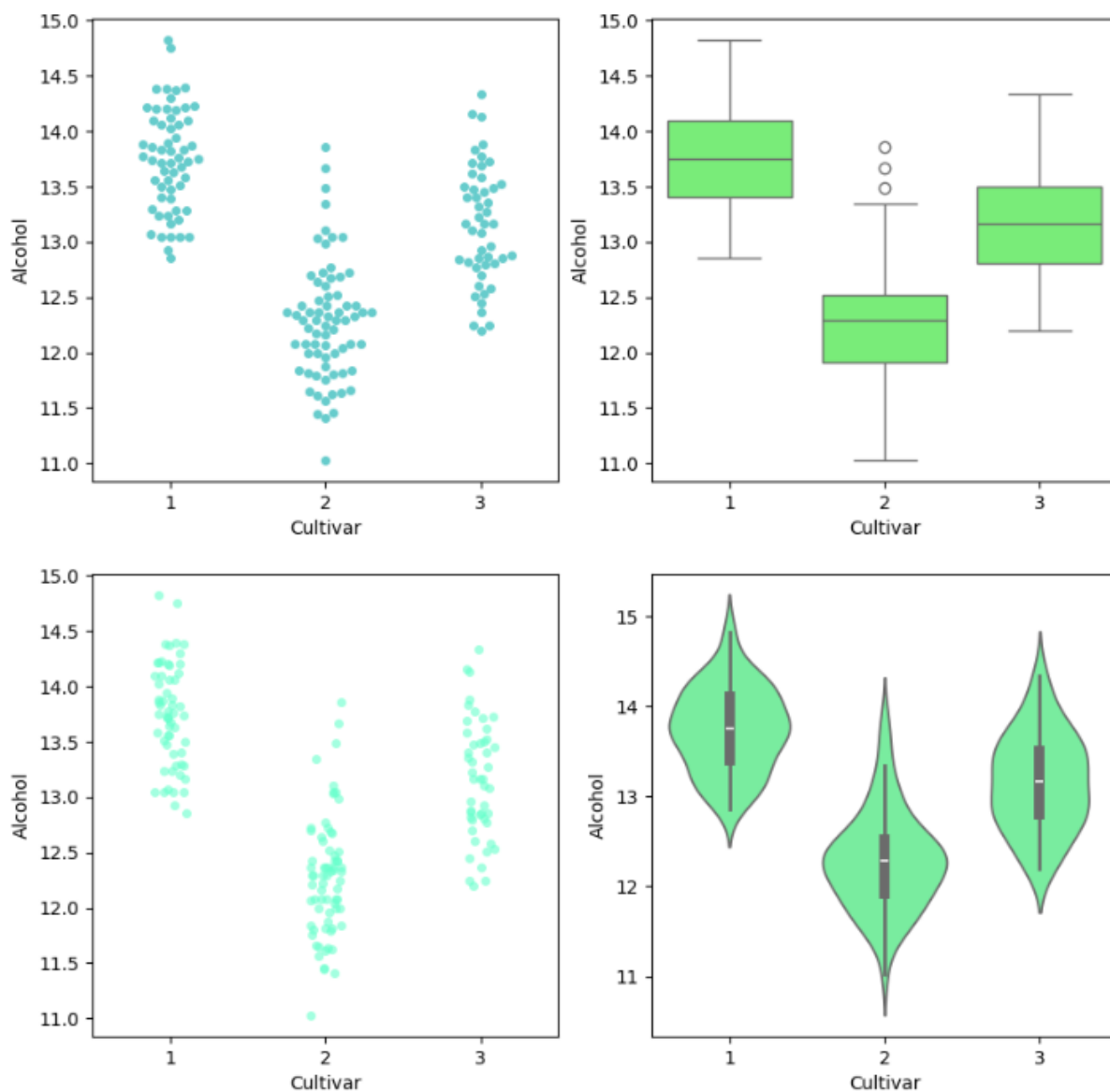


# **Analiza pomiarów wybranych parametrów win**

## 1. Analiza zawartości alkoholu w zależności od odmiany winogron

Wykresy przedstawione poniżej ukazują rozkład zawartości alkoholu w winach pochodzących z trzech różnych odmian winogron, z wykorzystaniem różnych typów wizualizacji: wykresu punktowego (scatter), wykresu pudełkowego (boxplot), wykresu rojowego (swarm) oraz violin plotu.



Na podstawie analizy graficznej można sformułować następujące wnioski:

- Średnia zawartość alkoholu jest najwyższa w winach pochodzących z odmiany 1, nieco niższa w odmianie 3, a najniższa w odmianie 2. Różnice te są wyraźnie widoczne we wszystkich typach wykresów.

