

# **Proiect Business Intelligence**

**Analiza companiilor ce au codul CAEN 5610  
din municipiul Satu Mare  
realizată pe anii 2017-2020**

Sabou Sergiu

## Contents

1. Descriere problemă de business .....	3
2. Pregătirea datelor .....	4
2.1. Incărcarea datelor și conectarea la seturile de date.....	4
2.2. Descrierea pașilor de curățare și a modificărilor făcute. ....	5
2.2. Descrierea pașilor de curățare și a modificărilor făcute. ....	7
3. Data Visualization & Analysis.....	9
3.1 Câmpuri calculate.....	9
3.2 Grafice și diagrame .....	11
3.2.1 Profit pe ultimii 2 ani.....	11
3.2.2 Clasificare companii în funcție de activele totale deținute ordonate de la mic la mare .....	12
3.2.3 Comparatie între rata profitului pe primii 2 ani analizați .....	12
3.2.4 Cifrele de afaceri pe ani .....	13
3.2.5 Capital total pozitiv și negativ .....	13
3.2.6 Evoluția numărului de angajați pe ani utilizând range.....	14
3.2.7 Tendința profitului net în funcție de timp.....	14
3.2.8 Distribuția cifrei de afaceri în funcție de codul poștal .....	15
4. Creare Dashboard cu date legate .....	15
5. Crearea unei povești (Story) .....	16

## 1. Descriere problemă de business

Având în vedere situația pandemică din ultimii ani, am optat pentru analiza situației financiare a companiilor considerate cele mai afectate din această perioadă, restaurantele. Am preluat datele necesare de pe [data.gov.ro/dataset](https://data.gov.ro/dataset).

Am analizat următoarele aspecte:

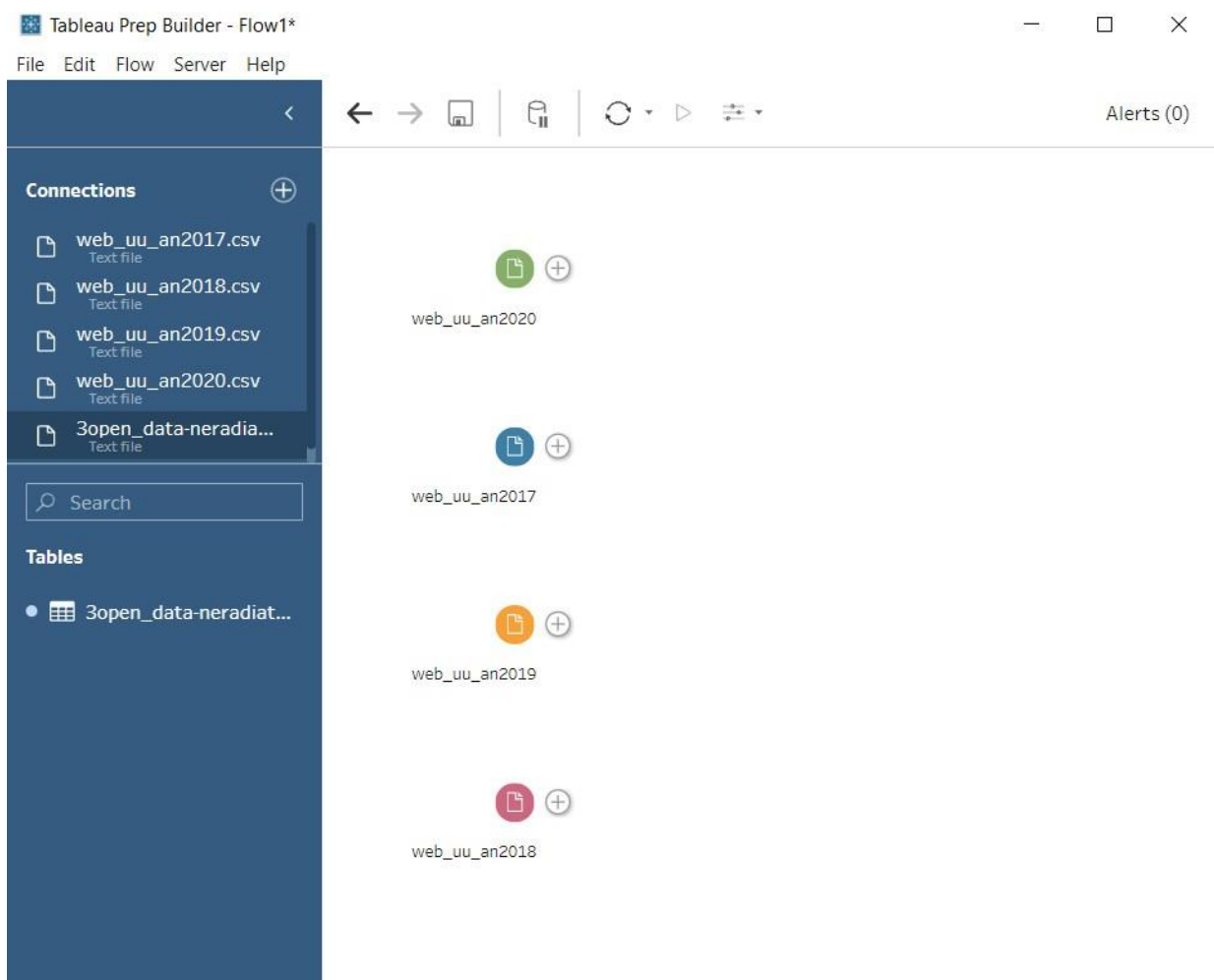
- profitul în 2019 și în anul pandemic (2020)
- clasificarea companiilor în funcție de activele totale în an pandemic
- rata profitului în 2 ani prepandemici
- suma cifrelor de afaceri pe ani
- companiile care au înregistrat capital total pozitiv/neativ
- evoluția în timp a numărului de angajați
- tendința cheltuielilor pe ani având în vedere media și constanta

