

INF379

Visualización de datos

Profesor: José Luis Martí Lara
Ayudante: Michell Concha Gaete
Universidad Técnica Federico Santa María
Departamento de Informática – Abril 03, 2023.

Nombre: Martín Ruiz

Rol: 201973093-2 **Paralelo:** 200

Nombre: Eduardo Pino

Rol: 201973613-2 **Paralelo:** 200

Tarea 1

Para ver el código ejecutado en Jupyter revisar lo adjuntado en el anexo.

1. Parte 1: Empresas

Las empresas que se eligieron para la tarea son: Apple y Amazon.
Cada una será mejor que la otra en algún criterio definido.

1.1. Apple

El criterio seleccionado para evaluar Apple y Amazon será la evolución del precio de las acciones con respecto a un intervalo de tiempo determinado, el cual será en el que se levantó y bajo la cuarentena en Chile, donde comenzó a principios de marzo del año 2020, y como bajada a principios de marzo del año 2022.

- Datos recolectados (ver anexo):

A partir de los datos obtenidos es posible notar que Apple con el paso de los meses, solo van aumentando el precio de sus acciones, y si había alguna caída del precio, este iba a ser algo mínimo, por el lado de Amazon, solo logró mantenerse e ir oscilando su precio con respecto a los 160 dólares.

- Visualización propuesta:

Utilizaremos la visualización Candlestick de la librería Plotly, donde vamos a tener 2 tipos de marcas, por una parte vamos a tener cada uno de los Candlestick por cada mes, que representaría si hubo una caída o alza del precio de la acción, por otra parte, una línea la cual irá recorriendo por cada uno de los precios al momento de cerrar el precio de la acción.

Los canales por su parte, corresponden al eje x tendremos una variable de tipo string, la cual representará el mes y año del precio de la acción, y en el eje y , tendremos el precio de

las acciones en dólares, donde los gráficos que resultaron fueron las siguientes:

Historical prices: Apple



Figura 1: Precios historicos de las acciones de Apple

Historical prices: Amazon

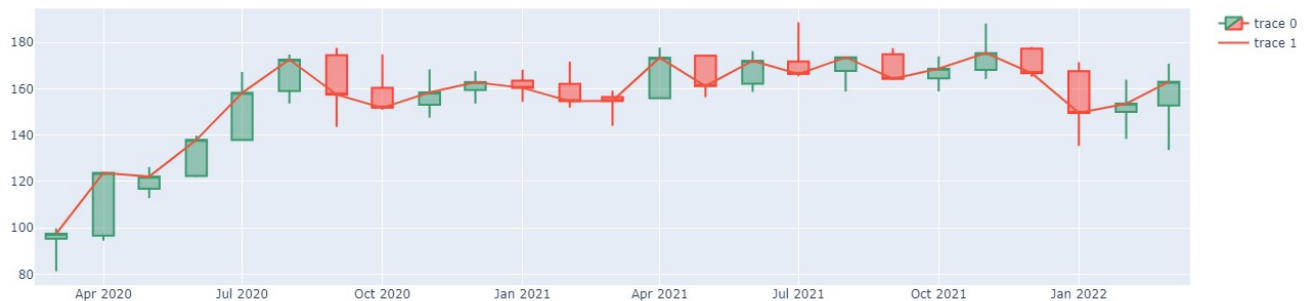


Figura 2: Precios historicos de las acciones de Amazon

- Resultados:

Del gráfico anterior es posible notar que Apple, en el momento en que comenzó la cuarentena en Chile (y en muchas partes del mundo) estaba a 60 dólares aproximadamente, y fue creciendo paulatinamente su precio hasta que se “terminó” la cuarentena en Chile, llegando a valer 180 dólares (habiendo triplicado su precio), por otro lado, se tiene a Amazon que comenzó en los 100 dólares aproximadamente, llegando a los 160 dólares, donde se quedó oscilando hasta el “terminó” de la cuarentena.

