

Modelarea si Analiza Datelor pentru Decizii de Management: Pregătirea datelor, antrenarea modelelor de predicție, vizualizarea datelor

I. Pregătirea datelor

Setul de date ce trebuie pregătit pentru aplicarea unui algoritm ML este:

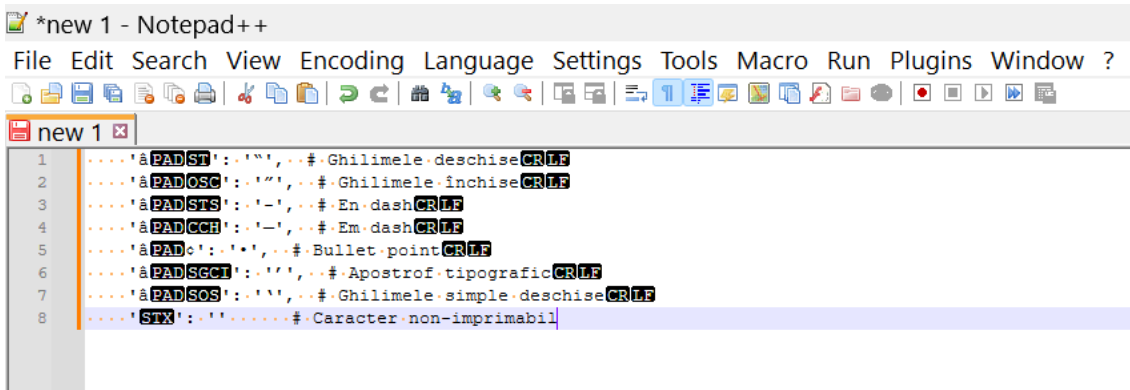
<https://www.kaggle.com/datasets/anoopjohny/new-york-city-job-dataset/data>.

Acest set de date, NYC Jobs, conține 30 de coloane și 6743 de intrări, cu valori lipsă și tipuri diferite de date.

A. Stabilim codificarea corecta

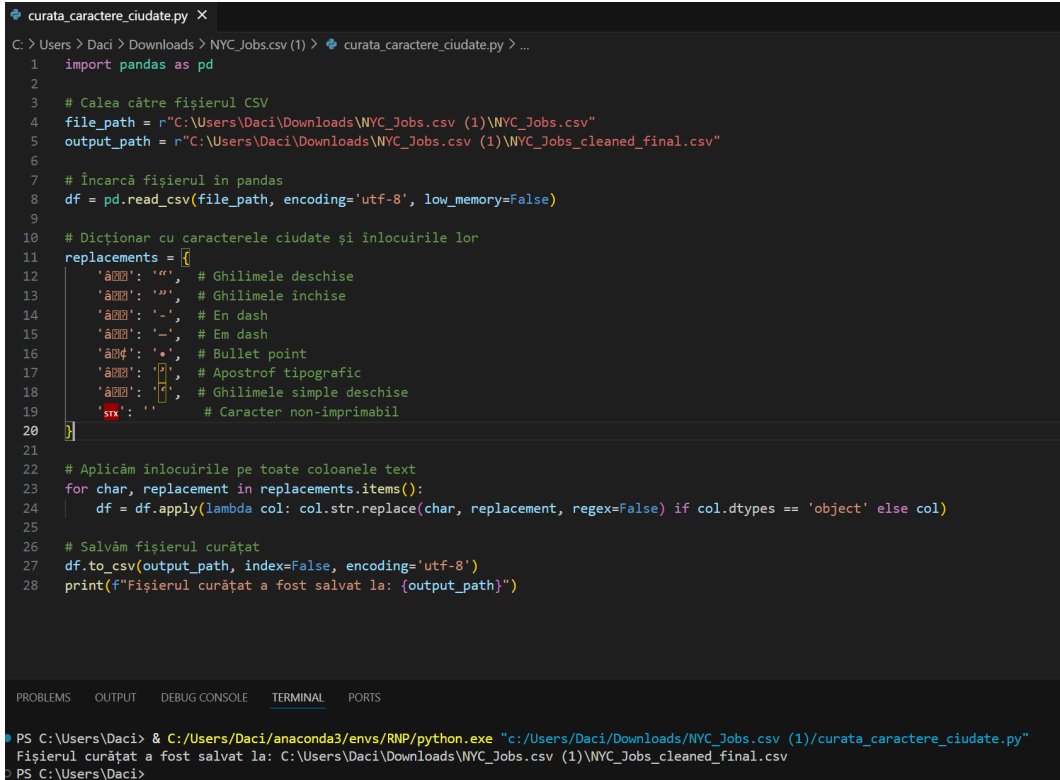
NYC_Jobs.csv, fișierul care conține setul de date este codat în UTF-8, dar conține caractere care au fost codificate diferit.

Am identificat caracterele:



```
*new 1 - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
new 1 x
1 .....âPADST'::'''',...#Ghilimele.deschiseCRLF
2 .....âPADOSC'::'''',...#Ghilimele.închiseCRLF
3 .....âPADSTS'::'''',...#En.dashCRLF
4 .....âPADCCH'::'''',...#Em.dashCRLF
5 .....âPADo'::'''',...#Bullet.pointCRLF
6 .....âPADSGCI'::'''',...#Apostrof.tipograficCRLF
7 .....âPADSOS'::'''',...#Ghilimele.simple.deschiseCRLF
8 .....STX'::'''',...#Caracter.non-imprimabil
```

Le-am înlocuit cu ajutorul următorului cod Python:



```
curata_caractere_ciudate.py X
C:\Users\Daci> Downloads\NYC_Jobs.csv (1) curata_caractere_ciudate.py > ...
1 import pandas as pd
2
3 # Calea către fișierul CSV
4 file_path = r"C:\Users\Daci\Downloads\NYC_Jobs.csv (1)\NYC_Jobs.csv"
5 output_path = r"C:\Users\Daci\Downloads\NYC_Jobs.csv (1)\NYC_Jobs_cleaned_final.csv"
6
7 # Încarcă fișierul în pandas
8 df = pd.read_csv(file_path, encoding='utf-8', low_memory=False)
9
10 # Dictionar cu caracterele ciudate și înlocuirile lor
11 replacements = {
12     'âPADST': "'", # Ghilimele deschise
13     'âPADOSC': '"', # Ghilimele închise
14     'âPADSTS': '-', # En dash
15     'âPADCCH': '-', # Em dash
16     'âPADo': '.', # Bullet point
17     'âPADSGCI': "'", # Apostrof tipografic
18     'âPADSOS': "'", # Ghilimele simple deschise
19     'STX': '' # Caracter non-imprimabil
20 }
21
22 # Aplicăm înlocuirile pe toate coloanele text
23 for char, replacement in replacements.items():
24     df = df.apply(lambda col: col.str.replace(char, replacement, regex=False) if col.dtypes == 'object' else col)
25
26 # Salvăm fișierul curățat
27 df.to_csv(output_path, index=False, encoding='utf-8')
28 print(f"Fișierul curățat a fost salvat la: {output_path}")

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
PS C:\Users\Daci> & C:\Users\Daci\anaconda3\envs\RNP\python.exe "C:\Users\Daci\Downloads\NYC_Jobs.csv (1)\curata_caractere_ciudate.py"
Fișierul curățat a fost salvat la: C:\Users\Daci\Downloads\NYC_Jobs.csv (1)\NYC_Jobs_cleaned_final.csv
PS C:\Users\Daci>
```

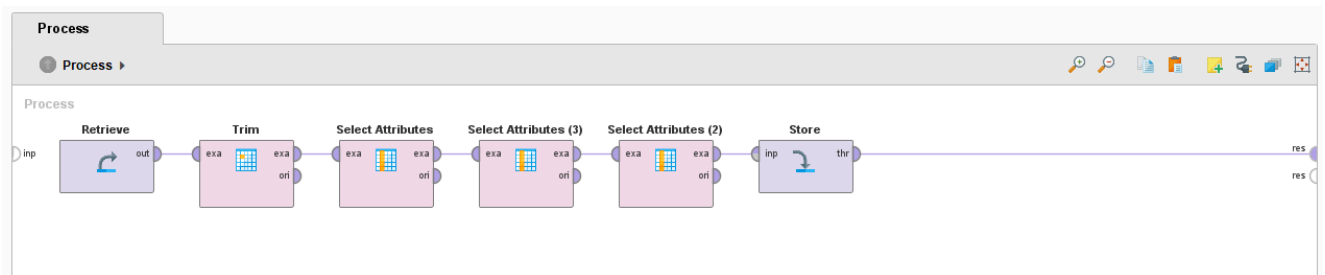
B. Curatarea datelor - valori lipsa

-Aplicam operatorul **Trim** din Altair AI Studio care elimină **spațiile albe** (leading și trailing spaces) doar din **coloanele cu valori nominale** (texte).

-Analizam datele in tab-ul Results - Statistics si observam ca trebuie sa:

1. Stergem coloana Recruitment Contact deoarece are toate valorile lipsa(o vom sterge prin operatorul Select Attributes).
2. Stergem coloana Process Date(Date when the hiring process was initiated) deoarece are o singura valoare pentru toate inregistrarile(10/03/2023) si nu pare sa fie utila.
3. Stergem coloana Work Location 1 deoarece este similara cu coloana Work Location (coloana Work Location 1 are multe valori lipsă pe cand Work Location nu are valori lipsa).

Procesul in Altair arata asa:



C. Transformarea tipului de date ale unor atribute

Am transformat cu Turbo prep coloana Post Until din tip category in tip date.

Job Category	Full-Time/Part-Time	Career Level	Salary Frequency	Work Location	Division/Work Group	Job Description	Minimum Qualifications	Preferred Skills	Additional Information	To Apply	Hours/Shift	Residency Requirements	Post Until	Job ID Number	# Of Fills
Technology, Data & Analytics	F	Experienced (n..)	Annual	42-09 28th Street	Network Techn...	ONLY PERMAN...	(1) A baccalaur...	0?Bachelor's d...	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	593004	1
Health Building...	F	Experienced (n..)	Annual	30-30 47TH AVE	Clinical Operati...	ONLY PERMAN...	Two years of ful...	ALL applicants ...	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	586225	1
Legal Affairs	F	Experienced (n..)	Annual	1 Centre St., N.Y.	Citywide Equity	DCAS's missio...	1. Admission to ...	0? In depth un...	?	Please go to w...	?	New York City ...	1-Dec-23	589884	1
Engineering, Ar...	F	Experienced (n..)	Annual	30-30 Thomso...	INFRAPRGRM	Only candidate...	1. A baccalaur...	Candidates mu...	?	For City Employ...	35 Hours	New York City ...	1-Dec-23	585249	1
Constituent Ser...	F	Experienced (n..)	Annual	42-09 28th Street	Early Intervention	ONLY PERMAN...	1. A baccalaur...	?	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	565420	1
Health Building...	F	Experienced (n..)	Annual	30-30 47TH AVE	Clinical Operati...	ONLY PERMAN...	Two years of ful...	ALL applicants ...	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	586225	1
Health	F	Entry-Level	Annual	42-09 28th Street	MNY Comm En...	This position m...	1. A baccalaur...	-The PMH-AAN...	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	586295	15
Legal Affairs	F	Experienced (n..)	Annual	1 Centre St., N.Y.	Citywide Equity	DCAS's missio...	1. Admission to ...	The candidate ...	?	Please go to w...	?	New York City ...	1-Dec-23	589906	1
Constituent Ser...	F	Experienced (n..)	Annual	42-09 28th Street	Early Intervention	Established in ...	Qualification R...	Preference will ...	**IMPORTANT...	Apply online wit...	?	New York City ...	1-Dec-23	570806	2
Constituent Ser...	F	Experienced (n..)	Annual	90-37 Parsons ...	Clinical Operati...	Established in ...	High school di...	?	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	581650	1
Legal Affairs	F	Experienced (n..)	Annual	1 Centre St., N.Y.	Citywide Equity	DCAS's missio...	1. Admission to ...	The candidate ...	?	Please go to w...	?	New York City ...	1-Dec-23	589906	1
Health	F	Experienced (n..)	Annual	NYC - All Borou...	BPHC Health E...	Established in ...	1. A baccalaur...	?	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	583450	1
Constituent Ser...	F	Experienced (n..)	Annual	42-09 28th Street	Early Intervention	ONLY PERMAN...	1. A baccalaur...	?	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	565420	1
Public Safety, L...	F	Experienced (n..)	Annual	125 Worth Ste...	Child Care	ONLY PERMAN...	1. A baccalaur...	Candidate mus...	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	590018	1
Constituent Ser...	F	Experienced (n..)	Annual	100 Gold Street	HCV Unit	About the Agen...	Qualification RL...	0?Excellent Co...	?	Apply Online	?	New York City ...	1-Dec-23	609271	1
Public Safety, L...	F	Experienced (n..)	Annual	125 Worth Ste...	Child Care	ONLY PERMAN...	1. A baccalaur...	Candidate mus...	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	590018	1
Technology, Da...	F	Experienced (n..)	Annual	42-09 28th Street	Network Techn...	ONLY PERMAN...	(1) A baccalaur...	Bachelor's deg...	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	593034	1
Administration ...	F	Experienced (n..)	Annual	90-27 Parsons ...	Early Intervention	ONLY PERMAN...	1. A baccalaur...	Proficiency with...	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	591592	1
Administration ...	F	Experienced (n..)	Annual	90-27 Parsons ...	Early Intervention	ONLY PERMAN...	1. A baccalaur...	Proficiency with...	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	591592	1
Health	F	Experienced (n..)	Annual	NYC - All Borou...	BPHC Health E...	Established in ...	1. A baccalaur...	?	**IMPORTANT ...	Apply online wit...	?	New York City ...	1-Dec-23	583450	1

D. Eliminarea înregistrărilor duplicate

Altair AI Studio Educational 2024.1.0 @ Piercing

Views Design Results Turbo Prep Auto Model Interactive Analysis

Turbo Prep

Cleanse

28 columns selected

AUTO CLEANING

REMOVE LOW QUALITY

REMOVE CORRELATED

REPLACE MISSING

NORMALIZATION

DISCRETIZATION

SUMMARY BACKLOGGING

PCA

REMOVE DUPLICATES

Test missing values as duplicates?

APPLY

DataPrep_2_change_type_date (2)

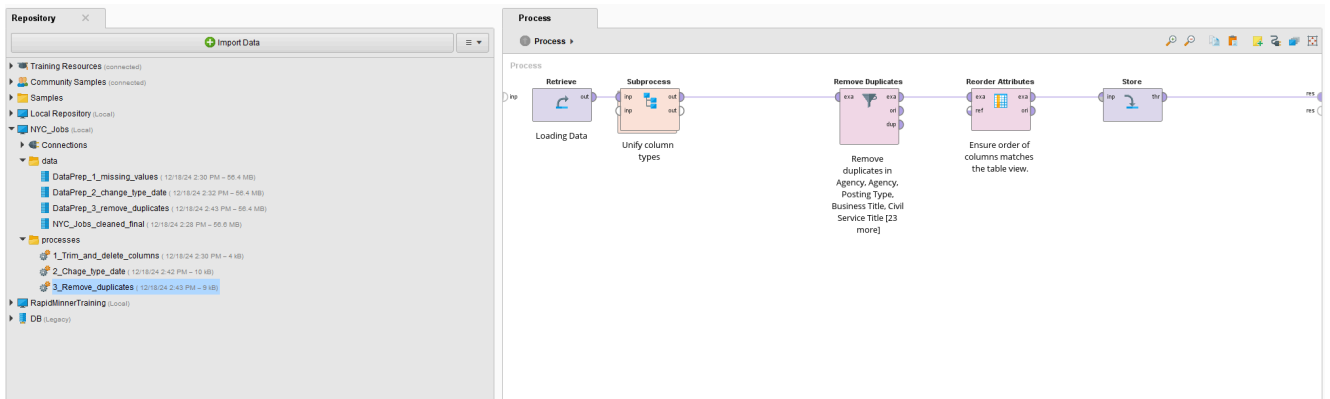
Select a column to clean (hold Shift for selecting a range of columns; Ctrl for (de)selecting multiple columns; Alt to select all columns of the same type; Ctrl-A for all columns). Make changes and commit them at the end.

