

1. Description -The business problem for which the project was carried out:

IT is one of the most important fields in our country. An increasing number of people choose to work in this field because it offers material and other benefits that are far superior to other fields.

In Romania, studies show that in 2022 there will be more than 200,000 employees in the IT sector, producing almost 7% of Romania's GDP. In the last 4 years, GDP has grown from 4.7% to 7% and the number of employees in the sector is increasing by 10,000 every year.

So the project I have chosen to carry out is focused on the two branches of this field (software development and hardware components production), which includes the following CAEN codes:

- CAEN 2611 - Manufacture of electronic sub-assemblies (modules)
- CAEN 2612 - Manufacture of other electronic components
- CAEN 2620 - Manufacture of computer and peripheral equipment
- CAEN 2630 - Manufacture of communication equipment
- CAEN 2640 - Manufacture of consumer electronics
- CAEN 2651 - Manufacture of instruments and appliances for measuring, checking and testing, control, navigation
- CAEN 2660 - Manufacture of irradiation, electrodiagnostic and electrotherapy equipment
- CAEN 2670 - Manufacture of optical instruments and photographic equipment
- CAEN 2680 - Manufacture of magnetic and optical recording media
- CAEN 6201 - Custom software development activities (client-oriented software)
- CAEN 6202 - Information technology consultancy activities
- CAEN 6203 - Management and operation of means of transport calculation
- CAEN 6209 - Other information technology service activities

Codes with 26 are represented by the equipment manufacturing industry hardware, while codes with 62 present the software development industry.

The 4 years I have chosen as reference are: 2017, 2018, 2019, 2020.

county (where I was born), but because it is currently not so well known as a to be developed I made a comparison with Cluj county.

I have taken into account economic aspects such as the comparison of revenues, net or gross profit of the companies, aspects related to the nature of the business (fixed and current assets or stocks), as well as social aspects such as the average number of employees and its evolution. The tables show the field values in all 4 years filtered by county, city or code

fall.

The most important year is 2020, when the global economy has been hit hard by the pandemic. In the graphs I have also shown that the IT sector has not been affected. On the contrary, both profits and the number of employees grew at a faster pace than in the previous years.

2. Data preparation

For this project I used the following datasets from the data.gov platform:

- WEB_BL_BS_SL_AN2017.csv
- WEB_BL_BS_SL_AN2018.csv
- WEB_BL_BS_SL_AN2019..csv
- WEB_BL_BS_SL_AN2020.csv
- 3open_data-neradiate_cu_sediu-08.02.2021.csv.

In the first step I added the 5 tables to Tableau Prep. I started with processing the balance sheet data for each column. I changed the data type from Decimal to String for the CAEN and CUI columns and dropped the following columns: I6,I8,I9,I10,I11,I12. I renamed the remaining columns as follows: Fixed assets 20xx., Current assets 20xx, Stocks 20xx, Receivables 20xx, House and bank accounts 20xx, Payables 20xx, Net turnover 20xx, Income 20xx, Expenses 20xx, Gross profit 20xx, Gross loss 20xx, Net profit a20xx,Net loss 20xx, Average number of employees 20xx.

