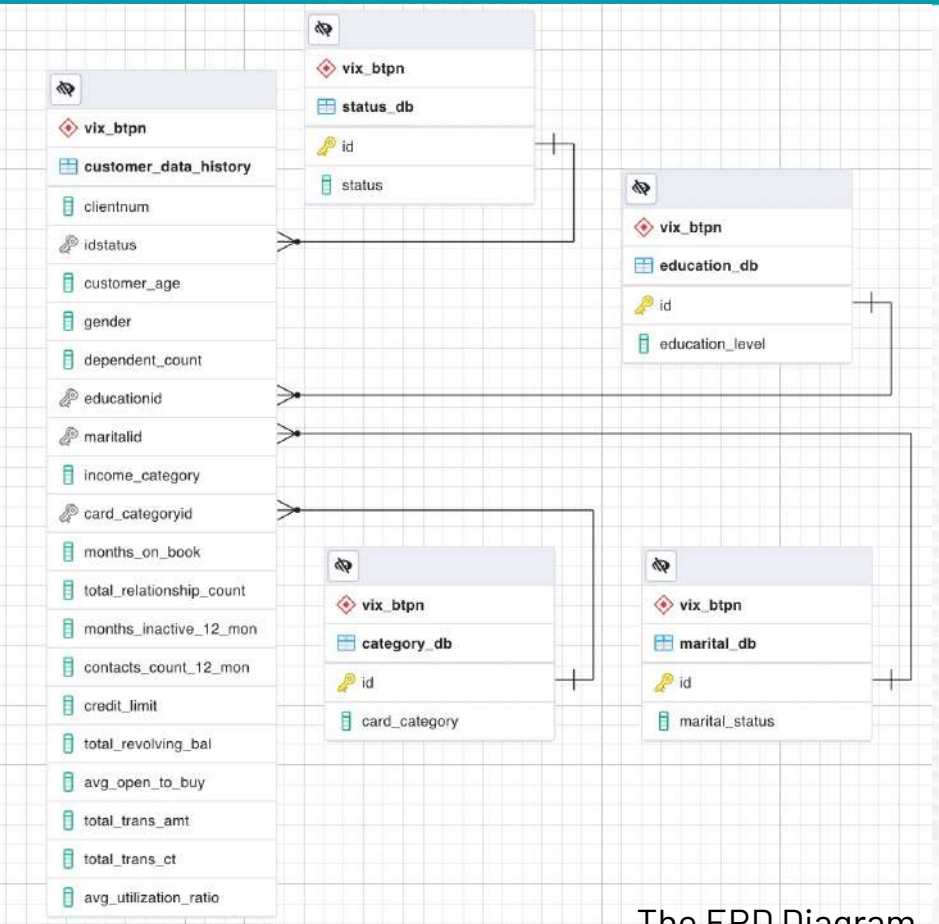
# Final Project
# VIX – Data Engineer
# Bank BTPN Syariah

**Created by:**
**Rheza Paleva Uyanto**
uyantorheza@gmail.com
https://www.linkedin.com/in/rheza-uyanto/

A data enthusiast who graduated from Data Science Bootcamp, Rakamin Academy. and also a registered pharmacist who graduated from Surabaya University since 2017, with 4+ years of experience in clinical pharmacy and quality data indicator in the hospital. Familiar with data preprocessing with Python, PostgreSQL and data visualization with Tableau. Interest with Data Analyst or Data Scientist for fulltime position especially in the health-care setting. With this project, I as Data Engineer, will find 4 items: ***business objective, data exploration, insight visualization and suggestion*** from the company dataset.

**Supported by:**
**Rakamin Academy**
Career Acceleration School
www.rakamin.com

# Overview




The ERD Diagram

As I Data Engineer, I need to find 4 things in this project : business objective, data exploration, insight visualization and suggestion for the company.

I get 5 tables from database vix_btpn,
they are :

- customer_data_history,
- status_db,
- category_db,
- education_db, and
- marital_db.

We use Dbeaver for importing dataset, PostgreeSQL for making a data agregation and Tableau for insight visualization.

# Business Objective

For Business Objective :

The company will want to know why the customer leaving their credit card services.

