Sol Bank

CREDIT DATA
DOCUMENTATION

Prepared By :

Abdallah Khaled
Amr Tamer
Youssef Elsayed
Asma Magdy
Shahd Mohamed
Habiba Mohsen

Supervised By :

Dr. Jakie Kamal

Contents:

1. Data source:

The data source for our Credit database originates from Kaggle, a prominent online community for data scientists and machine learning practitioners. Kaggle provides a vast collection of datasets contributed by data enthusiasts from around the world.

https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers

By utilizing this Kaggle dataset, we have access to a wealth of data relating to various aspects of the Credit Sector in a bank. This includes information about CLIENTNUM, Attrition Flag, Customer Age,, Gender, Dependent count, Education Level, Marital Status, Income Category, Card Category, Months on book, Months Inactive 12 months, Contacts Count 12 months, Credit Limit, Total Revolving Balance and the Average Utilization Ratio Through the integration of this diverse data, we aim to provide valuable insights and analysis that can assist in understanding the dynamics of the Credit Sector

2. Business Problem:

The main business problem in this project is "*Which customer segments are most likely to take a loan, and how can the bank target them more effectively?*"

To Answer that we must run the data through multiple phases which include (but not limited to): Data Cleaning, Loan Generation (as there is no loan variable), Data Visualization, and Data Analysis.

3. Dataset components:

The dataset consists of 20 columns:

1. CLIENTNUM : Unique identifier for the client.
2. Attrition Flag : Indicates the customer status: Existing Customer or Attired Customer (turned over).
3. Customer Age: The age of the customer in years.
4. Gender: Demographic variable - M=Male, F=Female
5. Dependent Count: Demographic variable - Number of dependents
6. Educational Level : The customer's highest level of education.
7. Marital Status: The customer's marital status.
8. Income Category: The customer's annual income bracket.
9. Card Category: The type or tier of credit card held by the customer
10. Months_on_book: The duration (in months) of the customer relationship with the bank.
11. Total Relationship count: Total no. of products held by the customer
12. Months Inactive 12 months: Number of months the customer was inactive in the last 12 months.
13. Contacts Count 12 months: Number of times the customer contacted the bank in the last 12 months.

14. Credit Limit: The customer's total credit line or limit.
15. Total Revolving Balance: The total revolving balance on the credit card account

16. Average Utilization Ratio : The average credit card utilization ratio

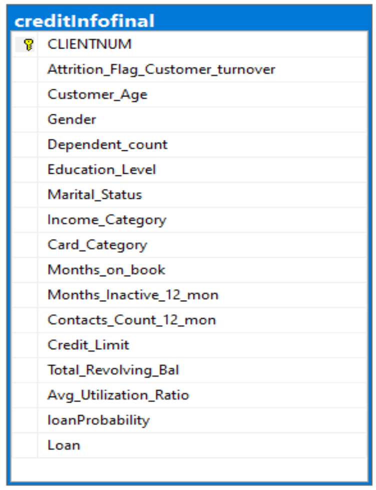
17. Average Open to Buy : Open to Buy Credit Line

18. Total Amount Change Q4-Q1: Change in Transaction Amount (Q4 over Q1)

19. Total Transactions Amount: Total Transaction Amount (Last 12 months)

20. Total Transaction Count: Total Transaction Count (Last 12 months)

4. Database schema :



5. Python:

We chose python for its libraries which will help us with Data Cleaning, Loan Output Generation, and ML Models.

<u>Libraries</u>:

- Pandas for reading, editing, and saving .CSV files
- Numpy for (bounded) random probability generation.
- Matplotlib & Seaborn for data visualization
- gdown and os for data loading.
- Sklearn for data scaling, train/validation/test data splitting, cross validation score, KNN, Logistic Regression, and model metrics.

5.1 Data Cleaning:

Upon loading the data from Google Drive, we began by doing an Exploratory Data Analysis to help us explore & understand the data and check for any missing/incorrect values.
After that we began by removing two irrelevant columns from our dataset, both beginning with "Naive_Bayes_Classifier……".
The rest of the variables were in a completely normal range and did not have any empty cells, outliers or any extreme values. So no further cleaning was required.

1.1 Exploratory Data Analysis (EDA) for `df_credit_info`

Descriptive Statistics for Numerical Features

```
print('Descriptive Statistics for Numerical Features:')
display(df_credit_info.describe())
```

Descriptive Statistics for Numerical Features:

	CLIENTNUM	Customer_Age	Dependent_count	Months_on_book	Total_Relationship_Count	Months_Inactive_12_mon	Contacts_Count_12_mon	Credit_Limit	Total_Revolving_Bal	Av
count	1.012700e+04	10127.000000	10127.000000	10127.000000	10127.000000	10127.000000	10127.000000	10127.000000	10127.000000	
mean	7.391776e+08	46.325960	2.346203	35.928409	3.812580	2.341167	2.455317	8631.953698	1162.814061	
std	3.690378e+07	8.016814	1.298908	7.986416	1.554408	1.010622	1.106225	9088.776650	814.987335	
min	7.080821e+08	26.000000	0.000000	13.000000	1.000000	0.000000	0.000000	1438.300000	0.000000	
25%	7.130368e+08	41.000000	1.000000	31.000000	3.000000	2.000000	2.000000	2555.000000	359.000000	
50%	7.179264e+08	46.000000	2.000000	36.000000	4.000000	2.000000	2.000000	4549.000000	1276.000000	
75%	7.731435e+08	52.000000	3.000000	40.000000	5.000000	3.000000	3.000000	11067.500000	1784.000000	
max	8.283431e+08	73.000000	5.000000	56.000000	6.000000	6.000000	6.000000	34516.000000	2517.000000	

5.2 Loan Output Generation:

Since this dataset has no 'Loan' feature while having everything else that fits a project with this business problem, we decided to generate the loan output ourselves using NumPy and Pandas while maintaining a good degree of random probability to not overfit the ML model and to maintain variation and realism within the dataset.

We began by defining two features; loanProbability and Loan.

Since we have no ‘Credit_Score’ feature, we’ve decided to use ‘credit_limit’ as an auxiliary variable to indicate whether a customer’s credit score is good or not. To do that we’ve defined the average credit limit for each credit card type.

```
avgCardLimit = df.groupby('Card_Category')['Credit_Limit'].mean()
avgCardLimit
```

Card_Category	Credit_Limit
Blue	7363.780002
Gold	28416.370690
Platinum	30283.450000
Silver	25277.836036

dtype: float64

The logic behind that is; if a customer with a credit card were to have a credit limit higher than the average credit limit for his/her card type, then it’s an indicator that they have a good credit score.

