

PROIECT BUSINESS INTELLIGENCE

Pentru realizarea proiectului am folosit o abordare agilă, care ne-a permis să lucrăm într-un mod flexibil și adaptabil. Am început prin explorarea și curățarea datelor, apoi am trecut la combinarea informațiilor din surse diferite și, pe măsură ce lucrurile au prins contur, am construit vizualizări în Tableau pentru a testa ce informații pot fi extrase. Pe parcurs, am făcut ajustări ori de câte ori a fost nevoie, în funcție de calitatea datelor și de ceea ce am descoperit. Acest mod de lucru ne-a ajutat să rămânem eficienți, să ne adaptăm din mers și să obținem rezultate relevante și bine structurate.

1. DESCRIEREA PROBLEMEI DE BUSINESS PENTRU CARE S-A REALIZAT PROIECTUL

În ultimii ani, tot mai multă atenție se îndreaptă spre companiile care au un impact semnificativ asupra economiei locale, mai exact cele care generează venituri ridicate, obțin profituri stabile și contribuie la crearea de locuri de munca. Pornind de la această idee, în acest proiect ne-am propus să analizăm cele mai importante 10 firme din județul Suceava, în funcție de performanța lor financiară în perioada 2020-2023. Prin această analiză am urmărit să răspundem la întrebări precum:

- Care sunt cele mai performante 10 firme din județul Suceava în perioada 2020–2023?
- Cum a evoluat numărul de angajați în aceste firme pe parcursul celor patru ani?
- Cum au evoluat ele de-a lungul celor patru ani?
- Ce firmă domină piața locală în ceea ce privește cifra de afaceri și profitul total?

Scopul final este de a obține o imagine clară asupra structurii economice locale și asupra tendințelor de dezvoltare ale firmelor de top din județ. Pentru această analiză am folosit Tableau Prep pentru prelucrarea și curățarea datelor, iar Tableau Desktop pentru vizualizarea și interpretarea acestora.

2. PREGĂTIREA DATELOR

2.1 Conectarea la seturile de date

Pentru realizarea acestui proiect, am utilizat mai multe seturi de date disponibile public pe platforma data.gov.ro, din surse oficiale precum Ministerul Finanțelor Publice și Oficiul Național al Registrului Comerțului (ONRC).

Datele utilizate au fost în primul rând situații financiare anuale ale firmelor, pentru anii 2020, 2021, 2022 și 2023, structurate în mai multe fișiere de tip txt: *web_bl_bs_sl_an[an]*, *web_ir_an[an]*, *web_uu_an[an]*. Aceste fișiere conțin informații legate de bilanț (active/pasive), conturi de profit și pierdere și alte date utile (de exemplu, număr de angajați, cod CAEN, etc.).

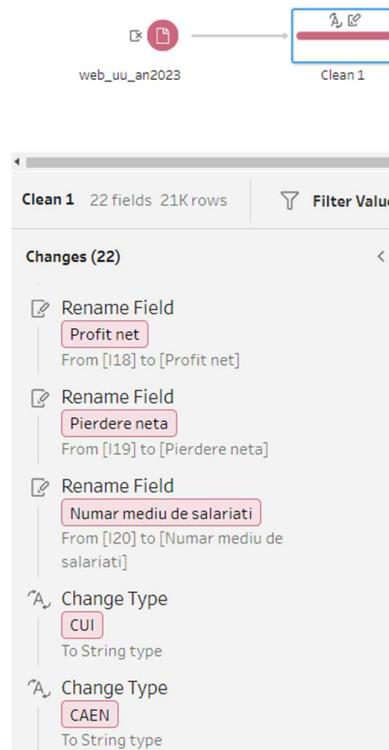
Am mai utilizat și un fișier care conține date specifice despre firme cum ar fi, codul unic de identificare (CUI), denumirea firmei, localitatea și alte informații de identificare. Acest fișier se numește *od_firme.csv*.

2.2 Crearea setului final de date consolidat

Toate aceste fișiere au fost încărcate în Tableau Prep, unde am aplicat, pentru fiecare în parte, un pas de tip Clean în care am realizat mai multe transformări esențiale.

În primul rând, am uniformizat denumirile coloanelor, astfel încât aceleași tipuri de informații să fie denumite identic în toate fișierele. Am redenumit câmpurile în funcție de conținutul lor real, pentru a reflecta corect tipul de date și a facilita analiza ulterioară. De exemplu, coloanele originale i1, i2, i10 sau i18 au fost redenumite în: active imobilizate - total, active circulante - total, venituri totale, cheltuieli totale, profit net, pierdere netă, număr mediu de salariați, caen etc.

