EDULYT INDIA

PROJECT REPORT

ON

Customer's Spending Psychology and Repayment Discipline

Submitted By:

Shubham Gupta

in.sg5447@gmail.com

AI/ML

Amity University Noida

Problem Statement

Sanity Checks – Data Cleaning

• Note – All tasks need to be completed for month basis i.e. write logic for monthly calculations not cumulative for entire period.

• Provide a meaningful treatment to all values where age is less than 18.

• Is there any customer who have spent more than his/her Credit Limit for any particular month.

Tasks

- Monthly spend of each customer.
- Monthly repayment of each customer.
- Highest paying 10 customers.
- People in which segment are spending more money.
- Which age group is spending more money?
- Which is the most profitable segment?
- In which category the customers are spending more money?
- Impose an interest rate of 2.9% for each customer for any due amount.
- Monthly profit for the bank.

TOOLS USED: Microsoft EXCEL and Python.

Link (Excel and Jupyter Nb): https://drive.google.com/drive/folders/1YSkM-K6N-18x4rHVTwwmbv3rLvtQgRt0?usp=sharing

Dataset Description

This dataset comprises three interrelated Excel spreadsheets: Customer Acquisition, Spending, and Repayment. Together, they provide a comprehensive view of customer credit card usage and repayment behavior.

1. Customer Acquisition Sheet:

This sheet contains the demographic and onboarding details of each customer:

Column Name	Description
Masked Name	Anonymized customer identifier (e.g., A1, A2, ..., An)
Age	Age of the customer
City	City of residence
Credit Card Product	Type of credit card assigned (Platinum, Gold, or Silver)
Limit	Monthly spending limit on the credit card
Masked Company Name	Anonymized company name (e.g., C1, C2, ..., Cn)
Segment	Employment sector (e.g., Government, Salaried_MNC, Salaried_Pvt, Self-employed)

2. Spending Sheet

This sheet records monthly credit card spending activity by each customer:

Column Name	Description
Name	Masked name of the customer
Month	Transaction month (in date format)
Type	Spending category (e.g., Bus Ticket, Shopping, Petrol, Train Ticket, Food, Camera, Jewellery, Movie Ticket)
Amount	Monetary amount spent

3. Repayment Sheet

This sheet captures the amount each customer is expected to repay to the bank:

Column Name	Description
Name	Masked name of the customer
Month	Repayment month
Amount	Amount to be repaid to the bank

