# Proiect DSAD

## Analiza de clusteri

Setul de date ales este preluat din sursa <https://archive.ics.uci.edu/dataset/186/wine+quality>, aici sunt incluse două seturi de date referitoare la probe de vin roșu și alb, din nordul Portugaliei.

Acest set de date se bazează pe rezultatele unor teste fizico-chimice și evaluări senzoriale ale vinurilor. Variabilele de intrare sunt rezultatele testelor fizico-chimice, dar și o variabilă de ieșire, o evaluare a calității vinului, bazată pe date senzoriale.

Descrierea variabilelor:

1. Aciditate fixă (**fixed acidity**): Cantitatea de aciditate non-volatile din vin, exprimată în grame de acid tartric per litru.
2. Aciditate volatilă (**volatile acidity**): Cantitatea de acid acetic din vin, exprimată în grame de acid acetic per litru de vin.
3. Acid citric (**citric acid**): Cantitatea de acid citric din vin, exprimată în grame de acid citric per litru de vin.
4. Zahăr rezidual (**residual sugar**): Cantitatea de zahăr rămas în vin după fermentație, exprimată în grame de zahăr per litru de vin.
5. Cloruri (**chlorides**): Cantitatea de cloruri din vin, exprimată în grame de cloruri de sodiu per litru de vin.
6. Dioxid de sulf liber (**free sulfur dioxide**): Cantitatea de dioxid de sulf liber din vin, exprimată în miligrame de dioxid de sulf liber per litru de vin.
7. Dioxid de sulf total (**total sulfur dioxide**): Cantitatea totală de dioxid de sulf liber și legat din vin, exprimată în miligrame de dioxid de sulf total per litru de vin.
8. Densitate (**density**): Densitatea vinului, exprimată în grame per centimetru cub.
9. **pH**: Măsură a acidității sau alcalinității vinului pe o scală de la 0 (acid) la 14 (alcalin).
10. Sulfați (**sulphates**): Cantitatea de sulfați din vin, exprimată în grame de potasiu de sulfat per litru de vin.
11. Alcool(**alcohol**): Conținutul de alcool din vin, exprimat în procente de volum.
12. Calitate (**quality**): O evaluare a calității vinului, dată de experți în degustare, pe o scală de la 0 la 10, unde 0 reprezintă cea mai slabă calitate și 10 cea mai înaltă calitate.

Pentru a putea analiza și interpreta dendograma rezultată în urmă aplicării algoritmului de clusterizare am subdimensionat și unit într-un singur fișier de intrare 50 de monstre de vin roșu și 32 de monstre de vin alb. În setul de date am adăugat un index pentru fiecare monstră “R” reprezentând vinurile roșii, iar “W” repreznită vinurile albe.

Setul de date analizat:

"index";"fixed acidity";"volatile acidity";"citric acid";"residual sugar";"chlorides";"free sulfur dioxide";"total sulfur dioxide";"density";"pH";"sulphates";"alcohol";"quality"  
Monstra R 1;7.4;0.7;0;1.9;0.076;11;34;0.9978;3.51;0.56;9.4;5  
Monstra R 2;7.8;0.88;0;2.6;0.098;25;67;0.9968;3.2;0.68;9.8;5  
Monstra R 3;7.8;0.76;0.04;2.3;0.092;15;54;0.997;3.26;0.65;9.8;5  
Monstra R 4;11.2;0.28;0.56;1.9;0.075;17;60;0.998;3.16;0.58;9.8;6  
Monstra R 5;7.4;0.7;0;1.9;0.076;11;34;0.9978;3.51;0.56;9.4;5  
Monstra R 6;7.4;0.66;0;1.8;0.075;13;40;0.9978;3.51;0.56;9.4;5  
Monstra R 7;7.9;0.6;0.06;1.6;0.069;15;59;0.9964;3.3;0.46;9.4;5  
Monstra R 8;7.3;0.65;0;1.2;0.065;15;21;0.9946;3.39;0.47;10;7  
Monstra R 9;7.8;0.58;0.02;2;0.073;9;18;0.9968;3.36;0.57;9.5;7  
Monstra R 10;7.5;0.5;0.36;6.1;0.071;17;102;0.9978;3.35;0.8;10.5;5  
Monstra R 11;6.7;0.58;0.08;1.8;0.097;15;65;0.9959;3.28;0.54;9.2;5  
Monstra R 12;7.5;0.5;0.36;6.1;0.071;17;102;0.9978;3.35;0.8;10.5;5  
Monstra R 13;5.6;0.615;0;1.6;0.089;16;59;0.9943;3.58;0.52;9.9;5  
Monstra R 14;7.8;0.61;0.29;1.6;0.114;9;29;0.9974;3.26;1.56;9.1;5  
Monstra R 15;8.9;0.62;0.18;3.8;0.176;52;145;0.9986;3.16;0.88;9.2;5  
Monstra R 16;8.9;0.62;0.19;3.9;0.17;51;148;0.9986;3.17;0.93;9.2;5  
Monstra R 17;8.5;0.28;0.56;1.8;0.092;35;103;0.9969;3.3;0.75;10.5;7  
Monstra R 18;8.1;0.56;0.28;1.7;0.368;16;56;0.9968;3.11;1.28;9.3;5  
Monstra R 19;7.4;0.59;0.08;4.4;0.086;6;29;0.9974;3.38;0.5;9;4  
Monstra R 20;7.9;0.32;0.51;1.8;0.341;17;56;0.9969;3.04;1.08;9.2;6  
Monstra R 21;8.9;0.22;0.48;1.8;0.077;29;60;0.9968;3.39;0.53;9.4;6  
Monstra R 22;7.6;0.39;0.31;2.3;0.082;23;71;0.9982;3.52;0.65;9.7;5  
Monstra R 23;7.9;0.43;0.21;1.6;0.106;10;37;0.9966;3.17;0.91;9.5;5  
Monstra R 24;8.5;0.49;0.11;2.3;0.084;9;67;0.9968;3.17;0.53;9.4;5  
Monstra R 25;6.9;0.4;0.14;2.4;0.085;21;40;0.9968;3.43;0.63;9.7;6  
Monstra R 26;6.3;0.39;0.16;1.4;0.08;11;23;0.9955;3.34;0.56;9.3;5  
Monstra R 27;7.6;0.41;0.24;1.8;0.08;4;11;0.9962;3.28;0.59;9.5;5  
Monstra R 28;7.9;0.43;0.21;1.6;0.106;10;37;0.9966;3.17;0.91;9.5;5  
Monstra R 29;7.1;0.71;0;1.9;0.08;14;35;0.9972;3.47;0.55;9.4;5  
Monstra R 30;7.8;0.645;0;2;0.082;8;16;0.9964;3.38;0.59;9.8;6  
Monstra R 31;6.7;0.675;0.07;2.4;0.089;17;82;0.9958;3.35;0.54;10.1;5  
Monstra R 32;6.9;0.685;0;2.5;0.105;22;37;0.9966;3.46;0.57;10.6;6  
Monstra R 33;8.3;0.655;0.12;2.3;0.083;15;113;0.9966;3.17;0.66;9.8;5  
Monstra R 34;6.9;0.605;0.12;10.7;0.073;40;83;0.9993;3.45;0.52;9.4;6  
Monstra R 35;5.2;0.32;0.25;1.8;0.103;13;50;0.9957;3.38;0.55;9.2;5  
Monstra R 36;7.8;0.645;0;5.5;0.086;5;18;0.9986;3.4;0.55;9.6;6  
Monstra R 37;7.8;0.6;0.14;2.4;0.086;3;15;0.9975;3.42;0.6;10.8;6  
Monstra R 38;8.1;0.38;0.28;2.1;0.066;13;30;0.9968;3.23;0.73;9.7;7  
Monstra R 39;5.7;1.13;0.09;1.5;0.172;7;19;0.994;3.5;0.48;9.8;4  
Monstra R 40;7.3;0.45;0.36;5.9;0.074;12;87;0.9978;3.33;0.83;10.5;5  
