**ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI**

**Facultatea**: Cibernetică, Statistică și Informatică Economică

**Specializarea**: Informatică economică

***Analiza calității produselor McDonald’s***

****

**Proiect realizat de:**

Crețu Costin-Răzvan

**Profesor coordonator:**

Vințe Claudiu

**Cuprins:**

[**Descrierea variabilelor și a observațiilor** 3](#_Toc124630144)

[**Abordarea privind analiza datelor** 3](#_Toc124630145)

[**Motivația de a opta pentru o anumită metodă de analiză a datelor** 4](#_Toc124630146)

[**Prezentarea rezultatelor** 5](#_Toc124630147)

[Varianță explicată de componente principale 5](#_Toc124630148)

[Corelaţiile dintre variabilele observate și componentele principale 6](#_Toc124630149)

[Calculul scorurilor 7](#_Toc124630150)

[Valorile cosinus 7](#_Toc124630151)

[Contribuţia instanțelor la varianța componentelor 8](#_Toc124630152)

[Comunalități 8](#_Toc124630153)

[**Interpretare** 9](#_Toc124630154)

[Interpretare varianță explicată de componente principale 9](#_Toc124630155)

[Interpretare corelaţiile dintre variabilele observate și componentele principale 9](#_Toc124630156)

[Interpretare calculul scorurilor 9](#_Toc124630157)

[Interpretare valori cosinus 10](#_Toc124630158)

[Interpretare contribuţia instanțelor la varianța componentelor 10](#_Toc124630159)

[Interpretare comunități 10](#_Toc124630160)

[**Discuții și concluzii** 11](#_Toc124630161)

# **Descrierea variabilelor și a observațiilor**

**Din**  punct de vedere al contextului, McDonald’s a fost lansat în anul 1996 pe piața din India în Bandra, Mumbai, printr-un singur restaurant. Astăzi, McDonald’s are aproximativ 480 de magazine în toată țara respectivă. Acest set de date oferă o analiză nutrițională a fiecărui element de meniu din meniul indian McDonald's, incluzând micul dejun, burgeri, cartofi prăjiți, salate, sifon, cafea și ceai, milkshake și deserturi. Am ales să iau în considerare variabilele următoare:

* Caloriile (kCal)
* Proteine (g)
* Grăsimi totale (g)
* Grăsimi saturate (g)
* Grăsimi nesaturate (g)
* Colesterol (mg)
* Carbohidrați totali (g)
* Zaharuri totale (g)
* Zaharuri adăugate (g)
* Sodiu (mg)

