

Proiect

**Analiza firmelor care activează în domeniul întreținerii
autovehiculelor din județul Bihor**

Student: Malan Andrei

Anul 3, grupa 3

Cuprins

Descrierea problemei de business	3
Pregătirea datelor	3
Data visualisation & Analysis	11
Venituri totale pe localitate pe an	11
Profiturile între 2018-2021 ale firmelor din Bihor grupate pe localitate	12
Profit raportat la numărul mediu de angajați	13
Evoluția profitului în cei 4 ani pentru Oradea & zona metropolitană	13
Harta în funcție de media profitului net	14
Cluster	15
Analiza de trend	16
Crearea de Dashboard	17
Interactivitate	17
Crearea poveștii	19

Descrierea problemei de business

Pentru realizarea acestui proiect s-a ținut cont de situațiile financiar pe 4 ani(2018, 2019, 2020, 2021) ale firmelor care prestează servicii în domeniul întreținerii autovehiculelor, aflate în județul Bihor, acestea având ca și cod CAEN codul 4520. Nevoia de analiză a plecat de la o problemă personală identificată, prin care de câte ori am nevoie de reparații sau mentenanță auto sunt nevoit să aștept o perioadă destul de mare de timp, fapt ce îmi creează un discomfort considerabil. Astfel, analiza pe care am facut-o se va axa pe câțiva factori pe care îi văd relevanți pentru a putea trage concluzii relevante:

- Cifra de afaceri de netă
- Venituri totale
- Cheltuieli totale
- Profit net
- Pierdere netă
- Număr mediu de angajați

Pregătirea datelor

Pentru a obține datele de care avem nevoie, am accesat platforma Data Gov, platformă a Guvernului României, disponibilă la linkul <https://data.gov.ro/>. Am utilizat următoarele seturi de date:

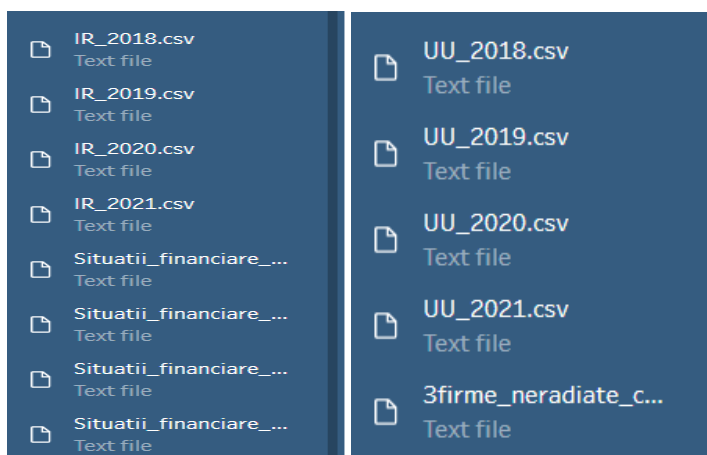
- web_uu_an2018 în format text
- web_uu_an2019 în format text
- web_uu_an2020 în format text
- web_uu_an2021 în format text
- web_bl_bs_sl_an2018 în format text
- web_bl_bs_sl_an2019 în format text
- web_bl_bs_sl_an2020 în format text
- web_bl_bs_sl_an2021 în format text

- web_ir_an2018 în format text
- web_ir_an2019 în format text
- web_ir_an2020 în format text
- web_ir_an2021 în format text

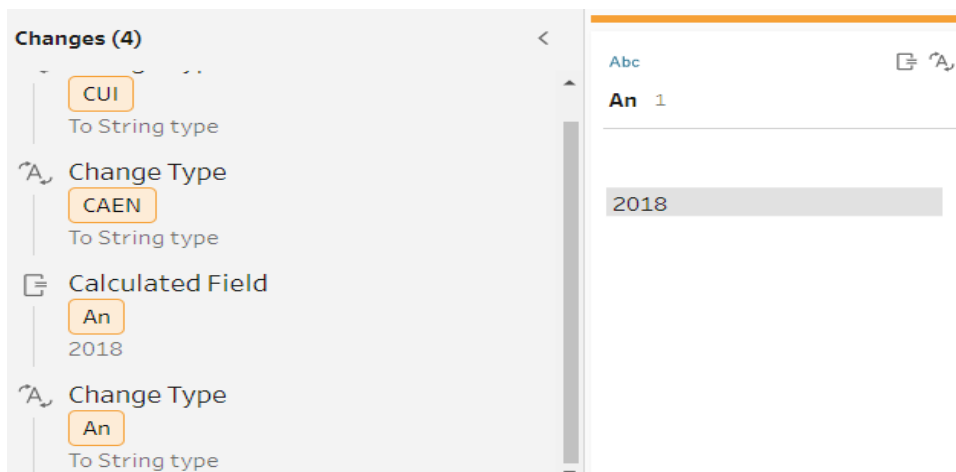
Aceste documente au ca sursă informațiile oferite de Ministerul Finanțelor

- 3firme_neradiate_cu_sediu_2022-07-07 în format text, document oferit de Registrul Comerțului

Următorul pas a fost să încărcăm documentele ce activează pe post de sursă în cadrul Tableau Prep, astfel încât să putem demara procesul de curățare a datelor. Datele inițiale erau în format text, iar în cadrul Tableau Prep le-am inclus sub forma unor documente cu extensia csv. Astfel, primul pas arată în felul următor:

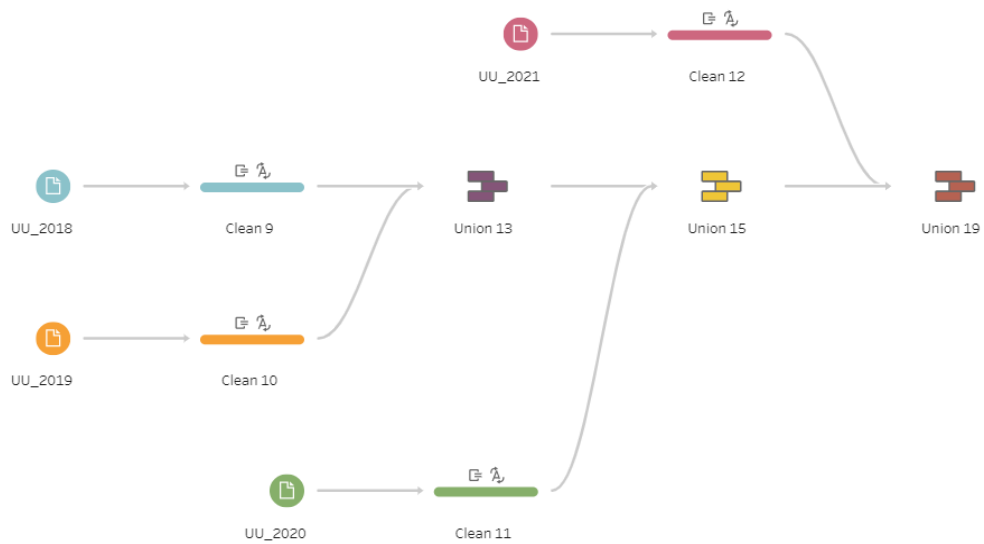
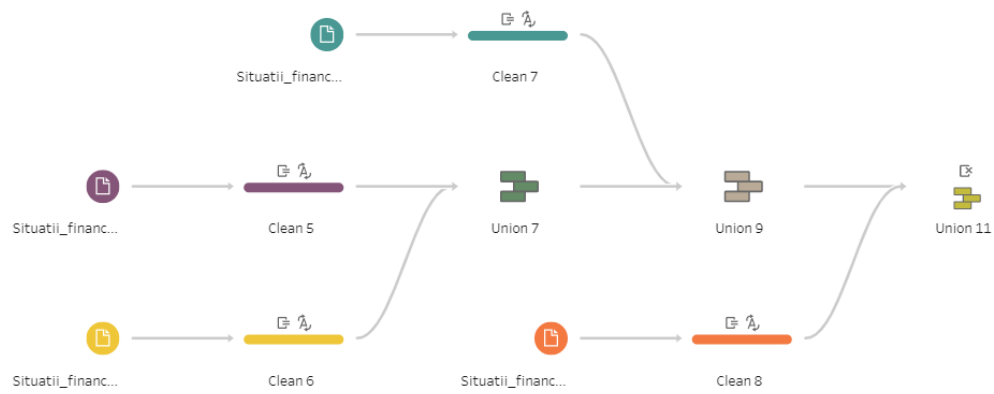
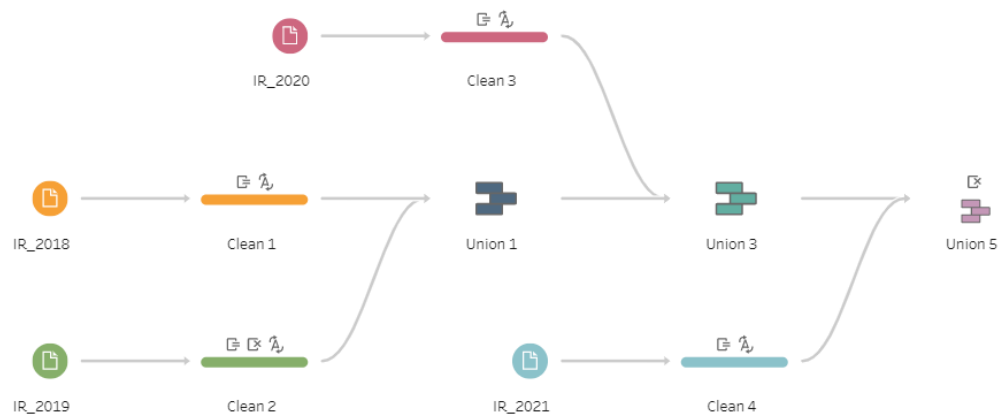


