Data Translator

Case Study

Equipo 3

José Jiménez Gonzalez

Iván López García

Data Translator - Resultados

Importamos las librerías a utilizar en jupyter y leemos nuestro archivo *"Estudycas/Usvideos.csv"*.

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

%matplotlib inline

from IPython.display import YouTubeVideo, display
from skimage import io #To read img from file
from urllib.error import HTTPError #for mi
```

[2]:
```
# Part1
```

Part 2

[4]:
```
dfv=pd.read_csv ("Estudycase/USvideos.csv")
dfv
```

[4]:

	video_id	trending_date	title	channel_title	category_id	publish_time	tags	views	likes	d
0	2kyS6SvSYSE	17.14.11	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	2017-11-13T17:13:01.000Z	SHANtell martin	748374	57527	
1	1ZAPwfrtAFY	17.14.11	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-11-13T07:30:00.000Z	last week tonight trump presidency|"last week ...	2418783	97185	

Data Translator - Resultados

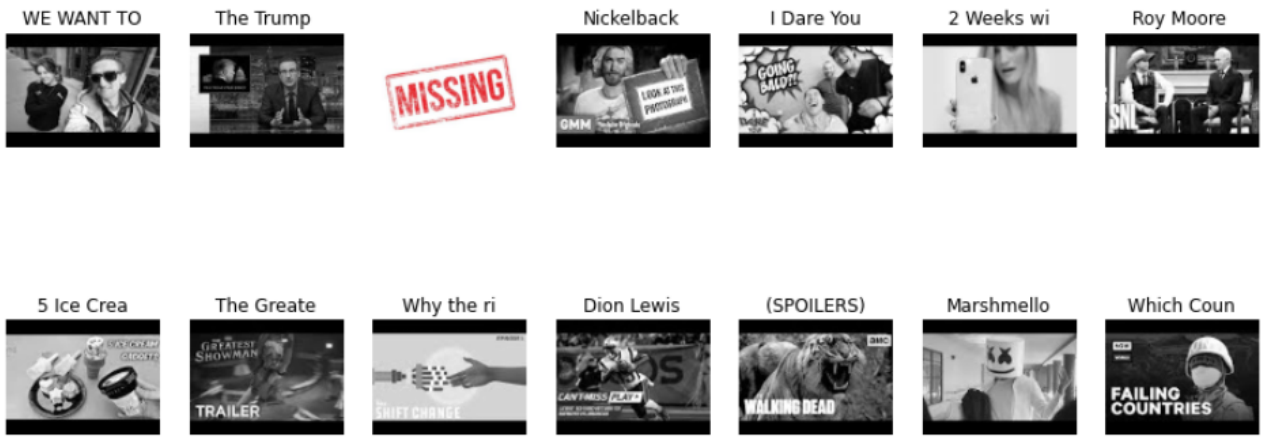
Part 3

[6]:

```
dfimg=dfv[:21]

imag=list(dfimg["thumbnail_link"]) #URL
fig, axs=plt.subplots(3,7, figsize=(15,10))
titles=list(dfimg["title"])

for ax, url, title in zip(axs.flatten(),imag,titles):
    try:
        ax.imshow(io.imread(url, as_gray=True), cmap="gray")
        ax.set_title(title[:10])
        ax.axis("off")
    except:
        ax.imshow(io.imread("http://www.dagdrivarn.se/BILDER/stambilder/MISSING.jpg"))
        ax.axis("off")
```



Data Translator - Resultados

Part 4

[9]:
```
dfv["tags"]
```

[9]:
```
0                                          SHANtell martin
1          last week tonight trump presidency|"last week ...
2          racist superman|"rudy"|"mancuso"|"king"|"bach"...
3          rhett and link|"gmm"|"good mythical morning"|"...
4          ryan|"higa"|"higatv"|"nigahiga"|"i dare you"|"...
                                ...
40944      aarons animals|"aarons"|"animals"|"cat"|"cats"...
40945                                                 [none]
40946      I gave safiya nygaard a perfect hair makeover ...
40947      Black Panther|"HISHE"|"Marvel"|"Infinity War"|...
40948          call of duty|"cod"|"activision"|"Black Ops 4"
Name: tags, Length: 40949, dtype: object
```

[10]:
```
dfv["tag_conteo"]=dfv["tags"].apply(lambda x:x.count("|")+1 if (x!="[none]") else 0)


