

Iceland
Liechtenstein
Norway grants



GHID DE **PRELUCRARE** A DATELOR

AUTORI: AVRAM CALIN, GLIGOR ADRIAN, NYULAS VICTORIA și DUMITRU ROMAN

Acest ghid a fost realizat în cadrul proiectului intitulat: **“Innovative methods needed by teaching and research staff for AI data analysis and processing”**, care s-a derulat în perioada februarie 2022-februarie 2023, având ca parteneri Universitatea de Medicină, Farmacie, Științe și Tehnologie “G.E.Palade” Targu Mures și SINTEF – Norvegia.

Disclaimer:

Material made with the financial support of the EEA Financial Mechanism 2014 - 2021. Its content (text, photos, video) does not reflect the official opinion of the Program Operator, the National Contact Point or the Office of the Financial Mechanism. The information and opinions expressed are the sole responsibility of the author(s).

CUPRINS

INTRODUCERE	4
1. DESCRIEREA ELEMENTELOR DE DATE. TIPURI DE DATE	5
2. PREZENTARE GENERALĂ A ANALIZEI DATELOR	7
3. PROCESUL DE ANALIZĂ A DATELOR.....	9
4. ANALIZA DATELOR CU EXCEL – PREZENTARE GENERALĂ.....	11
5. IMPORTUL DATELOR ÎN EXCEL	13
5.1 Importul datelor dintr-o bază de date	13
5.2 Importul datelor dintr-o foaie de lucru prin copiere și lipire.....	18
5.3 Importul datelor de pe site-uri web	21
6. CONSOLIDAREA DATELOR	26
6.1 Pregătirea datelor pentru consolidare	26
6.2 Consolidarea datelor în același registru de lucru.....	26
6.3 Consolidarea datelor din diferite registre de lucru.....	32
7. TABELE IN EXCEL.....	34
7.1 Distincția dintre tabel și interval (arie).....	35
7.2 Crearea unui tabel în Excel.....	35
7.3 Gestionarea numelor într-un tabel.....	38
7.4 Propagarea unei formule într-un tabel.....	41
7.5 Redimensionarea tabelului	42
7.6 Eliminarea datelor duplicate în Excel	44
7.7 Convertirea tabelului în interval (arie).....	46
8. CURĂȚAREA DATELOR CU FUNCȚII DE TEXT	48
8.1 Eliminarea caracterelor nedorite din text	48
8.2 Identificarea modelelor de text necesare cu funcțiile de text.....	50
8.3 Extragerea valorilor datelor din text.....	50
8.3.1 Extragerea valorilor datelor cu Text to Columns Wizard	51
8.3.2 Extragerea valorilor datelor cu funcții text	53
8.3.3 Extragerea valorilor datelor cu Flash Fill.....	54
8.4 Formatarea datelor cu funcții de text.....	56
9. CURĂȚAREA DATELOR CARE CONȚIN VALORI DATE.....	58
9.1 Formate de date	58

9.2 Conversia datelor în format serial în format Lună-Zi-An	59
9.3 Conversia datelor în format lună-zi-an în format serial.....	59
9.4 Calcularea datei curente	61
9.5 Identificarea unei zile lucrătoare după zile specificate.....	61
9.6 Extragerea Anului, Lună, Zi din Data	62
9.7 Extragerea zilei săptămânii din date.....	62
9.8 Obținerea datei din an, lună și zi.....	63
9.9 Calcularea anilor, lunilor și zilelor dintre două date	63
10. LUCRUL CU VALORI DE TIMP	64
10.1 Ora în format serial.....	64
10.2 Ora în format oră-minut-secundă	64
10.3 Conversia timpilor în format serial în format oră-minut-secundă	65
10.4 Conversia timpilor în format oră-minut-secundă în format serial	65
10.5 Obținerea timpului din oră, minut și secundă.....	67
10.6 Extragerea orei, minutelor și secundelor din timp	67
11. SORTAREA DATELOR.....	69
11.1 Sortarea după numere	69
11.2 Sortarea după text.....	71
11.3 Sortare personalizată	72
12. FILTRAREA DATELOR	74
12.1 Filtrarea după valorile selectate.....	75
12.2 Filtrarea după text	75
12.3 Filtrarea după dată	77
12.4 Ștergerea filtrului.....	79

INTRODUCERE

Pentru a realiza o cercetare validă și valabilă, un element critic esențial este reprezentat de colectarea și prelucrarea exactă și eficientă a datelor pentru analiză. Majoritatea cercetătorilor din domeniul medical dețin un nivel redus de pregătire referitor la modul de gestionare și prelucrare al datelor medicale colectate, aspect care poate duce la o creștere a timpului utilizat pentru curățarea datelor, precum și la riscul apariției erorilor de colectare sau înregistrare.

Studentii, doctoranzi, tineri cercetători, toți sunt sau pot fi implicați în colectarea datelor, planificarea, execuția, și analiza cercetării. Frecvent, datorită lipsei de finanțare, a experienței sau a ambelor situații, aceste date sunt introduse în computer folosind un proces ad-hoc care are ca urmare date codificate în mod eronat, valori incorect formate, informații incomplete și erori tipografice. Deseori, este o situație comună ca cercetătorii să afle că datele colectate trebuie „curățate” pe scară largă înainte de a putea fi analizate corect.

În consecință am realizat acest ghid care pune la dispoziție îndrumări practice, concise și ilustrative cu privire la modul de prelucrare a datelor pentru a crea un set de date bine conceput, omogen și verificat. Deoarece Microsoft Excel este cel mai folosit soft de către studenți, cercetători pentru colectarea și organizarea datelor, ghidul prezintă metode specifice care pot fi implementate în programul de calcul tabelar Microsoft Excel.

Prezentul ghid reprezintă un tutorial cuprinzător care oferă o perspectivă bună asupra celor mai recente și avansate funcții disponibile în Microsoft Excel. În cadrul ghidului am inclus o mulțime de capturi de ecran care explică modul de utilizare a unei anumite caracteristici, pas cu pas.

1. DESCRIEREA ELEMENTELOR DE DATE. TIPURI DE DATE

Colectarea exactă a datelor începe cu planificarea. Înainte de a colecta oricărui date, un investigator ar trebui să definească întrebările de cercetare și să determine ce măsurători sunt necesare pentru a le răspunde. De obicei, un set de date de cercetare include cel puțin o variabilă de rezultat (variabilă dependentă) și una sau mai multe variabile predictoare (independente). Alte măsuri demografice, covariabile sau de verificare pot fi, de asemenea, înregistrate. Este important ca fiecare subiect investigat într-un set de date să aibă atribuit un număr de identificare unic. Pentru fiecare studiu, cercetătorul ar trebui să folosească aceste informații pentru a verifica dacă toate variabilele necesare pentru a efectua o analiză sunt incluse. Este descurajant să realizezi prea târziu că o variabilă necesară pentru a finaliza o analiză nu a fost colectată.

Definirea Big Data

În literatură specifică domeniului, Big Data sunt caracterizate ca date care includ:

- Volum mare - atât în ceea ce privește elementele de date, cât și dimensionalitatea.
- Viteză mare - ajunge la o rată foarte mare, de obicei cu o ipoteză de latență scăzută între sosirea datelor și valoarea obținută.
- Varietate ridicată - cuprinde capacitatea de a evolua în timp pentru forma și sensul datelor.

Și care necesită:

- Procesare eficientă din punct de vedere al costurilor
- Tipuri inovatoare de analiză
- Valoare nouă de afaceri

Tabel 1. Tipuri de date – categorizarea datelor

Date	Descriere	Exemple
Date nominale sau categorice	Date care pot fi plasate în categorii care se exclud	Numele unei țări, Genul, Rezidența, Prezența unei boli, Status marital
Date ordinale	Datele sunt ordonate sau clasificate în funcție de anumite caracteristici	Scala Likert, Nivel educațional, Grupa sanguină, Nivelul de satisfacție al pacientului, Statusul socio-economic, Stadiul bolii
Date de tip interval	Date caracterizate și ordonate după o anumită distanță între fiecare observație, dar fără zero	Testare IQ, Temperatura în grade Fahrenheit, Celsius,
Date de tip raport	Date care au un zero natural	Vârsta, Înălțimea, Greutatea, Valoarea glicemiei bazale, Alura ventriculară

2. PREZENTARE GENERALĂ A ANALIZEI DATELOR

Analiza datelor reprezintă un proces de inspectare, curățare, transformare și modelare a datelor cu obiectivul de a descoperi informații necesare, de a sugera concluzii și de a sprijini luarea deciziilor.

Statisticianul John Tukey în 1961 definește procesul de analiză a datelor în felul următor „Proceduri de analiză a datelor, tehnici de interpretare a rezultatelor unor astfel de proceduri, modalități de planificare a culegerii datelor pentru a le face analiza mai ușoară, mai precisă sau mai exactă, precum și toate mașinile. și rezultatele statisticilor (matematice) care se aplică analizării datelor.” În consecință, analiza datelor este un proces pentru obținerea de date mari, nestructurate din diverse surse și transformarea acestora în informații utile pentru:

- Răspunderea la întrebări de cercetare
- Testarea ipotezelor
- Luarea deciziilor
- Infirmarea teoriilor

Tipuri de analiză a datelor

Există mai multe tehnici de analiză a datelor care cuprind diverse domenii, cum ar fi afaceri, știință, științe sociale etc., cu o varietate de nume. Principalele abordări de analiză și exploatare a datelor sunt:

- Data Mining
- Business Intelligence
- Analiză statistică a datelor
- Analiza predictivă a datelor
- Analiza textului

Data Mining este analiza unor cantități mari de date pentru a extrage modele de date interesante, necunoscute anterior, date neobișnuite și dependențe. Rețineți că scopul este extragerea de modele și cunoștințe din cantități mari de date și nu extragerea datelor în sine.

Analiza data mining implică metode informatice la intersecția dintre inteligența artificială, învățarea automată, statistica și sistemele de baze de date.

Business Intelligence

Tehnicile și instrumentele de Business Intelligence se folosesc pentru achiziția și transformarea unor cantități mari de date de afaceri nestructurate cu scopul de a folosi la identificarea, dezvoltarea și crearea de noi oportunități strategice de afaceri. Țelul Business Intelligence este de a permite interpretarea ușoară a unor volume mari de date pentru a identifica noi oportunități.

Analiza statistică a datelor

Statistica se ocupă cu studiul colectării, analizei, interpretării, prezentării și organizării datelor. În analiza statistică a datelor, se folosesc două metodologii statistice principale:

1. Statistică descriptivă: datele din întreaga populație sau un eșantion sunt rezumate cu indicatori descriptivi numerici.
2. Statistică inferențială: utilizează modele din eșantionul de date pentru a deduce inferențe despre populația reprezentată. Aceste inferențe pot fi: testarea ipotezelor, analize de estimare, corelație și regresie.

Analiza predictivă a datelor

Analiza predictivă a datelor apelează la modele statistice pentru a analiza datele actuale și anterioare pentru prognoză (predicții) despre evenimente viitoare sau necunoscute, putând astfel să ajute la luarea deciziilor.

Analiza textului

Analiza textului este procesul de obținere a informațiilor de înaltă calitate din text. Exploatarea textului implică de obicei procesul de structurare a textului de intrare, de derivare a modelelor în cadrul datelor structurate folosind mijloace precum învățarea modelelor statistice și, în final, evaluarea și interpretarea rezultatelor.

3. PROCESUL DE ANALIZĂ A DATELOR

Procesul de analiză a datelor constă din următoarele etape care sunt de natură iterativă:

- Specificații privind cerințele de date
- Colectarea datelor
- Procesarea datelor
- Curățarea datelor
- Analiza datelor
- Comunicarea datelor

Specificații privind cerințele de date

Datele necesare analizei se bazează pe o întrebare de cercetare sau un experiment. Pe baza cerințelor care direcționează analiza, sunt identificate datele necesare ca intrări în analiză (de exemplu, Populația pacienților cu hipertensiune arterială). Caracteristicile specifice referitoare la o populație (de exemplu, vârsta, sexul și venitul) pot fi specificate și obținute. Datele pot fi *numerice* sau *categorice*.

Colectarea datelor

Colectarea datelor reprezintă procesul de culegere a informațiilor despre caracteristicile țintă marcate ca cerințe de date. Accentul se pune pe asigurarea unei colectări exacte și corecte a datelor. Colectarea datelor se asigură că datele culese sunt exacte, astfel încât deciziile aferente sunt valide.

Procesarea datelor

Sursa datelor colectate este diversă, de la baze de date organizaționale la informațiile din paginile web. Datele obținute în acest mod, pot să nu prezinte o structură coerentă și pot conține informații irelevante. Prin urmare, datele colectate trebuie să fie supuse procesării și curățării datelor.

Curățarea datelor

Datele procesate și organizate pot fi incomplete, pot conține valori duplicate sau pot conține erori. Curățarea datelor este procesul de prevenire și corectare a acestor erori. În funcție de tipul datelor, putem apela la mai multe tipuri de curățare a datelor.

Analiza datelor

Datele care sunt procesate, organizate și curățate sunt pregătite și potrivite pentru următoarea etapă, aceea de analiză. Pe baza cerințelor, putem apela la diferite metode de analiză a datelor pentru a înțelege, interpreta și a trage concluzii.

Comunicarea datelor

Rezultatele analizei datelor trebuie raportate într-un format cerut de utilizatori pentru a-și susține deciziile și demersurile ulterioare. Feedbackul de la utilizatori ar putea duce la analize suplimentare. Pentru comunicarea datelor se poate apela la tehnici de vizualizare și ilustrare a datelor, cum ar fi tabele și diagrame, care contribuie la comunicarea mesajului clar, concis și eficient către utilizatori. Instrumentele de analiză pun la dispoziție modalități de a evidenția informațiile necesare cu coduri de culoare și formatare în tabele și diagrame.

4. ANALIZA DATELOR CU EXCEL – PREZENTARE GENERALĂ

Analiza datelor cu Excel

Microsoft Excel pune la dispoziție diferite și numeroase modalități de analiză și interpretare a datelor. Proveniența datelor poate fi din surse diverse, iar modurile de convertire și formatare a datelor sunt de asemenea diverse. Pentru analizarea datelor, Excel oferă următoarele tehnici, comenzi funcții și instrumente: formatare condiționată, intervale, tabele, funcții de text, funcții de dată, funcții de timp, funcții financiare, subtotaluri, analiză rapidă, formule auditare, instrumente de investigare, analiză IF, rezolvitor, modele de date, PowerPivot, PowerView, PowerMap etc.

După cum am menționat și enumerat, Excel conține și oferă comenzi, funcții și instrumente care ușurează sarcinile de analiză a datelor.

Intervale și tabele

Datele pot fi prezentate într-un interval sau într-un tabel. Anumite comenzi asupra datelor pot fi realizate indiferent de modul de prezentare a datelor, interval sau tabel. Însă, anumite comenzi sunt mai eficiente dacă datele sunt prezentate sub formă de tabel, decât în interval, de asemenea Excel oferă funcții exclusiv pentru tabele.

Curățarea datelor – Funcții text, dată și oră

Înainte de procesul de analiză a datelor, trebuie realizată etapa de curățare și structurare a datelor, mai jos sunt prezentate câte funcții prin care se pot structura datele:

- Cu funcții de text
- Cu funcții pentru valori date
- Cu funcții pentru valori de timp

Formatarea condițională

Pentru a colora anumite celule, fontul textului sau a insera pe baza unor criterii predefinite, diferite simboluri lângă valorile conținute de celule, Microsoft Excel pune la dispoziție comenzi pentru formatarea condițională.

Sortare și filtrare

În etapa de pregătire a analizei datelor, sau pentru a expune diverse anumite date, se poate apela la funcțiile de Filtrare și Sortare a datelor puse la dispoziție de Excel.

Subtotaluri cu intervale și PivotTable

Pentru a rezuma datele în Excel, se poate apela la instrumentul PivotTable, însă pentru gruparea/degruparea datelor sau rezumarea lor există un alt instrument pus la dispoziție de Excel, se poate folosi Subtotaluri cu intervale.

Vizualizarea datelor

Excel pune la dispoziție numeroase tehnici de vizualizare a datelor (diagrame tip bandă, gantt, diagramă în cascadă, boxplot, scatterplot etc.), toate le putem folosi de la secțiunea diagrame Excel din meniul Insert.

Validarea datelor

Se poate folosi validarea de date pentru a restricționa tipul de date sau valorile pe care utilizatorii le introduc în celule. Cu ajutorul comenzilor de validare a datelor, se pot seta cu ușurință valori de validare a datelor pentru o celulă, un mesaj de introducere care solicită utilizatorului ce se așteaptă să fie introdus în celulă, validarea valorile introduse cu criteriile definite și afișarea un mesaj de eroare în caz de intrări incorecte.

Lucrul cu mai multe foi de lucru

În cazul în care este necesară efectuarea unor calcule identice în mai multe foi de lucru, cu ajutorul funcțiilor din excel se poate face acest lucru în câțiva pași simpli.

5. IMPORTUL DATELOR ÎN EXCEL

Un număr ridicat de utilizatori folosesc o conexiune pentru a importa date externe în Excel ca un instantaneu actualizabil. Avantajul este reprezentat de faptul că se creează un document independent care poate fi utilizat pentru a lucra offline, dar actualizat cu date noi atunci când este online. Deoarece datele sunt conținute în Excel, utilizatorii le pot transforma și pentru a raționa propriul context sau necesitățile de analiză.

Câteva provocări cheie trebuie luate în considerare atunci când se realizează importul Big Data în Excel:

- *Interogarea Big Data* — Sursele de date concepute pentru Big Data, cum ar fi SaaS, HDFS și sursele relaționale mari, pot necesita uneori instrumente specializate. În Excel putem apela la instrumentul Power Query, care pune la dispoziție mai multe seturi moderne de conectori pentru utilizatorii Excel, inclusiv conectori pentru relațional, HDFS, SaaS (Dynamics CRM, Salesforce), etc.
- *Transformarea datelor*— similar cu alte tipuri de date, Big Data, nu prea des sunt absolut curate. Power Query oferă capacitatea de a crea un set ordonat și reproductibil de pași pentru prelucrarea datelor.

În Excel, se pot importa date din diverse surse, cum ar fi: baze de date Microsoft Access, pagini web, fișiere text, tabele SQL Server, cub de analiză SQL Server, fișiere XML etc.

Importarea datelor externe

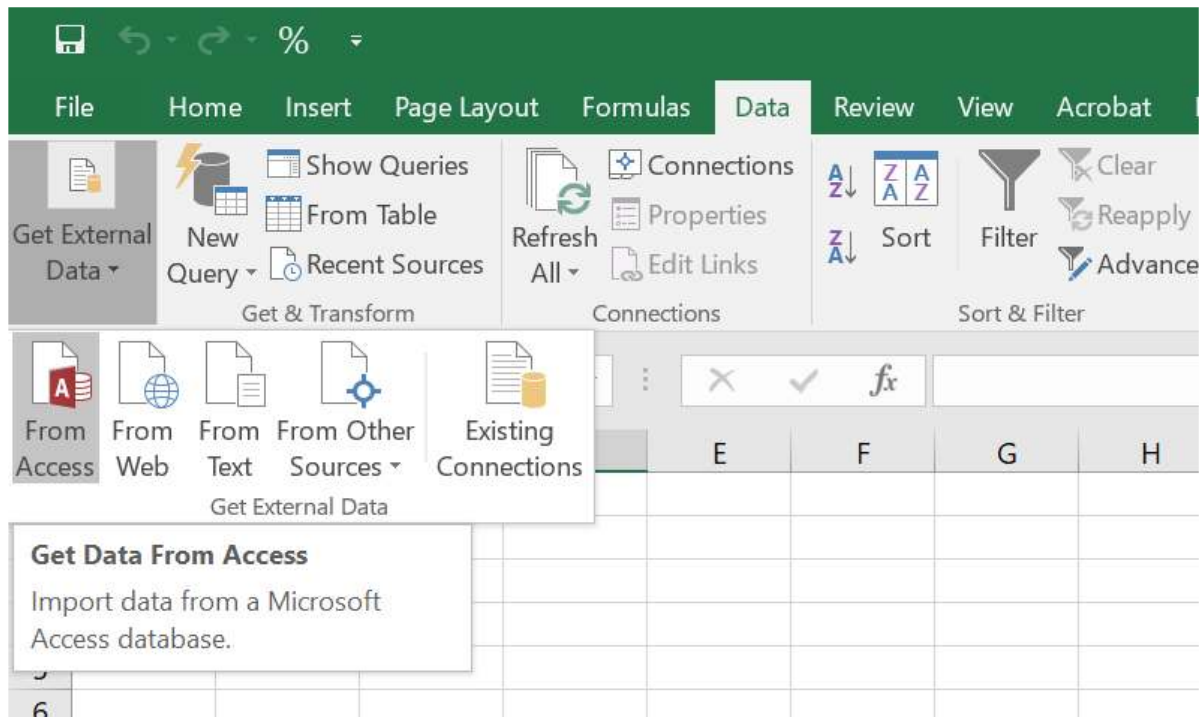
Datele aflate în fișiere externe compatibile pot fi importate în Excel fără a fi nevoie de reinserarea totală a datelor. În funcție de formatul datelor care se doresc a se importa, pot fi utilizate diferite tehnici, inclusiv deschiderea și salvarea în Excel, conectarea la baze de date, importarea datelor, copierea și lipirea lor în Excel.

5.1 Importul datelor dintr-o bază de date

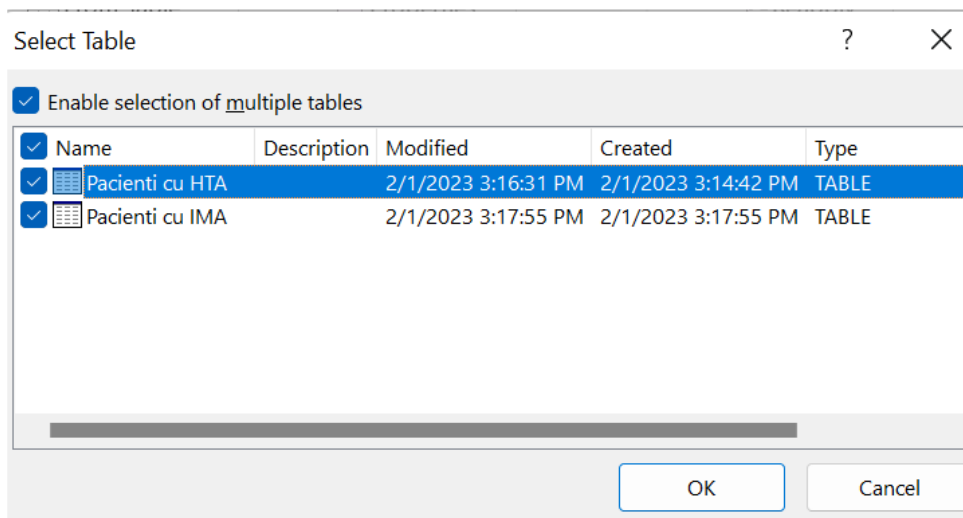
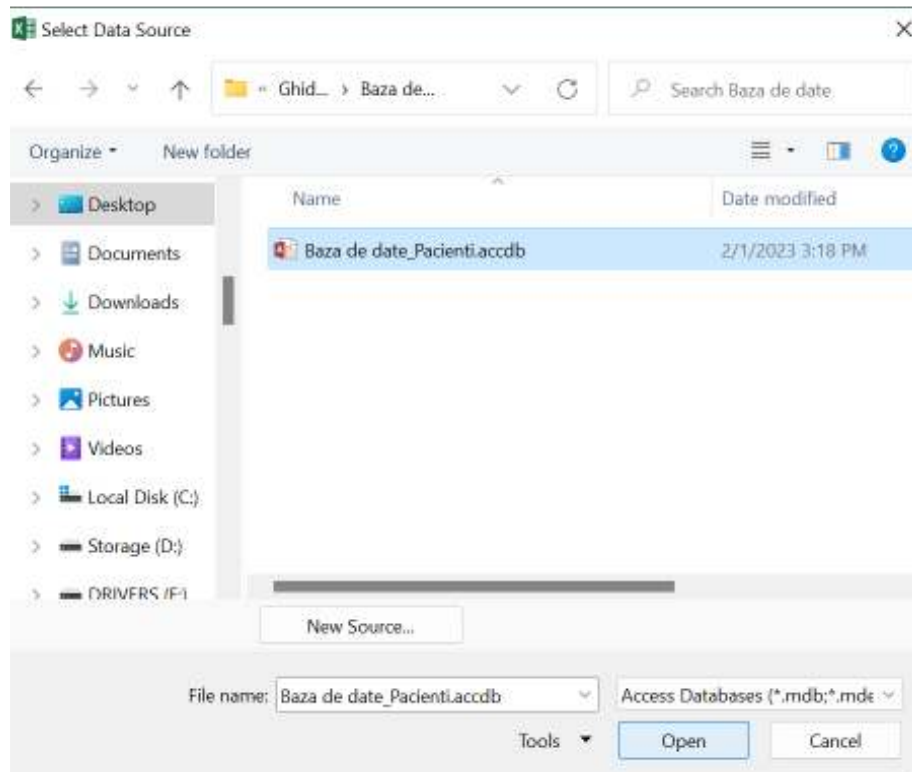
Vom exemplifica această acțiune într-o foaie de calcul fără conținut. Scopul acestei acțiuni este de a importa date în Microsoft Excel dintr-o bază de date Microsoft Access.

Într-un folder de pe calculator avem salvat un fișier care conține o bază de date Access, denumită **Bază de date_Pacienți**.

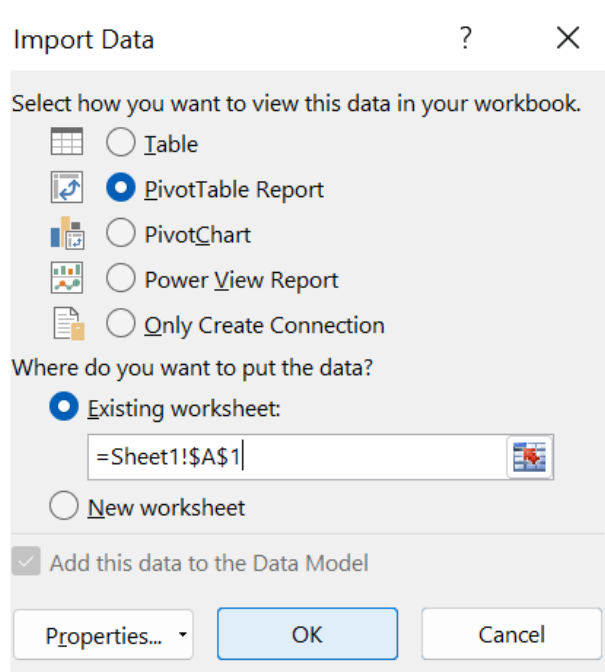
1. În Excel 2016, deschidem o foaie de lucru fără conținut.
2. Facem clic pe **DATA > Get External Data > From Access**.



