

TEMA 3

ALGORITMI DE CLUSTERIZARE

***Primele 6 cerinte au fost preluate integral din Tema 2.**

Analiza a fost realizata pe un set de date ce contine informatii despre 200 de companii listate la bursa, obtinute din sursa Yahoo Finance, sectiunea Financials (Income Statement & Cash Flow). Datele se refera la TTM (Trailing Twelve Months), oferind o privire de ansamblu asupra performantei financiare a acestor companii pe parcursul ultimelor 12 luni disponibile.

In ceea ce priveste structura setului de date, acesta nu prezinta valori lipsa, iar variabilele au fost notate de la X1 la X10, cu denumiri sugestive in fisierul Excel. In ceea ce priveste outlierii/valorile extreme, acestia au fost eliminati din setul de date, deoarece prezenta acestora afecta negativ analiza, distorsionand structura datelor si reducand relevanta rezultatelor obtinute. Dupa eliminarea valorilor extreme, in setul de date am ramas cu 81 de observatii.

Setul de date este compus din urmatoarele variabile, notate de la X1 la X10 si exprimate in mii USD (\$):

- **Total Revenue/Venit total** – ce reprezinta suma totală a veniturilor generate de o companie din vânzarea bunurilor sau serviciilor sale, fără a ține cont de costurile asociate.

- **Gross Profit/Profitul brut** – este calculat prin scăderea costurilor directe asociate producției bunurilor sau serviciilor din venitul total si reflectă eficiența în generarea profitului din activitatea principală.

- **Operating Income/Venitul operational** – este venitul generat din activitățile de bază ale companiei, excluzând veniturile și cheltuielile non-operaționale; acesta indica profitabilitatea operațiunilor zilnice ale companiei.

- **Net Income/ Venitul net** - reprezintă profitul total obținut de companie după scăderea tuturor cheltuielilor, inclusiv taxe și cheltuieli non-operaționale; este un indicator esențial al sănătății financiare a unei companii.

- **Earnings before interest and taxes (EBIT)** - acest indicator arată profitul companiei înainte de deducerea cheltuielilor cu dobânzile și impozitele, fiind util pentru compararea performanței între companii, indiferent de structura lor de capital.

- **Earnings per share (EPS)** - reprezintă venitul pe acțiune și oferă o măsură a profitabilității unei companii pe acțiune.

- **Operating Cash Flow/Fluxul de numerar operational** - acesta măsoară capacitatea companiei de a genera numerar din activitățile sale operaționale.

- **Investing Cash Flow/Fluxul de numerar din activitati de investitii** - reflectă numerarul cheltuit sau generat din activitățile de investiții ale companiei, inclusiv achiziții de active sau vânzări de active.

- **Financing Cash Flow/Fluxul de numerar din activitati de finantare** - acesta arată fluxurile de numerar rezultate din activitățile de finanțare, cum ar fi emisiunea de acțiuni, împrumuturile și rambursările de datorii.

- **Free Cash Flow/Fluxul de numerar liber** - este un indicator important al capacității unei companii de a genera numerar după ce a acoperit toate cheltuielile necesare.

În ceea ce privește **observatiile**, fiecare linie din setul de date corespunde unei companii listate la bursă, oferind o imagine de ansamblu asupra performanței financiare a acestora prin intermediul variabilelor de mai sus. Aceasta permite compararea companiilor în funcție de diferiți indicatori financiari.

Obiectivul general al analizei este de a explora și interpreta relațiile dintre variabilele financiare ale unui set de 200 de companii listate la bursă, cu scopul de a evidenția factorii esențiali care influențează performanța financiară. Analiza datelor permite reducerea dimensiunii setului de informații, sintetizarea indicatorilor-cheie și identificarea tiparelor relevante, contribuind astfel la o înțelegere aprofundată a dinamicii financiare și la sprijinirea deciziilor informate în domeniul investițional și managerial.

Interpretarea indicatorilor statistici

```
> summary(tema)
Companie
Length:81
Class :character
Mode :character

      X1      X2      X3      X4      X5      X6
Min.   : 4131  Min.   : 3966  Min.   :-636864  Min.   :-762367  Min.   :-727188  Min.   :-4.7200
1st Qu.:1275994 1st Qu.: 705928 1st Qu.: -14466 1st Qu.: -83497 1st Qu.: -24339 1st Qu.: -0.4200
Median : 2562440 Median :1107379 Median : 281000 Median : 116261 Median : 213000 Median : 0.5900
Mean   : 4474807 Mean   :1647343 Mean   : 293014 Mean   : 89902 Mean   : 239688 Mean   : 0.9617
3rd Qu.: 5320059 3rd Qu.:2123393 3rd Qu.: 506000 3rd Qu.: 280000 3rd Qu.: 439514 3rd Qu.: 2.4400
Max.   :23813905 Max.   :9646000 Max.   :1550863 Max.   : 837880 Max.   :1362945 Max.   : 8.8600

      X7      X8      X9      X10
Min.   :-586000 Min.   :-1209300 Min.   :-919000 Min.   :-614000
1st Qu.: 131885 1st Qu.: -383000 1st Qu.: -335144 1st Qu.: 23889
Median : 384670 Median : -147000 Median : -157094 Median : 181000
Mean   : 449780 Mean   : -242236 Mean   : -194710 Mean   : 233963
3rd Qu.: 712000 3rd Qu.: -51200 3rd Qu.: -12411 3rd Qu.: 383000
Max.   :1796100 Max.   : 314000 Max.   : 414345 Max.   : 823000
```

Figura 1. Rezultatul comenzii *summary*

```
> describe(tema[,-1])
vars  n    mean    sd median trimmed    mad    min    max    range skew kurtosis    se
X1    1  81  4474806.54 5230485.32 2562440.00 3345078.97 2334521.23 4131.00 23813905.00 23809774.00 2.07 3.92 581165.04
X2    2  81  1647342.80 1646814.18 1107379.00 1354668.31 975246.87 3966.00 9646000.00 9642034.00 2.25 6.56 182979.35
X3    3  81  293013.65 398911.76 281000.00 272393.32 391702.92 -636864.00 1550863.00 2187727.00 0.52 0.38 44323.53
X4    4  81  89902.35 313272.53 116261.00 100658.46 257585.44 -762367.00 837880.00 1600247.00 -0.32 0.27 34808.06
X5    5  81  239687.85 426923.00 213000.00 228556.34 351878.80 -727188.00 1362945.00 2090133.00 0.33 0.18 47435.89
X6    6  81  0.96 2.41 0.59 0.95 1.76 -4.72 8.86 13.58 0.41 1.27 0.27
X7    7  81  449780.25 447403.06 384670.00 410150.88 382605.69 -586000.00 1796100.00 2382100.00 0.84 0.78 49711.45
X8    8  81 -242235.63 282768.47 -147000.00 -199593.78 172682.87 -1209300.00 314000.00 1523300.00 -1.28 1.43 31418.72
X9    9  81 -194709.60 284965.00 -157094.00 -180020.72 236108.50 -919000.00 414345.00 1333345.00 -0.49 0.16 31662.78
X10  10  81 233962.54 298504.74 181000.00 228332.17 249804.76 -614000.00 823000.00 1437000.00 0.13 -0.13 33167.19
```

Figura 2. Rezultatul comenzii *describe*

Variabila X1 - Venitul total prezintă o distribuție între 4.131 USD și 23.813.905 USD, indicând variații semnificative între companiile analizate, de la firme mici la lideri ai pieței.

Majoritatea companiilor au venituri situate între 1.275.994 USD (1st Qu) și 5.320.059 USD (3rd Qu), dar media de 4.474.807 USD este semnificativ influențată de câteva companii foarte mari. Mediana de 2.562.440 USD, mai mică decât media, indică o distribuție asimetrică spre dreapta, aspect confirmat de coeficientul de asimetrie (skew) de 2,07. Kurtosis-ul de 3,92 sugerează o distribuție leptocurtică, cu o concentrație mare a companiilor în jurul valorilor centrale și câteva valori extreme notabile. Devierea standard de 5.230.485 USD reflectă variații mari între veniturile companiilor analizate, sugerând o piață eterogenă dominată de câțiva actori majori.

Variabila X2 - Profitul brut prezintă o distribuție cu valori între 3.966 USD și 9.646.000 USD, indicând variații semnificative între companiile analizate, de la profituri mici la valori foarte mari. Majoritatea companiilor au profituri între 705.928 USD și 2.123.393 USD. Media de 1.647.343 USD este influențată de câteva companii extrem de profitabile, în timp ce mediana de 1.107.379 USD, mai mică decât media, reflectă o distribuție asimetrică spre dreapta, confirmată de coeficientul de asimetrie de 2,25. Kurtosis-ul de 6,56 sugerează o distribuție leptocurtică, caracterizată prin concentrarea valorilor în jurul medianei și prezența unor valori extreme notabile. Devierea standard de 1.646.814 USD evidențiază o dispersie semnificativă între companii, confirmând că piața este dominată de câțiva actori majori cu profituri semnificative, în timp ce restul companiilor obțin profituri considerabil mai mici.

Variabila X3 - Venitul operațional prezintă o distribuție semnificativ variabilă, cu valori între -636.864 USD și 1.550.863 USD, indicând o gamă largă de performanțe între companiile analizate, de la pierderi semnificative până la profituri operaționale mari. Majoritatea companiilor au venituri operaționale între -14.466 USD și 506.000 USD. Media de 293.014 USD este influențată de câteva companii cu venituri operaționale semnificativ mai mari, în timp ce mediana de 281.000 USD este mai mică decât media, ceea ce sugerează o distribuție asimetrică moderată spre dreapta. Coeficientul de asimetrie (skew) de 0.52 indică o ușoară asimetrie, iar kurtosis-ul de 0.38 sugerează o distribuție plată, cu o concentrație mai scăzută în jurul medianei și mai puține valori extreme. Aceasta sugerează o piață unde majoritatea companiilor au venituri operaționale mai mici, cu câțiva jucători mari care generează venituri semnificativ mai mari.

În ceea ce privește **variabila X4**, reprezentând **venitul net**, valoarea medie în perioada analizată este de 89.902 USD, cu un minim de -762.367 USD și un maxim de 837.880 USD, ceea ce reflectă o variabilitate moderată între companiile incluse în analiză, de la pierderi până la profituri semnificative. Conform medianei, jumătate din observații au venituri nete sub 116.261 USD, în timp ce cealaltă jumătate le depășește. Diferența dintre medie și mediană sugerează o ușoară asimetrie spre stânga, indicată și de coeficientul de asimetrie de -0.32, ceea ce înseamnă că există câteva valori mai mici care influențează distribuția. Kurtosis-ul de 0.27 arată o distribuție relativ plată, cu o concentrație scăzută în jurul medianei și câteva valori extreme, dar nu foarte dominante. Abaterea standard de 313.272,53 USD indică o dispersie semnificativă a valorilor, subliniind diferențele dintre companiile analizate. Această analiză evidențiază faptul că, deși majoritatea companiilor au venituri nete moderate, există variații relevante datorate atât pierderilor, cât și unor performanțe mai ridicate.

Analizând **variabila X5 - EBIT (Earnings Before Interest and Taxes)**, observăm o distribuție moderat variabilă, cu valori cuprinse între -727.188 USD și 1.362.945 USD, ceea ce

indică diferențe semnificative în rentabilitatea operațională a companiilor analizate. Media EBIT este de 239.688 USD, iar mediana de 213.000 USD sugerează că jumătate dintre companii au valori sub această medie, iar cealaltă jumătate le depășește. Coeficientul de asimetrie de 0.33 indică o distribuție ușor asimetrică spre dreapta, ceea ce arată că există câteva companii cu valori mari care influențează media. Kurtosis-ul de 0.18 sugerează o distribuție ușor plată, cu o concentrare scăzută a datelor în jurul medianei. Abaterea standard de 426.923 USD reflectă o dispersie semnificativă între valorile EBIT ale companiilor. Analiza sugerează că, în timp ce majoritatea companiilor au rentabilități operaționale mai mici, există câteva cu performanțe mai ridicate, ceea ce evidențiază posibile strategii mai eficiente sau avantaje competitive.

Variabila X6, reprezentând venitul pe acțiune (EPS), prezintă o distribuție variabilă, cu valori cuprinse între -47.200 USD și 88.600 USD, majoritatea valorilor fiind concentrate în jurul medianei de 0.5900 USD. Coeficientul de asimetrie de 0.41 indică o distribuție ușor asimetrică spre dreapta, în care majoritatea valorilor sunt concentrate în partea stângă, dar există câteva valori mai mari care influențează media. Kurtosis-ul de 1.27 sugerează o distribuție moderat leptocurtică, ceea ce înseamnă că există o concentrație relativ mare de valori în jurul medianei, dar și câteva valori extreme ce se depărtează de medie. Din punct de vedere economic, acest lucru sugerează că majoritatea companiilor au un EPS mic, iar câteva reușesc să obțină profituri semnificative pe acțiune, influențând astfel media. Variabilitatea moderată sugerează o diversificare a performanțelor financiare pe piață, cu companii care pot avea rezultate modeste sau semnificativ mai mari.

Variabila X7, fluxul de numerar operațional, variază semnificativ, cu valori cuprinse între -586,000 USD și 1,796,100 USD, reflectând diferențe mari între companiile analizate, de la fluxuri de numerar negative până la valori semnificativ pozitive. Majoritatea companiilor au fluxuri de numerar între 131,885 USD și 712,000 USD, iar media de 449,780 USD este influențată de câteva companii cu performanțe financiare foarte bune. Mediana de 384,670 USD sugerează o ușoară asimetrie spre dreapta, iar coeficientul de asimetrie de 0.84 confirmă această tendință, indicând că există companii cu fluxuri de numerar extrem de mari care trag media în sus. Kurtosis-ul de 0.78 sugerează o distribuție platycurtică, ceea ce înseamnă că valorile sunt distribuite mai uniform, cu mai puține concentrații în jurul mediei. Aceste date sugerează o piață diversificată, cu performanțe financiare influențate de diferite strategii și condiții de piață ale companiilor.

Variabila X8, fluxul de numerar din activitățile de investiții, prezintă o distribuție variabilă, cu valori cuprinse între -1,209,300 USD și 314,000 USD, ceea ce reflectă diferențe semnificative între companiile analizate, de la cheltuieli mari până la fluxuri pozitive asociate unor câștiguri din investiții. 50% dintre observații au fluxuri de numerar între -383,000 USD și -51,200 USD, iar media de -242,236 USD este influențată de câteva companii care au înregistrat fluxuri pozitive, însă majoritatea companiilor au fluxuri negative. Mediana de -147,000 USD, mai mare decât media, confirmă o distribuție asimetrică spre stânga. Coeficientul de asimetrie de -1.28 indică o coadă lungă spre stânga, ceea ce arată că multe companii se confruntă cu cheltuieli semnificative pentru investiții, dar câteva înregistrează fluxuri pozitive care atenuază media. Kurtosis-ul de 1.43 sugerează o distribuție leptocurtică moderată, cu o concentrație mare de

valori în jurul medianei, dar și câteva valori extreme. Analiza arată că majoritatea companiilor se află într-o etapă de investiții semnificative, ceea ce ar putea reflecta strategii de creștere pe termen lung.

Variabila X9, fluxul de numerar din activitățile de finanțare, prezintă o distribuție variabilă, cu valori între -919,000 USD și 414,345 USD. Media de -194,710 USD și mediana de -157,094 USD sugerează că majoritatea fluxurilor de numerar din activitățile de finanțare sunt negative, ceea ce indică o tendință generală de finanțare negativă. Coeficientul de asimetrie de -0.49 indică o ușoară asimetrie spre stânga, cu o concentrare a valorilor în partea dreaptă și o coadă ușor extinsă în stânga. Kurtosis-ul de 0.16 sugerează o distribuție plată, cu o concentrație moderată în jurul medianei și cu câteva valori extreme. Abaterea standard de 284,965 USD subliniază variabilitatea semnificativă între fluxurile de numerar din activitățile de finanțare ale companiilor analizate. Aceste rezultate sugerează că majoritatea companiilor sunt implicate în activități financiare cu fluxuri negative, dar există și câteva companii care beneficiază de fluxuri financiare pozitive din activitățile de finanțare.

Variabila X10, fluxul de numerar liber, variază între -614,000 USD și 823,000 USD, cu o medie de 233,963 USD și o mediana de 181,000 USD. Coeficientul de asimetrie de 0.13 sugerează o distribuție ușor asimetrică spre dreapta, indicând că majoritatea companiilor au fluxuri de numerar mai mici, cu câteva excepții generând valori semnificativ mai mari. Kurtosis-ul de -0.13 arată o distribuție plată, cu o concentrație relativ uniformă a valorilor și puține valori extreme. Abaterea standard de 298,504.74 USD indică o variabilitate semnificativă între fluxurile de numerar liber ale companiilor. Din punct de vedere economic, fluxul de numerar liber sugerează că majoritatea companiilor au un flux de numerar relativ modest, dar există un număr mic de firme cu fluxuri de numerar mai mari, care pot reflecta o performanță financiară puternică și oportunități pentru investiții sau recompensarea acționarilor.

Asadar, concluzia analizei este că există o diversitate semnificativă între companiile analizate, cu o dominanță clară a unor actori mari pe piață. Variabilele financiare examinează diferențele considerabile în veniturile, profiturile și fluxurile de numerar ale companiilor, ceea ce reflectă o piață eterogenă. În ansamblu, analiza sugerează o piață financiară diversificată, unde companiile mari au un impact semnificativ asupra indicatorilor financiari, iar majoritatea companiilor mici și mijlocii au rezultate mai modeste, dar cu un potențial de creștere pe termen lung, în special pentru cele implicate în activități de investiții.

Matricea de corelație și matricea de covarianță

Pentru a observa mai bine rezultatele și a fi mai ușor de interpretat, am standardizat datele utilizând funcția scale. Observăm că după standardizarea datelor, matricea de covarianță este egală cu matricea de corelație, toate variabilele având aceeași deviație standard (1) și media 0, ceea ce le face comparabile direct între ele. Asadar, valorile observate reflectă doar relațiile dintre variabile, fără influența unității de măsură, astfel încât covarianțele și corelațiile devin mai ușor de interpretat.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.0000000	0.9045192	0.5718650	0.2988523	0.4717430	0.12605371	0.6700836	-0.53350678	-0.4915897	0.4187086
X2	0.9045192	1.0000000	0.5752673	0.2847471	0.4123124	0.11844546	0.6509549	-0.50014086	-0.4603218	0.4554152
X3	0.5718650	0.5752673	1.0000000	0.6695824	0.8032117	0.39782999	0.7554865	-0.66274626	-0.5823679	0.5261587
X4	0.2988523	0.2847471	0.6695824	1.0000000	0.9158698	0.74327828	0.5107917	-0.28602568	-0.5455754	0.5513663
X5	0.4717430	0.4123124	0.8032117	0.9158698	1.0000000	0.63149859	0.6425682	-0.45358593	-0.6531158	0.5789148
X6	0.1260537	0.1184455	0.3978300	0.7432783	0.6314986	1.00000000	0.2245358	-0.02663007	-0.3068137	0.3263840
X7	0.6700836	0.6509549	0.7554865	0.5107917	0.6425682	0.22453583	1.0000000	-0.71682955	-0.6125110	0.8096187
X8	-0.5335068	-0.5001409	-0.6627463	-0.2860257	-0.4535859	-0.02663007	-0.7168295	1.00000000	0.2874937	-0.3145959
X9	-0.4915897	-0.4603218	-0.5823679	-0.5455754	-0.6531158	-0.30681372	-0.6125110	0.28749371	1.0000000	-0.5262104
X10	0.4187086	0.4554152	0.5261587	0.5513663	0.5789148	0.32638399	0.8096187	-0.31459593	-0.5262104	1.0000000

Figura 3. Matricea de corelatie

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.0000000	0.9045192	0.5718650	0.2988523	0.4717430	0.12605371	0.6700836	-0.53350678	-0.4915897	0.4187086
X2	0.9045192	1.0000000	0.5752673	0.2847471	0.4123124	0.11844546	0.6509549	-0.50014086	-0.4603218	0.4554152
X3	0.5718650	0.5752673	1.0000000	0.6695824	0.8032117	0.39782999	0.7554865	-0.66274626	-0.5823679	0.5261587
X4	0.2988523	0.2847471	0.6695824	1.0000000	0.9158698	0.74327828	0.5107917	-0.28602568	-0.5455754	0.5513663
X5	0.4717430	0.4123124	0.8032117	0.9158698	1.0000000	0.63149859	0.6425682	-0.45358593	-0.6531158	0.5789148
X6	0.1260537	0.1184455	0.3978300	0.7432783	0.6314986	1.00000000	0.2245358	-0.02663007	-0.3068137	0.3263840
X7	0.6700836	0.6509549	0.7554865	0.5107917	0.6425682	0.22453583	1.0000000	-0.71682955	-0.6125110	0.8096187
X8	-0.5335068	-0.5001409	-0.6627463	-0.2860257	-0.4535859	-0.02663007	-0.7168295	1.00000000	0.2874937	-0.3145959
X9	-0.4915897	-0.4603218	-0.5823679	-0.5455754	-0.6531158	-0.30681372	-0.6125110	0.28749371	1.0000000	-0.5262104
X10	0.4187086	0.4554152	0.5261587	0.5513663	0.5789148	0.32638399	0.8096187	-0.31459593	-0.5262104	1.0000000

Figura 4. Matricea de covarianta

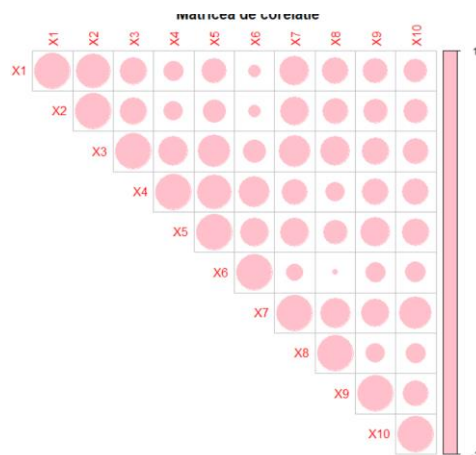


Figura 5. Reprezentarea grafică a matricei de corelatie

Analizând matricea de corelație, putem observa un set complex de relații între variabilele financiare, care oferă o perspectivă asupra interdependențelor dintre indicatori.

Variabila X1 (Venitul total generat) este puternic corelată cu X2 (Profitul brut) și X7 (Fluxul de numerar operațional), indicând o legătură strânsă între creșterea veniturilor și performanța financiară generală. Corelațiile moderat pozitive cu X3 (Venitul operațional) și X5 (EBIT) sugerează că veniturile susțin profitabilitatea operațională și profitul înainte de dobânzi și taxe.

