MARKETING ANALYTICS (MANM533)

Word count: 1976

Student Name: Eldhose Varghese

Student URN: 6793260

Table of Contents

List of Tables

List of Figures

1. Data preparation and pre-processing

1.1 Introduction

1.1.1 Aim

This project aims to develop a predictive model determining the likelihood of a customer responding to an international bank's marketing campaign for promoting a fixed-term savings account.

1.1.2 Data

Dataset 1 consists of campaign contact details (contact type, contact duration) and the corresponding response outcome for each client, whereas Dataset 2 consists of personalized attributes of the customers (age, region, education, balance, etc). The combined dataset comprises 33909 data points, each containing 13 unique attributes.

1.2 Pre-processing

Table 1 – Univariate Statistics for Dataset 1

Univariate Statistics

	N	Mean	Std. Deviation	Missing		No. of Extremes[a]	
				Count	Percent	Low	High
custID	33909	22667.10	13040.886	0	.0	0	0
duration	33909	257.61	256.435	0	.0	0	1568
contact	33909			0	.0		
response	33909			0	.0		

a. Number of cases outside the range (Mean - 2*SD, Mean + 2*SD).

All variables in the dataset are complete, with no missing entries. The 'duration' variable presents an average customer contact time of 257.61 seconds. With a substantial standard deviation of 256.435 for 'duration', it suggests there's a wide disparity in how long customers are in contact. Regarding 'duration', the data shows 1,568 instances that exceed the mean by more than double the standard deviation, pointing to a considerable number of unusually lengthy customer contacts, while there are no notably short durations

Table 2 - Univariate Statistics for Dataset 2

Univariate Statistics

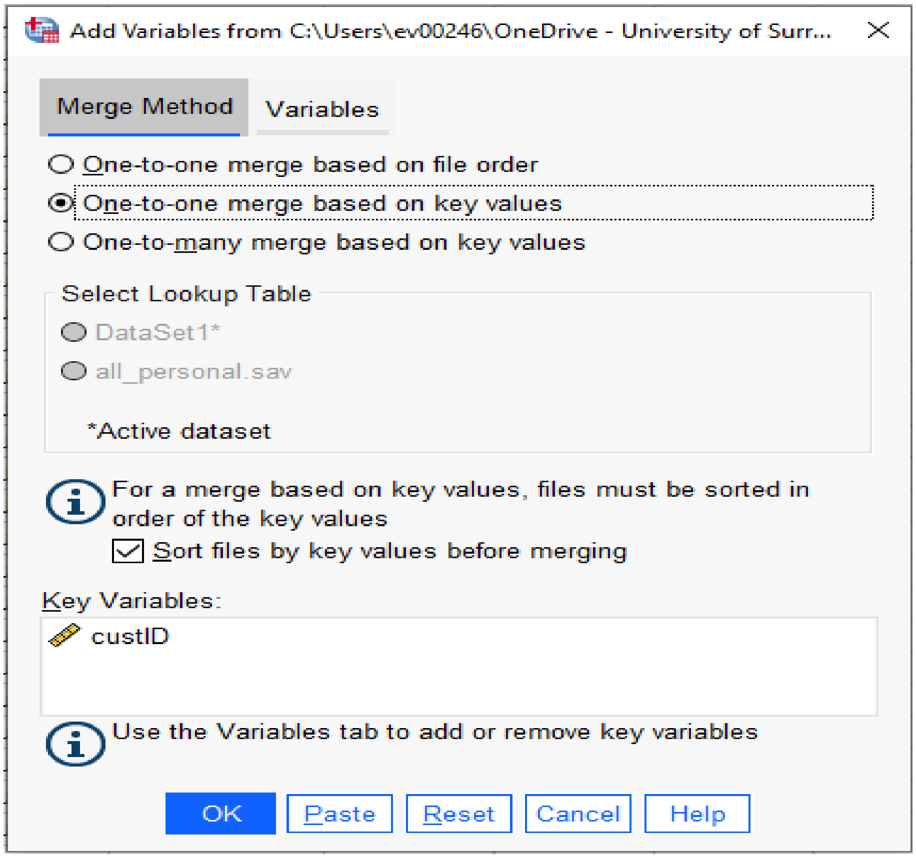
	N	Mean	Std. Deviation	Missing		No. of Extremes[a]	
				Count	Percent	Low	High
custID	33909	22667.10	13040.886	0	.0	0	0
age	33909	40.97	10.628	0	.0	32	720
balance	33909	1569.57	3420.725	0	.0	1	1159
region	33909			0	.0		
job	33909			0	.0		
marital	33909			0	.0		
education	33909			0	.0		
default	33909			0	.0		
housing	33909			0	.0		
loan	33909			0	.0		

a. Number of cases outside the range (Mean - 2*SD, Mean + 2*SD).

In Dataset 2, each variable is complete without any missing entries. The ‘balance’ variable averages at 1,569.57 pounds, indicating the typical account balance for a customer within this data. The substantial standard deviation of 3,420.725 pounds for 'balance' points to a significant diversity in the account balances among customers. For the ‘age’ variable, there are 32 instances lower and 720 instances higher than the usual age range, based on the mean adjusted by two standard deviations. Similarly, for ‘balance’, the data shows one extremely low account balance and 1,159 unusually high balances, highlighting notable deviations from the average balance.

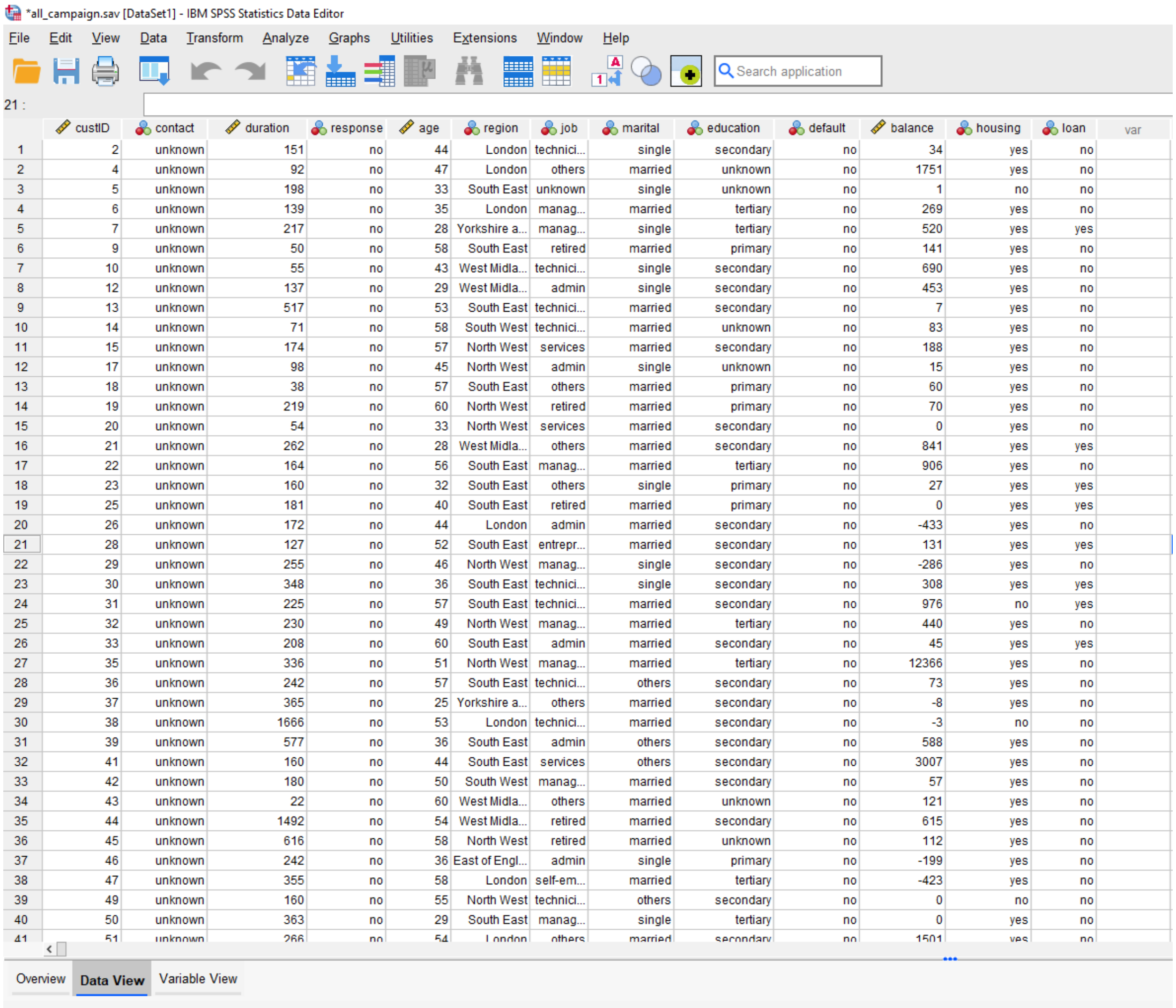
1.3 Merging the two datasets

Figure 1 – Merging Datasets



The two datasets 'all_campaign' and 'all_personal' were merged based on the common column 'custID' as shown above.

Figure 2 – Merged dataset



	custID	contact	duration	response	age	region	job	marital	education	default	balance	housing	loan	var
1	2	unknown	151	no	44	London	technici...	single	secondary	no	34	yes	no	
2	4	unknown	92	no	47	London	others	married	unknown	no	1751	yes	no	
3	5	unknown	198	no	33	South East	unknown	single	unknown	no	1	no	no	
4	6	unknown	139	no	35	London	manag...	married	tertiary	no	269	yes	no	
5	7	unknown	217	no	28	Yorkshire a...	manag...	single	tertiary	no	520	yes	yes	
6	9	unknown	50	no	58	South East	retired	married	primary	no	141	yes	no	
7	10	unknown	55	no	43	West Midla...	technici...	single	secondary	no	690	yes	no	
8	12	unknown	137	no	29	West Midla...	admin	single	secondary	no	453	yes	no	
9	13	unknown	517	no	53	South East	technici...	married	secondary	no	7	yes	no	
10	14	unknown	71	no	58	South West	technici...	married	unknown	no	83	yes	no	
11	15	unknown	174	no	57	North West	services	married	secondary	no	188	yes	no	
12	17	unknown	98	no	45	North West	admin	single	unknown	no	15	yes	no	
13	18	unknown	38	no	57	South East	others	married	primary	no	60	yes	no	
14	19	unknown	219	no	60	North West	retired	married	primary	no	70	yes	no	
15	20	unknown	54	no	33	North West	services	married	secondary	no	0	yes	no	
16	21	unknown	262	no	28	West Midla...	others	married	secondary	no	841	yes	yes	
17	22	unknown	164	no	56	South East	manag...	married	tertiary	no	906	yes	no	
18	23	unknown	160	no	32	South East	others	single	primary	no	27	yes	yes	
19	25	unknown	181	no	40	South East	retired	married	primary	no	0	yes	yes	
20	26	unknown	172	no	44	London	admin	married	secondary	no	-433	yes	no	
21	28	unknown	127	no	52	South East	entrepr...	married	secondary	no	131	yes	yes	
22	29	unknown	255	no	46	North West	manag...	single	secondary	no	-286	yes	no	
23	30	unknown	348	no	36	South East	technici...	single	secondary	no	308	yes	yes	
24	31	unknown	225	no	57	South East	technici...	married	secondary	no	976	no	yes	
25	32	unknown	230	no	49	North West	manag...	married	tertiary	no	440	yes	no	
26	33	unknown	208	no	60	South East	admin	married	secondary	no	45	yes	yes	
27	35	unknown	336	no	51	North West	manag...	married	tertiary	no	12366	yes	no	
28	36	unknown	242	no	57	South East	technici...	others	secondary	no	73	yes	no	
29	37	unknown	365	no	25	Yorkshire a...	others	married	secondary	no	-8	yes	no	
30	38	unknown	1666	no	53	London	technici...	married	secondary	no	-3	no	no	
31	39	unknown	577	no	36	South East	admin	others	secondary	no	588	yes	no	
32	41	unknown	160	no	44	South East	services	others	secondary	no	3007	yes	no	
33	42	unknown	180	no	50	South West	manag...	married	secondary	no	57	yes	no	
34	43	unknown	22	no	60	West Midla...	others	married	unknown	no	121	yes	no	
35	44	unknown	1492	no	54	West Midla...	retired	married	secondary	no	615	yes	no	
36	45	unknown	616	no	58	North West	retired	married	unknown	no	112	yes	no	
37	46	unknown	242	no	36	East of Engl...	admin	single	primary	no	-199	yes	no	
38	47	unknown	355	no	58	London	self-em...	married	tertiary	no	-423	yes	no	
39	49	unknown	160	no	55	North West	technici...	others	secondary	no	0	no	no	
40	50	unknown	363	no	29	South East	manag...	single	tertiary	no	0	yes	no	
41	51	unknown	266	no	54	London	others	married	secondary	no	1501	yes	no	