Monstra R 41;7.3;0.45;0.36;5.9;0.074;12;87;0.9978;3.33;0.83;10.5;5  
Monstra R 42;8.8;0.61;0.3;2.8;0.088;17;46;0.9976;3.26;0.51;9.3;4  
Monstra R 43;7.5;0.49;0.2;2.6;0.332;8;14;0.9968;3.21;0.9;10.5;6  
Monstra R 44;8.1;0.66;0.22;2.2;0.069;9;23;0.9968;3.3;1.2;10.3;5  
Monstra R 45;6.8;0.67;0.02;1.8;0.05;5;11;0.9962;3.48;0.52;9.5;5  
Monstra R 46;4.6;0.52;0.15;2.1;0.054;8;65;0.9934;3.9;0.56;13.1;4  
Monstra R 47;7.7;0.935;0.43;2.2;0.114;22;114;0.997;3.25;0.73;9.2;5  
Monstra R 48;8.7;0.29;0.52;1.6;0.113;12;37;0.9969;3.25;0.58;9.5;5  
Monstra R 49;6.4;0.4;0.23;1.6;0.066;5;12;0.9958;3.34;0.56;9.2;5  
Monstra R 50;5.6;0.31;0.37;1.4;0.074;12;96;0.9954;3.32;0.58;9.2;5  
Monstra W 1;7;0.27;0.36;20.7;0.045;45;170;1.001;3;0.45;8.8;6  
Monstra W 2;6.3;0.3;0.34;1.6;0.049;14;132;0.994;3.3;0.49;9.5;6  
Monstra W 3;8.1;0.28;0.4;6.9;0.05;30;97;0.9951;3.26;0.44;10.1;6  
Monstra W 4;7.2;0.23;0.32;8.5;0.058;47;186;0.9956;3.19;0.4;9.9;6  
Monstra W 5;7.2;0.23;0.32;8.5;0.058;47;186;0.9956;3.19;0.4;9.9;6  
Monstra W 6;8.1;0.28;0.4;6.9;0.05;30;97;0.9951;3.26;0.44;10.1;6  
Monstra W 7;6.2;0.32;0.16;7;0.045;30;136;0.9949;3.18;0.47;9.6;6  
Monstra W 8;7;0.27;0.36;20.7;0.045;45;170;1.001;3;0.45;8.8;6  
Monstra W 9;6.3;0.3;0.34;1.6;0.049;14;132;0.994;3.3;0.49;9.5;6  
Monstra W 10;8.1;0.22;0.43;1.5;0.044;28;129;0.9938;3.22;0.45;11;6  
Monstra W 11;8.1;0.27;0.41;1.45;0.033;11;63;0.9908;2.99;0.56;12;5  
Monstra W 12;8.6;0.23;0.4;4.2;0.035;17;109;0.9947;3.14;0.53;9.7;5  
Monstra W 13;7.9;0.18;0.37;1.2;0.04;16;75;0.992;3.18;0.63;10.8;5  
Monstra W 14;6.6;0.16;0.4;1.5;0.044;48;143;0.9912;3.54;0.52;12.4;7  
Monstra W 15;8.3;0.42;0.62;19.25;0.04;41;172;1.0002;2.98;0.67;9.7;5  
Monstra W 16;6.6;0.17;0.38;1.5;0.032;28;112;0.9914;3.25;0.55;11.4;7  
Monstra W 17;6.3;0.48;0.04;1.1;0.046;30;99;0.9928;3.24;0.36;9.6;6  
Monstra W 18;6.2;0.66;0.48;1.2;0.029;29;75;0.9892;3.33;0.39;12.8;8  
Monstra W 19;7.4;0.34;0.42;1.1;0.033;17;171;0.9917;3.12;0.53;11.3;6  
Monstra W 20;6.5;0.31;0.14;7.5;0.044;34;133;0.9955;3.22;0.5;9.5;5  
Monstra W 21;6.2;0.66;0.48;1.2;0.029;29;75;0.9892;3.33;0.39;12.8;8  
Monstra W 22;6.4;0.31;0.38;2.9;0.038;19;102;0.9912;3.17;0.35;11;7  
Monstra W 23;6.8;0.26;0.42;1.7;0.049;41;122;0.993;3.47;0.48;10.5;8  
Monstra W 24;7.6;0.67;0.14;1.5;0.074;25;168;0.9937;3.05;0.51;9.3;5  
Monstra W 25;6.6;0.27;0.41;1.3;0.052;16;142;0.9951;3.42;0.47;10;6  
Monstra W 26;7;0.25;0.32;9;0.046;56;245;0.9955;3.25;0.5;10.4;6  
Monstra W 27;6.9;0.24;0.35;1;0.052;35;146;0.993;3.45;0.44;10;6  
Monstra W 28;7;0.28;0.39;8.7;0.051;32;141;0.9961;3.38;0.53;10.5;6  
Monstra W 29;7.4;0.27;0.48;1.1;0.047;17;132;0.9914;3.19;0.49;11.6;6  
Monstra W 30;7.2;0.32;0.36;2;0.033;37;114;0.9906;3.1;0.71;12.3;7  
Monstra W 31;8.5;0.24;0.39;10.4;0.044;20;142;0.9974;3.2;0.53;10;6  
Monstra W 32;8.3;0.14;0.34;1.1;0.042;7;47;0.9934;3.47;0.4;10.2;6