Variabila X2 (Profitul brut) are corelații puternice cu X1 (Venitul total generat), X7 (Fluxul de numerar operațional) și X3 (Venitul operațional), indicând că o creștere a veniturilor și a fluxului de numerar operațional contribuie semnificativ la profitabilitatea brută și operațională a companiei. Corelațiile mai slabe cu X6 (Venitul pe acțiune) și X4 (Venitul net) sugerează că profitul brut nu influențează direct performanța pe acțiune sau profitabilitatea netă, aceste variabile fiind mai sensibile la alte factori economici sau strategii financiare.

Variabila X3 (Venitul operațional) prezintă corelații puternice cu X5 (EBIT) și X7 (Fluxul de numerar operațional), indicând o relație strânsă între performanța operațională și fluxul de numerar generat de activitățile de bază ale companiei. Corelațiile moderat pozitive cu celelalte variabile sugerează că venitul operațional este influențat într-o măsură semnificativă de veniturile totale, profitul brut și fluxul de numerar liber, dar și de variabilele financiare de bază.

Variabila X4 (Venitul net) este puternic corelată cu X5 (EBIT) și X6 (Venitul pe acțiune), indicând o influență semnificativă a profitabilității operaționale și performanței pe acțiune. Corelațiile moderate cu celelalte variabile sugerează o legătură mai slabă, dar încă semnificativă, între venitul net și veniturile totale, profitul brut și fluxurile de numerar. Corelațiile mai slabe cu X1 (Venitul total) și X2 (Profitul brut) sugerează că venitul net este mai puțin influențat direct de acestea din cauza cheltuielilor și taxelor.

Variabila X5 (EBIT) prezintă corelații puternice cu majoritatea variabilelor financiare, ceea ce subliniază faptul că profitul înainte de dobânzi și taxe are un impact semnificativ asupra altor indicatori cheie. Corelația cea mai puternică este cu **X4 (Venitul net)**, ceea ce sugerează că EBIT are un rol determinant în generarea veniturilor nete, indicând o relație strânsă între profitul operațional și profitul final al companiei. Aceasta evidențiază faptul că un EBIT mai mare conduce, în general, la un venit net mai mare, având în vedere impactul relativ mai mic al altor cheltuieli și taxe.

Variabila X6 (EPS - Venitul pe acțiune) are corelații slab și moderat pozitive cu majoritatea variabilelor financiare, ceea ce sugerează că venitul pe acțiune este influențat într-o măsură relativ mică de celelalte indicatori financiari. Cele mai puternice corelații sunt cu X4 (Venitul net) și X5 (EBIT), indicând că venitul pe acțiune este strâns legat de profitabilitatea generală a companiei.

Variabila X7 (Fluxul de numerar operațional) este puternic corelată pozitiv cu majoritatea variabilelor financiare, indicând o influență directă din veniturile totale, profitul brut, venitul operațional și EBIT. Corelația slab pozitivă cu X6 (Venitul pe acțiune) sugerează un impact indirect asupra veniturilor pe acțiune. Corelațiile negative cu X8 (Fluxul de numerar din activități de investiții) și X9 (Fluxul de numerar din activități de finanțare) sugerează că investițiile și finanțarea prin datorii afectează negativ fluxul de numerar operațional, indicând o utilizare mai puțin eficientă a numerarului.

Variabilele X8 (Investing Cash Flow - Fluxul de numerar din activități de investiții) și X9 (Financing Cash Flow - Fluxul de numerar din activități de finanțare) au corelații negative cu toate celelalte variabile financiare, ceea ce sugerează că, pe măsură ce fluxurile de numerar din investiții și finanțare cresc, performanțele financiare ale companiei (precum veniturile, profiturile

și fluxurile de numerar operaționale) tind să scadă. Totuși, X8 și X9 prezintă o corelație pozitivă moderată între ele, ceea ce sugerează că, de obicei, atunci când o companie alocă mai multe fonduri pentru activități de investiții, ea recurge și la surse de finanțare suplimentare (precum împrumuturi sau emisiuni de acțiuni), reflectând un comportament strategic comun de finanțare a expansiunii prin investiții.

Variabila X10 (Fluxul de numerar liber) are cea mai puternică corelație pozitivă cu X7 (Fluxul de numerar operațional), indicând că un flux operațional mai mare generează mai multe fonduri pentru cheltuielile de capital și pentru investiții sau plata datoriilor. De asemenea, X10 prezintă corelații moderat pozitive cu celelalte variabile financiare, sugerând o legătură între fluxul de numerar liber și performanțele financiare ale companiei, influențate de mai mulți factori economici.

Matricea de covarianță, având aceleași valori ca matricea de corelație după standardizare, oferă informații esențiale despre relațiile dintre variabilele economice analizate. Majoritatea covarianțelor sunt pozitive, ceea ce sugerează o relație directă între variabilele financiare. Astfel, pe măsură ce o variabilă crește, celelalte au tendința să crească și ele. Cu toate acestea, covarianțele negative între X8 (fluxul de numerar din activități de investiții) și X9 (fluxul de numerar din activități de finanțare) cu celelalte variabile financiare sugerează că o creștere a acestor fluxuri are un efect negativ asupra altor indicatori, cum ar fi venitul total și profitul brut. Aceasta poate semnala faptul că investițiile și finanțările externe sunt invers proporționale cu profitabilitatea pe termen scurt, resursele fiind alocate pe termen lung, ceea ce afectează negativ fluxurile de numerar operaționale și alte activități economice.

ANALIZA CLUSTER PRIN METODE IERARHICE

Ca un prim pas în analiza, am analizat matricea de corelație și am ales să elimin variabilele X8 și X9, deoarece au corelații negative cu toate celelalte variabile, acestea fiind mai puțin relevante pentru explicarea variabilității totale a datelor.

```
> tema3 = tema
> cor(tema3[-1])
```

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.0000000	0.9045192	0.5718650	0.2988523	0.4717430	0.12605371	0.6700836	-0.53350678	-0.4915897	0.4187086
X2	0.9045192	1.0000000	0.5752673	0.2847471	0.4123124	0.11844546	0.6509549	-0.50014086	-0.4603218	0.4554152
X3	0.5718650	0.5752673	1.0000000	0.6695824	0.8032117	0.39782999	0.7554865	-0.66274626	-0.5823679	0.5261587
X4	0.2988523	0.2847471	0.6695824	1.0000000	0.9158698	0.74327828	0.5107917	-0.28602568	-0.5455754	0.5513663
X5	0.4717430	0.4123124	0.8032117	0.9158698	1.0000000	0.63149859	0.6425682	-0.45358593	-0.6531158	0.5789148
X6	0.1260537	0.1184455	0.3978300	0.7432783	0.6314986	1.00000000	0.2245358	-0.02663007	-0.3068137	0.3263840
X7	0.6700836	0.6509549	0.7554865	0.5107917	0.6425682	0.22453583	1.0000000	-0.71682955	-0.6125110	0.8096187
X8	-0.5335068	-0.5001409	-0.6627463	-0.2860257	-0.4535859	-0.02663007	-0.7168295	1.00000000	0.2874937	-0.3145959
X9	-0.4915897	-0.4603218	-0.5823679	-0.5455754	-0.6531158	-0.30681372	-0.6125110	0.28749371	1.0000000	-0.5262104
X10	0.4187086	0.4554152	0.5261587	0.5513663	0.5789148	0.32638399	0.8096187	-0.31459593	-0.5262104	1.0000000

Figura 6. Matricea de corelație

Eliminarea acestora ajută la îmbunătățirea calității analizei, concentrându-se pe variabilele care au o corelație mai puternică cu celelalte, și, de asemenea, reduce dimensiunea datelor, facilitând interpretarea și identificarea principalilor factori care influențează variabilitatea setului de date.

Jordan Maria-Alexandra

Grupa 1080-A

Apoi, am standardizat noul set de date.

	X1	X2	X3	X4	X5	X6	X7	X10
RDDT	-0.667892429	-0.472650050	-2.33103597	-2.1192932313	-2.05318491	-1.841847057	-0.92638446	-0.70766227
SPOT	1.790693001	1.206971142	-1.04036456	-0.6381100433	-0.74179151	-0.649061628	0.85430741	1.97329346
DJT	-0.854734345	-0.997912589	-0.77455891	-0.4727236926	-0.60537112	-0.578408833	-1.02706773	-0.81639421
SOUN	-0.844918831	-0.977051219	-0.91891164	-0.6386271650	-0.75570736	-0.557628599	-1.17183875	-1.03482625
HIMS	-0.651501598	-0.467732676	-0.70450331	-0.2291913255	-0.53337219	-0.366450447	-0.72233356	-0.48394053
IONQ	-0.849558171	-0.989999857	-1.23203352	-0.8100625619	-1.02629012	-0.732182564	-1.21978210	-1.19973485
NGD	-0.690816688	-0.903163709	-0.47131640	-0.2228166842	-0.58719688	-0.391386727	-0.21452747	-0.45748869
PTON	-0.339243194	-0.267694321	-1.75129873	-2.0487029366	-1.59112500	-1.276624693	-1.15305479	-1.07121428
WULF	-0.832534512	-0.954947936	-0.78590478	-0.4505449204	-0.61844139	-0.474507663	-0.88796944	-1.11964567
HOOD	-0.427647992	-0.104652247	0.03004761	0.6387334898	0.15298344	-0.270861370	-2.31509422	-2.84070040
RIG	-0.267051040	0.416353710	-0.78968255	-1.4010240544	-0.52863831	-0.574252786	-0.77956607	-1.90939191
KGC	0.008468326	-0.237332669	1.45743096	1.2867315828	1.51599271	-0.233456949	3.00918761	1.51031924
GRAB	-0.361497343	-0.362119061	-1.18074649	-0.9668972531	-0.81908882	-0.570096739	0.23294376	0.78068259
BTG	-0.483880824	-0.530657808	0.94070516	-0.7356193919	-0.17883284	-0.449571382	1.42829991	-0.38827371
RELY	-0.648673753	-0.618868125	-0.99428670	-0.6184689977	-0.77458196	-0.628281394	-1.02163638	-0.85435675
CFLT	-0.690035116	-0.620326698	-1.76855318	-1.4668868359	-1.52760769	-0.890112342	-1.03702743	-0.90181328
PAYC	-0.515538306	-0.106385290	0.73089936	1.2201218478	0.85613366	3.041507920	0.14304272	0.24479161
TTMI	-0.412568894	-0.727318733	-0.38637281	-0.2324249318	-0.34302404	-0.333202072	-0.57600913	-0.70510284
CVNA	1.375243981	0.329519384	0.29577054	1.9921876349	2.55154245	0.913611965	0.81631036	1.68854086
NXT	-0.331630325	-0.431498473	0.95338966	1.0138445829	1.08100794	1.171286865	-0.28101785	0.27213456
MCW	-0.672282652	-0.592057571	-0.27388928	-0.0616471092	-0.13797535	-0.312421838	-0.54395525	-1.30953210
MBLY	-0.502784423	-0.466563145	-1.36374435	-1.0020104503	-1.03458435	-0.516068131	-0.40853598	-0.17742614
ETSY	-0.325480800	0.187777224	0.12348432	0.6199798504	0.14408956	0.493851239	0.63460619	1.52858026
LAUR	-0.56009015	-0.756130120	-0.13754256	-0.3367568021	-0.28660707	-0.119807028	-0.45687270	-0.19025340

Figura 7. Noul set de date standardizat

Distanța dintre forme/observatii prin metoda euclidiană:

În următoarea figură este prezentată matricea distanțelor, realizată prin metoda euclidiană:

```
> dist_euclidiană <- dist(as.matrix(date_3_std),method="euclidian")
> dist_euclidiană
```

	RDDT	SPOT	DJT	SOUN	HIMS	IONQ	NGD	PTON	WULF	HOOD
SPOT	5.12117007									
DJT	3.02638785	4.82735284								
SOUN	2.82350405	4.98936475	0.37496296							
HIMS	3.28583332	4.20666946	0.80006791	1.04973151						
IONQ	2.41591192	5.13046947	0.84147369	0.51068789	1.43674408					
NGD	3.47323062	4.27179699	1.00727270	1.30699747	0.71249440	1.68127884				
PTON	1.09891793	4.86006166	2.40347732	2.15792307	2.64982890	1.79511525	2.92380766			
WULF	3.08726118	4.92977003	0.35388627	0.40894357	0.88104747	0.82404677	1.03607568	2.42154277		
HOOD	5.21422172	6.60157599	3.04787802	2.99057482	3.16957747	3.18585739	3.51496585	4.35485917	2.89387702	
RIG	3.03964304	4.82701612	2.11407128	1.96062167	2.09481499	1.95756410	2.42371489	2.06919474	1.93833143	2.98523764
KGC	7.88672568	5.03333951	5.97705540	6.31299605	5.43870688	6.69873043	5.06399922	7.50158992	5.99928372	7.19747654
GRAB	3.07519623	3.00705824	2.28926617	2.46207546	1.87115112	2.60025672	1.81028651	2.81974827	2.43165219	4.97555184
	RIG	KGC	GRAB	BTG	RELY	CFLT	PAYC	TTMI	CVNA	NXT
SPOT										
DJT										
SOUN										
HIMS										
IONQ										
NGD										
PTON										
WULF										
HOOD										
RIG										
KGC	6.56677264									
GRAB	3.04998243	5.09873524								
	MCW	MBLY	ETSY	LAUR	IP	DNB	KTB	COMP	SFM	DJTW
SPOT										
DJT										
SOUN										
HIMS										
IONQ										
NGD										
PTON										
WULF										

Figura 8. Matricea distanțelor dintre observatii prin metoda euclidiană

Calcularea distantei dintre 2 observatii, utilizand formula de calcul a distantei euclidiene:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Figura 9. Formula distantei euclidiene

Aplicarea formulei:

```
> dist_euclidiana[1]
[1] 5.12117
> date_3_std[1:2,]
      X1      X2      X3      X4      X5      X6      X7      X10
RDDT -0.6678924 -0.472650 -2.331036 -2.119293 -2.0531849 -1.8418471 -0.9263845 -0.7076623
SPOT  1.7906930  1.206971 -1.040365 -0.638110 -0.7417915 -0.6490616  0.8543074  1.9732935
> dist_spot_rddt
[1] 5.238988
```

Figura 10. Rezultatul aplicarii formulei anterioare

Se observă că valoarea obținută prin aplicarea formulei distanței Euclidiene este mai mare decât cea calculată utilizând comanda *dist*.

O alta metoda de calcul a distantei dintre observatii este distanta Manhattan, exprimata prin formula:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

Figura 11. Formula distantei Manhattan

```
> dist_manhattan <- dist(as.matrix(date_3_std), method="manhattan")
> dist_manhattan
      RDDT      SPOT      DJT      SOUN      HIMS      IONQ      NGD      PTON      WULF
SPOT 13.87588765 10.15963907 0.87525263 2.74048418 3.90061810 4.12093049 2.18578306 5.83241429 7.28366364
DJT  6.83581828 10.08121900 1.96814579 2.74048418 3.90061810 4.12093049 2.18578306 5.83241429 7.28366364
SOUN 6.72853225 10.08121900 1.96814579 2.74048418 3.90061810 4.12093049 2.18578306 5.83241429 7.28366364
HIMS 6.96092474 9.38658384 1.96814579 2.74048418 3.90061810 4.12093049 2.18578306 5.83241429 7.28366364
IONQ 6.02927817 10.81558101 1.95865013 1.16013392 3.90061810 4.12093049 2.18578306 5.83241429 7.28366364
NGD  8.08811296 9.48787261 2.18845817 2.96079656 1.32732658 4.12093049 2.18578306 5.83241429 7.28366364
PTON 2.80143711 12.25489523 5.86320537 5.06708113 6.36452584 4.29505099 7.53433047 5.49822304 10.40162696
WULF 7.11330082 10.36029156 0.65801032 0.94465235 1.96547067 1.93514743 2.18578306 5.83241429 7.28366364
HOOD 13.02625447 15.13359078 7.61464511 7.66057737 6.92087501 8.39500740 7.76907849 10.40162696 7.28366364
RIG  7.69015666 9.66647838 4.36676183 4.37343028 4.22076902 4.82762783 5.49822304 5.34206021 3.97873985
KGC  19.43729119 12.94040894 14.44451223 15.21685062 12.47636644 16.37698455 12.25605406 17.81629876 14.44183712
GRAB 7.87313103 6.16068640 5.10850975 4.98459953 4.31883034 5.12649378 4.42011910 6.48549639 5.29253805
      HOOD      RIG      KGC      GRAB      BTG      RELY      CFLT      PAYC      TTMI
SPOT 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
DJT  6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
SOUN 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
HIMS 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
IONQ 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
NGD  6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
PTON 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
WULF 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
HOOD 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
RIG  6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
KGC  6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
GRAB 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
      CUNA      NXT      MCW      MBLV      ETSY      LAUR      IP      DNB      KTB
SPOT 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
DJT  6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
SOUN 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
HIMS 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
IONQ 6.99294043 15.45796659 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015 5.69530070 11.56416015
```

Figura 12. Rezultatul distantei Manhattan

Analizând rezultatele, constatăm că distanțele Manhattan sunt considerabil mai mari decât distanțele Euclidiene, deoarece suma absolută a diferențelor dintre coordonate tinde să depășească suma pătratelor diferențelor, urmată de extragerea rădăcinii pătrate. În general, ambele metode mențin o ordine similară a proximității între puncte, astfel încât observațiile apropiate în metrică Euclidiană rămân, de obicei, apropiate și în metrică Manhattan, deși pot exista mici diferențe. Aceste două tipuri de distanțe oferă perspective complementare asupra similarității între companii. Deoarece analiza se concentrează pe evidențierea diferențelor semnificative între valori, distanța Euclidiană este mai adecvată, datorită sensibilității sale la variații mari.

Algoritmul aglomerativ de CLUSTERIZARE IERARHICA – utilizand doua metode de calcul a distantei dintre clustere:

➤ Metoda Single Linkage

În primul rand, am aplicat algoritmul aglomerativ de clusterizare ierarhică utilizând **metoda „single”**, care calculează distanța dintre cele mai apropiate puncte ale două clustere.

Metoda Single este o tehnică de clusterizare aglomerativă ierarhică care determină distanța între două clustere ca fiind cea mai mică distanță dintre toate perechile de observații din fiecare cluster. Astfel, la fiecare pas al procesului de clusterizare, se combină cele două clustere care au distanța minimă între ele, iar distanța finală dintre aceste clustere este dată de cel mai apropiat punct din fiecare grup.

Mai întâi, am calculat matricea de distanță euclidiană între observațiile din setul de date (la punctul anterior). Apoi, am utilizat funcția `hclust()` pentru a aplica algoritmul de clusterizare ierarhică, specificând metoda „single” pentru a determina distanța dintre clustere pe baza celor mai apropiate puncte.

Rezultatele clusterizării, care includ informații despre ordinea în care clusterele au fost combinate și distanțele asociate, sunt afișate prin comanda `cbind(clust_single$merge, clust_single$height)`, care combină informațiile despre combinațiile de clustere și înălțimile acestora în dendograma.

[,1]	[,2]	[,3]	[38,]	35	36	0.80433096			
[1,]	-3	-30	0.08999484	[39,]	-54	38	0.84909568		
[2,]	-5	-38	0.23500668	[40,]	-32	39	0.84913946		
[3,]	-64	2	0.30431472	[41,]	-8	-16	0.88419316		
[4,]	-48	1	0.31871626	[42,]	-75	28	0.90834598		
[5,]	-39	-60	0.31953204	[43,]	40	42	0.91124632		
[6,]	-9	4	0.35388627	[44,]	37	43	0.93605439		
[7,]	-4	6	0.37496296	[45,]	34	44	0.94318858		
[8,]	-44	7	0.39165704	[46,]	-28	45	0.97717192		
[9,]	-18	3	0.39924134	[47,]	-1	41	1.09891793		
[10,]	-51	-79	0.40340310	[48,]	-50	46	1.12647914		
[11,]	-61	10	0.40490515	[49,]	47	48	1.13128024		
[12,]	-34	-63	0.42805641	[50,]	-47	49	1.14077488		
[13,]	-27	-41	0.43466481	[51,]	-13	50	1.20183933		
[14,]	8	9	0.44005490	[52,]	-31	51	1.20911067		
[15,]	-15	14	0.48308702	[53,]	-23	-53	1.25366005		
[16,]	-43	-74	0.48420282	[54,]	52	53	1.26641495		
[17,]	5	16	0.49173758	[55,]	-72	54	1.27216413		
[18,]	-24	11	0.49813078	[56,]	-37	-66	1.35385166		
[19,]	-6	15	0.51068789	[57,]	-14	55	1.42148821		
[20,]	-36	13	0.51610901	[58,]	-45	-55	1.43039072		
[21,]	17	18	0.52284093	[59,]	-49	57	1.51157283		
[22,]	12	19	0.53936608	[60,]	58	59	1.53215013		
[23,]	-7	22	0.56047853	[61,]	56	60	1.66203962		
[24,]	-22	-56	0.61999348	[62,]	-11	61	1.73822960		
[25,]	20	21	0.62384036	[63,]	-58	62	1.76810335		
[26,]	23	24	0.67170852	[64,]	-67	63	1.78101805		
[27,]	-29	-73	0.67397513	[65,]	-25	-81	1.80628811		
[28,]	-69	27	0.68938187	[66,]	-12	64	1.87316423	[75,]	72
[29,]	-26	-80	0.69501540	[67,]	-17	66	1.87791193	[76,]	-10
[30,]	-68	-77	0.69870671	[68,]	-19	67	1.92232553	[77,]	-59
[31,]	26	29	0.70760429	[69,]	-62	68	2.09891694	[78,]	-65
[32,]	25	30	0.71266192	[70,]	-35	69	2.13342290	[79,]	-42
[33,]	-57	-70	0.72400491	[71,]	-78	70	2.26224270	[80,]	-76
[34,]	-21	31	0.73239565	[72,]	-52	65	2.28273287		
[35,]	-40	32	0.74286424	[73,]	-71	71	2.41348978		
[36,]	-46	33	0.77171179	[74,]	-2	73	2.44736973		
[37,]	-20	-33	0.77193505	[75,]	72	74	2.51525057		

Figura 11. Rezultatul clusterizării prin metoda single

În final, am vizualizat dendrograma rezultatului clusterizării utilizând funcția plot(), care afișează cum au fost aglomerate observațiile și cum au evoluat clusterelor pe măsură ce algoritmul a progresat.

