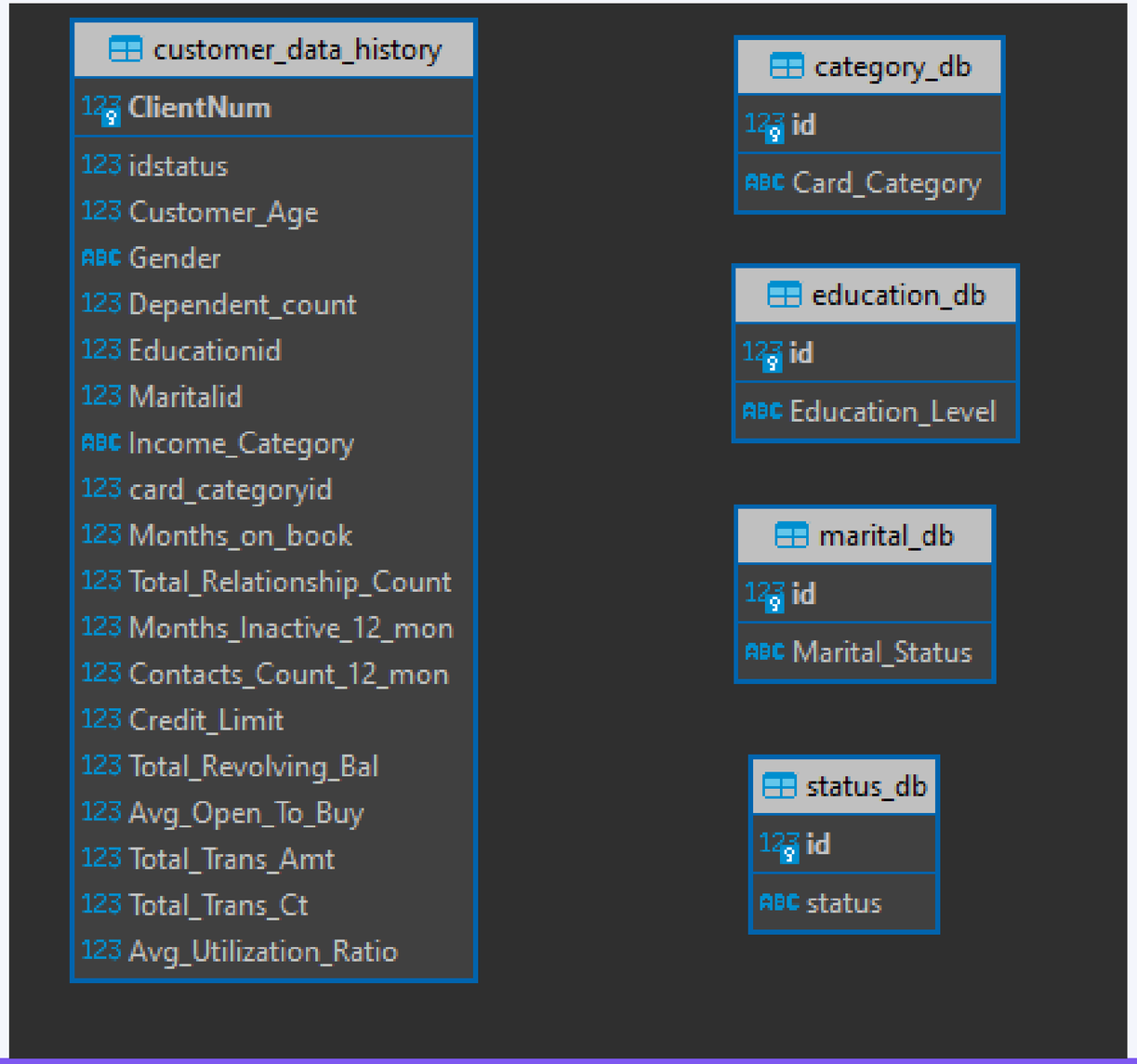
INVESTIGATION OF THE CAUSES OF CUSTOMER CHURN

