

Atividade Prática

Disciplina	MAP Bootcamp
Atividade	Identificar Outlier

1. Objetivos

Vamos rodar um programa em Python 3.x para identificar outliers em uma lista conforme abaixo:

Minha lista sem outlier:

[3246, 3476, 3724, 3773, 3837, 3968, 4198, 4048, 4170, 4226, 4788, 4009, 3568, 4357]

Minha lista com outlier:

[6799, 3476, 3724, 3773, 3837, 3968, 4198, 4048, 4170, 4226, 4788, 4009, 3568, 4357]

```

1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from scipy import stats
5
6 # Base de dados
7 xi = [3246, 3476, 3724, 3773, 3837, 3968, 4198, 4048, 4170, 4226, 4788, 4009, 3568, 4357]
8
9 # Base de dados com anomalia
10 xy = [6799, 3476, 3724, 3773, 3837, 3968, 4198, 4048, 4170, 4226, 4788, 4009, 3568, 4357]
11
12 def print_full(x):
13     pd.set_option('display.max_rows', len(x))
14     pd.set_option('display.max_columns', len(x))
15     print(x)
16     pd.reset_option('display.max_rows')
17     pd.reset_option('display.max_columns')
18
19 # Média
20 xi_mean = np.mean(xi)
21 xy_mean = np.mean(xy)
22
23 mediana_xi = np.median(xi)
24 mediana_xy = np.median(xy)
25 moda_xi = stats.mode(xi)
26 moda_xy = stats.mode(xy)
27
28 print("Minha lista sem anomalias: " + str(xi))
29 print("A mediana da minha lista é "+str(mediana_xi))
30 # print("A moda da minha lista é "+str(moda_xi))
31 print("Minha lista com anomalias: " + str(xy))
32 print("A mediana da minha lista é "+str(mediana_xy))
33 # print("A moda da minha lista é "+str(moda_xy))
34
35
36 # Desvio Padrão
37 xi_sd = np.std(xi, ddof = 1)
38 xy_sd = np.std(xy, ddof = 1)
39
40 # Z-score
41 z_score_i = (xi - xi_mean) / xi_sd
42 z_score_y = (xy - xy_mean) / xy_sd
43

```

```

43
44 df = pd.DataFrame({'Base (sem anomalias)':xi, 'Z-score (sem anomalias)':z_score_i,
45                  'Base (com anomalias)':xy, 'Z-score (com anomalias)':z_score_y})
46 df.round(2)
47 cols = ['Base (sem anomalias)', 'Z-score (sem anomalias)', 'Base (com anomalias)', 'Z-score (com anomalias)']
48 df = df[cols]
49 print_full(df)
50
51 # Plot da distribuição normal
52 stats.probplot (xi, dist="norm", plot=plt)
53 plt.show()
54
55 # Plot da distribuição normal
56 stats.probplot (xy, dist="norm", plot=plt)
57 plt.show()
58
59 # Create a figure instance
60 fig_xi = plt.figure(1, figsize=(6, 4))
61 # Create an axes instance
62 axi = fig_xi.add_subplot(111)
63 bp_xi = axi.boxplot(xi)
64 # Create a figure instance
65 fig_xy = plt.figure(2, figsize=(6, 4))
66 # Create an axes instance
67 axy = fig_xy.add_subplot(111)
68 bp_xy = axy.boxplot(xy)

```

2. Conclusão

Nesta atividade prática identificamos anomalia em uma lista pelo método z-Score e pela identificação via BoxPlot.