

Commander Exploratory Data Analysis

¿Más cartas mejor juego?

Itsazain M. Bilbao

Índice

Índice	2
Introducción y Contexto	3
¿Qué es Magic the Gathering?	3
¿Qué es Commander o EDH?	3
Datos y enfoque de Análisis	4
Limpieza Inicial de Datos	4
AnálisisCantidad.ipynb	5
Arquetipos y Sesgos	5
Preparación de Mazos	5
AnálisisCasosEspeciales.ipynb	6
Graficos.py	6
Resultados y Hallazgos Clave	7
Estudio por Colores	8
Conclusiones y Oportunidades	8

Introducción y Contexto

Para comprender de qué va el análisis y entender las conclusiones finales hay que entender lo que es Magic: The Gathering.

¿Qué es Magic the Gathering?

Es un juego de cartas coleccionables o TCG (trading card game) de más de treinta años de antigüedad. Como juego tiene varios formatos y como coleccionable existe en varios niveles de profundidad. Puede verse tanto como un mercado de valores o como objetos de coleccionismo, donde cada carta tiene tanto un valor monetario único como una función única dentro del juego. Y cada año salen varias colecciones que expanden el juego y su mercado.

¿Qué es Commander o EDH?

Es uno de los muchos formatos de Magic, como casi todos, nació de la comunidad para alejarse del formato oficial principal de su época. En este caso, fue un formato no oficial que surgió en los propios torneos entre aquellos que eran eliminados temprano o tenían que esperar a su ronda. Es un formato multijugador, de entre tres y cuatro jugadores, aunque pueden jugar hasta seis, y tiene sus propias diferencias a la hora de crear el mazo y sus propias reglas del formato.

Las principales diferencias al crear el mazo son las siguientes:

- Mazos de exactamente 100 cartas únicas, salvo tierras básicas (un tipo de carta).
- Todos los mazos tienen mínimo un comandante, o dos si tienen partner. Estos entran dentro del límite de cartas, pero están en su propia zona y son siempre jugables desde esa zona. (No te tienen que tocar del mazo)
- Los colores del comandante definen el color de las cartas que puedes llevar.
- Se empieza a 40 vidas en vez de 20.
- Se juegan cartas de todas las ediciones, salvo algunas pocas prohibidas (Casi 30.000 cartas disponibles).

Esto último lleva a que, a diferencia de otros formatos, la cantidad de estrategias y opciones sea prácticamente infinita. También hace que las nuevas opciones tengan más dificultad para encontrar juego, ya que sus competidores son cartas conocidas que pueden llevar décadas en mazos.

Y esto es precisamente lo que vamos a analizar, si las cartas nuevas son utilizadas.

Datos y enfoque de Análisis

Antes de comenzar con el análisis hay que definir las fuentes de datos. Las principales fuentes han sido Scryfall, una biblioteca de todas las cartas de Magic, que posee su propia API y que es la solución principal a la hora de buscar y conocer cartas. En `api_request.py` está el código que descarga en `.csv` toda la base de datos de cartas. A pesar de todas las columnas, las relevantes para el análisis son aquellas que ayudan en la limpieza de datos, las fechas de publicación de las cartas, si son reimpresiones, su color y su rango en `edhrec`. La limpieza se lleva a cabo en `PreparacionDFC.ipynb`. No es que la base de datos contenga muchos nulos, es más que nada porque gran parte de los datos son redundantes y para analizar las cartas de los mazos tener varias cartas con el mismo nombre de diferentes ediciones podría dar problemas.

Otras fuentes de datos son Moxfield, que es de donde han salido los mazos competitivos para el análisis y luego EDHrec que, a pesar de no haber sacado información directamente, contiene el rango de popularidad de las cartas y es de donde lo toma Scryfall.

