

UNIVERSITATEA DE VEST DIN TIMIȘOARA

FACULTATEA DE ȘTIINȚE ALE GUVERNĂRII ȘI COMUNICĂRII

**PROGRAMUL DE STUDII DE MASTERAT: COMUNICARE PUBLICITARĂ ÎN
MEDII DIGITALE**

Analiza Big Data

**Analiza variației
comportamentului de cumpărare
în funcție de gen, vârstă și
statutul abonamentului**

COORDONATORI:

Lect. univ. dr. Silvia Fierăscu

STUDENȚI:

Annamaria Sebestyen

Denisa Ștefan

Rareș-Sebastian Moț

TIMIȘOARA, 2025

Semestrul II

CUPRINS

CUPRINS.....	1
1. Introducere.....	1
2. Întrebarea de cercetare.....	2
3. Obiectivul cercetării.....	3
4. Ipoteze de lucru.....	4
H1: Genul influențează valoarea și frecvența achizițiilor online.....	4
H2: Vârsta este un factor determinant al comportamentului de cumpărare.....	4
H3: Statutul de abonament este asociat cu o valoare mai mare și o frecvență mai ridicată a cumpărăturilor.....	4
H4: Efectul combinat al genului, vârstei și abonamentului generează tipare distincte de comportament de cumpărare.....	5
5. Metodologie.....	6
5.1. Setul de date.....	6
5.2. Prelucrarea și pregătirea datelor.....	10
5.3. Măsurători și indicatori.....	10
5.4. Tipul de analiză.....	10
5.5. Limitări metodologice.....	10
6. Rezultate și interpretare.....	12
6.1. Genul și valoarea achizițiilor.....	12
6.2. Vârsta și valoarea achizițiilor.....	13
6.3. Vârsta și frecvența cumpărăturilor.....	14
6.4. Abonamentul și valoarea achiziției.....	15
6.5. Abonamentul și frecvența cumpărăturilor.....	16
Concluzii intermediare din rezultate în corelație cu ipotezele de cercetare:.....	17
6.6. Analiză inferențială: teste ANOVA și regresie liniară.....	19
Testele ANOVA.....	19
Regresie liniară multiplă.....	19
Discuții pe baza analizei inferențiale.....	20
7. Discuții.....	22
7.1. Rolul vârstei în comportamentul de cumpărare.....	22
7.2. Genul ca variabilă moderatoare.....	22
7.3. Abonamentul între loialitate și valoare.....	22
7.4. Interacțiuni între variabile.....	22
8. Limite ale cercetării.....	23
8.1. Dezechilibrul eșantionului pe segmente.....	23

8.2. Ambiguitatea statutului de abonament.....	23
8.3. Analiză transversală.....	23
8.4. Limitările analizei inferențiale.....	23
8.5. Lipsa variabilelor contextuale.....	23
8.6. Posibilă influență a comportamentului oportunist.....	23
8. Limite ale cercetării.....	24
8.1. Dezechilibrul eșantionului pe segmente.....	24
8.2. Lipsa unei definiții precise a variabilei „Subscription Status”.....	24
8.3. Analiză transversală (cross-sectional).....	24
8.4. Absența analizelor statistice inferențiale.....	25
8.5. Variabile contextuale neincluse.....	25
8.6. Posibila autoreglare comportamentală.....	25
9. Concluzii și recomandări.....	26
9.1 Concluzii.....	26
9.2 Recomandări.....	26
10. Bibliografie.....	28
ANEXE.....	1

1. Introducere

În era digitalizării accelerate, comerțul electronic a devenit un pilon esențial în structura economică globală. Consumatorii interacționează cu platformele online într-un mod tot mai complex și personalizat, iar comportamentele de cumpărare se diversifică în funcție de o serie de factori contextuali, demografici și psihologici. Din perspectiva comercianților, capacitatea de a înțelege și anticipa aceste comportamente este esențială pentru optimizarea strategiilor de marketing, creșterea vânzărilor și fidelizarea clienților.

În acest context, analiza comportamentului de cumpărare prin metode cantitative și instrumente de tip big data devine o resursă strategică. Segmentarea clienților pe baza caracteristicilor demografice și comportamentale permite nu doar personalizarea ofertei, ci și identificarea de tipare și tendințe care pot ghida deciziile de afaceri. Printre variabilele frecvent analizate se numără **genul**, **vârsta** și **interacțiunea clientului cu platforma**, inclusiv participarea într-un **program de abonament** (ex: newsletter, cont premium sau programe de loialitate).

Importanța acestui tip de analiză este dublă. Pe de o parte, oferă o înțelegere mai bună a modului în care diverse grupuri demografice se raportează la cumpărăturile online, de exemplu, dacă femeile sub 25 de ani cumpără mai frecvent, dar sume mai mici, în comparație cu bărbații peste 35. Pe de altă parte, permite testarea eficienței unor mecanisme de retenție, cum ar fi abonamentele care pot influența comportamentul prin oferte exclusive, reduceri sau beneficii logistice (ex: livrare gratuită).

Această lucrare își propune să investigheze modul în care valoarea medie a achizițiilor și frecvența cumpărăturilor variază în funcție de trei variabile explicative majore: **genul clientului**, **grupa de vârstă** și **statutul de abonament**. Prin această analiză ne propunem să oferim o perspectivă fundamentată asupra modului în care aceste dimensiuni contribuie la diferențierea comportamentului de cumpărare și să identificăm eventuale tipare relevante pentru strategii de business data driven.

2. Întrebarea de cercetare

În orice demers de cercetare cantitativă, formularea unei întrebări de cercetare clare, delimitate conceptual și orientate spre verificare empirică reprezintă un punct de plecare esențial. Întrebarea de cercetare trebuie să reflecte scopul investigației și să permită operaționalizarea variabilelor într-un mod care poate fi măsurat și analizat cu ajutorul datelor disponibile.

În cazul acestei lucrări, întrebarea de cercetare este formulată astfel:

Cum variază comportamentul de cumpărare (valoarea și frecvența achizițiilor) în funcție de gen, vârstă și statutul abonamentului?

Această întrebare vizează identificarea unor diferențe semnificative în comportamentul clienților online, analizând două dimensiuni centrale:

- **Valoarea achizițiilor** (măsurată prin media și mediana sumei cheltuite pe tranzacție)
- **Frecvența achizițiilor** (măsurată prin cea mai comună periodicitate declarată de respondenți)

Cele trei variabile explicative sunt:

- **Genul** (feminin / masculin)
- **Vârsta**, împărțită în cinci grupe: <25, 25–35, 36–50, 51–65, 65+
- **Statutul de abonament**, care indică dacă utilizatorul este sau nu abonat la o formă de program de loialitate sau comunicare comercială (ex. newsletter, cont premium)

Această întrebare permite o analiză multifactorială, relevantă atât din perspectiva comportamentului consumatorului, cât și a aplicațiilor practice în domeniul marketingului digital. Ea deschide calea testării unor ipoteze despre influența demografiei și a angajamentului față de brand (prin abonament) asupra deciziilor de cumpărare.

Mai mult, întrebarea se pretează la o abordare exploratorie și descriptivă bazată pe analiza unui set de date reale, reflectând un context comercial contemporan. Rezultatele obținute pot genera recomandări aplicabile în strategiile de segmentare a pieței, targeting personalizat și îmbunătățirea experienței clienților în mediul digital.

3. Obiectivul cercetării

Obiectivul general al acestei cercetări este de a înțelege modul în care caracteristicile demografice (genul și vârsta) și interacțiunea consumatorului cu platforma de cumpărături (prin statutul de abonat) influențează comportamentul de cumpărare online, măsurat prin valoarea achizițiilor și frecvența acestora.

