

Bank Customer Dataset Analysis



By: Yougender Chauhan

Lekha Ajit Kumar



California State University, Los Angeles

Submitted to Dr. Shilpa Balan

INDEX

Part A: DATASET(S)/ DATA SOURCE URL	2
Part B: Introduction	2
Part C: Data Description	4
Part D: Data cleaning	5
Part E: Analysis & Visualizations	12
Part F: Statistical summary	22
Part G: Statistical tests	24
CONCLUSION	28

Part A: DATASET(S)/ DATA SOURCE URL:

<https://www.kaggle.com/datasets/xanikalsen/bank-customers-dataset>

Part B: Introduction

The primary objective of this research is to conduct a comprehensive analysis of various facets of customer demographics and financial behavior within the context of banking services. It seeks to provide valuable insights into the intricate relationship between the demographics, financial behaviors, and socio-economic attributes of the bank's customer base. Through this investigation, we aim to unravel the ways in which factors such as education, gender, property ownership, marital status, income, age, and ways of living interplay with banking-related behaviors and outcomes. The ultimate goal is to offer a deeper understanding of how these aspects influence or correlate with the customers' interactions with our banking services. This analytical approach is not just an academic pursuit but a practical strategy to inform and optimize our banking operations, marketing endeavors, and customer service initiatives.

Demographics and Banking Behavior: By delving into customer demographics, this study provides a comprehensive view of the diverse backgrounds and socio-economic characteristics that define our customer base. Recognizing the rich tapestry of our clientele is pivotal, as it equips us with the knowledge of who our customers are and the financial landscapes they navigate.

Insights into Financial Behavior: The study endeavors to unearth discernible patterns and trends governing how customers engage with the bank. It explores critical factors such as credit status, the impact of promotional activities, property ownership, and income levels, all of which

are fundamental components of financial behavior. These insights enable us to tailor our services to better meet the diverse financial needs of our customers.

Identification of Correlations and Influences: A central mission of this research is to pinpoint correlations and influences that may exist within our customer base. It aspires to address questions such as how gender might influence property ownership or how education levels relate to marital status. Such discoveries are instrumental in understanding the underlying motivations and decision-making processes of our customers, aiding us in crafting more effective strategies and personalized services.

REFERENCES:

1. Article title: Statistical analysis of bank deposits dataset.

URL: <https://www.sciencedirect.com/science/article/pii/S2352340918303093>

Date published: March 26, 2018

2. Article title: An Analysis of Bank Financial Strength Ratings and Credit Rating Data.

URL: <https://www.mdpi.com/2227-9091/9/9/155>

Date published: August 26, 2021

3. Article title: How can banking data analysis mitigate financial risks?

URL: <https://www.linkedin.com/advice/0/how-can-banking-data-analysis>

Website title: How Data Analysis Can Mitigate Banking Risks.

Date published: September 18, 2023

Part C: Data Description

Screenshot of the dataset:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
ID	CREDITSTATUS	AGE	GENDER	CAMPAIGN	OWNCAR	OWNPROPERTY	CHILDRENCOUNT	FAMSIZE	INCOMETOTAL	INCOMETYPE	EDUCATION	MARITALSTA	HOUSINGTY	MOBILE	EMAIL	OCCUPATION
1	5008824	0	26 M	1	Y	Y	0	2	135000	Commercial	Secondary / Married	House / apartY	N	Labors		
2	5008842	0	25 M	2	N	Y	1	3	405000	Commercial	Higher educu	Married	House / apartY	N	Managers	
4	5008852	0	24 M	2	Y	Y	0	2	193471	Commercial	Secondary / Married	House / apartY	N	Drivers		
5	5008872	0	30 M	2	Y	Y	0	2	360000	Commercial	Secondary / Married	House / apartY	N	Security staff		
6	5008926	0	27 F	3	N	Y	0	1	297000	Commercial	Secondary / Single / not i	Rented apartY	N	Labors		

Data Description of Used Columns

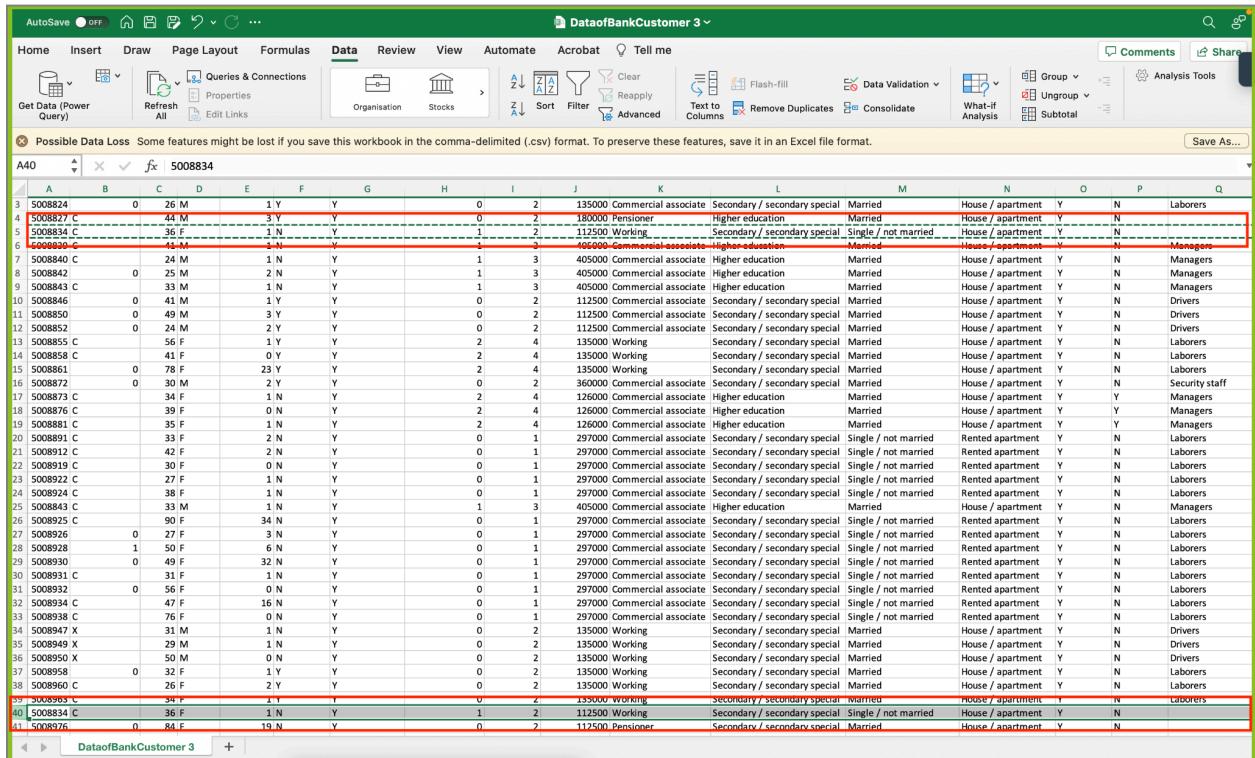
NAME	DESCRIPTION	Example value
CREDITSTATUS	Credit loan status 0: 1-29 days past due 1: 30-59 days past due 2: 60-89 days overdue 3: 90-119 days overdue 4: 120-149 days overdue 5: Bad debts / Write-offs C: paid off for that month X: no loan for the month	2
AGE	Age of the Client	59
GENDER	Gender (M = Male, F = Female)	F
CAMPAIGN	Number of Promotions	13
OWNPROPERTY	Owns a car (Y = Yes, N = No)	Y

CHILDRENCOUNT	Number of children	1
INCOMETOTAL	Annual Income	135000
INCOMETYPE	Income category	Working
EDUCATIONLEVEL	Education level	Higher education
MARITALSTATUS	Marital status	Married
HOUSINGTYPE	Way of living	House/apartment
MOBILE	Owns a mobile phone (1 = Yes, 0 = No)	1
EMAIL	Has an e-mail account (1 = Yes, 0 = No)	1
OCCUPATION	Occupation	Managers
FAMSIZE	Family size	5

Part D: Data cleaning

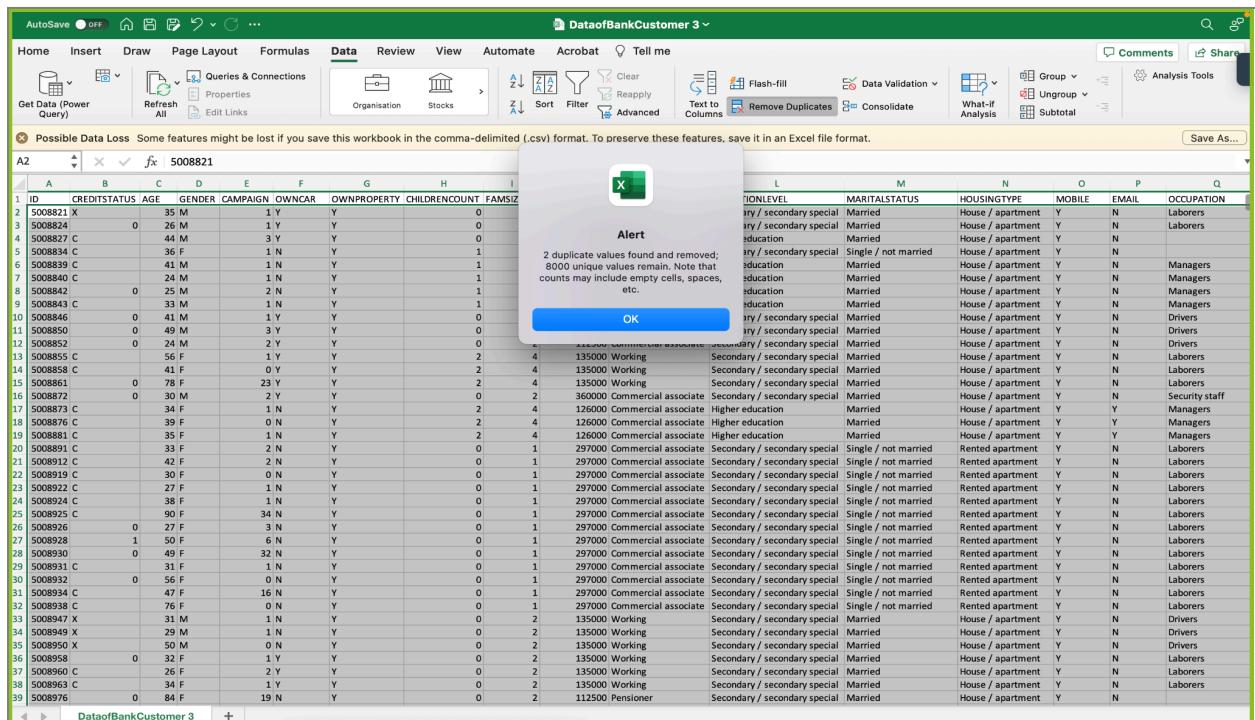
1. Drop Duplicates: Our data had a few duplicate records. In the below screenshot, it shows that rows 5 and 40 are the same. Duplicates can significantly impact the quality, accuracy, and reliability of our data and lead to inaccurate results in analysis. Hence we dropped the duplicate records.

