Lucrări de construcții a clădirilor rezidențiale și nerezidențiale (România)

1. Tematică și descrierea problemei de business

Tema aleasă pentru dezvoltarea acestui proiect este următoarea: Lucrări de construcții a clădirilor rezidențiale și nerezidențiale în România. Această tematică propusă de mine, se găsește sub codul CAEN 4120[[1]](#footnote-1), cod care face parte din clasa F – Construcții, subclasa 41 – Construcții de clădiri, sub-subclasa 412 – Lucrări de construcții a clădirilor rezidențiale și nerezidențiale.

Setul de date care ne va ajuta în realizarea acestui proiect, se găsește pe site-ul https://data.gov.ro/ , de unde vom descărca situațiile financiare aferente anilor 2016-2019. Seturile de date folosite pentru analiză sunt de forma:

* WEB\_BL\_BS\_SL\_AN20xx.txt – Societăți comerciale care depun bilanț lung, bilanț scurt și raportări anuale ,
* WEB\_IR\_20xx.txt – Raportări anuale ale societăților comerciale care depun bilanț IFRS,
* WEB\_UU\_AN20xx.txt – Societăți comerciale care depun bilanț format prescurtat.

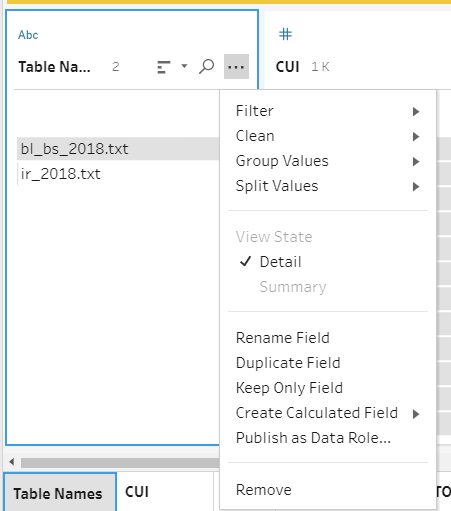
De pe același site menționat mai sus, vom lua date de la Oficiul Național al Registrului Comerțului, unde se găsesc informații cu privire la firmele alese. Setul de date ales se numește Firme înregistrate la Registrul Comerțului până la data de 20 noiembrie 2019, neradiate, cu sediu. Acesta se regăsește împărțit în 4 fișiere în format .csv.

După descărcarea datelor, pentru o mai bună înțelegere și pentru a ne ușura munca, consultând fișierele în format .csv aferente fiecărei categorii alese pentru fiecare an, vom pregăti datele cu ajutorul Tableau Prep Builder.

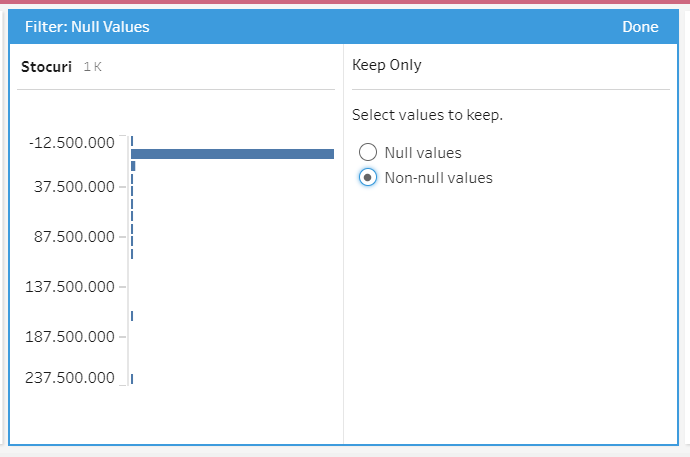
1. Pregătirea datelor

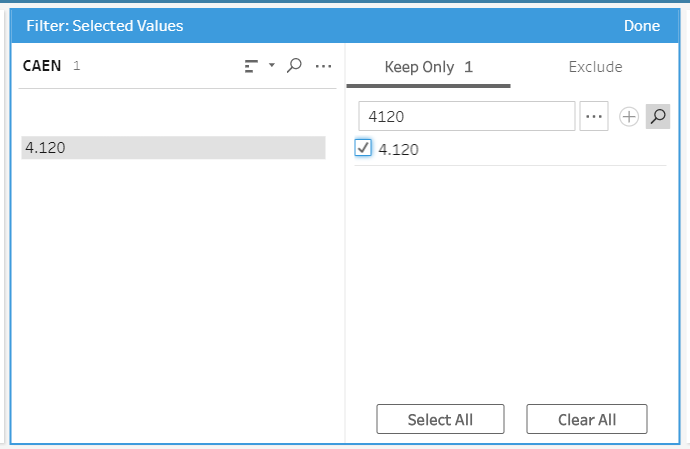
Primul pas pentru analiza propusă constă în pregătirea datelor în Tableau Prep. Vom conecta datele cu ajutorul butonului „Connect to data”. Trebuie aduse atât datele din fiecare categorie, pentru fiecare an, cât și datele de la Registrul Comerțului.

În primul pas de curățare fiecare câmp va fi redenumit conform documentelor .csv și eliminăm numele de tabele.



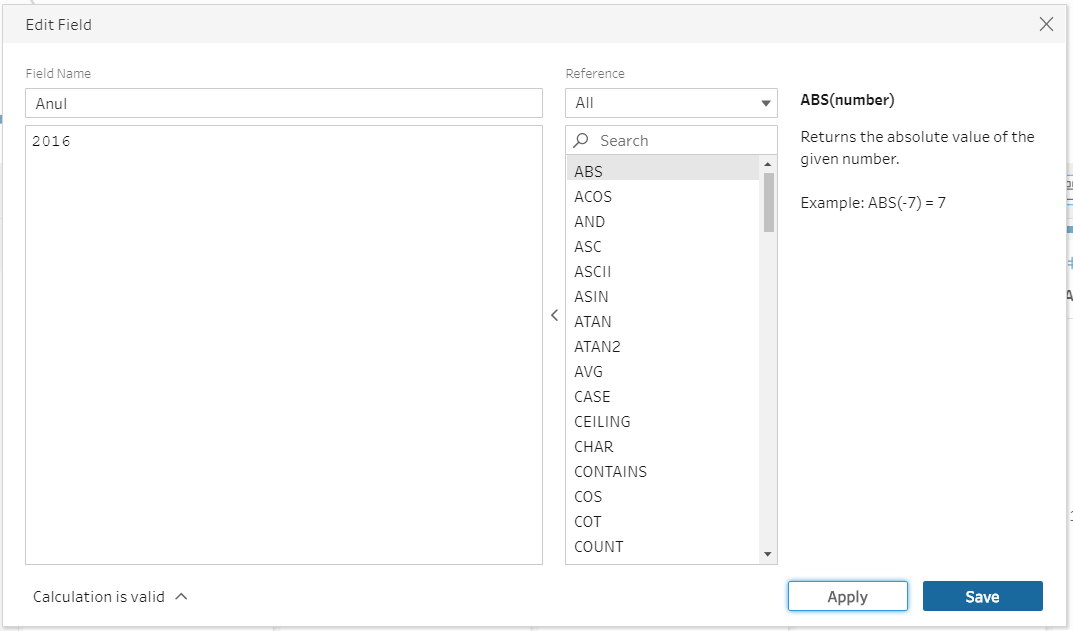
După redenumirea câmpurilor urmează pasul de Union pentru fiecare an. În acest fel vom avea datele corespunzătoare pentru fiecare an, urmând apoi un pas de curățare unde setăm un filtru pe câmpul „CAEN”, alegând CAEN-ul pentru care am decis să facem analiza: 4120 și eliminăm valorile nule pentru a avea o analiză cât mai precisă.





