**Raport de analiza a datelor**

**1.Identificarea nevoilor de business**

Domeniul de business: Vanzari supermarket

Problema specifica: Analiza performatei vanzarilor pentru a identifica tendintele si oportunitatile de crestere a vanzarilor.

Intrebare de cercetare: Care sunt produsele cele mai populare si care sunt tendintele de vanzare in functie de diferiti factori precum: locatia, data?

Pentru a stabili tipurile de analiza si rezultatele necesare vom tine cont de urmatori factori:

* Analiza descriptiva a vanzarilor: medie, mediana, modul, variatie, deviatie standard.
* Identificarea produselor cu cele mai mici si mai mari vanzari.
* Tendintele de vanzare pe diferite segmente precum: data, locatie.

**2.Colectarea datelor**

Sursa datelor este site-ul Kaggle Datasets de unde am furnizat fisierul supermaket\_sales-Sheet1.csv. Datele au fost colectate printr-un fisier CSV, iar mai apoi importate in Google Sheets. Acest lucru poate sugera utilizarea unui sistem de gestiune ce furnizeaza datele in astfel de format.

Nu am intampinat provocari specifice , dar provocarile tipice ar putea include date lipsa sau incorect.

**3.Evaluarea calitatii datelor**

Datele vor fi evaluate dupa 3 criterii importante: acuratete, completitudine si actualitate.

Acuratete: Datele sunt exportate dintr-un sistem de vanzari, probabilitatea ca fiind precise este destul de mare.

Completitudine: Vom verifica completitudinea datelor.

Actualitate: Verificarea actualitatii datelor in conformitate cu cerintele uzuale.

Nu am identificate probleme majore legate de calitatea datelor.

**4.Curatarea datelor in Google Sheets**

Tehnicile aplicate pentru a curata setul de date sunt:

* Eliminarea erorilor si valorilor lipsa
* Eliminarea duplicatelor
* Corectarea erorilor de formatare

**5. Normalizarea Datelor**

Datele vor fi normalizate dupa doua criterii:

* Standardizare: Formatele date vor fi standardizate pentru a asigura consistenta.
* Transformari necesare: Daca este necesar, se vor aplica transformari precum scalarea datelor.

**6.Aplicarea conceptelor statistice in Google Sheets**

Pentru a putea utiliza conceptele statistice in Google Sheets vom utiliza urmatoarele formule:

* Media: =AVERAGE(range)
* Mediana: =MEDIAN(range)
* Modulul: =MODE(range)
* Variatia: =VAR.S(range)
* Deviatia standard: STDEV.S(range)

**7.Activitate practica: Analiza unui set de date**

Pentru a analiza un set de date vom aplica urmatoarele tehnici:

* Curatarea si normalizarea datelor in Google Sheets
* Realizarea unei analize statistice realizate
* Prezentarea analizei intr-un raport final

Voi analiza si curata fisierul supermarket\_sales-Sheet1.csv .

Dupa examinarea intiala a datelor observam urmatoarele:

* Acuratetea dateor este ridicata, toate coloanele fiind complete.
* Fiecare coloanal are 500 de intrari neavand valori lipsa.
* Datele sunt din 2019, actualizate recent.

Nu am identificat probleme majore de calitate a datelor .

**Curatarea Datelor in Google Sheets**

Pentru a curata datele vom avea in vedere utilizarea urmatoareloe tehnici:

* Elimiarea erorilor si valorilor lipsa: Nu exista valori lipsa.
* Eliminarea duplicatelor: verificarea acestora se face in coloanal Invoice ID
* Corectarea erorilor de formatare: Verifcam formatul datelor in coloanele Date,Time si daca datele respecta scriere la incputul fiecarei celule cu majuscule.

Normalizarea datelor

1. Standardizare: Vom transforma coloanele Date si Time in formate standardizate pentru o analiza ulterioara.
2. Transformari necesare: Nu avem nevoie de transformari pentru acest de date
3. Scrierea cu majuscule a cuvintelor din celule.

**Analiza Statistică**

Vom calcula media, mediana, modul, varianța și deviația standard pentru variabilele de interes (Unit price, Quantity, Total, Rating) în Google Sheets.

**Pași Practici pentru Google Sheets**

1. **Importați fișierul CSV în Google Sheets.**
2. **Curățarea datelor:**
   * Verificați duplicatele și eliminați-le, dacă există.
   * Corectați formatele coloanelor Date și Time (de exemplu, transformați data în format AAAA-LL-ZZ).
3. **Normalizarea datelor:**
   * Standardizați formatele pentru date și ore.

**Calcule Statistice în Google Sheets:**

* **Media:** =AVERAGE(range)
* **Mediana:** =MEDIAN(range)
* **Modul:** =MODE(range)
* **Varianța:** =VAR.S(range)
* **Deviația standard:** =STDEV.S(range)

**Rezultatele Analizei Statistice**

**1. Unit Price (Preț pe unitate)**

* **Media:** 54,85
* **Mediana:** 52.43
* **Modul:** 40.30
* **Varianța:** 720.84
* **Deviația Standard:** 26.85

**2. Quantity (Cantitate)**

* **Media:** 5.69
* **Mediana:** 6.0
* **Modul:** 10
* **Varianța:** 8.41
* **Deviația Standard:** 2.899

**3. Total**

* **Media:** 361
* **Mediana:** 293
* **Modul:** 276.95
* **Varianța:** 67201,70
* **Deviația Standard:** 259,23

**4. Rating**

* **Media:** 7.019
* **Mediana:** 7.0
* **Modul:** 6.5
* **Varianța:** 2.96
* **Deviația Standard:** 1.72

**Interpretarea Rezultatelor**

1. **Unit Price:**
   * Media și mediana sunt foarte apropiate, indicând o distribuție echilibrată a prețurilor unitare.
   * Varianța și deviația standard relativ mari sugerează o dispersie semnificativă a prețurilor între produsele vândute.
2. **Quantity:**
   * Media și mediana indică faptul că majoritatea tranzacțiilor implică achiziționarea a aproximativ 5 unități.
   * Modulul la 10 sugerează că există un număr semnificativ de tranzacții cu exact 10 unități.
   * Varianța și deviația standard moderate sugerează o anumită variabilitate, dar nu extremă.
3. **Total:**
   * Media și mediana diferă semnificativ, indicând o distribuție asimetrică a totalurilor tranzacțiilor, cu câteva tranzacții cu sume foarte mari care ridică media.
   * Varianța și deviația standard mari confirmă această variabilitate mare a sumelor totale ale tranzacțiilor.
4. **Rating:**
   * Valorile de rating sunt destul de uniforme, cu o medie aproape de 7.
   * Varianța și deviația standard scăzute indică o distribuție compactă a ratingurilor.

### Concluzie

Acest raport detaliază procesul de curățare și analiză a datelor de vânzări ale unui supermarket, utilizând concepte statistice pentru a oferi perspective valoroase asupra performanței produselor și comportamentului clienților. Următorul pas este implementarea acestor analize în Google Sheets pentru o colaborare eficientă și actualizări dinamice.

Mai jos avem atasat link-ul spre documentul google sheets:

https://docs.google.com/spreadsheets/d/1cUoNWJIhcta0uiNW5QADJv5HqVTY5TU9V7Qc5jz-6sk/edit?usp=sharing