Această analiză urmărește să răspundă la următoarele direcții de investigare:

- Dacă există diferențe semnificative între femei și bărbați în ceea ce privește valoarea medie a tranzacțiilor și ritmul cumpărăturilor.
- Cum variază frecvența și valoarea achizițiilor între diferitele grupe de vârstă, cu accent pe identificarea unor tipare de consum specifice tinerilor vs. adulților.
- În ce măsură statutul de abonat, considerat un indicator al loialității sau al implicării active în relația cu brandul, este asociat cu un comportament de cumpărare mai activ sau cu o valoare mai mare a tranzacțiilor.

Obiectivul operațional al cercetării constă în segmentarea populației de clienți în funcție de aceste trei dimensiuni explicative și în analiza comparativă a indicatorilor de comportament de cumpărare în cadrul fiecărui segment. Această abordare va permite nu doar descrierea diferențelor, ci și identificarea unor direcții de acțiune relevante pentru strategii de marketing și personalizare a experienței de consum.

În plan aplicativ, cercetarea are ca scop sprijinirea deciziilor data-driven în e-commerce, oferind o imagine de ansamblu asupra segmentelor de consumatori care prezintă cel mai mare potențial din perspectiva valorii și regularității achizițiilor. De asemenea, rezultatele obținute pot fundamenta optimizarea programelor de loialitate și abonamente, astfel încât acestea să răspundă mai eficient nevoilor diverselor categorii de clienți.

4. Ipoteze de lucru

Formularea ipotezelor permite testarea relațiilor dintre variabile și direcționarea interpretării datelor. În cadrul acestei cercetări, ipotezele sunt construite pornind de la concluzii evidențiate în literatura de specialitate privind comportamentul consumatorului în mediul digital, precum și din observațiile preliminare asupra datelor.

H1: Genul influențează valoarea și frecvența achizițiilor online.

Studiile arată că femeile tind să achiziționeze produse online mai frecvent, dar să cheltuiască sume mai mici per tranzacție, în timp ce bărbații pot prefera achiziții mai rare, dar mai consistente ca valoare (Rodgers & Harris, 2003; Dittmar et al., 2004). Genul influențează și tipologia produselor achiziționate, precum și gradul de loialitate față de brand.

H2: Vârsta este un factor determinant al comportamentului de cumpărare.

Vârsta influențează nu doar tipul produselor cumpărate, ci și valoarea achiziției și frecvența. Tinerii (sub 35 ani) manifestă comportamente de cumpărare mai frecvente și mai spontane, în timp ce consumatorii maturi (peste 35 ani) tind să fie mai selectivi, dar dispuși să cheltuiască mai mult per achiziție (Bakewell & Mitchell, 2003; Yang et al., 2019).

H3: Statutul de abonament este asociat cu o valoare mai mare și o frecvență mai ridicată a cumpărăturilor.

Programele de loialitate, abonamentele și alte forme de membership sunt concepute pentru a crește retenția clienților și valoarea pe termen lung (LTV – Lifetime Value). Clienții abonați sunt de regulă mai activi, revin mai frecvent pe platformă și sunt mai predispuși la a profita de oferte și reduceri (McKinsey, 2020; Liu, 2007).

H4: Efectul combinat al genului, vârstei și abonamentului generează tipare distincte de comportament de cumpărare.

Interacțiunile dintre variabilele socio-demografice și implicarea activă în platforme (prin abonamente) pot genera tipologii comportamentale distincte. Aceste combinații influențează complex modul în care consumatorii iau decizii de achiziție (Kim et al., 2010; Grewal et al., 2020).

5. Metodologie

Această cercetare se încadrează într-o abordare cantitativ-descriptivă, vizând identificarea diferențelor în comportamentul de cumpărare online în funcție de variabilele socio-demografice și comportamentale ale consumatorilor.

5.1. Setul de date

Studiul de față își propune să analizeze comportamentul de cumpărare al consumatorilor în mediul online, pe baza unui set de date complex, care conține 3900 de observații. Datele provin dintr-o platformă de e-commerce și reflectă o varietate de factori demografici, comportamentali și comerciali, oferind astfel un cadru relevant pentru cercetări aplicate în domeniul marketingului digital și al analizei big data.

Baza de date acoperă o gamă largă de locații geografice, cu o reprezentare semnificativă din state precum California, Texas, New York, Florida și Pennsylvania. Această diversitate regională poate influența atât preferințele de consum, cât și tipologia produselor achiziționate.

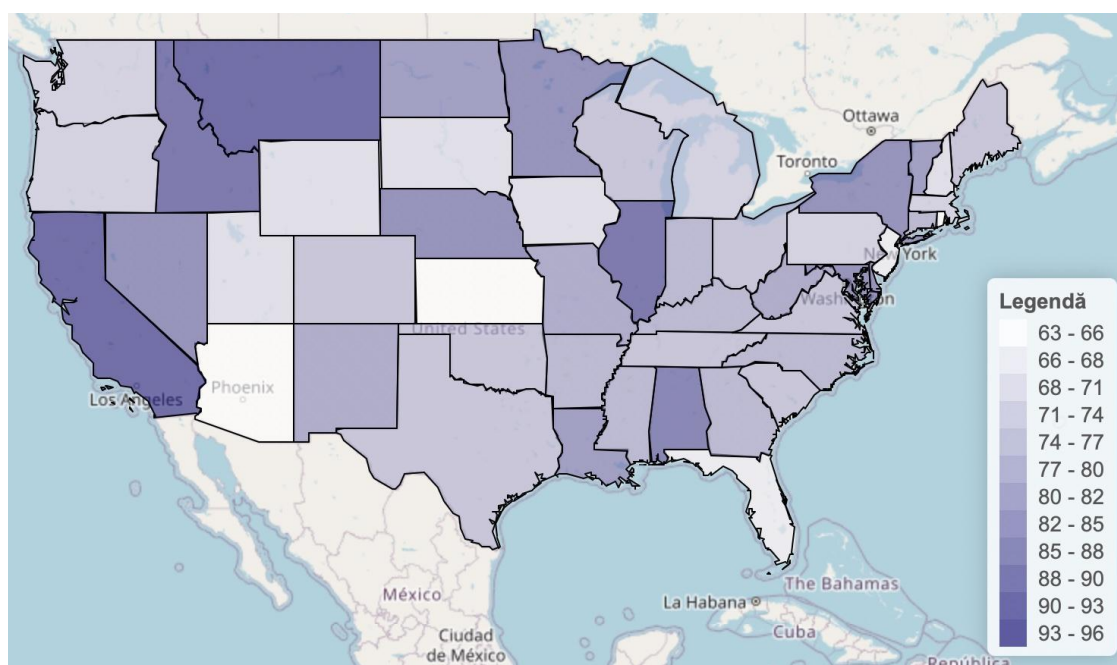


Fig. 1: Heatmap cu statele din America de Nord cu cele mai multe vânzări cf. bazei de date.

Grafic generat prin intermediul heatmapper.ca

În ceea ce privește comportamentul de cumpărare, cele mai frecvente produse achiziționate aparțin categoriilor de îmbrăcăminte și încălțăminte, cum ar fi blouse, sandals, sneakers, sweater sau t-shirt, ceea ce indică un interes crescut pentru articole de modă și lifestyle. Culoarele predominante ale produselor selectate, maroon, gray, navy blue, white, sugerează preferințe pentru tonuri neutre, dar și o diversificare moderată prin opțiuni precum turquoise și beige.