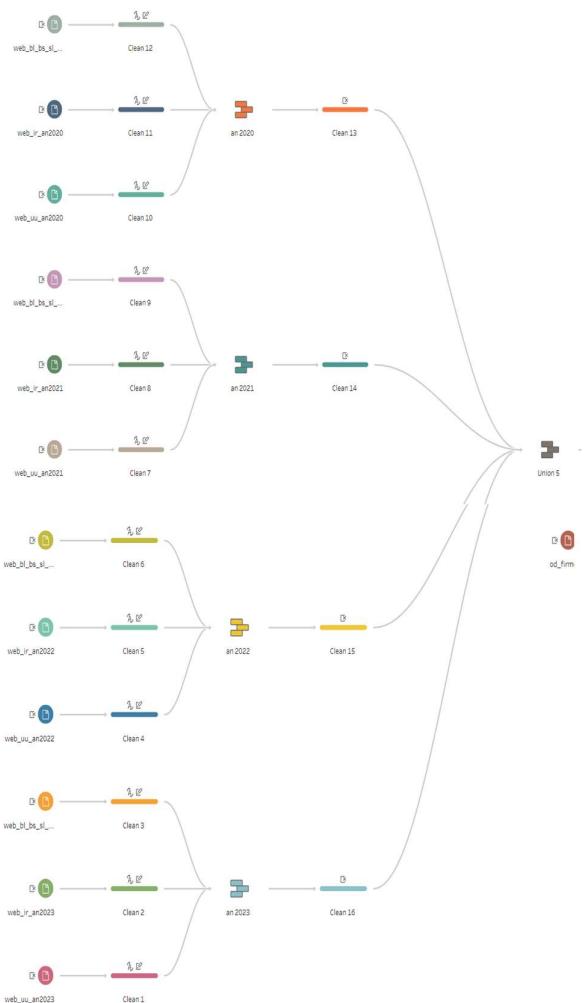
În continuare, am setat corect tipurile de date: câmpurile de identificare precum CUI și CAEN au fost convertite în format text (string), pentru a putea realiza joinuri corecte între seturi.



După curățarea individuală a fiecărui fișier pentru anii 2020, 2021, 2022 și 2023, am realizat câte un Union intern pentru fiecare an între cele trei fișie. Astfel, am reunit toate datele relevante pentru fiecare firmă într-un singur tabel anual (an 2020, an 2021, etc.).

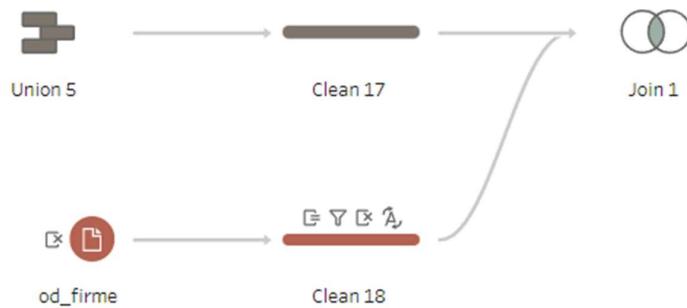
În continuare am aplicat un nou pas de tip Clean pentru fiecare Union obținut, în care am eliminat câmpul Table Names, adăugat automat de Tableau Prep în timpul combinării surselor. Această curățare finală a fost necesară pentru a păstra doar câmpurile relevante pentru analiză.

După ce am obținut cele patru tabele anuale, am realizat încă un Union, prin care am combinat toate aceste tabele într-un set de date consolidat ce conține informații pentru perioada 2020–2023. Această operațiune a permis centralizarea datelor într-o singură sursă de analiză. Imaginea de mai jos ilustrează această etapă de combinare și curățare.



2.3 Îmbinarea datelor financiare cu identificatorii ONRC

După consolidarea datelor financiare pentru perioada 2020–2023, am realizat un Join final cu fișierul od_firme.csv, care conține informații detaliate despre firme: denumire, localitate, data înmatriculare, formă juridică etc. Acest Join a fost făcut pe câmpul CUI, utilizând un Inner Join, pentru a păstra doar firmele care se regăsesc în ambele seturi de date.



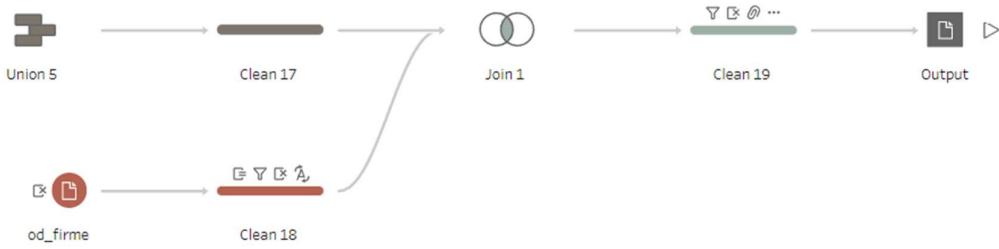
Înainte de realizarea join-ului final, în pasul Clean 18, am aplicat mai multe transformări asupra fișierului od_firme. Mai întâi, am convertit câmpul CUI în format text, pentru a asigura compatibilitatea cu celelalte surse de date. Apoi, am creat un câmp nou, denumit Denumire_clean, pentru a standardiza formatul denumirii firmelor. În final, am filtrat rezultatele, în funcție de județ, pentru a obține doar analiza firmelor din Suceava.

