



Engenharia de Computação

Iuri Ramon Cervantes Santos

Relatório : Mineração de dados - Algerian Forest Fires Daset.

Introdução

Este trabalho tem como objetivo aplicar os conhecimentos adquiridos na disciplina de Mineração de Dados em uma base de dados afim de extrair as informações nela contida e resolver o problema proposto pela base de dados.

1- Apresentação da base de dados

A base utilizada neste trabalho foi a Algerian Forest Fires Daset.

Esta base de dados apresenta informações de uma região florestal da Algéria, do período de junho de 2012 a setembro de 2012, a base de dados possui 11 atributos mais 1 atributo de saída (classe).

A base possui 122 objetos, classificados em “ocorreu incêndio” e “não ocorreu incêndio”.

Atributos:

- Data: Dia, mês e ano das informações coletadas;
- Temperatura: Temperatura máxima no dia em graus Celsius (22°C - 42°C);
- Humidade relativa: Humidade relativa no dia em percentagem (21% - 90%);
- Velocidade do vento: velocidade do vento em km/h (6km/h – 29km/h);
- Chuva: total de chuva no dia em milímetros (0mm – 16,8mm);
- FFMC: Índice de humidade de combustíveis finos, representa o teor de humidade dos combustíveis finos mortos, de secagem rápida, constituindo um bom indicador do seu grau de inflamabilidade. (28,6 – 92,5; índice FWI);
- DMC: Índice de húmus, representa o teor de humidade de uma fina e pouco compactada camada de solo (até cerca de 8 cm de profundidade), dando a indicação do

estado da matéria orgânica decomposta (húmus) e materiais lenhosos de tamanho médio que aí se encontram. (1,1 – 65,9; índice FWI);

- DC: Índice de Seca, representa o teor de humidade dos combustíveis florestais (húmus e materiais lenhosos de maiores dimensões), que se encontram abaixo da superfície do solo, entre 8 e 20 cm de profundidade e é um bom indicador dos efeitos do período seco sazonal. (7 – 220,4; índice FWI);
- ISI: Índice de Propagação Inicial, resulta da combinação do FFMC e da intensidade do vento, representando a taxa de propagação inicial do fogo, sem incluir a influência de quantidades variáveis de combustível. (0 – 18,5; índice FWI);
- BUI: Índice de Combustível Disponível, resulta da combinação do DMC e do DC, representando a quantidade total de combustíveis disponível para propagação do fogo. (1,1 – 68; índice FWI);
- Risco Incêndio - Índice de Perigo de Incêndio (FWI). (0 – 31,1);
- Classe: 1 para incêndio, 0 para não incêndio.

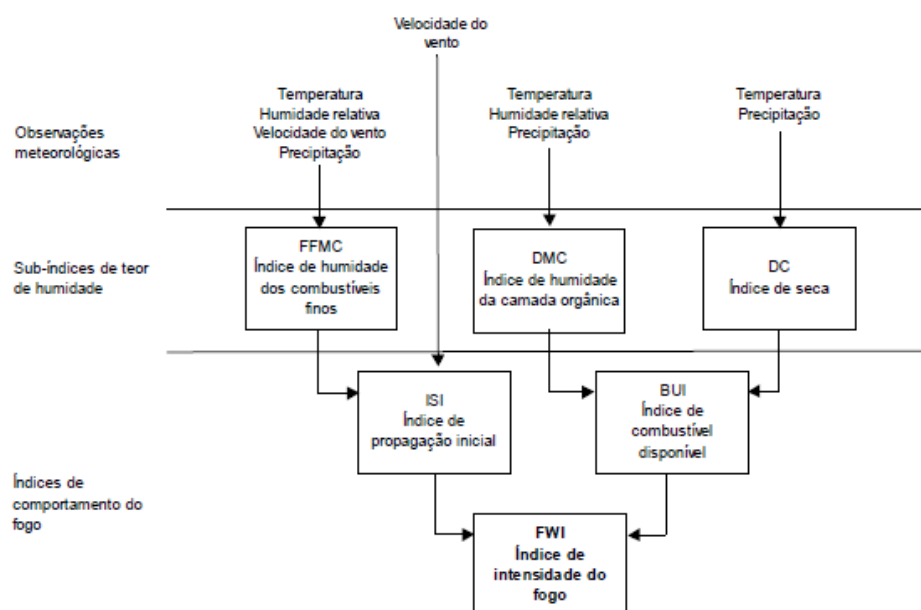


Figura 1 – Estrutura do índice meteorológico de risco de incêndio florestal FWI.

A base de dados apresentava dados faltantes em alguns atributos, tal problema é resolvido no pré-processamento, na fase de limpeza de dados.

2- Pré-Processamento

- Limpeza:

A limpeza da base de dados foi executada por meio do preenchimento dos valores faltantes, para tal foi considerada a média de cada atributo, média foi escolhida pois a sensibilidade aos valores extremos (desde que não sejam outliers) é importante para a detecção de um incêndio.

Seguem as informações sobre o procedimento:

```
INFORMAÇÕES GERAIS DOS DADOS

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 122 entries, 0 to 121
Data columns (total 14 columns):
#   Column          Non-Null Count  Dtype  
---  -
0   Dia              122 non-null   int64  
1   Mes              122 non-null   int64  
2   Ano              122 non-null   int64  
3   Temperatura      99 non-null    float64 
4   RH               117 non-null   float64 
5   WS               118 non-null   float64 
6   Chuva            122 non-null   float64 
7   FPMC             122 non-null   float64 
8   DMC              122 non-null   float64 
9   DC               122 non-null   float64 
10  ISI              122 non-null   float64 
11  BUI              122 non-null   float64 
12  FWI              122 non-null   float64 
13  Classes          122 non-null   int64  
dtypes: float64(10), int64(4)
memory usage: 13.5 KB
None
```

Figura 2 – Classificação geral dos atributos.

```
DESCRIÇÃO DOS DADOS

   Dia      Mes      Ano  Temperatura  ...      ISI      BUI      FWI      Classes
count 122.000000 122.000000 122.0    99.000000  ...  122.000000 122.000000 122.000000 122.000000
mean  15.754098   7.500000 2012.0    31.343434  ...   3.655738 15.426230   5.577869   0.483607
std    8.843274   1.115259   0.0     3.326338  ...   3.021768 14.474302   6.343051   0.501792
min    1.000000   6.000000 2012.0    22.000000  ...   0.000000  1.100000   0.000000   0.000000
25%    8.000000   7.000000 2012.0    29.000000  ...   1.125000  5.100000   0.500000   0.000000
50%   16.000000   7.500000 2012.0    32.000000  ...   2.650000 11.200000   3.000000   0.000000
75%   23.000000   8.000000 2012.0    34.000000  ...   5.600000 21.675000   8.700000   1.000000
max   31.000000   9.000000 2012.0    37.000000  ...  12.500000 67.400000  30.200000   1.000000

[8 rows x 14 columns]
```

Figura 3 – Descrição dos dados antes da limpeza.

	Dia	Mes	Ano	Temperatura	RH	WS	Chuva	FFMC	DMC	DC	ISI	BUI	FWI	Classes
0	1	6	2012	29.0	57.0	18.0	0.0	65.7	3.4	7.6	1.3	3.4	0.5	0
1	2	6	2012	29.0	61.0	13.0	1.3	64.4	4.1	7.6	1.0	3.9	0.4	0
2	3	6	2012	26.0	NaN	22.0	13.1	47.1	2.5	7.1	0.3	2.7	0.1	0
3	4	6	2012	25.0	89.0	13.0	2.5	28.6	1.3	6.9	0.0	1.7	0.0	0
4	5	6	2012	27.0	77.0	16.0	0.0	64.8	3.0	14.2	1.2	3.9	0.5	0
5	6	6	2012	31.0	67.0	14.0	0.0	82.6	5.8	22.2	3.1	7.0	2.5	1
6	7	6	2012	33.0	NaN	13.0	0.0	88.2	9.9	30.5	6.4	10.9	7.2	1
7	8	6	2012	30.0	NaN	15.0	0.0	86.6	12.1	38.3	5.6	13.5	7.1	1
8	9	6	2012	NaN	88.0	13.0	0.2	52.9	7.9	38.8	0.4	10.5	0.3	0
9	10	6	2012	28.0	79.0	12.0	0.0	73.2	9.5	46.3	1.3	12.6	0.9	0
10	11	6	2012	31.0	65.0	14.0	0.0	84.5	12.5	54.3	4.0	15.8	5.6	1
11	12	6	2012	26.0	81.0	19.0	0.0	84.0	13.8	61.4	4.8	17.7	7.1	1
12	13	6	2012	27.0	84.0	21.0	1.2	50.0	6.7	17.0	0.5	6.7	0.2	0
13	14	6	2012	NaN	78.0	20.0	0.5	59.0	4.6	7.8	1.0	4.4	0.4	0
14	15	6	2012	28.0	80.0	17.0	3.1	49.4	3.0	7.4	0.4	3.0	0.1	0

Figura 4 – 15 Linhas iniciais antes de executar a limpeza.

	Dia	Mes	Ano	Temperatura	...	ISI	BUI	FWI	Classes
count	122.000000	122.000000	122.0	122.000000	...	122.000000	122.000000	122.000000	122.000000
mean	15.754098	7.500000	2012.0	31.278689	...	3.655738	15.426230	5.577869	0.483607
std	8.843274	1.115259	0.0	2.996588	...	3.021768	14.474302	6.343051	0.501792
min	1.000000	6.000000	2012.0	22.000000	...	0.000000	1.100000	0.000000	0.000000
25%	8.000000	7.000000	2012.0	30.000000	...	1.125000	5.100000	0.500000	0.000000
50%	16.000000	7.500000	2012.0	31.000000	...	2.650000	11.200000	3.000000	0.000000
75%	23.000000	8.000000	2012.0	33.000000	...	5.600000	21.675000	8.700000	1.000000
max	31.000000	9.000000	2012.0	37.000000	...	12.500000	67.400000	30.200000	1.000000

[8 rows x 14 columns]

Figura 5 – Descrição dos dados pós limpeza.

	Dia	Mes	Ano	Temperatura	RH	WS	Chuva	FFMC	DMC	DC	ISI	BUI	FWI	Classes
0	1	6	2012	29.0	57.0	18.0	0.0	65.7	3.4	7.6	1.3	3.4	0.5	0
1	2	6	2012	29.0	61.0	13.0	1.3	64.4	4.1	7.6	1.0	3.9	0.4	0
2	3	6	2012	26.0	78.0	22.0	13.1	47.1	2.5	7.1	0.3	2.7	0.1	0
3	4	6	2012	25.0	89.0	13.0	2.5	28.6	1.3	6.9	0.0	1.7	0.0	0
4	5	6	2012	27.0	77.0	16.0	0.0	64.8	3.0	14.2	1.2	3.9	0.5	0
5	6	6	2012	31.0	67.0	14.0	0.0	82.6	5.8	22.2	3.1	7.0	2.5	1
6	7	6	2012	33.0	78.0	13.0	0.0	88.2	9.9	30.5	6.4	10.9	7.2	1
7	8	6	2012	30.0	78.0	15.0	0.0	86.6	12.1	38.3	5.6	13.5	7.1	1
8	9	6	2012	31.0	88.0	13.0	0.2	52.9	7.9	38.8	0.4	10.5	0.3	0
9	10	6	2012	28.0	79.0	12.0	0.0	73.2	9.5	46.3	1.3	12.6	0.9	0
10	11	6	2012	31.0	65.0	14.0	0.0	84.5	12.5	54.3	4.0	15.8	5.6	1
11	12	6	2012	26.0	81.0	19.0	0.0	84.0	13.8	61.4	4.8	17.7	7.1	1
12	13	6	2012	27.0	84.0	21.0	1.2	50.0	6.7	17.0	0.5	6.7	0.2	0
13	14	6	2012	31.0	78.0	20.0	0.5	59.0	4.6	7.8	1.0	4.4	0.4	0
14	15	6	2012	28.0	80.0	17.0	3.1	49.4	3.0	7.4	0.4	3.0	0.1	0

Figura 6 – 15 Linhas iniciais após executar a limpeza.

