

Business Intelligence

Analiza hotelurilor si altor facilități de cazare
similare din municipiul Alba

COD CAEN: 5510

Studenți: Tătar Gabriela, Tămășan Raluca, Vlad Irina

Facultatea de Științe Economice și Gestiunea Afacerilor

Informatică Economică

Anul III, promoția 2020-2023

1. Descrierea problemei de business pentru care s-a realizat proiectul

În acest proiect am realizat *un studiu asupra firmelor din municipiul Alba Iulia, care se concentrează pe hoteluri și alte facilități similare de cazare ca obiect de activitate principală*. Datele care aparțin firmelor neradiate cu un sediu vor fi filtrate pe baza câmpului „adresa”, acolo unde acest câmp conține secvența „Municipiul Alba Iulia”. În ceea ce privește domeniul de activitate, vom continua filtrarea prin intermediul câmpului „CAEN”. Având în vedere tema aleasă, codul CAEN în funcție de care vom realiza filtrarea va fi 5510 – Hoteluri și alte facilități de cazare similare.

Pentru a realiza un studiu complet, am realizat grafice pentru a analiza cifrele de afaceri a fiecărei firme, profitabilitatea și riscul de credit, precum și evoluția acestora pe parcursul celor 4 ani aleși (2017, 2018, 2019, 2020).

2. Pregătirea datelor

2.1. Conectarea la seturile de date

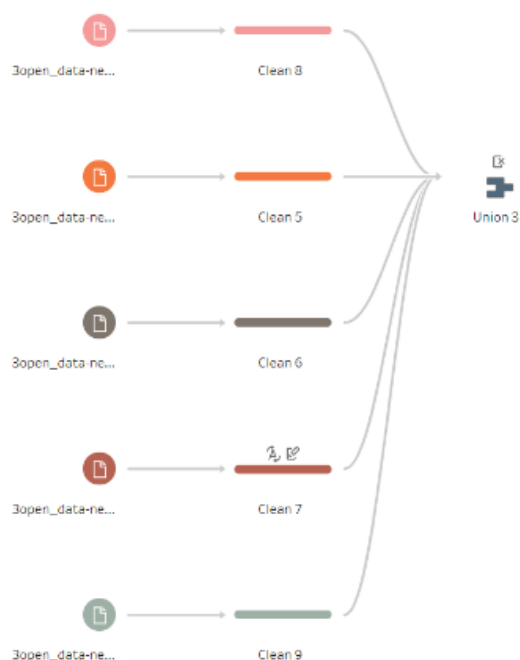
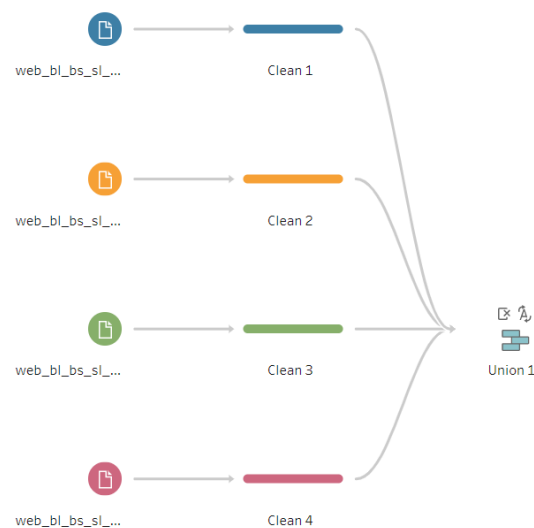
Seturile de date utilizate în cadrul acestui proiect au fost preluate de pe site-ul <https://data.gov.ro/dataset>:

- De la **Ministerul Finanțelor Publice** am preluat fișierele de tip **.txt** cu numele **WEB_BL_BS_SL_Anul**, unde „Anul” ia valorile 2017, 2018, 2019, 2020. Acestea reprezintă situațiile financiare din anii respectivi.
- De la **Oficiul Național al Registrului Comerțului** am preluat fișierele de tipul **.csv**, care conțin date despre firmele neradiate cu sediu înregistrate până la data de 08.12.2020. Acestea sunt în număr de 5, iar denumirea este de forma **3open_data-neradiate_cu_sediu-08.12.2020.001.csv**. În fișierele extrase, putem afla informații despre denumire, adresă, cod de înmatriculare etc.

Am încărcat atât fișierele de tip **.txt**, cât și cele de tip **.csv** în Tableau Prep Builder, pentru a prelucra datele conținute și a crea setul de date necesar analizei noastre.

