

Mini Project - Advanced SAS

TELECOM COMPANY CUSTOMER DISTRIBUTION & DEACTIVATION ANALYSES

INSTRUCTOR – HAMID RAJAEE

BY : VANDANA MURALIDHARAN

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

Importing Data Set Into SAS

```
Title " Importing Data Set ";

data telecom.Details;
infile "C:\Users\vanda\OneDrive\Desktop\Telecom
company\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;
```

	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	117695400288	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	117700067271	23DEC1999			0 1	B1		75	ON	\$134.00
10	1177010940613	09DEC1999			1 2	A1		42	NS	\$11.00
11	1177025997013	09NOV1999			1 1	A1		26	BC	\$154.00
12	1177027515760	19OCT1999			1 1	B1		73	BC	\$16.00
13	1177028996676	21SEP2000			0 1	C1		QC		\$179.00
14	1177038747105	14MAR2000			0 1	C1		41	ON	\$705.00
15	1177045857516	22JUN2000			1 1	A1		53	QC	\$83.00
16	1177057406016	21SEP2000			0 1	C1		50	ON	\$529.00
17	1177066422248	26APR1999	15JAN2001	NEED	0 1	A2		55	NS	\$44.00
18	1177089399155	17AUG2000			1 3	A1		56	BC	\$548.00
19	117713886410	13SEP2000	08JAN2001	COMP	0 1	C1		45	ON	\$63.00
20	1177128264924	10DEC1999			1 1	B1		38	ON	\$178.00
21	1177135400264	09DEC2000			0 1	A1		42	ON	\$345.00
22	1177169300842	24MAR2000			0 3	B1		54	ON	\$331.00
23	1177177440963	15FEB2000			0 1	B1		23	BC	\$97.00
24	1177177784000	30JUL2000			1 1	B1		41	ON	\$42.00
25	1177184144512	11OCT1999			1 1	A1		57	ON	
26	1177196331390	20DEC1999			0 3	A1		26	ON	\$773.00
27	1177200000723	12AUG2000	26OCT2000	DEBT	0 2	C1		NS		\$71.00

NOTE: 102255 records were read from the infile "C:\Users\vanda\OneDrive\Desktop\Telecom company\New_Wireless_Fixed.txt".

The minimum record length was 92.

The maximum record length was 92.

NOTE: The data set TELECOM.DETAILS has 102255 observations and 10 variables.

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 the Descriptive portion

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

The CONTENTS Procedure

Data Set Name	TELECOM.DETAILS	Observations	102255
Member Type	DATA	Variables	10
Engine	V9	Indexes	0
Created	03/20/2022 12:56:55	Observation Length	64
Last Modified	03/20/2022 12:56:55	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

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

There are 102255 observations and 10 variables in the given dataset

FINDING HEAD OF DATA FOR 10 OBSERVATION

```
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

```
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;
```

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

Number of missing values

Province	deactreason	rateplan	dealertype
5907	83162	0	0

Number of missing values

The MEANS Procedure

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

Finding Number Of Unique levels in each Variables

Number Of Unique/Distinct Values In All Variables

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

Number of unique distinct values in each variables			
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	

GoodCredit	Frequency
0	31253
1	71002

DealerType	Frequency
A1	56132
A2	11255
B1	20670
C1	14198

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

DESCRIPTIVE ANALYSIS OF CONTINOUS VARIABLES

```
*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;
```

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	1.0000000	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 Activated_Accounts,
       COUNT(Deactdt) AS Deactivated_Accounts
FROM telecom.Data;
QUIT;
```

Number of Activated and Deactivated Accounts

Total_Accounts	Activated_Accounts	Deactivated_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 ?

Age distributions of active customers

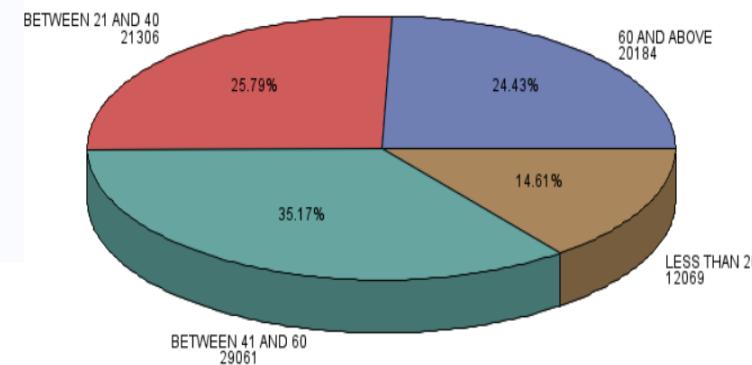
FREQUENCY of Age_Group

```
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;
```

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



```
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;
```

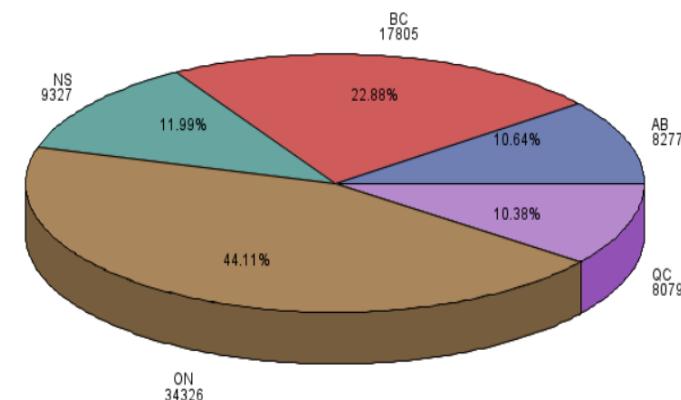
Distribution of Active Customers By Age and Province

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

Distribution of Active Customers By Province

Obs	Province	TOTAL_ACTIVE_CUSTOMERS	TOTAL_SALES
2	AB	8277	1370399
3	BC	17805	2970948
4	NS	9327	1522474
5	ON	34326	5715280
6	QC	8079	1345707

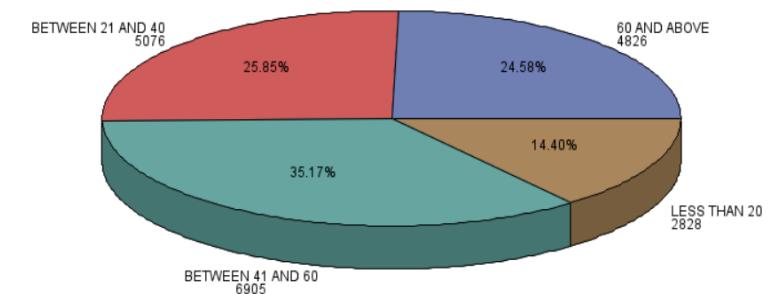
Distribution of Active Customers By Province