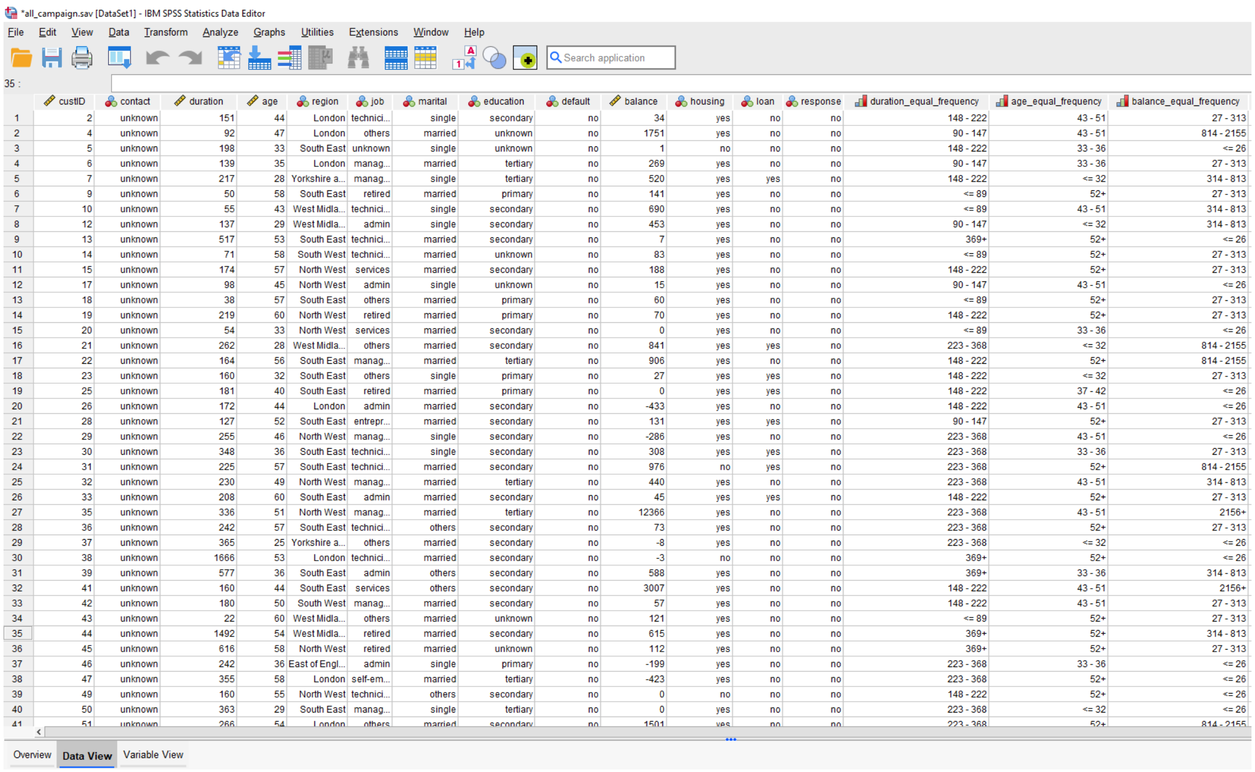
Figure 2 shows the final merged dataset.

1.4 Binning Numeric Variables [Approach 2 - Equal Frequency Binning]

Equal Frequency Binning

Equal frequency binning ensures that each bin has the same number of observations.

Figure 3 – Dataset after Equal Frequency Binning



	custID	contact	duration	age	region	job	marital	education	default	balance	housing	loan	response	duration_equal_frequency	age_equal_frequency	balance_equal_frequency
1	2	unknown	151	44	London	technici...	single	secondary	no	34	yes	no	no	148 - 222	43 - 51	27 - 313
2	4	unknown	92	47	London	others	married	unknown	no	1751	yes	no	no	90 - 147	43 - 51	814 - 2155
3	5	unknown	198	33	South East	unknown	single	unknown	no	1	no	no	no	148 - 222	33 - 36	<= 26
4	6	unknown	139	35	London	manag...	married	tertiary	no	269	yes	no	no	90 - 147	33 - 36	27 - 313
5	7	unknown	217	28	Yorkshire a...	manag...	single	tertiary	no	520	yes	yes	no	148 - 222	<= 32	314 - 813
6	9	unknown	50	58	South East	retired	married	primary	no	141	yes	no	no	<= 89	52+	27 - 313
7	10	unknown	55	43	West Midla...	technici...	single	secondary	no	690	yes	no	no	<= 89	43 - 51	314 - 813
8	12	unknown	137	29	West Midla...	admin	single	secondary	no	453	yes	no	no	90 - 147	<= 32	314 - 813
9	13	unknown	517	53	South East	technici...	married	secondary	no	7	yes	no	no	369+	52+	<= 26
10	14	unknown	71	58	South West	technici...	married	unknown	no	83	yes	no	no	<= 89	52+	27 - 313
11	15	unknown	174	57	North West	services	married	secondary	no	188	yes	no	no	148 - 222	52+	27 - 313
12	17	unknown	98	45	North West	admin	single	unknown	no	15	yes	no	no	90 - 147	43 - 51	<= 26
13	18	unknown	38	57	South East	others	married	primary	no	60	yes	no	no	<= 89	52+	27 - 313
14	19	unknown	219	60	North West	retired	married	primary	no	70	yes	no	no	148 - 222	52+	27 - 313
15	20	unknown	54	33	North West	services	married	secondary	no	0	yes	no	no	<= 89	33 - 36	<= 26
16	21	unknown	262	28	West Midla...	others	married	secondary	no	841	yes	yes	no	223 - 368	<= 32	814 - 2155
17	22	unknown	164	56	South East	manag...	married	tertiary	no	906	yes	no	no	148 - 222	52+	814 - 2155
18	23	unknown	160	32	South East	others	single	primary	no	27	yes	yes	no	148 - 222	<= 32	27 - 313
19	25	unknown	181	40	South East	retired	married	primary	no	0	yes	yes	no	148 - 222	37 - 42	<= 26
20	26	unknown	172	44	London	admin	married	secondary	no	-433	yes	no	no	148 - 222	43 - 51	<= 26
21	28	unknown	127	52	South East	entrepr...	married	secondary	no	131	yes	yes	no	90 - 147	52+	27 - 313
22	29	unknown	255	46	North West	manag...	single	secondary	no	-286	yes	no	no	223 - 368	43 - 51	<= 26
23	30	unknown	348	36	South East	technici...	single	secondary	no	308	yes	yes	no	223 - 368	33 - 36	27 - 313
24	31	unknown	225	57	South East	technici...	married	secondary	no	976	no	yes	no	223 - 368	52+	814 - 2155
25	32	unknown	230	49	North West	manag...	married	tertiary	no	440	yes	no	no	223 - 368	43 - 51	314 - 813
26	33	unknown	208	60	South East	admin	married	secondary	no	45	yes	yes	no	148 - 222	52+	27 - 313
27	35	unknown	336	51	North West	manag...	married	tertiary	no	12366	yes	no	no	223 - 368	43 - 51	2156+
28	36	unknown	242	57	South East	technici...	others	secondary	no	73	yes	no	no	223 - 368	52+	27 - 313
29	37	unknown	365	25	Yorkshire a...	others	married	secondary	no	-8	yes	no	no	223 - 368	<= 32	<= 26
30	38	unknown	1666	53	London	technici...	married	secondary	no	-3	no	no	no	369+	52+	<= 26
31	39	unknown	577	36	South East	admin	others	secondary	no	588	yes	no	no	369+	33 - 36	314 - 813
32	41	unknown	160	44	South East	services	others	secondary	no	3007	yes	no	no	148 - 222	43 - 51	2156+
33	42	unknown	180	50	South West	manag...	married	secondary	no	57	yes	no	no	148 - 222	43 - 51	27 - 313
34	43	unknown	22	60	West Midla...	others	married	unknown	no	121	yes	no	no	<= 89	52+	27 - 313
35	44	unknown	1492	54	West Midla...	retired	married	secondary	no	615	yes	no	no	369+	52+	314 - 813
36	45	unknown	616	58	North West	retired	married	unknown	no	112	yes	no	no	369+	52+	27 - 313
37	46	unknown	242	36	East of Engl...	admin	single	primary	no	-199	yes	no	no	223 - 368	33 - 36	<= 26
38	47	unknown	355	58	London	self-em...	married	tertiary	no	-423	yes	no	no	223 - 368	52+	<= 26
39	49	unknown	160	55	North West	technici...	others	secondary	no	0	no	no	no	148 - 222	52+	<= 26
40	50	unknown	363	29	South East	manag...	single	tertiary	no	0	yes	no	no	223 - 368	<= 32	<= 26
41	51	unknown	266	54	London	others	married	secondary	no	1501	yes	no	no	223 - 368	52+	814 - 2155

Approach 2 - numeric variables 'duration', 'age', and 'balance' are binned based on the equal frequency binning method.