Clean 18 0 fields | Filter Values...

Changes (4)

- Change Type: CUI To String type
- Calculated Field: Denumire_clean
REPLACE(REPLACE(TRIM([DENUMIRE]), ",,,"), ",,", ",")
- Remove Field: DENUMIRE
- Filter: Selected Values: ADR_JUDET
Keep only: "Suceava"

După realizarea Join-ului, am aplicat un ultim pas de curățare în cadrul Clean 19. În această etapă, am eliminat câmpurile considerate irelevante pentru analiză. Am continuat prin aplicarea unor filtre pentru a exclude înregistrările care conțineau valori lipsă în câmpurile esențiale, precum Număr mediu de salariați, Venituri, Creanțe sau Datorii. La final, am redenumit coloana Table Names în AN și am setat-o ca dimensiune temporală, pentru a marca anul.



| Clean 19 29 fields 47 rows | | Filter Values. | |
|--|----------------------------|---|--------------------------|
| Changes (26) | | | |
| <input type="checkbox"/> Remove Field | PROVIZIOANE | <input type="checkbox"/> Filter: Null Values | ... |
| <input type="checkbox"/> Remove Field | COD_INMATRICULARE | <input type="checkbox"/> Keep only: non-null values | |
| <input type="checkbox"/> Remove Field | DATA_INMATRICULARE | <input type="checkbox"/> Filter: Null Values | Stocuri |
| <input type="checkbox"/> Remove Field | EUID | <input type="checkbox"/> Keep only: non-null values | Create |
| <input type="checkbox"/> Remove Field | FORMA_JURIDICA | <input type="checkbox"/> Filter: Null Values | Casa si conturi la bani |
| <input type="checkbox"/> Group Values | Table Names | <input type="checkbox"/> Keep only: non-null values | CHELTUIELI IN AVANS |
| | 4 values replaced | <input type="checkbox"/> Filter: Null Values | VENITURI IN AVANS |
| <input type="checkbox"/> Rename Field | AN | <input type="checkbox"/> Keep only: non-null values | PROVIZIOANE |
| | From [Table Names] to [AN] | <input type="checkbox"/> Filter: Null Values | Numar mediu de salariati |
| <input type="checkbox"/> Change Role | ADR_JUDET | <input type="checkbox"/> Keep only: non-null values | ADR_BLOC |
| | To State/Province | <input type="checkbox"/> Filter: Null Values | ADR_SCARA |
| <input type="checkbox"/> Remove Field | ADR_SCARA | <input type="checkbox"/> No option selected | |
| <input type="checkbox"/> Remove Field | ADR_ETAJ | <input type="checkbox"/> Remove Field | ADR_SECTOR |
| <input type="checkbox"/> Remove Field | ADR_APARTAMENT | | |
| <input type="checkbox"/> Filter: Null Values | ADR_COD_POSTAL | | |
| | Keep only: non-null values | | |
| <input type="checkbox"/> Remove Field | ADR_COMPLETARE | | |

3. DATA VISUALIZATION & ANALYSIS

a) Câmp calculat, Filtre, Marcatori personalizați

Rank Profit per salariat

Results are computed along Denumire firme.

```
RANK_UNIQUE(([Profit per salariat]), 'desc')
```

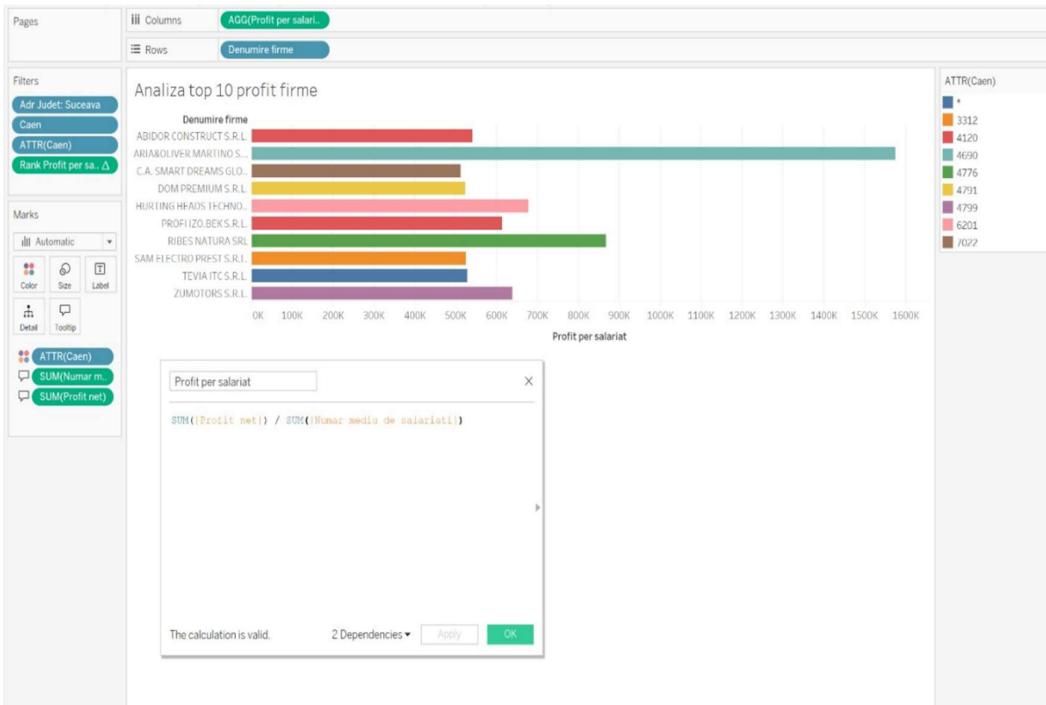
The calculation is valid.

