

ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI
FACULTATEA DE CIBERNETICĂ, STATISTICĂ ȘI INFORMATICĂ ECONOMICĂ
SPECIALIZAREA CIBERNETICĂ ECONOMICĂ

PROIECT LA DISCIPLINA PACHETE SOFTWARE

Studiu Analitic asupra Vânzărilor Walmart: Fundamente pentru Expansiune

Profesori coordonatori:

PROF. UNIV. DR. UȚĂ ILEANA ADINA

ASIST.UNIV. NICULAE C ANDREEA-MIHAELA

Student:

LUNGU VANESA-DENISA

București

2025

CUPRINS

Introdúcere	3
CAPITOLUL 1: PROGRAMARE PYTHON.....	5
CAPITOLUL 2: PROGRAMARE SAS	18
CAPITOLUL 3: PROGRAMARE SAS EG.....	28
CONCLUZII FINALE	41

Introducere

Într-un context economic tot mai competitiv, analiza performanței comerciale a devenit esențială pentru companiile care doresc să-și optimizeze activitatea și să identifice noi oportunități de creștere. Alegerea temei acestui proiect – analiza activității unei organizații de retail de mari dimensiuni (Walmart) – este motivată de nevoia reală a companiilor de a înțelege impactul factorilor interni și externi asupra vânzărilor, precum și de a fundamenta decizii strategice legate de extinderea geografică sau diversificarea serviciilor.

Fișierul utilizat în cadrul acestui studiu conține date săptămânale privind vânzările înregistrate de mai multe magazine Walmart, alături de variabile relevante precum temperatura, rata șomajului, indicele prețurilor de consum (CPI) și prezența sărbătorilor. Acest set de date oferă o bază solidă pentru explorarea relațiilor dintre condițiile economice și comportamentul de consum.

Analiza se va desfășura utilizând instrumente din pachetele Python și SAS (SAS EG), care permit procesarea, vizualizarea și modelarea datelor la scară largă. Studiul va include:

Programare Python:

1. utilizarea listelor și a dicționarelor, incluzând metode specifice acestora;
2. utilizarea seturilor și a tuplurilor, incluzând metode specifice acestora;
3. definirea și apelarea unor funcții;
4. utilizarea structurilor condiționale;
5. utilizarea structurilor repetitive;
6. importul unei fișier csv sau json în pachetul pandas;
7. accesarea datelor cu loc și iloc;
8. modificarea datelor în pachetul pandas;
9. utilizarea funcțiilor de grup;
10. tratarea valorilor lipsă;
11. ștergerea de coloane și înregistrări;
12. prelucrări statistice, gruparea și agregarea datelor în pachetul pandas;
13. prelucrarea seturilor de date cu merge / join;
14. reprezentare grafică a datelor cu pachetul matplotlib;
15. utilizarea pachetului scikit-learn (clusterizare, regresie logistică)
16. utilizarea pachetului statmodels (regresie multiplă).

Programare SAS Enterprise Guide:

1. Importul unui fisier non-SAS;
2. Interogări;
3. Joncțiune;
4. Folosirea parametrilor;
5. Rapoarte;
6. Grafice;
7. Prelucrări statistice;
8. Formate definite de utilizator;
9. Crearea unui document compus;
10. Stiluri create de utilizator.

Programare SAS:

1. crearea unui set de date SAS din fișiere externe;
2. crearea și folosirea de formate definite de utilizator;
3. procesarea iterativă și condițională a datelor;
4. crearea de subseturi de date;
5. utilizarea de funcții SAS;
6. combinarea seturilor de date prin proceduri specifice SAS și SQL;
7. utilizarea de masive;
8. utilizarea de proceduri pentru raportare;
9. folosirea de proceduri statistice;
10. generarea de grafice.

Scopul final al proiectului este de a oferi recomandări strategice bazate pe date concrete, care să ajute la optimizarea operațiunilor curente și la conturarea unei direcții de dezvoltare sustenabilă pentru rețeaua analizată.

CAPITOLUL 1

PROGRAMARE PYTHON

#6. importul unei fișier csv sau json în pachetul pandas;

#12.prelucrări statistice, gruparea și agregarea datelor în pachetul pandas;

#13.prelucrarea seturilor de date cu merge / join;

Se dorește importul datelor brute (SET_DE_DATE_WALMART_PS.csv' 'SET_DE_DATE_WALMART_PS_2.csv') într-un format accesibil pentru analiză, iar mai apoi imbinarea informațiilor despre prețul combustibilului (Fuel_Price) cu cele despre vânzări.

	Store	Date	Weekly_Sales	...	CPI	Unemployment	Fuel_Price
0	1	05-02-2010	1643690.90	...	211.096358	8.106	2.572
1	1	05-02-2010	1643690.90	...	211.096358	8.106	2.548
2	1	05-02-2010	1643690.90	...	211.096358	8.106	2.514
3	1	05-02-2010	1643690.90	...	211.096358	8.106	2.561
4	1	05-02-2010	1643690.90	...	211.096358	8.106	2.625
...
3956	10	18-06-2010	1837636.24	...	126.114000	9.524	3.120
3957	10	18-06-2010	1837636.24	...	126.114000	9.524	3.058
3958	10	18-06-2010	1837636.24	...	126.114000	9.524	2.941
3959	10	18-06-2010	1837636.24	...	126.114000	9.524	3.057
3960	10	18-06-2010	1837636.24	...	126.114000	9.524	2.935

Dupa imbinarea celor doua seturi de date, se dorește aflarea valorilor medii pentru combustibil si valorilor maxime pentru vanzarile saptamanale si temperature (respective minime) pentru fiecare magazin.

	Fuel_Price	Weekly_Sales	Temperature	
	mean	max	min	max
Store				
1	2.708316	1643690.90	38.51	80.69
2	2.704750	2137809.50	38.49	85.81
3	2.704750	461622.22	45.71	83.52
4	2.716500	2188307.39	28.84	81.53
5	2.704750	337825.89	39.70	83.91
6	2.704750	1857380.09	40.43	83.89
7	2.695650	574450.23	10.53	63.59
8	2.704750	1004137.09	33.34	82.21
9	2.704750	586061.46	37.08	83.75
10	3.010600	2193048.75	49.96	89.67

Prețul combustibilului este relativ constant pentru magazinele 1–9: aproximativ 2.70 USD/gallon.

Magazinul 10 are un preț semnificativ mai mare: 3.01 USD/gallon, ceea ce sugerează o locație diferită (probabil urbană sau într-o zonă cu taxe mai mari).

Cele mai mari vânzări maxime sunt la:

Magazinul 10 – 2,193,048.75 USD

Magazinul 2 – 2,137,809.50 USD

Magazinul 4 – 2,188,307.39 USD

Cele mai mici sunt la:

Magazinul 5 – doar 337,825.89 USD, probabil un magazin mic sau cu performanță scăzută.

Cele mai scăzute temperaturi minime:

Magazinul 7 – 10.53°F (~ -11.9°C), deci probabil în nordul SUA.

Cele mai ridicate temperaturi maxime:

Magazinul 10 – 89.67°F (~ 32°C), zonă caldă sau sudică.

Magazinele cu temperaturi scăzute (Magazinul 7) au vânzări maxime semnificativ mai mici (~570.000 USD). Poate reflecta sezonabilitate sau consum redus iarna.

Magazinul 10, cu cel mai mare preț la combustibil, are și cele mai mari vânzări — posibil zonă urbană densă, unde clienții sunt mai puțin sensibili la prețul carburantului.

Magazinele cu temperaturi moderate și preț scăzut la carburant (Magazinele 2, 4) au performanțe bune – pot fi modele pentru expansiune.

Secvența de cod:

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
df = pd.read_csv('SET_DE_DATE_WALMART_PS.csv')
df1 = pd.read_csv('SET_DE_DATE_WALMART_PS_2.csv')
```

```
jonctiune = pd.merge(df,
                     df1[['Store','Fuel_Price']],
                     on='Store')
```