Limpieza Inicial de Datos

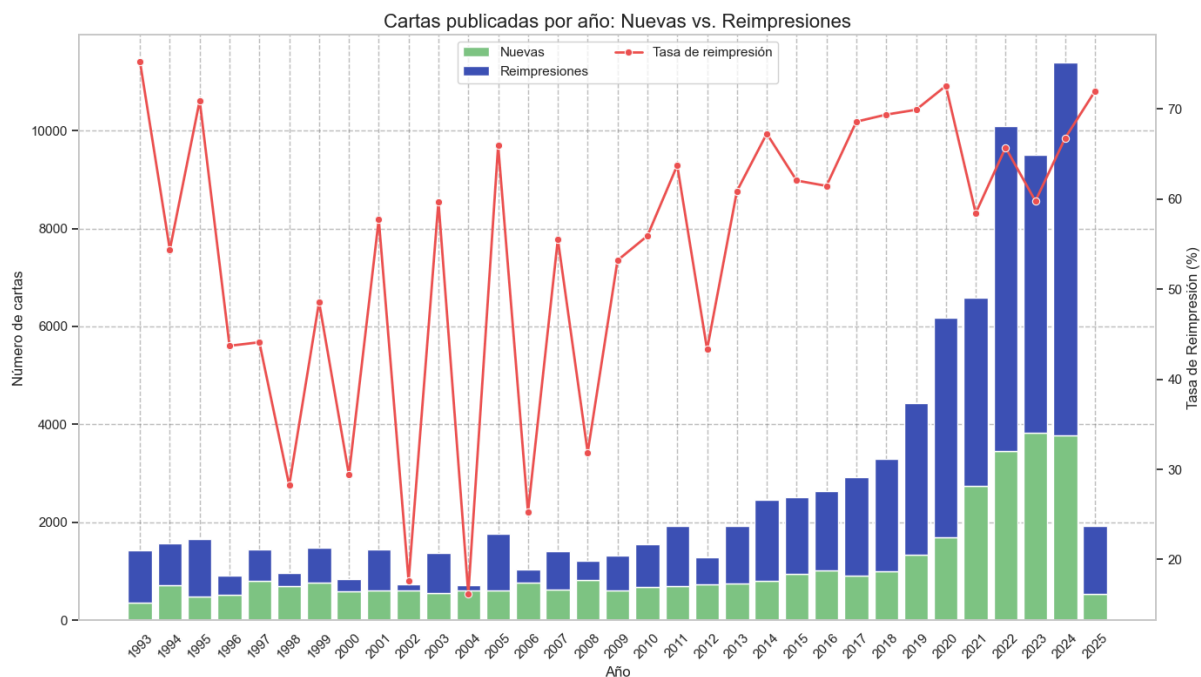
Igualmente, para el análisis de cantidad y reimpresiones, que está en `AnalisisCantidad.ipynb`, se utiliza el `.csv` descargado con la limpieza mínima, simplemente quedarse con aquellas que son relevantes para el formato o la construcción de mazo. Filas que se eliminan, por ejemplo, independientemente de qué queramos analizar sólo queremos aquellas que en `legality` contengan que son legales en el formato `commander`. También, las tierras básicas, a pesar de que son imprescindibles en la construcción de un mazo, son ruido en el análisis, ya que todas las tierras básicas (salvo los yermos) salieron en la primera edición y en total hay 6 tierras básicas.

También, cartas que aparecen como legales pero no son parte de la construcción del mazo o son consecuencia de alguna interacción. Por ejemplo los emblemas, tokens, los stickers, las cartas fusionadas y demás cartas mecánicas. Por lo general esa limpieza se ha hecho de dos formas, comprobando su tipo y en `"type_line"` y eliminando las coincidan con ese tipo o eliminando aquellas cuyo `edhrec_rank` sea nulo, ya que como no entran en la construcción directa del mazo nadie las puede seleccionar y por lo tanto su rango de popularidad es nula.

Después, se han aplicado varias transformaciones a los datos para su manejo más adelante. La fecha de publicación era un `object (string)` y sólo se quería su año de publicación, por lo que se ha sacado el año de cada carta y añadido en otra columna como `"release_year"`. En caso de los colores igual, para análisis posteriores, se toman las columnas de `"colors"` y `"color_identity"`, se imputan los nulos, y luego se construye otra columna que indica en un `string` categórico el color o colores de la carta.

