



Mini Project – Advanced SAS

Telecom Company Customer Distribution & Deactivation Analysis

By : Maya Babu

INTRODUCTION

Now a days telecommunications industries has become very popular with large number of operators around the world. This has led to an increased level of competition between the providers. It is extremely hard for companies to survive in this competitive market. They need to implement multiple marketing strategies depending on the customer behavioral pattern. Some of the approaches these companies implement to generate more revenue are to getting more new customers, holding on to existing customers with promotions and other benefits to make sure they don't leave the company searching for another provider. Customer churn is one of the most important problem every telecom company faces nowadays. Prediction of potential customer churn is very important because most of the time the costs related with acquiring a new customer is much greater when compared to retaining an existing customer. Therefore, by predicting potential customers who are more likely to churn from the company on an early stage will prove to be more profitable and there by increasing the overall revenue of the company. Good Customer Retention Strategy is needed by the companies to minimize customer attrition. This can be achieved by investigating and analyzing the customer distribution and behavioral pattern to get a proper understanding of the customer and the factors that influence customer churn.

UNDERSTANDING BUSINESS

The data used in this project contains a CRM data of wireless company for 2 years starting from January 1999 to January 2001. Each row represents a customer with a ClientID, each column contains the customer's attributes such as the date when the customer activated his account with the company, If the customer is still active with the company or has the customer deactivated his account and the reason for de activation Customer credit status , age, demographic information and sales customer has had during his active period with the company In this project, we will analyze customer level data of a telecom company, investigate their behavioral pattern to identify customers at high risk of churn and recognize the main indicators or reason for their churn. This behavioral analysis will help the company to categorize customers and design strategies that will in turn help them to minimize customer attrition ,

retain their most valuable customers and help to forecast the deactivation trends for the next 6 months.

BUSINESS QUESTIONS

- What is the number of accounts activated and deactivated? When is the earliest and latest activation/deactivation dates available?
- What is the age and province distributions of active and deactivated customers?
- What is the tenure in days for each account ?
- How many accounts are deactivated each month in total? Compare it with the accounts getting deactivate by month each year. Is there any increase in customer attrition over the years ?
- Is there any association between tenure and credit status , Rate plan and Dealer Type?
- Is there any association between the account status and the tenure segments?
- Does Sales amount differ among different account status, GoodCredit, and customer age segments?

ATTRIBUTES

Acctno: Account number.

Actdt: Account activation date

Deactdt: Account deactivation date

DeactReason: Reason for deactivation.

GoodCredit: Customer's credit (Good or Bad)

RatePlan: Rate plan for the customer.

DealerType: Dealer type.

Age: Customer age.

Province: Province.

Sales: The amount of sales to a customer

IMPORTING AND GETTING FAMILIAR WITH DATA

Title " Importing Data Set in to SAS";

```
data telecom.Details;
infile "C:\DSA\Advanced SAS\Final Project\New_Wireless_Fixed.txt";
input acctno 1-14
@15 actdt mmddyy10.
@26 Deactdt mmddyy10.
Deactreason $ 41-45
GoodCredit 53
Rateplan $62-63
DealerType $65-66
Age 74 -75
Province $ 80-81
Sales dollar11.2
;
format acctno 14.0 actdt date9. Deactdt date9. ;
format Sales DOLLAR11.2;
run;
```



```
proc print data = telecom.Details(obs=10);
run;
```

Importing Data Set in to SAS

Obs	acctno	actdt	Deactdt	Deactreason	GoodCredit	Rateplan	DealerType	Age	Province	Sales
1	1176913194483	20JUN1999	.		0 1		A1	58	BC	\$128.00
2	1176914599423	04OCT1999	15OCT1999	NEED	1 1		A1	45	AB	\$72.00
3	1176951913656	01JUL2000	.		0 1		A1	57	BC	\$593.00
4	1176954000288	30MAY2000	.		1 2		A1	47	ON	\$83.00
5	1176969186303	13DEC2000	.		1 1		C1	82	BC	.
6	1176991056273	31AUG1999	18SEP2000	MOVE	1 1		C1	92	QC	\$1,041.00
7	1176991866552	24MAY2000	.		1 1		A1	77	ON	.
8	1176992889500	28NOV2000	.		1 1		C1	68	AB	\$72.00
9	1177000067271	23DEC1999	.		0 1		B1	75	ON	\$134.00
10	1177010940613	09DEC1999	.		1 2		A1	42	NS	\$11.00

Analysis Requests:

Question 1.1 Explore and describe the dataset briefly. For example, is the acctno unique? What is the number of accounts activated and deactivated? When is the earliest and latest activation/deactivation dates available?

1.1 a) Explore and describe the dataset briefly

```
*=====
Browsing Descriptive Portion
=====
|proc contents  data = telecom.Details order = varnum;
run;
```

The CONTENTS Procedure				
		Variables in Creation Order		
#	Variable	Type	Len	Format
1	acctno	Num	8	14.
2	actdt	Num	8	DATE9.
3	Deactdt	Num	8	DATE9.
4	Deactreason	Char	5	
5	GoodCredit	Num	8	
6	Rateplan	Char	2	
7	DealerType	Char	2	
8	Age	Num	8	
9	Province	Char	2	
10	Sales	Num	8	DOLLAR11.2

The CONTENTS Procedure

Data Set Name	TELECOM.DETAILS	Observations	102255
Member Type	DATA	Variables	10
Engine	V9	Indexes	0
Created	2024-01-27 00:49:17	Observation Length	64
Last Modified	2024-01-27 00:49:17	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

There are 102255 observations and 10 variables in the given dataset

FINDING HEAD OF DATA FOR 10 OBSERVATION

```
Browsing Head of data set
=====
Title "Browsing Head of Data Set";
proc print data = telecom.Details (obs = 10);
run;
```

Browsing Head of Data Set

Obs	acctno	actdt	Deactdt	Deactreason	GoodCredit	Rateplan	DealerType	Age	Province	Sales
1	1176913194483	20JUN1999	.		0	1	A1	58	BC	\$128.00
2	1176914599423	04OCT1999	15OCT1999	NEED	1	1	A1	45	AB	\$72.00
3	1176951913656	01JUL2000	.		0	1	A1	57	BC	\$593.00
4	1176954000288	30MAY2000	.		1	2	A1	47	ON	\$83.00
5	1176969186303	13DEC2000	.		1	1	C1	82	BC	.
6	1176991056273	31AUG1999	18SEP2000	MOVE	1	1	C1	92	QC	\$1,041.00
7	1176991866552	24MAY2000	.		1	1	A1	77	ON	.
8	1176992889500	28NOV2000	.		1	1	C1	68	AB	\$72.00
9	1177000067271	23DEC1999	.		0	1	B1	75	ON	\$134.00
10	1177010940613	09DEC1999	.		1	2	A1	42	NS	\$11.00

FINDING TAIL OF DATA FOR 10 OBSERVATION

```
Browsing tail of data set
=====
Title "Browsing Tail of Data Set";
proc print data = telecom.Details (obs = 102255 firstobs =102246);
run;
```

Browsing Tail of Data Set

Obs	acctno	actdt	Deactdt	Deactreason	GoodCredit	Rateplan	DealerType	Age	Province	Sales
102246	2673080989261	06JAN2001	.		1	2	B1	68	ON	\$25.00
102247	2673104899783	12JAN2001	.		0	2	C1	68	ON	.
102248	2673110609830	12JAN2001	.		1	1	A2	63	ON	\$78.00
102249	2673216477998	12JAN2001	.		1	2	C1	53	ON	\$316.00
102250	2673579485183	31DEC2000	.		0	1	A1	.	NS	\$130.00
102251	2673974127660	29DEC2000	.		1	1	A2	50		\$112.00
102252	2674189951308	15JAN2001	.		1	2	A1	40	BC	\$87.00
102253	2674548796918	15JAN2001	.		1	1	A1	16	NS	\$316.00
102254	2675119766018	15JAN2001	.		1	2	B1	76	ON	.
102255	2675135410256	17JAN2001	.		1	1	A1	46	BC	\$319.00

1.1 (b) Finding if all the account numbers are unique?

```
*Removing Duplicated Data if any;
proc sort data = telecom.Details out = telecom.Data nodupkey;
by Acctno;
run;

NOTE: There were 102255 observations read from the data set TELECOM.DETAILS.
NOTE: 0 observations with duplicate key values were deleted.
NOTE: The data set TELECOM.DATA has 102255 observations and 10 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time            0.03 seconds
      cpu time             0.01 seconds

No duplicated values for account number were found so zero
observations were deleted
```

Number of unique accounts

```
PROC SQL OUTOBS=20;
SELECT COUNT(*) AS TOTAL_COUNT,
       COUNT(DISTINCT Acctno) AS UNIQUE_ACCOUNTS
FROM TELECOM.DATA
;
QUIT;          Number of Unique Accounts
```

TOTAL_COUNT	UNIQUE_ACCOUNTS
102255	102255

Checking For Missing Values in the Dataset

```
Title "Number of missing values";
proc means data = telecom.Data nmiss;
run;

proc sql;
select nmiss(Province) as Province,nmiss (deactreason) as deactreason,
nmiss(rateplan) as rateplan,nmiss(dealertype) as dealertype
from telecom.Data;
quit;
```

Number of missing values

The MEANS Procedure

Number of missing values			
Province	deactreason	rateplan	dealertype
5907	83162	0	0

Variable	N Miss
acctno	0
actdt	0
Deactdt	82620
GoodCredit	0
Age	7708
Sales	8605

Here missing values in the Deactivation Date is considered that those accounts are still active with the company.

Finding Number Of Unique levels in each Variables

Number Of Unique/Distinct Values In All Variable

```
Title "Number of unique distinct values in each variables";
proc freq data =telecom.Details nlevels;
ods exclude onewayfreqs;
run;
```

The FREQ Procedure

Number of Variable Levels			
Variable	Levels	Missing Levels	Nonmissing Levels
acctno	102255	0	102255
actdt	729	0	729
Deactdt	686	1	685
Deactreason	6	1	5
GoodCredit	2	0	2
Rateplan	3	0	3
DealerType	4	0	4
Age	101	1	100
Province	6	1	5
Sales	1202	1	1201

Number Of Unique value in each Variables Group

Deactreason	Frequency
COMP	4722
DEBT	4020
MOVE	1696
NEED	6888
TECH	1767
Frequency Missing = 83162	

DealerType	Frequency
A1	56132
A2	11255
B1	20670
C1	14198

GoodCredit	Frequency
0	31253
1	71002

Province	Frequency
AB	10277
BC	22040
NS	11529
ON	42500
QC	10002
Frequency Missing = 5907	

DESCRIPTIVE ANALYSIS OF CONTINUOUS VARIABLES

```
TITLE "DESCRIPTIVE ANALYSIS OF CONTINUOUS";
PROC MEANS DATA = Telecom.Data N NMISS MIN Q1 MEDIAN Q3 MAX qrange mean std cv clm;
RUN;
```

DESCRIPTIVE ANALYSIS OF CONTINUOUS															
The MEANS Procedure															
Variable	N	N Miss	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Quartile Range	Mean	Std Dev	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean		
acctno	102255	0	1.1769132E12	1.5426519E12	1.9073763E12	2.2757324E12	2.6751354E12	733080460318	1.9104261E12	423614424868	22.1738194	1.9078296E12	1.9130225E12		
actdt	102255	0	14264.00	14522.00	14660.00	14855.00	14995.00	333.0000000	14675.35	200.1245116	1.3636777	14674.13	14676.58		
Deactdt	19635	82620	14269.00	14697.00	14853.00	14941.00	14995.00	244.0000000	14801.92	163.9246301	1.1074553	14799.63	14804.21		
GoodCredit	102255	0	0	0	1.0000000	1.0000000	1.0000000	0.0000000	0.6943621	0.4606793	66.3456859	0.6915385	0.6971858		
Age	94547	7708	0	34.0000000	48.0000000	60.0000000	99.0000000	26.0000000	47.5922240	18.5554517	38.9884107	47.4739467	47.7105012		
Sales	93650	8605	0	52.0000000	91.0000000	190.0000000	1200.00	138.0000000	181.2461612	233.9710402	129.0902045	179.7476434	182.7446791		

1.1 (c) What is the number of accounts activated and deactivated?

```
Title "Number of Activated and Deactivated Accounts";
PROC SQL;
SELECT COUNT(Acctno) AS Total_Accounts,
       (COUNT(Actdt) - COUNT(Deactdt)) AS Activted_Accounts,
       COUNT(Deactdt) AS Deactvated_Accounts
FROM telecom.Data;
QUIT;
```

Number of Activated and Deactivated Accounts

Total_Accounts	Activted_Accounts	Deactvated_Accounts
102255	82620	19635

When is the earliest and latest activation dates available?

```
Title "Earliest and Latest Activation Date";
proc sql;
select min(actdt) as Earliest_Activation_Date format = date9.,
       max(actdt) as Latest_Activation_Date format = date9.
from telecom.Data;
quit;
```

Earliest and Latest Activation Date

Earliest_Activation_Date	Latest_Activation_Date
20JAN1999	20JAN2001

When is the earliest and latest deactivation dates available?

```
Title "Earliest and Latest DeActivation Date";
proc sql;
select min(deactdt) as Earliest_DeActivation_Date format = date9.,
max(deactdt) as Latest_DeActivation_Date format = date9.
from telecom.Data;
quit;
```

Earliest and Latest DeActivation Date

Earliest_DeActivation_Date	Latest_DeActivation_Date
25JAN1999	20JAN2001

1.2 What is the age and province distributions of active customers ?

```
data telecom.Account_segment ;
set telecom.Data;
length Age_Group $25;
IF AGE <= 20 THEN AGE_GROUP = "LESS THAN 20";
ELSE IF 21<= AGE<=40 THEN AGE_GROUP = "BETWEEN 21 AND 40 ";
ELSE IF 41<=AGE <=59 THEN AGE_GROUP ="BETWEEN 41 AND 60";
ELSE IF AGE >= 60 THEN AGE_GROUP = "60 AND ABOVE";
RUN;

proc print data = telecom.Account_segment (obs=20);
run;

Title"Age distributions of active customers";

PROC SQL ;
CREATE TABLE AGEDIST AS
SELECT AGE_GROUP,
(COUNT(Actdt) - COUNT(Deactdt))AS TOTAL_ACTIVE_CUSTOMERS,
SUM(SALES) AS TOTAL_SALES
FROM TELECOM.ACCT_SEGMENT
WHERE DEACTDT IS NULL
GROUP BY AGE_GROUP
ORDER BY AGE_GROUP;
QUIT;

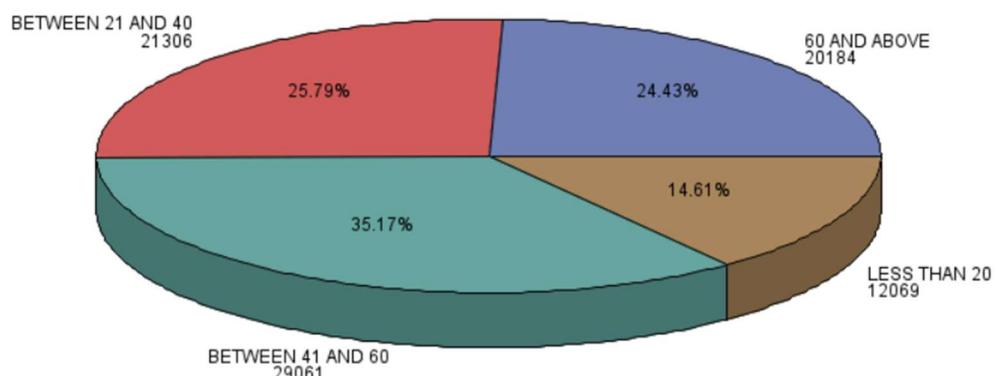
PROC PRINT DATA = AGEDIST;
RUN;
```

