**ITA-PROJECT DATA SCIENCE**

**INFORME**

**Titulo: Cryptopunks**

**Estudiantes: Tomas Lucaccini & Manel Forcales**

**Supervisor: Kevin Mamaqi**

Ultima Actualización : 19/07/2021

**Índice**

[Introducción del Proyecto 3](#__RefHeading___Toc209_743788306)

[Objetivos 3](#__RefHeading___Toc211_743788306)

[Metodología 4](#__RefHeading___Toc213_743788306)

[Fase 1: Extracción, Transformación y Carga de datos (ETL) 4](#__RefHeading___Toc215_743788306)

[Fase 2: Análisis visual exploratorio (EDA) 4](#__RefHeading___Toc217_743788306)

[Fase 3: Predicciones 4](#__RefHeading___Toc219_743788306)

[Desarrollo Técnico 5](#__RefHeading___Toc221_743788306)

[Análisis exploratorio 6](#__RefHeading___Toc382_763797363)

[Conclusiones / Recomendaciones 14](#__RefHeading___Toc223_743788306)

# **Introducción del Proyecto**

Los ***CryptoPunks*** son 10,000 personajes coleccionables generados de forma única con certificado de propiedad almacenados en “Ethereum Blockchain”. No hay exactamente dos iguales como se muestra en la siguiente imagen.



Originalmente creados en el 2017, podían ser reclamados de forma gratuita por cualquier persona con una cuenta en “Ethereum” y los 10,000 fueron reclamados rápidamente. Ahora deben comprarse a alguien a través de “blockchain”. En este mercado se puede ofertar, pujar y ofrecer punks para la transacción.

La información de cada punk puede encontrarse en las siguientes webs:

* https://www.larvalabs.com/cryptopunks
* <https://defypunk.com/>

Los punks con fondo azul no están a la venta y no tienen ofertas actuales.

Los punks con un fondo rojo están disponibles para la venta por su propietario.

Por último, los punks con un fondo púrpura tienen una oferta activa en ellos.

# **Objetivos**

El objetivo principal de este proyecto es analizar la información de todos los ***CryptoPunks*** posibles y averiguar si hay diferencia en el precio de estos punks con sus distintas agrupaciones.

Para ello tendremos que validar o rechazar la hipótesis Nula (H0)

H0 (nula= No existen diferencias en el precio de los punks y sus agrupaciones

HA (alternativa) = Existen diferencias

Además vamos a intentar predecir mediante regresiones el precio de un posible punk mediante sus características y/o agrupaciones.

# **Metodología**

La metodología de trabajo se ha definido en diferentes fases, vamos a explicar brevemente cada una de ellas:

## **Fase 1: Extracción, Transformación y Carga de datos (ETL)**

En esta fase vamos a crear un modelo Entidad-Relación donde definimos los distintos parámetros que vamos a utilizar. Para este fin utilizaremos MySQL work-bench.

Utilizaremos la técnica de “Webscapping” para descargar los datos de las diferentes páginas web:

* <https://www.larvalabs.com/cryptopunks>
* <https://defypunk.com/>

Una vez tenemos la información descargada, la limpiaremos y la transformaremos para cargar los datos en la base de dados

## **Fase 2: Análisis visual exploratorio (EDA)**

En esta fase vamos a hacer un análisis descriptivo de los datos tanto si son variables categóricas como si son numéricas, visualizamos los datos y miraremos los diferentes parámetros de sus distribuciones (media, desviación estándar, rangos,…) y si es necesario analizaremos posibles “outliers” y los filtraremos.

