

IT ACADEMY - DATA SCIENCE PROJECT

Adrià Guerrero, Fernando Nogueira

Abstract (Resumen)

El mercado de compra y venta de criptopunks, figuras digitales clasificadas como cripto arte, ha tenido una “explosión” mediática trás las compras millonarias de algunas personas públicas y grandes empresas. Según algunas fuentes especializadas en criptomercados, el valor percibido de estas figuras es la representación de su pionerismo en el mundo del cripto arte y de las NFT’s (“*non fungible tokens*”). Estas figuras se lanzaron a la red de criptografía “Ethereum” en 2017. En el presente informe se aborda la determinación del valor específico de cada punk en base a sus características con el objetivo de evaluar una estrategia de inversión en este mercado que produzca beneficios económicos. Para ello se ha buscado y almacenado datos relativos a las características de cada “Punk” y sus movimientos de compraventa. Una vez obtenidos los datos se han realizado un análisis uni y multivariante de los datos para la detección de los elementos que mayor condicionamiento tienen sobre el valor económico de los “Punks” y evaluados por un algoritmo de “Machine Learning” no supervisado para la detección de grupos comunes. La característica más definitoria el valor económico de los punks es su tipo generando dos grupos bien diferenciados según su rareza y valor económico. El primer grupo formado por punks tipo “Alien”, “Ape” y “Zombie” (AAZ) contiene los tipos más exclusivos (<1,5%) y por ende con un valor económico más elevado; y el segundo, formado por los punks de tipo Humano (“Male y Female” >98,5%). Dentro de este segundo grupo no se ha podido determinar con exactitud la influencia de las diferentes características en relación a su valor económico, aunque sí que se han detectado ciertas tendencias. en función de su piel y alguno de sus atributos, siempre motivados por la escasez de estos frente a otros.

Índice

Introducción	3
PARTE 1	
Cryptopunks, Bienes Digitales y Blockchain	4
1.1 - Qué es un Cryptopunk	4
1.2 - Qué es el Blockchain	4
1.3 - Red Ethereum	5
1.4 - Crypto Arte	5
1.5 - Características de los CryptoPunks	5
1.6 - Clasificación de los Cryptopunks	6
PARTE 2	
Recolección y Almacenamiento de Datos	8
2.1 - Adquisición de datos vía Web Scraping	8
2.2 - Almacenamiento de Datos	8
2.2.1 Diseño de la base de datos:	8
PARTE 3	
Análisis Exploratorio y Tratamiento de Datos Preliminar	10
3.1 - Consideraciones previas	10
3.2 - Características de los Punks:	10
3.2.1 - Tipos y Pieles:	10
3.2.2. - Atributos:	13
3.2.2.1 - Número de atributos:	13
3.2.2.2 - Atributos más exclusivos:	14
3.3.3 - Índice de rareza (“Rarity”):	15
3.3.4 - Análisis de características internas del grupo AAZ	17
3.3 - Características de los Propietarios	18
3.4 Evolución de ventas de los Cryptopunks.	19
3.4.1 - Número de ventas:	19
3.4.2 - Precio de las ventas:	20
PARTE 4	
Machine Learning	23
4.1 Introducción	23
4.2 Data Frame	23
4.3. Ingeniería de variables.	24
4.3.1 - Tratamiento de variables categóricas:	24
4.3.2 - Tratamiento de outliers	24
4.3.3 - Estandarización	24

4.3.4 - PCA	25
4.4 - Modelos de clusterización	25
4.4.1 - Modelo inicial.	25
4.4.1 - Iteraciones de mejora del modelo.	27
4.4.1 - Resultados.	27
PARTE 5	
Conclusiones	29

Introducción

Este trabajo ha sido desarrollado como proyecto final del curso de Data Science de IT Academy (Barcelona Activa), con el objetivo de emular un proyecto de Data Science en un contexto real.

En el proyecto se realiza un informe con recomendaciones para un gran inversor que desea aplicar recursos en Cryptopunks, un activo digital negociado en el blockchain de Ethereum.

Usando técnicas de recolección (scraping), almacenamiento (BBDD), exploración y visualización de datos, juntamente con técnicas de Machine Learning, prioritariamente de clustering, se tratarán de identificar grupos y/o características con mayores posibilidades de apreciación futura en sus precios.

PARTE 1

Cryptopunks, Bienes Digitales y Blockchain

1.1 - Qué es un *Cryptopunk*

Un *Cryptopunk* es un objeto de arte original, de tipo digital, garantizado por un contrato digital que tiene su veracidad autenticada por una estructura digital de datos llamada *Blockchain*.

Un *Cryptopunk*, como cualquier objeto de arte, puede ser negociado entre particulares. En su caso, los *Cryptopunks* hacen uso de la red Ethereum y su criptomoneda ETH (Ether) para valorar su precio.

1.2 - Qué es el Blockchain

Una cadena de bloques¹, conocida en inglés como “*blockchain*” es una estructura de datos cuya información se agrupa en conjuntos (bloques) a los que se les añade meta-informationes relativas a otro bloque de la cadena anterior en una línea temporal. De esta forma, gracias a técnicas criptográficas, la información contenida en un bloque solo puede ser repudiada o editada modificando todos los bloques posteriores. Esta propiedad permite su aplicación en un entorno distribuido de manera que la estructura de datos *blockchain* puede ejercer de base de datos pública no relacional que contenga un histórico irrefutable de información.

El *blockchain* se considera una tecnología en la que la "verdad" (estado confiable del sistema) es construida, alcanzada y fortalecida por los propios miembros; incluso en un entorno en el que exista una minoría de nodos en la red con comportamiento malicioso (nodos sybil) dado que, en teoría, para comprometer los datos, un atacante requeriría de una mayor potencia de cómputo y presencia en la red que el resultante de la suma de todos los restantes nodos combinados. Por las razones anteriores, la tecnología blockchain es especialmente adecuada para escenarios en los que se requiera almacenar de forma creciente datos ordenados en el tiempo, sin posibilidad de modificación ni revisión y cuya confianza pretenda ser distribuida en lugar de residir en una entidad certificadora.

¹ https://es.wikipedia.org/wiki/Cadena_de_bloques#cite_note-1

1.3 - Red Ethereum

La red Ethereum² es una tecnología inspirada en las innovaciones del Bitcoin³, que permite hacer transacciones de criptomonedas entre personas por una pequeña tarifa. En este caso la red Ethereum es programable, por lo que no solo se utiliza como red de pagos si no que también puede como mercado de servicios financieros, juegos y aplicaciones, certificando el origen y la autenticidad⁴.

1.4 - Crypto Arte

Emergiendo como un género de nicho de trabajo artístico después del desarrollo de redes blockchain como Bitcoin y Ethereum a mediados y finales de la década de 2010, el arte criptográfico creció rápidamente en popularidad en gran parte debido a la capacidad sin precedentes que ofrece la tecnología subyacente para que las obras de arte puramente digitales sean compradas, vendidas o recolectadas por cualquier persona de manera descentralizada.

Si bien no hay una definición acordada para el término, actualmente existen dos interpretaciones comunes entre los artistas criptográficos y sus coleccionistas. El primero, con respecto a las obras de arte con temas criptográficos, o aquellos con temas que se centran en la cultura, la política, la economía o la filosofía que rodea a la tecnología blockchain y criptomonedas. La segunda definición, y más popularizada, incluye obras de arte digitales que se publican directamente en una cadena de bloques en forma de un token no fungible (NFT), que hace posible la propiedad, transferencia y venta de una obra de arte de una manera criptográficamente segura y verificable⁵.

1.5 - Características de los *CryptoPunks*

Los *CryptoPunks*⁶ son una colección de 10.000 imágenes de 24 x 24 píxeles que muestran retratos de personajes ficticios creados aleatoriamente por un algoritmo a partir de muestran diferentes características. (Raza, Sexo, Color de piel, Accesorios, etc.).

El resultado es un conjunto de expresiones artísticas⁷ ligadas a un código blockchain permitiendo identificar al propietario de la obra. Por lo tanto, pueden ser compradas y vendidas como cualquier otra pieza de arte.

² <https://ethereum.org/en/>

³ <https://bitcoin.org/es/>

⁴ <https://ethereum.org/en/what-is-ethereum/>

⁵ https://en.wikipedia.org/wiki/Crypto_art

⁶ <https://larvalabs.com/cryptopunks>

⁷ <https://mashable.com/article/cryptopunks-ethereum-art-collectibles>

Fueron creados en junio de 2017 por el Estudio americano Larva Labs (Link), formado por Matt Hall y John Watkinson, siendo uno de los primeros “NFT’s”⁸ de la red Ethereum⁹.

Dentro de la individualidad de cada obra, existen varios tipos de *CryptoPunks*. Cada uno es diferente respecto a los otros, aunque en muchos casos pueden compartir una o varias características. Dependiendo de la característica, puede ser más frecuente o más rara. Por ejemplo, clasificándolos por Tipo, los más raros son los Aliens, Simios (apes) y Zombies. Según palabras de los creadores. Esta colección fue concebida como un experimento para evaluar algunas de las dinámicas de la exclusividad y la demanda.

Durante su lanzamiento, cada Punk podía ser adquirido gratuitamente, a un precio simbólico, por cualquier individuo con una cartera en Ethereum agotando su disponibilidad rápidamente.

Actualmente pueden ser adquiridos a través del marketplace de Ethereum¹⁰, en el cual los Punks pueden ser vendidos, ofertados o pujados.

La pertenencia a la red Ethereum certifica la autenticidad de cada fichero e impide que puedan añadirse más piezas a la colección.