<input type="checkbox"/>	Type	Field Name	Original Field Name	Changes	Preview
<input checked="" type="checkbox"/>	Abc	CUI	CUI	Abc	13609073, 13698860, 13854617
<input checked="" type="checkbox"/>	Abc	CAEN	CAEN	Abc	121, 1413, 2562
<input checked="" type="checkbox"/>	#	Active imobiliza...	I1	I1	29,284,037, 304,402, 24,509,123
<input checked="" type="checkbox"/>	#	Active circulant...	I2	I2	24,089,162, 240,643, 4,726,050
<input checked="" type="checkbox"/>	#	Stocuri 2019	I3	I3	8,827,332, 38,554, 3,680,670
<input checked="" type="checkbox"/>	#	Creante 2019	I4	I4	14,885,840, 188,093, 627,916
<input checked="" type="checkbox"/>	#	Casa si conturi l...	I5	I5	375,990, 13,996, 417,464
<input type="checkbox"/>	#	I6	I6	I6	942,473, null, 79,480
<input checked="" type="checkbox"/>	#	Datorii 2019	I7	I7	27,859,966, 1,959,544, 11,675,645
<input type="checkbox"/>	#	I8	I8	I8	9,303,183, null, 5,716
<input type="checkbox"/>	#	I9	I9	I9	null, 102,460
<input type="checkbox"/>	#	I10	I10	I10	17,152,523, -1,414,499, 17,530,832
<input type="checkbox"/>	#	I11	I11	I11	100,000, 500, 3,543,280
<input type="checkbox"/>	#	I12	I12	I12	null
<input checked="" type="checkbox"/>	#	Cifra de afaceri ...	I13	I13	34,305,632, 8,364, 24,516,658
<input checked="" type="checkbox"/>	#	Venituri 2019	I14	I14	40,672,784, 8,364, 24,793,474
<input checked="" type="checkbox"/>	#	Cheltuieli 2019	I15	I15	39,084,534, 681,444, 24,732,764
<input checked="" type="checkbox"/>	#	Profitul brut 20...	I16	I16	1,588,250, 0, 60,710
<input checked="" type="checkbox"/>	#	Pierdere bruta 2...	I17	I17	0, 673,080
<input checked="" type="checkbox"/>	#	Profitul neta 20...	I18	I18	1,357,294, 0, 15,082
<input checked="" type="checkbox"/>	#	Pierdere neta 2...	I19	I19	0, 673,080
<input checked="" type="checkbox"/>	#	Numar mediu de...	I20	I20	153, null, 89

Fig 1- Renaming fields

I performed a cleaning step, in which I filtered the CAEN column by the chosen codes: 2611, 2612, 2629, 2630, 2640, 2651, 2652, 2660, 2670, 620, 6201, 6203, 6209.

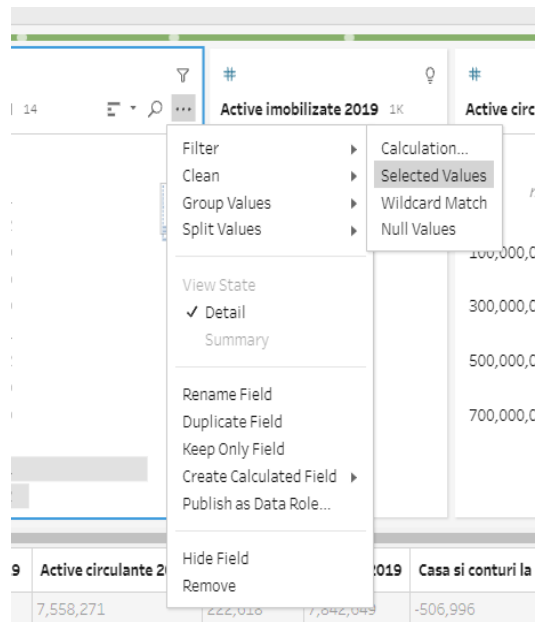


Fig 2- Field cleaning

After making these changes I made 3 consecutive joins which left me with 963 records. All these joins kept only common records (inner join) based on the condition CUI = CUI. I deleted the repeating columns for CUI and CAEN.

CUI	CAEN	Active immobilizate 2017	Active circulante 2017	Stocuri 2017	Creante 2017	Casa si conturi la banci 2017	Cifra de afaceri neta 2017	Venituri 2017
10026547	2611	12,500,000	150,000,000	75,000,000	325,000,000			
10144143	2612	50,000,000	200,000,000	325,000,000				
10175867	2620	87,500,000	200,000,000	325,000,000				
10184334	2630	125,000,000	200,000,000	325,000,000				
10210166	2640	162,500,000	200,000,000	325,000,000				
10228872	2652	200,000,000	200,000,000	325,000,000				
10303400	2660	237,500,000	200,000,000	325,000,000				
10329907	620							
10372559	6201							
10441497	6202							
10608391	6203							
10672947	6209							

Fig 3- Result join

I went on to clean up the data. I unchecked the EUID and FIRM_STATUS fields and converted the CUI from Decimal to String.

