Program Summary - Insurance Assignment2.sas

Execution Environment

Author:	u61856037
File:	/home/u61856037/sasuser.v94/Insurance Assignment2.sas
SAS Platform:	Linux LIN X64 3.10.0-1062.12.1.el7.x86_64
SAS Host:	ODAWS02-APSE1-2.ODA.SAS.COM
SAS Version:	9.04.01M6P11072018
SAS Locale:	en_US
Submission Time:	7/17/2022, 12:32:17 PM
Browser Host:	157.47.48.2
User Agent:	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/103.0.0.0 Safari/537.36
Application Server:	ODAMID00-APSE1-2.ODA.SAS.COM

Code: Insurance Assignment2.sas

```
/*-------------------------------------------------------------------------------------------------------------------*/
/*-------------------------------------------------SAS GRADED PROJECT------------------------------------------------*/
/*-------------------------------------------------DOMAIN: INSURANCE-------------------------------------------------*/
/*-------------------------------------------------NAME: PANAGAM MOHITHA---------------------------------------------*/
/*-------------------------------------------------------------------------------------------------------------------*/
/*Question1: Import dataset in the SAS environment and check top 10 record of import dataset*/

FILENAME REFFILE '/home/u61856037/sasuser.v94/Life+Insurance+Dataset.csv';
PROC IMPORT DATAFILE=REFFILE
    DBMS=CSV
    OUT=Life_Insurance_data;
    GETNAMES=YES;
RUN;
PROC CONTENTS DATA=Life_Insurance_data; RUN;


/*Question2: Check variable type of the import dataset*/
proc contents data=Life_Insurance_data varnum;
run;

/*Question3: Checks if any variables have missing values, if yes then do treatment?*/
proc means data=Life_Insurance_data nmiss;
run;

/*Question4: Check summary and percentile distribution of all numerical variables
for churners and non-churners?*/
proc means data=Life_Insurance_data n nmiss min p1 p5 p10 p25 p50 p75 p90 p95 p99 max maxdec=0;
var Age Cust_Tenure Overall_cust_satisfation_score CC_Satisfation_score Cust_Income Agent_Tenure
YTD_contact_cnt Due_date_day_cnt Existing_policy_count Miss_due_date_cnt ;
run;

/*Question5: Check for outlier, if yes then do treatment?*/
proc univariate data=Life_Insurance_data;
var Age Cust_Tenure Overall_cust_satisfation_score CC_Satisfation_score Cust_Income Agent_Tenure
YTD_contact_cnt Due_date_day_cnt Existing_policy_count Miss_due_date_cnt;
run;
/*we have some outlier and below is the flooring and cappping for those variables*/
data Life_Insurance_data;
set Life_Insurance_data;
if Cust_Income > 35999 then Cust_Income = 35999;
run;
/*checking distribution after flooring and capping*/
proc univariate data=Life_Insurance_data;
var Age Cust_Tenure Overall_cust_satisfation_score CC_Satisfation_score Cust_Income Agent_Tenure
YTD_contact_cnt Due_date_day_cnt Existing_policy_count Miss_due_date_cnt;
run;

/*Question6: Check the proportion of all categorical variables and
extract percentage contribution of each class in respective variables?*/
proc freq data=Life_Insurance_data;
table Payment_Period Product EducationField Gender Cust_Designation Cust_MaritalStatus Complaint/ nocum;
run;


/*Question7: Customer service management want you to create a macro where they will just put mobile number
and they will get all the important information like Age, Education, Gender, Income and CustID*/
/*Created Marcro*/
%MACRO Customer_info();
DATA output (keep = Age EducationField Gender Cust_Income CustID);
SET Life_Insurance_data;
where Mobile_num in (&Mobile_num.);
```

```
RUN;
proc print data=output;
run;
%MEND;
/*Provided input mobile number*/
%let Mobile_num = 9878913773,9898819662,9904978124,9887638137,9882200862;
/*run macro for output*/
%Customer_info;


/*Question8: Check correlation of all numerical variables before building model,
because we cannot add correlated variables in model?*/
proc corr data=Life_Insurance_data NOPROB;
var Age Cust_Tenure Overall_cust_satisfation_score CC_Satisfation_score Cust_Income Agent_Tenure
YTD_contact_cnt Due_date_day_cnt Existing_policy_count Miss_due_date_cnt ;
run;


/*Question9: Create train and test (70:30) dataset from the existing data set. Put seed 1234?*/
proc freq data=Life_Insurance_data;
table Churn /nocum;
run;
proc surveyselect data= Life_Insurance_data method = srs rep=1
sampsize=600 seed = 1234 out =test;
RUN;
proc contents data=test varnum;
run;
proc freq data=test;
table Churn /nocum;
run;
proc sql;
create table train as select t1.* from Life_Insurance_data as t1
where CustID not in (select CustID from test);
quit;
proc freq data=train;
table Churn /nocum;
run;


/*Question10: Develop linear regression model first on the target variable
to extract VIF information to check multicollinearity?*/
proc reg data=Life_Insurance_data;
model Churn=Age Cust_Tenure Overall_cust_satisfation_score CC_Satisfation_score Cust_Income Agent_Tenure
YTD_contact_cnt Due_date_day_cnt Existing_policy_count Miss_due_date_cnt / vif tol collin;
title 'Life_Insurance_data - Multicollinearity Investigation of VIF';
run;
quit;


/*Question11: Create clean logistic model on the target variables?*/
%let var = Age Cust_Tenure Overall_cust_satisfation_score CC_Satisfation_score Cust_Income Agent_Tenure
YTD_contact_cnt Due_date_day_cnt Existing_policy_count Miss_due_date_cnt ;
proc logistic data=train descending outmodel=model;
model Churn = &var / lackfit;
output out = train_output xbeta = coeff stdxbeta = stdcoeff predicted = prob;
run;



/*Question12: Create a macro and take a KS approach to take a cut off on the calculated scores?*/
proc univariate data=Life_Insurance_data;
var Overall_cust_satisfation_score;
   histogram Overall_cust_satisfation_score / normal(mu=est sigma=est);
run;




/*Question13: Predict test dataset using created model?*/
/*Predicting by equation, you can use score statment, in my version of SAS score function is not
present*/
proc reg data=Life_Insurance_data outest=test1;
model Churn=Age Cust_Tenure Overall_cust_satisfation_score CC_Satisfation_score Cust_Income Agent_Tenure
YTD_contact_cnt Due_date_day_cnt Existing_policy_count Miss_due_date_cnt ;
run;

proc score data=Life_Insurance_data score=test1 type=parms predict out=test2;
var Age Cust_Tenure Overall_cust_satisfation_score CC_Satisfation_score Cust_Income Agent_Tenure
YTD_contact_cnt Due_date_day_cnt Existing_policy_count Miss_due_date_cnt ;
run;
```

Question 02:

The CONTENTS Procedure

Data Set Name	WORK.LIFE_INSURANCE_DATA	**Observations**	1924
Member Type	DATA	**Variables**	20
Engine	V9	**Indexes**	0
Created	07/17/2022 10:08:49	**Observation Length**	184
Last Modified	07/17/2022 10:08:49	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	3
First Data Page	1
Max Obs per Page	711
Obs in First Data Page	687
Number of Data Set Repairs	0
Filename	/saswork/SAS_work5A1D000014AE_odaws01-apse1-2.oda.sas.com/SAS_workE9AE000014AE_odaws01-apse1-2.oda.sas.com/life_insurance_data.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	134321073
Access Permission	rw-r--r--
Owner Name	u61856037
File Size	512KB
File Size (bytes)	524288

Variables in Creation Order

#	Variable	Type	Len	Format	Informat
1	CustID	Num	8	BEST12.	BEST32.
2	Mobile_num	Num	8	BEST12.	BEST32.
3	Churn	Num	8	BEST12.	BEST32.
4	Age	Num	8	BEST12.	BEST32.
5	Payment_Period	Char	9	$9.	$9.
6	Product	Char	14	$14.	$14.
7	Cust_Tenure	Num	8	BEST12.	BEST32.
8	EducationField	Char	17	$17.	$17.
9	Gender	Char	6	$6.	$6.
10	Overall_cust_satisfation_score	Num	8	BEST12.	BEST32.
11	Cust_Designation	Char	14	$14.	$14.
12	CC_Satisfation_score	Num	8	BEST12.	BEST32.
13	Cust_MaritalStatus	Char	8	$8.	$8.
14	Cust_Income	Num	8	BEST12.	BEST32.
15	Agent_Tenure	Num	8	BEST12.	BEST32.
16	Complaint	Num	8	BEST12.	BEST32.
17	YTD_contact_cnt	Num	8	BEST12.	BEST32.
18	Due_date_day_cnt	Num	8	BEST12.	BEST32.
19	Existing_policy_count	Num	8	BEST12.	BEST32.
20	Miss_due_date_cnt	Num	8	BEST12.	BEST32.

