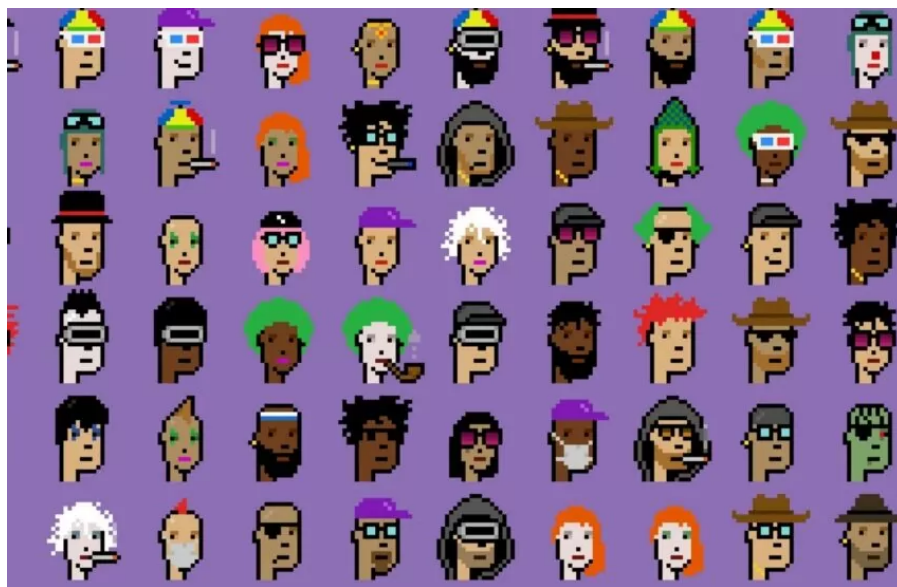


Crypto Punks:

Análisis de Ventas y patrones de precio



Resumen

Los crypto punks son NFT (Non Fungible Tokens) un estilo de arte digital que fueron puestos en circulación en el 2017 y han llegado a precios de venta de 7.5 millones de dólares. Cada uno de ellos cuenta con características propias distintas que lo diferencian de los otros. Estudiando estas características mediante análisis descriptivo y aplicando las técnicas de Machine Learning de Random Forest Classification y Support Vector Machines se estudian posibles relaciones entre características y precios que expliquen el valor asociado a los Cripto Punks.

Al hacer el análisis descriptivo se aprecia una clara relación entre un tipo concreto de Crypto Punks muy poco frecuente menos del 2% del total y un precio elevado pero no se ve una relación que indique las diferencias de precio entre el grueso de los otros crypto punks. Tras analizar las variables para identificar outliers y hacer una limpieza de datos se utilizan los algoritmos de Machine Learning para detectar patrones que expliquen el precio asociado.

Los algoritmos no són capaces de identificar las características que determinan el precio del crypto punk y hacen distribuciones aleatorias en función de la frecuencia de clase.

No se ha podido determinar que haya características específicas que dicten el valor de mercado de los Crypto Punks a parte del grupo minoritario.

Abstract

Crypto Punks are a type of NFT (Non Fungible Tokens), a new type of digital art, that were first put on the market in 2017 and have reached prices of over 7.5 million dollars. There are 10.000 Crypto Punks each one of them with a unique set of characteristics. In the present study we aim to find whether there is a relation between these unique characteristics and their sales price. We use exploratory analysis and machine learning techniques to study the possible relations between the set of characteristics and the sale price.

On a first analysis we can see a clear relation between a very low frequency type of Crypto Punks, less than 2% of the total amount, and a really high price but fail to see a relation between the big set of Crypto Punks and the price. After analyzing variables to search for outliers and clean the data, we apply different Machine Learning techniques to identify patterns that explain the difference in price of the Crypto Punks.

The algorithms are unable to extract any patterns that could explain the different price points and end up doing random distributions mostly based on frequency of classes.

We couldn't find any specific characteristics that dictate the market value of the Crypto Punks aside from the minority group.

ÍNDICE

Introducción	3
2.Extracción y organización del dato	4
2.1 Extracción	4
2.2 Organización	5
3.Análisis descriptivo	6
3.1 Análisis univariable	6
3.1.1 Type	6
3.1.2 Skin	6
3.1.3 Traits	7
3.1.4 Hidden Traits	8
3.1.5 Number of Traits	9
3.1.6 Transaction Type	10
3.1.7 Amount	12
3.1.8 Transaction date	13
3.2 Análisis multivariable	13
3.2.1 Type transaction y amount	13
3.2.2 Amount y transaction date:	14
3.2.3 Precio medio por id	15
3.2.4 Skins y precio	18
3.2.5 Types y precio	19
3.2.6 Relación entre tipo de transacción	20
3.2.7 Análisis de Outliers y precios	22
4.Machine Learning	26
4.1 Random Forest Classification	26
4.2 XGBoost Gradient	30
5.Conclusiones	32
6. Referències:	33

1. Introducción

El presente trabajo es el proyecto final del curso de DataScience impartido por el Cibernàrium. El objetivo de este trabajo es el de realizar un estudio sobre el mercado de Crypto Punks aplicando todos los conocimientos adquiridos durante el curso.

Crypto Punks fue lanzado por primera vez en junio de 2017 como el primer NFT(non-fungible token) dentro de la blockchain de ethereum. Consisten en 10.000 caracteres con un conjunto de atributos que los hace únicos.

El proyecto desarrollado por [Larva Labs](#) sirvió como precedente para lo que hoy en día se conoce como el mercado NFT y el cryptoArt. Un mercado en plena expansión que consiste en la compra y venta de arte digital dentro del ecosistema descentralizado de criptomonedas.

La acción de compra-venta se realiza mediante contratos inteligentes. Estos contratos emulan los contratos tradicionales exceptuando que a diferencia de los contratos tradicionales, en los contratos inteligentes no hay un agente o institución externa que valide dicho contrato si no que el acuerdo entre el comprador y vendedor queda inscrito en la cadena de bloques, lo único que sí es validado por un tercero es la transacción puesto que para efectuar-la los mineros deben validar-la.

Las crypto monedas junto con los NFT son un mercado relativamente nuevo. Bitcoin fue creada en 2009 y ethereum en 2015 apenas hace 6 años. Durante los últimos años hay un creciente interés en estos mercados si se analiza con [google trends](#) las tendencias de los últimos 5 años se nos encontramos con un interés estable con algunos picos hasta 2020 donde el interés se dispara de forma casi exponencial. Las búsquedas relacionadas con los Crypto Punks siguen un patrón de interés muy parecido a una escala más pequeña. También se ve reflejado este interés por el aumento del coste de las Crypto Monedas. Por ejemplo ethereum tiene una [capitalización de mercado actual](#) de casi 500 mil millones de dólares y analizando la gráfica de precio nos encontramos con un patrón parecido con un nivel de crecimiento muy grande desde mediados de 2020.

Esto deja en evidencia la importancia e impacto que está teniendo este mercado actualmente y sus [posibilidades de desarrollo](#).

Los Crypto Punks como parte de este mercado también han aumentado en exposición al público y capacidad de mover capital. Se han llegado a vender Crypto Punks por valor de 7.5 millones de dólares y [estrellas](#) como Jay-z, Snoop Dogg o Serena Williams han comprado Crypto Punks.

1.1 Objetivos y estructura.

El objetivo del siguiente estudio consiste en encontrar mediante análisis estadístico y machine learning posibles patrones en los datos que nos ayuden a entender que condiciones o atributos tienen los diferentes Crypto Punks que puedan explicar su valor en el mercado así como su incidencia en el precio para finalmente implementar un método que de posible oportunidad de inversión.

Para realizar dicho estudio, necesitamos llevar a cabo la siguiente serie de fases:

- Realizar una recolección de datos mediante webscrapping
- Construcción de una base de datos a partir de la data recolectada

- Primera exploración de los datos
- Descripción estadística de los datos
- Implementación de modelos de machine learning
- Análisis de resultados y conclusión

1.2 Presentacion de la hipotesis.

H_0 = el precio por criptopunk no es condicionado por sus atributos

$$\forall x \in D, P(x)$$

$$D = \{\text{cryptopunks}\}$$

$$P(x) = k$$

siendo k un precio constante

H_a = el precio por criptopunk es condicionado por sus atributos

$$\exists x \in D, \neg P(x)$$

2.Extracción y organización del dato

2.1 Extracción

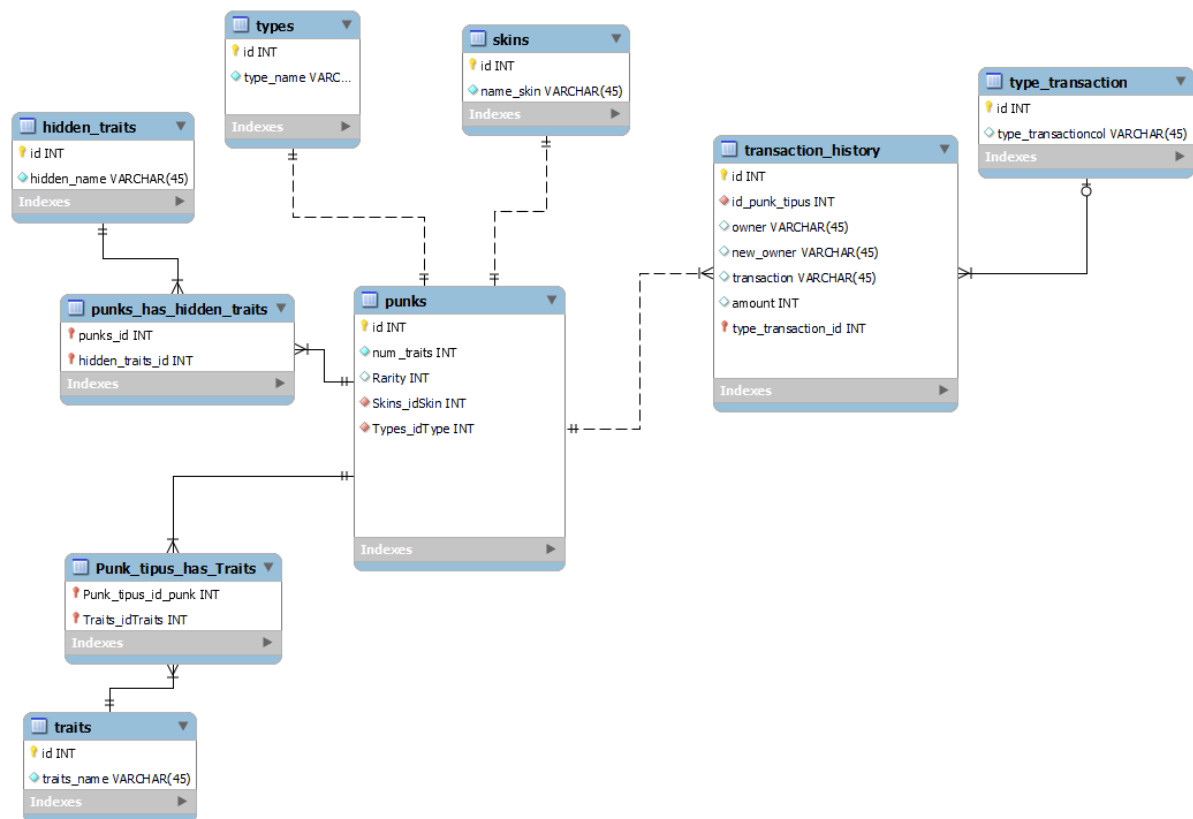
Los datos necesarios para realizar el análisis se extraen a partir de las diferentes secciones de la página web [DEFYPUNK](https://defypunk.com).

Para realizar la extracción, se genera mediante la librería Selenium que ofrece Python un web scrapping que itera por todas las secciones de la web extrayendo los datos candidatos de forma automática.

Más concretamente, mediante una request por crypto punk se accede a la url [https://defypunk.com/punks/\(n\)](https://defypunk.com/punks/(n)) donde n pertenece al conjunto de id's de todos los criptopunks donde por cada Crypto Punk se extrae sus atributos y su historico de transacciones.

