

1 Descrizione dati iniziali

Il dataset che abbiamo analizzato contiene dati sulle carriere accademiche degli studenti del corso di laurea di informatica dell'universit degli studi di Firenze e il loro voto conseguito al test di ingresso. In particolare, le informazioni presenti sono:

- Coorte: Anno di immatricolazione
- Crediti totali: Numero crediti complessivi dello studente
- Crediti con voto: Numero di crediti assegnati allo studente per esami con votazione in trentesimi(tutti tranne inglese)
- Voto medio: Media pesata dei voti degli esami sostenuti

In seguito ci sono sei coppie di attributi che indicano:

- Nome dell'esame
- Data in cui lo studente ha sostenuto l'esame

Gli esami sono Algoritmi e strutture dati (ASD), Architetture degli elaboratori(ARC), Programmazione(PRG), Analisi I(ANI), Matematica discreta e logica(MDL) e Inglese.

- Punteggio conseguito al test di ingresso.

Per quanto riguarda gli attributi relativi ai voti degli esami i valori che questi assumono sono: il voto conseguito dallo studente nel caso in cui abbia sostenuto l'esame; oppure zero se lo studente non ha sostenuto l'esame durante la sessione estiva del proprio anno di immatricolazione.

2 Gestione dei dati

Abbiamo deciso di effettuare le seguenti operazioni sul dataset:

- eliminazione degli studenti che hanno sostenuto solo inglese
- riportare tutti gli attributi relativi alle date degli esami nel formato YYYY-MM-DD

Il motivo per cui abbiamo deciso di eseguire l'operazione del primo punto che gli studenti in questione non avendo voti in trentesimi non ci sono pari significativi per analisi. Per quanto riguarda il secondo punto invece, nel caso in cui uno studente non avesse sostenuto un particolare esame, erano presenti valori pari a zero in alcuni casi e valori pari a "0000-00-00" in altri. Abbiamo quindi deciso di trattare questa incosistenza dei dati ponendo i valori pari a "0000-00-00".

Per effettuare queste due operazioni abbiamo importato il dataset in database tramite l'operazione di import riportata nel codice.

CODICE IMPORT

Per quanto riguarda la gestione dell'incosistenza relativa alle date degli esami abbiamo risolto il problema specificando le query riportate nel codice

```
update dmo.studenti set data_ARC = '0000-00-00' where data_ARC = '0';  
update dmo.studenti set data_ASD = '0000-00-00' where data_ASD = '0';  
update dmo.studenti set data_PRG = '0000-00-00' where data_PRG = '0';  
update dmo.studenti set data_ANI = '0000-00-00' where data_ANI = '0';  
update dmo.studenti set data_MDL = '0000-00-00' where data_MDL = '0';  
update dmo.studenti set data_INGLESE = '0000-00-00' where data_INGLES
```