

Reporte de práctica 4: Visualización de información con Plotly

Plotly, también conocida por su URL, [Plot.ly \(https://plot.ly/\)](https://plot.ly), es una empresa de computación técnica con sede en Montreal, Quebec, que desarrolla herramientas de visualización y análisis de datos en línea. Plotly proporciona herramientas de gráficos, análisis y estadísticas en línea para individuos y colaboración, así como bibliotecas de gráficos científicos para Python, R, MATLAB, Perl, Julia, Arduino y REST.

Objetivos

- Incluir al menos tres diferentes tipos de gráficos usando Plotly e intentar concluir algo sobre los datos.

Los datos

Debido a que la base de datos con la que contamos se encuentra fragmentada en años, se utilizarán los datos que cuente con la información más completa.

Importando datos

```
In [1]: import pandas as pd
df2018 = pd.read_excel('https://raw.githubusercontent.com/SamatarouKami/CIENCIA_DE_DATOS/master/2018.xlsx', index_col=None, header=0, sheet_name=0)
df2018 = df2018[['Categoria', 'Edad', 'Pais', 'Titulo', 'Genero', 'Duracion', 'Marca', 'Referencia', 'Dias', 'Marcas', 'Personas']]
print(len(df2018))
```

452

Gráficas

Una vez cargada la información, se procede a graficar y verificar si se puede concluir algo relevante.

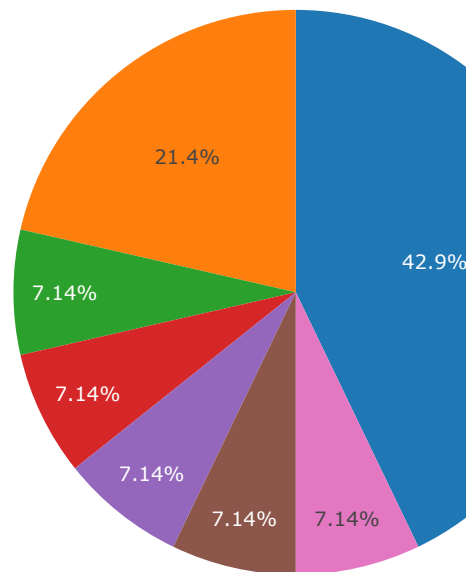
Extranjeros

Se busca saber cual es la proporción de extranjeros que participan en el festival de acuerdo a su país de procedencia. Se retiran a los Colombianos de la lista de participantes para obtener solo extranjeros. Se realiza una categorización por país, luego se hace un conteo y graficamos.

```
In [2]: import plotly.plotly as py
import plotly.figure_factory as ff
import statsmodels.api as sm
import numpy as np
import plotly.graph_objs as go
from numpy import isnan

cine = df2018
países = cine['País'].unique()
in1 = np.argwhere(países=='Colombia')
países = np.delete(países,in1)
listaPaíses = []
for country in países:
    listaPaíses.append([country,cine[cine['País']==country].count()['País']])
dfpais = pd.DataFrame(data=listaPaíses)
dfpais.columns = ['País', 'Total']
trace = go.Pie(labels=dfpais.País, values=dfpais.Total)
py.iplot([trace], filename='P4_1')
```

Out[2]:

[EDIT CHART](#)