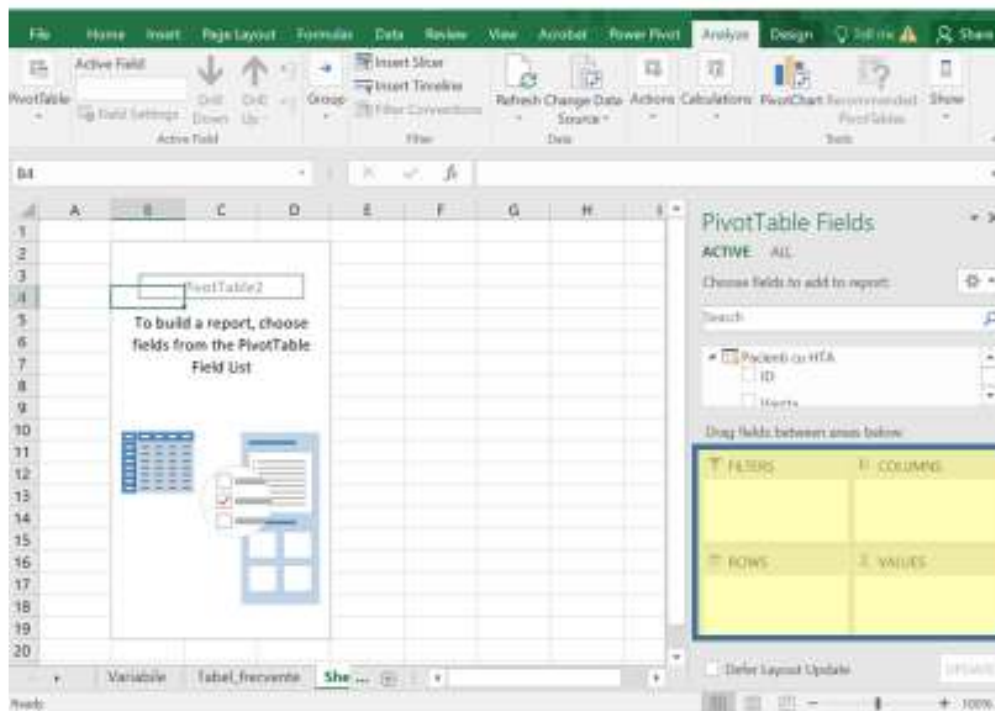
3. Selectăm fișierul **Baza de date_Pacienti.accdb** din folder și facem clic pe **Open**. Se afișează o altă fereastră care arată tabelele găsite în baza de date Access. Tabelele dintr-o bază de date sunt similare cu foile de lucru din Excel. În continuare bifăm caseta de selectare **Enable selection of multiple tables** și selectăm toate tabelele. Apoi clic pe **OK**.



4. Apare fereastra **Import Data**. Selectăm căsuța **PivotTable Report**, care importă tabelele în Excel și pregătește un PivotTable pentru analiza tabelor importate, și bifăm căsuța cu **Existing worksheet**, apoi facem clic pe **OK**.



5. După ce datele sunt importate, este creat un PivotTable folosind tabelele importate. Explorarea datelor importate este ușoară dacă se folosește un PivotTable. Într-un PivotTable, glisați câmpuri (similare cu coloanele din Excel) din tabele (cum ar fi tabelele tocmai importate din baza de date Access) în diferite **zone** ale raportului PivotTable pentru a regla modul în care prezintă datele. Un PivotTable are patru zone: **FILTERS**, **COLUMNS**, **ROWS** și **VALUES**.



Poate o să fie necesare mai multe încercări pentru a identifica în ce zonă ar trebui selectat un câmp. În funcție de dorințele utilizatorului se pot selecta cât de multe sau cât de puține câmpuri din tabele, până când PivotTable afișează datele conform dorințelor. Un exemplu de raport PivotTable este reprezentat în figura de mai jos.

The screenshot shows a PivotTable report titled "Sum of Educație". The report is organized with "Stare civilă" and "Mediu urban" as filters, "Educație" as rows, and "Sum of Educație" as values. The report shows the sum of education levels for different categories of people (Stare civilă and Mediu urban).

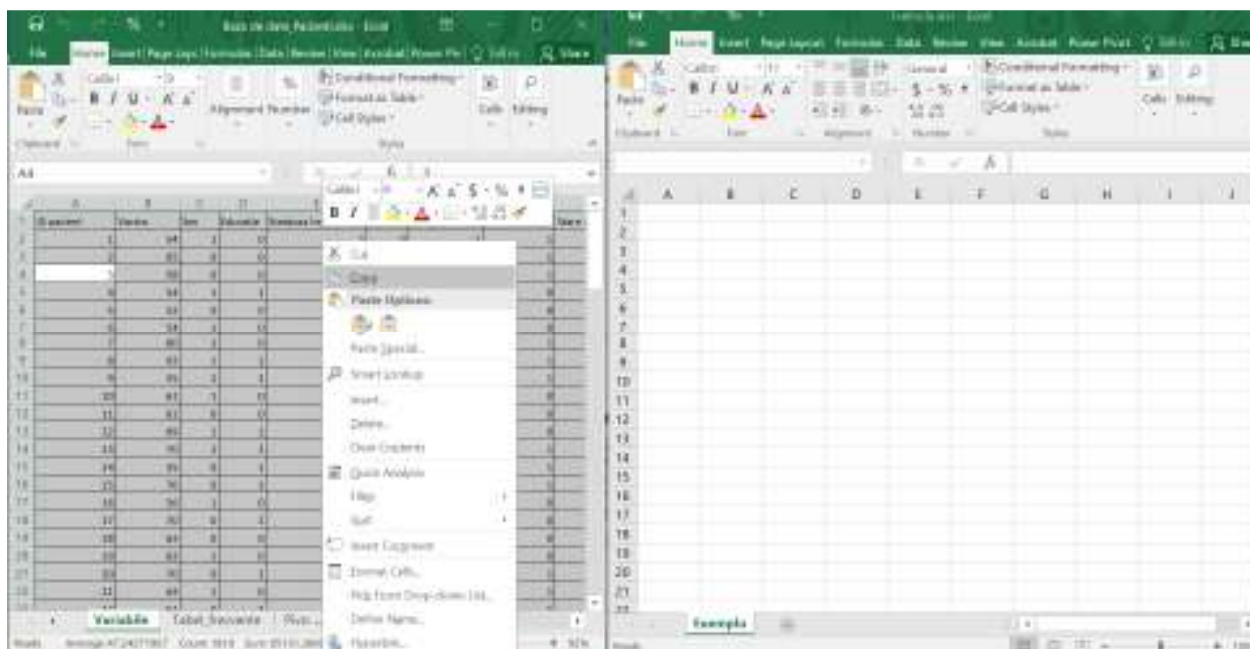
Column Labels	(blank)	0	1	
Row Labels	Count of Varsta	Sum of Educație	Count of Varsta	Sum of Educație
0		0		
1		7		
Grand Total		7		

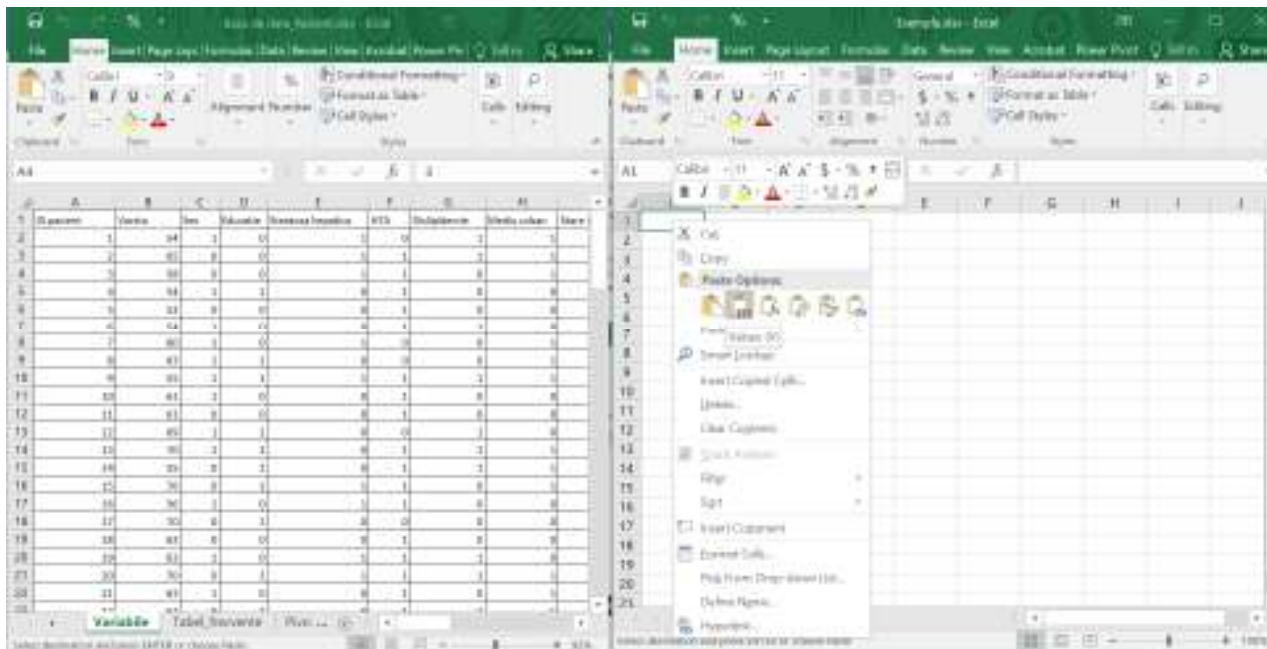
5.2 Importul datelor dintr-o foaie de lucru prin copiere și lipire

Acum, să importăm date din altă sursă, de data aceasta dintr-un registru de lucru existent, apoi să specificăm relațiile dintre datele noastre existente și noile date. Relațiile vă permit analizarea colecțiilor de date în Excel și crearea de vizualizări interesante și captivante din datele importate.

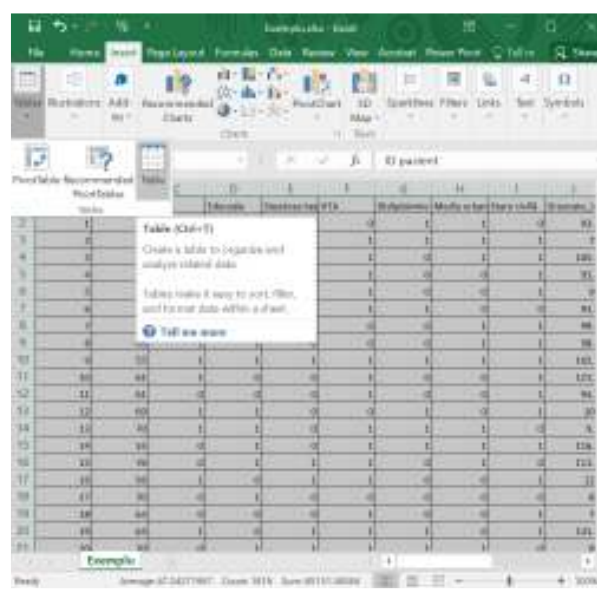
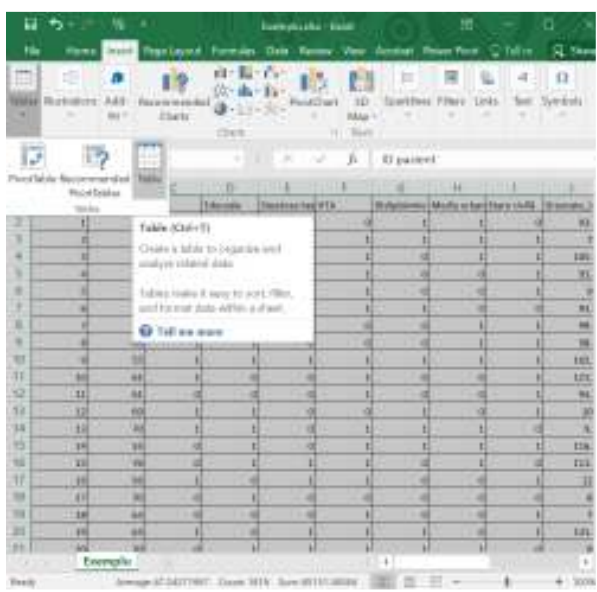
Să începem prin crearea unei foi de lucru necompletate, apoi importăm date dintr-un registru de lucru Excel. Pentru a realiza importul datelor dintr-o foaie de lucru Excel existent în calculator, se poate ține cont de pașii următori:

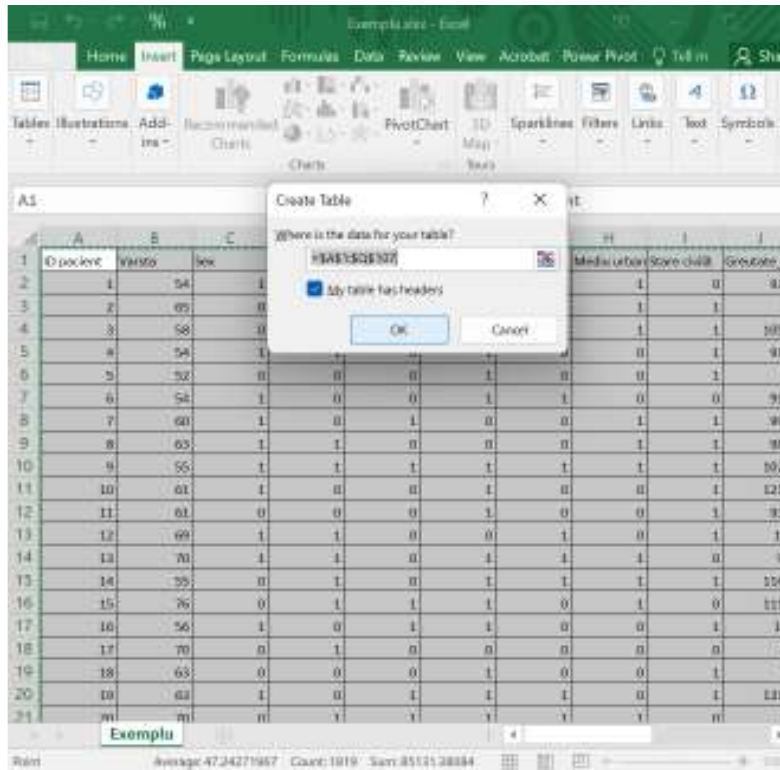
1. Inserăm un nou registru de lucru Excel și îl denumim **Exemplu**.
2. Deschidem fișierul Excel numit **Baza de date_Pacienti.xlsx** din folderul de pe calculator.
3. Selectăm și copiem datele din foaia de calcul **Variabile**. Pentru selectarea întregului conținut din registru de lucru folosim combinația de taste CTRL+A.
4. Pe foaia de lucru **Exemplu**, poziționăm cursorul în **celula A1** și lipim datele (Clic dreapta-Paste sau CTRL+V) selectate din foaia de calcul anterioară.



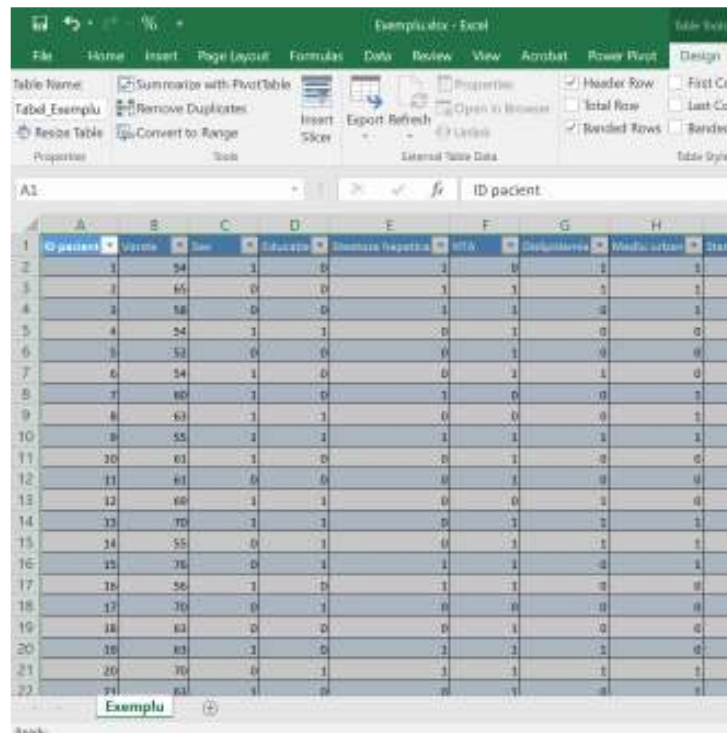


5. În continuare, cu datele selectate și evidențiate în registrul denumit **Exemplu**, folosim combinația de taste CTRL+T pentru a formata datele inserate sub formă de tabel. De asemenea, putem formata datele sub formă de tabel din panglică, prin selectarea **INSERT – TABLES – CREATE A TABLE**. Deoarece fiecare coloană deține antet bifăm **My table has headers** în fereastra **Create a table** care se afișează ulterior și la final clic **OK**.





6. Pentru denumirea tabelului, în meniul **TABLE TOOLS - DESIGN - PROPERTIES**, găsim câmpul **TABLE NAME** și tastă **Exemplu**. Iar la final salvăm fișierul Excel.



5.3 Importul datelor de pe site-uri web

Datele și informațiile site-urilor web și din alte surse pot fi importate în Excel dacă sunt scrise într-un format adecvat pe care îl recunoaște și Microsoft Excel.

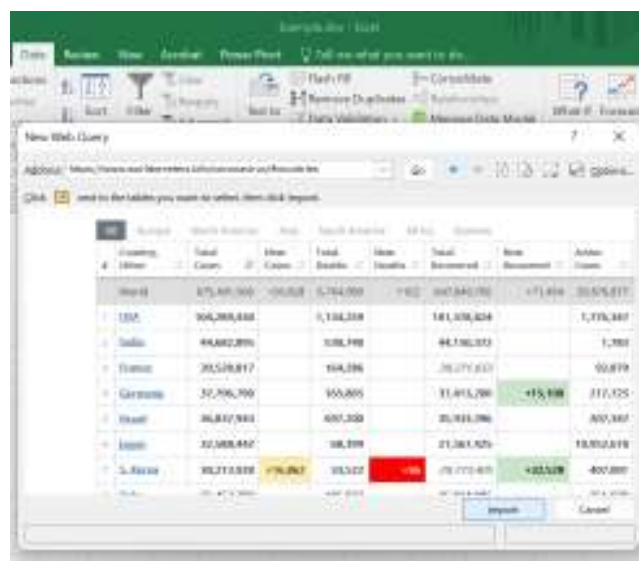
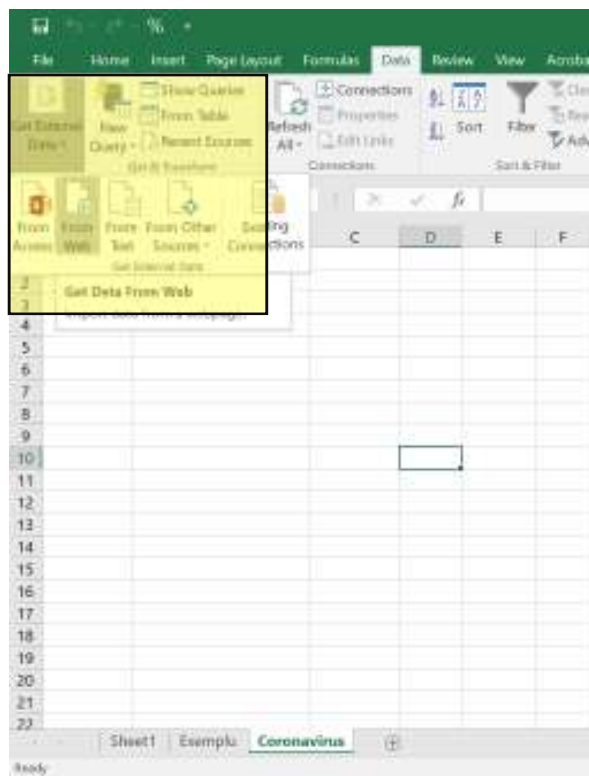
Pentru a importa informații din site-uri web vom urmări pașii explicați mai jos:

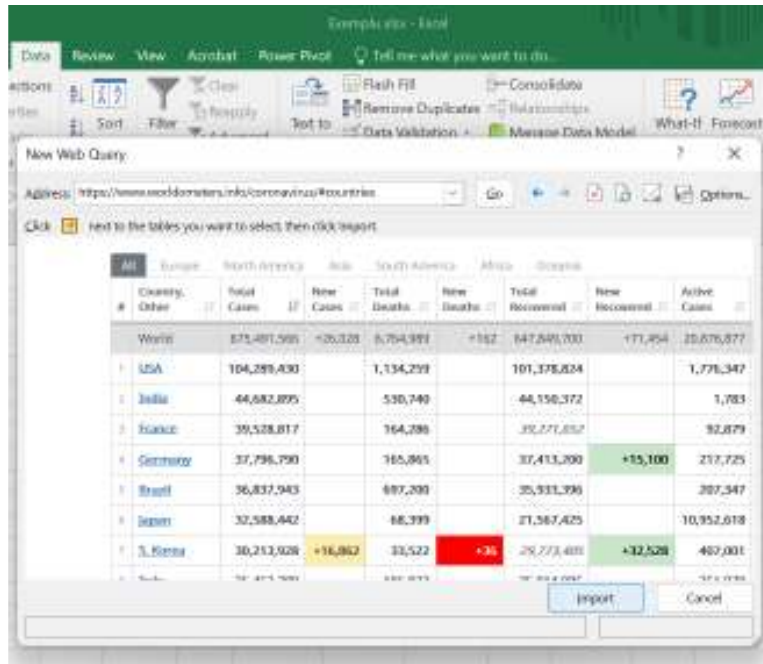
1. Întrăm pe site-ul www.worldometers.info, care conține informații și date referitoare la numărul de cazuri COVID-19 la nivel mondial, copiem adresa URL a paginii web: <https://www.worldometers.info/coronavirus/#countries>.

2. Deschidem un registru Excel pe care îl denumim **Coronavirus**.

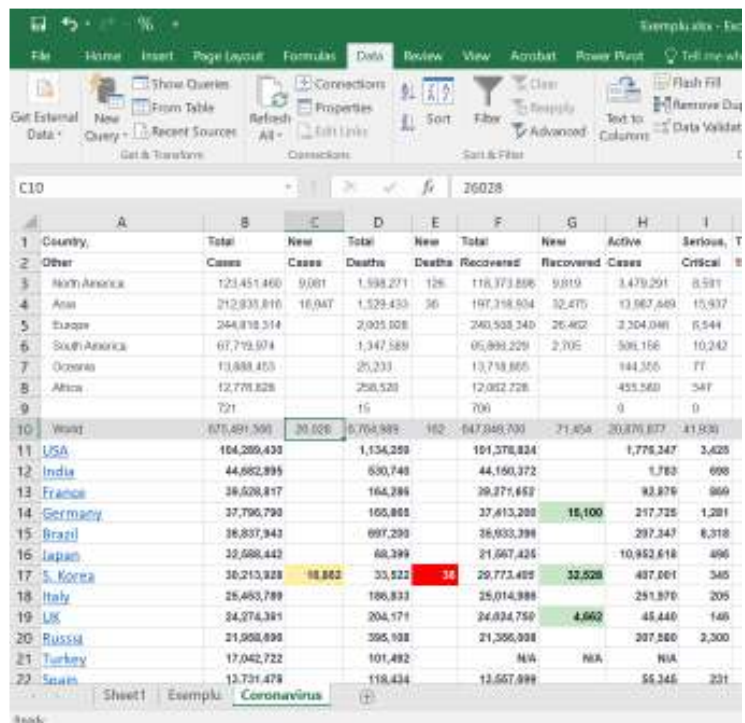
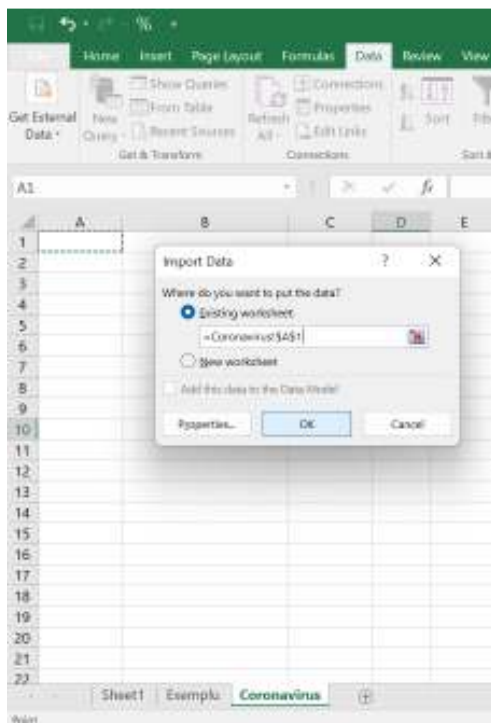
4. Din meniul **Data**, selectăm **Get External Data – From Web**. După acești pași se deschide caseta de dialog pentru a introduce adresa URL a paginii web cu datele pe care dorim să le importăm.

5. Lipim adresa URL în caseta de dialog **From Web** și facem clic pe **OK**. Panoul de navigare se va deschide cu o listă de date care pot fi importate în Excel.