The company want to know the characteristic that customer

1. Identifying the factor for decreasing the Existing Customer **based on their background.**
    - Education Level, Marital Status, Age,
    - Gender, Dependent Count, Income category
2. Increasing total and frequency of **Customer' transaction using Credit Card**
    - Transaction total , transaction frequency, and ratio of utilization.
3. Identifying the **Customer' Credit Card Information**
    - Card Category, Contact_count_12_mon, total_relationship_count ,
    - credit limit, revolving balance

# Data Exploration

## Customer Background

- Education Level, Marital Status, Age,
- Gender, Dependent Count, Income category

## Common Table Expressions (CTE)

```
with btpn_agregat as
(
select *,
case when customer_age <=30 then '30 years below'
     when customer_age between 31 and 35 then '31-35 years'
     when customer_age between 36 and 40 then '36-40 years'
     when customer_age between 41 and 45 then '41-45 years'
     when customer_age between 46 and 50 then '46-50 years'
     when customer_age between 51 and 55 then '51-55 years'
     when customer_age between 56 and 60 then '56-60 years'
     else '>60 years'
end as age_category
    from vix_btpn.customer_data_history cdh
    left join vix_btpn.category_db cd on cdh.card_categoryid = cd.id
    left join vix_btpn.education_db edu on cdh.educationid = edu.id
    left join vix_btpn.marital_db md on cdh.maritalid = md.id
    left join vix_btpn.status_db st on cdh.idstatus = st.id
)
```

We join 4 other tables with the customer_data_history, and make age_category

# Data Exploration

## 1. Customer Background

- Education Level, Marital Status, Age, Gender, Dependent Count, Income category

### Total row

```
select count (clientnum) from btpn_agregat
```

| count bigint |
|---|
| 10127 |

### Total customer

```
select count (distinct clientnum) from btpn_agregat
```

| count bigint |
|---|
| 10127 |

From total of row and total of customer, we don't find any duplicated data.

### Marital Status

```
select marital_status, count (marital_status) from btpn_agregat
group by marital_status
order by count (marital_status) desc
```

The Married is the biggest amount of marital status of customer

| | marital_status character varying (50) | count bigint |
|---|---|---|
| 1 | Married | 4687 |
| 2 | Single | 3943 |
| 3 | Unknown | 749 |
| 4 | Divorced | 748 |

### Education level

```
select education_level, count (education_level)
from btpn_agregat group by education_level
order by count (education_level) desc
```

The biggest amount of customer from Education level is graduated

| | education_level character varying (50) | count bigint |
|---|---|---|
| 1 | Graduate | 3128 |
| 2 | High School | 2013 |
| 3 | Unknown | 1519 |
| 4 | Uneducated | 1487 |
| 5 | College | 1013 |
| 6 | Post-Graduate | 516 |
| 7 | Doctorate | 451 |

# Data Exploration

## 1. Customer Background

- Education Level, Marital Status, Age, Gender, Dependent Count, Income category

### Age

```
select min (customer_age), max (customer_age),
count (distinct customer_age), avg(customer_age) from btpn_agregat
```

| | min integer | max integer | count bigint | avg numeric |
|---|---|---|---|---|
| 1 | 26 | 73 | 45 | 46.32596030 |

The minimum customer age is 26 years old.
The maximum customer age is 73 years old.
They are 45 unique value of age from 26-73 year.
The average age of customer is 46,33 years

```
select customer_age, count (customer_age) from btpn_agregat
group by customer_age
order by count(customer_age) desc limit 1
```

| | customer_age integer | count bigint |
|---|---|---|
| 1 | 44 | 500 |

The biggest number of people (500 people) is 44 years old.

```
select customer_age, count (customer_age) from btpn_agregat
group by customer_age
order by count(customer_age) asc limit 1
```

| | customer_age integer | count bigint |
|---|---|---|
| 1 | 73 | 1 |

And only 1 person who is 73 years old.

```
select age_category, count(age_category) from btpn_agregat
group by age_category order by age_category asc
```

Based on category, The biggest number of customer is around 46-50 years old, followed by category 41-45 years old.

| | age_category text | count bigint |
|---|---|---|
| 1 | 30 years below | 265 |
| 2 | 31-35 years | 654 |
| 3 | 36-40 years | 1478 |
| 4 | 41-45 years | 2264 |
| 5 | 46-50 years | 2388 |
| 6 | 51-55 years | 1747 |
| 7 | 56-60 years | 926 |
| 8 | >60 years | 405 |

# Data Exploration

## 1. Customer Background

- Education Level, Marital Status, Age, Gender, Dependent Count, Income category

### Gender

```
select gender, count (gender)
from btpn_agregat
group by gender order by count (gender) desc
```

| | gender character varying (50) | count bigint |
|---|---|---|
| 1 | F | 5358 |
| 2 | M | 4769 |

