Telecom Company

Customer Distribution and Deactivation Analyses

Project By-Nandini Desai
Supervised By-Mr. Hamid Rajaee

Agenda

- Objective
- Background
- Appendix
  - Data Sources
  - Data Methodology
  - Data Model Assumptions
- Key Findings
- Capture Audience Attention
- Recommendations

Objective

- Analyzing the CRM data of a wireless company for 2 years to investigate the customer distribution and business behaviors.
- Gain insightful understanding about the customers, and to forecast the deactivation trends for the next 6 months.

Background

The attached CRM data encompasses two years' worth of customer interactions and transactions for a wireless company. In an increasingly competitive telecommunications landscape, understanding customer distribution and business behaviors is critical for enhancing customer retention and optimizing service offerings. This analysis aims to provide a comprehensive overview of customer demographics and usage patterns, enabling the company to identify key insights into customer preferences and needs. Furthermore, by examining historical data, we seek to forecast deactivation trends over the next six months. This proactive approach will empower the company to implement targeted strategies that enhance customer satisfaction and reduce churn, ultimately driving business growth and profitability.

Data Sources

Obs	Acctno	Actdt	Deactdt	DeactReason	GoodCredit	RatePlan	DealerType	Age	Province	Sales
1	1176913194483	06/20/1999	.		0	1	A1	58	BC	$128.00
2	1176914599423	10/04/1999	10/15/1999	NEED	1	1	A1	45	AB	$72.00
3	1176951913656	07/01/2000	.		0	1	A1	57	BC	$593.00
4	1176954000288	05/30/2000	.		1	2	A1	47	ON	$83.00
5	1176969186303	12/13/2000	.		1	1	C1	82	BC	.
6	1176991056273	08/31/1999	09/18/2000	MOVE	1	1	C1	92	QC	$1,041.00
7	1176991866552	05/24/2000	.		1	1	A1	77	ON	.
8	1176992889500	11/28/2000	.		1	1	C1	68	AB	$72.00
9	1177000067271	12/23/1999	.		0	1	B1	75	ON	$134.00
10	1177010940613	12/09/1999	.		1	2	A1	42	NS	$11.00
11	1177025997013	11/09/1999	.		1	1	A1	26	BC	$154.00
12	1177027515760	10/19/1999	.		1	1	B1	73	BC	$16.00
13	1177028996676	09/21/2000	.		0	1	C1	.	QC	$179.00
14	1177038747105	03/14/2000	.		0	1	C1	41	ON	$705.00
15	1177045857516	06/22/2000	.		1	1	A1	53	QC	$83.00

Data Methodology

1.1 Explore and describe the dataset briefly

1.2 Customer's distribution

1.3 Segmentation

1.4 Statistic analysis

- 1.4.1 Tenure
- 1.4.2 Inactive distribution
- 1.4.3 Tenure segmentation
- 1.4.4 Status distribution by segment
- 1.4.5 Relationship between tenure segments vs. GoodCredit, RatePlan, DealerType
- 1.4.6 Sales study

Exploring Data

```
DATA Nandini.Telcm;
INFILE 'E:\Metro College\SAS\Advanced sas\New_Wireless_Fixed.txt' truncover;
INPUT Acctno: $ 15. Actdt:mmddyy10. Deactdt:mmddyy10. @36 DeactReason: $ 6.
@48 GoodCredit  @58 RatePlan: $ 2. @65 DealerType: $ 2. @72 Age @78 Province: $ 2.
@86 Sales: dollar10.2;
format Actdt mmddyy10. Deactdt mmddyy10. Sales dollar10.2 ;
RUN;
proc print Data=Nandini.Telcm (obs=15);run;
   proc contents Data=Nandini.telcm;run;
```

```
proc sql;
select count(distinct(Acctno))as Total_Actno
from Nandini.Telcm;
quit;
```

Total_Actno
102255

No Missing Account Numbers

The CONTENTS Procedure

Data Set Name	NANDINI.TELCM	Observations	102255
Member Type	DATA	Variables	10
Engine	V9	Indexes	0
Created	2024-10-08 12:25:00	Observation Length	72
Last Modified	2024-10-08 12:25:00	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
1	Acctno	Char	15	
2	Actdt	Num	8	MMDDYY10.
8	Age	Num	8	
4	DeactReason	Char	6	
3	Deactdt	Num	8	MMDDYY10.
7	DealerType	Char	2	
5	GoodCredit	Num	8	
9	Province	Char	2	
6	RatePlan	Char	2	
10	Sales	Num	8	DOLLAR10.2

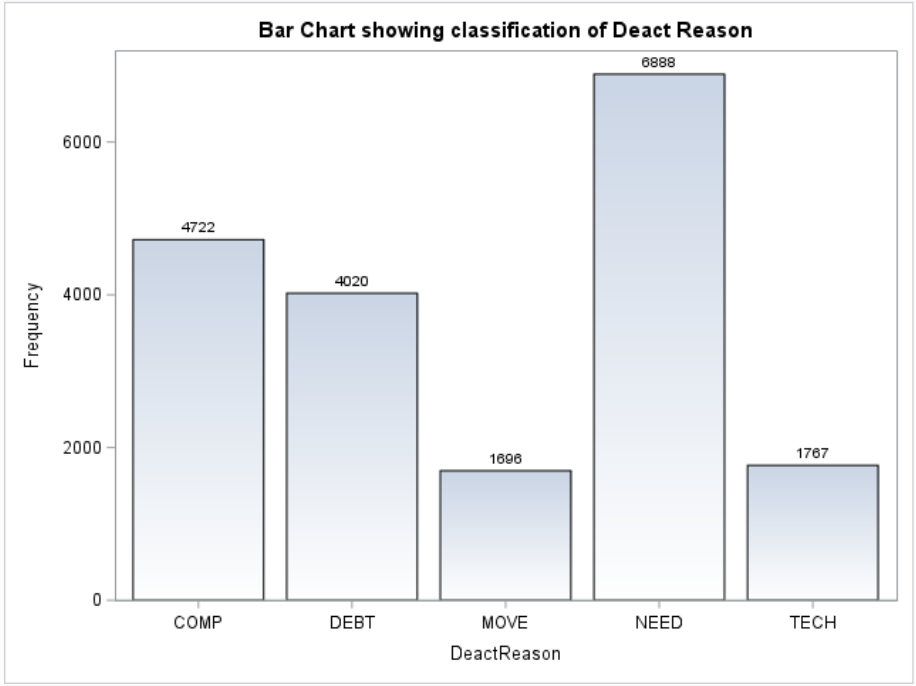
Distribution Of Categorical Variables

```
proc freq Data=Nandini.Telcm;
table DeactReason DealerType Province RatePlan/ missing;
run;
```

```
title"UNIVARIATE ANALYSIS";
title"Pie Chart showing classification of Deact Reason";
proc sgplot data=Nandini.Telcm;
vbar DeactReason/filltype=gradiant groupdisplay=cluster datalabel;
run;
```

DeactReason	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	83162	81.33	83162	81.33
COMP	4722	4.62	87884	85.95
DEBT	4020	3.93	91904	89.88
MOVE	1696	1.66	93600	91.54
NEED	6888	6.74	100488	98.27
TECH	1767	1.73	102255	100.00

Deact Reason



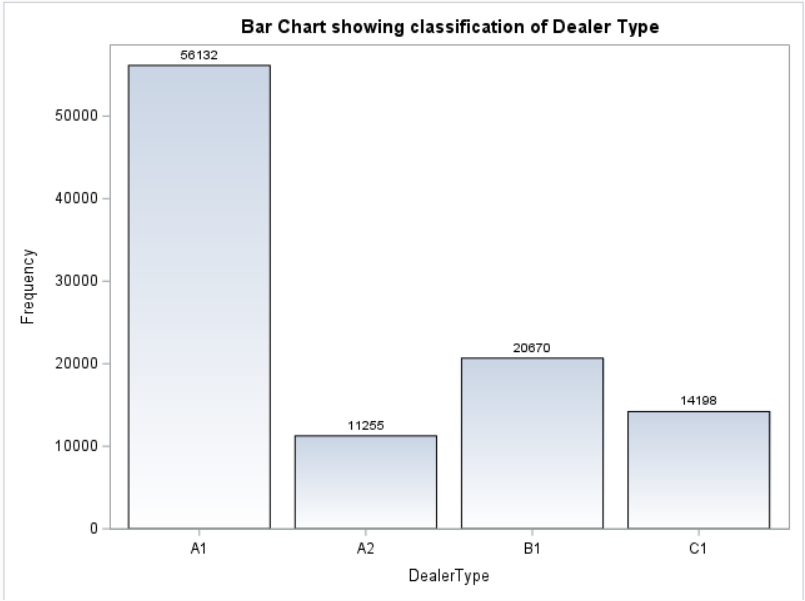
Distribution Of Categorical Variables

```
proc freq Data=Nandini.Telcm;
table DeactReason DealerType Province RatePlan/ missing;
run;
```

```
title"UNIVARIATE ANALYSIS";
title"Bar Chart showing classification of Dealer Type";
proc sgplot data=Nandini.Telcm;
vbar DealerType/filltype=gradiant groupdisplay=cluster datalabel;
run;
title;
```

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

Dealer Type



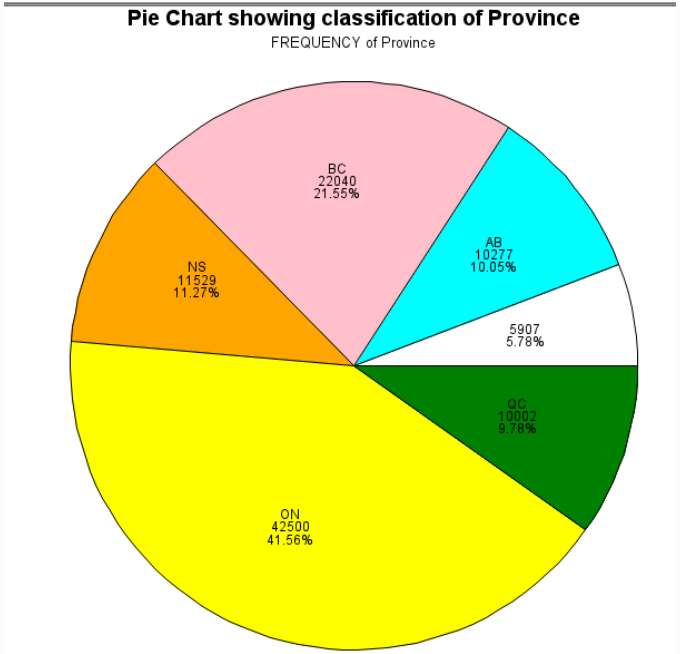
Distribution Of Categorical Variables

```
proc freq Data=Nandini.Telcm;
table DeactReason DealerType Province RatePlan/ missing;
run;
```

```
title"UNIVARIATE ANALYSIS";
title"Pie Chart showing classification of Province";
proc gchart data=Nandini.Telcm;
pie Province/missing discrete value= inside percent=inside;
goption colors=(white,Cyan,Pink, orange,Yellow,green);
run;
```

Province	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	5907	5.78	5907	5.78
AB	10277	10.05	16184	15.83
BC	22040	21.55	38224	37.38
NS	11529	11.27	49753	48.66
ON	42500	41.56	92253	90.22
QC	10002	9.78	102255	100.00

Province



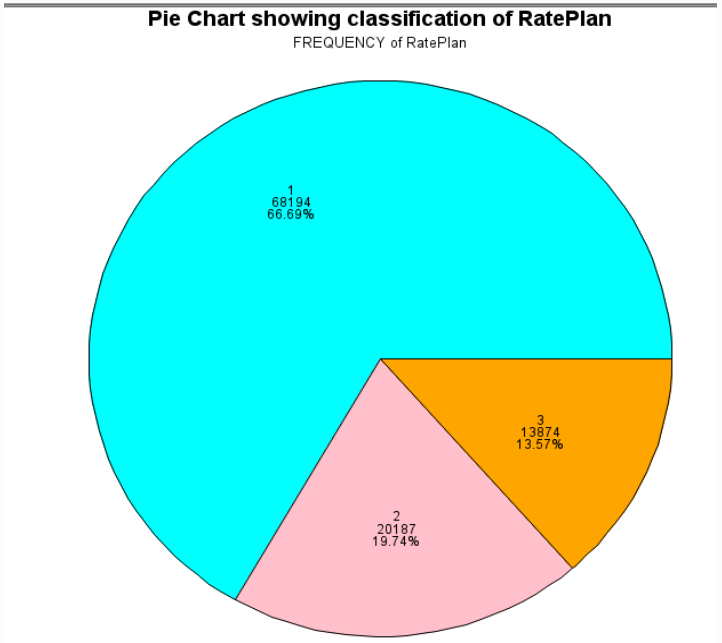
Distribution Of Categorical Variables

```
proc freq Data=Nandini.Telcm;
table DeactReason DealerType Province RatePlan/ missing;
run;
```

```
title"UNIVARIATE ANALYSIS";
title"Pie Chart showing classification of RatePlan";
proc gchart data=Nandini.Telcm;
pie Rateplan/discrete value= inside percent=inside;
goption colors=(Cyan,Pink, orange);
run;
```

RatePlan	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	68194	66.69	68194	66.69
2	20187	19.74	88381	86.43
3	13874	13.57	102255	100.00