Din punct de vedere al tipului, datele sunt atât valorice, cât și numerice. Pentru fiecare câmp vom seta tipul de date potrivit.

Tot la acest pas de curățare, vom adăuga un câmp nou pentru fiecare Union, denumit „Anul”, de tipul date și vom seta valoarea corespunzătoare pentru fiecare set de date. Deoarece suntem constrânși să alegem anul sub forma DD-MM-YYY, vom seta pentru fiecare an, data de 31 Decembrie, întrucât pe noi ne interesează doar anul, lucru ce nu ne afectează în vreo formă analiza pe care urmează să o facem.



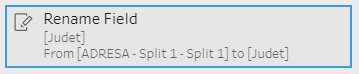
Vom elimina câmpul „Patrimoniul regiei” deoarece nu este relevant pentru analiza noastră.

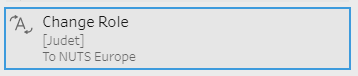
După acest pas de curățare, unim datele printr-un „Union” pentru toți cei patru ani, apoi printr-un pas de curățare eliminăm numele tabelelor componente. Acum avem datele necesare pentru codul CAEN ales de noi, pentru perioada 2016-2019, curățate.

Datele provenite de la Registrul Comerțului vor fi unite printr-un Union, iar mai apoi curățate printr-un pas de Clean: ștergem câmpul cu numele tabelelor, eliminăm valorile nule, schimbăm tipurile de date în funcție de datele componente pentru fiecare câmp.

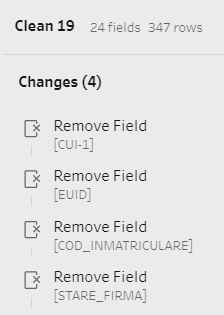
Avem nevoie de un câmp „Județ” pe care îl vom creea cu ajutorul câmpului „Adresă”. În acest câmp „Judet” se vor regăsi toate județele din România. Îi vom schimba tipul pentru a fi recunoscut de tableau ca fiind regiune, lucru ce ne va ajuta în partea de Tableau Desktop pentru crearea de hărți.



