

oooo

ARVIO ANANDI



BANK CUSTOMER CHURN ANALYSIS

oooo

TABLE OF CONTENTS

- Executive Summary
- Introduction
- Data Wrangling
- EDA
- Data Visualization
- Key Findings and Insights
- Conclusion & Recommendations



EXECUTIVE SUMMARY

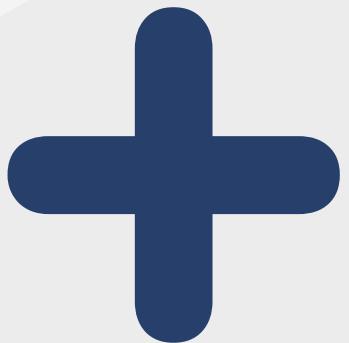
BUSINESS BACKGROUND

I've taken on the role of a freelance data analyst, partnering with a prominent EU bank. This bank is renowned for its extensive customer base and innovative financial solutions. Their aim is simple yet crucial: reduce customer churn. The challenge is to understand why customers are leaving and devise modest, effective solutions to retain them. This project is about enhancing customer experiences and fostering loyalty, all while staying true to the bank's values and reputation.

BUSINESS OBJECTIVES

The core objectives of this analysis project are to uncover why customers are leaving the bank, pinpoint which customer segments are most at risk of churn, and provide actionable strategies to retain customers while enhancing overall satisfaction. This project strives to strengthen customer relationships, reduce churn rates, and empower data-driven decision-making within the bank.

OUTPUT



The output of this project is a dashboard to visualize the customer churn data



INTRODUCTION

DATA SOURCE

The Kaggle logo is displayed in a large, bold, blue sans-serif font. The word "kaggle" is written in lowercase, with a small blue trademark symbol (TM) positioned above the letter "e".

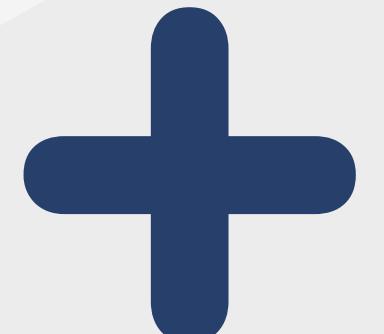
kaggle™



The dataset was acquired from Kaggle, with 10000 rows and 18 columns.

COLUMN EXPLANATION

	Column Name	Description	Relevance to Churn
0	RowNumber	Record number, no impact on churn	Not Relevant
1	CustomerId	Random values, no impact on churn	Not Relevant
2	Surname	Customer surname, no impact on churn	Not Relevant
3	CreditScore	Credit score's impact on churn	Relevant
4	Geography	Customer location's impact on churn	Relevant
5	Gender	Gender's impact on churn	Relevant
6	Age	Age's impact on churn	Relevant
7	Tenure	Number of years as a client, impact on churn	Relevant
8	Balance	Impact on churn based on account balance	Relevant
9	NumOfProducts	Number of products owned by the customer	Relevant
10	HasCrCard	Impact on churn based on credit card ownership	Relevant
11	IsActiveMember	Impact on churn based on customer activity	Relevant
12	EstimatedSalary	Impact on churn based on estimated salary	Relevant
13	Exited	Whether the customer left the bank (churn)	Relevant
14	Complain	Customer complaint status	Relevant
15	Satisfaction Score	Customer satisfaction score for complaint reso...	Relevant
16	Card Type	Type of card held by the customer	Relevant
17	Points Earned	Points earned by the customer for using a cred...	Relevant



DATA WRANGLING

```
11    -- delete missing values
12 •  delete
13    from customer.`customer-churn-records`
14    where RowNumber is NULL
15    or CustomerId is NULL
16    or Surname is NULL
17    or CreditScore is NULL
18    or Geography is NULL
19    or Age is NULL
20    or Tenure is NULL
21    or Balance is NULL
22    or NumOfProducts is NULL
23    or HasCrCard is NULL
24    or IsActiveMember is NULL
25    or EstimatedSalary is NULL
26    or Exited is NULL
27    or Complain is NULL
28    or SatisfactionScore is NULL
29    or CardType is NULL
30    or PointEarned is NULL;
```

Removed missing values

```
32      -- check for duplicate values (there are none)
33 •   select
34     (select distinct count(*) from customer.`customer-churn-records`) as distinct_values,
35     (select count(*) from customer.`customer-churn-records` where count(*)
36       not in (select distinct count(*) from customer.`customer-churn-records`)) as non_distinct_values;
37
```