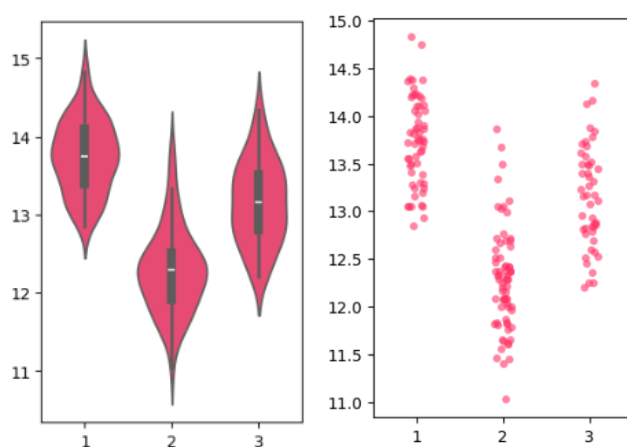
- Zmienność zawartości alkoholu (rozproszenie danych) również różni się pomiędzy odmianami. Odmiana 2 charakteryzuje się największym zróżnicowaniem (rozciągnięcie wykresów pionowo), podczas gdy odmiana 1 wykazuje bardziej skupione wartości.
- Rozkład wartości w przypadku odmiany 2 jest wyraźnie skośny w prawo – co szczególnie dobrze widać na wykresie violin plot. Oznacza to, że chociaż większość próbek ma niższą zawartość alkoholu, pojawiają się też próbki o wyższej zawartości.
- Wartości odstające (outliers) są widoczne głównie w odmianie 2, co potwierdza wykres pudełkowy – kilka punktów znajduje się znacznie powyżej górnej granicy pudełka.
- Potencjał diagnostyczny: zawartość alkoholu wydaje się być jednym z parametrów, które mogą pomóc w rozróżnieniu win pochodzących z różnych odmian winogron. W szczególności odmiana 2 odróżnia się niską średnią i większym zróżnicowaniem zawartości alkoholu.

## 2. Porównanie wybranych cech chemicznych win trzech szczepów

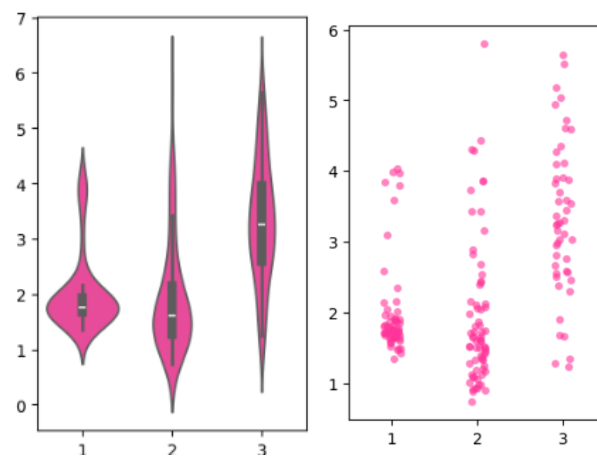
W tej części analizy porównano sześć cech chemicznych win (każda na osobnym wykresie), biorąc pod uwagę ich rozkład w zależności od szczepu winorośli (Cultivar). Do wizualizacji wykorzystano wykresy typu violinplot oraz stripplot, co pozwoliło jednocześnie zobaczyć gęstość rozkładu oraz indywidualne obserwacje.

Cechy poddane analizie:

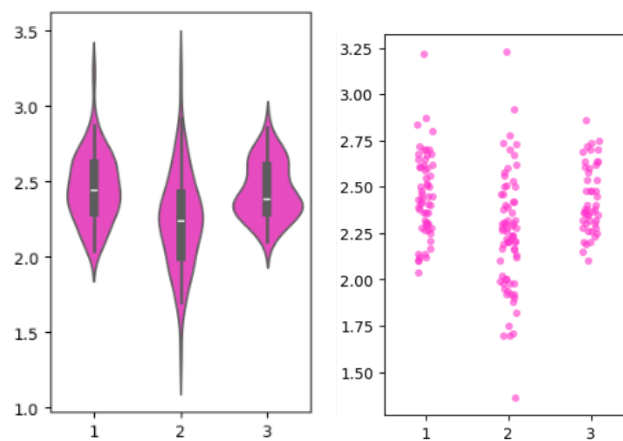
### 1. Malic Acid



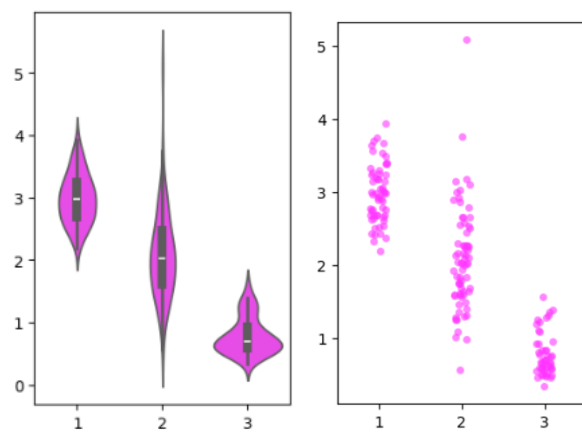
## 2. Ash



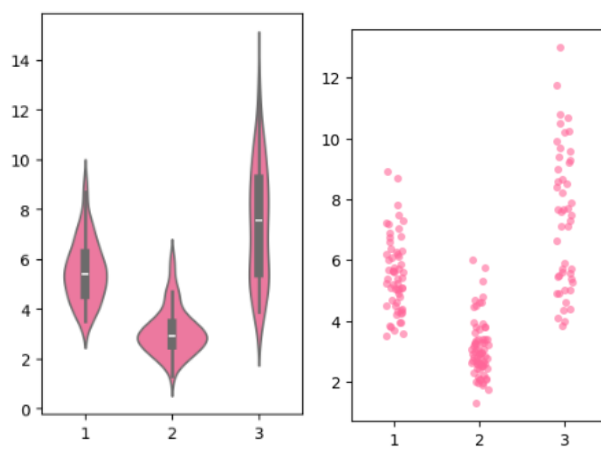
## 3. Alkalinity of Ash



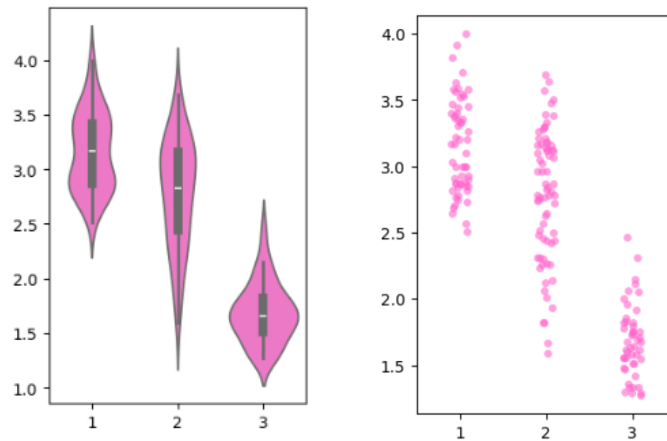
## 4. Magnesium



## 5. Flavanoids



## 6. Color Intensity



Wnioski z analizy poszczególnych cech:

- **Malic Acid:**  
Szczep 2 ma wyraźnie najniższe wartości tej cechy, natomiast szczep 3 wykazuje największe zróżnicowanie. Szczep 1 ma wartości umiarkowane i względnie stabilne.
- **Ash:**  
Szczep 1 i 3 mają zbliżone wartości mediany, ale szczep 2 wykazuje nieco niższe wartości i szerszy rozkład. Różnice nie są jednak bardzo wyraźne.
- **Alcalinity of Ash:**  
Tu najbardziej zróżnicowany jest szczep 3, który ma największy rozrzut. Szczep 2 ma niższą medianę niż pozostałe, co może wskazywać na różnice technologiczne w produkcji wina.
- **Magnesium:**  
Wina ze szczepu 1 są najbardziej bogate w magnez (wyższe mediany), podczas gdy szczep 3 ma wyraźnie niższe wartości – również widoczny jest mniejszy rozrzut.
- **Flavanoids:**  
Szczep 1 wyróżnia się zdecydowanie wyższą zawartością flawonoidów. Szczep 3 ma ich zdecydowanie najmniej. Jest to jedna z cech najlepiej różnicujących szczepy.

- Color Intensity:  
Szczep 3 ma największe zróżnicowanie intensywności barwy, a także najwyższe wartości. Szczep 2 wykazuje najniższą średnią intensywność barwy, natomiast szczep 1 prezentuje wartości pośrednie.



### 3. Statystyczna charakterystyka wybranych cech win

W niniejszym rozdziale przeprowadzono analizę sześciu wybranych cech fizykochemicznych win z wykorzystaniem rozszerzonych statystyk opisowych. Uwzględniono miary tendencji centralnej (średnia, mediana), rozproszenia (odchylenie standardowe, zakres, kwartyle), a także kształt rozkładu (skośność i kurtoza).

[18]:

	Cultivar	Alcohol	Malic acid	Ash	Flavanoids	Color intensity	OD280/OD315
count	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000
mean	1.938202	13.000618	2.336348	2.366517	2.029270	5.058090	2.611685
std	0.775035	0.811827	1.117146	0.274344	0.998859	2.318286	0.709990
min	1.000000	11.030000	0.740000	1.360000	0.340000	1.280000	1.270000
25%	1.000000	12.362500	1.602500	2.210000	1.205000	3.220000	1.937500
50%	2.000000	13.050000	1.865000	2.360000	2.135000	4.690000	2.780000
75%	3.000000	13.677500	3.082500	2.557500	2.875000	6.200000	3.170000
max	3.000000	14.830000	5.800000	3.230000	5.080000	13.000000	4.000000
skew	0.106524	-0.051047	1.030869	-0.175207	0.025129	0.861248	-0.304690
kurt	-1.319440	-0.862260	0.257348	1.078576	-0.889365	0.337370	-1.089675

#### 1. Alcohol

Zawartość alkoholu w analizowanych próbkach wina waha się od 11.03% do 14.83%, ze średnią wartością 13.00%. Odchylenie standardowe wynosi 0.81, co wskazuje na stosunkowo niewielką zmienność tej cechy. Rozkład jest zbliżony do symetrycznego (skośność = -0.05), a ujemna wartość kurtozy (-0.86) sugeruje nieco bardziej spłaszczony rozkład w porównaniu do rozkładu normalnego. Mediana i średnia są bardzo zbliżone (odpowiednio 13.05 i 13.00), co dodatkowo potwierdza brak silnej asymetrii.

#### 2. Malic acid

Cechę tę charakteryzuje największa skośność spośród analizowanych zmiennych (1.03), co wskazuje na silnie prawoskośny rozkład. Wartości wahają się od 0.74 do 5.80, a średnia (2.34) jest wyraźnie wyższa od mediany (1.87), co również świadczy o obecności wyższych, mniej typowych wartości. Odchylenie standardowe wynosi 1.12, co oznacza stosunkowo dużą zmienność. Kurtoza dodatnia (0.26) sugeruje lekką koncentrację wyników wokół wartości średniej.

#### 3. Ash

Wartości popiołu mieszczą się w przedziale od 1.36 do 3.23, ze średnią 2.37 i medianą 2.36. Odchylenie standardowe (0.27) świadczy o niskim zróżnicowaniu tej cechy. Skośność jest lekko ujemna (-0.18), co oznacza niewielkie odchylenie rozkładu w lewo. Kurtoza (1.08) wskazuje na bardziej szczytowy rozkład w porównaniu do rozkładu normalnego. Cecha ta wydaje się dość jednorodna w całej próbce.