Rate Plan



Distribution Of Categorical Variables

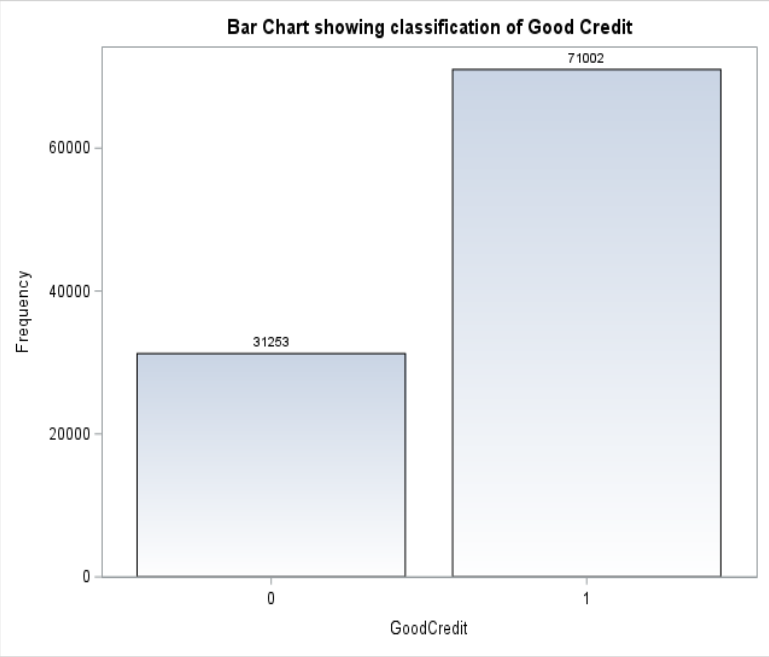
```
proc freq Data=Nandini.Telcm;
 table GoodCredit;run;
```

```
title"UNIVARIATE ANALYSIS";
title"Bar Chart showing classification of Good Credit";
proc sgplot data=Nandini.Telcm;
vbar GoodCredit/filltype=gradiant groupdisplay=cluster datalabel;
run;
title;
```

The FREQ Procedure

GoodCredit	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	31253	30.56	31253	30.56
1	71002	69.44	102255	100.00

Good Credit



Distribution Of Numerical Variables

Activation and Deactivation Dates

The MEANS Procedure

Variable	N	N Miss	Variance	Std Dev	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	Mean	Sum	Minimum	Maximum
Actdt	102255	0	40049.82	200.12	1.36	14674.13	14676.58	14675.35	1500628189.0	14264.00	14995.00
Deactdt	19635	82620	26871.28	163.92	1.11	14799.63	14804.21	14801.92	290635668.00	14269.00	14995.00

```
proc means Data=Nandini.Telcm n nmiss var std cv clm mean sum min max maxdec=2;
var Actdt Deactdt;
run;
```

```
/*Latest Activation Date*/
title;
proc sort Data=Nandini.Telcm out=Nandini.Act  nodupkey;
by descending Actdt;
run;
proc print Data=Nandini.Telcm (obs=1);
format Actdt mmddyy10.;
run;
```

Obs	Acctno	Actdt
1	1184263635198	01/20/2001

Latest Activation Date

```
/*Latest Deactivation Date*/
proc sort Data=Nandini.Telcm out=Nandini.Deact nodupkey;
by descending Deactdt;
run;
proc print Data=Nandini.Deact (obs=1);
format Deactdt mmddyy10.;
run;
```

Obs	Acctno	Actdt	Deactdt
1	1218085964217	11/29/1999	01/20/2001

Latest Deactivation Date

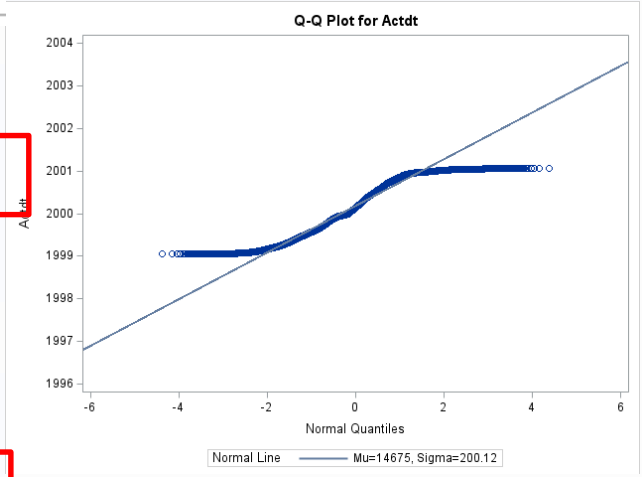
Distribution Of Numerical Variables

```
proc univariate Data=Nandini.telcm normal;
 var Actdt Deactdt;
 qqplot /normal (mu=est sigma=est);
 run;
```

Activation and Deactivation Dates Distribution

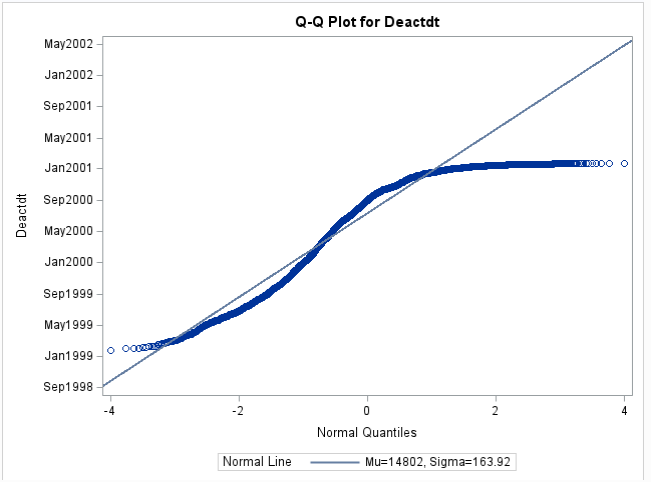
Activation

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.070832	Pr > D	<0.0100
Cramer-von Mises	W-Sq	153.9566	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1107.889	Pr > A-Sq	<0.0050



Deactivation

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.130351	Pr > D	<0.0100
Cramer-von Mises	W-Sq	103.8733	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	632.7164	Pr > A-Sq	<0.0050



- P-value for both categories of account status is <0.05 significance level
- Activation date and Deactivation dates are not normally distributed

Distribution Of Numerical Variables

```
proc means Data=Nandini.Telcm n nmiss var std cv clm mean sum min max qrange maxdec =2;
var Age;
run;
```

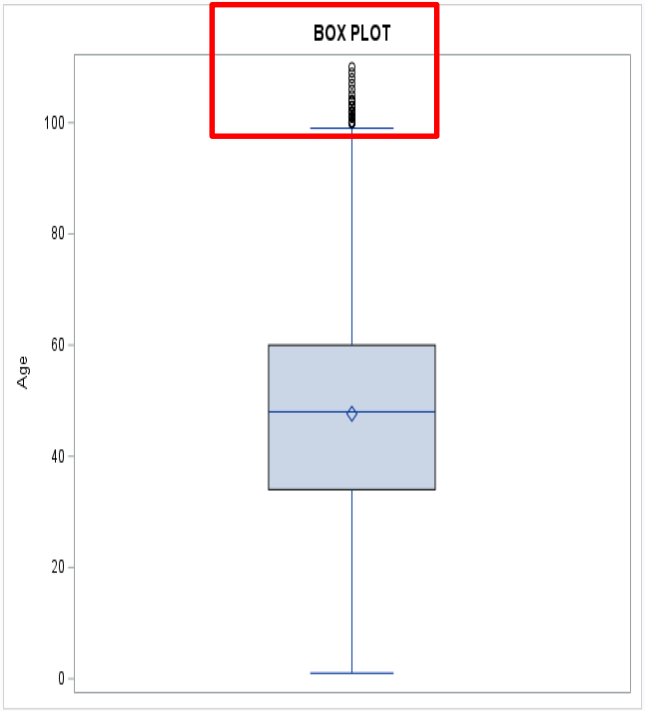
The MEANS Procedure

Analysis Variable : Age											
N	N Miss	Variance	Std Dev	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	Mean	Sum	Minimum	Maximum	Quartile Range
94547	7708	344.81	18.57	38.97	47.53	47.77	47.65	4504902.00	1.00	110.00	26.00

```
TITLE'BOX PLOT';
proc sgplot data = Nandini.Telcm;;
vBOX Age ;
run;
```

- Age has outliers . But within Outer fence.(Q3+3IQR)
- Age is not normally distributed

Age



Distribution Of Numerical Variables

```
proc univariate Data=Nandini.Telcm normal plot;
var Age;
qqplot /normal (mu=est sigma=est);
run;
```

The UNIVARIATE Procedure
Variable: Age

Moments			
N	94547	Sum Weights	94547
Mean	47.6472231	Sum Observations	4504902
Std Deviation	18.5690003	Variance	344.807771
Skewness	0.04374618	Kurtosis	-0.4066355
Uncorrected SS	247246266	Corrected SS	32600195.5
Coeff Variation	38.9718415	Std Error Mean	0.06038995

Basic Statistical Measures			
Location		Variability	
Mean	47.64722	Std Deviation	18.56900
Median	48.00000	Variance	344.80777
Mode	48.00000	Range	109.00000
		Interquartile Range	26.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	788.9926	Pr > |t|	<.0001
Sign	M	47273.5	Pr >= |M|	<.0001
Signed Rank	S	2.2348E9	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.02295	Pr > D	<0.0100
Cramer-von Mises	W-Sq	6.471068	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	49.27553	Pr > A-Sq	<0.0050

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	98548	106	55165
1	94136	107	84990
1	93661	108	85043
1	92838	109	29209
1	92571	110	21615

Age

Quantiles (Definition 5)	
Level	Quantile
100% Max	110
99%	89
95%	79
90%	72
75% Q3	60
50% Median	48
25% Q1	34
10%	23
5%	17
1%	7
0% Min	1

Distribution Of Numerical Variables

```
proc means Data=Nandini.Telcm n nmiss var std cv clm mean sum min max Q1 Q3 qrange maxdec =2;
var Sales;
run;
```

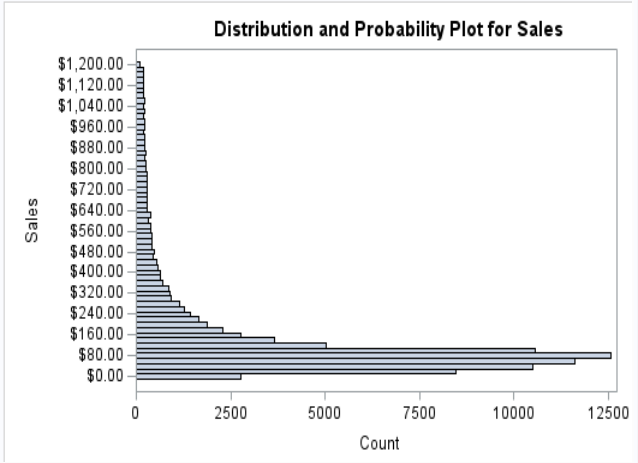
Sales

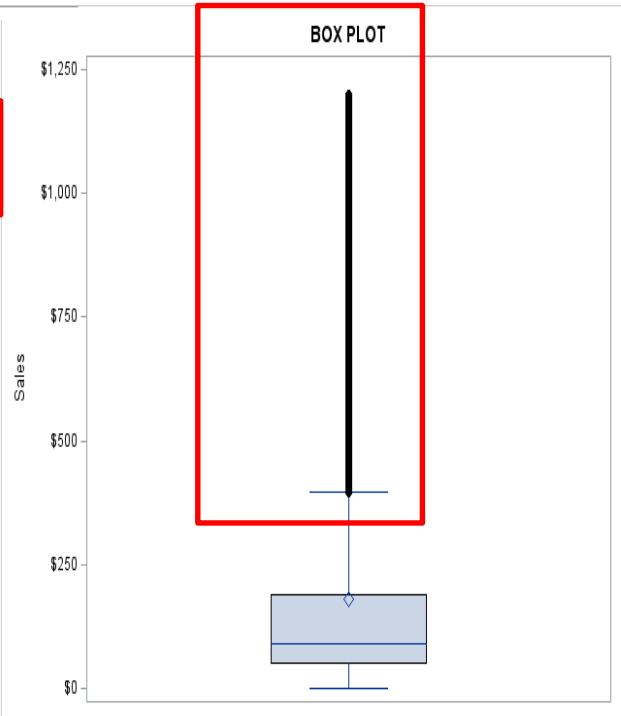
The MEANS Procedure

Analysis Variable : Sales

N	N Miss	Variance	Std Dev	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	Mean	Sum	Minimum	Maximum	Lower Quartile	Upper Quartile	Quartile Range
93650	8605	54742.45	233.97	129.09	179.75	182.74	181.25	16973703.00	0.00	1200.00	52.00	190.00	138.00

```
TITLE'BOX PLOT';
proc sgplot data = Nandini.Telcm;;
vBOX sales ;
run;
```





- Sales has outliers .

more than Q3+3IQR i.e $604

- Data is skewed
- Sales is not normally distributed

Distribution Of Numerical Variables

```
proc univariate Data=Nandini.Telcm normal plot;
var Sales;
qqplot /normal (mu=est sigma=est);
run;
```

The UNIVARIATE Procedure
Variable: Sales

Moments			
N	93650	Sum Weights	93650
Mean	181.246161	Sum Observations	16973703
Std Deviation	233.97104	Variance	54742.4477
Skewness	2.36652039	Kurtosis	5.28183679
Uncorrected SS	8202993991	Corrected SS	5126575480
Coeff Variation	129.090205	Std Error Mean	0.76455409

Basic Statistical Measures			
Location		Variability	
Mean	181.2462	Std Deviation	233.97104
Median	91.0000	Variance	54742
Mode	92.0000	Range	1200
		Interquartile Range	138.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	237.0613	Pr > |t|	<.0001
Sign	M	46790.5	Pr >= |M|	<.0001
Signed Rank	S	2.1894E9	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.250025	Pr > D	<0.0100
Cramer-von Mises	W-Sq	2131.825	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	11236.89	Pr > A-Sq	<0.0050

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	101798	1200	49191
0	99796	1200	50411
0	97151	1200	50506
0	89286	1200	65146
0	87254	1200	100136

Sales

Quantiles (Definition 5)	
Level	Quantile
100% Max	1200
99%	1100
95%	768
90%	490
75% Q3	190
50% Median	91
25% Q1	52
10%	26
5%	15
1%	4
0% Min	0

Bivariate Analysis

```
Data Nandini.Status;
set Nandini.Telcm;
length Status $ 12.;
If Deactdt eq . then Status="Active";
else if Deactdt ne . then Status="Deactivated";
proc print Data=Nandini.status (obs=20);run;

proc means Data=Nandini.Status n min max std mean cv clm maxdec=2;
class Status/missing;
var Age;
run;

proc univariate Data=Nandini.Status normal plot;
var Age;
Class Status;
qqplot /normal (mu=est sigma=est);
run;
```

The MEANS Procedure

Analysis Variable : Age									
Status	N Obs	N	Minimum	Maximum	Std Dev	Mean	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean
Active	82620	76377	1.00	109.00	18.58	47.63	39.00	47.50	47.76
Deactivated	19635	18170	1.00	110.00	18.53	47.71	38.84	47.44	47.98

Age Distribution of active and Deactivate Customers

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.022454	Pr > D	<0.0100
Cramer-von Mises	W-Sq	5.069319	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	38.36078	Pr > A-Sq	<0.0050

Age-Active

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.025526	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.444658	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	11.25617	Pr > A-Sq	<0.0050

Age-Deactivated

- P value of Age for both status is <0.05
- We reject null hypothesis of normal distribution.
- But as per CLT, we consider data is normally distributed
- Mean and Std Dev in both status groups appears

Approximately similar.