Result Grid | Filter Rows: Export: Wrap Cell Content:

distinct_values	non_distinct_values
10000	10000

There seems to be no duplicates as the distinct values and non distinct value comparison is the same

```
40      -- delete unneeded features
41 • alter table customer.`customer-churn-records`
42      drop column RowNumber
43      , drop column CustomerId
44      , drop column Surname;
45
```

Delete unneeded features which are rownumber, customer id, and surname as they don't contribute much to this project.

```
91      -- Add the Age_Segment column to the table and populate it using CASE statement
92 • alter table customer.`customer-churn-records`
93   add column Age_Segment varchar(50)
94   generated always as (
95     case
96       when Age < 30 then 'Under 30'
97       when Age between 30 and 39 then '30-39'
98       when Age between 40 and 49 then '40-49'
99       when Age between 50 and 59 then '50-59'
00       else '60 and Over'
01     end
02   ) stored;
```

Feature engineer a new feature called Age_segment to segment the age groups to 5 groups (Under 30, 30-39, 40-49, 50-59, and 60 and Over).

```
104    -- Add the creditscore_segment column to the table and populate it using CASE statement
105 • alter table customer.`customer-churn-records`
106   add column CreditScore_segment varchar(50)
107   generated always as (
108     case
109       when creditscore < 500 then 'Very Low'
110       when creditscore between 500 and 599 then 'Low'
111       when creditscore between 600 and 699 then 'Medium'
112       when creditscore between 700 and 799 then 'High'
113       else 'Very High'
114     end
115   ) stored;
116 }
```

Feature engineer a new feature called Creditscore_segment to segment the CreditScore to group of 5 categories (Very Low, Low, Medium, High, and Very High).

```
117    -- Add the EstimatedSalary_segment column to the table and populate it using CASE statement
118 • alter table customer.`customer-churn-records`
119 add column Salary_segment varchar(50)
120     generated always as (
121         case
122             when EstimatedSalary < 60000 then 'Low Income'
123             when EstimatedSalary between 60000 and 140000 then 'Middle Income'
124             when EstimatedSalary > 140000 then 'High Income'
125         end
126     ) stored;
```

::::: Feature engineered a new feature called Salary_segment to segment the Salary to groups of 3 categories (Low Income, Middle Income, and High Income).

The screenshot shows the Power BI Power Query Editor interface. On the left is a table with 11 rows of data. The columns are labeled: NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary, Exited, and Complain. The IsActiveMember column contains values like 1, 0, Yes, and No. The EstimatedSalary column contains numerical values. The Exited and Complain columns contain binary values (0 or 1). The formula bar at the top of the editor window displays the formula: = Table.ReplaceValue(#"Changed Type", "1", "Yes", Replacer.ReplaceText, {"IsActiveMember"}). To the right of the table is the 'Query Settings' pane, which includes sections for 'PROPERTIES' (Name: customer customer-churn-records) and 'APPLIED STEPS'. The 'APPLIED STEPS' section lists several steps, with the 'Replaced Value' step highlighted, indicating it was used to change the value '1' to 'Yes' in the IsActiveMember column.

In Power BI, using the Power Query Transform, i changed the values of column IsActiveMember from 1 and 0s to Yes and No. This was done to create clearer labels for visualization later.

EXPLORATORY DATA ANALYSIS

```
46      -- check for churn ratio
47 •   select Exited, count(*) as count
48   from customer.`customer-churn-records`
49   group by Exited;
50
```

Result Grid | Filter Rows: Export:

	Exited	count
1	2038	
0	7962	

This table shows that there seems to be data imbalance for churned and not churned customers. This explains that roughly 80% of customers chose not to churn from the bank.

```
51    -- checking on geographical correlation
52 • select
53     Geography
54     , avg(Age) as Avg_Age
55     , count(*) as Total_Customers
56     , sum(EstimatedSalary) as Total_Salary
57     , sum(Exited) as churned
58 from
59     customer.`customer-churn-records`
60 group by
61     Geography;
62
```

Result Grid | Filter Rows: Export: Wrap Cell Content:

Geography	Avg_Age	Total_Customers	Total_Salary	churned
France	38.5118	5014	500894492.6000007	811
Spain	38.8910	2477	246314297.53999978	413
Germany	39.7716	2509	253693608.6700002	814

The table shows the distribution of churned customers from different countries. This tells us that people from France and Germany are more likely to churn than people from Spain.

```
63      -- checking on gender's correlation
64 •   select
65       Gender
66 , avg(Age) as Avg_Age
67 , count(*) as Total_Customers
68 , sum(EstimatedSalary) as Total_Salary
69 , sum(Exited) as churned
70 from
71     customer.`customer-churn-records`
72 group by
73     Gender;
```