6,743 rows - 27 columns (20 nominal, 4 numerical, 3 date)

Agency	Posting Type	Business Title	Civil Service ...	Title Classif...	Title Code No	Level	Job Category	Full-Time/Part...	Career Level	Salary Freq...	Work Location	Division/Work...	Job Descript...	Minimum Qu...	Preferred SKI...	Additional Inf...	To Apply	Hours/Shift
DEPT OF HEAL...	External	IT Asset and D...	COMPUTER S...	Competitive-1	13632	2	Technology, Da...	F	Experienced (In...	Annual	42-09 28th Street	Network Techn...	ONLY PERMAN...	(1) A baccalaur...	2 Bachelor's d...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	Internal	Stock Worker S...	SUPERVISOR	Competitive-1	12202	2	Health Building	F	Experienced (In...	Annual	30-30 47TH AVE	Clinical Operat...	ONLY PERMAN...	Two years of ful...	All applicants ...	***IMPORTANT ...	Apply online wit...	?
DEPT OF CITY...	Internal	Investigations ...	AGENCY ATTO...	Non-Competit...	30087	3	Legal Affairs	F	Experienced (In...	Annual	1 Centre St. N.Y.	Citywide Equi...	DCAS's missio...	1 Admission to...	27 In depth un...	?	Please go to w...	?
DEPT OF DESI...	Internal	Deputy Director	ADMINISTRAT...	Competitive-1	83008	0	Engineering, Ar...	F	Experienced (In...	Annual	30-30 Thoms...	INFRASTRUCTUR...	Only candidate...	1 A baccalaur...	Candidates mu...	?	For City Emplo...	35 Hours
DEPT OF HEAL...	External	NYSE Help De...	PRINCIPAL AD...	Competitive-1	10124	1	Constituent Ser...	F	Experienced (In...	Annual	42-09 28th Street	Early Intervent...	ONLY PERMAN...	1 A baccalaur...	?	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	External	Stock Worker S...	SUPERVISOR	Competitive-1	12202	2	Health Building	F	Experienced (In...	Annual	30-30 47TH AVE	Clinical Operat...	ONLY PERMAN...	Two years of ful...	All applicants ...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	External	PMH-A&MPT C...	COMMUNITY C...	Non-Competit...	56058	0	Health	F	Entry-Level	Annual	42-09 28th Street	MHY Comm En...	This position m...	1 A baccalaur...	-The PMH-A&M...	***IMPORTANT ...	Apply online wit...	?
DEPT OF CITY...	External	Investigations ...	AGENCY ATTO...	Non-Competit...	30087	2	Legal Affairs	F	Experienced (In...	Annual	1 Centre St. N.Y.	Citywide Equi...	DCAS's missio...	1 Admission to...	The candidate ...	?	Please go to w...	?
DEPT OF HEAL...	Internal	Developmental...	SR HEALTHCA...	Non-Competit...	83052	1	Constituent Ser...	F	Experienced (In...	Annual	42-09 28th Street	Early Intervent...	Established in...	Qualification R...	Preference wil...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	External	Laboratory Ass...	LABORATORY	Competitive-1	21512	1	Constituent Ser...	F	Experienced (In...	Annual	90-37 Persons...	Clinical Operat...	Established in...	High school d...	?	***IMPORTANT ...	Apply online wit...	?
DEPT OF CITY...	Internal	Investigations ...	AGENCY ATTO...	Non-Competit...	30087	2	Legal Affairs	F	Experienced (In...	Annual	1 Centre St. N.Y.	Citywide Equi...	DCAS's missio...	1 Admission to...	The candidate ...	?	Please go to w...	?
DEPT OF HEAL...	Internal	Navigation Sup...	COMMUNITY C...	Non-Competit...	56058	0	Health	F	Experienced (In...	Annual	NYC - All Boros...	BP&HC Health E...	Established in...	1 A baccalaur...	?	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	Internal	NYSE Help De...	PRINCIPAL AD...	Competitive-1	10124	1	Constituent Ser...	F	Experienced (In...	Annual	42-09 28th Street	Early Intervent...	ONLY PERMAN...	1 A baccalaur...	?	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	External	Childcare Insp...	ASSOCIATE P...	Competitive-1	31220	2	Public Safety, L...	F	Experienced (In...	Annual	125 Worth Stre...	Child Care	ONLY PERMAN...	1 A baccalaur...	Candidate mus...	***IMPORTANT ...	Apply online wit...	?
HOUSING PRE...	Internal	Team Leader f...	COMMUNITY C...	Non-Competit...	56057	0	Constituent Ser...	F	Experienced (In...	Annual	100 Gold Street	HCV Unit	About the Agen...	Qualification R...	27Excellent Co...	?	Apply Online	?
DEPT OF HEAL...	Internal	Childcare Insp...	ASSOCIATE P...	Competitive-1	31220	2	Public Safety, L...	F	Experienced (In...	Annual	125 Worth Stre...	Child Care	ONLY PERMAN...	1 A baccalaur...	Candidate mus...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	Internal	IT Desktop and...	COMPUTER S...	Competitive-1	13632	2	Technology, Da...	F	Experienced (In...	Annual	42-09 28th Street	Network Techn...	ONLY PERMAN...	(1) A baccalaur...	Bachelor's deg...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	External	Scheduler	PRINCIPAL AD...	Competitive-1	10124	1	Administration	F	Experienced (In...	Annual	90-27 Persons...	Early Intervent...	ONLY PERMAN...	1 A baccalaur...	Proficiency with...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	Internal	Scheduler	PRINCIPAL AD...	Competitive-1	10124	1	Administration	F	Experienced (In...	Annual	90-27 Persons...	Early Intervent...	ONLY PERMAN...	1 A baccalaur...	Proficiency with...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	External	Navigation Sup...	COMMUNITY C...	Non-Competit...	56058	0	Health	F	Experienced (In...	Annual	NYC - All Boros...	BP&HC Health E...	Established in...	1 A baccalaur...	?	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	Internal	IT Asset and D...	COMPUTER S...	Competitive-1	13632	2	Technology, Da...	F	Experienced (In...	Annual	42-09 28th Street	Network Techn...	ONLY PERMAN...	(1) A baccalaur...	27Bachelor's d...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	External	IT Desktop and...	COMPUTER S...	Competitive-1	13632	2	Technology, Da...	F	Experienced (In...	Annual	42-09 28th Street	Network Techn...	ONLY PERMAN...	(1) A baccalaur...	Bachelor's deg...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	Internal	Laboratory Ass...	LABORATORY	Competitive-1	21512	1	Constituent Ser...	F	Experienced (In...	Annual	90-37 Persons...	Clinical Operat...	Established in...	High school d...	?	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	Internal	PMH-A&MPT C...	COMMUNITY C...	Non-Competit...	56058	0	Health	F	Entry-Level	Annual	42-09 28th Street	MHY Comm En...	This position m...	1 A baccalaur...	-The PMH-A&M...	***IMPORTANT ...	Apply online wit...	?
DEPT OF DESI...	External	Deputy Director	ADMINISTRAT...	Competitive-1	83008	0	Engineering, Ar...	F	Experienced (In...	Annual	30-30 Thoms...	INFRASTRUCTUR...	Only candidate...	1 A baccalaur...	Candidates mu...	?	For City Emplo...	35 Hours
DEPT OF HEAL...	External	Developmental...	SR HEALTHCA...	Non-Competit...	83052	1	Constituent Ser...	F	Experienced (In...	Annual	42-09 28th Street	Early Intervent...	Established in...	Qualification R...	Preference wil...	***IMPORTANT ...	Apply online wit...	?
DEPT OF CITY...	External	Investigations ...	AGENCY ATTO...	Non-Competit...	30087	3	Legal Affairs	F	Experienced (In...	Annual	1 Centre St. N.Y.	Citywide Equi...	DCAS's missio...	1 Admission to...	27 In depth un...	?	Please go to w...	?
DEPT OF HEAL...	Internal	Data Governan...	CITY RESEAR...	Non-Competit...	21744	3	Constituent Ser...	F	Experienced (In...	Annual	42-09 28th Street	Bureau of Epid...	Established in...	1 For Assign...	3 to 5 years of...	***IMPORTANT ...	Apply online wit...	?
DEPT OF HEAL...	External	Data Governan...	CITY RESEAR...	Non-Competit...	21744	3	Constituent Ser...	F	Experienced (In...	Annual	42-09 28th Street	Bureau of Epid...	Established in...	1 For Assign...	3 to 5 years of...	***IMPORTANT ...	Apply online wit...	?
ADMIN FOR C...	External	ELU Agency MB...	AGENCY ATTO...	Non-Competit...	30087	2	Legal Affairs S...	F	Experienced (In...	Annual	110 William St...	Employment L...	THE SELECTE...	1 Admission to...	Experience in E...	THE SELECTE...	APPLICATIONS	?

Au fost sterse 79 de inregistrari.

lata cum arata aceste procese si datele aferente pana in acest moment:



Am importat NYC_Jobs_cleaned_final.csv în Local Repository, în folderul data.

Primul proces, 1_Trim_and_delete_columns:

- Am aplicat operatorul Trim pentru a elimina spațiile albe.
- Am șters coloanele nedorite.
- Am salvat rezultatul ca DataPrep_1_missing_values.csv în Local Repository.

Al doilea proces, 2_Change_type_date:

Am schimbat tipul de date al coloanei Post Until folosind formatul DD-MMM-YY.

Am salvat rezultatul ca DataPrep_2_change_type_date.csv.

Al treilea proces, 3_Remove_duplicates:

Am salvat procesul din Turbo Prep aplicat datelor din DataPrep_2_change_type_date.csv pentru a curata inregistrarile duplicat. Am salvat rezultatul in DataPrep_3_remove_duplicates.

II. Clasificare(Decision Tree) in Altair

Construim un model care prezice dacă un job are salariu „mic”, „mediu” sau „mare”, folosind anumite informații despre job.

Predicțiile modelului (Salary Level prezis) sunt comparate cu valorile reale ale Salary Level(calculate din Salary Range To).

Folosim informațiile din coloanele Agency, Business Title, Job Category, Work Location, Posting Type si Salary Range To ca date de intrare initiale. Am ales aceste coloane deoarece sunt date ce influenteaza nivelul de salariu.

Agentia (Agency): Departamentul guvernamental sau organizația care oferă postul. Diferite agenții pot avea politici salariale distincte, influențând astfel salariul oferit pentru poziții similare.

Tipul postării (Posting Type): Indică dacă anunțul este intern (adresat angajaților existenți) sau extern (deschis publicului larg). Postările interne pot avea structuri salariale diferite față de cele externe, în funcție de politicile organizației.

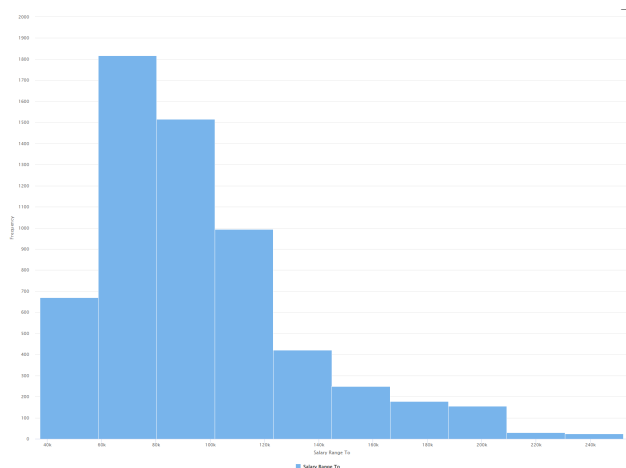
Titlul postului (Business Title): Denumirea utilizată în cadrul organizației pentru postul respectiv. Titluri diferite pot reflecta niveluri de responsabilitate și, implicit, intervale salariale variate.

Categoria postului (Job Category): Categoria de muncă în care se încadrează postul. Anumite categorii pot avea grile salariale specifice, influențând remunerația oferită.

Locația de muncă (Work Location): Locația fizică unde se desfășoară activitatea. Costul vieții și piața muncii locale pot determina variații semnificative în intervalele salariale pentru aceeași poziție în locații diferite.

Verificam in Turbo Prep daca datele din aceste coloane necesita transformari.

Pentru Salary Range To, stergem inregistrarile<10000 pentru a vizualiza si intelege mai bine datele sariale. Datele arata asa:



Având în vedere distribuția salarială și concluziile obținute din grafic, putem propune următoarele praguri optime pentru clasificarea salariilor în mici, medii și mari:

Mic: 40k - 80k
Mediu: 80k - 120k
Mare: >120k

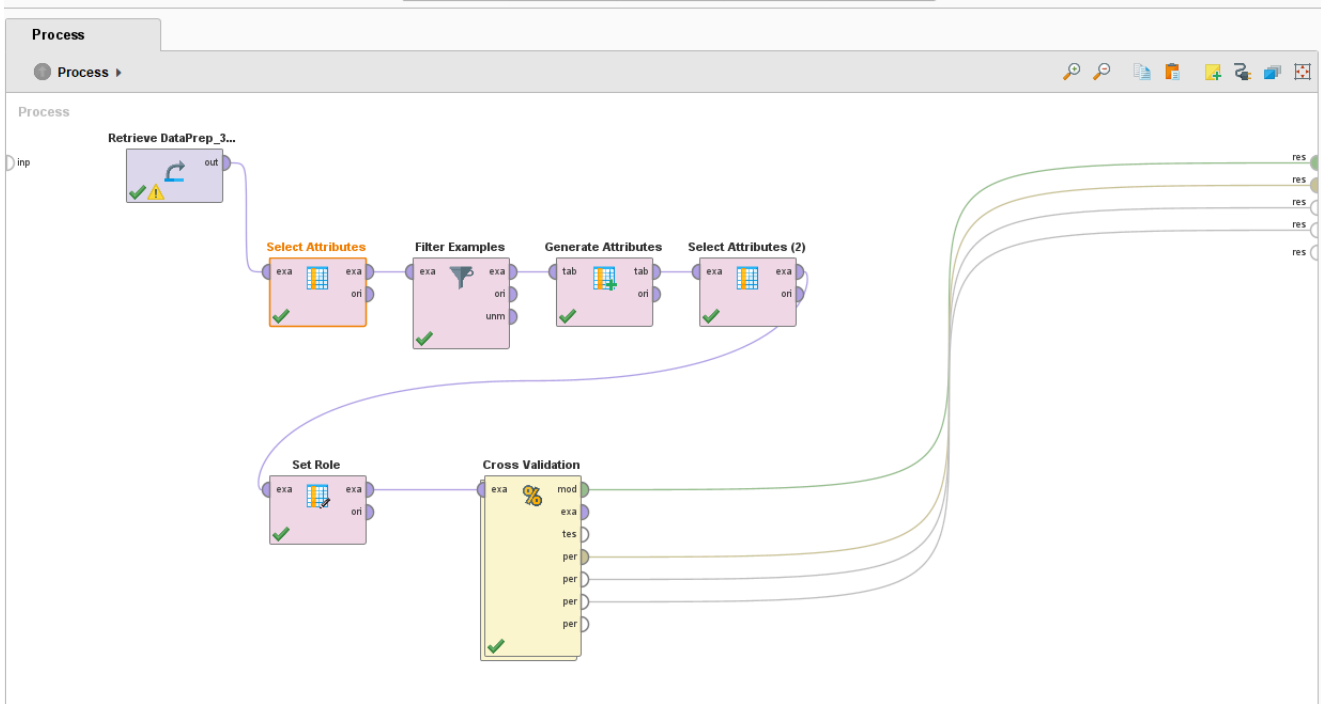
Creăm o coloană nouă, Salary Level, cu valorile „mic”, „mediu” si „mare”.
Salary Level lua valorile date de urmatoarea expresie:

```
if([Salary Range To] <= 80000, "Low",  
  if([Salary Range To] <= 120000, "Medium", "High"))
```

Asa arata datele ce includ si coloana Salary Level calculat din Salary Range to.