BEFORE



Screenshot of an Excel spreadsheet titled "DataofBankCustomer 3". The data consists of approximately 40 rows and 18 columns. The columns include ID, Age, Gender, and various status and occupation codes. Several rows are highlighted with red boxes, indicating data that needs to be cleaned or processed. The rows highlighted are: Row 3 (ID 5008824), Row 4 (ID 5008827), Row 5 (ID 5008834), Row 6 (ID 5008830), Row 7 (ID 5008840), Row 8 (ID 5008842), Row 9 (ID 5008843), Row 10 (ID 5008846), Row 11 (ID 5008850), Row 12 (ID 5008852), Row 13 (ID 5008855), Row 14 (ID 5008858), Row 15 (ID 5008861), Row 16 (ID 5008872), Row 17 (ID 5008873), Row 18 (ID 5008875), Row 19 (ID 5008881), Row 20 (ID 5008891), Row 21 (ID 5008912), Row 22 (ID 5008919), Row 23 (ID 5008922), Row 24 (ID 5008924), Row 25 (ID 5008843), Row 26 (ID 5008925), Row 27 (ID 5008926), Row 28 (ID 5008928), Row 29 (ID 5008930), Row 30 (ID 5008958), Row 31 (ID 5008960), Row 32 (ID 5008963), Row 33 (ID 5008976), Row 34 (ID 5008976), Row 35 (ID 5008949), Row 36 (ID 5008950), Row 37 (ID 5008958), Row 38 (ID 5008960), Row 39 (ID 5008976), and Row 40 (ID 5008834). The rows are highlighted in red boxes, suggesting they are part of a specific subset or require attention.

AFTER



Screenshot of the same Excel spreadsheet after data cleaning. The red boxes from the previous screenshot are no longer present, indicating the data has been processed. An alert dialog box is visible in the center, stating "2 duplicate values found and removed; 8000 unique values remain. Note that counts may include empty cells, spaces, etc." with an "OK" button. The data structure remains the same, with approximately 40 rows and 18 columns, but the rows highlighted in red are no longer present.

2. Fill with Average: We had missing values for the column “Income Total” for a few of the rows, so we took an average of the remaining data for that column and added in the values via an Excel formula.

BEFORE

Possible Data Loss		Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.																	
J36		Clear																	
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S				
1	GENDER	CAMPAGN	OWNCNR	OWNPROPERTY	CHILDRENCOUNT	FAMSIZE	INCOMETOTAL	INCOMETYPE	EDUCATIONLEVEL	MARITALSTATUS	HOUSINGTYPE	MOBILE	EMAIL	OCCUPATION					
2	M	1	Y	Y	0	2	135000	Commercial associate	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
3	M	1	Y	Y	0	2	135000	Commercial associate	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
4	M	1	N	Y	1	3	45000	Commercial associate	Higher education	Married	House / apartment	Y	N	Managers					
5	M	2	N	Y	1	3	45000	Commercial associate	Higher education	Married	House / apartment	Y	N	Managers					
6	M	1	N	Y	1	3	40500	Commercial associate	Higher education	Married	House / apartment	Y	N	Managers					
7	M	2	Y	Y	0	2	4	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
8	F	1	Y	Y	0	2	4	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
9	F	0	N	Y	0	2	4	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
10	M	2	Y	Y	0	2	360000	Commercial associate	Secondary / secondary special	Married	House / apartment	Y	N	Staff					
11	F	1	N	Y	2	4	126000	Commercial associate	Higher education	Married	House / apartment	Y	Y	Managers					
12	F	0	N	Y	2	4	126000	Commercial associate	Higher education	Married	House / apartment	Y	Y	Managers					
13	F	0	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers					
14	F	1	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers					
15	F	1	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers					
16	F	3	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers					
17	F	3	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers					
18	F	6	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers					
19	F	32	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers					
20	F	1	N	Y	0	1	1	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers					
21	F	1	N	Y	0	1	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers				
22	F	15	N	Y	0	1	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers				
23	F	0	N	Y	0	1	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers				
24	M	1	N	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
25	M	1	N	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
26	M	0	N	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
27	F	1	Y	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
28	F	2	Y	Y	0	2	330000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
29	F	1	Y	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
30	M	1	Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
31	M	1	Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
32	M	1	Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
33	M	52	Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
34	M	1	Y	Y	0	2	235000	Commercial associate	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					
35	F	1	N	Y	0	1	315000	Working	Secondary / secondary special	Single / not married	House / apartment	Y	N	Accountants					
36	F	3	Y	Y	1	3	1	State servant	Higher education	Married	House / apartment	Y	Y	Core staff					
37	F	1	N	Y	0	2	112500	Working	Higher education	Civil marriage	House / apartment	Y	Y	Drivers					
38	F	2	N	Y	0	2	112500	Working	Higher education	Civil marriage	House / apartment	Y	Y	Drivers					
39	M	0	N	Y	0	2	90000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers					

AFTER

Data from DataBankCustomer 3																		
Home		Insert		Draw		Page Layout		Formulas		Data		Review		View		Automate		
Calibri (Body)	12	A ¹	A ²	Number	Conditional Formatting	Insert	Format as Table	Format Cells	Format	Sort & Filter	Find & Select	Sensitivity	Comments	Share				
Paste																		
B	I	U	C															
Possible Data Loss: Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.																		
J7																		
1	GENDER	CAMPAIGN	OWNCNR	OWNPROPERTY	CHILDRENCOUNT	FAMSIZE	INCOMETOTAL	INCOMEME	EDUCATIONLEVEL	MARITALSTATUS	HOUSINGTYPE	MOBILE	EMAIL	OCCUPATION	R	S		
2	1	Y	Y	Y	0	2	135000	Commercial associate	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
3	M	1	Y	Y	0	2	135000	Commercial associate	Secondary / secondary special	Married	House / apartment	Y	N	Labors				
4	M	1	N	Y	1	3	405000	Commercial associate	Higher education	Married	House / apartment	Y	N	Managers				
5	M	2	N	Y	1	3	405000	Commercial associate	Higher education	Married	House / apartment	Y	N	Managers				
6	M	1	N	Y	1	3	405000	Commercial associate	Higher education	Married	House / apartment	Y	N	Managers				
7	M	2	Y	Y	0	3	193475	Commercial associate	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
8	F	1	Y	Y	2	4	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Labors				
9	F	0	Y	Y	2	4	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Labors				
10	M	2	Y	Y	0	2	280000	Commercial associate	Secondary / secondary special	Married	House / apartment	Y	N	Security staff				
11	F	1	N	Y	2	4	126000	Commercial associate	Higher education	Married	House / apartment	Y	N	Managers				
12	F	0	N	Y	2	4	126000	Commercial associate	Higher education	Married	House / apartment	Y	N	Managers				
13	F	0	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
14	F	1	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
15	F	1	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
16	F	34	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
17	F	3	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
18	F	5	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
19	F	32	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
20	F	1	N	Y	0	1	193475	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
21	F	0	N	Y	0	2	280000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
22	F	16	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
23	F	0	N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors				
24	M	1	N	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
25	M	1	N	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
26	M	1	N	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
27	F	1	V	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Labors				
28	F	2	Y	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Labors				
29	F	1	Y	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Labors				
30	M	1	Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
31	M	1	Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
32	M	1	Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
33	M	52	Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				
34	F	1	V	Y	0	2	225000	Commercial associate	Secondary / secondary special	Single / not married	House / apartment	Y	N	Labors				
35	F	1	N	Y	0	2	225000	Commercial associate	Secondary / secondary special	Single / not married	House / apartment	Y	N	Accountants				
36	F	3	Y	Y	1	3	193475	State servant	Higher education	Married	House / apartment	Y	Y	Core staff				
37	F	1	N	Y	0	2	122000	Working	Higher education	Civil marriage	House / apartment	Y	Y	Labors				
38	F	2	N	Y	0	2	112500	Working	Higher education	Civil marriage	House / apartment	Y	Y	Labors				
39	M	0	N	Y	0	2	90000	Working	Secondary / secondary special	Married	House / apartment	Y	N	Drivers				

3. Drop Nulls: We had missing values in our records, hence we dropped the rows with missing values using Excel functionality. We clicked Find & Select -> Then Clicked Go to Special ->Choose Blanks -> Clicked OK, and then all the blank rows/cells were highlighted -> Choose the Delete under Cells section on the Home Tab -> Then Clicked Delete Sheet Rows. Hence we were able to delete rows with empty cells in our dataset.

BEFORE

The screenshot shows a Microsoft Excel spreadsheet titled 'DataofBankCustomer 3'. The data consists of 39 rows and 16 columns. The columns are labeled A through Q. The data includes various personal and financial information such as ID, age, gender, credit status, campaign details, property ownership, family size, income, education level, marital status, housing type, mobile number, email, and occupation. Many cells in the 'CREDITSTATUS' column are empty (pink highlighted) and are being deleted.

AFTER

4. Data Formatting: Formatting the Total Income column, we noticed that in our data some rows had extra decimal points, and some had dollar signs. We wanted the data to be consistent; hence we removed the decimal points and dollar signs using the options “Accounting Number Format” and “Decrease Decimal” options available in Excel.

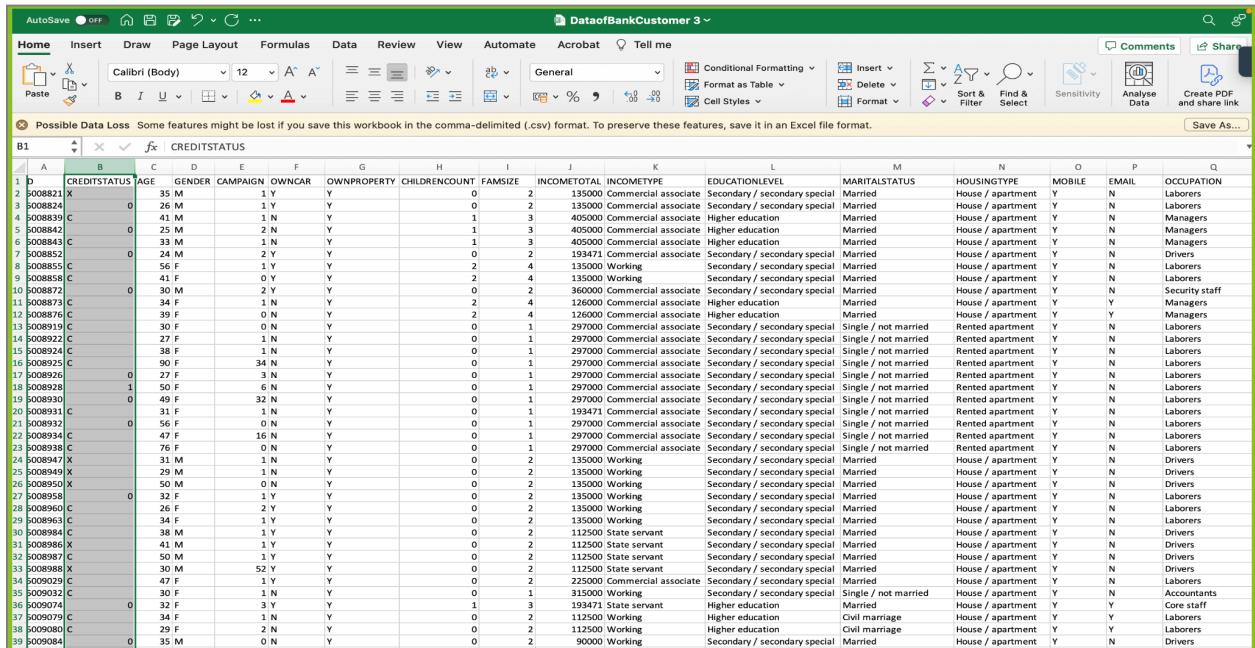
BEFORE

AFTER

The screenshot shows a Microsoft Excel spreadsheet titled 'DataofBankCustomer 3'. The data is sorted by the 'CREDITSTATUS' column. The 'CREDITSTATUS' column contains values 'C' and 'X'. After sorting, 'C' is converted to 6 and 'X' is converted to 7. The table includes columns for ID, CREDITSTATUS, AGE, GENDER, CAMPAIGN, OWNCAR, OWNPROPERTY, CHILDRENCOUNT, FAMSIZE, INCOMETOTAL, INCOMETYPE, EDUCATIONLEVEL, MARITALSTATUS, HOUSINGTYPE, MOBILE, EMAIL, OCCUPATION, and various job titles like Commercial associate, State servant, Working, etc. The table has 39 rows of data.