## 2. Pregătirea datelor

### 2.1. Incărcarea datelor și conectarea la seturile de date

Pentru realizarea flow-ului în Tableau Prep am avut nevoie de fișierele web\_uu\_an<<an>>.csv care conțin câmpurile: CUI, CAEN, Active imobilizate, Active circulante, Stocuri, Creanțe, Casa și conturi la bănci, Cheltuieli în avans, Datorii, Venituri în avans, Provizioane, Capital, Capital subscris vărsat, Patrimoniul regiei, Cifra de afaceri netă, Venituri totale, Cheltuieli totale, Profit brut, Pierdere brută, Profit net, Pierdere netă, Număr mediu de salariați și de și fișierul 3open\_data-neradiate\_cu\_sediu-08.02.2021, cu datele: Denumirea firmei, CUI, Cod înmatriculare, EUID, Stare firmă și Adresă.

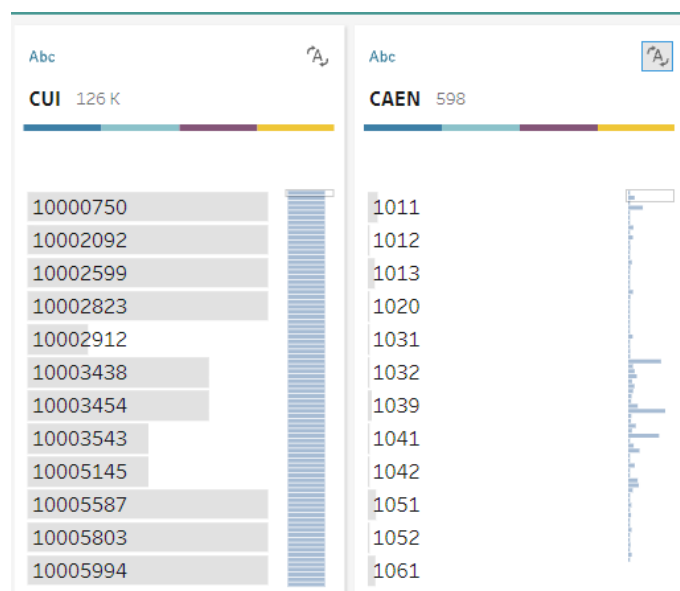
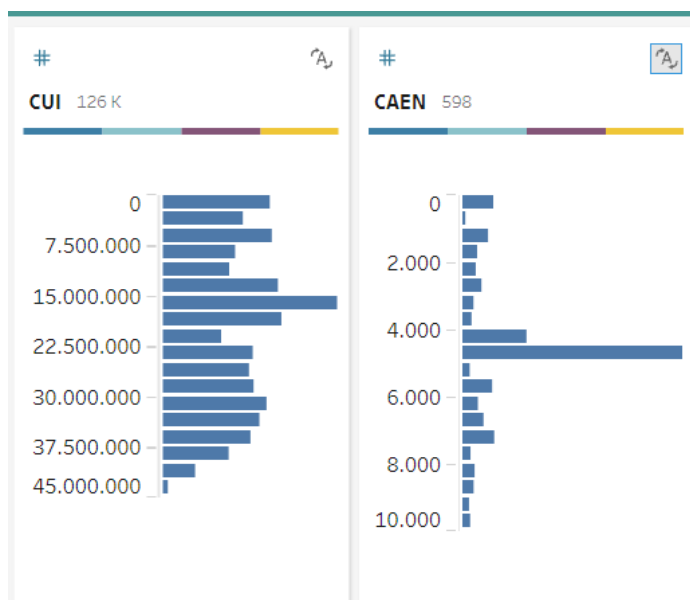
Am descărcat fișierele de pe siteul [data.gov.ro/dataset](https://data.gov.ro/dataset) și le-am adăugat la Connections.