```
print(jonctiune)
```

```
print('-----'*20)
```

```
print(jonctiune.groupby("Store").agg({
    "Fuel_Price": "mean",
    "Weekly_Sales": "max",
    "Temperature": ["min","max"]
})))
```

#9. utilizarea funcțiilor de grup;

#7. accesarea datelor cu loc și iloc;

#14.reprezentare grafică a datelor cu pachetul matplotlib;

#1. utilizarea listelor și a dicționarelor, incluzând metode specifice acestora;

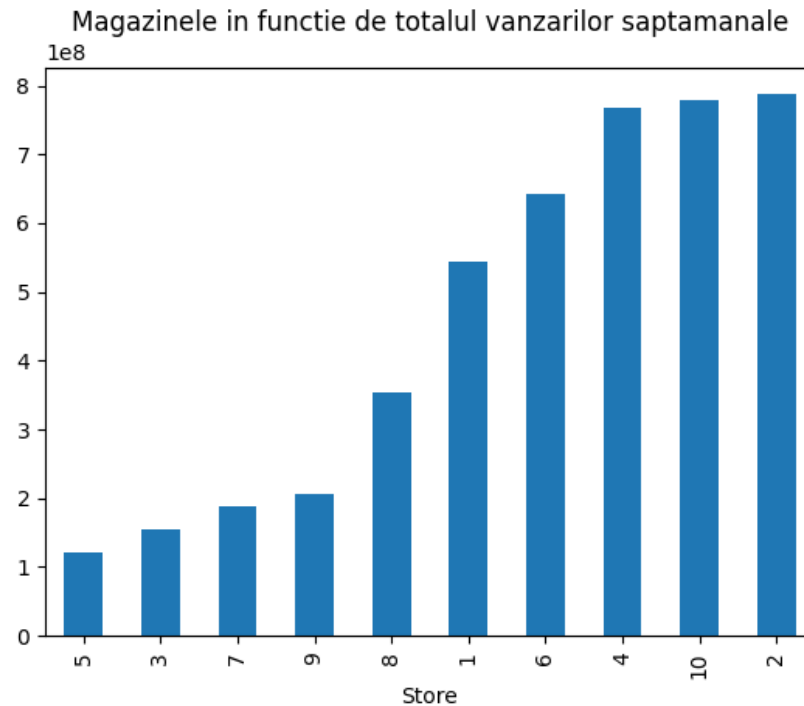
Se dorește vizualizarea primelor 20 de înregistrări pentru magazinul 1 și informațiile despre rata somajului pentru magazinele 2 și 5.

	Store	Unemployment
361	2	8.324
362	2	8.324
363	2	8.324
364	2	8.324
365	2	8.324
...
1956	5	6.465
1957	5	6.465
1958	5	6.465
1959	5	6.465
1960	5	6.465

Se poate observa ca rata somajului are aceeasi valoare pentru fiecare inregistrare pentru fiecare magazin.

	Store	Date	Weekly_Sales	...	CPI	Unemployment	Fuel_Price
0	1	05-02-2010	1643690.90	...	211.096358	8.106	2.572
1	1	05-02-2010	1643690.90	...	211.096358	8.106	2.548
2	1	05-02-2010	1643690.90	...	211.096358	8.106	2.514
3	1	05-02-2010	1643690.90	...	211.096358	8.106	2.561
4	1	05-02-2010	1643690.90	...	211.096358	8.106	2.625
5	1	05-02-2010	1643690.90	...	211.096358	8.106	2.667
6	1	05-02-2010	1643690.90	...	211.096358	8.106	2.720
7	1	05-02-2010	1643690.90	...	211.096358	8.106	2.732
8	1	05-02-2010	1643690.90	...	211.096358	8.106	2.719
9	1	05-02-2010	1643690.90	...	211.096358	8.106	2.770
10	1	05-02-2010	1643690.90	...	211.096358	8.106	2.808
11	1	05-02-2010	1643690.90	...	211.096358	8.106	2.795
12	1	05-02-2010	1643690.90	...	211.096358	8.106	2.780
13	1	05-02-2010	1643690.90	...	211.096358	8.106	2.835
14	1	05-02-2010	1643690.90	...	211.096358	8.106	2.854
15	1	05-02-2010	1643690.90	...	211.096358	8.106	2.826
16	1	05-02-2010	1643690.90	...	211.096358	8.106	2.759
17	1	05-02-2010	1643690.90	...	211.096358	8.106	2.705
18	1	05-02-2010	1643690.90	...	211.096358	8.106	2.668
19	1	12-02-2010	1641957.44	...	211.242170	8.106	2.572

Se dorește să se analizeze grafic totalul vânzărilor săptămânale pentru fiecare magazin. Pentru acest lucru vom avea nevoie de informațiile despre vânzările săptămânale grupate pe magazine, pe care le vom însuma pentru a afla totalul.



Magazinele 2, 10 și 4 se remarcă prin cele mai mari vânzări totale (~800 milioane USD fiecare).

Magazinul 5 are cele mai mici vânzări (sub 150 milioane USD), urmat de Magazinele 3 și 7.

Există o variație semnificativă între magazine, ceea ce sugerează că:

- localizarea și condițiile externe (preț carburant, șomaj, temperatură) pot influența puternic performanța.
- Magazinele din dreapta graficului (cu vânzări mari) indică:
 - potențial de extindere,
 - cerere stabilă și profitabilitate crescută.
- Magazinele din stânga pot necesita:
 - optimizare operațională,
 - analiza cererii locale,
 - reevaluarea poziționării pe piață.

Secventa de cod:

```
print(jonctiune.loc[jonctiune['Store'].isin([5, 2]),['Store','Unemployment']])
```

```
plot=jonctiune.groupby('Store')['Weekly_Sales'].sum()
plot.sort_values().plot(kind='bar')
plt.title("Magazinele in functie de totalul vanzarilor saptamanale")
plt.show()
```

```
print('-----'*20)
print(jonctiune.iloc[:20])
```

#3. definirea și apelarea unor funcții;

Se dorește crearea unei funcții care să calculeze media pentru prețul combustibilului pentru un anumit magazin pe care îl vom specifica (mai exact, magazinul 8).

Pentru acest lucru vom avea nevoie de evidența fiecărui magazin și prețul combustibilului pentru acestea. Astfel, vom crea o funcție care va avea ca parametru de intrare id-ul magazinului pentru care vrem să aflăm prețul mediu al combustibilului.

Asadar, pentru magazinul 8:

```
Media pretului pentru magazinul 8: 2.7
```

Secventa de cod:

```
def pret_mediu_combustibil(store_id):
    store_data = jonctiune[jonctiune["Store"] == store_id]
    return round(store_data["Fuel_Price"].mean(), 2)

print('-----'*20)
print("Media pretului pentru magazinul 8:", pret_mediu_combustibil(8))
```

#10.tratarea valorilor lipsă;

Pentru a evita prezenta valorilor lipsa, lucru ce ar putea influenta rezultatele analizei, vom inlocui valorile lipsa (daca exista), cu media variabilei respective:

```
Nu exista valori lipsa.
```

Secventa de cod:

```
if jonctiune["Fuel_Price"].isnull().any():  
    jonctiune["Fuel_Price"].fillna(jonctiune["Fuel_Price"].mean(), inplace=True)  
    print("Valorile lipsa au fost inlocuite cu media.")  
else:  
    print("Nu exista valori lipsa.")
```

#5. utilizarea structurilor repetitive;

Pentru a usura codul si a maximiza eficienta acestui, se doreste utilizarea unei structuri repetitive pentru a calcula totalul vanzarilor pentru toate magazinele. Pentru acest lucru vom avea nevoie de toate valorile inregistrate de toate magazinele pentru vanzarile saptamanale.

```
Totalul vanzarilor: 4548140394.759995
```

Secventa de cod:

```
print('-----'*20)  
total_vanzari = 0  
for vanzari in jonctiune['Weekly_Sales']:  
    total_vanzari += vanzari  
print(f"Totalul vanzarilor: {total_vanzari}")
```

#4. utilizarea structurilor condiționale;

Se doreste analiza preturile combustibilului pentru toate magazinele cu ajutorul unei structuri conditionale care va clasifica preturile in acest fel:

- Daca pretul este mai mare de 3 USD/gallon, va fi clasificat ca fiind ridicat

- Daca pretul este mai mic de 2.5 USD/gallon, va fi clasificat ca fiind scazut
- Altfel, va fi clasificat ca fiind standard

Asadar, vom analiza preturile combustibilului pentru magazinul cu numarul 10:

```
Pretul combustibilului este ridicat.
Pretul combustibilului este standard.
Pretul combustibilului este ridicat.
Pretul combustibilului este standard.
Pretul combustibilului este standard.
Pretul combustibilului este standard.
Pretul combustibilului este standard.
Pretul combustibilului este standard.
Pretul combustibilului este standard.
Pretul combustibilului este ridicat.
Pretul combustibilului este ridicat.
Pretul combustibilului este standard.
Pretul combustibilului este ridicat.
Pretul combustibilului este ridicat.
Pretul combustibilului este ridicat.
Pretul combustibilului este ridicat.
Pretul combustibilului este ridicat.
Pretul combustibilului este ridicat.
Pretul combustibilului este ridicat.
Pretul combustibilului este ridicat.
Pretul combustibilului este standard.
```

Se poate observa ca exista fluctuatii in pretul combustibilului, fiind majoritar ridicate, posibil fiind zonă urbană densă, unde clienții sunt mai puțin sensibili la prețul carburantului.

#8. modificarea datelor în pachetul pandas;

#11.ștergerea de coloane și înregistrări;

Se dorește adăugarea unei noi coloane care să conțină vânzarile săptămânale exprimate în mii de dolari și să se ajusteze vânzarile magazinului 5 cu 10%.”

De asemenea, să se ștergă înregistrările pentru vânzarile săptămânale care nu depășesc suma de 1000 USD și coloana Holiday_Flags, care nu are o așa mare influență în analiză.

	Store	Date	Weekly_Sales	...	Unemployment	Fuel_Price	Weekly_Sales_2
1561	5	05-02-2010	317173.1	...	6.566	2.572	317.1731
1562	5	05-02-2010	317173.1	...	6.566	2.548	317.1731
1563	5	05-02-2010	317173.1	...	6.566	2.514	317.1731
1564	5	05-02-2010	317173.1	...	6.566	2.561	317.1731
1565	5	05-02-2010	317173.1	...	6.566	2.625	317.1731
...
1956	5	18-06-2010	313795.6	...	6.465	2.826	313.7956
1957	5	18-06-2010	313795.6	...	6.465	2.759	313.7956
1958	5	18-06-2010	313795.6	...	6.465	2.705	313.7956
1959	5	18-06-2010	313795.6	...	6.465	2.668	313.7956
1960	5	18-06-2010	313795.6	...	6.465	2.637	313.7956

	Store	Date	Weekly_Sales	...	Unemployment	Fuel_Price	Weekly_Sales_2
1561	5	05-02-2010	348890.41	...	6.566	2.572	348.89041
1562	5	05-02-2010	348890.41	...	6.566	2.548	348.89041
1563	5	05-02-2010	348890.41	...	6.566	2.514	348.89041
1564	5	05-02-2010	348890.41	...	6.566	2.561	348.89041
1565	5	05-02-2010	348890.41	...	6.566	2.625	348.89041
...
1956	5	18-06-2010	345175.16	...	6.465	2.826	345.17516
1957	5	18-06-2010	345175.16	...	6.465	2.759	345.17516
1958	5	18-06-2010	345175.16	...	6.465	2.705	345.17516
1959	5	18-06-2010	345175.16	...	6.465	2.668	345.17516
1960	5	18-06-2010	345175.16	...	6.465	2.637	345.17516

	Store	Date	Weekly_Sales	...	Unemployment	Fuel_Price	Weekly_Sales_2
0	1	05-02-2010	1643690.90	...	8.106	2.572	1643.69090
1	1	05-02-2010	1643690.90	...	8.106	2.548	1643.69090
2	1	05-02-2010	1643690.90	...	8.106	2.514	1643.69090
3	1	05-02-2010	1643690.90	...	8.106	2.561	1643.69090
4	1	05-02-2010	1643690.90	...	8.106	2.625	1643.69090
...
3956	10	18-06-2010	1837636.24	...	9.524	3.120	1837.63624
3957	10	18-06-2010	1837636.24	...	9.524	3.058	1837.63624
3958	10	18-06-2010	1837636.24	...	9.524	2.941	1837.63624
3959	10	18-06-2010	1837636.24	...	9.524	3.057	1837.63624
3960	10	18-06-2010	1837636.24	...	9.524	2.935	1837.63624

Secventa de cod:

```
print('-----'*20)
jonctiune['Weekly_Sales_2'] = jonctiune['Weekly_Sales'] / 1000
print(jonctiune.loc[(jonctiune['Store']==5)])
jonctiune.loc[(jonctiune['Store']==5), 'Weekly_Sales'] = jonctiune.loc[(jonctiune['Store']==5), 'Weekly_Sales'] * 1.1
jonctiune.loc[(jonctiune['Store']==5), 'Weekly_Sales_2'] = jonctiune.loc[(jonctiune['Store']==5), 'Weekly_Sales_2'] * 1.1
print(jonctiune.loc[(jonctiune['Store']==5)])
print('-----'*20)
jonctiune.drop(columns=['Holiday_Flag'], inplace=True)
jonctiune = jonctiune[jonctiune['Weekly_Sales'] >= 1000]
print(jonctiune)
```

#2.utilizarea seturilor și a tuplurilor, incluzând metode specifice acestora;

Se creează un set pentru temperaturi în care să se specifice și data și temperatura respectivă și un tuplu în care să se specifice data și vânzarea pe care a făcut-o magazinul respectiv în acea săptămână pentru prima înregistrare (respectiv primul magazin).

```
Temperaturile în zilele respective: {('07-05-2010', 73.47), ('26-02-2010', 47.1), ('23-04-2010', 64.11), ('23-04-2010', 61.26)}
-----
Magazinul a înregistrat vânzări săptămânale până în ziua respectivă în valoare de: (1, '05-02-2010', 1643690.9)
```

Secventa de cod:

```
set_temperaturi = set(zip(jonctiune['Date'], jonctiune['Temperature']))
print("Temperaturile în zilele respective:", set_temperaturi)

print('-----'*20)
tuplu = (int(jonctiune.loc[1, 'Store']), jonctiune.loc[1, 'Date'], float(jonctiune.loc[1, 'Weekly_Sales']))
print("Magazinul a înregistrat vânzări săptămânale până în ziua respectivă în valoare de:", tuplu)
```

#15.utilizarea pachetului scikit-learn (clusterizare, regresie logistică)

#16.utilizarea pachetului statmodels (regresie multiplă).

Se dorește clusterizarea magazinelor Walmart pe baza vânzărilor săptămânale și ratei șomajului pentru a identifica grupuri naturale cu caracteristici similare și modelarea impactului șomajului și temperaturii asupra vânzărilor săptămânale pentru previziuni și analiză a factorilor critici.

Pentru clusterizare vom avea:

- Variabile de intrare: Weekly_Sales, Unemployment
- Număr de clustere: 3 (ales pe baza logicii de segmentare a pieței)
- Date curățate (eliminarea valorilor lipsă)

iar pentru regresie vom avea:

- Variabilă dependentă: Weekly_Sales
- Variabile independente: Unemployment, Temperature
- Date curățate (eliminarea valorilor lipsă)

	Weekly_Sales	Unemployment	Cluster
0	1643690.90	8.106	0.0
1	1643690.90	8.106	0.0
2	1643690.90	8.106	0.0
3	1643690.90	8.106	0.0
4	1643690.90	8.106	0.0
...
3956	1837636.24	9.524	2.0
3957	1837636.24	9.524	2.0
3958	1837636.24	9.524	2.0
3959	1837636.24	9.524	2.0
3960	1837636.24	9.524	2.0

Concluzii și Recomandări de Acțiune

1. Pentru Cluster 0:
 - Oferă promoții pentru a menține cererea în condiții economice instabile.
 - Monitorizează stocurile pentru a evita surplusul.
2. Pentru Cluster 2:
 - Introdu produse premium pentru clienții cu venituri mari.
 - Colaborează cu autoritățile locale pentru programe de reducere a șomajului.

OLS Regression Results						
=====						
Dep. Variable:	Weekly_Sales	R-squared:	0.322			
Model:	OLS	Adj. R-squared:	0.315			
Method:	Least Squares	F-statistic:	46.56			
Date:	Sun, 01 Jun 2025	Prob (F-statistic):	2.86e-17			
Time:	15:36:17	Log-Likelihood:	-2914.7			
No. Observations:	199	AIC:	5835.			
Df Residuals:	196	BIC:	5845.			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-2.108e+06	3.5e+05	-6.019	0.000	-2.8e+06	-1.42e+06
Unemployment	3.532e+05	3.74e+04	9.454	0.000	2.79e+05	4.27e+05
Temperature	9285.6103	2576.365	3.604	0.000	4204.655	1.44e+04
=====						
Omnibus:	43.988	Durbin-Watson:	0.220			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	12.578			
Skew:	-0.332	Prob(JB):	0.00186			
Kurtosis:	1.962	Cond. No.	548.			
=====						

R-squared = 0.322, deci modelul explică aproximativ 32.2% din variația vânzărilor săptămânale.

Adj. R-squared = 0.315: valoare ajustată, ține cont de numărul de predictorii. Este destul de apropiată de R^2 , ceea ce sugerează că predictorii incluși adaugă valoare modelului.

F-statistic = 46.56 cu $p < 0.001$: Modelul este global semnificativ; există o relație semnificativă între cel puțin unul dintre predictorii independenți și vânzările săptămânale

Jarque-Bera (JB) = 12.578, $p = 0.0018$: Reziduurile nu sunt perfect distribuite normal.

Skew = -0.332, **Kurtosis = 1.962**: Ușoară asimetrie la stanga și o curbă puțin mai plată decât distribuția normală.

Durbin-Watson = 0.22: Posibilă autocorelație pozitivă între erori, ceea ce poate indica o problemă în cazul datelor temporale.

Creșterea șomajului duce la creșterea vânzărilor – un rezultat contraintuitiv. Posibile explicații:

- Oamenii cumpără mai mult din magazine de discount sau supermarketuri în loc de restaurante sau alte servicii.
- Poate reflecta stimulente guvernamentale sau ajutoare sociale în perioada de șomaj.

Temperatura are un impact pozitiv asupra vânzărilor – poate indica sezonabilitate:

- Vânzările cresc în perioadele calde (ex: produse de vară, turism, băuturi reci etc.).

Modelul explică o parte moderată din variația vânzărilor (32%), deci alte variabile (precum promoții, sărbători, concurență, etc.) ar trebui explorate pentru îmbunătățirea modelului.

Secvența de cod:

```
from sklearn.cluster import KMeans
print('-----'*20)
clusterizare = jonctiune[['Weekly_Sales', 'Unemployment']].dropna()

kmeans = KMeans(n_clusters=3, random_state=0)
jonctiune.loc[clusterizare.index, 'Cluster'] = kmeans.fit_predict(clusterizare)
print(jonctiune[['Weekly_Sales', 'Unemployment', 'Cluster']])

import statsmodels.api as sm
print('-----'*20)

X = df[['Unemployment', 'Temperature']].dropna()
y = df['Weekly_Sales'].loc[X.index]

X = sm.add_constant(X)
model = sm.OLS(y, X).fit()

print(model.summary())
```

CAPITOLUL 2

PROGRAMARE SAS

/*1. crearea unui set de date SAS din fișiere externe;*/

Se dorește importul datelor brute (SET_DE_DATE_WALMART_PS.txt' 'SET_DE_DATE_WALMART_PS_2.txt') într-un format accesibil pentru analiză.

Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	CPI	Unemployment
1	1	05/02/2010	1643690.90	0	42.31	211.096	8.106
2	1	12/02/2010	1641957.44	1	38.51	211.242	8.106
3	1	19/02/2010	1611968.17	0	39.93	211.289	8.106
4	1	26/02/2010	1409727.59	0	46.63	211.320	8.106
5	1	05/03/2010	1554806.68	0	46.50	211.350	8.106
6	1	12/03/2010	1439541.59	0	57.79	211.381	8.106
7	1	19/03/2010	1472515.79	0	54.58	211.216	8.106
8	1	26/03/2010	1404429.92	0	51.45	211.018	8.106
9	1	02/04/2010	1594968.28	0	62.27	210.820	7.808
10	1	09/04/2010	1545418.53	0	65.86	210.623	7.808
11	1	16/04/2010	1466058.28	0	66.32	210.489	7.808
12	1	23/04/2010	1391256.12	0	64.84	210.439	7.808
13	1	30/04/2010	1425100.71	0	67.41	210.390	7.808
14	1	07/05/2010	1603955.12	0	72.55	210.340	7.808
15	1	14/05/2010	1494251.50	0	74.78	210.337	7.808
16	1	21/05/2010	1399662.07	0	76.44	210.617	7.808
17	1	28/05/2010	1432069.95	0	80.44	210.897	7.808
18	1	04/06/2010	1615524.71	0	80.69	211.176	7.808
19	1	11/06/2010	1542561.09	0	80.43	211.456	7.808
20	2	05/02/2010	2136989.46	0	40.19	210.753	8.324
21	2	12/02/2010	2137809.50	1	38.49	210.898	8.324
22	2	19/02/2010	2124451.54	0	39.69	210.945	8.324
23	2	26/02/2010	1865097.27	0	46.10	210.976	8.324
24	2	05/03/2010	1991013.13	0	47.17	211.007	8.324
25	2	12/03/2010	1990483.78	0	57.56	211.038	8.324

Obs	Store	Fuel_Price
1	1	2.572
2	1	2.548
3	1	2.514
4	1	2.561
5	1	2.625
6	1	2.667
7	1	2.720
8	1	2.732
9	1	2.719
10	1	2.770
11	1	2.808
12	1	2.795
13	1	2.780
14	1	2.835
15	1	2.854
16	1	2.826
17	1	2.759
18	1	2.705
19	1	2.668
20	2	2.572
21	2	2.548
22	2	2.514
23	2	2.561
24	2	2.625
25	2	2.667

Secventa de cod:

```
data date1;  
    infile '/home/u64205077/PROIECT/SET_DE_DATE_WALMART_PS.txt' dlm='09'x dsd  
    firstobs=2;  
    input Store Date : ddmmyy10. Weekly_Sales Holiday_Flag Temperature CPI Unemployment;  
    format Date ddmmyy10.;  
run;
```

```
proc print data=date1;  
run;
```

```
data date2;  
    infile '/home/u64205077/PROIECT/SET_DE_DATE_WALMART_PS_2.txt' dlm='09'x dsd  
    firstobs=2;  
    input Store Fuel_Price;  
run;
```

/*2. crearea și folosirea de formate definite de utilizator;

4. crearea de subseturi de date;

6. combinarea seturilor de date prin proceduri specifice SAS și SQL;*/

Se dorește crearea unui format personalizat în care zilele de sărbătoare vor fi transformate din valori (0,1) în “E sărbătoare” sau “Nu e sărbătoare”. Acest lucru se va face prin folosirea unui subset a setului inițial pe care îl vom îmbina mai apoi cu celălalt set de date care conține date legate de prețul combustibilului.

Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	CPI	Unemployment	Fuel_Price
1	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.572
2	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.668
3	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.705
4	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.759
5	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.826
6	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.854
7	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.835
8	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.780
9	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.795
10	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.808
11	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.770
12	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.719
13	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.732
14	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.720
15	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.667
16	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.625
17	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.561
18	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.514
19	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.548
20	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.572
21	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.668
22	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.705
23	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.759
24	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.826
25	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.854
26	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.835

Secventa de cod:

```
proc print data=date2;
```

```
run;
```

```
proc format;
```

```
value formatare
```

```
1 = 'E sarbatoare'
```

```
0 = 'Nu e sarbatoare';
```

```
run;
```

```
data date_format;
```

```
set date1;
```

```
format Holiday_Flag formatare.;
```

```
run;
```

```
proc print data=date_format;
```

```
run;
```

```
proc sql;
```

```
create table jonctiune as
```

```
select a.*, b.Fuel_Price
```

```
from date_format a
```

```
inner join date2 b
```

```
on a.Store = b.Store;
```

```
quit;
```

```
proc print data=jonctiune;
```

```
run;
```

/*3. Procesare iterativă și condițională */

Se dorește crearea unei noi variabile numită Tip_vanzare, cu ajutorul căreia se vor clasifica prin procesare iterativă și condițională vânzarile săptămânale, astfel încât vor fi încadrate în: vânzări ridicate dacă depășesc 2000000 USD, vânzări medii dacă se încadrează între 1000000 și 2000000 USD, și vânzări mici dacă nu depășesc pragul de 1000000 USD.

Astfel, vom avea nevoie de vânzarile săptămânale ale tuturor magazinelor din setul de date imbinat, pe care le vom procesa printr-o structură condițională de tip if:

Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	CPI	Unemployment	Fuel_Price	Tip_vanzare
1	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.572	Mediu
2	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.668	Mediu
3	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.705	Mediu
4	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.759	Mediu
5	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.826	Mediu
6	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.854	Mediu
7	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.835	Mediu
8	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.780	Mediu
9	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.795	Mediu
10	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.808	Mediu
11	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.770	Mediu
12	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.719	Mediu
13	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.732	Mediu
14	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.720	Mediu
15	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.667	Mediu
16	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.625	Mediu
17	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.561	Mediu
18	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.514	Mediu
19	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.548	Mediu
20	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.572	Mediu
21	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.668	Mediu
22	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.705	Mediu
23	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.759	Mediu
24	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.826	Mediu

Secventa de cod:

```
data vanzari;
```

```
set jonctiune;
```

```
length Tip_vanzare $10;
```

```
if Weekly_Sales > 2000000 then Tip_vanzare = "Ridicat";
```

```
else if Weekly_Sales > 1000000 then Tip_vanzare = "Mediu";
```

```
else Tip_vanzare = "Mic";
```

```
run;
```

```
proc print data=vanzari;
```

```
run;
```

/*7. Utilizarea de masive*/

Se doreste convertirea temperaturii din grade Fahrenheit in Celsius pentru a usura interpretarea rezultatelor, cu ajutorul unui masiv.

Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	CPI	Unemployment	Fuel_Price	Temperatura_Celsius
1	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.572	5.7278
2	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.668	5.7278
3	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.705	5.7278
4	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.759	5.7278
5	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.826	5.7278
6	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.854	5.7278
7	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.835	5.7278
8	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.780	5.7278
9	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.795	5.7278
10	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.808	5.7278
11	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.770	5.7278
12	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.719	5.7278
13	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.732	5.7278
14	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.720	5.7278
15	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.667	5.7278
16	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.625	5.7278
17	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.561	5.7278
18	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.514	5.7278
19	1	05/02/2010	1643690.90	Nu e sarbatoare	42.31	211.096	8.106	2.548	5.7278
20	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.572	3.6167
21	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.668	3.6167
22	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.705	3.6167
23	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.759	3.6167
24	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.826	3.6167
25	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.854	3.6167
26	1	12/02/2010	1641957.44	E sarbatoare	38.51	211.242	8.106	2.835	3.6167

Secventa de cod:

```
data masiv;  
    set jonctiune;  
    array temp[1] Temperature;  
    array tempC[1] Temperatura_Celsius;  
    do i = 1 to 1;  
        tempC[i] = (temp[i] - 32) * 5 / 9;  
    end;  
    drop i;  
run;  
  
