

# Untitled

Mateusz Wizner

2025-05-25

## a) Dane: Wine (biblioteka HDclassif).

Opis zmiennych w zbiorze danych **Wine**

Kolumna	Nazwa zmiennej	Opis
V1	Alcohol	Zawartość alkoholu (%)
V2	Malic acid	Zawartość kwasu jabłkowego (g/l)
V3	Ash	Zawartość popiołu (g/l)
V4	Alcalinity of ash	Zasadowość popiołu (g/l)
V5	Magnesium	Zawartość magnezu (mg/l)
V6	Total phenols	Zawartość fenoli ogółem (g/l)
V7	Flavanoids	Zawartość flawonoidów (g/l)
V8	Nonflavanoid phenols	Zawartość fenoli nienależących do flawonoidów (g/l)
V9	Proanthocyanins	Zawartość proantocyjaninów (g/l)
V10	Color intensity	Intensywność koloru (od 0 do 13)
V11	Hue	Odcień barwy (od 0 do 1)
V12	OD280/OD315 of diluted wines	Absorbancja przy długości fali 280 nm do 315 nm (rozcieńczone wino)
V13	Proline	Zawartość proliny (mg/l)

Pierwsze 10 rekordów zbioru danych.

class	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92	1065
1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050
1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185
1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45	1480
1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93	735
1	14.20	1.76	2.45	15.2	112	3.27	3.39	0.34	1.97	6.75	1.05	2.85	1450
1	14.39	1.87	2.45	14.6	96	2.50	2.52	0.30	1.98	5.25	1.02	3.58	1290
1	14.06	2.15	2.61	17.6	121	2.60	2.51	0.31	1.25	5.05	1.06	3.58	1295
1	14.83	1.64	2.17	14.0	97	2.80	2.98	0.29	1.98	5.20	1.08	2.85	1045
1	13.86	1.35	2.27	16.0	98	2.98	3.15	0.22	1.85	7.22	1.01	3.55	1045

Zbiór danych ma **178** przypadków i **14** zmiennych

Table 3: Liczba unikalnych wartości w każdej zmiennej

	Liczba unikalnych
class	3
V1	126
V2	133
V3	79
V4	63
V5	53
V6	97
V7	132
V8	39
V9	101
V10	132
V11	78
V12	122
V13	121

Zmienna **class** zawiera informacje o przynależności obiektów do grup (klas), czyli pełni funkcję etykiety klas. Potwierdza to zarówno jej nazwa, jak i liczba unikalnych wartości, która wskazuje, ile klas znajduje się w zbiorze danych.

Ilość rekordów przypisanych do odpowiedniej klasy.

Klasa	Ilość
1	59
2	71
3	48

Ilość danych oznaczonych jako **Na** w danych kolumnach.

```
## class    V1    V2    V3    V4    V5    V6    V7    V8    V9    V10   V11   V12
##      0      0      0      0      0      0      0      0      0      0      0      0
##   V13
##      0
```

**Analizując dane**, można zauważyć, że **zbiór nie zawiera żadnych braków** — ani oznaczonych jako **NA**, ani zapisanych w inny sposób. **Wszystkie obserwacje** wydają się być **poprawnie wprowadzone**.

Jeśli chodzi o **wartości nietypowe**, to w kolumnie **V10** znajduje się **jedna obserwacja** o wartości **9.899999**, która ma **aż sześć miejsc po przecinku**.

Dla porównania, **pozostałe wartości** w tej kolumnie mają **najwyżej dwie cyfry po przecinku**, co może sugerować **błąd w zapisie** tej konkretnej danej.

W kolumnie **V5** pojawia się również wartość **162**, która **znaczaco odstaje od reszty obserwacji** i może być wynikiem **błędu pomiaru** lub **wprowadzenia danych**.

---

**Typ** danych jaki przyjmują wartości z danej kolumny

	Typ danych
class	integer
V1	numeric
V2	numeric
V3	numeric
V4	numeric
V5	integer
V6	numeric
V7	numeric
V8	numeric
V9	numeric
V10	numeric
V11	numeric
V12	numeric
V13	integer

Widać, że **wszystkie zmienne mają prawidłowo przypisane typy danych**, z wyjątkiem **naszej etykiety klas** — kolumny **class**.

Obecnie ma ona typ **integer**, co jest zrozumiałe, ponieważ wartości to liczby całkowite **{1, 2, 3}**.

Jednak **dla poprawnej analizy** i właściwego traktowania tej zmiennej jako **zbioru kategorii**, powinna zostać **przekonwertowana na typ factor**.

---