Presentation by :

- Sushant Mane

Workflow of the Project

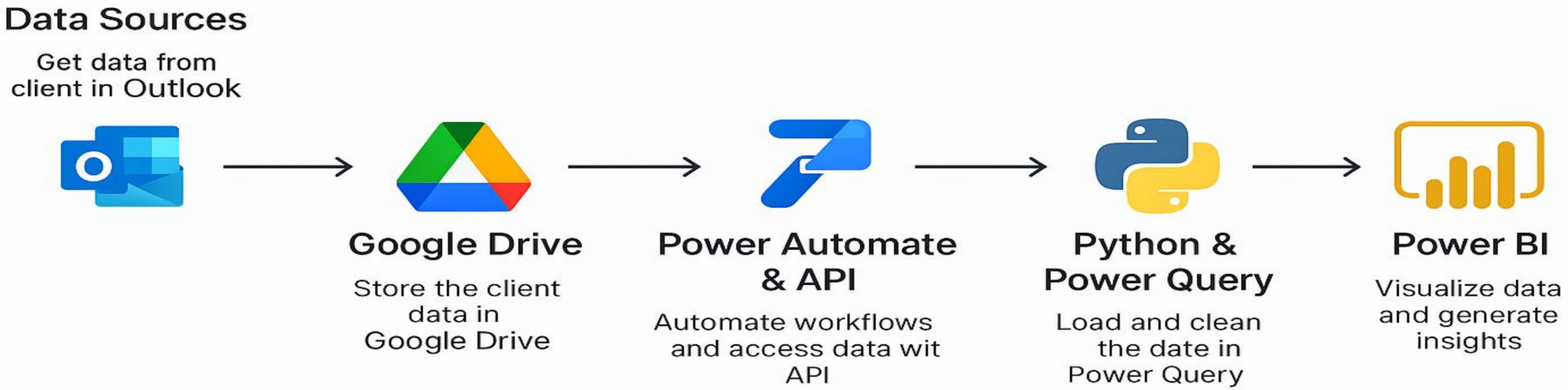


Table of Contents

- Loading Data From Email to Power BI automatically.
- Cleaning Data in Power Query
- Solving 10 Business Problems

TOOLS USED


EMAIL


MICROSOFT®
AUTOMATE


Power BI
SERVICE


Google Cloud
Platform


Google Drive


Power Query

PROBLEM STATEMENT 1

FINANCIAL ANALYTICS

Current Workflow Overview

Each day around 3 PM, we receive ~25 data files via email from a geographically distributed survey team. These files are critical for building a dashboard, which must be completed and delivered to our client (the US government) by 8 PM. The process involves:

- Downloading files from email
- Combining them into one consolidated file
- Cleaning the data for consistency and accuracy
- Building and updating the dashboard
- Delivering final insights to the client

Key Challenges:

- Time-Consuming: Significant time is spent on downloading, merging, and cleaning files, causing delays. • High Costs: Hiring two additional staff to handle this has added $12,000/month in expenses.
- Error-Prone: Frequent data manipulation errors delay output and affect accuracy.
- Workload Pressure: The intensive daily workflow is affecting other project timelines and overall team efficiency.

Objective:

We need solutions to:

- Reduce turnaround time for file handling and processing • Minimize errors in data manipulation
- Cut costs by reducing dependency on manual labor
- Streamline workload to improve overall team productivity

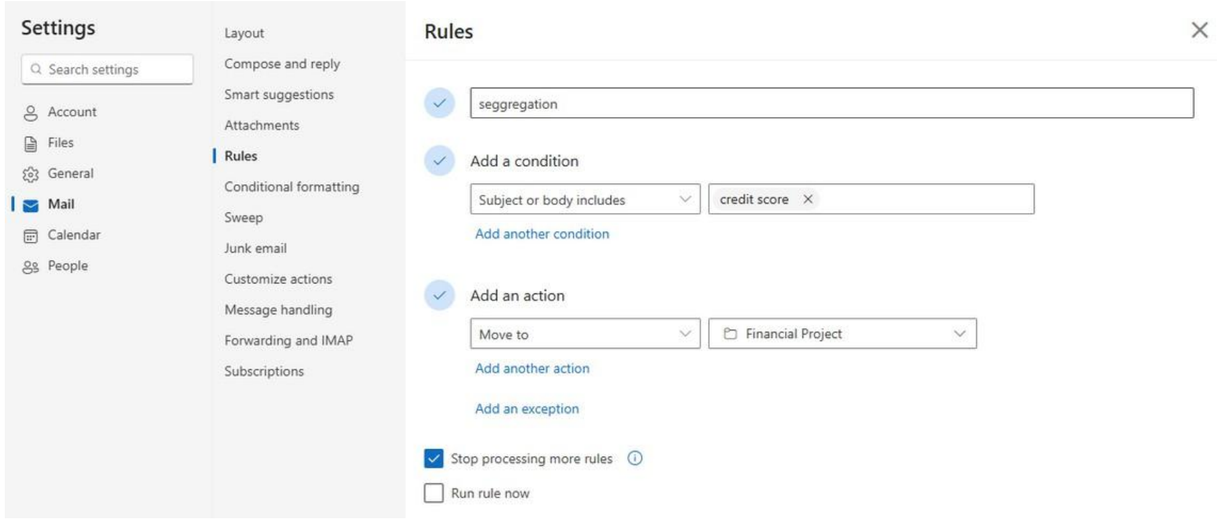
Your insights on automating or optimizing this workflow would be immensely helpful in boosting efficiency and performance.