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

Age distributions of Deactive customers

FREQUENCY of Age_Group



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

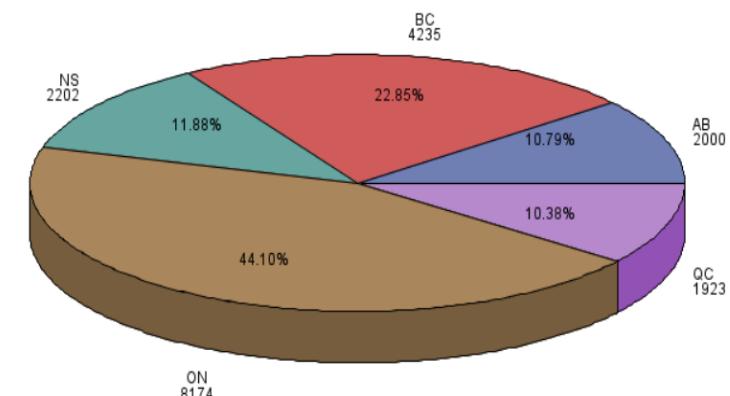
Distribution of DeActive Customers By Age and Province

Obs	Province	Age_Group	TOTAL_DEACTIVATED_CUSTOMERS	TOTAL_SALES
5	AB	60 AND ABOVE	485	74532
6	AB	BETWEEN 21 AND 40	493	76407
7	AB	BETWEEN 41 AND 60	727	120005
8	AB	LESS THAN 20	295	50783
9	BC	60 AND ABOVE	1095	177446
10	BC	BETWEEN 21 AND 40	1072	193347
11	BC	BETWEEN 41 AND 60	1457	229478
12	BC	LESS THAN 20	611	98336
13	NS	60 AND ABOVE	545	85430
14	NS	BETWEEN 21 AND 40	584	99553
15	NS	BETWEEN 41 AND 60	772	146156
16	NS	LESS THAN 20	301	47350
17	ON	60 AND ABOVE	1987	320500
18	ON	BETWEEN 21 AND 40	2144	364614
19	ON	BETWEEN 41 AND 60	2860	459417
20	ON	LESS THAN 20	1183	198592
21	QC	60 AND ABOVE	458	68739
22	QC	BETWEEN 21 AND 40	492	82273
23	QC	BETWEEN 41 AND 60	693	111439
24	QC	LESS THAN 20	280	48818

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



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	117700067271	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

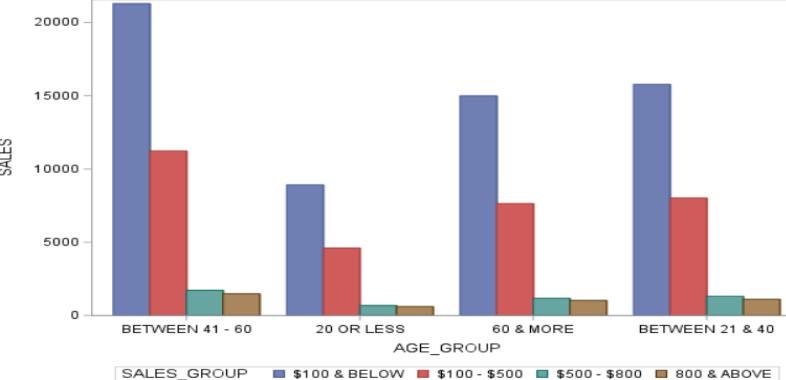
SALES BASED ON AGE GROUP

SALES BASED ON DIFFERENT AGE GROUPS						
The FREQ Procedure						
Table of AGE_GROUP by SALES_GROUP						
AGE_GROUP	SALES_GROUP					
	\$100 & BELOW	\$100 - \$500	\$500 - \$800	800 & ABOVE	Total	
BETWEEN 41 - 60	21202 20.95 59.56 34.92	11241 11.06 31.45 35.67	1728 1.70 4.83 35.19	1486 1.46 4.14 35.04	35747 35.17	
20 OR LESS	8916 8.77 60.15 14.62	4604 4.53 31.06 14.61	689 0.68 4.65 14.03	614 0.60 4.14 14.48	14823 14.58	
60 & MORE	14994 14.75 60.33 24.59	7645 7.52 30.76 24.26	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

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	1.6219	0.2028
Phi Coefficient		0.0095	
Contingency Coefficient		0.0095	
Cramer's V		0.0055	

SALES BASED ON DIFFERENT AGE GROUPS



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

Here from chi-square test we can see the p value >0.05 in all three cases As p value is more than 5% ,we fail to reject null hypothesis and can say that there is no statistical association between the variables.

SALES BASED ON PROVINCE

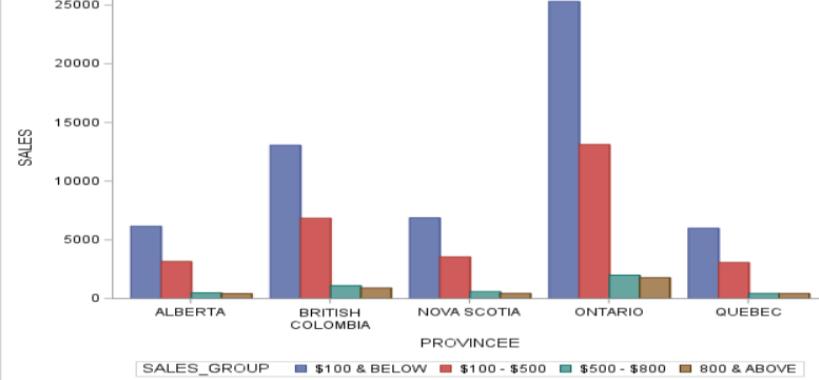
The FREQ Procedure						
Table of PROVINCE by SALES_GROUP						
PROVINCE	SALES_GROUP					
	\$100 & BELOW	\$100 - \$500	\$500 - \$800	800 & ABOVE	Total	
ALBERTA	6161 6.43 60.25 10.73	3158 3.30 30.88 10.61	483 0.50 4.72 10.47	424 0.44 4.15 10.65	10226 10.68	
BRITISH COLOMBIA	13061 13.64 59.06 22.75	6850 7.15 31.08 23.01	1101 1.15 5.02 23.88	901 0.65 4.11 22.63	21913 22.88	
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 11.97	
ONTARIO	25321 26.44 59.96 44.10	13127 13.71 31.08 44.10	1999 2.09 4.73 43.35	1783 1.88 4.22 44.78	42230 44.09	
QUEBEC	5998 6.26 60.32 10.45	3074 3.21 30.91 10.33	439 0.65 4.41 9.52	433 0.55 4.35 10.87	9944 10.38	
Total	57423 59.95	29765 31.08	4611 4.81	3982 4.16	95781 100.00	

