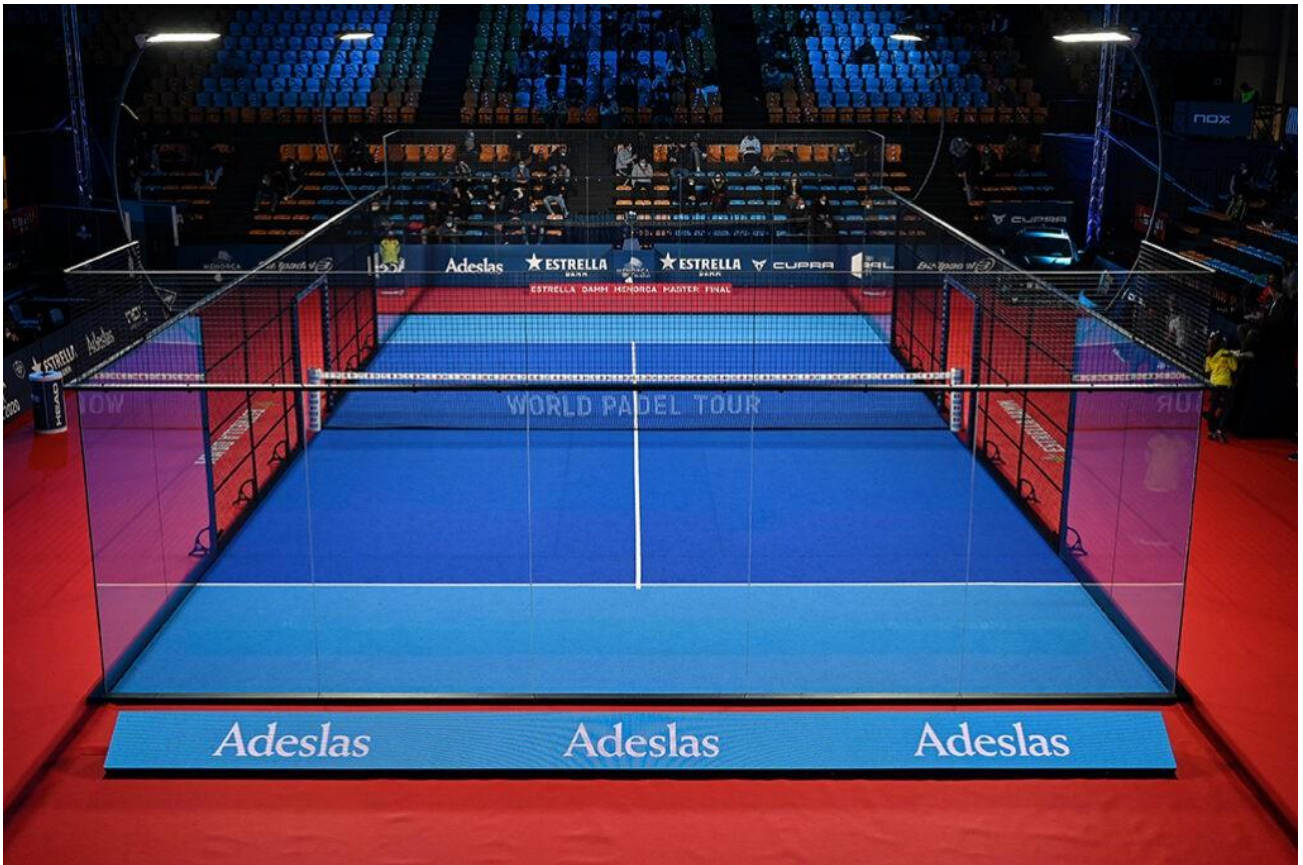


Proyecto final:

## Padel: Estilos de juego y probabilidades de éxito



Curso Data scientist – Coder house

Fecha 31/10/23

Profesor: **Russo Locati**, Ignacio

Tutor: **Dignani**, Federico

Autor: **Lambre**, Santiago

INDICE

Introducción ..... 3

Descripción de la temática de los datos ..... 3

Nivel de aplicación ..... 3

Objetivos ..... 4

Hallazgos encontrados por el EDA..... 4

Modelo elegido ..... 6

Conclusiones

## Introducción

El padel es un deporte en pleno crecimiento, el cual se juega 2 vs 2, y cada equipo cuenta con un entrenador sentado en el banquillo, quien puede hablar con los jugadores cada 2 games cuando cambian de lado de cancha. Se juega al mejor de 3 sets, con partidos que pueden durar desde 45 minutos a 3 horas.

