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

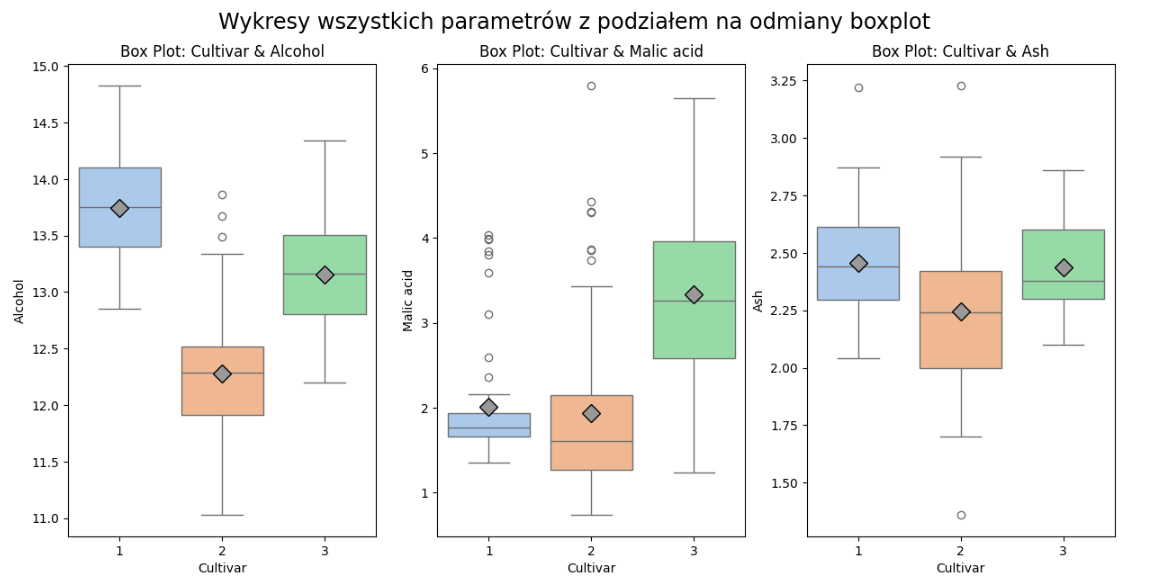
Podobieństwa:

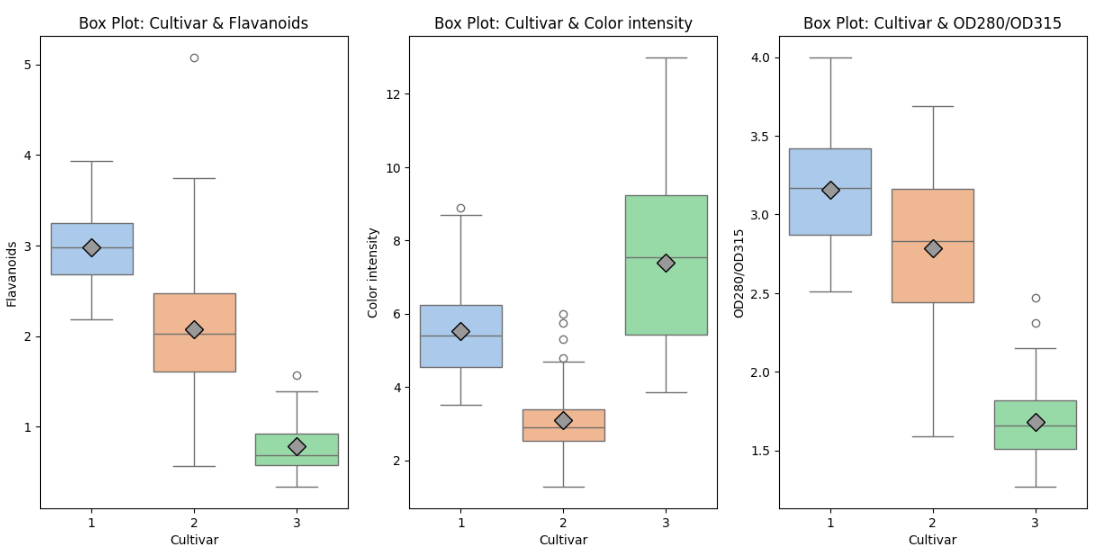
Zauważyć możemy podobieństwa w niektórych miejscach analizując wykres box plot oraz strip plot. Głównie zauważamy, że:

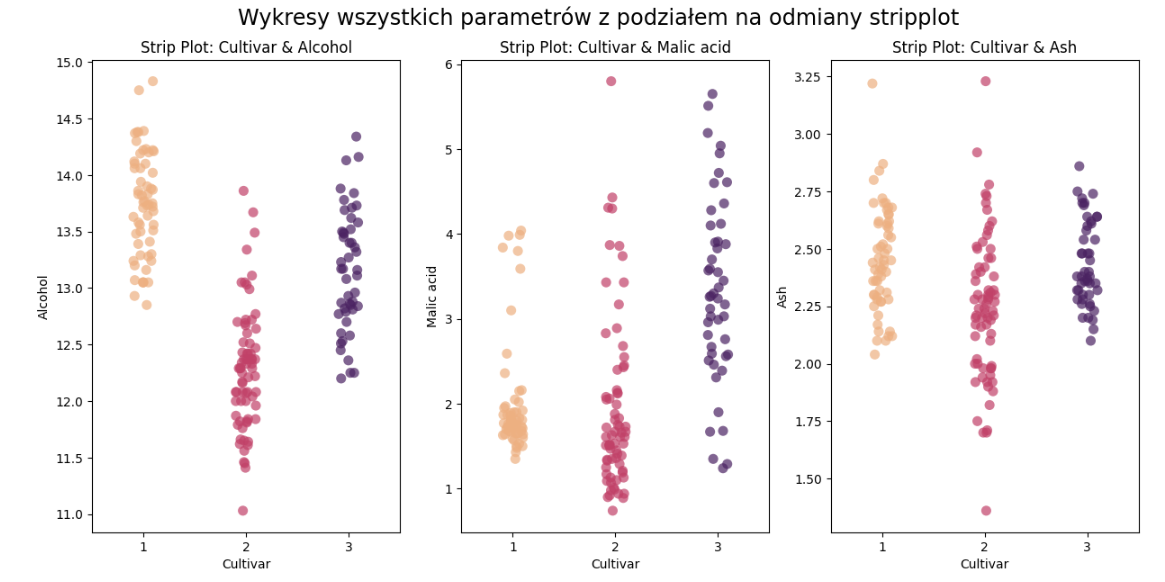
* Odmiana 1 i 2 są stosunkowo podobne pod względem kwasu jabłkowego (Meliac acid), ich wartości średnie wizualnie są bardzo podobne.
* Wszystkie odmiany 1,2 oraz 3 są podobne do siebie w kwestii popiołu (Ash), najlepiej jest to widoczne na boxplocie gdzie zarówno wartości średnie nie odstają od siebie mocno oraz zakresy Q1 oraz Q3 często się pokrywają.
* Widzimy również podobieństwo w absorpcji między odmianami 1 oraz 2. Tutaj wartości średnie są oddalone bardziej niż w poprzednich przykładach ale zakres Q1 oraz Q3 również często się pokrywa tak samo jak wąsy wykresów.

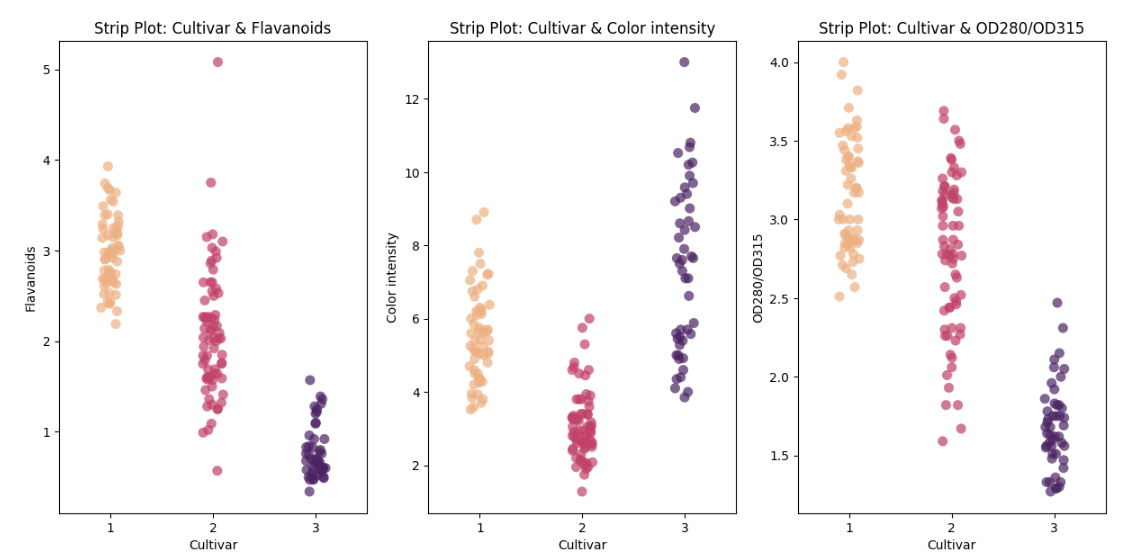
Główne różnice na które można zwrócić uwagę to:

* Poziom alkoholu ze względu na odmianę, zauważamy że odmiana 1 zawiera średnio najwięcej alkoholu w przeciwieństwie do odmiany 2 która ma go średni najmniej, jednak ma też największą wariację co widzimy po wartościach odstających i długich wąsach wykresu. Odmiana 3 klasyfikuje się pomiędzy 1 a 2
* Kwas jabłkowy mimo, że jego wartość średnia jest podobna dla odmiany 1 oraz 2 to można zauważyć, ze wartości dla odmiany 1 są mniej rozrzucone co wskazuje bardzo mały przedział Q1 – Q3. Odmiana 2 ma już ten przedział większy co wskazuje, że ilość kwasu jabłkowego się mocniej wacha. Jednak odmiana 3 najbardziej odstaje od reszty, jego wartość średnia jest dużo większa od odmian 1 i 2 oraz zakres wartości prawdopodobnych jest największy.
* W wykresie flawonoidów zauważamy że każda z odmian ma zupełnie inna wartość średnią. Odmiana 1 ma wartość średnią największą, następnie jest odmiana 2 oraz 3. Dodatkowo największą wariację wartości posiada odmiana 2 a najmniejszą 3 co sugeruje, że odmiana 3 jest najbardziej stabilna pod względem ilości flawonoidów.
* Intensywność koloru natomiast posiada wartości średnie rozrzucone stosunkowo nie mocno, jednak różnice są zauważalne. Dodatkowo odmiana 3 ma najbardziej rozrzucone wartości co dobrze widać choćby na strip plocie, jednak jego średnia intensywność koloru jest największa. Odmiana 2 jest najbardziej stabilna pod względem intensywności koloru, jednak jest też średnio najmniej intensywna.
* Na zestawieniu odmian oraz absorbcji zauważamy, że mimo że wartości dla odmiany 1 i 2 są bardzo podobne zarówno pod względem średniej jak i wariacji to wartości odmiany 3 zupełnie odbiegają od reszty. Jej absorbcji średnio jest dużo niższa od pozostałej dwójki a dodatkowo ma najmniejszą wariancje, dobrze widać rozrzucenie wyników na wykresie Strip Plot gdzie można zauważyć od razu tą różnicę.









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

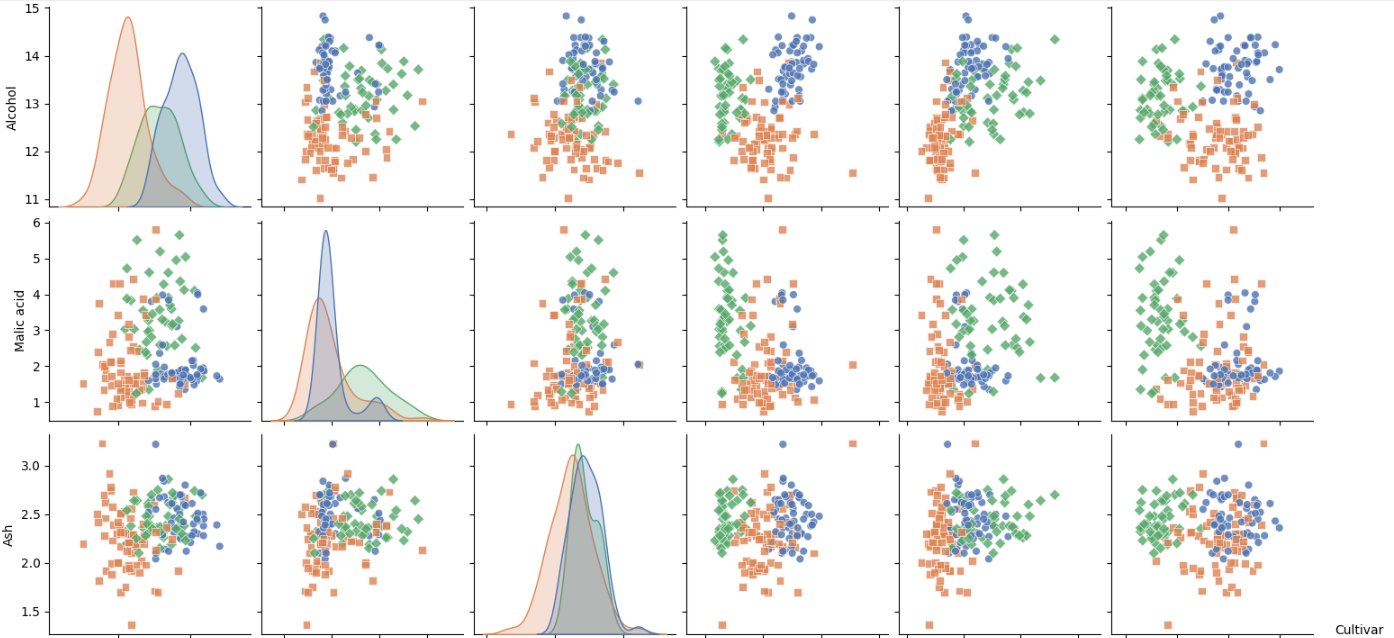
Do znalezienia które zestawienia parametrów pozwalają na rozróżnienie win z poszczególnych odmian skorzystamy z pairplota z podziałem na odmiany winogron.

