

UNIVERSITATEA BABEȘ-BOLYAI  
FACULTATEA DE ȘTIINȚE ECONOMICE ȘI GESTIUNEA AFACERILOR

## **PROIECT BUSINESS INTELLIGENCE**

Serețan Daria Emanuela  
Silaghi Roxana Cristina  
Vilău Paula Irina

Grupa 6  
Informatică Economică

## **I. Descriere Problema de Business pentru care s-a realizat proiectul**

În cadrul acestui proiect am analizat situațiile financiare ale mai multor firme prin combinarea unor seturi de date preluate de pe site-ul Ministerului Finanțelor Publice, <https://data.gov.ro/dataset>. Scopul acestei lucrări este de a prelucra datele firmelor care au ca domeniu de activitate realizarea serviciilor în tehnologia informației din județul Cluj, cu codul CAEN 620. Acest grup cuprinde:

CAEN 6201 - Activități de realizare a soft-ului la comandă (software orientat client)

CAEN 6202 - Activități de consultanță în tehnologia informației

CAEN 6203 - Activități de management (gestiune și exploatare) a mijloacelor de calcul

CAEN 6209 - Alte activități de servicii privind tehnologia informației

Elementele firmelor pe care am ales să ne concentrăm sunt:

- Cifra de afaceri netă
- Activele totale
  - Active imobilizate
  - Active circulante
- Profitul net
- Veniturile totale
- Datorii
- Cheltuieli totale
- Număr mediu de salariați

## **II. Pregătirea datelor**

Pentru analiza datelor, avem nevoie de mai multe seturi de date:

- web\_uu\_an2017 (.txt);
- web\_uu\_an2018 (.txt);
- web\_uu\_an2019 (.txt);
- web\_uu\_an2020 (.txt, .csv);
- web\_ir\_an2017 (.txt);
- web\_ir\_an2018 (.txt);
- web\_ir\_an2019 (.txt);
- web\_ir\_an2020 (.txt, .csv);
- web\_bl\_bs\_sl\_an 2017 (.txt);
- web\_bl\_bs\_sl\_an 2018 (.txt);
- web\_bl\_bs\_sl\_an 2019 (.txt);
- web\_bl\_bs\_sl\_an 2020 (.txt, .csv);
- Date\_identificare\_2020\_cluj (.txt, .csv);

Antetul coloanelor din situațiile financiare (primele trei tipuri de seturi de date) sunt explicate în .csv-ul care acompaniază fișierul text unde sunt informațiile propriu-zise (am preluat doar câte un fișier .csv din fiecare tip deoarece este același pentru toți cei patru ani).

În sensul procesării situațiilor financiare pentru societățile comerciale care depun bilanț lung, bilanț scurt și raportări anuale din cei patru ani, este necesar un union pentru a aduce toate datele din toți anii la un loc și pentru a putea schimba antetul conform fișierului .CSV.

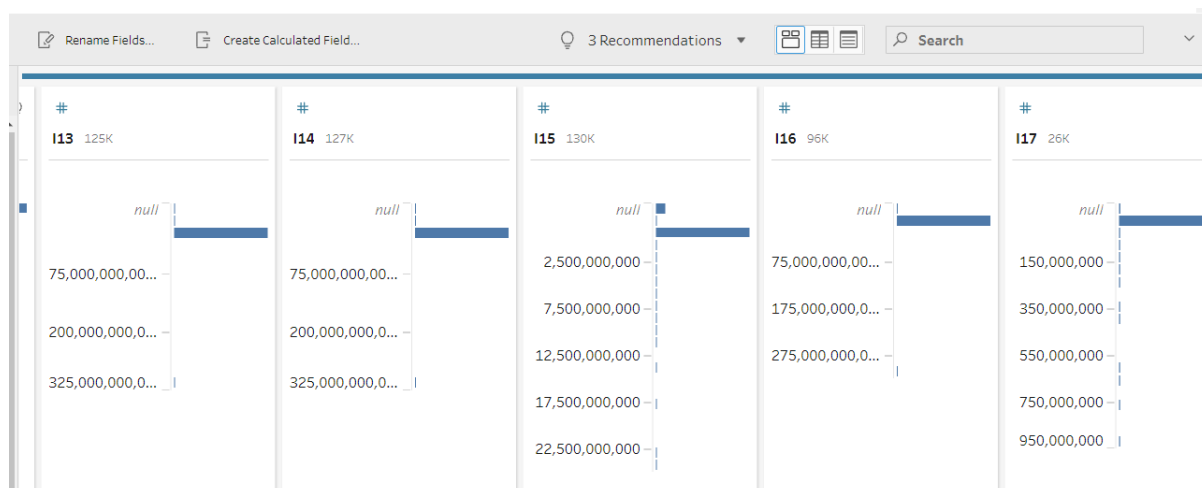


Figura 1. Antetele coloanelor din setul de date nu au nume sugestive

	A	B	C	D	E
1	CUI;CUI				
2	CAEN;CAEN				
3	ACTIVE IMOBILIZATE - TOTAL;i1				
4	ACTIVE CIRCULANTE - TOTAL din care;;i2				
5	Stocuri;i3				
6	Creante;i4				
7	Casa si conturi la banci;i5				
8	CHELTUIELI IN AVANS;i6				
9	DATORII;i7				
10	VENITURI IN AVANS;i8				
11	PROVIZIOANE;i9				
12	CAPITALURI - TOTAL din care;;i10				
13	Capital subscris varsat;i11				
14	Patrimoniul regiei;i12				
15	Cifra de afaceri neta;i13				
16	VENITURI TOTALE;i14				
17	CHELTUIELI TOTALE;i15				
18	Profitul brut;i16				
19	Pierdere bruta;i17				
20	Profitul net;i18				
21	Pierdere neta;i19				
22	Numar mediu de salariat;i20				