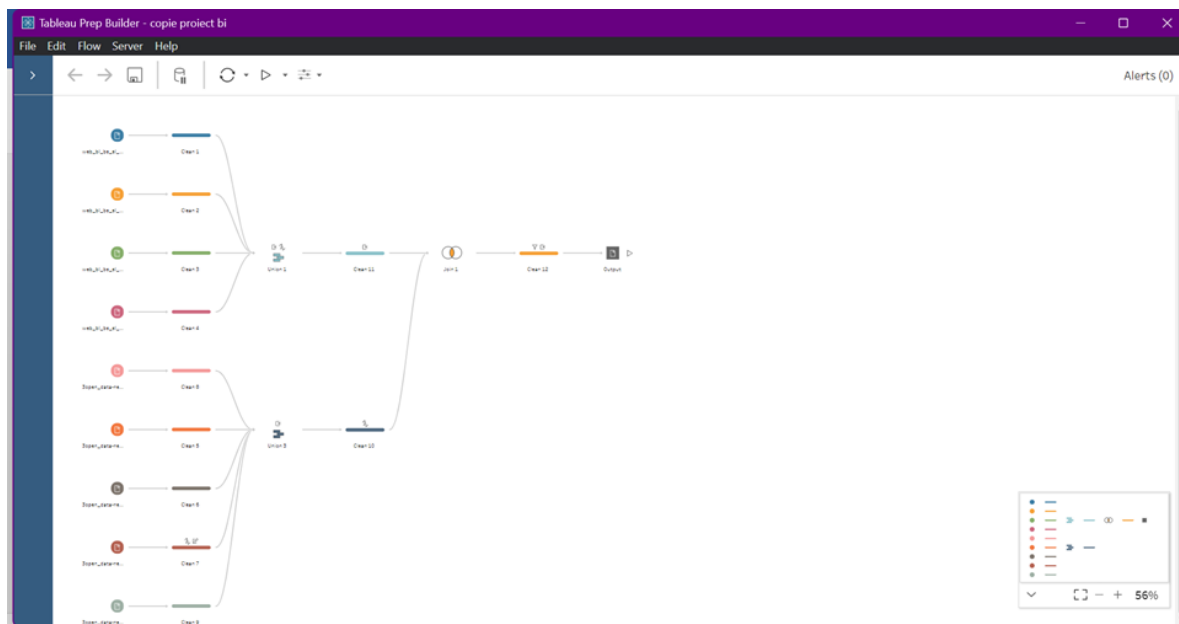
- Primul pas a fost sa aducem in Tableau Prep toate fişierele aferente situaţiilor financiare din cei 4 ani, iar după pasul de *cleaning* să facem un *union* între acestea.

Aceste fişiere conţin date precum CUI, CAEN si elementele definatorii ale bilanţului economic.



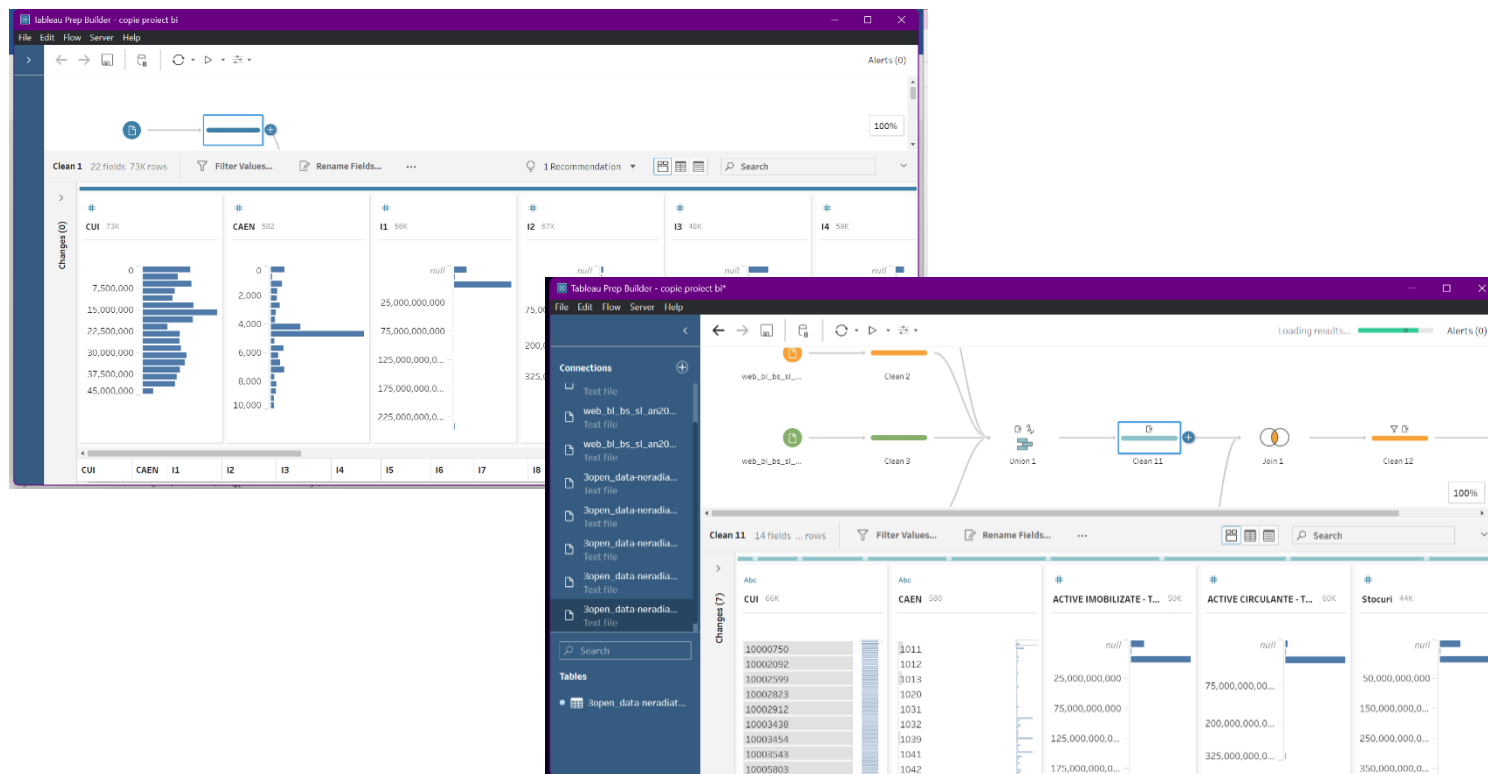
- Al doilea pas a fost de a aduce toate fişierele responsabile cu evidenţa firmelor neradiate cu sediu din cei 4 ani. Şi aceste fişiere au trecut printr-un proces de *cleaning*, urmând mai apoi să se realizeze un *union*. În aceste fişiere regăsim informaţii despre adresă şi denumirea firmei.