Final Task VIX Data Engineer
BTPN Syariah

By: Nur Almar'atussaliha

Data Overview

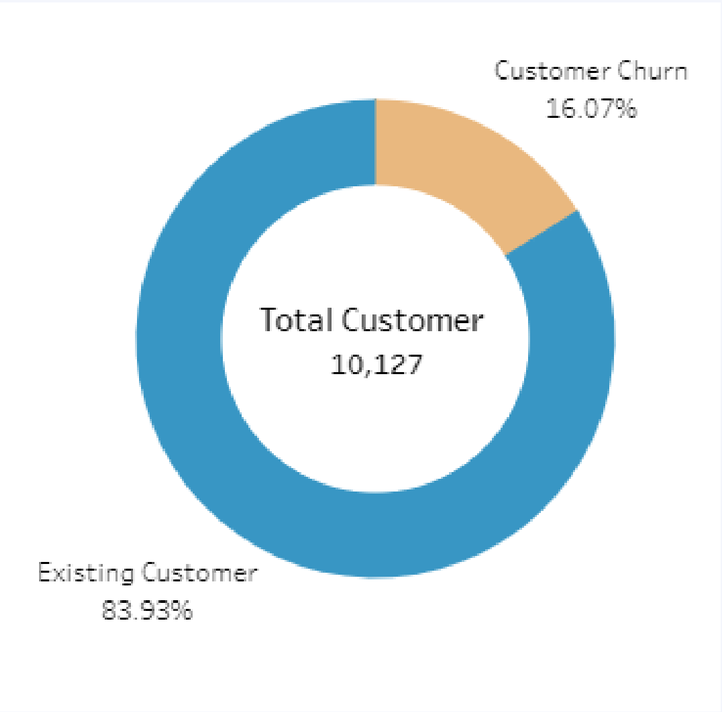


Column Descriptions

Customer data history table: Previous historical customer data
Clientnum :customer ID number
Idstatus :Customer Status Information
Customer_age : age of customer
Gender : gender of the customer
Dependent count : how many dependants the customer has
Educationid : customer's education description
Maritalid : marital status of customer
Income_category : Customer income category
Card_categoryid : type of customer credit card
Month_on_book : period of contact with the bank
Relationship_in_count : Total number of products held by the customer
Months_inactive_in_12_month : Number of inactive months in the last 12 months.
Contacts_Count_12_mon : Total contacts with bank in last 12 months.
Credit limit : Credit limit
Total Revolving Balance on the Credit Card :Total revolving balance on the credit card.
Avg_open_to_buy :Credit card purchases in the last 12 months.
Total_trans_amt : total number of transactions
Total_trans_ct : Transaction frequency
Avg_utilization_ratio : Average card usage rate
Category_db table; credit card service category used
Education_db table: customer education level
Marital_db table: marital status of the customer
Status_db table: customer existing/owned status data

Problem Statement

Bank X has 10271 customers, of which 8500 or 84% are existing customers and 1627 or 16% are customer churn. Bank X wants to reduce customer churn by analyzing the data they have and finding proactive steps to increase their customers and revenue.,



	status	cnt_customer	pct_cust_status
1	Existing Customer	8,500	83.93
2	Attrited Customer	1,627	16.07

Business Objective

Customer Backgrounds

Analyze the causes of customer churn based on customer background such as gender, age, marital status, education, and income.

Customers Transaction

Analyze customer transactions by credit card, total transaction amount, and transaction frequency.

Bank Customer Interaction

Analyze the customer's interaction with the bank, such as the number of months of inactivity and the number of times the bank has contacted the customer in a year.

Data Exploratory

Create New Table

Query:

```
CREATE TABLE customers_data AS
 SELECT cdh.*, cd.Card_Category, ed.Education_Level, md.Marital_Status, sd.status  from
customer_data_history cdh
 LEFT JOIN category_db cd ON cdh.card_categoryid = cd.id
 LEFT JOIN education_db ed ON cdh.Educationid = ed.id
 LEFT JOIN marital_db md ON cdh.Maritalid = md.id
 LEFT JOIN status_db sd ON cdh.idstatus = sd.id
SELECT *FROM customers_data cd  LIMIT 5
```

	ClientNum	idstatus	Customer_Age	Gender	Dependent_count	Educationid	Maritalid	Income_Category	card_categoryid	Months_on_book	Total_Relationship_Count	Months_Inactive_12_mon	Contacts_Count_12_mon	Credit_Limit	Total_Revolving_Bal
1	708,082,083	1	45	F	3	1	1	Less than $40K	1	36	4	3	3	3,544	1,661
2	708,083,283	2	58	M	0	4	2	$40K - $60K	1	45	3	1	3	3,421	2,517
3	708,084,558	2	46	M	3	7	0	$80K - $120K	1	38	6	3	3	8,258	1,771
4	708,085,458	1	34	F	2	3	2	Less than $40K	1	24	6	2	2	1,438.3	0
5	708,086,958	1	49	F	2	3	1	Unknown	1	41	3	5	2	3,128	749

Avg_Open_To_Buy	Total_Trans_Amt	Total_Trans_Ct	Avg_Utilization_Ratio	Card_Category	Education_Level	Marital_Status	status
1,883	15,149	111	0.469	Blue	High School	Married	Existing Customer
904	992	21	0.736	Blue	Unknown	Single	Attrited Customer
6,487	1,447	23	0.214	Blue	Doctorate	Divorced	Attrited Customer
1,438.3	3,940	82	0	Blue	Uneducated	Single	Existing Customer
2,379	4,369	59	0.239	Blue	Uneducated	Married	Existing Customer

We create a new table named '**customers_data**' and join all existing tables with left join.