- Normalização:

Após a efetuação da limpeza dos dados, partiu-se para a etapa de normalização, para critério de estudo, foram implementados dois métodos diferentes de normalização, Z-score e Min-Max, que depois acompanharão a base de dados nos procedimentos seguintes, e estes foram os resultados:

Normalização Z-Score

	Dia	Mes	Ano	Temperatura	...	ISI	BUI	FWI	Classes
count	1.220000e+02	122.000000	122.0	1.220000e+02	...	1.220000e+02	1.220000e+02	1.220000e+02	122.000000
mean	5.824121e-17	0.000000	0.0	5.642117e-16	...	5.824121e-17	2.821059e-17	-4.004083e-17	0.483607
std	1.004124e+00	1.004124	0.0	1.004124e+00	...	1.004124e+00	1.004124e+00	1.004124e+00	0.501792
min	-1.675278e+00	-1.350526	0.0	-3.109186e+00	...	-1.214790e+00	-9.938515e-01	-8.829931e-01	0.000000
25%	-8.804515e-01	-0.450175	0.0	-4.284744e-01	...	-8.409559e-01	-7.163601e-01	-8.038416e-01	0.000000
50%	2.792130e-02	0.000000	0.0	-9.338545e-02	...	-3.342034e-01	-2.931856e-01	-4.080842e-01	0.000000
75%	8.227475e-01	0.450175	0.0	5.767925e-01	...	6.460720e-01	4.334951e-01	4.942426e-01	1.000000
max	1.731120e+00	1.350526	0.0	1.917148e+00	...	2.938920e+00	3.605569e+00	3.897756e+00	1.000000

Figura 6 – Dados da normalização Z-score.

[8 rows x 14 columns]													
	Dia	Mes	Ano	Temperatura	RH	WS	...	DMC	DC	ISI	BUI	FWI	Classes
0	-1.675278	-1.350526	0.0	-0.763563	-1.039680	0.730077	...	-0.793971	-0.883547	-0.782804	-0.834294	-0.803842	0
1	-1.561731	-1.350526	0.0	-0.763563	-0.674670	-1.030188	...	-0.731627	-0.883547	-0.882493	-0.799608	-0.819672	0
2	-1.448135	-1.350526	0.0	-1.768830	0.876622	2.138290	...	-0.874127	-0.893243	-1.115101	-0.882855	-0.867163	0
3	-1.334638	-1.350526	0.0	-2.103919	1.880399	-1.030188	...	-0.981003	-0.897122	-1.214790	-0.952228	-0.882993	0
4	-1.221091	-1.350526	0.0	-1.433741	0.785370	0.025971	...	-0.829596	-0.755555	-0.816034	-0.799608	-0.803842	0
5	-1.107545	-1.350526	0.0	-0.093385	-0.127155	-0.678135	...	-0.580221	-0.600413	-0.184670	-0.584552	-0.487236	1
6	-0.993998	-1.350526	0.0	0.576702	0.876622	-1.030188	...	-0.215064	-0.439453	0.011909	-0.313997	0.256788	1
7	-0.880452	-1.350526	0.0	-0.428474	0.876622	-0.326082	...	-0.019127	-0.288189	0.646072	-0.133628	0.240958	1
8	-0.766905	-1.350526	0.0	-0.093385	1.789147	-1.030188	...	-0.393189	-0.278493	-1.081871	-0.341747	-0.835502	0
9	-0.653358	-1.350526	0.0	-1.098652	0.967874	-1.382241	...	-0.250689	-0.133047	-0.782804	-0.196064	-0.740520	0

Figura 7 – Dados normalizados (Z-Score).

	Dia	Mes	Ano	Temperatura	...	ISI	BUI	FWI	Classes
count	122.000000	122.000000	122.0	122.000000	...	122.000000	122.000000	122.000000	122.000000
mean	0.491803	0.500000	0.0	0.618579	...	0.292459	0.216082	0.184698	0.483607
std	0.294776	0.371753	0.0	0.199773	...	0.241741	0.218315	0.210035	0.501792
min	0.000000	0.000000	0.0	0.000000	...	0.000000	0.000000	0.000000	0.000000
25%	0.233333	0.333333	0.0	0.533333	...	0.090000	0.060332	0.016556	0.000000
50%	0.500000	0.500000	0.0	0.600000	...	0.212000	0.152338	0.099338	0.000000
75%	0.733333	0.666667	0.0	0.733333	...	0.448000	0.310332	0.288079	1.000000
max	1.000000	1.000000	0.0	1.000000	...	1.000000	1.000000	1.000000	1.000000

Figura 8 – Dados da normalização Min-Max.

[8 rows x 14 columns]													
	Dia	Mes	Ano	Temperatura	RH	WS	...	DMC	DC	ISI	BUI	FWI	Classes
0	0.000000	0.0	0.0	0.466667	0.272727	0.466667	...	0.050467	0.003279	0.104	0.034691	0.016556	0
1	0.033333	0.0	0.0	0.466667	0.363636	0.133333	...	0.063551	0.003279	0.080	0.042232	0.013245	0
2	0.066667	0.0	0.0	0.266667	0.750000	0.733333	...	0.033645	0.000937	0.024	0.024133	0.003311	0
3	0.100000	0.0	0.0	0.200000	1.000000	0.133333	...	0.011215	0.000000	0.000	0.009050	0.000000	0
4	0.133333	0.0	0.0	0.333333	0.727273	0.333333	...	0.042991	0.034192	0.096	0.042232	0.016556	0
5	0.166667	0.0	0.0	0.600000	0.500000	0.200000	...	0.095327	0.071663	0.248	0.088989	0.082781	1
6	0.200000	0.0	0.0	0.733333	0.750000	0.133333	...	0.171963	0.110539	0.512	0.147813	0.238411	1
7	0.233333	0.0	0.0	0.533333	0.750000	0.266667	...	0.213084	0.147073	0.448	0.187029	0.235099	1
8	0.266667	0.0	0.0	0.600000	0.977273	0.133333	...	0.134579	0.149415	0.032	0.141780	0.009934	0
9	0.300000	0.0	0.0	0.400000	0.772727	0.066667	...	0.164486	0.184543	0.104	0.173454	0.029801	0

Figura 9 – Dados normalizados (Min-Max).

- Redução:

Para o processo de redução de dados, foram consideradas ambas as estratégias de normalização mencionadas anteriormente, e ambas apresentaram resultados diferentes quando expostas ao procedimento PCA, a variável alvo foi a Classe, responsável pela classificação do objeto em incêndio ou não.

- Min-Max

None													
	Dia	Mes	Ano	Temperatura	...	ISI	BUI	FWI	Classes				
count	122.000000	122.000000	122.0	122.000000	...	122.000000	122.000000	122.000000	122.000000				
mean	0.491803	0.500000	0.0	0.618579	...	0.292459	0.216082	0.184698	0.483607				
std	0.294776	0.371753	0.0	0.199773	...	0.241741	0.218315	0.210035	0.501792				
min	0.000000	0.000000	0.0	0.000000	...	0.000000	0.000000	0.000000	0.000000				
25%	0.233333	0.333333	0.0	0.533333	...	0.090000	0.060332	0.016556	0.000000				
50%	0.500000	0.500000	0.0	0.600000	...	0.212000	0.152338	0.099338	0.000000				
75%	0.733333	0.666667	0.0	0.733333	...	0.448000	0.310332	0.288079	1.000000				
max	1.000000	1.000000	0.0	1.000000	...	1.000000	1.000000	1.000000	1.000000				

[8 rows x 14 columns]

	Dia	Mes	Ano	Temperatura	RH	WS	...	DMC	DC	ISI	BUI	FWI	Classes
0	0.000000	0.0	0.0	0.466667	0.272727	0.466667	...	0.050467	0.003279	0.104	0.034691	0.016556	0
1	0.033333	0.0	0.0	0.466667	0.363636	0.133333	...	0.063551	0.003279	0.080	0.042232	0.013245	0
2	0.066667	0.0	0.0	0.266667	0.750000	0.733333	...	0.033645	0.000937	0.024	0.024133	0.003311	0
3	0.100000	0.0	0.0	0.200000	1.000000	0.133333	...	0.011215	0.000000	0.000	0.009050	0.000000	0
4	0.133333	0.0	0.0	0.333333	0.727273	0.333333	...	0.042991	0.034192	0.096	0.042232	0.016556	0
5	0.166667	0.0	0.0	0.600000	0.500000	0.200000	...	0.095327	0.071663	0.248	0.088989	0.082781	1
6	0.200000	0.0	0.0	0.733333	0.750000	0.133333	...	0.171963	0.110539	0.512	0.147813	0.238411	1
7	0.233333	0.0	0.0	0.533333	0.750000	0.266667	...	0.213084	0.147073	0.448	0.187029	0.235099	1
8	0.266667	0.0	0.0	0.600000	0.977273	0.133333	...	0.134579	0.149415	0.032	0.141780	0.009934	0
9	0.300000	0.0	0.0	0.400000	0.772727	0.066667	...	0.164486	0.184543	0.104	0.173454	0.029801	0

[10 rows x 14 columns]

Explained variance per component:

[0.46547484489138274, 0.20856619913002605, 0.1273079308858925, 0.07889625937821567, 0.047112566499467295, 0.031163647098296787, 0.022924237605538682, 0.01900811813439192, 0.010294716614675278, 0.0023620012431202217, 0.000252455015754255, 2.7023530205935622e-05, 2.3435614194409474e-37]

Figura 10 – Normalização Min-Max

Pela variância por componente podemos perceber que os atributos que apresentaram uma maior variância após a normalização foram BUI (Índice de Combustível Disponível) e FWI (Índice de risco de incêndio).

Após a normalização foi aplicada a técnica PCA, proporcionando os seguintes resultados:

```
DataFrame PCA
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 122 entries, 0 to 121
Data columns (total 3 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   principal component 1                 122 non-null    float64
1   principal component 2                 122 non-null    float64
2   Classes                              122 non-null    int64
dtypes: float64(2), int64(1)
memory usage: 3.0 KB
None
```

	principal component 1	principal component 2	Classes
count	1.220000e+02	1.220000e+02	122.000000
mean	1.274026e-17	-1.456030e-17	0.483607
std	5.713452e-01	3.750415e-01	0.501792
min	-1.006894e+00	-5.472171e-01	0.000000
25%	-4.887426e-01	-2.242153e-01	0.000000
50%	-6.202398e-02	-2.355026e-02	0.000000
75%	3.507589e-01	2.439139e-01	1.000000
max	1.791852e+00	5.644910e-01	1.000000

	principal component 1	principal component 2	Classes
0	-0.503866	-0.443315	0
1	-0.524675	-0.450052	0
2	-0.893066	-0.460303	0
3	-1.006894	-0.523028	0
4	-0.601269	-0.508310	0
5	-0.234971	-0.483853	1
6	-0.006708	-0.523956	1
7	-0.040869	-0.527228	1
8	-0.540021	-0.542371	0
9	-0.336427	-0.524466	0

Figura 11 – PCA para Min-Max.

