



# MACHINE LEARNING PROJECT

**MARÍA  
CAGIGAS**

Presentación



**CONTEXTO**



# ¿QUÉ TENEMOS?

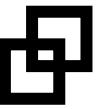
## 01. DATA SET

Tenemos un dataset con datos sobre **género** de videojuego, su **plataforma** de videojuego y su **desarrollador**.

También incluye las **ventas globales**, NUESTRO TARGET.

## 02. HIPÓTESIS

Creemos posible entrenar un **algoritmo** que obtenga **un MAE** menor a **0.5** en base a una predicción de ventas sobre el tipo de videojuego.

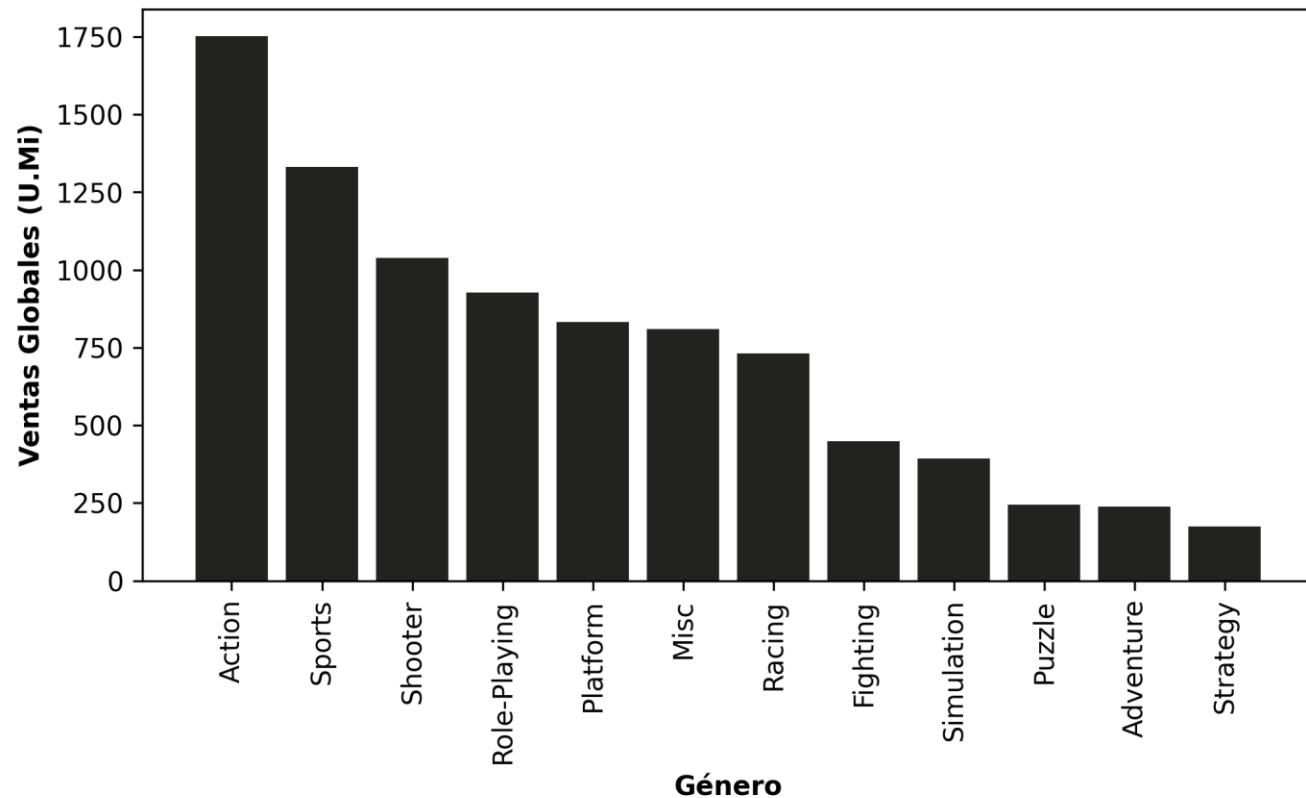


# DATOS



# ENTENDIENDO LOS DATOS

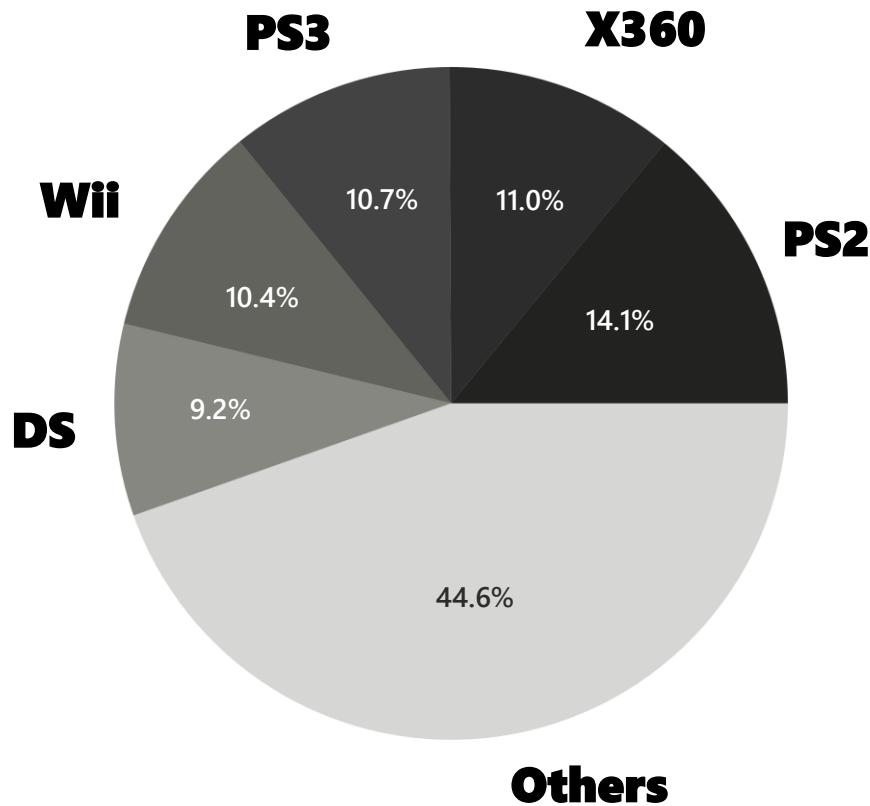
GÉNERO DE VIDEOJUEGOS





# ENTENDIENDO LOS DATOS

PLATAFORMA DE VIDEOJUEGOS





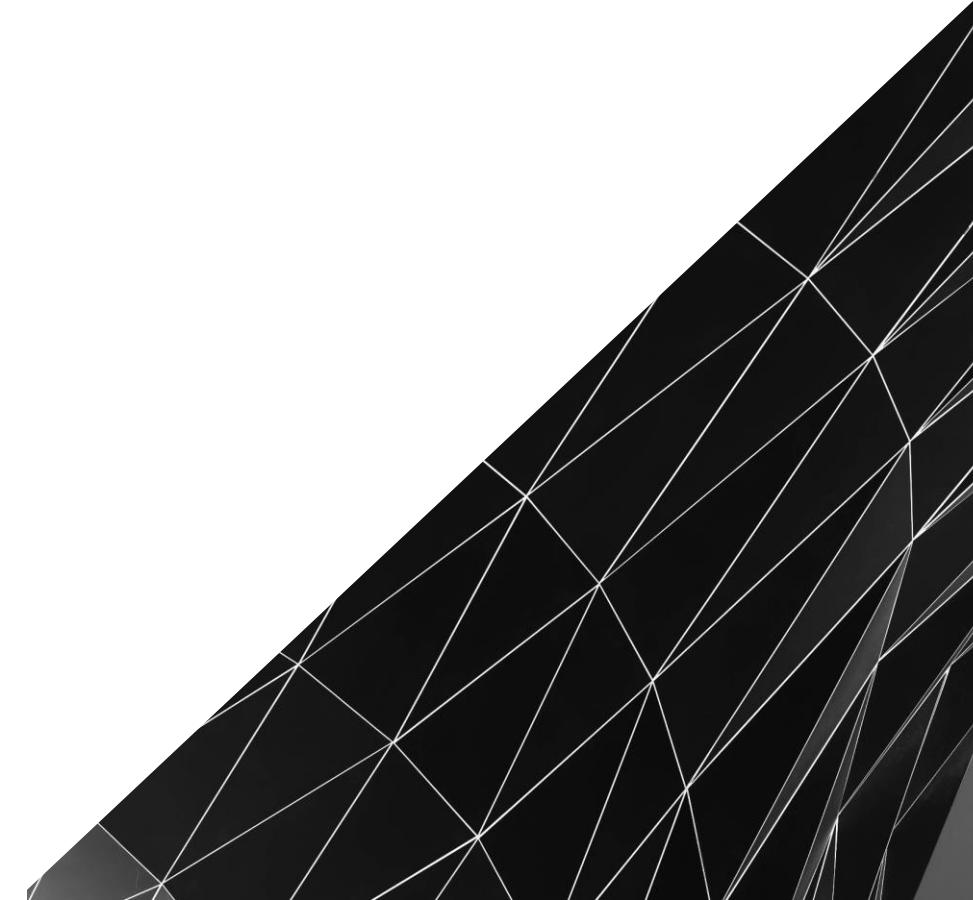