- Código:

```
df_aapl = pd.read_csv('https://raw.githubusercontent.com/PinoEdu/Tarea1-Visualizacion-Datos/master/AAPL.csv')

fig_aapl = go.Figure(data=[go.Candlestick(x=df_aapl['Date'],
                                           open=df_aapl['Open'], high=df_aapl['High'],
                                           low=df_aapl['Low'], close=df_aapl['Close'])
                           ])

fig_aapl.add_trace(
    go.Scatter(
        x=df_aapl['Date'].to_list(),
        y=df_aapl['Close'].to_list()
    )
)

fig_aapl.update_layout(xaxis_rangeslider_visible=False,
                       title = 'Historical prices: Apple',
                       )
fig_aapl.show()
```

Python

Figura 3: Precios historicos de las acciones de Apple - Código

```
df_amzn = pd.read_csv('https://raw.githubusercontent.com/PinoEdu/Tarea1-Visualizacion-Datos/master/AMZN.csv')

fig_amzn = go.Figure(data=[go.Candlestick(x=df_amzn['Date'],
                                           open=df_amzn['Open'], high=df_amzn['High'],
                                           low=df_amzn['Low'], close=df_amzn['Close'])
                           ])

fig_amzn.add_trace(
    go.Scatter(
        x=df_amzn['Date'].to_list(),
        y=df_amzn['Close'].to_list()
    )
)

fig_amzn.update_layout(xaxis_rangeslider_visible=False,
                       title = 'Historical prices: Amazon',
                       )
fig_amzn.show()
```

Python

Figura 4: Precios historicos de las acciones de Amazon - Código

1.2. Amazon

El criterio utilizado para realizar esta visualización es el de la cantidad de empleados por año de Amazon y Apple desde el 2000 hasta el 2022.

- Datos recolectados (ver anexo):

En este caso se utiliza un dataset propio, que se compone de dos datasets proporcionados por el sitio [Trading Economics](#) que contienen información histórica de la cantidad de empleados de dichas empresas.

- Visualización propuesta:

Se utiliza un Heatmap de la librería Seaborn. Los canales para este caso son en el eje horizontal el tiempo, este va avanzando cada 5 años y en el eje vertical están las compañías a visualizar. Cada celda contiene el número de empleados de ese año de la respectiva empresa en miles de personas, las marcas en este caso son las celdas con su respectivo color. Al costado derecho se observa una franja que indica la intensidad del color que toma la celda, a medida que aumenta la cantidad de empleados. El gráfico obtenido es el siguiente:

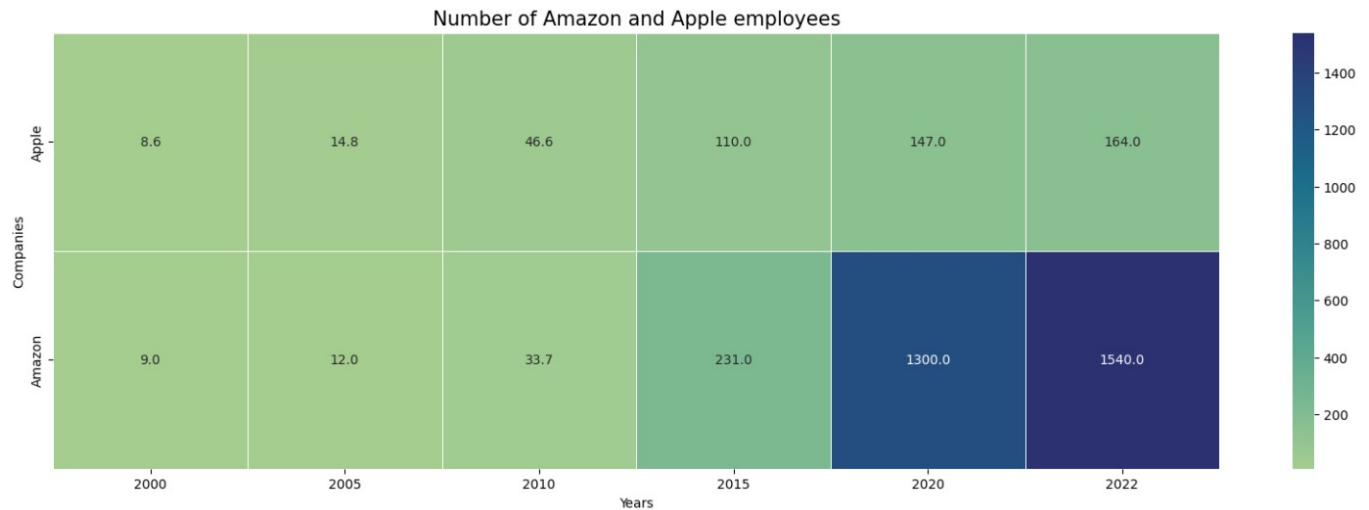


Figura 5: Número de empleados Amazon y Apple

- Resultados:

Se puede apreciar que en Amazon el número de empleados en los años 2020 y 2022 aumenta exponencialmente, esto puede ser debido al aumento de la demanda de productos online debido al confinamiento generado por la pandemia del virus Covid-19, mientras que Apple muestra un crecimiento constante de empleados.