Age distributions of active customers

Obs	Age_Group	TOTAL_ACTIVE_CUSTOMERS	TOTAL_SALES
1	60 AND ABOVE	20184	3359273
2	BETWEEN 21 AND 40	21306	3540947
3	BETWEEN 41 AND 60	29061	4861388
4	LESS THAN 20	12069	1977941

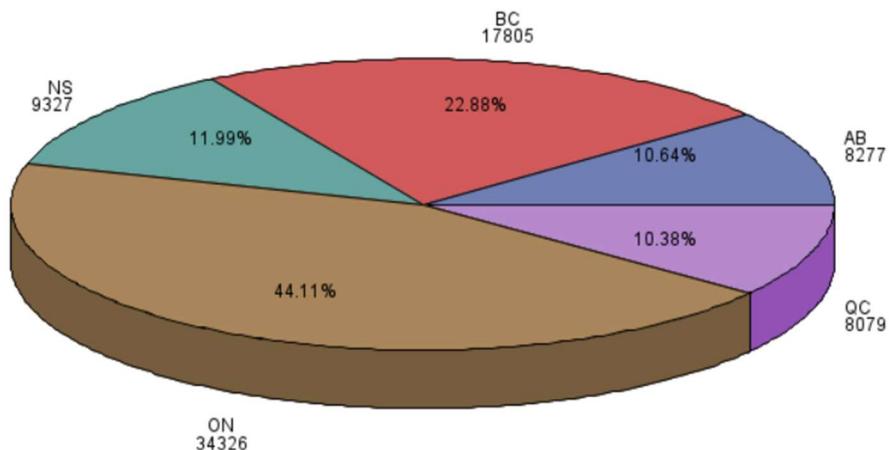
Age distributions of active customers

FREQUENCY of Age_Group



Distribution of Active Customers By Province

FREQUENCY of Province



1.2 What is the age and province distributions of Deactivated customers ?

```
Title"Distribution of DeActive Customers By Age and Province ";
PROC SQL ;
CREATE TABLE AGEPROVINDIST1 AS
SELECT PROVINCE,AGE_GROUP,
COUNT(Deactdt)AS TOTAL_DEACTIVATED_CUSTOMERS,
SUM(SALES) AS TOTAL_SALES
FROM TELECOM.ACCTOUNT_SEGMENT
WHERE DEACTDT IS NOT NULL
GROUP BY PROVINCE,AGE_GROUP
ORDER BY PROVINCE,AGE_GROUP;
QUIT;
```

Distribution of Active Customers By Age and Province

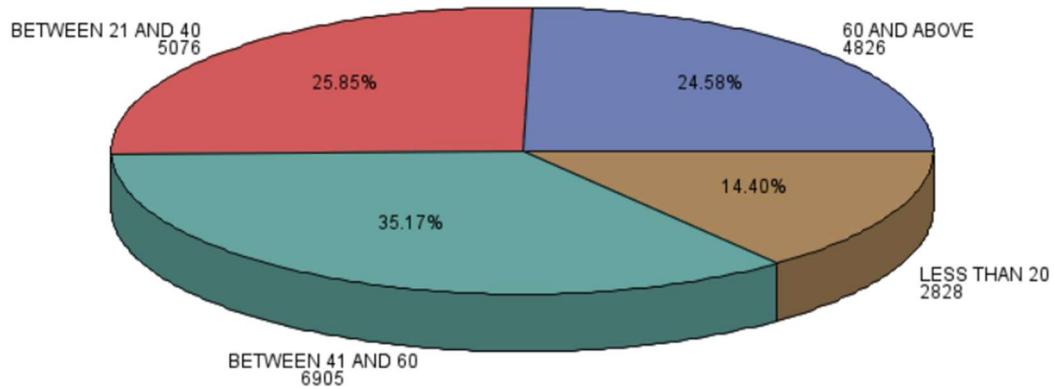
Obs	Province	Age_Group	TOTAL_ACTIVE_CUSTOMERS	TOTAL_SALES
5	AB	60 AND ABOVE	2032	344433
6	AB	BETWEEN 21 AND 40	2109	340328
7	AB	BETWEEN 41 AND 60	2931	492806
8	AB	LESS THAN 20	1205	192832
9	BC	60 AND ABOVE	4460	730822
10	BC	BETWEEN 21 AND 40	4587	762399
11	BC	BETWEEN 41 AND 60	6142	1039681
12	BC	LESS THAN 20	2616	438046
13	NS	60 AND ABOVE	2252	376702
14	NS	BETWEEN 21 AND 40	2458	396486
15	NS	BETWEEN 41 AND 60	3317	540012
16	NS	LESS THAN 20	1300	209274
17	ON	60 AND ABOVE	8311	1375838
18	ON	BETWEEN 21 AND 40	8799	1472867
19	ON	BETWEEN 41 AND 60	12123	2035345
20	ON	LESS THAN 20	5093	831230
21	QC	60 AND ABOVE	1961	332063
22	QC	BETWEEN 21 AND 40	2124	355626
23	QC	BETWEEN 41 AND 60	2822	465868
24	QC	LESS THAN 20	1172	192150

Age distributions of Deactive customers

Obs	Age_Group	TOTAL_DEACTIVE_CUSTOMERS	TOTAL_SALES
1	60 AND ABOVE	4826	766549
2	BETWEEN 21 AND 40	5076	864655
3	BETWEEN 41 AND 60	6905	1130531
4	LESS THAN 20	2828	472419

Age distributions of Deactive customers

FREQUENCY of Age_Group

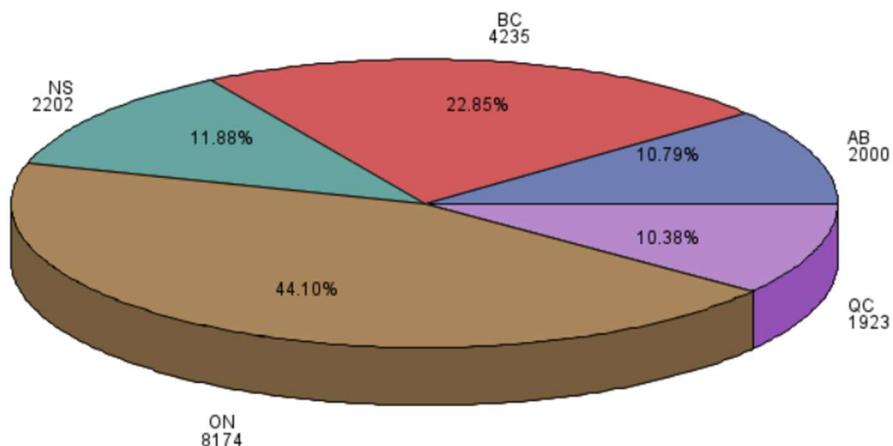


Distributions of Deactive customers by Province

Obs	Province	TOTAL_DEACTIVE_CUSTOMERS	TOTAL_SALES
2	AB	2000	321727
3	BC	4235	698607
4	NS	2202	378489
5	ON	8174	1343123
6	QC	1923	311269

Distribution of DeActive Customers By Province

FREQUENCY of Province



1.3 Segment the customers based on age, province and sales amount: Sales segment: < \$100, \$100---500, \$500-\$800, \$800 and above. Age segments: < 20, 21-40, 41-60, 60 and above. Do Analysis of sales based on segmentation

```
TITLE"SEGMENTATION BASED ON AGE , SALES and PROVINCE";
DATA TELECOM.SEGMENTS;
SET TELECOM.DATA;
DROP NEW_AGE;
LENGTH AGE_GROUP $25;
LENGTH SALES_GROUP $25;
LENGTH PROVINCEE $25;

IF AGE <= 20 THEN AGE_GROUP = "20 OR LESS";
ELSE IF 21<= AGE<=40 THEN AGE_GROUP= "BETWEEN 21 & 40 ";
ELSE IF 41<=AGE<=59 THEN AGE_GROUP =" BETWEEN 41 - 60";
ELSE IF AGE >= 60 THEN AGE_GROUP = "60 & MORE";
```

```
*Sales segment: < $100, $100---500, $500-$800, $800 and above.;
```

```
IF SALES<100 THEN SALES_GROUP ="$100 & BELOW";
ELSE IF 100 <SALES<500 THEN SALES_GROUP ="$100 - $500";
ELSE IF 500 <SALES<800 THEN SALES_GROUP ="$500 - $800";
ELSE IF SALES >=800 THEN SALES_GROUP="800 & ABOVE";
```

```
*Province Segmentation;
```

```
IF PROVINCE = "AB" THEN PROVINCEE = "ALBERTA";
ELSE IF PROVINCE ="BC" THEN PROVINCEE ="BRITISH COLOMBIA";
ELSE IF PROVINCE ="NS" THEN PROVINCEE = "NOVA SCOTIA";
ELSE IF PROVINCE ="ON" THEN PROVINCEE = "ONTARIO";
ELSE IF PROVINCE ="QC" THEN PROVINCEE ="QUEBEC";
RUN;
```

```
PROC PRINT DATA = TELECOM.SEGMENTS (OBS = 50);
```

```
RUN;
```

SEGMENTATION BASED ON AGE , SALES and PROVINCE

Obs	acctno	actdt	Deactdt	Deactreason	GoodCredit	Rateplan	DealerType	Age	Province	Sales	AGE_GROUP	SALES_GROUP	PROVINCEE
1	1176913194483	20JUN1999	.		0 1	A1		58	BC	\$128.00	BETWEEN 41-60	\$100 - \$500	BRITISH COLOMBIA
2	1176914599423	04OCT1999	15OCT1999	NEED		1 1	A1	45	AB	\$72.00	BETWEEN 41-60	\$100 & BELOW	ALBERTA
3	1176951913656	01JUL2000	.		0 1	A1		57	BC	\$593.00	BETWEEN 41-60	\$500 - \$800	BRITISH COLOMBIA
4	1176954000288	30MAY2000	.		1 2	A1		47	ON	\$83.00	BETWEEN 41-60	\$100 & BELOW	ONTARIO
5	1176969186303	13DEC2000	.		1 1	C1		82	BC	.	60 & MORE	\$100 & BELOW	BRITISH COLOMBIA
6	1176991056273	31AUG1999	18SEP2000	MOVE		1 1	C1	92	QC	\$1,041.00	60 & MORE	800 & ABOVE	QUEBEC
7	1176991866552	24MAY2000	.		1 1	A1		77	ON	.	60 & MORE	\$100 & BELOW	ONTARIO
8	1176992889500	28NOV2000	.		1 1	C1		68	AB	\$72.00	60 & MORE	\$100 & BELOW	ALBERTA
9	1177000067271	23DEC1999	.		0 1	B1		75	ON	\$134.00	60 & MORE	\$100 - \$500	ONTARIO
10	1177010940613	09DEC1999	.		1 2	A1		42	NS	\$11.00	BETWEEN 41-60	\$100 & BELOW	NOVA SCOTIA
11	1177025997013	09NOV1999	.		1 1	A1		26	BC	\$154.00	BETWEEN 21 & 40	\$100 - \$500	BRITISH COLOMBIA
12	1177027515760	19OCT1999	.		1 1	B1		73	BC	\$16.00	60 & MORE	\$100 & BELOW	BRITISH COLOMBIA
13	1177028996676	21SEP2000	.		0 1	C1		.	QC	\$179.00	20 OR LESS	\$100 - \$500	QUEBEC
14	1177038747105	14MAR2000	.		0 1	C1		41	ON	\$705.00	BETWEEN 41-60	\$500 - \$800	ONTARIO
15	1177045857516	22JUN2000	.		1 1	A1		53	QC	\$83.00	BETWEEN 41-60	\$100 & BELOW	QUEBEC
16	1177057406016	21SEP2000	.		0 1	C1		50	ON	\$529.00	BETWEEN 41-60	\$500 - \$800	ONTARIO
17	1177066422248	26APR1999	15JAN2001	NEED		0 1	A2	55	NS	\$44.00	BETWEEN 41-60	\$100 & BELOW	NOVA SCOTIA
18	1177089399155	17AUG2000	.		1 3	A1		56	BC	\$548.00	BETWEEN 41-60	\$500 - \$800	BRITISH COLOMBIA
19	1177113886410	13SEP2000	08JAN2001	COMP		0 1	C1	45	ON	\$63.00	BETWEEN 41-60	\$100 & BELOW	ONTARIO
20	1177128264924	10DEC1999	.		1 1	B1		38	ON	\$178.00	BETWEEN 21 & 40	\$100 - \$500	ONTARIO

SALES BASED ON AGE GROUP

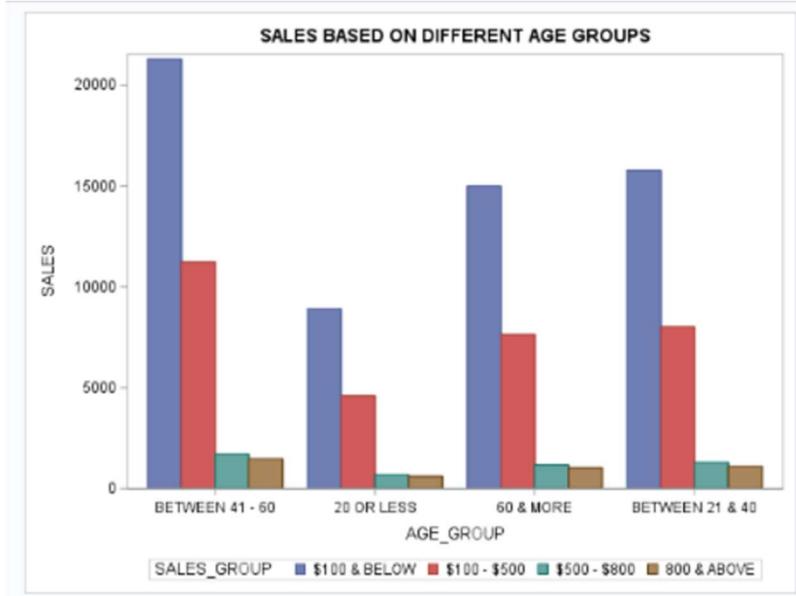
SALES BASED ON DIFFERENT AGE GROUPS

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of AGE_GROUP by SALES_GROUP					
	AGE_GROUP	SALES_GROUP				
		\$100 & BELOW	\$100 - \$500	\$500 - \$800	\$800 & ABOVE	Total
BETWEEN 41 - 60		21292 20.95 59.56 34.92	11241 11.08 31.45 35.87	1728 1.70 4.83 35.19	1486 1.46 4.16 35.04	35747 35.17
20 OR LESS		8916 8.77 60.15 14.62	4804 4.53 31.08 14.81	689 0.68 4.65 14.03	814 0.60 4.14 14.48	14823 14.58
60 & MORE		14994 14.75 60.33 24.59	7845 7.52 30.78 24.28	1181 1.16 4.75 24.05	1032 1.02 4.15 24.33	24852 24.45
BETWEEN 21 & 40		15779 15.52 60.18 25.88	8021 7.89 30.59 25.45	1312 1.29 5.00 26.72	1109 1.09 4.23 26.15	26221 25.80
Total		60981 60.00	31511 31.00	4910 4.83	4241 4.17	101643 100.00
Frequency Missing = 612						