1.6 - Clasificación de los Cryptopunks

Los Cryptopunks se dividen por:

- Tipos (5): Es una característica excluyente, es decir, cada punk solo puede ser de un tipo. Los tipos son: Alien - 0,09%, Ape - 0,24% , Zombie - 0,88%, Female - 38,40%, Male - 60,39%.
- Atributos (87): Son características no excluyentes, es decir, un punk puede tener más de un atributo¹¹. Cada atributo tiene diferente grado de rareza, Por ejemplo: Earring - 24.59%, Mole - 6.44%, Hidden Smile - 0.32%. Para este estudio los atributos que aparecen ocultos debajo de otros atributos como puede ser la sonrisa (*smile*) oculta debajo de una barba (*beard*) se consideran atributos independientes con el prefijo *Hidden*, por ejemplo: “*HiddenSmile*”.
- Número de Atributos (0-9): Cantidad de atributos que definen al punk.
- Índice de Rareza o Rarity (1-10.000): Es un índice creado por la web *Defypunks* con la intención de clasificar a los punks según su rareza evaluando la interacción de cada una de sus características. El índice de cada punk se genera comenzando por su característica más rara, ya sea el tipo (alien, ape, zombie, female, male.), el número de

⁸ https://en.wikipedia.org/wiki/Non-fungible_token

⁹ <https://en.wikipedia.org/wiki/CryptoPunks>

¹⁰ <https://etherscan.io/>

¹¹ <https://larvalabs.com/cryptopunks/attributes>

rasgos o *traits* (0 a 7), los rasgos particulares de cada punk (de *beanies* a *earrings*) o el color de la piel (*Skin*).

Ante un empate, el desempate se determina en función de la siguiente característica más rara

PARTE 2

Recolección y Almacenamiento de Datos

2.1 - Adquisición de datos vía Web Scraping

Los datos fueron adquiridos utilizando técnicas de “webscraping” utilizando las librerías de python BeautifulSoup y Selenium.

Para adquirir los datos de los *CryptoPunks* se “escraperon” dos páginas:

- Larva Labs para la obtención de los datos relativos a los:
 - Propietarios (“Owners”): <https://www.larvalabs.com/cryptopunks/leaderboard>
 - Tipos (Types): <https://www.larvalabs.com/cryptopunks/attributes>
- DefyPunk (<https://defypunk.com/punks/0...>):
 - Descripción de cada punk: id, índice de rareza, tipo, color de piel, *traits*, dueño actual.
 - Histórico de transacciones por punk: Id_transacción, Índice de rareza, tipo, piel, atributos (*traits*), *comprador*, *vendedor*, tipo de transacción, vendedor (*from*), comprador (*to*), precio en ETH, fecha de la transacción.

El histórico de precios del ETH en relación al dólar fueron obtenidos de la página:

- <https://www.nasdaq.com/market-activity/cryptocurrency/eth/historical>

2.2 - Almacenamiento de Datos

Los datos generados por el *webscrapping* se modificaron posteriormente para adecuarlos para ser guardados en una base de datos relacional tipo SQL y guardados en diferentes documentos .CSV.

La herramienta para diseñar y generar la base de datos ha sido MySql Workbench por su practicidad y amplio uso en el mercado.

2.2.1 Diseño de la base de datos:

Para elaborar la base de datos, en el modelo relacional se ha generado una tabla para cada uno de los diferentes scrapings realizados (**Fig. 1**).

De este conjunto de tablas las dos principales son *punks* y *transactions*.

La tabla *punks* recoge las características individuales de cada uno de los 10000 punks (*id, rarity, type, skin, traits, owned by*). La tabla *transactions* recoge el registro de todos los movimientos (ofertas, contraofertas y ventas) de cada uno de los punks desde que se sacaron al mercado hasta la fecha de elaboración del informe.

Las tablas *owners*, *types* y *traits*, se constituyen como tablas auxiliares para la validación de los datos de las tablas principales. En este punto la tabla *punks_has_traits* se constituye como una tabla derivada de la unión de las tablas *punks* i *traits* mediante los identificadores de id para el punk y para cada característica, generando una tabla desglosada de las características que posee cada punk.

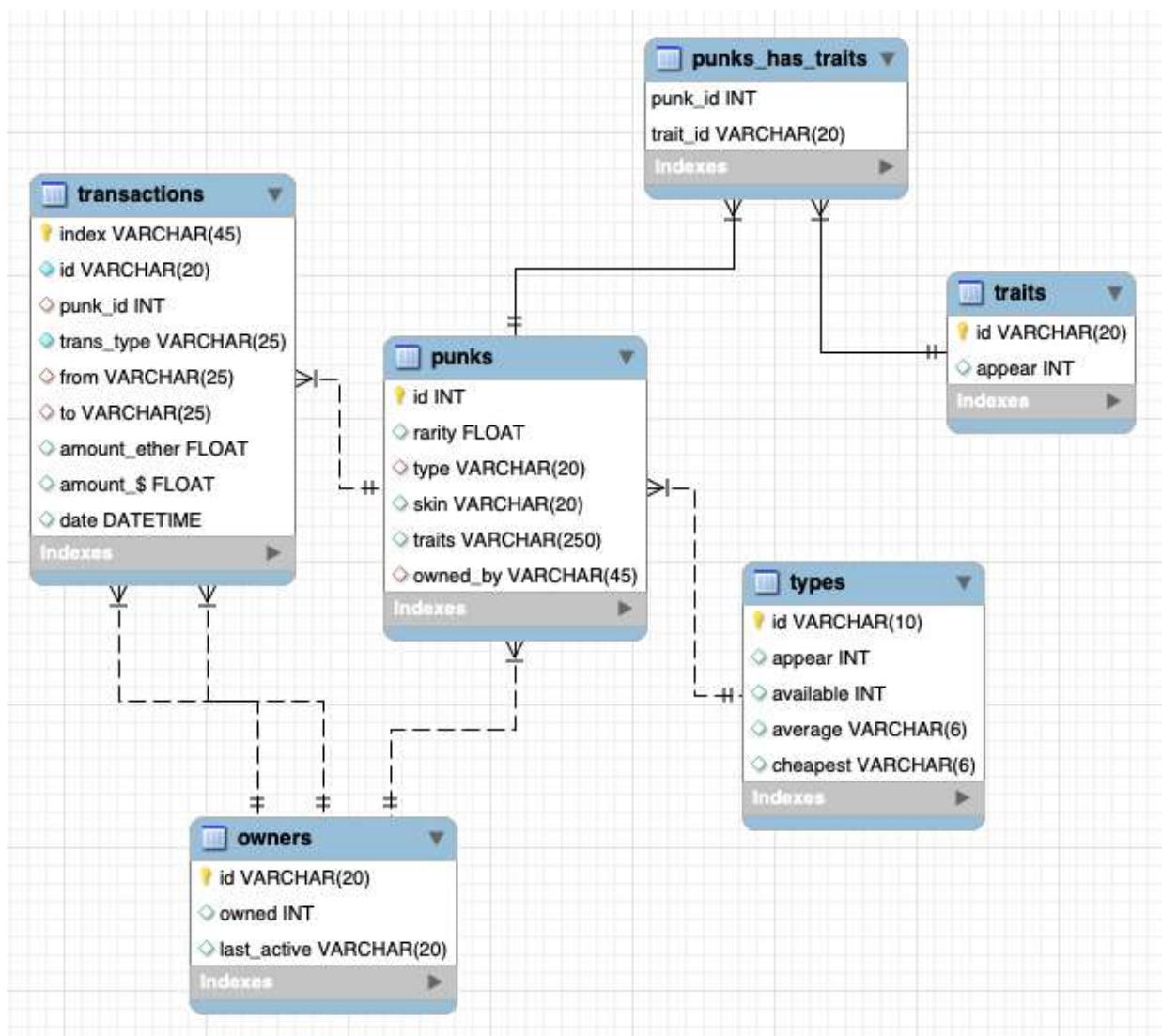


Fig.1.- Modelo Relacional de la base de datos diseñada para el alojamiento de los datos obtenidos mediante webscrapping.

PARTE 3

Análisis Exploratorio y Tratamiento de Datos Preliminar

3.1 - Consideraciones previas

Para trabajar con los datos se han generado dos peticiones (“queries”) a la base de datos:

- La primera *query* recoge el total de punks así como sus características y el propietario actual
- La segunda recoge todas las transacciones realizadas, anotando el identificador de la transacción, la fecha, el comprador, el vendedor y el precio de venta en ETH, así como las características del punk asociado a la transacción (igual que en la tabla anterior).

Las tablas obtenidas se han procesado posteriormente para obtener una columna donde se indica el número de atributos de cada punk y una “*dummy table*” de los atributos de los punks, es decir, una tabla desglosada donde se indica para cada atributo si un punk lo posee o no.

A la tabla de transacciones también se le ha añadido la conversión de ETH a dólares según la tasa de cambio en la fecha de la transacción (**Fig 2.**).

	index	punk_id	trans_type	from	to	amount_ether	amount_\$	date	rarity	type	...	'TassleHat'	'Tiara'	'TopHat'	'VR'
0	0	1	Bid withdrawn	0xE1E492		10.00	33222.10	2021-08-13	4352.0	Male	...	0	0	0	0
1	1	1	Bid	0xE1E492		10.00	30434.10	2021-08-12	4352.0	Male	...	0	0	0	0
2	10	1	Offer withdrawn			0.00	0.00	2020-09-14	4352.0	Male	...	0	0	0	0
3	100	3	Bid	0x717403		0.06	16.36	2017-06-26	7896.0	Male	...	0	0	0	0
4	1000	138	Offered			47.49	149903.00	2021-08-16	9806.0	Male	...	0	0	0	0
...
166396	99995	6400	Bid withdrawn	0x034B46		55.00	121639.00	2021-04-24	4925.0	Male	...	0	0	0	0
166397	99996	6400	Bid	0x034B46		55.00	129997.00	2021-04-23	4925.0	Male	...	0	0	0	0
166398	99997	6400	Offered			86.00	183984.00	2021-04-12	4925.0	Male	...	0	0	0	0
166399	99998	6400	Offered			75.00	155408.00	2021-04-09	4925.0	Male	...	0	0	0	0
166400	99999	6400	Offered			65.00	136053.00	2021-04-04	4925.0	Male	...	0	0	0	0

Fig 2.- Extracto parcial de la tabla final donde se recogen las características de las transacciones de *CryptoPunks*.