Type	Field Name	Original Field Name	Changes	Preview
<input checked="" type="checkbox"/>	Abc	DENUMIRE	DENUMIRE	DONOIU ANA DONOIU MARIA ÎNTEPRINDERE FAMILIALĂ, NIMARA DAN ÎNTEPRINDERE FAMILIALĂ, ILI...
<input checked="" type="checkbox"/>	Abc	CUI	CUI	4500254, 2343340, 20481030
<input checked="" type="checkbox"/>	Abc	COD_INMATRICULARE	COD_INMATRICULARE	F40/71/1993, F40/90/1993, F40/107/1993
<input type="checkbox"/>	Abc	EUID	EUID	ROONRC.F40/71/1993, ROONRC.F40/90/1993, ROONRC.F40/107/1993
<input type="checkbox"/>	Abc	STARE_FIRMA	STARE_FIRMA	1074,2080,2120, 1048,2080
<input checked="" type="checkbox"/>	Abc	ADRESA	ADRESA	București Sectorul 1, Str. PIETII, Nr. 55, Cod poștal 84541, București Sectorul 2, Sos. COLENTINA, Nr. 2C, ...

Fig 4 - Cleaning of the unadjusted data table

After that, I performed a cleanup step in which I deleted records that had the CUI field null. From the address field with the help of a Custom Split I was able to get the County, City and Postcode field. I performed a filtering with Selected Values in which I chose Bihor and Cluj counties.

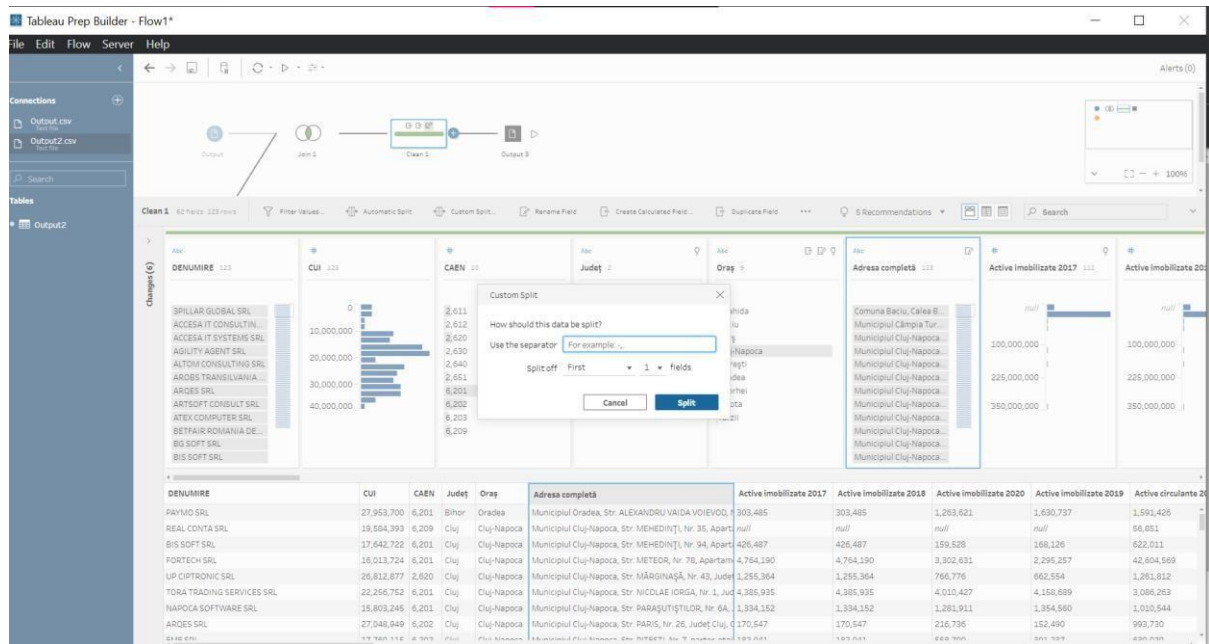
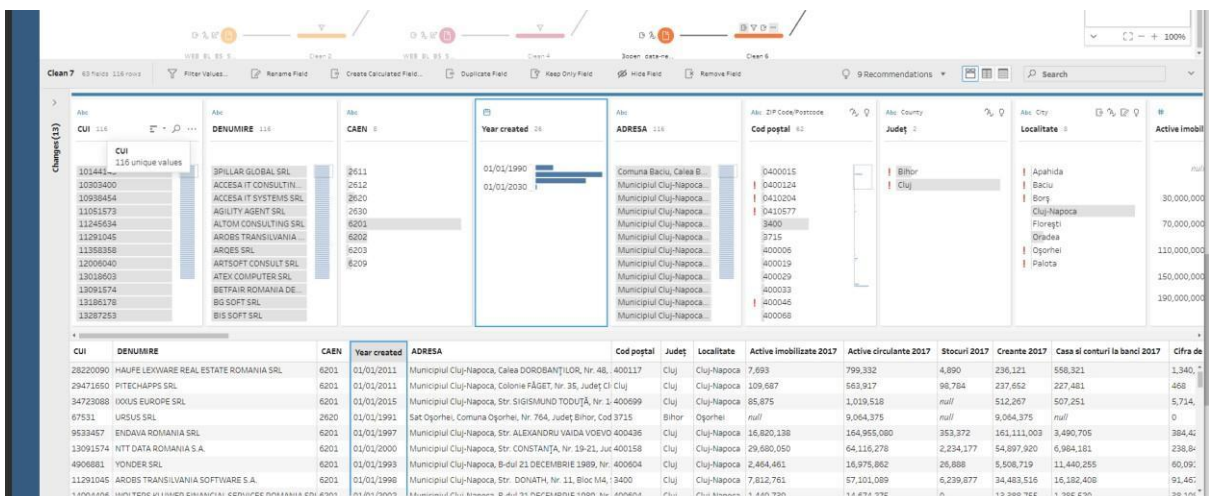