În continuare vom discuta despre procesul de filtrare a datelor pe baza codului CAEN si al municipiului din câmpul „adresa”, pentru a ajunge la acest rezultat:

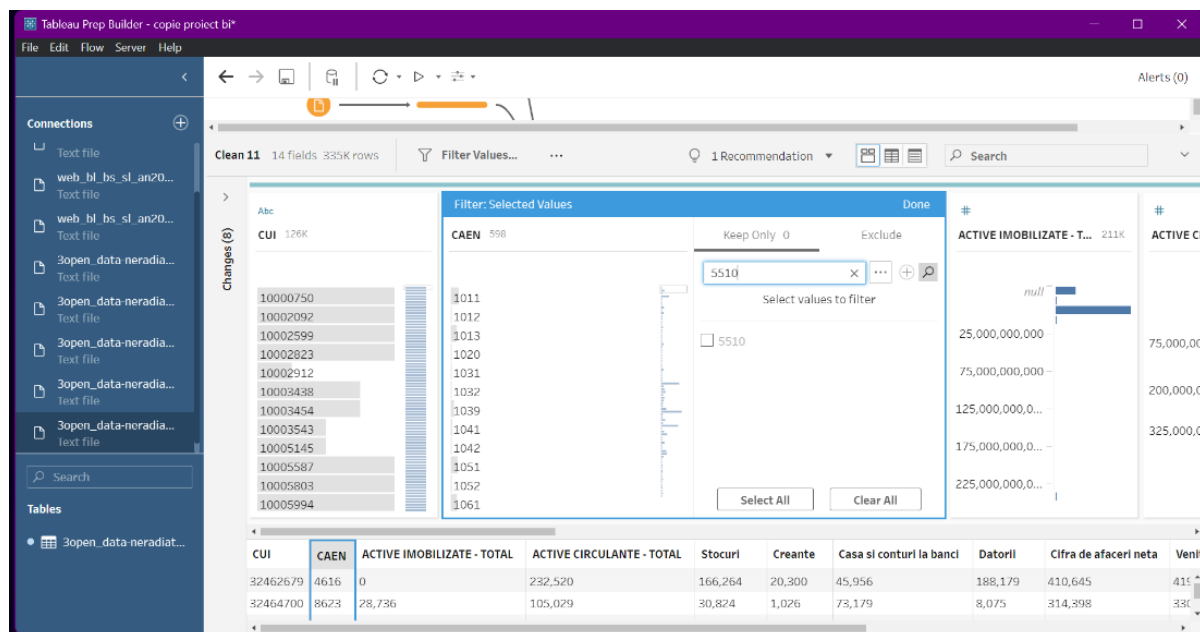


2.2. Descrierea tipurilor de date și a modificărilor făcute

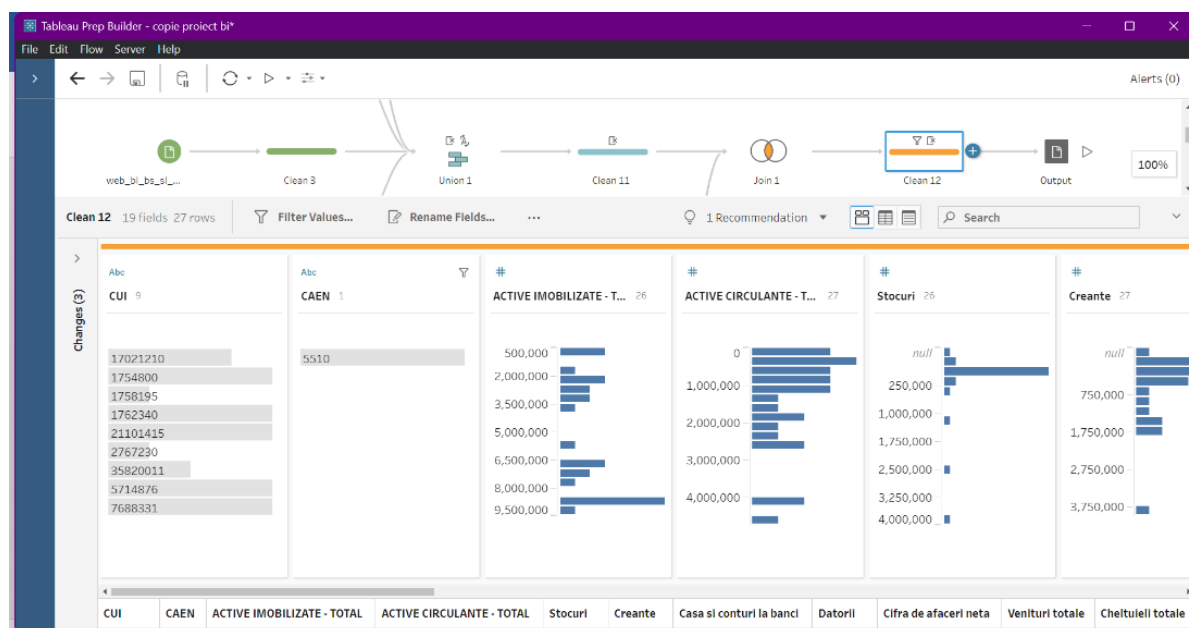
Am modificat tipul de date al coloanelor CUI si CAEN în tipul de date *string*, pentru toate fișierele cu *situații financiare*, astfel:



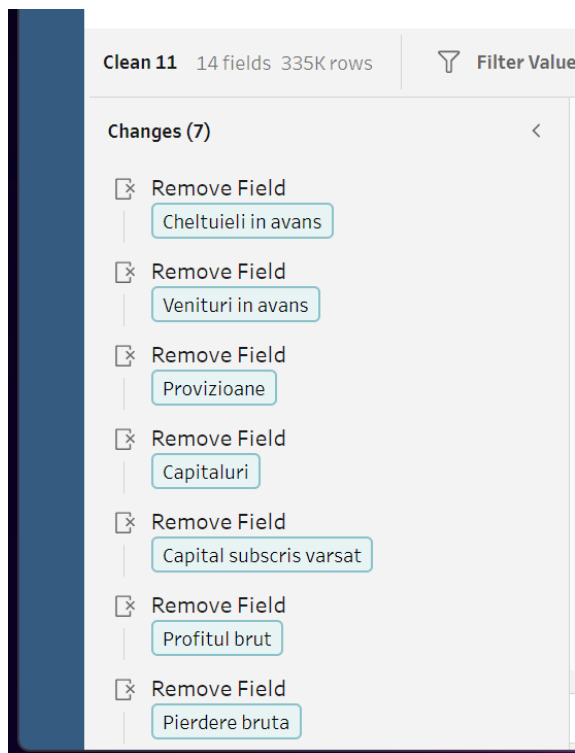
În continuare, am realizat o filtrare pentru a extrage doar firmele care activează în domeniul ales de noi, astfel că filtrarea s-a realizat *pe baza codului CAEN 5510*. Se poate observa în a doua captură de ecran afișarea doar a firmelor cu codul CAEN ales de noi.



În cazul filtrării datelor fișierelor care conțin informații despre firmele neradiate cu sediu, aceasta s-a realizat pe baza numelui municipiului. Rezultatul final este poziționat mai jos.



2.3. Descriere curățare date cu exemplificare captură ecran flow Tableau Prep



Cu ajutorul *cleaning steps*, am realizat o curățare a datelor nefolositoare în procesul nostru de analiză, prin ștergerea anumitor campuri atât din fișierele care contin informații despre situațiile financiare ale firmelor, cât și din cele care conțin informații despre firmele neradiate cu sediu.

De asemenea, cu ajutorul Tableau Prep, s-au realizat redenumiri de câmpuri (pentru o gestiune mai accesibila a datelor) și splits (pentru a extrage doar datele folositoare dintr-un câmp).

2.4. Descriere Union si Join pentru seturile de date utilizate

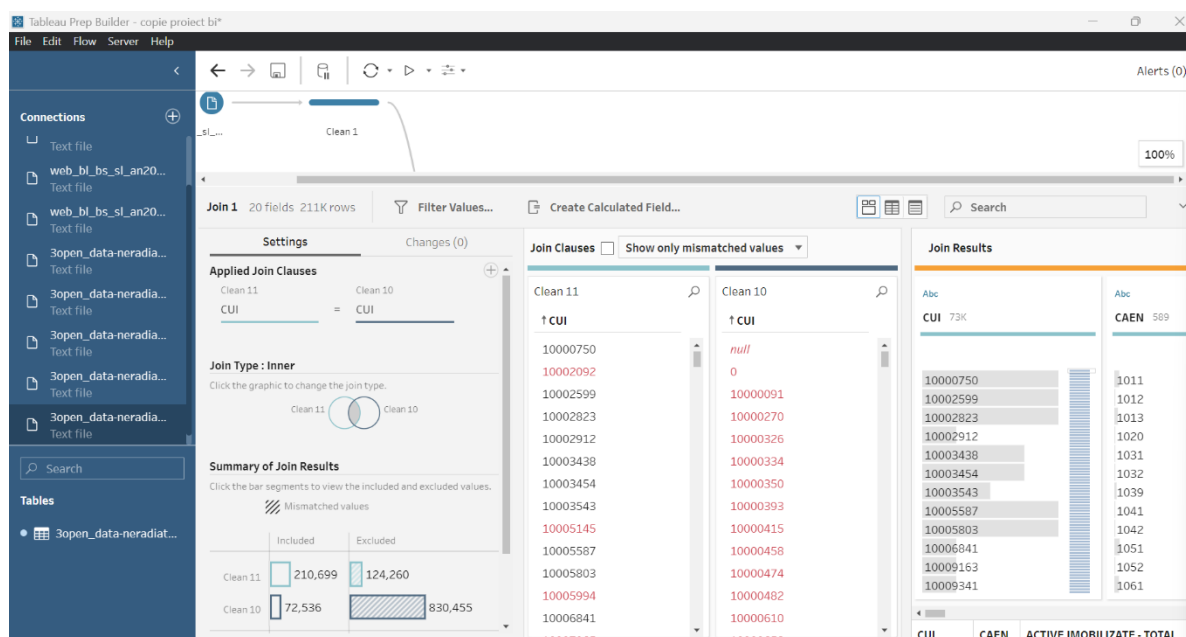
Pentru a ajunge la setul final de date, s-au realizat doua union-uri: unul între fișierele cu numele *WEB_BL_BS_SL_Anul*, reprezentând situațiile financiare ale firmelor, și unul între fișierele cu denumirea *3open_data-neradiate_cu_sediu-08.12.2020.001.csv*, reprezentând firmele neradiate cu sediu.

În prealabil acestor unions, s-au curățat toate datele fiecărui fișier, iar coloanele CUI si CAEN au devenit string (dupa cum am exemplificat la pasul 2.2.).

