1. **Pod jakimi względami wina wykonane z poszczególnych odmian winogron są podobne a pod jakimi się różnią.**

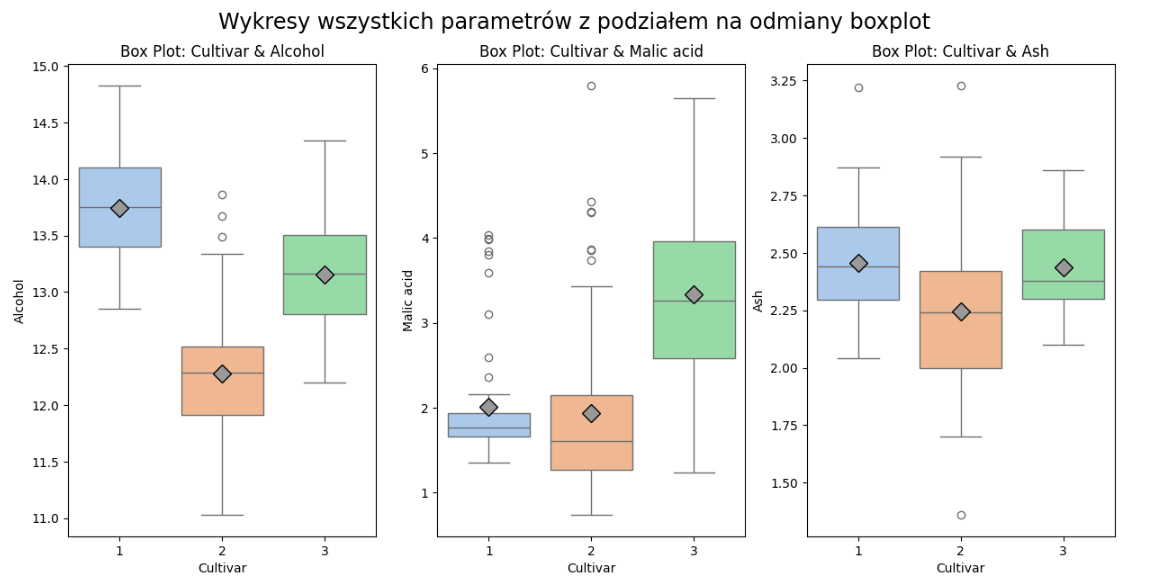
Analizując wykresy **boxplot** i **stripplot** możemy zauważyć podobieństwa oraz różnice.

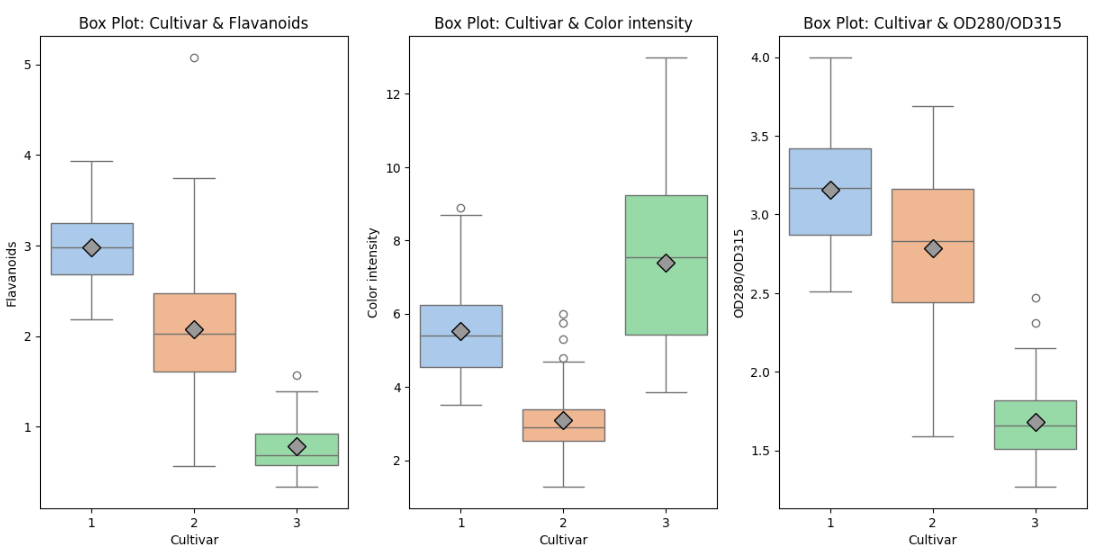
**Podobieństwa**:

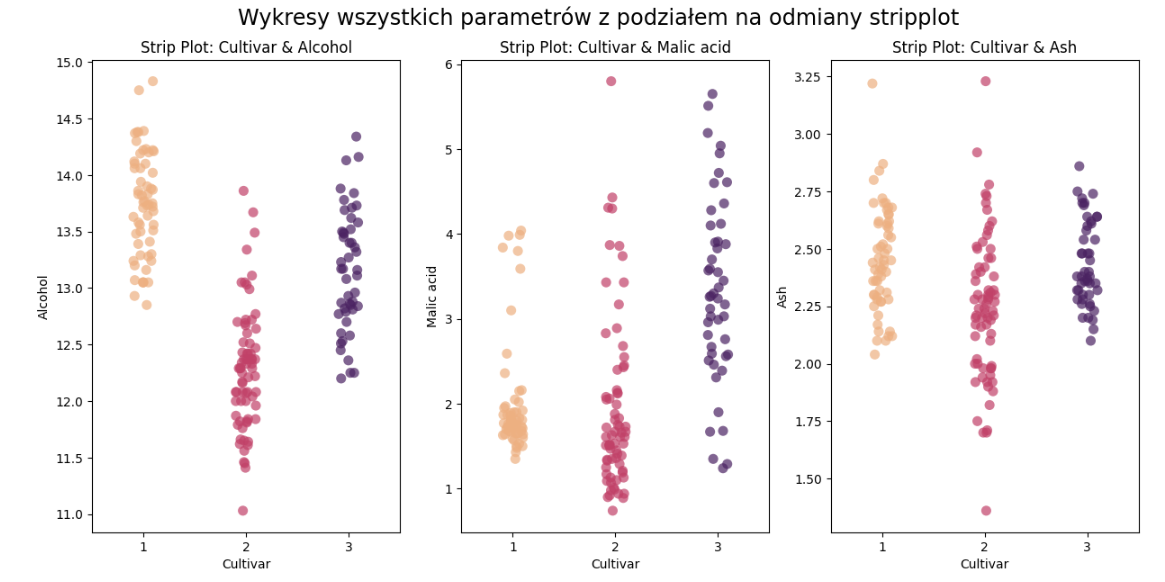
* **Kwas** **Jabłkowy**: Odmiana 1 i 2 mają podobne **średnie** **wartości**.
* **Popiół**: Wszystkie odmiany 1,2,3 mają zbliżone zarówno **wartości** **średnie** i zakresy **Q1** – **Q3**.
* **Absorbancja**: Dla odmian 1 oraz 2 **wartości** **średnie** są oddalone bardziej niż w poprzednich przykładach ale **rozrzut** **danych** jest podobny.

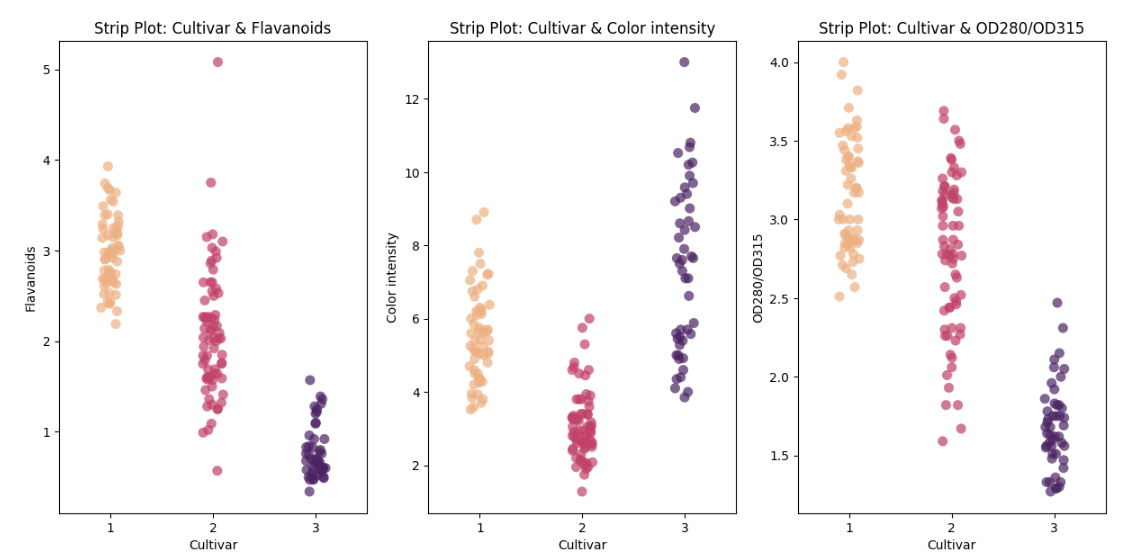
**Różnice**:

* **Alkohol**: Odmiana 1 ma największą średnią zawartość alkoholu, w przeciwieństwie do odmiany 2 która ma tą zawartość najmniejszą, a odmiana 3 klasyfikuje się pomiędzy 1 a 2
* **Kwas Jabłkowy:** Odmiana 3 ma wyższe średnie wartości i szerszy zakres niż odmiany 1 i 2
* **Flawonoidy**: każda z odmian ma inna wartość średnią, gdzie 1 ma ją największą. Największą wariację posiada odmiana 2 a najmniejszą 3.
* **Intensywność Koloru:** Wartości średnie każdej z odmian stosunkowo nie bardzo się różnią. Odmiana 3 ma najwyższą wartość średnią i największą zmienność. Odmiana 2 jest najbardziej stabilna.
* **Absorbancja**: Odmiana 3 ma niższe wartości średnie od odmian 1 oraz 2 i najmniejszą wariancję.









1. **Które zestawienia parametrów pozwalają na rozróżnienie win wykonanych z poszczególnych odmian winogron, które parametry ze sobą korelują, a które nie itp.**

**Zestawienia parametrów**

Do znalezienia które zestawienia parametrów pozwalają na rozróżnienie win z poszczególnych odmian skorzystamy z **pairplota** z podziałem na **odmiany** **winogron**.

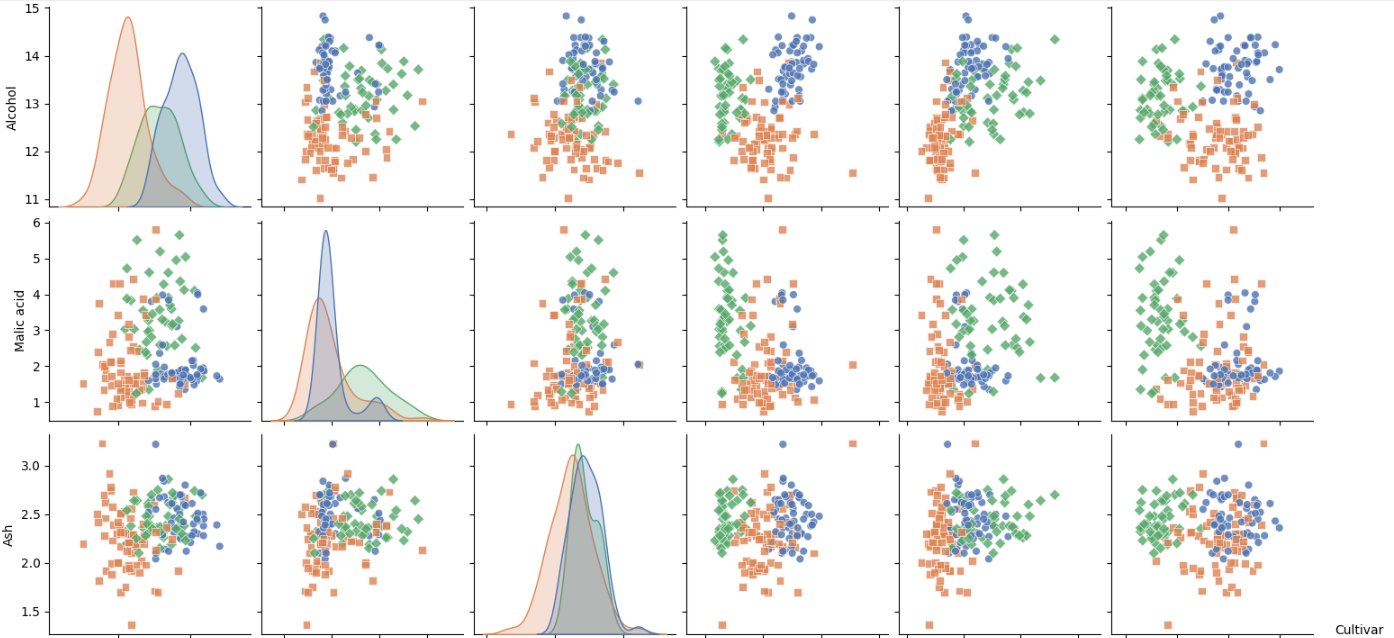
By znaleźć odpowiednie zestawienia **parametrów** szukamy na **pairplocie** tych połączeń, które mają **jak** **najmniej** **nachodzących** na siebie **punktów**.

