

# **DATA ENGINEER: FINAL PROJECT**



Data-driven information systems student at Institut Teknologi Sepuluh Nopember and Shibaura Institute of Technology, interning at BTPN Syariah in Data Engineering and a multi-talented individual with experience in volunteering, founding a startup, serving as a programmer and ground control systems in a national robotic competition team. Achiever in business competitions, especially startups and a senior staff at the Startup Academy.

# MATERIAL

BACKGROUND & OVERVIEW

BUSINESS OBJECTIVE

DATA EXPLORATION

INSIGHT VISUALIZATION

BUSINESS RECOMENDATION

# BACKGROUND

The manager of a bank has noticed an increasing number of customers leaving the card service, and is concerned about losing valuable business. In an effort to prevent further attrition, the bank is seeking a way to identify which customers are at risk of leaving so that they can proactively approach them with better service and incentives to keep them onboard.

As a team of data engineers, our goal is to analyze customer data and determine the key factors that contribute to customer churn. Specifically, we will be focusing on customers who have previously held a credit card with the bank and are still active, as well as those who have already churned and are no longer interested in using a credit card. By identifying the characteristics and behaviors of both groups, we hope to uncover insights that will allow the bank to improve their customer retention strategies and reduce overall churn rates.

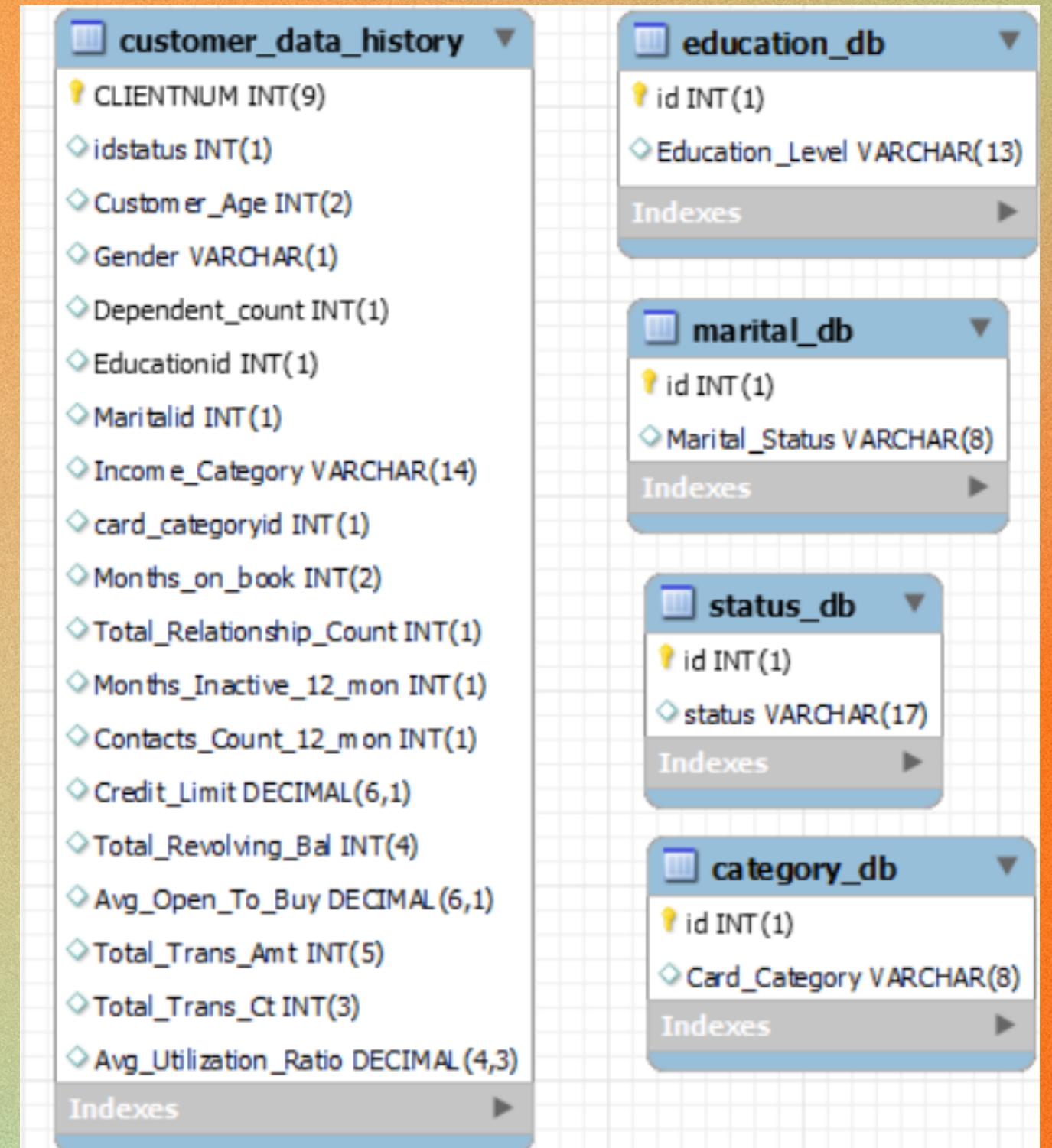
# OVERVIEW

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 FP\_BTPN\_SYARIAH, THEY ARE :

- CUSTOMER\_DATA\_HISTORY,
- STATUS\_DB,
- CATEGORY\_DB,
- EDUCATION\_DB, AND
- MARITAL\_DB.