Fig 5- Field extraction - Locality



Also using Custom Split I obtained from the EUID code, the year the company was founded and changed it to type Date. After all the changes I made, I joined the two tables, obtaining 116 records. This join keeps only the common records from the two tables on the basis of CUI = CUI (Unique Registration Code).

Fig 6 - Final Table

3. Data Visualization & Analysis

In the first analyses I compared the number of companies in the software development industry and in the hardware components industry in Bihor and Cluj Counties. It can be seen that in the case of companies with CAEN code 62., the number of employees is higher in Cluj than in Bihor. However, in the case of companies with CAEN codes 26., there are more companies in Bihor probably due to the fact that there is a very developed industrial park around Oradea.

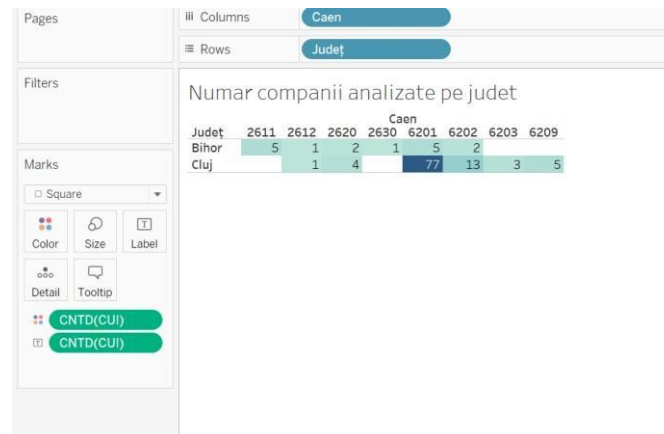


Fig 7 - Number of companies analysed by County

In the following two analyses you can see the companies and the year of establishment in both Cluj and Bihor counties. In both cases the increase in the number of companies was progressive and started in 1991 and 1992 respectively.