The female customer is the bigger population than the male customer (5358 vs 4769)

### Dependent Count

The number of children that the customer has

```
select dependent_count, count(dependent_count)
from btpn_agregat
group by dependent_count order by dependent_count asc
```

| | dependent_count integer | count bigint |
|---|---|---|
| 1 | 0 | 904 |
| 2 | 1 | 1838 |
| 3 | 2 | 2655 |
| 4 | 3 | 2732 |
| 5 | 4 | 1574 |
| 6 | 5 | 424 |

The amount of children that the customer have from no child until 5 children. The most customer have 2-3 children.

### Status

```
select status, count(status)
from btpn_agregat
group by status order by count(status) desc
```

| | status character varying (50) | count bigint |
|---|---|---|
| 1 | Existing Customer | 8500 |
| 2 | Attrited Customer | 1627 |

The most customer is exiting customer

### Income Category

```
select income_category, count (income_category)
from btpn_agregat
group by income_category order by income_category
```

| | income_category character varying (50) | count bigint |
|---|---|---|
| 1 | $120K + | 727 |
| 2 | $40K - $60K | 1790 |
| 3 | $60K - $80K | 1402 |
| 4 | $80K - $120K | 1535 |
| 5 | Less than $40K | 3561 |
| 6 | Unknown | 1112 |

Most of our customer have income less than $40K

# Data Exploration

## 2. Customer' Transaction using credit card

- Transaction total , transaction frequency, and ratio of utilization.

### Transaction total

```
select sum(total_trans_amt), min(total_trans_amt),
max(total_trans_amt), avg(total_trans_amt) from btpn_agregat
```

| | sum bigint | min integer | max integer | avg numeric |
|---|---|---|---|---|
| 1 | 44600182 | 510 | 18484 | 4404.086303 |

The total customer transaction using credit card is $44,600,162 and every customer spend $4,404 / transaction using credit card

### Transaction frequency

```
select sum(total_trans_ct), min(total_trans_ct),
max(total_trans_ct), avg(total_trans_ct)from btpn_agregat

)
select total_trans_ct, count(total_trans_ct) from btpn_agregat
group by total_trans_ct order by count(total_trans_ct) desc limit 1
```

| | sum bigint | min integer | max integer | avg numeric |
|---|---|---|---|---|
| 1 | 656824 | 10 | 139 | 64.85869457 |

| | total_trans_ct integer | count bigint |
|---|---|---|
| 1 | 81 | 208 |

The total customer transaction using credit card is 656,824 times, and every customer use credit card for 64 times in average.
The big amount of customer (208 customer) use the credit card for 81 times.

### Ratio of Utilization

```
select min(avg_utilization_ratio), max (avg_utilization_ratio),
avg(avg_utilization_ratio) from btpn_agregat
```

| | min real | max real | avg double precision |
|---|---|---|---|
| 1 | 0 | 0.999 | 0.27489355180339 |

The average ratio utilization of credit card is 0,27. It means in 10 transaction, 2 -3 transactions using credit card.

# Data Exploration

## 3. Customer' Credit Card Information

- Card Category, Contact_count_12_mon, total_relationship_count,
- credit limit, revolving balance

### Card Category

```
select card_category, count(card_category) from btpn_agregat
group by card_category order by count(card_category) desc
```

The most of customer use their BLUE credit card when transaction (9,436 customers)

| | card_category character varying (50) | count bigint |
|---|---|---|
| 1 | Blue | 9436 |
| 2 | Silver | 555 |
| 3 | Gold | 116 |
| 4 | Platinum | 20 |

### Contact count in 12 months

* total dihubungi bank dalam 12 bulan terakhir

```
select contacts_count_12_mon,
count(contacts_count_12_mon) from btpn_agregat
group by contacts_count_12_mon
order by contacts_count_12_mon asc
```

The bank should customer 2-3 times each customer for reminder their payment.

| | contacts_count_12_mon integer | count bigint |
|---|---|---|
| 1 | 0 | 399 |
| 2 | 1 | 1499 |
| 3 | 2 | 3227 |
| 4 | 3 | 3380 |
| 5 | 4 | 1392 |
| 6 | 5 | 176 |
| 7 | 6 | 54 |

```
select avg(contacts_count_12_mon)
from btpn_agregat
```

| | avg numeric |
|---|---|
| 1 | 2.455317468 |

# Data Exploration

## 3. Customer' Credit Card Information

- Card Category, Contact_count_12_mon, total_relationship_count,
- credit limit, revolving balance