Figura 2. Fișierul .csv

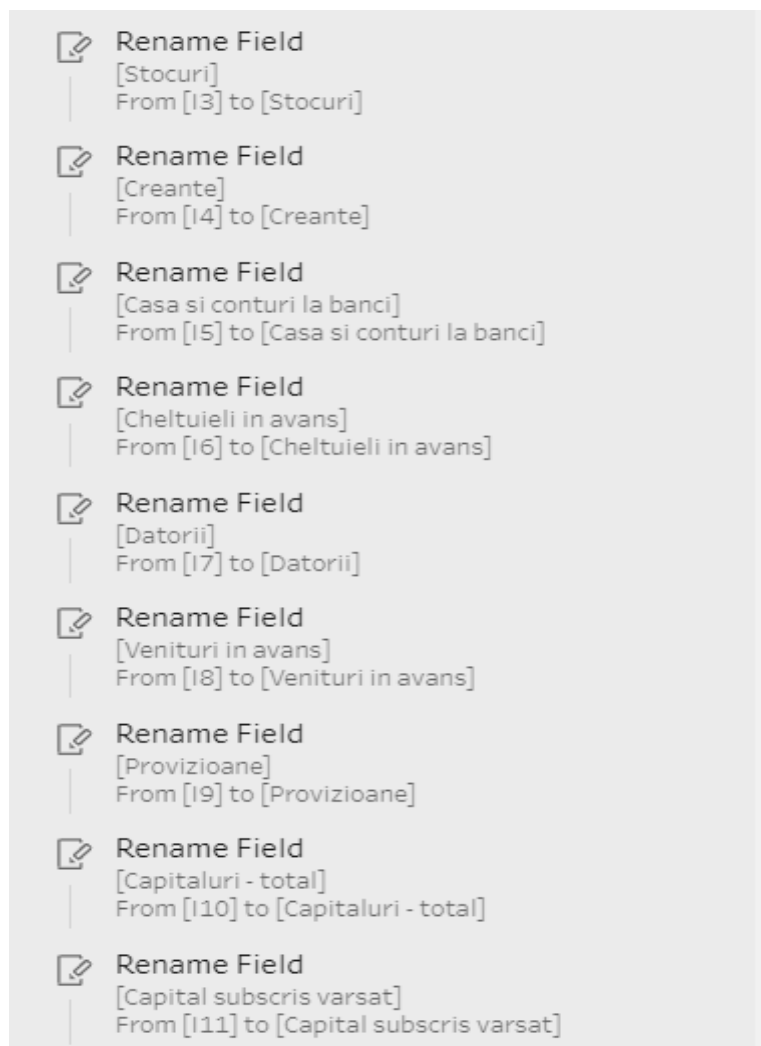


Figura 3. Modificări

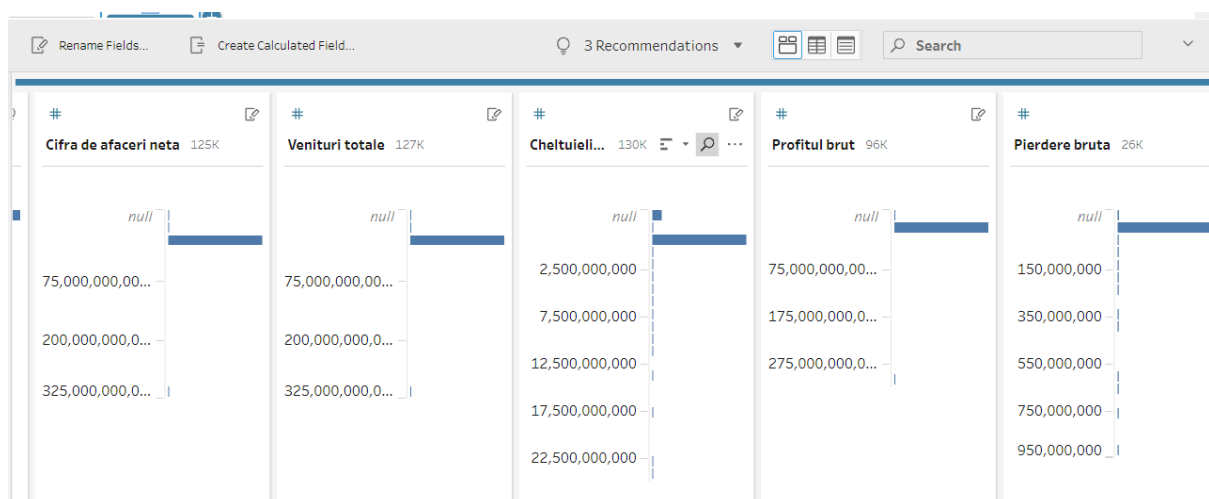


Figura 4. Rezultatul modificărilor

Am procedat la fel cu seturile de date cu situațiile financiare pentru societățile comerciale care depun bilanț format prescurtat și cu raportările anuale ale societăților comerciale care depun bilanț IFRS, rezultând alte două union-uri.

În cazul acestor union-uri am aplicat un filtru pentru firmele care au ca activitate grupul CAEN 602, obținând următoarele rezultate:

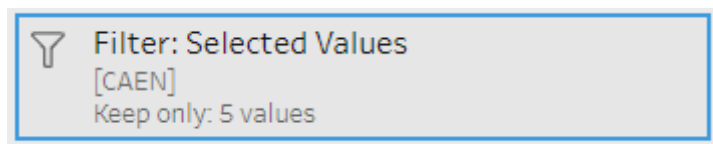


Figura 5. Filtru CAEN

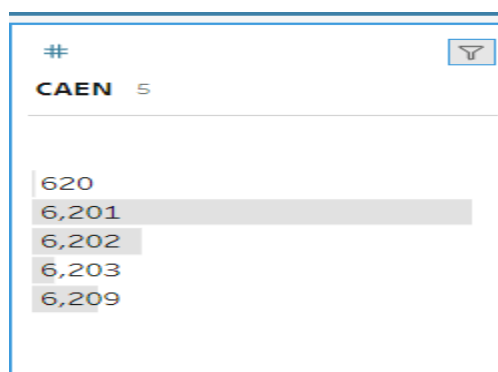


Figura 6. Rezultat firme cu bilanț lung, bilanț scurt și raportări anuale

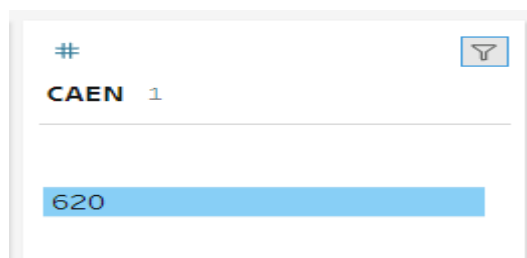


Figura 7. Rezultat firme bilanț format prescurtat și bilanț IFRS

Pentru fiecare union am redenumit table names cu „An”.

În scopul obținerii firmelor din România care au domeniul de activitate inclus în codul CAEN 602, am realizat un union alcătuit din union-urile precedente în care firmele sunt clasificate în funcție de tipul de bilanț, pe care l-am redenumit „Firme IT Ro”.

În cazul câmpului „Firme IT Ro”, am redenumit table name cu „Uniune bilanturi”.