AnalisisCantidad.ipynb

En este notebook lo único que se hace, con los datos preparados, es contar cuantas cartas se han publicado por año dividiendo aquellas que sean reimpresiones de las que no. Además de un cálculo sencillo que comprueba por año cuál es el porcentaje de reimpresión global o general. De esta forma se puede ver como cada año se imprimen y se reimprimen más cartas.



Arquetipos y Sesgos

Antes de comenzar a analizar si las cartas nuevas se emplean en los mazos hay que estudiar la posibilidad de que haya mazos muy concretos que toman sus cartas de ediciones o colecciones muy específicas, lo que implica posible ruido en el análisis.

Preparación de Mazos

Todo este análisis se hace en AnalisisCasosEspeciales.ipynb con funciones varias en la carpeta utils que se encargan de dibujar los gráficos, preparar los mazos y demás. Para empezar los mazos se descargan desde Moxfield (de forma manual), en un formato .txt. Todos los mazos de Moxfield tienen el mismo formato y en edh_requests.py hay funciones que se encargan de tomar la dirección de un fichero o carpeta y procesar los .txt con formato definido, línea por línea, va construyendo el dataframe básico que consiste en dos columnas: el comandante (Deck Name) y cada carta (Card Name). Después se les aplica una función que, tomando el mazo y la base de datos de cartas limpiada, añade columnas al dataframe. Estas columnas son relevantes para el análisis.