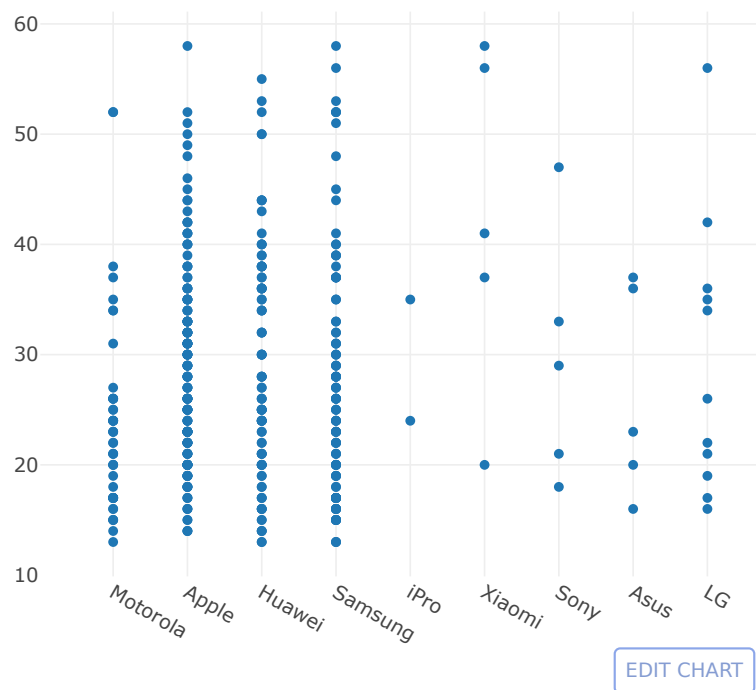
Se puede concluir que en el año 2018, la mayor parte de extranjeros procedía de España.

Marca de smartphones

Ahora se comprueba los rangos de edades donde los participantes utilizan cierta marca de celular.

```
In [3]: datos = [go.Scatter(x = df2018.Marca, y = df2018.Edad, mode = 'markers')]  
py.iplot(datos, filename='P4_2')
```

Out[3]:



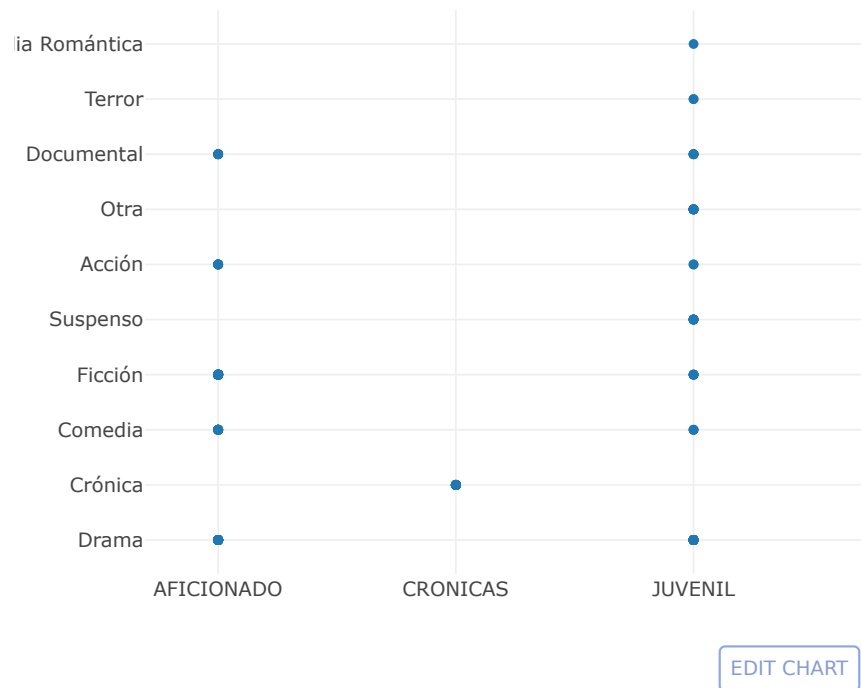
Se puede concluir que la mayoría de los concursantes utiliza smartphones de la marca Apple, Samsung y Huawei.

Género por categoría

Ahora se verifica si hay alguna relación entre los géneros seleccionados y la categoría de participación.

```
In [4]: datos = [go.Scatter(x = df2018.Categoria, y = df2018.Genero, mode = 'markers')]  
py.iplot(datos, filename='P4_3')
```

Out[4]:



Se puede concluir en los géneros de Ficción, Comedia, Drama, Acción y Documental existen participantes de todas las categorías, excepto Crónicas que parece ser un género únicamente relacionado con la categoría.

Conclusión

En esta práctica pude aprender a utilizar la plataforma de Plotly. Se pudo determinar tres cualidades sobre nuestros datos simplemente graficando la información.

--04 de junio 2019-- Luis Angel Gutiérrez Rodríguez 1484412