Frequency Missing = 6474

Statistics for Table of PROVINCE by SALES_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	0.0671	0.7956
Phi Coefficient		0.0121	
Contingency Coefficient		0.0121	
Cramer's V		0.0070	

SALES BASED ON DIFFERENT PROVINCE



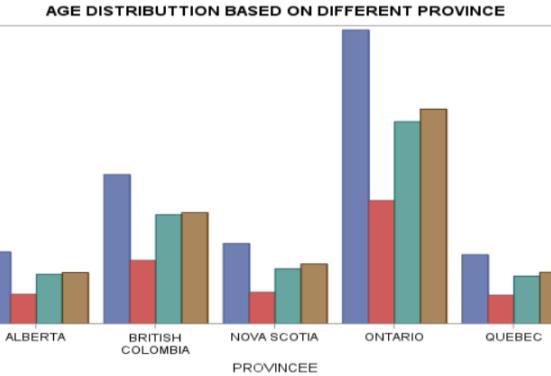
Irrespective of location/province the sales is highest for the category \$100 and below

AGE DISTRIBUTION IN PROVINCES

Table of PROVINCE by AGE_GROUP						
PROVINCE	AGE_GROUP					
	BETWEEN 41 - 60	20 OR LESS	60 & MORE	BETWEEN 21 & 40	Total	
ALBERTA	3658 3.80 35.59 10.81	1500 1.56 14.60 10.67	2517 2.61 24.49 10.67	2602 2.70 25.32 10.47	10277 10.67	
BRITISH COLOMBIA	7559 7.69 34.48 22.45	3227 3.35 14.64 22.96	5555 5.77 25.20 23.55	22040 5.87 25.68 22.76	22040 11.97	
NOVA SCOTIA	4089 4.24 35.47 12.08	1601 1.66 13.89 11.39	2797 2.90 24.26 11.86	3042 3.16 26.39 12.24	11529 11.97	
ONTARIO	14963 15.65 35.25 44.27	6276 6.51 14.77 44.65	10295 10.69 24.23 43.66	10545 11.36 25.75 44.01	42500 44.11	
QUEBEC	3515 3.65 35.14 10.39	1452 1.51 14.52 10.33	2419 2.51 24.19 10.26	2616 2.72 26.15 10.52	10002 10.38	
Total	33844 35.13	14056 14.59	23586 24.46	24862 25.80	96348 100.00	

Frequency Missing = 5907

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	

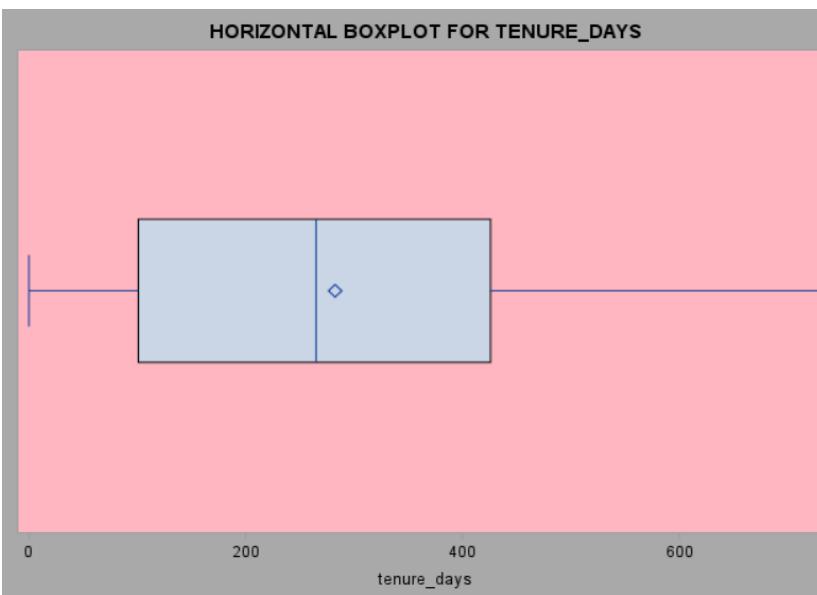


In each province maximum number of customers belong to the age group between 41 and 60

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



Tenure in days for each account														tenure_days
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	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	1176954000288	30MAY2000			1 2	A1		47	ON	\$83.00	BETWEEN 41 - 60	\$100 & BELOW	ONTARIO	235
5	1176969186303	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	117700067271	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		26	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			QC	\$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

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% CL for Mean	Upper 95% CL for Mean
102255	0	0	101.0000000	265.0000000	426.0000000	731.0000000	325.0000000	282.5717960	197.3237106	69.8313538	281.3623396	283.7812524