După ce am încărcat sursele de date, am ales să fac un pas de Clean Step, pentru a putea include un câmp care să conțină anul calendaristic pentru care s-a realizat procesul. Motivația a fost faptul că, în cadrul analizei pe care doresc să o fac, temporalitatea poate juca un rol important și dezvălui tendințe relevante pentru un context de business. Alături de adăugarea anului, am schimbat tipul acestuia în string, lucru realizat și pentru codul CAEN(starea lor precedentă era în forma de number)

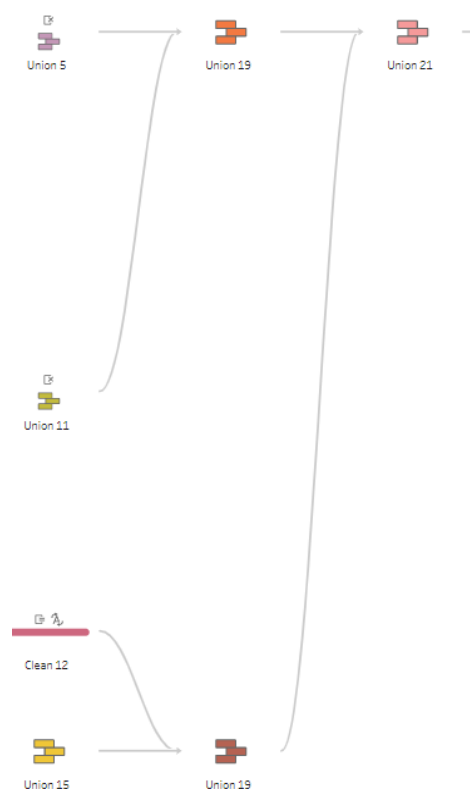


În continuare, am demarat procesul de Union aferent fiecărui document. Acest lucru a presupus aducerea laolaltă a celor 4 documente de bl_bs, în cadrul unui singur pas de tip union, acest proces repetându-se și în cazul documentelor de tip ir sau uu. Ca și pași intermediari, a existat câte o operație de tip Union între fiecare Union precedent și Clean Stepul aferent unui an neinclus deja în cadrul uniunii. Pentru exemplificare, voi folosi următorul exemplu: pentru documentele IR 2018 și 2019 am creat Union1(care deci va conține în acest moment toate informațiile financiare aferente anilor 2018 și 2019), pe baza căruia am mai făcut Union3, ce va cuprinde și informațiile financiare ale anului 2020. Procesul se repetă și pentru anul 2021, incluzând Union 5. Aceste ultime union au fost create având la bază pașii de Clean Step aferenți anilor 2020, și 2021. Practic, în final, va exista un singur Union(Union5), care va include toate informațiile financiare ale anilor 2018-2021, acesta stând la baza viitoarelor operațiuni de union și join pe care le-am făcut.

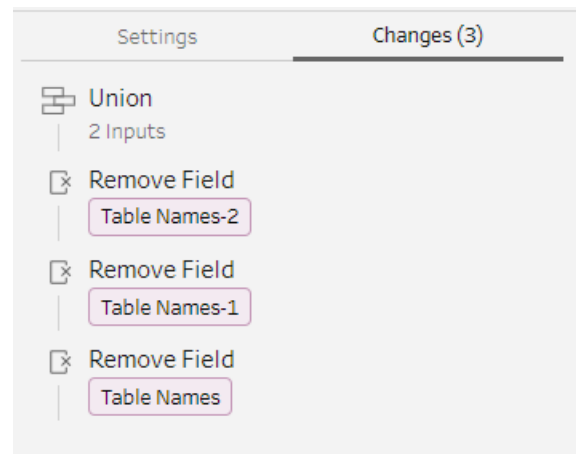
Procesul va fi identic și pentru documentele de tip bilanț lung-bilanț scurt sau ir, astfel că la final vom avea 3 Union-uri(Union 5, 11, 19) cu care vom merge mai departe în prelucrarea datelor.



Având astfel cei 3 pași de union care conțin toate informațiile necesare de care avem nevoie, am ales să fac un pas care să unească aceste 3 documente, și de această dată cu pas intermediar(între primele 2 Union-Union 5 și 11 am făcut Union 19, care s-a unit ulterior cu al treilea Union), rezultând uniunea finală a datelor financiare de care am nevoie pentru a analiza zona de business aleasă.



Ca și pași de curățare, am șters coloanele care conțineau titlurile documentelor precedente, astfel încât pasul final să conțină strict date cu caracter și relevanță financiară.



Având documentele financiare, trebuie să le legăm de firmele care le-au realizat de-a lungul acestor ani. Aici intră în joc datele obținute de la Registrul Comerțului, acestea conținând informații despre firmele neradiate cu sediu de la data de 7 aprilie 2023. După ce am încărcat fișierul, am ales să fac un pas de Clean Step, care să îmi permită să fac anumite modificări asupra datelor. Astfel, am schimbat tipul de date al CUI-ului în String și am eliminat o serie de coloane pe care nu le-am considerat relevante pentru analiza mea(EUID, Cod Înmatriculare, Strada, Bloc, Nr înregistrare etc).

Changes (3)

- Change Type
 - CUI
 - To String type
- Remove Fields
 - COD_INMATRICULARE
 - EUID
 - ...
- Remove Field
 - F19

Mai departe, având acest pas de Clean Step făcut, care conține informații despre pe care le consider importante firme, plus Union 21, care conține datele financiare despre toate aceste firme în intervalul 2018-2021, am aplicat un pas de JOIN, având la bază câmpul comun ce se referă la CUI-ul firmelor.



Join_final1 32 fields 493 K rows | Filter Values... | Create Calculated Field...

Settings | Changes (0)

Applied Join Clauses

Union 21 = Clean 13

CUI = CUI

Join Type : Inner

Click the graphic to change the join type.

Union 21 Clean 13

Summary of Join Results

Click the bar segments to view the included and excluded values.