Pasul următor a fost de a crea JOIN-ul dintre cele doua rezultate ale anilor bilanțieri si al firmelor neradiate cu sediu. Înainte de acest pas, s-a realizat cleaning-ul de la pasul 2.2., pentru a selecta doar firmele cu codul CAEN 5510.

JOIN-ul s-a realizat pe baza câmpului CUI, deoarece reprezinta un câmp comun pentru toate aceste date adunate pâna acum. După acest JOIN, a mai urmat un cleaning step final, pentru prelua doar firmele din Municipiul Alba Iulia.

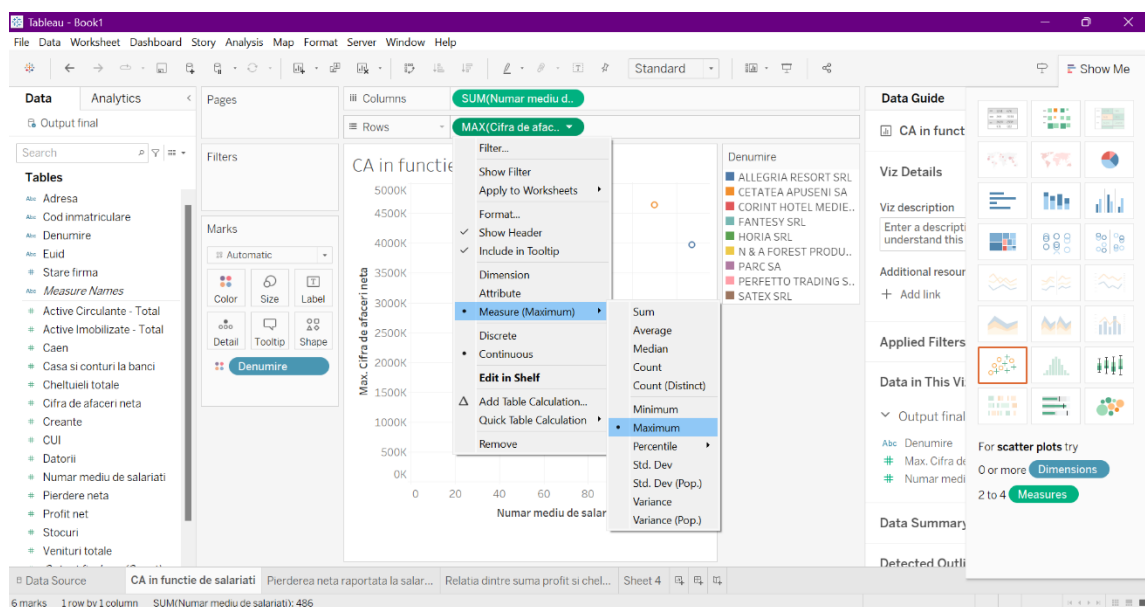
Acest JOIN reprezintă unul de tip INNER. Datele rezultate din Tableau Prep sunt stocate în fișierul „output.csv”.



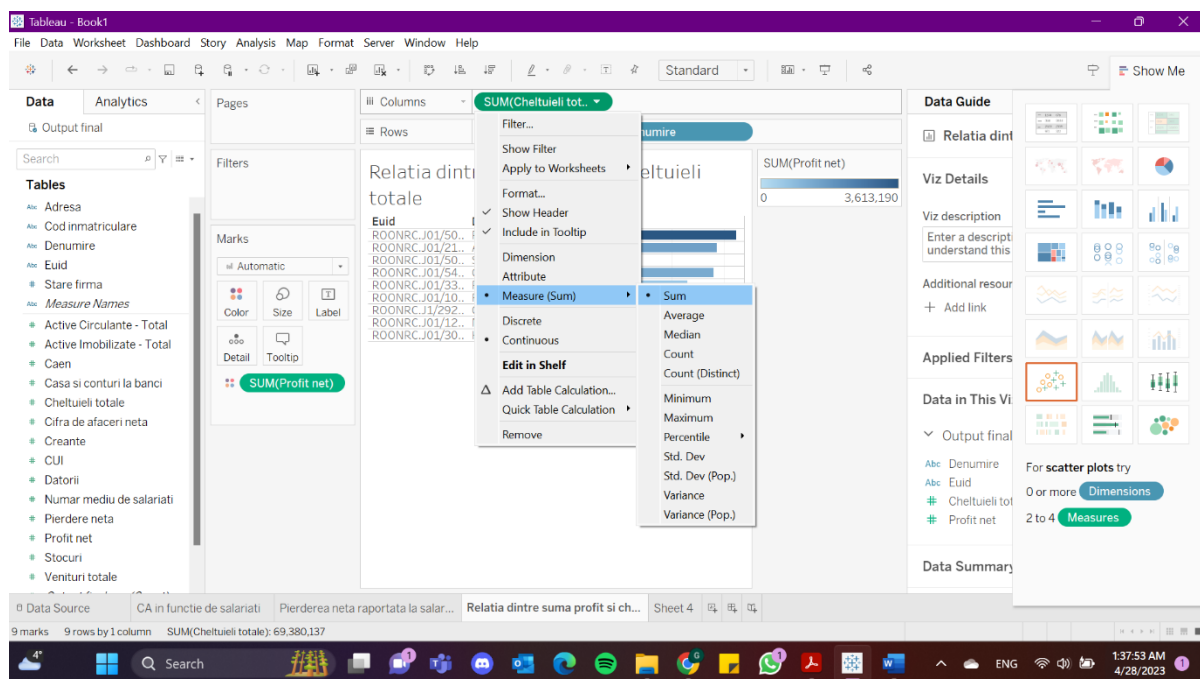
3. Data visualization & analysis

3.1. Câmp calculat, filtre, marcatori personalizați

Am realizat un câmp calculat cu ajutorul funcției agregate MAX asupra datelor care contin cifra de afaceri netă, iar apoi am raportat acest rezultat la numarul mediu de salariați ai firmei în cauză.