Binning details

Duration – [bins = 5, width % = 20]

Age – [bins = 5, width % = 20]

Balance – [bins = 5, width % = 20]

Figure 4 – Distribution of duration after Equal Frequency Binning

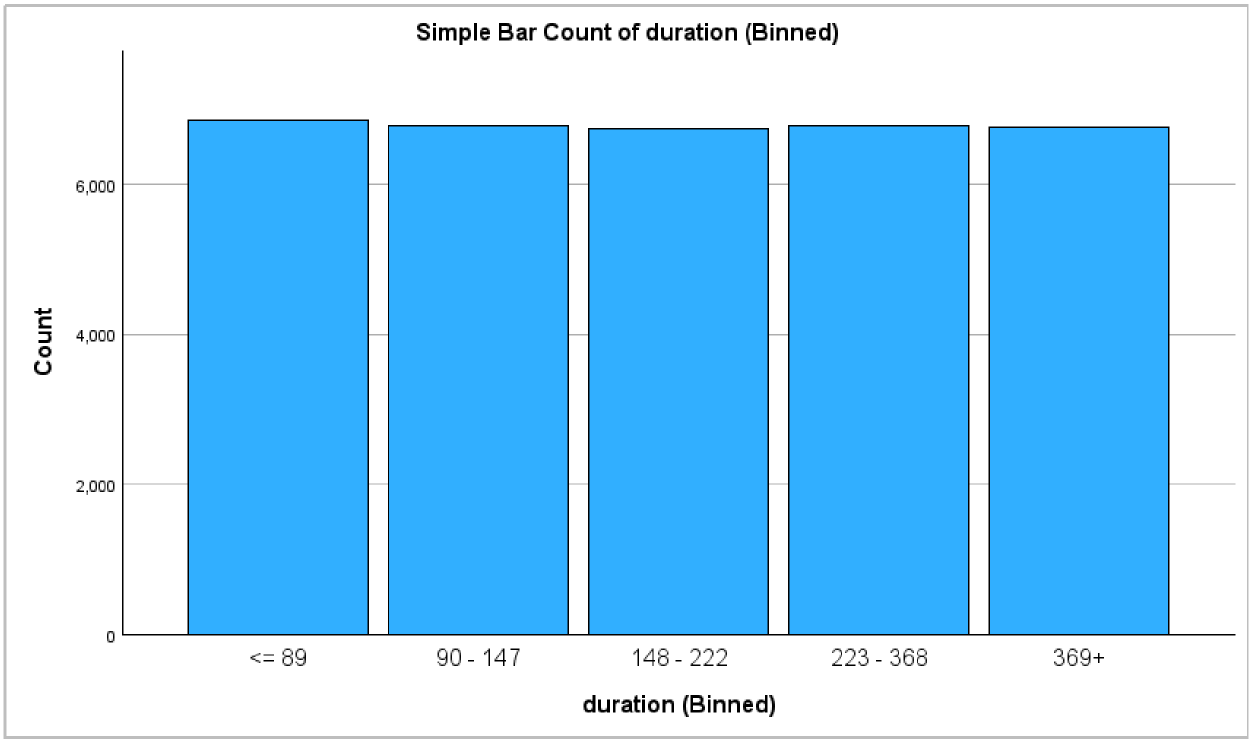


Figure 5 – Distribution of age after Equal Frequency Binning

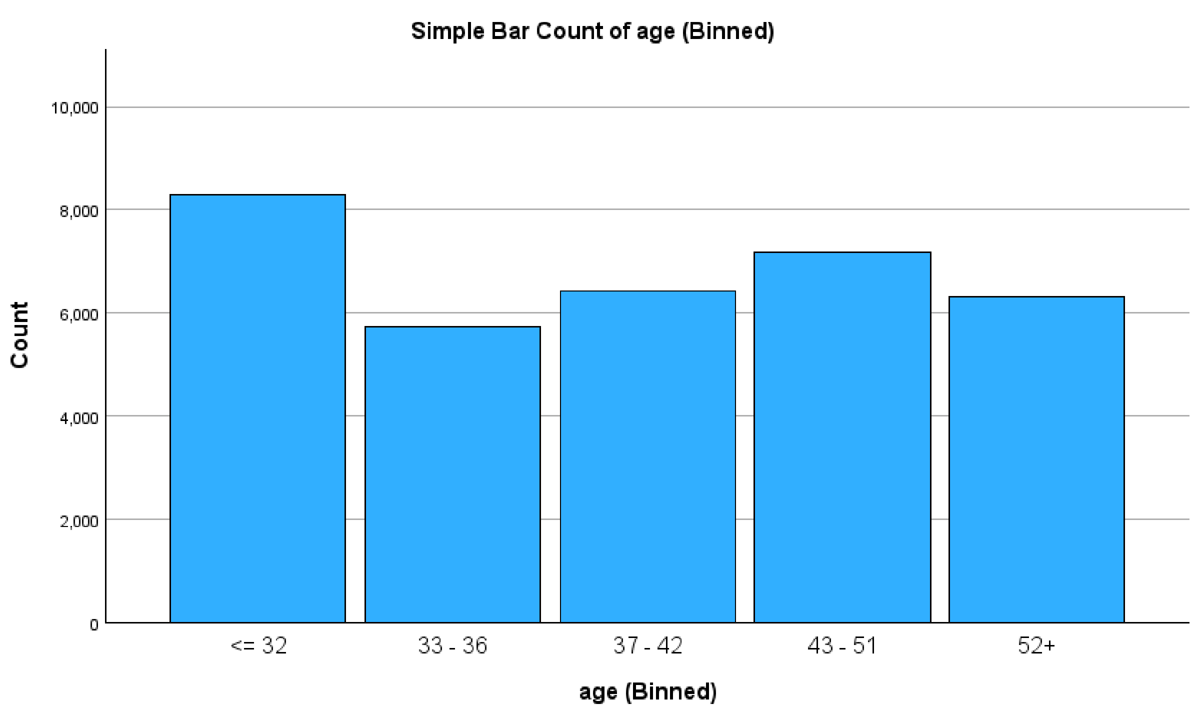
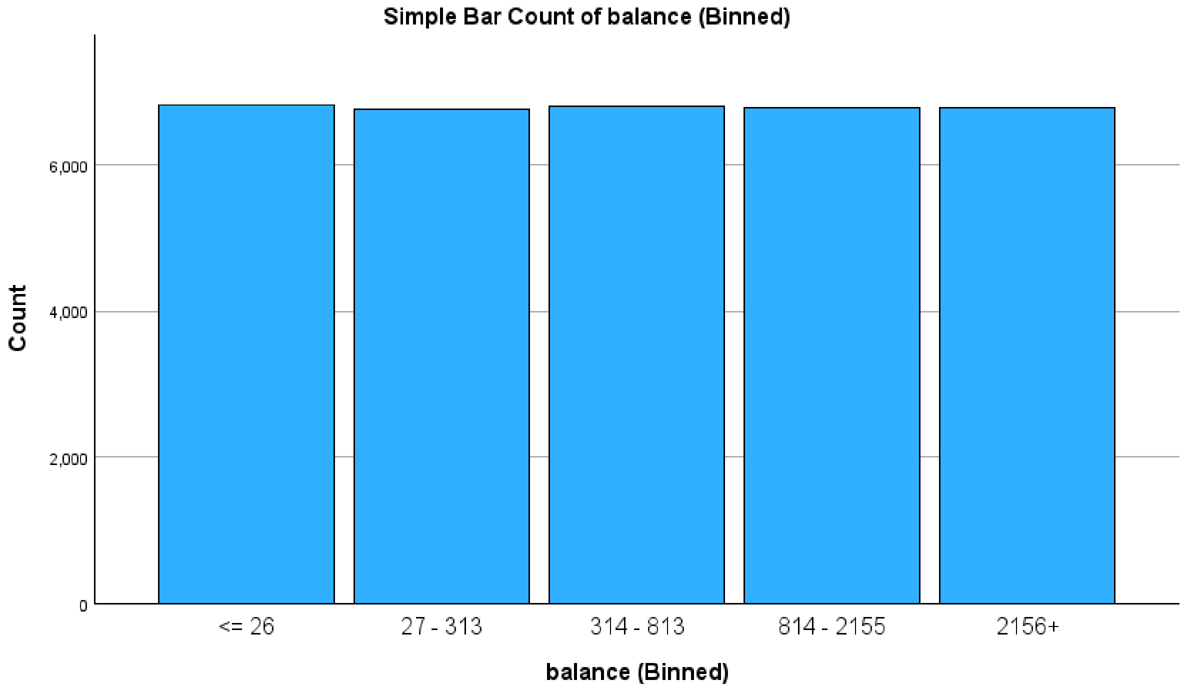


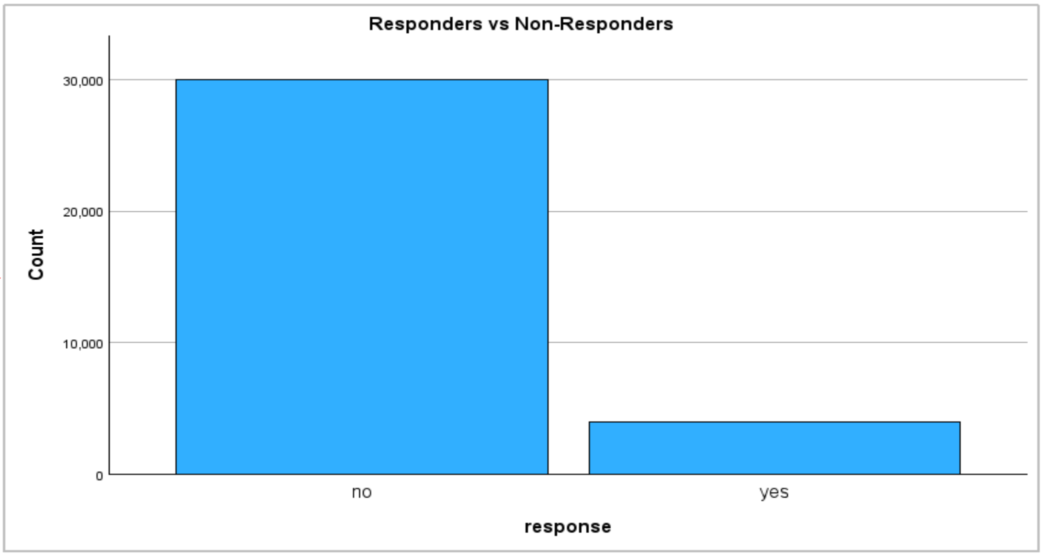
Figure 6 – Distribution of balance after Equal Frequency Binning



[Please refer to appendix section 5.1 to see the explanation for approach 1 – Equal interval Binning]

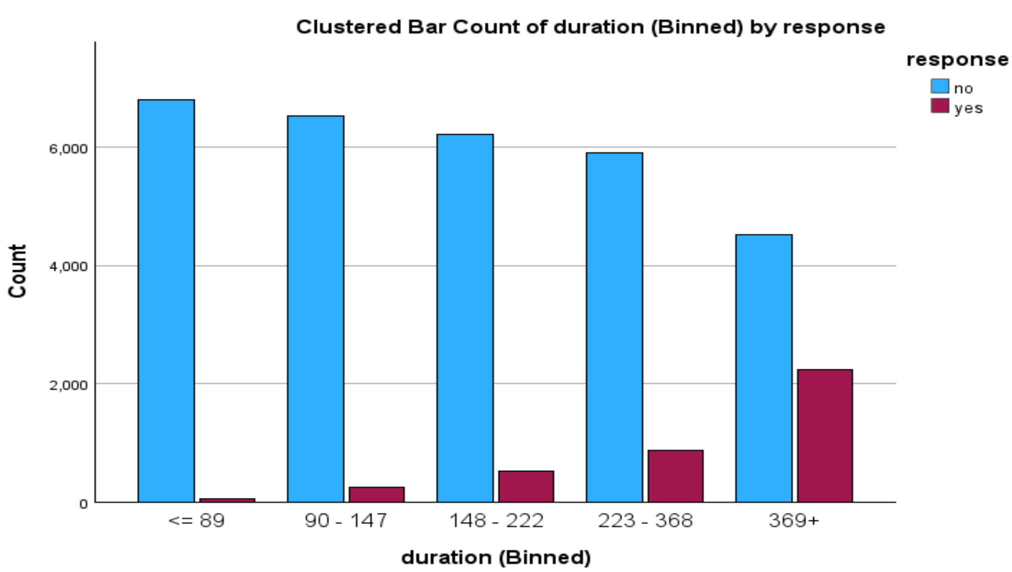
1.5 Exploratory data analysis

Figure 7 – Responders vs Non - Responders



From Figure 7, it is evident that a larger number of customers did not respond to the marketing campaign, as indicated by the "no" bar, which reaches close to the 30,000 mark. In contrast, the "yes" bar, representing the customers who responded positively to the campaign, is significantly lower, suggesting that the campaign had a comparatively low response rate.

