**Academia de Studii Economice București**

****

**Proiect Pachete Software**

**Excel, SAS, SAS EG, Python**

**Facultatea de Cibernetică, Statistică și Informatică Economică**

**Studenți: Murgeanu Andrei Profesor coordonator: Ștefan Preda**

**Ivasișin Andrei Anul 3, Semestrul II**

**Grupa: 1062 Specializarea Cibernetică Economică**

Prezentarea firmei

Colgate-Palmolive este o companie înființată în anul 1992, în România. Este cunoscută la nivel mondial și are o notorietate foarte mare, fiind lider de piață (cel puțin în România) la vânzarea de pastă de dinți.

Cele mai importante brand-uri sunt: Colgate, Palmolive, Protex, ajax, Axion, Mennen Speed Stick, Lady Speed Stick. Un brand nou, adoptat în anul 2019 este Elmex, o nouă categorie de pastă de dinți.

În România compania deține cc. 50 de angajați. Firma este recunoscută ca făcând parte din “Cele mai bune 100 de branduri globale”.

Valorile principale ale companiei sunt susținute prin programe de știință & inovație, reducerea consumului de apă și durabilitate.

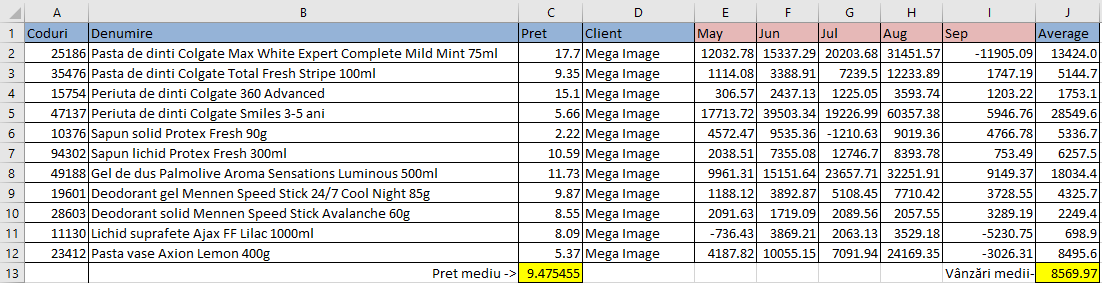
Scopul principal al companiei este menținerea succesului la nivel regional, dar și la nivel global; iar, de asemenea, dorința de a reduce cât mai mult până în anul 2020 fabricarea de ambalaje din plastic nenecesare.

#MakingSmilesTogether

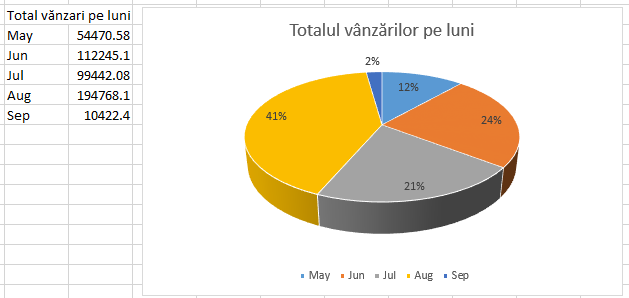
**2. Pachetul Excel**

**2.1 Realizarea unei foi de calcul**

Am realizat o foaie de calcul în care am introdus date referitoare la vânzarile anumitor produse din lunile Mai, Iunie, Iulie, August, Septembrie. Aceasta conține si codurile produselor, denumirile, prețul, cât și clientul căruia au fost vândute. În cazul de față am selectat doar produsele distribuite clientului Mega Image.



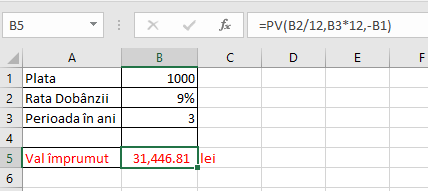
Totodată, am realizat un grafic care ilustrează totalul vânzarilor in fiecare luni, din care reiese că luna cea mai profitabilă pentru companie a fost August cu o valoare de 41% din totalul vânzărilor, iar cea mai puțin profitabilă a fost Septembrie, cu doar 2% din totalul vânzărilor.



**2.2 Functii financiare**

Problemă în care utilizăm **funcția PV**, care returneaza valoarea prezentă a unei investiții

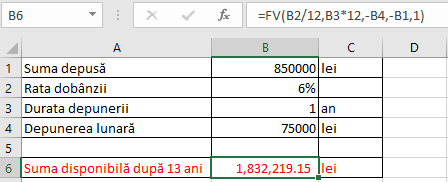
Compania a luat în considerare închirierea unui spațiu de depozitare a produselor venite din afara țării. Prin această modalitate vor ușura transportul realizat în magazinele partenere, urmând să plătească un împrumut lunar de 1.000 lei. Perioada de închiriere este de 3 ani, iar pentru realizarea închirierii se va înregistra o rată a dobânzii de 9%. Care va fi valoarea împrumutului?



În urma efectuării calculelor rezultă că împrumutul va costa în total 31.446,81 lei.

Problemă în care utilizăm **funcția FV**, care returneaza valoarea viitoare a unei investiții bazate pe plăți periodice și constante și cu o rată a dob. constante.

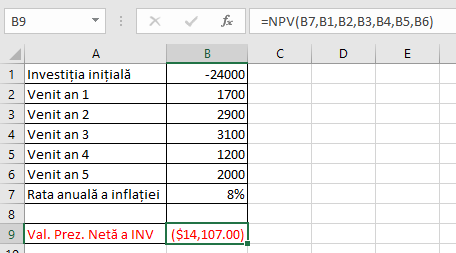
Firma, după o vânzare din care s-au obținut 850000 de lei, vrea să depună banii la bancă/ Rata anuală a dobânzii este de 6%. De asemenea se mai depune in fiecare lună suma de 75000 lei, o parte din veniturile fixe. Care este valoarea contului după un an de zile?



După perioada de 12 luni, în cont se va regăsi valoarea de 1832219,15 lei.

Problemă în care utilizăm **funcția NPV**, care calculează valoarea prezentă netă a unei investiții utilizând rata inflației si o serie de plați și venituri viitoare.

Presupunem că firma face o investiție de 24000 lei, iar în următorii 5 ani primește sumele: 1700 Ron; 2900 Ron; 3100 Ron; 1200 Ron și 2000 Ron. Dacă rata anuală a inflației este 8%, care va fi valoarea prezentă netă a investiției?



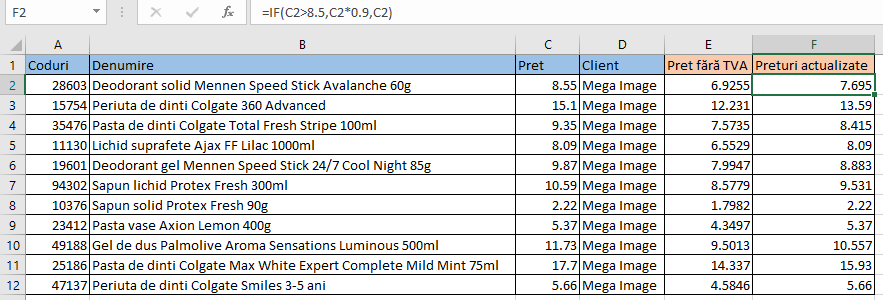
**2.3 Funcții logice**

Se dorește calcularea prețurilor produselor fără a se lua în calcul taxa pe valoare adăugată.



Prețurile calculate fără TVA se încadreaza în coloana E.

Se dorește ca produsele care au prețul mai mare de 8,50 lei să primească o reducere de 10%.

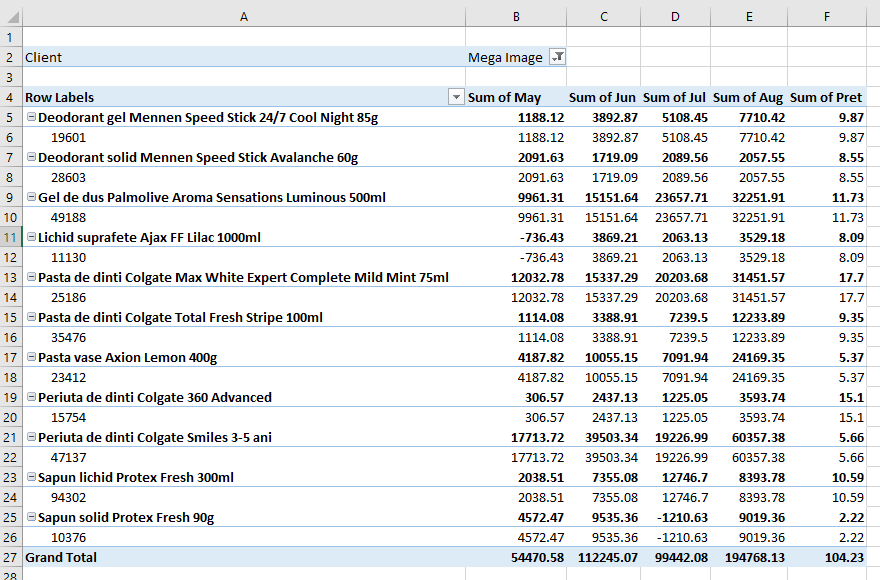


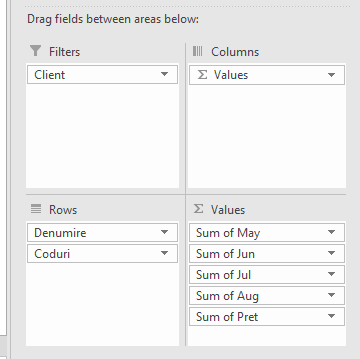
Prețurile noi (atât cele ale produselor ieftinite, cât și a produselor care nu au suferit o ieftinire) se pot observa in coloana F.

Formulele folosite pentru realizarea funcțiilor logice se pot observa în print screen-urile realizate.

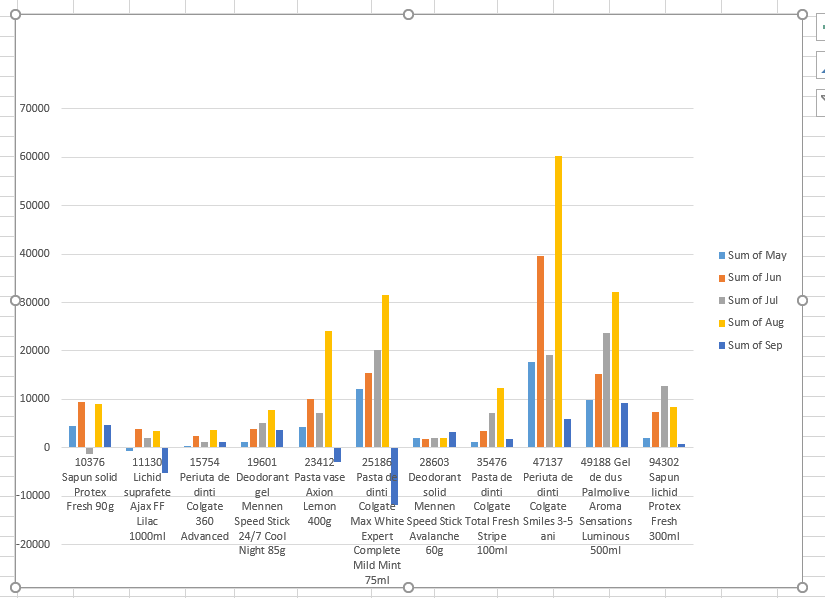
**2.4 Funcții definite de utilizator**

**2.5 Pivot table și pivot chart**





Raportul Pivot Table este un tabel interactiv care se utilizează pentru însumarea cantitătiților mari de date. Pentru a putea observa diferitele rezumate se pot roti rândurile si coloanele.

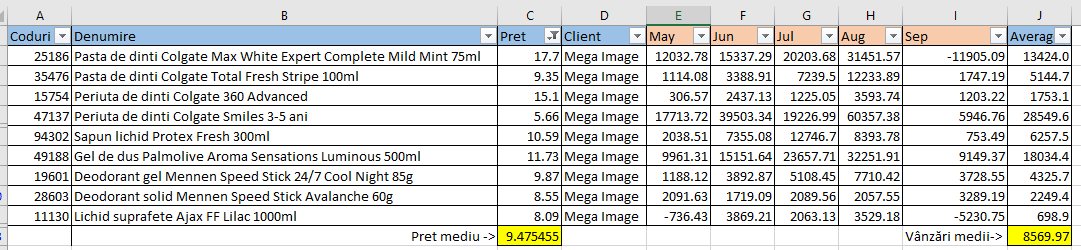


Graficul de mai sus evidențiază vânzările din lunile Mai, Iunie, Iulie, August, Septembrie. Se observă de asemenea, denumirea fiecărui produs și codul intern al produsului.

Analiza vânzărilor a fost realizată la nivelul clientului Mega Image.

Pentru studierea valorilor individuale pe produs se pot folosi filtrele existente în grafic.

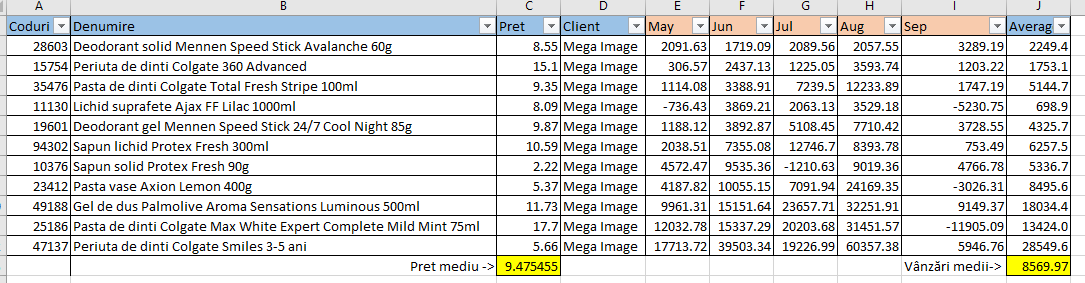
**Filtrări**



Am realizat o filtrare care să conțină doar primele 10 cele mai scumpe produse.

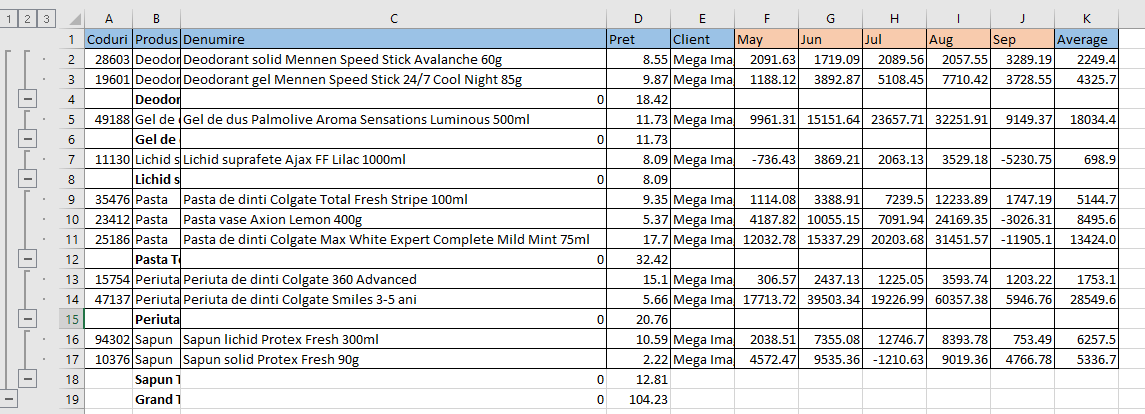
**Sortare**

Vom realiza o sortare a produselor în funcție de vânzările din luna Iunie, în ordine crescătoare.