## 2.2. Descrierea pașilor de curățare și a modificărilor făcute.

Înainte de a face Union pe fișiere am efectuat pași de clean pentru a eficientiza procesul fișierelor de foarte mari dimensiuni. În acești pași de clean:

- Am schimbat tipul datelor pentru coloanele CUI și CAEN din number whole în string pentru a ușura vizualizarea.
- Nu am exclus valorile null pentru a nu omite anumite companii (de exemplu sunt companii care nu au active imobilizate, însă au active circulante, iar noi vrem să le includem și pe acestea în analiza noastră).
- Am șters câmpurile Stocuri, Creanțe, respectiv Casa și conturi la alte bănci pentru că: Stocuri + Creanțe + Casa și conturi la bănci = Active circulante.
- Am filtrat după codul CAEN 5610 pentru a avea o evidență a companiilor care activează în domeniul restaurantelor.



## &gt;

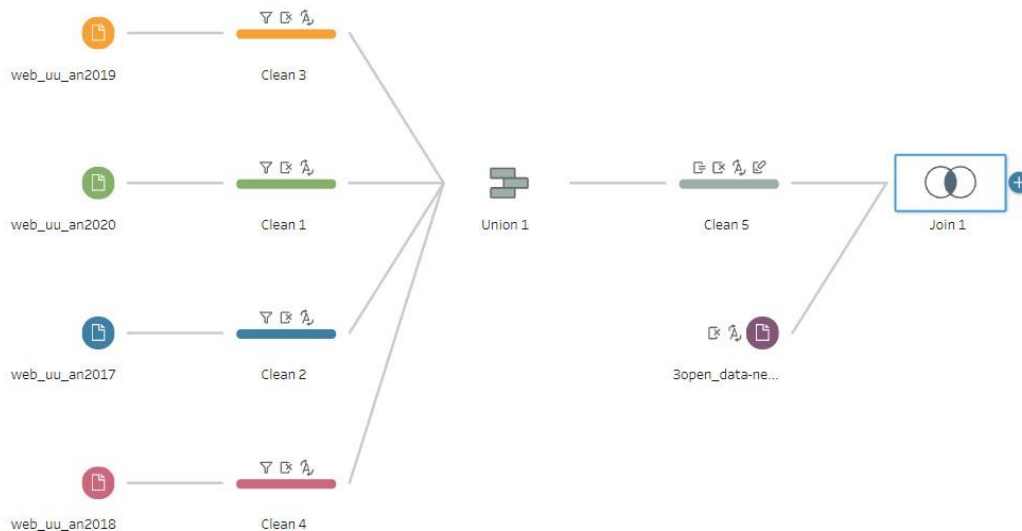
## 2.2. Descrierea pașilor de curățare și a modificărilor făcute.

Realizăm Union dintre cei patru pași de clean, în care reunim datele din cele 4 fișiere inițiale. După Union, numărul coloanelor și denumirea rămân aceleași.



Creăm un nou pas de clean pentru Union și efectuăm următoarele:

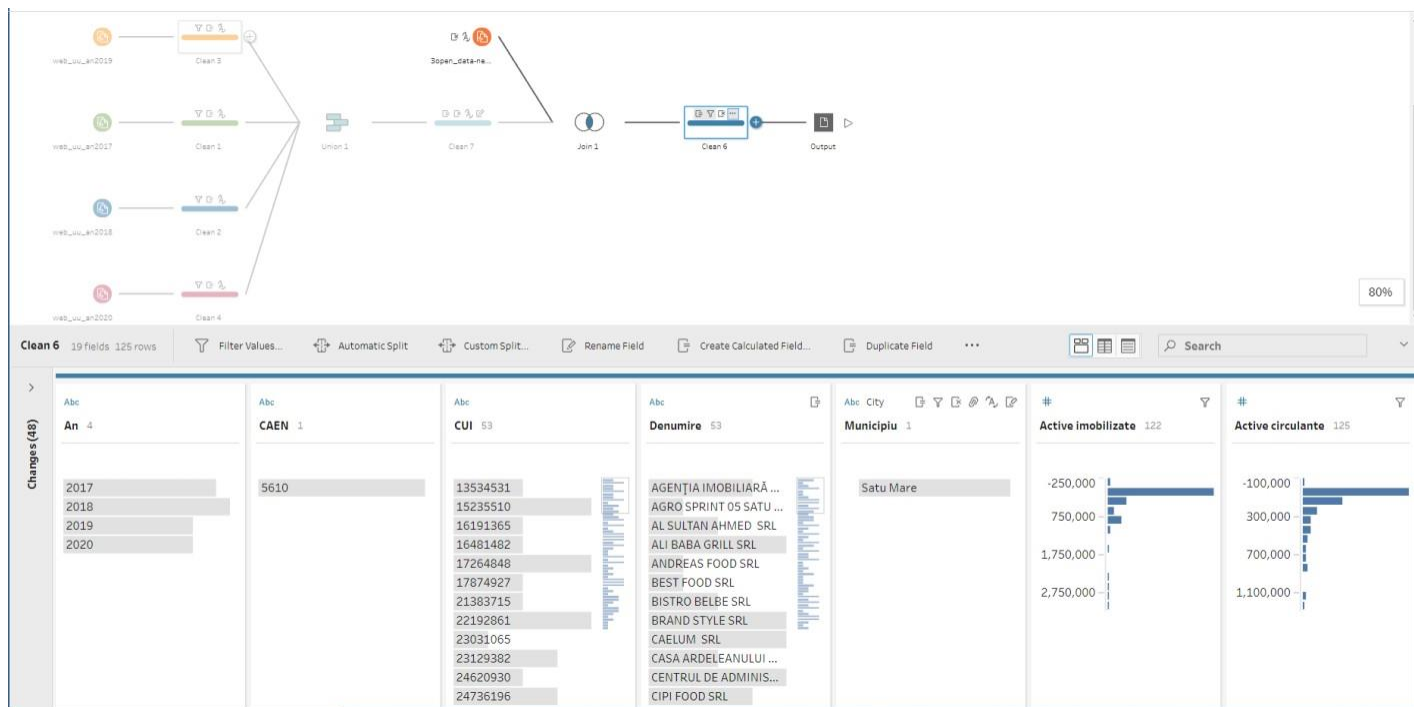
- Am redenumit câmpurile după capul de tabel disponibil pe site-ul [data.gov.ro/dataset](https://data.gov.ro/dataset).
- Am efectuat Automatic Split asupra câmpului Table Names pentru a extrage anul pentru care analizăm înregistrările. Coloana nouă am numit-o An și am schimbat tipul datelor (din number whole în string).
- Am șters coloanele cu multe valori nule: Cheltuieli în avans, Venituri în avans, Provizioane și Patrimoniul regiei.