We’ve also decided to increase the chance of being given a loan for the following cases:
- The customer has 0-2 dependents
- The customer is Married
- The customer has bought products from the bank(linear relationship)
- The customer is from the working-force age demographic
- The customer has a higher income category
- The customer has a good education
- The customer has a considerably higher credit limit than the average for his credit card type

We’ve also kept the given probability random but bounded within a certain range to introduce variation and realism to this dataset.

```
    #now by credit card limit
    if row['Credit_Limit'] > 1.3*avgCardLimit[row['Card_Category']]:
      p+=np.random.uniform(0.15,0.25)
    elif row['Credit_Limit'] > 1.2*avgCardLimit[row['Card_Category']]:
      p+=np.random.uniform(0.07,0.12)
    elif row['Credit_Limit'] > 1.1*avgCardLimit[row['Card_Category']]:
      p+=np.random.uniform(0.02,0.07)

    #marital status: if ur married then ur more likely to get a loan
    if row['Marital_Status']=='Married':
      p+=np.random.uniform(0.05,0.09)
    #education level, self explanatory.
    if row['Education_Level']=='High School':
      p+=np.random.uniform(0.03, 0.05)
    elif row['Education_Level']=='College':
      p+=np.random.uniform(0.05, 0.1)
    elif row['Education_Level']=='Post-Graduate':
      p+=np.random.uniform(0.1, 0.16)
    elif row['Education_Level']=='Doctorate':
      p+=np.random.uniform(0.1, 0.2)

    return min(p,1)
```

The ‘Loan’ variable must be binary, representing whether a customer gets a loan (1) or not (0).

We treat it as a Bernoulli trial, which is a special case of the binomial distribution with a single trial (n=1).

For each customer we draw a random outcome based on their predicted loan probability (loan Probability).

5.3 Machine Learning Rationale & Business Impact:

5.3.1 Purpose of Loan Eligibility Prediction:

Predicting loan eligibility is a fundamental requirement in financial institutions due to risk management, operational efficiency, customer targeting, and regulatory compliance. Accurate prediction helps Sol Bank reduce default risk, accelerate decision-making, and improve customer segmentation.

5.3.2 Technical Motivation for Using Machine Learning:

Machine Learning models capture nonlinear relationships between customer behavior, demographic attributes, and financial parameters. Unlike static rule-based systems, ML models adapt, scale, and generalize across thousands of clients.

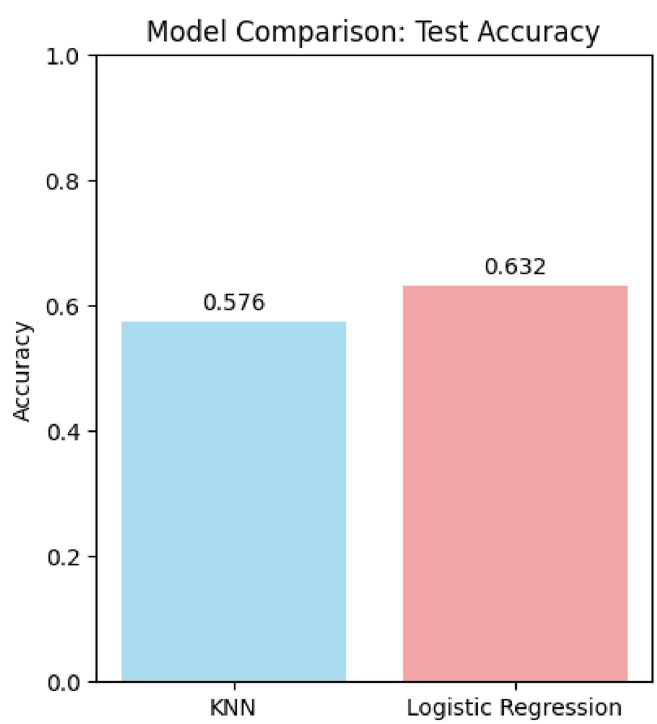
5.3.3 Selected Models: KNN and Logistic Regression:

KNN:
- Captures local patterns and similarity between customers
- Effective for non-linear trends
- Useful as a baseline classifier

Logistic Regression:
- Highly interpretable
- Produces probabilistic outputs
- Suitable for binary loan approval decisions

5.3.4 Model Comparison:

The comparison illustrates how each model performs in interpretability, speed, scalability, and suitability for loan decisions.



6 Importing data in SQL Server:

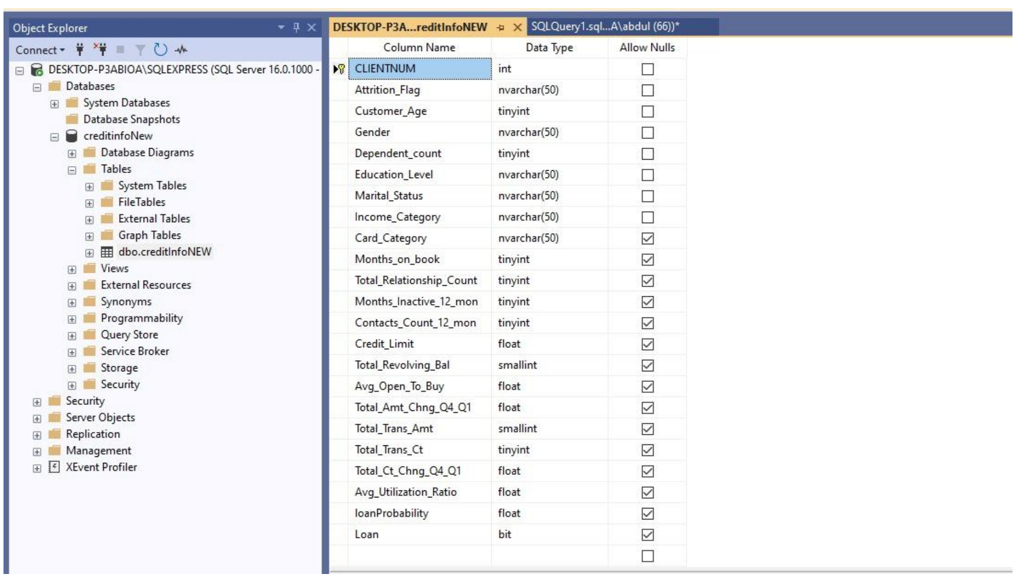
After being done with our ML Model we exported the dataset as a .csv file using the Pandas function '.to_csv' to import it in Microsoft SQL Server.

➔ Why SQL?

1. To clean & structure the data properly before running analysis
2. Perform transformations & data checks more efficiently
3. Create a reliable centralized data source for Power BI
4. Allow new entries under a proper, structured format

6.1 Defining constrains & Primary Key

After being done with our ML Model we exported the dataset as a .csv file using the Pandas function '.to_csv' to import it in Microsoft SQL Server.