Data Exploratory

Customer Backgrounds

GENDER BY CUSTOMER STATUS

Query:

```
SELECT Gender , status, COUNT(Gender)
cust_by_gender from customers_data cd
GROUP BY 1,2
ORDER BY 3  DESC
```

	Gender	status	cust_by_gender
1	F	Existing Customer	4,428
2	M	Existing Customer	4,072
3	F	Attrited Customer	930
4	M	Attrited Customer	697

MARITAL BY CUSTOMER STATUS

Query:

```
SELECT Marital_Status, status,
 COUNT(Marital_Status) cust_by_marital
from customers_data cd
GROUP BY 1,2
ORDER BY 3  DESC
```

	Marital_Status	status	cust_by_marital
1	Married	Existing Customer	3,978
2	Single	Existing Customer	3,275
3	Married	Attrited Customer	709
4	Single	Attrited Customer	668
5	Divorced	Existing Customer	627
6	Unknown	Existing Customer	620
7	Unknown	Attrited Customer	129
8	Divorced	Attrited Customer	121

Data Exploratory

EDUCATION BY CUSTOMER STATUS

Query:

```
SELECT Education_Level , status ,
COUNT(Education_Level) cust_by_edu from
customers_data cd
GROUP BY 1,2
ORDER BY 3  DESC
```

	Education_Level	status	cust_by_edu
1	Graduate	Existing Customer	2,641
2	High School	Existing Customer	1,707
3	Unknown	Existing Customer	1,263
4	Uneducated	Existing Customer	1,250
5	College	Existing Customer	859
6	Graduate	Attrited Customer	487
7	Post-Graduate	Existing Customer	424
8	Doctorate	Existing Customer	356
9	High School	Attrited Customer	306
10	Unknown	Attrited Customer	256
11	Uneducated	Attrited Customer	237
12	College	Attrited Customer	154
13	Doctorate	Attrited Customer	95
14	Post-Graduate	Attrited Customer	92

Most of the customers are graduate and high school educated and so are the customers of churn.

Data Exploratory

AGE BY CUSTOMER STATUS

Query:

```
WITH age_segmentation AS (
 SELECT *, CASE WHEN Customer_Age < 30 THEN '  Twenty'
     WHEN Customer_Age < 40 THEN 'Thirty'
     WHEN Customer_Age < 50 THEN 'Forty'
     WHEN Customer_Age < 60 THEN 'Fifty'
     WHEN Customer_Age < 70 THEN 'Sixty'
     ELSE 'Seventy'
    END AS age_group
 FROM customers_data
)
SELECT age_group, COUNT(age_group) tot_group_age
 FROM age_Segmentation
GROUP BY 1
ORDER BY 2 DESC
```

	age_group	tot_group_age
1	Forty	4,561
2	Fifty	2,998
3	Thirty	1,841
4	Sixty	530
5	Twenty	195
6	Seventy	2

Data Exploratory

AGE BY CUSTOMER STATUS

Query:

```
SELECT age_group, status, COUNT(age_group)
tot_group_age  FROM age_segmentation
WHERE status = 'Existing Customer'
GROUP BY 1
ORDER BY 3 DESC
```

	age_group	status	tot_group_age
1	Forty	Existing Customer	3,789
2	Fifty	Existing Customer	2,492
3	Thirty	Existing Customer	1,580
4	Sixty	Existing Customer	459
5	Twenty	Existing Customer	178
6	Seventy	Existing Customer	2

```
SELECT age_group, status, COUNT(age_group)
tot_group_age  FROM age_segmentation
WHERE status = 'Attrited Customer'
GROUP BY 1
ORDER BY 3 DESC
```

	age_group	status	tot_group_age
1	Forty	Attrited Customer	772
2	Fifty	Attrited Customer	506
3	Thirty	Attrited Customer	261
4	Sixty	Attrited Customer	71
5	Twenty	Attrited Customer	17

Most customers are in their forties, fifties, and thirties, and so the customers who churn are also in their forties and fifties.

Data Exploratory

INCOME CATEGORY

Query:

```
SELECT Income_Category,
COUNT(Income_Category) tot_cust FROM
customers_data cd
GROUP BY 1
ORDER BY 2 DESC
```

	Income_Category	tot_cust
1	Less than $40K	3,561
2	$40K - $60K	1,790
3	$80K - $120K	1,535
4	$60K - $80K	1,402
5	Unknown	1,112
6	$120K +	727

INCOME BY CUSTOMER STATUS

```
SELECT Income_Category, status,
COUNT(Income_Category) tot_cust_att FROM
customers_data cd
WHERE status ='Attrited Customer'
GROUP BY 1
ORDER BY 3 DESC
```

	Income_Category	status	tot_cust_att
1	Less than $40K	Attrited Customer	612
2	$40K - $60K	Attrited Customer	271
3	$80K - $120K	Attrited Customer	242
4	$60K - $80K	Attrited Customer	189
5	Unknown	Attrited Customer	187
6	$120K +	Attrited Customer	126

The lower the income, the more likely the customer is to churn.

Data Exploratory

Customers Transaction

CARD CATEGORY

Query:

```
WITH card_cat AS (
SELECT Card_Category, status,
COUNT(Card_Category) as cnt_category
FROM customers_data
GROUP BY 1, 2
ORDER BY 2  DESC
)
SELECT *FROM card_cat
```

	Card_Category	status	cnt_category
1	Blue	Existing Customer	7,917
2	Gold	Existing Customer	95
3	Platinum	Existing Customer	15
4	Silver	Existing Customer	473
5	Blue	Attrited Customer	1,519
6	Gold	Attrited Customer	21
7	Platinum	Attrited Customer	5
8	Silver	Attrited Customer	82

The majority of customers have a blue credit card. Both existing customers and churn.