SOLUTION OF PROBLEM STATEMENT 1

FINANCIAL ANALYTICS

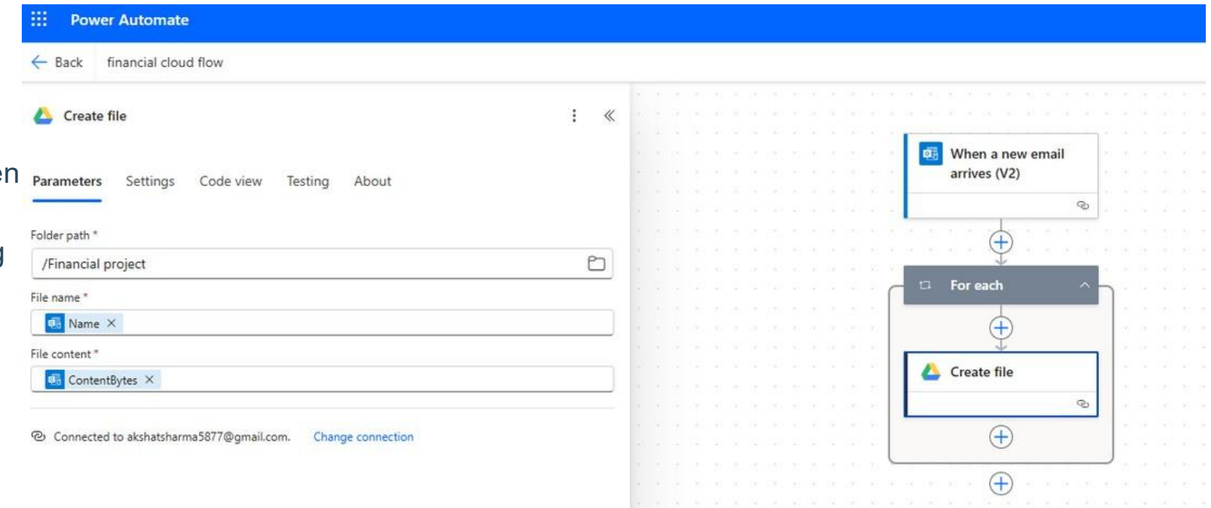
STEP 1:

Create an Outlook account and set up a mail rule (Settings > Rules > Add new rule) to automatically move emails containing **"credit score"** in the subject or body to a dedicated folder named **"Financial Project"**. This ensures all relevant daily analytics emails are stored in one place.



STEP 2:

Create a folder in Google Drive named "Financial Project", and then use Power Automate to transfer attachments from the Outlook "Financial Project" folder to this Google Drive folder, streamlining data collection for analysis.

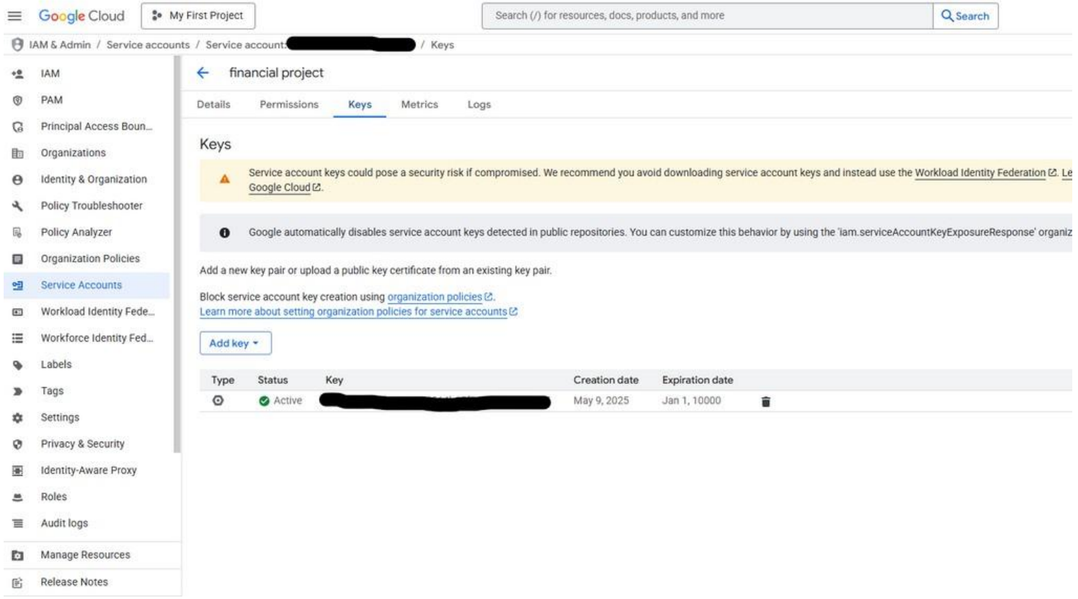


SOLUTION OF PROBLEM STATEMENT 1

FINANCIAL ANALYTICS

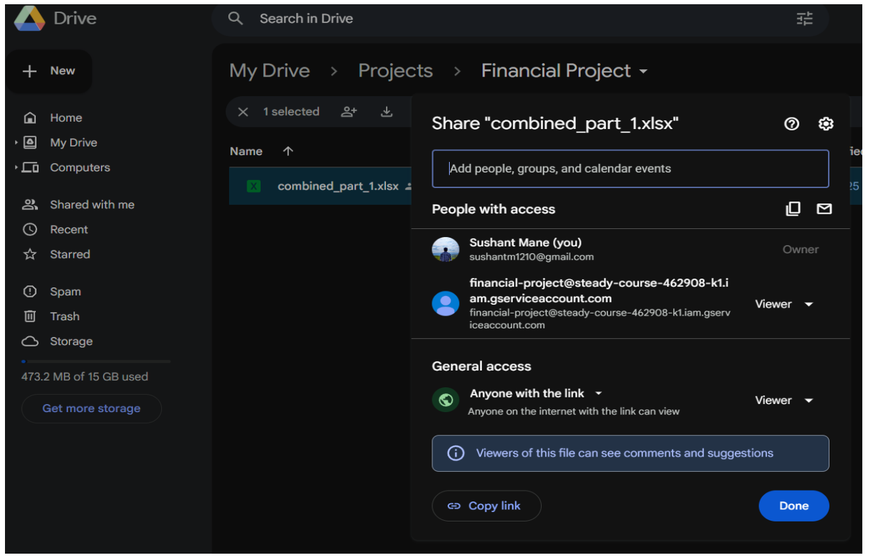
STEP 3:

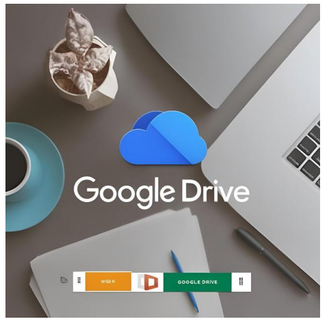
To get the data from Google Drive into Power BI Desktop, a direct connection isn't possible since they are separate platforms. To bridge this, we'll create an **API** using **Google Cloud Platform (GCP)**. By enabling the Google Drive API and generating the necessary **credentials**, we can securely access and fetch the Drive-stored data into Power BI Desktop for analysis.