# **Abordarea privind analiza datelor**

**Am** ales să analizez acest set de date dat fiind faptul că am luat în considerare următoarele ipoteze:

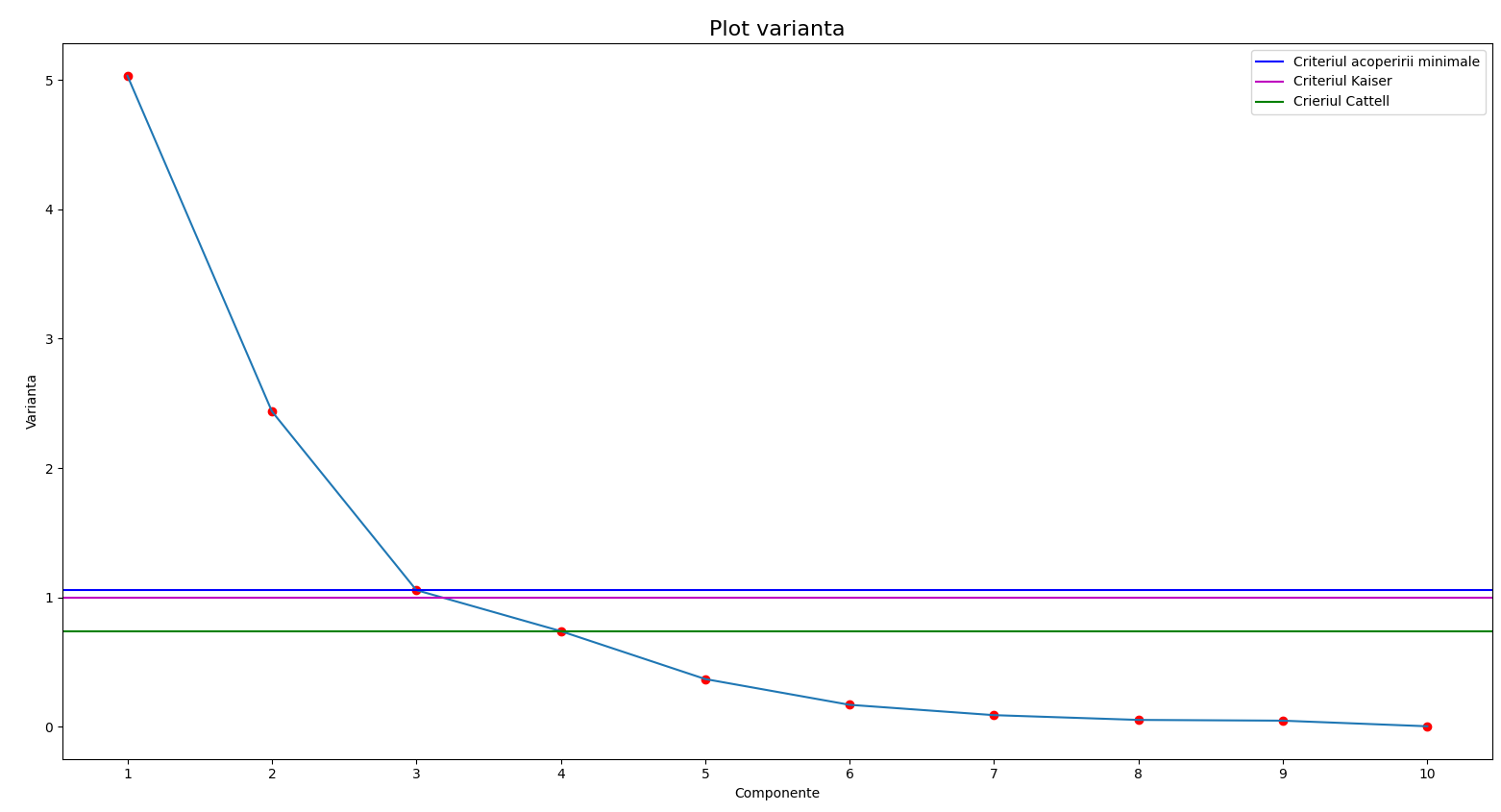
* Câte calorii ar putea avea un meniu de tip mediu McDonald's ?
* Cât de mult se iau în considerare băuturile atunci când vine vorba de sifon sau cafea la aportul caloric general ?
* Valoarea nutrițională a unui sandviș crește atunci când puiul este la grătar în loc de crocant ?
* Pentru a satisface cerințele nutriționale în decursul unei zile, care ar fi cel mai mic număr de produse pe care le-am putea comanda din meniu ?