**2.6 Subtotaluri**

Utilizând funcția subtotal, am obținut următorul tabel:

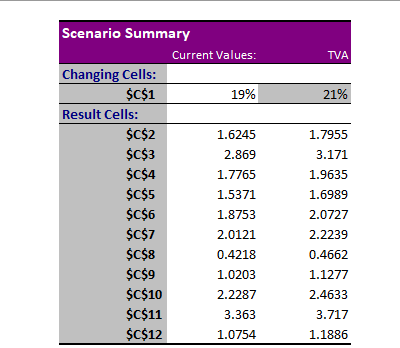


Din tabel reies prețurile fiecărui produs, cât și prețul total al fiecărui pachet de produse. Spre exemplu pachetul de Deodorante conține două produse Deodorant solid Mennen Speed Stick Avalanche 60g și Deodorant gel Mennen Speed Stick 24/7 Cool Night 85g, iar prețul pachetului este de 18,42 lei.

**2.7 Șcenarii**

Pe baza datelor anterioare vom realiza un șcenariu unde se va urmări efectul creșterii valorii TVA de la 19% la 21%.

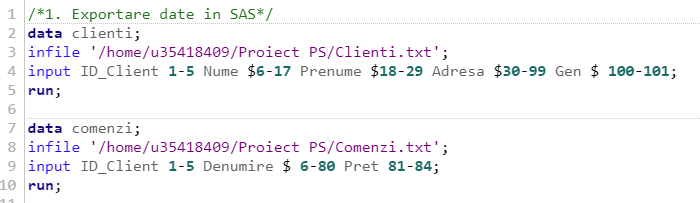
După realizarea șcenariului, putem observa schimbarea TVA-ului aplicata fiecarui produs, cât și o comparație între vechea valoare si noua valoare.



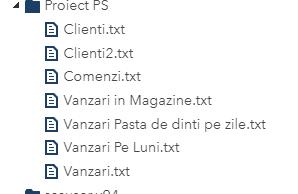
**3. Pachetul SAS STUDIO**

**3.1 Importarea datelor in SAS STUDIO**

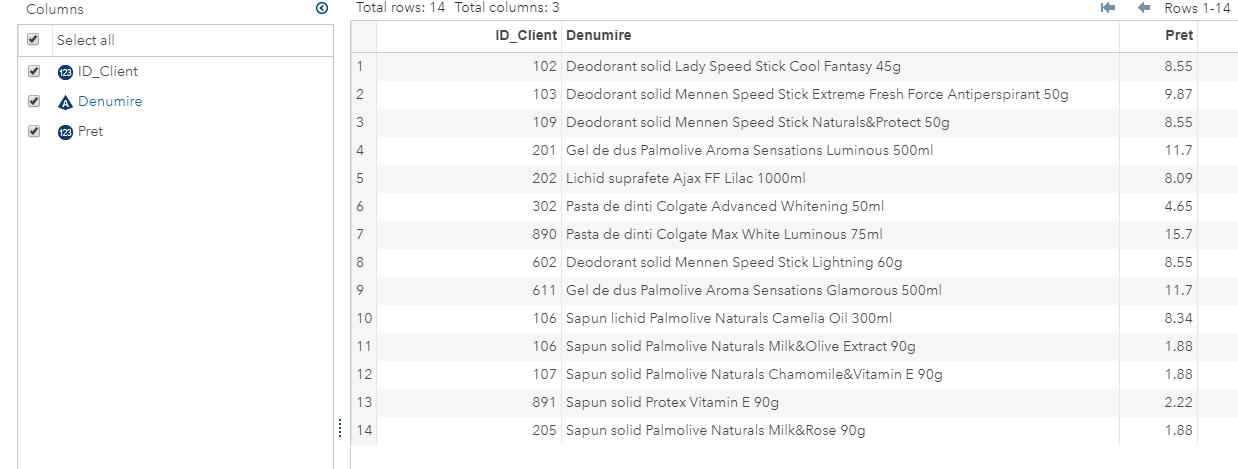
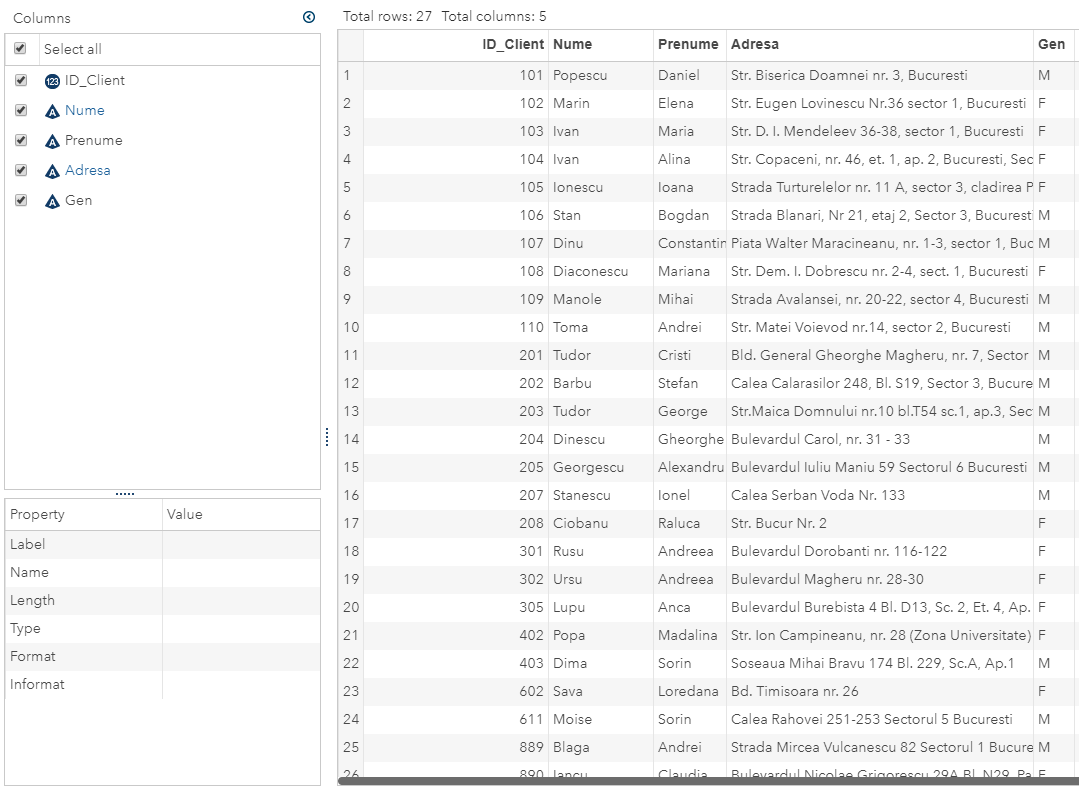
Pentru început am creat două fișiere text, primul referitor la clienții Colgate Palmolive, iar al doilea referitor la comenzile realizate de aceștia, pe care le-am importat în SAS conform codului următor:



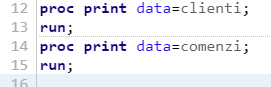
În vederea rezolvării cerințelor următoare, am creat mai multe fișiere text:



Conținutul fișierelor pe care le vom folosi se pot observa in print screen-ul următor:



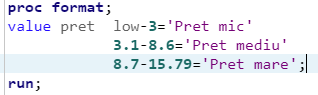
Folosim funcția print pentru a afișa și în același timp încadrarea datelor într-un tabel.



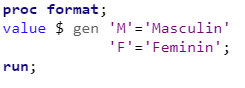


**3.2 Formate definite de utilizator**

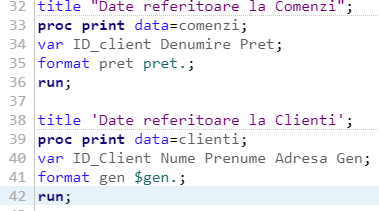
Pentru a gestiona mai ușor prețurile, o să realizăm un format care să le încadreze în prețuri mici, medii și mari.

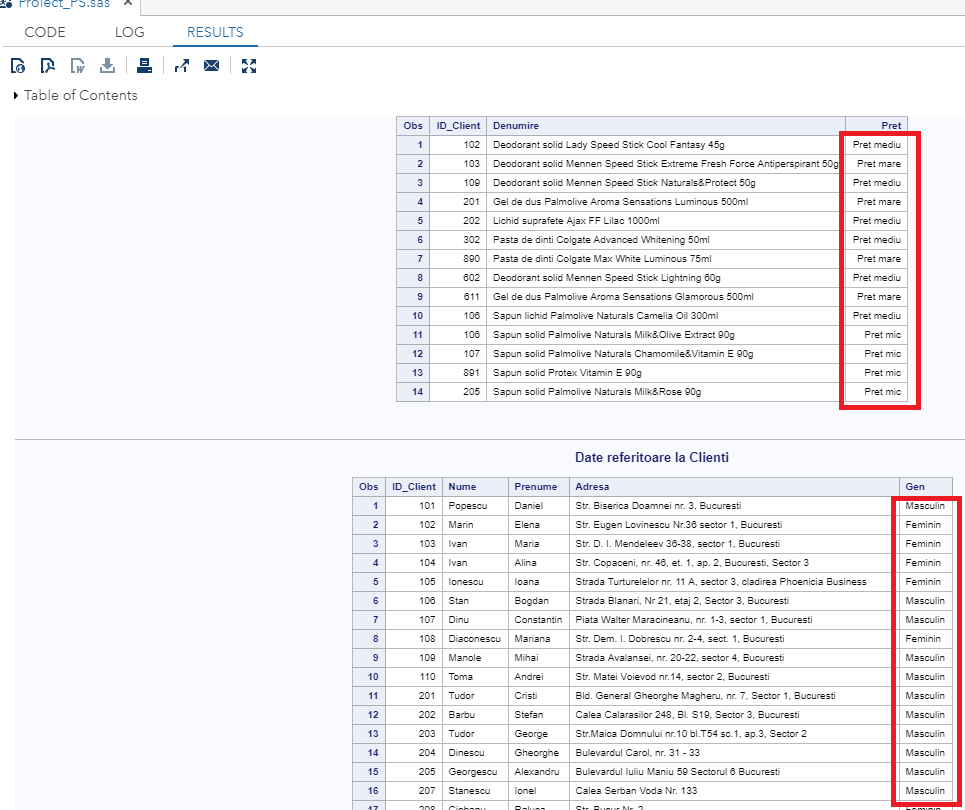


Totodată mai realizăm un format pentru gen:



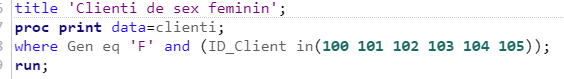
În printscreen-urile de mai jos se pot observa formatele utilizate în rularea efectivă.





**3.3 Procesarea iterativă si condiționată**

Să se selecteze clienții de sex feminin a căror cod se încadrează între valorile de 100 și 105

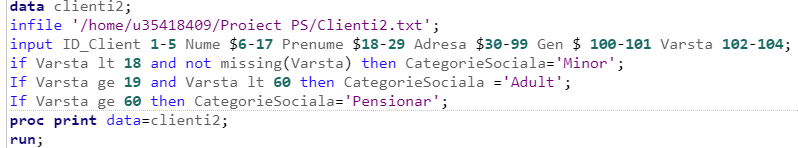


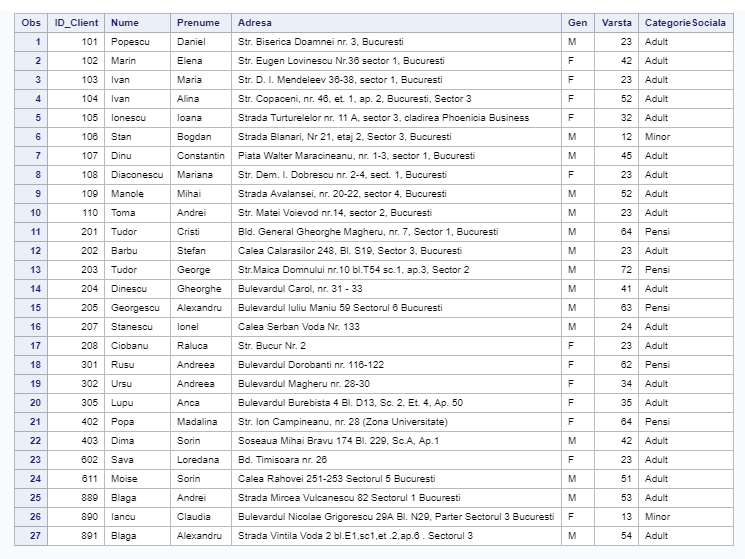


Să se încadreze datele în categorii bazate pe vârsta clienților astfel:

-Minor pentru persoanele cu vârsta mai mică decât 18 ani

-Adult pentru persoanele cu vârsta cuprinsă între 19-60 de ani

-Pensionar pentru persoanele cu vârsta peste 60 de ani

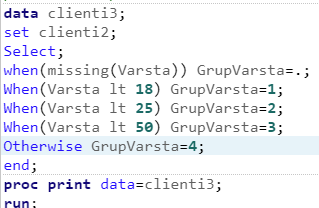


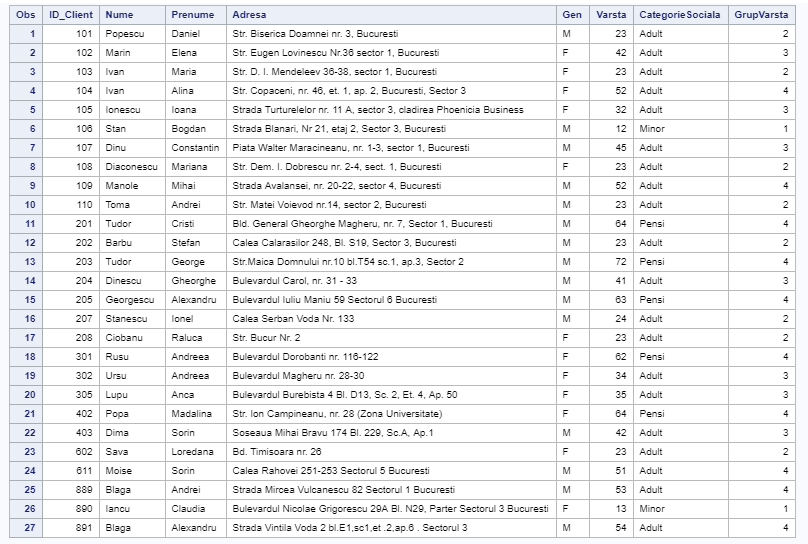
Să se încadreze săpunul într-o gamă special dedicată săpunului, iar restul produselor unei alte game, ținând cont de faptul că săpunul este cel mai ieftin produs cu un preț mai mic de 3 lei.



Funcția Select

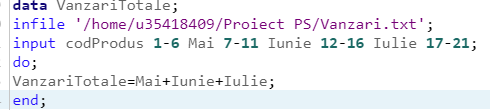
Să se atribuie fiecărui client o grupă de vârstă de la 1 la 4 bazată pe vârstă, folosindu-se funcția select.





Funcția DO

Folosindu-se fișierul Vanzari.txt, care conține date referitoare la vânzările pe lunile Mai, Iunie și Iulie, să se calculeze vânzările totale pentru fiecare produs în parte.

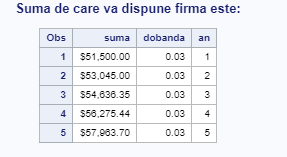




Firma dorește să facă o investiție pe 5 ani, în valoare de 50000 de lei, cu o dobândă de 3%. Care va fi valoarea în lei de care o să dispună firma după această perioadă?