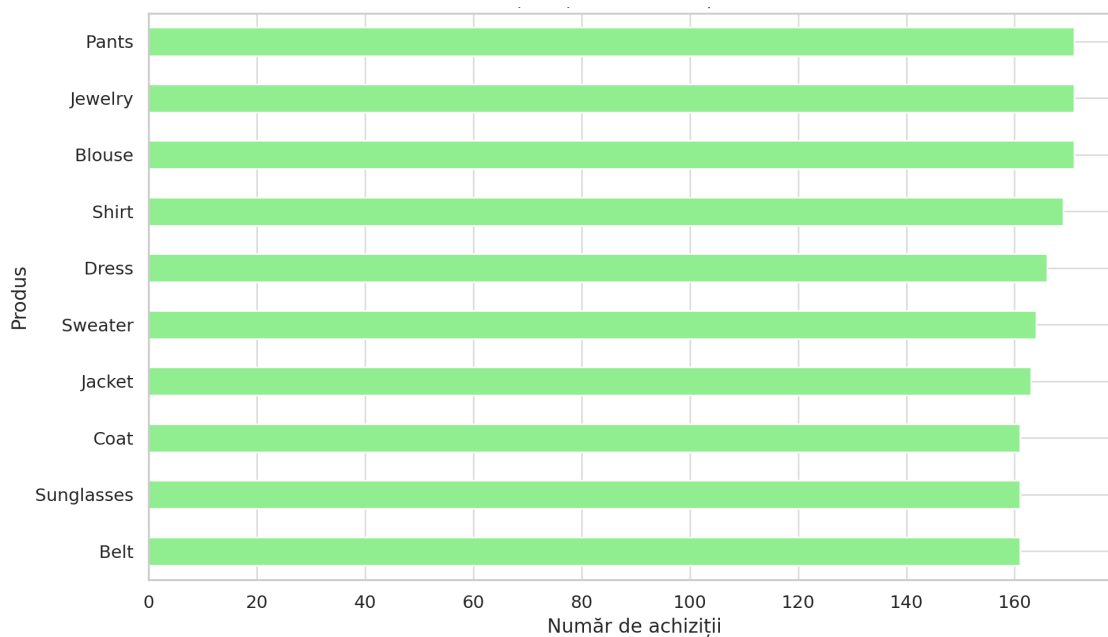


Fig. 2: Top 10 produse achiziționate.

Grafic generat prin intermediul Chat GPT 4o printr-o comandă explicită și în concordanță cu baza de date.

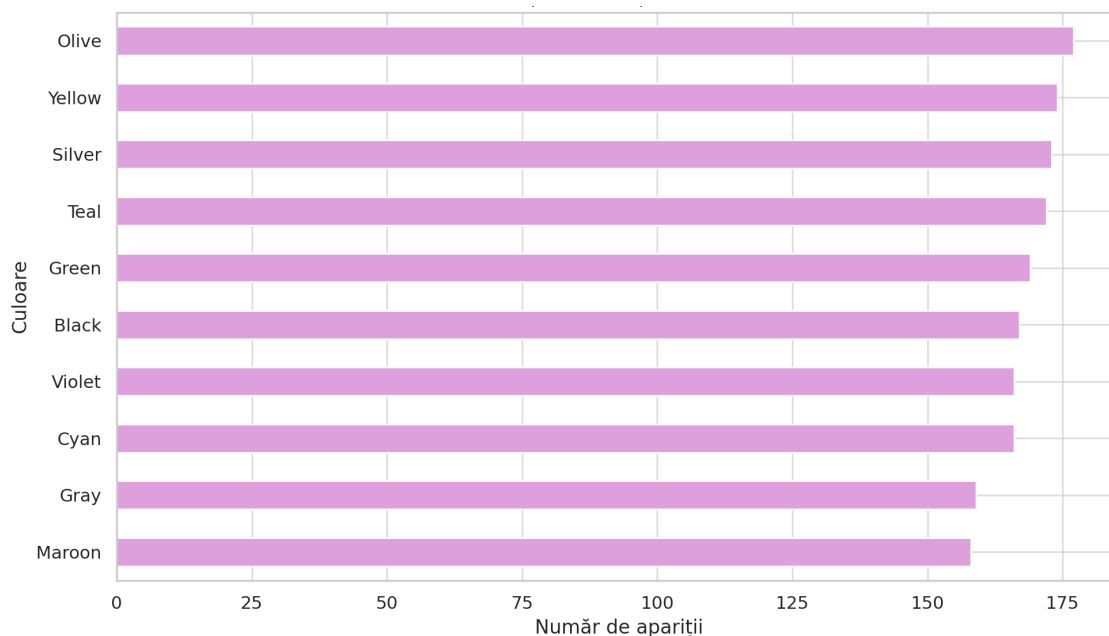


Fig. 3: Top 10 culori preferate.

Grafic

generat prin intermediul Chat GPT 4o printr-o comandă explicită și în concordanță cu baza de date.

Achizițiile sunt distribuite relativ uniform între cele patru anotimpuri, însă primăvara și vara înregistrează o activitate crescută, posibil datorită lansării colecțiilor noi și a campaniilor sezoniere.

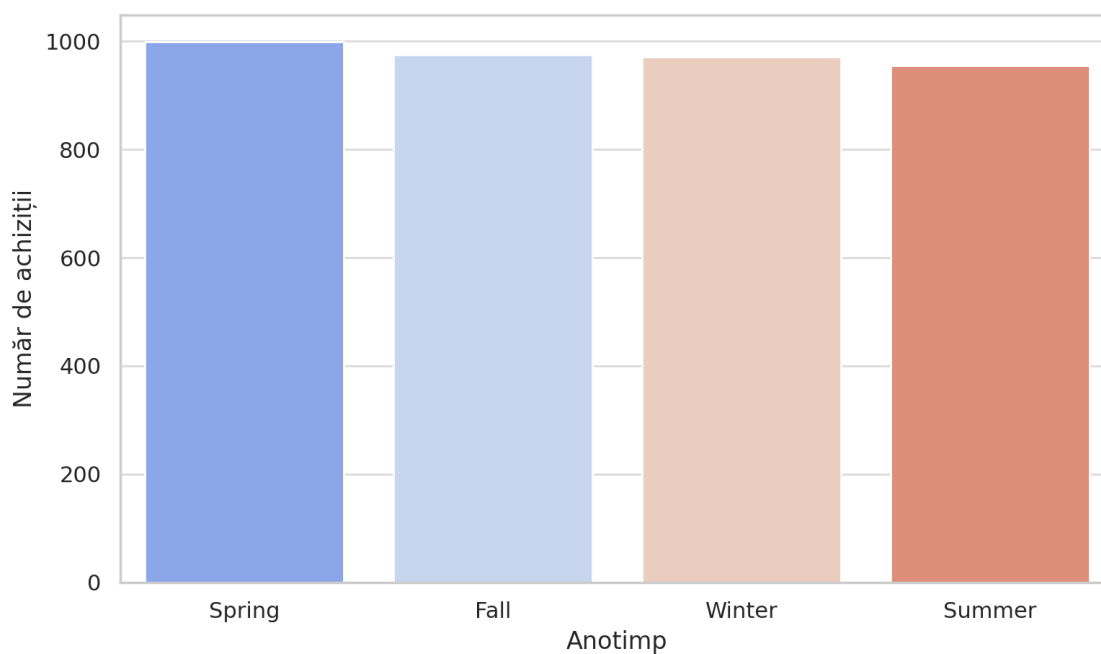


Fig. 4: Distribuția cumpărăturilor pe anotimpuri.

Grafic generat prin intermediul Chat GPT 4o printr-o comandă explicită și în concordanță cu baza de date.