# PROCESO



# ¿WORKFLOW?



**01. DATA**





# ¿WORKFLOW?



**01. DATA**



**02. DATA WRANGLING  
DATA CLEANING**





# ¿WORKFLOW?



**01. DATA**



**02. DATA WRANGLING  
DATA CLEANING**



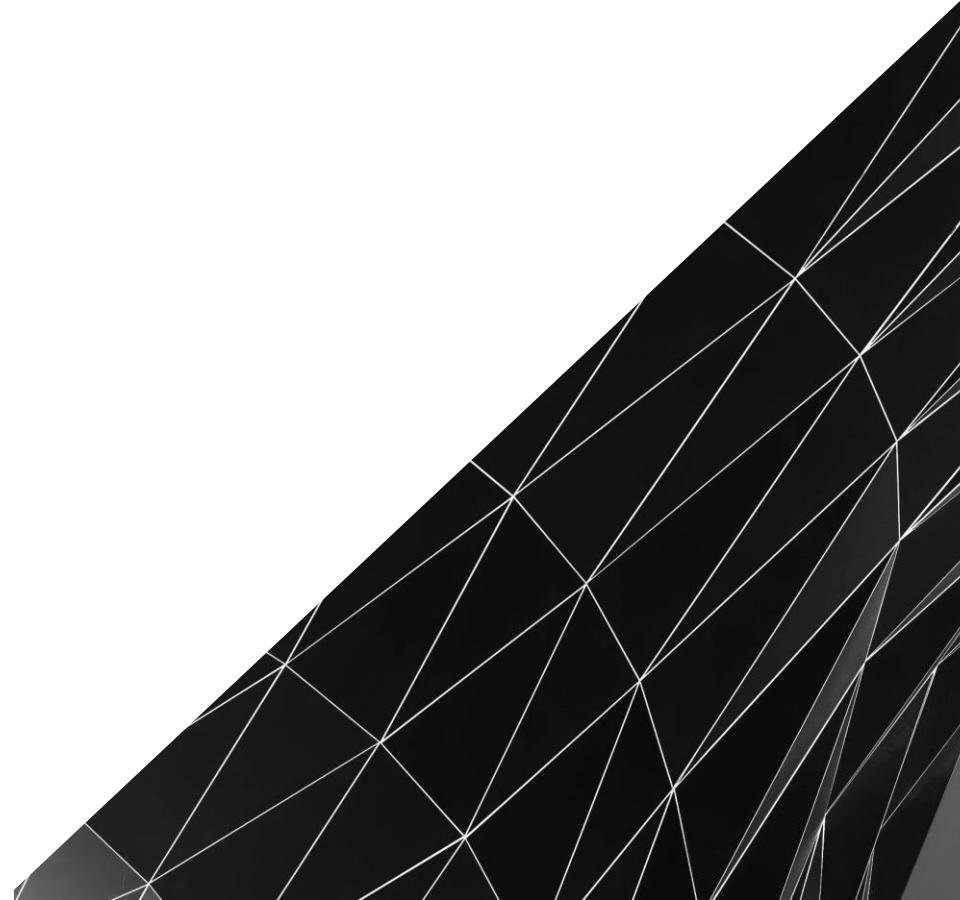
**03. VISUALIZACIÓN**





# ¿WORKFLOW?

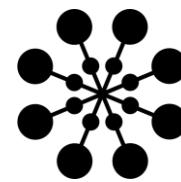
## ? 04. HIPÓTESIS



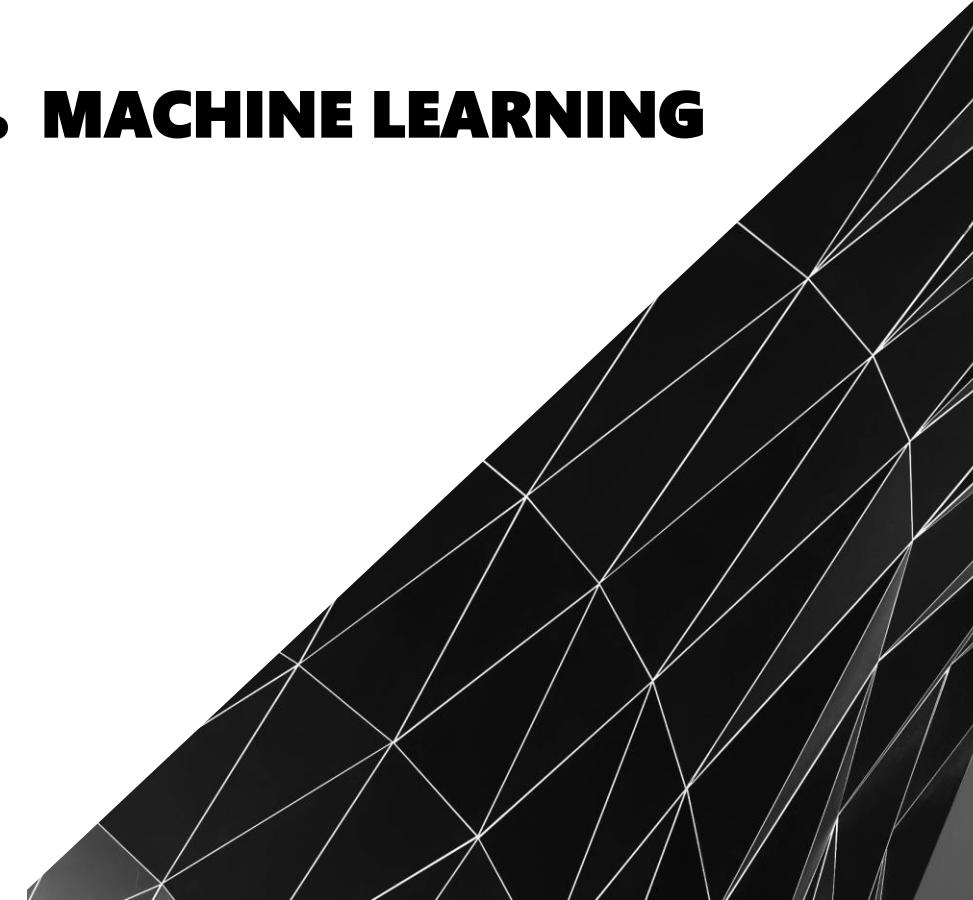


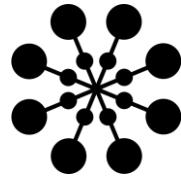
# ¿WORKFLOW?