# **Motivația de a opta pentru o anumită metodă de analiză a datelor**

**Dintre** metodele de analiză puse la dispoziție, am ales să utilizez Analiza componentelor principale(ACP) dat fiind faptul că, este una din cele mai utilizate tehnici de analiza a seturilor de date, în special cele de mari dimensiuni, având ca scop sintetizarea informației. Din punct de vedere al analizei datelor, prima componentă principală a unui set de n variabile este variabila derivată formată ca o combinație liniară a variabilelor originale ce explică cea mai mare varianță. Urmatoarea lămurește cea mai mare variație a ceea ce ramâne odată ce efectul celei de-a întâia este eliminat, astfel trecându-se prin n iterații până când varianța este explicată. Analiza factorilor (AF) este o colecție de tehnici statistice al căror obiectiv este acela de a simplifica seturi complexe de variabile, ceea ce setul de date utilizat nu prezintă o astfel de caracteristică. De asemenea, Analiza corelațiilor canonice (ACC) se referă la o modalitate de a deduce informații din matricea de covarianță încrucișată, acest lucru nefiind de folos în cadrul analizei alese, Analiza discriminantă, este o tehnică multivariată utilizată pentru a găsi contribuția fiecărei variabile în separarea grupurilor de observații pe baza variabilelor măsurate pe fiecare eșantion, ceea ce nu este semnificativ într-un set de date legat de valori nutriționale în cadrul unor produse. Analiza de cluster (AC) se referă la gruparea unui set de obiecte în așa fel încât obiectele din același cluster să fie mai asemănătoare într-un anumit sens între ele decât cu cele din alte grupuri, fapt ce nu se regăsește în setul de date curent. În final, Analiza corespondențelor multiple (ACM) este o tehnică de analiză pentru date categorice nominale, utilizată pentru a detecta și reprezenta structurile subiacente într-un set de date neavând legătură cu acest set de date.