3.2 - Características de los Punks:

3.2.1 - Tipos y Pieles:

Como ya se ha indicado en la descripción inicial de los punk, estos se subdividen en 5 tipos. Los más frecuentes son los hombres (“Male”) y mujeres (“Female”) conformando el 98,8% de la población total (**Fig. 3**). Los otros tres tipos “Zombie”, “Ape”(simio) y “Alien” conjuntamente forman el 1,2% restante.

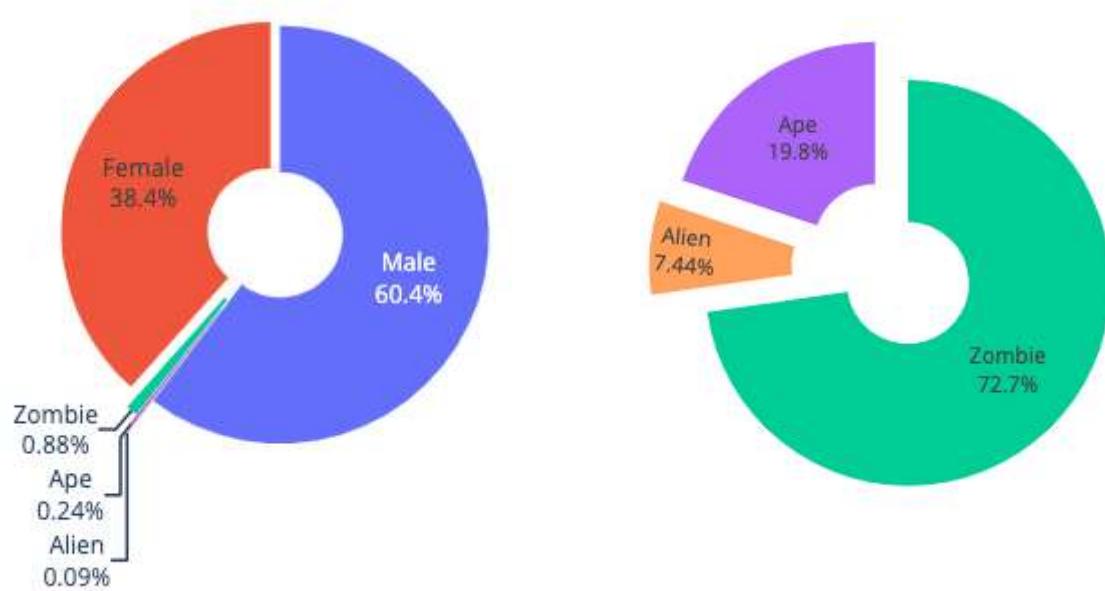


Fig 3.- Distribución del tipo de criptopunks según su tipo.

Dentro del grupo marginal estos tres tipos se distribuyen tal y como se observa en la Fig. 3. Dentro de este subgrupo los Zombies son mayoritarios, conformando el 72.7% del subgrupo, seguidos por los Simios y por los Aliens, siendo este último el tipo más exclusivo.

En cuanto al color de la piel, éste únicamente es distintivo para los punks Humanos. Vemos que en este caso la distribución de los tipos de piel es uniforme, excepto para la piel albina que ocupa un porcentaje mucho menor en relación a los otros tipos de piel (Fig 4). En esta figura también podemos observar como la proporción de tipos de piel se mantiene similar para Machos y Hembras.



Fig 4.- Distribución de color de piel para criptopunks de tipo *Male* y *Female*. Se aprecia como la proporción *Male* / *Female* es similar para cada tipo de piel.

Analizando la distribución de los punks por tipo, en relación a sus precios de venta (**Fig. 5**). Podemos notar que los tipos con menor cantidad de punks (*Alien*, *Ape* y *Zombie*) tienen valores medios más altos en sus ventas. y una desviación estándar menor (con menos outliers). que los punks del grupo (*Male*” y *Female*). A estos dos grupos nos referiremos a partir de ahora como AAZ (por las siglas de *Alien*, *Ape* y *Zombie*) y Humanos (*Male* y *Female*).

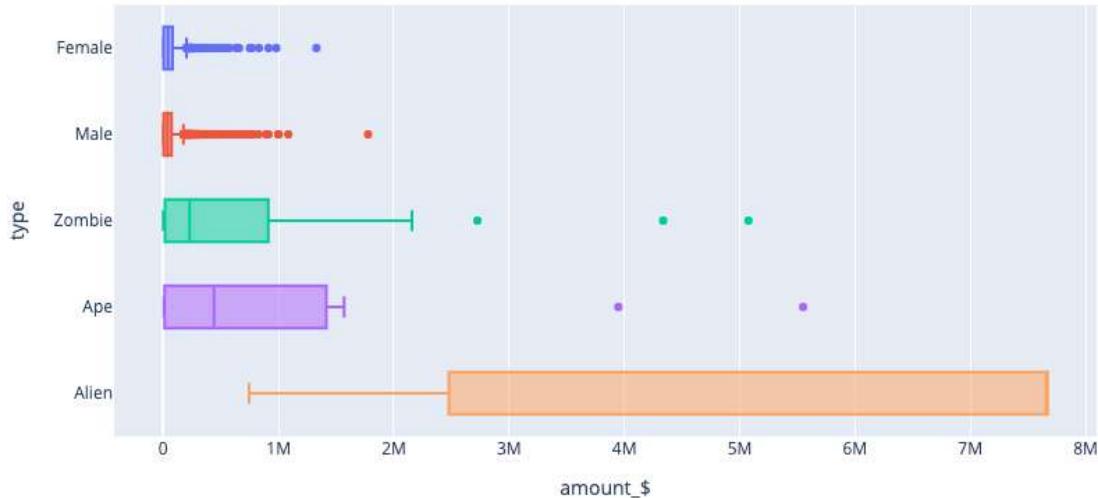


Fig 5.- Distribución del precio de venta se los precios según el tipo de punk. Periodo 01/01/2019 - 30/09/2021.

Aunque no tan notorias como las diferencia entre los punks AAZ. Los punks Humanos muestran diferencias internas en función de su tipo y piel. En la **Fig. 6** se aprecia que los precios mayores son preferentemente en los punks Hembra, frente a los Machos y en los de piel albina o clara frente a los de piel oscura.

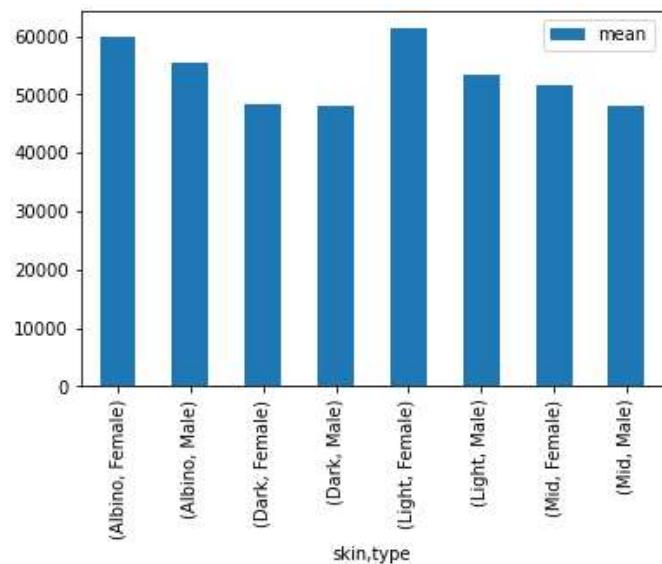


Fig. 6.- Valor medio de ventas por *type* y *skin* de los punks tipo Humano (*Male* , *Female*). Periodo 01/01/2019 - 30/09/2021.

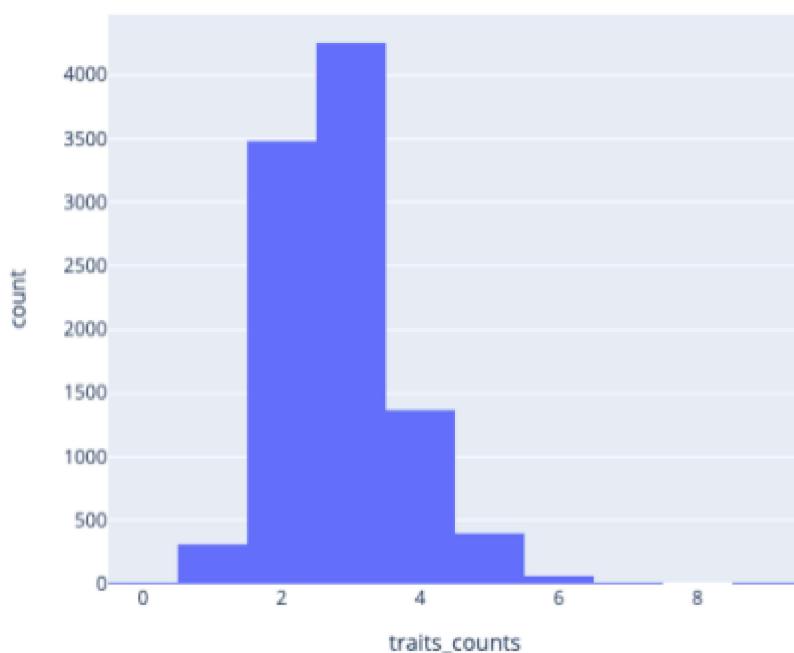
La gran diferencia de precios de venta entre los dos grupos es lo suficientemente significativa como para poder afirmar que el grupo AAZ forma por sí solo un grupo con características de mercado únicas e independientes del resto, debido también a que forman un porcentaje menor al 2% de la población de punks.

A partir de este punto los análisis se presentarán generalmente para punks de tipo humano. Los punks AAZ serán estudiados independientemente, mostrando los resultados en un apartado dedicado exclusivamente a ellos.

3.2.2. - Atributos:

3.2.2.1 - Número de atributos:

La distribución de punks según su número de características se asemeja a una distribución normal con un ligero “Skew” hacia la derecha tal y como se puede observar en la Fig. 5.