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

Total accounts deactivated each month from 1999 to 2021

Number of monthly deactivation from 1999 - 2001

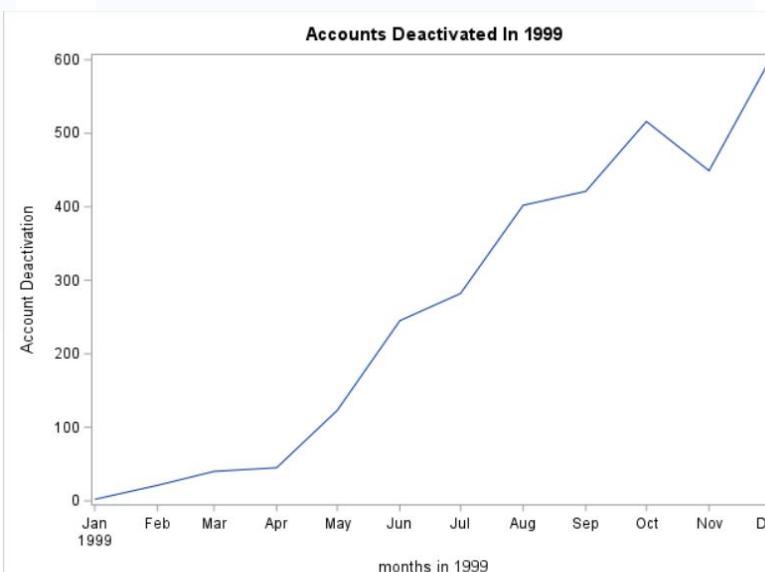
The FREQ Procedure

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		

YEAR - 1999

Number Of Accounts Deactivated For Each Month In 1999

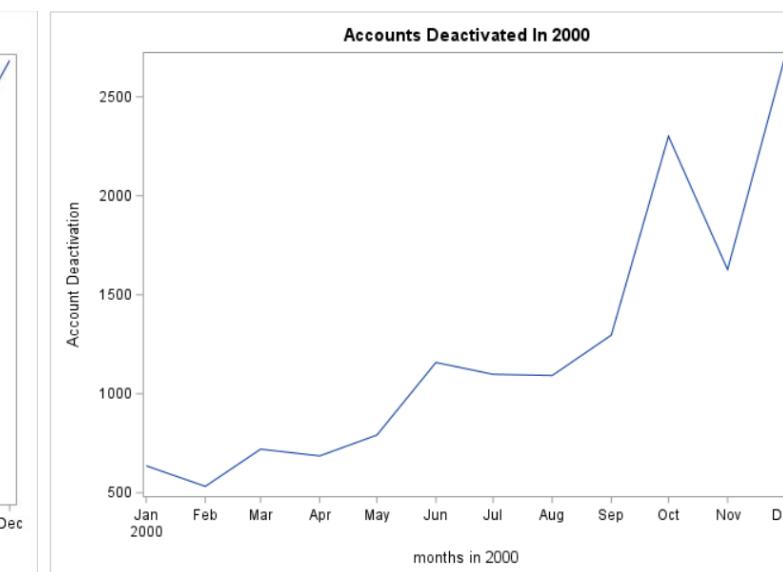
Obs	deact_month	Deact_Number
1	Jan	2
2	Feb	21
3	Mar	40
4	Apr	45
5	May	123
6	Jun	245
7	Jul	282
8	Aug	402
9	Sep	421
10	Oct	516
11	Nov	449
12	Dec	599



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.

```

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;

```

Account Status

The FREQ Procedure

Account_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Active	82620	80.80	82620	80.80
DeActivated	19635	19.20	102255	100.00

Tenure Segmentation

The FREQ Procedure

Tenure_Segment	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Between 31 and 60 days	8586	8.41	8586	8.41
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				

Obs	acctno	actdt	Deactdt	Deactreason	GoodCredit	Rateplan	DealerType	Age	Province	Sales	AGE_GROUP	SALES_GROUP	PROVINCEE	tenure_days	Account_Status	Tenure_Segment	Credit_Type
1	1176913194483	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	1176951913656	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 days and 1 year	Good
5	1176989186303	13DEC2000			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	1177000067271	23DEC1999			0	1	B1	75	ON	\$134.00	60 & MORE	\$100 - \$500	ONTARIO	394	Active	Over 1 year	Bad
10	1177010940613	09DEC1999			1	2	A1	42	NS	\$11.00	BETWEEN 41 - 60	\$100 & BELOW	NOVA SCOTIA	403	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

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

Segmentation of goodcredit

```
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;

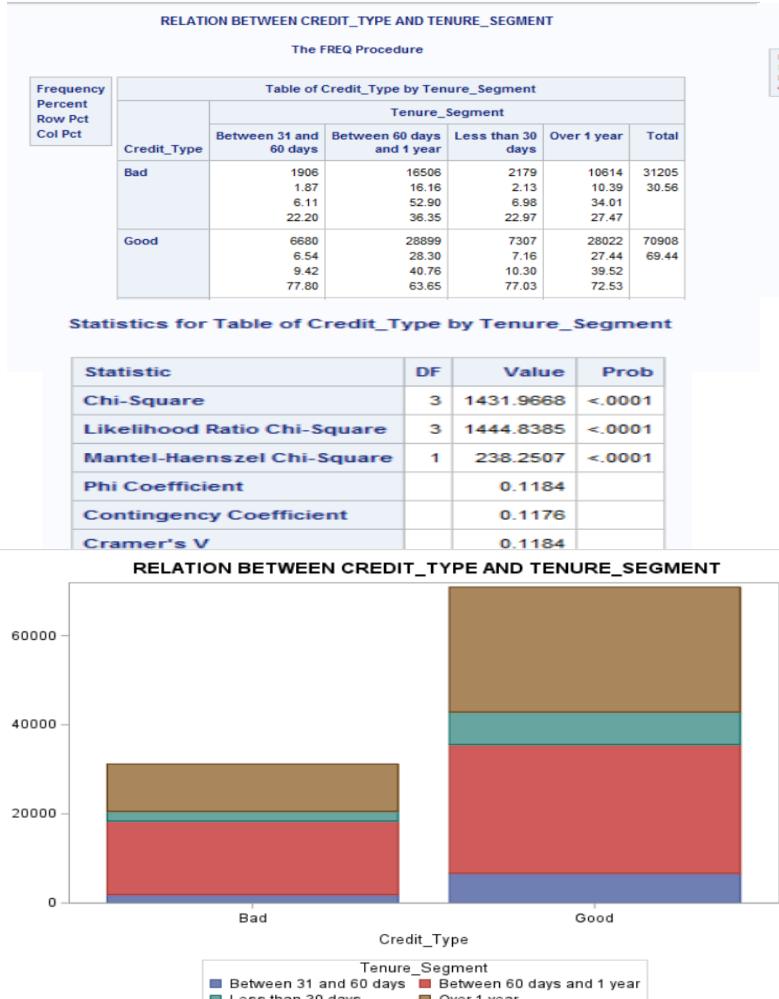
%MACRO BI_ANALYSIS_CAT_CAT (DSN = ,CLASS= ,
VAR= );
PROC FREQ DATA =&DSN;
TITLE " RELATION BETWEEN &VAR. AND &CLASS .";
TABLE &VAR.*&CLASS/chisq;
PROC SGPLOT DATA = &DSN;
VBAR &VAR/GROUP = &CLASS GROUPDISPLAY =
STACK;
RUN;
%MEND BI_ANALYSIS_CAT_CAT;

%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);
```

