

Wyniki Analizy Danych

1. Charakterystyka danych

Zmienne w zbiorze danych (dane zaznaczone na czerwono nie zostały poddane dyskretyzacji):

- **X** - zmienna zawiera się w przedziale liczb całkowitych od 1 do 9
- **Y** - zmienna zawiera się w przedziale liczb całkowitych od 2 do 9
- month
- day
- FFMC
- DMC
- DC
- ISI
- temp
- RH
- wind
- rain
- area - Atrybut Decyzyjny

Month — zmienna zawiera jedną z 12 wartości tekstowych określających miesiąc (skrótowa nazwa w języku angielskim). Każda nazwa miesiąca została zmieniona na liczbę całkowitą według schematu:

“jan” = 1

“feb” = 2

...

“dec” = 12

Day – zmienna zawiera jedną z 7 wartości tekstowych określających dzień tygodnia (skrótowa nazwa w języku angielskim). Każda nazwa tygodnia została zmieniona na liczbę całkowitą według schematu:

“mon” = 1

“tue” = 2

...

“sun” = 7

FFMC – zmienna ciągła, zmiennoprzecinkowa. Jej podział został opracowany na podstawie tabeli ze strony: [link](#) . Podział:

<0-76) – 1 (Low)

<76-84) – 2 (Moderate)

<84-88) – 3 (High)

<88-91) – 4 (Very High)

<91-inf) – 5 (Extreme)

DMC – zmienna ciągła, zmiennoprzecinkowa. Jej podział został opracowany na podstawie tabeli ze strony: [link](#) . Podział:

- <0-21) – 1 (Low)
- <21-27) – 2 (Moderate)
- <27-40) – 3 (High)
- <40-60) – 4 (Very High)
- <60-inf) – 5 (Extreme)

DC – zmienna ciągła, zmiennoprzecinkowa. Jej podział został opracowany na podstawie tabeli ze strony: [link](#) . Podział:

- <0-79) – 1 (Low)
- <79-189) – 2 (Moderate)
- <189-299) – 3 (High)
- <299-424) – 4 (Very High)
- <424-inf) – 5 (Extreme)

ISI – zmienna ciągła, zmiennoprzecinkowa. Jej podział został opracowany na podstawie tabeli ze strony: [link](#) . Podział:

- <0-1.5) – 1 (Low)
- <1.5-4.0) – 2 (Moderate)
- <4.0-8.0) – 3 (High)
- <8.0-15) – 4 (Very High)
- <15-inf) – 5 (Extreme)

temp – zmienna ciągła, zmiennoprzecinkowa. Jej podział został zdefiniowany subiektywnie. Aktualny przedział wartości został podzielony na 4 równe części, dzięki czemu powstały następujące grupy:

- <-inf - 11.5) – 0 (Cold)
- <11.5-19) – 1 (Moderate)
- <19-26) – 2 (Warm)
- <26-inf) – 3 (Hot)

RH – zmienna całkowita. Ponieważ zmienna przedstawia poziom wilgotności, została ona podzielona na 5 równych poziomów, które się zmieniają co 20 punktów procentowych:

- <0-20) – 1
- <20-40) – 2
- <40-60) – 3
- <60-80) – 4
- <80-100) – 5

wind – zmienna ciągła przedstawiająca prędkość wiatru w km/h. Do jej podziału została wykorzystana skala Beauforta:

- <0-1) – 1 (Calm)
- <1-5) – 2 - (Light Air)
- <5-11) – 3 - (Light Breeze)

rain – zmienna ciągła. Podział wartości również jak w przypadku temperatury, został podzielony subiektywnie:

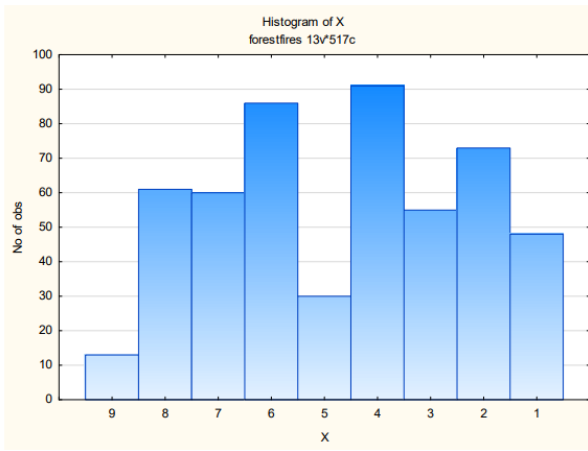
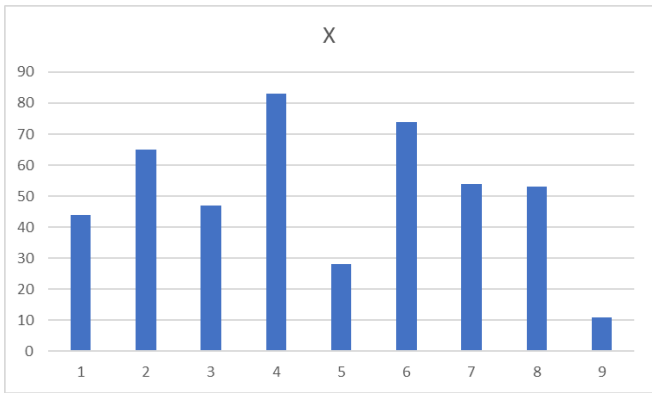
(-inf - 0> – 0 (No rain)
 (0 – 1> – 1 (Little rain)
 (1 – 3> – 2 (Normal rain)
 (3 – 6> – 3 (Heavy rain)

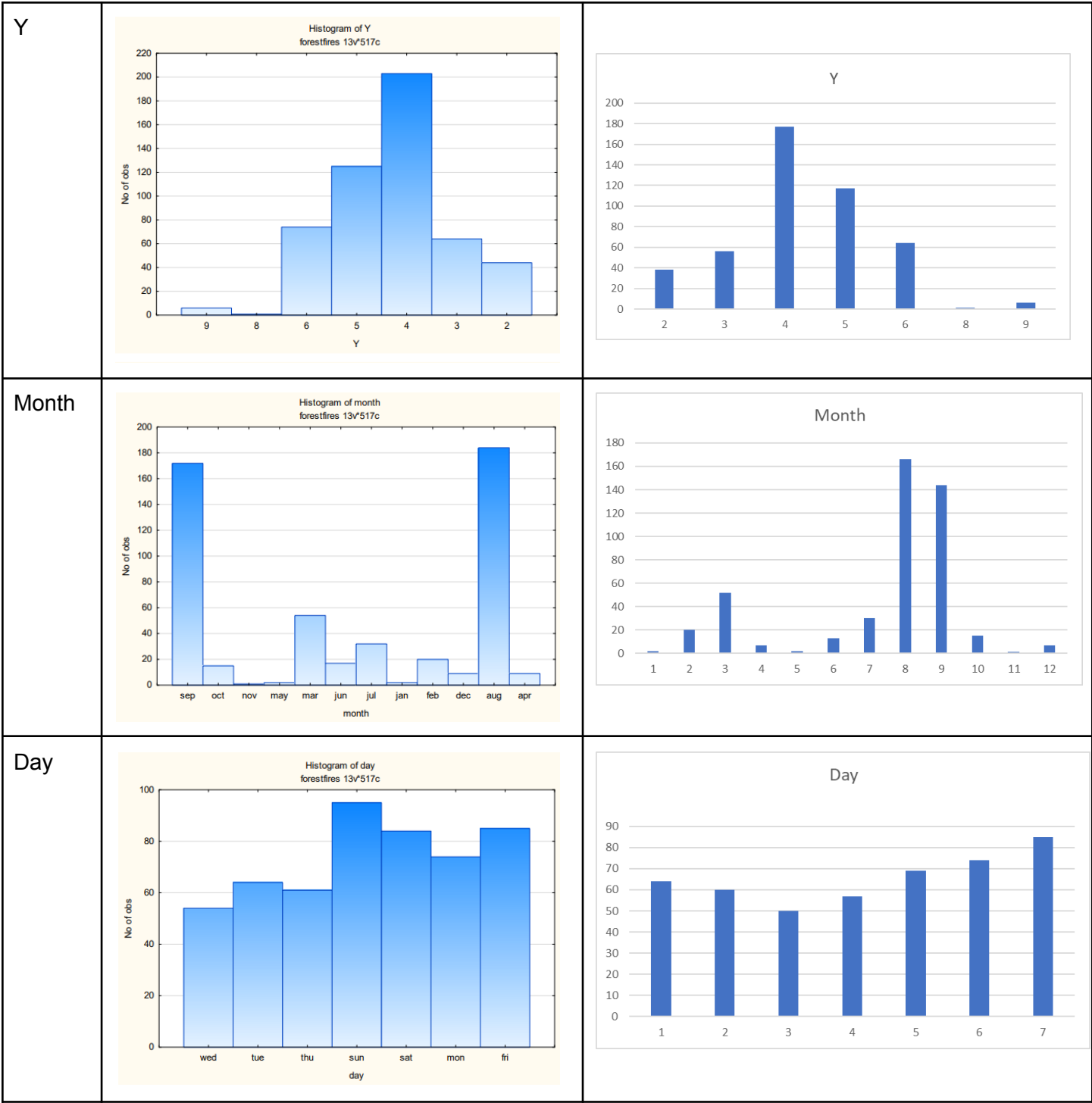
area – zmienna ciągła. Na podstawie posiadanych danych, podział również został podzielony subiektywnie. Ze względu na posiadane wartości, podział nie jest równomierny, starano się go podzielić logarymicznie. tzn: kolejne kategorie posiadają większy zakres zmiennej niż poprzednie:

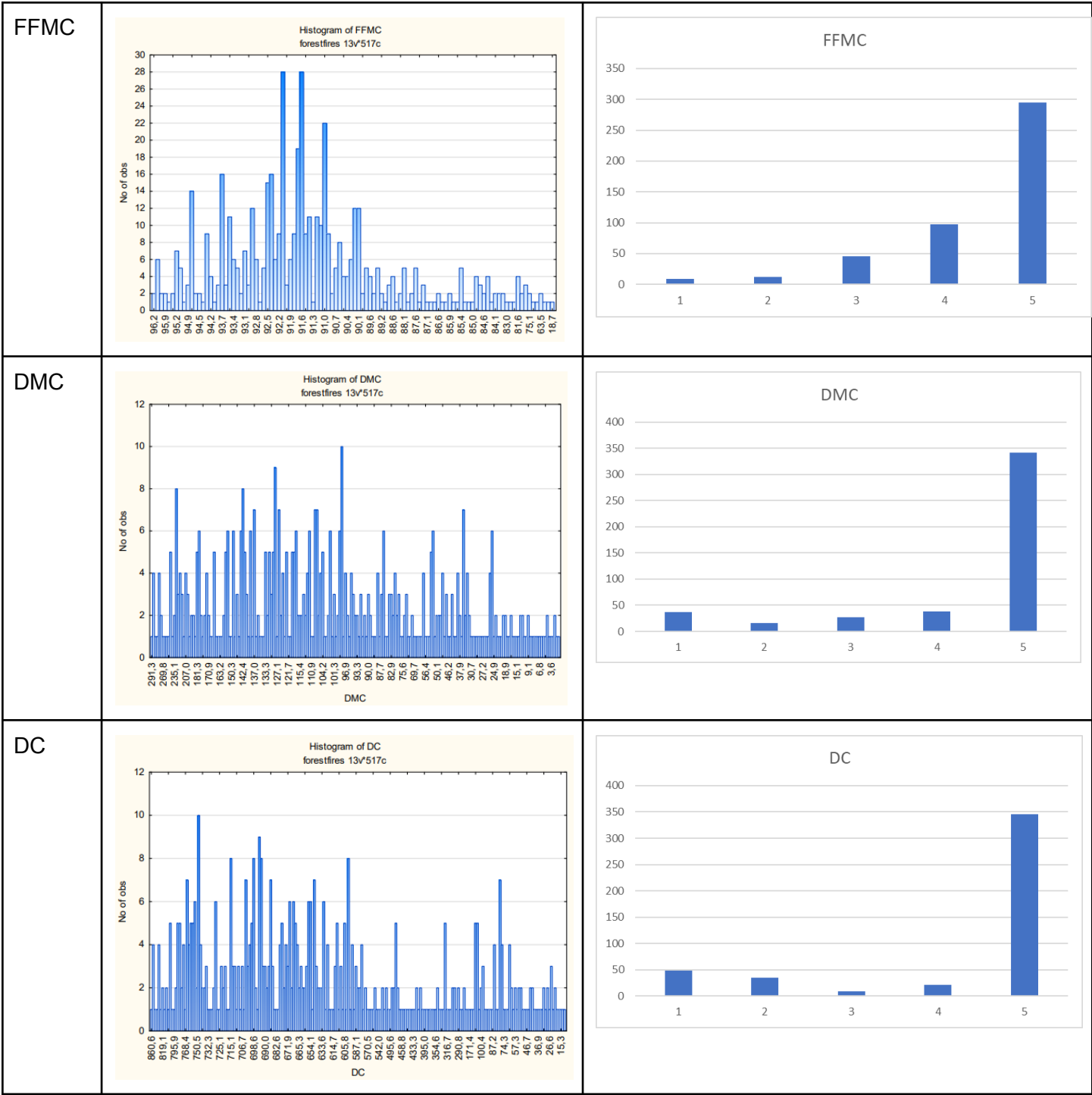
(-inf - 0> – 0 (no area)
 (0 - 5> – 1 (small area)
 (5 - 20> – 2 (medium area)
 (20 - 100> – 3 (big area)
 (100 - inf> – 4 (catastrophic big area)