WE USE MYSQL WORKBENCH FOR IMPORTING DATASET & MAKING A DATA AGREGATION  
TABLEAU FOR INSIGHT VISUALIZATION



# OVERVIEW

THE DATA CONSISTS OF 10127 ROWS

## COLUMN DESCRIPTION:

CLIENTNUM: CLIENT ID NUMBER

IDSTATUS: DESCRIPTION OF CUSTOMER STATUS

CUSTOMER\_AGE : CUSTOMER AGE

GENDER: CUSTOMER GENDER

NUMBER OF DEPENDENTS: THE RESPONSIBILITY OF THE CUSTOMER

EDUCATIONID : INFORMATION ON CUSTOMER EDUCATION LEVEL

MARRIAGE: MARRIAGE CUSTOMER STATUS

INCOME\_CATEGORY : CUSTOMER INCOME CATEGORY

CARD\_CATEGORYID : TYPE OF CUSTOMER'S CREDIT CARD

MONTH\_ON\_BOOK : PERIOD RELATED TO BANK

RELATIONSHIP\_IN\_COUNT : TOTAL PRODUCT HELD BY CUSTOMER

MONTHS\_INACTIVE\_IN\_12\_MONTH : NUMBER OF MONTHS OF INACTIVITY IN THE LAST 12

MONTHS

CONTACTS\_COUNT\_12\_MON : TOTAL CONTACTED BY BANK IN THE LAST 12 MONTHS

CREDIT LIMIT: CREDIT LIMIT

TOTAL REVOLVING CREDIT CARD BALANCE: TOTAL REVOLVING BALANCE ON CREDIT CARDS

AVG\_OPEN\_TO\_BUY : BOUGHT BY CREDIT CARD IN THE LAST 12 MONTHS

TOTAL\_TRANS\_AMT : NUMBER OF TRANSACTIONS

TOTAL\_TRANS\_CT : TRANSACTION FREQUENCY

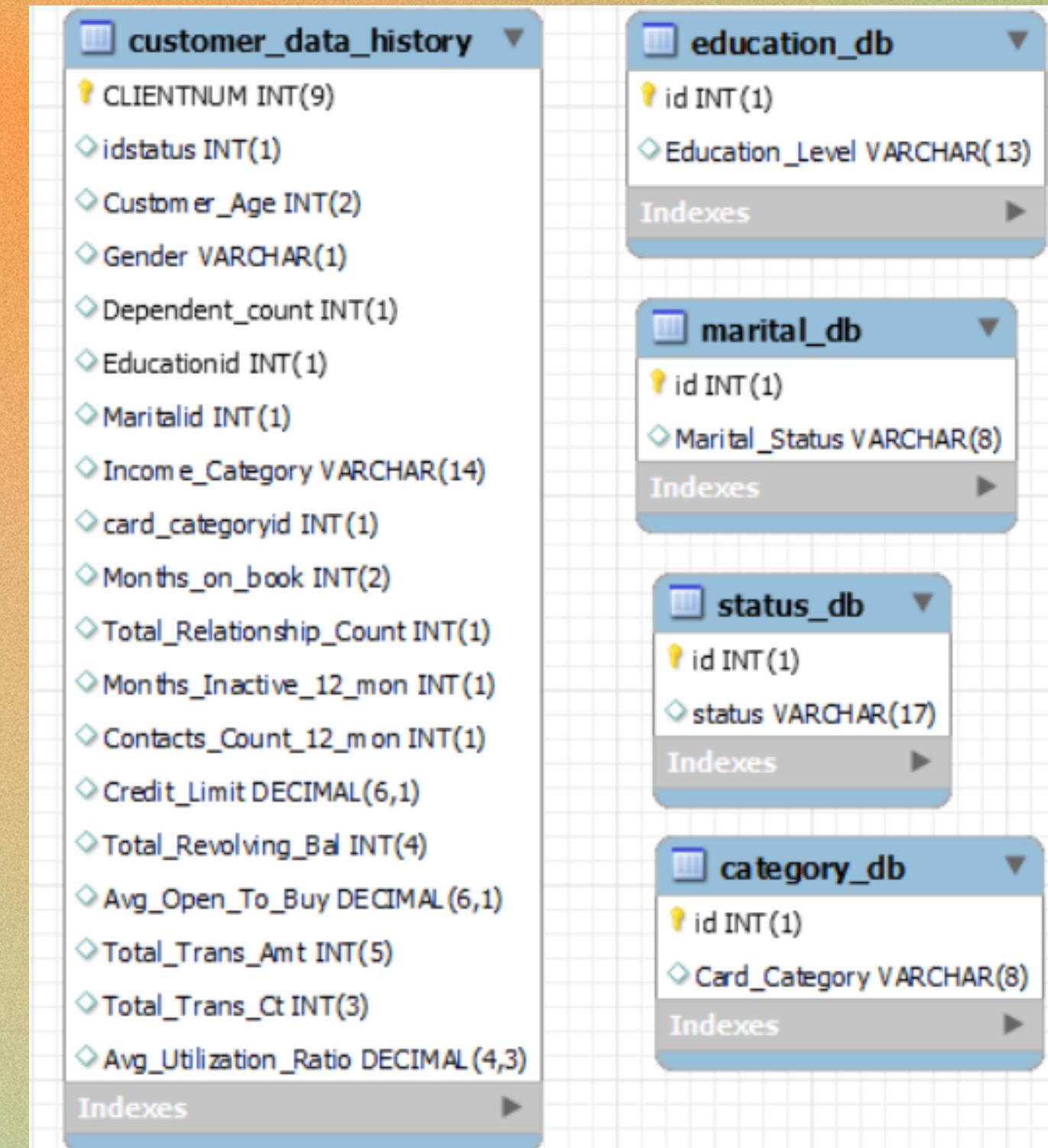
AVG\_UTILIZATION\_RATIO : AVERAGE CREDIT CARD USAGE RATIO

DATABASE CATEGORY\_DB = DATA CATEGORY OF THE CREDIT CARD SERVICE USED

DATABASE EDUCATION\_DB = EDUCATION CUSTOMER LEVEL DATA

MARRIAGE DATABASE\_DB = MARRIAGE CUSTOMER STATUS DATA

DATABASE STATUS\_DB = DATA ON THE STATUS OF EXISTING/DRESSED CUSTOMERS



# BUSINESS OBJECTIVE

1. OUR PROJECT AIMS TO IDENTIFY WHY CUSTOMERS ARE LEAVING OUR CREDIT CARD SERVICES BY ANALYZING FACTORS SUCH AS EDUCATION LEVEL, MARITAL STATUS, AGE, GENDER, DEPENDENT COUNT, AND INCOME CATEGORY.
2. WE WILL DEVELOP PREVENTATIVE MEASURES TO RETAIN CUSTOMERS WHO MAY BE AT RISK OF LEAVING BY USING INSIGHTS GAINED FROM OUR ANALYSIS.
3. OUR GOAL IS TO INCREASE CREDIT CARD USAGE AND TRANSACTION FREQUENCY BY IDENTIFYING THE TRANSACTION TOTAL, TRANSACTION FREQUENCY, AND UTILIZATION RATIO OF CUSTOMERS.
4. WE WILL ANALYZE CUSTOMER CREDIT CARD INFORMATION, SUCH AS CARD CATEGORY, CONTACT COUNT, TOTAL RELATIONSHIP COUNT, CREDIT LIMIT, AND REVOLVING BALANCE, TO BETTER UNDERSTAND OUR CUSTOMER BASE AND IMPROVE OUR SERVICES.

BY ACHIEVING THESE GOALS, WE AIM TO IMPROVE OUR CUSTOMER RETENTION RATES AND ULTIMATELY DRIVE BUSINESS GROWTH.