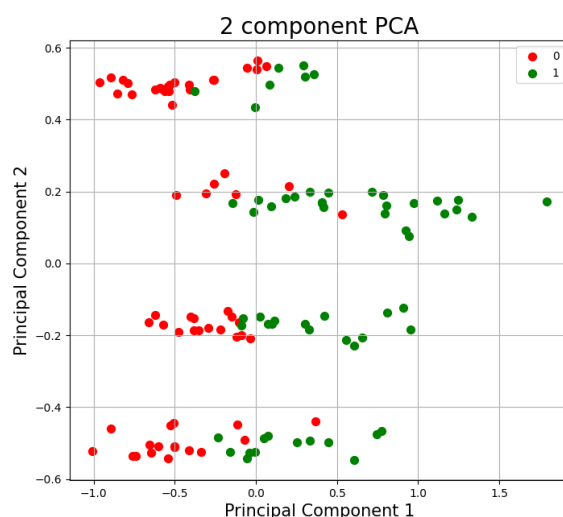


Figura 12- Plot de PCA para Min-Max.

Analisando estes resultados, é perceptível a divisão dos resultados, enquanto os valores classificados como 0 (não incêndio), se encontram em sua maioria com valores negativos para o componente 1, enquanto em relação ao componente 2, variam ao longo de todo o eixo.

○ Z-Score

None	Dia	Mes	Ano	Temperatura	RH	WS	...	DMC	DC	ISI	BUI	FWI	Classes
count	1.220000e+02	122.000000	122.0	1.220000e+02	1.220000e+02	1.220000e+02	...	1.220000e+02	1.220000e+02	1.220000e+02	1.220000e+02	1.220000e+02	122.000000
mean	5.824121e-17	0.000000	0.0	5.642117e-16	4.204287e-16	7.280151e-18	...	-1.528832e-16	7.007145e-17	5.824121e-17	2.821059e-17	-4.004083e-17	0.483607
std	1.004124e+00	1.004124	0.0	1.004124e+00	1.004124e+00	1.004124e+00	...	1.004124e+00	1.004124e+00	1.004124e+00	1.004124e+00	1.004124e+00	0.501792
min	-1.675278e+00	-1.350526	0.0	-3.109186e+00	-2.134709e+00	-1.734295e+00	...	-1.034440e+00	-8.971220e-01	-1.214790e+00	-9.938515e-01	-8.829931e-01	0.000000
25%	-8.804515e-01	-0.458175	0.0	-4.284744e-01	-6.746098e-01	-6.781351e-01	...	-7.650258e-01	-8.368348e-01	-8.409559e-01	-7.163601e-01	-8.038415e-01	0.000000
50%	2.792110e-02	0.000000	0.0	9.338545e-02	-3.590261e-02	-3.260820e-01	...	-2.551424e-01	-3.415193e-01	-3.342034e-01	-2.931050e-01	-4.008042e-01	0.000000
75%	8.227475e-01	0.458175	0.0	5.767925e-01	8.766220e-01	7.300774e-01	...	3.549363e-01	5.015812e-01	6.460720e-01	4.334051e-01	4.942426e-01	1.000000
max	1.731120e+00	1.350526	0.0	1.917148e+00	1.880399e+00	3.546502e+00	...	3.730409e+00	3.243233e+00	2.938920e+00	3.605569e+00	3.897756e+00	1.000000

(8 rows x 14 columns)	Dia	Mes	Ano	Temperatura	RH	WS	Chuva	FFHC	DMC	DC	ISI	BUI	FWI	Classes
0	-1.675278	-1.350526	0.0	-0.763563	-1.039600	0.730077	-0.351193	-0.570004	-0.793971	-0.883547	-0.782804	-0.834204	-0.803842	0
1	-1.561731	-1.350526	0.0	-0.763563	-0.674670	-1.030188	0.190628	-0.662993	-0.731627	-0.883547	-0.882493	-0.799608	-0.819672	0
2	-1.448185	-1.350526	0.0	-1.768830	0.876622	2.138290	5.108701	-1.779495	-0.874127	-0.893243	-1.115101	-0.882855	-0.867163	0
3	-1.334638	-1.350526	0.0	-2.103919	1.880399	-1.030188	0.690771	-2.973443	-0.981003	-0.897122	-1.214790	-0.952228	-0.882993	0
4	-1.221091	-1.350526	0.0	-1.433741	0.785370	0.025971	-0.351193	-0.637178	-0.829596	-0.755555	-0.816034	-0.799608	-0.803842	0
5	-1.107545	-1.350526	0.0	-0.093385	-0.127155	-0.678135	-0.351193	0.511594	-0.580221	-0.600413	-0.184670	-0.584552	-0.487236	1
6	-0.993998	-1.350526	0.0	0.576792	0.876622	-1.030188	-0.351193	0.873005	-0.215064	-0.439453	0.911900	-0.313997	0.256788	1
7	-0.880452	-1.350526	0.0	-0.428474	0.876622	-0.326082	-0.351193	0.760744	-0.019127	-0.288189	0.646072	-0.133628	0.240958	1
8	-0.766905	-1.350526	0.0	-0.093385	1.789147	-1.030188	-0.267836	-1.405176	-0.393189	-0.278493	-1.081871	-0.341747	-0.835502	0
9	-0.653358	-1.350526	0.0	-1.098652	0.967874	-1.382241	-0.351193	-0.095061	-0.250689	-0.133047	-0.782804	-0.196064	-0.740520	0

Explained variance per components:
[0.5160226730013792, 0.1472780837124512, 0.09257878025955606, 0.06758859138176727, 0.0530411213242776386, 0.048473698096576805, 0.03242136948373998, 0.02217683524195548, 0.01664897994338949, 0.002632684305223293, 0.00030371661486126125, 3.337471712373089e-05, 2.4827254685507823e-33]

Figura 13 – Normalização Z-Score.

Pela variância por componente podemos perceber que os atributos que BUI (Índice de Combustível Disponível) e FWI (Índice de risco de incêndio), continuam sendo os que apresentaram uma variância maior do que os outros atributos.

Após a normalização foi aplicada a técnica PCA, proporcionando os seguintes resultados:

```
Dataframe PCA
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 122 entries, 0 to 121
```

```
Data columns (total 3 columns):
```

#	Column	Non-Null Count	Dtype
0	principal component 1	122 non-null	float64
1	principal component 2	122 non-null	float64
2	Classes	122 non-null	int64

```
dtypes: float64(2), int64(1)
```

```
memory usage: 3.0 KB
```

```
None
```

	principal component 1	principal component 2	Classes
count	1.220000e+02	1.220000e+02	122.000000
mean	3.640075e-17	7.280151e-18	0.483607
std	2.500625e+00	1.334894e+00	0.501792
min	-4.341686e+00	-2.845961e+00	0.000000
25%	-2.026179e+00	-8.772041e-01	0.000000
50%	-3.076748e-01	-2.600732e-01	0.000000
75%	1.438445e+00	8.344961e-01	1.000000
max	7.874032e+00	4.900078e+00	1.000000

	principal component 1	principal component 2	Classes
0	-2.001236	-0.928025	0
1	-2.106809	-1.417235	0
2	-4.341686	3.445404	0
3	-4.170996	0.652135	0
4	-2.463878	-0.302742	0
5	-0.862813	-1.494485	1
6	0.226868	-1.369866	1
7	-0.011712	-0.507242	1
8	-2.034657	-0.222111	0
9	-1.343563	-0.629996	0

Figura 14 – PCA para Z-Score.

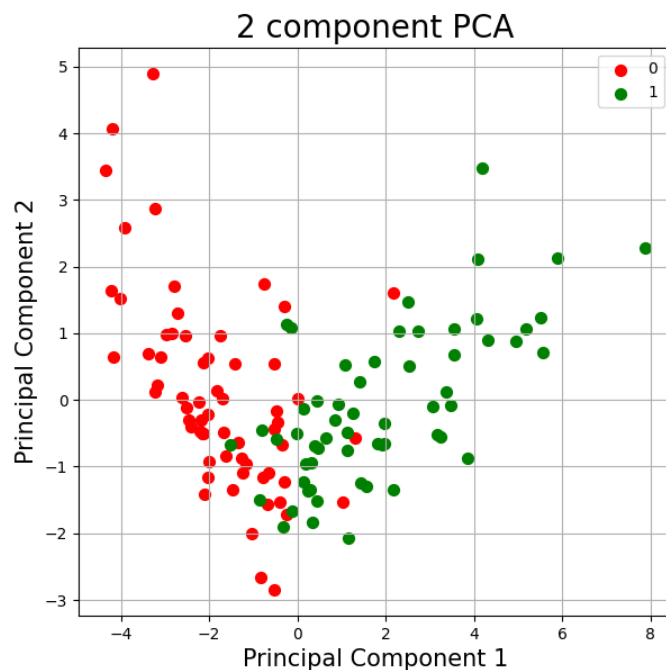


Figura 15 – Plot de PCA para Z-Score.

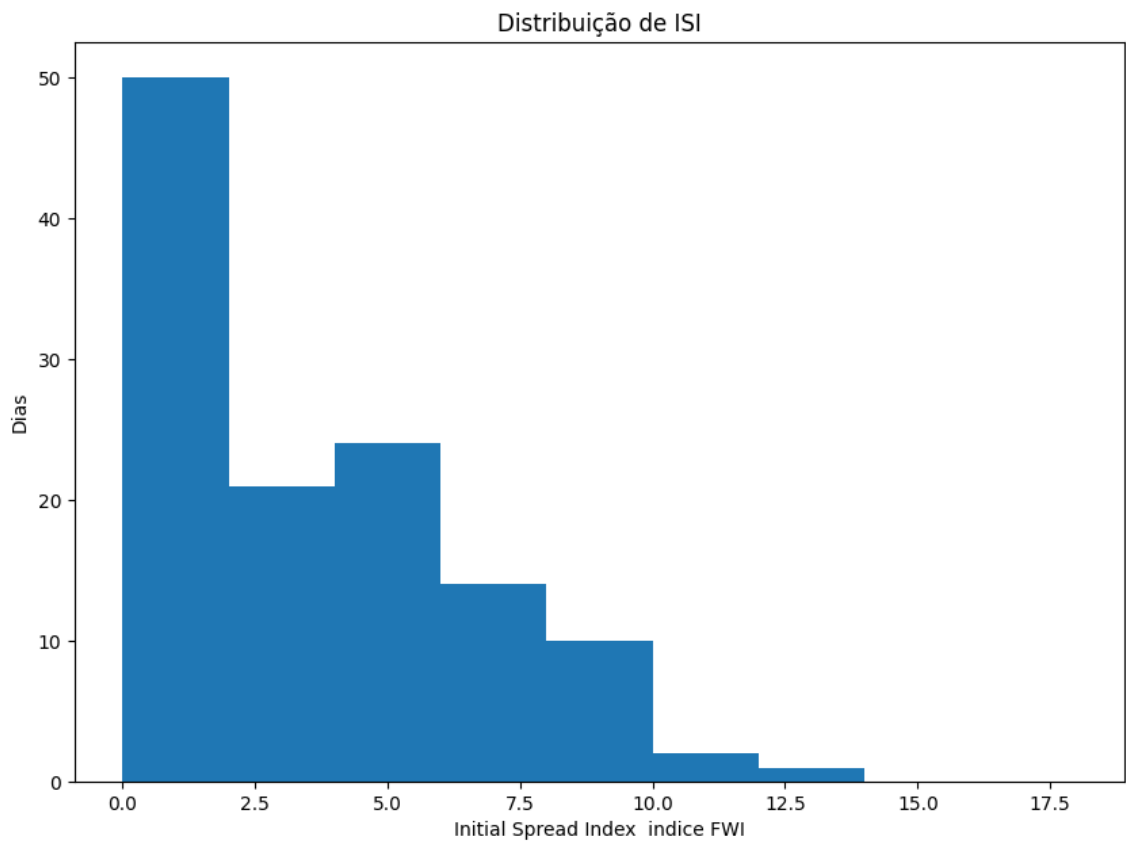
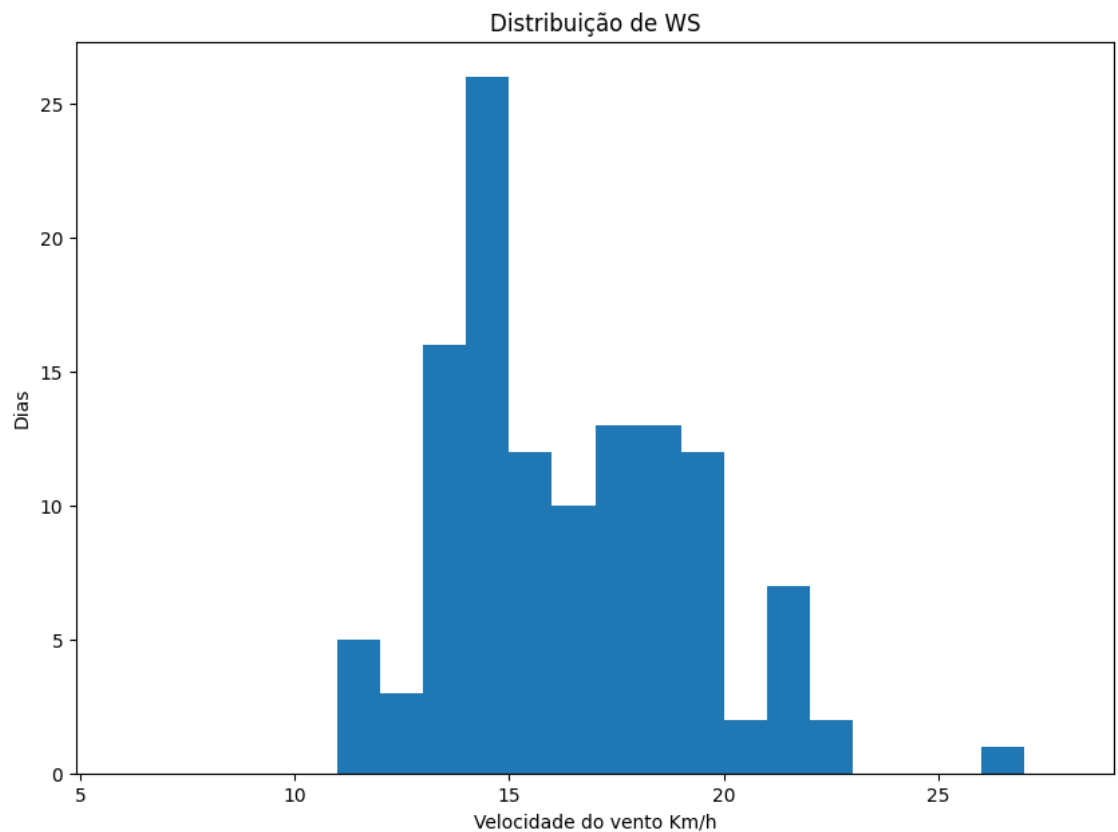
Analisando os resultados obtidos podemos perceber que a normalização Z-Score proporciona uma distribuição diferente dos pontos, enquanto no gráfico de PCA para Min-Max, podemos notar um agrupamento em patamares em relação ao componente 2, na normalização Z-Score temos os pontos mais distribuídos por esse eixo.