## **Fase 3: Predicciones**

En esta fase vamos a aplicar los diferentes técnicas de “machine learning” para analizar en detalle los datos y predecir resultados. Utilizaremos técnicas de Clustering (“unsupervised learning”) y regresiones (“supervised learning”)

*Nota*:

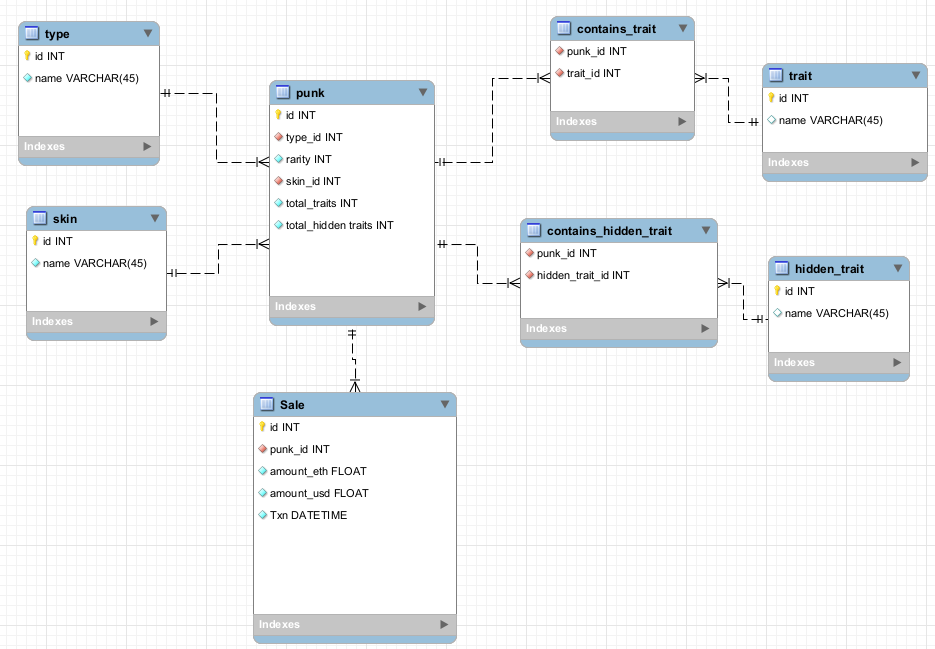
Toda la información relativa al código utilizado en este proyecto se encuentra en el repositorio de GITHUB:

https://github.com/ManelForcales/IT\_academy\_DataScience\_project

# **Desarrollo Técnico**

## **Base de datos**

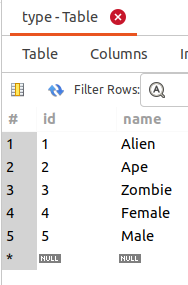
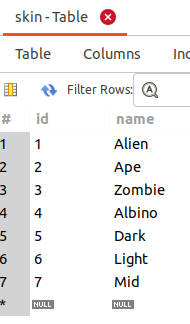
Hemos creado el siguiente modelo relacional (ERD) de la Figura 1 para estructurar los datos.



*Figura 1. diagrama Entidad-Relación para las diferentes tablas del proyecto*

La tabla “punk” contiene el ID de cada punk, el identificador ID asociado con el tipo de punk, la rareza del punk, el identificador ID asociado con la piel del punk y el numero total de atributos no ocultos y ocultos.

Las tablas “type” y “skin” de la Figura 2,contienen el ID y el correspondiente nombre. Como son pocas filas, se han introducido a mano a la BBDD directamente



*Figura 2. ejemplo de la tabla “skin” y “type”*

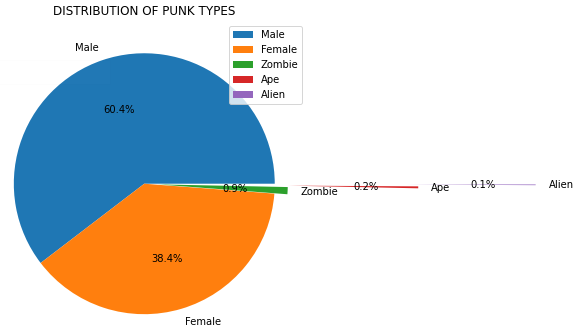
Las tablas “contains\_trait” y “contains\_hidden\_trait” se relacionan con el ID del punk de la tabla “punk” y con un identificador ID asociado con el nombre de cada trait/hidden\_trait