Question 03:

The MEANS Procedure

Variable	N Miss
CustID	0
Mobile_num	0
Churn	0
Age	0
Cust_Tenure	0
Overall_cust_satisfation_score	0
CC_Satisfation_score	0
Cust_Income	0
Agent_Tenure	0
Complaint	0
YTD_contact_cnt	0
Due_date_day_cnt	0
Existing_policy_count	0
Miss_due_date_cnt	0

Question 04:

The MEANS Procedure

Variable	N	N Miss	Minimum	1st Pctl	5th Pctl	10th Pctl	25th Pctl	50th Pctl	75th Pctl	90th Pctl	95th Pctl	99th Pctl	Maximum
Age	1924	0	21	22	27	30	34	42	51	57	59	60	60
Cust_Tenure	1924	0	1	1	3	4	7	12	19	23	24	25	25
Overall_cust_satisfation_score	1924	0	1	1	2	2	2	3	4	5	5	5	5
CC_Satisfation_score	1924	0	1	1	1	1	2	3	4	5	5	5	5
Cust_Income	1924	0	16009	17001	17296	17619	18556	20392	23768	29582	33159	35331	96000
Agent_Tenure	1924	0	0	0	0	1	1	2	5	7	8	10	10
YTD_contact_cnt	1924	0	16	16	16	17	18	20	23	27	28	30	31
Due_date_day_cnt	1924	0	0	1	2	3	6	10	16	23	28	34	41
Existing_policy_count	1924	0	1	1	1	2	4	8	12	14	15	15	15
Miss_due_date_cnt	1924	0	0	0	0	0	0	1	2	5	8	10	10

Question 05:

The UNIVARIATE Procedure
Variable: Age

Moments			
N	1924	Sum Weights	1924
Mean	42.6242204	Sum Observations	82009
Std Deviation	10.0113121	Variance	100.226371
Skewness	0.00576962	Kurtosis	-0.9543936
Uncorrected SS	3688305	Corrected SS	192735.311
Coeff Variation	23.4873789	Std Error Mean	0.22823827

Basic Statistical Measures			
Location		Variability	
Mean	42.62422	Std Deviation	10.01131
Median	42.00000	Variance	100.22637
Mode	38.00000	Range	39.00000
		Interquartile Range	17.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	186.7532	Pr > |t|	<.0001
Sign	M	962	Pr >= |M|	<.0001
Signed Rank	S	925925	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	60
99%	60
95%	59
90%	57
75% Q3	51
50% Median	42
25% Q1	34
10%	30
5%	27
1%	22
0% Min	21

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
21	1764	60	1777
21	1712	60	1823
21	1648	60	1877
21	1637	60	1903
21	1596	60	1905

The UNIVARIATE Procedure
Variable: Cust_Tenure

Moments			
N	1924	Sum Weights	1924
Mean	12.6486486	Sum Observations	24336
Std Deviation	7.01534187	Variance	49.2150216
Skewness	0.18921756	Kurtosis	-1.2031407
Uncorrected SS	402458	Corrected SS	94640.4865
Coeff Variation	55.4631729	Std Error Mean	0.15993603

Basic Statistical Measures			
Location		Variability	
Mean	12.64865	Std Deviation	7.01534
Median	12.00000	Variance	49.21502
Mode	8.00000	Range	24.00000
		Interquartile Range	12.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	79.08567	Pr > |t|	<.0001
Sign	M	962	Pr >= |M|	<.0001
Signed Rank	S	925925	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	25
99%	25

Quantiles (Definition 5)	
Level	Quantile
95%	24
90%	23
75% Q3	19
50% Median	12
25% Q1	7
10%	4
5%	3
1%	1
0% Min	1

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	1921	25	1820
1	1882	25	1837
1	1853	25	1886
1	1834	25	1903
1	1814	25	1905

The UNIVARIATE Procedure
Variable: Overall_cust_satisfation_score

Moments			
N	1924	Sum Weights	1924
Mean	3.39553015	Sum Observations	6533
Std Deviation	1.18053232	Variance	1.39365656
Skewness	-0.1158404	Kurtosis	-1.1124382
Uncorrected SS	24863	Corrected SS	2680.00156
Coeff Variation	34.767246	Std Error Mean	0.02691382

Basic Statistical Measures			
Location		Variability	
Mean	3.395530	Std Deviation	1.18053
Median	3.000000	Variance	1.39366
Mode	4.000000	Range	4.00000

Basic Statistical Measures			
Location		Variability	
		Interquartile Range	2.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	126.1631	Pr > |t|	<.0001
Sign	M	962	Pr >= |M|	<.0001
Signed Rank	S	925925	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	5
99%	5
95%	5
90%	5
75% Q3	4
50% Median	3
25% Q1	2
10%	2
5%	2
1%	1
0% Min	1

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	1888	5	1911
1	1885	5	1915
1	1874	5	1918
1	1856	5	1919
1	1788	5	1920

The UNIVARIATE Procedure
Variable: CC_Satisfation_score

Moments

Moments			
N	1924	Sum Weights	1924
Mean	3.0514553	Sum Observations	5871
Std Deviation	1.36631832	Variance	1.86682575
Skewness	-0.123559	Kurtosis	-1.1147222
Uncorrected SS	21505	Corrected SS	3589.90593
Coeff Variation	44.7759572	Std Error Mean	0.03114938

Basic Statistical Measures			
Location		Variability	
Mean	3.051455	Std Deviation	1.36632
Median	3.000000	Variance	1.86683
Mode	3.000000	Range	4.00000
		Interquartile Range	2.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	97.962	Pr > |t|	<.0001
Sign	M	962	Pr >= |M|	<.0001
Signed Rank	S	925925	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	5
99%	5
95%	5
90%	5
75% Q3	4
50% Median	3
25% Q1	2
10%	1
5%	1
1%	1
0% Min	1

Extreme Observations	
Lowest	Highest

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	1923	5	1895
1	1922	5	1901
1	1916	5	1905
1	1911	5	1915
1	1910	5	1921

The UNIVARIATE Procedure
Variable: Cust_Income

Moments			
N	1924	Sum Weights	1924
Mean	21964.42	Sum Observations	42259544
Std Deviation	4729.40327	Variance	22367255.3
Skewness	1.35423264	Kurtosis	1.00394092
Uncorrected SS	9.71219E11	Corrected SS	4.30122E10
Coeff Variation	21.5321109	Std Error Mean	107.821114

Basic Statistical Measures			
Location		Variability	
Mean	21964.42	Std Deviation	4729
Median	20391.50	Variance	22367255
Mode	17177.00	Range	19990
		Interquartile Range	5212

Note: The mode displayed is the smallest of 6 modes with a count of 4.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	203.7117	Pr > |t|	<.0001
Sign	M	962	Pr >= |M|	<.0001
Signed Rank	S	925925	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile

Quantiles (Definition 5)	
Level	Quantile
100% Max	35999.0
99%	35331.0
95%	33159.0
90%	29582.0
75% Q3	23767.5
50% Median	20391.5
25% Q1	18556.0
10%	17619.0
5%	17296.0
1%	17001.0
0% Min	16009.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
16009	160	35943	1728
16051	732	35999	13
16051	241	35999	983
16091	935	35999	1030
16102	559	35999	1509

The UNIVARIATE Procedure
Variable: Agent_Tenure

Moments			
N	1924	Sum Weights	1924
Mean	3.16320166	Sum Observations	6086
Std Deviation	2.50124822	Variance	6.25624268
Skewness	0.9674889	Kurtosis	0.04843355
Uncorrected SS	31282	Corrected SS	12030.7547
Coeff Variation	79.073309	Std Error Mean	0.05702355