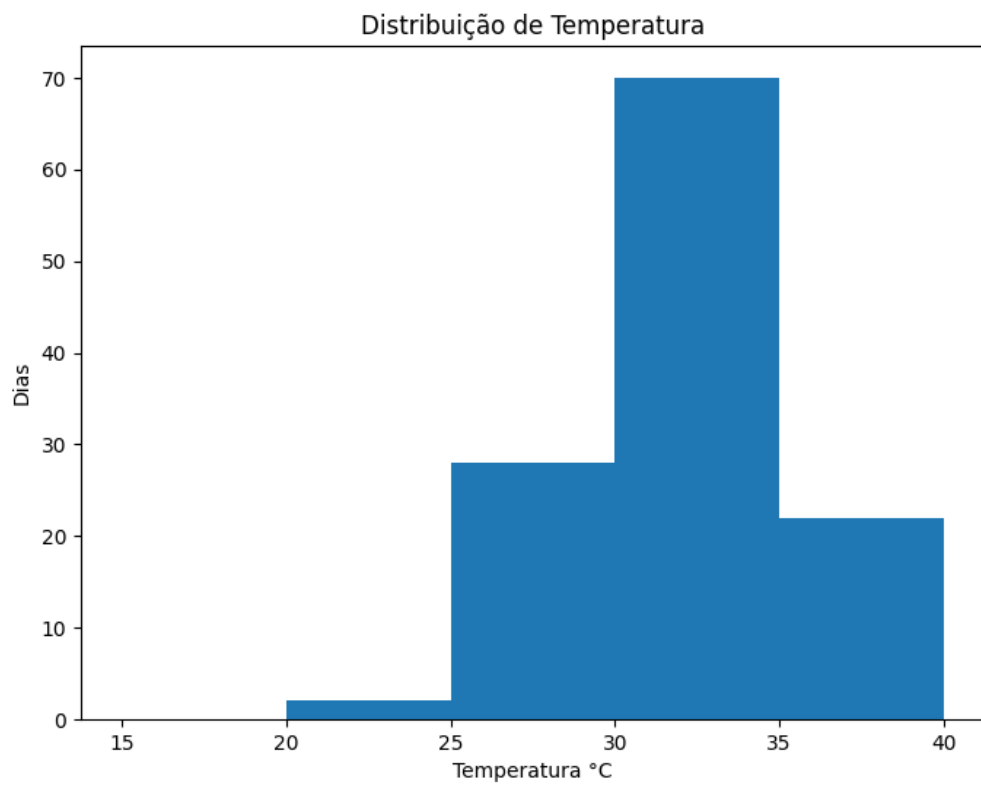
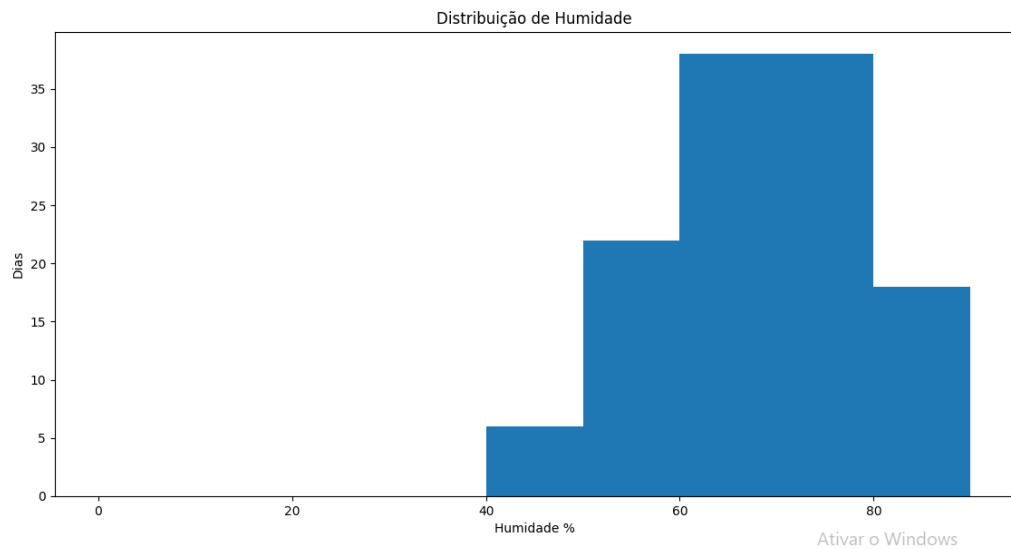
Porém o padrão de classificação permaneceu o mesmo, 0 (não incêndio), assumindo valores negativos para o componente 1, enquanto 1 (incêndio) apresentou valores positivos para este componente.

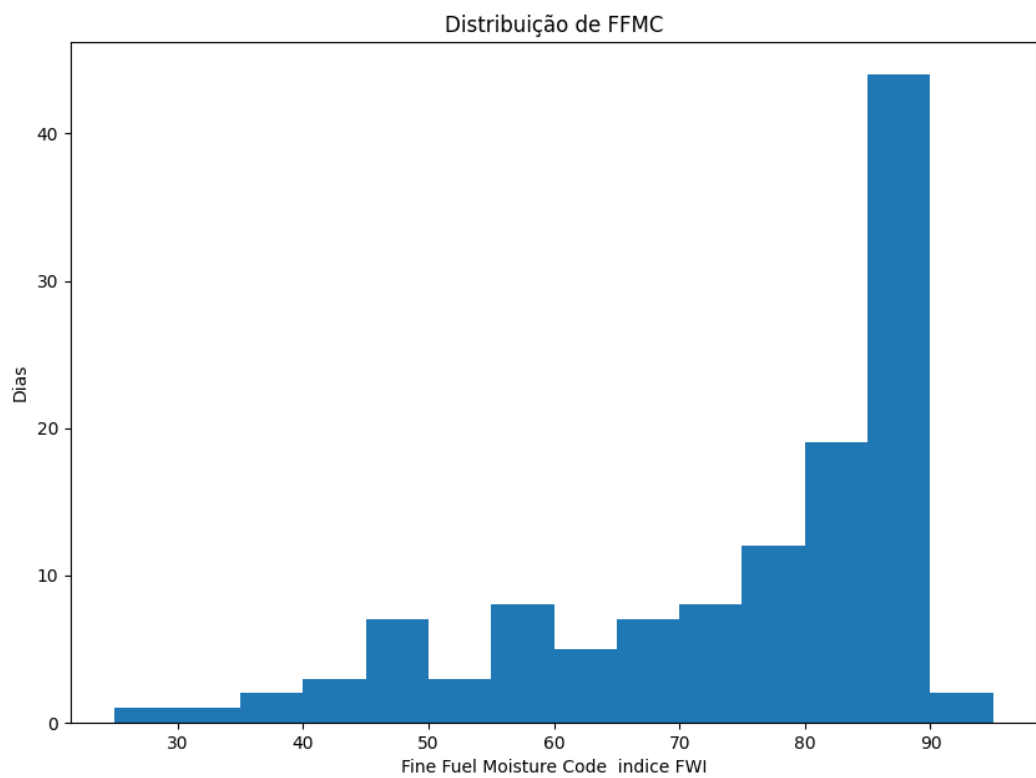
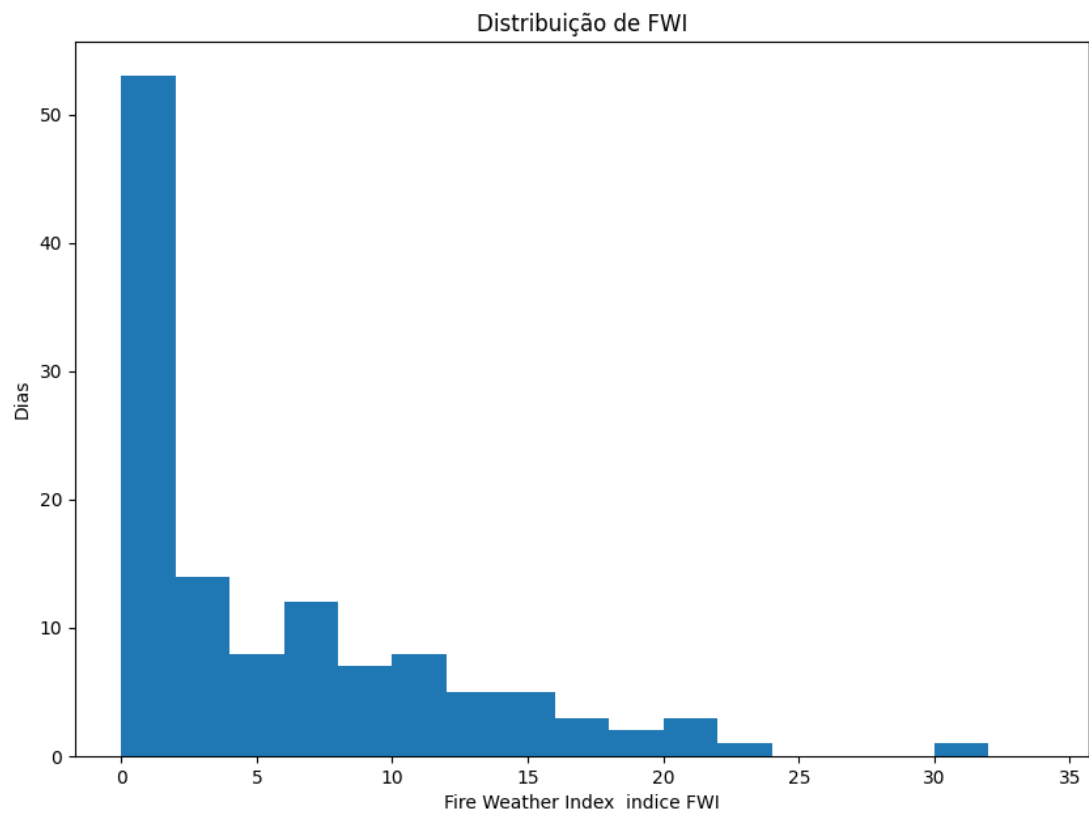
3- Análise descritiva dos dados

