Home Credit –Data Analyst test 2020

1. **SQL test**

## Data Structure

### Table with credit’s data

select \* from view\_credit

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **DATE** | **CREDIT\_AMOUNT** | **ID\_PERSON** | **PAYMENT\_NUM** | **INTEREST** |
| 1001 | 03.09.2006 | 40000 | 5001 | 6 | 0,15 |
| 1002 | 10.09.2006 | 30000 | 5002 | 7 | 0,19 |
| 1003 | 13.09.2006 | 30000 | 5003 | 8 | 0,25 |
| 1004 | 23.09.2006 | 25000 | 5004 | 9 | 0,17 |
| 1011 | 22.09.2007 | 30000 | 5009 | 6 | 0,22 |
| 1012 | 11.11.2007 | 62000 | 5003 | 11 | 0,25 |
| 1005 | 23.09.2006 | 40000 | 5005 | 5 | 0,18 |
| 1006 | 26.09.2006 | 30000 | 5006 | 6 | 0,13 |
| 1007 | 03.10.2006 | 30000 | 5007 | 5 | 0,19 |
| 1008 | 09.10.2006 | 25000 | 5008 | 10 | 0,22 |
| 1009 | 11.10.2006 | 50000 | 5009 | 12 | 0,25 |
| 1010 | 12.10.2006 | 55000 | 5010 | 12 | 0,11 |

ID – identification of credit Date – date of credit issue,

Credit\_amount – amount of loan, Id\_person – identification of person,

Payment\_num – number of monthly installments, Interest – year interest rate.

### Table with client’s data.

select \* from view\_person

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID\_PERSON** | **DATE\_BIRTH** | **NAME1** | **NAME2** | **NAME3** | **CHILD\_NUM** |
| 5001 | 03.09.1985 | Igor | Levy |  | 0 |
| 5002 | 13.05.1954 | Ilja | Muromec | Blazen | 0 |
| 5003 | 26.09.1980 | Martin | Zahalka |  | 1 |
| 5004 | 11.04.1949 | Frantisek | Pravy | Lotr | 2 |
| 5005 | 23.08.1974 | Sergey | Panda | Zlobr | 3 |
| 5006 | 26.09.1974 | Maxim | Gross |  | 2 |
| 5007 | 03.12.1981 | Vladimir | Havel |  | 1 |
| 5008 | 09.10.1965 | Lumir | Robinson | Zubejda | 1 |
| 5009 | 21.06.1977 | Jiri | Marat | Hele | 1 |
| 5000 | 30.01.1993 | John | Smith |  | 0 |
| 5011 | 25.11.1932 | Martin | Schmidt |  | 2 |
| 5010 | 12.08.1980 | Jan | Marek |  | 0 |

Id\_person – identification of person, Date\_birth – date of birth,

Name1 – name, Name2 – surname,

Name3 – second name, Child\_num – number of children.

## Question:

### Which query returns detailed information about all credits and clients who got that credits?

1. select \* from view\_credit c right join view\_person p on c.id = p.id\_person
2. select \* from view\_credit c join view\_person p on p.id\_person = c. id\_person
3. **select \* from view\_credit c left join view\_person p on p.id\_person = c. id\_person**
4. select \* from view\_credit c full outer join view\_person p on p.id\_person = c. id\_person

**select** \*

**from** view\_credit **as** *c*

**left** **join** view\_person *p*

**on** *c*.ID\_PERSON = *p*.ID\_PERSON;

### Which query returns detailed information about all clients and credits they got (i.e. it show all clients and credits data for those who have it)?