STEP 4:

To ensure the API only accesses the **Financial Project** folder in Google Drive, we need to restrict its access. After creating the API, copy the service account email generated in GCP, then go to **Drive > Financial Project > Share, paste the service email, and click "Add"**. This grants the API permission to access only that specific folder.

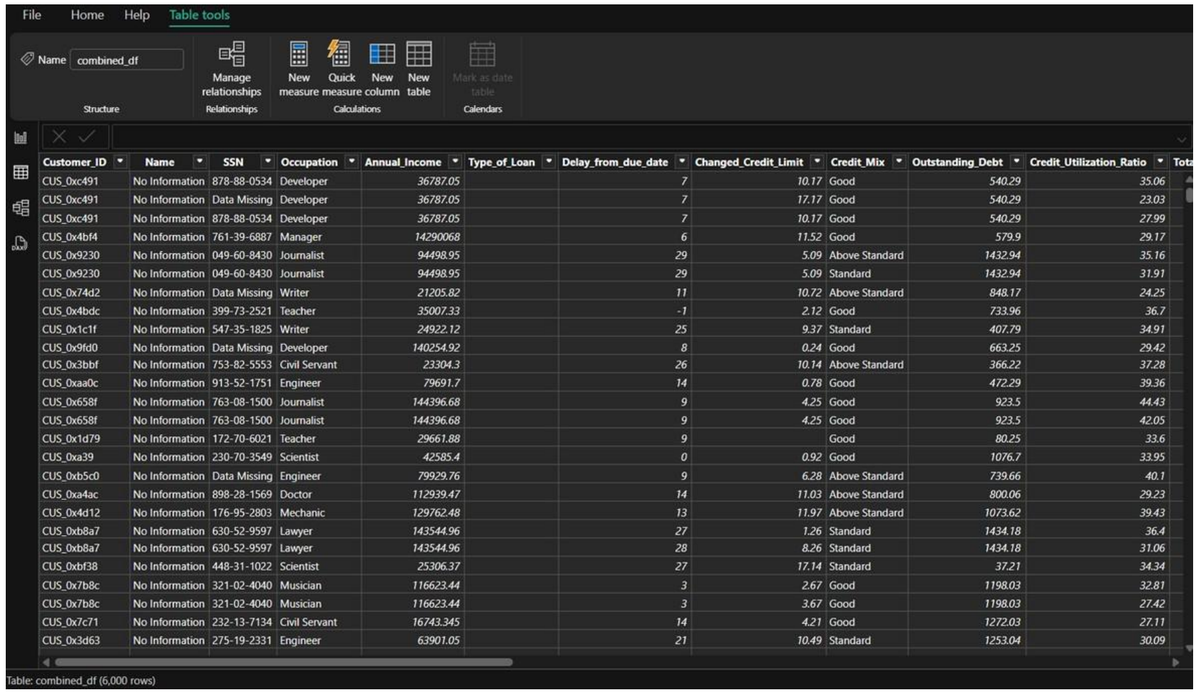




SOLUTION OF PROBLEM STATEMENT 1

STEP 3:

Now, open Power BI Desktop and go to:
Get Data > More > Python Script, then add the script saved in resources folder.
This Python script uses the **credentials** downloaded during API creation to access the **"Financial Project"** folder in Google Drive, fetch all the files, and **concatenate** them into a single **combined_df** for further analysis in Power BI.



Customer_ID	Name	SSN	Occupation	Annual_Income	Type_of_Loan	Delay_from_due_date	Changed_Credit_Limit	Credit_Mix	Outstanding_Debt	Credit_Utilization_Ratio	Tota
CUS_0xc491	No Information	878-88-0534	Developer	36787.05		7	10.17	Good	540.29	35.06	
CUS_0xc491	No Information	Data Missing	Developer	36787.05		7	17.17	Good	540.29	23.03	
CUS_0xc491	No Information	878-88-0534	Developer	36787.05		7	10.17	Good	540.29	27.99	
CUS_0x4bf4	No Information	761-39-6887	Manager	14290068		6	11.52	Good	579.9	29.17	
CUS_0x9230	No Information	049-60-8430	Journalist	94498.95		29	5.09	Above Standard	1432.94	35.16	
CUS_0x9230	No Information	049-60-8430	Journalist	94498.95		29	5.09	Standard	1432.94	31.91	
CUS_0x74d2	No Information	Data Missing	Writer	21205.82		11	10.72	Above Standard	848.17	24.25	
CUS_0x4bdc	No Information	399-73-2521	Teacher	35007.33		-1	2.12	Good	733.96	36.7	
CUS_0x1c1f	No Information	547-35-1825	Writer	24922.12		25	9.37	Standard	407.79	34.91	
CUS_0x9fd0	No Information	Data Missing	Developer	140254.92		8	0.24	Good	663.25	29.42	
CUS_0x3bbf	No Information	753-82-5553	Civil Servant	23304.3		26	10.14	Above Standard	366.22	37.28	
CUS_0xaa0c	No Information	913-52-1751	Engineer	79691.7		14	0.78	Good	472.29	39.36	
CUS_0x658f	No Information	763-08-1500	Journalist	144396.68		9	4.25	Good	923.5	44.43	
CUS_0x658f	No Information	763-08-1500	Journalist	144396.68		9	4.25	Good	923.5	42.05	
CUS_0x1d79	No Information	172-70-6021	Teacher	29661.88		9		Good	80.25	33.6	
CUS_0xa39	No Information	230-70-3549	Scientist	42585.4		0	0.92	Good	1076.7	33.95	
CUS_0xb5c0	No Information	Data Missing	Engineer	79929.76		9	6.28	Above Standard	739.66	40.1	
CUS_0xa4ac	No Information	898-28-1569	Doctor	112939.47		14	11.03	Above Standard	800.06	29.23	
CUS_0x4d12	No Information	176-95-2803	Mechanic	129762.48		13	11.97	Above Standard	1073.62	39.43	
CUS_0xb8a7	No Information	630-52-9597	Lawyer	143544.96		27	1.26	Standard	1434.18	36.4	
CUS_0xb8a7	No Information	630-52-9597	Lawyer	143544.96		28	8.26	Standard	1434.18	31.06	
CUS_0xbf38	No Information	448-31-1022	Scientist	25306.37		27	17.14	Standard	37.21	34.34	
CUS_0x7b8c	No Information	321-02-4040	Musician	116623.44		3	2.67	Good	1198.03	32.81	
CUS_0x7b8c	No Information	321-02-4040	Musician	116623.44		3	3.67	Good	1198.03	27.42	
CUS_0x7c71	No Information	232-13-7134	Civil Servant	16743.345		14	4.21	Good	1272.03	27.11	
CUS_0x3d63	No Information	275-19-2331	Engineer	63901.05		21	10.49	Standard	1253.04	30.09	