Statistics for Table of AGE_GROUP by SALES_GROUP

Statistic	DF	Value	Prob
Chi-Square	9	9.2197	0.4172
Likelihood Ratio Chi-Square	9	9.2131	0.4178
Mantel-Haenszel Chi-Square	1	0.4647	0.4954
Phi Coefficient		0.0095	
Contingency Coefficient		0.0095	
Cramer's V		0.0055	

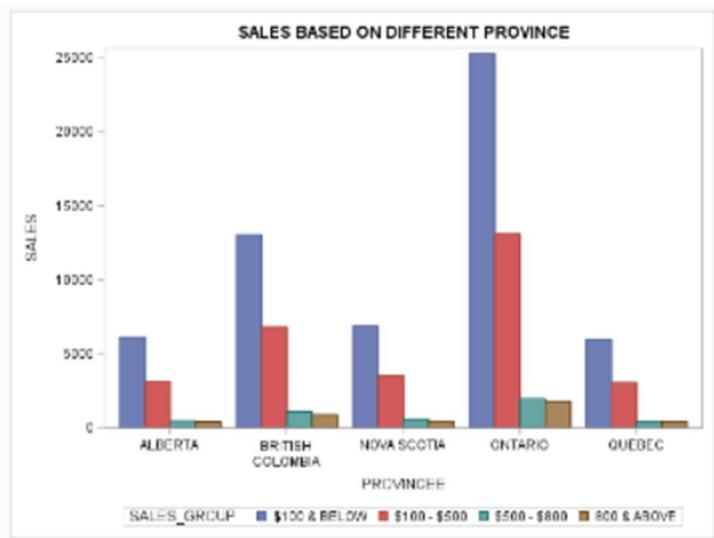


Irrespective of age the sales is highest for the category \$100 and below

SALES BASED ON PROVINCE

SALES BASED ON DIFFERENT PROVINCE						
The FREQ Procedure						
Frequency Percent Row Pct Col Pct	Table of PROVINCE by SALE\$_GROUP					
PROVINCE	SALE\$_GROUP					
	\$100 & BELOW	\$100 - \$600	\$600 - \$800	\$800 & ABOVE	Total	
ALBERTA	6161 6.43 60.25 10.73	3158 3.30 30.88 10.61	493 0.50 4.72 10.47	424 0.44 4.15 10.65	10226	
BRITISH COLUMBIA	13061 13.64 59.60 22.75	6850 7.15 31.26 23.01	1101 1.15 5.02 23.88	901 0.94 4.11 22.63	21913	
NOVA SCOTIA	6882 7.19 60.01 11.98	3556 3.71 31.01 11.95	589 0.61 5.14 12.77	441 0.46 3.85 11.07	11468	
ONTARIO	25321 26.44 59.96 44.10	13127 13.71 31.08 44.10	1999 2.09 4.73 43.35	1783 1.86 4.22 44.78	42230	
QUEBEC	5998 6.26 60.32 10.45	3074 3.21 30.91 10.33	439 0.46 4.41 9.52	433 0.45 4.35 10.87	9944	
Total	57423 59.95	29765 31.08	4611 4.81	3982 4.16	95781	100.00
Frequency Missing = 8474						

Statistics for Table of PROVINCE by SALE\$_GROUP			
Statistic	DF	Value	Prob
Chi-Square	12	13.9580	0.3034
Likelihood Ratio Chi-Square	12	14.0293	0.2988
Mantel-Haenszel Chi-Square	1	0.0571	0.7956
Phi Coefficient		0.0121	
Contingency Coefficient		0.0121	
Cramer's V		0.0070	



Irrespective of location/province the sales is highest for the category \$100 and below

AGE DISTRIBUTION IN PROVINCES

AGE DISTRIBUTION BASED ON DIFFERENT PROVINCE

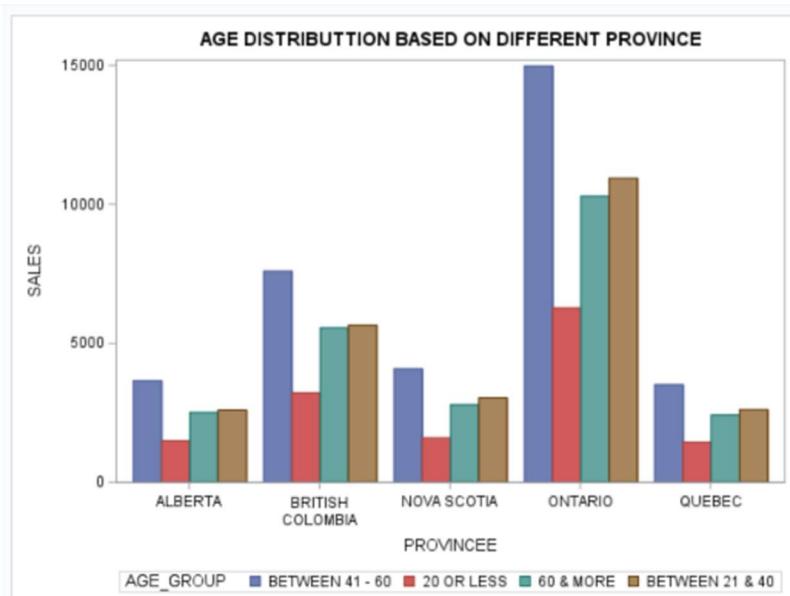
The FREQ Procedure

Frequency Percent Row Pct Col Pct	PROVINCE	Table of PROVINCE by AGE_GROUP				
		BETWEEN 41 - 60	20 OR LESS	60 & MORE	BETWEEN 21 & 40	Total
ALBERTA	3658	1500	2517	2602	10277	
	3.80	1.56	2.61	2.70	10.67	
	36.59	14.60	24.49	25.32		
	10.81	10.67	10.67	10.47		
BRITISH COLOMBIA	7599	3227	5555	5659	22040	
	7.89	3.35	5.77	5.87	22.88	
	34.43	14.64	25.20	25.68		
	22.45	22.96	23.55	22.76		
NOVA SCOTIA	4089	1601	2797	3042	11529	
	4.24	1.66	2.90	3.16	11.97	
	35.47	13.89	24.26	26.39		
	12.08	11.39	11.86	12.24		
ONTARIO	14983	6276	10298	10943	42500	
	15.55	6.51	10.89	11.36	44.11	
	35.25	14.77	24.23	25.75		
	44.27	44.65	43.66	44.01		
QUEBEC	3515	1452	2419	2616	10002	
	3.65	1.51	2.51	2.72	10.38	
	35.14	14.52	24.19	26.15		
	10.39	10.33	10.26	10.52		
Total	33844	14056	23686	24862	96348	
	35.13	14.59	24.48	25.80	100.00	

Frequency Missing = 6807

Statistics for Table of PROVINCE by AGE_GROUP

Statistic	DF	Value	Prob
Chi-Square	12	18.2553	0.1082
Likelihood Ratio Chi-Square	12	18.2805	0.1074
Mantel-Haenszel Chi-Square	1	0.0329	0.8560
Phi Coefficient		0.0138	
Contingency Coefficient		0.0138	
Cramer's V		0.0079	



In each province maximum number of customers belong to the age group between 41 and 60

1.4.Statistical Analysis:

1.4(a) Calculate the tenure in days for each account and give its simple statistics.

```
Title "Tenure in days for each account";
data telecom.tenure;
set telecom.segments;
*reference_date = "20JAN2001"d;
if deactdt = '' then tenure_days = intck('day',actdt,"20JAN2001"d);
else
tenure_days = intck('day',actdt,deactdt);
run;
```

Tenure in days for each account

Obs	acctno	actdt	Deactdt	Deactreason	GoodCredit	Rateplan	DealerType	Age	Province	Sales	AGE_GROUP	SALES_GROUP	PROVINCE	tenure_days
1	1176913194483	20JUN1999	.		0 1	A1	58 BC	\$128.00	BETWEEN 41 - 60	\$100 - \$500	BRITISH COLOMBIA		580	
2	1176914599423	04OCT1999	15OCT1999	NEED	1 1	A1	45 AB	\$72.00	BETWEEN 41 - 60	\$100 & BELOW	ALBERTA		11	
3	1176951913656	01JUL2000	.		0 1	A1	57 BC	\$593.00	BETWEEN 41 - 60	\$500 - \$800	BRITISH COLOMBIA		203	
4	117695400288	30MAY2000	.		1 2	A1	47 ON	\$83.00	BETWEEN 41 - 60	\$100 & BELOW	ONTARIO		235	
5	117699186303	13DEC2000	.		1 1	C1	82 BC	.	60 & MORE	\$100 & BELOW	BRITISH COLOMBIA		38	
6	1176991056273	31AUG1999	18SEP2000	MOVE	1 1	C1	92 QC	\$1,041.00	60 & MORE	800 & ABOVE	QUEBEC		384	
7	1176991866552	24MAY2000	.		1 1	A1	77 ON	.	60 & MORE	\$100 & BELOW	ONTARIO		241	
8	1176992889500	28NOV2000	.		1 1	C1	68 AB	\$72.00	60 & MORE	\$100 & BELOW	ALBERTA		53	
9	1177000067271	23DEC1999	.		0 1	B1	75 ON	\$134.00	60 & MORE	\$100 - \$500	ONTARIO		394	
10	1177010940613	09DEC1999	.		1 2	A1	42 NS	\$11.00	BETWEEN 41 - 60	\$100 & BELOW	NOVA SCOTIA		408	
11	1177025997013	09NOV1999	.		1 1	A1	28 BC	\$154.00	BETWEEN 21 & 40	\$100 - \$500	BRITISH COLOMBIA		438	
12	1177027515760	19OCT1999	.		1 1	B1	73 BC	\$16.00	60 & MORE	\$100 & BELOW	BRITISH COLOMBIA		459	
13	1177028996676	21SEP2000	.		0 1	C1	.	\$179.00	20 OR LESS	\$100 - \$500	QUEBEC		121	
14	1177038747105	14MAR2000	.		0 1	C1	41 ON	\$705.00	BETWEEN 41 - 60	\$500 - \$800	ONTARIO		312	
15	1177045857516	22JUN2000	.		1 1	A1	53 QC	\$83.00	BETWEEN 41 - 60	\$100 & BELOW	QUEBEC		212	
16	1177057406016	21SEP2000	.		0 1	C1	50 ON	\$529.00	BETWEEN 41 - 60	\$500 - \$800	ONTARIO		121	
17	1177068422248	26APR1999	15JAN2001	NEED	0 1	A2	55 NS	\$44.00	BETWEEN 41 - 60	\$100 & BELOW	NOVA SCOTIA		630	
18	1177089399155	17AUG2000	.		1 3	A1	58 BC	\$548.00	BETWEEN 41 - 60	\$500 - \$800	BRITISH COLOMBIA		156	
19	1177113886410	13SEP2000	08JAN2001	COMP	0 1	C1	45 ON	\$83.00	BETWEEN 41 - 60	\$100 & BELOW	ONTARIO		117	
20	1177128264924	10DEC1999	.		1 1	B1	38 ON	\$178.00	BETWEEN 21 & 40	\$100 - \$500	ONTARIO		407	

Tenure in days for each account

The MEANS Procedure

Analysis Variable : tenure_days

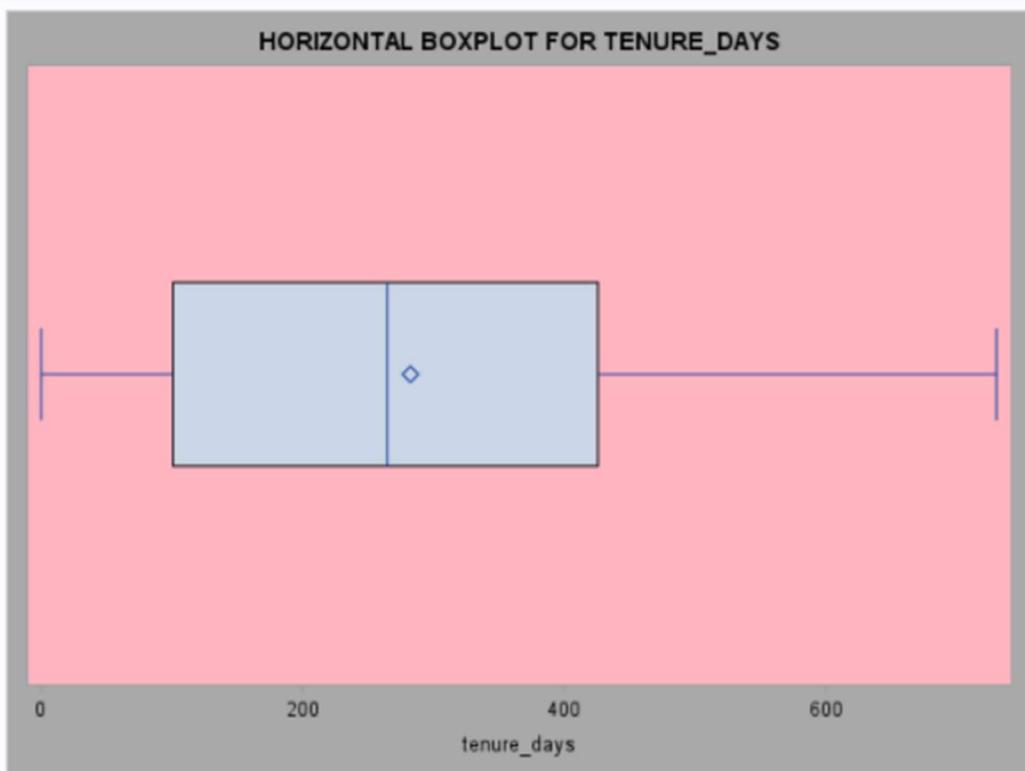
N	N Miss	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Quartile Range	Mean	Std Dev	Coeff of Variation	Lower 95%	Upper 95%
												CL for Mean
102255	0	0	101.0000000	285.0000000	428.0000000	731.0000000	325.0000000	282.5717960	197.3237108	69.8313538	281.3623396	283.7812524

```

%MACRO UNI_ANALYSIS_NUM(DATA,VAR);
  TITLE "HORIZONTAL BOXPLOT FOR &VAR";
  PROC SGPLOT DATA=&DATA;
    HBOX &VAR;
    STYLEATTRS
      BACKCOLOR=DARKGREY
      WALLCOLOR=LIGHTPINK
    ;
  RUN;
  TITLE "UNIVARIATE ANALYSIS FOR &VAR";
  proc means data=&DATA  N NMISS MIN Q1 MEDIAN MEAN Q3 MAX qrange cv clm maxdec=2 ;
  var &var;
  run;
  %MEND;

```

```
%UNI_ANALYSIS_NUM(telecom.tenure,TENURE_DAYS);
```



UNIVARIATE ANALYSIS FOR TENURE_DAYS

The MEANS Procedure

Analysis Variable : tenure_days												
N	N Miss	Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum	Quartile Range	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	
102255	0	0.00	101.00	265.00	282.57	426.00	731.00	325.00	69.83	281.36	283.78	

2) Calculate the number of accounts deactivated for each month?