Pentru a prelucra curat datele, am inclus un cleaning step la „Firme IT Ro”. Am șters câmpurile „Provizioane”, „Venituri în avans”, „Cheltuieli în avans” și „Patrimoniul regiei”, deoarece conținea valori nule.

După ce am adăugat conexiunea „Date\_identificare\_2020\_cluj”, am creat un inner join pe baza câmpurilor cod\_fiscal (din „Date\_identificare\_2020\_cluj”) și CUI (din „Firme IT Ro”) pentru a obține firmele din Cluj-Napoca cu grupa de activitate 602. Ulterior, am adăugat un cleaning step pentru a șterge câmpurile considerate nefolositoare (ex. „etaj”).

Am adăugat un câmp calculat unde am introdus câmpurile referitoare la adresă („str” și „nr”) pe care l-am redenumit sugestiv „Adresa”.

Am observat că la câmpul „localitate” apare o redundanță, adică existența valorilor „Cluj-Napoca” și „CLUJ-NAPOCA”. Am grupat aceste valori sub numele de „Cluj-Napoca”.

După toate aceste prelucrări am obținut un output numit „Firme\_IT\_Cluj”(format .csv).

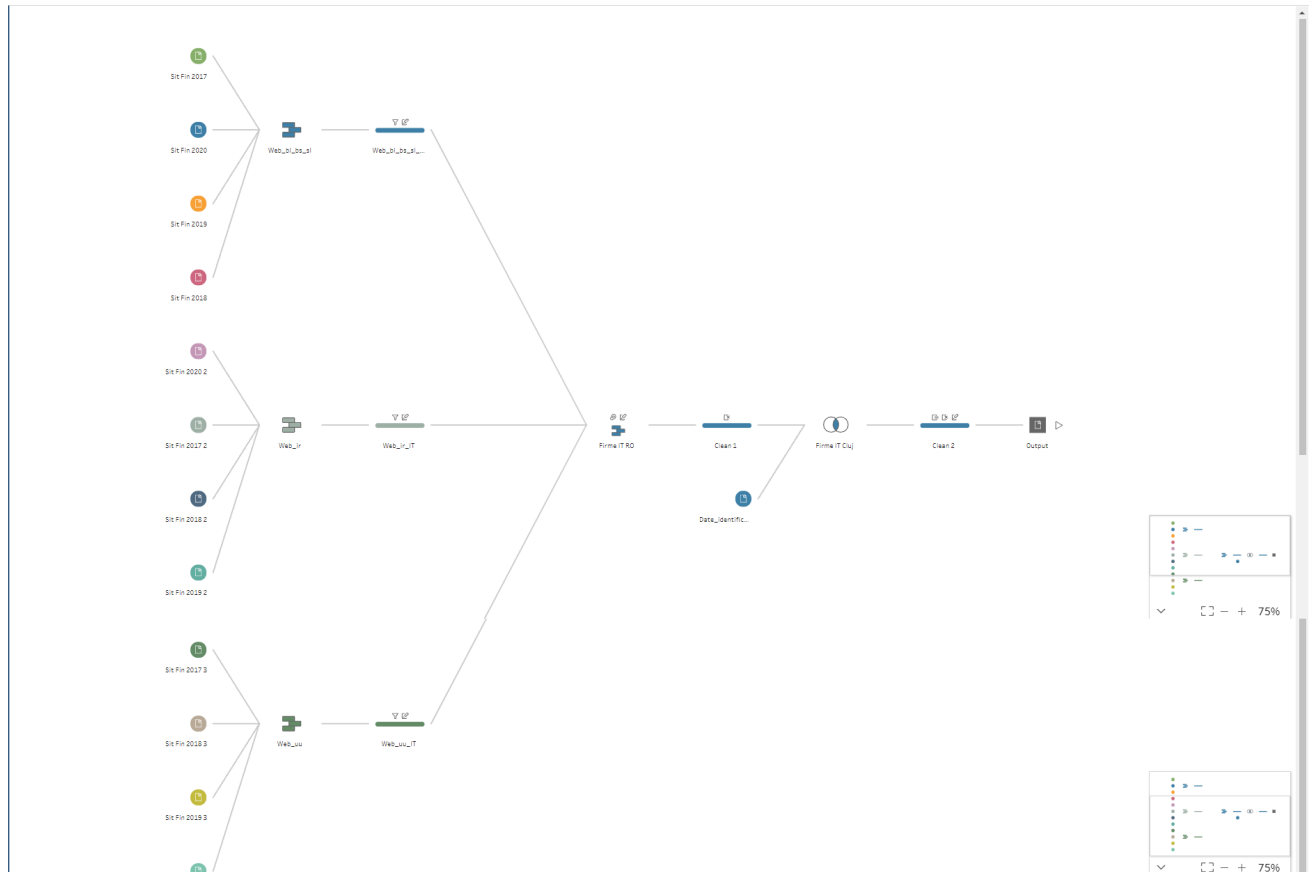


Figura 8. Reprezentarea flow-ului final

### III. Data Visualization & Analysis

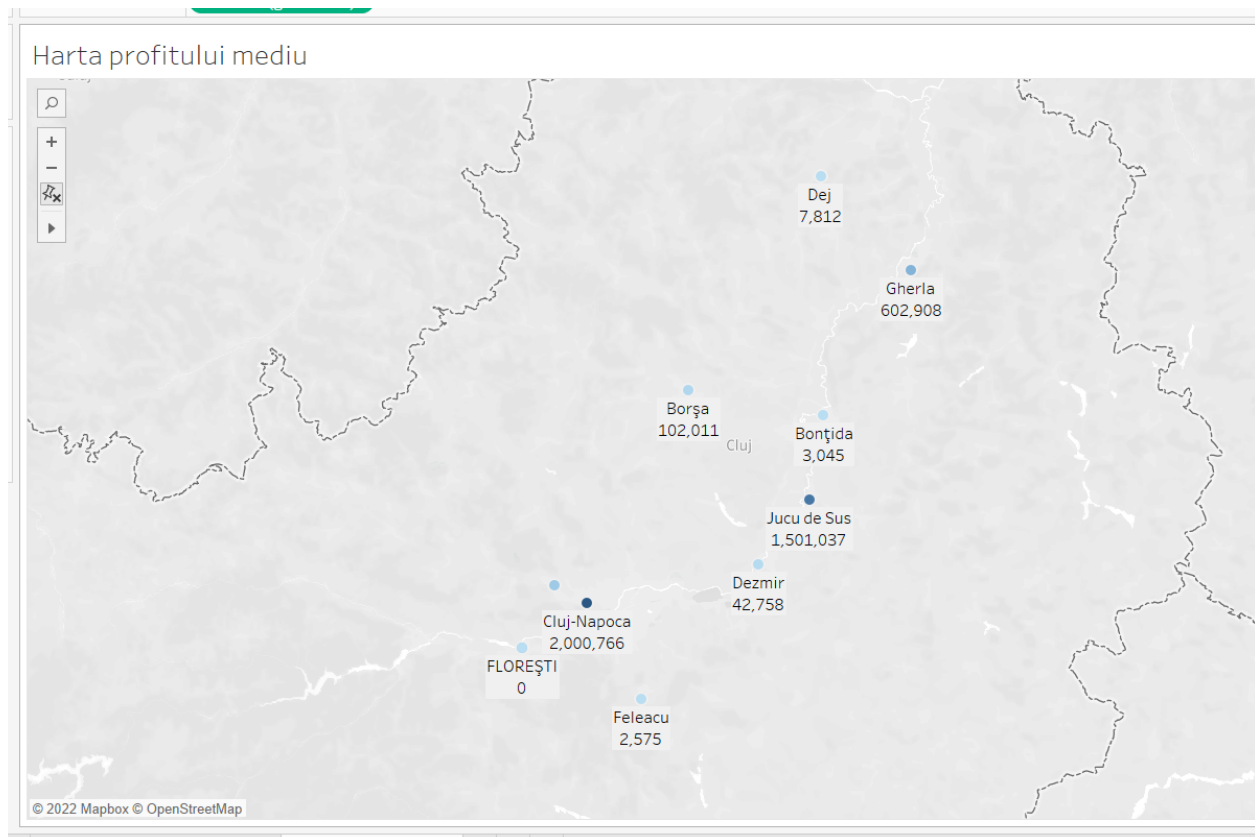