We used the CLIENTNUM as a primary key as it is a unique code for every client so it will not be repetitive.

There are some data fields which accept Null Values as In banking, not all data is collected at the exact same moment.

6.2 SQL Queries:

We made segmentation for the Avg. Credit limit using Income Category

```sql
-- Average Credit Limit by Income Category
SELECT
    Income_Category,
    ROUND(AVG(Credit_Limit), 2) AS Avg_Credit_Limit
FROM creditInfoNew
GROUP BY Income_Category
ORDER BY Avg_Credit_Limit DESC;
```

Results | Messages

	Income_Category	Avg_Credit_Limit
1	$120K +	19717.32
2	$80K - $120K	15809.89
3	$60K - $80K	10758.77
4	Unknown	9516.58
5	$40K - $60K	5462.26
6	Less than $40K	3754.4

We made segmentation for the Credit limit using Card Category

```sql
-- Credit Limit Ordered by Card Category
SELECT CLIENTNUM, Credit_Limit, Card_Category
FROM creditInfoNew
ORDER BY Credit_Limit DESC;
```

100 % No issues found

Results | Messages

	CLIENTNUM	Credit_Limit	Card_Category
1	708153558	34516	Blue
2	708160008	34516	Blue
3	708163758	34516	Blue
4	708195633	34516	Silver
5	708226833	34516	Gold
6	708273858	34516	Blue
7	708358683	34516	Silver
8	708428958	34516	Gold
9	708501783	34516	Blue
10	708564858	34516	Gold
11	708672708	34516	Blue
12	708697983	34516	Blue
13	708710283	34516	Silver
14	708758658	34516	Blue
15	708848208	34516	Blue
16	708894183	34516	Silver
17	708948483	34516	Silver
18	709015833	34516	Silver
19	709066983	34516	Blue
20	709104333	34516	Silver
21	709163358	34516	Gold
22	709188108	34516	Silver

We requested a table with a detailed Credit limit for every Client

```sql
-- Credit Limit
SELECT CLIENTNUM, Credit_Limit
FROM creditInfoNew
ORDER BY Credit_Limit DESC;
```

100 % No issues found

Results | Messages

	CLIENTNUM	Credit_Limit
1	708153558	34516
2	708160008	34516
3	708163758	34516
4	708195633	34516
5	708226833	34516
6	708273858	34516
7	708358683	34516
8	708428958	34516
9	708501783	34516
10	708564858	34516
11	708672708	34516
12	708697983	34516
13	708710283	34516
14	708758658	34516
15	708848208	34516
16	708894183	34516
17	708948483	34516
18	709015833	34516
19	709066983	34516
20	709104333	34516
21	709163358	34516
22	709188108	34516

We requested a table with Clients who has high probability to get a loan but they didn't get one yet (High potential Customers)

```sql
-- High Loan Probability Customers (No Current Loan)
SELECT
    CLIENTNUM,
    Customer_Age,
    LoanProbability,
    Credit_Limit
FROM creditInfoNew
WHERE Loan = 0
  AND LoanProbability > 0.70
ORDER BY LoanProbability DESC;
```

100 % No issues found

Results | Messages

	CLIENTNUM	Customer_Age	LoanProbability	Credit_Limit
1	721012983	46	0.89229428768158	12935
2	787489083	55	0.881252229213715	21351
3	827114058	46	0.879927575588226	11634
4	708869283	50	0.854896426200867	23209
5	710297133	47	0.847214996814728	30210
6	718848183	50	0.838253438472748	9873
7	715471833	50	0.838176071643829	28253
8	778871508	54	0.828406512737274	13048
9	716966808	47	0.81315153837204	11722
10	709861008	47	0.811393857002258	27720
11	708664008	57	0.81131237745285	14270
12	780715683	41	0.810438990592957	13602
13	717491058	42	0.797808527946472	20803
14	715506258	50	0.797247707843781	16388
15	708793008	46	0.796253621578217	9949
16	718673358	35	0.796036601066589	13590
17	788786208	46	0.795978248119354	9881
18	794843883	41	0.791999936103821	34516
19	718542858	50	0.79071319103241	10588
20	718899708	55	0.783041954040527	17642
21	719242533	41	0.777929186820984	22770
22	772076208	58	0.771627843379974	26108

We requested a table with the percentage and counting of Existing and Attired Clients

```sql
-- Customer Attrition Breakdown (Count + Percentage)
SELECT
    Attrition_Flag,
    COUNT(CLIENTNUM) AS Customer_Count,
    (COUNT(CLIENTNUM) * 100.0 / (SELECT COUNT(*) FROM creditInfoNew)) AS Percentage
FROM creditInfoNew
GROUP BY Attrition_Flag;
```

100 % No issues found

Results | Messages

	Attrition_Flag	Customer_Count	Percentage
1	Attrited Customer	1627	16.064375987361
2	Existing Customer	8501	83.935624012638

We requested a table with the Loan probability of each Client using the Client Number

```sql
-- Loan Probability
SELECT CLIENTNUM, loanProbability
FROM creditInfoNew
ORDER BY loanProbability DESC;
```

100 % No issues found

Results | Messages

	CLIENTNUM	loanProbability
1	788943333	0.9223912358284
2	721325133	0.909323692321777
3	803148033	0.90581202507019
4	719475633	0.893334448337555
5	721012983	0.89229428768158
6	715216908	0.89096736907959
7	787489083	0.881252229213715
8	717975333	0.881193935871124
9	827114058	0.879927575588226
10	789412458	0.878727197647095
11	779557533	0.868985116481781
12	718699683	0.867621123790741
13	808747233	0.866858839988708
14	796336833	0.862685739994049
15	718785483	0.860981941223145
16	816596583	0.857225239276886
17	708869283	0.854896426200867
18	711048783	0.854389011859894
19	716791908	0.850802779197693
20	709967358	0.848528027534485
21	710297133	0.847214996814728
22	718623783	0.847208440303802

Inserting & Selecting a New Client Data:

```sql
-- Inserting Row
INSERT INTO creditInfoNew (
    CLIENTNUM,
    Attrition_Flag,
    Customer_Age,
    Gender,
    Dependent_count,
    Education_Level,
    Marital_Status,
    Income_Category,
    Card_Category
)
VALUES (
    999000001,
    'Existing Customer',
    32,
    'F',
    0,
    'Doctorate',
    'Married',
    '$80K - $120K',
    'Platinum'
);
```

```sql
-- Selecting new row
Select *
FROM creditinfoNew
where CLIENTNUM LIKE '999000001'
```

100 % No issues found

Results | Messages

	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category	Card_Category	Months_on_book	Total_Relationship_Count	Months_Inactive_12_mon	Contacts_Count_12_mon
1	999000001	Existing Customer	32	F	0	Doctorate	Married	$80K - $120K	Platinum	NULL	NULL	NULL	NULL