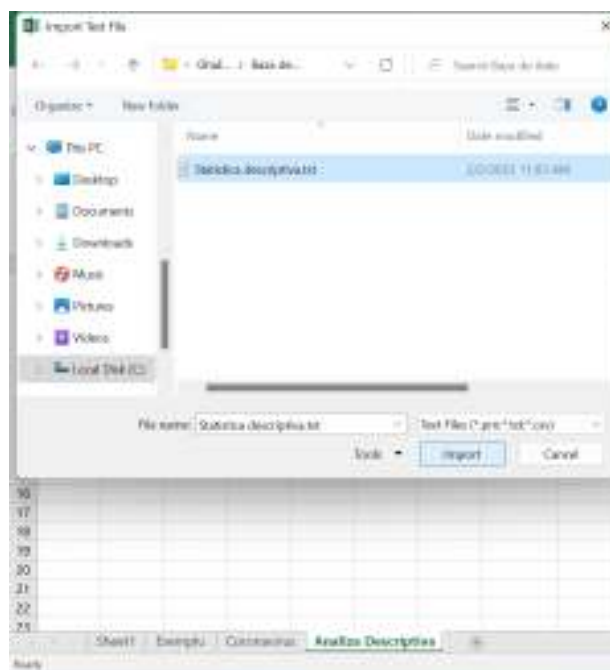
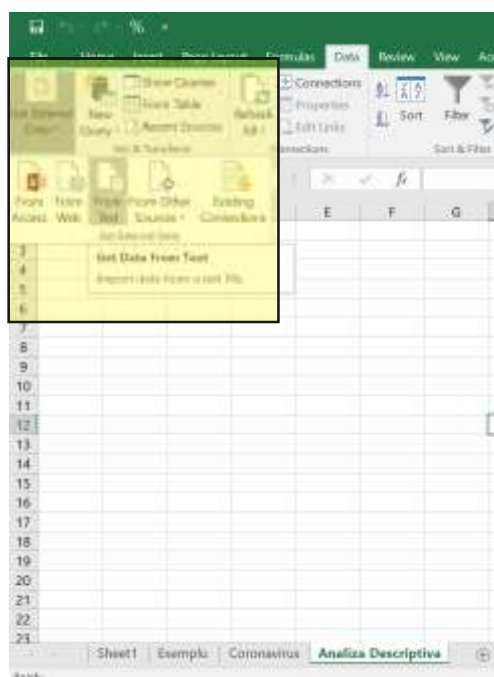


6. În fereastra următoare selectăm foaia de calcul în care dorim să fie importate datele și clic pe OK.

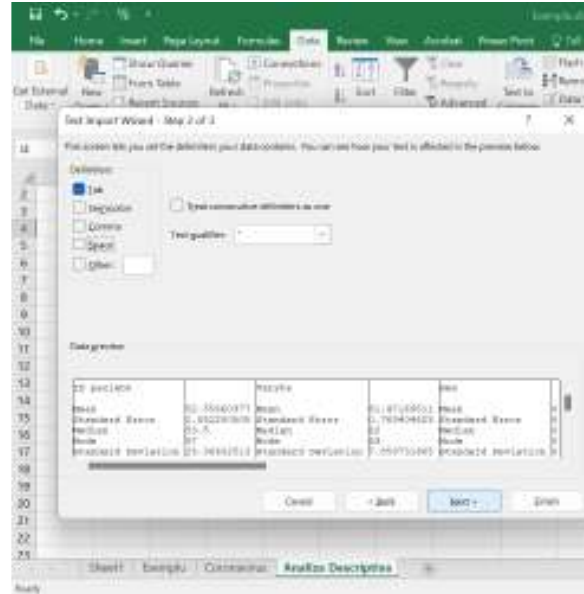
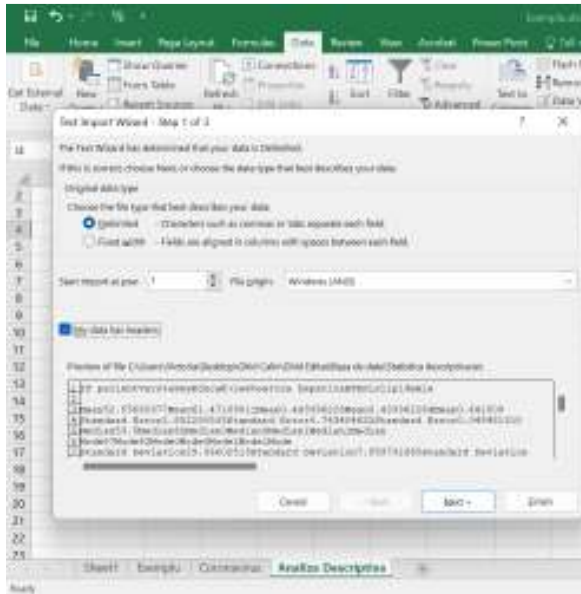


5.4 Importul datelor dintr-un fișier

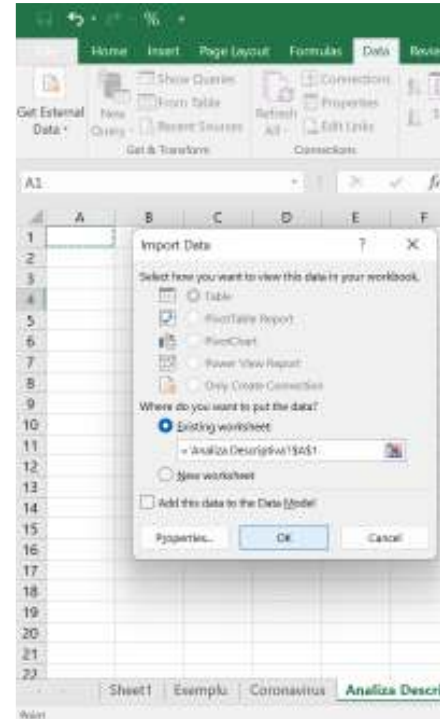
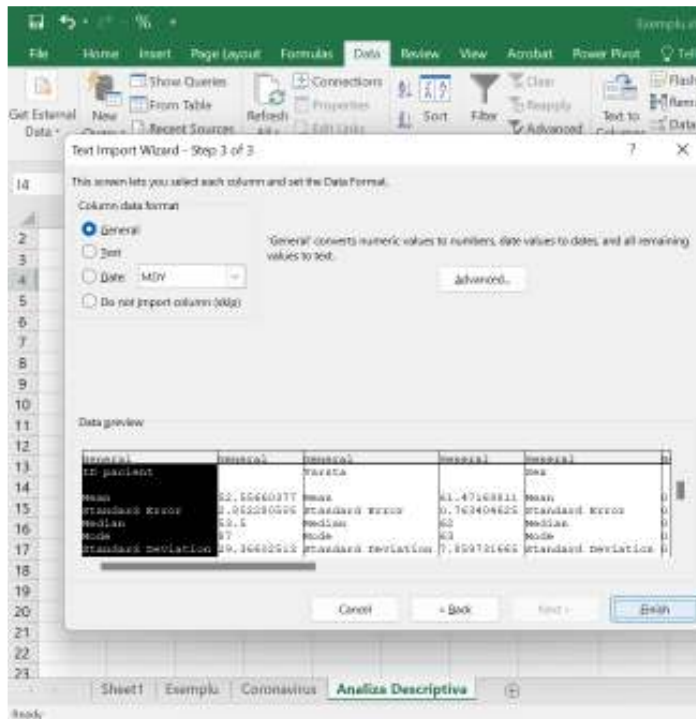
1. Pentru importul unor date dintr-un fișier, am creat o foaie de calcul goală denumită ***Analiza Descriptiva***, iar de la meniul **Data** selectăm **Get External Data – From Text**. În pasul următor se va afișa fereastra **Import Text File** unde va trebui să localizăm și să selectăm fișierul denumit ***Statistica descriptiva.txt*** din care dorim să importăm datele.



2. Facem clic pe **Import**, iar după aceea în fereastra **Text Import Wizard** care se afișează bifăm **Delimited**, clic pe **Next**, în următoarea fereastră bifăm **Tab** și **General** și clic pe **Finish**.



3. În fereastra **Import Data**, selectăm foaia de calcul în care dorim să se realizeze importul datelor. Clic pe **OK**.



4. Datele importate din fișierul .txt se fișează în modul următor:

6. CONSOLIDAREA DATELOR

Consolidarea datelor se referă la colectarea și integrarea datelor din mai multe surse într-o singură destinație. În timpul acestui proces, diferite surse de date sunt reunite sau consolidate într-un singur depozit de date. Deoarece datele provin dintr-o gamă largă de surse, consolidarea permite utilizatorilor să prezinte mai ușor datele, facilitând în același timp o analiză eficientă a datelor.

Tehnicile de consolidare a datelor reduc ineficiența, cum ar fi duplicarea datelor, costurile legate de dependența de mai multe baze de date și mai multe puncte de gestionare a datelor.

În consecință poate fi de un real folos să explicăm pașii care trebuie urmați pentru configurarea unei foi de lucru rezumată care consolidează datele din sursele existente configurate. Excel pune la dispoziție această tehnică de consolidare în câteva etape simple și rapide.

6.1 Pregătirea datelor pentru consolidare

Primul lucru de care trebuie să ne asigurăm înainte de a apela la tehnica de consolidare, este existența coerenței între sursele de date deținute. Datele trebuie să fie aranjate într-un anumit fel și să aibă o logică, acest lucru fiind valabil pentru fiecare sursă de date.

Caracteristici ale datelor ce urmează a fi consolidate:

- Fiecare arie de date să fie pe o foaie de lucru separată, sau registru separat.
- Fiecare arie de date să aibă format listă, cu etichete în primul rând.
- Toate ariile de date au același aspect.
- Toate ariile de date conțin acțiuni asemănătoare
- Să nu existe conținut gol (rânduri și coloane) în fiecare arie.

6.2 Consolidarea datelor în același registru de lucru

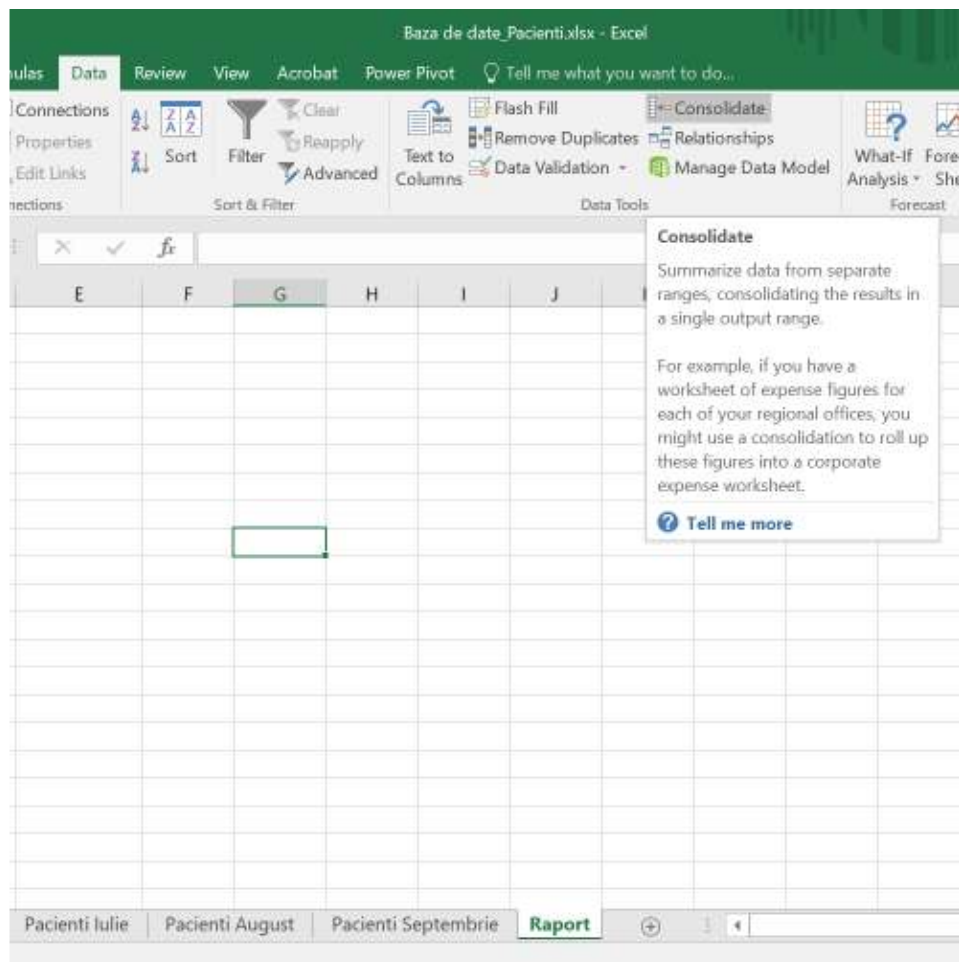
Pentru consolidarea datelor care se află în același fișier Excel, urmărim pașii care urmează:

1. Am creat câte o foaie de lucru separată pentru pacienții spitalizați în lunile Iulie, August, Septembrie din câteva județe ale țării, și anume **Pacienți Iulie**, **Pacienți August**, **Pacienți Septembrie**, de asemenea am creat o foaie de lucru necompletată denumită **Raport**.
2. În continuare o să lucrăm în foaia de calcul **Raport**, unde vom face clic pe celula în care dorim să se afișeze rezultatele consolidării, după care clic pe meniul **DATA**, urmărind secțiunea **Data Tools** – clic pe **CONSOLIDATE**.

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The ribbon includes options like 'Get External Data', 'Connections', 'Sort & Filter', and 'Data Validation'. The active cell is N16. Below the ribbon is a table with the following data:

	A	B	C	D	E	F	G	H	I
1	Judet	Numar pacienti diagnosticati cu IMA	Numar pacienti externati	Numar pacienti decedati					
2	Bistrita Nasaud	345	122	14					
3	Mures	432	233	16					
4	Iasi	123	89	11					
5	Bucuresti	324	32	23					
6	Sibiu	128	67	6					
7	Brasov	543	211	8					
8	Timisoara	422	333	5					
9	Bihor	166	78	8					
10	Satu Mare	342	321	4					
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									

At the bottom of the window, the sheet tabs are visible: 'Tabel_frecvente', 'PivotTable', 'Pacienți Iulie', 'Pacienți August', 'Pacienți Septembrie', and 'Raport'. The 'Pacienți Iulie' tab is currently selected and highlighted in yellow.



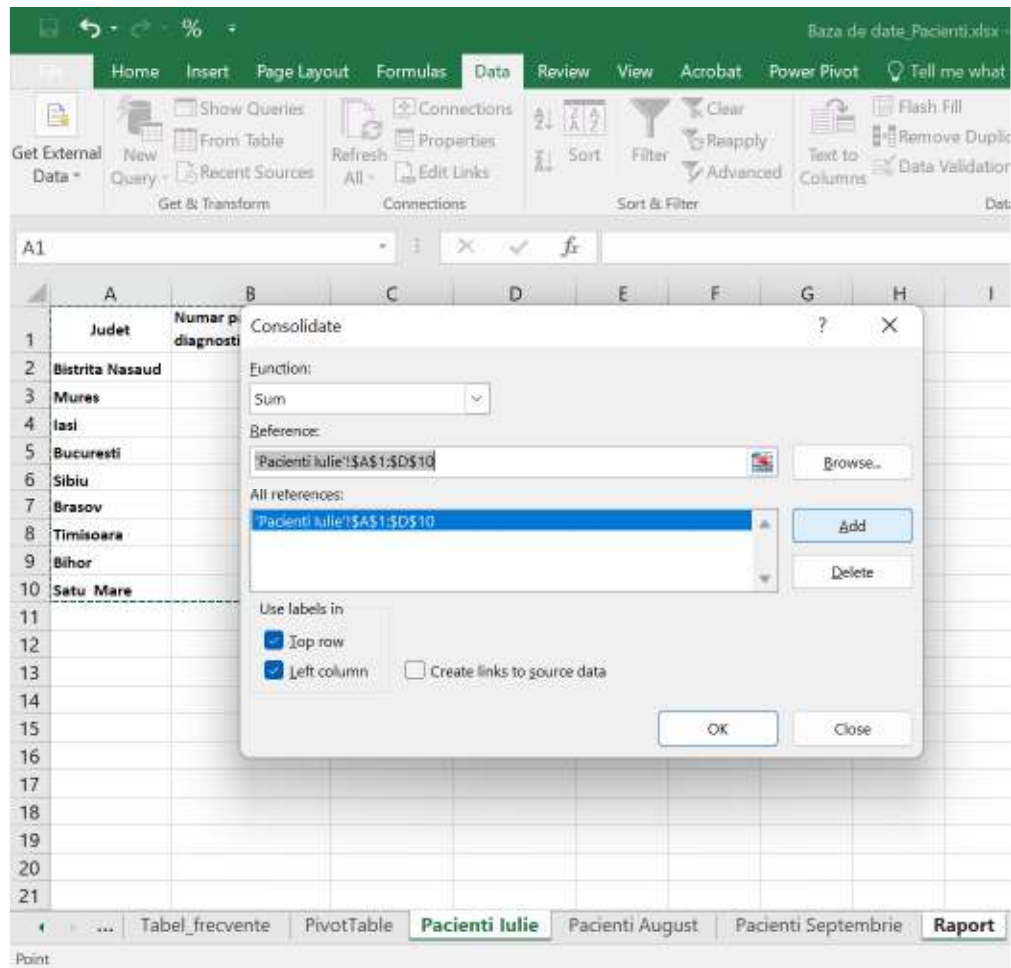
3. Se afișează fereastra de dialog **CONSOLIDATE** în care putem selecta și adăuga ariile de referință din cele 3 foi de calcul menționate mai sus. În câmpul **Function** selectăm opțiunea care ne interesează, și anume SUM, COUNT etc., bifăm **Use labels** în **First row** și **Left column**, pentru o bună structură a raportului care dorim să îl creăm, după care clic **OK**.

Microsoft Excel ribbon: Data > Connections > Refresh All

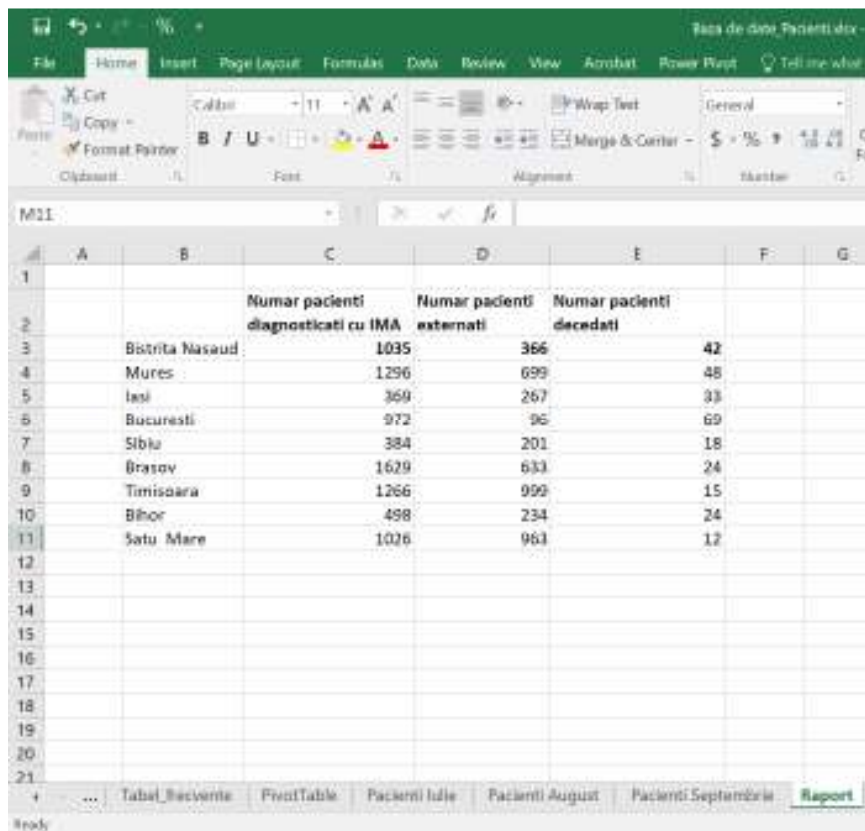
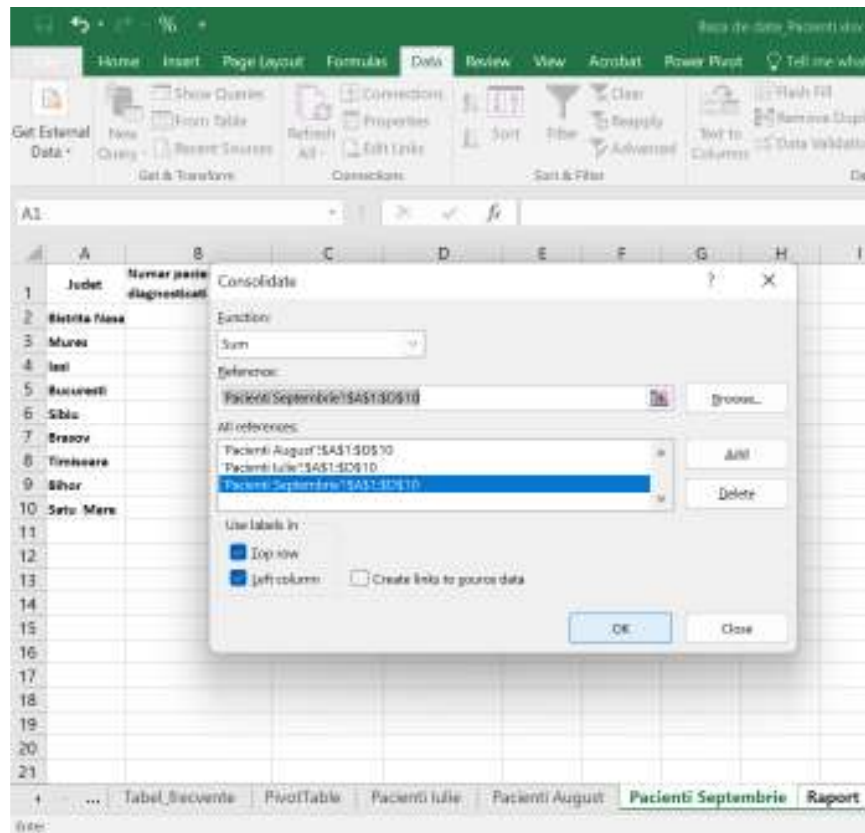
	A	B	C	D	E	F
1	Judet	Numar pacienti diagnosticati cu IMA	Numar pacienti externati	Numar pacienti decedati		
2	Bistrita Nasaud	345	122	14		
3	Mures	432	233	16		
4	Iasi	123	89	11		
5	Bucuresti	324	32	23		
6	Sibiu	128	67	6		
7	Brasov	543	211	8		
8	Timisoara	422	333	5		
9	Bihor	166	78	8		
10	Satu Mare	342	321	4		

Consolidate - Reference: 'Pacienti Iulie'!\$A\$1:\$D\$10

Worksheet tabs: Tabel_frecvente, PivotTable, Pacienti Iulie (selected), Pacienti August, Pacien...



4. Rezultatele comenzii **CONSOLIDATE** le putem vizualiza în imaginile de mai jos.



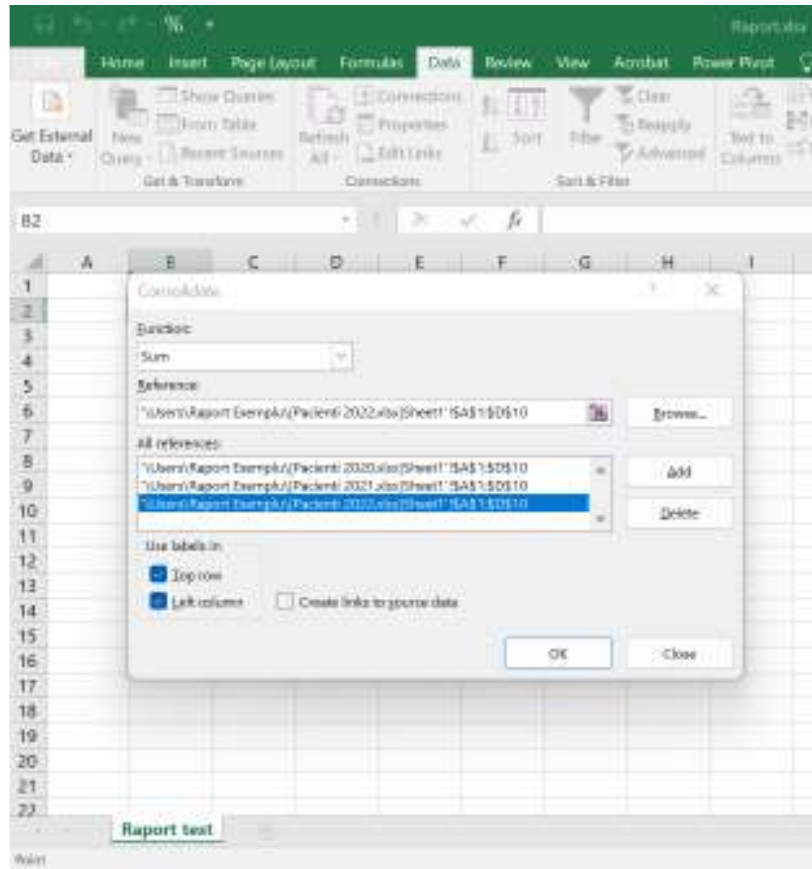
6.3 Consolidarea datelor din diferite registre de lucru

În situația în care avem nevoie să consolidăm anumite date și variabile care sunt stocate în registre diferite, putem să apelăm la următorii pași pentru realizarea raportului de consolidare.

1. Deschidem registrele de lucru care conțin datele și informațiile, denumite **Pacienti 2020, Pacienti 2021, Pacienti 2022**.
2. Deschidem un registru de lucru Excel necompletat numit **Raport**, creăm o foaie de lucru denumită **Raport test**, facem clic în celula în care dorim inserarea raportului, după care clic pe meniul **DATA**, urmărim secțiunea **Data Tools** – clic pe **CONSOLIDATE**.



3. Se afișează fereastra de dialog **CONSOLIDATE** în care putem selecta și adăuga ariile de referință din cele 3 registre de lucru Excel menționate mai sus. În câmpul **Function** selectăm opțiunea care ne interesează, și anume SUM, COUNT etc., bifăm **Use labels** în **First row** și **Left column**, pentru o bună structură a raportului pe care dorim să îl creăm, după care clic **OK**, și putem vizualiza Raportul sumar.



	Bistrita Nasaud	Mures	Iasi	Bucuresti	Sibiu	Brasov	Timisoara	Bihor	Satu Mare
Numar pacienti diagnosticati cu IMA	1035	1206	369	972	384	1629	1266	498	1026
Numar pacienti externati	366	609	267	96	201	633	999	234	963
Numar pacienti decedati	42	48	33	69	18	24	15	24	12

7. TABELE IN EXCEL

Când ne referim la un tabel, putem menționa că acesta reprezintă o arie dreptunghiulară de date structurate. Câteva dintre particularitățile cheie ale unui tabel sunt:

- Fiecare rând din tabel corespunde unei singure înregistrări a datelor. Un exemplu: informații despre pacienți.

- Fiecare coloană conține o anumită informație. Un exemplu: coloanele pot conține date precum codul de identificare, genul, nivelul educațional, rezidența și vârsta pacientului etc.

- Primul rând de sus expune informațiile conținute în fiecare coloană și este denumit rând antet.

- Fiecare intrare din rândul de sus este numită antet de coloană.