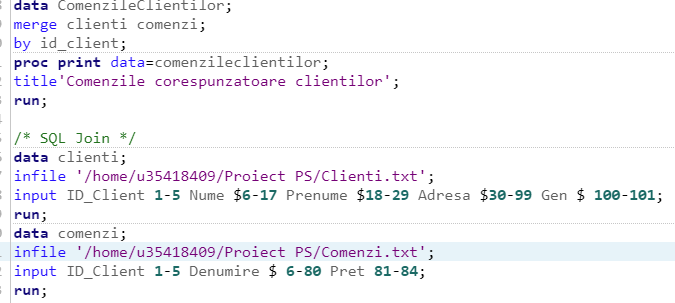
Rezultatul rulării codului de mai sus:



După cei 5 ani, firma o să dispună de o valoare de 57.963,70 de lei.

**3.4 Combinarea seturilor de date**

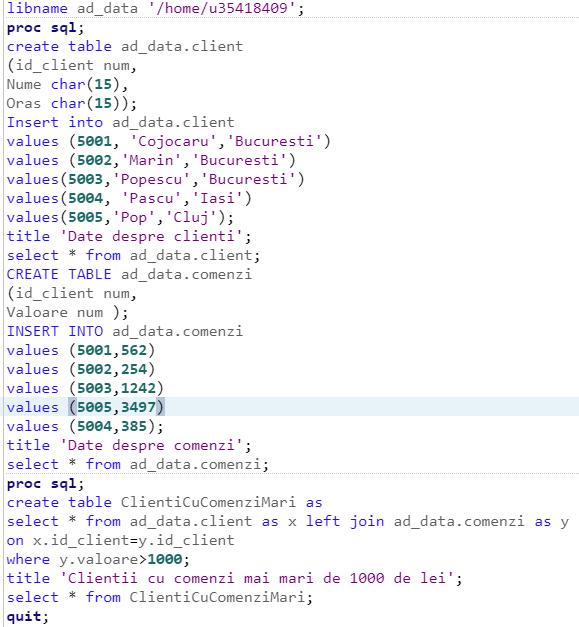
Folosind datele text de la primul subpunct, să se realizeze combinarea datelor atât prin intermediul funcției merge cât și prin funcțiile SQL.

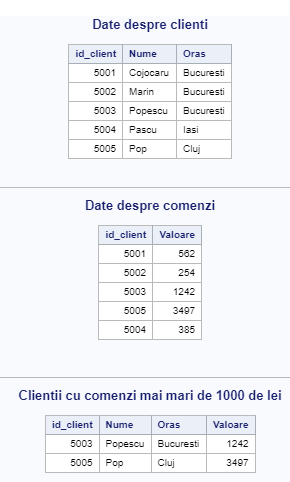






Dorim să creăm o nouă bază de date prin intermediul funcțiilor SAS și SQL în care să introducem noi valori referitoare la comenzi și clienți. Să se folosească funcția join și se afișeze doar clienții care au realizat comenzi cu prețul mai mare de 1000 de lei .

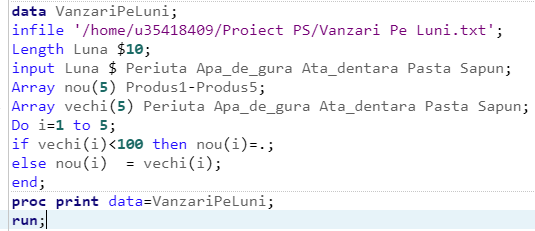


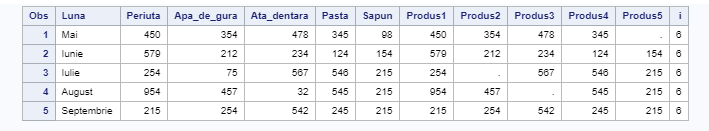


Clienții care au plasat comenzi mai mari de 1000 de lei sunt Popescu din București și Pop din Cluj.

**3.5 Utilizarea de masive**

Pe baza fișierului text Vanzari Pe Luni în care sunt încărcate vânzările produselor: Periuță, apă de gură, ață dentară, pastă, săpun, se dorește să se scoată din calcul vânzările cu o valoare mai mică de 100.

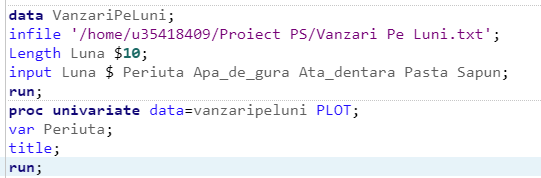


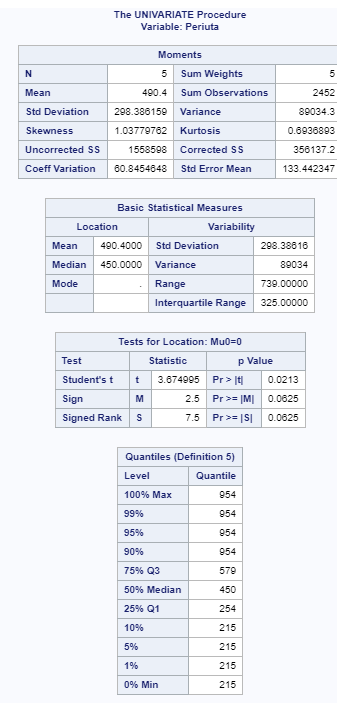


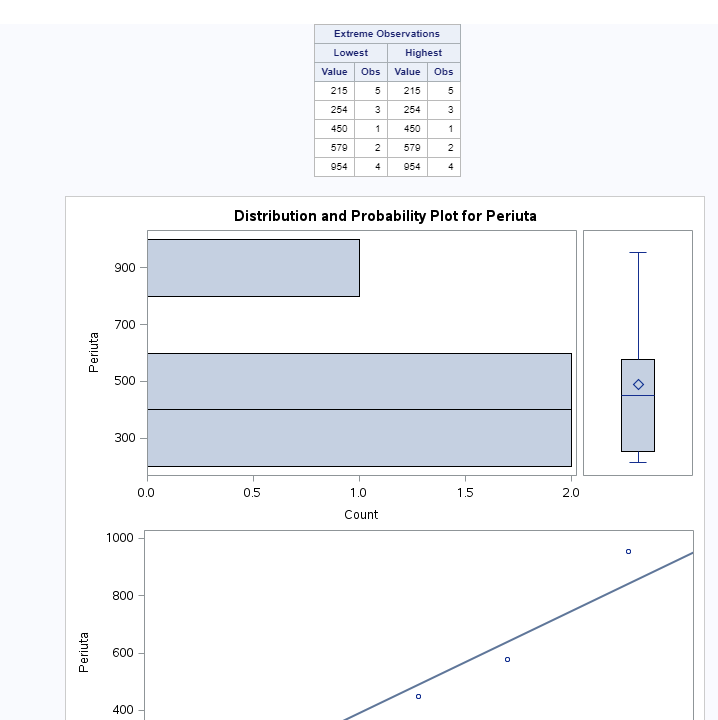
În tabelul rezultat se pot observa cele 3 valori mai mici decât 100 înlocuite cu Semnul punct și produsele a căror denumire a fost schimbată generic în produs1-produs2.

**3.6 Utilizarea procedurilor pentru rapoarte și aplicarea de analize statistice**

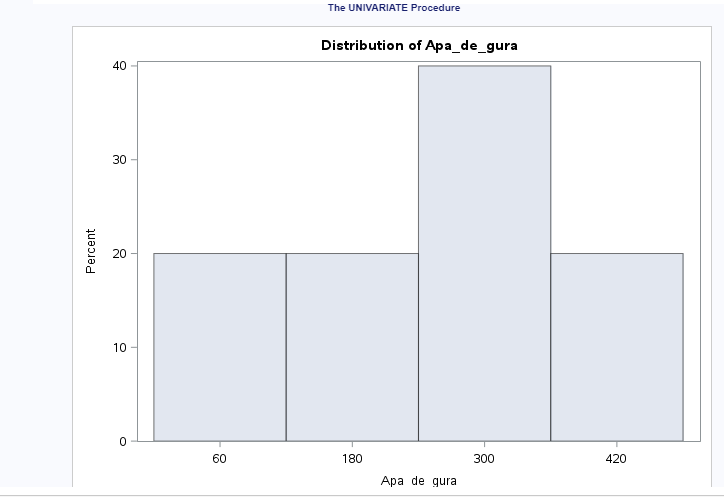
Să se examineze distribuțiile datelor, normalitatea lor și să se identifice valorile extreme din cadrul fișierului Vanzari pe luni.txt, utilizând funcția UNIVARIATE.





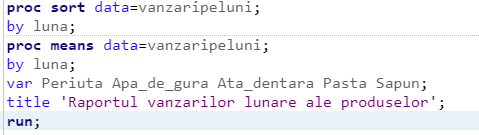


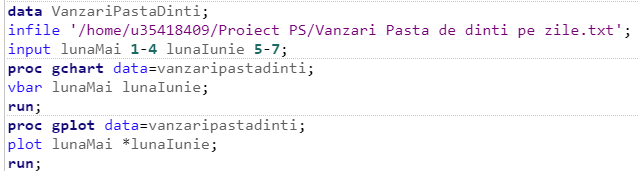
Histograma din care reiese procentul din vânzări înregistrat în fiecare lună.

****

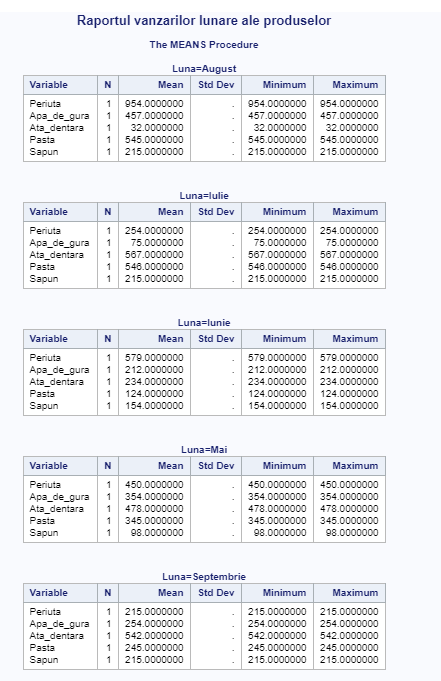
Să se realizeze raportul vânzărilor lunare ale produselor.

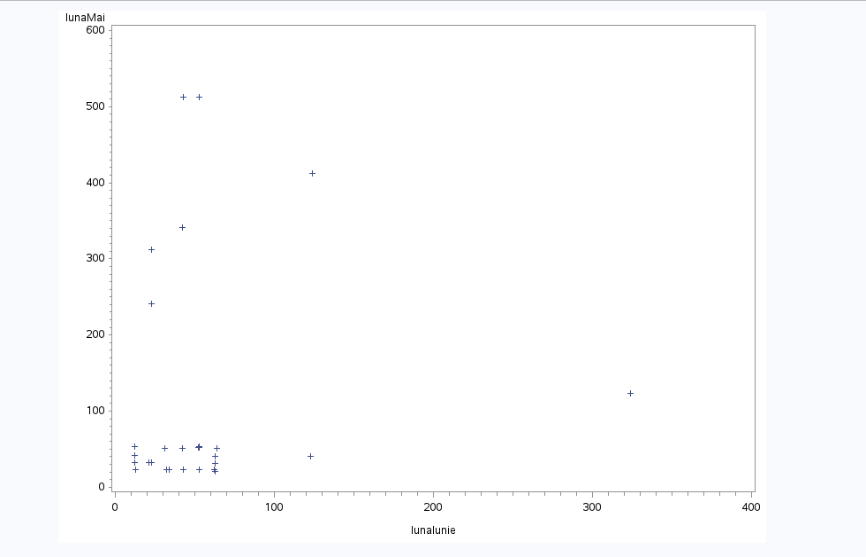
Pentru a realiza raportul, mai întâi sortam datele dupa lună.

****



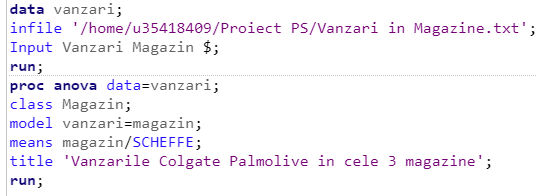
Rezultatul rulării codului:

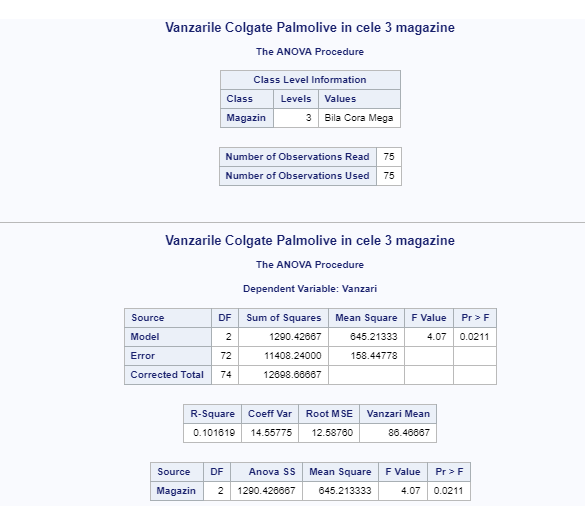


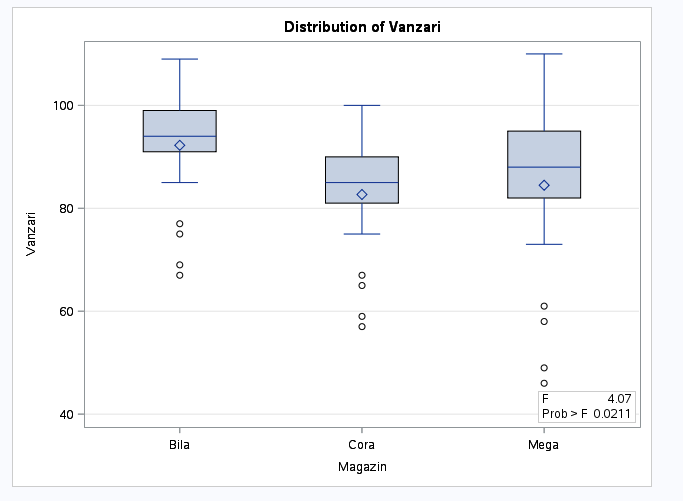


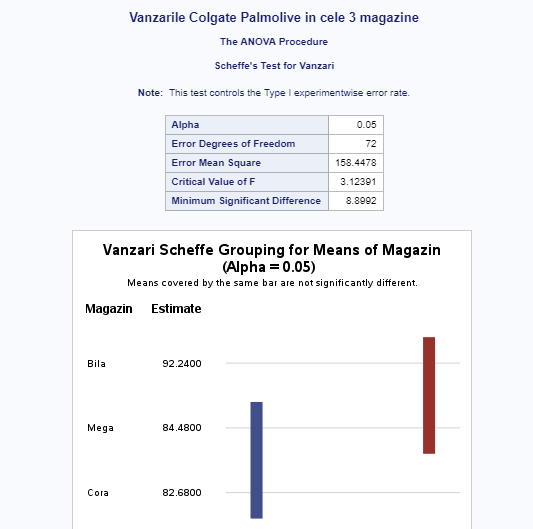
Se poate observa că valorile predominante se află în jurul valorii 0-100, dar se regasesc și valori extreme în luna Mai în jur de 500-550 de vânzări, dar și de 300-350 în luna Iunie.