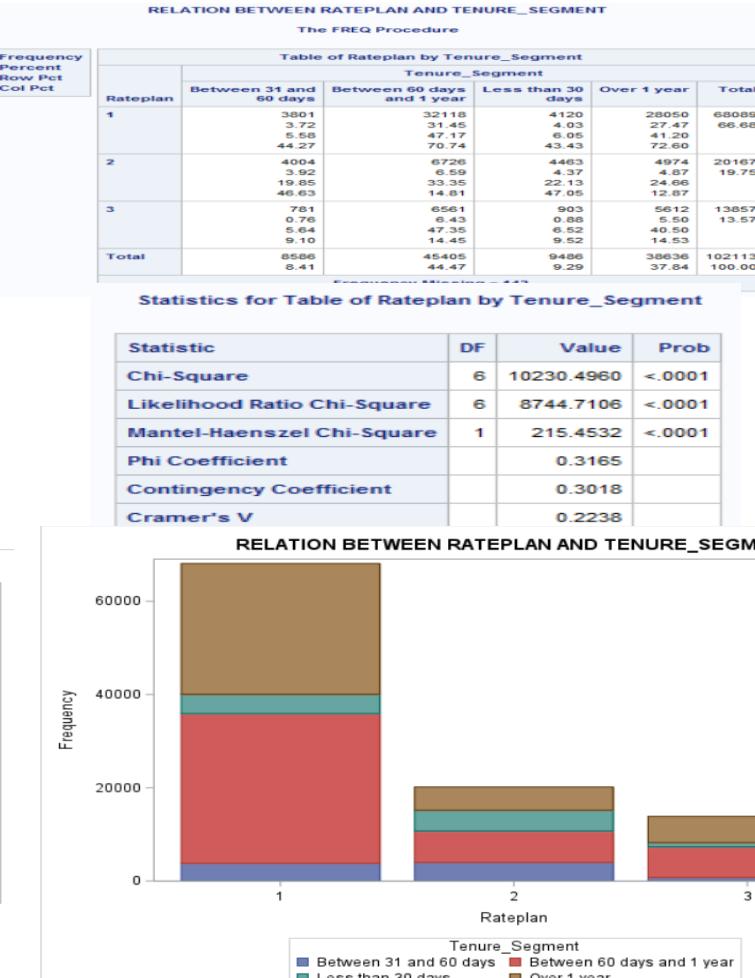
Obs	acctno	actdt	Deactdt	Deactreason	GoodCredit	Rateplan	DealerType	Age	Province	Sales	AGE_GROUP	SALES_GROUP	PROVINCEE	tenure_days	Account_Status	Tenure_Segment	Credit_Type	
1	1176913194483	20JUN1999	.	.	0	1	A1	58	BC	\$128.00	BETWEEN 41 - 60	\$100 - \$500	BRITISH COLOMBIA	580	Active	Over 1 year	Bad	
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	1176951913656	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	1176954000288	30MAY2000	.	.		1	2	A1	47	ON	\$83.00	BETWEEN 41 - 60	\$100 & BELOW	ONTARIO	235	Active	Between 60 days and 1 year	Good
5	1176969186303	13DEC2000	.	.		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	1177000067271	23DEC1999	.	.		0	1	B1	75	ON	\$134.00	60 & MORE	\$100 - \$500	ONTARIO	394	Active	Over 1 year	Bad
10	1177010940613	09DEC1999	.	.		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

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

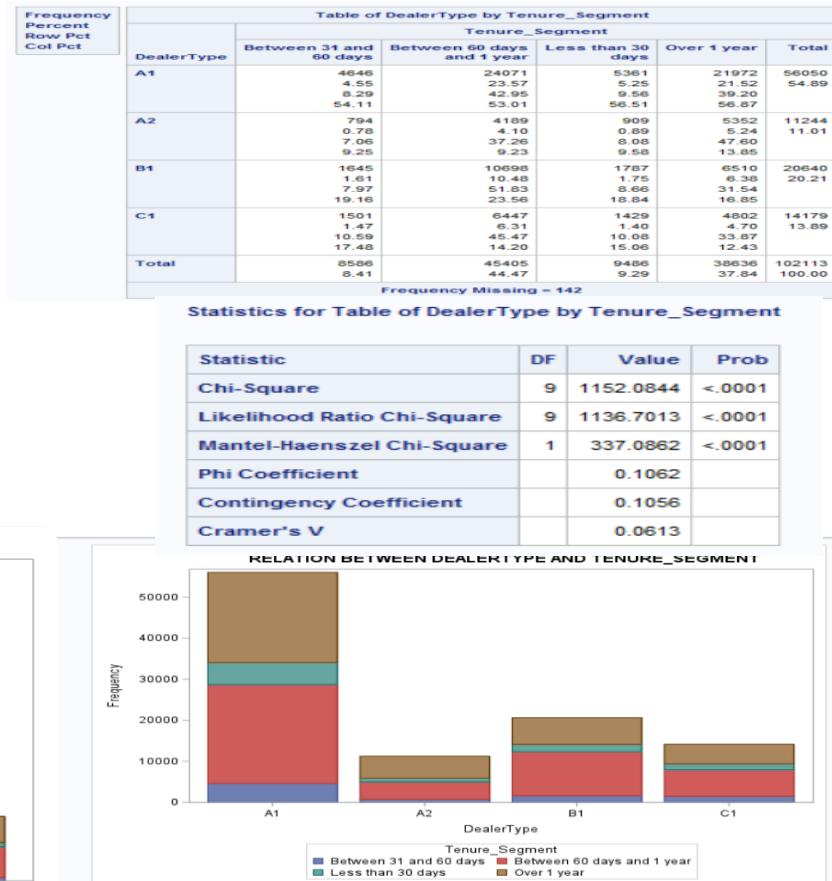
Tenure Segment vs Credit type



Tenure Segment vs Rate Plan



Tenure Segment vs Dealer 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.

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.