Al ser un deporte en crecimiento, los premios, el dinero específicamente, no son tan importantes como en otros deportes, por lo que para todas las parejas es importante tener buenas actuaciones, cada cierto tiempo al menos, para obtener premios del torneo y para acaparar auspiciantes.

Una forma de maximizar las oportunidades de tener resultados, o cambiar ante los adversos, es con el cambio de pareja. Estos cambios se pueden llevar a cabo varias veces en el año, solo hace falta el acuerdo entre los nuevos integrantes de la pareja.

Luego del análisis del deporte, confirmado con el modelo de machine learning no supervisado, concluimos que los jugadores tienen 3 tipos de estilo de juego:

- **Definidores:** se destacan por tener alta cantidad de puntos totales y remates junto con moderados a bajos errores no forzados
- **Oportunistas:** valores intermedios entre los definidores y los defensivos. Suelen tener pocos errores no forzados. Según el desarrollo del partido pueden atacar o defender mas
- **Defensivos:** en este grupo se juntan jugadores que tienen un estilo de juego defensivo, suelen armar el punto para que defina el compañero y tener pocos remates. En el contexto del estudio, también se suman los jugadores que comenten gran cantidad de errores no forzados, que por su estilo de juego deberían estar en otro grupo

## Descripción de la temática de los datos

A partir de 2 bases de datos que contiene estadísticas de partidos de padel masculinos de la World Padel Tour (WPT), se evaluará cuáles datos pueden servir para realizar un modelo de Machine Learning

El dataset para cada jugador consta de 1256 filas y 28 columnas, siendo cada fila un set de un jugador. Los datos son del tipo numérico y categórico nominal u ordinal. En total se evaluaron 52 jugadores

Las características del dataset por jugadores son: nombre, edad, altura, posición, torneo, set, saque, NF\_total, w\_total, sm\_total, nsmtotal, precisión sm%, csm, fondo, m3, m5, m7, puntos ganados, juegos jugados, circuito, fase, mano, spcsm, bloqueo, partido, nfr, puntos netos.

El dataset de los equipos tiene 196 filas y 28 columnas, siendo cada fila las estadísticas de una pareja en un set. Los datos son del tipo numérico y categórico nominal u ordinal. Lo importante del dataset es la característica de quién ganó el set. En total se evaluaron 18 parejas

Las características del dataset por equipo son: nombre, nf, w, sm, ns, precisión, csm, fondo, m3, m5, m7, puntos totales, total snf, edad, spcsm, bloqueo, set, saque primario, juegos jugados, saque oro, resto oro, oportunidades de quiebre, circuito, torneo, fase, ganó set, nfr, partido, puntos netos.

## Nivel de aplicación

**Estratégico:** los resultados podrían ser de utilidad para el armado de una pareja de padel, para el entrenador que pueda servirse de los datos para enfocar el entrenamiento en los puntos flojos o altos del equipo, el estilo de juego de los jugadores según sus estadísticas y como meta para jugadores que quieran llegar a jugar al más alto nivel

Objetivos:

- 1. Generar un modelo que permita evaluar:
  - Los tipos de estilos de juego de los jugadores
  - Las probabilidades de una pareja de ganar un set, para así evaluar cuáles podrían ser combinaciones, probablemente, exitosas y cuáles tendrían menos probabilidades, junto con mejorar ayudar al análisis y entrenamiento de la pareja y evaluar su performance
- 2. Analizar la regla del cambio de compañero varias veces en el año teniendo en cuenta las probabilidades y los estilos de juego

EDA

Dataset de los jugadores

Luego de la limpieza del dataset, se eligieron las 3 columnas mas relevantes al analizar a los jugadores, como son los puntos por game (pxg), errores no forzados por game (nfxg), puntos netos (pts\_netos) y remates por game (rem). Las 3 primeras tuvieron una alta correlación con el resultado (cuando se analizó el dataset de los equipos con el resultado), y la cantidad de remates ayuda a diferenciar el estilo de juego de los jugadores.

Las características que no fueron tenidas en cuenta no aportaban al estudio por estar ya contabilizadas en una de las variables elegidas (por ejemplo hacer un punto después de 7 golpes está incluida en puntos totales) o por no tener utilidad específicamente para el estudio, por ejemplo el torneo

Dataset de los equipos

Luego de la limpieza del dataset, se utilizaron 4 características para predecir quienes ganan o pierden el set, las cuales tenían una correlación con el resultado mayor a 0.4. A los valores de las 3 columnas numéricas se las dividió por el promedio de games jugados por el jugador/equipo, así se evitó el sesgo de tener mayores valores por tener mas games jugados.