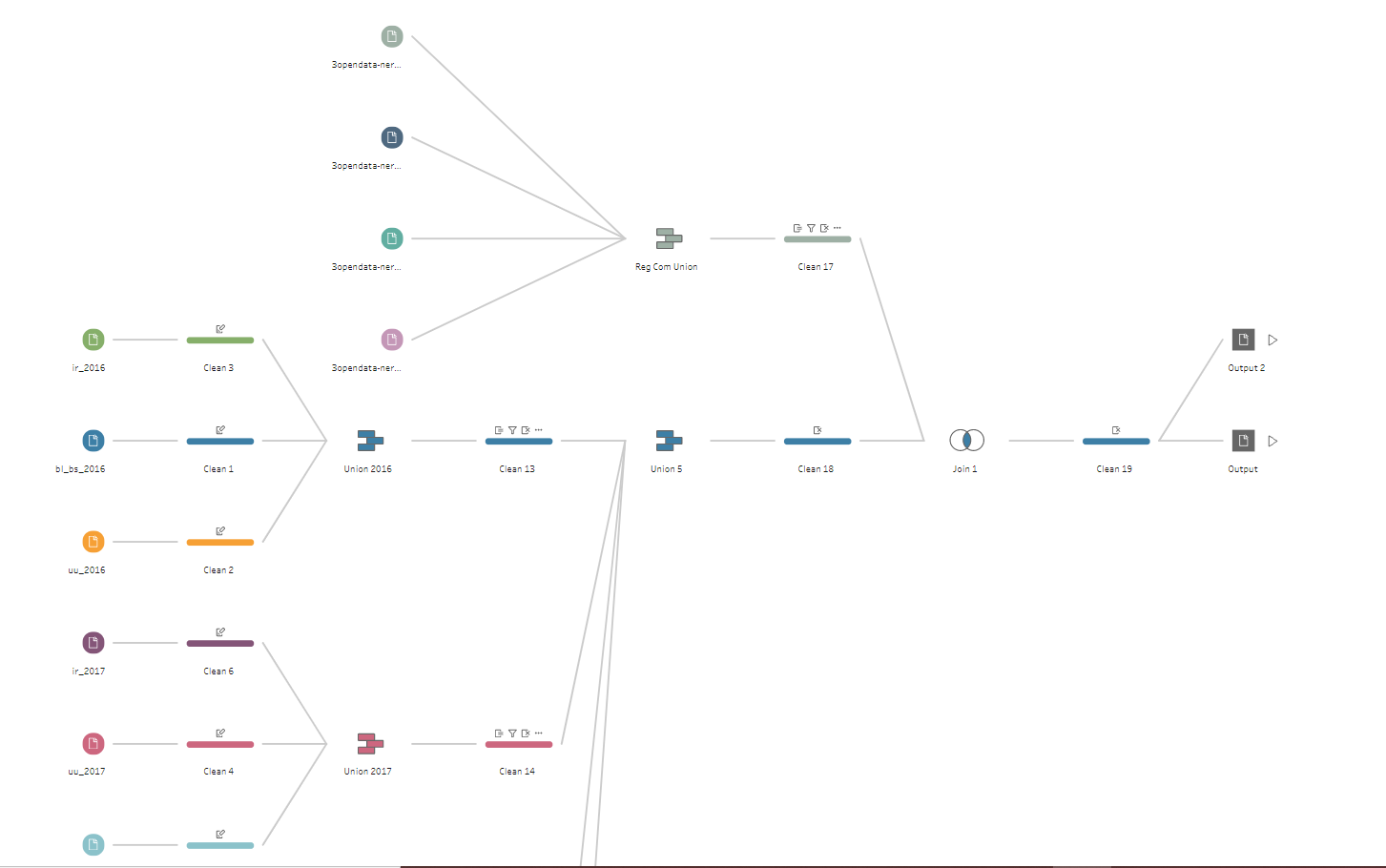


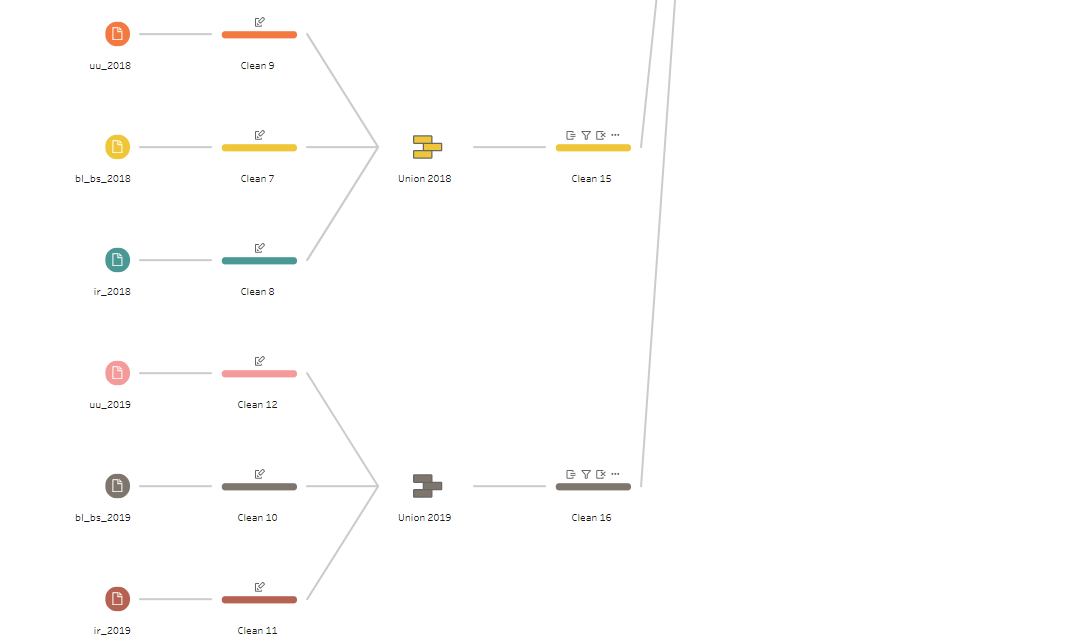


Acum că avem și datele de la Registrul Comerțului curățate, vom face un pas de Join după CUI (câmpul comun) între acestea și datele pentru toți anii, urmând să le curățăm. Curățarea de la acest pas constă în ștergerea câmpurilor care nu sunt relevante pentru analiza noastră: EUID, COD\_INMATRICULARE, STARE\_FIRMA și duplicatul de la CUI.



După tot acest proces putem să formăm output-ul.