În urma aplicării scriptului de Python care utilizează metoda ward pentru a implementa un algoritm de clusterizare ierarhică aglomerativă în scopul analizei setului de date referitor la vinuri am obținut:

Având în vedere că "R" reprezintă vinuri roșii și "W" reprezintă vinuri albe, putem explora și mai departe diferențele dintre cele două tipuri de vinuri în cadrul clusterelor identificate.

Clusterul 1 (Vinuri Roșii) – culoarea figura 1

Observațiile (Monstrele): Acest cluster este dominat de vinurile roșii, eticheta “R” având caracteristici fizico-chimice și organoleptice variate, dar predominanța este dată de acest tip de vin. Caracteristicile acestui cluster ar putea indica o diversitate semnificativă în tipurile de vinuri roșii, care pot fi asociate cu diferite caracteristici precum aciditatea, zahărul rezidual, etc.

Clusterul 2 (Vinuri Albe) – culoarea figura 1

Observațiile (Monstrele): Acest cluster pare să conțină în principal vinuri albe, eticheta “W” dar și câteva vinuri roșii. Caracteristicile vinurilor albe din acest cluster ar putea să fie diferite de cele din clusterul 1, reflectând variații în procesele de producție și caracteristicile organoleptice specifice acestui tip de vin. De asemenea, prezența unor vinuri roșii în acest cluster poate indica similarități în anumite caracteristici cu vinurile albe, sugerând posibile suprapuneri în profilul aromatic și gustativ al acestor vinuri.

Această interpretare ajustată ține cont de diferențele semnificative între vinurile roșii și cele albe și explică modul în care aceste diferențe pot influența gruparea lor în cadrul clusterelor identificate.

