

# Analiza parametrów win

Celem poniższej pracy jest **przedstawienie wyników** analizy statystycznej dokonanej na próbce danych, zawierającej wyniki pomiarów wybranych parametrów **win** pochodzących z tego samego regionu Włoch, ale wyprodukowanych z trzech różnych **odmian**.

**Odmiany** winogron oznaczone są kolejnymi liczbami: 1, 2, 3

Inne parametry to:

- Zawartość alkoholu
- Zawartość kwasu jabłkowego
- Popiół (składniki mineralne będące pozostałościami po fermentacji winogron)
- Flawonoidy
- Intensywność koloru
- Stosunek absorpcji światła przy długościach fali 280 nm i 315 nm.

W pracy zawarte są odpowiedzi na **trzy** zasadnicze pytania:

- 1) Pod jakimi względami wina wykonane z poszczególnych odmian winogron są podobne, a pod jakimi się różnią?
- 2) Które zestawienia parametrów pozwalają na rozróżnienie win wykonanych z poszczególnych odmian winogron?
- 3) Jakie związki między parametrami win możemy zaobserwować, z uwzględnieniem kierunku i siły relacji dla całości zestawu danych oraz dla win poszczególnych odmian?

Do wykonania analizy posłużył język **Python** oraz biblioteki: pandas, matplotlib, seaborn, scipy, numpy.

Numer Zespołu: 43

Autorzy:

- Kacper Potaczała 425724
- Robert Skulik 428339
- Maja Piątek 427763

# 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 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, odmiana 2 ma ją najmniejszą.
- **Kwas Jabłkowy:** Odmiana 3 ma wyższe średnie wartości i znacznie szerszy zakres niż odmiany 1 i 2. Odmiany 1 i 2 posiadają dużą liczbę wartości odstających, których obecność wpływa na wartość średniej w znacznym stopniu.
- **Flawonoidy:** Każda z odmian ma inną wartość średnią, gdzie 1 ma ją największą. Największy rozrzut wartości wykazuje odmiana 2, a najmniejszy odmiana 3.
- **Intensywność Koloru:** Odmiana 3 ma najwyższą wartość średnią i największe zróżnicowanie. Odmiana 2 jest najbardziej stabilna.
- **Absorbancja:** Odmiana 3 ma niższe wartości średnie od odmian 1 oraz 2 i najmniejsze rozproszenie.

## Analiza

Na podstawie **danych powyżej** możemy zauważyć następujące **interesujące zależności**.

Jeżeli zależy nam na **ilości alkoholu** to najlepszym wyborem byłyby wina z odmiany **1**, ponieważ średnio zawierają one **najwięcej** alkoholu - około **13.75 %**.

Podobnie należy wybrać odmianę **1**, gdybyśmy chcieli **zminimalizować** ilość **kwasu jabłkowego**, chociaż w tym wypadku odmiana **2** również jest **dobrym wyborem**. Należy tutaj jednak **zwrócić uwagę**, że jej **wartości odstające** są **większe**, co wiąże się z **większym ryzykiem** kupienia wina z **wyższą ilością kwasu jabłkowego**, niż gdyby wybrać wino z odmiany **1**.

Jeżeli chodzi o zawartość **popiołu**, to **nie** ma **dużego znaczenia** z jakiej odmiany zostanie wybrane wino, ponieważ ich **wartości średnie** są **bardzo zbliżone** do siebie, więc lepiej jest oprzeć wybór na **innych parametrach**, które są dla nas **ważne**.

