



METRO COLLEGE OF TECHNOLOGY

Project for investigating the Customer Distribution, Business Behaviors and Deactivation of a Telecommunication Company using SAS Programming

Submitted by:

Mohammad Monjur-E-Elahi

Course: Advanced SAS Programming [DSA05]

Program: Data Science and Application - Advanced Diploma [6060]

Metro College of Technology

Date: 24 April, 2021

Introduction:

In this project we will be analyzing and investing the CRM data of a Wireless Telecommunication Company for 2 years. The wireless company would like us to investigate the customer distribution and business behaviors, and then gain insightful understanding about the customers, and to forecast the deactivation trends for the next 6 months. We will be performing the EDA, Distribution analysis, visualizaion and some statistical significant association analysis.

Data:

Acctno: account number.

Actddt: account activation date

Deactddt: account deactivation date

DeactReason: reason for deactivation.

GoodCredit: customer's credit is good or not.

RatePlan: rate plan for the customer.

DealerType: dealer type.

Age: customer age.

Province: province.

Sales: the amount of sales to a customer.

Task1: Let us first import the data from the provided text file to SAS in the custom permament library name Project that I have created to handle this project.

First look at the data in .txt format

New_Wireless_Fixed.txt - Notepad

File Edit Format View Help

1176913194483	06/20/1999			0	1	A1	58	BC	\$128.00
1176914599423	10/04/1999	10/15/1999	NEED	1	1	A1	45	AB	\$72.00
1176951913656	07/01/2000			0	1	A1	57	BC	\$593.00
1176954000288	05/30/2000			1	2	A1	47	ON	\$83.00
1176969186303	12/13/2000			1	1	C1	82	BC	
1176991056273	08/31/1999	09/18/2000	MOVE	1	1	C1	92	QC	\$1041.00
1176991866552	05/24/2000			1	1	A1	77	ON	
1176992889500	11/28/2000			1	1	C1	68	AB	\$72.00
117700067271	12/23/1999			0	1	B1	75	ON	\$134.00
1177010940613	12/09/1999			1	2	A1	42	NS	\$11.00
1177025997013	11/09/1999			1	1	A1	26	BC	\$154.00
1177027515760	10/19/1999			1	1	B1	73	BC	\$16.00
1177028996676	09/21/2000			0	1	C1		QC	\$179.00
1177038747105	03/14/2000			0	1	C1	41	ON	\$705.00
1177045857516	06/22/2000			1	1	A1	53	QC	\$83.00
1177057406016	09/21/2000			0	1	C1	50	ON	\$529.00
1177066422248	04/26/1999	01/15/2001	NEED	0	1	A2	55	NS	\$44.00
1177089399155	08/17/2000			1	3	A1	56	BC	\$548.00
1177113886410	09/13/2000	01/08/2001	COMP	0	1	C1	45	ON	\$63.00
1177128264924	12/10/1999			1	1	B1	38	ON	\$178.00
1177135400264	12/09/2000			0	1	A1	42	ON	\$345.00
1177169300842	03/24/2000			0	3	B1	54	ON	\$331.00
1177177440963	02/15/2000			0	1	B1	23	BC	\$97.00

Task2: Let us import the dataset to SAS first. We bring the dataset to the Permanent custome Library named Project.

```
DATA Project.CustomerBehavior;
Infile "C:\Users\ruzdomain\Desktop\ASAS\Project\New_Wireless_Fixed.txt";

Input @1 Account_Number $13.
      @15 Account_Activation_Date MMDDYY10.
      @26 Account_Deactivation_Date MMDDYY10.
      @37 Reason_For_Deactivation $8.
      @53 Customer_Credit_Status 1.
      @62 Rate_Plan 1.
      @65 Dealer_Type $2.
      @74 Customer_Age 2.
      @80 Province $2.
      @83 Sales_Amount dollar8.2
;
Format Account_Activation_Date Account_Deactivation_Date MMDDYY10. Sales_Amount dollar8.2;
RUN;
```

Task3: Let us run the below code to get the top 20 rows of the dataset in SAS.

```
PROC PRINT DATA = Project.CustomerBehavior (OBS =20);
RUN;
```

Obs	Account_Number	Account_Activation_Date	Account_Deactivation_Date	Reason_For_Deactivation	Customer_Credit_Status	Rate_Plan	Dealer_Type	Customer_Age	Province	Sales_Amount
1	1176913194483	06/20/1999	-		0	1	A1	58	BC	\$128.00
2	1176914599423	10/04/1999	10/15/1999	NEED	1	1	A1	45	AB	\$72.00
3	1176951913656	07/01/2000	-		0	1	A1	57	BC	\$593.00
4	1176954000288	05/30/2000	-		1	2	A1	47	ON	\$83.00
5	1176969186303	12/13/2000	-		1	1	C1	82	BC	-
6	1176991056273	08/31/1999	09/18/2000	MOVE	1	1	C1	92	QC	\$1041.00
7	1176991866552	05/24/2000	-		1	1	A1	77	ON	-
8	1176992889500	11/28/2000	-		1	1	C1	68	AB	\$72.00
9	1177000067271	12/23/1999	-		0	1	B1	75	ON	\$134.00
10	1177010940613	12/09/1999	-		1	2	A1	42	NS	\$11.00
11	1177025997013	11/09/1999	-		1	1	A1	26	BC	\$154.00
12	1177027515760	10/19/1999	-		1	1	B1	73	BC	\$16.00
13	1177028996676	09/21/2000	-		0	1	C1	-	QC	\$179.00
14	1177038747105	03/14/2000	-		0	1	C1	41	ON	\$705.00
15	1177045857516	06/22/2000	-		1	1	A1	53	QC	\$83.00
16	1177057406016	09/21/2000	-		0	1	C1	50	ON	\$529.00
17	1177066422248	04/26/1999	01/15/2001	NEED	0	1	A2	55	NS	\$44.00
18	1177089399155	08/17/2000	-		1	3	A1	56	BC	\$548.00
19	1177113886410	09/13/2000	01/08/2001	COMP	0	1	C1	45	ON	\$63.00
20	1177128264924	12/10/1999	-		1	1	B1	38	ON	\$178.00

Task4: We examine the content of the dataset using the below code.

```
PROC CONTENTS DATA=Project.CustomerBehavior varnum;
RUN;
```

The CONTENTS Procedure

Data Set Name	PROJECT.CUSTOMERBEHAVIOR	Observations	102255
Member Type	DATA	Variables	10
Engine	V9	Indexes	0
Created	04/23/2021 21:39:28	Observation Length	80
Last Modified	04/23/2021 21:39:28	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	126
First Data Page	1
Max Obs per Page	817
Obs in First Data Page	791
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	C:\Users\ruzdomain\Desktop\ASAS\Project\customerbehavior.sas7bdat
Release Created	9.0401M6
Host Created	X64_10PRO
Owner Name	DINARTRIOSEDUCA\ruzdomain
File Size	8MB
File Size (bytes)	8323072

Variables in Creation Order				
#	Variable	Type	Len	Format
1	Account_Number	Char	13	

2	Account_Activation_Date	Num	8	MMDDYY10.
3	Account_Deactivation_Date	Num	8	MMDDYY10.
4	Reason_For_Deactivation	Char	8	
5	Customer_Credit_Status	Num	8	
6	Rate_Plan	Num	8	
7	Dealer_Type	Char	2	
8	Customer_Age	Num	8	
9	Province	Char	2	
10	Sales_Amount	Num	8	DOLLAR8.2

Task5: We came to know from the output of the content that total number of rows were 102255 and we go ahead to check the bottom 20 row.

```
PROC PRINT DATA = Project.CustomerBehavior (FIRSTOBS=102236) ;  
RUN;
```

Obs	Account_Number	Account_Activation_Date	Account_Deactivation_Date	Reason_For_Deactivation	Customer_Credit_Status	Rate_Plan	Dealer_Type	Customer_Age	Province	Sales_Amount
102236	2671603280496	01/13/2001	-	-	1	2	A1	49	QC	\$141.00
102237	2671907586819	12/23/2000	-	-	1	2	A1	74	ON	\$81.00
102238	2672144380608	12/16/2000	-	-	1	2	A1	49	ON	\$47.00
102239	2672331366304	12/24/2000	-	-	1	2	B1	71	ON	\$67.00
102240	2672370281824	12/12/2000	-	-	1	2	B1	-	ON	\$85.00
102241	2672600712600	12/20/2000	-	-	1	1	C1	70	BC	\$434.00
102242	2672684650146	12/18/2000	-	-	1	2	A1	72	ON	\$345.00
102243	2672761336176	12/20/2000	-	-	0	1	C1	18	ON	\$210.00
102244	2672981197684	12/28/2000	12/28/2000	-	0	1	C1	44	QC	\$12.00
102245	2672987766282	01/02/2001	-	-	1	2	A1	-	ON	\$40.00
102246	2673080989261	01/06/2001	-	-	1	2	B1	68	ON	\$25.00
102247	2673104899783	01/12/2001	-	-	0	2	C1	68	ON	-
102248	2673110609830	01/12/2001	-	-	1	1	A2	63	ON	\$78.00
102249	2673216477998	01/12/2001	-	-	1	2	C1	53	ON	\$316.00
102250	2673579485183	12/31/2000	-	-	0	1	A1	-	NS	\$130.00
102251	2673974127660	12/29/2000	-	-	1	1	A2	50	-	\$112.00
102252	2674189951308	01/15/2001	-	-	1	2	A1	40	BC	\$87.00
102253	2674548796918	01/15/2001	-	-	1	1	A1	16	NS	\$316.00
102254	2675119766018	01/15/2001	-	-	1	2	B1	76	ON	-
102255	2675135410256	01/17/2001	-	-	1	1	A1	46	BC	\$319.00

Task6: Let us attempt to find the distribution of missing and non-missing values in each column.

```
✉DATA WORK.CustomerBehavior_00 ;
  SET Project.CustomerBehavior ;
  RUN;


---


✉proc format ;
  value $missfmt ' ' = 'Missing' other = 'Not Missing';
  value missfmt . = 'Missing' other = 'Not Missing';
  RUN;


---


✉PROC FREQ DATA = CustomerBehavior_00;
  format _CHAR_ $missfmt.;
  format _NUMERIC_ missfmt.|;
  tables _CHAR_ / missing nocum nopercent;
  tables _NUMERIC_ / missing nocum nopercent;
  RUN;
```

The FREQ Procedure

Account_Number	Frequency
Not Missing	102255

Reason_For_Deactivation	Frequency
Missing	83162
Not Missing	19093

Dealer_Type	Frequency
Not Missing	102255

Province	Frequency
Missing	5907
Not Missing	96348

Account_Activation_Date	Frequency
Not Missing	102255

Account_Deactivation_Date	Frequency
Missing	82620
Not Missing	19635

Customer_Credit_Status	Frequency
Not Missing	102255

Rate_Plan	Frequency
Not Missing	102255

Customer_Age	Frequency
Missing	7708
Not Missing	94547

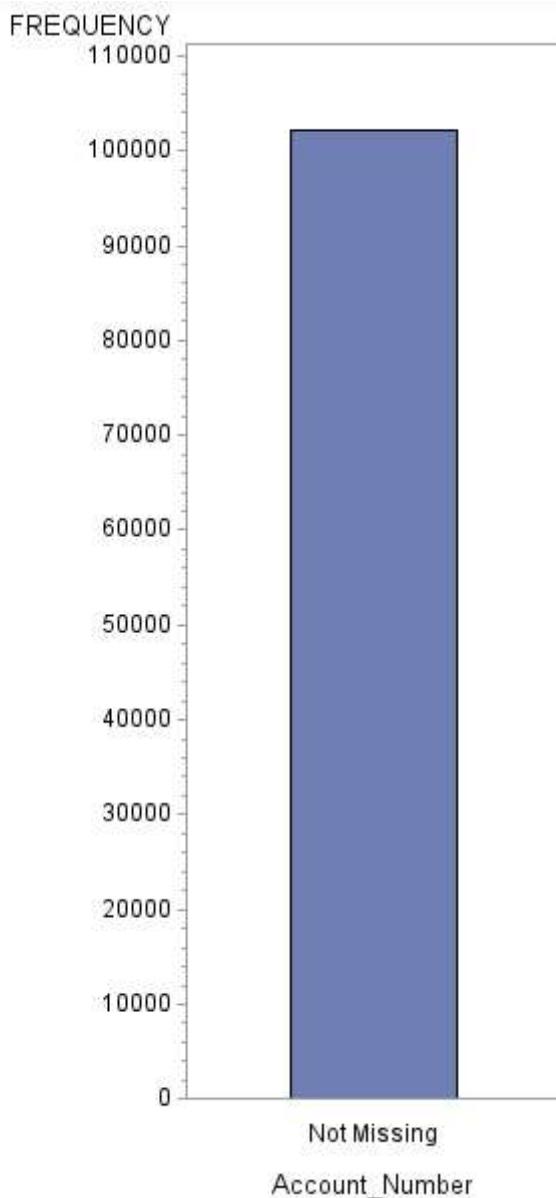
Sales_Amount	Frequency
Missing	8605
Not Missing	93650

Task7: Now let us try to visualize the missing values in each columns.

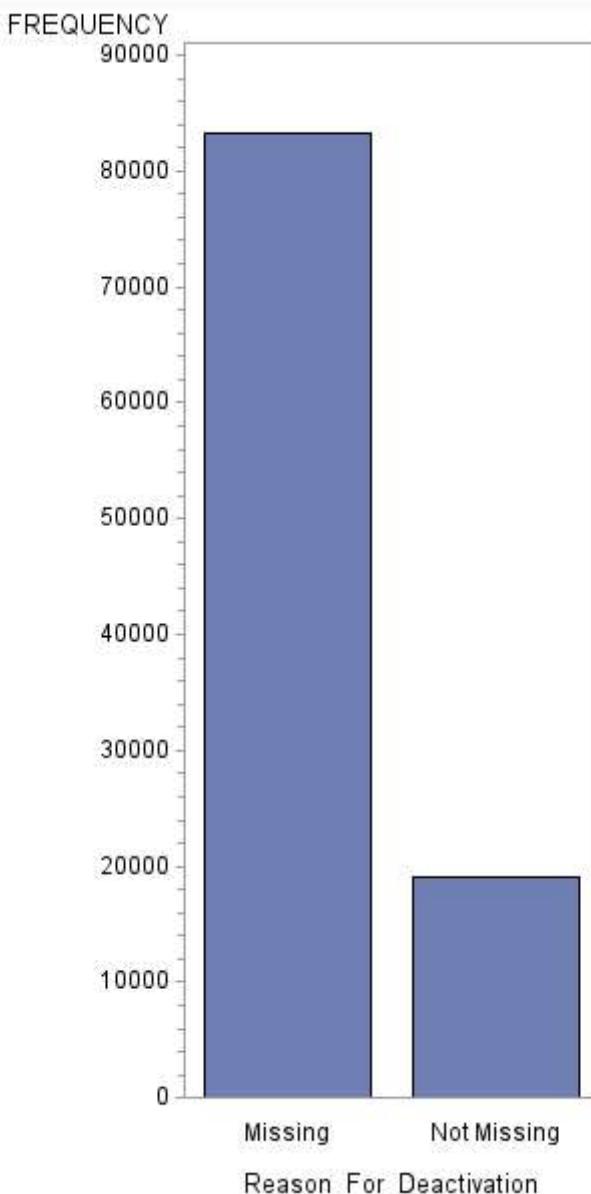
```
DATA CustomerBehavior_Missing;
SET Project.CustomerBehavior;
format Account_Number Reason_For_Deactivation Dealer_Type Province $missfmt.
      Account_Number Account_Activation_Date Account_Deactivation_Date Customer_Credit_Status Rate_Plan Customer_Age Sales_Amount missfmt.;

RUN;
```

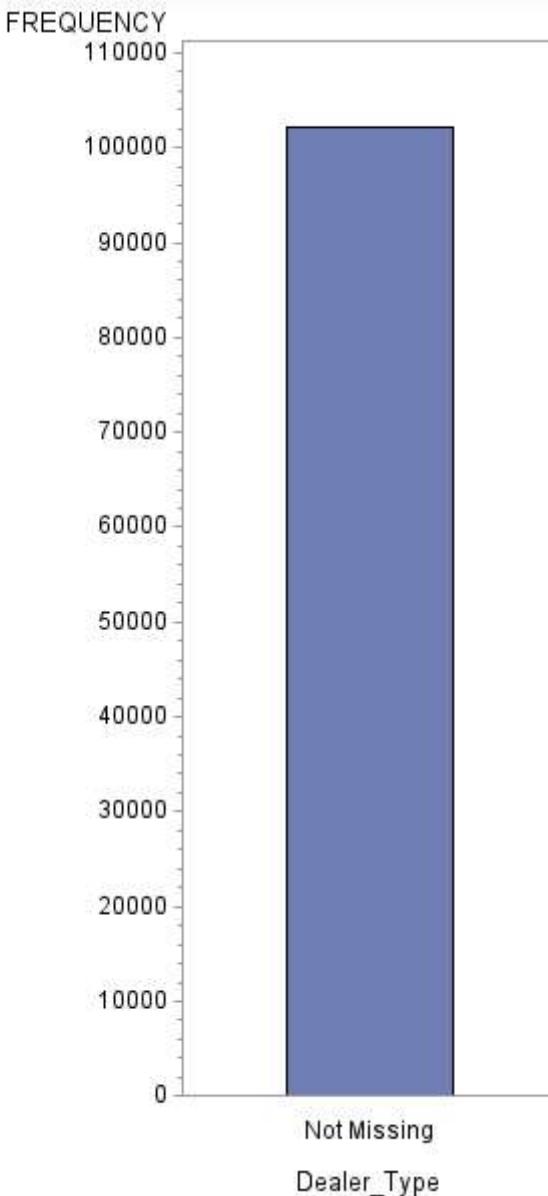
```
PROC GCHART DATA = CustomerBehavior_Missing;
vbar Account_Number;
RUN;
```