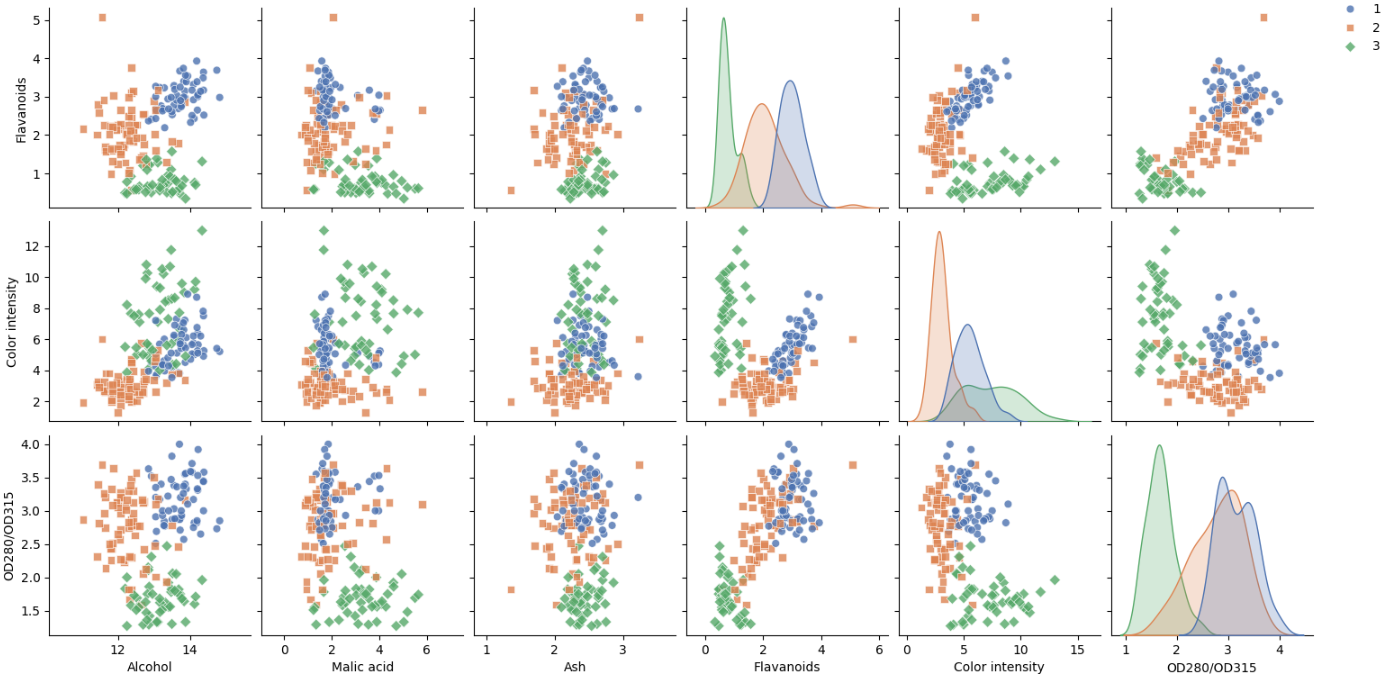
By znaleźć odpowiednie zestawienia parametrów szukamy na pairplocie tych połączeń które mają jak najmniej nachodzących na siebie punktów. I dzięki temu możemy zauważyć, że:

* Zestawienie Alkoholu oraz flawonoidów punkty nie nachodzą na siebie co sugeruje że możemy skorzystać z tego zestawienia do rozróżniania odmian
* Zestawienie Alkoholu i absorbcji również ma zauważalne grupy co sugeruje że możemy z niego korzystać.
* Zestawienie kwasu jabłkowego i flawonoidów również jest stosunkowo dobrym zestawieniem parametrów.
* Zestawienie flawonoidów oraz intensywności kolorów też jest dobry
* Zestawienie intensywności kolorów oraz absorpcja też ma punkty które w większości się nie mieszają.