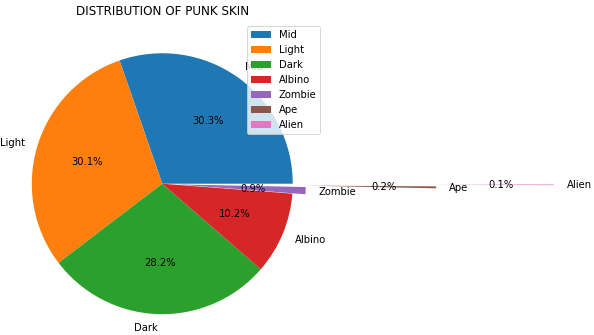
Finalmente la table “Sale” identifica las transacciones que ha existido de cada punk ya sean ventas/ofertas o pujas, con el precio tanto en ETH como en $ y la fecha de esta venta.

## Análisis exploratorio

La Figura 3 muestra la proporción de tipos de punk y tipos de piel del total de todos los punks.

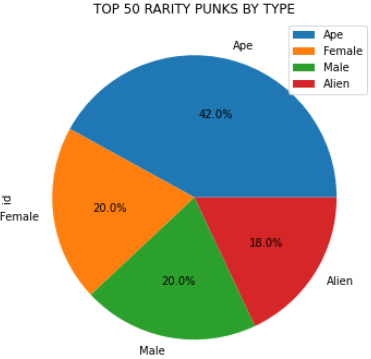
Se puede observar que una parte muy pequeña de datos corresponden a tipo “Alien”,”Ape” y “Zombie” mientras que un 98.8% corresponden a “Male” y “Female”. Entre Male y Female el tipo de piel se reparten bastante uniformemente entre Dark, Light, Mid y Albino, siendo Albino la menor distribución para este tipo de piel.

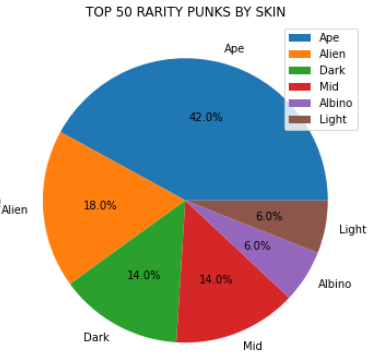




*Figura 3. distribución de tipos y piel de los punks*

Si clasificamos los punks por su rareza, siendo el 1er punk en más raro y analizamos los primeros 50 punks vemos en la Figura 4 que el porcentaje de punks en función del tipo y piel cambian respecto a las graficas anteriores.



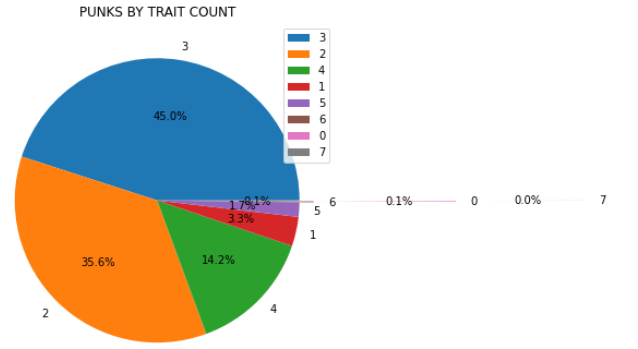


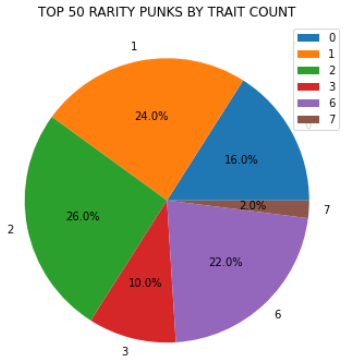
*Figura 4. distribución de tipos y piel de los punks de los top 50 punks mas raros*

Es interesante observar que “Ape” y “Alien” destacan por su alta rareza mientras que “Zombie” no aparecen en esta lista, por lo que no deben ser considerados raros.

Más tarde analizaremos si hay una correlación entre rareza/tipo de punk y precio asociado.