Mismatched values

	Included	Excluded
Union 21	492,616	2,231,436
Clean 13	163,113	115,956
Join Result	492,616	

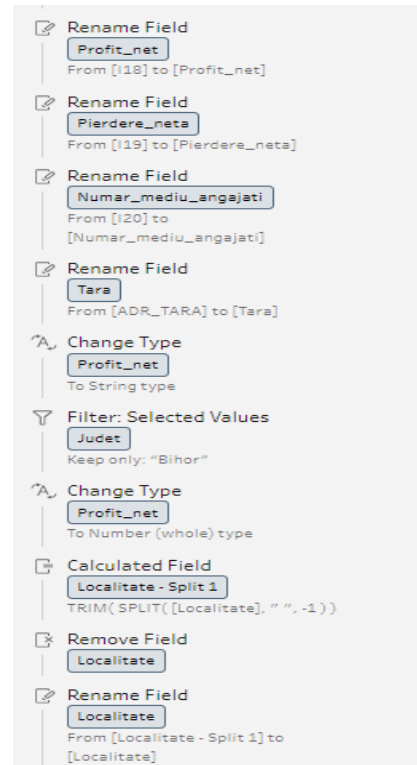
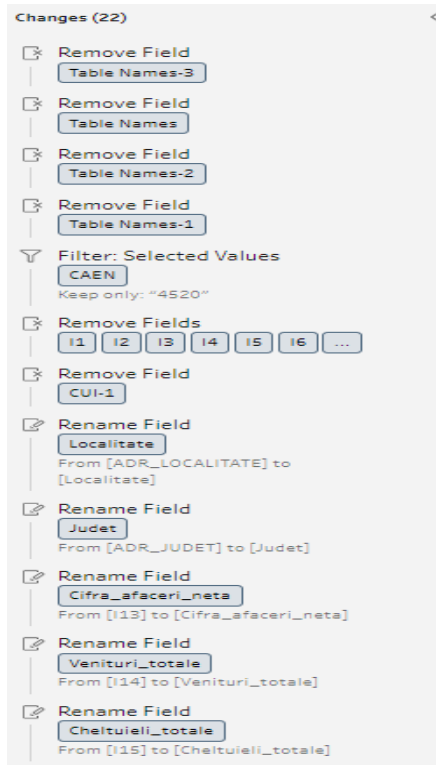
Join Clause Recommendations

Join Clauses ☐ Show only mismatched values

Union 21	Clean 13
↑ CUI	↑ CUI
10000270	null
10000326	10024511
10000334	10028726
10000350	10029969
10000393	10035324
10000415	10043262
10000458	10088728
10000474	10114800
10000482	10144097
10000610	10152464
10000652	10173777
10000679	10181079
10000733	10197563
10000750	10204636
10000784	10220992
10000890	10239812
10000970	10257762
10001127	10311551
10001186	10318568
10001194	10335704

După cum se poate observa, am realizat un Inner Join, pe baza CUI-ului, din care am obținut 492.616 rezultate care se potrivesc. Acestea ulterior vor fi filtrate, iar numărul va fi mult mai mic în momentul în care vom face analiza de business.

Mai departe, am făcut un nou pas de Clean Step(Clean 14), în care vom realiza ultimele curățări necesare pentru a pregăti fișierul de Outcome, cel care va fi folosit ca și sursă de date pentru vizualizări.



Aceste modificări presupun:

- Înlăturarea coloanelor care conțineau numele precedentelor documente
- Filtrarea pentru un singur cod CAEN, specific zonei de business pe care o urmărim(4520)
- Ștergerea câmpurilor financiare(I1, I2, I3, etc) care conțineau informații mai degrabă de natură contabilă, de genul active imobilizate, provizioane sau creanțe, lucruri pe care nu le consider relevante într-o analiză de business
- Înlăturarea câmpului CUI-1, apărut ca urmare a operației de JOIN(păstrarea lui ar fi însemnat existența unei coloane duplicat în tabel)

- Redenumirea unor câmpuri(Localitate, Județ, Cifra_afaceri_netă, Venituri_totale, Cheltuieli_totale, Profit_net, Pierdere_netă, Număr_mediu_angajați, Țara)
- Schimbarea tipului coloanei Profit_net în tip numeric
- Filtrarea doar pe județul pe care doresc să-l analizez(Bihor)
- Realizarea unui split pe câmpul Localitate, pentru a obține doar numele localității, urmată de ștergera coloanei inițiale.

De menționat faptul că datele în câmpul de localitate erau destul de neregulate. Existau așezări umane care aveau denumiri compuse din 2 sau 3 cuvinte, însă majoritatea aveau un singur cuvânt, și acesta era ultimul. Pentru a găsi o variantă de compromis, am realizat regula prin care doar ultimul cuvânt să fie păstrat, iar pentru acele localități care au nume compus, am decis să completez eu în fișierul excel de output cuvintele lipsă, acest fapt putând fi motivat și de faptul că exista un număr mic de astfel de localități si rânduri care au fost afectate.

În final, ultimul pas a fost acela de creare de Output, unde am generat un fișier de tip excel, care urmează să fie folosit în Tableau Desktop pentru a realiza vizualizări și analize specifice domeniului de business.

Save output to		Save to Output_project_final.xlsx												
File		An	DENUMIRE	CUI	CAEN	Cifra_afaceri_neta	Venituri_totale	Cheltuieli_totale	Profit_net	Pierdere_neta	Numar_mediu_angajati	Tara	Localitate	Judet
Browse		2018	GEOBARA SERVICE SRL	34206383	4520	4.849.245	5.590.967	5.042.064	470.321	0	20	România	Borș	Bihor
Name		2020	EVIO BENI S.R.L.	35447606	4520	239.660	239.658	145.024	92.994	0	1	România	Aleșd	Bihor
Output_project_final		2020	AGO AUTO EXPRES S.R.L.	34850316	4520	285.751	286.357	277.500	6.166	0	1	România	Oradea	Bihor
Location		2020	TUNA GO WASH S.R.L.	41162441	4520	31.070	40.070	35.245	4.522	0	1	România	Aștileu	Bihor
C:\Users\User\Documents\My Tableau Prep Repository\Datasources		2020	TP TECHNIK PULS SRL	37648770	4520	214.038	233.403	230.158	1.836	0	1	România	Oradea	Bihor
Output type		2020	ALEMIR AUTOSERVICE SRL	37833492	4520	248.499	248.501	170.946	71.718	0	1	România	Sînmartin	Bihor
Microsoft Excel (.xlsx)		2020	EFICRIS TRUCK SRL	38380560	4520	229.381	229.389	180.070	47.190	0	1	România	Borș	Bihor
		2020	NIVA & MIO S.R.L.	41517650	4520	131.162	131.162	90.225	39.256	0	1	România	Oradea	Bihor
Worksheet		2020	CAR WASH & GO S.R.L.	41203959	4520	340.422	340.422	22.099	308.111	0	1	România	Oradea	Bihor
Output_project_final		2020	AWD PERFORMANCE DEVELOPMENTS S.R.L.-D.	38452038	4520	null	0	237	0	237	1	România	Oradea	Bihor
		2020	JENUSHAO S.R.L.	38706144	4520	45.199	45.204	48.899	0	4.254	1	România	Răbăgani	Bihor
Write Options		2020	TUNING DEN SRL	37254121	4520	12.305	12.305	22.742	0	10.713	1	România	Oradea	Bihor
Select an option to create or update your output table.		2020	C&B CARS CLINING S.R.L.	40988969	4520	38.631	38.631	50.212	0	12.197	1	România	Oradea	Bihor
		2020	DEMIOAL TUNING SRL	37966790	4520	18.937	19.176	45.524	0	26.539	1	România	Derna	Bihor
Full refresh		2020	CIUCU PROJECT SRL	38049907	4520	null	3.620	65.160	0	61.556	1	România	Borș	Bihor
Append to table		2020	AIDA ANVELOPE CENTER S.R.L.	41523856	4520	35.060	42.046	40.623	371	0	1	România	Diosig	Bihor
		2020	RAREȘ SPEED CENTER S.R.L.	36235709	4520	2.206	2.206	7.112	0	4.915	1	România	Tileagd	Bihor
		2020	SMARTGAMA SRL	37966707	4520	365.484	367.389	225.344	138.390	0	1	România	Oradea	Bihor
		2020	TDR GABRIEL TRANS SRL	35635588	4520	37.218	37.218	46.609	0	10.435	1	România	Jos	Bihor
		2020	SERVILAR SRL	38380586	4520	15.168	206.818	161.510	39.724	0	1	România	Sîntandrei	Bihor
Run Flow		2020	OSMOSIS STAR SRL	37860683	4520	263.825	263.825	118.897	137.013	0	1	România	Oradea	Bihor

Se poate observa și structura tabelului care va rezulta, cu toate datele curățate și pregătite pentru a fi analizate.