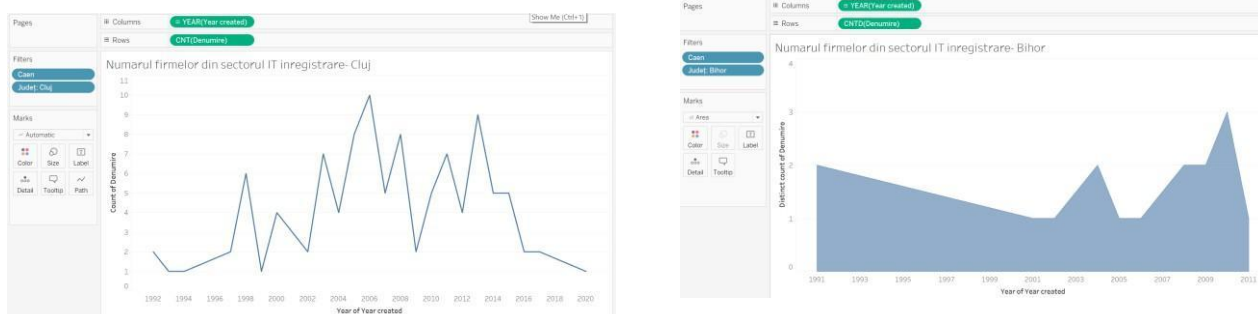


Fig 8 - Company set-up chart

To illustrate that for the software development industry the assets are small compared to the manufacturing industry I have created two graphs. For better organizing the data I created two Dashboards: one that does a comparative analysis between the two counties and a more detailed one for the Bihor county (but in which I used all the data from the selected counties).

In the first dashboard, it can be seen that for Bihor county the number of employees in production is much higher than the other categories.

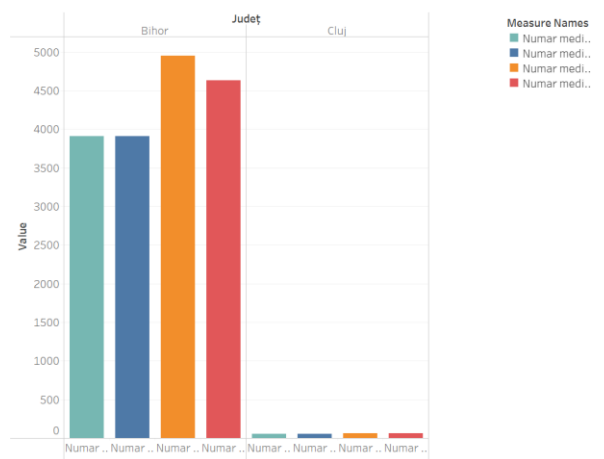


Fig 9- Number of production employees Cluj vs Bihor

In Cluj county, the ratio is exactly the opposite. A very high number of employees for software development, over 15,000 for a single CAEN code.

