**Procesul de gandire (POC Veridion)**

Am inceput cu o abordare de tip fuzzy matching intre input\_company\_name si company\_name. A functionat decent, dar am realizat ca nu este suficient pentru a identifica entitatea corecta in toate cazurile, asa ca am creat un scor compus care include mai multi factori relevanti:

* Numele companiei
* Tara
* Orasul
* Website-ul
* LinkedIn-ul

Apoi m-am gandit ca nu este corect ca un camp necompletat sa penalizeze scorul, asa ca am modificat logica: daca input\_main\_country lipseste, nu penalizez scorul, ba chiar adaug puncte ca sa nu dezavantajez entitatea. (M-am gandit ca, daca toate campurile se potrivesc in afara de unul singur, entitatea va fi mult dezavantajata, desi cel mai probabil este un MATCH.)

Am adaugat si o coloana numita score\_justification pentru a documenta clar de ce fiecare companie a primit scorul respectiv (ex: "Fuzzy: 41.5p | Country: +20p (missing input)", etc).

**Verificarea calitatii datelor**

Apoi am vrut sa verific daca datele pe care le-am selectat ca "MATCHED" sunt si calitativ utile. M-am intrebat:

Cate companii au website? Cate au venituri sau coduri de industrie? Merita aceasta baza de date livrata clientului?

Am construit o analiza de Quality Control (QC) care masoara gradul de completitudine pentru fiecare camp relevant si genereaza un raport CSV.

Am tratat si cazurile in care coloanele nu exista (de exemplu, dupa merge(), unele coloane devin \*\_x sau \*\_y).

Cand am recitit cerinta, am observat si partea de documentatie si mi-am dat seama ca nu am citit initial instructiunile legate de API. Am vazut ca Veridion ofera un API care intoarce un scor de incredere si un ID unic pentru fiecare companie.

Totusi, nu am primit o cheie de autentificare si am presupus ca ideea challenge-ului este sa lucrez fara API, folosind doar fisierele puse la dispozitie.

**Ce am invatat din acest proiect:**

* Cum sa construiesc un scor personalizat de potrivire a entitatilor, adaptat la scenarii reale
* Cum sa evaluez datele nu doar pentru acuratete, ci si pentru utilitatea practica
* Sa citesc intotdeauna documentatia inainte