

# Projeto Final Programação e Algoritmos II



Motivation.txt

```
(Docente () {  
João Orvalho();  
}();  
(Discentes ()  
José Pedro Gomes;  
Rui Carreira();  
}();  
#Quote #Programming #Selfcare
```

# *índice*

INTRODUÇÃO

ANÁLISE DE DADOS

DICIONÁRIO DE DADOS

CÓDIGO

CONCLUSÃO

# *introdução*

O presente trabalho foi realizado no âmbito da unidade curricular de Programação e Algoritmos II, lecionada pelo docente João Orvalho, da Licenciatura de Comunicação e Design Multimédia. Foi-nos proposto pelo docente, explorar e adquirir conhecimentos de conjuntos de dados. Posto isto, decidimos avaliar o top 1000 de influencers no TikTok. A razão pela qual escolhemos este tema para o trabalho, deve-se ao facto do TikTok ser uma das redes sociais mais utilizadas no momento e com curiosidade decidimos estudar alguns dados da aplicação. Para iniciar o nosso trabalho, começámos por fazer a pesquisa dos dados. Procuramos por vários websites e páginas e acabamos por escolher os dados que são apresentados pelo Kaggle. Optamos por usar os dados de 2022, pois no site estes já estavam apresentados de forma clara e fácil para usar. Este site já continha o ficheiro csv, o que agilizou um pouco o processo, pois apenas tivemos de aperfeiçoar algumas partes. Pensamos que seja um tema importante para conseguir analisar a quantidade de utilizadores que existem no TikTok, conseguindo assim perceber quantos subscritores, likes, comentários, partilhas e visualizações existem. Os dados para este trabalho foram obtidos através do Kaggle, uma fonte confiável que disponibiliza dados de forma gratuita e ética, visível em:

<https://www.kaggle.com/datasets/prasertk/top-1000-tiktok-influencers-ranking>

# dicionário de dados

Country	Rank	Account	Title	Link	Subscriberscount	Viewsavg	Likesavg	Commentsavg	Sharesavg
All	1.0	billieeilish	BILLIE EILISH	<a href="https://www.tiktok.com/@billieeilish">https://www.tiktok.com/@billieeilish</a>	41600000.0	95000000.0	18100000.0	151000.0	54800.0
All	2.0	badbunny	Bad Bunny	<a href="https://www.tiktok.com/@badbunny">https://www.tiktok.com/@badbunny</a>	13400000.0	33300000.0	7300000.0	101900.0	106800.0
All	3.0	chipmunksoftiktok	Chipmunks of TikTok	<a href="https://www.tiktok.com/@chipmunksoftiktok">https://www.tiktok.com/@chipmunksoftiktok</a>	11400000.0	25000000.0	3000000.0	21800.0	64200.0
All	4.0	charlidamelio	charli d'amelio	<a href="https://www.tiktok.com/@charlidamelio">https://www.tiktok.com/@charlidamelio</a>	135400000.0	18700000.0	2600000.0	54700.0	35200.0
All	5.0	yessicadcs	Yessica y Moises 🇵🇷	<a href="https://www.tiktok.com/@yessicadcs">https://www.tiktok.com/@yessicadcs</a>	263700.0	16700000.0	1400000.0	16200.0	138400.0
...	...	...	...	...	...	...	...	...	...
All	996.0	leleburnier	LELÉ BURNIER	<a href="https://www.tiktok.com/@leleburnier">https://www.tiktok.com/@leleburnier</a>	1100000.0	1200000.0	284300.0	2800.0	968.0
All	997.0	sofiadelmonstro	Cat Sofia	<a href="https://www.tiktok.com/@sofiadelmonstro">https://www.tiktok.com/@sofiadelmonstro</a>	5200000.0	2600000.0	161200.0	1300.0	1600.0
All	998.0	jena	Jena	<a href="https://www.tiktok.com/@jena">https://www.tiktok.com/@jena</a>	10500000.0	2800000.0	320300.0	856.0	512.0
All	999.0	christiandaloi_	Christiandaloi_	<a href="https://www.tiktok.com/@christiandaloi_">https://www.tiktok.com/@christiandaloi_</a>	2700000.0	1100000.0	188600.0	3000.0	2900.0
All	1000.0	davidgetial	deivis	<a href="https://www.tiktok.com/@davidgetial">https://www.tiktok.com/@davidgetial</a>	13200000.0	2300000.0	337500.0	1400.0	212.0