Figura 9. Harta profitului net mediu pe localități

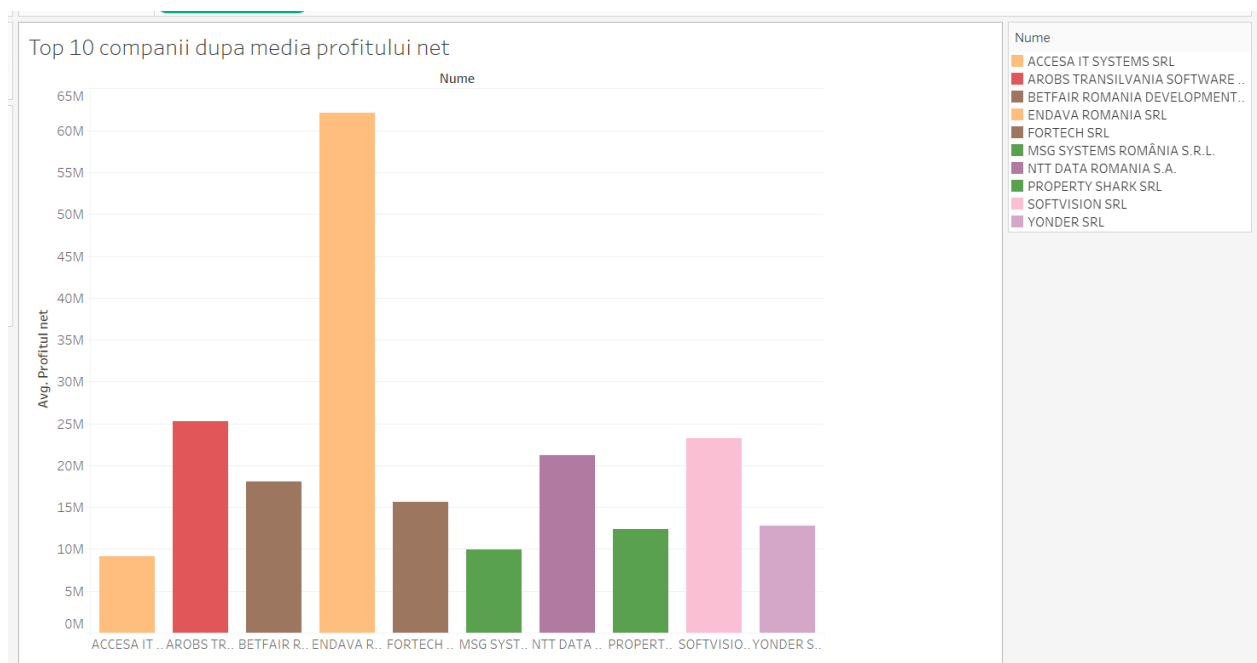


Figura 10. Top 10 companii în funcție de media profitului net

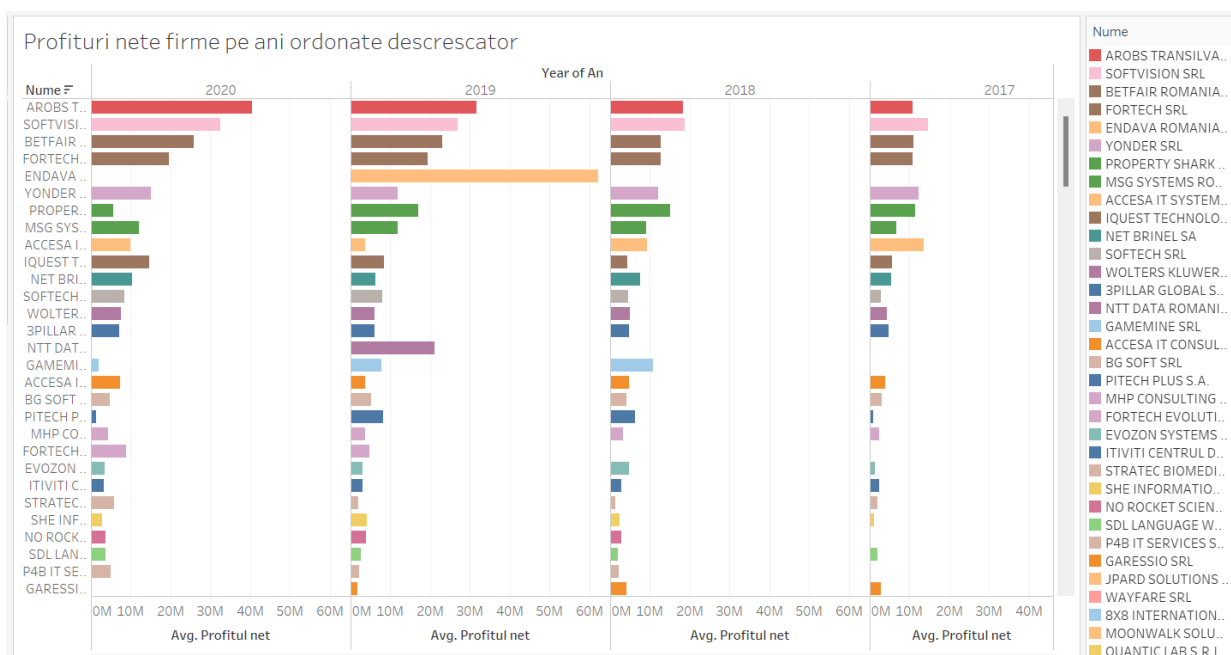


Figura 11. Profiturile nete ordonate descrescator în funcție de an

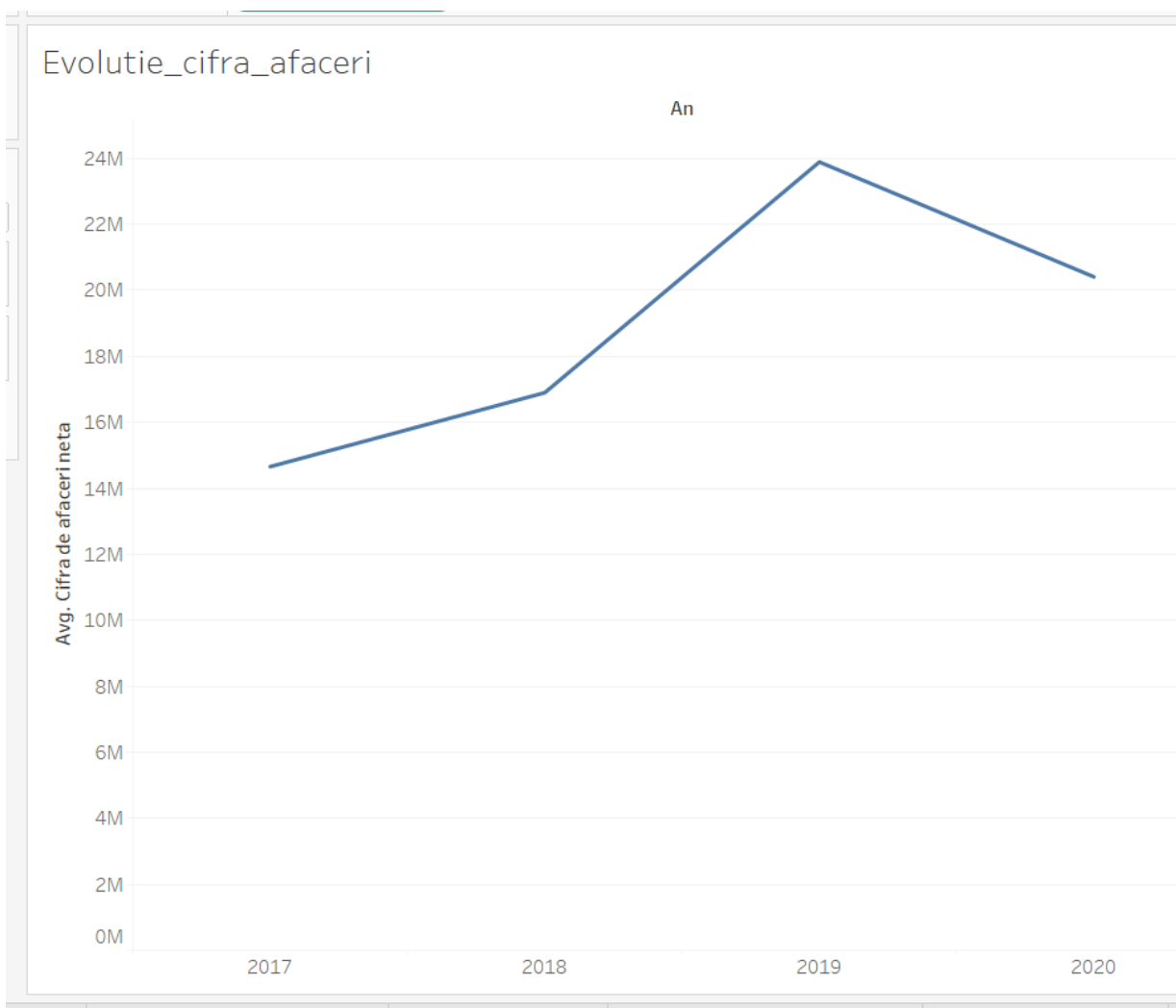


Figura 12. Evoluția mediei cifrei de afaceri netă



Active totale

×

SUM([Active circulante]+[Active imobilizate])

The calculation is valid.

Apply

OK

Figura 13. Câmp calculat „Active totale”