ID pacient	Vârsta	Sex	Educația	Starea hepatică	HTA	Diabetul	Medicamentul	Înălțimea	Greutate_cilindru
1	34	1	0	1	0	1	1	1	93.4
2	2	83	0	0	1	1	1	1	78
3	3	38	0	0	1	1	0	1	103.8
4	4	34	1	1	0	1	0	0	91.3
5	5	32	0	0	0	1	0	0	81
6	6	34	1	0	0	1	1	0	91.4
7	7	40	1	0	1	0	0	1	99.3
8	8	43	1	1	0	0	0	1	98.1
9	9	53	1	1	1	1	1	1	102.8
10	10	42	1	0	0	1	0	0	121.7
11	11	45	0	0	0	1	0	1	93.5
12	12	49	1	1	0	0	1	0	100
13	13	29	1	1	0	1	1	0	87
14	14	33	0	1	0	1	1	1	116.4
15	15	33	0	1	0	1	1	1	116.4

Realizarea tabelor în Excel reprezintă un mod foarte rapid și ușor pentru prelucrarea, administrarea și analiza datelor deținute. Tehnicile de Filtrare, Sortare și Umbrire reprezintă modalități de economisire a timpului și facilitează mult activitățile de raportare și sinteză a datelor.

De asemenea o altă caracteristică cheie a utilizării tabelor în Excel o reprezintă extinderea formulelor și funcțiilor de la o celulă la alta.

7.1 Distincția dintre tabel și interval (arie)

Mai jos vom prezenta câte aspecte referitoare la diferențe dintre un tabel și un interval sau arie din Excel:

- Utilizarea tabelului face ca modul de lucru să fie mai structurat și ordonat, decât intervalele.
- În câțiva pași putem realiza conversia unui interval într-un tabel, și prin urmare automat datele vor fi mai structurate, tabelul Excel conținând:
 - Nume tabel
 - Nume antet coloane
 - Formatarea datelor (culoarea fontului și a celulei) pentru vizualizare și evidențiere
- Existența unor caracteristici adiționale pentru tabele care nu sunt disponibile pentru intervale, precum:
 - Instrumentele de editare ale tabelului în panglică, de la proprietăți la stiluri.
 - Existența butonului Filtru în fiecare antet de coloană pentru a sorta datele sau a filtra tabelul astfel încât să fie afișate numai rândurile care îndeplinesc criteriile dorite.
 - Antetul tabelului se păstrează atunci când se realizează derularea rândurilor sau a foi de calcul.
 - Extinderea formulelor și funcțiilor de la o celulă la alta.
 - Extinderea dimensiunii tabelului adăugând mai multe rânduri sau mai multe coloane făcând clic și trăgând micul control triunghiular din colțul din dreapta jos al celulei din dreapta jos.

7.2 Crearea unui tabel în Excel

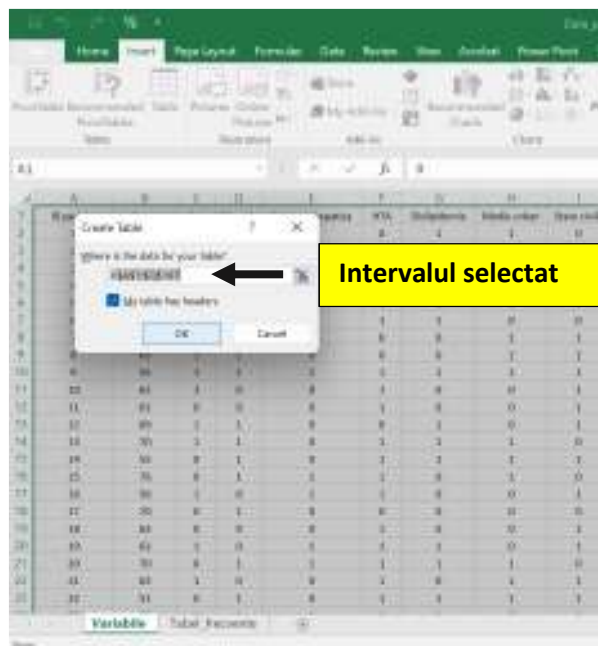
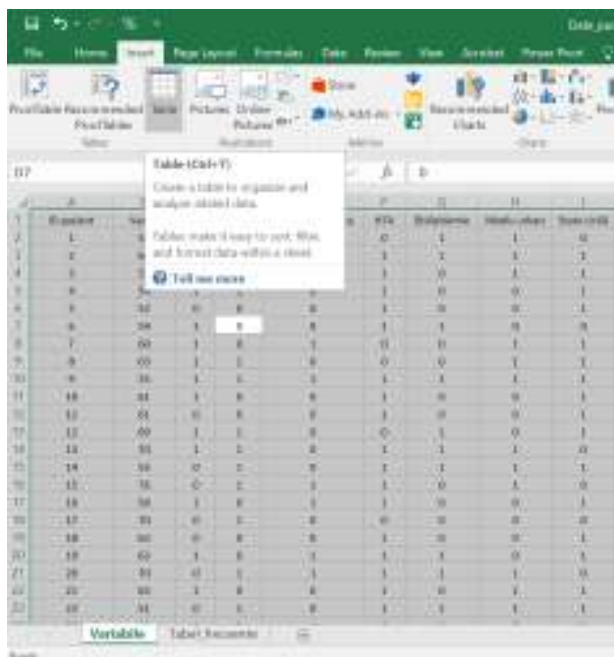
Mai jos vom prezenta pașii care trebuie urmați pentru a crea un tabel cu datele existente sub formă de interval:

1. Selectăm intervalul de celule care dorim să îl convertim în tabel. Celulele pot conține informații sau pot fi fără conținut. În imaginea următoare, vom prezenta o foaie de

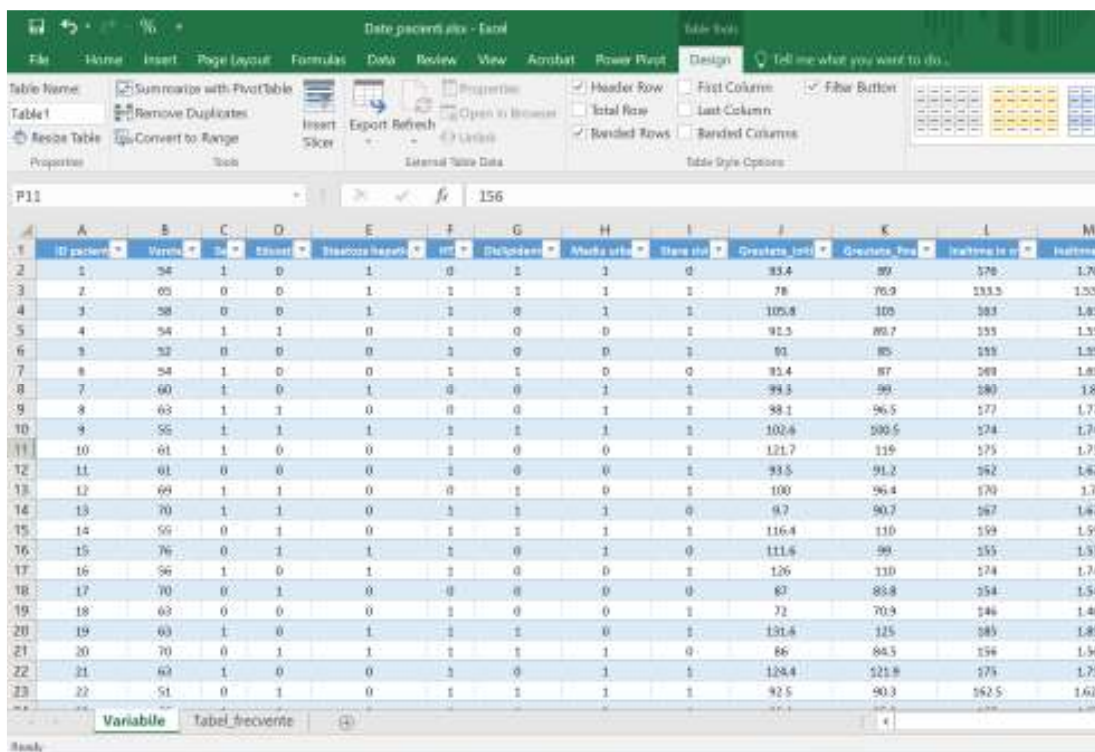
calcul care conține un interval cu 107 rânduri de date despre pacienți. Primul rând conține eticheta coloanei (antetul).

ID_pacient	Varsta	Sex	Educatie	Steatoza/hepatita	HTA	Diabetul	Mediu_urban	Stare_civila	Greutate_inital	Greutate_finala	Inaltime_in_cm	Inaltime_in_m	POC_vital
1	54	1	0	1	0	1	1	0	93.4	89	176	1.76	30.15
2	65	0	0	1	1	1	1	1	78	76.9	153.5	1.535	33.10
3	58	0	0	1	1	0	1	1	105.8	103	163	1.63	35.82
4	54	1	1	0	1	0	0	1	91.3	89.7	155	1.55	36.00
5	52	0	0	0	1	0	0	1	91	85	155	1.55	37.88
6	54	1	0	0	1	1	0	0	91.4	87	169	1.69	32.00
7	60	1	0	1	0	0	1	1	96.3	99	180	1.8	30.65
8	63	1	1	0	0	0	1	1	98.1	96.5	177	1.77	31.31
9	55	1	1	1	1	1	1	1	102.6	100.5	174	1.74	33.89
10	61	1	0	0	1	0	0	1	121.7	119	175	1.75	39.74
11	61	0	0	0	1	0	0	1	93.5	91.2	162	1.62	35.63
12	69	1	1	0	0	1	0	1	100	96.4	170	1.7	34.60
13	70	1	1	0	1	1	1	0	9.7	90.7	167	1.67	3.48
14	55	0	1	0	1	1	1	1	116.4	110	159	1.59	46.04
15	76	0	1	1	1	0	1	0	111.6	99	155	1.55	46.45
16	56	1	0	1	1	0	0	1	126	110	174	1.74	41.62
17	70	0	1	0	0	0	0	0	87	83.8	154	1.54	36.68
18	68	0	0	0	1	0	0	1	72	70.9	146	1.46	33.78
19	68	1	0	1	1	1	0	1	131.6	125	185	1.85	38.45
20	70	0	1	1	1	1	1	0	86	84.5	156	1.56	35.34
21	68	1	0	0	1	0	1	1	124.4	121.9	175	1.75	40.62
22	51	0	1	0	1	1	1	1	92.5	90.3	162.5	1.625	35.03

2. Vom selecta întregul interval, după care de la meniul **Insert** vom face clic pe **Table**, în fereastra următoare denumită **Create Table**, verificăm dacă intervalul de date selectat în caseta dialog **Where is the data for your table?**. Bifăm **My table has headers**, și clic pe **OK**.



3. Intervalul selectat este convertit în Tabel cu stilul implicit.



4. În meniul **Table Tools - Design**, putem atribui nume pentru tabelul creat. În secțiunea **Properties**, va fi afișată caseta **Table Name** și de acolo se poate modifica.

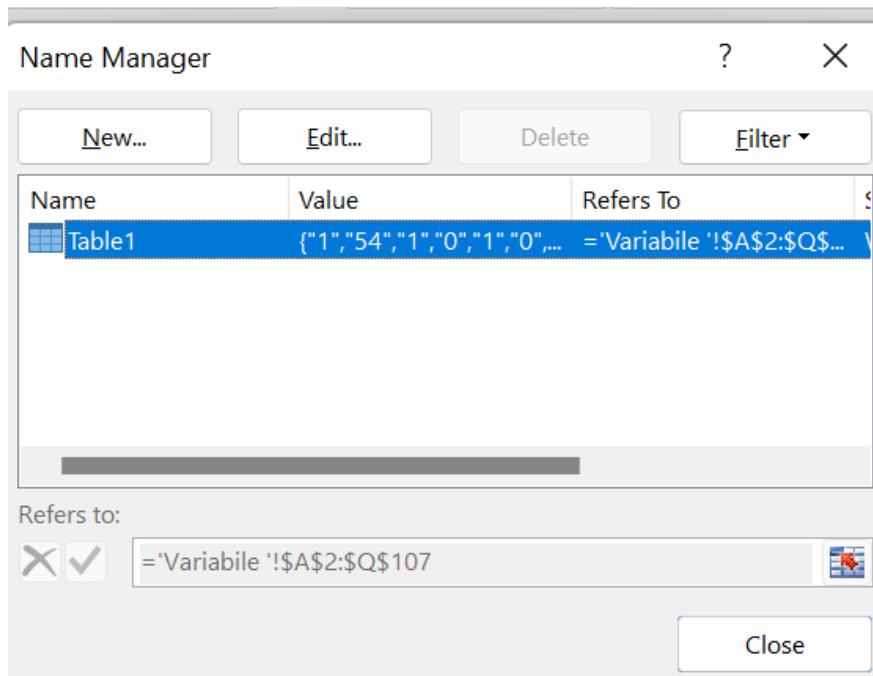
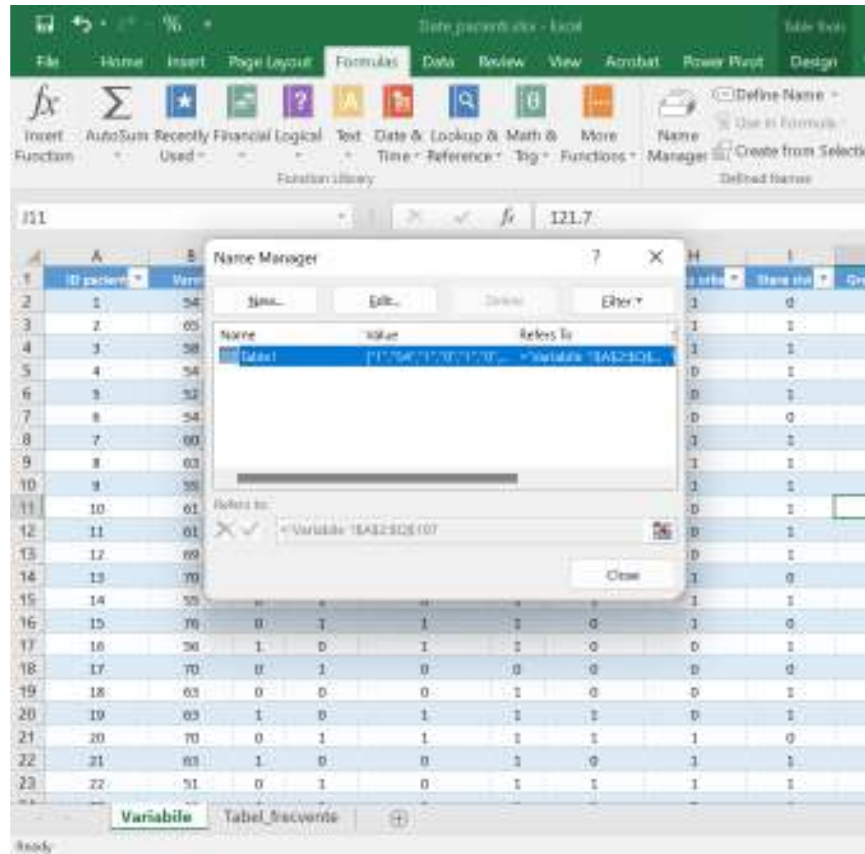
	A	B	C	D	E	F	G	H	I
1	ID pacient	Varsta	Sex	Educat	Streatoza hepatic	HT	Dialpidem	Mediu urbe	Stare civil
2	1	54	1	0	1	0	1	1	0
3	2	65	0	0	1	1	1	1	1
4	3	58	0	0	1	1	0	1	1
5	4	54	1	1	0	1	0	0	1
6	5	52	0	0	0	1	0	0	1
7	6	54	1	0	0	1	1	0	0
8	7	60	1	0	1	0	0	1	1
9	8	63	1	1	0	0	0	1	1
10	9	55	1	1	1	1	1	1	1
11	10	61	1	0	0	1	0	0	1
12	11	61	0	0	0	1	0	0	1
13	12	69	1	1	0	0	1	0	1
14	13	70	1	1	0	1	1	1	0
15	14	55	0	1	0	1	1	1	1
16	15	76	0	1	1	1	0	1	0
17	16	56	1	0	1	1	0	0	1
18	17	70	0	1	0	0	0	0	0
19	18	63	0	0	0	1	0	0	1
20	19	63	1	0	1	1	1	0	1
21	20	70	0	1	1	1	1	1	0
22	21	63	1	0	0	1	0	1	1
23	22	51	0	1	0	1	1	1	1

7.3 Gestionarea numelor într-un tabel

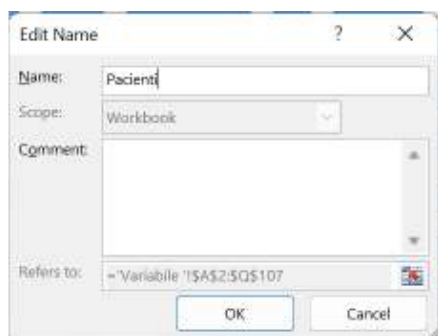
Cu ajutorul **Name Manager** putem gestiona numele tabelelor create la fel ca și cum gestionați numele intervalelor.

1. Facem clic pe tabelul inserat, după care clic pe **Formulas - Defined Names - Name Manager**.

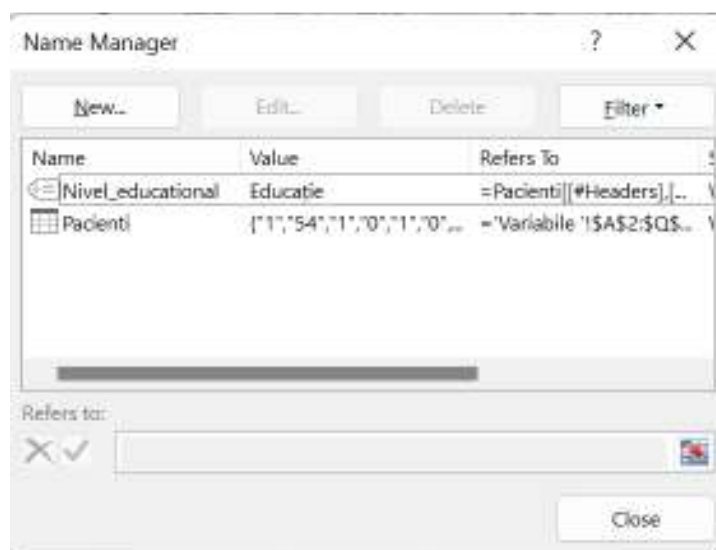
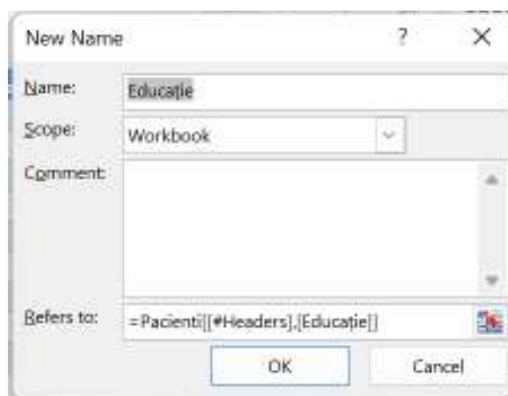
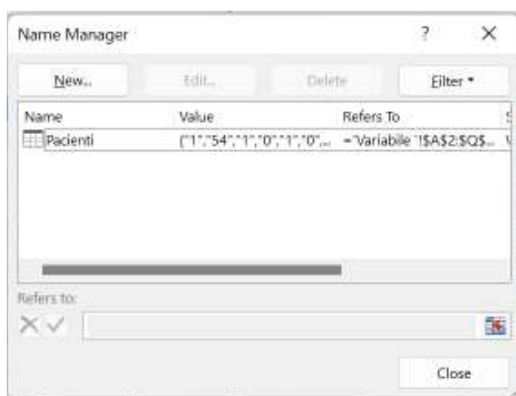
Se afișează caseta de dialog **Name Manager** și putem găsi numele tabelelor din foaia de calcul denumită Variabile.



2. Tabelul creat de noi are denumirea **Table1**, după cum se poate vedea și în imaginile de mai sus, însă dorim să modificăm numele din **Table1** în **Pacienți**, pentru a face acest lucru în fereastra **Name Manager** facem clic pe **Edit**.



De asemenea pe lângă numele tabelului mai putem modifica și numele antetelor de coloană. De exemplu antetul de coloană numit **Educație** am dori să îl modificăm în **Nivel_educational**, iar pentru a face asta trebuie să: facem clic pe antetul de coloană **Educație** din tabel, după aceea clic pe **Name Manager** din meniul **Formulas** (ca în pașii anteriori). După ce se afișează caseta de dialog, facem clic pe **New**. În caseta **Name**, regăsim Antetul coloanei pe care dorim să o edităm.



7.4 Propagarea unei formule într-un tabel

În tabelul denumit **Pacienți**, avem înregistrați și indicatorii antropometrici **Înălțime** și **Greutate**, iar în exemplul următor vom demonstra cum calculăm **Indicele de masă corporală** (IMC) prin propagarea formulei la toți pacienții din tabel.

1. Creăm 2 coloane noi (necompletate) în tabel, denumite **IMC_initial** și **IMC_final**, iar în oricare dintre celulele coloanei **IMC_initial** tastăm formula de calcul pentru IMC, $=[@[Greutate_initial]]/([@[Inaltime\ in\ m]]^2)$ și tastăm **Enter**.

The screenshot shows an Excel spreadsheet with a table containing patient data. The table has columns for 'Stare civilă', 'Greutate_initial', 'Greutate_final', 'Inaltime in cm', 'Inaltime in m', 'IMC_initial', 'IMC_final', 'Glicemie 1', and 'Glicemie 2'. The data is as follows:

Stare civilă	Greutate_initial	Greutate_final	Inaltime in cm	Inaltime in m	IMC_initial	IMC_final	Glicemie 1	Glicemie 2
0	93.4	89	176	1.76			149	109
1	78	76.9	153.5	1.535			73	100
1	105.8	103	163	1.63			121	94
1	91.3	89.7	155	1.55			105	93
1	91	85	155	1.55			105	93
0	91.4	87	169	1.69			123	90
1	99.3	99	180	1.8			131	100
1	98.1	96.5	177	1.77			70	70
1	102.6	100.5	174	1.74			103	85
1	121.7	119	175	1.75			156	70
1	93.5	91.2	162	1.62			117	92
1	100	96.4	170	1.7			91	91
0	9.7	90.7	167	1.67			88	80
1	116.4	110	159	1.59			115	107
0	111.6	99	155	1.55			103	102
1	126	110	174	1.74			206	145
0	87	83.8	154	1.54			157	132
1	72	70.9	146	1.46			144	125
1	131.6	125	185	1.85			103	90
0	86	84.5	156	1.56			139	139
1	124.4	121.9	175	1.75			104	104
1	92.5	90.3	162.5	1.625			118	120

The screenshot shows an Excel spreadsheet with a table containing patient data. The table has columns for 'Stare civil', 'Greutate_initial', 'Greutate_final', 'Inaltime in cm', 'Inaltime in', 'IMC_initial', and 'IMC_final'. A formula bar at the top displays the formula: $=[@[Greutate_initial]]/([@[Inaltime\ in\ m]]^2)$. The formula is applied to the 'IMC_initial' column, and the value 1.69 is visible in cell M3.

Stare civil	Greutate_initial	Greutate_final	Inaltime in cm	Inaltime in	IMC_initial	IMC_final
0	93.4	89	176	1.76	30.15	
1	78	76.9	153.5	1.535	33.10	
1	105.8	103	163	1.63	39.82	
1	91.3	89.7	155	1.55	38.00	
1	91	85	155	1.55	37.98	
0	91.4	87	169	1.69	30.65	
1	99.3	99	180	1.8	30.65	
1	98.1	96.5	177	1.77	31.31	
1	102.6	100.5	174	1.74	33.89	
1	121.7	119	175	1.75	39.74	
1	93.5	91.2	162	1.62	35.63	
1	100	96.4	170	1.7	34.60	
0	9.7	90.7	167	1.67	3.48	
1	116.4	110	159	1.59	46.04	
0	111.6	99	155	1.55	46.45	
1	126	110	174	1.74	41.62	
0	87	83.8	154	1.54	36.68	
1	72	70.9	146	1.46	33.78	
1	131.6	125	185	1.85	38.45	
0	86	84.5	156	1.56	35.34	
1	124.4	121.9	175	1.75	40.62	
1	92.5	90.3	162.5	1.625	35.03	

Formula de calcul pentru IMC se propagă automat în celelalte celule din aceea coloană a tabelului.

7.5 Redimensionarea tabelului

Excel pune la dispoziție și opțiunea de redimensionare a unui tabel inserat, putem elimina sau adăuga coloane/rânduri în cadrul tabelului. Avem tabelul denumit **Pacienți** care conține 107 rânduri, și dorim să mai adăugăm 3 pacienți, adică vom insera încă 3 rânduri.

1. Facem clic pe tabel și vom derula pagina până la finalul tabelului, iar în colțul dreapta jos al tabelului se găsește un punctuleț denumit **Control**, vom face clic pe **Control** și vom trage în jos pentru a insera 3 rânduri și în dreapta pentru inserarea unei noi coloane.

O	P	Q	R
1.06	106	92	
1.09	264	164	
1.06	172	102	
1.08	114	120	
1.07	124	123	
1.10	148	104	
1.06	153	113	
1.06	292	203	Control

Inaltime ii	IMC_initial	IMC_final	Glicemie 1	Glicemie 2
1.73	44.44	0.06	106	92
1.73	33.81	0.09	264	164
1.6	43.83	0.06	172	102
1.66	30.66	0.08	114	120
1.485	33.28	0.07	124	123
1.8	29.94	0.10	148	104
1.7	41.90	0.06	153	113
1.65	41.10	0.06	292	203

După inserarea rândurilor și coloanei necompletate, tabelul va arăta precum în imaginea de mai jos.

The screenshot shows the Microsoft Excel interface with the 'Formulas' tab selected. The formula bar displays 'R100'. Below it is a data table with the following columns: Inaltime (m), IMC_initial, IMC_final, Glicemie 1, Glicemie 2, and Column1. The data rows are numbered 100 to 112.

	Inaltime (m)	IMC_initial	IMC_final	Glicemie 1	Glicemie 2	Column1
100	1.73	44.44	0.06	106	92	
101	1.73	33.81	0.09	264	164	
102	1.6	43.83	0.06	172	102	
103	1.66	30.66	0.08	114	120	
104	1.485	33.28	0.07	124	123	
105	1.8	29.94	0.10	148	104	
106	1.7	41.90	0.06	153	113	
107	1.65	41.10	0.06	292	203	
108						
109						
110						
111						
112						

7.6 Eliminarea datelor duplicate în Excel

În funcție de sursa datelor colectate și de modul de colectarea al acestora, este foarte probabil ca bazele de date să conțină date și informații duplicate. Pentru o analiză corectă și validă a datelor, eliminarea valorilor duplicate și curățarea bazei de date este absolut necesară.