### Total relationship count

* Jumlah produk yang dipegang customer

```
select total_relationship_count, count (total_relationship_count)
from btpn_agregat group by total_relationship_count
order by total_relationship_count asc
```

Most of customer have 3 products in their account

| | total_relationship_count integer | count bigint |
|---|---|---|
| 1 | 1 | 910 |
| 2 | 2 | 1243 |
| 3 | 3 | 2305 |
| 4 | 4 | 1912 |
| 5 | 5 | 1891 |
| 6 | 6 | 1866 |

### Credit limit

```
select min (credit_limit), max (credit_limit),
avg (credit_limit) from btpn_agregat
```

| | min integer | max integer | avg numeric |
|---|---|---|---|
| 1 | 1438 | 34516 | 8631.938678 |

The range of credit limit is $ 1,438 until $ 34,516, and the average is $ 8,631 each customer

### Revolving Balance

* Total saldo bergulir pada kartu kredit

```
select min (total_revolving_bal), max (total_revolving_bal),
avg (total_revolving_bal) from btpn_agregat
```

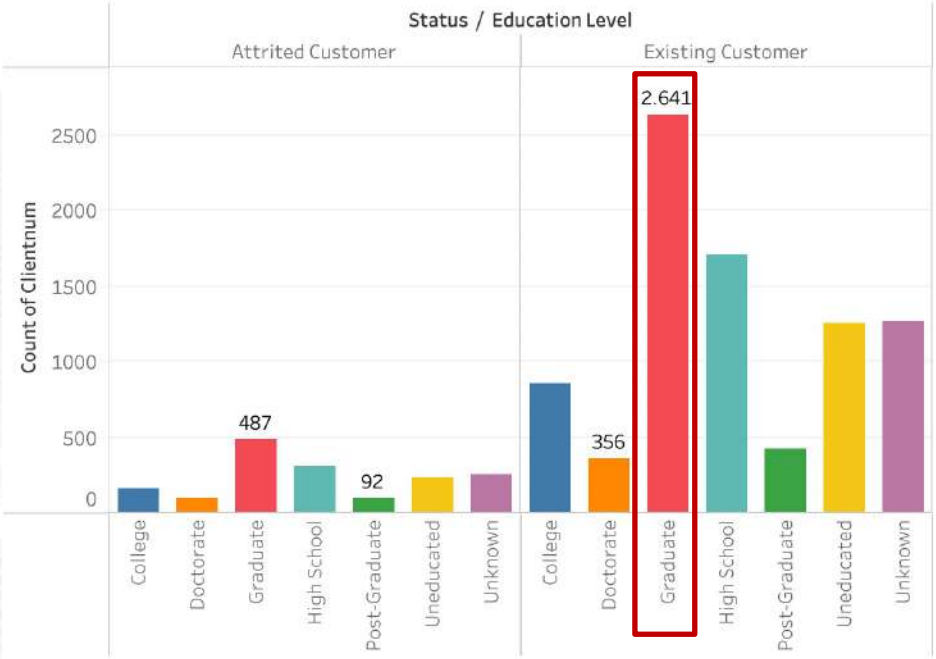
| | min integer | max integer | avg numeric |
|---|---|---|---|
| 1 | 0 | 2517 | 1162.8140614199664264 |

The range of credit limit is $ 0 until $ 2,517, and the average is $ 1,162 each customer