? **04. HIPÓTESIS**



**05. MACHINE LEARNING**





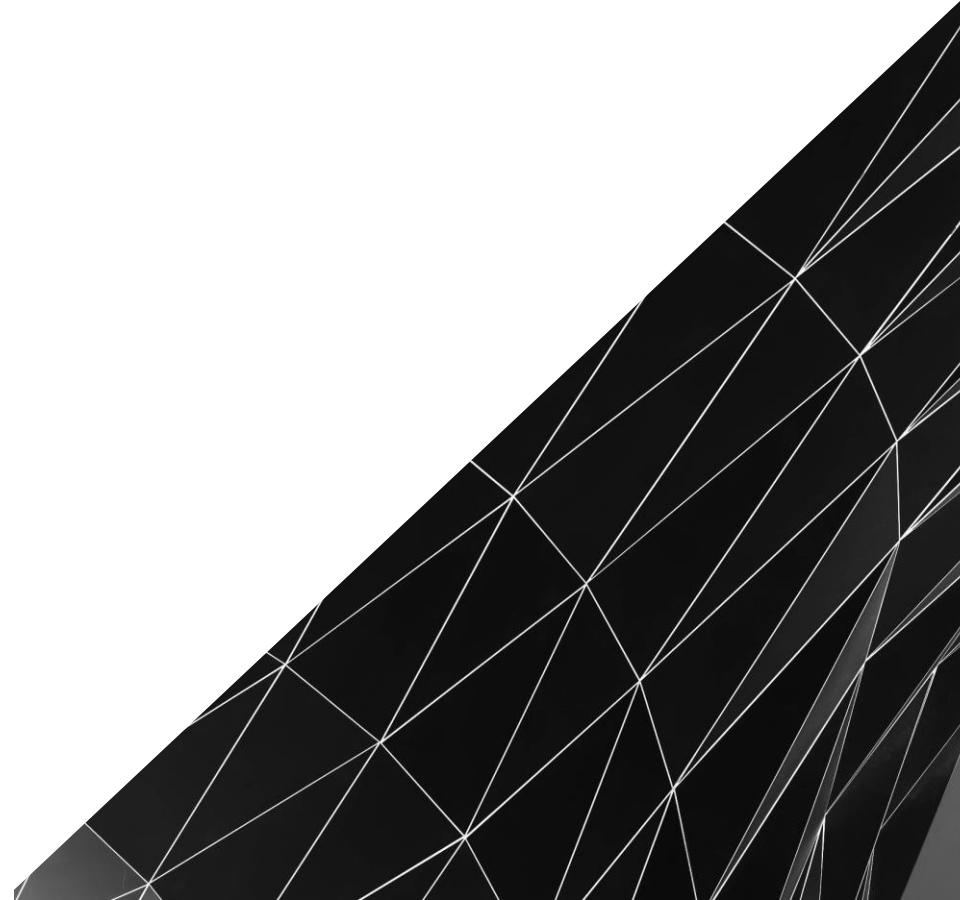
## 05. MACHINE LEARNING

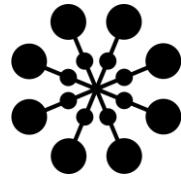


0110  
1001  
1010

### 05.1 GET DUMMIES

3 COLUMNAS PRINCIPALES





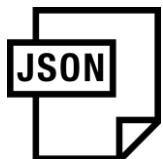
## 05. MACHINE LEARNING



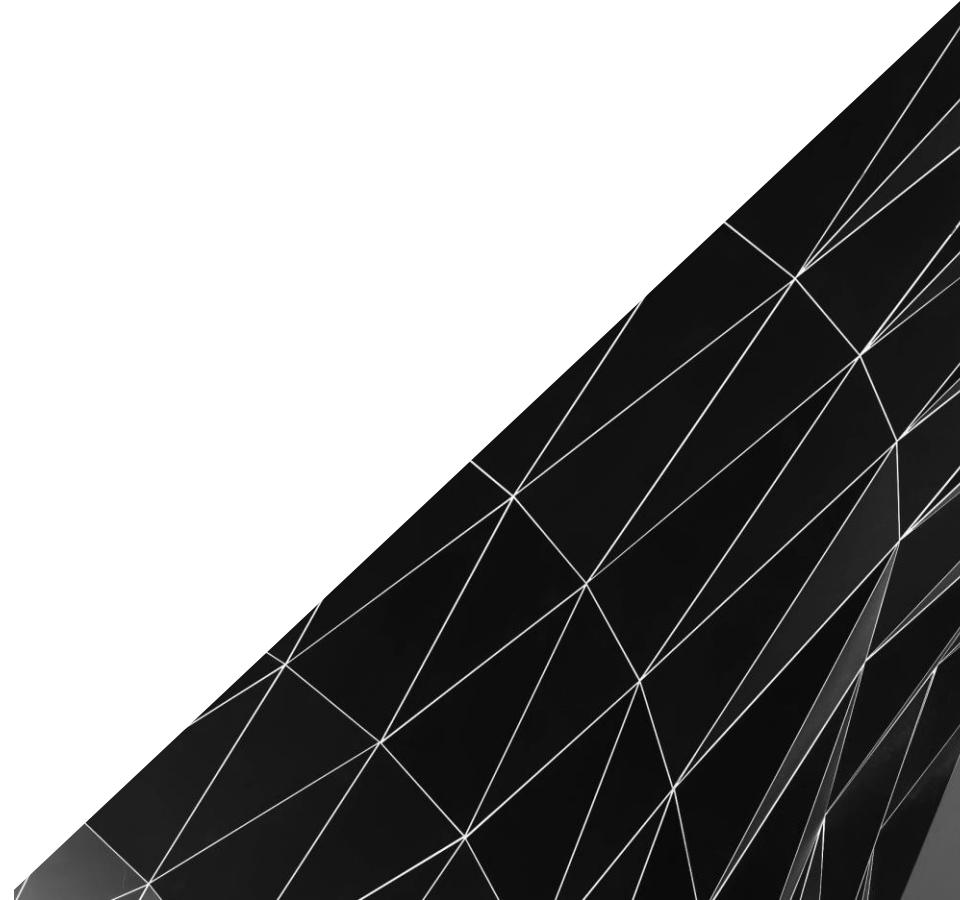
0110  
1001  
1010

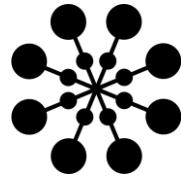
### 05.1 GET DUMMIES

3 COLUMNAS PRINCIPALES



SACAMOS UN JSON CON LA  
POSICIÓN EN LA QUE SE ENCUENTRA  
CADA VARIABLE





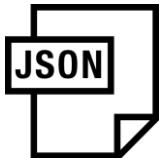
## 05. MACHINE LEARNING



0110  
1001  
1010

### 05.1 GET DUMMIES

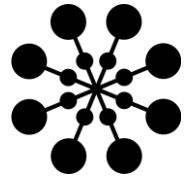
3 COLUMNAS PRINCIPALES



SACAMOS UN JSON CON LA  
POSICIÓN EN LA QUE SE ENCUENTRA  
CADA VARIABLE

**TOTAL = 621 COLUMNAS**

<b>VERDE</b>	<b>1</b>	<b>0</b>	<b>0</b>
<b>AZUL</b>	<b>0</b>	<b>1</b>	<b>0</b>
<b>ROJO</b>	<b>0</b>	<b>0</b>	<b>1</b>