#### 4. Flavanoids

Średnia zawartość flawonoidów wynosi 2.03, a rozrzut danych jest umiarkowany (odchylenie standardowe 1.00). Zakres wartości to 0.34–5.08. Rozkład tej cechy jest praktycznie symetryczny (skośność = 0.025), a kurtoza ujemna (-0.89) sugeruje bardziej płaski rozkład. Mediana (2.14) jest zbliżona do średniej, co wzmacnia wniosek o braku znacznych odstępstw od symetrii.

#### 5. Color intensity

Intensywność barwy win wykazuje znaczne zróżnicowanie: wartości mieszczą się w zakresie od 1.28 do 13.00, przy średniej 5.06 i odchyleniu standardowym 2.32. Rozkład jest prawoskośny (skośność = 0.86), a dodatnia kurtoza (0.34) oznacza, że dane są lekko skupione wokół średniej, ale z obecnością wartości skrajnie wysokich. Mediana (4.69) jest niższa od średniej, co potwierdza obecność takich wartości.

#### 6. OD280/OD315

Wskaźnik ten, powiązany z zawartością fenoli, przyjmuje wartości od 1.27 do 4.00. Średnia wynosi 2.61, a odchylenie standardowe to 0.71, co wskazuje na umiarkowaną zmienność. Rozkład jest lekko lewoskośny (skośność = -0.30), co oznacza, że nieco więcej próbek ma wartości powyżej średniej. Kurtoza ujemna (-1.09) wskazuje na stosunkowo płaski rozkład bez wyraźnie dominujących wartości.

#### 4. Analiza korelacji właściwości chemicznych win

a) Obliczony współczynnik korelacji liniowej pomiędzy parametrami OD280/OD315 (na osi X) a Flavanoids (na osi Y) wynosi 0.7872. Oznacza to silną dodatnią korelację, wskazującą, że wzrost zawartości flawonoidów wiąże się ze wzrostem wskaźnika OD280/OD315. Tak wysoka wartość współczynnika sugeruje, że między tymi cechami istnieje wyraźny związek chemiczny lub technologiczny, istotny dla charakterystyki wina.

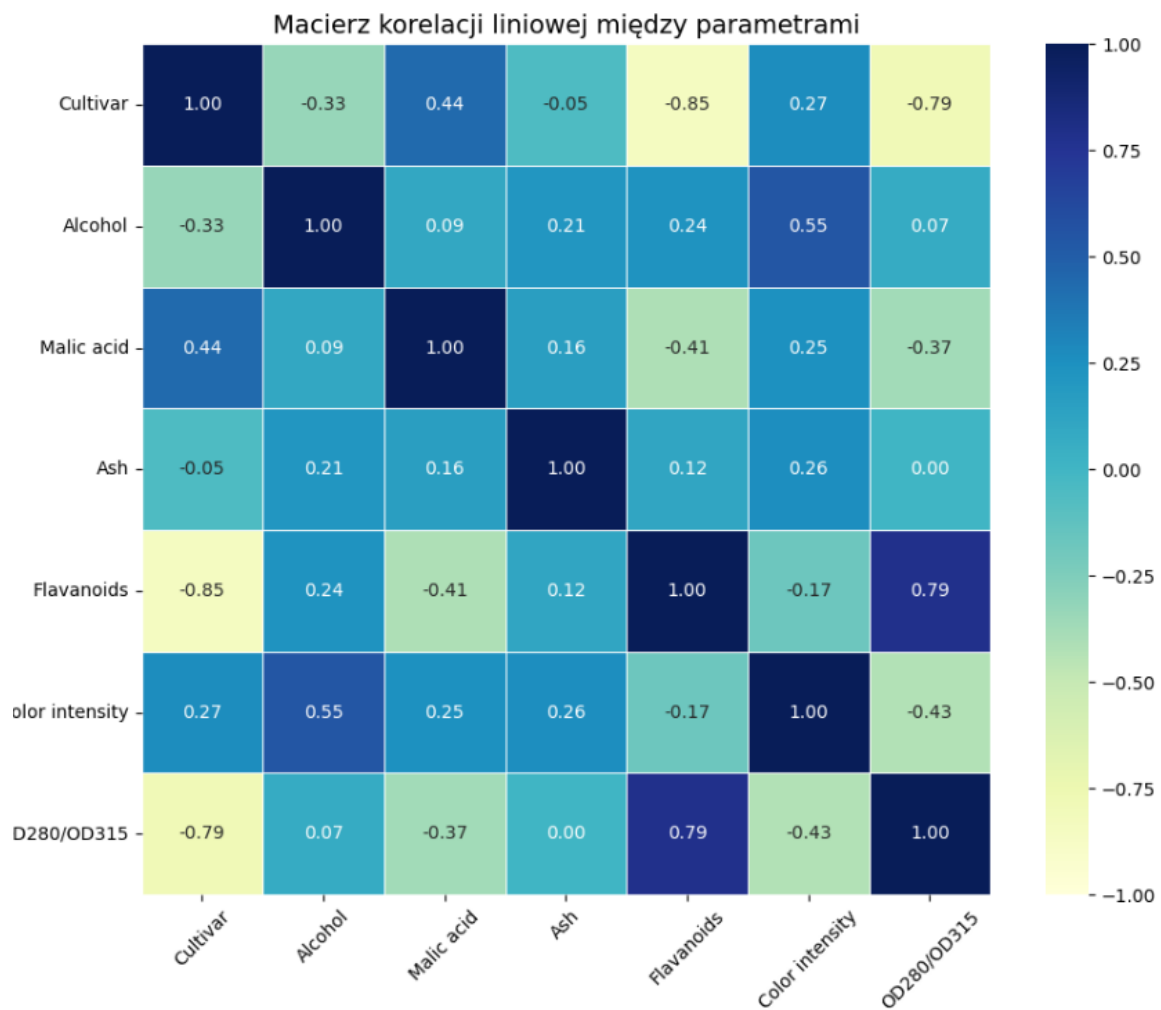
b) Tablica współczynników korelacji przedstawia wszystkie możliwe pary analizowanych zmiennych.