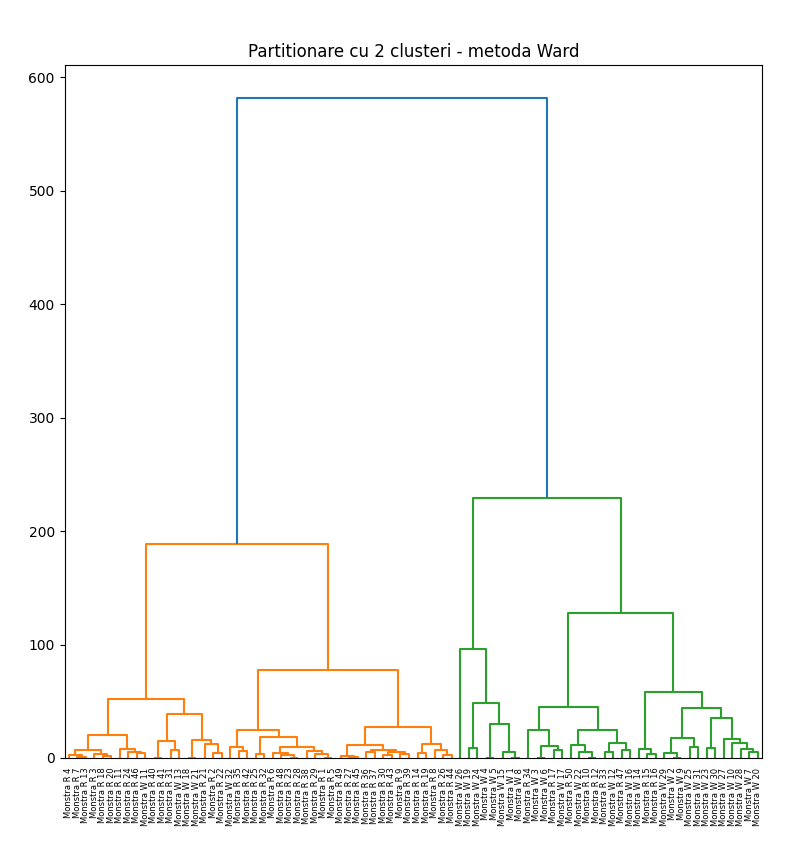


Figura 1

## Analiza în componente principale

Setul de date este preluat din sursa <https://archive.ics.uci.edu/dataset/59/letter+recognition> și conțin date despre identificarea literelor în baza descrierii pixelilor și a poziției acestora. Acestea includ poziția orizontală și verticală a cadrului care conține litera, lățimea și înălțimea cadrului, numărul total de pixeli negri, diverse momente statistice și numărul de muchii. Fiecare cadru este reprezentat printr-un număr mare de afișaje rectangulare de pixeli alb-negru care conține una dintre cele 26 de litere, majuscule din alfabetul englez.

Ca și variabile analizate dintre acestea se numără 17, incluzând categoria literei și 16 caracteristice numerice:

1. **lettr** (litera): Această variabilă reprezintă litera majusculă asociată imaginii respective. Are 26 de valori posibile, de la A la Z.
2. **x-box** (poziția orizontală a cadrului): Aceasta indică poziția orizontală a cadrului care conține litera în imaginea dată. Este exprimată ca un număr întreg.
3. **y-box** (poziția verticală a cadrului): Similar cu x-box, dar indică poziția verticală a cadrului care conține litera în imaginea dată. Este, de asemenea, exprimată ca un număr întreg.
4. **width** (lățimea cadrului): Acesta reprezintă lățimea cadrului care conține litera și este exprimată ca un număr întreg.
5. **high** (înălțimea cadrului): Similar cu width, dar indică înălțimea cadrului care conține litera și este exprimată ca un număr întreg.
6. **onpix** (numărul total de pixeli activați): Acesta indică numărul total de pixeli care sunt activați (în acest caz, pixelii negri) în imaginea cadrului care conține litera.
7. **x-bar** (media pozițiilor pe axa x a pixelilor activați): Reprezintă media pozițiilor pe axa x a pixelilor activați în cadru. Este exprimată ca un număr întreg.
8. **y-bar** (media pozițiilor pe axa y a pixelilor activați): Similar cu x-bar, dar indică media pozițiilor pe axa y a pixelilor activați în cadrului. Este, de asemenea, exprimată ca un număr întreg.
9. **x2bar** (variația mediei pe axa x): Acesta reprezintă variația mediei pozițiilor pe axa x a pixelilor activați. Este exprimată ca un număr întreg.
10. **y2bar** (variația mediei pe axa y): Similar cu x2bar, dar indică variația mediei pozițiilor pe axa y a pixelilor activați. Este, de asemenea, exprimată ca un număr întreg.
11. **xybar** (corelația dintre mediile pozițiilor x și y): Reprezintă corelația dintre media pozițiilor pe axa x și media pozițiilor pe axa y a pixelilor activați în cadru. Este, de asemenea, exprimată ca un număr întreg.
12. **x2ybr** (media produsului x \* x \* y): Acesta indică media produsului dintre pozițiile pe axa x, pătratul acestora și pozițiile pe axa y a pixelilor activați în cadru. Este exprimată ca un număr întreg.
13. **xy2br** (media produsului x \* y \* y): Similar cu x2ybr, dar indică media produsului dintre pozițiile pe axa x, pozițiile pe axa y, pătratul acestora a pixelilor activați în cadru. Este, de asemenea, exprimată ca un număr întreg.
14. **x-ege** (numărul mediu de margini de la stânga la dreapta): Reprezintă numărul mediu de margini din imaginea cadrului care sunt orientate de la stânga la dreapta. Este exprimată ca un număr întreg.
15. **xegvy** (corelația dintre x-ege și y): Acesta indică corelația dintre numărul de margini orientate de la stânga la dreapta și coordonata y a pixelilor activați în cadru. Este exprimată ca un număr întreg.
16. **y-ege** (numărul mediu de margini de jos în sus): Similar cu x-ege, dar indică numărul mediu de margini din imaginea cadru care sunt orientate de jos în sus. Este, de asemenea, exprimată ca un număr întreg.
17. **yegvx** (corelația dintre y-ege și x): Acesta indică corelația dintre numărul de margini orientate de jos în sus și coordonata x a pixelilor activați în cutie. Este, de asemenea, exprimată ca un număr întreg.