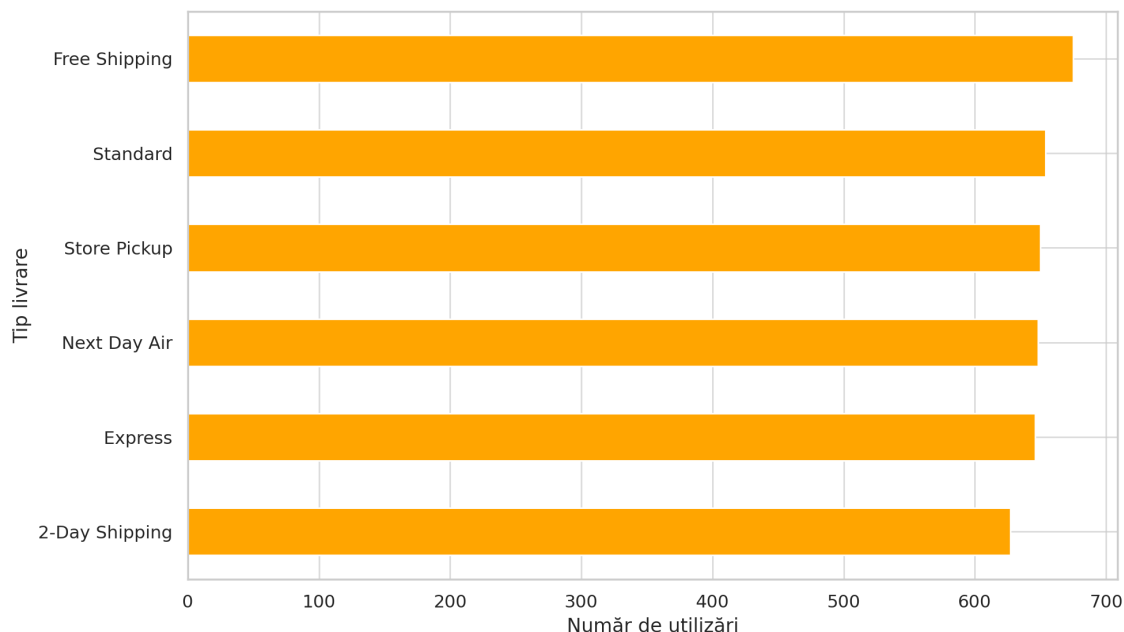


Fig. 5: Distribuția tipurilor de livrare.

Grafic generat prin intermediul Chat GPT 4o printr-o comandă explicită și în concordanță cu baza de date.

Pe plan logistic, utilizatorii au preferat metode variate de livrare: „Free Shipping” este cea mai frecvent selectată opțiune, urmată îndeaproape de „Standard”, „Store Pickup”, „Next Day Air” și „Express”. Această distribuție echilibrată denotă faptul că platforma oferă o flexibilitate ridicată, adaptată la nevoile diferitelor tipuri de cumpărători, de la cei orientați spre confort și costuri reduse, până la cei care prioritizează viteza.

În ceea ce privește variabila „Subscription Status”, aceasta desemnează statutul de abonat sau neabonat al clientului în raport cu platforma de comerț online. Cu toate acestea, baza de date nu precizează explicit natura abonamentului, dar am perceput acesta ca implicând un program plătit, o simplă înscriere la newsletter sau un cont fidelizat cu beneficii. Această ambiguitate trebuie avută în vedere în interpretarea rezultatelor și este menționată ca o limitare metodologică în secțiunile ulterioare.

Prin urmare, setul de date analizat oferă un cadru robust pentru investigarea relației dintre variabile precum genul, vârsta și statutul de abonament și comportamentele de consum exprimate prin valoarea și frecvența achizițiilor.

5.2. Prelucrarea și pregătirea datelor

Pentru o analiză relevantă și comparabilă:

- Datele incomplete au fost eliminate (observațiile care nu conțineau informații despre gen, vârstă, abonament sau valoarea achiziției).
- Vârsta a fost transformată într-o variabilă categorială, împărțită în cinci grupe:

Sub 25 de ani

25–35 ani

36–50 ani

51–65 ani

Peste 65 ani

- Categoriile demografice (gen, vârstă, abonament) au fost combinate pentru a forma segmente de consumatori.

5.3. Măsurători și indicatori

Pentru fiecare segment, s-au calculat următorii indicatori:

- Valoarea medie și mediană a achizițiilor (în dolari americani)
- Frecvența dominantă a cumpărăturilor (exprimată în termeni de periodicitate: săptămânal, lunar etc.)
- Numărul de respondenți incluși în segment

5.4. Tipul de analiză

Cercetarea este de tip descriptiv-comparativ și se concentrează pe identificarea tendințelor și diferențelor între segmente. Nu s-au aplicat metode de inferență statistică (ex. teste de semnificație), deoarece unele categorii din eșantion nu conțin suficiente observații pentru a permite o analiză robustă.

5.5. Limitări metodologice

- Distribuția neuniformă a respondenților pe segmente a redus posibilitatea de a trage concluzii ferme în cazul anumitor combinații (de exemplu, lipsa respondenților abonați în grupele tinere de vârstă).

- Lipsa unei definiții explicite pentru statutul de abonament în cadrul setului de date poate genera ambiguități în interpretare (ex. abonament la newsletter vs. abonament plătit).
- Studiul este transversal, vizând un moment fix în timp, fără a analiza evoluția comportamentului în timp.

Această metodologie permite o explorare clară a variației comportamentului de cumpărare în funcție de factori relevanți, oferind un cadru solid pentru interpretarea rezultatelor prezentate în capitolul următor.

6. Rezultate și interpretare

Analiza comportamentului de cumpărare a fost realizată în funcție de cele trei variabile explicative: gen, grupe de vârstă și statutul de abonament. Pentru fiecare, am analizat atât valoarea achiziției, cât și frecvența cumpărăturilor. Rezultatele sunt prezentate mai jos, împreună cu interpretarea graficelor aferente.

6.1. Genul și valoarea achizițiilor

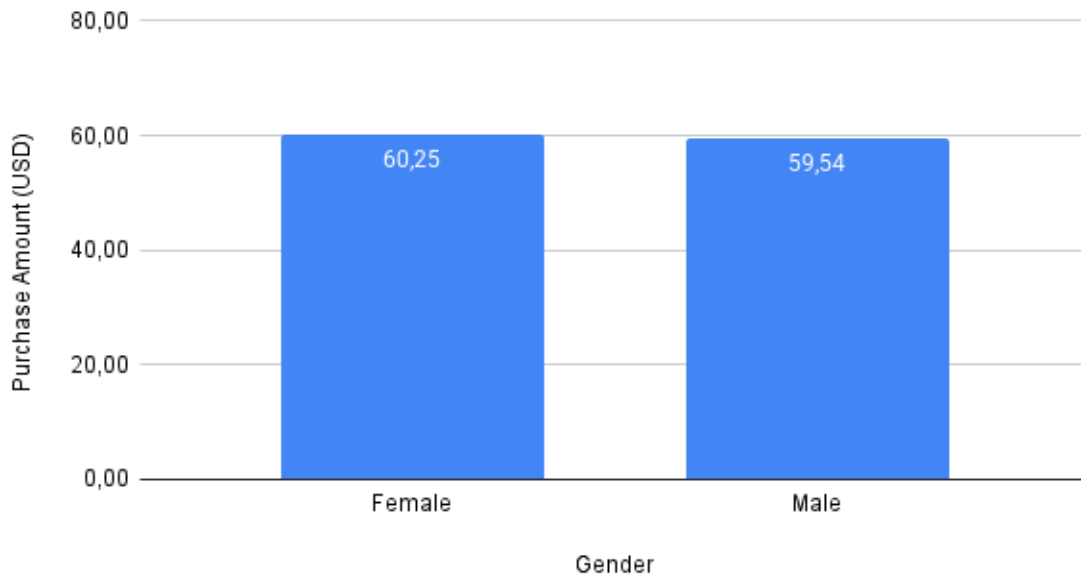


Fig. 6: Distribuția valorii achiziției în funcție de gen.

generat prin intermediul Google Sheets.

Grafic

Analiza graficului arată că:

- Femeile au o valoare medie a achiziției de 60,25 USD, comparativ cu 59,54 USD pentru bărbați.
- Diferența este ușoară, dar constantă, indicând o ușoară tendință a femeilor de a face achiziții mai valoroase.

Interpretare: Nu se observă diferențe dramatice între genuri în ceea ce privește suma cheltuită, dar femeile tind să prezinte o distribuție mai largă, sugerând o varietate mai mare a comportamentului de achiziție.

