**ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI**

**FACULTATEA DE CIBERNETICĂ, STATISTICĂ ȘI INFORMATICĂ ECONOMICĂ**

**INFORMATICĂ-ECONOMICĂ**

Logo

Description automatically generated

**PROIECT PACHETE SOFTWARE**

**Analiza Datelor Comerciale și Comportamentale în Retail:**

**Studiu de Caz Walmart**

Studenți: Neacșu David

Patachi Amalia

Seria D, grupa 1096

# **Introducere**

## Organizația Analizată

Walmart Inc. este una dintre cele mai mari și mai cunoscute companii de retail din lume, cu sediul în Statele Unite ale Americii. Fondată în 1962 de Sam Walton, compania s-a dezvoltat rapid și a devenit lider mondial în domeniul comerțului cu amănuntul, operând mii de magazine în numeroase țări. Walmart oferă o gamă largă de produse, de la alimente și articole de uz casnic, până la electronice și haine, adresându-se unui segment vast de consumatori.

Prin strategia sa axată pe prețuri competitive și eficiență operațională, Walmart a reușit să devină un jucător-cheie în industrie, influențând semnificativ piața globală. Totodată, compania investește constant în tehnologii și soluții digitale pentru optimizarea lanțului de aprovizionare și îmbunătățirea experienței clienților.

Acest proiect analizează datele de vânzări și stocuri ale Walmart, cu scopul de a evidenția modalitățile prin care vizualizarea eficientă a informațiilor poate sprijini deciziile economice legate de gestionarea inventarului și optimizarea operațiunilor.

## Problema Economică

În contextul actual al pieței retail, companiile se confruntă cu provocarea de a gestiona volume mari de date legate de vânzări, stocuri și comportamentul clienților. Aceste date sunt esențiale pentru optimizarea operațiunilor, dar, în lipsa unor instrumente adecvate de analiză și vizualizare, devin dificil de interpretat și utilizat eficient.

Astfel, problema economică centrală este reprezentată de:

* **Transformarea datelor complexe în informații clare și acționabile** pentru sprijinirea deciziilor strategice și operaționale.
* **Vizualizarea eficientă a evoluției stocurilor și a vânzărilor** pe produse și regiuni, pentru identificarea rapidă a eventualelor probleme precum lipsa stocurilor sau fluctuațiile cererii.
* **Asigurarea unei baze solide pentru luarea deciziilor** privind reaprovizionarea, campaniile promoționale și gestionarea portofoliului de produse.

Prin adresarea acestei probleme, companiile pot reduce costurile asociate cu stocurile neoptime, pot crește satisfacția clienților și pot îmbunătăți performanța economică generală.

## Tehnologii utilizate

* **Python** — pentru procesarea și analiza datelor, precum și pentru crearea de scripturi de automatizare.
* **Streamlit** — pentru dezvoltarea rapidă a interfețelor web interactive destinate vizualizării și explorării datelor.
* **SAS** — pentru gestionarea și prelucrarea datelor complexe, folosind proceduri statistice și SQL.
* **SAS Studio** — mediul integrat de dezvoltare folosit pentru scrierea, testarea și executarea codului SAS.

# **Vizualizarea Datelor în Streamlit**

1. Încărcare date

@st.cache\_data

def load\_data():

return pd.read\_csv("./Walmart.csv", parse\_dates=['transaction\_date'])

df = load\_data()

st.title("Dashboard Vânzări Walmart")

1. Dashboard principal cu filtre

st.sidebar.header("Filter")

store = st.sidebar.multiselect("Store Location", df["store\_location"].unique())

category = st.sidebar.multiselect("Category", df["category"].unique())

filtered\_df = df.copy()

if store:

filtered\_df = filtered\_df[filtered\_df["store\_location"].isin(store)]

if category:

filtered\_df = filtered\_df[filtered\_df["category"].isin(category)]

A screenshot of a computer

AI-generated content may be incorrect.