Setul de date conține 16 caracteristici numerice extrase din imagini ale literelor. ACP poate comprima aceste caracteristici într-un număr mai mic de componente principale, păstrând în același timp cât mai multă variație posibilă. Astfel, se reduce complexitatea datelor și se poate facilita procesul de analiză și interpretare.

Setul de date analizat subdimensionat:

'lettr', 'x-box', 'y-box', 'width', 'high', 'onpix', 'x-bar', 'y-bar', 'x2bar' , 'y2bar', 'xybar', 'x2ybr', 'xy2br', 'x-ege', 'xegvy', 'y-ege', 'yegvx'  
T,2,8,3,5,1,8,13,0,6,6,10,8,0,8,0,8  
I,5,12,3,7,2,10,5,5,4,13,3,9,2,8,4,10  
D,4,11,6,8,6,10,6,2,6,10,3,7,3,7,3,9  
N,7,11,6,6,3,5,9,4,6,4,4,10,6,10,2,8  
G,2,1,3,1,1,8,6,6,6,6,5,9,1,7,5,10  
S,4,11,5,8,3,8,8,6,9,5,6,6,0,8,9,7  
B,4,2,5,4,4,8,7,6,6,7,6,6,2,8,7,10  
A,1,1,3,2,1,8,2,2,2,8,2,8,1,6,2,7  
J,2,2,4,4,2,10,6,2,6,12,4,8,1,6,1,7  
M,11,15,13,9,7,13,2,6,2,12,1,9,8,1,1,8  
X,3,9,5,7,4,8,7,3,8,5,6,8,2,8,6,7  
O,6,13,4,7,4,6,7,6,3,10,7,9,5,9,5,8  
G,4,9,6,7,6,7,8,6,2,6,5,11,4,8,7,8  
M,6,9,8,6,9,7,8,6,5,7,5,8,8,9,8,6  
R,5,9,5,7,6,6,11,7,3,7,3,9,2,7,5,11  
F,6,9,5,4,3,10,6,3,5,10,5,7,3,9,6,9  
O,3,4,4,3,2,8,7,7,5,7,6,8,2,8,3,8  
C,7,10,5,5,2,6,8,6,8,11,7,11,2,8,5,9  
T,6,11,6,8,5,6,11,5,6,11,9,4,3,12,2,4  
J,2,2,3,3,1,10,6,3,6,12,4,9,0,7,1,7  
J,1,3,2,2,1,8,8,2,5,14,5,8,0,7,0,7  
H,4,5,5,4,4,7,7,6,6,7,6,8,3,8,3,8  
S,3,2,3,3,2,8,8,7,5,7,5,7,2,8,9,8  
O,6,11,7,8,5,7,6,9,6,7,5,9,4,8,5,5  
J,3,6,4,4,2,6,6,4,4,14,8,12,1,6,1,6  
C,6,11,7,8,3,7,8,7,11,4,7,14,1,7,4,8  
M,7,11,11,8,9,3,8,4,5,10,11,10,10,9,5,7  
W,12,14,12,8,5,9,10,4,3,5,10,7,10,12,2,6  
H,6,9,8,7,6,8,6,6,7,7,7,9,6,8,4,8  
G,3,6,4,4,2,6,6,5,5,6,6,9,2,8,4,8  
L,2,3,3,4,1,0,1,5,6,0,0,6,0,8,0,8  
L,1,3,3,1,1,6,4,1,7,8,3,10,0,7,2,9  
X,8,12,8,6,4,3,10,4,7,12,11,9,3,7,3,4  
B,5,9,7,7,10,9,8,4,4,6,8,6,6,11,8,7  
M,6,9,9,7,6,5,6,3,5,10,9,9,8,5,2,7  
G,4,7,6,5,3,6,6,6,8,6,5,9,3,10,4,8  
O,4,7,5,5,3,7,7,8,6,7,6,8,3,8,3,8  
P,3,6,4,4,2,4,14,8,1,11,6,3,0,10,4,8  
G,4,9,5,6,6,8,5,4,3,7,5,11,6,8,5,11  
E,3,4,3,6,2,3,8,6,10,7,6,15,0,8,7,8  
X,5,11,8,8,4,8,8,1,8,10,5,7,3,8,4,8  
E,3,7,4,5,4,7,7,5,8,8,8,9,3,9,6,9  
X,4,6,7,4,3,9,7,2,8,11,3,7,3,8,4,9  
G,4,5,5,8,3,7,6,8,8,6,6,10,1,8,6,11  
V,7,9,6,5,2,8,10,4,5,8,10,5,4,12,3,8  
X,4,10,5,7,5,8,7,3,8,5,6,6,3,8,6,8  
W,5,9,6,7,8,7,9,5,3,7,9,8,6,8,3,8  
G,1,3,2,1,1,7,7,5,6,7,5,10,1,9,3,9  
R,8,10,8,6,6,7,7,3,5,8,4,8,6,6,7,7  
S,3,9,4,7,2,7,5,6,10,5,6,10,0,9,9,8