## COMMON TABLE EXPRESSION (CTE)

```

WITH TABEL_AGGREGAT 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 FP_BTPN_SYARIAH.CUSTOMER_DATA_HISTORY CDH
    LEFT JOIN FP_BTPN_SYARIAH.CATEGORY_DB CD ON CDH.CARD_CATEGORYID = ID_CATEGORY
    LEFT JOIN FP_BTPN_SYARIAH.EDUCATION_DB EDU ON CDH.EDUCATIONID = ID_EDUCATION
    LEFT JOIN FP_BTPN_SYARIAH.MARITAL_DB MD ON CDH.MARITALID = ID_MARITAL
    LEFT JOIN FP_BTPN_SYARIAH.STATUS_DB ST ON CDH.IDSTATUS = ID_STATUS
)

```

```

11      ELSE '>60 years'
12  END AS age_category
13  FROM fp_btpn_syariah.customer_data_history cdh
14  LEFT JOIN fp_btpn_syariah.category_db cd ON cdh.card_categoryid = id_category
15  LEFT JOIN fp_btpn_syariah.education_db edu ON cdh.Educationid = id_education
16  LEFT JOIN fp_btpn_syariah.marital_db md ON cdh.Maritalid = id_marital
17  LEFT JOIN fp_btpn_syariah.status_db st ON cdh.idstatus = id_status
18 ) SELECT * FROM tabel_aggregat;
19
20

```

vg_Open_To_Buy	Total_Trans_Amt	Total_Trans_Ct	Avg_Utilization_Ratio	id_category	Card_Category	id_education	Education_Level	id_marital	Marital_Status	id_status	status	age_category
6.0	1336	30	0.788	1	Blue	1	High School	1	Married	1	Existing Customer	>60 years
90.0	931	18	0.152	1	Blue	1	High School	1	Married	1	Existing Customer	41-45 years
23.0	1045	38	0.285	1	Blue	1	High School	1	Married	1	Existing Customer	46-50 years
06.0	1756	32	0.000	1	Blue	1	High School	1	Married	1	Existing Customer	46-50 years
023.0	1217	27	0.074	1	Blue	1	High School	1	Married	1	Existing Customer	46-50 years
2.0	1946	38	0.687	1	Blue	1	High School	1	Married	1	Existing Customer	>60 years
61.0	1553	38	0.355	1	Blue	1	High School	1	Married	1	Existing Customer	51-55 years
426.0	1659	38	0.150	1	Blue	1	High School	1	Married	1	Existing Customer	51-55 years
62.0	1303	39	0.118	1	Blue	1	High School	1	Married	1	Existing Customer	41-45 years
28.0	1651	28	0.359	1	Blue	1	High School	1	Married	1	Existing Customer	56-60 years
38.3	2184	53	0.000	1	Blue	1	High School	1	Married	1	Existing Customer	51-55 years

USING CTE TO JOINING 4 TABLES WITH THE CUSTOMER\_DATA\_HISTORY AND MAKE AGE CATEGORY

# DATA EXPLORATION

THERE IS NONE DUPLICATE  
DATA

```
SELECT count(DISTINCT CLIENTNUM) FROM tabel_aggregat;
```

```
SELECT count(CLIENTNUM) FROM tabel_aggregat;
```

	count(DISTINCT CLIENTNUM)
▶	10127

	count(CLIENTNUM)
▶	10127

GRADUATE IS THE LEADING  
GROUP

	Education_Level	count(Education_Level)
▶	Graduate	3128
	High School	2013
	Unknown	1519
	Uneducated	1487
	College	1013
	Post-Graduate	516
	Doctorate	451

```
SELECT
Education_Level, count(Education_Level)
FROM tabel_aggregat
GROUP BY
Education_Level
ORDER BY count(Education_level) desc;
```

MARRIED IS THE LEADING  
GROUP

	Marital_Status	count(Marital_Status)
▶	Married	4687
	Single	3943
	Unknown	749
	Divorced	748

```
SELECT
Marital_Status,
count(Marital_Status)
FROM tabel_aggregat
GROUP BY
Marital_Status
ORDER BY
count(Marital_Status) DESC;
```

# DATA EXPLORATION

## OVERVIEW CUSTOMER AGE

41-50 YEARS IS THE LEADING GROUP

FEMALE IS THE LEADING GROUP

<code>min(customer_age)</code>	<code>max(customer_age)</code>	<code>Avg(customer_age)</code>	<code>count(DISTINCT customer_age)</code>
26	73	46.3260	45

```

SELECT
    min(customer_age), max(customer_age), Avg(customer_age), count(DISTINCT customer_age)
FROM tabel_aggregat;
  