Reviewing the Whole Data:

```sql
-- Selecting Query
SELECT *
FROM creditInfoNew;
```

100 % No issues found

Results | Messages

	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category	Card_Category	Months_on_book	Total_Relationship_Count	Months_Inactive_12_mon	Contacts_Count_12_mon	Credit_Limit
1	708082083	Existing Customer	45	F	3	High School	Married	Less than $40K	Blue	36	4	3	3	3544
2	708083283	Attrited Customer	58	M	0	Unknown	Single	$40K - $60K	Blue	45	3	1	3	3421
3	708084558	Attrited Customer	46	M	3	Doctorate	Divorced	$80K - $120K	Blue	38	6	3	3	8258
4	708085458	Existing Customer	34	F	2	Uneducated	Single	Less than $40K	Blue	24	6	2	2	1438.30004882813
5	708086958	Existing Customer	49	F	2	Uneducated	Married	Unknown	Blue	41	3	5	2	3128
6	708095133	Existing Customer	43	M	4	Unknown	Unknown	$120K +	Blue	34	5	2	2	33304
7	708098133	Existing Customer	32	F	0	Graduate	Married	Less than $40K	Blue	19	6	1	0	2834
8	708099183	Existing Customer	37	F	2	High School	Single	Less than $40K	Blue	36	4	2	2	5723
9	708100533	Existing Customer	55	F	3	College	Single	Less than $40K	Blue	36	3	3	3	2679
10	708103608	Existing Customer	52	M	1	High School	Single	$60K - $80K	Blue	45	1	5	1	11898
11	708104658	Existing Customer	46	M	1	Graduate	Unknown	$40K - $60K	Blue	36	4	3	2	1438.30004882813
12	708108333	Attrited Customer	47	F	3	Graduate	Married	Unknown	Blue	36	3	3	1	5590
13	708112008	Existing Customer	56	M	2	Graduate	Married	$80K - $120K	Blue	36	4	2	3	23510
14	708113208	Existing Customer	53	F	1	High School	Married	Less than $40K	Blue	36	3	2	2	1688
15	708117933	Attrited Customer	44	F	2	Graduate	Divorced	Less than $40K	Blue	36	3	2	3	1880
16	708119658	Existing Customer	49	F	4	Graduate	Married	$40K - $60K	Blue	38	4	2	1	12836
17	708121908	Attrited Customer	48	M	4	Unknown	Married	$80K - $120K	Blue	36	2	3	2	22917

7. Power BI:

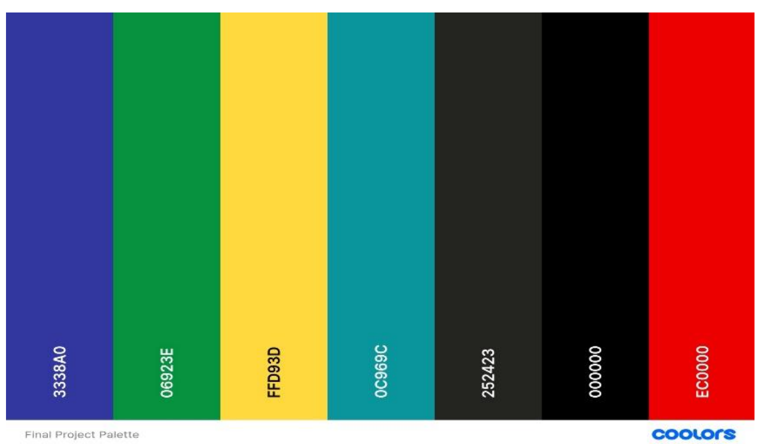
Why we chose Power BI for this task?

Power BI provides a seamless workflow with SQL Server and is optimized for interactive dashboards. With 10K+ rows, creative charts & high-interactivity, Power BI seems like the perfect fit for this project.

7.1 Power Bi: Color Palette & Design:

For the color palette we chose the following:

- Forest Green: related to money and financial matters
- Mustard Yellow: refers to the Gold (Premium)
- Egyptian Blue: refers to Trust, Security and Logic
- Dark Cyan: refers to Modernity and Innovation
- Pure Red
- Carbon Black
- Pure Black

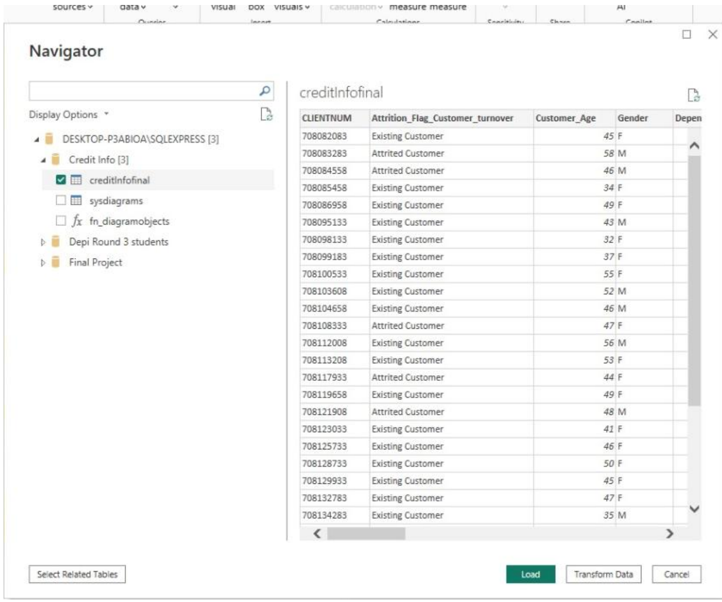


While keeping using black text on a white canvas to ensure maximum readability, clarity, letting visuals stand out clearly without straining the eye

We chose this color palette because it reflects a professional financial theme, provides strong readability through high contrast, and uses intuitive colors to guide the user's attention while maintaining a clean, modern aesthetic.