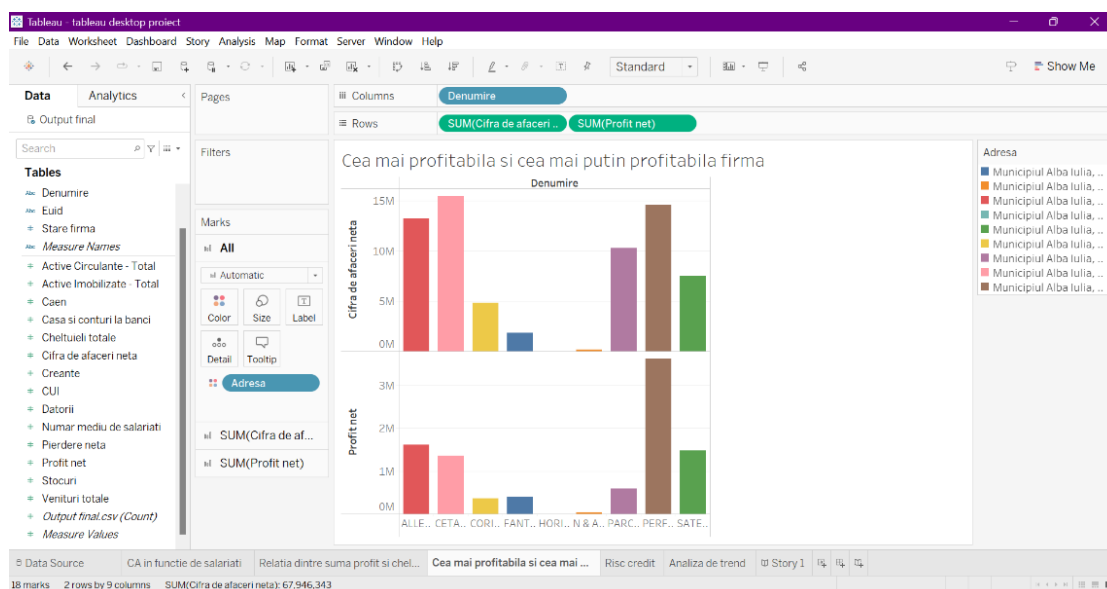
Folosind funcția SUM, am încercat să stabilim o relație între suma cheltuielilor și suma profitului, pentru a stabili care este cea mai profitabilă firmă.



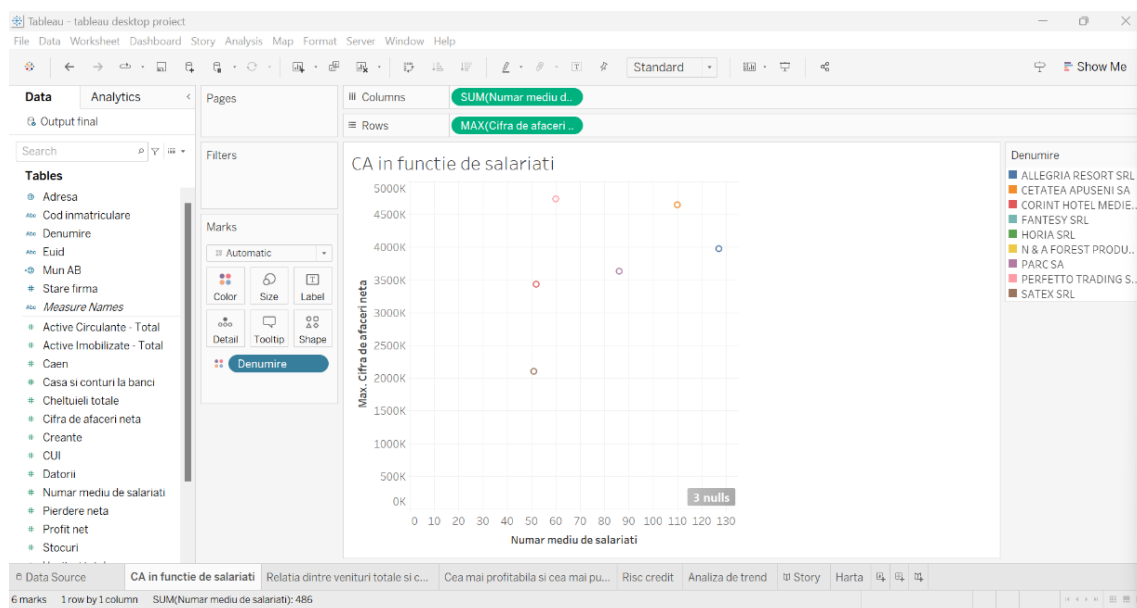
Marcatorii personalizați folosiți în graficele conținute de proiect au fost de tipul *color*, în funcție de diferite proprietăți cum ar fi *denumirea*, *adresa*, *suma profitului net*, sau pe baza filtrului *Measure Names*.

De asemenea, proiectul conține și un marcator de tip *size*, pe baza *profitului net*, în diagramele analizei de trend.