5. Sorting Data: We sorted the data in the Credit Status columns cause data sorting helped us to arrange the data into a meaningful order to make it easier to understand, analyze, and visualize. We had both numeric and non-numeric values in this column. After sorting we also converted the non-numeric column into numeric values, which we personally assigned. All "C" were converted to 6 and all the "X" were converted to 7.

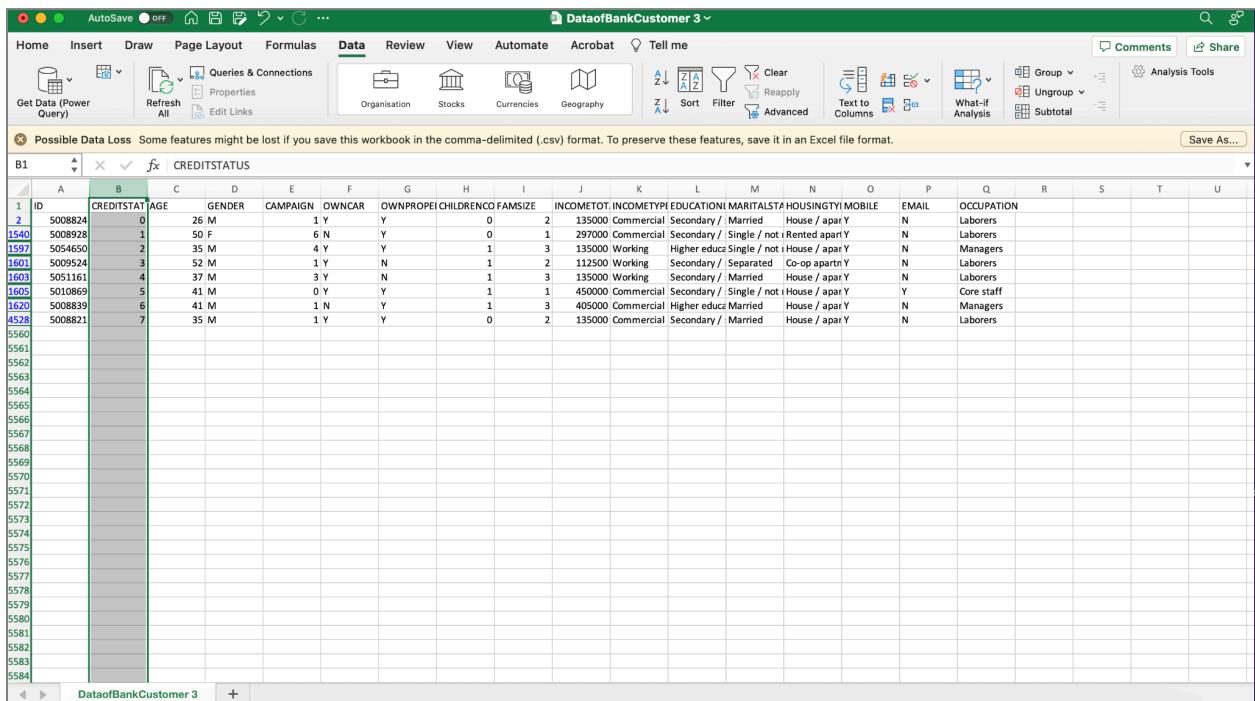
BEFORE



B1	CREDITSTATUS	AGE	GENDER	CAMPAIGN	OWNCAR	OWNPROPERTY	CHILDRENCOUNT	FAMSIZE	INCOMETOTAL	INCOMETYPE	EDUCATIONLEVEL	MARITALSTATUS	HOUSINGTYPE	MOBILE	EMAIL	OCCUPATION
1	b	35 M	1 Y	Y	0	2	135000	Commercial associate	Secondary / secondary special	Married	House / apart Y	Y	N	Labors		
2	5008821	X	26 M	1 Y	Y	0	2	135000	Commercial associate	Secondary / secondary special	Married	House / apart Y	Y	N	Labors	
3	5008824	0	41 M	1 N	Y	1	3	405000	Commercial associate	Higher education	Married	House / apart Y	Y	N	Managers	
4	5008839	C	21 M	0 N	Y	1	3	405000	Commercial associate	Higher education	Married	House / apart Y	Y	N	Managers	
5	5008842	0	33 M	1 N	Y	1	3	405000	Commercial associate	Higher education	Married	House / apart Y	Y	N	Managers	
6	5008843	C	24 M	2 Y	Y	0	2	193471	Commercial associate	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
7	5008852	0	56 F	1 Y	Y	2	4	135000	Working	Secondary / secondary special	Married	House / apart Y	Y	N	Labors	
8	5008855	C	41 F	0 Y	Y	2	4	135000	Working	Secondary / secondary special	Married	House / apart Y	Y	N	Labors	
9	5008858	C	30 M	2 Y	Y	0	2	360000	Commercial associate	Secondary / secondary special	Married	House / apart Y	Y	N	Security staff	
10	5008872	0	34 F	0 N	Y	2	4	126000	Commercial associate	Higher education	Married	House / apart Y	Y	Y	Managers	
11	5008873	C	39 F	27 F	0 N	Y	0	4	126000	Commercial associate	Higher education	Married	House / apart Y	Y	Y	Managers
12	5008876	C	29 F	0 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
13	5008891	C	27 F	1 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
14	5008922	C	38 F	1 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
15	5008924	C	90 F	34 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
16	5008925	C	27 F	3 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
17	5008926	0	50 F	6 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
18	5008928	1	49 F	32 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
19	5008930	0	41 F	1 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
20	5008931	0	56 F	1 N	Y	0	1	193471	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
21	5008932	0	56 F	0 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
22	5008934	C	47 F	16 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
23	5008938	C	76 F	0 N	Y	0	1	297000	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Labors	
24	5008947	X	31 M	1 N	Y	0	2	135000	Working	Secondary / secondary special	Single / not married	Rented apartment	Y	N	Drivers	
25	5008949	X	29 M	1 N	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
26	5008950	X	50 M	0 N	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
27	5008958	0	32 F	1 Y	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
28	5008960	C	36 F	2 Y	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
29	5008963	C	34 F	1 Y	Y	0	2	135000	Working	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
30	5008984	C	38 M	1 Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
31	5008986	X	41 M	1 Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
32	5008987	C	50 M	1 Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
33	5008989	X	30 M	52 Y	Y	0	2	112500	State servant	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	
34	5008992	C	37 F	1 Y	Y	0	2	220000	Commercial associate	Secondary / secondary special	Single / not married	House / apart Y	Y	N	Drivers	
35	5008993	C	30 F	1 N	Y	0	1	193471	Working	Secondary / secondary special	Single / not married	House / apart Y	Y	N	Accountants	
36	5008974	0	32 F	3 Y	Y	1	3	193471	State servant	Higher education	Married	House / apart Y	Y	Y	Core staff	
37	5008979	C	34 F	1 N	Y	0	2	112500	Working	Higher education	Civil marriage	House / apart Y	Y	Y	Labors	
38	5008980	C	29 F	2 N	Y	0	2	112500	Working	Higher education	Civil marriage	House / apart Y	Y	Y	Labors	
39	5008984	0	35 M	0 N	Y	0	2	90000	Working	Secondary / secondary special	Married	House / apart Y	Y	N	Drivers	

AFTER

Note: In this Image, Only the unique records are shown. It is to show that all the non-numeric data is being converted into numeric.



B1	CREDITSTATUS	AGE	GENDER	CAMPAIGN	OWNCAR	OWNPROPERTY	CHILDRENCOUNT	FAMSIZE	INCOMETOTAL	INCOMETYPE	EDUCATIONLEVEL	MARITALSTATUS	HOUSINGTYPE	MOBILE	EMAIL	OCCUPATION
1	5008824	0	26 M	1 Y	Y	0	2	135000	Commercial	Secondary /	Married	House / apart Y	Y	N	Labors	
1540	5008928	1	50 F	6 N	Y	0	1	297000	Commercial	Secondary /	Single / not	Rented apart Y	Y	N	Labors	
1597	5054650	2	35 M	4 Y	Y	1	3	135000	Working	Higher educ	Single / not	House / apart Y	Y	N	Managers	
1601	5009524	3	52 M	1 Y	N	1	2	112500	Working	Secondary /	Separated	Co-op apart Y	Y	N	Labors	
1603	5051161	4	37 M	3 Y	N	1	3	135000	Working	Secondary /	Married	House / apart Y	Y	N	Labors	
1605	5010869	5	41 M	0 Y	Y	1	1	450000	Commercial	Secondary /	Single / not	House / apart Y	Y	N	Core staff	
1620	5008839	6	41 M	1 N	Y	1	3	405000	Commercial	Higher educ	Married	House / apart Y	Y	N	Managers	
4528	5008821	7	35 M	1 Y	Y	0	2	135000	Commercial	Secondary /	Married	House / apart Y	Y	N	Labors	

Part E: Analysis & Visualizations

1. **Question of Analysis** - How can we tailor our banking services to better meet the distinct financial needs and preferences of male and female customers based on their educational backgrounds?