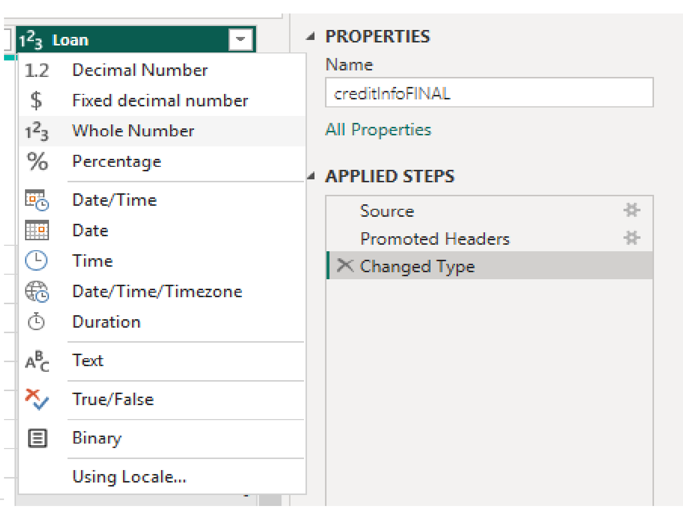
7.2 Power Bi: Importing data from SQL:

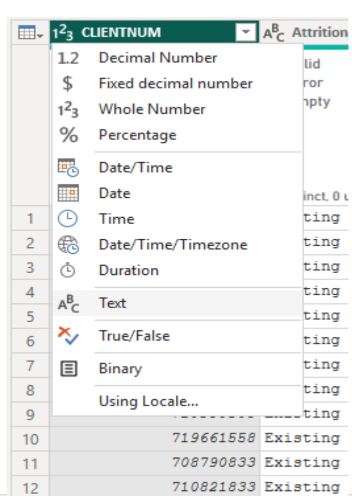
We imported the data from Microsoft SQL Server to Power Bi.



7.3 Power Bi: Data Transformation:

During the transformation phase we've had to change the data type of both 'Loan' and 'CLIENTNUM' as the first got turned into True/False during SQL Importation stages and the latter turned into a whole number rather than a Text/String. After correcting their data types and checking data values we found nothing wrong and proceeded with our dashboard design and analysis

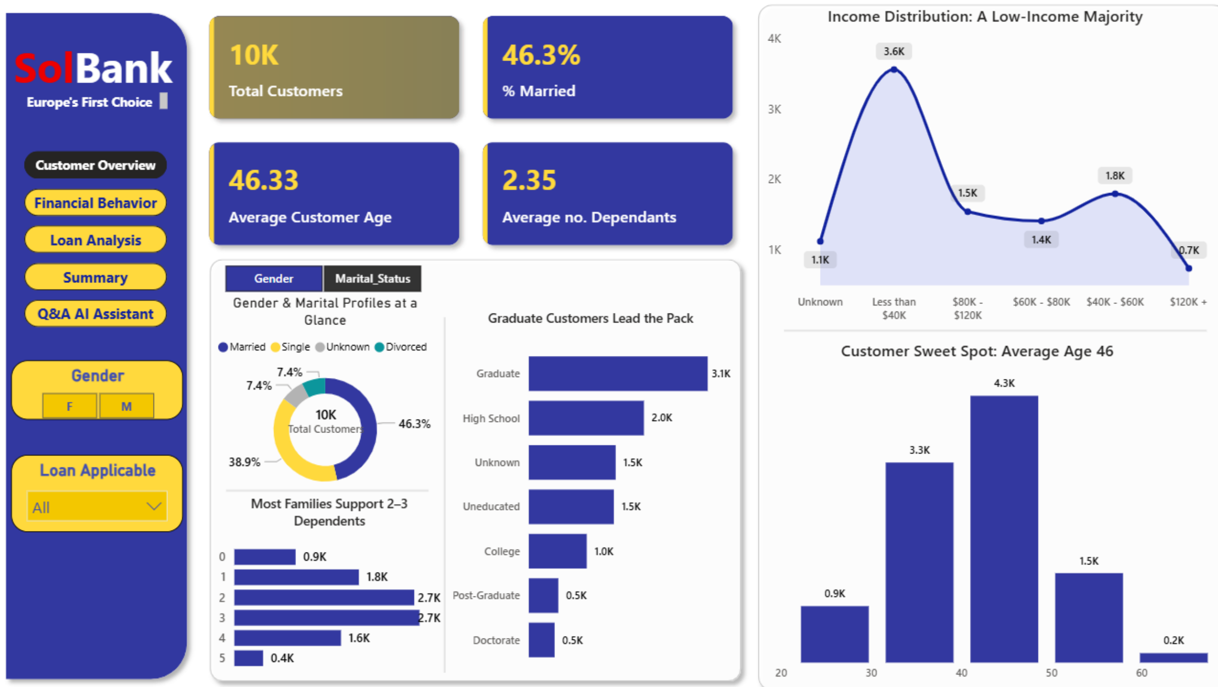




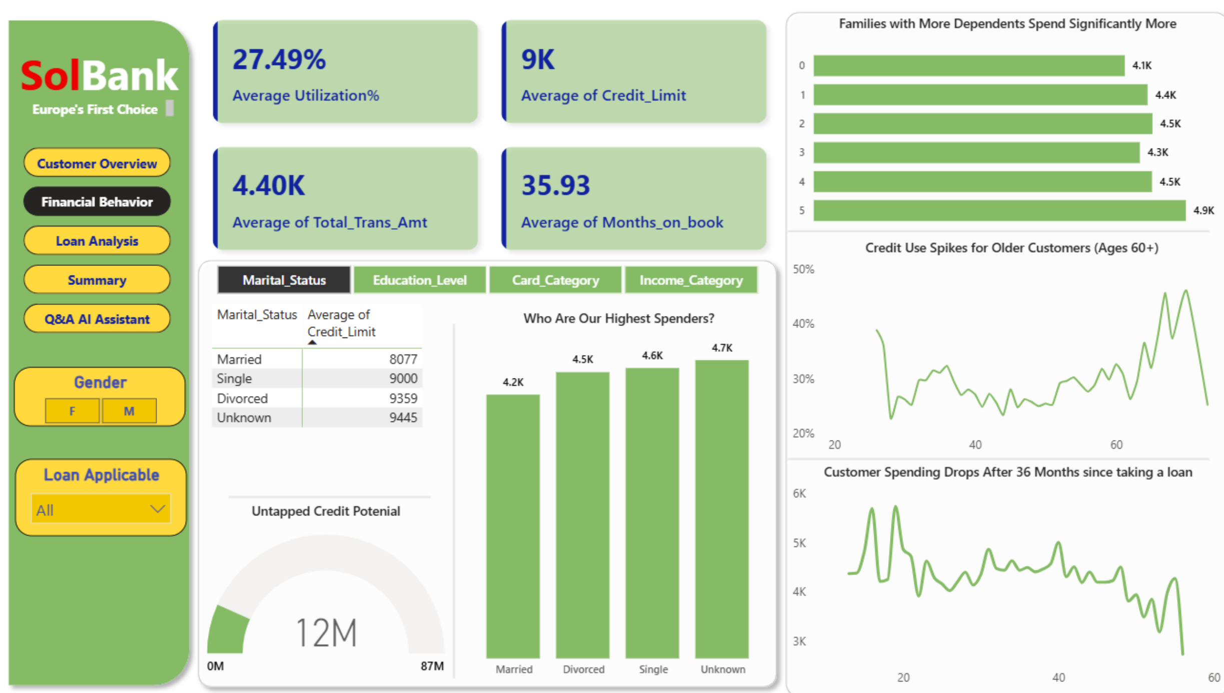
7.4 Power Bi: Building Dashboards

We divided our Business Problem into Four Sectors the first sector is a Customer Analysis Overview, the second sector is Customers Financial Behavior, the Third Sector is Loan Analysis and a Quick Summary that Summarizes the Dashboards