	Cultivar	Alcohol	Malic acid	Ash	Flavanoids	Color intensity	OD280/OD315
Cultivar	1.000000	-0.328222	0.437776	-0.049643	-0.847498	0.265668	-0.788230
Alcohol	-0.328222	1.000000	0.094397	0.211545	0.236815	0.546364	0.072343
Malic acid	0.437776	0.094397	1.000000	0.164045	-0.411007	0.248985	-0.368710
Ash	-0.049643	0.211545	0.164045	1.000000	0.115077	0.258887	0.003911
Flavanoids	-0.847498	0.236815	-0.411007	0.115077	1.000000	-0.172379	0.787194
Color intensity	0.265668	0.546364	0.248985	0.258887	-0.172379	1.000000	-0.428815
OD280/OD315	-0.788230	0.072343	-0.368710	0.003911	0.787194	-0.428815	1.000000

Z analizy  $r$  table można wysnuć kilka istotnych obserwacji:

- Odmiana winogron (Cultivar) jest silnie i ujemnie skorelowana z zawartością flawonoidów (-0.85) oraz wskaźnikiem OD280/OD315 (-0.79), co wskazuje na duże różnice między odmianami w kontekście tych parametrów.
- Zawartość alkoholu wykazuje umiarkowaną dodatnią korelację z intensywnością koloru (Color intensity, 0.55), sugerując, że wyższa zawartość alkoholu może iść w parze z bardziej intensywną barwą wina.
- Flawonoidy są umiarkowanie ujemnie skorelowane z kwasem jabłkowym (-0.41), co może świadczyć o wpływie kwasowości na skład związków fenolowych.
- Intensywność koloru oraz OD280/OD315 wykazują umiarkowaną, ujemną korelację (-0.43), co oznacza, że bardziej intensywnie zabarwione wina niekoniecznie mają wyższy stosunek absorpcji przy długościach fal 280/315 nm.

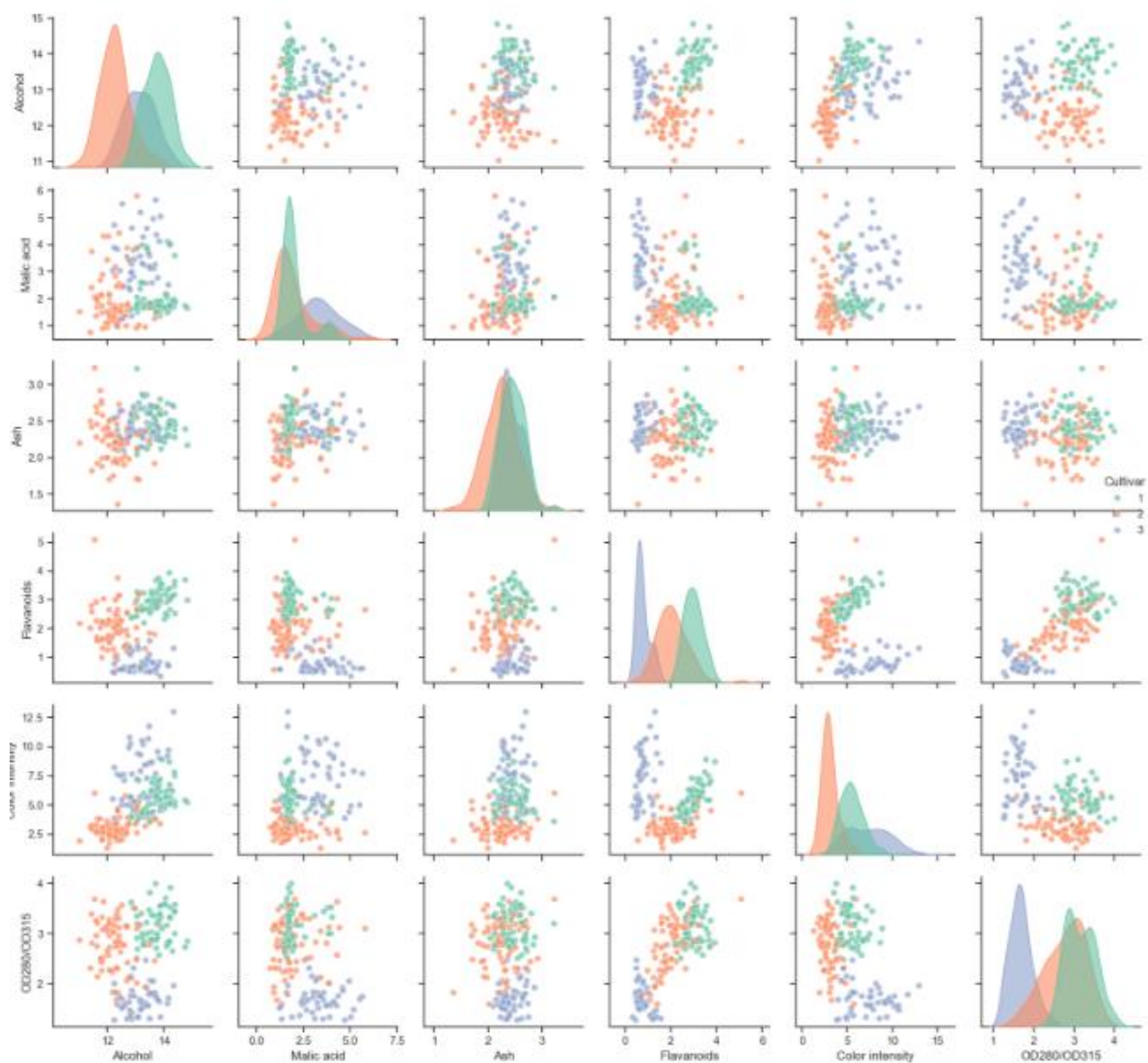
c) W celu łatwiejszej interpretacji korelacji między parametrami przygotowano graficzną reprezentację macierzy korelacji przy użyciu funkcji `seaborn.heatmap()`.