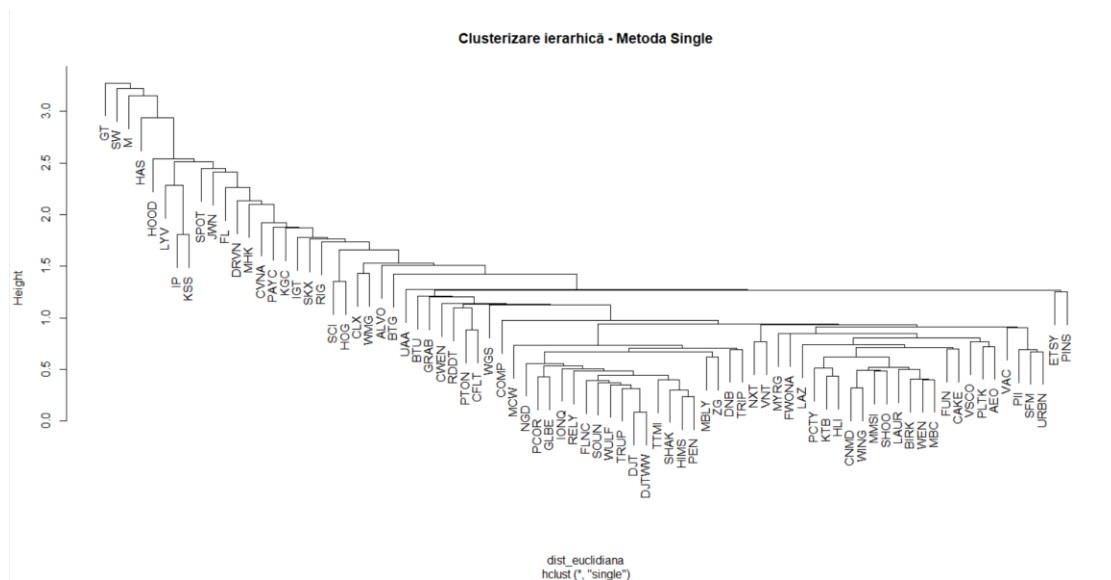


Figura 12. Dendrograma clusterizării prin metoda single

În rezultatele clusterizării aglomerative ierarhice, fiecare linie reprezintă o etapă în procesul de combinare a observațiilor în cluster. În prima etapa, observațiile 3 și 30 sunt combinate într-un cluster, având o înălțime de aproximativ 0.08. Aceasta înseamnă că la acest nivel, distanța dintre cele două observații este relativ mică, ceea ce le face să fie grupate

împreună. În a doua etapa, observațiile 5 și 38 formează un alt cluster, la o înălțime de 0.23. Acesta este un alt grup de observații care sunt mai apropiate între ele comparativ cu celelalte observații, motiv pentru care sunt combinate. În a treia etapa, observația 64 se atașează clasei 2, care a fost formată în etapa anterioară. Aceasta înseamnă că observația 64 este mai apropiată de observațiile din clasa 2 decât de alte observații și, prin urmare, este adăugată la acest cluster. Înălțimea în acest pas este de aproximativ 0.30. Aceste etape continuă pe parcursul celor 80 de etape de clusterizare, iar în fiecare linie, putem observa cum se formează și se extind grupuri de observații pe baza distanțelor dintre ele. În general, observațiile sunt reunite în clustere la fiecare pas, iar înălțimea (valoarea din a treia coloană) reflectă distanța la care are loc această fuziune. Pe măsură ce algoritmul progresează, distanțele dintre clusterelor formate devin mai mari, iar structura dendogramei se lărgeste.

Un avantaj semnificativ al acestei metode de clusterizare este capacitatea de a evidenția grupurile de observații similare și, totodată, de a identifica observațiile izolate sau diferite, pe baza valorilor înălțimii. Înălțimea reprezintă distanța dintre două clusterelor, iar valori mici ale înălțimii indică faptul că observațiile sau clusterelor sunt foarte apropiate între ele, sugerând similitudini economice ridicate.

De exemplu, în prima etapă a procesului de clusterizare, atunci când observațiile 3 și 30 sunt unite la o înălțime de 0.08, acest lucru sugerează că aceste două firme au indicatori financiari similari, ceea ce le face să fie strâns legate din punct de vedere economic, având o distanță mică între ele.

Pe măsură ce înălțimea crește în etapele ulterioare ale clusterizării, distanțele dintre observații devin tot mai mari, ceea ce semnifică faptul că observațiile devin mai diferite unele de altele sau se izolează mai mult față de restul pieței. De exemplu, la o înălțime de 1.14, observațiile 47 și 49 sunt unite într-un cluster, ceea ce sugerează că aceste firme au caracteristici economice semnificativ diferite față de celelalte din setul de date. Aceasta ar putea indica, de exemplu, performanțe financiare mai volatile sau mai instabile față de celelalte companii.

Astfel, analiza înălțimii în clusterizarea ierarhică permite identificarea nu doar a grupurilor de firme cu comportamente economice similare, dar și a celor care sunt mai izolate sau mai diferite, în funcție de indicatorii financiari. Aceste informații sunt esențiale pentru investitori și analiști financiari, care pot folosi aceste rezultate pentru a distinge între firme stabile și cele cu potențial de risc mai mare, ajutând astfel la luarea unor decizii mai bine fundamentate în gestionarea portofoliilor de investiții.

➤ Metoda Complete Linkage

Ca a doua metoda, am aplicat metoda de clusterizare ierarhică folosind **metoda „Complete”** (cunoscută și sub denumirea de „complete linkage”), care presupune că distanța dintre două clusterelor este determinată de distanța maximă dintre orice două puncte, câte unul din fiecare cluster.

Metoda „Complete” de clusterizare ierarhică presupune că distanța dintre două clusterelor este definită de cea mai mare distanță dintre orice două puncte, câte unul din fiecare cluster. Această

metodă favorizează gruparea observațiilor care sunt apropiate la nivelul celor mai îndepărtate puncte din fiecare cluster, conducând la formarea unor clustere mai compacte și bine definite.

Utilizand matricea distantelor calculate anterior, am aplicat funcția `hclust()` pentru a efectua procesul de clusterizare ierarhică, specificând metoda „complete”. După ce am obținut rezultatele clusterizării, am combinat informațiile despre etapele de fuziune ale clusterelor (din `clust_complete$merge`) și înălțimile acestora (din `clust_complete$height`), pentru a vizualiza detaliile procesului de fuziune.

	[,1]	[,2]	[,3]				
[1,]	-3	-30	0.08999484	[39,]	-13	-72	1.27216413
[2,]	-5	-38	0.23500668	[40,]	-37	-66	1.35385166
[3,]	-64	2	0.31233619	[41,]	18	31	1.37645490
[4,]	-39	-60	0.31953204	[42,]	-1	27	1.41290005
[5,]	-48	1	0.33015436	[43,]	-14	-47	1.42148821
[6,]	-51	-79	0.40340310	[44,]	-45	-55	1.43039072
[7,]	-4	-9	0.40894357	[45,]	-31	24	1.47599708
[8,]	-34	-63	0.42805641	[46,]	32	41	1.53190442
[9,]	-27	-41	0.43466481	[47,]	35	36	1.70398243
[10,]	-43	-74	0.48420282	[48,]	22	34	1.71954145
[11,]	-44	3	0.49121767	[49,]	-25	-81	1.80628811
[12,]	-24	6	0.53094249	[50,]	33	37	1.83485464
[13,]	-15	5	0.57956519	[51,]	-12	-67	1.87316423
[14,]	7	13	0.60576161	[52,]	29	48	2.04296219
[15,]	-7	-18	0.60880096	[53,]	-49	42	2.04977900
[16,]	-22	-56	0.61999348	[54,]	46	47	2.19628403
[17,]	-29	-73	0.67397513	[55,]	40	44	2.22865138
[18,]	-26	-80	0.69501540	[56,]	-11	50	2.23477483
[19,]	-68	-77	0.69870671	[57,]	45	52	2.29226329
[20,]	11	15	0.71249440	[58,]	-35	-78	2.30612466
[21,]	-57	-70	0.72400491	[59,]	-19	-58	2.37604427
[22,]	-36	9	0.73010772	[60,]	-17	-62	2.39042138
[23,]	-40	-61	0.74286424	[61,]	-52	49	2.64832603
[24,]	-20	-33	0.77193595	[62,]	39	54	2.70077924
[25,]	4	10	0.79432340	[63,]	55	59	3.04090056
[26,]	-69	17	0.87953174	[64,]	53	56	3.08741706
[27,]	-8	-16	0.88419316	[65,]	43	62	3.10504685
[28,]	8	16	0.97118418	[66,]	-71	61	3.21464818
[29,]	-54	21	1.01542620	[67,]	-2	38	3.28501990
[30,]	-6	14	1.02288891	[68,]	51	63	3.35635578
[31,]	-46	19	1.07846720	[69,]	58	64	3.67090099
[32,]	12	23	1.13879099	[70,]	-65	66	4.11180772
[33,]	-28	28	1.19217011	[71,]	57	60	4.28357902
[34,]	-75	26	1.19431100	[72,]	-42	68	4.42121749
[35,]	-32	25	1.19972535	[73,]	-59	67	4.59521120
[36,]	-21	20	1.24022451	[74,]	-76	70	4.77079488
[37,]	-50	30	1.25202544	[75,]	-10	65	4.97555184
[38,]	-23	-53	1.25366005	[76,]	69	75	5.21422172
				[77,]	71	72	6.56328013
				[78,]	73	76	7.19192548

Figura 13. Rezultatul clusterizării prin metoda Complete

În final, am generat un grafic care reprezintă dendrograma clusterizării ierarhice, folosind `14isual14 plot()`, pentru a observa cum s-au format clusterele și la ce înălțimi s-au realizat aceste fuziuni între observații.

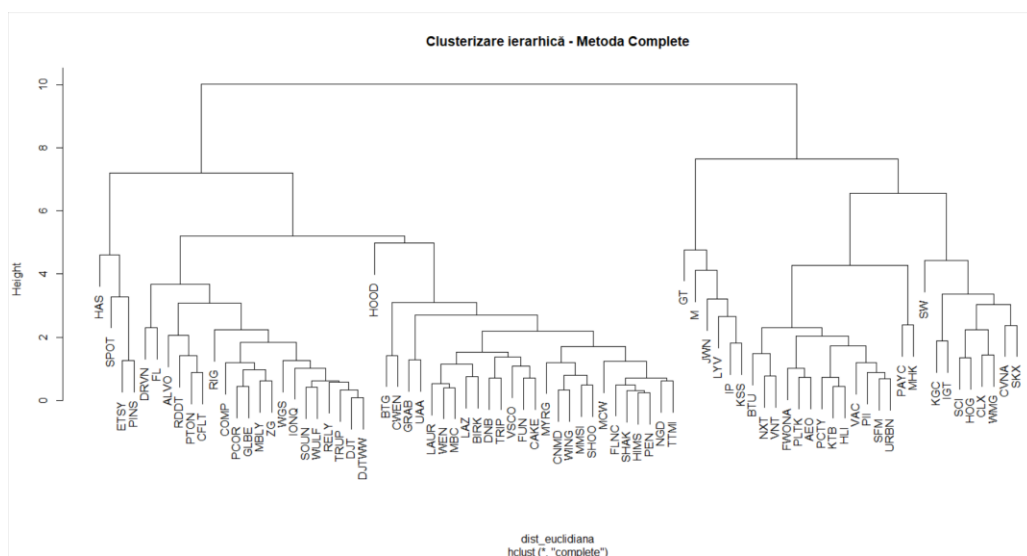


Figura 14. Dendrograma clusterizării prin metoda complete

Comparand rezultatele obtinute prin metodele de clusterizare Complete și Single, putem observa câteva similarități în etapele inițiale ale procesului de grupare, dar diferențele devin mai evidente pe măsură ce procesul avansează. În primele etape, ambele metode unesc observațiile cu caracteristici economice similare, iar distanțele dintre acestea sunt mici, indicând similarități economice ridicate. De exemplu, observațiile 3 și 30 sunt unite la distanțe mici în ambele metode, sugerând o apropiere economică. Totuși, pe măsură ce procesul continuă, diferențele devin mai clare: în Complete Linkage, distanțele dintre clusterelor fuzionate cresc mult mai rapid, ceea ce reflectă o distanțiere semnificativă între grupurile de observații. De exemplu, observațiile 57 și 60 sunt unite într-un cluster la o înălțime de 4.28, indicând o diferență considerabilă între acestea și restul pieței, iar fuziunile târzii, cum ar fi fuziunea observațiilor 71 și 72 la o înălțime de 6.56, sugerează un comportament economic unic, mai riscant sau instabil. În schimb, în metoda Single Linkage, distanțele nu cresc atât de repede și pot apărea fuziuni între grupuri mai diverse, creând clusterelor mai largi și mai puțin omogene. Astfel, în timp ce în primele etape ale procesului ambele metode sunt similare, Complete Linkage tinde să creeze grupuri mai compacte și mai omogene, reflectând diferențe economice semnificative pe măsură ce distanțele cresc. Aceste diferențe sunt utile pentru identificarea firmelor cu comportamente economice distincte și pentru evaluarea riscurilor financiare.

Determinarea numărului optim de clusterelor

Prin urmare, în continuarea analizei, voi opta pentru metoda Complete Linkage pentru a determina numărul optim de clusterelor, deoarece aceasta generează grupuri compacte și bine delimitate, facilitând o interpretare mai clară a datelor și identificarea unor segmente distincte cu relevanță economică. Spre deosebire de metoda Single Linkage, care poate produce clusterelor neomogene, aceasta metoda oferă o structură mai coerentă și mai ușor de analizat.

Variante de decizie a numărului optim de clusterelor:

➤ Criteriul general al clasificării, utilizand dendograma

Criteriul general al clasificării presupune identificarea unui număr optim de clusterelor pe baza dendogramei rezultate. Dendograma este un grafic care ilustrează ierarhia grupării observațiilor, construită de sus în jos.

Pentru a determina numărul optim de clusterelor, vizualizam distanțele dintre etapele succesive de grupare, identificând distanța cea mai mare dintre două etape consecutive.

În figura următoare, este evidențiată etapa în care această distanță maximă a fost identificată.

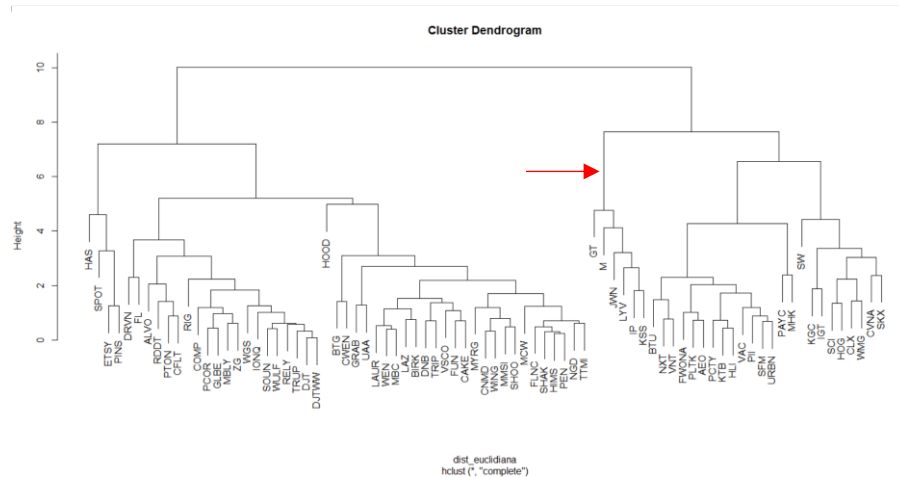


Figura 15. Identificarea celei mai mari distante

Prin urmare, vom trasa o linie orizontală la acest nivel și vom număra grupurile separate de linie pentru a determina numărul de clustere. În acest caz, observăm că linia de tăiere separă 5 grupuri.

În figura următoare, este ilustrată această metodă, utilizând funcția `rect.hclust` pentru a delimita vizual cele 5 clustere identificate în dendrogramă.

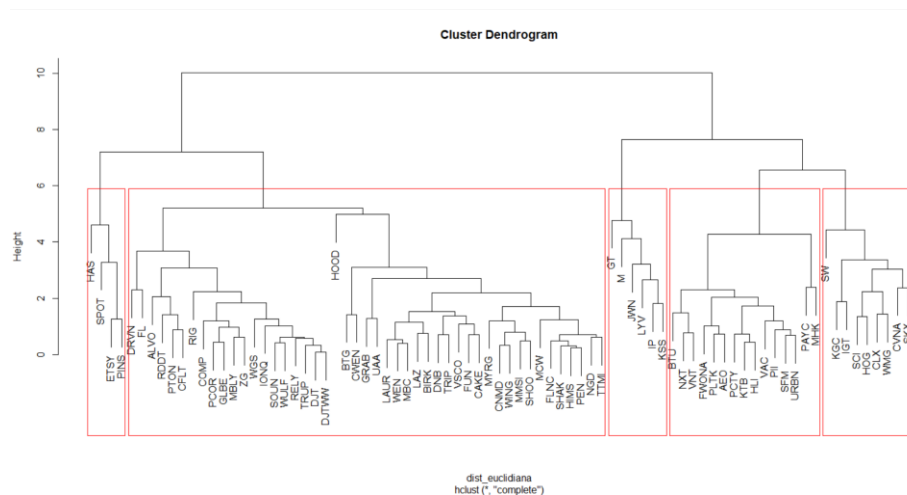


Figura 16. Separarea celor 5 clustere identificate

➤ Graficul "Elbow"

Următoarea metodă utilizată pentru determinarea numărului optim de clustere a fost metoda graficului "elbow". Aceasta este o tehnică vizuală ce ajută la identificarea punctului în care creșterea numărului de clustere nu aduce îmbunătățiri semnificative ale sumei pătratelor distanțelor (WSS - Within-Cluster Sum of Squares).

În acest caz, am utilizat funcția `fviz_nbclust` din pachetul `factoextra` pentru a crea un grafic "elbow", unde pe axa orizontală sunt reprezentate numerele de clustere, iar pe axa verticală valoarea WSS (suma pătratelor distanțelor intra-cluster), care trebuie să fie cât mai mică

posibil. Linia verticală, marcată cu linii punctate, indică punctul de "cotitură" sau "elbow", care corespunde numărului optim de clustere.

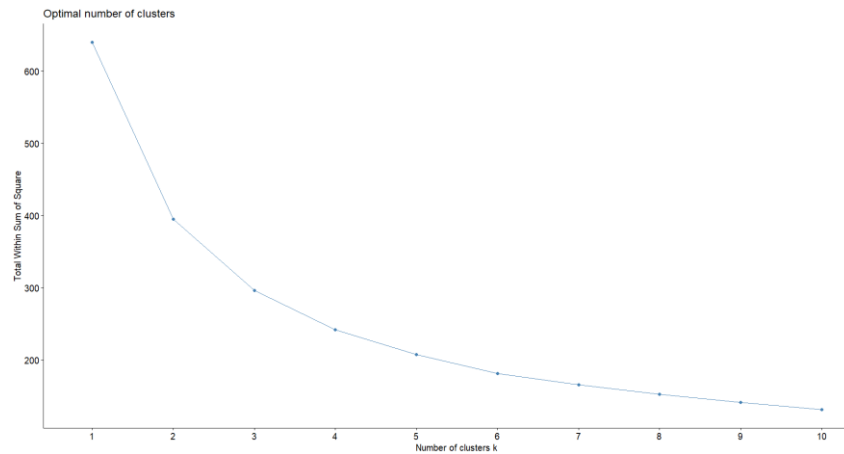


Figura 17. Graficul Elbow

Asadar, am ales 6 clustere, deoarece, după acest punct, graficul arată o stabilizare a valorii WSS, sugerând că adăugarea unor clustere suplimentare nu mai reduce semnificativ dispersia internă a grupurilor. Acest punct este esențial pentru a selecta un număr adecvat de clustere, evitând atât sub-gruparea, cât și supra-gruparea observațiilor.

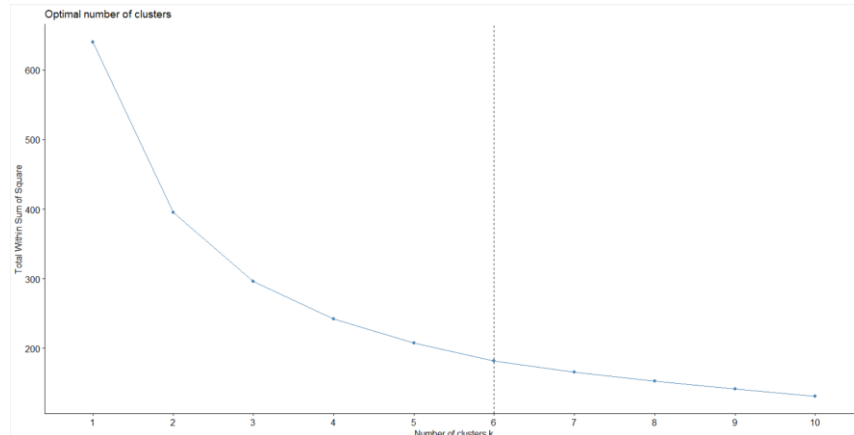


Figura 18. Identificarea punctului de cotitura

➤ Calcularea unor indici

O alta varianta este calcularea indicilor de evaluare a numărului optim de clustere, utilizând pachetul *NbClust*. Această metodă combină mai multe tehnici statistice pentru a determina numărul ideal de clustere, având în vedere atât distanțele dintre observații, cât și diversitatea internă a acestora. Funcția *NbClust* aplică algoritmul de grupare ierarhică (prin metoda "complete linkage") și calculează un set larg de indici de validare care ajută la alegerea numărului optim de clustere.

În acest caz, datele sunt standardizate, iar distanța utilizată este distanța Euclidiană. Parametrii *min.nc* și *max.nc* definesc intervalul de posibile numere de clustere (de la 3 la 10), iar rezultatele sunt obținute prin evaluarea tuturor indicilor disponibile (index = "all"). Aceste rezultate ajută să se stabilească un număr optim de clustere pe baza celor mai potrivite criterii de clasificare.