Mean	2.83
Std Dev	0.91
Min	0.00
1%	1.00
5%	2.00
25%	2.00
50% (Median)	3.00
75%	3.00
95%	4.00
99%	5.00
Max	9.00

Fig 5.- Histograma de distribución de los punks según su número de traits y descripción estadística.

Entrando más en detalle y analizando la descripción estadística se aprecia que la desviación estándar es relativamente pequeña y que el 98% aproximadamente de los punks están aglutinados entre 2 y 4 atributos, en porcentaje superior al 95%. El resto se distribuye entre 0-1 atributos y entre 5-9 atributos. Esto hace que los punks con 0,1, 6, 7, 8 o 9 atributos sean extremadamente raros.

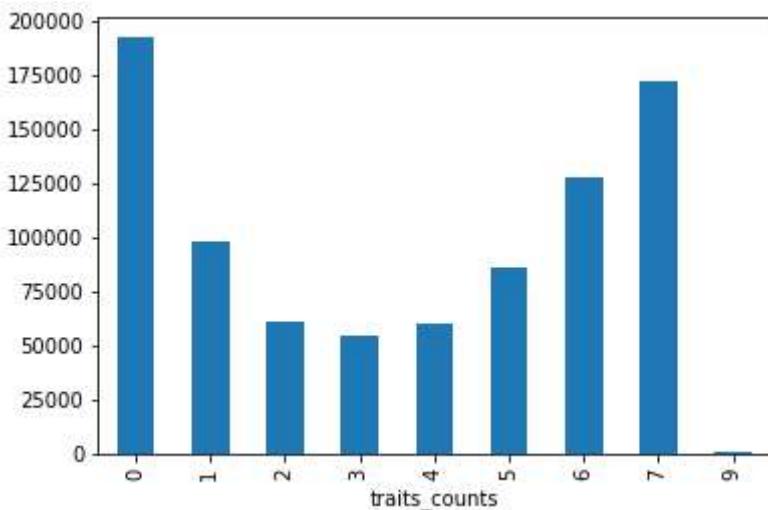


Fig 6.- Histograma de precios medios de venta en US\$ para Punk Humanos según su número de traits.

La última vez que se vendió el único punk con 9 características (Punk #5949) fué en Abril de 2019 por U\$ 307,49. Actualmente el propietario lo oferta por un valor superior a los 16.000 millones de U\$ y recibe pujas de potenciales compradores por valores cercanos a los 2 Millones de U\$.

3.2.2.2 - Atributos más exclusivos:

Como cualquiera de las demás características de los punks, los atributos no se distribuyen de manera uniforme. En el siguiente gráfico de barras (**Fig. 7**) se observa la gran diferencia entre el atributo que aparece con mayor frecuencia (“Earring”) con un 24,59% y los atributos más raros, con la característica “Hidden”, siendo el más raro (“HiddenBuckTeeth”) con un 0.02%.

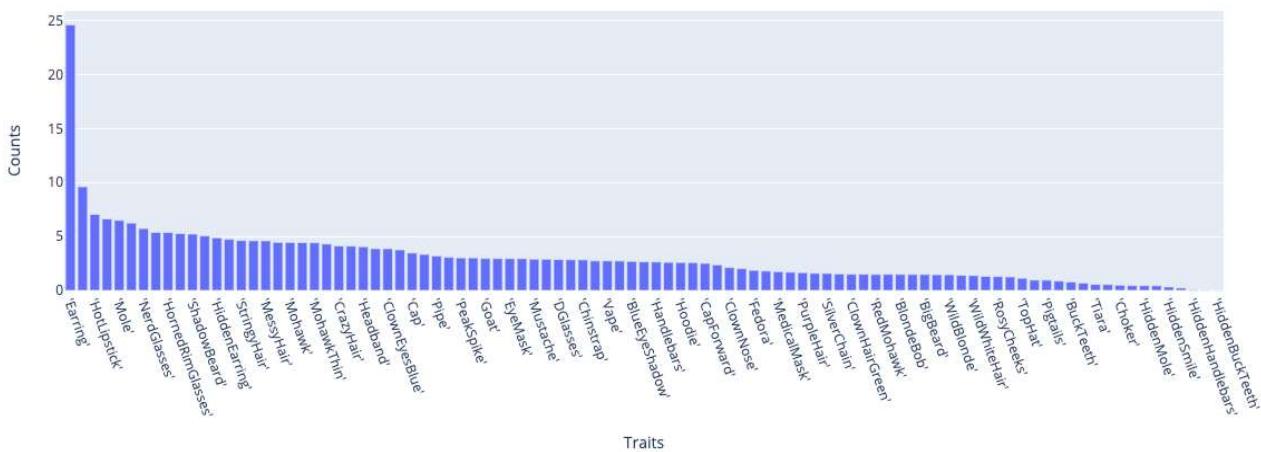


Fig 7.- Frecuencia de aparición de los atributos en %.

Con la intención de identificar si algunos atributos aparecen en los punks vendidos por valores más altos, se han separado los punks en 2 grupos: a) ventas sumadas, agrupadas por Punk Id, por encima de U\$ 1.000.000; b) ventas sumadas, agrupadas por Punk Id, por debajo de U\$ 10.000 (**Fig. 8**).

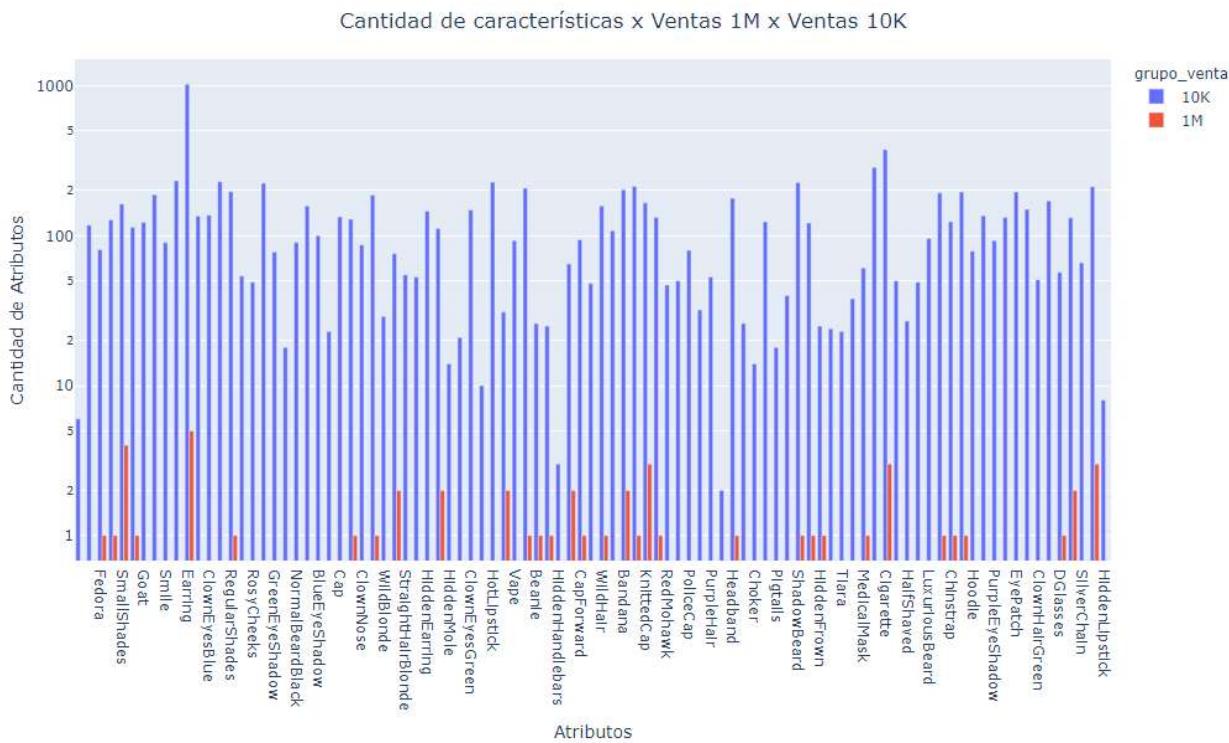


Fig 8.- Comparativa de características entre grupo de ventas > de 1M x < 10K U\$

En el gráfico en cuestión no se revela ninguna característica destacada en el grupo con ventas mayores en comparación con el otro grupo, es decir, las características que están en el grupo más caro, también aparecen en el grupo más barato.

3.3.3 - Índice de rareza (“Rarity”):

El “rarity” clasifica a todas las muestras ordenandolas de manera que todas tienen un valor pero este no se puede repetir por lo que juega un papel doble: a) como etiqueta para ordenar las muestras; b) como valor numérico que se relaciona con su exclusividad.

En este segundo aspecto podemos relacionar el valor de este índice con el precio de venta de los punks. De este modo vemos que si existe una relación aunque con un coeficiente relativamente bajo y negativo (-0.175).

Si confrontamos estas dos variables en un gráfico de dispersión (**Fig 9a**) estas dos variables se visualiza claramente cómo para índices de rareza más pequeños la distribución del valor de venta de los punks se ensancha hacia valores más altos formando una especie de embudo que cierra hacia la derecha. En esta figura también se detectan ciertos picos donde el valor de venta se dispara alrededor de un índice de rareza (“rarity” = 5000, 6000, 8000 aprox.). El estudio de estos picos no se incluye en este informe pero se considera que sería un factor a tener en cuenta para estudiar en un futuro.

El gráfico **9b** muestra el detalle de la zona con menor índice de rareza. Aquí se puede observar con más claridad el aumento progresivo en la dispersión de los valores. Otra cuestión que se observa es la existencia de un hueco entre el rango de “rareza” entre 475 y 600 y una zona menos poblada en el rango de 0 a 50 aprox. Estos huevos se forman ya que en este gráfico no se representan los punks de tipo AAZ, que se muestran en el siguiente apartado (**3.3.4, Fig. 10**).