Odcienie koloru w macierzy wskazują siłę i kierunek powiązań:

- Ciemnoniebieskie pola odpowiadają wysokim dodatnim korelacjom (np. Flavanoids i OD280/OD315).
- Jasnożółte wskazują na silne korelacje ujemne (np. Cultivar i Flavanoids).

d) Za pomocą funkcji `seaborn.pairplot()` przygotowano zestaw wykresów korelacyjnych dla każdej pary zmiennych, z dodatkowym rozróżnieniem kolorystycznym odmian winogron (Cultivar).



Dzięki temu możliwe było zaobserwowanie, że:

- Poszczególne odmiany tworzą wyraźne klastry, szczególnie widoczne w zależnościach takich jak Flavanoids vs. OD280/OD315.
- W niektórych przypadkach, jak dla pary Alcohol i Color intensity, widoczny jest trend wzrostowy, niezależnie od odmiany.
- Rozkłady gęstości (na przekątnej macierzy wykresów) pokazują odmienne charakterystyki poszczególnych odmian, co sugeruje, że każda z odmian winogron ma unikalny profil chemiczny.

## 5. Regresja liniowa

a) W pierwszym etapie zadania, korzystając z metody najmniejszych kwadratów, ręcznie obliczono współczynniki regresji liniowej wyrażającej zależność parametru Flavanoids (oś Y) od wskaźnika OD280/OD315 (oś X). Obliczenia prowadzą do uzyskania równania prostej regresji w postaci:

$$Y = a_1 \cdot X + b_1,$$

gdzie:

- $a_1 = 1.1075$  — współczynnik kierunkowy,
- $b_1 = -0.8631$  — wyraz wolny.

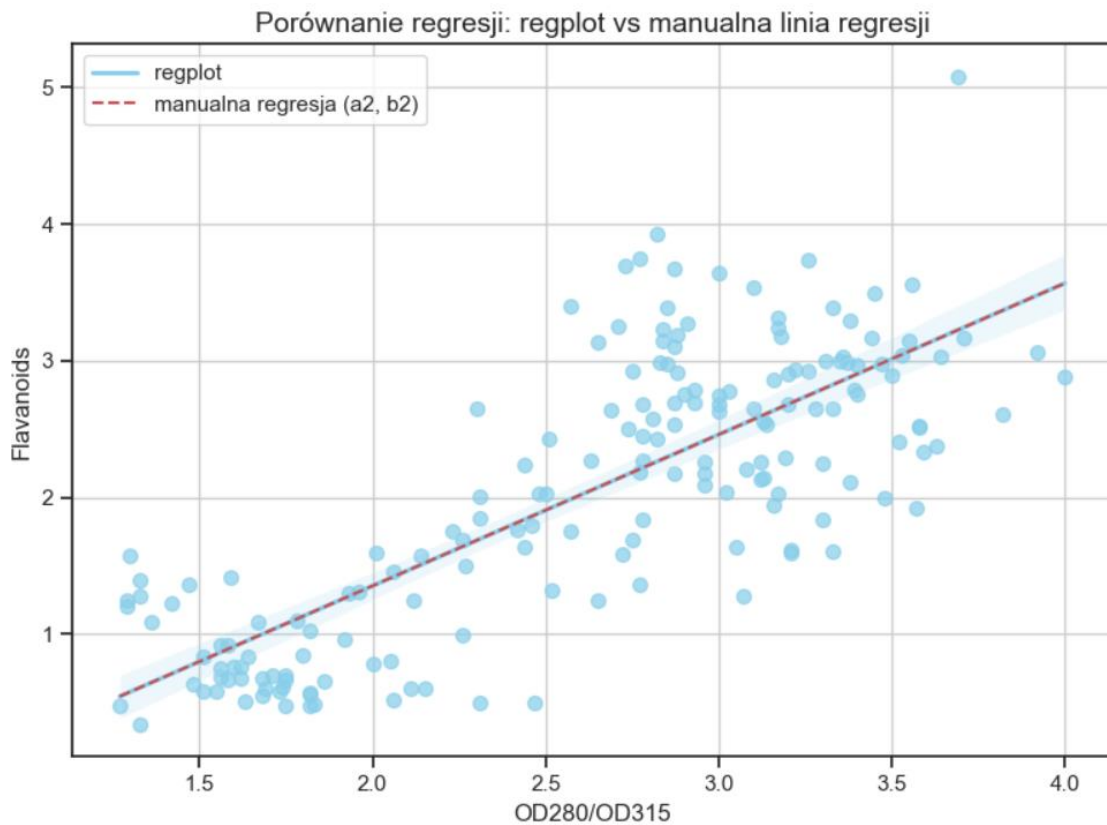
Współczynniki te uzyskano zgodnie z klasyczną formułą regresji liniowej, przy użyciu średnich wartości zmiennych oraz iloczynu odchyleń od średnich.