Bivariate Analysis

Age Distribution of active and Deactivate Customers

```
/*Equality of variance*/
proc glm data=Nandini.Status;
class Status;
model Age = Status;
means Status / hovtest=levene(type=abs) welch;
run;
```

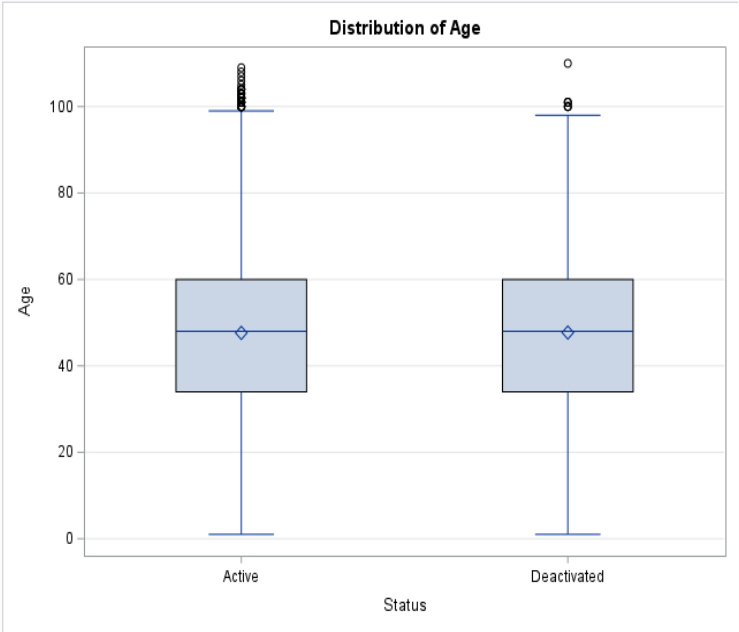
The GLM Procedure

Levene's Test for Homogeneity of Age Variance
ANOVA of Absolute Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Status	1	5.4265	5.4265	0.05	0.8307
Error	94545	11219747	118.7		

Welch's ANOVA for Age

Source	DF	F Value	Pr > F
Status	1.0000	0.26	0.6080
Error	27524.8		



- More outliers in active status as compared to Deactivated status
- P value of for Homocedacity is 0.8307.>0.05
- We fail to reject null hypothesis of Homocedacity.
- We conclude variance of Age is equal in both status.

Bivariate Analysis

Age Distribution of active and Deactivate Customers

```
proc ttest Data=Nandini.Status;
Var Age;
Class Status;
run;
```

The TTEST Procedure

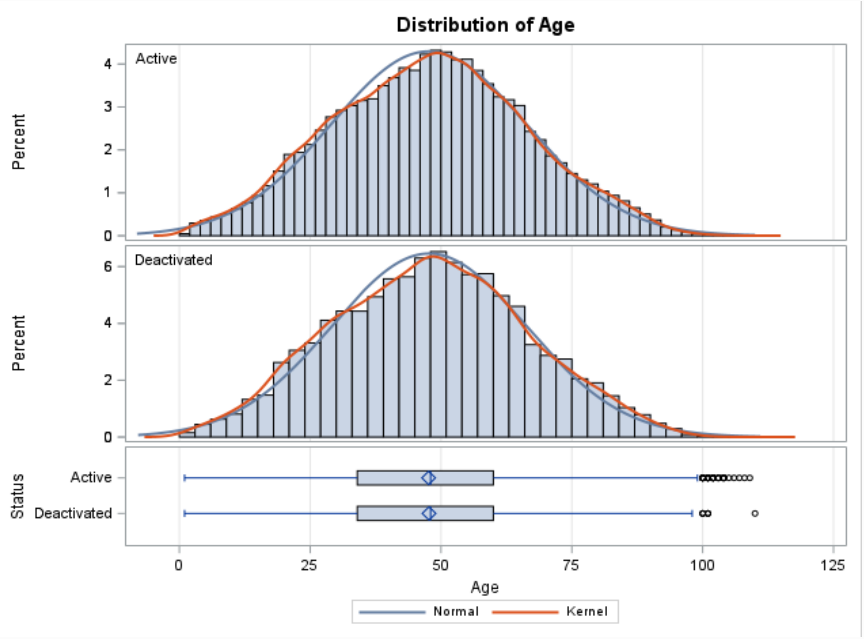
Variable: Age

Status	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Active		76377	47.6321	18.5786	0.0672	1.0000	109.0
Deactivated		18170	47.7106	18.5290	0.1375	1.0000	110.0
Diff (1-2)	Pooled		-0.0785	18.5691	0.1533		
Diff (1-2)	Satterthwaite		-0.0785		0.1530		

Status	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Active		47.6321	47.5004	47.7639	18.5786	18.4859	18.6722
Deactivated		47.7106	47.4412	47.9801	18.5290	18.3405	18.7215
Diff (1-2)	Pooled	-0.0785	-0.3789	0.2219	18.5691	18.4858	18.6532
Diff (1-2)	Satterthwaite	-0.0785	-0.3784	0.2214			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	94545	-0.51	0.6086
Satterthwaite	Unequal	27525	-0.51	0.6080

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	76376	18169	1.01	0.6492



- Homocedacity pvalue-0.6492 >0.05- Variances are equal
- Pooled test P value is 0.6086>0.05- Means are equal
- Age is equally distributed in both active and Deactive status

Bivariate Analysis

Province Distribution of active and Deactivate Customers

```
proc freq Data=Nandini.Status;
table Province*Status/missing chisq norow nocol;
run;
title"Comparison between Province and Status";
proc SGplot Data=Nandini.Status;
vbar Province/group=status filltype=Gradiant groupdisplay=cluster datalabel;
run;
```

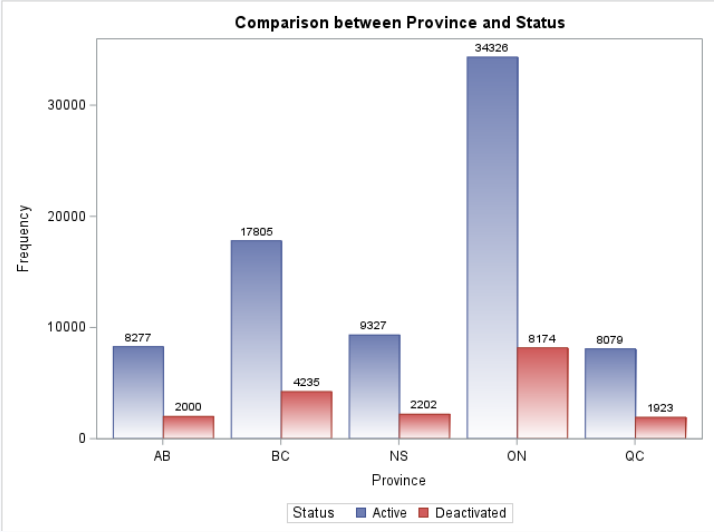
The FREQ Procedure

Frequency Percent	Table of Province by Status		
	Status		
Province	Active	Deactivated	Tota
	4806 4.70	1101 1.08	5907 5.78
AB	8277 8.09	2000 1.96	10277 10.05
BC	17805 17.41	4235 4.14	22040 21.55
NS	9327 9.12	2202 2.15	11529 11.27
ON	34326 33.57	8174 7.99	42500 41.56
QC	8079 7.90	1923 1.88	10002 9.78
Total	82620 80.80	19635 19.20	102255 100.00

Statistics for Table of Province by Status

Statistic	DF	Value	Prob
Chi-Square	5	1.7616	0.8811
Likelihood Ratio Chi-Square	5	1.7693	0.8801
Mantel-Haenszel Chi-Square	1	0.1233	0.7255
Phi Coefficient		0.0042	
Contingency Coefficient		0.0042	
Cramer's V		0.0042	

Sample Size = 102255



- P value of >0.05- We fail to reject null hypothesis of Independency
- Province and Status(active/Deactivated) are not Statistically significantly associated with each other.
- Province and Status are Independent of each other

Bivariate Analysis

Age Distribution for Province

```
/*Age Vs Province Descriptive Analysis*/
proc means Data=Nandini.Telcm n min max std mean cv clm maxdec=2;
var Age;
class Province/missing;
run;

/*Age Vs. province Normality Test*/
proc univariate Data=Nandini.Telcm normal plot;
var Age;
class Province;
```

The MEANS Procedure

Analysis Variable : Age

Province	N Obs	N	Minimum	Maximum	Std Dev	Mean	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean
	5907	5459	1.00	106.00	18.29	47.74	38.31	47.25	48.22
AB	10277	9500	1.00	102.00	18.57	47.63	38.98	47.26	48.01
BC	22040	20437	1.00	109.00	18.79	47.77	39.33	47.51	48.02
NS	11529	10692	1.00	103.00	18.38	47.57	38.64	47.22	47.92
ON	42500	39222	1.00	110.00	18.55	47.61	38.97	47.42	47.79
QC	10002	9237	1.00	102.00	18.53	47.60	38.94	47.22	47.97

- P value of Age for both status is <0.05
- We reject null hypothesis of normal distribution.
- But as per CLT, we consider data is normally distributed
- Mean and Std Dev in both status groups appears approximately similar.

Tests for Normality — Age-AB

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.025901	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.8705	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	6.17301	Pr > A-Sq	<0.0050

Tests for Normality — Age-BC

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.023013	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.571222	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	11.64046	Pr > A-Sq	<0.0050

Tests for Normality — Age-NS

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.024863	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.799578	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	5.561931	Pr > A-Sq	<0.0050

Tests for Normality — Age-ON

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.022965	Pr > D	<0.0100
Cramer-von Mises	W-Sq	2.610763	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	20.40963	Pr > A-Sq	<0.0050

Tests for Normality — Age-QC

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.02473	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.682977	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	5.321143	Pr > A-Sq	<0.0050

Bivariate Analysis

```
/*CHecking Equality of Variances */
proc glm data=Nandini.Telcm;
class Province;
model Age = Province;
means Province / hovtest=levene(type=abs) welch;
run;
```

The GLM Procedure

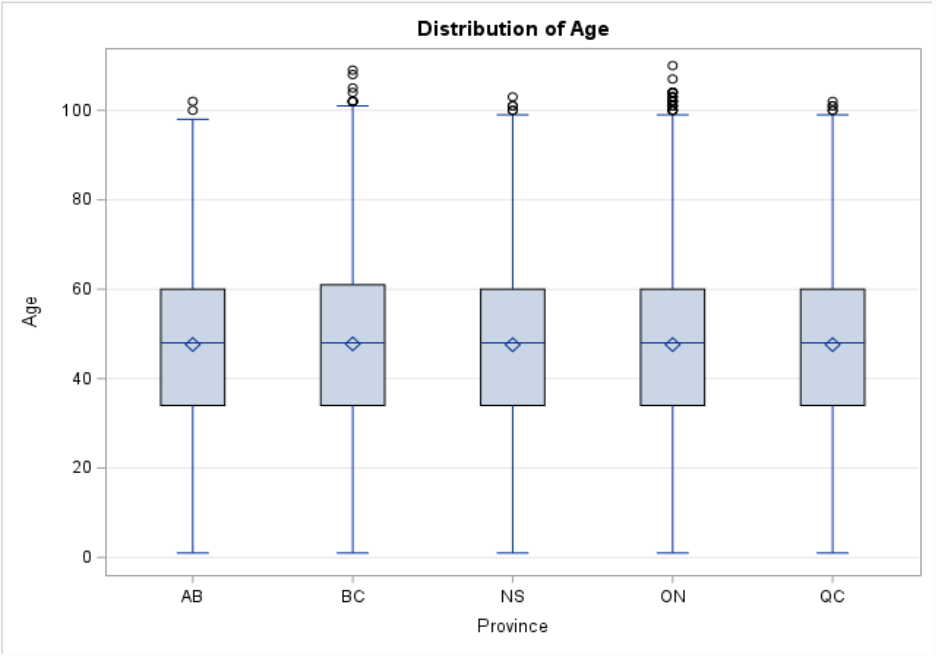
Levene's Test for Homogeneity of Age Variance
ANOVA of Absolute Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Province	4	933.4	233.3	1.96	0.0970
Error	89083	10583759	118.8		

Welch's ANOVA for Age

Source	DF	F Value	Pr > F
Province	4.0000	0.31	0.8724
Error	28548.3		

Age Distribution for Province



- Outliers exists in all province groups.
- P value of for Homocedacity is 0.0970.>0.05
- We fail to reject null hypothesis of Homocedacity.
- We conclude variance of Age is equal in all provinces.