**Country** - Países onde são vistos os influencers;

**Rank** - Lugar que os Influencers ocupam na tabela;

**Account** - Nome da conta do Influencer;

**Title** - Nome original do Influencer;

**Subscriberscount** - Número de subscritores;

**Viewsavg** - Média de visualizações;

**Likesavg** - Média de Likes;

**Commentsavg** - Média de Comentários;

**Sharesavg** - Média de Partilhas;

# código

```
In [21]: import pandas as pd #importar os pandas como pd
import matplotlib.pyplot as plt #importar o matplotlib.pyplot para os gráficos como plt
import numpy as np #importar os numpy como np
```

Nestas primeiras linhas de código são importadas três bibliotecas que irão ser utilizadas no restante código, na primeira linha importámos a biblioteca pandas como pd, na segunda a biblioteca matplotlib.pyplot como plt, e na terceira numpy como np

## Top 1000 Influencers TikTok

```
In [192]: top = pd.read_csv("tiktok_top_1000.csv") #ler o primeiro conjunto de dados através dos pandas
top.columns = top.columns.str.replace(' ', '') #esta linha serve para retirar os espaços dos títulos das colunas
top
```

```
Out[192]:
```

	Country	Rank	Account	Title	Link	Subscriberscount	Viewsavg	Likesavg	Commentsavg	Shares
	All	1.0	billieeilish	BILLIE EILISH	<a href="https://www.tiktok.com/@billieeilish">https://www.tiktok.com/@billieeilish</a>	41600000.0	95000000.0	18100000.0	151000.0	5480
	All	2.0	badbunny	Bad Bunny	<a href="https://www.tiktok.com/@badbunny">https://www.tiktok.com/@badbunny</a>	13400000.0	33300000.0	7300000.0	101900.0	1080
	All	3.0	chipmunksoftiktok	Chipmunks of TikTok	<a href="https://www.tiktok.com/@chipmunksoftiktok">https://www.tiktok.com/@chipmunksoftiktok</a>	11400000.0	25000000.0	3000000.0	21800.0	6420
	All	4.0	charlidamelio	charli d'amelio	<a href="https://www.tiktok.com/@charlidamelio">https://www.tiktok.com/@charlidamelio</a>	135400000.0	18700000.0	2800000.0	54700.0	3520
	All	5.0	yessicadcs	Yessica y Moises	<a href="https://www.tiktok.com/@yessicadcs">https://www.tiktok.com/@yessicadcs</a>	263700.0	16700000.0	1400000.0	16200.0	13840
...	...	...	...	...	...	...	...	...	...	...
	All	996.0	leleburnier	LELÉ BURNIER	<a href="https://www.tiktok.com/@leleburnier">https://www.tiktok.com/@leleburnier</a>	1100000.0	1200000.0	284300.0	2800.0	90
	All	997.0	sofiadelmonstro	Cat Sofia	<a href="https://www.tiktok.com/@sofiadelmonstro">https://www.tiktok.com/@sofiadelmonstro</a>	5200000.0	2600000.0	161200.0	1300.0	160
	All	998.0	jena	Jena	<a href="https://www.tiktok.com/@jena">https://www.tiktok.com/@jena</a>	10500000.0	2800000.0	320300.0	856.0	510
	All	999.0	christandaloi_	Christandaloi_	<a href="https://www.tiktok.com/@christandaloi_">https://www.tiktok.com/@christandaloi_</a>	2700000.0	1100000.0	188800.0	3000.0	290
	All	1000.0	davidgetal	deivis	<a href="https://www.tiktok.com/@davidgetal">https://www.tiktok.com/@davidgetal</a>	13200000.0	2300000.0	337500.0	1400.0	210

3 rows x 11 columns

Nesta parte do código foi criada a variável "top" que compreende o ficheiro csv utilizado, na segunda linha foram substituídos os espaços em branco dos títulos das colunas, a terceira linha serviu apenas para invocar a variável top, para comprovar que foram feitas as devidas alterações

## Retirar definitivamente o campo "Scraped"

```
In [23]: #a seguinte linha de código serve o propósito de retirar a coluna "Scraped" pois não será utilizada no decorrer do trabalho
top.drop('Scraped', axis=1, inplace=True)
top.head(10)
```

Out[23]:

	Country	Rank	Account	Title	Link	Subscriberscount	Viewsavg	Likesavg	Commentsavg	Sharesavg
0	All	1.0	billieeilish	BILLIE EILISH	<a href="https://www.tiktok.com/@billieeilish">https://www.tiktok.com/@billieeilish</a>	41600000.0	95000000.0	18100000.0	151000.0	54800.0
1	All	2.0	badbunny	Bad Bunny	<a href="https://www.tiktok.com/@badbunny">https://www.tiktok.com/@badbunny</a>	13400000.0	33300000.0	7300000.0	101900.0	106800.0
2	All	3.0	chipmunksoftiktok	Chipmunks of TikTok	<a href="https://www.tiktok.com/@chipmunksoftiktok">https://www.tiktok.com/@chipmunksoftiktok</a>	11400000.0	25000000.0	3000000.0	21800.0	64200.0
3	All	4.0	charlidamelio	charli d'amelio	<a href="https://www.tiktok.com/@charlidamelio">https://www.tiktok.com/@charlidamelio</a>	135400000.0	18700000.0	2600000.0	54700.0	35200.0
4	All	5.0	yessicadcs	Yessica y Moises 🇵🇷	<a href="https://www.tiktok.com/@yessicadcs">https://www.tiktok.com/@yessicadcs</a>	263700.0	16700000.0	1400000.0	16200.0	138400.0
5	All	6.0	therock	The Rock	<a href="https://www.tiktok.com/@therock">https://www.tiktok.com/@therock</a>	49700000.0	24700000.0	3000000.0	30900.0	16400.0
6	All	7.0	the_pruid	ThePruid	<a href="https://www.tiktok.com/@the_pruid">https://www.tiktok.com/@the_pruid</a>	433300.0	12200000.0	2700000.0	36300.0	56600.0
7	All	8.0	lukedonkin	Luke Donkin	<a href="https://www.tiktok.com/@lukedonkin">https://www.tiktok.com/@lukedonkin</a>	2000000.0	14000000.0	2400000.0	34000.0	54900.0
8	All	9.0	masteringus	masteringus	<a href="https://www.tiktok.com/@masteringus">https://www.tiktok.com/@masteringus</a>	1300000.0	13400000.0	2000000.0	32500.0	44100.0
9	All	10.0	txt.bighitent	TOMORROW X TOGETHER	<a href="https://www.tiktok.com/@txt.bighitent">https://www.tiktok.com/@txt.bighitent</a>	15300000.0	7200000.0	2200000.0	57200.0	31000.0

Nesta parte do código foi utilizado `.drop` aplicado à variável `top` para eliminar a coluna "Scraped", `axis=1` significa que será eliminada uma coluna e `"inplace"` para retirar a devida coluna, `top.head(10)` invoca a variável `top`, ou seja mostra a tabela de dados, para comprovar que a coluna "Scraped" foi eliminada, como previsto

```
In [24]: top.dtypes #de modo a descobrir o tipo de dados implementamos esta função
```

```
Out[24]: Country      object
Rank                float64
Account             object
Title               object
Link                object
Subscriberscount    float64
Viewsavg            float64
Likesavg            float64
Commentsavg         float64
Sharesavg           float64
dtype: object
```

Nesta parte do código foi utilizado `".dtypes"` aplicado à variável `"top"` para conseguirmos ver qual o tipo de dado utilizado em cada coluna, por exemplo, na coluna `"viewavg"`, `float64` significa que os dados são números, que podem ser decimais, ou seja separados com vírgulas.

## Qual a média mais alta e mais baixa de visualizações no TikTok ?

```
In [25]: max_visua = top.Viewsavg.max() #calcular o valor máximo da coluna Viewsavg
min_visua = top.Viewsavg.min() #calcular o valor mínimo da coluna Viewsavg

max_visua,min_visua
```

Out[25]: (95000000.0, 585300.0)

Nesta parte do código foi definida a variável "max\_visua" que compreende o valor máximo atribuído à média de visualizações e a variável "min\_visua", que compreende o valor mínimo que a média de visualizações na coluna "Viewsavg" assume, depois são invocadas ambas as variáveis para que estas aparecem no output

## Qual é o influencer com menos subscritores ?

```
In [26]: #ordenar os dados por ordem ascendente a partir de uma coluna para ver qual o influencer com menos subscritores
menos = top.sort_values(by=['Subscriberscount']).head(1)
#head(1) = mostra apenas a primeira linha, ou seja, o influencer com menos subscritores

menos
```

Out[26]:

	Country	Rank	Account	Title	Link	Subscriberscount	Viewsavg	Likesavg	Commentsavg	Sharesavg
844	All	845.0	bipbrome	Lena 🇧🇷	<a href="https://www.tiktok.com/@bipbrome">https://www.tiktok.com/@bipbrome</a>	4300.0	750700.0	213400.0	4300.0	5900.0

```
In [27]: nome=menos.iloc[0]["Account"]
nome
```

Out[27]: 'bipbrome'

Nesta parte do código foi definida a variável "menos", que compreende a linha correspondente à conta de tiktok com menos subs do top 1000, na linha 27 do código localizámos a coluna com o título "Account" pelo método iloc(0) ou seja a linha 0 da parte da tabela que foi mostrada no output da linha 26 e criámos uma variável para que compreende apenas o nome da conta apresentada nessa linha.



