Outline of hands-on training session 4 (Questions)

The following questions are developed on a basis of the table *cfpb_complaints_2500*. Please import the file [*cfpb_compliants 2500.sql*] to your database and then perform the analysis:

➢ **Task 1**: please retrieve consumer complaints that happened (Data_received) on 2012-01-01, 2012-02-01, 2012-03-01, 2012-04-01 or 2012-05-01, using function 'IN'.

➢ **Task 2:** please retrieve the records relating to the company *Bank of America*, in which the *issue* is related to *ATM* (string includes ATM).

➢ **Task 3:** please retrieve the records relating to the company *Bank of America*, in which the *issue* is related to *ATM* or *theft.*

➢ **Task 4**: please retrieve the top 3 companies that have most *issues* relating to *ATM.*

➢ **Task 5:** please retrieve the records in which *Issue* includes 7 characters.

➢ **Task 6:** please retrieve the records in which *Issue* are related to loan, saving or credit (please use **regexp** function)

➢ **Task 7:** please list the names of companies that are included in the dataset. No duplicated names should be returned.

➢ **Task 8:** please produce a list of different products (column "products") and count their frequencies respectively.

➢ **Task 9:** please list different ways of submitting a complaint (*Submitted via*) and count their respective frequencies.

➢ **Task 10:** please retrieve the first 5 characters as well as last 5 characters of *issue,* respectively.

➢ **Task 11**: please return the last two words of Issue

➢ **Task 12**: Please return a new column by replacing the word "fee" to be "cost" in the column **Issue**, only showing the records in which the **Issue** includes the word "fee".

The following questions are based on the exploring the table [*tripadvisor_data_for_handson_assignment_ONLY*]. Please import the file [tripadvisor_data_for_handson_assignment_ONLY.sql] to your database and perform your analysis.

➢ **Task 13: Please** count the number of words for the title of each review.

➢ **Task 14: Please** find French reviews in the table.

As we know, être, avoir, je, de are popular French words. If one of the above-mentioned French words appears at the review, we say it is a French review. Please identify those French reviews based on the column title. No need to consider punctuation, like [de.].

➢ **Task 15:** Run the following command to create a table representing customers' comment:

```
create table Con_comment (Consumer_comment varchar(1000));
insert into Con_comment values ('first sentence. second sentece. third sentence.');
insert into Con_comment values ('first sentence. second sentece. third sentence.');
```

Please write one query to get the following result

Result #1 (3×2)		
sen1	sen2	sen3
first sentence	second sentece	third sentence.
first sentence	second sentece	third sentence.

● If we want to save the result to be a new table of 'Comment_sentence', what command should be used?

➢ **Advance tasks - Task 16:**

A fake review detection project: We would like to develop something to detect fake reviews online. However, the first thing we need to do is to obtain a sample of fake reviews so that we can analyze their characteristics. We know that fake reviews with exactly the same content may be submitted to different hotels.

- To make it simple, we thereafter only focus on and analyze **review title**, instead of review content.

You need to use 'count(*)' function to do the following tasks.

● Please detect which review title has been most often used by users in the database, and please order the results by the frequency of titles in a descending way.
● Please detect which review title has been most often used by users in the database if the difference in the use of capital character matters, and please order the results by the frequency of repeated titles in a descending way.
● Let's assume that the use of **punctuations,** like '.' and '!' is not important, redo the above analysis by dropping the punctuations '.' and '!' from the titles.
● What is the most popular last word used in the titles, **ONLY** consider those review titles that have at least 2 words?

Hope now you can see the beauty of MySQL! ☺