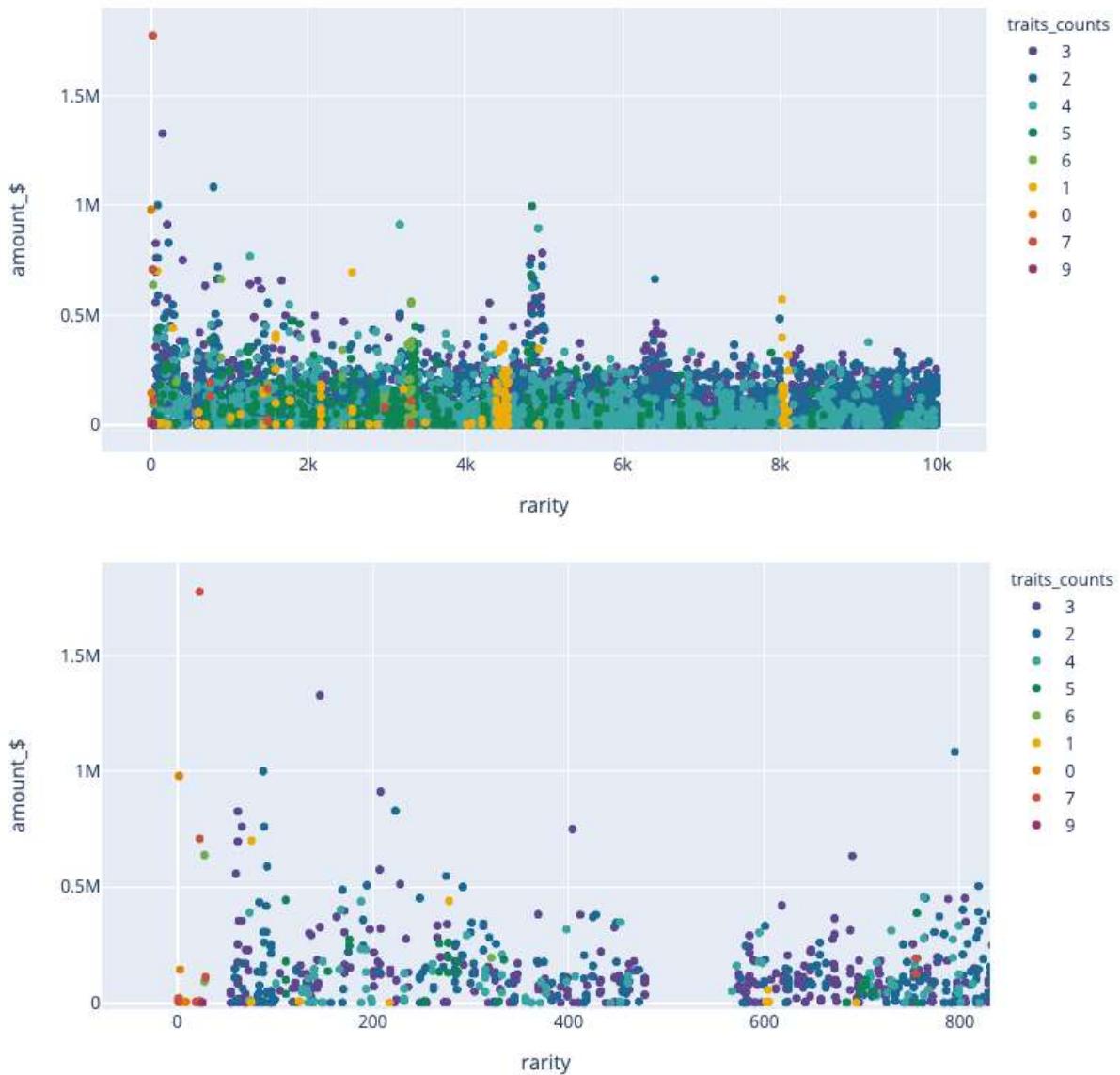


Fig. 9.- a) Precio de venta de los cryptopunks tipo Humano en función de su índice de rareza; **b)** Detalle del gráfico anterior que muestra ampliada la zona con los punks con un índice de rareza menor (más exclusivos). Periodo 01/01/2019 - 30/09/2021.

3.3.4 - Análisis de características internas del grupo AAZ

La correlación entre los diferentes parámetros que clasifican a los punks del grupo AAZ en relación al precio de venta quedan resumidas en la figura siguiente (**Fig.10**)

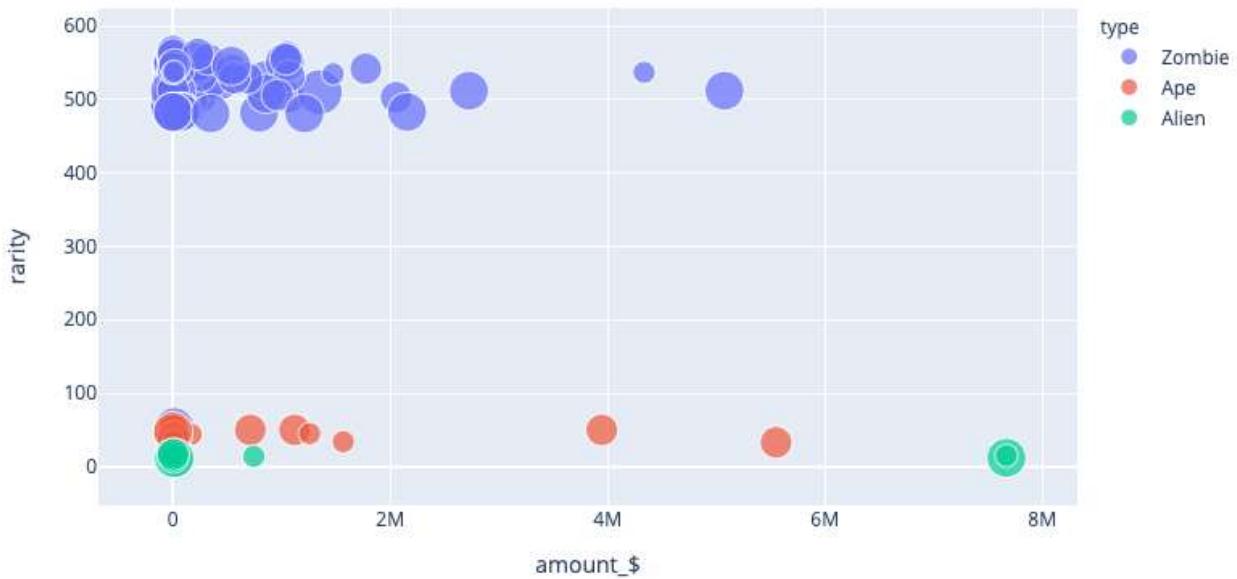


Fig 10.- Precio de venta de los cryptopunks tipo AAZ en función de su índice de rareza.

Si atendemos al eje vertical (*rarity*) se observa como el rango de rareza de este grupo es significativamente bajo (0-600) - cuanto menor es el índice, más raro es el punk - en relación al rango general (0-10000) por lo que, en general todos los punks tipo AAZ se consideran raros, aun así los tipo *Ape* y *Zombie* tienen un índice de rareza mucho menor.

En cuanto al número de atributos (tamaño de los círculos), no se observa una tendencia definida en relación al precio de venta.

En resumen podríamos definir que el único factor relevante para el precio de este grupo de punks es su tipo ya que este por sí mismo, al ser tan minoritario respecto a los Humanos (apartado 3.2.1) se auto-atribuye un grado de exclusividad elevado (bajo índice de rareza).

3.3 - Características de los Propietarios

El total de criptopunks se encuentra repartido entre 2933 propietarios, de estos existen 100 (3.41%) que acumulan la mitad de los punks (**Fig.11**). Esta tendencia de ciertos propietarios a acumular punks indica que los punks han adquirido un valor de objetos coleccionables.

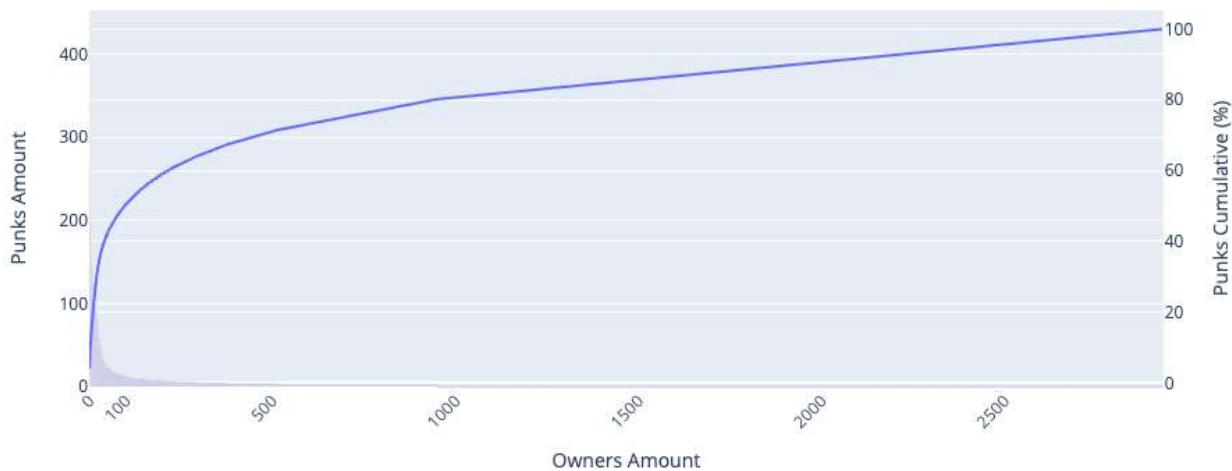


Fig 11.- Distribución de los punks entre sus propietarios.

No obstante, al mirar la distribución por tipo y piel de los punks de los 100 propietarios con mayor número de punks observamos que estos siguen un patrón muy similar a la distribución original de los punks (**Fig. 12**), indicando que no existe una determinada preferencia por determinados punks entre estos propietarios. La tendencia acumulativa de estos punks también puede explicarse con la finalidad de acumularlos por su valor financiero. Este punto se desarrollará posteriormente.

