

## Estudio de mercado y proyecciones de NFT's: Crypto Punks.

Informe Proyecto Final Data Science - Antón Sobrino, Aram Puig

Supervisión: Kevin Mamaqi



# Índice

<b>1. Introducción</b>	<b>3</b>
1.1 Información preliminar	3
¿Qué son los Cryptopunks?	3
¿Qué es un NFT?	3
1.2 Definición del problema y estudio de mercado	3
Hipótesis del trabajo y objetivo final	4
1.3 Definición de métodos y herramientas empleadas	4
<b>2. Análisis descriptivo y exploración</b>	<b>5</b>
2.1 Características de los punks	5
2.1.1 Tipos de punks	5
2.1.2 Atributos de los punks	7
2.1.3 Número de atributos	10
2.2 Historial de transacciones	11
2.2.1 Categóricas: Tipo de transacciones	12
Conclusiones tipo de transacciones	12
2.2.2 Categóricas: Usuarios (emisario, receptor y usuario asociado a la transacción)	13
Conclusiones users	13
2.2.3 Numéricas: cantidad de transacciones por id de punk	14
Conclusiones número de transacciones por punk	14
2.2.4 Numéricas: estudio de ventas	15
2.2.5 Numéricas: Precio ETH-Dólares (Adj. Close)	17
<b>3. Análisis factorial</b>	<b>19</b>
<b>4. Clustering no-supervisado</b>	<b>20</b>
4.1 Examen del Tercer Cluster (2)	21
4.1.1 Punks seleccionados	24
<b>5. Regresión lineal y proyecciones</b>	<b>25</b>
5.1 Target Precio de Venta en Eth.	25
5.1.1 Predicciones	25
5.2 Target variable Rendimiento	26
5.2.1 Predicciones	26
<b>6. Confirmación/Refutación de la Hipótesis inicial</b>	<b>27</b>
6.1 Demostración estadística ANOVA y F-Test	27
<b>7. ¿El punk más atractivo? En qué punks invertir y conclusiones del informe.</b>	<b>31</b>
<b>Anexo</b>	<b>33</b>
A1. Obtención de datos	33
Fuentes (URLS) y métodos de scraping	33
A2. Diseño base de datos relacional	33
Descripción Tablas	33
Atributos	34
Esquema Relacional	36

# 1. Introducción

## 1.1 Información preliminar

### ¿Qué son los Cryptopunks?

Según la página de Larvalabs (sitio oficial de los cryptopunks), los cryptopunks son 10.000 personajes únicos coleccionables con prueba de propiedad almacenada en la blockchain de Ethereum. Es el proyecto que inspiró el movimiento moderno del criptoarte <sup>1</sup>, basado en los NFT (Non-Fungible Tokens).

No hay dos cryptopunks iguales. Estos 10.000 personajes (sus rostros) fueron creados mediante un algoritmo, combinando una serie de tipos principales - male, female, ape, zombie and alien - y accesorios - 87 accesorios distintos como barba, pendientes, gafas de sol, cigarro etc.

### ¿Qué es un NFT?<sup>2</sup>

NFT significa Non-Fungible Token<sup>3</sup> en inglés. En economía, un activo fungible es algo con unidades que pueden intercambiarse fácilmente, como el dinero. En el caso del dinero, se puede cambiar un billete de 10 libras esterlinas por dos billetes de 5 libras esterlinas y tendrá el mismo valor.

Sin embargo, si algo no es fungible, esto es imposible: significa que tiene propiedades únicas, por lo que no puede intercambiarse con otra cosa. Un ejemplo de un activo no fungible puede ser una casa o un cuadro como la Mona Lisa, que es único. Se puede hacer una foto del cuadro o comprar una impresión, pero sólo habrá un cuadro original.

Por lo tanto, los NFT son activos "únicos" en el mundo digital que pueden comprarse y venderse como cualquier otra propiedad, pero no tienen forma tangible propia. Los tokens digitales pueden considerarse como certificados de propiedad de activos virtuales o físicos.

## 1.2 Definición del problema y estudio de mercado

Estos tokens se almacenan criptográficamente en una red blockchain para su autenticación y validación. Incluso cuando se realizan transacciones, cada ficha se vende como un todo, a diferencia de las criptomonedas, que son divisibles en unidades más pequeñas. Los NFTs pierden su valor cuando se dividen en porciones más pequeñas, sin embargo, pueden ser duplicados por el artista y cada copia adicional puede ser vendida por separado (pero el número total de duplicados será entonces siempre conocido).

---

<sup>1</sup> <https://www.larvalabs.com/cryptopunks>

<sup>2</sup> <https://www.bbc.com/news/technology-56371912>

<sup>3</sup> Token puede significar ficha, identificador o símbolo en inglés, entre otros. No tiene una traducción directa al español.

El problema planteado para el análisis y su desarrollo es el estudio de los precios de este grupo de NFT's, los *CryptoPunks*, ligados al blockchain de Ethereum. Los *CryptoPunks* aparecieron en 2017 en forma de un posible obsequio para aquellos que tuvieran activos de Ethereum, sus usuarios podían reclamarlos de forma gratuita. Desde entonces sus propietarios han podido comerciar con ellos y sacar grandes beneficios, pues es un mercado que no ha dejado de crecer. El precio récord de compra por un solo *cryptopunk* se alcanzó el pasado marzo, cuando se vendió por 7,58 millones de dólares (4,2K eth.) el punk número #3100 y por un precio parejo el punk #7804. Este fenómeno ha causado sensación y atrae la atención de inversores a un mercado potencialmente rentable, buscando respuestas a qué hace más valioso a un punk u otro, y por cuál debería uno apostar. Aquí es donde empieza nuestro trabajo como data scientists, intentando responder a estas preguntas.



CryptoPunk #3100 vendido por 7,58M.

## Hipótesis del trabajo y objetivo final

- ( $H_0$ ) : El precio medio de los punks en ethereum es el mismo en cada uno de los cinco tipos principales ('Male', 'Female', 'Ape', 'Zombie' y 'Alien')
- ( $H_A$ ) : El precio medio de los punks en ethereum difiere en alguno (o en todos) de los cinco tipos principales ('Male', 'Female', 'Ape', 'Zombie' y 'Alien')
- Valor de significación  $\alpha = 0.05$ .

La notación estadística para la hipótesis nula es la siguiente, donde M = 'Male', F = 'Female', Ap = 'Ape', Z = 'Zombie' y Al = 'Alien'.

$$\mu_M = \mu_F = \mu_{Ap} = \mu_Z = \mu_{Al}$$

La hipótesis nos sirve como punto de partida para enmarcar el estudio aunque, como hemos dicho, el objetivo final es identificar a los punks potencialmente más rentables, o identificar nuevos grupos (*clusters*) no evidentes de punks, por los que sería más seguro apostar (comprar o pujar) esperando beneficios.

## 1.3 Definición de métodos y herramientas empleadas

Para llevar a cabo el estudio usaremos los datos disponibles en la página oficial: <https://www.larvalabs.com/cryptopunks> y la página de refuerzo no oficial <https://defypunk.com/>.

Mediante Selenium hemos escrapeado los atributos de cada punk y el historial de transacciones. [Por falta de tiempo los compañeros Julio Abril y Francisco Gil nos han facilitado los datos económicos complementarios, el historial de transacciones de cada punk.]

Para normalizar este conjunto de datos diseñaremos una base de datos relacional usando MySQL y Python.

Una vez disponibles todos los datos necesarios pasaremos al análisis estadístico de cada variable, para pasar después a su análisis factorial y selección de variables para clusterizar. Tras la agrupación no-supervisada de los datos intentaremos encontrarle alguna lógica, por medio del estudio estadístico de los clusters para, finalmente, distinguir qué punks o qué grupos de punks son los indicados para invertir.

Tras comparar los grupos creados según las conclusiones del análisis estadístico y los creados según métodos no-supervisados, escogeremos las variables (features y target) que sean más prometedores para el modelado.

A continuación aplicaremos los modelos que hemos aprendido durante el itinerario a los grupos que nos interesan, presentando finalmente, las proyecciones resultantes de estos. Dado que trabajaremos con series temporales y valores monetarios (continuos) nos centraremos en modelos y predicciones de regresión lineal.

## 2. Análisis descriptivo y exploración

### 2.1 Características de los punks

Para el análisis exploratorio y descriptivo de las características físicas de los punks hemos usado los datos escrapeados de la página oficial.

“Cada CryptoPunk tiene una serie de atributos específicos. [Dejando de lado su tipo,] estos atributos también son responsables de hacer que un CryptoPunk sea raro y de aumentar su valor en función de ello. Por ejemplo, sólo 44 punks de los 10.000 llevan gorro. Por lo tanto, un punk con gorro tendrá obviamente más valor en comparación con otros atributos.

Algunos de los atributos más raros pueden determinar el valor de los CryptoPunks. Por ejemplo, el CryptoPunk #2964 es un punk masculino, que no es un género raro. Sin embargo, el sólo hecho de que un gorro, hizo que se vendiera a un precio de 660.121 dólares.”<sup>4</sup>



punk #2964

Según la página web citada, ciertos atributos raros hacen que el valor del punk aumente, independientemente de que su tipo sea común. Este tipo de cuestiones son las que intentaremos clarificar con la siguiente exploración. Empezamos por los distintos tipos de punks.

#### 2.1.1 Tipos de punks

---

<sup>4</sup> <https://101blockchains.com/cryptopunks/>

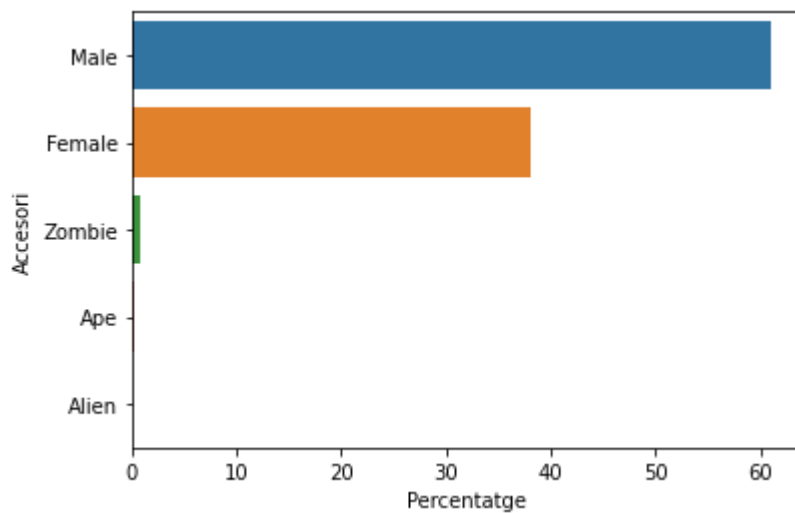


Fig 1. Porcentaje de los tipos de punks sobre el total

<u>Tipo</u>	<u>Total</u>	<u>%</u>
Male	6039	60.39
Female	3840	38.40
Zombie	88	0.88
Ape	24	0.24
Alien	9	0.09

Como vemos hay dos tipos principales que predominan: 'Male' y 'Female', agrupando el 98, 79% de los punks. A parte de estos encontramos tres tipos residuales: 'Zombie' (88 punks), 'Ape' (24 punks) y 'Alien' (9 punks).