- Código:

```
df = pd.read_csv('https://raw.githubusercontent.com/PinoEdu/Tarea1-Visualizacion-Datos/master/Number_employees_2000-2020.csv', index_col=0)
f, ax = plt.subplots(figsize=(20, 6))

sns.heatmap(df.T, cmap="crest", annot=True, fmt=".1f", linewidth=.5)
ax.set_title('Number of Amazon and Apple employees', fontsize = 15)
plt.xlabel('Years', fontsize = 10)
plt.ylabel('Companies', fontsize = 10)

plt.show()
```

Figura 6: Número de empleados - Código

2. Parte 2: Redes sociales

Las redes sociales que se eligieron para la tarea son: Facebook y Twitter. Cada una será mejor que la otra en algún criterio definido.

2.1. Twitter

El criterio usado para esta visualización es el carácter emocional que tienen los post realizados en Twitter y Facebook. Para esta visualización se utilizan emociones positivas, negativas y neutrales, donde se considera como emocionales las de carácter positivas y negativas.

- Datos recolectados (ver anexo):

El dataset utilizado es una composición de dos datasets, cada uno de estos contiene 10.000 post realizados Facebook o Twitter y una emoción atribuida a cada post. El dataset de Facebook fue extraído de la página -

- Visualización propuesta:

Para la visualización se usa un Polar Chart de la librería Plotly. La marca a la figura geométrica, que indica el área generada por la cantidad del tipo de emoción para cada post. Por otro lado, los canales de esta visualización son el color de la figura, que nos permite identificar a qué red social se refiere, y el nivel de los puntos, que nos indica la cantidad de mensajes para una emoción correspondiente. El gráfico obtenido es el siguiente:

Sentiment of people expressed in social networks

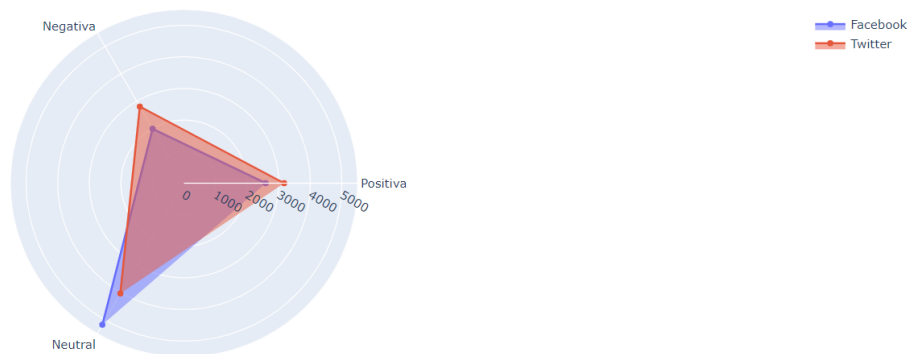


Figura 7: Número de empleados Amazon y Apple

- Resultados:

Se puede concluir a través de este gráfico que los usuarios de Twitter tienden a postear mensajes que contienen un carácter más emocional, mientras que los post de los usuarios de Facebook tienden a tener un carácter más neutral.

- Código:

```

df = pd.read_csv('https://raw.githubusercontent.com/PinoEdu/Tarea1-Visualizacion-Datos/master/sentiment_dataset.csv', index_col=0)

categories = ['Positiva', 'Negativa', 'Neutral']

fig = go.Figure()

fig.add_trace(go.Scatterpolar(
    r=[df['Positiva']['Facebook'], df['Negativa']['Facebook'], df['Neutral']['Facebook']],
    theta=categories,
    fill='toself',
    name='Facebook'
)))

fig.add_trace(go.Scatterpolar(
    r=[df['Positiva']['Twitter'], df['Negativa']['Twitter'], df['Neutral']['Twitter']],
    theta=categories,
    fill='toself',
    name='Twitter'
)))

fig.update_layout(
    title = 'Sentiment of people expressed in social networks',
    autosize=True,
    polar=dict(
        radialaxis=dict(
            visible=True,
            range=[0, 5500]
        ),
    ),
)

fig.show()

```

Figura 8: Número de empleados - Código

2.2. Facebook

El criterio seleccionado en la que Facebook le gana a Twitter es en la cantidad de usuarios activos que tiene cada uno, este análisis lo llevaremos a las 50 ciudades más pobladas de España (aproximadamente 17.4 millones de personas).