\$All.index																		
	KL	CH	Hartigan	CCC	Scott	Marriot	TrCovW	TraceW	Friedman	Rubin	Cindex	DB	Silhouette					
3	2.5540	39.5894	10.1118	-3.2376	195.8447	220753085270	1592.7374	317.6003	33.0628	2.0151	0.3372	1.1356	0.3304					
4	0.5286	32.7596	13.1580	-4.4138	244.7881	214469832956	1185.7192	281.1521	37.3334	2.2763	0.3386	1.1987	0.3270					
5	0.6568	31.6414	19.0858	-4.2839	296.9853	175923071908	932.0064	240.1196	42.4636	2.6653	0.4046	1.1763	0.2649					
6	8.3404	35.0202	4.8165	-1.8189	352.5464	127580736671	692.1530	191.9223	48.3732	3.3347	0.3582	1.2453	0.2145					
7	0.6306	31.4357	5.5377	-3.2540	389.5634	109952838438	609.2761	180.3408	50.9155	3.5488	0.3594	1.0902	0.2302					
8	1.1273	29.3503	4.9146	-3.2293	435.2107	81741992063	572.6381	167.7848	55.5846	3.8144	0.3692	1.0067	0.2408					
9	0.7786	27.6409	5.5422	-3.2449	463.3000	73138309262	474.7694	157.2014	58.4213	4.0712	0.3817	0.9260	0.2529					
10	0.7571	26.7007	6.7414	-3.0245	503.8214	54751744966	417.3906	145.9657	62.0938	4.3846	0.3882	0.8734	0.2654					
\$All.CriticalValues																		
	CritValue_Duda	CritValue_PseudoT2	Fvalue_Beale															
3	0.7080	20.2060	0.2615															
4	0.6108	14.0177	0.0008															
5	0.6992	19.3613	0.0033															
6	0.6285	14.7761	0.9995															
7	0.3012	9.2802	1.0000															
8	0.1620	10.3432	1.0000															
9	0.4148	9.8744	1.0000															
10	0.5288	11.5843	0.0004															
\$Best.nc																		
	KL	CH	Hartigan	CCC	Scott	Marriot	TrCovW	TraceW	Friedman	Rubin	Cindex	DB	Silhouette					
Number_clusters	6.0000	3.0000	6.0000	6.0000	6.0000	6	4.0000	6.0000	6.0000	6.0000	3.0000	10.0000						
Value_Index	8.3404	39.5894	14.2693	-1.8189	55.5612	307144370004	407.0183	36.6158	5.9096	-0.4552	0.3372	0.8734						
	KL	CH	Hartigan	CCC	Scott	Marriot	TrCovW	TraceW	Friedman	Rubin	Cindex	DB	Silhouette					
Number_clusters	3.0000	3.000	3.0000	3.0000	3.0000	4.0000	5.000	2	3.0000	5.0000	0	9.0000						
Value_Index	0.3304	0.804	11.9471	1.2629	0.4046	35.5787	0.572	NA	0.5894	0.1196	0	1.1566						
\$Best.partition																		
	Index	Sdbw																
Number_clusters	0	10.0000																
Value_Index	0	0.2056																
\$Best.partition																		
RDDT	SPOT	DJT	SOUN	HIMS	IONQ	NGD	PTON	WULF	HOOD	RIG	KGC	GRAB	BTU	RELY	CFLT	PAYC	TTMI	CVNA
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NXT	MCW	MBLY	ETSY	LAUR	IP	DNB	KTB	COMP	SFM	DJTWW	BTU	MYRG	VNT	PCOR	DRVN	PCTY	SCI	PEN
2	1	1	1	1	3	1	1	2	1	2	1	2	1	2	1	2	2	1
CNMD	LAZ	HLI	SW	MMSI	FLNC	CLX	VSCO	CWEN	TRUP	ALVO	WGS	WEN	LVY	PINS	FWONA	WMG	ZG	PLTK
1	1	2	2	1	1	2	1	1	1	1	1	1	3	1	1	2	1	2
SKX	HAS	WING	BIRK	MHK	GLBE	SHAK	M	HOG	IGT	FUN	PII	AEO	JWN	UAA	URBN	SHOO	VAC	GT
1	1	1	1	1	2	1	3	2	2	1	2	2	3	1	2	1	2	3
CAKE	FL	MCB	TRIP	KSS														
1	1	1	1	1	3													

Figura 19. Rezultatul funcției NbClust

Rezultatele obținute sugerează o evaluare detaliată a performanței mai multor metode de clustering pentru un set de date. Indicatorii utilizați, precum KL, CH, Hartigan, CCC, Scott și alții, indică variații semnificative în funcție de numărul de clustere ales. De exemplu, pentru majoritatea indicilor, cel mai bun număr de clustere este de 6, cu excepția câtorva care sugerează 3 sau 4 clustere. Cele mai mari valori pentru indicii CH și Scott se obțin pentru 6 clustere, ceea ce sugerează o bună separare a datelor în acest caz. Silhouette, un indicator important al calității clusteringului, variază, având o valoare maximă de 0.33 pentru 6 clustere, indicând o separare relativ bună a grupurilor. În ceea ce privește „Best.partition”, rezultatele sugerează o repartizare diversă a valorilor în diverse clustere, cu unele date care sunt mai bine grupate în clustere specifice, cum ar fi 1 sau 2, pe când altele sunt mai distribuite.

În concluzie, analiza sugerează că 6 clustere sunt cel mai bine definite pe baza evaluărilor făcute de majoritatea indicilor, dar există și dovezi pentru 3 clustere.

Concluzie

Conform primului grafic, numărul optim de clustere este 5, în timp ce al doilea grafic sugerează 6 clustere ca fiind cel mai adecvat. Totuși, având în vedere că majoritatea indicilor indică 6 ca fiind opțiunea cea mai bună, vom opta pentru această valoare.

În continuare, voi calcula **indicele siluetei** pentru fiecare observatie și voi realiza **graficul siluetei** pentru a evalua calitatea clusteringului realizat.

	cluster	neighbor	sil_width
[1,]	1	3	0.39083616
[2,]	2	3	0.25011814
[3,]	1	3	0.19022823
[4,]	1	3	0.30360100
[5,]	3	1	0.14828405
[6,]	1	3	0.40039770
[7,]	3	1	0.22793956
[8,]	1	3	0.43667370
[9,]	1	3	0.16880212
[10,]	3	1	0.02725244
[11,]	1	3	0.19927383
[12,]	4	5	0.29961727
[13,]	3	1	0.03925234
[14,]	3	5	0.15442431
[15,]	1	3	0.29896339
[16,]	1	3	0.44782234
[17,]	5	3	0.34989041
[18,]	3	1	0.25997030
[19,]	4	5	0.17904494
[20,]	5	3	0.38559157
[21,]	3	1	0.21607932
[22,]	1	3	0.21804751
[23,]	2	5	-0.30982086
[24,]	3	5	0.26897676
[25,]	6	4	0.12267992
[26,]	3	1	0.37352807
[27,]	5	3	0.25763275
[28,]	1	3	0.31770351
[29,]	5	3	0.33942933
[30,]	1	3	0.16629663
[31,]	5	3	0.32229064
[32,]	3	1	0.32222109
[33,]	5	3	0.36805626
[34,]	1	3	0.29739898
[35,]	1	3	0.29300953
[36,]	5	3	0.01193652
[37,]	4	5	0.08336877
[38,]	3	1	0.21864549
[39,]	3	5	0.28730153
[40,]	3	5	0.11371112
[41,]	5	3	0.30646648
[42,]	4	6	0.21663663
[43,]	3	5	0.40156724
[44,]	3	1	-0.02110568
[45,]	4	5	-0.14327682
[46,]	3	5	0.01555481
[47,]	3	5	0.06255887
[48,]	1	3	0.08719481
[49,]	1	3	0.39023775
[50,]	1	3	0.32463348
[51,]	3	5	0.16822785
[52,]	6	4	0.26542847
[53,]	2	3	0.16269923
[54,]	5	3	0.14472773
[55,]	4	5	0.20720609
[56,]	1	3	0.13229313
[57,]	5	3	0.03421751
[58,]	4	5	0.30659582
[59,]	2	1	0.13253742
[60,]	3	5	0.26610701
[61,]	3	5	0.23198174
[62,]	5	4	-0.01692629
[63,]	1	3	0.20621541
[64,]	3	1	0.13605552
[65,]	6	4	0.37221681
[66,]	4	5	0.11079324
[67,]	4	5	0.29271139
[68,]	3	5	0.16006153
[69,]	5	3	0.28981643
[70,]	5	3	0.13621448
[71,]	6	5	0.08930757
[72,]	3	1	0.15185818
[73,]	5	3	0.34343535
[74,]	3	5	0.21659943
[75,]	5	3	0.24667908
[76,]	6	3	0.13207253
[77,]	3	5	0.31194886
[78,]	1	3	0.19017973
[79,]	3	5	0.25678686
[80,]	3	1	0.34689622
[81,]	6	4	0.26131455

Figura 20. Indicele siluetei pt fiecare observatie

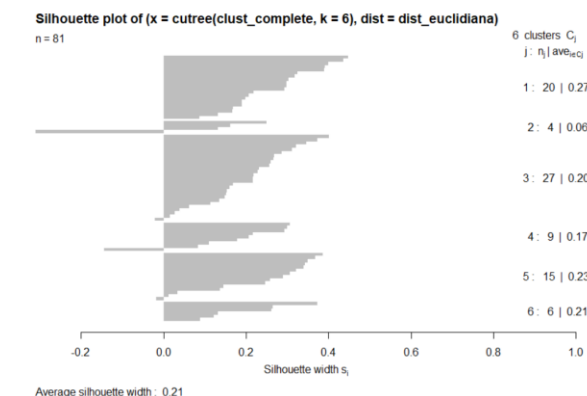


Figura 21. Graficul siluetei

Graficul generat vizualizează aceste valori, iar valorile siluetei au următoarele semnificații: dacă $s(i)$ este aproape de 1, observația este bine încadrarea în clusterul său; dacă este aproape de 0, observația se află la limita dintre două clustere, indicând incertitudine; iar dacă

$s(i)$ este negativ, înseamnă că observația este greșit plasată într-un cluster și ar trebui să facă parte din altul.

Din analiza graficului, se poate observa că fiecare cluster are valorile lui $s(i)$ mai mari decât 0, ceea ce sugerează că observațiile sunt corect încadrate în clusterelor respective. De asemenea, în urma calculării indicilor, se remarcă faptul că un număr foarte mic de observații au un indice negativ.

În concluzie, având în vedere că majoritatea observațiilor au fost corect alocate în cluster, cu valori pozitive ale indicelui $s(i)$, clusterizarea poate fi considerată relativ corectă. Indicele negativ este întâlnit doar pentru un număr foarte mic de observații, ceea ce sugerează că erorile de încadrare sunt minime. Aceste rezultate indică o separare eficientă între cluster și o distribuție corespunzătoare a datelor în grupurile formate.

Calcularea indicatorilor statistici per cluster

Pentru a calcula indicatorii statistici per fiecare cluster, mai întâi am asignat fiecare observație la clusterul corespunzător folosind funcția cutree, care returnează numărul clusterului pentru fiecare observație din setul de date. Apoi, am separat datele în 6 grupuri distincte pe baza rezultatelor obținute și am calculat indicatorii statistici pentru fiecare grup în parte utilizând funcțiile summary și describe. Astfel vom putea înțelege caracteristicile fiecărui cluster în mod individual, formându-ne o imagine de ansamblu asupra distribuției și comportamentului datelor în fiecare grup.

	X1	X2	X3	X4	X5	X6	X7	X10
RDDT	-0.6678924	-0.47265005	-2.33103597	-2.1192932	-2.0531849	-1.8418471	-0.9263845	-0.7076623
DJT	-0.8547343	-0.99791259	-0.77455891	-0.4727237	-0.6053711	-0.5784088	-1.0270677	-0.8163942
SOUN	-0.8449188	-0.97705122	-0.91891164	-0.6386272	-0.7557074	-0.5576286	-1.1718388	-1.0348262
IONQ	-0.8495582	-0.98999986	-1.23203352	-0.8100626	-1.0262901	-0.7321826	-1.2197821	-1.1997348
PTON	-0.3392432	-0.26769432	-1.75129873	-2.0487029	-1.5911250	-1.2766247	-1.1530548	-1.0712143
WULF	-0.8325345	-0.95494794	-0.78590478	-0.4505449	-0.6184414	-0.4745077	-0.8879694	-1.1196457
RIG	-0.2670510	0.41635371	-0.78968255	-1.4010241	-0.5286383	-0.5742528	-0.7795661	-1.9093919
RELY	-0.6486738	-0.61886812	-0.99428670	-0.6184690	-0.7745820	-0.6282814	-1.0216364	-0.8543568
CFLT	-0.6900351	-0.62032670	-1.76855318	-1.4668868	-1.5276077	-0.8901123	-1.0370274	-0.9018133
MBLY	-0.5027844	-0.46656315	-1.36374435	-1.0020105	-1.0345844	-0.5160681	-0.4085360	-0.1774261
COMP	0.1364488	-0.43632294	-1.29229998	-1.0380813	-1.0933772	-0.5991891	-0.9384832	-0.7251561
DJTWW	-0.8547344	-0.99791271	-0.68456407	-0.4727237	-0.6053711	-0.5784088	-1.0270677	-0.8163942
PCOR	-0.6632070	-0.49878961	-1.14896001	-0.7249769	-0.8748249	-0.8028354	-0.7105344	-0.4977025
DRVN	-0.4122569	-0.26606329	0.00959196	-2.7205365	-2.2647547	-2.3613529	-0.4961348	-1.4689266
TRUP	-0.6255969	-0.90814727	-0.77079617	-0.3473504	-0.5729250	-0.5867209	-0.9199428	-0.7037528
ALVO	-0.7968740	-0.91005277	-1.04114669	-2.2609175	-1.7429697	-1.5010512	-1.6986098	-1.9328388
WGS	-0.8044337	-0.90380373	-0.87908077	-0.5535096	-0.6964953	-1.6839173	-1.1286272	-0.9847433
ZG	-0.4595762	-0.03299875	-1.28853975	-0.7370654	-0.7886384	-0.6532177	-0.3437175	-0.3650278
GLBE	-0.7345245	-0.83421057	-1.01212273	-0.6370950	-0.8208081	-0.6739979	-0.7148481	-0.3575037
FL	0.7030310	0.36898953	-0.38608452	-1.4520978	-1.6623322	-2.0122450	-0.1090298	-0.3415776

Showing 1 to 20 of 20 entries. 8 total columns

Figura 22. Observatiile din Cluster 1

```
> summary(cluster1)
      X1      X2      X3      X4      X5      X6
Min.   :-0.8547 Min.   :-0.9979 Min.   :-2.331036 Min.   :-2.7205 Min.   :-2.2648 Min.   :-2.3614
1st Qu.: -0.8115 1st Qu.: -0.9213 1st Qu.: -1.289480 1st Qu.: -1.4558 1st Qu.: -1.5435 1st Qu.: -1.3327
Median :-0.6655 Median :-0.6196 Median :-1.003205 Median :-0.7736 Median :-0.8478 Median :-0.6636
Mean   :-0.5505 Mean   :-0.5684 Mean   :-1.060201 Mean   :-1.0986 Mean   :-1.0819 Mean   :-0.9761
3rd Qu.: -0.4477 3rd Qu.: -0.3942 3rd Qu.: -0.783068 3rd Qu.: -0.6022 3rd Qu.: -0.6770 3rd Qu.: -0.5784
Max.    : 0.7030 Max.    : 0.4164 Max.    : 0.009592 Max.    : -0.3474 Max.    : -0.5286 Max.    : -0.4745

      X7      X10
Min.   :-1.6986 Min.   :-1.9328
1st Qu.: -1.0599 1st Qu.: -1.0833
Median :-0.9324 Median :-0.8354
Mean   :-0.8860 Mean   :-0.8993
3rd Qu.: -0.7138 3rd Qu.: -0.6522
Max.    : -0.1090 Max.    : -0.1774
```

Figura 23. Summary pentru Cluster 1

```
> describe(cluster1)
      vars  n mean  sd median trimmed mad  min  max range skew kurtosis  se
X1      1 20 -0.55 0.39 -0.67 -0.63 0.26 -0.85 0.70 1.56 1.84 3.07 0.09
X2      2 20 -0.57 0.44 -0.62 -0.63 0.46 -1.00 0.42 1.41 0.89 -0.27 0.10
X3      3 20 -1.06 0.51 -1.00 -1.05 0.34 -2.33 0.01 2.34 -0.42 0.45 0.11
X4      4 20 -1.10 0.70 -0.77 -1.01 0.45 -2.72 -0.35 2.37 -0.88 -0.58 0.16
X5      5 20 -1.08 0.53 -0.85 -1.01 0.36 -2.26 -0.53 1.74 -0.83 -0.72 0.12
X6      6 20 -0.98 0.58 -0.66 -0.88 0.18 -2.36 -0.47 1.89 -1.07 -0.37 0.13
X7      7 20 -0.89 0.36 -0.93 -0.90 0.31 -1.70 -0.11 1.59 0.18 0.06 0.08
X10     8 20 -0.90 0.48 -0.84 -0.85 0.39 -1.93 -0.18 1.76 -0.66 -0.20 0.11
```

Figura 24. Describe pentru Cluster 1

Clusterul 1, ce contine 20 de observatii, include companii cu venituri și profituri relativ scăzute, așa cum indică valorile negative ale variabilelor. De exemplu, venitul total generat (X1) are o medie de -0,55 și o mediană de -0,67, sugerând o predominanță a companiilor cu performanțe financiare reduse. Distribuția este asimetrică spre dreapta, cu un skew de 1,84 și câteva valori semnificativ mai mari, ceea ce indică prezența unor companii care depășesc considerabil performanța majorității. În cazul profitului brut (X2), media de -0,57 și mediana de -0,62 reflectă o distribuție relativ echilibrată, cu o ușoară asimetrie pozitivă (skew 0,89) și o dispersie moderată (range 1,41). Venitul operațional (X3) prezintă o medie de -1,06 și o mediană de -1,00, ceea ce indică o distribuție aproape simetrică, dar cu variații semnificative între companii, sugerate de un range de 2,34. Venitul net (X4) are o medie de -1,10 și o mediană de -0,77, cu o distribuție ușor asimetrică spre stânga (skew -0,88), reflectând pierderi financiare în rândul multor companii din cluster. În mod similar, profitul înainte de dobânzi și taxe (X5) și venitul pe acțiuni (X6) au valori medii negative (-1,08 și -0,98, respectiv), indicând performanțe slabe în ceea ce privește rentabilitatea. Fluxurile de numerar operațional (X7) și liber (X10) confirmă această tendință, având medii de -0,89 și -0,90, respectiv.

Per ansamblu, companiile din primul cluster sunt caracterizate de o performanță financiară slabă, cu pierderi semnificative și o variabilitate ridicată în rândul indicatorilor cheie.

	X1	X2	X3	X4	X5	X6	X7	X10
SPOT	1.79069300	1.2069711	-1.0403646	-0.6381100	-0.7417915	-0.6490616	0.8543074	1.973293
ETSY	-0.32548080	0.1877772	0.1234843	0.6199799	0.1440896	0.4938512	0.6346062	1.528580
PINS	-0.21731799	0.5925041	-0.4452605	0.3258621	-0.2911388	-0.2791735	0.8484000	1.920353
HAS	-0.02906165	0.5732020	0.4737548	-2.3337583	-1.7724223	-2.3281045	1.1813056	1.839962

Showing 1 to 4 of 4 entries. 8 total columns

Figura 25. Observatiile din Cluster 2

```
> summary(cluster2)
```

	X1	X2	X3	X4	X5	X6
Min.	:-0.3255	Min. :0.1878	Min. : -1.0404	Min. : -2.3338	Min. : -1.7724	Min. : -2.32810
1st Qu.	:-0.2444	1st Qu.:0.4768	1st Qu.: -0.5940	1st Qu.: -1.0620	1st Qu.: -0.9994	1st Qu.: -1.06882
Median	:-0.1232	Median :0.5829	Median : -0.1609	Median : -0.1561	Median : -0.5165	Median : -0.46412
Mean	: 0.3047	Mean :0.6401	Mean : -0.2221	Mean : -0.5065	Mean : -0.6653	Mean : -0.69062
3rd Qu.	: 0.4259	3rd Qu.:0.7461	3rd Qu.: 0.2111	3rd Qu.: 0.3994	3rd Qu.: -0.1823	3rd Qu.: -0.08592
Max.	: 1.7907	Max. :1.2070	Max. : 0.4738	Max. : 0.6200	Max. : 0.1441	Max. : 0.49385

	X7	X10
Min.	:0.6346	Min. :1.529
1st Qu.	:0.7950	1st Qu.:1.762
Median	:0.8514	Median :1.880
Mean	:0.8797	Mean :1.816
3rd Qu.	:0.9361	3rd Qu.:1.934
Max.	:1.1813	Max. :1.973

Figura 26. Summary pentru Cluster 2

```
> describe(cluster2)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	4	0.30	1.00	-0.12	0.30	0.22	-0.33	1.79	2.12	0.72	-1.71	0.50
X2	2	4	0.64	0.42	0.58	0.64	0.30	0.19	1.21	1.02	0.30	-1.85	0.21
X3	3	4	-0.22	0.66	-0.16	-0.22	0.68	-1.04	0.47	1.51	-0.15	-2.10	0.33
X4	4	4	-0.51	1.33	-0.16	-0.51	0.93	-2.33	0.62	2.95	-0.43	-1.95	0.67
X5	5	4	-0.67	0.82	-0.52	-0.67	0.66	-1.77	0.14	1.92	-0.35	-1.93	0.41
X6	6	4	-0.69	1.19	-0.46	-0.69	0.85	-2.33	0.49	2.82	-0.39	-1.86	0.60
X7	7	4	0.88	0.23	0.85	0.88	0.16	0.63	1.18	0.55	0.28	-1.85	0.11
X10	8	4	1.82	0.20	1.88	1.82	0.10	1.53	1.97	0.44	-0.59	-1.80	0.10

Figura 27. Describe pentru Cluster 2

Al doilea cluster, ce contine doar 4 observații, se caracterizează printr-o performanță financiară mixtă, cu indicatori care variază de la pierderi moderate la câștiguri pozitive. Variabila X1 (Venitul total generat) are o medie pozitivă de 0,30, dar o mediană ușor negativă de -0,12, indicând o tendință generală de creștere a veniturilor, dar cu o dispersie semnificativă între companii (range de 2,12). De asemenea, distribuția este ușor asimetrică spre dreapta (skew 0,72), ceea ce sugerează prezența unor companii cu performanțe remarcabile. Profitul brut (X2) prezintă valori medii și mediane pozitive (0,64 și 0,58, respectiv), indicând o tendință generală pozitivă în performanțele operaționale. În schimb, venitul operațional (X3) are o medie negativă de -0,22, dar valoarea maximă pozitivă (0,47) sugerează că unele companii din cluster au obținut câștiguri operaționale notabile, în timp ce altele au avut pierderi semnificative (minimul de -1,04). Venitul net (X4) reflectă o medie de -0,51, ceea ce indică pierderi moderate, dar distribuția asimetrică spre stânga (skew -0,43) sugerează că pierderile sunt mai frecvente decât câștigurile. Similar, profitul înainte de dobânzi și taxe (X5) și venitul pe acțiuni (X6) au valori medii negative (-0,67 și -0,69, respectiv), cu o variație semnificativă (range de 1,92 pentru X5 și 2,82 pentru X6), ceea ce evidențiază diferențe notabile în performanțele individuale ale companiilor.

În contrast, fluxurile de numerar operațional (X7) și liber (X10) au medii și mediane pozitive (0,88 și 1,82, respectiv), indicând o capacitate mai bună a companiilor din Clusterul 2 de a genera lichidități și de a gestiona fluxurile financiare. Aceste valori pozitive pot reprezenta un avantaj competitiv pentru companiile din acest cluster, chiar și în condițiile în care alte aspecte financiare sunt mai slabe.

Per ansamblu, Clusterul 2 este caracterizat de o performanță financiară neuniformă, cu companii care prezintă câștiguri operaționale și fluxuri de numerar pozitive, dar care suferă de pierderi la nivelul altor indicatori. Această diversitate sugerează oportunități și riscuri variabile în cadrul clusterului.

	X1	X2	X3	X4	X5	X6	X7	X10
HIMS	-0.6515016	-0.467732676	-0.70450331	-0.2291913255	-0.53337219	-0.366450447	-0.72233356	-0.48394053
NGD	-0.6908167	-0.903163709	-0.47131640	-0.2228166042	-0.58719688	-0.391386727	-0.21452747	-0.45748869
HOOD	-0.4276480	-0.104652247	0.03004761	0.6387934998	0.15296344	-0.270861370	-0.21509422	-2.84070040
GRAB	-0.3614973	-0.362119061	-1.18074649	-0.9668972531	-0.81908882	-0.570096739	0.23294376	0.78068259
BTG	-0.4838808	-0.530657808	0.94070516	-0.7356193919	-0.17883284	-0.449571382	1.42829991	-0.38827371
TTMI	-0.4125689	-0.727318733	-0.38637281	-0.2324249318	-0.34302404	-0.333202072	-0.57600913	-0.70510284
MCW	-0.6722827	-0.592057571	-0.27388928	-0.0616471092	-0.13797535	-0.312421838	-0.54395525	-1.30953210
LAUR	-0.5600290	-0.758130120	0.13754256	0.3367568021	0.28660707	0.119807028	-0.45687270	-0.19025340
DNB	-0.4043997	-0.086799596	-0.28155012	-0.3939137176	-0.16370131	-0.432947195	-0.03705886	-0.05749504
MYRG	-0.1695218	-0.807849980	-0.56376041	-0.1298752434	-0.39093432	0.813866842	-0.81411433	-0.79525216
PEN	-0.6387789	-0.569988293	-0.54914314	-0.2415415946	-0.58523166	-0.254237183	-0.71919321	-0.41302038
CNMD	-0.6115709	-0.571658184	-0.32690602	0.0277415144	-0.18054978	0.905299871	-0.61428110	-0.25087556
LAZ	-0.3164516	-0.458417114	-0.17098933	0.1971563065	0.06539153	0.273580759	0.02583521	0.66333437
MMSI	-0.6067411	-0.632615274	-0.35626590	0.0878680764	-0.16926905	0.435666584	-0.51796303	-0.19062861
FLNC	-0.4457352	-0.842505378	-0.80671640	-0.3567543787	-0.62887887	-0.470351616	-0.74220379	-0.44472842
VSCO	0.3154953	0.369596765	-0.03011607	0.1535329472	0.09676721	0.327609367	-0.09561903	-0.12717568
CWEN	-0.6167318	-0.539431111	0.80465502	0.0003117232	-0.06251210	-0.083839265	0.71572991	0.49928003
WIEN	-0.4334525	-0.531670673	0.30295051	0.3316462358	0.35743014	-0.009030423	-0.19492099	0.15717156
WING	-0.7425245	-0.827557122	-0.34951002	0.0347162722	-0.19782924	1.0021689181	-0.58573190	-0.31608725

Figura 28. Observatiile din Cluster 3