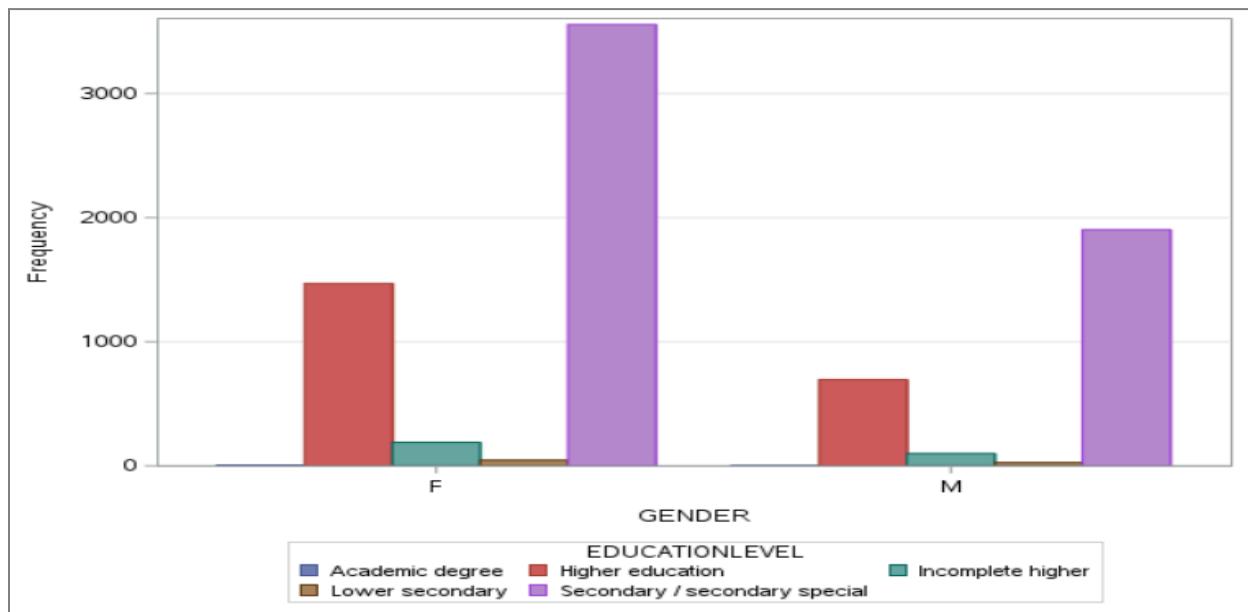
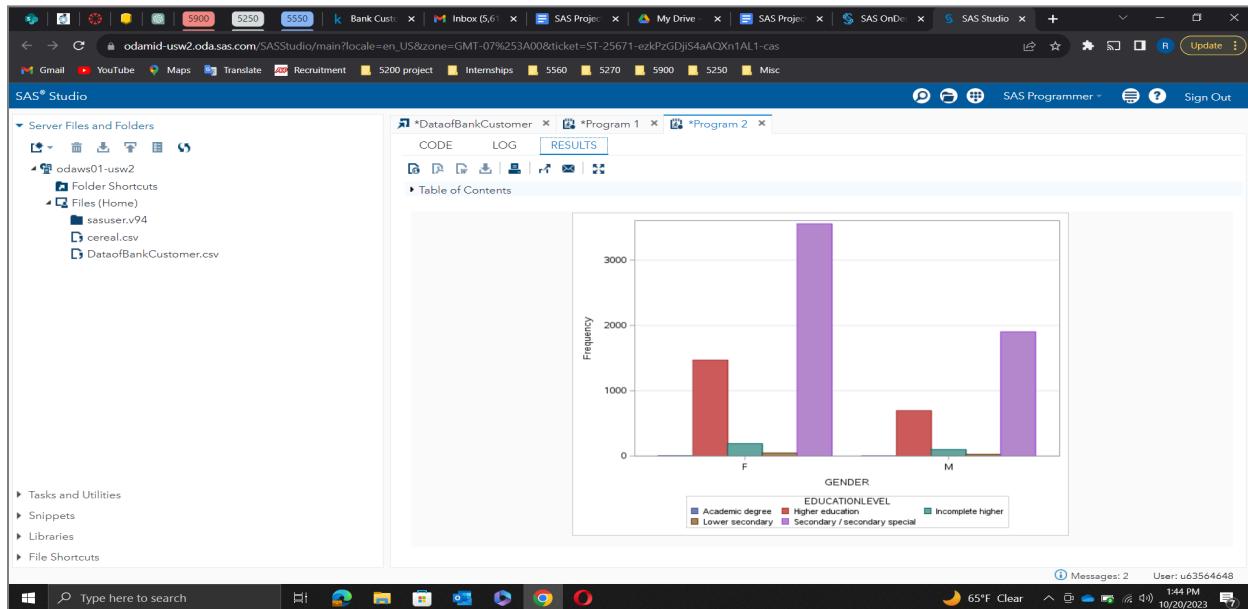


Figure 1. Frequency of customer's education level by gender.

The vertical bar chart in Figure 1 shows the customer's frequency count of each education level by gender. It is clear that female customers surpass the males in each category of education level, from holding an academic degree to graduating from secondary or special secondary schools. By using data-driven tools and tailored services, banks can help their customers avoid financial instability and achieve their savings goals. As financial coaching becomes increasingly important, banks that are proactive in supporting their customers are likely to see increased customer satisfaction and loyalty.

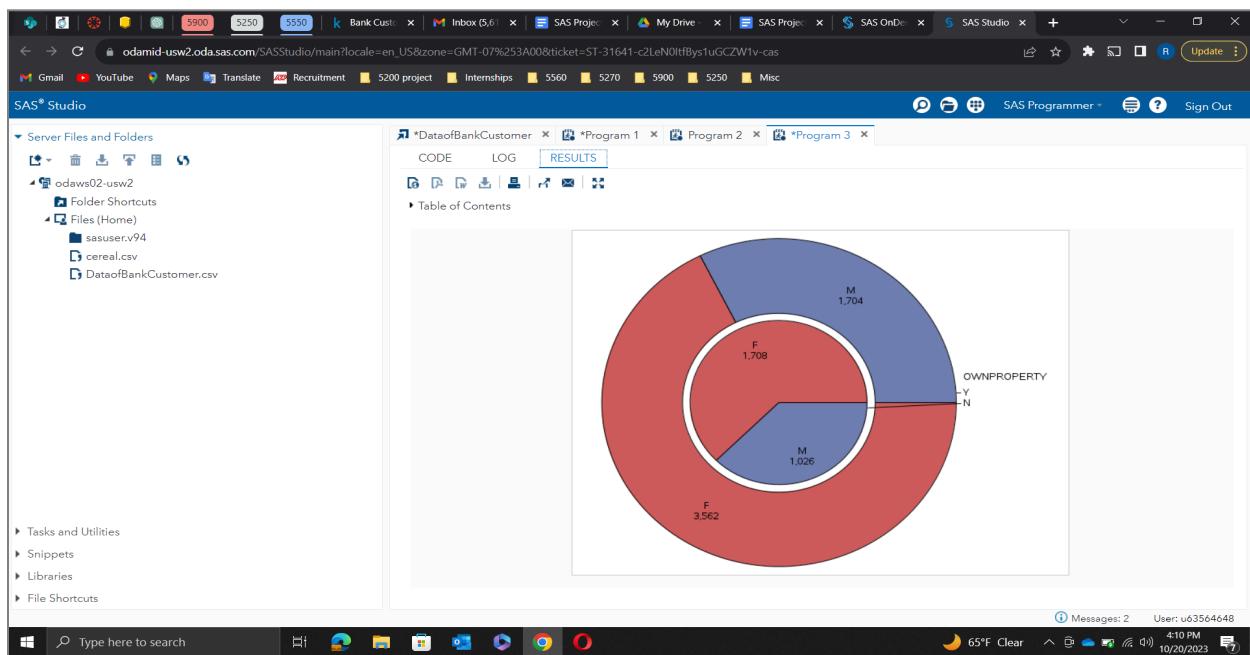
Reference -

Article title: How banks can offer tailored support to different consumer groups.

URL: <https://tink.com/blog/open-banking/tailored-financial-support-consumer-groups/>

Date published: March 20, 2022.

2. Question of Analysis - How can we tailor property-related financial services to the distinct property ownership patterns between male and female customers?



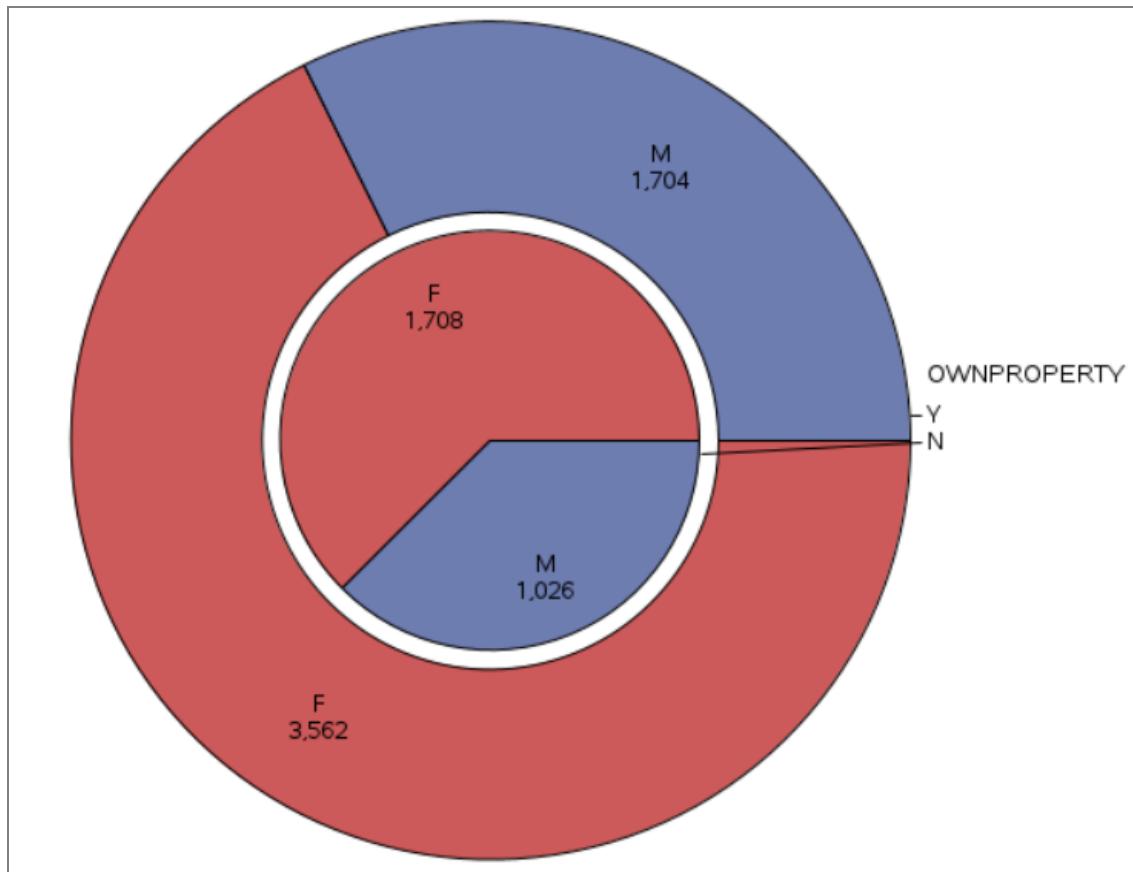


Figure 2. Percentage of customers who own property by gender.

The pie chart in Figure 2 shows the percentage of customers who own a property by gender. The pie chart indicates over 67% of female customers and 32% of male customers who own a property. Opposingly, over 62% of female customers and 37% of male customers do not own a property. This can be interpreted as although the majority of female customers own a property, the majority of customers who do not own a single property are also female. A basic reason the gender gap in homeownership among single Americans has narrowed in recent years is that single women no longer so heavily outnumber single men among older household heads.