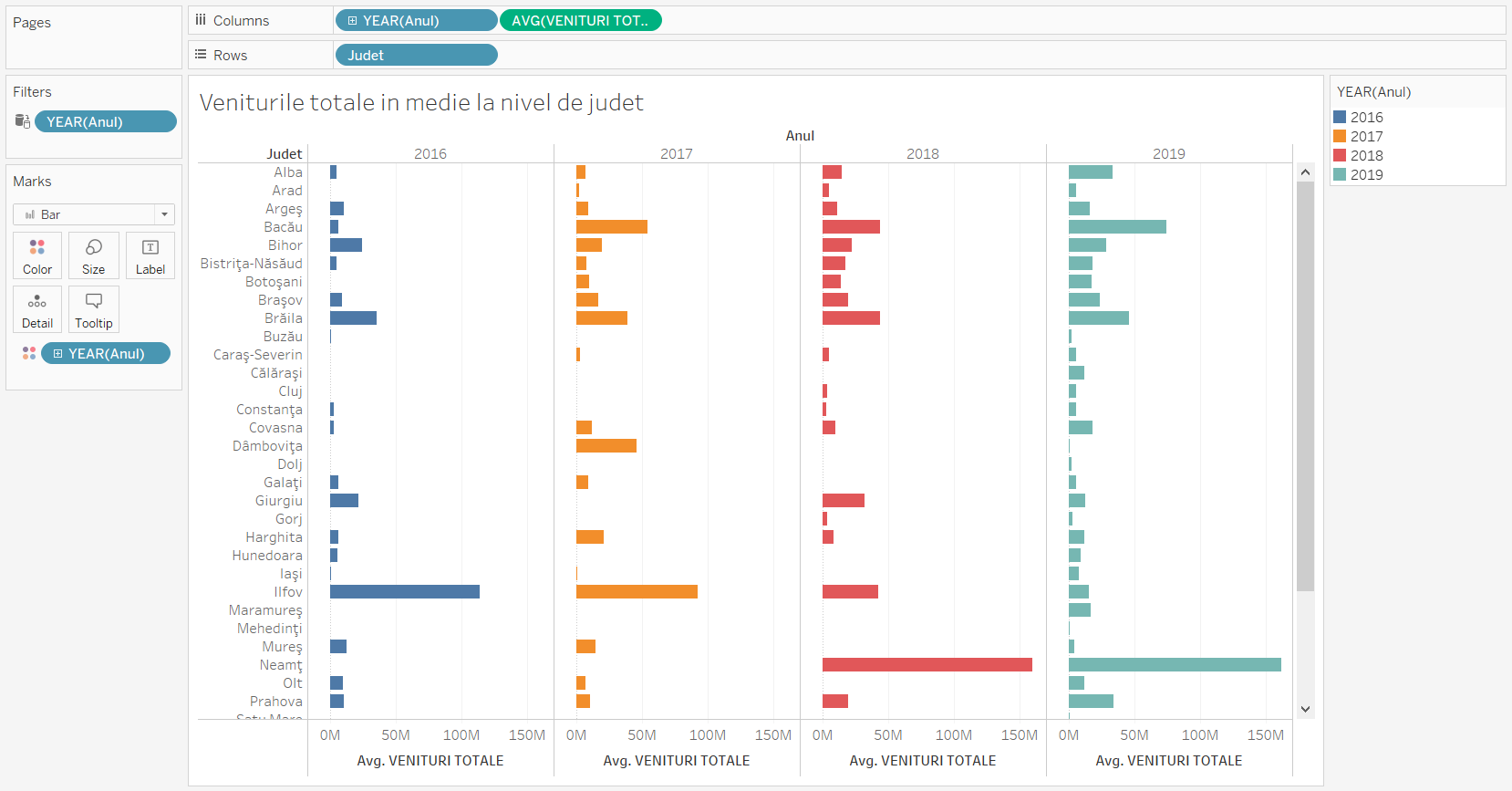
1. Data Visualization & Analysis

Veniturile și cheltuielile totale pentru fiecare an

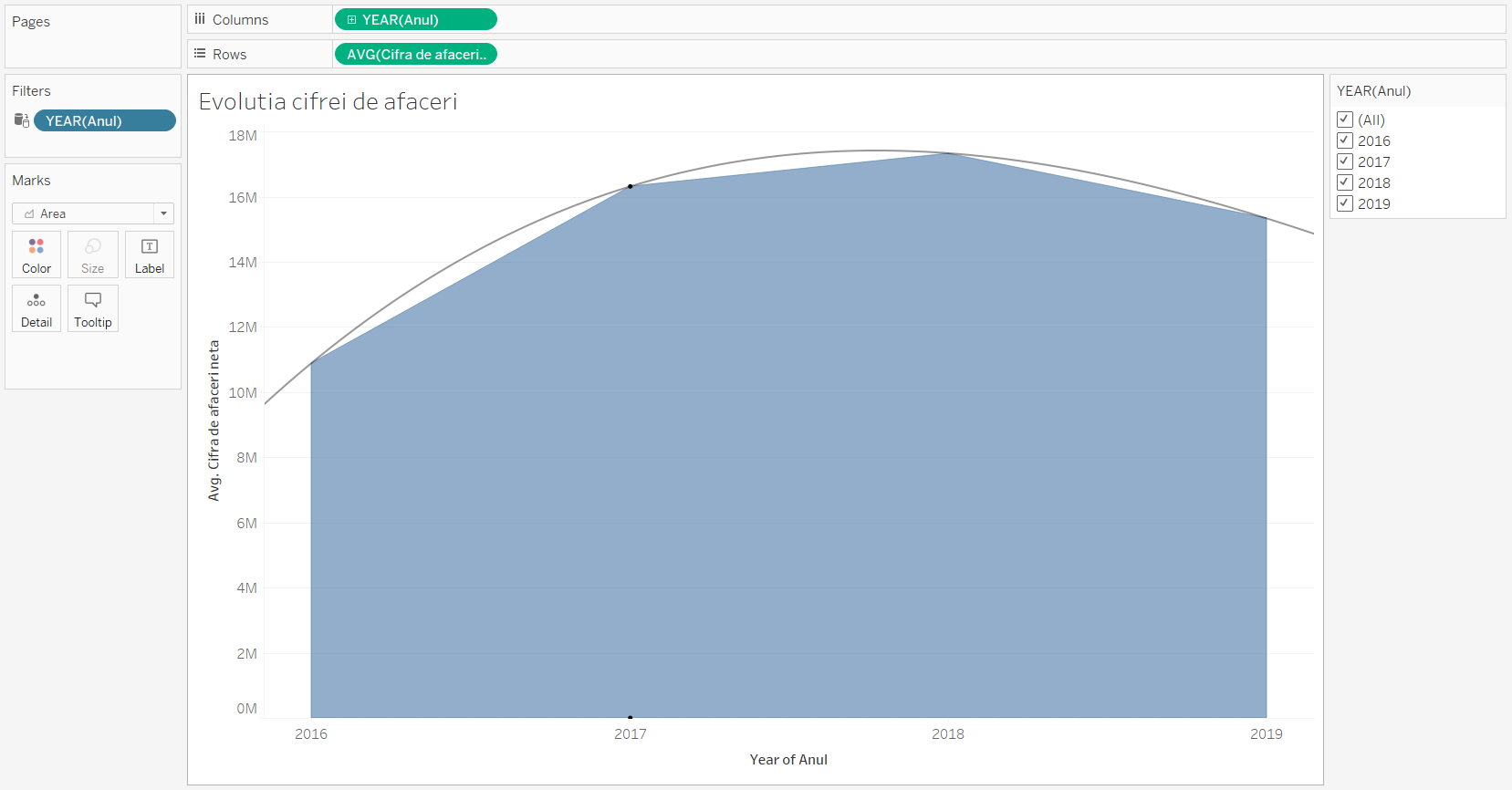


Se poate observa faptul că, la nivel de România, în perioada 2016-2018, veniturile, dar și cheltuielile cresc, în timp ce în 2019 scad. În anul 2018 cheltuielile totale depășesc pragul veniturilor totale, în timp ce în restul anilor veniturile se mențin mai mari decât cheltuielile.

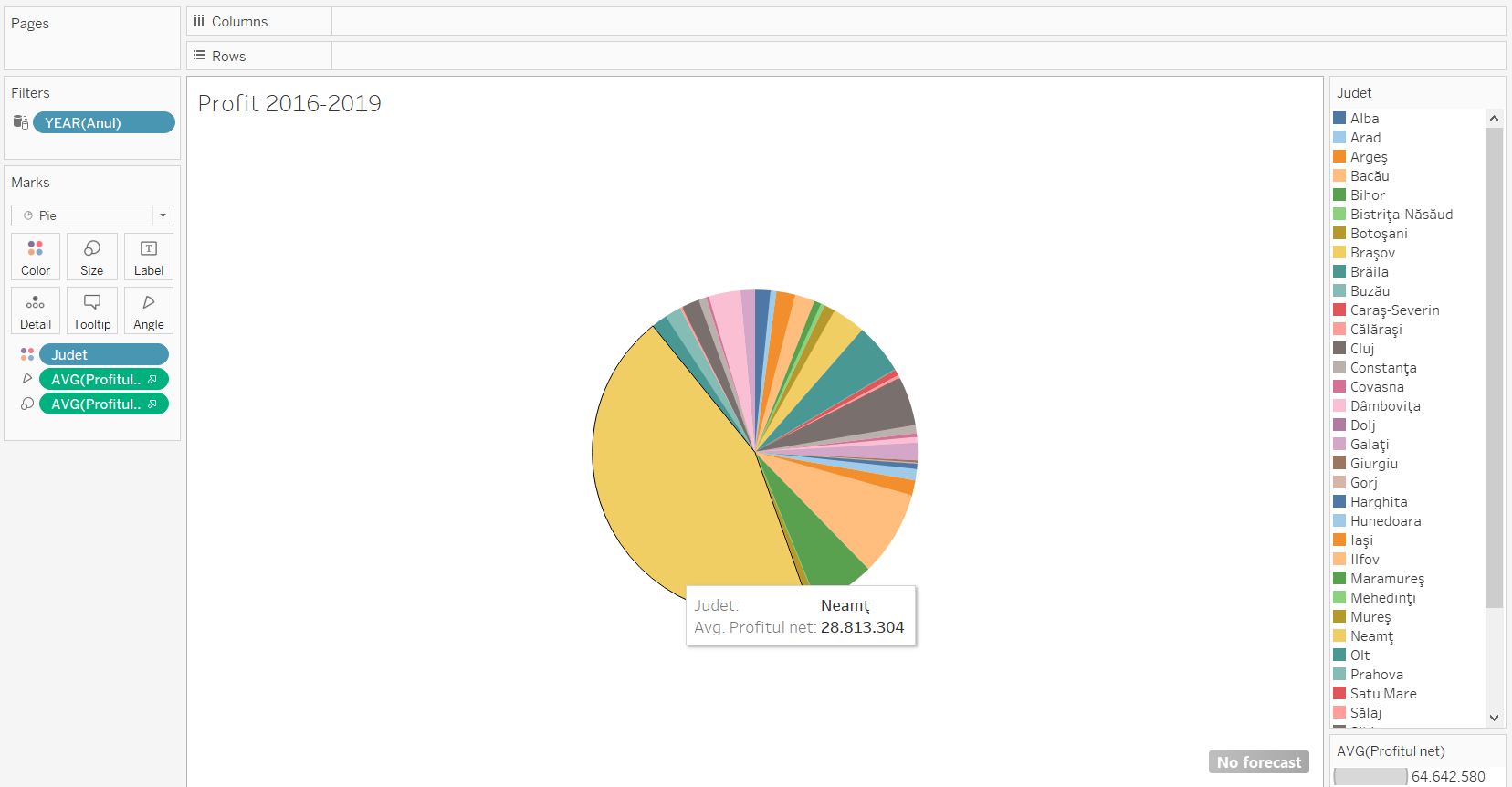
Venituri totale în medie la nivel de județ pentru fiecare an