Am efectuat Join între Clean 5 și data-neradiate\_cu\_sediu-08.02.2021, iar apoi am creat un clean step în care:

- am filtrat datele după Municipiul Satu Mare
- am șters CUI-1, pentru reducerea redundanței
- am eliminat toate valorile null
- am realizat Custom Split pe câmpul Adresa pentru a extrage separat municipiul Satu Mare, respectiv codul poștal și am șters coloana Adresă
- Am șters extra spaces din cod poștal
- Am înlocuit “S.R.L.” și “S R L” cu “SRL” în denumirea firmelor pentru a avea aceeași format
- am pus ca role City la coloana Municipiu, respectiv rol ZIP Code/Postcode la coloana Cod poștal





### 3. Data Visualization & Analysis

#### 3.1 Câmpuri calculate

Pentru analiză am calculat 3 indicatori financiari: Active totale, Rata profitului net și Rata de îndatorare globală.

Active totale

×

[Active circulante]+[Active imobilizate]

▶

The calculation is valid.

2 Dependencies ▼

Apply

OK

Rata profitului net

×

[Profit net]/[Cifra de afaceri netă]

▶

The calculation is valid.

Apply

OK

ata de îndatorare globală

[Datorii]/[Active totale]

The calculation is valid.

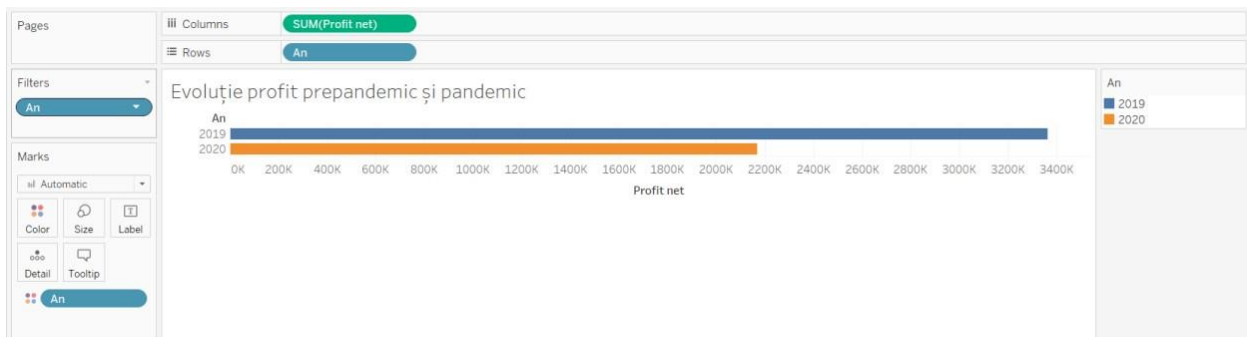
ApplyOK

## 3.2 Grafice și diagrame

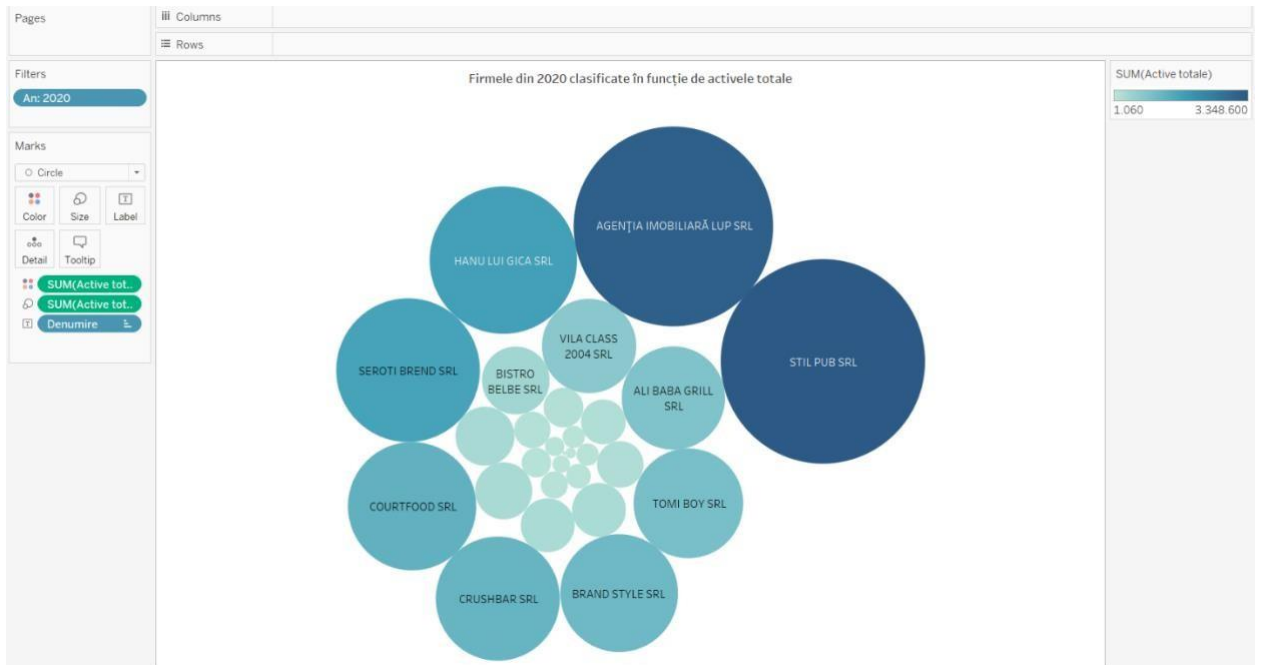
Am folosit marcatori (an - culoare, active totale, denumire etc.) și am **filtrat după diverse coloane**.

### 3.2.1 Profit pe ultimii 2 ani

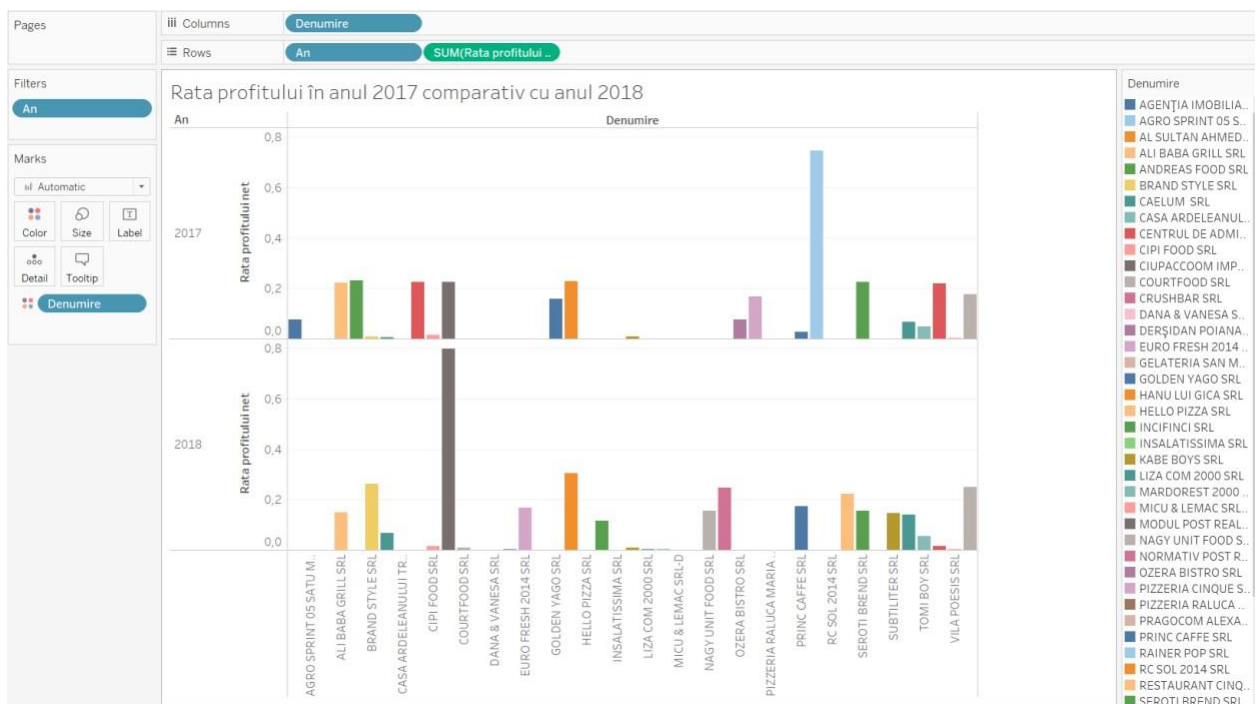
Din graficul de mai jos reiese faptul că în anul pandemic 2020 profitul companiilor per total a scăzut.