## Qual é o influencer com mais subscritores ?

```
In [28]: #ordenar os dados por ordem descendente a partir de uma coluna para ver qual o influencer com menos subscritores
mais = top.sort_values(by=['Subscriberscount'], ascending=False).head(1)

#sort_values é uma função que ordena por defeito os dados por ordem ascendente através de uma coluna
#.head(1) = mostra apenas a primeira linha, ou seja, o influencer com mais subscritores
mais
```

Out[28]:

	Country	Rank	Account	Title	Link	Subscriberscount	Viewsavg	Likesavg	Commentsavg	Sharesavg
3	All	4.0	charlidamelio	charli d'amelio	https://www.tiktok.com/@charlidamelio	135400000.0	18700000.0	2600000.0	54700.0	35200.0

```
In [29]: nome=mais.iloc[0]["Account"]
nome
```

Out[29]: 'charlidamelio'

Nesta parte do código foi definida a variável "mais", dentro dela utilizámos o método "sort.values" associado à coluna "Subscriberscount", ascending=false serve para começar a organização de forma descendente o que nos permite obter a linha com maior valor no parâmetro "Subscriberscount", head(1) serviu para que aparecesse apenas a primeira linha da tabela, de seguida invocámos a variável "mais". De seguida, na linha 29 através do método iloc(0) aplicado à coluna Account foi feito com que apenas aparecesse o nome da conta com mais subs.

## Top 10 influencers com mais subscritores no TikTok

```
In [30]: #aqui criamos uma variavel que compreende os 10 influencers com mais subs do tiktok
top10subs = top.sort_values(by=['Subscriberscount'], ascending=False).head(10)
top10subs
#a função sort_values organiza por defeito os dados por ordem ascendente, logo recorremos à função ascending=false para que fizesse
#head(10) mostra as 10 primeiras linhas, neste caso os 10 influencers com mais subscritores
```

Out[30]:

	Country	Rank	Account	Title	Link	Subscriberscount	Viewsavg	Likesavg	Commentsavg	Sharesavg
3	All	4.0	charlidamelio	charli d'amelio	https://www.tiktok.com/@charlidamelio	135400000.0	18700000.0	2600000.0	54700.0	35200.0
27	All	28.0	khaby.lame	Khabane lame	https://www.tiktok.com/@khaby.lame	131100000.0	13800000.0	1600000.0	12700.0	5900.0
20	All	21.0	bellapoarch	Bella Poarch	https://www.tiktok.com/@bellapoarch	87900000.0	15400000.0	1800000.0	16400.0	3100.0
43	All	44.0	addisonre	Addison Rae	https://www.tiktok.com/@addisonre	86500000.0	9500000.0	1100000.0	12900.0	6900.0
119	All	120.0	zachking	Zach King	https://www.tiktok.com/@zachking	67100000.0	10600000.0	709800.0	3200.0	2200.0
72	All	73.0	willsmith	Will Smith	https://www.tiktok.com/@willsmith	65600000.0	4800000.0	736600.0	6400.0	20100.0
23	All	24.0	kimberly.loaiza	Kimberly Loaiza	https://www.tiktok.com/@kimberly.loaiza	59000000.0	10400000.0	1600000.0	22500.0	4900.0
127	All	128.0	dixiedamelio	dixie	https://www.tiktok.com/@dixiedamelio	57000000.0	6200000.0	705400.0	5900.0	3000.0
855	All	856.0	cznburak	cznburak	https://www.tiktok.com/@cznburak	56000000.0	2900000.0	170600.0	2000.0	1200.0
639	All	640.0	lorengray	Loren Gray	https://www.tiktok.com/@lorengray	54400000.0	2400000.0	354800.0	1500.0	1400.0

Nesta parte do código foi definida a variável "top10subs", que compreende de forma organizada de forma descendente quanto ao número de subs apresentados na coluna "Subscriberscount" através do método ascending=false. De seguida invocámos a variável.



