# **PROIECT BUSINESS**

# **INTELLICENCE**

### Autor,

## darius-f96

## **Cuprins**

Descriere scenariu	2
Pregatirea datelor	3
Colectarea seturilor de date	3
Conectarea la seturile de date	4
Curatare, transformare si calculare a datelor	5
Analiza datelor	9
Crearea graficelor	11
Dashboard	15
Story	15

### Descriere scenariu

Alba este un judet situat in partea centrala a Romaniei, ocupa 2.6% din suprafata Romaniei si are aproximativ 416.000 de locuitori. Se invecineaza cu jud Cluj, Bihor, Arad, Valcea, Sibiu si Mures. Resedinta de judet este la Alba-Iulia,

Analiza de mai jos a fost realizata folosind datele din perioada 2016-2020. Scopul acesteia este sa aflam mai multe despre companiile cele mai prospere din judetul Alba, dar si domeniile in care acestea activeaza. In urma acestei analiza va fi mai clar ce domenii sunt profitabile si in care regiuni din judetul Alba s-au concentrat.

#### Analiza contine:

- Top 10 domenii (profit minim vs profit maxim pe perioada 2016-2020)
- Top 10 companii (profit 2016 vs profit 2020)
- Evolutia companiilor din top 10
- Localizarea companiilor din top 10
- Localizare generala a companiilor din jud Alba
- Localizare a top 10 domenii profitabile

### Pregatirea datelor

#### Colectarea seturilor de date

Fiind de origine din judetul Alba, am ales sa analizez datele companiilor din acest judet. Datele au fost luate de pe <a href="https://data.gov.ro/dataset">https://data.gov.ro/dataset</a> :

- Situatii Financiare
  - o Link:

https://data.gov.ro/dataset?q=situatii+financiare&sort=score+desc %2C+metadata\_modified+desc

- o Extensie sursa de date: web\_uu\_an20XX.txt
- Observatie: s-au folosit fisierele anilor 2016-2020
- Lista firme judetul Alba:
  - o Link: https://data.gov.ro/dataset/informatii\_fiscale\_mai\_2020
  - o Extensie sursa de date: alba.csv
- Coduri CAEN:
  - o Link: https://github.com/StemateF/CAEN\_db
  - o Extensie sursa db: Caen\_DB.sql

Perioada de referinta este 2016-2020, au fost descarcate cele mai recente fisiere, anul 2020 fiind ultimul disponibil. In aceste fisiere coloanele au fost denumite generic (i1, i2, i3, etc). Am descarcat si un fisier cu extensia web\_uu\_an2020.csv, care contine legenda coloanelor, folosita ulterior la pasul de curatare pentru a redenumi coloanele.

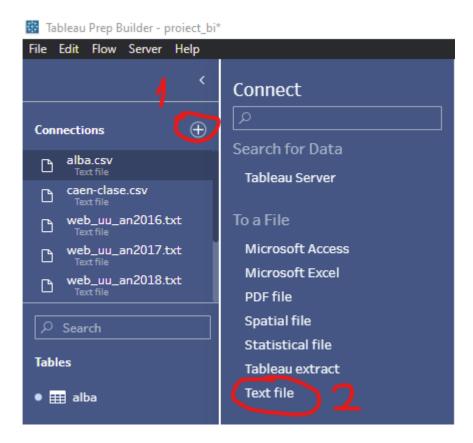
Pentru fisierul cu codurile caen am importat caen\_db.sql descarcat anterior intr-o baza de date, pe urma am facut export (in format .csv) datelor din tabela caen\_clase.

Alternativ se putea face o conexiune direct in tableau catre baza de date mysql, dupa descarcarea pluginului ODBC, insa am considerat ca este mai eficient sa fac un export si sa opresc baza de date ulterior, fiind nevoie doar de acel tabel.