Se vor analiza vânzările produselor firmei Colgate Palmolive în 3 magazine: Cora, Mega Image și Bila. Pentru realizarea acestei analize se va utiliza fișierul Vanzari in magazine.txt cu ajutorul testului ANOVA.









**4. Utilizarea pachetului SAS EG**

**4.1 Importarea fișierelor non-SAS (fișier Excel)**

Să se importe fișierul ClientiEG.xlsx cu următoarele specificații: primul rând al fișierului reprezintă numele câmpului, iar coloanele se vor formata în funcție de parametrii conținuți.

Pași de rezolvare:

File- Import Data Selectăm fișierul dorit



**4.2 Interogări(Select și filter)**

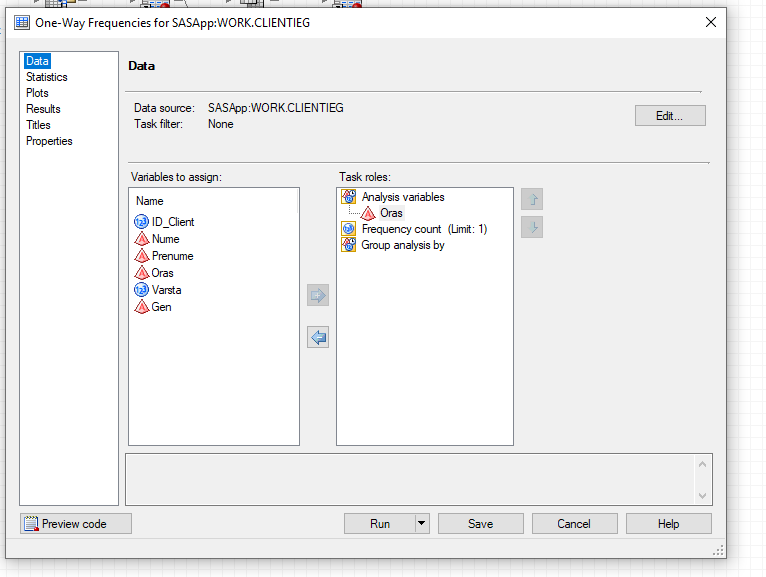
Să se creeze un raport care să afișeze Clienții stabiliți în București.

Pași de rezolvare:

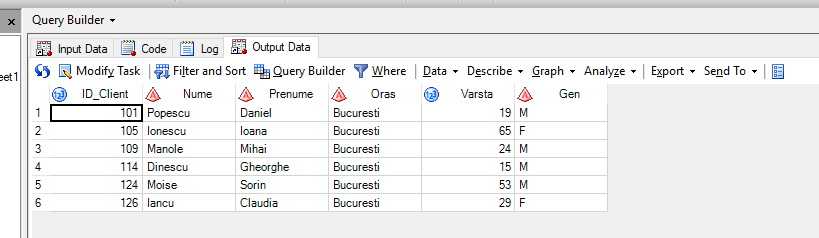
Selectăm Data imported from CleintiEG.xlsx;

Utilizăm funcția Query Builde, unde vom încărca toate coloanele pe care o să le utilizăm;

La Filter Data creăm un nou filtru, unde selectăm variabila Oraș



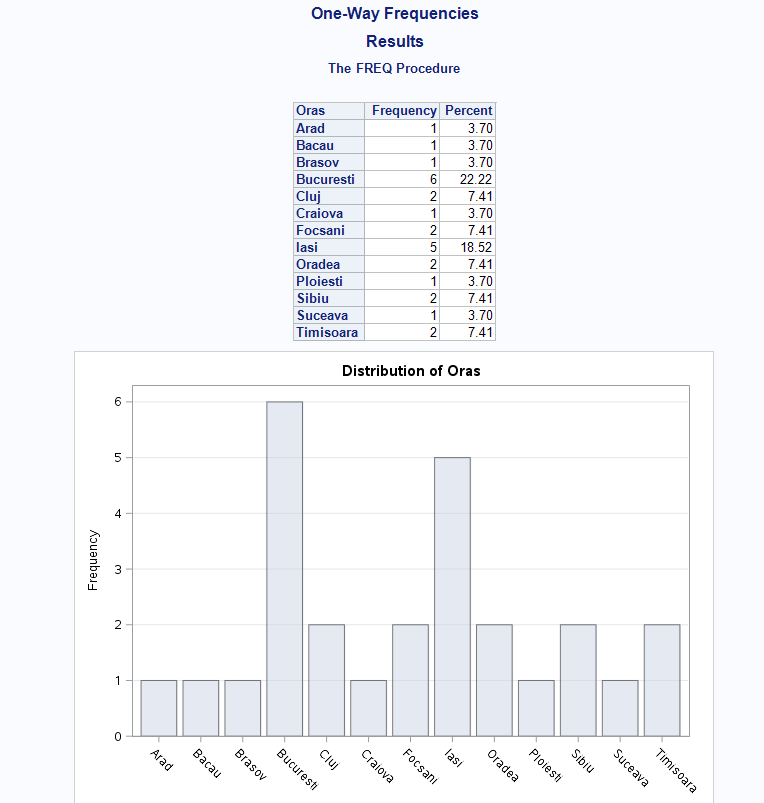
Deasupra avem selectarea variabilei, iar jos rezultatul obținut după filtrare.



De asemenea o să realizăm si o descriere a frecvențelor de apartenentă a clienților la orașele din care fac parte.

Pentru acest lucru o să selectăm din optiunea Task optiunea One Way Frequencies.

Rezultatul se poate observa mai jos.



Din summary observăm că București este orașul cu cei mai mulți clienți care au depus comenzi, cu o frecvență de 22,22%, iar din barplot observăm numărul total al clienților și orașele din care fac parte.

**4.3 Joncțiunea**

**Joncțiune internă**

Dorim să cream o joncțiune a tabelelor Clienți și Comenzi.

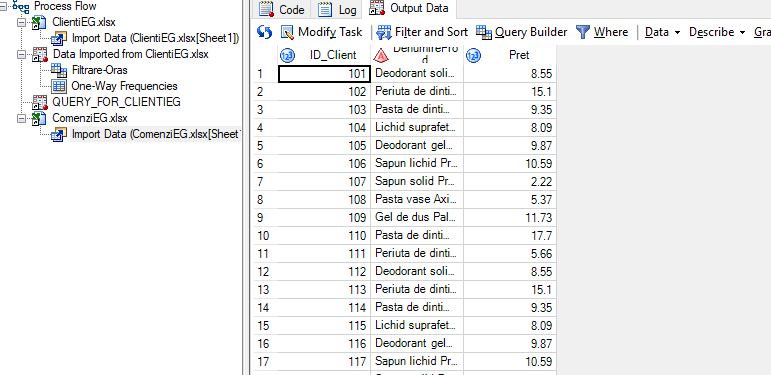
Pași de rezolvare:

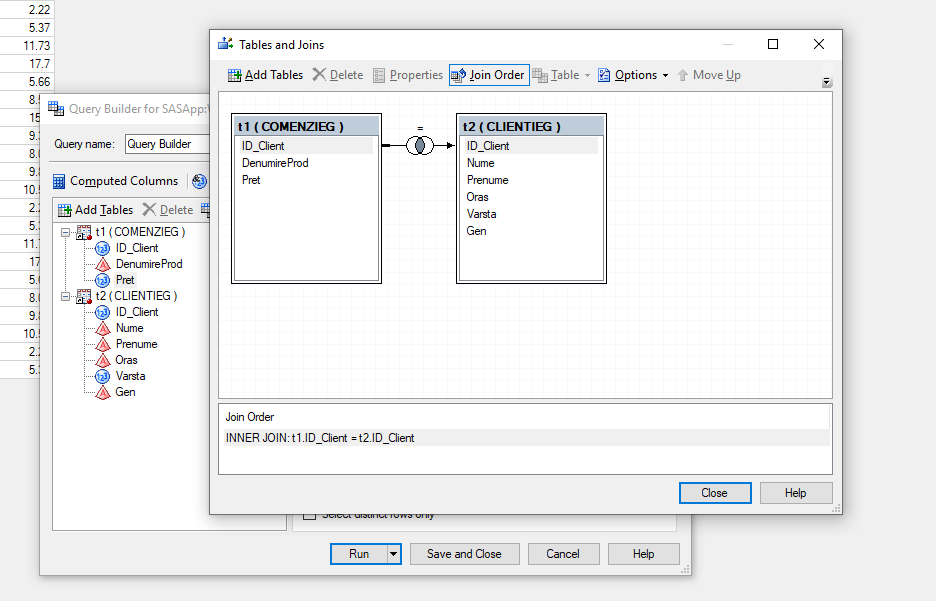
Pentru început trebuie să importăm noua foaie cu comenzile efectuate de clienți.

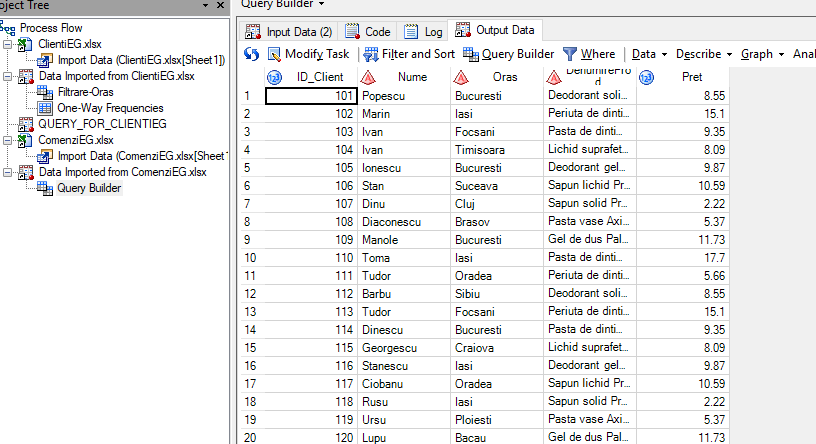
Selectăm Data imported from Clienti, Query Builder, unde vom adăuga noua tabela cu comenzile la Add tables.

În secțiunea Join Tables putem oserva legătura care se formeaza automat dintre cele două tabele, datorită faptului că fiecare tabelă conține variabila comună ID\_Client.

Introducem toate variabilele pe care le dorim utilizate și apasăm Run.







**Joncțiune externă:**

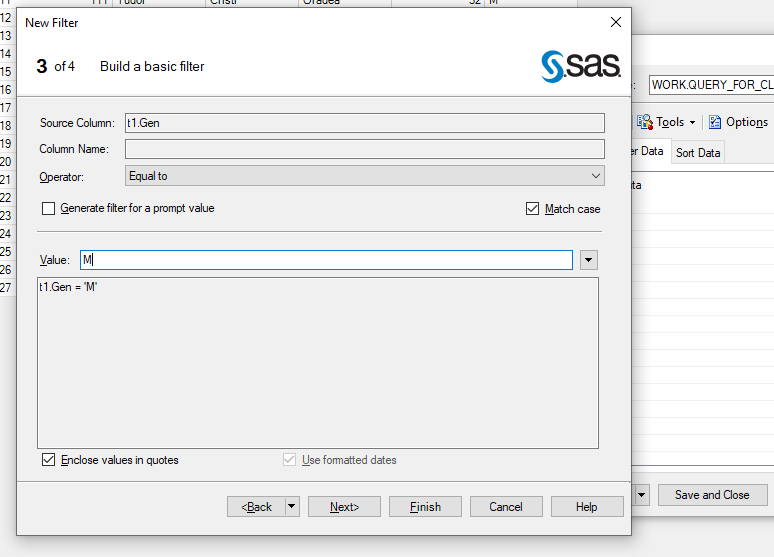
Dorim realizarea unei Jonctiuni de tip right join. Firma doreste sa identifice comenzile care au fost plasate doar de către persoanele de sex M.

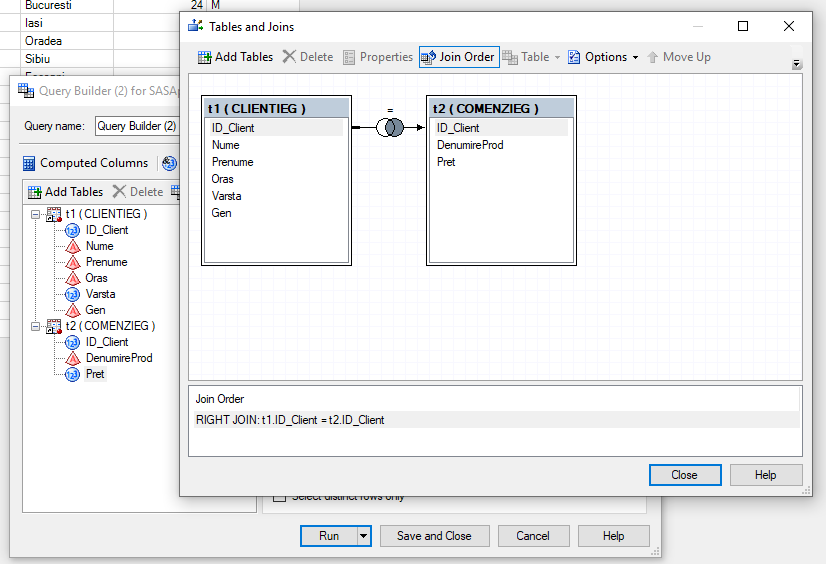
Pași de rezolvare:

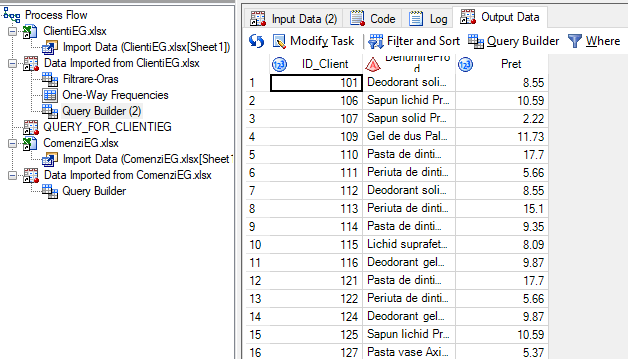
Selectăm tabela Clienți și accesăm Query Builder;

La secțiunea Filter Data, o să introducem prin New Filter variabila Gen, iar la Value scriem M pentru a lua în calcul doar persoanele de sex masculin.

La secțiunea Join Tables, o să modificăm tipul de joncțiune în Right Outer Join.







În noua tabelă avem doar persoanele de sex masculin care au plasat comenzi.

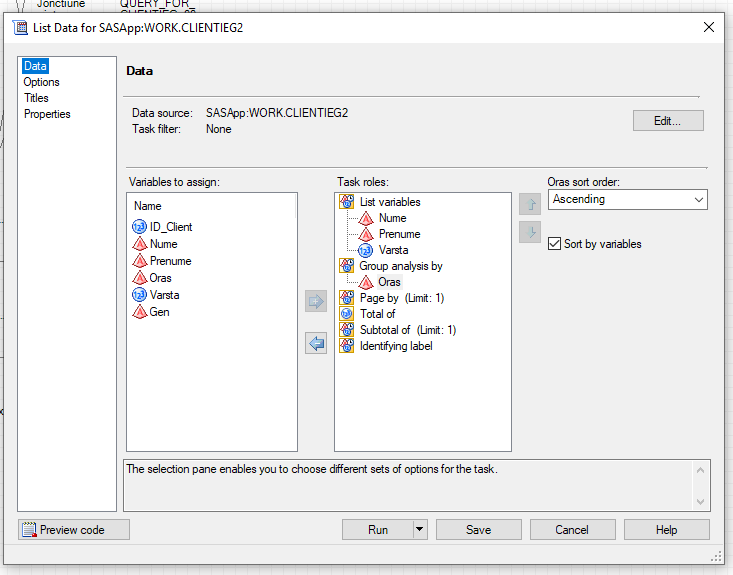
**4.5 Raport**

Firma dorește să posteze o listă a clienților săi din ultima perioadă. Lista trebuie să aibă următoarele caracteristici: clienții din același oraș trebuie să fie grupați impreună, iar numele Orașului trebuie să apară doar o singură data în raport.