Características elegidas:

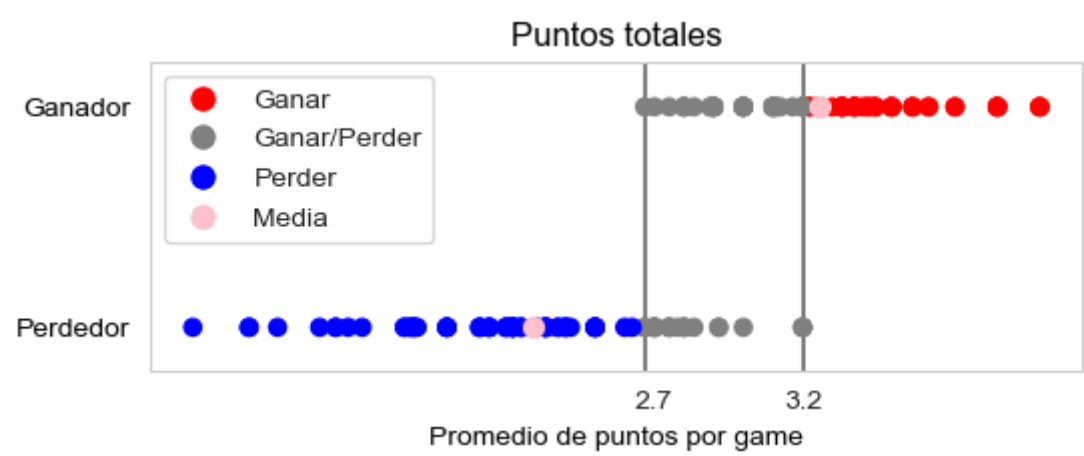
- **Puntos totales por game** – pto\_tot: abarca varias características como los puntos hechos luego de 3, 5 o mas golpes, winners, remates, además que, elimina las condiciones de la cancha, siendo canchas rápidas tendientes a puntos más rápidos y las lentas a puntos mas largos.
- **No forzados por game** – nf: Son los errores que comete el jugador sin que el otro equipo tenga mérito por el error.
- **Puntos netos por game** – pts\_netos: la resta entre los puntos que hizo el jugador y sus errores no forzados.
- **Resultado** – result: 1 para el ganador y 0 para el perdedor

Correlaciones

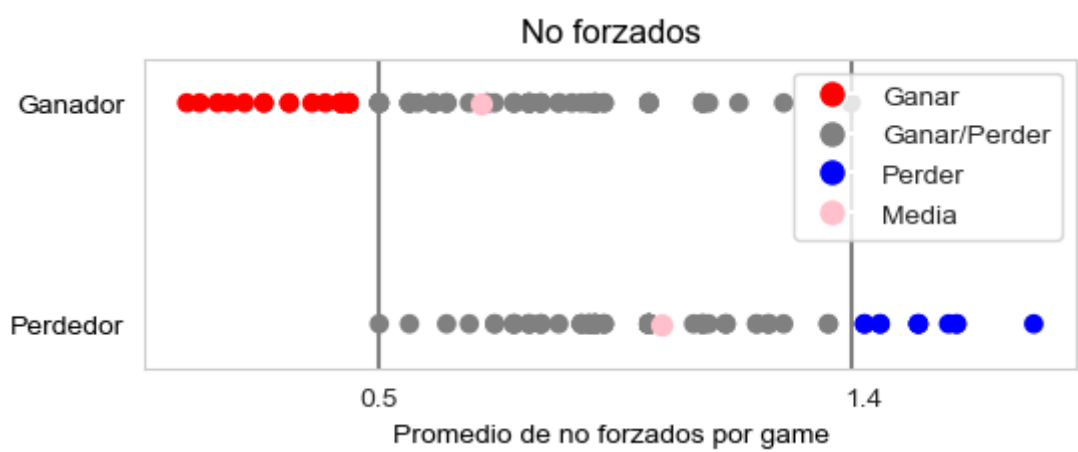
	nf	pto_tot	pts_netos
result	-0.536903	0.784272	0.775189

Las 3 características influyen en el resultado, siendo los errores no forzados inversamente proporcionales a las probabilidades de ganar

