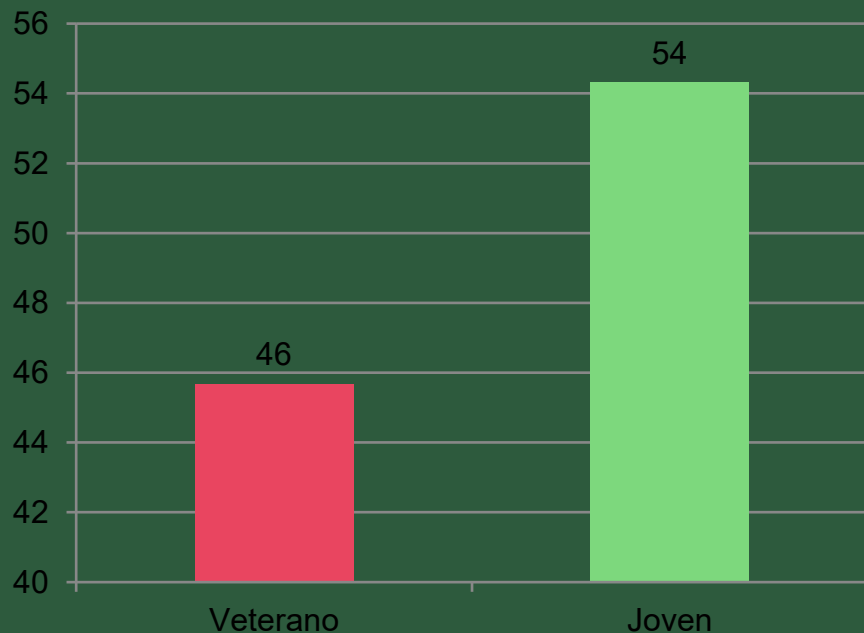
Nadal ganó 22 Grand Slams. La creencia se sostiene sobre su caso, no sobre el conjunto: en estos 2.671 duelos los zurdos ganan menos de la mitad.

Nadal es la excepción, no la regla

4. ANÁLISIS BIVARIANTE

¿Los veteranos rinden mejor en Grand Slams?

Partidos con diferencia de edad > 5 años



EL MITO VS LA REALIDAD

Se cree que la experiencia ayuda en partidos a 5 sets...

45,68 %

Veteranos en Grand Slams

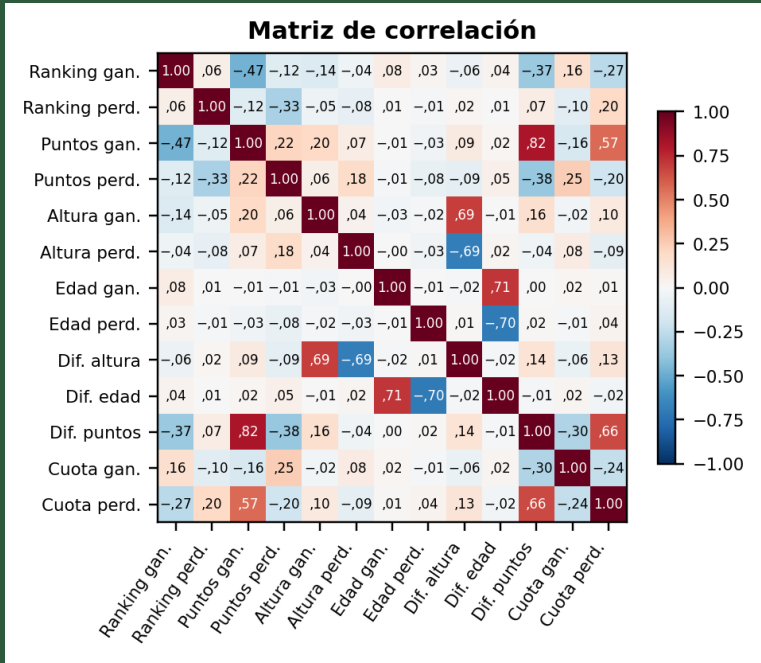
¿Y en partidos cortos?

Y no es que los partidos largos les penalicen: a 3 sets ganan el 45,27 %, prácticamente lo mismo. La duración no cambia nada.

1.031 partidos a 5 sets con más de 5 años de diferencia

5. Correlaciones y Simulación de Apuestas

Matriz de Correlación



SIMULACIÓN: Apostar 1€ por partido

Alto en hierba	56,96 %	-3,31 %
Apostar al favorito	68,18 %	-4,54 %
Todas pierden. La casa se queda un 5,46 % de margen medio.		

6. Conclusiones

¿Altos en superficies rápidas?

Sí (56,96 %)

No rentable (-3,31 %)

¿Zurdos con ventaja oculta?

No (47,62 %)

No rentable

¿Apostar siempre al favorito?

Sí (68,18 %)

No rentable (-4,54 %)

Las casas de apuestas son muy eficientes

Los patrones existen, pero la casa se guarda un 5,46 % de margen: apostando al favorito, al no favorito o al más alto en hierba, se pierde igual.



¿Preguntas?

Gracias por su atención