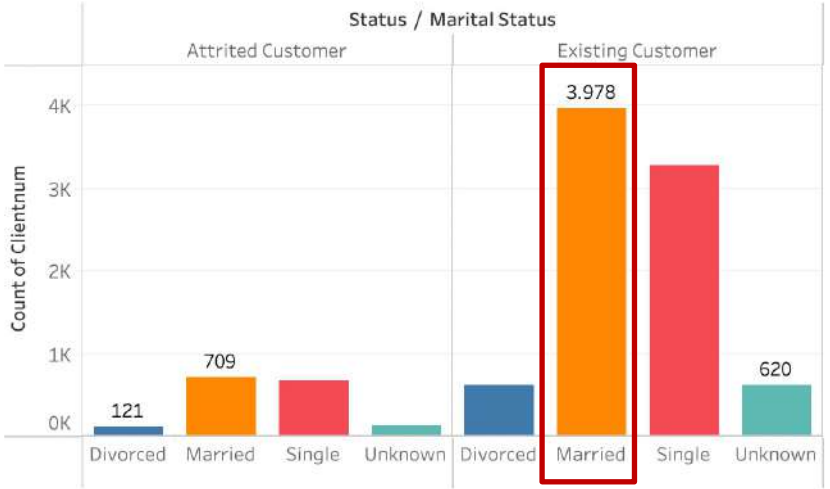
# Insight Visualization

## 1. Customer Background

### Education




### Marital Status

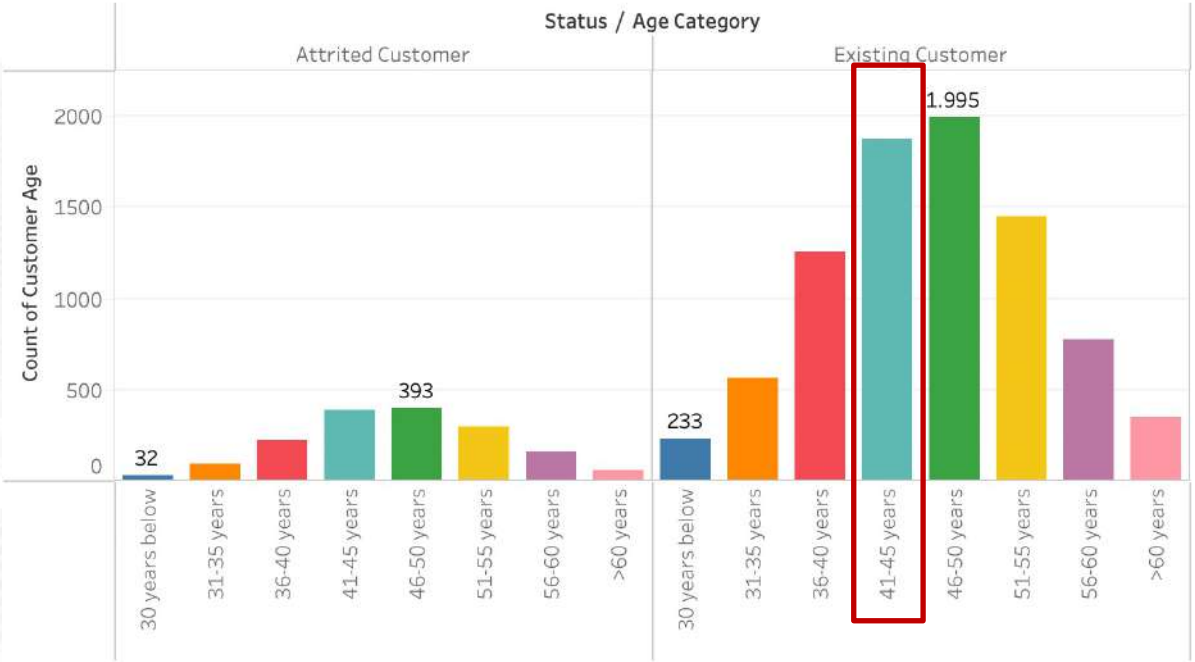



Based on the education and marital status graphic, The customer who existing is graduated and married