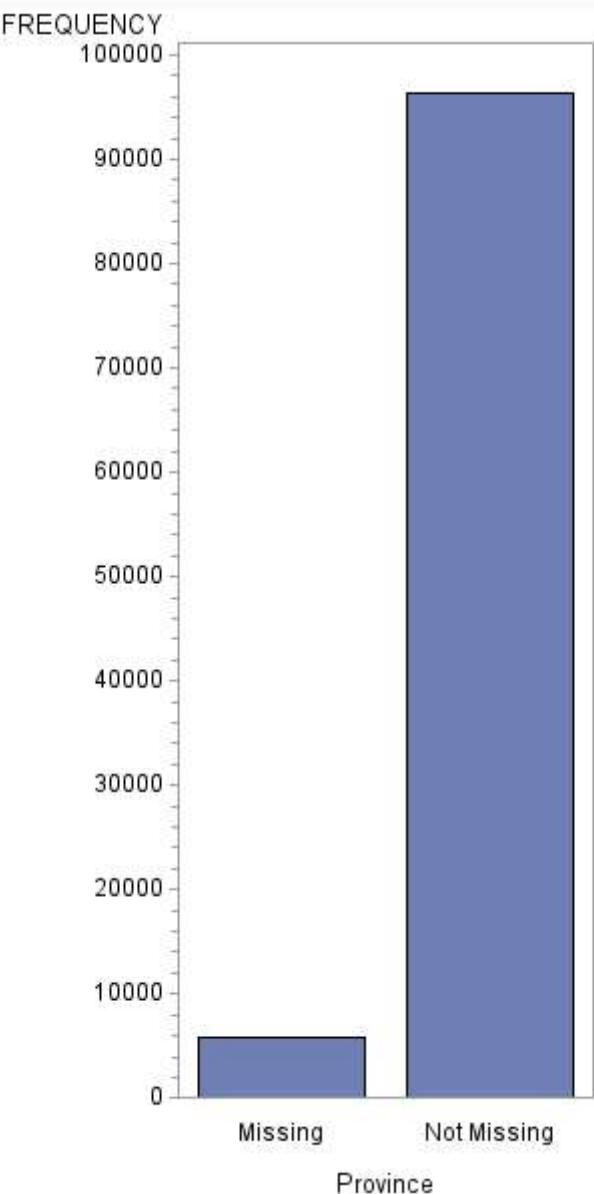
```
PROC GCHART DATA = CustomerBehavior_Missing;
vbar Reason_For_Deactivation;
RUN;
```



```
PROC GCHART DATA = CustomerBehavior_Missing;  
vbar Dealer_Type;  
RUN;
```

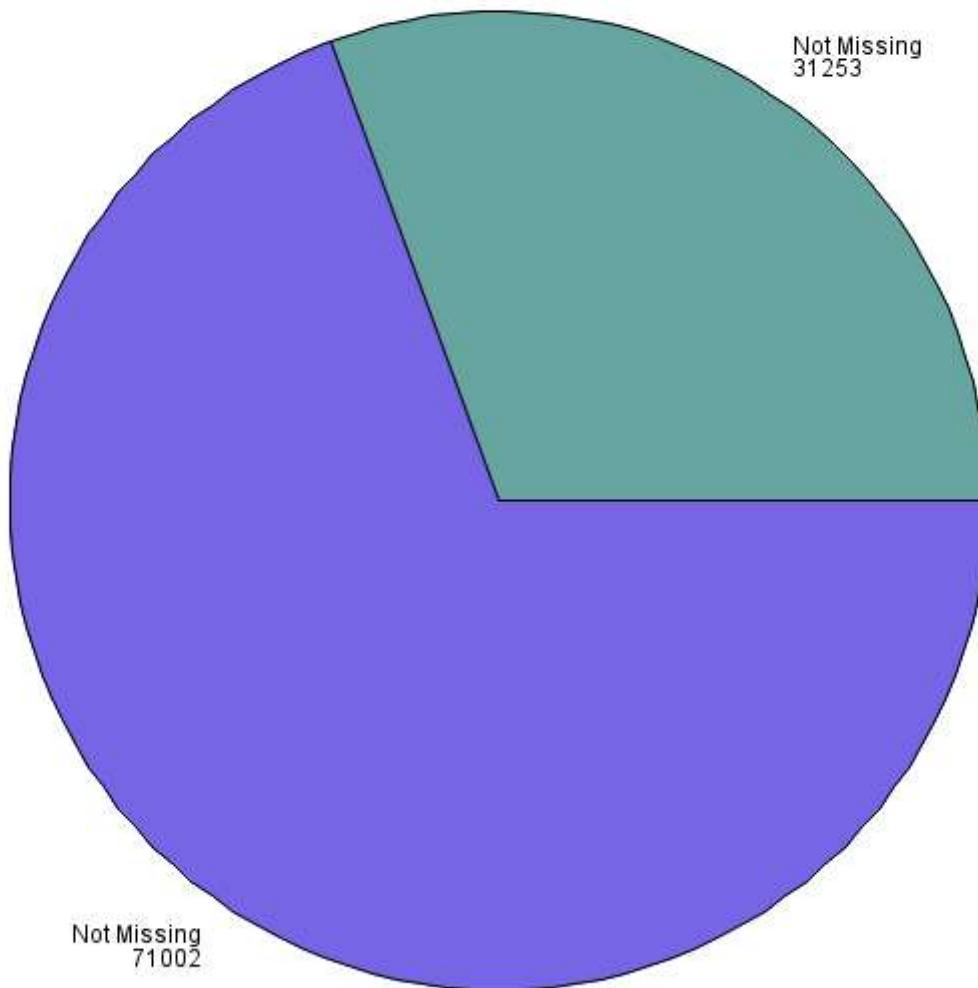


```
PROC GCHART DATA = CustomerBehavior_Missing;  
vbar Province;  
RUN;
```



```
PROC GCHART DATA = CustomerBehavior_Missing;  
PIE Customer_Credit_Status ;  
RUN;
```

FREQUENCY of Customer_Credit_Status



Task8: Let us now attempt to vizualize the data as it is found just after loadinbg the data to SAS.

```
□PROC GCHART DATA = Project.CustomerBehavior;  
    vbar Reason_For_Deactivation;  
RUN;

---


```

```
□PROC GCHART DATA = Project.CustomerBehavior;  
    vbar Dealer_Type;  
RUN;

---


```

```
□PROC GCHART DATA = Project.CustomerBehavior;  
    vbar Province;  
RUN;

---


```

```
□PROC GCHART DATA = Project.CustomerBehavior;  
    PIE Customer_Credit_Status ;  
RUN;

---


```

```
□PROC GCHART DATA = Project.CustomerBehavior;  
    PIE Rate_Plan ;  
RUN;

---


```

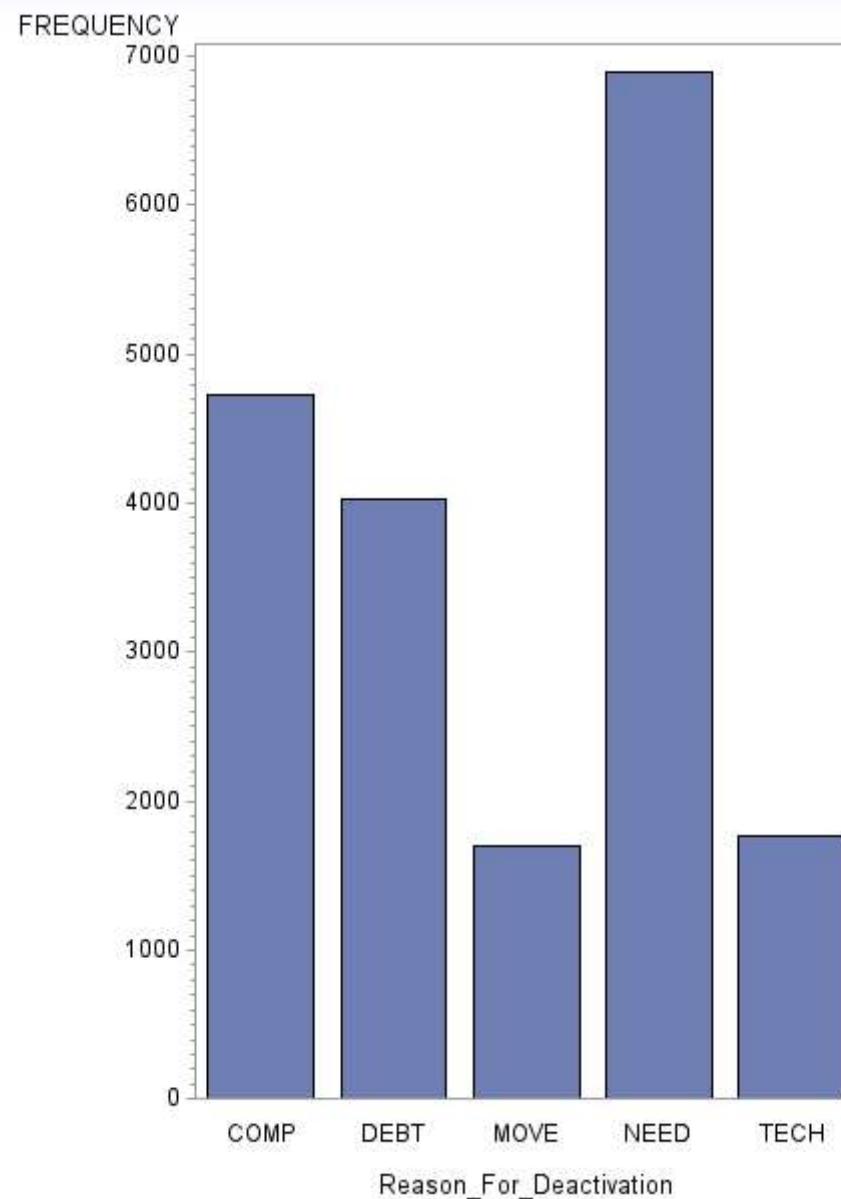
```
□PROC UNIVARIATE DATA = Project.CustomerBehavior;  
    var Customer_Age;  
    histogram/normal;  
RUN;

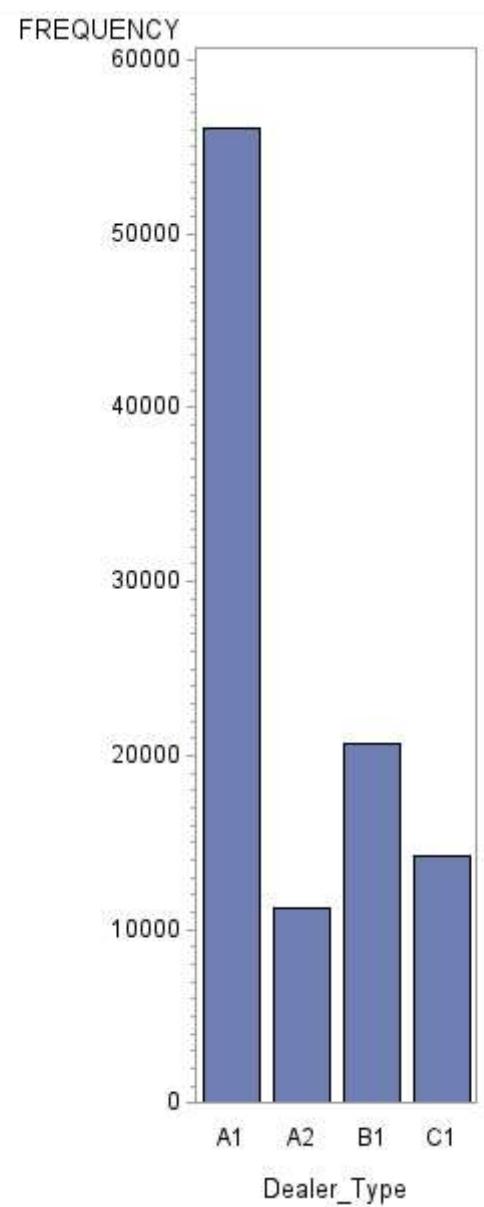
---


```

```
□PROC UNIVARIATE DATA = Project.CustomerBehavior;  
    var Sales_Amount;  
    histogram/normal;  
RUN;
```

The visualization is as shown below.