Figure 8 – Customer contact duration vs customer response



For shorter contact durations (<= 89 seconds), there's a high number of contacts, but the response rate is very low, suggesting that very short calls are not very effective in generating a positive response. On the other hand, the longest calls (369+ seconds) have a much better chance of yielding a positive response.

Figure 9 – Default Credit vs Response

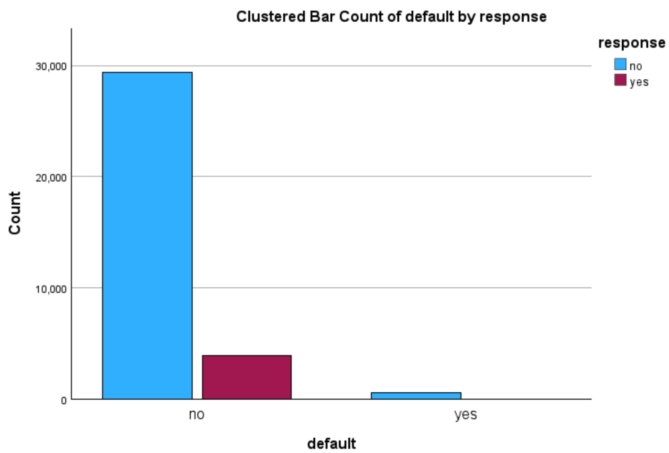
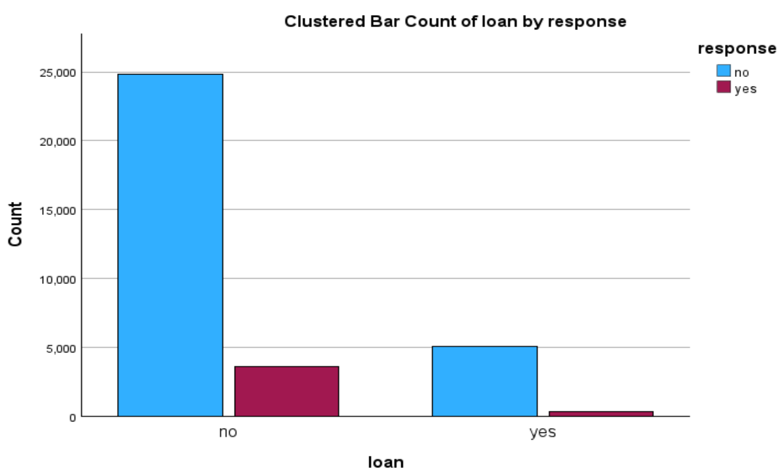


Figure 9 shows that customers without a credit history are more likely to respond to the campaign compared to customers with a credit history.

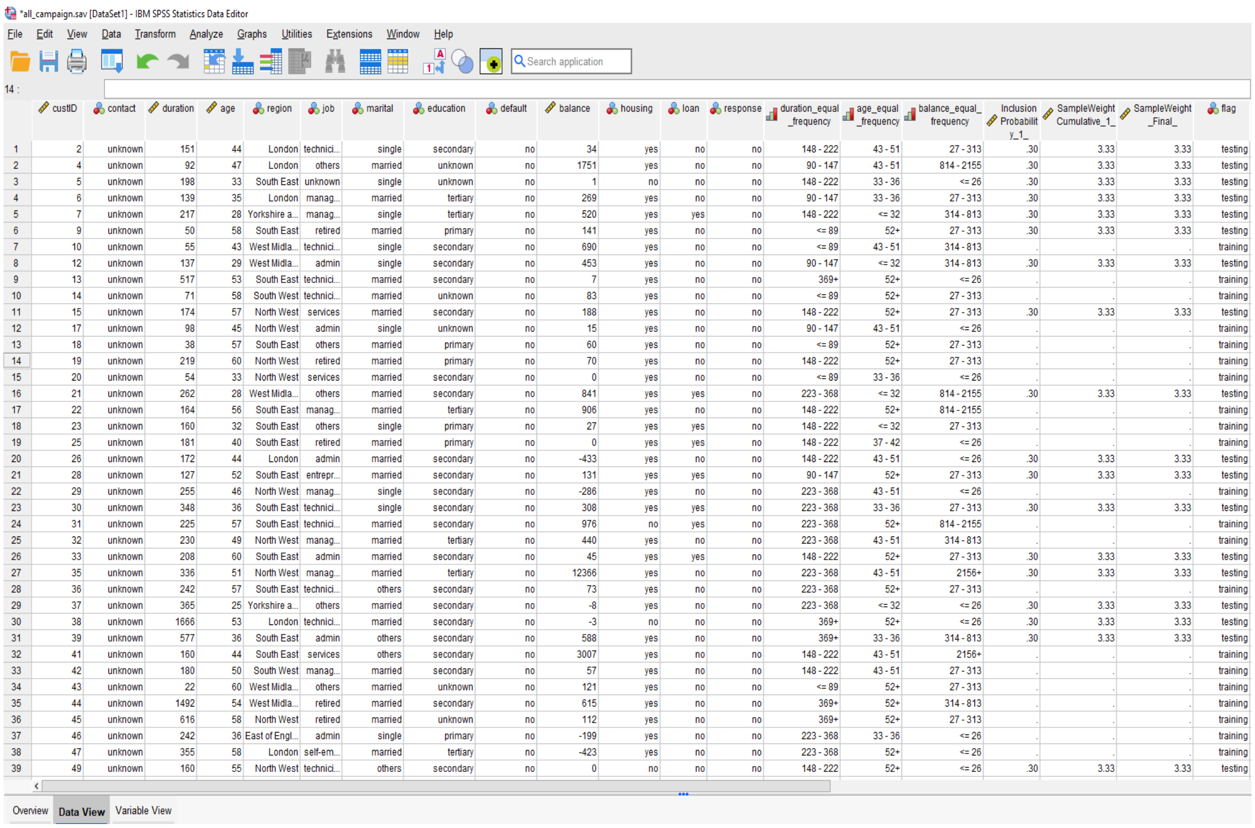
Figure 10 – Personal Loan vs Response



According to Figure 10, there is a higher likelihood of campaign response among consumers who do not possess a personal loan, as opposed to customers who do possess a personal loan.

1.6 Data Splitting

Figure 11 – Train - Test Split



	custID	contact	duration	age	region	job	marital	education	default	balance	housing	loan	response	duration_equal_frequency	age_equal_frequency	balance_equal_frequency	Inclusion Probability_1_	SampleWeight Cumulative_1_	SampleWeight_Final_	flag
1	2	unknown	151	44	London	technici...	single	secondary	no	34	yes	no	no	148 - 222	43 - 51	27 - 313	.30	3.33	3.33	testing
2	4	unknown	92	47	London	others	married	unknown	no	1751	yes	no	no	90 - 147	43 - 51	814 - 2155	.30	3.33	3.33	testing
3	5	unknown	198	33	South East	unknown	single	unknown	no	1	no	no	no	148 - 222	33 - 36	<= 26	.30	3.33	3.33	testing
4	6	unknown	139	35	London	manag...	married	tertiary	no	269	yes	no	no	90 - 147	33 - 36	27 - 313	.30	3.33	3.33	testing
5	7	unknown	217	28	Yorkshire a...	manag...	single	tertiary	no	520	yes	yes	no	148 - 222	<= 32	314 - 813	.30	3.33	3.33	testing
6	9	unknown	50	58	South East	retired	married	primary	no	141	yes	no	no	<= 89	52+	27 - 313	.30	3.33	3.33	testing
7	10	unknown	55	43	West Midla...	technici...	single	secondary	no	690	yes	no	no	<= 89	43 - 51	314 - 813	.	.	.	training
8	12	unknown	137	29	West Midla...	admin	single	secondary	no	453	yes	no	no	90 - 147	<= 32	314 - 813	.30	3.33	3.33	testing
9	13	unknown	517	53	South East	technici...	married	secondary	no	7	yes	no	no	369+	52+	<= 26	.	.	.	training
10	14	unknown	71	58	South West	technici...	married	unknown	no	83	yes	no	no	<= 89	52+	27 - 313	.	.	.	training
11	15	unknown	174	57	North West	services	married	secondary	no	188	yes	no	no	148 - 222	52+	27 - 313	.30	3.33	3.33	testing
12	17	unknown	98	45	North West	admin	single	unknown	no	15	yes	no	no	90 - 147	43 - 51	<= 26	.	.	.	training
13	18	unknown	38	57	South East	others	married	primary	no	60	yes	no	no	<= 89	52+	27 - 313	.	.	.	training
14	19	unknown	219	60	North West	retired	married	primary	no	70	yes	no	no	148 - 222	52+	27 - 313	.	.	.	training
15	20	unknown	54	33	North West	services	married	secondary	no	0	yes	no	no	<= 89	33 - 36	<= 26	.	.	.	training
16	21	unknown	262	28	West Midla...	others	married	secondary	no	841	yes	yes	no	223 - 368	<= 32	814 - 2155	.30	3.33	3.33	testing
17	22	unknown	164	56	South East	manag...	married	tertiary	no	906	yes	no	no	148 - 222	52+	814 - 2155	.	.	.	training
18	23	unknown	160	32	South East	others	single	primary	no	27	yes	yes	no	148 - 222	<= 32	27 - 313	.	.	.	training
19	25	unknown	181	40	South East	retired	married	primary	no	0	yes	yes	no	148 - 222	37 - 42	<= 26	.	.	.	training
20	26	unknown	172	44	London	admin	married	secondary	no	-433	yes	no	no	148 - 222	43 - 51	<= 26	.30	3.33	3.33	testing
21	28	unknown	127	52	South East	entrepr...	married	secondary	no	131	yes	yes	no	90 - 147	52+	27 - 313	.30	3.33	3.33	testing
22	29	unknown	255	46	North West	manag...	single	secondary	no	-286	yes	no	no	223 - 368	43 - 51	<= 26	.	.	.	training
23	30	unknown	348	36	South East	technici...	single	secondary	no	308	yes	yes	no	223 - 368	33 - 36	27 - 313	.30	3.33	3.33	testing
24	31	unknown	225	57	South East	technici...	married	secondary	no	976	no	yes	no	223 - 368	52+	814 - 2155	.	.	.	training
25	32	unknown	230	49	North West	manag...	married	tertiary	no	440	yes	no	no	223 - 368	43 - 51	314 - 813	.	.	.	training
26	33	unknown	208	60	South East	admin	married	secondary	no	45	yes	yes	no	148 - 222	52+	27 - 313	.30	3.33	3.33	testing
27	35	unknown	336	51	North West	manag...	married	tertiary	no	12366	yes	no	no	223 - 368	43 - 51	2156+	.30	3.33	3.33	testing
28	36	unknown	242	57	South East	technici...	others	secondary	no	73	yes	no	no	223 - 368	52+	27 - 313	.	.	.	training
29	37	unknown	365	25	Yorkshire a...	others	married	secondary	no	-8	yes	no	no	223 - 368	<= 32	<= 26	.30	3.33	3.33	testing
30	38	unknown	1666	53	London	technici...	married	secondary	no	-3	no	no	no	369+	52+	<= 26	.30	3.33	3.33	testing
31	39	unknown	577	36	South East	admin	others	secondary	no	588	yes	no	no	369+	33 - 36	314 - 813	.30	3.33	3.33	testing
32	41	unknown	160	44	South East	services	others	secondary	no	3007	yes	no	no	148 - 222	43 - 51	2156+	.	.	.	training
33	42	unknown	180	50	South West	manag...	married	secondary	no	57	yes	no	no	148 - 222	43 - 51	27 - 313	.	.	.	training
34	43	unknown	22	60	West Midla...	others	married	unknown	no	121	yes	no	no	<= 89	52+	27 - 313	.	.	.	training
35	44	unknown	1492	54	West Midla...	retired	married	secondary	no	615	yes	no	no	369+	52+	314 - 813	.	.	.	training
36	45	unknown	616	58	North West	retired	married	unknown	no	112	yes	no	no	369+	52+	27 - 313	.	.	.	training
37	46	unknown	242	36	East of Engl...	admin	single	primary	no	-199	yes	no	no	223 - 368	33 - 36	<= 26	.	.	.	training
38	47	unknown	355	58	London	self-em...	married	tertiary	no	-423	yes	no	no	223 - 368	52+	<= 26	.	.	.	training
39	49	unknown	160	55	North West	technici...	others	secondary	no	0	no	no	no	148 - 222	52+	<= 26	.30	3.33	3.33	testing