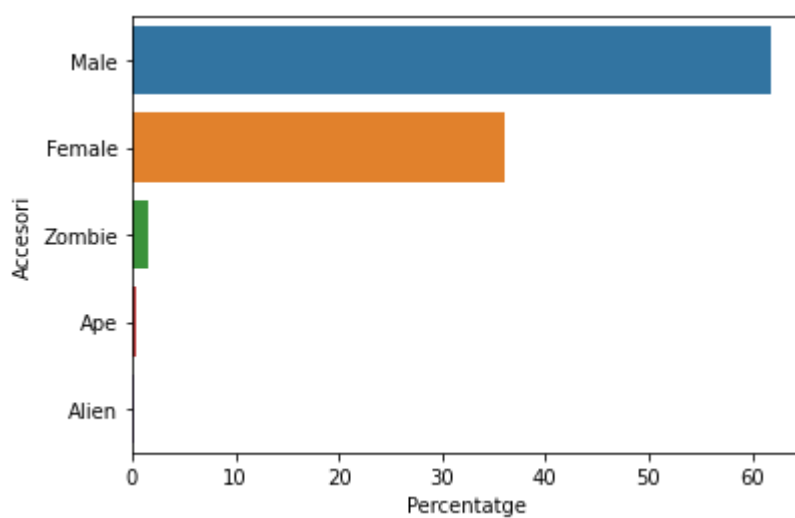


Fig 2. Porcentaje de los tipos de punk sobre los 1000 punks vendidos más caros

<u>Tipo</u>	<u>%</u>
-------------	----------

Male	61.77
Female	36.05
Zombie	1.61
Ape	0.40
Alien	0.16

En el caso de los 1000 con el precio más elevado, vemos que la distribución aumenta en los 3 tipos más raros, siendo más significativo el caso de los 'Zombie', que doblan su presencia.

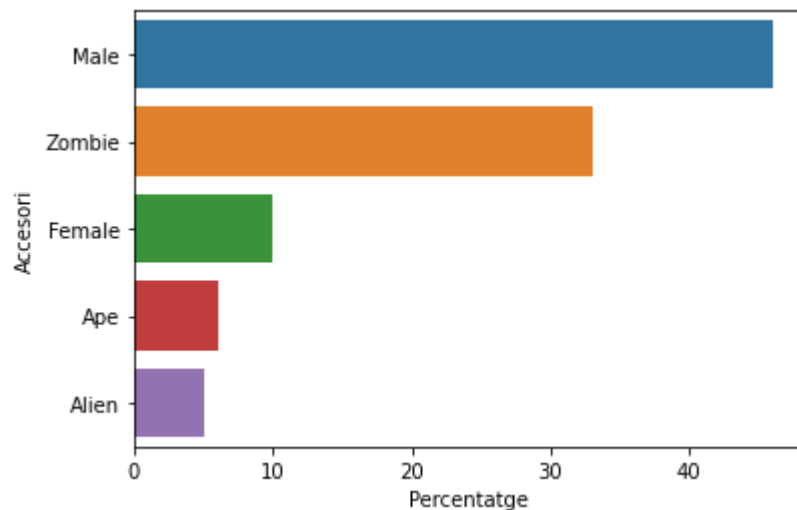


Fig 3. Porcentaje de los tipos de punk sobre los 1000 punks vendidos más caros

## 2.1.2 Atributos de los punks

Pasamos ahora a los 87 atributos o accesorios distintos que pueden tener los punks:

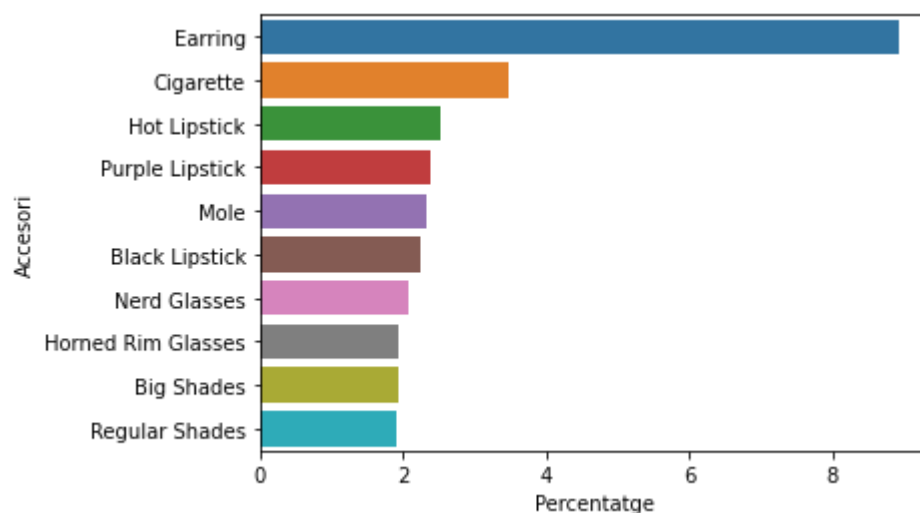


Fig 4. Porcentaje de los 10 atributos más repetidos

Observamos que 'earring' es el atributo más común de todos (8,93%), apareciendo casi el triple de veces más que el siguiente accesorio más frecuente ('cigarette', con un 3,49%). A partir de este punto el descenso en la distribución es muy paulatino:

Atributo	%
Hot Lipstick	2.53
Purple Lipstick	2.38
Mole	2.34
Black Lipstick	2.24
Nerd Glasses	2.08
Horned Rim Glasses	1.95
Big Shades	1.95
Regular Shades	1.92

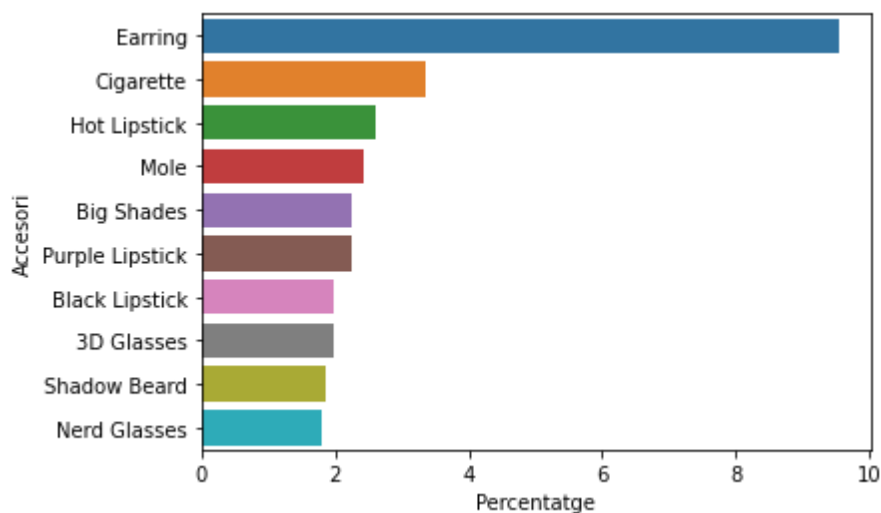


Fig 5. Porcentaje de los 10 atributos más repetidos entre los 1000 pinks vendidos a un precio más elevado

Comparando la distribución de los atributos más comunes de todos los pinks con la correspondiente a los 1000 pinks con el precio más alto observamos algunos cambios. Los tres atributos más frecuentes se mantienen, pero a partir de ahí 'Mole' pasa de quinta a cuarta posición y 'Big shades' pasa de novena a quinta. 'Purple Lipstick' desciende a sexta posición y 'Black lipstick' a séptima. Aparecen los atributos '3D Glasses' y 'Shadow beard' en octavo y noveno puesto respectivamente. Por último 'Nerd Glasses' desciende del sexto al décimo lugar.



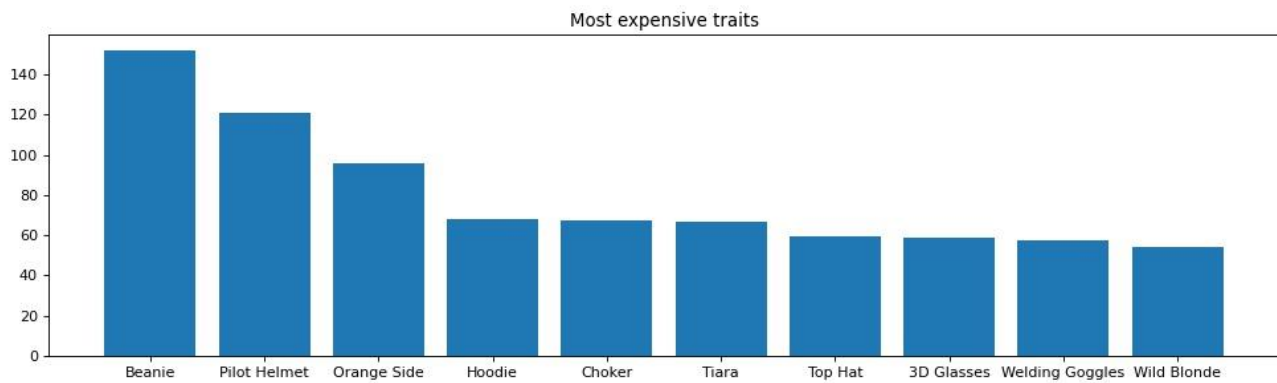


Fig 6. Atributos con la media de precio más elevado (en Ethereum).