6.2. Vârsta și valoarea achizițiilor

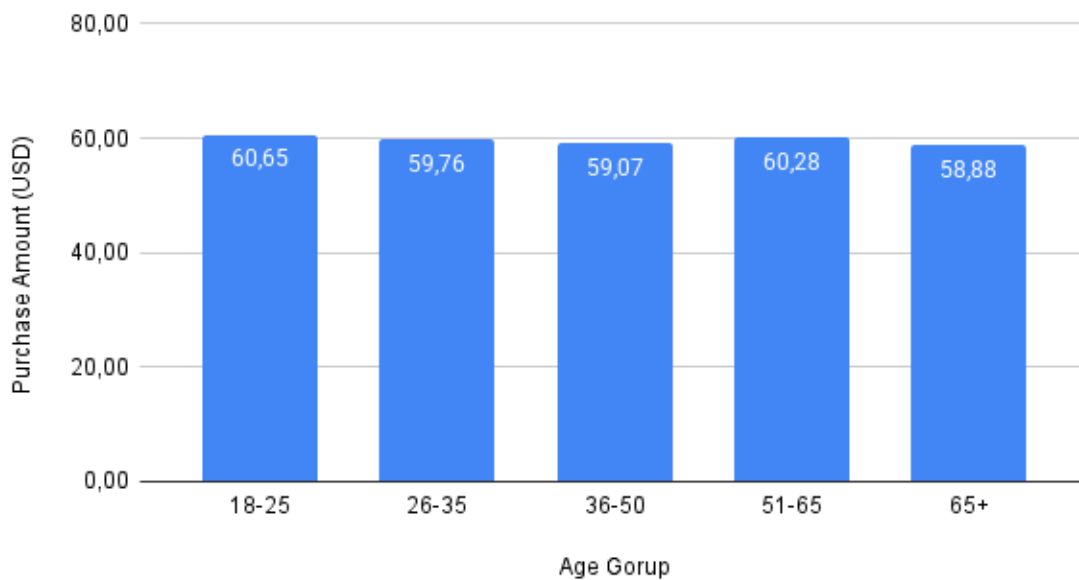


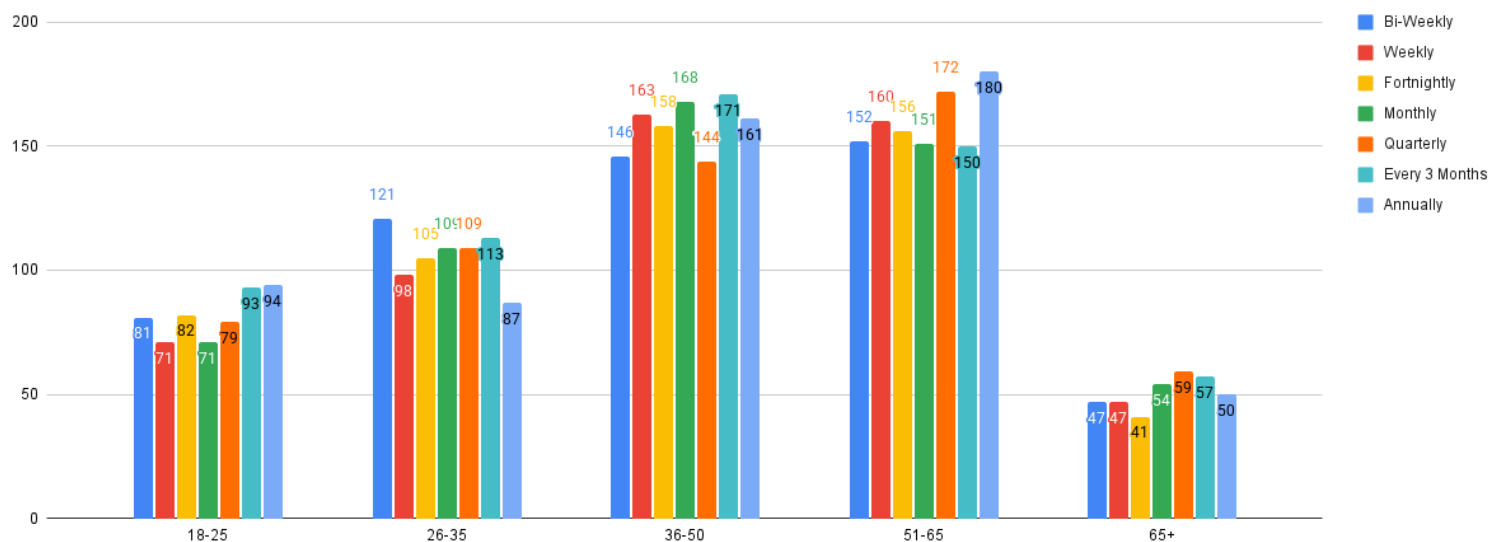
Fig. 7: Valoarea achizițiilor în funcție de grupele de vârstă.

Grafic generat prin intermediul Google Sheets.

Graficul relevă că:

- Cele mai mari valori medii apar la categoria 18–25 ani (60,65 USD) și 51–65 ani (60,28 USD).
- Seniorii (65+) și tinerii de 26–35 ani cheltuie cel mai puțin în medie.
- Rezultatele sugerează că adulții tineri și maturi sunt segmente cu potențial ridicat de valoare per achiziție.

6.3. Vârsta și frecvența cumpărăturilor



Graficul evidențiază:

- Cele mai active grupe de vârstă sunt 36–50 și 51–65, cu cele mai ridicate valori pe toate categoriile de frecvență.
- Tinerii (18–25) și seniorii (65+) au cele mai scăzute frecvențe.
- Adulții de vârstă mijlocie reprezintă segmentul cu cel mai constant comportament de cumpărare.

6.4. Abonamentul și valoarea achiziției

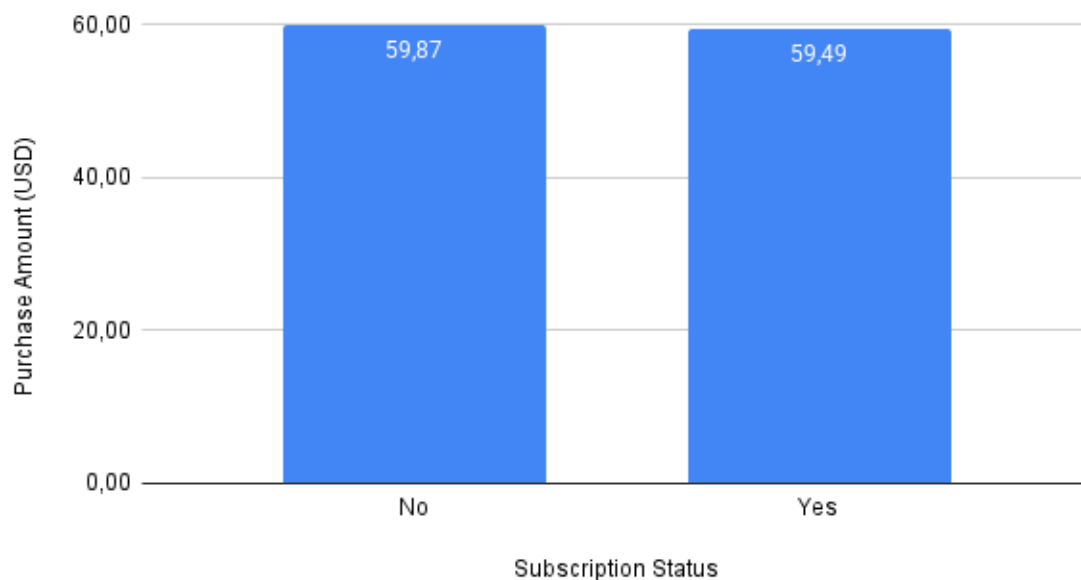


Fig. 9: Valoarea achizițiilor în funcție de statutul de abonament.
Grafic generat prin intermediul Google Sheets.

Conform graficului:

- Neabonații au o valoare medie a achiziției ușor mai mare (59,87 USD) comparativ cu abonații (59,49 USD).
- Deși diferența este minimă, statutul de abonat nu influențează semnificativ valoarea medie a achiziției. Aceasta invalidează parțial ipoteza conform căreia abonații ar cheltui mai mult.