**Interpretare:**

Dashboard-ul urmărește să ofere o privire de ansamblu asupra performanței vânzărilor Walmart — cu posibilitatea de filtrare pe:

* locația magazinului (Store Location)
* categoria produsului (Category)

1) Total vânzări: $15,263,601.45

* Reprezintă suma totală a veniturilor generate din vânzări.
* Probabil este calculată pe întregul dataset (adică fără niciun filtru aplicat momentan).
* Sugerează un volum mare de vânzări.

2) Mărimea medie a coșului de cumpărături: 2.98 produse

* Fiecare tranzacție conține în medie ~3 produse.
* Acest indicator este util pentru înțelegerea comportamentului de cumpărare al clienților.

3)Număr de tranzacții: 5000

* Totalul tranzacțiilor înregistrate în baza de date.

1. Vânzări totale de-a lungul timpului – strat sezonier

# Adăugăm o coloană pentru anotimp

def get\_season(date):

month = date.month

if month in [12, 1, 2]:

return "Iarnă"

elif month in [3, 4, 5]:

return "Primăvară"

elif month in [6, 7, 8]:

return "Vară"

else:

return "Toamnă"

sales\_by\_date = (filtered\_df.groupby('date')['total\_sales']

.sum().reset\_index())

sales\_by\_date['season'] = pd.to\_datetime(sales\_by\_date['date']).apply(get\_season)

# Afișăm cu colorare după sezon

fig\_time = px.line(

sales\_by\_date,

x='date',

y='total\_sales',

color='season',

line\_group='season',

title='Vânzări Totale de-a lungul Timpului - Strat Sezonier',

labels={'date': 'Dată', 'total\_sales': 'Vânzări', 'season': 'Anotimp'},

color\_discrete\_sequence=["#2f77c3", "#f8b3ec", "#a4d13a", "#fc5e1d"] # Iarnă, Primăvară, Vară, Toamnă

)

st.plotly\_chart(fig\_time)

A screen shot of a graph

AI-generated content may be incorrect.

**Interpretare:**

**Structura graficului:**

* Axa X (Dată): Zilele din 2024 (ianuarie – septembrie).
* Axa Y (Vânzări): Suma vânzărilor zilnice (total\_sales).
* Culoare: Segmentele sunt colorate în funcție de anotimp.

**Ce observăm:**

Iarnă (albastru):

* Variații mari, dar vânzări relativ constante între 40k–90k.
* Posibile vârfuri după sărbători (ianuarie).

Primăvară (roz):

* Scăderi frecvente și o variabilitate ridicată.
* Vânzările sunt mai mici în medie decât în iarnă, dar mai multe fluctuații (valuri de cumpărături).

Vară (verde):

* Se mențin vânzări mari și stabile, între 60k–100k.
* Poate fi considerată o perioadă de vârf (sezon de cumpărături active — turism, vacanțe, aer condiționat, băuturi răcoritoare etc.).

Toamnă (portocaliu):

* Deși e parțială (doar septembrie), pare să scadă ușor.
* Ritmul de cumpărături pare în scădere după vară.

**Concluzii importante:**

1. Vară = sezon de vârf al vânzărilor, urmat de iarnă.
2. Primăvara are instabilitate mare (posibilă legătură cu promoții ocazionale, schimbare sezonieră de stocuri).
3. Toamna începe cu un trend descendent — poate influențat de sfârșitul sezonului turistic.
4. Vânzări dupa grupe de vârstă și sex

gender = filtered\_df["customer\_gender"].unique().tolist()

if gender:

filtered\_df = filtered\_df[filtered\_df["customer\_gender"].isin(gender)]

st.subheader("Vânzări după Grupe de Vârstă și Sex")

age\_gender = (

filtered\_df.groupby(['age\_group', 'customer\_gender'])['total\_sales']

.sum().reset\_index()

)

fig\_age = px.bar(age\_gender, x='age\_group', y='total\_sales',

color='customer\_gender', barmode='group',

labels={'age\_group': 'Grupa de Vârstă',

'total\_sales': 'Vânzări Totale',

'customer\_gender': 'Sex'},

title='Vânzări în funcție de vârstă și sex',

color\_discrete\_map={"Male": "#1126a5", "Female": "#ff0000 ", "Other": "#999999"})

st.plotly\_chart(fig\_age)

A graph of different colored bars

AI-generated content may be incorrect.