Default Table Calculation

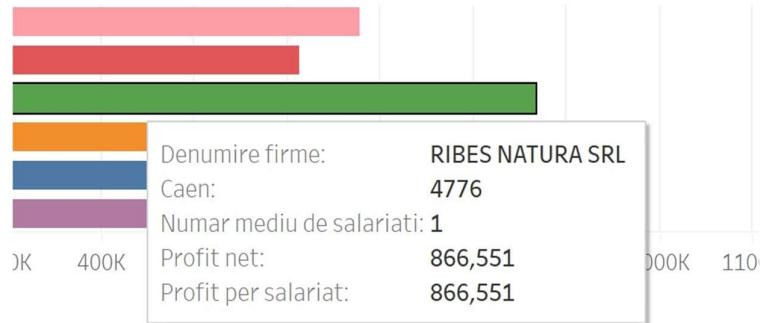
1 Dependency ▾ Apply OK

Pentru a analiza eficiența firmelor din Suceava, am realizat un câmp calculat denumit *Profit per salariat*, folosind formula: $\text{SUM}([\text{Profit net}]) / \text{SUM}([\text{Numar mediu de salariati}])$. Pe baza acestuia, am creat un alt doilea câmp, *Rank Profit per salariat*, prin funcția *RANK_UNIQUE*, pentru a determina ordinea firmelor în funcție de acest indicator.

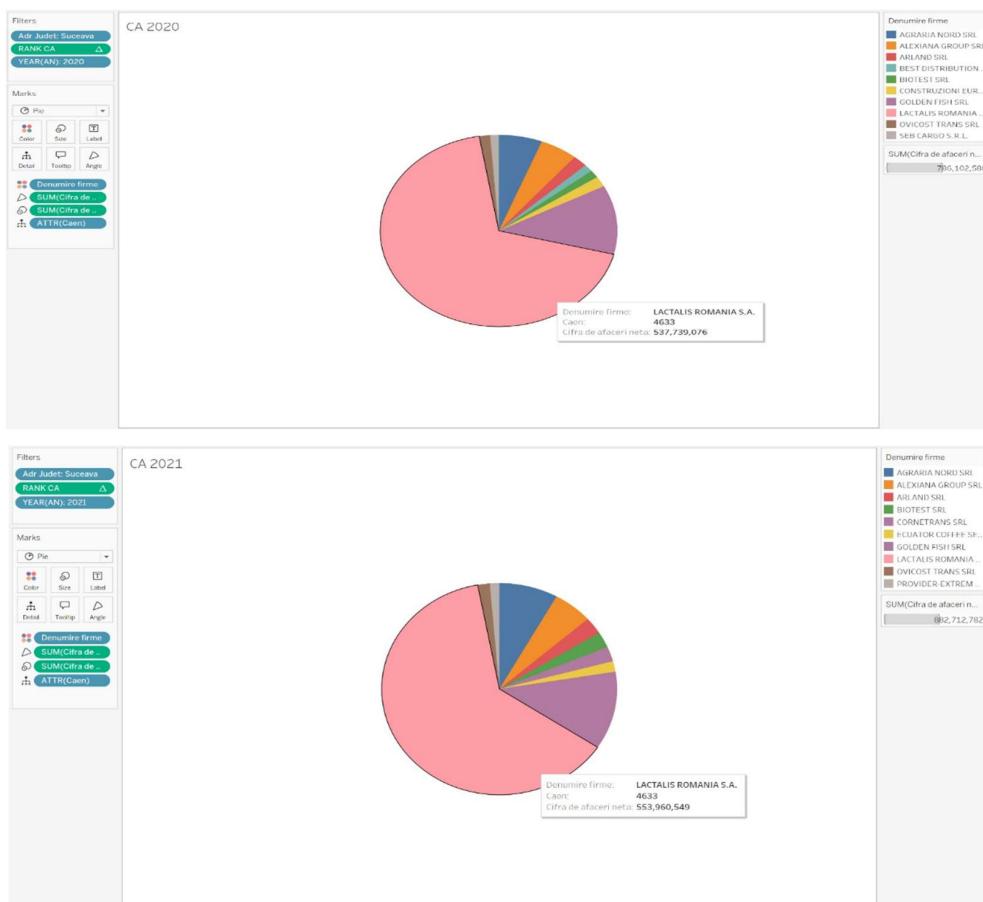
Am aplicat un filtru care păstrează doar top 10 firme din județ. Pentru o reprezentare clară, am utilizat marcatori personalizați: culoare în funcție de codul CAEN, etichete cu valoarea profitului per salariat și detaliu pentru denumirea firmei. Graficul rezultat este un bar chart care evidențiază cele mai performante firme din județ în perioada 2020–2023.