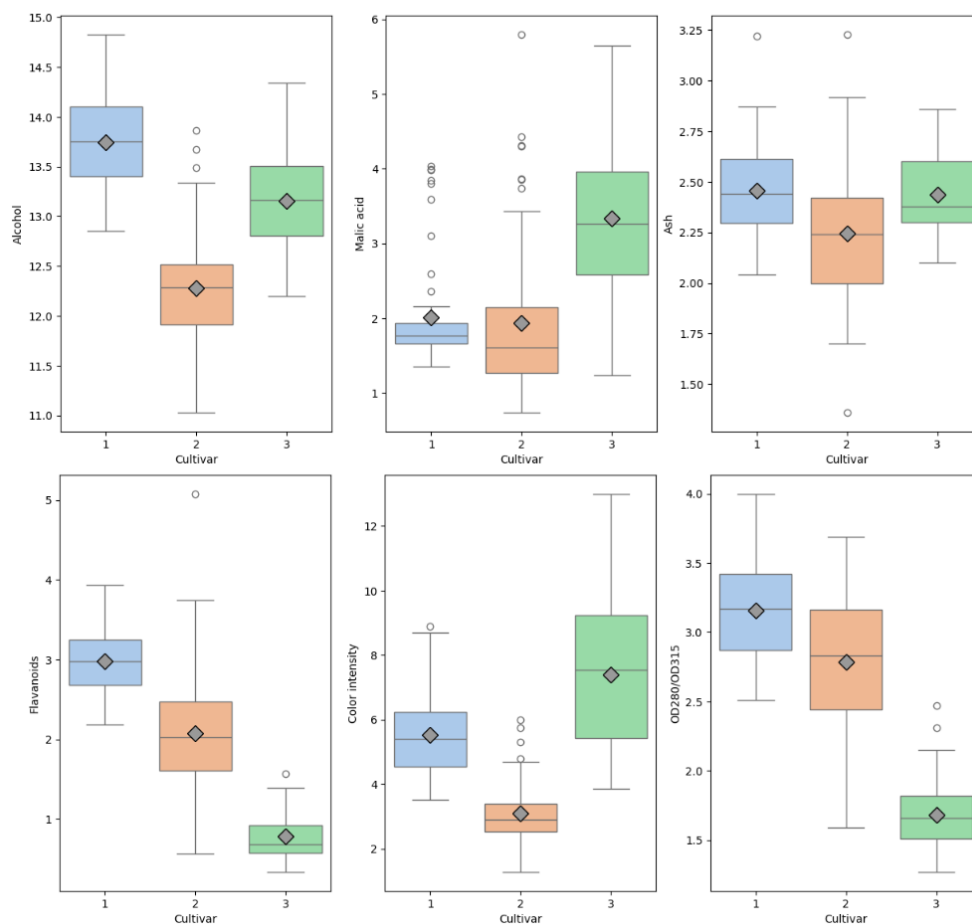
**Flawonoidy** bardzo dobrze pokazują **różnice** w winie. Każda z odmian ma inną **wartość średnią** - od mniej niż 1 dla odmiany **3**, aż do wartości 3 dla odmiany **1**, zatem jeżeli znamy swoje **preferencje**, to ten **parametr** świetnie wskazuje, jakie wino **powinniśmy wybrać**.

**Intensywność koloru** również różnicuje odmiany **wyraźnie**. Jeżeli preferujemy **ciemniejsze wina**, to odmiana **3** będzie **dobrym wyborem** ze względu na **najwyższą wartość średnią**. Odmiana **2** natomiast ma **najmniejszą wartość średnią**, więc jest odpowiednia dla osób wolących **stonowane kolory**.

Warto także zwrócić uwagę na **absorbancję**, która przypomina **trendy flawonoidów**, czyli **najmniejszą średnią absorbancję** ma odmiana **3** a **największą** odmiana **1**. **Możliwe**, że istnieje **zależność** między tymi **parametrami**, jednak na tym etapie analizy **nie możemy** tego jednoznacznie **potwierdzić**.

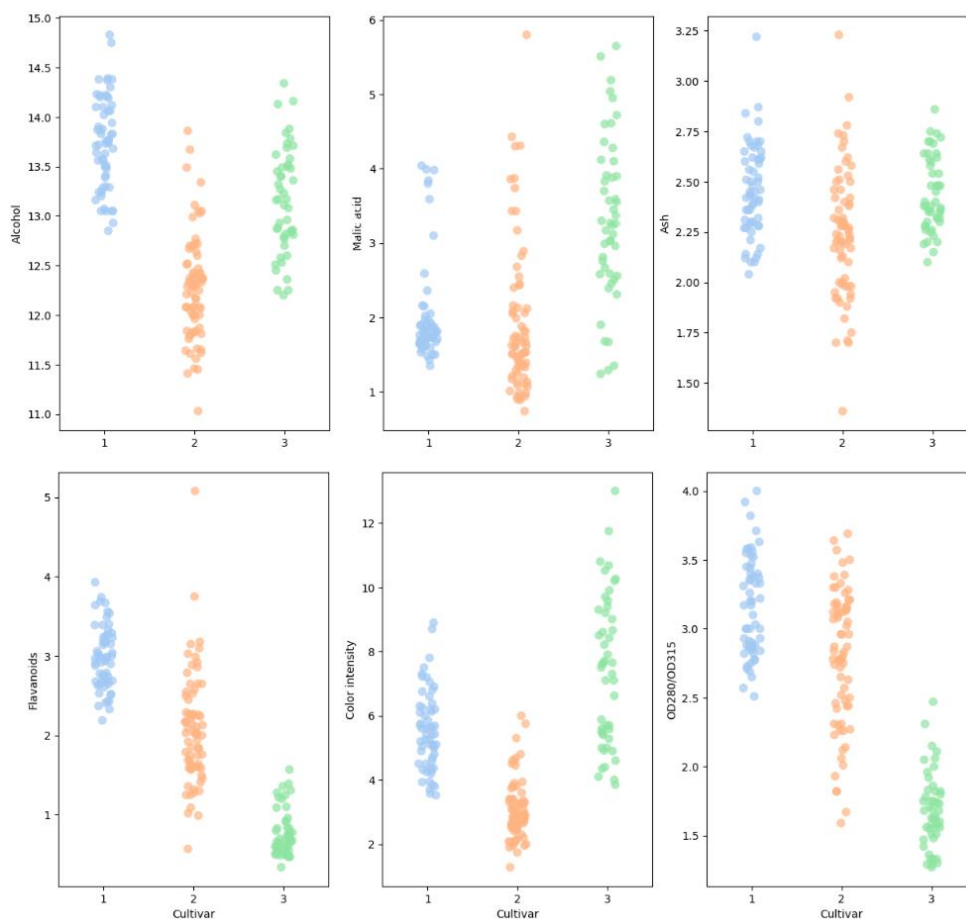
## Boxploty z podziałem na odmiany, każdy boxplot ma inny parametr na osi Y