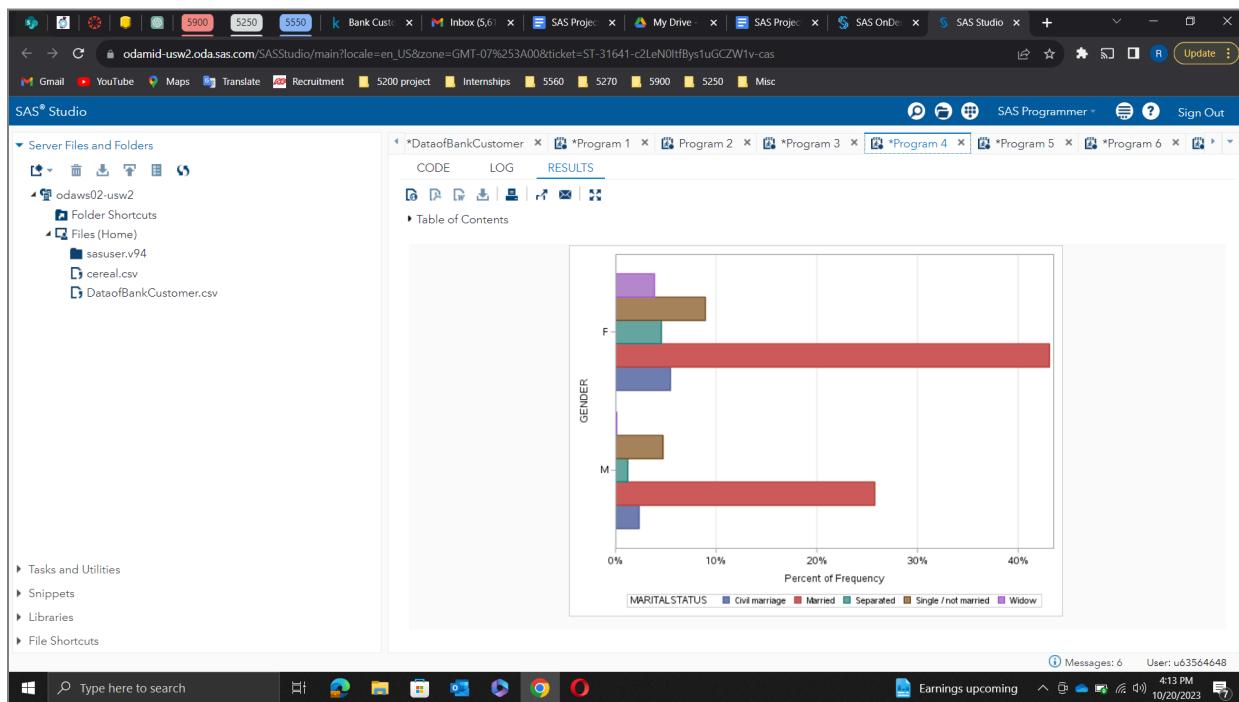
Reference -

Article title: Single women own more homes than single men in the U.S., but that edge is narrowing

URL:<https://www.pewresearch.org/short-reads/2023/06/12/single-women-own-more-homes-than-single-men-in-the-us-but-that-edge-is-narrowing/>

Date published: June 12, 2023

3. Question of Analysis - How can we better align banking services with the varying marital status of male and female customers, especially in addressing the unique financial needs of widows?



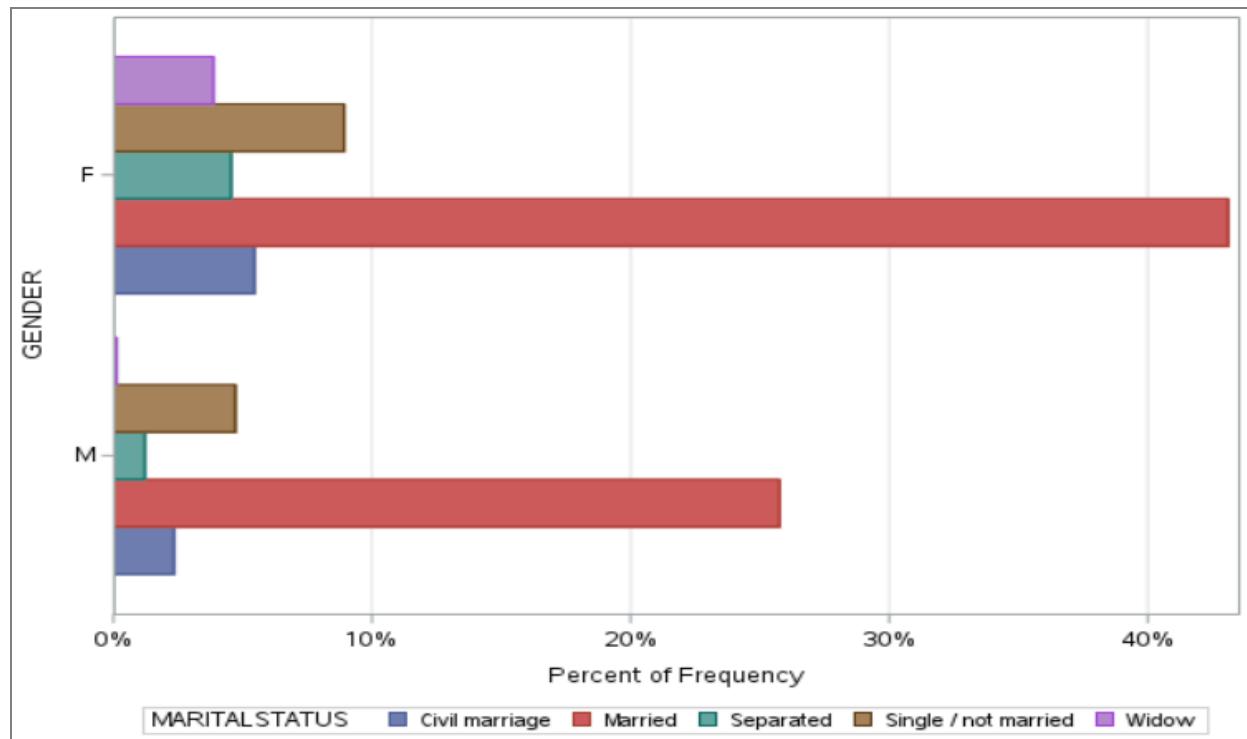


Figure 3. Percentage of customer's marital status by gender.

The horizontal bar chart in Figure 3 indicates the percentage of customer's marital status by gender. The chart shows that over 40% of female customers and 25% of male customers are married. On the other hand, about 5% of females are widows, while less than 1% of males are widows. Today, women control a third of total US household financial assets—more than \$10 trillion (Exhibit 1). But over the next decade, large sums of money are expected to change hands. The biggest driver of this shift is demographics. Today, roughly 70 percent of US affluent-household investable assets are controlled by baby boomers.

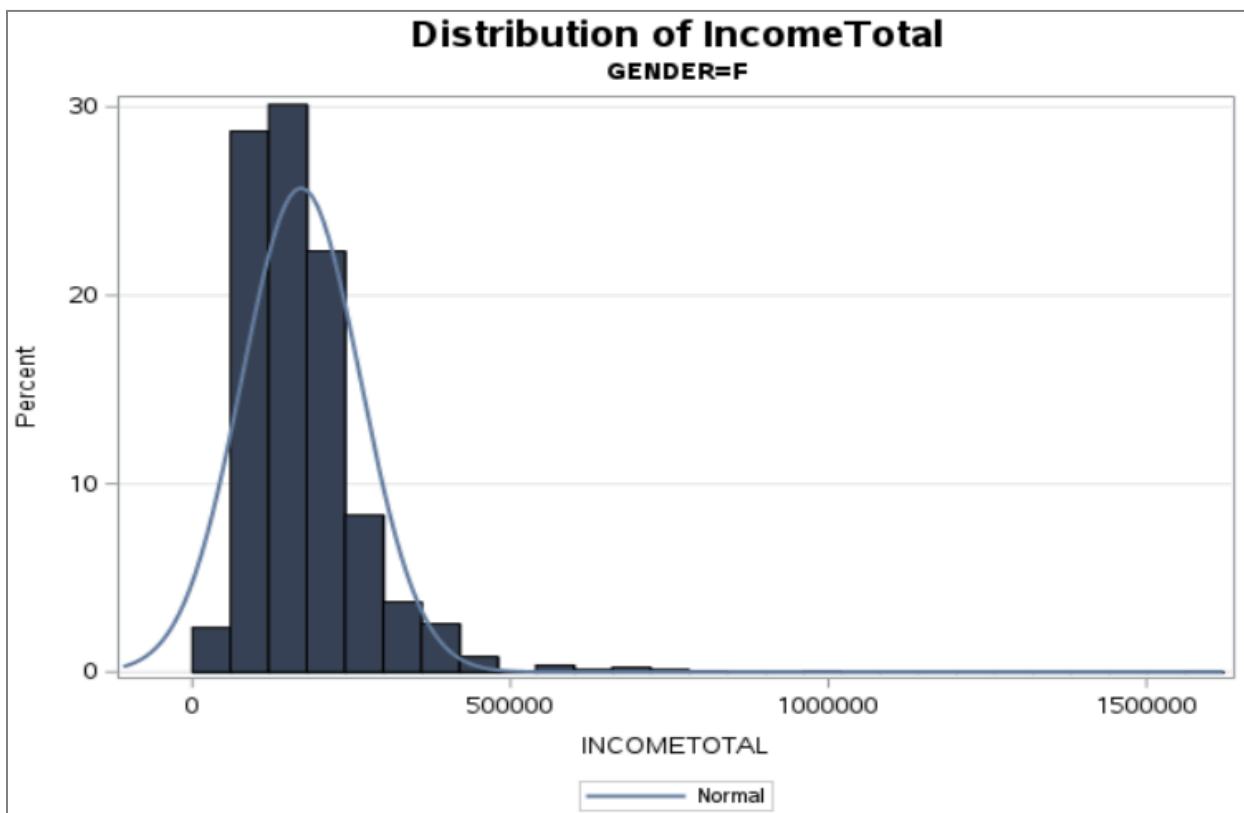
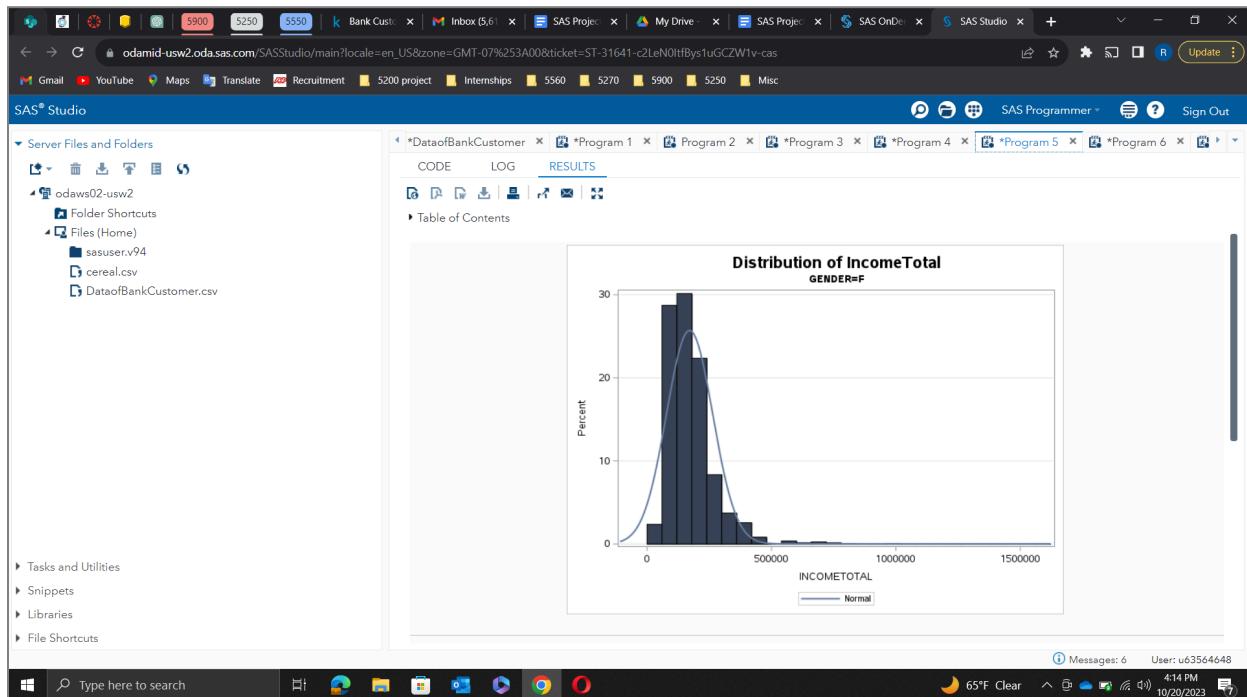
Reference -

Article title: Women as the Next Wave of Growth in US Wealth Management.