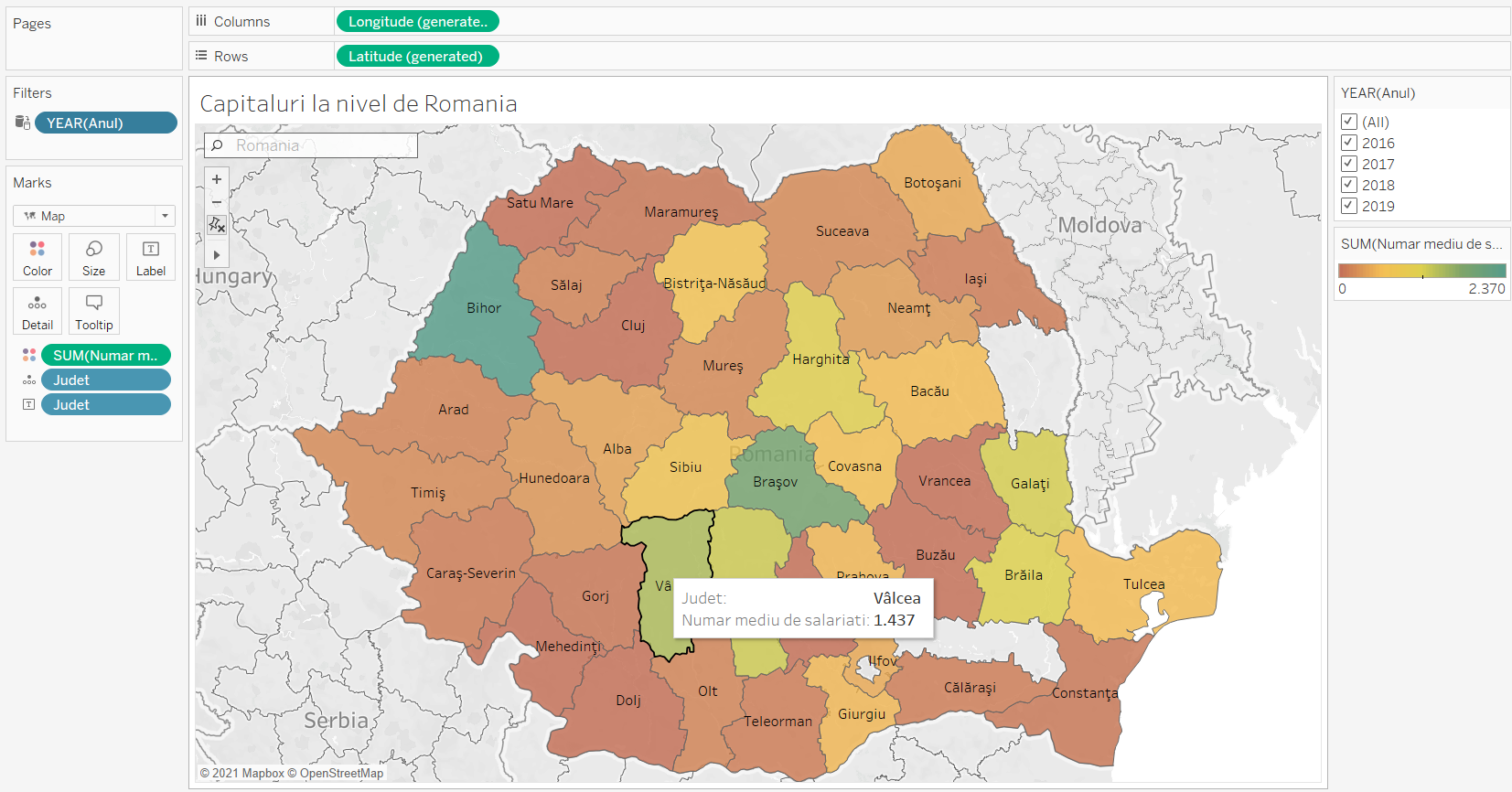
Evoluția cifrei de afaceri în timp la nivel de România



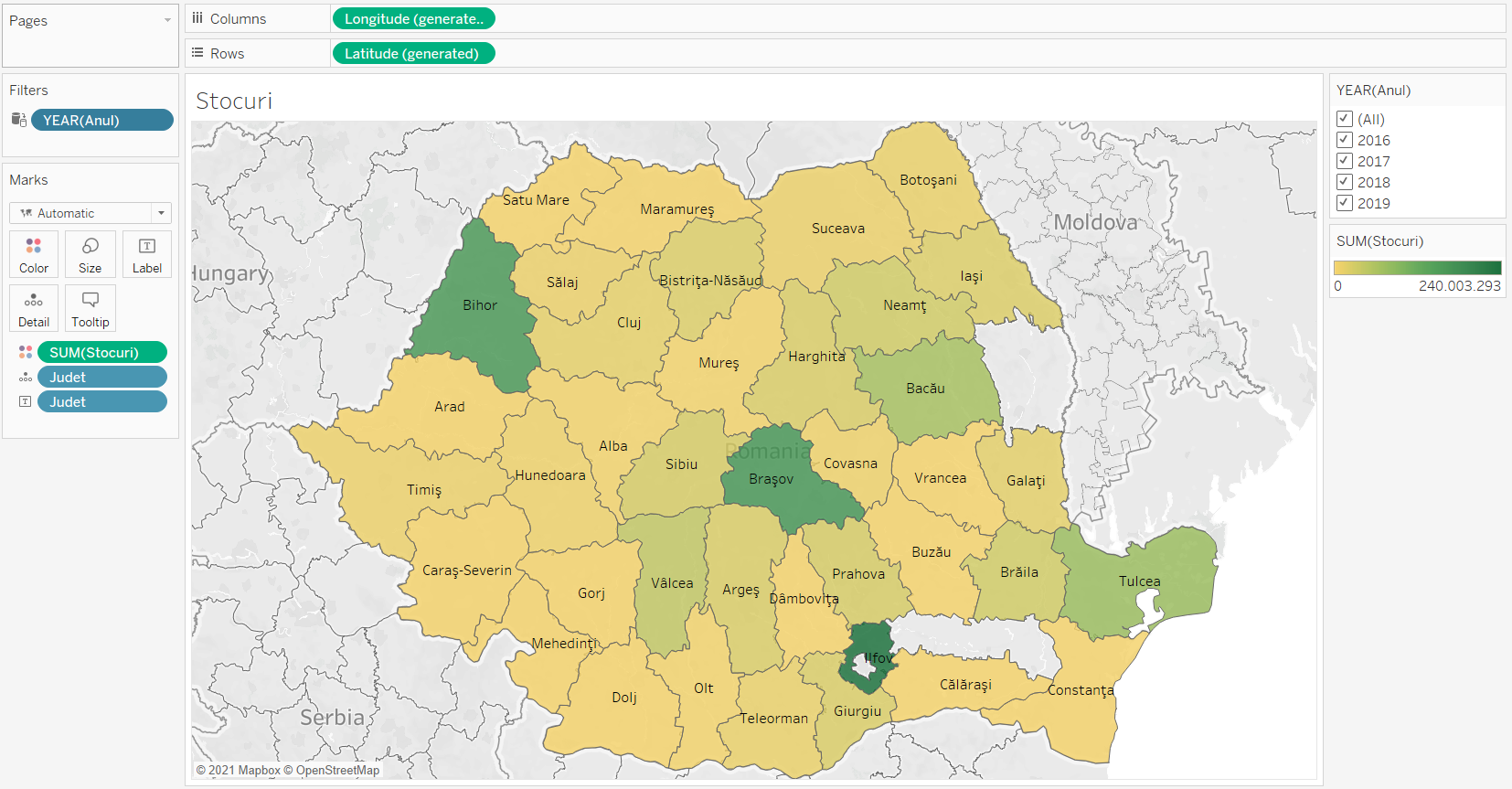
Profitul net la nivel de județ pe perioada 2016-2019