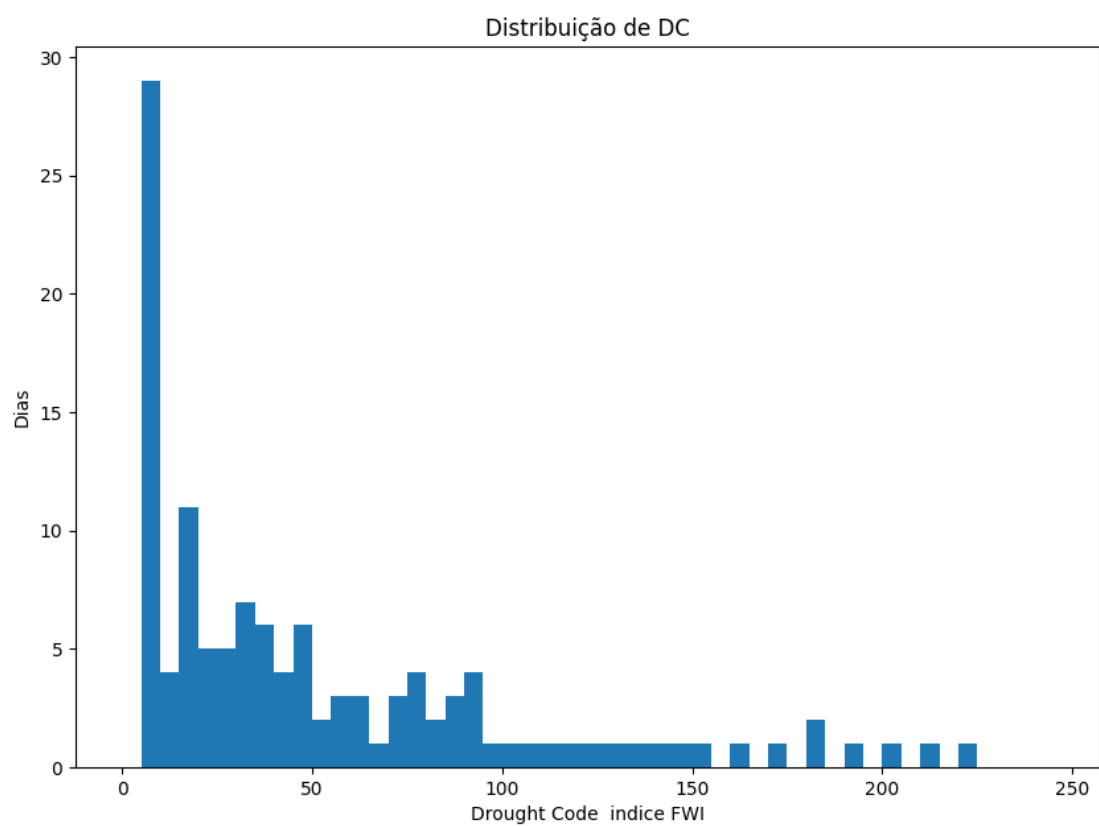
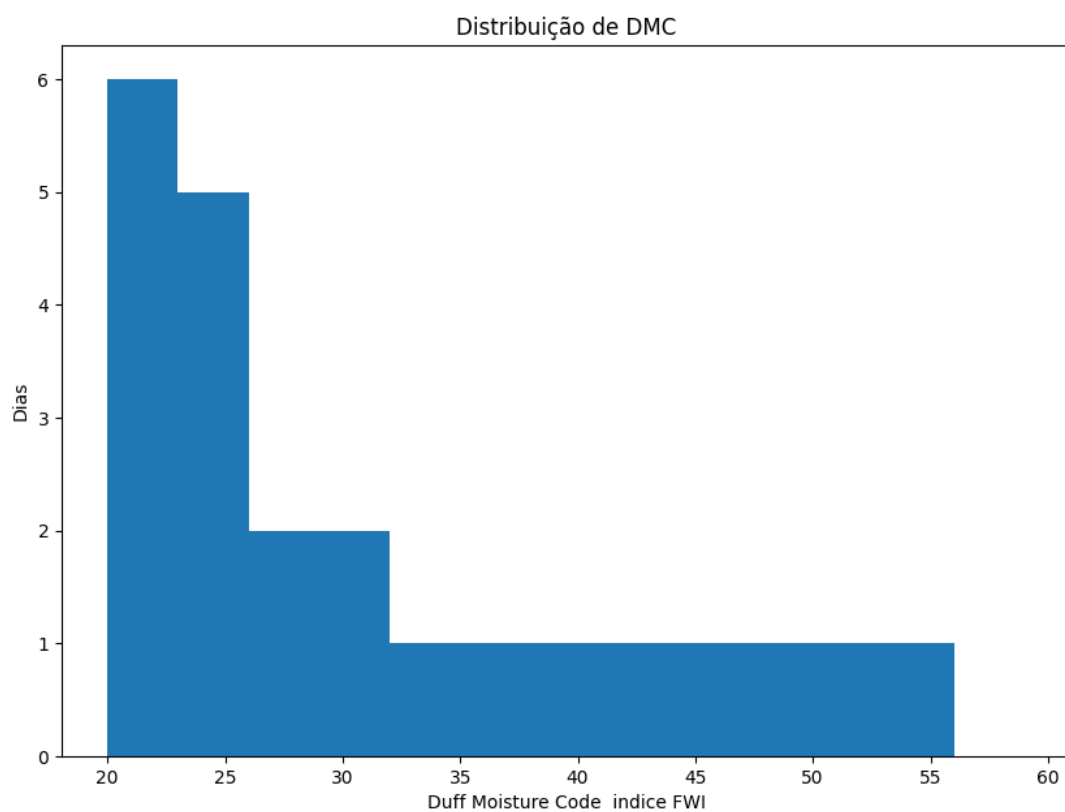
Para uma melhor análise, a base de dados foi exposta à diversas medidas de resumo:

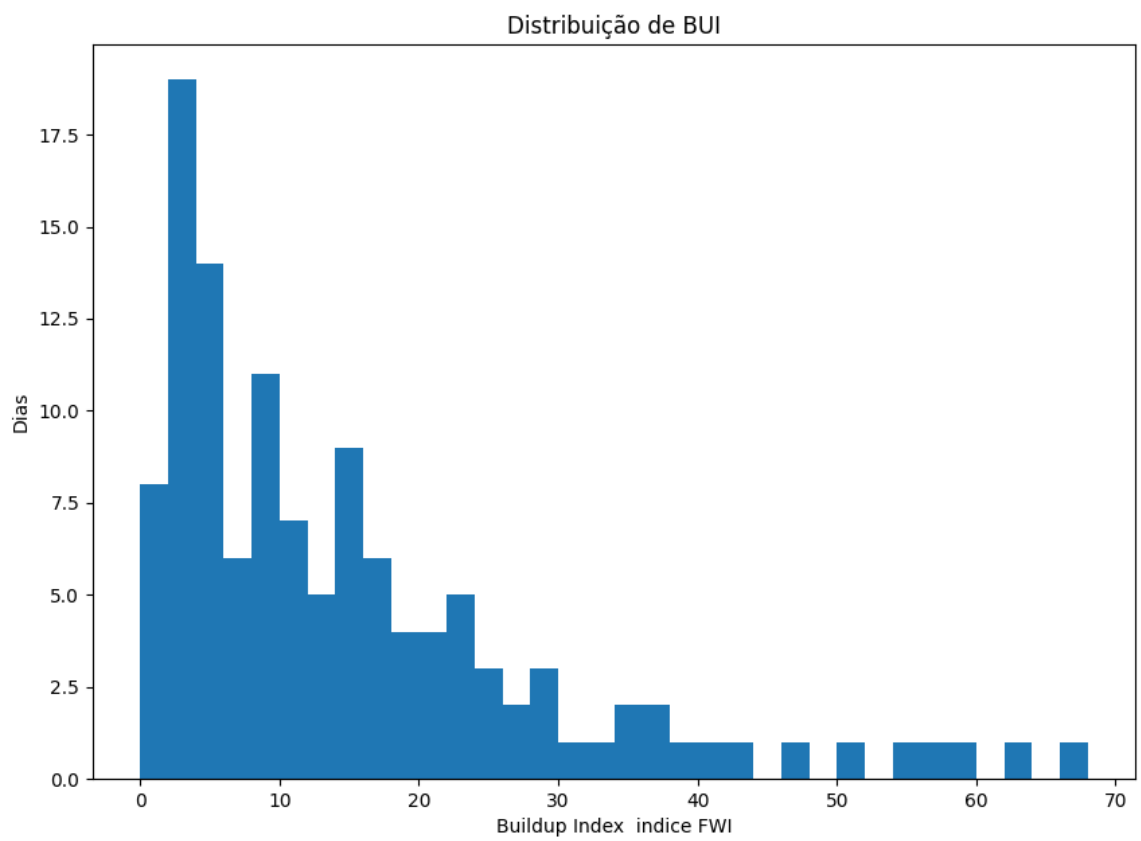
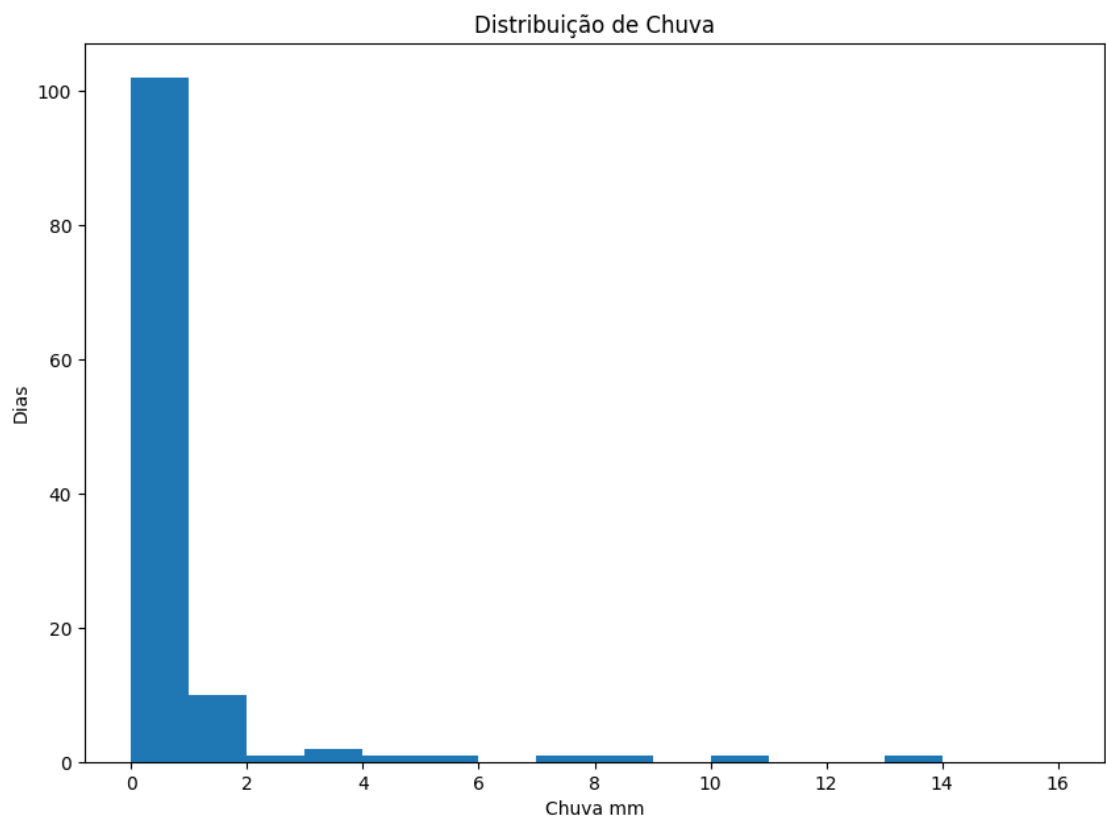
Distribuição de frequência dos dados











Medidas de tendência central

Para medidas de tendência central foram escolhidas média e moda, pois com a média podemos analisar como os extremos estão afetando a medição em relação à moda.