dfv["tag_conteo"]
```

[10]:
```
0         1
1         4
```

Data Translator - Resultados

Parte 5

[11]:
```
#Part 5
dfv["Likes"]=dfv["likes"]>dfv["dislikes"]

dfv.head()
```

[11]:

:ount	thumbnail_link	comments_disabled	ratings_disabled	video_error_or_removed	description	tag_conteo	Likes
5954	https://i.ytimg.com/vi/2kyS6SvSYSE/default.jpg	False	False	False	SHANTELL'S CHANNEL - https://www.youtube.com/s...	1	True
2703	https://i.ytimg.com/vi/1ZAPwfrtAFY/default.jpg	False	False	False	One year after the presidential election, John...	4	True
8181	https://i.ytimg.com/vi/5qpjK5DgCt4/default.jpg	False	False	False	WATCH MY PREVIOUS VIDEO ▶️ \n\nSUBSCRIBE ► http...	23	True
2146	https://i.ytimg.com/vi/puqaWrEC7tY/default.jpg	False	False	False	Today we find out if Link is a Nickelback amat...	27	True
7518	https://i.ytimg.com/vi/d380meD0W0M/default.jpg	False	False	False	I know it's been a while since we did this sho...	14	True

Data Translator - Resultados

Parte 6

[12]:
```
import json

with open("Estudycase/US_category_id.json","r") as f:data=json.loads(f.read())
data
```

[12]:
```
{'items': [{'kind': 'youtube#videoCategory',
   'etag': '"m2yskBQFythfE4irbTIeOgYYfBU/Xy1mB4_yLrHy_BmKmPBggty2mZQ"',
   'id': '1',
   'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
    'title': 'Film & Animation',
    'assignable': True}},
  {'kind': 'youtube#videoCategory',
   'etag': '"m2yskBQFythfE4irbTIeOgYYfBU/UZ1oLIIz2dxIhO45ZTFR3a3NyTA"',
   'id': '2',
   'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
    'title': 'Autos & Vehicles',
    'assignable': True}},
  {'kind': 'youtube#videoCategory',
   'etag': '"m2yskBQFythfE4irbTIeOgYYfBU/nqRIq97-xe5XRZTxbknKFVe5Lmg"',
   'id': '10',
   'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
    'title': 'Music',
    'assignable': True}},
  {'kind': 'youtube#videoCategory',
   'etag': '"m2yskBQFythfE4irbTIeOgYYfBU/HwXKamM1Q20q9BN-oBJavSGkfDI"',
   'id': '15',
   'snippet': {'channelId': 'UCBR8-60-B28hp2BmDPdntcQ',
    'title': 'Pets & Animals',
```

Data Translator - Resultados

Parte 7

```
#Parte 7 convertir la "CATEGORY_TITLE" a int
dfv["category_id"]=dfv["category_id"].astype(int)
arch2["category_id"]=arch2["category_id"].astype(int)
```

```
dfvarch2=pd.merge(dfv,arch2, how="inner", left_on="category_id", right_on="category_id")
dfvarch2=dfvarch2[['title','channel_title','category_title','views','likes','dislikes','comment_count','description','tag_conteo','Like
dfvarch2
```

	title	channel_title	category_title	views	likes	dislikes	comment_count	description	tag_conteo	Likes
0	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	People & Blogs	748374	57527	2966	15954	SHANTELL'S CHANNEL - https://www.youtube.com/s...	1	True
1	Me-O Cats Commercial	Nobrand	People & Blogs	98966	2486	184	532	Kittens come out of the eggs in a Thai commerc...	4	True
2	AFFAIRS, EX BOYFRIENDS, $18MILLION NET WORTH -...	Shawn Johnson East	People & Blogs	321053	4451	1772	895	Subscribe for weekly videos ▶️ http://bit.ly/sj...	44	True
3	BLIND(folded) CAKE DECORATING CONTEST (with Mo...	Grace Helbig	People & Blogs	197062	7250	217	456	Molly is an god damn amazing human and she cha...	12	True
4	Wearing Online Dollar Store Makeup For A Week	Safiya Nygaard	People & Blogs	2744430	115426	1110	6541	I found this online dollar store called ShopMi...	25	True
...	...	...	...	...	...	...	...	...	...	...

Data Translator - Resultados

Parte 8