```

	age_category	count(age_category)
▶	46-50 years	2388
	41-45 years	2264
	51-55 years	1747
	36-40 years	1478
	56-60 years	926
	31-35 years	654
	>60 years	405
	30 years below	265

```

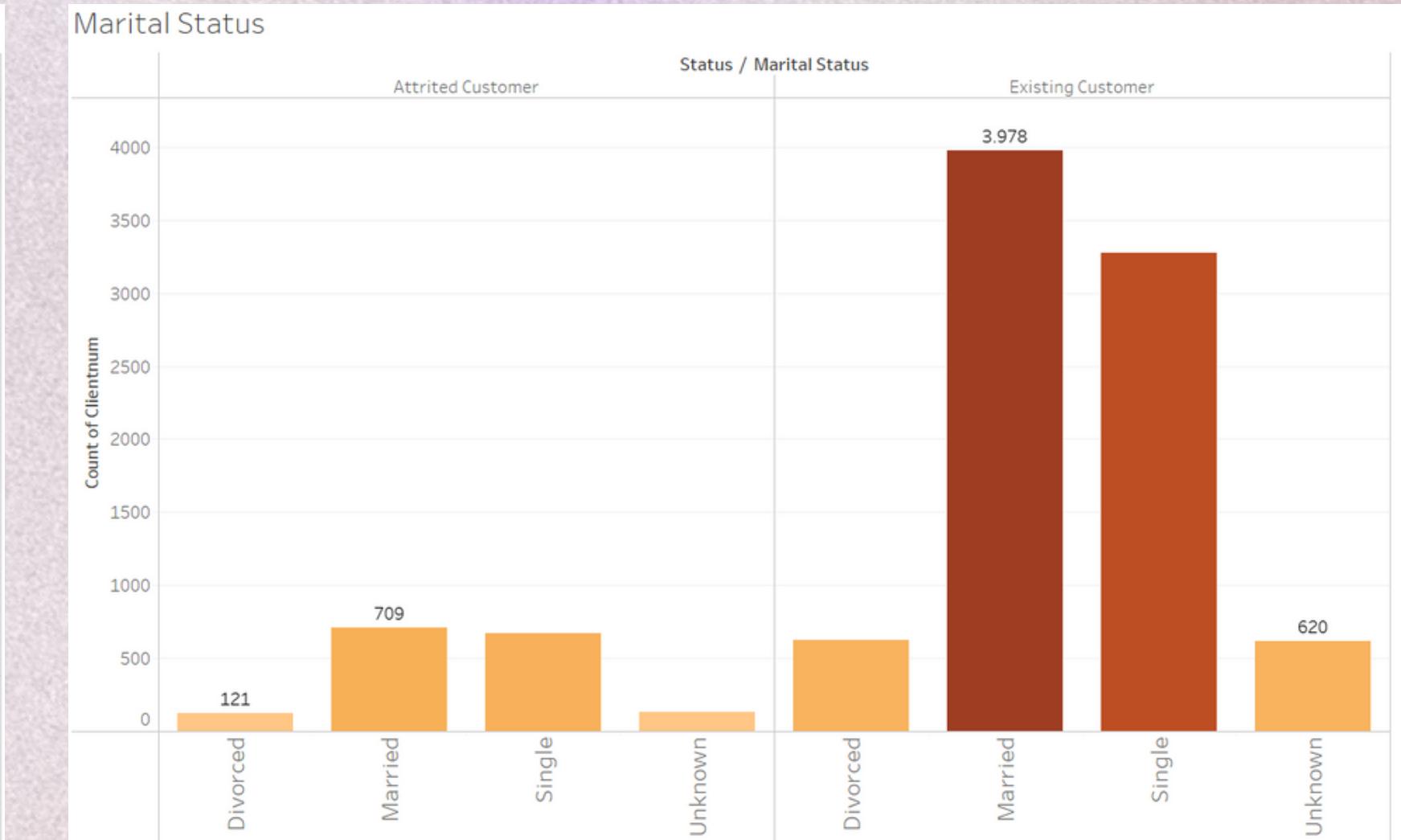
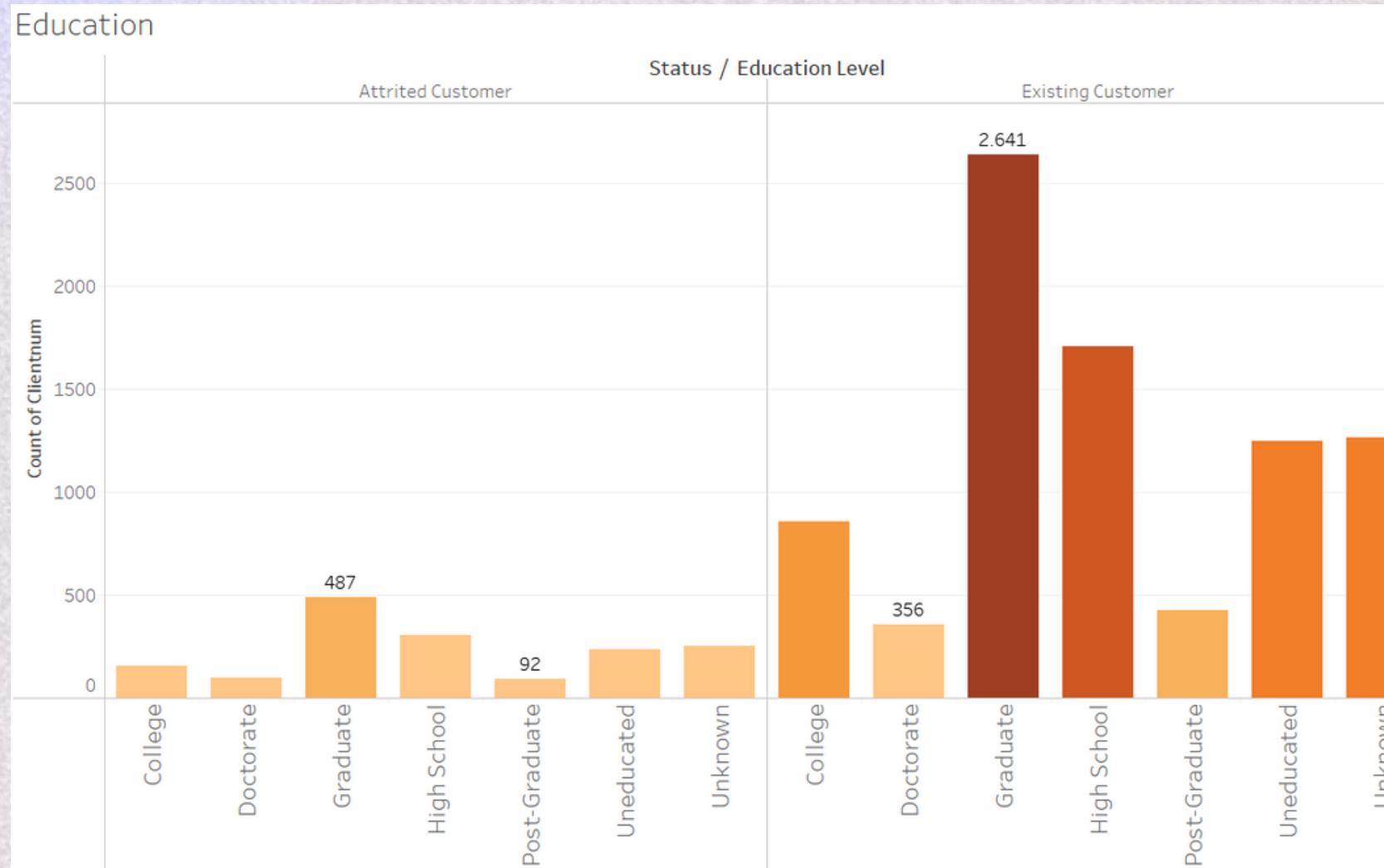
SELECT
    age_category,
    count(age_category)
FROM tabel_aggregat
GROUP BY age_category
ORDER BY count(age_category) DESC;
  