Basic Statistical Measures			
Location		Variability	
Mean	3.163202	Std Deviation	2.50125

Basic Statistical Measures			
Location		Variability	
Median	2.000000	Variance	6.25624
Mode	1.000000	Range	10.00000
		Interquartile Range	4.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	55.47185	Pr > |t|	<.0001
Sign	M	897	Pr >= |M|	<.0001
Signed Rank	S	805057.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	10
99%	10
95%	8
90%	7
75% Q3	5
50% Median	2
25% Q1	1
10%	1
5%	0
1%	0
0% Min	0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	962	10	1687
0	952	10	1701
0	951	10	1721
0	942	10	1840
0	941	10	1856

Variable: YTD_contact_cnt

Moments			
N	1924	Sum Weights	1924
Mean	20.6689189	Sum Observations	39767
Std Deviation	3.63693577	Variance	13.2273018
Skewness	0.83112558	Kurtosis	-0.1898013
Uncorrected SS	847377	Corrected SS	25436.1014
Coeff Variation	17.5961587	Std Error Mean	0.082915

Basic Statistical Measures			
Location		Variability	
Mean	20.66892	Std Deviation	3.63694
Median	20.00000	Variance	13.22730
Mode	19.00000	Range	15.00000
		Interquartile Range	5.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	249.2784	Pr > |t|	<.0001
Sign	M	962	Pr >= |M|	<.0001
Signed Rank	S	925925	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	31
99%	30
95%	28
90%	27
75% Q3	23
50% Median	20
25% Q1	18
10%	17
5%	16
1%	16
0% Min	16

Extreme Observations

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
16	961	31	1430
16	947	31	1486
16	945	31	1506
16	943	31	1531
16	931	31	1592

The UNIVARIATE Procedure
Variable: Due_date_day_cnt

Moments			
N	1924	Sum Weights	1924
Mean	11.6496881	Sum Observations	22414
Std Deviation	7.56699631	Variance	57.2594331
Skewness	1.06616338	Kurtosis	0.77727348
Uncorrected SS	371226	Corrected SS	110109.89
Coeff Variation	64.9544967	Std Error Mean	0.17251267

Basic Statistical Measures			
Location		Variability	
Mean	11.64969	Std Deviation	7.56700
Median	10.00000	Variance	57.25943
Mode	10.00000	Range	41.00000
		Interquartile Range	9.50000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	67.52947	Pr > |t|	<.0001
Sign	M	959.5	Pr >= |M|	<.0001
Signed Rank	S	921120	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	41.0

Quantiles (Definition 5)	
Level	Quantile
99%	34.0
95%	28.0
90%	23.0
75% Q3	15.5
50% Median	10.0
25% Q1	6.0
10%	3.0
5%	2.0
1%	1.0
0% Min	0.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	732	37	642
0	598	37	1324
0	595	38	1594
0	241	38	1870
0	141	41	1011

The UNIVARIATE Procedure
Variable: Existing_policy_count

Moments			
N	1924	Sum Weights	1924
Mean	8.09303534	Sum Observations	15571
Std Deviation	4.32748996	Variance	18.7271694
Skewness	-0.0334004	Kurtosis	-1.1849708
Uncorrected SS	162029	Corrected SS	36012.3467
Coeff Variation	53.4717788	Std Error Mean	0.09865828

Basic Statistical Measures			
Location		Variability	
Mean	8.09304	Std Deviation	4.32749
Median	8.00000	Variance	18.72717

Basic Statistical Measures			
Location		Variability	
Mode	12.00000	Range	14.00000
		Interquartile Range	8.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	82.03098	Pr > |t|	<.0001
Sign	M	962	Pr >= |M|	<.0001
Signed Rank	S	925925	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	15
99%	15
95%	15
90%	14
75% Q3	12
50% Median	8
25% Q1	4
10%	2
5%	1
1%	1
0% Min	1

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	1905	15	1868
1	1897	15	1869
1	1864	15	1876
1	1857	15	1889
1	1856	15	1906

Moments			
N	1924	Sum Weights	1924
Mean	1.8004158	Sum Observations	3464
Std Deviation	2.25117519	Variance	5.06778974
Skewness	2.09467831	Kurtosis	4.07297366
Uncorrected SS	15982	Corrected SS	9745.35967
Coeff Variation	125.036405	Std Error Mean	0.05132238

Basic Statistical Measures			
Location		Variability	
Mean	1.800416	Std Deviation	2.25118
Median	1.000000	Variance	5.06779
Mode	2.000000	Range	10.00000
		Interquartile Range	2.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	35.08052	Pr > |t|	<.0001
Sign	M	687.5	Pr >= |M|	<.0001
Signed Rank	S	473000	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	10
99%	10
95%	8
90%	5
75% Q3	2
50% Median	1
25% Q1	0
10%	0
5%	0
1%	0
0% Min	0

Extreme Observations	
Lowest	Highest

Value	Extremes Obs	Value	Obs
Lowest		Highest	

Value	Obs	Value	Obs
0	1923	10	1781
0	1920	10	1797
0	1918	10	1873
0	1917	10	1879
0	1916	10	1913

Question 06:

The FREQ Procedure

Payment_Period	Frequency	Percent
Monthly	345	17.93
Quarterly	189	9.82
Yearly	1390	72.25

Product	Frequency	Percent
Market Link	81	4.21
Pure Term Plan	560	29.11
Traditional	1283	66.68

EducationField	Frequency	Percent
CA	583	30.30
Engineer	188	9.77
MBA	30	1.56
Marketing Diploma	219	11.38
Other	110	5.72
Statistics	794	41.27

Gender	Frequency	Percent
Female	732	38.05
Male	1192	61.95

Cust_Designation	Frequency	Percent
AVP	139	7.22
Executive	723	37.58
Manager	679	35.29
Senior Manager	298	15.49
VP	85	4.42

Cust_MaritalStatus	Frequency	Percent
Divorced	313	16.27
Married	968	50.31
Single	643	33.42

Complaint	Frequency	Percent
0	1368	71.10
1	556	28.90

Question 07:

Obs	CustID	Age	EducationField	Gender	Cust_Income
1	10046	40	Engineer	Female	34545
2	10159	57	Marketing Diploma	Male	25855
3	10344	36	Marketing Diploma	Female	23268
4	13498	30	Statistics	Female	21093
5	15213	37	Statistics	Female	20302

Question 8:

The CORR Procedure

10 Variables:	Age Cust_Tenure Overall_cust_satisfation_score CC_Satisfation_score Cust_Income Agent_Tenure YTD_contact_cnt Due_date_day_cnt Existing_policy_count Miss_due_date_cnt

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Age	1924	42.62422	10.01131	82009	21.00000	60.00000
Cust_Tenure	1924	12.64865	7.01534	24336	1.00000	25.00000
Overall_cust_satisfation_score	1924	3.39553	1.18053	6533	1.00000	5.00000
CC_Satisfation_score	1924	3.05146	1.36632	5871	1.00000	5.00000
Cust_Income	1924	21964	4729	42259544	16009	35999
Agent_Tenure	1924	3.16320	2.50125	6086	0	10.00000
YTD_contact_cnt	1924	20.66892	3.63694	39767	16.00000	31.00000
Due_date_day_cnt	1924	11.64969	7.56700	22414	0	41.00000
Existing_policy_count	1924	8.09304	4.32749	15571	1.00000	15.00000
Miss_due_date_cnt	1924	1.80042	2.25118	3464	0	10.00000