### 3.2.2 Clasificare companii în funcție de activele totale deținute ordonate de la mic la mare



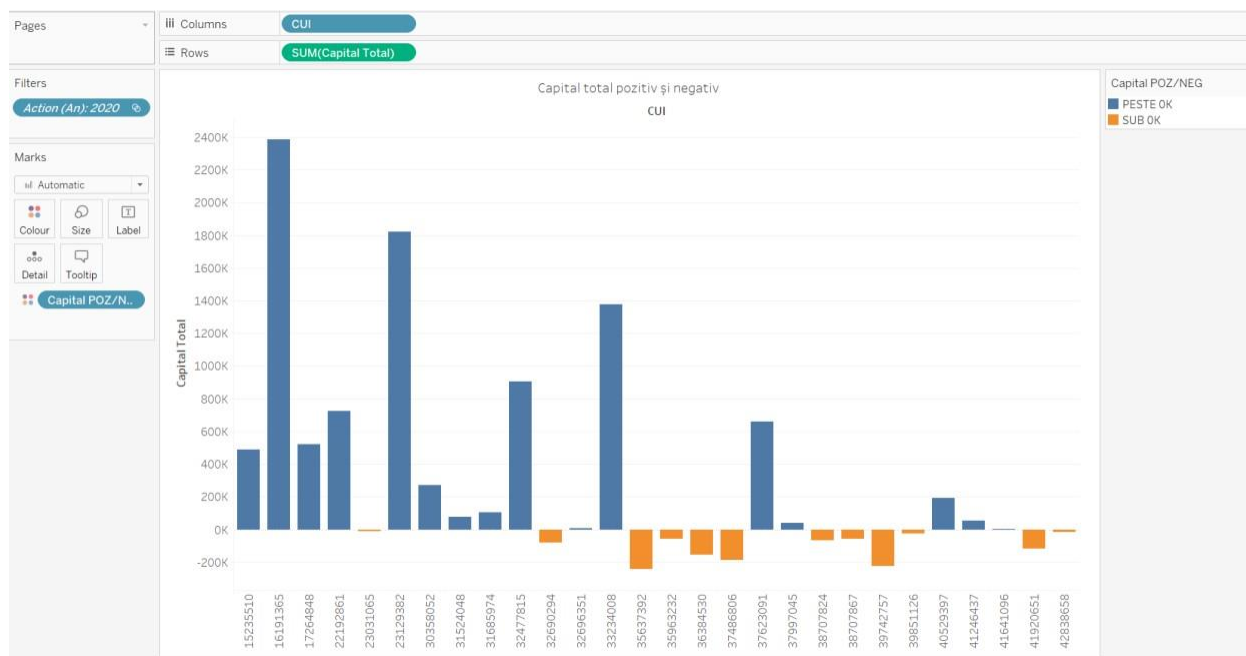
### 3.2.3 Comparație între rata profitului pe primii 2 ani analizați