## Gráfico de Subscritores

```
In [31]: import numpy as np #importar os numpy como np
import matplotlib.pyplot as plt #importar da biblioteca dados de gráficos

#dados necessários para o gráfico
conta = top10subs['Account'].tolist() #transformar os valores da coluna Account presentes numa lista numa única variável
subs = top10subs['Subscriberscount'].tolist() #transformar os valores da coluna Subscriberscount presentes numa lista numa única

#definições do gráfico
plt.bar(conta, subs, color="blue") #criação de um gráfico de barras azul com os dados das colunas Subscriberscount
plt.gcf().set_size_inches(20, 6) #função que permite trabalhar as dimensões do gráfico

#estilização do gráfico
plt.xlabel("Contas") #título do eixo do x (horizontal)
plt.title("Número de subscritores") #título do gráfico
plt.xticks(conta) #valores do eixo do x (todos os valores da coluna Account)
plt.yticks([500000, 50000000, 100000000, 150000000]) #escala de valores do eixo y (vertical)
plt.show() #invocar o gráfico de barras
```



Nesta parte do código foi criada a variável "conta" e a variável "subs" que transformam os valores de "Account" e "Subscriberscount" numa lista numa única variável. As duas próximas linhas configuram o gráfico, uma para que as barras correspondentes aos valores de "Subscriberscount" apareçam em azul e a outra trabalha as dimensões do gráfico.

As últimas linhas: "plt.xlabel(Contas)" faz com que o eixo x tenha o título "Contas", a linha seguinte faz com que o título do gráfico seja "Número de subscritores", "plt.xticks(Conta)" faz com que o eixo x apresente os nomes das contas, "plt.yticks(.....)" personaliza os valores da escala do eixo y. "plt.show()" invoca o gráfico.

## Comparação entre as views e subs do top 10 influencers com mais subscritores no tiktok.

```
In [32]: import numpy as np #importar os numpy como np
import matplotlib.pyplot as plt #importar da biblioteca dados de gráficos

#dados para o gráfico
conta = top10subs ['Account'].tolist() #transformar os valores da coluna Account numa lista numa única variável
subs = top10subs ['Subscriberscount'].tolist() #transformar os valores da coluna Subscriberscount numa lista numa única variável
conta = top10subs ['Account'].tolist() #transformar os valores da coluna Account em uma lista numa única variável
views = top10subs ['Viewsavg'].tolist() #transformar os valores da coluna Viewsavg em uma lista numa única variável

#definições do gráfico
plt.bar(conta, subs, color="blue", label="subscritores") #criar um gráfico de barras azul com os dados das colunas Subscriberscount
plt.gcf().set_size_inches(20, 6) #função para trabalhar as dimensões do gráfico
plt.bar(conta, views, color="red", label="visualizações") #criar um gráfico de barras azul com os dados das colunas Viewsavg
plt.gcf().set_size_inches(20, 6) #função para trabalhar as dimensões do gráfico

#estilização do gráfico
plt.xlabel("Contas") #título do eixo do x (horizontal)
plt.title("Número de subscritores perante visualizações") #título do gráfico
plt.xticks(conta) #valores do eixo do x (todos os valores da coluna Account)
plt.yticks([500000,500000000,1000000000,1500000000]) #escala para os valores do eixo y (vertical)
plt.legend() #mostrar Legenda do gráfico
plt.show() #mostrar o gráfico
```



Nesta parte do código foi criada a variável foram invocados de novo os dados necessários para a criação do gráfico: as variáveis: conta, subs, conta e views. Nas próximas 4 linhas foi atribuída a cor azul às barras correspondentes aos subscritores e vermelho às barras correspondentes às visualizações. A última parte do código anteriormente apresentado faz de uma forma sucinta o mesmo que as ultimas linhas da imagem anterior a esta.

## Agora vamos iniciar uma comparação entre o top 10 influencers com uma menor média de likes e a média de comentários dos mesmos.

```
In [33]: top10likes = top.sort_values(by=['Likesavg'], ascending=True).head(10) #criação da variável "top10likes" que compreende os valores
top10likes
# a função sort_values organiza por defeito os dados por ordem ascendente, logo recorremos à função ascending=true
```

Out[33]:

	Country	Rank	Account	Title	Link	Subscriberscount	Viewsavg	Likesavg	Commentsavg	Share
926	All	927.0	weirdryanstuff	Weird Ryan	<a href="https://www.tiktok.com/@weirdryanstuff">https://www.tiktok.com/@weirdryanstuff</a>	1200000.0	1700000.0	31000.0	26400.0	
463	All	464.0	naple_13	☆Awys☆	<a href="https://www.tiktok.com/@naple_13">https://www.tiktok.com/@naple_13</a>	187400.0	4200000.0	47000.0	11500.0	11
881	All	882.0	viajeademor	Viaje de Amor	<a href="https://www.tiktok.com/@viajeademor">https://www.tiktok.com/@viajeademor</a>	257700.0	988100.0	48500.0	538.0	31
813	All	814.0	truenorthoutdoors	HOOO-LY THAT'S GOOD	<a href="https://www.tiktok.com/@truenorthoutdoors">https://www.tiktok.com/@truenorthoutdoors</a>	781700.0	2100000.0	62800.0	286.0	13
41	All	42.0	mero_428	Mero_428	<a href="https://www.tiktok.com/@mero_428">https://www.tiktok.com/@mero_428</a>	1300000.0	585300.0	70900.0	3000.0	5300
596	All	597.0	dustinvogelbacher	Dustin vogelbacher	<a href="https://www.tiktok.com/@dustinvogelbacher">https://www.tiktok.com/@dustinvogelbacher</a>	624200.0	1300000.0	78300.0	1300.0	26
601	All	602.0	yamandihati	Yaman Dihati	<a href="https://www.tiktok.com/@yamandihati">https://www.tiktok.com/@yamandihati</a>	267700.0	2700000.0	78900.0	8200.0	4
778	All	779.0	noted99	AMOY99	<a href="https://www.tiktok.com/@noted99">https://www.tiktok.com/@noted99</a>	21300.0	1100000.0	85400.0	30.0	20
904	All	905.0	daleyzabeauty	Daleyzaa_beauty	<a href="https://www.tiktok.com/@daleyzabeauty">https://www.tiktok.com/@daleyzabeauty</a>	1900000.0	3100000.0	86700.0	144.0	5
878	All	879.0	deiveleonardooficial	Deive Leonardo	<a href="https://www.tiktok.com/@deiveleonardooficial">https://www.tiktok.com/@deiveleonardooficial</a>	5300000.0	853100.0	86700.0	777.0	19

Nesta parte do código foi criada a função top10likes, que compreende a tabela de dados organizada de forma descendente quanto ao número de likes, foi utilizado o método ascending=true. De seguida foi invocada a variável para comprovar que foram feitas as alterações pretendidas

## Gráfico comparação entre likes e comentários dos influencers com a menor média de likes

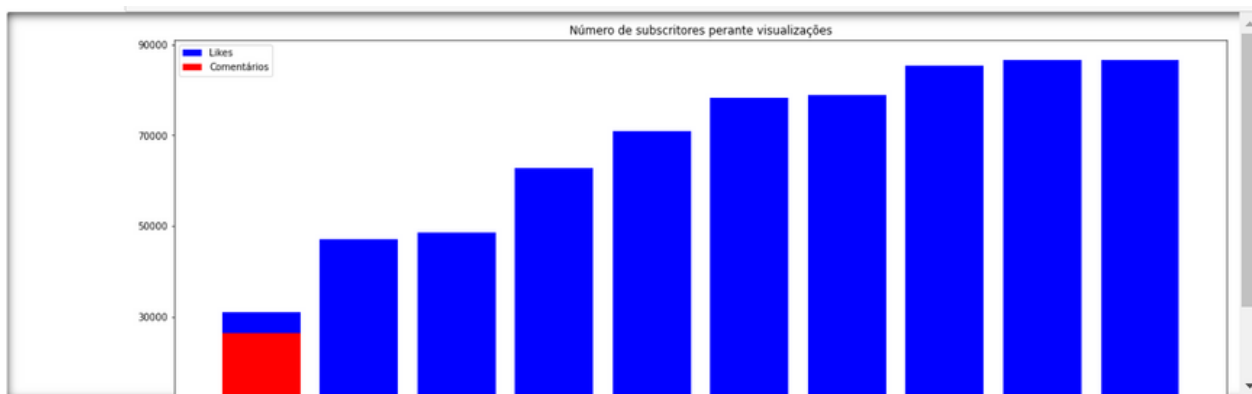
```
In [34]: import numpy as np #importar os numpy como np
import matplotlib.pyplot as plt #importar da biblioteca dados de gráficos

#dados para o gráfico
conta = top10likes ['Account'].tolist() #transformar os valores da coluna Account numa lista numa única variável
likes = top10likes ['Likesavg'].tolist() #transformar os valores da coluna Likesavg numa lista numa única variável
conta = top10likes ['Account'].tolist() #transformar os valores da coluna Account numa lista numa única variável
comments = top10likes ['Commentsavg'].tolist() #transformar os valores da coluna Commentavg numa lista numa única variável

#definições do gráfico
plt.bar(conta, likes, color="blue", label="Likes") #criar um gráfico de barras azul com os dados das colunas Likesavg
plt.gcf().set_size_inches(20, 8) #função para trabalhar as dimensões do gráfico
plt.bar(conta, comments, color="red", label="Comentários") #criar um gráfico de barras azul com os dados das colunas Commentavg
plt.gcf().set_size_inches(20, 8) #função para trabalhar as dimensões do gráfico

#estilização do gráfico
plt.xlabel("Contas") #título do eixo do x (horizontal)
plt.title("Número de subscritores perante visualizações") #título do gráfico
plt.xticks(conta) #valores do eixo do x (todos os valores da coluna Account)
plt.yticks([10000, 30000, 50000, 70000, 90000]) #escala para os valores do eixo y (vertical)
plt.legend() #Legenda do gráfico
plt.show() #mostrar o gráfico
```