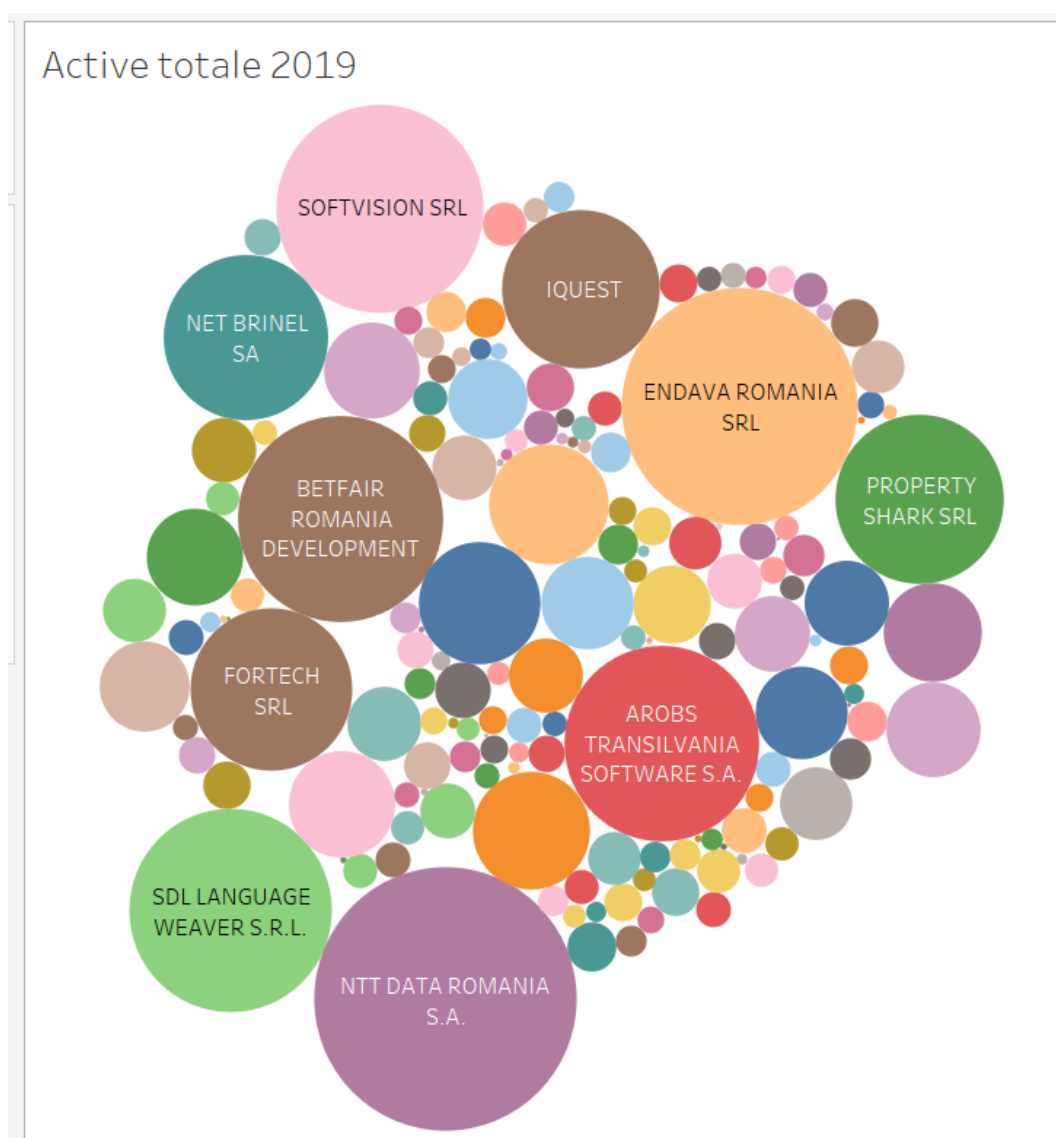


Figura 14. Activele totale pe firme 2019

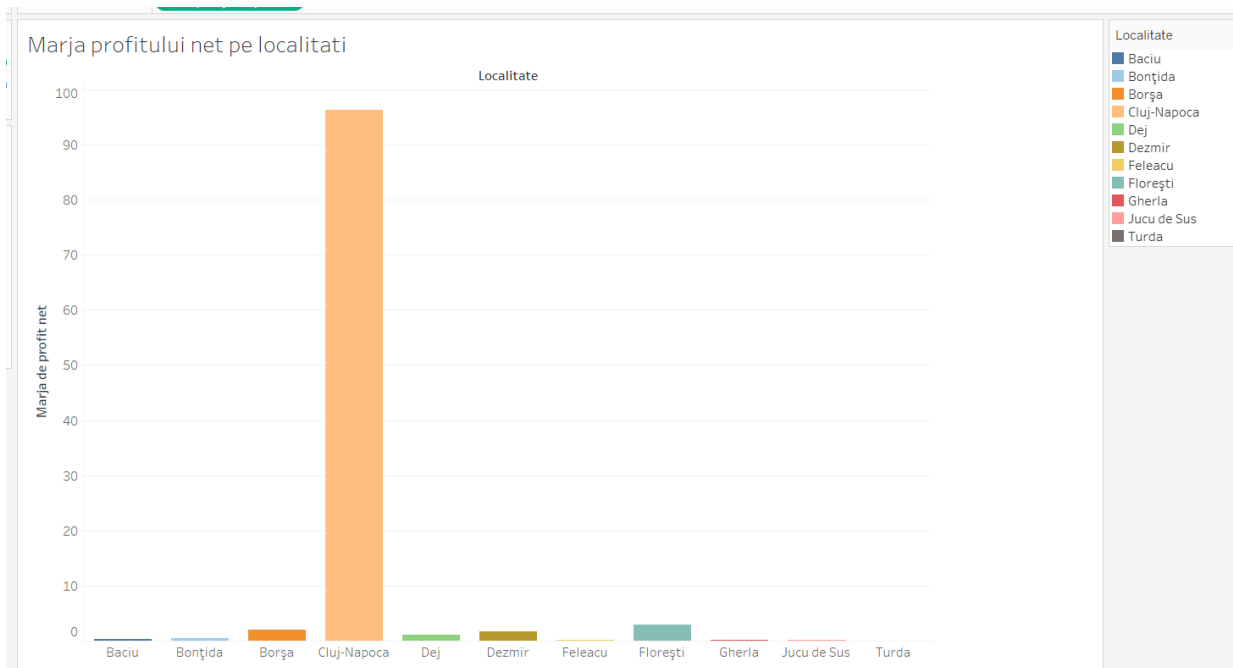


Figura 15. Marja de profit net pe localități



Figura 16. Top 10 firme în funcție de datorii

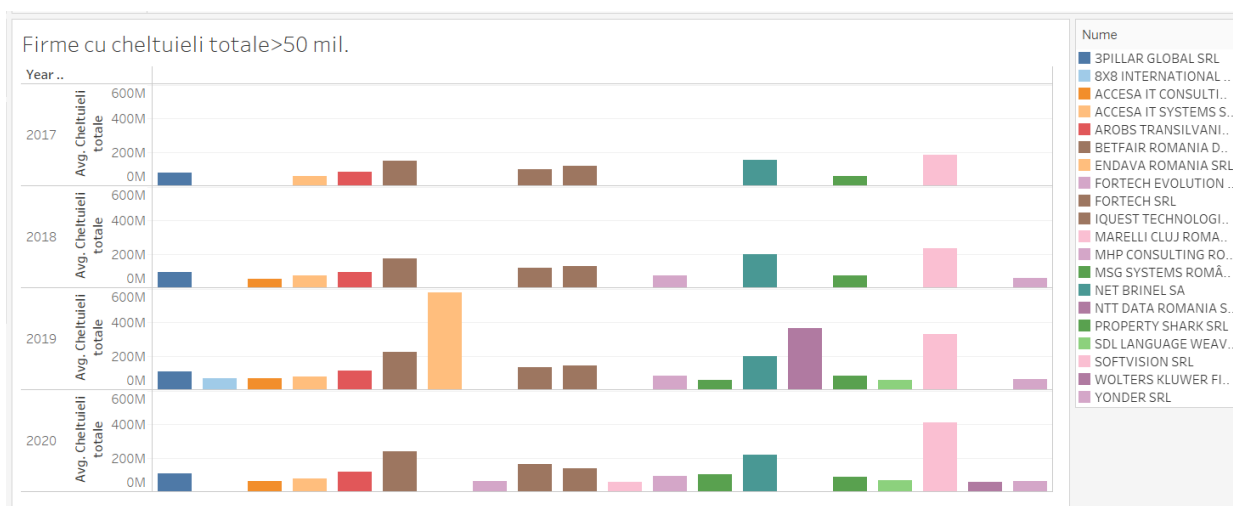


Figura 17. Graficul firmelor care au cheltuielile totale mai mari de 50 mil lei

×

[Cifra de afaceri neta]/[Numar mediu de salariati]

The calculation is valid.