Pearson Correlation Coefficients, N = 1924

	Age	Cust_Tenure	Overall_cust_satisfation_score	CC_Satisfation_score	Cust_Income	Agent_Tenure	YTD_contact_cnt	Due_date_day_cnt	Existing_policy_count	Miss_due_date_cnt
Age	1.00000	0.26821	0.18660	-0.11028	0.06829	-0.03420	-0.00126	0.08963	0.01701	-0.44226
Cust_Tenure	0.26821	1.00000	0.17760	-0.07049	0.07440	-0.01564	0.02964	0.08561	0.01271	-0.37719
Overall_cust_satisfation_score	0.18660	0.17760	1.00000	-0.05454	0.07038	-0.03825	-0.01599	0.03811	0.00643	-0.26320
CC_Satisfation_score	-0.11028	-0.07049	-0.05454	1.00000	0.01686	0.07712	0.00573	0.02523	0.02944	0.12659
Cust_Income	0.06829	0.07440	0.07038	0.01686	1.00000	0.18961	-0.00097	0.76659	0.02487	-0.13632
Agent_Tenure	-0.03420	-0.01564	-0.03825	0.07712	0.18961	1.00000	0.02561	0.25318	-0.00866	0.03396
YTD_contact_cnt	-0.00126	0.02964	-0.01599	0.00573	-0.00097	0.02561	1.00000	0.01173	0.05125	-0.01474
Due_date_day_cnt	0.08963	0.08561	0.03811	0.02523	0.76659	0.25318	0.01173	1.00000	0.02785	-0.13424
Existing_policy_count	0.01701	0.01271	0.00643	0.02944	0.02487	-0.00866	0.05125	0.02785	1.00000	0.00591
Miss_due_date_cnt	-0.44226	-0.37719	-0.26320	0.12659	-0.13632	0.03396	-0.01474	-0.13424	0.00591	1.00000

Question 9:

The FREQ Procedure

Churn	Frequency	Percent
0	1607	83.52
1	317	16.48

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	LIFE_INSURANCE_DATA
Random Number Seed	1234
Sample Size	600
Selection Probability	0.31185
Sampling Weight	3.206667
Number of Replicates	1
Total Sample Size	600
Output Data Set	TEST

The CONTENTS Procedure

Data Set Name	WORK.TEST	Observations	600
Member Type	DATA	Variables	21
Engine	V9	Indexes	0
Created	07/17/2022 12:32:18	Observation Length	192
Last Modified	07/17/2022 12:32:18	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1

Engine/Host Dependent Information	
First Data Page	1
Max Obs per Page	682
Obs in First Data Page	600
Number of Data Set Repairs	0
Filename	/saswork/SAS_work5A1D000014AE_odaws01-apse1-2.oda.sas.com/SAS_workE9AE000014AE_odaws01-apse1-2.oda.sas.com/test.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	134347725
Access Permission	rw-r--r--
Owner Name	u61856037
File Size	256KB
File Size (bytes)	262144

Variables in Creation Order						
#	**Variable**	**Type**	**Len**	**Format**	**Informat**	**Label**
1	Replicate	Num	8			Sample Replicate Number
2	CustID	Num	8	BEST12.	BEST32.	
3	Mobile_num	Num	8	BEST12.	BEST32.	
4	Churn	Num	8	BEST12.	BEST32.	
5	Age	Num	8	BEST12.	BEST32.	
6	Payment_Period	Char	9	$9.	$9.	
7	Product	Char	14	$14.	$14.	
8	Cust_Tenure	Num	8	BEST12.	BEST32.	
9	EducationField	Char	17	$17.	$17.	
10	Gender	Char	6	$6.	$6.	
11	Overall_cust_satisfation_score	Num	8	BEST12.	BEST32.	
12	Cust_Designation	Char	14	$14.	$14.	
13	CC_Satisfation_score	Num	8	BEST12.	BEST32.	
14	Cust_MaritalStatus	Char	8	$8.	$8.	
15	Cust_Income	Num	8	BEST12.	BEST32.	
16	Agent_Tenure	Num	8	BEST12.	BEST32.	
17	Complaint	Num	8	BEST12.	BEST32.	
18	YTD_contact_cnt	Num	8	BEST12.	BEST32.	
19	Due_date_day_cnt	Num	8	BEST12.	BEST32.	
20	Existing_policy_count	Num	8	BEST12.	BEST32.	
21	Miss_due_date_cnt	Num	8	BEST12.	BEST32.	

The FREQ Procedure

Churn	Frequency	Percent
0	511	85.17
1	89	14.83

The FREQ Procedure

Churn	Frequency	Percent
0	1096	82.78
1	228	17.22

Question 10:

Life_Insurance_data - Multicollinearity Investigation of VIF

The REG Procedure
Model: MODEL1
Dependent Variable: Churn

Number of Observations Read	1924
Number of Observations Used	1924

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	196.38570	19.63857	549.37	<.0001
Error	1913	68.38509	0.03575		
Corrected Total	1923	264.77079			

Root MSE	0.18907	R-Square	0.7417
Dependent Mean	0.16476	Adj R-Sq	0.7404
Coeff Var	114.75432		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	0.47687	0.04562	10.45	<.0001	.	0
Age	1	-0.00689	0.00048656	-14.17	<.0001	0.78344	1.27642
Cust_Tenure	1	-0.00818	0.00067153	-12.18	<.0001	0.83760	1.19389
Overall_cust_satisfation_score	1	-0.02368	0.00382	-6.20	<.0001	0.91472	1.09323
CC_Satisfation_score	1	0.00649	0.00320	2.03	0.0427	0.97301	1.02774
Cust_Income	1	-0.00000283	0.00000142	-1.99	0.0469	0.40954	2.44174
Agent_Tenure	1	0.00248	0.00179	1.38	0.1670	0.92496	1.08113
YTD_contact_cnt	1	0.00155	0.00119	1.30	0.1934	0.99492	1.00510
Due_date_day_cnt	1	-0.00105	0.00090437	-1.16	0.2448	0.39694	2.51927
Existing_policy_count	1	-0.00149	0.00099900	-1.49	0.1371	0.99464	1.00539
Miss_due_date_cnt	1	0.10690	0.00229	46.61	<.0001	0.69731	1.43408