Uitați-vă la următoarele date unde aveți informații despre diverse produse de diferite mărci. Să presupunem că doriți să eliminați duplicatele din aceste date.

1. Avem un tabel care conține substanța activă și denumirea medicamentului OTC, și dorim să verificăm și să eliminăm valorile duplicate. Facem clic pe tabelul care conține datele de interes, iar din meniul **Table Tools - DESIGN**, clic pe **Remove Duplicates** din secțiunea **Tools**.
2. Se afișează Fereastra **Remove Duplicates** care conține numele coloanelor existente în tabel, bifăm coloanele și clic **OK**.

	A	B	C	D
1		Substanta activa	Medicament OTC	
2		Paracetamol	Panadol	
3		Acid acetil-salicilic	Aspirina	
4		Ibuprofen	Advil	
5		Dexketoprofen	Tador	
6		Acid acetil-salicilic	Aspacardin	
7		Ibuprofen	Nurofen forte	
8		Vitamina C	Gripovit Max C	
9		Ibuprofen	Advil	
10		Paracetamol	Panadol	
11		Paracetamol	Panadol	
12				

Table Name: Table24

Tools:

- Summarize with PivotTable
- Remove Duplicates
- Convert to Range

Remove Duplicates

To delete duplicate values, select one or more columns that contain duplicates.

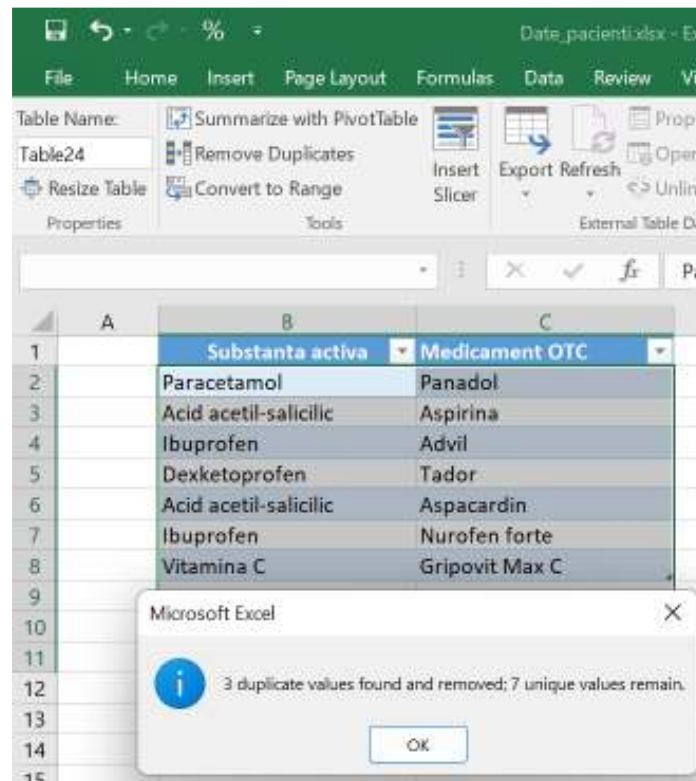
Select All
 Unselect All
 My data has headers

Columns:

- Substanta activa
- Medicament OTC

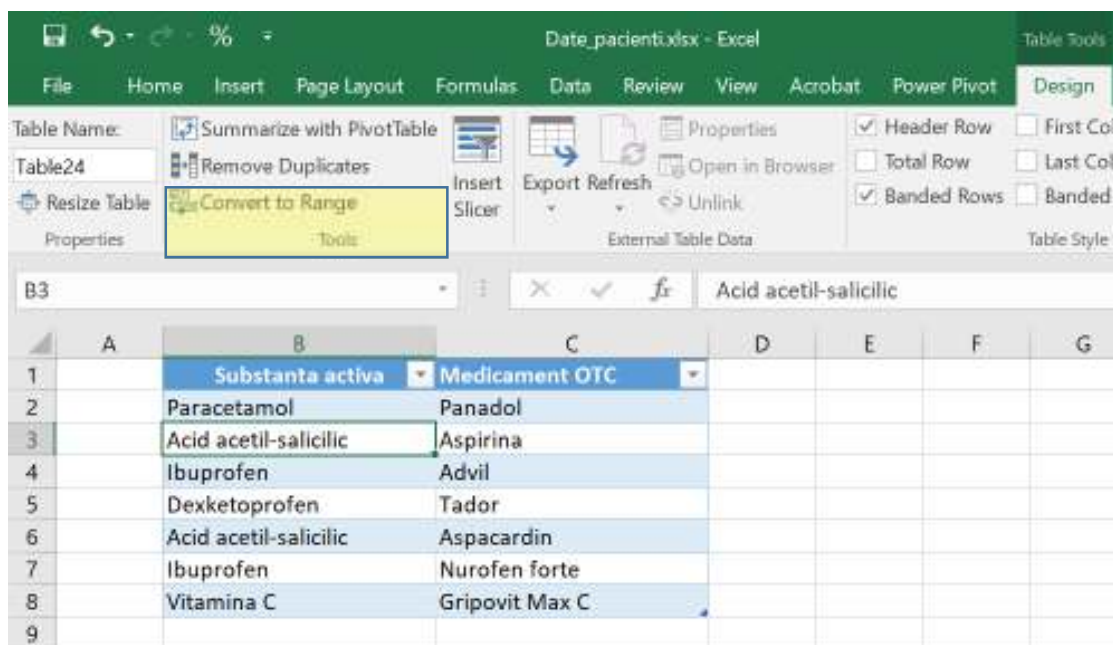
OK Cancel

În urma analizării duplicatelor, se va afișa un mesaj despre câte rânduri cu valori duplicate vor fi eliminate și câte valori unice rămân.

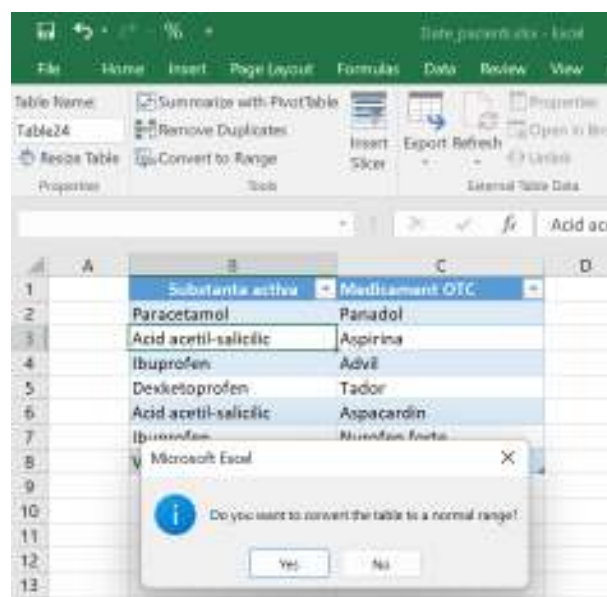
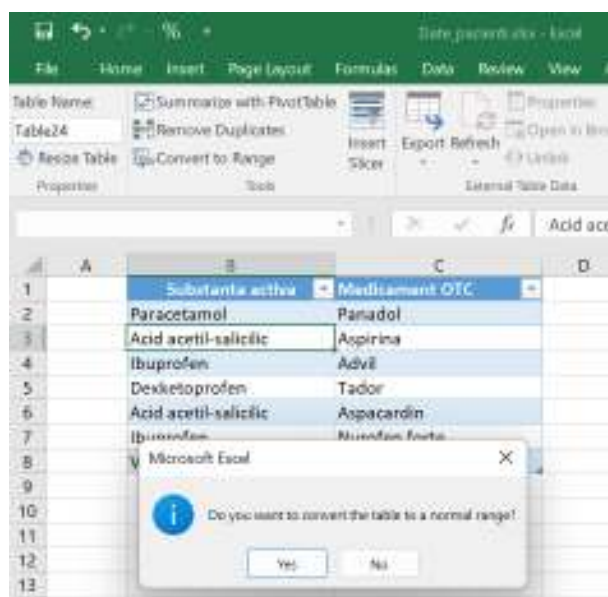


7.7 Convertirea tabelului în interval (arie)

Dacă Excel pune la dispoziție conversia unui interval în tabel, avem desigur și opțiunea inversă, conversia unui tabel în interval. Ca și exemplu, avem un tabel care conține substanța activă și denumirea medicamentului OTC, și dorim să îl convertim în interval. Facem clic pe tabelul care conține datele de interes, iar din meniul **Table Tools - DESIGN**, clic pe **Convert to Range** din secțiunea **Tools**.



Rezultatul se poate vizualiza în figurile de mai jos.



8. CURĂȚAREA DATELOR CU FUNCȚII DE TEXT

După cum am menționat de mai multe ori pe parcursul acestui ghid, este foarte probabil ca datele care le obținem din diferite surse de colectare a datelor să nu prezinte un format pregătit pentru o analiză corespunzătoare.

Odată colectate în forma brută, datele trebuie curățate, iar procesul de curățare a datelor reprezintă eliminarea caracterelor nedorite din text. Procedee de structurare ale datelor pentru analiză sunt:

- Identificarea modelelor de text necesare prin utilizarea funcțiilor de text.
- Extragerea valorilor datelor din text.
- Formatarea datelor cu ajutorul funcțiilor text.
- Realizarea operațiunilor de date cu funcțiile text.

8.1 Eliminarea caracterelor nedorite din text

În unele cazuri se întâmplă ca datele importate să prezinte caractere neprintabile și/sau spații în exces. Spațiile în exces pot fi spații de conducere și/sau spații suplimentare între cuvinte.

Un exemplu de astfel de date prezentăm în imaginea de mai jos, în care se poate vizualiza informațiile brute ale unui raport referitor la substanța activă, numele medicamentului, brandul și prețul medicamentelor iar în pașii următori vom prezenta tehnici de curățare a datelor de caracterele nedorite. Datele sunt importate într-o singură coloană iar caracterul „|” separă câmpul din fiecare rând, iar unele rânduri conțin spații în exces, care trebuie eliminate.

Raport	
Substanta activa	Medicament OTCBrandPret
Paracetamol Panadol	Sandoz34.5
Acid acetil-salicilic	AspirinaGee on55.6
Ibuprofen	Adv ilBiopharma123
Dexketoprofen	TadorFarmexim22
Acid acetil-salicilic	AspacardinFarmexpert134
Ibuprofen	Nurofen fort eSandoz322
Vitamina C	Gripovit Max CBioe el123

Pentru eliminarea spațiilor în exces putem folosi CLEAN și funcția TRIM.

- **CLEAN** - Elimină toate caracterele care nu pot fi imprimate din text
- **TRIM** - Elimină spațiile din text

F		G	
Raport		Date curatate	
Substanta activa	Medicament OTCBrandPret	=TRIM(CLEAN(F5))	
Paracetamol Panadol	Sandoz34.5		
Acid acetil-salicilic	AspirinaGee on55.6		
Ibuprofen	Adv ilBiopharma123		
Dexketoprofen	TadorFarmexim22		
Acid acetil-salicilic	AspacardinFarmexpert134		
Ibuprofen	Nurofen fort eSandoz322		
Vitamina C	Gripovit Max CBioe el123		

Selectăm aria din coloana adiacentă corespondentă ariei care conține datele colectate, tastăm =TRIM (CLEAN (F5)) și apoi apăsăm CTRL+Enter, iar formula se propagă în toată aria, curățând datele de spațiile în exces.

Raport	Date curatate
Substanta activaMedicament OTCBrandPret	Substanta activaMedicament OTCBrandPret
Paracetamol PanadolSandoz34.5	Paracetamol PanadolSandoz34.5
Acid acetil-salicilicAspirinaGee on55.6	Acid acetil-salicilicAspirinaGee on55.6
IbuprofenAdv ilBiopharma123	IbuprofenAdv ilBiopharma123
DexketoprofenTadorFarmexim22	DexketoprofenTadorFarmexim22
Acid acetil-salicilicAspacardinFarmexpert134	Acid acetil-salicilicAspacardinFarmexpert134
IbuprofenNurofen fort eSandoz322	IbuprofenNurofen fort eSandoz322
Vitamina CGripovit Max CBioe el123	Vitamina CGripovit Max CBioe el123

8.2 Identificarea modelelor de text necesare cu funcțiile de text

Pentru efectuarea procesului de structurare a datelor, uneori este necesar să realizăm anumite potriviri de model de text pe baza cărora să putem extrage valorile datelor. În acest scop putem apela la funcțiile de text:

- **EXACT** - Verifică dacă două valori de text sunt identice
- **FIND** - Găsește o valoare de text într-o altă valoare (sensibilă la majuscule și minuscule)
- **SEARCH** - Găsește o valoare de text într-o altă valoare (fără diferențiere între majuscule și minuscule)

8.3 Extragerea valorilor datelor din text

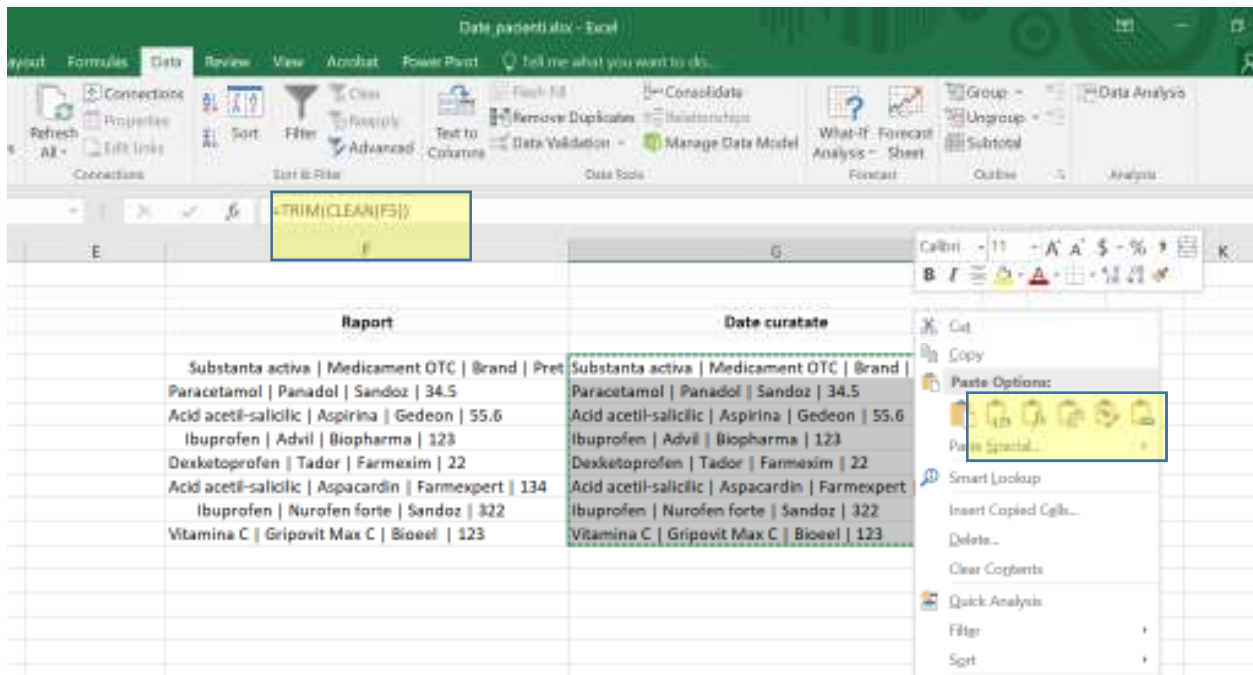
În exemplul de pe punctul 8.1, avem importate datele într-o singură coloană, însă ca să le putem analiza, trebuie să le prelucrăm să fie afișate pe 4 coloane: Substanta activa, Medicament OTC, Brand și Preț. Pentru extragerea datelor putem apela la următoarele moduri:

- Extragerea valorilor datelor cu **Text to Columns Wizard** în coloane
- Extragerea valorilor de date cu funcții de text
- Extragerea valorilor datelor cu **Flash Fill**

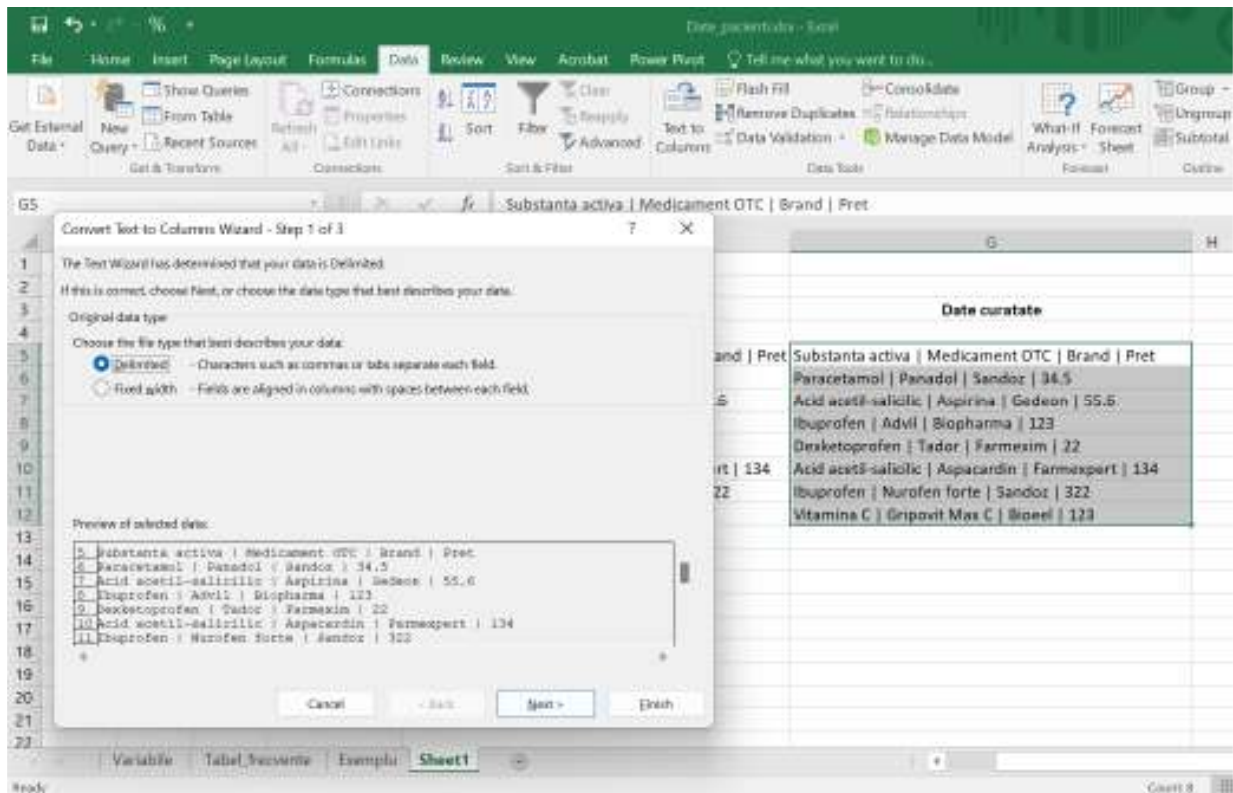
8.3.1 Extragerea valorilor datelor cu Text to Columns Wizard

Folosim **Text to Columns Wizard** pentru a extrage valorile datelor în coloane Excel dacă câmpurile sunt delimitate de un caracter sau sunt aliniate în coloane cu spații între fiecare câmp. În exemplul nostru, câmpurile sunt separate de caracterul „|”.

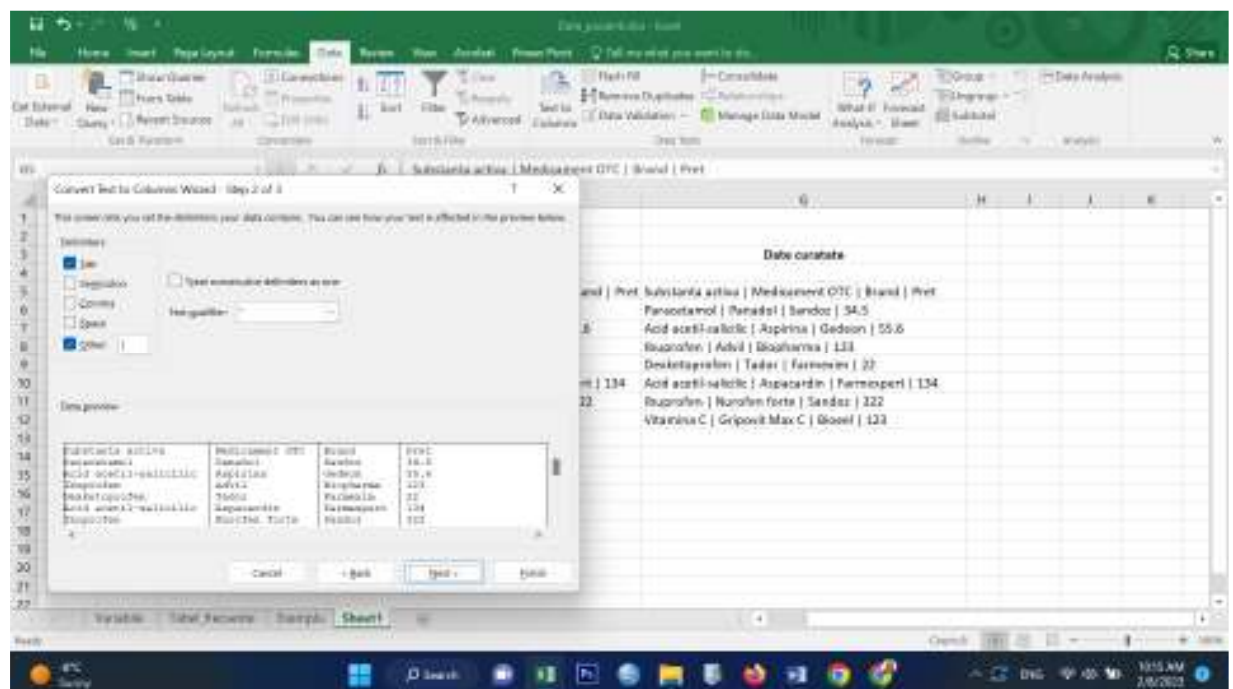
1. Selectăm datele, le copiem și le lipim în același loc sub forma de **Paste as values** pentru ca să dispară formula aplicată anterior pentru curățarea datelor.



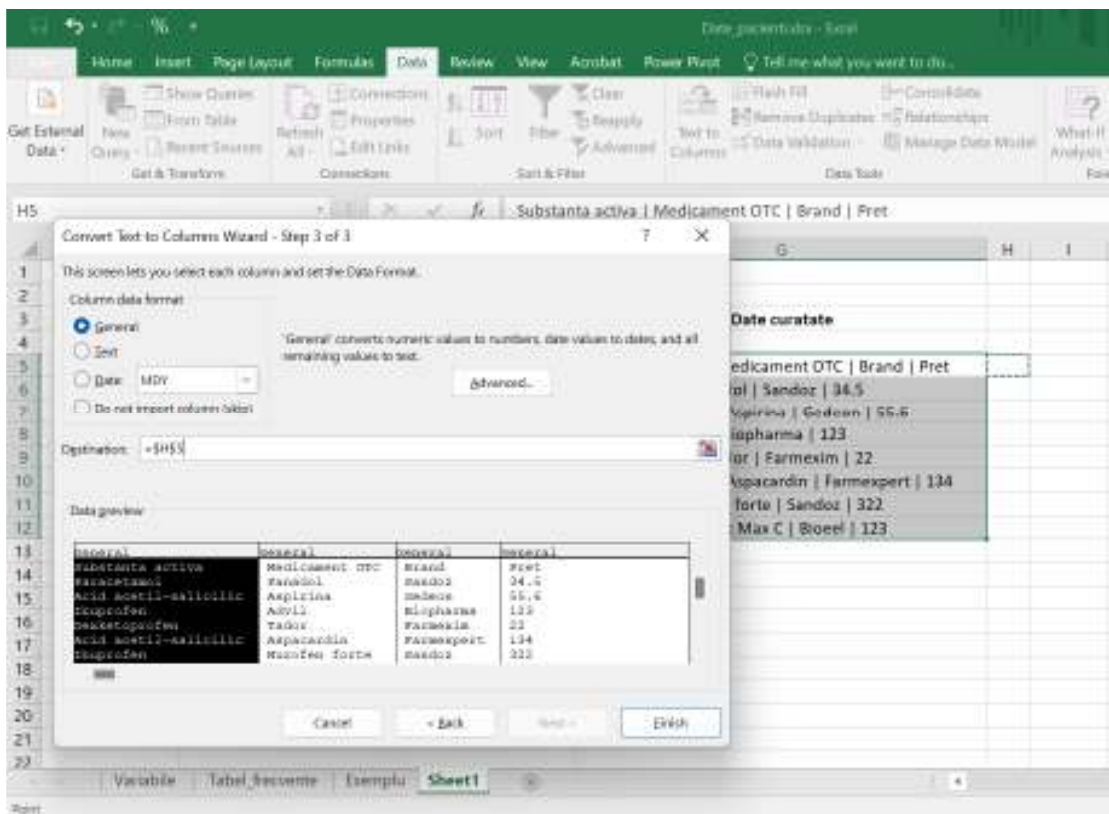
2. După care selectăm din nou datele, iar din meniul **Data** – secțiunea **Data Tools**, selectăm **Text to Columns Wizard**. Iar în prima casetă de dialog apărută bifăm **Delimited** și clic pe **Next**.



3. În pasul următor selectăm delimitatorul conținut de datele noastre, în acest caz vom bifa **Other**, și vom introduce manual caracterul "|", după care clic pe **Next**.



4. În ultimul pas, selectăm destinația unde dorim inserarea datelor prelucrate și putem vizualiza un preview al datelor ce urmează să fie afișate după ce facem clic pe **Finish**.



5. În figura de mai jos putem vizualiza rezultatul, textul dintr-o singură coloană transpus în cele 4 coloane, iar fiecare dintre cele 4 coloane prezintă un antet.