The data is split into training and test sets (70% training and 30% testing) using a stratified sampling approach with a **seed value of 260**. The result of this step is we get 3 new variables and out of 3, the variable of interest is Inclusion probability. If Inclusion probability = 0.3, then it's test data, and If Inclusion probability =. (missing) then its training data. The inclusion probability values are then coded into a new variable named flag. The outcome of this step is we get a new variable flag with values "testing" and "training". Now, the 3 new unwanted columns are removed while the flag column is kept intact. Finally, the data is split into 2 files by putting all flag 1 records into the training dataset and all flag 0 records into the test dataset. Also, the training dataset has 23,736 observations and the test dataset has 10,173 observations respectively.

2. Build a Response Model

2.1 Logistic Regression Model 2 (final model)

Table 3 – Frequency table for the target variable (response)

response

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	20959	88.3	88.3	88.3
	yes	2777	11.7	11.7	100.0
	Total	23736	100.0	100.0	

Table 3 indicates a response rate of 11.7% from customers to the campaign, demonstrating a relatively low level of engagement.

Table 4 – Descriptive Statistics

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
duration	23736	0	3785	258.08	255.136
age	23736	18	93	40.98	10.587
balance	23736	-3852	94423	1558.56	3351.981
Valid N (listwise)	23736				

Table 4 presents descriptive statistics for three variables, duration, age, and balance across a dataset of 23,736 entries. The duration varies from 0 to 3,785 seconds with an average of about 258, suggesting a wide range of values with substantial variation, as indicated by the standard deviation of approximately 255. The age of participants ranges from 18 to 93 years, with the average age being around 41 years and relatively less variability (standard deviation of approximately 10.6). The balance has an extensive range from a deficit of -3,852£ to a positive balance of 94,423£, averaging at 1,558.56, but with a very high standard deviation of about 3,351.981, pointing to significant disparities in financial standing among the participants.

Table 5 – Categorical Variable Coding for Logistic Regression Model 2

Categorical Variables Codings

			Parameter coding					
		Frequency	(1)	(2)	(3)	(4)	(5)	(6)
job_update	admin	2672	1.000	.000	.000	.000	.000	.000
	management	4964	.000	1.000	.000	.000	.000	.000
	retired	1170	.000	.000	1.000	.000	.000	.000
	student	471	.000	.000	.000	1.000	.000	.000
	technician	4027	.000	.000	.000	.000	1.000	.000
	unemployed	689	.000	.000	.000	.000	.000	1.000
	others	9743	.000	.000	.000	.000	.000	.000
contact	mobile	15467	1.000	.000				
	telephone	1531	.000	1.000				
	unknown	6738	.000	.000				
marital	others	2708	1.000	.000				
	married	14330	.000	1.000				
	single	6698	.000	.000				
default	no	23333	1.000					
	yes	403	.000					
education_update	tertiary	6962	1.000					
	others	16774	.000					
loan	no	19918	1.000					
	yes	3818	.000					
housing	no	10506	1.000					
	yes	13230	.000					

Table 6 - Logistic Regression Model 2

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	contact			354.701	2	<.001	
	contact(1)	1.443	.077	354.624	1	<.001	4.234
	contact(2)	1.268	.115	121.986	1	<.001	3.554
	duration	.004	.000	2210.022	1	<.001	1.004
	marital			44.059	2	<.001	
	marital(1)	-.211	.082	6.646	1	.010	.810
	marital(2)	-.355	.054	44.004	1	<.001	.701
	default(1)	.468	.233	4.022	1	.045	1.597
	balance	.000	.000	14.401	1	<.001	1.000
	housing(1)	.714	.050	203.122	1	<.001	2.043
	loan(1)	.617	.077	63.378	1	<.001	1.853
	job_update			177.637	6	<.001	
	job_update(1)	.547	.079	48.177	1	<.001	1.728
	job_update(2)	.270	.077	12.413	1	<.001	1.310
	job_update(3)	.983	.093	111.013	1	<.001	2.674
	job_update(4)	1.084	.127	72.910	1	<.001	2.956
	job_update(5)	.213	.072	8.772	1	.003	1.237
	job_update(6)	.368	.131	7.922	1	.005	1.444
	education_update(1)	.221	.061	13.017	1	<.001	1.247
	Constant	-6.060	.256	560.287	1	<.001	.002

a. Variable(s) entered on step 1: contact, duration, marital, default, balance, housing, loan, job_update, education_update.

Using a significance threshold of 0.1, the logistic regression model under discussion includes 7 categorical variables (contact, job, marital, education, default, housing, loan) and 2 numerical variables (duration, balance). The insignificant variables ($p > 0.1$) ‘age’ and ‘region’ were removed from the model. The insignificant subcategories ‘others’, ’entrepreneur’, ‘domestic worker’, ‘self-employed’, and ‘services’ in the ‘job’ variable are merged with the reference category ‘unknown’ to form a new variable ‘job_update’ with only 7 subcategories. Similarly, the insignificant subcategories ‘primary’, and ‘secondary’ in the ‘education’ variable are merged with the reference category ‘unknown’ to form a new variable education_update which now has only 2 subcategories.

Impact of Input Variables on the Target

Duration

- Significant relationship with response (at 0.1 level)
- A higher "duration" of contact is more likely to respond to the offer.
- Keeping all other variables constant, for a one-unit increase in contact duration, it is estimated to see a 0.4% increase in the odds of response.

Balance

- Keeping all other variables constant, for each unit increase in balance, the odds of response increases by 0%, indicating a minimal effect.

Contact

- There is a significant difference among different categories of contact (at 0.1 level)
- As all regression coefficients are positive and significant (at 0.1), all other categories are more likely to respond than "Unknown". Comparing the magnitudes, it is estimated that customers contacted through "Mobile" are the most likely to respond.
- Keeping other variables constant, the odds of "Mobile" customers responding are 4.234 times greater than those of "Unknown" customers. For the other group of customers, the odds of customers in "Telephone" are 3.554 times higher than those of "Unknown" customers.

Marital

- "Single" is the reference category of "Marital". As all regression coefficients are negative and significant (at 0.1), all other categories are less likely to respond than "single".
- marital(1) represents "others". As the regression coefficient is significant (i.e. 0.010) and negative (i.e. -.211), this means: that compared with the reference category (i.e. single), customers with ''others" marital status are less likely to respond.
- Compared to single individuals (reference category), being married (marital(2)) decreases the odds of response by a factor of 0.701.

Housing

- Not having a housing loan (housing(1)) increases the odds of response by a factor of 2.043 compared to having one.

Loan

- Not having a personal loan (loan(1)) increases the odds of response by a factor of 1.853 compared to having one.

Job

- Significant difference among different categories (at 0.1 level)
- As all regression coefficients are positive and significant (at 0.1), all other categories are more likely to respond than the reference category “others”. Comparing the magnitudes, it is estimated that “student” is the most likely to respond. The second group is those in “retired” followed by “admin”.
- Keeping other variables constant, the odds of “student” customers responding are 2.956 times greater than those of “other” customers. The odds of customers in “retired” and “admin” are 2.674 and 1.728 times higher than those of “other” customers.

Education

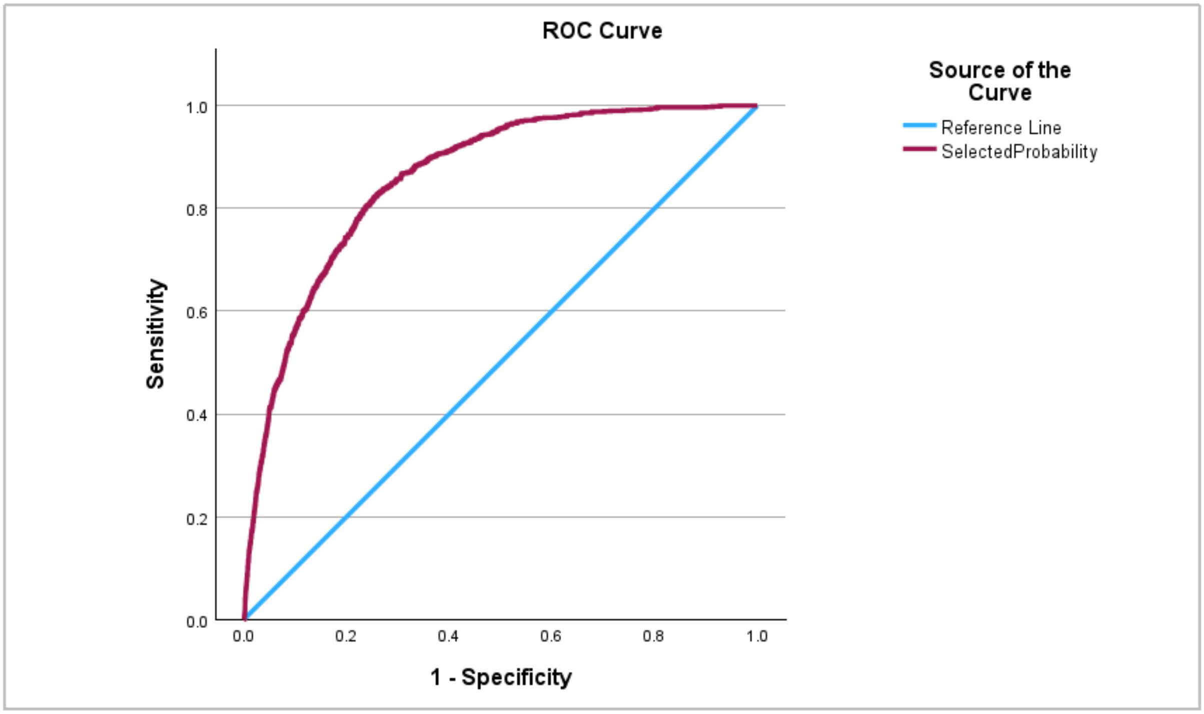
- Compared to the reference education level (ie others), having a tertiary education (education_update(1)) increases the odds of response by a factor of 1.247.