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?

```

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";

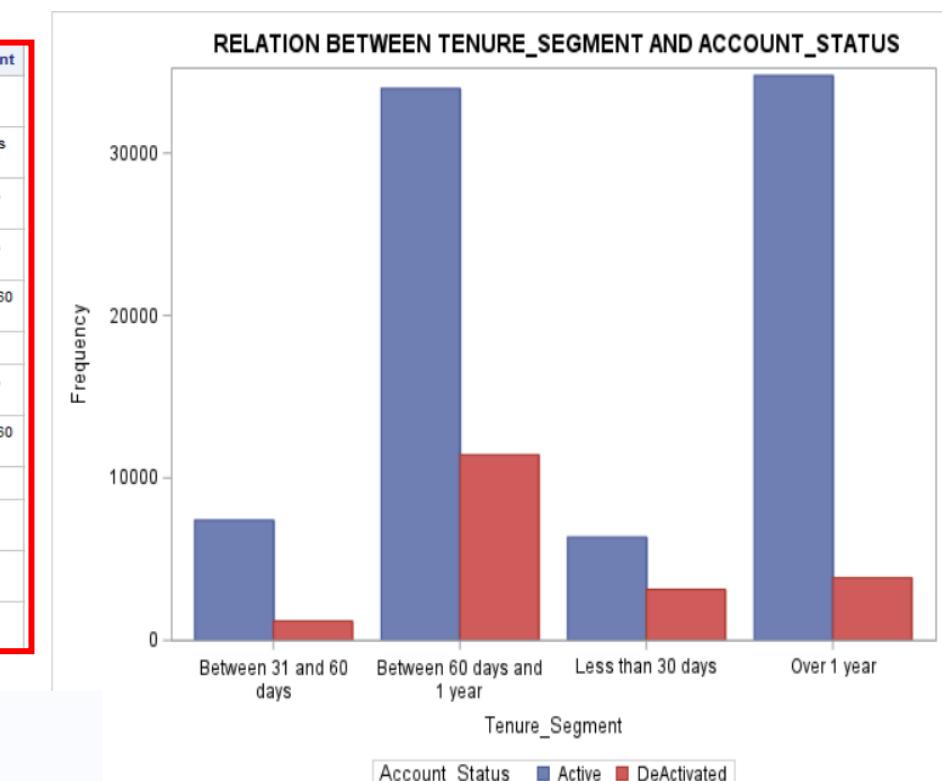
%MACRO BI_ANALYSIS_CAT_CAT (DSN = ,CLASS= ,
VAR= );
PROC FREQ DATA =&DSN;
TITLE " RELATION BETWEEN &VAR. AND &CLASS .";
TABLE &VAR.*&CLASS/chisq;
PROC SGPlot DATA = &DSN;
VBAR &VAR/GROUP = &CLASS GROUPDISPLAY =
CLUSTER;
RUN;
%MEND BI_ANALYSIS_CAT_CAT;

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

```

Sales	AGE_GROUP	SALES_GROUP	PROVINCE	tenure_days	Account_Status	Tenure_Segment
\$128.00	BETWEEN 41 - 60	\$100 - \$500	BRITISH COLOMBIA	580	Active	Over 1 year
\$72.00	BETWEEN 41 - 60	\$100 & BELOW	ALBERTA	11	DeActivated	Less than 30 days
\$593.00	BETWEEN 41 - 60	\$500 - \$800	BRITISH COLOMBIA	203	Active	Between 60 days and 1 year
\$83.00	BETWEEN 41 - 60	\$100 & BELOW	ONTARIO	235	Active	Between 60 days and 1 year
60 & MORE	\$100 & BELOW	BRITISH COLOMBIA		38	Active	Between 31 and 60 days
\$1,041.00	60 & MORE	800 & ABOVE	QUEBEC	384	DeActivated	Over 1 year
60 & MORE	\$100 & BELOW	ONTARIO		241	Active	Between 60 days and 1 year
\$72.00	60 & MORE	\$100 & BELOW	ALBERTA	53	Active	Between 31 and 60 days
\$134.00	60 & MORE	\$100 - \$500	ONTARIO	394	Active	Over 1 year
\$11.00	BETWEEN 41 - 60	\$100 & BELOW	NOVA SCOTIA	408	Active	Over 1 year
\$154.00	BETWEEN 21 & 40	\$100 - \$500	BRITISH COLOMBIA	438	Active	Over 1 year
\$16.00	60 & MORE	\$100 & BELOW	BRITISH COLOMBIA	459	Active	Over 1 year

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		Account_Status			Total
	Between 31 and 60 days		Active	7398 7.24 86.16 8.97	DeActivated	1188 1.16 13.84 6.06
	Between 60 days and 1 year		Active	33977 33.27 74.83 41.18	DeActivated	11428 11.19 25.17 58.28
	Less than 30 days		Active	6353 6.22 66.97 7.70	DeActivated	3133 3.07 33.03 15.98
	Over 1 year		Active	34777 34.06 90.01 42.15	DeActivated	3859 3.78 9.99 19.68
	Total		Active	82505 80.80	DeActivated	19608 19.20
	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	

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

```

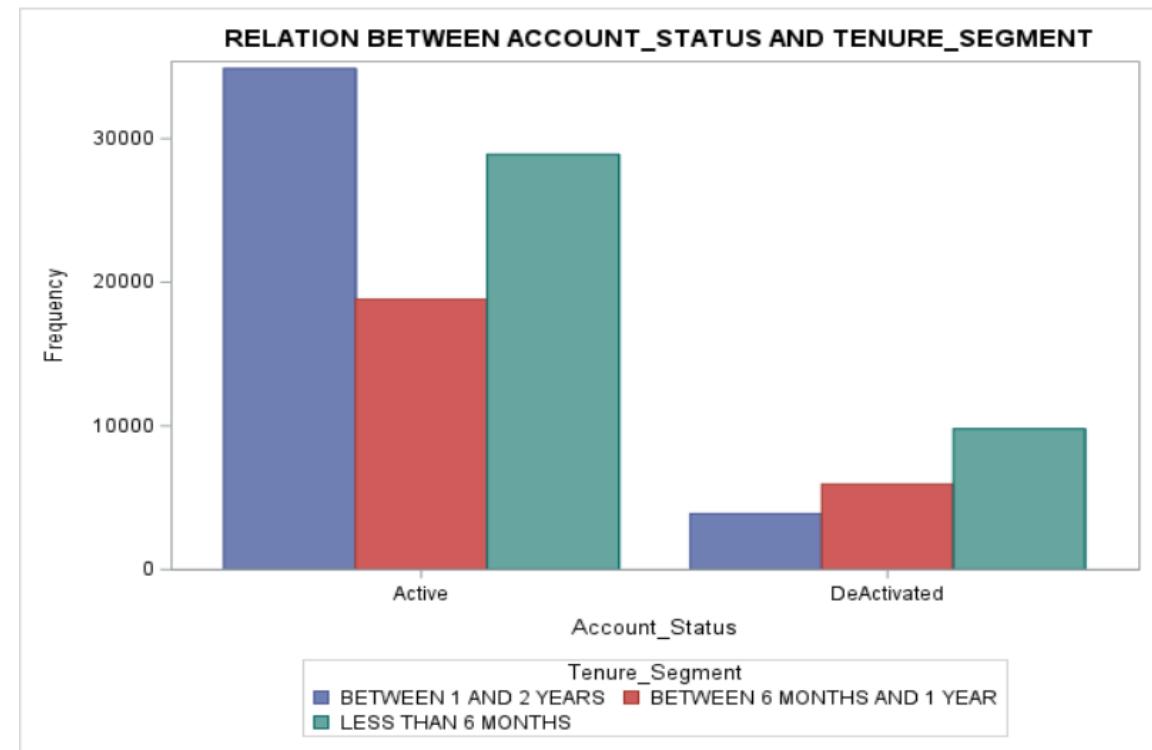
DATA telecom.TENURE_SEGNEW;
SET telecom.credit;
IF TENURE_DAYS < 183 THEN TENURE_SEGMENT ="LESS THAN 6 MONTHS";
ELSE IF TENURE_DAYS <365 then TENURE_SEGMENT ="BETWEEN 6 MONTHS AND 1 YEAR";
ELSE IF TENURE_DAYS <=731 then TENURE_SEGMENT ="BETWEEN 1 AND 2 YEARS";
RUN;

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

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