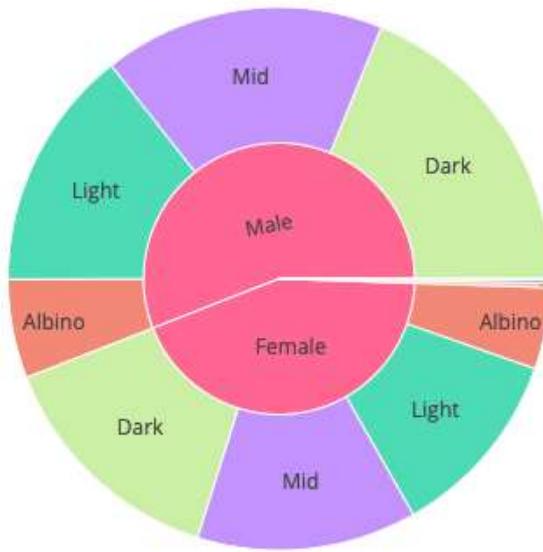


Fig. 12.- Distribución de los punks por tipo y raza coleccionados por los 100 mayores propietarios. Se puede observar como la distribución es muy semejante a la distribución original de la **Fig. 4**.

3.4 Evolución de ventas de los Cryptopunks.

3.4.1 - Número de ventas:

En la **Fig.13** se puede observar como desde su lanzamiento el número de ventas de Cryptopunks ha sido más o menos estable hasta mediados de 2019 donde se empieza a notar un ligero aumento que será mucho más notable a partir del 2021. Este aumento abrupto durante el último año coincide con la rápida revalorización del ETH durante este mismo periodo producida por la entrada en el mercado de grandes personalidades e instituciones (BBVA, VISA, etc.)¹² que confían en el valor de esta moneda. El valor del ethereum se prevé que siga en aumento, igualando cada vez más al bitcoin¹³.

Más en detalle, al comparar las ventas acumuladas por año (**Fig. 14**) se puede observar fácilmente como las ventas de CriptoPunks en 2021 superan a las del resto de años juntas.

¹² <https://www.bsc.news/post/crypto-punks-continue-mainstream-interest-stemming-from-visa-and-odell-beckham-jr-purchases>
<https://www.bsc.news/post/nft-madness-accelerates-on-ethereum-after-user-purchases-88-punks-for-5-000-000>

¹³ <https://www.ipgroup.com/economia-digital/20339-ethereum-alcanzo-un-nuevo-maximo-historico-que-paso>

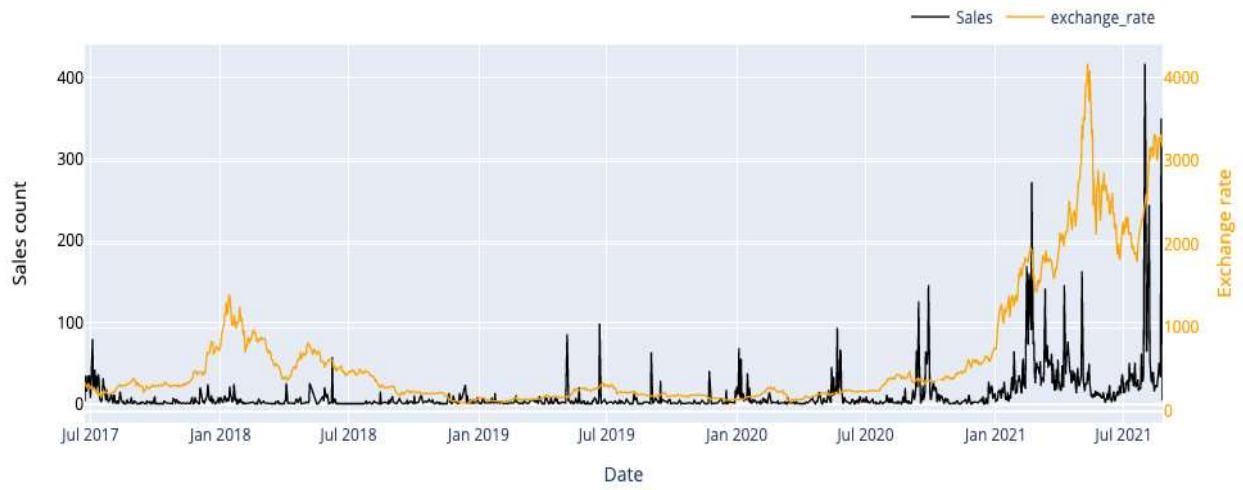


Fig 13.- Evolución del número de ventas de *CryptoPunks* desde su lanzamiento hasta la actualidad.

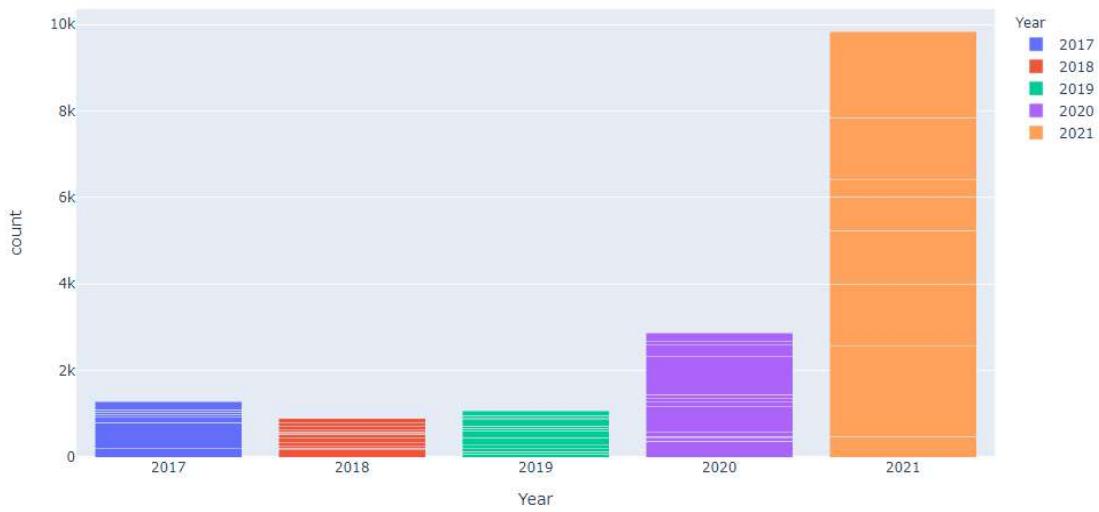


Fig 14.- Ventas de *CryptoPunks* acumuladas por año desde su lanzamiento hasta la actualidad.

3.4.2 - Precio de las ventas:

El histórico de precios de venta (**Fig. 15**) muestra una tendencia más o menos estable hasta la primavera de 2020. Dentro de este período aparecen dos ciclos estacionales de subidas que alcanzan su máximo en enero 2018 y julio de 2019. A partir de mayo de 2020 el precio de venta de los criptopunks empieza a aumentar de forma constante, provocada por el aumento de la tasa de cambio US\$/ETH que se observa paralela al subida de precio del punk y del incremento del número de transacciones.

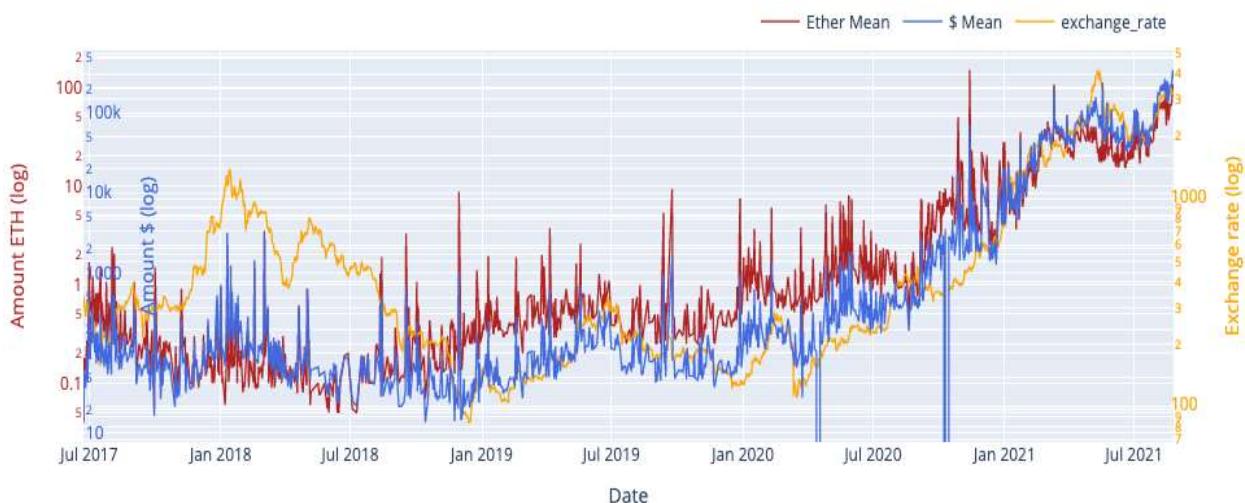


Fig. 15.- Histórico de precios de venta en US\$ y ETH y evolución de la tasa de cambio US\$/ETH. Los ejes verticales están marcados en escala logarítmica. Las líneas roja y azul representan el valor medio por venta y las sombras asociadas los valores máximos y mínimos.

El aumento de la actividad en el mercado de los *CryptoPunks* ligado al aumento del valor del ETH es un claro indicador del valor especulativo de los cryptopunks.

En el gráfico de evolución de precios según tipo durante el 2021 (**Fig. 16**) se observa como los punk tipo AAZ tienen una acusada tendencia a revalorizarse. Este crecimiento es especialmente notorio en los tipo Alien.