Raport	Date curatate	Substanta activa	Medicament OTC	Brand	Pret
Substanta activa Medicament OTC Brand Pret	Substanta activa Medicament OTC Brand Pret	Substanta activa	Medicament OTC	Brand	Pret
Paracetamol Panadol Sandoz 34.5	Paracetamol Panadol Sandoz 34.5	Paracetamol	Panadol	Sandoz	34.5
Acid acetil-salicilic Aspirina Gedeon 55.6	Acid acetil-salicilic Aspirina Gedeon 55.6	Acid acetil-salicilic	Aspirina	Gedeon	55.6
Ibuprofen Advil Biopharma 123	Ibuprofen Advil Biopharma 123	Ibuprofen	Advil	Biopharma	123
Dexketoprofen Tador Farmexim 22	Dexketoprofen Tador Farmexim 22	Dexketoprofen	Tador	Farmexim	22
Acid acetil-salicilic Aspacardin Farmexpert 134	Acid acetil-salicilic Aspacardin Farmexpert 134	Acid acetil-salicilic	Aspacardin	Farmexpert	134
Ibuprofen Nurofen forte Sandoz 322	Ibuprofen Nurofen forte Sandoz 322	Ibuprofen	Nurofen forte	Sandoz	322
Vitamina C Gripovit Max C Bioeel 123	Vitamina C Gripovit Max C Bioeel 123	Vitamina C	Gripovit Max C	Bioeel	123

8.3.2 Extragerea valorilor datelor cu funcții text

Să luăm un exemplu în care datele colectate nu sunt delimitate de caractere și nici nu sunt alinierte în coloane cu spații între fiecare câmp, în acest caz, putem apela la funcții de text care extrag valorile datelor. Aceste funcții sunt:

- **LEFT** - Returnează caracterele cele mai din stânga dintr-o valoare text

- **RIGHT** - Returnează caracterele cele mai din dreapta dintr-o valoare text
- **MID** - Returnează un anumit număr de caractere dintr-un șir de text începând cu poziția pe care o specificați
- **LEN** - Returnează numărul de caractere dintr-un șir de text

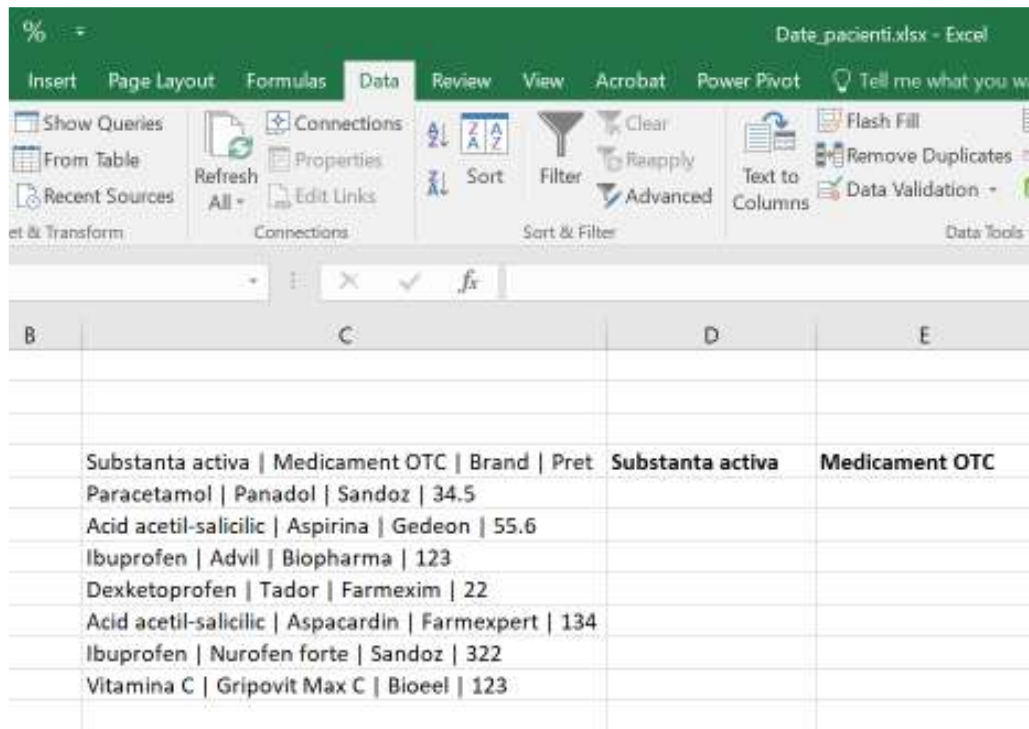
În funcție de datele deținute, funcțiile de mai sus se pot combina pentru a putea extrage valorile necesare.

8.3.3 Extragerea valorilor datelor cu Flash Fill

O altă tehnică de extragere a valorilor datelor din text o reprezintă Flash Fill, însă această tehnică este utilă atunci când Excel poate să identifice un model în datele respective.

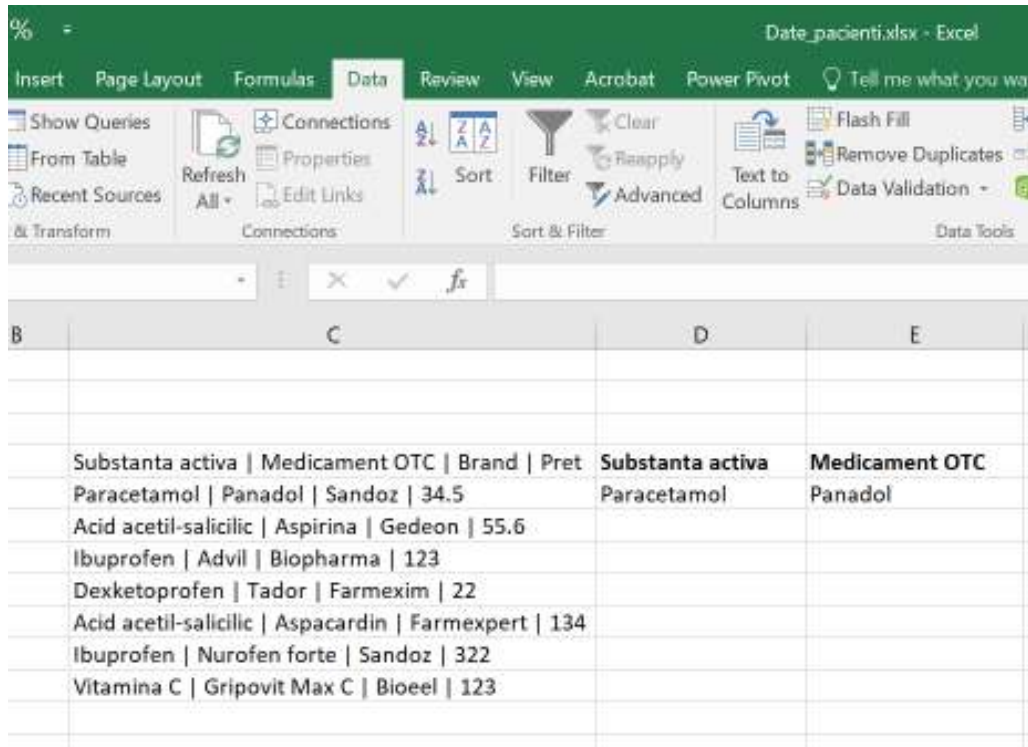
În cele ce urmează vom realiza un exemplu.

1. Vom crea 2 coloane cu numele **Substanta activa** și **Medicament OTC**.

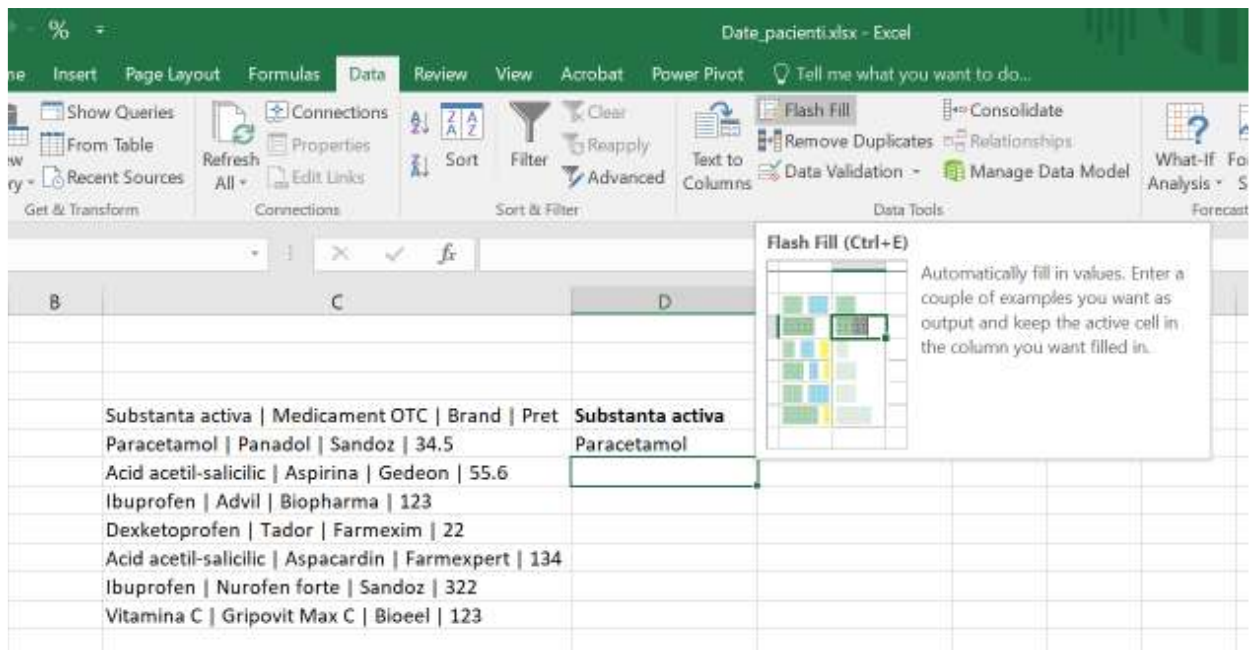


B	C	D	E
	Substanta activa Medicament OTC Brand Pret	Substanta activa	Medicament OTC
	Paracetamol Panadol Sandoz 34.5		
	Acid acetil-salicilic Aspirina Gedeon 55.6		
	Ibuprofen Advil Biopharma 123		
	Dexketoprofen Tador Farmexim 22		
	Acid acetil-salicilic Aspacardin Farmexpert 134		
	Ibuprofen Nurofen forte Sandoz 322		
	Vitamina C Gripovit Max C Bioeel 123		

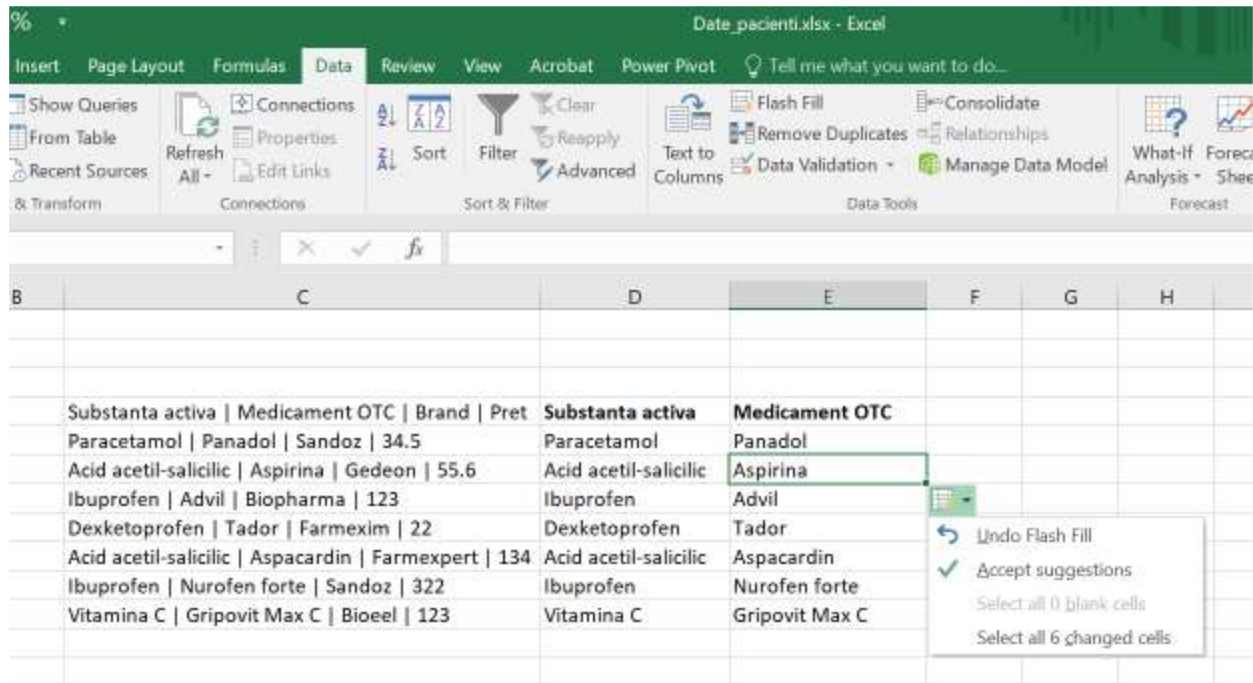
2. După care vom completa primul rând de la coloanele create cu informațiile din datele brute.



3. Vom face clic în al doilea rând, iar din meniul **Data - Data Tools**, vom face clic pe **Flash Fill**.



4. Cu ajutorul **Flash Fill**, coloana Substanta activa a fost completată automat cu informațiile din coloana cu date brute. Am repetat acești pași și pentru coloana Medicament OTC.



8.4 Formatarea datelor cu funcții de text

Pentru formatarea datelor care conțin text în funcție de necesități, Microsoft Excel pune la dispoziție următoarele funcții de text:

- **LOWER** - Convertește textul în minuscule
- **UPPER** - Convertește textul în majuscule
- **PROPER** - Scrie cu majuscule prima literă din fiecare cuvânt al unei valori de text
- **DOLLAR** - Convertește un număr în text, folosind formatul de monedă \$ (dolar).
- **FIXED** - Formatează un număr ca text cu un număr fix de zecimale
- **TEXT** - Formatează un număr și îl convertește în text
- **VALUE** - Convertește un argument text într-un număr

Funcții de text sunt folosite la operațiuni de text asupra datelor care conțin text:

- **REPLACE** - Înlocuiește caracterele din text
- **SUBSTITUTE** - Înlocuiește textul nou cu textul vechi într-un șir de text
- **CONCATENATE** - Unește mai multe elemente de text într-un singur element de text
- **CONCAT** - Combină textul din mai multe intervale și/sau șiruri de caractere, dar nu furnizează argumentele delimitare sau IgnoreEmpty.
- **TEXTJOIN** - Combină textul din mai multe intervale și/sau șiruri de caractere și include un delimitator pe care îl specificați între fiecare valoare de text care va fi combinată. Dacă delimitatorul este un șir de text gol, această funcție va concatena efectiv intervalele.
- **REPT** - Repetă textul de un anumit număr de ori

9. CURĂȚAREA DATELOR CARE CONȚIN VALORI DATE

Diversitatea surselor de obținere a datelor brute pot să determine diferite formate pentru date și valori. În continuare vom prezenta metode de prelucrare datele care conțin valori ale datei, precum conversia datelor în format serial – format lună-zi-an și invers, extragerea anului, lunii, zilei din dată, calcularea numărului de ani, luni și zile între două date și așa mai departe.

9.1 Formate de date

Microsoft Excel acceptă valorile date în două tipuri:

- Format Serial
- În diferite formate An-Lună-Zi

Putem să transformăm:

- O dată în format serial la o dată în format an-lună-zi
- O dată în format an-lună-zi la o dată în format serial

Data în format serial

Data în format serial este reprezentată de un număr întreg pozitiv care reprezintă numărul de zile dintre Data curentă și 1 ianuarie 2019. Atât Data curentă, cât și 1 ianuarie 2019 sunt incluse în numărare. De exemplu, 43341 este o dată care reprezintă 29/08/2018.

Data în formatele lună-zi-an

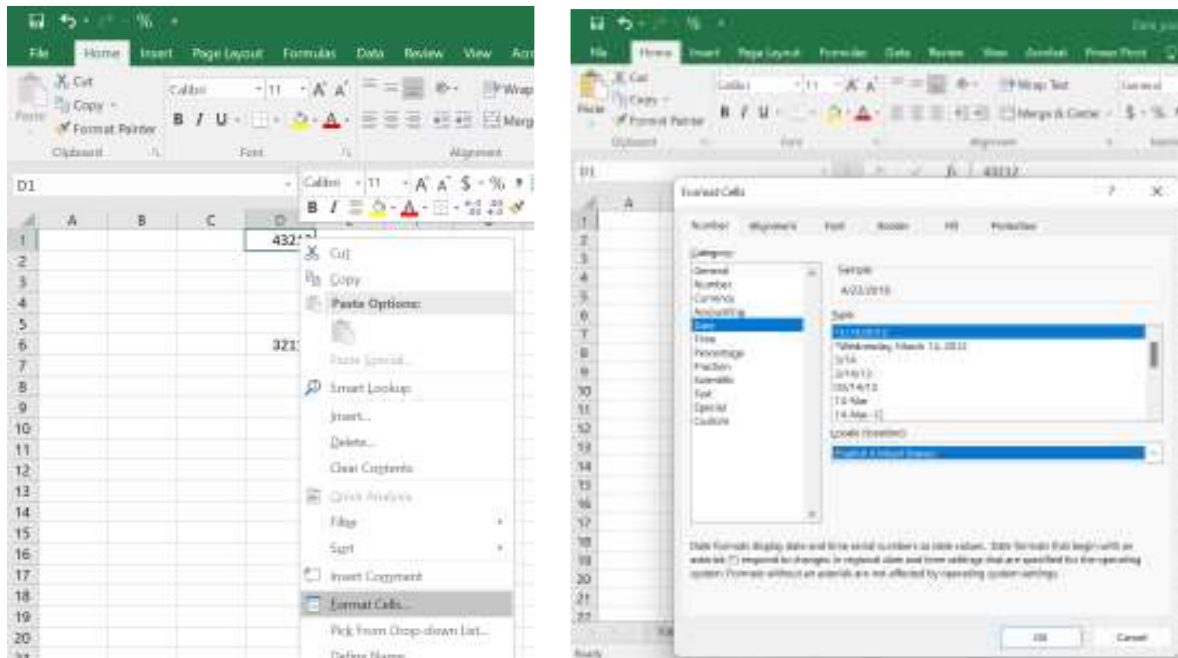
Excel acceptă diferite formate de dată în funcție de localitatea (locația) pe care o alegeți. Prin urmare, trebuie să determinăm mai întâi compatibilitatea formatelor de date pe care le deținem și analiza datelor la îndemână. Anumite formate de dată sunt prefixate cu *(asterisc), asta înseamnă că:

- Formatele de dată care încep cu *(asterisc) răspund la modificări ale setărilor regionale de dată și oră care sunt specificate pentru sistemul de operare.
- Formatele de date fără *(asterisc) nu sunt afectate de setările sistemului de operare.

9.2 Conversia datelor în format serial în format Lună-Zi-An

Conversia datelor din formatul Serial în formatul Lună-Zi-An, se realizează în felul următor:

Selectăm data (numărul 43212), clic dreapta pe număr, clic pe Format Cells după care clic pe **Number** din caseta de dialog **Format cells**. La **Category** selectăm **Date**, alegem un format tip dată și locația după care **OK**. În câmpul **Sample** putem vizualiza rezultatul conversiei datei de tip serial 43212 în format Lună-Zi-An, 04/22/2018.



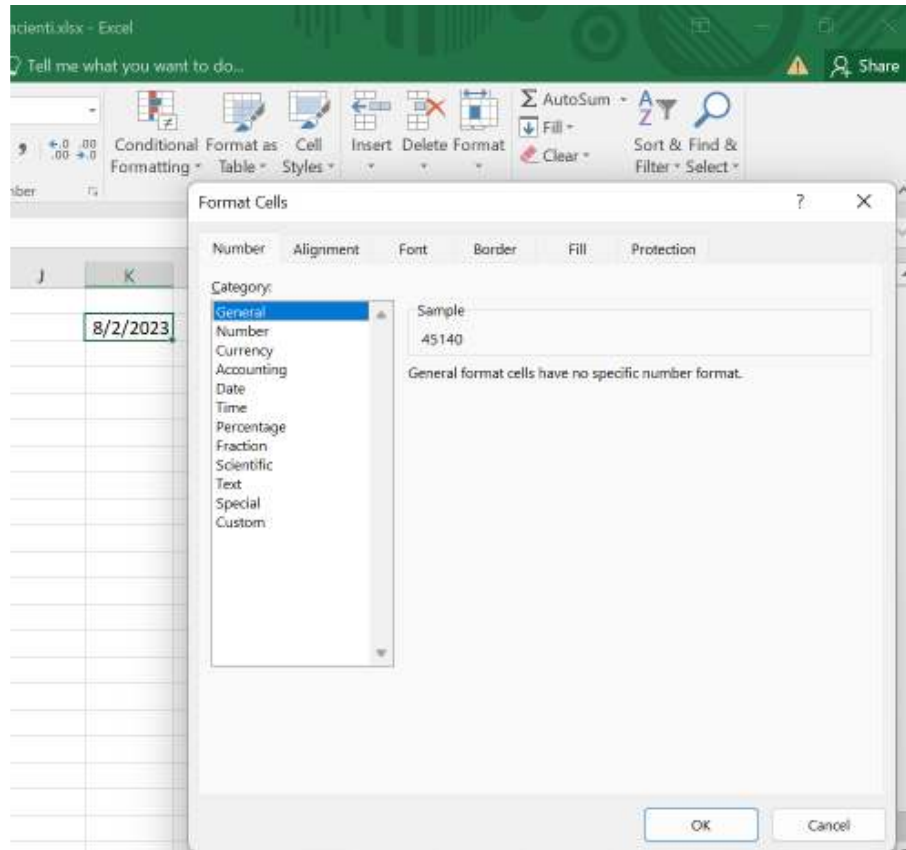
9.3 Conversia datelor în format lună-zi-an în format serial

Conversia datelor din format Lună-Zi-An în format Serial se poate efectua în două maniere:

- Prin utilizarea casetei de dialog **Format Cells**
- Prin folosirea funcției Excel **DATEVALUE**

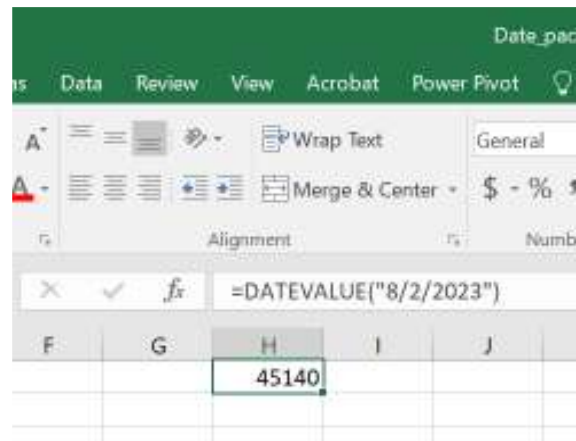
➤ **Caseta de dialog Format Cells**

Selectăm data 8/2/2023, clic dreapta pe dată, clic pe Format Cells, la **Category** selectăm **General**. În câmpul **Sample** putem vizualiza rezultatul conversiei datei Zi-Lună-An, 8/2/2023 în format Serial 45140.



- **Utilizarea funcției Excel DATEVALUE**

Funcția Excel DATEVALUE o folosim la realizarea conversiei unui format de dată în număr serial. De reținut! Trebuie data trebuie introdusă între caracterele " " (ghilimele), =DATEVALUE ("8/2/2023") are ca rezultat 45140.



9.4 Calcularea datei curente

Data curentă o putem calcula cu ajutorul funcției TODAY.

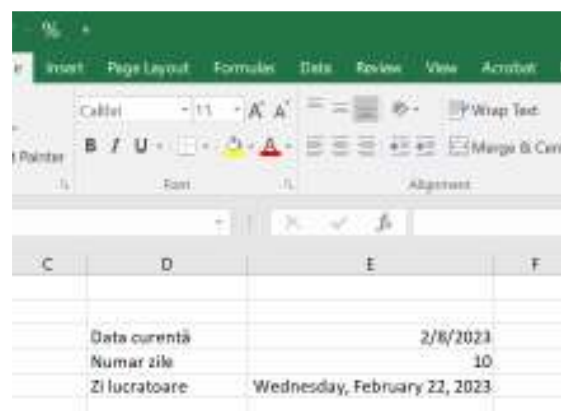
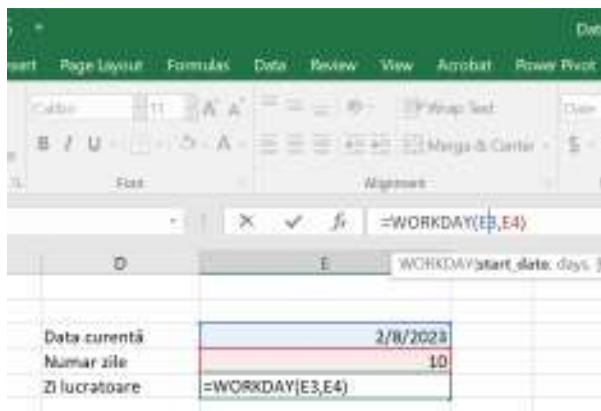


9.5 Identificarea unei zile lucrătoare după zile specificate

Putem utiliza funcțiile Excel **WORKDAY** și **WORKDAY.INTL** pentru identificarea unei zile lucrătoare după zile specificate.

- **WORKDAY** - Returnează numărul numărul de serie al zilelor lucrătoare ale datei dinainte sau după o anumită
- **WORKDAY.INTL** - Returnează numărul de serie al datei dinaintea sau după un anumit număr de zile lucrătoare folosind parametrii pentru a indica care și câte zile sunt zile de weekend.

Un exemplu ar fi să calculăm următoarea zi lucrătoare, peste 10 zile începând cu data de astăzi.



9.6 Extragerea Anului, Lună, Zi din Data

Putem extrage din fiecare dată dintr-o listă de date, ziua, luna și anul corespunzătoare folosind funcțiile excel **DAY**, **MONTH** și **YEAR**.

Avem următoarele date din figura de mai jos, din fiecare poziție putem extrage ziua, luna și anul.