Data Exploratory

TRANSACTION AMOUNT

Query:

```
WITH trx_amt AS (
 SELECT status , MIN(Total_Trans_Amt) min_trx, MAX(Total_Trans_Amt) max_trx,
SUM(Total_Trans_Amt) tot_trx,
 MEDIAN(Total_Trans_Amt) median_trx, AVG(Total_Trans_Amt) avg_trx FROM customers_data
GROUP BY 1
ORDER BY 1
)
SELECT *FROM trx_amt
```

	status	min_trx	max_trx	tot_trx	median_trx	avg_trx
1	Attrited Customer	510	10,583	5,035,607	2,329	3,095.0258143823
2	Existing Customer	816	18,484	39,564,575	4,100	4,654.6558823529

The total transaction of customers is $44,600,182. If we look at the transaction by status, 50% of the customer churn transaction about $2,000 and spent about $3,000 (average).

Data Exploratory

TRANSACTION FREQUENCY

Query:

```
WITH freq_trx AS (
 SELECT  MIN(Total_Trans_Ct) min_freq, MAX(Total_Trans_Ct) max_freq, SUM(Total_Trans_Ct)
tot_freq,
 MEDIAN(Total_Trans_Ct) median_freq, ROUND(AVG(Total_Trans_Ct) , 2) avg_freq
 FROM customers_data cd
 GROUP BY 1
 ORDER BY 1
)
SELECT *FROM freq_trx
```

	min_freq	max_freq	tot_freq	median_freq	avg_freq
1	10	139	656,824	67	64.86

*all customers

	status	min_freq	max_freq	tot_freq	median_freq	avg_freq
1	Attrited Customer	10	94	73,107	43	44.93
2	Existing Customer	11	139	583,717	71	68.67

*customers by status

The customer's transaction frequency is 656,824 times with an average of 65. By status, the maximum transaction frequency of churned customers is 94, where 50% of them have transaction frequency below 43.

Data Exploratory

Bank Customer Interaction

NUMBER OF MONTHS OF INACTIVITY

Query:

```
SELECT Months_Inactive_12_mon, COUNT(Months_Inactive_12_mon)
cnt_inactive_mnt FROM customers_data cd
GROUP BY 1
ORDER BY 2 DESC
```

	Months_Inactive_12_mon	cnt_inactive_mnt
1	3	3,846
2	2	3,282
3	1	2,233
4	4	435
5	5	178
6	6	124
7	0	29

```
SELECT Months_Inactive_12_mon, status,
COUNT(Months_Inactive_12_mon) cnt_inactive_mnt FROM
customers_data cd
WHERE status = 'Attrited Customer'
GROUP BY 1,2
ORDER BY 3 DESC
```

	Months_Inactive_12_mon	status	cnt_inactive_mnt
1	3	Attrited Customer	826
2	2	Attrited Customer	505
3	4	Attrited Customer	130
4	1	Attrited Customer	100
5	5	Attrited Customer	32
6	6	Attrited Customer	19
7	0	Attrited Customer	15

Most of the customers are inactive for 3 months, for 2 months and for 1 month. By status, the inactivity of the customer churn is also for 3 months and for 2 months.

Data Exploratory

THE NUMBER OF TIMES THE BANK HAS CONTACTED THE CUSTOMER

Query:

```
SELECT Contacts_Count_12_mon, COUNT(Contacts_Count_12_mon)
cnt_contact FROM customers_data cd
GROUP BY 1
ORDER BY 2 DESC
```

	Contacts_Count_12_mon	cnt_contact
1	3	3,380
2	2	3,227
3	1	1,499
4	4	1,392
5	0	399
6	5	176
7	6	54