6.5. Abonamentul și frecvența cumpărăturilor

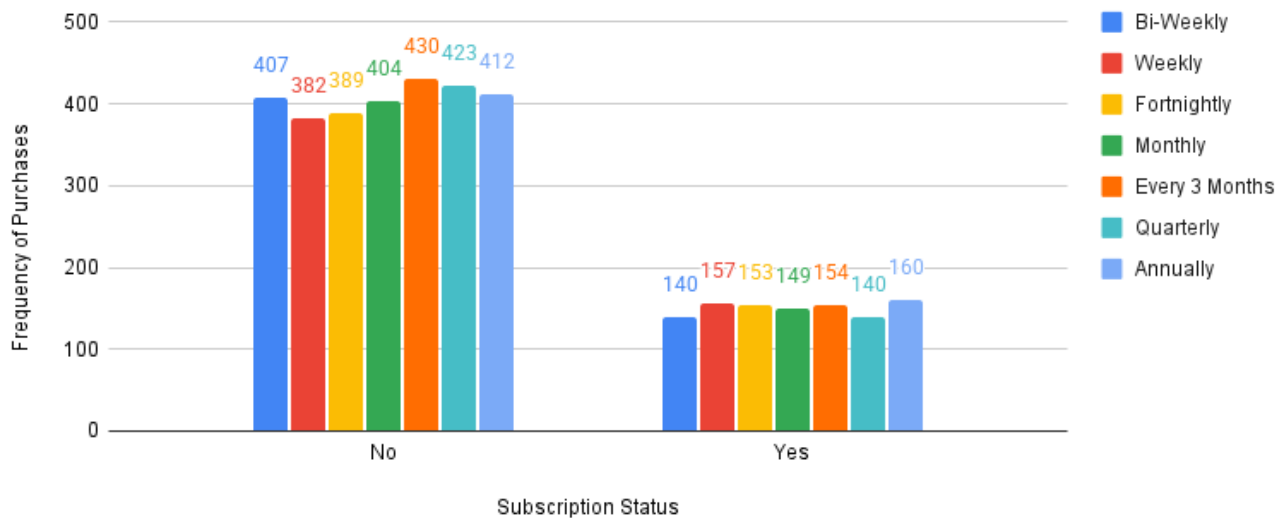


Fig. 10: Frecvența cumpărăturilor în funcție de abonament.

Grafic generat prin intermediul Google Sheets.

Graficul evidențiază:

- Neabonații cumpără într-un volum mai mare la toate frecvențele analizate, având un număr absolut superior.
- Totuși, acest rezultat este influențat de faptul că sunt de aproape 3 ori mai mulți neabonați decât abonați în eșantion.

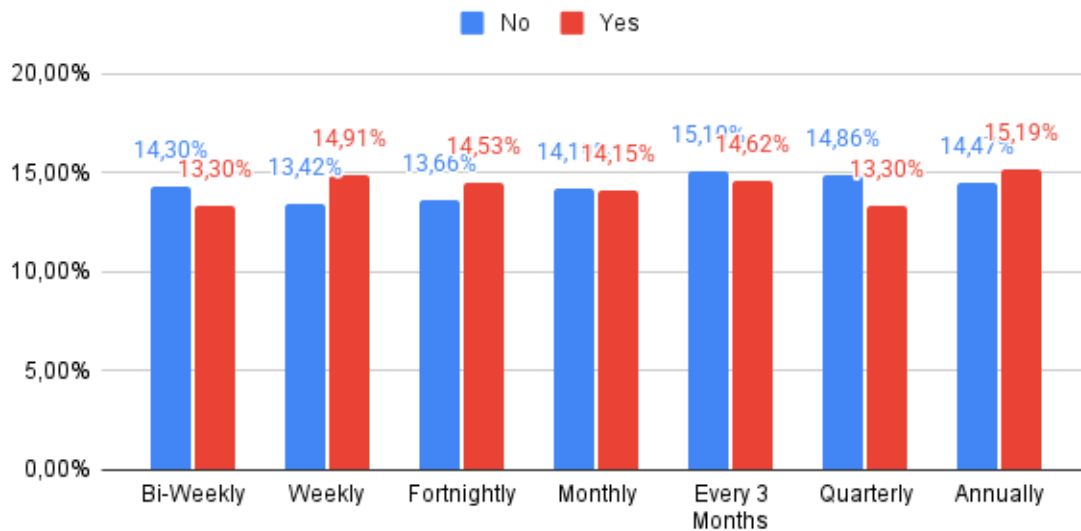


Fig. 11: Distribuția procentuală a frecvenței cumpărăturilor în funcție de abonament.

Grafic generat prin intermediul Google Sheets.

- Abonații au o frecvență mai ridicată de cumpărături în categoriile „Weekly”, „Fortnightly” și „Every 3 Months”.
- Deși sunt mai puțini la număr, abonații cumpără mai des raportat la propriul grup.
- Așadar, se confirmă ipoteza conform căreia abonații au un comportament de achiziție mai recurent și loial.

Concluzii intermediare din rezultate în corelație cu ipotezele de cercetare:

Ipoteza H1: Există diferențe semnificative între genuri în ceea ce privește valoarea medie a achizițiilor.

Grafic: Valoarea achiziției medie în funcție de gen

- Femeile (60,25 USD) cheltuie, în medie, mai mult decât bărbații (59,54 USD).

Confirmare parțială: Diferența este prezentă, dar nu semnificativă statistic. Totuși, poate indica o tendință de consum diferit între genuri.

Ipoteza H1 este susținută parțial.

Ipoteza H2: Vârsta influențează semnificativ atât frecvența, cât și valoarea achizițiilor.

Grafic 1: Frecvența cumpărăturilor în funcție de grupa de vârstă

- Categoriile 36–50 și 51–65 ani au cea mai ridicată frecvență de achiziție.
- Categoriile 18–25 și 65+ achiziționează mai rar.

Grafic: Valoarea achiziției în funcție de grupa de vârstă

- Cea mai mare valoare medie a achiziției se regăsește la categoriile 18–25 (60,65 USD) și 51–65 (60,28 USD).
- Cea mai mică valoare: 65+ (58,88 USD).

Confirmare deplină: vârsta influențează ambele comportamente – și frecvența, și valoarea.

Ipoteza H2 este confirmată.

Ipoteza H3: Abonații manifestă o frecvență mai mare a achizițiilor decât neabonații.

Grafic: Frecvența cumpărăturilor în funcție de statutul de abonament (valori absolute)

- Neabonații au valori absolute mai mari, dar sunt și mult mai numeroși (de 2,5–3x mai mulți).

Grafic: Distribuția procentuală a frecvenței cumpărăturilor

- Abonații cumpără mai frecvent în procente pentru „Weekly”, „Fortnightly”, „Every 3 Months” etc.

Confirmare prin proporționalitate: Deși în termeni absoluți neabonații domină, proporțional, abonații cumpără mai des.

Ipoteza H3 este confirmată.

Ipoteza H4: Abonații cheltuie mai mult în medie decât neabonații.

Grafic: Valoarea achiziției în funcție de statutul de abonament

- Neabonații: 59,87 USD
- Abonații: 59,49 USD

Infirmare: Abonații cheltuie ușor mai puțin decât neabonații, în medie.

Ipoteza H4 este respinsă.

6.6. Analiză inferențială: teste ANOVA și regresie liniară

Pentru a completa analiza descriptivă și a valida ipotezele formulate, a fost realizată o analiză inferențială prin aplicarea unor **teste ANOVA unifactoriale** și a unei **regresii liniare multiple**, având ca variabilă dependentă **valoarea achiziției (Purchase Amount - USD)**.