#### Conectarea la seturile de date

Pentru a conecta seturile de date am folosit software-ul Tableau Prep Builder si am incarcat diferitele surse de date mentionate mai sus:

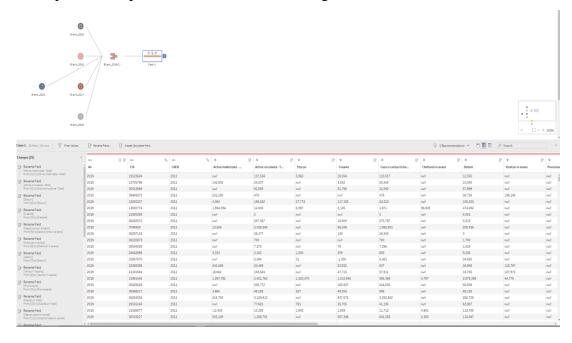
- 1. Situatiile financiare
- 2. Lista companiilor din judetul Alba
- 3. Codurile CAEN



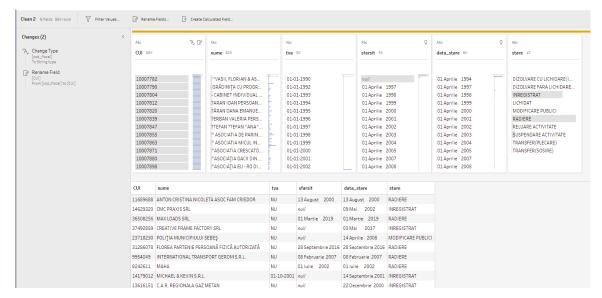
### Curatare, transformare si calculare a datelor

Dupa ce am incarcat toate datele mentionate la punctul anterior, am trecut la urmatoarul pas in procesul de analiza datelor: curatare, transformare si calculare a datelor:

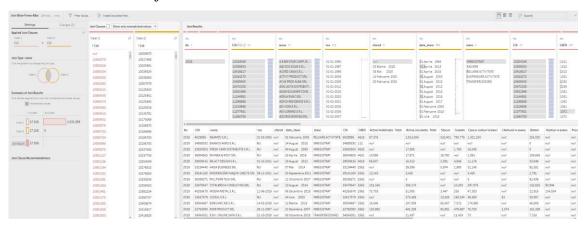
- 1. Pentru situatiile financiare am creat un Union, denumit Bilant\_2016-2020 in care am inclusiv toate cele 5 tabele aferente anilor 2016-2020. Ulterior am adaugat un pas de curatare in care am efectuat urmatoarele:
  - Redenumirea tuturor coloanelor format iX folosindu-ma de legenda coloanelor mentionata mai sus
  - Redenumirea coloanei "Tablename" in "An"
  - Stergerea literelor si a semnelor de punctuatie din coloana "An"
  - Tipul de date pentru CUI number -> string
  - Tipul de date pentru CAEN number -> string



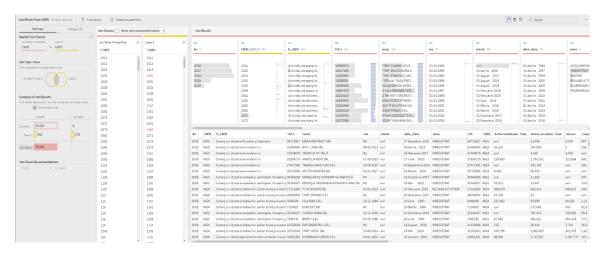
2. Companii judet: Pentru prelucrarea datelor din fisierul Firme\_Alba am creat un pas de agregare pentru cod fiscal, nume, tva, data\_stare, stare, loc si sfarsit, ulterior intr-un pas de curatare am redenumit coloana "cod fiscal" in "CUI" si i-am schimbat tipul de date de la number -> string.



3. Ulterior am facut un join intre datele din Bilant\_2016-2020 si Firme\_Alba:

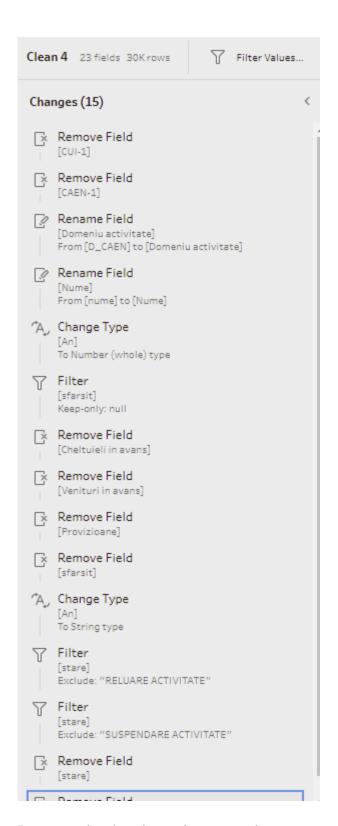


4. Pentru a avea informatiile despre domeniul de activitate este nevoie de un JOIN intre datele din Join-Bilant-Firme-Alba si datele din clase\_caen, pe cod CAEN, dar nu inainte de a schimba tipul de date pentru coloana CAEN din number in string si a selecta doar coloanele CAEN si D\_CAEN printr-o functie de agregare:

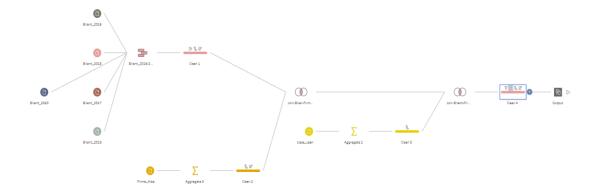


5. Inainte de output, am mai adaugat un pas de cleaning, unde am sters coloanele duplicate, redenumit unele coloane, schimbat tipul de date pentru An din string -> number, sters firmele care si-au incetat activitatea in aceasta perioada.





In urma celor descrise mai sus a rezultat urmatorul flux:



In urma executarii acestui flow a rezultat un fisier cu 25 de coloane si 35820 de randuri.

### Analiza datelor

Pentru a analiza datele companiilor din judetul Alba s-au realizat urmatoarele:

- Campuri calculate pentru profitul net (maxim si minim) pe toata perioada 2016-2020:
  - MaxProfitNet / MinProfitNet

```
ProfitNet-Maxim

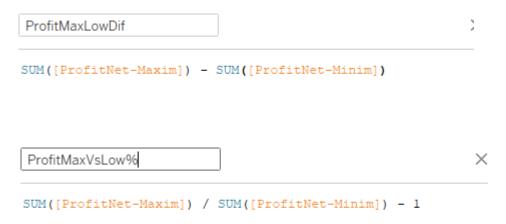
IF {
    FIXED [Domeniu activitate]:MAX([An])} = [An]
THEN [Profit net] END

ProfitNet-Minim

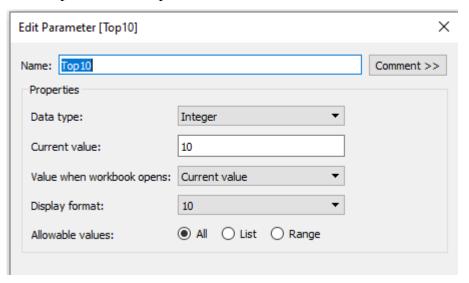
X

IF {
    FIXED [Domeniu activitate]:MIN([An])} = [An]
THEN [Profit net] END
```

o Profit minim vs maxim, atat valoric cat si procentual



- Parametrii calculati care ne vor ajuta in realizarea unui top 10 cele mai profitabile companii/domenii:
  - o Crearea parametrului top 10:



o Crearea campului calculat Top 10 domenii profitabile:

```
Top-10-Domenii X

Results are computed along Table (across).

IF RANK_UNIQUE([ProfitMaxLowDif]) <= [Top10]

THEN ATTR([Domeniu activitate]) ELSE NULL END
```

o Crearea campului calculat Top 10 companii profitabile

```
Top-10-Companii X

Results are computed along Table (across).

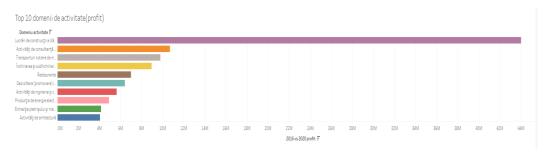
IF RANK_UNIQUE([Profit growth 2016-2020])<=[Top10]

THEN ATTR([Nume]) ELSE NULL END
```

## Crearea graficelor

Pentru reprezentarea rezultatelor prin intermediul graficelor am ales sa prezentam:

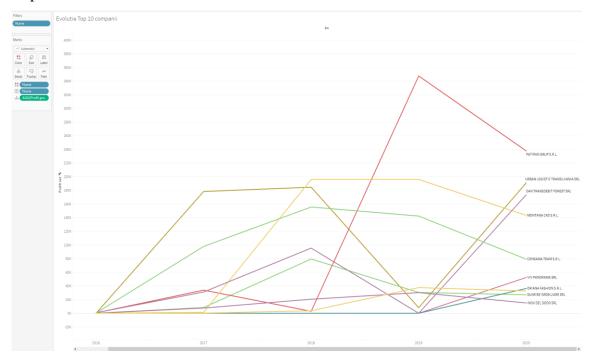
• Top 10 domenii min vs max profit:



• Top 10 companii profit 2016 vs profit 2020:



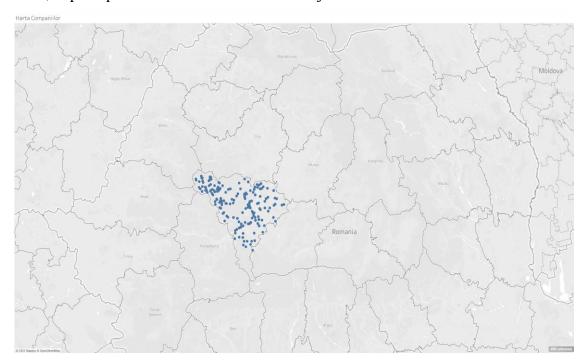
 Reprezentarea grafica in timp a rezultatelor inregistrate de companiile respective intre 2016 si 2020:



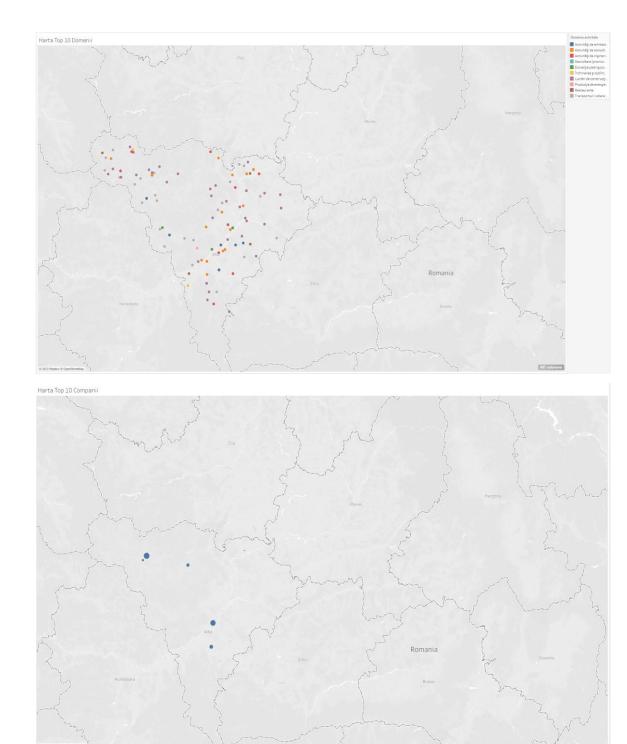
Se poate observa faptul ca Patipan Grup SRL are profituri mult mai mari decat Urban Logistic Transilvania, insa motivul pentru care anterior nu a fost pe locul 1 este ca nu a inregistrat o crestere procentuala la fel de mare comparativ cu Urban Logistic Grup Transilvania SRL.

#### • Reprezentarea geografica:

Pentru a genera o harta corecta a fost nevoie de un import custom geocoding, dat fiind faptul ca Tableau nu cunoaste majoritatea localitatilor din judetul Alba, dupa import harta cu locatia firmelor din judetul Alba arata astfel:



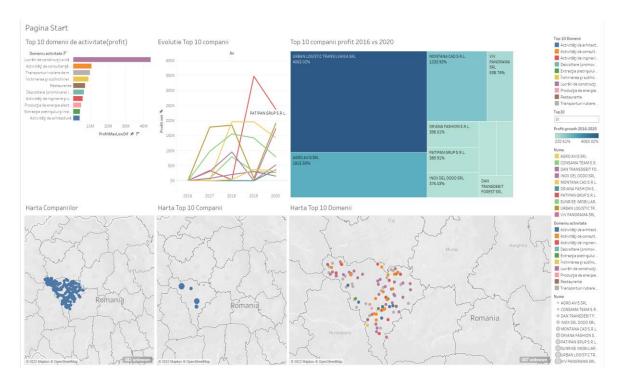
Pentru a obtine Geocodes am facut un script in Python, care preia datele din fisierul alba.csv si cauta coordonatele geografice ale localitatilor pe <a href="https://locationiq.com/">https://locationiq.com/</a>



O buna parte din companii se afla in Alba Iulia si Bistra

### **Dashboard**

Am creat un dashboard intitulat "Pagina Start", care ofera un rezumat a tot ce s-a realizat in cadrul acestui studiu:



## Story

In ultima etapa a proiectului am creat si un story care sa contina cele mentionate mai sus:

#### Analiza Judet Alba



Alba

Localizare Top 10 Domenii

Top 10 companii profitabile din judetul Alba

Evolutie Top 10 companii 2016-2020 Localizare Top 10 Companii