Row No.	Agency	Posting Type	Business Title	Job Category	Work Locati...	# Of Positions	Salary Level
1	DEPT OF HE...	External	IT Asset and ...	Technology, ...	42-09 28th St...	1	High
2	DEPT OF HE...	Internal	Stock Worker ...	Health Buildi...	30-30 47TH A...	1	Low
3	DEPT OF CIT...	Internal	Investigation...	Legal Affairs	1 Centre St., ...	1	Medium
4	DEPT OF DE...	Internal	Deputy Director	Engineering, ...	30-30 Thoms...	1	High
5	DEPT OF HE...	External	NYEIS Help ...	Constituent S...	42-09 28th St...	1	Low
6	DEPT OF HE...	External	Stock Worker ...	Health Buildi...	30-30 47TH A...	1	Low
7	DEPT OF HE...	External	PMH-AANHPL...	Health	42-09 28th St...	15	Low
8	DEPT OF CIT...	External	Investigation...	Legal Affairs	1 Centre St., ...	1	Medium
9	DEPT OF HE...	Internal	Development...	Constituent S...	42-09 28th St...	2	Low
10	DEPT OF HE...	External	Laboratory As...	Constituent S...	90-37 Parson...	1	Low
11	DEPT OF CIT...	Internal	Investigation...	Legal Affairs	1 Centre St., ...	1	Medium
12	DEPT OF HE...	Internal	Navigation S...	Health	NYC - All Bor...	1	Low
13	DEPT OF HE...	Internal	NYEIS Help ...	Constituent S...	42-09 28th St...	1	Low
14	DEPT OF HE...	External	Childcare Ins...	Public Safety...	125 Worth Str...	1	Medium
15	HOUSING P...	Internal	Team Leader...	Constituent S...	100 Gold Stre...	1	Low
16	DEPT OF HE...	Internal	Childcare Ins...	Public Safety...	125 Worth Str...	1	Medium
17	DEPT OF HE...	Internal	IT Desktop a...	Technology, ...	42-09 28th St...	1	High
18	DEPT OF HE...	External	Scheduler	Administratio...	90-27 Parson...	1	Low
19	DEPT OF HE...	Internal	Scheduler	Administratio...	90-27 Parson...	1	Low
20	DEPT OF HE...	External	Navigation S...	Health	NYC - All Bor...	1	Low
21	DEPT OF HE...	Internal	IT Asset and ...	Technology, ...	42-09 28th St...	1	High
22	DEPT OF HE...	External	IT Desktop a...	Technology, ...	42-09 28th St...	1	High
23	DEPT OF HE...	Internal	Laboratory As...	Constituent S...	90-37 Parson...	1	Low
24	DEPT OF HE...	Internal	PMH-AANHPL...	Health	42-09 28th St...	15	Low
25	DEPT OF DE...	External	Deputy Director	Engineering, ...	30-30 Thoms...	1	High
26	DEPT OF HE...	External	Development...	Constituent S...	42-09 28th St...	2	Low
27	DEPT OF CIT...	External	Investigation...	Legal Affairs	1 Centre St., ...	1	Medium
28	DEPT OF HE...	Internal	Data Governa...	Constituent S...	42-09 28th St...	1	Medium
29	DEPT OF HE...	External	Data Governa...	Constituent S...	42-09 28th St...	1	Medium

Mai jos este procesul care va genera setul de date pentru modelul de clasificare:



1. Retrieve (Retrieve DataPrep_3):

Importă setul de date pregătit la tema 2.

2. Select Attributes (Primul Select Attributes):

Selectează doar atributele relevante pentru model: Agency, Business Title, Job Category, Work Location, Posting Type si Salary Range To.

3. Filter Examples:

Aplică un filtru pentru a ștergem înregistrările cu valori <10000 din coloana Salary Range To.

4. Generate Attributes:

Creează un nou atribut (Salary Level) bazat pe Salary Range To.

Utilizează o expresie pentru a clasifica salariile în 3 categorii (Low, Medium, High), în funcție de pragurile stabilite:

Low: 40k - 80k.

Medium: 80k - 120k.

High: >120k.

5. Select Attributes (Al doilea Select Attributes):

Elimină coloana Salary Range To, care nu mai este necesară pentru antrenarea modelului(a fost folosit doar pentru a crea Salary Level).

6. Set Role:

Setează Salary Level ca label (variabila țintă pe care modelul trebuie să o prezică).

Atributele (restul coloanelor) sunt informațiile pe care modelul le folosește pentru a învăța să facă predicția.

7. Cross Validation:

Evaluează performanța modelului folosind o metodă de validare 10-fold.

Ce face:

Zona Training:

Construiește modelul de clasificare folosind Decision Tree.

Modelul „învăță” relațiile dintre datele despre job și nivelurile salariale (Salary Level) folosind algoritmul Decision Tree.

De exemplu:

„Manager” dintr-o agenție financiară → Salariu „Mare”.

„Asistent” dintr-o agenție educațională → Salariu „Mic”.

Modelul prezice nivelul salarial („Mic”, „Mediu”, „Mare”) pentru fiecare job.

Zona Testing:

Aplică modelul pe datele de test și compară predicțiile cu valorile reale (Salary Level).

Modelul compară Salary Level prezis (din Apply Model) cu Salary Level real (etichetă existentă din Set Role).

Compararea este realizată automat de operatorul Performance, care calculează metricile precum Accuracy(procentul total de predicții corecte), Precision(din predicțiile unei categorii, câte sunt corecte) și Recall(din toate valorile reale ale unei categorii, câte sunt identificate corect).

Dacă predicția este identică cu valoarea reală → Corect. Dacă nu, → Greșit.

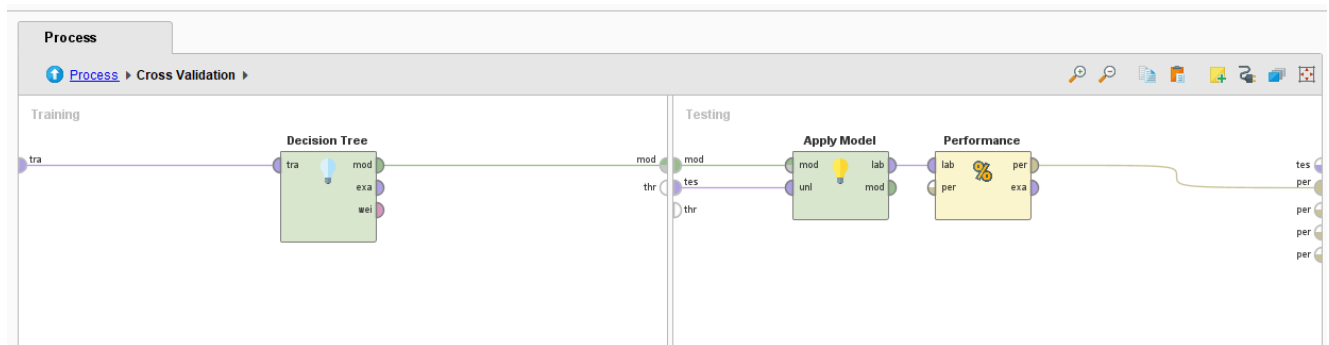
8. Ieșirile Cross Validation:

Ce se trimite către Results:

Performance Vector: Metricile generale ale modelului.

Confusion Matrix: Performanța pe fiecare clasă (Low, Medium, High).

Cross Validation arata asa:



Rulam procesul si obtinem urmatoarii **indici de performanță**:

	true High	true Low	true Medium	class precision
pred. High	32	41	28	31.68%
pred. Low	357	1127	797	49.41%
pred. Medium	814	1280	1574	42.91%
class recall	2.66%	46.04%	65.61%	

Precision (Precizia pe clasă):

High: 31.68% – Din toate predicțiile pentru clasa High, doar 31.68% sunt corecte.

Low: 49.41% – Aproape jumătate din predicțiile pentru clasa Low sunt corecte.

Medium: 42.91% – Modelul confundă frecvent clasa Medium cu celelalte clase.

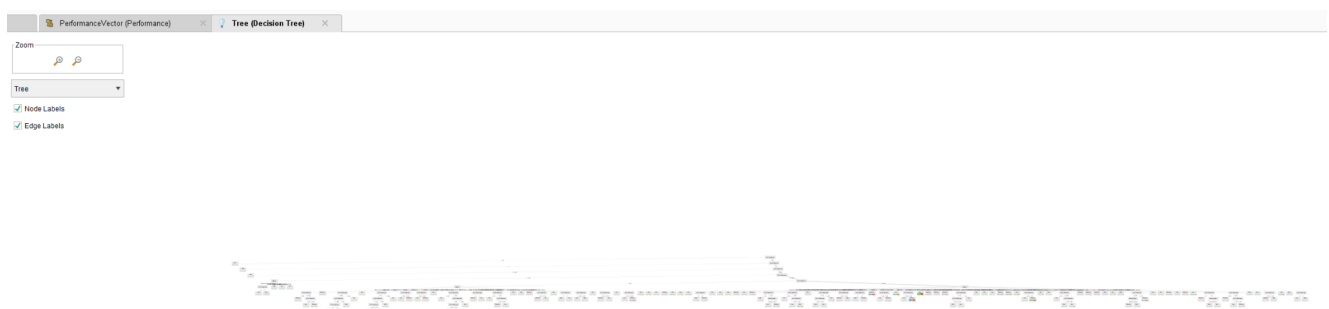
Recall (Rata de identificare a clasei):

High: 2.66% – Modelul detectează extrem de puține exemple din clasa High.

Low: 46.04% – Aproximativ jumătate din exemplele reale Low sunt detectate corect.

Medium: 65.61% – Modelul recunoaște bine exemplele din clasa Medium.

Asa arata arborele:



Pentru a imbunatati performanta modelului putem face urmatoarele modificari:

Modificarea parametrilor arborelui de decizie:

- Maximum Depth (Adâncimea maximă):

Reducerea adâncimii arborelui poate preveni overfitting (modelul care "memorizează" datele în loc să generalizeze). Un arbore prea adânc tinde să captureze zgomotul din date și să performeze mai prost pe setul de testare.

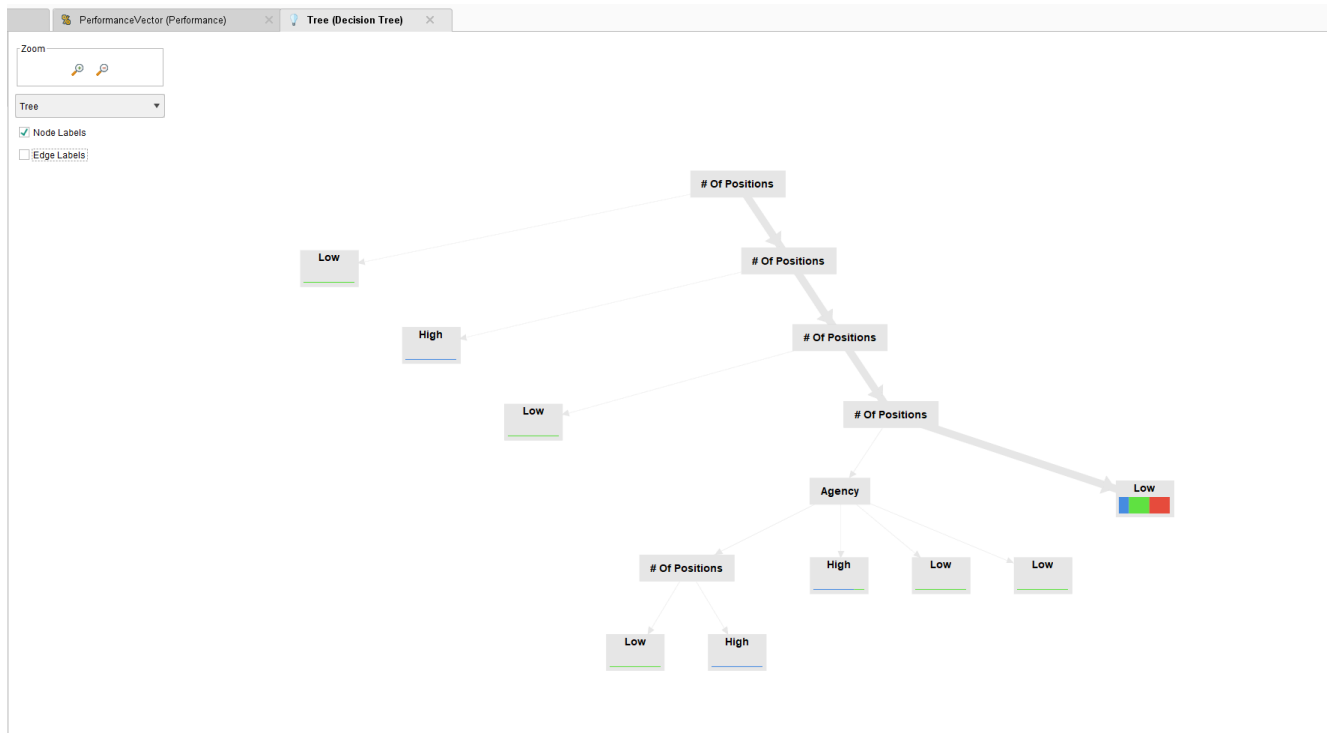
Creșterea dimensiunii minime a frunzelor obligă arborele să creeze frunze care acoperă mai multe exemple. Astfel, modelul devine mai general.

- Minimal Gain (Câștig minim):

Creșterea acestui parametru va reduce numărul de divizări în arbore. Dacă câștigul de informație este prea mic, divizarea nu se va realiza.

Test 1

Am modificat Minimal Gain - de la 0.01 la 0.5. Am obtinut predictii echilibrate pentru cele 3 clase.



Parameters

Decision Tree

criterion

gain ratio

maximal depth

10

☒ apply pruning

confidence

0.1

☒ apply prepruning

minimal gain

0.05

minimal leaf size

2

minimal size for split

4

number of prepruning alternatives

3

PerformanceVector (Performance)

Tree (Decision Tree)

Table View

Plot View

accuracy: 40.30% +/- 0.27% (micro average: 40.30%)

	true High	true Low	true Medium	class precision
pred High	4	5	0	40.00%
pred Low	839	1715	1680	40.51%
pred Medium	360	727	719	39.81%
class recall	0.33%	70.06%	29.97%	

PerformanceVector (Performance)

Tree (Decision Tree)

Table View

Plot View

accuracy: 45.17% +/- 3.64% (micro average: 45.17%)

	true High	true Low	true Medium	class precision
pred High	32	41	28	31.68%
pred Low	357	1127	797	49.41%
pred Medium	814	1280	1574	42.91%
class recall	2.66%	40.94%	65.61%	

Modelul anterior:

Performanța generală a fost mai echilibrată între clase, cu un recall mai mare pentru clasa Medium. A avut o acuratețe globală mai bună (45.17%).

Modelul actual:

Îmbunătățește precizia pentru clasa High (de la 31.68% la 40%).

Identifică mult mai bine exemplele din clasa Low (recall de 70.06%).

Însă suferă în ceea ce privește performanța pe clasa Medium (recall scăzut la 29.97%).

Acuratețea totală este mai slabă (40.30%).

III. Clustering in Altair

Pasul 1.