Data visualisation & Analysis

Primul pas pentru a obține vizualizările necesare analizei de business este procurarea setului de date. Cu setul obținut și menționat anterior din Tableau Prep, vom realiza ceea ce ne propunem. Astfel, la început, încărcăm setul de date, și menținem conexiunea Live, astfel încât să lucrăm direct cu fișierul și eventuale modificări pe acesta, nu cu niște copii, în cazul în care am alege varianta de Extract.

Tableau - Proiect_Malan_Andrei

File Data Server Window Help

Connections Add

Output_proiect Microsoft Excel

Sheets

Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.

Output_proiect_final

New Union

New Table Extension

Output_proiect (Output_proiect)

Connection ☒ Live ☐ Extract

Filters 0 Add

Output_proiect_final

Need more data?
Drag tables here to relate them. [Learn more](#)

Output_proiect_final 16 fields 460 rows

100 rows

Name	Output_proiect_final
Denumire	Outputproiectfinal
CUI	Outputproiectfinal
CAEN	Outputproiectfinal
Cifra_afaceri_neta	Outputproiectfinal
Venituri_totale	Outputproiectfinal
Cheltuieli_total	Outputproiectfinal

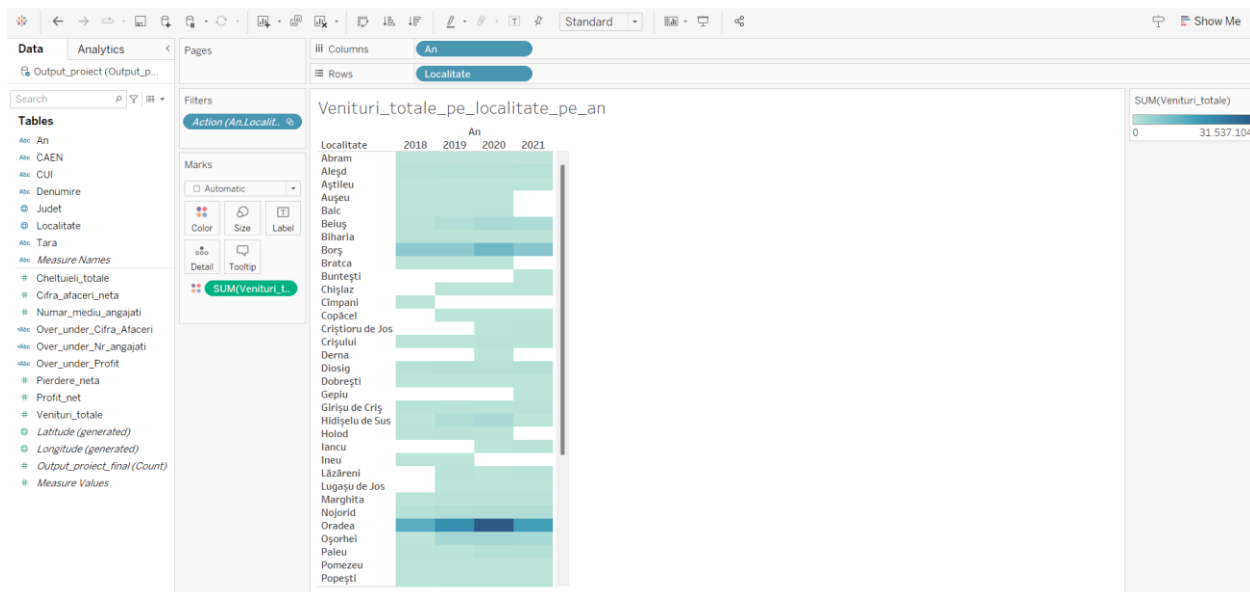
Type	Field Name	Physical Ta...	Rem...
Abc	An	Outputproie...	An
Abc	Denumire	Outputproie...	Denu...
Abc	CUI	Outputproie...	CUI

Abc	Abc	Abc	#	#	#
Outputproiectfinal	Outputproiectfinal	Outputproiectfinal	Outputproiectfinal	Outputproiectfinal	Outputproiectfinal
GEOBARA SERVICE SRL	34206383	4520	5.352.575	9.922.816	9.27
GEOBARA SERVICE SRL	34206383	4520	4.849.245	5.590.967	5.04
GEOBARA SERVICE SRL	34206383	4520	4.849.245	5.590.967	5.04
GEOBARA SERVICE SRL	34206383	4520	6.055.958	6.360.637	5.8
IOSIANKA SRL	34880473	4520	null	0	null
DOSA CAMEX S.R.L.	35679878	4520	null	0	null

Observăm și că avem la dispoziție o previzualizare a setului de date.

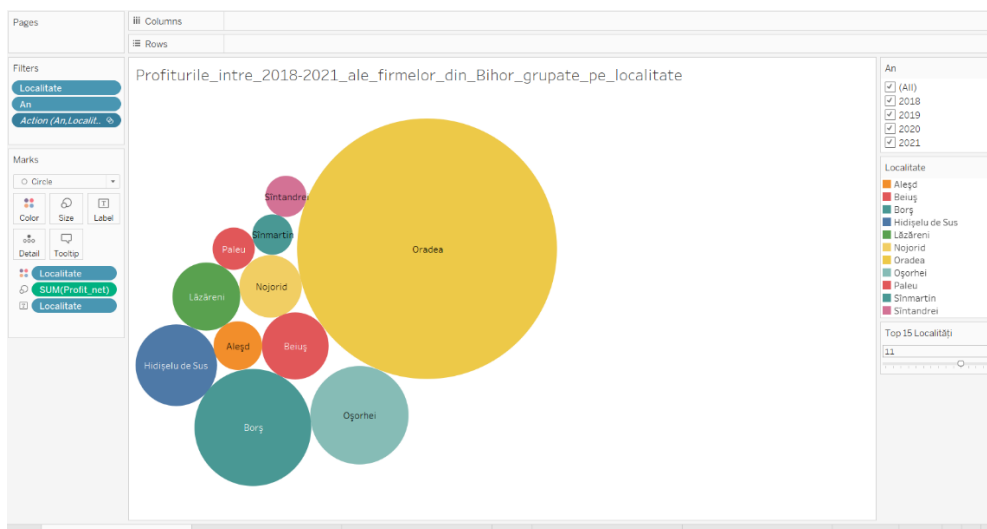
Venituri totale pe localitate pe an

Prima vizualizare pe care o vom realiza relevă distribuția veniturilor în cei 4 ani, pentru fiecare localitate. Am inclus anul pe coloane, localitatea pe rânduri, un filtru grupat ca și set între localități și an, iar cu culori, asemănător unui Heat Map, o măsură de tip numeric care calculează automat suma veniturilor totale pentru fiecare localitate.