proc print data=masiv;  
run;
```

/*5. utilizarea de funcții SAS;

8. utilizarea de proceduri pentru raportare;

9. folosirea de proceduri statistice

10.generarea de grafice.*/

Se cere un raport statistic pentru vanzarile saptamanale din luna mai si pretul combustibilului. De asemenea, sa se genereze histograma pentru magazinul cu numarul 5 din luna mai si graficul evolutiei vanzarilor saptamanale.

The UNIVARIATE Procedure
Variable: Weekly_Sales

Moments			
N	796	Sum Weights	796
Mean	1124013.46	Sum Observations	894714712
Std Deviation	663893.726	Variance	4.40755E11
Skewness	0.01264813	Kurtosis	-1.7460964
Uncorrected SS	1.35607E15	Corrected SS	3.504E14
Coeff Variation	59.064571	Std Error Mean	23531.0892

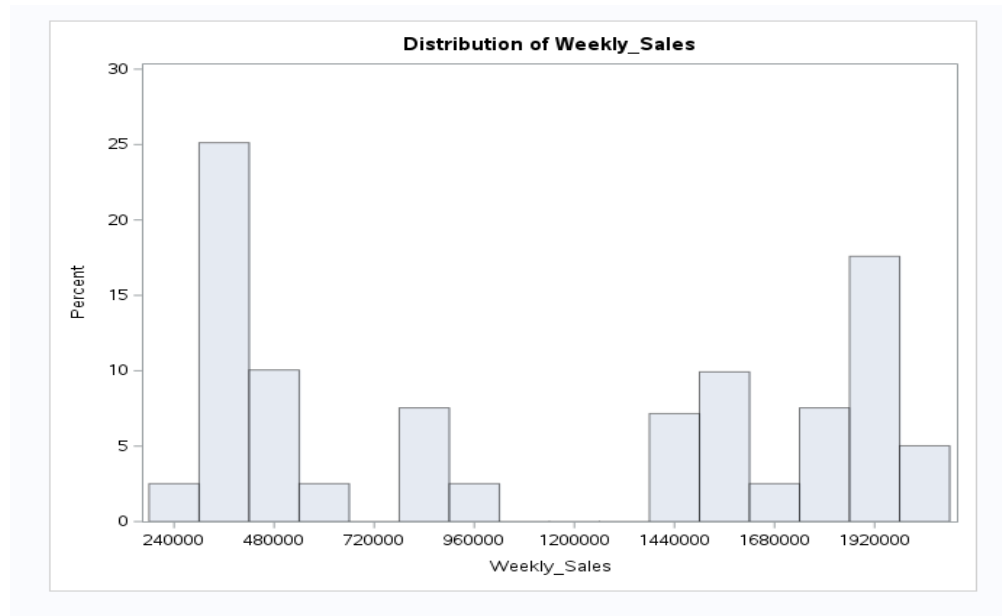
Basic Statistical Measures			
Location		Variability	
Mean	1124013	Std Deviation	663894
Median	916821	Variance	4.40755E11
Mode	296674	Range	1745908
		Interquartile Range	1462080

Note: The mode displayed is the smallest of 36 modes with a count of 20.

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	47.76717	Pr > t 	<.0001
Sign	M	398	Pr >= M 	<.0001
Signed Rank	S	158603	Pr >= S 	<.0001

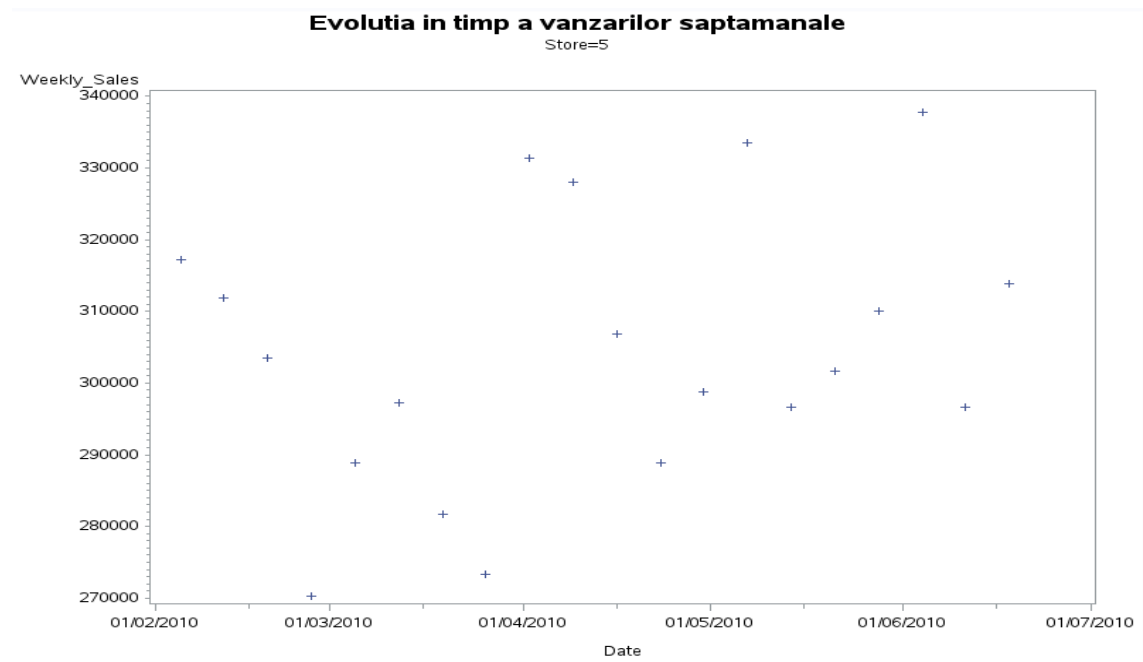
Quantiles (Definition 5)	
Level	Quantile
100% Max	2042582
99%	2042582
95%	2000626
90%	1921432
75% Q3	1857534
50% Median	916821
25% Q1	395454
10%	333523
5%	301615
1%	296674
0% Min	296674

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
296674	356	2042582	92
296674	355	2042582	93
296674	354	2042582	94
296674	353	2042582	95
296674	352	2042582	96



Store=5

Analysis Variable : Fuel_Price				
N	Mean	Std Dev	Minimum	Maximum
400	2.7047500	0.1000646	2.5140000	2.8540000



Vânzări Săptămânale:

- Medie și Variabilitate: Vânzările medii săptămânale sunt de aproximativ 1,124,013 USD, cu o abatere standard mare (663,894 USD), indicând fluctuații semnificative în vânzări de la o săptămână la alta.
- Distribuție Asimetrică: Mediana (916,821 USD) este mai mică decât media, sugerând o distribuție cu coadă lungă spre dreapta (valori extreme ridicate). Kurtosis negativ (-1.75) confirmă o distribuție mai plată decât cea normală, cu mai puține valori extreme.
- Cuartile:
 - 25% din săptămâni au vânzări sub 395,454 USD, iar 75% sub 1,857,534 USD.
 - Vânzările maxime (2,042,582 USD) sunt de aproape 7 ori mai mari decât minimele (296,674 USD), evidențiind sezonabilitate sau promoții excepționale.
- Teste Statistice: Toate testele (t-test, sign test, signed rank) confirmă că vânzările sunt semnificativ diferite de zero ($p < 0.0001$), ceea ce este normal pentru un magazin funcțional.

Prețul Combustibilului pentru Magazinul 5

- Stabilitate Relativă: Prețul mediu al combustibilului este 2.70 USD, cu o abatere standard mică (0.10 USD) și un interval restrâns (2.51–2.85 USD). Aceasta sugerează o politică de prețuri constante sau o perioadă fără fluctuații majore pe piața energiei.

Concluzii

- Vânzări: Variabilitatea mare în vânzări indică oportunități de optimizare (ex.: promoții pentru săptămânile cu vânzări slabe, gestionarea stocurilor în perioadele de vârf).
- Combustibil: Prețul stabil al combustibilului poate fi un factor de predictibilitate pentru costurile logistice ale magazinului.
- Acțiuni Posibile: Analiza cauzelor vânzărilor extreme (ex.: săptămâna cu 2,042,582 USD) și investigarea modei frecvente (296,674 USD, apărută de 20 de ori) pentru a identifica pattern-uri (ex.: perioade de leneș economic).

Concluzie Finală: Walmart ar trebui să se concentreze pe reducerea variabilității vânzărilor prin strategii de marketing adaptate și să monitorizeze factori externi (ex.: prețul combustibilului) care ar putea afecta indirect performanța. Investițiile în analiza datelor în timp real ar putea îmbunătăți capacitatea de anticipare a tendințelor.

Secventa de cod:

```
data luna_mai;
```

```
    set masiv;
```

```
    where month(Date) = 5;
```

```
run;
```

```
proc UNIVARIATE data=luna_mai;
```

```
    var Weekly_Sales;
```

```
    histogram Weekly_Sales;
```

```
run;
```

```
PROC MEANS DATA = jonctiune;
```

```
    BY Store;
```

```
    VAR Fuel_Price;
```

```
    TITLE 'Raportul pretului la combustibil per magazin';
```

```
RUN;
```

```
TITLE "Evolutia in timp a vanzarilor saptamanale";
```

```
PROC GPLOT data=jonctiune;
```

```
    by Store;
```

```
    PLOT Weekly_Sales * Date;
```

```
RUN;
```

```
QUIT;
```

CAPITOLUL 3

PROGRAMARE SAS ENTERPRISE GUIDE

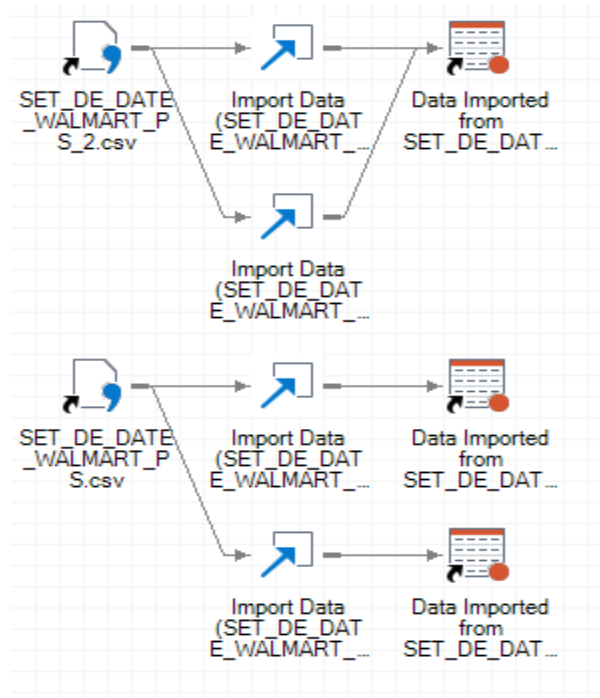
Pentru acest capitol, ne vom ajuta de comenzile SQL utilizabile în SAS, cât și de comenzi SAS:

1. Importul unui fisier non-SAS;

Se dorește importul datelor brute (SET_DE_DATE_WALMART_PS.csv' 'SET_DE_DATE_WALMART_PS_2.csv') într-un format accesibil pentru analiză.

Acest lucru se va face prin accesarea File->Import data-> “SET_DE_DATE_WALMART_PS.csv”->Open

File->Import data-> “SET_DE_DATE_WALMART_PS_2.csv”->Open



2. Jonctiune;

Se doreste crearea unui nou set de date, prin jonctiune, care sa contina atat informatii despre performanta magazinului, cat si informatiile despre pretul combustibilului.

	⊕ Store	📅 Date	⊕ Weekly_Sales	⊕ Holiday_Flag	⊕ Temperature	⊕ CPI	⊕ Unemployment	⊕ Fuel_Price
1	1	05/02/2010	1643690.9	0	42.31	211.0963582	8.106	2.572
2	1	11/06/2010	1542561.09	0	80.43	211.4560951	7.808	2.572
3	1	09/04/2010	1545418.53	0	65.86	210.6228574	7.808	2.572
4	1	19/03/2010	1472515.79	0	54.58	211.215635	8.106	2.572
5	1	28/05/2010	1432069.95	0	80.44	210.8967606	7.808	2.572
6	1	07/05/2010	1603955.12	0	72.55	210.3399684	7.808	2.572
7	1	30/04/2010	1425100.71	0	67.41	210.3895456	7.808	2.572
8	1	26/03/2010	1404429.92	0	51.45	211.0180424	8.106	2.572
9	1	14/05/2010	1494251.5	0	74.78	210.3374261	7.808	2.572
10	1	05/03/2010	1554806.68	0	46.5	211.3501429	8.106	2.572
11	1	02/04/2010	1594968.28	0	62.27	210.8204499	7.808	2.572
12	1	19/02/2010	1611968.17	0	39.93	211.2891429	8.106	2.572
13	1	26/02/2010	1409727.59	0	46.63	211.3196429	8.106	2.572
14	1	16/04/2010	1466058.28	0	66.32	210.4887	7.808	2.572
15	1	12/02/2010	1641957.44	1	38.51	211.2421698	8.106	2.572
16	1	21/05/2010	1399662.07	0	76.44	210.6170934	7.808	2.572
17	1	04/06/2010	1615524.71	0	80.69	211.1764278	7.808	2.572
18	1	23/04/2010	1391256.12	0	64.84	210.4391228	7.808	2.572
19	1	12/03/2010	1439541.59	0	57.79	211.3806429	8.106	2.572
20	1	05/02/2010	1643690.9	0	42.31	211.0963582	8.106	2.668
21	1	11/06/2010	1542561.09	0	80.43	211.4560951	7.808	2.668
22	1	09/04/2010	1545418.53	0	65.86	210.6228574	7.808	2.668
23	1	19/03/2010	1472515.79	0	54.58	211.215635	8.106	2.668
24	1	28/05/2010	1432069.95	0	80.44	210.8967606	7.808	2.668
25	1	07/05/2010	1603955.12	0	72.55	210.3399684	7.808	2.668
26	1	30/04/2010	1425100.71	0	67.41	210.3895456	7.808	2.668
27	1	26/03/2010	1404429.92	0	51.45	211.0180424	8.106	2.668
28	1	14/05/2010	1494251.5	0	74.78	210.3374261	7.808	2.668
29	1	05/03/2010	1554806.68	0	46.5	211.3501429	8.106	2.668
30	1	02/04/2010	1594968.28	0	62.27	210.8204499	7.808	2.668

Secventa de cod:

```
proc sql;  
create table WORK.jonctiune as
```

```

select ps1.*, ps2.Fuel_Price
from WORK.SET_DE_DATE_WALMART_PS as ps1
left join WORK.SET_DE_DATE_WALMART_PS_2 as ps2
on ps1.Store = ps2.Store;
quit;

```

3. Interogări;

Sa se afle zilele in care pretul combustibilului a depasit 2.719 USD/gallon.

	⊕ Store	📅 Date	⊕ Fuel_Price
1	1	05/02/2010	2.77
2	1	11/06/2010	2.77
3	1	09/04/2010	2.77
4	1	19/03/2010	2.77
5	1	28/05/2010	2.77
6	1	07/05/2010	2.77
7	1	30/04/2010	2.77
8	1	26/03/2010	2.77
9	1	14/05/2010	2.77
10	1	05/03/2010	2.77
11	1	02/04/2010	2.77
12	1	19/02/2010	2.77
13	1	26/02/2010	2.77
14	1	16/04/2010	2.77
15	1	12/02/2010	2.77
16	1	21/05/2010	2.77
17	1	04/06/2010	2.77
18	1	23/04/2010	2.77
19	1	12/03/2010	2.77
20	1	05/02/2010	2.72
21	1	11/06/2010	2.72
22	1	09/04/2010	2.72
23	1	19/03/2010	2.72
24	1	28/05/2010	2.72
25	1	07/05/2010	2.72
26	1	30/04/2010	2.72
27	1	26/03/2010	2.72
28	1	14/05/2010	2.72
29	1	05/03/2010	2.72
30	1	02/04/2010	2.72

Secventa de cod:

```
proc sql;  
    create table WORK.interogare as  
    select Store, Date, Fuel_Price  
    from WORK.jonctiune  
    where Fuel_Price > 2.719;  
quit;
```

4. Folosirea parametrilor;

Folosind un parametru pentru numarul magazinului, sa se aduca informatii despre vanzarile saptamanale pentru magazinul 1.

Vânzări săptămânale pentru magazinul 1

Date	Weekly_Sales
05/02/2010	1643690.9
11/06/2010	1542561.09
09/04/2010	1545418.53
19/03/2010	1472515.79
28/05/2010	1432069.95
07/05/2010	1603955.12
30/04/2010	1425100.71
26/03/2010	1404429.92
14/05/2010	1494251.5
05/03/2010	1554806.68
02/04/2010	1594968.28
19/02/2010	1611968.17
26/02/2010	1409727.59
16/04/2010	1466058.28
12/02/2010	1641957.44
21/05/2010	1399862.07
04/06/2010	1615524.71
23/04/2010	1391256.12
12/03/2010	1439541.59
05/02/2010	1643690.9
11/06/2010	1542561.09
09/04/2010	1545418.53
19/03/2010	1472515.79
28/05/2010	1432069.95
07/05/2010	1603955.12
30/04/2010	1425100.71
26/03/2010	1404429.92
14/05/2010	1494251.5
05/03/2010	1554806.68
02/04/2010	1594968.28
19/02/2010	1611968.17
26/02/2010	1409727.59
16/04/2010	1466058.28
12/02/2010	1641957.44
21/05/2010	1399862.07
04/06/2010	1615524.71
23/04/2010	1391256.12

Secventa de cod:

```
%let store_id = 1;
```

```
title "Vânzări săptămânale pentru magazinul 1";
```

```
proc sql;
```

```
select Date, Weekly_Sales
```

```
from WORK.jonctiune
```

```
where Store = &store_id;
```

```
quit;
```

```
title;
```

5. Rapoarte;

Sa se genereze un raport despre magazinele care au avut vanzari mai mari de 2 milioane USD.

Pentru acest lucru, vom avea nevoie de numarul magazinului, data si valoarea vanzarilor saptamanale.

Raport - Vânzări săptămânale peste 2 milioane

Magazin	Data	Vânzări săptămânale
2	02/04/2010	\$2,086,188
	19/02/2010	\$2,124,452
	12/02/2010	\$2,137,810
	05/02/2010	\$2,136,989
	11/06/2010	\$2,025,539
	04/06/2010	\$2,102,540
	18/06/2010	\$2,001,637
	07/05/2010	\$2,042,582
	02/04/2010	\$2,086,188
	19/02/2010	\$2,124,452
	12/02/2010	\$2,137,810
	05/02/2010	\$2,136,989
	11/06/2010	\$2,025,539
	04/06/2010	\$2,102,540
	18/06/2010	\$2,001,637
	07/05/2010	\$2,042,582
	02/04/2010	\$2,086,188
	19/02/2010	\$2,124,452
	12/02/2010	\$2,137,810
	05/02/2010	\$2,136,989
2	11/06/2010	\$2,025,539
	04/06/2010	\$2,102,540
	18/06/2010	\$2,001,637
	07/05/2010	\$2,042,582
	02/04/2010	\$2,086,188
	19/02/2010	\$2,124,452
	12/02/2010	\$2,137,810
	05/02/2010	\$2,136,989
	11/06/2010	\$2,025,539
	04/06/2010	\$2,102,540

Secventa de cod:

```
title "Raport - Vânzări săptămânale peste 2 milioane";
```

```
proc report data=WORK.jonctiune nowd;
```

```
columns Store Date Weekly_Sales;
```

```
where Weekly_Sales > 2000000;
```

```
define Store / group 'Magazin';
```

```
define Date / display 'Data';
```

```
define Weekly_Sales / sum format=dollar12. 'Vânzări săptămânale';
```

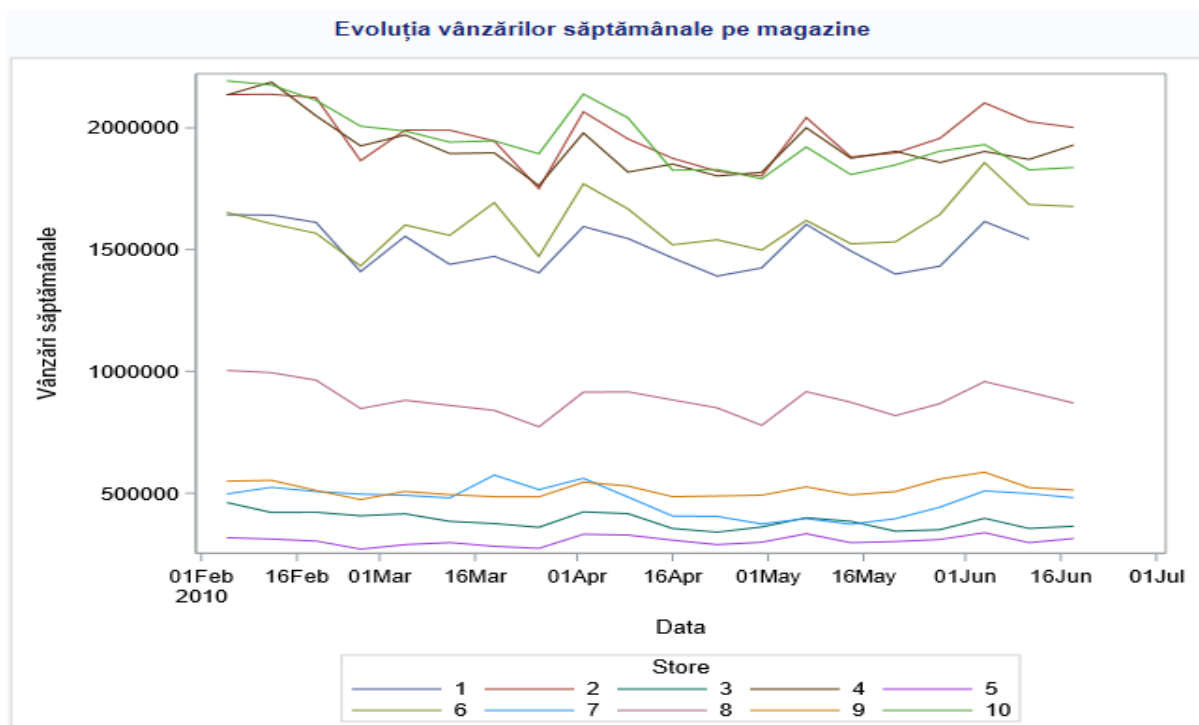
```
run;
```