Data Cleaning and Preprocessing

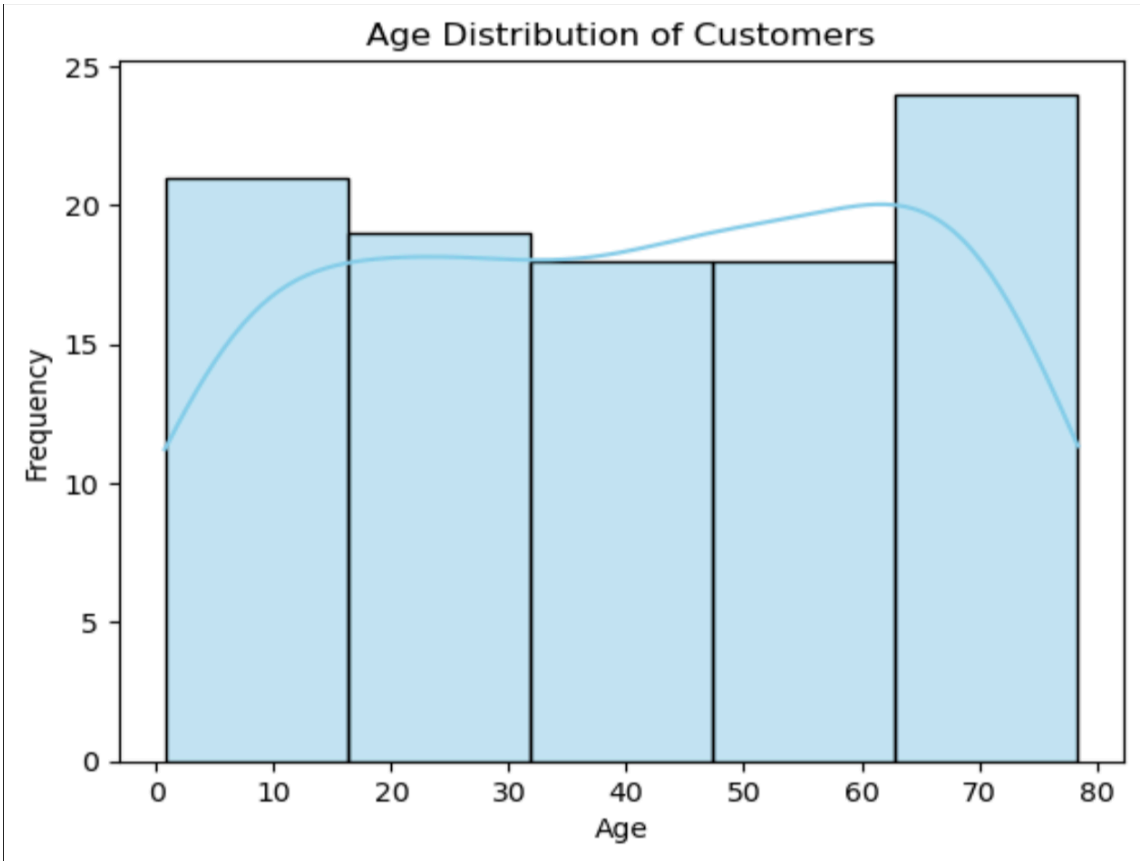
1. Handling Invalid Age Values

Upon inspection of the Age column in the **Customer Acquisition** sheet, it was observed that a few customers had age values less than 18 — an unrealistic scenario in the context of credit card ownership. To handle this:

- The mean age of all customers aged **18 and above** was calculated to be **49.29**.
- All entries with age below 18 were replaced with this mean value.

This approach was chosen because:

- The age distribution in the dataset is approximately **normal** (bell-shaped), with most values clustered around the center.
- There were **no significant outliers**, making the **mean** an appropriate measure for imputation.



2. **Age Group Categorization**

To enrich the demographic insights, customers were grouped into four meaningful age categories:

Age Range	**Category**
18–29	Young Adult
30–44	Adult
45–59	Middle Aged
60 and above	Senior

3. **Identifying Credit Limit Exceedance**

To determine whether any customer exceeded their **monthly credit limit**, the following steps were performed:

- Extracted **month and year** from the Month column in the **Spending** sheet and stored it in a new Month_Extracted column (e.g., *Jan 2024*).
- Grouped spending data by **Customer Name** and **Month_Extracted**, then calculated the **total monthly spend per customer**.
- Retrieved the credit limit of each customer from the Customer Acquisition sheet.
- Computed **exceedance** by subtracting the customer's total monthly spend from their credit limit.

If the result was **negative**, it indicated that the customer had exceeded their allowed limit for that month.