## 05. MACHINE LEARNING



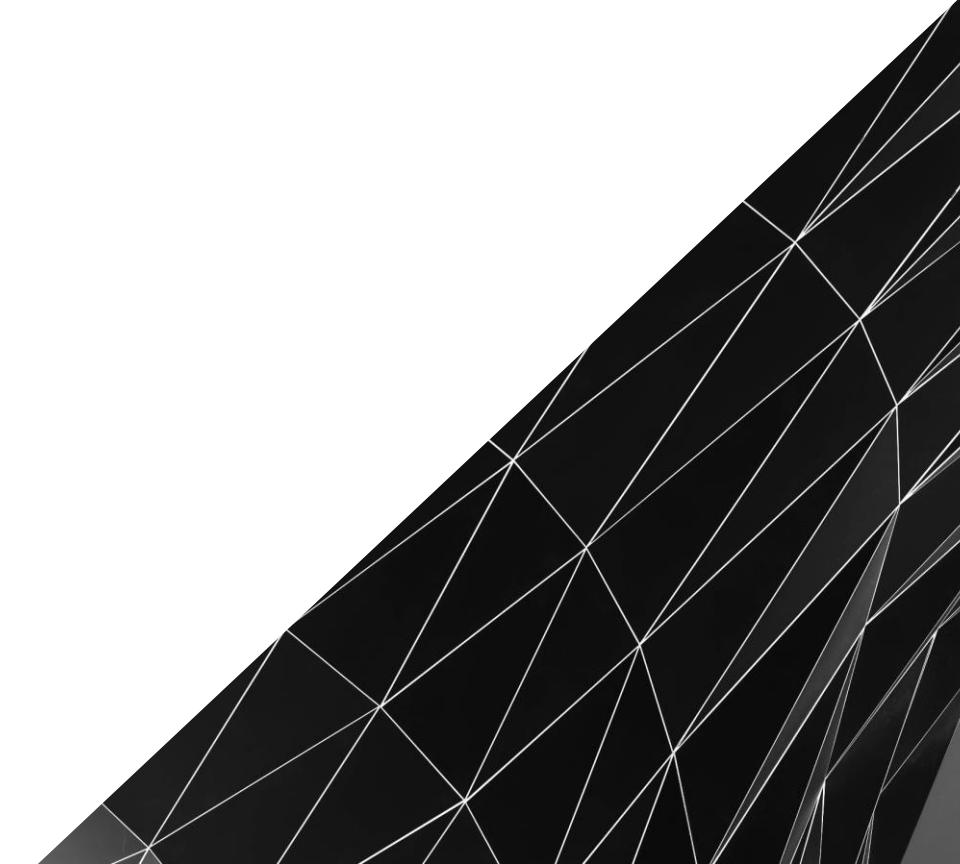
### 05.2 DEFINIMOS X e Y



### 05.3 PCA

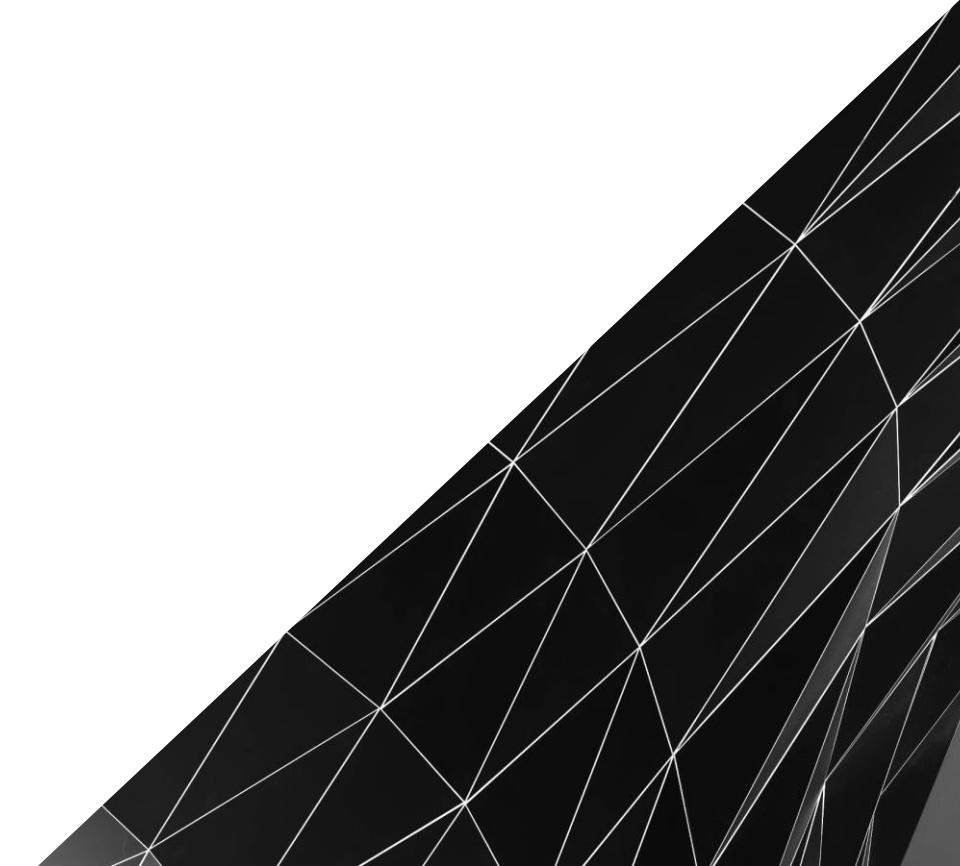


### 05.4 ENTRENAMOS EL MODELO





# 06. FLASK