```
> summary(cluster3)
```

	X1	X2	X3	X4	X5	X6
Min.	:-0.7425	Min. :-0.9032	Min. :-1.18075	Min. :-0.9668973	Min. :-0.81909	Min. :-0.57010
1st Qu.	:-0.6142	1st Qu. :-0.6506	1st Qu. :-0.51023	1st Qu. :-0.2369833	1st Qu. :-0.46215	1st Qu. :-0.34983
Median	:-0.4839	Median :-0.5307	Median :-0.28155	Median : 0.0003117	Median :-0.17883	Median :-0.15449
Mean	:-0.4394	Mean :-0.4237	Mean :-0.23658	Mean :-0.0722356	Mean :-0.19051	Mean :-0.02406
3rd Qu.	:-0.3829	3rd Qu. :-0.2401	3rd Qu. : 0.02566	3rd Qu. : 0.1338009	3rd Qu. : 0.08108	3rd Qu. : 0.38164
Max.	: 0.3155	Max. : 0.5672	Max. : 0.94071	Max. : 0.6387335	Max. : 0.35743	Max. : 1.02167

	X7	X10
Min.	:-2.31509	Min. :-2.840700
1st Qu.	:-0.60305	1st Qu. :-0.470715
Median	:-0.46714	Median :-0.316087
Mean	:-0.35310	Mean :-0.306984
3rd Qu.	:-0.09914	3rd Qu. :-0.007693
Max.	: 1.42830	Max. : 0.780683

Figura 29. Summary pentru Cluster 3

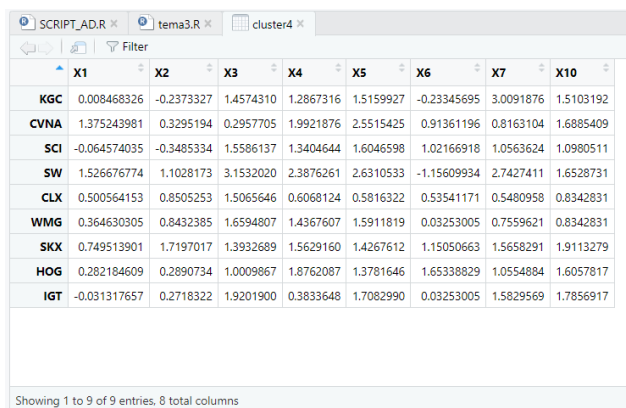
```
> describe(cluster3)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	27	-0.44	0.25	-0.48	-0.48	0.19	-0.74	0.32	1.06	1.55	2.16	0.05
X2	2	27	-0.42	0.36	-0.53	-0.46	0.25	-0.90	0.57	1.47	1.02	0.45	0.07
X3	3	27	-0.24	0.47	-0.28	-0.26	0.42	-1.18	0.94	2.12	0.53	0.47	0.09
X4	4	27	-0.07	0.35	0.00	-0.05	0.34	-0.97	0.64	1.61	-0.49	0.17	0.07
X5	5	27	-0.19	0.33	-0.18	-0.19	0.41	-0.82	0.36	1.18	-0.15	-1.09	0.06
X6	6	27	0.02	0.48	-0.15	-0.01	0.44	-0.57	1.02	1.59	0.61	-1.07	0.09
X7	7	27	-0.35	0.63	-0.47	-0.37	0.37	-2.32	1.43	3.74	-0.09	3.34	0.12
X10	8	27	-0.31	0.69	-0.32	-0.24	0.40	-2.84	0.78	3.62	-1.60	4.51	0.13

Figura 30. Describe pentru Cluster 3

Al treilea cluster, cu 27 de observații, este caracterizat de performanțe financiare în general sub medie, cu valori negative predominante în majoritatea variabilelor. De exemplu, venitul total generat (X1) are o medie de -0,44 și o mediană de -0,48, ceea ce sugerează că majoritatea companiilor din acest cluster au înregistrat venituri scăzute sau pierderi. Profitul brut (X2) urmează o tendință similară, având o medie de -0,42 și o mediană de -0,53, indicând o performanță slabă în această zonă. În ceea ce privește venitul operațional (X3), media de -0,24 și mediană de -0,28 sugerează o ușoară îmbunătățire pentru unele companii, dar majoritatea au rezultate sub nivelul așteptărilor. Venitul net (X4) prezintă o valoare aproape de zero, cu o mediană de 0, ceea ce indică faptul că unele companii au reușit să își mențină stabilitatea financiară. Profitul înainte de dobânzi și taxe (X5) și fluxul de numerar operațional (X6) au, de asemenea, valori negative, ceea ce reflectă dificultăți financiare, dar cu o anumită variație între companii, indicând câțiva participanți mai performanți. Fluxul de numerar liber (X7) este predominant negativ, cu o medie de -0,35, ceea ce sugerează o gestionare dificilă a acestui indicator. Venitul pe acțiuni (X10) prezintă o valoare medie de -0,31 și o variație semnificativă, ceea ce indică o performanță financiară slabă în majoritatea companiilor din cluster, dar și câteva companii care au obținut rezultate mai bune.

Per ansamblu, Clusterul 3 reflectă o performanță financiară slabă, dar cu o mare variabilitate între companii, ceea ce sugerează că, în ciuda dificultăților comune, există și câteva exemple de succes în acest grup.



	X1	X2	X3	X4	X5	X6	X7	X10
KGC	0.008468326	-0.2373327	1.4574310	1.2867316	1.5159927	-0.23345695	3.0091876	1.5103192
CVNA	1.375243981	0.3295194	0.2957705	1.9921876	2.5515425	0.91361196	0.8163104	1.6885409
SCI	-0.064574035	-0.3485334	1.5586137	1.3404644	1.6046598	1.02166918	1.0563624	1.0980511
SW	1.526676774	1.1028173	3.1532020	2.3876261	2.6310533	-1.15609934	2.7427411	1.6528731
CLX	0.500564153	0.8505253	1.5065646	0.6068124	0.5816322	0.53541171	0.5480958	0.8342831
WMG	0.364630305	0.8432385	1.6594807	1.4367607	1.5911819	0.03253005	0.7559621	0.8342831
SKX	0.749513901	1.7197017	1.3932689	1.5629160	1.4267612	1.15050663	1.5658291	1.9113279
HOG	0.282184609	0.2890734	1.0009867	1.8762087	1.3781646	1.65338829	1.0554884	1.6057817
IGT	-0.031317657	0.2718322	1.9201900	0.3833648	1.7082990	0.03253005	1.5829569	1.7856917

Figura 31. Observatiile din Cluster 4

```
> summary(cluster4)
      X1      X2      X3      X4      X5      X6
Min.   : -0.064574   Min.   : -0.3485   Min.   : 0.2958   Min.   : 0.3834   Min.   : 0.5816   Min.   : -1.15610
1st Qu.: 0.008468    1st Qu.: 0.2718    1st Qu.: 1.3933    1st Qu.: 1.2867    1st Qu.: 1.4268    1st Qu.: 0.03253
Median : 0.364630    Median : 0.3295    Median : 1.5066    Median : 1.4368    Median : 1.5912    Median : 0.53541
Mean   : 0.523488    Mean   : 0.5356    Mean   : 1.5495    Mean   : 1.4303    Mean   : 1.6655    Mean   : 0.43890
3rd Qu.: 0.749514    3rd Qu.: 0.8505    3rd Qu.: 1.6595    3rd Qu.: 1.8762    3rd Qu.: 1.7083    3rd Qu.: 1.02167
Max.   : 1.526677    Max.   : 1.7197    Max.   : 3.1532    Max.   : 2.3876    Max.   : 2.6311    Max.   : 1.65339

      X7      X10
Min.   : 0.5481    Min.   : 0.8343
1st Qu.: 0.8163    1st Qu.: 1.0981
Median : 1.0564    Median : 1.6058
Mean   : 1.4592    Mean   : 1.4357
3rd Qu.: 1.5830    3rd Qu.: 1.6885
Max.   : 3.0092    Max.   : 1.9113
```

Figura 32. Summary pentru Cluster 4


```
> describe(cluster4)
vars n mean sd median trimmed mad min max range skew kurtosis se
X1 1 9 0.52 0.59 0.36 0.52 0.57 -0.06 1.53 1.59 0.60 -1.32 0.20
X2 2 9 0.54 0.66 0.33 0.54 0.77 -0.35 1.72 2.07 0.28 -1.19 0.22
X3 3 9 1.55 0.76 1.51 1.55 0.23 0.30 3.15 2.86 0.51 0.05 0.25
X4 4 9 1.43 0.64 1.44 1.43 0.65 0.38 2.39 2.00 -0.24 -1.22 0.21
X5 5 9 1.67 0.62 1.59 1.67 0.24 0.58 2.63 2.05 0.13 -0.82 0.21
X6 6 9 0.44 0.86 0.54 0.44 0.75 -1.16 1.65 2.81 -0.36 -1.08 0.29
X7 7 9 1.46 0.88 1.06 1.46 0.75 0.55 3.01 2.46 0.72 -1.18 0.29
X10 8 9 1.44 0.41 1.61 1.44 0.27 0.83 1.91 1.08 -0.48 -1.60 0.14
```

Figura 33. Describe pentru Cluster 4

Al patrulea cluster, cu 8 observații, se caracterizează printr-o performanță financiară semnificativ mai bună comparativ cu alte grupuri, cu valori pozitive pentru majoritatea variabilelor. În ceea ce privește venitul total generat (X1), acest cluster prezintă o medie pozitivă de 0,52 și o medie a valorii medii a observațiilor (mediană) de 0,36, indicând un comportament financiar mai sănătos în acest cluster. De asemenea, variabilele precum profitul brut (X2) și venitul operațional (X3) au valori medii pozitive (0,54 și 1,55), iar venitul net (X4) se află și el în zone pozitive, cu o medie de 1,43 și o mediană de 1,44, ceea ce sugerează o performanță operațională consistentă. Profitul înainte de dobânzi și taxe (X5) are o valoare medie de 1,67, ceea ce confirmă stabilitatea financiară a companiilor din acest cluster. Fluxul de numerar operațional (X6) are o medie mai mică (0,44), dar încă pozitivă, iar fluxul de numerar liber (X7) are o medie de 1,46, indicând o gestiune mai bună a capitalului liber. Venitul pe acțiuni (X10) este, de asemenea, pozitiv, cu o medie de 1,44 și o valoare mediană de 1,61, ceea ce sugerează că companiile din acest cluster au o valoare considerabilă per acțiune.

În ansamblu, acest cluster include companii cu performanțe financiare robuste, iar distribuția variabilelor arată o tendință generală de creștere a valorilor financiare. Cu toate acestea, există și unele variații, cum ar fi variabilitatea fluxului de numerar operațional, care sugerează unele diferențe între companiile din cluster. De asemenea, dispersia relativ mare în jurul mediilor indică o diversitate a performanțelor financiare, cu unele companii înregistrând performanțe extrem de bune, în timp ce altele pot avea fluctuații.

	X1	X2	X3	X4	X5	X6	X7	X10
PAYC	-0.51553831	-0.10638529	0.73089936	1.2201218	0.85613366	3.04150792	0.14304272	0.24479161
NXT	-0.33163032	-0.43149847	0.95338966	1.0138446	1.08100794	1.17128687	-0.28101785	0.27213456
KTB	-0.36561933	-0.32788387	0.06914398	0.4777810	0.18118290	1.35830897	-0.11675433	0.45581673
SFM	0.51894753	0.63100149	0.38473257	0.7496625	0.46806134	0.92192406	0.07108077	0.04042635
BTU	-0.02586883	-0.36582318	0.69711243	1.4715547	1.34500168	1.20869129	-0.06075114	-0.71242601
VNT	-0.28070178	-0.12736276	0.63218579	1.0016763	0.80228086	0.68087334	-0.03504725	0.40916421
PCTY	-0.58738174	-0.41689998	-0.08252616	0.3730415	0.04779585	1.02582523	-0.14552928	0.24104628
HLI	-0.47082104	-0.55584948	0.23079878	0.6957158	0.34056293	1.47467828	-0.11681915	0.34037803
FWONA	-0.14966232	-0.30868255	0.27571598	0.6036203	0.70577634	0.04499819	0.58609289	0.64333134
PLTK	-0.36872421	0.12093483	0.51235979	0.3986869	0.62028083	-0.15864811	0.04117038	0.43194442
MIHK	1.21159091	0.64415840	1.12461550	1.5133202	1.18151787	3.28255863	1.30407858	0.88888858
PII	0.61837349	-0.03737083	0.18672387	0.3629353	0.34013663	1.08400988	0.36191025	0.07181613

Figura 34. Observatiile din Cluster 5

```
> summary(cluster5)
      X1      X2      X3      X4      X5      X6
Min.   :-0.58738 Min.   :-0.55585 Min.   :-0.08253 Min.   :0.2270 Min.   :0.0478 Min.   :-0.1586
1st Qu.: -0.36717 1st Qu.: -0.34685 1st Qu.: 0.24007 1st Qu.:0.4382 1st Qu.:0.3645 1st Qu.: 0.8014
Median : -0.14966 Median : -0.10639 Median : 0.38473 Median :0.6957 Median :0.4681 Median : 1.0840
Mean   : -0.02399 Mean   : -0.05582 Mean   : 0.45678 Mean   :0.7544 Mean   :0.6152 Mean   : 1.1721
3rd Qu.: 0.17067 3rd Qu.: 0.10386 3rd Qu.: 0.66465 3rd Qu.:1.0078 3rd Qu.:0.8292 3rd Qu.: 1.3625
Max.   : 1.21159 Max.   : 0.64416 Max.   : 1.12462 Max.   :1.5133 Max.   :1.3450 Max.   : 3.2826

      X7      X10
Min.   :-0.47335 Min.   :-0.71243
1st Qu.: -0.11679 1st Qu.: 0.05612
Median : 0.02726 Median : 0.24479
Mean   : 0.09048 Mean   : 0.21907
3rd Qu.: 0.10706 3rd Qu.: 0.42055
Max.   : 1.30408 Max.   : 0.88889
```

Figura 35. Summary pentru Cluster 5

```
> describe(cluster5)
      vars  n  mean  sd median trimmed mad  min max range skew kurtosis  se
X1      1 15 -0.02 0.50 -0.15 -0.08 0.46 -0.59 1.21 1.80 0.97 0.04 0.13
X2      2 15 -0.06 0.37 -0.11 -0.07 0.34 -0.56 0.64 1.20 0.56 -0.87 0.10
X3      3 15 0.46 0.33 0.38 0.45 0.29 -0.08 1.12 1.21 0.37 -0.81 0.09
X4      4 15 0.75 0.41 0.70 0.74 0.45 0.23 1.51 1.29 0.59 -1.04 0.10
X5      5 15 0.62 0.38 0.47 0.60 0.35 0.05 1.35 1.30 0.45 -1.00 0.10
X6      6 15 1.17 0.95 1.08 1.11 0.42 -0.16 3.28 3.44 0.84 0.12 0.25
X7      7 15 0.09 0.42 0.03 0.04 0.21 -0.47 1.30 1.78 1.52 2.19 0.11
X10     8 15 0.22 0.38 0.24 0.24 0.28 -0.71 0.89 1.60 -0.67 0.51 0.10
```

Figura 36. Describe pentru Cluster 5

Al cincilea cluster, cu 15 observații, se distinge prin valori financiare negative sau apropiate de zero pentru majoritatea variabilelor, indicând o performanță financiară mai puțin favorabilă comparativ cu alte grupuri. De exemplu, venitul total generat (X1) are o medie negativă de -0,02 și o mediană de -0,15, ceea ce sugerează o tendință generală de scădere a veniturilor. De asemenea, profitul brut (X2) și venitul operațional (X3) au valori negative sau apropiate de zero (medii de -0,06 și 0,46, respectiv), ceea ce sugerează o performanță financiară mai slabă. Cu toate acestea, venitul net (X4) și profitul înainte de dobânzi și taxe (X5) prezintă valori pozitive (medii de 0,75 și 0,62), ceea ce poate indica o oarecare stabilitate operațională în acest cluster. Fluxul de numerar operațional (X6) are o medie de 1,17, ceea ce sugerează că, în ciuda performanței financiare generale mai slabe, companiile din acest cluster au un flux de numerar relativ mai bun, dar cu o dispersie semnificativă (de la -0,16 la 3,28), ceea ce poate reflecta fluctuații în gestiunea capitalului. Fluxul de numerar liber (X7) are o medie de 0,09, cu o variabilitate mare, iar venitul pe acțiuni (X10) este, de asemenea, pozitiv, cu o medie de 0,22, dar cu o dispersie semnificativă între valori, indicând o performanță inconstantă pe termen lung.

În ansamblu, clusterul 5 include companii care prezintă o performanță financiară mixtă, cu unele fluctuații semnificative, dar și cu o posibilă creștere a fluxului de numerar în unele cazuri.

	X1	X2	X3	X4	X5	X6	X7	X10
IP	2.694242	2.130573	0.5564798	1.1015893	0.90956016	0.11149493	2.8390949	1.6885409
LYV	3.697381	2.489806	1.6665499	0.5291963	2.37308170	0.05746633	1.5089721	1.0526448
M	3.639087	4.857049	2.5368677	0.3003700	0.27946995	-0.12955578	1.6120134	0.2547278
JWN	2.006352	2.320637	0.4236184	0.6195808	0.65658714	0.31929727	0.5235095	-0.3281775
GT	2.846809	1.299271	1.5667283	-1.4106004	0.01010053	-0.91504862	1.1135815	-1.4537878
KSS	2.418359	3.167727	0.9325028	0.6195808	0.99622684	0.66009311	1.6477754	1.5913900

Figura 37. Observatiile din Cluster 6

```
> summary(cluster6)
      X1      X2      X3      X4      X5      X6
Min.   :2.006  Min.   :1.299  Min.   :0.4236  Min.   : -1.4106  Min.   :0.0101  Min.   : -0.91505
1st Qu.:2.487  1st Qu.:2.178  1st Qu.:0.6505  1st Qu.: 0.3576  1st Qu.:0.3737  1st Qu.: -0.08280
Median :2.771  Median :2.405  Median :1.2496  Median : 0.5744  Median :0.7831  Median : 0.08448
Mean   :2.884  Mean   :2.711  Mean   :1.2805  Mean   : 0.2933  Mean   :0.8708  Mean   : 0.01729
3rd Qu.:3.441  3rd Qu.:2.998  3rd Qu.:1.6416  3rd Qu.: 0.6196  3rd Qu.:0.9746  3rd Qu.: 0.26735
Max.   :3.697  Max.   :4.857  Max.   :2.5369  Max.   : 1.1016  Max.   :2.3731  Max.   : 0.66009

      X7      X10
Min.   :0.5235  Min.   : -1.4538
1st Qu.:1.2124  1st Qu.: -0.1825
Median :1.5605  Median : 0.6537
Mean   :1.5408  Mean   : 0.4676
3rd Qu.:1.6388  3rd Qu.: 1.4567
Max.   :2.8391  Max.   : 1.6885
```

Figura 38. Summary pentru Cluster 6

```
> describe(cluster6)
      vars n mean  sd median trimmed mad  min max range skew kurtosis  se
X1      1 6 2.88 0.67  2.77    2.88 0.83  2.01 3.70  1.69  0.10   -1.85 0.27
X2      2 6 2.71 1.21  2.41    2.71 0.77  1.30 4.86  3.56  0.65   -1.04 0.49
X3      3 6 1.28 0.80  1.25    1.28 0.82  0.42 2.54  2.11  0.33   -1.63 0.33
X4      4 6 0.29 0.87  0.57    0.29 0.24 -1.41 1.10  2.51 -1.08   -0.47 0.36
X5      5 6 0.87 0.83  0.78    0.87 0.53  0.01 2.37  2.36  0.75   -0.93 0.34
X6      6 6 0.02 0.53  0.08    0.02 0.33 -0.92 0.66  1.58 -0.58   -1.03 0.22
X7      7 6 1.54 0.76  1.56    1.54 0.40  0.52 2.84  2.32  0.40   -1.07 0.31
X10     8 6 0.47 1.22  0.65    0.47 1.42 -1.45 1.69  3.14 -0.38   -1.66 0.50
```

Figura 39. Describe pentru Cluster 6

Al șaselea și ultimul cluster, care conține doar 6 observații, este caracterizat printr-o diversitate semnificativă în valorile variabilelor financiare, cu o performanță mai pozitivă comparativ cu clusterul 5. De exemplu, venitul total generat (X1) are o medie de 2,88 și o gamă de la 2,01 la 3,70, ceea ce sugerează o performanță solidă și o stabilitate relativă în veniturile companiilor din acest cluster. Profitul brut (X2) are, de asemenea, o valoare medie de 2,71, cu o dispersie mare, indicând o performanță financiară în general pozitivă, dar cu unele fluctuații semnificative. Venitul operațional (X3) prezintă o medie de 1,28 și variază de la 0,42 la 2,54, ceea ce sugerează o stabilitate relativă, dar cu o tendință generală de creștere. Totuși, venitul net (X4) are o valoare medie scăzută de 0,29 și o gamă largă, ceea ce poate indica o performanță financiară mai slabă în anumite aspecte ale operațiunilor. Fluxul de numerar operațional (X5) are o medie de 0,87 și o variabilitate semnificativă (de la 0,01 la 2,37), ceea ce sugerează fluctuații importante, dar în general o tendință pozitivă. Fluxul de numerar liber (X6) are o medie aproape de zero (0,02), ceea ce sugerează o performanță mixtă în acest domeniu, cu unele companii având dificultăți în generarea unui flux de numerar pozitiv. Totuși, venitul pe acțiuni (X7) are o medie de 1,54 și o gamă între 0,52 și 2,84, ceea ce sugerează o performanță solidă în ceea ce privește rentabilitatea pe acțiuni. În final, venitul pe acțiuni ajustat (X10) prezintă o medie de 0,47 și o variabilitate mare, indicând o performanță inconstantă în această componentă.