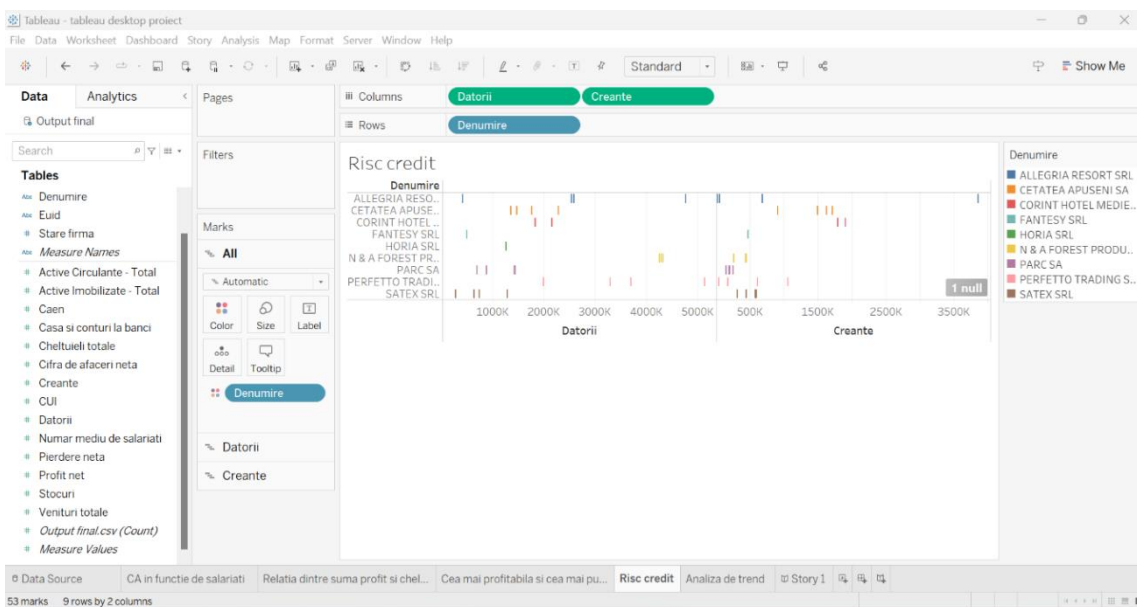
3.2. Cel puțin 3 tipuri de grafice din meniul ShowMe



Primul tip de grafic folosit a fost *Stacked Bars*, pentru a evidenția care firme sunt cele mai profitabile și care sunt cele mai puțin profitabile din setul nostru de date.



Al doilea tip de grafic folosit a fost *Scatter Plots*, pentru a oferi o privire de ansamblu mai clară asupra cifrei de afaceri în funcție de numărul de angajați.



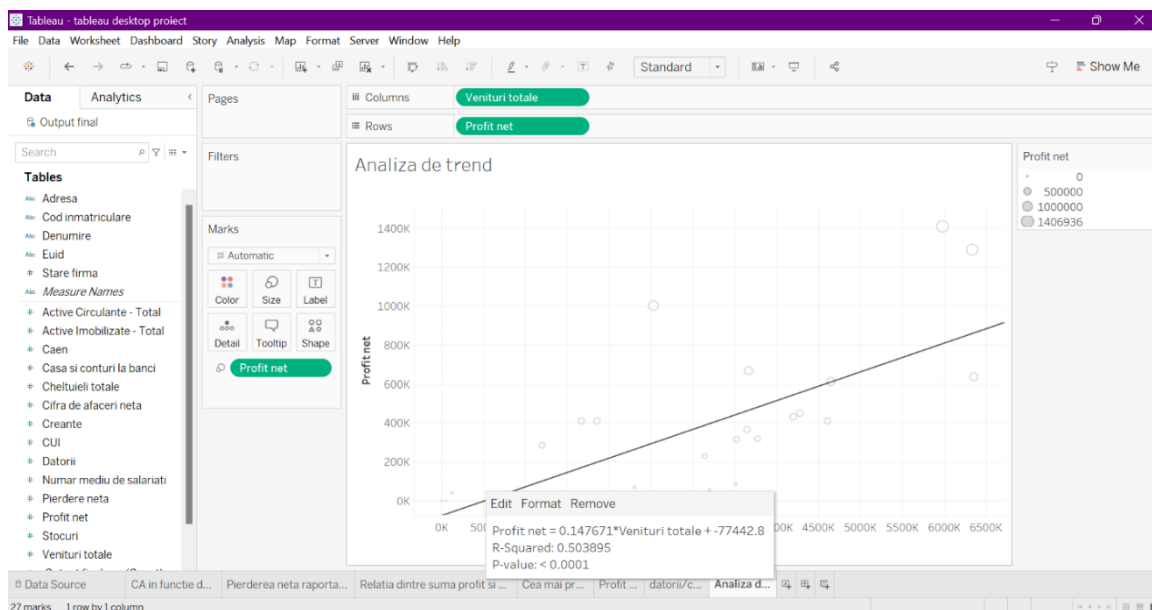
În ce privește al treilea grafic, am optat pentru unul de tip *Grantt*. Acesta este util pentru a evalua riscul de credit al fiecărei firme din setul de date.

Pentru a avea o imagine mai clară asupra firmelor despre care discutăm, am creat o harta care localizează sediile acestora în Alba Iulia. Pentru a realiza aceasta hartă, am folosit marcatorul de tip color, reprezentat de Adresa1.