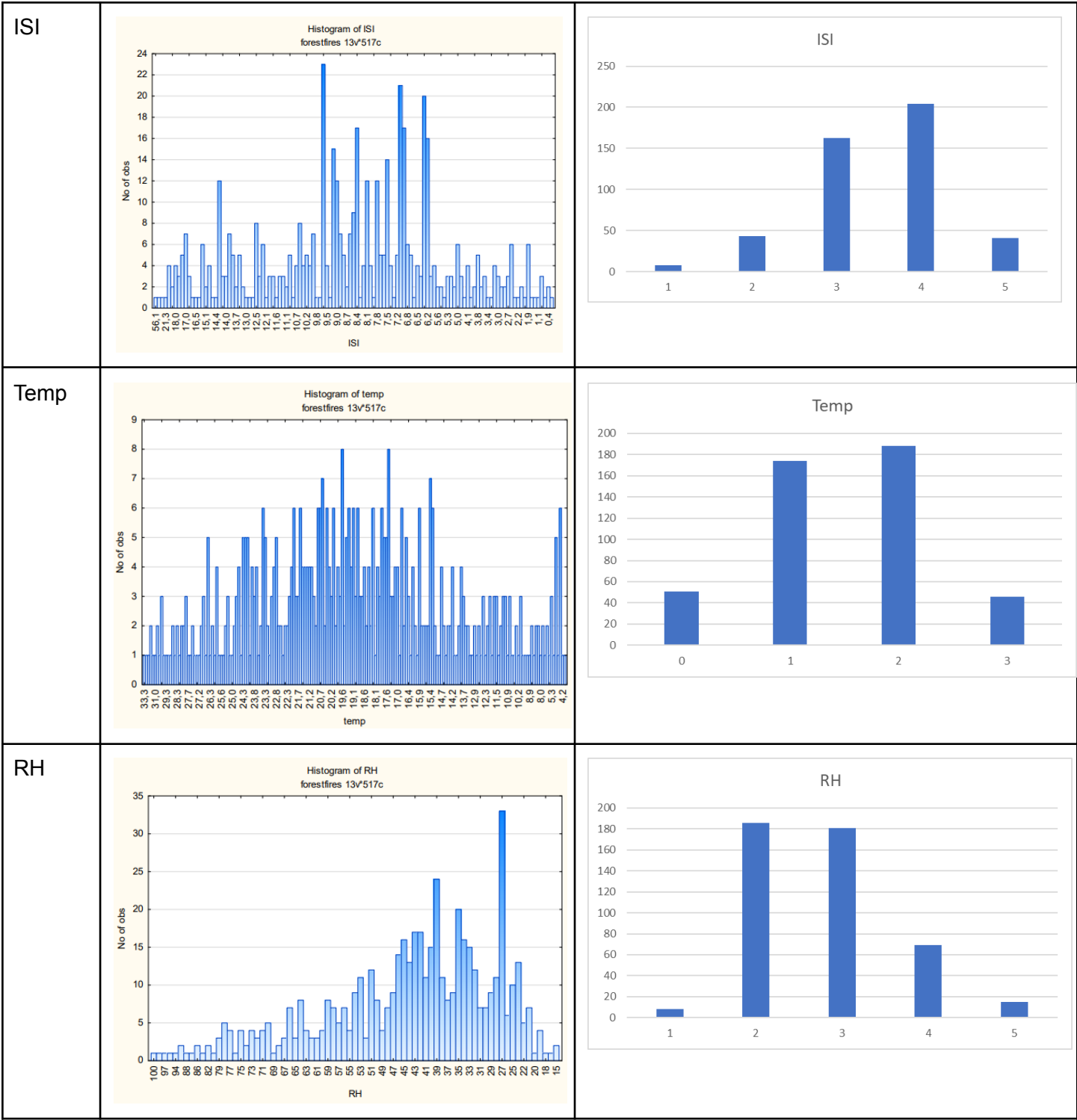
W danych wystąpiło 13 duplikatów oraz 16 sprzeczności, które zostały usunięte. Nie wystąpiły braki w danych.