Są jeszcze takie zestawienia jak flawonoidy i absorpcja czy alkohol i intensywność kolorów które też moglibyśmy wykorzystać do rozróżnienia odmian jednak podane wyżej są dużo mocniejsze.

Dodatkowo możemy zauważyć że same flawonoidy czy intensywność koloru są dobrymi parametrami do rozróżniania odmiany winogron, co można zauważyć na przekątnej osi gdzie wykresy nie są bardzo nałożone na siebie.





Zbadamy jeszcze korelację poszczególnych parametrów, w tym celu skorzystamy z heatmapy która jest wizualizacją korelacji uzyskanych z danych. Warto zaznaczyć tutaj że badamy korelacje dodatnie, ujemne oraz brak korelacji gdzie wartości blisko -1 to mocna korelacja ujemna co oznacza że gdy jeden parametr maleje to drugi rośnie, odwrotność tej korelacji to wartości bliskie 1 co oznacza mocną korelację dodatnią, która charakteryzuje się tym, że gdy jedna cecha rośnie to druga też rośnie. Ostatnie wartości to wartości w okolicach 0 co wskazuje na brak korelacji, czyli nie możemy jednoznacznie wskazać czy wzrost lub spadek jednej cechy powoduje zmiany w innej. 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 nie jest znacząca statystycznie dla nas.

Korelacje dodatnie:

* Odmiana i Kwas jabłkowy (0.44 – umiarkowana korelacja)
* Odmiana i Intensywność koloru (0.27 – słaba korelacja)
* Alkohol i popiół (0.21 – słaba korelacja)
* Alkohol i flawonoidy (0.24 – słaba korelacja)
* Alkohol i intensywność koloru (0.55 – umiarkowana korelacja)
* Kwas jabłkowy i Intensywność koloru (0.25 – słaba korelacja)
* Popiół i intensywność koloru (0.26 – słaba korelacja)
* Flawonoidy i absorpcja (0.79 – mocna korelacja)

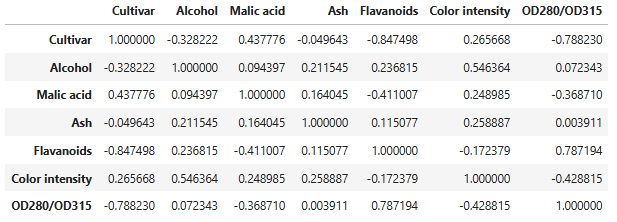
Korelacje ujemne:

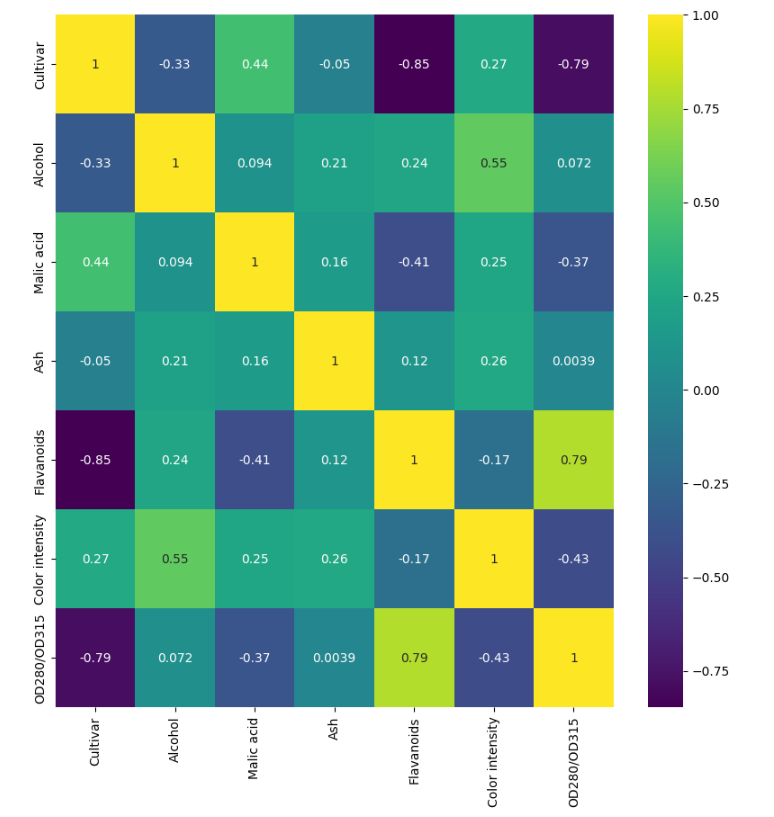
* Odmiana i alkohol (-0.33 – słaba korelacja)
* Odmiana i flawonoidy (-0.85 – mocna korelacja)
* Odmiana i absorpcja (-0.79 – mocna korelacja)
* Kwas jabłkowy i flawonoidy (-0.41 – umiarkowana korelacja)
* Kwas jabłkowy i absorpcja (-0.37 – słaba korelacja)
* Intensywność koloru i absorpcja (-0.43 – umiarkowana korelacja)