```
#Part 8
dfv20=dfv.nlargest(20,"views")

for v,t, in enumerate (dfv20.index):
    print(f"{v+1}: views:{dfv20.views[t]}, title:{dfv20.title[t]}")
    print("-"*80)
```

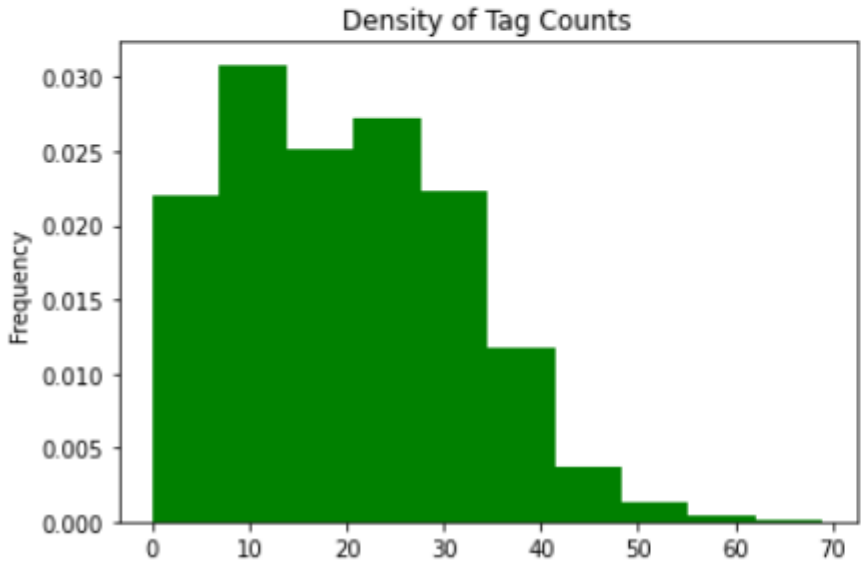
```
1: views:225211923, title:Childish Gambino - This Is America (Official Video)
--------------------------------------------------------------------------------
2: views:220490543, title:Childish Gambino - This Is America (Official Video)
--------------------------------------------------------------------------------
3: views:217750076, title:Childish Gambino - This Is America (Official Video)
--------------------------------------------------------------------------------
4: views:210338856, title:Childish Gambino - This Is America (Official Video)
--------------------------------------------------------------------------------
5: views:205643016, title:Childish Gambino - This Is America (Official Video)
--------------------------------------------------------------------------------
6: views:200820941, title:Childish Gambino - This Is America (Official Video)
--------------------------------------------------------------------------------
7: views:196222618, title:Childish Gambino - This Is America (Official Video)
--------------------------------------------------------------------------------
8: views:190950401, title:Childish Gambino - This Is America (Official Video)
--------------------------------------------------------------------------------
9: views:184446490, title:Childish Gambino - This Is America (Official Video)
--------------------------------------------------------------------------------
```

Data Translator - Resultados

Parte 9

```
fig, ax=plt.subplots()
ax.hist(dfvarch2["tag_conteo"], color="green", density=True) #Density es la densidad de probabilidad
ax.set_title('Density of Tag Counts')
ax.set_ylabel('Frequency')
```

```
Text(0, 0.5, 'Frequency')
```



Data Translator - Resultados

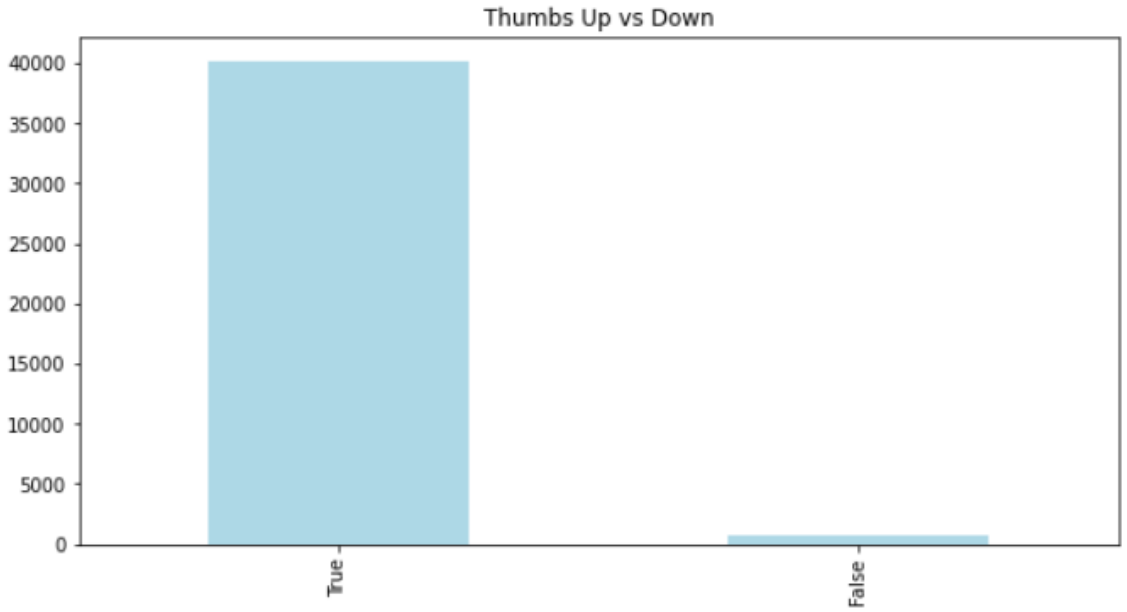
Parte10