Média

```
Media Temperatura: 31.278688524590162
Media RH: 68.39344262295081
Media WS: 15.926229508196721
Media Chuva: 0.8426229508196721
Media FFMC: 74.67295081967214
Media DMC: 12.314754098360657
Media DC: 53.160655737704914
Media ISI: 3.6557377049180326
Media BUI: 15.426229508196721
Media FWI: 5.577868852459017
```

Moda

```
Moda Temperatura: 0 31.0
dtype: float64
Moda RH: 0 78.0
dtype: float64
Moda WS: 0 14.0
dtype: float64
Moda Chuva: 0 0.0
dtype: float64
Moda FFMC: 0 88.9
1 89.4
dtype: float64
Moda DMC: 0 1.9
1 12.5
dtype: float64
Moda DC: 0 7.5
dtype: float64
Moda ISI: 0 1.1
dtype: float64
Moda BUI: 0 3.0
dtype: float64
Moda FWI: 0 0.0
1 0.1
2 0.5
dtype: float64
```

Medidas de dispersão

Desvio Padrão

```
Desvio padrão Temperatura: 2.9965884047480618
Desvio padrão RH: 11.003799026203511
Desvio padrão WS: 2.852193683184015
Desvio padrão Chuva: 2.4092081629368325
Desvio padrão FFMC: 15.558713308488693
Desvio padrão DMC: 11.274359648167286
Desvio padrão DC: 51.77826462107204
Desvio padrão ISI: 3.021768109757118
Desvio padrão BUI: 14.47430212582062
Desvio padrão FWI: 6.343051280961404
```

Variância

```
Variância Temperatura: 8.979542067470533
Variância RH: 121.08359300907735
Variância WS: 8.135008806394797
Variância Chuva: 5.804283972361468
Variância FFMC: 242.07355981574315
Variância DMC: 127.11118547622276
Variância DC: 2680.9886871697604
Variância ISI: 9.131082509145104
Variância BUI: 209.5054220295353
Variância FWI: 40.23429955290611
```

Medidas de posição relativa

```
Data:
  Temperatura  RH  WS  Chuva  FFMC  DMC  DC  ISI  BUI  FWI  Classes
0      29.0  57.0  18.0   0.0  65.7  3.4  7.6  1.3  3.4  0.5      0
1      29.0  61.0  13.0   1.3  64.4  4.1  7.6  1.0  3.9  0.4      0
2      26.0  78.0  22.0  13.1  47.1  2.5  7.1  0.3  2.7  0.1      0
3      25.0  89.0  13.0   2.5  28.6  1.3  6.9  0.0  1.7  0.0      0
4      27.0  77.0  16.0   0.0  64.8  3.0  14.2  1.2  3.9  0.5      0
..      ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
117     31.0  54.0  11.0   0.0  82.0  6.0  16.3  2.5  6.2  1.7      0
118     31.0  66.0  11.0   0.0  85.7  8.3  24.9  4.0  9.0  4.1      1
119     32.0  47.0  14.0   0.7  77.5  7.1  8.8  1.8  6.8  0.9      0
120     26.0  78.0  16.0   1.8  47.4  2.9  7.7  0.3  3.0  0.1      0
121     25.0  78.0  14.0   1.4  45.0  1.9  7.5  0.2  2.4  0.1      0

[122 rows x 11 columns]
Kurtosis:
0      0.761623
1      0.811790
2      2.749274
3      6.922108
4      1.306951
..      ...
117     2.350221
118     1.548945
119     2.619934
120     2.971443
121     3.398807
Length: 122, dtype: float64
```


Medidas de associação

Covariância												
	Temperatura	RH	WS	Chuva	FFMC	DMC	DC	ISI	BUI	FWI	Classes	
Temperatura	8.979542	-19.151876	-2.384230	-2.987183	29.913386	17.438829	81.826758	5.576074	22.700068	11.061591	0.674028	
RH	-19.151876	121.083593	6.731744	8.027720	-105.569435	-41.206679	-175.171169	-18.154342	-51.980653	-31.332543	-1.844736	
WS	-2.384230	6.731744	8.135009	2.459369	-10.252425	-0.141878	9.122693	-0.548747	0.835835	-0.031398	-0.211963	
Chuva	-2.987183	8.027720	2.459369	5.804284	-21.496276	-7.208072	-35.384012	-2.667602	-9.483359	-4.476239	-0.404254	
FFMC	29.913386	-105.569435	-10.252425	-21.496276	242.073560	106.741890	463.399588	37.145735	135.280054	66.670140	5.805751	
DMC	17.438829	-41.206679	-0.141878	-7.208072	106.741890	127.111185	561.696040	25.554708	162.462502	64.488263	3.490327	
DC	81.826758	-175.171169	9.122693	-35.384012	463.399588	561.696040	2680.988687	112.275682	736.379140	285.706808	15.289432	
ISI	5.576074	-18.154342	-0.548747	-2.667602	37.145735	25.554708	112.275682	9.131083	32.661832	18.095045	1.261252	
BUI	22.700068	-51.980653	0.835835	-9.483359	135.280054	162.462502	736.379140	32.661832	209.505422	82.668437	4.449194	
FWI	11.061591	-31.332543	-0.031398	-4.476239	66.670140	64.488263	285.706808	18.095045	82.668437	40.234300	2.418229	
Classes	0.674028	-1.844736	-0.211963	-0.404254	5.805751	3.490327	15.289432	1.261252	4.449194	2.418229	0.251795	

Correlação												
	Temperatura	RH	WS	Chuva	FFMC	DMC	DC	ISI	BUI	FWI	Classes	
Temperatura	1.000000	-0.580820	-0.278960	-0.413771	0.641601	0.516177	0.527376	0.615801	0.523362	0.581959	0.448257	
RH	-0.580820	1.000000	0.214489	0.302813	-0.616626	-0.332149	-0.307449	-0.545980	-0.326363	-0.448905	-0.334093	
WS	-0.278960	0.214489	1.000000	0.357907	-0.231033	-0.004412	0.061773	-0.063670	0.020246	-0.001736	-0.148101	
Chuva	-0.413771	0.302813	0.357907	1.000000	-0.573476	-0.265371	-0.283652	-0.366425	-0.271951	-0.292914	-0.334392	
FFMC	0.641601	-0.616626	-0.231033	-0.573476	1.000000	0.608512	0.575221	0.790086	0.600707	0.675553	0.743637	
DMC	0.516177	-0.332149	-0.004412	-0.265371	0.608512	1.000000	0.962193	0.750098	0.995551	0.901759	0.616951	
DC	0.527376	-0.307449	0.061773	-0.283652	0.575221	0.962193	1.000000	0.717591	0.982554	0.869911	0.588464	
ISI	0.615801	-0.545980	-0.063670	-0.366425	0.790086	0.750098	0.717591	1.000000	0.746761	0.944062	0.831796	
BUI	0.523362	-0.326363	0.020246	-0.271951	0.600707	0.995551	0.982554	0.746761	1.000000	0.900417	0.612576	
FWI	0.581959	-0.448905	-0.001736	-0.292914	0.675553	0.901759	0.869911	0.944062	0.900417	1.000000	0.759758	
Classes	0.448257	-0.334093	-0.148101	-0.334392	0.743637	0.616951	0.588464	0.831796	0.612576	0.759758	1.000000	

4- Análise de grupos

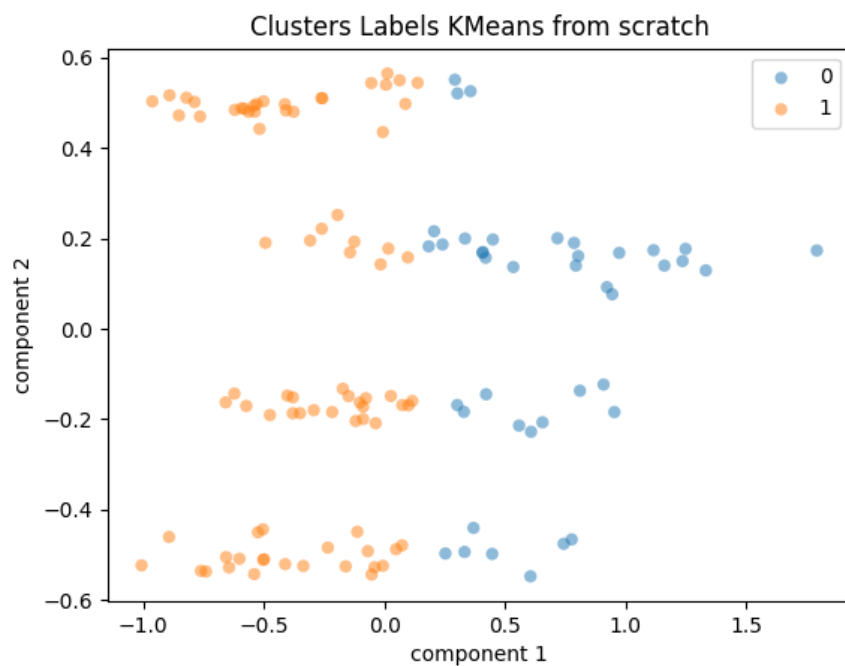
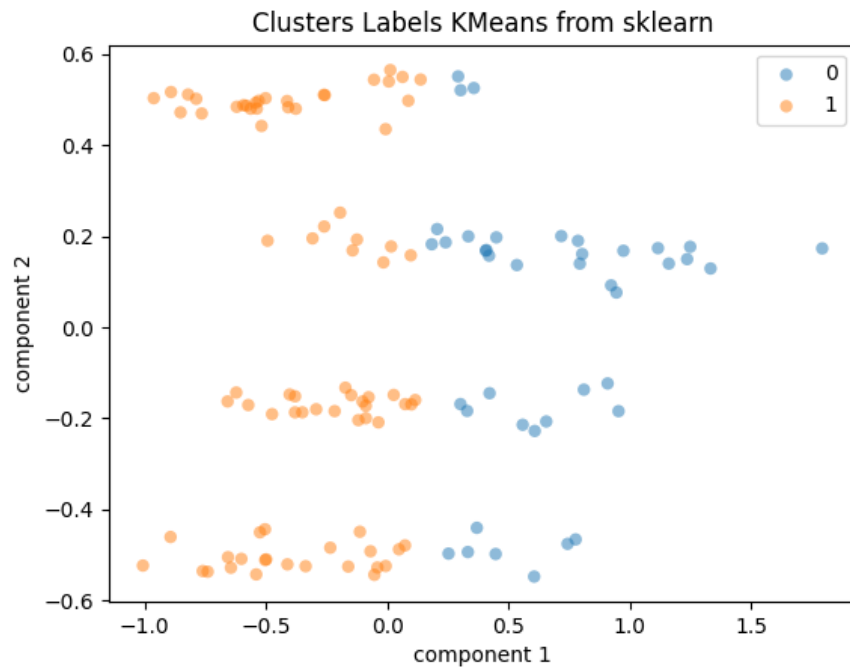
Para o agrupamento, foi utilizado o algoritmo K-Means, este algoritmo foi aplicado na base de dados, tanto normalizada pelo método Min-Max, quanto pelo método Z-Score.

