

## **Task: Familiarizarea cu datele magazinului online**

**Curs: Data Analysis and Processing using Python**

**Modul: Familiarizarea cu datele și structura acestora**

### **Obiectivul sarcinii:**

Dezvoltați un script Python care permite analiza inițială și orientarea într-un set de date ce conține informații despre produsele unui magazin online. Accentul se pune pe utilizarea funcționalităților de bază ale bibliotecii Pandas pentru a încărca, vizualiza, filtra și sorta datele.

### **Contextul sarcinii:**

Alex dorește să aplice tot ceea ce a învățat în analiza datelor bibliotecii și la locul său de muncă, pentru a îmbunătăți activitățile de marketing și întregul proces de afaceri al magazinului. Prin urmare, a reușit să extragă cele mai relevante date despre toate produsele disponibile în oferta magazinului online unde lucrează. Noi îl vom ajuta să se familiarizeze mai mult cu aceste date.

### **Sarcina:**

În cadrul activităților sale zilnice, Alex a reușit să colecteze următorul set de date despre produsele magazinului online:

[online\\_store\\_data.csv](#)

Familiarizați-vă cu datele cu care veți lucra:

## Tutorial video: Familiarizarea cu datele magazinului online

Scrieți un script Python care să răspundă la următoarele întrebări folosind **biblioteca Pandas**:

- **Câte produse există în setul de date?**
  - Afișați numărul total de rânduri din DataFrame.
- **Care este cel mai bine vândut produs din întregul magazin online?**
  - Sortați DataFrame-ul după numărul de unități vândute și afișați primul rând.
- **Care sunt cele mai bine vândute 5 telefoane mobile?**
  - Filtrați produsele din categoria "**Smartphones**", apoi sortați-le după numărul de unități vândute și selectați primele 5.
- **Care este prețul celui mai scump și celui mai ieftin laptop**
  - Filtrați produsele din categoria "**Laptops**", eliminați rândurile cu valori lipsă și găsiți prețul maxim și minim.

## Ghid pentru rezolvarea sarcinii:

Mai jos sunt pașii care vă pot ajuta în realizarea sarcinilor:

### 1. Determinarea numărului total de produse din setul de date al magazinului online

- Pentru a afla numărul total de produse din setul de date, trebuie să determinați numărul de rânduri. În acest scop, se utilizează o proprietate a obiectelor DataFrame, discutată în prima lecție. Aceasta returnează numărul total de rânduri și coloane, iar sarcina voastră este să izolați doar numărul de rânduri și să afișați această valoare în terminal.

## 2. Identificarea celui mai vândut produs din întregul magazin online

- Cel mai vândut produs poate fi găsit prin sortarea setului de date după coloana care conține numărul total de vânzări. După sortarea în ordine descrescătoare, trebuie să extrageți primul rând și să îl afișați în terminal. Acesta va fi cel mai vândut produs.

## 3. Identificarea celor mai vândute cinci telefoane mobile

- Pentru a găsi cele mai vândute 5 telefoane mobile, este necesar să aplicați filtrarea astfel încât să obțineți doar datele care au categoria *Smartphones*. Apoi, trebuie să sortați setul de date obținut în ordine descrescătoare și să extrageți primele 5 rânduri. Acestea vor reprezenta cele 5 cele mai vândute telefoane mobile.

## 4. Găsirea prețului celui mai ieftin și celui mai scump laptop

- Pentru a găsi prețurile celui mai ieftin și celui mai scump laptop, mai întâi este necesar să filtrați datele astfel încât să rămână doar cele care aparțin categoriei laptopurilor. Apoi, setul de date obținut poate fi sortat în ordine descrescătoare. Urmează să preluați prima și ultima valoare, iar acestea vor fi laptopul cel mai scump și cel mai ieftin, respectiv. La final, este necesar să izolați prețurile acestora. Dacă există valori lipsă în setul de date care îngreunează obținerea celui mai ieftin laptop, extindeți filtrarea pentru a include doar laptopurile ale căror prețuri sunt mai mari decât zero.

## Predarea sarcinii:

După finalizarea lucrării, codul sursă al proiectului complet trebuie transmis instructorului cursului.

- Numele fișierului: `task1_online_store_analysis.py`
- Salvați fișierul în format `.zip`, `.rar` sau `.7z` și trimiteți-l pe platformă.
- Fișierul trebuie să conțină comentarii și să fie formatat lizibil.

## Cerințe pentru evaluare

Instructorul va evalua sacina după următoarele criterii:

| Criteriu   | Procent din notă |
|--|------------------|
| Setul de date încărcat corect și afișarea numărului de produse                         | 20%              |
| Cel mai bine vândut produs găsit corect  | 20%              |
| Cele mai bine vândute 5 telefoane mobile filtrate și afișate corect                    | 20%              |
| Laptop-uri filtrate corect și prețuri afișate ale celui mai ieftin și mai scump laptop | 20%              |
| Claritate, structură și comentarii în cod  | 20%              |

## Mod de evaluare

□□□□ (90-100%): Sarcina este completă, toate calculele și intrările de date sunt corecte, rezultatele sunt afișate clar, iar codul este lizibil și scris eficient.

□□□ (75-89%): Erori minore la introducerea datelor sau la calcule, dar majoritatea funcționalității funcționează corect. Prezentarea rezultatelor este clară, cu câteva erori minore.

□□ (60-74%): Mai multe erori în introducerea datelor, calcule sau afișarea rezultatelor. Codul conține erori minore de logică sau de sintaxă.

□□ (50-59%): Erori majore în introducerea datelor și calcule. Programul funcționează doar parțial, iar afișarea rezultatelor nu este corectă.

□ (45-49%): Sarcina este parțial finalizată, dar există erori semnificative în introducerea datelor și în calcule, care afectează acuratețea rezultatelor.

Programul are o funcționalitate limitată și necesită îmbunătățiri pentru a afișa corect informațiile solicitate.

□ **(0-44%)**: Programul nu este funcțional, conține erori grave care interferează cu execuția sa sau sarcina nu a fost înțeleasă corect.

**Succes cu sarcina!**