```
title;
```

6. Grafice;

Se cere analiza evolutiei vanzarilor saptamanale pentru toate magazinele cu ajutorul unui grafic.

Pentru acest lucru, vom avea nevoie de informatii despre numarul magazinului, data si valoarea vanzarilor saptamanale pentru fiecare in parte.



1. Tendințe Generale în Vânzări

- Vânzările fluctuează semnificativ pe parcursul perioadei analizate (februarie – iulie 2010), cu valori cuprinse între 500,000 USD și 2,000,000 USD.
- Maximele (aproape 2 milioane USD) și minimele (sub 500,000 USD) sugerează o sezonabilitate pronunțată sau evenimente punctuale (ex.: sărbători, promoții).

2. Observații Cheie

- Perioade de Vârf:
 - Creșteri bruște ale vânzărilor (ex.: pe 16 februarie, 1 aprilie, 1 iunie) pot fi legate de campanii promoționale (ex.: Valentine's Day, Paste) sau lansări de produse.
- Perioade de Scădere:
 - Declinuri în săptămânile din martie și mai ar putea indica sezonabilitate negativă (ex.: perioade post-sărbători) sau factori externi (ex.: vreme nefavorabilă).

3. Comparație între Magazine (Magazinul 1 vs. Magazinele 2, 6, 7)

- Magazinele 1 și 2:
 - Store 1 pare să aibă vânzări mai volatile (creșteri și scăderi mai abrupte) comparativ cu Store 2, care ar putea avea o strategie de prețuri sau o localizare mai stabilă.
- Magazinele 6 și 7:
 - Store 6 atinge vânzări mai mari decât Store 7 în unele săptămâni, ceea ce ar putea indica diferențe în mărimea magazinului, oferta de produse sau traficul de clienți.

4. Implicații Economice

- Optimizarea Stocurilor:
 - Magazinele cu vânzări volatile (ex.: Magazinul 1) ar beneficia de inventar dinamic pentru a evita surplusul sau lipsurile.
- Strategii Personalizate:
 - Store 6 și 7 ar putea testa tactici de marketing diferite (ex.: promoții locale) pentru a uniformiza performanța.
- Analiza Sezonității:
 - Identificarea pattern-urilor recurente (ex.: creșteri în pre-sărbători) poate îmbunătăți planificarea bugetară.

Walmart ar trebui să capitalizeze pe perioadele de vârf (ex.: promoții planificate) și să atenueze scăderile prin stimulente pentru clienți.

Investigația factorilor care afectează performanța magazinelor (ex.: locație, dimensiune, concurență locală) poate descoperi oportunități de îmbunătățire.

Secvența de cod;

```
proc sgplot data=WORK.SET_DE_DATE_WALMART_PS;
    series x=Date y=Weekly_Sales / group=Store;
    title "Evoluția vânzărilor săptămânale pe magazine";
    xaxis label="Data";
    yaxis label="Vânzări săptămânale";
run;
title;
```

8. Formate definite de utilizator;

Se dorește crearea unui format personalizat cu ajutorul căruia sărbătorile nu vor mai fi afișate ca 0,1, ci “Nu e sărbătoare” sau “E sărbătoare”.

Setul de date cu noul format

Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	CPI	Unemployment	Fuel_Price
1	1	05/02/2010	1643690.9	Nu e sarbatoare	42.31	211.0963582	8.106	2.572
2	1	11/06/2010	1542561.09	Nu e sarbatoare	80.43	211.4580951	7.808	2.572
3	1	09/04/2010	1545418.53	Nu e sarbatoare	65.86	210.6228574	7.808	2.572
4	1	19/03/2010	1472515.79	Nu e sarbatoare	54.58	211.215635	8.106	2.572
5	1	28/05/2010	1432069.95	Nu e sarbatoare	80.44	210.8967606	7.808	2.572
6	1	07/05/2010	1603955.12	Nu e sarbatoare	72.55	210.3399684	7.808	2.572
7	1	30/04/2010	1425100.71	Nu e sarbatoare	67.41	210.3895456	7.808	2.572
8	1	26/03/2010	1404429.92	Nu e sarbatoare	51.45	211.0180424	8.106	2.572
9	1	14/05/2010	1494251.5	Nu e sarbatoare	74.78	210.3374261	7.808	2.572
10	1	05/03/2010	1554806.68	Nu e sarbatoare	46.5	211.3501429	8.106	2.572

Secvența de cod:

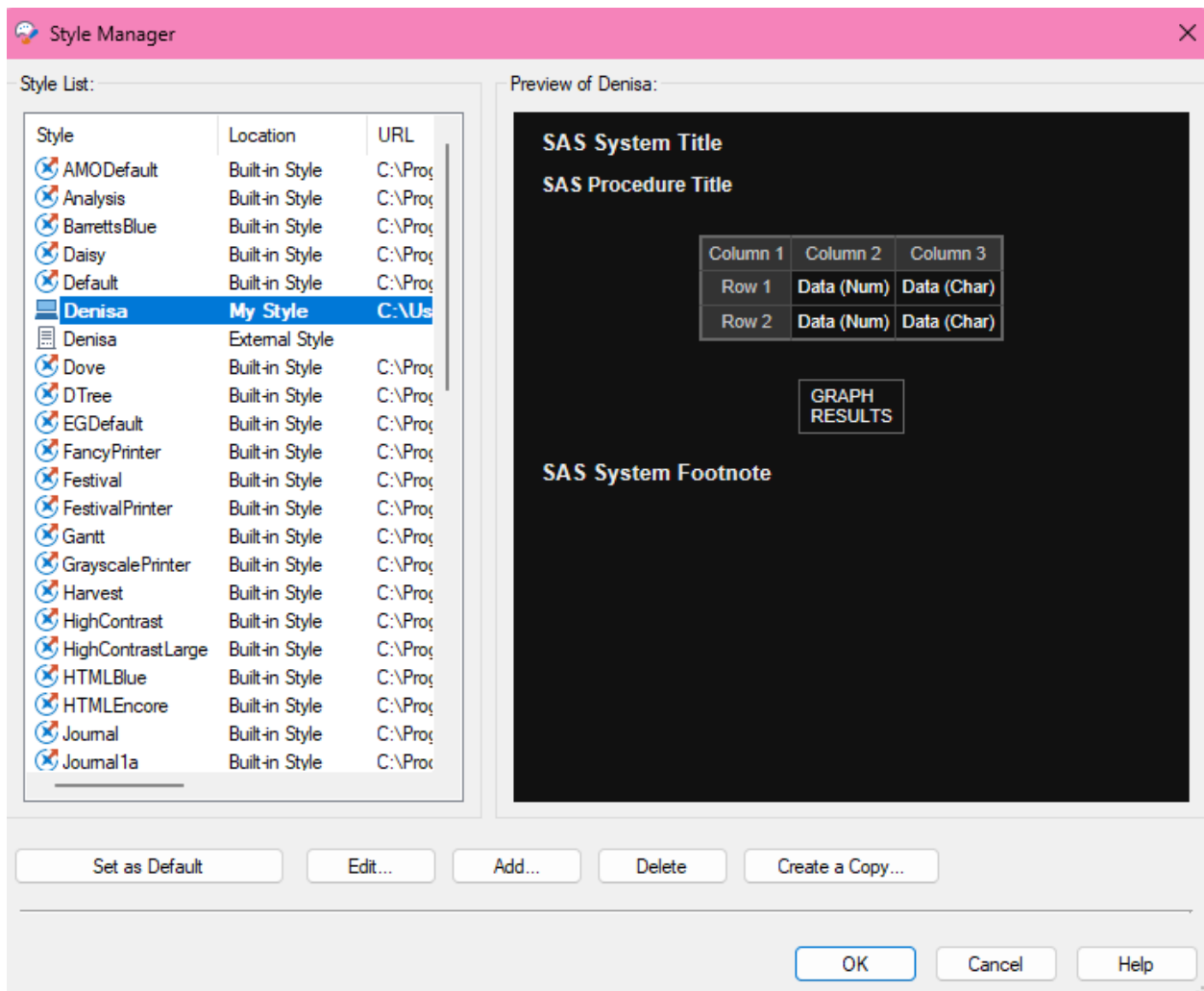
```
proc format;
    value holidayflag 0='Nu e sarbatoare' 1='E sarbatoare';
run;
```

