

# Home Credit –BI Analyst test

**Name: Tran Trong Hoang Tuan**

**Date: July 28<sup>th</sup>, 2023**

**Result:**

## A. SQL test

### 1. 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.

## 2. Question:

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

- a) select \* from view\_credit c right join view\_person p on c.id = p.id\_person
- b) select \* from view\_credit c join view\_person p on p.id\_person = c.id\_person
- c) select \* from view\_credit c left join view\_person p on p.id\_person = c.id\_person
- d) select \* from view\_credit c full outer join view\_person p on p.id\_person = c.id\_person

2. 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)?

- a) select \* from view\_person p join view\_credit c on p.id\_person = c.id\_person
- b) select \* from view\_credit c join view\_person p on p.id\_person = c.id\_person
- c) select \* from view\_credit c left join view\_person p on p.id\_person = c.id\_person
- d) select \* from view\_person p left join view\_credit c on p.id\_person = c.id\_person

3. Which query returns information about all clients who don't have any credit?

- a) select c.\* from view\_person p left join view\_credit c on p.id\_person = c.id\_person where c.id\_person is null
- b) select p.\* from view\_person p left join view\_credit c on p.id\_person = c.id\_person where c.id is null
- c) select p.\* from view\_person p where p.id\_person not in (select p.id\_person from view\_credit)
- d) 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)

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

- a) select p.\* from view\_person p where (select count(c.id) from view\_credit c where c.id\_person=p.id\_person)>1
- b) select p.\* from view\_person p where (select count(c.id) from view\_credit c where c.id\_person=p.id\_person)=2
- c) 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
- d) select p.\* from view\_person p join (select count(\*) k, c.id\_person from view\_credit c group by c.id\_person where count(\*) > 1) g on p.id\_person=g.id\_person



- c) `select sum(p.child_num)/count(*) from view_person p where p.id_person in (select id_person from view_credit c)`
- d) `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)`

## B. Logical test

- I. 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?

Total annual salary cost =  $(2 * \$11,000) + \$8,000 + (3 * \$14,000) = \$22,000 + \$8,000 + \$42,000 = \$72,000$

The average monthly cost for salary for one employee in this company = Total annual salary cost/6/12=\$1,000

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

Conclusions:

- I. All flowers are fans.
  - II. Some fans are flowers.
- a) Only conclusion I follows
  - b) Only conclusion II follows
  - c) Either I or II follows
  - d) Neither I nor II follows
  - e) Both I and II follow

### 3. 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?

- a) Median age of men is bigger than median age of women. (true/false) (the distribution of ages within each group is not given. The median is the middle value in a sorted list of numbers)
- b) If we sort both groups by age and create 10 pairs each containing a man & woman (youngest man & youngest woman, 2<sup>nd</sup> youngest man & 2<sup>nd</sup> youngest woman, etc.), there will be 5 or more pairs having a man older or same age as a woman. (true/false) (If the value of one sample is too big, it will not be the case)
- c) Groups can be rearranged by moving people from one group to another so average age increases in both groups. (true/false) (The average age of a group is calculated by summing all the ages and dividing by the number of people in the group. Moving people from one group to another would not increase the average age in both groups, because the total sum of the ages is fixed. Moving a person from one group to another would make the average age in one group go up and the other go down, depending on the age of the person moved and the existing average ages of the groups)

- a) Here's an example to counter this:

Let's say the ages of the males are: [20, 21, 22, 23, 24, 25, 26, 27, 28, 29] (median = 24.5)  
And the ages of the females are: [15, 16, 17, 18, 19, 30, 31, 32, 33, 34] (median = 24.5)

In this case, the median age of both groups is the same, so the statement is not true.

