Template for the Assignment

Table of Contents

CHAPTER 1: Introduction (100 words)

As data scientist, I need to develop a model to solve the missing values in two variables, MARITAL_STATUS and LOAN_AMOUNT variables. The prediction model used was Logistic regression. Before the model is apply, the data need to be cleansed first then replace the missing values with appropriate values.

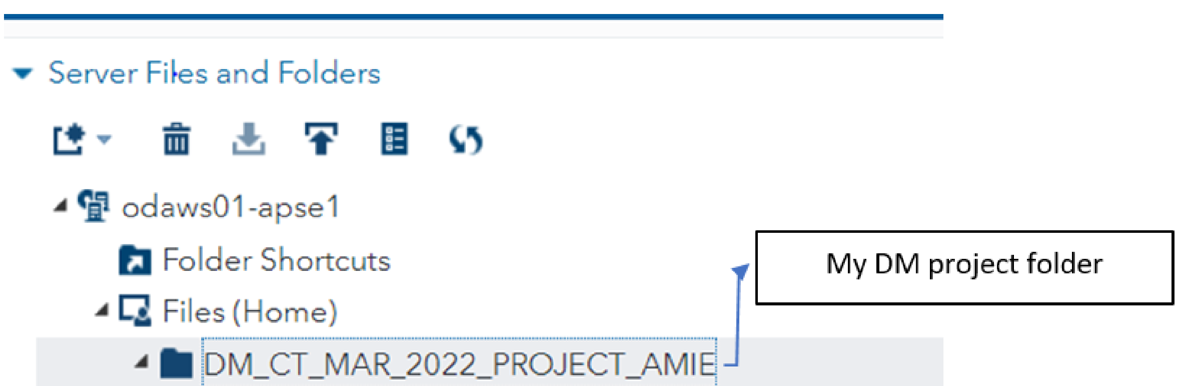
CHAPTER 2: Experimentation

2.1 Create a folder on SAS

2.1.1 Explanation

First create a permanent folder dataset. This folder stores the uploaded dataset(s) that use for this project. This dataset will be temporarily store in the created folder.

2.1.2 Screenshot(s)

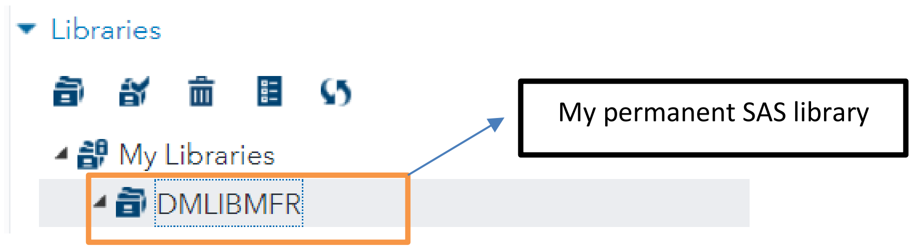


2.2 Create a permanent library on SAS

2.2.1 Explanation

The SAS datasets/files are stored permanently on the created library while creating SAS datasets. Only eight characters is allowed for naming the library. Linked the created folder to the new created permanent library.

2.2.2 Screenshot

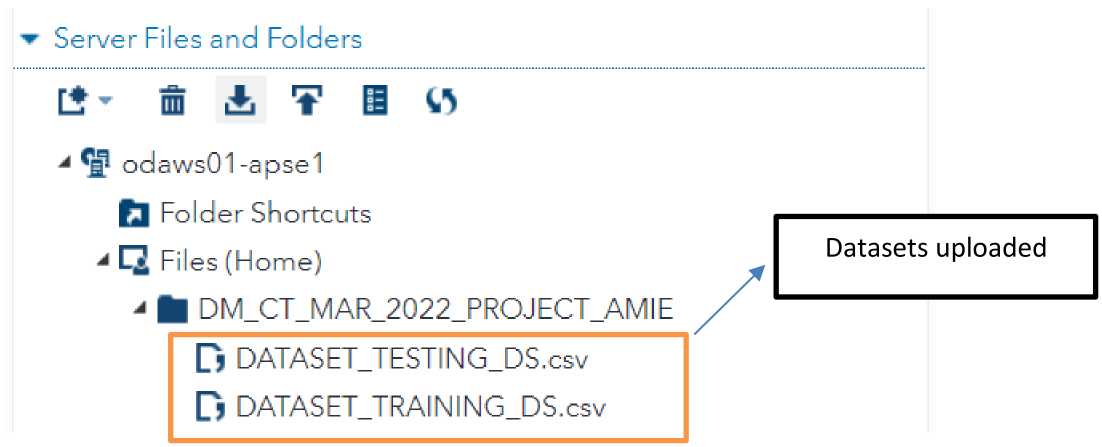


2.3 Upload the datasets DATASET_TRAINING_DS & DATASET_TESTING_DS to the folder DAP----

2.3.1 Explanation

We have to upload the datasets to the created project folder. Before starts to code.

2.3.2 Screenshots

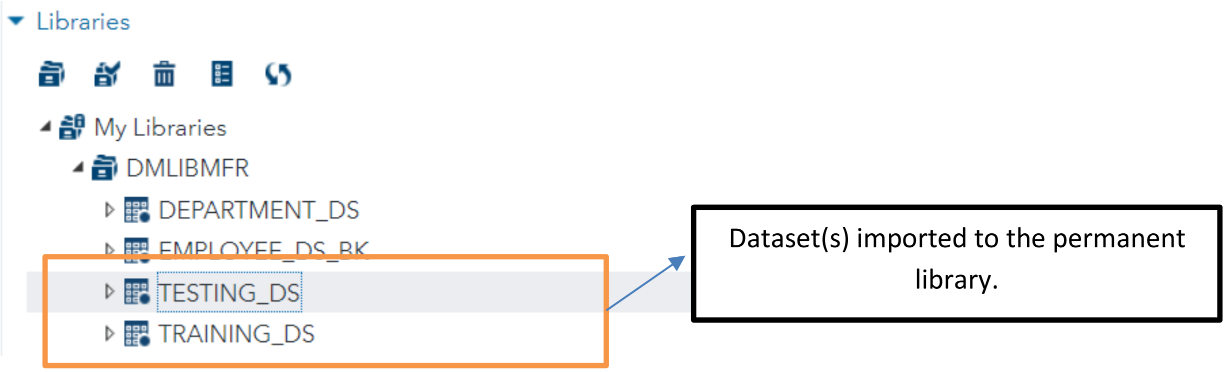


2.4 Import the datasets DATASET_TRAINING_DS & DATESET_TESTING_DS to the newly created library.

2.4.1. Explanation

Before proceeding with the code, let's import the datasets DATASET_TRAINING_DS & DATESET_TESTING_DS to the newly created library which is DMLIBMFR.

2.4.2 Screenshots



2.5 Display the structure (data dictionary) of the training dataset: TRAINING_DS

2.5.1 SAS Codes

```
TITLE1 'Structure/Data Dictionary of the dataset - DMLIBMFR.TRAINING_DS';
PROC CONTENTS DATA = DMLIBMFR.TRAINING_DS;
RUN;
```

2.5.2 Screenshot(s)/Output(s)

Structure/Data Dictionary of the dataset - DMLIBMFR.TRAINING_DS

The CONTENTS Procedure

Data Set Name	DMLIBMFR.TRAINING_DS	Observations	614
Member Type	DATA	Variables	10
Engine	V9	Indexes	0
Created	03/16/2022 14:54:22	Observation Length	80
Last Modified	03/16/2022 14:54:22	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1635
Obs in First Data Page	614
Number of Data Set Repairs	0
Filename	/home/u61014881/DM_CT_MAR_2022_PROJECT_AMIE/training_ds.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	31590072
Access Permission	rw-r--r--
Owner Name	u61014881
File Size	256KB

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
5	CANDIDATE_INCOME	Num	8	BEST12.	BEST32.
2	GENDER	Char	6	$6.	$6.
6	LOAN_AMOUNT	Num	8	BEST12.	BEST32.
10	LOAN_APPROVAL_STATUS	Char	1	$1.	$1.
7	LOAN_DURATION	Num	8	BEST12.	BEST32.
8	LOAN_HISTORY	Num	8	BEST12.	BEST32.
9	LOAN_LOCATION	Char	7	$7.	$7.
3	MARITAL_STATUS	Char	11	$11.	$11.
4	QUALIFICATION	Char	14	$14.	$14.
1	SME_LOAN_ID_NO	Char	8	$8.	$8.

2.5.3 Explanation

The detail structure / data dictionary of the dataset – DMLIBMFR.TRAINING_DS. Furthermore, it also provided the list of attributes. It also shows our folder location.

2.6 Univariate Analysis of variables found in the dataset DMLIBMFR.TRAINING_DS.

2.6.1 Univariate Analysis of the categorical variable: GENDER

2.6.1 SAS Codes

```
TITLE1 'Figure 1 Univariate Analysis of the Categorical variable: Gender';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;
TABLE GENDER;
RUN;

/* This code display a barchart of Gender*/
ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
VBAR GENDER;
Title 'Figure 2 Univariate Analysis of the Categorical variable: Gender';
RUN;
```

2.6.2 Screenshot(s)/Output(s)

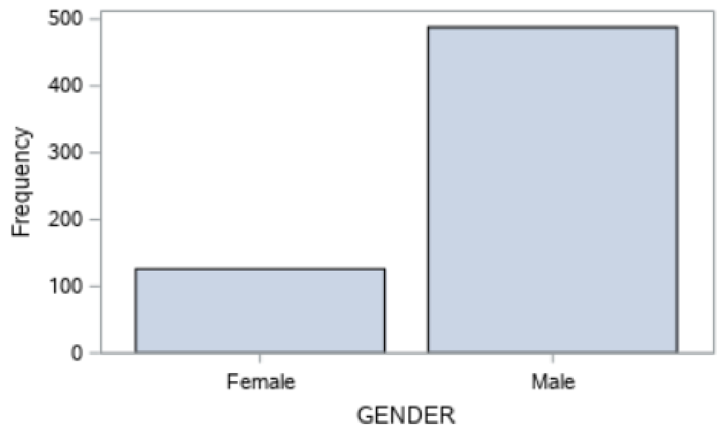
Figure 1 Univariate Analysis of the Categorical variable: Gender

The FREQ Procedure

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	126	20.52	126	20.52
Male	488	79.48	614	100.00

----End----

Figure 2 Univariate Analysis of the Categorical variable: Gender



----End----

2.6.3 Explanation

In Figure 1 shows there are two categorical variable of GENDER which are According to Figures 1 and 2, the frequency of Male is higher than females. Besides that, Figure 1 shows Male have a high percentage cumulative frequency and percentage than females.