2.2 Organización

Una vez se ha recolectado toda la información necesaria, se procede a organizar los datos mediante la creación de una base de datos relacional en MySQL.



3. Análisis descriptivo

3.1 Análisis univariable

Variables propias de los Crypto Punks

Cada crypto punk está definido por un conjunto distinto de características, estas características se agrupan en cuatro conjuntos. A continuación hacemos una descripción de las distintas variables de esta categoría.

3.1.1 Type

Es una variable categórica que hace referencia al tipo de Crypto Punk son Alien, Ape, Zombie, Male o Female.

Hay 10000 Crypto Punks, cada uno con un type. El type con la frecuencia más elevada es Male con un total de 6039, seguido por female con 3840 seguido de los tres siguientes que son outliers en la categoría, con una frecuencia combinada de 121 de los 10.000 totales. Zombies: 88, Ape: 24 y Alien: 9

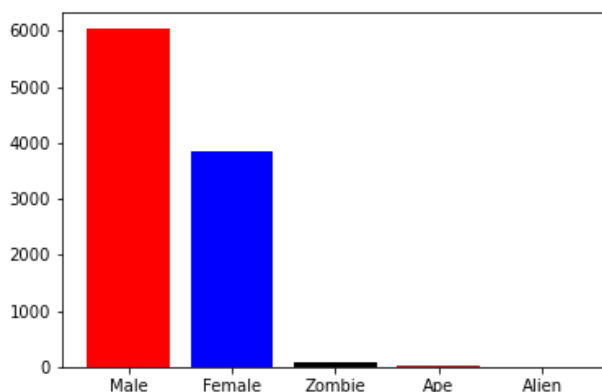


FIG 1. Frecuencia de los types en los Crypto Punks

3.1.2 Skin

Variable categórica que hace referencia al estilo de la piel de cada crypto punk son Alien, Ape, Zombie, Albino, Dark, Light y Mid. Es una variable que tiene el mismo comportamiento que Type, pero toma 7 valores distintos. Las frecuencias son Mid: 3031, Light: 3006, Dark: 2824, Albino: 1018, Zombi: 88, Ape: 24 y Alien: 9. Como en la variable Types, Zombie, Ape y Alien son outliers con una frecuencia acumulada del 121 de los 10.000 Crypto Punks.

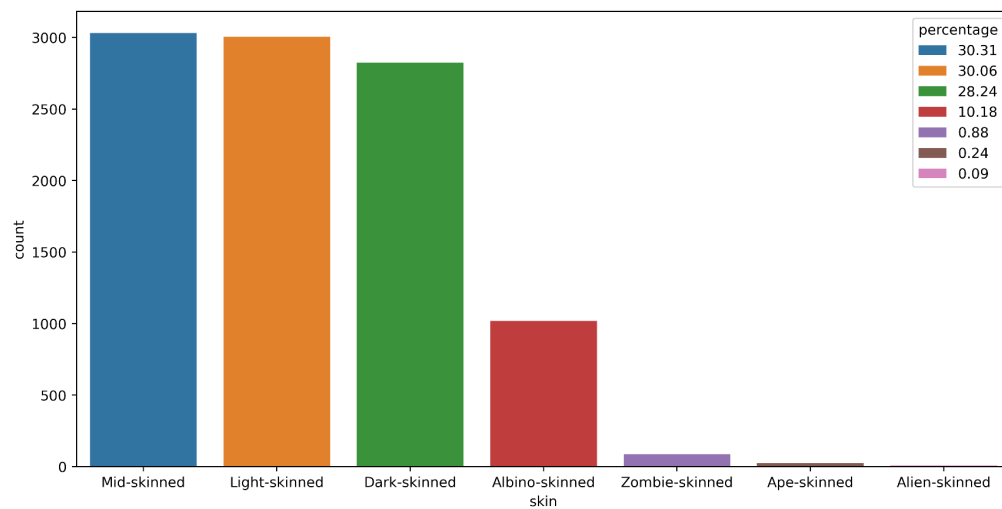


Fig 2. Frecuencia de skins

3.1.3 Traits

Variable categórica que identifica los rasgos de cada Crypto Punk, hay un conjunto muy variado de rasgos distintos. A diferencia de las características anteriores estos no son rasgos únicos, cada Crypto Punk puede tener un conjunto de rasgos distintos o no tener ninguno. A continuación analizamos la frecuencia de aparición de los distintos rasgos.

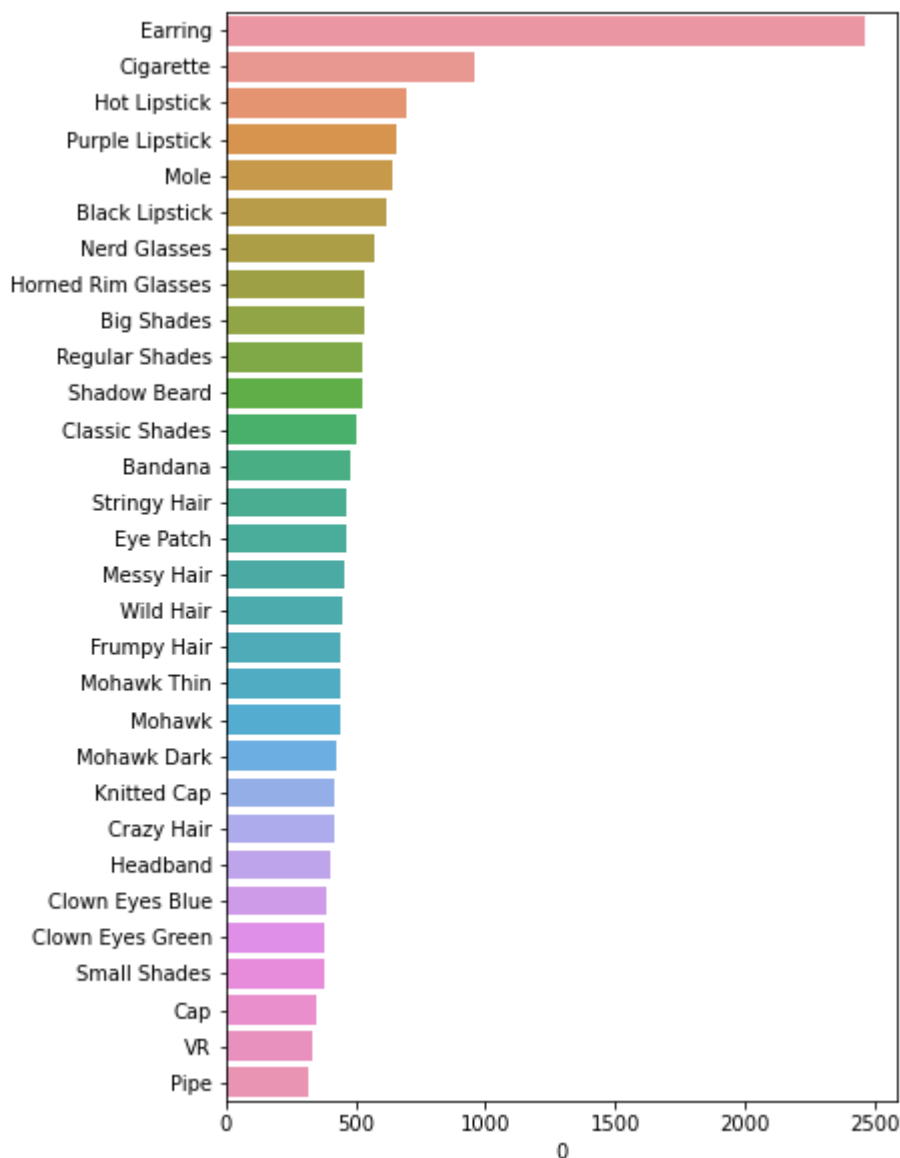


FIG 3. Frecuencia de los primeros 30 rasgos

Hay un total de 87 rasgos distintos con una presencia combinada de 27539, el rasgo con presencia más elevada es earring con 2459 apariciones i el rasgo con menos presencia es Beanie con 44, la media de aparición es de 317 veces por rasgo con una desviación típica de 289.8. El dato que más destaca y se considera como outlier es el rasgo earring, ya que su presencia de aparición es 7 veces superior a la normal.

3.1.4 Hidden Traits

Variable categórica que indica otro estilo de rasgos que puede tener un Crypto Punk, son rasgos más raros que los Traits, y hay un menor número de ellos. Como en los Traits un Crypto Punk puede tener diferentes Hidden Traits o no tener ninguno. A continuación analizamos la frecuencia de aparición de los distintos Hidden Traits.

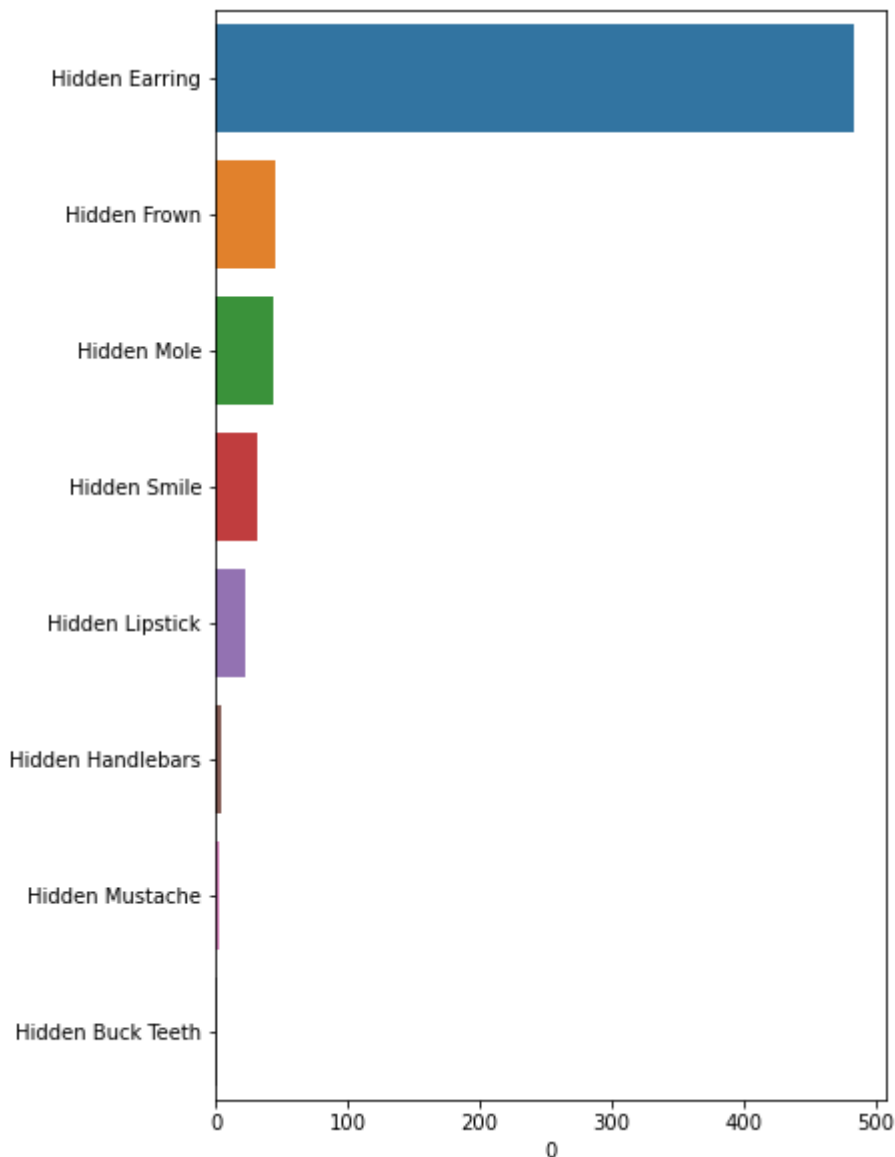


Fig 4. Frecuencia Hidden traits

Existen 8 rasgos escondidos distintos, con una presencia combinada de 636, el rasgo escondido más común es Hidden Earring con un total de 483 apariciones. El rasgo menos frecuente es Hidden Buck Teeth con 2 apariciones

La presencia media de los rasgos escondidos es de 79.5 con una desviación típica de 163.98. Podemos destacar que el rasgo Hidden Earring es un outlier en la categoría debido a que tiene una presencia más de 10 veces superior al siguiente rasgo escondido más común.