URL: <https://www.mckinsey.com/industries/financial-services/our-insights/women-as-the-next-wave-of-growth-in-us-wealth-management>

Date published: July 29, 2020.

4. Question of Analysis - How can the bank develop income-tailored financial services to address the earnings gap between male and female customers?



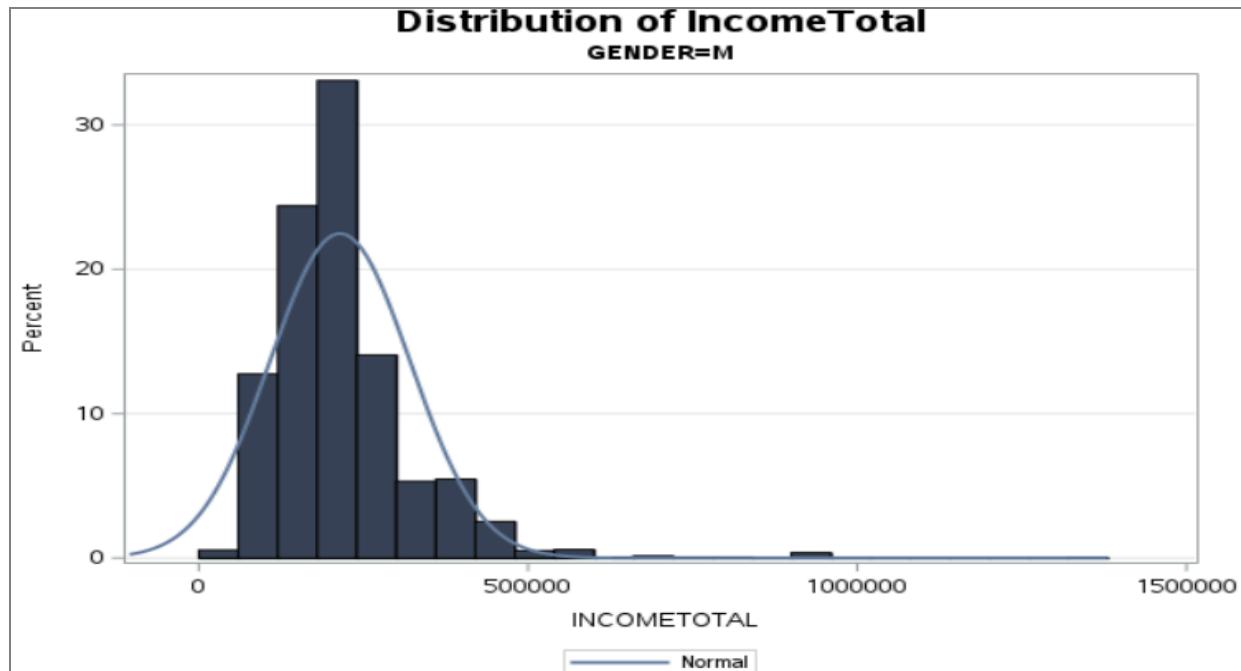


Figure 4. Distribution of annual income by gender.

Figure 4 displays two histograms showing the customer's annual income distribution by gender. It shows that over 30% of males and 20% of females have an annual income of 250000, which indicates that most of the males earned more at its peak compared to females. However, there are 3% of females that earn more during the first quartile of annual income compared to only less than 1% of males. There simply isn't a "one-size-fits-all" path to achieve greater gender equity. However, organizations that prioritize and advance gender equity internally can position themselves to influence public policy, shape or reshape cultural norms, and differentiate themselves across their spheres of influence the marketplace, and society as a whole.

Reference -

Article title: Advancing More Women Leaders in financial services: A Global Report

URL: <https://www2.deloitte.com/xe/en/insights/industry/financial-services/gender-diversity-in-global-financial-services.html>

Date published: 16 June 2022.

5. Question of Analysis - What are the income distribution characteristics and outlier patterns among different housing types for male and female customers?

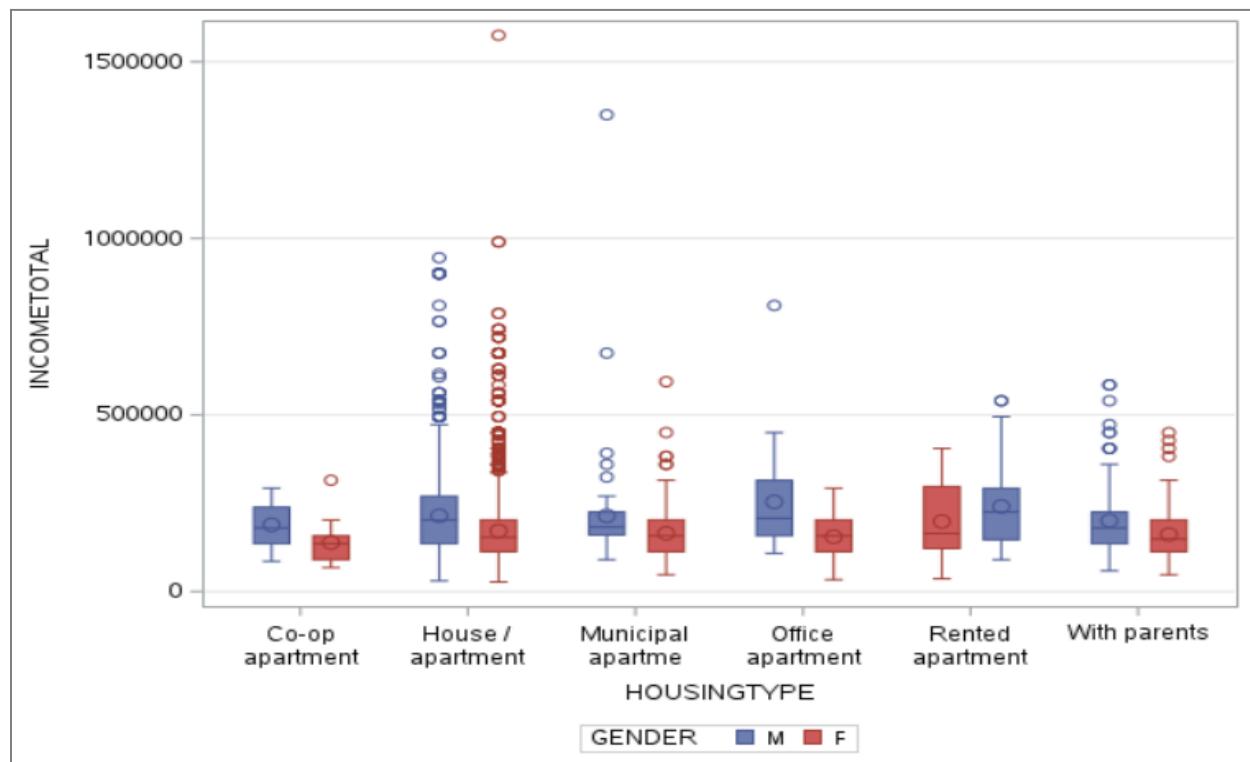
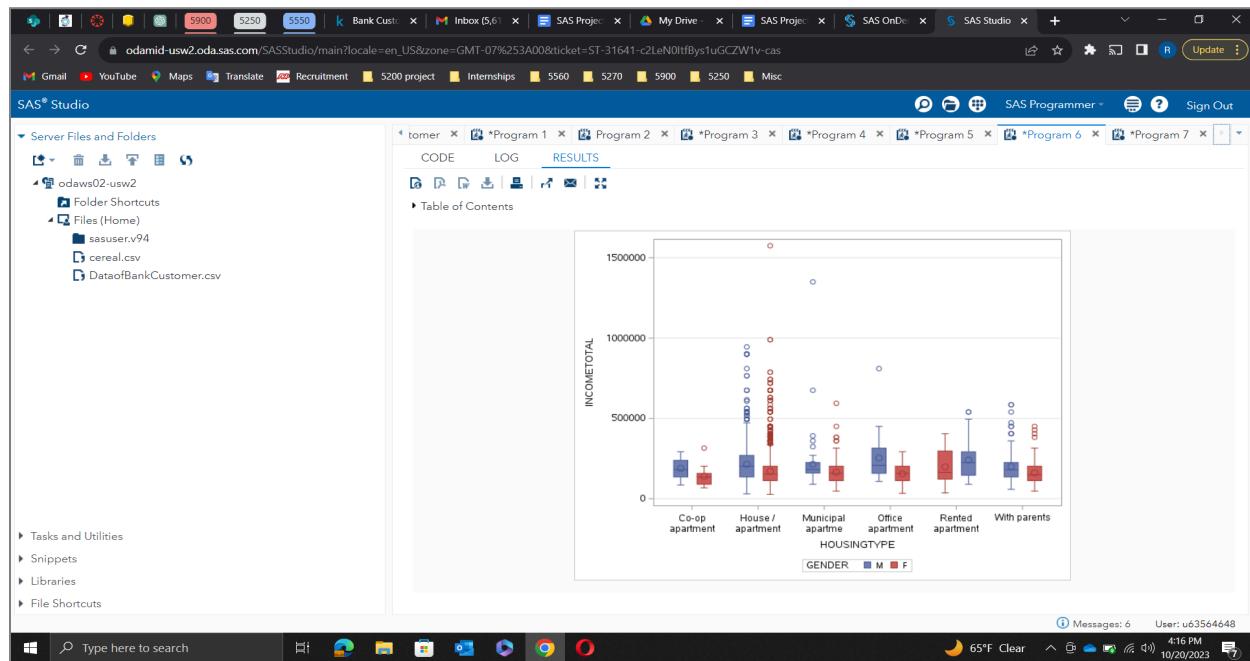


Figure 5. Frequency of customer's annual income by ways of living ordered by gender.