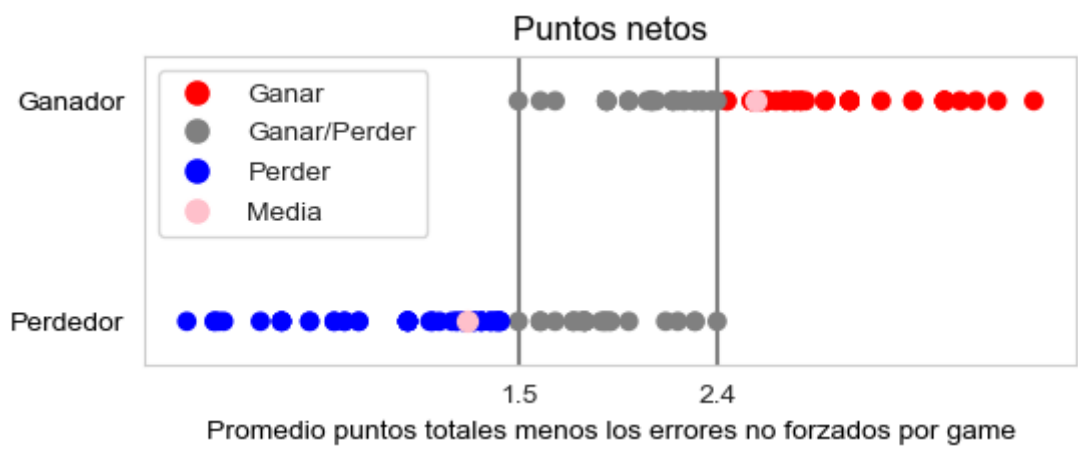
A partir del nuevo dataset con las 4 características se analizaron los datos según se haya perdido o ganado el game, evidenciando cuantos puntos y errores no forzados suelen hacer los que ganan y pierden.



- Pareja que promedie más de 3.2 puntos por game, gana el partido el 100% de las veces. Media 3.25
- Pareja que promedie menos de 2.7 puntos por game pierde el 100% de las veces. Media 2.29



- Pareja que promedie menos de 0.5 errores no forzados por game, gana el partido el 100% de las veces. Media 0.69
- Pareja que promedie más de 1.4 errores no forzados por game, pierde el partido el 100% de las veces. Media 1.02



- Pareja que promedie más de 2.4 puntos netos por game, gana el partido el 100% de las veces. Media 2.56
- Pareja que promedie menos de 1.5 puntos netos por game pierde el 100% de las veces. Media 1.27

Un dato importante es que los puntos hechos por los jugadores tienen en cuenta los errores no forzados del rival, por lo que, un equipo que cometa muchos errores no forzados, hará que el otro equipo, sin hacer gran cantidad de puntos por mérito propio, tenga un promedio alto de puntos totales y puntos netos. Es la alta cantidad de errores no forzados lo que complica juntar a dos jugadores con muchos puntos que suelen hacer muchos errores no forzados

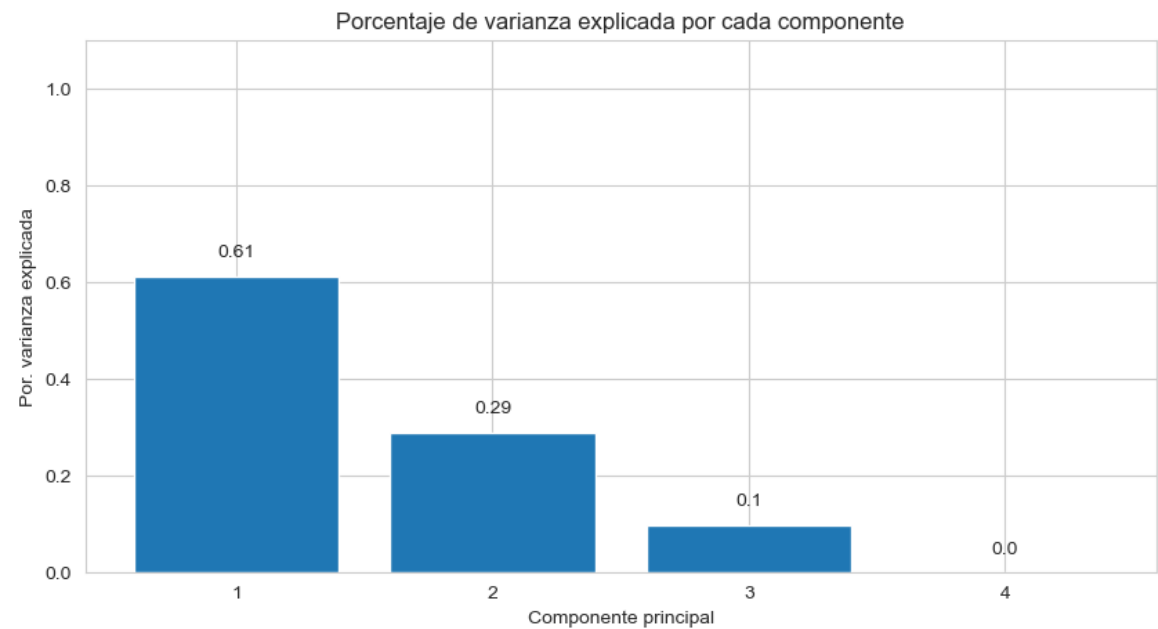
Estos datos podrían ser tenidos en cuenta a la hora de armar las parejas, para el entrenamiento y para analizar la performance y claves del partido, de manera muy superficial