În ansamblu, clusterul 6 cuprinde companii care au o performanță financiară relativ puternică, dar cu fluctuații semnificative în anumite variabile, ceea ce sugerează o diversitate de strategii și condiții financiare în cadrul acestui grup.

Concluziile analizei realizate

Nr. cluster	Nr. de observatii	Caracteristici principale
1	20	<p>Venituri scazute - Majoritatea companiilor din cluster au venituri negative, dar există câteva excepții cu venituri semnificativ mai mari, sugerând asimetrie pozitivă (skewness de 1,84).</p> <p>Profituri negative și echilibrate - Profitul brut are o medie de -0,57, iar distribuția este relativ echilibrată, cu o ușoară asimetrie pozitivă, indicând atât pierderi, cât și câteva profituri moderate.</p> <p>Variabilitate mare în performanță - Venitul operațional prezintă o variabilitate semnificativă, indicând diferențe considerabile între companii în ceea ce privește pierderile și profitabilitatea.</p>
2	4	<p>Venituri mixte - Venitul total generat are o medie pozitivă, dar o mediană negativă, indicând o tendință generală de creștere, dar cu o mare variabilitate între companii și o ușoară asimetrie pozitivă.</p> <p>Performanță operațională mixtă - Profitul brut este în general pozitiv, cu o medie de 0,64 și mediană de 0,58, dar venitul operațional arată o tendință negativă, cu fluctuații semnificative.</p>

		<p>Fluxuri de numerar pozitive</p> <p>- Companiile au fluxuri de numerar operațional și liber pozitive sugerând o capacitate bună de a genera lichidități și de a gestiona fluxurile financiare.</p>
3	27	<p>Performanță financiară sub medie - Majoritatea variabilelor au valori negative, indicând venituri scăzute sau pierderi pentru majoritatea companiilor.</p> <p>Variație semnificativă - Deși performanțele financiare sunt în general slabe, există o mare variabilitate între companii. Fluxul de numerar liber și venitul pe acțiuni au, de asemenea, valori negative, dar cu o diferență considerabilă între companiile care au înregistrat rezultate mai bune.</p> <p>Stabilitate pentru unele companii - Venitul net și venitul operațional sugerează că unele companii au reușit să își mențină stabilitatea financiară, având valori mai apropiate de zero, deși în general performanțele sunt sub așteptări.</p>
4	9	<p>Performanță financiară superioară - Clusterul 4 este caracterizat de valori pozitive pentru majoritatea variabilelor, indicând o performanță financiară semnificativ mai bună decât în alte grupuri.</p>

		<p>Gestionează eficient capitalul - Fluxul de numerar liber este ridicat (1,46), ceea ce sugerează o bună gestionare a resurselor financiare, iar venitul pe acțiune este de asemenea pozitiv (1,44), indicând o valoare bună a acțiunilor.</p> <p>Variație între companii - Deși performanțele financiare sunt robuste, există o variabilitate semnificativă, în special la nivelul fluxului de numerar operațional, ceea ce sugerează diferențe în modul de gestionare financiară între companiile din acest cluster.</p>
5	15	<p>Performanță financiară mixtă - Clusterul 5 se distinge prin valori negative sau aproape de zero pentru majoritatea indicatorilor, cum ar fi venitul total generat și profitul brut, indicând o performanță mai slabă. Totuși, venitul net și profitul înainte de dobânzi și taxe sunt pozitive, ceea ce sugerează o oarecare stabilitate operațională.</p> <p>Fluxuri de numerar variabile - Fluxul de numerar operațional este pozitiv, dar cu o mare dispersie (de la - 0,16 la 3,28), ceea ce sugerează fluctuații semnificative în gestionarea capitalului, dar și unele cazuri de performanță mai bună.</p> <p>Inconstanță în performanța pe termen lung - Venitul pe acțiune are o valoare medie</p>

		pozitivă, dar cu o variabilitate mare, indicând fluctuații semnificative și o performanță financiară inconstantă pe termen lung.
6	6	<p>Performanță financiară relativ puternică - Performanța financiară solidă, cu valori pozitive și relativ stabile pentru majoritatea indicatorilor, cum ar fi venitul total generat și profitul brut. Totuși, există o variabilitate semnificativă în aceste valori, ceea ce sugerează diferențe între companiile din cluster.</p> <p>Fluctuații semnificative în fluxurile de numerar - Fluxul de numerar prezintă o variabilitate semnificativă, cu unele companii având performanțe excelente, iar altele înregistrează fluctuații mai mari.</p> <p>Rentabilitate pe acțiune - Venitul pe acțiune este pozitiv și indică o rentabilitate solidă, cu valori medii semnificative. Totuși, venitul pe acțiune ajustat arată o performanță inconstantă, ceea ce sugerează volatilitate.</p>

În urma caracteristicilor relevate din analiza, putem denumi clusterelor sugestiv după cum urmează:

- Cluster 1 – Companii cu performanță financiară scăzută
- Cluster 2 – Companii cu performanță financiară bună
- Cluster 3 – Companii în recuperare
- Cluster 4 – Companii cu performanță financiară excelentă
- Cluster 5 – Companii cu performanță financiară moderată
- Cluster 6 – Companii cu performanță financiară solidă

În concluzie, analiza clusterizării a evidențiat diversitatea performanțelor financiare ale companiilor, grupate în clustere ce reflectă stadii diferite de dezvoltare și stabilitate. Clusterele variază de la companii cu performanțe scăzute și expunere la riscuri financiare, până la entități cu o performanță solidă și potențial de creștere. Aceste informații sunt valoroase pentru identificarea oportunităților de îmbunătățire și adaptare a strategiilor financiare, dar și pentru estimarea riscurilor asociate diverselor segmente ale pieței. Astfel, analiza clusterizării oferă un cadru util pentru deciziile strategice pe termen lung și investițiile în companii cu potențial variabil.

ALGORITMUL K-MEANS

În această etapă, am aplicat algoritmul de învățare nesupervizată K-means pentru a efectua o clusterizare pe setul de date standardizate, folosind numărul de clustere identificat anterior.

Algoritmul K-means grupează observațiile în funcție de similaritatea lor, astfel încât elementele din fiecare cluster să fie mai apropiate între ele decât de cele din alte clustere.

```
K-means clustering with 6 clusters of sizes 19, 19, 22, 6, 9, 6

Cluster means:
      X1      X2      X3      X4      X5      X6      X7      X10
1 -0.1059988 -0.05639083 0.4189851 0.59898787 0.4669719 0.83399086 0.09903907 0.3050211
2 -0.6164306 -0.61778747 -1.0956804 -1.08003164 -1.0513524 -0.92161080 -0.92688572 -0.9286585
3 -0.5042478 -0.50076847 -0.2230507 -0.03551288 -0.1684132 -0.05387701 -0.47972275 -0.4768965
4 0.3491603 0.49112636 -0.5803297 -0.93636367 -1.0009795 -1.05358352 0.52485392 1.0932896
5 0.6024908 0.51271944 1.5070621 1.53106446 1.7321303 0.74413761 1.54321294 1.4417508
6 2.8837052 2.71084350 1.2804578 0.29328613 0.8708377 0.01729121 1.54082447 0.4675564

Clustering vector:
RDDT SPOT DJT SOUN HIMS IONQ NGD PTON WULF HOOD RIG KGC GRAB BTG RELY CFLT PAYC TTMI CVNA
 2 4 2 2 3 2 3 2 2 3 2 5 4 3 2 2 1 3 5
NXT MCW MBLY ETSY LAUR IP DNB KTB COMP SFM DJTWW BTU MYRG VNT PCOR DRVN PCTY SCI PEN
 1 3 2 1 3 6 3 1 2 1 2 1 3 1 2 2 1 5 3
CNMD LAZ HLI SW MMSI FLNC CLX VSCO CWEN TRUP ALVO WGS WEN LYV PINS FWONA WMG ZG PLTK
 3 1 1 5 3 3 1 1 2 2 2 2 3 6 4 1 5 2 1
SKX HAS WING BIRK MHK GLBE SHAK M HOG IGT FUN PII AEO JWN UAA URBN SHOO VAC GT
 5 4 3 3 5 2 3 6 5 5 3 1 1 6 4 1 3 1 6
CAKE FL MBC TRIP KSS
 3 4 3 3 6

Within cluster sum of squares by cluster:
[1] 29.44011 33.68566 30.83429 21.89426 35.60451 31.81857
(between_SS / total_SS = 71.4 %)

Available components:
[1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss" "size"
[8] "iter" "ifault"
```

Figura 40. Rezultatul aplicării algoritmului K-means

Analizând rezultatele, putem observa că algoritmul K-means a efectuat o clusterizare a observațiilor în 6 clustere, cu următoarele măriri: 19, 19, 22, 6, 9, 6. Comparând cu metoda anterioară (clusterizarea ierarhică), putem observa că algoritmul k-means tinde să integreze mai puține observații în cadrul unui cluster, singura excepție fiind clusterul 2 ce prezintă mai multe observații în acest caz.

Cluster means/Mijloacele pentru fiecare cluster reprezintă mediile variabilelor pentru fiecare cluster în parte. Analizând valorile, putem deduce că clusterul 1 include companii cu performanțe financiare moderate, cu valori ușor negative pentru venituri și profituri, dar cu o gestionare relativ bună a fluxului de numerar și o rentabilitate pe acțiune moderată. Clusterul 2 reflectă companii cu performanțe financiare foarte slabe, cu valori negative pentru majoritatea indicatorilor, sugerând dificultăți financiare serioase și pierderi semnificative. Clusterul 3 include

companii cu performanțe financiare sub medie, cu venituri scăzute și fluctuații în profituri, dar unele companii din acest cluster prezintă o oarecare stabilitate financiară. Clusterul 4 cuprinde companii cu performanțe financiare moderate spre slabe, prezentând atât valori pozitive, cât și negative. Clusterul 5 include companii cu performanțe mixte, cu fluctuații semnificative, dar cu unele indicii de stabilitate operațională, în special în fluxul de numerar. Clusterul 6 conține companii cu performanțe financiare puternice, dar cu fluctuații notabile între variabile, indicând o diversitate a strategiilor și condițiilor financiare.

În cadrul vectorului de clusterizare sunt atribuite fiecărei observații (companie) un număr de cluster corespunzător. De exemplu, companiile RDDT și DJT sunt atribuite clusterului 2, în timp ce companiile NXT și ETSY sunt în clusterul 1. Acesta este un tabel de tip vector care indică la care cluster fiecare observație (companie) a fost atribuită.

Suma pătratică internă a erorii indică un anumit grad de dispersie în interiorul fiecărui cluster, adică cât de mult variază observațiile din interiorul fiecărui cluster față de centrul acestuia. Cu cât valoarea este mai mică, cu atât observațiile din cluster sunt mai apropiate între ele. Putem observa că clusterul 4 are cea mai mică valoare, ceea ce îl face cel mai compact, cu o performanță financiară mai uniformă și mai puțin variabilă față de celelalte cluster. Acesta este urmat de clusterul 1, care prezintă o valoare aproape la fel de mică, indicând un grup compact, dar cu o ușoară variabilitate mai mare comparativ cu clusterul 4.

Procentul de 71.4% indică faptul că 71.4% din variabilitatea totală a datelor este explicată de diferențele între clusterele identificate. Acest lucru sugerează că modelul K-means a realizat o clusterizare destul de bună, întrucât majoritatea variabilității din date poate fi atribuită diferențelor dintre cluster, nu fluctuațiilor din interiorul acestora.

Indicatorii variabilitatii (BSS, WSS, TSS) si calitatea partitiei K-MEANS

În analiza clusterizării, indicatorii variabilitatii ofera informații despre modul în care sunt distribuite observațiile între și în interiorul clusterelor.

```
> variab
      tss      wss      bss calit_part
[1,] 640 183.2774 456.7226   71.36291
> k_means$withinss #variabilitatea intraclassa pt fiecare clasa in parte
[1] 29.44011 33.68566 30.83429 21.89426 35.60451 31.81857
> #suma acestor variabilitati da variabilitatea totala intraclassa (wss)
> sum(k_means$withinss)
[1] 183.2774
> |
```

Figura 41. Indicatorii variabilitatii si calitatea partitiei

Variabilitatea totală (TSS) este de 640, ceea ce reprezintă suma pătratelor diferențelor între valorile observațiilor și media globală a întregului set de date.

Variabilitatea intraclassa (WSS), care măsoară dispersia observațiilor în interiorul fiecărui cluster, este de 183.2774, ceea ce indică o oarecare concentrare a datelor în jurul centrului fiecărui cluster.

Variabilitatea interclasa (BSS), de 456.7226, arată cât de mult variază centrele clusterelor între ele, ceea ce sugerează o separare clară între grupuri. **Calitatea partitiei** este egală cu raportul dintre BSS și TSS care este de aproximativ 71.4%, ceea ce înseamnă că 71.4% din variabilitatea totală este explicată de diferențele dintre cluster, indicând astfel o bună calitate a partitiei și o separare eficientă a datelor. Aceste valori sugerează că algoritmul K-means a reușit să împartă datele într-un mod semnificativ, cu cluster distincte și relativ compacte.

În final, am analizat impactul creșterii numărului de cluster (k) asupra calității rezultatelor și am constatat că, într-adevăr, odată cu creșterea valorii lui k, calitatea partitiei se îmbunătățește, indicând o mai bună separare și o mai mare coerență între clusterele formate.

Spre exemplu, acestea sunt rezultatele atunci când $k = 7$ (când avem 7 cluster):

```
tss      wss      bss calit_part
[1,] 640 162.4849 477.5151 74.61173
```

Observăm că calitatea partitiei a crescut cu 3%, ceea ce este o îmbunătățire evidentă a rezultatului. Așadar, cu cât creștem numărul de cluster, cu atât calitatea clusterizării va crește.

Reprezentarea grafică a clusterelor

Mai întâi am extras informațiile despre clasele obținute în urma algoritmului K-means folosind `k_means$cluster`, care indică apartenența fiecărei observații la un anumit cluster. Apoi, am creat un nou obiect, `c`, în care am combinat aceste clase cu valorile normalizate ale datelor (`data_3_std`), rotunjite la 6 zecimale. Ulterior, am transformat acest obiect într-un `data.frame` numit `m`.

SCRIPT_AD.R

tema3.R

m

Filter

	clasa	X1	X2	X3	X4	X5	X6	X7	X10
RDDT	2	-0.667892	-0.472650	-2.331036	-2.119293	-2.053185	-1.841847	-0.926384	-0.707662
SPOT	4	1.790693	1.206971	-1.040365	-0.638110	-0.741792	-0.649062	0.854307	1.973293
DJT	2	-0.854734	-0.997913	-0.774559	-0.472724	-0.605371	-0.578409	-1.027068	-0.816394
SOUN	2	-0.844919	-0.977051	-0.918912	-0.638627	-0.755707	-0.557629	-1.171839	-1.034826
HIMS	3	-0.651502	-0.467733	-0.704503	-0.229191	-0.533372	-0.366450	-0.722334	-0.483941
IONQ	2	-0.849558	-0.990000	-1.232034	-0.810063	-1.026290	-0.732183	-1.219782	-1.199735
NGD	3	-0.690817	-0.903164	-0.471316	-0.222817	-0.587197	-0.391387	-0.214527	-0.457489
PTON	2	-0.339243	-0.267694	-1.751299	-2.048703	-1.591125	-1.276625	-1.153055	-1.071214
WULF	2	-0.832535	-0.954948	-0.785905	-0.450545	-0.618441	-0.474508	-0.887969	-1.119646
HOOD	3	-0.427648	-0.104652	0.030048	0.638733	0.152983	-0.270861	-2.315094	-2.840700
RIG	2	-0.267051	0.416354	-0.789683	-1.401024	-0.528638	-0.574253	-0.779566	-1.909392
KGC	5	0.008468	-0.237333	1.457431	1.286732	1.515993	-0.233457	3.009188	1.510319
GRAB	4	-0.361497	-0.362119	-1.180746	-0.966897	-0.819089	-0.570097	0.232944	0.780683
BTG	3	-0.483881	-0.530658	0.940705	-0.735619	-0.178833	-0.449571	1.428300	-0.388274
RELY	2	-0.648674	-0.618868	-0.994287	-0.618469	-0.774582	-0.628281	-1.021636	-0.854357
CFLT	2	-0.690035	-0.620327	-1.768553	-1.466887	-1.527608	-0.890112	-1.037027	-0.901813
DAVE	2	-0.655508	-0.606508	-0.700000	-0.300000	-0.055431	-0.345508	-0.410000	-0.147000

Showing 1 to 17 of 81 entries. 9 total columns

Figura 42. Data-frame cu observațiile și clasele de care aparțin

După pregătirea acestui `data.frame`, am generat o reprezentare grafică a clusterelor folosind funcția `fviz_cluster`, care afișează vizual apartenența fiecărei observații la clusterelor respective.

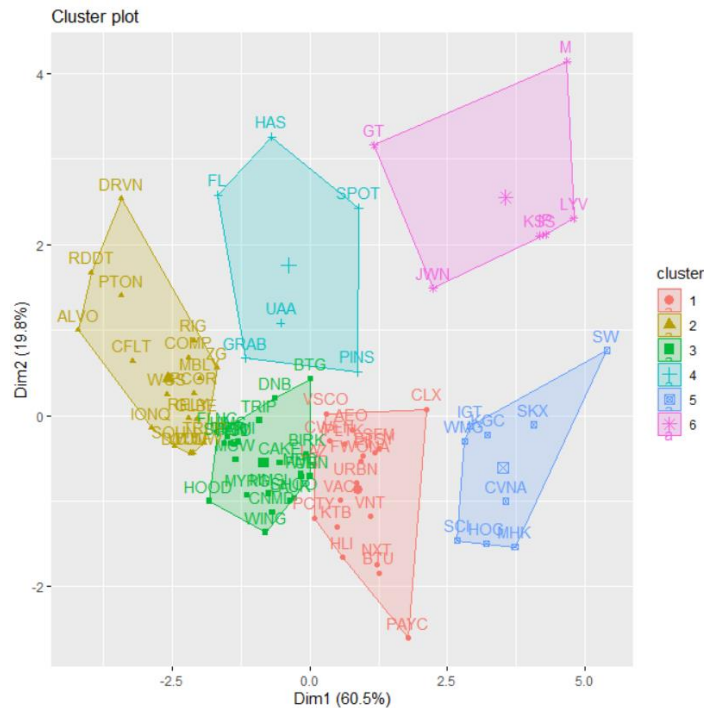


Figura 43. Reprezentarea grafica a clusterelor

Având în vedere că avem mai mult de două variabile, acestea au fost proiectate pe două dimensiuni: Dimensiunea 1, care explică 60.5% din variabilitatea datelor, și Dimensiunea 2, care explică 19.8%.

Analizând graficul, se observă o dispersie semnificativă a clusterelor, majoritatea observațiilor fiind concentrate în clusterelor 1 și 3, care au valori moderate pentru Dimensiunea 1 (companii cu performanțe financiare medii, moderate) și relativ scăzute pentru Dimensiunea 2. De asemenea, clusterul 2 conține multe observații, cu valori scăzute pe Dimensiunea 1, dar destul de mari pe Dimensiunea 2, reprezentând asadar companii cu performanțe slabe. În clusterelor 4, 5 și 6, numărul observațiilor este redus, acestea distingându-se de celelalte prin performanțe financiare superioare, acestea reprezintă companiile ce au cele mai ridicate valori de pe piața bursieră, având valori ridicate pentru Dimensiunea 1, care reflectă performanța financiară conform analizei ACP.