Atributos	Precio (eth.)	N_punks
Beanie	152.00	44
Pilot Helmet	121.20	54
Orange Side	96.03	68
Hoodie	67.75	259
Choker	67.33	48
Tiara	67.00	55
Top Hat	59.48	115
3D Glasses	58.56	286
Welding Goggles	57.40	86
Wild Blonde	54.43	144

Con este último gráfico, en vez de tomar a los 1000 punks con el precio más elevado y buscar los atributos más frecuentes, tomamos el precio medio de cada atributo en Ethereum, sumando los precios de los punks con dicho elemento y dividiéndolos por su número. Como decía el fragmento del principio de esta sección, el accesorio 'Beanie' nos permite ilustrar el hecho de que un atributo menos frecuente hace más 'atractivo' al punk, aumentando su precio. Lo ilustraremos con un histograma de los 10 atributos más raros para ver si coinciden con los más cotizados:

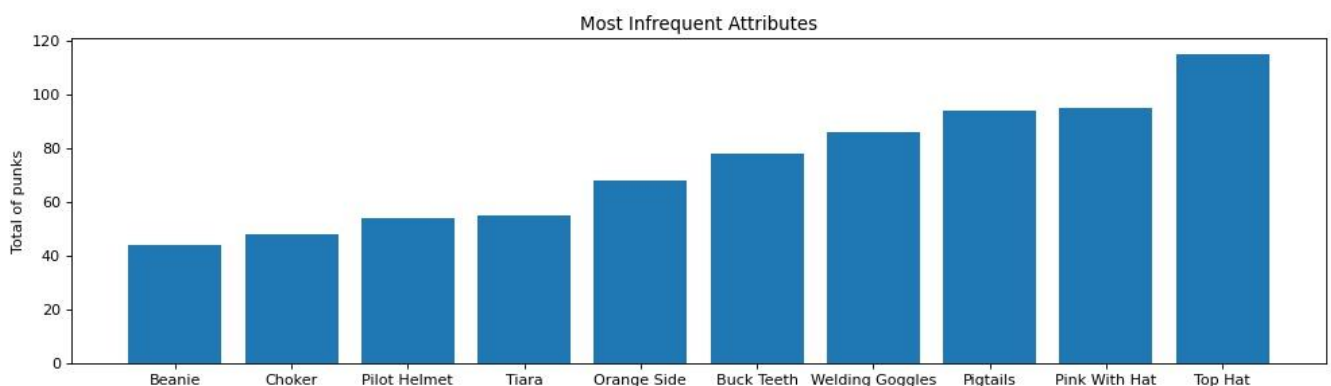


Fig 7. Histograma de los atributos menos frecuentes en orden ascendente.

A excepción de 'Buck Teeth', 'Pigtales' y 'Pink with Hat', los otros siete atributos menos frecuentes están también en la lista de los 10 atributos más cotizados. Podemos establecer una cierta correlación positiva entre atributos menos frecuentes y su precio medio entre los punks.

Ahora vamos a tener en cuenta el número de accesorios por cada punk.

### 2.1.3 Número de atributos

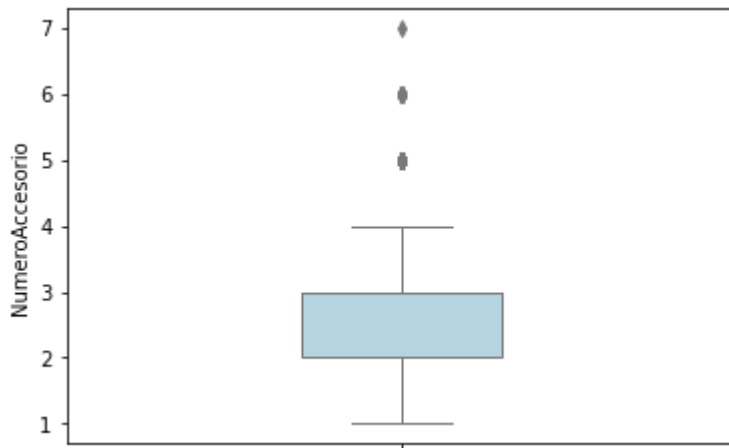


Fig 8. Boxplot del número de accesorios por punk. Cuartiles y outliers.

<u>N_atributos</u>	<u>N_punks</u>	<u>%</u>
3	4501	45.01
2	3560	35.60
4	1420	14.20
1	333	3.33
5	166	1.66
6	11	0.11
0	8	0.08
7	1	0.01

El número de atributos por punk oscila entre los 0 y los 7, siendo los más comunes de 2 a 4 atributos (94,81%). Los punks que tienen 0, 6 o 7 atributos son minoritarios, con 8, 11 y un solo punk, respectivamente. Hay 333 punks con un solo atributo y 166 con 5 atributos.

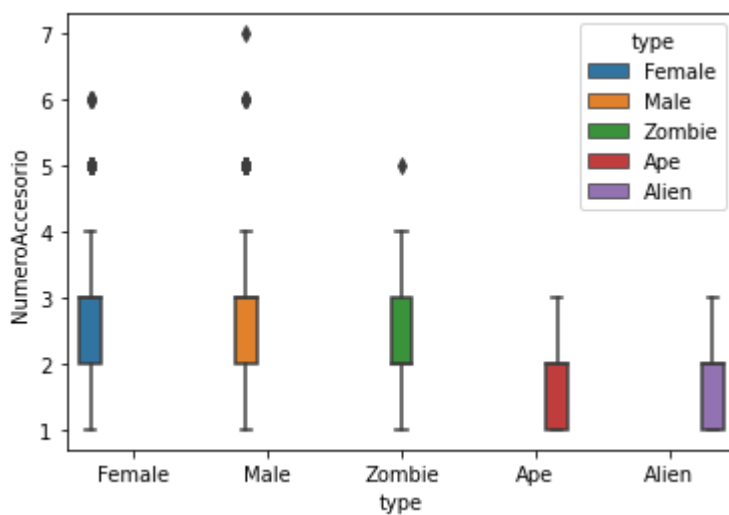


Fig. 9 Boxplot del número de atributos según el tipo del punk.

Podemos ver como varían los cuartiles y los outliers del número de atributos según los tipos de los punks. El único punk con siete atributos es un 'Male'. En estos boxplots no aparecen los punks con 0 atributos, aunque posiblemente serían considerados outliers también, por su leve incisión. Lo que más destaca es la diferencia en la distribución de los cuartiles de los tipos 'Ape' y 'Alien', que en vez de ir de 1 a 4, con el intercuartil de 2 a 3, van de 1 a 3 atributos, con el intercuartil de 1 a 2. A diferencia del resto, estos dos no tienen outliers.

Hacemos el mismo análisis para los 1000 punks con precios más elevados:

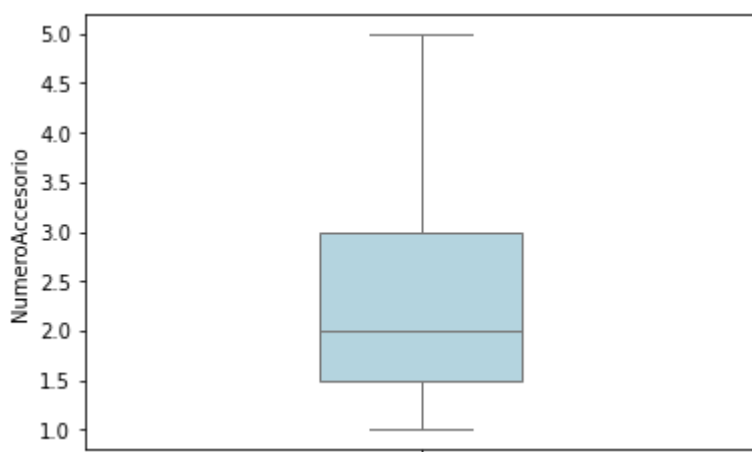


Fig 10. Boxplot del número de atributos de los 1000 punks más caros

<u>N_atributos</u>	<u>%</u>
2	38.98
3	30.51
1	25.42
4	3.39
5	1.69

En el caso de los 1000 punks con el precio más elevado desaparecen los outliers del boxplot. La gran mayoría tienen 1, 2 o 3 atributos (94,91%), mientras que un 3,39% de estos tiene 4 atributos y sólo un 1,69% tiene 5.

## 2.2 Historial de transacciones

Para el análisis descriptivo y exploratorio de los datos económicos hemos usado el historial de transacciones asociado a cada punk. Cada transacción tiene una fecha, el id del punk en cuestión (de 0 a 9999), qué tipo de transacción es (puja, venta, oferta, transferencia...), el usuario que hace la transacción. En el caso de cierto tipo de transacciones también encontramos el usuario emisor y el usuario receptor, así como también el importe en ethereum.

A estos datos, que son los escrapeados del portal defypunk.com, nosotros hemos añadido el precio ajustado de cierre del valor de ethereum en dólares, así como la varianza del mismo, para tener una referencia más inmediata y encontrar posibles relaciones.

El análisis incluye estudio de nulos, estadística descriptiva de variables categóricas y numéricas y visualizaciones. A su vez daremos conclusiones de cada apartado y posibles nichos para investigar.

## 2.2.1 Categóricas: Tipo de transacciones

El **estudio de nulos** nos indica que hay 44.047 registros de transacciones **sin valor monetario**, estos están divididos porcentualmente en los siguientes tipos de transacciones (%):

<u>Tipos</u>	<u>%</u>
Offer withdrawn	41,86
Transfer	25,08
Claimed	22,70
Transfer (Wrap)	5,48
Transfer (Unwrap)	4,88

Podemos observar qué tipo de transacciones no precisan valor monetario, y cómo se distribuyen: las ofertas retiradas, las transferencias en el Marketplace oficial, las transferencias en mercados no oficiales (wrap y unwrap) como *opensea.io*, y claimed. Los comparamos con la distribución de los tipos de transacciones sobre el total:

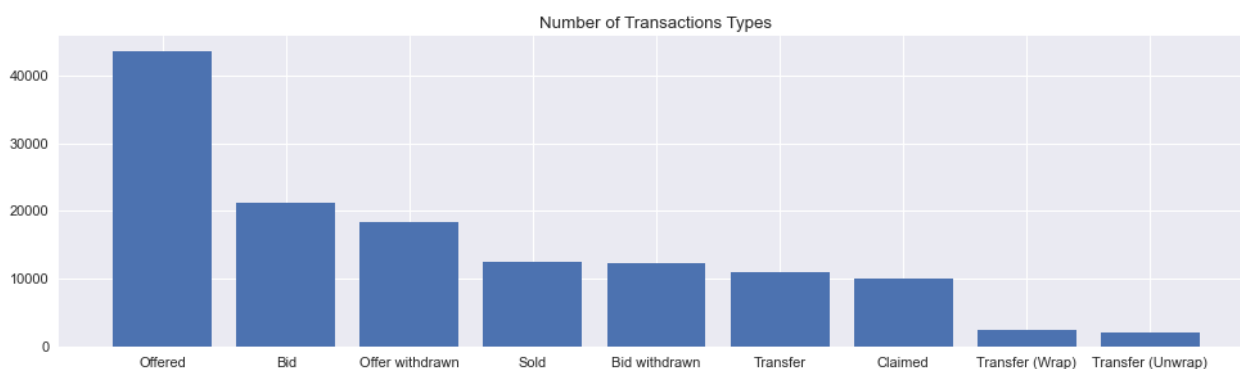


Fig.11, Number of transactions by type

### Conclusiones tipo de transacciones

Podemos dividir los tipos de transacciones de punks entre usuarios por si van acompañadas de un valor monetario o no. Por una lado tenemos las transacciones con precio: 'Offered', 'Bid', 'Sold' y 'Bid withdrawn'. Por otro lado las transacciones o operaciones sin precio: 'Offer withdrawn', 'Transfer', 'Claimed', 'Transfer (Wrap)' y 'Transfer (Unwrap)'.