Costomer	Month_Extracted	SUMofSpend	Limit	LimitCheck
A1	Jan-04	1511173	500000	Exceeded
A1	Jan-05	398404	500000	Within Limit
A1	Feb-05	1404194	500000	Exceeded
A2	Jan-04	735264	100000	Exceeded
A3	Jan-04	647217	10000	Exceeded
A8	Feb-04	60302	100002	Within Limit
A4	Jan-04	1103216	10001	Exceeded
A9	Feb-05	810326	100003	Exceeded
A10	Feb-05	357397	500000	Within Limit
A11	Feb-05	492694	500000	Within Limit
A12	Feb-05	1208185	500000	Exceeded
A13	Feb-06	116944	500000	Within Limit
A14	Mar-06	1161758	500000	Exceeded
A15	Mar-06	774680	500000	Exceeded
A16	Mar-06	1881078	500000	Exceeded
A17	Mar-04	473416	500000	Within Limit
A18	Mar-06	1545245	100000	Exceeded
A19	Apr-05	484196	10000	Exceeded
A20	Apr-06	177517	10001	Exceeded
A21	Apr-04	659456	10002	Exceeded
A22	Apr-06	1569312	100000	Exceeded
A23	Apr-04	644108	100001	Exceeded
A24	Apr-05	561654	100002	Exceeded
A25	May-04	351505	100003	Exceeded
A26	May-06	623713	500000	Exceeded
A27	May-05	1282529	500000	Exceeded
A28	May-06	1533698	500000	Exceeded
A29	May-05	1028970	500000	Exceeded
A30	May-06	530869	500000	Exceeded
A31	Jul-06	176296	500000	Within Limit
A32	Aug-05	69630	500000	Within Limit
A33	Sep-04	337923	10002	Exceeded
A34	Nov-05	711559	100000	Exceeded
A35	Nov-06	65440	100001	Within Limit
A36	Nov-05	111362	100002	Exceeded
A37	Nov-04	369290	100003	Exceeded
A38	May-05	1146403	500000	Exceeded
A39	May-05	1079058	500000	Exceeded
A40	May-05	1362125	500000	Exceeded
A41	May-05	1744899	500000	Exceeded
A42	May-05	984958	500000	Exceeded
A43	May-05	1286680	500000	Exceeded
A44	May-05	1331513	500000	Exceeded
A45	May-05	805153	500000	Exceeded
A46	Jun-05	1245877	100000	Exceeded
A47	Jun-05	632810	10000	Exceeded
A48	Jun-05	404768	10001	Exceeded
A49	Jul-05	229824	10002	Exceeded

Count of LimitCheck		LimitCheck	
Costomer	Month_Extracted	Exceeded	Within Limit
A1	Apr-06	Yes	
	Aug-05		Yes
	Feb-04		Yes
	Feb-05	Yes	
	Jan-04	Yes	
	Jan-05		Yes
	May-04		Yes
	Nov-05		Yes
	Oct-06		Yes
A10	Apr-06		Yes
	Aug-06		Yes
	Feb-05		Yes
	Jan-04	Yes	
	Jul-06		Yes
	Mar-04		Yes
	Mar-06		Yes
	May-04		Yes
	May-05	Yes	
	May-06		Yes
	Nov-05		Yes
A100			Yes
A11	Apr-04		Yes
	Aug-05		Yes
	Feb-04	Yes	
	Feb-05		Yes
	Jan-04	Yes	
	Jun-05		Yes
	Jun-06		Yes
	Mar-06		Yes
	May-05		Yes
	May-06		Yes
	Nov-06		Yes
	Sep-06		Yes
A12	Apr-05	Yes	
	Apr-06		Yes
	Aug-06		Yes
	Feb-04		Yes
	Feb-05	Yes	
	Jan-04	Yes	
	Jan-05		Yes
	Jul-06		Yes
	Jun-05	Yes	
	Mar-06	Yes	
	May-05	Yes	
	Nov-05		Yes
	Oct-06		Yes
	Sep-04		Yes

NOTE: ***Advanced*** *Microsoft Excel* ***tools like Power Query and Pivot Tables are used for Data transformation, merging, and analysis***

The attached screenshots illustrate only a portion of the results. For the complete analysis and full results, please refer to the link provided above.

CONCEPT USED, METHODOLOY AND RESULTS

1. Monthly Spend of Each Customer

- **Concept Used:**

  To determine how much each customer spends in a month, data from the Spending sheet was utilized.

- **Methodology:**
  - Extracted Month_Extracted from the Month column.
  - Grouped data by Customer Name and Month_Extracted.
  - Aggregated total Amount spent by each customer for each month.