# **Prezentarea rezultatelor**

## Varianță explicată de componente principale

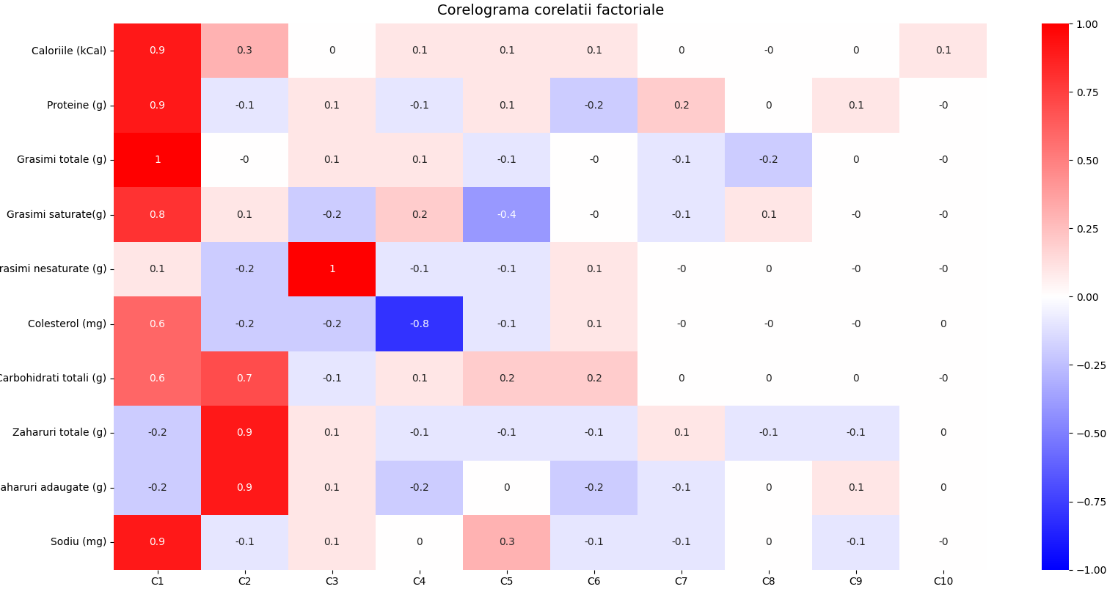


*Figura 1. Graficul valorilor proprii*



*Figura 2. Distribuția variației*

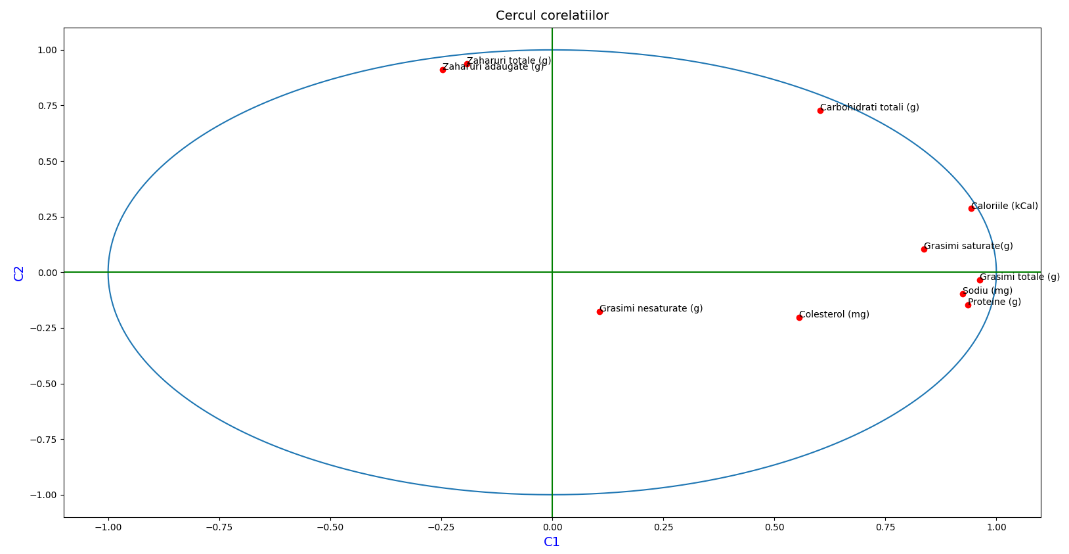
## Corelaţiile dintre variabilele observate și componentele principale