FREQUENCY

50000

40000

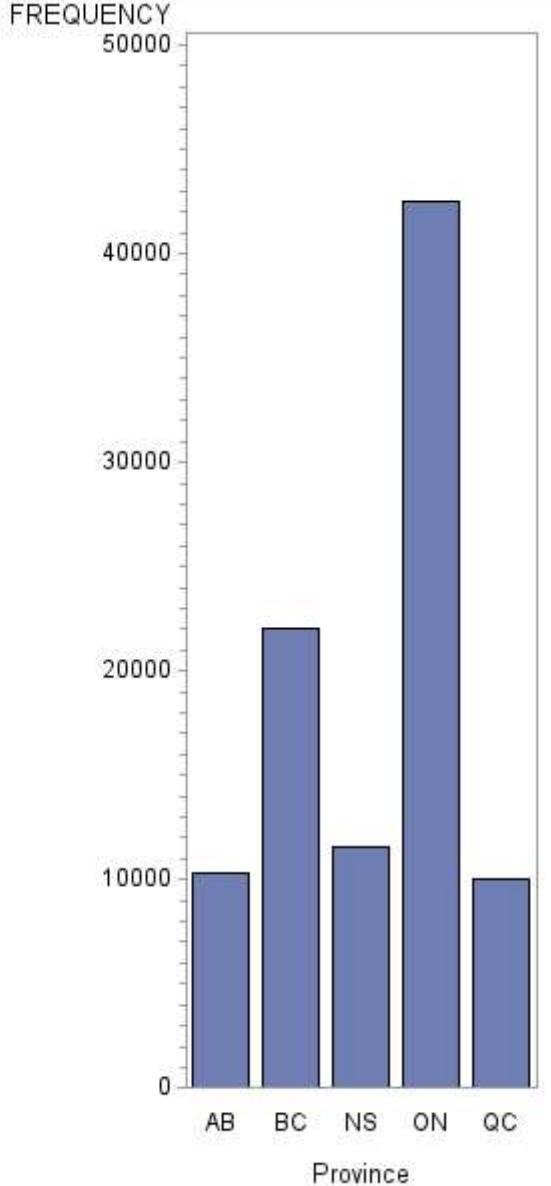
30000

20000

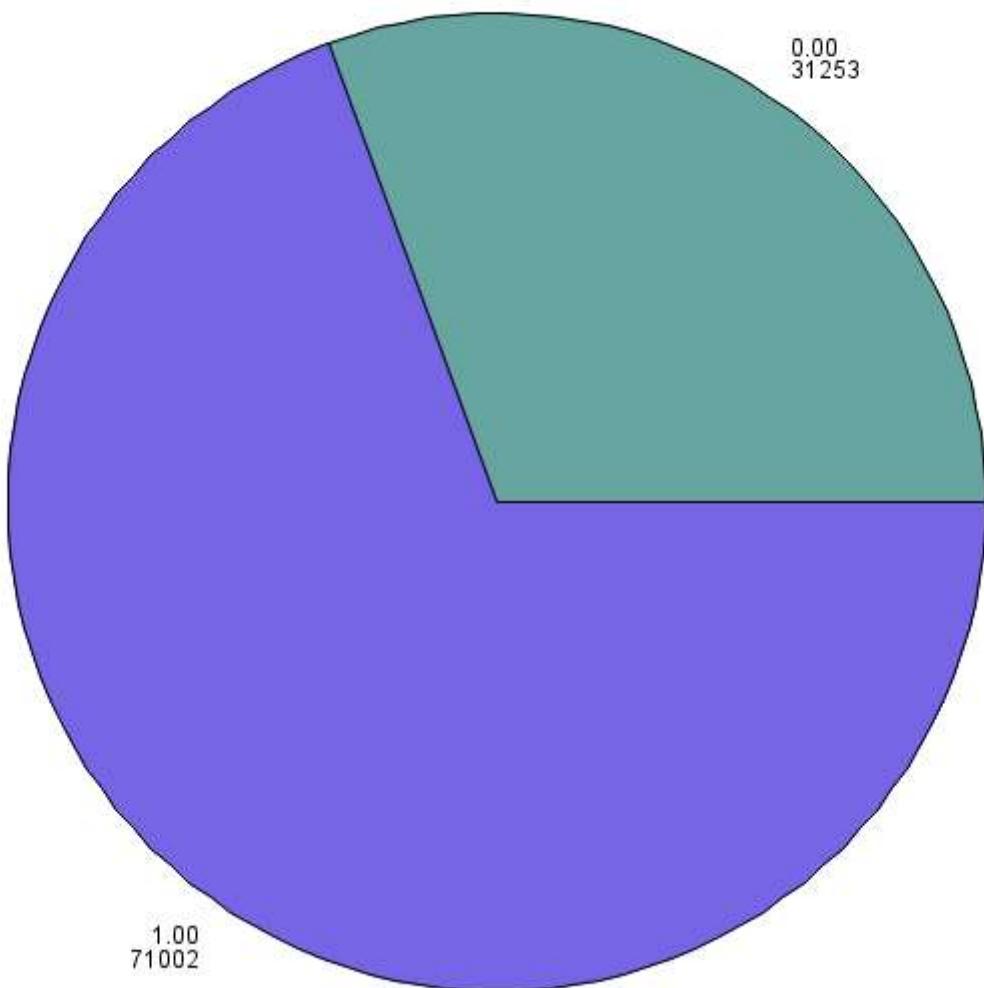
10000

AB BC NS ON QC

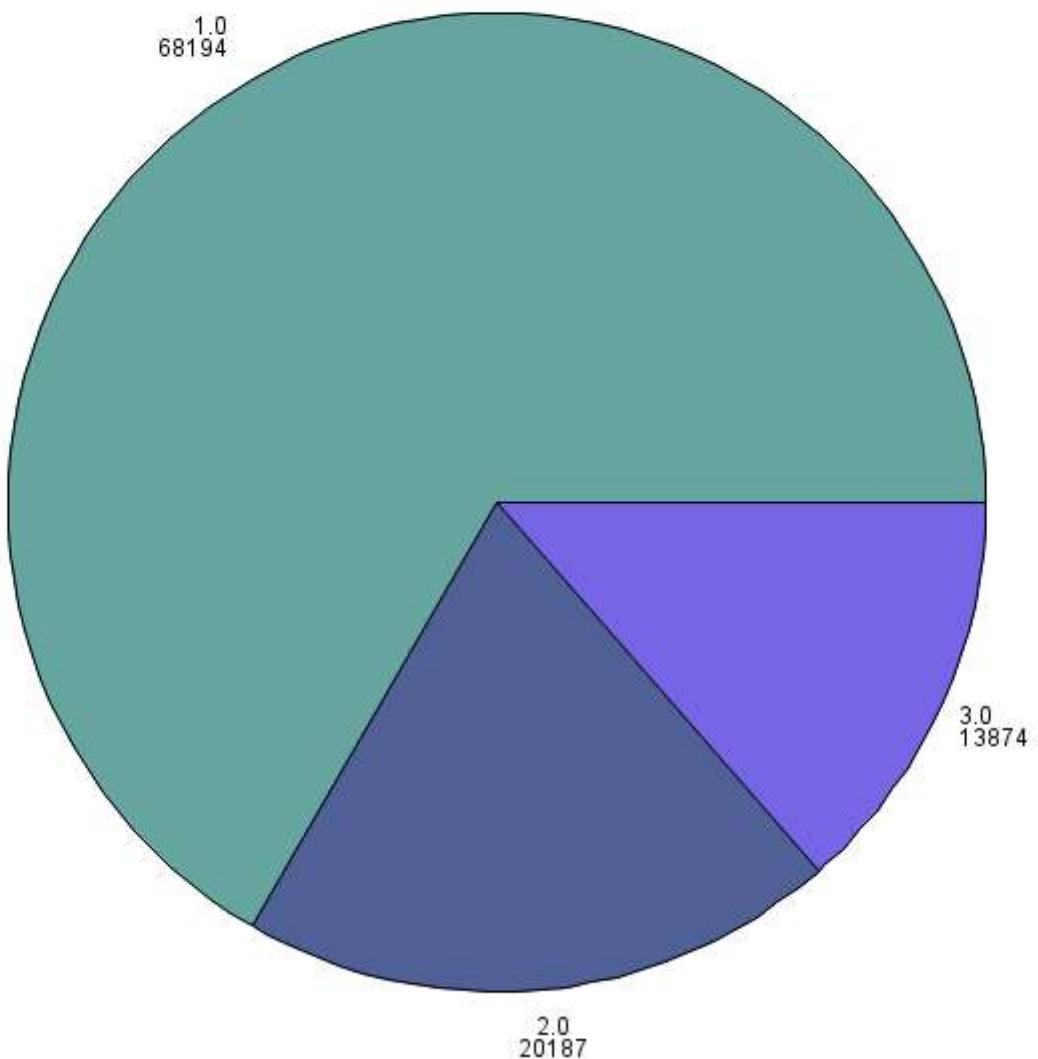
Province



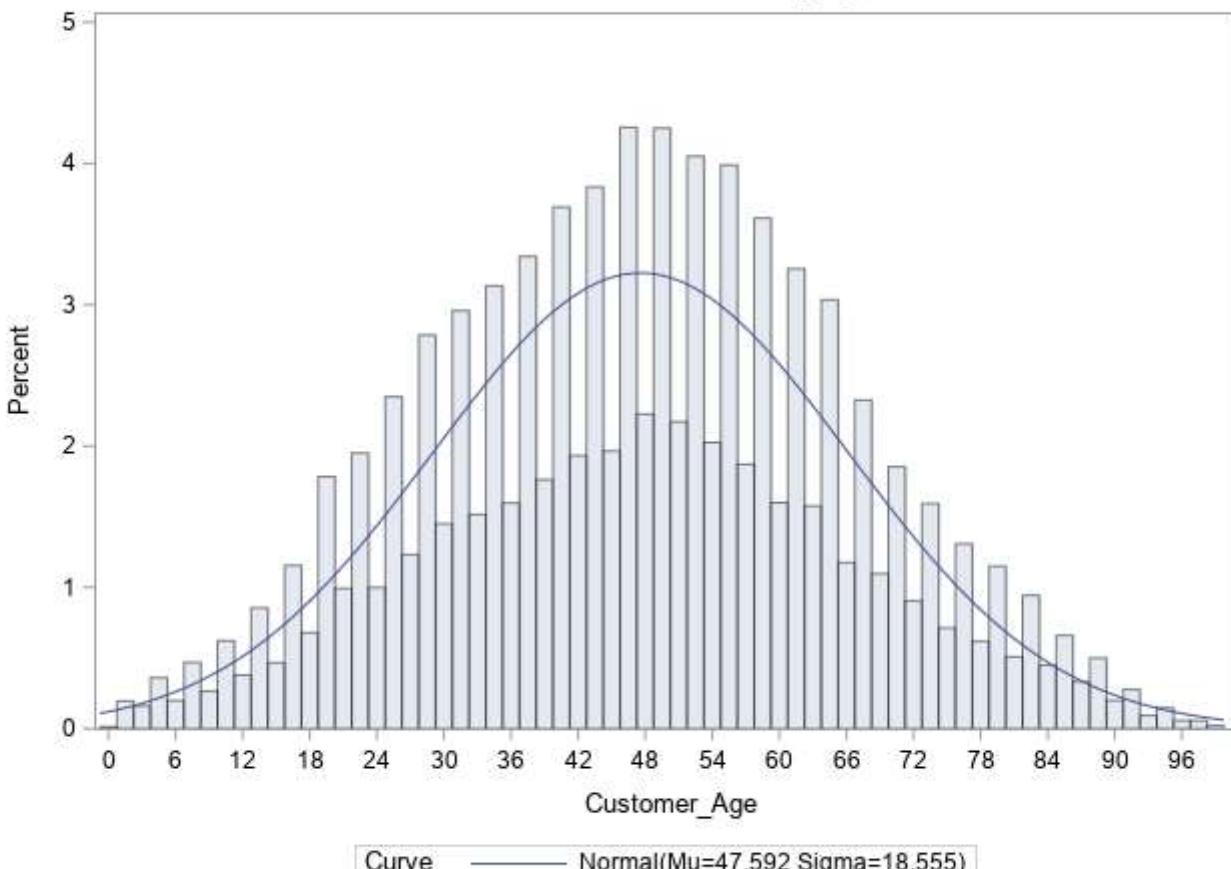
FREQUENCY of Customer_Credit_Status

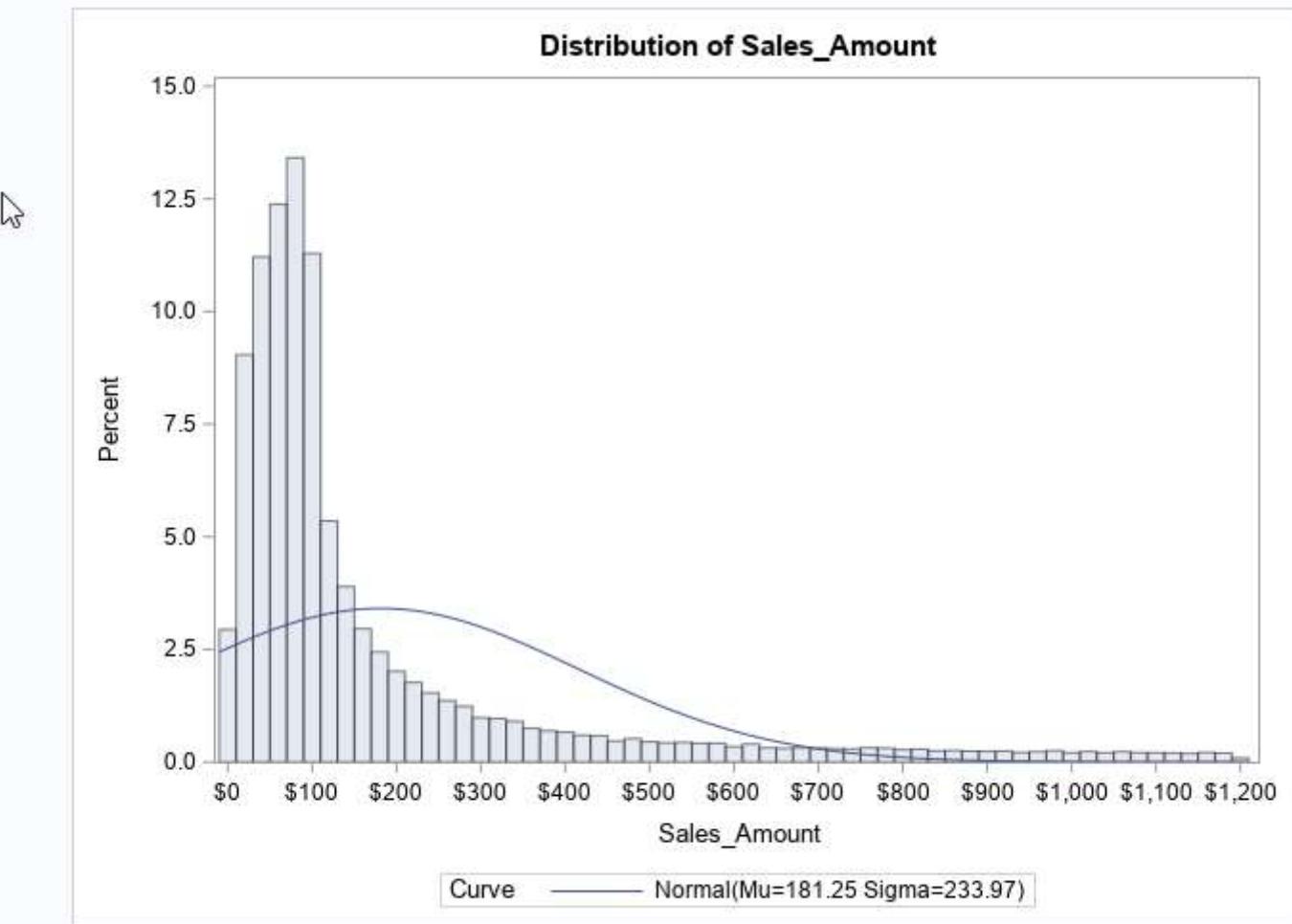


FREQUENCY of Rate_Plan



Distribution of Customer_Age





Task9: Let us now convert the missing values in the Account_Deactivation_Date column to churn value RETAIN meaning that churn did not take place and non missing values to churn value CHURN meaning that churn took place.

```
□ PROC SQL;
  CREATE TABLE Project.CustomerBehavior_01 AS
  SELECT * ,
    CASE
      WHEN Account_Deactivation_Date EQ . THEN "RETAIN"
      ELSE "CHURN"
    END AS Churn_Status
  FROM Project.CustomerBehavior
  ;
QUIT;
```

```
□ PROC PRINT DATA = Project.CustomerBehavior_01 (OBS =20) ;
  RUN;
```

Obs	Account_Number	Account_Activation_Date	Account_Deactivation_Date	Reason_For_Deactivation	Customer_Credit_Status	Rate_Plan	Dealer_Type	Customer_Age	Province	Sales_Amount	Churn_Status
1	1176913194483	06/20/1999			0	1	A1	58	BC	\$128.00	RETAIN
2	1176914599423	10/04/1999	10/15/1999	NEED	1	1	A1	45	AB	\$72.00	CHURN
3	1176951913656	07/01/2000			0	1	A1	57	BC	\$593.00	RETAIN
4	1176954000288	05/30/2000			1	2	A1	47	ON	\$83.00	RETAIN
5	1176969186303	12/13/2000			1	1	C1	82	BC		RETAIN
6	1176991056273	08/31/1999	09/18/2000	MOVE	1	1	C1	92	QC	\$1041.00	CHURN
7	1176991866552	05/24/2000			1	1	A1	77	ON		RETAIN
8	1176992889500	11/28/2000			1	1	C1	68	AB	\$72.00	RETAIN
9	1177000067271	12/23/1999			0	1	B1	75	ON	\$134.00	RETAIN
10	1177010940613	12/09/1999			1	2	A1	42	NS	\$11.00	RETAIN
11	1177025997013	11/09/1999			1	1	A1	26	BC	\$154.00	RETAIN
12	1177027515760	10/19/1999			1	1	B1	73	BC	\$16.00	RETAIN
13	1177028996676	09/21/2000			0	1	C1		QC	\$179.00	RETAIN
14	1177038747105	03/14/2000			0	1	C1	41	ON	\$705.00	RETAIN
15	1177045857516	06/22/2000			1	1	A1	53	QC	\$83.00	RETAIN
16	1177057406016	09/21/2000			0	1	C1	50	ON	\$529.00	RETAIN
17	1177066422248	04/26/1999	01/15/2001	NEED	0	1	A2	55	NS	\$44.00	CHURN
18	1177089399155	08/17/2000			1	3	A1	56	BC	\$548.00	RETAIN
19	1177113886410	09/13/2000	01/08/2001	COMP	0	1	C1	45	ON	\$63.00	CHURN
20	1177128264924	12/10/1999			1	1	B1	38	ON	\$178.00	RETAIN

Task10: Trying to convert the missing values to MEAN for Customer_Age

```

PROC SQL;
  CREATE TABLE Project.CustomerBehavior_01a AS
  SELECT * ,
    CASE
      WHEN Customer_Age EQ . THEN MEAN(Customer_Age)
      ELSE Customer_Age
    END AS NM_Customer_Age
  FROM Project.CustomerBehavior_01
  ;
QUIT;

```

```

PROC PRINT DATA = Project.CustomerBehavior_01a (OBS =20) ;
RUN;

```

Obs	Account_Number	Account_Activation_Date	Account_Deactivation_Date	Reason_For_Deactivation	Customer_Credit_Status	Rate_Plan	Dealer_Type	Customer_Age	Province	Sales_Amount	Churn_Status	NM_Customer_Age
1	1176913194483		06/20/1999		0	1	A1	58	BC	\$128.00	RETAIN	58.0000
2	1176914599423		10/04/1999	10/15/1999 NEED	1	1	A1	45	AB	\$72.00	CHURN	45.0000
3	1176951913656		07/01/2000		0	1	A1	57	BC	\$593.00	RETAIN	57.0000
4	1176954000288		05/30/2000		1	2	A1	47	ON	\$83.00	RETAIN	47.0000
5	1176969186303		12/13/2000		1	1	C1	82	BC		RETAIN	82.0000
6	1176991056273		08/31/1999	09/18/2000 MOVE	1	1	C1	92	QC	\$1041.00	CHURN	92.0000
7	1176991866552		05/24/2000		1	1	A1	77	ON		RETAIN	77.0000
8	1176992889500		11/28/2000		1	1	C1	68	AB	\$72.00	RETAIN	68.0000
9	1177000067271		12/23/1999		0	1	B1	75	ON	\$134.00	RETAIN	75.0000
10	1177010940613		12/09/1999		1	2	A1	42	NS	\$11.00	RETAIN	42.0000
11	1177025997013		11/09/1999		1	1	A1	26	BC	\$154.00	RETAIN	26.0000
12	1177027515760		10/19/1999		1	1	B1	73	BC	\$16.00	RETAIN	73.0000
13	1177028996676		09/21/2000		0	1	C1		QC	\$179.00	RETAIN	47.5922
14	1177038747105		03/14/2000		0	1	C1	41	ON	\$705.00	RETAIN	41.0000
15	1177045857516		06/22/2000		1	1	A1	53	QC	\$83.00	RETAIN	53.0000
16	1177057406016		09/21/2000		0	1	C1	50	ON	\$529.00	RETAIN	50.0000
17	1177066422248		04/26/1999	01/15/2001 NEED	0	1	A2	55	NS	\$44.00	CHURN	55.0000
18	1177089399155		08/17/2000		1	3	A1	56	BC	\$548.00	RETAIN	56.0000
19	1177113886410		09/13/2000	01/08/2001 COMP	0	1	C1	45	ON	\$63.00	CHURN	45.0000
20	1177128264924		12/10/1999		1	1	B1	38	ON	\$178.00	RETAIN	38.0000

Task11: Trying to convert the missing values to MEAN for Sales_Amount.

```
PROC SQL;
  CREATE TABLE Project.CustomerBehavior_01b AS
  SELECT * ,
    CASE
      WHEN Sales_Amount EQ . THEN MEAN(Sales_Amount)
      ELSE Sales_Amount
    END AS NM_Sales_Amount
  FROM Project.CustomerBehavior_01a
  ;
QUIT;

PROC PRINT DATA = Project.CustomerBehavior_01b (OBS =20) ;
RUN;
```

Account_Number	Account_Activation_Date	Account_Deactivation_Date	Reason_For_Deactivation	Customer_Credit_Status	Rate_Plan	Dealer_Type	Customer_Age	Province	Sales_Amount	Churn_Status	NM_Customer_Age	NM_Sales_Amount
1176913194483	06/20/1999	-		0	1	A1	58	BC	\$128.00	RETAIN	58.0000	128.00
1176914599423	10/04/1999	10/15/1999	NEED	1	1	A1	45	AB	\$72.00	CHURN	45.0000	72.00
1176951913656	07/01/2000	-		0	1	A1	57	BC	\$593.00	RETAIN	57.0000	593.00
1176954000288	05/30/2000	-		1	2	A1	47	ON	\$83.00	RETAIN	47.0000	83.00
1176969186303	12/13/2000	-		1	1	C1	82	BC	-	RETAIN	82.0000	181.25
1176991056273	08/31/1999	09/18/2000	MOVE	1	1	C1	92	QC	\$1041.00	CHURN	92.0000	1041.00
1176991866552	05/24/2000	-		1	1	A1	77	ON	-	RETAIN	77.0000	181.25
1176992889500	11/28/2000	-		1	1	C1	68	AB	\$72.00	RETAIN	68.0000	72.00
1177000067271	12/23/1999	-		0	1	B1	75	ON	\$134.00	RETAIN	75.0000	134.00
1177010940613	12/09/1999	-		1	2	A1	42	NS	\$11.00	RETAIN	42.0000	11.00
1177025997013	11/09/1999	-		1	1	A1	26	BC	\$154.00	RETAIN	26.0000	154.00
1177027515760	10/19/1999	-		1	1	B1	73	BC	\$16.00	RETAIN	73.0000	16.00
1177028996676	09/21/2000	-		0	1	C1	-	QC	\$179.00	RETAIN	47.5922	179.00
1177038747105	03/14/2000	-		0	1	C1	41	ON	\$705.00	RETAIN	41.0000	705.00
1177045857516	06/22/2000	-		1	1	A1	53	QC	\$83.00	RETAIN	53.0000	83.00
1177057406016	09/21/2000	-		0	1	C1	50	ON	\$529.00	RETAIN	50.0000	529.00
1177066422248	04/26/1999	01/15/2001	NEED	0	1	A2	55	NS	\$44.00	CHURN	55.0000	44.00
1177089399155	08/17/2000	-		1	3	A1	56	BC	\$548.00	RETAIN	56.0000	548.00
1177113886410	09/13/2000	01/08/2001	COMP	0	1	C1	45	ON	\$63.00	CHURN	45.0000	63.00
1177128264924	12/10/1999	-		1	1	B1	38	ON	\$178.00	RETAIN	38.0000	178.00

Task12: Trying to convert the 1 and 0 to "GOOD CREDIT" and "NOT GOOD" for Customer_Credit_Status.

```
PROC SQL;
  CREATE TABLE Project.CustomerBehavior_01c AS
  SELECT * ,
    CASE
      WHEN Customer_Credit_Status EQ 1 THEN "GOOD CREDIT"
      ELSE "NOT GOOD"
    END AS NM_Customer_Credit_Status
  FROM Project.CustomerBehavior_01b
  ;
QUIT;

PROC PRINT DATA = Project.CustomerBehavior_01c (OBS =20) ;
RUN;
```