Table: combined_df (6,000 rows)

DATA OVERVIEW

Now, if any team sends data via email, it will automatically flow into Power BI through the following streamlined pipeline:
Email → Power Automate → Google Drive → API → Python Script → Power BI.
This end-to-end automation ensures real-time data ingestion with minimal manual effort.

CLEANING DATA

FINANCIAL ANALYTICS

To prepare the dataset for analysis, I carried out a series of data cleaning steps using Power Query:

1. **Name Column:** I began by replacing any blank entries in the Name column with the phrase "No Information" to handle missing data.
2. **Age Column:** Underscores (_) were removed and the data type was converted to whole numbers. I then created a custom column to handle outliers—ages below 12 or above 95 were replaced with the median age. The original column was dropped and the new one renamed to Age.
3. **SSN Column:** Invalid or garbage values were standardized by replacing them with "Data Missing" for clarity.
4. **Annual Income:** Similar to the age column, underscores were removed and the data type was set to whole number.
5. **Monthly Salary Calculation:** I introduced a new column, monthly_in_hand_salary, calculated by deducting 10% tax from the annual income and dividing the result by 12.
6. **Bank Accounts:** To address any unrealistic values in Num_Bank_Accounts, I replaced values outside the 0–10 range with the column's median.
7. **Credit Cards:** Likewise, for Num_Credit_Card, values equal to or above 25 were considered outliers and replaced with the median.
8. **Interest Rate:** Any value exceeding 34 was treated as invalid and substituted with the median interest rate.
9. **Number of Loans:** I extracted the number of loans by splitting the Type_of_Loan field using a comma delimiter and counting the resulting items. If the field was empty, the count was set to 0.
10. **Credit Limit & Monthly Investment:** I cleaned the Changed_Credit_Limit and Amount_Invested_Monthly columns by removing underscores and converting the values to whole numbers.
11. **Credit History Age:** A new column was created to convert the credit history from text (e.g., "2 Years and 5 Months") to total months for consistency in analysis.
12. **Age Binning:** Finally, I created a conditional column to categorize customers into age bins, making it easier to group and analyze data by age segments.

PROBLEM STATEMENT 2

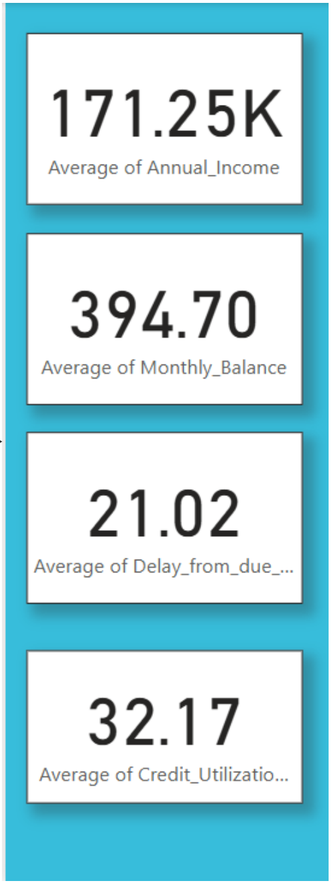
Show information's like

- Average Annual Income.
- Average Monthly balance.
- Average number of payment delays.
- Average Credit Utilization.

SOLUTION OF PROBLEM STATEMENT 2

To answer this question,
I used the **Card visual** in Power BI to display the following key metrics:

- **Average Annual Income:** ₹171.25K
- **Average Monthly Balance:** ₹394.70
- **Average Delay in Payment:** 21.02 days
- **Average Credit Utilization:** 32.17%

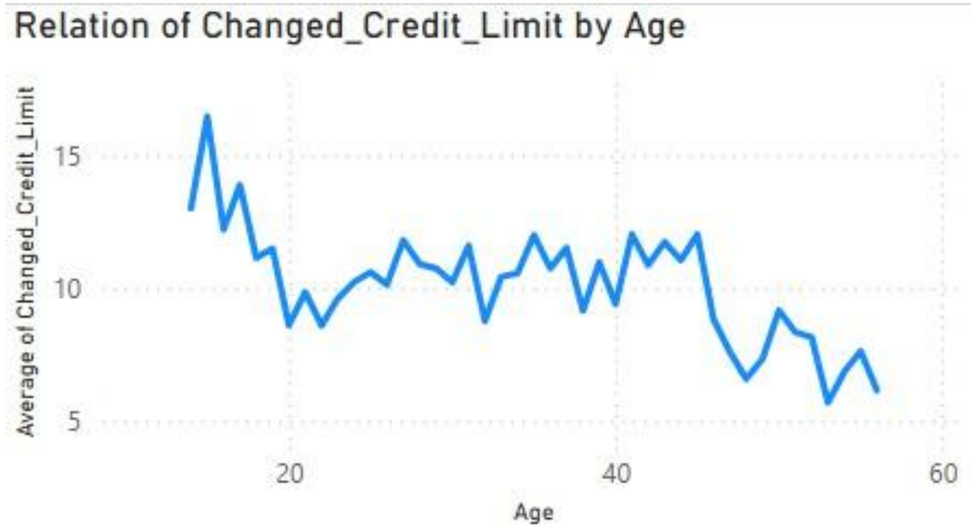


PROBLEM STATEMENT 3

Investigate the **relationship between age and changes in credit limit.** Analyze how variations in customer age correlate with adjustments in their credit limits to understand if and how age-related factors influence credit limit changes. This insight can help in tailoring credit products and strategies based on different age demographics. How is the payment behaviour of people different for different credit mix categories?