Brak korelacji:

* Odmiana i popiół (-0.05)
* Alkohol i kwas jabłkowy (0.094)
* Alkohol i absorpcja (0.072)
* kwas jabłkowy i popiół (0.16)
* Popiół i flawonoidy (0.12)
* Popiół i absorpcja (0.0039)
* Flawonoidy i intensywność kolory (-0.17)

Korelacja w postaci tabeli



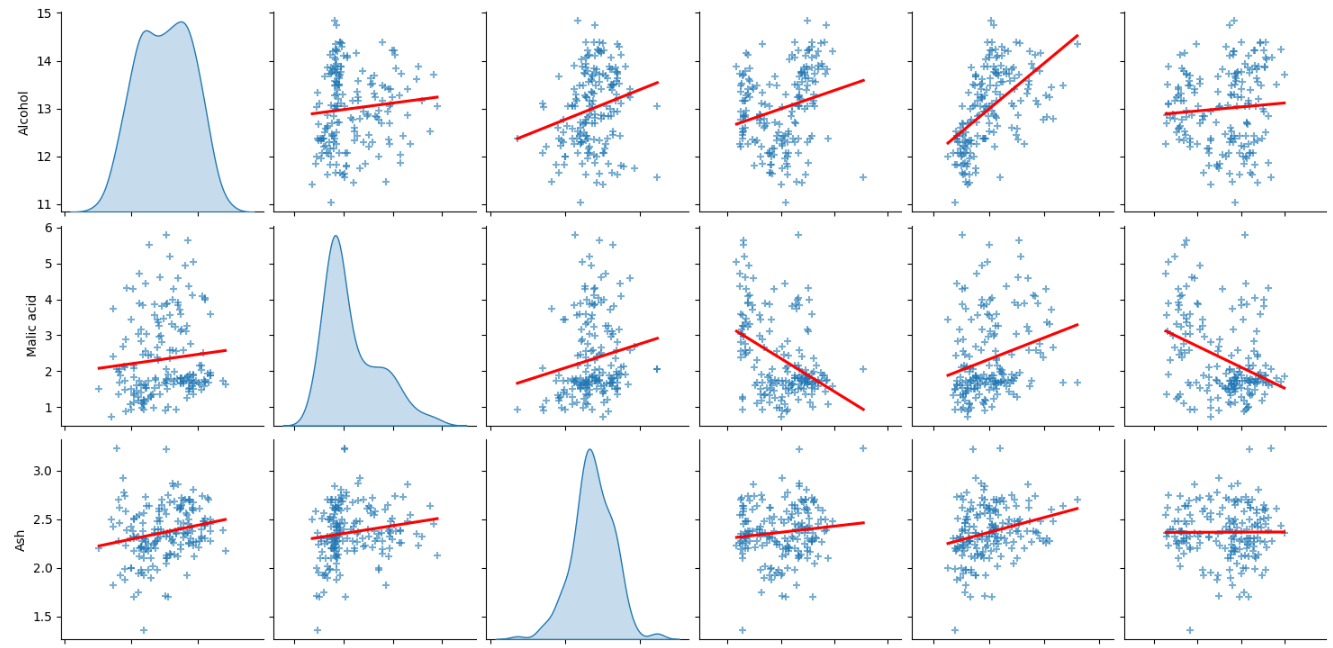
Wizualizacja tabeli jako heatmap 

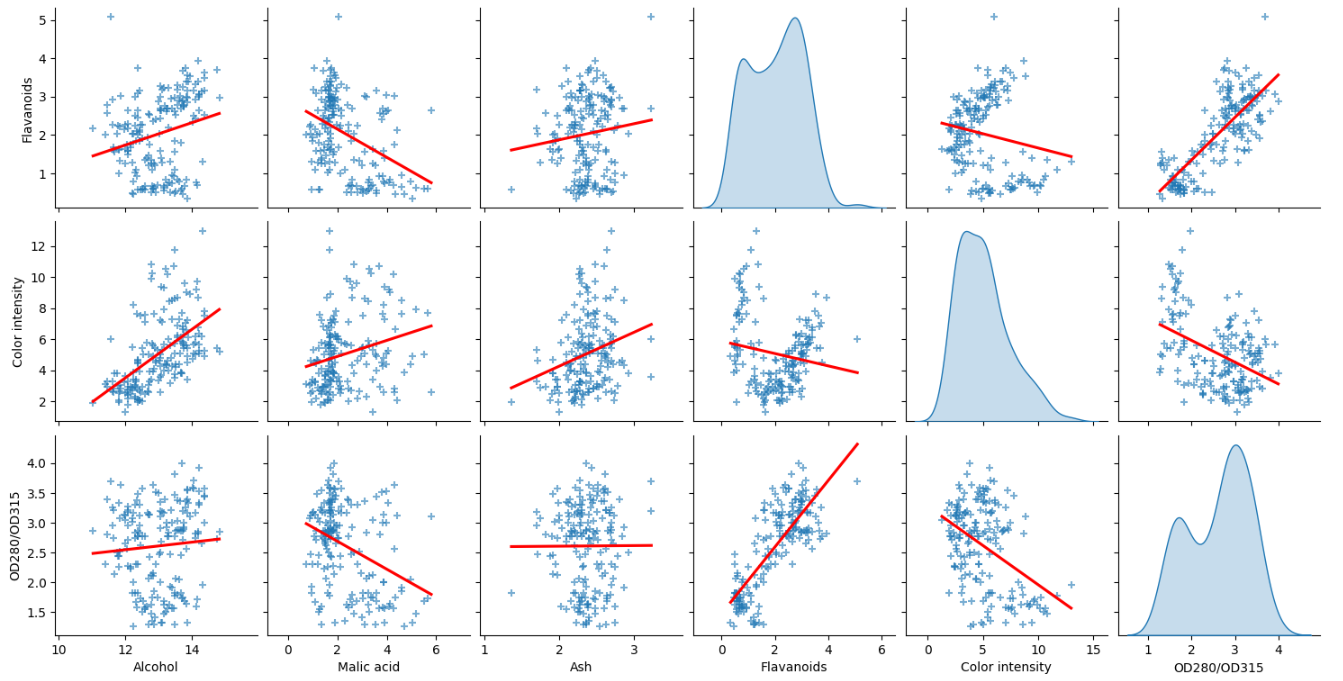
1. **Jakie związki między poszczególnym parametrami win możemy zaobserwować, z uwzględnieniem kierunku i siły relacji dla całości zestawu danych oraz dla win wykonanych z poszczególnych odmian winogron, a także różnic w trendach obserwowanych dla całości zestawu danych oraz dla win wykonanych z poszczególnych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

Zaczniemy naszą analizę dla całego zestawu danych. W tym celu skorzystamy z pairplot z dodatkowa prostą regresji która wskaże nam kierunek i siłę relacji, później te proste regresji wykorzystamy do porównania trendów między całym zestawem danych a danymi z podziałem na odmiany.

**Analiza kierunku i siły relacji:**

* Zauważamy że dla alkoholu praktycznie wszystkie linie regresji są dodatnie, gdzie najbardziej stroma czyli najmocniejsza jest korelacja alkoholu z intensywnością koloru, natomiast alkohol i kwas Jabłkowy czy alkohol i absorpcja przypominają bardziej linię prostą co wskazuje na brak zauważalnej korelacji