Account_Deactivation_Date	Reason_For_Deactivation	Customer_Credit_Status	Rate_Plan	Dealer_Type	Customer_Age	Province	Sales_Amount	Churn_Status	NM_Customer_Age	NM_Sales_Amount	NM_Customer_Credit_Status
		0	1	A1	58	BC	\$128.00	RETAIN	58.0000	128.00	NOT GOOD
10/15/1999	NEED	1	1	A1	45	AB	\$72.00	CHURN	45.0000	72.00	GOOD CREDIT
		0	1	A1	57	BC	\$593.00	RETAIN	57.0000	593.00	NOT GOOD
		1	2	A1	47	ON	\$83.00	RETAIN	47.0000	83.00	GOOD CREDIT
		1	1	C1	82	BC		RETAIN	82.0000	181.25	GOOD CREDIT
09/18/2000	MOVE	1	1	C1	92	QC	\$1041.00	CHURN	92.0000	1041.00	GOOD CREDIT
		1	1	A1	77	ON		RETAIN	77.0000	181.25	GOOD CREDIT
		1	1	C1	68	AB	\$72.00	RETAIN	68.0000	72.00	GOOD CREDIT
		0	1	B1	75	ON	\$134.00	RETAIN	75.0000	134.00	NOT GOOD
		1	2	A1	42	NS	\$11.00	RETAIN	42.0000	11.00	GOOD CREDIT
		1	1	A1	26	BC	\$154.00	RETAIN	26.0000	154.00	GOOD CREDIT
		1	1	B1	73	BC	\$16.00	RETAIN	73.0000	16.00	GOOD CREDIT
		0	1	C1		QC	\$179.00	RETAIN	47.5922	179.00	NOT GOOD
		0	1	C1	41	ON	\$705.00	RETAIN	41.0000	705.00	NOT GOOD
		1	1	A1	53	QC	\$83.00	RETAIN	53.0000	83.00	GOOD CREDIT
		0	1	C1	50	ON	\$529.00	RETAIN	50.0000	529.00	NOT GOOD
01/15/2001	NEED	0	1	A2	55	NS	\$44.00	CHURN	55.0000	44.00	NOT GOOD
		1	3	A1	56	BC	\$548.00	RETAIN	56.0000	548.00	GOOD CREDIT
01/08/2001	COMP	0	1	C1	45	ON	\$63.00	CHURN	45.0000	63.00	NOT GOOD
		1	1	B1	38	ON	\$178.00	RETAIN	38.0000	178.00	GOOD CREDIT

Task13: Trying to convert the 1, 2 and 3 to "LEVEL1", "LEVEL2" and "LEVEL3" respectively for Customer_Credit_Status.

```

DATA Project.CustomerBehavior_01d;
SET Project.CustomerBehavior_01c;
IF Rate_Plan EQ 1 THEN CH_Rate_Plan = "LEVEL1";
ELSE IF Rate_Plan EQ 2 THEN CH_Rate_Plan = "LEVEL2";
ELSE CH_Rate_Plan = "LEVEL3";
RUN;

```

```

PROC PRINT DATA = Project.CustomerBehavior_01d (OBS =20) ;
RUN;

```

Reason_For_Deactivation	Customer_Credit_Status	Rate_Plan	Dealer_Type	Customer_Age	Province	Sales_Amount	Churn_Status	NM_Customer_Age	NM_Sales_Amount	NM_Customer_Credit_Status	CH_Rate_Plan
	0	1	A1	58	BC	\$128.00	RETAIN	58.0000	128.00	NOT GOOD	LEVEL1
NEED 	1	1	A1	45	AB	\$72.00	CHURN	45.0000	72.00	GOOD CREDIT	LEVEL1
	0	1	A1	57	BC	\$593.00	RETAIN	57.0000	593.00	NOT GOOD	LEVEL1
	1	2	A1	47	ON	\$83.00	RETAIN	47.0000	83.00	GOOD CREDIT	LEVEL2
	1	1	C1	82	BC	.	RETAIN	82.0000	181.25	GOOD CREDIT	LEVEL1
MOVE	1	1	C1	92	QC	\$1041.00	CHURN	92.0000	1041.00	GOOD CREDIT	LEVEL1
	1	1	A1	77	ON	.	RETAIN	77.0000	181.25	GOOD CREDIT	LEVEL1
	1	1	C1	68	AB	\$72.00	RETAIN	68.0000	72.00	GOOD CREDIT	LEVEL1
	0	1	B1	75	ON	\$134.00	RETAIN	75.0000	134.00	NOT GOOD	LEVEL1
	1	2	A1	42	NS	\$11.00	RETAIN	42.0000	11.00	GOOD CREDIT	LEVEL2
	1	1	A1	26	BC	\$154.00	RETAIN	26.0000	154.00	GOOD CREDIT	LEVEL1
	1	1	B1	73	BC	\$16.00	RETAIN	73.0000	16.00	GOOD CREDIT	LEVEL1
	0	1	C1	.	QC	\$179.00	RETAIN	47.5922	179.00	NOT GOOD	LEVEL1
	0	1	C1	41	ON	\$705.00	RETAIN	41.0000	705.00	NOT GOOD	LEVEL1
	1	1	A1	53	QC	\$83.00	RETAIN	53.0000	83.00	GOOD CREDIT	LEVEL1
	0	1	C1	50	ON	\$529.00	RETAIN	50.0000	529.00	NOT GOOD	LEVEL1
NEED	0	1	A2	55	NS	\$44.00	CHURN	55.0000	44.00	NOT GOOD	LEVEL1
	1	3	A1	56	BC	\$548.00	RETAIN	56.0000	548.00	GOOD CREDIT	LEVEL3
COMP	0	1	C1	45	ON	\$63.00	CHURN	45.0000	63.00	NOT GOOD	LEVEL1
	1	1	B1	38	ON	\$178.00	RETAIN	38.0000	178.00	GOOD CREDIT	LEVEL1

Task14: Let us drop the unnecessary columns and rearrange columns name for now.

```
DATA Project.CustomerBehavior_01e (DROP=Customer_Credit_Status Rate_Plan Customer_Age Sales_Amount) ;
SET Project.CustomerBehavior_01d;

rename NM_Customer_Credit_Status=Customer_Credit_Status ;
rename CH_Rate_Plan=Rate_Plan ;
rename NM_Customer_Age=Customer_Age ;
rename NM_Sales_Amount=Sales_Amount ;
```

RUN;

```
PROC PRINT DATA = Project.CustomerBehavior_01e (OBS =20) ;
RUN;
```

Obs	Account_Number	Account_Activation_Date	Account_Deactivation_Date	Reason_For_Deactivation	Dealer_Type	Province	Churn_Status	Customer_Age	Sales_Amount	Customer_Credit_Status	Rate_Plan	
1	1176913194483		06/20/1999	-	A1	BC	RETAIN	58.0000	128.00	NOT GOOD	LEVEL1	
2	1176914599423		10/04/1999	10/15/1999	NEED	A1	AB	CHURN	45.0000	72.00	GOOD CREDIT	LEVEL1
3	1176951913656		07/01/2000	-	A1	BC	RETAIN	57.0000	593.00	NOT GOOD	LEVEL1	
4	1176954000288		05/30/2000	-	A1	ON	RETAIN	47.0000	83.00	GOOD CREDIT	LEVEL2	
5	1176969186303		12/13/2000	-	C1	BC	RETAIN	82.0000	181.25	GOOD CREDIT	LEVEL1	
6	1176991056273		08/31/1999	09/18/2000	MOVE	C1	QC	CHURN	92.0000	1041.00	GOOD CREDIT	LEVEL1
7	1176991866552		05/24/2000	-	A1	ON	RETAIN	77.0000	181.25	GOOD CREDIT	LEVEL1	
8	1176992889500		11/28/2000	-	C1	AB	RETAIN	68.0000	72.00	GOOD CREDIT	LEVEL1	
9	1177000067271		12/23/1999	-	B1	ON	RETAIN	75.0000	134.00	NOT GOOD	LEVEL1	
10	1177010940613		12/09/1999	-	A1	NS	RETAIN	42.0000	11.00	GOOD CREDIT	LEVEL2	
11	1177025997013		11/09/1999	-	A1	BC	RETAIN	26.0000	154.00	GOOD CREDIT	LEVEL1	
12	1177027515760		10/19/1999	-	B1	BC	RETAIN	73.0000	16.00	GOOD CREDIT	LEVEL1	
13	1177028996676		09/21/2000	-	C1	QC	RETAIN	47.5922	179.00	NOT GOOD	LEVEL1	
14	1177038747105		03/14/2000	-	C1	ON	RETAIN	41.0000	705.00	NOT GOOD	LEVEL1	
15	1177045857516		06/22/2000	-	A1	QC	RETAIN	53.0000	83.00	GOOD CREDIT	LEVEL1	
16	1177057406016		09/21/2000	-	C1	ON	RETAIN	50.0000	529.00	NOT GOOD	LEVEL1	
17	1177066422248		04/26/1999	01/15/2001	NEED	A2	NS	CHURN	55.0000	44.00	NOT GOOD	LEVEL1
18	1177089399155		08/17/2000	-	A1	BC	RETAIN	56.0000	548.00	GOOD CREDIT	LEVEL3	
19	1177113886410		09/13/2000	01/08/2001	COMP	C1	ON	CHURN	45.0000	63.00	NOT GOOD	LEVEL1
20	1177128264924		12/10/1999	-	B1	ON	RETAIN	38.0000	178.00	GOOD CREDIT	LEVEL1	

Task15: Let us check the number of Account_Number, Reason_For_Deactivation, Customer_Credit_Status, Rate_Plan, Dealer_Type and Province.

```

PROC SQL;
  SELECT count(distinct Account_Number) as Total_Unique_Account_Number,
         count(distinct Reason_For_Deactivation) as Unique_Reason,
         count(distinct Customer_Credit_Status) as Credit_Status,
         count(distinct Rate_Plan) as Rate_Plan,
         count(distinct Dealer_Type) as Dealer_Type,
         count(distinct Province) as Province
  FROM Project.CustomerBehavior_01e;
QUIT;

```

Total_Unique_Account_Number	Unique_Reason	Credit_Status	Rate_Plan	Dealer_Type	Province
102255	5	2	3	4	5

Task16: Let us determine Earliest/Latest Activation, Earliest/Latest Deactivation dates and also minimum and maximum of Customer_age and Sales_Amount.

□ **PROC SQL;**

```

SELECT min(Account_Activation_Date) as Earliest_Activation format=MMDDYY10.,
       max(Account_Activation_Date) as Latest_Activation format=MMDDYY10.,
       min(Account_Deactivation_Date) as Earliest_Deactivation format=MMDDYY10.,
       max(Account_Deactivation_Date) as Latest_Deactivation format=MMDDYY10.,
       min(Customer_Age) as Min_Customer_Age,
       max(Customer_Age) as Max_Customer_Age,
       min(Sales_Amount) as Min_Sales_Amount format=dollar8.2,
       max(Sales_Amount) as Max_Sales_Amount format=dollar8.2

FROM Project.CustomerBehavior_01e;
QUIT;|

```

Earliest_Activation	Latest_Activation	Earliest_Deactivation	Latest_Deactivation	Min_Customer_Age	Max_Customer_Age	Min_Sales_Amount	Max_Sales_Amount
01/20/1999	01/20/2001	01/25/1999	01/20/2001	0	99	\$0.00	\$1200.00

Task17: Number of total activation and deactivation during the time period of the study.

□ **PROC SQL;**

```

SELECT count(Account_Activation_Date) as Total_Activation,
       count(Account_Deactivation_Date) as Total_Deactivation

FROM Project.CustomerBehavior_01e;
QUIT;|

```

Total_Activation	Total_Deactivation
102255	19635

Task18: Let us have a look at the table for each categorical variable.

```
PROC FREQ DATA = Project.CustomerBehavior_01e;
table Churn_Status;
RUN;
```

```
PROC FREQ DATA = Project.CustomerBehavior_01e;
table Reason_For_Deactivation;
RUN;
```

```
PROC FREQ DATA = Project.CustomerBehavior_01e;
table Customer_Credit_Status;
RUN;
```

```
PROC FREQ DATA = Project.CustomerBehavior_01e;
table Rate_Plan;
RUN;
```

```
PROC FREQ DATA = Project.CustomerBehavior_01e;
table Dealer_Type;
RUN;
```

```
PROC FREQ DATA = Project.CustomerBehavior_01e;
table Province;
RUN;
```

The FREQ Procedure

Churn_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
CHURN	19635	19.20	19635	19.20
RETAIN	82620	80.80	102255	100.00

The FREQ Procedure

Reason_For_Deactivation	Frequency	Percent	Cumulative Frequency	Cumulative Percent
COMP	4722	24.73	4722	24.73
DEBT	4020	21.05	8742	45.79
MOVE	1696	8.88	10438	54.67
NEED	6888	36.08	17326	90.75
TECH	1767	9.25	19093	100.00
Frequency Missing = 83162				

The FREQ Procedure

Customer_Credit_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
GOOD CREDIT	71002	69.44	71002	69.44
NOT GOOD	31253	30.56	102255	100.00

The FREQ Procedure

Rate_Plan	Frequency	Percent	Cumulative Frequency	Cumulative Percent
LEVEL1	68194	66.69	68194	66.69
LEVEL2	20187	19.74	88381	86.43
LEVEL3	13874	13.57	102255	100.00

The FREQ Procedure

Dealer_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A1	56132	54.89	56132	54.89
A2	11255	11.01	67387	65.90
B1	20670	20.21	88057	86.12
C1	14198	13.88	102255	100.00

The FREQ Procedure

Province	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AB	10277	10.67	10277	10.67
BC	22040	22.88	32317	33.54
NS	11529	11.97	43846	45.51
ON	42500	44.11	86346	89.62
QC	10002	10.38	96348	100.00
Frequency Missing = 5907				

Task19: Let us use the below user defined format to convert few numerical column values into some group of values.

```
proc format ;
    value salesg
        low-<100='LT100'
        100- <500='LT500'
        500- <800='LT800'
        800-high='GT800'
    ;
    value ageg
        low-<20='TEEN'
        20- <40='LT40'
        40- <60='LT60'
        60-high='Senior'
    ;
    value tenureg
        low-<31='UPTO_MONTH'
        31- <61='UPTO2MONTHS'
        61- <366='UPTO_YEAR'
        366-high='YEAR+'
    ;
run;

DATA CustomerBehavior_02;
    SET Project.CustomerBehavior_01e;
    format Customer_Age ageg. Sales_Amount salesg.;
RUN;

PROC PRINT DATA = CustomerBehavior_02 (OBS =20);
RUN;
```

Obs	Account_Number	Account_Activation_Date	Account_Deactivation_Date	Reason_For_Deactivation	Dealer_Type	Province	Churn_Status	Customer_Age	Sales_Amount	Customer_Credit_Status	Rate_Plan
1	1176913194483		06/20/1999	-	A1	BC	RETAIN	LT60	LT500	NOT GOOD	LEVEL1
2	1176914599423		10/04/1999	10/15/1999	NEED	A1	AB	CHURN	LT60	LT100	GOOD CREDIT
3	1176951913656		07/01/2000	-	A1	BC	RETAIN	LT60	LT800	NOT GOOD	LEVEL1
4	1176954000288		05/30/2000	-	A1	ON	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL2
5	1176969186303		12/13/2000	-	C1	BC	RETAIN	Senior	LT500	GOOD CREDIT	LEVEL1
6	1176991056273		08/31/1999	09/18/2000	MOVE	C1	QC	CHURN	Senior	GT800	GOOD CREDIT
7	1176991866552		05/24/2000	-	A1	ON	RETAIN	Senior	LT500	GOOD CREDIT	LEVEL1
8	1176992889500		11/28/2000	-	C1	AB	RETAIN	Senior	LT100	GOOD CREDIT	LEVEL1
9	1177000067271		12/23/1999	-	B1	ON	RETAIN	Senior	LT500	NOT GOOD	LEVEL1
10	1177010940613		12/09/1999	-	A1	NS	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL2
11	1177025997013		11/09/1999	-	A1	BC	RETAIN	LT40	LT500	GOOD CREDIT	LEVEL1
12	1177027515760		10/19/1999	-	B1	BC	RETAIN	Senior	LT100	GOOD CREDIT	LEVEL1
13	1177028996676		09/21/2000	-	C1	QC	RETAIN	LT60	LT500	NOT GOOD	LEVEL1
14	1177038747105		03/14/2000	-	C1	ON	RETAIN	LT60	LT800	NOT GOOD	LEVEL1
15	1177045857516		06/22/2000	-	A1	QC	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL1
16	1177057406016		09/21/2000	-	C1	ON	RETAIN	LT60	LT800	NOT GOOD	LEVEL1
17	1177066422248		04/26/1999	01/15/2001	NEED	A2	NS	CHURN	LT60	LT100	NOT GOOD
18	1177089399155		08/17/2000	-	A1	BC	RETAIN	LT60	LT800	GOOD CREDIT	LEVEL3
19	1177113886410		09/13/2000	01/08/2001	COMP	C1	ON	CHURN	LT60	LT100	NOT GOOD
20	1177128264924		12/10/1999	-	B1	ON	RETAIN	LT40	LT500	GOOD CREDIT	LEVEL1

Task20: Let us check the distribution of Customer_Age and Sales_Amount after converting those to categorical variables.

```
PROC FREQ DATA = CustomerBehavior_02;
table Customer_Age;
RUN;
```

```
PROC FREQ DATA = CustomerBehavior_02;
table Sales_Amount;
RUN;
```

The FREQ Procedure

Customer_Age	Frequency	Percent	Cumulative Frequency	Cumulative Percent
TEEN	6318	6.18	6318	6.18
LT40	25518	24.96	31836	31.13
LT60	45409	44.41	77245	75.54
Senior	25010	24.46	102255	100.00

The FREQ Procedure

Sales_Amount	Frequency	Percent	Cumulative Frequency	Cumulative Percent
LT100	52376	51.22	52376	51.22
LT500	40705	39.81	93081	91.03
LT800	4933	4.82	98014	95.85
GT800	4241	4.15	102255	100.00

Task21: Let us calculate the Tenure of the customer in number of days.We have considered partial days as well.

```
PROC SQL;
CREATE TABLE CustomerBehavior_03 AS
SELECT * ,
CASE
WHEN Account_Deactivation_Date EQ . THEN (max(Account_Activation_Date) - Account_Activation_Date + 1)
ELSE (Account_Deactivation_Date - Account_Activation_Date + 1)
END AS Tenure_in_Days
FROM CustomerBehavior_02
;
QUIT;

PROC PRINT DATA = CustomerBehavior_03 (OBS =20) ;
RUN;
```