```

	Gender	count(Gender)
	F	5358
	M	4769

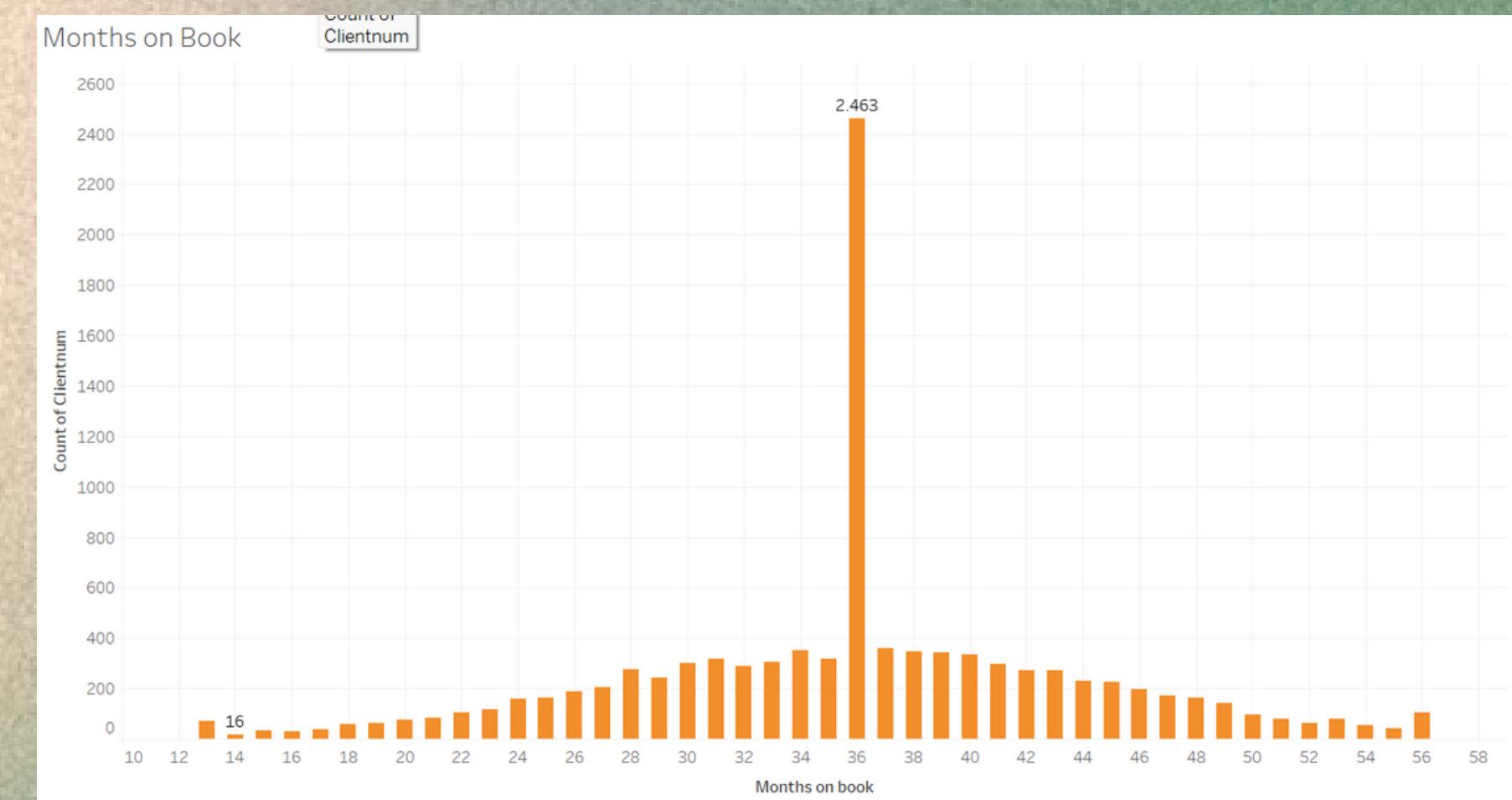
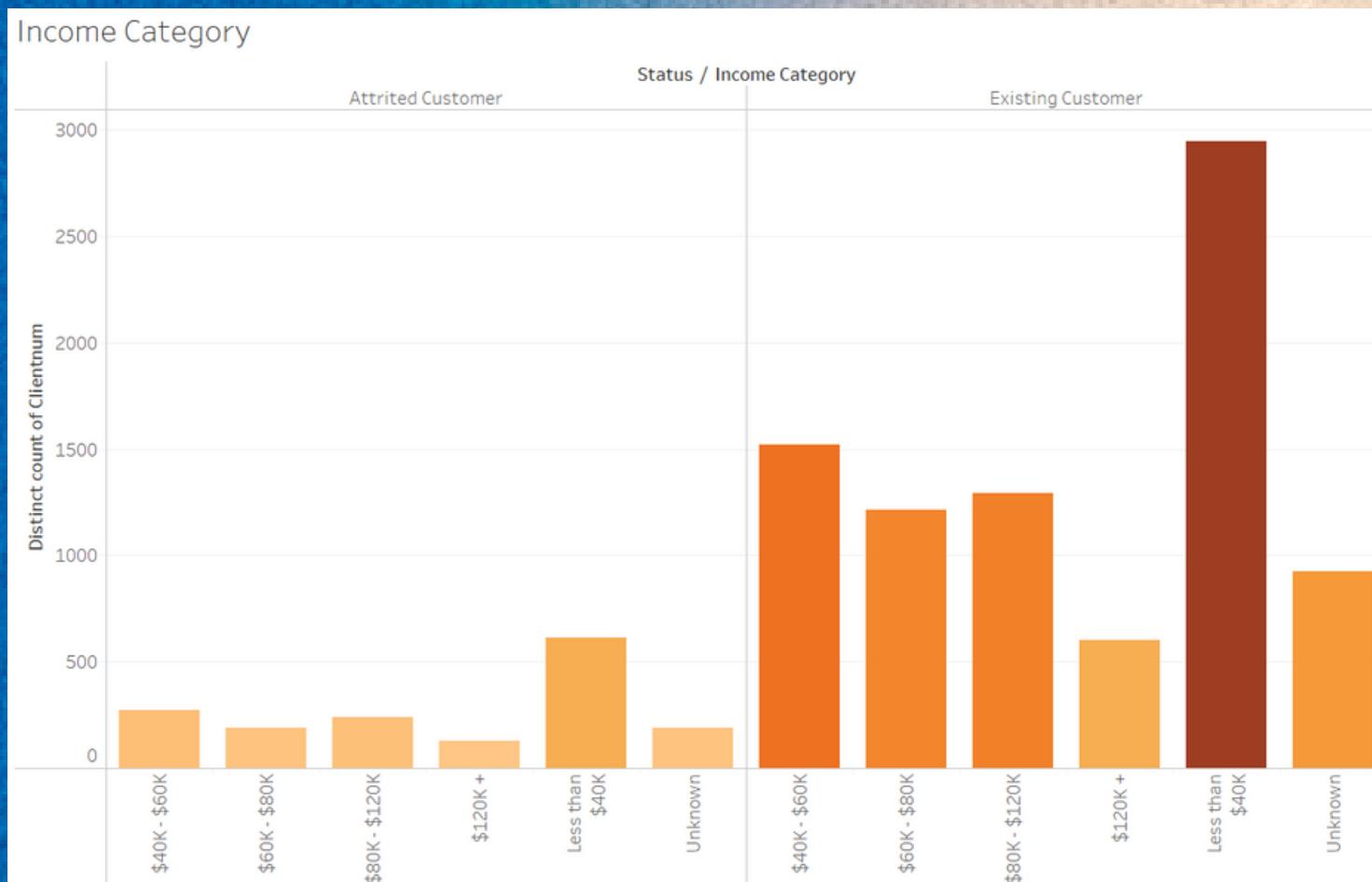
```

SELECT Gender, count(Gender)
FROM tabel_aggregat
GROUP BY Gender
ORDER BY count(Gender) DESC;
  