* Kwas jabłkowy posiada dwie korelacje ujemne oraz trzy korelacje które są dodatnie bądź zbliżone do dodatnich. Złączenie kwasu Jabłkowego oraz flawonoidów czy absorbcji wskazuje na korelacje ujemną, stosunkowo mocną, bądź umiarkowaną ze względu na nachylenie krzywych. Natomiast kwas jabłkowy i intensywność koloru są skorelowane dodatnio, jednak słabo. Natomiast korelacja z alkoholem czy popiołem już bardziej przypomina linię prostą przez co uznajemy to za brak korelacji, mimo, że Kwas jabłkowy i popiół są na granicy z korelacją słabą.
* Popiół zawiera bardzo dużo linii regresji podobnych do prostej przez co wszystkie złączenia cech nie są skorelowane po za intensywnością koloru która jest trochę bardziej stroma od reszty jednak jest to korelacja słaba.
* Flawonoidy zawiera zarówno proste regresji rosnące jak i malejące, które są ważne statystycznie oraz większość z nich jest wystarczająco silne by mówić tutaj o korelacji dodatniej lub ujemnej. Zauważamy dodatnią korelację dla porównania z alkoholem czy absorbcją, gdzie z absorpcja jest mocno skorelowana co widać po bardzo stromej prostej. Natomiast Kwas jabłkowy jest skorelowany ujemnie. Popiół oraz intensywność koloru są na granicy bycia słabymi korelacjami jednak przypominają już za bardzo linie proste.
* Intensywność koloru zawiera 3 pierwsze parametry skorelowane pozytywnie czyli alkohol, kwas jabłkowy i popiół, wszystkie są korelacjami słabymi lub umiarkowanymi. Absorbcja natomiast jest skorelowana ujemnie słabo / umiarkowanie. Tylko flawonoidy nie są skorelowane.
* Absorbcja zawiera dwie wartości nie skorelowane, czyli alkohol i popiół, natomiast flawonoidy są mocno skorelowane pozytywnie. Kwas jabłkowy czy Intensywność koloru z drugiej strony są skorelowane negatywnie na poziomie umiarkowanym\mocnym