1. select \* from view\_person p join view\_credit c on p.id\_person =c.id\_person
2. select \* from view\_credit c join view\_person p on p.id\_person =c.id\_person
3. select \* from view\_credit c left join view\_person p on p.id\_person =c.id\_person
4. **select \* from view\_person p left join view\_credit c on p.id\_person =c.id\_person**

**select** \*

**from** view\_person **as** *p*

**left** **join** view\_credit **as** *c*

**on** *c*.ID\_PERSON = *p*.ID\_PERSON;

### Which query returns information about all clients who don’t have any credit?

1. select c.\* from view\_person p left join view\_credit c on p.id\_person =c.id\_person where c. id\_person is null
2. select p.\* from view\_person p left join view\_credit c on p.id\_person =c.id\_person where c.id is null
3. **select p.\* from view\_person p where p.id\_person not in (select p.id\_person from view\_credit)**
4. select p.\* from view\_person p where p.id\_person not in (select c.id\_person from view\_credit c where p.id\_person = c.id\_person)

**select** \*

**from** view\_person

**where** ID\_PERSON **not** **in** (**select** **distinct** ID\_PERSON **from** view\_credit);

### Which query returns information about all clients who have more than 1 credit?

1. select p.\* from view\_person p where (select count(c.id) from view\_credit c where c.id\_person=p.id\_person)>1
2. select p.\* from view\_person p where (select count(c.id) from view\_credit c where c.id\_person=p.id\_person)=2
3. select p.\* from view\_person p join (select count(\*) k, c.id\_person from view\_credit c group by c.id having count(\*) > 1) g on p.id\_person=g.id\_person
4. **select p.\* from view\_person p join (select count(\*) k, c.id\_person from view\_credit c group by c.id\_person having count(\*) > 1) g on p.id\_person=g.id\_person**

**select** \*

**from** view\_person

**where** ID\_PERSON **in** (**select** ID\_PERSON

**from** view\_credit

**group** **by** ID\_PERSON

**having** **count**(ID) > 1)

### What query returns this result:

|  |  |  |  |
| --- | --- | --- | --- |
| INTEREST | CREDIT\_AMOUNT\_SUM | CREDIT\_AMOUNT\_MAX | Accs |
| 0,11 | 6200000 | 80000 | 512 |
| 0,13 | 3000000 | 50000 | 148 |
| 0,15 | 4000000 | 30000 | 212 |
| 0,17 | 3250000 | 60000 | 303 |

INTEREST - all possible interest rates from data *(listed lowest to highest)*

CREDIT\_AMOUNT\_SUM - sum of amounts with this interest CREDIT\_AMOUNT\_MAX - maximum amount with this interest Accs - number of accounts with this interest

**select** INTEREST

, **sum**(CREDIT\_AMOUNT) **as** *CREDIT\_AMOUNT\_SUM*

, **max**(CREDIT\_AMOUNT) **as** *CREDIT\_AMOUNT\_MAX*

, **count**(**distinct** ID\_PERSON) **as** *Accs*

**from** view\_credit

**group** **by** INTEREST

**order** **by** INTEREST **asc**

### Which query returns average age of clients who have credits?

1. select avg((sysdate-p.date\_birth)/365.25) from view\_person p
2. **select avg((sysdate-p.date\_birth)/365.25) from view\_person p where p.id\_person in (select id\_person from view\_credit c)**
3. select avg((sysdate-p.date\_birth)/365) from view\_person p left join view\_credit c on p.id\_person = c.id\_person
4. select avg((sysdate-p.date\_birth)/365) from view\_person p

**select** **avg**(**datediff**(**now**(), DATE\_BIRTH ) / 365.25) **as** *avg\_age*

**from** view\_person

**where** ID\_PERSON **in** (**select** ID\_PERSON **from** view\_credit)

### Which query returns average number of children for all clients

1. **select avg(p.child\_num) from view\_person p**
2. select avg(p.child\_num) from view\_person p where p.id\_person in (select id\_person from view\_credit c)
3. select sum(p.child\_num)/count(\*) from view\_person p where p.id\_person in (select id\_person from view\_credit c)
4. select avg(p.child\_num) from view\_person p where p.id\_person in (select id\_person from view\_credit c where c.id\_person=p.id\_person)

**select** **avg**(CHILD\_NUM) **as** *avg\_chil*

**from** view\_person

# Logical test

### There are 6 employees in company. Two of them have 11000 USD year salary. One of them gets 8000 USD per year. Others get 14000 USD. What is the average monthly cost for salary for one employee in this company? Why?