Fig. 16.- Evolución de precios según tipo durante el año 2021

Atendiendo a los punks de tipo Humano, estos no parecen acusar diferencias muy grandes respecto a su género, no obstante, sí que existe cierta tendencia a una mayor revalorización en los punks con un rarity menor (**Fig.17**).

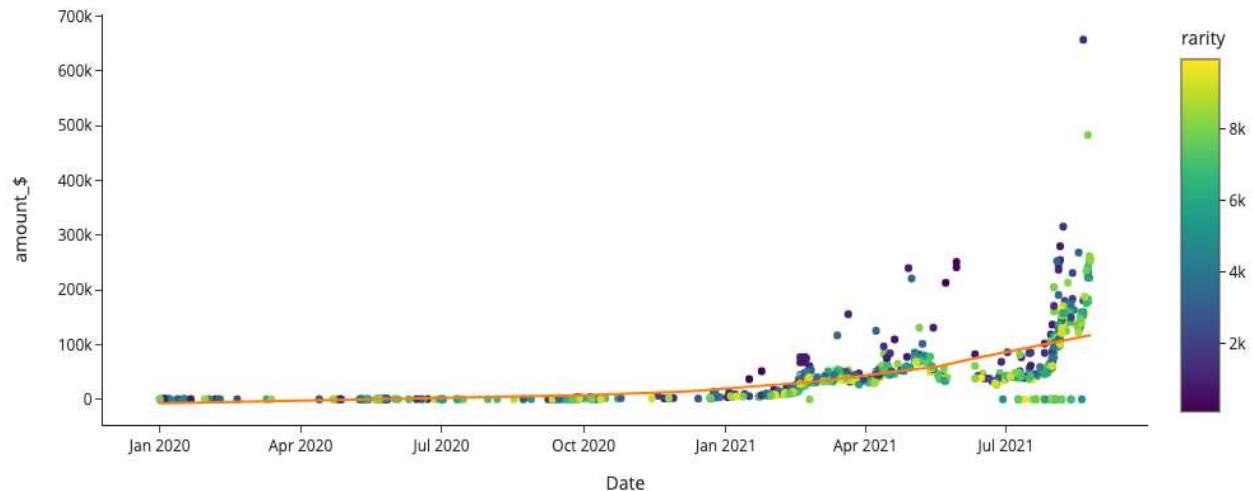


Fig 17.- Evolución del precio de venta de los 100 punks Humanos más vendidos en el periodo Enero 2020 - presente.

PARTE 4

Machine Learning

4.1 Introducción

Para la realización del análisis predictivo se han considerado únicamente las ventas de los punks del tipo Humano ya que se considera que con los resultados obtenidos en la **parte 3** queda suficientemente demostrada la diferenciación de los Humanos con los punks de tipo AAZ.

El análisis predictivo se realiza con la finalidad de determinar, si fuese el caso, en qué proporción afectan los diferentes parámetros de los punks en relación a su valor económico. Para ello se propone un análisis de clustering no supervisado mediante el algoritmo *K-Means*¹⁴ para visualizar posibles similitudes entre punks y posteriormente. Los algoritmos utilizados se obtienen de la librería *Sci-Kit Learn*, ampliamente utilizada en Machine Learning.

4.2 Data Frame

El análisis predictivo se realiza a partir de un *dataframe* formado originalmente por 106 columnas y 15892 filas, que se corresponden con todas las ventas realizadas para los cryptopunks tipo Humano.

Cada columna se corresponde con:

- Punk_id: Identificador del Punk
- amount_\$: precio de venta en dólares
- traits_counts : número de atributos
- Dummies: 96 columnas para los atributos y 4 para la piel
- year: año de venta (2017-2021)
- month: mes de venta (1-12)
- day: dia de la venta (1-31)
- Type.1_Male: identificador del tipo de punk; 1:*Male*, 0:*Female*.

¹⁴ <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html?highlight=k%20means#sklearn.cluster.KMeans>

Durante las diferentes iteraciones que se realizan durante todo el proceso para “afinar” el modelo, el número de filas y de columnas irá variando afectado por el tratamiento de outliers, eliminación de los dummies, etc.

4.3. Ingeniería de variables.

4.3.1 - Tratamiento de variables categóricas:

Dependiendo del tipo de variable categórica se han realizado diferentes procesos de transformación.

Para el punk ID se ha hecho una conversión de tipo de datos de categórico a numérico, ya que el identificador de los punks está formado únicamente por dígitos.

El los atributos y el color de piel se han transformado a variables tipo dummy

El número de atributos se ha reordenado en función de su exclusividad después de comprobar que esta transformación mejoraba los parámetros del *clustering* (**Tabla 1**).. La transformación ha sido la siguiente:

Número de traits	0	1	2	3	4	5	6	7	8	9
Orden de Exclusividad	2	5	8	9	7	6	4	3	-	1

Tabla 1.- conversión de número de atributos a orden de exclusividad. El orden de exclusividad se ha realizado teniendo en cuenta el % de punks que tiene este número de atributos (**Fig. 5**).

4.3.2 - Tratamiento de outliers

Se han considerado como outliers todos aquellos precios de venta superiores a 101000 US\$ que conforman el percentil 85% y los inferiores a 36 US\$ que conforman el percentil de 5%.

4.3.3 - Estandarización

Se ha realizado la estandarización de los datos utilizando el algoritmo “*StandardScaler*”¹⁵ de *Sci-kit Learn*. Este algoritmo se ha aplicado con la intención de homogeneizar todas las

15

<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html?highlight=standardscaler#sklearn.preprocessing.StandardScaler>

variables dada su diversidad en cuanto distribución de valores, adaptándolas a distribuciones normales con valor de media 0 y desviación estándar de 1.

4.3.4 - PCA

Para intentar mejorar el algoritmo de clusterización se ha realizado un análisis de componentes principales o PCA. En este proceso, primeramente se ha realizado un PCA con una cantidad de componentes igual a la cantidad de columnas para visualizar el % de varianza que explica cada uno. En última instancia se ha escogido un número de componentes principales lo más reducido posible pero con una varianza acumulada significativa.

4.4 - Modelos de clusterización

4.4.1 - Modelo inicial.

El punto de partida en el proceso de predicción ha sido aplicando el algoritmo al *dataframe* con las variables categóricas ya transformadas y estandarizadas.

En este punto se observa una correlación muy baja entre las diferentes variables, tal y como podemos observar en la (**Fig. 18**). Los valores de correlación mayores se dan entre el año de la venta y el precio.

Para determinar el número óptimo de clusters se ha realizado un análisis de los parámetros *inertia* y *silhouette* en un rango de K entre 2 y 11 clusters. El número de clusters óptimo se ha decidido utilizando el método del codo o “elbow method”¹⁶, para acabar de afinar se busca el valor de *silhouette* mayor para los puntos cercanos a la zona del “codo”.

¹⁶ [https://en.wikipedia.org/wiki/Elbow_method_\(clustering\)](https://en.wikipedia.org/wiki/Elbow_method_(clustering))

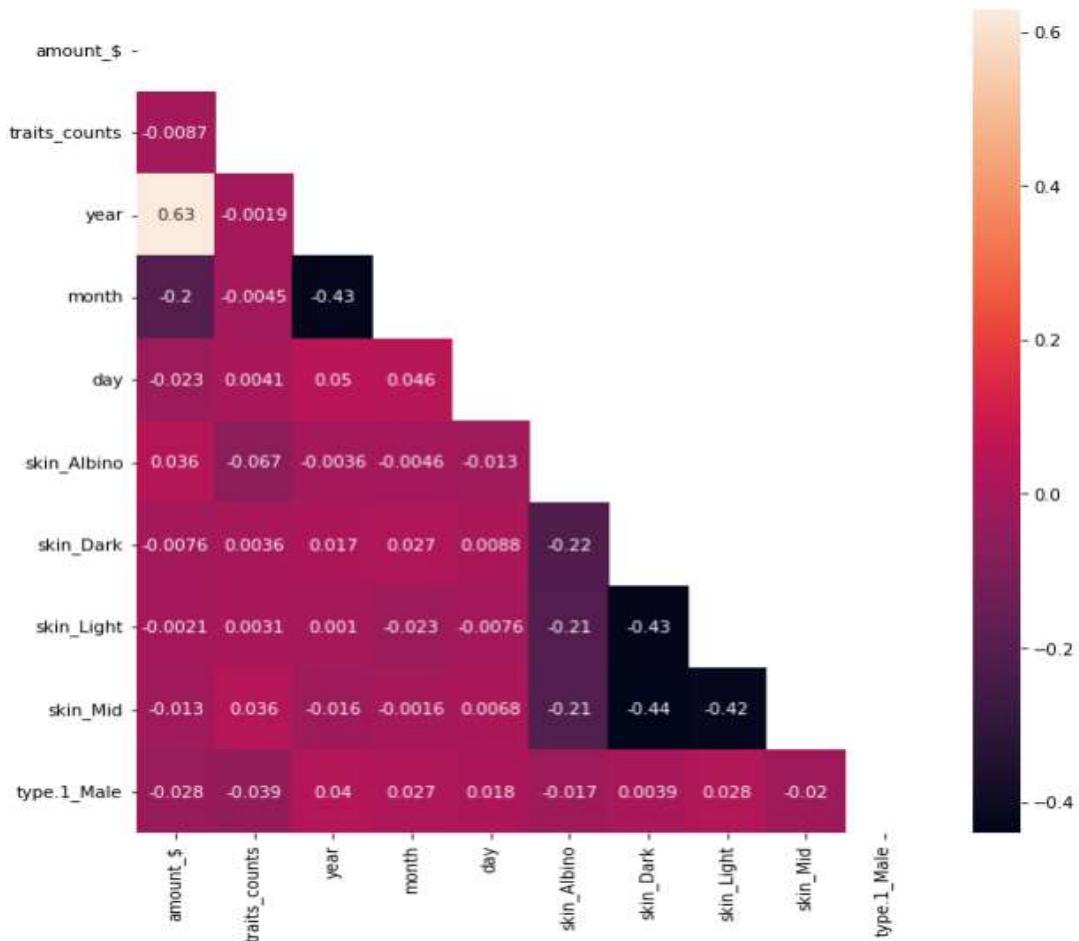


Fig 18.- Extracto de la matriz de correlación entre las diferentes variables del data frame de entrada. Los valores dummy por lo general tienen valores de correlación inferiores a ± 0.05 .