- **Result-**

Sum of Amount	Costomer				
Month_Extracted	A1	A10	A100	A11	A12
Jan-04	1511173	747428	42254	819545	744971
Feb-04	41381			501940	414392
Mar-04		435159			
Apr-04				493104	
May-04	131197	480729			
Sep-04					190053
Nov-04					
Jan-05	398404				287857
Feb-05	1404193	357396		492694	1208185
Apr-05					524412
May-05		696067		337785	859622
Jun-05				412747	929412
Jul-05					
Aug-05	129388			453981	
Sep-05					
Oct-05					
Nov-05	457317	484426			64787
Dec-05					
Jan-06					
Feb-06					
Mar-06		233741		148870	877582
Apr-06	564507	60213			179585
May-06		117296		103940	
Jun-06				462785	
Jul-06		267437			133727
Aug-06		230894			347129
Sep-06				92871	
Oct-06	220735				86873
Nov-06				261662	
Dec-06					
Grand Total	4858294	4110786	42254	4581924	6848587

2. Monthly Repayment of Each Customer

- **Concept Used:**

  To track how much each customer repaid to the bank each month, the Repayment sheet was used.

- **Methodology:**
  - Extracted Month_Extracted from the Month column.
  - Grouped data by Customer Name and Month_Extracted.
  - Aggregated total Amount repaid.

- **Result:**

Sum of Amount	Month_Extracted							
Costomer	Jan-04	Feb-04	Mar-04	Apr-04	May-04	Sep-04	Nov-04	Jan-05
A1	1362775.235	191180.0116						1581.969829
A10	1149997.084		266929.3785					446068.3285
A100	151141.3919							
A11	1052386.318							

3. Top 10 Highest Paying Customers

- **Concept Used:**

  Identified the top 10 customers who paid the highest cumulative repayment amounts month wise.

- **Methodology:**
  - Grouped the Repayment data by customer.
  - Summed the total amount paid.
  - Sorted in descending order and selected the top 10 customers.

- **Result:**

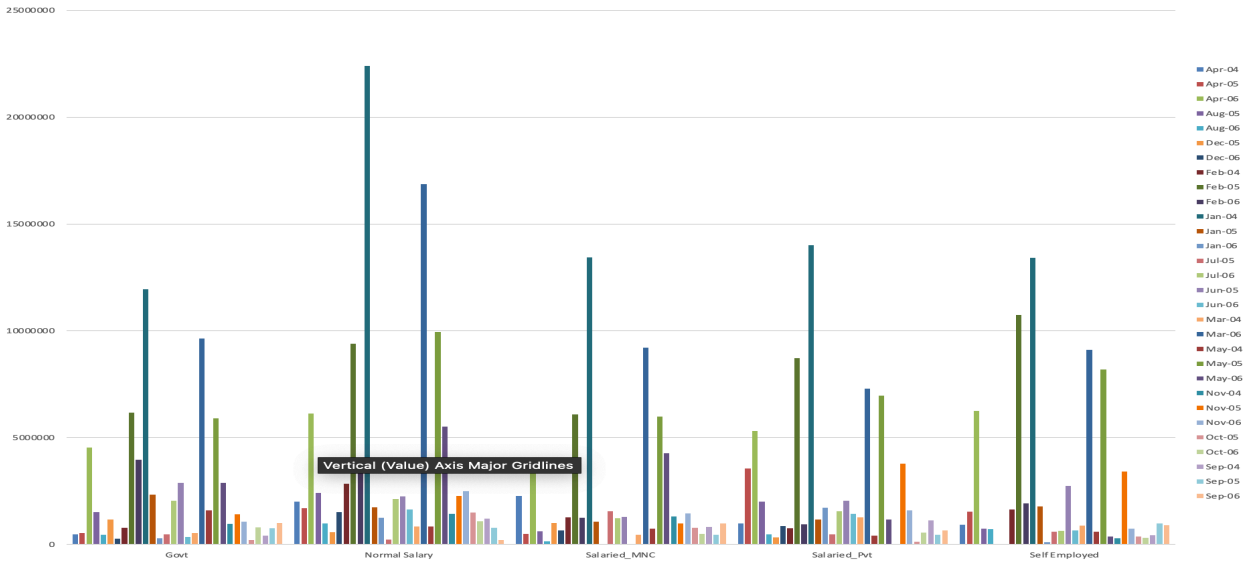
MonthlyAmount		
Month_Extracted	Costomer	Total
Jan-04	A18	1760599.631
	A22	1921997.848
	A39	2020852.819
	A40	2496299.125
	A41	3026775.837
	A42	1947732.961
	A43	2362765.475
	A5	1802038.712
	A60	1985479.054
	A7	1646591.586
Jan-04 Total		20971133.06
Feb-04	A27	341428.7783
	A30	463060.7497
	A32	463114.5575
	A33	382097.5135
	A36	253815.0194
	A44	329739.1897
	A48	294439.8826
	A49	638576.0721
	A57	512881.454
	A9	384894.321
Feb-04 Total		4064047.538
Mar-04	A10	266929.3785
	A14	309081.9588
	A17	256148.1189
	A20	489900.5599
	A40	449060.9196
	A41	465141.0571
	A51	302975.5031
	A57	491491.1122
	A58	522999.3065
	A59	222355.3626