The average yearly cost for salary for one employee in this company is:

= (11.000\*2 + 8.000 + 14.000\*3)/6 = 12.000

### Statements: All flowers are cats. All fans are cats.

Conclusions:

* + 1. All flowers are fans.
    2. Some fans are flowers.

1. Only conclusion I follows
2. Only conclusion II follows
3. Either I or II follows
4. **Neither I nor II follows** (tệp flowers và fans đều là tệp con của cats, 2 tệp này có thể không giao nhau luôn > nên cả 2 đáp án có thể sai
5. Both I and II follow

### Mean & median

Assume there are 2 groups of people, first containing 10 males, second containing 10 females. Average age in group of males is 25 and average age in group of females is 20. There are twins within groups - a male and a female. Which of following conclusions are generally valid?

1. Median age of men is bigger than median age of women. (**true**/false)

Trung bình lớn hơn không có nghĩa median sẽ lớn hơn, nhưng trong tình huống này, với cùng số lượng (10 người) và có cặp song sinh cùng tuổi, thường thì median nam sẽ lớn hơn hoặc ít nhất bằng median nữ > Chọn true

1. If we sort both groups by age and create 10 pairs each containing a man & woman (youngest man & youngest woman, 2nd youngest man & 2nd youngest woman, etc.), there will be 5 or more pairs having a man older or same age as a woman. (**true**/false)

Trung bình tuổi nam cao hơn nữ, tức tổng tuổi nam lớn hơn tổng tuổi nữ.

Khi ghép cặp theo thứ tự tăng dần, tổng chênh lệch tuổi giữa nam và nữ sẽ là tổng (tuổi nam - tuổi nữ) của 10 cặp.Tổng chênh lệch này phải dương (vì tổng tuổi nam lớn hơn tổng tuổi nữ).

Để tổng chênh lệch dương, ít nhất một nửa số cặp phải có nam lớn tuổi hơn hoặc bằng nữ, vì nếu ít hơn 5 cặp như vậy thì tổng chênh lệch sẽ âm hoặc bằng 0 (không đủ để tạo tổng tuổi nam lớn hơn).

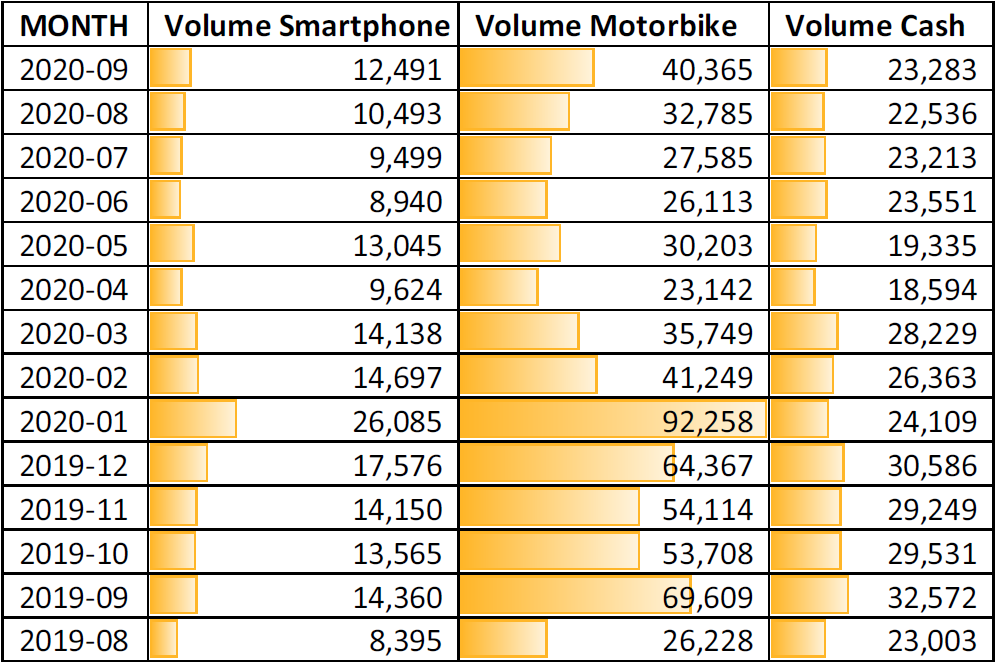
1. Groups can be rearranged by moving people from one group to another so average age increases in both groups. (true/**false**)