Articolul "Explainable k-means clustering: Theory and practice" (Dasgupta, Sanjoy, et al., 2020)

<http://interpretable-ml.org/icml2020workshop/pdf/06.pdf> analizează problema lipsei de

interpretabilitate a algoritmilor tradiționali de clustering, precum k-means. Aceste metode grupează datele pe baza distanțelor față de centrele clusterelor, dar funcționează ca o „cutie neagră”, fără a oferi explicații clare despre logica grupării, limitând astfel utilizarea lor în decizii practice.

Autorii propun algoritmul **Explainable k-Means Clustering (ExKMC)**, care combină avantajele k-means cu interpretabilitatea, generând clustere explicate prin reguli simple și intuitive, similare cu arborii de decizie. De exemplu:

- „Joburile din clusterul 1 au salariu < 100k și aparțin categoriei **Finance & Administration**.”
- „Joburile din clusterul 2 au salariu > 80k și necesita nivel de educație **Master**.”

Această metodă transformă clustering-ul într-un proces transparent, facilitând aplicarea în domenii precum analiza pieței muncii sau luarea deciziilor administrative, unde interpretabilitatea este esențială.

Pasul 2.

Aplicarea algoritmului Explainable k-Means Clustering (ExKMC) pe setul de date NYC_Jobs, în Altair nu este direct posibil deoarece ExKMC este un algoritm specific, descris în literatura științifică, și nu este integrat nativ în Altair.

Deoarece **ExKMC combină k-means și arbori de decizie**, putem simula acest proces printr-un workflow structurat.

În cadrul acestui proiect, utilizăm clustering-ul pentru a grupa joburile din setul de date în funcție de asemănări relevante între diferite atribute. Obiectivul este să identificăm tipare ascunse și să segmentăm datele în grupuri (clustere) semnificative, care să ofere informații utile despre structura și distribuția joburilor.

Următorii pași în Altair:

1. Selectăm atributele relevante din setul de date.
2. Aplicăm un algoritm de clustering k-means pentru a genera centrele.
3. Construim un arbore de decizie explicativ care să detalieze clusterelor generate.
4. Verificăm și ajustăm interpretabilitatea și precizia modelului.

Selectăm atributele relevante din setul de date.

- **Salary Range To:**
Reprezintă principalele variabile numerice care diferențiază joburile pe baza salariului.
Pot fi utilizate pentru a identifica clustere de joburi cu salarii mici, medii sau mari.
- **Job Category:**
Ajută la identificarea grupurilor de joburi similare (e.g., Social & Health Services, Technical Services, Finance & Administration, Public Safety & Legal si Policy & Research).
- **Posting Type:**
Indică dacă jobul este Internal sau External, ceea ce poate influența modul în care sunt grupate joburile.
Variabilă nominală cu două valori (binevenită în clustering).
- **# Of Positions:**
Variabilă numerică care indică numărul de poziții disponibile pentru un job.

Este utilă pentru a diferenția joburile individuale de cele cu multiple poziții.

- **Career level**
Ne arata nivelul de experienta necesar.
- **Minimum qualification requirements**
Putem extrage nivelul de educatie formala necesar.

data_attribute_selectate_pt_ExKMC

Add new data sets on the left. Details for the selected data are shown below. You can change the data with the following actions. ⓘ

TRANSFORM CLEANSE GENERATE PIVOT MERGE

MODEL CHARTS CREATE PROCESS HISTORY

Posting Type	Job Category	Career Level	Minimum Qual Requirements	# Of Positions	Salary Range To
External	Technology, Data & Innovation	Experienced (non-manager)	1) A baccalaureate degree from an accredi...	1	120190
Internal	Health Building Operations & Maintenance	Experienced (non-manager)	Two years of full-time satisfactory experience...	1	71614
Internal	Legal Affairs	Experienced (non-manager)	1. Admission to the New York State Bar; and ...	1	117541
Internal	Engineering, Architecture, & Planning	Experienced (non-manager)	1. A baccalaureate degree from an accredite...	1	123150
External	Constituent Services & Community Programs Social Services	Experienced (non-manager)	1. A baccalaureate degree from an accredite...	1	69462
External	Health Building Operations & Maintenance	Experienced (non-manager)	Two years of full-time satisfactory experience...	1	71614
External	Health	Entry Level	1. A baccalaureate degree from an accredite...	15	67983
External	Legal Affairs	Experienced (non-manager)	1. Admission to the New York State Bar; and ...	1	87763
Internal	Constituent Services & Community Programs Health	Experienced (non-manager)	Qualification Requirements 1. A Baccalaure...	2	56625
External	Constituent Services & Community Programs Health Policy, Research & Analysis	Experienced (non-manager)	High school diploma or its equivalent and: 1...	1	50000
Internal	Legal Affairs	Experienced (non-manager)	1. Admission to the New York State Bar; and ...	1	87763
Internal	Health	Experienced (non-manager)	1. A baccalaureate degree from an accredite...	1	70000
Internal	Constituent Services & Community Programs Social Services	Experienced (non-manager)	1. A baccalaureate degree from an accredite...	1	69462
External	Public Safety, Inspections, & Enforcement	Experienced (non-manager)	1. A baccalaureate degree from an accredite...	1	111321
Internal	Constituent Services & Community Programs	Experienced (non-manager)	Qualification Requirements 1. High school ...	1	65249
Internal	Public Safety, Inspections, & Enforcement	Experienced (non-manager)	1. A baccalaureate degree from an accredite...	1	111321

Vom prelucra datele.

A.) Analizand coloana Job Category vom crea coloana Job_Domain, in care avem ca valori domenii largi de activitate:

ExampleSet (Filter Examples)

Open in Turbo Prep Auto Model Interactive Analysis

Row No.	Job Category	count(Job Category)
1	Administration & Human Resources	314
2	Administration & Human Resources Social Services	150
3	Building Operations & Maintenance	307
4	Communications & Intergovernmental Affairs	89
5	Constituent Services & Community Programs	269
6	Constituent Services & Community Programs Health	84
7	Constituent Services & Community Programs Health Policy, Research & Analysis	291
8	Constituent Services & Community Programs Social Services	61
9	Engineering, Architecture, & Planning	729
10	Engineering, Architecture, & Planning Public Safety, Inspections, & Enforcement	150
11	Finance, Accounting, & Procurement	355
12	Finance, Accounting, & Procurement Policy, Research & Analysis	158
13	Health	477
14	Health Policy, Research & Analysis	90
15	Health Public Safety, Inspections, & Enforcement	63
16	Legal Affairs	353
17	Legal Affairs Public Safety, Inspections, & Enforcement	54
18	Policy, Research & Analysis	145
19	Public Safety, Inspections, & Enforcement	253
20	Social Services	349
21	Technology, Data & Innovation	546

Grupam astfel:

1. Social & Health Services

- Administration & Human Resources Social Services
- Constituent Services & Community Programs
- Constituent Services & Community Programs Health
- Constituent Services & Community Programs Health Policy, Research & Analysis
- Constituent Services & Community Programs Social Services
- Health
- Health Policy, Research & Analysis

- Social Services

2. Technical Services

- Engineering, Architecture, & Planning
- Engineering, Architecture, & Planning Public Safety, Inspections, & Enforcement
- Technology, Data & Innovation

3. Finance & Administration

- Administration & Human Resources
- Building Operations & Maintenance
- Communications & Intergovernmental Affairs
- Finance, Accounting, & Procurement
- Finance, Accounting, & Procurement Policy, Research & Analysis

4. Public Safety & Legal

- Public Safety, Inspections, & Enforcement
- Health Public Safety, Inspections, & Enforcement
- Legal Affairs
- Legal Affairs Public Safety, Inspections, & Enforcement

5. Policy & Research

- Policy, Research & Analysis

In python facem acest lucru folosind codul:

```
import pandas as pd

def categorize_job_domain(job_category):
    if any(keyword in job_category for keyword in [
        "Administration & Human Resources Social Services",
        "Constituent Services & Community Programs",
        "Constituent Services & Community Programs Health",
        "Constituent Services & Community Programs Health Policy, Research & Analysis",
        "Constituent Services & Community Programs Social Services",
        "Health",
        "Health Policy, Research & Analysis",
        "Social Services"
    ]):
        return "Social & Health Services"

    if any(keyword in job_category for keyword in [
        "Engineering, Architecture, & Planning",
        "Engineering, Architecture, & Planning Public Safety, Inspections, & Enforcement",
        "Technology, Data & Innovation"
    ]):
        return "Technical Services"

    if any(keyword in job_category for keyword in [
        "Administration & Human Resources",
        "Building Operations & Maintenance",
        "Communications & Intergovernmental Affairs",
        "Finance, Accounting, & Procurement",
        "Finance, Accounting, & Procurement Policy, Research & Analysis"
    ]):
        return "Finance & Administration"

    if any(keyword in job_category for keyword in [
        "Public Safety, Inspections, & Enforcement",
        "Health Public Safety, Inspections, & Enforcement",
        "Legal Affairs",
        "Legal Affairs Public Safety, Inspections, & Enforcement"
    ]):
        return "Public Safety & Legal"
```

```

    if "Policy, Research & Analysis" in job_category:
        return "Policy & Research"

    return "Other"

# Citire fișier CSV
input_file = r"C:\\Users\\Daci\\Downloads\\NYC_Jobs.csv (1)\\data_attribute_selectate_pt_ExKMC.csv"
try:
    data = pd.read_csv(input_file, encoding='utf-8', sep=';', on_bad_lines='skip')
except pd.errors.ParserError as e:
    print(f"Eroare la citirea fișierului CSV: {e}")
    raise

# Curățare nume coloane
data.columns = data.columns.str.strip()

# Verificare nume coloane
print("Coloanele disponibile în fișierul CSV sunt:", data.columns)

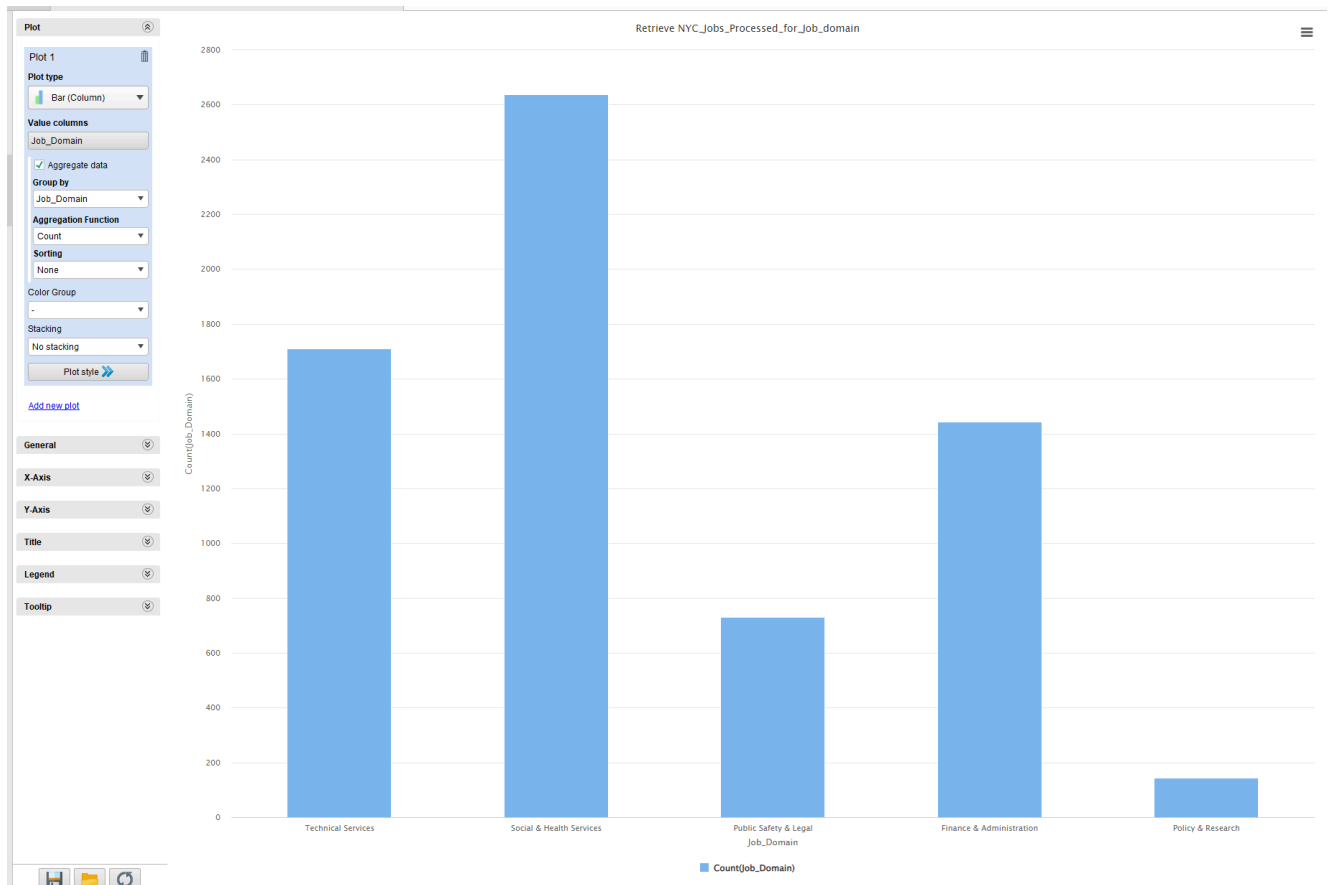
# Adăugare coloană nouă
if 'Job Category' in data.columns:
    data['Job_Domain'] = data['Job Category'].apply(categorize_job_domain)
else:
    raise ValueError("Coloana 'Job Category' nu există în fișierul de intrare.")

# Salvare fișier rezultat
output_file = r"C:\\Users\\Daci\\Downloads\\NYC_Jobs_Processed.csv"
data.to_csv(output_file, index=False, encoding='utf-8')

print(f"Fișier procesat salvat la: {output_file}")

```

Obținem:



B.) Creăm o coloană nouă, Salary Level, cu valorile „mic”, „mediu” si „mare” cu Generate Atributes. Salary Level lua valorile date de urmatoarea expresie:

```
if([Salary Range To] <= 80000, "Low",
  if([Salary Range To] <= 120000, "Medium", "High"))
```