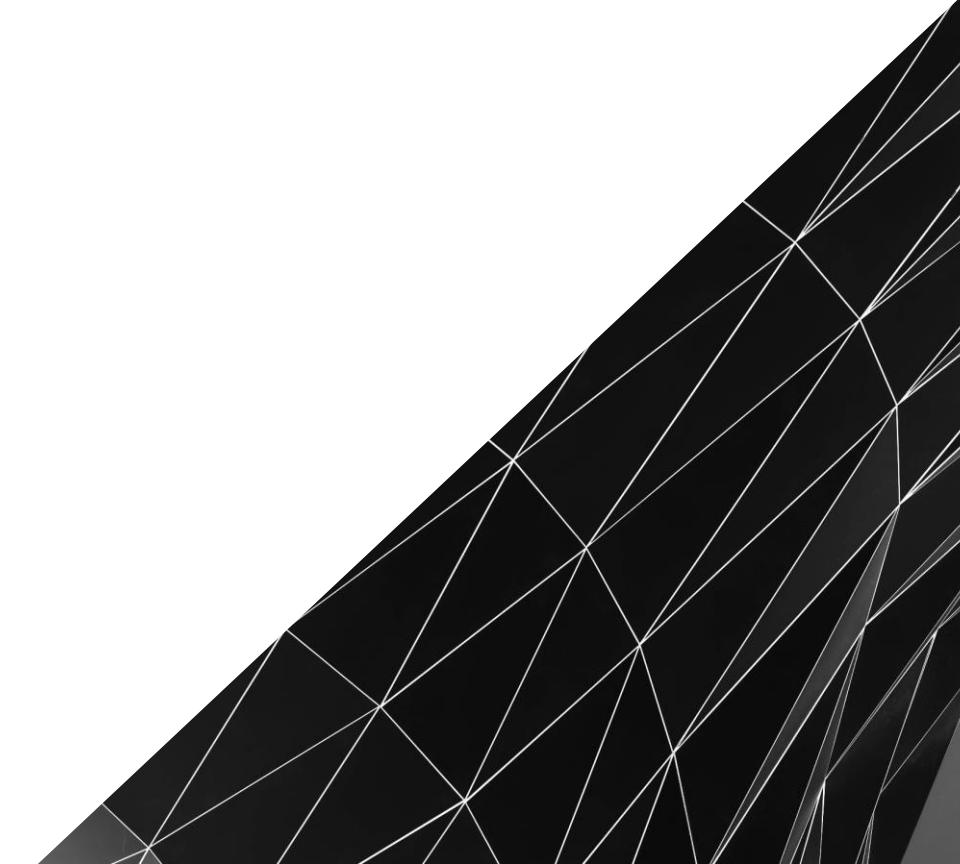




# **06. FLASK**



## **06.1 3 DESPLEGABLES DE OPCIONES**





# 06. FLASK



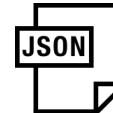
## 06.1 3 DESPLEGABLES DE OPCIONES

CUANDO ELIGE UNA OPCIÓN:

- COMPARA ESA OPCIÓN CON EL JSON
- LO GUARDA EN UN ARRAY



**VS.**





## **06. FLASK**



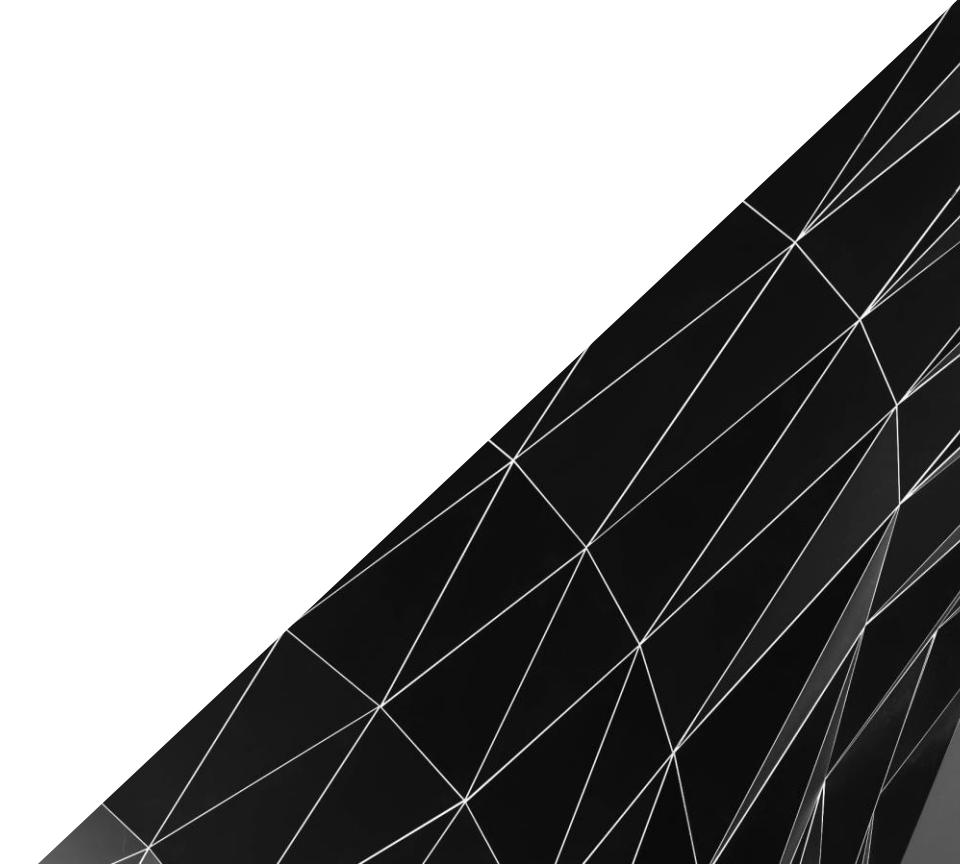
### **06.2 ARRAY DEFINITIVO**



### **06.3 CARGA EL MODELO**



### **05.4 PREDICCIÓN**





# ÚLTIMOS PASOS



# ÚLTIMOS PASOS

La última parte del trabajo se basa en definir los pasos a futuro y las conclusiones.

## FUTUROS PASOS

Data engineer, haciendo **3 grupos** sobre **género** más vendidos, los intermedios y los que menos.

## ¿QUÉ MEJORARÍAS?

Intentaría ponerlo en un **entorno cloud** para poder acceder desde la nube.

## CONCLUSIONES

Hipótesis **confirmada**.

La **columna género** tiene mucho **impacto** para predecir el numero de ventas globales.

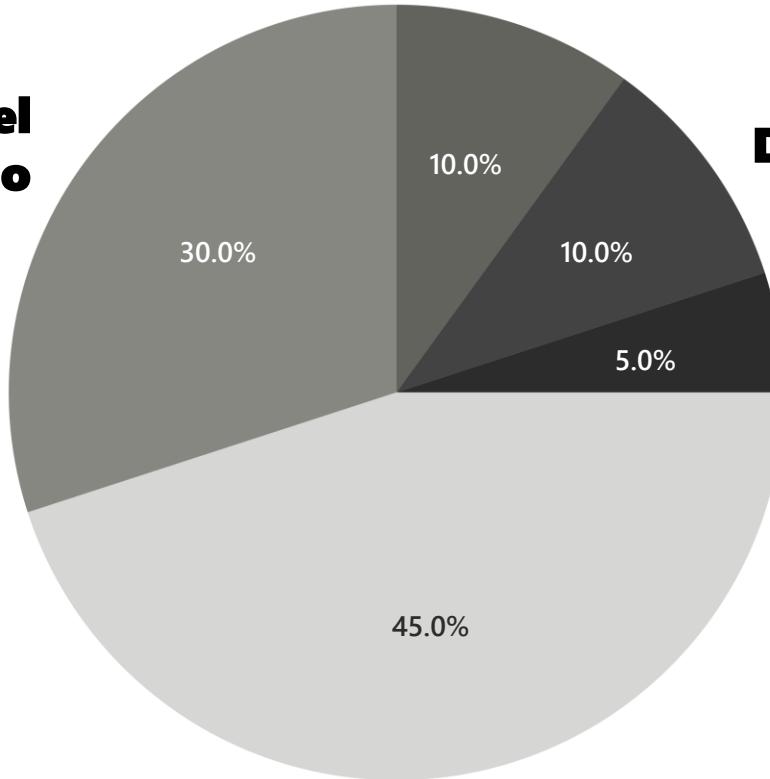


# TIMING PIE CHART



# PIE CHART

**Entrenamiento del modelo**



**Visualización**

**Data Wrangling/Mining**

**Obtener Datos**

**Flask**

# GRACIAS!

Email      **Mariacagigas.s@gmail.com**

LinkedIn      **/mariacagigas**

Github      **@Mariacagi**