**Interpretare:**

**Structura graficului:**

* Axa X: Grupa de vârstă (6 intervale: 18-25, 26-35, 36-45, 46-55, 56-65, 65+)
* Axa Y: Total vânzări (în unități monetare)
* Culoare: Sexul clientului.

**Ce observăm:**

Grupa 26–35:

* Este cea mai activă categorie de vârstă din punct de vedere al vânzărilor.
* Femeile (roșu) au vânzări cele mai ridicate, urmate de Other, apoi bărbați.

Grupele 36–65:

* Vânzări ridicate și destul de echilibrate între genuri.
* Categoriile 46–55 și 56–65 ating vârfuri foarte apropiate.

Grupa 65+:

* Cel mai mic volum de vânzări total.
* Ușoară predominanță a bărbaților.

Grupa 18–25:

* Vânzări mai scăzute decât restul, dar vizibil mai apropiate între genuri.

**Concluzii cheie:**

1. Segmentul 26–35 ani este cel mai profitabil — potențial target ideal pentru campanii.
2. Femeile au cele mai mari contribuții în aproape toate grupele de vârstă.
3. Categoria Other are o prezență notabilă, ceea ce înseamnă că incluziunea în marketing e importantă.
4. Clienții peste 65 ani au un impact comercial redus — pot necesita abordări specifice sau pot fi puțin digitalizați.
5. Vânzări în funcție de condiții meteo și timp

st.subheader("Vânzări în funcție de Condiții de Mediu și Timp")

*# a) Weather*

weather\_sales = (filtered\_df.groupby('weather\_conditions')['total\_sales']

.sum().reset\_index())

fig\_weather = px.bar(weather\_sales, x='weather\_conditions', y='total\_sales',

color='weather\_conditions',

labels={'weather\_conditions':'Condiții Meteo',

'total\_sales':'Vânzări Totale'},

title='Vânzări pe Condiții Meteo',

color\_discrete\_sequence=["#ADD8E6", "#9c9ce4", "#808080", "#f9cd48"]) # cloudy, rainy, stormy, sunny

st.plotly\_chart(fig\_weather)

A graph with different colored squares

AI-generated content may be incorrect.

**Interpretare:**

**Ce afișează graficul:**

* Axa X: Condiții meteo – Cloudy, Rainy, Stormy, Sunny
* Axa Y: Vânzări totale cumulate în acele condiții
* Barele sunt colorate sugestiv.

**Concluzii:**

1. Zilele însorite generează cele mai mari vânzări — clienții ies mai des la cumpărături în vreme bună.
2. Vremea înnorată nu descurajează semnificativ cumpărăturile — e comparabilă cu sunny.
3. Furtunile nu reduc vânzările drastic — posibil datorită comenzilor online sau urgențelor.
4. Zilele ploioase scad ușor traficul comercial.

*# b) Holiday vs normal day*

holiday\_sales = (filtered\_df.groupby('holiday\_indicator')['total\_sales']

.sum().reset\_index())

holiday\_sales['holiday\_indicator'] = holiday\_sales['holiday\_indicator'] \

.map({True: 'Sărbătoare', False: 'Zi normală'})

fig\_holiday = px.pie(holiday\_sales, names='holiday\_indicator',

values='total\_sales',

title='Vânzări în Zile de Sărbătoare vs. Zile Normale',

color\_discrete\_sequence=["#fc4545", "#ADD8E6"]) # sărbătoare, zile normale

st.plotly\_chart(fig\_holiday)

A screen shot of a graph

AI-generated content may be incorrect.

**Interpretare:**

**Date afișate:**

* Roșu: Sărbătoare → 50.2% din totalul vânzărilor
* Albastru deschis: Zi normală → 49.8% din totalul vânzărilor

**Ce observăm:**

* Vânzările în zilele de sărbătoare sunt ușor mai mari, dar diferența este foarte mică (~0.4%).
* Reiese că vânzările sunt constante, indiferent de tipul zilei.