### Elección de modelos

Para agrupar a los jugadores según su estilo de juego y performance, se utilizó un modelo de machine learning no supervisado, el cual confirmó que hay 3 tipos de estilo de juego para los jugadores.

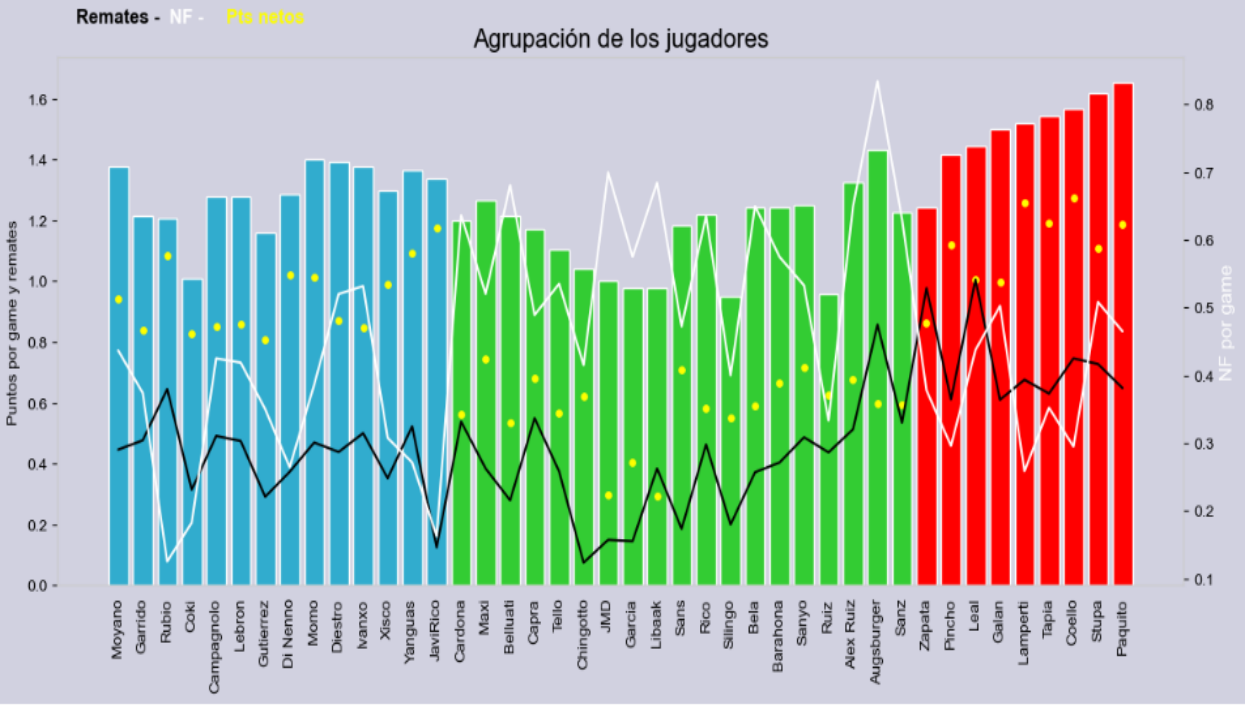
Se llevó a cabo mediante el **KMeans**, eligiendo el número de clúster igual a 3 luego de realizar el método del codo

Al realizar el PCA los primeros 2 componentes de 4 explicaban el 90% de la varianza por lo que se entrenó con dichos componentes, descartando los otros 2



La métrica para evaluarlo fue de silhouette, que fue de 0.29 para un k=3. Las valores de k anteriores y posteriores eran casi idénticos, por lo que dado el análisis previo del deporte, agrupar en 3 grupos a los jugadores fue lo más conveniente

### Agrupación de los jugadores según sus características



En el gráfico de arriba se observan los 3 grupos que agrupan a los jugadores según sus estadísticas por game que juegan; estadísticas que son un reflejo del rol del jugador o de su performance en los partidos analizados. Las barras indican el promedio de puntos que hacen, la línea blanca los errores no forzados y la línea negra los remates que realizan.