Total accounts deactivated each month from 1999 to 2001

Number of monthly deactivation from 1999 - 2001

The FREQ Procedure

deact_month	Frequency	Percent
Jan	2494	12.70
Feb	553	2.82
Mar	760	3.87
Apr	731	3.72
May	914	4.65
Jun	1403	7.15
Jul	1380	7.03
Aug	1494	7.61
Sep	1717	8.74
Oct	2817	14.35
Nov	2076	10.57
Dec	3296	16.79
Frequency Missing = 82620		

Yearly Account deactivation based on months

YEAR - 1999

Number Of Accounts Deactivated For Each Month In 1999

Obs	deact_month	Deact_Number
1	Jan	2
2	Feb	21
3	Mar	40
4	Apr	45
5	May	123
6	Jun	245
7	Jul	282
8	Aug	402
9	Sep	421
10	Oct	516
11	Nov	449
12	Dec	599

YEAR - 2000

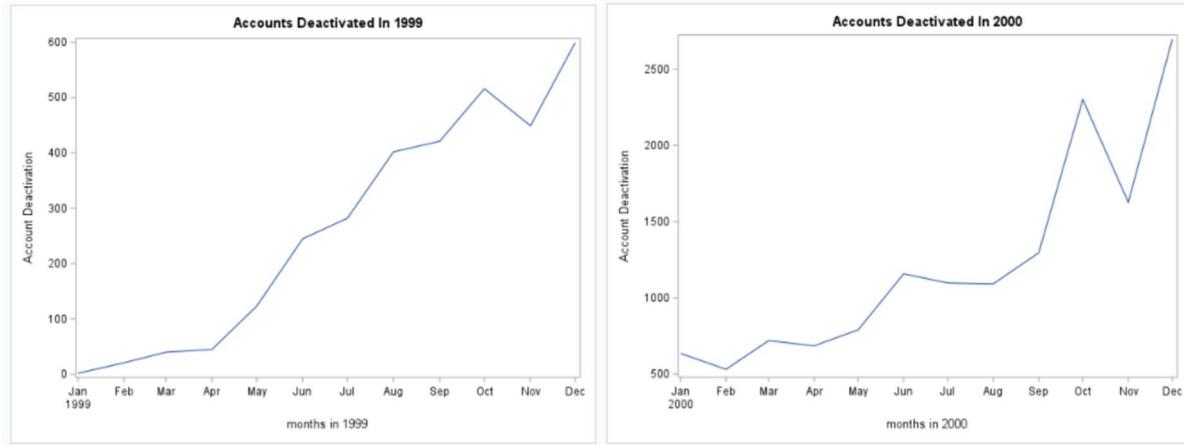
Number Of Accounts Deactivated For Each Month In 2000

Obs	deact_month	Deact_Number
1	Jan	636
2	Feb	532
3	Mar	720
4	Apr	686
5	May	791
6	Jun	1158
7	Jul	1098
8	Aug	1092
9	Sep	1296
10	Oct	2301
11	Nov	1627
12	Dec	2697

YEAR - 2001

Number Of Accounts Deactivated For Each Month In 2001

Obs	deact_month	Deact_Number
1	Jan	1856



3) Segment the account, first by account status “Active” and “Deactivated”, then by Tenure: < 30 days, 31---60 days, 61 days---one year, over one year. Report the number of accounts of percent of all for each segment.

```
Title "Tenure in days for each account";
data telecom.tenure;
set telecom.segments;
*reference_date = "20JAN2001"d;
if deactdt = '' then tenure_days = intck('day',actdt,"20JAN2001"d);
else
tenure_days = intck('day',actdt,deactdt);
run;
```

Tenure in days for each account															
Obs	acctno	actdt	Deactdt	Deactreason	GoodCredit	Rateplan	DealerType	Age	Province	Sales	AGE_GROUP	SALES_GROUP	PROVINCE	tenure_days	
1	117691319483	20JUN1999	.		0	1	A1	58	BC	\$128.00	BETW EEN 41 - 60	\$100 - \$500	BRITISH COLOMBIA	580	
2	1176914659423	04OCT1999	15OCT1999	NEED	1	1	A1	45	AB	\$72.00	BETW EEN 41 - 60	\$100 & BELOW	ALBERTA	11	
3	1176951913856	01JUL2000	.		0	1	A1	57	BC	\$593.00	BETW EEN 41 - 60	\$500 - \$800	BRITISH COLOMBIA	203	
4	1176954000288	30MAY2000	.		1	2	A1	47	ON	\$83.00	BETWEEN 41 - 60	\$100 & BELOW	ONTARIO	235	
5	117695186303	13DEC2000	.		1	1	C1	82	BC	.	60 & MORE	\$100 & BELOW	BRITISH COLOMBIA	38	
6	1176991056273	31AUG1999	18SEP2000	MOVE	1	1	C1	92	QC	\$1,041.00	60 & MORE	800 & ABOVE	QUEBEC	384	
7	1176991866552	24MAY2000	.		1	1	A1	77	ON	.	60 & MORE	\$100 & BELOW	ONTARIO	241	
8	1176952889600	28NOV2000	.		1	1	C1	68	AB	\$72.00	60 & MORE	\$100 & BELOW	ALBERTA	53	
9	1177000067271	23DEC1999	.		0	1	B1	75	ON	\$134.00	60 & MORE	\$100 - \$500	ONTARIO	394	
10	1177010940813	09DEC1999	.		1	2	A1	42	NS	\$11.00	BETW EEN 41 - 60	\$100 & BELOW	NOVA SCOTIA	408	
11	1177025897013	09NOV1999	.		1	1	A1	26	BC	\$154.00	BETW EEN 21 & 40	\$100 - \$500	BRITISH COLOMBIA	438	
12	1177027515760	19OCT1999	.		1	1	B1	73	BC	\$16.00	60 & MORE	\$100 & BELOW	BRITISH COLOMBIA	459	
13	1177028996676	21SEP2000	.		0	1	C1	70	QC	\$179.00	20 OR LESS	\$100 - \$500	QUEBEC	121	
14	1177038747105	14MAR2000	.		0	1	C1	41	ON	\$705.00	BETW EEN 41 - 60	\$500 - \$800	ONTARIO	312	
15	1177045657516	22JUN2000	.		1	1	A1	53	QC	\$83.00	BETW EEN 41 - 60	\$100 & BELOW	QUEBEC	212	
16	1177057408016	21SEP2000	.		0	1	C1	50	ON	\$529.00	BETW EEN 41 - 60	\$500 - \$800	ONTARIO	121	
17	1177066422248	26APR1999	15JAN2001	NEED	0	1	A2	55	NS	\$44.00	BETW EEN 41 - 60	\$100 & BELOW	NOVA SCOTIA	630	
18	1177089399155	17AUG2000	.		1	3	A1	56	BC	\$548.00	BETW EEN 41 - 60	\$500 - \$800	BRITISH COLOMBIA	156	
19	1177113886410	13SEP2000	08JAN2001	COMP	0	1	C1	45	ON	\$63.00	BETW EEN 41 - 60	\$100 & BELOW	ONTARIO	117	
20	1177128264924	10DEC1999	.		1	1	B1	38	ON	\$178.00	BETW EEN 21 & 40	\$100 - \$500	ONTARIO	407	

```
data telecom.Account_segment ;
set telecom.tenure;
length Account_Status $25;
length Tenure_Segment $30;
if deactdt = '' then Account_Status = "Active";
else Account_Status = "DeActivated";

if tenure_days <30 then Tenure_Segment ="Less than 30 days";
else if tenure_days <60 then Tenure_Segment = "Between 31 and 60 days";
else if tenure_days <365 then Tenure_Segment = "Between 60 days and 1 year";
else if tenure_days > 365 then Tenure_Segment = "Over 1 year";
run;
```

Tenure in days for each account																
Obs	acctno	acctdt	Deacctdt	Deactreason	GoodCredit	Ratplan	DealerType	Age	Province	Sales	AGE_GROUP	SALES_GROUP	PROVINCE	tenure_days	Account_Status	Tenure_Segment
1	117691319483	20JUN1999	-	-	0	1	A1	58	BC	\$128.00	BETWEEN 41-60	\$100 - \$500	BRITISH COLUMBIA	580	Active	Over 1 year
2	1176914589423	04OCT1999	15OCT1999	NEED	1	1	A1	45	AB	\$72.00	BETWEEN 41-60	\$100 & BELOW	ALBERTA	11	DeActvated	Less than 30 days
3	1176951913656	01JUL2000	-	-	0	1	A1	57	BC	\$593.00	BETWEEN 41-60	\$500 - \$800	BRITISH COLUMBIA	203	Active	Between 60 days and 1 year
4	117695400288	30MAY2000	-	-	1	2	A1	47	ON	\$83.00	BETWEEN 41-60	\$100 & BELOW	ONTARIO	235	Active	Between 60 days and 1 year
5	1176959186303	13DECE2000	-	-	1	1	C1	82	BC	-	60 & MORE	\$100 & BELOW	BRITISH COLUMBIA	38	Active	Between 31 and 60 days
6	117699106273	31AUG1999	18SEP2000	MOVE	1	1	C1	92	QC	\$1,041.00	60 & MORE	\$800 & ABOVE	QUEBEC	384	DeActvated	Over 1 year
7	117699186652	24MAY2000	-	-	1	1	A1	77	ON	-	60 & MORE	\$100 & BELOW	ONTARIO	241	Active	Between 60 days and 1 year
8	1176992889500	28NOV2000	-	-	1	1	C1	68	AB	\$72.00	60 & MORE	\$100 & BELOW	ALBERTA	53	Active	Between 31 and 60 days
9	1177000067271	23OCTC1999	-	-	0	1	B1	75	ON	\$134.00	60 & MORE	\$100 - \$500	ONTARIO	394	Active	Over 1 year
10	1177010490613	09DEC1999	-	-	1	2	A1	42	NS	\$11.00	BETWEEN 41-60	\$100 & BELOW	NOVA SCOTIA	408	Active	Over 1 year
11	1177025997013	09NOV1999	-	-	1	1	A1	26	BC	\$154.00	BETWEEN 21 & 40	\$100 - \$500	BRITISH COLUMBIA	438	Active	Over 1 year
12	1177027515760	19OCT1999	-	-	1	1	B1	73	BC	\$16.00	60 & MORE	\$100 & BELOW	BRITISH COLUMBIA	459	Active	Over 1 year
13	1177028996676	21SEP2000	-	-	0	1	C1	-	QC	\$179.00	20 OR LESS	\$100 - \$500	QUEBEC	121	Active	Between 60 days and 1 year
14	1177038747105	14MARA2000	-	-	0	1	C1	41	ON	\$70.00	BETWEEN 41-60	\$500 - \$800	ONTARIO	312	Active	Between 60 days and 1 year
15	1177045857516	22JUN2000	-	-	1	1	A1	53	QC	\$83.00	BETWEEN 41-60	\$100 & BELOW	QUEBEC	212	Active	Between 60 days and 1 year
16	1177057408016	21SEP2000	-	-	0	1	C1	50	ON	\$529.00	BETWEEN 41-60	\$500 - \$800	ONTARIO	121	Active	Between 60 days and 1 year
17	117706422248	26APR1999	15JAN2001	NEED	0	1	A2	55	NS	\$44.00	BETWEEN 41-60	\$100 & BELOW	NOVA SCOTIA	630	DeActvated	Over 1 year
18	1177089399155	17AUG2000	-	-	1	3	A1	56	BC	\$54.00	BETWEEN 41-60	\$500 - \$800	BRITISH COLUMBIA	156	Active	Between 60 days and 1 year
19	1177113866100	13SEP2000	08JAN2001	COMP	0	1	C1	45	ON	\$63.00	BETWEEN 41-60	\$100 & BELOW	ONTARIO	117	DeActvated	Between 60 days and 1 year
20	1177128264924	10DEC1999	-	-	1	1	B1	38	ON	\$178.00	BETWEEN 21 & 40	\$100 - \$500	ONTARIO	407	Active	Over 1 year

```
title "Account Status";
|proc freq data = telecom.Account_Segment;
table Account_Status;
run;
title "Tenure Segmentation";
|proc freq data = telecom.Account_Segment;
table Tenure_Segment;
run;
```

Tenure Segmentation

The FREQ Procedure

Account Status					The FREQ Procedure				
Account_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent	Tenure_Segment	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Active	82620	80.80	82620	80.80	Between 31 and 60 days	8586	8.41	8586	8.41
DeActivated	19635	19.20	102255	100.00	Between 60 days and 1 year	45405	44.47	53991	52.87
					Less than 30 days	9486	9.29	63477	62.16
					Over 1 year	38636	37.84	102113	100.00
					Frequency Missing = 142				

4) Test the general association between the tenure segments and “Good Credit” “RatePlan ” and “DealerType.”

```

data telecom.credit ;
set telecom.Account_Segment;
if goodcredit = 1 then Credit_Type = "Good";
else Credit_Type = "Bad";
run;

proc print data = telecom.credit (obs = 100);
run;

PROC OPTIONS OPTION = MACRO;
RUN;

%MACRO BI_ANALYSIS_CAT_CAT (DSN = ,CLASS= , VAR= );
PROC FREQ DATA =&DSN;
TITLE " RELATION BETWEEN &VAR. AND &CLASS.";
TABLE &VAR.*&CLASS/chisq;
PROC SGPOINT DATA = &DSN;
VBAR &VAR/GROUP = &CLASS GROUPDISPLAY = STACK;
RUN;
%MEND BI_ANALYSIS_CAT_CAT;

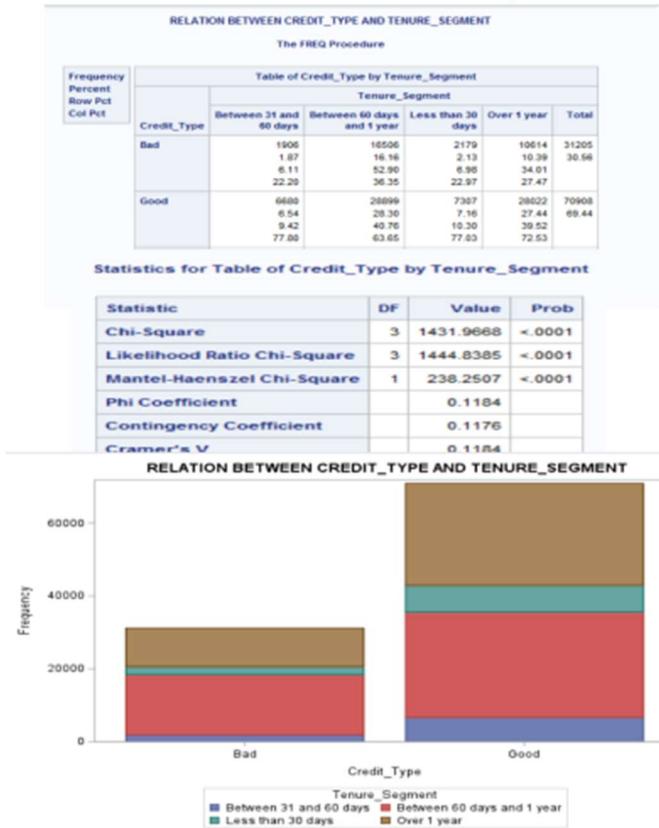
%BI_ANALYSIS_CAT_CAT(DSN =telecom.credit ,CLASS = TENURE_SEGMENT, VAR = CREDIT_TYPE);
%BI_ANALYSIS_CAT_CAT(DSN =telecom.credit ,CLASS = TENURE_SEGMENT, VAR = RATEPLAN);
%BI_ANALYSIS_CAT_CAT(DSN =telecom.credit ,CLASS = TENURE_SEGMENT, VAR = DEALERTYPE);