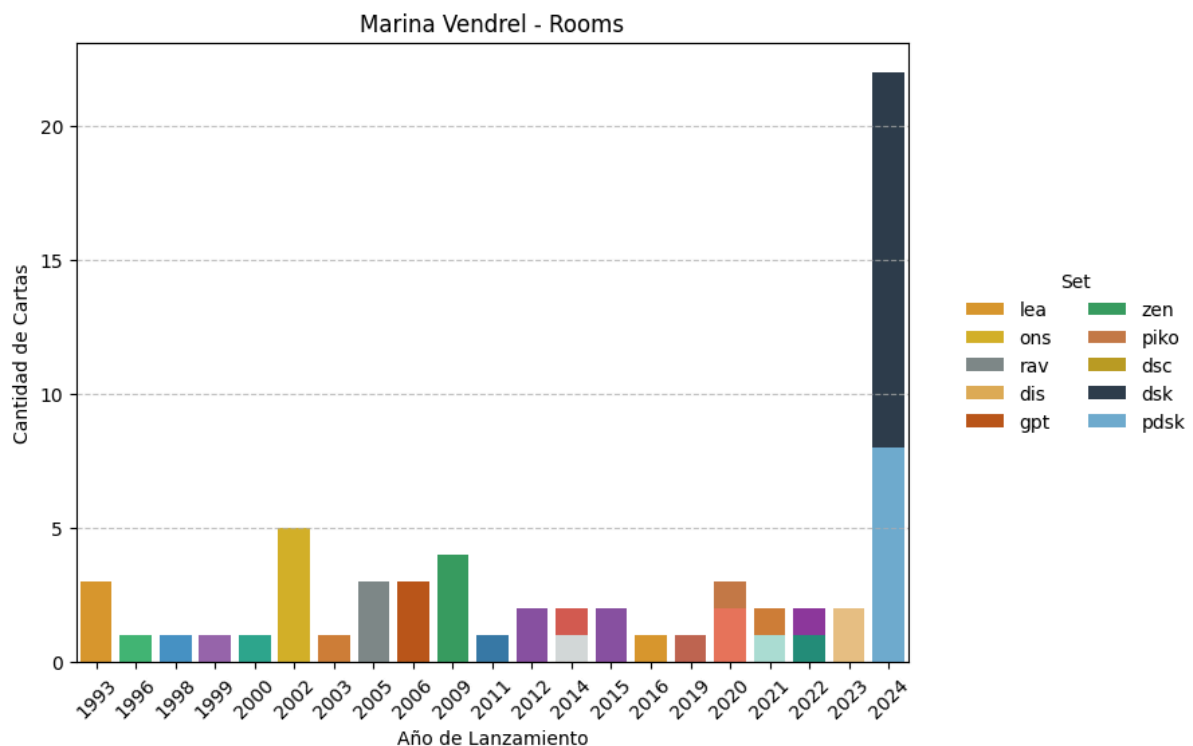
AnalisisCasosEspeciales.ipynb

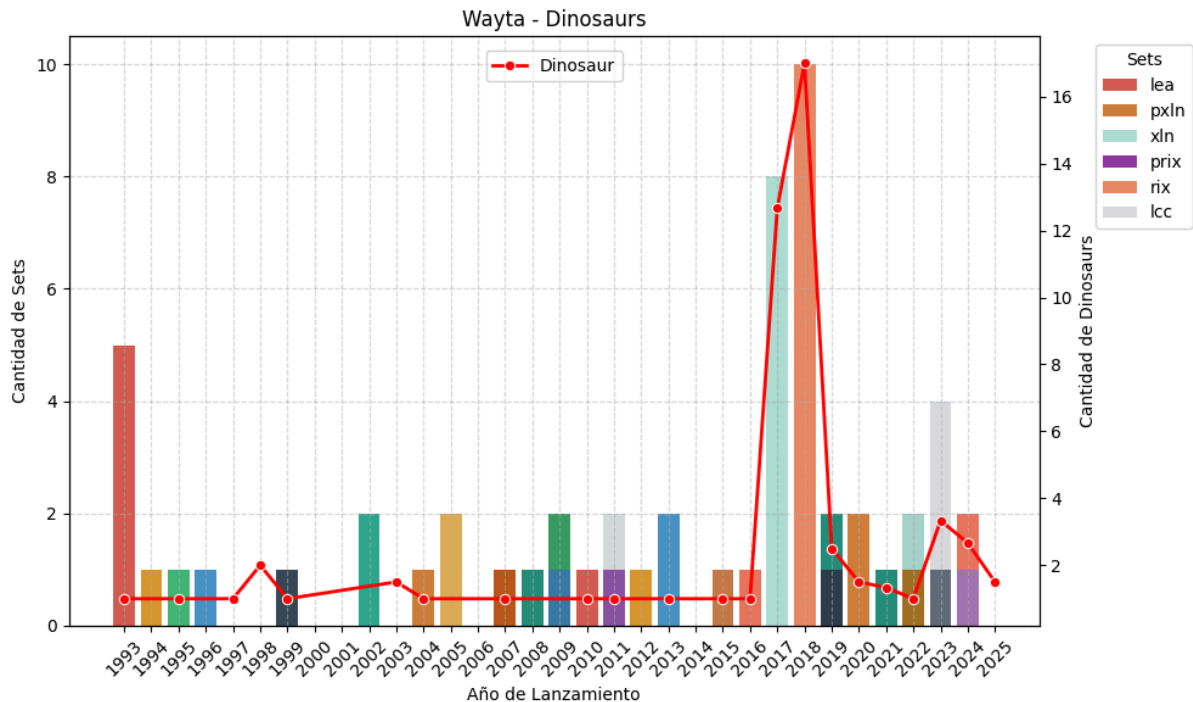
Para los casos especiales se toman dos mazos muy especiales, pero relativamente populares en el entorno casual, y se comprueba que, efectivamente, esos mazos son propensos a tomar cartas de ediciones específicas publicadas en años específicos. Por ejemplo, el de Dinosaurios es evidente que su mayor cantidad de cartas son en las colecciones de Ixalan, mientras que con Marin Vendrel se ve que la mayoría de sus cartas son las que se publicaron con ella en su misma edición en 2024 (La edición en la que salieron las cartas de tipo Puerta/Room, que es el arquetipo que ella emplea).