- Datos recolectados (ver anexo): Se tomó la información desde EPData. Donde los datos recolectados muestran que Facebook desde el 2016 hasta el 2020 siempre fueron superiores a las demás redes sociales.

- Visualización propuesta:

Se utilizó la gráfica Funnel de la librería Plotly, donde vamos a tener una marca que será encargado de decirnos como se va comportando Funnel, si se encuentra en el mismo nivel, con respecto a la de cercanía a la pantalla con otra valor, nos dirá que tienen el mismo valor, y si está más profundo en la pantalla, nos dirá que disminuyó.

Los canales por su parte, corresponde a un eje x sin definir sus valores, ya que este solo aumentará o disminuirá en proporción al total que ocupe dentro del gráfico, y en un eje y , que nos dirá los años asociados a cada valor.

- Resultados: En la gráfica se puede observar que Facebook desde el 2016 hasta el 2020 siempre fue superior en la cantidad de usuarios activos, donde se puede [extrapolar](#) hasta la actualidad que Facebook seguirá siendo superior a Twitter, pero no con la misma cantidad de usuarios activos.

- Código:

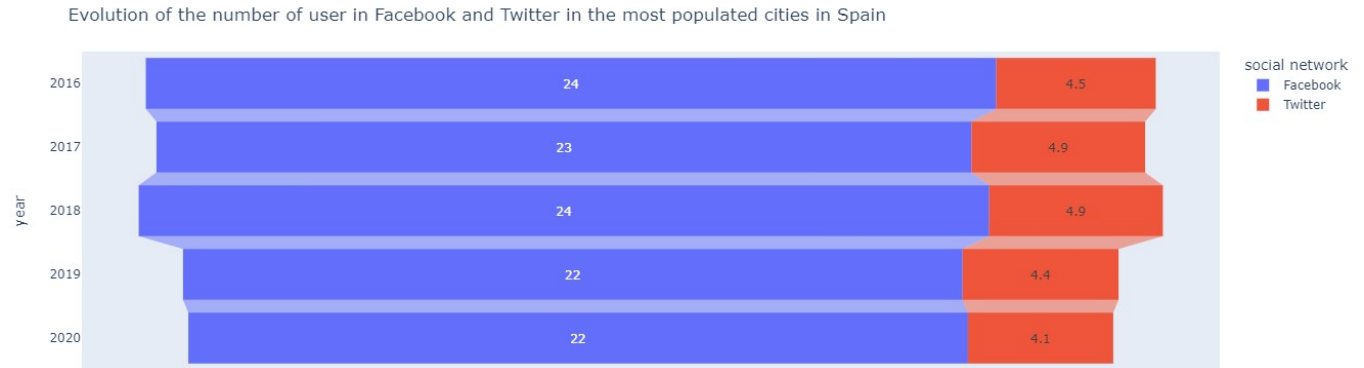


Figura 9: Evolución del número de usuarios en Facebook y Twitter

```
df = pd.read_csv('https://raw.githubusercontent.com/PinoEdu/Tarea1-Visualizacion-Datos/master/evolution-of-numbers.csv')

df_fb = pd.DataFrame(dict(number=df['Facebook'].to_list(), year=map(lambda x: str(x), df['Año'].to_list())))
df_fb['social network'] = 'Facebook'

df_tw = pd.DataFrame(dict(number=df['Twitter'].to_list(), year=map(lambda x: str(x), df['Año'].to_list())))
df_tw['social network'] = 'Twitter'

df = pd.concat([df_fb, df_tw], axis=0)
fig = px.funnel(df, x='number', y='year', color='social network')

fig.update_layout(
    title = 'Evolution of the number of user in Facebook and Twitter in the most populated cities in Spain',
)

fig.show()
```

Figura 10: Evolución del número de usuarios en Facebook y Twitter - Código

3. Anexo

- Apple - Historical prices
- Amazon - Historical prices
- Amazon - Employees Total Number
- Apple - Employees Total Number
- Twitter sentiment dataset
- 50 ciudades más pobladas de España
- Redes sociales con más usuarios en el mundo (2023)
- Facebook sentiment dataset
- Visualización de los códigos y gráficos (Google Colab)