

SQL Mini-Proiect: Wine Quality Dataset

Acest proiect este bazat pe setul de date „Wine Quality” disponibil pe UCI Machine Learning Repository. Scopul proiectului este să îți exersezi abilitățile SQL aplicate pe un set de date real, folosind comenzi de tip SELECT, WHERE, GROUP BY, HAVING, JOIN, CTE, subquery, agregări și funcții analitice.

Setul de date conține informații despre caracteristicile fizico-chimice și scorurile de calitate pentru diferite vinuri. Poți folosi un singur tabel (wine_data), dar în unele exerciții vom simula un model relațional cu tabele auxiliare.

1. Inspectarea datelor

Afișează primele 10 rânduri din tabelul wine_data. Ce coloane sunt disponibile?

2. Statistici de bază

Afișează valorile minime, maxime și medii pentru coloanele „alcohol”, „pH”, „residual sugar”.

3. Filtrare

Selectează toate vinurile cu alcohol > 12.5 și pH < 3.2.

4. Grupare și Agregare

Grupează vinurile după „quality” și calculează media pentru „alcohol”, „pH” și „density” pentru fiecare grup.

5. Condiții pe grupuri

Grupează după „quality”, dar afișează doar acele grupuri care au peste 50 de vinuri.

6. CTE + subquery

Creează un CTE care grupează vinurile după „quality” și calculează media alcoolului. Folosește apoi acest CTE pentru a selecta doar acele calități unde media alcoolului > 11.

7. Clasificare binară

Creează o coloană calculată care să eticheteze vinurile ca „High Quality” dacă quality >= 7, altfel „Low Quality”. Calculează apoi câte vinuri sunt în fiecare categorie.

8. Window Functions

Adaugă o coloană care să arate rangul fiecărui vin în funcție de „alcohol”, în ordine descrescătoare.

9. Simulare JOIN (cu tabel suplimentar)

Creează un tabel auxiliar numit wine_types.csv cu coloanele: - quality (int) - type (text: 'premium', 'standard', 'economy') Asociază acest tabel cu wine_data pe baza coloanei quality. Afișează pentru fiecare tip de vin: numărul de vinuri, media alcoolului și media acidității volatile.

10. Export & Vizualizare

Exportă rezultatul exercițiului 9 într-un fișier CSV sau folosește Pandas pentru vizualizare.