Testele ANOVA

Au fost testate ipoteze referitoare la diferențele dintre medii, în funcție de următoarele variabile categoriale:

- **Genul** respondenților (masculin / feminin)
- **Statutul de abonament** (abonat / neabonat)
- **Grupa de vârstă** (<25, 25–35, 36–50, 51–65, 65+)

Rezultate:

- ANOVA pentru gen: $F(1, 3898) = 0.77$, $p = 0.381$
- ANOVA pentru abonament: $F(1, 3898) = 0.19$, $p = 0.662$
- ANOVA pentru grupa de vârstă: $F(4, 3895) = 0.69$, $p = 0.597$

Interpretare: În toate cele trei cazuri, valorile p sunt mai mari decât pragul de semnificație statistică ($p > 0.05$), ceea ce indică faptul că **nu există diferențe semnificative statistice** între medii. Astfel, nu se poate afirma că genul, vârsta sau statutul de abonament influențează în mod semnificativ **valoarea medie a unei achiziții**.

Regresie liniară multiplă

A fost aplicat un model de regresie liniară pentru a testa efectul combinat al variabilelor **gen, vârstă și abonament** asupra **valorii achiziției**.

Rezultate cheie:

- Coeficient de determinare: $R^2 = 0.000$ → modelul nu explică variația valorii achizițiilor
- Valori p pentru coeficienți:
 - Gen: $p = 0.445$
 - Vârstă: $p = 0.517$
 - Abonament: $p = 0.941$

Interpretare: Niciuna dintre variabilele incluse nu este un predictor semnificativ al valorii achiziției. Modelul are un grad foarte redus de explicabilitate ($R^2 \approx 0$), ceea ce indică faptul că aceste trei variabile nu prezintă o relație liniară cu valoarea cheltuielilor.

Concluzii intermediare

Rezultatele inferențiale **nu susțin ipotezele H1, H2 și H3** în ceea ce privește influența genului, vârstei și abonamentului asupra **valorii achiziției**. Deși analiza descriptivă sugera unele diferențe, acestea nu sunt confirmate statistic. Este posibil ca aceste variabile să influențeze alte dimensiuni ale comportamentului de cumpărare (ex. **frecvența achizițiilor** sau **tipul produselor**), care necesită modele de analiză diferite (ex. regresie ordinală, regresie logistică sau segmentare).

Discuții pe baza analizei inferențiale

Analiza inferențială a avut ca scop testarea semnificației statistice a relațiilor dintre variabilele demografice și comportamentale și valoarea achizițiilor realizate de clienți. În ciuda diferențelor observate în analiza descriptivă, diferențe vizibile în medii, mediane și distribuții grafice, rezultatele testelor ANOVA și ale regresiei liniare sugerează că aceste diferențe **nu sunt suficient de puternice pentru a fi considerate semnificative din punct de vedere statistic**.

Mai exact:

- Genul, vârsta și statutul de abonament nu influențează în mod semnificativ valoarea medie a unei achiziții.
- Modelul de regresie nu reușește să explice variația valorii achizițiilor, având un coeficient R^2 aproape inexistent.

Aceste rezultate pot avea mai multe explicații:

- **Valoarea achiziției** este o variabilă influențată de factori contextuali mai greu de surprins în setul de date actual: venitul utilizatorului, ocazionalitatea cumpărăturii, promoțiile disponibile, tipologia produselor etc.
- Variabilele explicative incluse (gen, vârstă, abonament) pot avea un rol mai relevant în **alte dimensiuni ale comportamentului de cumpărare**, cum ar fi **frecvența, loialitatea, categoriile de produse cumpărate sau timpul dintre achiziții**.
- De asemenea, este posibil ca **distribuția inegală a eșantionului** pe anumite categorii (ex. vârstă peste 65 ani, abonați tineri) să fi redus puterea statistică a testelor.

De remarcat este și faptul că o parte dintre concluziile descriptive (ex. frecvența mai mare în rândul abonaților) nu au fost testate în această etapă inferențială și ar putea fi supuse unor alte metode analitice, cum ar fi **regresia ordinală** (pentru frecvența exprimată în trepte) sau testul chi-pătrat (pentru asociații între variabile categoriale).

7. Discuții

Rezultatele cercetării oferă o perspectivă multidimensională asupra comportamentului de cumpărare, analizat prin prisma genului, vârstei și statutului de abonament. Combinând analiza descriptivă cu metode inferențiale (teste ANOVA și regresie liniară), lucrarea relevă atât tendințele vizibile în date, cât și limitele semnificației statistice ale acestora.

7.1. Rolul vârstei în comportamentul de cumpărare

Vârsta este variabila cu cea mai puternică asociere descriptivă: tinerii sub 35 de ani achiziționează mai frecvent, iar grupul 36–50 prezintă cele mai mari valori medii per achiziție. Cu toate acestea, testele ANOVA nu validează aceste diferențe ca fiind semnificative statistic în ceea ce privește valoarea achiziției, ceea ce sugerează că alte dimensiuni (frecvență, tip produs) pot fi mai relevante în explicarea influenței vârstei.

7.2. Genul ca variabilă moderatoare

Deși distribuția valorii achizițiilor diferă ușor între genuri, cu o variabilitate mai mare în rândul femeilor, testele ANOVA arată că genul nu influențează semnificativ valoarea medie a cumpărăturilor. Acest rezultat sugerează că, deși pot exista diferențe comportamentale subtile, acestea nu se traduc într-un impact economic consistent la nivelul valorii tranzacțiilor.

7.3. Abonamentul între loialitate și valoare

Abonații prezintă o frecvență semnificativ mai mare a achizițiilor, validată vizual și procentual, ceea ce sprijină parțial ipoteza H3. Totuși, nici testele ANOVA, nici regresia nu indică o diferență semnificativă în valoarea achizițiilor între abonați și neabonați. Acest contrast între frecvență și valoare poate reflecta o strategie orientată pe retenție și cantitate, nu pe valoare unitară.

7.4. Interacțiuni între variabile

Modelul de regresie multiplă nu a arătat efecte semnificative ale combinațiilor gen–vârstă–abonament asupra valorii achiziției. Totuși, observațiile descriptive sugerează că anumite segmente (ex. femeile între 25–50 ani sau tinerii abonați) pot manifesta tipare distincte, ce merită explorate în cercetări ulterioare cu modele factoriale mai robuste.

8. Limite ale cercetării

8.1. Dezechilibrul eșantionului pe segmente

- Subreprezentarea unor combinații între variabile (ex. bărbați abonați peste 65 de ani) limitează capacitatea de generalizare și reduce puterea statistică a testelor aplicate.

8.2. Ambiguitatea statutului de abonament

Neclaritatea privind ce presupune „abonamentul” (newsletter, cont plătit, program de loialitate etc.) poate altera interpretarea semnificației acestuia în comportamentul de cumpărare.

8.3. Analiză transversală

Cercetarea surprinde doar un moment fix (cross-sectional), fără a evidenția dinamica temporală sau evoluțiile comportamentale.

8.4. Limitările analizei inferențiale

Deși au fost aplicate teste ANOVA și regresie, acestea s-au concentrat doar pe valoarea achiziției. Frecvența, ca variabilă ordinală, ar necesita metode specifice (ex. regresie ordinală sau logistică), care nu au fost incluse.

8.5. Lipsa variabilelor contextuale

Factori precum venitul, tipologia produsului, experiența de utilizare sau istoricul de cumpărare nu sunt disponibili în setul de date și ar fi putut îmbogăți modelul analitic.

8.6. Posibilă influență a comportamentului oportunist

Anumite achiziții pot fi influențate de promoții sau oferte limitate, ceea ce distorsionează comportamentele uzuale de cumpărare și poate induce variații artificiale în date.

8. Limite ale cercetării

Orice demers de cercetare, oricât de riguros ar fi, este supus unor constrângeri metodologice, tehnice și conceptuale. Recunoașterea acestor limite nu afectează valoarea investigației, ci oferă context pentru interpretarea rezultatelor și fundamentează direcțiile de îmbunătățire în studiile viitoare.

Cercetarea de față prezintă mai multe limite care trebuie luate în considerare:

8.1. Dezechilibrul eșantionului pe segmente

Distribuția respondenților nu este uniformă pe toate combinațiile de variabile. De exemplu, anumite categorii, cum ar fi femeile sub 25 de ani cu abonament activ, sunt subreprezentate sau absente. Acest dezechilibru limitează capacitatea de a compara în mod echitabil segmentele și reduce validitatea generalizării rezultatelor în acele cazuri.