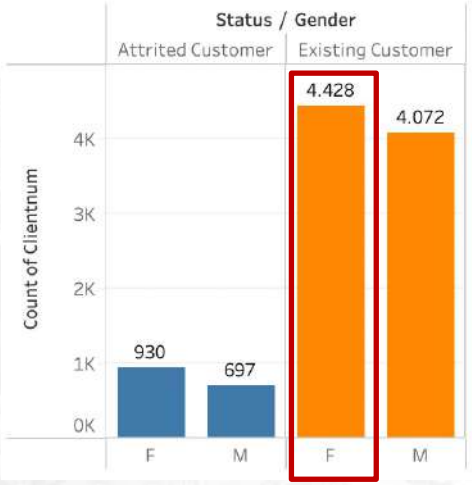
# Insight Visualization

## 1. Customer Background

### Age Category



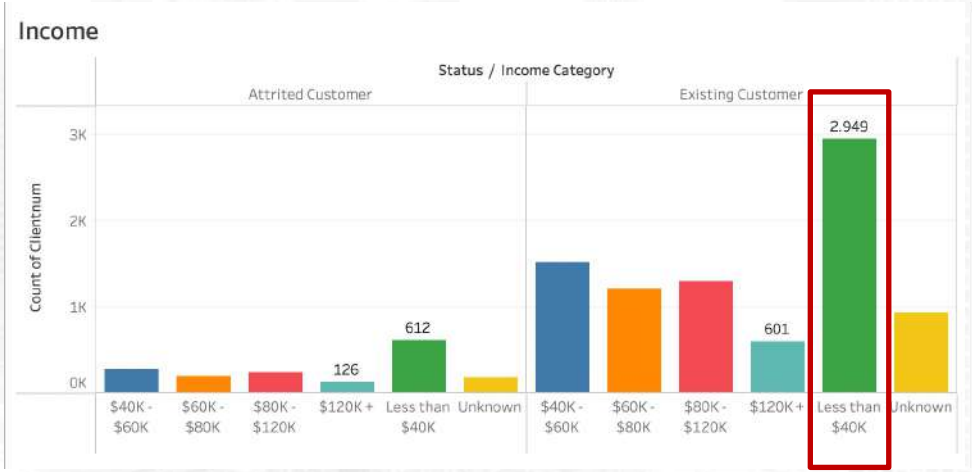

### Gender

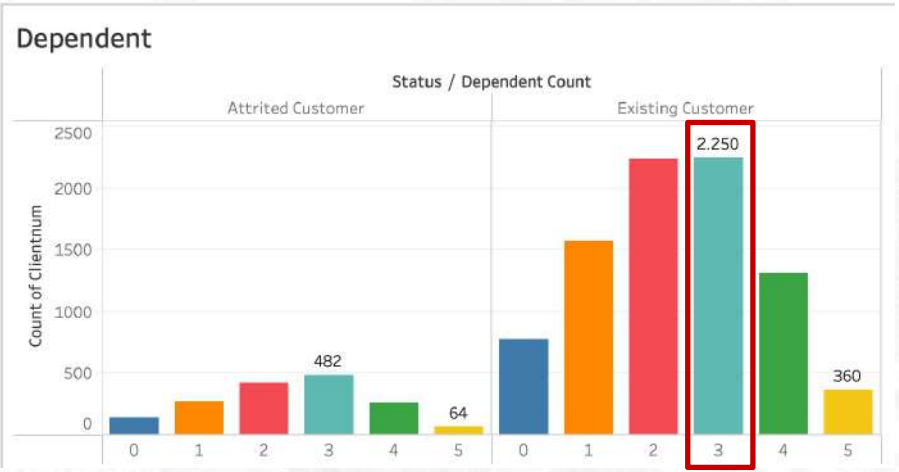



Beside education and marital status, the customer who existing also is in 46-50 years and female gender.