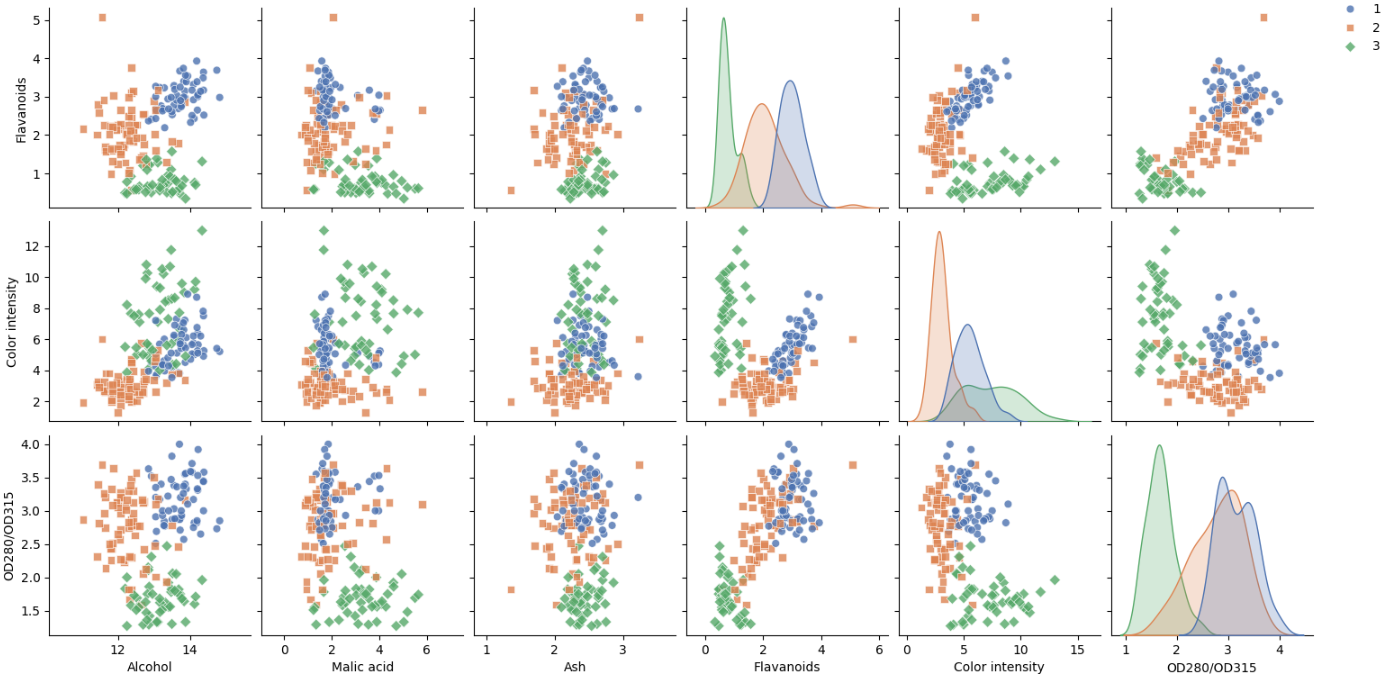
Stosując to znajdujemy **zestawienia** **parametrów** pozwalających rozróżnić odmiany:

* **Alkohol** & **Flawonoidy:** wyraźne grupy
* **Alkohol** & **Absorbancja**: bardzo wyraźne grupy
* **Kwas** **Jabłkowy** & **Flawonoidy**: stosunkowo dobrze wyraźne grupy
* **Flawonoidy** & **Intensywność** **Koloru**: wyraźne grupy
* **Intensywność** **Koloru** & **Absorbancja:** stosunkowo dobrze wyraźne grupy

Są jeszcze takie zestawienia jak **Flawonoidy** & **Absorbancja** czy **Alkohol** & **Intensywność** **Koloru** które też moglibyśmy wykorzystać do rozróżnienia win wykonanych z poszczególnych odmian jednak podane **wyżej** będą **bardziej** **przydatne**.

Dodatkowo możemy zauważyć, że same **Flawonoidy** czy **Intensywność** **Koloru** są dobrymi **parametrami** do rozróżniania win wykonanych z odmian winogron, co widać na **przekątnej** **osi** gdzie wykresy **nie** są bardzo **nałożone** na siebie.