Obs	Account_Number	Account_Activation_Date	Account_Deactivation_Date	Reason_For_Deactivation	Dealer_Type	Province	Churn_Status	Customer_Age	Sales_Amount	Customer_Credit_Status	Rate_Plan	Tenure_in_Days	
1	1176913194483		06/20/1999	-	A1	BC	RETAIN	LT60	LT500	NOT GOOD	LEVEL1	581	
2	1176914599423		10/04/1999	10/15/1999	NEED	A1	AB	CHURN	LT60	LT100	GOOD CREDIT	LEVEL1	12
3	1176951913656		07/01/2000	-	A1	BC	RETAIN	LT60	LT800	NOT GOOD	LEVEL1	204	
4	117695400288		05/30/2000	-	A1	ON	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL2	236	
5	1176969186303		12/13/2000	-	C1	BC	RETAIN	Senior	LT500	GOOD CREDIT	LEVEL1	39	
6	1176991056273		08/31/1999	09/18/2000	MOVE	C1	QC	CHURN	Senior	GT800	GOOD CREDIT	LEVEL1	385
7	1176991866552		05/24/2000	-	A1	ON	RETAIN	Senior	LT500	GOOD CREDIT	LEVEL1	242	
8	1176992889500		11/28/2000	-	C1	AB	RETAIN	Senior	LT100	GOOD CREDIT	LEVEL1	54	
9	1177000067271		12/23/1999	-	B1	ON	RETAIN	Senior	LT500	NOT GOOD	LEVEL1	395	
10	1177010940613		12/09/1999	-	A1	NS	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL2	409	
11	1177025997013		11/09/1999	-	A1	BC	RETAIN	LT40	LT500	GOOD CREDIT	LEVEL1	439	
12	1177027515760		10/19/1999	-	B1	BC	RETAIN	Senior	LT100	GOOD CREDIT	LEVEL1	460	
13	1177028996676		09/21/2000	-	C1	QC	RETAIN	LT60	LT500	NOT GOOD	LEVEL1	122	
14	1177038747105		03/14/2000	-	C1	ON	RETAIN	LT60	LT800	NOT GOOD	LEVEL1	313	
15	1177045857516		06/22/2000	-	A1	QC	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL1	213	
16	1177057406016		09/21/2000	-	C1	ON	RETAIN	LT60	LT800	NOT GOOD	LEVEL1	122	
17	1177066422248		04/26/1999	01/15/2001	NEED	A2	NS	CHURN	LT60	LT100	NOT GOOD	LEVEL1	631
18	1177089399155		08/17/2000	-	A1	BC	RETAIN	LT60	LT800	GOOD CREDIT	LEVEL3	157	
19	1177113886410		09/13/2000	01/08/2001	COMP	C1	ON	CHURN	LT60	LT100	NOT GOOD	LEVEL1	118
20	1177128264924		12/10/1999	-	B1	ON	RETAIN	LT40	LT500	GOOD CREDIT	LEVEL1	408	

Task22: Let us check the maximum and minimum tenure considering the last date of the dataset.

```

PROC SQL;
  SELECT min(Tenure_in_Days) as Min_Tenure_in_Days,
         max(Tenure_in_Days) as Max_Tenure_in_Days
  FROM CustomerBehavior_03;
QUIT;

```

Min_Tenure_in_Days	Max_Tenure_in_Days
1	732

Task23: Let us convert the Tenure in number of days to categories as instructed.

```
DATA CustomerBehavior_04;
  SET CustomerBehavior_03;
  format Tenure_in_Days tenureeg.;

RUN;

PROC PRINT DATA = CustomerBehavior_04 (OBS =20) ;
RUN;

PROC FREQ DATA = CustomerBehavior_04;
  table Tenure_in_Days;
RUN;|
```

Obs	Account_Number	Account_Activation_Date	Account_Deactivation_Date	Reason_For_Deactivation	Dealer_Type	Province	Churn_Status	Customer_Age	Sales_Amount	Customer_Credit_Status	Rate_Plan	Tenure_in_Days
1	1176913194483	06/20/1999	-		A1	BC	RETAIN	LT60	LT500	NOT GOOD	LEVEL1	YEAR+
2	1176914599423	10/04/1999	10/15/1999	NEED	A1	AB	CHURN	LT60	LT100	GOOD CREDIT	LEVEL1	UPTO_MONTH
3	1176951913656	07/01/2000	-		A1	BC	RETAIN	LT60	LT800	NOT GOOD	LEVEL1	UPTO_YEAR
4	117695400288	05/30/2000	-		A1	ON	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL2	UPTO_YEAR
5	1176969186303	12/13/2000	-		C1	BC	RETAIN	Senior	LT500	GOOD CREDIT	LEVEL1	UPTO2MONTHS
6	1176991056273	08/31/1999	09/18/2000	MOVE	C1	QC	CHURN	Senior	GT800	GOOD CREDIT	LEVEL1	YEAR+
7	1176991866552	05/24/2000	-		A1	ON	RETAIN	Senior	LT500	GOOD CREDIT	LEVEL1	UPTO_YEAR
8	1176992889500	11/28/2000	-		C1	AB	RETAIN	Senior	LT100	GOOD CREDIT	LEVEL1	UPTO2MONTHS
9	1177000067271	12/23/1999	-		B1	ON	RETAIN	Senior	LT500	NOT GOOD	LEVEL1	YEAR+
10	1177010940613	12/09/1999	-		A1	NS	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL2	YEAR+
11	1177025997013	11/09/1999	-		A1	BC	RETAIN	LT40	LT500	GOOD CREDIT	LEVEL1	YEAR+
12	1177027515760	10/19/1999	-		B1	BC	RETAIN	Senior	LT100	GOOD CREDIT	LEVEL1	YEAR+
13	1177028996676	09/21/2000	-		C1	QC	RETAIN	LT60	LT500	NOT GOOD	LEVEL1	UPTO_YEAR
14	1177038747105	03/14/2000	-		C1	ON	RETAIN	LT60	LT800	NOT GOOD	LEVEL1	UPTO_YEAR
15	1177045857516	06/22/2000	-		A1	QC	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL1	UPTO_YEAR
16	1177057406016	09/21/2000	-		C1	ON	RETAIN	LT60	LT800	NOT GOOD	LEVEL1	UPTO_YEAR
17	1177066422248	04/26/1999	01/15/2001	NEED	A2	NS	CHURN	LT60	LT100	NOT GOOD	LEVEL1	YEAR+
18	1177089399155	08/17/2000	-		A1	BC	RETAIN	LT60	LT800	GOOD CREDIT	LEVEL3	UPTO_YEAR
19	1177113886410	09/13/2000	01/08/2001	COMP	C1	ON	CHURN	LT60	LT100	NOT GOOD	LEVEL1	UPTO_YEAR
20	1177128264924	12/10/1999	-		B1	ON	RETAIN	LT40	LT500	GOOD CREDIT	LEVEL1	YEAR+

The FREQ Procedure

Tenure_in_Days	Frequency	Percent	Cumulative Frequency	Cumulative Percent
UPTO_MONTH	9486	9.28	9486	9.28
UPTO2MONTHS	8586	8.40	18072	17.67
UPTO_YEAR	45405	44.40	63477	62.08
YEAR+	38778	37.92	102255	100.00

Task23: Let us perform some bivariate analysis in visualization.

```
PROC GCHART DATA=CustomerBehavior_04;
vbar Customer_Credit_Status/subgroup=Churn_Status type=percent;
RUN;
```

```
PROC GCHART DATA=CustomerBehavior_04;
vbar Rate_Plan/subgroup=Churn_Status type=percent;
RUN;
```

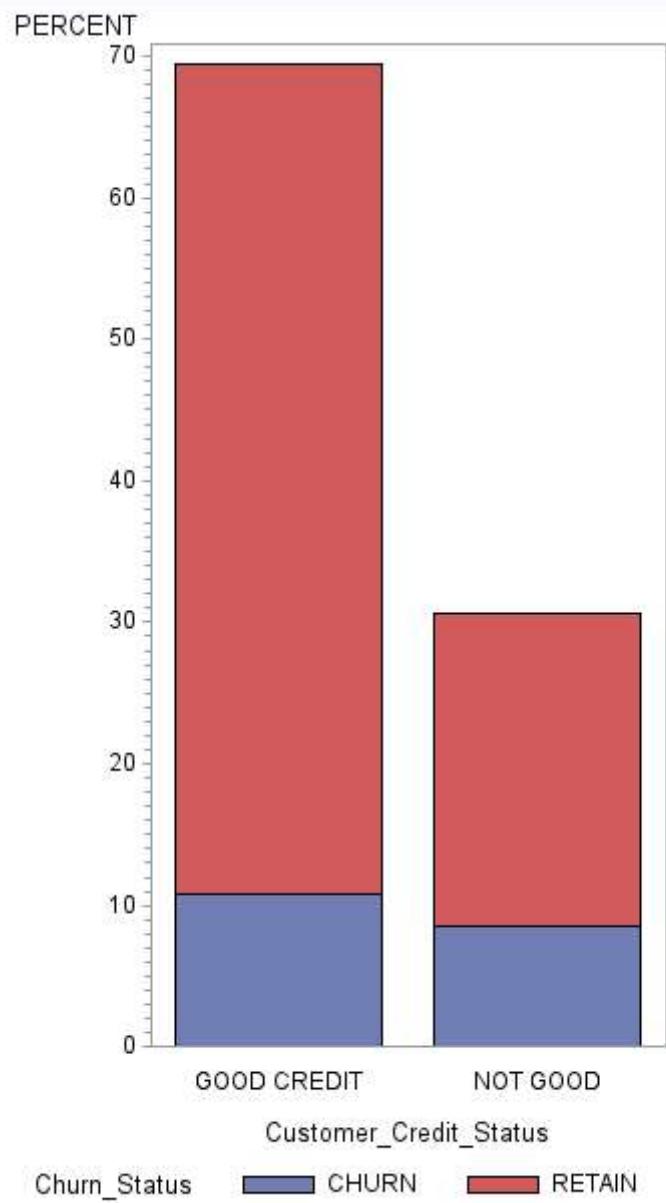
```
PROC GCHART DATA=CustomerBehavior_04;
vbar Dealer_Type/subgroup=Churn_Status type=percent;
RUN;
```

```
proc freq DATA=CustomerBehavior_04 order=freq;
tables Churn_Status*Customer_Age /
plots=freqplot(twoway=stacked orient=horizontal);
run;
```

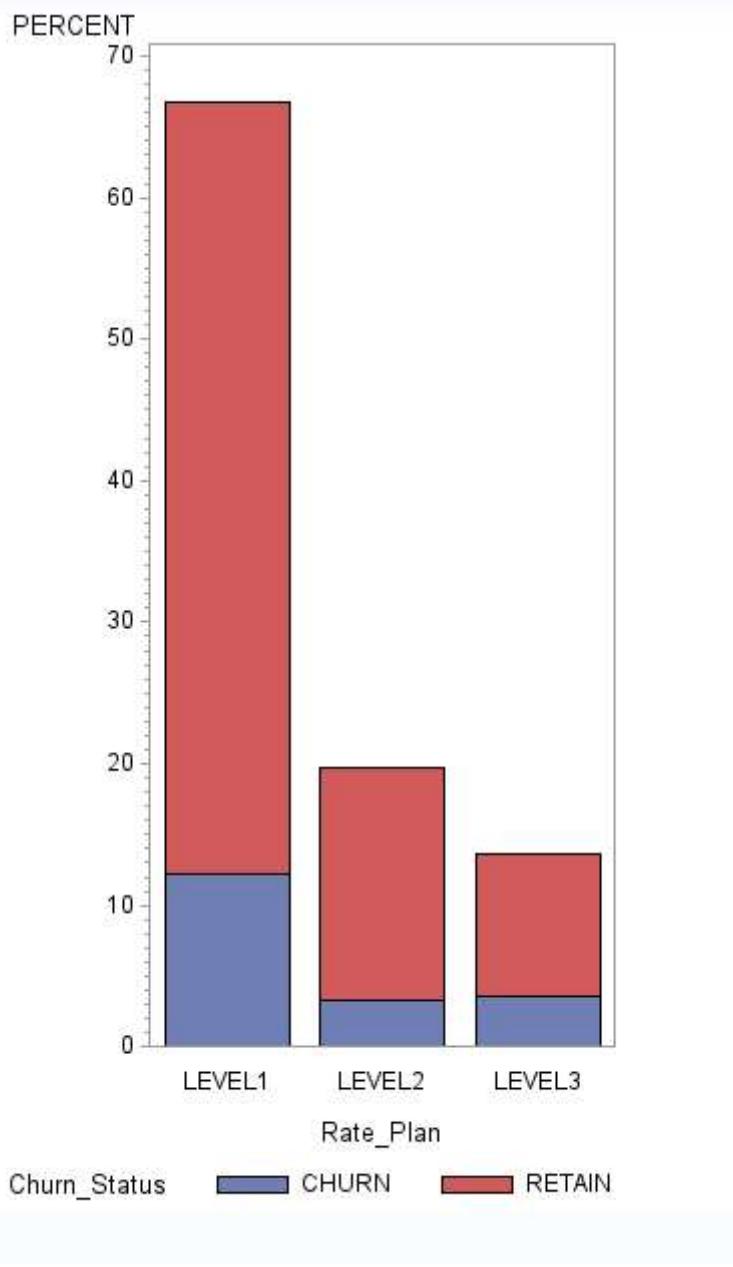
```
proc freq DATA=CustomerBehavior_04 order=freq;
tables Churn_Status*Customer_Age /
plots=freqplot(twoway=stacked orient=horizontal);
run;
```

```
proc freq DATA=CustomerBehavior_04 order=freq;
tables Churn_Status*Sales_Amount /
plots=freqplot(twoway=stacked orient=horizontal);
run;|
```

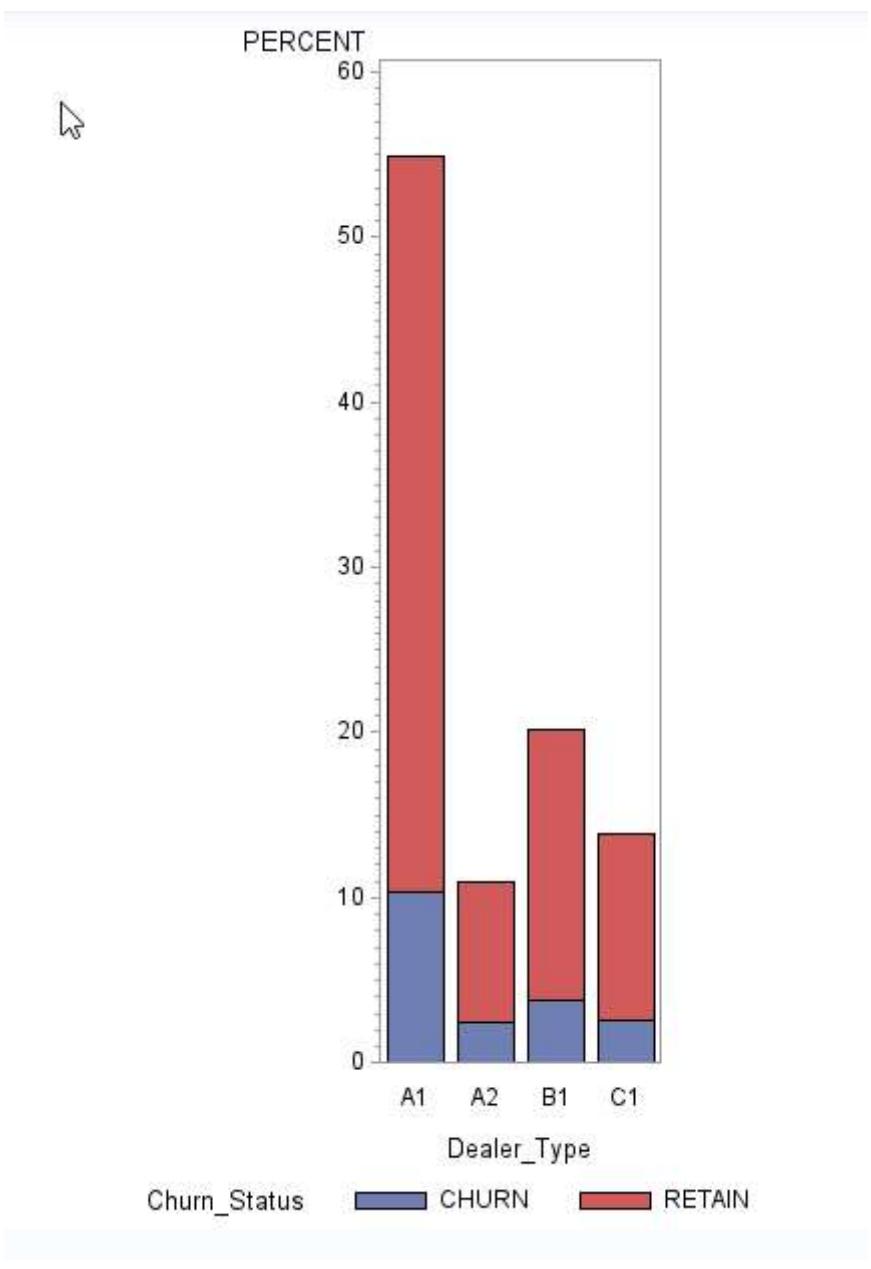
In the below plot it is evident that Churn is much less in case of "GOOD CREDIT"



It seems from the below chart that LEVEL3 is the worst for churn and other two levels are almost equal in percentage.

