A.1. Să se creeze un set de date SAS din fișierul extern “Electronice\_vanzari2.csv”.

B. Setul de date folosit este “Electronice\_vanzari2.csv”

C.

data Electronice\_vanzari2;

infile '/home/u63371627/Proiect/Electronice\_vanzari2.csv' dsd firstobs=2 truncover;

input Nume\_Produs $ Categorie $ Unitati\_vandute Pret\_pe\_unitate Vanzari\_totale Data\_lansare mmddyy10.;

format Data\_lansare date9.;

proc print data=Electronice\_vanzari2;

run;

D.

A screenshot of a table

Description automatically generated with low confidence

1. 2. Să se creeze un set de date numit "Unități" care conține informații despre vânzările unor categorii de produse. Setul de date are două variabile: Categorie (numele categoriei de produs) și Vânzare (valoarea vânzărilor pentru fiecare categorie). Se dorește să se afișeze datele despre vânzări și să se aplice un format pentru variabilă Vânzare, astfel încât să clasifice vânzările în categorii "Scăzute", "Medii" și "Ridicate" în funcție de valoarea lor.
2. Am creat un set de date numit "Unități" din “Electronice\_vanzari2.csv”

proc format;

value Vanzare

low -< 1000 = 'Scăzute'

1000 -< 3000 = 'Medii'

3000 - high = 'Ridicate' ;

run;

data Unitati;

input Categorie$ Vanzare;

datalines;

iphone 4454.8

samsung 2316.5

sony 6377.1

lg 472.24

dell 902.6

;

title "Date despre numarul de vanzari";

proc print data=Unitati;

var Categorie Vanzare;

format Categorie $Categorie.

Vanzare Vanzare.;

run;

D.



1. Interpretare

Setul de date "Unitati" conține informații despre vânzările înregistrate pentru diverse categorii de produse. Utilizând formatul personalizat "Vanzare", valorile variabilei sunt afișate ca "Scăzute", "Medii" sau "Ridicate", în funcție de intervalul în care se încadrează.

1. 3. Să se creeze un set de date numit "vanzari\_angajati" care conține informații despre vânzările lunare ale mai multor angajați. Setul de date are trei variabile: Numele angajatului, lună în care s‑au realizat vânzările și valoarea vânzărilor. Dorim să identificăm și să afișăm numele și vânzările maxime obținute de un angajat într‑o singură lună.
2. Am creat un set de date numit "vanzari\_angajati"

data vanzari\_angajati;

infile datalines dlm=' ';

input Nume $ Luna $ Vanzari;

datalines;

Ion ianuarie 5000

Maria ianuarie 7000

Mihai ianuarie 6000

Emilia ianuarie 8000

Ion februarie 9000

Maria februarie 6000

Mihai februarie 8000

Emilia februarie 7500

;

run;

data Vanzari\_maxime;

set vanzari\_angajati;

if Vanzari > 7000 then do;

max\_vanzari = Vanzari;

Nume\_max = Nume;

end;

title "Date despre vanzarile angajatilor";

proc print data=Vanzari\_maxime;

run;

A picture containing text, screenshot, number, font

Description automatically generated

1. Interpretare

În coloană "max\_vanzari" observăm vânzările care au trecurt de pragul stabilit, iar în coloană "nume\_max" numele asociat angajaților care au obținut rezultatele vânzărilor.

1. 4. Să se creeze un subset de date numit "subset\_vanzari" din setul de date existent "vanzari\_angajati" care conține doar înregistrările în care valoarea de "Vânzări" este mai mare de 7000.
2. Am creat un subset de date numit "subset\_vanzari" din "vanzari\_angajati"

data subset\_vanzari;

set vanzari\_angajati;

where Vanzari > 7000;

title "Subsetul din Vanzari\_Angajati";

proc print data=subset\_vanzari;

run;

A screenshot of a computer

Description automatically generated with low confidence

1. Interpretare

In aest subset am folosit datele create anterior și se aplică o condiție “where” pentru a selecta doar înregistrările în care valoarea variabilei "Vanzari" este mai mare decât 7000. Aceste înregistrări care îndeplinesc condiția sunt incluse în noul set de date "subset\_vanzari", împreuna cu numele angajatului.

1. 5. Folosind funcțiile șum (), intck () și count () în SAS, generați următoarele informații:

• Să se calculeze suma totală a vânzărilor pentru toate produsele electronice.