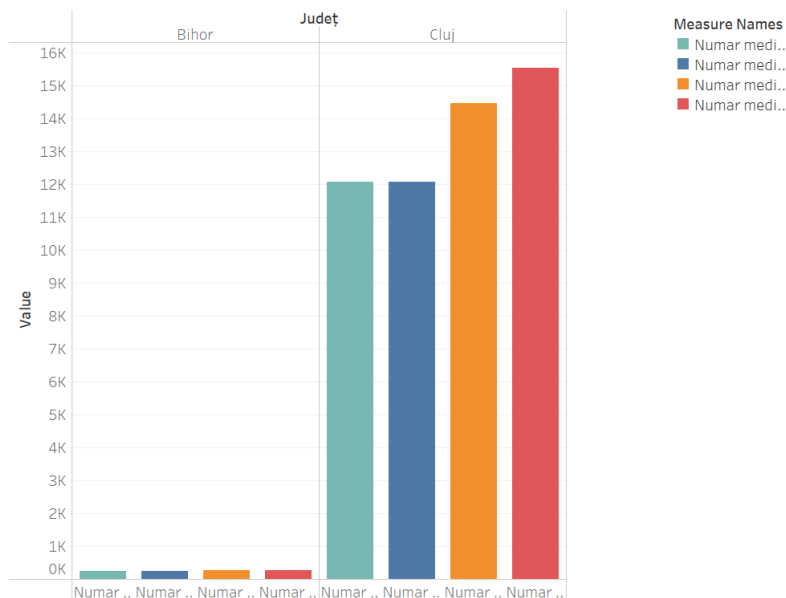


Fig 10- Number of software development employees Cluj vs Bihor

After that, I analysed the net income according to caen codes: in Bihor the highest income comes from companies with caen codes 2611 and 2622, and in Cluj county the highest share is for caen code 6201.

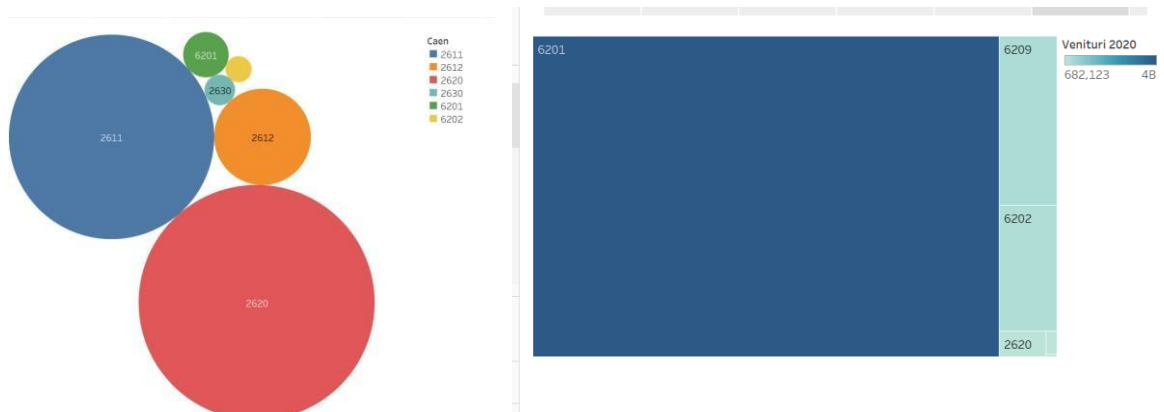


Fig 11 - Net income filtered by caen codes

The following graph shows the average number of employees for each county over the four years analysed and is noticeably higher in Cluj county.

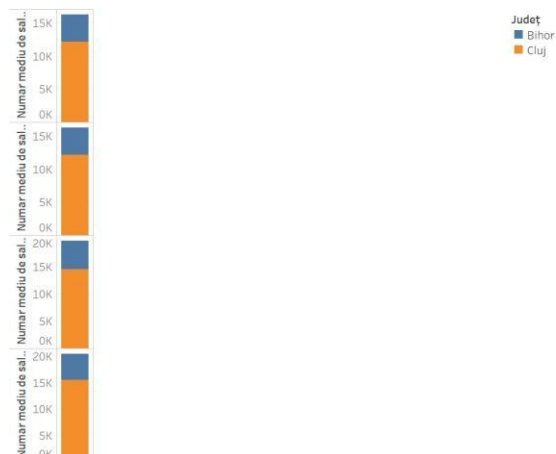


Fig 12 - Average number of employees per county

The map is composed of 2 layers: one showing the amount of gross profit for the two counties in 2020 and one showing the amount of gross profit for the localities in the county in 2020. In order to use the Locality field, it was necessary to add the geographical coordinates of some localities.