Descripción de los grupos:

El grupo 1 y 2 tienen como similitud rematar menos y hacer menos puntos totales que el grupo 3.

**Grupo 1, celeste, oportunistas:** son jugadores que presentan valores de puntos y remates por game intermedios entre los 3 grupos. Suelen cometer pocos errores no forzados. Por estas 2 razones es que la cantidad de puntos netos es alta. En cuanto al rol en el partido, la mayoría son jugadores que arman el juego y que también definen los puntos; o que definen los puntos pero no rematan tanto como el grupo 3

**Grupo 2, verde, defensivos:** son jugadores que estadísticamente se diferencian por la alta cantidad de errores no forzados que cometen. La mayoría realiza pocos remates, aunque hay algunos jugadores que rematan frecuentemente. Suelen ser de los que menos puntos por game realizan por lo que tienen valores bajos de puntos netos. En cuanto a su rol en el partido suelen ser jugadores que arman el juego para que defina su compañero; o jugadores que no se encuentran en un alto nivel, hecho que se ve influido por la cantidad de partidos analizados y contra cuál pareja jugaron cuando se analizó.

**Grupo 3, rojo, definidores:** son los jugadores que realizan muchos puntos y remates junto con pocos o moderada cantidad de errores no forzados, por lo que suelen tener alta cantidad de puntos netos. Su rol en el partido es de definición del punto

Análisis de probabilidades de ganar

**Regresión logística:** obtuvo el puntaje mas alto de accuracy para obtener probabilidades, con un 0.94

	precision	recall	f1-score	support
0	0.94	0.94	0.94	16
1	0.94	0.94	0.94	17
accuracy			0.94	33
macro avg	0.94	0.94	0.94	33
weighted avg	0.94	0.94	0.94	33

Métrica

La métrica elegida fue **accuracy** debido a que el valor del modelo es por error y no por el tipo de error, ya que en este deporte equivocarse no trae consecuencias distintas según se haya predicho triunfo o derrota.

Análisis de las top 5 parejas

	pts_totales	pts_netos	no forzados	probabilidad	ranking	grupo
Tapia-Coello	3.095960	2.442761	0.653199	0.865	1	3 - 3
Galan-Lebron	2.791439	1.871784	0.919655	0.430	2	3 - 1
Stupa-Di Nenno	2.855037	2.066339	0.788698	0.591	3	3 - 1
Momo-Sanyo	2.686124	1.790843	0.895281	0.330	4	1 - 2
Paquito-Chingotto	2.658249	1.781066	0.877183	0.313	5	3 - 2

- En la característica “grupo”, el primer número es el grupo del primer jugador nombrado en la pareja, y el que juega en la posición izquierda, lo mismo con el segundo nombre y el segundo número, es el jugador de la derecha

La **pareja número 1 del ranking** tuvo un comienzo de año histórico, solo perdiendo un partido, por lo que la diferencia en la probabilidad con el resto es lógica. Es la única pareja que juega con dos jugadores que tienen el mismo estilo de juego, situación que es difícil de lograr ya que “compiten” por hacer el mismo punto. Además de hacer tantos puntos ambos, su otra virtud es los pocos



errores no forzados que cometen los 2, lo cual suele ser una debilidad cuando se juntan dos jugadores que son definidores

Las **parejas 2 y 3 del ranking** tienen el mismo estilo de juego 3-1, siendo los jugadores de izquierda definidores y los dos jugadores de derecha del grupo 1, oportunistas, que suelen armar el juego y definirlo cuando tienen la oportunidad. Siendo uno, Lebron, de los mejores jugadores del circuito, y el otro, Di Nenno, de los mejores defensores

La **pareja 4 del ranking** tiene un jugador del grupo uno, Momo con las típicas características de los oportunistas, y el otro jugador, Sanyo, que es penalizado por la cantidad de errores no forzados, siendo su estilo de juego tanto del 1 como del 3, pero promediando estar en el 2 por los errores no forzados

La **pareja 5 del ranking** tiene un estilo distinto a las demás, teniendo un jugador de izquierda, Paquito, netamente definidor, y un compañero que es de los mejores defensores y armadores de juego, sirviéndole el punto a su compañero. Luego de un mal comienzo mejoraron notoriamente, escalando posiciones en el ranking.