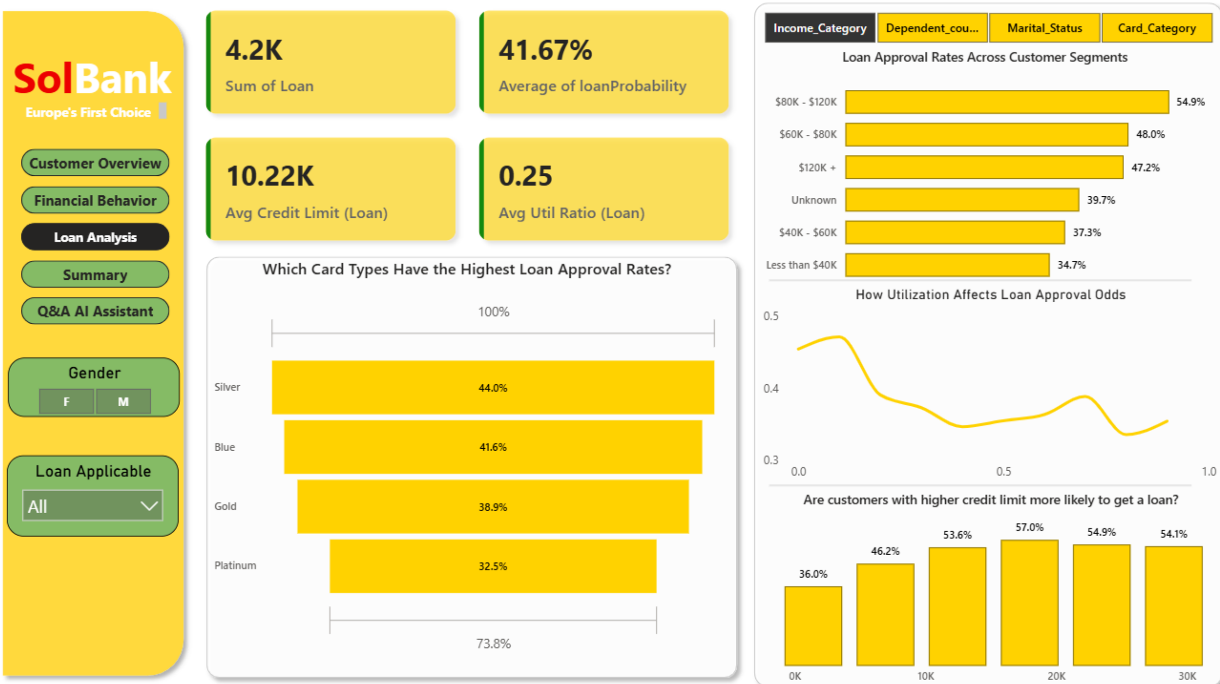
First Customers Overview:



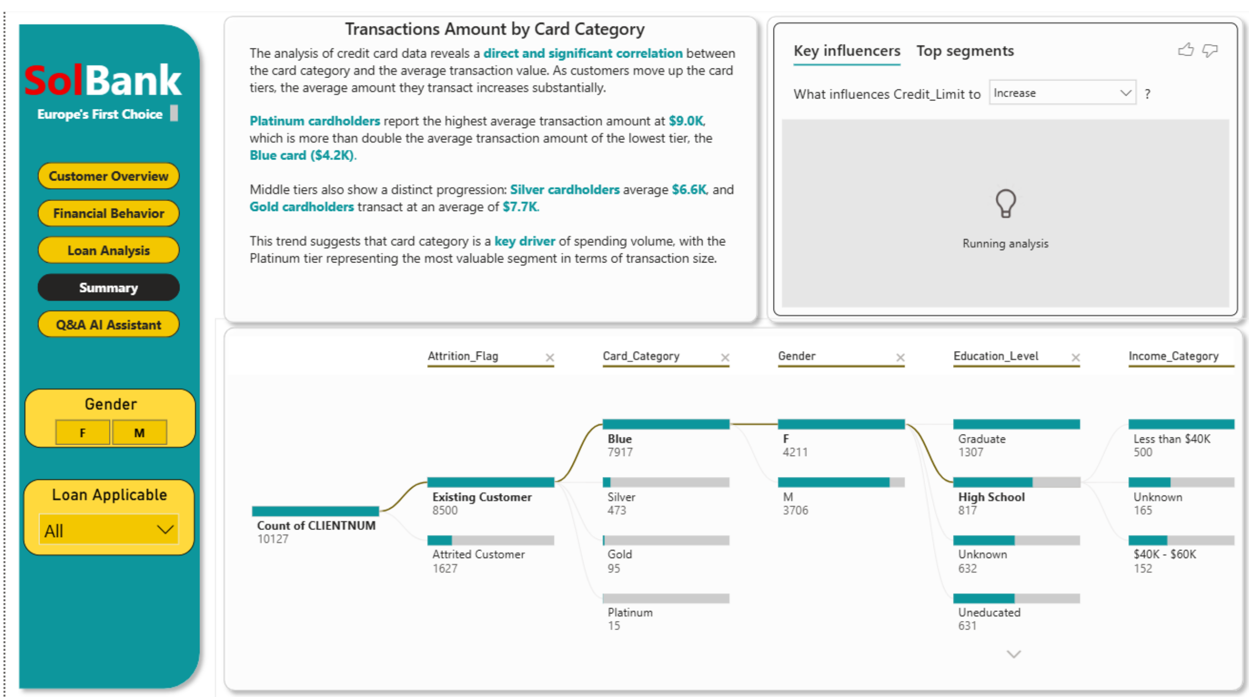
Second Financial Behavior:



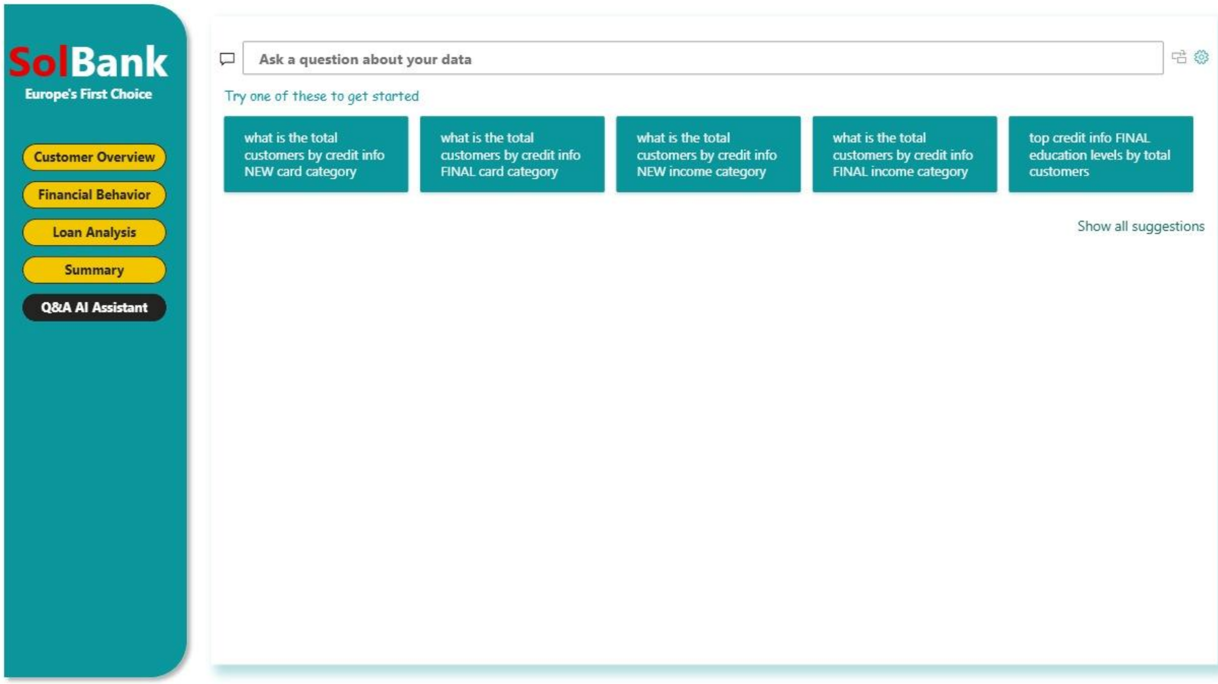
Third Loan Analysis:



Fourth Summary :



Then We Made an AI Q&A Agents to help in getting Insights from the Dashboards



7.5 Power Bi: Business Questions

1. How Many Client is enrolled in our Bank ?
   10,000 Clients
2. What is The Average Customer Age ?
   46.33Years
3. What's the Martial Status of The most of Our clients ?
   46.3% are Married
4. On Average, how many dependents do families have ?
   2.35 Dependents
5. What is the Distribution of the Martial Status of Our Clients ?
   46.3% Married
   38.9% Single
   7.4% Divorced
   7.4% Unknown

6. What is the Distribution of the Gender Profile of Our Clients ?
   47% Males
   53% Females
7. What is the Distribution of the Number of Dependents for Our Clients ?
   Between 0 and 5 with a semi bell shaped Graph
8. How educated is our Clients Base ?
   Only 3500 clients are either Having a High School Degree or Uneducated
9. What is the Target Segment of the Bank ?
   Most of our Clients are from the Low-Income Level (Less Than $40K)
10. What is the Average Age of Our Customer Base ?
    From the Age of 40 till the age of 50 represents 43% of our Clients
11. What is the Average Utilization Percentage ?
    27.49%
12. What is the Average Credit Limit for our Clients ?
    9000 $
13. What is the Average Transactions Amount of Our Clients ?
    4400 $
14. Average Months on book ?
    35.93 Months
15. What is the credit limit for each marital status ?
    The Unknown & Divorced Leads The Average Credit Limit
16. Does The Married clients spend more than the Single ones
    No, As the Married spend on average 400$ less than the Single ones
17. Who leads the Spending Cycle by Educational Level ?
    The post graduates
18. What is the card category Spends most ?
    The platinum Cards
19. How is Credit limit represented across different income categories ?
    +120K $ with approximately credit limit 20,000 $
    Less than 40K with a credit limit 3,700 $
    60K - 80K with a credit limit 10,000 $
20. Is The Credit limit Used efficiently ?
    No, as the Credit used by clients is only 12 Millions from 87 Millions \

21. How does spending differ across different family sizes?
    At 5 dependents families spend $400 more than families who only have 4 dependents, who spend $200 more than families with 3 dependents.
22. What's the relation between credit utilization and Age?
    Younger clients utilize it more to adapt to their life needs, usage spikes for older clients as they need medication.
23. Do customers spend more since taking a loan?
    No, as a matter of fact spending shows a downward trend since taking a loan
24. How many clients got approved for a loan?
    4200 clients
25. What's the average probability at getting a loan?
    Approximately 41.7%
26. What's the average credit limit for clients who got a loan?
    Around $10,220.
27. How much do clients with a loan utilize their credit limit?
    The average utilization ratio is 0.25, which is a healthy range.
28. Which Card Types have the highest loan approval rates?
    Silver takes the lead with 44% which is higher than the average by 3% while blue has 41.6% as it is the most common card type, representing the average customer.
29. What financial class leads in Loan Approval rates?
    $80K-$120K is in the lead with 54.9% approval rate, followed by $60K-$80K with 48%
30. What's the relation between loan approval rates and dependent count?
    Loan approval rates show a some-what linear relationship; as dependent count increases the loan approval rates increase EXCEPT for those with 5 dependents, which have only 35% approval rate.
31. In terms of marital affairs, who are the most and least approved?
    The most approved are married people 45% while Divorced people only have 37%
32. How does utilization affect loan approval odds?
    As utilization levels increase to unhealthy amounts, loan approval odds decrease.

33. Which credit card category have the highest probability to take a loan?
    Silver takes the lead with 44% followed by blue with 41.6%
34. Which credit card categories ACTUALLY get approved for a loan?
    Gold with 44% and Silver are tied 44% followed by blue at 41.3% and platinum at 35%
35. Are customers with higher credit limit more likely to get a loan?
    Yes, as there's an upward trend linking between credit limit and loan approval, with a spike of 57% in the middle for those with a credit limit between $15K~$20K

7.6 Insights and Actionable Solution

- **Dashboard 1:**

  Sol Bank's current customer base is firmly rooted in the **middle-aged, middle-class family demographic**. The bank serves an educated population that is nonetheless earning on the lower end of the income spectrum ($40K or less). This suggests a strategy focused on stability, family financial planning, and volume-based retail banking rather than high-net-worth wealth management.

1. The average customer is **46 years old**, with the vast majority falling between the ages of 30 and 50. This is a prime age for financial stability needs (mortgages, insurance, retirement planning).
2. The most striking insight is the income distribution. The largest segment (**3.6K customers**) earns **less than $40K**, followed by the $40K–$60K bracket. Only a tiny fraction (0.7K) earns over $120K These customers may have a higher sensitivity to fees and a higher risk of loan default if economic conditions worsen.