2.7 Univariate Analysis of the categorical variable: LOAN_LOCATION

2.7.1 SAS Codes

```
TITLE1 'Figure 3 Univariate Analysis of the Categorical variable: LOAN_LOCATION';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;
TABLE LOAN_LOCATION;
RUN;

/* This code display a barchart of LOAN_LOCATION*/
ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
VBAR LOAN_LOCATION;
Title 'Figure 4 Univariate Analysis of the Categorical variable: LOAN_LOCATION';
RUN;
```

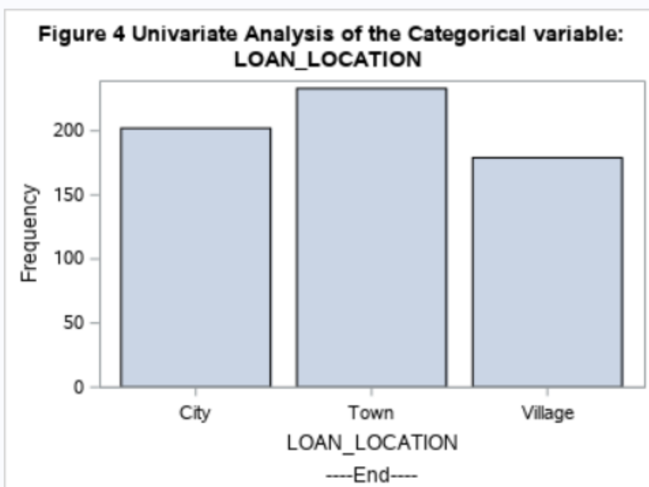
2.7.2 Screenshot(s)/Output(s)

Figure 3 Univariate Analysis of the Categorical variable: LOAN_LOCATION

The FREQ Procedure

LOAN_LOCATION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
City	202	32.90	202	32.90
Town	233	37.95	435	70.85
Village	179	29.15	614	100.00

----End----



2.7.3 Explanation

In Figure3 shows there are three categorical variables of LOAN_LOCATION which are City, Town and village. From the figures above, the LOAN_LOCATION variable has a high frequency at Town, followed by City and then the Village.

2.8 Univariate Analysis of the categorical variable: MARITAL_STATUS

2.8.1 SAS Codes

```
TITLE1 'Figure 5 Univariate Analysis of the Categorical variable: MARITAL_STATUS';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;
TABLE MARITAL_STATUS;
RUN;

/* This code display a barchart of MARITAL_STATUS*/
ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
VBAR MARITAL_STATUS;
Title 'Figure 6 Univariate Analysis of the Categorical variable: MARITAL_STATUS';
RUN;
```

2.8.2 Screenshot(s)/Output(s)

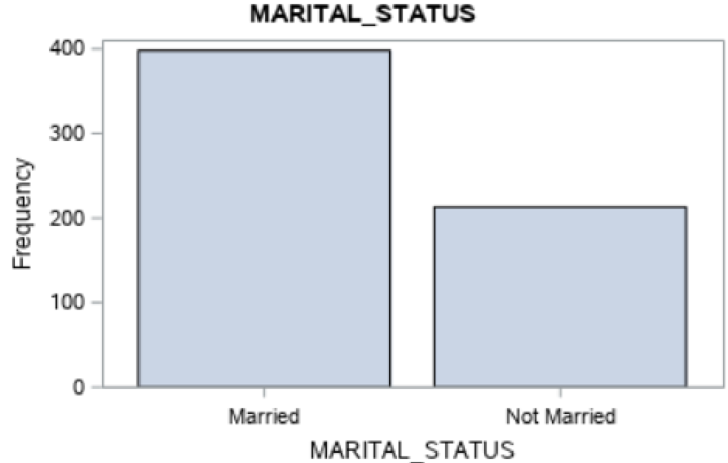
Figure 5 Univariate Analysis of the Categorical variable: MARITAL_STATUS

The FREQ Procedure

MARITAL_STATUS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Married	398	65.14	398	65.14
Not Married	213	34.86	611	100.00
Frequency Missing = 3				

----End----

Figure 6 Univariate Analysis of the Categorical variable: MARITAL_STATUS



----End----

2.8.3 Explanation

The Figure 5 shows there are two categorical variables of MARITAL_STATUS which are married and not married. In this analysis, we encounter some issue. The Figure 5 shows among 614 observations there are 3 observations is missing.

2.9 Univariate Analysis of the categorical variable: QUALIFICATION

2.9.1 SAS Codes

```
TITLE1 'Figure 7 Univariate Analysis of the Categorical variable: QUALIFICATION';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;
TABLE QUALIFICATION;
RUN;

/* This code display a barchart of QUALIFICATION*/
ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
VBAR QUALIFICATION;
Title 'Figure 8 Univariate Analysis of the Categorical variable: QUALIFICATION';
RUN;
```

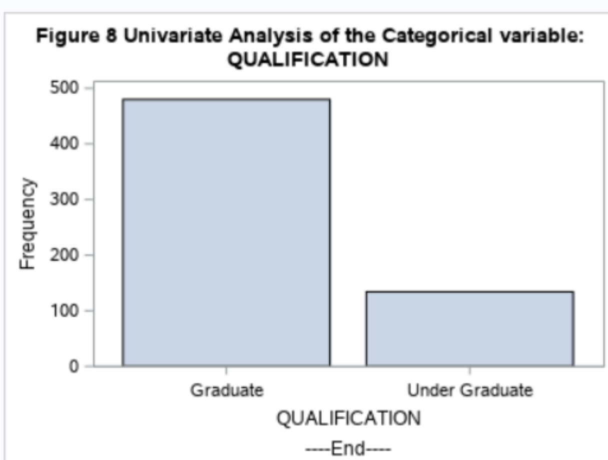
2.9.2 Screenshot(s)/Output(s)

Figure 7 Univariate Analysis of the Categorical variable: QUALIFICATION

The FREQ Procedure

QUALIFICATION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Graduate	480	78.18	480	78.18
Under Graduate	134	21.82	614	100.00

----End----



2.9.3 Explanation

Figure 7 shows there are two categorical variables of QUALIFICATION: graduate and undergraduate. It shows the graduate has a high frequency than the undergraduate. Followed by percent and cumulative percent. But low cumulative frequency than undergraduate.

2.10 Univariate Analysis of the categorical variable: LOAN_HISTORY

2.10.1 SAS Codes

```
TITLE1 'Figure 9 Univariate Analysis of the Categorical variable: LOAN_HISTORY';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;
TABLE LOAN_HISTORY;
RUN;

/* This code display a barchart of LOAN_HISTORY*/
ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
VBAR LOAN_HISTORY;
Title 'Figure 10 Univariate Analysis of the Categorical variable: LOAN_HISTORY';
RUN;
```

2.10.2 Screenshot(s)/Output(s)

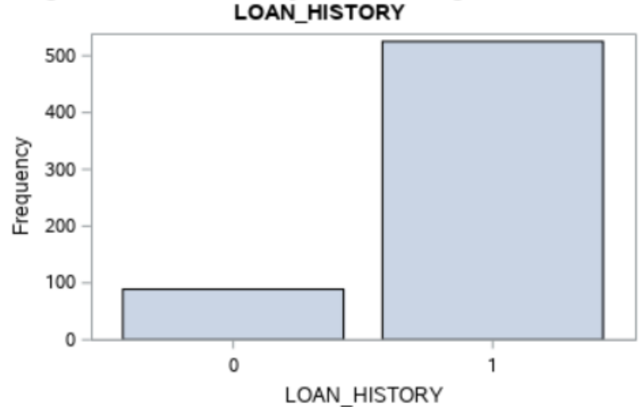
Figure 9 Univariate Analysis of the Categorical variable: LOAN_HISTORY

The FREQ Procedure

LOAN_HISTORY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	89	14.50	89	14.50
1	525	85.50	614	100.00

----End----

Figure 10 Univariate Analysis of the Categorical variable: LOAN_HISTORY



----End----

2.10.3 Explanation

Figure 9 shows there are two categorical variables of LOAN_HISTORY: two numeric values, 0 and 1. It shows that LOAN_HISTORY at 0 has 14.50 % that might have loan, not yet settle. In addition, the LOAN_HISTORY at 1 shows 85.5% settled their loan payment.