C.) Avem nevoie de coloana Education. Rulam urmatorul cod python:

```
import pandas as pd
import numpy as np

def extract_education_level(qualification):
    qualification = qualification.lower()
    if any(keyword in qualification for keyword in ["phd", "doctorate", "juris doctor", "admission to the bar"]):
        return "Doctorat"
    elif any(keyword in qualification for keyword in ["master's degree", "mba"]):
        return "Master"
    elif any(keyword in qualification for keyword in ["bachelor", "baccalaureate", "registered nurse", "architect", "engineer"]):
        return "Bachelor"
    elif any(keyword in qualification for keyword in ["high school", "diploma", "ged"]):
        return "High school diploma"
    else:
        return "No formal education specified"

# Citire fișier CSV
input_file = r"C:\\Users\\Daci\\Downloads\\NYC_Jobs.csv (1)\\data_attribute_selectate_pt_ExKMC_ed_level.csv"
try:
    data = pd.read_csv(input_file, encoding='utf-8', sep=';', on_bad_lines='skip')
```

```

except pd.errors.ParserError as e:
    print(f"Eroare la citirea fișierului CSV: {e}")
    raise

# Curățare nume coloane
data.columns = data.columns.str.strip()

# Verificare nume coloane
print("Coloanele disponibile în fișierul CSV sunt:", data.columns)

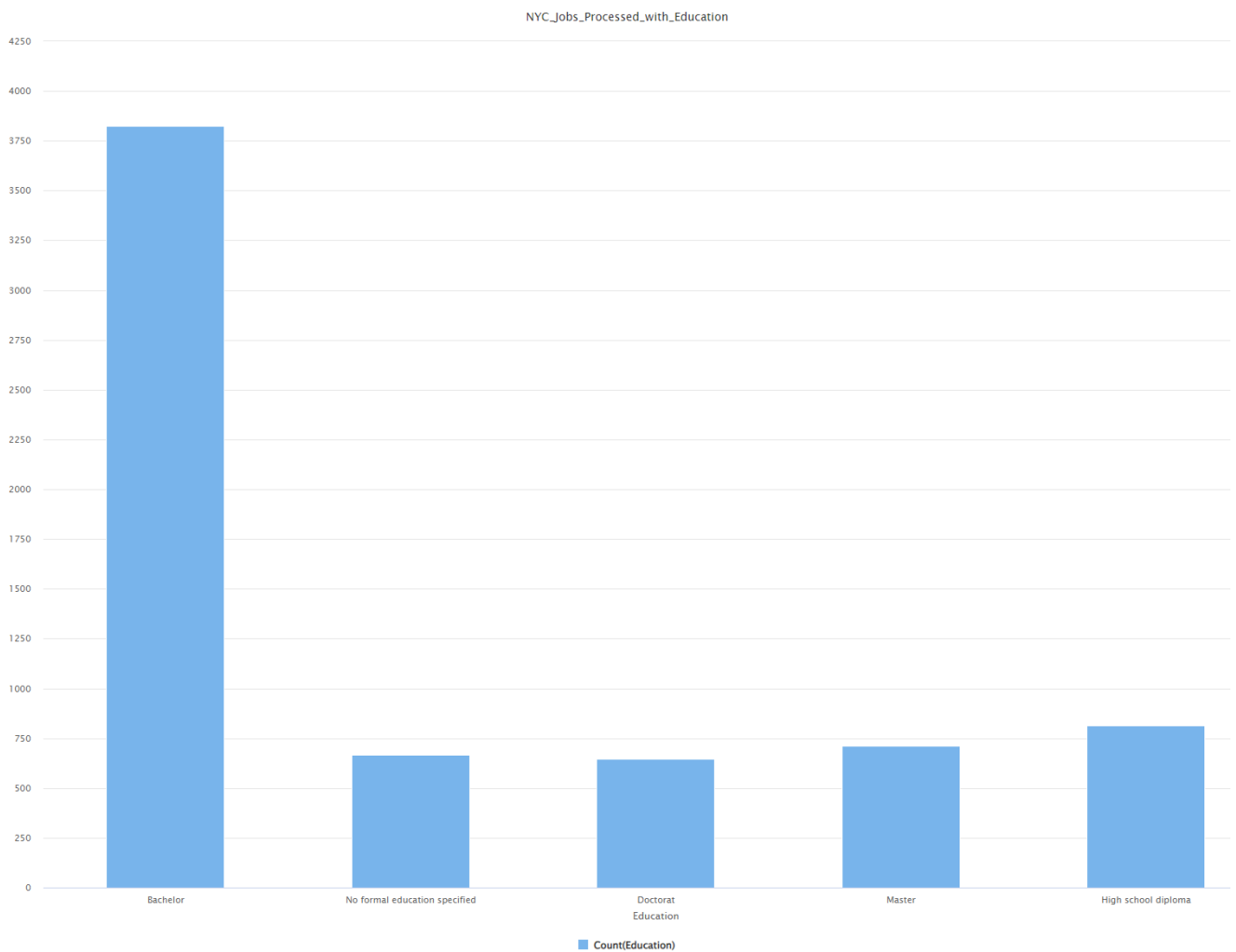
# Adăugare coloană Nivel Educație
if 'Minimum Qual Requirements' in data.columns:
    data['Education'] = data['Minimum Qual Requirements'].fillna('').apply(extract_education_level)
else:
    raise ValueError("Coloana 'Minimum Qual Requirements' nu există în fișierul de intrare.")

# Salvare fișier rezultat
output_file = r"C:\\Users\\Daci\\Downloads\\NYC_Jobs_Processed_with_Education.csv"
data.to_csv(output_file, index=False, encoding='utf-8')

print(f"Fișier procesat salvat la: {output_file}")

```

Obținem:



Asa arata datele pe care vom aplica algoritmul **ExKMC** ce combină **k-means** și **arbori de decizie**

ExampleSet (Select Attributes) X

Open in

Turbo Prep

Auto Model

Interactive Analysis

Row No.	Posting Type	Career Level	# Of Positions	Job_Domain	Salary Level	Education
1	External	Experienced (non-manager)	1	Technical Services	High	Bachelor
2	Internal	Experienced (non-manager)	1	Social & Health Services	Low	No formal education specified
3	Internal	Experienced (non-manager)	1	Public Safety & Legal	Medium	Doctorat
4	Internal	Experienced (non-manager)	1	Technical Services	High	Master
5	External	Experienced (non-manager)	1	Social & Health Services	Low	Bachelor
6	External	Experienced (non-manager)	1	Social & Health Services	Low	No formal education specified
7	External	Entry-Level	15	Social & Health Services	Low	Bachelor
8	External	Experienced (non-manager)	1	Public Safety & Legal	Medium	Doctorat
9	Internal	Experienced (non-manager)	2	Social & Health Services	Low	Master
10	External	Experienced (non-manager)	1	Social & Health Services	Low	High school diploma
11	Internal	Experienced (non-manager)	1	Public Safety & Legal	Medium	Doctorat
12	Internal	Experienced (non-manager)	1	Social & Health Services	Low	Bachelor
13	Internal	Experienced (non-manager)	1	Social & Health Services	Low	Bachelor
14	External	Experienced (non-manager)	1	Public Safety & Legal	Medium	Bachelor
15	Internal	Experienced (non-manager)	1	Social & Health Services	Low	High school diploma
16	Internal	Experienced (non-manager)	1	Public Safety & Legal	Medium	Bachelor
17	Internal	Experienced (non-manager)	1	Technical Services	High	Bachelor
18	External	Experienced (non-manager)	1	Social & Health Services	Low	Bachelor
19	Internal	Experienced (non-manager)	1	Social & Health Services	Low	Bachelor
20	External	Experienced (non-manager)	1	Social & Health Services	Low	Bachelor
21	Internal	Experienced (non-manager)	1	Technical Services	High	Bachelor
22	External	Experienced (non-manager)	1	Technical Services	High	Bachelor
23	Internal	Experienced (non-manager)	1	Social & Health Services	Low	High school diploma
24	Internal	Entry-Level	15	Social & Health Services	Low	Bachelor
25	External	Experienced (non-manager)	1	Technical Services	High	Master
26	External	Experienced (non-manager)	2	Social & Health Services	Low	Master
27	External	Experienced (non-manager)	1	Public Safety & Legal	Medium	Doctorat
28	Internal	Experienced (non-manager)	1	Social & Health Services	Medium	Doctorat
29	External	Experienced (non-manager)	1	Social & Health Services	Medium	Doctorat
30	External	Experienced (non-manager)	1	Social & Health Services	Medium	Doctorat

Pregatirea datelor pentru aplicarea k-means

Datele noastre arata astfel:

Posting Type - Tip date Nominal Valori: **Internal, External**

Career level - Tip date Nominal Valori: **Experienced (non-manager), Entry-Level, Manager, Student, Executive**

Of Positions - Tip date Real

Job_Domain - Tip date Nominal Valori: **Social & Health Services, Technical Services, Finance & Administration, Public Safety & Legal, Policy & Research**

Salary Level - Tip date Nominal Valori: **Low, Medium, High**

Education - Tip date Nominal Valori: **Bachelor, High school diploma, Master, No formal education specified, Doctorat**

În Altair, folosim aceste metode pentru a converti datele nominale in numerice:

a. Codificare ordinală (Ordinal Encoding)

Aceasta presupune atribuirea unui număr fiecărei categorii.

Este potrivita pentru datele categorice care au o ordine clară, cum ar fi Career Level sau Salary Level.

Exemplu: Career Level:

Student → 0

Entry-Level → 1

Experienced (non-manager) → 2

Manager → 3

Executive → 4

b. Codificare one-hot (One-Hot Encoding)

Aceasta creează mai multe coloane binare (0 sau 1) pentru fiecare categorie.

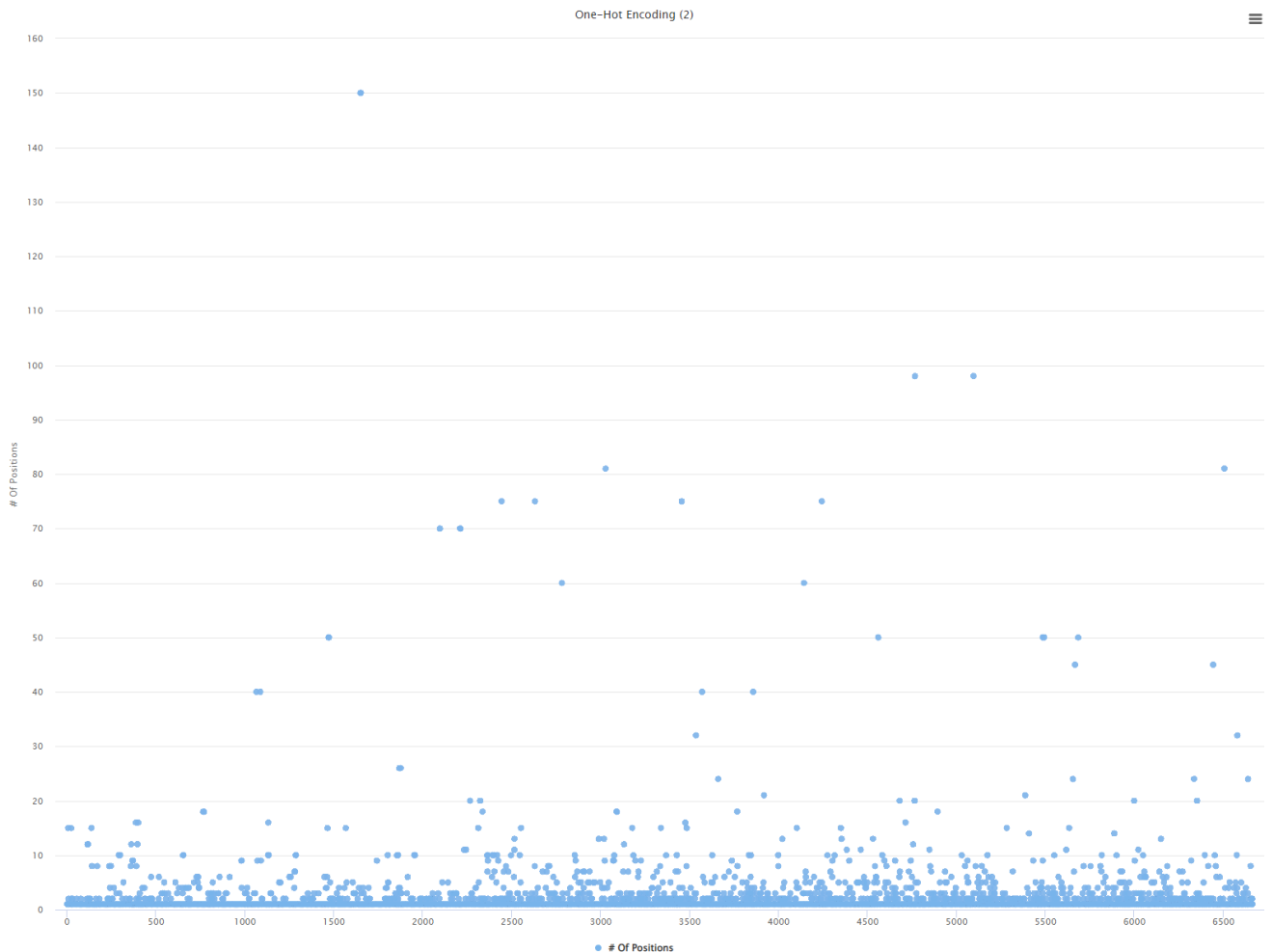
Este utilă pentru date fără ordine, cum ar fi Job Domain sau Education.

Exemplu: Job Domain:

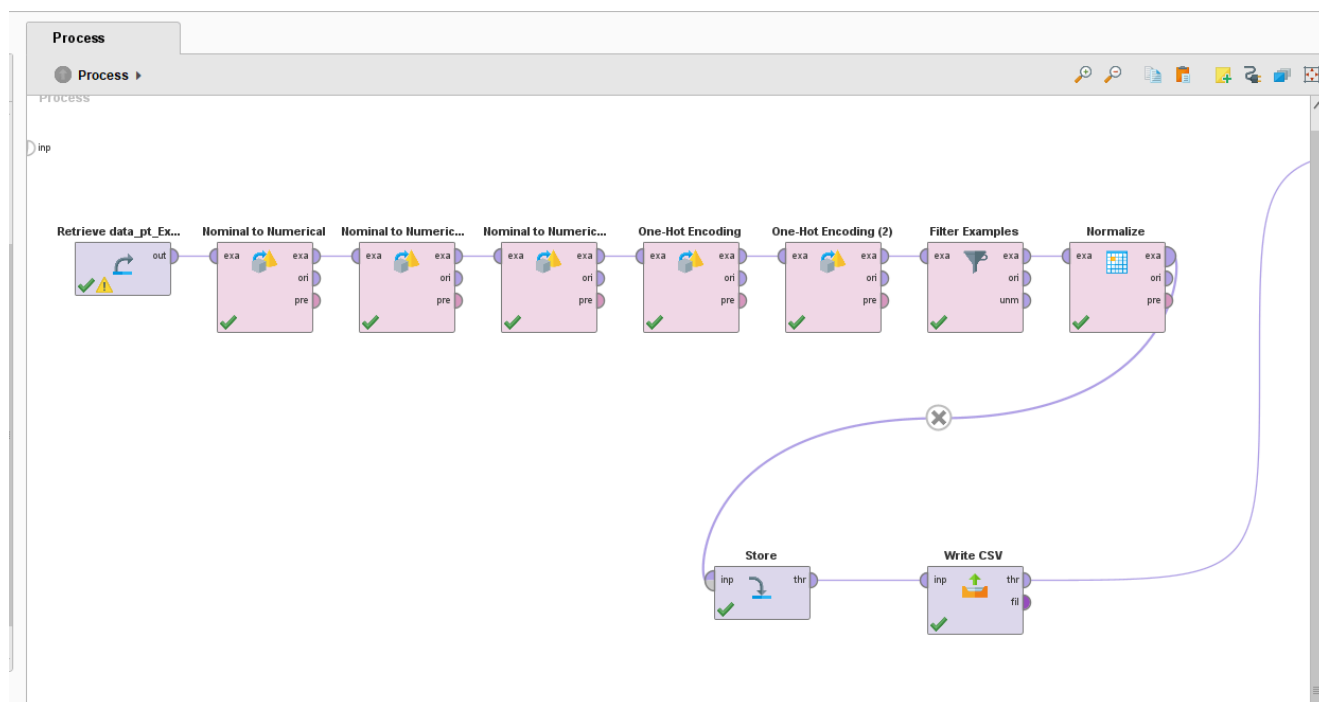
Social & Health Services → [1, 0, 0, 0, 0]

Technical Services → [0, 1, 0, 0, 0]

Coloana # Of Positions este deja numerică, dar **trebuie scalată (normalizată)**. Folosim Z-Score Standardization care reduce impactul valorilor extreme (min val 1.0 max val 150), ceea ce ajută K-Means să creeze clusteruri mai uniforme și mai relevante.



Vom elimina val de 150 deoarece e prea mare fata de restul val si poate distorsiona clusteringul. Asa arata procesul in Altair.