La Figura 5 muestra la proporción de punks con su numero total de atributos tanto de la muestra completa como de los 50 punks mas raros. Cabria esperar que un mayor numero de atributos iría asociado con una mayor rareza.



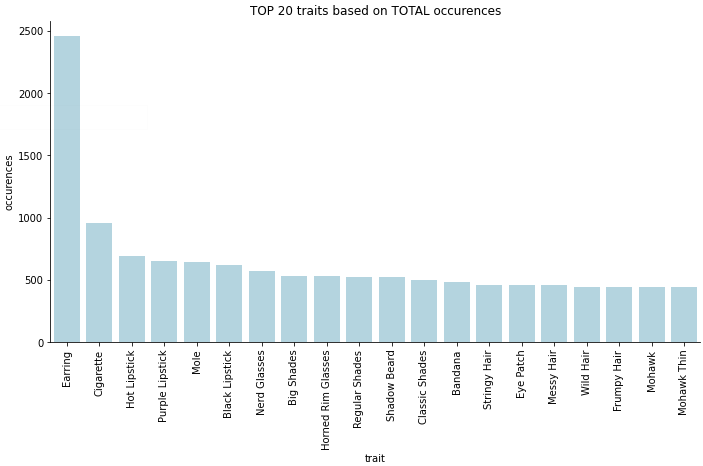


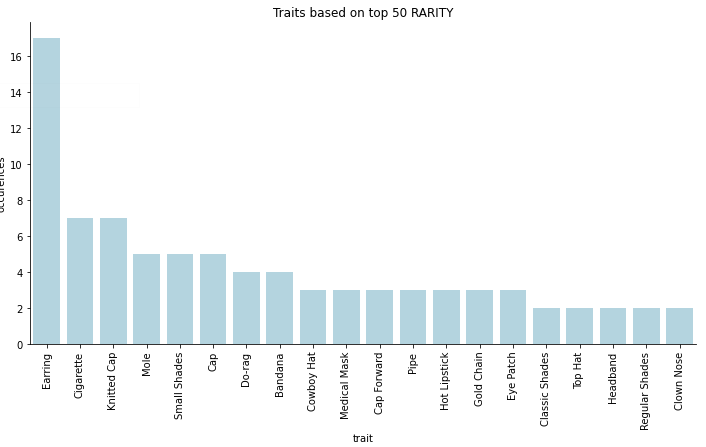
*Figura 5. distribución de numero total de atributos de todos los punks y de los 50 punks mas raros*

En el caso de tener 7 atributos solo existe 1 punk y está entre los 50 mas raros. De manera similar también ocurre si tienen 6 atributos, pero esta tendencia se rompe para punks con 5 o 4 atributos. El caso contrario también existe, hay pocos punks con 0 atributos pero aun así se encuentran entre los mas raros. Aparentemente no hay una correlación clara entre numero de atributos y rareza.

Existe un patrón similar para los atributos ocultos pero no los mostramos en este informe.

Tenemos un total de 89 atributos diferentes. Los 20 atributos que aparecen mas a menudo se muestran en la Figura 6. Claramente “Earring” es el atributo más común con diferencia. Si nos basamos solamente en los atributos que aparecen en los 50 punks más raros vemos que los mas comunes se mantienen pero aparecen otros





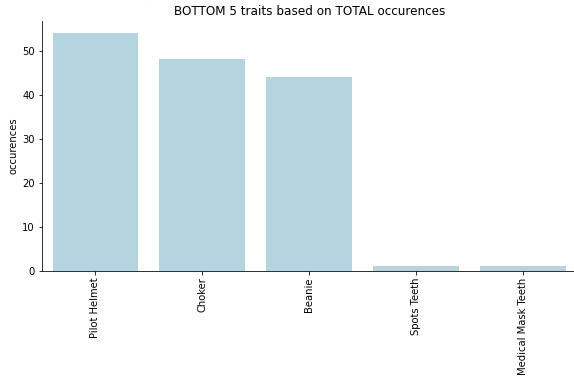
*Figura 6. ocurrencia de los atributos* entre todos los punks y los 50 punks mas raros

