Assignment for the course MySQL for Data Analytics

Name: Pham Thi Thu Ha

- **Result is more important than process.** If the final answer is wrong, you will get only a few points even if most of your MySQL codes are correct.
- **Please do not copy codes from other students.** Cheating (such as copying) on assignments will lead to course failure.
- **You can add new columns to the table or create additional tables to support your analysis.**
- **The assignment has a total of 100 points.**

1. In the table orders at the car retailer (**classicmodels**) database, what is the customerNumber of the customer who has the highest frequency of placing orders to the company in 2004 [order-Date in 2004]? You don't need to consider whether the products have been finally shipped or not.

   The customerNumber of the customer: 141(6 points)

   The highest frequency is: 9 (6 points)

2. In the **classicmodels** database, customer names and customer numbers can be found [see table "customers"]. Customers make different payments on different dates [See table "payments"]. Please specify the names of two customers who are most **often** (count frequency) to make payments during the weekend [12 points].
   Name of Customer 1: Mini Gifts Distributors Ltd.
   Name of Customer 2: Marseille Mini Autos
   Note: i) Please provide "customerName", not the contact names from table "customers"

3. In the classicmodels database, one sale representative is responsible for one or several customers [see "salesRepEmployeeNumber" in table "customers"]. Customers make different payments on different dates [See table "payments"]. In other words, we can say sale representatives help the company get customers to make payments. Now the question is who is the sale representative that brings the most **revenue** [or **the total amount of the payments from customers in the table payments**] to the company?
   Note i) Ignoring those customers who are not assigned to any sale representative – they did not make any purchases yet.
   The salesRepEmployeeNumber of the sale representative who brings the most revenue is: 1370 [12 points].

4. In the **cfpb_complaints_2500** database, you can find "closed with relief" in **Company_response** column. Please find the **name of the company** which has the highest **ratio** of cases that is '**closed with relief**'. Only those companies with more than 30 cases [all different kinds of cases, no matter whether they are featured with 'closed with relief' or not] in the database are considered (10 points).

   - **ratio** [or percentage] for an individual company is calculated as:

   The number of cases featured with 'closed with relief' / the number of total cases

   The name of the company: Barclays (5 points)

   The ratio of the cases for the company: 42.8571% (5 points)

5. In the **cfpb_complaints_2500** database, there are many cases that are related to 'loan' (those cases in which the word 'loan' is included in the Issue). Please specify **the name of the company** that has the most issues related to '**loan**' on **Wednesday (DATA_received)**. Only companies located in the '**State**' starting with character "**A**" is considered (8 points).

   The name of the company: Bank of America (8 points)

6. In the Chile database, let's assume that an income less than 10,000 is a low_income; an income between 10,000 and 100,000 is a middle_income; an income higher than 100,000 is a high_income. We would like to know whether the income level and the statusquo have a certain impact on females who voted yes to Pinochet. To answer this question, you need to provide the average statusquo value for the females who voted yes to Pinochet in correspondence to their different income levels. (9 points)

   [Please carefully read the question so that you will not miss any important condition when answering the question; Please provide **three** digits after the decimal point in the results]

Suggestion: you may need to update the table by adding a new column of income_level in order to answer the question.

Income level	Mean statusquo
High_income	1.077
Middle_income	0.927
Low_income	0.937

7. Based on the use of cfpb consumer complaint database (2500 rows), please count the frequency of the complaints that satisfy the following three conditions at the same time: i) consumer finally disputed with the company (*Consumer_disputed*); and ii) were received on Friday (*Data_received*) and iii) the difference between *Data_received* and *Data_sent_to_company* are more than 5 days. (10 points)

Please specify the name of the company that has the biggest amount of the abovementioned complaints.

The name of the company: Bank of America (5 points)

Frequency of the mentioned complaints of the company: 13 (5 points)

8. In the car retails database, please find the **name** of the product that generates the most profit for the company (7 points). *Assume* all the orders were accepted and shipped with no cancellation!

    Guidance:
    - In the table orderdetails, we can find the information of productCode, the selling price [priceEach] as well as the quantity of sold product [quantityOrdered].
    - In the table products, we can find the information of productCode, the price [buyPrice] that the company used to purchase product.
    - Therefore, by comparing the purchasing price and selling price as well as considering the quantity of the sold, you will be able to find the most profitable product [productCode].

- You can find the product name from the table products.

The name of the product: 2003 Harley-Davidson Eagle Drag Bike (7 point)