Wine parameters by grape cultivar (Boxplots)



## Stripploty z podziałem na odmiany, każdy stripplot ma inny parametr na osi Y

Wine parameters by grape cultivar (Stripplots)



## 2. Które zestawienia parametrów pozwalają na rozróżnienie win wykonanych z poszczególnych odmian winogron?

### Zestawienia parametrów

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

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

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

- **Alkohol & Popiół**: można wydzielić grupę dla odmiany 2
- **Alkohol & Flawonoidy**: wyraźne grupy dla wszystkich odmian
- **Alkohol & Intensywność koloru**: można wydzielić grupę dla odmiany 2
- **Alkohol & Absorbancja**: bardzo wyraźne grupy dla wszystkich odmian
- **Kwas Jabłkowy & Flawonoidy**: stosunkowo dobrze wyraźne grupy dla wszystkich odmian
- **Kwas Jabłkowy & Absorbancja**: można wydzielić grupę dla odmiany 3
- **Popiół & Absorbancja**: można wydzielić grupę dla odmiany 3
- **Flawonoidy & Intensywność Koloru**: wyraźne grupy dla wszystkich odmian
- **Flawonoidy & Absorbancja**: można wydzielić grupę dla odmiany 3
- **Intensywność Koloru & Absorbancja**: stosunkowo dobrze wyraźne grupy dla wszystkich odmian

### Analiza

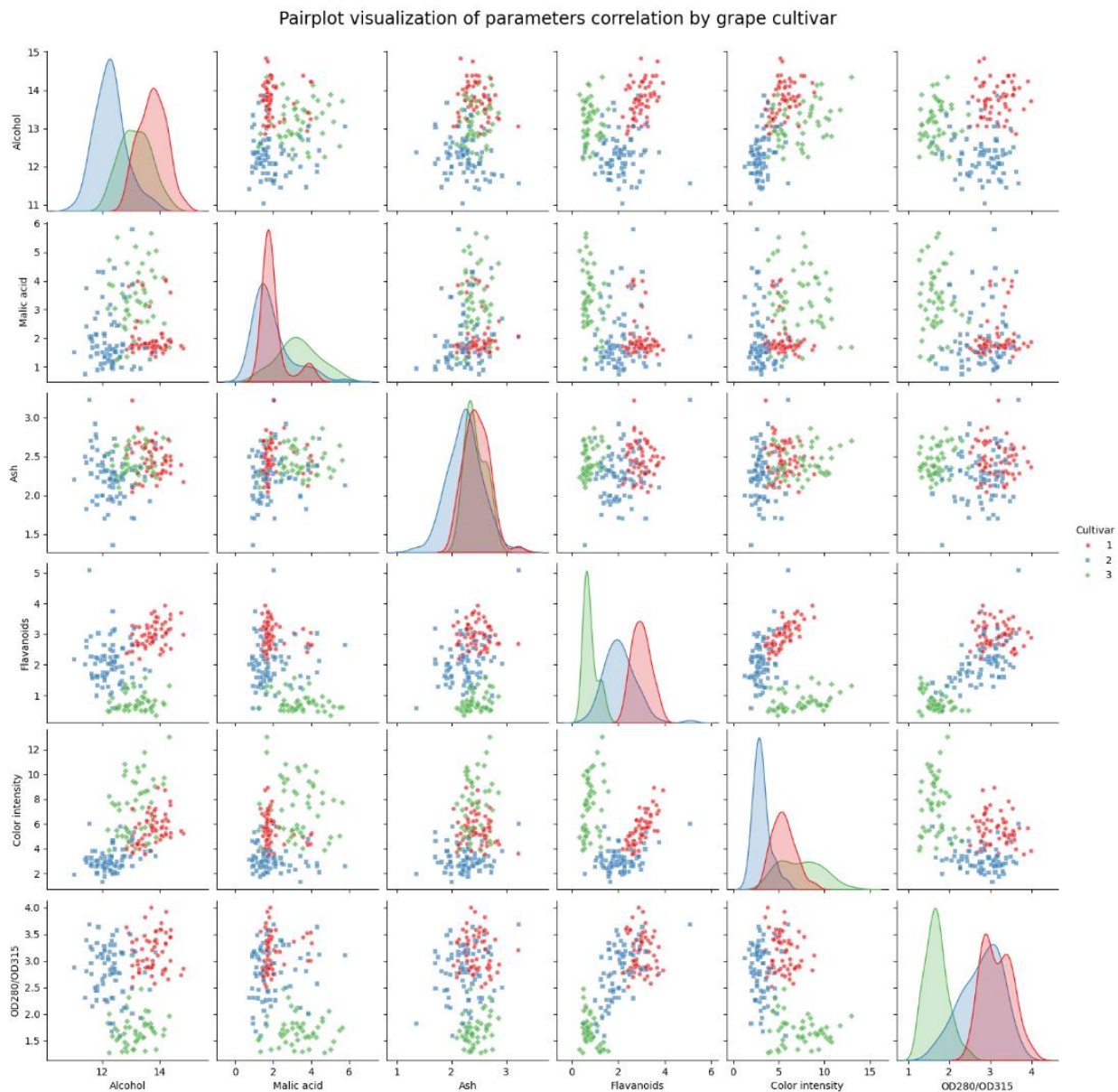
Dodatkowo, możemy zauważyć, że same **Flawonoidy** czy **Intensywność Koloru** są dobrymi **parametrami** do rozróżniania win wykonanych z badanych odmian winogron, co widać na **przekątnej osi**, gdzie wykresy **nie** są znacznie **nałożone** na siebie. Szczególnie we Flawonoidach wykres dla odmiany 3 nie nachodzi na resztę wykresów.