Nesta parte do código foi criado o gráfico do mesmo modo que no gráfico criado anteriormente porém foram relacionados os likes e comentários.



Este é o output das linhas de código anteriores.

## Top 10 menos partilhas

```
In [36]: top10menospartilhas = top.sort_values(by=['Sharesavg'], ascending=True).head(10) #criação da variável "top10menospartilhas" que
top10menospartilhas
#a função sort_values organiza por defeito os dados por ordem ascendente, Logo recorremos à função ascending=true
#head(10) mostra as 10 primeiras linhas, neste caso os 10 influencers com menos partilhas
```

Out[36]:

	Country	Rank	Account	Title	Link	Subscriberscount	Viewsavg	Likesavg	Commentsavg	Sharesavg
999	All	1000.0	davidgetial	devis	<a href="https://www.tiktok.com/@davidgetial">https://www.tiktok.com/@davidgetial</a>	13200000.0	2300000.0	337500.0	1400.0	212.0
717	All	718.0	baldybrobryxzx	celloszcx	<a href="https://www.tiktok.com/@baldybrobryxzx">https://www.tiktok.com/@baldybrobryxzx</a>	10100000.0	3200000.0	279300.0	2100.0	260.0
586	All	587.0	ianboggs	IAN	<a href="https://www.tiktok.com/@ianboggs">https://www.tiktok.com/@ianboggs</a>	10200000.0	3100000.0	371500.0	2300.0	264.0
783	All	784.0	fabiola.baglieri	Fabiola	<a href="https://www.tiktok.com/@fabiola.baglieri">https://www.tiktok.com/@fabiola.baglieri</a>	11200000.0	2700000.0	384100.0	1500.0	265.0
923	All	924.0	xllsoup	Soup	<a href="https://www.tiktok.com/@xllsoup">https://www.tiktok.com/@xllsoup</a>	1700000.0	2000000.0	299300.0	2100.0	274.0
980	All	981.0	daniela_giraldo1	Daniela Giraldo	<a href="https://www.tiktok.com/@daniela_giraldo1">https://www.tiktok.com/@daniela_giraldo1</a>	17700000.0	2400000.0	360400.0	1200.0	288.0
984	All	985.0	minhnggia13522	Min Minn	<a href="https://www.tiktok.com/@minhnggia13522">https://www.tiktok.com/@minhnggia13522</a>	2400000.0	2200000.0	264400.0	1900.0	289.0
732	All	733.0	janeydm	จันทน์ วัฒนกุล	<a href="https://www.tiktok.com/@janeydm">https://www.tiktok.com/@janeydm</a>	14000000.0	3200000.0	415400.0	1300.0	293.0
808	All	809.0	indrakenz	Indra Kesuma	<a href="https://www.tiktok.com/@indrakenz">https://www.tiktok.com/@indrakenz</a>	5300000.0	2700000.0	163100.0	3700.0	301.0
946	All	947.0	corvii_ff	EL CORVIL	<a href="https://www.tiktok.com/@corvii_ff">https://www.tiktok.com/@corvii_ff</a>	10300000.0	1600000.0	259400.0	2900.0	307.0

Aqui foi criada a variável "top10menospartilhas" que compreende a tabela de dados organizados de forma ascendente utilizando o método ascending=true quanto à coluna "Sharesavg" e de seguida foi invocada a variável para comprovar que foram feitas as alterações pretendidas.