```

## BASED ON EDUCATIONAL AND MARITAL STATUS, GRADUATED AND MARRIED ARE THE LEADING GROUP

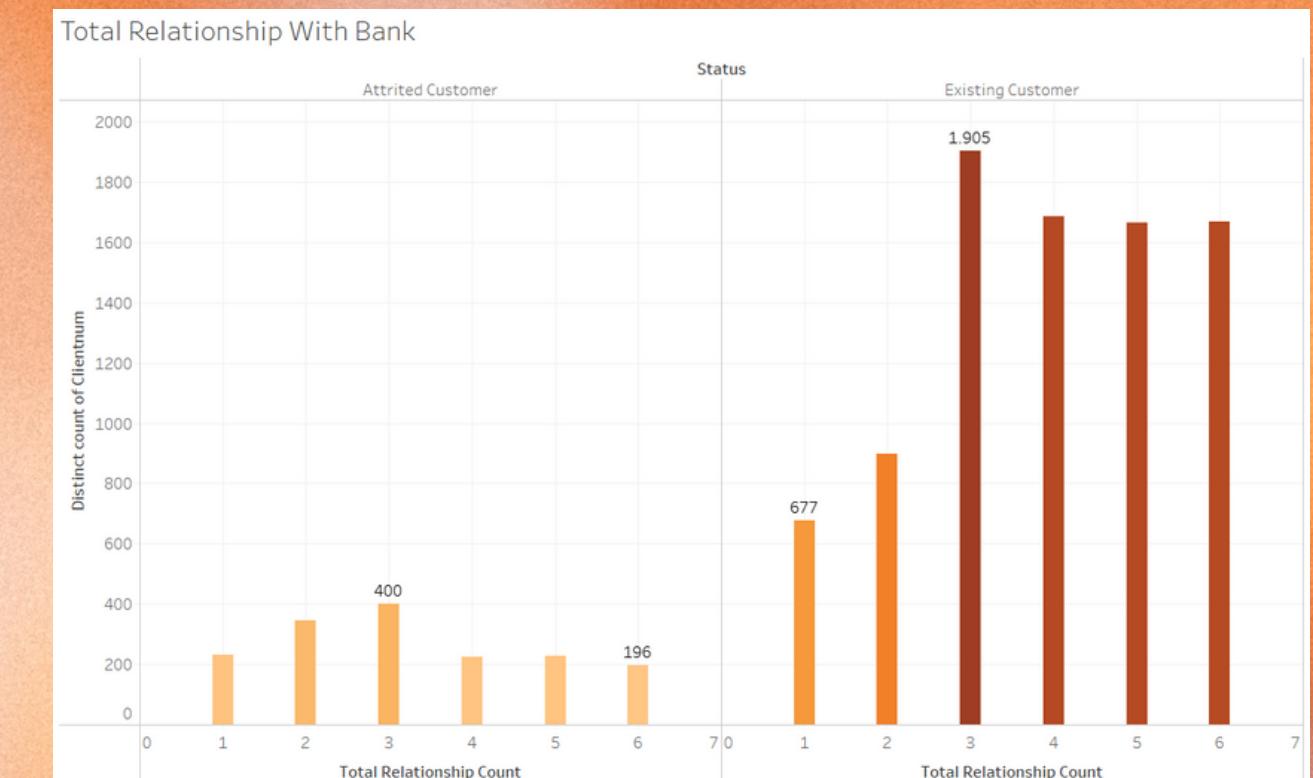
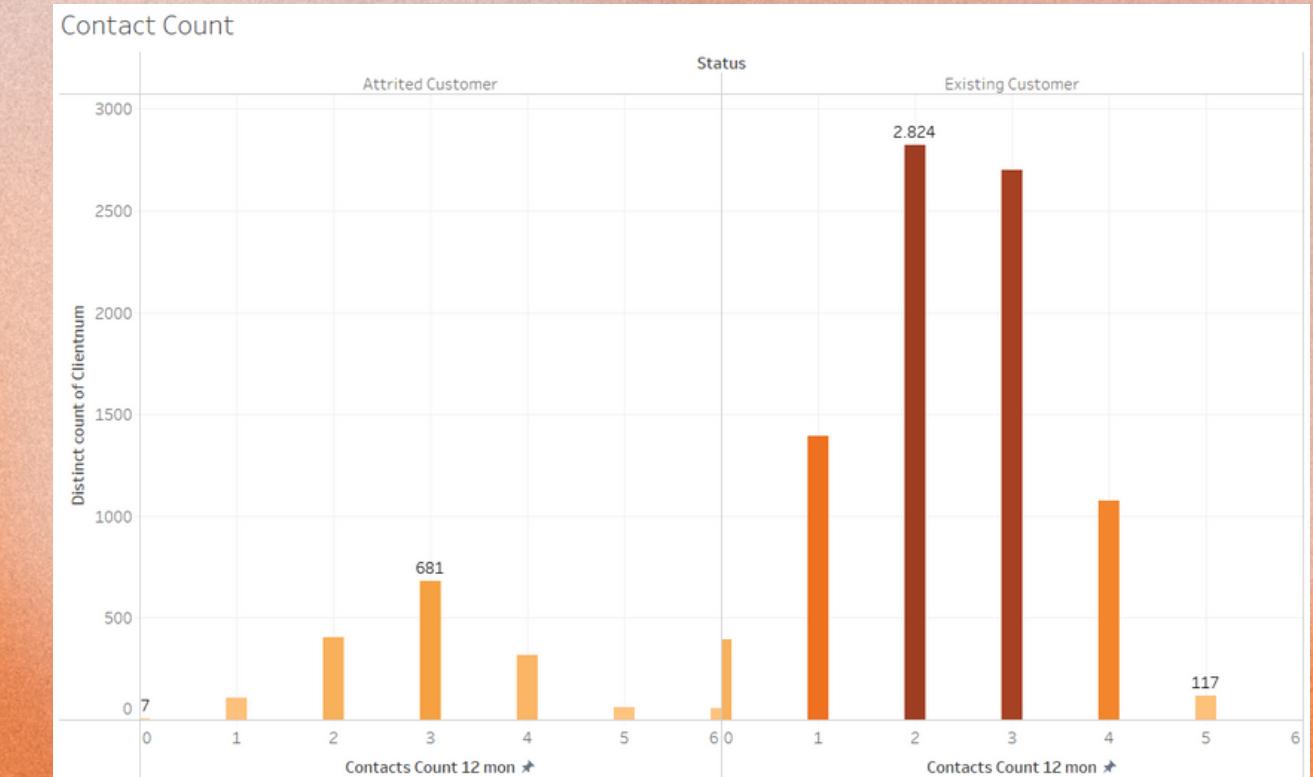
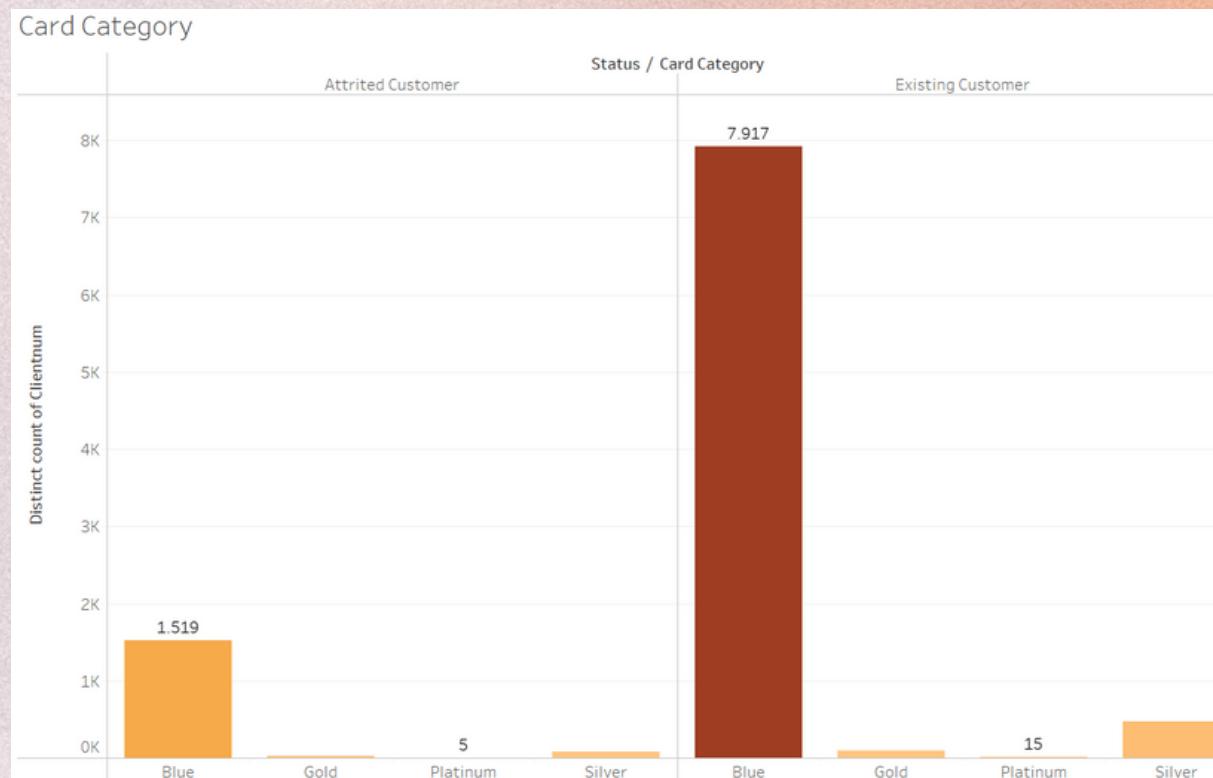