3.1.5 Number of Traits

Variable numérica que indica el número total de rasgos por cada Crypto Punk, es una estilo de característica distinta de las anteriores, que se usa para analizar la relación entre número de Traits (hidden y normales) con la otras variables.

La media de traits por Crypto Punk es de 2.81, con una desviación típica de 0.91.

Aquí podemos ver cómo se clasifican los Crypto Punks en función del número de rasgos que tengan.

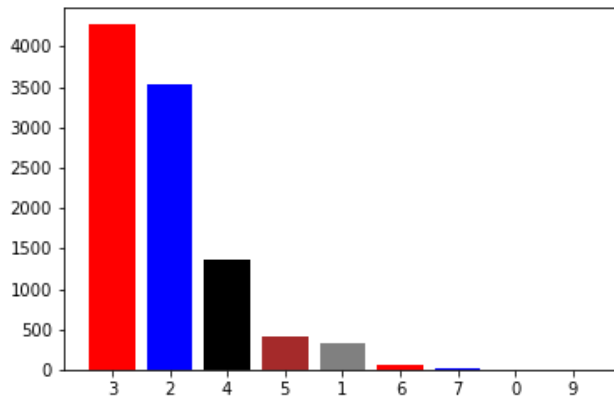


FIG 5. Frecuencia de Number of traits

Son outliers los valores del 5 al 9, que tienen un porcentaje de frecuencia acumulado del 8,14%

Variables transacciones

Cada Crypto Punk tiene asociado un tabla de transacciones que detalla la información de cada operación que se ha realizado con ese Crypto punk. A continuación hay una descripción de las distintas variables que contienen las transacciones.

3.1.6 Transaction Type

Variable categórica que indica el Tipo de operación realizada asociada al Crypto punk, esta puede ser:

- Offered: El propietario del punk pone el punk a la venta con un precio fijo para la compra o cambia el precio de venta.
- Bid: Un usuario pone una puja en un Crypto Punk concreto para ver si el propietario acepta la oferta y se lleva a cabo la compra del crypto punk.
- Offer withdrawn: El propietario del Crypto Punk retira la oferta de venta del Crypto Punk.
- Sold: El propietario acepta una oferta de puja y vende el Crypto Punk
- Bid withdrawn: El usuario que tenía una puja activa sobre un Punk concreto retira la oferta.
- Transfer: Crypto Punk transferido entre owners sin un pago asociado a la transferencia
- Claimed: Esta es la primera operación de todos los Crypto Punks. Se realizó cuando salieron por primera vez los Crypto Punks y los distintos usuarios de la plataforma

los reclamaban sin coste alguno y pasaban a ser los primeros propietarios de los Punks. Esta operación no da ninguna información relevante respecto al Crypto Punk.

- Transfer (Wrap) y (Unwrap): Transferencia de llave segura para acceder al Punk y Recibimiento de llave de transferencia y apertura. El resultado de estas operaciones es el mismo que una transferencia normal y se considerara la secuencia de operaciones wrap y unwrap como una transferencia normal.

Aquí vemos la frecuencia de distribución de los distintos tipos de transacciones.

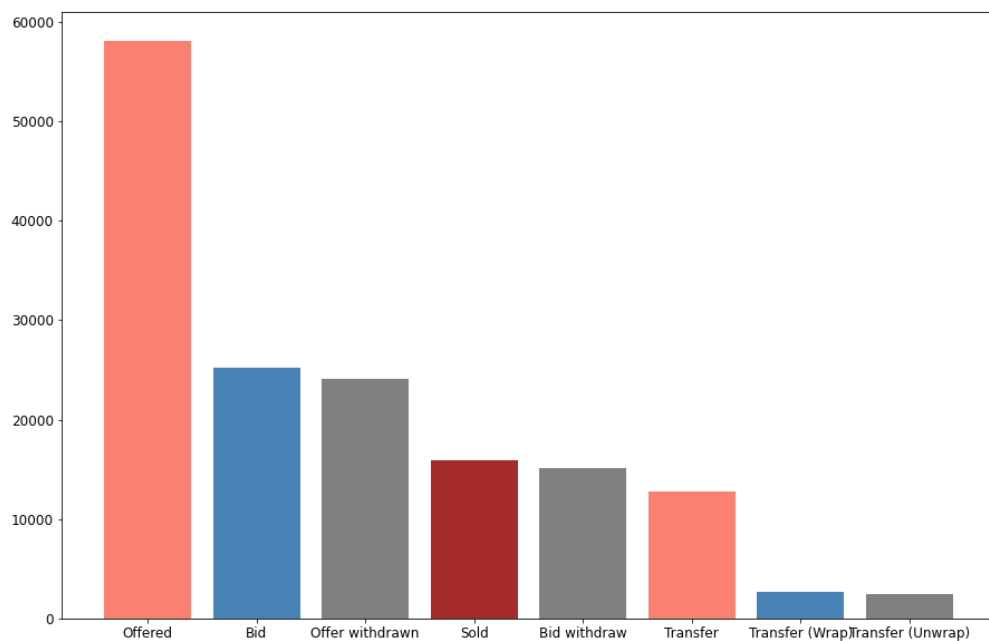


FIG 6. Frecuencia de distribución por estilo de transacción

3.1.7 Amount

Variable numérica que indica la cantidad de ethereum asociada a la transferencia. Se realiza el análisis tanto con ethereum como con la conversión a dólares.

	amount_eth	Amount_dollars
count	113797.00	1.137970e+05
mean	109.60	1.942798e+05
std	6669.76	9.168200e+06
min	0.00	0.000000e+00
25%	1.70	5.263000e+02
50%	19.99	3.768294e+04
75%	40.00	9.080170e+04
max	1000000.00	1.777278e+09

Fig 7. tabla datos estadísticos sobre la cantidad de ethereum y dollars

Podemos observar en el índice max, que indica el valor máximo registrado en una transacción, un indicio de presencia de outliers dado que la cantidad es desorbitada

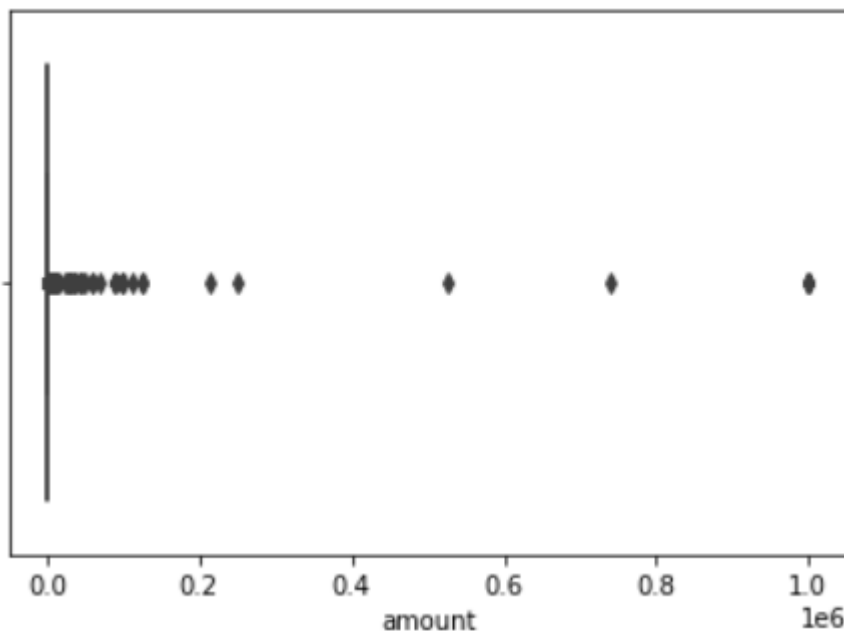


Fig 8. distribución sobre amount

En el gráfico anterior apreciamos que la mayoría de los datos se concentran al lado izquierdo del gráfico dejando ver en outliers en la parte derecha

3.1.8 Transaction date

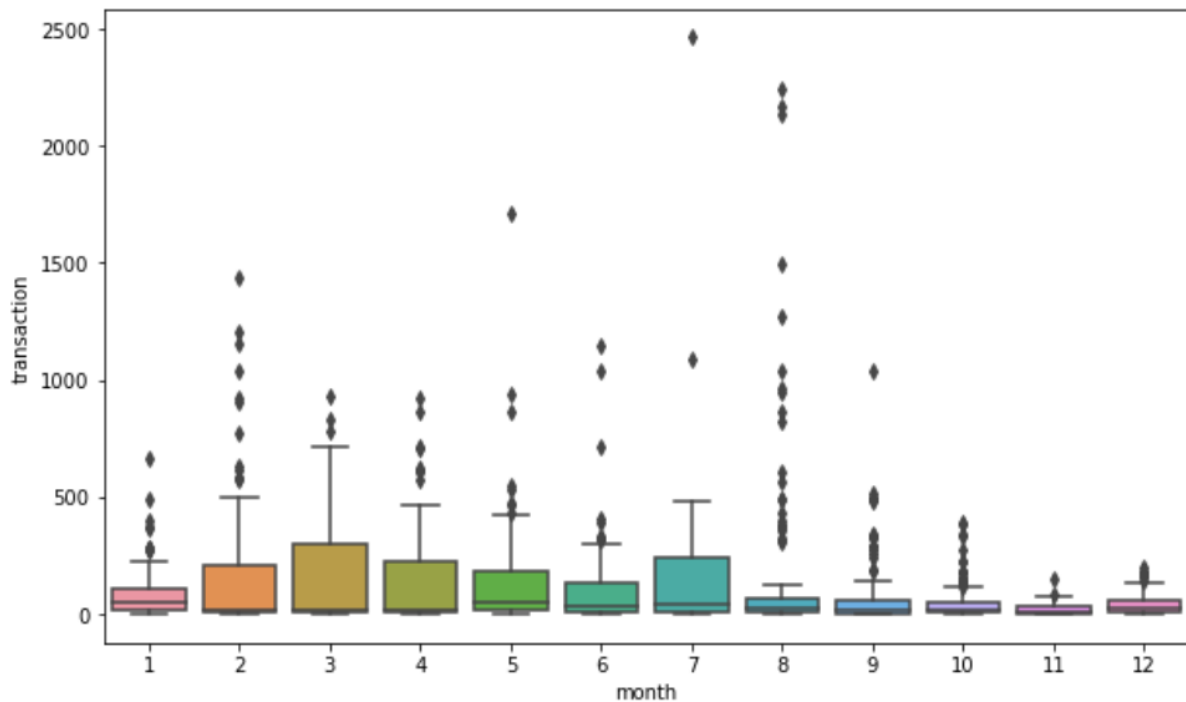


Fig 9. frecuencia de transacciones por meses

3.2 Análisis multivariable

Análisis entre distintas variables para descubrir posibles relaciones.

3.2.1 Type transaction y amount

Análisis de la posible relación entre tipo de transacción realizada y la cantidad de dinero adjunta a la transacción. Para hacer este análisis a priori se han eliminado los outliers más extremos del pago asociado en dólares. Se han identificado los valores a eliminar calculando el rango intercuartil y eliminando todos los valores que estaban 1.5 veces el rango intercuartil por encima del cuartil 3. No había valores que cumplieran esta condición por debajo del cuartil uno.