2.11 Univariate Analysis of the continuous variable: LOAN_APPROVAL_STATUS

2.11.1 SAS Codes

```
TITLE1 'Figure 11 Univariate Analysis of the Categorical variable: LOAN_APPROVAL_STATUS';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;
TABLE LOAN_APPROVAL_STATUS;
RUN;

/* This code display a barchart of LOAN_APPROVAL_STATUS*/
ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
VBAR LOAN_APPROVAL_STATUS;
Title 'Figure 12 Univariate Analysis of the Categorical variable: LOAN_APPROVAL_STATUS';
RUN;
```

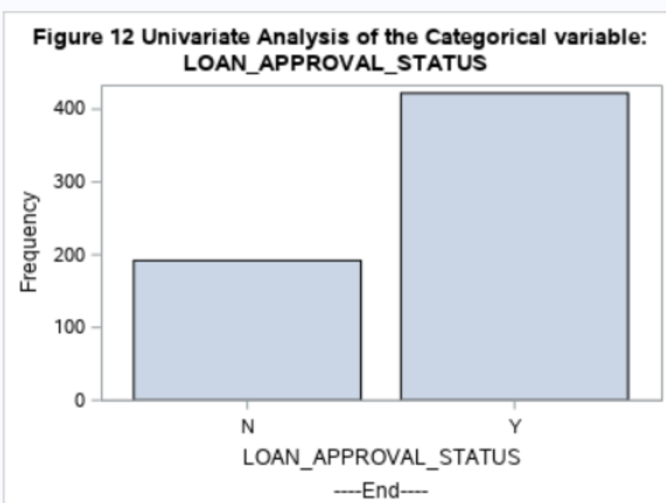
2.11.2 Screenshot(s)/Output(s)

Figure 11 Univariate Analysis of the Categorical variable: LOAN_APPROVAL_STATUS

The FREQ Procedure

LOAN_APPROVAL_STATUS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
N	192	31.27	192	31.27
Y	422	68.73	614	100.00

----End----



2.11.3 Explanation

Figure 11 shows there are two categorical variables of LOAN_APPROVAL_STATUS: two values, N and Y. It shows that at Y have high amounts of LOAN_APPROVAL_STATUS than at N.

2.12 Univariate Analysis of the continuous variable: CANDIDATE_INCOME

2.12.1 SAS Codes

```
TITLE1 'Figure 13 Univariate Analysis of the Categorical variable: CANDIDATE_INCOME';
PROC MEANS DATA = DMLIBMFR.TRAINING_DS N NMISS MIN MAX MEAN MEDIAN STD;
VAR CANDIDATE_INCOME;
RUN;

ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
HISTOGRAM CANDIDATE_INCOME;
Title 'Figure 14 Univariate Analysis of the Categorical variable: CANDIDATE_INCOME';
RUN;
```

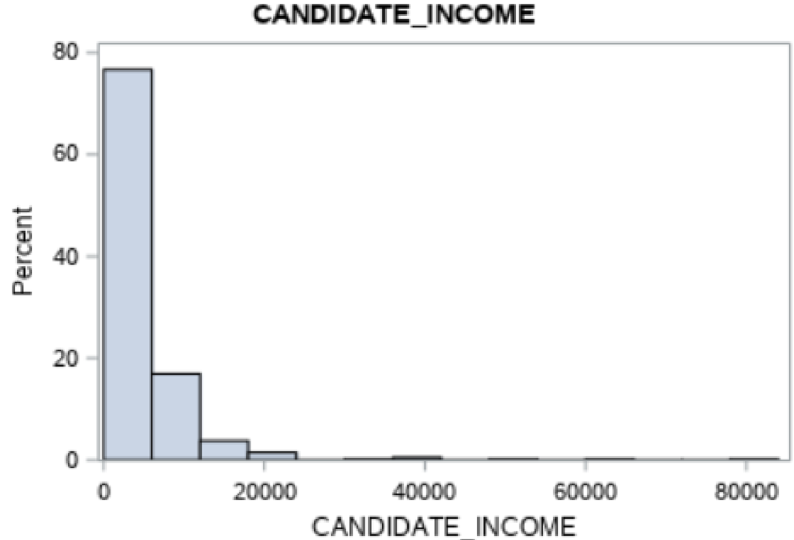
2.12.2 Screenshot(s)/Output(s)

Figure 13 Univariate Analysis of the Categorical variable: CANDIDATE_INCOME

The MEANS Procedure

Analysis Variable : CANDIDATE_INCOME						
N	N Miss	Minimum	Maximum	Mean	Median	Std Dev
614	0	150.0000000	81000.00	5403.46	3812.50	6109.04

Figure 14 Univariate Analysis of the Categorical variable: CANDIDATE_INCOME



2.12.3 Explanation

There is no missing value from the observations. The bar chart shows the CANDIDATE_INCOME shows Min value, Max value, Mean, Median and standard deviation in Figure 13.

2.13 Univariate Analysis of the continuous variable: LOAN_AMOUNT

2.13.1 SAS Codes

```
TITLE1 'Figure 15 Univariate Analysis of the Categorical variable: LOAN_AMOUNT';
PROC MEANS DATA = DMLIBMFR.TRAINING_DS N NMISS MIN MAX MEAN MEDIAN STD;
VAR LOAN_AMOUNT;
RUN;

ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
HISTOGRAM LOAN_AMOUNT;
Title 'Figure 16 Univariate Analysis of the Categorical variable: LOAN_AMOUNT';
RUN;
```

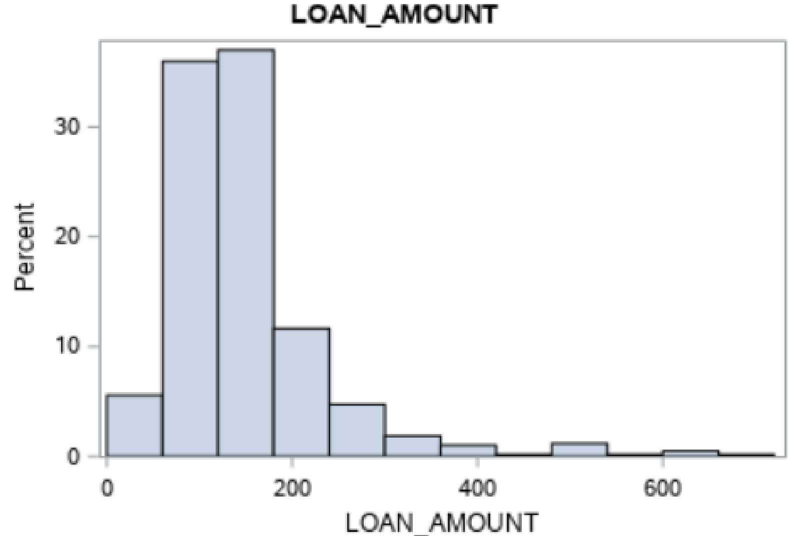
2.13.2. Screenshot(s)/Output(s)

Figure 15 Univariate Analysis of the Categorical variable: LOAN_AMOUNT

The MEANS Procedure

Analysis Variable : LOAN_AMOUNT						
N	N Miss	Minimum	Maximum	Mean	Median	Std Dev
592	22	9.0000000	700.0000000	146.4121622	128.0000000	85.5873252

Figure 16 Univariate Analysis of the Categorical variable: LOAN_AMOUNT



2.13.3 Explanation

This variable shows twenty-two (22) observations from 614 observations are missing in the LOAN_AMOUNT variable. Same thing happened at MARITAL_STATUS variable but different number of missing values.

2.14 Univariate Analysis of the continuous variable: LOAN_DURATION

2.14.1 SAS Codes

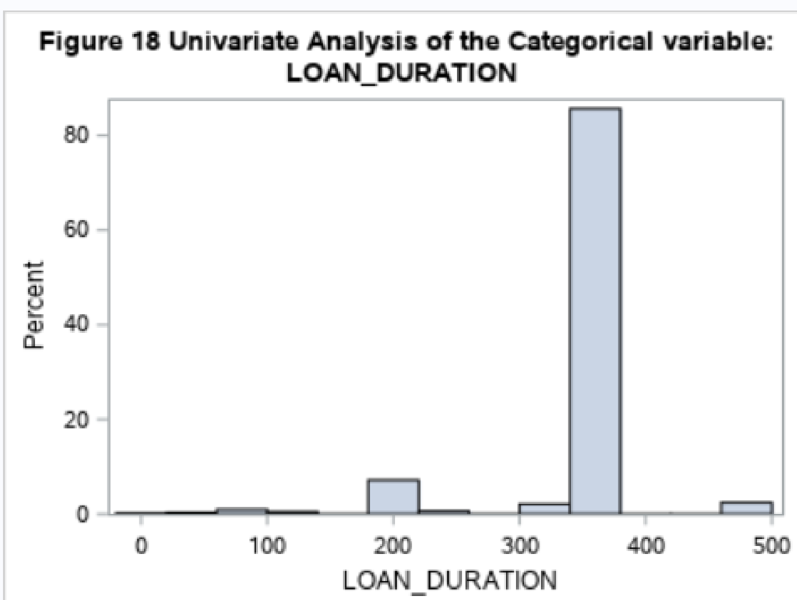
```
TITLE1 'Figure 17 Univariate Analysis of the Categorical variable: LOAN_DURATION';
PROC MEANS DATA = DMLIBMFR.TRAINING_DS N NMISS MIN MAX MEAN MEDIAN STD;
VAR LOAN_DURATION;
RUN;

ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
HISTOGRAM LOAN_DURATION;
Title 'Figure 18 Univariate Analysis of the Categorical variable: LOAN_DURATION';
RUN;
```

2.14.2 Screenshot(s)/Output(s)

Figure 17 Univariate Analysis of the Categorical variable: LOAN_DURATION

The MEANS Procedure

Analysis Variable : LOAN_DURATION						
N	N Miss	Minimum	Maximum	Mean	Median	Std Dev
614	0	12.0000000	480.0000000	342.4104235	360.0000000	64.4286291



2.14.3 Explanation

In this variable, there is no missing value. Since this variable also a numeric variable, it clearly shows the minimum, maximum, mean, median and standard deviation value.

2.15 Bivariate Analysis of the variables found in the DMLIBMFR.TRAINING_DS.

2.15.1 Bivariate Analysis of the variables (Categorical vs Categorical) or (Categorical vs Numeric) or both same categorical found in the DMLIBMFR.TRAINING_DS.