Bivariate Analysis

Age Distribution for Province

```
PROC ANOVA DATA = Nandini.Telcm PLOTS(MAXPOINTS=20 );
 CLASS Province;
 MODEL Age = Province;
 MEANS Province/scheffe;
TITLE "Age distribution across Province";
RUN;
QUIT;
```

The ANOVA Procedure

Dependent Variable: Age

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	431.97	107.99	0.31	0.8697
Error	89083	30774067.45	345.45		
Corrected Total	89087	30774499.41			

R-Square	Coeff Var	Root MSE	Age Mean
0.000014	39.01293	18.58639	47.64161

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Province	4	431.9676449	107.9919112	0.31	0.8697

Comparisons significant at the 0.05 level are indicated by ***.

Province Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
BC - AB	0.1317	-0.5792	0.8426
BC - ON	0.1580	-0.3359	0.6519
BC - QC	0.1697	-0.5481	0.8875
BC - NS	0.1941	-0.4892	0.8774
AB - BC	-0.1317	-0.8426	0.5792
AB - ON	0.0263	-0.6283	0.6810
AB - QC	0.0380	-0.7986	0.8746
AB - NS	0.0624	-0.7448	0.8696
ON - BC	-0.1580	-0.6519	0.3359
ON - AB	-0.0263	-0.6810	0.6283
ON - QC	0.0117	-0.6504	0.6738
ON - NS	0.0361	-0.5885	0.6607
QC - BC	-0.1697	-0.8875	0.5481
QC - AB	-0.0380	-0.8746	0.7986
QC - ON	-0.0117	-0.6738	0.6504
QC - NS	0.0244	-0.7889	0.8377
NS - BC	-0.1941	-0.8774	0.4892
NS - AB	-0.0624	-0.8696	0.7448

Difference in means of pairwise province is very small

P value-0.8697 >0.05- We fail to reject null hypothesis

Means of age are equal in all province

Age is equally distributed in all province

Customer Segmentation

Age

```
proc format;
 value Agegroup
 low-20='<20'
 21- 40='21-40'
 41-60='41-60'
 61-High='60 and above';
 run;
```

Sales

```
proc format;
 value SalesGroup
 low-100='<$100'
 101- 500='$100-$500'
 501-800='$500-$800'
 801-High='$800 and above'
 ;
 RUN;
```

```
Data Nandini.Telcml;
 set Nandini.Telcm;
 Agesegment=Age;
 Salessegment=Sales;
 format Agesegment Agegroup. Salessegment SalesGroup.;
proc print Data=Nandini.Telcml (obs=20);
 run;
```

```
/*Customer Segemetation based on Province, Agegroup and Sales group*/
proc freq Data=Nandini.Telcml;
 table Province Agesegment Salessegment/missing;
 run;
 title;
```

Province	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	5907	5.78	5907	5.78
AB	10277	10.05	16184	15.83
BC	22040	21.55	38224	37.38
NS	11529	11.27	49753	48.66
ON	42500	41.56	92253	90.22
QC	10002	9.78	102255	100.00

Agesegment	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	7708	7.54	7708	7.54
<20	7137	6.98	14845	14.52
21-40	26382	25.80	41227	40.32
41-60	37478	36.65	78705	76.97
60 and above	23550	23.03	102255	100.00

Salessegment	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	8605	8.42	8605	8.42
<$100	52965	51.80	61570	60.21
$100-$500	31534	30.84	93104	91.05
$500-$800	4920	4.81	98024	95.86
$800 and above	4231	4.14	102255	100.00

Statistical Analysis

Tenure

```
Title"Latest Activation and Deactivation Dates";
proc sql;
Create table Nandini.Dates as
select max(Actdt) as Latest_Activation_Date,
       max(Deactdt) as Latest_Deactivation_Date
from Nandini.Telcm;
run;
proc print Data=Nandini.Dates;
format Latest_Activation_Date Latest_Deactivation_Date mmddyy10.;run;
title;
```

```
Data Nandini.Tenure;
set Nandini.Telcm;
dl="20JAN2001"D;
if Deactdt eq . then Tenuredays=intck('day', Actdt, Dl);
if Deactdt ne . then Tenuredays=intck('day', Actdt, Deactdt);
RUN;
PROC PRINT DATA=Nandini.Tenure (obs=20);
FORMAT Dl DATE9.;
RUN;

proc means Data=Nandini.Tenure maxdec=2;
var Tenuredays;run;
```

Obs	Acctno	Actdt	Deactdt	DeactReason	GoodCredit	RatePlan	DealerType	Age	Province	Sales	d1	Tenuredays
1	1176913194483	06/20/1999	.		0	1	A1	58	BC	$128.00	20JAN2001	580
2	1176914599423	10/04/1999	10/15/1999	NEED	1	1	A1	45	AB	$72.00	20JAN2001	11
3	1176951913656	07/01/2000	.		0	1	A1	57	BC	$593.00	20JAN2001	203
4	1176954000288	05/30/2000	.		1	2	A1	47	ON	$83.00	20JAN2001	235
5	1176969186303	12/13/2000	.		1	1	C1	82	BC	.	20JAN2001	38
6	1176991056273	08/31/1999	09/18/2000	MOVE	1	1	C1	92	QC	$1,041.00	20JAN2001	384
7	1176991866552	05/24/2000	.		1	1	A1	77	ON	.	20JAN2001	241
8	1176992889500	11/28/2000	.		1	1	C1	68	AB	$72.00	20JAN2001	53
9	1177000067271	12/23/1999	.		0	1	B1	75	ON	$134.00	20JAN2001	394
10	1177010940613	12/09/1999	.		1	2	A1	42	NS	$11.00	20JAN2001	408

Latest Activation and Deactivation Dates

Obs	Latest_Activation_Date	Latest_Deactivation_Date
1	01/20/2001	01/20/2001

The MEANS Procedure

Analysis Variable : Tenuredays				
N	Mean	Std Dev	Minimum	Maximum
102255	282.57	197.32	0.00	731.00

Statistical Analysis

Accounts Deactivated Each Month

```
Data Nandini.Deact;
set Nandini.Telcm;
Month=month(Deactdt);
format Deactdt date9.;
proc print Data=Nandini.Deact (OBS=20);run

proc sql;
select actdt,deactdt from Nandini.Telcm
where actdt=deactdt
;
quit;
```

```
proc sql;
select month,count(Acctno) as Total_Deactivated
from Nandini.Deact
where not missing(Deactdt)
group by Month
order by Month
;
quit;
```

Obs	Acctno	Actdt	Deactdt	DeactReason	GoodCredit	RatePlan	DealerType	Age	Province	Sales	Month
1	1176913194483	06/20/1999	.		0	1	A1	58	BC	$128.00	.
2	1176914599423	10/04/1999	15OCT1999	NEED	1	1	A1	45	AB	$72.00	10
3	1176951913656	07/01/2000	.		0	1	A1	57	BC	$593.00	.
4	1176954000288	05/30/2000	.		1	2	A1	47	ON	$83.00	.
5	1176969186303	12/13/2000	.		1	1	C1	82	BC	.	.
6	1176991056273	08/31/1999	18SEP2000	MOVE	1	1	C1	92	QC	$1,041.00	9
7	1176991866552	05/24/2000	.		1	1	A1	77	ON	.	.
8	1176992889500	11/28/2000	.		1	1	C1	68	AB	$72.00	.
9	1177000067271	12/23/1999	.		0	1	B1	75	ON	$134.00	.
10	1177010940613	12/09/1999	.		1	2	A1	42	NS	$11.00	.

Actdt	Deactdt
12/28/2000	12/28/2000
09/13/2000	09/13/2000
01/03/2000	01/03/2000
01/14/2001	01/14/2001
08/11/2000	08/11/2000
10/21/1999	10/21/1999
05/15/2000	05/15/2000
09/29/2000	09/29/2000
01/15/2000	01/15/2000
05/16/2000	05/16/2000
09/13/1999	09/13/1999
05/26/2000	05/26/2000
01/28/2000	01/28/2000
12/16/2000	12/16/2000
12/02/2000	12/02/2000

Month	Total_Deactivated
1	2494
2	553
3	760
4	731
5	914
6	1403
7	1380
8	1494
9	1717
10	2817
11	2076
12	3296

- 340 Customers Deactivated service on the same day
- Maximum Deactivation occurred in Winter.

Statistical Analysis

Segmentation

```
Data Nandini.Status_Tenure;
set Nandini.Tenure;
length Acct_Status $ 12. Tenure $ 25.;
If Deactdt eq . then Acct_Status="Active";
else if Deactdt ne . then Acct_Status="Deactivated";
if Tenuredays <30 then Tenure="0-30 Days";
else if Tenuredays >=31 and Tenuredays<60 then Tenure="31--60Days";
else if Tenuredays >=61 and Tenuredays<366 then Tenure="61 days --One Year";
else Tenure="Over One Year";
run;
proc print Data=Nandini.Status_Tenure (obs=20);run;

proc freq Data=Nandini.Status_Tenure;
table Acct_Status Tenure/missing;run;
```

The FREQ Procedure

Acct_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Active	82620	80.80	82620	80.80
Deactivated	19635	19.20	102255	100.00

Tenure	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0-30 Days	9486	9.28	9486	9.28
31--60Days	7980	7.80	17466	17.08
61 days --One Year	45389	44.39	62855	61.47
Over One Year	39400	38.53	102255	100.00

Obs	Acctno	Actdt	Deactdt	DeactReason	GoodCredit	RatePlan	DealerType	Age	Province	Sales	d1	Tenuredays	Acct_Status	Tenure
1	1176913194483	06/20/1999	.		0	1	A1	58	BC	$128.00	14995	580	Active	Over One Year
2	1176914599423	10/04/1999	10/15/1999	NEED	1	1	A1	45	AB	$72.00	14995	11	Deactivated	0-30 Days
3	1176951913656	07/01/2000	.		0	1	A1	57	BC	$593.00	14995	203	Active	61 days --One Year
4	1176954000288	05/30/2000	.		1	2	A1	47	ON	$83.00	14995	235	Active	61 days --One Year
5	1176969186303	12/13/2000	.		1	1	C1	82	BC	.	14995	38	Active	31--60Days
6	1176991056273	08/31/1999	09/18/2000	MOVE	1	1	C1	92	QC	$1,041.00	14995	384	Deactivated	Over One Year
7	1176991866552	05/24/2000	.		1	1	A1	77	ON	.	14995	241	Active	61 days --One Year
8	1176992889500	11/28/2000	.		1	1	C1	68	AB	$72.00	14995	53	Active	31--60Days
9	1177000067271	12/23/1999	.		0	1	B1	75	ON	$134.00	14995	394	Active	Over One Year
10	1177010940613	12/09/1999	.		1	2	A1	42	NS	$11.00	14995	408	Active	Over One Year
11	1177025997013	11/09/1999	.		1	1	A1	26	BC	$154.00	14995	438	Active	Over One Year

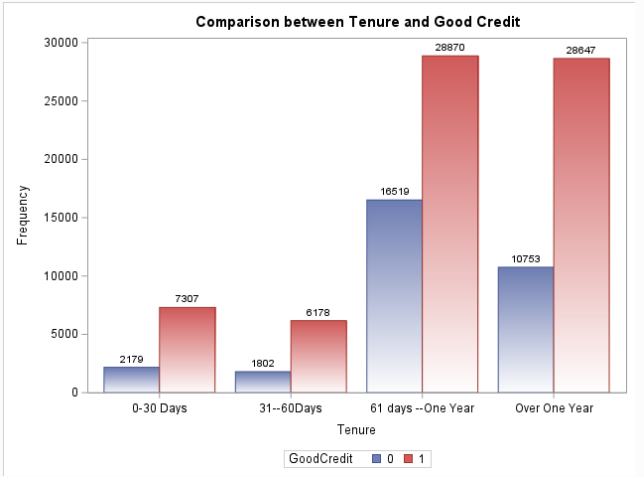
Statistical Analysis

Association Tenure Vs Good Credit

```
proc freq Data=Nandini.Status_Tenure;
table Tenure*GoodCredit/Missing chisq norow nocol;
run;
proc freq Data=Nandini.Status;
table Status*goodcredit/missing;
run;
title"Comparison between Tenure and Good Credit ";
proc SGplot Data=Nandini.Status_tenure;
vbar Tenure/group=GoodCredit filltype=Gradient
groupdisplay=cluster datalabel;
run;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Status by GoodCredit		
	GoodCredit		
Status	0	1	Total
Active	22596 22.10 27.35 72.30	60024 58.70 72.65 84.54	82620 80.80
Deactivated	8657 8.47 44.09 27.70	10978 10.74 55.91 15.46	19635 19.20
Total	31253 30.56	71002 69.44	102255 100.00



Frequency Percent	Table of Tenure by GoodCredit		
	GoodCredit		
Tenure	0	1	Total
0-30 Days	2179 2.13	7307 7.15	9486 9.28
31--60Days	1802 1.76	6178 6.04	7980 7.80
61 days --One Year	16519 16.15	28870 28.23	45389 44.39
Over One Year	10753 10.52	28647 28.02	39400 38.53
Total	31253 30.56	71002 69.44	102255 100.00

Statistics for Table of Tenure by GoodCredit

Statistic	DF	Value	Prob
Chi-Square	3	1423.1037	<.0001
Likelihood Ratio Chi-Square	3	1432.8657	<.0001
Mantel-Haenszel Chi-Square	1	298.0330	<.0001
Phi Coefficient		0.1180	
Contingency Coefficient		0.1172	
Cramer's V		0.1180	

- P value-<0.05- reject null hypothesis Of Independency. Means there is statistically significant association between Tenure and Good Credit
- Customers with greater tenure have Good Credit
- Majority Active customers are likely to have Good Credit.

Statistical Analysis

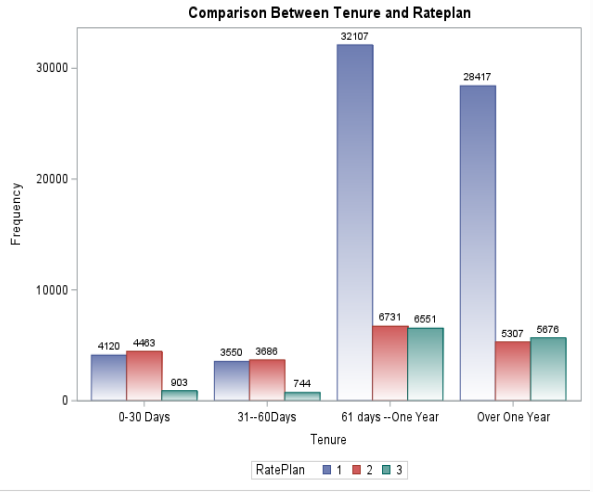
Association Tenure Vs Rate Plan