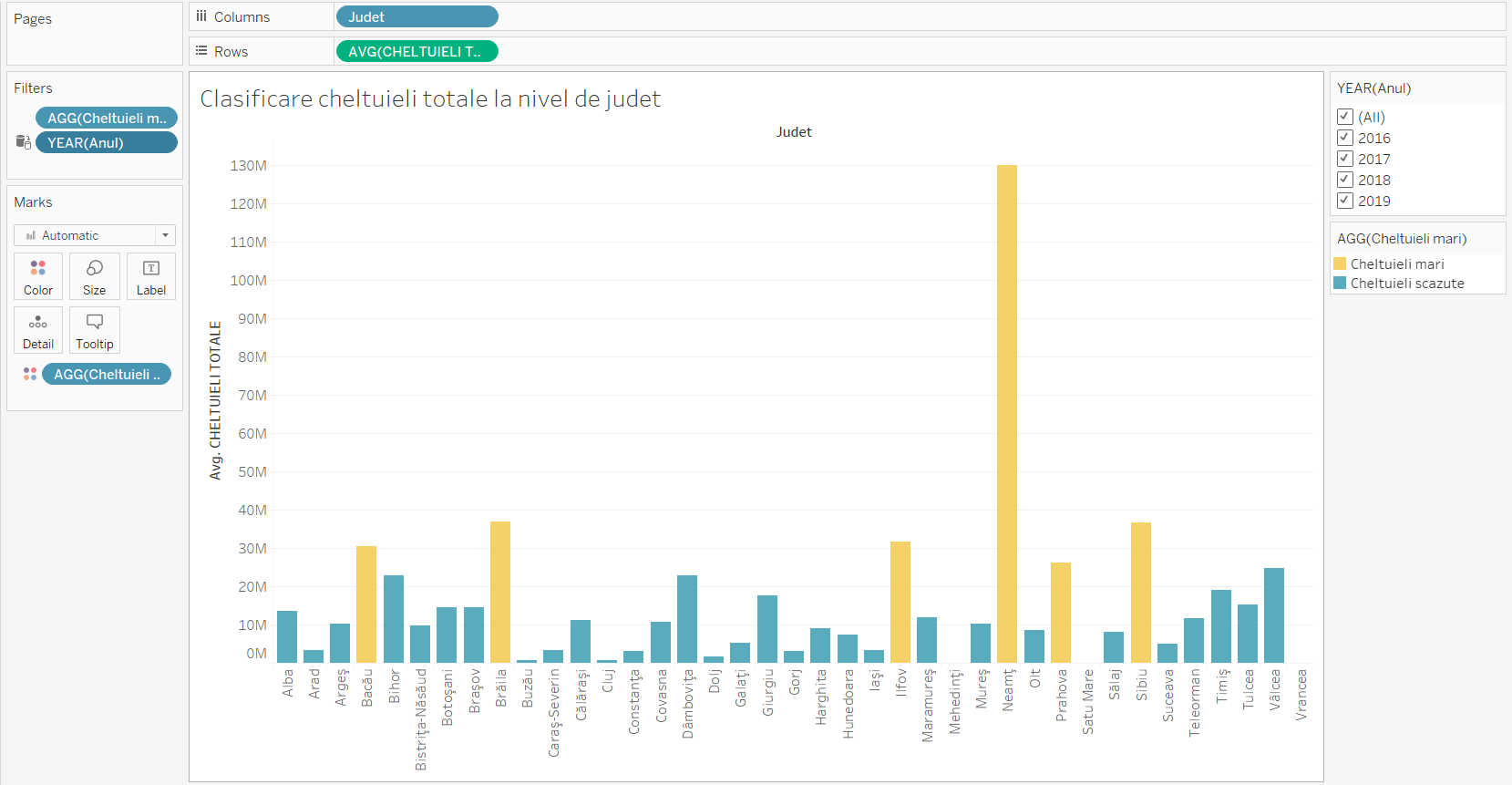
Numărul mediu de salariați în România pentru 2016-2019



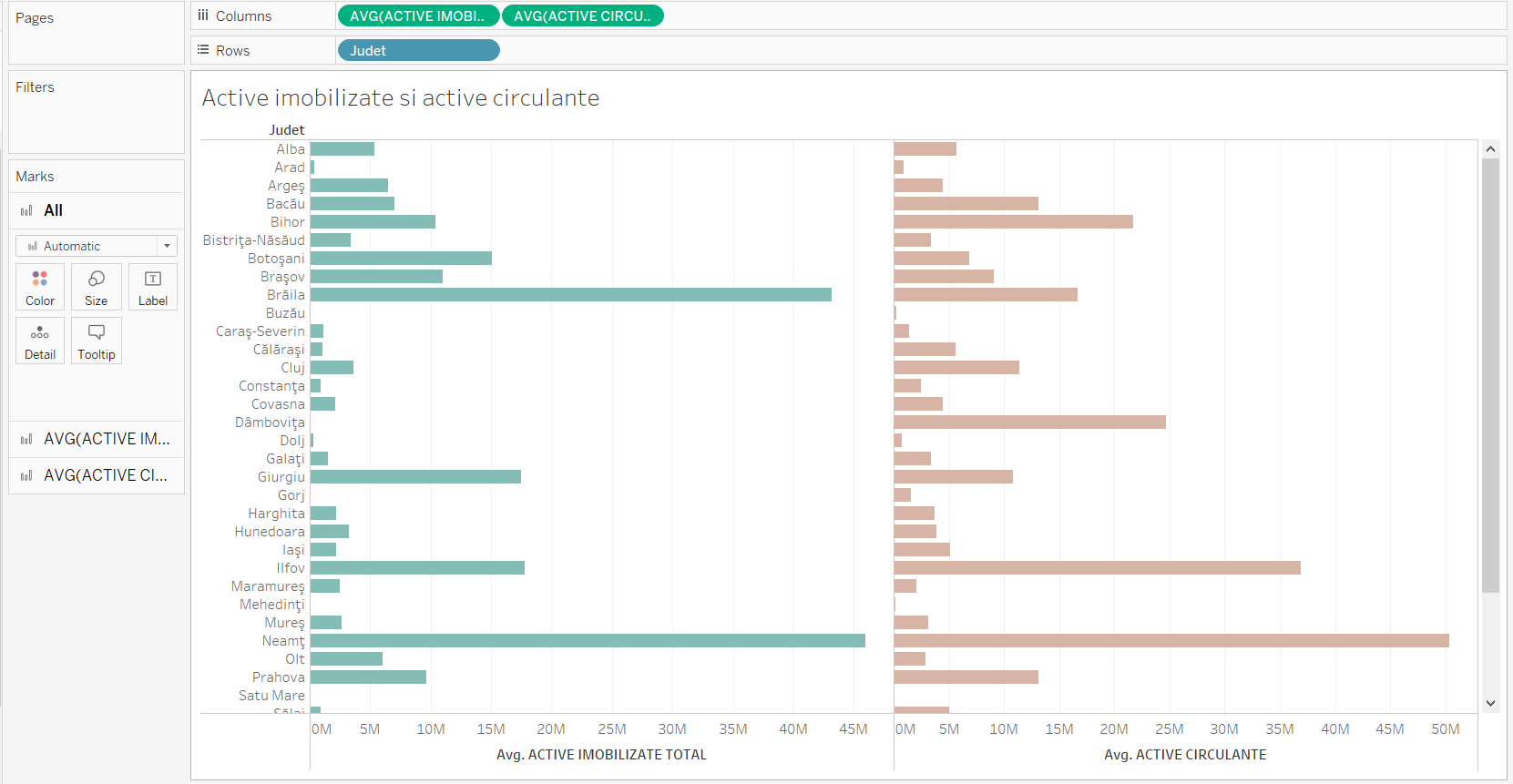
Stocurile pentru fiecare județ raportate la perioada 2016-2019