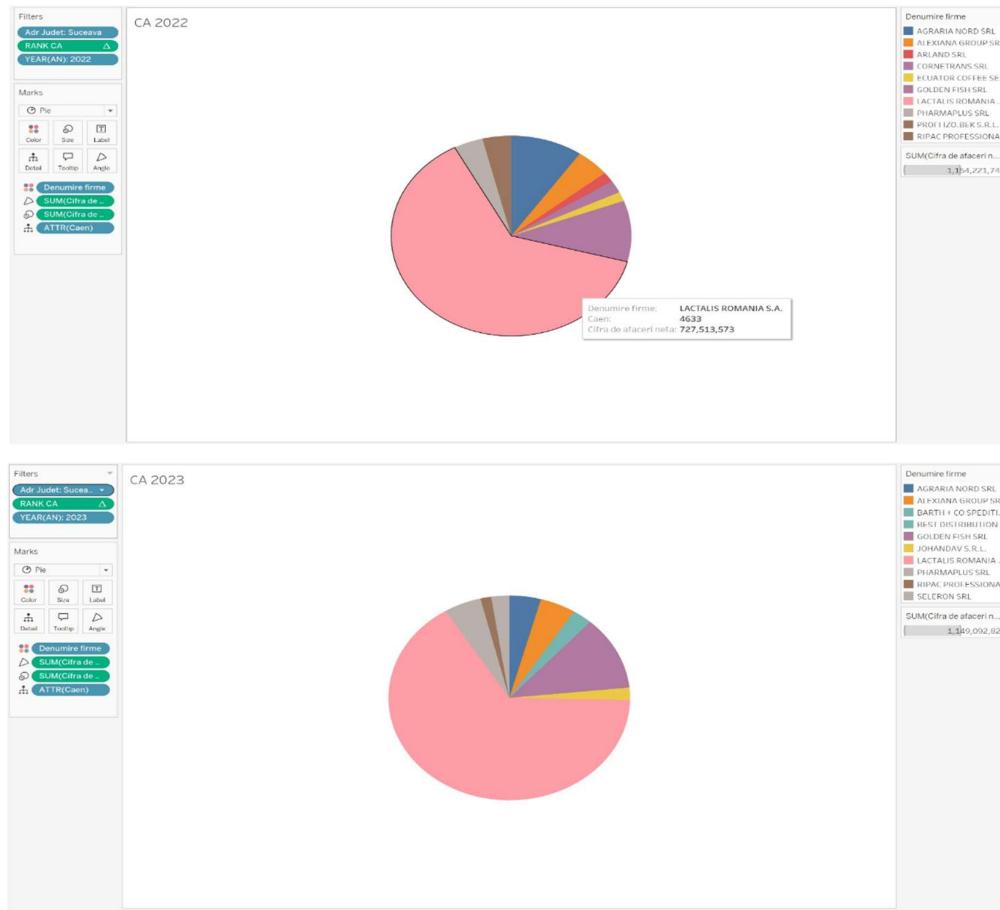


Tooltip-ul oferă informații detaliate despre firma selectată, incluzând denumirea, codul CAEN, numărul mediu de salariați, profitul net și profitul pe salariat. Aceasta permite o vizualizare clară și concisă a datelor specifice fără a încărca vizual graficul principal.



Pentru a evidenția evoluția cifrei de afaceri în județul Suceava, am realizat grafice de tip distribuție pentru fiecare an din perioada 2020–2023. Acestea arată clar că firma Lactalis Romania S.A. deține o cotă importantă din total, iar cifra ei de afaceri crește constant în acești ani. În schimb, restul firmelor au pondere mult mai mică, ceea ce arată o piață concentrată, dar cu o ușoară creștere a diversității firmelor active.





b) Tip de grafic din meniul ShowMe

Graficul prezentat arată evoluția numărului mediu de angajați în județul Suceava pentru fiecare an analizat. În fiecare bară sunt reprezentate top 10 firme din acel an, iar segmentele colorate corespund firmelor care au făcut parte în mod specific din acel top în anul respectiv.



Această vizualizare ne permite să urmărim atât evoluția generală a numărului de angajați, cât și schimbările în topul firmelor, evidențiind dinamica forței de muncă în timp.