Si analizamos que atributos aparecen en la 2a grafica de la Figura 6 pero no aparecen en la 1a grafica veremos que son:

* ['Medical Mask', 'Top Hat', 'Clown Nose', 'Headband', 'Knitted Cap', 'Gold Chain', 'Cap', 'Do-rag', 'Pipe', 'Cowboy Hat', 'Cap Forward', 'Small Shades']

Habría que indagar si el precio de estos atributos es diferente a los mas comunes

Entre los atributos que menos aparecen son:



Como los 2 últimos atributos solamente aparecen 1 vez se puede considerar unificar con otros atributos similares. De la misma manera se podrían hacer agrupaciones de atributos que se llaman de manera muy similar.

**Ventas históricas**

La incorporación de este dataset al análisis nos nutre de tres flujos con valiosa información:

* Cantidad de transacciones
* Nivel de precios
* Fecha de transacción

Mediante el estudio de la cantidad de transacciones, puede observarse que la proporción de punks que han sido transaccionados al menos una vez respecto al total es relativamente baja. Tan solo un 30% del total de punks han sido vendidos al menos una vez:

-- INSERT GRAPH --

Como complemento del gráfico anterior, también podemos saber qué cantidad de veces se ha transaccionado cada uno. La cantidad de punks por cantidad de transacciones responde a la siguiente curva.

Chart, line chart, histogram

Description automatically generated

La evolución de la cantidad de las transacciones nos indica que hay una tendencia creciente marcada. El año 2017, quizás por ser un elemento innovador y por sus tan bajos precios, registra niveles superiores al año 2018. Desde aquel año en adelante, la dirección es siempre hacia arriba.

Chart, bar chart

Description automatically generated

Es también interesante observar el registro de transacciones por día desde el momento de inicio hasta la fecha, ya que pueden identificarse los picos de registro por día, alcanzando en el último año días con más de 150 transacciones realizadas.

Chart, histogram

Description automatically generated

Si analizamos tales transacciones por “Punk Type”, entonces la cantidad para cada uno de ellos resulta de la siguiente manera:

Chart, bar chart

Description automatically generated

Como puede verse, la mayor cantidad de transacciones se concentra entre los “Male” y los “Female”, hecho esperable dada la proporción inicial de punks bajo mismos atributos.

El análisis de precio merece un tratamiento particular. Por una parte, decir que tal variable posee dos unidades de medida: por una parte, el precio del punk en porcentajes de Ethereum; por otra parte, el precio universal del punk expresado en dólares. Cada una de las transacciones realizadas tiene como correlato una fecha específica de realización que permite contrastar en cada caso, el valor de cotización entre dólar y Ethereum.

Dos de las principales características de las criptomonedas son su volatilidad y su tendencia alcista en el tiempo. Esto significa que este tipo de activos financieros suele convalidar cambios bruscos en su cotización respecto a otras monedas, ya sea en sentido creciente como decreciente, pero en el largo plazo alcistas. Como los cripto-punks son comercializados mediante la crypto-moneda Ethereum, los elementos recién descriptos afectan de la misma manera al comportamiento de su precio en el tiempo.

Así se ve la evolución del precio promedio de los crypto-punks año a año.

Chart, bar chart

Description automatically generated

Mediante este gráfico es visible el crecimiento exponencial de tal activo, creciendo en proporciones aún más elevadas que las identificadas para la cantidad de transacciones. Es decir, que el efecto precio es aún más grande que el efecto cantidad. De hecho si se observa en detalle, el precio en dólares crece aún con más vigor gracias al efecto de valorización de la ciptomoneda Ethereum.

A continuación, puede verse también el precio de los principales 50 punks más caros:

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Si desglosamos el nivel de precio medio por “punk\_type”, entonces podemos ver que este se comporta de la siguiente manera:

Chart

Description automatically generated

Los Punks del tipo Alien sobresalen como los más caros por una buena distancia.

# **Conclusiones / Recomendaciones**

tbd