SOLUTION OF PROBLEM STATEMENT 3

To explore the link between age and changes in credit limit, I used a **line chart** in Power BI.



The chart shows that younger customers (under 25) experience **higher average changes**, while changes decline steadily with age, becoming **more stable after 30.**

This suggests that younger users face more credit adjustments due to evolving financial behavior, while older customers have more stable credit profiles, guiding age-specific credit strategies.

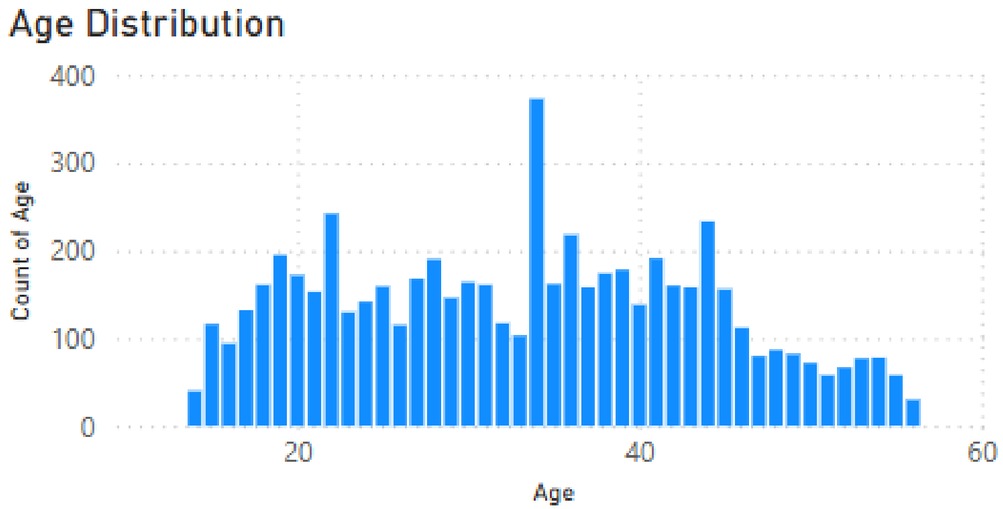
PROBLEM STATEMENT 4

Generate **a distribution plot to visualize the age demographics** of the customer base. This plot should illustrate the frequency and distribution of various ages to provide insights into the age profile of the population served, aiding in targeted marketing and service strategies.

SOLUTION OF PROBLEM STATEMENT 4

To visualize the age demographics of the customer base, I used a **bar chart** as a count plot.

The plot shows a wide age spread, with noticeable spikes around ages 22, 34, and 44.



PROBLEM STATEMENT 5

Client want to see **a number of people of different age groups having different kinds of credit scores.**

for making the age group consider the following rule: - •

14-19 “Teen”

- 19-25 “Young Adult”
- 25-35 “Old Adult”
- 35-45 “Old1”
- >45 “Old2”

SOLUTION OF PROBLEM STATEMENT 5

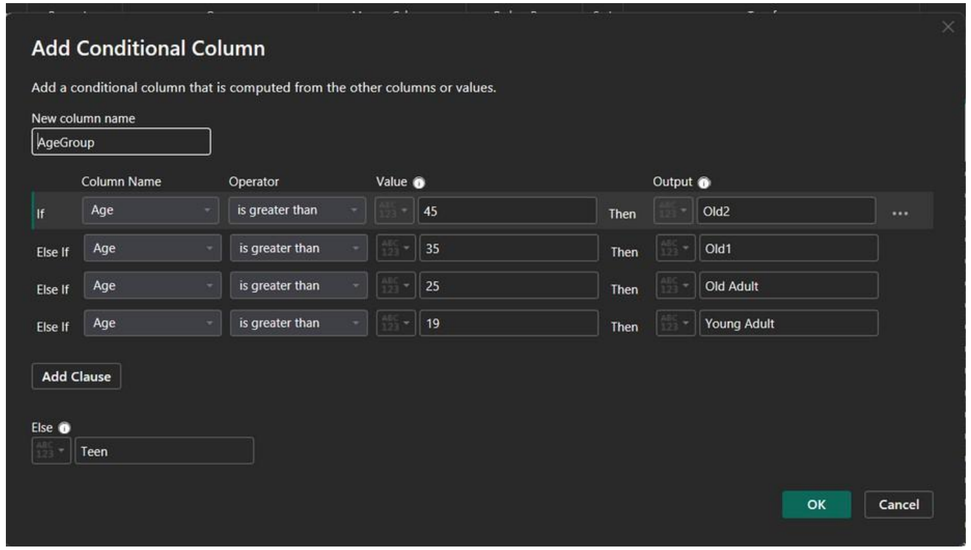
To analyze how credit scores vary across different age groups, I first created a **conditional age group column** to create above bins.

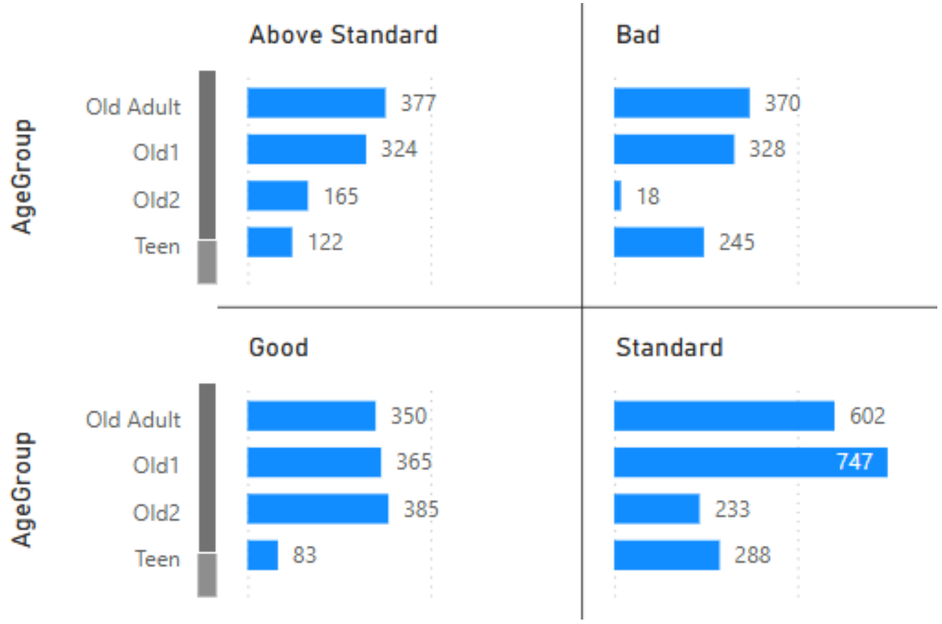
Then, I used a **bar chart** with small multiples by Credit Score categories to show the count of people in each age group for each credit score type (Above Standard, Good, Standard, Bad).