2.15.1 Bivariate Analysis of the variables (LOAN_LOCATION – categorical variable vs LOAN_APPROVAL_STATUS – categorical variable) found in the DMLIBMFR.TRAINING_DS.

2.15.1. SAS Codes

```
TITLE1 'Figure 19 Bivariate Analysis of the variables:)';
TITLE2 ' ( LOAN_location - Categorical variable vs LOAN_APPROVAL_STATUS - Categorical variable ) ';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;

TABLE  LOAN_LOCATION * LOAN_APPROVAL_STATUS /
PLOTS = FREQPLOT( TWOWAY = STACKED SCALE = GROUPPCT );
RUN;
```

2.15.2 Screenshot(s)/Output(s)

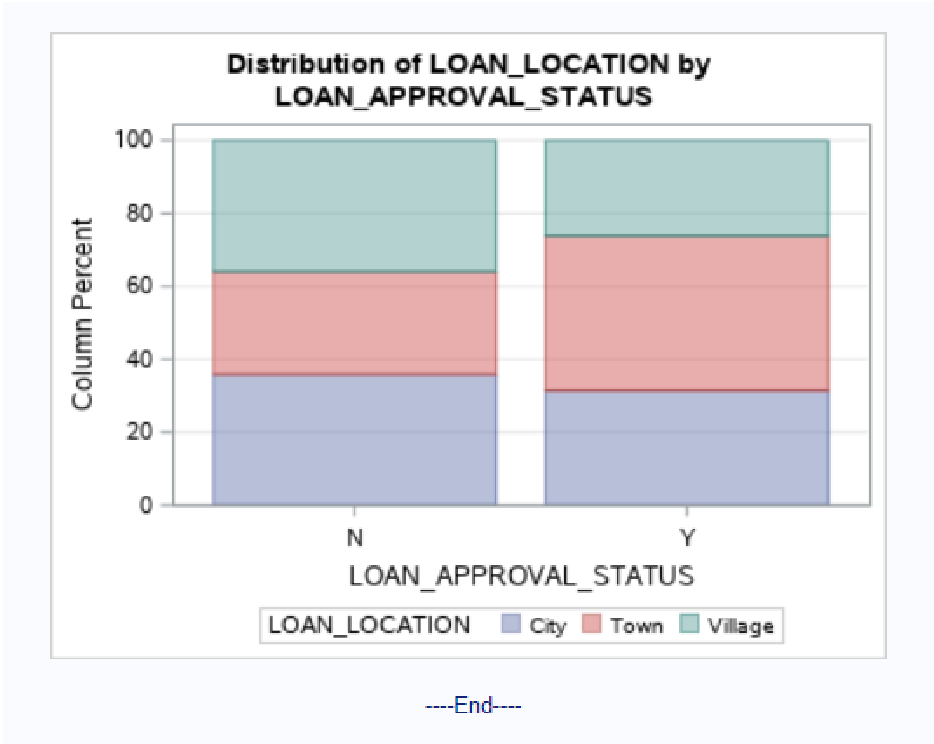
Figure 19 Bivariate Analysis of the variables:)
(LOAN_location - Categorical variable vs LOAN_APPROVAL_STATUS - Categorical variable)

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of LOAN_LOCATION by LOAN_APPROVAL_STATUS

	LOAN_APPROVAL_STATUS		
LOAN_LOCATION	N	Y	Total
City	69 11.24 34.16 35.94	133 21.66 65.84 31.52	202 32.90
Town	54 8.79 23.18 28.13	179 29.15 76.82 42.42	233 37.95
Village	69 11.24 38.55 35.94	110 17.92 61.45 26.07	179 29.15
Total	192 31.27	422 68.73	614 100.00



2.15.3 Explanation

This analysis shows the correlation between LOAN_LOCATION and LOAN_APPROVAL_STATUS. This analysis wants to show the impacts of LOAN_LOCATION on LOAN_APPROVAL_STATUS. It tells us how many applicants get approval and not approval based on different types of locations.

2.16 Bivariate Analysis of the variables (GENDER– categorical variable vs LOAN_APPROVAL_STATUS – categorical variable) found in the DMLIBMFR.TRAINING_DS.

2.16.1 SAS Codes

```
TITLE1 'Figure 20 Bivariate Analysis of the variables:)';
TITLE2 ' ( GENDER - Categorical variable vs LOAN_APPROVAL_STATUS - Categorical variable ) ';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;

TABLE  GENDER * LOAN_APPROVAL_STATUS /
PLOTS = FREQPLOT( TWOWAY = STACKED SCALE = GROUPPCT );
RUN;
```

2.16.2 Screenshot(s)/Output(s)

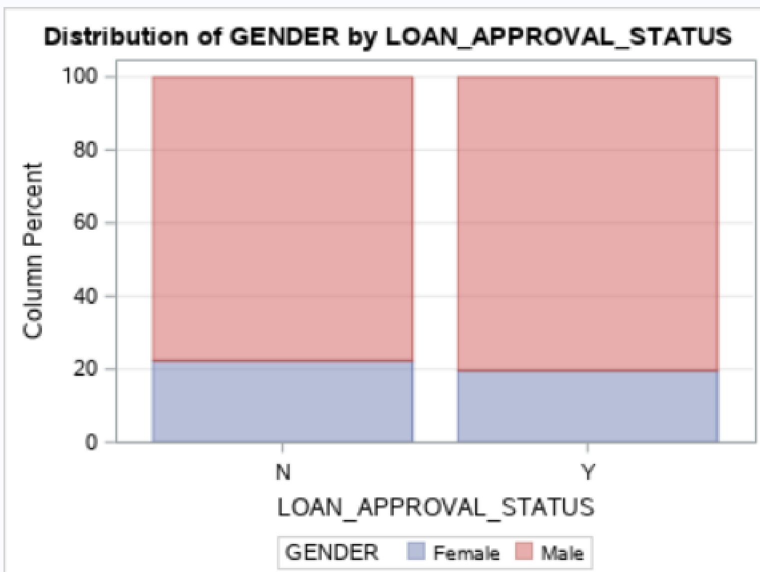
Figure 20 Bivariate Analysis of the variables:)
(GENDER - Categorical variable vs LOAN_APPROVAL_STATUS - Categorical variable)

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of GENDER by LOAN_APPROVAL_STATUS

	LOAN_APPROVAL_STATUS		
GENDER	N	Y	Total
Female	43 7.00 34.13 22.40	83 13.52 65.87 19.67	126 20.52
Male	149 24.27 30.53 77.60	339 55.21 69.47 80.33	488 79.48
Total	192 31.27	422 68.73	614 100.00



----End----

2.16.3 Explanation

In this relationship, it shows the total of Male and Female get approval for the loan. From this analysis shows that Male have more approval (Y) and rejected (N) than Female.

2.17 Bivariate Analysis of the variables (QUALIFICATION– categorical variable vs LOAN_APPROVAL_STATUS – categorical variable) found in the LAPPDK.TRAINING_DS.

2.17.1 SAS Codes

```
TITLE1 'Figure 21 Bivariate Analysis of the variables:)';
TITLE2 ' ( QUALIFICATION - Categorical variable vs LOAN_APPROVAL_STATUS  - Categorical variable ) ';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;

TABLE  QUALIFICATION * LOAN_APPROVAL_STATUS /
PLOTS = FREQPLOT( TWOWAY = STACKED SCALE = GROUPPCT );
RUN;
```

2.17.2 Screenshot(s)/Output(s)

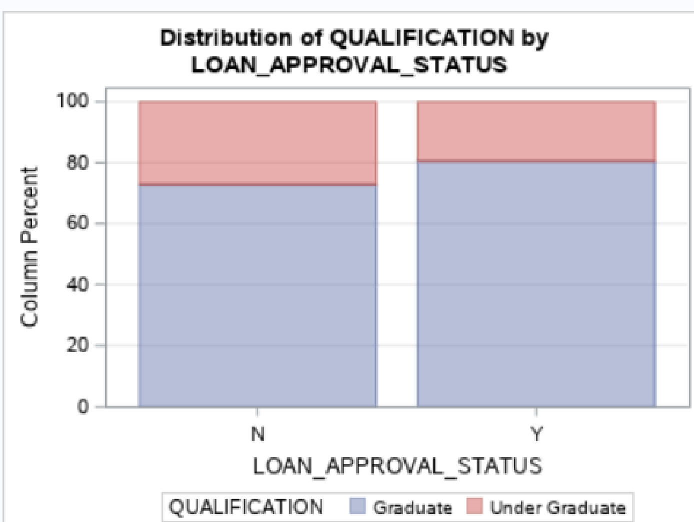
Figure 21 Bivariate Analysis of the variables:)
(QUALIFICATION - Categorical variable vs LOAN_APPROVAL_STATUS - Categorical variable)

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of QUALIFICATION by LOAN_APPROVAL_STATUS

QUALIFICATION	LOAN_APPROVAL_STATUS N	LOAN_APPROVAL_STATUS Y	Total
Graduate	140 22.80 29.17 72.92	340 55.37 70.83 80.57	480 78.18
Under Graduate	52 8.47 38.81 27.08	82 13.36 61.19 19.43	134 21.82
Total	192 31.27	422 68.73	614 100.00



----End----

2.17.3 Explanation

This section shows the relationship between the QUALIFICATION and the LOAN_APPROVAL_STATUS. The graduate has 340 for Y and 140 for N, but undergraduate has 82 for Y and 52 for N. It shows that approval and rejected status is high at Graduate than Undergraduate.