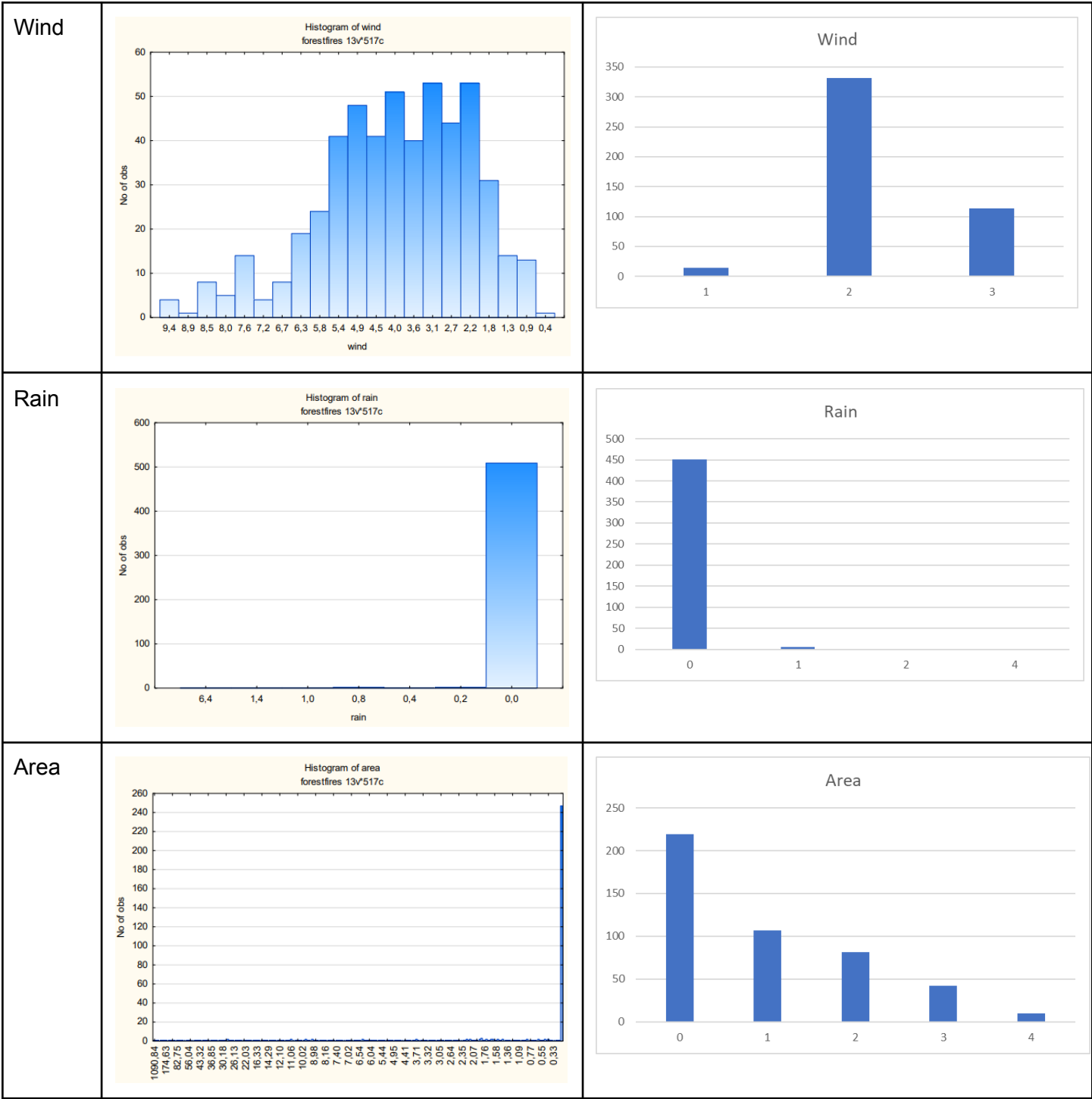
2. Rozkład danych

Atrybut	Przed dyskretyzacją	Po dyskretyzacji																																								
X	<div><p>Histogram of X forestfires 13V517c</p><table><thead><tr><th>X</th><th>No of obs</th></tr></thead><tbody><tr><td>9</td><td>13</td></tr><tr><td>8</td><td>61</td></tr><tr><td>7</td><td>60</td></tr><tr><td>6</td><td>86</td></tr><tr><td>5</td><td>30</td></tr><tr><td>4</td><td>91</td></tr><tr><td>3</td><td>55</td></tr><tr><td>2</td><td>73</td></tr><tr><td>1</td><td>48</td></tr></tbody></table></div>	X	No of obs	9	13	8	61	7	60	6	86	5	30	4	91	3	55	2	73	1	48	<div><p>X</p><table><thead><tr><th>X</th><th>No of obs</th></tr></thead><tbody><tr><td>1</td><td>44</td></tr><tr><td>2</td><td>65</td></tr><tr><td>3</td><td>47</td></tr><tr><td>4</td><td>83</td></tr><tr><td>5</td><td>28</td></tr><tr><td>6</td><td>74</td></tr><tr><td>7</td><td>54</td></tr><tr><td>8</td><td>53</td></tr><tr><td>9</td><td>11</td></tr></tbody></table></div>	X	No of obs	1	44	2	65	3	47	4	83	5	28	6	74	7	54	8	53	9	11
X	No of obs																																									
9	13																																									
8	61																																									
7	60																																									
6	86																																									
5	30																																									
4	91																																									
3	55																																									
2	73																																									
1	48																																									
X	No of obs																																									
1	44																																									
2	65																																									
3	47																																									
4	83																																									
5	28																																									
6	74																																									
7	54																																									
8	53																																									
9	11																																									