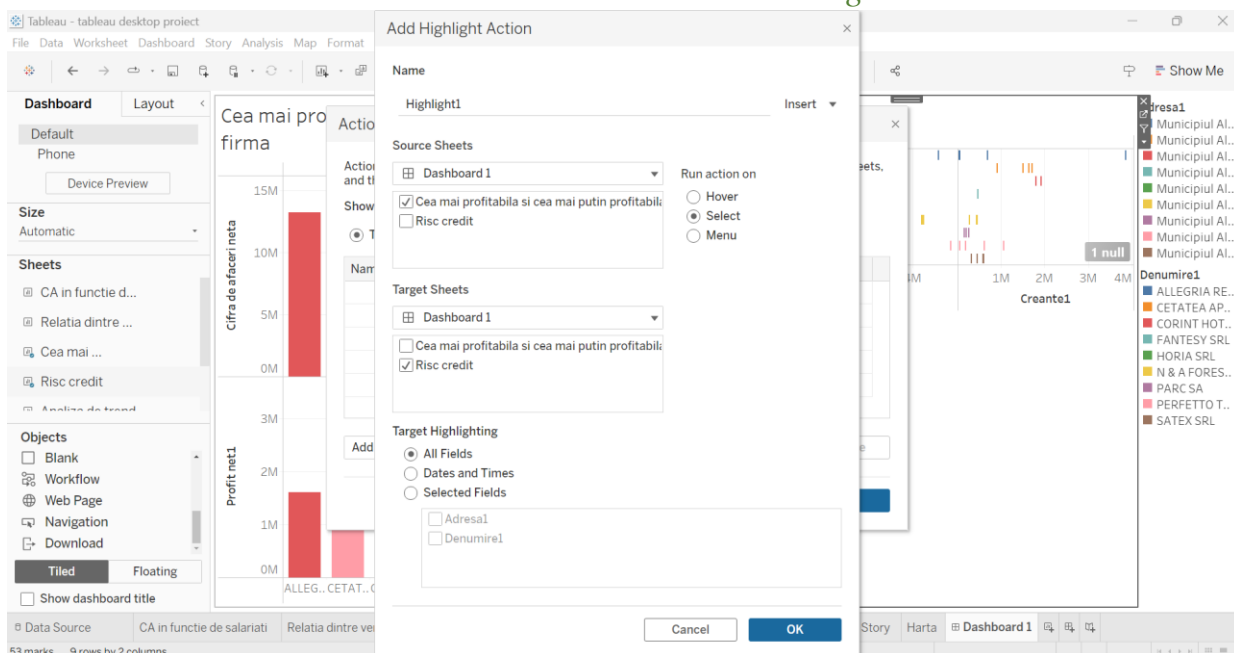
3.5. Analiza de trend pentru evoluția în timp a datelor

În realizarea analizei de trend, am ales datele care privesc cifra de afaceri netă și veniturile totale. Din graficul rezultat, obținem următoarele valori: pentru R-Squared - 0.503895, iar pentru p-value <0.0001.

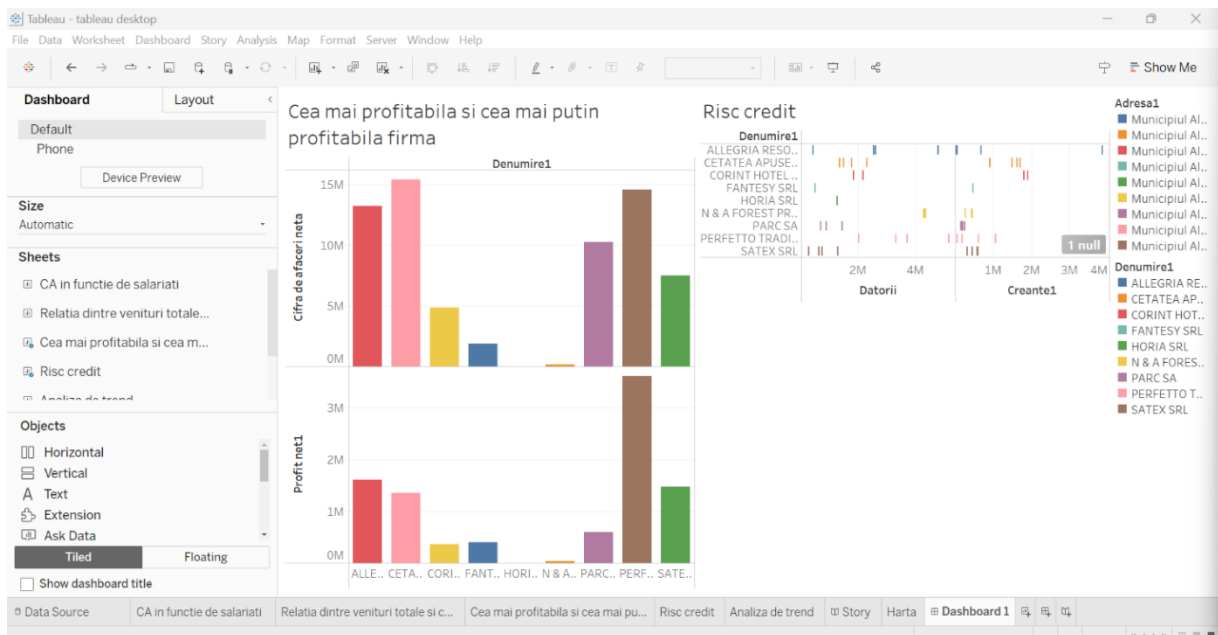
Din aceste date rezultate, deducem ca exista o corelație semnificativă între cele doua variabile, iar valoarea foarte mica a lui p-value (aproape de 0) întărește aceasta constatare.



4. Crearea unui Dashboard cu date legate



Pentru acest proiect am creat un Dashboard care se bazează pe graficul aferent celor mai profitabile si celor mai puțin profitabile firme din setul nostru de date.



5. Crearea unei povești (Story)

Pentru a grupa toate graficele create pana acum, am realizat un Story. Acesta oferă un acces mai ușor si mai organizat asupra tuturor acestor grafice, pentru o analiză mai bună.

Story



Story