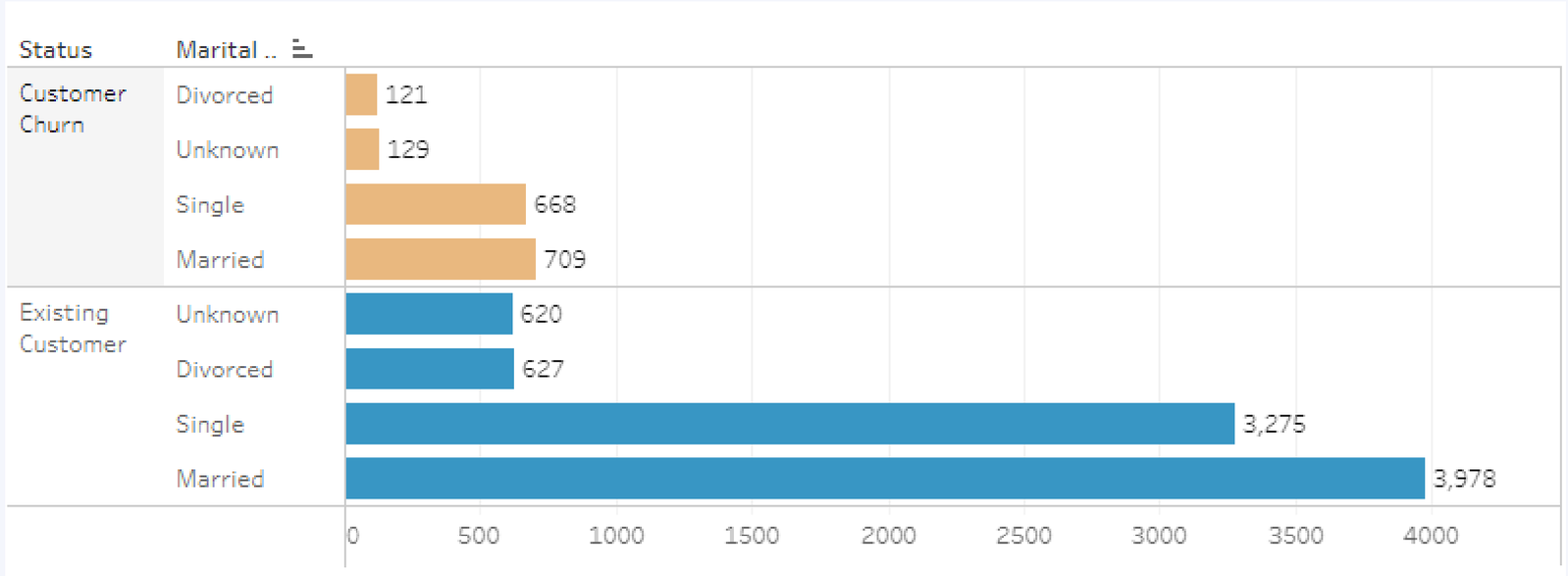
```
SELECT Contacts_Count_12_mon, status ,
COUNT(Contacts_Count_12_mon) cnt_contact FROM
customers_data cd
WHERE status = 'Attrited Customer'
GROUP BY 1,2
ORDER BY 3 DESC
```

	Contacts_Count_12_mon	status	cnt_contact
1	3	Attrited Customer	681
2	2	Attrited Customer	403
3	4	Attrited Customer	315
4	1	Attrited Customer	108
5	5	Attrited Customer	59
6	6	Attrited Customer	54
7	0	Attrited Customer	7

In most cases, the bank contacted all of its customers 2 to 4 times a year. the less contact the bank has with the customer, the more the customer will churn.