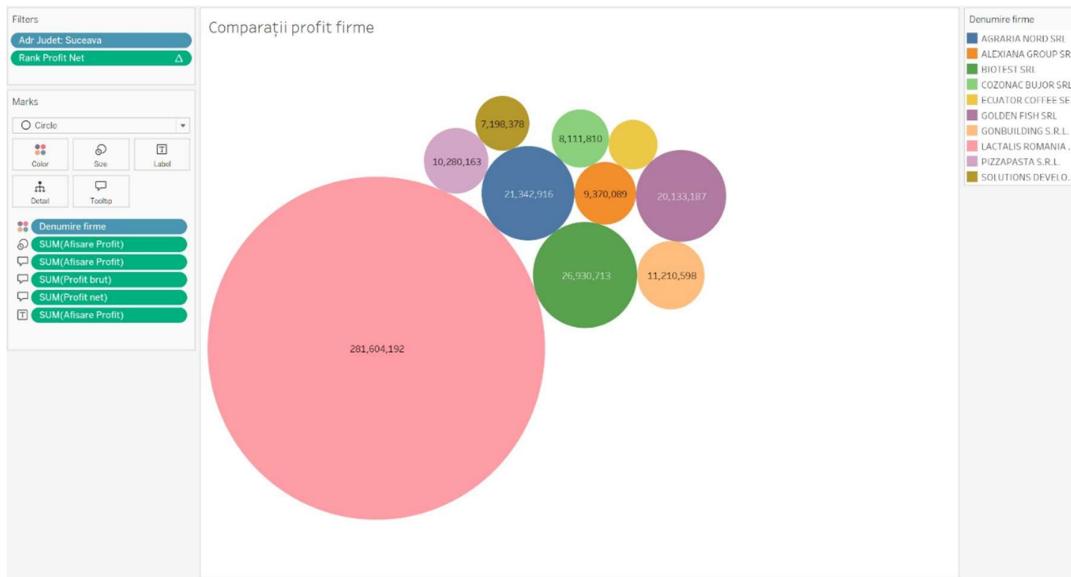
c) Crearea diagramelor cu date categorice și numerice

În realizarea vizualizărilor, am utilizat atât date categorice, cât și numerice. Dintre datele categorice s-au evidențiat: denumirea firmei, codul CAEN, anul (AN) și localitatea, care au fost folosite pentru etichetare, segmentare și filtrare. În ceea ce privește datele numerice, am folosit profitul net, cifra de afaceri netă, cheltuielile totale, numărul mediu de salariați și profitul per salariat. Aceste câmpuri numerice au stat la baza calculelor, comparațiilor și analizelor de tip trend, clustering sau ranking. Prin combinarea acestor două tipuri de date, am putut crea grafice relevante care oferă o imagine completă asupra performanței firmelor analizate.

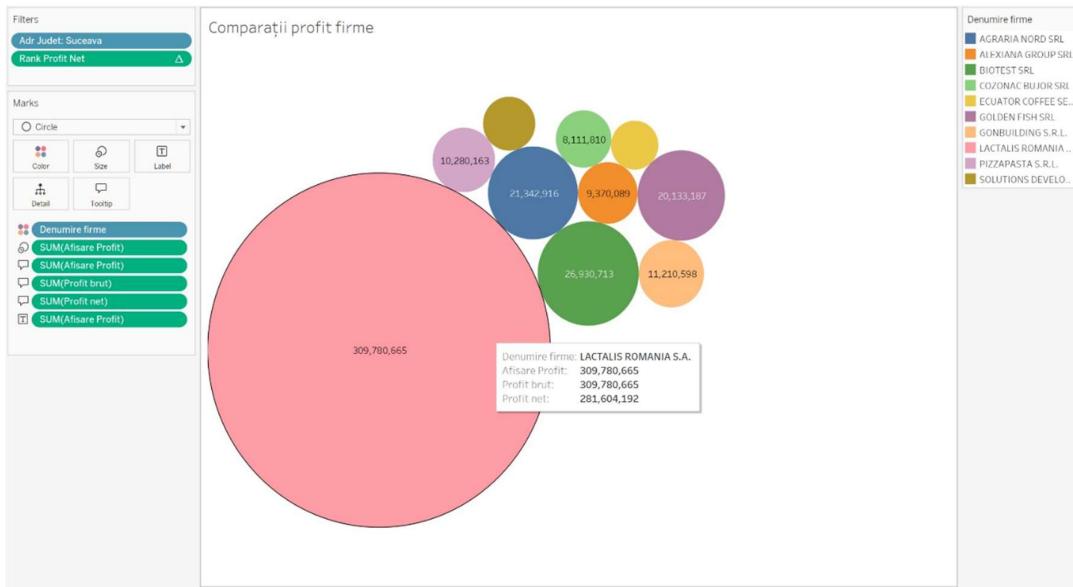
d) Action

Pentru explorarea datelor, am utilizat funcționalitatea Action. Astfel, utilizatorul poate plasa cursorul pe un anumit element din grafic pentru a afișa detalii suplimentare relevante despre firmă, cum ar fi denumirea firmei, afișarea profitului și valorile corespunzătoare fiecărui tip de profit.