```
proc freq Data=Nandini.Status_Tenure;
table Tenure*Rateplan/Missing chisq norow nocol;
run;
proc freq Data=Nandini.Status;
table Status*Rateplan/missing;
run;
title"Comparison Between Tenure and Rateplan";
proc sgplot Data=Nandini.Status_tenure;
vbar Tenure/group=Rateplan filltype=Gradient
Groupdisplay=cluster datalabel;
run;
```

Frequency
Percent
Row Pct
Col Pct

Table of Status by RatePlan

Status	RatePlan 1	2	3	Total
Active	55725 54.50 67.45 81.72	16748 16.38 20.27 82.96	10147 9.92 12.28 73.14	82620 80.80
Deactivated	12469 12.19 63.50 18.28	3439 3.36 17.51 17.04	3727 3.64 18.98 26.86	19635 19.20
Total	68194 66.69	20187 19.74	13874 13.57	102255 100.00



Frequency
Percent

Table of Tenure by RatePlan

Tenure	RatePlan 1	2	3	Total
0-30 Days	4120 4.03	4463 4.36	903 0.88	9486 9.28
31--60Days	3550 3.47	3686 3.60	744 0.73	7980 7.80
61 days --One Year	32107 31.40	6731 6.58	6551 6.41	45389 44.39
Over One Year	28417 27.79	5307 5.19	5676 5.55	39400 38.53
Total	68194 66.69	20187 19.74	13874 13.57	102255 100.00

Statistics for Table of Tenure by RatePlan

Statistic	DF	Value	Prob
Chi-Square	6	9661.6962	<.0001
Likelihood Ratio Chi-Square	6	8227.3953	<.0001
Mantel-Haenszel Chi-Square	1	179.3125	<.0001
Phi Coefficient		0.3074	
Contingency Coefficient		0.2938	
Cramer's V		0.2174	

- P value-<0.05- reject null hypothesis

Of Independency. Means there is statistically significant association between Tenure and Rate Plan

- Customers with Rate plan 1 have tenure more than 60 days.
- Majority Active customers are likely to have Rate plan 1
- Rate Plan 2 has minimum churn followed by rate plan 1
- Rate Plan 3 has Maximum Churn.
- Over all rate plan 1 is better than other 2

Statistical Analysis

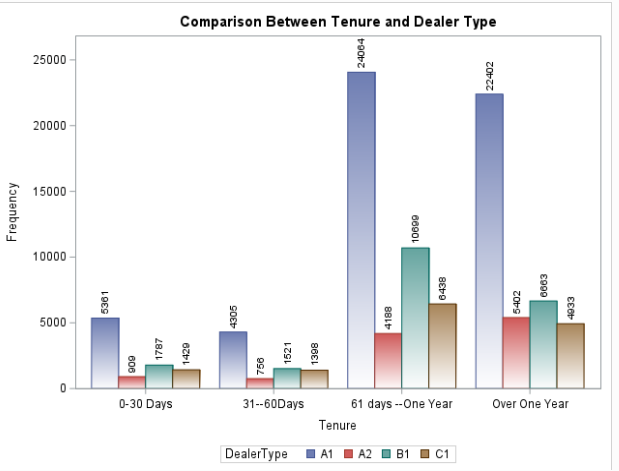
Association Tenure Vs Dealer Type

```
proc freq Data=Nandini.Status_Tenure;
table Tenure*DealerType/Missing chisq norow nocol;
run;
proc freq Data=Nandini.Status;
table Status*Dealertype/missing norow nocol;
run;
Proc sgplot Data=Nandini.Status_Tenure;
vbar Tenure/group=DealerType filltype=Gradient
Groupdisplay=cluster datalabel;
run;
```

Frequency
Percent

Table of Status by DealerType

Status	DealerType A1	A2	B1	C1	Total
Active	45501 44.50	8706 8.51	16791 16.42	11622 11.37	82620 80.80
Deactivated	10631 10.40	2549 2.49	3879 3.79	2576 2.52	19635 19.20
Total	56132 54.89	11255 11.01	20670 20.21	14198 13.88	102255 100.00



Frequency
Percent

Table of Tenure by DealerType

Tenure	DealerType A1	A2	B1	C1	Total
0-30 Days	5361 5.24	909 0.89	1787 1.75	1429 1.40	9486 9.28
31--60Days	4305 4.21	756 0.74	1521 1.49	1398 1.37	7980 7.80
61 days --One Year	24064 23.53	4188 4.10	10699 10.46	6438 6.30	45389 44.39
Over One Year	22402 21.91	5402 5.28	6663 6.52	4933 4.82	39400 38.53
Total	56132 54.89	11255 11.01	20670 20.21	14198 13.88	102255 100.00

Statistic	DF	Value	Prob
Chi-Square	9	1110.5992	<.0001
Likelihood Ratio Chi-Square	9	1096.4370	<.0001
Mantel-Haenszel Chi-Square	1	328.9110	<.0001
Phi Coefficient		0.1042	
Contingency Coefficient		0.1037	
Cramer's V		0.0602	

- P value-<0.05- reject null hypothesis Of Independency. Means there is statistically significant association between Tenure and Dealer Type
- Customers with Dealer Type A1 have tenure more than 60 days.
- Majority Active customers are likely to deal with A1 Dealer

Statistical Analysis

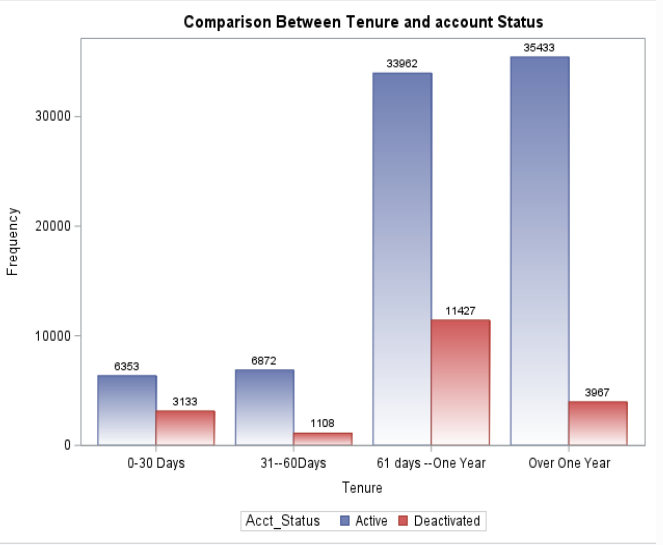
Association Tenure Vs Account Status

```
proc freq Data=Nandini.Status_tenure;
table acct_Status*Tenure/chisq missing norow nocol;
run;
```

```
Title"Comparison Between Tenure and account Status";
Proc sgplot Data=Nandini.Status_Tenure;
vbar Tenure/Group=acct_Status filltype=Gradient
Groupdisplay=CLuster datalabel;
run;
```

Statistics for Table of Acct_Status by Tenure

Statistic	DF	Value	Prob
Chi-Square	3	4476.5710	<.0001
Likelihood Ratio Chi-Square	3	4611.5562	<.0001
Mantel-Haenszel Chi-Square	1	1716.4870	<.0001
Phi Coefficient		0.2092	
Contingency Coefficient		0.2048	
Cramer's V		0.2092	



Frequency Percent	Table of Acct_Status by Tenure				
	Tenure				
Acct_Status	0-30 Days	31--60Days	61 days --One Year	Over One Year	Total
Active	6353 6.21	6872 6.72	33962 33.21	35433 34.65	82620 80.80
Deactivated	3133 3.06	1108 1.08	11427 11.18	3967 3.88	19635 19.20
Total	9486 9.28	7980 7.80	45389 44.39	39400 38.53	102255 100.00

- P value-<0.05- reject null hypothesis

Of Independency. Means there is statistically significant association between Tenure and Account Status

- Majority Active Customers have tenure more than 60 days
- Customer churn is seen more with tenure <30 days.(33%)

Statistical Analysis

Alternate Tenure Strategy

```
/*Alternate tenure strateagy*/
Data Nandini.AltTenure;
set Nandini.Status_Tenure;
length AltTenure $ 20.;
if Tenuredays <30 then AltTenure="1 month and less";
else if Tenuredays >=31 and Tenuredays<60 then AltTenure="2months";
else if Tenuredays >=61 and Tenuredays<90 then AltTenure="3 months";
else if Tenuredays >=91 and Tenuredays<180 then AltTenure="3 to 6 months";
else if Tenuredays >=181 and Tenuredays<366 then AltTenure="6Months-1 yr";
else AltTenure="year and above";
run;
proc print Data=Nandini.AltTenure (obs=20);run;
```

- Majority Active Customers have tenure more than 1 year
- Customer churn is seen more in first month (33%).Least in One year and Above
- Churn trend

Frequency
Percent

Table of Acct_Status by AltTenure

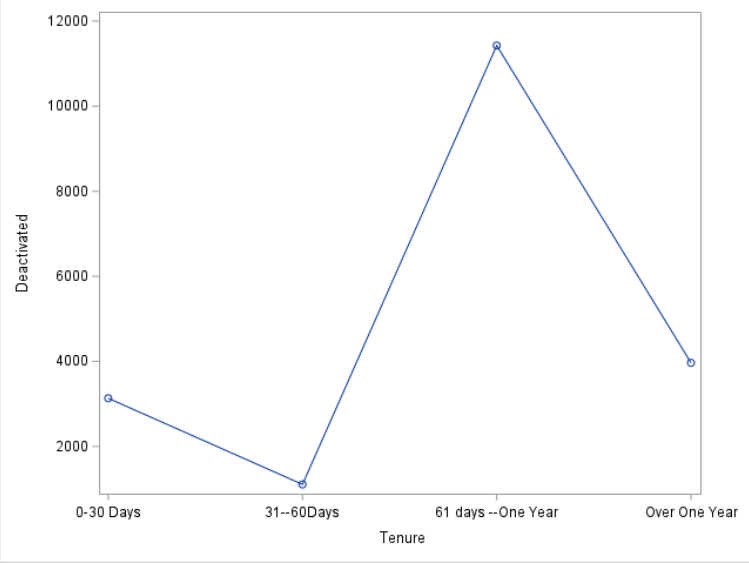
Acct_Status	AltTenure						
	1 month and less	2months	3 months	3 to 6 months	6Months-1 yr	year and above	Total
Active	6353 6.21	6872 6.72	3942 3.86	10716 10.48	19121 18.70	35616 34.83	82620 80.80
Deactivated	3133 3.06	1108 1.08	1253 1.23	4051 3.96	6051 5.92	4039 3.95	19635 19.20
Total	9486 9.28	7980 7.80	5195 5.08	14767 14.44	25172 24.62	39655 38.78	102255 100.00

Statistic	DF	Value	Prob
Chi-Square	5	4497.4179	<.0001
Likelihood Ratio Chi-Square	5	4607.2354	<.0001
Mantel-Haenszel Chi-Square	1	2002.8305	<.0001
Phi Coefficient		0.2097	
Contingency Coefficient		0.2053	
Cramer's V		0.2097	

Statistical Analysis

```
proc freq Data=Nandini.AltTenure ;
 table acct_Status*AltTenure/chisq missing norow nocol out=Nandini.freq;
 run;
proc print Data=Nandini.freq;run;
```

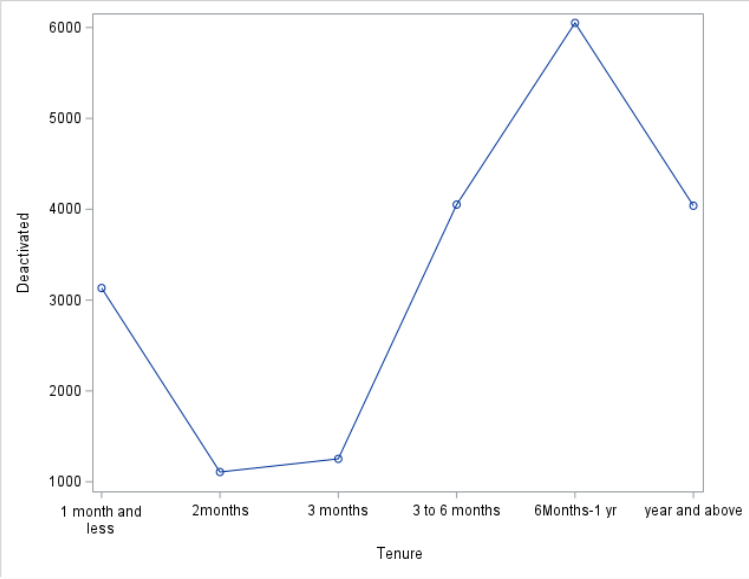
Original Tenure Segmentation



Alternate Tenure Strategy

```
proc sgplot Data=Nandini.freq;
where acct_status="Deactivated";
series x=Alttenure y= count/markers;
xaxis label="Tenure";
yaxis label="Deactivated";
run;
```

Alternate Tenure Segmentation



Statistical Analysis

Sales Distribution Across Age segments