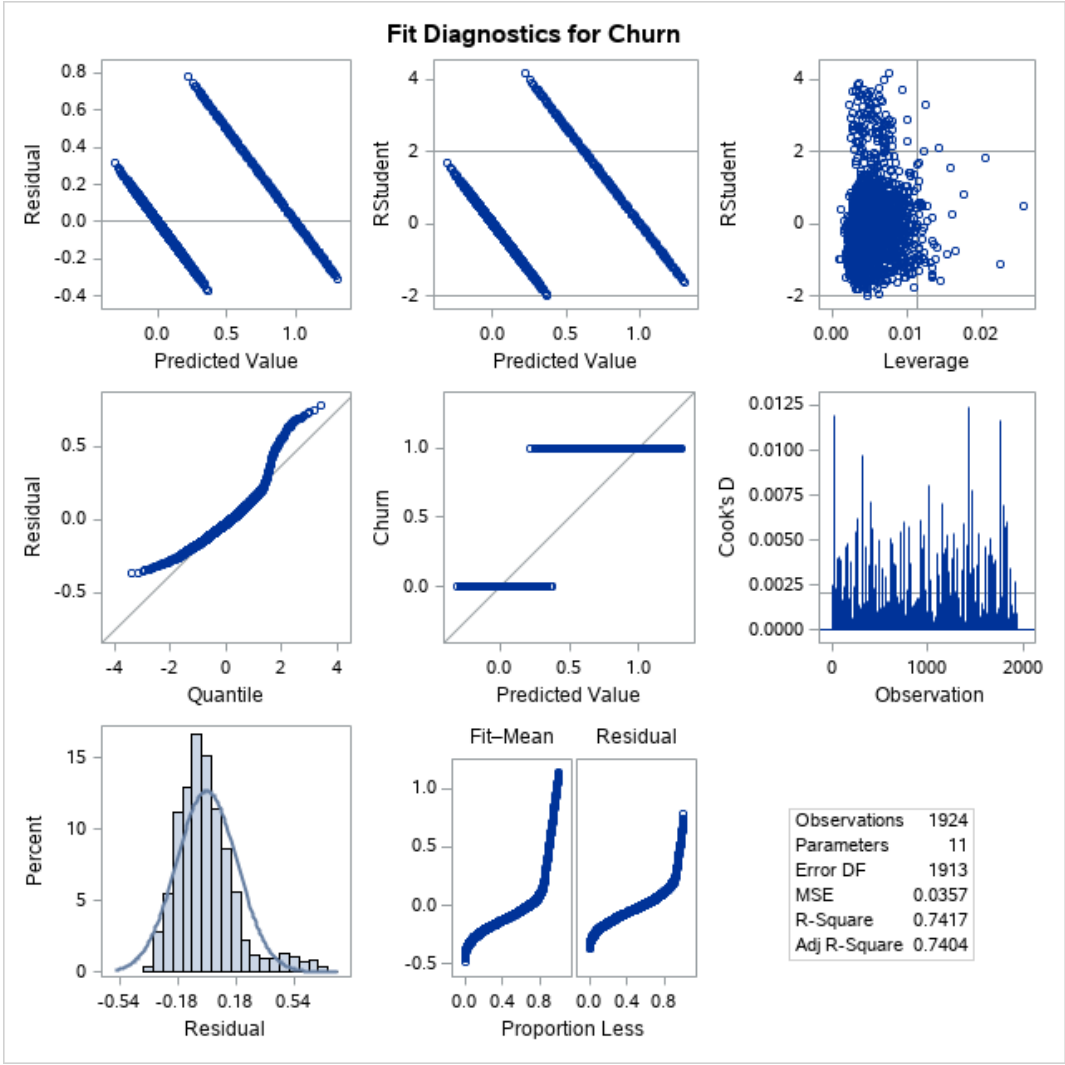
Collinearity Diagnostics

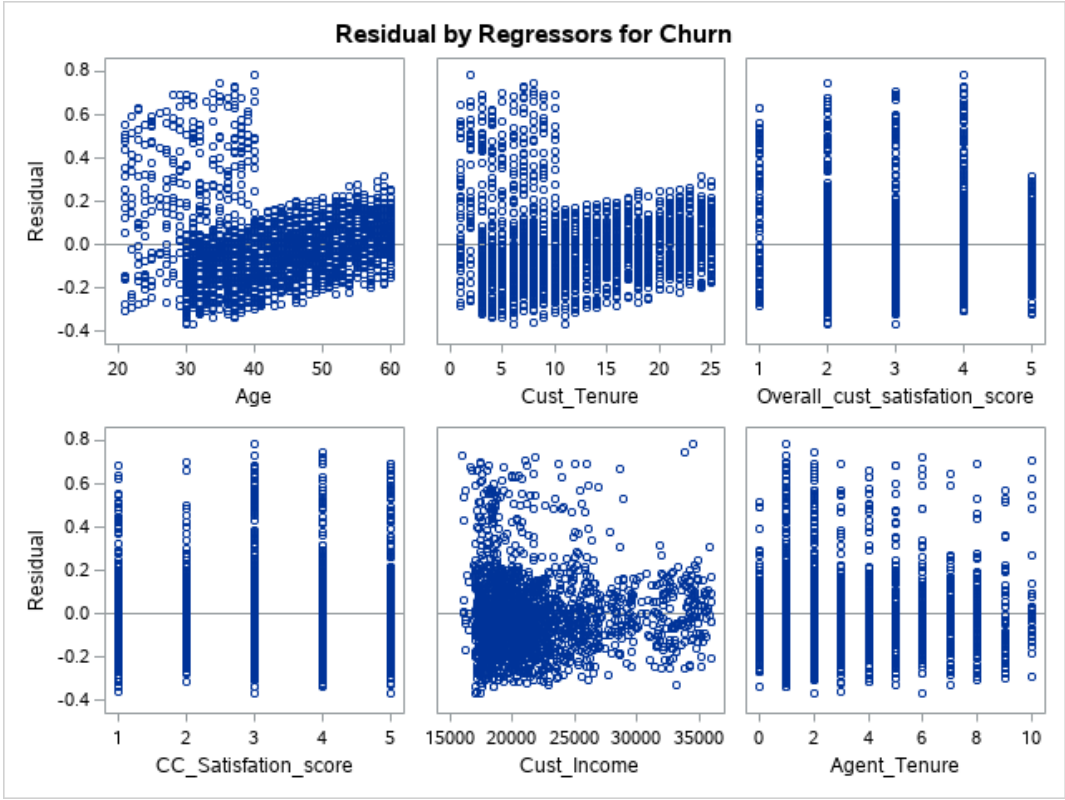
Number	Eigenvalue	Condition Index	Proportion of Variation										
			Intercept	Age	Cust_Tenure	Overall_cust_satisfation_score	CC_Satisfation_score	Cust_Income	Agent_Tenure	YTD_contact_cnt	Due_date_day_cnt	Existing_policy_count	Miss_due_date_cnt
1	9.02839	1.00000	0.00010731	0.00047094	0.00190	0.00108	0.00170	0.00021537	0.00296	0.00035118	0.00111	0.00219	0.00206
2	0.69471	3.60499	3.282394E-7	0.00097551	0.02644	0.00209	0.00181	0.00006689	0.00037261	8.897059E-8	0.00387	0.00001351	0.51507
3	0.38387	4.84967	0.00027796	0.00187	0.02310	0.00730	0.00307	0.00003504	0.63284	0.00099093	0.04063	0.03278	0.00449
4	0.24428	6.07946	0.00005966	0.00050091	0.03065	0.00238	0.00545	0.00364	0.33840	0.00033817	0.30789	0.01180	0.01264
5	0.20778	6.59173	0.00013276	0.00086779	0.17461	0.00662	0.00535	0.00006573	0.00469	0.00030708	0.00113	0.80204	0.05328
6	0.17069	7.27271	0.00017983	0.00050140	0.34676	0.00441	0.53762	0.00007335	0.00951	0.00057063	0.00472	0.05913	0.15028
7	0.13344	8.22564	0.00171	0.01793	0.35038	0.22625	0.34480	0.00140	0.00095531	0.00702	0.01468	0.04967	0.00132
8	0.07357	11.07747	0.00667	0.10750	0.03778	0.69187	0.04921	0.00543	0.00389	0.06017	0.02047	0.03142	0.02778
9	0.03587	15.86542	0.00280	0.60815	0.00098294	0.00008221	0.01220	0.01078	0.00228	0.34949	0.00753	0.00331	0.10587

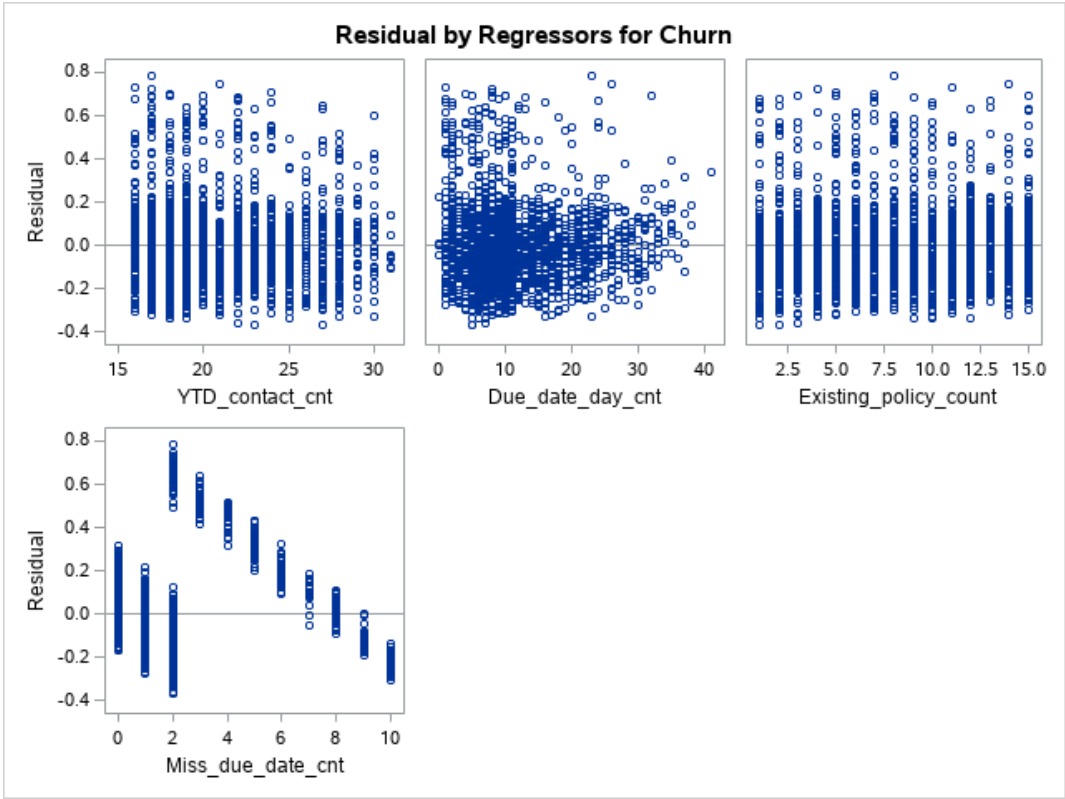
Collinearity Diagnostics													
			Proportion of Variation										
Number	Eigenvalue	Condition Index	Intercept	Age	Cust_Tenure	Overall_cust_satisfation_score	CC_Satisfation_score	Cust_Income	Agent_Tenure	YTD_contact_cnt	Due_date_day_cnt	Existing_policy_count	Miss_due_date_cnt
10	0.02066	20.90341	0.02133	0.07383	0.00004188	0.02753	0.00858	0.54948	8.330998E-9	0.30569	0.38642	0.00060483	0.02430
11	0.00674	36.59776	0.96673	0.18741	0.00736	0.03038	0.03022	0.42881	0.00409	0.27506	0.21154	0.00704	0.10291