Default credit

- Not having a credit default (default(1)) increases the odds of response by a factor of 1.597 compared to having a credit default.

Performance of logistic regression model

Figure 12 – ROC Curve



Area Under the ROC Curve

Test Result Variable(s): SelectedProbability

Area
.859

The test result variable (s): SelectedProbability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Figure 12 shows that the model has an AUC value of 0.859, which indicates a very good ability to discriminate between respondents and non-respondents. The curve and the AUC indicate that the logistic regression model has a high level of performance in predicting customer responses to the marketing campaign.

[Refer to section 5.2 in the appendix to see the explanation for the baseline model]

3. Marketing Campaign

Based on the findings from the logistic regression analysis, the following marketing campaign plan is designed to target the customer segments most likely to respond.

3.1 Target Customers

Engagement Focus: Prioritize customers who have engaged longer in calls (369+ seconds), as the data indicates a higher chance of a positive response from this group. These are likely to be individuals who have shown interest in the conversation about the offer.

Marital Status: Target single individuals more aggressively than married ones, as singles have shown a greater likelihood to respond to the campaign.

Housing and Personal Loans: Segment customers based on their loan status. Those without housing or personal loans are more responsive and thus should be a focus.

Occupation: Direct specific campaign efforts towards students and retired individuals, who are more inclined to respond than other job categories.

Education: Aim at customers with tertiary education, who have been identified as more likely to engage with the campaign.

Credit History: Customers with no credit default should be considered safer targets for the campaign since they are more likely to respond.

3.2 Channels

Mobile Marketing: Utilize mobile communication as the primary channel for outreach, capitalizing on the higher response rates observed from customers contacted through mobile phones. Leverage SMS and mobile app notifications, which can be personalized and are known to have a high open rate of 98% and a response rate of 45% (Memud, 2023).

Social Media: Deploy targeted ads on social media platforms popular among students and younger demographics, as well as platforms frequented by retirees.

3.3 Strategies

Personalized Messaging: Create messages that resonate with each segment, like offering financial planning tips for students and retirees.

Value Proposition: Emphasize freedom and flexibility in the offer, which is often more appealing to single individuals and those without the financial burden of loans.

Educational Content: Develop content that educates customers on financial health, which aligns with the interests of those with tertiary education and demonstrates value beyond the product.

Incentives: Introduce incentives for early responses to encourage quick decision-making among target demographics

Data-Driven Duration Analysis: Implement analytics to identify calls reaching the critical 369+ second threshold and flag them for follow-up communications.

Training: Equip call center staff with insights on how to extend productive calls and provide training on dealing with target segments like students and retired individuals.

4. Reference

Memud, S. (2023). Customer Engagement via Email and SMS for Financial Services. [online] www.linkedin.com. Available at: https://www.linkedin.com/pulse/customer-engagement-via-email-sms-financial-services-shina-memud/ [Accessed 31 Mar. 2024].

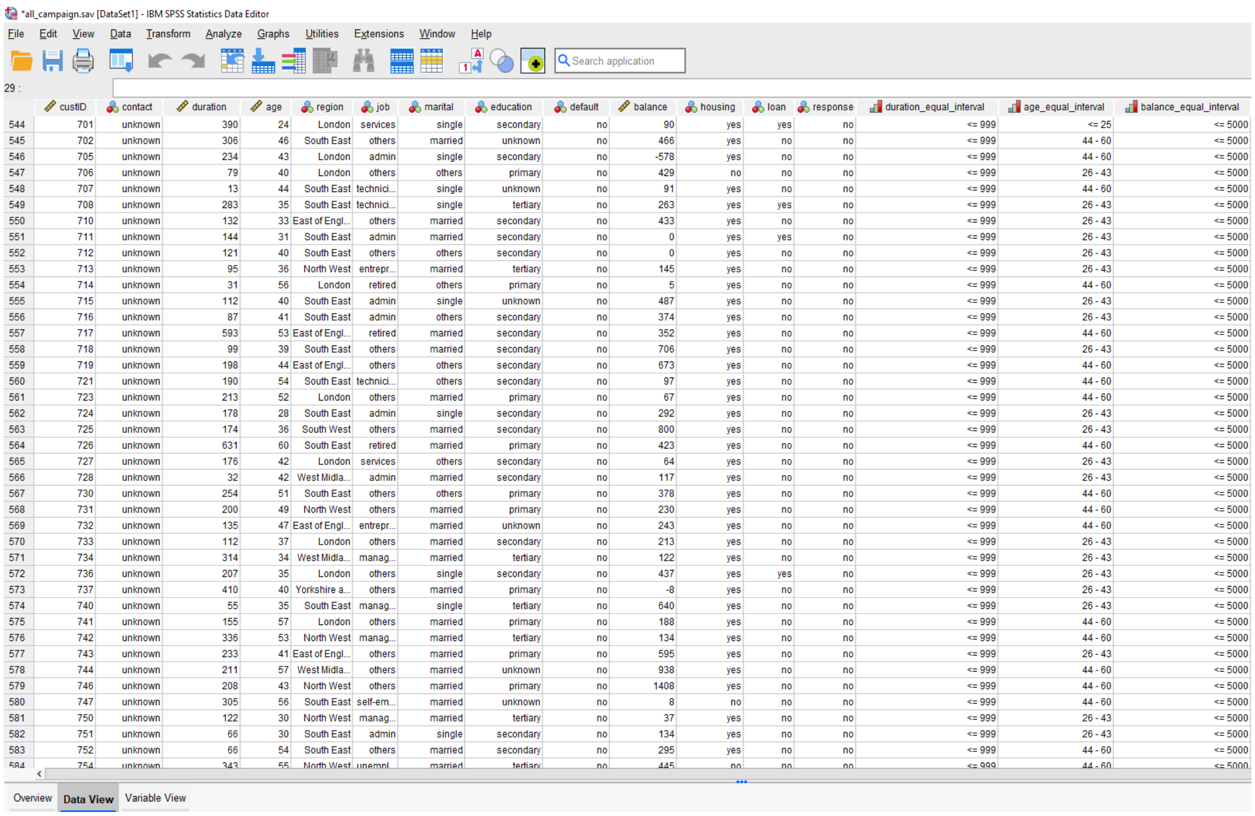
5. Appendix

5.1 Binning Numeric Variables [Approach 1 - Equal Interval Binning]

Equal Interval Binning

This method divides the range of data into intervals of equal size

Figure 13 – Dataset after Equal Interval Binning



	custID	contact	duration	age	region	job	marital	education	default	balance	housing	loan	response	duration_equal_interval	age_equal_interval	balance_equal_interval
544	701	unknown	390	24	London	services	single	secondary	no	90	yes	yes	no	<= 999	<= 25	<= 5000
545	702	unknown	306	46	South East	others	married	unknown	no	466	yes	no	no	<= 999	44 - 60	<= 5000
546	705	unknown	234	43	London	admin	single	secondary	no	-578	yes	no	no	<= 999	44 - 60	<= 5000
547	706	unknown	79	40	London	others	others	primary	no	429	no	no	no	<= 999	26 - 43	<= 5000
548	707	unknown	13	44	South East	technici...	single	unknown	no	91	yes	no	no	<= 999	44 - 60	<= 5000
549	708	unknown	283	35	South East	technici...	single	tertiary	no	263	yes	yes	no	<= 999	26 - 43	<= 5000
550	710	unknown	132	33	East of Engl...	others	married	secondary	no	433	yes	no	no	<= 999	26 - 43	<= 5000
551	711	unknown	144	31	South East	admin	married	secondary	no	0	yes	yes	no	<= 999	26 - 43	<= 5000
552	712	unknown	121	40	South East	others	others	secondary	no	0	yes	no	no	<= 999	26 - 43	<= 5000
553	713	unknown	95	36	North West	entrepr...	married	tertiary	no	145	yes	no	no	<= 999	26 - 43	<= 5000
554	714	unknown	31	56	London	retired	others	primary	no	5	yes	no	no	<= 999	44 - 60	<= 5000
555	715	unknown	112	40	South East	admin	single	unknown	no	487	yes	no	no	<= 999	26 - 43	<= 5000
556	716	unknown	87	41	South East	admin	others	secondary	no	374	yes	no	no	<= 999	26 - 43	<= 5000
557	717	unknown	593	53	East of Engl...	retired	married	secondary	no	352	yes	no	no	<= 999	44 - 60	<= 5000
558	718	unknown	99	39	South East	others	married	secondary	no	706	yes	no	no	<= 999	26 - 43	<= 5000
559	719	unknown	198	44	East of Engl...	others	others	secondary	no	673	yes	no	no	<= 999	44 - 60	<= 5000
560	721	unknown	190	54	South East	technici...	others	secondary	no	97	yes	no	no	<= 999	44 - 60	<= 5000
561	723	unknown	213	52	London	others	married	primary	no	67	yes	no	no	<= 999	44 - 60	<= 5000
562	724	unknown	178	28	South East	admin	single	secondary	no	292	yes	no	no	<= 999	26 - 43	<= 5000
563	725	unknown	174	36	South West	others	married	secondary	no	800	yes	no	no	<= 999	26 - 43	<= 5000
564	726	unknown	631	60	South East	retired	married	primary	no	423	yes	no	no	<= 999	44 - 60	<= 5000
565	727	unknown	176	42	London	services	others	secondary	no	64	yes	no	no	<= 999	26 - 43	<= 5000
566	728	unknown	32	42	West Midla...	admin	married	secondary	no	117	yes	no	no	<= 999	26 - 43	<= 5000
567	730	unknown	254	51	South East	others	others	primary	no	378	yes	no	no	<= 999	44 - 60	<= 5000
568	731	unknown	200	49	North West	others	married	primary	no	230	yes	no	no	<= 999	44 - 60	<= 5000
569	732	unknown	135	47	East of Engl...	entrepr...	married	unknown	no	243	yes	no	no	<= 999	44 - 60	<= 5000
570	733	unknown	112	37	London	others	married	secondary	no	213	yes	no	no	<= 999	26 - 43	<= 5000
571	734	unknown	314	34	West Midla...	manag...	married	tertiary	no	122	yes	no	no	<= 999	26 - 43	<= 5000
572	736	unknown	207	35	London	others	single	secondary	no	437	yes	yes	no	<= 999	26 - 43	<= 5000
573	737	unknown	410	40	Yorkshire a...	others	married	primary	no	-8	yes	no	no	<= 999	26 - 43	<= 5000
574	740	unknown	55	35	South East	manag...	single	tertiary	no	640	yes	no	no	<= 999	26 - 43	<= 5000
575	741	unknown	155	57	London	others	married	primary	no	188	yes	no	no	<= 999	44 - 60	<= 5000
576	742	unknown	336	53	North West	manag...	married	tertiary	no	134	yes	no	no	<= 999	44 - 60	<= 5000
577	743	unknown	233	41	East of Engl...	others	married	primary	no	595	yes	no	no	<= 999	26 - 43	<= 5000
578	744	unknown	211	57	West Midla...	others	married	unknown	no	938	yes	no	no	<= 999	44 - 60	<= 5000
579	746	unknown	208	43	North West	others	married	primary	no	1408	yes	no	no	<= 999	44 - 60	<= 5000
580	747	unknown	305	56	South East	self-em...	married	unknown	no	8	no	no	no	<= 999	44 - 60	<= 5000
581	750	unknown	122	30	North West	manag...	married	tertiary	no	37	yes	no	no	<= 999	26 - 43	<= 5000
582	751	unknown	66	30	South East	admin	single	secondary	no	134	yes	no	no	<= 999	26 - 43	<= 5000
583	752	unknown	66	54	South East	others	married	secondary	no	295	yes	no	no	<= 999	44 - 60	<= 5000
584	754	unknown	343	55	North West	unempl...	married	tertiary	no	445	no	no	no	<= 999	44 - 60	<= 5000

Approach 1 - numeric variables 'duration', 'age', and 'balance' are binned based on the equal interval binning method.