Figura 18. Câmp calculat „Cifra\_afaceri/angajati”

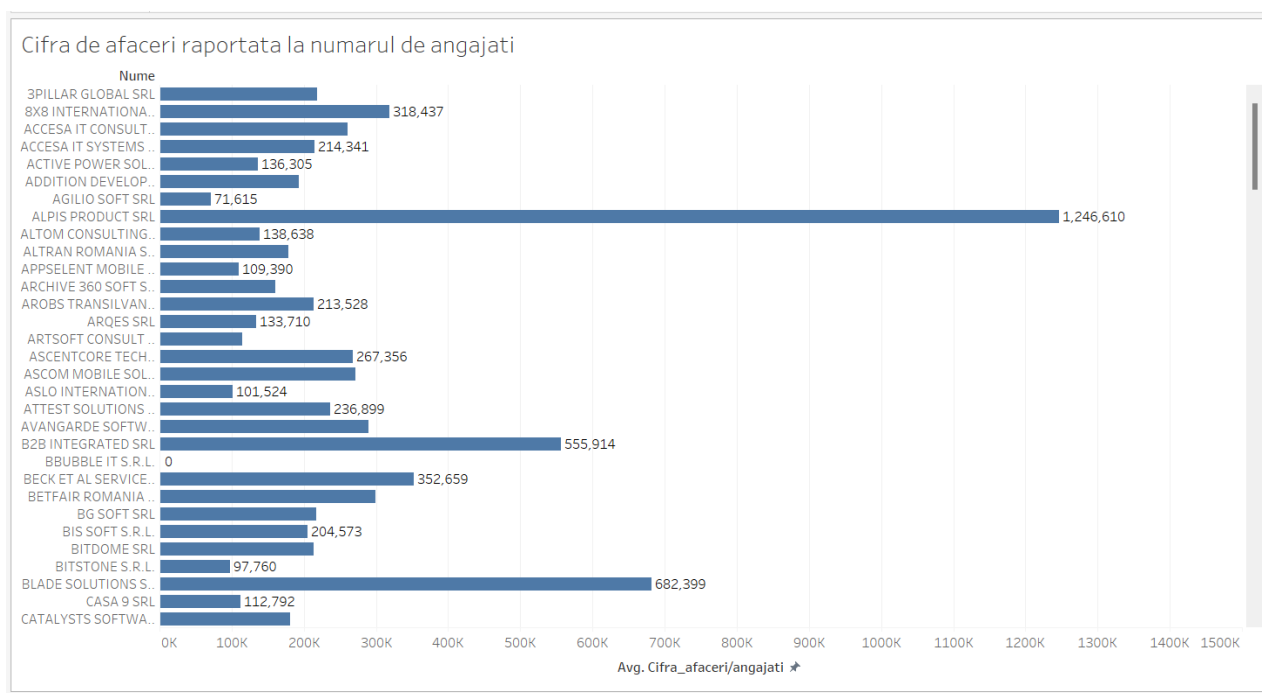


Figura 19. Graficul cifrei de afaceri raportată la numărul de angajați

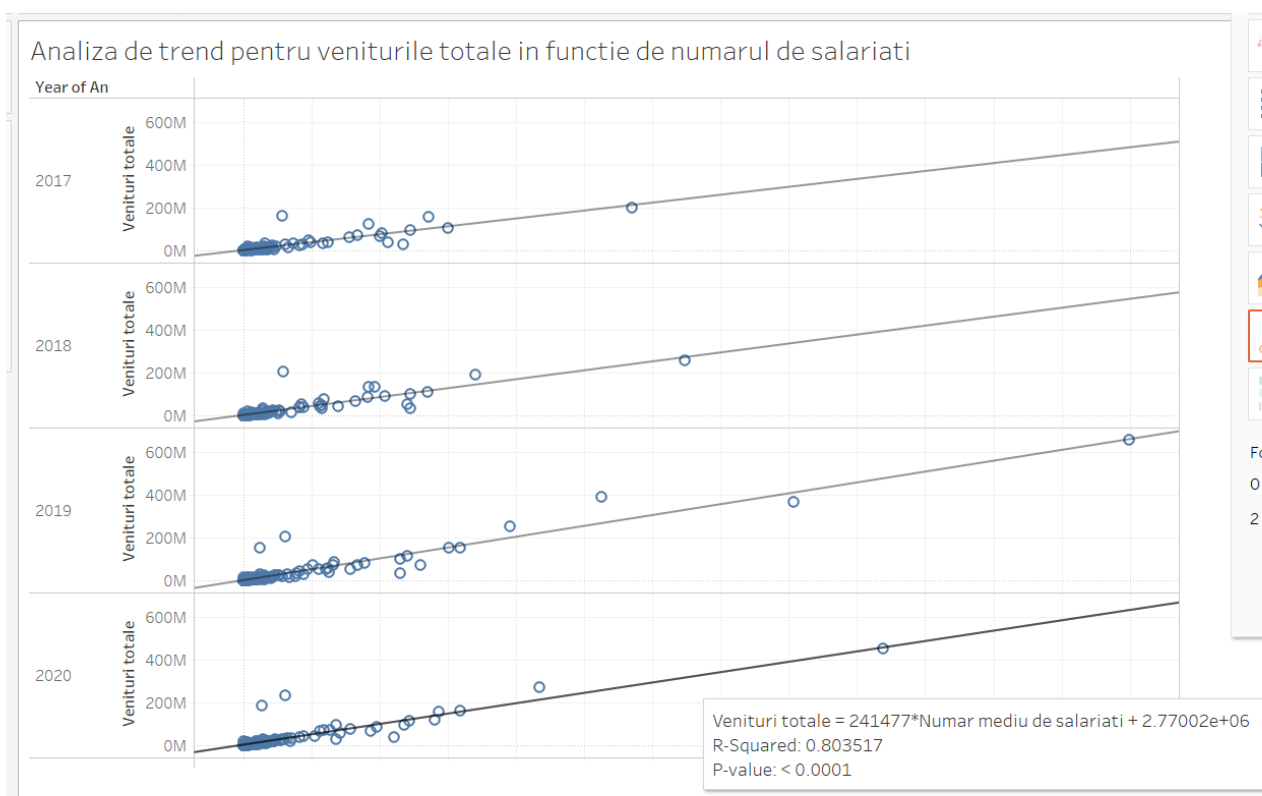


Figura 20. Analiza de trend pentru evoluția în timp a veniturilor totale în funcție de numărul de salariați

Cu ajutorul valorii P decidem dacă trendul se aplică și la scară largă, nu numai pe eșantionul de mai sus. Deoarece valoarea P este mai mică de 0.0001 în cazul tuturor regresiilor, concluzia este că graficul reprezentat este relevant.