Life_Insurance_data - Multicollinearity Investigation of VIF

The REG Procedure
Model: MODEL1
Dependent Variable: Churn


Fit Diagnostics for Churn
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Churn
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 1924
Parameters 11
Error DF 1913
MSE 0.0357
R-Square 0.7417
Adj R-Square 0.7404


Residual by Regressors for Churn
Residual
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
20 30 40 50 60
Age
0 5 10 15 20 25
Cust_Tenure
1 2 3 4 5
Overall_cust_satisfation_score
1 2 3 4 5
CC_Satisfation_score
15000 20000 25000 30000 35000
Cust_Income
0 2 4 6 8 10
Agent_Tenure


Residual by Regressors for Churn
Residual
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
15
20
25
30
YTD_contact_cnt
0
10
20
30
40
Due_date_day_cnt
2.5
5.0
7.5
10.0
12.5
15.0
Existing_policy_count
Residual
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
0
2
4
6
8
10
Miss_due_date_cnt

Question 11:

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TRAIN
Response Variable	Churn
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1324
Number of Observations Used	1324

Response Profile		
Ordered Value	Churn	Total Frequency
1	1	228
2	0	1096

Probability modeled is Churn='1'.

Model Convergence Status
Quasi-complete separation of data points detected.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1218.401	89.328
SC	1223.590	146.401
-2 Log L	1216.401	67.328

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1149.0729	10	<.0001
Score	979.4505	10	<.0001
Wald	25.9167	10	0.0039

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.1713	89.2349	0.0065	0.9359
Age	1	-0.3222	0.0796	16.3745	<.0001
Cust_Tenure	1	-0.4777	0.1159	16.9729	<.0001
Overall_cust_satisfa	1	-0.8933	0.3419	6.8278	0.0090
CC_Satisfation_score	1	0.4810	0.2655	3.2829	0.0700
Cust_Income	1	0.000010	0.000108	0.0089	0.9250
Agent_Tenure	1	0.0629	0.1476	0.1815	0.6701
YTD_contact_cnt	1	-0.1296	0.0998	1.6843	0.1944
Due_date_day_cnt	1	-0.0565	0.0725	0.6080	0.4355
Existing_policy_coun	1	-0.1055	0.0785	1.8090	0.1786
Miss_due_date_cnt	1	12.5768	44.5882	0.0796	0.7779

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Age	0.725	0.620	0.847
Cust_Tenure	0.620	0.494	0.778
Overall_cust_satisfa	0.409	0.209	0.800
CC_Satisfation_score	1.618	0.961	2.722
Cust_Income	1.000	1.000	1.000
Agent_Tenure	1.065	0.797	1.422
YTD_contact_cnt	0.878	0.722	1.068
Due_date_day_cnt	0.945	0.820	1.089
Existing_policy_coun	0.900	0.772	1.049
Miss_due_date_cnt	>999.999	<0.001	>999.999

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	99.9	Somers' D	0.998
Percent Discordant	0.1	Gamma	0.998
Percent Tied	0.0	Tau-a	0.285
Pairs	249888	c	0.999

Partition for the Hosmer and Lemeshow Test					
		Churn = 1		Churn = 0	
Group	Total	Observed	Expected	Observed	Expected

Partition for the Hosmer and Lemeshow Test					
		Churn = 1		Churn = 0	
Group	Total	Observed	Expected	Observed	Expected
1	789	0	0.00	789	789.00
2	132	0	0.00	132	132.00
3	132	0	0.64	132	131.36
4	85	42	41.35	43	43.65
5	186	186	186.00	0	0.00

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
0.6695	3	0.8804

Note: In calculating the Expected values, predicted probabilities less than 1E-6 and greater than 0.999999 were changed to 1E-6 and 0.999999 respectively.

Question 12:

The UNIVARIATE Procedure
Variable: Overall_cust_satisfation_score

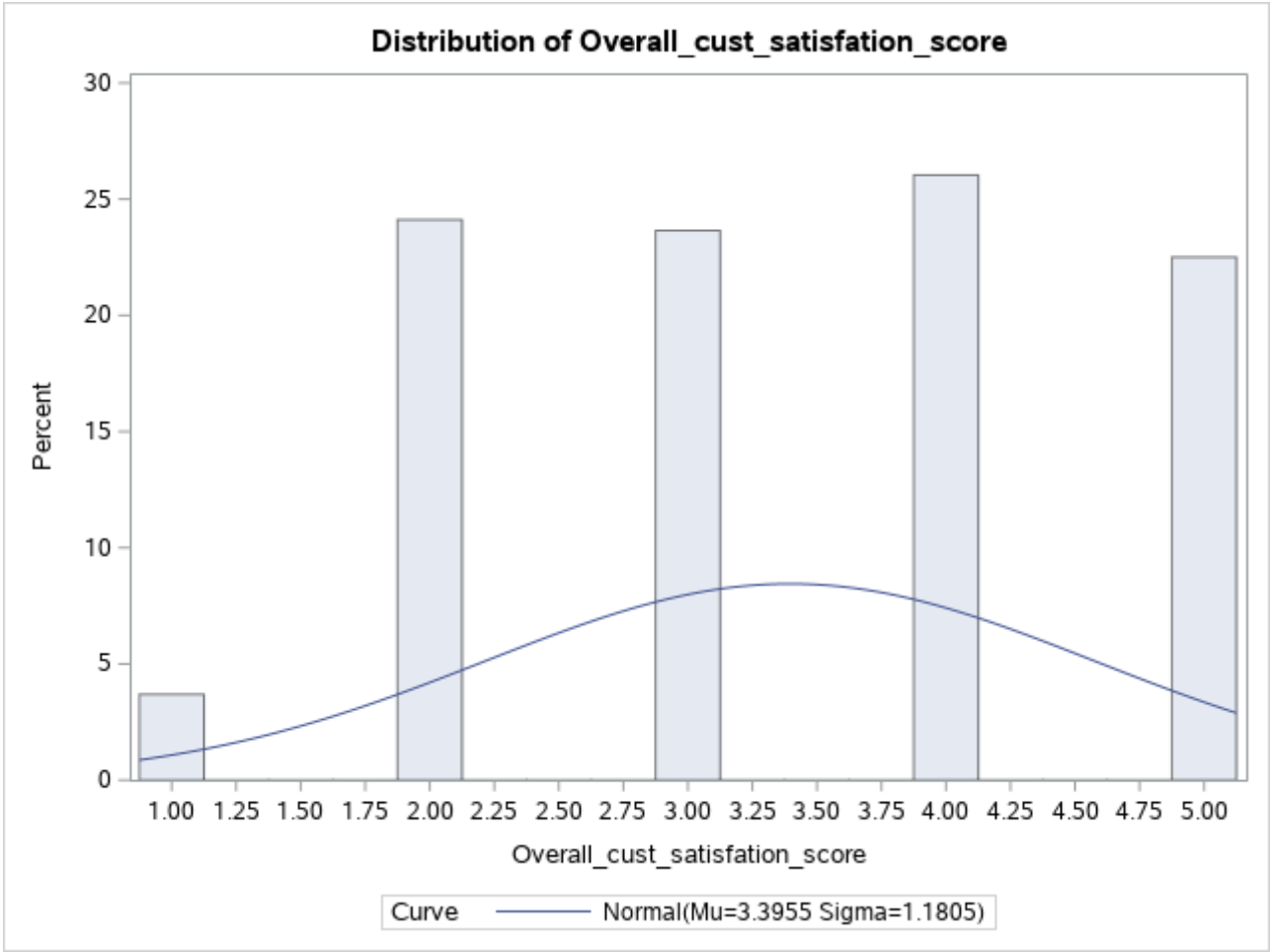
Moments			
N	1924	Sum Weights	1924
Mean	3.39553015	Sum Observations	6533
Std Deviation	1.18053232	Variance	1.39365656
Skewness	-0.1158404	Kurtosis	-1.1124382
Uncorrected SS	24863	Corrected SS	2680.00156
Coeff Variation	34.767246	Std Error Mean	0.02691382