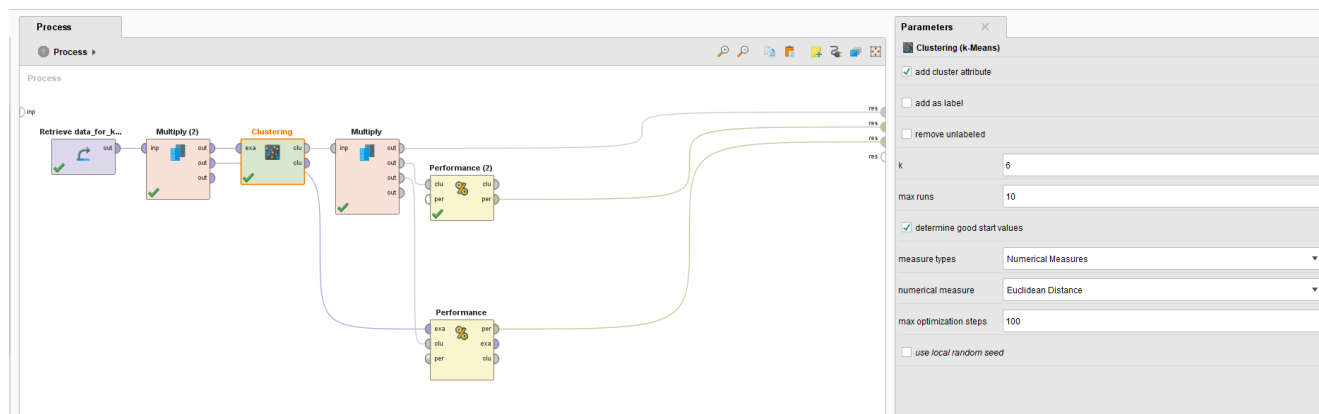
Asa arata datele dupa codificare

ExampleSet (Normalize)												
Open in Turbo Prep Auto Model Interactive Analysis												
Row No.	# Of Positions	Job_Domain...	Job_Domain...	Job_Domain...	Job_Domain...	Education = ...	Education = ...	Education = ...	Education = ...	Salary Level	Posting Type	Career Level
1	-0.199	1	0	0	0	1	0	0	0	0	0	0
2	-0.199	0	1	0	0	0	1	0	0	1	1	0
3	-0.199	0	0	1	0	0	0	0	0	2	1	0
4	-0.199	1	0	0	0	0	0	1	0	0	1	0
5	-0.199	0	1	0	0	1	0	0	0	1	0	0
6	-0.199	0	1	0	0	0	1	0	0	1	0	0
7	3.416	0	1	0	0	1	0	0	0	1	0	1
8	-0.199	0	0	1	0	0	0	0	0	2	0	0
9	0.060	0	1	0	0	0	0	1	0	1	1	0
10	-0.199	0	1	0	0	0	0	0	1	1	0	0
11	-0.199	0	0	1	0	0	0	0	0	2	1	0
12	-0.199	0	1	0	0	1	0	0	0	1	1	0
13	-0.199	0	1	0	0	1	0	0	0	1	1	0
14	-0.199	0	0	1	0	1	0	0	0	2	0	0
15	-0.199	0	1	0	0	0	0	0	1	1	1	0
16	-0.199	0	0	1	0	1	0	0	0	2	1	0
17	-0.199	1	0	0	0	1	0	0	0	0	1	0
18	-0.199	0	1	0	0	1	0	0	0	1	0	0
19	-0.199	0	1	0	0	1	0	0	0	1	1	0
20	-0.199	0	1	0	0	1	0	0	0	1	0	0
21	-0.199	1	0	0	0	1	0	0	0	0	1	0
22	-0.199	1	0	0	0	1	0	0	0	0	0	0
23	-0.199	0	1	0	0	0	0	0	1	1	1	0
24	3.416	0	1	0	0	1	0	0	0	1	1	1
25	-0.199	1	0	0	0	0	0	1	0	0	0	0
26	0.060	0	1	0	0	0	0	1	0	1	0	0
27	-0.199	0	0	1	0	0	0	0	0	2	0	0
28	-0.199	0	1	0	0	0	0	0	0	2	1	0

Operatorul One-Hot Encoding funcționează astfel în Altair Studio AI:

Atributul nominal original va fi eliminat, iar în locul său vor fi adăugate un set de noi attribute binare. Va exista o coloană nouă pentru fiecare valoare nominală posibilă a coloanei originale, cu excepția uneia (așa-numitul grup de comparație).

Aplicam k means:



Folosim 2 operatori de performanta:

Cluster Distance Performance si Item Distribution Performance

Testam pt

k=6,8,10,15

1. Avg. Within Centroid Distance (Compactitatea clusterelor)

Aceasta măsoară cât de aproape sunt punctele din fiecare cluster de centroidul lor:

Valorile mai mici (în valoare absolută) indică o compactitate mai mare (clusterare mai bine grupate).

La valori mai mari de k, valorile medii tind să scadă (clusterarele devin mai mici și mai compacte).

Interpretare:

Pe măsură ce k crește, clusterarele devin mai mici și mai compacte (ceea ce este de așteptat).

k=6 are cea mai mare valoare absolută, indicând clusterare mai mari și mai puțin compacte, ceea ce poate reflecta o grupare mai generală.

Compactitatea crește semnificativ la k=15, dar acest lucru poate duce la supraîmpărțirea datelor.

2. Davies-Bouldin Index (Separarea și coeziunea clusterelor)

Aceasta măsoară raportul dintre distanțele dintre clusterare și dimensiunile lor:

Valorile mai mici indică clusterare bine separate și compacte.

Interpretare:

k=8 are cel mai mic Davies-Bouldin Index (-1.466), indicând o separare mai bună a clusterelor.

La k=10, valoarea crește (devine mai puțin negativă), sugerând că separarea și coeziunea clusterelor se deteriorează.

k=6 și k=15 au valori similare, dar mai slabe decât k=8.

3. Example Distribution (Uniformitatea distribuției datelor între clusterare)

Aceasta măsoară cât de bine sunt distribuite exemplele între clusterare:

Valorile apropiate de 1 indică o distribuție uniformă.

Valorile apropiate de 0 indică o distribuție foarte dezechilibrată.

Interpretare:

k=6 are cea mai uniformă distribuție a datelor între clusterare.

Distribuția devine din ce în ce mai dezechilibrată pe măsură ce crește numărul de clusterare.

La k=15, unele clusterare pot fi foarte mici sau subreprezentate, indicând supraîmpărțirea datelor.

Concluzii generale:

1. Alegerea k depinde de scopul analizei:

- k=6: Clusterare mai generale, bine distribuite (uniformitate bună). Recomandat dacă scopul este o segmentare de nivel înalt.

- k=8: Oferă un echilibru bun între compactitate, separare și distribuție. Este cel mai recomandat din perspectiva Davies-Bouldin Index.
- k=10: Compactitate crescută, dar separarea clusterelor se deteriorează (Davies-Bouldin Index mai slab).
- k=15: Compactitate maximă, dar distribuția datelor este foarte dezechilibrată, ceea ce poate duce la clusterle inutile sau redundante.

2. Recomandare optimă: k=8

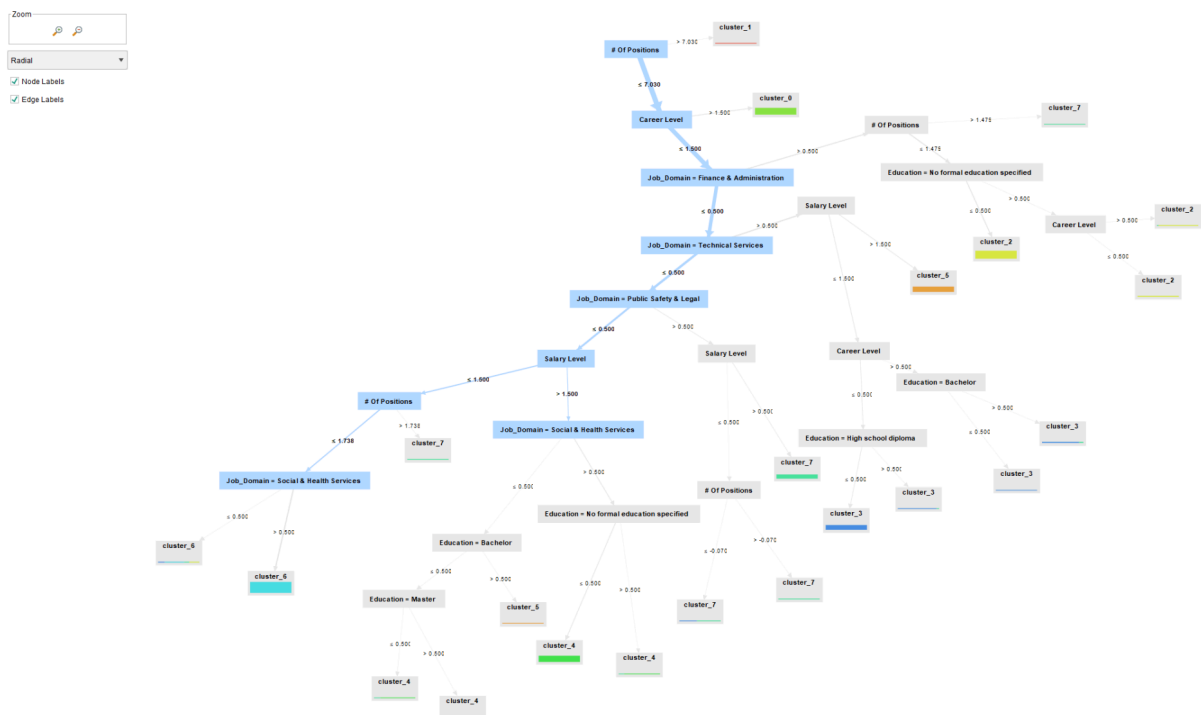
Separare bună a clusterelor (Davies-Bouldin Index scăzut).

Distribuție relativ uniformă și compactitate echilibrată.

Aplicam arborii de decizie:



Detalii complete despre fiecare cluster, așa cum rezultă din arborele de decizie:



Cluster 1 (# Of Positions > 7.030)

- Caracteristici:
Număr mare de poziții disponibile: peste 7.030.
Exclusiv 24 exemple.
- Interpretare:
Acest cluster reprezintă joburi cu cerere extrem de mare.
Probabil joburi generale sau care necesită forță de muncă mare (ex.: retail, logistică).

Cluster 0 (Career Level > 1.500 și # Of Positions ≤ 7.030)

- Caracteristici:
Joburi avansate, cu nivel de carieră peste 1.500.
Include 893 exemple.
Domeniile joburilor și alte caracteristici nu sunt specificate în arbore pentru acest cluster.
- Interpretare:
Roluri de management sau senior-level, cerând experiență și calificări superioare.
Poate include poziții din diverse domenii, dar clar distinse prin nivelul ridicat de carieră.

Cluster 7 (Job_Domain = Finance & Administration și Career Level ≤ 1.500)

- Segment 1 (# Of Positions > 1.479):
Include 26 exemple.
Joburi financiare sau administrative cu cerere moderată (peste 1.479 poziții disponibile).
- Segment 2 (# Of Positions ≤ 1.479):
Nu este direct detaliat în arbore.
- Interpretare:
Joburi în domeniul financiar sau administrativ, probabil la nivel intermediar.

Cluster 2 (Job_Domain = Finance & Administration, Career Level ≤ 1.500, # Of Positions ≤ 1.479)

- Subdiviziuni pe educație:
Fără educație formală: 1062 exemple.
Majoritatea joburilor din acest cluster.
- Alt nivel de educație:
Educația și nivelul de carieră influențează joburile (48 exemple cu Career Level > 0.500 și educație limitată).
- Interpretare:
Roluri de bază sau entry-level în domeniul financiar sau administrativ.

Cluster 5 (Job_Domain = Technical Services și Salary Level > 1.500)

- Caracteristici:
Include 690 exemple.
Joburi tehnice cu salarii mari.
- Interpretare:
Roluri bine plătite în domeniul tehnic (ex.: inginerie, IT, proiectare).
Clusterul sugerează un segment premium al pieței muncii.

Cluster 3 (Job_Domain = Technical Services și Salary Level ≤ 1.500)

- Subdiviziuni pe educație:
Diplomă de licență > 0.500: 70 exemple.
Joburi tehnice de nivel mediu.
Diplomă de liceu: 612 exemple.
Joburi tehnice entry-level.
- Interpretare:
Clusterul conține joburi mai accesibile în domeniul tehnic, cu salarii mai mici comparativ cu Cluster 5.

Cluster 7 (Job_Domain = Public Safety & Legal, Career Level ≤ 1.500, Salary Level > 0.500)

- Caracteristici:
Include 529 exemple.
Joburi în domeniul public (ex.: poliție, asistență juridică), cu salarii peste medie.
- Interpretare:

Clusterul reflectă poziții cu responsabilități ridicate în domeniul public.

Cluster 4 (Job_Domain = Social & Health Services, Salary Level > 1.500)

- Subdiviziuni pe educație:
Educație fără specificare formală:
Include 811 exemple.
Alt nivel de educație:
Roluri mai bine definite, 26 exemple cu educație înaltă.
- Interpretare:
Joburi din domeniul sănătății cu salarii mari.
Clusterul reflectă un mix de specialiști și roluri mai generale.

Cluster 6 (Job_Domain = Social & Health Services, Salary Level ≤ 1.500)

- Subdiviziuni:
Majoritatea joburilor sunt entry-level sau intermediare, cu salarii mai mici.
Include 1495 exemple.
- Interpretare:
Joburi diverse, cu cerințe mai puțin stricte, reflectând probabil o gamă largă de roluri.
Cluster 7 (subclustere suplimentare pentru Job_Domain = Public Safety & Legal și # Of Positions > -0.070)
- Caracteristici:
Roluri care combină salarii medii și cerințe moderate pentru numărul de poziții.
Poate include joburi de suport sau operaționale.

Concluziile statistice despre joburile din sectorul public din orasul New York, așa cum reies din analiza setului de date, pot fi valoroase pentru mai multe categorii de persoane și organizații:

1. Viitorii angajați și studenți

Cluster 1: Joburi cu cerere foarte mare (>7.030 poziții).

Recomandare: Alegerea unor domenii generale (ex.: logistică, retail) pentru oportunități rapide de angajare.

Cluster 0: Joburi avansate, senior-level (893 poziții).

Recomandare: Investiți în studii superioare și experiență pentru roluri de management sau specializare avansată.

Cluster 7 (Finanțe & Administrație): Joburi cu cerere moderată/intermediară.

Recomandare: Studiați economie, contabilitate sau administrație publică pentru oportunități în domeniu.

Cluster 2: Joburi entry-level în finanțe/administrativ (1062 poziții fără cerințe educaționale).

Recomandare: Accesibile imediat, dar avansarea necesită educație formală suplimentară.

Cluster 5: Joburi tehnice bine plătite (690 poziții).

Recomandare: Optați pentru educație STEM (IT, inginerie) pentru salarii competitive și cerere crescută.

Cluster 3: Joburi tehnice entry-level (612 poziții cu liceu, 70 cu licență).

Recomandare: Începeți cu calificări tehnice de bază, apoi avansați prin programe de licență sau certificări.

Cluster 7 (Siguranță Publică & Legal): Joburi în siguranță publică (529 poziții).

Recomandare: Alegeți domenii precum dreptul sau serviciile publice pentru stabilitate și salarii bune.

Cluster 4: Joburi bine plătite în sănătate (811 fără cerințe, 26 specialiști).

Recomandare: Profitați de oportunități entry-level, iar specializarea medicală poate aduce câștiguri semnificative.

Cluster 6: Joburi entry-level în sănătate (1495 poziții).

Recomandare: Oportunități accesibile imediat; dezvoltarea profesională prin programe de formare continuă.

Studentii și viitorii angajați pot alege domeniile și nivelurile educaționale în funcție de aspirațiile lor, orientându-se spre clusterelor cu cerințe și beneficii compatibile cu planurile lor de carieră.

2. Organizațiile guvernamentale și agențiile publice

Cluster 1: Joburi cu cerere foarte mare (>7.030 poziții).