**Korelacja między parametrami**.

Do badania **korelacji** między parametrami korzystamy z **heatmapy** która jest **wizualizacją** **korelacji** uzyskanych z **danych**.

Warto zaznaczyć tutaj, że badamy korelacje **dodatnie** (gdy jedna wartość **rośnie** druga też **rośnie**), **ujemne** (gdy jedna wartość **rośnie** druga **maleje**) oraz **brak** korelacji. Gdzie zastosujemy przedstawiony niżej przedział wartości:

|Kor| < 0.2 -> **brak** korelacji

0.2 <= |Kor| < 0.4 -> **słaba** korelacja

0.4 <= |Kor| < 0.7 -> **umiarkowana** korelacja

0.7 <= |Kor| < 0.9 -> **mocna** korelacja

|Kor| >= 0.9 -> **bardzo** mocna korelacja

Dodatkowo na naszym wykresie **przekątna** **główna** składa się z samych **1** co jest prawidłowe bo jest to korelacja danej cechy z **samą** **sobą**, **nie** **uwzględniamy** jej w analizie bo nic **nie** **wnosi** do naszej analizy.

**Korelacje** **dodatnie**:

* **Odmiana** & **Kwas** **Jabłkowy**: 0.44 – **umiarkowana** korelacja
* **Odmiana** & **Intensywność** **Koloru**: 0.27 – **słaba** korelacja
* **Alkohol** & **Popiół**: 0.21 – **słaba** korelacja
* **Alkohol** & **Flawonoidy**: 0.24 – **słaba** korelacja
* **Alkohol** & **Intensywność** **Koloru**: 0.55 – **umiarkowana** korelacja
* **Kwas** **Jabłkowy** & **Intensywność** **Koloru**: 0.25 – **słaba** korelacja
* **Popiół** & **Intensywność** **Koloru**: 0.26 – **słaba** korelacja
* **Flawonoidy** & **Absorbancja**: 0.79 – **mocna** korelacja

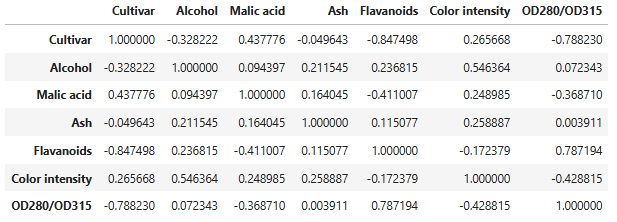
**Korelacje** **ujemne**:

* **Odmiana** & **Alkohol:** -0.33 – **słaba** korelacja
* **Odmiana** & **Flawonoidy:** -0.85 – **mocna** korelacja
* **Odmiana** & **Absorbancja:** -0.79 – **mocna** korelacja
* **Kwas** **Jabłkowy** & **Flawonoidy:** -0.41 – **umiarkowana** korelacja
* **Kwas** **Jabłkowy** & **Absorbancja:** -0.37 – **słaba** korelacja
* **Intensywność** **Koloru** & **Absorbancja:** -0.43 – **umiarkowana** korelacja