## Qual é a média de partilhas do top 1000 de influencers no TikTok?

```
In [37]: media_partilhas = top["Sharesavg"].mean() #média de todos os valores da coluna Sharesavg
media_partilhas=round(media_partilhas,2) #arredondar o resultado para 2 casas decimais
print("No top 1000 do TikTok, existe em média ",media_partilhas , "de partilhas.")
```

No top 1000 do TikTok, existe em média 10562.57 de partilhas.

## Qual é a média de comentários do top 1000 de influencers no TikTok?

```
In [38]: media_comentarios = top["Commentsavg"].mean() #média de todos os valores da coluna Commentsavg
media_comentarios=round(media_comentarios,2) #arredondar o resultado para 2 casas decimais
print("No top 1000 do TikTok, existe em média ",media_comentarios , "de comentários.")
```

No top 1000 do TikTok, existe em média 4562.24 de comentários.

## Qual é a média de likes do top 1000 de influencers no TikTok?

```
In [39]: media_likes = top["Likesavg"].mean() #média de todos os valores da coluna Likesavg
media_likes=round(media_likes,2) #arredondar o resultado para 2 casas decimais
print("No top 1000 do TikTok, existe em média ",media_likes , "de likes.")
```

No top 1000 do TikTok, existe em média 453433.0 de likes.

Nas primeiras linhas de código foi criada a variável "media\_partilhas" que compreende o valor da média de todos os valores associados à coluna "Sharesavg" através do método "mean()", de seguida através de "round(media\_partilhas,2)" faz com que os valores da média sejam apresentados com 2 casas decimais. de seguida foi feito um print que enuncia o número médio de partilhas no output. Nas linhas de código dentro de In 38, 39, 40 e 41 são utilizados os mesmos métodos e são criadas as variáveis: media\_comentarios, media\_likes, media\_views e media\_subs.

## Gráfico das médias de likes, comentários, visualizações, partilhas e subscritores

In [19]:

```
#dados do gráfico
labels = ['Subscritores', 'Visualizações', 'Likes', 'Comentários', 'Partilhas'] #dar nome as legendas do gráfico
Medias = [media_subs, media_views, media_likes, media_comentarios, media_partilhas] #definir quais os dados a serem colocados

plt.axis("equal") #define uma escala igual

plt.pie(Medias, labels=labels, autopct='%1.1f%%', shadow=True, radius=1.5) #fazer um gráfico circular com os
#dados das Médias calculadas. foram postas labels - subscritores, visualizações, likes, comentários e partilhas.
#Autopct serve para atribuir uma legenda a cada uma das "fatias" do gráfico
#o seu respetivo número. shadow cria uma sombra no gráfico quando igualdo a true
#Radius é o raio do gráfico que inicialmente é atribuido automaticamente o valor de 1.

plt.legend(loc='lower right', bbox_to_anchor=(1.8,-0.1)) #plt.legend coloca a legenda do gráfico no sítio escolhido, neste caso co
#loc='upper left' = localização mias precisa da legenda - canto superior esquerdo
plt.title('Influencers', pad=70) #plt.title = (titulo do gráfico que neste caso foi subido quanto ao eixo superior em 70 pontos

plt.show() #mostrar o gráfico de barras
```

Esta parte do código tem o objetivo final de criar um gráfico de "pizza" que apresenta a comparação das medias de cada parâmetro: subs, views, likes, partilhas e comentários. Nas primeiras linhas de código são atribuídos nomes às "labels" do gráfico através do método labels=....., Medias compreende os dados colocados, neste caso os dados das variáveis das médias criadas nas linhas de código presentes na imagem anterior à de cima. plt.axis("equal") faz com que a escala seja sempre igual. plt.pie cria o gráfico circular, labels=labels vai fazer com que as labels sejam criadas com os títulos atribuídos à variável labels anteriormente, radius=1.5 define o raio do gráfico circular. plt.legend(loc="lower right"...., coloca a legenda do gráfico no canto inferior direito. plt.title("Influencers", pads=70) faz com que o título do gráfico seja influencers e que este seja subido 70 pontos sobre o eixo vertical, o eixo y. plt.show faz com que o gráfico seja revelado

# conclusão

Com a realização deste trabalho, conseguimos obter mais informações acerca do tiktok e dos seus usuários assim como o seu desempenho relativamente aos likes, partilhas e comentários. E concluimos assim que, existem em alguns casos mais visualizações do que subscritores e em outros casos mais subscritores do que likes. Para além disto este trabalho serviu principalmente o propósito de nos ajudar a consolidar as nossas "skills" trabalhadas em aula no que toca ao tratamento de dados, através de pandas com recurso a jupyter notebook