4. Segment-Wise Spending Analysis

- **Concept Used:**

  To identify which customer segment (e.g., Government, Salaried_MNC, Salaried_Pvt, Self-employed) spends the most.

- **Methodology:**
  - Merged Customer Acquisition and Spending sheets using Customer Name.
  - Grouped by Segment.
  - Summed total Amount spent by each segment.
- **Result:**



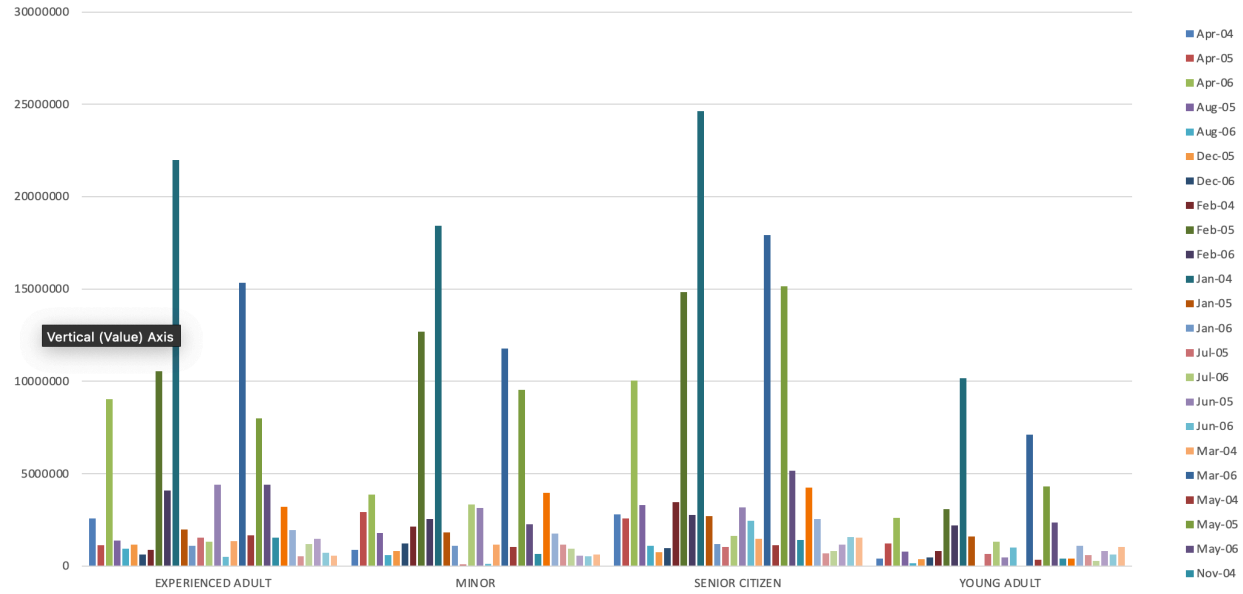
5. Age Group-Wise Spending Analysis

- **Concept Used:**

  To determine which age group contributes the most to spending.

- **Methodology:**
  - Used the earlier age group categorization (Young Adult, Adult, Middle Aged, Senior).
  - Merged with Spending data.
  - Grouped by Age Group and summed total Amount spent.

- **Result:**



6. Most Profitable Segment

- **Concept Used:**

  The most profitable segment is identified based on repayment minus spend.