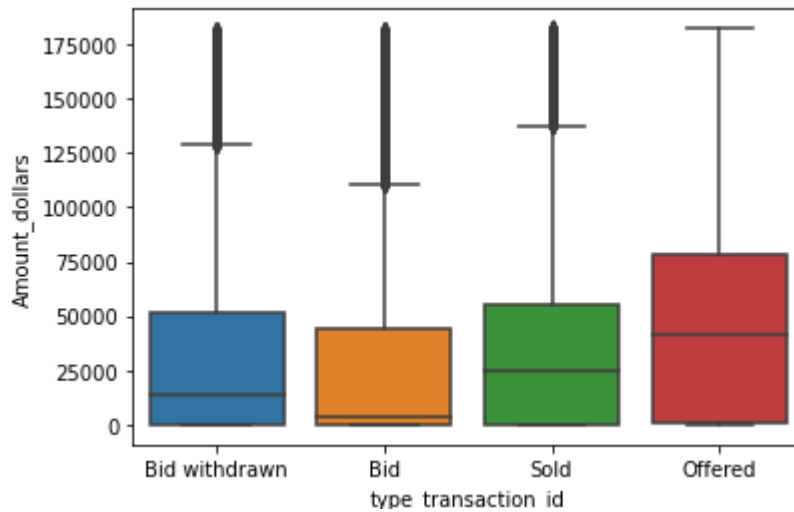


FIG 10. Cantidad Dólares por Tipo de transacción

Se ve una distribución similar entre los distintos tipos de transacciones y la cantidad de dólares asociados a ellas. El tipo de transacción con más dinero asociado es la oferta y la que menos la puja. El valor asociado a la venta del Crypto punk es un valor intermedio entre las ofertas y las pujas. Las pujas retiradas tienen un valor medio ligeramente más elevado que las pujas pudiendo indicar que las pujas tienden a ser retiradas más rápidamente cuando son más elevadas y no dar opción a completar la venta y en cambio a permanecer más tiempo cuando son más bajas terminar completando la venta.

3.2.2 Amount y transaction date:

Análisis de la evolución del precio a lo largo del tiempo.

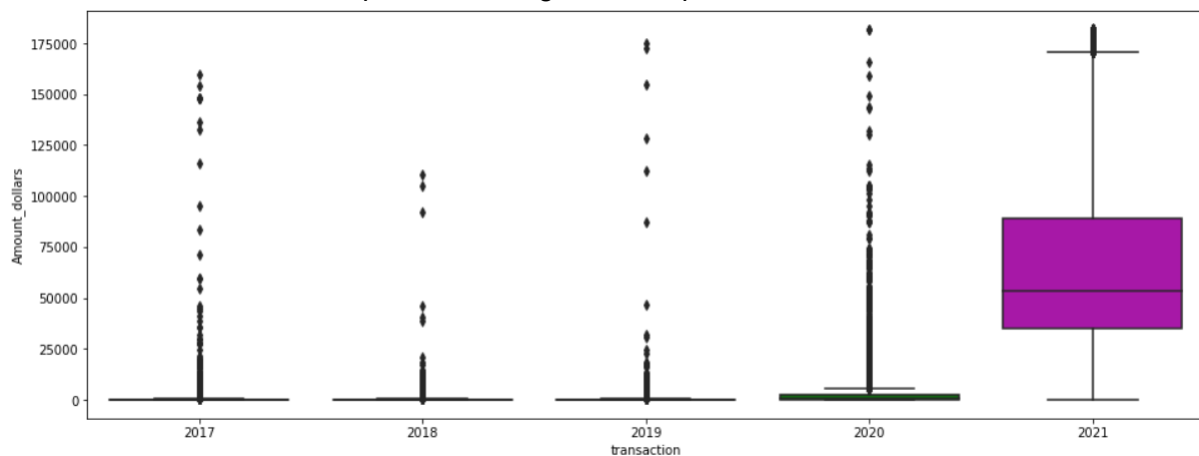


Fig 11. cantidad dólares por año

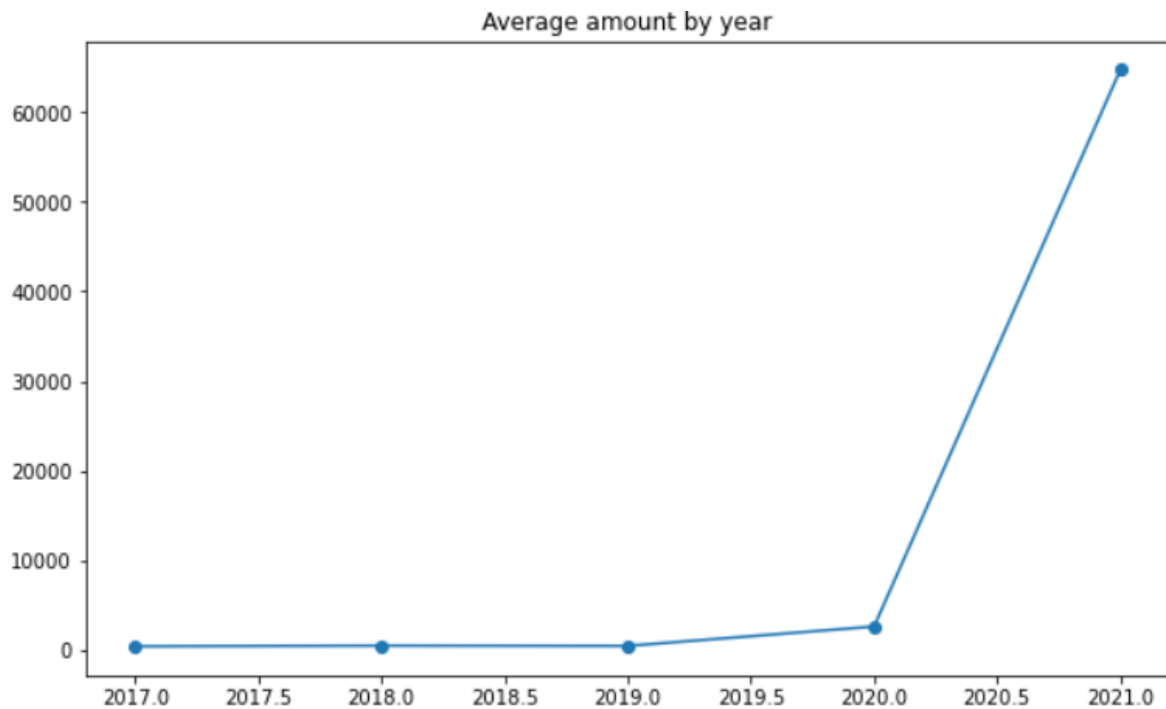


Fig 12. media de dólares por año

Podemos observar tanto como en el gráfico 11 y 12 la evolución del precio ha ido incrementando, se puede deber a dos factores, por un lado un incremento en la demanda y por otro lado el aumento del precio de ethereum respecto al dólar.

3.2.3 Precio medio por id

El valor medio asociado a los distintos tipos de transacciones por todos los Crypto Punks es de 31410 \$ con una desviación típica de 34311. Representamos a continuación los Crypto Punks con un valor medio más elevado y los de valor medio más bajo. Permite ver las grandes diferencias de precio entre los Crypto Punks e identificar Crypto Punks concretos con relevancia respecto a patrones de rasgos que puedan influir al precio.

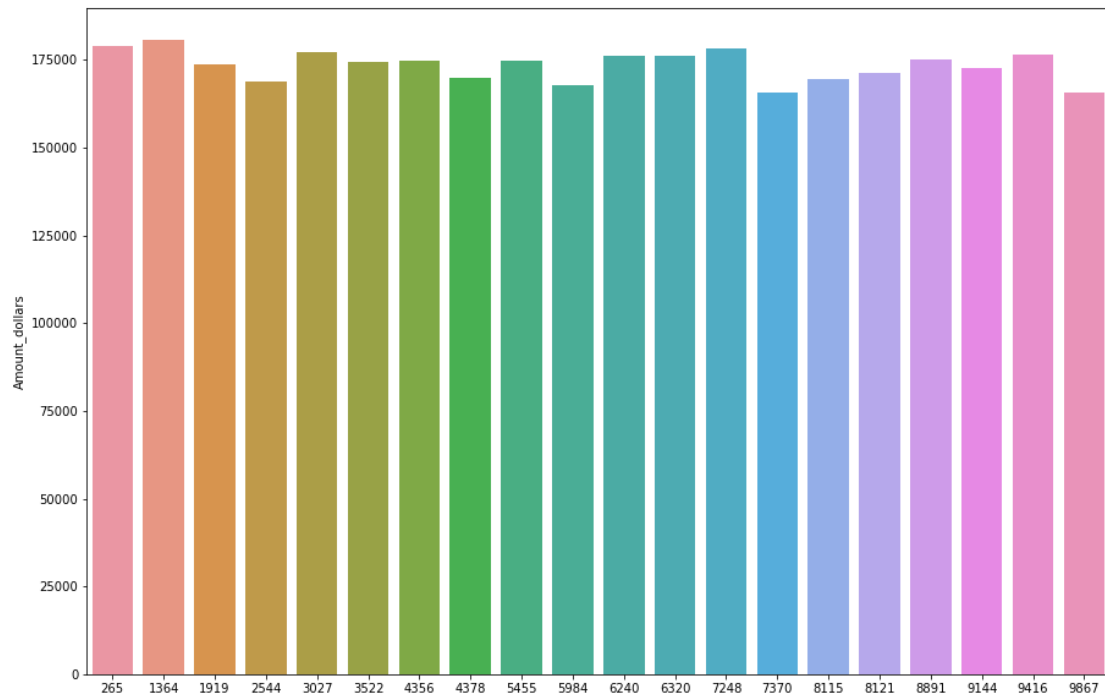


Fig 13, 20 crypto punks más caros

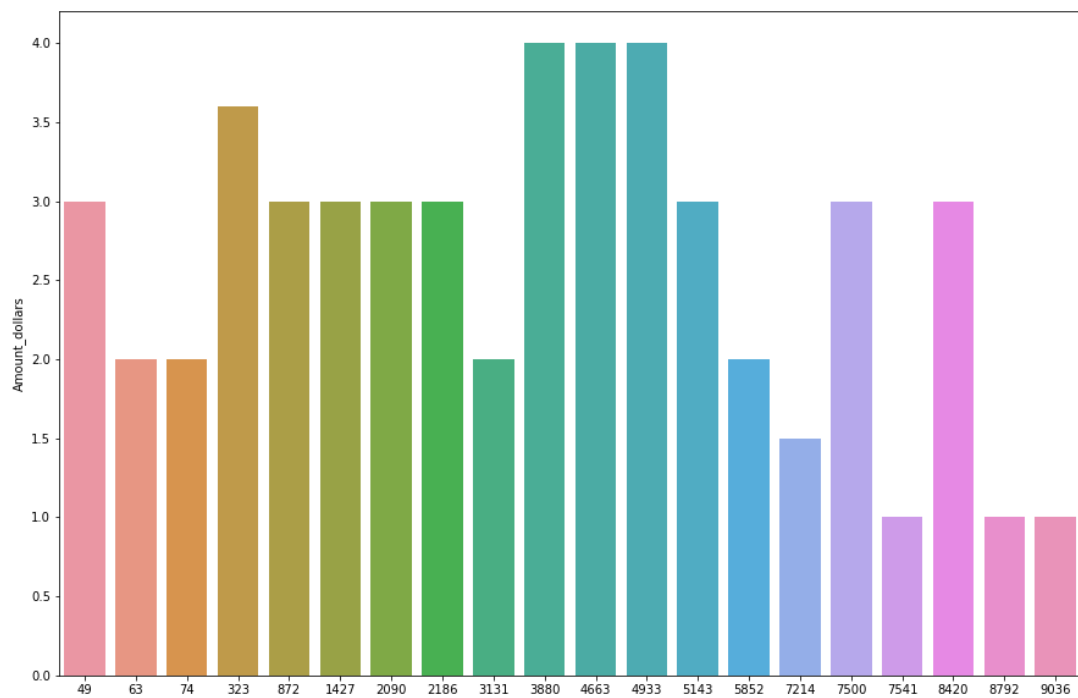


Fig 14. 20 crypto punks más baratos.

3.2.4 Número de rasgos y precio

Analizamos si hay relación entre la cantidad de rasgos que tiene el crypto punk y el precio asociado a las transacciones.

Primer análisis con los crypto punk que tienen entre 2 y 4 rasgos ya que los que tienen más o menos rasgos son considerados outliers. Se analizarán por separado para ver si hay relaciones entre valores extremos de número de rasgos y precio la cual cosa podría indicar un patrón.