Acțiuni: Campanii de recrutare masive și parteneriate pentru formare rapidă.

Cluster 0: Joburi avansate, senior-level (893 poziții).

Acțiuni: Pregătirea internă pentru promovarea angajaților și atragerea de talente senior.

Cluster 7: Joburi în Finanțe & Administrație (cerere moderată/intermediară).

Acțiuni: Traineeship-uri pentru tineri absolvenți și digitalizarea proceselor repetitive.

Cluster 2: Joburi entry-level în finanțe/administrativ (1062 poziții fără cerințe educaționale).

Acțiuni: Accesibilitate crescută pentru comunități defavorizate.

Cluster 5: Joburi tehnice bine plătite (690 poziții).

Acțiuni: Sprijinirea educației STEM și programe de relocare pentru talente.

Cluster 3: Joburi tehnice entry-level (612 poziții cu liceu, 70 cu licență).

Acțiuni: Parteneriate cu licee tehnice și programe de certificare.

Cluster 7: Joburi în Siguranță Publică & Legal (529 poziții, salarii peste medie).

Acțiuni: Recrutare activă și programe de formare continuă.

Cluster 4: Joburi bine plătite în sănătate (811 fără cerințe, 26 specialiști).

Acțiuni: Investiții în educația medicală și retenția talentelor.

Cluster 6: Joburi entry-level în sănătate (1495 poziții).

Acțiuni: Diversitate în angajare și instruire pentru avansare.

Aceste date sprijină recrutarea, formarea și optimizarea forței de muncă publice.

IV. Vizualizarea datelor

Exportam fișierul cu datele preprocesate în Altair Studio Ai.



În python vom adauga coloana Education:

```
import pandas as pd
import numpy as np

def extract_education_level(qualification):
    qualification = qualification.lower()
    if any(keyword in qualification for keyword in ["phd", "doctorate", "juris doctor", "admission to the bar"]):
        return "Doctorat"
    elif any(keyword in qualification for keyword in ["master's degree", "mba"]):
        return "Master"
    elif any(keyword in qualification for keyword in ["bachelor", "baccalaureate", "registered nurse", "architect", "engineer"]):
        return "Bachelor"
    elif any(keyword in qualification for keyword in ["high school", "diploma", "ged"]):
        return "High school diploma"
    else:
        return "No formal education specified"

# Calea fișierului de intrare
input_file = r"C:\Users\Daci\Downloads\NYC_Jobs.csv (1)\nyc_jobs_data_for_powerbi.csv"

# Detectarea separatorului corect
try:
    # Detectăm automat separatorul
    data = pd.read_csv(input_file, encoding='utf-8', sep=None, engine='python', on_bad_lines='skip')
    print("Separator detectat automat. Coloanele disponibile în fișierul CSV sunt:")
    print(data.columns)
except pd.errors.ParserError as e:
    print(f"Eroare la citirea fișierului CSV: {e}")
    raise

# Curățare nume coloane
data.columns = data.columns.str.strip()
```



```

# Adăugare coloană Nivel Educație
if 'Minimum Qual Requirements' in data.columns:
    data['Education'] = data['Minimum Qual Requirements'].fillna('').apply(extract_education_level)
else:
    raise ValueError("Coloana 'Minimum Qual Requirements' nu există în fișierul de intrare.")

# Calea fișierului de ieșire
output_file = r"C:\Users\Daci\Downloads\NYC_Jobs.csv (1)\nyc_jobs_data_for_powerbi_add_education.csv"

# Salvare fișier rezultat cu separator definit explicit
data.to_csv(output_file, index=False, encoding='utf-8', sep=',')
print(f"Fișier procesat salvat la: {output_file}")

```

In python vom adauga si coloana Job_Domain.

```

import pandas as pd

def categorize_job_domain(job_category):
    if any(keyword in job_category for keyword in [
        "Administration & Human Resources Social Services",
        "Constituent Services & Community Programs",
        "Constituent Services & Community Programs Health",
        "Constituent Services & Community Programs Health Policy, Research & Analysis",
        "Constituent Services & Community Programs Social Services",
        "Health",
        "Health Policy, Research & Analysis",
        "Social Services"
    ]):
        return "Social & Health Services"

    if any(keyword in job_category for keyword in [
        "Engineering, Architecture, & Planning",
        "Engineering, Architecture, & Planning Public Safety, Inspections, & Enforcement",
        "Technology, Data & Innovation"
    ]):
        return "Technical Services"

    if any(keyword in job_category for keyword in [
        "Administration & Human Resources",
        "Building Operations & Maintenance",
        "Communications & Intergovernmental Affairs",
        "Finance, Accounting, & Procurement",
        "Finance, Accounting, & Procurement Policy, Research & Analysis"
    ]):
        return "Finance & Administration"

    if any(keyword in job_category for keyword in [
        "Public Safety, Inspections, & Enforcement",
        "Health Public Safety, Inspections, & Enforcement",
        "Legal Affairs",
        "Legal Affairs Public Safety, Inspections, & Enforcement"
    ]):
        return "Public Safety & Legal"

    if "Policy, Research & Analysis" in job_category:
        return "Policy & Research"

    return "Other"

```

```

# Calea fișierului de intrare
input_file = r"C:\Users\Daci\Downloads\NYC_Jobs.csv (1)\nyc_jobs_data_for_powerbi_add_education.csv"

try:
    # Detectarea separatorului corect
    data = pd.read_csv(input_file, encoding='utf-8', sep=None, engine='python', on_bad_lines='skip')
    print("Separator detectat automat. Coloanele disponibile în fișierul CSV sunt:")
    print(data.columns)
except pd.errors.ParserError as e:
    print(f"Eroare la citirea fișierului CSV: {e}")
    raise

# Curățare nume coloane
data.columns = data.columns.str.strip()

# Adăugare coloană nouă
if 'Job Category' in data.columns:
    data['Job_Domain'] = data['Job Category'].apply(categorize_job_domain)
else:
    raise ValueError("Coloana 'Job Category' nu există în fișierul de intrare.")

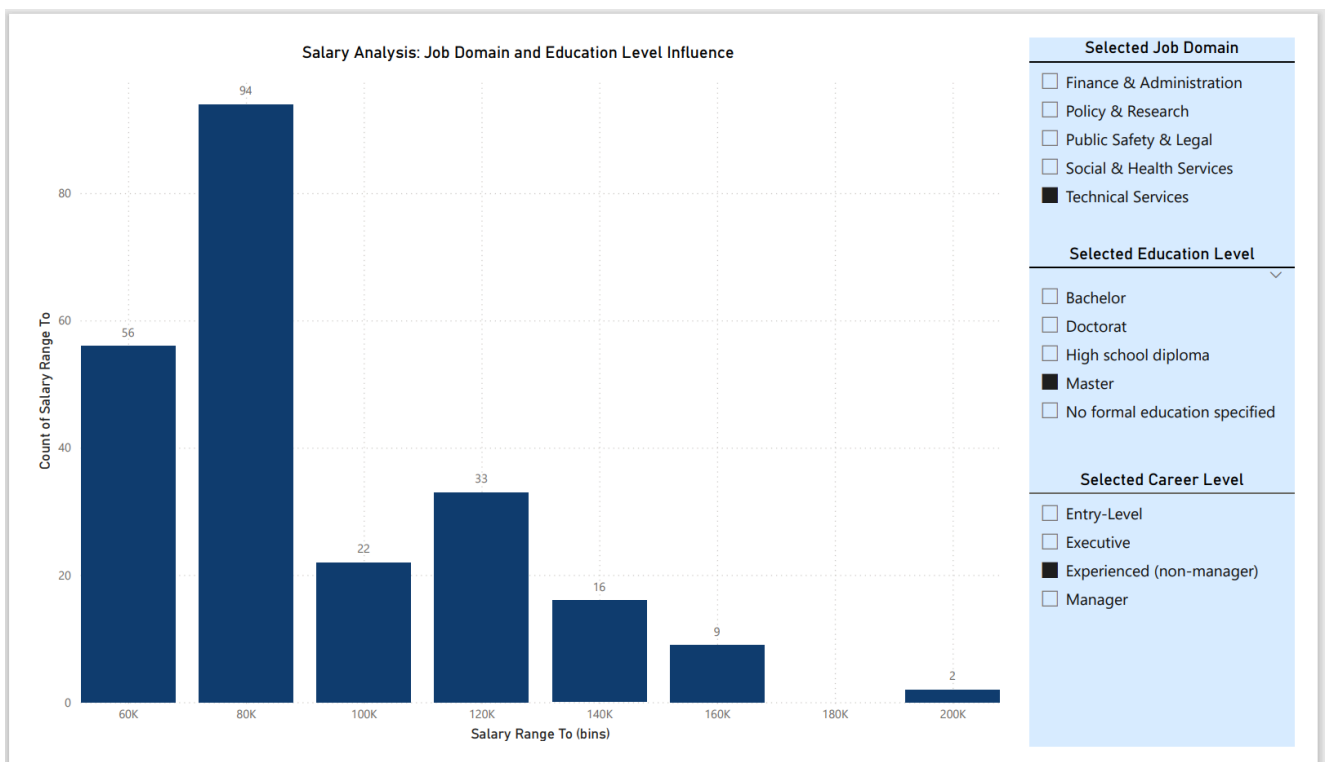
# Calea fișierului de ieșire
output_file = r"C:\Users\Daci\Downloads\NYC_Jobs.csv (1)\nyc_jobs_data_for_powerbi_add_education_job_domain.csv"

# Salvare fișier rezultat cu separator definit explicit
data.to_csv(output_file, index=False, encoding='utf-8', sep=',')
print(f"Fișier procesat salvat la: {output_file}")

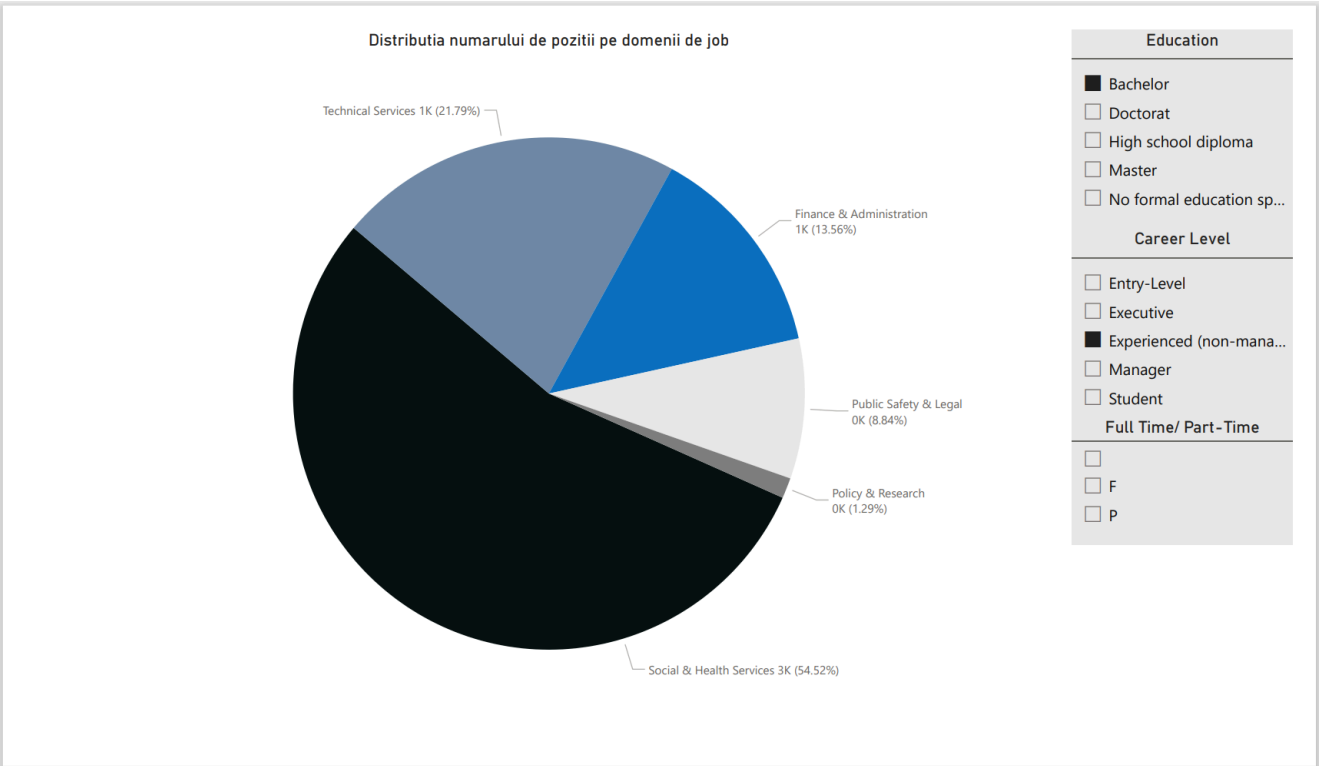
```

Vizualizări:

1. Distribuția salariilor in functie de nivelul de cariera, nivelul educatiei si domeniul.

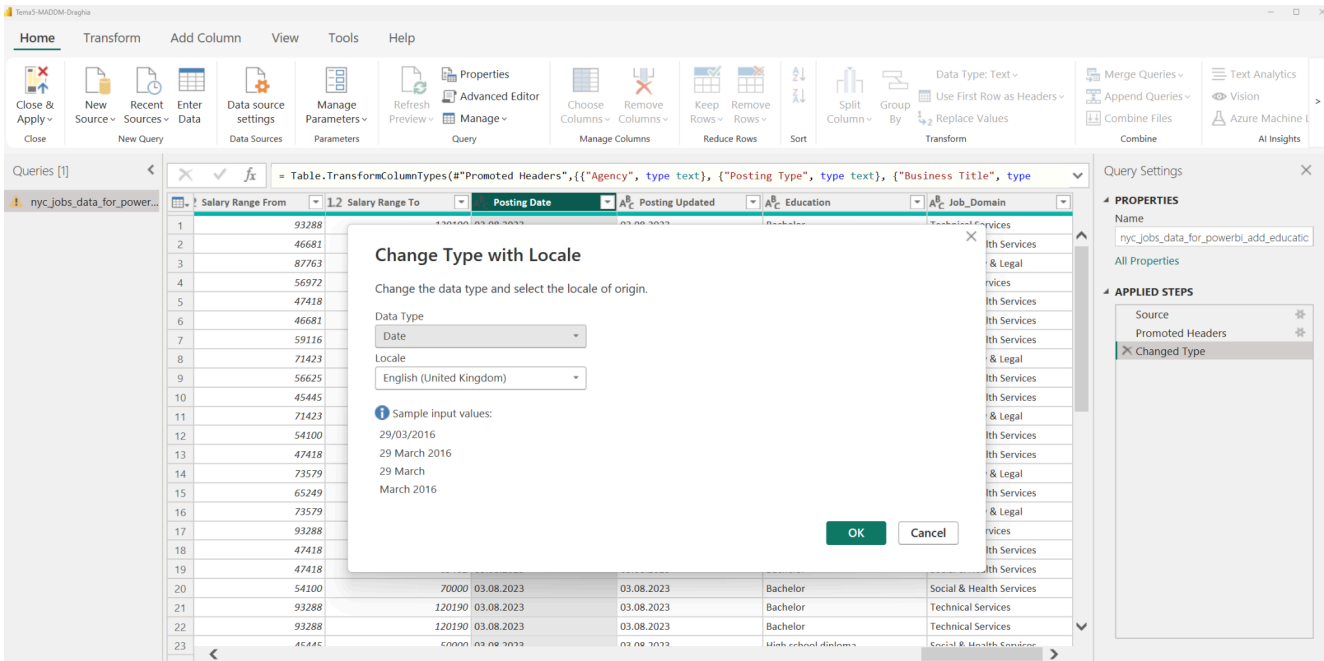


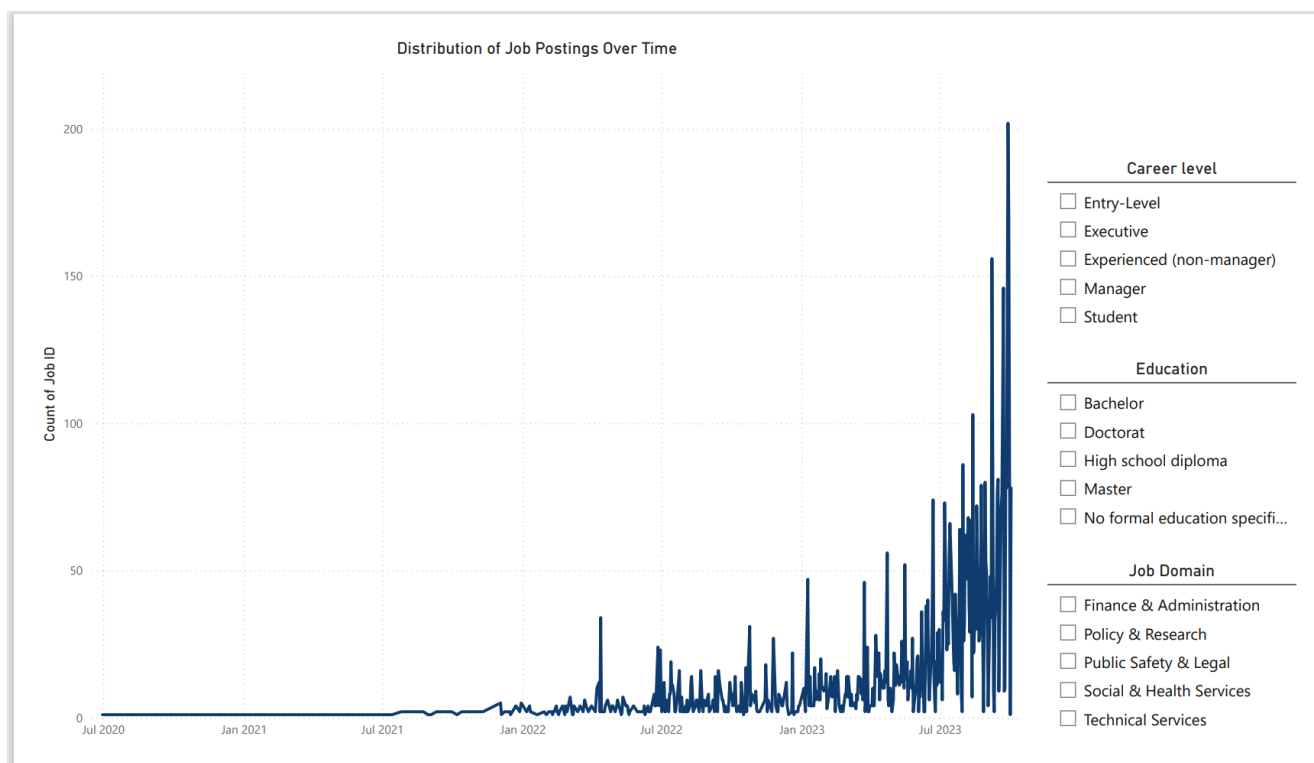
2. Diagramă circulară pentru a arăta distribuția totală a numărului de poziții (# Of Positions) pe categorii de job (Job_Domain).



3. Tendințe în postările de joburi

Grafic liniar pentru a evidenția numărul de joburi postate în timp (Posting Date).
Convertim coloanele de tip Date care au fost identificate ca si text in tipul Date.





4. Rezidență necesară

Grafic de tip pie chart pentru a evidenția cerințele de rezidență (Residency Requirement).

Analizăm datele din coloana Residency Requirement și observăm că deși avem 96 de formulare distincte pentru Residency Requirement, ele reflectă doar 6 categorii distincte:

1. No Residency Requirement – Nu există cerințe de rezidență.
2. Residency Required in 90 Days – Rezidență necesară în NYC în termen de 90 de zile.
3. Flexible Residency Requirements – Rezidență permisă în alte locații după 2 ani.
4. Residency in New York State Required – Rezidență necesară în New York State.
5. NYC Residency Required – Rezidență obligatorie în New York City.
6. Temporary COVID-19 Waiver – Cerințe flexibile din cauza pandemiei COVID-19.

În Power BI, accesăm Transform Data pentru a deschide Power Query.

Selectăm coloana Residency Requirement, Add Column, Custom Column.

Introducem formula pentru noua coloană Residency.

```
= if Text.Contains([Residency Requirement], "no residency requirement",
Comparer.OrdinalIgnoreCase) or
Text.Contains([Residency Requirement], "no residency required", Comparer.OrdinalIgnoreCase)
or
Text.Contains([Residency Requirement], "not required", Comparer.OrdinalIgnoreCase) then "No
Residency Requirement"
else if Text.Contains([Residency Requirement], "90 days", Comparer.OrdinalIgnoreCase) then
"Residency Required in 90 Days"
else if Text.Contains([Residency Requirement], "New York State", Comparer.OrdinalIgnoreCase)
then "Residency in New York State Required"
else if Text.Contains([Residency Requirement], "New York City", Comparer.OrdinalIgnoreCase) and
not Text.Contains([Residency Requirement], "90 days", Comparer.OrdinalIgnoreCase) then
```

"NYC Residency Required"

else if Text.Contains([Residency Requirement], "2 years", Comparer.OrdinalIgnoreCase) or
Text.Contains([Residency Requirement], "two continuous years",

Comparer.OrdinalIgnoreCase) then "Flexible Residency Requirements"

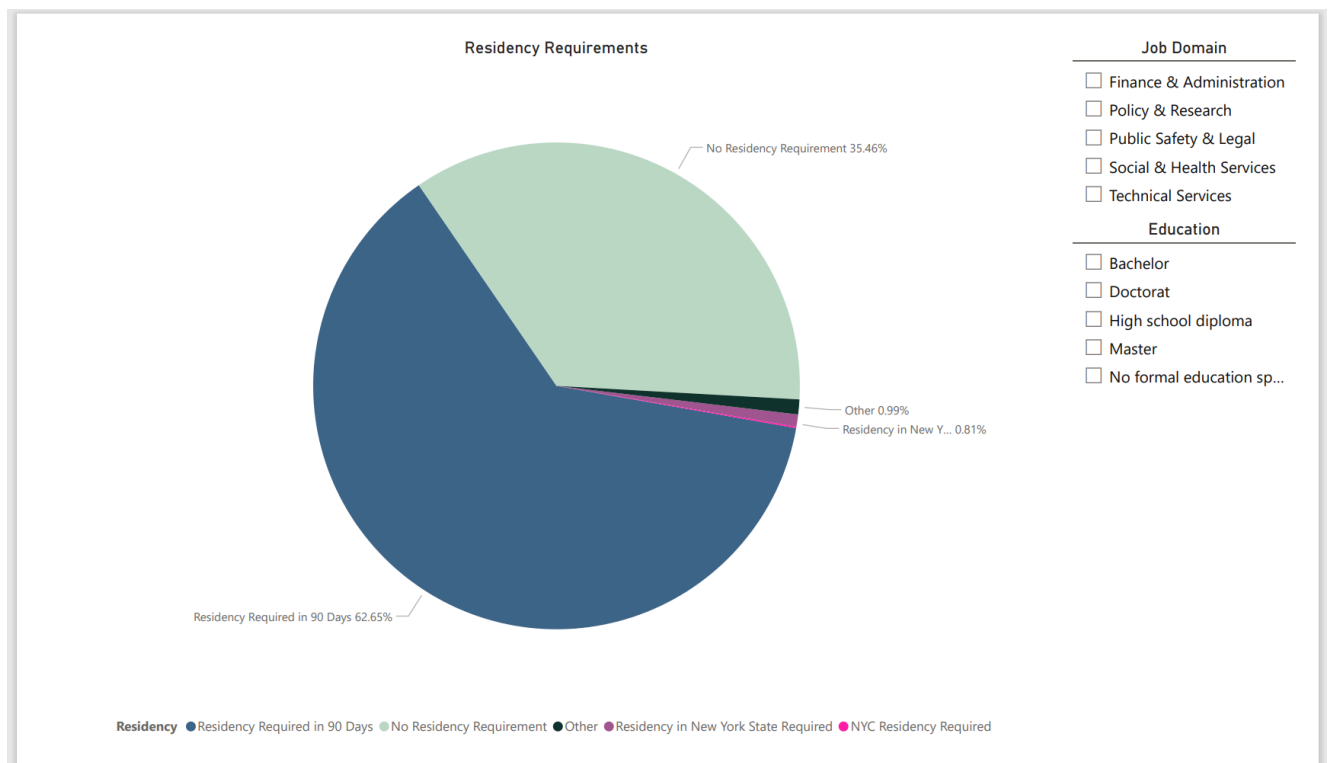
else if Text.Contains([Residency Requirement], "COVID-19", Comparer.OrdinalIgnoreCase) or

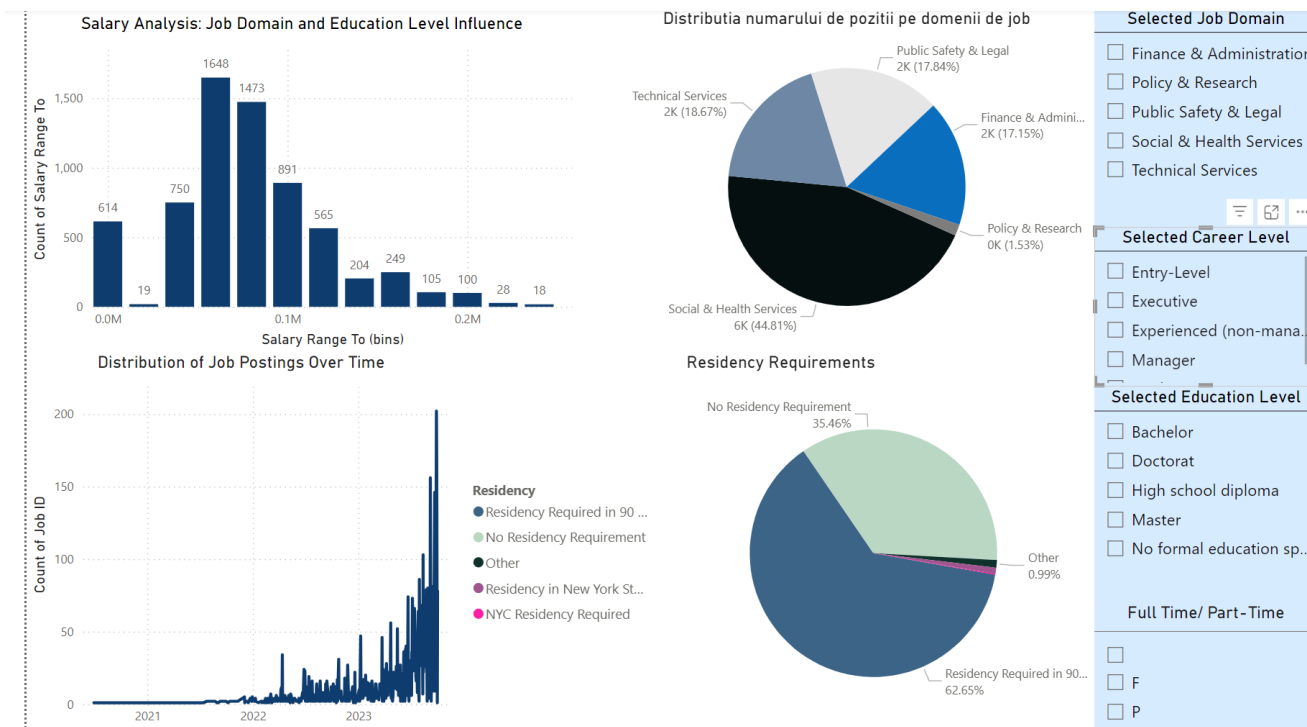
Text.Contains([Residency Requirement], "state of emergency", Comparer.OrdinalIgnoreCase)

then "Temporary COVID-19 Waiver"

else "Other"

Table.AddColumn("Changed Type with Locale2", "Residency", each if Text.Contains([Residency Requirement], "no residency", Comparer.OrdinalIgnoreCase) then "No Residency Requirement" else if Text.Contains([Residency Requirement], "2 years", Comparer.OrdinalIgnoreCase) or Text.Contains([Residency Requirement], "two continuous years", Comparer.OrdinalIgnoreCase) then "Flexible Residency Requirements" else if Text.Contains([Residency Requirement], "COVID-19", Comparer.OrdinalIgnoreCase) or Text.Contains([Residency Requirement], "state of emergency", Comparer.OrdinalIgnoreCase) then "Temporary COVID-19 Waiver" else "Other"						
Id	Salary Range To	Posting Date	Posting Updated	Education	Job_Domain	Residency
28	97138	22/09/2023	22/09/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
29	97138	22/09/2023	22/09/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
30	107032	02/10/2023	02/10/2023	Doctorat	Social & Health Services	No Residency Requirement
31	91768	02/10/2023	02/10/2023	Bachelor	Social & Health Services	Residency Required in 90 Days
32	208826	02/10/2023	02/10/2023	Bachelor	Finance & Administration	No Residency Requirement
33	103548	02/10/2023	02/10/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
34	99694	02/10/2023	02/10/2023	Bachelor	Social & Health Services	No Residency Requirement
35	100000	04/05/2023	06/09/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
36	121296	02/10/2023	02/10/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
37	66054	02/10/2023	02/10/2023	Bachelor	Social & Health Services	Residency Required in 90 Days
38	66054	02/10/2023	02/10/2023	Bachelor	Social & Health Services	Residency Required in 90 Days
39	100000	04/05/2023	06/09/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
40	107032	02/10/2023	02/10/2023	Doctorat	Social & Health Services	No Residency Requirement
41	117541	02/10/2023	02/10/2023	Doctorat	Social & Health Services	No Residency Requirement
42	46104	02/10/2023	02/10/2023	No formal education specified	Social & Health Services	Residency Required in 90 Days
43	117541	02/10/2023	02/10/2023	Doctorat	Social & Health Services	No Residency Requirement
44	121296	02/10/2023	02/10/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
45	103548	02/10/2023	02/10/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
46	107032	02/10/2023	02/10/2023	Doctorat	Social & Health Services	No Residency Requirement
47	69709	02/10/2023	02/10/2023	High school diploma	Social & Health Services	Residency Required in 90 Days
48	208826	02/10/2023	02/10/2023	Bachelor	Finance & Administration	No Residency Requirement
49	103548	02/10/2023	02/10/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
50	103548	02/10/2023	02/10/2023	Doctorat	Social & Health Services	Residency Required in 90 Days
51						





Ce observații despre setul vostru de date în graficele realizate?

Analiza nivelelor salariale:

- Cele mai multe salarii sunt în intervalul 40-60k(1648 salarii).
- În intervalul 60-80k sunt 1473 de salarii.
- Cele mai puține se afla în intervalul 220-240k(18 salarii).
- Dacă selectăm domeniul tehnic, nivel de educație - master și career level - entry observăm că salariile sunt în intervalul 80k.

Distributia joburilor postate în timp:

- Începând cu luna iulie 2023 numărul joburilor postate a crescut substanțial față de perioada precedentă, în 28 septembrie s-au postat 200 de noi joburi.

Distributia joburilor pe domenii:

- Cele mai multe joburi sunt în Social and Health Services 44,81%, iar cele mai puține în Policy and Research: 1,53%.
- Cele mai multe joburi ca manager se găsesc în domeniul tehnic: 29,71%.

Cerintele de rezidență:

- 62.65% din joburi cer candidaților să poată obține rezidență în următoarele 3 luni de la data angajării
- 35,46% din joburi specifică că nu e necesară rezidența