b) W kolejnym kroku wykorzystano gotową funkcję `scipy.stats.linregress()`, która automatycznie oblicza parametry regresji liniowej. Wyniki uzyskane za pomocą tej metody są identyczne jak w przypadku obliczeń manualnych:

$$a_2 = 1.1075 \text{ oraz } b_2 = -0.8631.$$

Zgodność wyników potwierdza poprawność implementacji ręcznych obliczeń i stanowi dobrą walidację procesu. Dzięki funkcji `linregress()` dodatkowo możliwe byłoby uzyskanie takich parametrów jak wartość  $R^2$ , błąd standardowy czy p-wartość, co mogłoby posłużyć do dalszej, bardziej zaawansowanej analizy statystycznej.

c) Dla zweryfikowania wyników graficznie, stworzono wykres z wykorzystaniem funkcji `seaborn.regplot()`, który automatycznie dopasowuje linię regresji do danych. Na tym samym wykresie naniesiono również prostą regresji wyznaczoną manualnie. Obie linie praktycznie się pokrywają, co jednoznacznie potwierdza, że zarówno obliczenia ręczne, jak i funkcja `linregress()` prowadzą do tego samego rozwiązania.



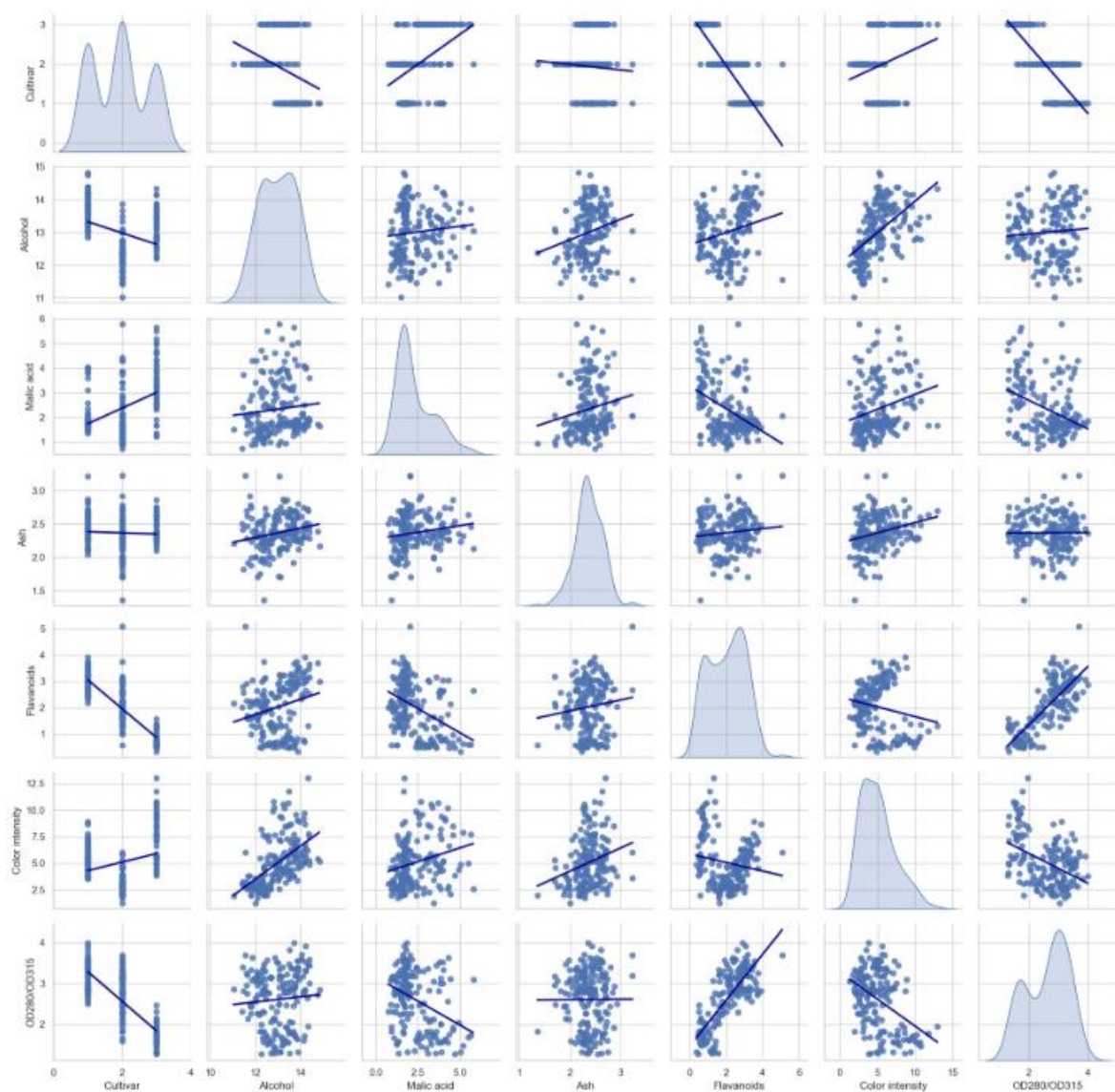
- Niebieskie punkty odpowiadają obserwacjom rzeczywistym.
- Czarna linia reprezentuje regresję uzyskaną przez `regplot()`.
- Przerywana linia odpowiada regresji wyznaczonej ręcznie ( $a_2$ ,  $b_2$ ).

d) W ostatniej części zadania przygotowano rozszerzone wykresy korelacyjne z naniesioną linią regresji dla wszystkich par parametrów.