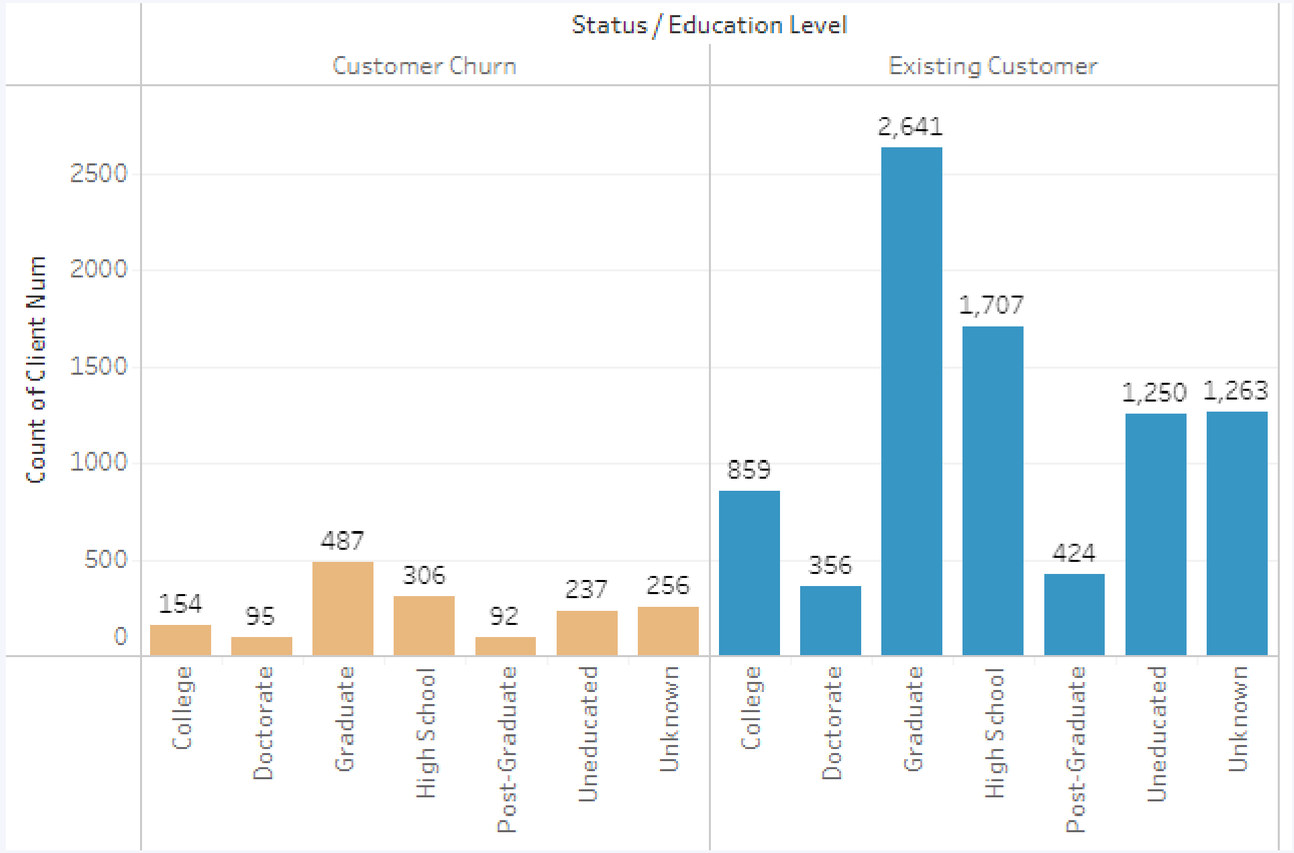
Insight Visualization 📈

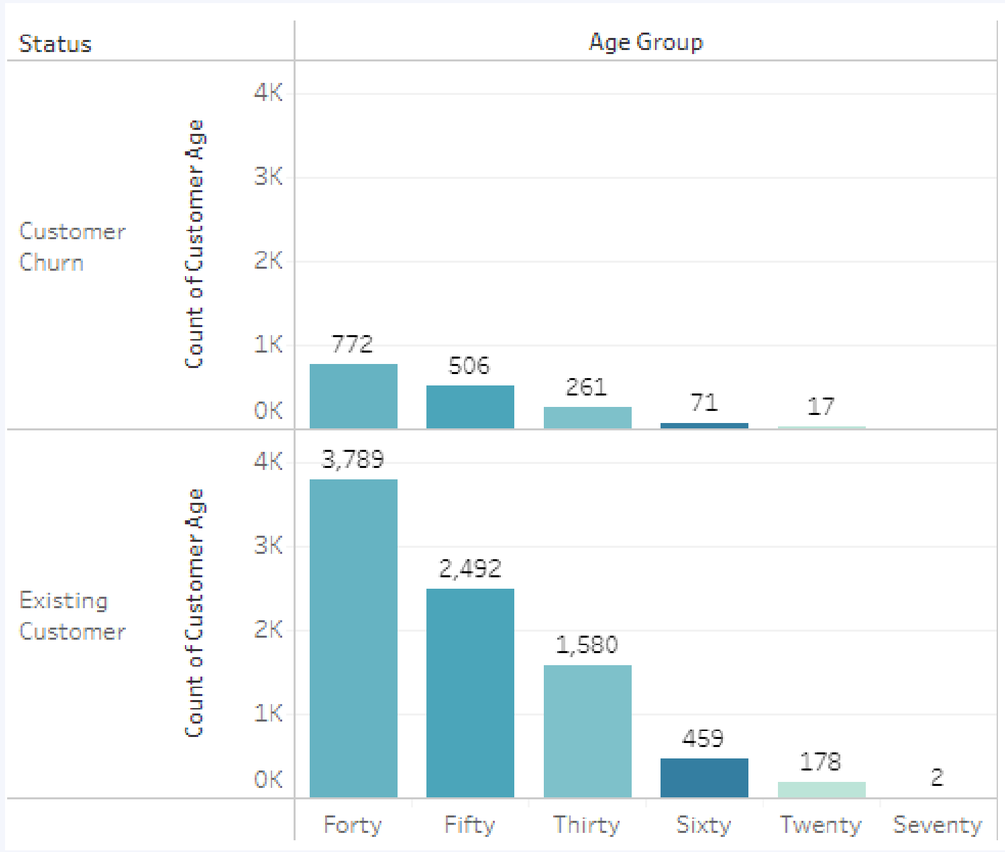
Gender	Customer Churn	Existing Customer
F	930	4,428
M	697	4,072



- ✓ Based on gender, the customer churn is 930, as most of the customers are women.
- ✓ Based on marital status, most customers are married and single, and most customers of churn are from married and single.