Round	Sin Outliers	Estandarización	Dummies	PCA	num_K	Inertia	Sil.
1	No	Si	Si	No	4	115916.73	0.0948
2	Si	Si	Si	No	4	91607.11	0.0910
3	Si	Si	No	No	4	46039.71	0.1717
4	Si	Si	Si	Si	4	58006.70	0.1454
5	Si	Si	No	Si	4	21960.93	0.2778

Tabla 2.- Procesos aplicados y parámetros de evaluación obtenidos en las diferentes iteraciones del modelo de clusterización.

4.4.1 - Iteraciones de mejora del modelo.

Durante las diferentes iteraciones el valor de la *inertia* ha ido decreciendo a la vez que aumenta el *silhouette*, Indicando la mejora progresiva de los resultados en función del aumento del procesado de los datos de entrada.

Para todas las iteraciones el número de clusters óptimo ha sido $K=4$

Cómo se puede observar en la **tabla 2** el mayor incremento se da entre las iteraciones 2 y 3 donde prácticamente los valores de las métricas se doblan y reducen a la mitad respectivamente. Este salto se produce al eliminar de la tabla de entrada las columnas relativas a los *dummies* de los atributos, indicando que los atributos de los punkts no aportan valor en cuanto a la determinación de grupos afines.

En segunda instancia el análisis de componentes principales y su posterior aplicación también han contribuido considerablemente a la mejora de estos parámetros.

Finalmente el último modelo (5) con mejor puntuación responde a la siguiente configuración:

- Sin outliers
- Estandarizado
- Sin dummies
- Con PCA = 3 (67,19% de varianza)
- $K = 4$ clusters

4.4.1 - Resultados.

En el modelo 5, encontramos un valor de *Inertia* de 21960.93 y el Silhouette de 0.27 con una clusterización en 4 grupos.

label	count	amount_ \$	traits_counts	year	month	day	S_Albino	S_Dark	S_Light	S_Mid	Male
0	3817	43248.17	2.36	2020.94	3.71	16.01	0.12	0.31	0.28	0.29	0.67
1	2637	929.87	2.92	2019.22	7.46	23.78	0.08	0.31	0.29	0.32	0.68
2	3874	43151.33	3.30	2020.92	3.67	15.55	0.08	0.31	0.31	0.30	0.66
3	2327	411.52	2.77	2018.55	7.19	8.63	0.10	0.31	0.29	0.30	0.66

Tabla 3.- Procesos aplicados y parámetros de evaluación obtenidos en las diferentes iteraciones del modelo de clusterización.

El modelo generó 4 grupos en los cuales destacamos algunas características:

- **Label 0:** Precio medio alto, cantidad media de traits más baja .
- **Label 1:** Precio medio medio-bajo, cantidad de traits media.
- **Label 2:** Precio medio alto, cantidad de traits alta.
- **Label 3:** Precio medio bajo, cantidad de traits media.

Si bien estos valores de *inertia* y *silhouette* no podrían considerarse óptimos para afirmar la existencia de grupos de punks bien diferenciados si que se observan ciertas características con cierto interés. Lo más destacable es que los grupos 0 y 2, y 1 y 4 tienen características más similares entre ellos que con el otro grupo.

Tanto el grupo 0 como 2 tienen un número de punks similar, superior a 3800 unidades. También el precio medio de venta (*amount_*\$), alrededor de 43.200 US\$ y ambos están formados por ventas realizadas durante el 2020.

Los rasgos distintivos para estos dos grupos “más caros y más recientes” es que el grupo 0 tiene un número medio de punks albino mayor que el grupo 2 y un número de atributos promedio cercano a 2, mientras que el número de atributos promedio del grupo es superior a 3. Hay que recordar que tal y como se comenta en el apartado anterior. El tipo de piel albino es el más exclusivo dentro de los punks Humanos y también el índice de rareza, se aleja del valor modal (3) (**apartados 3.2.1 y 3.2.2.1**). Esto explicaría también que el precio medio de venta del grupo 0 es ligeramente superior.

En el grupo 1 y 3 caracterizado la principal diferencia entre los precios de venta medios viene marcada por el año de las transacciones, a venta más reciente, mayor el precio. El hecho de que el número de punks de estos grupos aumente con el año también es un reflejo de un aumento progresivo del mercado de *Cryptopunks*.

En cuanto al tipo de distribución por tipo y piel todos están muy cercanos a la distribución original.

En base a estas consideraciones podría concluirse que el factor principal que afecta al precio de los punks humanos es la tendencia alcista del valor del ETH potenciado por los factores que se detallan en la introducción de este trabajo. Más allá de esta tendencia general que afecta al mercado en general el algoritmo de clasificación ninguna correlación entre las características de los punks sirvan para predecir el valor determinado de un punk, aunque se deduce una cierta valoración de los punks con tipos de piel más exclusivos.

PARTE 5

Conclusiones

Se concluye que existen dos grupos muy bien diferenciados de punks respecto a su evolución en cuanto a su valor económico.

El primer factor diferencial de estos dos grupos es el tipo. Por un lado el grupo que en este trabajo denominamos AAZ por las siglas de los tipos que lo componen (**Alien, Ape y Zombie**). Este grupo tiene una representación inferior al 1.25% respecto al total de punks. El segundo grupo, llamado Humanos, tipo “*Male*” y “*Female*”, es el mayoritario (>98.75%).

En cuanto al valor económico, las ventas de los punks AAZ se mueven en un rango aproximado de U\$ 1M a 7M, mientras que el precio medio de las ventas para los punks humanos es de miles de dólares.

Dentro del grupo AAZ no se han encontrado otros factores relevantes que definan tendencias en la variación de los precios de venta más allá de la exclusividad que les proporcionan su propio “tipo”.

El análisis de las características de los punks Humanos ha revelado cierta influencia del número de atributos que afecta a su precio en variaciones de decenas a cientos de miles, seguido del color de piel (más claro, más valor) y el tipo (mujeres más valor que hombres) que implican cambios de decenas de miles, siendo los valores más altos los que coinciden con configuraciones menos frecuentes.

Mediante la aplicación del algoritmo de clusterización no se han podido confirmar con rotundidad las tendencias expuestas en el párrafo anterior ya no se han encontrado correlaciones lo suficientemente elevadas entre las características de los punks Humanos y su precio de venta

No obstante se ha detectado que al eliminar del algoritmo de predicción las columnas relativas a la frecuencia de los atributos, el resultado de los parámetros de evaluación ha mejorado considerablemente indicando que los atributos tienen una influencia muy escasa o nula en el valor del punk y que incluso pueden introducir ruido en el modelo.

Otro dato a resaltar del proceso de clusterización ha sido que los grupos obtenidos están agrupados principalmente en relación al año de venta, destacando que el valor de los criptopunks está ligado principalmente a la tendencia creciente del valor del ETH y al aumento de la actividad del mercado.

En cuanto a la evolución del mercado de compraventa de cryptopunks se observa un aumento progresivo de la actividad generado por el aumento continuo del valor del ETH desde mayo de 2020. Hay que tener en cuenta que un 30% aproximadamente en los cryptopunks nunca ha sido vendido desde su creación.

Dentro de esta subida general del valor del ETH el mercado de venta de punks tipo AAZ muestra una evolución de precios más acusada, aunque es menos activo debido a la escasez de este tipo de punks. Por el contrario, el crecimiento del valor de los punks humanos es más suave pero con una mayor actividad en la compraventa.

Actualmente el 50% de los punks se encuentran repartidos entre 100 propietarios un 3,5% del total. Este dato sugiere la existencia de cierto ánimo de colección entre los inversores o de la estrategia “*buy and hold*”, que predica la compra y manutención sin plazo, hasta que se valore al punto que desea el comprador.

Expuesto esto, se considera que el análisis exploratorio no está agotado, se podría profundizar en otros aspectos con el objeto de encontrar más informaciones relevantes que pudiesen ayudar en la toma de decisión. La corta edad de la red Ethereum y de los cryptopunks en general hacen que todavía sea un mercado muy inmaduro y por lo tanto difícil de predecir. A medida que avance el tiempo y la cantidad de datos generada crezca se presume que un estudio de características similares a éste podrá proporcionar conclusiones más específicas.

A modo de resumen general, los resultados obtenidos nos permiten destacar los siguientes puntos en cuanto a las opciones de inversión:

1. De manera general los punks parecen seguir la lógica de precios de mercado: cuanto más raro (escaso) más valorado.
2. El grupo AAZ presenta posibilidades de ganancia más alta debido a exclusividad, pero al mismo tiempo conlleva riesgos relacionados con alta volatilidad y bajo volumen de negociaciones.
3. El grupo de Humanos presenta negociaciones más modestas en su valor medio, pero mayores en cantidad. Estos factores disminuyen la probabilidad de una gran subida de precio y por consecuencia de ganancias extraordinarias, pero también del riesgo de no lograr vender el punk en el momento en que se desea.
4. No hemos podido identificar ningún factor con fuerte correlación con el precio, aunque la cantidad de características (*traits*) parece indicar que punks con 0, 1, 6 y 7 características son los que tienen precios medios más altos.
5. Consideramos que la participación de un curador profesional del mercado de arte y/o criptoarte puede agregar información importante para los análisis, ya que el mercado de arte tiene movimientos con características propias.
6. Finalmente, ante las tendencias alcistas del mercado del ETH y la creciente popularidad de los *Cryptopunks*, se indicaría ante una posible inversión lo ideal sería formar una cartera diversificada donde se figure algún punk del grupo AAZ y también punks del grupo Humanos con orientación hacia Mujeres con piel clara. Si bien la tendencia del

mercado es alcista tiene características de burbuja por lo que habría un cierto riesgo a una pérdida de valor abrupta.