Để tăng tuổi trung bình của nhóm nam thì cần đổi 1 người nhóm nữ già hơn nam vào nhóm nam. Tuy nhiên đổi vậy lại làm tuổi trung bình của nhóm nữ bị giảm do bị mất đi người lớn tuổi thay bằng 1 người nam tuổi nhỏ hơn > Chọn false

# Analyst test

### Analytics 1

Given the table as below:



1. Why the volume in January is significantly higher than other months?

Volume January tháng 1 cao hơn hẳn các tháng khác đến từ sự tăng volume Smartphone & Motorbikes hơn các tháng khác, còn volume cash thì không thay đổi thậm chí thấp hơn.

1. Why the volume of Motorbike in the September is normally higher than August and October?

Tháng 9 là thời điểm tựu trường, học sinh, sinh viên bắt đầu đi học, thường sẽ có nhu cầu sắm sửa xe máy cho con em đi học > nhiều bên cũng có các chương trình back to school rất lớn

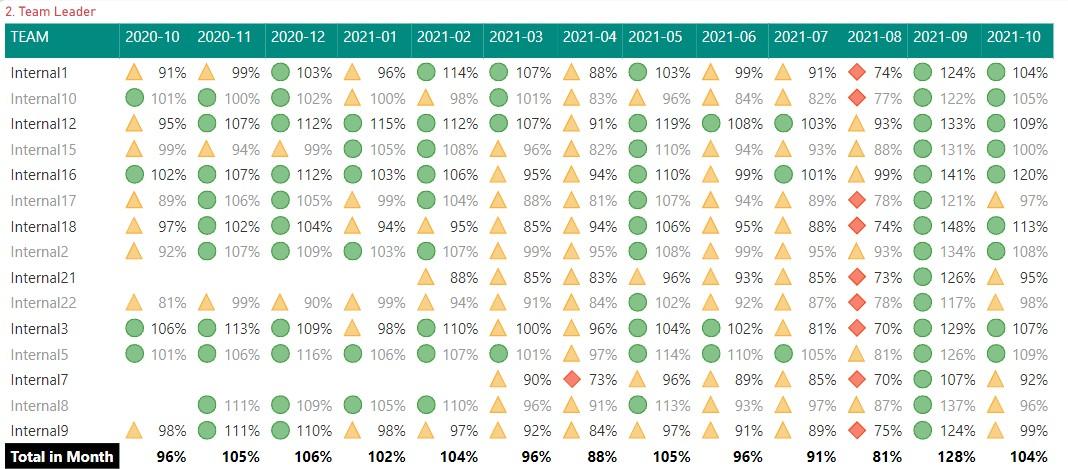
Tháng 8 là lúc vẫn còn đang nghỉ hè, thường sẽ là tháng du lịch, nghỉ ngơi, chưa mua sắm nhiều

Tháng 10 là dịp cuối năm nhưng thường ko phải mùa cao điểm mua sắm (thường cao điểm là blackfriday tháng 11 và christmast tháng 12)

1. Why the trend of Cash volume in June and July go up while others go down?

### Tháng 6, 7 là đầu hè, người ta có nhu cầu đi du lịch nhiều > sẽ cần tiền mặt để chi trả cho các hoạt động này nhiều hơn vì vậy mà xu hướng vay tiền mặt tăng

### Analytics 2

Below table is Cash Loan Target Achievement of Team Leaders during last 13 months:

1. From your point of view, who are the best and the worst Team Leader? Why?

Internal 05 là the best: có 11/13 tháng vượt goal, 2 tháng đạt goal >90%

Internal 07 là the worst: Có 2 tháng đạt goal < 70%, chỉ có 1 tháng vượt goal

1. Is there any unusual result in above table? Why?

Trong tháng 8 có một sự kiện gì đấy khiến cho tất cả các team đều có performance thấp > có thể do yếu tố mùa vụ