```

Tenure Segmentation

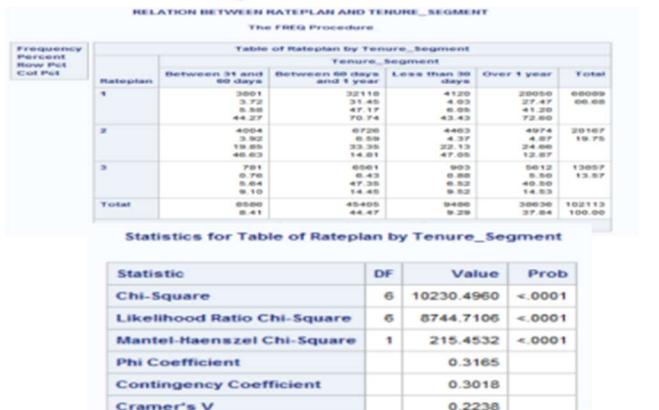
Obs	acctno	acctd	Deactdt	Deactreason	GoodCredit	Rateplan	DealerType	Age	Province	Sales	AGE_GROUP	SALES_GROUP	PROVINCE	tenure_days	Account_Status	Tenure_Segment	Credit_Type
1	117691319448	20JUN1999	.	.	0 1	A1	58	BC	\$128.00	BETWEEN 41-60	\$100 - \$500	BRITISH COLOMBIA	580	Active	Over 1 year	Bad	
2	1176914599423	04OCT1999	15OCT1999	NEED	1 1	A1	45	AB	\$72.00	BETWEEN 41-60	\$100 & BELOW	ALBERTA	11	DeActivated	Less than 30 days	Good	
3	117691913656	01JUL2000	.	.	0 1	A1	57	BC	\$593.00	BETWEEN 41-60	\$500 - \$800	BRITISH COLOMBIA	203	Active	Between 60 days and 1 year	Bad	
4	117695400288	30MAY2000	.	.	1 2	A1	47	ON	\$83.00	BETWEEN 41-60	\$100 & BELOW	ONTARIO	235	Active	Between 60 day s and 1 year	Good	
5	1176969186303	13OE2000	.	.	1 1	C1	82	BC	.	60 & MORE	\$100 & BELOW	BRITISH COLOMBIA	38	Active	Between 31 and 60 days	Good	
6	1176991056273	31AUG1999	18SEP2000	MOVE	1 1	C1	92	QC	\$1,041.00	60 & MORE	800 & ABOVE	QUEBEC	384	DeActivated	Over 1 year	Good	
7	1176991866552	24MAY2000	.	.	1 1	A1	77	ON	.	60 & MORE	\$100 & BELOW	ONTARIO	241	Active	Between 60 days and 1 year	Good	
8	1176992889500	28NOV2000	.	.	1 1	C1	68	AB	\$72.00	60 & MORE	\$100 & BELOW	ALBERTA	53	Active	Between 31 and 60 days	Good	
9	117700006727	23OE1999	.	.	0 1	B1	75	ON	\$134.00	60 & MORE	\$100 - \$500	ONTARIO	394	Active	Over 1 year	Bad	
10	1177010940613	09OE1999	.	.	1 2	A1	42	NS	\$11.00	BETWEEN 41-60	\$100 & BELOW	NOVA SCOTIA	408	Active	Over 1 year	Good	
11	1177025997013	09NOV1999	.	.	1 1	A1	26	BC	\$154.00	BETWEEN 21 & 40	\$100 - \$500	BRITISH COLOMBIA	438	Active	Over 1 year	Good	
12	1177027515760	19OCT1999	.	.	1 1	B1	73	BC	\$16.00	60 & MORE	\$100 & BELOW	BRITISH COLOMBIA	459	Active	Over 1 year	Good	
13	1177028996676	21SEP2000	.	.	0 1	C1	.	QC	\$179.00	20 OR LESS	\$100 - \$500	QUEBEC	121	Active	Between 60 day s and 1 year	Bad	
14	1177038747103	14MAR2000	.	.	0 1	C1	41	ON	\$705.00	BETWEEN 41-60	\$500 - \$800	ONTARIO	312	Active	Between 60 day s and 1 year	Bad	
15	1177045857516	22JUN2000	.	.	1 1	A1	53	QC	\$83.00	BETWEEN 41-60	\$100 & BELOW	QUEBEC	212	Active	Between 60 day s and 1 year	Good	
16	1177057406016	.	.	.	0 1	C1	50	ON	\$529.00	BETWEEN 41-60	\$500 - \$800	ONTARIO	121	Active	Between 60 day s and 1 year	Bad	
17	117706642248	26APR1999	15JAN2001	NEED	0 1	A2	55	NS	\$44.00	BETWEEN 41-60	\$100 & BELOW	NOVA SCOTIA	630	DeActivated	Over 1 year	Bad	
18	117708939915	17AUG2000	.	.	1 3	A1	56	BC	\$548.00	BETWEEN 41-60	\$500 - \$800	BRITISH COLOMBIA	156	Active	Between 60 days and 1 year	Good	
19	1177113886410	13SEP2000	08JAN2001	COMP	0 1	C1	45	ON	\$63.00	BETWEEN 41-60	\$100 & BELOW	ONTARIO	117	DeActivated	Between 60 days and 1 year	Bad	
20	1177128264924	10OE1999	.	.	1 1	B1	38	ON	\$178.00	BETWEEN 21 & 40	\$100 - \$500	ONTARIO	407	Active	Over 1 year	Good	

Tenure Segment vs Credit type



Majority of the customers are between the tenure of 60 days and 1 year, and over 1 year, and they all have good credit.

Tenure Segment vs Rate Plan



Majority of the customers are between the tenure of 60 days and 1 year, and over 1 year, and they are all with rate plan 1.

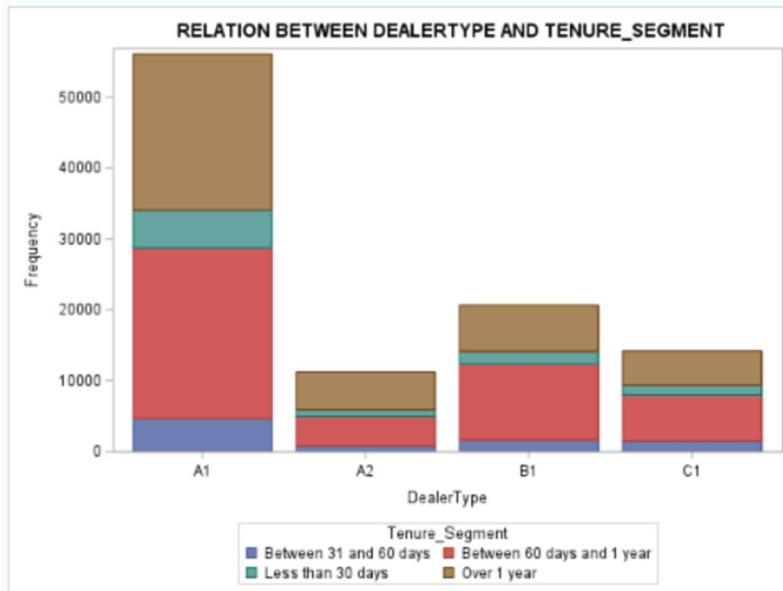
Tenure Segment vs Dealer Type

	Frequency	Percent	Row Pct	Col Pct	
	Table of DealerType by Tenure_Segment				
	Tenure_Segment				
DealerType	Between 31 and 60 days	Between 60 days and 1 year	Less than 30 days	Over 1 year	Total
A1	4646 4.55 8.29 54.11	24071 23.57 42.95 53.01	5361 5.25 9.56 56.51	21972 21.52 39.20 56.87	56050 54.89
A2	794 0.76 7.06 9.25	4189 4.10 37.26 9.23	909 0.09 0.08 9.58	5352 5.24 47.60 13.85	11244 11.01
B1	1645 1.01 7.97 19.16	10698 10.48 51.03 23.56	1787 1.75 8.66 16.84	6510 6.50 31.54 16.85	20640 20.21
C1	1501 1.47 10.59 17.48	6447 6.31 45.47 14.20	1429 1.40 10.08 15.06	4802 4.70 33.87 12.43	14179 13.89
Total	8586 8.41	45405 44.47	9486 9.29	38636 37.84	102113 100.00

Frequency Missing = 142

Statistics for Table of DealerType by Tenure_Segment

Statistic	DF	Value	Prob
Chi-Square	9	1152.0844	<.0001
Likelihood Ratio Chi-Square	9	1136.7013	<.0001
Mantel-Haenszel Chi-Square	1	337.0862	<.0001
Phi Coefficient		0.1062	
Contingency Coefficient		0.1056	
Cramer's V		0.0613	



Majority of the customers are between the tenure of 60 days and 1 year, and over 1 year, and they are all with dealer A1.

5) Is there any association between the account status and the tenure segments?

```
%MACRO BI_ANALYSIS_CAT_CAT (DSN = ,CLASS= , VAR= );
PROC FREQ DATA =&DSN;
TITLE "| RELATION BETWEEN &VAR. AND &CLASS.";
TABLE &VAR.*&CLASS/chisq;
PROC SGPlot DATA = &DSN;
VBAR &VAR/GROUP = &CLASS GROUPDISPLAY = CLUSTER;
RUN;
%MEND BI_ANALYSIS_CAT_CAT;

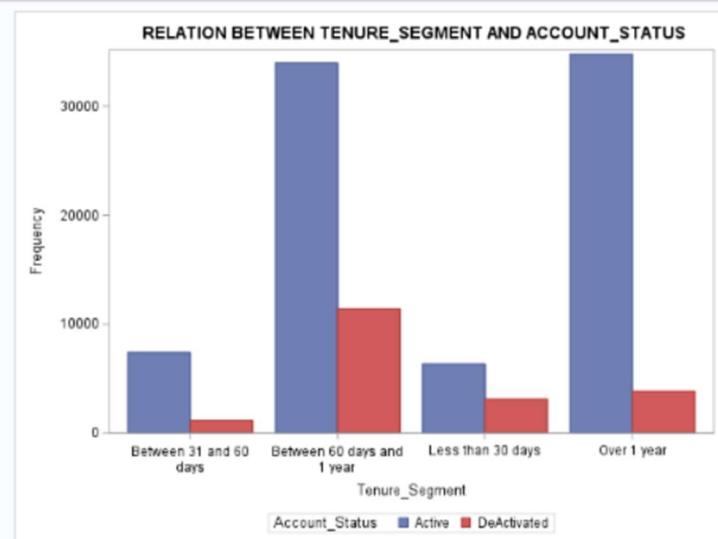
%BI_ANALYSIS_CAT_CAT(DSN =telecom.credit ,CLASS =ACCOUNT_STATUS , VAR = TENURE_SEGMENT);
```

RELATION BETWEEN TENURE_SEGMENT AND ACCOUNT_STATUS				
The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of Tenure_Segment by Account_Status			
	Tenure_Segment	Active	DeActivated	Total
Between 31 and 60 days	7398 7.24 88.18 8.97	1188 1.16 13.84 6.06		8586 8.41 86.18 8.06
Between 60 days and 1 year	33977 33.27 74.83 41.18	11428 11.19 25.17 58.28		45405 44.47
Less than 30 days	6353 6.22 68.97 7.70	3133 3.07 33.03 15.98		9486 9.29
Over 1 year	34777 34.08 90.01 42.15	3859 3.78 9.99 19.66		38636 37.84
Total	82505 80.80	19808 19.20		102113 100.00
Frequency Missing = 142				

Statistics for Table of Tenure_Segment by Account_Status

Statistic	DF	Value	Prob
Chi-Square	3	4484.1283	<.0001
Likelihood Ratio Chi-Square	3	4624.2009	<.0001
Mantel-Haenszel Chi-Square	1	1634.5256	<.0001
Phi Coefficient		0.2096	
Contingency Coefficient		0.2051	
Cramer's V		0.2096	

Sample Size = 102113
Frequency Missing = 142



Number of customers de-activated is highest between tenure of 60 days and 1 year. Deactivated accounts are lower when the tenure is over 1 year.

Could you find out a better tenure segmentation strategy that is more associated with the account status?

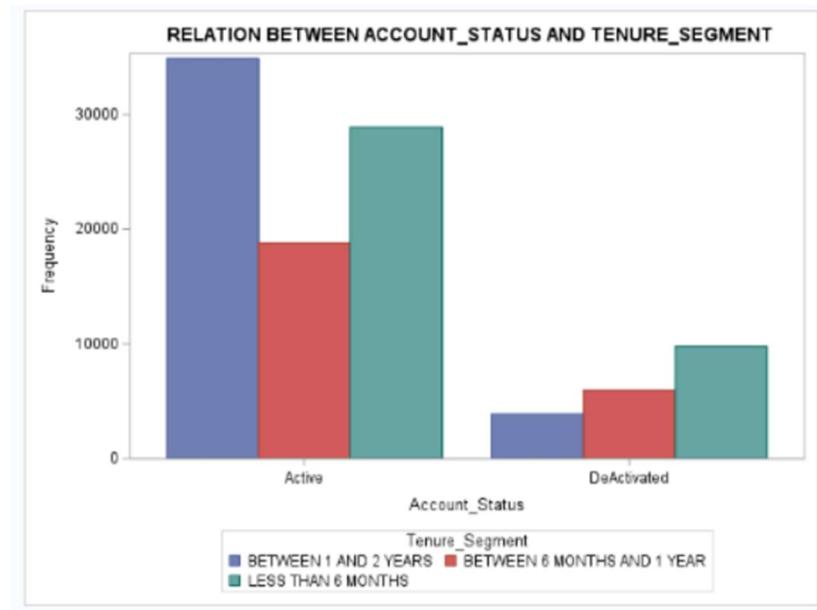
```

DATA telecom.TENURE_SEGNEW;
SET telecom.credit;
IF TENURE_DAYS < 183 THEN TENURE_SEGMENT ="LESS THAN 6 MONTHS";
ELSE IF TENURE_DAYS <365 then TENURE_SEGMENT ="BETWEEN 6 MONTHS AND 1 YEAR";
ELSE IF TENURE_DAYS <=731 then TENURE_SEGMENT ="BETWEEN 1 AND 2 YEARS";
RUN;

proc print data = telecom.TENURE_SEGNEW (obs=100);
run;

%BI_ANALYSIS_CAT_CAT(DSN =telecom.TENURE_SEGNEW ,CLASS =TENURE_SEGMENT , VAR = ACCOUNT_STATUS);

```



RELATION BETWEEN ACCOUNT_STATUS AND TENURE_SEGMENT

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Account_Status by Tenure_Segment				Statistics for Table of Account_Status by Tenure_Segment	
	Account_Status	Tenure_Segment				
		BETWEEN 1 AND 2 YEARS	BETWEEN 6 MONTHS AND 1 YEAR	LESS THAN 6 MONTHS		
Active	Active	34892	18820	28908	82520	
		34.12	18.40	28.27	80.80	
		42.23	22.78	34.99		
		89.98	75.98	74.68		
DeActivated	DeActivated	3888	5950	9799	19635	
		3.80	5.82	9.58	19.20	
		19.79	30.30	49.91		
		10.02	24.02	25.32		
Total	Total	38778	24770	38707	102255	
		37.92	24.22	37.85	100.00	

Sample Size = 102255

Number of active customers is highest for tenure segment between 1 and 2 years. Where as, number of de-activated customers is lowest for the tenure between 1 and 2 years and highest for tenure segment less than 6 months.