```
proc print data=WORK.jonctiune (obs=10);
    format Holiday_Flag holidayflag.;
    title "Setul de date cu noul format";
run;
title;
```

10. Stiluri create de utilizator.

Se se creeze un nou stil personalizat de afisare in HTML.

Pentru acest lucru, se va accesa Tools->Style Manager->Add New Style->Add new based on existing style (selectam stilul dorit si ii dam un nume personalizat)->Ok->Ok



7. Prelucrări statistice;

Sa se genereze un raport statistic pentru pretul combustibilului, in functie de numarul magazinului.

Vom analiza spre exemplu, magazinul 2.

Statistici pentru prețul combustibilului

The UNIVARIATE Procedure

Variable: Fuel_Price

Store = 2

Moments			
N	400	Sum Weights	400
Mean	2.70475	Sum Observations	1081.9
Std Deviation	0.10006458	Variance	0.01001292
Skewness	-0.3509787	Kurtosis	-1.0076751
Uncorrected SS	2930.26418	Corrected SS	3.995155
Coeff Variation	3.69958695	Std Error Mean	0.00500323

Basic Statistical Measures			
Location		Variability	
Mean	2.704750	Std Deviation	0.10006
Median	2.719500	Variance	0.01001
Mode	2.514000	Range	0.34000
		Interquartile Range	0.15650

Note: The mode displayed is the smallest of 20 modes with a count of 20.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	540.6009	Pr > t	<.0001
Sign	M	200	Pr >= M	<.0001
Signed Rank	S	40100	Pr >= S	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	2.8540
99%	2.8540
95%	2.8445
90%	2.8305
75% Q3	2.7875
50% Median	2.7195
25% Q1	2.6310
10%	2.5545
5%	2.5310
1%	2.5140
0% Min	2.5140

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
2.514	441	2.854	697
2.514	440	2.854	698
2.514	439	2.854	699
2.514	438	2.854	700
2.514	437	2.854	701

1. Tendințe Centrale și Variabilitate

- Prețul mediu al combustibilului este 2.70475 USD, cu o abatere standard mică (0.10006 USD), indicând stabilitate relativă pe perioada analizată.
- Mediana (2.7195 USD) este ușor mai mare decât media, sugerând o asimetrie ușoară spre stânga (confirmată de skewness negativ: -0.35), adică mai multe valori sub medie.
- Moda (2.514 USD, cea mai mică dintre 20 de valori modale) apare de 20 de ori, marcând un preț frecvent întâlnit la limita inferioară.

2. Distribuție și Valori Extreme

- Interval restrâns: Prețurile variază între 2.514 USD (minim) și 2.854 USD (maxim), cu un range de doar 0.34 USD.
- Cuartile:
 - 75% din observații sunt sub 2.7875 USD, iar 25% sub 2.631 USD.
 - 95% din prețuri sunt sub 2.8445 USD, iar doar 1% ating minimul (2.514 USD).
- Kurtosis negativ (-1.008) arată o distribuție mai plată decât cea normală, cu valori mai puțin concentrate în cozi.

3. Teste de Semnificație

- Toate testele (t-test, sign test, signed rank) confirmă că prețul mediu este semnificativ diferit de zero ($p < 0.0001$), ceea ce este de așteptat pentru o variabilă economică.

4. Observații Extreme

- Valorile minime (2.514 USD) și maxime (2.854 USD) apar în blocuri (ex.: observațiile 437–441 pentru minim, 697–701 pentru maxim), sugerând schimbări bruște în preț (ex.: ajustări la nivel macroeconomic sau variații sezoniere).

- Prețul combustibilului pentru magazinul 2 a fost relativ stabil (coeficient de variație: 3.7%), ceea ce este benefic pentru:
 - Planificarea bugetară (costuri logistice anticipate).
 - Strategii de prețuri (impact redus al fluctuațiilor combustibilului asupra costurilor de transport).
- Minimele (2.514 USD) și maximele (2.854 USD) pot fi legate de:
 - Schimbări în prețurile petrolului la nivel global.
 - Promoții locale sau taxe aplicabile în anumite perioade.
 - Sezonalitate (ex.: cerere crescută în vacanțe).

Recomandări pentru Walmart:

- Monitorizare continuă: Deoarece prețul combustibilului afectează costurile de aprovizionare, este utilă analiza în timp real a datelor pentru a anticipa tendințe.
- Optimizare logistică: Pe perioadele cu prețuri mai mici (ex.: sub 2.63 USD, Q1), Walmart poate intensifica livrările pentru a reduce costurile.
- Comparatie între magazine: Dacă alte magazine au variații mai mari, ar putea fi necesară investigarea diferențelor regionale (ex.: taxe locale, concurență).

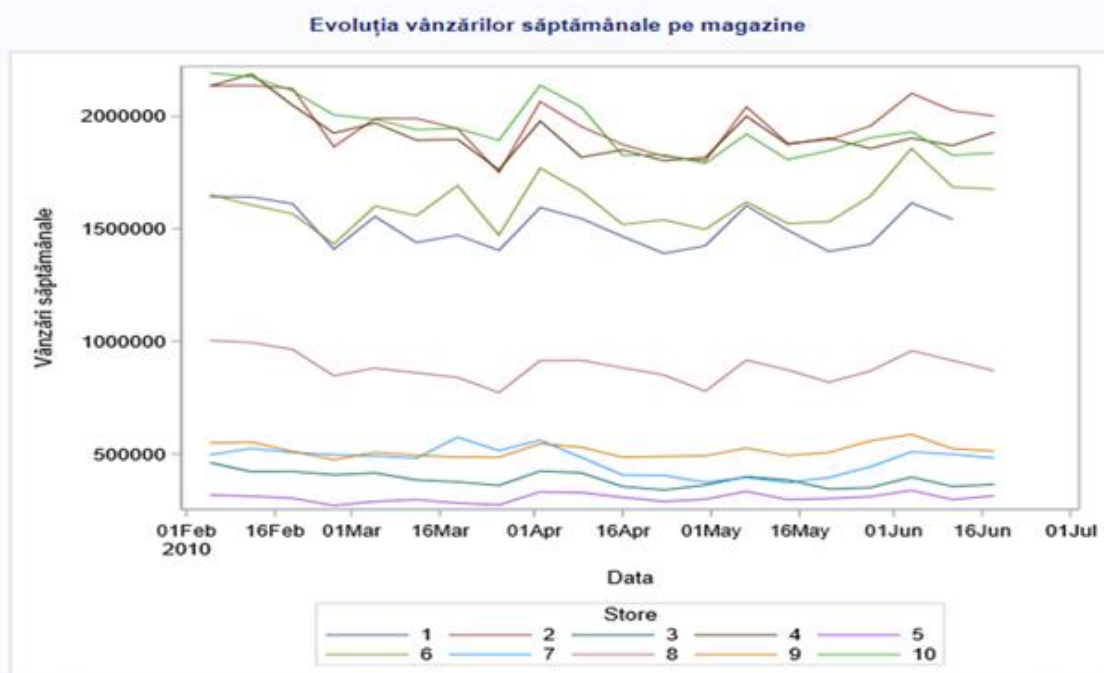
Secventa de cod:

```
proc UNIVARIATE data=WORK.jonctiune;
  var Fuel_Price;
  class Store;
  title "Statistici pentru prețul combustibilului";
run;
title;
```

9. Crearea unui document compus;

Se dorește crearea unui document compus de tip raport care să conțină Evoluția vânzărilor săptămânale pe magazine și magazinele care ating pragul de vânzări săptămânale de 2 milioane USD.

Se va proceda in urmatoarul fel: File->New->Report->Insert image (Se vor selecta imaginile necesare si se vor incarca)->Ok



Raport - Vânzări săptămânale peste 2 milioane

Magazin	Data	Vânzări săptămânale
2	02/04/2010	\$2.066,188
	16/02/2010	\$2.124,452
	12/02/2010	\$2.137,810
	05/02/2010	\$2.136,989
	11/06/2010	\$2.025,539
	04/06/2010	\$2.102,540
	18/06/2010	\$2.001,637
	07/05/2010	\$2.042,582
	02/04/2010	\$2.066,188
	16/02/2010	\$2.124,452
	12/02/2010	\$2.137,810
	05/02/2010	\$2.136,989
	11/06/2010	\$2.025,539
	04/06/2010	\$2.102,540
	18/06/2010	\$2.001,637
	07/05/2010	\$2.042,582
	02/04/2010	\$2.066,188
	16/02/2010	\$2.124,452
	12/02/2010	\$2.137,810
	05/02/2010	\$2.136,989
	11/06/2010	\$2.025,539
	04/06/2010	\$2.102,540
	18/06/2010	\$2.001,637
	07/05/2010	\$2.042,582
	02/04/2010	\$2.066,188
	16/02/2010	\$2.124,452
	12/02/2010	\$2.137,810

CONCLUZII FINALE

1. Performanța Vânzărilor și Factori de Influență

- Vânzările săptămânale prezintă variații semnificative (296,674 USD – 2,193,048 USD), evidențiind:
 - Sezonalitate puternică, cu vârfuri asociate sărbătorilor (ex.: Black Friday) și scăderi în perioadele post-sărbători.
 - Impactul factorilor externi: Temperaturile scăzute (ex.: Magazinul 7, -11.9°C) reduc vânzările, în timp ce cele ridicate (ex.: Magazinul 10, 32°C) le stimulează.
 - Diferențe regionale: Magazinele din zone urbane (ex.: Magazinul 10, preț combustibil 3.01 USD) au vânzări maxime, în timp ce cele din zone rurale (ex.: Magazinul 5, 337,825 USD) au performanțe slabe.

2. Clusterizarea și Segmentarea Pieței

- Clusterizarea bazată pe vânzări și șomaj a identificat 3 grupuri:
 - Cluster 0 (Vânzări mici, șomaj ridicat): Necesită promoții și gestionare dinamică a stocurilor.
 - Cluster 2 (Vânzări mari, șomaj scăzut): Potențial pentru produse premium și colaborări cu autoritățile locale.
- Recomandare: Adaptarea strategiilor de marketing și a sortimentului în funcție de cluster.

3. Regresia Multiplă și Predictorii

- Modelul statistic explică 32.2% din variația vânzărilor, cu următoarele concluzii:
 - Șomajul are un impact pozitiv (contraintuitiv), sugerând că clienții optează pentru Walmart în loc de alternative mai scumpe.
 - Temperatura influențează pozitiv vânzările, reflectând sezonabilitatea (ex.: creșterea cererii de produse de vară).
 - Limitări: Alți factori (promoții, concurență) ar putea îmbunătăți acuratețea modelului.

4. Prețul Combustibilului și Logistica

- Stabilitate (media 2.70 USD, abatere standard 0.10 USD) în majoritatea magazinelor, cu excepția Magazinului 10 (3.01 USD, zonă urbană).

- Implicații: Costuri logistice predictibile, cu oportunități de optimizare a aprovizionării în perioadele cu prețuri scăzute.

5. Recomandări Strategice pentru Expansiune

- Optimizarea Operațională:
 - Pentru magazinele cu vânzări slabe (ex.: Magazinul 5): Reevaluarea poziționării pe piață, promoții agresive.
 - Pentru magazinele cu vânzări mari (ex.: Magazinele 2, 4, 10): Extinderea rețelei în zone similare (urbane, cu cerere stabilă).
- Îmbunătățirea Analizei Datelor:
 - Implementarea de modele predictive pentru a anticipa sezonalitya și impactul promoțiilor.
 - Monitorizarea în timp real a factorilor externi (temperatură, șomaj).
- Diversificarea Ofertei:
 - Introducerea de produse adaptate climatului (ex.: articole de iarnă în zonele reci).
 - Colaborări cu furnizorii locali pentru a reduce costurile logistice.

6. Concluzie Finală

Studiul a evidențiat oportunități clare de expansiune pentru Walmart, bazate pe:

1. Capitalizarea pe sezonalitya și evenimente cheie.
2. Extinderea în zone urbane cu cerere ridicată și prețuri stabile la combustibil.
3. Personalizarea strategiilor pe clusteruri de magazine.
4. Investiții în analiza datelor pentru a îmbunătăți predictibilitatea și eficiența operațională.

Prin abordarea acestor recomandări, Walmart poate consolida poziția pe piață, reduce riscurile asociate variațiilor de vânzări și exploata oportunitățile de creștere sustenabilă.