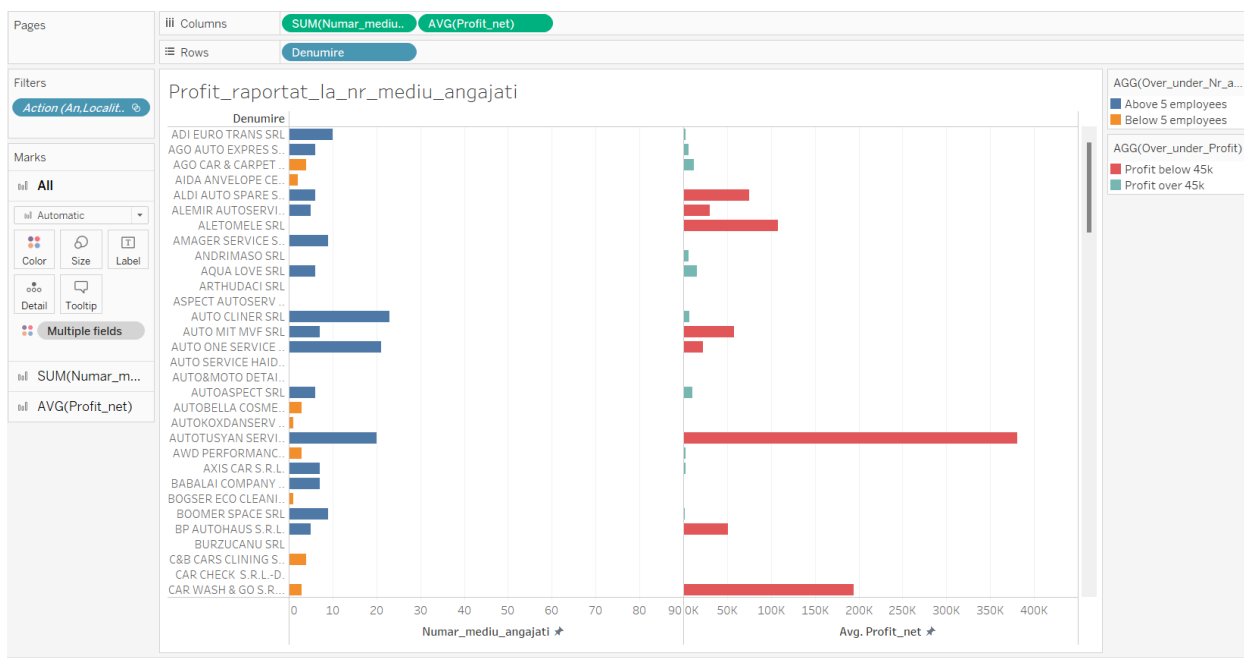
Profiturile între 2018-2021 ale firmelor din Bihor grupate pe localitate

O a doua vizualizare urmărește relația dintre profitul declarat și localitate, pentru cei 4 ani la care ne referim (2018-2021). Am încorporat Profitul net sub formă de size, Localitatea sub formă de culoare și text, iar la filtru am adăugat setul dintre localitate și an menționat și anterior, dar și cele 2 variabile luate separat. Astfel, putem vizualiza ce localități au generat cel mai mare profit pentru zona de întreținere de autovehicule în intervalul de timp analizat. Graficul este de tip circle, astfel vom vedea aceste date sub forma unor cercuri de mărime diferită în funcție de mărimea profitului. Suplimentar, am adăugat un parametru prin care să-mi pot lua primele maxim 15 înregistrări, printr-o bară prin care beneficiarul analizei își poate alege numărul dorit de localități analizate.



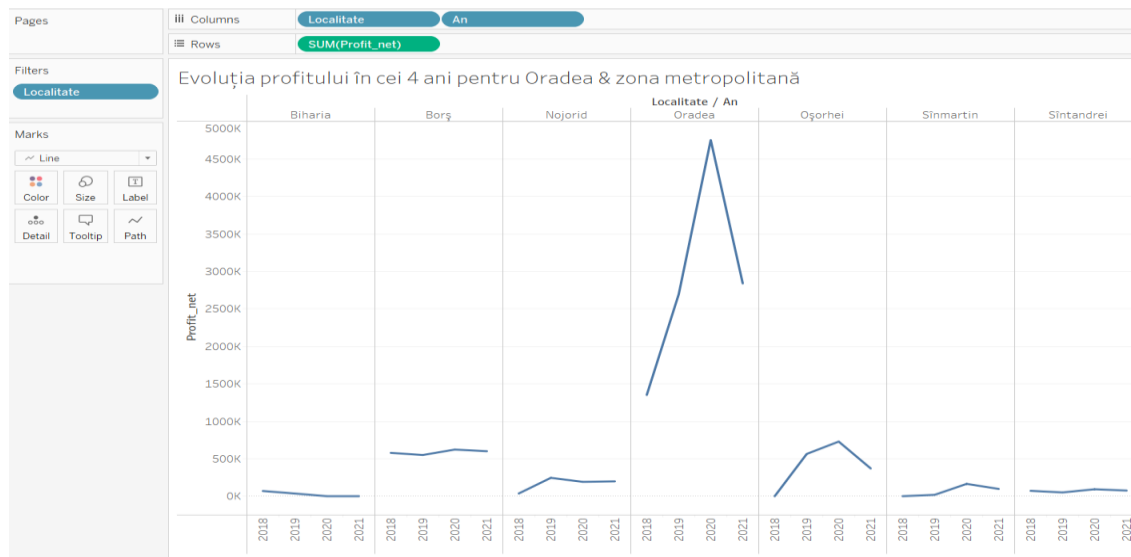
Profit raportat la numărul mediu de angajați

În continuare, vom realiza un grafic care ilustrează relația dintre profitul mediu și numărul de angajați. Pe coloane am introdus o metrică care calculează suma numărului de angajați și media profitului realizat de fiecare firmă, denumirea companiilor fiind pe linii. În plus, am adăugat 2 câmpuri calculate, care verifică pe de-o parte dacă numărul de angajați este peste sau sub 5, iar pe de altă parte dacă media profitul este peste sau sub 45 de mii de lei, fiecare dintre aceste cazuri fiind colorate distinct.



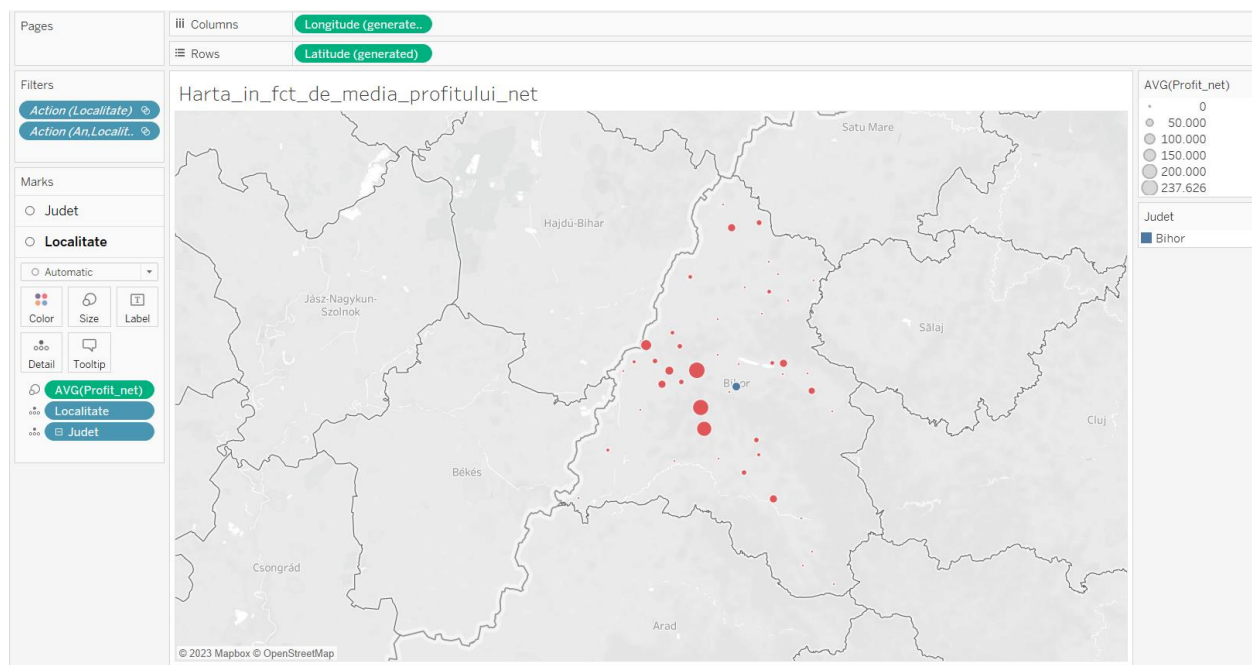
Evoluția profitului în cei 4 ani pentru Oradea & zona metropolitană