Moreover, if we change the ages of the females to: [15, 16, 17, 18, 19, 31, 32, 33, 34, 35] (median = 25.5) and calculate the median again, we get:

Median age of males: 24.5  
Median age of females: 25.5

In this case, the median age of females is bigger than the median age of males, which directly contradicts the original statement. So, the statement is not always true. It depends on the specific ages within each group.

b) Here's an example to counter this:

Let's say the ages of the males are: [18, 19, 20, 21, 22, 23, 24, 25, 35, 36] (average = 25)

And the ages of the females are: [17, 18, 19, 20, 21, 22, 23, 24, 35, 36] (average = 20)

Pair them up again, we get:

- (18M, 17F)
- (19M, 18F)
- (20M, 19F)
- (21M, 20F)
- (22M, 21F)
- (23M, 22F)
- (24M, 23F)
- (25M, 24F)
- (35M, 35F)
- (36M, 36F)

In this case, there are only 8 pairs where the man is older or the same age as the woman.

.....

.....

.....

.....

.....

.....

.....

## C. Analyst test

### 1. Analytics 1

Given the table as below:

MONTH	Volume Smartphone	Volume Motorbike	Volume Cash
2020-09	12,491	40,365	23,283
2020-08	10,493	32,785	22,536
2020-07	9,499	27,585	23,213
2020-06	8,940	26,113	23,551
2020-05	13,045	30,203	19,335
2020-04	9,624	23,142	18,594
2020-03	14,138	35,749	28,229
2020-02	14,697	41,249	26,363
2020-01	26,085	92,258	24,109
2019-12	17,576	64,367	30,586
2019-11	14,150	54,114	29,249
2019-10	13,565	53,708	29,531
2019-09	14,360	69,609	32,572
2019-08	8,395	26,228	23,003

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

**Post-Holiday Sales and Returns:** January is typically a month of high sales volume due to post-holiday sales and gift card redemptions. Additionally, returns from the holiday season may also contribute to an increase in volume as returned items are resold.

**New Year Resolutions and Aspirations:** The start of a new year often brings with it resolutions and aspirations for improvement. This could lead to increased purchases of items like smartphones and motorbikes as people invest in new technology or transportation for the new year.

**Year-End Bonuses and Increased Spending Power:** Many companies distribute year-end bonuses to their employees in December, which could lead to increased spending power in January. This could result in higher sales volumes for discretionary items like smartphones and motorbikes.

**New Product Releases:** Many tech and motorbike companies release new models at the beginning of the year to capitalize on the post-holiday sales rush and the new year's resolutions phenomenon. This could lead to an increase in sales volume.

**Tax Returns and Financial Planning:** Individuals begin receiving their tax returns in January if their company completes the finalization of tax in January. This influx of cash can lead to increased spending. Additionally, people often do financial planning at the beginning of the year, which might involve large cash transactions.

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

**End of Summer Sales:** September marks the end of summer. Retailers often hold end-of-season sales during this time to clear out inventory, which could lead to an increase in motorbike sales.



**Back to School/Work:** September is when students return to school and adults return to work after summer vacations. This could lead to an increase in motorbike purchases as a means of transportation.

**New Model Releases:** Many motorbike manufacturers release new models in the fall to capitalize on the end-of-summer sales season. This could lead to an increase in sales in September.

**Weather Conditions:** September offers optimal weather conditions for riding, which could stimulate motorbike sales.

**Financial Planning:** As the end of the year approaches, businesses may make purchases for tax purposes or to spend remaining budget allocations.

.....

.....

.....

.....

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

**Seasonal Spending:** The summer months often see an increase in consumer spending due to vacations, travel, and other recreational activities. This could lead to an increase in cash volume during this period.

**Mid-Year Bonuses:** Some companies provide mid-year bonuses to their employees, which could lead to an increase in cash volume as people have more disposable income.

**Tax Refunds:** In some countries, tax refunds are distributed around this time, leading to an increase in cash volume as people have more money to spend.

**Seasonal Sales:** Retailers often hold sales during the summer months to clear out inventory for the fall season. This could lead to an increase in cash volume as more transactions are taking place.

**Financial Planning:** Some people may engage in financial planning around the middle of the year, leading to larger cash transactions.

**Some important notes to the data:** high phone volumes are associated with lower cash volumes and high motorbike volumes are associated with higher cash volumes

**Substitution effect:** It could be that when people are buying more smartphones, they are substituting away from other purchases that involve cash. For example, if a large portion of the cash volume is from small, day-to-day transactions, then people might be using the money they would normally use for these transactions to buy smartphones instead. On the other hand, motorbike purchases might be associated with an increase in economic activity, leading to more cash usage.