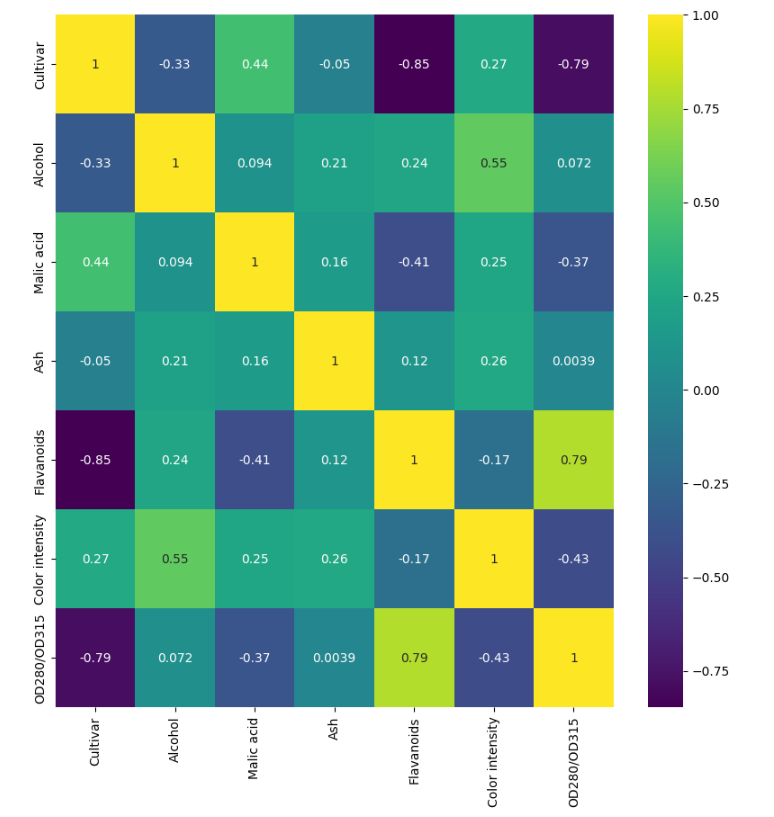
**Brak korelacji:**

* **Odmiana** & **Popiół**: -0.05
* **Alkohol** & **Kwas** **Jabłkowy**: 0.094
* **Alkohol** & **Absorbancja**: 0.072
* **Kwas** **Jabłkowy** & **Popiół**: 0.16
* **Popiół** & **Flawonoidy**: 0.12
* **Popiół** & **Absorbancja**: 0.0039
* **Flawonoidy** & **Intensywność** **Koloru**: -0.17

**Korelacja w postaci tabeli**



**Wizualizacja tabeli jako heatmapa**



1. **Jakie związki między poszczególnym parametrami win możemy zaobserwować, z uwzględnieniem kierunku i siły relacji dla całości zestawu danych oraz dla win wykonanych z poszczególnych odmian winogron, a także różnic w trendach obserwowanych dla całości zestawu danych oraz dla win wykonanych z poszczególnych odmian winogron.**

**Cały zestaw danych.**

Skorzystamy z **pairplot** z dodatkową **prostą** **regresji** która wskaże nam **kierunek**:

* Prosta regresji skierowana w **górę** -> korelacja **dodatnia**
* Prosta regresji skierowana w **dół** -> korelacja **ujemna**
* Prosta regresji przypominająca prostą **poziomą** -> **brak** korelacji

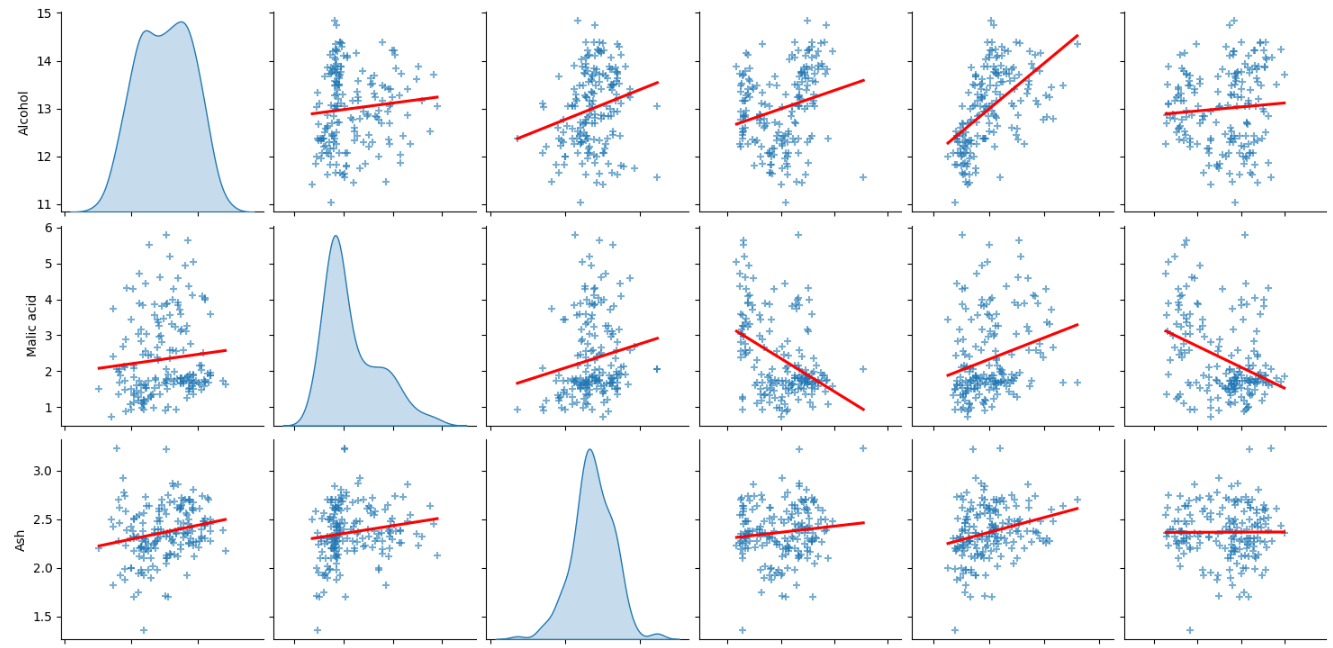
Oraz **siłę** korelacji:

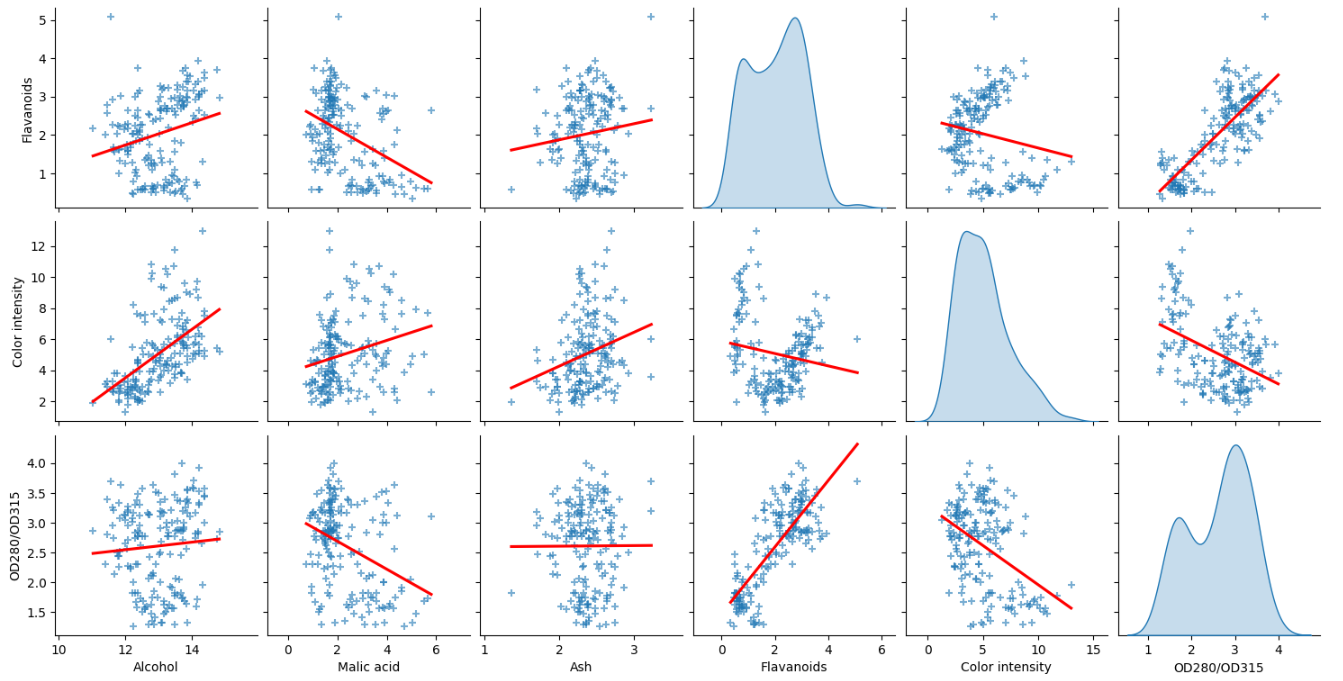
* Bardziej **pionowa** prosta regresji -> **duża** siła korelacji
* Bardziej **pozioma** prosta regresji -> **mała** siła korelacji / **brak** korelacji

Te proste regresji wykorzystamy również do **porównania** **trendów** między **całym** zestawem danych a danymi z **podziałem** na **odmiany**.

**Całość danych:**

* **Alkohol**:
  + **Dodatnia słaba** korelacja z **Popiołem** oraz **Flawonoidami**
  + **Dodatnia umiarkowana** korelacja z **Intensywnością** **Koloru**
  + **Brak** korelacji z **Kwasem** **Jabłkowy** oraz **Absorbancją**
* **Kwas Jabłkowy:**
  + **Dodatnia słaba** korelacja z **Intensywnością** **Koloru**
  + **Ujemna** **słaba** / **umiarkowana** korelacja z **Flawonoidami** oraz **Absorbancją**
  + **Brak** korelacji z **Popiołem**
* **Popiół**:
  + **Dodatnia słaba** korelacja z **Intensywnością** **Koloru**
  + **Brak** korelacji z **Flawonoidami** oraz **Absorbancją**
* **Flawonoidy**:
  + **Dodatnia mocna** korelacja z **Absorbancją**
  + **Brak** korelacji z **Intensywnością Koloru**
* **Intensywność Koloru:**
  + **Ujemna umiarkowana** korelacja z **Absorbancją**





**Zestaw danych z podziałem na odmiany.**

Dla **pairplotu** z dodatkowym podziałem na **odmiany** pojawią się dodatkowe **linie** **regresji** (łącznie 3) oraz punkty odpowiadające **odmianom**.