Pentru analiza de trend, am propus analiza vânzărilor din perspectiva profitului în cei 4 ani analizați pentru municipiul Oradea și zona metropolitană a acestuia. Am inclus pe coloane localitatea și anul, iar pe linii am pus profitul anual realizat la nivelul localității în anul respectiv. Pentru a putea analiza doar acele localități dorite, am pus un filtru unde am inclus doar Oradea și localitățile limitrofe, și totodată, pentru a avea forma această de linie, ce se pretează foarte bine pe analiza seriilor de timp și a trendurilor în timp, de la Marks vom alege opțiunea de Line.

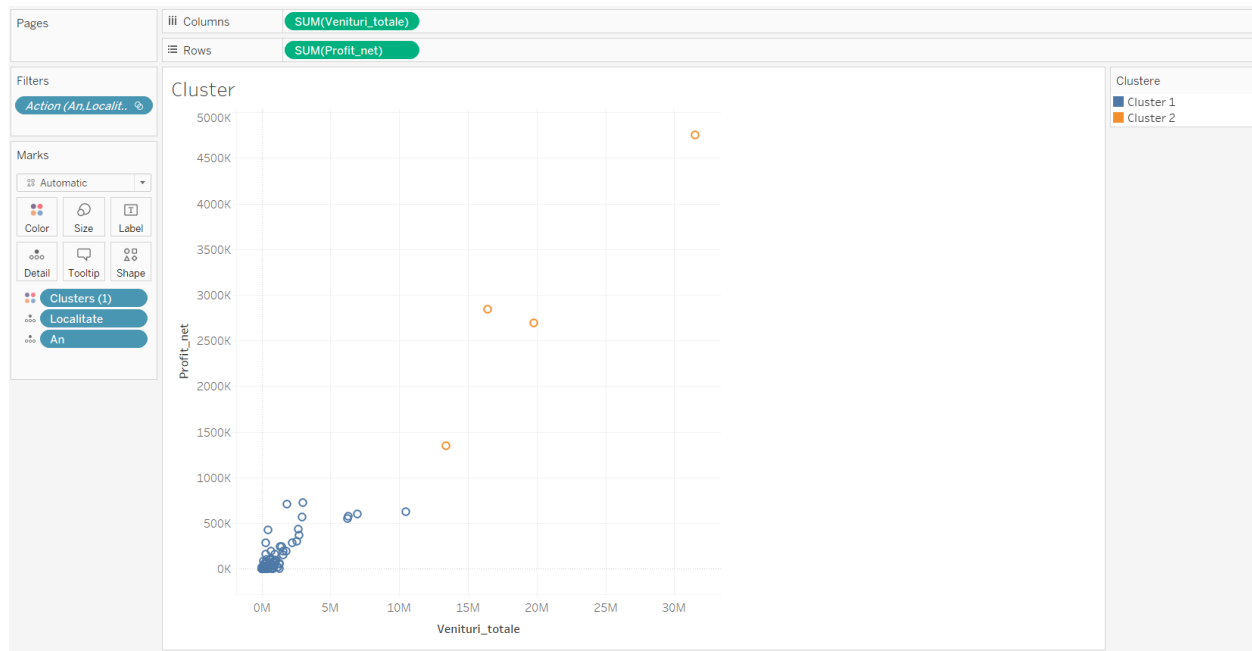


Harta în funcție de media profitului net

Penultima vizualizare făcută este o hartă, care va conține 2 straturi(județ și localitățile). Pe rânduri și coloane am introdus coordonatele geografice(latitudine și longitudine) în mod manual pentru fiecare localitate. Pentru a diferenția mărimea profitului în funcție de localitate, am plasat media profitului net în zona de Size. Stratul aferent județului a fost posibil ca introducere datorită setării ca tip geografic de State, astfel că Tableau va recunoaște automat coordonatele județului Bihor.



Cluster



Pentru a realiza analiza de cluster, am realizat un grafic în care mi-am ales ca număr 2 cluster, având în vedere că un cluster presupune un volum mare de activitate economică, chiar la nivel național sau global, și tocmai de aceea am considerat suficient un număr de 2 cluster. Tehnic, am pus pe coloane suma veniturilor totale, pe rânduri suma profitului net, iar în secțiunea Marks localitatea și anul la rubrica detalii. Din meniul de Analytics, am glisat opțiunea de Cluster, obținând ce se poate observa în imaginea de mai sus. Pentru a putea analiza această zonă, vom alege zona de describe clusters, unde vom avea 2 opțiuni.

Describe Clusters

SummaryModels

Inputs for Clustering

Variables:

Sum of Profit_net

Sum of Venituri_totale

Level of Detail:

An, Localitate

Scaling:

Normalized

Summary Diagnostics

Number of Clusters:

2

Number of Points:

148

Between-group Sum of Squares:

2.9043

Within-group Sum of Squares:

0.85907

Total Sum of Squares:

3.7633

Centers

Clusters	Number of Items	Sum of Profit_net	Sum of Venituri_totale
Cluster 1	144	68517.0	6.1823e+05
Cluster 2	4	2.9115e+06	2.0284e+07
Not Clustered	0		

Describe Clusters

SummaryModels

Analysis of Variance:

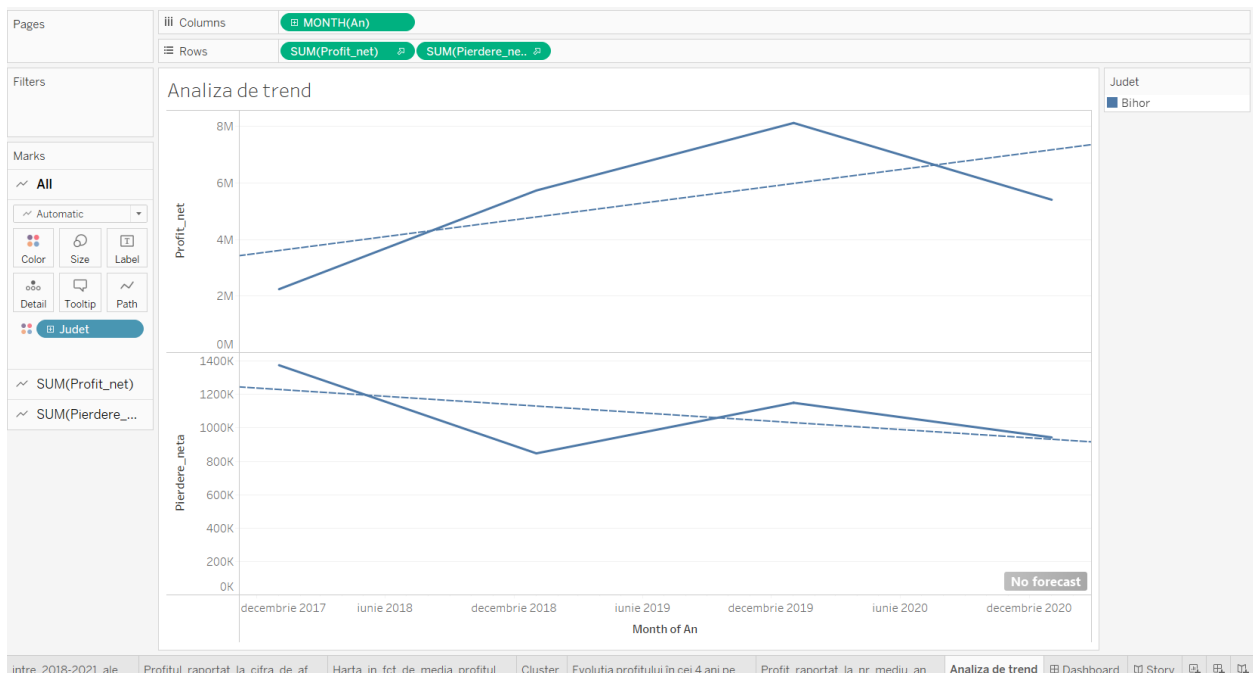
Variable	F-statistic	p-value	Model		Error	
			Sum of Squares	DF	Sum of Squares	DF
Sum of Profit_net	113.2	0.0	1.391	1	1.795	146
Sum of Venituri_totale	112.2	0.0	1.513	1	1.969	146