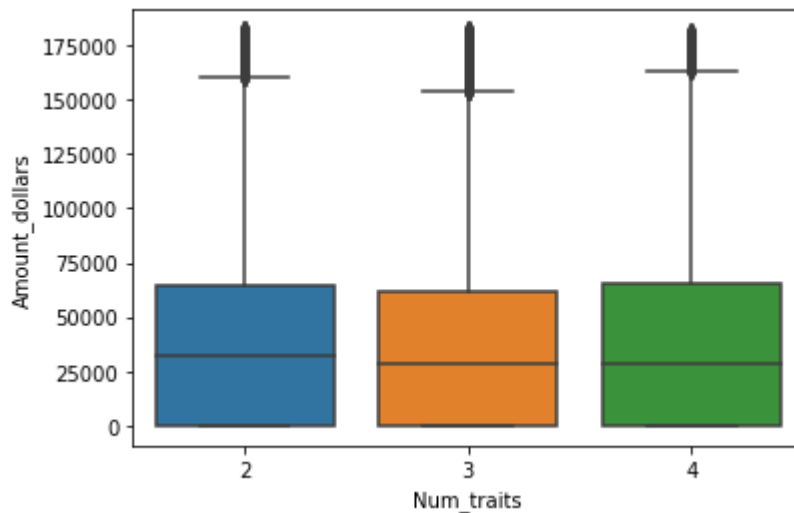


Fig 15. Precio por número de rasgos

Como vemos no hay diferencias significativas entre el precio asociado a las transacciones y el número de rasgos. En todos hay una distribución muy parecida.

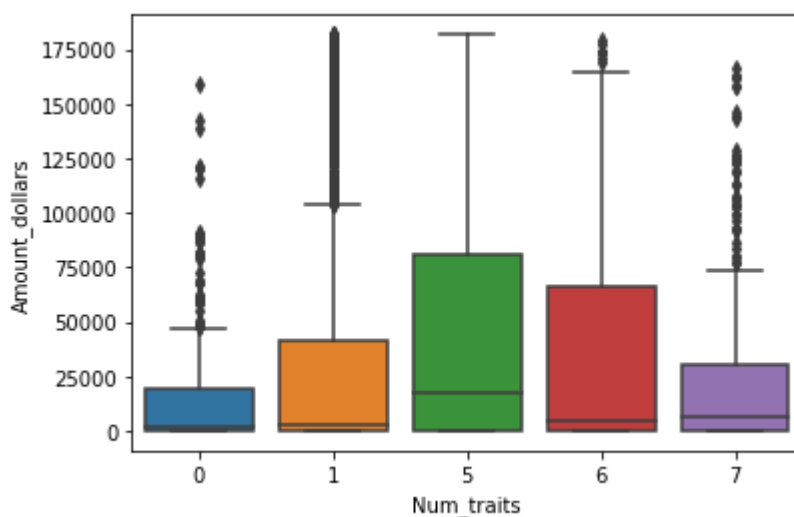
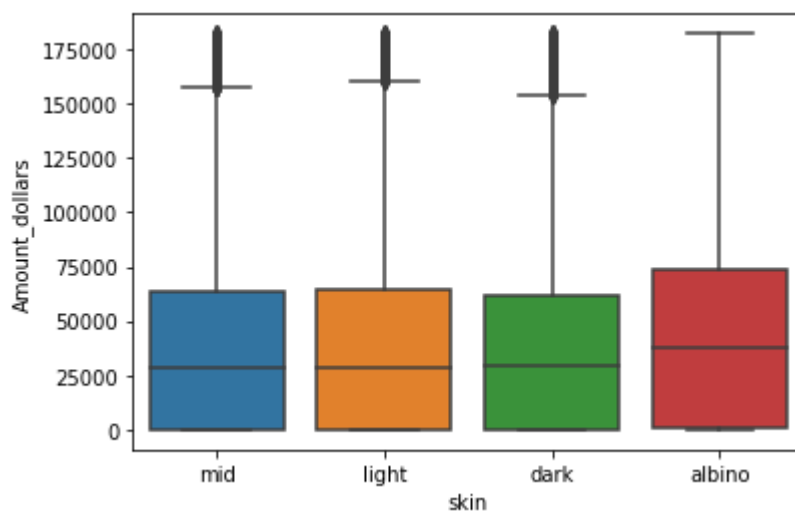


Fig 16. Precio por número de rasgos outliers

En el gráfico con la representación de los outliers en cuanto a número de rasgos vemos distribuciones un poco distintas a las vistas anteriormente pero no se aprecia que haya ninguno de ellos que tenga un precio medio mucho más elevado o alguna otra característica que destaque, esto junto a los pocos datos que hay de estos conjuntos no nos indica ninguna relación especial entre el número de rasgos y el precio.

3.2.4 Skins y precio

Análisis de la relación entre el tipo de skin de los Crypto Punks y el precio.

**Fig 17. Tipo skin i precio**

Las distribuciones entre los distintos tipos de skin y el precio siguen la misma forma. Con valores medios y cuartiles muy similares. La skin albino tiene un valor un poco más elevado que las otras pero no se aleja lo suficiente como para indicar una diferencia significativa.

Se analizan las skins poco comunes (outliers) para ver si hay algún patrón diferente.

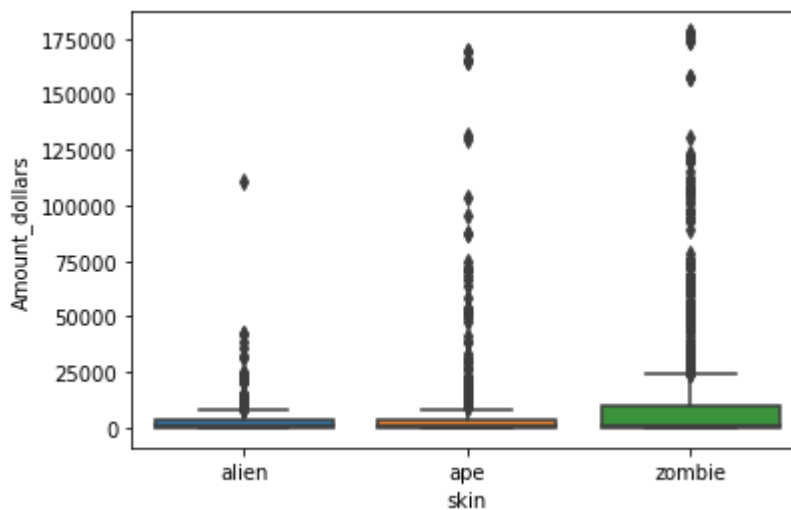


Fig 18. Alien, ape y zombie con precio

Se ve la distribución de precio por las skins outliers. Debido a la falta de datos asociados a este tipo de skins la información no es concluyente. La gráfica indica un valor menor asociado a estas skins que el que se ve con las skins más comunes.

3.2.5 Types y precio

A continuación valoramos la relación entre el precio del crypto punk y su tipo.

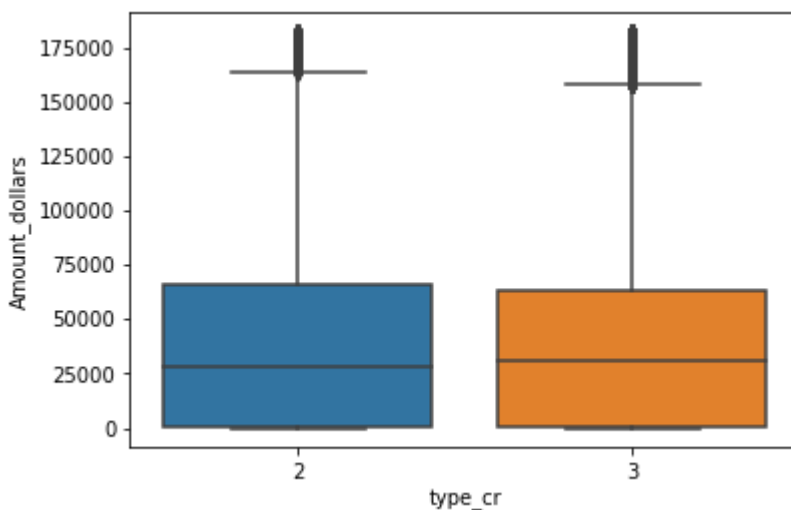


Fig 19. Female (2) y Male(2) con precio

La distribución es prácticamente la misma en los dos valores significativos de los tipos de crypto punks. No se aprecia ninguna diferencia significativa que permita decir que uno de los dos tipos conlleva un precio mayor o menor.

A continuación vemos los 3 tipos de más que hay, que han sido clasificados como outliers para ver si se puede apreciar algún patrón.

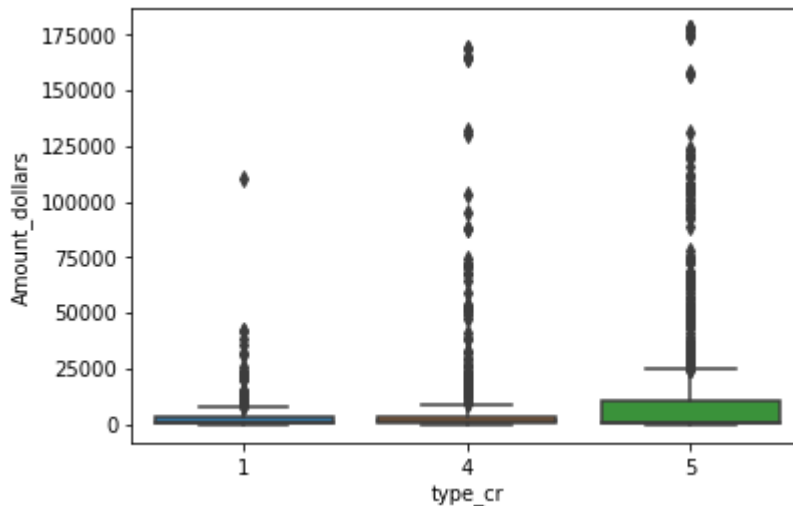


Fig 20. Alien (1), Ape(4), Zombie (5) con precio.

Debido a la falta de datos con estas 3 muestras comparadas con las anteriores la información que se extrae no es tan significativa, se puede apreciar como las distribuciones son similares en cuanto el conjunto más grande de precios se centra por debajo de los 25.000 dólares y luego hay ciertos valores extremos en cada uno. Vendría a indicar más bien una disminución de precios cuando un Crypto tiene uno de estos tipos que no un aumento debido a su rareza.

3.2.6 Relación entre tipo de transacción

Analizamos posibles relaciones entre el tipo de transacción y los diferentes atributos descartando las transacciones catalogadas como outliers.

Transacciones realizadas por número de traits

Observamos que la mayor cantidad de transacciones vendidas se encuentran en el grupo de Crypto Punks con 3 traits . No resulta un dato de relevancia puesto que los Crypto Punks con 3 traits son el grupo con más predominancia.