Basic Statistical Measures			
Location		Variability	
Mean	3.395530	Std Deviation	1.18053
Median	3.000000	Variance	1.39366
Mode	4.000000	Range	4.00000
		Interquartile Range	2.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	126.1631	Pr > |t|	<.0001
Sign	M	962	Pr >= |M|	<.0001
Signed Rank	S	925925	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	5
99%	5
95%	5
90%	5
75% Q3	4
50% Median	3
25% Q1	2
10%	2
5%	2
1%	1
0% Min	1

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	1888	5	1911
1	1885	5	1915
1	1874	5	1918
1	1856	5	1919
1	1788	5	1920



The UNIVARIATE Procedure
Fitted Normal Distribution for Overall_cust_satisfation_score

Parameters for Normal Distribution		
Parameter	**Symbol**	**Estimate**
Mean	Mu	3.39553
Std Dev	Sigma	1.180532

Goodness-of-Fit Tests for Normal Distribution				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.1811325	**Pr > D**	<0.010
Cramer-von Mises	**W-Sq**	11.6638584	**Pr > W-Sq**	<0.005
Anderson-Darling	**A-Sq**	77.6893845	**Pr > A-Sq**	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	**Observed**	**Estimated**
1.0	1.00000	0.64920
5.0	2.00000	1.45373
10.0	2.00000	1.88262
25.0	2.00000	2.59927
50.0	3.00000	3.39553
75.0	4.00000	4.19179
90.0	5.00000	4.90844
95.0	5.00000	5.33733
99.0	5.00000	6.14186

Question 13:

The REG Procedure
Model: MODEL1
Dependent Variable: Churn

Number of Observations Read	1924
Number of Observations Used	1924

Analysis of Variance

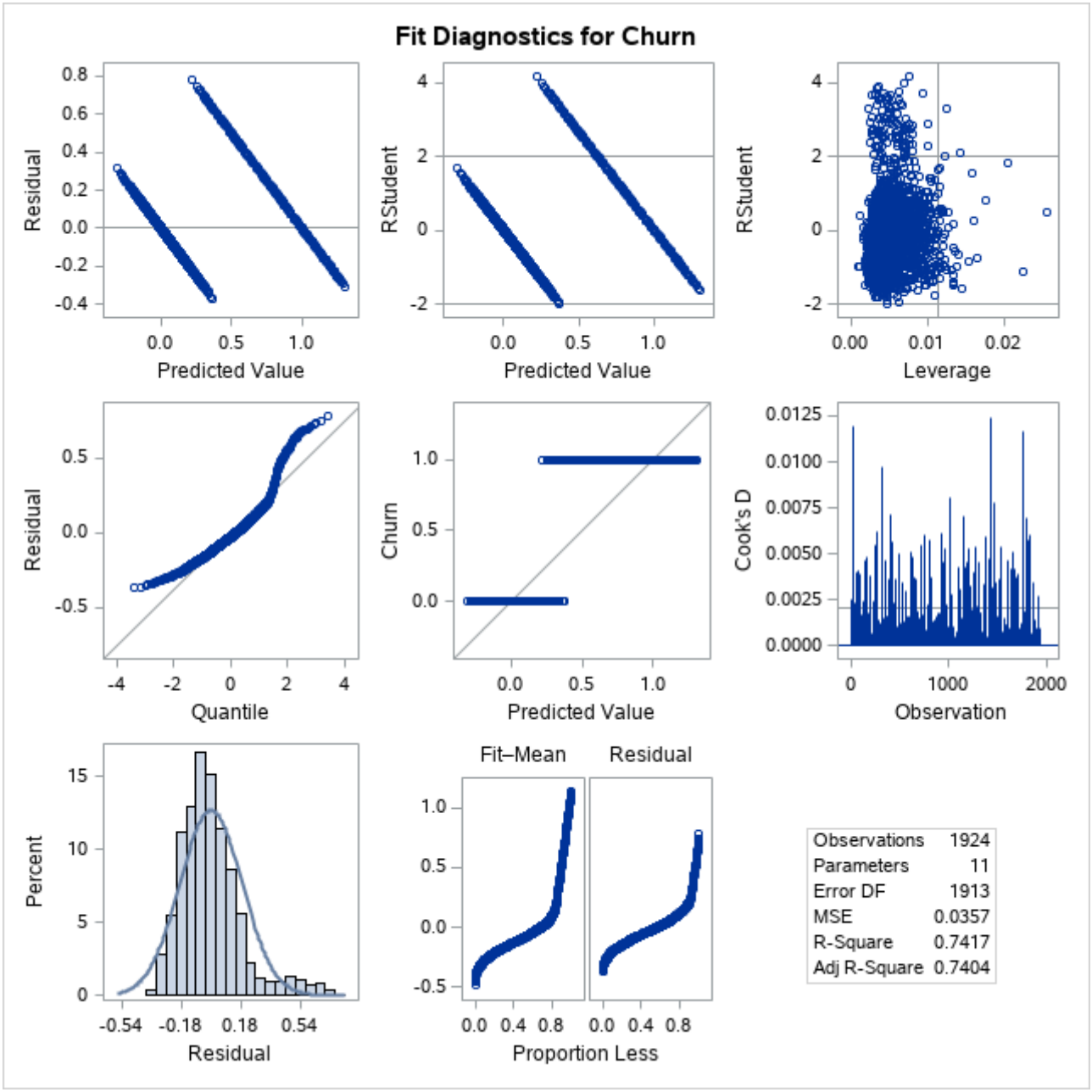
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	196.38570	19.63857	549.37	<.0001
Error	1913	68.38509	0.03575		
Corrected Total	1923	264.77079			