Important de menționat că dacă scoatem anul din analiza noastră, valoarea lui p-value va lua o valoare apropiată de 0, în prezent fiind de 0.

Analiza de trend

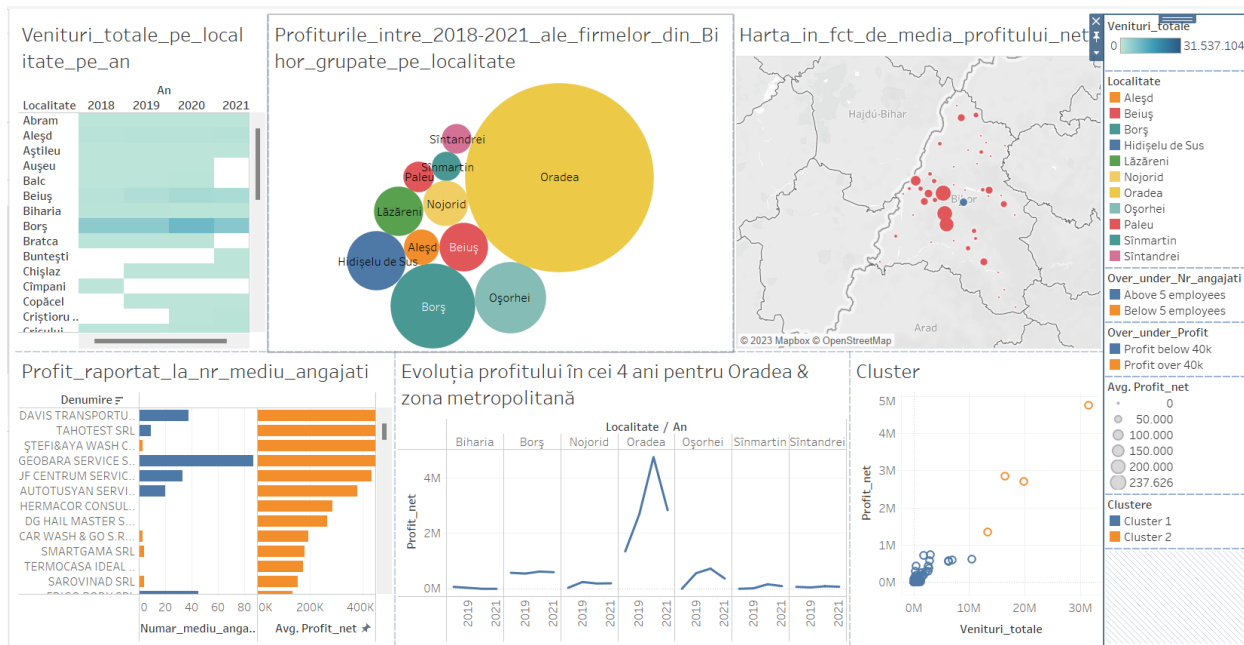
Pentru a realiza analiza de trend la nivelul județului Bihor, am decis să urmăresc trendul profitului net și a pierderii nete în paralel. Astfel, am transformat variabila An în tip dată calendaristică, pentru a putea vedea trendul raportat la luni. Pe rânduri, am introdus sumele profiturilor nete, respectiv a pierderilor nete, iar la nivelul culorilor am plasat județul asupra căruia fac analiza. Ulterior, cu click dreapta, trend lines, și opțiunea show trend lines, am putut să vizualizez trendul sumelor pe care le-am introdus.

Valorile statistice pentru R^2 și P-value sunt pe rând 0,4 și 0,3, respectiv 0,36 și 0,45, ceea ce reprezintă faptul că, din punct de vedere statistic, modelul nu este extrem de relevant, dar nici de neglijat, pe rând 40% sau 30% din varianța profitului putând fi explicat prin analiza variabilei dependente, și anume timpul.



Crearea de Dashboard

Ca aceste vizualizări să aibă un sens, le vom pune împreună în cadrul unui dashboard, astfel încât să poată fi vizibile laolaltă.