Pași de rezolvare:

Selectăm tabela pe care dorim să facem raportul.

Din opțiunea Task selectăm Data List, iar în noua fereastră introducem datele din cerință





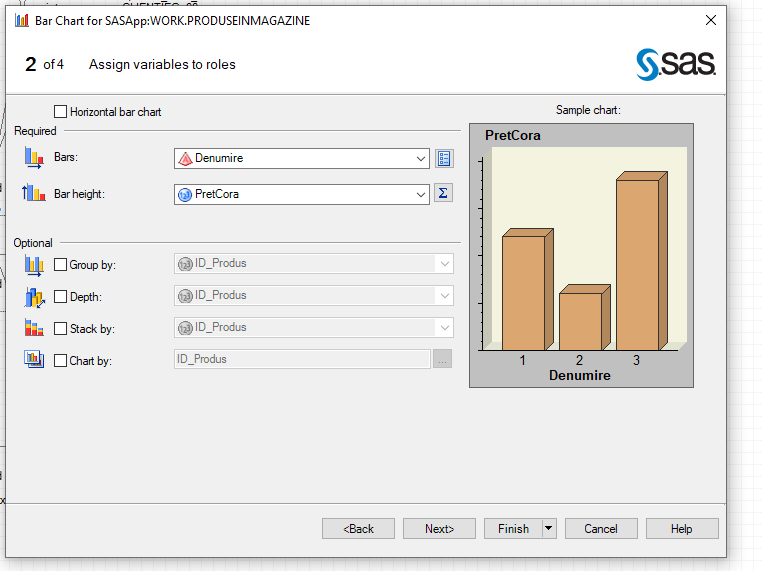
**4.6 Grafic**

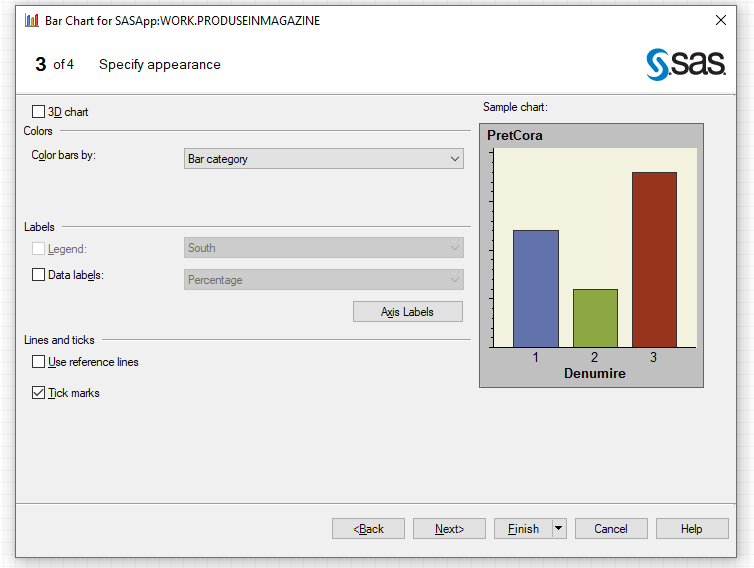
Firma dorește să creeze un grafic care să ilustreze fiecare produs vândut în ultima perioadă, cât și prețurile la care au fost vândute în supermarketul Cora.

Pași de rezolvare:

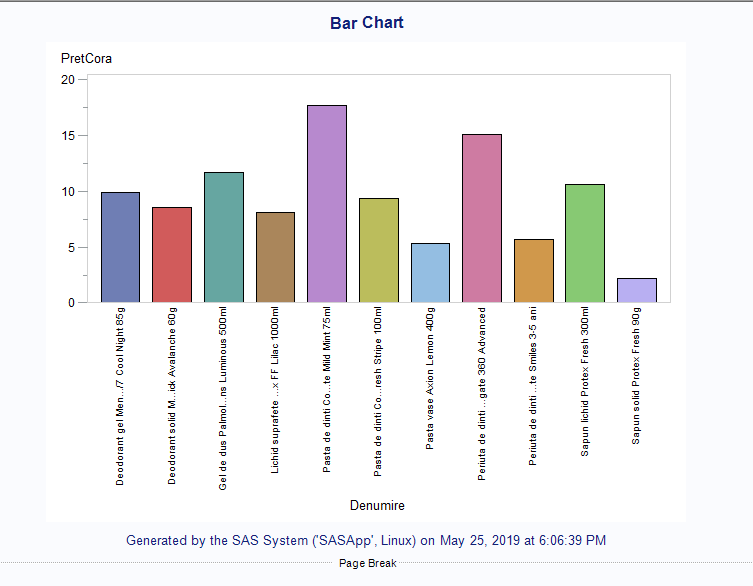
Selectăm tabelul pentru care dorim să realizăm graficul (ProduseInMagazine);

Intrăm la Task-Graph-Bar Chart Wizard, unde vom introduce variabile necesare pe care le luăm în considerare: prețurile din supermarketul Cora și numele produselor cărora le aparțin.





Rezultatul obținut :



Din rezultat observăm că cel mai scump produs vândut in magazinul Cora este Pasta de dinți Co.. 75ml, iar cel mai ieftin Sapun solid..90g.

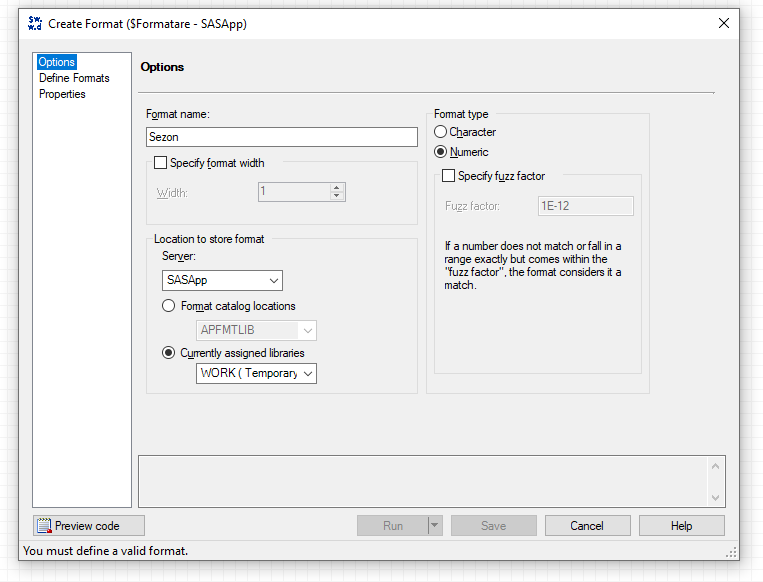
**4.7 Formate definite de utilizator**

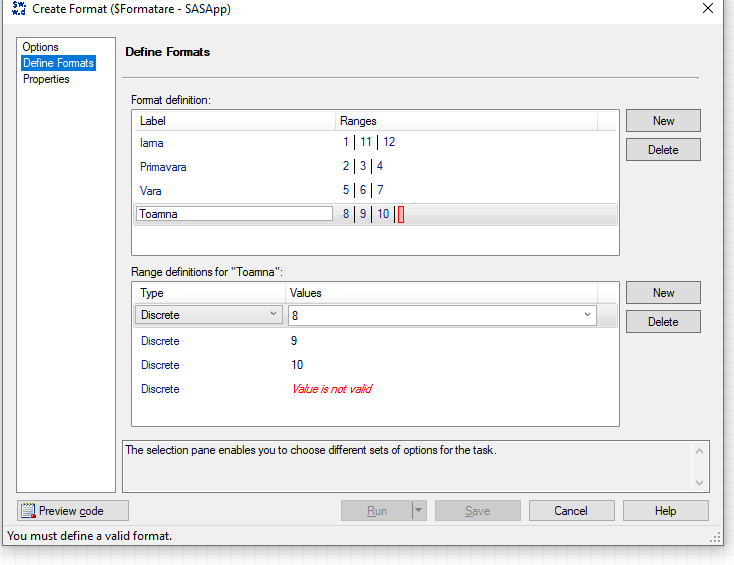
Pentru a gestiona mai ușor produsele pe luni se dorește crearea unui nou format care să încadreze lunile în sezoane :Primavara,Vara,Toamna,Iarna.

Pași de rezolvare:

Se selectează orice Query realizat până acum si se deschide tab-ul Data;

Selectăm Create format și introducem datele conform screen shoturilor de mai jos



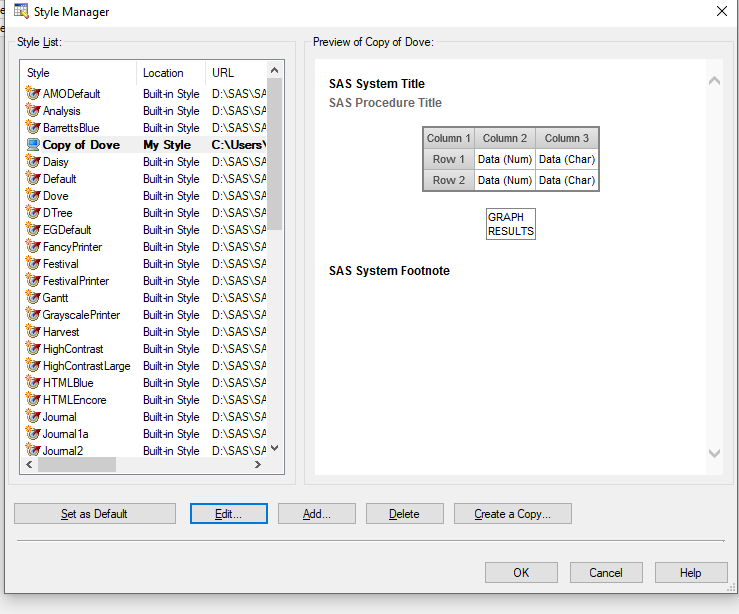


**4.8 Stiluri create de utilizator**

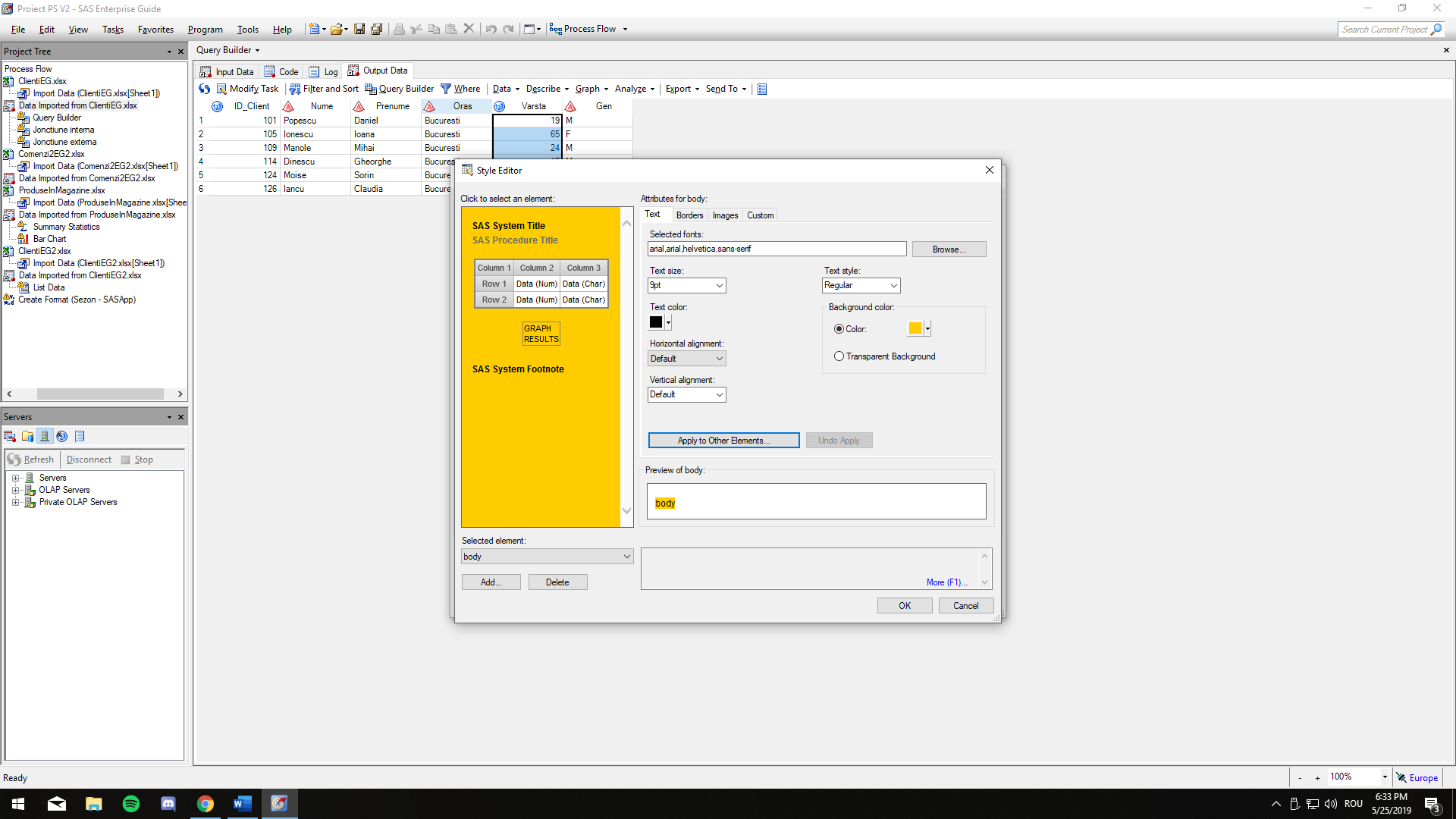
Pentru a crea un nou stil urmăm pașii:

Deschidem Tools – Style Manager;

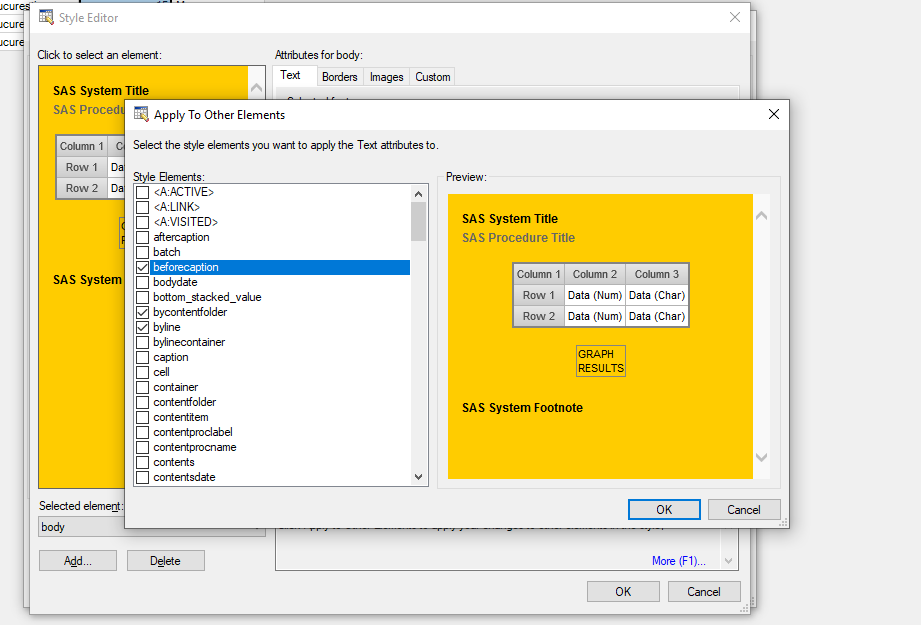
Selectăm un stil de referință pe care urmează să lucrăm și să îl modificăm.;



Selectăm Edit;



Putem de asemenea adăuga alte elemente;



**4.9 Prelucrări statistice**

Firma urmărește să scadă prețul unui produs cu o unitate monetară în fiecare zi timp de o lună pentru a putea observa trendul manifestat de cumpărătorii acestui produs.