Możemy z tego wysnuć następujące **wnioski**: gdy znamy **zależności** między parametrami, np. takimi jak **Alkohol** i **Flawonoidy**, czy **Intensywność Koloru** i **Absorbancja**, możemy z dużym **prawdopodobieństwem** określić z jakiej **odmiany winogron** zostało wykonane wino.

Natomiast jeżeli chcielibyśmy **sprawdzić**, czy wino pochodzi z odmiany **3**, to **najlepszym** wyborem byłoby sprawdzenie **parametrów związanych z Flawonoidami**, ponieważ ta odmiana w tym aspekcie **wyraźnie** różni się od **pozostałych**, co można łatwo zauważyć na **przekątnej** wykresu pairplot, gdzie wykres zielony (odmiana 3) **nie nachodzi** na inne wykresy w **dużym stopniu**.

**Problematyczna** natomiast może być odmiana **1**, ponieważ **nie tworzy** oddzielnej grupy, gdzie odmiany 2 i 3 się **mieszają**, tak jak ma to miejsce w przypadku wymienionych **wcześniej** odmian, np. w zestawieniu **Alkoholu** i **Intensywności Koloru** możemy jedynie wydzielić grupę **2**. W takim wypadku musimy  **badać** te zestawienia, gdzie **wszystkie trzy** odmiany tworzą **oddzielne grupy**. Nie jest to bardzo duży kłopot, ale pozostawia nam on **mniej możliwości**.

## Pairplot z podziałem na odmiany, każdy kolor to inna odmiana



### 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 także), **ujemne** (gdy jedna wartość **rośnie**, druga **maleje**) oraz **brak** korelacji. Zastosujemy przedstawione niżej przedziały wartości:

|                        |    |                               |
|------------------------|----|-------------------------------|
| $ Kor  < 0.2$          | -> | <b>brak</b> korelacji         |
| $0.2 \leq  Kor  < 0.4$ | -> | <b>słaba</b> korelacja        |
| $0.4 \leq  Kor  < 0.7$ | -> | <b>umiarkowana</b> korelacja  |
| $0.7 \leq  Kor  < 0.9$ | -> | <b>mocna</b> korelacja        |
| $ Kor  \geq 0.9$       | -> | <b>bardzo</b> mocna korelacja |

Dodatkowo, na naszym wykresie **przekątna główna** składa się z samych **1**, co jest prawidłowe, ponieważ jest to korelacja danej cechy ze **samą sobą**. **Nie uwzględniamy** jej w analizie, gdyż nic do niej nie wnosi. Tak samo **nie uwzględniamy** korelacji z **odmianą** mimo że **znajduje** się na **wykresie**. Robimy tak, ze względu że jej **liczby** tylko **informują** nas o **rodzaju odmiany winogron** i **nie są znaczące statystycznie**.

## Korelacje dodatnie:

- 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:

- Kwas Jabłkowy & Flawonoidy: -0.41 – **umiarkowana** korelacja
- Kwas Jabłkowy & Absorbancja: -0.37 – **słaba** korelacja
- Intensywność Koloru & Absorbancja: -0.43 – **umiarkowana** korelacja

## Brak korelacji:

- 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

## Analiza

Patrząc na **korelacje** zebrane powyżej, możemy zauważyć kilka **istotnych zależności**.