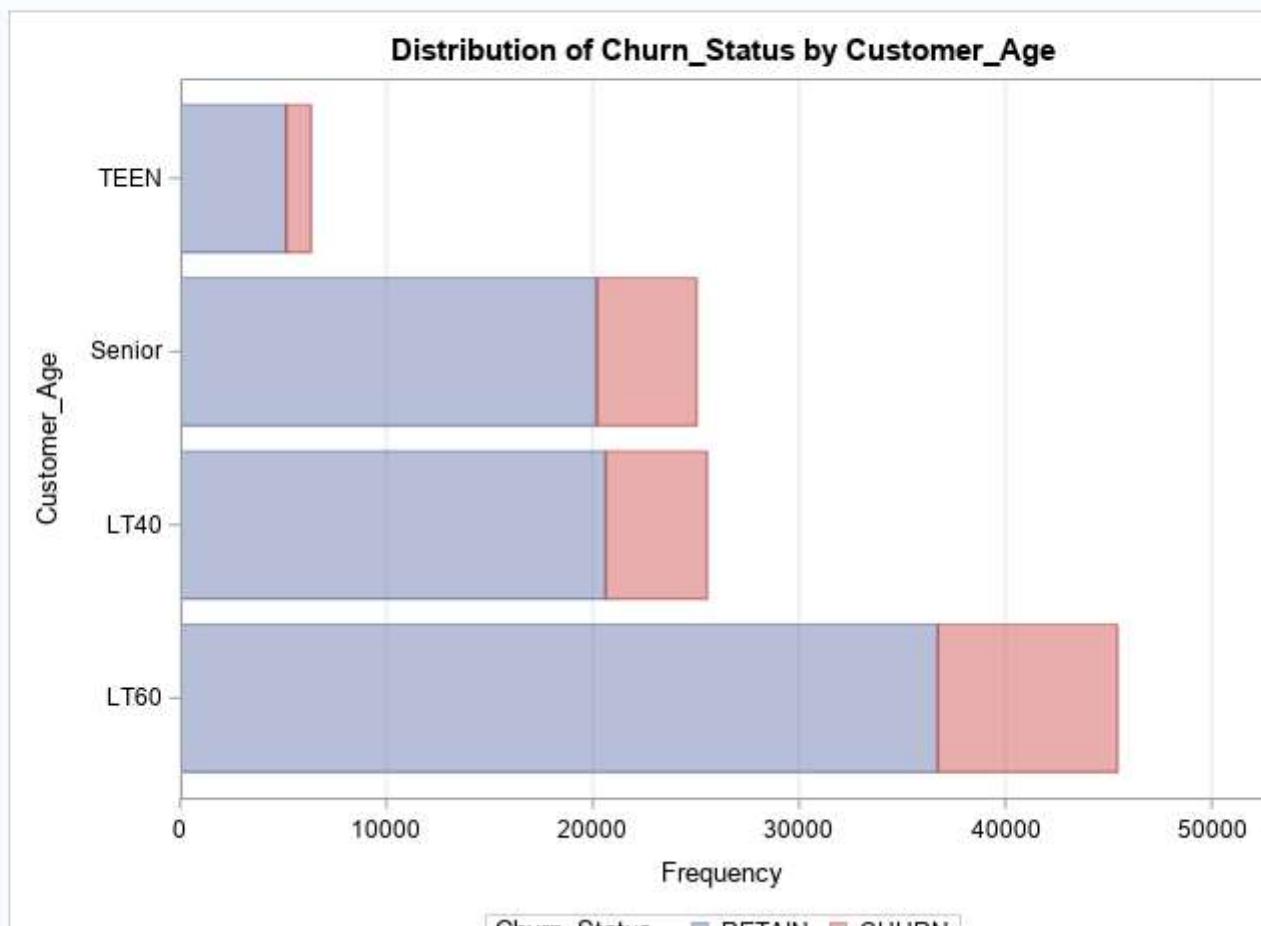


Relation between Dealer_Type and Churn is depicted below:



The FREQ Procedure

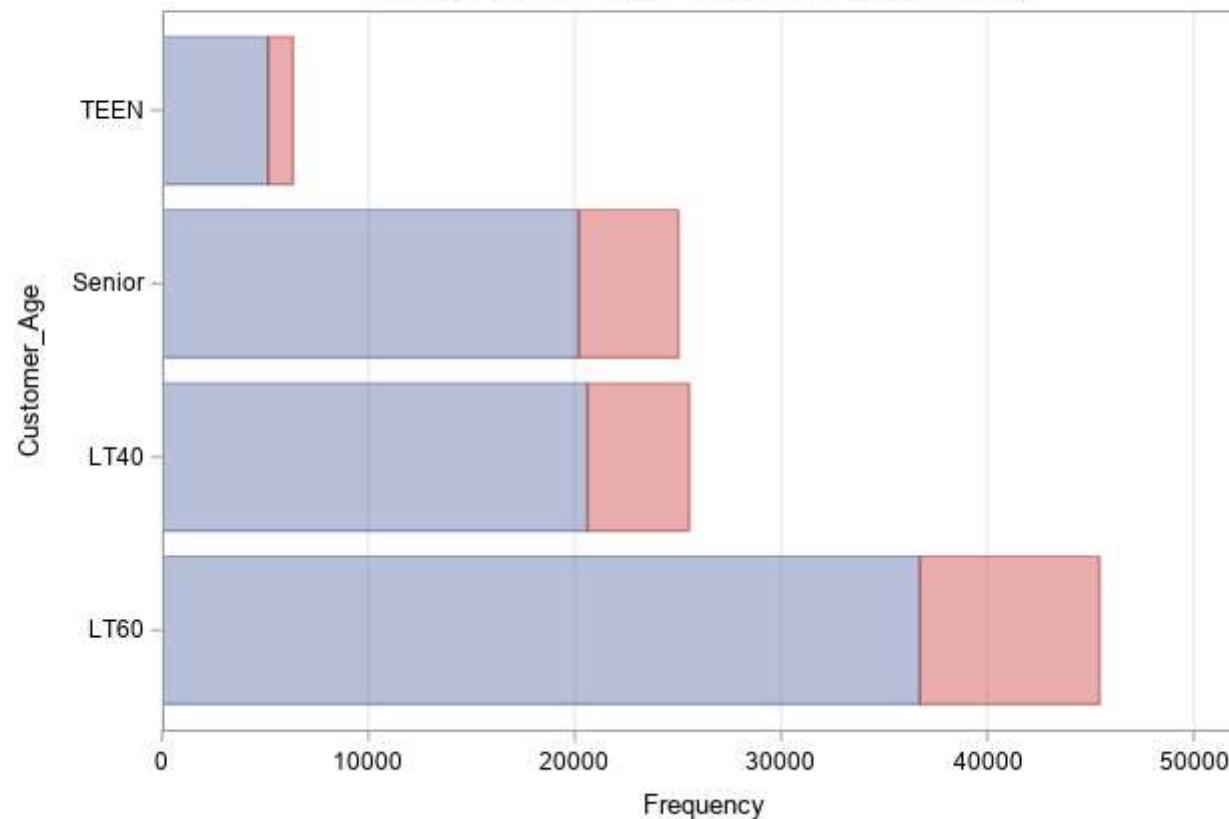
Frequency	Table of Churn_Status by Customer_Age				
	Customer_Age				
Churn_Status	LT60	LT40	Senior	TEEN	Total
RETAIN	36708	20603	20184	5125	82620
	35.90	20.15	19.74	5.01	80.80
	44.43	24.94	24.43	6.20	
	80.84	80.74	80.70	81.12	
CHURN	8701	4915	4826	1193	19635
	8.51	4.81	4.72	1.17	19.20
	44.31	25.03	24.58	6.08	
	19.16	19.26	19.30	18.88	
Total	45409	25518	25010	6318	102255
	44.41	24.96	24.46	6.18	100.00



The FREQ Procedure

Frequency	Table of Churn_Status by Customer_Age					
Percent	Customer_Age					
Row Pct	Churn_Status	LT60	LT40	Senior	TEEN	Total
RETAIN	RETAIN	36708	20603	20184	5125	82620
		35.90	20.15	19.74	5.01	80.80
		44.43	24.94	24.43	6.20	
		80.84	80.74	80.70	81.12	
CHURN	CHURN	8701	4915	4826	1193	19635
		8.51	4.81	4.72	1.17	19.20
		44.31	25.03	24.58	6.08	
		19.16	19.26	19.30	18.88	
Total	Total	45409	25518	25010	6318	102255
		44.41	24.96	24.46	6.18	100.00

Distribution of Churn_Status by Customer_Age



Churn_Status

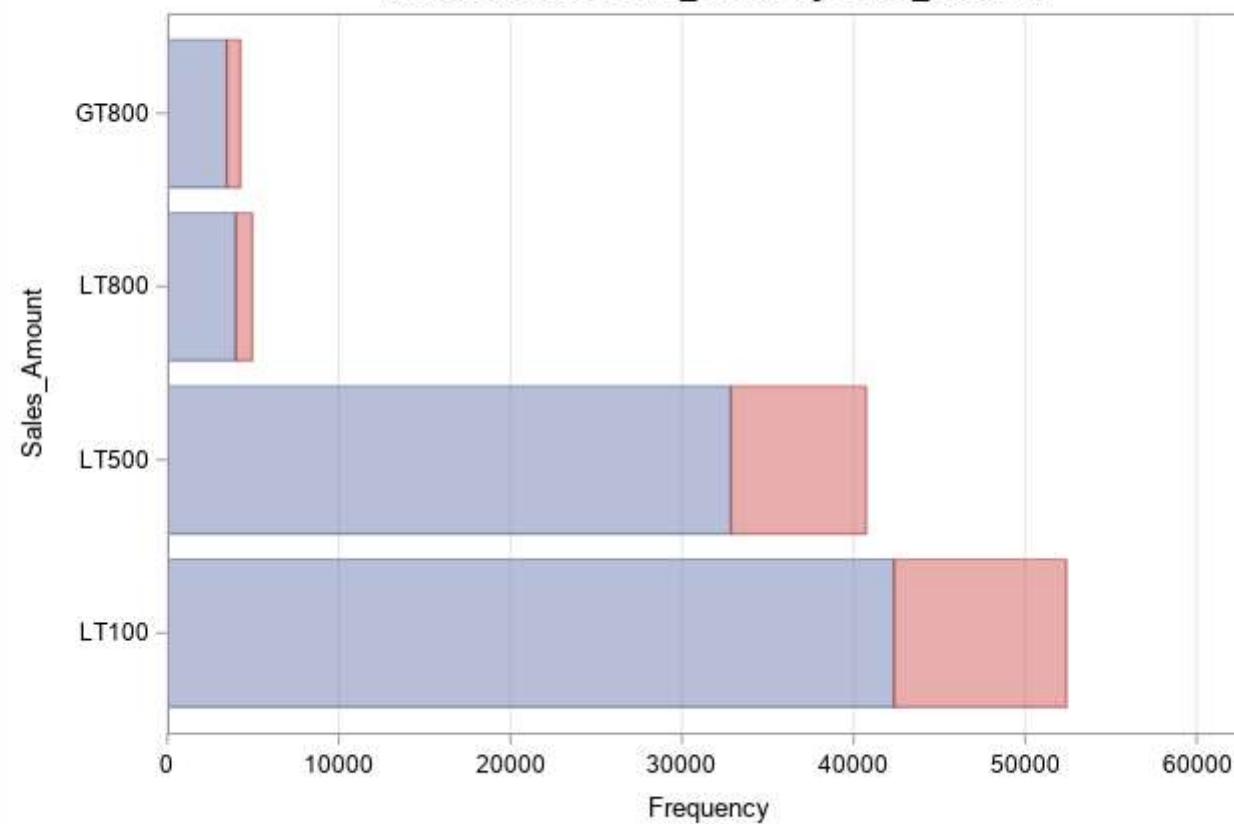
RETAIN

CHURN

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Churn_Status by Sales_Amount					
	Churn_Status	Sales_Amount				
		LT100	LT500	LT800	GT800	Total
	RETAIN	42348	32835	3984	3453	82620
		41.41	32.11	3.90	3.38	80.80
		51.26	39.74	4.82	4.18	
		80.85	80.67	80.76	81.42	
	CHURN	10028	7870	949	788	19635
		9.81	7.70	0.93	0.77	19.20
		51.07	40.08	4.83	4.01	
		19.15	19.33	19.24	18.58	
	Total	52376	40705	4933	4241	102255
		51.22	39.81	4.82	4.15	100.00

Distribution of Churn_Status by Sales_Amount



Task24: Let us now drop the AccountActivationDate and AccountDeactivationDate as the tenure of churned and retained customers have been captured in Tenure in Days column already.

```
✉ DATA CustomerBehavior_05 (DROP=Account_Activation_Date Account_Deactivation_Date) ;  
SET CustomerBehavior_04;  
RUN;  
  
✉ PROC PRINT DATA = CustomerBehavior_05 (OBS =20);  
RUN;|
```

Obs	Account_Number	Reason_For_Deactivation	Dealer_Type	Province	Churn_Status	Customer_Age	Sales_Amount	Customer_Credit_Status	Rate_Plan	Tenure_in_Days
1	1176913194483		A1	BC	RETAIN	LT60	LT500	NOT GOOD	LEVEL1	YEAR+
2	1176914599423	NEED	A1	AB	CHURN	LT60	LT100	GOOD CREDIT	LEVEL1	UPTO_MONTH
3	1176951913656		A1	BC	RETAIN	LT60	LT800	NOT GOOD	LEVEL1	UPTO_YEAR
4	1176954000288		A1	ON	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL2	UPTO_YEAR
5	1176969186303		C1	BC	RETAIN	Senior	LT500	GOOD CREDIT	LEVEL1	UPTO2MONTHS
6	1176991056273	MOVE	C1	QC	CHURN	Senior	GT800	GOOD CREDIT	LEVEL1	YEAR+
7	1176991866552		A1	ON	RETAIN	Senior	LT500	GOOD CREDIT	LEVEL1	UPTO_YEAR
8	1176992889500		C1	AB	RETAIN	Senior	LT100	GOOD CREDIT	LEVEL1	UPTO2MONTHS
9	1177000067271		B1	ON	RETAIN	Senior	LT500	NOT GOOD	LEVEL1	YEAR+
10	1177010940613		A1	NS	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL2	YEAR+
11	1177025997013		A1	BC	RETAIN	LT40	LT500	GOOD CREDIT	LEVEL1	YEAR+
12	1177027515760		B1	BC	RETAIN	Senior	LT100	GOOD CREDIT	LEVEL1	YEAR+
13	1177028996676		C1	QC	RETAIN	LT60	LT500	NOT GOOD	LEVEL1	UPTO_YEAR
14	1177038747105		C1	ON	RETAIN	LT60	LT800	NOT GOOD	LEVEL1	UPTO_YEAR
15	1177045857516		A1	QC	RETAIN	LT60	LT100	GOOD CREDIT	LEVEL1	UPTO_YEAR
16	1177057406016		C1	ON	RETAIN	LT60	LT800	NOT GOOD	LEVEL1	UPTO_YEAR
17	1177066422248	NEED	A2	NS	CHURN	LT60	LT100	NOT GOOD	LEVEL1	YEAR+
18	1177089399155		A1	BC	RETAIN	LT60	LT800	GOOD CREDIT	LEVEL3	UPTO_YEAR
19	1177113886410	COMP	C1	ON	CHURN	LT60	LT100	NOT GOOD	LEVEL1	UPTO_YEAR
20	1177128264924		B1	ON	RETAIN	LT40	LT500	GOOD CREDIT	LEVEL1	YEAR+

Task25: Let us try to analyze relationships among tenure segments and “Good Credit” “RatePlan ” and “DealerType, RatePlan and DealerType.

```
proc freq DATA=CustomerBehavior_05 order=freq;
  tables Customer_Credit_Status*Tenure_in_Days /
    plots=freqplot(twoway=stacked orient=horizontal);
run;
```

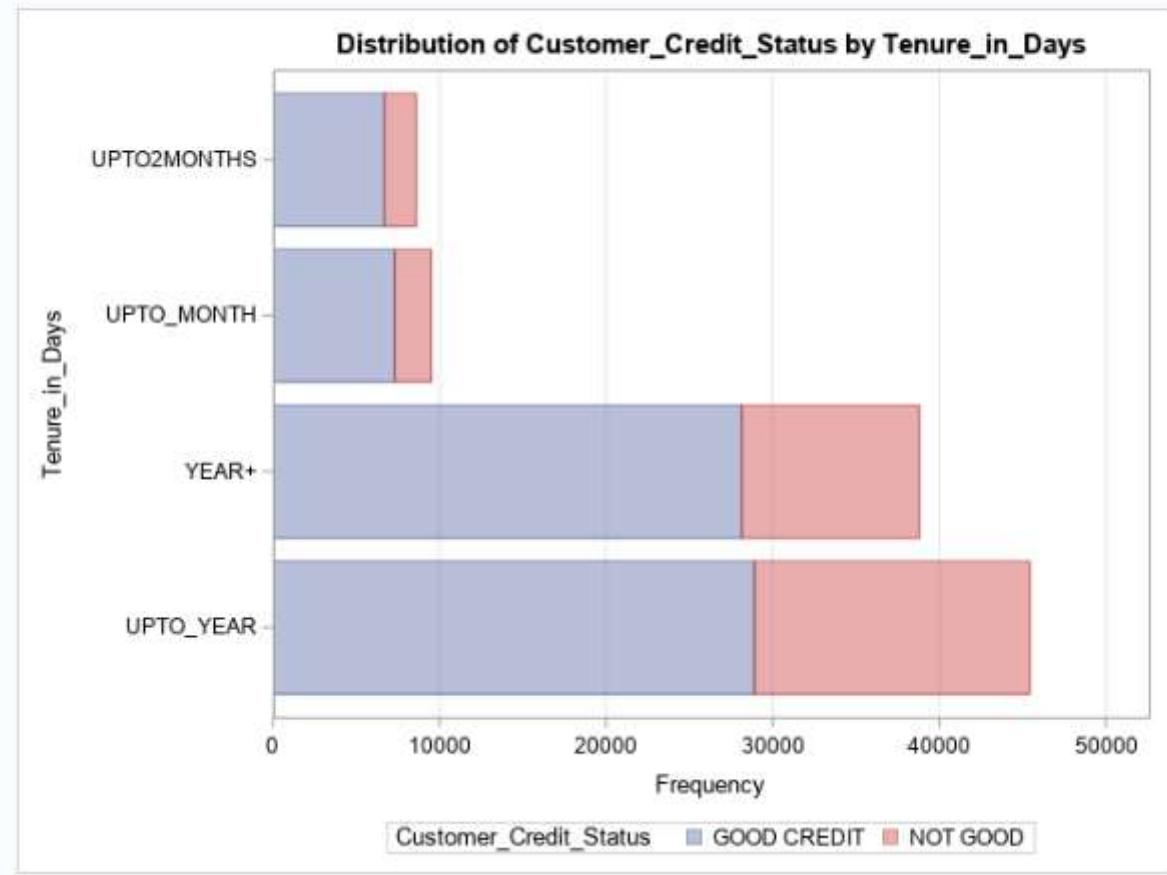
```
proc freq DATA=CustomerBehavior_05 order=freq;
  tables Rate_Plan*Tenure_in_Days /
    plots=freqplot(twoway=stacked orient=horizontal);
run;
```

```
proc freq DATA=CustomerBehavior_05 order=freq;
  tables Dealer_Type*Tenure_in_Days /
    plots=freqplot(twoway=stacked orient=horizontal);
run;
```

```
proc freq DATA=CustomerBehavior_05 order=freq;
  tables Dealer_Type*Rate_Plan /
    plots=freqplot(twoway=stacked orient=horizontal);
run;|
```

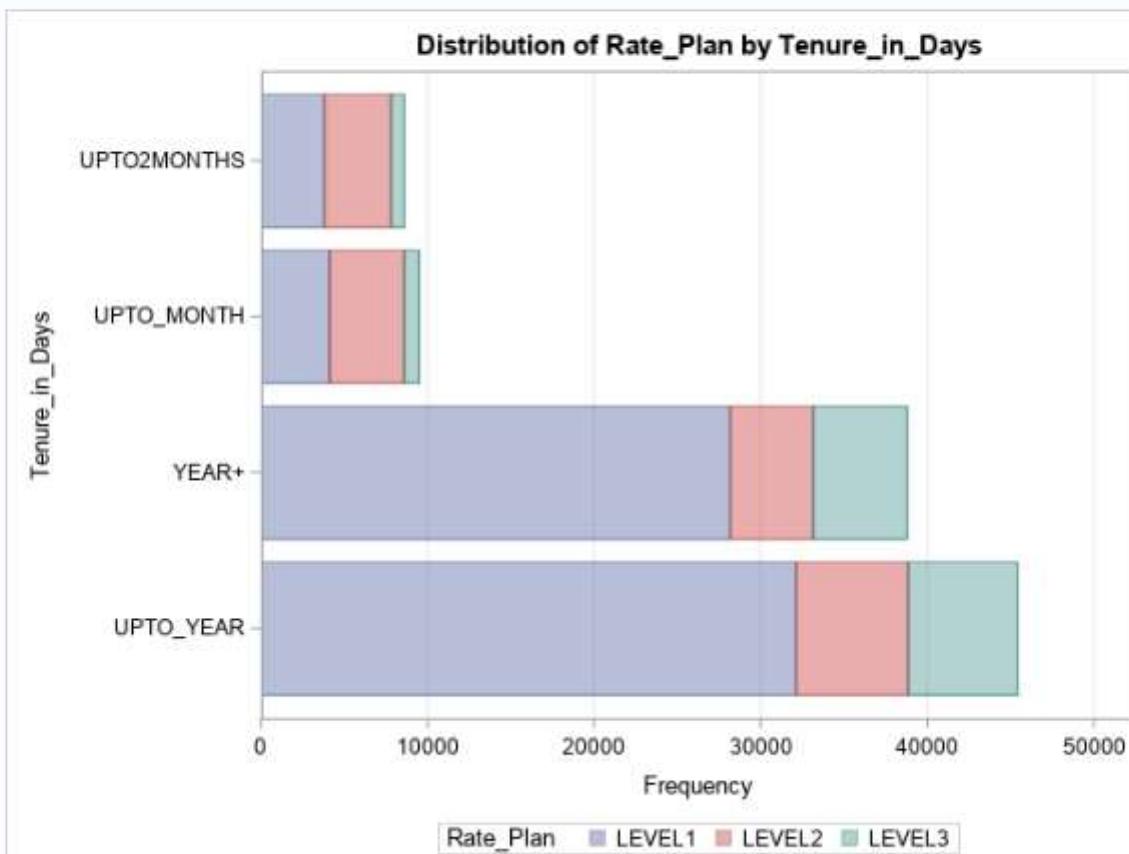
The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Customer_Credit_Status by Tenure_in_Days					
	Customer_Credit_Status	Tenure_in_Days				
		UPTO_YEAR	YEAR+	UPTO_MONTH	UPTO2MONTHS	Total
		28899	28116	7307	6680	71002
28.26	27.50	7.15	6.53	9.41	69.44	
40.70	39.60	10.29	6.10	9.41		
63.65	72.51	77.03	22.20	77.80		
NOT GOOD		16506	10662	2179	1906	31253
16.14	10.43	2.13	1.86	6.10	30.56	
52.81	34.12	6.97				
36.35	27.49	22.97				
Total		45405	38778	9486	8586	102255
		44.40	37.92	9.28	8.40	100.00