6. Does Sales amount differ among different account status?

The UNIVARIATE Procedure				
Variable: Sales				
Account_Status = DeActivated				
Moments				
N	17975	Sum Weights	17975	
Mean	179.925118	Sum Observations	3234154	
Std Deviation	231.770287	Variance	53717.4661	
Skewness	2.38570044	Kurtosis	5.38837505	
Uncorrected SS	1547423276	Corrected SS	985517735	
Coeff Variation	128.814859	Std Error Mean	1.72871464	
Basic Statistical Measures				
Location		Variability		
Mean	179.9251	Std Deviation	231.77029	
Median	91.0000	Variance	53717	
Mode	55.0000	Range	1199	
		Interquartile Range	135.00000	
Tests for Location: Mu0=0				
Test	Statistic	p Value		
Student's t	t	104.0803	Pr > t	<.0001
Sign	M	8981.5	Pr >= M	<.0001
Signed Rank	S	80871833	Pr >= S	<.0001
Tests for Normality				
Test	Statistic	p Value		
Kolmogorov-Smirnov	D	0.247594	Pr > D	<0.0100
Cramer-von Mises	W-Sq	409.5501	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2162.105	Pr > A-Sq	<0.0050

Test of Equality of variance

Levene's Test for Homogeneity of Variances is produced by PROC GLM to test if variances are considered equal across all groups .

From Levene's test result we find that p value is >0.0505. As the P value is greater than 5%, we accept null hypothesis and conclude variances are equal.

The GLM Procedure					
Levene's Test for Homogeneity of Sales Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	FValue	Pr > F
Account_Status	1	112914	112914	3.83	0.0505
Error	93648	2.7641E9	29516.2		

Since p value <0.05. As the p value is less than 5% we conclude that data is not normally distributed

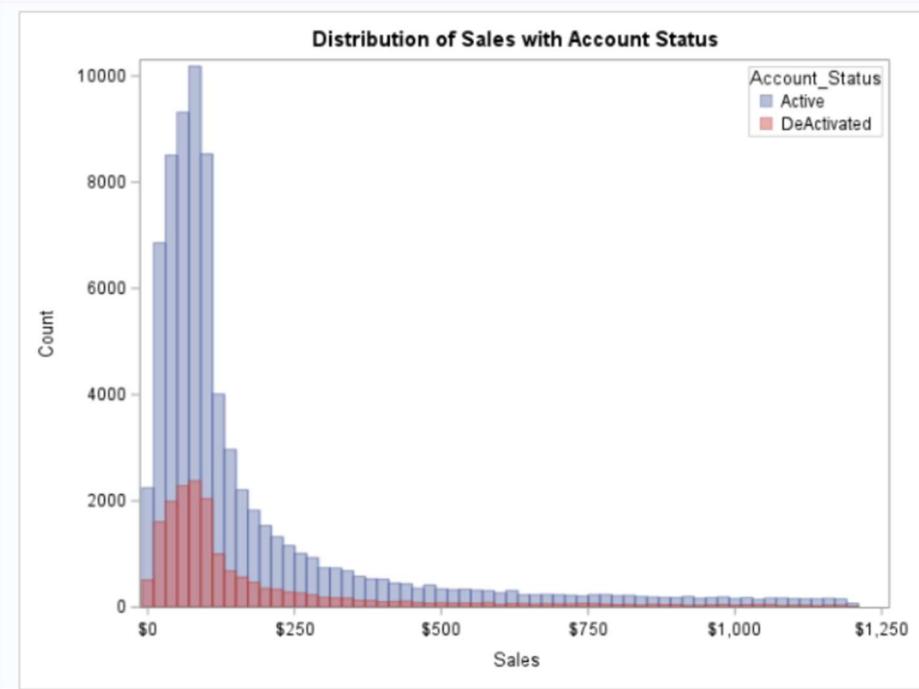
Test of Independency

Account_Status	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Active		75675	181.6	234.5	0.8524	0	1200.0
DeActivated		17975	179.9	231.8	1.7287	0	1199.0
Diff (1-2)	Pooled		1.6348	234.0	1.9414		
Diff (1-2)	Satterthwaite		1.6348		1.9274		

Account_Status	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
Active		181.6	179.9	183.2	234.5
DeActivated		179.9	176.5	183.3	231.8
Diff (1-2)	Pooled	1.6348	-2.1702	5.4399	234.0
Diff (1-2)	Satterthwaite	1.6348	-2.1431	5.4127	235.0

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	93648	0.84	0.3997
Satterthwaite	Unequal	27392	0.85	0.3963

Equality of Variances					
Method	Num DF	Den DF	F Value	Pr> F	
Folded F	75674	17974	1.02	0.0475	



Visualization is done using grouped histogram. Both active and de-activated accounts follow the same pattern in terms of sales, where majority of the sales are happening under \$250

As our categorical variable Account status has 2 levels ,we perform T test for test of independency. Since the Levene's test p-value > 0.05 and variance are found to be equal pooled T test is appropriate for testing independency

Based on the result from pooled T test , the p value >0.05, hence we fail to reject null hypothesis and find that there is no statistical relationship between sales and Account status.

Does Sales amount differ among age group?

The UNIVARIATE Procedure				
Variable: Sales				
AGE_GROUP = BETWEEN 21 & 40				
Moments				
N		24146	Sum Weights	24146
Mean		182.456804	Sum Observations	4405602
Std Deviation		235.583154	Variance	55499.4225
Skewness		2.3438425	Kurtosis	5.1517088
Uncorrected SS		2143865620	Corrected SS	1340033557
Coeff Variation		129.11722	Std Error Mean	1.51607831
Basic Statistical Measures				
Location		Variability		
Mean	182.4568	Std Deviation	235.58315	
Median	91.0000	Variance	55499	
Mode	86.0000	Range	1200	
		Interquartile Range	142.00000	
Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student'st		t	120.3479	Pr > t <.0001
Sign		M	12062.5	Pr >= M <.0001
Signed Rank		S	1.4551E8	Pr >= S <.0001

Test of Normality

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.24979	Pr > D	<0.0100
Cramer-von Mises	W-Sq	547.8163	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2886.84	Pr > A-Sq	<0.0050

p value is found to be <0.05. As the p value is less than 5% we conclude that data is not normally distributed

Test of Equality of variance

Levene's Test for Homogeneity of Variances is produced by PROC GLM to test if variances are considered equal across all groups . From Levene's test result we find that p value is <0.2203. As the P value is more than 5%, we fail to reject null hypothesis and conclude variances are equal.

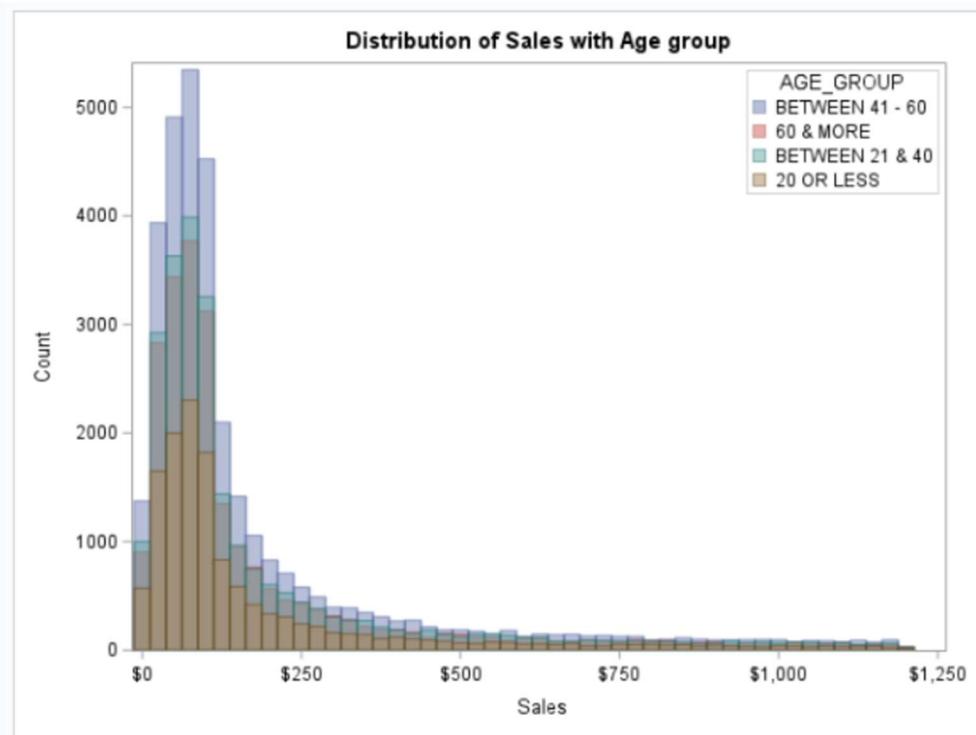
The GLM Procedure					
Levene's Test for Homogeneity of Sales Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
AGE_GROUP	3	130221	43407.0	1.47	0.2203
Error	93646	2.7641E9	29516.5		

Test of Independency using One Way Anova

The ANOVA Procedure					
Dependent Variable: Sales					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	81794	27265	0.50	0.6836
Error	93646	5126493686	54743		
Corrected Total	93649	5126575480			

R-Square	Coeff Var	Root MSE	Sales Mean
0.000016	129.0912	233.9729	181.2462

Source	DF	Anova SS	Mean Square	F Value	Pr > F
AGE_GROUP	3	81794.34973	27264.78324	0.50	0.6836



Similarly, all the age groups follow the same pattern in terms of sales, where majority of the sales are happening under \$250

As groups have equal variance we consider p-value of the standard one-way ANOVA results for test of independency

Based on the result from Anova, the p value is 0.683 , as it is more than 5 % we fail to reject null hypothesis there is a no statistical relationship between sales and Age group.

Does Sales amount differ among credit type

```
proc univariate data=telecom.TENURE_SEGNEW normal;
  class credit_type;
  var sales;
  run;

%BI_ANALYSIS_NUMS_CAT (DSN = telecom.TENURE_SEGNEW ,CLASS=CREDIT_TYPE , VAR=sales );
```

The UNIVARIATE Procedure

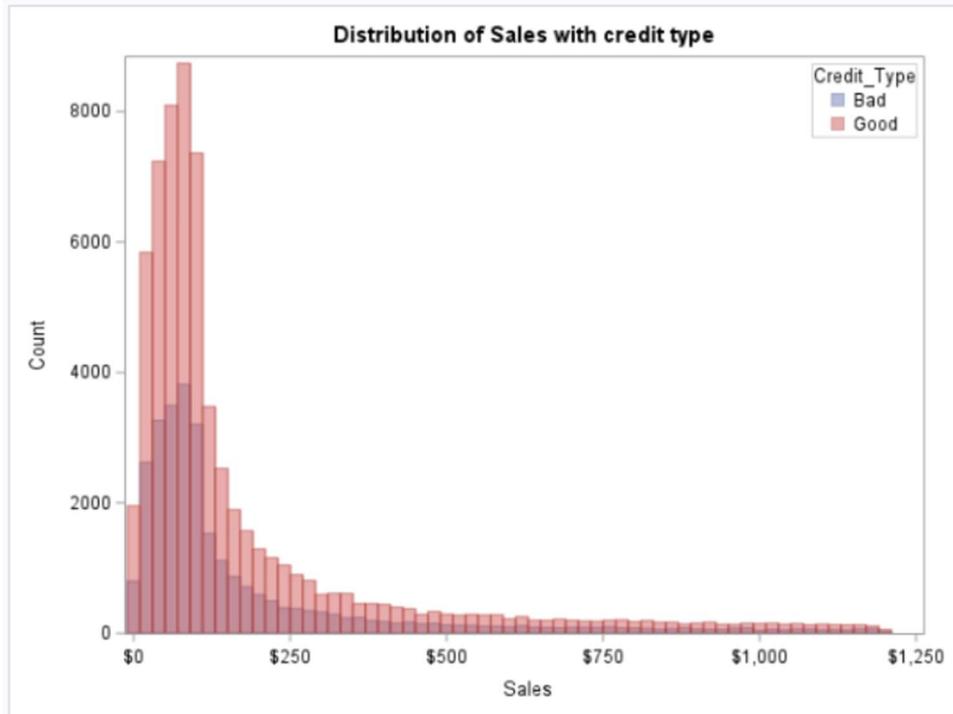
Variable: Sales

Credit_Type = Good

Moments			
N	65051	Sum Weights	65051
Mean	181.103795	Sum Observations	11780983
Std Deviation	233.590787	Variance	54564.6555
Skewness	2.36258058	Kurtosis	5.25896457
Uncorrected SS	5683011579	Corrected SS	3549430843
Coeff Variation	128.981718	Std Error Mean	0.91585922

Basic Statistical Measures			
Location		Variability	
Mean	181.1038	Std Deviation	233.59079
Median	91.0000	Variance	54565
Mode	82.0000	Range	1200
		Interquartile Range	137.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	197.742	Pr > t	<.0001
Sign	M	32503	Pr >= M	<.0001
Signed Rank	S	1.0565E9	Pr >= S	<.0001



Similarly, both bad and good credit types follow the same pattern in terms of sales, where majority of the sales are happening under \$250

Test of Normality

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.250311	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1479.829	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	7799.709	Pr > A-Sq	<0.0050

Since p value is less than 0.0001 which is <0.05 . As the p value is less than 5% we conclude that data is not normally distributed

Test of Equality of Variance

Levene's Test for Homogeneity of Variances is produced by PROC GLM to test if variances are considered equal across all groups . From Levene's test result we find that p value is >0.6795. As the P value is greater than 5%, we accept null hypothesis and conclude variances are equal.