```
Data Nandini.Sales;
set Nandini.Status;
Agesegment=Age;
format Agesegment Agegroup.;
run;
proc print data=Nandini.Sales (obs=20);run;

/*Descriptive Analysis Sales Vs Agesegment*/
proc means Data=Nandini.sales n nmiss var std cv clm mean sum min Q1 Q3 qrange max ;
var Sales;
class Agesegment;
run;
```

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.246737	Pr > D	<0.0100
Cramer-von Mises	W-Sq	147.2475	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	779.516	Pr > A-Sq	<0.0050

Age <20-P value <0.05

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

Age 21-40-P value <0.05

```
/*Normality test Sales Vs Agesegment*/
proc univariate Data=Nandini.Sales normal plot;
var Sales;
Class Agesegment;
qqplot /normal (mu=est sigma=est);
run;
```

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.250593	Pr > D	<0.0100
Cramer-von Mises	W-Sq	783.9127	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4130.056	Pr > A-Sq	<0.0050

Age 41-60-P value <0.05

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.24986	Pr > D	<0.0100
Cramer-von Mises	W-Sq	490.1939	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2586.639	Pr > A-Sq	<0.0050

Age 60&above-P value <0.05

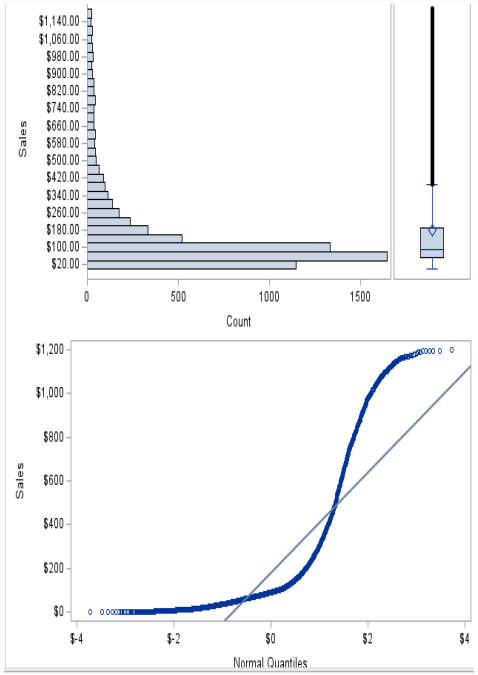
Analysis Variable : Sales

Agesegment	N Obs	N	N Miss	Variance	Std Dev	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	Mean	Sum	Minimum	Lower Quartile	Upper Quartile	Quartile Range	Maximum
<20	7137	6514	623	52866.53	229.93	129.33	172.20	183.37	177.78	1158080.00	0.00	52.00	188.00	136.00	1198.00
21-40	26382	24146	2236	55499.42	235.58	129.12	179.49	185.43	182.46	4405602.00	0.00	52.00	194.00	142.00	1200.00
41-60	37478	34385	3093	54735.26	233.96	129.04	178.83	183.78	181.30	6234135.00	0.00	53.00	189.00	136.00	1200.00
60 and above	23550	21564	1986	54277.03	232.97	129.03	177.45	183.67	180.56	3893685.00	0.00	52.00	190.00	138.00	1200.00

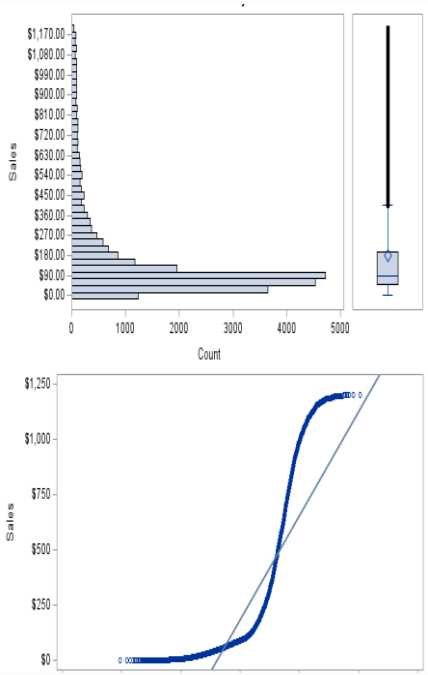
- Sales is not normally distributed across
- Any age group
- However as per CLT
- We assume Sales is
- Normally distributed

Statistical Analysis

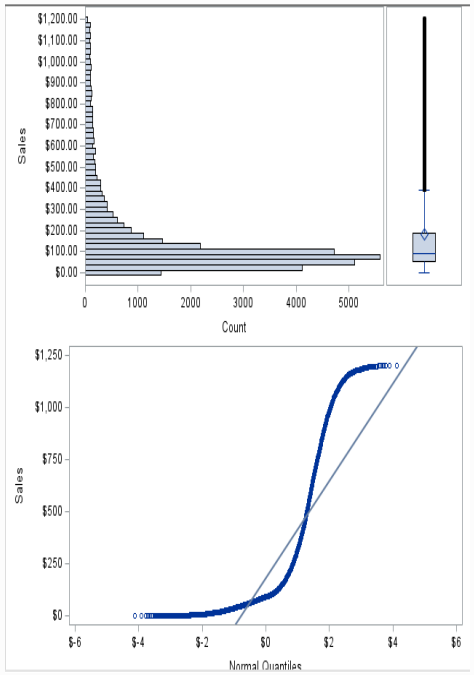
Sales Distribution Across Age segments



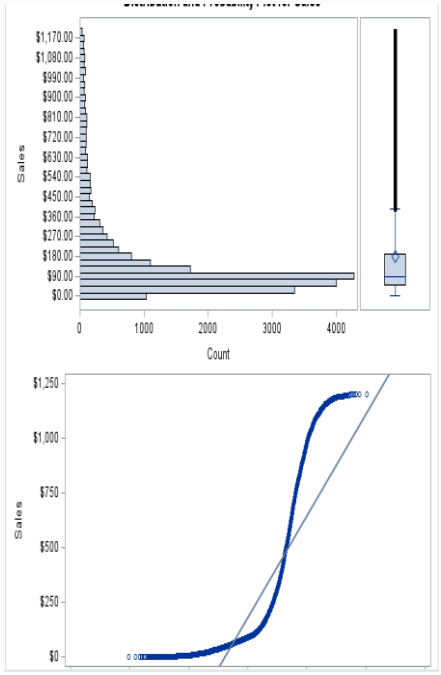
Age <20-P value <0.05



Age 21-40-P value <0.05



Age 41-60-P value <0.05



Age 60&above-P value <0.05

- No bell shape visible -Data is highly skewed-Sales not normally distributed in any age group

Statistical Analysis

Sales Distribution Across Age segments

```
/*Equality of variance Sales Vs Agesegement*/
proc glm data=Nandini.Sales;
class Agesegment;
model Sales = Agesegment;
means Agesegment / hovtest=levene(type=abs) welch;
run;
```

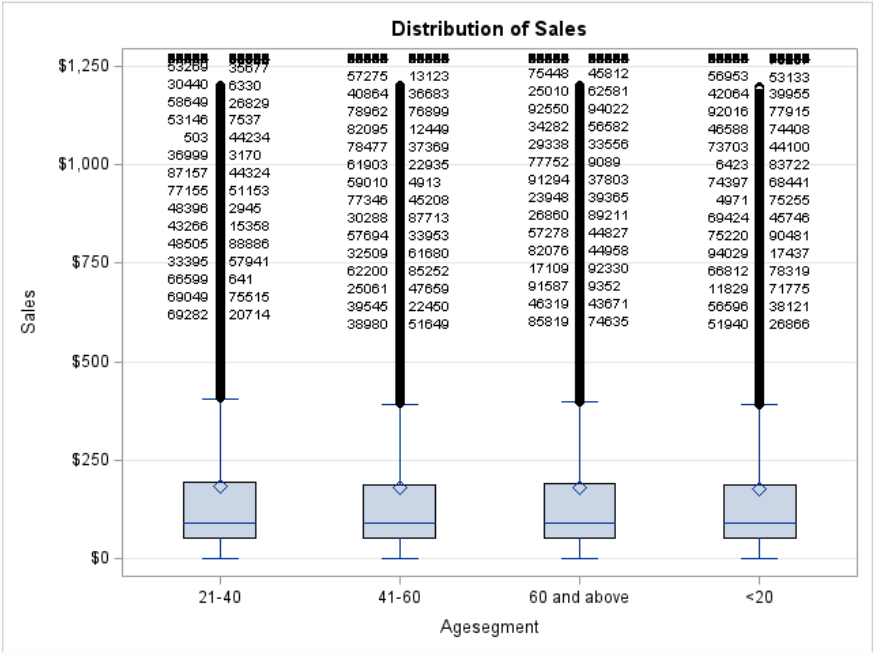
The GLM Procedure

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
Agesegment	3	214477	71492.5	2.42	0.0638
Error	86605	2.5549E9	29500.1		

Welch's ANOVA for Sales

Source	DF	F Value	Pr > F
Agesegment	3.0000	0.76	0.5143
Error	26457.8		



- P value for Homocedacity is >0.05-We fail to Reject Ho
- Sales in All age group have equal variance.
- Presence of extreme outliers in all age groups.(Higher than upper outer fence Q3+3IQR)

Statistical Analysis

Sales Distribution Across Age segments

```
TITLE "Sales distribution across Age Segements";
PROC ANOVA DATA = Nandini.Sales;
 CLASS Agesegment;
 MODEL Sales = Agesegment;
 MEANS Agesegment/scheffe;
RUN;
QUIT;
title;
```

Sales distribution across Age Segements

The ANOVA Procedure

Dependent Variable: Sales

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	123228	41076	0.75	0.5216
Error	86605	4736746122	54694		
Corrected Total	86608	4736869350			

R-Square	Coeff Var	Root MSE	Sales Mean
0.000026	129.0824	233.8668	181.1763

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Agesegment	3	123227.8921	41075.9640	0.75	0.5216

Comparisons significant at the 0.05 level are indicated by ***.

Agesegment Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
21-40 - 41-60	1.153	-4.336	6.642
21-40 - 60 and above	1.893	-4.233	8.018
21-40 - <20	4.674	-4.454	13.802
41-60 - 21-40	-1.153	-6.642	4.336
41-60 - 60 and above	0.740	-4.939	6.419
41-60 - <20	3.521	-5.314	12.355
60 and above - 21-40	-1.893	-8.018	4.233
60 and above - 41-60	-0.740	-6.419	4.939
60 and above - <20	2.781	-6.462	12.024
<20 - 21-40	-4.674	-13.802	4.454
<20 - 41-60	-3.521	-12.355	5.314
<20 - 60 and above	-2.781	-12.024	6.462

- P value for 0.516 is >0.05-We fail to Reject Ho
- Means of Sales is equal in all Age groups
- Sales is equal in All age group

Statistical Analysis

Sales Distribution for Active and Deactivated Customers

```
/*Descriptive Analysis Sales Vs Status*/
proc means Data=Nandini.sales n nmiss var std cv clm mean sum min Q1 Q3 qrange max maxdec=2 ;
var Sales;
class Status;
run;
```

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.250605	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1722.213	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	9074.765	Pr > A-Sq	<0.0050

Status-Active

```
/*Normality test Sales Vs Status*/
proc univariate Data=Nandini.Sales normal plot;
var Sales;
Class Status;
qqplot /normal (mu=est sigma=est);
run;
```

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

Status- Deactivated

```
/*Total sales ineach account status*/
proc sql;
select sum(sales)as Total_Sales,acct_status
from Nandini.Status_Tenure
group by Acct_Status;
quit;
```

Total_Sales	Acct_Status
13739549	Active
3234154	Deactivated

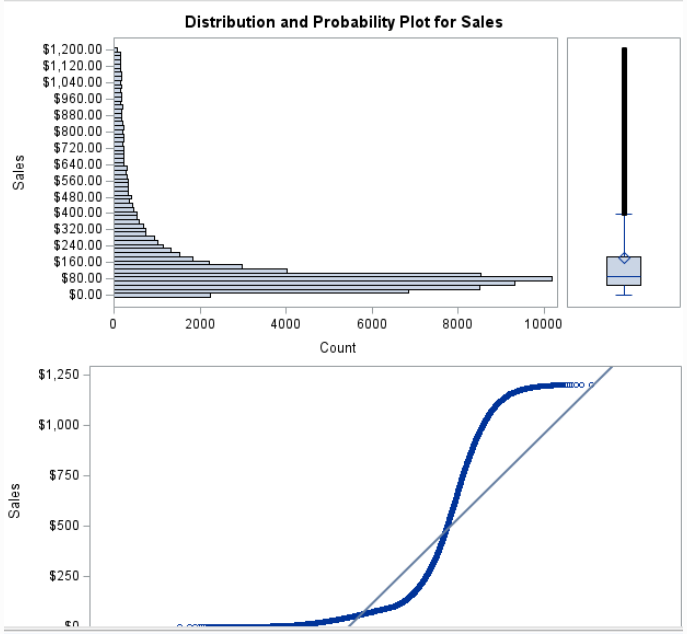
- Sales is not normally distributed across Any account status
- However as per CLT
- We assume Sales is
- Normally distributed

Analysis Variable : Sales														
Status	N Obs	N	N Miss	Variance	Std Dev	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	Mean	Sum	Minimum	Lower Quartile	Upper Quartile	Quartile Range	Maximum
Active	82620	75675	6945	54986.11	234.49	129.15	179.89	183.23	181.56	13739549.00	0.00	52.00	191.00	139.00	1200.00
Deactivated	19635	17975	1660	53717.47	231.77	128.81	176.54	183.31	179.93	3234154.00	0.00	53.00	188.00	135.00	1199.00

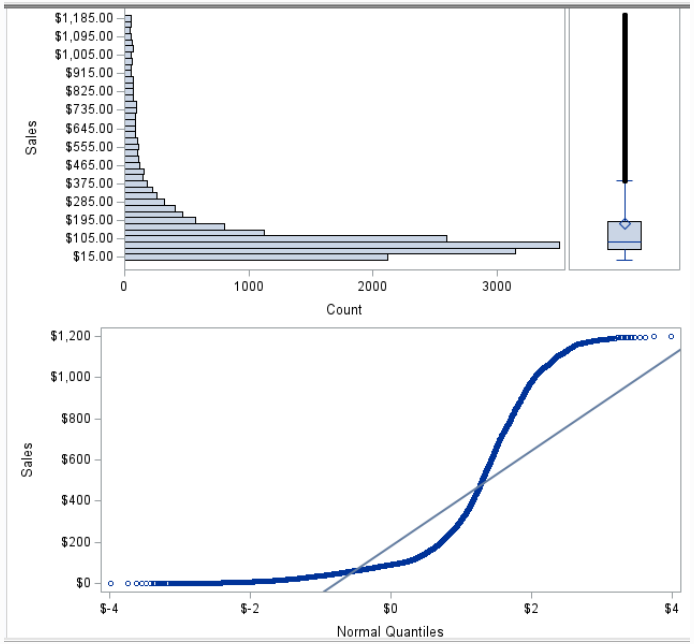
Statistical Analysis

Sales Distribution for Active and Deactivated Customers

Status-Active



Status- Deactivated



- No bell shape visible
- Data is highly skewed
- Sales not normally distributed in any Account status
- Presence of extreme outliers

Statistical Analysis

Sales Distribution for Active and Deactivated Customers