**Najsilniejsza dodatnia** korelacja występuje między **flawonoidami** a **absorbancją**. Oznacza to, że gdy ilość flawonoidów **rośnie**, **zazwyczaj** również **rośnie** absorbancja. Co ciekawe, **pokrywa** się to z naszymi **podejrzeniami** z analizy pierwszego pytania. W **praktyce** może to okazać się **użyteczne** - jeżeli **nie znamy** wartości **absorbancji**, ale **znamy** ilość **flawonoidów**, to możemy z pewnym **przybliżeniem** oszacować **absorbancję**.

Dodatkowo, widzimy, że **ilość alkoholu** jest **umiarkowanie** i **dodatnio** skorelowana z **intensywnością koloru**, co oznacza, że wina o **wyższej zawartości alkoholu** mają **zazwyczaj intensywniejsze kolory**. Warto mieć to na **uwadze** przy wyborze wina. Również wino o **wysokiej zawartości alkoholu**, ale o **mało intensywnym kolorze** może być **trudniejsze** do znalezienia.

Spójrzmy jeszcze na **umiarkowaną ujemną** korelację **kwasu jabłkowego** i **flawonoidów**. Sugeruje to, że gdy zostanie wybrane wino z **większą ilością kwasu jabłkowego**, **może** wiązać się to ze **spadkiem flawonoidów**. Ponieważ **flawonoidy** są **dodatnio** skorelowane z **absorbancją**, to ta zmiana będzie też miała **wpływ** na **absorbancję**. Więc zmiana tego **jednego parametru** może mieć konsekwencje dla **innych właściwości** wina. Jednak należy pamiętać, że nie oznacza to **bezpośredniego związku przyczynowego**.

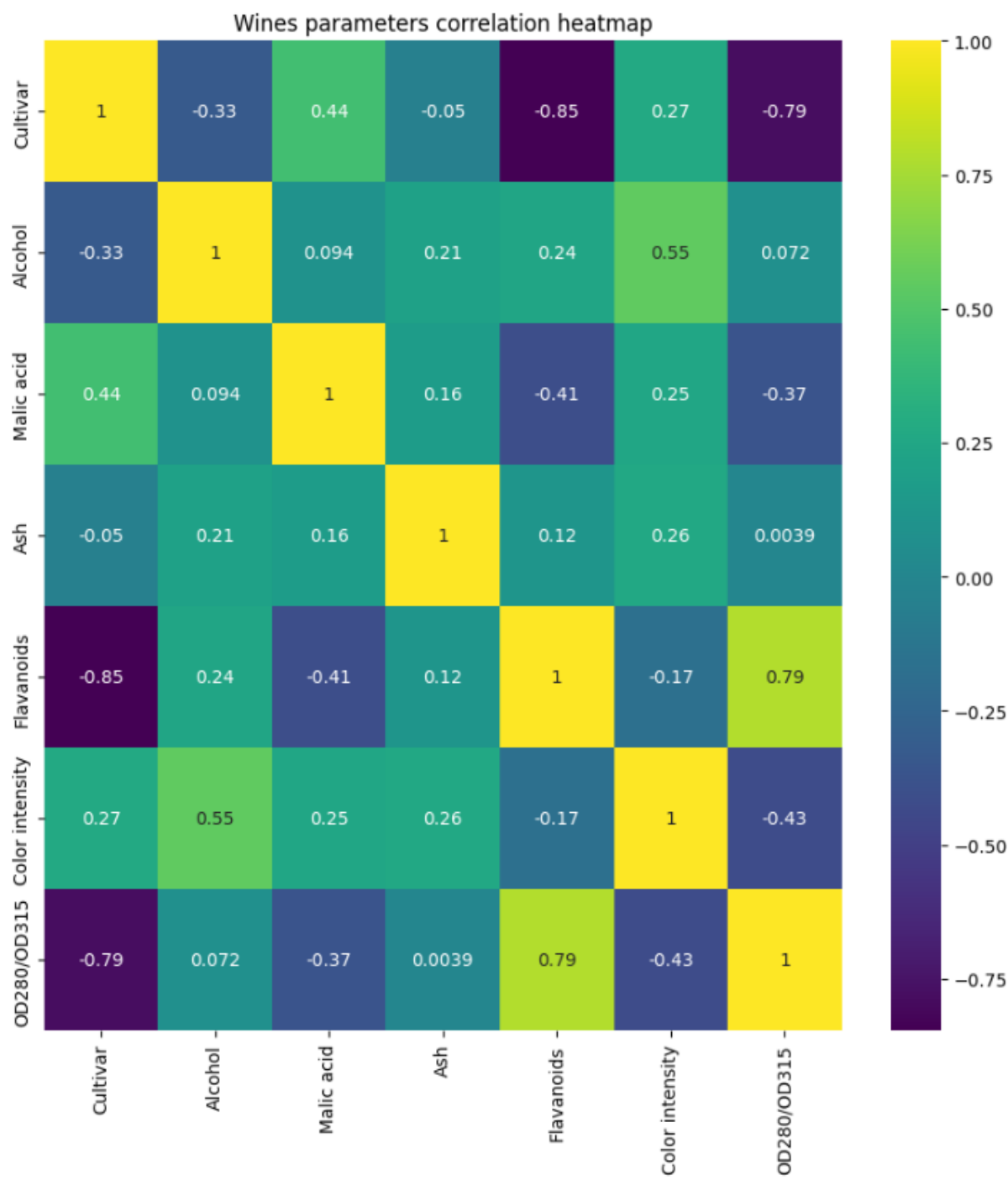
Podobną zależność zauważamy dla **intensywności koloru** i **absorbancji**. Również jest to korelacja **ujemna umiarkowana** i tak samo jak dla **kwasu jabłkowego** i **flawonoidów**, obserwujemy jak **jedna właściwość** może wpływać na **inne**.