The GLM Procedure					
Levene's Test for Homogeneity of Sales Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Credit_Type	1	5039.1	5039.1	0.17	0.6795
Error	93648	2.7643E9	29517.5		

Test of Independency – T test

The TTEST Procedure							
Variable: Sales							
Account_Status	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Active		75675	181.6	234.5	0.8524	0	1200.0
DeActivated		17975	179.9	231.8	1.7287	0	1199.0
Diff (1-2)	Pooled		1.6348	234.0	1.9414		
Diff (1-2)	Satterthwaite		1.6348		1.9274		

Account_Status	Method	Mean	95% CL Mean	Std Dev	95% CL	Std Dev	
Active		181.6	179.9	183.2	234.5	233.3	235.7
DeActivated		179.9	176.5	183.3	231.8	229.4	234.2
Diff (1-2)	Pooled	1.6348	-2.1702	5.4399	234.0	232.9	235.0
Diff (1-2)	Satterthwaite	1.6348	-2.1431	5.4127			

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	93648	0.84	0.3997
Satterthwaite	Unequal	27392	0.85	0.3963

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	75674	17974	1.02	0.0475

As our categorical variable credit type has 2 levels ,we perform T test for test of independency. Since the Levene's test p-value > 0.05 and variance are found to be equal, pooled T test is appropriate for testing independency

Based on the result from pooled T test , the p value >0.05, hence we accept null hypothesis and that there is no statistical relationship between sales and Account status.

CONCLUSIONS

- Based on the data, it can be noticed that majority of the customers are between age group of 21- 40 and 41-60 years, and they are from Ontario and British Columbia. This is the same case for both active and de-activated customers.
- Irrespective of age and province, sales was highest for category \$100 and below. In each province, maximum number of customers belong to the age group between 41 and 60.
- Number of de-activated accounts increased over the months and, year on year.
- Majority of the customers are between the tenure of 60 days and 1 year, and over 1 year, and they all have good credit. Also, majority of the customers had taken rate plan 1 and with dealer A1.
- Number of active customers is highest for tenure segment between 1 and 2 years. Where as, number of deactivated customers is lowest for the tenure between 1 and 2 years and highest for tenure segment less than 6 months.

Recommendations

- The middle-aged population makes up the majority of the population overall. Once the middle-aged group has stabilized, we can cultivate new users within the 20–40 year old demographic.
- 30.56% of accounts have bad credit, while 69.44% of accounts have good credit. Too many people have bad credit accounts. In the future, a thorough credit check should be performed on the customer when opening an account.
- RatePlan #1's percentage of the total data is twice the total of RatePlans #2 and #3. This indicates that RatePlan#1 is more well-liked, so we might think about releasing additional plans similar to it in the future.
- Number of deactivated accounts can be reduced by offering promotions and discounts for customers who are less than 6 months tenure.
- By providing continued excellent customer service, it may possible to retain the active customers.

Appendix – SAS code

*Small Project for the Telecom company

Customer Distribution and Deactivation Analyses
Objective: Analyzing and performing EDA on a CRM data of a wireless company for 2 years.
investigating the customer distribution and business behaviors, and then gaining insightful understanding about the customers, and to forecast the deactivation trends for the next 6 months.

Data:

Acctno: account number.
Actdt: account activation date
Deactdt: 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;

```
LIBNAME TELECOM "C:\DSA\Advanced SAS\Final Project";  
=====  
==  
IMPORTING DATA SET IN TO SAS  
=====  
=;  
  
Title " Importing Data Set in to SAS";  
  
data telecom.Details;  
infile "C:\DSA\Advanced SAS\Final Project\New_Wireless_Fixed.txt";  
input acctno 1-14  
@15 actdt mmddyy10.  
@26 Deactdt mmddyy10.  
Deactreason $ 41-45  
GoodCredit 53  
Rateplan $62-63  
DealerType $65-66  
Age 74 -75  
Province $ 80-81  
Sales dollar11.2  
;  
format acctno 14.0 actdt date9. Deactdt date9. ;  
format Sales DOLLAR11.2;  
run;  
  
proc print data = telecom.Details(obs=10);  
run;
```

```

title;
*****
***** Analyzing the data set and Performing EDA *****
*****;

=====
Browsing Descriptive Portion
=====

=====;
proc contents data = telecom.Details order = varnum;
run;

=====
=====;
Browsing Head of data set
=====

=====;
Title "Browsing Head of Data Set";
proc print data = telecom.Details (obs = 10);
run;

=====
=====;
Browsing tail of data set
=====

=====;
Title "Browsing Tail of Data Set";
proc print data = telecom.Details (obs = 102255 firstobs =102246);
run;

=====
=====;
*Finding number of unique distinct values
=====

=====;
Title "Number of unique distinct values in each variables";
proc freq data =telecom.Details nlevels;
ods exclude onewayfreqs;
run;

proc freq data = telecom.Details;
table deactreason Dealertype province goodcredit/ nopercnt nocum;
run;

*****
***** *EDA AND DATA CLEANING; *****
*****;

*Removing Duplicated Data if any;
proc sort data = telecom.Details out = telecom.Data nodupkey;
by Acctno;

```

```

run;
*No duplicated account numbers were found so zero observations were deleted
;

=====
=====Checking for missing values=====
=====;
Title "Number of missing values";
proc means data = telecom.Data nmiss;
run;

proc sql;
select nmiss(Province) as Province,nmiss (deactreason) as deactreason,
nmiss(rateplan) as rateplan,nmiss(dealertype) as dealertype
from telecom.Data;
quit;

* Replacing missing values in age with mean age in new column as New_Age ;
PROC SQL;
CREATE TABLE TELECOM.WIRELESS_NOMISS AS
SELECT*,
COALESCE(AGE,MEDIAN(AGE)) as New_age
      FROM TELECOM.Details
;
QUIT;

PROC PRINT DATA = TELECOM.WIRELESS_NOMISS (OBS=50);
RUN;

=====
=====Descriptive Analysis of continous variables=====
=====;
TITLE"DESCRIPTIVE ANALYSIS OF CONTINUOUS";
PROC MEANS DATA = Telecom.Data N NMISS MIN Q1 MEDIAN Q3 MAX qrange mean std
cv clm;
RUN;

PROC UNIVARIATE DATA = Telecom.Data;RUN;
-----
-----
QUESTION 1
1.1 Explore and describe the dataset briefly. For example, is the acctno
unique? What
is the number of accounts activated and deactivated? When is the earliest and
latest activation/deactivation dates available?
-----
-----;

*1.1a Explore and describe the dataset briefly;
TITLE "DESCRIPTION OF DATASET";
PROC CONTENTS DATA = TELECOM.DATA;
RUN;

```

```

*Checking if All Account Numbers are Unique ?;
*Removing Duplicated Data if any;
proc sort data = telecom.Details out = telecom.Data nodupkey;
by Acctno;
run;

*Number of unique accounts;
PROC SQL OUTOBS=20;
SELECT COUNT(*) AS TOTAL_COUNT,
       COUNT(DISTINCT Acctno) AS UNIQUE_ACCOUNTS
FROM TELECOM.DATA
;
QUIT;
*No duplicated account numbers were found so zero observations were deleted
;

*what is the number of accounts activated and deactivated ?;
Title "Number of Activated and Deactivated Accounts";
PROC SQL;
SELECT COUNT(Acctno) AS Total_Accounts,
       (COUNT(Actdt) - COUNT(Deactdt)) AS Activted_Accounts,
       COUNT(Deactdt) AS Deactvated_Accounts
FROM telecom.Data;
QUIT;

*When is the earliest and latest activation/deactivation dates available?;
*Earliest and Latest Activation Date;

Title "Earliest and Latest Activation Date";
proc sql;
select min(actdt) as Earliest_Activation_Date format = date9.,
       max(actdt) as Latest_Activation_Date format = date9.
from telecom.Data;
quit;

*Earliest and Latest DeActivation Date;

Title "Earliest and Latest DeActivation Date";
proc sql;
select min(deactdt) as Earliest_DeActivation_Date format = date9.,
       max(deactdt) as Latest_DeActivation_Date format = date9.
from telecom.Data;
quit;
title;
*QUESTION 2;
*1.2 What is the age and province distributions of active customers?;

data telecom.Account_segment ;
set telecom.Data;
length Age_Group $25;
IF AGE <= 20 THEN AGE_GROUP = "LESS THAN 20";
ELSE IF 21<= AGE<=40 THEN AGE_GROUP = "BETWEEN 21 AND 40 ";
ELSE IF 41<=AGE <=59 THEN AGE_GROUP ="BETWEEN 41 AND 60";
ELSE IF AGE >= 60 THEN AGE_GROUP = "60 AND ABOVE";
RUN;

proc print data = telecom.Account_segment (obs=20);

```

```

run;

Title"Age distributions of active customers";

PROC SQL ;
CREATE TABLE AGEDIST AS
SELECT AGE_GROUP,
       (COUNT(Actdt) - COUNT(Deactdt))AS TOTAL_ACTIVE_CUSTOMERS,
       SUM(SALES) AS TOTAL_SALES
FROM TELECOM.ACCTSEGMENT
WHERE DEACTDT IS NULL
GROUP BY AGE_GROUP
ORDER BY AGE_GROUP;
QUIT;

PROC PRINT DATA = AGEDIST;
RUN;

PROC GCHART DATA = TELECOM.ACCTSEGMENT;
PIE3D AGE_GROUP/PERCENT =INSIDE;
WHERE DEACTDT IS NULL;
RUN;
*=====;
PROC SQL ;
CREATE TABLE PROVINDIST AS
SELECT PROVINCE,
       (COUNT(Actdt) - COUNT(Deactdt))AS TOTAL_ACTIVE_CUSTOMERS,
       SUM(SALES) AS TOTAL_SALES
FROM TELECOM.ACCTSEGMENT
WHERE DEACTDT IS NULL ,PROVINCE IS NOT NULL
GROUP BY PROVINCE
ORDER BY PROVINCE;
QUIT;

*PROC PRINT DATA = PROVINDIST;
*WHERE PROVINCE IS NOT NULL;
*RUN;

Title"Distribution of Active Customers By Province ";
PROC GCHART DATA = TELECOM.ACCTSEGMENT;
PIE3D PROVINCE/PERCENT =INSIDE;
WHERE DEACTDT IS NULL;
RUN;

*=====;
PROC SQL ;
CREATE TABLE AGEPROVINDIST AS
SELECT PROVINCE,AGE_GROUP,
       (COUNT(Actdt) - COUNT(Deactdt))AS TOTAL_ACTIVE_CUSTOMERS,
       SUM(SALES) AS TOTAL_SALES
FROM TELECOM.ACCTSEGMENT
WHERE DEACTDT IS NULL
GROUP BY PROVINCE,AGE_GROUP
ORDER BY PROVINCE,AGE_GROUP;

```

```

QUIT;

Title"Distribution of Active Customers By Age and Province ";
PROC PRINT DATA = AGEPROVINDIST;
WHERE PROVINCE IS NOT NULL;
RUN;

*What is the age and province distributions of deactivated customers?;
Title"Distribution of DeActive Customers By Age and Province ";
PROC SQL ;
CREATE TABLE AGEPROVINDIST1 AS
SELECT PROVINCE,AGE_GROUP,
COUNT(Deactdt) AS TOTAL_DEACTIVATED_CUSTOMERS,
SUM(SALES) AS TOTAL_SALES
FROM TELECOM.ACCT_SEGMENT
WHERE DEACTDT IS NOT NULL
GROUP BY PROVINCE,AGE_GROUP
ORDER BY PROVINCE,AGE_GROUP;
QUIT;

PROC PRINT DATA = AGEPROVINDIST1;
WHERE PROVINCE IS NOT NULL;
RUN;
*VISUALIZATION;
Title"Age distributions of Deactive customers";

PROC SQL ;
CREATE TABLE AGEDIST1 AS
SELECT AGE_GROUP,
COUNT(DEACTDT) AS TOTAL_DEACTIVE_CUSTOMERS,
SUM(SALES) AS TOTAL_SALES
FROM TELECOM.ACCT_SEGMENT
WHERE DEACTDT IS NOT NULL
GROUP BY AGE_GROUP
ORDER BY AGE_GROUP;
QUIT;

PROC PRINT DATA = AGEDIST1;
RUN;

PROC GCHART DATA = TELECOM.ACCT_SEGMENT;
PIE3D AGE_GROUP/ PERCENT =INSIDE;
WHERE DEACTDT IS NOT NULL;
RUN;
=====
=====
Title" Distributions of Deactive customers by Province";
PROC SQL ;
CREATE TABLE PROVINDIST AS
SELECT PROVINCE,
COUNT(DEACTDT) AS TOTAL_DEACTIVE_CUSTOMERS,
SUM(SALES) AS TOTAL_SALES
FROM TELECOM.ACCT_SEGMENT
WHERE DEACTDT IS NOT NULL
GROUP BY PROVINCE
ORDER BY PROVINCE;

```

```

QUIT;

PROC PRINT DATA = PROVINDIST;
WHERE PROVINCE IS NOT NULL;
RUN;

Title"Distribution of DeActive Customers By Province ";
PROC GCHART DATA = TELECOM.ACCTOUNT_SEGMENT;
PIE3D PROVINCE/ PERCENT =INSIDE;
WHERE DEACTDT IS NOT NULL;
RUN;

*QUESTION- 3;
*Segment the customers based on age, province and sales amount:;

*Sales segment: < $100, $100---500, $500-$800, $800 and above.;

TITLE"SEGMENTATION BASED ON AGE , SALES and PROVINCE";
DATA TELECOM.SEGMENTS;
SET TELECOM.DATA;
DROP NEW_AGE;
LENGTH AGE_GROUP $25;
LENGTH SALES_GROUP $25;
LENGTH PROVINCE $25;

IF AGE <= 20 THEN AGE_GROUP = "20 OR LESS";
ELSE IF 21<= AGE<=40 THEN AGE_GROUP= "BETWEEN 21 & 40 ";
ELSE IF 41<=AGE<=59 THEN AGE_GROUP =" BETWEEN 41 - 60";
ELSE IF AGE >= 60 THEN AGE_GROUP = "60 & MORE";

*Sales segment: < $100, $100---500, $500-$800, $800 and above.;

IF SALES<100 THEN SALES_GROUP ="$100 & BELOW";
ELSE IF 100 <SALES<500 THEN SALES_GROUP ="$100 - $500";
ELSE IF 500 <SALES<800 THEN SALES_GROUP ="$500 - $800";
ELSE IF SALES >=800 THEN SALES_GROUP="800 & ABOVE";

*Province Segmentation;

IF PROVINCE = "AB" THEN PROVINCE = "ALBERTA";
ELSE IF PROVINCE ="BC" THEN PROVINCE ="BRITISH COLOMBIA";
ELSE IF PROVINCE ="NS" THEN PROVINCE = "NOVA SCOTIA";
ELSE IF PROVINCE ="ON" THEN PROVINCE = "ONTARIO";
ELSE IF PROVINCE ="QC" THEN PROVINCE ="QUEBEC";
RUN;

PROC PRINT DATA = TELECOM.SEGMENTS (OBS = 50);

RUN;

*Create Analysis Report Based on Segmentation;
TITLE "SALES BASED ON DIFFERENT AGE GROUPS";
PROC FREQ DATA = TELECOM.SEGMENTS;
TABLE AGE_GROUP*SALES_GROUP/CHISQ;
run;