This visual helps compare how credit score distribution shifts with age, revealing that:

- Teens and Old2 show lower counts, while Old1 and Old Adult dominate in most categories.
- Bad scores are noticeably higher among Teens and Old Adults, highlighting segments with potential risk.

FINANCIAL ANALYTICS



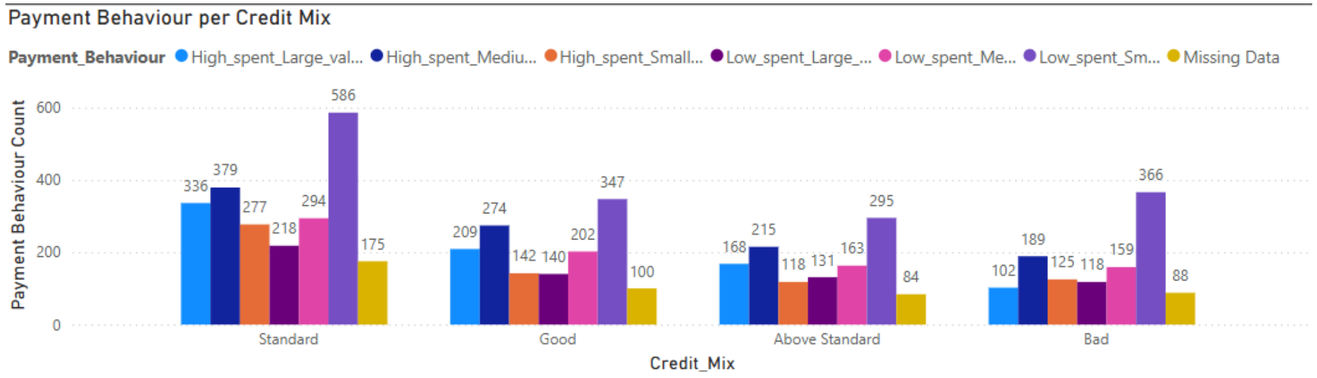


PROBLEM STATEMENT 6

Analyse the frequency of various **payment behaviours within each credit mix category.** Examine how often different payment behaviours occur across different types of credit mixes .This analysis will provide insights into payment behaviour trends associated with each credit mix, helping to tailor risk management and credit strategies.

SOLUTION OF PROBLEM STATEMENT 6

To analyze payment behavior across credit mix categories, I used a clustered **bar chart** that displays the count of people for each payment behavior within every credit mix category.



PROBLEM STATEMENT 7

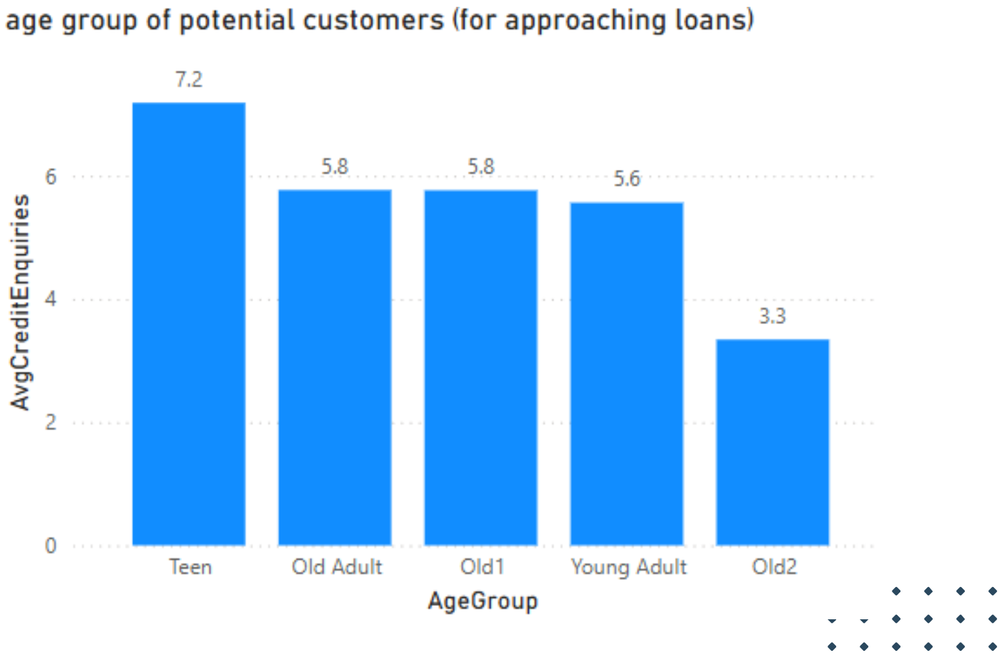
From the data available, is it possible to know the **age group of my potential customers** (for approaching loans), if yes then please show how.