<u>Tipos</u>	<u>Número</u>	<u>%</u>	<u>Precio s/n</u>
Offered	43556	32,61	Sí
Bid	21270	15,92	Sí
Offer withdrawn	18440	13,80	No
Sold	12494	9,35	Sí
Bid withdrawn	12215	9,14	Sí
Transfer	11046	8,27	No

Claimed	10000	7,49	No
Transfer (Wrap)	2413	1,81	No
Transfer (Unwrap)	2148	1,61	No

-> Principalmente a nosotros nos interesan aquellas transacciones que vienen acompañadas de un precio, aunque también pueden llegar a ser interesante investigar porque se dan ciertas transferencias y cómo varía el precio de los punks después de una transferencia o puja.

## 2.2.2 Categóricas: Usuarios (emisario, receptor y usuario asociado a la transacción)

Comparando los valores de los tipos de transacciones entre 'From' (emisario) y 'To' (receptor) podemos adivinar su función. Podemos separar los tipos de transacción en si coinciden cuantitativamente entre 'From' y 'To' o no:

### Tipos

Coinciden: Sold (12494),  
Transfer (11046),  
Transfer (Wrap) (2413),  
Transfer (Unwrap) (2148)

Únicos de 'From': Bid (21270),  
Bid withdrawn (12215)

Únicos de 'To': Claimed (10000),  
Offered (383)

'Claimed' fue el primer propietario del punk y coincide en número con los 10000 punks únicos. De las pujas solo tenemos el emisario. En el caso de 'Offered', vemos que tenemos 383 con receptor, pero parecen casos aislados ya que hay un total de 43.556 ofertas, y solo estas tienen destinatario.

### **Conclusiones users**

Tenemos más de 110.000 usuarios únicos en el historial de transacciones. En el caso de emisarios y destinatarios el número se reduce con 4.212 y 4.537 respectivamente.

### 2.2.3 Numéricas: cantidad de transacciones por id de punk

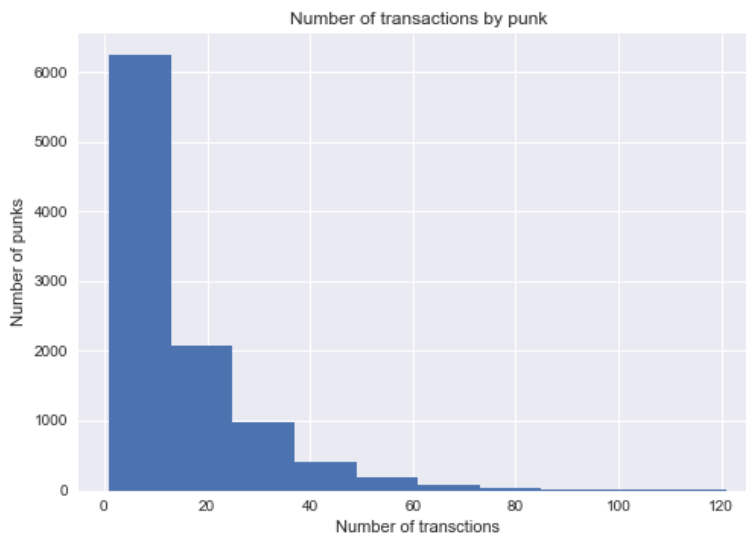
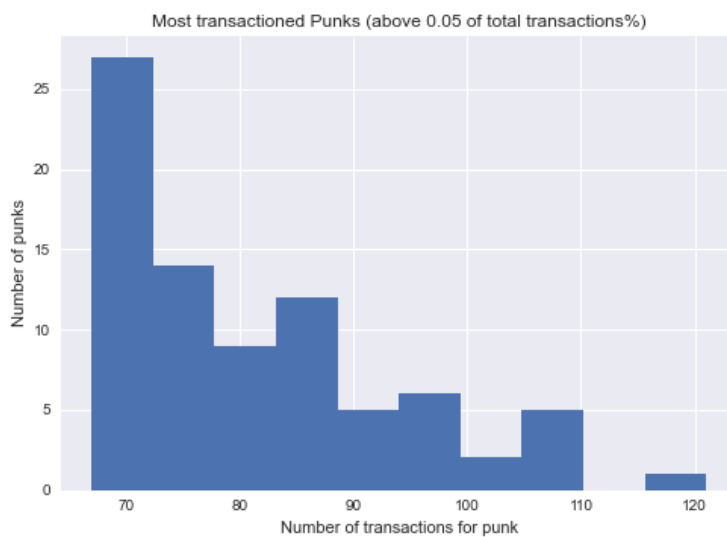


Fig.2 Number of transactions for each punk

count	10000
mean	13.36
median	8.00
std	13.84
min	1.00
25%	4.00
50%	8.00
75%	19.00
max	121.00



count	81
mean	80.88
median	77.00
std	12.28
min	67.00
25%	71.00
50%	77.00
75%	88.00
max	121.00

Fig. 12 Number of transactions for the most transactioned punks. In the second, only those that individually represent more than the 0.05% of the total transactions.

### Conclusiones número de transacciones por punk

Las visualizaciones y estadística descriptiva anteriores son resultado de agrupar por valores únicos los datos de la variable 'number' o id\_punk. En las visualizaciones en forma de histograma observamos una distribución muy asimétrica, ladeada a la derecha (right-skewed), con una pendiente inversamente exponencial. Cogiendo solo los ids que supongan por lo menos un 0.05% de las transacciones totales, la cantidad de valores únicos disminuye notablemente (de 10.000 a 81) y la pendiente se hace menos pronunciada.

De este análisis podemos concluir que, a excepción de pocos punks, la mayoría tiene un número de transacciones similar. La media se encuentra en las 13,38 transacciones por punk, con una

desviación estándar de 13,86. El mínimo es 1 con un máximo de 121, encontrándose el 25 percentil a 4, el 50 a 8 y el 75 a 19 transacciones. Hay un gran salto entre el 75 percentil y el máximo (de 19 a 121 transacciones).

## 2.2.4 Numéricas: estudio de ventas

Elegimos sólo las ventas de entre los tipos de transacciones. Su estadística descriptiva:

count	5699
mean	2.19
std	1.47
min	1.00
25%	1.00
50%	2.00
75%	3.00
max	13.00

La media de veces que se ha vendido un punk es 2.19, con una desviación estándar de 1,47. Hay 4301 punks que NO han sido vendidos nunca y 5699 punks que han sido vendidos al menos una vez. De los punks vendidos, el máximo de veces es 13, estando el 25 percentil en 1 vez, el 50 en 2 veces y el 75 percentil en 3.

De estos datos podemos concluir que, de los punks que han sido vendidos alguna vez (56,99%), un 75% de ellos ha sido vendido un máximo de 3 veces.

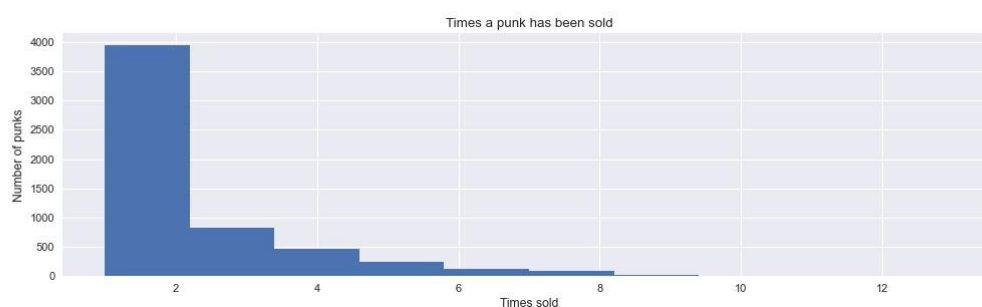


Fig. 13 Histogram of how many times punks are sold.

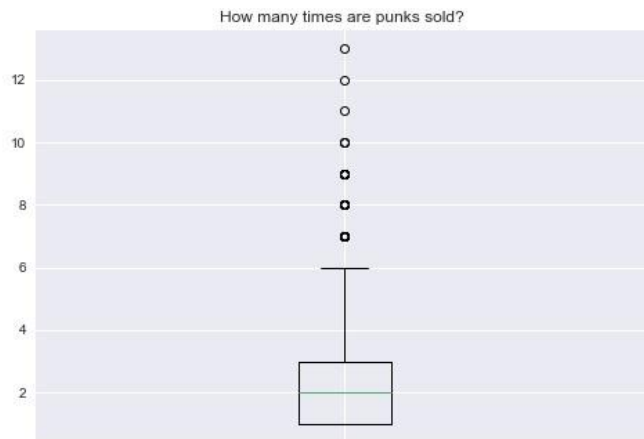


Fig. 14 Boxplot of how many times punks are sold. Quartiles and outliers.

Mediante las visualizaciones observamos que a partir de 6 veces, ese punk se considera un outlier. -> Podemos estudiar los beneficios reportados de estos outliers con el resto y si el beneficio es incremental con el número de ventas.

Hacemos un 'zoom', agrupando sólo los punks que han sido vendidos más de 6 veces:

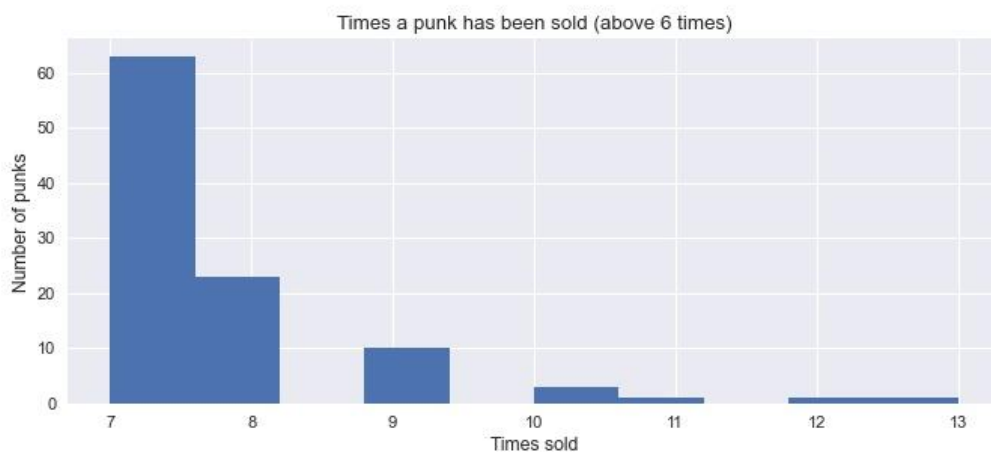


Fig. 15 Histogram of punks sold over 6 times (outliers)

Y su estadística descriptiva:

count	102
mean	7.66
std	1.10
min	7.00
25%	7.00
50%	7.00
75%	8.00
max	13.00

El total de punks en este grupo es de 102, un 1'02% del total. De igual forma la mayoría se encuentran en el número de 7 ventas, seguidos de 8 y 9, de forma descendiente. El resto de veces son residuales.