2.18 Bivariate Analysis of the variables (LOAN_LOCATION vs CANDIDATE_INCOME) found in the DMLIBMFR.TRAINING_DS.

2.18.1 SAS Codes

```
TITLE1 'Figure 22 Bivariate Analysis of the variables:)';
TITLE2 ' ( LOAN_LOCATION - Categorical variable vs CANDIDATE_INCOME  - Numeric variable ) ';
FOOTNOTE '----End----';

PROC MEANS DATA = DMLIBMFR.TRAINING_DS;

CLASS LOAN_LOCATION; /* It is a Categorical variable */
VAR CANDIDATE_INCOME; /* Numeric variable : Continous variable */
RUN;

PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;

VBOX CANDIDATE_INCOME / CATEGORY = LOAN_LOCATION;
/* LOAN_LOCATION X-AXIS CANDIDATE_INCOME Y-AXIS */
TITLE1 'Figure 23 Bivariate Analysis of variables LOAN_LOCATION vs CANDIDATE_INCOME';
RUN;
```

2.18.2 Screenshot(s)/Output(s)

Figure 22 Bivariate Analysis of the variables:)
(LOAN_LOCATION - Categorical variable vs CANDIDATE_INCOME - Numeric variable)

The MEANS Procedure

Analysis Variable : CANDIDATE_INCOME						
LOAN_LOCATION	N Obs	N	Mean	Std Dev	Minimum	Maximum
City	202	202	5398.25	6392.93	416.0000000	63337.00
Town	233	233	5292.26	5279.63	210.0000000	39999.00
Village	179	179	5554.08	6782.66	150.0000000	81000.00

----End----

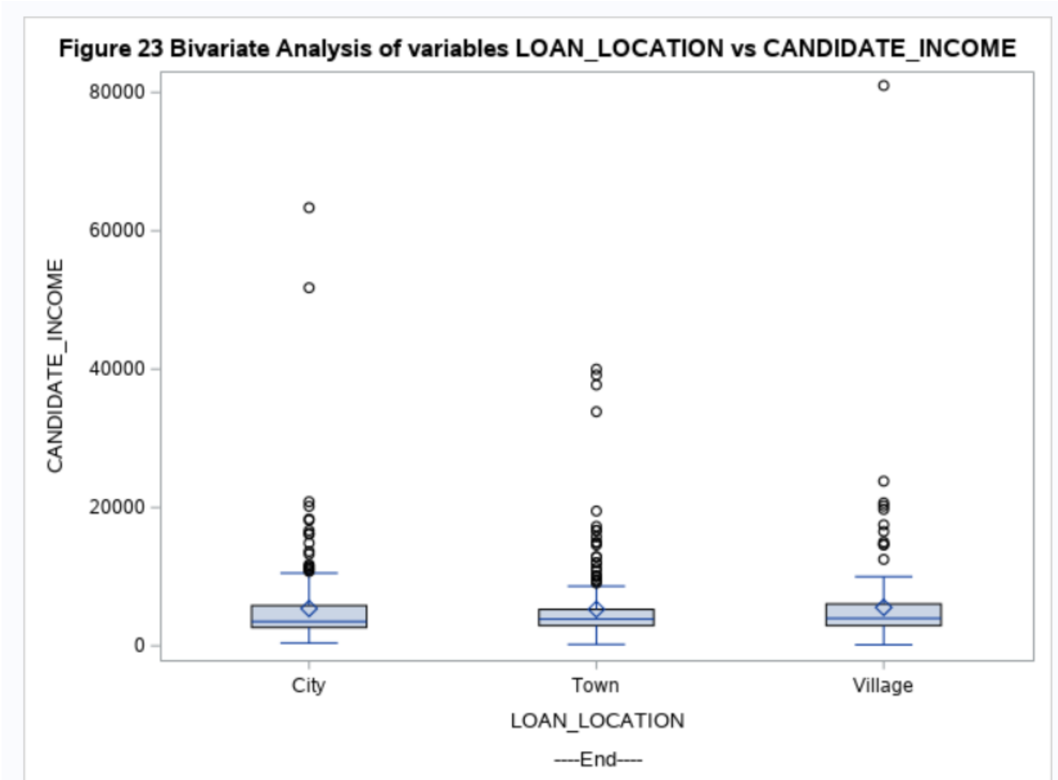


Figure 23 Bivariate Analysis of variables LOAN_LOCATION vs CANDIDATE_INCOME

2.18.3 Explanation

This time we want to observe the impacts of LOAN_LOCATION on CANDIDATE_INCOME. It clearly differentiates the CANDIDATE_INCOME variable based on the LOAN_LOCATION variable. The village shows high maximum in CANDIDATE_INCOME compared to the others location like town and city.

2.19 Bivariate Analysis of the variables (MARITAL_STATUS vs CANDIDATE_INCOME) found in the DMLIBMFR.TRAINING_DS.

2. 19.1 SAS Codes

```
TITLE1 'Figure 24 Bivariate Analysis of the variables:)';
TITLE2 ' ( MARITAL_STATUS - Categorical variable vs CANDIDATE_INCOME  - Continous ) ';
FOOTNOTE '----End----';

PROC MEANS DATA = DMLIBMFR.TRAINING_DS;

CLASS MARITAL_STATUS; /* It is a Categorical variable */
VAR CANDIDATE_INCOME; /* Numeric variable */
RUN;

PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;

VBOX CANDIDATE_INCOME / CATEGORY = MARITAL_STATUS;
/* MARITAL_STATUS X-AXIS CANDIDATE_INCOME Y-AXIS */
TITLE1 'Figure 25 Bivariate Analysis of variables MARITAL_STATUS vs CANDIDATE_INCOME';
RUN;
```

2. 19.2 Screenshot(s)/Output(s)

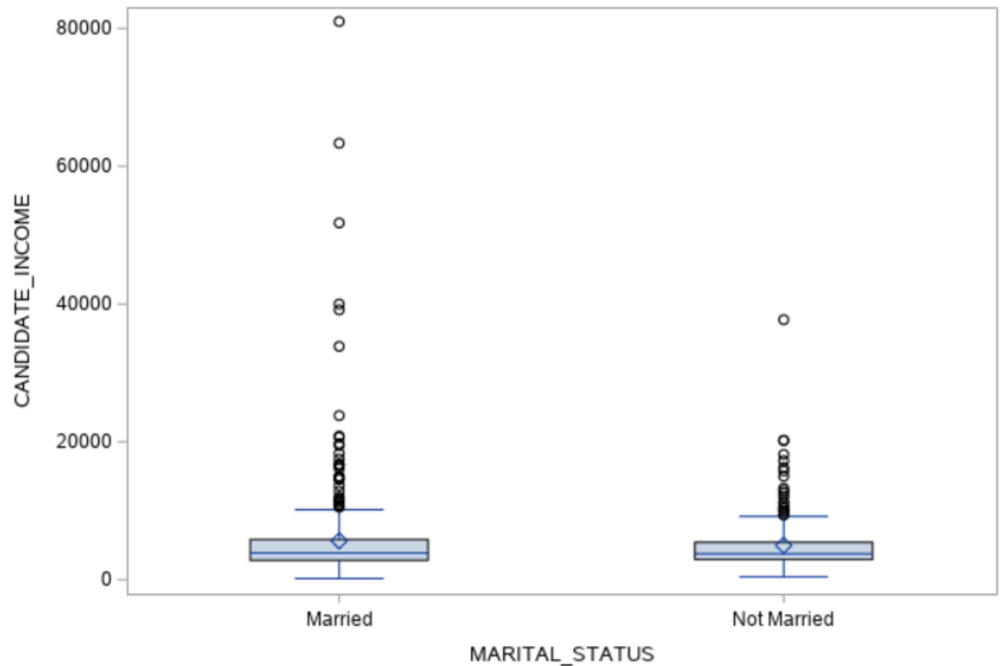
Figure 24 Bivariate Analysis of the variables:)
(MARITAL_STATUS - Categorical variable vs CANDIDATE_INCOME - Continous)

The MEANS Procedure

Analysis Variable : CANDIDATE_INCOME						
MARITAL_STATUS	N Obs	N	Mean	Std Dev	Minimum	Maximum
Married	398	398	5629.17	6989.25	150.0000000	81000.00
Not Married	213	213	4970.38	4004.33	416.0000000	37719.00

----End----

Figure 25 Bivariate Analysis of variables MARITAL_STATUS vs CANDIDATE_INCOME



----End----

2. 19.3 Explanation

This time we want to observe the impacts of MARITAL_STATUS on CANDIDATE_INCOME. There are no missing values. The married applicants have high income compared to not married applicants.

2.20 Imputing (replacing) missing values

2.20.1 Imputing missing values found in the variables:

2.20.2 Make a copy of the dataset: DMLIBMFR.TRAINING_DS

2.20.2 SAS Codes

```
/* Make a back-up copy of the DMLIBMFR.TRAINING_DS before do cleansing treatment or data cleaning*/

TITLE1 'Make a back-up copy of the DMLIBMFR.TRAINING_DS';

PROC SQL;

CREATE TABLE DMLIBMFR.TRAINING_DS_BK AS
SELECT *
FROM DMLIBMFR.TRAINING_DS;

QUIT;
```

2.21.1 Before Imputing ….List The Observations With Missing Values In The MARITAL_STATUS Variable

2.21.1 SAS Codes

```
TITLE1 'Before imputing missing values';
TITLE2 'STEP 1: List the observation(s) with the missing values in the MARITAL_STATUS';
FOOTNOTE '----End----';

PROC SQL;

SELECT *
FROM DMLIBMFR.TRAINING_DS
WHERE ( ( marital_status IS NULL ) OR
        ( marital_status IS MISSING ) OR
        ( marital_status EQ '' ) );

QUIT;
```

2.21.2 Screenshot(s)/Output(s)

Before imputing missing values
STEP 1: List the observation(s) with the missing values in the MARITAL_STATUS

SME_LOAN_ID_NO	GENDER	MARITAL_STATUS	QUALIFICATION	CANDIDATE_INCOME	LOAN_AMOUNT	LOAN_DURATION	LOAN_HISTORY	LOAN_LOCATION	LOAN_APPROVAL_STATUS
LP001357	Male		Graduate	3816	160	360	1	City	Y
LP001760	Male		Graduate	4758	158	480	1	Town	Y
LP002393	Female		Graduate	10047	.	240	1	Town	Y

----End----

2.21.3 Explanation

Before imputing or replacing the missing values, the first step is to list the observation(s) with the missing values in the MARITAL_STATUS.