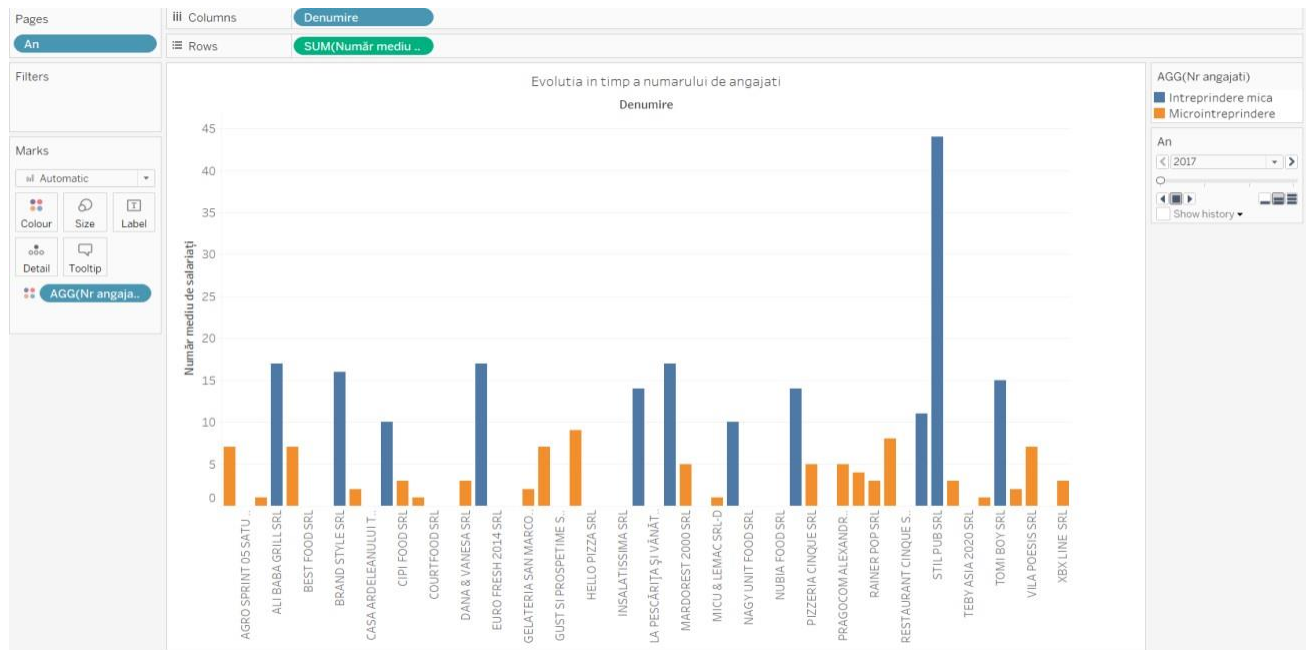
### 3.2.4 Cifrele de afaceri pe ani



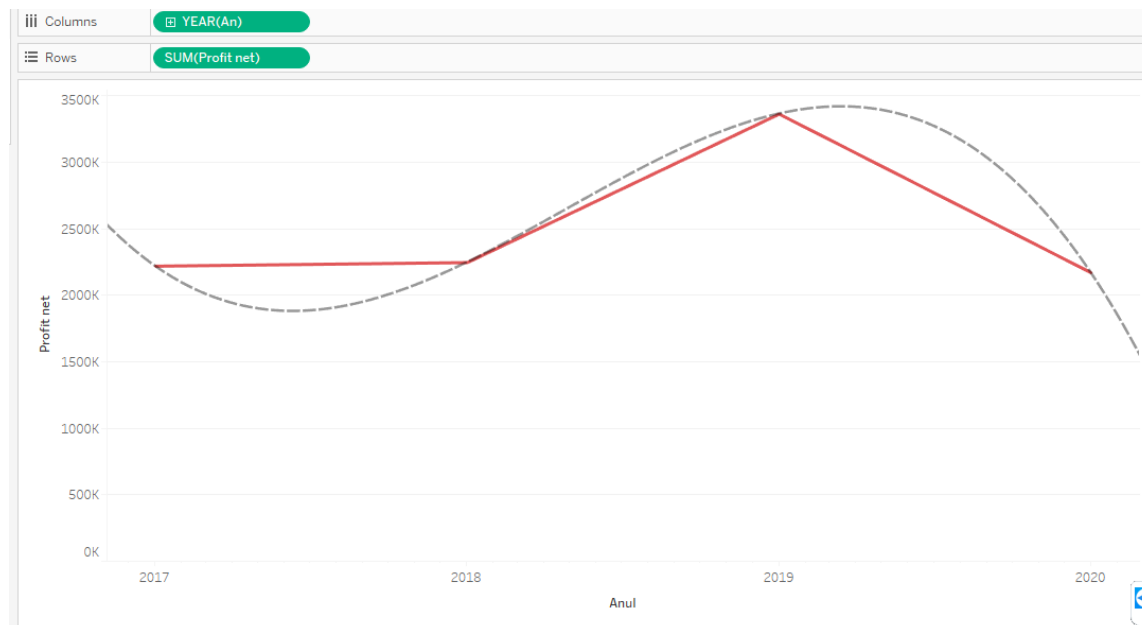
### 3.2.5 Capital total pozitiv și negativ