În urma aplicării algoritmului de analiză în componente principale obținem următoarele:

Numar componente principale semnificative conform crit. **Kaiser**: **5** - Acest criteriu înseamnă că 5 componente principale au o variație suficient de mare pentru a justifica includerea lor în analiza ulterioară.

Numar componente principale semnificative conform crit. **Procente acoperire**: **7** - Criteriul procentelor de acoperire identifică 7 componente principale semnificative. Acest lucru sugerează că primele 7 componente principale explică cel puțin 80% din variația totală a datelor.

Numar componente principale semnificative conform crit. **Cattell**: **2 -** În conformitate cu acest criteriu, sunt identificate 2 componente principale semnificative. Aceasta înseamnă că 2 componente principale aduc contribuții semnificative la explicarea variației din setul de date, iar diferența de varianță între ele este semnificativă.

Valoarea valorilor proprii rezultate: **alpha [4.99045235 2.35832044 2.03260843 1.40621249 1.08792592 0.9585624 0.77221394 0.70303975 0.52724826 0.39694788 0.37186836 0.29059642 0.22968228 0.09670236 0.05304861 0.05110071]** Valorile proprii mai mari indică că componenta principală corespunzătoare are o contribuție semnificativă la variația totală a datelor. Astfel, componentele principale cu valorile proprii cele mai mari sunt cele mai importante în explicarea structurii și variației setului de date.Valorile proprii mici indică că componenta principală corespunzătoare explică o cantitate redusă de variație în setul de date. În general, componentele principale cu valori proprii mici pot fi considerate mai puțin relevante sau mai puțin semnificative în explicarea structurii datelor.

A graph with blue dots

Description automatically generated

A red and yellow squares with white lines

Description automatically generated

A graph with blue dots

Description automatically generated

Din scatter plot al corelațiilor dintre componenta C1 și componenta C2 se observă că punctele sunt dispuse pe o linie diagonală ascendentă de la stânga jos la dreapta sus, aceasta indică o corelație pozitivă între componenta C1 și C2. Cu alte cuvinte, o creștere a valorii unei componente este asociată cu o creștere a celeilalte componente, iar o descreștere a uneia este asociată cu o descreștere a celeilalte.

A graph with blue dots

Description automatically generated

Scatter plot al corelațiilor dintre componenta C5 și C6 - faptul că punctele sunt distribuite orizontal sugerează că nu există o relație liniară sau o asociere între valorile acestor două componente. Cu alte cuvinte, variația într-o componentă nu este asociată cu variația în cealaltă componentă. Într-o interpretare mai detaliată, absența unei relații între componentele este indicată de faptul că schimbarea într-o componentă nu influențează în mod constant sau semnificativ schimbarea în cealaltă componentă. Acest lucru poate fi înțeles ca fiecare componentă reprezentând un aspect distinct sau independent al datelor, care nu este asociat în mod direct cu celălalt aspect.

În concluzie, analiza în componente principale și reprezentările grafice ale corelațiilor ne oferă o înțelegere mai profundă a structurii și relațiilor din setul de date, identificând componente semnificative și evidențiind modul în care acestea contribuie la variația datelor. Aceste informații pot fi valoroase în analizele ulterioare și în luarea deciziilor bazate pe date.

POP Laura-Nicoleta

Grupa 1094E

An III