<u>Veces</u>	<u>N punks</u>
7	63 (0,63%)
8	23 (0,23%)
9	10 (0.1%)
10	3 (0.03%)
11	1 (0.01%)
12	1 (0.01%)
13	1 (0.01%)

Agrupamos los registros por ventas y sacamos la estadística descriptiva de las variables de precio, en ethereum y en dólares:

	Ethereum	Dolars
count	12494	12447
mean	14.78	29323.05
median	6.50	6754.52
std	59.19	112329.78
min	0.00	0.00
25%	0.60	134.06
50%	6.50	6754.52
75%	21.95	42648.71
max	4200.00	7670018.77

La media del precio histórico de venta en dólares es 29.360,04. El mínimo es 0 y el máximo son 7.670.018,77, con una desviación estándar de 112.284,38. El 50 percentil del precio histórico de venta en dólares se encuentra en los 6.784,11, teniendo el 25 percentil a 134,18 y el 75 a 42.714,17.

La media del precio histórico de venta en Ethereum es de 59,16. El mínimo son 0 y el máximo son 4200, con una desviación estándar de 59,16. El 50 percentil del precio histórico de venta en dólares se encuentra en los 6,5. El 25 percentil a 0.6 y el 75 a 21,95.

## 2.2.5 Numéricas: Precio ETH-Dólares (Adj. Close)

Para finalizar el estudio de transacciones nos fijaremos en el precio histórico del ethereum y de su varianza.

“Aproximadamente cada 3 meses, la actividad del mercado de criptopunks aumenta<sup>5</sup>. El servicio analítico Dune llamó la atención sobre los ciclos de hype de los tokens NFT más famosos, que tiene una duración de tres meses.

<sup>5</sup> <https://u.today/crypto-punks-market-activity-grows-quarterly-why-exactly>

En particular, el último pico local del coste medio de Crypto Punks NFT se observó en marzo. Al mismo tiempo, coincidió con el precio de ETH, la unidad de pago más común para los 'punks'.

Más tarde, en abril, el coste de la NFT de Crypto Punks marcó un máximo histórico. Sin embargo, el precio de ETH en la misma época resultó ser significativamente inferior a las cotizaciones de marzo.”

Como el último pico del coste medio de Crypto Punks se observó en marzo dividiremos el precio histórico en esa fecha:

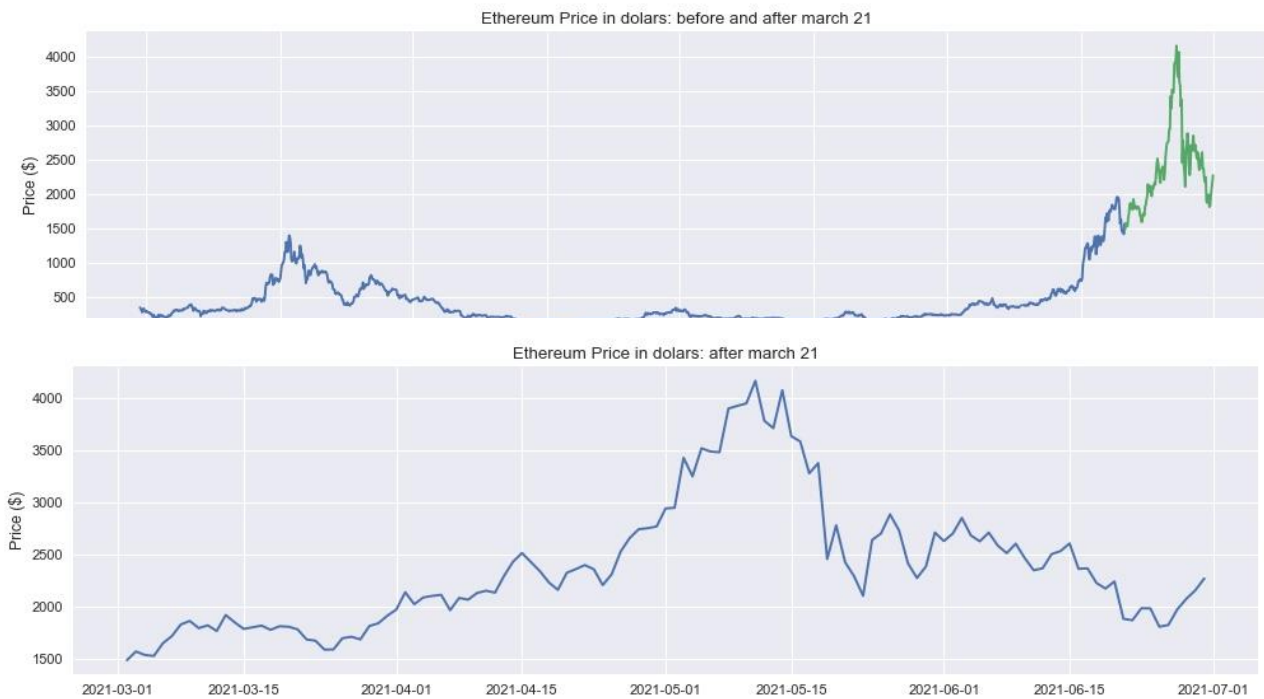


Fig.16 Ethereum historic price in dollars dividing by color the highest peak of its history after march '21.

Fig. 17 Zoom of Ethereum historic price in dollars from march to july '21.

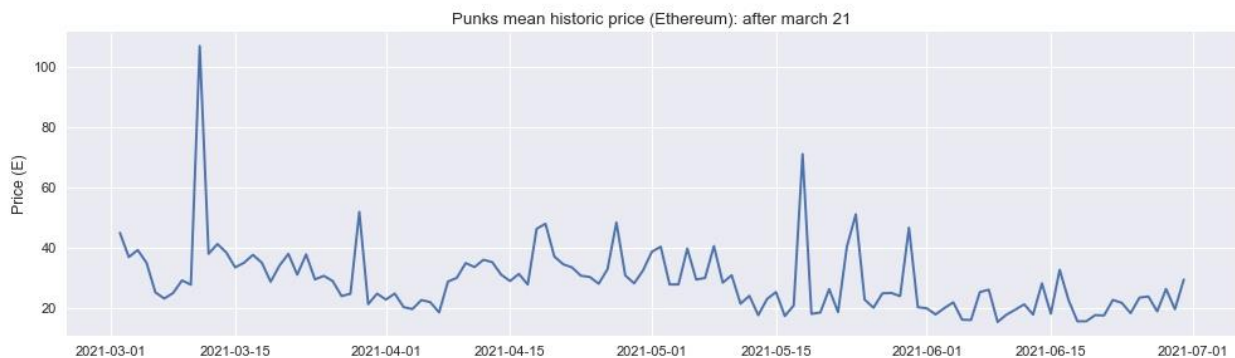


Fig 18. Historic mean price for punks in ethereum (from march to july '21)

Observamos un incremento importante del precio de Ethereum durante el último trimestre de 2017, alcanzando un pico histórico, solo alcanzado por durante el primer trimestre de 2021 y su subsiguiente máximo histórico el 12 de mayo del mismo año (4168\$). Desde el 1 de marzo a mediados de mayo el precio de Ethereum aumentó más de un 200% - de 1492 hasta 4168 dólares -, hasta bajar por debajo de los 2000 durante la tercera semana de junio. Estos datos nos indican la potencial volatilidad de la moneda con la que se transaccionan los cryptopunks y su potencial influencia en sus transacciones. Finalmente visualizamos la varianza diaria en comparación a la evolución del precio. Observamos a simple vista una cierta correlación positiva entre los picos en el incremento de precio del ethereum y su variación diaria. En los días en que el precio sube más, también aumenta la variación. En concreto, usando la correlación de Pearson obtenemos una correlación positiva de 0.77 entre variación y precio de cierre del ethereum.

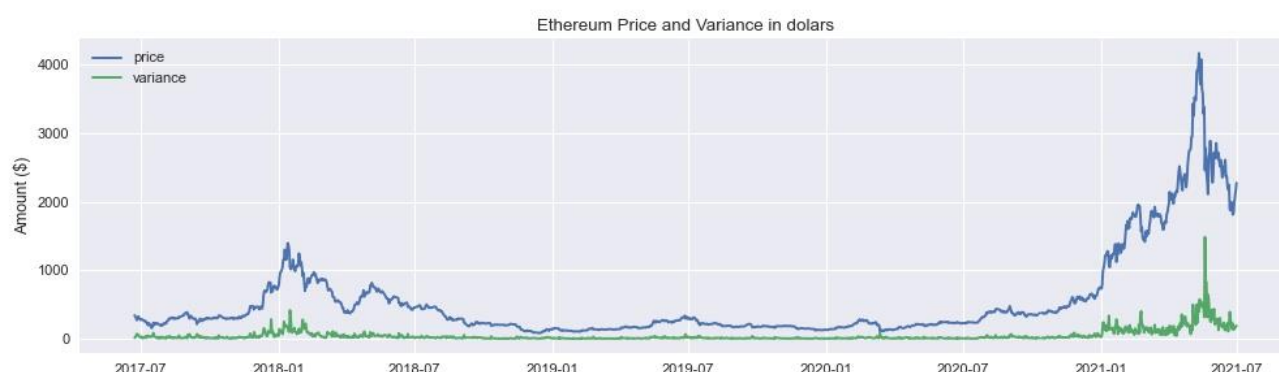


Fig 19. Comparison between Ethereum historic price in dollars and its historic daily variance.

### 3. Análisis factorial

Finalmente nuestro análisis factorial ha consistido en seleccionar y agrupar las variables de las dos fuentes de datos - las características de los punks y sus transacciones - en diferentes modos para ver como los agrupaba a posteriori un algoritmo de clustering no-supervisado.

En este proceso hemos creado la variable rendimiento, clave para nuestro estudio. Visto que el precio de los punks definitivamente varía según sus características y sus tipos, queremos ver por qué punk o grupo de punks vale la pena invertir. Para esto hemos hecho un contador de ventas por cada punk, de manera que tomando la venta anterior y la subsiguiente hemos calculado el beneficio o rendimiento que ha traído la compra y venta de ese punk al vendedor. Su cálculo es el siguiente:

$$\text{'Rendimiento'} = \text{precio de venta (\$/)} / \text{precio de compra (\$/)}$$

Con esta variable ya tenemos un indicador del beneficio reportado por el punk, que en definitiva es nuestro objetivo. No sólo comprarlo, si no comprarlo para venderlo más caro después. Un resultado de 1 en el rendimiento de una venta indicaría un beneficio neutro, igual a 0. Menos de 1 indicaría pérdidas y más de 1 beneficios.

Finalmente hemos reunido la rarity del punk<sup>6</sup>, el precio de venta de los punks en ethereum, la 'id' de cada punk, el precio de cierre del ethereum en dólares el día de la transacción, el número de venta (1a vez que se vende ese punk, 2a, 5a etc.) y el tipo del punk (pasado a numérico usando label encoder). Los atributos los hemos descartado en este paso para simplificar y ver qué agrupaciones obteníamos.