Valoarea R-Pătrat sugerează dacă modelul redat se potrivește cu datele pe care dorim să le analizăm. Aceste valori pot varia între 0 și 1, iar având în vedere că valorile noastre de

R-Pătrat se află în intervalul [0.744556 - 0.900092], putem spune că avem o corelație puternică spre foarte puternică.

## IV. Creare Dashboard cu date legate

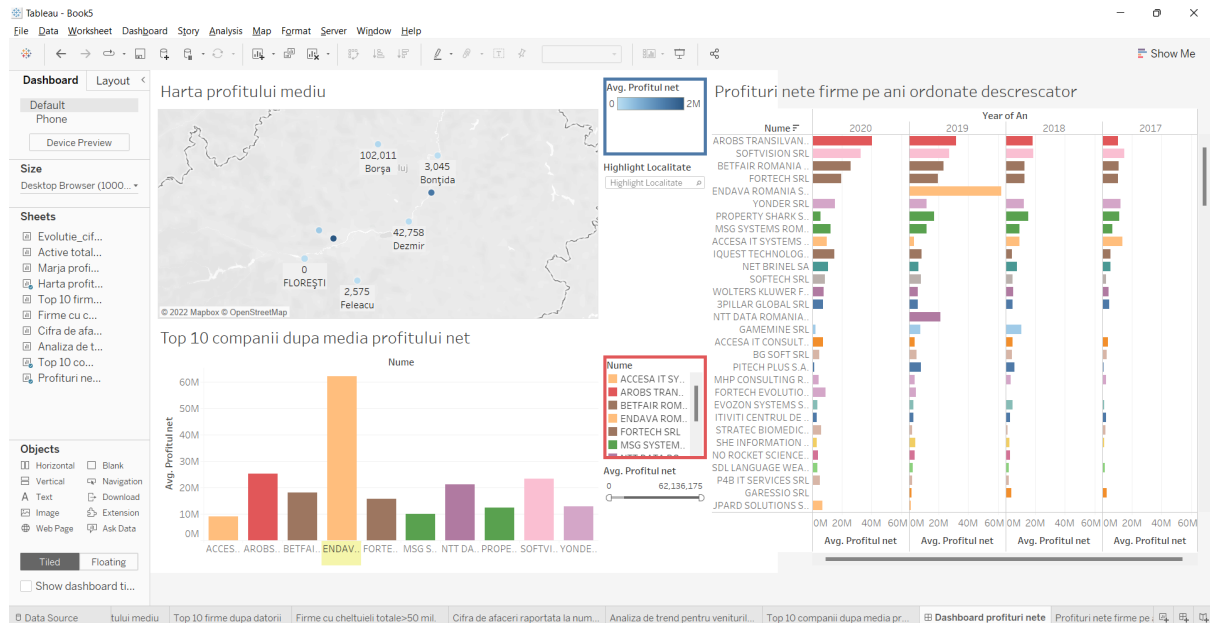


Figura 21. Dashboard profituri nete

## V. Crearea unei povești (Story)

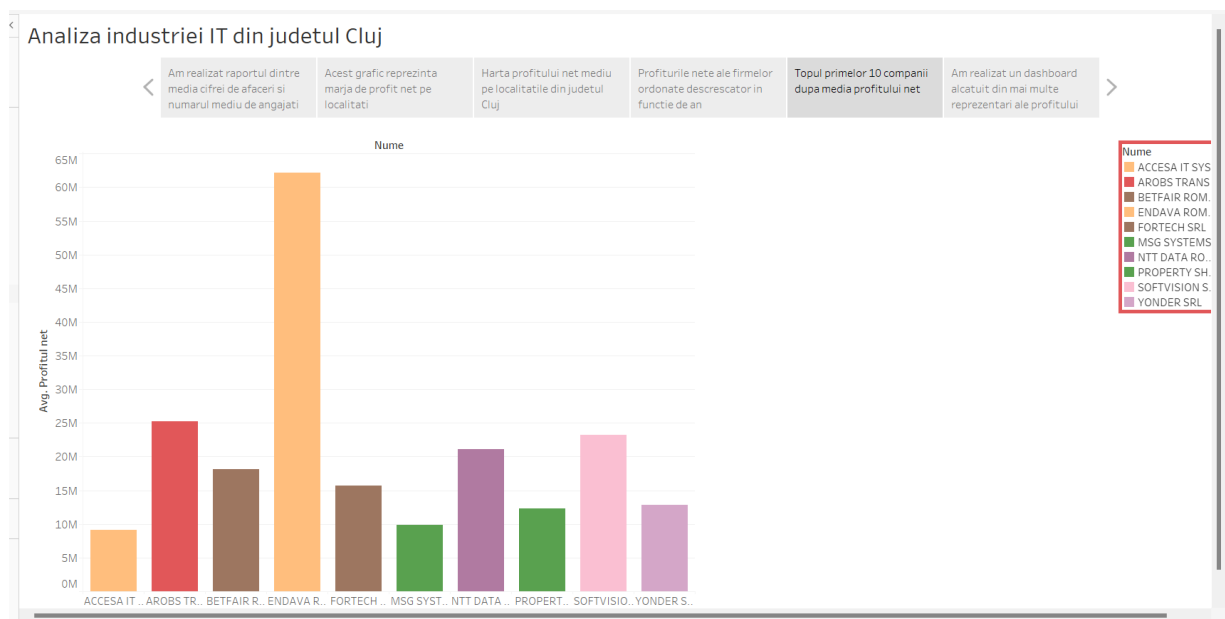


Figura 22. Story pentru analiza industriei IT din județul Cluj