Indicatorii statistici per cluster

Pentru a genera indicatorii statistici pentru fiecare cluster, mai intai am împărțit datele originale în șase subseturi, câte unul pentru fiecare cluster, utilizând funcția subset(). Astfel, am creat variabile pentru fiecare cluster, pe baza valorii corespunzătoare din coloana „clasa”.

	clasa	X1	X2	X3	X4	X5	X6	X7	X10
PAYC	1	-0.515538	-0.106385	0.730899	1.220122	0.856134	3.041508	0.143043	0.244792
NXT	1	-0.331630	-0.431498	0.953390	1.013845	1.081008	1.171287	-0.281018	0.272135
ETSY	1	-0.325481	0.187777	0.123484	0.619980	0.144090	0.493851	0.634606	1.528580
KTB	1	-0.365619	-0.327884	0.069144	0.477781	0.181183	1.358309	-0.116754	0.455817
SFM	1	0.518948	0.631001	0.384733	0.749662	0.468061	0.921924	0.071081	0.040426
BTU	1	-0.025869	-0.365823	0.697112	1.471555	1.345002	1.208691	-0.060751	-0.712426
VNT	1	-0.280702	-0.127363	0.632186	1.001676	0.802281	0.680873	-0.035047	0.409164
PCTY	1	-0.587382	-0.416900	-0.082526	0.373042	0.047796	1.025825	-0.145529	0.241046
LAZ	1	-0.316452	-0.458417	-0.170989	0.197156	0.065392	0.273581	0.025835	0.663334
HLI	1	-0.470821	-0.555849	0.230799	0.695716	0.340563	1.474678	-0.116819	0.340378
CLX	1	0.500564	0.850525	1.506565	0.606812	0.581632	0.535412	0.548096	0.834283
VSCO	1	0.315495	0.369597	-0.030116	0.153533	0.096767	0.327609	-0.095619	-0.127176
CWEN	1	-0.616732	-0.539431	0.804655	0.000312	-0.062512	-0.083839	0.715730	0.499280
FWONA	1	-0.149662	-0.308683	0.275716	0.603620	0.705776	0.044998	0.586093	0.643331
PLTK	1	-0.368724	0.120935	0.512360	0.398687	0.620281	-0.158648	0.041170	0.431944
PII	1	0.618373	-0.037371	0.186724	0.362935	0.340137	1.084010	0.361910	0.071816

Figura 44. Observatiile din Cluster 1

```
> summary(cluster_1)
```

	clasa	X1	X2	X3	X4	X5
Min. :	1	Min. : -0.6167	Min. : -0.55585	Min. : -0.1710	Min. : 0.000312	Min. : -0.06251
1st Qu. :	1	1st Qu. : -0.3672	1st Qu. : -0.39136	1st Qu. : 0.1551	1st Qu. : 0.367988	1st Qu. : 0.16264
Median :	1	Median : -0.2807	Median : -0.10638	Median : 0.3533	Median : 0.603620	Median : 0.41533
Mean :	1	Mean : -0.1060	Mean : -0.05639	Mean : 0.4190	Mean : 0.598988	Mean : 0.46697
3rd Qu. :	1	3rd Qu. : 0.1707	3rd Qu. : 0.15436	3rd Qu. : 0.6646	3rd Qu. : 0.726143	3rd Qu. : 0.66303
Max. :	1	Max. : 0.6184	Max. : 0.85052	Max. : 1.5066	Max. : 1.471555	Max. : 1.34500

	X6	X7	X10
Min. :	-0.1586	Min. : -0.47335	Min. : -0.71243
1st Qu. :	0.3006	1st Qu. : -0.10619	1st Qu. : 0.05612
Median :	0.9219	Median : 0.02726	Median : 0.27214
Mean :	0.8340	Mean : 0.09904	Mean : 0.30502
3rd Qu. :	1.1900	3rd Qu. : 0.25248	3rd Qu. : 0.47755
Max. :	3.0415	Max. : 0.71573	Max. : 1.52858

Figura 45. Summary pentru Cluster 1

```
> describe(cluster_1)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
clasa	1	19	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	NaN	NaN	0.00
X1	2	19	-0.11	0.39	-0.28	-0.12	0.38	-0.62	0.62	1.24	0.47	-1.21	0.09
X2	3	19	-0.06	0.40	-0.11	-0.08	0.44	-0.56	0.85	1.41	0.61	-0.63	0.09
X3	4	19	0.42	0.41	0.35	0.39	0.41	-0.17	1.51	1.68	0.80	0.33	0.09
X4	5	19	0.60	0.38	0.60	0.58	0.34	0.00	1.47	1.47	0.58	-0.37	0.09
X5	6	19	0.47	0.37	0.42	0.45	0.40	-0.06	1.35	1.41	0.64	-0.38	0.09
X6	7	19	0.83	0.74	0.92	0.76	0.65	-0.16	3.04	3.20	1.08	1.56	0.17
X7	8	19	0.10	0.32	0.03	0.10	0.21	-0.47	0.72	1.19	0.49	-0.79	0.07
X10	9	19	0.31	0.46	0.27	0.29	0.34	-0.71	1.53	2.24	0.37	1.05	0.11

Figura 46. Describe pentru Cluster 1

Clusterul 1, ce contine 19 companii, prezintă o performanță financiară moderată și stabilă. Veniturile totale (X1) sunt ușor sub media globală, iar venitul operațional (X3) arată o tendință pozitivă. Profitul brut (X2) este mai slab, însă venitul net și profitul înainte de dobanzi și taxe (X4, X5) sunt solide. Fluxul de numerar operațional (X7) este modest, însă fluxul de numerar liber (X10) indică o performanță mai echilibrată. În concluzie, clusterul 1 grupează companii cu o performanță financiară moderată, caracterizată prin stabilitate și echilibru, având venituri apropiate de media globală și indicatori financiari care sugerează o gestionare eficientă a resurselor, în special prin fluxul de numerar liber consistent.

	clasa	X1	X2	X3	X4	X5	X6	X7	X10
RDDT	2	-0.667892	-0.472650	-2.331036	-2.119293	-2.053185	-1.841847	-0.926384	-0.707662
DJT	2	-0.854734	-0.997913	-0.774559	-0.472724	-0.605371	-0.578409	-1.027068	-0.816394
SOUN	2	-0.844919	-0.977051	-0.918912	-0.638627	-0.755707	-0.557629	-1.171839	-1.034826
IONQ	2	-0.849558	-0.990000	-1.232034	-0.810063	-1.026290	-0.732183	-1.219782	-1.199735
PTON	2	-0.339243	-0.267694	-1.751299	-2.048703	-1.591125	-1.276625	-1.153055	-1.071214
WULF	2	-0.832535	-0.954948	-0.785905	-0.450545	-0.618441	-0.474508	-0.887969	-1.119646
RIG	2	-0.267051	0.416354	-0.789683	-1.401024	-0.528638	-0.574253	-0.779566	-1.909392
RELY	2	-0.648674	-0.618868	-0.994287	-0.618469	-0.774582	-0.628281	-1.021636	-0.854357
CFLT	2	-0.690035	-0.620327	-1.768553	-1.466887	-1.527608	-0.890112	-1.037027	-0.901813
MBLY	2	-0.502784	-0.466563	-1.363744	-1.002010	-1.034584	-0.516068	-0.408536	-0.177426
COMP	2	0.136449	-0.436323	-1.292300	-1.038081	-1.093377	-0.599189	-0.938483	-0.725156
DJTW	2	-0.854734	-0.997913	-0.684564	-0.472724	-0.605371	-0.578409	-1.027068	-0.816394

Showing 1 to 13 of 19 entries, 9 total columns

Figura 47. Observatiile din Cluster 2

```
> summary(cluster_2)
```

clasa	X1	X2	X3	X4	X5
Min. :2	Min. :-0.8547	Min. :-0.9979	Min. :-2.331036	Min. :-2.7205	Min. :-2.2648
1st Qu.:2	1st Qu. :-0.8185	1st Qu. :-0.9325	1st Qu. :-1.290420	1st Qu. :-1.4340	1st Qu. :-1.3105
Median :2	Median :-0.6679	Median :-0.6203	Median :-1.012123	Median :-0.7371	Median :-0.8208
Mean :2	Mean :-0.6164	Mean :-0.6178	Mean :-1.095681	Mean :-1.0800	Mean :-1.0514
3rd Qu.:2	3rd Qu. :-0.4812	3rd Qu. :-0.4514	3rd Qu. :-0.787794	3rd Qu. :-0.5860	3rd Qu. :-0.6575
Max. :2	Max. : 0.1364	Max. : 0.4164	Max. : 0.009592	Max. :-0.3473	Max. :-0.5286

X6	X7	X10
Min. :-2.3614	Min. :-1.6986	Min. :-1.9328
1st Qu. :-1.0834	1st Qu. :-1.0828	1st Qu. :-1.0954
Median :-0.6532	Median :-0.9385	Median :-0.8544
Mean :-0.9216	Mean :-0.9269	Mean :-0.9287
3rd Qu. :-0.5784	3rd Qu. :-0.7472	3rd Qu. :-0.7057
Max. :-0.4745	Max. :-0.3437	Max. :-0.1774

Figura 48. Summary pentru Cluster 2

```
> describe(cluster_2)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
clasa	1	19	2.00	0.00	2.00	2.00	0.00	2.00	2.00	0.00	NaN	NaN	0.00
X1	2	19	-0.62	0.26	-0.67	-0.65	0.24	-0.85	0.14	0.99	1.30	1.25	0.06
X2	3	19	-0.62	0.39	-0.62	-0.66	0.43	-1.00	0.42	1.41	0.97	0.25	0.09
X3	4	19	-1.10	0.50	-1.01	-1.09	0.35	-2.33	0.01	2.34	-0.40	0.67	0.11
X4	5	19	-1.08	0.72	-0.74	-1.03	0.39	-2.72	-0.35	2.37	-0.94	-0.56	0.16
X5	6	19	-1.05	0.53	-0.82	-1.01	0.32	-2.26	-0.53	1.74	-0.99	-0.40	0.12
X6	7	19	-0.92	0.54	-0.65	-0.86	0.14	-2.36	-0.47	1.89	-1.31	0.43	0.12
X7	8	19	-0.93	0.31	-0.94	-0.92	0.28	-1.70	-0.34	1.35	-0.14	0.20	0.07
X10	9	19	-0.93	0.47	-0.85	-0.91	0.32	-1.93	-0.18	1.76	-0.63	-0.19	0.11

Figura 49. Describe pentru Cluster 2

Clusterul 2 cuprinde, de asemenea, 19 companii, și se caracterizează prin indicatori economici negativi pe majoritatea variabilelor analizate. În special, rentabilitatea și fluxurile financiare sunt mult sub media globală. Indicatorii sunt în mare parte în teritoriul negativ sau apropiat de zero, ceea ce sugerează o performanță economică slabă pentru companiile din acest cluster, cu o tendință de scădere a veniturilor și profitabilității. De asemenea, companiile din acest grup ar putea întâmpina dificultăți financiare, având valori sub medie la indicatorii de rentabilitate și flux de numerar, ceea ce indică o performanță financiară mai puțin favorabilă în comparație cu alte grupuri de companii.

SCRIPT_AD.R

tema3.R

cluster_3

Filter

	clasa	X1	X2	X3	X4	X5	X6	X7	X10
HIMS	3	-0.651502	-0.467733	-0.704503	-0.229191	-0.533372	-0.366450	-0.722334	-0.483941
NGD	3	-0.690817	-0.903164	-0.471316	-0.222817	-0.587197	-0.391387	-0.214527	-0.457489
HOOD	3	-0.427648	-0.104652	0.030048	0.638733	0.152983	-0.270861	-2.315094	-2.840700
BTG	3	-0.483881	-0.530658	0.940705	-0.735619	-0.178833	-0.449571	1.428300	-0.388274
TTMI	3	-0.412569	-0.727319	-0.386373	-0.232425	-0.343024	-0.333202	-0.576009	-0.705103
MCW	3	-0.672283	-0.592058	-0.273889	-0.061647	-0.137975	-0.312422	-0.543955	-1.309532
LAUR	3	-0.560029	-0.756130	0.137543	0.336757	0.286607	0.119807	-0.456873	-0.190253
DNB	3	-0.404400	-0.086800	-0.281550	-0.393914	-0.163701	-0.432947	-0.037059	-0.057495
MYRG	3	-0.169522	-0.807850	-0.563760	-0.129875	-0.390934	0.813867	-0.814114	-0.795252
PEN	3	-0.638779	-0.569988	-0.549143	-0.241542	-0.585232	-0.254237	-0.719193	-0.413020
CNMD	3	-0.611571	-0.571658	-0.326906	0.027742	-0.180550	0.905300	-0.614281	-0.250876
MMSI	3	-0.606741	-0.632615	-0.356266	0.087868	-0.169269	0.435667	-0.517963	-0.190629

Showing 1 to 13 of 22 entries. 9 total columns

Figura 50. Observatiile din Cluster 3

```
> summary(cluster_3)
```

clasa	X1	X2	X3	X4	X5
Min. :3	Min. : -0.7425	Min. : -0.90316	Min. : -0.80672	Min. : -0.73562	Min. : -0.62888
1st Qu.:3	1st Qu.: -0.6210	1st Qu.: -0.71264	1st Qu.: -0.45008	1st Qu.: -0.23162	1st Qu.: -0.37896
Median :3	Median : -0.5085	Median : -0.55083	Median : -0.30423	Median : -0.01695	Median : -0.17969
Mean :3	Mean : -0.5042	Mean : -0.50077	Mean : -0.22305	Mean : -0.03551	Mean : -0.16841
3rd Qu.:3	3rd Qu.: -0.4288	3rd Qu.: -0.39202	3rd Qu.: 0.02785	3rd Qu.: 0.10752	3rd Qu.: 0.09982
Max. :3	Max. : -0.1695	Max. : 0.00465	Max. : 0.94071	Max. : 0.63873	Max. : 0.35743

X6	X7	X10
Min. : -0.47035	Min. : -2.3151	Min. : -2.8407
1st Qu.: -0.33216	1st Qu.: -0.6266	1st Qu.: -0.5727
Median : -0.20436	Median : -0.5385	Median : -0.4006
Mean : 0.05388	Mean : -0.4797	Mean : -0.4769
3rd Qu.: 0.55723	3rd Qu.: -0.2932	3rd Qu.: -0.1903
Max. : 1.02167	Max. : 1.4283	Max. : 0.3090

Figura 51. Summary pentru Cluster 3

```
> describe(cluster_3)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
clasa	1	22	3.00	0.00	3.00	3.00	0.00	3.00	3.00	0.00	NaN	NaN	0.00
X1	2	22	-0.50	0.15	-0.51	-0.52	0.15	-0.74	-0.17	0.57	0.56	-0.32	0.03
X2	3	22	-0.50	0.28	-0.55	-0.52	0.26	-0.90	0.00	0.91	0.49	-1.03	0.06
X3	4	22	-0.22	0.39	-0.30	-0.25	0.37	-0.81	0.94	1.75	1.03	1.32	0.08
X4	5	22	-0.04	0.31	-0.02	-0.03	0.32	-0.74	0.64	1.37	0.02	-0.20	0.07
X5	6	22	-0.17	0.31	-0.18	-0.17	0.39	-0.63	0.36	0.99	0.12	-1.21	0.07
X6	7	22	0.05	0.50	-0.20	0.01	0.35	-0.47	1.02	1.49	0.61	-1.26	0.11
X7	8	22	-0.48	0.62	-0.54	-0.49	0.27	-2.32	1.43	3.74	0.17	4.85	0.13
X10	9	22	-0.48	0.63	-0.40	-0.38	0.31	-2.84	0.31	3.15	-2.29	6.22	0.13

Figura 52. Describe pentru Cluster 3

Clusterul 3 include 22 de companii și se caracterizează prin valori medii negative pentru majoritatea variabilelor economice, dar cu o tendință de îmbunătățire în comparație cu Clusterul 2. Venitul total (X1) și profitul brut (X2) prezintă medii negative, de aproximativ -0,50, în timp ce venitul operațional (X3) și venitul net (X4) rămân sub zero, dar cu valori mai apropiate de pragul pozitiv, având o medie de -0,22, respectiv -0,04. Profitul înainte de dobânzi și taxe (X5) înregistrează o valoare medie negativă de -0,17, dar mai aproape de zero decât în alte grupuri. Pe de altă parte, venitul pe acțiuni (X6) are o medie pozitivă de 0,05, sugerând o performanță mai bună pentru unele companii din acest cluster. Fluxul de numerar operațional (X7) și fluxul de numerar liber (X10) rămân în teritoriu negativ, însă variabilitatea mare între valori indică faptul că există companii cu o performanță financiară relativ mai bună. În ansamblu, Clusterul 3 reflectă o performanță financiară moderat slabă, dar cu semne clare de îmbunătățire și o diversitate semnificativă în performanțele economice ale companiilor analizate.

	clasa	X1	X2	X3	X4	X5	X6	X7	X10
SPOT	4	1.790693	1.206971	-1.040365	-0.638110	-0.741792	-0.649062	0.854307	1.973293
GRAB	4	-0.361497	-0.362119	-1.180746	-0.966897	-0.819089	-0.570097	0.232944	0.780683
PINS	4	-0.217318	0.592504	-0.445261	0.325862	-0.291139	-0.279173	0.848400	1.920353
HAS	4	-0.029062	0.573202	0.473755	-2.333758	-1.772422	-2.328105	1.181306	1.839962
UAA	4	0.209115	0.567210	-0.903277	-0.553181	-0.719104	-0.482820	0.141197	0.387024
FL	4	0.703031	0.368990	-0.386085	-1.452098	-1.662332	-2.012245	-0.109030	-0.341578

Figura 53. Observatiile din Cluster 4

```
> summary(cluster_4)
```

clasa	X1	X2	X3	X4	X5
Min. :4	Min. : -0.36150	Min. : -0.3621	Min. : -1.1807	Min. : -2.3338	Min. : -1.7724
1st Qu.:4	1st Qu.: -0.17025	1st Qu.: 0.4185	1st Qu.: -1.0061	1st Qu.: -1.3308	1st Qu.: -1.4515
Median :4	Median : 0.09003	Median : 0.5702	Median : -0.6743	Median : -0.8025	Median : -0.7804
Mean :4	Mean : 0.34916	Mean : 0.4911	Mean : -0.5803	Mean : -0.9364	Mean : -1.0010
3rd Qu.:4	3rd Qu.: 0.57955	3rd Qu.: 0.5877	3rd Qu.: -0.4009	3rd Qu.: -0.5744	3rd Qu.: -0.7248
Max. :4	Max. : 1.79069	Max. : 1.2070	Max. : 0.4738	Max. : 0.3259	Max. : -0.2911

X6	X7	X10
Min. : -2.3281	Min. : -0.1090	Min. : -0.3416
1st Qu.: -1.6714	1st Qu.: 0.1641	1st Qu.: 0.4854
Median : -0.6096	Median : 0.5407	Median : 1.3103
Mean : -1.0536	Mean : 0.5249	Mean : 1.0933
3rd Qu.: -0.5046	3rd Qu.: 0.8528	3rd Qu.: 1.9003
Max. : -0.2792	Max. : 1.1813	Max. : 1.9733

Figura 54. Summary pentru Cluster 4


```
> describe(cluster_4)
vars n mean sd median trimmed mad min max range skew kurtosis se
clasa 1 6 4.00 0.00 4.00 4.00 0.00 4.00 0.00 0.00 0.00 NaN NaN 0.00
X1 2 6 0.35 0.80 0.09 0.35 0.56 -0.36 1.79 2.15 0.80 -1.07 0.33
X2 3 6 0.49 0.51 0.57 0.49 0.17 -0.36 1.21 1.57 -0.33 -0.98 0.21
X3 4 6 -0.58 0.61 -0.67 -0.58 0.49 -1.18 0.47 1.65 0.62 -1.26 0.25
X4 5 6 -0.94 0.90 -0.80 -0.94 0.67 -2.33 0.33 2.66 -0.18 -1.36 0.37
X5 6 6 -1.00 0.59 -0.78 -1.00 0.41 -1.77 -0.29 1.48 -0.29 -1.85 0.24
X6 7 6 -1.05 0.88 -0.61 -1.05 0.34 -2.33 -0.28 2.05 -0.52 -1.88 0.36
X7 8 6 0.52 0.51 0.54 0.52 0.53 -0.11 1.18 1.29 0.02 -1.98 0.21
X10 9 6 1.09 0.97 1.31 1.09 0.94 -0.34 1.97 2.31 -0.31 -1.88 0.39
```

Figura 55. Describe pentru Cluster 4

Clusterul 4, ce cuprinde 6 companii, prezinta o performanță financiară mixta. Venitul total (X1) este pozitiv cu o medie de 0.35, dar cu o variabilitate ridicată (de la -0.36 la 1.79), indicând performanțe destul de diverse între companii. Profitul brut (X2) are, de asemenea, o medie pozitivă (0.49), cu o dispersie similară (între -0.36 și 1.21), indicând o ușoară tendință de creștere a acestui indicator, dar cu fluctuații semnificative. Venitul operational (X3) și venitul net (X4) sunt negative, cu medii de -0.58 și respectiv -0.94, ceea ce sugerează o performanță mai slabă în aceste domenii. Profitul înainte de dobanzi și taxe (X5) este, de asemenea, negativ (-1.00), iar venitul pe acțiune (X6) rămâne sub zero cu o medie de -1.05, semnalând dificultăți financiare în aceste aspecte. Fluxul de numerar operational (X7) și fluxul de numerar liber (X10) sunt pozitive, cu medii de 0.52 și respectiv 1.09, semnificând o performanță mai bună în aceste domenii. În ansamblu, companiile din Clusterul 4 au o performanță mixtă, cu indicatori economici negativi semnificativi în unele domenii, dar și semne de creștere în altele, indicând o situație financiară mai instabilă comparativ cu celelalte cluster.

	clasa	X1	X2	X3	X4	X5	X6	X7	X10
KGC	5	0.008468	-0.237333	1.457431	1.286732	1.515993	-0.233457	3.009188	1.510319
CVNA	5	1.375244	0.329519	0.295771	1.992188	2.551542	0.913612	0.816310	1.688541
SCI	5	-0.064574	-0.348533	1.558614	1.340464	1.604660	1.021669	1.056362	1.098051
SW	5	1.526677	1.102817	3.153202	2.387626	2.631053	-1.156099	2.742741	1.652873
WMG	5	0.364630	0.843239	1.659481	1.436761	1.591182	0.032530	0.755962	0.834283
SKX	5	0.749514	1.719702	1.393269	1.562916	1.426761	1.150507	1.565829	1.911328
MHK	5	1.211591	0.644158	1.124616	1.513320	1.181518	3.282559	1.304079	0.888889
HOG	5	0.282185	0.289073	1.000987	1.876209	1.378165	1.653388	1.055488	1.605782
IGT	5	-0.031318	0.271832	1.920190	0.383365	1.708299	0.032530	1.582957	1.785692

Figura 56. Observatiile din Cluster 5


```
> summary(cluster_5)
      clasa      X1      X2      X3      X4      X5
Min.   :5      Min.   :-0.064574      Min.   :-0.3485      Min.   :0.2958      Min.   :0.3834      Min.   :1.182
1st Qu.:5      1st Qu.: 0.008468      1st Qu.: 0.2718      1st Qu.:1.1246      1st Qu.:1.3405      1st Qu.:1.427
Median :5      Median : 0.364630      Median : 0.3295      Median :1.4574      Median :1.5133      Median :1.591
Mean   :5      Mean   : 0.602491      Mean   : 0.5127      Mean   :1.5071      Mean   :1.5311      Mean   :1.732
3rd Qu.:5      3rd Qu.: 1.211591      3rd Qu.: 0.8432      3rd Qu.:1.6595      3rd Qu.:1.8762      3rd Qu.:1.708
Max.   :5      Max.   : 1.526677      Max.   : 1.7197      Max.   :3.1532      Max.   :2.3876      Max.   :2.631