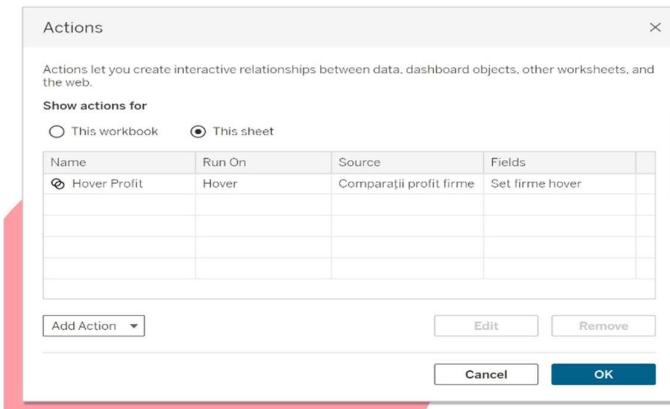
Graficul inițial prezintă valorile profitului net pentru firmele din Suceava, reprezentate prin cercuri de dimensiuni proporționale cu mărimea profitului.



Prin poziționarea cursorului asupra oricărui cerc, se activează acest action creat pentru a arăta valoarea profitului brut.

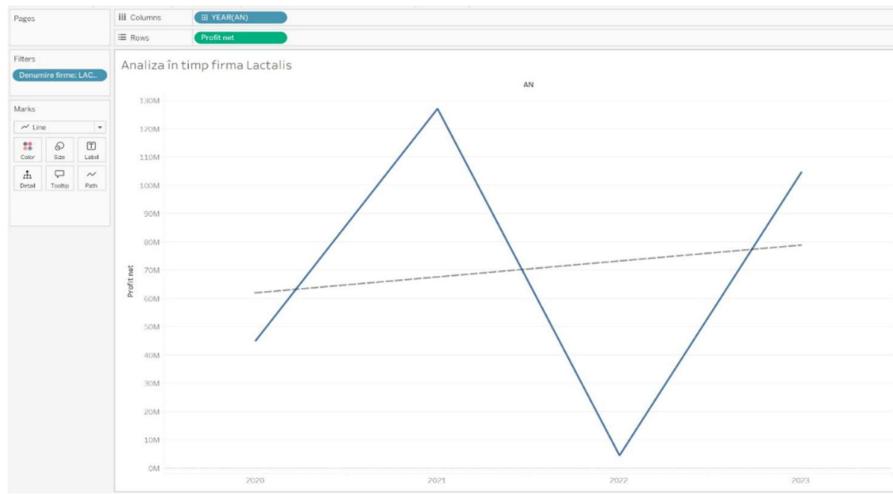


Am configurat această funcționalitate în setările de acțiune, creând o acțiune numită „Hover Profit” care se activează atunci când pozitionăm cursorul peste elementele din foaia „Comparări profit firme” și folosește un grup special de firme pentru afișarea informațiilor la hover.



e) Analiza de trend pentru evolutia in timp a datelor

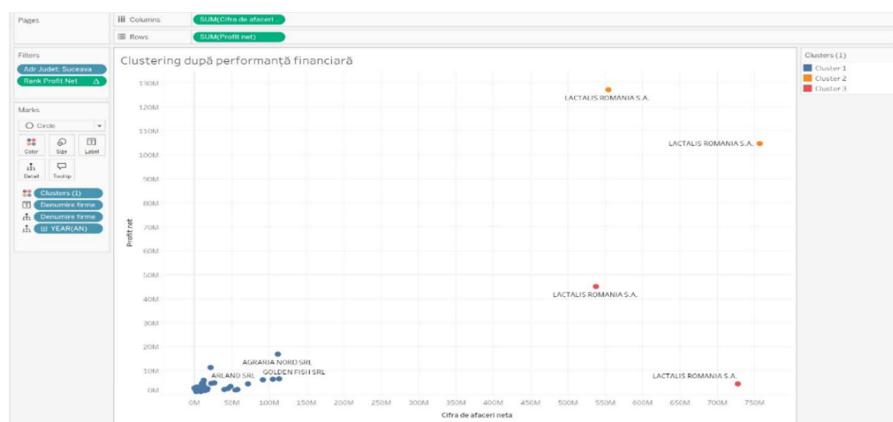
Graficul prezintă evoluția profitului net al firmei Lactalis România pe parcursul anilor analizați. Se observă variații semnificative în profit, însă în ansamblu, firma rămâne liderul județului în ceea ce privește performanța financiară, având cel mai ridicat profit net comparativ cu celelalte companii.



f) Data mining algorithm

Clasificarea firmelor din județul Suceava este ilustrată printr-un grafic ce conține trei clustere diferite, pe baza a două variabile financiare: cifra de afaceri netă și profitul net.