3 Tablica Korelacji

Atrybut	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
X	1	0,5299	-0,0682	-0,0187	-0,0688	-0,0689	-0,0664	-0,0349	-0,0203	0,087	0,001	0,0928	0,0895
Y	0,5299	1	-0,0483	-0,0104	-0,0567	-0,0639	-0,0504	-0,0202	0,0122	0,0441	-0,0184	0,0534	0,0476
month	-0,0682	-0,0483	1	-0,043	0,3952	0,6938	0,8914	0,2757	0,3826	-0,0854	-0,1269	0,0051	0,1398
day	-0,0187	-0,0104	-0,043	1	0,011	0,0255	-0,0056	0,054	0,0399	0,0835	0,0088	-0,0528	-0,0012
FFMC	-0,0688	-0,0567	0,3952	0,011	1	0,5839	0,498	0,7264	0,4446	-0,2566	-0,0749	0,0687	0,0332
DMC	-0,0689	-0,0639	0,6938	0,0255	0,5839	1	0,8686	0,4909	0,5258	-0,058	-0,1512	0,0412	0,0703
DC	-0,0664	-0,0504	0,8914	-0,0056	0,498	0,8686	1	0,3916	0,5198	-0,0708	-0,1611	0,0321	0,1088
ISI	-0,0349	-0,0202	0,2757	0,054	0,7264	0,4909	0,3916	1	0,3747	-0,1652	0,0029	0,0441	0,0253
temp	-0,0203	0,0122	0,3826	0,0399	0,4446	0,5258	0,5198	0,3747	1	-0,4921	-0,1761	0,0788	0,1181
RH	0,087	0,0441	-0,0854	0,0835	-0,2566	-0,058	-0,0708	-0,1652	-0,4921	1	0,0671	0,183	-0,048
wind	0,001	-0,0184	-0,1269	0,0088	-0,0749	-0,1512	-0,1611	0,0029	-0,1761	0,0671	1	0,0265	0,0051
rain	0,0928	0,0534	0,0051	-0,0528	0,0687	0,0412	0,0321	0,0441	0,0788	0,183	0,0265	1	-0,0115
area	0,0895	0,0476	0,1398	-0,0012	0,0332	0,0703	0,1088	0,0253	0,1181	-0,048	0,0051	-0,0115	1

Współczynnik niezbalansowania jest liczony w taki sposób że połowa zbiorów bardziej licznych jest zestawiana z połową zbiorów mniej licznych. W przypadku nieparzystej liczby elementów zbiór większościowy posiada jeden element więcej.

Atrybut	IR
X	2,53
Y	9,2
month	13,34
day	1,75
FFMC	20,86
DMC	9,67
DC	14,3
ISI	8,37
temp	3,73
RH	18,96
wind	31,79
rain	228,5
area	7,83

Wnioski

Niektóre z atrybutów naszego zbioru są w miarę zbalansowane. Wynika to z tego iż pożar nie zależy od dnia tygodnia, więc atrybut "day" posiada niski współczynnik niezbalansowania. Natomiast pożar występuje znacznie rzadziej w przypadku opadów dlatego współczynnik niezbalansowania dla atrybutu "rain" jest wysoki.