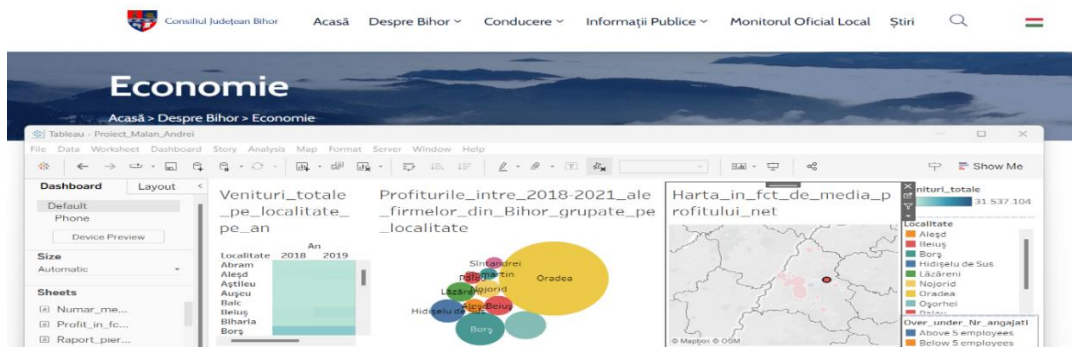
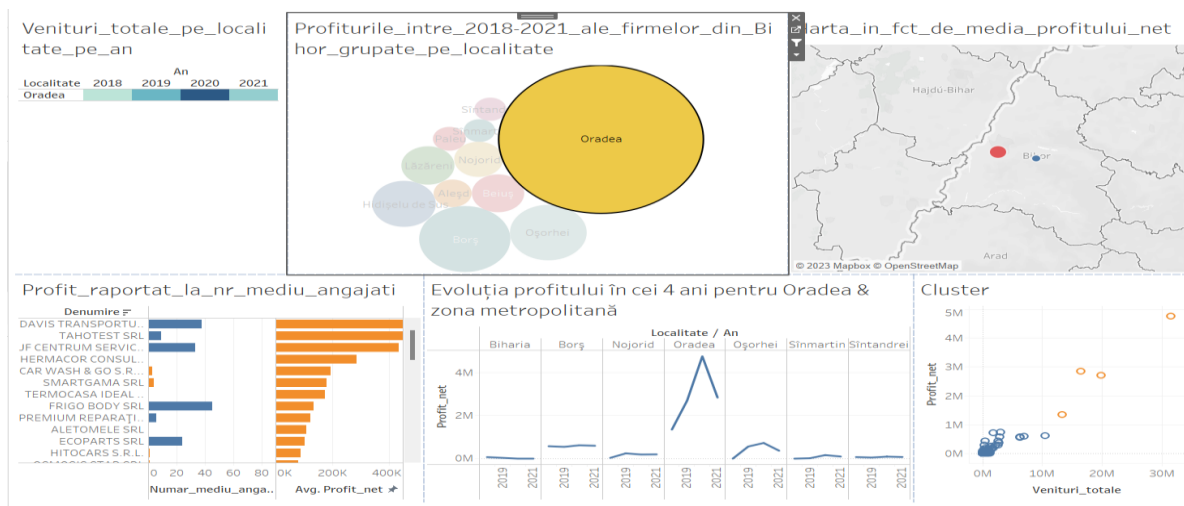
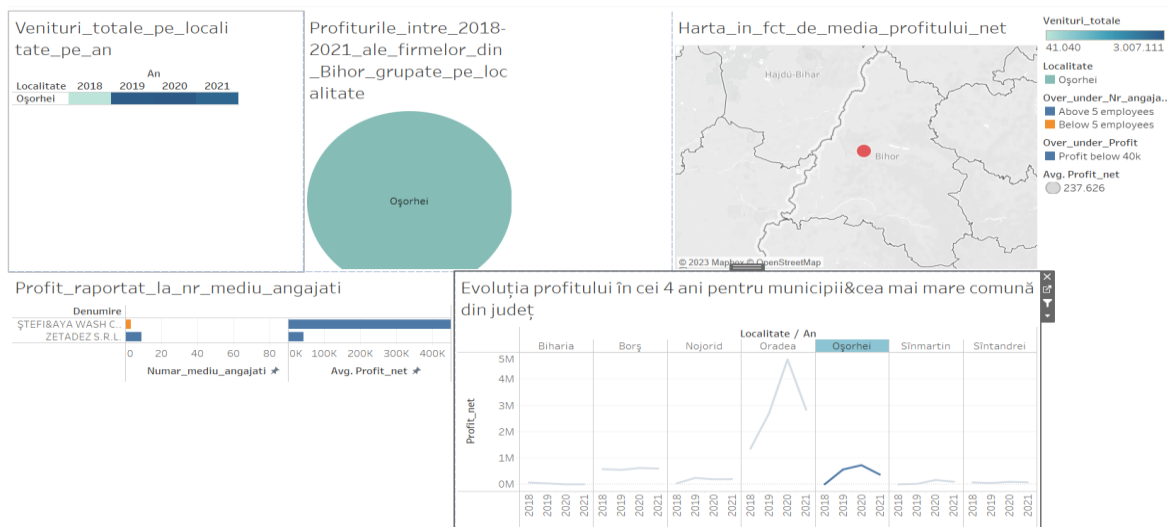


Interactivitate

Există 3 metode prin care se asigură interactivitatea în cadrul proiectului:

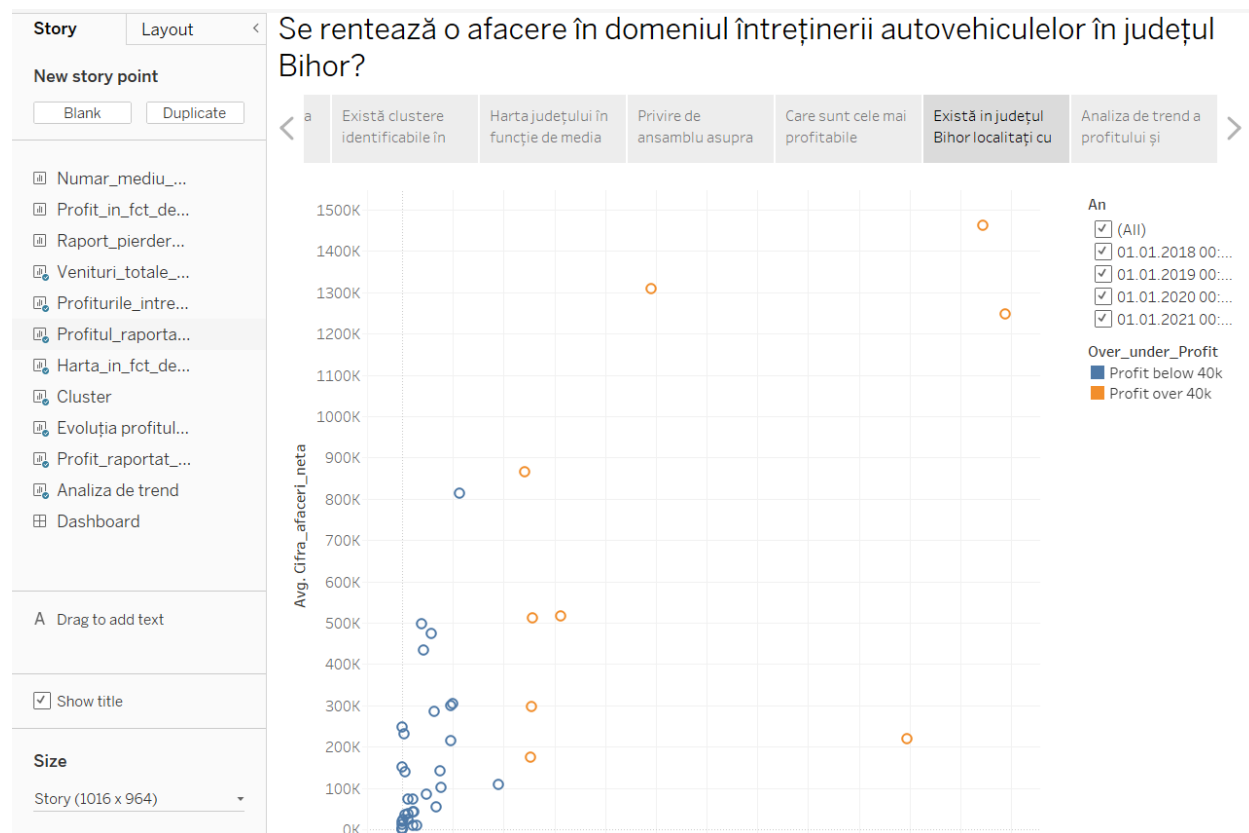
- Filtru 1, numit Interactivitate, ce asigură interactivitatea dashboard-ului la apăsarea pe un cerc din graficul Profiturile între 2018-2021 ale firmelor din Bihor grupate pe localitate.
- Filtru 2, numit Interactivitate1, ce asigură interactivitatea dashboard-ului la apăsarea pe o coloană din graficul Evoluția profitului în cei 4 ani pentru Oradea și zona metropolitană.
- Adresă URL, numit Web, ce asigură interactivitatea dashboard-ului prin apăsarea pe o localitate de pe hartă, iar acest lucru determină apariția unei pagini web în care se găsesc informații economice de la nivelul județului Bihor.

Dacă ar fi să acționăm unul dintre filtre, vom avea următoarele figuri:



Crearea poveștii

Această parte se realizează din partea de story a Tableau Desktop. Ca urmare a graficelor făcute, putem să creem un story relevant, și care să summarizeze efortul nostru de business: Se rentează o afacere în domeniul întreținerii autovehiculelor în județul Bihor?



Observăm că este posibil să filtrăm dacă dorim anii, astfel încât povestea să poată fi extinsă și analizată în profunzime. Totodată, observăm că am inclus toate ilustrările din Dashboard în Story, astfel încât prezentarea rezultatelor să fie cât mai complexă și ușor de citit, acest fapt fiind datorat și opțiunii de a ne muta de la o vizualizare la alta, prin acel carusel de opțiuni vizibil sub întrebarea de business pe care ne-am pus-o.

În poveste, suplimentar față de Dashboard merită menționat faptul că am pus analiza de trend, motivul principal fiind de a nu încărca suplimentar pagina de dashboard, care trebuie să fie clară și lizibilă.