SOLUTION OF PROBLEM STATEMENT 7

We used a **bar chart** with a **DAX measure** to calculate the average number of credit enquiries per age group. The Teen (14–19) group had the highest average (7.2), making them the most likely to approach for loans. The Old2 (>45) group had the lowest (3.3).

```
AvgCreditEnquiries =
AVERAGEX(
    FILTER(
        combined_df,
        NOT(ISERROR(VALUE(combined_df[Num_Credit_inquiry])))
    ),
    VALUE(combined_df[Num_Credit_inquiry])
    )

```

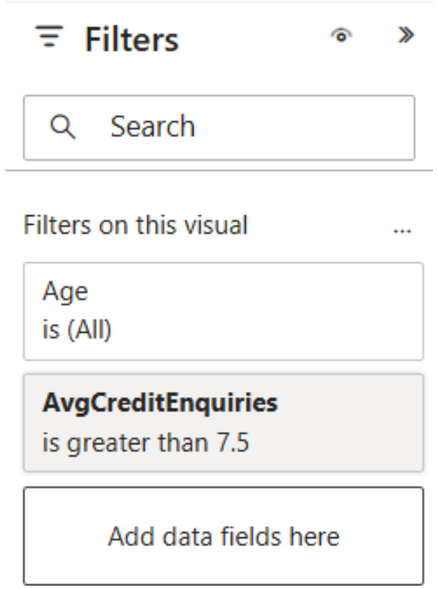


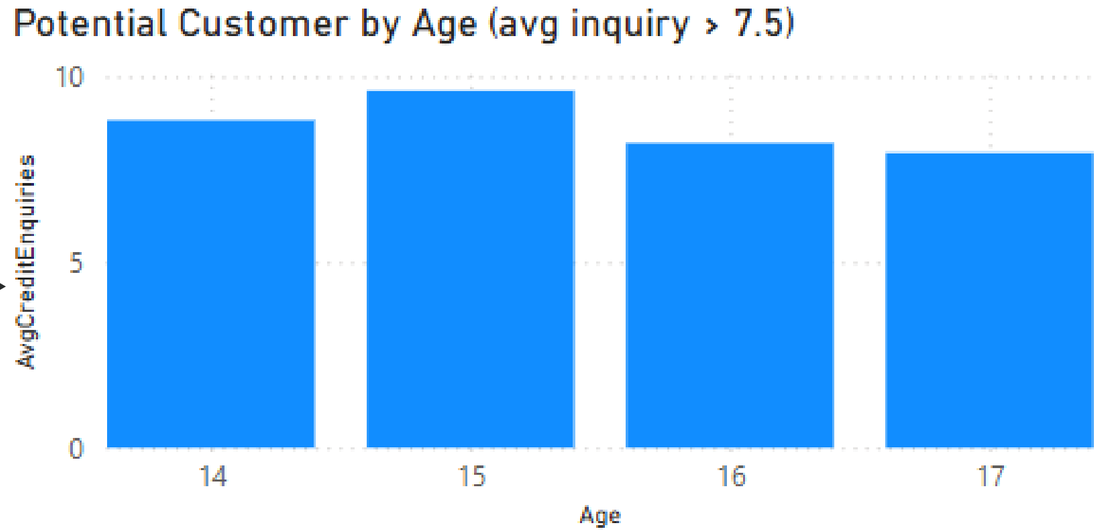
PROBLEM STATEMENT 8

From the data of previous task, I got clear that, I have few ages group as potential customers, however now I want in depth study of it, **all those ages will be my potential customers where age average inquiry is more than 7.5**

SOLUTION OF PROBLEM STATEMENT 8

To identify specific potential customers, I created a **bar chart** using the same **AvgCreditEnquiries** measure from the previous task and **applied a filter where average inquiries > 7.5**.
This helped highlight specific ages (14–17) with high interest in loans, making them strong potential customer segments.





PROBLEM STATEMENT 9

FINANCIAL ANALYTICS

I shared your insights on the basis of age groups to higher management and they are very happy with your insights. They have asked to find LTV scores for each year group, and based on the LTV score roll out following promotions.

if LTV > 80000 - “30% off on online purchases +home loan at 4% interest”

LTV between (60000-80000) - “15% off on online purchases+10000 worth gift hampers”

LTV between (50000-60000) - “Any loan at 5% interest rate”

to calculate LTV- use following formula -

LTV= (0.3*Average annual Income of the age)-(0.15*Average days in delay of payment from due date for that age)+0.4(Average of credit score[to calculate credit score go down good-3 bad -0])+(0.075*Average Amount Invested)+(0.075*Average Monthly balance)

```
LTV_Score =
var avg_annual_income = AVERAGE(combined_df[Annual_Income])
var avg_day_delay_payment = AVERAGE(combined_df[Delay_from_due_date])
var avg_amount_invested = AVERAGE(combined_df[Amount_invested_monthly])
var avg_monthly_balance = AVERAGE(combined_df[Monthly_Balance])
RETURN
(0.3*avg_annual_income) - (0.15*avg_day_delay_payment) + (0.4*credit_score
[avg_credit_score]) + (0.075*avg_amount_invested) + (0.075*avg_monthly_balan
```

SOLUTION OF PROBLEM STATEMENT 9

I calculated the LTV score using a measure and then created another measure called 'Promotion' to assign offers based on LTV. This helped identify which promotion to give each age group.

I used **table chart** for this question.

```
promotions =
SWITCH(TRUE(),[LTV_Score]>80000,"30% off on online purchases +home loan at 4% interest",
               [LTV_Score] > 60000, "15% off on online purchases+10000 worth gift hampers",
               [LTV_Score] > 50000, "Any Loan at 5% interest Rate",
               BLANK())
```