## 4. Clustering no-supervisado

El objetivo de nuestro modelo no-supervisado era identificar si las divisiones resultantes de las transacciones de venta variaban significativamente en relación al rendimiento.

Para el modelo usamos el algoritmo K-means con las variables id, type, rendimiento, rarity, adj\_close, amount\_eth y num\_venta. Usando elbow method hemos elegido 3 clusters. Estos quedan divididos de la manera siguiente:

<u>Cluster</u>	<u>Núm. registros</u>
0	12089
1	49
2	342

De primeras ya vemos un desbalance muy importante en cómo se han dividido los clusters. La gran mayoría de registros pertenecen al primer cluster (96,87%), habiendo 0,4% en el segundo cluster y un 2,73% en el tercero. Esta división de primeras puede parecer muy problemática porque parece como si prácticamente no hubiera distinguido grupos de punks diferentes. Pero recordamos que estos registros son transacciones de ventas, tenemos que ahondar más si queremos distinguir los punks y sus grupos. Vamos a profundizar un poco más observando la variable 'rendimiento'. Vemos la media de rendimiento de cada cluster:

<u>Cluster</u>	<u>Avg. Rendimiento</u>	<u>eliminando las primeras ventas</u>
0	9.16	17.27
1	82.71	144.74
2	1000.15	1000.15

Hay una diferencia de más de cien veces entre la media de rendimiento de los punks del primer cluster(0), con el tercer cluster(2). Aún eliminando los registros de las primeras ventas (en las que el rendimiento es igual a 0) la diferencia sigue siendo de más de 50 veces entre el primer y el tercer cluster. Veamos la media de rarity por cluster:

<u>Cluster</u>	<u>Avg. Rarity score</u>
0	0.102
1	0.104
2	0.105

No hay una gran diferencia significativa pero el tercer cluster tiene la mayor media de rarity score. Veamos la media del precio de las ventas en eth. :

---

<sup>6</sup> Variable obtenida del cálculo de 'rareza' de los atributos de cada punk. Esta métrica la obtuvimos del portal: <https://0xtycoon.github.io/punk-ranks/>, que crearon su propio score de rarity, diferente al ranking de defypunk.com.

<u>Cluster</u>	<u>Avg. Amount Eth</u>
0	13.60
1	177.08
2	33.51

A diferencia del rendimiento, la media en el precio de venta es aproximadamente 5 veces más elevado en el segundo cluster(1) que en el tercero(2). En cierta forma, este factor puede hacer aún más atractivos los punks del tercer cluster ya que implican un desembolso menor y un mayor rendimiento. Por último miremos la distribución de los tipos de punk en cada cluster:

	<u>Type</u>	<u>N</u>	<u>%</u>
<u>Cluster 0:</u>	Male	8020	66,34
	Female	3954	32,71
	Zombie	80	0,66
	Ape	28	0,23
	Alien	7	0,06
<u>Cluster 1:</u>	Male	29	59,18
	Female	18	36,73
	Alien	2	4,08
<u>Cluster 2:</u>	Male	237	69,30
	Female	101	29,53
	Zombie	2	0,58
	Ape	1	0,29
	Alien	1	0,29

Siendo el rendimiento de cada venta uno de los factores más determinantes a la hora de elegir invertir en un punk, vamos a profundizar en el estudio del tercer cluster.

## 4.1 Examen del Tercer Cluster (2)

Este cluster contiene 342 punks distintos y 953 registros de transacciones, con un mínimo de una y un máximo de 5 ventas. Tenemos que entender primero de todo, que hay una estrecha relación entre el

rendimiento y el número de venta en el historial de ventas del punk:

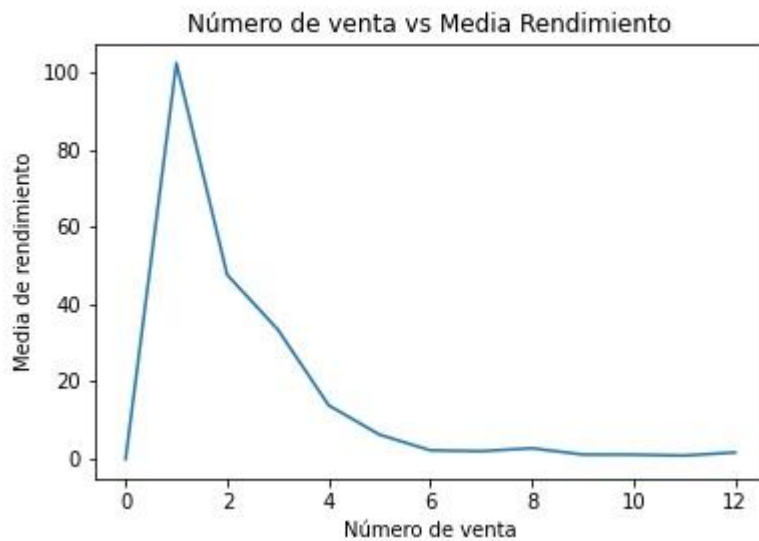


Fig. 20 Relación rendimiento medio por cada vez que se vende un punk de todas las ventas.

Esta correlación negativa (cuantas más veces se vende un punk más disminuye su rendimiento medio) es importante tenerla en cuenta en relación a nuestro cluster, ya que podría darse el caso que se hubieran agrupado por el número de venta, hecho que no nos daría tanta ventaja a la hora de elegir el punk. Estamos hablando de rendimiento medio, e ir tan sólo a por los punks que han sido vendidos una vez (aunque sin duda no sería la peor estrategia) no creemos una estrategia con muchas garantías. Tendríamos que investigar la evolución del rendimiento de las segundas ventas en la historia de los cryptopunks, pues podría ser que esta hubiera ido disminuyendo desde el principio y que ahora mismo no fuera una estrategia tan rentable.

Hecho este aparte, vayamos a ello. En este cluster solo tenemos registros de segundas a quintas ventas. De la primera venta (0) no se puede hacer el cálculo de rendimiento pues no hay precio de compra, los punks se 'reclamaron' ('Claimed') gratuitamente cuando aparecieron.

<u>Núm Venta</u>	<u>N</u>
1	700
2	182
3	59
4	12

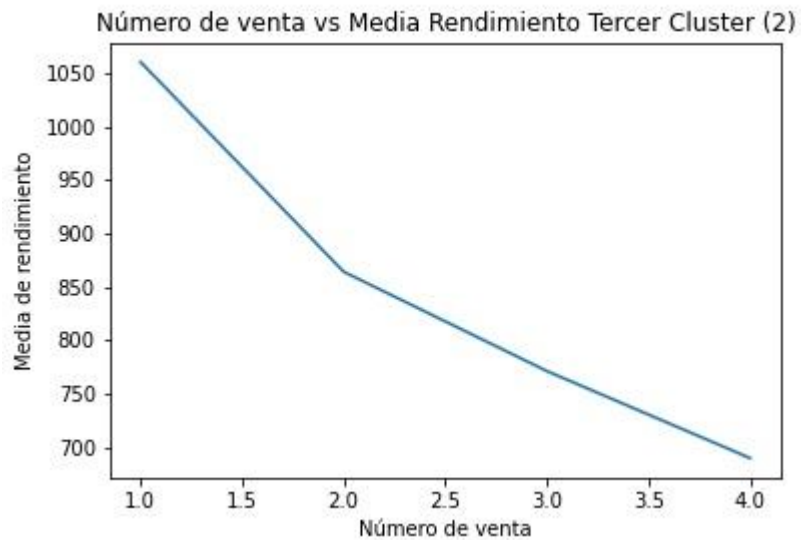


Fig 21. Rendimiento medio por número de ventas en el tercer cluster (2).

Aunque observamos la misma tendencia de correlación negativa, en este cluster es menos aguda y el rendimiento sigue siendo muy superior a la media en cada una de ellas:

1	1060.73
2	864.24
3	771.25
4	689.77

Miramos el rendimiento medio por tipo dentro de este cluster:

<u>Type</u>	<u>Avg. Rend.</u>	<u>N</u>
Male	941.86	661
Female	899.51	286
Zombie	25195.51	3
Ape	1220.92	2
Alien	344.09	1

Las ventas de los tipos más raros no nos interesan, son residuales. Veamos cuantos punks hay de cada tipo:

Male	237
Female	101
Ape	2
Zombie	1
Alien	1

### 4.1.1 Punks seleccionados

La id del alien es #2890, la del zombie #2424 y la de los ape: #4156 y #6965. Son los siguientes:

	id	type_x	type_le	Redimiento	rarity	adj_close	Num_venta	amount_eth	Kmeans_Clustering_types
0	2424	Zombie	4	25195.513196	0.008871	2643.59	1	325.0	2
73	4156	Ape	1	1934.440331	0.002851	1937.45	3	650.0	2
748	6965	Ape	1	507.393482	0.001236	1960.16	3	800.0	2
1023	2890	Alien	0	344.091869	0.000677	1230.99	1	605.0	2



Vamos a ver las clases mayoritarias, Male y Female y cinco punks de cada una de los cuales se haya sacado más rendimiento en alguna de sus ventas:

- Male: punks #6552, #3134, #6095, #6721 y #2376:

	id	type_x	type_le	Redimiento	rarity	adj_close	Num_venta	amount_eth	Kmeans_Clustering_types
3	6552	Male	3	8447.378898	0.056414	2157.66	1	59.0	2
5	3134	Male	3	7134.473684	0.067263	1594.76	1	15.3	2
10	6095	Male	3	3551.240591	0.067273	2316.06	1	55.0	2
13	6721	Male	3	3385.286889	0.094198	2662.87	1	150.0	2
15	2376	Male	3	3101.031875	0.104356	3587.51	1	16.0	2



- Female: punks #382, #2000, #3162, #4218 y #4072:

	id	type_x	type_le	Redimiento	rarity	adj_close	Num_venta	amount_eth	Kmeans_Clustering_types
8	382	Female	2	3676.938687	0.021581	3902.65	1	130.00	2
17	2000	Female	2	3092.043035	0.021970	2299.19	1	120.00	2
41	3162	Female	2	2371.523342	0.107588	1575.85	1	24.50	2
48	4218	Female	2	2185.377144	0.057266	2431.95	1	99.00	2
51	4072	Female	2	2120.230203	0.151551	2643.59	2	17.42	2





## 5. Regresión lineal y proyecciones

Nuestro ejercicio de regresión lineal no nos ha servido para el propósito del informe. Primero probamos con todas las variables y teníamos un R2 score negativo muy alto. Después probamos de reducir las dimensiones de los features para ver si simplificando los resultados mejoraban. No hemos tenido suerte y los resultados dejan mucho que desear. Viendo que el tiempo era un factor importante hemos decidido centrar esfuerzos en el análisis de clusters y posibilidades de estudio. Aún así dejamos los resultados, con las predicciones y las variables utilizadas para ello.