# Insight Visualization

## 1. Customer Background

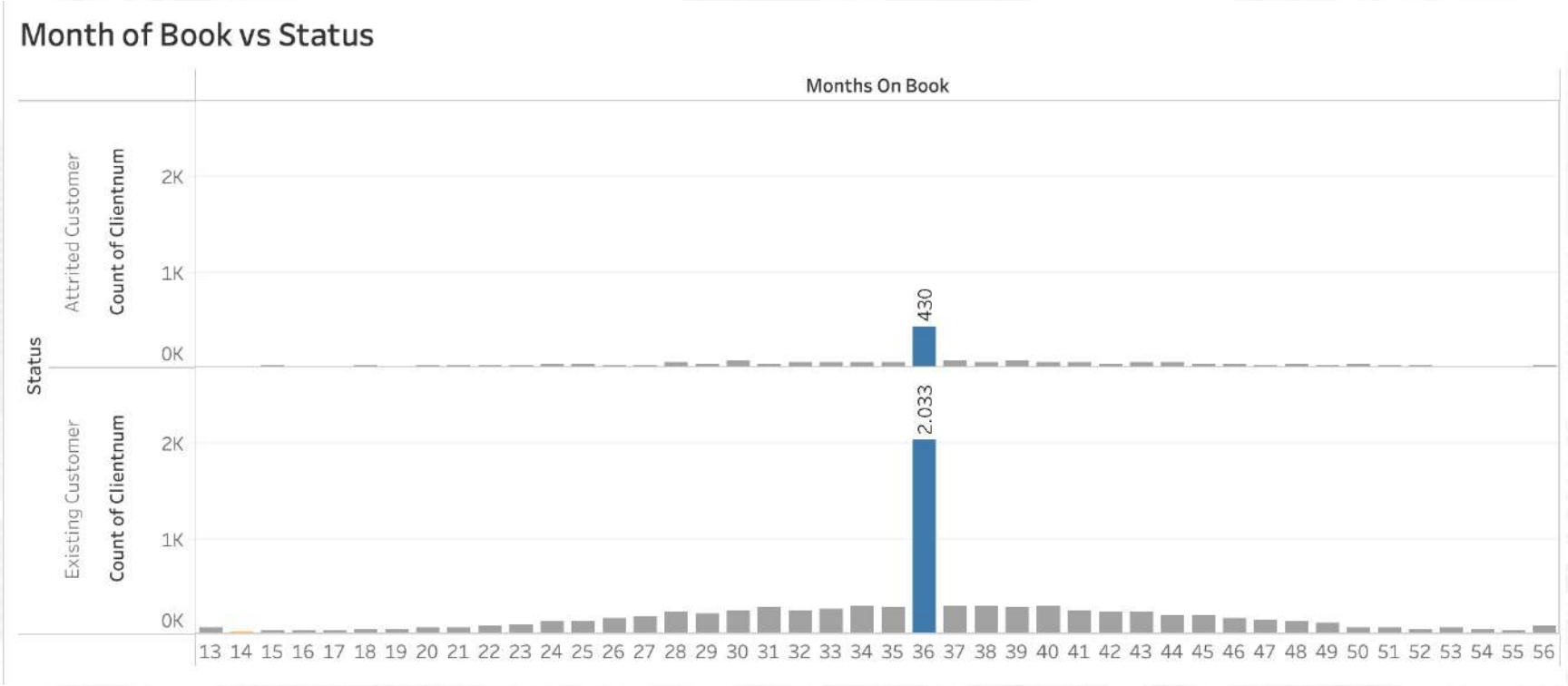



Most of customer who leaving the credit card services has income less than $40K, and have 3 children

# Insight Visualization

## 1. Customer Background




In addition, Most customer have the 36-months-plan for their credit card, for attrited and existing customer

# Insight Visualization

## 2. Customer' Transaction using credit card

### Transaction total

| Status | |
|---|---|
| Attrited Customer | 5.035.607 |
| Existing Customer | 39.564.575 |

### Average Transaction

| Status | |
|---|---|
| Attrited Customer | 3,095 |
| Existing Customer | 4,655 |

The total and the average of transaction of customer who existing is larger than attrited customer

### Transaction Frequency

| Status | |
|---|---|
| Attrited Customer | 1.627 |
| Existing Customer | 8.500 |

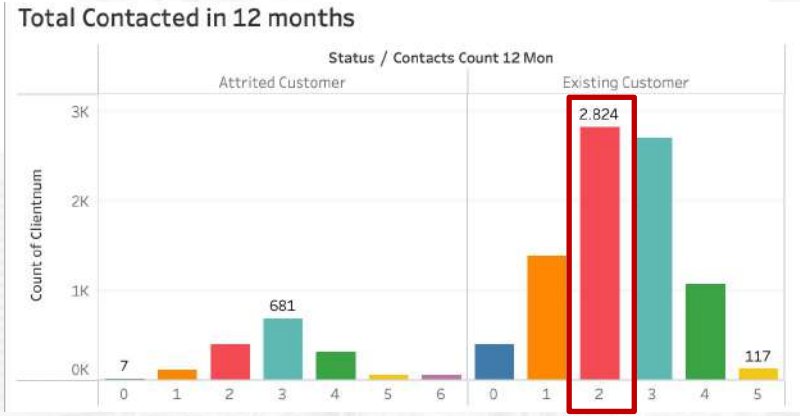
### Transaction Frequency Average

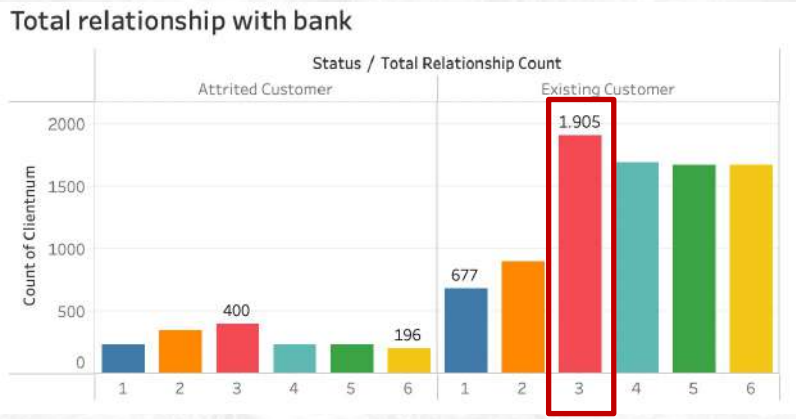
| Status | |
|---|---|
| Attrited Customer | 44.93 |
| Existing Customer | 68.67 |