**Analiza kierunku i siły relacji z podziałem na odmiany – w tej części dokonamy analizy korelacji z podziałem na odmiany, będziemy teraz posiadać 3 krzywe regresji dla każdej z odmian.**

Analiza:

* Alkohol oraz Kwas jabłkowy widocznie nie są skorelowane we wszystkich odmianach winogron, wskazuje na to, że wszystkie 3 linie regresji są praktycznie poziome. Jeżeli chodzi o popiół to zauważamy że o ile dla odmiany 1 i 2 linia jest skierowana trochę w dół co wskazuje na brak albo słabą korelację ujemną, tak dla odmiany 3 zauważamy korelacje dodatnią na poziomie słabym, co wskazuje, że nie trend z linii dla wszystkich danych nie pokrywa się dla odmiany 3 w tym zestawieniu. To samo tyczy się flawonoidów gdzie tym razem odmiana 3 jest dodatnia a reszta nie wskazuje korelacji. Natomiast dla intensywności koloru widzimy mocną korelację dla wszystkich 3 odmian co pokrywa się z trendem linii głównej. Dla absorbcji zachowujemy trend linii głównej i jak widzimy wszystkie 3 regresje przypominają linie poziome
* Kwas jabłkowy oraz popiół pokazuje że dla każdej z odmian linia regresji przypomina linie poziomą co zgadza się z trendami, natomiast interesujący jest porównanie kwasu jabłkowego z flawonoidami, tam odmiany 1 i 3 są skorelowane ujemnie gdzie 1 jest mocno skorelowana, natomiast 2 jest skorelowana dodatnio ale bardzo słabo. Wiec dla odmiany 2 trend nie jest zachowany. Jeżeli zobaczymy intensywność koloru to trend ujemny dobrze widać na każdej z odmian co kompletnie nie pokrywa się ogólnym trendem gdzie jest on dodatni. To samo zauważamy dla absorbcji gdzie każda z odmian ma trend dodatni albo bliski dodatniemu a ogólny trend jest bliżej trendu ujemnego.
* Popiół i Flawonoidy również ma ciekawy wygląd, gdzie dla odmian 2 i 3 linia jest dodatnia a dla 1 jest prawie pozioma jednak delikatnie skierowana w dół, nie mniej jednak pokrywa się to z ogólnym trendem. Taką samą sytuacje zauważamy dla intensywności koloru tylko nachylenia linii dodatnich są mniejsze co by sugerowało że regresja nie zachodzi a jednak trend ogólny jest dodatni co się nie pokrywa. Również tą samą sytuacje zauważamy dla absorbcji jednak tutaj trendy się pokrywają bo zarówno trend ogólny jak i pojedyncze trendy dla odmian przypominają linie poziomą
* Flawonoidy oraz intensywność koloru dla każdego z odmian jest dodatni, gdzie dla odmiany 3 jest bardziej przybliżony do linii poziomej. Co zupełnie nie pokrywa się z trendem ogólnym gdzie linia regresji jest raczej ujemna. Dla absorbcji sytuacja wygląda inaczej, jeżeli chodzi o odmianę 2 jest ona dodatnie skorelowana umiarkowanie / mocno natomiast dla odmiany 1 i 3 zauważamy brak korelacji / korelacje słabą ujemną, więc dla odmiany 2 trend ogólny się pokrywa jednak dla odmian 1 i 3 już nie.
* Intensywność koloru tutaj badamy absorbcję, dla wszystkich odmian korelacja jest słaba ujemna / brak korelacji co pokrywa się z ogólnym trendem wskazującym na korelacje ujemną