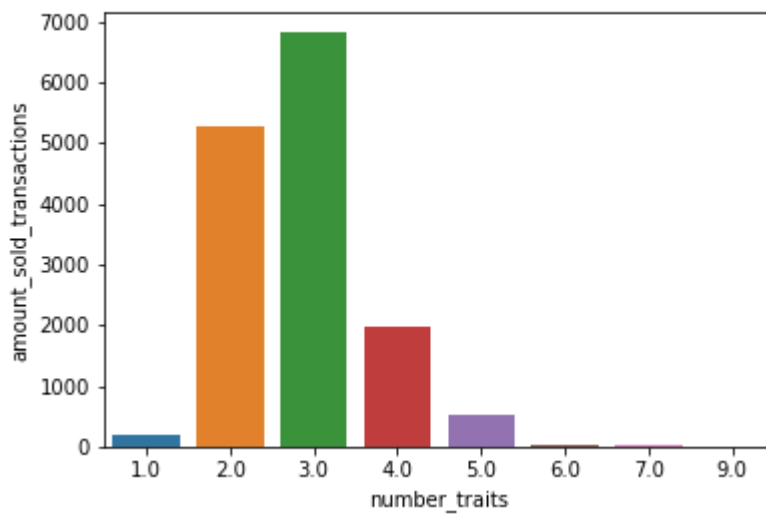


Fig 21. Transacciones realizadas por número de traits

Transacciones realizadas por tipo

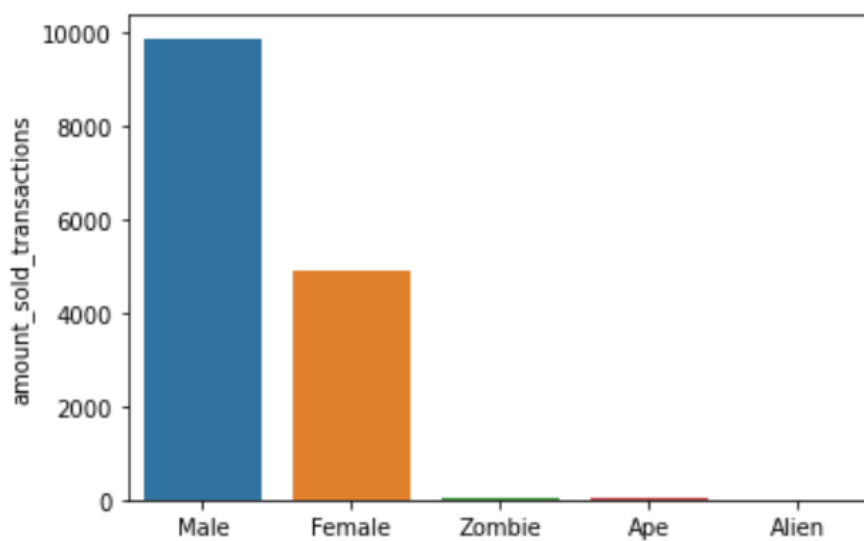


Fig 22. Transacciones realizadas por types

Transacciones realizadas por skin

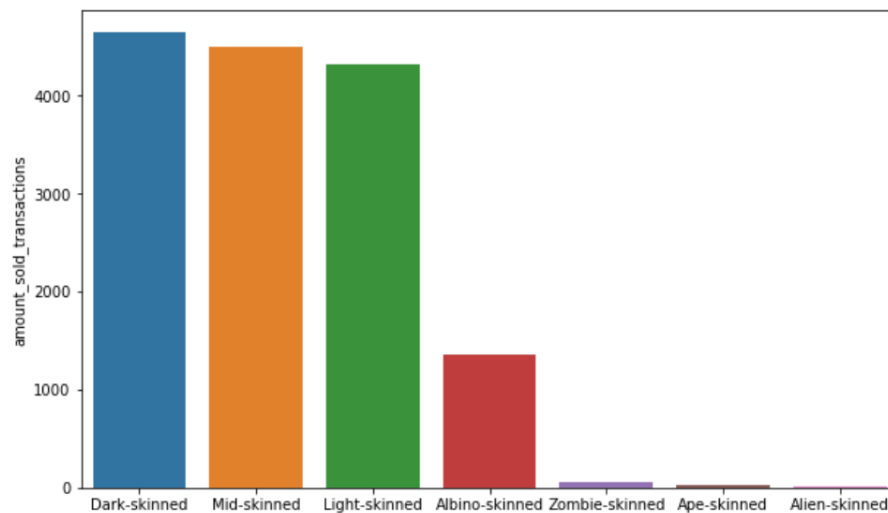


Fig 23. Transacciones realizadas por skins

3.2.7 Análisis de Outliers y precios

Precios de venta:

Analizamos únicamente el precio de venta de los Crypto Punks. De los que se han vendido por precios medios dentro de los valores normales y los que destacan como outliers por tener un precio muy elevado.

Por las ventas outliers tenemos un total de 1037 ventas realizadas con una media de precio de 340371 y una desviación típica de 470000 dólares el valor de mínimo de venta considerado outlier es de 182.240 dólares y el valor máximo de 7.668.318 dólares.

En las ventas consideradas normales tenemos un total de 14665 ventas con una media de 36214 y una desviación típica de 43612 el valor de venta mínima es de 3 dólares y el valor de venta máximo es de 182.000 dólares.

Distribuciones: Vemos el porcentaje de skins y types considerados como outliers en los dos análisis de transacciones para ver si la distribución se mantiene o encontramos un porcentaje muy distinto en las dos distribuciones.

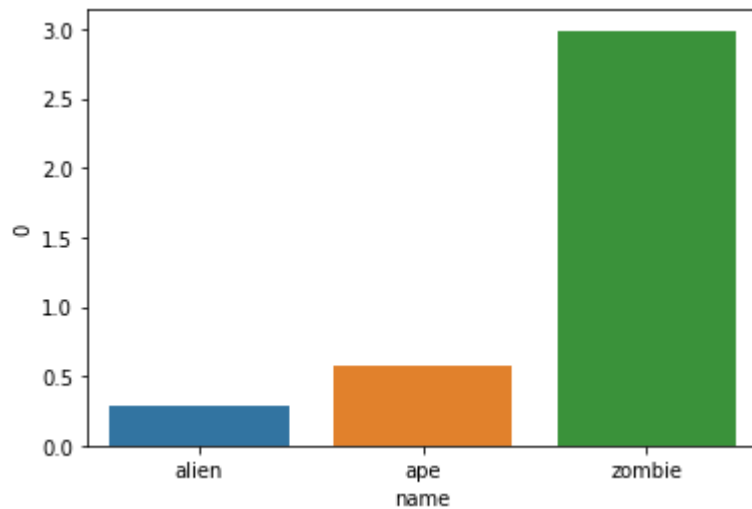


Fig 24. Porcentaje types outliers en ventas outliers

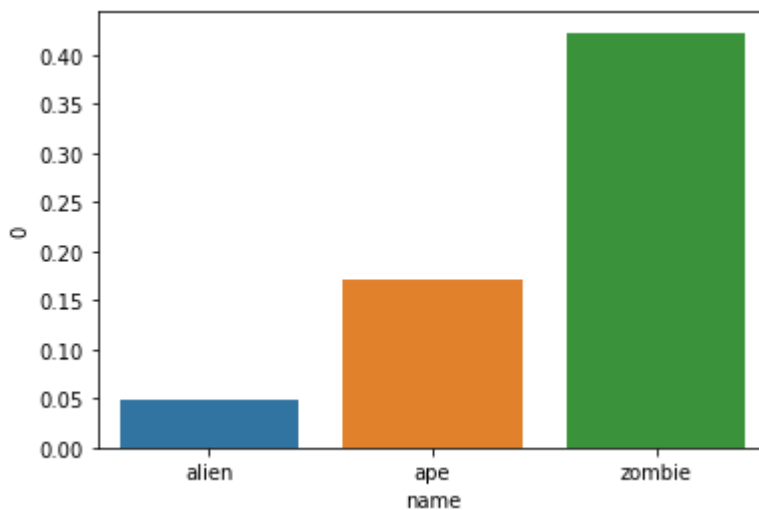


Fig 25. Porcentaje types outliers en ventas

Porcentajes de aparición de los rasgos muy bajos en las dos pero interesante. Los rasgos outliers son hasta 7.5 veces más frecuentes en las transacciones con precios más elevados que en el conjunto medio de transacciones. Si analizamos las 30 ventas con el precio más alto

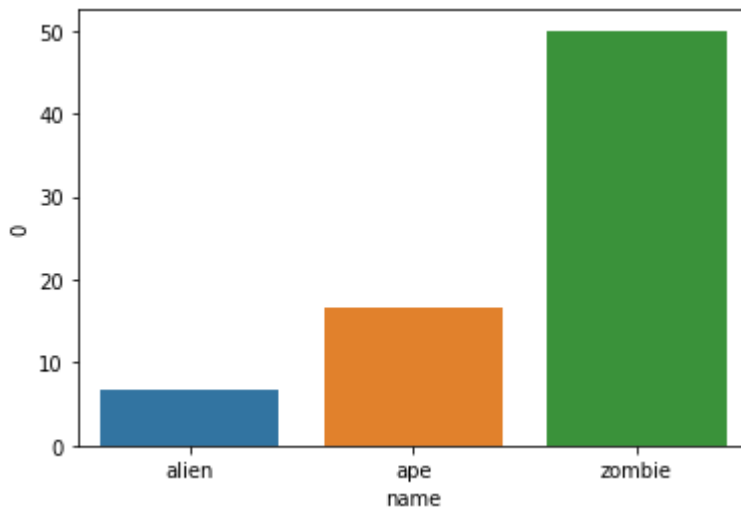


Fig 26. Porcentaje tipos outliers en top 30 ventas

La distribución en las 30 ventas con precio más alto toma valores distintos. El conjunto de Crypto Punks con tipos considerados outliers forman el 73.33 % de estas ventas. Se ve cómo a medida que se analizan precios de venta más altos la frecuencia de aparición de cryptos con rasgos menos frecuentes aumenta.

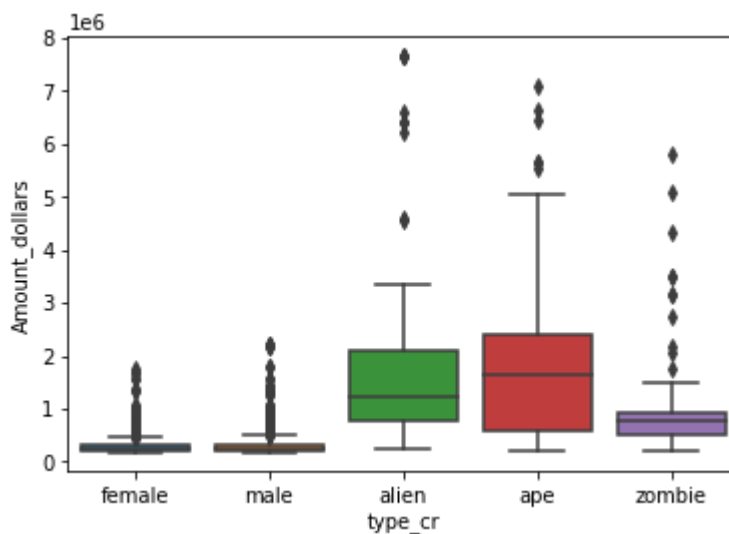


Fig 27. Amount for types en transacciones (Bid and sold) outliers.

La presencia de Alien, Ape y Zombie en las transacciones consideradas outliers por su alto valor es muy baja como se ve en la figura Fig 25. Pero como vemos en este gráfico (Fig 27.) que indica la cantidad de dólares en pujas y en ventas la cantidad asociada a estos tipos es muy superior a la cantidad asociada a Male y Female.

Crypto Punks con número de rasgos outliers por rangos de transacciones.

Analizando el número de traits por crypto punk en las dos distribuciones de transacciones.

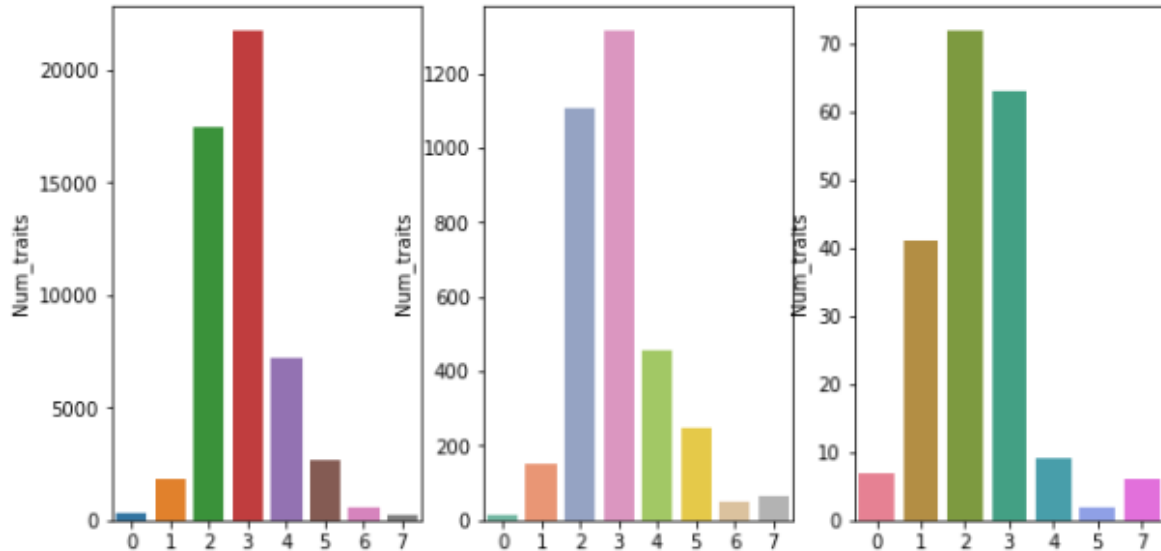


FIG 28. Num traits per normal transactions, outlier transaction y top 200 transactions.