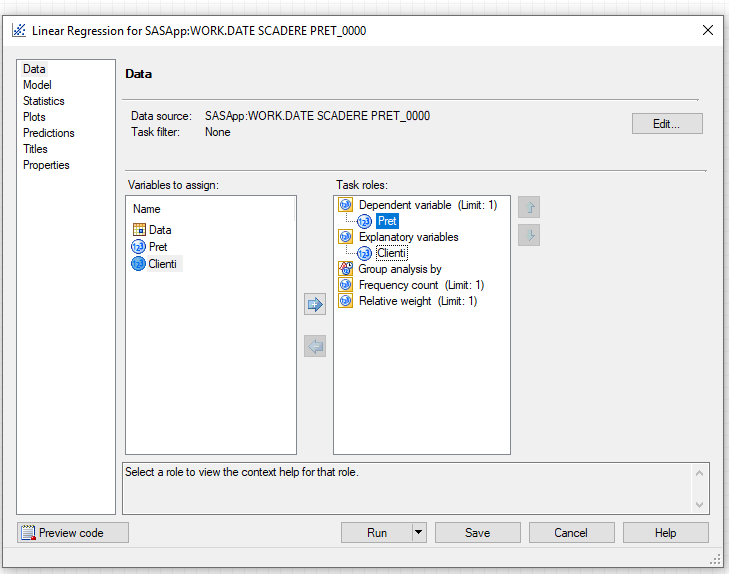
Se vor utiliza datele înregistrate în luna respectivă Date scadere pret.xlsx.

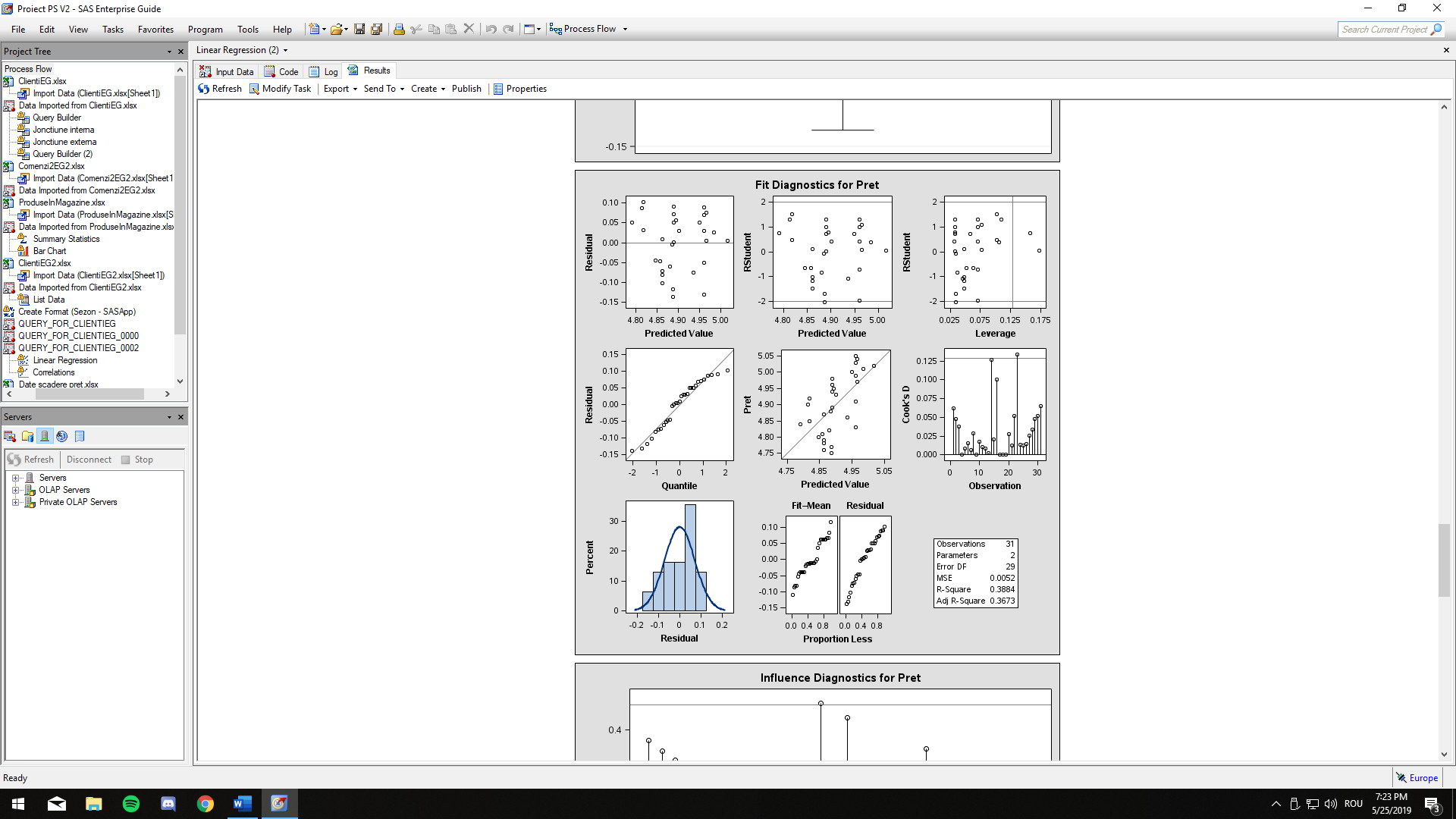
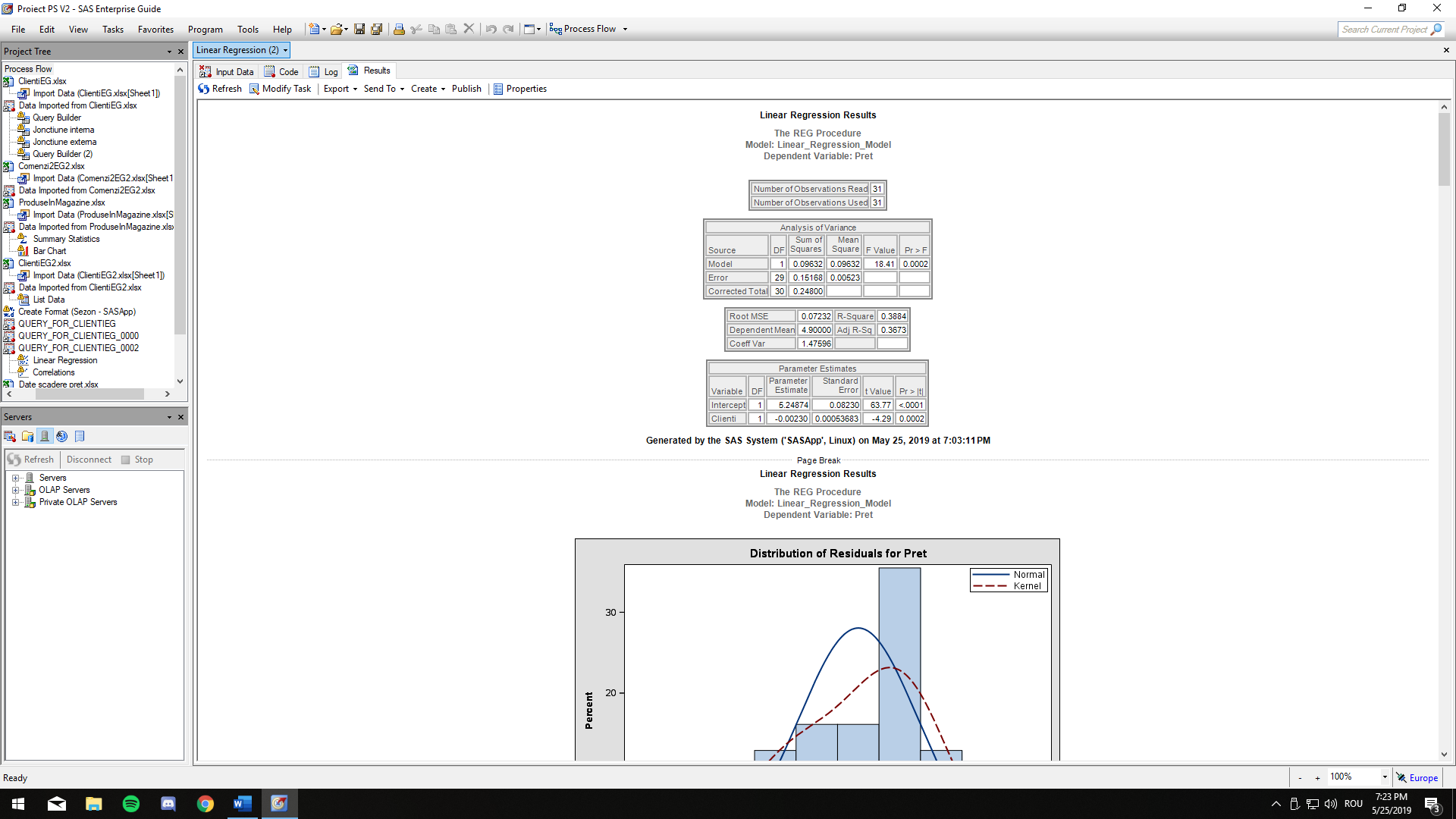
Pași de rezolvare:

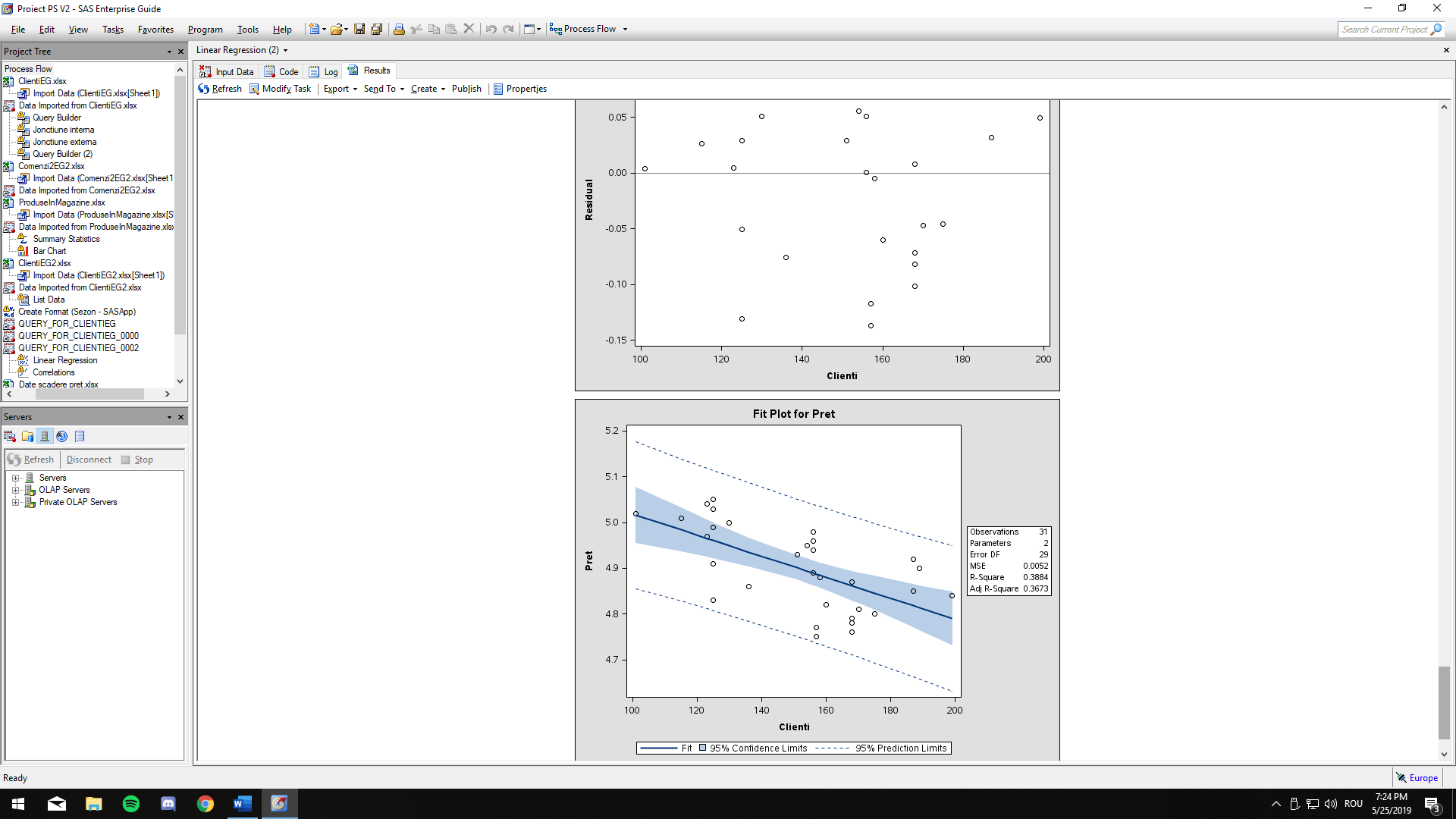
Selectăm Data imported from Date scadere pret.xlsx.