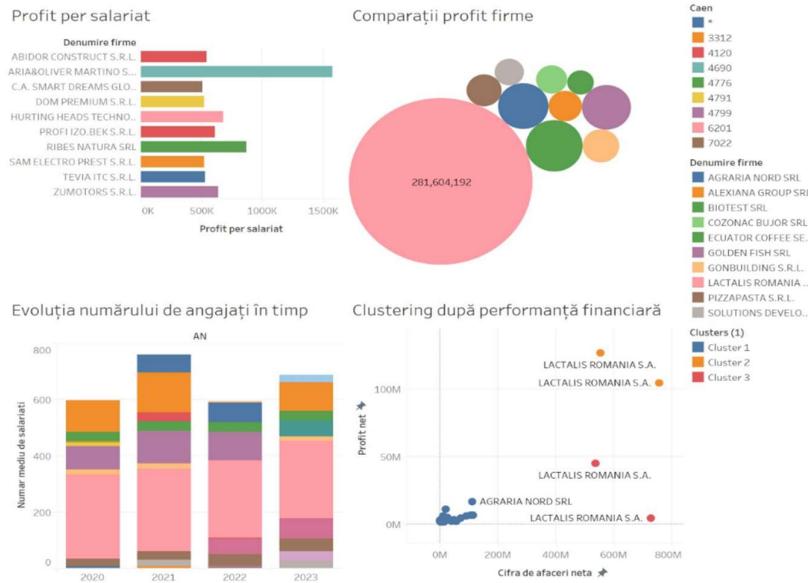
- Cluster 1 (albastru) reunește majoritatea firmelor, cu valori mai mici atât ale cifrei de afaceri, cât și ale profitului net, reprezentând segmentul de companii mici și mijlocii.
- Cluster 2 (portocaliu) include firme cu valori intermediare, ce au o cifră de afaceri și un profit net mai ridicat, comparativ cu clusterul 1.
- Clusterul 3 (roșu) cuprinde firmele de top, cu valori foarte mari pentru ambele variabile, reprezentând liderii de piață din județ. Astfel, se observă că firma Lactalis România S.A. apare în acest cluster, având o performanță financiară ridicată.



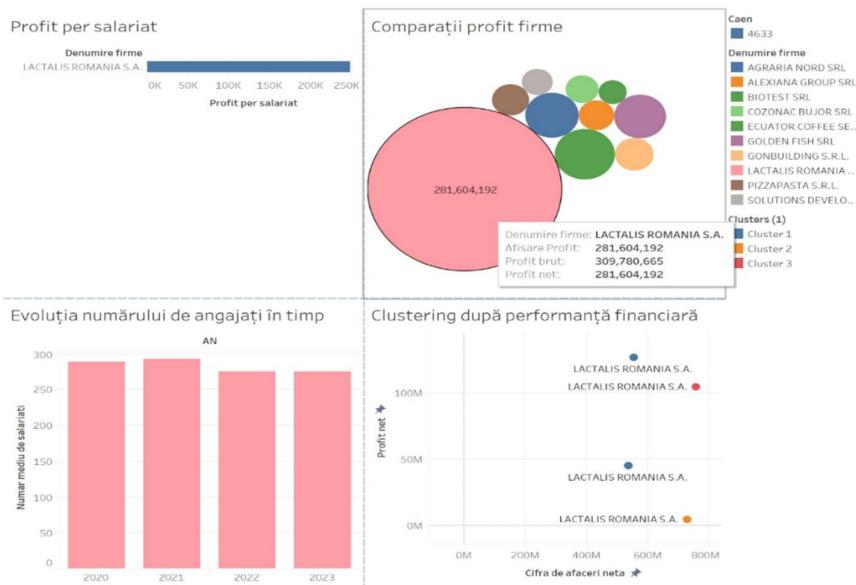
Această analiză ne oferă o imagine clară asupra diferențelor niveluri de performanță financiară ale firmelor din județ, evidențiind clar liderii și diferențele dintre companii.

4. CREARE DASHBOARD CU DATE LEGATE

Prima imagine prezintă patru grafice complementare care analizează firmele din top 10 în funcție de profitul net. Fiecare grafic oferă informații diferite: profitul pe angajat, comparația profitului total al firmelor, evoluția numărului de angajați pe parcursul anilor și gruparea firmelor după performanță financiară.



În cea de-a doua imagine este ilustrată funcționalitatea interactivă a dashboard-ului. Atunci când cursorul este poziționat peste o bulă din graficul „Comparații profit firme”, se afișează automat informații detaliate despre firma selectată, actualizând în același timp datele din celelalte grafice.



5. CREAREA UNEI POVEŞTI (STORY)

Pentru a oferi o prezentare clară și structurată a analizei firmelor din județul Suceava, am realizat un story care reunește principalele grafice și etape ale investigației. Story-ul conține o serie de pagini care evidențiază diverse aspecte ale performanței financiare. Această metodă facilitează o înțelegere progresivă și interactivă a datelor, permitând utilizatorului să parcurgă pas cu pas concluziile și să exploreze informațiile într-un mod intuitiv.