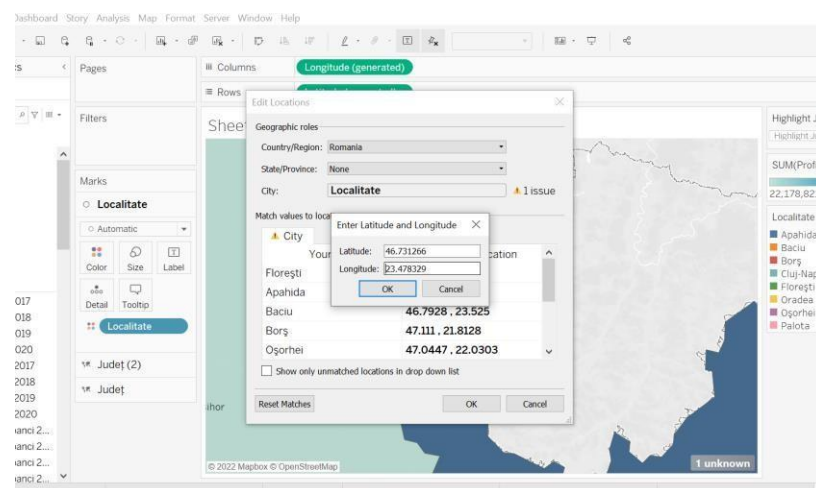


Fig 12 - Select coordinates for map

Last but not least, with the help of a story I made 4 graphs showing the net profit for each category of enterprise over the 4 years analysed.

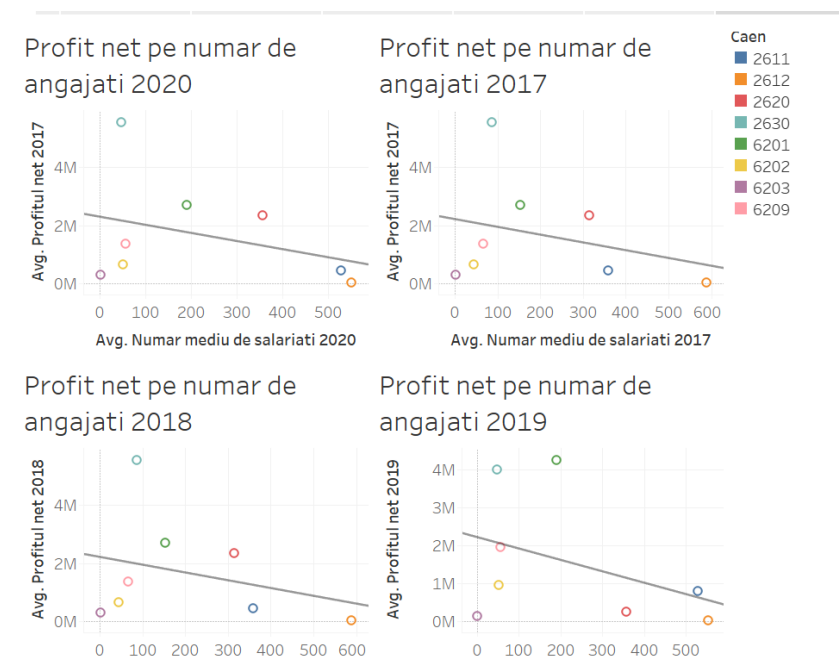


Fig 13 - Net profit for the four years analysed

For Bihor county I have made an extensive analysis, examining the net profit for the four localities where the companies in this sector are located. Most of the profit is generated in Oradea, this being the county seat.

Localit..	Profitul net 2017	Profitul net 2018	Profitul net 2019	Profitul net 2020
Borș	12,147,373	12,147,373	0	0
Oradea	16,539,584	16,539,584	20,205,102	18,502,427
Oșorhei	0	0	0	0
Palota	250,026	250,026	499,542	941,805

Fig 14 - Net profit by locality Bihor

In the next story I compared the turnover for the caen codes in the 4 years analysed and I noticed that it increases progressively from year to year and that it is the highest for the caen code 2620 in all four years. One last interesting observation is that although in 2020 the labour market was affected by the pandemic and the unemployment rate increased by more than 13% percent, in the area studied the number of employees increased in both counties. It can also be seen that the net profit has increased in 2020 for all CAEN codes.