```
/*Equality of variance Sales Vs Status*/
proc glm data=Nandini.Sales;
class Status;
model Sales = Status;
means Status / hovtest=levene(type=abs) welch;
run;
```

The GLM Procedure

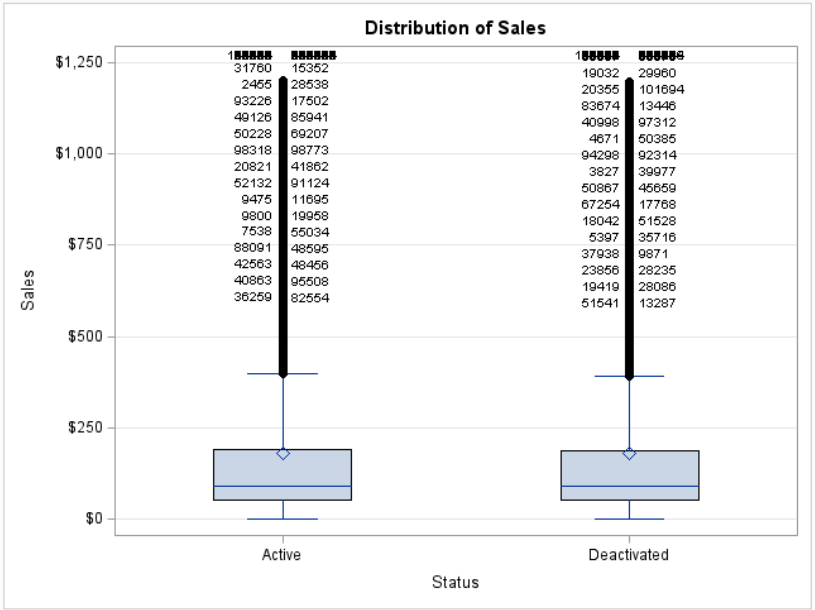
Levene's Test for Homogeneity of Sales Variance
ANOVA of Absolute Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Status	1	112914	112914	3.83	0.0505
Error	93648	2.7641E9	29516.2		

Welch's ANOVA for Sales

Source	DF	F Value	Pr > F
Status	1.0000	0.72	0.3963
Error	27392.3		

The GLM Procedure



- P value for Homocedacity is >0.05-We fail to Reject Ho
- Sales in both account status have equal variance.
- Presence of extreme outliers both account status.(Higher than upper outer fence Q3+3IQR)

Statistical Analysis

Sales Distribution for Active and Deactivated Customers

```
proc ttest Data=Nandini.Sales;
Var Sales;
Class status;
run;
```

The TTEST Procedure

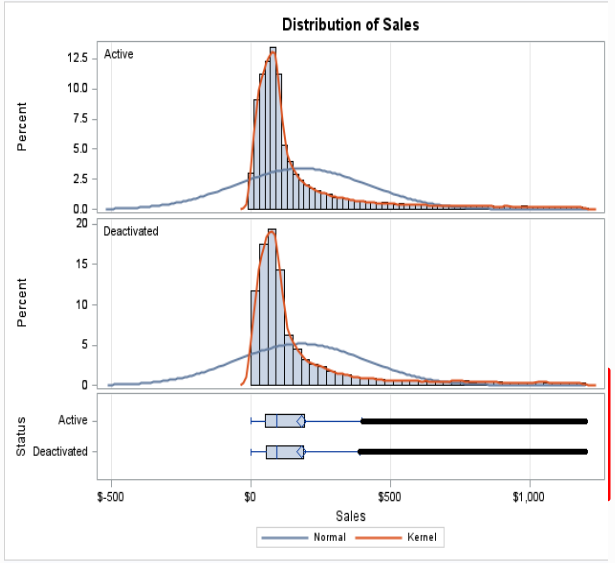
Variable: Sales

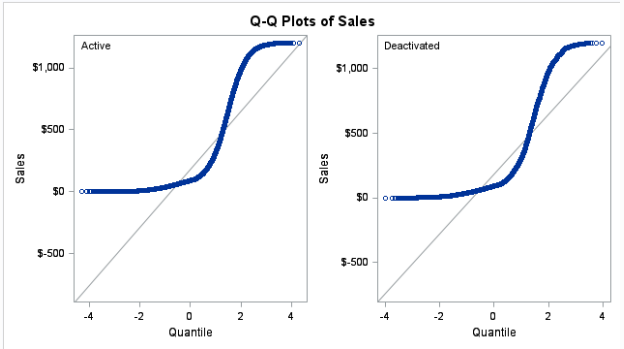
Status	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Active		75675	181.6	234.5	0.8524	0	1200.0
Deactivated		17975	179.9	231.8	1.7287	0	1199.0
Diff (1-2)	Pooled		1.6348	234.0	1.9414		
Diff (1-2)	Satterthwaite		1.6348		1.9274		

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





- P value for folded F 0.0475 is <0.05-We Reject H0 for Homocedacity
- Sathherwaite P value 0.3963 >0.05- we fail to reject H0
- Means of Sales is equal for both Active and Deactivated Customers

Statistical Analysis

```
/*Descriptive Analysis Sales Vs Goodcredit*/
proc means Data=Nandini.sales n nmiss var std cv clm mean sum min Q1 Q3 qrange max maxdec=2 ;
var Sales;
class Goodcredit;
run;
/*Normality Test Sales Vs Agesegment*/
proc univariate Data=Nandini.Sales normal plot;
var Sales;
Class Goodcredit;
qqplot /normal (mu=est sigma=est);
run;
/*Total Sales classified between Good credit categories*/
proc sql;
select sum(sales )as Total_Sales_credit,Goodcredit
from Nandini.Sales
group by goodcredit;
quit;
```

Sales Distribution for Good Credit

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.249606	Pr > D	<0.0100
Cramer-von Mises	W-Sq	652.0233	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3437.394	Pr > A-Sq	<0.0050

Good Credit -0-NO

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.25031	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

Good Credit -1-Yes

Total_Sales_credit	GoodCredit
5192720	0
11780983	1

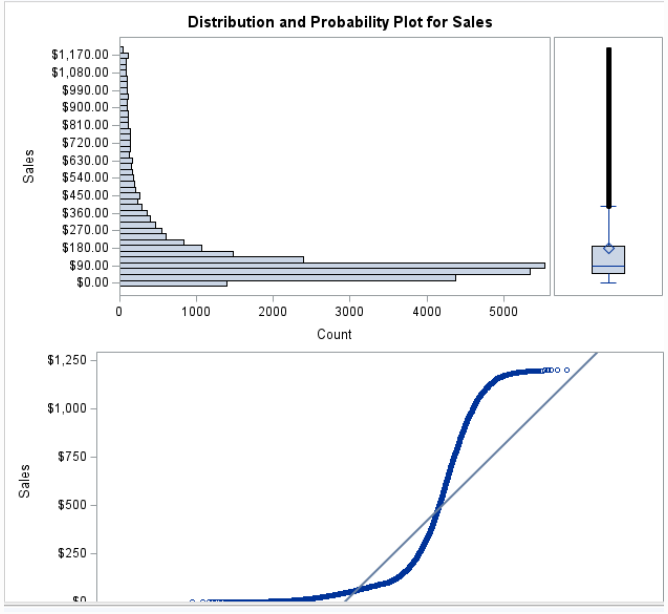
- Sales is not normally distributed across Any Good Credit Category
- However as per CLT We assume Sales is Normally distributed

Analysis Variable : Sales														
GoodCredit	N Obs	N	N Miss	Variance	Std Dev	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	Mean	Sum	Minimum	Lower Quartile	Upper Quartile	Quartile Range	Maximum
0	31253	28599	2654	55148.62	234.84	129.34	178.85	184.29	181.57	5192720.00	0.00	52.00	190.00	138.00	1200.00
1	71002	65051	5951	54564.66	233.59	128.98	179.31	182.90	181.10	11780983.00	0.00	53.00	190.00	137.00	1200.00

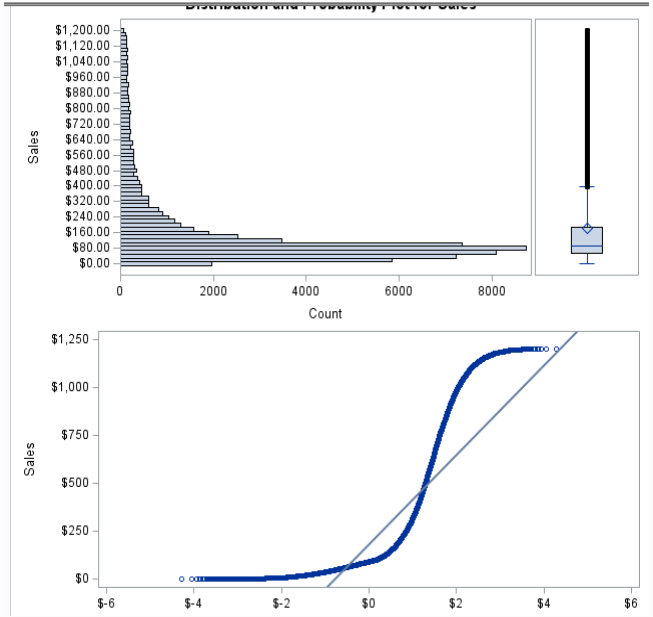
Statistical Analysis

Sales Distribution for Good Credit

No Good Credit



Has Good Credit



- No bell shape visible
- Data is highly skewed
- Sales not normally distributed in Good Credit category
- Presence of extreme outliers

Statistical Analysis

Sales Distribution for Good Credit

```
/*Equality of variance*/
proc glm data=Nandini.Sales;
class Goodcredit;
model Sales = Goodcredit;
means Goodcredit / hovtest=levene(type=abs) welch;
run;
```

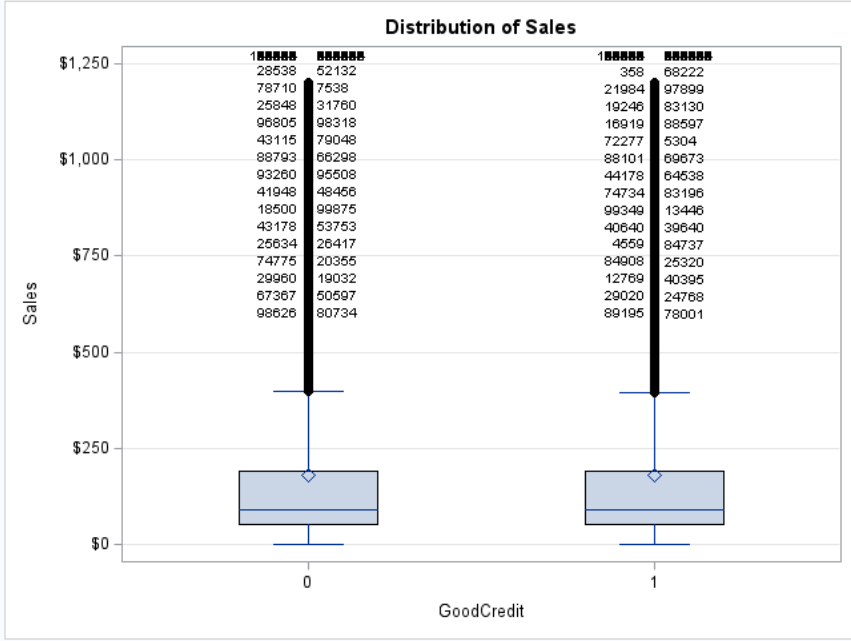
The GLM Procedure

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
GoodCredit	1	5039.1	5039.1	0.17	0.6795
Error	93648	2.7643E9	29517.5		

Welch's ANOVA for Sales

Source	DF	F Value	Pr > F
GoodCredit	1.0000	0.08	0.7793
Error	54366.1		



- P value for Homocedacity is >0.05-We fail to Reject H0
- Sales in both Good Credit Category have equal variance.
- Presence of extreme outliers both account status.(Higher than upper outer fence Q3+3IQR)

Statistical Analysis

Sales Distribution for Good Credit