Para cada uma das normalizações o algoritmo foi rodado duas vezes, uma fazendo o agrupamento em dois grupos, e outro em quatro grupos ($k = 2$ e $k = 4$).

- K-Means com normalização MinMax, com $k = 2$

```
Explained variance per component:
[0.46547484 0.2005662 ]
(122, 14)
(122, 2)

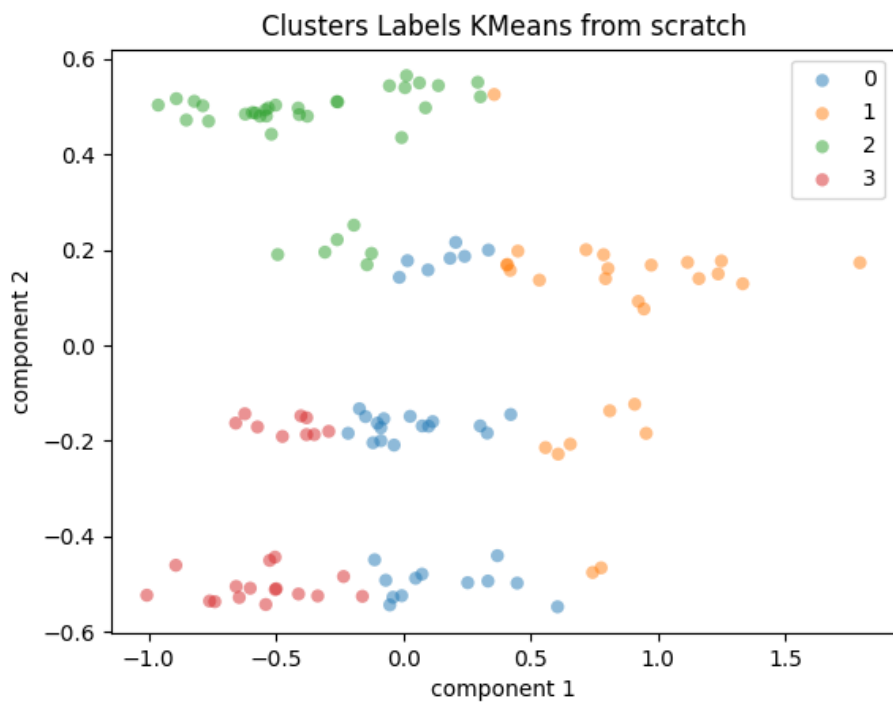
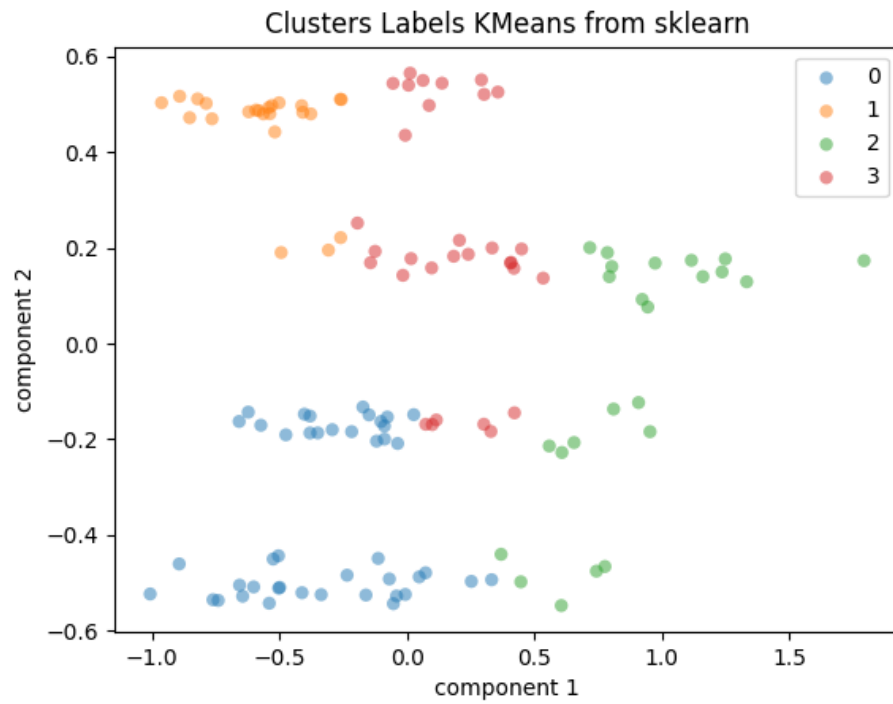
29.709202507870565
For n_clusters = 2, silhouette score is 0.41447923021889954)
```



- K-Means com normalização MinMax, com k=4.

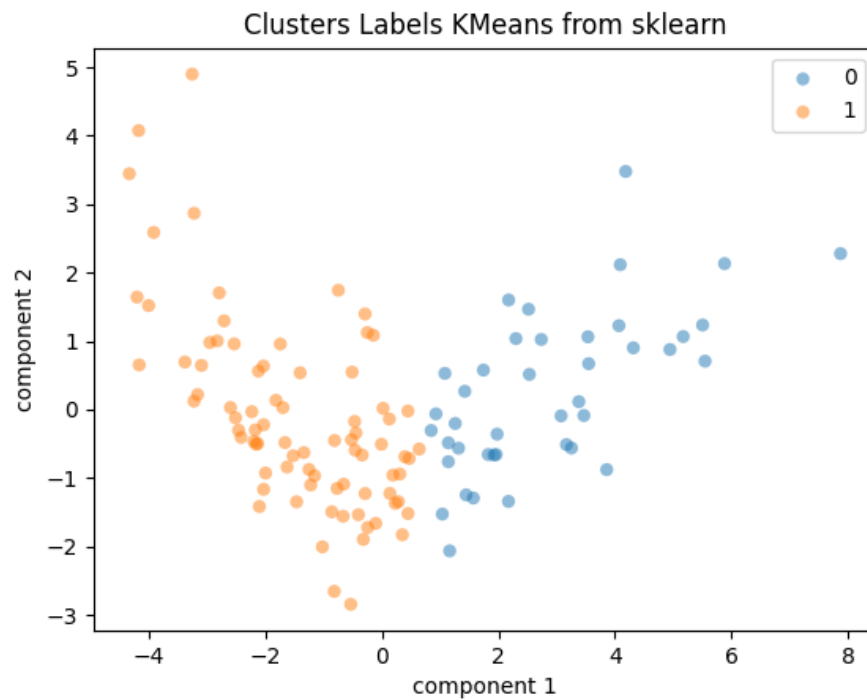
```
Explained variance per component:
[0.46547484 0.2005662 ]
(122, 14)
(122, 2)

13.236946205690224
For n_clusters = 4, silhouette score is 0.43474283483216786)
```



- K-Means com normalização Z-Score, com $k=2$.

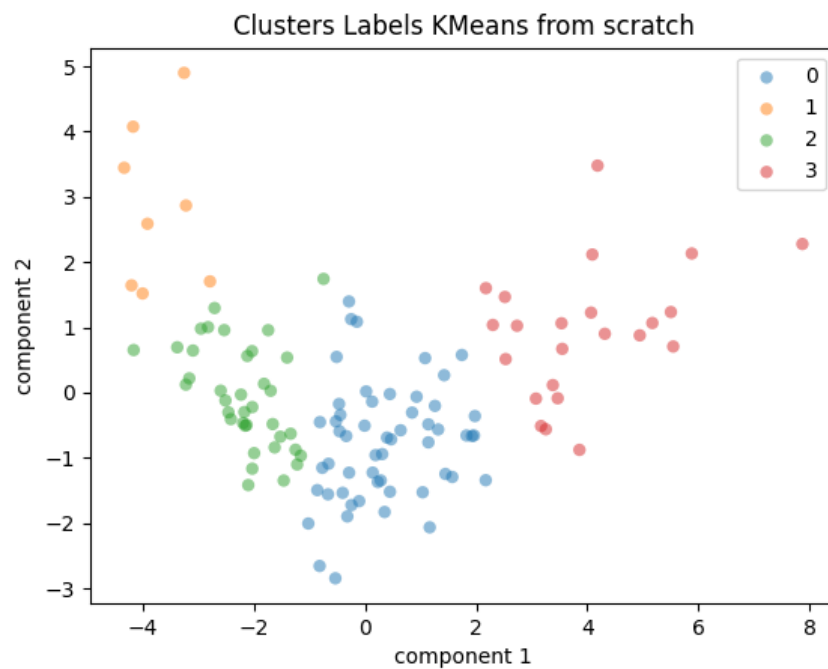
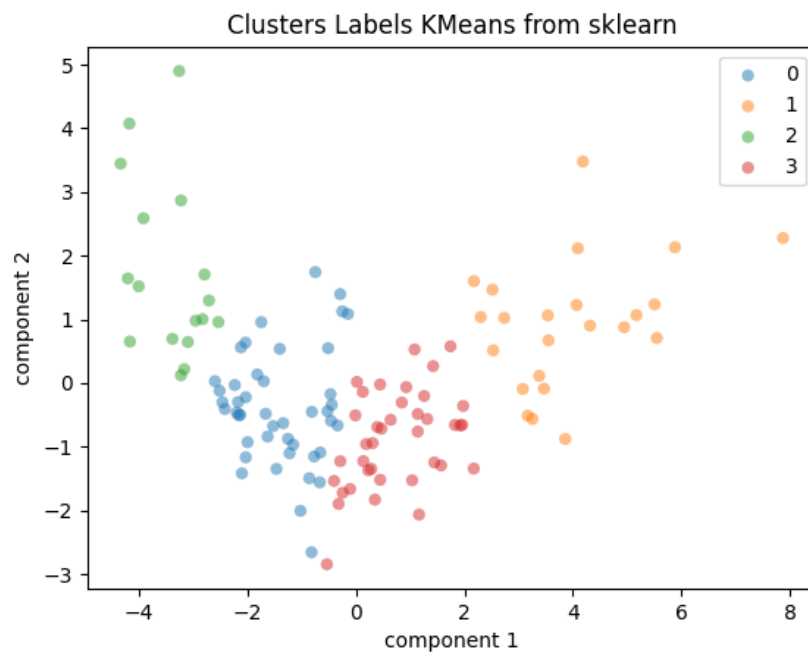
```
Explained variance per component:  
[0.51682267 0.14727808]  
(122, 14)  
(122, 2)  
  
465.0331655100531  
For n_clusters = 2, silhouette score is 0.46587213094764907)
```



- K-Means com normalização Z-Score, com k=4.

```
Explained variance per component:
[0.51682267 0.14727808]
(122, 14)
(122, 2)

201.0132484630516
For n_clusters = 4, silhouette score is 0.3833762440863893)
```



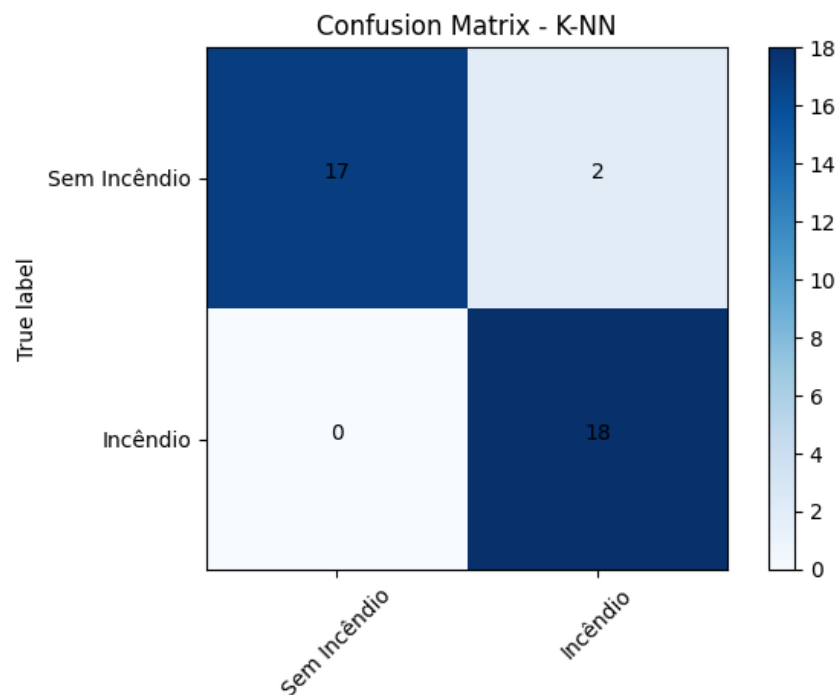
Após a análise dos resultados, é possível observar que a execução que mostrou uma maior efetividade, dada pelo valor de silhueta apresentado (0,46), foi a normalização Z-Score com dois grupos (k=2).

