

Introdução à Ciência de Dados

Questionário 2 - Análise exploratória de dados

Prof. Francisco Rodrigues

1 - Leia os dados do seguinte link:

<https://gist.github.com/aishwarya8615/89d9f36fc014dea62487f7347864d16a>

Construa um gráfico de setores (pizza) para a variável Status. Qual a porcentagem de países desenvolvidos?

Use o código a seguir para ler os dados:

```
import pandas as pd
```

```
url = "https://gist.github.com/aishwarya8615/89d9f36fc014dea62487f7347864d16a"
```

```
# Read the table from the URL into a dataframe
```

```
df = pd.read_html(url)[0]
```

2 - Considerando a base anterior, construa um histograma para a variável Life_expectancy. Qual a média e variância desse atributo?

3 - Mostre a série temporal da variável Life_Expectancy para os países: Ireland, Finland, Netherlands, Zambia, Angola. Os países com maior e menor expectativas de vida em 2014 são:

4 - Faça um boxplot para a variável Schooling. Considere os países: 'Ireland', 'Finland', 'Netherlands', 'Zambia', 'Angola'. Em termos da mediana, qual deles apresenta maior nível de escolaridade?

5 - Importe os dados do link:

<https://www.worldometers.info/world-population/population-by-country>

Construa a matriz de correlação. Quais as variáveis menos correlacionadas?

Use o código abaixo para ler os dados.

```
import pandas as pd
import requests
from bs4 import BeautifulSoup

url = "https://www.worldometers.info/world-population/population-by-country/"

# Send a GET request to the URL
response = requests.get(url)

# Parse the HTML content with BeautifulSoup
soup = BeautifulSoup(response.content, 'html.parser')

# Find the table containing the population data
table = soup.find('table', {'id': 'example2'})

# Convert the table to a Pandas DataFrame
df = pd.read_html(str(table))[0]

# Print the DataFrame
df.head()
```