Binning details

Duration – [bins = 5, bin width = 979 seconds]

Age – [bins = 5, bin width = 17 years]

Balance – [bins = 5, bin width = 27360 £]

Figure 14 – Distribution of duration

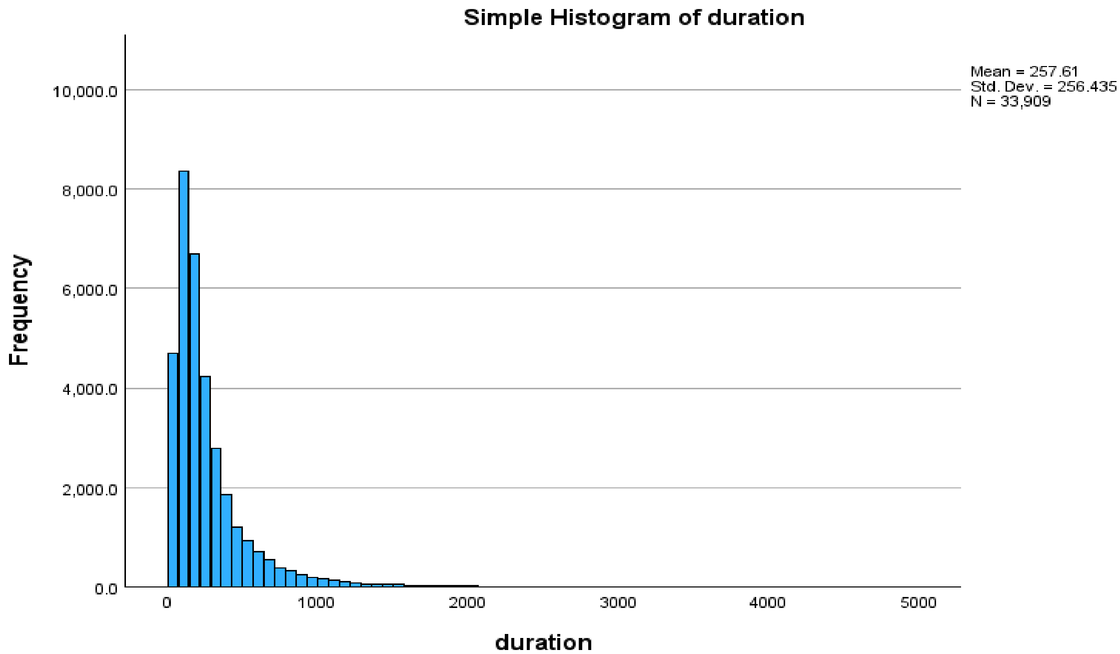


Figure 15 – Distribution of age

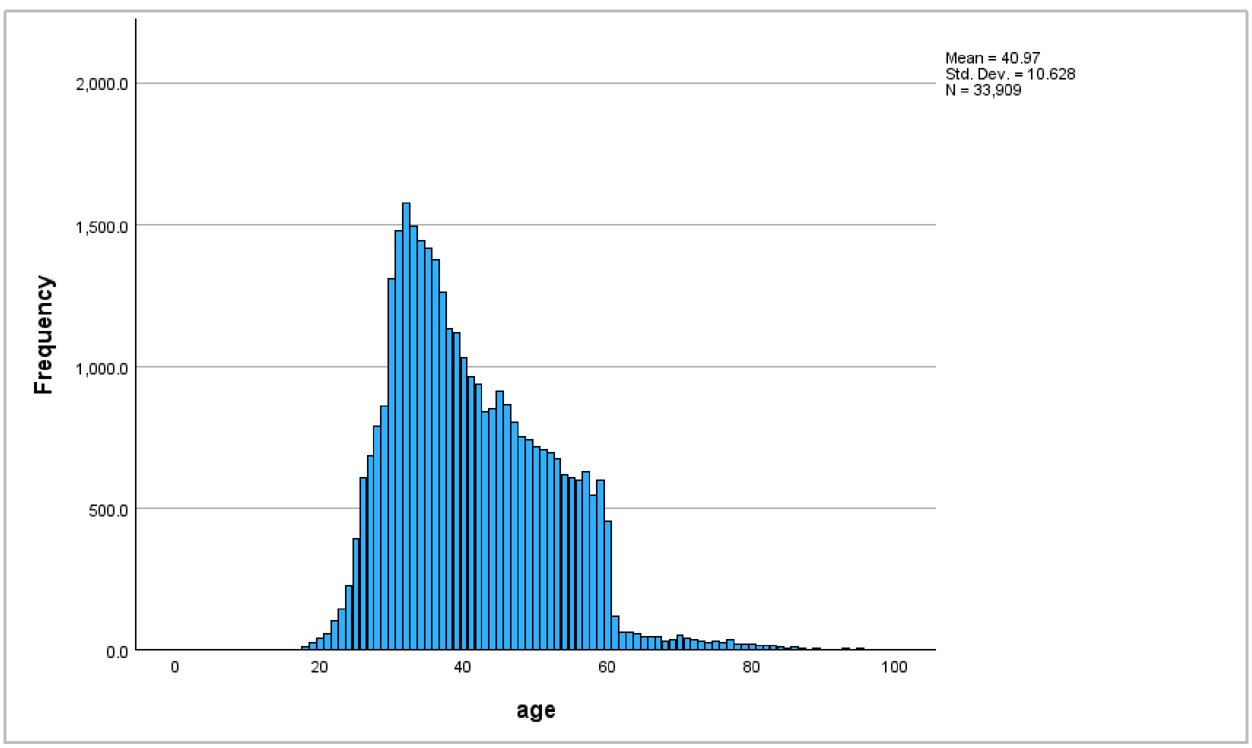


Figure 16 – Distribution of balance

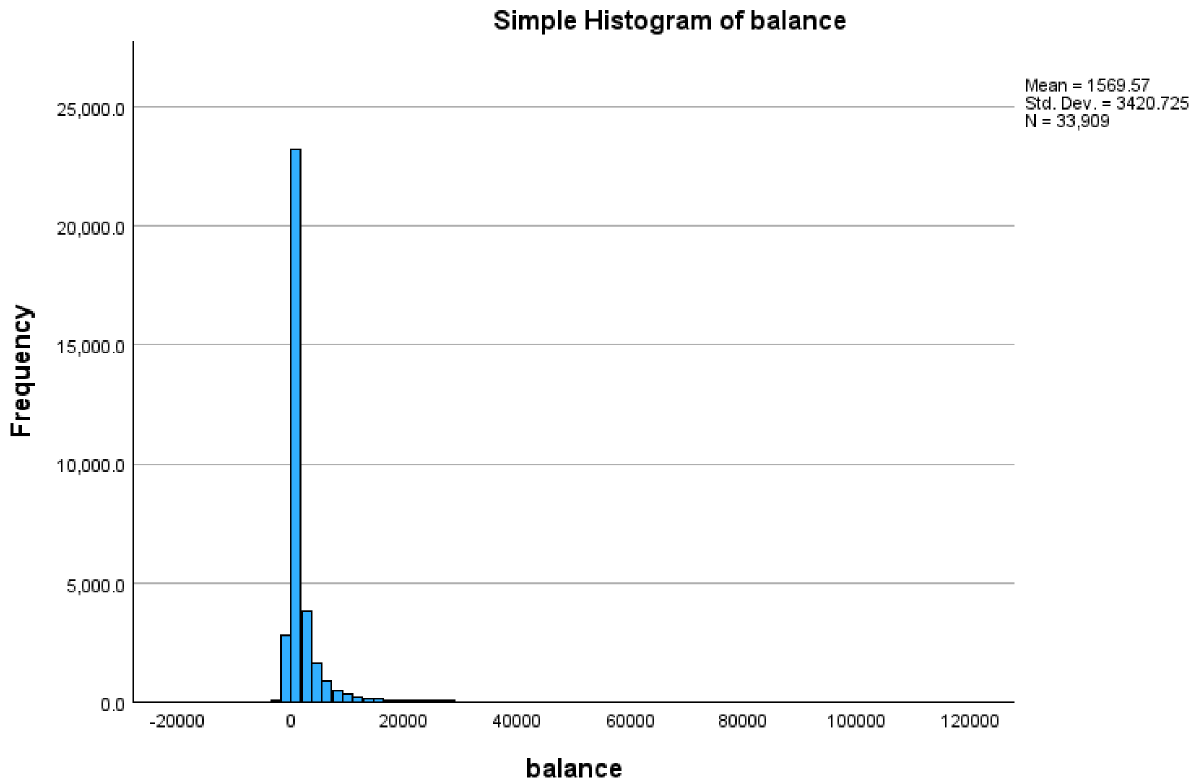


Figure 17 – Distribution of duration after binning

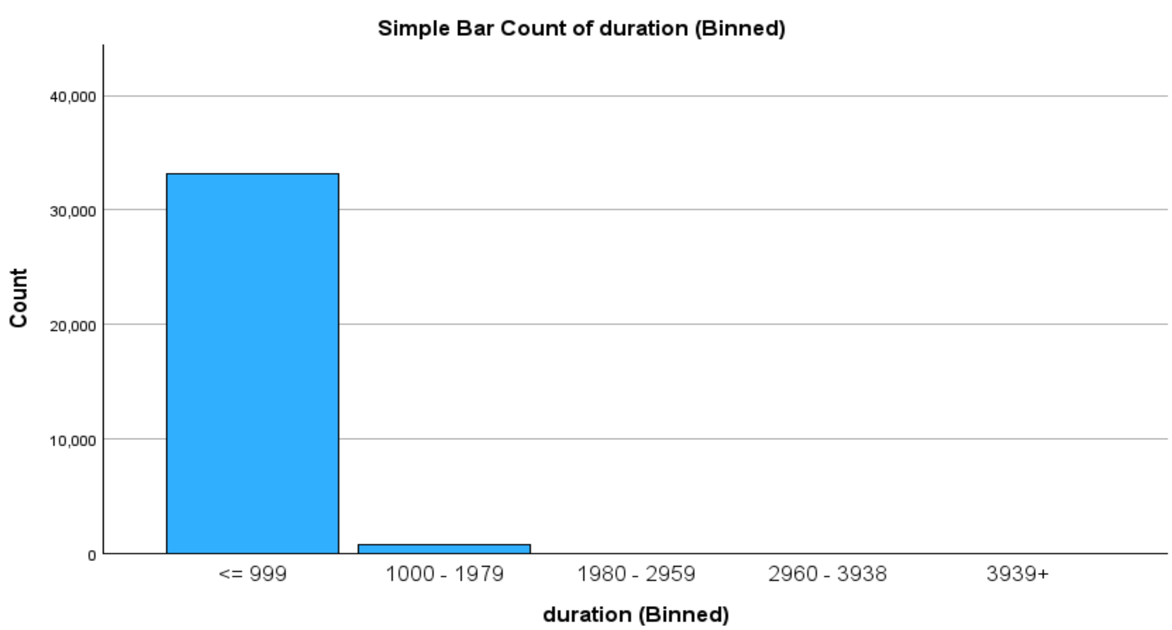


Figure 18 – Distribution of age after binning

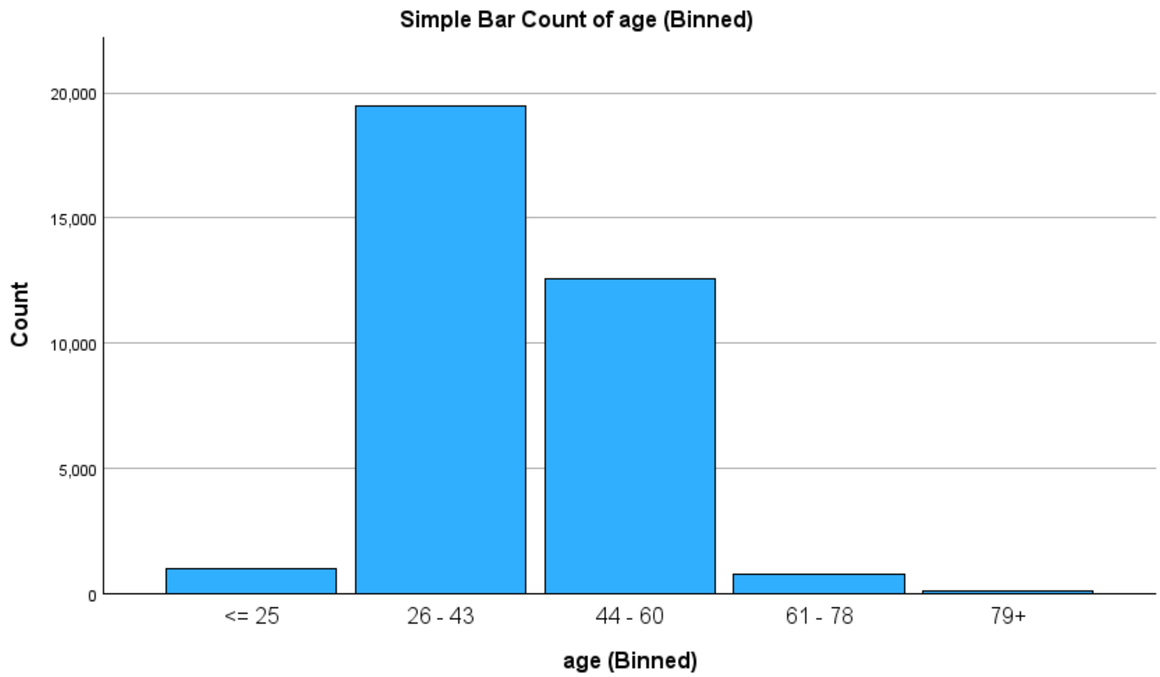
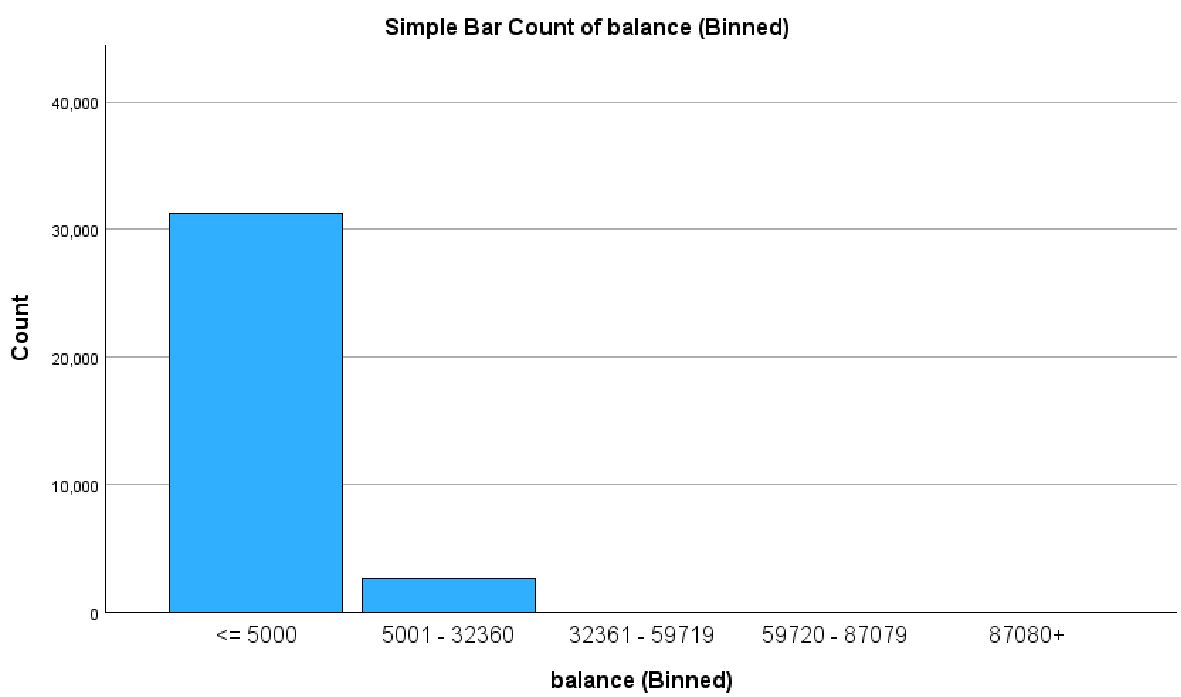


Figure 19 – Distribution of balance after binning



Since the numeric variables ‘duration’, ’age’, and ‘balance’ have skewed or long-tailed distributions it is not appropriate to use this binning method as these lead to many data points being concentrated in a few bins while leaving others empty or sparsely populated. Equal frequency binning, on the other hand, ensures that each bin has the same number of observations, which can be more appropriate for handling outliers and skewed distributions.

5.2 Logistic Regression Model 1 (Initial baseline model)

Table 7 – Categorical Variable Coding for Logistic Regression Model 1

Categorical Variables Codings

			Parameter coding										
		Frequency	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
job	admin	2672	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	others	5120	.000	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	entrepreneur	760	.000	.000	1.000	.000	.000	.000	.000	.000	.000	.000	.000
	domestic worker	681	.000	.000	.000	1.000	.000	.000	.000	.000	.000	.000	.000
	management	4964	.000	.000	.000	.000	1.000	.000	.000	.000	.000	.000	.000
	retired	1170	.000	.000	.000	.000	.000	1.000	.000	.000	.000	.000	.000
	self-employed	872	.000	.000	.000	.000	.000	.000	1.000	.000	.000	.000	.000
	services	2176	.000	.000	.000	.000	.000	.000	.000	1.000	.000	.000	.000
	student	471	.000	.000	.000	.000	.000	.000	.000	.000	1.000	.000	.000
	technician	4027	.000	.000	.000	.000	.000	.000	.000	.000	.000	1.000	.000
	unemployed	689	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	1.000
	unknown	134	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
region	North East	115	1.000	.000	.000	.000	.000	.000	.000	.000			
	South West	725	.000	1.000	.000	.000	.000	.000	.000	.000			
	East of England	2604	.000	.000	1.000	.000	.000	.000	.000	.000			
	London	5141	.000	.000	.000	1.000	.000	.000	.000	.000			
	South East	6501	.000	.000	.000	.000	1.000	.000	.000	.000			
	North West	5205	.000	.000	.000	.000	.000	1.000	.000	.000			
	West Midlands	2575	.000	.000	.000	.000	.000	.000	1.000	.000			
	Yorkshire and the Humber	782	.000	.000	.000	.000	.000	.000	.000	1.000			
	East Midlands	88	.000	.000	.000	.000	.000	.000	.000	.000			
education	primary	3598	1.000	.000	.000								
	secondary	12224	.000	1.000	.000								
	tertiary	6962	.000	.000	1.000								
	unknown	952	.000	.000	.000								
marital	others	2708	1.000	.000									
	married	14330	.000	1.000									
	single	6698	.000	.000									
contact	mobile	15467	1.000	.000									