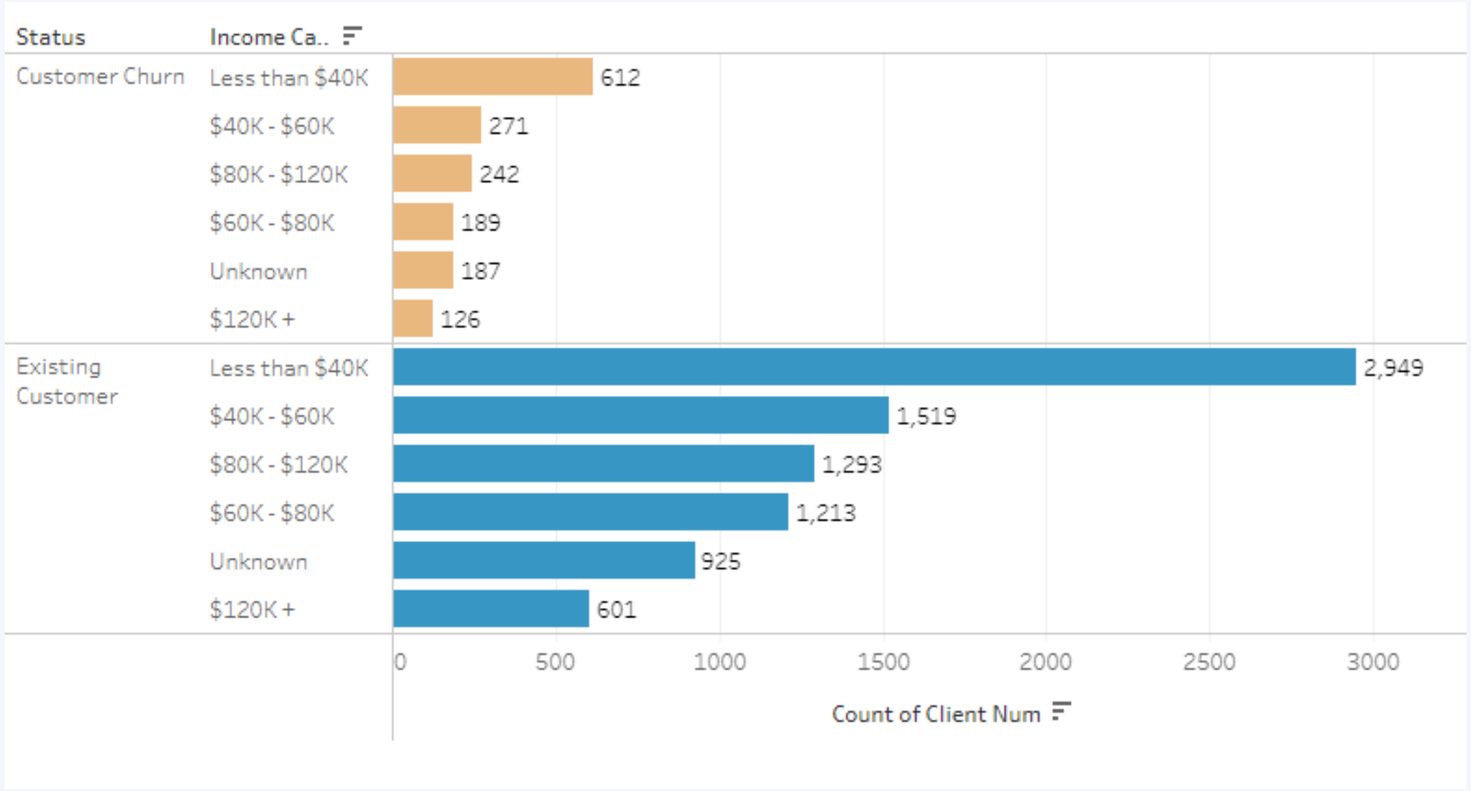
Insight Visualization

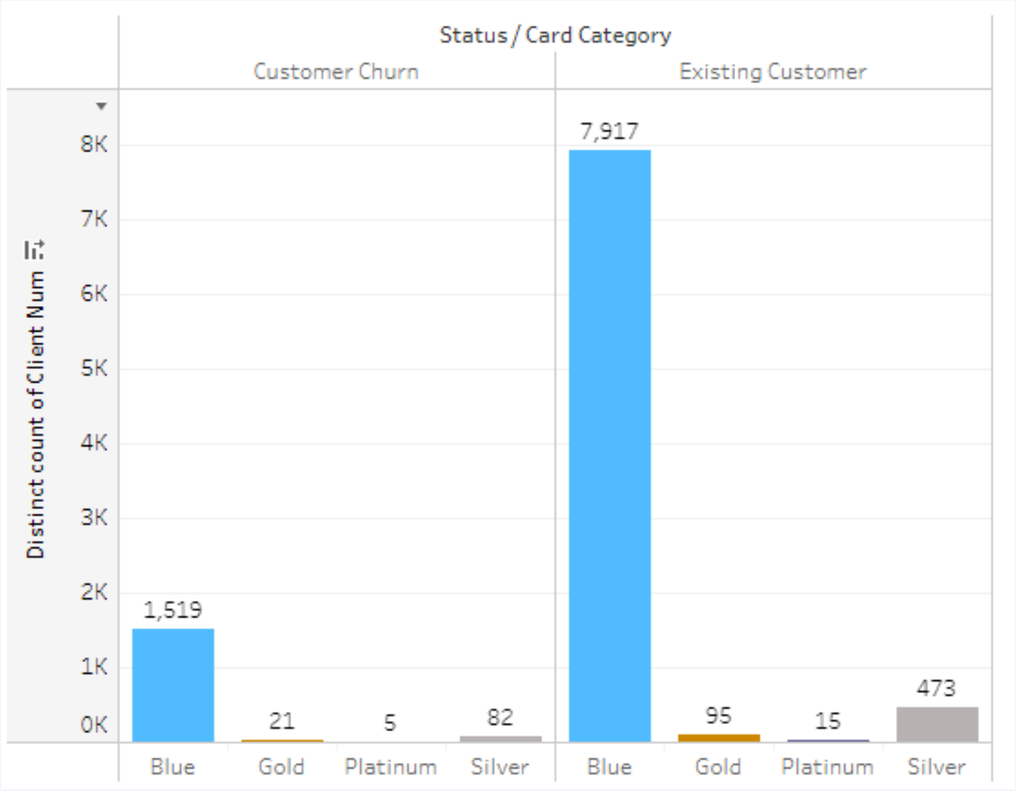




- ✓ Students and graduates are the biggest customers of Bank X. Most of the churn customers are graduates, students and unknown education.
- ✓ After grouping customers by age, we can see that most customers are in their forties and fifties, both in terms of existing customers and customer attrition.