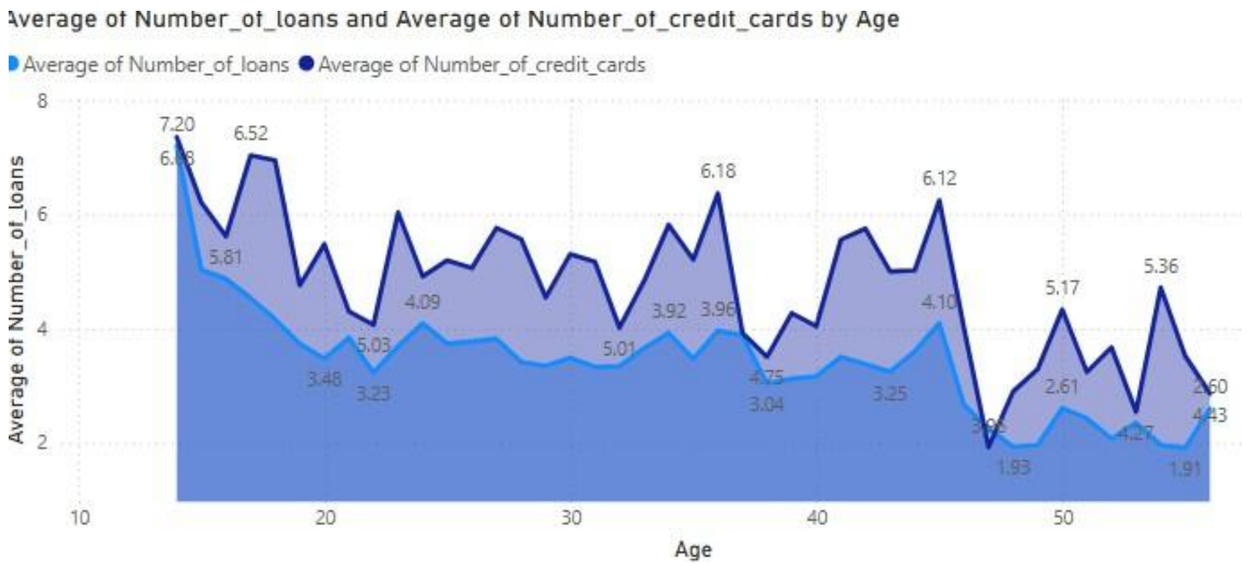
Age	LTV_Score	promotions
55	335510	30% off on online purchases +home loan at 4% interest
54	186420	30% off on online purchases +home loan at 4% interest
46	126391	30% off on online purchases +home loan at 4% interest
52	106173	30% off on online purchases +home loan at 4% interest
45	94001	30% off on online purchases +home loan at 4% interest
21	93941	30% off on online purchases +home loan at 4% interest
43	92903	30% off on online purchases +home loan at 4% interest
40	75755	15% off on online purchases+10000 worth gift hampers
20	67626	15% off on online purchases+10000 worth gift hampers
29	64975	15% off on online purchases+10000 worth gift hampers
22	61747	15% off on online purchases+10000 worth gift hampers
25	61492	15% off on online purchases+10000 worth gift hampers
15	60829	15% off on online purchases+10000 worth gift hampers
37	59282	Any Loan at 5% interest Rate
24	56889	Any Loan at 5% interest Rate
26	54735	Any Loan at 5% interest Rate
38	54002	Any Loan at 5% interest Rate
33	52549	Any Loan at 5% interest Rate
34	51167	Any Loan at 5% interest Rate
31	50445	Any Loan at 5% interest Rate
28	50179	Any Loan at 5% interest Rate

PROBLEM STATEMENT 10

Calculate the **average number of loans and credit cards held by customers across different age.** Present these averages and analyze the data to uncover trends and insights related to credit card ownership by age. This analysis will help in understanding how credit card usage varies with age, which can inform targeted credit card offers and marketing strategies.

SOLUTION OF PROBLEM STATEMENT 10

I analyzed the average number of loans and credit cards by age using an **area chart**.
This helped uncover how credit card ownership varies across age groups, supporting better targeting for credit card offers.



PROBLEM STATEMENT 11

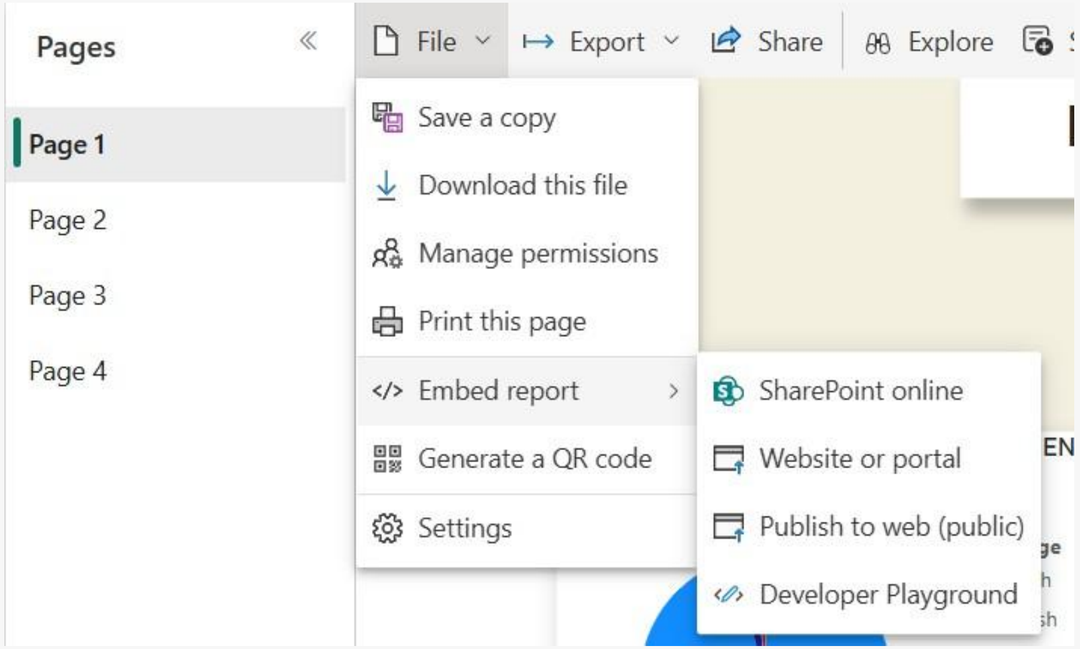
Please Create a visual that will show **the count of each type of loans** that are dispersed till now, this will help the company to keep a record of which is the most popular loans and roll out offers Accordingly.

SOLUTION OF PROBLEM STATEMENT 11

To solve this, I had a column containing multiple loan types for each customer ID, separated by commas. I used Power Query to handle this by duplicating the dataset and keeping only the Customer ID and Loan Type columns. Then, I split the Loan Type column using a comma delimiter and applied the unpivot feature to convert the separated columns into individual rows. Finally, I trimmed the values to remove any extra spaces. This allowed me to accurately count the total number of loans dispersed and identify the most popular loan types with help of **bar chart.**

PUBLISHING TO POWER BI SERVICES

PUBLISH < FILE < EMBED REPORT < WEBSITE OR PORTAL



LINK (you will need a microsoft account to access this dashboard)

https://app.powerbi.com/view?r=eyJrIjoiN2IwYTY5NmEtMDk2NC00NTcyLWI2YmItOTNkNmVmYTc0OTQxIiwidCI6ImNhZWFiMzBhLTJjODUtNDc3Ny1iYjgwLTM1ZWQwNmU0ODFkZSJ9

Thank you