2.21.4 SAS Codes

```
TITLE1 'STEP 2: Count the number of observations with missing values in the MARITAL_STATUS';
FOOTNOTE '----End----';

PROC SQL;

SELECT COUNT(*) LABEL = 'Number of observations'
FROM DMLIBMFR.TRAINING_DS
WHERE ( ( marital_status IS NULL ) OR
        ( marital_status IS MISSING ) OR
        ( marital_status EQ '' ) );

QUIT;
```

2.21.5 Screenshot(s)/Output(s)

STEP 2: Count the number of observations with missing values in the MARITAL_STATUS

Number of observations
3

----End----

2.21.6 Explanation

Then, for the second step is to count the number of observations with missing values in the MARITAL_STATUS.

2.21.7 Create a dataset to hold the MARITAL_STATUS and the number of applicants.

2.21.7 SAS Codes

```
TITLE1 'STEP 3: Create a small dataset to hold or keep the intermediate results of marital_status & its counts.';
FOOTNOTE '----End----';

PROC SQL;

CREATE TABLE DMLIBMFR.TRAINING_DS_MARITAL_STATUS AS
SELECT marital_status, COUNT(*) AS COUNTS
FROM DMLIBMFR.TRAINING_DS
WHERE ( ( marital_status IS NOT NULL ) OR
        ( marital_status IS NOT MISSING ) OR
        ( marital_status NE '' ) )
GROUP BY marital_status;

QUIT;
```

2.21.8 Explanation

Then, for the third step is to create a small dataset to hold or keep the intermediate results of MARITAL_STATUS and the number of counts of applicants.

2.21.9 SAS Codes

```
TITLE1 'STEP 4: Display the contents of the dataset - DMLIBMFR.TRAINING_DS_MARITAL_STATUS';
FOOTNOTE '----End----';

PROC SQL;

SELECT *
FROM DMLIBMFR.TRAINING_DS_MARITAL_STATUS;

QUIT;
```

2.21.10 Screenshot(s)/Output(s)

STEP 4: Display the contents of the dataset - DMLIBMFR.TRAINING_DS_MARITAL_STATUS

MARITAL_STATUS	COUNTS
Married	398
Not Married	213

----End----

2.21.11 Explanation

Then, for the fourth step is to display the contents or the observations inside the dataset - DMLIBMFR.TRAINING_DS_MARITAL_STATUS. In the created tiny dataset shows there are two observations, Married and Not Married.

2.21.12 SAS Codes

```
TITLE1 'STEP 5: Find the MOD and impute the missing values found in the dataset DMLIBMFR.TRAINING_DS';
FOOTNOTE '----End----';

PROC SQL;

UPDATE DMLIBMFR.TRAINING_DS
SET marital_status = ( SELECT marital_status Label = 'MOD_MARITAL_STATUS'
                       FROM DMLIBMFR.TRAINING_DS_MARITAL_STATUS
                       WHERE ( counts EQ ( SELECT MAX(counts) Label = 'Highest Counts'
                                           FROM DMLIBMFR.TRAINING_DS_MARITAL_STATUS ) ) )
                                           /* Above is a sub-program to find the MOD of MARITAL_STATUS*/
WHERE ( ( marital_status IS NULL ) OR
        ( marital_status IS MISSING ) OR
        ( marital_status EQ '' ) );

QUIT;
```

2.21.13 Screenshot(s)/Output(s)

2.21.14 Explanation

Then, for the fifth step is to find the MOD and impute the missing values found in the dataset DMLIBMFR.TRAINING_DS. The sub-program is used to instantly update the MOD of MARITAL_STATUS automatically. Then replaced the MARITAL_STATUS with the outcome of program, Null, Missing, or blank space (‘’).

2.22 STEP 6 & 7: After imputing missing values: list the observations with missing values in MARITAL_STATUS variable

2.22.1 SAS Codes

```
TITLE1 'After imputing missing values';
TITLE2 'STEP 6: List the observation(s) with missing values in the MARITAL_STATUS';
FOOTNOTE '----End----';

PROC SQL;

SELECT *
FROM DMLIBMFR.TRAINING_DS
WHERE ( ( marital_status IS NULL ) OR
        ( marital_status IS MISSING ) OR
        ( marital_status EQ '' ) );

QUIT;
```

2.22.2 Screenshot(s)/Output(s)

After imputing missing values
STEP 6: List the observation(s) with missing values in the MARITAL_STATUS

----End----

2.22.3 Explanation

This section shows the after process of cleansing the MARITAL_STATUS variable. The results show an empty or blank, meaning the data cleansing is succeeded.

2.22.4 SAS Codes

```
TITLE1 'STEP 7: Count the number of observations with missing values in the MARITAL_STATUS';
FOOTNOTE '----End----';

PROC SQL;

SELECT COUNT(*) LABEL = 'Number of observations'
FROM DMLIBMFR.TRAINING_DS
WHERE ( ( marital_status IS NULL ) OR
        ( marital_status IS MISSING ) OR
        ( marital_status EQ '' ) );

QUIT;
```

2.22.2 Screenshot(s)/Output(s)

STEP 7: Count the number of observations with missing values in the MARITAL_STATUS

Number of observations
0

----End----

2.22.3 Explanation

The number of observations turns zero. Meaning there is no missing values detected anymore.

Since its already being replaced on previous program.

2.23 After imputation: Univariate Analysis of the categorical variable: MARITAL_STATUS

2.23.1 SAS Codes

```
TITLE1 'Figure 26 After Imputation: Univariate Analysis of the Categorical variable: MARITAL_STATUS';
FOOTNOTE '----End----';

PROC FREQ DATA = DMLIBMFR.TRAINING_DS;
TABLE MARITAL_STATUS;
RUN;

ODS GRAPHICS / RESET WIDTH = 4.0 IN HEIGHT = 3.0 IN IMAGEMAP;
PROC SGPLOT DATA = DMLIBMFR.TRAINING_DS;
VBAR MARITAL_STATUS;
Title 'Figure 27 After Imputation: Univariate Analysis of the Categorical variable: MARITAL_STATUS';
RUN;
```

2.23.1 Screenshot(s)/Output(s)

Figure 26 After Imputation: Univariate Analysis of the Categorical variable: MARITAL_STATUS

The FREQ Procedure

MARITAL_STATUS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Married	401	65.31	401	65.31
Not Married	213	34.69	614	100.00

----End----

Figure 27 After Imputation: Univariate Analysis of the Categorical variable: MARITAL_STATUS

2.23. 3 Explanation

Previously, there are 3 missing values. Now after the imputation process, the missing values is clear and clean from missing values.

2.24 Impute missing values found in the variable – LOAN_AMOUNT

2.24.1 SAS Codes

```
TITLE1 'STEP 1: Make a back-up copy of the dataset - DMLIBMFR.TRAINING_DS';
FOOTNOTE '----End----';

PROC SQL;

CREATE TABLE DMLIBMFR.TRAINING_DS_BK AS
SELECT *
FROM DMLIBMFR.TRAINING_DS;

QUIT;
```

```
TITLE1 'STEP 2: ( Before imputation ) List the observations with missing values in the variable: LOAN_AMOUNT ';
FOOTNOTE '----End----';

PROC SQL;

SELECT *
FROM DMLIBMFR.TRAINING_DS
WHERE ( ( loan_amount EQ . ) OR
        ( loan_amount IS MISSING ) );

QUIT;
```

2.24.2 Screenshot(s)/Output(s)

STEP 2: (Before imputation) List the observations with missing values in the variable: LOAN_AMOUNT

SME_LOAN_ID_NO	GENDER	MARITAL_STATUS	QUALIFICATION	CANDIDATE_INCOME	LOAN_AMOUNT	LOAN_DURATION	LOAN_HISTORY	LOAN_LOCATION	LOAN_APPROVAL_STATUS
LP001002	Male	Not Married	Graduate	5849	.	360	1	City	Y
LP001106	Male	Married	Graduate	2275	.	360	1	City	Y
LP001213	Male	Married	Graduate	4945	.	360	0	Village	N
LP001266	Male	Married	Graduate	2395	.	360	1	Town	Y
LP001326	Male	Not Married	Graduate	6782	.	360	1	City	N
LP001350	Male	Married	Graduate	13650	.	360	1	City	Y
LP001356	Male	Married	Graduate	4652	.	360	1	Town	Y
LP001392	Female	Not Married	Graduate	7451	.	360	1	Town	Y
LP001449	Male	Not Married	Graduate	3865	.	360	1	Village	Y
LP001682	Male	Married	Under Graduate	3992	.	180	1	City	N
LP001922	Male	Married	Graduate	20667	.	360	1	Village	N
LP001990	Male	Not Married	Under Graduate	2000	.	360	1	City	N
LP002054	Male	Married	Under Graduate	3601	.	360	1	Village	Y
LP002113	Female	Not Married	Under Graduate	1830	.	360	0	City	N
LP002243	Male	Married	Under Graduate	3010	.	360	0	City	N
LP002393	Female	Married	Graduate	10047	.	240	1	Town	Y
LP002401	Male	Married	Graduate	2213	.	360	1	City	Y
LP002533	Male	Married	Graduate	2947	.	360	1	City	N
LP002697	Male	Not Married	Graduate	4680	.	360	1	Town	N
LP002778	Male	Married	Graduate	6633	.	360	0	Village	N
LP002784	Male	Married	Under Graduate	2492	.	360	1	Village	Y
LP002960	Male	Married	Under Graduate	2400	.	180	1	City	N

----End----

2.24.3 Explanation

Before imputation, make a back-up copy of the dataset - DMLIBMFR.TRAINING_DS and list the observations with missing values in the variable: LOAN_AMOUNT.