En las figuras anteriores se ve la frecuencia del número de rasgos en distintas categorías de transacciones. La frecuencia es muy similar en las transacciones normales y outliers. Se aprecia un cambio de frecuencia en las 200 transacciones con más dinero asociado. Especialmente para los Crypto Punks con 1 trait.

Cuando se analizan los punks individuales coincide que muchos punks con 1 trait en el rango de las 200 transacciones son del tipo ape, alien o zombie.

4. Machine Learning

Para analizar las características del conjunto de Crypto Punks con transacciones asociadas dentro de los valores considerados no outliers y ver si encontramos patrones de tipos que describen el comportamiento de estas se usan los algoritmos de Machine Learning de Random Forest i XGBoost.

Este análisis se realiza con un conjunto de datos que tiene como target la diferencia de dinero asociado a la venta inicial y final con los valores comunes descartando todas las transacciones con un valor más elevado del tercer cuartil más 1.5 por el rango intercuartil.

El conjunto de datos lo forman todos los punks que sean de tipo Male y Female, con skins Albino, Dark, Light y Mid. Con número de rasgos de 2, 3 o 4 y sin tener en cuenta el rasgo de Earing y hidden earing.

Características:

- Type male y female
- Skins: Albino, Dark, Light, Mid.
- Cantidad de rasgos (sin hidden traits)
- Diferencia de valor entre transacción inicial y final
- Rarity
- Número de Pujas y ventas por Crypto Punk

Este procedimiento nos permite ver si hay un patrón distintivo a la hora de llevar a cabo las transacciones.

4.1 Random Forest Classification

Algoritmo de Machine Learning basado en árboles predictores que crea una colección de árboles de decisión y devuelve la clase seleccionada por más árboles. En el caso actual clasifica los precios y atributos para encontrar patrones de decisión. Se evalúa un modelo base de Random Forest Classifier y se lleva a cabo una cerca aleatoria de parámetros para ver si se puede mejorar el modelo base.

El dataset usado para entrenar el modelo se basa en los Cripto Punks que cumplen las características anteriores, como target se ha calculado el beneficio de compra asociado a estos y se ha categorizado en 4 clases dependiendo de su valor, si el balance era negativo, bajo, medio o alto.

El modelo inicial ha resultado con una precisión del 37%, la clase mayoritaria es un 35% del total de la distribución indicando que se podría llegar a un valor similar de forma completamente aleatoria. También el f1_score y el recall_score son muy bajos indicando una capacidad muy baja de predicción.

```
Accuracy: 0.3774480712166172
f1_score: 0.33526037500640693
recall score: 0.3774480712166172
```

Fig 29. Métricas Random Forest Classifier.

Para mejorar el modelo se ha realizado un Randomized Search CV, que valora distintos parámetros para el modelo y devuelve los parámetros con mejores métricas y se ha hecho un upsample de las clases minoritarias.

Los resultados del nuevo modelo han mejorado en un 15% la capacidad de predicción del modelo. Esto indica que el modelo es capaz de predecir de forma acurada el 50% de las clases a las que pertenecen los Crypto Punks. Vemos que supera la predicción anterior pero no es capaz de predecir de forma acurada las categorías de las ventas. Seguimos con unas métricas muy bajas, todas alrededor del 0.51 indicando la aleatoriedad de la elección.

Accuracy: 0.5178346621905161
 f1_score: 0.5216143085047621
 recal_socre: 0.5178346621905161

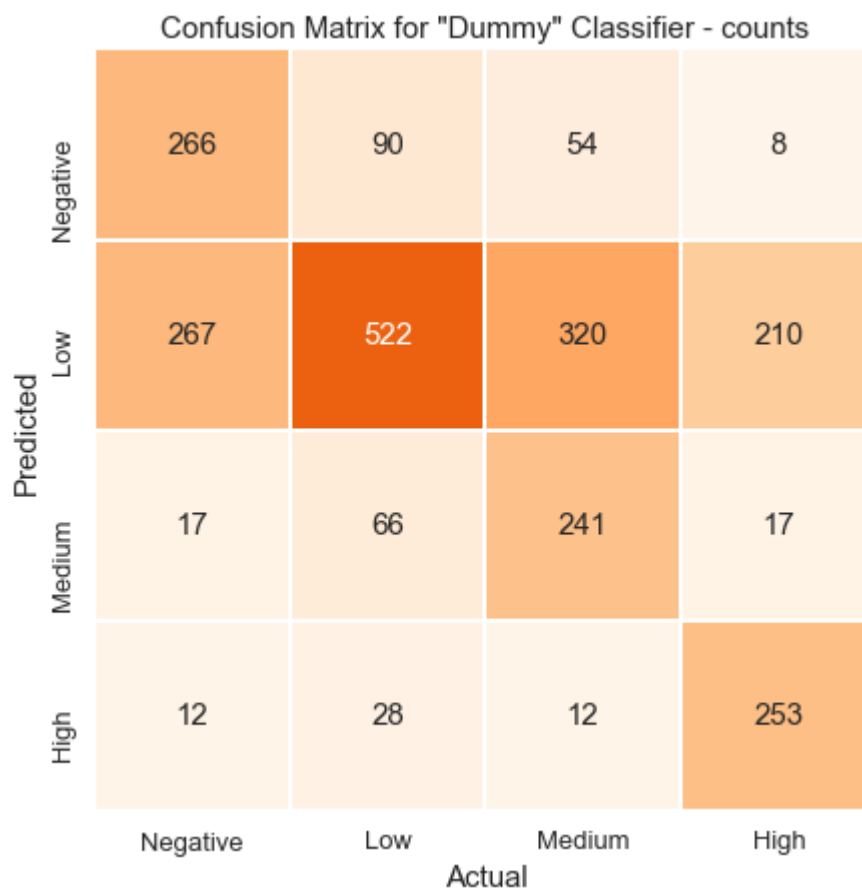


Fig 30. Métricas y Matriz de confusion Random Forest Classifier.

La matriz de correlación nos muestra como las predicciones se concentran en la categoría low, la más frecuente en el ds usado y como los valores reales positivos son inferiores en las predicciones a las suma de los falsos, excepto en la columna low. Por ejemplo en la categoría Medium tenemos 241 valores reales positivos y nos ha predecido 380 valores de esta categoría como valores pertenecientes a otras categorías. Se da el mismo supuesto en las categorías High y Low.

Dado el resultado del estudio inicial se ha incluido la variable de número de pujas y ventas por Crypto Punk para ver si hay una posible relación entre estas y el precio de venta, se ha

aplicado de nuevo el Random forest classification para ver si aumentaba la capacidad de predicción. Los resultados obtenidos han sido similares a los anteriores.

```
Accuracy: 0.5646515533165407
f1_score: 0.5441164733409053
recall score: 0.5646515533165407
```

La variable de número de pujas ayuda ligeramente a mejorar el modelo de predicción. Ha mejorado el modelo en un 5% llegando al 55% de precisión.

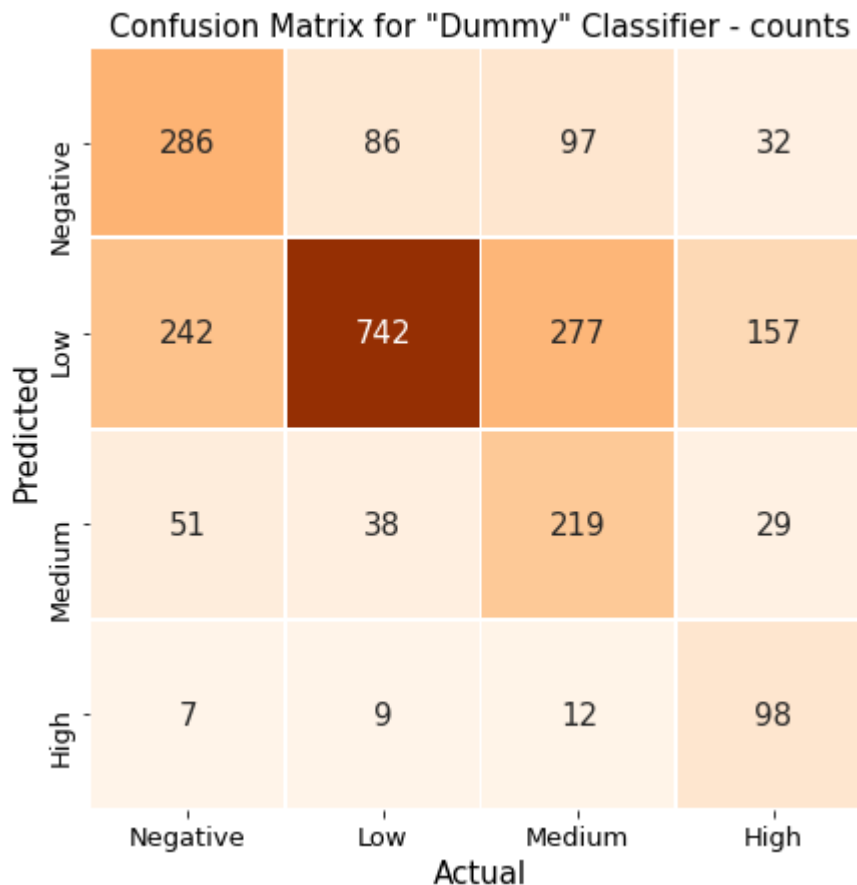


Fig 31. Matriz de confusión con número de transacciones

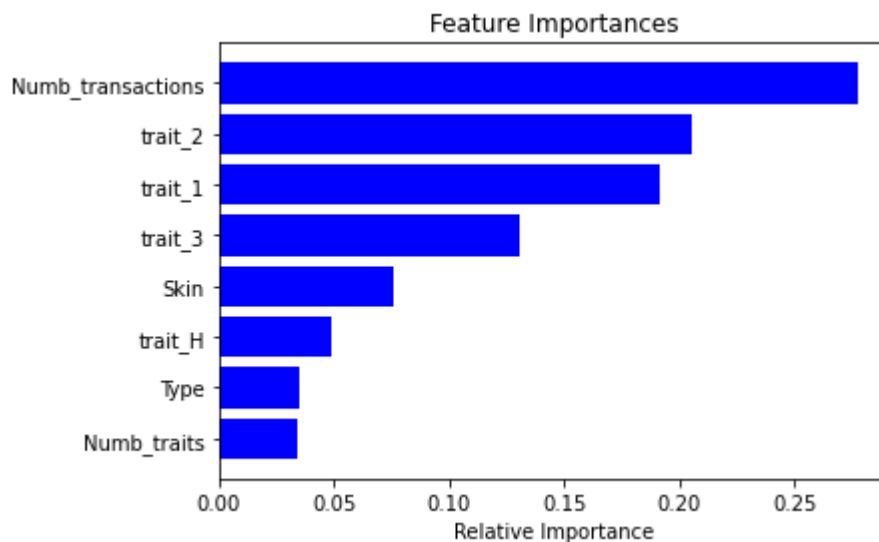


Fig 32. Importancia de variables.

La variable de número de transacciones (ventas y pujas) es la variable con más significación a la hora de predecir la categoría del Crypto Punk. Seguida por la clase de los dos primeros traits. Esto indica que hay una cierta relación entre la cantidad de operaciones realizadas con un Crypto Punk y la diferencia de dinero entre la primera y la última transacción realizada con este. Aun así no es suficiente para predecir de forma acurada su valor.

