Alegeți un set de date care să conțină diverse tipuri de variabile (de exemplu, cantitative, calitative, atributive, spațiale, temporale). Pentru diversitatea variabilelor, se pot folosi mai multe seturi de date în același proiect. Documentul final va include și codul R utilizat.

Cerințele vor fi adăugate săptămânal în documentul de proiect. La finalul semestrului, fiecare echipă, formată din 2-3 studenți, va preda documentul completat săptămânal.

Pentru exemplificare, se va folosi setul de date „SuperMarket\_Analysis”. Exemplele vor fi prezentate schematic, proiectul fiind completat cu noțiuni teoretice preluate din curs.

Penultima și ultima săptămână din semestru vor fi dedicate predărilor (ianuarie). Predările vor fi programate într-un document Google Sheet (decembrie).

**Coperta**

Student 1:

Student 2:

Student 3:

Screenshot cu toate variabilele și primele 10 observații

A screenshot of a computer

Description automatically generated

A screenshot of a table

Description automatically generated

**Săptămâna 21-25 octombrie**

Cerința1. Clasificarea variabilelor

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variabila** | **Invoice\_ID** | **Branch** | **City** | **Customer\_type** | **Gender** | **Product\_line** | **Unit\_price** | **Quantity** | **Tax\_5%** |
| Natura | atributivă | atributivă | de spațiu | atributivă | atributivă | atributivă | atributivă | atributivă | atributivă |
| Exprimare | calitativă | calitativă | calitativă | calitativă | calitativă | calitativă | cantitativă | cantitativă | cantitativă |
| Variație | - | - | - | - | - | - | continuă | discretă | continuă |
|  | | | | | | | | | |
| **Variabila** | **Sales** | **Date** | **Time** | **Payment** | **Cogs** | **Gross\_margin** | **Gross\_income** | **Rating** |  |
| Natura | atributivă | de timp | de timp | atributivă | atributivă | atributivă | atributivă |  |  |
| Exprimare | cantitativă | calitativă | calitativă | calitativă | cantitativă | cantitativă | cantitativă |  |  |
| Variație | continuă | - | - | - | continuă | continuă | continuă |  |  |

Cerința 2. Reprezentări grafice și comentarii/interpretări.

2.1. O histogramă

|  |
| --- |
|  |

Histograma arată că prețurile unitare sunt distribuite uniform între 0 și 10.000, fără să existe o concentrație foarte mare într-un anumit interval. Aceasta sugerează că datele analizate conțin o varietate de prețuri, iar acest lucru ar putea fi util pentru a înțelege variația prețurilor în contextul respectiv.

2.2. O diagramă prin benzi

|  |
| --- |
|  |

Comentariu: Se observă că numărul persoanelor de gen feminin (aproximativ 600) este mai mare decât cel al persoanelor de gen masculin (aproximativ 450). Diferența dintre cele două categorii este vizibilă, sugerând o pondere mai mare a femeilor în acest set de date. Această reprezentare grafică subliniază o distribuție inegală între genuri, cu un număr semnificativ mai mare de femei.

2.3. Un cerc de sructură (engl.piechart)

A pie chart with different colored circles

Description automatically generated

Comentariu:

|  |
| --- |
|  |

Cerința 3. Reprezentarea grafică a unei relații dintre două variabile și comentarii/interpetări

3.1. Diagrama norului de puncte

A graph of a sales curve

Description automatically generated with medium confidence

Comentariu:

3.2. Boxplot

A diagram of a payment method

Description automatically generated

Comentariu:

**Cod R**

# Săptămâna 21-25 octombrie

data = SuperMarket\_Analysis

colnames(SuperMarket\_Analysis)

hist(SuperMarket\_Analysis$Unit\_price, main = "Histograma Prețului Unitar", xlab ="Prețul Unitar", ylab="Frecvența", ylim = c(0,110))

barplot(table(SuperMarket\_Analysis$Gender),main = "Reprezentarea Genului", xlab ="Genul", ylab="Frecventa", ylim =c(0,600), col = "blue")

pie(table(SuperMarket\_Analysis$City),main ="Reprezentarea Orașelor")

plot(SuperMarket\_Analysis$Gross\_income ~ SuperMarket\_Analysis$Sales, xlab ="Gross Income", ylab="Sales",ylim=c(0,600000), main ="Gross Income versus Sales")

boxplot(SuperMarket\_Analysis$Unit\_price ~ SuperMarket\_Analysis$Payment,xlab ="Payment", ylab="Unit Price", main ="Unit price per Payment")