Pozostałe **połączenia parametrów** mają korelacje **słabe** lub **nieistotne statystycznie**. **Nie** oznacza to, że **nie istnieją** między nimi jakieś **zależności**, jednak ich **wpływ** jest **mniejszy** niż w przypadkach **wymienionych wyżej**. W takim wypadku, przy **wyborze win** warto skupić się na tych **parametrach** które wykazują **silniejsze korelacje**.

Korelacja w postaci tabeli

|                 | 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    |

Wizualizacja tabeli jako heatmapa



### 3. Jakie związki między parametrami win możemy zaobserwować, z uwzględnieniem kierunku i siły relacji dla całości zestawu danych oraz dla win poszczególnych odmian?

#### 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 **stroma** prosta regresji -> **duża** siła korelacji
- Mniej **stroma** prosta regresji -> **mała** siła korelacji / **brak** korelacji

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 Jabłkowym** 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**

#### Analiza

Po zbadaniu **wizualizacji** można wyciągnąć następujące **wnioski**.

Dla **alkoholu** **potwierdzają** się korelacje analizowane w odpowiedzi na **drugie** pytanie. W szczególności wykres **liniowej regresji** dla **intensywności koloru** pokazuje **umiarkowaną dodatnią** zależność. Prosta regresji jest zauważalnie **dodatnie nachylona**.

Tak samo potwierdzają się **wyniki** dla **kwasu jabłkowego**, chociaż tu można zauważyć, że **korelacja** z **absorbancją** i **flawonoidami** jest **ujemna** i **słaba / umiarkowana**, jednakże to też **pokrywa się** z poprzednimi wynikami, ze względu na to, że wartości te występują na **pograniczu przedziałów**, które ustaliliśmy wcześniej. Proste regresji mają **zauważalne nachylenie w dół**.