2.24.4 SAS Codes

```
TITLE1 'STEP 3: ( Before imputation ) Number of observations with missing values in the variable: LOAN_AMOUNT ';
FOOTNOTE '----End----';

PROC SQL;

SELECT COUNT(*) Label = 'Number of Observations'
FROM DMLIBMFR.TRAINING_DS
WHERE ( ( loan_amount EQ . ) OR
        ( loan_amount IS MISSING ) );

QUIT;
```

2.24.5 Screenshot(s)/Output(s)

STEP 3: (Before imputation) Number of observations with missing values in the variable: LOAN_AMOUNT

Number of Observations
22

----End----

2.24.6 Explanation

Then, cross check to confirm that there are 22 missing values or observations in LOAN_AMOUNT variable.

2.24.7 SAS Codes

```
TITLE1 'STEP 4: Impute the missing values found in the variable - LOAN_AMOUNT';
FOOTNOTE '---End---';

PROC STDIZE DATA = DMLIBMFR.TRAINING_DS REPONLY

METHOD = MEAN OUT = DMLIBMFR.TRAINING_DS;
var LOAN_AMOUNT;

QUIT;
```

2.24.8 Screenshot(s)/Output(s)

Total rows: 614 Total columns: 10 Rows 1-100

	SME_LOAN_ID...	GEND...	MARITAL_STA...	QUALIFICATION	CANDIDATE_INCOME	LOAN_AMOUNT	LOAN_DURATION
1	LP001002	Male	Not Married	Graduate	5849	146.41216216	360
2	LP001003	Male	Married	Graduate	4583	128	360
3	LP001005	Male	Married	Graduate	3000	66	360
4	LP001006	Male	Married	Under Graduate	2583	120	360
5	LP001008	Male	Not Married	Graduate	6000	141	360
6	LP001011	Male	Married	Graduate	5417	267	360
7	LP001013	Male	Married	Under Graduate	2333	95	360
8	LP001014	Male	Married	Graduate	3036	158	360
9	LP001018	Male	Married	Graduate	4006	168	360
10	LP001020	Male	Married	Graduate	12841	349	360
11	LP001024	Male	Married	Graduate	3200	70	360
12	LP001027	Male	Married	Graduate	2500	109	360

2.24.9 Explanation

In the outputs, the missing value in LOAN_AMOUNT variable at first observation has been replaced with suitable value through the program. Since previously, it only shows '.' And now it has a value in it.

2.24.10 SAS Codes

```
TITLE1 'STEP 5: ( After imputation ) List the observations with missing values in the variable: LOAN_AMOUNT ';
FOOTNOTE '----End----';

PROC SQL;

SELECT *
FROM DMLIBMFR.TRAINING_DS
WHERE ( ( loan_amount EQ . ) OR
        ( loan_amount IS MISSING ) );

QUIT;
```

2.24.11 Screenshot(s)/Output(s)

STEP 5: (After imputation) List the observations with missing values in the variable: LOAN_AMOUNT

----End----

2.24.12 Explanation

After imputation is completed, the result shows an empty or blank on observations with missing values in the variable of LOAN_AMOUNT.

2.24.13 SAS Codes

```
TITLE1 'STEP 6: ( After imputation ) Number of observations with missing values in the variable: LOAN_AMOUNT ';
FOOTNOTE '----End----';

PROC SQL;

SELECT COUNT(*) Label = 'Number of Observations'
FROM DMLIBMFR.TRAINING_DS
WHERE ( ( loan_amount EQ . ) OR
        ( loan_amount IS MISSING ) );

QUIT;
```

2.24.14 Screenshot(s)/Output(s)

STEP 6: (After imputation) Number of observations with missing values in the variable: LOAN_AMOUNT

Number of Observations
0

----End----

2.24.15 Explanation

This part shows the number of observations of missing values is zero in the variable of LOAN_AMOUNT.

2.25 Univariate Analysis of the variables found in the dataset DMLIBMFR.TESTING_DS.

2.25.1 Introduction

MACRO is an advanced functions feature from SAS to shorten or minimize the length of codes by write the code once but can call it many times.

2.25.2 SAS Codes

```
/* Univariate Analysis of Variables found in the DMLIBMFR.TESTING_DS using MACRO

MACRO BEGINS HERE */

%MACRO MACRO_UVA_TESTING_DS(PDS_NAME, PVARI_NAME, PTITLE_1, PTITLE_2);

TITLE1 &PTITLE_1; /* PASSING VALUE */
TITLE2 &PTITLE_2;
FOOTNOTE '----End----';

PROC FREQ DATA = &PDS_NAME;

TABLE &PVARI_NAME;

QUIT;

%MEND MACRO_UVA_TESTING_DS;

/* MACRO ENDS HERE */
```

2.25.1 Run the SAS Macro

```
/* CALL/RUN THE SAS MACRO */

%MACRO_UVA_TESTING_DS(DMLIBMFR.TESTING_DS, GENDER, 'UNIVARIATE ANALYSIS', 'OF THE CATEGORICAL VARIABLE - GENDER');
%MACRO_UVA_TESTING_DS(DMLIBMFR.TESTING_DS, MARITAL_STATUS, 'UNIVARIATE ANALYSIS', 'OF THE CATEGORICAL VARIABLE - MARITAL_STATUS');
%MACRO_UVA_TESTING_DS(DMLIBMFR.TESTING_DS, QUALIFICATION, 'UNIVARIATE ANALYSIS', 'OF THE CATEGORICAL VARIABLE - QUALIFICATION');
%MACRO_UVA_TESTING_DS(DMLIBMFR.TESTING_DS, LOAN_LOCATION, 'UNIVARIATE ANALYSIS', 'OF THE CATEGORICAL VARIABLE - LOAN_LOCATION');
%MACRO_UVA_TESTING_DS(DMLIBMFR.TESTING_DS, LOAN_HISTORY, 'UNIVARIATE ANALYSIS', 'OF THE CATEGORICAL VARIABLE - LOAN_HISTORY');
```

2.25.3 Screenshot(s)/Output(s)

UNIVARIATE ANALYSIS
OF THE CATEGORICAL VARIABLE - LOAN_LOCATION

The FREQ Procedure

LOAN_LOCATION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
City	202	32.90	202	32.90
Town	233	37.95	435	70.85
Village	179	29.15	614	100.00

----End----

UNIVARIATE ANALYSIS
OF THE CATEGORICAL VARIABLE - MARITAL_STATUS

The FREQ Procedure

MARITAL_STATUS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Married	233	63.49	233	63.49
Not Married	134	36.51	367	100.00

----End----

UNIVARIATE ANALYSIS
OF THE CATEGORICAL VARIABLE - QUALIFICATION

The FREQ Procedure

QUALIFICATION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Graduate	283	77.11	283	77.11
Under Graduate	84	22.89	367	100.00

----End----

UNIVARIATE ANALYSIS
OF THE CATEGORICAL VARIABLE - LOAN_LOCATION

The FREQ Procedure

LOAN_LOCATION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
City	140	38.15	140	38.15
Town	116	31.61	256	69.75
Village	111	30.25	367	100.00

----End----

UNIVARIATE ANALYSIS
OF THE CATEGORICAL VARIABLE - LOAN_HISTORY

The FREQ Procedure

LOAN_HISTORY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	59	16.08	59	16.08
1	308	83.92	367	100.00

----End----

2.25.4 Explanation

This part shows the result of each minimize code done in SAS Macro for univariate analysis. Since there are 5 SAS Macro codes, so there are 5 results for each of the macro code.

2.26 Bivariate Analysis of the variables found in the DMLIBMFR.TESTING_DS using Macro.

2.26.1 SAS Codes

```
/* Bivariate Analysis of variables found in the - Using MACRO

MACRO BEGINS HERE */

%MACRO MACRO_BVA_CATE_TESTING_DS(PDS_NAME, PVARI_1, PVARI_2, PTITLE_1, PTITLE_2);

TITLE1 &PTITLE_1;
TITLE2 &PTITLE_2;
FOOTNOTE '----End---';

PROC FREQ DATA = &PDS_NAME;

TABLE &PVARI_1 * &PVARI_2 /
PLOTS = FREQPLOT( TWOWAY = STACKED SCALE = GROUPPCT );

RUN;

%MEND MACRO_BVA_CATE_TESTING_DS;

/* MACRO ENDS HERE */
```

```
/* To run/call the MACRO - MACRO_BVA_CATE_TESTING_DS */
%MACRO_BVA_CATE_TESTING_DS(DMLIBMFR.TESTING_DS, LOAN_LOCATION, GENDER, 'BIVARIATE ANALYSIS', 'LOAN_LOCATION - Cate variable vs GENDER - Cate variable');
%MACRO_BVA_CATE_TESTING_DS(DMLIBMFR.TESTING_DS, QUALIFICATION, GENDER, 'BIVARIATE ANALYSIS', 'QUALIFICATION - Cate variable vs GENDER - Cate variable');
```