	telephone	1531	.000	1.000									
	unknown	6738	.000	.000									
default	no	23333	1.000										
	yes	403	.000										
housing	no	10506	1.000										
	yes	13230	.000										
loan	no	19918	1.000										
	yes	3818	.000										

Table 8 - Logistic Regression Model 1

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	contact			352.800	2	<.001	
	contact(1)	1.442	.077	352.733	1	<.001	4.230
	contact(2)	1.267	.116	119.649	1	<.001	3.550
	duration	.004	.000	2205.863	1	<.001	1.004
	age	.003	.003	1.358	1	.244	1.003
	region			5.547	8	.698	
	region(1)	.235	.632	.138	1	.711	1.264
	region(2)	.635	.497	1.634	1	.201	1.887
	region(3)	.662	.483	1.877	1	.171	1.939
	region(4)	.697	.481	2.102	1	.147	2.008
	region(5)	.665	.480	1.919	1	.166	1.945
	region(6)	.627	.481	1.700	1	.192	1.872
	region(7)	.746	.483	2.386	1	.122	2.110
	region(8)	.623	.495	1.588	1	.208	1.865
	job			163.200	11	<.001	
	job(1)	1.071	.385	7.720	1	.005	2.917
	job(2)	.585	.385	2.316	1	.128	1.796
	job(3)	.415	.408	1.038	1	.308	1.515
	job(4)	.366	.411	.796	1	.372	1.442
	job(5)	.803	.383	4.387	1	.036	2.232
	job(6)	1.485	.388	14.614	1	<.001	4.413
	job(7)	.638	.398	2.562	1	.109	1.892
	job(8)	.602	.390	2.390	1	.122	1.826
	job(9)	1.643	.397	17.134	1	<.001	5.172
	job(10)	.737	.383	3.697	1	.054	2.090
	job(11)	.914	.398	5.270	1	.022	2.494
	marital			39.496	2	<.001	
	marital(1)	-.233	.087	7.117	1	.008	.792
	marital(2)	-.367	.059	39.211	1	<.001	.693
	education			16.369	3	<.001	
	education(1)	-.182	.137	1.764	1	.184	.834
	education(2)	-.017	.121	.021	1	.886	.983
	education(3)	.189	.127	2.207	1	.137	1.208
	default(1)	.468	.234	4.009	1	.045	1.596
	balance	.000	.000	13.367	1	<.001	1.000
	housing(1)	.716	.051	195.808	1	<.001	2.045
	loan(1)	.618	.078	63.232	1	<.001	1.854
	Constant	-7.350	.677	117.865	1	<.001	.001

a. Variable(s) entered on step 1: contact, duration, age, region, job, marital, education, default, balance, housing, loan.

Using a significance threshold of 0.1, the logistic regression model under discussion includes 8 categorical variables (contact, region, job, marital, education, default, housing, loan) and 3 numerical variables (duration, age, balance). The 'age' variable is not significant since its p-value exceeds 0.1, suggesting that it can be excluded from the final model. The remaining two numerical variables show statistical significance under this threshold.

The categorical variable 'region', along with all its subcategories, does not show significance at the 0.1 level, leading to its removal from the final model. The 'job' category is significant overall, however, some specific subcategories (job(2), job(3), job(4), job(7), and job(8)) do not meet the significance criterion. These non-significant job categories will be merged with the baseline "unknown" category in a revised final model.

Likewise, the 'education' variable is significant as a whole, but the subcategories 'education(1)' and 'education(2)' are not. These will also be combined with the "unknown" reference category for the final refined model. All other variables are considered significant at the 0.1 significance level.