*# c) Weekday*

weekday\_order = ["Monday","Tuesday","Wednesday",

"Thursday","Friday","Saturday","Sunday"]

weekday\_sales = (filtered\_df.groupby('weekday')['total\_sales']

.sum().reindex(weekday\_order).reset\_index())

fig\_weekday = px.line(weekday\_sales, x='weekday', y='total\_sales',

markers=True,

labels={'weekday':'Zi','total\_sales':'Vânzări'},

title='Vânzări după Ziua Săptămânii',

color\_discrete\_sequence=["#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#9467bd", "#8c564b", "#e377c2"]) # Fiecare zi săptămână cu o culoare diferită

st.plotly\_chart(fig\_weekday)

A graph with blue lines

AI-generated content may be incorrect.

**Interpretare:**

**Ce arată graficul:**

* Axa X: Zilele săptămânii (Luni → Duminică).
* Axa Y: Totalul vânzărilor.
* Stil: Linie cu marcatori pentru fiecare punct de date.

**Concluzii:**

1. Vârfuri de vânzări apar Lunea și Joia — posibil corelate cu:
   * aprovizionare de început de săptămână
   * oferte de tip „midweek”
2. Cele mai slabe zile sunt Vineri și Duminică — potențial:
   * Vineri: tranziție spre weekend → scade consumul
   * Duminică: magazine închise sau trafic redus
3. Sâmbăta are o revenire, dar nu suficientă pentru a egala Luni/Joi.
4. Trendul vânzărilor în funcție de promoții

st.subheader("Trendul Vânzărilor în Funcție de Promoții")

*# a) Cu vs. Fără promoție*

promo\_trend = (filtered\_df.groupby(['date', 'promotion\_applied'])['total\_sales']

.sum().reset\_index())

promo\_trend['promotion\_applied'] = promo\_trend['promotion\_applied'] \

.map({True: 'Cu Promoție', False: 'Fără Promoție'})

fig\_promo = px.line(promo\_trend, x='date', y='total\_sales',

color='promotion\_applied',

labels={'date':'Dată','total\_sales':'Vânzări'},

title='Trend zilnic: Cu vs. Fără Promoție',

color\_discrete\_sequence=["#2f77c3", "#fc4545"]) # fără promoție, cu promoție

st.plotly\_chart(fig\_promo)

A graph of red and blue lines

AI-generated content may be incorrect.

**Interpretare:**

**Structura graficului:**

* Axa X (Dată): Zilele din ianuarie până în septembrie 2024.
* Axa Y (Vânzări): Totalul vânzărilor zilnice.
* Linii:
  + Albastru = Fără Promoție
  + Roșu = Cu Promoție

**Ce observăm:**

1. Fluctuații mari în ambele cazuri – nu există un trend dominant evident.
2. Zilele cu promoții (roșu) tind să aibă vârfuri mai înalte, în special în anumite perioade (februarie, mai, august).
3. Există perioade (ex. martie-aprilie) în care zilele fără promoții depășesc ocazional vânzările cu promoții.
4. Spre final (august-septembrie), promoțiile par să aibă un impact mai clar — vârfurile roșii sunt mai frecvente și mai înalte.

**Concluzii:**

1. Promoțiile au potențialul de a genera vârfuri de vânzări, dar nu constant — depinde de moment și context.
2. Nu toate promoțiile aduc un beneficiu clar — unele zile fără promoții pot genera vânzări comparabile sau mai mari.
3. Poate fi nevoie de optimizarea momentului în care sunt aplicate promoțiile și de analizat tipul de produs vizat.

*# b) Pe tipuri de promoție (doar unde există promoții)*

type\_trend = (filtered\_df[filtered\_df['promotion\_applied']]

.groupby(['date','promotion\_type'])['total\_sales']

.sum().reset\_index())

if not type\_trend.empty:

fig\_type = px.line(type\_trend, x='date', y='total\_sales',

color='promotion\_type',

labels={'date':'Dată','total\_sales':'Vânzări'},

title='Trend zilnic pe Tipuri de Promoție',

color\_discrete\_sequence=["#2f77c3", "#fc4545"]) # Set1 oferă culori distincte pentru fiecare tip de promoție

st.plotly\_chart(fig\_type)

else:

st.info("Nu există date cu promoții pentru filtrele selectate.")