BY INCOME CATEGORY AND MONTHS ON BOOK,  
CUSTOMER THAT HAS LESS THAN \$40K INCOME AND HAS THE 36-  
MONTHS MEMBERSHIP IS THE LEADING GROUP



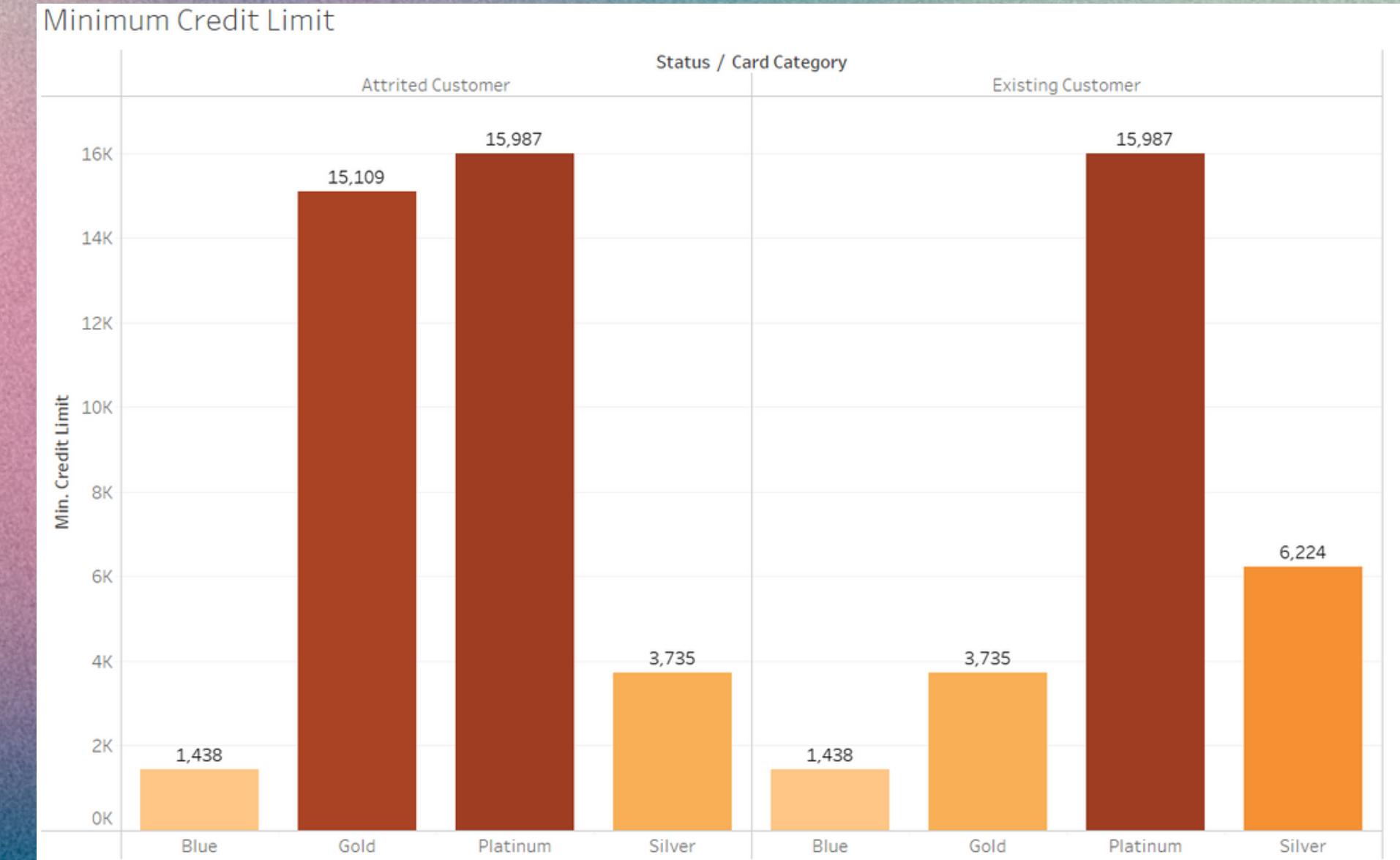
# DATA EXPLORATION

1. CUSTOMER CONTACT: BANKS USUALLY CONTACT EXISTING CUSTOMERS 5 TIMES MORE OFTEN THAN NEW CUSTOMERS. MOST EXISTING CUSTOMERS ARE CONTACTED BY THE BANK ONLY TWICE, INDICATING A TARGETED APPROACH TO OUTREACH.
2. CUSTOMER RELATIONSHIP: MOST EXISTING CUSTOMERS HAVE 3 OR MORE PRODUCTS WITH THE BANK, SHOWING A STRONG LEVEL OF ENGAGEMENT AND TRUST.
3. BLUE CARD: THE MOST POPULAR PRODUCT AMONG EXISTING CUSTOMERS IS THE BLUE CARD, SUGGESTING THAT IT MEETS THEIR NEEDS WELL AND SHOULD CONTINUE TO BE PROMOTED.



# DATA EXPLORATION

- 1.CREDIT LIMIT: ATTRITED AND EXISTING CUSTOMERS HAVE SIMILAR AVERAGE CREDIT LIMITS.
- 2.REVOLVING BALANCE: EXISTING CUSTOMERS TEND TO USE THEIR CREDIT CARDS MORE THAN ATTRITED CUSTOMERS, WITH A HIGHER REVOLVING BALANCE.
- 3.GOLD CREDIT CARD: EXISTING CUSTOMERS HAVE A LOWER MINIMUM CREDIT LIMIT THAN ATTRITED CUSTOMERS FOR THE GOLD CREDIT CARD, WHICH MAY NEED TO BE REVIEWED TO ENSURE CUSTOMERS ARE GETTING THE CREDIT THEY NEED.



## Average Credit Limit

Status	
Attrited Customer	8,136.0
Existing Customer	8,726.9

## Average Revolving Balance

Status	
Attrited Customer	672.8
Existing Customer	1,256.6

## SUGESSTION

1. CONDUCT A THOROUGH ANALYSIS OF THE CUSTOMER REQUIREMENTS FOR A 36-MONTH CREDIT PLAN AND THE BLUE CARDS PROGRAM, ENSURING THAT THESE OFFERINGS ARE VIABLE FOR THE CUSTOMER BASE.
2. EVALUATE THE CREDIT LIMIT FOR EACH CARD, WITH A FOCUS ON MAINTAINING THE SAME LIMIT FOR GOLD AND PLATINUM CARDS WHILE MONITORING PERFORMANCE OVER A 3-MONTH PERIOD.
3. DEVELOP ADDITIONAL PROGRAMS FOR CUSTOMERS WHO UTILIZE MORE THAN THREE PRODUCTS, WITH THE GOAL OF INCREASING THEIR ENGAGEMENT AND LOYALTY.
4. IMPLEMENT A TIERED INTEREST RATE SYSTEM BASED ON CUSTOMER INCOME, PROVIDING LOWER INTEREST RATES FOR CUSTOMERS WITH AN INCOME BELOW \$40,000 AND FOR THE CUSTOMER THAT ALREADY MARRIED.
5. RESTRICT CUSTOMER SERVICE CONTACT TO A MAXIMUM OF THREE ATTEMPTS, AS RESEARCH INDICATES THAT EXCESSIVE CONTACT CAN LEAD TO CUSTOMER CHURN. BY MINIMIZING BANK OUTREACH, THE COMPANY CAN MAINTAIN A LOW CHURN RATE AND ENSURE THAT CUSTOMERS FEEL RESPECTED AND VALUED.

1. INCOME AND NUMBER OF DEPENDENTS IMPACT CUSTOMER ATTRITION: THE ANALYSIS SUGGESTS THAT CUSTOMERS WITH LOWER INCOME AND MORE DEPENDENTS ARE MORE LIKELY TO LEAVE. THIS IS BECAUSE THEY MAY HAVE DIFFICULTY PAYING THEIR CREDIT CARD BILLS, LEADING TO MISSED PAYMENTS AND LATE FEES. TO ADDRESS THIS, THE COMPANY SHOULD CONSIDER SCREENING CUSTOMERS' FINANCIAL PROFILES MORE CAREFULLY TO MINIMIZE THE RISK OF DEFAULT.
2. FILTER CREDIT APPROVALS TO MINIMIZE RISK OF DEFAULT: TO PREVENT CUSTOMERS FROM LEAVING DUE TO PAYMENT DIFFICULTIES, THE COMPANY SHOULD CONSIDER SETTING MORE CONSERVATIVE CREDIT LIMITS AND CONDUCTING MORE THOROUGH CREDIT CHECKS. BY DOING SO, IT CAN REDUCE THE RISK OF CUSTOMERS DEFAULTING ON THEIR PAYMENTS AND IMPROVE OVERALL CUSTOMER RETENTION.
3. PROVIDE PROMOTIONS TO MAINTAIN CUSTOMER LOYALTY: TO ENCOURAGE CUSTOMERS TO REMAIN LOYAL AND CONTINUE USING THEIR CREDIT CARDS, THE COMPANY SHOULD OFFER TARGETED PROMOTIONS AND REWARDS PROGRAMS. THIS CAN INCENTIVIZE CUSTOMERS TO USE THEIR CARDS MORE FREQUENTLY, BOOSTING REVENUE AND IMPROVING CUSTOMER SATISFACTION.