The box plot in Figure 5 shows the frequency of customer's annual income by ways of living ordered separated by gender. Firstly, the plot shows that the median annual income for all ways of living seems to be close to each other. Also, the plot shows that customers living in a house or apartment consist of the most outliers that are a few times more than the upper quartile or maximum value of annual income. Furthermore, customers living in a municipal apartment and with their parents are also shown to have few outliers a few times more than the upper quartile or maximum value of annual income. Lastly, both male and female customers living in a rented apartment show to have the highest maximum annual income across all ways of living. we observe that the median annual incomes are quite similar across various ways of living for both male and female customers. However, strikingly, individuals living in houses or apartments, as well as those in municipal apartments or with their parents, exhibit a significant number of outliers, surpassing the upper quartile or maximum income values by a notable margin. This suggests notable income disparities in these living situations. This analysis unveils intriguing insights into income distribution among different ways of living, highlighting income disparities and outlier patterns within these groups.

Reference -

Article title: Cost of Living Variation, Nonmetropolitan America, and Implications for the Supplemental Poverty Measure.

URL:https://www.researchgate.net/publication/358546641_Cost_of_Living_Variation_Nonmetropolitan_America_and_Implications_for_the_Supplemental_Poverty_Measure

Date published: August 2022.

6. Question of Analysis - How does family size relate to credit loan status and income category among customers?

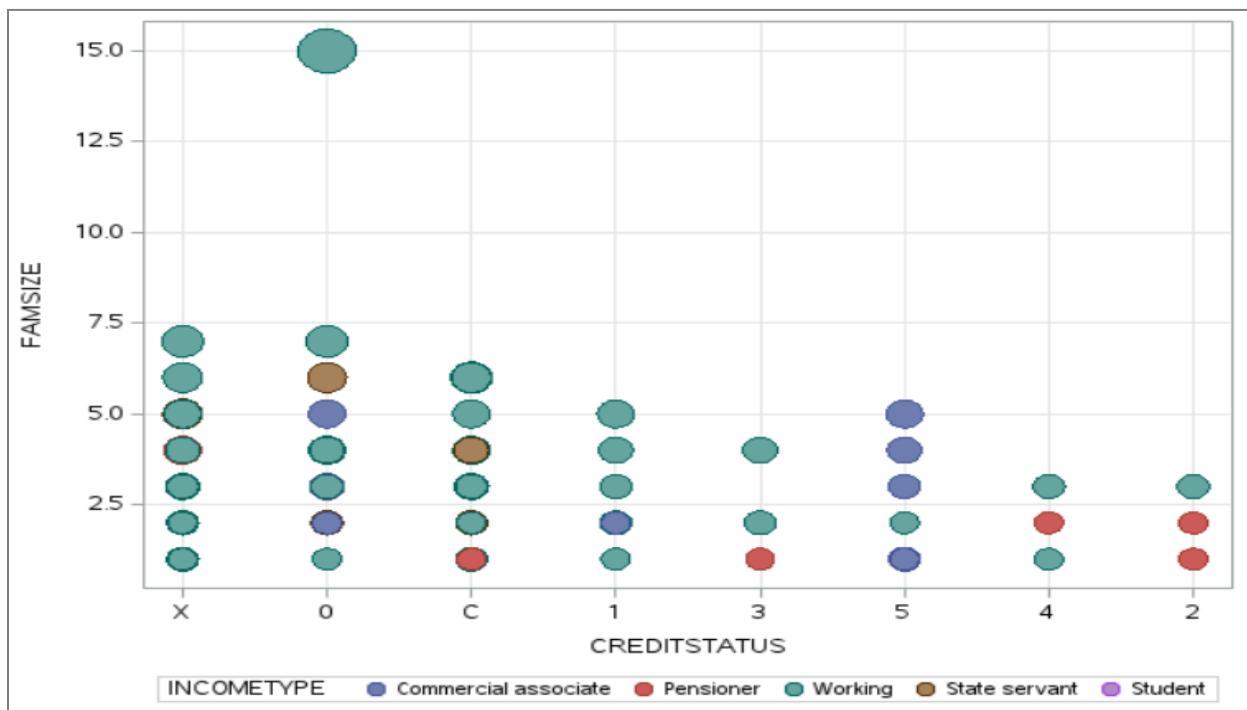
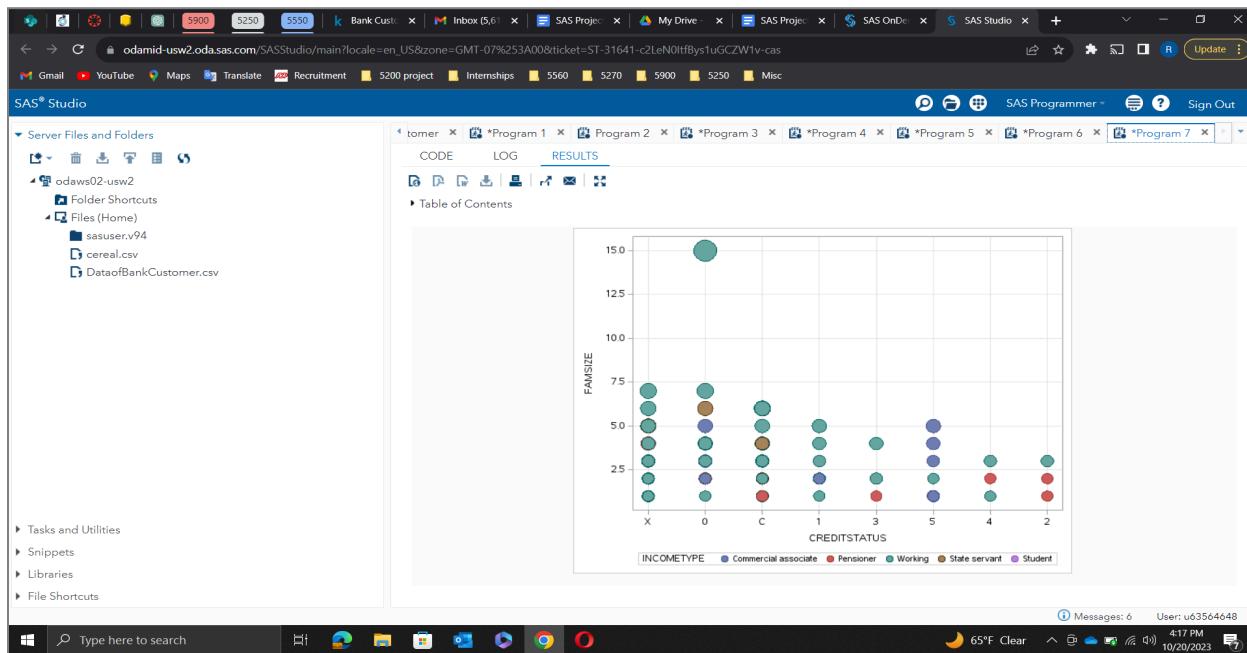


Figure 6. Frequency of customer's family size by credit loan status controlling for income category and number of children.

The bubble plot in Figure 6 shows the frequency of customer's family size by credit loan status, where the bubble size represents the number of children and the bubble color for the income category. Firstly, most customers with no loan for the month (X) for their credit loan status seem to be working-class people with a maximum family size of 7. Next, most of the customers with bad debts or write-offs (5) earn their income as commercial associates with a maximum family size of 5. Lastly, the plot shows that as the number of family sizes increases, the number of children (bubble size) also increases.

Reference -

Article title: Credit and the Family: The Economic Consequences of Closing the Credit Gap of US Couples

URL:<https://www.jpmorganchase.com/institute/research/household-income-spending/credit-and-the-family-truth-in-lending-act>

Date published: July 2022

Part F: Statistical summary

The screenshot shows the SAS Studio interface. The left sidebar contains a tree view of 'Server Files and Folders' and 'Tasks and Utilities', with 'List Data' currently selected. The main pane displays the results of a statistical analysis. The 'RESULTS' tab is active, showing a table of contents and a detailed table of summary statistics for three variables:

Variable	N	N Miss	Minimum	Maximum	Mode	Median	Mean	Std Dev
CHILDRENCOUNT	8800	0	0	14,000000	0	0	0	0,7539289
INCOMETOTAL	8800	0	27000,00	157500,00	136000,00	157500,00	155721,85	100015,4
FAMSIZE	8800	0	1,000000	15,000000	2,000000	2,000000	2,1983750	0,9200020

Variable	N	N Miss	Minimum	Maximum	Mode	Median	Mean	Std Dev
CHILDRENCOUNT	8000	0	0	14.0000000	0	0	0.4215000	0.7539289
INOMETOTAL	8000	0	27000.00	1575000.00	135000.00	157500.00	185712.85	100015.41
FAMSIZE	8000	0	1.0000000	15.0000000	2.0000000	2.0000000	2.1883750	0.9200520

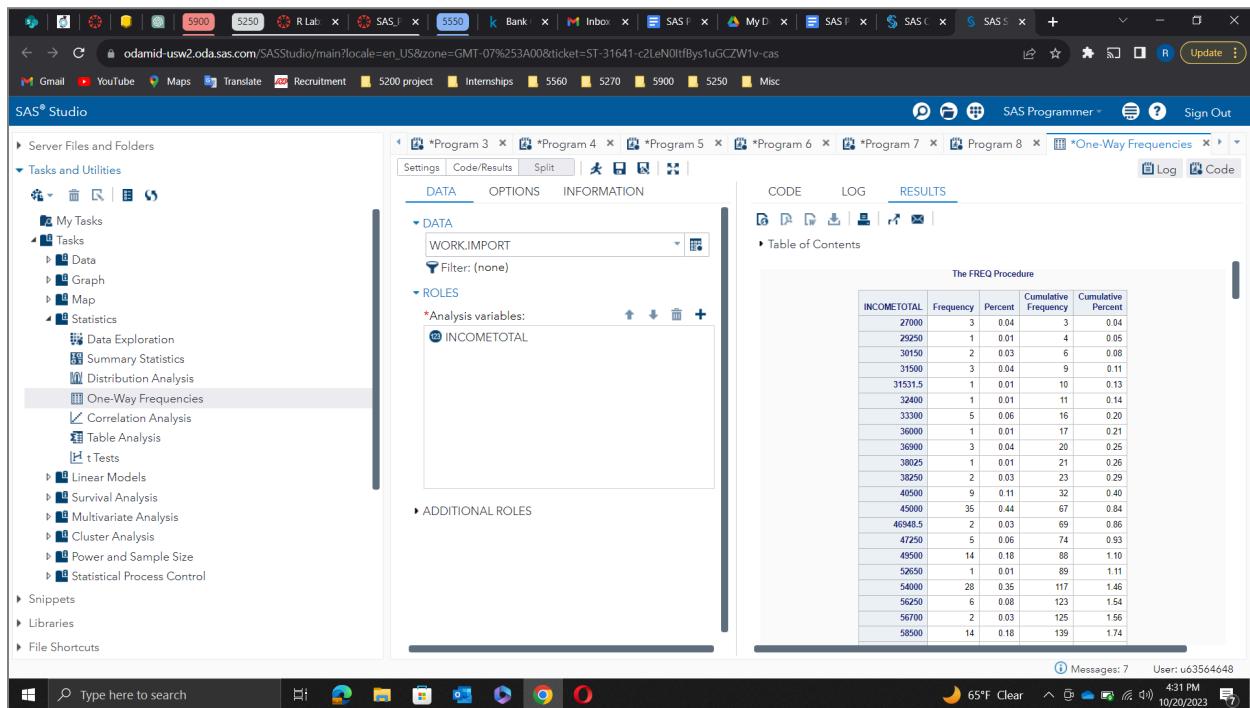
Figure - Descriptive Statistics for assignment data.