	A	B	C	D	E	F	G
1		Marcaj de timp	Ziua	Luna	An		
2		11/18/2022 13:24:43	18	11	2022		
3		11/10/2022 13:28:10	10	11	2022		
4		11/22/2022 13:33:10	22	11	2022		
5		11/18/2022 13:58:27	18	11	2022		
6		11/20/2022 14:03:51	20	11	2022		
7		11/18/2022 14:04:13	18	11	2022		
8		11/18/2022 14:06:11	18	11	2022		
9		11/21/2022 14:07:38	21	11	2022		
10		11/18/2022 14:10:15	18	11	2022		
11		11/18/2022 14:10:27	18	11	2022		
12		11/18/2022 14:12:14	18	11	2022		
13		11/11/2022 14:14:23	11	11	2022		
14		11/12/2022 14:16:51	12	11	2022		
15		11/18/2022 14:23:13	18	11	2022		
16		11/18/2022 14:23:16	18	11	2022		

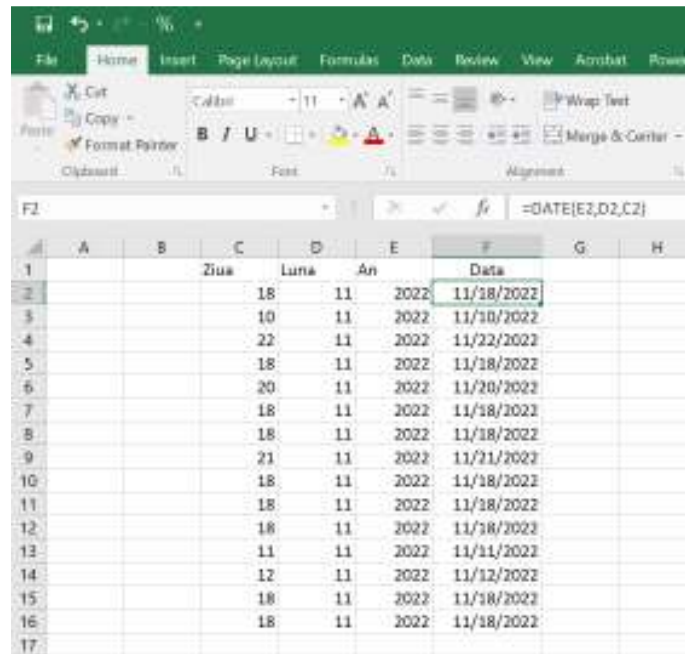
9.7 Extragerea zilei săptămânii din date

Cu ajutorul funcției **WEEKDAY** se poate extrage din fiecare dată dintr-o listă de date, ziua corespunzătoare a săptămânii.

	A	B	C	D	E	F
1		Marcaj de timp	Ziua	Luna	An	Ziua săptămânii
2		11/18/2022 13:24:43	18	11	2022	6
3		11/10/2022 13:28:10	10	11	2022	5
4		11/22/2022 13:33:10	22	11	2022	3
5		11/18/2022 13:58:27	18	11	2022	6
6		11/20/2022 14:03:51	20	11	2022	1
7		11/18/2022 14:04:13	18	11	2022	6
8		11/18/2022 14:06:11	18	11	2022	6
9		11/21/2022 14:07:38	21	11	2022	2
10		11/18/2022 14:10:15	18	11	2022	6
11		11/18/2022 14:10:27	18	11	2022	6
12		11/18/2022 14:12:14	18	11	2022	6
13		11/11/2022 14:14:23	11	11	2022	6
14		11/12/2022 14:16:51	12	11	2022	7
15		11/18/2022 14:23:13	18	11	2022	6
16		11/18/2022 14:23:16	18	11	2022	6

9.8 Obținerea datei din an, lună și zi

Cu ajutorul funcției DATE putem obține valorile datei, dacă ziua, luna și anul sunt scrise în coloane separate.



	A	B	C	D	E	F	G	H
1			Ziua	Luna	An	Data		
2			18	11	2022	11/18/2022		
3			10	11	2022	11/10/2022		
4			22	11	2022	11/22/2022		
5			18	11	2022	11/18/2022		
6			20	11	2022	11/20/2022		
7			18	11	2022	11/18/2022		
8			18	11	2022	11/18/2022		
9			21	11	2022	11/21/2022		
10			18	11	2022	11/18/2022		
11			18	11	2022	11/18/2022		
12			18	11	2022	11/18/2022		
13			11	11	2022	11/11/2022		
14			12	11	2022	11/12/2022		
15			18	11	2022	11/18/2022		
16			18	11	2022	11/18/2022		
17								

9.9 Calcularea anilor, lunilor și zilelor dintre două date

Poate fi necesar să calculăm timpul petrecut de la o dată la altă dată. Este posibil să avem nevoie de aceste informații sub formă de ani, luni și zile. Ca și exemplu, vom calcula vârsta curentă a unei persoane, prin operația de diferență dintre data nașterii și ziua de azi. Apelăm la funcțiile Excel **DATEDIF**, **TODAY** și **CONCATENATE**.

10. LUCRUL CU VALORI DE TIMP

Excel acceptă valorile timpului în două moduri:

- Format Serial
- În diverse formate oră-minut-secundă

Putem converti:

- Ora în format serial până la ora în format oră-minut-secundă
- Ora în format oră-minut-secundă până la ora în format serial

10.1 Ora în format serial

Ora în format serial este un număr pozitiv care reprezintă Ora ca o fracțiune dintr-o zi de 24 de ore, punctul de plecare fiind miezul nopții.

În Excel, putem combina Data și Ora în aceeași celulă. Numărul de serie este numărul de zile după 1 ianuarie 1900 și fracțiunea de timp asociată cu ora dată.

10.2 Ora în format oră-minut-secundă

Excel ne permite să specificăm ora în format oră-minut-secundă cu două puncte (:) după oră și alte două puncte înaintea secundelor. De exemplu, 7:40 AM, 7:40 PM sau doar 7:40 folosind formatul de 12 ore sau ca 7:40, 19:40 în format de 24 de ore. Ora 7:40:40 AM reprezintă 7 ore, 40 de minute și 40 de secunde.

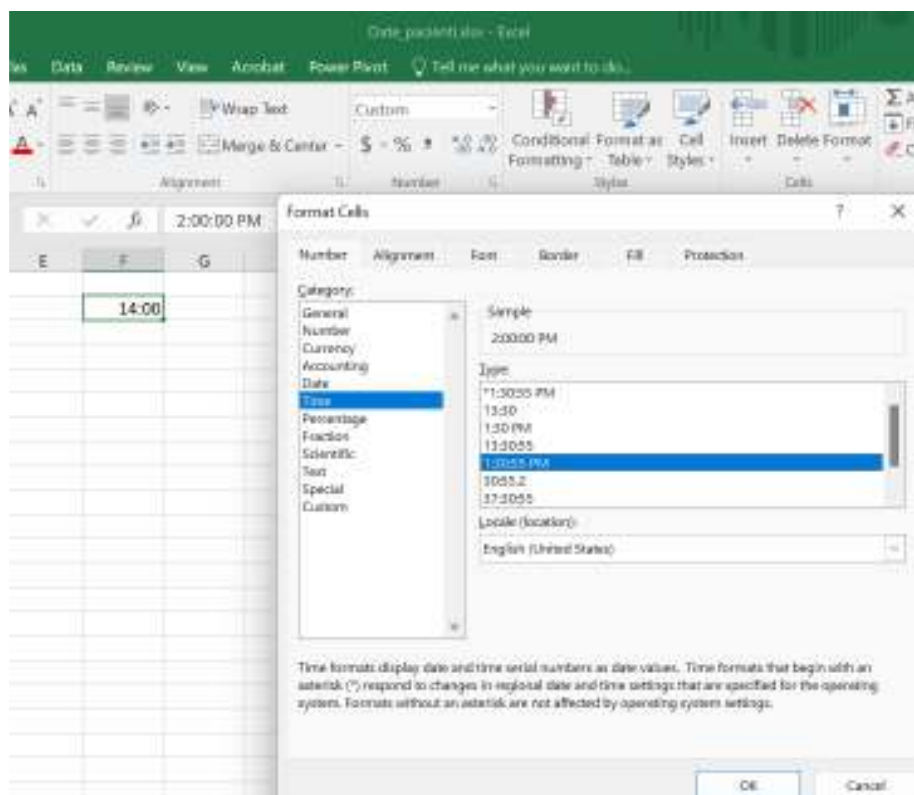
De asemenea, putem specifica data și ora împreună. De exemplu, dacă tastăm 8 februarie 2023 9:25 într-o celulă, aceasta va fi afișată ca 2/8/2023 9:25 și reprezintă 2/8/2023 9:25:00 AM.

Microsoft Excel acceptă diferite formate de oră în funcție de localitatea (locația) pe care o alegem. Prin urmare, trebuie să determinați mai întâi compatibilitatea formatelor de timp și analiza datelor la îndemână.

10.3 Conversia timpilor în format serial în format oră-minut-secundă

Pentru a converti formatul de timp serial în format oră-min-sec, vom realiza un exemplu.

Intr-o celulă am tastat ora 14:00, după care clic dreapta pe celulă, alegem **Format Cells – Number – Category – Time**, și din lista derulantă alegem un format și o locație pentru oră, care poate fi vizualizat în câmpul **Sample**, după care clic **OK**.



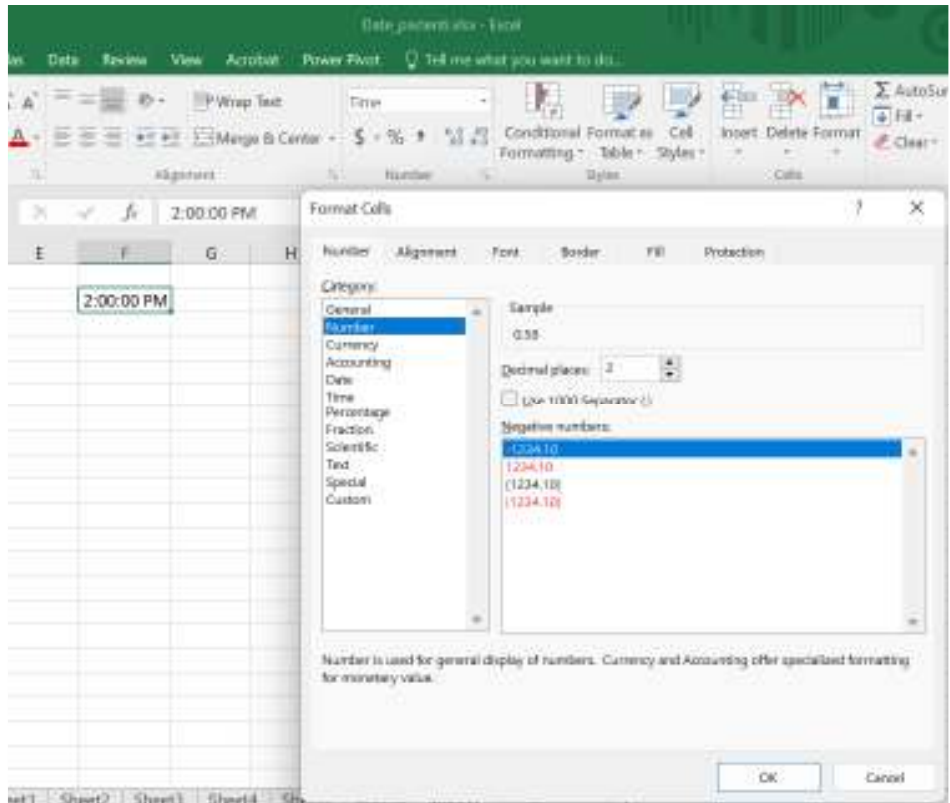
10.4 Conversia timpilor în format oră-minut-secundă în format serial

Putem converti ora în format oră-minut-secundă în format serial în două maniere:

- Cu ajutorul casetei de dialog **Format Cells**
- Cu ajutorul funcției Excel **TIMEVALUE**

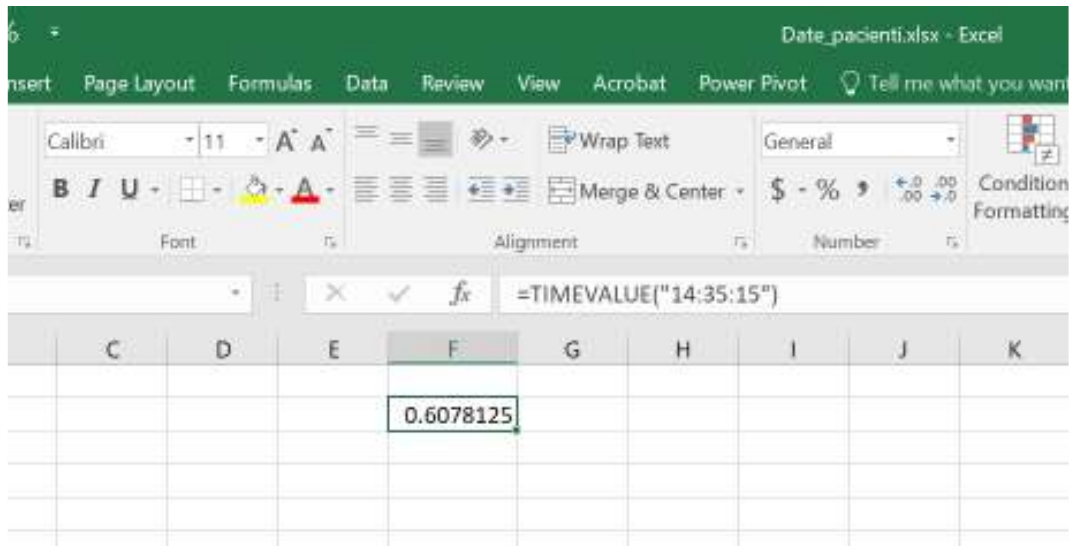
Conversia cu ajutorul casetei de dialog Format Cells

Intr-o celulă am tastat ora 2:00:00 PM, după care clic dreapta pe celulă, alegem **Format Cells – Number – Category – General**, iar ora de tip serial care poate fi vizualizată în câmpul **Sample**, după care clic **OK**.



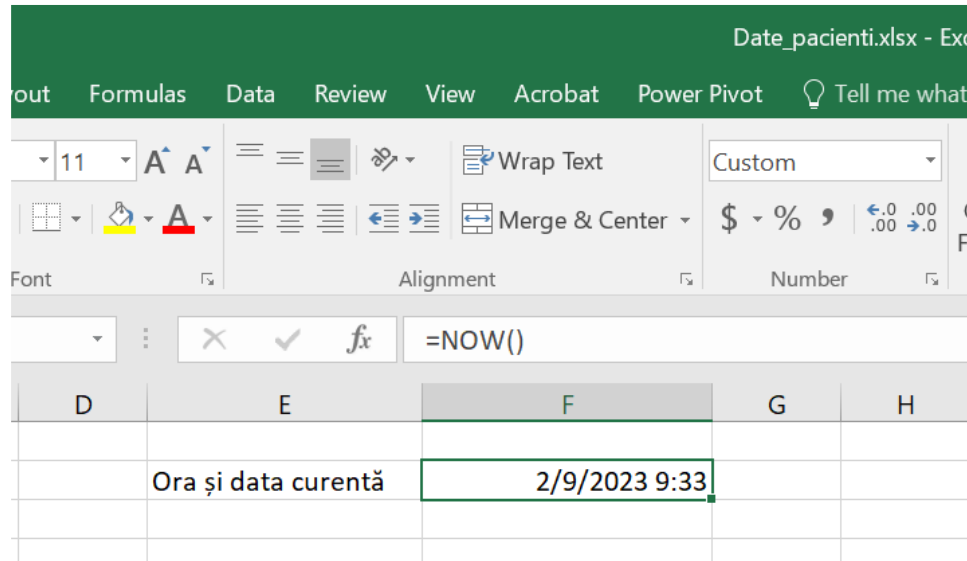
Conversia cu ajutorul funcției Excel TIMEVALUE

Un exemplu este următorul: TIMEVALUE ("14:35:15") are ca rezultat 0.6078125, de reținut inserarea caracterelor " " (ghilimele).



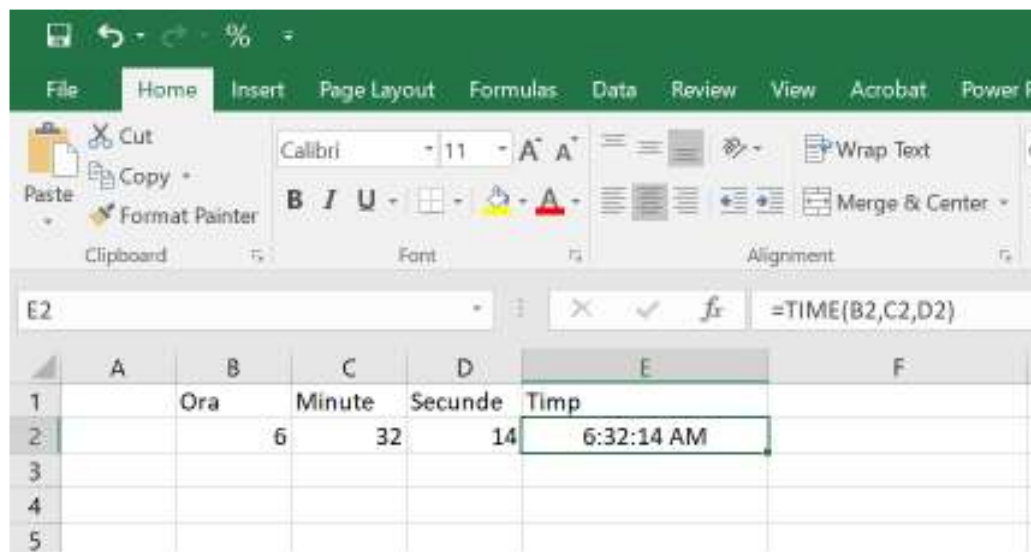
Obținerea orei curente

Pentru calcularea orei curente folosim funcția Excel **NOW()**. Rezultatul reflectă data și ora la care este utilizată funcția.



10.5 Obținerea timpului din oră, minut și secundă

În cazul în care datele noastre referitoare la timp, sunt importate în coloane separate, putem uni aceste informații pentru a afla timpul, cu ajutorul funcției **TIME**.



10.6 Extragerea orei, minutelor și secundelor din timp

În funcție de analiza care o realizăm, uneori este posibil să avem nevoie de informații despre ore, minute și secunde în mod separat. Putem utiliza funcțiile Excel **hour**, **MINUT**, **SECOND** pentru a obține valorile timpului în coloane separate.

Date_pacienti.xlsx - Excel

ulas Data Review View Acrobat Power Pivot Tell me what you want to do...

Wrap Text Merge & Center

General \$ - % +.0 .00 -0.0

Alignment Number Conditional Form Formatting Tab Styles

\times \checkmark f_x =HOUR(E2)

E	F	G	H	I	J
Time	Ora	Minute	Secunde		
6:32:14 AM	6	32	14		

11. SORTAREA DATELOR

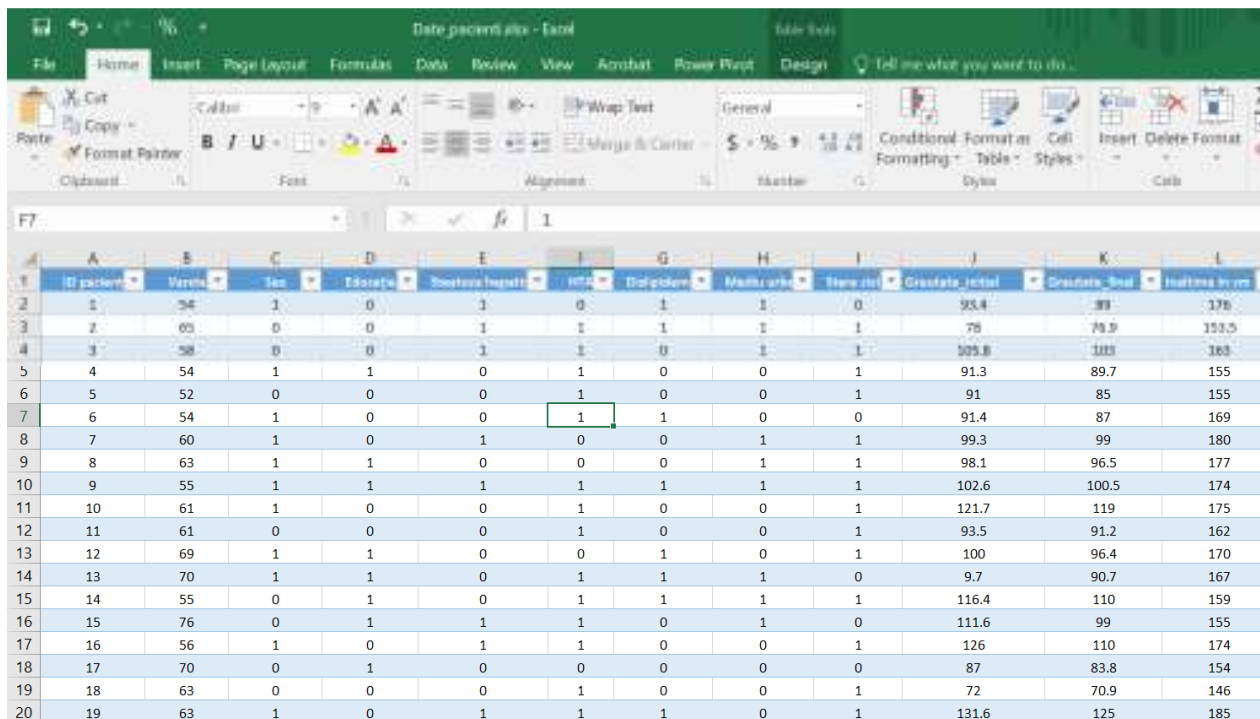
O etapă importantă în procesul de analiză și prelucrare a datelor o reprezintă operația de sortare. Prin sortare putem vizualiza o listă în ordine alfabetică, sau de la cel mai mare la cel mai mic și invers, în ordine crescătoare sau descrescătoare, sau putem realiza o sortare personalizată în funcție de necesități. Sortarea se poate efectua după coloane sau după rânduri.

Criterii de sortare a datelor într-una sau mai multe coloane în funcție de:

- text (de la A la Z sau de la Z la A)
- numere (de la cel mai mic la cel mai mare, crescător sau de la cel mai mare la cel mai mic, descrescător)
- date și ore (de la cel mai vechi la cel mai nou și de la cel mai nou la cel mai vechi)
- listă personalizată (de exemplu, mare, medie și mică)
- format, inclusiv culoarea celulei, culoarea fontului sau setul de pictograme.

11.1 Sortarea după numere

Putem sorta un tabel folosind o coloană care conține text. În tabelul din imaginea de mai jos sunt colectate informații despre pacienții internați într-o clinică privată.



	A	B	C	D	E	F	G	H	I	J	K	L
	ID pacient	Varsta	Sex	Etapa	Starea fizică	HTA	Diabet	Mănușă	Stare psi	Gradul de recuperare	Gradul de stres	Înălțimea în cm
2	1	54	1	0	1	0	1	1	0	95.4	89	176
3	2	65	0	0	1	1	1	1	1	76	79.9	153.5
4	3	58	0	0	1	1	0	1	1	105.8	103	163
5	4	54	1	1	0	1	0	0	1	91.3	89.7	155
6	5	52	0	0	0	1	0	0	1	91	85	155
7	6	54	1	0	0	1	1	0	0	91.4	87	169
8	7	60	1	0	1	0	0	1	1	99.3	99	180
9	8	63	1	1	0	0	0	1	1	98.1	96.5	177
10	9	55	1	1	1	1	1	1	1	102.6	100.5	174
11	10	61	1	0	0	1	0	0	1	121.7	119	175
12	11	61	0	0	0	1	0	0	1	93.5	91.2	162
13	12	69	1	1	0	0	1	0	1	100	96.4	170
14	13	70	1	1	0	1	1	1	0	9.7	90.7	167
15	14	55	0	1	0	1	1	1	1	116.4	110	159
16	15	76	0	1	1	1	0	1	0	111.6	99	155
17	16	56	1	0	1	1	0	0	1	126	110	174
18	17	70	0	1	0	0	0	0	0	87	83.8	154
19	18	63	0	0	0	1	0	0	1	72	70.9	146
20	19	63	1	0	1	1	1	0	1	131.6	125	185

Ca și exemplu, dorim să sortăm coloana ID_pacient, să se afișeze datele în ordine descrescătoare (de la mic la mare).

Pentru a realiza sortarea, facem clic în coloana de interes, iar din meniul **HOME**, secțiunea **Editing**, vom selecta **Sort& Filter**, și bifăm sortare de la Z to A (de la mare la mic).

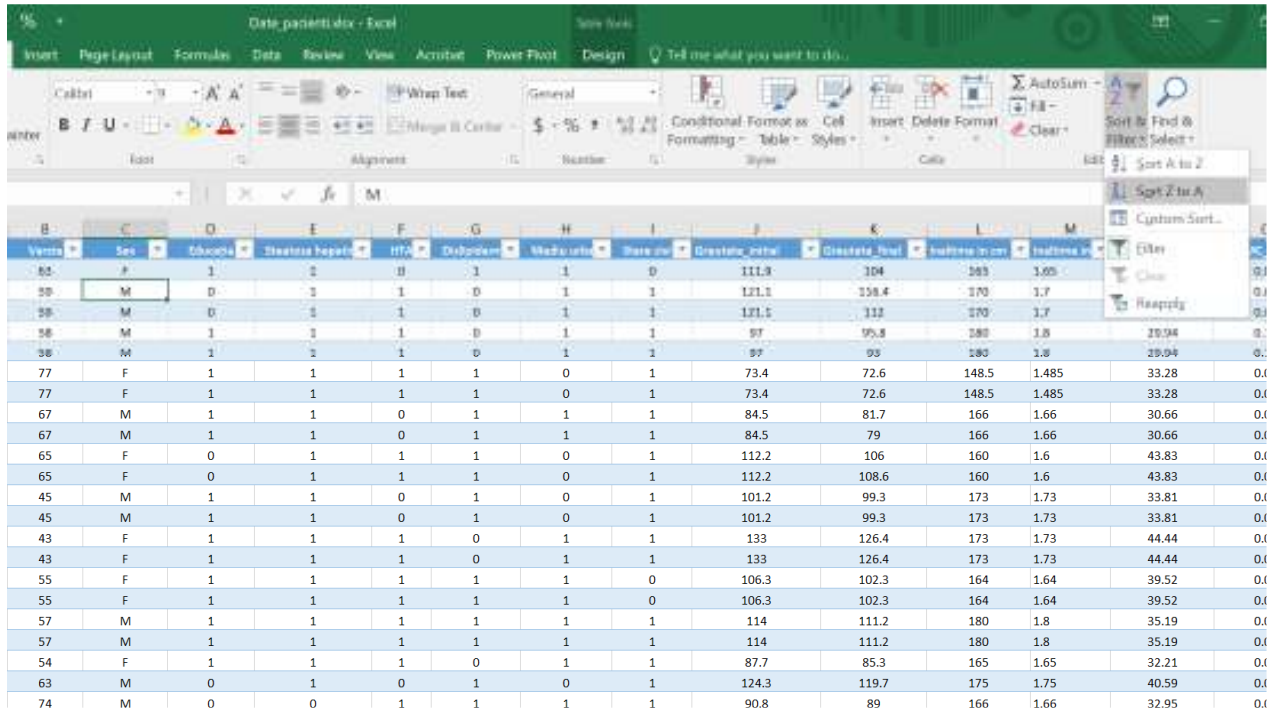
ID_pacient	Nume	Sex	Educatie	Stare_de_sanatate	IMC	Dolgiu	Mușchi_ani	Stare_cel	Gratificat	Gratificat_Sex		
1	84	1	0	1	1	1	1	0	84.4	85	216	1.16
2	86	0	0	1	1	1	1	1	18	18.9	1019	1.019
3	88	0	0	1	1	0	1	1	101.8	101	85	1.01
4	84	1	1	0	1	0	0	1	81.6	80.7	84	1.06
5	52	0	0	0	1	0	0	1	91	85	155	1.55
6	54	1	0	0	1	1	0	0	91.4	87	169	1.69
7	60	1	0	1	0	0	1	1	99.3	99	180	1.8
8	63	1	1	0	0	0	1	1	98.1	96.5	177	1.77
9	55	1	1	1	1	1	1	1	102.6	100.5	174	1.74
10	61	1	0	0	1	0	0	1	121.7	119	175	1.75
11	61	0	0	0	1	0	0	1	93.5	91.2	162	1.62
12	69	1	1	0	0	1	0	1	100	96.4	170	1.7
13	70	1	1	0	1	1	1	0	9.7	90.7	167	1.67
14	55	0	1	0	1	1	1	1	116.4	110	159	1.59
15	76	0	1	1	1	0	1	0	111.6	99	155	1.55
16	56	1	0	1	1	0	0	1	126	110	174	1.74
17	70	0	1	0	0	0	0	0	87	83.8	154	1.54
18	63	0	0	1	1	0	0	1	72	70.9	146	1.46
19	63	1	0	1	1	1	0	1	131.6	125	185	1.85
20	63	1	0	1	1	1	1	1	131.6	125	185	1.85
21	70	0	1	1	1	1	1	0	86	84.5	156	1.56

Rezultatul se poate vizualiza în figura de mai jos, unde putem vedea coloana ID pacient are valorile ordonate descrescător. Același lucru se poate efectua și pentru o listă text.