```
df4=dfvarch2["Likes"].value_counts()
df4
```

```
True     40192
False      757
Name: Likes, dtype: int64
```

```
dfvarch2["Likes"].value_counts().plot(kind="bar",color='lightblue',figsize=(10,5)).set_title('Thumbs Up vs Down')
```

```
Text(0.5, 1.0, 'Thumbs Up vs Down')
```

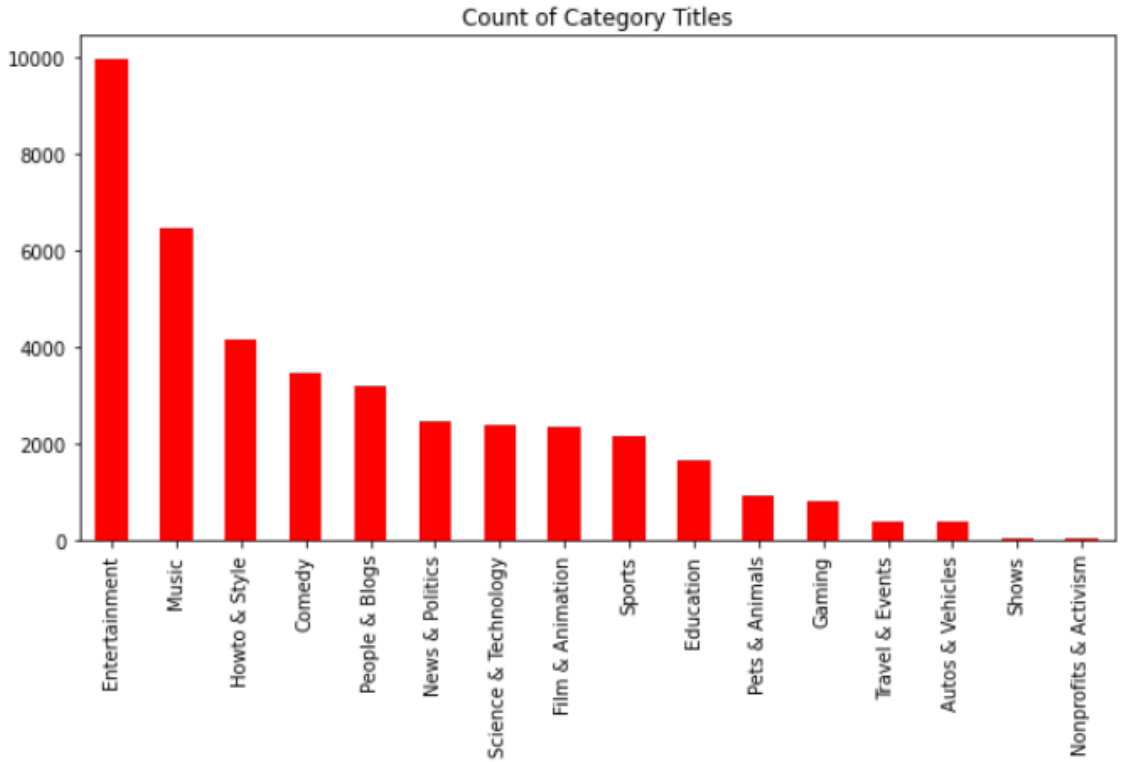


Data Translator - Resultados

Parte 11

```
dfvarch2["category_title"].value_counts().plot.bar(color="red",figsize=(10,5)).set_title('Count of Category Titles')
```