### 3.2.6 Evoluția numărului de angajați pe ani utilizând range



### 3.2.7 Tendința profitului net în funcție de timp

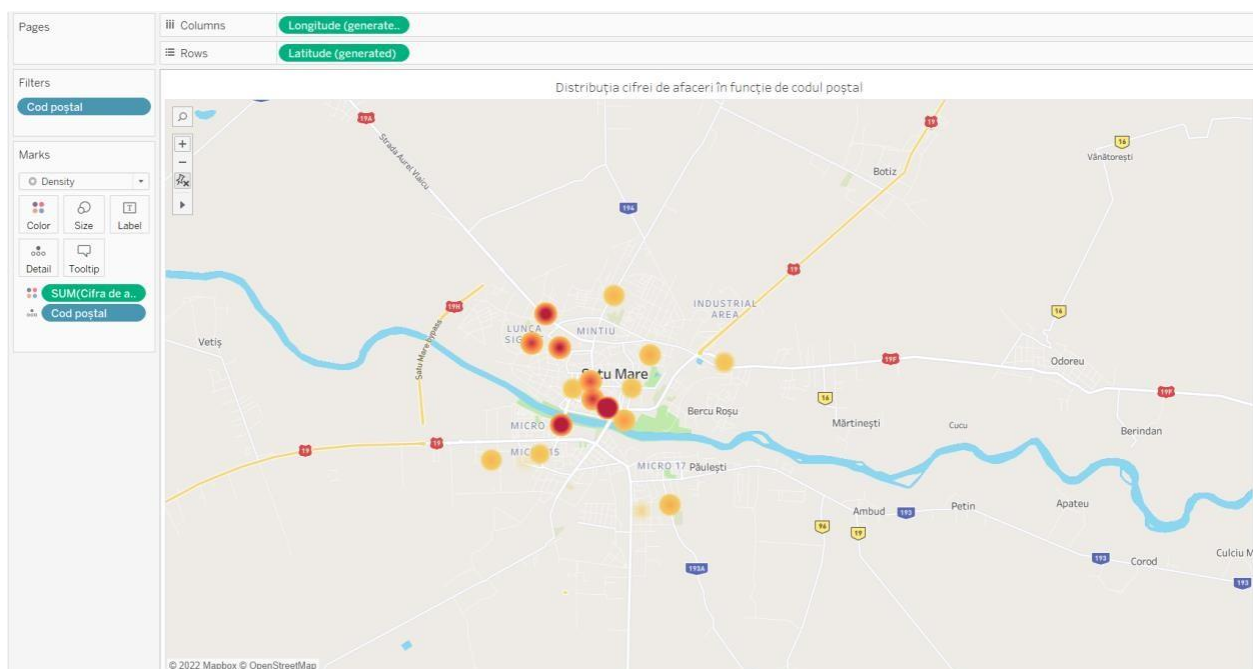


Pentru crearea tendinței am ales modelul de regresie pe baza a 4 observații (Profitul net total din anii 2017, 2018, 2019, 2020). S-a generat o tendință polinomială, cu  $R^2 = 1$ , ceea ce demonstrează o legătură foarte puternică între variația profitului net și timp.

Se remarcă o tendință negativă în perioadele 2017 – 2018 și 2019 – 2020, iar în 2018 – 2019 o tendință pozitivă.

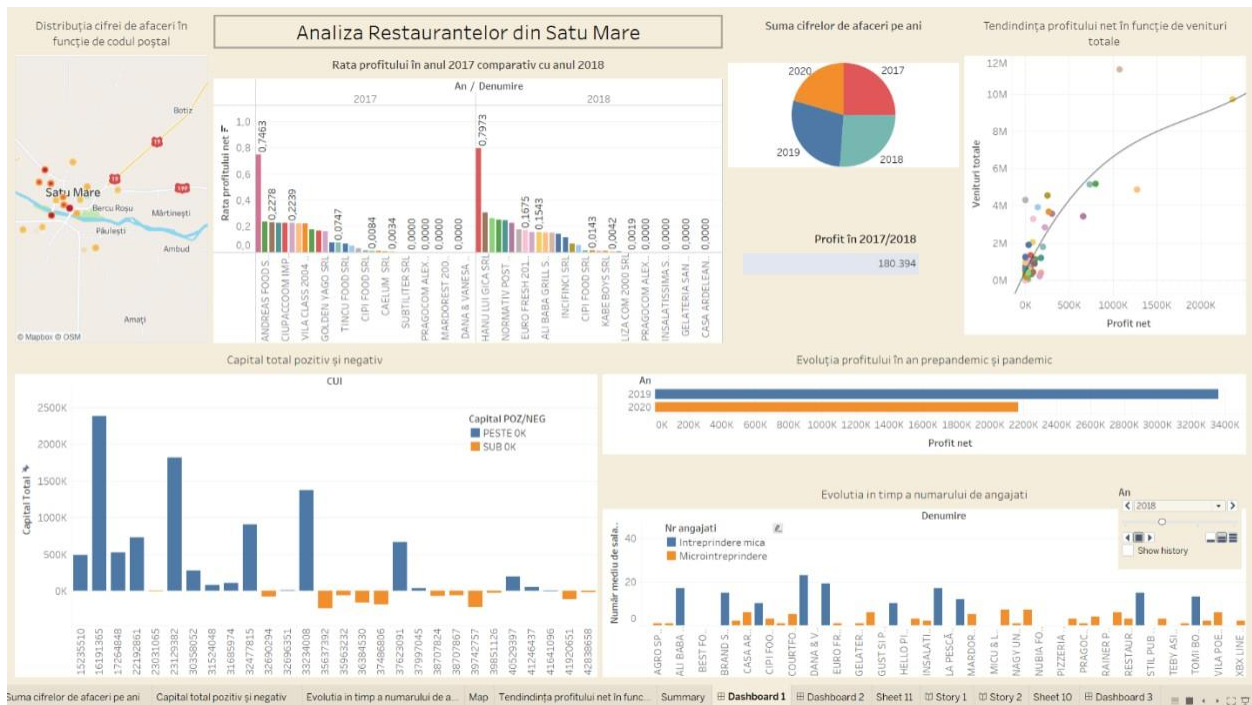
Scăderea profitului în anul 2020 coincide și cu perioada pandemică din România.

### 3.2.8 Distribuția cifrei de afaceri în funcție de codul poștal



## 4. Creare Dashboard cu date legate

Dashboard-ul creat afișează cele mai relevante informații în analiza financiară a restaurantelor analizate. Scopul acestuia este de a facilita înțelegerea datelor.



## 5. Crearea unei povești (Story)



Pentru prezentarea analizei financiare a restaurantelor din municipiul Satu Mare pe perioada 2017-2020 am realizat un story pe baza graficelor și diagramelor cuprinse în dashboard-urile Capital total pozitiv/negativ și cifrele de afaceri pe ani, Evoluția profitului și a ratei profitului ale firmelor analizate în anii 2017/2018 și Analiza Restaurantelor din Satu Mare.