4.2 XGBoost Gradient

XGBoost es un algoritmo de machine learning supervisado de código abierto que se presenta como una versión mejorada del gradiente de árboles reforzados.

Como punto a destacar de este algoritmo es su función paralelizable que nos permite usar de forma eficiente toda la capacidad de procesamiento que se dispone ganando una mayor rapidez.

En este caso, utilizamos el modelo dirigido a la clasificación no binaria junto con grid search cross validation que nos servirá para encontrar los mejores parámetros que se adapten a nuestro dataset

el dataset a entrenar será el mismo que hemos utilizado con Random Forest con una pequeña modificación respecto al tratamiento de los tipos Male y Female ya que en vez de utilizar labelcoder hemos utilizado onehotencoder. Respecto a la clasificación del target se trata de la misma que en el random forest 4 etiquetas que representan el estado del balance.

Los resultados de este modelo han dado peor puntuación respecto al modelo anterior con un máximo de accuracy del 48% y f1-score del 0.42%. Estos resultados pueden ser debido al tamaño del dataset puesto que este algoritmo funciona mejor con datasets de gran tamaño.

	precision	recall	f1-score	support
0	0.43	0.41	0.42	281
1	0.50	0.88	0.64	487
2	0.00	0.00	0.00	281
3	0.00	0.00	0.00	73
accuracy			0.48	1122
macro avg	0.23	0.32	0.26	1122
weighted avg	0.32	0.48	0.38	1122

Fig 33. Métricas Xgboost.

En la matriz de confusión no aporta mejores métricas que random forest incluso vemos que en la predicción no hay ningún valor en la categoría high aunque acertando el 39% de los valores en la categoría low

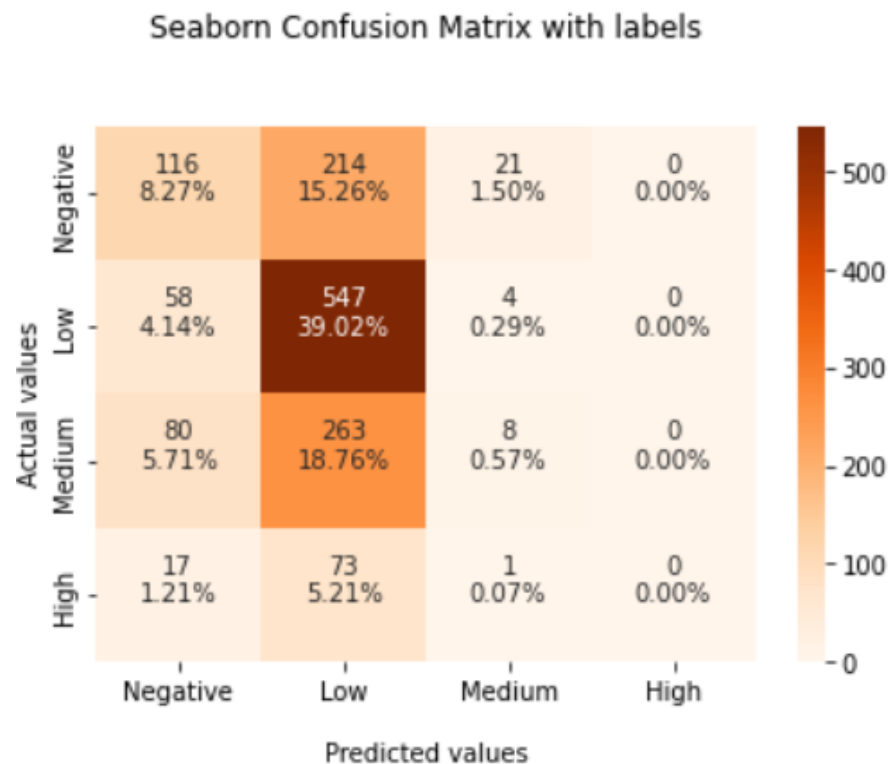


Fig 34. Matriz de confusion Xgboost.

En cuanto a las variables de mayor importancia xgboost consideró variables diferentes a las que considero random forest

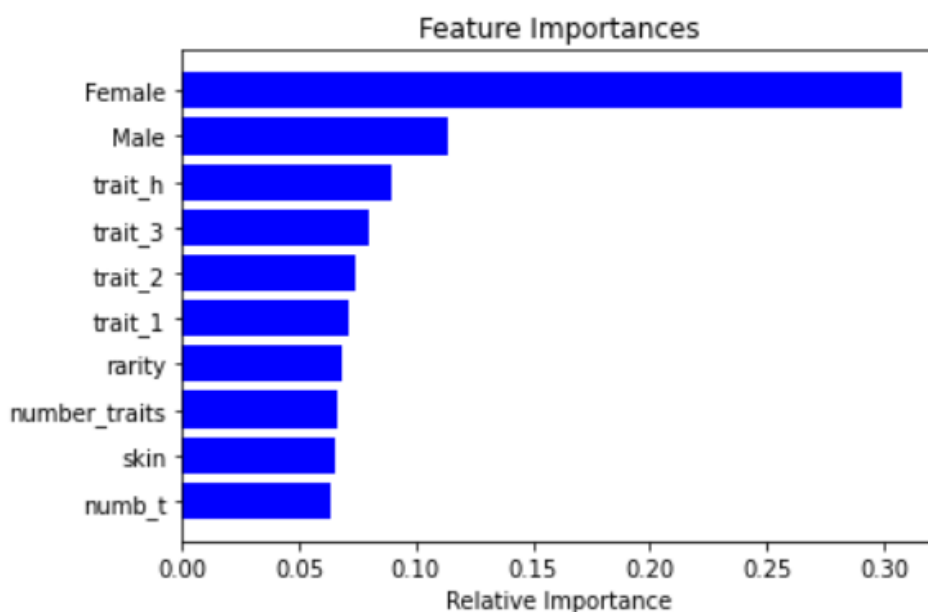


Fig 35. Importancia de variables de Xgboost.

Dando el mayor peso sobre la variables de tipo Female y Male seguido de los atributos y dando el menor peso en el número de transacciones, variable que en random forest fue la de mayor peso.

Los resultados entre los dos modelos muestran que no parece haber correlación entre atributos y valor asociado puesto que si fuera así al menos veríamos una semejanza a la hora de elegir las variables de mayor peso.

Reafirmando aún más la hipótesis nula, no parece haber relación entre atributos que expliquen el valor.

5.Conclusiones

Al analizar los Crypto Punks, sus características y los precios asociados se ven dos grupos muy diferenciados. El grupo de Crypto Punks con Type poco común, Aliens, Zombies, Ape y los más comunes que conforman más del 98% de los Crypto Punks de Male y Female.

Las diferencias de precio entre estos dos grupos son muy elevadas e indican un interés de compraventa mucho más elevado por el grupo menos frecuente.

Al analizar el grueso de Crypto Punks de tipo Male y Female hay distintos conjuntos de características, número de rasgos skins y tipos de rasgos. Una vez hecho el análisis exploratorio no se aprecia ningún patrón que indique un conjunto de características relacionadas con el precio ni ningún factor relevante que esté correlacionado con el precio. Después del análisis exploratorio se han descartado todas aquellas variables outliers que pudieran influir de forma más significativa a desbalancear los datos y se han agrupado en cuatro categorías en función de los beneficios en ventas producidos por cada punk.

Se ha trabajado con dos algoritmos de ML de clasificación el Random Forest Classifier y el XGBoost tanto los resultados inicial como los obtenidos después de aplicar modificaciones a los datos y mejorar los parámetros de los modelos han sido negativos, sin indicar ninguna correlación entre las características de los Crypto Punks y sus beneficios generados.

Se ha visto que la inclusión del número de operaciones realizadas en un Crypto Punk aumenta la capacidad de predicción del modelo. indicando una ligera relación entre este número y el precio. Pero no se puede concluir que este sea un factor determinante. Esto tampoco nos dice si las características influyen a aumentar el número de pujas, puede ser que el simple hecho de que la gente pueje por un Crypto Punk lleve a un factor llamada que genere más pujas para ese Crypto Punk concreto sin tener en cuenta sus características.

Visto esto se descarta la hipótesis alternativa: el precio viene dado por las características de los Crypto Punks, por el conjunto más genérico de los Crypto Punks y se mantiene la hipótesis nula que el precio no tiene relación a sus características.

Se destacan las diferencias entre el conjunto de Crypto Punks más frecuentes y aquellos de tipo Zombie, Alien y Ape que sí que muestran una clara tendencia a precios más elevados. Es una situación en la que se cumple la hipótesis alternativa, pero debido a su baja frecuencia no se puede generalizar al conjunto de Crypto Punks.

7. Mejoras para próximo estudio

Al quedar de manifiesto que los atributos no tienen una correlación en su valor, hemos querido replantear el enfoque del modelo aunque por falta de tiempo no hemos podido desarrollarlo más.

Para esto, hemos descartado que el motivo del precio de los crypto punks se deben a los atributos y nos hemos fijado en algunos aspectos de las transacciones :

- solo analizar criptopunks con 3 o más transacciones
- generar una nueva variable con un coeficiente el sumatorio de incrementos de las transacciones
- coger como punto de partida la penúltima transacción como precio inicial de inversión (prueba para modelo regresivo)
- generar otra nueva variable con la media de transacciones descartando el último valor que fijaremos como target (prueba para modelo regresivo)

Al realizar estos nuevos cambios las métricas han mejorado considerablemente con un 62% de accuracy y 72% de f1-score .

	precision	recall	f1-score	support
0	0.74	0.70	0.72	159
1	0.58	0.71	0.64	97
2	0.53	0.53	0.53	137
3	0.45	0.27	0.34	33
accuracy			0.62	426
macro avg	0.58	0.55	0.56	426
weighted avg	0.62	0.62	0.61	426

Por lo tanto, los datos demuestran que las características de los criptopunks sujetos a este estudio no tienen mucha relación con el precio si no que el precio se puede deber más a técnicas de especulación financiera.

Dado los resultados de la clasificación, hemos querido probar un algoritmo de regresión para predecir el precio cogiendo como punto inicial la penúltima transacción y el target como la última transacción realizada.

Las pruebas con algoritmos regresivos, muestran un buen punto de partida para posibles estudios en un futuro puesto que en la primera prueba realizada el modelo ha arrojado un resultado bastante aceptable con un 84% en el r2 score aunque un mean_squared_error demasiado alto

```
print(r2_score(y_test,rfPred))
print(mean_squared_error(y_test,rfPred))
```

```
0.8454733591447071
431457088.2157409
```

```
PrintActualPredicted(y_test,rfPred)
```

	Actual	Predicted
7400	54619.9625	48459.574666
2412	41186.2640	40283.467712
6819	33455.3200	39854.355208
3152	50552.4614	47614.979457
7284	115030.3500	86036.389684
...
1512	70.4301	504.424780
351	111736.8900	102345.589611
8196	36296.5300	40662.229295
8590	7515.5436	10275.676210
8654	180631.5500	92450.782010

426 rows × 2 columns

De esta forma queremos demostrar que aún no haber encontrado relaciones entre características y precios podría haber formas distintas de predecir hasta cierto punto el precio, sin dar tanta importancia a estas características y dando más importancia a la parte económica del proyecto, ya habíamos visto cómo incluir el número de transacciones mejoraba el modelo de predicción y como esto lo corrobora.

6. Referencias:

Larva Labs(2021) *Crypto Punks* <https://www.larvalabs.com/cryptopunks>

Google trends (2021) <https://trends.google.es/trends/?geo=ES>

Precio Ethereum (2021), *Coin Marker* <https://coinmarketcap.com/es/currencies/ethereum/>

McNamara, Ryan. Benzinga (2021), *Crypto Punks Watchlists: Celebrities and influencers who own Punks.*

<https://www.benzinga.com/money/cryptopunks-watchlist-celebrities-and-influencers-who-own-punks/>

DefyPunk(2021), *Crypto Punks information.* <https://defypunk.com/>

Browne, Ryan. CNBC (2021). *Crypto collectibles are selling for thousands.*
<https://www.cnbc.com/2021/02/25/nfts-why-digital-art-and-sports-collectibles-are-suddenly-so-popular.html>