ID_pacient	Nume	Sex	Educatie	Stare_de_sanatate	IMC	Dolgiu	Mușchi_ani	Stare_cel	Gratificat	Gratificat_Sex
96	82	0	1	1	0	1	1	0	121.9	104
95	59	1	0	1	1	0	1	1	121.1	116.4
95	59	1	0	1	1	0	1	1	121.1	112
94	58	1	1	1	1	0	1	1	97	85.8
94	58	1	1	1	1	0	1	1	97	81
93	77	0	1	1	1	1	0	1	73.4	72.6
93	77	0	1	1	1	1	0	1	73.4	72.6
92	87	1	1	1	0	1	1	1	84.1	81.7
92	87	1	1	1	0	1	1	1	84.1	79
91	85	0	0	1	1	1	0	1	112.2	100
91	85	0	0	1	1	1	0	1	112.2	106.6
90	45	1	1	1	0	1	0	1	101.2	99.1
90	45	1	1	1	0	1	0	1	101.2	99.1
89	43	0	1	1	1	0	1	1	131	126.4
89	43	0	1	1	1	0	1	1	131	126.4

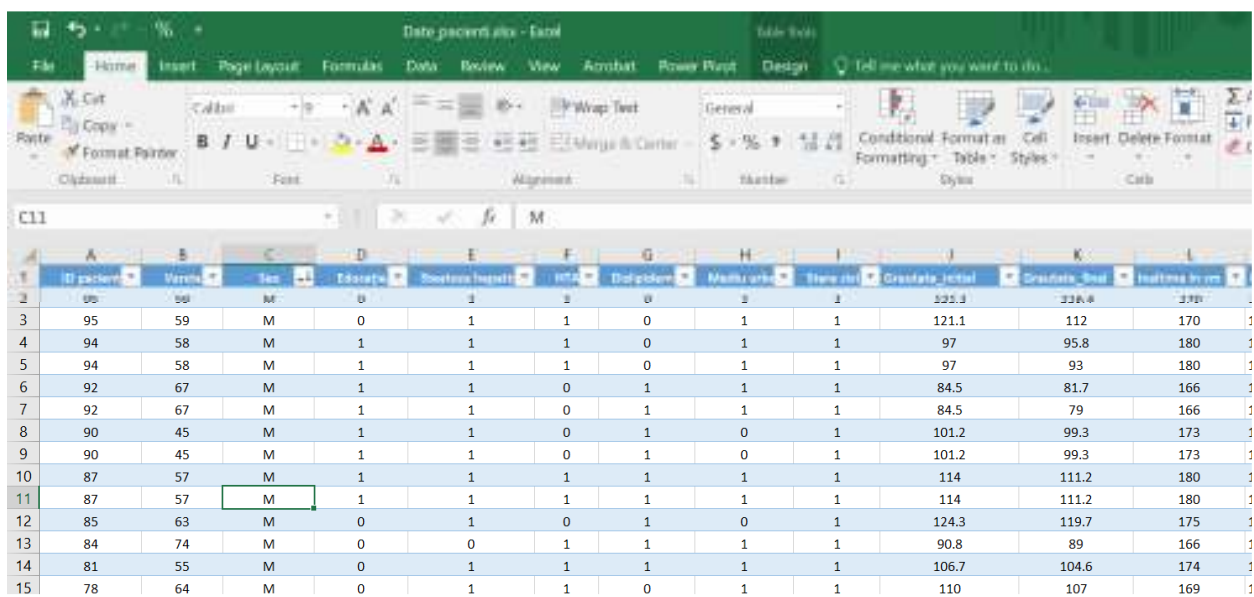
11.2 Sortarea după text

Dorim să sortăm coloana sex, ca sexul M să apară primul în listă. Pentru a realiza sortarea, facem clic în coloana de interes, iar din meniul **HOME**, secțiunea **Editing**, vom selecta **Sort & Filter**, și bifăm sortare de la Z to A.



Vârsta	Sex	Cholesterol	Tratata hepat	HFA	Doliprom	Medicatie	Stare de	Diastolic	Diastolic	Diastolic	Diastolic	Diastolic	Diastolic
95	F	1	1	1	1	1	0	111.3	104	183	1.05		
90	M	0	1	1	0	1	1	121.1	138.4	170	1.7		
98	M	0	1	1	0	1	1	121.1	112	170	1.7		
98	M	1	1	1	0	1	1	97	95.8	180	1.8	20.94	
98	M	1	1	1	0	1	1	97	95	180	1.8	20.94	
77	F	1	1	1	1	0	1	73.4	72.6	148.5	1.485	33.28	
77	F	1	1	1	1	0	1	73.4	72.6	148.5	1.485	33.28	
67	M	1	1	0	1	1	1	84.5	81.7	166	1.66	30.66	
67	M	1	1	0	1	1	1	84.5	79	166	1.66	30.66	
65	F	0	1	1	1	0	1	112.2	106	160	1.6	43.83	
65	F	0	1	1	1	0	1	112.2	108.6	160	1.6	43.83	
45	M	1	1	0	1	0	1	101.2	99.3	173	1.73	33.81	
45	M	1	1	0	1	0	1	101.2	99.3	173	1.73	33.81	
43	F	1	1	1	0	1	1	133	126.4	173	1.73	44.44	
43	F	1	1	1	0	1	1	133	126.4	173	1.73	44.44	
55	F	1	1	1	1	1	0	106.3	102.3	164	1.64	39.52	
55	F	1	1	1	1	1	0	106.3	102.3	164	1.64	39.52	
57	M	1	1	1	1	1	1	114	111.2	180	1.8	35.19	
57	M	1	1	1	1	1	1	114	111.2	180	1.8	35.19	
54	F	1	1	1	0	1	1	87.7	85.3	165	1.65	32.21	
63	M	0	1	0	1	0	1	124.3	119.7	175	1.75	40.59	
74	M	0	0	1	1	1	1	90.8	89	166	1.66	32.95	

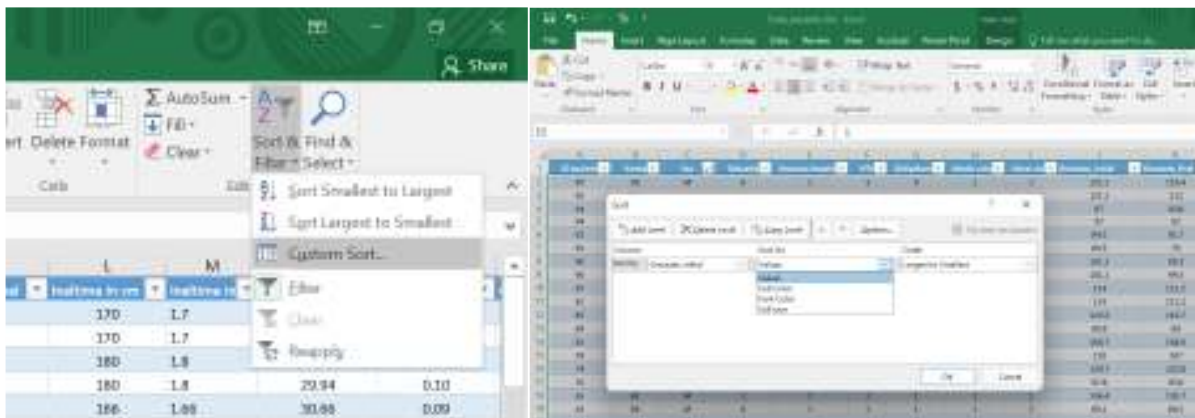
Rezultatul se poate vizualiza în figura de mai jos, unde putem vedea coloana Sex are valorile ordonate alfabetică descrescătoare, de la M la F.



ID patient	Vârsta	Sex	Cholesterol	Tratata hepat	HFA	Doliprom	Medicatie	Stare de	Diastolic	Diastolic	Diastolic	Diastolic	Diastolic
95	59	M	0	1	1	0	1	1	1	1	121.1	138.4	170
94	58	M	1	1	1	0	1	1	1	1	97	95.8	180
94	58	M	1	1	1	0	1	1	1	1	97	93	180
92	67	M	1	1	0	1	1	1	1	1	84.5	81.7	166
92	67	M	1	1	1	0	1	1	1	1	84.5	79	166
90	45	M	1	1	0	1	0	1	1	1	101.2	99.3	173
90	45	M	1	1	0	1	0	1	1	1	101.2	99.3	173
87	57	M	1	1	1	1	1	1	1	1	114	111.2	180
87	57	M	1	1	1	1	1	1	1	1	114	111.2	180
85	63	M	0	1	0	1	0	1	1	1	124.3	119.7	175
84	74	M	0	0	1	1	1	1	1	1	90.8	89	166
81	55	M	0	1	1	1	1	1	1	1	106.7	104.6	174
78	64	M	0	1	1	0	1	1	1	1	110	107	169

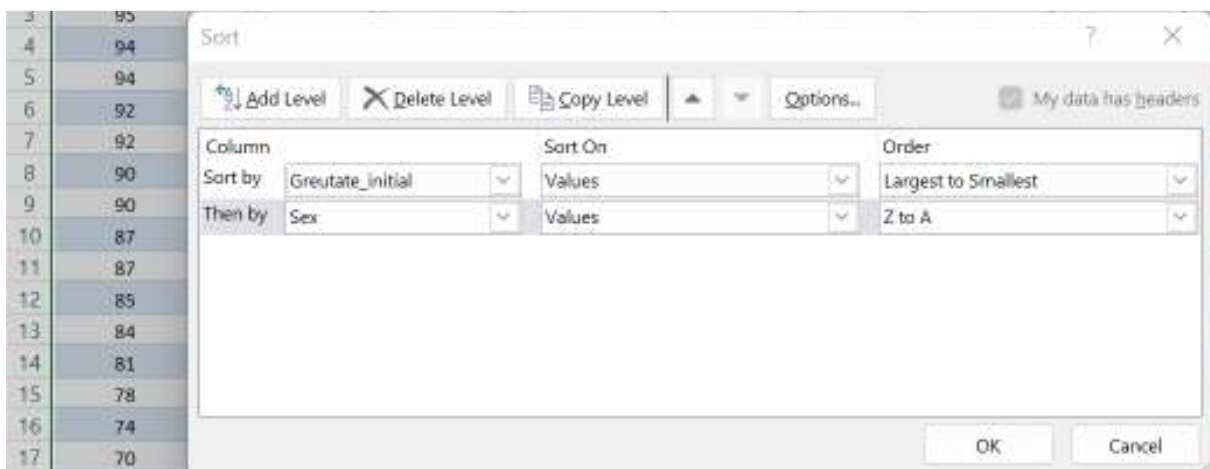
11.3 Sortare personalizată

Pentru a realiza o sortare personalizată accesăm meniul **HOME**, secțiunea **Editing**, vom selecta **Sort & Filter**, și bifăm **Custom Sort**.



Dorim să facem o sortare dublă, în sensul că dorim să se afișeze pacienții de sex masculin care au greutatea cea mai mare prima dată, și după aceea să apară pacienții de sex feminin.

După ce am realizat pașii de mai sus, se deschide **Caseta de dialog Sort**, iar la **Sort By**, selectăm din lista derulantă coloanele care dorim să le sortăm, dacă avem mai multe coloane bifăm **Add level**, de câte ori dorim să adăugăm o coloană nouă la sortare, la **Sort on**, bifăm în funcție de ce anume să se facă sortarea iar la **Order**, selectăm ordinea de apariție a datelor.



Și astfel avem realizată sortarea personalizată, în figura de mai jos.

ID person	Varos	Sex	Edzaine	Startaps nepars	int	Disiplins	Mudu arns	Stava drs	Grasats_sistal	Grasats_base	Valstima in on
31	52	F	0	0	1	0	0	1	143.5	140.7	160
89	43	F	1	1	1	0	1	1	135	136.4	173
89	43	F	1	1	1	0	1	1	135	136.4	173
19	63	M	0	1	1	1	0	1	231.6	125	285
79	61	F	0	1	1	1	1	1	130	125	157
41	48	M	0	1	0	0	1	1	127	123.4	180
16	56	M	0	1	1	0	0	1	126	110	174
45	61	F	0	0	1	1	1	0	124.8	123	163
74	70	M	0	1	1	0	1	1	124.7	120.8	178
21	63	M	0	0	1	0	1	1	124.4	121.9	175
85	63	M	0	1	0	1	0	1	124.3	119.7	175
10	61	M	0	0	1	0	0	1	121.7	119	175
95	59	M	0	1	1	0	1	1	121.1	116.4	170
95	59	M	0	1	1	0	1	1	121.1	112	170
34	40	F	0	1	0	1	1	0	120.6	117.2	164
69	52	F	0	1	1	1	1	1	120	114.9	194
32	71	M	0	1	1	1	0	1	119.6	116.3	165
14	55	F	1	0	1	1	1	1	116.4	110	159
64	67	F	1	0	1	1	1	0	116.1	114.9	173
87	57	M	1	1	1	1	1	1	114	111.2	180
87	57	M	1	1	1	1	1	1	114	111.2	180

12. FILTRAREA DATELOR

Pentru extragerea unor date care îndeplinesc anumite criterii definite dintr-un interval sau tabel, apelăm la operația de Filtrare. Filtrarea reprezintă o metodă rapidă și ușoară de afișare doar a informațiilor de care avem nevoie dintr-o bază de date de dimensiuni mari.

Metoda de filtrare se poate realiza după:

- Valori selectate
- Text filtre dacă coloana selectată conține text
- Date filtre dacă coloana selectată conține date
- Filtre de numere dacă coloana selectată conține numere
- Culoarea celulei dacă coloana selectată conține celule cu culoare
- Culoare font dacă coloana selectată conține font cu culoare
- Pictogramă celulă dacă coloana selectată conține pictograme pentru celulă
- Filtru avansat

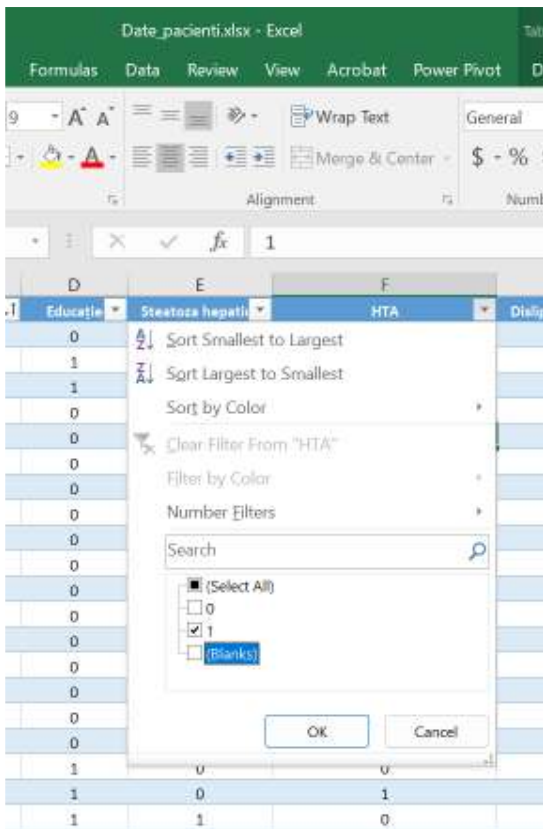
Dacă datele noastre sunt aranjate sub forma unui tabel în Excel, antetele coloanelor tabelului prezintă automat eticheta filtru, denumită AutoFilter. Filtrul automat este reprezentat de săgeata de lângă antetul coloanei. Fiecare filtru automat are opțiuni de filtrare în funcție de tipul de date pe care le deținem în coloana respectivă.

Dacă datele noastre sunt organizate sub formă de interval, putem eticheta fiecare coloană cu filtru accesând meniul Home – Editing – Sort&Filter – bifăm Filter, acest lucru va activa Filtrul automat pentru interval. Tot în acest mod se poate dezactiva opțiunea Filter.



12.1 Filtrarea după valorile selectate

În exemplul ce urmează dorim să facem o filtrare în care să se afișeze doar pacienții care prezintă HTA=1. Facem clic pe săgeata Filtru din dreptul antetului coloanei HTA, se va deschide caseta de dialog care prezintă opțiunile de filtrare, și debifăm 0, Blanks, după care clic pe OK. După ce s-a realizat filtrarea se poate observa, că lângă săgeata Filtru apare un simbol nou, o pâlnie, acest simbol ne arată că pe coloana respectivă exista un filtru activ.

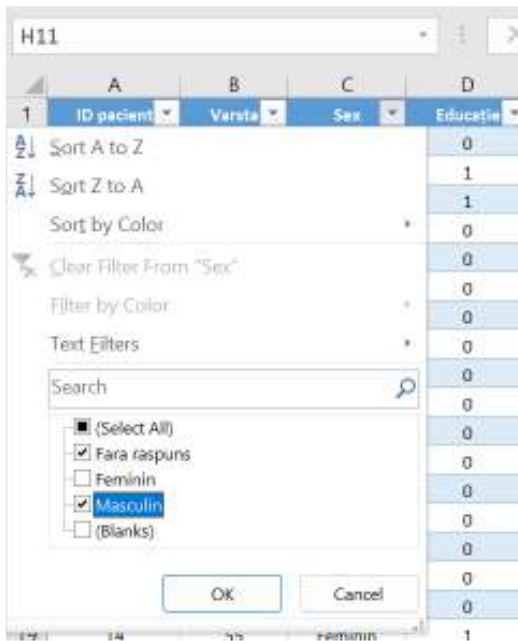


	F	G	H
	HTA	Dislipidem	Mediu urba
	1	0	0
	1	0	1
	1	0	1
	1	1	0
	1	1	1
	1	0	0
	1	1	1
	1	0	1
	1	0	0
	1	0	1
	1	1	1
	1	1	0
	1	1	1
	1	1	0
	1	1	0
	1	1	1
	1	0	1
	1	1	1
	1	0	1
	1	0	1

12.2 Filtrarea după text

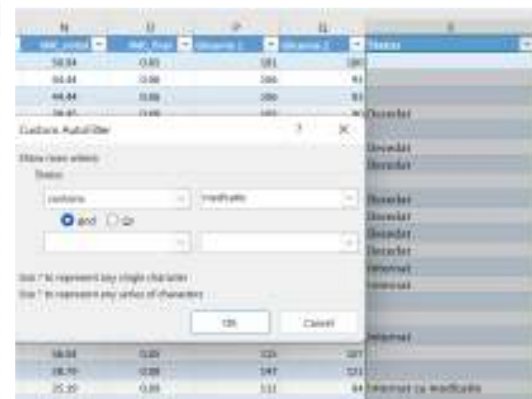
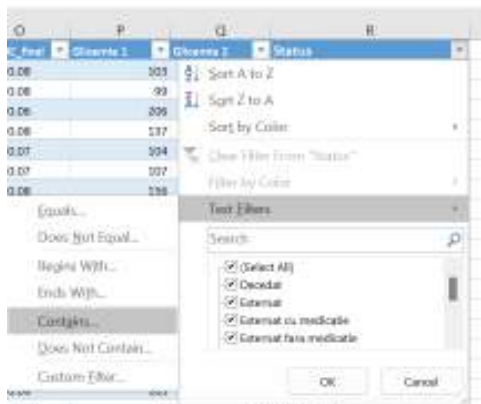
Dorim să filtrăm baza de date astfel încât la coloana Sex să se afișeze doar pacienții de sexmasculin și pacienții care nu au răspuns, deci să excludem pacienții de sex feminin.

Facem clic pe săgeata Filtru din dreptul antetului coloanei Sex, se va deschide caseta de dialog care prezintă opțiunile de filtrare, și debifăm **Select All**, după care bifăm doar opțiunile de interes, **Masculin** și **Fără răspuns**.



	A	B	C	D
1	ID pacient	Varsta	Sex	Educatie
5	10	63	Masculin	0
7	41	40	Masculin	0
8	18	36	Masculin	0
10	74	70	Masculin	0
11	21	63	Masculin	0
12	85	63	Masculin	0
13	10	61	Masculin	0
14	95	59	Fara raspuns	0
15	95	59	Fara raspuns	0
18	32	71	Fara raspuns	0
21	87	57	Fara raspuns	1
22	87	57	Fara raspuns	1
25	59	61	Fara raspuns	1
29	78	64	Fara raspuns	0
32	81	55	Fara raspuns	0
33	65	62	Fara raspuns	1
38	8	50	Fara raspuns	1
40	30	50	Masculin	1
41	90	40	Masculin	1
42	90	40	Masculin	1
43	12	69	Masculin	1
44	7	60	Masculin	0

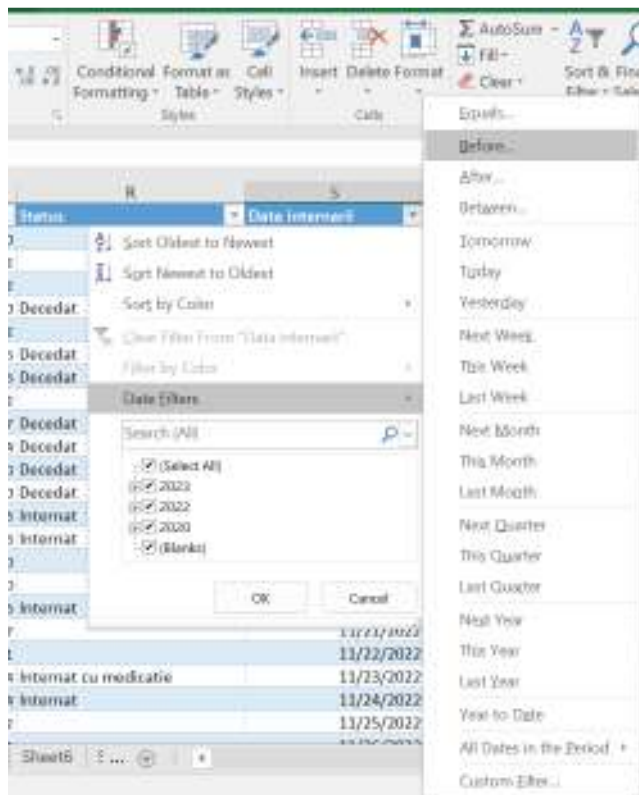
De asemenea Filtrare după text conține multe alte opțiuni, un alt exemplu este următorul. La coloana Status dorim să se afișeze doar pacienții care au primit sau primesc medicatie, indiferent de statusul lor. Pentru acest lucru, vom selecta **Text Filters – Contains**, iar în caseta de dialog **Custom AutoFilter**, la coloana Status vom tasta *medicatie*, clic pe OK. Vor fi afișate doar rândurile în care Titlul conține Medicatie.



	R	S
	Status	
84	Internat cu medicatie	
77	Externat fara medicatie	
131	Externat cu medicatie	
85	Fara detalii despre medicatie	
164	Fara detalii despre medicatie	

12.3 Filtrarea după dată

Dorim să realizăm o filtrare după data internării pacienților, și anume să se afișeze toți pacienții internați înainte de anul 11/18/2022. Pașii și rezultatul se pot vedea în imaginile de mai jos.



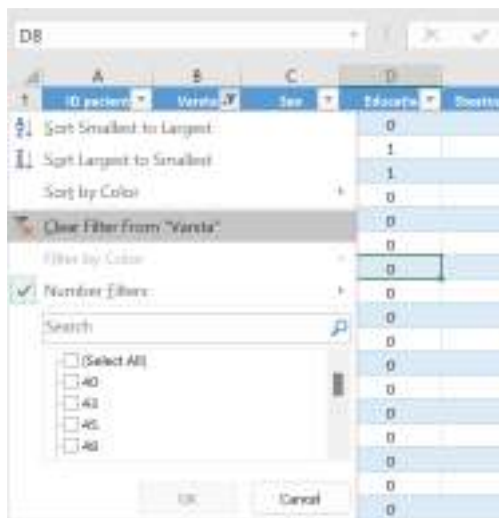
10.3 Filtrarea după numere

Dorim să vizualizăm doar pacienții cu vârsta mai mare de 35 ani. Pentru a realiza această filtrare facem clic pe săgeata de lângă antetul coloanei Vârsta. Din lista derulantă, facem clic **Number Filters**. Apar opțiunile **Number Filters**. Faceți clic pe **Greater than**. Se deschide caseta de dialog Filtru automat personalizat. Tastăm 35 în caseta de lângă **Greater than**.

12.4 Ștergerea filtrului

Eliminarea filtrelor este denumită Clear Filter în Excel. Putem să ștergem filtru aplicat dintr-o singură coloană sau toate filtrele aplicate în foaia de calcul, în mod simultan.

Pentru a elimina un filtru dintr-o anumită coloană, facem clic pe săgeata din antetul tabelului al acelei coloane. Din meniul drop-down, faceți clic pe **Clear Filter** from "Varsta".



Filtrul din coloana Varsta este îndepărtat. Iar pentru a elimina filtrarea din întreaga foaie de lucru, selectați **Clear** la grupul de editare **Sort&Filter** din meniul **Home**. Toate filtrele din foaia de lucru sunt eliminate simultan.