• Să se determine diferența în zile între dată lansării fiecărui produs și dată curentă. Această diferență să fie stocată într‑un nou set de date numit "dif\_data".

• Să se calculeze numărul de înregistrări pentru fiecare categorie de produse și să se afișeze rezultatul.

1. Am creat un set de date numit "Functii" din “Electronice\_vanzari2.csv”

data Functii;

infile '/home/u63371627/Proiect/Electronice\_vanzari2.csv' dsd firstobs=2 truncover;

input Nume\_Produs $ Categorie $ Unitati\_vandute Pret\_pe\_unitate Vanzari\_totale Data\_lansare mmddyy10.;

format Data\_lansare mmddyy10.;

/\* sum() \*/

proc sql;

select sum(Vanzari\_totale) as Suma\_totala\_vanzari

from Functii;

quit;

/\* intck() \*/

data dif\_data;

set Functii;

Zile\_diferenta = intck('day', Data\_lansare, today());

title "Functia intck() - diferenta de zile";

proc print data=dif\_data;

run;

/\*count() \*/

proc sql;

select Categorie, count(\*) as Numar\_inregistrari

from Functii

group by Categorie;

quit;



A picture containing text, screenshot, number, font

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

1. Interpretare

Folosind funcția sum(), s-a creat o mică tabelă cu rezultatul sumei tuturor produselor din “Electronice\_vanzari2.csv” . Folosind funcția intck(), s-a creat coloana “zile\_diferenta” sunt determinate zilele între dată lansării fiecărui produs și dată curentă. Folosind funcția count(), s-a creat o tabelă unde s-au calculat numărul de înregistrări pentru fiecare categorie de produse și categoria acestora.

1. 6. Utilizând proceduri specifice de SQL și SAS combinați setul de date "Funcții" împreună cu "Electronice\_vanzari2". Având setul de date combinat "CombinedData".
2. Am creat setul de date “CombinedData" din "Funcții" si "Electronice\_vanzari2.csv"

data Functii;

infile '/home/u63371627/Proiect/Electronice\_vanzari2.csv' dsd firstobs=2 truncover;

input Nume\_Produs $ Categorie $ Unitati\_vandute Pret\_pe\_unitate Vanzari\_totale Data\_lansare mmddyy10.;

format Data\_lansare mmddyy10.;

run;

/\* Setul de date rezultat din combinația utilizând procedura SQL \*/

proc sql;

create table CombinedData as

select A.\*, B.Numar\_inregistrari

from Functii A

left join

(

select Categorie, count(\*) as Numar\_inregistrari

from Functii

group by Categorie

) B

on A.Categorie = B.Categorie;

quit;

proc print data=CombinedData;

run;

A picture containing text, screenshot, number

Description automatically generated

1. Interpretare

Utilizând proceduri specifice de SQL și SAS s-au combinat setul de date "Funcții" împreună cu "Electronice\_vanzari2". Noul set "CombinedData" conține informații despre vânzările de produse electronice, inclusiv numele produsului, categoria, unitățile vândute, prețul pe unitate, vânzările totale și numărul de înregistrări pentru fiecare categorie.

1. 7. Porinind de la "Electronice\_vanzari2", să se creeze un nou set de date numit "Masive" care să calculeze suma vânzărilor dar și media prețurilor.
2. Am creat setul de date “Masive din “Electronice\_vanzari2.csv”

data Masive;

infile '/home/u63371627/Proiect/Electronice\_vanzari2.csv' dsd firstobs=2 truncover;

input Nume\_Produs $ Categorie $ Unitati\_vandute Pret\_pe\_unitate Vanzari\_totale Data\_lansare mmddyy10.;

format Data\_lansare date9.;

run;

/\* Calcularea sumei totale a vânzărilor \*/

data sum\_masiv;

set Masive;

sum\_vanzari + Vanzari\_totale;

run;

/\* Calcularea mediei prețului pe unitate pentru fiecare categorie \*/

proc sql;

create table Medii\_preturi as

select Categorie, mean(Pret\_pe\_unitate) as Medie\_pret\_unitate

from Masive

group by Categorie;

quit;

title "Suma totală a vânzărilor";

proc print data=sum\_masiv;

run;

title "Media prețului pe unitate pentru fiecare categorie";

proc print data=Medii\_preturi;

run;

A picture containing text, screenshot, number, font

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

1. 8. Analizați un set de date despre vânzările de produse electronice și generați următoarele informații:

• Numărați frecvența de apariție a fiecărei categorii de produse din setul de date.