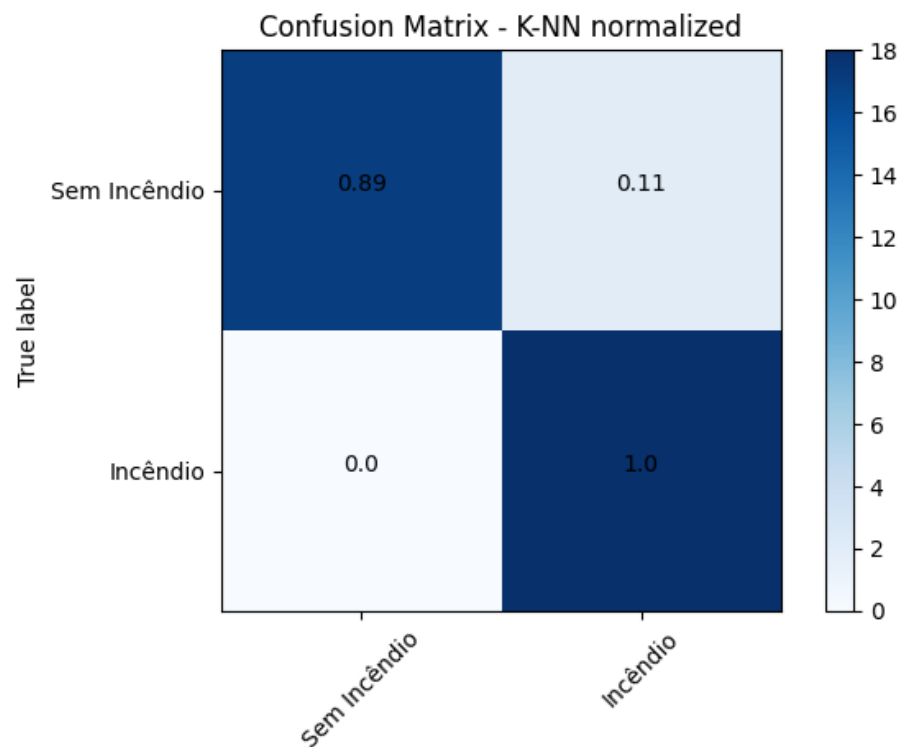
5- Classificação

Para a classificação foi utilizado o método de normalização Z-Score, pois o mesmo apresentou melhores resultados nas etapas realizadas previamente.

KNN

```
Total samples: 122
Total train samples: 85
Total test samples: 37
Acurracy K-NN from scratch: 94.59%
F1 Score K-NN from scratch: 0.95%
Confusion matrix, without normalization
[[17  2]
 [ 0 18]]
Normalized confusion matrix
[[0.89473684 0.10526316]
 [0.         1.         ]]
Acurracy K-NN from sk-learn: 94.59%
F1 Score K-NN from sk-learn: 0.95%
Confusion matrix, without normalization
[[17  2]
 [ 0 18]]
Normalized confusion matrix
[[0.89473684 0.10526316]
 [0.         1.         ]]
```



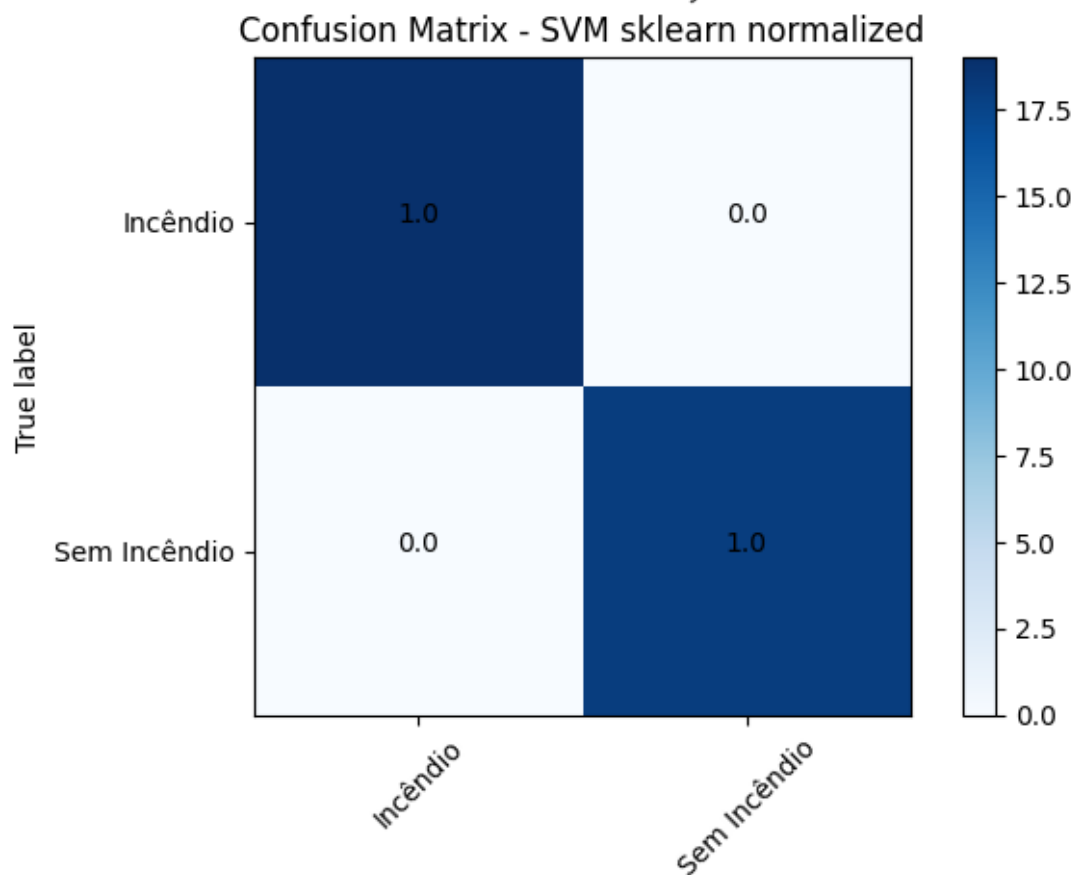
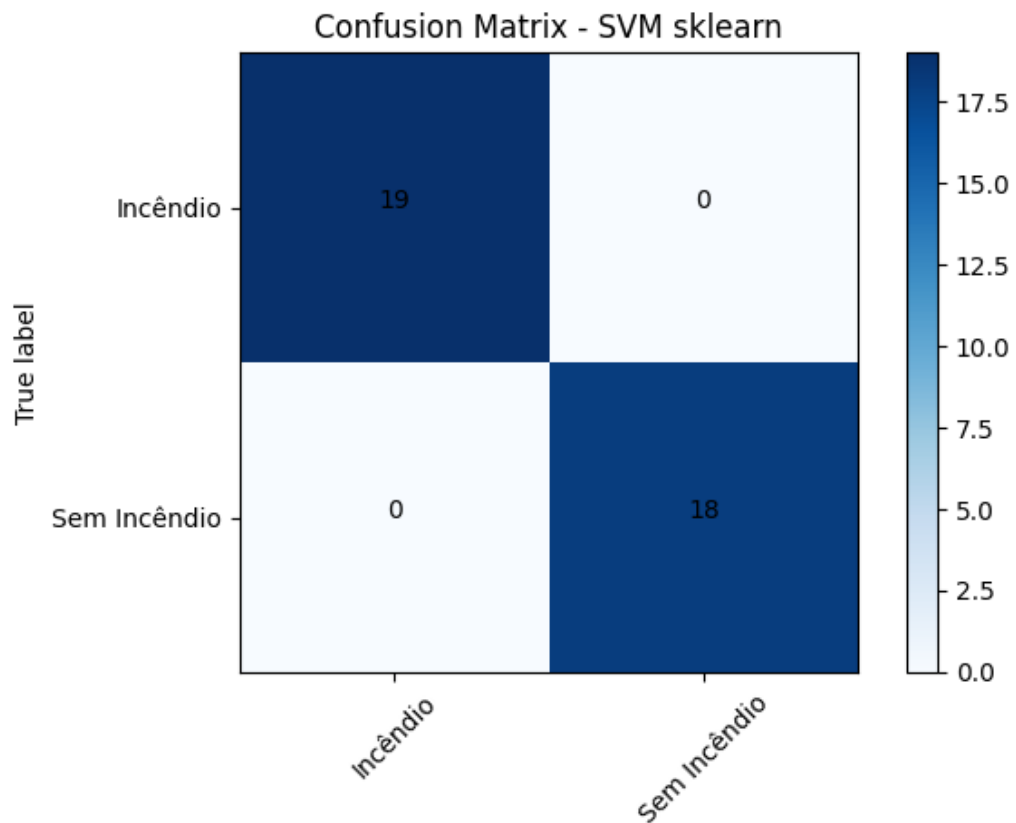


Como foi possível observar nos dados obtidos, a matriz de confusão gerada utilizando sklearn, para k=5, apresentou a mesma acurácia na predição.

SVM

```
Total samples: 122
Total train samples: 85
Total test samples: 37
```

```
Qtd Support vectors:
[21 17]
Acurracy SVM from sk-learn: 100.00%
F1 Score SVM from sk-learn: 1.00%
Confusion matrix, without normalization
[[19  0]
 [ 0 18]]
Normalized confusion matrix
[[1. 0.]
 [0. 1.]]
[0.53846154 0.53846154 0.58333333 0.5      0.5      0.5
 0.5      0.5      0.5      0.5      ]
Média cross validation = 0.5160256410256411
```



Como foi possível observar nos dados obtidos, o algoritmo obteve acurácia de 100% dos testes de predição de incêndios feitos, utilizando o kernel = rfb como classificador SVM.

6- Discussão

Como podemos comparar durante todo os processos realizados durante o trabalho, alguns métodos se sobressaíram a outros, para o tratamento desta base de dados específica, para este problema tivemos uma vantagem do método de normalização Z-Score, sobre o Min-Max, quando aplicado o agrupamento K-Means.

Outra comparação que podemos fazer é a da acurácia fornecida pelo método SVM, que alcançou 100% de acurácia nas predições, quando configurado com kernel=rfb, enquanto o KNN, com k=5 (n° de objetos próximos para comparação), alcançou 94,59 %.