### 5.1 Target Precio de Venta en Eth.

**Target (y) =** amount\_eth (Precio de la venta en Ethers)

**Features (X) =** type\_le (Tipo de punk)  
 Redimiento (Precio de venta/Precio de compra en dólares)  
 adj\_close (Precio de cierre Ethereum día de la venta en dólares)

	<u>MAE</u>	<u>RMSE</u>	<u>R2</u>
LinearRegression	321398.33	106056834382.99	-1729449913.08
Ridge	13869.64	205653478.06	-3353553.65
Lasso	5.93	61.44	-0.002
DecisionTreeRegressor	17.62	371.07	-5.05
RandomForestRegressor	11.41	175.66	-1.86

#### 5.1.1 Predicciones

	Actual	Predicted		Actual	Predicted
0	400.585	3.618850e+09	0	400.585	672.386
1	862.766	4.856962e+09	1	862.766	672.386
2	543.739	3.203200e+09	2	543.739	672.386
3	533.664	4.841180e+09	3	533.664	672.386
4	664.153	3.966089e+09	4	664.153	672.386
...	...	...	...	...	...
204	376.336	3.263805e+09	204	376.336	672.386
205	729.449	3.682934e+09	205	729.449	672.386
206	495.117	3.929364e+09	206	495.117	672.386
207	420.175	3.648041e+09	207	420.175	672.386
208	656.538	3.647874e+09	208	656.538	672.386

Predicciones del 'y\_test' de los modelos LinearRegression y RandomForestRegressor, respectivamente, para el target 'amount\_eth'.

## 5.2 Target variable Rendimiento

**Target (y) =** Rendimiento (Precio de venta/Precio de compra en dólares)

**Features (X) =** type\_le (Tipo de punk)  
amount\_eth (Precio de la venta en Ethers)  
adj\_close (Precio de cierre Ethereum día de la venta en dólares)

	<u>MAE</u>	<u>RMSE</u>	<u>R2</u>
LinearRegression	3828202311.94	1.4954+19	-156273400087696.25
Ridge	453.67	287129.43	-2.0005
Lasso	243.77	96133.63	-0.005
DecisionTreeRegressor	363.66	227646.91	-1.38
RandomForestRegressor	248.62	96937.016	-0.013

### 5.2.1 Predicciones

	Actual	Predicted		Actual	Predicted
0	400.585	3.618850e+09	0	400.585	688.634
1	862.766	4.856962e+09	1	862.766	688.634
2	543.739	3.203200e+09	2	543.739	688.634
3	533.664	4.841180e+09	3	533.664	688.634
4	664.153	3.966089e+09	4	664.153	688.634
...	...	...	...	...	...
204	376.336	3.263805e+09	204	376.336	688.634
205	729.449	3.682934e+09	205	729.449	688.634
206	495.117	3.929364e+09	206	495.117	688.634
207	420.175	3.648041e+09	207	420.175	688.634
208	656.538	3.647874e+09	208	656.538	688.634

Predicciones del 'y\_test' de los modelos LinearRegression y RandomForestRegressor, respectivamente, para el target 'Rendimiento'.

## 6. Confirmación/Refutación de la Hipótesis inicial

Nuestro trabajo consistía primeramente en probar la hipótesis nula, según la cual:

- ( $H_0$ ) : El precio de venta medio (Eth.) es igual en cada uno de los cinco tipos de punks ('Male', 'Female', 'Ape', 'Zombie' y 'Alien')
- ( $H_A$ ) : El precio de venta medio (Eth.) difiere en uno o más de los cinco tipos de punks ('Male', 'Female', 'Ape', 'Zombie' y 'Alien')
- Valor de significación  $\alpha = 0.05$ .

La notación estadística para la hipótesis nula es la siguiente, donde M = 'Male', F = 'Female', Ap = 'Ape', Z = 'Zombie' y Al = 'Alien'.

$$\mu_M = \mu_F = \mu_{Ap} = \mu_Z = \mu_{Al}$$

### 6.1 Demostración estadística ANOVA y F-Test

El método de análisis de la varianza en este contexto se centra en responder a la pregunta:

‘¿Es la variabilidad de las medias muestrales tan grande que parece improbable que se deba únicamente al azar?’

Consideraremos simultáneamente los cinco grupos y evaluaremos si sus medias muestrales difieren más de lo que cabría esperar de la variación natural.

Llamamos a esta variabilidad Mean Square between Groups (MSG), y tiene un grado de libertad asociado,  $df_G = k - 1$  cuando hay  $k$  grupos. El MSG puede considerarse como una fórmula de varianza escalada para las medias. Si la hipótesis nula es verdadera, cualquier variación en las medias de la muestra se debe al azar y no debería ser demasiado grande.

El MSG es, por sí solo, bastante inútil en una prueba de hipótesis. Necesitamos un valor de referencia para saber cuánta variabilidad debe esperarse entre las medias de la muestra si la hipótesis nula es verdadera. Por lo tanto, calculamos una estimación de la varianza agrupada, el Mean Square Error (MSE), que tiene un valor asociado de grados de libertad  $df_E = n - k$ . El MSE funciona como una medida de la variabilidad dentro de los grupos.

Como hemos dicho, cuando la hipótesis nula es verdadera, cualquier diferencia entre las medias de la muestra se debe sólo al azar, y la MSG y el MSE deberían ser aproximadamente iguales. Como estadística de prueba para el ANOVA, examinamos la fracción entre el MSG y MSE :

$$F = MSG / MSE$$

Si nuestra hipótesis nula es correcta de que todas las medias de los grupos son iguales, el numerador y el denominador deberían ser aproximadamente iguales y el resultado estadístico debería estar cerca de 1,0. Una ratio más elevada - cuando el resultado estadístico de F es muy superior a 1,0 - implica que las muestras proceden de poblaciones con valores medios muy diferentes.

Utilizaremos el paquete `scipy f_oneway()` para probar nuestra hipótesis nula de que dos o más grupos tienen la misma media poblacional.

Necesitamos evaluar el F-Statistic contra un punto crítico en la F-Distribution para determinar si nuestro resultado es significativo. Los datos iniciales de las medias de la población y los tipos son los siguientes:

	N	Mean	SD	SE	95% Conf.	Interval
<b>type_x</b>						
<b>Alien</b>	18	973.833	1779.903	419.527	88.708	1858.958
<b>Ape</b>	47	95.250	215.743	31.469	31.906	158.595
<b>Female</b>	11362	15.172	19.358	0.182	14.816	15.528
<b>Male</b>	23232	12.799	21.116	0.138	12.527	13.070
<b>Zombie</b>	196	89.621	137.686	9.835	70.225	109.017

<b>Población</b>	34855.0	14.612	51.478	0.276	14.072	15.152
------------------	---------	--------	--------	-------	--------	--------

Sacamos el F-Statistic:

```
fstats, pvalue = stats.f_oneway(df['amount_eth'][df['type_x'] == 'Male'],
                                df['amount_eth'][df['type_x'] == 'Female'],
                                df['amount_eth'][df['type_x'] == 'Ape'],
                                df['amount_eth'][df['type_x'] == 'Zombie'],
                                df['amount_eth'][df['type_x'] == 'Alien'])

print('f-statistics value = ', fstats.round(3), ' p-value = ', pvalue.round())
```

executed in 37ms, finished 11:11:42 2021-07-10

f-statistics value = 2116.28 p-value = 0.0

Nuestro F-Statistic de 2116.28 sugiere que la varianza entre grupos es 2116 veces la varianza dentro de un mismo grupo.

Ahora sabemos que la relación de nuestras varianzas no es igual a uno, y RECHAZAMOS la Hipótesis Nula de que las medias de los grupos sean iguales debido a que el valor p-valor es menor que nuestro nivel de significación ( $p < 0,05$ ).

Hagamos un repaso de la F-Distribution, ya que la estamos utilizando para evaluar nuestro F-Statistic. Evaluamos el F-Statistic con un intervalo de confianza del 95%, y para determinar la forma de nuestra distribución necesitamos los *dfn* (numerador de los grados de libertad) y *dfd* (denominador de los grados de libertad).

Donde *k* es el número de grupos de comparación - en nuestro caso cinco - y *N* es el número total de observaciones en el análisis, que es de 34850.

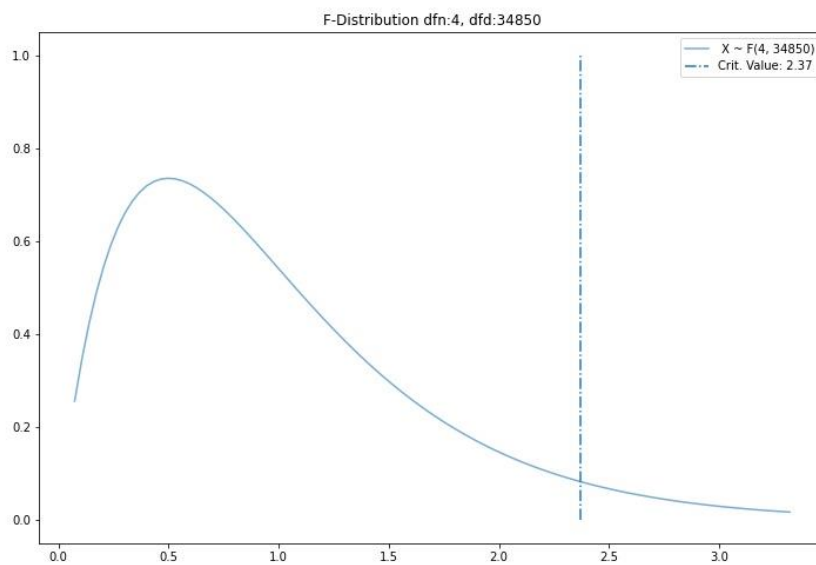


Fig.22 F-distribution para evaluar nuestro F-statistic con un 95% de intervalo de confianza.

El análisis de la varianza (ANOVA) asignará la variación total de cada variable independiente, y probará la significación de cada una frente a las variables dependientes. Y nos iluminará parte de la abstracción que el one way F-Test no nos permite ver.

Para ello, ejecutamos una regresión 'ols' regular con 'amount\_eth' como la variable dependiente, y los diferentes tipos de 'type\_x' como las variables independientes categóricas.

<b>Dep. Variable:</b>	amount_eth	<b>R-squared:</b>	0.195
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.195
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	2116.
<b>Date:</b>	Sat, 10 Jul 2021	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	12:57:21	<b>Log-Likelihood:</b>	-1.8304e+05
<b>No. Observations:</b>	34855	<b>AIC:</b>	3.661e+05
<b>Df Residuals:</b>	34850	<b>BIC:</b>	3.661e+05
<b>Df Model:</b>	4		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
<b>Intercept</b>	973.8333	10.884	89.474	0.000	952.500	995.166
<b>C(type_x)[T.Ape]</b>	-878.5829	12.800	-68.641	0.000	-903.671	-853.495
<b>C(type_x)[T.Female]</b>	-958.6611	10.893	-88.010	0.000	-980.011	-937.311
<b>C(type_x)[T.Male]</b>	-961.0346	10.888	-88.264	0.000	-982.376	-939.693
<b>C(type_x)[T.Zombie]</b>	-884.2125	11.373	-77.748	0.000	-906.504	-861.921

<b>Omnibus:</b>	98683.925	<b>Durbin-Watson:</b>	1.063
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	11548241977.681
<b>Skew:</b>	37.152	<b>Prob(JB):</b>	0.00
<b>Kurtosis:</b>	2821.903	<b>Cond. No.</b>	125.