The transaction frequency of customer who existing is also larger than attrited customer

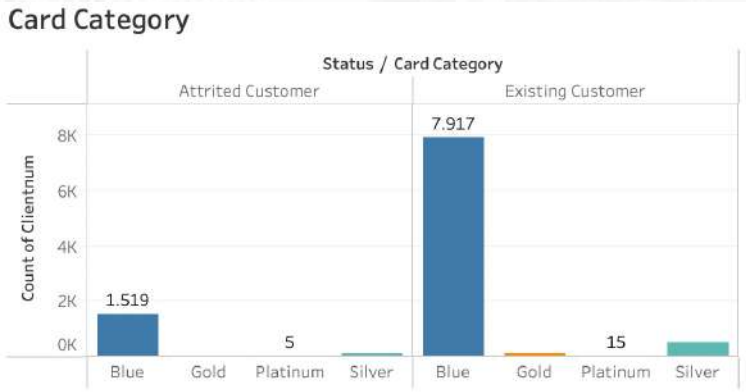
# Insight Visualization

## 3. Customer' Credit Card Information




- Based on credit card facility, the bank usually contact the customer 5 times more on existing customer. The most of existing customer contacted by bank for 2 times.
- The most existing customer have 3 product (the relationship with bank) and the the blue card is the biggest existing customer have.

# Insight Visualization

## 3. Customer' Credit Card Information

### Credit limit

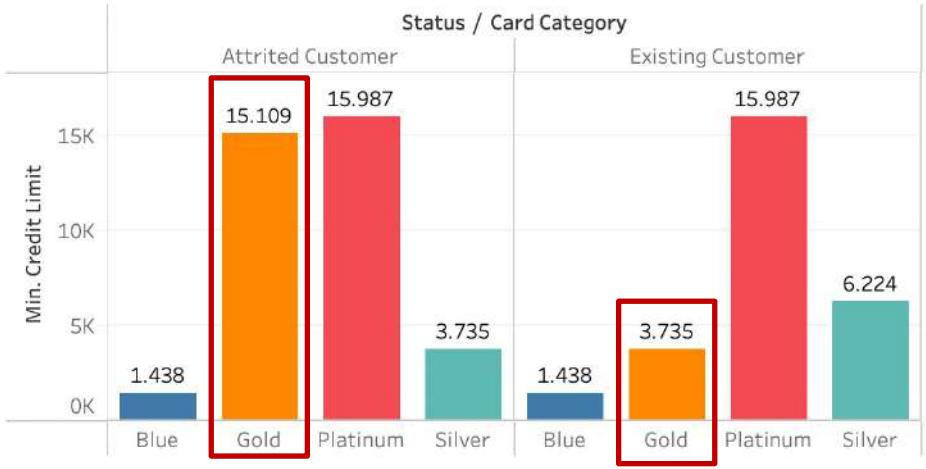
| Status | |
|---|---|
| Attrited Customer | 8,136.0 |
| Existing Customer | 8,726.9 |

### Revolving Balance

| Status | |
|---|---|
| Attrited Customer | 672.8 |
| Existing Customer | 1,256.6 |

- There is not different of average credit limit from attrited and existing customer, but for revolving balance, the existing customer have high than attrited customer.
- For the Gold credit card, the existing customer have minimal credit limit lower than attrited customer.

### Credit limit vs Status



# Suggestion

**Some suggestion :**

1. Review the requirement of customer of 36-months credit plan, and blue cards program. Make sure that this program is achievable for them.
2. Review the credit limit for every card. For the gold card, keep the credit limit same as the platinum card, and review again for 3 months period.
3. Add more program for the customer who use more than 3 products
4. Based on the income, priority get the customer who their income < $40 K, use the light interest should better.
5. For customer service, can contact the customer maximal 3 times, because > 3 times the customer used to leave. The less call from the bank show the low churn rate of customer.