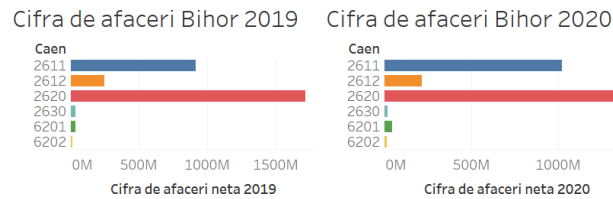
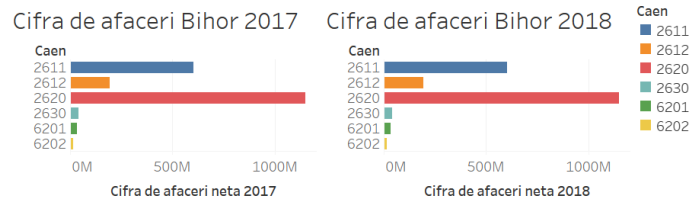


Fig 15- Turnover Bihor

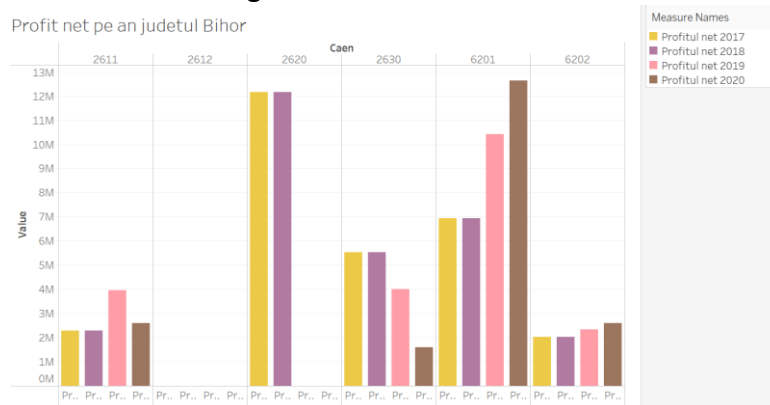


Fig 16 - Net profit Bihor

In the last story there is a comparison between income and expenditure.

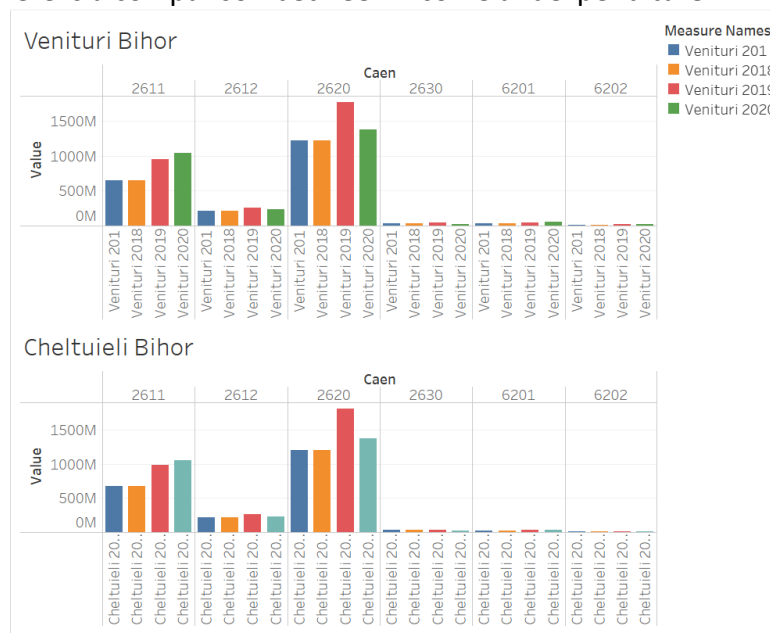


Fig 18- Income vs expenditure