Graficos.py

En graficos.py existen funciones que sirven para dibujar esas gráficas. Existen dos, una de ellas dibuja la gráfica de barras de cantidad de cartas por año y luego las divide por set. Además, en la leyenda sólo aparecen los Sets que tengan cierta aparición, para que en una lista de tantas cartas no aparezcan cincuenta o más sets en la leyenda.

La segunda función hace lo mismo, pero tiene un añadido que es que dibuja una línea que, dando un tipo, mira todas las cartas en la base de datos de ese tipo dado y muestra la cantidad que se ha impreso por año. Es decir, dibuja una línea que si le dices “Dinosaur”, te dice la cantidad de dinosaurios por año. En caso, como con Marina Vendrel, sólo hay “Rooms” en 2024, sólo dibuja un punto rojo en ese año. Por eso con Marina Vendrel tiro de la función básica.

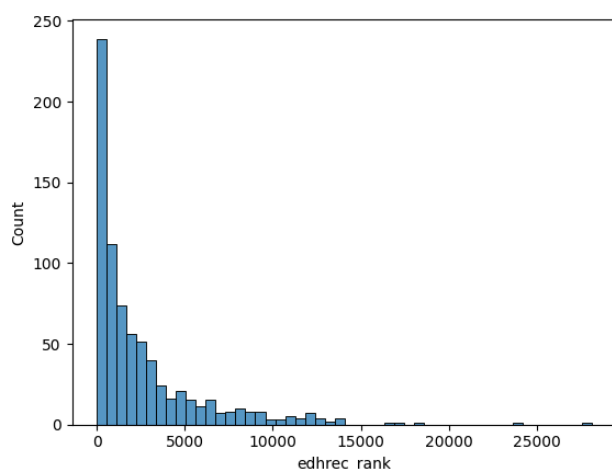




Resultados y Hallazgos Clave

Para seguir con el análisis se han tomado más de 70 mazos competitivos que han quedado finalistas en algún torneo en los últimos años y se han unido en un dataframe. De aquí, al igual que con los casos especiales, se les ha añadido columnas relevantes para el análisis y también se ha hecho una copia a la hora de eliminar las cartas duplicadas (ya que diferentes mazos pueden tener cartas en común). Esto es para comprobar la cantidad de cartas únicas por año publicado y también la cantidad de cartas en general que se usan de ese año.

Luego, para estudiar más a fondo el impacto de esas cartas se toma en cuenta el `edhrec_rank` que mide la popularidad (cuanto más bajo más popular), es decir, cuantos

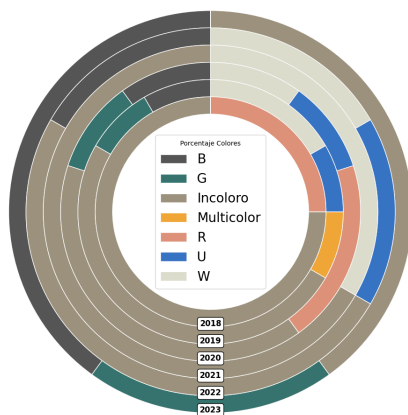
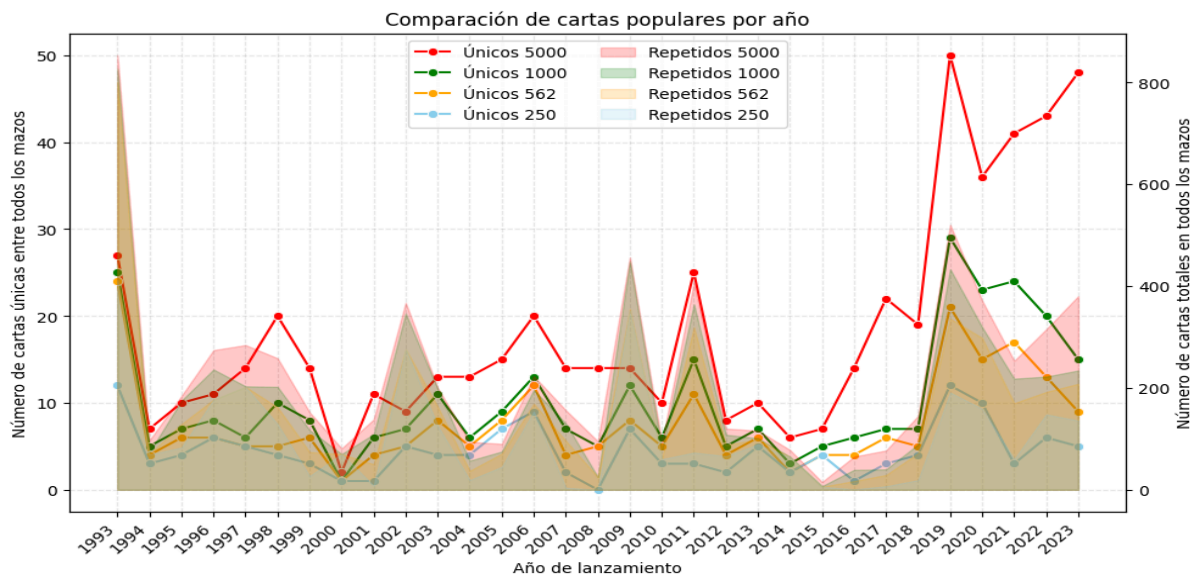