      X6      X7      X10
Min.   :-1.15610      Min.   :0.756      Min.   :0.8343
1st Qu.: 0.03253      1st Qu.:1.055      1st Qu.:1.0981
Median : 0.91361      Median :1.304      Median :1.6058
Mean   : 0.74414      Mean   :1.543      Mean   :1.4418
3rd Qu.: 1.15051      3rd Qu.:1.583      3rd Qu.:1.6885
Max.   : 3.28256      Max.   :3.009      Max.   :1.9113
```

Figura 57. Summary pentru Cluster 5

```
> describe(cluster_5)
      vars  n mean  sd median trimmed mad  min max range skew kurtosis  se
clasa   1  9 5.00 0.00  5.00  5.00 0.00  5.00 5.00 0.00  NaN  NaN 0.00
X1      2  9 0.60 0.63  0.36  0.60 0.59 -0.06 1.53 1.59 0.30 -1.81 0.21
X2      3  9 0.51 0.65  0.33  0.51 0.76 -0.35 1.72 2.07 0.37 -1.05 0.22
X3      4  9 1.51 0.77  1.46  1.51 0.49  0.30 3.15 2.86 0.62 -0.02 0.26
X4      5  9 1.53 0.56  1.51  1.53 0.34  0.38 2.39 2.00 -0.49 -0.32 0.19
X5      6  9 1.73 0.51  1.59  1.73 0.24  1.18 2.63 1.45 0.85 -1.00 0.17
X6      7  9 0.74 1.28  0.91  0.74 1.31 -1.16 3.28 4.44 0.46 -0.66 0.43
X7      8  9 1.54 0.81  1.30  1.54 0.41  0.76 3.01 2.25 0.78 -1.09 0.27
X10     9  9 1.44 0.40  1.61  1.44 0.27  0.83 1.91 1.08 -0.46 -1.62 0.13
```

Figura 58. Describe pentru Cluster 5

Clusterul 5 include 9 companii, cu performanțe financiare bune. Venitul total (X1) are o medie de 0.60, iar profitul brut (X2) este de 0.51. Indicatorii venitului operational (X3) și venitului net (X4) sunt ridicați (1.51 și 1.53), iar EBIT (X5) are o medie de 1.73. Venitul pe actiune (X6) și fluxul de numerar operational (X7) sunt pozitive, cu variații mai mari, iar fluxul de numerar liber (X10) are o medie de 1.44. În general, companiile din acest cluster au o performanță financiară solidă, cu creștere constantă, dar cu anumite fluctuații între ele.

clasa	X1	X2	X3	X4	X5	X6	X7	X10
IP	6 2.694242	2.130573	0.556480	1.101589	0.909560	0.111495	2.839095	1.688541
LYV	6 3.697381	2.489806	1.666550	0.529196	2.373082	0.057466	1.508972	1.052645
M	6 3.639087	4.857049	2.536868	0.300370	0.279470	-0.129556	1.612013	0.254728
JWN	6 2.006352	2.320637	0.423618	0.619581	0.656587	0.319297	0.523509	-0.328178
GT	6 2.846809	1.299271	1.566728	-1.410600	0.010101	-0.915049	1.113581	-1.453788
KSS	6 2.418359	3.167727	0.932503	0.619581	0.996227	0.660093	1.647775	1.591390

Figura 59. Observatiile din Cluster 6

```
> summary(cluster_6)
      clasa      X1      X2      X3      X4      X5      X6
Min.   :6   Min.   :2.006   Min.   :1.299   Min.   :0.4236   Min.   :-1.4106   Min.   :0.0101   Min.   :-0.91505
1st Qu.:6   1st Qu.:2.487   1st Qu.:2.178   1st Qu.:0.6505   1st Qu.:0.3576   1st Qu.:0.3737   1st Qu.:-0.08280
Median :6   Median :2.771   Median :2.405   Median :1.2496   Median :0.5744   Median :0.7831   Median :0.08448
Mean   :6   Mean   :2.884   Mean   :2.711   Mean   :1.2805   Mean   :0.2933   Mean   :0.8708   Mean   :0.01729
3rd Qu.:6   3rd Qu.:3.441   3rd Qu.:2.998   3rd Qu.:1.6416   3rd Qu.:0.6196   3rd Qu.:0.9746   3rd Qu.:0.26735
Max.   :6   Max.   :3.697   Max.   :4.857   Max.   :2.5369   Max.   :1.1016   Max.   :2.3731   Max.   :0.66009

      X7      X10
Min.   :0.5235   Min.   :-1.4538
1st Qu.:1.2124   1st Qu.: -0.1825
Median :1.5605   Median :0.6537
Mean   :1.5408   Mean   :0.4676
3rd Qu.:1.6388   3rd Qu.:1.4567
Max.   :2.8301   Max.   :1.6885
```

Figura 60. Summary pentru Cluster 6

```
> describe(cluster_6)
      vars  n mean  sd median trimmed mad  min max range skew kurtosis  se
clasa    1  6 6.00 0.00   6.00   6.00 0.00  6.00 6.00  0.00   NaN      NaN 0.00
X1       2  6 2.88 0.67   2.77   2.88 0.83  2.01 3.70  1.69  0.10   -1.85 0.27
X2       3  6 2.71 1.21   2.41   2.71 0.77  1.30 4.86  3.56  0.65   -1.04 0.49
X3       4  6 1.28 0.80   1.25   1.28 0.82  0.42 2.54  2.11  0.33   -1.63 0.33
X4       5  6 0.29 0.87   0.57   0.29 0.24 -1.41 1.10  2.51 -1.08   -0.47 0.36
X5       6  6 0.87 0.83   0.78   0.87 0.53  0.01 2.37  2.36  0.75   -0.93 0.34
X6       7  6 0.02 0.53   0.08   0.02 0.33 -0.92 0.66  1.58 -0.58   -1.03 0.22
X7       8  6 1.54 0.76   1.56   1.54 0.40  0.52 2.84  2.32  0.40   -1.07 0.31
X10      9  6 0.47 1.22   0.65   0.47 1.42 -1.45 1.69  3.14 -0.38   -1.66 0.50
```

Figura 61. Describe pentru Cluster 6

Clusterul 6 include 6 companii, care prezintă o performanță financiară variabilă. Indicatorul Venit total (X1) are o medie de 2.88 și variază între 2.01 și 3.70, semnificând fluctuații semnificative, dar și o creștere constantă. Profitul brut (X2) are o medie de 2.71, cu o dispersie mai mare (1.21), indicând o performanță financiară variabilă între companii. Venitul operational (X3) este relativ stabil, având o medie de 1.28 și o varianță de 0.80. Venitul net (X4) este scăzut, cu o medie de 0.29 și fluctuații mari (0.87), ceea ce sugerează o performanță financiară mai slabă în comparație cu alți indicatori. EBIT (X5) se situează la o medie de 0.87 și o varianță de 0.83, indicând performanțe moderate. EPS (X6) are o medie aproape de zero (0.02), dar cu o dispersie mare (0.53), reflectând instabilitate în fluxurile de numerar ale companiilor din acest cluster. Fluxul de numerar operational (X7) prezintă o medie de 1.54 și o varianță de 0.76, iar valorile sale variază între 0.52 și 2.84, semnificând o performanță financiară mai consistentă, dar cu unele fluctuații. În final, fluxul de numerar liber (X10) are o medie de 0.47 și o variabilitate mare (1.22), variind de la -1.45 la 1.69. Per ansamblu, companiile din acest cluster demonstrează o performanță financiară instabilă, cu unele fluctuații notabile în profitabilitate și fluxuri de numerar.

Concluziile analizei realizate

Nr. cluster	Nr. de observatii	Caracteristici principale
1	19	<p>Performanță financiară moderată și stabilă - Veniturile totale (X1) sunt ușor sub media globală, iar fluxul de numerar operațional (X7) este modest, sugerând o activitate financiară echilibrată.</p> <p>Indicatori pozitivi de profitabilitate - Venitul operațional (X3) arată o tendință de creștere, iar profitul net și cel înainte de dobânzi și taxe (X4, X5) sunt solide, ceea ce reflectă o profitabilitate robustă.</p> <p>Flux de numerar liber echilibrat - Fluxul de numerar liber (X10) indică o gestionare eficientă a resurselor, ceea ce contribuie la stabilitatea generală a clusterului.</p>
2	19	<p>Performanță financiară slabă - Indicatorii economici sunt în mare parte negativi sau aproape de zero, ceea ce sugerează o performanță financiară slabă pentru companiile din acest cluster, cu o tendință de scădere a veniturilor și profitabilității.</p> <p>Rentabilitate și fluxuri financiare sub media globală - Acestea sunt mult sub media globală, indicând o capacitate redusă de a genera profituri și de a gestiona eficient resursele financiare.</p>

		<p>Dificultăți financiare - Valori sub medie pentru indicatorii de rentabilitate și flux de numerar sugerează că aceste companii ar putea întâmpina dificultăți financiare și ar avea o performanță mai puțin favorabilă comparativ cu alte grupuri de companii.</p>
3	22	<p>Tendință de îmbunătățire - Deși majoritatea indicatorilor economici sunt negativi, există semne de îmbunătățire comparativ cu Clusterul 2, cu valori mai apropiate de pragul pozitiv pentru venitul operațional și venitul net.</p> <p>Performanță financiară moderat slabă, dar diversificată - Indicatorii financiare sunt, în general, sub zero, dar cu o mare variabilitate între companii, indicând că, deși majoritatea companiilor au o performanță slabă, unele dintre ele au performanțe financiare relativ mai bune, reflectate în valoarea pozitivă a veniturii pe acțiune (X6).</p> <p>Fluxuri financiare sub pragul zero - Atât fluxul de numerar operațional cât și fluxul de numerar liber sunt negative, sugerând dificultăți financiare, dar cu o dispersie semnificativă între companiile din acest cluster, unele având performanțe mai bune decât altele.</p>
4	6	<p>Performanță financiară mixtă - Clusterul 4 prezintă o performanță financiară mixtă,</p>

		<p>cu indicatori economici negativi semnificativi în domenii precum venitul operațional și venitul net, dar și valori pozitive în alte domenii, cum ar fi venitul total și profitul brut, care sugerează o ușoară tendință de creștere.</p> <p>Variabilitate ridicată între companii - Există o dispersie mare între companiile din acest cluster, cu valori ale veniturilor și profitului brut variind semnificativ (de la valori negative la valori mari pozitive), ceea ce indică performanțe economice foarte diferite între acestea.</p> <p>Fluxuri financiare pozitive - În ciuda performanței mai slabe în anumite domenii, fluxul de numerar operațional și fluxul de numerar liber sunt pozitive, indicând că unele companii din acest cluster au o performanță financiară mai bună în gestionarea resurselor financiare.</p>
5	9	<p>Performanță financiară excelentă – Acest cluster include companii cu performanțe financiare foarte bune, având valori ridicate pentru toți indicatorii principali, cum ar fi venitul operațional (X3), venitul net (X4) și EBIT (X5), ceea ce sugerează o creștere constantă a veniturilor și profitabilității.</p> <p>Stabilitate financiară cu fluctuații moderate - Deși majoritatea indicatorilor sunt</p>

		<p>pozitivi și indică o performanță solidă, există unele fluctuații între companii, reflectând o diversitate în performanțele financiare ale acestui grup.</p> <p>Fluxuri financiare puternice - Atât fluxul de numerar operațional (X7), cât și fluxul de numerar liber (X10) sunt pozitive și indică o gestionare eficientă a resurselor financiare, cu un flux consistent de numerar, ce susține creșterea companiilor din acest cluster.</p>
6	6	<p>Fluctuații semnificative în venituri și profit - Clusterul 6 prezintă fluctuații mari în indicatorii de venit, cum ar fi venitul total (X1) și profitul brut (X2), indicând o performanță financiară variabilă între companiile din acest grup, dar cu o tendință generală de creștere.</p> <p>Instabilitate în fluxurile financiare - Indicatorii precum venitul net (X4), EBIT (X5) și EPS (X6) reflectă o performanță mai slabă și o instabilitate semnificativă, cu fluctuații mari care sugerează o gestionare mai puțin eficientă a resurselor financiare în comparație cu alte cluster.</p> <p>Performanță financiară relativ constantă, dar cu fluctuații - Deși fluxul de numerar operațional (X7) și fluxul de numerar liber (X10) sunt relativ consistenti,</p>

		variabilitatea mare a acestora sugerează că, deși există o performanță acceptabilă, companiile din acest cluster se confruntă cu o instabilitate notabilă în fluxurile lor financiare.
--	--	--

În urma caracteristicilor relevate din analiza, putem denumi clusterelor sugestiv după cum urmează:

- Clusterul 1 – Companii cu performanță financiară stabilă
- Clusterul 2 – Companii cu performanță financiară slabă
- Clusterul 3 – Companii în recuperare
- Clusterul 4 – Companii cu performanță financiară mixtă
- Clusterul 5 – Companii cu performanță financiară solidă
- Clusterul 6 – Companii instabile/fluctuante

În urma analizei clusterelor financiare, am identificat o diversitate semnificativă în performanțele economice ale companiilor analizate, oferind o perspectivă detaliată asupra diferențelor dintre acestea.

Spre deosebire de analiza anterioară, în care Clusterul 1 reprezenta companii cu performanță slabă, de această dată acesta grupează companii cu o performanță echilibrată și o gestionare eficientă a resurselor, având venituri apropiate de media globală. În schimb, Clusterul 2 ("*Companii cu performanță financiară slabă*") include firme cu indicatori sub media globală și o tendință de scădere a veniturilor și profitabilității, corespunzând astfel Clusterului 1 din analiza precedentă. Clusterul 3 ("*Companii în recuperare*") rămâne similar cu analiza anterioară, evidențiind companii aflate în dificultate, dar cu semne clare de îmbunătățire. Clusterul 4 ("*Companii cu performanță financiară mixtă*") reflectă o situație financiară instabilă, cu o combinație de performanțe pozitive și negative. Clusterul 5 ("*Companii cu performanță financiară solidă*") include firme caracterizate prin creștere constantă și stabilitate, confirmând poziția Clusterului 2 din analiza precedentă. În același timp, Clusterul 6 ("*Companii instabile/fluctuante*") evidențiază firme cu fluctuații semnificative în fluxurile de numerar și profitabilitate, dar și semne de creștere.

Această analiză subliniază diferențele majore în performanțele financiare ale companiilor și oferă o bază solidă pentru înțelegerea comportamentului economic și strategic al firmelor. Rezultatele pot ghida luarea deciziilor informate, atât pentru investitori, cât și pentru management, facilitând dezvoltarea de strategii personalizate pentru îmbunătățirea performanței financiare, adaptate specificului fiecărui cluster.

Iordan Maria-Alexandra
Grupa 1080-A

ANEXA

tema <- Date_AD

View(tema)

Eliminarea outlierilor:

Iterăm prin coloanele setului de date (excluzând prima coloană)

for (col in colnames(tema)[-1]) { # Excludem prima coloană folosind [-1]

 # Identificăm outlierii pentru fiecare coloană

 outliers <- boxplot(tema[[col]], plot = F)\$out

 # Excludem liniile care conțin outlieri

 tema <- tema[-which(tema[[col]] %in% outliers),]

}

View(tema)

Calcularea indicatorilor statistici

summary(tema)

install.packages("psych")

library(psych)

Iordan Maria-Alexandra

Grupa 1080-A

describe(tema[-1])

Matricea de corelatie si matricea de covarianta

cor(tema[-1])

cov(tema[-1])

#Pentru a observa mai bine rezultatele, vom standardiza datele

tema_std = scale(tema[-1], scale = T)

View(tema_std)

Recalculam corelatia si covarianta

matrice_corelatie <- cor(tema_std)

matrice_covarianta <- cov(tema_std)

View(matrice_corelatie)

View(matrice_covarianta)

Reprezentarea grafica a matricei de corelatie

install.packages("corrplot")

library(corrplot)

windows()

Iordan Maria-Alexandra

Grupa 1080-A

```
corrplot(matrice_corelatie, method = "circle", type = "upper", col = "pink", title = "Matricea de  
corelatie" )
```

```
#Analiza cluster prin metode ierarhice
```

```
tema3 = tema
```

```
cor(tema3[-1])
```

```
#Elimin pe X8 si X9
```

```
date_3 = cbind(tema3[,2:8], tema3[,11])
```

```
# Standardizarea datelor
```

```
date_3_std = scale(date_3, scale=TRUE)
```

```
rownames(date_3_std)=tema3$Companie
```

```
View(date_3_std)
```

```
#Distanta dintre forme/observatii (similaritatea/disimilaritatea)
```

```
#Matricea distantelor (de proximitate) prin metoda euclidiană
```

```
dist_euclidiană <- dist(as.matrix(date_3_std),method="euclidian")
```

```
dist_manhattan <- dist(as.matrix(date_3_std),method="manhattan")
```

Iordan Maria-Alexandra

Grupa 1080-A

dist_euclidiană

dist_manhattan

#Calcularea distanței prin met. euclidiană, utilizând alta metodă de calcul:

#Distanța dintre firma SPOT și firma RDDT:

dist_euclidiană[1]

date_3_std[1:2,]

dist_spot_rddt=sqrt((date_3_std[1,1]-date_3_std[2,1])²+

(date_3_std[1,2]-date_3_std[2,2])²+

(date_3_std[1,3]-date_3_std[2,3])²+

(date_3_std[1,4]-date_3_std[2,4])²+

(date_3_std[1,5]-date_3_std[2,5])²+

(date_3_std[1,6]-date_3_std[2,6])²+

(date_3_std[1,7]-date_3_std[2,7])²+

(date_3_std[1,7]-date_3_std[2,8])²)

dist_spot_rddt

#Algoritmul aglomerativ de clusterizare ierarhică - utilizând două metode de calcul a distanței dintre clustere

#Metoda single (distanța dintre cele mai apropiate puncte)

clust_single=hclust(dist_euclidiană,method="single")

Iordan Maria-Alexandra

Grupa 1080-A

```
cbind(clust_single$merge,clust_single$height)
```

```
windows()
```

```
plot(clust_single, main = "Clusterizare ierarhică - Metoda Single")
```

```
#Metoda complete (distanța dintre cele mai îndepărtate puncte)
```

```
clust_complete=hclust(dist_euclidiana,method="complete")
```

```
cbind(clust_complete$merge,clust_complete$height)
```

```
windows()
```

```
plot(clust_complete, main = "Clusterizare ierarhică - Metoda Complete")
```

```
#In continuare, vom folosi rezultatele clusterizarii prin metoda Complete Linkage
```

```
#Criterii de alegere a nr de clustere
```

```
#Criteriul general al clasificarii
```

```
#Dendograma-ne uitam de sus in jos,gasim distanta cea mai mare intre 2 etape succesive si realizam o taietura paralela cu cea orizontala
```

```
windows()
```

```
plot(clust_complete,labels=rownames(date_3_std))
```

```
rect.hclust(clust_complete,k=5,border="red")
```

```
#Grafic elbow
```

```
install.packages("factoextra")
```

Iordan Maria-Alexandra

Grupa 1080-A

```
library(factoextra)
```

```
fviz_nbclust(date_3_std, hcut, method = "wss")+  
  geom_vline(xintercept = 6, linetype = 2)+  
  labs(substitute = "Elbow method - STD")
```

```
#Calculul unor indici
```

```
install.packages("NbClust")
```

```
library(NbClust)
```

```
res <- NbClust(date_3_std, distance = "euclidean", min.nc = 3, max.nc = 10, method =  
"complete", index = "all")
```

```
res
```

```
#Graficul siluetei
```

```
library(cluster)
```

```
si_std <- silhouette(cutree(clust_complete, k = 6), dist_euclidian)
```

```
plot(si_std, cex.names = 0.5)
```

```
si_std
```

```
#Daca s(i) -> 0 => obs. i este intre 2 clase
```

```
#Daca s(i) -> 1 => obs. i este bine incadrata in clasa
```

Iordan Maria-Alexandra

Grupa 1080-A

#Daca $s(i) < 0 \Rightarrow$ obs. i este eronat incadrata

#Calcularea indicatorilor statistici per cluster

```
cutree_result <- cutree(clust_complete, k = 6)
```

```
cluster1 <- date_3_std[cutree_result == 1, ]
```

```
cluster2 <- date_3_std[cutree_result == 2, ]
```

```
cluster3 <- date_3_std[cutree_result == 3, ]
```

```
cluster4 <- date_3_std[cutree_result == 4, ]
```

```
cluster5 <- date_3_std[cutree_result == 5, ]
```

```
cluster6 <- date_3_std[cutree_result == 6, ]
```

```
summary(cluster1)
```

```
summary(cluster2)
```

```
summary(cluster3)
```

```
summary(cluster4)
```

```
summary(cluster5)
```

```
summary(cluster6)
```

```
describe(cluster1)
```

```
describe(cluster2)
```

```
describe(cluster3)
```

```
describe(cluster4)
```

Iordan Maria-Alexandra

Grupa 1080-A

describe(cluster5)

describe(cluster6)

#Algoritmul K-means

k_means = kmeans(date_3_std, 6)

k_means

#Calculul indicatorilor

tss = k_means\$totss #variabilitatea totala/suma patratelor variatiilor

wss = k_means\$tot.withinss #variabilitatea intraclassa/suma patratelor variatiilor intraclassa

bss = k_means\$betweenss #variabilitatea interclassa/suma patratelor variatiilor interclassa

calit_part = bss/tss * 100 #calitatea partitiei

variab = cbind(tss, wss, bss, calit_part)

variab

k_means\$withinss #variabilitatea intraclassa pt fiecare clasa in parte

#suma acestor variabilitati da variabilitatea totala intraclassa (wss)

sum(k_means\$withinss)

#Reprezentare grafica a claselor

Iordan Maria-Alexandra
Grupa 1080-A

```
clasa = k_means$cluster
```

```
c = cbind(clasa, round(date_3_std,6))
```

```
c
```

```
m = data.frame(c)
```

```
m
```

```
library(factoextra)
```

```
windows()
```

```
fviz_cluster(list(data = date_3_std, cluster = clasa))
```

```
#Calcularea indicatorilor statistici per cluster
```

```
cluster_1 <- subset(m, clasa == 1)
```

```
cluster_2 <- subset(m, clasa == 2)
```

```
cluster_3 <- subset(m, clasa == 3)
```

```
cluster_4 <- subset(m, clasa == 4)
```

```
cluster_5 <- subset(m, clasa == 5)
```

```
cluster_6 <- subset(m, clasa == 6)
```


Iordan Maria-Alexandra
Grupa 1080-A

```
summary(cluster_1)
```

```
summary(cluster_2)
```

```
summary(cluster_3)
```

```
summary(cluster_4)
```

```
summary(cluster_5)
```

```
summary(cluster_6)
```

```
library(psych)
```

```
describe(cluster_1)
```

```
describe(cluster_2)
```

```
describe(cluster_3)
```

```
describe(cluster_4)
```

```
describe(cluster_5)
```

```
describe(cluster_6)
```