Result Grid | Filter Rows: Export: Wrap Cell Content:

Gender	Avg_Age	Total_Customers	Total_Salary	churned
Female	39.2384	4543	457032802.50000006	1139
Male	38.6582	5457	543869596.3100008	899

From the given dataset, female customers tend to churn more than male customers. This could provide some insights and be explored further.

```
76 •  select
77   case
78     when Age < 30 then 'Under 30'
79     when Age between 30 and 39 then '30-39'
80     when Age between 40 and 49 then '40-49'
81     when Age between 50 and 59 then '50-59'
82     else '60 and Over'
83   end as Age_Group,
84   SUM(Exited) / COUNT(*) AS churned_ratio
85   FROM
86     customer.`customer-churn-records`
87   GROUP BY
88     age_group
89   ORDER BY
90     churned_ratio;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Age_Group	churned_ratio
Under 30	0.0756
30-39	0.1088
60 and Over	0.2795
40-49	0.3083
50-59	0.5604

::::::::::
People from the youngest and oldest group of age have a fair share for not churning. For the young group, this could be explained by their digitalized environment where information are easily consumed which explains their financial literacy that helps to determine their long-term bank decisions whereas for the seniors, this could be explained by a more stable financial decisions, and their long-term relationships with banks which corresponds to them not switching banks

DATA VISUALIZATION



This dashboard is created using Microsoft Power BI and Power Query Transform. You can slice the data using the slicers provided. The two slicers were provided to avoid visualization misunderstandings.

KEY FINDINGS & INSIGHTS

1. Customers with **very high** and **very low credit score** tend **not** to churn.
2. Customers who **rarely complain** about bank products tend **not** to churn.
3. Customers who subscribes to **more bank products** tend **not** to churn.
4. Customers with **Low** and **High Income** tend **not** to churn.
5. **The youngest** and **oldest group** of customers tend **not** to churn.
6. **Males** are **less likely** to churn than females.
7. **Active members** are **less likely** to churn than non-active members.
8. Customers from **Spain** are **less likely** to churn than customers from Germany or France.

CONCLUSION & RECOMMENDATIONS

CONCLUSION

The analysis has provided valuable insights into customer churn patterns for the bank.

Customers with extreme credit scores, low complaint frequency, higher product subscriptions, diverse income levels, and belonging to specific age groups, genders, or active member status are less likely to churn. Understanding these patterns is crucial because retaining existing customers is more cost-effective than acquiring new ones. Recognizing the factors that influence customer decisions to leave allows the bank to develop proactive strategies for churn prevention.

RECOMMENDATIONS

1. Tailored Loyalty Programs: Develop personalized loyalty programs based on customer segments. Offer incentives and rewards to high-churn-risk groups to encourage their loyalty.
2. Improved Customer Service: Enhance customer service, especially for customers who frequently complain. Swift and effective complaint resolution can significantly improve customer satisfaction and retention rates.
3. Product Bundling: Encourage customers to subscribe to multiple bank products by offering bundled services. Create packages that cater to various needs, providing an incentive for customers to stay.
4. Income-Adjusted Services: Customize services based on income levels. Offer financial products that align with the financial capacity of customers, ensuring affordability and value for money.
5. Age-Specific Engagement: Understand and address the unique requirements of different age groups. Develop age-specific products and services, ensuring relevance and appeal across generations.

RECOMMENDATIONS

- 6. Gender-Inclusive Services: Create banking services that address the specific needs of both genders. Recognize diverse financial goals and preferences, making the bank more inclusive.
- 7. Active Member Recognition: Acknowledge and reward active members. Offer exclusive services, discounts, or privileges to keep active members engaged and loyal.
- 8. Cultural Sensitivity: Recognize and respect cultural differences, especially for customers from different countries. Offer services that resonate with their cultural values, building trust and customer loyalty.
- 9. Regular Feedback Mechanism: Implement a systematic feedback system. Regularly collect feedback from customers to understand their evolving needs and concerns, allowing the bank to adapt its strategies accordingly.
- 10. Data-Driven Decisions: Invest in advanced analytics and AI tools to continually analyze customer behavior. Predictive models can identify potential chasers early, allowing the bank to take proactive measures.



THANK YOU



CONTACTS

+6287764260779



arvioanandi@gmail.com

PROJECT LINK



[Bank Customer Churn Analysis](#)