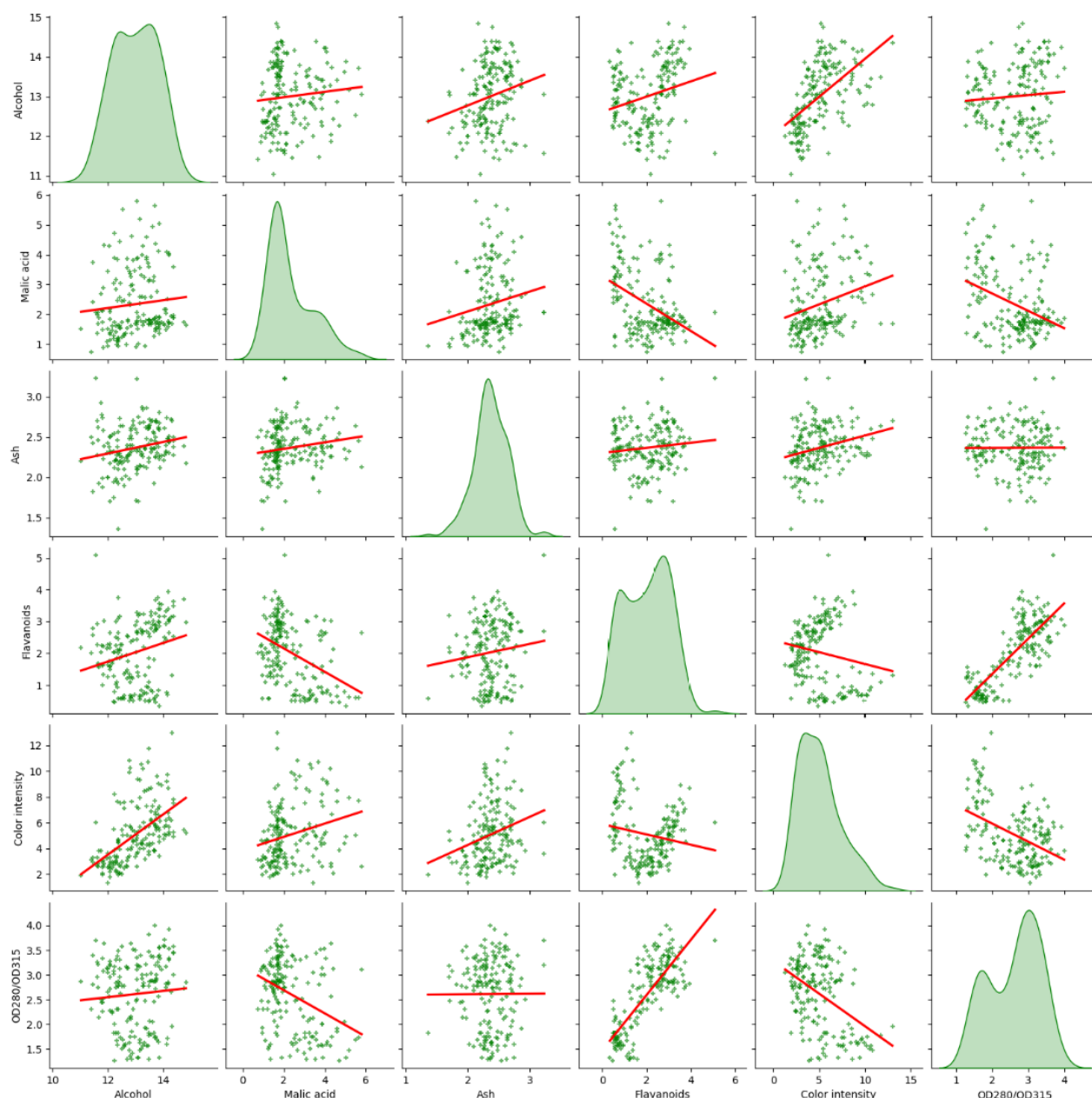
Dla **popiołu** wyniki **pokrywają** się w całości z **wcześniejszymi wnioskami** i bardzo dobrze można to **zauważyć** na **wykresach**.

**Flawonoidy** są **mocno** skorelowane z **absorbancją**, co widoczne jest na **wykresie**, **prosta regresji** jest **stroma** i skierowana w **górę**, co wskazuje na **silną liniową zależność** między tymi **cechami**. Dodatkowo, zauważamy, że korelacja z **intensywnością koloru** jest **ujemna**, jednak wydaje się ona **słaba** – **nachylenie** prostej regresji jest **niewielkie**.

**Intensywność koloru** wskazuje **umiarkowaną** i **ujemną** korelację z **absorbancją**, prosta regresji skierowana jest w **dół**, a **nachylenie** wskazuje już na korelację **umiarkowaną**.

### Pairplot z prostą regresji bez podziału na odmiany

Parameters with linear regression on Pairplot



### Zestaw danych z podziałem na odmiany.

Dla wykresów **pairplot** 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**, czyli jej kierunek i siła będą się mniej więcej zgadzać, to 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**, przy czym mają **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 **mniejsza siła** relacji