Figure shows descriptive statistics for three numeric variables in the assignment data. The descriptive statistics include the number of observations (N), number of missing values (N Miss), minimum, maximum, mode, median, mean, and standard deviation values of the variables. Moreover, the binary variables; EMAIL, MOBILE, GENDER, OWNCAR, and OWNPROPERTY are ignored in this table because their central tendency doesn't mean anything because all their values just have '0', '1', 'N', and 'Y' in each of the variables. Likewise, the categorical variables; EDUCATIONLEVEL, MARITALSTATUS, HOUSINGTYPE, INCOMETYPE, and CREDITSTATUS are also excluded from the Figure because the values are all character or object format with no numeric data.

- For CHILDRENCOUNT, it appears that most customers have no children (mode and median of 0), and the distribution is positively skewed.
- For INOMETOTAL, the income distribution is right-skewed, with a higher mean income compared to the median, indicating that a relatively small number of customers have very high incomes.
- For FAMSIZE, a family size of 2 is most common, and the distribution is slightly right-skewed.

Part G: Statistical tests

One-way frequency:



The screenshot shows the SAS Studio interface with the 'One-Way Frequencies' analysis open. The left sidebar shows various tasks and utilities, including 'One-Way Frequencies' which is currently selected. The main workspace displays the 'DATA' and 'RESULTS' tabs. The 'RESULTS' tab shows the output of the 'The FREQ Procedure' for the 'INCOMETOTAL' column. The output table includes columns for INCOMETOTAL, Frequency, Percent, Cumulative Frequency, and Cumulative Percent. The data is as follows:

INCOMETOTAL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
27000	3	0.04	3	0.04
29250	1	0.01	4	0.05
30150	2	0.03	6	0.08
31500	3	0.04	9	0.11
31531.5	1	0.01	10	0.13
32400	1	0.01	11	0.14
33300	5	0.06	16	0.20
36000	1	0.01	17	0.21
36900	3	0.04	20	0.25
38025	1	0.01	21	0.26
38250	2	0.03	23	0.29
40500	9	0.11	32	0.40
45000	35	0.44	67	0.84
46948.5	2	0.03	69	0.86
47250	5	0.06	74	0.93
49500	14	0.18	88	1.10
52650	1	0.01	89	1.11
54000	28	0.35	117	1.46
56250	6	0.08	123	1.54
56700	2	0.03	125	1.56
58500	14	0.18	139	1.74

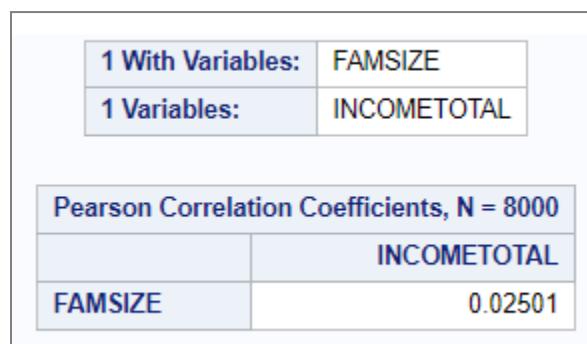
The FREQ Procedure				
INCOMETOTAL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
27000	3	0.04	3	0.04
29250	1	0.01	4	0.05
30150	2	0.03	6	0.08
31500	3	0.04	9	0.11
31531.5	1	0.01	10	0.13
32400	1	0.01	11	0.14
33300	5	0.06	16	0.20
36000	1	0.01	17	0.21
36900	3	0.04	20	0.25
38025	1	0.01	21	0.26
38250	2	0.03	23	0.29
40500	9	0.11	32	0.40
45000	35	0.44	67	0.84
46948.5	2	0.03	69	0.86
47250	5	0.06	74	0.93

Figure - One-way frequency for Incometotal column.

One-way frequency analysis, also known as univariate analysis, is a fundamental technique used in data analysis to explore the distribution of a single categorical or numerical variable. In the context of banking customer data analysis, you can use one-way frequency analysis to examine the distribution of the total income of customers. The insights gained from this one-way frequency analysis can be useful for various purposes, such as marketing strategy, credit card approval, credit card limit, or risk assessment, and loan approval in the banking industry. It helps in identifying the income groups that are most prevalent among the customer base and can inform decision-making processes.

Correlation Analysis:

The screenshot shows the SAS Studio interface with a correlation analysis open. The left sidebar shows various tasks and utilities, with 'Scatter Plot' selected. The main workspace displays the 'RESULTS' tab of the correlation analysis. The 'Analysis variables' section shows 'INCOMETOTAL' selected. The 'Correlate with' section shows 'FAMSIZE' selected. The results table shows the Pearson Correlation Coefficients for N = 8000, with a value of 0.02501 for the correlation between FAMSIZE and INCOMETOTAL. A scatter plot is also visible in the results pane.



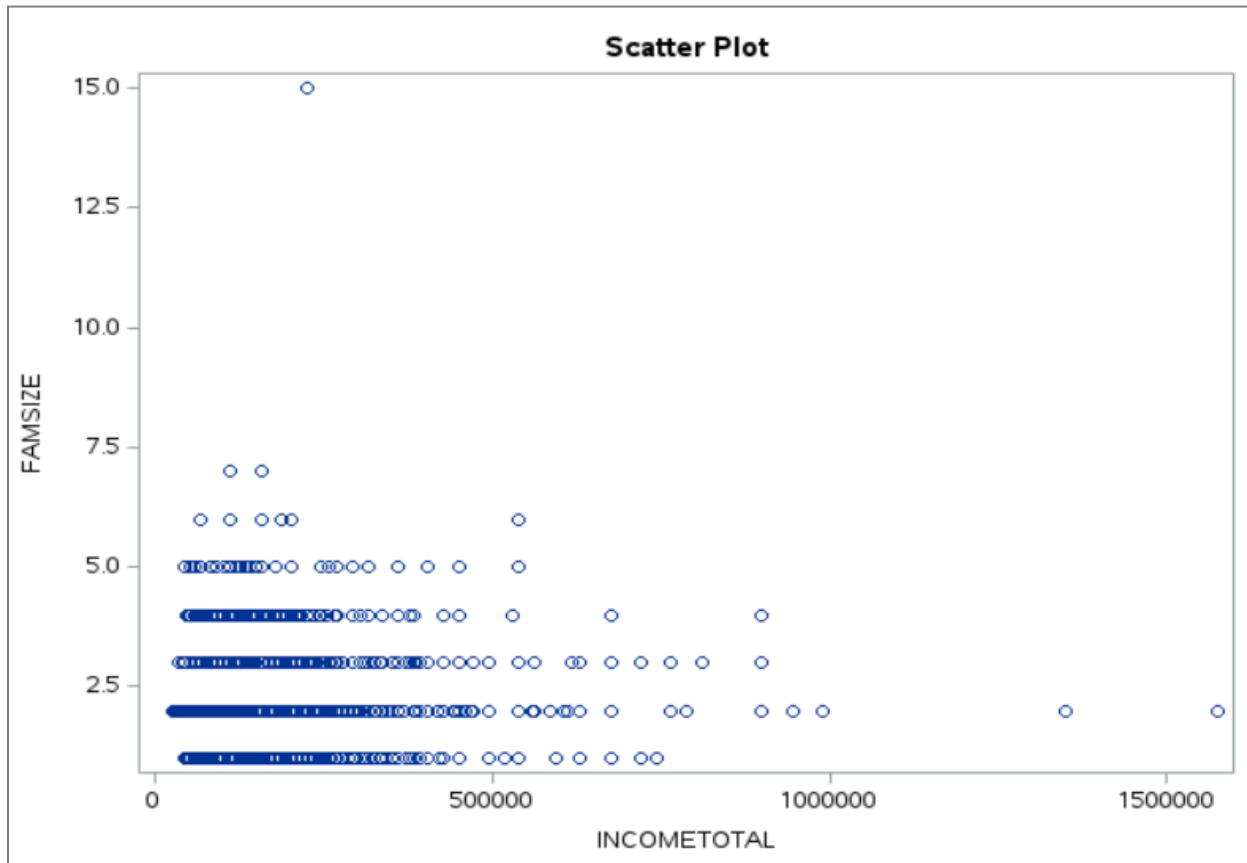


Figure - Correlation for Family size with Income total.

Correlation is a statistical measure that quantifies the strength and direction of the linear relationship between two variables. Here's the correlation value is 0.025. In summary, a correlation value of 0.025 means that there is a very weak positive linear relationship between the two variables. While it may be statistically significant, it is unlikely to be practically meaningful for making predictions or drawing strong conclusions about the relationship between the variables.

T-Test:

The screenshot shows the SAS Studio interface with multiple windows open. The main window displays the T-Test results for the variable FAMSIZE. The left sidebar shows the navigation menu with the 't Tests' option selected under 'Linear Models'. The T-Test window has tabs for 'DATA', 'OPTIONS', 'INFORMATION', 'CODE', 'LOG', and 'RESULTS'. The 'RESULTS' tab is active, showing the 'Tests for Normality' table and the 'Distribution of FAMSIZE' table. The 'Tests for Normality' table includes results for Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests. The 'Distribution of FAMSIZE' table includes summary statistics like Mean, Std Dev, and 95% Confidence Interval.

Variable: FAMSIZE					
Tests for Normality					
Test	Statistic		p Value		
Kolmogorov-Smirnov	D	0.314364	Pr > D	<0.0100	
Cramer-von Mises	W-Sq	121.9573	Pr > W-Sq	<0.0050	
Anderson-Darling	A-Sq	584.5417	Pr > A-Sq	<0.0050	

Variable: FAMSIZE					
N	Mean	Std Dev	Std Err	Minimum	Maximum
8000	2.1884	0.9201	0.0103	1.0000	15.0000

Mean	95% CL Mean	Std Dev	95% CL Std Dev	
2.1884	2.1682	2.2085	0.9060	0.9345

DF	t Value	Pr > t
7999	212.74	<0.0001

Figure - T-Test for Family size variable.

Mean of Family Size (2.18): The mean represents the average family size in the banking dataset.

In this case, the calculated mean is 2.18. Value of the Single-Tailed Test (t-value: 212.7): The t-value is a statistic that results from conducting a t-test. In hypothesis testing, a one-tailed t-test is used to compare the sample mean to a reference value. A high t-value, such as 212.7, indicates a substantial difference between the sample mean and the T-test value. The test has detected a statistically significant difference, which means that the family size is significantly greater (or lesser, depending on the direction of your test) than what might be expected. Hence there is a significant difference with respect to the mean, therefore we reject the hypothesis.

CONCLUSION:

In conclusion, it is determined that different factors are related to the customer's annual income (INCOMETOTAL) shows there is enough evidence to conclude that the distribution of change in annual income (INCOMETOTAL) for different education levels (EDUCATIONLEVEL) is significantly different while ways of living (HOUSINGTYPE) are not significantly different. Bankers should make an effort to take initiative at the education level (EDUCATIONLEVEL) instead of focusing on the ways of living (HOUSINGTYPE).