Obtuvimos el mismo F-statistic que antes - seguimos rechazando la hipótesis nula, así que nada ha cambiado - sin embargo, somos capaces de desglosarlo más, ya que tenemos más información sobre cómo se calculó nuestro resultado.

## 7. ¿El punk más atractivo? En qué punks invertir y conclusiones del informe.

Tras la exploración, selección de variables, agrupación y modelaje concluimos:

1. En general el precio de los punks depende principalmente de su tipo y secundariamente de sus atributos. Hemos reunido evidencias suficientes para refutar la hipótesis nula según la cual, la media de precios entre tipos era igual. Con la exploración y gracias a la rarity score, hemos visto cómo la rareza de los atributos puede hacer aumentar en gran medida el precio de un punk - de forma más notoria con un tipo común (Male o Female) - pero siempre bajo la influencia del tipo.
2. Una cosa es el precio y la otra dónde invertir. Cuando queremos invertir entran en juego muchos más factores, como por ejemplo la cartera del inversor y cuánto esté dispuesto a arriesgar. Para buscar los punks 'más atractivos' en el sentido económico, hemos creado una variable que calculara el rendimiento asociado a cada venta (precio de venta/precio de compra), para así observar con qué punk y en qué momento se habían sacado más rendimiento. El rendimiento indica los beneficios proporcionales a la compra, no el beneficio neto, por este motivo creemos que son muy importantes las características del inversor para llegar a determinar un punk o un grupo de punks en los que será más seguro obtener un rendimiento o un beneficio positivos.
3. Creemos firmemente que el rendimiento va asociado al momento de compra y venta, que a su vez está relacionado con el número de veces que ha sido vendido el mismo punk. El mayor rendimiento medio lo hemos encontrado en la segunda venta (aunque de la primera no podemos tener este dato porque se consiguieron gratuitamente) y desciende gradualmente a medida que aumentan sus ventas. Eso sí, un menor rendimiento no indica un menor beneficio, eso dependerá de las cantidades monetarias específicas de cada caso.

Aún así, trataremos de dar los datos de aquellos punks que parecen más rentables, considerando que la cifra actual (en Ethereums o en dólares) no es problema para el inversor. Gracias a la clusterización hemos observado un grupo de punks, el Tercer Cluster(2), con un rendimiento medio muy superior a los otros dos grupos y formado mayoritariamente por punks de tipos comunes (Male y Female). Creemos que el hecho de que sean de tipo común es una ventaja pues estos no suelen tener precios tan elevados y el riesgo, por lo tanto, es menor, pues se puede 'diversificar' la cartera de punks, comprando un grupo en vez de uno muy caro (como los Aliens). En el apartado de clusterización hemos enumerado punks en concreto con alto rendimiento.

4. Aunque el rendimiento es un buen indicador es importante la manera como lo usamos. Un punk que ha tenido un buen rendimiento no implica que tenga un buen rendimiento en el futuro (se debería estudiar a fondo). Puede que para obtener mayores beneficios se deban estudiar aquellos que no parezcan interesantes, ni 'atractivos'. Punks que hayan tenido rendimientos inferiores para comprarlos en el momento en el que no tienen un precio elevado o aún no cuentan con transacciones, (como es el caso del punk #200). Para esta estrategia se debería de buscar hasta dónde puede llegar el interés bursátil teniendo en cuenta las limitaciones físicas o la poca popularidad del punk.

Nos gustaría añadir unas conclusiones laterales al trabajo presentado. Algunas hacen referencia a caminos que se podrían haber seguido con más tiempo o otras son reflexiones y aspectos que creemos que pueden llegar a contribuir en el estudio.

#### *Enmarcar el problema*

1. Creemos que el estudio de mercado de los punks con intención de invertir requeriría también un estudio paralelo de otros mercados de NFT's y hasta de otras cryptomonedas para ver como afectan a las proyecciones y tener más información de este tipo de mercados y su comportamiento.
2. Siguiendo el estudio de los NFT's en general, reunir más información para apoyar o refutar la teoría de que el mercado es una burbuja económica.
3. Creemos que el beneficio reportado por los activos bursátiles como los NFT's dependen, en gran medida, del momento de su compra, un componente que puede ser tan importante como qué punk o grupo de punk elegir para invertir. Para ello se debería tener en cuenta el precio del Ethereum y su varianza, además de la consideración pública del valor de los NFT's.

#### *Otras rutas*

4. Creemos que se debería profundizar en el estudio de las veces que ha sido vendido cada punk en relación a su rendimiento económico.
5. Se deberían de explotar las diferentes agrupaciones posibles de los punks, ya sean supervisadas o no-supervisadas, para probar modelos de clasificación en caso de que sean más efectivos que los de regresión lineal.
6. Creemos que un estudio de los usuarios con más actividad, su cartera de punks e historial de compra-venta, podría ser beneficiario para entender su comportamiento y cómo han generado sus beneficios.

#### *Profundizar*

7. Nos ha faltado mejorar el modelo de regresión lineal y sus predicciones por falta de tiempo. Creemos que otro tipo de feature engineering como el one hot encoding de todos los atributos podría mejorar su R2 score.
8. Mejorar el diseño de la hipótesis de manera que la distribución de la población sea más gaussiana y que las muestras para el ANOVA se ajusten más a los requisitos previos de este test.
9. Profundizar en el análisis temporal en todas las variables, principalmente en la evolución del rendimiento en función del número de venta. Creemos que un estudio de la retención temporal de los punks y su incremento de precio entre venta y venta podría traer conclusiones significativas para saber cuando comprar (o ofertar) y vender.



# Anexo

## A1. Obtención de datos

### Fuentes (URLS) y métodos de scraping

- <https://www.larvalabs.com/cryptopunks/details/> + id Cripto punk (Selenium).
- <https://www.larvalabs.com/cryptopunks/topsales?sortByUSD=false> (Selenium)
- <https://www.larvalabs.com/cryptopunks/attributes> (Selenium)
- <https://defypunk.com/punks/> (Selenium)
- <https://www.larvalabs.com/cryptopunks/attributes> (Selenium)
- Historial de transacciones de cada punk: <https://defypunk.com/punks/0>.  
Escapeado y facilitado por los compañeros Julio Abril y Francisco Gil.

## A2. Diseño base de datos relacional

La construcción de la base de datos relacional está pensada para evitar la redundancia de datos, y reunir unos criterios de normalización para todos los datos usados en el proyecto, así como mejorar la eficacia de su acceso.

### Descripción Tablas

- 1.1. **master\_punks2** : Esta tabla es una tabla que nos permite dar robustez a la base de datos, en ella se encuentran registrados el id de todos los Criptopunks.
- 1.2. **topsales**: Esta tabla posee una lista de los Criptopunks con importe mayor de venta.
- 1.3. **types**: En esta tabla se encuentran registrados los tipos de Criptopunks que podemos encontrar.

- 1.4. **attributes:** En esta tabla están registrados los diferentes accesorios que podemos añadir a un Criptopunk para personalizarlo.
- 1.5. **rarity:** En esta tabla está registrado un dato que ofrece la web, que a través de un coeficiente nos permite valorar la rareza del criptopunk. La rareza puede influir en el valor del criptopunk.
- 1.6. **punks\_attributes:** En esta Tabla están registrados los atributos que tiene cada Criptopunk.
- 1.7. **Historico\_trans:** En esta base de datos están registrados los movimientos financieros que ha habido en cada criptopunks a lo largo del tiempo, por otro podemos encontrar el valor por el que fueron puestos a la venta así como su valor de compra final.

## Atributos

- 1.8. **Tabla “master\_punks2” :**
- 1.9.
  - 1.9.1. **id:** Tipo de dato INT y se ha definido como Primary Key.
  - 1.9.2. **Idpunks** tipo de dato INT. Este campo es un Key index para poder crear Foreign Keys.
- 1.10. **Tabla “topsals”:**
  - 1.10.1. **id:** Tipo de dato INT y se ha definido como Primary Key
  - 1.10.2. **Ranking:** Tipo de dato INT.
  - 1.10.3. **Valor\_ETH:** Dato de tipo FLOAT.
- 1.11. **Tabla “Types”:**
  - 1.11.1. **id:** Tipo de dato INT y se ha definido como Primary Key.
  - 1.11.2. **Attributes:** Tipo de dato VARCHAR(45).
  - 1.11.3. **num\_punks:** Tipo de dato INT.
- 1.12. **Tabla “Skin”:**
  - 1.12.1. **id:** Tipo de dato INT y se ha definido como Primary Key
  - 1.12.2. **skin :** Tipo de dato VARCHAR(45).
  - 1.12.3. **num\_p:** Tipo de dato INT.
- 1.13. **Tabla “Attributes”:**
  - 1.13.1. **id:** Tipo de dato INT y se ha definido como Primary Key
  - 1.13.2. **attributes :** Tipo de dato VARCHAR(45).
  - 1.13.3. **num\_punks:** Tipo de dato INT.

1.14. **Tabla “rarity”:**

- 1.14.1. **id:** Tipo de dato INT y se ha definido como Primary Key
- 1.14.2. **rarity :** Tipo de dato FLOAT

1.15. **Tabla “punks\_attributes”:**

- 1.15.1. **id:** Tipo de dato INT y se ha definido como Primary Key
- 1.15.2. **idpunks:** Tipo de dato INT.
- 1.15.3. **typecryptopunk:** Tipo de dato VARCHAR(45).
- 1.15.4. **counter:** Tipo de dato INT.
- 1.15.5. **attributes:** Tipo de dato VARCHAR(45).
- 1.15.6. **idattributes :** Tipo de dato INT.

1.16. **Tabla “Historico\_trans”:**

- 1.16.1. **id:** Tipo de dato INT y se ha definido como Primary Key
- 1.16.2. **transaccion\_date :** Tipo de dato “Date”
- 1.16.3. **cryptopunk\_number:** Tipo de dato INT
- 1.16.4. **trs\_type:** Tipo de dato VARCHAR(45).
- 1.16.5. **de:** Tipo de dato VARCHAR(45).
- 1.16.6. **para:** Tipo de dato VARCHAR(45).
- 1.16.7. **amount\_dolars:** Tipo de dato FLOAT.
- 1.16.8. **amount\_eth:** Tipo de dato FLOAT
- 1.16.9. **user\_transaction:** Tipo de dato VARCHAR(45).
- 1.16.10. **adj\_close:** Tipo de dato FLOAT

## Esquema Relacional