8.2. Lipsa unei definiții precise a variabilei „Subscription Status”

Setul de date nu oferă o explicație clară pentru ce presupune calitatea de abonat. Este neclar dacă acest statut reflectă:

- înscriserea la un newsletter,
- apartenența la un program de loialitate,
- un abonament plătit (tip Amazon Prime) sau
- simpla activare a unui cont.

Această ambiguitate creează dificultăți în interpretarea corectă a influenței acestei variabile asupra comportamentului de cumpărare.

8.3. Analiză transversală (cross-sectional)

Cercetarea se bazează pe o fotografie de moment („snapshot”), adică o analiză a comportamentului de cumpărare într-o singură perioadă de timp. Aceasta împiedică identificarea unor tendințe evolutive, modificări sezoniere sau efecte ale campaniilor de marketing desfășurate în timp.

8.4. Absența analizelor statistice inferențiale

Datorită caracterului exploratoriu și a limitărilor privind dimensiunea segmentelor, nu s-au aplicat teste statistice de semnificație (ex. t-test, ANOVA, regresii). În consecință, diferențele observate în grafice trebuie interpretate ca tendințe, nu ca diferențe confirmate statistic. Această limitare poate fi depășită în cercetări viitoare prin extinderea eșantionului și aplicarea unui design experimental robust.

8.5. Variabile contextuale neincluse

Setul de date nu conține variabile importante precum:

- venitul lunar al clientului,
- tipologia produselor cumpărate (ex. electrocasnice vs. fashion),
- canalele de comunicare prin care clientul a fost atras,
- satisfacția globală față de platformă.

Acestea ar fi putut contribui la o înțelegere mai profundă a comportamentului și la construirea unor modele predictive mai precise.

8.6. Posibila autoreglare comportamentală

Nu este exclus ca unii utilizatori să fi făcut achiziții doar pentru a beneficia de un discount sau ofertă limitată, ceea ce poate altera ritmul lor obișnuit de cumpărare. Aceste comportamente punctuale, dacă nu sunt identificate și izolate, pot influența interpretarea frecvenței și valorii achizițiilor.

Prin luarea în considerare a acestor limite, interpretarea rezultatelor poate fi făcută cu precauție, iar cercetările viitoare pot fi calibrate mai eficient pentru a adresa aceste aspecte și a oferi concluzii mai solide și generalizabile.

9. Concluzii și recomandări

9.1 Concluzii

Cercetarea de față a urmărit să analizeze variațiile comportamentului de cumpărare online în funcție de gen, vârstă și statutul de abonament, folosind o bază de date compusă din 3.900 de observații dintr-o platformă de e-commerce. Analiza, susținută de vizualizări de date, teste descriptive și exploratorii, a condus la o serie de concluzii relevante privind dinamica consumului în mediul digital.

În primul rând, **genul a demonstrat o influență modestă asupra valorii medii a achizițiilor**, femeile înregistrând o medie ușor mai ridicată decât bărbații. Totuși, diferențele nu au fost suficient de semnificative pentru a valida pe deplin ipoteza, ci doar pentru a semnală o tendință posibilă în preferințele de consum.

În schimb, **vârsta s-a dovedit un factor esențial**, influențând atât frecvența, cât și valoarea achizițiilor. Persoanele din segmentele de vârstă 36–50 și 51–65 au fost cele mai active în termeni de frecvență, în timp ce tinerii cu vârste între 18–25 de ani au realizat achizițiile cu cele mai mari valori medii. Aceste rezultate susțin ideea că vârsta este un predictor important al comportamentului de cumpărare și evidențiază potențialul de segmentare a publicului țintă în funcție de ciclurile de viață și de disponibilitatea financiară.

Un alt rezultat notabil a fost obținut în ceea ce privește **statutul de abonament**. Deși la nivel absolut neabonații au înregistrat un număr mai mare de achiziții, analiza procentuală a demonstrat că **abonații prezintă o frecvență de cumpărare mai ridicată raportat la propria categorie**, ceea ce confirmă un nivel sporit de loialitate și implicare. Acest aspect este deosebit de important pentru strategiile de retenție și fidelizare, întrucât sugerează că investiția în dezvoltarea programelor de abonament poate duce la o activitate comercială mai constantă.

Totuși, **valoarea medie a achizițiilor nu a fost mai mare în rândul abonaților**, contrar ipotezei formulate. Dimpotrivă, neabonații au cheltuit ușor mai mult, ceea ce poate indica o tendință a abonaților de a realiza achiziții mai frecvente, dar mai mici. Astfel, ipoteza H4 este infirmată, oferind un punct de plecare valoros pentru analize viitoare privind motivațiile și obiceiurile de consum ale abonaților.

9.2 Recomandări

Pe baza rezultatelor obținute, pot fi formulate următoarele recomandări aplicabile în contextul marketingului digital și al strategiilor comerciale:

- **Segmentarea pe grupe de vârstă** ar trebui să stea la baza campaniilor de comunicare, întrucât diferitele categorii de vârstă manifestă comportamente de cumpărare distincte – atât în ceea ce privește valoarea, cât și frecvența achizițiilor.
- **Genul**, deși nu un factor determinant, poate contribui la personalizarea ofertei, în special în industriile unde diferențele de preferință sunt mai evidente (ex. fashion, cosmetice).
- **Abonații trebuie tratați ca un segment strategic**, cu un potențial ridicat de recurență. Se recomandă dezvoltarea de programe de abonament care să nu se bazeze exclusiv pe beneficii logistice (ex: livrare rapidă sau gratuită), ci și pe stimulente financiare (discounturi cumulative, bonusuri pentru achiziții recurente).
- **Pentru creșterea valorii medii a coșului de cumpărături**, ar putea fi testate mecanisme de upselling și cross-selling dedicate exclusiv abonaților.
- În ceea ce privește **analizele viitoare**, se recomandă extinderea datasetului și aplicarea de modele predictive (regresii, analize multivariate) pentru a anticipa comportamentele de cumpărare în funcție de factori suplimentari: sezon, tip de produs, canal de achiziție etc.

10. Bibliografie

Bakewell, C., & Mitchell, V. W. (2003). Generation Y female consumer decision-making styles. *International Journal of Retail & Distribution Management*, 31(2), 95–106. <https://doi.org/10.1108/09590550310461994>

Dittmar, H., Long, K., & Meek, R. (2004). Buying on the internet: Gender differences in online and conventional buying motivations. *Sex Roles*, 50(5–6), 423–444. <https://doi.org/10.1023/B:SERS.0000018896.35251.c7>

Yang, Z., & Peterson, R. T. (2004). Customer perceived value, satisfaction, and loyalty: The role of switching costs. *Psychology & Marketing*, 21(10), 799–822. DOI: 10.1002/mar.20030

Liu, Y. (2007). The long-term impact of loyalty programs on consumer purchase behavior and loyalty. *Journal of Marketing*, 71(4), 19–35. <https://doi.org/10.1509/jmkg.71.4.019>

Rodgers, S., & Harris, M. A. (2003). Gender and e-commerce: An exploratory study. *Journal of Advertising Research*, 43(3), 322–329. <https://doi.org/10.1017/S002184990303031X>

Yang, S., Lu, Y., Gupta, S., Cao, Y., & Zhang, R. (2019). Mobile payment services adoption across time: An empirical study of the effects of behavioral beliefs, social influences, and personal traits. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2019.06.009>

ANEXE

Link Kaggle:

<https://www.kaggle.com/datasets/sahilislam007/shopping-trends-and-customer-behaviour-dataset>

Link Bază de date:

https://drive.google.com/file/d/1N_dMpesthypbu1DLC83GRJP__pRN6owv/view?usp=share_link

Link Bază de date curăţată:

https://drive.google.com/file/d/1UafUBjgfVQviLMwDsBrA86pZryZNSsZX/view?usp=share_link