*Figura 3. Corelograma corelații factoriale*

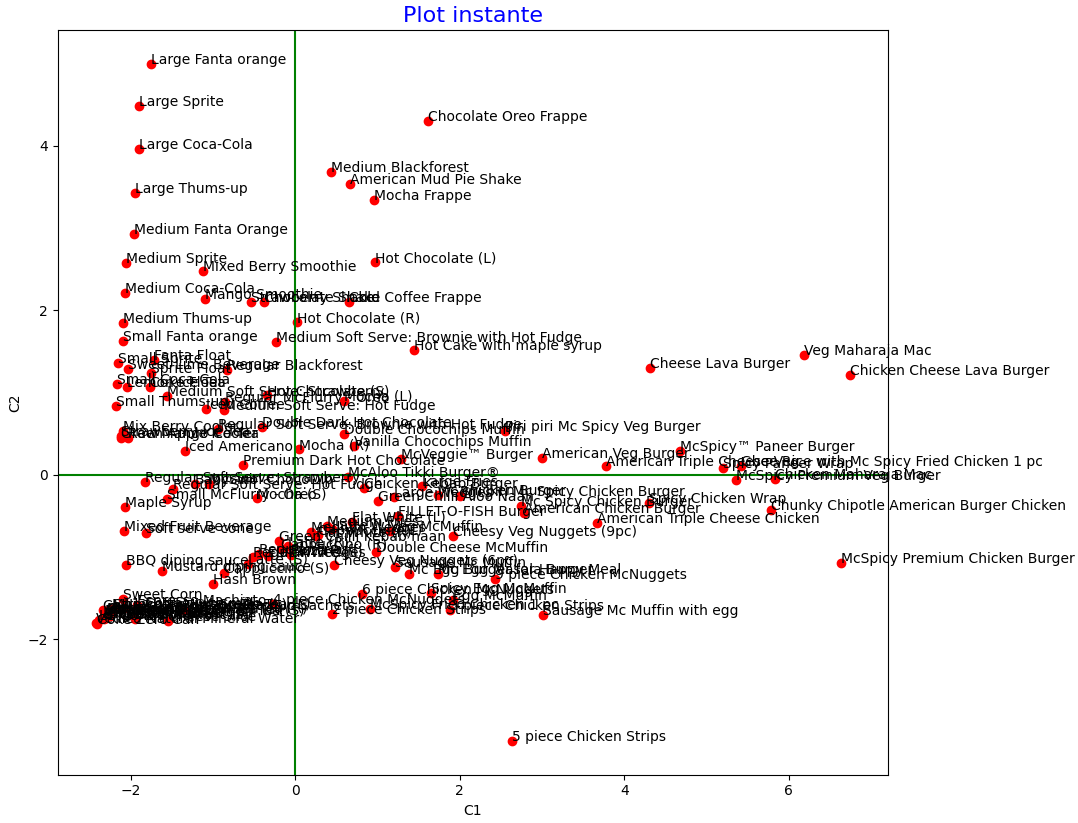


*Figura 4. Corelații factoriale*



*Figura 5. Cercul corelațiilor*

## Calculul scorurilor



*Figura 6. Plot scoruri în primele 2 componente*

## Valorile cosinus



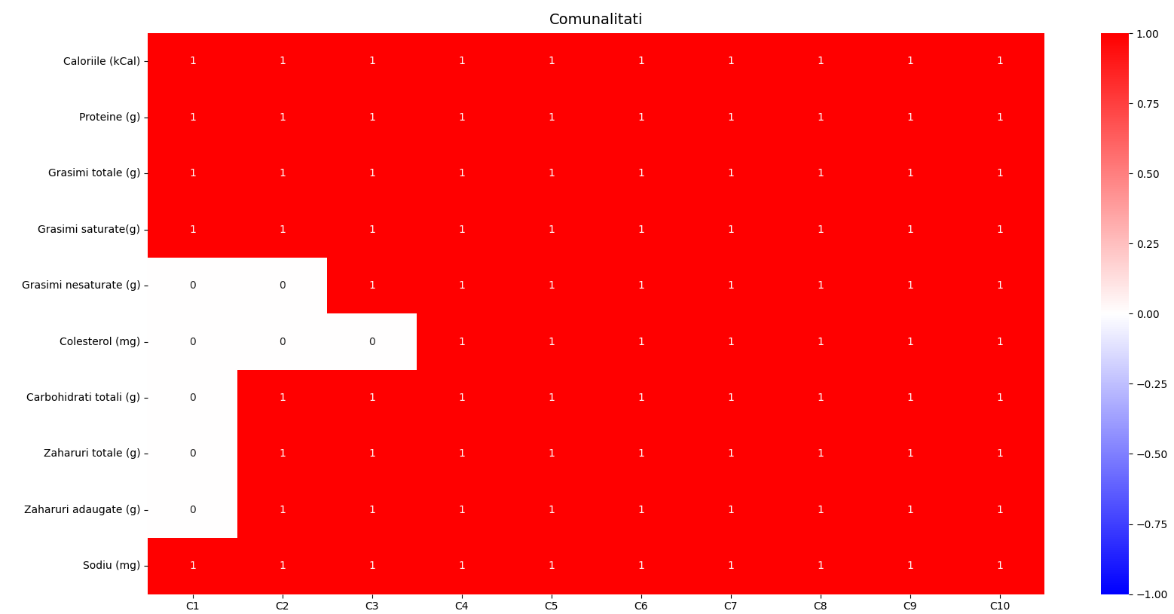
*Figura 7. Valori cosinus*

## Contribuţia instanțelor la varianța componentelor



*Figura 8. Contribuţia instanțelor la varianța componentelor*

## Comunalități



*Figura 9. Comunalități*

# **Interpretare**

## Interpretare varianță explicată de componente principale

**Din** figura este evediențiat faptul că sunt semnificative doar primele 3 componente, lucru evidențiat și de criteriile Cattel și Kaiser, deoarece au o varianță mai mare decât 1. De asemenea, din figura 2 este evidențiat faptul că prima componentă acoperă 50.31%, iar următoarele două acoperă între 24.37% și 10.56%.