Zastosowano funkcję `seaborn.pairplot()` w wersji:

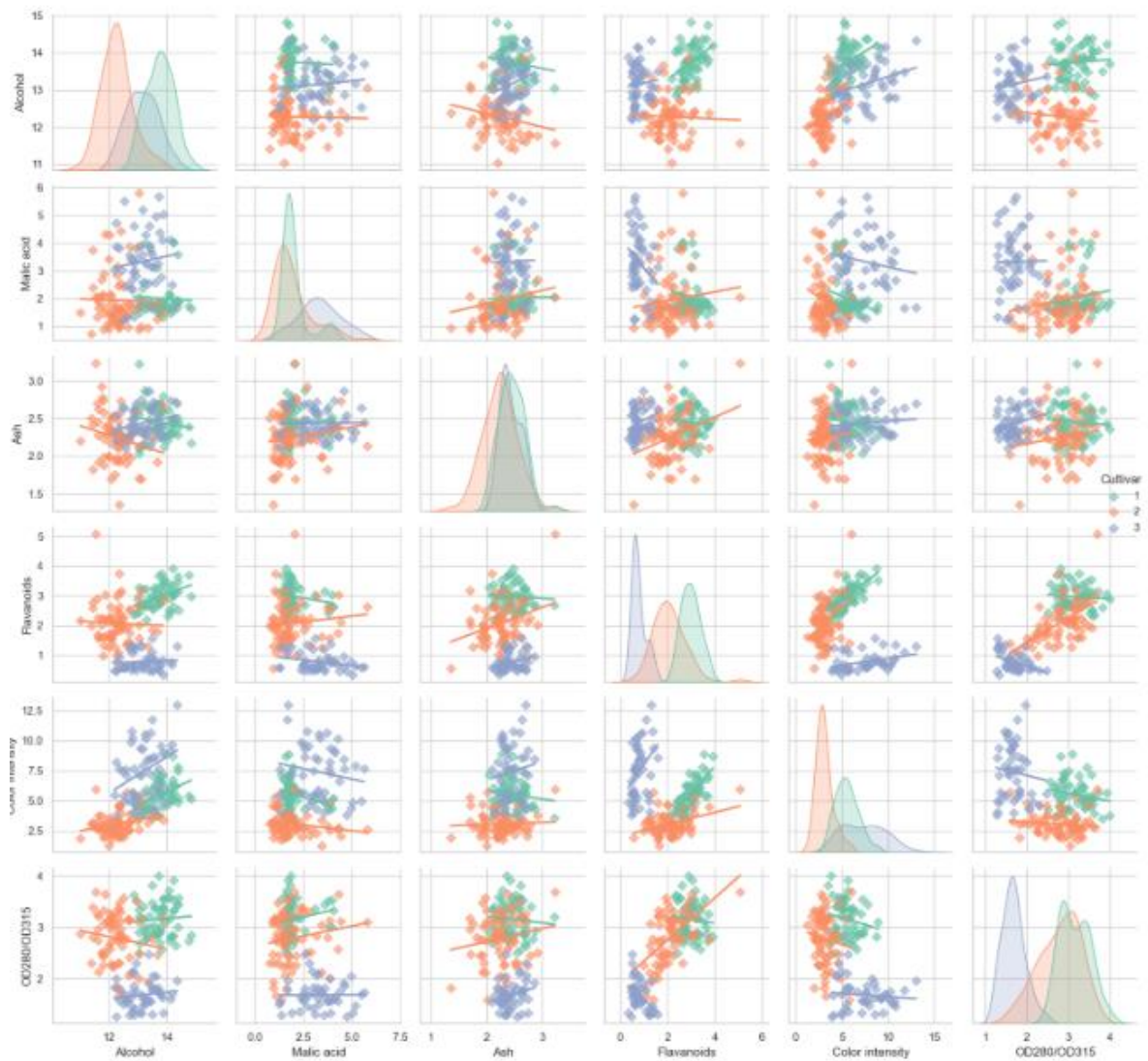
- dla całego zbioru danych — z jednolitym stylem punktów i niebieską linią regresji,
- oraz z podziałem na odmiany winogron (zmienna `Cultivar`) — z różnymi kolorami dla lepszego rozróżnienia.







Wykresy korelacyjne z prostą regresji (z podziałem na odmiany)



Wykresy te pozwalają zaobserwować nie tylko liniową zależność między zmiennymi, ale również to, jak rozkładają się te zależności w obrębie poszczególnych odmian.

- Niektóre pary cech, takie jak OD280/OD315 i Flavanoids, wykazują silną liniową zależność, widoczną jako wyraźnie nachylona linia.
- Inne relacje są znacznie słabsze lub wykazują duże rozproszenie punktów.