- **Dashboard 2:**

  This "Financial Behavior" dashboard reveals crucial details about how Sol Bank's customers interact with their money. While the previous dashboard established who they are (low-income families), this view highlights spending habits, lifecycle risks, and massive missed revenue opportunities.

1. Sol Bank has extended a massive amount of credit that customers are **not using**. With an average utilization of only **27.49%**, the bank is missing out on significant interest revenue and interchange fees.
   You don't need new customers to grow revenue; you need existing customers to use their cards more. Consider loyalty programs or "spend-and-get" campaigns to encourage card usage for daily purchases, not just emergencies.

2. The middle-right chart ("Credit Use Spikes for Older Customers") shows a dramatic rise in credit utilization for customers over age 60, reaching nearly **45%** , Older customers are relying heavily on credit, potentially to bridge the gap between retirement income and living expenses. This is a **high-risk segment**. If these are unsecured credit lines, the default risk increases significantly as these customers age on fixed incomes. However, it also signals a need for products tailored to seniors

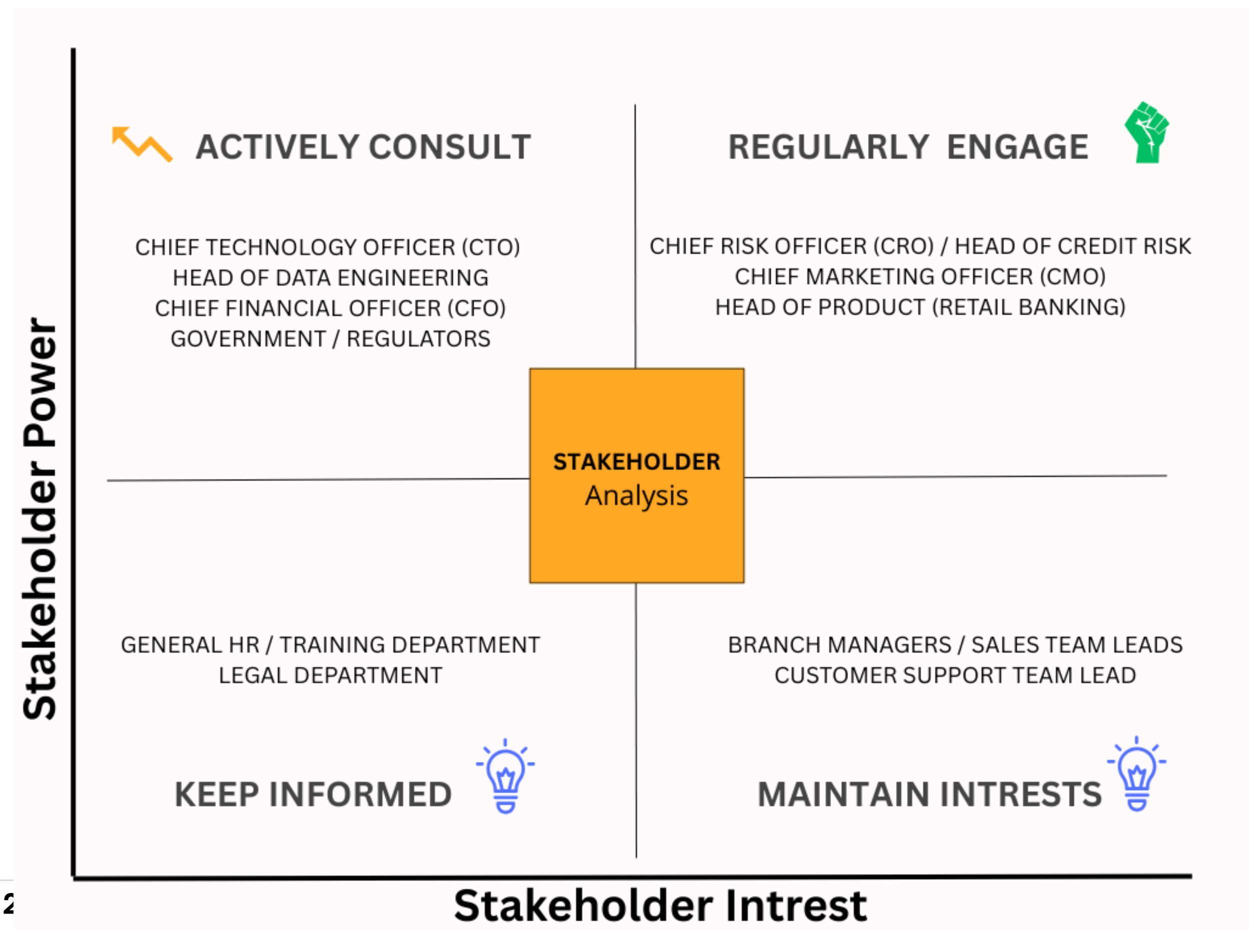
3. The top-right bar chart confirms that spending correlates directly with family size. Families with **5 dependents** have the highest transaction volume (~4.9K). Despite spending the most, the "Marital Status" table shows that **Married** customers have the **lowest average credit limit** ($8,077) compared to Single ($9,000) or Divorced ($9,359). **So the Bank Should Review their underwriting criteria.** Increasing limits for married couples with dependents could immediately unlock higher transaction volumes.

4. The data is inconsistent so the bank should Audit and update its database

- **Dashboard 3:**

1. Income Strongly Influences Loan Approvals as The $80K–$120K group has highest approval (54.9%) Followed by $60K–$80K (48%) while Low-income (<$40K) is the largest group but lowest average limit.
2. The probability of getting a loan creates a "cliff" drop-off once a customer uses just **10-20%** of their credit limit. You are rejecting the very people who have demonstrated they need credit and are willing to use it. This explains the low overall utilization (25%) seen in the previous dashboard—the system selects for non-spenders. So the bank should review their Credit system.

8. Stakeholder Analysis



9 Conclusion

This project set out to analyze the end-to-end customer lifecycle at Sol Bank from demographic profiling to spending behavior and loan uptake. The analysis of our 10,000-customer sample reveals a fundamental disconnect: Sol Bank’s customer base is "Mass Market," but our operational logic is "Premium Exclusive."

We are currently leaving approximately $75M in credit potential untapped due to overly conservative risk models and misaligned product offerings.

Sol Bank possesses a stable, educated, and family-oriented customer base with high loyalty potential. The stagnation in revenue is not due to a lack of market demand, but due to internal friction in our **Risk Logic** and **Data Quality**. By aligning our underwriting rules with the reality of our "Family" demographic, we can unlock significant value from the assets we already hold.