## Interpretare corelaţiile dintre variabilele observate și componentele principale

**Corelațiile** factoriale se calculează pornind de la vectorii proprii ai matricei de corelație, reprezintând coeficienții de corelație liniară dintre variabilele observate și componentele principale. Sunt calculate din vectorii proprii, iar din acest fapt rezultă contribuția fiecărei variabile observate la constituirea unei componente principale. Aceste corelații au scopul de a denomina componentele principale, practic de a asocia o semnificație concretă componentelor principale. Așadar, componenta C1 este puternic influențată în mod direct de către calorii, proteine, grăsimi totale, sodiu și grăsimi nesaturate. Componenta C2 este influențată pozitiv de zaharuri totale, adăugate și carbohidrați totali. Astfel aceasta poate fi un factor decisiv în ceea ce privește calitatea unui produs McDonald’s. Cea de a treia componenta, C3, este realizată prin grăsimi nesaturate.

## Interpretare calculul scorurilor

**Valorile** mari pentru componenta 1 reflectă valori mari pentru calorii, proteine, grăsimi totale, sodiu și grăsimi nesaturate, deoarece sunt puternic puternic corelate, iar variabilele zaharuri totale și adăugate sunt invers corelate. Pentru componenta 2 putem observa faptul că produse precum băuturile carbogazoase prezintă calorii mai puține, astfel fiind considerate mai calitative. În emisfera cealaltă, Produsele de pui sunt considerate ca fiind calitative datorită faptului că sunt bogate în proteine.

## Interpretare valori cosinus

Valorile cosinus reprezintă importanța relativă a unui produs pentru o componentă în raport cu celelalte, iar în folder-ul de date de tip output al analizei, mai specific, fișierul „cosin.csv” sunt prezentate aceste valori. Prin acest tabel se pot evidenția produsele ce au o importanță mai mare pentru anumite componente, precum produsul Regular Soft Serve: Strawberry în componenta 1 și produsul Medium Blackforest în componenta 2.

## Interpretare contribuţia instanțelor la varianța componentelor

Contribuția instanțelor în cadrul varianței componentelor reflectă importanța relativă a unui produs pentru o componentă în raport cu ceilalți indivizi, astfel, acestea pot fi regăsite în fișierul de date de tip output al analizei, mai specific, „contrib.csv”. Spre exemplu, pentru prima componentă, produsele Hot Chocolate (R) și Pizza Puff au valori foarte mici, deci au calorii scăzute. Large Fanta orange ia o valoare foarte mare pentru componenta 2, deci pare cel mai calitativ.

## Interpretare comunități

Scoaterea în evidență a „transferului” gradual de informație din variabilele observate asupra componentelor principale se face prin comunități, acestea fiind prezentate prin corelograma din figura 9, unde, se pot observa variabilele ce au legături mai puternice cu componentele mai puțin semnificative.

# **Discuții și concluzii**

În concluzie, analiza componentelor principale este una dintre cele mai utilizate metode de analiză factorială multidimensională, dat fiind faptul că reflectă un sistem de axe factoriale ce reprezintă informația din tabelul inițial cu scopul vizionării mai bune. Determinarea unui număr mai mic de variabile noi, numite componente principale, reflectat de această analiză, conturează informația și variabilitatea existentă din punctul de vedere al colectivității studiate, fiind folosit ca instrument de analiză statistică în multe domenii de activitate.

Având în vedere faptul că McDonald’s este un restaurant de tip fast-food, este dovedit științific că dependența de acest tip de mâncare duce la diferite efecte negative ale corpului uman, precum obezitatea, posibilitatea de creștere a riscului de cancer, cât și efecte mai puțin semnificative, precum dureri de cap, acnee și balonare.