Conformación de nuevas parejas

A partir de sus datos individuales, se generaron nuevas parejas ordenándolas de mayor a menor probabilidad de ganar el set. Las 20 primeras parejas son:

	NF_total	pts_ganados	pts_netos	probabilidad	grupo
Lamperti-Coello	0.572173	3.092864	2.520691	0.887	3 - 3
Paquito-Coello	0.764272	3.184202	2.419930	0.877	3 - 3
Tapia-Coello	0.653199	3.095960	2.442761	0.865	3 - 3
Stupa-Coello	0.816407	3.147830	2.331422	0.840	3 - 3
Pincho-Coello	0.594590	2.984335	2.389745	0.816	3 - 3
Galan-Coello	0.792485	3.082584	2.290099	0.805	3 - 3
Momo-Coello	0.672026	2.991904	2.319878	0.789	1 - 3
Leal-Coello	0.716584	3.002467	2.285882	0.777	3 - 3
Lamperti-Yanguas	0.573044	2.900074	2.327030	0.761	3 - 1
Lamperti-JaviRico	0.450664	2.826376	2.375712	0.759	3 - 1
Moyano-Coello	0.760345	2.990714	2.230368	0.746	1 - 3
Paquito-Yanguas	0.765143	2.991412	2.226269	0.744	3 - 1
Paquito-JaviRico	0.642763	2.917713	2.274950	0.741	3 - 1
Lamperti-Rubio	0.424194	2.782258	2.358065	0.733	3 - 1
Tapia-Yanguas	0.654069	2.903170	2.249100	0.722	3 - 1
Tapia-JaviRico	0.531689	2.829471	2.297782	0.719	3 - 1
Paquito-Rubio	0.616292	2.873596	2.257303	0.715	3 - 1
Ivanxo-Coello	0.831960	2.978082	2.146121	0.695	1 - 3
Tapia-Rubio	0.505219	2.785354	2.280135	0.692	3 - 1
Lamperti-Di Nenno	0.544464	2.800071	2.255608	0.684	3 - 1

Como límite inferior se estableció tener una probabilidad de ganar el set mayor o igual que a 0,430 (43%), ya que es el peor ranking de las top 3, que son las únicas parejas consolidadas; las demás alternan buenos y malos torneos.

Se armaron 97 parejas con las siguientes observaciones:

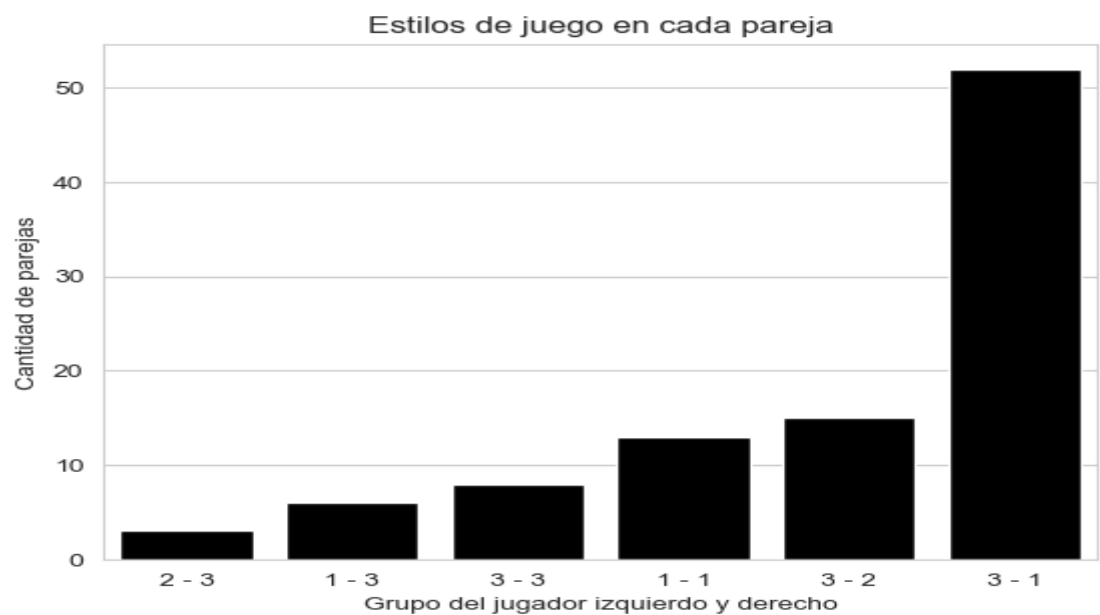
**A nivel jugadores:** las 8 primeras parejas con mayor probabilidad de ganar tenían al jugador Coello como jugador derecho, lo que denota su gran temporada. También se observa que jugadores que no están en el top 10, con un compañero que se complemente con sus estadísticas, pueden tener niveles muy altos de probabilidad de ganar, como es el caso de Leal, Ivaxo y JaviRico