```

```

PROC SGPLOT DATA =TELECOM SEGMENTS;
VBAR AGE_GROUP/ group = SALES_GROUP groupdisplay = CLUSTER;
YAXIS LABEL = "SALES";
RUN;

TITLE "SALES BASED ON DIFFERENT PROVINCE";

PROC FREQ DATA = TELECOM SEGMENTS;
TABLE PROVINCEE*SALES_GROUP/ chisq;
run;

PROC SGPLOT DATA =TELECOM SEGMENTS;
VBAR PROVINCEE/ group =SALES_GROUP groupdisplay = CLUSTER;
YAXIS LABEL = "SALES";
RUN;

TITLE "AGE DISTRIBUTION BASED ON DIFFERENT PROVINCE";

PROC FREQ DATA = TELECOM SEGMENTS;
TABLE PROVINCEE*AGE_GROUP/ chisq;
run;

PROC SGPLOT DATA =TELECOM SEGMENTS;
VBAR PROVINCEE/ group =AGE_GROUP groupdisplay = CLUSTER;
YAXIS LABEL = "SALES";
RUN;

PROC OPTIONS OPTION = MACRO;
RUN;

*METHOD USING MACRO;

%MACRO BI_ANALYSIS_CAT_CAT (DSN = ,VAR1= , VAR2= );
PROC FREQ DATA =&DSN;
TITLE " RELATION BETWEEN &VAR1. AND &VAR2." ;
TABLE &VAR1.*&VAR2/chisq;
PROC SGPLOT DATA = &DSN;
VBAR &VAR1/GROUP = &VAR2 GROUPDISPLAY = CLUSTER;
RUN;
%MEND BI_ANALYSIS_CAT_CAT;

%BI_ANALYSIS_CAT_CAT(DSN =TELECOM SEGMENTS,VAR1= AGE_GROUP, VAR2 =
SALES_GROUP);
%BI_ANALYSIS_CAT_CAT(DSN =TELECOM SEGMENTS ,VAR1 =SALES_GROUP , VAR2
=PROVINCE );
%BI_ANALYSIS_CAT_CAT(DSN =TELECOM SEGMENTS ,VAR1 AGE_GROUP , VAR2 =PROVINCE
);

*1.4.Statistical Analysis:;
*1) Calculate the tenure in days for each account and give its simple
statistics;

proc sort data = telecom.segments out=telecom.sort ;
by descending actdt ;
run;

```

```

proc print data = telecom.sort (obs=50);
run;

Title "Tenure in days for each account";
data telecom.tenure;
set telecom.segments;
*reference_date = "20JAN2001"d;
if deactdt = '' then tenure_days = intck('day',actdt,"20JAN2001"d);
else
tenure_days = intck('day',actdt,deactdt);
run;

proc print data = telecom.tenure(obs = 50);
run;

proc means data = telecom.tenure N NMISS MIN Q1 MEDIAN Q3 MAX qrange mean
std cv clm;
var tenure_days;
run;

%MACRO UNI_ANALYSIS_NUM(DATA,VAR);
TITLE "HORIZONTAL BOXPLOT FOR &VAR";
PROC SGPLOT DATA=&DATA;
HBOX &VAR;
STYLEATTRS
BACKCOLOR=DARKGREY
WALLCOLOR=LIGHTPINK
;
RUN;
TITLE "UNIVARIATE ANALYSIS FOR &VAR";
proc means data=&DATA N NMISS MIN Q1 MEDIAN MEAN Q3 MAX qrange cv clm
maxdec=2 ;
var &var;
run;
%MEND;

%UNI_ANALYSIS_NUM(telecom.tenure,TENURE_DAYS);

*) Calculate the number of accounts deactivated for each month;

data telecom.deact_yrmonth;
set telecom.segments;
deact_year = year(deactdt);
deact_month = intnx('month',deactdt,0,'b');
run;

proc print data = telecom.deact_yrmonth (obs=50);
format deact_month monname3.;
run;

*To find monthly deactivation in all the years from 1999 - 2001;
Title "Number of monthly deactivation from 1999 - 2001";
proc freq data = telecom.deact_yrmonth;
table deact_month/NOCUM;
format deact_month monname3.;
run;

```

```

*for year 1999;
DATA telecom.deact_1999;
  SET telecom.deact_yrmonth;
  KEEP acctno deactdt deact_year deact_month;
  where deact_year = 1999;
RUN;
proc print data = telecom.deact_1999 (obs=100);
format deact_month monname3.;
run;
proc sort data = telecom.deact_1999 out = telecom.deact1_1999;
by deact_month;
run;
proc print data = telecom.deact1_1999 (obs=100);
format deact_month monname3.;
run;
Title"Number Of Accounts Deactivated For Each Month In 1999";
proc sql;
create table telecom.deactNo_1999 as
select deact_month,
count(*) as Deact_Number
from telecom.deact1_1999
group by deact_month;
quit;

proc print data = telecom.deactNo_1999;
format deact_month monname3.;
run;

*for year 2000;

DATA telecom.deact_2000;
  SET telecom.deact_yrmonth;
  KEEP acctno deactdt deact_year deact_month;
  where deact_year = 2000;
RUN;
proc print data = telecom.deact_2000 (obs=100);
format deact_month monname3.;
run;
proc sort data = telecom.deact_2000 out = telecom.deact1_2000;
by deact_month;
run;
proc print data = telecom.deact1_2000 (obs=100);
format deact_month monname3.;
run;
Title"Number Of Accounts Deactivated For Each Month In 2000";
proc sql;
create table telecom.deactNo_2000 as
select deact_month,
count(*) as Deact_Number
from telecom.deact1_2000
group by deact_month;
quit;

proc print data = telecom.deactNo_2000;
format deact_month monname3.;

```

```

run;

*for year 2001;
DATA telecom.deact_2001;
  SET telecom.deact_yrmonth;
  KEEP acctno deactdt deact_year deact_month;
  where deact_year = 2001;
RUN;
proc print data = telecom.deact_2001 (obs=100);
format deact_month monname3.;
run;
proc sort data = telecom.deact_2001 out = telecom.deact1_2001;
by deact_month;
run;
proc print data = telecom.deact1_2001 (obs=100);
format deact_month monname3.;
run;
Title"Number Of Accounts Deactivated For Each Month In 2001";
proc sql;
create table telecom.deactNo_2001 as
select deact_month,
count(*) as Deact_Number
from telecom.deact1_2001
group by deact_month;
quit;

proc print data = telecom.deactNo_2001;
format deact_month monname3.;
run;
*visualization;
Title"Accounts Deactivated In 1999";
proc sgplot data=telecom.deactno_1999;
series x=deact_month y=deact_number;
xaxis label = "months in 1999";
yaxis label ="Account Deactivation";
format deact_month monname3.;
run;

Title"Accounts Deactivated In 2000";
proc sgplot data=telecom.deactno_2000;
series x=deact_month y=deact_number;
xaxis label = "months in 2000";
yaxis label ="Account Deactivation";
format deact_month monname3.;
run;

*3) Segment the account, first by account status "Active" and "Deactivated",
then by
Tenure: < 30 days, 31---60 days, 61 days--- one year, over one year. Report
the
number of accounts of percent of all for each segment.;

Title "Tenure in days for each account";
data telecom.tenure;
set telecom.segments;
*reference_date = "20JAN2001"d;

```

```

if deactdt = '' then tenure_days = intck('day',actdt,"20JAN2001'd);
else
tenure_days = intck('day',actdt,deactdt);
run;

proc print data = telecom.tenure(obs = 50);
run;

data telecom.Account_segment ;
set telecom.tenure;
length Account_Status $25;
length Tenure_Segment $30;
if deactdt = '' then Account_Status = "Active";
else Account_Status = "DeActivated";

if tenure_days <30 then Tenure_Segment ="Less than 30 days";
else if tenure_days <60 then Tenure_Segment = "Between 31 and 60 days";
else if tenure_days <365 then Tenure_Segment = "Between 60 days and 1 year";
else if tenure_days > 365 then Tenure_Segment = "Over 1 year";
run;

proc print data = telecom.Account_Segment(obs = 100);
run;

title"Account Status";
proc freq data = telecom.Account_Segment;
table Account_Status;
run;
title"Tenure Segmentation";
proc freq data = telecom.Account_Segment;
table Tenure_Segment;
run;

*4) Test the general association between the tenure segments and "Good Credit"
"RatePlan " and "DealerType.";
data telecom.credit ;
set telecom.Account_Segment;
if goodcredit = 1 then Credit_Type = "Good";
else Credit_Type = "Bad";
run;

proc print data = telecom.credit (obs = 100);
run;

PROC OPTIONS OPTION = MACRO;
RUN;

%MACRO BI_ANALYSIS_CAT_CAT (DSN = ,CLASS= , VAR= );
PROC FREQ DATA =&DSN;
TITLE " RELATION BETWEEN &VAR. AND &CLASS.";
TABLE &VAR.*&CLASS/chisq;
PROC SGPLOT DATA = &DSN;
VBAR &VAR/GROUP = &CLASS GROUPDISPLAY = STACK;
RUN;
%MEND BI_ANALYSIS_CAT_CAT;

```

```

%BI_ANALYSIS_CAT_CAT(DSN =telecom.credit ,CLASS = TENURE_SEGMENT, VAR =
CREDIT_TYPE);
%BI_ANALYSIS_CAT_CAT(DSN =telecom.credit ,CLASS = TENURE_SEGMENT, VAR =
RATEPLAN);
%BI_ANALYSIS_CAT_CAT(DSN =telecom.credit ,CLASS = TENURE_SEGMENT, VAR =
DEALERTYPE);

*5) Is there any association between the account status and the tenure
segments?
Could you find out a better tenure segmentation strategy that is more
associated
with the account status?;

%MACRO BI_ANALYSIS_CAT_CAT (DSN = ,CLASS= , VAR= );
PROC FREQ DATA =&DSN;
TITLE " RELATION BETWEEN &VAR. AND &CLASS.";
TABLE &VAR.*&CLASS/chisq;
PROC SGLOT DATA = &DSN;
VBAR &VAR/GROUP = &CLASS GROUPDISPLAY = CLUSTER;
RUN;
%MEND BI_ANALYSIS_CAT_CAT;

%BI_ANALYSIS_CAT_CAT(DSN =telecom.credit ,CLASS =ACCOUNT_STATUS , VAR =
TENURE_SEGMENT);

*better tenure segmentation strategy;
DATA telecom.TENURE_SEGNEW;
SET telecom.credit;
IF TENURE_DAYS < 183 THEN TENURE_SEGMENT ="LESS THAN 6 MONTHS";
ELSE IF TENURE_DAYS <365 then TENURE_SEGMENT ="BETWEEN 6 MONTHS AND 1 YEAR";
ELSE IF TENURE_DAYS <=731 then TENURE_SEGMENT ="BETWEEN 1 AND 2 YEARS";
RUN;

proc print data = telecom.TENURE_SEGNEW (obs=100);
run;

%BI_ANALYSIS_CAT_CAT(DSN =telecom.TENURE_SEGNEW ,CLASS =TENURE_SEGMENT , VAR
= ACCOUNT_STATUS);

*6 Does Sales amount differ among different account status, GoodCredit, and
customer age segments?;

*SALES VS ACCOUNT STATUS
=====
*summarization using proc univariate and test of normality;
proc univariate data=telecom.TENURE_SEGNEW normal;
class account_status;
var sales;
run;

%MACRO BI_ANALYSIS_NUMS_CAT (DSN = ,CLASS= , VAR=,VAR1= );
%LET N = %SYSFUNC(COUNTW(&VAR));
%DO I = 1 %TO &N;
%LET X = %SCAN(&VAR,&I);

```

```

PROC MEANS DATA = &DSN. N NMISS MIN Q1 MEDIAN MEAN Q3 MAX qrange cv clm
maxdec=2 ;
  TITLE " RELATION BETWEEN &X. AND &CLASS." ;
  CLASS &CLASS. ;
  VAR &X. ;
  OUTPUT OUT= OUT_&CLASS._&X. MIN = MEAN= STD = MAX = /AUTONAME ;
  RUN;
%END;
%MEND BI_ANALYSIS_NUMs_CAT;

%BI_ANALYSIS_NUMs_CAT (DSN = telecom.TENURE_SEGNEW ,CLASS=ACCOUNT_STATUS ,
VAR=sales,VAR1=PROVINCEE );
%BI_ANALYSIS_NUMs_CAT (DSN = telecom.TENURE_SEGNEW ,CLASS=CREDIT_TYPE ,
VAR=sales );
%BI_ANALYSIS_NUMs_CAT (DSN = telecom.TENURE_SEGNEW ,CLASS=AGE_GROUP,
VAR=sales );

*Visualization using histogram;
Title "Distribution of Sales with Account Status ";
proc sgplot data=telecom.TENURE_SEGNEW;
  histogram sales / group=account_status transparency=0.5 scale=count;
  keylegend / location=inside position=topright across=1;
run;
title;
*for test of equality of variance;
proc glm data=telecom.TENURE_SEGNEW;
class account_status;
model sales = account_status;
means account_status / hovtest=levene(type=abs);
run;
*T test for test of independency;
%MACRO BI_ANALYSIS_NUMs_CAT_TTEST (DSN = ,CLASS= , VAR= );
TITLE "TTEST for &VAR. grouped by &CLASS. in &DSN.";
proc ttest data=&DSN.;
var &VAR. ;
class &CLASS. ;
run;
QUIT;
%MEND BI_ANALYSIS_NUMs_CAT_TTEST;

%BI_ANALYSIS_NUMs_CAT_TTEST (DSN =telecom.TENURE_SEGNEW ,CLASS=account_status
, VAR=sales);

*SALES VS GOOD CREDIT

*summarization using proc univariate and test of normality;
proc univariate data=telecom.TENURE_SEGNEW normal;
class credit_type;
var sales;
run;

%BI_ANALYSIS_NUMs_CAT (DSN = telecom.TENURE_SEGNEW ,CLASS=CREDIT_TYPE ,
VAR=sales );

*Visualization using histogram;
Title "Distribution of Sales with credit type ";

```

```

proc sgplot data=telecom.TENURE_SEGNEW;
  histogram sales / group=credit_type transparency=0.5 scale=count;
  keylegend / location=inside position=topright across=1;
run;
title;
*for test of equality of variance;
proc glm data=telecom.TENURE_SEGNEW;
class credit_type;
model sales = credit_type;
means credit_type / hovtest=levene(type=abs);
run;

%BI_ANALYSIS_NUMS_CAT_TTEST (DSN =telecom.TENURE_SEGNEW ,CLASS=credit_type ,
VAR=sales);

*Does Sales amount differ among age group?;

proc univariate data=telecom.TENURE_SEGNEW normal;
class Age_group;
var sales;
run;

%BI_ANALYSIS_NUMS_CAT (DSN = telecom.TENURE_SEGNEW ,CLASS=CREDIT_TYPE ,
VAR=sales );

*Visualization using histogram;
Title "Distribution of Sales with Age group ";
proc sgplot data=telecom.TENURE_SEGNEW;
  histogram sales / group=age_group transparency=0.5 scale=count;
  keylegend / location=inside position=topright across=1;
run;
title;

*for test of equality of variance;
proc glm data=telecom.TENURE_SEGNEW;
class age_group;
model sales = age_group;
means age_group / hovtest=levene(type=abs);
run;

*Test of independancy using one way anova ;
PROC ANOVA DATA = TELECOM.TENURE_SEGNEW;
CLASS AGE_GROUP;
MODEL SALES = AGE_GROUP;
RUN;
*=====
=====;

```