# Data Science And Database Technology
# Homework 1

A TV production company is interested in analyzing some statistics related to a famous Italian music contest. During the last 30 years, the company stored in a database information about televoting from its viewers. Suppose that you have to use the database for realizing a data warehouse and analyzing trends related to the viewers' activity during the last years.
The aim of the analysis consists in analyzing the statistics separately for the participants of the contest. For each participant you have to store his/her name, residence city (or foundation city for bands). Consider the participant name as unique inside the database.
Votes in favor of each participant are given by viewers. The age group (i.e. 18-25, 25-30, 30-50, >50), gender, category ('experts jury' or 'audience') and residence city are known for the viewers.
The designed system stores the date and time when votes are expressed. Time is encoded in the following format: 'hh:mm' (hour, minutes). Additionally, the televoting mode is known (i.e. Phone, Facebook, Instagram, TV program website).
Each vote is associated to a specific edition of the TV program (e.g. '2018', '2019', ...). The TV program is conducted in February. Since it is possible to vote only for the ongoing edition, the year when the vote is expressed corresponds to the year of the edition. Finally, for each edition the name of the presenter is known.
The statistics are made on the number of expressed votes, the average vote and the total incomes obtained by the production company by introducing advertisement in the applications used for voting. The value of a vote is a number in the interval 0-5. The analysis must be conducted considering the following information:

- Participant name, city, province, region
- Age group, gender of the viewer, city, province, region
- Viewer category
- Date, day of the week when the vote has been expressed
- Time (e.g. 22.15), hour of the day (1-24) when the vote has been expressed
- Televoting mode (Phone, Facebook, Instagram, TV program website)
- Edition of the TV program (e.g. '2018', '2019')
- Presenter of the edition

**Homework tasks**

1. Design the data warehouse to address the specifications and to efficiently answer to all the provided frequent queries. Draw the conceptual schema of the data warehouse and the logical schema (fact and dimension tables).
2. Write the following frequent queries using the extended SQL language.
   (a) Consider the votes expressed with the 'Instagram' platform. Separately for edition and viewer city, compute:
      i. the cumulative total incomes obtained during the different editions
      ii. the percentage of incomes brought by viewers of a specific city with respect to the total incomes
      iii. the daily average incomes
   (b) Consider the editions between 2000 and 2010. Separately for edition, execute the following analyses:
      i. the total number of votes.
      ii. the percentage of number of votes of each participant with respect to the total number of votes.
      iii. assign a rank to the participants by decreasing average vote (0-5).