A graph with red and blue lines

AI-generated content may be incorrect.

**Interpretare:**

**Ce arată graficul:**

* Axa X: Data calendaristică (ianuarie – septembrie 2024)
* Axa Y: Total vânzări în acea zi (doar când a fost aplicată o promoție)
* Două linii de trend:
  + Albastru: BOGO
  + Roșu: Reducere procentuală

**Ce observăm:**

1. BOGO (albastru):
   * Atinge vârfuri mari izolate, de până la 25.000.
   * Pare să genereze vârfuri ocazionale mai puternice, dar cu distribuție mai rarefiată.
2. Reducere procentuală (roșu):
   * Are prezență mai frecventă pe axa timpului.
   * Vânzări mai stabile, dar cu vârfuri mai mici decât BOGO în general.

**Concluzii cheie:**

1. BOGO creează un efect de impuls — vârfuri mari în anumite zile → eficient pe termen scurt, dar instabil.
2. Reducerile procentuale sunt mai consecvente, dar cu impact moderat per zi.
3. Eficiența unei promoții depinde de strategia aplicată:
   * BOGO = excelent pentru lichidări de stoc sau atragere masivă.
   * Discount % = bun pentru stimulare constantă, mai ales în perioade cu cerere normală.

# **Generarea de Rapoarte în SAS Studio**

proc import datafile="/home/u64222860/proiectPachete/Walmart.csv"

out=walmart

dbms=csv

replace;

getnames=yes;

run;

/\* stocuri epuizate per locatie \*/

data stockouts;

set walmart;

if stockout\_indicator = 'TRUE';

run;

proc freq data=stockouts;

tables store\_location / out=stockouts\_per\_location;

run;

/\* Venit per tip promoție\*/

data promo;

set walmart;

if promotion\_applied = 'TRUE' then revenue = quantity\_sold \* unit\_price;

else delete;

run;

proc means data=promo noprint;

class promotion\_type;

var revenue;

output out=revenue\_by\_promo\_type sum=revenue;

run;

proc sql;

select product\_name, sum(quantity\_sold) as total\_sold,

sum(quantity\_sold \* unit\_price) as total\_revenue\_per\_product

from walmart

group by product\_name;

/\* venituri per categorie \*/

proc sql;

select category, sum(quantity\_sold \* unit\_price) as total\_revenue

from walmart

group by category;

/\* venituri totale \*/

proc sql;

select store\_location, sum(quantity\_sold) as total\_sold

from walmart

group by store\_location;

/\* venituri per magazin \*/

proc sql;

select store\_id, sum(quantity\_sold \* unit\_price) as revenue

from walmart

group by store\_id

order by revenue desc;  
  
**The FREQ Procedure**

| **store\_location** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **Chicago, IL** | 519 | 20.02 | 519 | 20.02 |
| **Dallas, TX** | 496 | 19.13 | 1015 | 39.14 |
| **Los Angeles, CA** | 542 | 20.90 | 1557 | 60.05 |
| **Miami, FL** | 489 | 18.86 | 2046 | 78.90 |
| **New York, NY** | 547 | 21.10 | 2593 | 100.00 |

| **product\_name** | **total\_sold** | **total\_revenue\_per\_product** |
| --- | --- | --- |
| Camera | 1873 | 1895104 |
| Fridge | 1967 | 1938013 |
| Headphones | 1816 | 1846334 |
| Laptop | 1685 | 1709159 |
| Smartphone | 1876 | 1931310 |
| TV | 1926 | 2049494 |
| Tablet | 1964 | 1996253 |
| Washing Machine | 1807 | 1897934 |

| **category** | **total\_revenue** |
| --- | --- |
| Appliances | 7321970 |
| Electronics | 7941632 |

| **store\_location** | **total\_sold** |
| --- | --- |
| Chicago, IL | 3085 |
| Dallas, TX | 2892 |
| Los Angeles, CA | 3168 |
| Miami, FL | 2834 |
| New York, NY | 2935 |

