

AKTIVITETI 03

GJUHETARI I IDENTITETEVE TE DYFISHTA

Logjika e Deduplikimit me AI

MISIONI JUAJ

Identifikoni 15 cifte rekordesh duplikate te fshehura ne nje dataset me 200 rekorde qytetaresh shqiptare. Perdorni AI per fuzzy matching dhe vleresoni besueshmerine. **Koha:** 60 minuta

TIPET E VARIACIONEVE PER T'U KERKUAR

Tipi	Pershkrimi	Shembull
NAME_CASE	Kapitalizim i ndryshem	Agron -> AGRON
NAME TYPO	Gabim shkrimi	Gentiana -> Genntiana
NAME VARIANT	Variant emri	Gëzim -> Gezim
DATE_FORMAT	Format tjetër date	1987-03-15 -> 15/03/1987
ADDRESS_ABBREV	Adrese e shkurtuar	Rruga -> Rr.
PHONE_FORMAT	Format tjetër telefoni	+355 -> 0
EMAIL VARIANT	Email i ngjashëm	a.b@ -> a.b1@
CITY VARIANT	Variant qyteti	Tirane -> Tirana

HAPAT E USHTRIMIT

- Hapni A03_rekordet_qytetareve.csv (200 rekorde)
- Ngarkoni ne Claude/Gemini
- Perdorni prompt-in e deduplikimit
- Per cdo cift te gjetur, plotesoni tabelen e rezultateve
- Vleresoni besueshmerine: HIGH / MEDIUM / LOW

SHABLLON PROMPT-I

Kam dataset me 200 rekorde qytetaresh. Identifiko ciftet e mundshme duplikate duke krahasuar emrin, mbiemrin, datelindjen, adresen, telefonin, email-in. Konsidero variacione ne kapitalizim, gabime shkrimi, formate te ndryshme. Per cdo cift raporto: ID-te, fushat qe perputhen, fushat qe ndryshojne, besueshmerine (HIGH/MEDIUM/LOW), rekomandimin (BASHKO/VERIFIKO/INJORO).

TABELA E REZULTATEVE

Plotesoni per cdo cift duplikat te gjetur:

Nr.	Rekord 1	Rekord 2	Tipi Variacionit	Besueshmeria
1				
2				
3				
4				
5				
6				
7				
8				