- **Methodology:**
  - Merged Customer Acquisition, Spending, and Repayment data.
  - Calculated total repayment and total spend per segment.
  - Profit = Total Repayment – Total Spend
  - Segment with the highest net profit was marked as the most profitable.

- **Result:**

Sum of Profit		
Month_Ex	Segment	Total
Apr-04	Salaried_MNC	847719.5527
	Self Employed	64776.03606
	Normal Salary	52133.5248
	Salaried_Pvt	47121.74553
	Govt	0
Apr-04 Total		1011750.859
Apr-05	Salaried_Pvt	651256.5103
	Normal Salary	218290.2773
	Self Employed	207493.9039
	Salaried_MNC	161846.2611
Apr-05 Total		1238886.953
Apr-06	Govt	1854571.55
	Normal Salary	1711076.401
	Self Employed	1570844.942
	Salaried_Pvt	1482324.869
	Salaried_MNC	864070.5025
Apr-06 Total		7482888.265
Aug-05	Normal Salary	729605.7016
	Salaried_MNC	443779.9014
	Salaried_Pvt	270343.6379
	Govt	98530.0876
	Self Employed	26893.41791
Aug-05 Total		1569152.746

7. Category-Wise Spending Analysis

- **Concept Used:**

  To find which spending category (e.g., Food, Shopping, Petrol) customers spend the most on.

- **Methodology:**
  - Grouped Spending data by Type.
  - Aggregated the total Amount for each category.
  - Sorted in descending order to identify top categories.
- **Result:**

Sum of Amount		
Month_Extracted	Type	Total
Apr-04	BUS TICKET	1705406
	SHOPPING	1197078
	PETRO	1039469
	TRAIN TICKET	663563
	FOOD	630585
	CAMERA	493104
	JEWELLERY	478293
	MOVIE TICKET	447983
Apr-04 Total		6655481
Apr-05	TRAIN TICKET	3456029
	CAMERA	1771783
	FOOD	983073
	JEWELLERY	744476
	AIR TICKET	469661
	CLOTHES	257587
	PETRO	122290
	BIKE	29709
Apr-05 Total		7834608
Apr-06	PETRO	3958966
	SHOPPING	3585639
	JEWELLERY	3299756
	AIR TICKET	3268221
	RENTAL	2396511
	MOVIE TICKET	2240242
	CAMERA	1823903
	TRAIN TICKET	1390284
	FOOD	1125744
	BIKE	863625
	CLOTHES	431184
	CAR	337909
	AUTO	328781
	SANDALS	310482
	BUS TICKET	185061
Apr-06 Total		25546308

8. Interest Calculation for Due Amount

- **Concept Used:**

  A 2.9% interest was applied to any amount a customer failed to repay.

- **Methodology:**
  - For each customer per month, calculated:

    Due = Spend – Repayment

- If Due > 0, then:

  Interest = Due × 2.9%

- Net Payable Amount = Due Amount + Interest Amount

- **Result:**

Repayment		Spend		DueAMt (Spend- Repay)	Interest	NetAmountPayble
Costomer	Total	Costomer	Total			
A1	3831937.75	A1	4858294	1026356.58	29764.34	1056120.92
A10	5230311.56	A10	4110786	0.00	0.00	0.00
A100	151141.39	A100	42254	0.00	0.00	0.00
A11	4735300.27	A11	4581924	0.00	0.00	0.00
A12	7572754.17	A12	6848587	0.00	0.00	0.00
A13	7919565.93	A13	8042339	122772.88	3560.41	126333.29
A14	6778903.68	A14	7247760	468855.95	13596.82	482452.78
A15	4075423.32	A15	4766802	691378.25	20049.97	711428.22
A16	4539255.16	A16	5944043	1404787.52	40738.84	1445526.36

9. Monthly Profit for the Bank

- **Concept Used:**

  The bank's monthly profit is derived from:

  - The positive difference between what customers repay and what they spend.
  - Plus, interest earned from due payments.

- **Methodology:**

  - For each customer per month:

    Profit = (Repayment – Spend) + Interest (if any)

  - Aggregated monthly across all customers.

- **Result:**