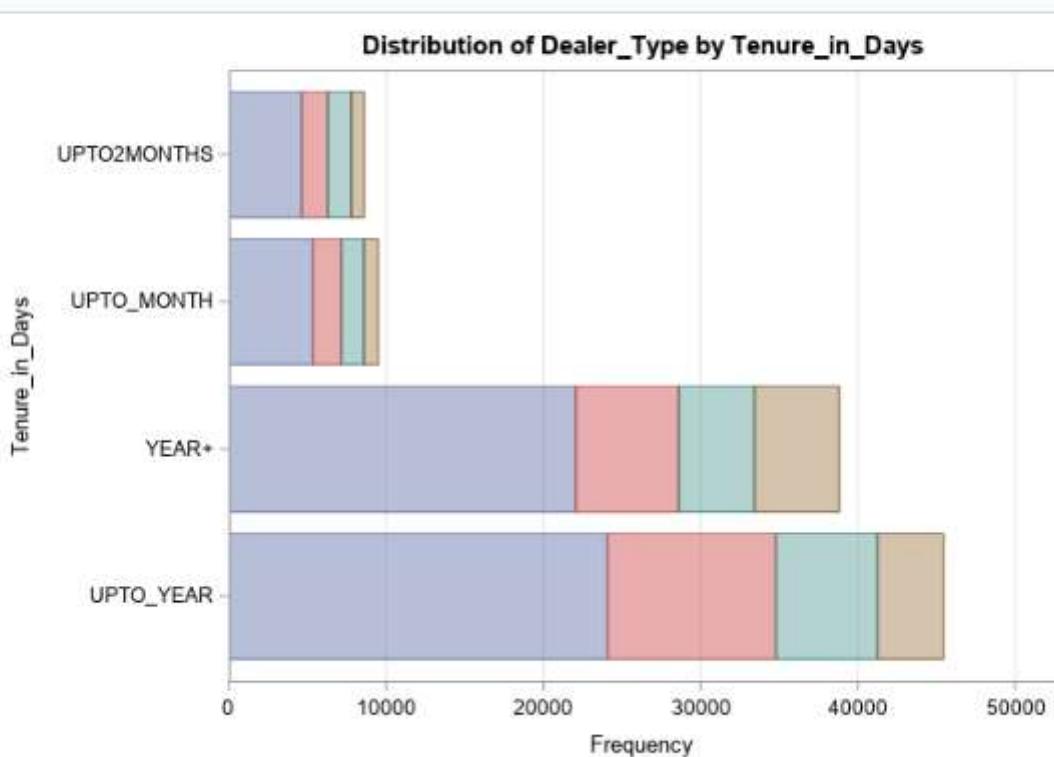
The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Rate_Plan by Tenure_in_Days					
	Rate_Plan	Tenure_in_Days				
		UPTO_YEAR	YEAR+	UPTO_MONTH	UPTO2MONTHS	Total
LEVEL1	LEVEL1	32118	28155	4120	3801	68194
		31.41	27.53	4.03	3.72	66.69
		47.10	41.29	6.04	5.57	
		70.74	72.61	43.43	44.27	
LEVEL2	LEVEL2	6726	4994	4463	4004	20187
		6.58	4.88	4.36	3.92	19.74
		33.32	24.74	22.11	19.83	
		14.81	12.88	47.05	46.63	
LEVEL3	LEVEL3	6561	5629	903	781	13874
		6.42	5.50	0.88	0.76	13.57
		47.29	40.57	6.51	5.63	
		14.45	14.52	9.52	9.10	
Total	Total	45405	38778	9486	8586	102255
		44.40	37.92	9.28	8.40	100.00



The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Dealer_Type by Tenure_in_Days					
	Dealer_Type	Tenure_in_Days				
		UPTO_YEAR	YEAR+	UPTO_MONTH	UPTO2MONTHS	Total
		24071	22054	5361	4646	56132
23.54	21.57	5.24	4.54	54.89		
42.88	39.29	9.55	8.28			
53.01	56.87	56.51	54.11			
A1						
B1	10698	6540	1787	1645	20670	
	10.46	6.40	1.75	1.61	20.21	
	51.76	31.64	8.65	7.96		
	23.56	16.87	18.84	19.16		
C1	6447	4821	1429	1501	14198	
	6.30	4.71	1.40	1.47	13.88	
	45.41	33.96	10.06	10.57		
	14.20	12.43	15.06	17.48		
A2	4189	5363	909	794	11255	
	4.10	5.24	0.89	0.78	11.01	
	37.22	47.65	8.08	7.05		
	9.23	13.83	9.58	9.25		
Total	45405	38778	9486	8586	102255	
	44.40	37.92	9.28	8.40	100.00	

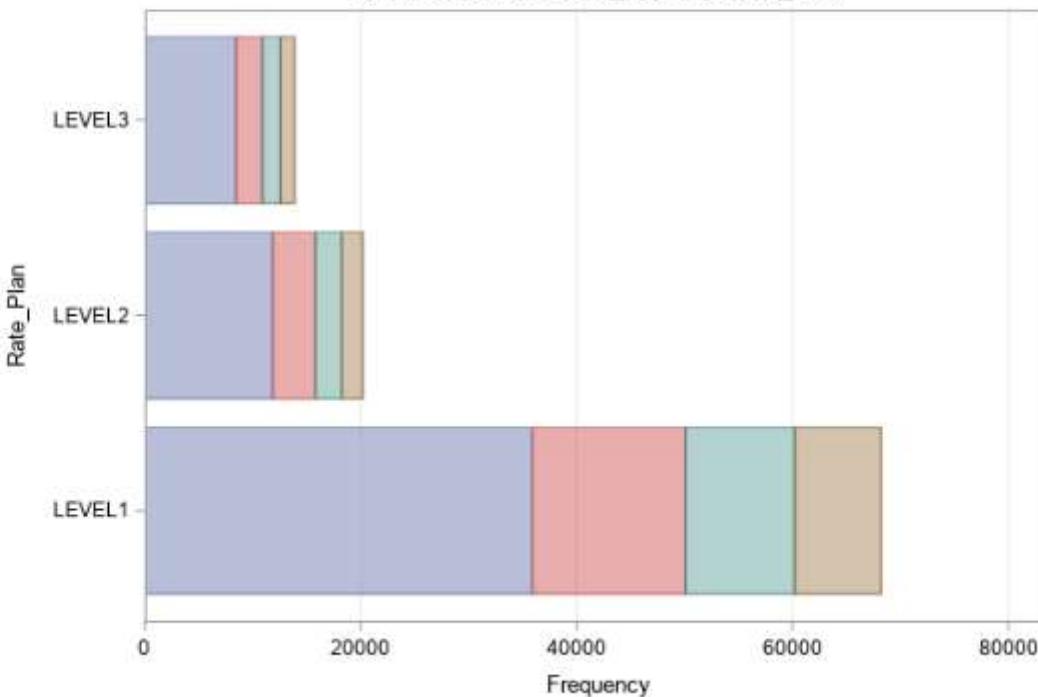


Dealer_Type A1 B1 C1 A2

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Dealer_Type by Rate_Plan				
	Dealer_Type	Rate_Plan			
		LEVEL1	LEVEL2	LEVEL3	Total
	A1	35875 35.08 63.91 52.61	11837 11.58 21.09 58.64	8420 8.23 15.00 60.69	56132 54.89
	B1	14239 13.92 68.89 20.88	3952 3.86 19.12 19.58	2479 2.42 11.99 17.87	20670 20.21
	C1	10105 9.88 71.17 14.82	2448 2.39 17.24 12.13	1645 1.61 11.59 11.86	14198 13.88
	A2	7975 7.80 70.86 11.69	1950 1.91 17.33 9.66	1330 1.30 11.82 9.59	11255 11.01
	Total	68194 66.69	20187 19.74	13874 13.57	102255 100.00

Distribution of Dealer_Type by Rate_Plan

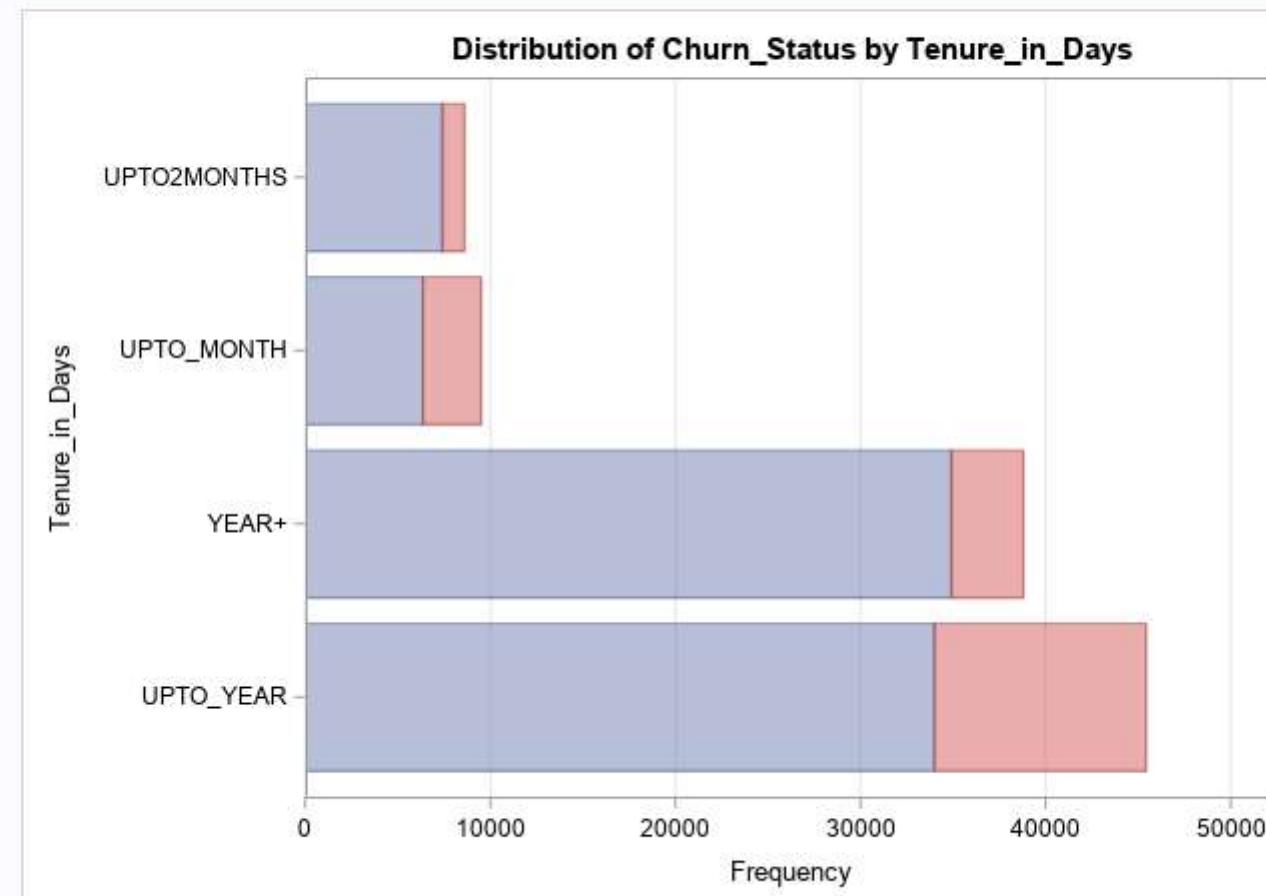


Task26: Let us try to analyze relationships between account status(Churn/Retain) and Tenure Segments.

```
proc freq DATA=CustomerBehavior_05 order=freq;
  tables Churn_Status*Tenure_in_Days /
    plots=freqplot(twoway=stacked orient=horizontal);
run;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Churn_Status by Tenure_in_Days					
	Churn_Status	Tenure_in_Days				
		UPTO_YEAR	YEAR+	UPTO_MONTH	UPTO2MONTHS	
		33977	34892	6353	7398	82620
33.23	34.12	6.21	7.23	80.80		
41.12	42.23	7.69	8.95			
74.83	89.98	66.97	86.16			
CHURN	11428	3886	3133	1188	19635	
11.18	3.80	3.06	1.16	19.20		
58.20	19.79	15.96	6.05			
25.17	10.02	33.03	13.84			
Total	45405	38778	9486	8586	102255	
44.40	37.92	9.28	8.40	100.00		



Task27: Does Sales amount differ among different account status, GoodCredit, and customer age segments?

```
proc freq DATA=CustomerBehavior_05 order=freq;
  tables Churn_Status*Sales_Amount /
    plots=freqplot(twoway=stacked orient=horizontal);
run;

proc freq DATA=CustomerBehavior_05 order=freq;
  tables Customer_Credit_Status*Sales_Amount /
    plots=freqplot(twoway=stacked orient=horizontal);
run;

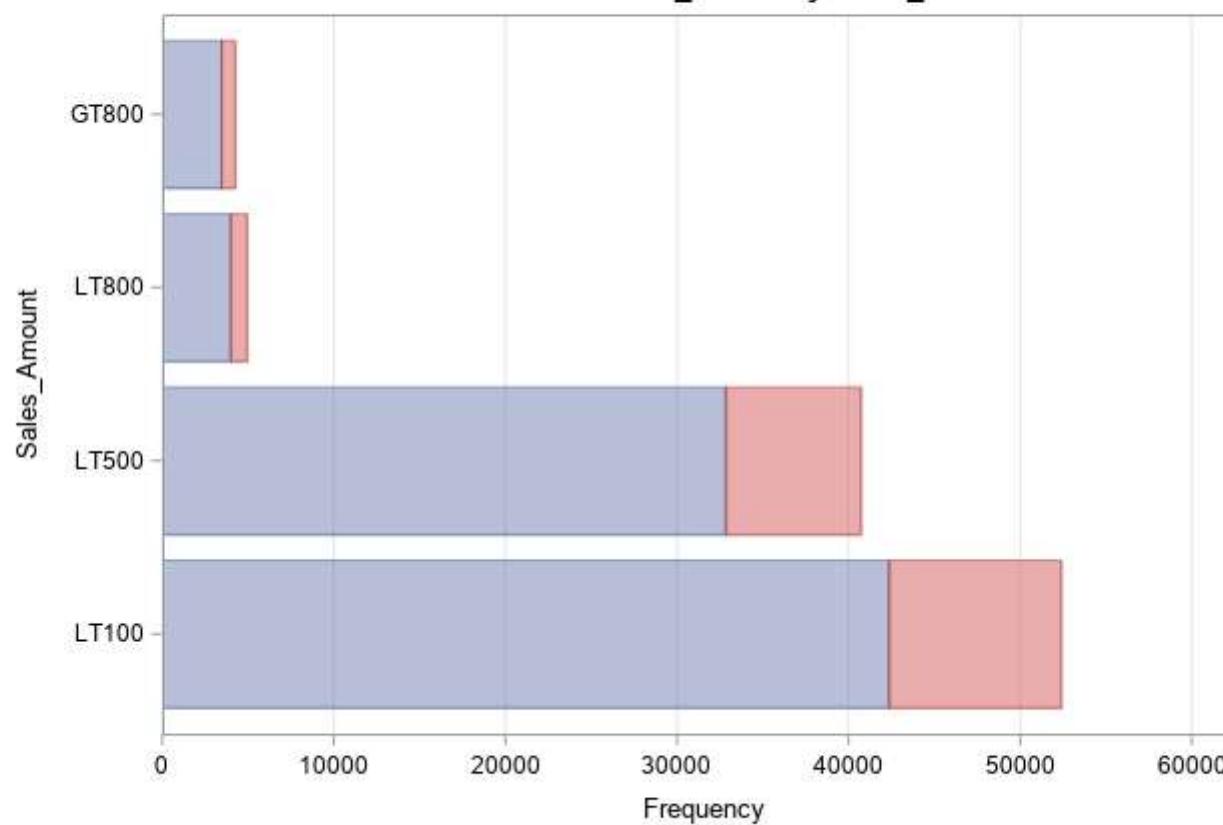
proc freq DATA=CustomerBehavior_05 order=freq;
  tables Customer_Age*Sales_Amount /
    plots=freqplot(twoway=stacked orient=horizontal);
run;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Churn_Status by Sales_Amount					
	Churn_Status	Sales_Amount				
		LT100	LT500	LT800	GT800	Total
		RETAIN	42348	32835	3984	3453
41.41	32.11	3.90	3.38		80.80	
51.26	39.74	4.82	4.18			
80.85	80.67	80.76	81.42			
CHURN	10028	7870	949	788	19635	
9.81	7.70	0.93	0.77		19.20	
51.07	40.08	4.83	4.01			
19.15	19.33	19.24	18.58			
Total	52376	40705	4933	4241	102255	
51.22	39.81	4.82	4.15		100.00	