| **store\_id** | **revenue** |
| --- | --- |
| 1 | 847019 |
| 11 | 831769.4 |
| 5 | 830071.7 |
| 17 | 822912.1 |
| 9 | 813169.5 |
| 2 | 796410.6 |
| 10 | 791735.6 |
| 20 | 789721.4 |
| 15 | 781436.4 |
| 8 | 775187.4 |
| 19 | 750516.5 |
| 16 | 739372.9 |
| 3 | 738115 |
| 14 | 726846.6 |
| 13 | 722058.9 |
| 7 | 720463.9 |
| 18 | 719431.6 |
| 12 | 704529.5 |
| 4 | 700333.7 |
| 6 | 662499.9 |

**/\* 1. Definirea formatelor \*/**

proc format;

value age\_groupe

low - <25 = 'Sub 25'

25 - <35 = '25-34'

35 - <45 = '35-44'

45 - <60 = '45-59'

60 - high = '60+';

value income\_groupe

low - <30000 = 'Sub 30k'

30000 - <60000 = '30k-59k'

60000 - <100000 = '60k-99k'

100000 - high = '100k+';

run;

**/\* 2. Creare set nou cu variabile grupate \*/**

data walmart\_grouped;

set walmart;

age\_group = put(customer\_age, age\_groupe.);

income\_group = put(customer\_income, income\_groupe.);

run;

**/\* 3. Rapoarte PROC TABULATE \*/**

proc tabulate data=walmart\_grouped format=6.2;

class customer\_loyalty\_level customer\_gender;

table

customer\_loyalty\_level='Nivel Loialitate' all='Total',

customer\_gender='Gen' \* (N='Număr' colpctn ='Procent (%)'\*f=8.2)

/ misstext='0%';

run;

proc tabulate data=walmart\_grouped format=6.2;

class customer\_loyalty\_level age\_group;

table

customer\_loyalty\_level='Nivel Loialitate' all='Total',

age\_group='Grupă de vârstă' \* (N='Număr' colpctn ='Procent (%)'\*f=8.2)

/ misstext='0%';

run;

proc tabulate data=walmart\_grouped format=6.2;

class customer\_loyalty\_level income\_group;

table

customer\_loyalty\_level='Nivel Loialitate' all='Total',

income\_group='Venit' \* (N='Număr' colpctn ='Procent (%)'\*f=8.2)

/ misstext='0%';

run;

| **Gen** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Female** | | **Male** | | **Other** | |
| **Număr** | **Procent (%)** | **Număr** | **Procent (%)** | **Număr** | **Procent (%)** |
| **Nivel Loialitate** | 396 | 23.91 | 418 | 24.85 | 439 | 26.41 |
| **Bronze** |
| **Gold** | 384 | 23.19 | 406 | 24.14 | 370 | 22.26 |
| **Platinum** | 422 | 25.48 | 425 | 25.27 | 452 | 27.20 |
| **Silver** | 454 | 27.42 | 433 | 25.74 | 401 | 24.13 |
| **Total** | 1656 | 100.00 | 1682 | 100.00 | 1662 | 100.00 |

**Concluzii:**

* Genul „Other” are cele mai multe persoane în nivelul Platinum, ceea ce poate indica o loialitate ridicată în acest segment.
* Femeile au un procent mare la Silver, iar bărbații sunt distribuiți relativ uniform.
* Distribuția este echilibrată, dar Platinum pare să concentreze cei mai implicați clienți din toate genurile.

|  | **Grupă de vârstă** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **25-34** | | **35-44** | | **45-59** | | **60+** | | **Sub 25** | |
| **Număr** | **Procent (%)** | **Număr** | **Procent (%)** | **Număr** | **Procent (%)** | **Număr** | **Procent (%)** | **Număr** | **Procent (%)** |
| **Nivel Loialitate** | 239 | 25.16 | 232 | 25.61 | 364 | 25.45 | 272 | 25.68 | 146 | 22.29 |
| **Bronze** |
| **Gold** | 200 | 21.05 | 223 | 24.61 | 328 | 22.94 | 245 | 23.14 | 164 | 25.04 |
| **Platinum** | 260 | 27.37 | 219 | 24.17 | 384 | 26.85 | 272 | 25.68 | 164 | 25.04 |
| **Silver** | 251 | 26.42 | 232 | 25.61 | 354 | 24.76 | 270 | 25.50 | 181 | 27.63 |
| **Total** | 950 | 100.00 | 906 | 100.00 | 1430 | 100.00 | 1059 | 100.00 | 655 | 100.00 |