```
Text(0.5, 1.0, 'Count of Category Titles')
```



Data Translator - Resultados

Parte 12

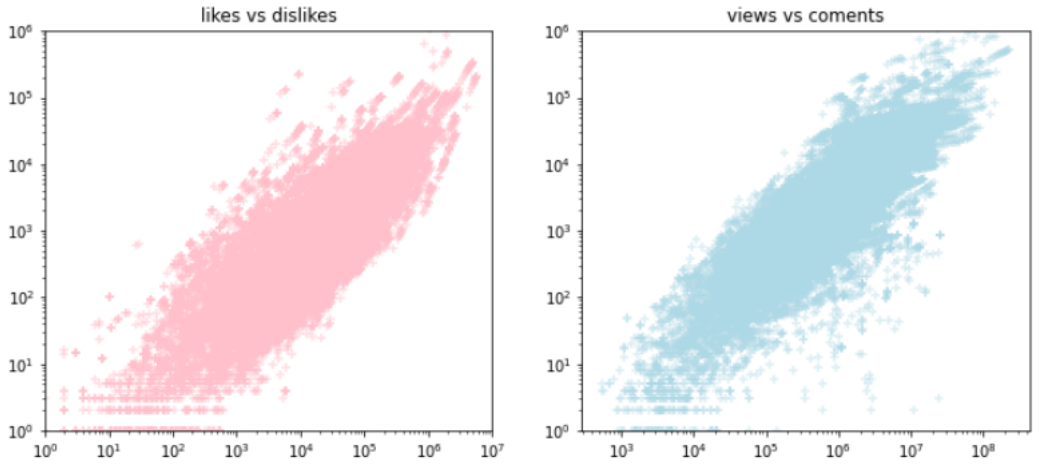
```
fig, axs = plt.subplots(1,2,figsize=(12,5))

axs[0].scatter(dfvarch2["likes"],dfvarch2["dislikes"],alpha=0.5,color='pink', marker="+")
axs[0].set_title('likes vs dislikes')
axs[0].set_xscale('log')
axs[0].set_yscale('log')

axs[1].scatter(dfvarch2["views"],dfvarch2["comment_count"],alpha=0.4,color='lightblue',marker="+")
axs[1].set_title('views vs coments')
axs[1].set_xscale('log')
axs[1].set_yscale('log')

axs[0].set_xlim(1,10**7)
axs[0].set_ylim(1,10**6)
axs[1].set_ylim(1,10**6)
#fig.tight_layout()
```

```
(1, 1000000)
```




```
dfvarch2.to_csv("dfvarch2.csv")
```

Data Translator

Case Study Dashboard

Data Translator - Resultados

Importamos las librerías a utilizar y leemos nuestro nuevo archivo de data frame que previamente habíamos creado.

```
1  import dash
2  from dash import dcc
3  from dash import html
4  from dash.dependencies import Input, Output
5  import plotly.express as px
6  import pandas as pd
7  import numpy as np
8
9  df=pd.read_csv('dfvarch2.csv')
```

Data Translator - Resultados

Declaramos la aplicación y posteriormente, en la primera fila agregamos el tag H1 que contiene el nombre del gráfico que se mostrará, H2 el nombre del dashboard y finalmente H3 contiene los nombres de los integrantes del equipo.

```
app=dash.Dash(__name__)

app.layout=html.Div([

    #first row
    html.Div(children=[
        html.H1(children='Gráfico de Likes vs Dislikes',style={'color':'skyblue'}),
        html.H2(children='Case Study, Data Translator',style={'color':'skyblue'}),
        html.H3(children='José Jiménez González & Iván López García',style={'color':'skyblue'}),
    ]),
```

Data Translator - Resultados

En la segunda fila se tienen dos subprocesos, en el primero se agrega el link de la imagen que esta alojada en internet y se mostrará en el dashboard y la escala con la que se ajustara .