Distribution of Churn_Status by Sales_Amount

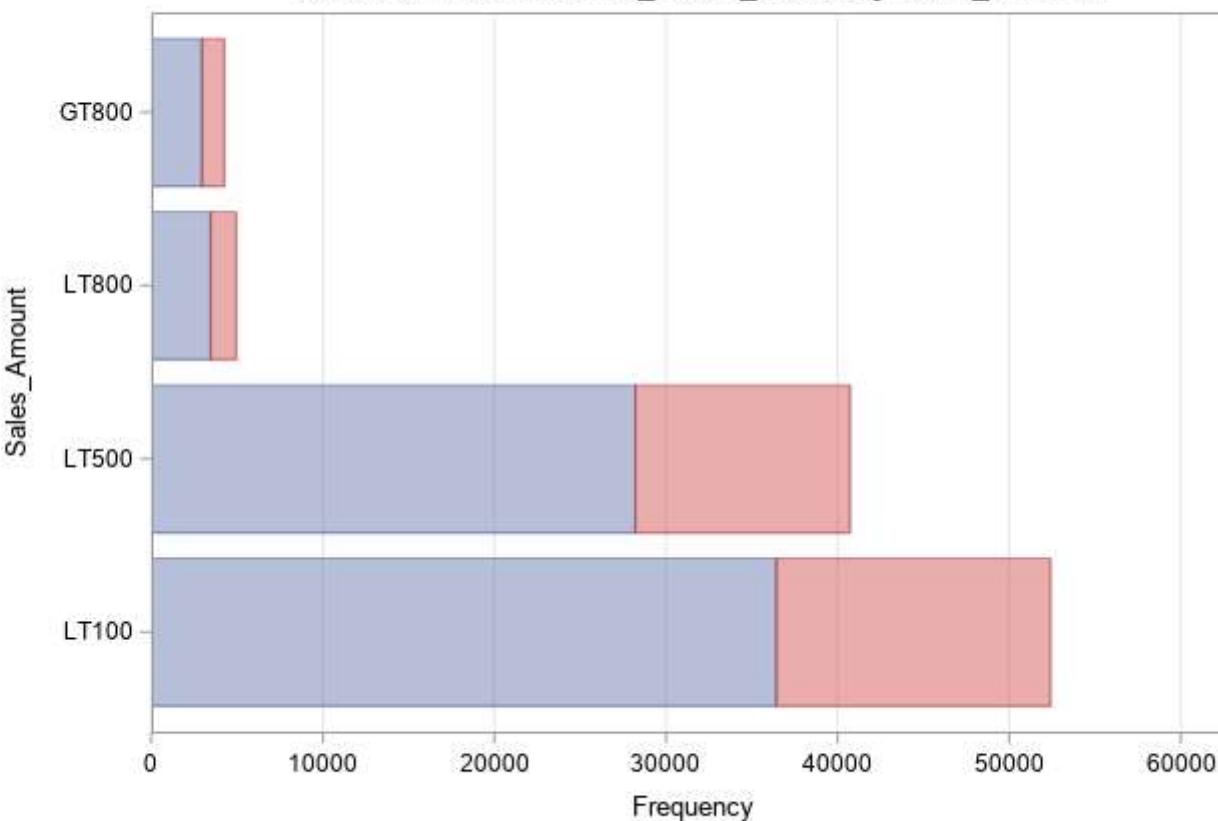


Churn_Status RETAIN CHURN

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Customer_Credit_Status by Sales_Amount					
	Customer_Credit_Status	Sales_Amount				
		LT100	LT500	LT800	GT800	Total
GOOD CREDIT	GOOD CREDIT	36419	28198	3443	2942	71002
		35.62	27.58	3.37	2.88	69.44
		51.29	39.71	4.85	4.14	
		69.53	69.27	69.80	69.37	
NOT GOOD	NOT GOOD	15957	12507	1490	1299	31253
		15.61	12.23	1.46	1.27	30.56
		51.06	40.02	4.77	4.16	
		30.47	30.73	30.20	30.63	
Total	Total	52376	40705	4933	4241	102255
		51.22	39.81	4.82	4.15	100.00

Distribution of Customer_Credit_Status by Sales_Amount



Customer_Credit_Status

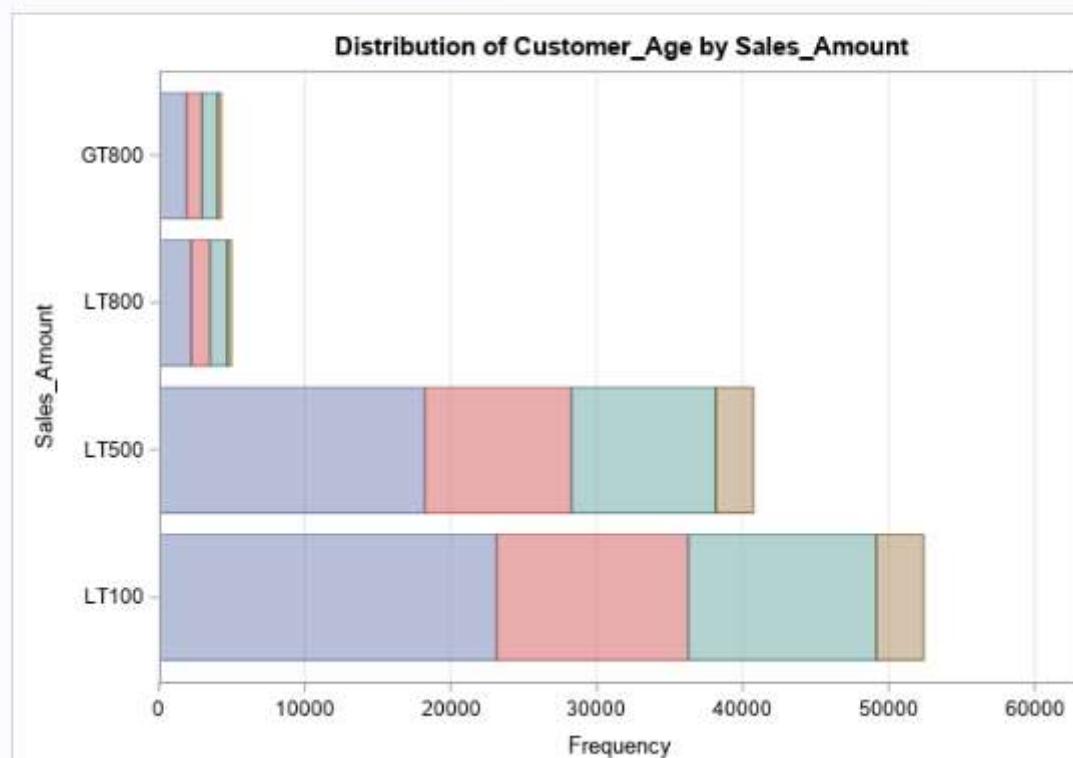
GOOD CREDIT

NOT GOOD

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

	Table of Customer_Age by Sales_Amount					
	Customer_Age	Sales_Amount				
Customer_Age		LT100	LT500	LT800	GT800	Total
LT60	23134	18190	2202	1883	45409	
	22.62	17.79	2.15	1.84	44.41	
	50.95	40.06	4.85	4.15		
	44.17	44.69	44.64	44.40		
Customer_Age	LT40	13124	10049	1279	1066	25518
		12.83	9.83	1.25	1.04	24.96
		51.43	39.38	5.01	4.18	
		25.06	24.69	25.93	25.14	
Customer_Age	Senior	12884	9904	1190	1032	25010
		12.60	9.69	1.16	1.01	24.46
		51.52	39.60	4.76	4.13	
		24.60	24.33	24.12	24.33	
Customer_Age	TEEN	3234	2562	262	260	6318
		3.16	2.51	0.26	0.25	6.18
		51.19	40.55	4.15	4.12	
		6.17	6.29	5.31	6.13	
Customer_Age	Total	52376	40705	4933	4241	102255
		51.22	39.81	4.82	4.15	100.00



Task28: Let us now find the statistical significant association between our target variable which is Churn_Status and other variables.

```
/* Let us now find the statistical significant association between our target variable which is Churn_Status and other variables
```

```
PROC FREQ DATA=CustomerBehavior_05;
TITLE "RELATIONSHIP BETWEEN Reason_For_Deactivation AND Churn_Status";
TABLE Reason_For_Deactivation * Churn_Status/CHISQ OUT=OUT_Deactivation_Churn_Status;
RUN;
```

```
PROC FREQ DATA=CustomerBehavior_05;
TITLE "RELATIONSHIP BETWEEN Dealer_Type AND Churn_Status";
TABLE Dealer_Type * Churn_Status/CHISQ OUT=OUT_Dealer_Type_Churn_Status;
RUN;
```

```
PROC FREQ DATA=CustomerBehavior_05;
TITLE "RELATIONSHIP BETWEEN Province AND Churn_Status";
TABLE Province * Churn_Status/CHISQ OUT=OUT_Province_Churn_Status;
RUN;
```

```
PROC FREQ DATA=CustomerBehavior_05;
TITLE "RELATIONSHIP BETWEEN Customer_Age AND Churn_Status";
TABLE Customer_Age * Churn_Status/CHISQ OUT=OUT_Customer_Age_Churn_Status;
RUN;
```

```
PROC FREQ DATA=CustomerBehavior_05;
TITLE "RELATIONSHIP BETWEEN Sales_Amount AND Churn_Status";
TABLE Sales_Amount * Churn_Status/CHISQ OUT=OUT_Sales_Amount_Churn_Status;
RUN;
```

```
RUN;
```

```
PROC FREQ DATA=CustomerBehavior_05;
TITLE "RELATIONSHIP BETWEEN Customer_Credit_Status AND Churn_Status";
TABLE Customer_Credit_Status * Churn_Status/CHISQ OUT=OUT_Credit_Status_Churn_Status;
RUN;
```

```
□ PROC FREQ DATA=CustomerBehavior_05;  
  TITLE "RELATIONSHIP BETWEEN Rate_Plan AND Churn_Status";  
  TABLE Rate_Plan * Churn_Status/CHISQ OUT=OUT_Rate_Plan_Churn_Status;  
  RUN;
```

```
□ PROC FREQ DATA=CustomerBehavior_05;  
  TITLE "RELATIONSHIP BETWEEN Tenure_in_Days AND Churn_Status";  
  TABLE Tenure_in_Days * Churn_Status/CHISQ OUT=OUT_Tenure_Churn_Status;  
  RUN;
```

RELATIONSHIP BETWEEN BETWEEN Reason_For_Deactivation AND Churn_Status

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Reason_For_Deactivation by Churn_Status			
	Reason_For_Deactivation	Churn_Status		
		CHURN	RETAIN	Total
		4722	0	4722
COMP	24.73	0.00	24.73	
DEBT	100.00	0.00		
MOVE	24.73			
NEED	4020	0	4020	
TECH	21.05	0.00	21.05	
Total	100.00	0.00		
	1696	0	1696	
	8.88	0.00	8.88	
	100.00	0.00		
	8.88			
	6888	0	6888	
	36.08	0.00	36.08	
	100.00	0.00		
	36.08			
	1767	0	1767	
	9.25	0.00	9.25	
	100.00	0.00		
	9.25			
	19093	0	19093	
	100.00	0.00	100.00	
Frequency Missing = 83162				

Statistics for Table of Reason_For_Deactivation by Churn_Status

Row or column sum zero. No statistics computed for this table.

Sample Size = 19093

Frequency Missing = 83162

RELATIONSHIP BETWEEN DEALER_TYPE AND CHURN_STATUS

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Dealer_Type by Churn_Status			
	Dealer_Type	Churn_Status		
		CHURN	RETAIN	Total
A1		10631 10.40 18.94 54.14	45501 44.50 81.06 55.07	56132 54.89
A2		2549 2.49 22.65 12.98	8706 8.51 77.35 10.54	11255 11.01
B1		3879 3.79 18.77 19.76	16791 16.42 81.23 20.32	20670 20.21
C1		2576 2.52 18.14 13.12	11622 11.37 81.86 14.07	14198 13.88
Total		19635 19.20	82620 80.80	102255 100.00

Statistics for Table of Dealer_Type by Churn_Status

Statistic	DF	Value	Prob
Chi-Square	3	101.4114	<.0001
Likelihood Ratio Chi-Square	3	97.9982	<.0001
Mantel-Haenszel Chi-Square	1	2.8637	0.0906
Phi Coefficient		0.0315	
Contingency Coefficient		0.0315	

Sample Size = 102255

RELATIONSHIP BETWEEN Province AND Churn_Status

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Province by Churn_Status			
	Province	Churn_Status		
		CHURN	RETAIN	Total
	AB	2000 2.08 19.46 10.79	8277 8.59 80.54 10.64	10277 10.67
BC	4235 4.40 19.22 22.85	17805 18.48 80.78 22.88	22040 22.88	
NS	2202 2.29 19.10 11.88	9327 9.68 80.90 11.99	11529 11.97	
ON	8174 8.48 19.23 44.10	34326 35.63 80.77 44.11	42500 44.11	
QC	1923 2.00 19.23 10.38	8079 8.39 80.77 10.38	10002 10.38	
Total	18534 19.24	77814 80.76	96348 100.00	
Frequency Missing = 5907				

Statistics for Table of Province by Churn_Status

Statistic	DF	Value	Prob
Chi-Square	4	0.4797	0.9754
Likelihood Ratio Chi-Square	4	0.4790	0.9755
Mantel-Haenszel Chi-Square	1	0.0920	0.7616
Phi Coefficient		0.0022	
Contingency Coefficient		0.0022	
Cramer's V		0.0022	

Sample Size = 96348
 Frequency Missing = 5907

RELATIONSHIP BETWEEN Customer_Age AND Churn_Status

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Customer_Age by Churn_Status			
	Customer_Age	Churn_Status		
		CHURN	RETAIN	Total
TEEN	TEEN	1193 1.17 18.88 6.08	5125 5.01 81.12 6.20	6318 6.18
	LT40	4915 4.81 19.26 25.03	20603 20.15 80.74 24.94	25518 24.96
	LT60	8701 8.51 19.16 44.31	36708 35.90 80.84 44.43	45409 44.41
	Senior	4826 4.72 19.30 24.58	20184 19.74 80.70 24.43	25010 24.46
Total		19635 19.20	82620 80.80	102255 100.00

Statistics for Table of Customer_Age by Churn_Status

Statistic	DF	Value	Prob
Chi-Square	3	0.6642	0.8816
Likelihood Ratio Chi-Square	3	0.6657	0.8812
Mantel-Haenszel Chi-Square	1	0.2081	0.6483
Phi Coefficient		0.0025	
Contingency Coefficient		0.0025	
Cramer's V		0.0025	

Sample Size = 102255

RELATIONSHIP BETWEEN Sales_Amount AND Churn_Status

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Sales_Amount by Churn_Status				
	Sales_Amount	Churn_Status			
		CHURN	RETAIN	Total	
		LT100	10028 9.81 19.15 51.07	42348 41.41 80.85 51.26	52376 51.22
	LT500	7870 7.70 19.33 40.08	32835 32.11 80.67 39.74	40705 39.81	
		LT800	949 0.93 19.24 4.83	3984 3.90 80.76 4.82	4933 4.82
		GT800	788 0.77 19.59	3453 3.38 81.49	4241 4.15

	18.58	81.42	
	4.01	4.18	
Total	19635	82620	102255
	19.20	80.80	100.00

Statistics for Table of Sales_Amount by Churn_Status

Statistic	DF	Value	Prob
Chi-Square	3	1.6238	0.6540
Likelihood Ratio Chi-Square	3	1.6319	0.6522
Mantel-Haenszel Chi-Square	1	0.4175	0.5182
Phi Coefficient		0.0040	
Contingency Coefficient		0.0040	
Cramer's V		0.0040	

Sample Size = 102255

RELATIONSHIP BETWEEN Customer_Credit_Status AND Churn_Status

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Customer_Credit_Status by Churn_Status			
	Customer_Credit_Status	Churn_Status		
		CHURN	RETAIN	Total
		10978	60024	71002
GOOD CREDIT		10.74	58.70	69.44
		15.46	84.54	
		55.91	72.65	
NOT GOOD		8657	22596	31253
		8.47	22.10	30.56
		27.70	72.30	
		44.09	27.35	
Total	19635	82620	102255	
	19.20	80.80	100.00	

Statistics for Table of Customer_Credit_Status by Churn_Status

Statistic	DF	Value	Prob
Chi-Square	1	2094.9138	<.0001
Likelihood Ratio Chi-Square	1	1997.7962	<.0001
Continuity Adj. Chi-Square	1	2094.1251	<.0001
Mantel-Haenszel Chi-Square	1	2094.8933	<.0001
Phi Coefficient		-0.1431	
Contingency Coefficient		0.1417	
Cramer's V		-0.1431	

Sample Size = 102255

RELATIONSHIP BETWEEN RATE_PLAN AND CHURN_STATUS

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Rate_Plan by Churn_Status			
	Rate_Plan	Churn_Status		
		CHURN	RETAIN	Total
		12469 12.19 18.28 63.50	55725 54.50 81.72 67.45	68194 66.69
LEVEL1	3439 3.36 17.04 17.51	16748 16.38 82.96 20.27	20187 19.74	
LEVEL2	3727 3.64 26.86 18.98	10147 9.92 73.14 12.28	13874 13.57	
LEVEL3	Total	19635 19.20	82620 80.80	102255 100.00

Statistics for Table of Rate_Plan by Churn_Status

Statistic	DF	Value	Prob
Chi-Square	2	622.9181	<.0001
Likelihood Ratio Chi-Square	2	581.3056	<.0001
Mantel-Haenszel Chi-Square	1	345.3448	<.0001
Phi Coefficient		0.0781	
Contingency Coefficient		0.0778	
Cramer's V		0.0781	

Sample Size = 102255

RELATIONSHIP BETWEEN BETWEEN Tenure_in_Days AND Churn_Status

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Tenure_in_Days by Churn_Status			
	Tenure_in_Days	Churn_Status		
		CHURN	RETAIN	Total
UPTO_MONTH	3133	6353	9486	
	3.06	6.21	9.28	
	33.03	66.97		
	15.96	7.69		
	UPTO2MONTHS	1188	7398	8586
		1.16	7.23	8.40
		13.84	86.16	
		6.05	8.95	
	UPTO_YEAR	11428	33977	45405
		11.18	33.23	44.40
		25.17	74.83	
		58.20	41.12	
	YEAR+	3886	34892	38778
		3.80	34.12	37.92
		10.02	89.98	

	19.79	42.23	
Total	19635	82620	102255
	19.20	80.80	100.00

Statistics for Table of Tenure_in_Days by Churn_Status

Statistic	DF	Value	Prob
Chi-Square	3	4476.7489	<.0001
Likelihood Ratio Chi-Square	3	4613.7901	<.0001
Mantel-Haenszel Chi-Square	1	3461.2464	<.0001
Phi Coefficient		0.2092	
Contingency Coefficient		0.2048	
Cramer's V		0.2092	

Sample Size = 102255

Conclusion:

We performed quite a few Exploratory Data Analysis consisting of Handling missing values, Segmentation, Visualization and testing of statistical significant association between our target variable churn and other independent variables. We found that Dealer_Type, Customer_Credit_Status, Rate_Plan, Tenure_in_Days have significant statistical association with the customer churn which is our target. So, after more analysis and exploration we should be able to design a classification model to predict customers who are prone to churn in future.