**Payment methods:** Another possibility could be related to the payment methods used for smartphones and motorbikes. If people tend to buy smartphones using digital payment methods (like credit cards or mobile payments) and motorbikes with cash, that could explain the observed relationship.

**Income effect:** It could also be a matter of income or wealth. Motorbikes are generally more expensive than smartphones, so an increase in motorbike volumes could indicate that people have more money to spend (and hence more cash volume), whereas an increase in smartphone volumes does not necessarily indicate the same.

**Market dynamics:** The market dynamics of smartphones and motorbikes could also play a role. For instance, if there's a new, highly desirable smartphone that many people want, it could cause a temporary spike in smartphone volumes and a corresponding drop in cash volumes as people divert their spending to this new product.

## 2. Analytics 2

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

z. team Leader

TEAM	2020-10	2020-11	2020-12	2021-01	2021-02	2021-03	2021-04	2021-05	2021-06	2021-07	2021-08	2021-09	2021-10
Internal1	▲ 91%	▲ 99%	● 103%	▲ 96%	● 114%	● 107%	▲ 88%	● 103%	▲ 99%	▲ 91%	◆ 74%	● 124%	● 104%
Internal10	● 101%	● 100%	● 102%	100%	▲ 98%	● 101%	▲ 83%	▲ 96%	▲ 84%	▲ 82%	◆ 77%	● 122%	● 105%
Internal12	▲ 95%	● 107%	● 112%	● 115%	● 112%	● 107%	▲ 91%	● 119%	● 108%	● 103%	▲ 93%	● 133%	● 109%
Internal15	▲ 99%	▲ 94%	▲ 99%	● 105%	● 108%	▲ 96%	▲ 82%	● 110%	▲ 94%	▲ 93%	▲ 88%	● 131%	● 100%
Internal16	● 102%	● 107%	● 112%	● 103%	● 106%	▲ 95%	▲ 94%	● 110%	▲ 99%	● 101%	▲ 99%	● 141%	● 120%
Internal17	▲ 89%	● 106%	● 105%	▲ 99%	● 104%	▲ 88%	▲ 81%	● 107%	▲ 94%	▲ 89%	◆ 78%	● 121%	▲ 97%
Internal18	▲ 97%	● 102%	● 104%	▲ 94%	▲ 95%	▲ 85%	▲ 94%	● 106%	▲ 95%	▲ 88%	◆ 74%	● 148%	● 113%
Internal2	▲ 92%	● 107%	● 109%	● 103%	● 107%	▲ 99%	▲ 95%	● 108%	▲ 99%	▲ 95%	▲ 93%	● 134%	● 108%
Internal21					▲ 88%	▲ 85%	▲ 83%	▲ 96%	▲ 93%	▲ 85%	◆ 73%	● 126%	▲ 95%
Internal22	▲ 81%	▲ 99%	▲ 90%	▲ 99%	▲ 94%	▲ 91%	▲ 84%	● 102%	▲ 92%	▲ 87%	◆ 78%	● 117%	▲ 98%
Internal3	● 106%	● 113%	● 109%	▲ 98%	● 110%	▲ 100%	▲ 96%	● 104%	● 102%	▲ 81%	◆ 70%	● 129%	● 107%
Internal5	● 101%	● 106%	● 116%	● 106%	● 107%	● 101%	▲ 97%	● 114%	● 110%	● 105%	▲ 81%	● 126%	● 109%
Internal7						▲ 90%	◆ 73%	▲ 96%	▲ 89%	▲ 85%	◆ 70%	● 107%	▲ 92%
Internal8		● 111%	● 109%	● 105%	● 110%	▲ 96%	▲ 91%	● 113%	▲ 93%	▲ 97%	▲ 87%	● 137%	▲ 96%
Internal9	▲ 98%	● 111%	● 110%	▲ 98%	▲ 97%	▲ 92%	▲ 84%	▲ 97%	▲ 91%	▲ 89%	◆ 75%	● 124%	▲ 99%
<b>Total in Month</b>	<b>96%</b>	<b>105%</b>	<b>106%</b>	<b>102%</b>	<b>104%</b>	<b>96%</b>	<b>88%</b>	<b>105%</b>	<b>96%</b>	<b>91%</b>	<b>81%</b>	<b>128%</b>	<b>104%</b>

a. From your1 point of view, who are the best and the worst Team Leader? Why?

