

```

1  /*****
2  Name of DS: Mr. NG KIM CHYE -TP074086
3  Job Position: Data Manager
4  Name of SAS program: 2023-06-09.sas
5  Description:
6  Date first written: Fri,09-Jun-2023
7  Date last updated: Fri,09-Jun-2023
8  Folder name: MYDM_MAY_PT_4086_2023
9  Library name: MYLIB086
10 *****/
11
12 /*****
13
14 Data Cleansing (Empty/Null Value)
15
16 *****/
17
18 TITLE '8.1 List the observations with missing values in the variable - la_maital_status';
19 FOOTNOTE '-----End-----';
20
21 PROC SQL;
22
23 SELECT *
24 FROM MYLIB086.LOAN_DS 1
25 WHERE ( ( 1.la_maital_status eq '' ) OR
26         ( 1.la_maital_status IS MISSING ) );
27
28 QUIT;
29
30 TITLE '8.2 Count the total number of observations with missing values in the variable -
31 la_maital_status';
32 FOOTNOTE '-----End-----';
33
34 PROC SQL;
35
36 SELECT COUNT(*) Label 'Number of Applicants'
37 FROM MYLIB086.LOAN_DS 1
38 WHERE ( ( 1.la_maital_status eq '' ) OR
39         ( 1.la_maital_status IS MISSING ) );
40
41 QUIT;
42
43 /* 8.3 Store the statistic (la_maital_status, number of applications) in a dataset */
44
45 PROC SQL;
46
47 CREATE TABLE MYLIB086.LOAN_MAITAL_STATUS_DS AS
48 SELECT 1.la_maital_status AS la_maital_status,
49        COUNT(*) AS number_of_applicants
50 FROM MYLIB086.LOAN_DS 1
51 GROUP BY 1.la_maital_status;
52
53 QUIT;
54
55 /* 8.4 Make a backup copy of the ds */
56
57 PROC SQL;
58
59 CREATE TABLE MYLIB086.LOAN_BK_DS AS
60 SELECT *
61 FROM MYLIB086.LOAN_DS;
62
63 QUIT;
64
65 /* 8.5 Impute the missing values found in the variable - la_maital_status */

```

```

65
66 PROC SQL;
67
68 UPDATE MYLIB086.LOAN_DS
69 SET la_maital_status = (SELECT l.la_maital_status AS la_maital_status
70                          FROM MYLIB086.LOAN_MAITAL_STATUS_DS l
71                          WHERE l.number_of_applicants eq ( SELECT MAX(l.number_of_applicants) AS
72                                                                highest_count
73                                                                FROM MYLIB086.LOAN_MAITAL_STATUS_DS l
74                                                                ) )
75 /*The above is a sub-program to find
76 the highest count*/
77
78 WHERE ( ( la_maital_status eq '' ) OR
79         ( la_maital_status IS MISSING ) );
80
81 QUIT;
82
83 TITLE '8.6 (After cleanse) List the observations with missing values in the variable -
84 la_maital_status';
85 FOOTNOTE '-----End-----';
86
87 PROC SQL;
88
89 SELECT *
90 FROM MYLIB086.LOAN_DS l
91 WHERE ( ( l.la_maital_status eq '' ) OR
92         ( l.la_maital_status IS MISSING ) );
93
94 QUIT;
95
96 /*****
97
98 Data Cleansing (Unwanted Symbol)
99
100 *****/
101
102 TITLE '9.1 List the details of the observations with [3+]';
103 FOOTNOTE '-----End-----';
104
105 PROC SQL;
106
107 SELECT *
108 FROM MYLIB086.LOAN_DS l
109 WHERE ( l.la_family_members eq '3+' );
110
111 QUIT;
112
113 TITLE '9.2 Count the total number of applicants with [3+] family members';
114 FOOTNOTE '-----End-----';
115
116 PROC SQL;
117
118 SELECT COUNT(*) Label = 'Number of applicants'
119 FROM MYLIB086.LOAN_DS l
120 WHERE ( l.la_family_members eq '3+' );
121
122 QUIT;
123
124 TITLE '9.3 Remove the [+] found in the variable - la_family_members';
125 FOOTNOTE '----End----';
126
127 PROC SQL;
128
129 UPDATE MYLIB086.LOAN_DS

```

```
126 SET la_family_members = SUBSTR( la_family_members, 1, 1 )
127 WHERE ( la_family_members eq '3+' );
128
129 QUIT;
130
131 TITLE '9.5 (After cleanse) Count the total number of applicants with [3+] family members';
132 FOOTNOTE '-----End-----';
133
134 PROC SQL;
135
136 SELECT COUNT(*) Label = 'Number of applicants'
137 FROM MYLIB086.LOAN_DS 1
138 WHERE ( 1.la_family_members eq '3+' );
139
140 QUIT;
141
142
```