```
#second row
html.Div([
    #first col of second row
    html.Div([
        html.Img(
            src='https://upload.wikimedia.org/wikipedia/commons/thumb/5/54/Bot%C3%B3n_Me_gusta.svg/1200px-Bot%C3%B3n_Me_gusta.svg.png',
            style={'width':'50%'}),
     ],style={'display':'inline-block','vertical-align':'top','margin-left':'3vw','margin-top':'3vw'}),
    ],style={'display':'inline-block','width':'30%','vertical-align':'top'}),
```

Data Translator - Resultados

En el segundo proceso de la segunda fila, en el menú de Dropdown se configuran los ejes del gráfico, a partir de la columna del data frame “category_title” y se deja por default preseleccionado la categoría “*Entertainment”.* Y se establece el ID de la gráfica como *“likes_plot*”

```
    #second col of second row
        html.Div([
            dcc.Dropdown(
                id='category-dropdown',
                options=[{'label':i,'value':i} for i in np.sort(df['category_title'].unique())],
                value='Entertainment'
                ),
            dcc.Graph(id='likes_plot')
        ],style={'display':'inline-block','width':'65%','vertical-align':'top'})#,style={'display':'inline-block','vertical-align':'top','margin-left':'3vw','margin-top':'3vw'})
    ])
])
```

Data Translator - Resultados

Se agrega 1) el callback de la gráfica con *input* y *output*, 2) la función para actualizar la figura en función de la categoría que se seleccione.

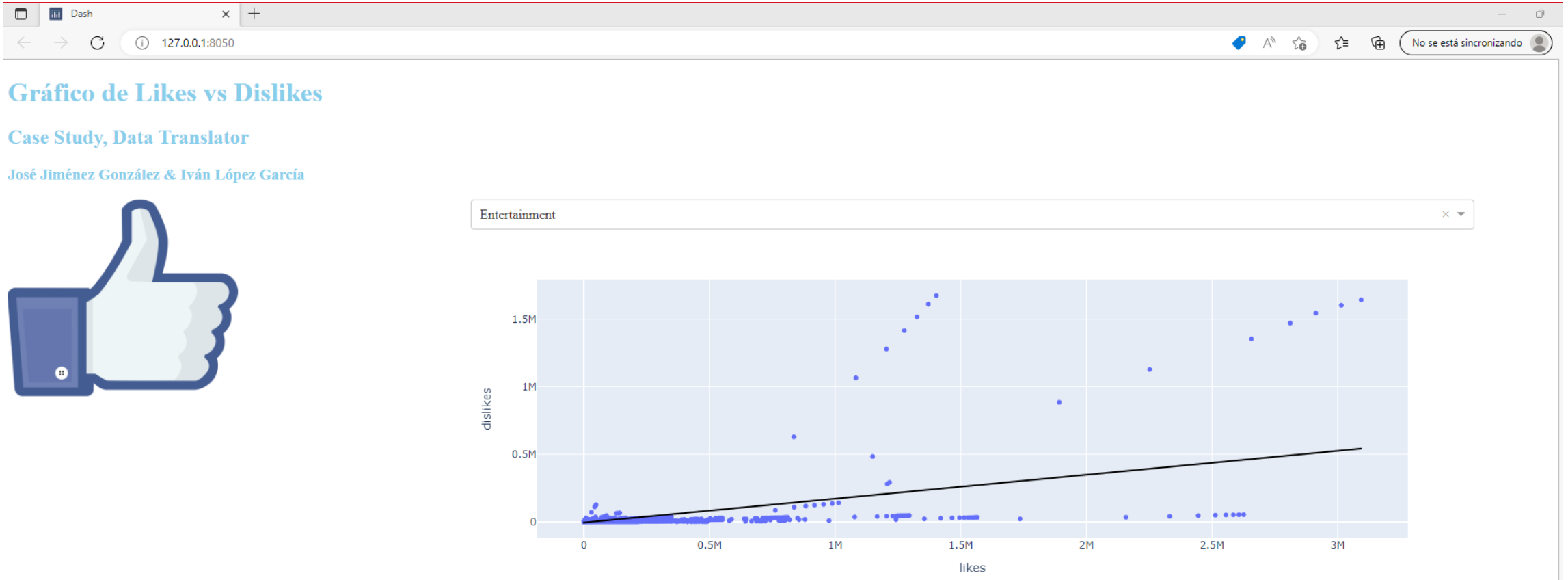
```
@app.callback(
    Output('likes_plot','figure'),
    Input('category-dropdown','value')
)

def update_figure(selected_category):
    filtered_df=df[df.category_title==selected_category]
    fig=px.scatter(filtered_df,x='likes',y='dislikes',trendline='ols',trendline_color_override='black')
    fig.update_layout(transition_duration=400)
    return fig

if __name__=='__main__':
    app.run_server(debug=True)
```

Data Translator - Resultados

Dashboard con la selección predeterminada



Data Translator - Resultados

Dashboard con lista desplegable y previamente seleccionado la categoría *Music*

¡Gracias!