

Outline of hands-on training session 3 (Questions)

- **Task 1:** Please import the CSV file, named “csv file for upload (dataset Chile Plebiscite 1988).csv”, to your database and name the new table as ‘chile’. You can download the CSV file from MyCourse.

Please read the description of the dataset (Description of Chile election 1988.docx) so that you can continue with doing following tasks, which are developed on a basis of this dataset.

- **Task 2:** Export the dataset [the one you just imported at task 1] to be a SQL file.
- **Task 3:** In the classicmodels database, please drop the foreign key relating to the table **orders**. You have imported the tables belonging to the classicmodels database through importing the file “classicmodels (using PCs in the lab).sql” at the first hands-on session.
- **Task 4:** Performing the following calculation via MySQL:
What is the result of $2.345 \times 3.456 \div 4.567$

Please try a few different calculations in MySQL that come to your mind.
- **Task 5:** Please retrieve records from the table ‘chile’ pertinent to the columns of ID, sex, income and vote. Change the sequence of the column names in the commands and check how the changes affect the presentation of the results.
- **Task 6:** Write a simple command [using the * function] to show all the columns and rows of table ‘chile’ without specifying any column name.
- **Task 7:** How many rows of records are there in the table ‘chile’?
- **Task 8:** When conducting a research on ‘chile’ data, your research group would like to study people who were living in Metropolitan Santiago area. In other words, the group wants

to export the records, of which values in the column 'region' is 'M', to be a CSV file, so that the file can be used later by e.g. a data mining software for further analysis. Please retrieve the records and then export them into a CSV file.

- **Task 9:** You are very interesting in comparing the voting choices of people in good status quo (statusquo more than 1) to the choices of those in bad status quo (statusquo less than -1). Thus, you want to create a new table to save the records of those people specified above. In this vein, please create a new table (Statusquo_research) that includes the columns of age, sex, statusquo and vote of those people who satisfy the requirements mentioned-above. Thereafter, copy the commands to the annotation window of the table so that you could remember how you developed this new table in the future.

Note: Please using 'less than' or 'more than' functions to create the table. After that please using 'not between and' function to create table again.

- **Task 10:** Export the table 'Statusquo_research' to be a csv file.

Note: **there are over 1000 rows of record in the database.**

- **Task 11:** One of your friends in the research group argues that, in Chilean 1988 Plebiscite, young and middle-age users (age less than 50) with a good education level of secondary education, and an income of over 75,000 Pesos are more likely to vote for Pinochet than the young and middle-age users (age less than 50), who have a limited education (primary education) and with a limited income (less than 10,000 Pesos). You want to test whether it is true or not. Please calculate the difference in the percentage of people voting for Pinochet for these two groups respectively.

Note:

1. Please first obtain the numbers of rows/records that satisfy the requirement and then compute the percentage.
2. MySQL have a special keyword to count row number which will be introduced in future class. However, you could still get the number of rows through the skills introduced at task 7.

- **Task 12:** Please import the file "tripadvisor_review_sample without review text.sql" to your database.

In the TripAdvisor data [table: tripadvisor_review_sample_without_reviewtext], you would like to find the hotels that received very bad overall rating from very experienced travelers. Specifically, could you retrieve records based on the following conditions?

- Overall_rating is 1;
- author_num_cities_visited is over 10;
- author_num_reviews is over 10;
- author_num_hotelreviews is over 10;
- the date of review is after 2012-10-21 but before 2012-12-21
- Include the review only from the users whose username does not start with the character 'a', 'b', 'c', or 'd'.
- Order the result by review date in a descending manner.

Advance question:

- **Task 13:** Your lecturer found that you are progressing so fast and so he decided to give a crazy task to you. The task is as follows:

Based on the chile data, please create a new table including the people who are:

- From either North or South area of Santiago;
- Living in a region with a population in the range of 5,000 and 90,000
- With age between 23-31 or 35-46, or 55-65
- With monthly income of between 2,500 and 10,000, or over 75,000
- With statusquo less than -0.5 or more than 0.5
- Have no secondary education
- Whose voting decisions are abstain or undecided.



Please order the result first by income (descending order) and then by age (ascending order) and by statusquo (descending order). Please return number 5-10 rows of the results.

Task completed!