Gdy **prosta** **regresji** danej **odmiany** będzie **pokrywać** się z prostą regresji dla **całości** **danych** możemy stwierdzić, że **trendy** są **takie** **same**, w przeciwnym wypadku będą się **różnić**.

**Dane z podziałem na odmiany:**

* **Alkohol**:
  + **Kwas** **Jabłkowy** => **Brak** korelacji dla odmian 1,2,3
    - 3 trend **taki** **sam**; 1,2 trend **delikatnie** **inny**
  + **Popiół** => **Brak** / **ujemna** **słaba** korelacja dla odmian 1,2; **dodatnia** **słaba** / **umiarkowana** korelacja dla odmiany 3
    - 1,2 trend **inny**; dla 3 trend **taki** **sam**.
  + **Flawonoidy** => **Brak** korelacji dla odmiany 2; **brak** / **słaba dodatnia** dla odmiany 3; **słaba dodatnia** dla odmiany 1
    - 1,3 trend **taki** **sam** przy czym 1 **większa** **siła** relacji**;** dla 2 trend **inny**
  + **Intensywność Koloru** => **Słaba** / **umiarkowana dodatnia** korelacja dla odmian 1,2,3
    - 1,2,3 trend **taki sam** przy czym **delikatnie** **mniejsza** **siła** relacji
  + **Absorbancja** => **Brak / słaba dodatnia** dla odmian 1,3; **Brak / słaba ujemna** dla odmiany 2
    - 1,3 trend **taki sam;** 2 trend **inny**
* **Kwas Jabłkowy**:
  + **Popiół** => **Brak** korelacji dla odmian 1,3; **Brak** / **słaba dodatnia** dla odmiany 2
    - 1,2,3 trend **taki sam**
  + **Flawonoidy** => **Brak** / **słaba ujemna** korelacja dla odmiany 1; **Brak / słaba dodatnia** korelacja dla odmiany 2; **Umiarkowana ujemna** korelacja dla odmiany 3
    - 1,3 trend **taki sam** przy czym 3 **większa siła** relacji; dla 2 trend **inny**
  + **Intensywność Koloru** => **Słaba** ujemna korelacja dla odmian 1,2,3
    - 1,2,3 trend **inny**
  + **Absorbancja** => **Brak** korelacji dla odmian 1,2,3
    - 1,2,3 trend **inny**
* **Popiół:**
  + **Flawonoidy** => **Brak** korelacji dla odmiany 1; **Słaba / umiarkowana dodatnia** korelacja dla odmian 2,3
    - 1 trend **inny** ; 2,3 trend **taki sam** przy czym mają **większą siłę** relacji
  + **Intensywność Koloru** => **Brak / słaba ujemna** korelacja dla odmiany 1; **Brak / słaba dodatnia** korelacja dla odmian 2,3
    - 1 trend **inny**; 2,3 trend **taki** **sam** przyczymmają **mniejszą siłę** relacji
  + **Absorbancja** => **Brak** korelacji dla odmiany 1; **Brak / słaba dodatnia** korelacja dla odmian 2,3
    - 1 trend **delikatnie** **inny**; 2,3 trend **taki** **sam**
* **Flawonoidy**:
  + **Intensywność Koloru** => **Umiarkowana / mocna dodatnia** korelacja dla odmian 1,2; **Brak / słaba dodatnia** korelacja dla odmiany 3
    - 1,2 3 trend **inny**
  + **Absorbancja** => **Brak / Słaba ujemna** korelacja dla odmian 1,3; **Umiarkowana / mocna dodatnia** korelacja dla odmiany 2
    - 1,3 trend **inny**; 2 trend **taki** **sam**
* **Intensywność Koloru:**
  + **Absorbancja => Brak** korelacji z odmianami 1,2,3
    - 1,2,3 trend **taki sam** przy czym **mniejszą** **siłę** relacji