## Analiza

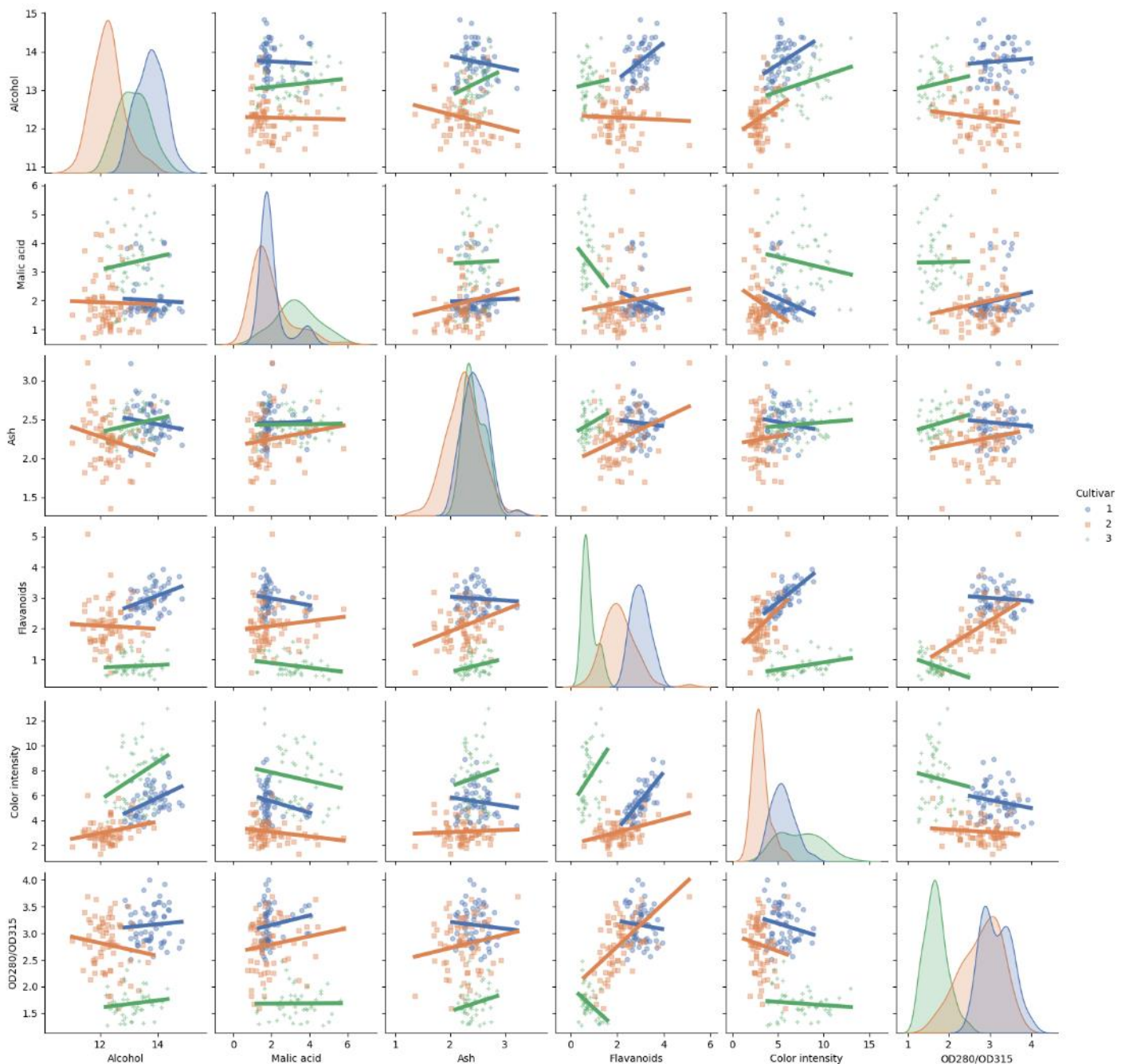
Ilość danych powyżej jest duża, dlatego skupimy się na **trendach** i tym, co możemy dzięki temu **zauważyć** i na co **uważać**.

Ważne jest tutaj co **oznacza** dla nas, że trend jest **taki sam** albo **podobny**, a co oznacza, że trend jest **inny**. Gdy trendy są **podobne**, to możemy do analizowania korelacji korzystać z **prostej regresji** wygenerowanej na podstawie **całości danych bez** podziału na **odmiany** i nasz **wynik** będzie mocno **przybliżony** do wyniku z **podziałem na odmiany**. Natomiast gdy trendy są **inne**, możemy zrobić **bardzo duży błąd** w analizie, jeżeli skorzystamy z **trendów ogólnych**, a mamy zbadać korelację dla **danej odmiany**, która akurat **nie pokrywa** się z nimi. Możemy np. stwierdzić, że **korelacja** jest **dodatnia** i to **umiarkowana**, ale dla danej **odmiany** może ona być **ujemna** co skutkować może zupełnie **źle** przeanalizowanymi **parametrami**, dlatego jest to tak **ważne**.

Podsumowując, musimy zdecydować, kiedy **bardziej opłaca** nam się zastosować **korelację ogólną** a kiedy musimy **rozdzielić** ją na **poszczególne odmiany**, dlatego właśnie **powyższe obserwacje** są kluczowe dla **poprawnej analizy**.

### Pairplot z prostą regresji oraz podziałem na odmiany

Parameters divided by cultivar with linear regression on Pairplot



Bibliografia:

Źródło danych:

Aeberhard S., Forina M. Leardi R.; Wine; DOI:10.24432/C5PC7J; <https://archive.ics.uci.edu/dataset/109/wine>

Biblioteki:

<https://pandas.pydata.org/>

<https://numpy.org/>

<https://seaborn.pydata.org/>

<https://matplotlib.org/>

<https://scipy.org/>

Użyte funkcje:

[https://pandas.pydata.org/docs/reference/api/pandas.read\\_csv.html](https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html)

[https://matplotlib.org/stable/api/\\_as\\_gen/matplotlib.pyplot.subplots.html](https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.subplots.html)

<https://seaborn.pydata.org/generated/seaborn.boxplot.html>

[https://matplotlib.org/stable/api/\\_as\\_gen/matplotlib.pyplot.show.html](https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.show.html)

<https://seaborn.pydata.org/generated/seaborn.stripplot.html>

<https://seaborn.pydata.org/generated/seaborn.pairplot.html>

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html>

<https://seaborn.pydata.org/generated/seaborn.heatmap.html>