```
proc ttest Data=Nandini.Sales;
Var Sales;
Class Goodcredit;
run;
```

The TTEST Procedure

Variable: Sales

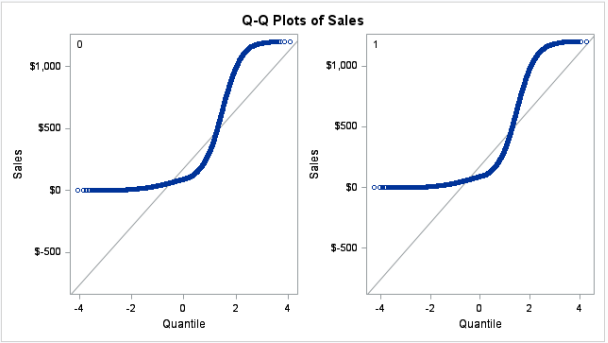
GoodCredit	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		28599	181.6	234.8	1.3886	0	1200.0
1		65051	181.1	233.6	0.9159	0	1200.0
Diff (1-2)	Pooled		0.4662	234.0	1.6600		
Diff (1-2)	Satterthwaite		0.4662		1.6635		

GoodCredit	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		181.6	178.8	184.3	234.8	232.9	236.8
1		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	94388	0.28	0.7795

Equality of Variances

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



- P value for folded F 0.2878 is <0.05-We fail to Reject H0 for Homocedacity. Means Sales have equal variances in both good credit categories
- Pooled P value 0.7788 >0.05- we fail to reject H0
- Mean of Sales is equal for both Good Credit categories
- Sales is equal in both categories

Statistical Analysis

Sales Vs Tenure Correlation

```
proc surveyselect data=Nandini.Status_Tenure out=Nandini.Salecorr method=srs n=100;
run;
proc print data=Nandini.Salecorr;run;
```

```
proc corr Data=Nandini.Salecorr;
var Tenuredays;
with Sales;
run;
proc reg data=Nandini.Salecorr;
model Sales= Tenuredays;
run;
```

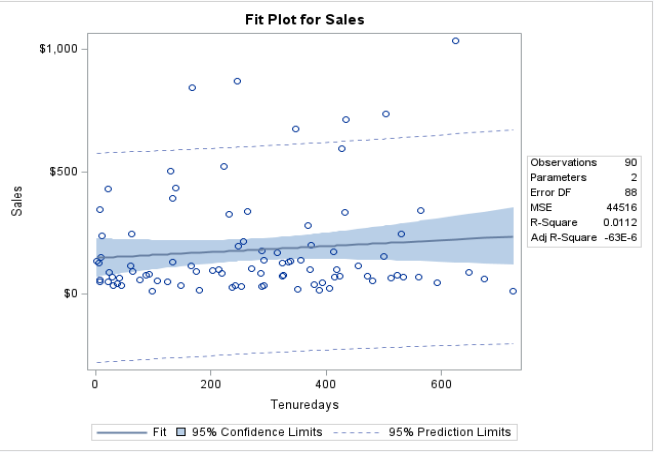
1 With Variables:	Sales
1 Variables:	Tenuredays

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Sales	90	181.20000	210.98156	16308	14.00000	1033
Tenuredays	100	267.34000	191.29126	26734	1.00000	724.00000

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Tenuredays
Sales	0.10570 0.3214 90



- R square is very low and RMSE is high
- Very week correlation found between Sale and Tenure days
- Model is not a best fit for the analysis between Sales and Age

The REG Procedure
Model: MODEL1
Dependent Variable: Sales

Number of Observations Read	100
Number of Observations Used	90
Number of Observations with Missing Values	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	44265	44265	0.99	0.3214
Error	88	3917411	44516		
Corrected Total	89	3961676			

Root MSE	210.98823	R-Square	0.0112
Dependent Mean	181.20000	Adj R-Sq	-0.0001
Coeff Var	116.43942		

Parameter Estimates

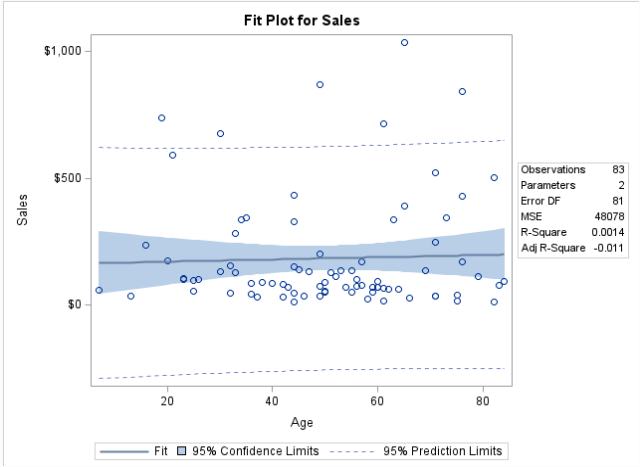
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	148.56703	39.56722	3.75	0.0003
Tenuredays	1	0.11972	0.12006	1.00	0.3214

Statistical Analysis

Sales Vs Age Correlation

```
proc surveyselect data=Nandini.Status_Tenure out=Nandini.Salecorr method=srs n=100;
run;
proc print data=Nandini.Salecorr;run;

proc corr Data=Nandini.Salecorr;
 var Age;
 with Sales;
 run;
proc reg data=Nandini.Salecorr;
 model Sales= Age ;
 run;
```



The CORR Procedure

1 With Variables:	Sales
1 Variables:	Age

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Sales	90	181.20000	210.98156	16308	14.00000	1033
Age	93	49.49462	18.39519	4603	7.00000	84.00000

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Age
Sales	0.03690 0.7405 83

- R square is very low and RMSE is high
- Very week correlation found between Sale and Age
- Model is not a best fit for the analysis between Sales and Age

The REG Procedure
Model: MODEL1
Dependent Variable: Sales

Number of Observations Read	100
Number of Observations Used	83
Number of Observations with Missing Values	17

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5308.68060	5308.68060	0.11	0.7405
Error	81	3894345	48078		
Corrected Total	82	3899653			

Root MSE	219.26771	R-Square	0.0014
Dependent Mean	185.24096	Adj R-Sq	-0.0110
Coeff Var	118.36891		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	163.11626	70.79875	2.30	0.0238
Age	1	0.43942	1.32241	0.33	0.7405

Statistical Analysis

Sales –Tenure-Age

```
proc sql;
 select sales,tenuredays
 from Nandini.Status_Tenure
 where Sales>900 and tenuredays<10;
 quit;
```

```
proc sql;
 select sales,Age
 from Nandini.Status_Tenure
 where Sales>900 and Age<5 and age ne .;
 quit;
```

```
proc sql;
 select sales,tenuredays
 from Nandini.Status_Tenure
 where Sales<100 and tenuredays>300;
 quit;
```

- Sale is missing or very low for tenure more than 1 year
- Sale is very high for even 1 days
- Sale is very high for age of 1 year
- Most places data looks unrealistic
- More accurate information needed to make this model perfect

Sales	Tenuredays
$11.00	408
$16.00	459
$44.00	630
$97.00	340
.	467
$71.00	440
$67.00	550
$76.00	407
$24.00	304
$33.00	575
$74.00	486
$27.00	449
$23.00	453
.	541
$11.00	325
$50.00	589
$90.00	540
$60.00	730
.	434
$92.00	407

Sales	Tenuredays
$965.00	6
$1,096.00	3
$951.00	3
$1,164.00	1
$1,135.00	8
$1,188.00	8
$956.00	2
$1,064.00	3
$1,026.00	4
$1,186.00	4
$917.00	4
$962.00	7
$1,072.00	7
$906.00	9
$1,053.00	9
$932.00	8

Sales	Age
$917.00	2
$945.00	4
$947.00	2
$1,023.00	1
$1,098.00	1
$1,057.00	2
$908.00	4
$1,114.00	2
$1,095.00	2
$913.00	2
$1,082.00	4
$973.00	4
$1,155.00	4
$1,123.00	1
$1,193.00	2
$1,047.00	1
$1,066.00	4
$927.00	2

Statistical Analysis

Sales –Tenure-Age

```
proc corr Data=Nandini.Status_Tenure pearson spearman kendall
plots(maxpoints=none) = matrix(histogram);
var Sales Age Tenuredays;
run;
```

3 Variables:	Sales Age Tenuredays

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
Sales	93650	181.24616	233.97104	91.00000	0	1200
Age	94547	47.64722	18.56900	48.00000	1.00000	110.00000
Tenuredays	102255	282.57180	197.32371	265.00000	0	731.00000

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Sales	Age	Tenuredays
Sales	1.00000 93650	0.00147 0.6656 86609	-0.00391 0.2310 93650
Age	0.00147 0.6656 86609	1.00000 94547	-0.00329 0.3123 94547
Tenuredays	-0.00391 0.2310 93650	-0.00329 0.3123 94547	1.00000 102255

Spearman Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Sales	Age	Tenuredays
Sales	1.00000 93650	0.00195 0.5668 86609	-0.00040 0.9022 93650
Age	0.00195 0.5668 86609	1.00000 94547	-0.00259 0.4261 94547
Tenuredays	-0.00040 0.9022 93650	-0.00259 0.4261 94547	1.00000 102255

Kendall Tau b Correlation Coefficients
Prob > |tau| under H0: Tau=0
Number of Observations

	Sales	Age	Tenuredays
Sales	1.00000 93650	0.00130 0.5685 86609	-0.00028 0.8998 93650
Age	0.00130 0.5685 86609	1.00000 94547	-0.00174 0.4249 94547
Tenuredays	-0.00028 0.8998 93650	-0.00174 0.4249 94547	1.00000 102255

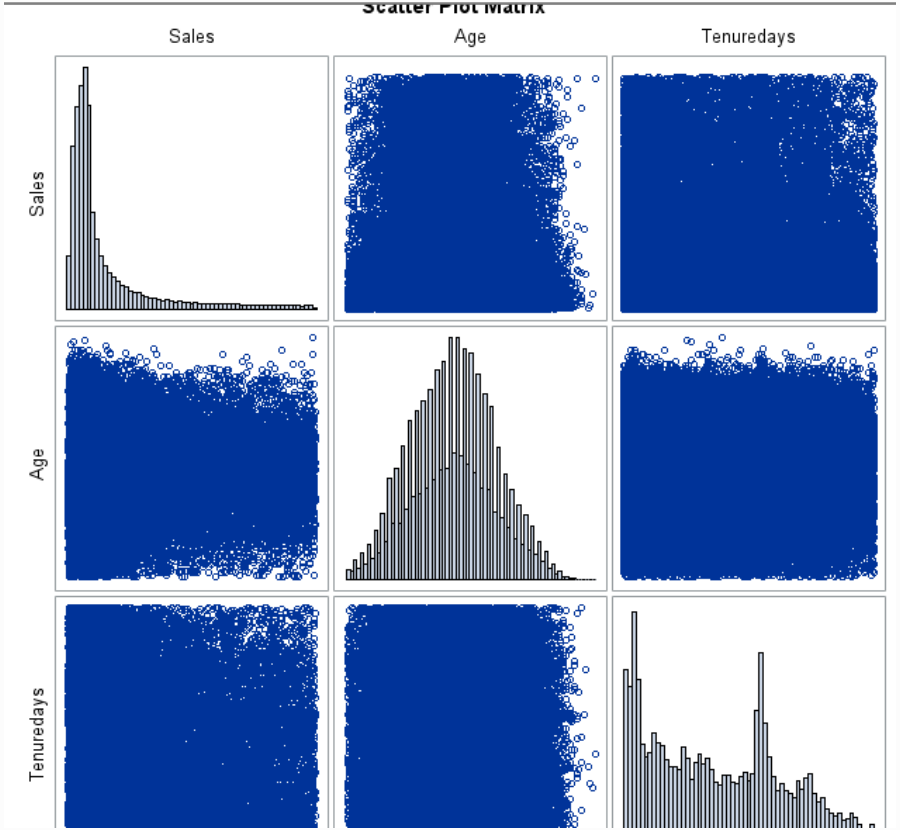
- No correlation found between Sales Age and Tenure

Statistical Analysis

Sales –Tenure-Age

```
proc corr Data=Nandini.Status_Tenure pearson spearman kendall
plots(maxpoints=none) = matrix(histogram);
var Sales Age Tenuredays;
run;
```

- No Pattern or linear correlation found between any of 3 independent variables.



Summary

Findings

- Age and Sales has Majority of Outliers, Missing Data, Unrealistic figures
- ON has maximum Customer Base where as QC has Minimum
- Count of Age<20 (Young Customers) is very low
- 52% customers has taken service of less than $100.Only 4 % customers have taken service of more than $800.
- Maximum Deactivation Occurred in Winters. Reason can be interrupted service, outage due to harsh weather
- 340 customers have deactivated service on the same day
- Customers with greater Tenure have good credit
- Majority active customers prefer Rate plan 1 .Rate plan 2 has minimum customer churn. Rate Plan 3 has maximum customer churn
- Dealer type A1 has maximum customer base.
- Customer churn is seen more with tenure <30 days.(33%)
- Customer churn went increasing after March till year end .
- Sales is Maximum in age group 41-60 and minimum in <20.It indicates young people tend to prefer better deals by competitors
- Sales is maximum for active customers than Deactivated Customers.
- Maximum Sales is from the customers with Good Credit
- Model is not best fit to explain sales with the help of Tenure and Age data.

Recommendations

- Need More accurate information for effective analysis.
- More promotions needed in QC
- Good Promotions, offers, deals need to be arranged to attract Young crowd
- Need to reach out to customer for their feedback and understand their need. Offer better solutions. Initiate Rewards for loyal customers.
- Need to find way outs to avoid interruption in winters.
- Deactivation on same day is a major point of concern.Need investigation
- Arrange Loyalty rewards for Good Credit Customers
- Rate Plan 2 can be improved to match rate plan 1.Rate plan 3 needs attention.
- Other Dealers need to match the service of Dealer A1
- Provide best service to avoid losing customers in first month.
- Competitor analysis required .
- Age and Sales figures need to be rechecked .
- Overall the available data is not sufficient to make strong conclusions regarding sale and customer churn. Details like customer income, customer feedback may add value to the analysis.

Thank You