2.26.2 Screenshot(s)/Output(s)

BIVARIATE ANALYSIS

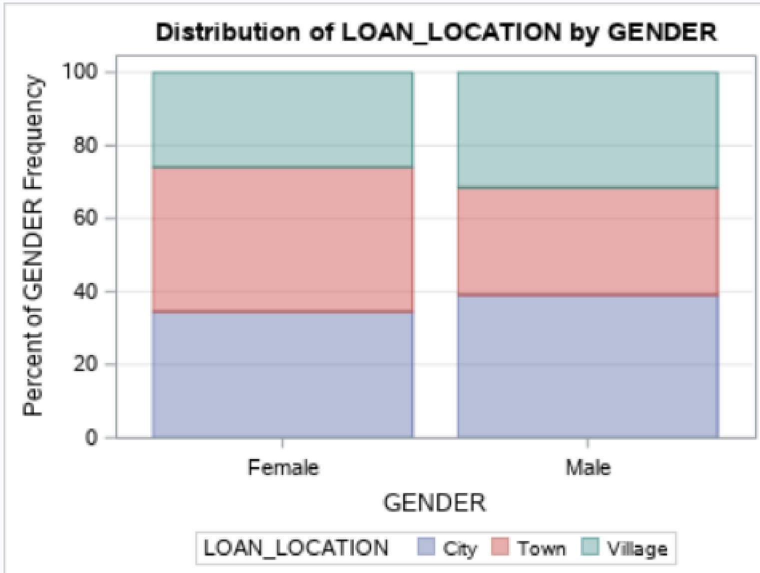
LOAN_LOCATION - Cate variable vs GENDER - Cate variable

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of LOAN_LOCATION by GENDER

	GENDER		
LOAN_LOCATION	Female	Male	Total
City	28 7.63 20.00 34.57	112 30.52 80.00 39.16	140 38.15
Town	32 8.72 27.59 39.51	84 22.89 72.41 29.37	116 31.61
Village	21 5.72 18.92 25.93	90 24.52 81.08 31.47	111 30.25
Total	81 22.07	286 77.93	367 100.00



----End---

BIVARIATE ANALYSIS

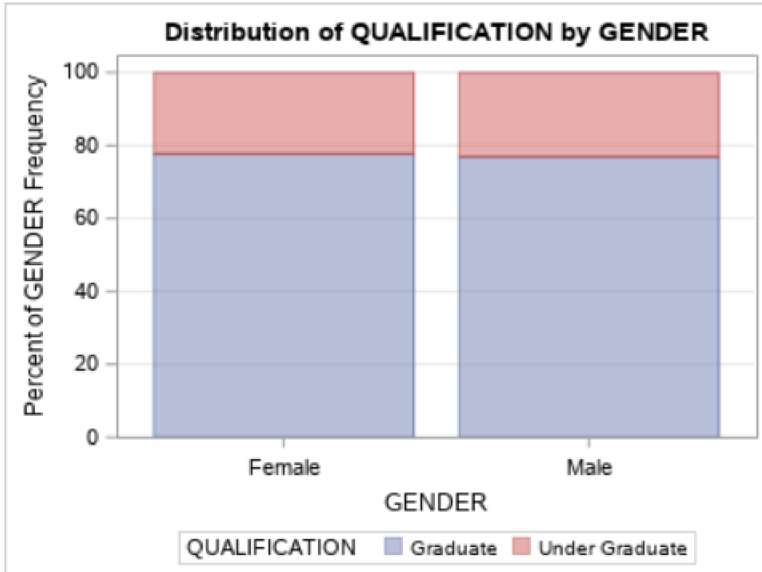
QUALIFICATION - Cate variable vs GENDER - Cate variable

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of QUALIFICATION by GENDER

QUALIFICATION	GENDER: Female	GENDER: Male	Total
Graduate	63 17.17 22.26 77.78	220 59.95 77.74 76.92	283 77.11
Under Graduate	18 4.90 21.43 22.22	66 17.98 78.57 23.08	84 22.89
Total	81 22.07	286 77.93	367 100.00



----End---

2.26.3 Explanation

This part shows the result of each minimize code done in SAS Macro for bivariate analysis. Since there are 2 SAS Macro codes, so there are 2 results for each of the macro code.

2.27 Create Logistic regression model

2.27.1 SAS Codes

```
/* Create a model */

PROC LOGISTIC DATA = DMLIBMFR.TRAINING_DS OUTMODEL = DMLIBMFR.TRAINING_DS_LRMODEL; /* Linear regression */
CLASS
GENDER
LOAN_LOCATION
MARITAL_STATUS
QUALIFICATION
LOAN_HISTORY;
/* Above are categorical variables found inside the DMLIBMFR.TRAINING_DS */
MODEL LOAN_APPROVAL_STATUS = /* place here all independent variables */
/* LOAN_APPLICATION_STATUS is a dependent variable */
GENDER
MARITAL_STATUS
QUALIFICATION
LOAN_AMOUNT
LOAN_DURATION
LOAN_HISTORY
CANDIDATE_INCOME
LOAN_LOCATION;
OUTPUT OUT = DMLIBMFR.TRAINING_DS_LR_OUT P = PRED_PROB;
/* PRED_PROB -> Predicted probability - variable to hold predicted probability
OUT -> the output will be stored in the dataset
Akaike Information Criterion must (AIC) < SC (Schwarz Criterion)
*/
RUN;
```

2.27.2 Screenshot(s)/Output(s)

Number of Observations Read	614
Number of Observations Used	614

2.27.3 Explanation

This result shows that have cleansed dataset 100 percentage. The TRAINING_DS turns pure and clamping cleansed well. Logistic regression highlighted and processed it.

2.27.4 Screenshot(s)/Output(s)

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

2.27.5 Explanation

This result shows the model convergence status fulfilled the criterion and satisfied.

2.27.6 Screenshot(s)/Output(s)

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	764.891	583.601
SC	769.311	627.801
-2 Log L	762.891	563.601

AIC < SC

2.27.7 Explanation

Akaike Information Criterion must (AIC) < SC (Schwarz Criterion). The value of SC must lower than AIC.

2.27.8 Screenshot(s)/Output(s)

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.2759	0.6583	0.1757	0.6751
GENDER	Female	1	0.0133	0.1391	0.0091	0.9238
MARITAL_STATUS	Married	1	-0.2894	0.1168	6.1435	0.0132
QUALIFICATION	Graduate	1	-0.1918	0.1283	2.2333	0.1351
LOAN_AMOUNT		1	0.00275	0.00151	3.3172	0.0686
LOAN_DURATION		1	0.000682	0.00175	0.1512	0.6973
LOAN_HISTORY	0	1	1.9452	0.2088	86.8297	<.0001
CANDIDATE_INCOME		1	-0.00002	0.000023	0.9613	0.3269
LOAN_LOCATION	City	1	0.1810	0.1497	1.4625	0.2265
LOAN_LOCATION	Town	1	-0.5272	0.1567	11.3178	0.0008

2.27.9 Explanation

If Pr > ChiSq is <= 0.05, it means that independent variable is an important variable and as is truly contributing to predict dependent variable. It means MARITAL_STATUS, QUALIFICATION, are contributing more other independent. Only few variables are contributed in this process but others is not.

2.28 Display the contents of the created model.

2.28.1 SAS Codes

```
TITLE1 'To view the contents of the model';

PROC SQL;

SELECT *
FROM DMLIBMFR.TRAINING_DS_LRMODEL;

QUIT;
```

2.28.2 Screenshot(s)/Output(s)

To view the contents of the model

TYPE	_NAME_	_CATEGORY_	_NAMEIDX_	_CATIDX_	_MISC_
L			.	.	0
M	NYYNYNNN		.	.	10
G	GENDER	Female	0	0	2
G	GENDER	Male	0	1	-2
G	GENDER		-1	-1	0
G	GENDER		-1	-2	-6
G	LOAN_LOCATION	City	1	0	2
G	LOAN_LOCATION	Town	1	1	2
G	LOAN_LOCATION	Village	1	2	-2
G	LOAN_LOCATION		-2	-1	0
G	LOAN_LOCATION		-2	-2	-7
G	MARITAL_STATUS	Married	2	0	2
G	MARITAL_STATUS	Not Married	2	1	-2
G	MARITAL_STATUS		-3	-1	0
G	MARITAL_STATUS		-3	-2	-11
G	QUALIFICATION	Graduate	3	0	2
G	QUALIFICATION	Under Graduate	3	1	-2
G	QUALIFICATION		-4	-1	0
G	QUALIFICATION		-4	-2	-14
G	LOAN_HISTORY	0	4	0	1
G	LOAN_HISTORY	1	4	1	-1
G	LOAN_HISTORY		-5	-1	0
G	LOAN_HISTORY		-5	-2	-12
G	LOAN_APPROVAL_STATUS	N	5	0	12
G	LOAN_APPROVAL_STATUS	Y	5	1	-12

2.29 Predict the LOAN_APPROVAL_STATUS using the Logistic Algorithm or model.

2.29.1 SAS Codes

```
/* Program to predict the LOAN_APPROVAL_STATUS using the Model created by the LRA (Logistic Regression Algorithm) */

PROC LOGISTIC INMODEL = DMLIBMFR.TRAINING_DS_LRMODEL; /* It is a model you created */
SCORE DATA = DMLIBMFR.TESTING_DS
OUT = DMLIBMFR.TESTING_DS_LAS_PREDICTED; /* Location of output */
QUIT;


/* To view the LOAN_APPROVAL_STATUS */

TITLE1 'LOAN_APPROVAL_STATUS';
FOOTNOTE '---End---';

PROC SQL;

SELECT *
FROM DMLIBMFR.TESTING_DS_LAS_PREDICTED;

QUIT;
```

2.29.1 Screenshot(s)/Output(s)

LOAN_APPROVAL_STATUS

LOAN_DURATION	LOAN_HISTORY	LOAN_LOCATION	LOAN_APPROVAL_STATUS	From: LOAN_APPROVAL_STATUS	Into: LOAN_APPROVAL_STATUS	Predicted Probability: LOAN_APPROVAL_STATUS=N	Predicted Probability: LOAN_APPROVAL_STATUS=Y
360	1	City			Y	0.173495	0.826505
360	1	City			Y	0.188719	0.811281
360	1	City			Y	0.218375	0.781625
360	1	City			Y	0.180398	0.819602
360	1	City			Y	0.346929	0.653071
360	1	City			Y	0.272307	0.727693
360	1	Town			Y	0.206971	0.793029
360	0	Village			N	0.953504	0.046496