Din tabul Tasks selectăm Regressions – Linear regression

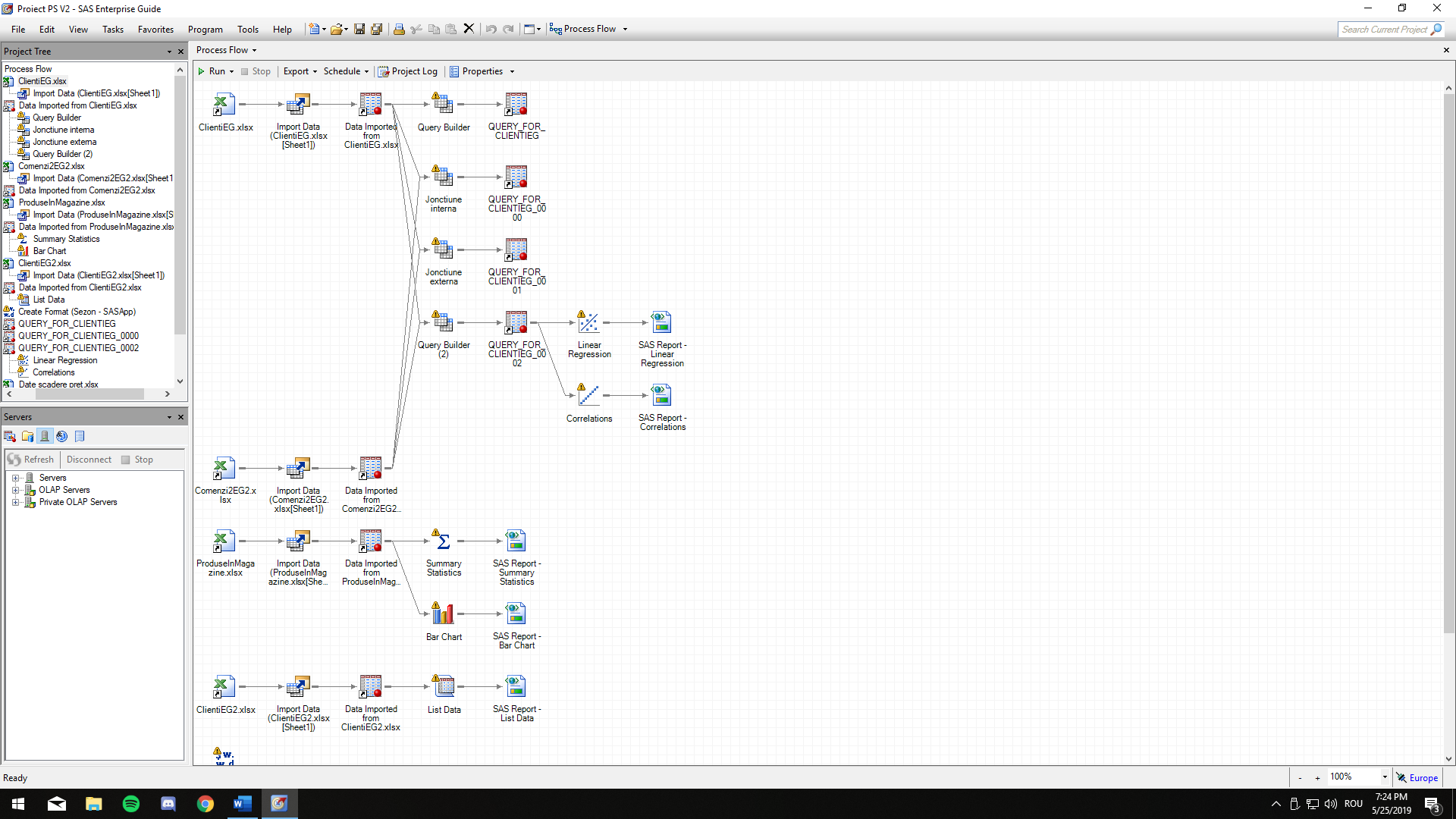
Includem Prețul și nr-ul de clienți în procesul de regresie și apasăm Run.







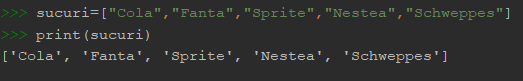
Process Flow final:



**5.Python**

**Utilizarea listelor și a dicționarelor**

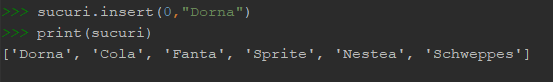
Vom crea o listă cu câteva elemente, iar pe aceasta vom exemplifica câteva funcții.



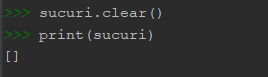
Lungimea listei, sau mai degrabă numărul de elemente este următorul:



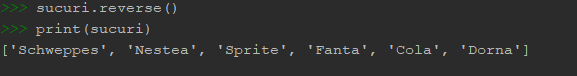
Am adăugat cu ajutorul insert un nou obiect în listă, la poziția 0:



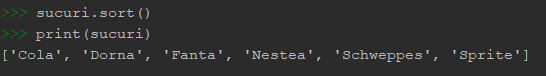
Prin clear am eliminat elementele din listă:



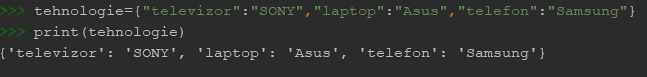
Cu ajutorul reverse am inversat ordinea elementelor în listă:



Iar cu ajutorul sort a fost realizată o sortare alfabetică a produselor:



Vom crea un dicționar pe baza căruia vom exemplifica câteva metode:



Prin metoda get vom afișa Asus din categoria laptop:



Metoda items returnează liste cu tuplu pentru fiecare pereche din dicționar:



Metoda keys afișează cheile dicționarului:

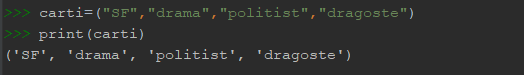


Iar prin metoda values se afișează valorile dicționarului:



**Utilizarea seturilor și a tuplurilor**

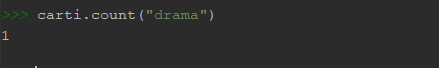
Vom crea un tuplu referitor la genurile de cărți:



Accesarea elementelor din tuplu se face astfel:



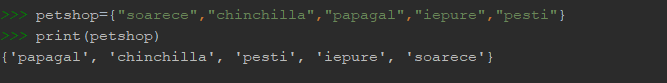
Cu ajutorul count se numără aparițiile unui element:



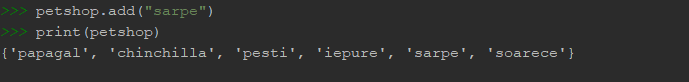
Cu ajutorul index se precizează poziția unui obiect în tuplu:



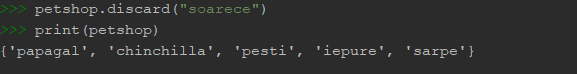
Vom crea un set de date și vom aplica câteva metode specifice:



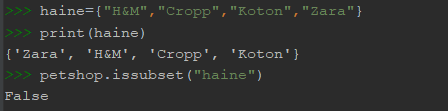
Cu ajutorul add adăugăm înregistrări noi în set:



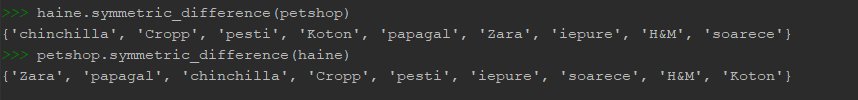
Discard se folosește pentru eliminarea unui obiect din set:



Issubset arată dacă un set e inclus în altul sau nu. În scopul folosirii metodei am mai creat un set de date – haine:



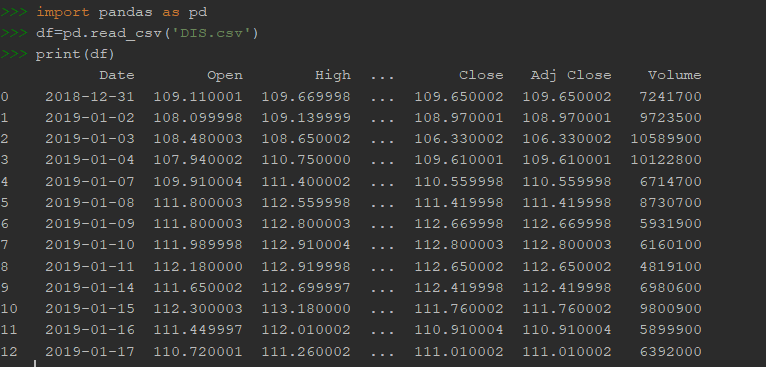
Symetric difference returnează un set cu elementele necomune celor două seturi. Am analizat din ambele puncte de vedere, și la haine, și la petshop; rezultatul fiind același :



Union returnează un set reprezentând reuniunea a 2 seturi (în cazul de față, rezultatul e la fel cu cel de la symetric difference):

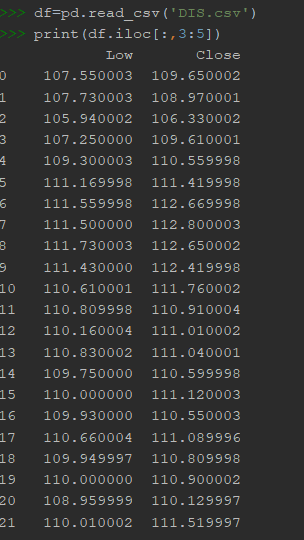


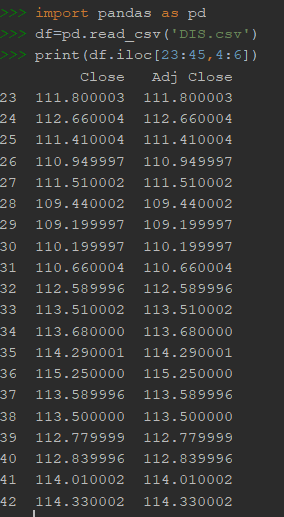
**Importul unui fișier csv sau json în pachetul pandas**

 Fișierul importat este unul csv, acesta conține date statistice ale companiei Walt Disney și reprezintă valori corespunzătoare perioadei 1.1.2019-26.5.2019.

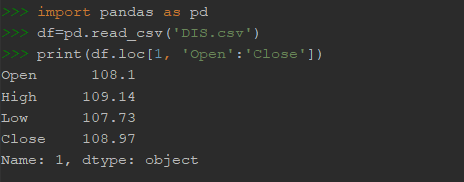
**Accesarea datelor cu loc și iloc**

Afișarea doar a anumitor coloane din datele inițiale cu ajutorul iloc.

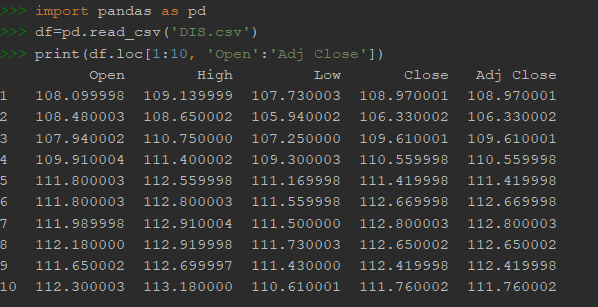


 Afișarea anumitor liste și anumitor înregistrări cu iloc

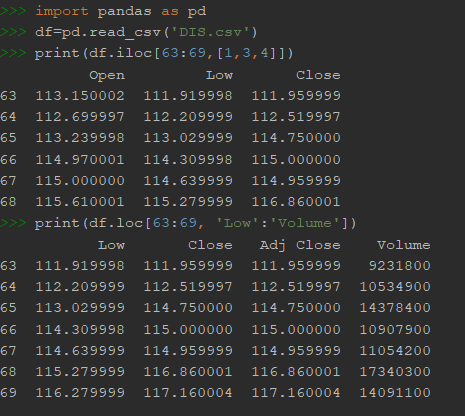
Accesarea unui anumit rând, cu specificarea coloanelor ce vor fi afișate prin loc:



Afișarea mai multor rânduri cu numărul de coloane specificate prin loc:

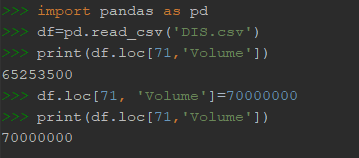


Afișarea cu iloc și loc a unor coloane din dataframe inițial:

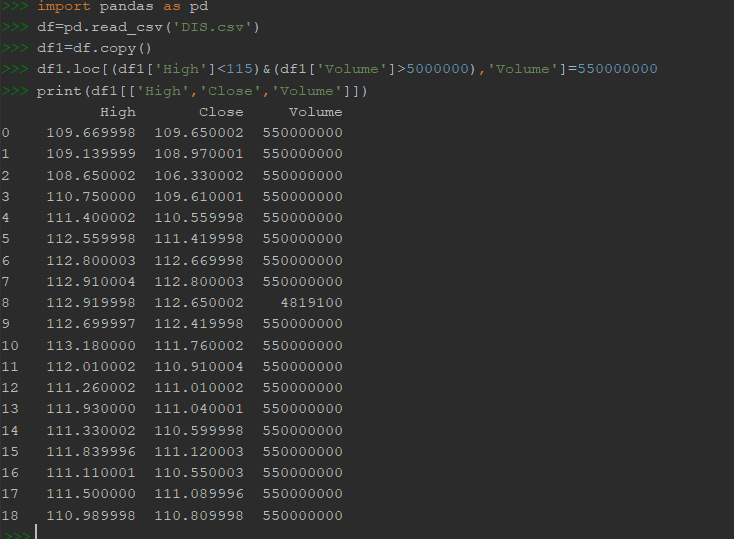


**Modificarea datelor în pachetul pandas**

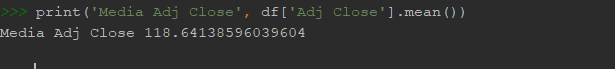
Modificarea volumului pentru înregistrarea numărul 71:



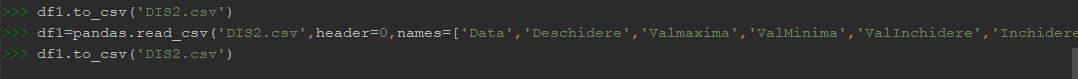
Modificarea volumului dacă valoarea High < 115 și Volumul > 5000000, toate acestea introduse într-un dataframe nou:

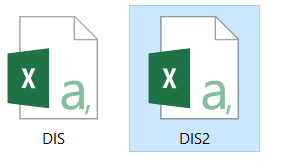


Calcularea valorii medii pentru Adj Close în cadrul datelor inițiale:

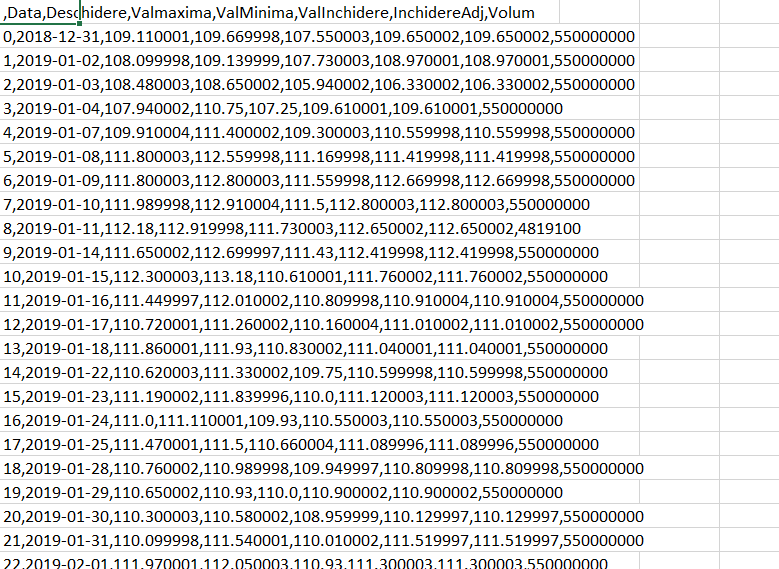


Am salvat dataframeul creat anterior (df1) într-un fișier csv:

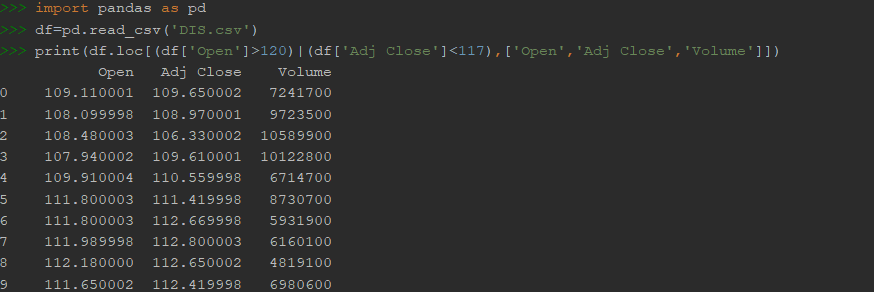


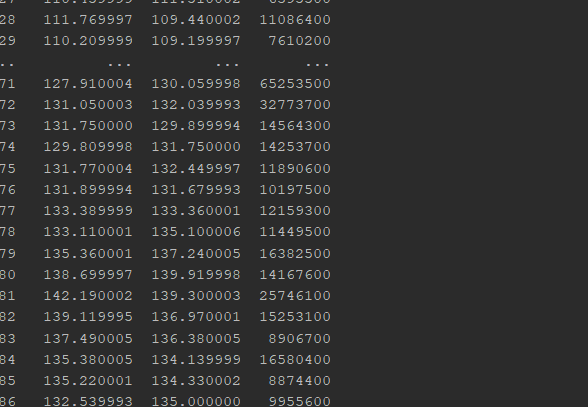


Și am redenumit coloanele în română:

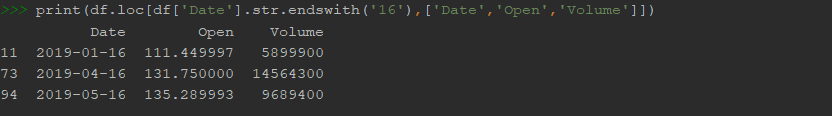


Să se afișeze înregistrările pentru care valoarea Open > 120 sau Adj Close < 117:

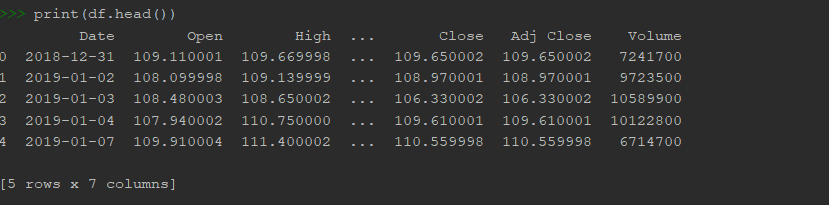




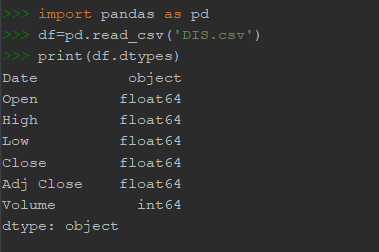
Afișarea înregistrărilor din zilele de 16 existene în baza de date:



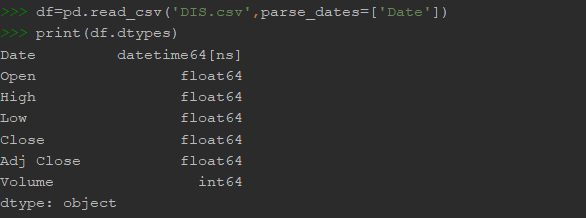
Afișarea primelor 5 înregistrări cu ajutorul head:



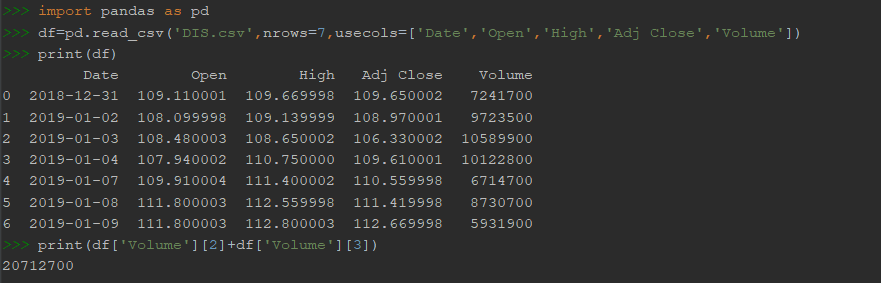
Identificarea tipului de date al fiecărei variabile:



Și modificarea formatului de dată:

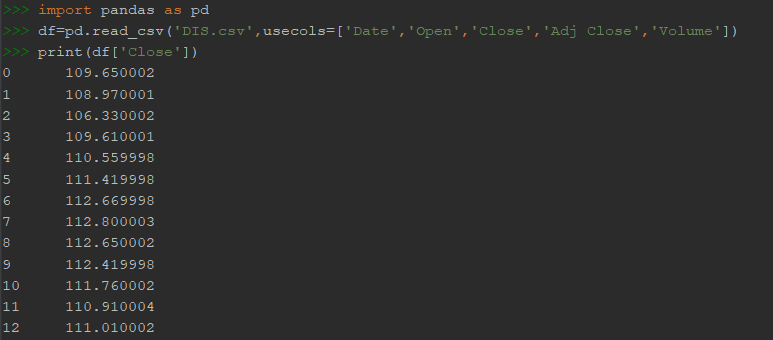


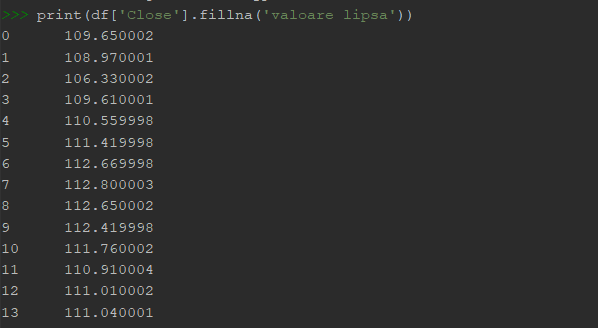
Însumarea a 2 valori diferite:

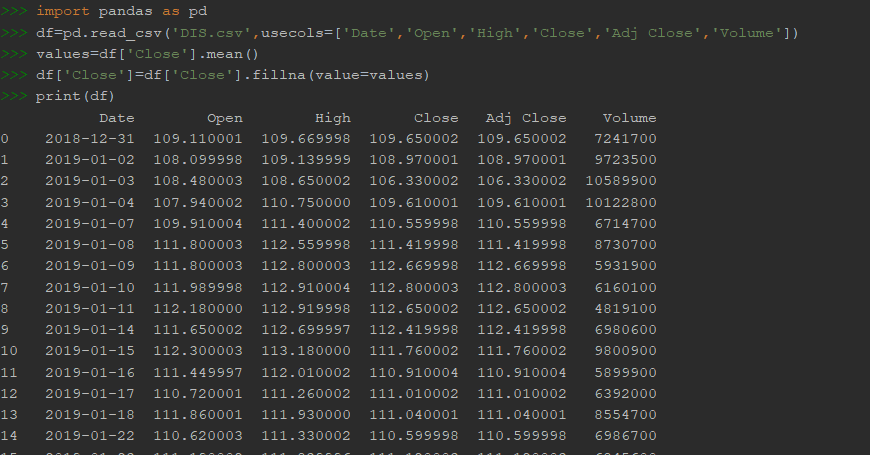


**Tratarea valorilor lipsă**

Am luat în considerare coloana Close și am analizat-o cu ajutorul fillna. Nu au existat valori lipsă:

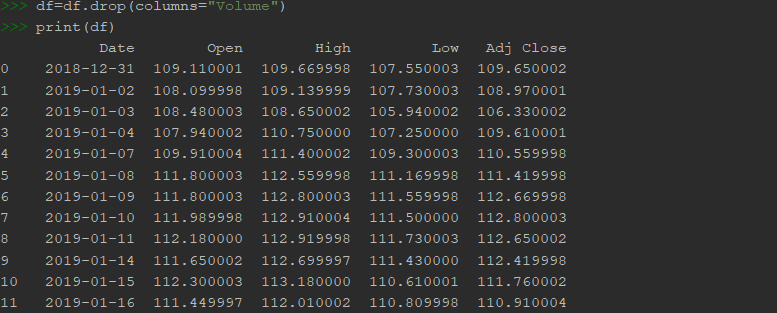




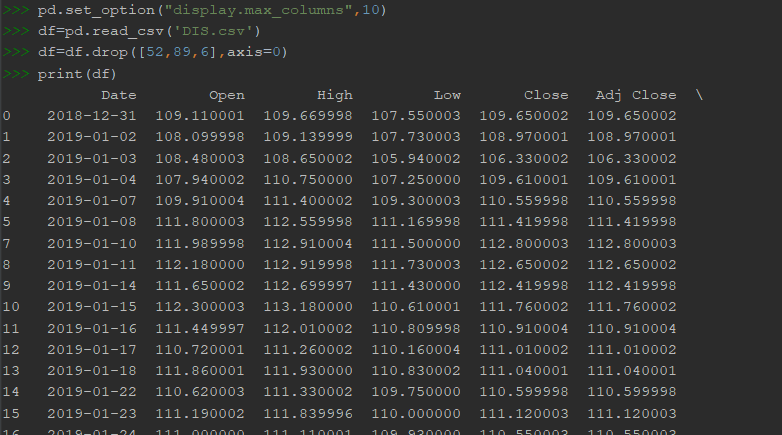


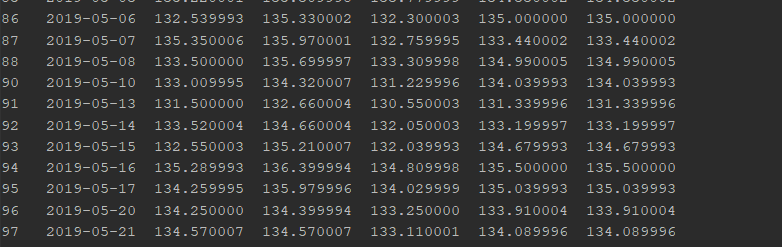
**Ștergerea de coloane și înregistrări**

Eliminarea unei coloane cu opțiunea columns:



Eliminarea de înregistrări:



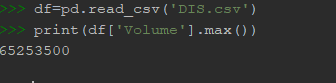


Ștergerea primelor 7 linii prin iloc:

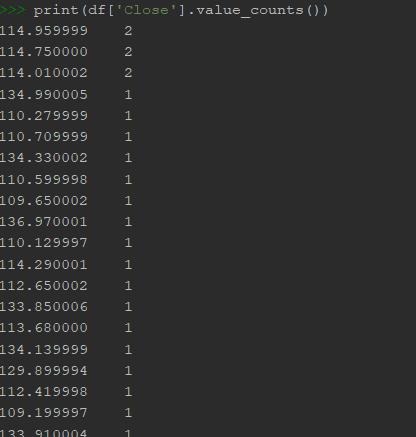


**Prelucrări statistice, gruparea și agregarea datelor în pachetul pandas**

Valoarea maximă a volumului:

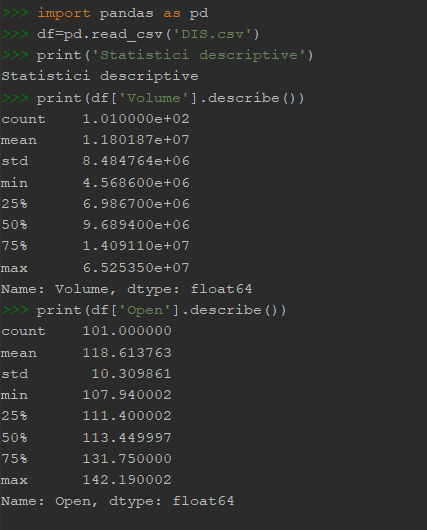


Numărul de valori de închidere în coloana Close

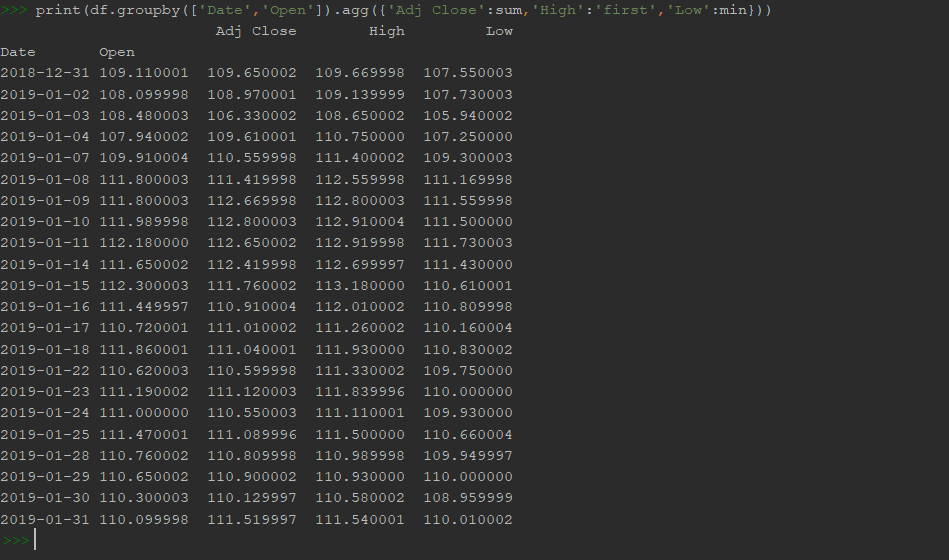


Eliminarea duplicatelor din coloana Adj Close



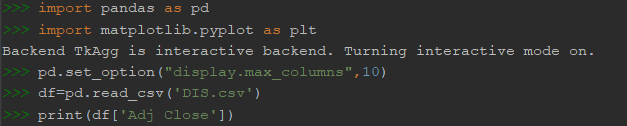
 Folosirea funcției describe pentru date statistice

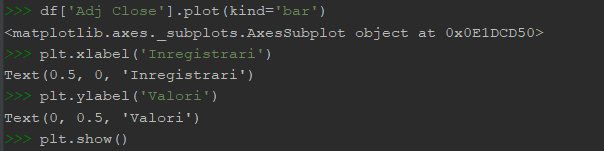
Gruparea și agregarea datelor în funcție de Date și Open



**Reprezentarea grafică a datelor cu pachetul matplotlib**

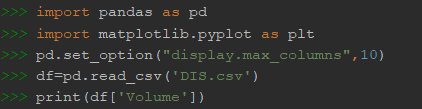
Am folosit coloana Adj Close, iar în reprezentare se observă numărul de înregistrări pe OX și valorile aferente pe OY

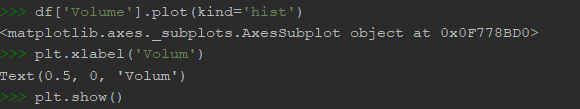
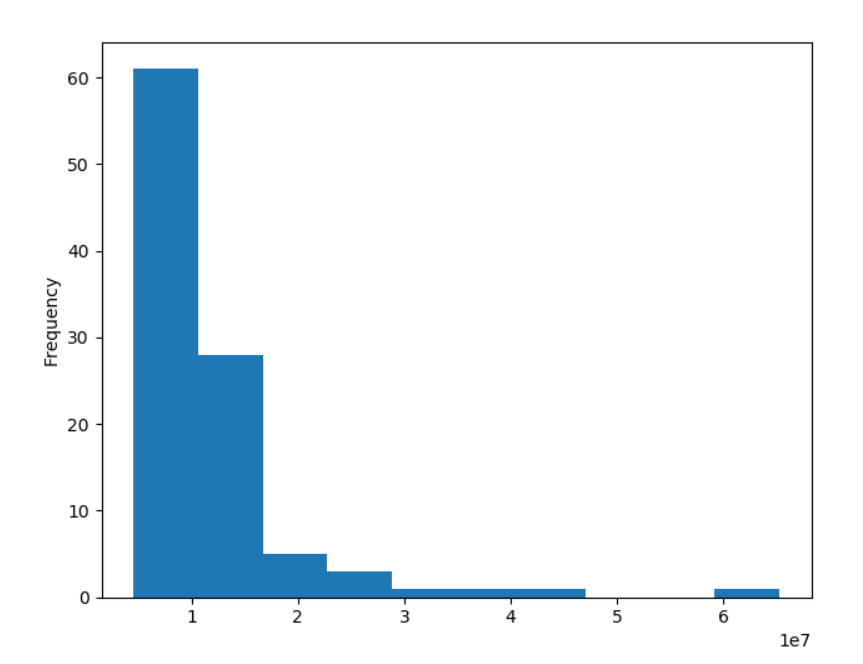






Vom realiza și o histogramă.





Se observă frecvența de apariție a volumelor respective.

Vom face un grafic cu gruparea și sortarea datelor prin matplotlib. Acesta va avea 5 înregistrări, echivalentul zilei de 15 din lunile existente în baza de date, vor fi sortate în baza valorii Open și vor afișa suma Volumului lor.

