

# Data Science e Tecnologie per le Basi di Dati

## Homework 1

Una società di produzione televisiva vuole analizzare alcune statistiche riguardanti un famoso programma di musica italiana. Nel corso degli ultimi 30 anni la società ha memorizzato in un database informazioni riguardanti il televoto dei propri telespettatori. Si supponga di utilizzare la base dati per realizzare un data warehouse ed analizzare i trend riguardanti l'attività dei telespettatori negli anni.

L'analisi ha come obiettivo quello di analizzare statistiche separatamente per i partecipanti che concorrono nel programma. Per ogni partecipante si memorizzano il nome, e la città di residenza (o fondazione per i gruppi). Si consideri il nome del partecipante come univoco nella base dati.

I voti in favore di ogni partecipante vengono conferiti dai telespettatori. Per i telespettatori sono note la fascia di età (18-25, 25-30, 30-50, >50), il sesso, la categoria ('giuria esperti' oppure 'pubblico') ed il luogo di residenza.

Il sistema memorizza la data e l'orario in cui vengono espresse le votazioni. L'orario viene registrato nel formato 'hh:mm' (ora, minuti). E' inoltre nota la modalità di voto (Telefono, Facebook, Instagram, Sito Web del programma). Ogni voto è associato ad un'edizione del programma (es: '2018', '2019', ...). Il programma si svolge nel mese di Febbraio. Siccome è possibile votare solo per l'edizione in corso, l'anno in cui viene espresso il voto corrisponde all'anno dell'edizione. Infine per ogni edizione è noto il nome del conduttore.

Le statistiche verranno effettuate sul numero di voti effettuati, sul voto medio assegnato e sugli incassi ottenuti dalla produzione tramite l'introduzione di pubblicità nelle applicazioni usate per il voto. Il valore di un voto è un numero compreso nell'intervallo 0-5. L'analisi deve essere condotta considerando le seguenti informazioni:

- Nome partecipante, città, provincia, regione
- Fascia età, sesso del telespettatore, città, provincia, regione
- Categoria telespettatore
- Data e giorno della settimana in cui è stato effettuato il voto
- Orario (es: 22.15), ora del giorno (1-24) in cui è stato effettuato il voto
- Modalità di voto (Facebook, Instagram, Telefono)
- Edizione (es: '2018', '2019', ...)
- Conduttore edizione

### Homework tasks

1. Progettare il data warehouse in modo da soddisfare le richieste descritte nelle specifiche del problema. Disegnare lo schema concettuale del datawarehouse e definire lo schema logico (tabelle dei fatti e delle dimensioni).
2. Esprimere le due interrogazioni seguenti utilizzando il linguaggio SQL esteso.
  - (a) Considerare i voti effettuati mediante la piattaforma 'Instagram'. Separatamente per edizione e città del telespettatore, calcolare:
    - i. il totale cumulativo degli incassi ottenuti durante le varie edizioni
    - ii. la percentuale di incassi apportati da telespettatori di una città rispetto al totale degli incassi
    - iii. l'incasso medio giornaliero
  - (b) Considerare le edizioni svolte tra il 2000 ed il 2010. Separatamente per edizione eseguire le seguenti analisi:
    - i. il numero totale di voti
    - ii. la percentuale del numero di voti di ogni partecipante rispetto al numero di voti totale
    - iii. assegnare un rank ai partecipanti per voto medio (0-5) decrescente