

Atividade 3 – Projeto Covid19

O mundo está sofrendo uma pandemia causada pela infecção generalizada pelo vírus Covid-19, também conhecido por Corona vírus. Assumindo que trabalhamos numa empresa de Data Science voltada a Helthcare, levantamos alguns dados relacionados ao número de casos. Diante disto, algumas análises precisam ser realizadas.

1 – Agrupar a soma de infectados e mortes de todos os países por data. (Ao final você terá um dataframe contendo 90.624 linhas)

2 – Criar duas colunas contendo a variação bruta diária de infectados e mortos. Recomenda-se realizar esta tarefa conforme o seguinte exemplo:

```
df[['diff_var1', 'diff_var2']] = df.groupby('group')[['var1', 'var2']].diff()
```

O dataframe resultante será algo similar a:

	country	date	infected	dead	diff_infected	diff_dead
0	Afghanistan	2020-01-22	0	0	NaN	NaN
1	Afghanistan	2020-01-23	0	0	0.0	0.0
2	Afghanistan	2020-01-24	0	0	0.0	0.0
3	Afghanistan	2020-01-25	0	0	0.0	0.0
4	Afghanistan	2020-01-26	0	0	0.0	0.0
...
90619	Zimbabwe	2021-05-03	38293	1573	12.0	3.0
90620	Zimbabwe	2021-05-04	38327	1574	34.0	1.0
90621	Zimbabwe	2021-05-05	38357	1574	30.0	0.0
90622	Zimbabwe	2021-05-06	38398	1575	41.0	1.0
90623	Zimbabwe	2021-05-07	38403	1576	5.0	1.0

90624 rows × 6 columns

3 – Criar duas colunas contendo a variação percentual diária de infectados e mortos. Recomenda-se realizar esta tarefa conforme o seguinte exemplo:

```
df[['diff_var1', 'diff_var2']] = df.groupby('group')[['var1', 'var2']].pct_change()
```

O dataframe resultante será algo similar a:

	country	date	infected	dead	diff_infected	diff_dead	pct_diff_infected	pct_diff_dead
0	Afghanistan	2020-01-22	0	0	NaN	NaN	NaN	NaN
1	Afghanistan	2020-01-23	0	0	0.0	0.0	NaN	NaN
2	Afghanistan	2020-01-24	0	0	0.0	0.0	NaN	NaN
3	Afghanistan	2020-01-25	0	0	0.0	0.0	NaN	NaN
4	Afghanistan	2020-01-26	0	0	0.0	0.0	NaN	NaN
...
90619	Zimbabwe	2021-05-03	38293	1573	12.0	3.0	0.000313	0.001911
90620	Zimbabwe	2021-05-04	38327	1574	34.0	1.0	0.000888	0.000636
90621	Zimbabwe	2021-05-05	38357	1574	30.0	0.0	0.000783	0.000000
90622	Zimbabwe	2021-05-06	38398	1575	41.0	1.0	0.001069	0.000635
90623	Zimbabwe	2021-05-07	38403	1576	5.0	1.0	0.000130	0.000635

90624 rows × 8 columns

4 – Filtrar os países: Brasil, Estados Unidos, Itália e Canadá. (Problema secundário: como você resolverá o problema dos valores “NaN” e alguns valores “inf”? Dica: existem algumas funções que podem auxiliar, como `dropna`, `fillna`, `replace`. Escolher o que melhor atender a solução.)

5 – Obter a média mensal da variação bruta de infectados e mortos para cada país. Descreva os resultados. (Dica: transformar a data que está no formato “yyyy-mm-dd” para ano-mês, isto é, “yyyy-mm”, pesquisar o comando “`df['date'].dt.to_period('M')`”)

6 – Obter a média mensal da variação percentual de infectados e mortos para cada país. Descreva os resultados.

7 – Gerar os seguintes gráficos:

- Gráfico de barras contendo a evolução média mensal da variação **BRUTA** de infectados para Brasil, E.U.A e Itália (Um gráfico para cada país separadamente)
- Gráfico de linhas contendo a evolução média mensal da variação **PERCENTUAL** de mortes para Brasil, E.U.A e Itália (Um gráfico para cada país separadamente)