**Concluzii:**

* Tinerii (Sub 25) tind să fie mai loiali în nivelul Silver → poate au fost recent recrutați în programe.
* Platinum domină la vârstele 25–34 și 45–59, adică segmente mature și active financiar.
* Segmentul 60+ este bine reprezentat în Platinum și Bronze, deci e un public valoros.

|  | **Venit** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **100k+** | | **30k-59k** | | **60k-99k** | | **Sub 30k** | |
| **Număr** | **Procent (%)** | **Număr** | **Procent (%)** | **Număr** | **Procent (%)** | **Număr** | **Procent (%)** |
| **Nivel Loialitate** | 253 | 24.88 | 350 | 24.20 | 517 | 25.80 | 133 | 24.95 |
| **Bronze** |
| **Gold** | 235 | 23.11 | 321 | 22.20 | 485 | 24.20 | 119 | 22.33 |
| **Platinum** | 287 | 28.22 | 402 | 27.80 | 469 | 23.40 | 141 | 26.45 |
| **Silver** | 242 | 23.80 | 373 | 25.80 | 533 | 26.60 | 140 | 26.27 |
| **Total** | 1017 | 100.00 | 1446 | 100.00 | 2004 | 100.00 | 533 | 100.00 |

**Concluzii:**

* Clienții cu venituri mari (100k+) sunt preponderent în Platinum → programul reușește să fidelizeze segmente profitabile.
* Cei cu venituri mici (sub 30k) sunt și ei bine reprezentați în Platinum, semn că loialitatea nu depinde doar de venit.
* Silver domină segmentele de venit mediu, deci aici se află majoritatea clienților activi.

proc import datafile="/home/u64232466/proiectPachete/Walmart.csv"

out=walmart

dbms=csv

replace;

getnames=yes;

run;

proc logistic data=walmart;

class

promotion\_type(ref='None')

customer\_loyalty\_level (ref='Bronze') / param=ref;

model customer\_loyalty\_level = promotion\_type

/ link=glogit stb;

run;

**The LOGISTIC Procedure**

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.WALMART |
| **Response Variable** | customer\_loyalty\_level |
| **Number of Response Levels** | 4 |
| **Model** | generalized logit |
| **Optimization Technique** | Newton-Raphson |

| **Number of Observations Read** | 5000 |
| --- | --- |
| **Number of Observations Used** | 5000 |

| **Response Profile** | | |
| --- | --- | --- |
| **Ordered Value** | **customer\_loyalty\_level** | **Total Frequency** |
| **1** | Bronze | 1253 |
| **2** | Gold | 1160 |
| **3** | Platinum | 1299 |
| **4** | Silver | 1288 |

**Logits modeled use customer\_loyalty\_level='Bronze' as the reference category.**

| **Class Level Information** | | | |
| --- | --- | --- | --- |
| **Class** | **Value** | **Design Variables** | |
| **promotion\_type** | **BOGO** | 1 | 0 |
|  | **None** | 0 | 0 |
|  | **Percentage Discount** | 0 | 1 |

| **Model Convergence Status** |
| --- |
| Convergence criterion (GCONV=1E-8) satisfied. |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 13859.255 | 13858.242 |
| **SC** | 13878.807 | 13916.897 |
| **-2 Log L** | 13853.255 | 13840.242 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 13.0127 | 6 | 0.0428 |
| **Score** | 12.9186 | 6 | 0.0443 |
| **Wald** | 12.8949 | 6 | 0.0447 |

| **Type 3 Analysis of Effects** | | | |
| --- | --- | --- | --- |
| **Effect** | **DF** | **Wald Chi-Square** | **Pr > ChiSq** |
| **promotion\_type** | 6 | 12.8949 | 0.0447 |