mazos utilizan esas cartas. Sin embargo, no se puede tomar un umbral de forma aleatoria, para eso hay que estudiar la distribución del rango por cartas. En el gráfico se ve claramente como el primer bin está repleto de cartas y que a medida que son menos populares hay menos cartas. Lo cual es normal cuando se trabaja con rangos, uno espera que haya más cartas populares que impopulares.

Se calcula cuánto mide cada bin y luego se buscan más umbrales lógicos con los cuartiles y se decide comprobar la cantidad de cartas que se usan por año

que estén dentro de diferentes umbrales. Esto da una visión bastante amplia de la distribución de las cartas. Además, la distribución se comprueba con cartas únicas y en

general (es decir, con cartas duplicadas). De esta forma se ve claramente como hay gran cantidad de cartas únicas a partir de 2019 (que tiene sentido, pues en 2019 oficializaron commander como formato y sacaron cartas exclusivas para el formato), pero que luego en general gran parte de las cartas son de las ediciones antiguas.



Estudio por Colores

Luego, para profundizar más en el motivo por el cual podrían ser populares se decide estudiar los colores de esas cartas de los últimos años. Para poder sacar conclusiones más concisas habría que analizar carta por carta, pero con conocimientos generales del juego y un análisis básico se puede concluir que las cartas que encuentran hueco esos años son aquellas que tienen uso general o son de colores que de normal necesitaban un impulso para funcionar mejor.

Conclusiones y Oportunidades

Las conclusiones generales que han salido son las siguientes:

- Es cierto que cada vez se imprimen y se reimprimen más cartas. Las tasas de reimpresión son altas, pero lo han sido siempre en proporción. A pesar de que es cierto de que hay años que apenas tienen reimpressiones suele ir dando saltos y habría que analizarlo con más detalle para comprobar si esas cartas reimpresas son del año anterior, del mismo o mucho más atrás.
- Se demuestra que las cartas nuevas ven juego tomando como ejemplo mazos competitivos. A pesar de que se sigan utilizando más las cartas antiguas, es cierto que se incorporan más cartas únicas nuevas en general, aunque no es que sean

muy populares. Eso se puede deber a que tampoco llevan tanto tiempo publicadas y no ha habido tiempo para que se creen o actualicen los suficientes mazos.

- Y el color de las cartas nuevas incorporadas es incoloro, que es un color general que puede utilizarse en cualquier mazo, o de colores poco versátiles que han visto incrementadas sus opciones estos últimos años.

Se podría profundizar más en el tema, analizando cada carta y mirando estadísticas para desarrollar y descubrir más información y sacar más conclusiones, pero eso requeriría tratar en más profundidad el diseño de juego y salud del metajuego.