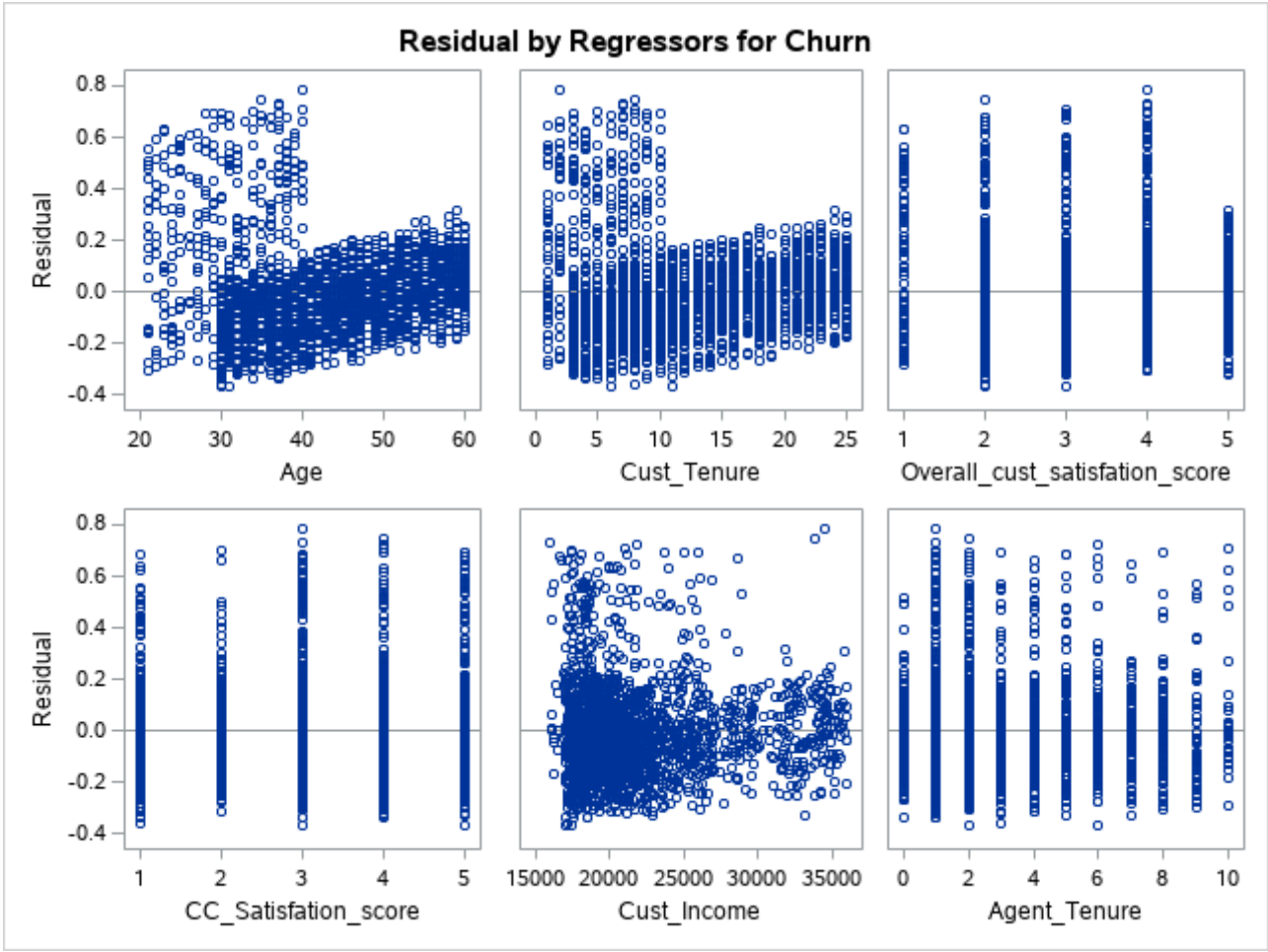
Root MSE	0.18907	R-Square	0.7417
Dependent Mean	0.16476	Adj R-Sq	0.7404
Coeff Var	114.75432		

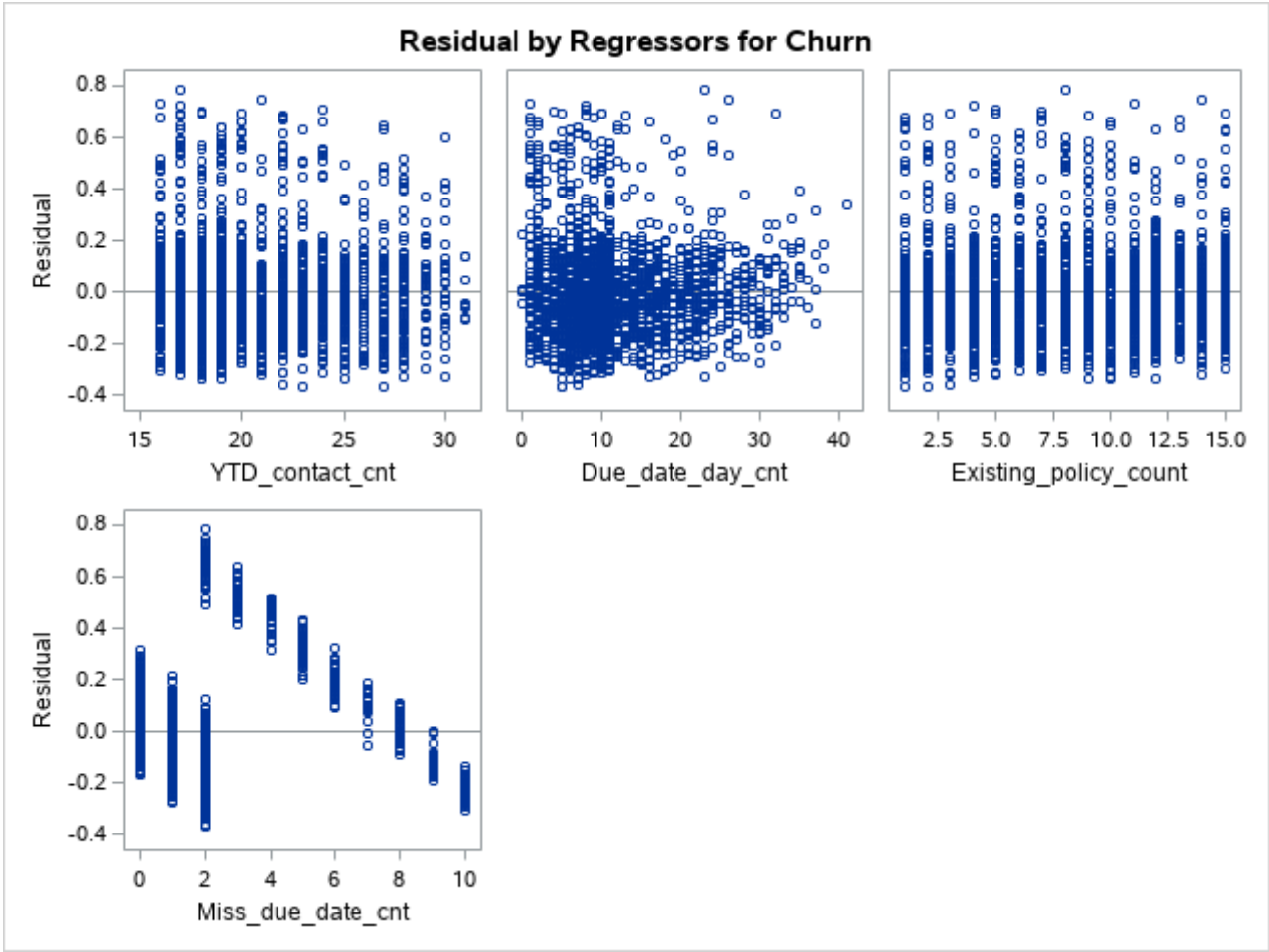
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.47687	0.04562	10.45	<.0001
Age	1	-0.00689	0.00048656	-14.17	<.0001
Cust_Tenure	1	-0.00818	0.00067153	-12.18	<.0001
Overall_cust_satisfation_score	1	-0.02368	0.00382	-6.20	<.0001
CC_Satisfation_score	1	0.00649	0.00320	2.03	0.0427
Cust_Income	1	-0.00000283	0.00000142	-1.99	0.0469
Agent_Tenure	1	0.00248	0.00179	1.38	0.1670
YTD_contact_cnt	1	0.00155	0.00119	1.30	0.1934
Due_date_day_cnt	1	-0.00105	0.00090437	-1.16	0.2448
Existing_policy_count	1	-0.00149	0.00099900	-1.49	0.1371
Miss_due_date_cnt	1	0.10690	0.00229	46.61	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: Churn


Fit Diagnostics for Churn
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Churn
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 1924
Parameters 11
Error DF 1913
MSE 0.0357
R-Square 0.7417
Adj R-Square 0.7404


Residual by Regressors for Churn
Residual
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
20 30 40 50 60
Age
0 5 10 15 20 25
Cust_Tenure
1 2 3 4 5
Overall_cust_satisfation_score
1 2 3 4 5
CC_Satisfation_score
15000 20000 25000 30000 35000
Cust_Income
0 2 4 6 8 10
Agent_Tenure


Residual by Regressors for Churn
Residual
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
15
20
25
30
YTD_contact_cnt
0
10
20
30
40
Due_date_day_cnt
2.5
5.0
7.5
10.0
12.5
15.0
Existing_policy_count
Residual
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
0
2
4
6
8
10
Miss_due_date_cnt