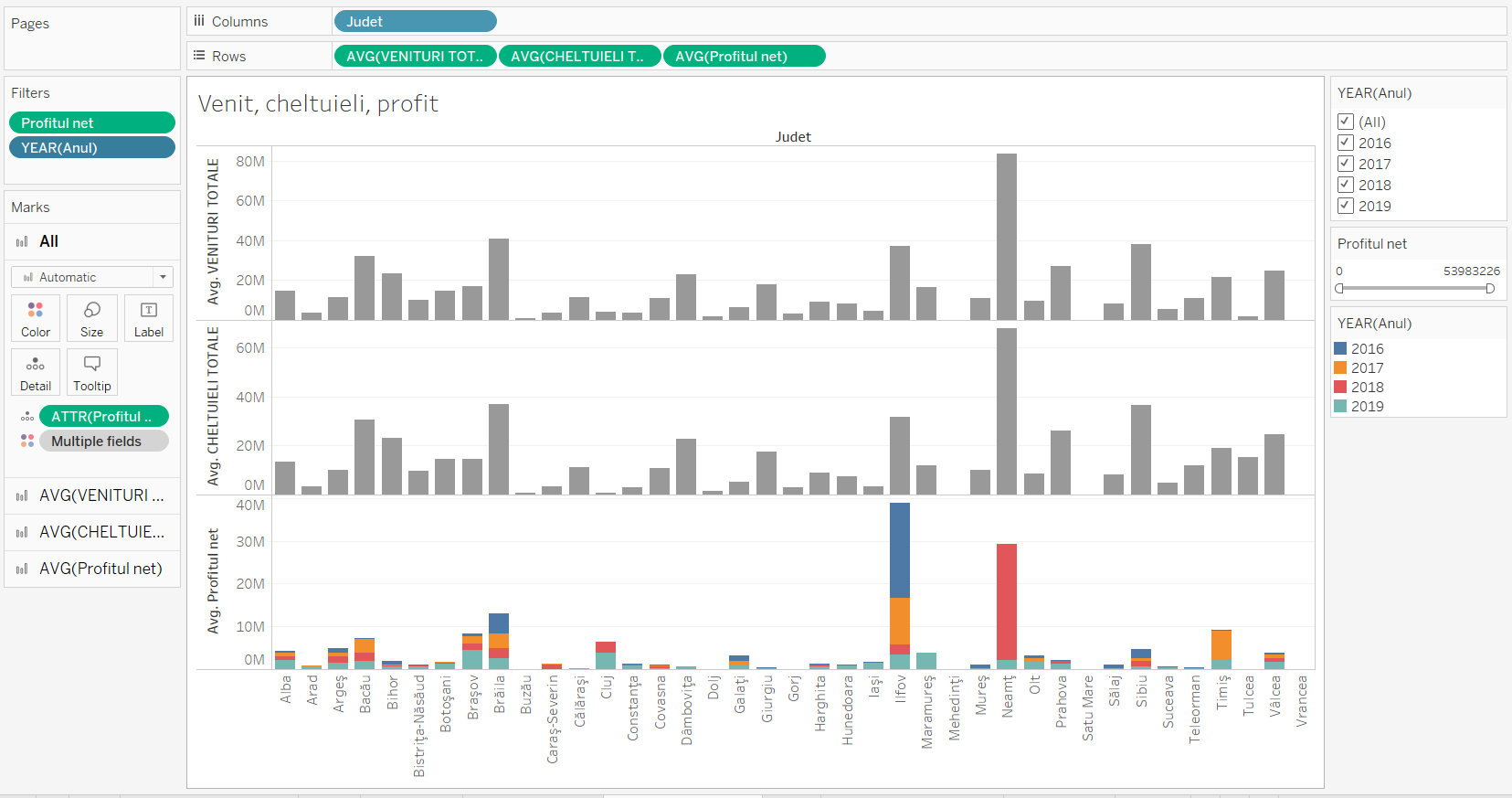
Clasificare cheltuieli totale la nivel de județ, pentru fiecare an, în cheltuieli mari și cheltuieli scăzute



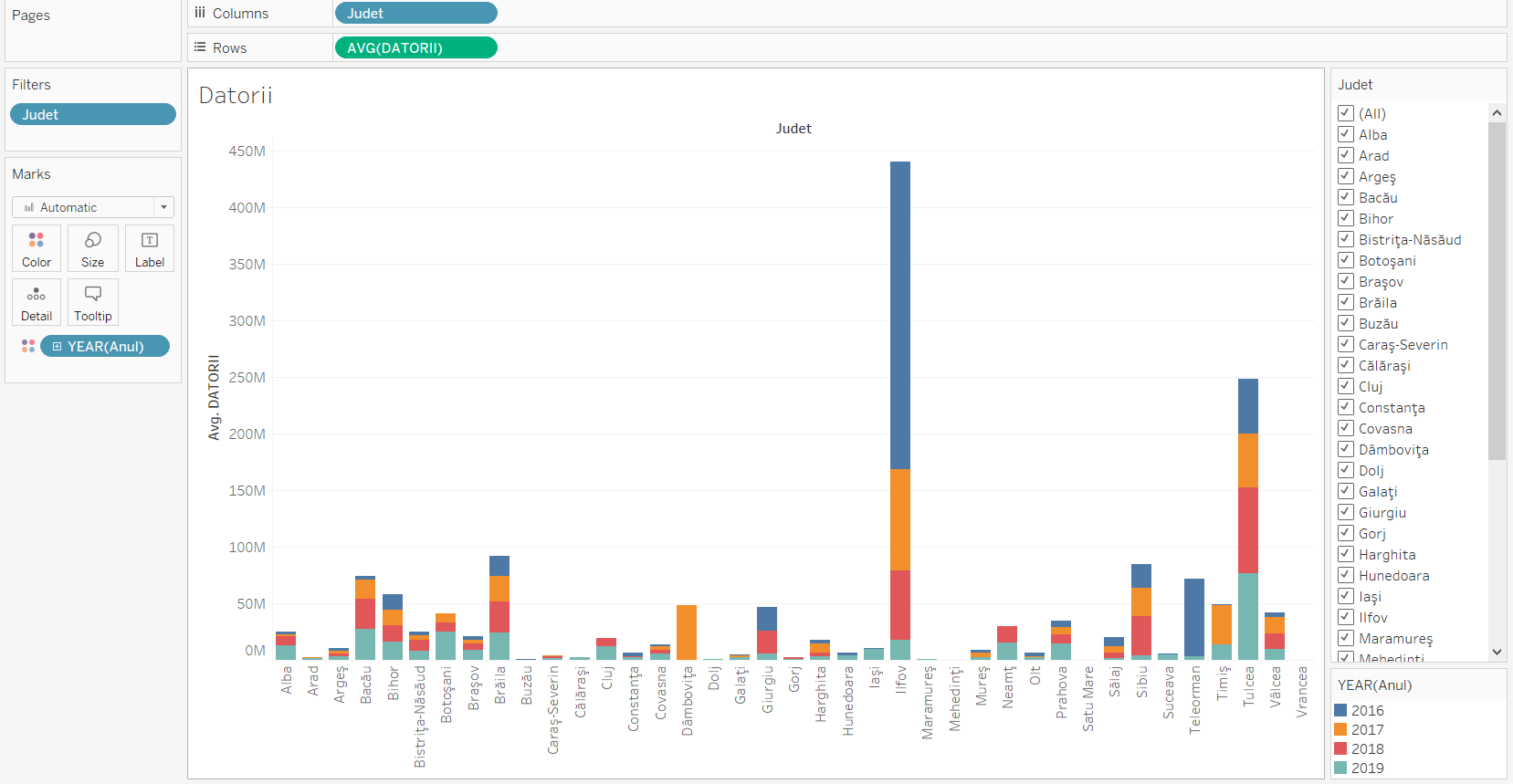
Active imobilizate și active circulante la nivel de județ