```

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



RELATION BETWEEN ACCOUNT_STATUS AND TENURE_SEGMENT

The FREQ Procedure

Account_Status	Table of Account_Status by Tenure_Segment			
	Tenure_Segment		Total	
	BETWEEN 1 AND 2 YEARS	BETWEEN 6 MONTHS AND 1 YEAR	LESS THAN 6 MONTHS	Total
Active	34892	18820	28908	82620
	34.12	18.40	28.27	80.80
	42.23	22.78	34.99	
	89.98	75.98	74.68	
DeActivated	3886	5950	9799	19635
	3.80	5.82	9.58	
	19.79	30.30	49.91	
	10.02	24.02	25.32	
Total	38778	24770	38707	102255
	37.92	24.22	37.85	100.00

Statistics for Table of Account_Status by Tenure_Segment

Statistic	DF	Value	Prob
Chi-Square	2	3410.0132	<.0001
Likelihood Ratio Chi-Square	2	3673.9521	<.0001
Mantel-Haenszel Chi-Square	1	2921.7801	<.0001
Phi Coefficient		0.1826	
Contingency Coefficient		0.1796	
Cramer's V		0.1826	

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	965517735	
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	80671833	Pr >= S 	<.0001
Account_Status N Obs Sum				
Active	82620	13739549.00		
DeActivated	19635	3234154.00		

Test of Normality

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

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 .

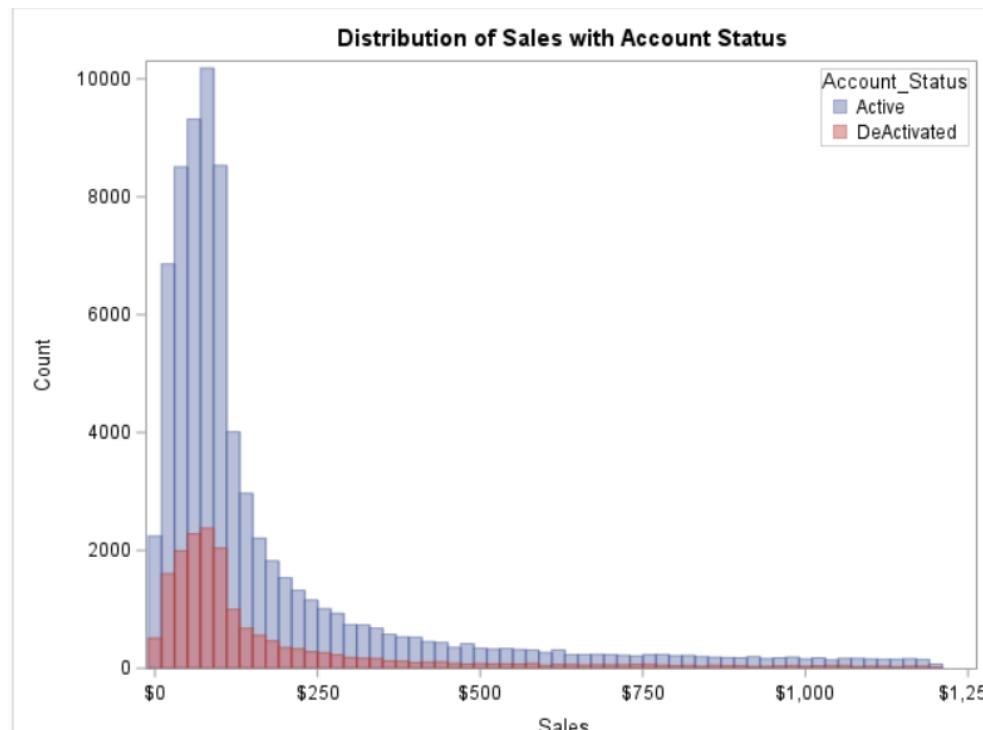
From Levene's test result we find that p value is >0.0505. As the P value is greater than 5%, we accept null hypothesis and conclude variances are equal.

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	F Value	Pr > F
Account_Status	1	112914	112914	3.83	0.0505
Error	93648	2.7641E9	29516.2		

Test of Independency – T test

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	0.3983			
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's t 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			

Since 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

Distribution of Sales with Age group

The ANOVA Procedure

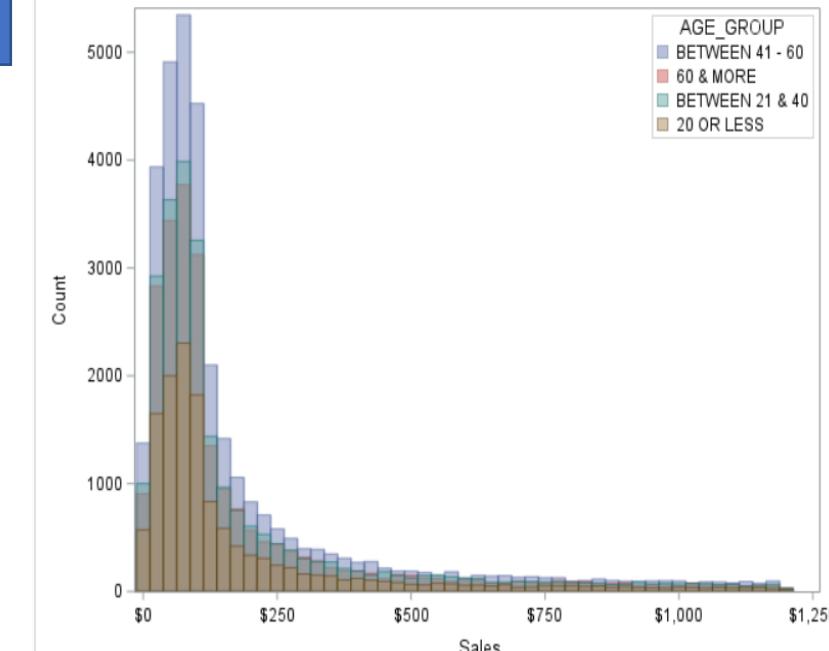
Dependent Variable: Sales

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

Distribution of Sales with Age group



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;