Here's the table with the average percentage achieved for each Team Leader:

Internal	2020-10	2020-11	2020-12	2021-01	2021-02	2021-03	2021-04	2021-05	2021-06	2021-07	2021-08	2021-09	2021-10	Average	Std
Internal18	97%	102%	104%	94%	95%	85%	94%	106%	95%	88%	74%	148%	113%	100%	0.18
Internal21					88%	85%	83%	96%	93%	85%	73%	126%	95%	92%	0.15
Internal3	106%	113%	109%	98%	110%	100%	96%	104%	102%	81%	70%	129%	107%	102%	0.15
Internal8		111%	109%	105%	110%	96%	91%	113%	93%	97%	87%	137%	96%	104%	0.14
Internal16	102%	107%	112%	103%	106%	95%	94%	110%	99%	101%	99%	141%	120%	107%	0.13
Internal9	98%	111%	110%	98%	97%	92%	84%	97%	91%	89%	75%	124%	99%	97%	0.13
Internal1	91%	99%	103%	96%	114%	107%	88%	103%	99%	91%	74%	124%	104%	99%	0.12
Internal15	99%	94%	99%	105%	108%	96%	82%	110%	94%	93%	88%	131%	100%	100%	0.12
Internal10	101%	100%	102%	100%	98%	101%	83%	96%	84%	82%	77%	122%	105%	96%	0.12
Internal7						90%	73%	96%	89%	85%	70%	107%	92%	88%	0.12
Internal17	89%	106%	105%	99%	104%	88%	81%	107%	94%	89%	78%	121%	97%	97%	0.12
Overall	96%	105%	106%	102%	104%	96%	88%	105%	96%	91%	81%	128%	104%	100%	0.11
Internal12	95%	107%	112%	115%	112%	107%	91%	119%	108%	103%	93%	133%	109%	108%	0.11
Internal2	92%	107%	109%	103%	107%	99%	95%	108%	99%	95%	93%	134%	108%	104%	0.11
Internal5	101%	106%	116%	106%	107%	101%	97%	114%	110%	105%	81%	126%	109%	106%	0.11
Internal22	81%	99%	90%	99%	94%	91%	84%	102%	92%	87%	78%	117%	98%	93%	0.10

Based on the average percentage achieved, the best Team Leader is "Internal12", the worst Team Leader is "Internal7"

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

Team 18 has a very high standard deviation. This indicates that the performance of Team Leader 18 fluctuated significantly across the last 13 months, showing considerable variability in achieving the Cash Loan Target. This irregularity might be a cause for further investigation to understand the reasons behind such fluctuations and to identify any potential areas for improvement.

The overall performance also exhibits a significant surge from August 2021 to September 2021. It appears that during this period, there was a notable increase in performance for the company as a whole, resulting in a higher average percentage achieved. Possible explanations for this surge could be that it was near the end of the financial year for the company, or it was the time when finalizing the Key Performance Indicators (KPIs) for the year, and all teams were striving to meet their targets to achieve company-wide objectives.

Other factor must be considered:

Marketing Campaigns: There might have been effective marketing campaigns during that period which led to an increase in awareness and interest in cash loans.

Economic Factors: Changes in the economy, such as lower interest rates, could have made cash loans more attractive.

Seasonal Factors: Some months are traditionally stronger for certain financial products. For example, people might take out loans in anticipation of holiday spending.

Operational Improvements: The team leaders might have implemented new strategies or improved their approach to achieving their targets.

Competitive Landscape: Changes in the offerings of competitors could have made your cash loan product more attractive.

Regulatory Changes: Sometimes changes in laws or regulations can impact the demand for certain financial products.