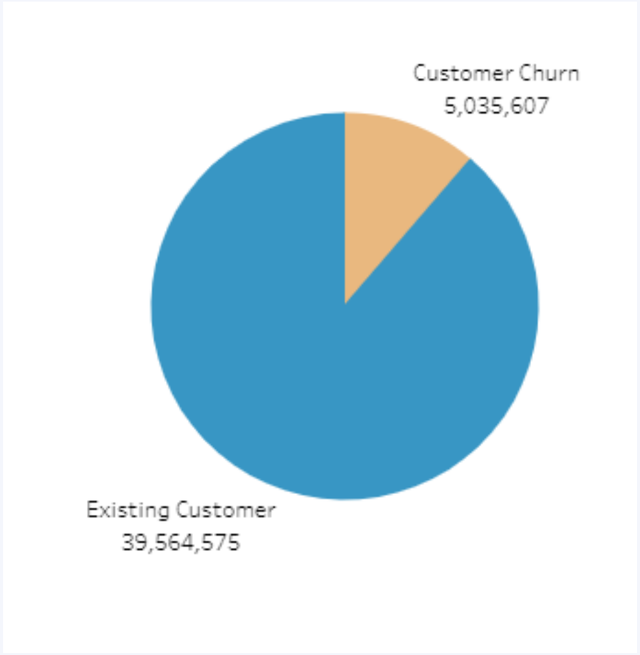
Insight Visualization 📈



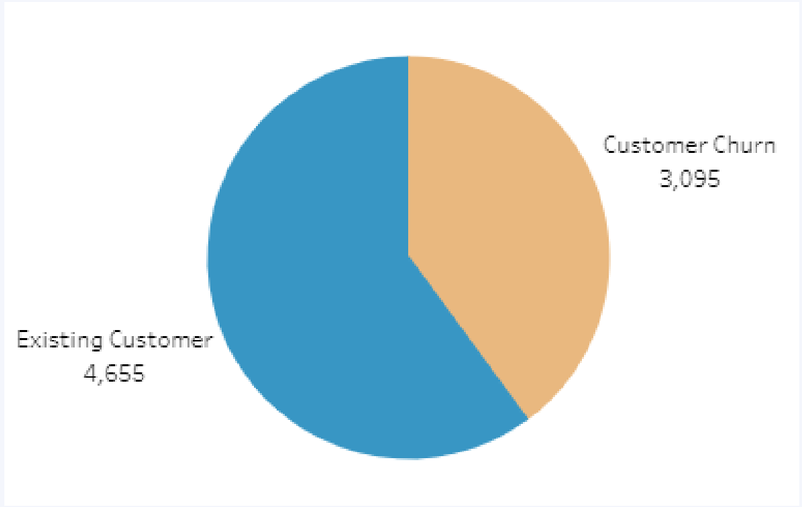


- ✓ Based on income, most customers have an income of less than $400,000 and the highest income is around $120,000. The lower the income, the higher the churn rate.
- ✓ Based on the credit card, most customers have blue credit card. This is the blue card is also considered the highest churn.

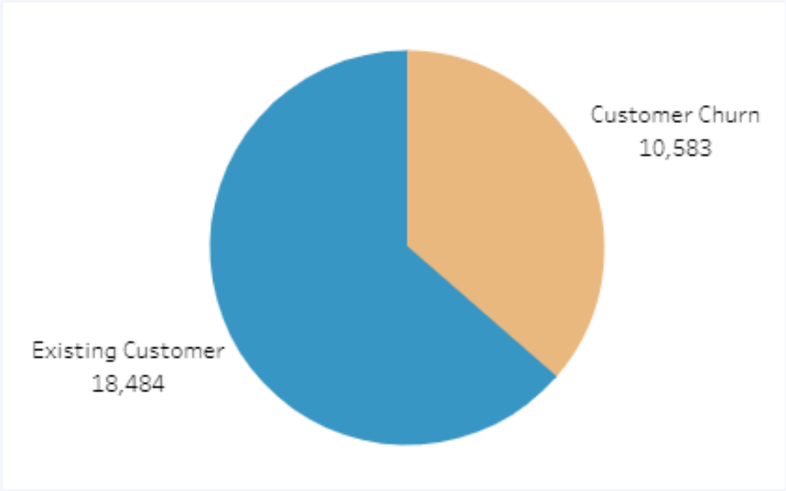
Insight Visualization 📈



Total Transaction



Avg Transaction



Max Transaction

✓ Both existing and churn customers have $44,000,000 in transactions. The average transaction from existing customers is $4655 and the average churn is $3095.

Conclusion and Suggestion

Conclusion

Income and the number of times the bank contacts customers have an impact on customer churn. The lower the income, the higher the churn. The less the bank contacts the customer, the more the customer will churn.

Suggesstion

- Closely monitor what customers buy, how they buy, and what they do with your products.
- Customer retention strategies, such as sending targeted messages via email or Whatsapp to customers who are about to leave, could be just the sign they need to stay with your company. It doesn't always have to be a big deal, sometimes just letting them know you still care is enough.