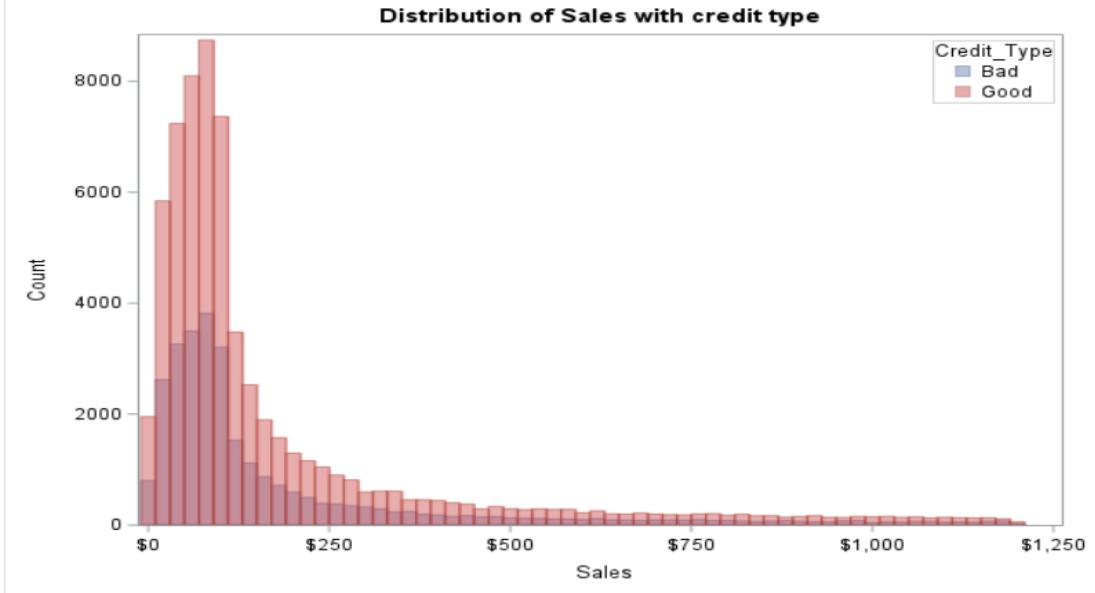
%BI_ANALYSIS_NUMs_CAT (DSN = telecom.TENURE_SEGNEW
,CLASS=CREDIT_TYPE , VAR=sales );
```

Summarization

The SAS System			
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

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

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

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.

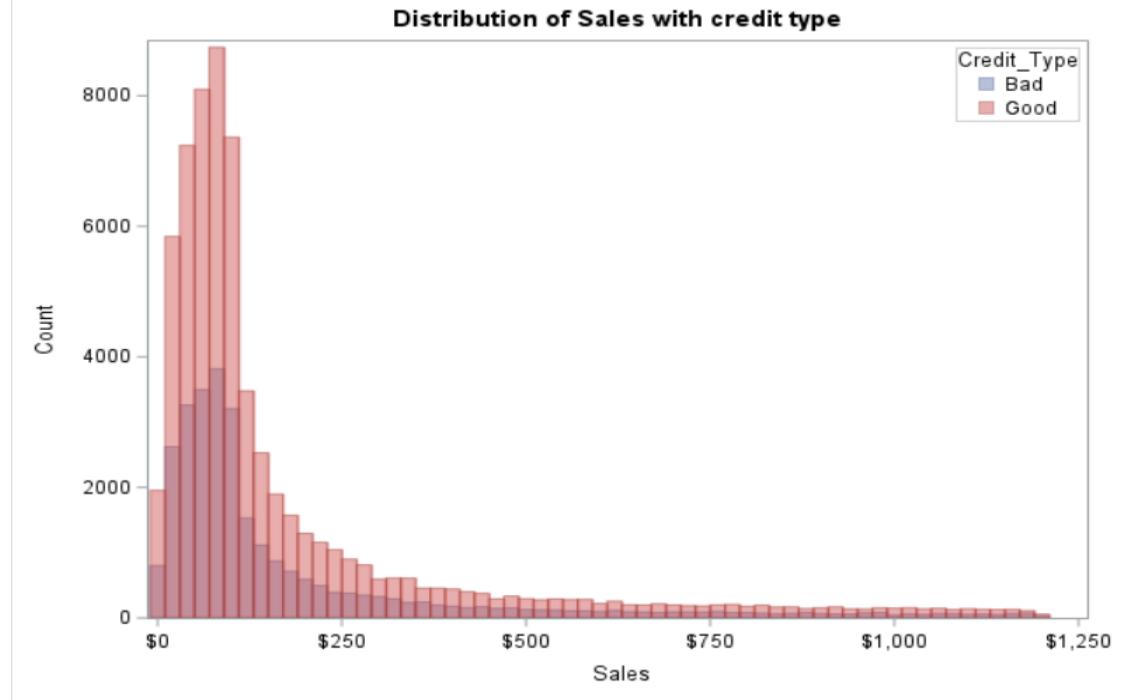
The GLM Procedure

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

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.



The TTEST Procedure

Variable: Sales

Credit_Type	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Bad		28599	181.6	234.8	1.3886	0	1200.0
Good		65051	181.1	233.6	0.9159	0	1200.0
Diff (1-2)	Pooled		0.4662	234.0	1.6600		
Diff (1-2)	Satterthwaite		0.4662		1.6635		

Credit_Type	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
Bad		181.6	178.8 184.3	234.8	232.9 236.8
Good		181.1	179.3 182.9	233.6	232.3 234.9
Diff (1-2)	Pooled	0.4662	-2.7874 3.7198	234.0	232.9 235.0
Diff (1-2)	Satterthwaite	0.4662	-2.7942 3.7266		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	93648	0.28	0.7788
Satterthwaite	Unequal	54366	0.28	0.7793

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	28598	65050	1.01	0.2878

Conclusion

- 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 Colombia. 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 de-activated customers is lowest for the tenure between 1 and 2 years and highest for tenure segment less than 6 months.

Recommendation

- In order to improve the business and minimize customer attrition, it is important to reduce the number of de-activated accounts which are under 6 months in tenure. This can be done by offering promotions and discounts for customers who are less than 6 months in tenure. Also, it is important to retain the active customers who are between 1- and 2-years tenure, by providing continued excellent customer service for them.