Jugadores que mas se repiten:

Coello: 17,Paquito: 14,Lamperti: 14, Tapia: 12,Stupa: 11,Galan: 8



A nivel pareja: distribución de los estilos de juego



Las 6 primeras parejas tenían a jugadores 3-3, es decir, dos jugadores con estilo de juego de definición. Sin embargo, la mayoría de las parejas tienen una formación 3-1. Esto ocurre, que la mejor pareja probabilísticamente no sea la más frecuente, porque es complicado que 2 jugadores de un mismo equipo hagan gran cantidad de puntos, como para entrar en el grupo de los definidores, ya que “compiten” entre ellos por hacer el punto. También, hay más jugadores derechos que zurdos, siendo más complicado rematar desde la posición derecha si sos derecho. Otro punto a tener en cuenta es que cuando se juntan dos jugadores con estilo de definidores, al no tener la vocación defensiva y de armar el juego, como los del grupo 1 y 2, suelen cometer más errores no forzados; que tenían una correlación negativa con el resultado de -0.53

Conclusiones

Encontramos mediante el análisis del deporte y la aplicación de la IA, que en el padel se pueden distinguir tres tipos de estilos de juegos: oportunista (grupo 1), defensivo (grupo 2), y definidor (grupo 3). De estas, la combinación mas frecuente es la 3-1, teniendo al jugador izquierdo como definidor y al derecho como oportunista. La combinación más efectiva, estadísticamente hablando, es la 3-3, pero en la realidad es difícil de lograr que funcione esta combinación porque requiere que ambos jugadores hagan muchos puntos y remates, y pocos errores no forzados.

Confirmamos que la combinación más efectiva es la 3-3, con 2 jugadores definidores, con los resultados que obtuvo la pareja 1 del ranking, logrando estar casi medio año invicta, solo perdiendo un partido. Sin embargo, se han juntado otros jugadores con estilo de juego de definidores y suelen quedarse los dos, o uno de los dos, clasificados como del grupo 2, por el alto nivel de errores no forzados y los pocos puntos totales y netos ya que “competían” entre ellos por los puntos de su equipo

Es importante aclarar que casi todas las combinaciones de estilos pueden ser efectivas para lograr una buena temporada a nivel elite, ya que, por ejemplo, en el top 5 quedaron representadas la mayoría de las combinaciones: tener a un definidor con otro definidor, oportunista o defensivo, y tener a un oportunista con un defensivo. La pareja con un estilo que no fue representada es la 2-2, que sería tener dos jugadores netamente defensivos, pero si hay una pareja dentro del top 10 que está clasificada como 2-2 pero es por el alto nivel de errores no forzados y los pocos puntos que hacen cada uno, ya que ambos tienen rol de definidores.

Acerca de la posibilidad de cambiar de parejas, teniendo en cuenta que casi todos los estilos de juego pueden tener éxito a nivel de elite, el poder cambiar de compañero varias veces en el año tiene su lado bueno y malo. Es beneficioso ya que con una probabilidad de ganar el set de 0.430(43%) esa pareja podría llegar a estar en el top 5, lo que le significaría ganar premios y mejores sponsors. Sin embargo, pocas parejas han jugado a su mejor nivel, y lo han mantenido, en sus primeros partidos juntas, siendo el cambio un arma de doble filo.

Quienes sí se benefician de los cambios son el deporte y el público. El deporte se va renovando constantemente, y se pueden ver uniones de jugadores de todo tipo, siempre es pos de mejorar. Esto, a su vez, renueva el desafío a los que estén mejor en el ranking, quienes tienen que ir mejorando también porque las parejas de abajo no dejarán de realizar cambios que los acerquen a ellos.

**Limitaciones del estudio**

- La cantidad de partidos analizados para cada pareja/jugador es diversa, habiendo parejas con 1 o 2 partidos, o que tuvieron un torneo muy bueno o muy malo, sobredimensionando o subestimando sus estadísticas
- Contra qué pareja fueron los partidos analizados
- Los distintos roles de los jugadores dentro de la pareja en donde no es lo mismo “competir” por hacer el punto con tu compañero que tener un compañero que tienda a armar el punto para que el otro defina.