• Calculați valorile medii pentru variabilele "Unitati\_vandute", "Pret\_pe\_unitate" și "Vanzari\_totale".

• Generați un raport care să afișeze categoria de produse și suma vânzărilor totale pentru fiecare categorie.

1. Am folosit setul de date “Electronice\_vanzari2.csv”

PROC IMPORT DATAFILE='/home/u63371627/Proiect/Electronice\_vanzari2.csv'

OUT=work.raport

DBMS=CSV;

RUN;

/\*frecvența de apariție \*/

PROC FREQ DATA=work.raport;

TABLES Categorie;

RUN;

/\*valorile medii pentru variabilele "Unitati\_vandute", "Pret\_pe\_unitate" și "Vanzari\_totale" \*/

PROC MEANS DATA=work.raport;

VAR Unitati\_vandute Pret\_pe\_unitate Vanzari\_totale;

RUN;

/\*raport pentru categoria de produse și suma vânzărilor \*/

PROC REPORT DATA=work.raport;

COLUMN Categorie Vanzari\_totale;

DEFINE Categorie / GROUP;

DEFINE Vanzari\_totale / ANALYSIS SUM;

RUN;

A picture containing text, screenshot, font, number

Description automatically generated

A picture containing text, screenshot, font, line

Description automatically generated

A screenshot of a computer

Description automatically generated with low confidence

1. Interpretare

Am analizați un set de date despre vânzările de produse electronice și am generat frecvența de apariție a fiecărei categorii de produse din setul de date, valorile medii pentru variabilele "Unitati\_vandute", "Pret\_pe\_unitate" și "Vanzari\_totale" si un raport care să afișeze categoria de produse și suma vânzărilor totale pentru fiecare categorie.

1. 9. Folosind procedurile statistice, să se analizeze relația dintre prețul pe unitate și vânzările totale și să se determine dacă există o corelație semnificativă între aceste două variabile.
2. Am folosit setul de date “Electronice\_vanzari2.csv”

PROC IMPORT DATAFILE='/home/u63371627/Proiect/Electronice\_vanzari2.csv'

OUT=work.raport

DBMS=CSV;

RUN;

/\* medie\*/

PROC MEANS DATA=work.raport SUM;

VAR Vanzari\_totale;

CLASS Categorie;

OUTPUT OUT=Suma\_vanzari\_per\_categorie SUM=Suma\_vanzari;

RUN;

PROC SORT DATA=Suma\_vanzari\_per\_categorie;

BY descending Suma\_vanzari;

RUN;

DATA Categorie\_max\_vanzari;

SET Suma\_vanzari\_per\_categorie;

IF \_N\_ = 1; /\* IF FIRST.Categorie; \*/

RUN;

PROC PRINT DATA=Categorie\_max\_vanzari;

RUN;

/\* corelatie\*/

PROC CORR DATA=work.raport;

VAR Unitati\_vandute Pret\_pe\_unitate Vanzari\_totale;

RUN;

/\* regresie\*/

PROC REG DATA=work.raport;

MODEL Vanzari\_totale = Unitati\_vandute Pret\_pe\_unitate;

RUN;

A screenshot of a computer screen

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated with medium confidence

A picture containing text, screenshot, diagram, line

Description automatically generated

1. Interpretare
2. 10. Generați grafice de tip Histogramă și diagrama de Dispersie pentru vânzările totale în funcție de unitățile din fiecare categorie de produse.
3. Am folosit setul de date “Electronice\_vanzari2.csv”

PROC IMPORT DATAFILE='/home/u63371627/Proiect/Electronice\_vanzari2.csv'

OUT=work.raport

DBMS=CSV;

RUN;

/\* Histograma \*/

PROC SGPLOT DATA=work.raport;

HISTOGRAM Vanzari\_totale;

RUN;

/\* Diagrama de dispersie \*/

PROC SGPANEL DATA=work.raport;

PANELBY Categorie;

SCATTER x=Pret\_pe\_unitate y=Vanzari\_totale;

RUN;

A picture containing diagram, screenshot, text, rectangle

Description automatically generated

A picture containing text, screenshot, diagram, number

Description automatically generated

1. Interpretare