| **Analysis of Maximum Likelihood Estimates** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **customer\_loyalty\_level** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** | **Standardized Estimate** |
| **Intercept** |  | **Gold** | 1 | -0.00365 | 0.0493 | 0.0055 | 0.9410 |  |
| **Intercept** |  | **Platinum** | 1 | 0.1027 | 0.0481 | 4.5614 | 0.0327 |  |
| **Intercept** |  | **Silver** | 1 | 0.0346 | 0.0489 | 0.5020 | 0.4786 |  |
| **promotion\_type** | **BOGO** | **Gold** | 1 | -0.2810 | 0.1137 | 6.1091 | 0.0134 | -0.0574 |
| **promotion\_type** | **BOGO** | **Platinum** | 1 | -0.1823 | 0.1081 | 2.8434 | 0.0918 | -0.0372 |
| **promotion\_type** | **BOGO** | **Silver** | 1 | -0.0168 | 0.1064 | 0.0249 | 0.8747 | -0.00342 |
| **promotion\_type** | **Percentage Discount** | **Gold** | 1 | -0.1806 | 0.1141 | 2.5044 | 0.1135 | -0.0360 |
| **promotion\_type** | **Percentage Discount** | **Platinum** | 1 | -0.2362 | 0.1123 | 4.4235 | 0.0354 | -0.0471 |
| **promotion\_type** | **Percentage Discount** | **Silver** | 1 | -0.0251 | 0.1094 | 0.0525 | 0.8187 | -0.00500 |

| **Odds Ratio Estimates** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **customer\_loyalty\_level** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **promotion\_type BOGO vs None** | **Gold** | 0.755 | 0.604 | 0.943 |
| **promotion\_type BOGO vs None** | **Platinum** | 0.833 | 0.674 | 1.030 |
| **promotion\_type BOGO vs None** | **Silver** | 0.983 | 0.798 | 1.211 |
| **promotion\_type Percentage Discount vs None** | **Gold** | 0.835 | 0.667 | 1.044 |
| **promotion\_type Percentage Discount vs None** | **Platinum** | 0.790 | 0.634 | 0.984 |
| **promotion\_type Percentage Discount vs None** | **Silver** | 0.975 | 0.787 | 1.208 |

**Interpretare generală:**

* Promoțiile reduc probabilitatea ca un client să aparțină nivelurilor superioare (Gold/Platinum) comparativ cu Bronze.
* Acest lucru contrazice așteptările — poate indica:
  + clienții loiali nu au nevoie de promoții pentru a cumpăra;
  + promoțiile atrag clienți ocazionali, nu fideli.

Analiza logistică multinomială sugerează că aplicarea promoțiilor (BOGO sau reduceri procentuale) este asociată cu o probabilitate mai mică ca un client să fie Gold sau Platinum față de Bronze. Promoțiile par mai atrăgătoare pentru clienți ocazionali, nu pentru cei loiali. Din perspectiva marketingului, acest rezultat indică necesitatea de a personaliza promoțiile pentru segmentele loiale și de a folosi promoțiile standard pentru atragerea de noi clienți.

# **Concluzie**

Acest proiect a avut ca scop înțelegerea mai clară a modului în care datele despre vânzări, promoții și comportamentul clienților pot fi transformate în informații utile pentru decizii de business. Folosind Python și SAS, am analizat tendințele de cumpărare în funcție de anotimp, vârstă, gen, ziua săptămânii sau tipul de promoție, folosindu-ne de datele reale ale firmei Walmart.

Rezultatele arată că vânzările sunt cele mai mari vara și în zilele de luni și joi, iar clienții din grupa 26–35 de ani, în special femeile, contribuie semnificativ la venituri. În același timp, promoțiile nu par să atragă clienți loiali, ci mai degrabă pe cei ocazionali, ceea ce poate fi util pentru planificarea unor strategii de fidelizare mai personalizate.

Prin combinarea vizualizărilor interactive din Python cu analiza statistică din SAS, am reușit să obținem o imagine completă și ușor de interpretat a realității comerciale de la Walmart.