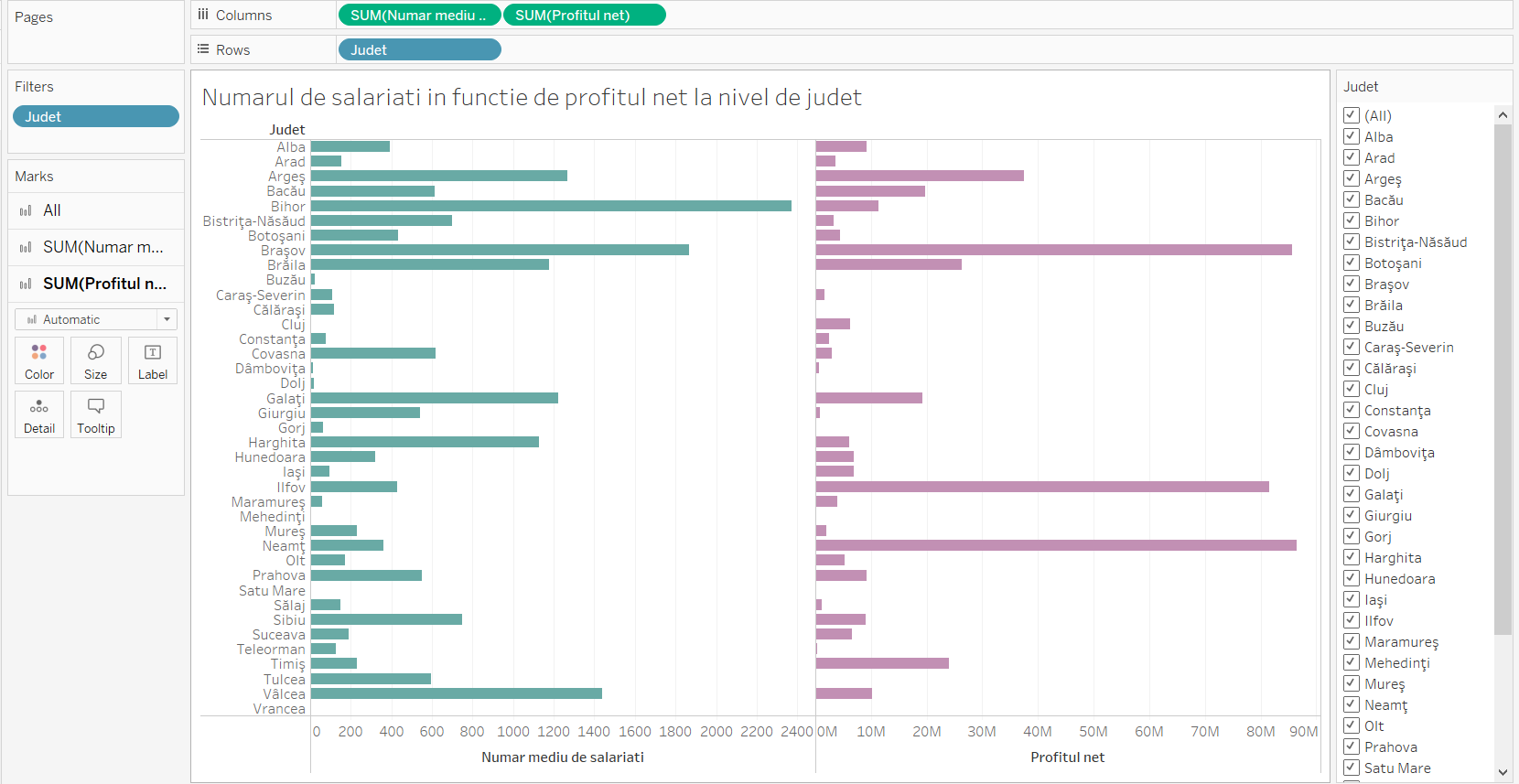
Venituri, cheltuieli și profit pe județe, raportate pentru fiecare perioadă



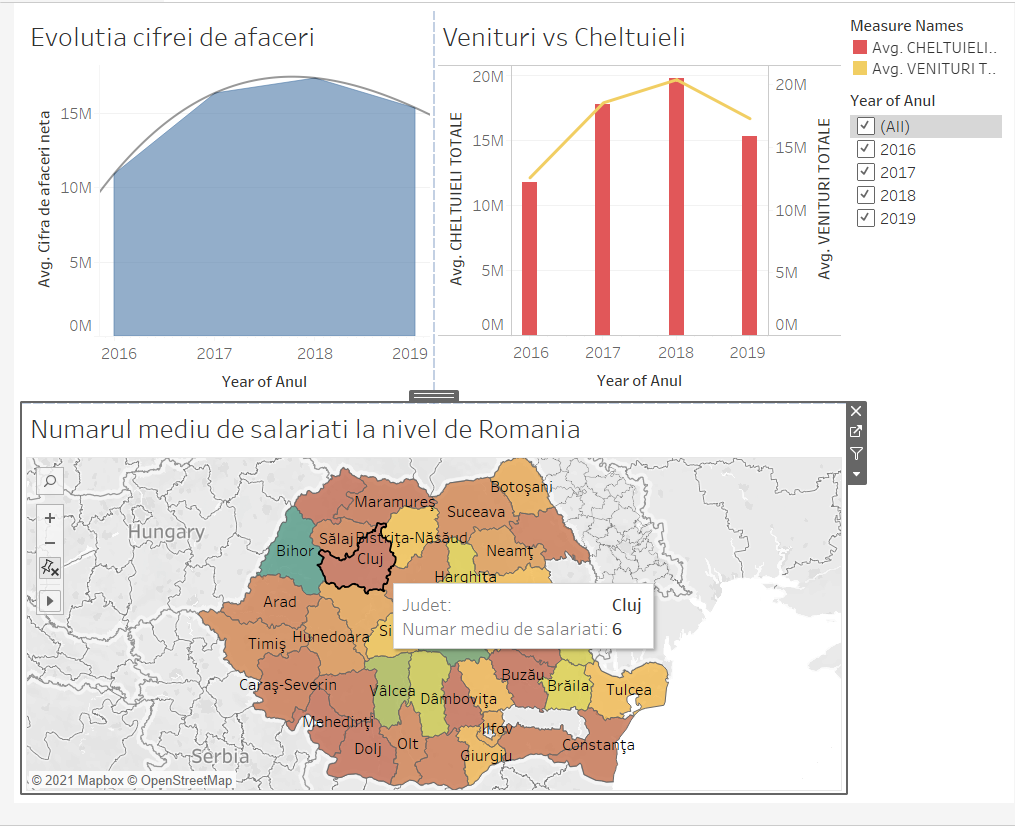
Datoriile raportate la nivel de județ pentru fiecare an



Legătura dintre numărul de salariați și profitul net la nivel de județ



Dashboard: privire de ansamblu asupra legăturii dintre numărul mediu de salariați, cifra de afaceri, veniturile și cheltuielile raportate pentru fiecare an



Ultima parte, cea de story, va prezenta, într-o manieră cât mai prietenoasă cu auditoriu, câteva aspecte considerate esențiale pentru această arie de interes studiată în cadrul proiectului.

1. <https://caen.ro/> [↑](#footnote-ref-1)