

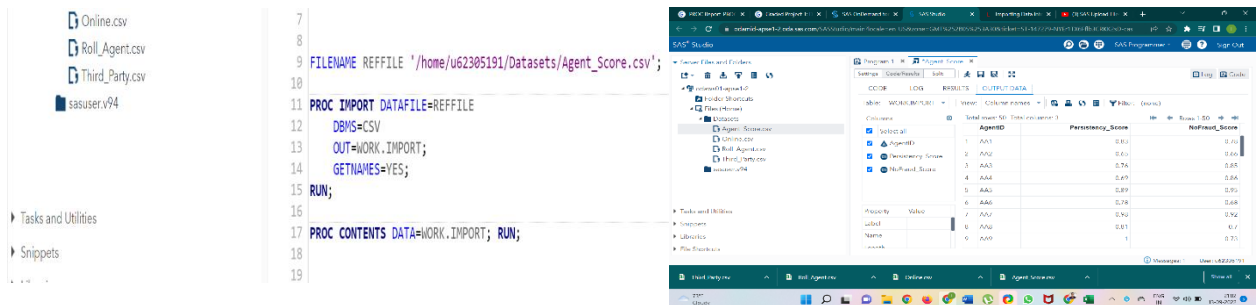
Great Learning SAS Assessment

Ishita Sarkar

1. Import all the 4 files in SAS data environment

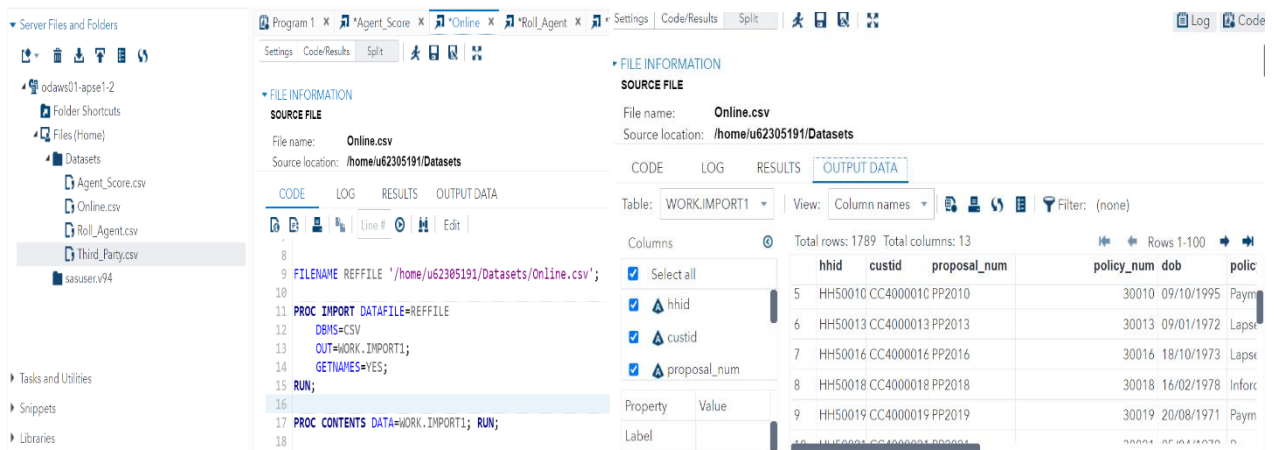
Importing First file

```
FILENAME REFFILE '/home/u62305191/Datasets/Agent_Score.csv';  
PROC IMPORT DATAFILE=REFFILE  
    DBMS=CSV  
    OUT=SAS.Agent;  
    GETNAMES=YES;  
RUN;  
PROC CONTENTS DATA= SAS.Agent; RUN;
```



Importing Second file

```
FILENAME REFFILE '/home/u62305191/Datasets/Online.csv';  
PROC IMPORT DATAFILE=REFFILE  
    DBMS=CSV  
    OUT=SAS.Online;  
    GETNAMES=YES;  
RUN;  
PROC CONTENTS DATA= SAS.Online; RUN;
```



Great Learning SAS Assessment

Ishita Sarkar

Importing Third file

```
FILENAME REFFILE '/home/u62305191/Datasets/Roll_Agent.csv';  
PROC IMPORT DATAFILE=REFFILE  
    DBMS=CSV  
    OUT=SAS.Roll;  
    GETNAMES=YES;  
RUN;  
PROC CONTENTS DATA= SAS.Roll; RUN;
```

FILE INFORMATION

SOURCE FILE

File name: Roll_Agent.csv
Source location: /home/u62305191/Datasets

CODE LOG RESULTS OUTPUT DATA

7
8
9 FILENAME REFFILE '/home/u62305191/Datasets/Roll_Agent.csv';
10
11 PROC IMPORT DATAFILE=REFFILE
12 DBMS=CSV
13 OUT=WORK.IMPORT2;
14 GETNAMES=YES;
15 RUN;
16
17 PROC CONTENTS DATA=WORK.IMPORT2; RUN;
18

SUMMARY OUTPUT DATA

View: Column names Filter: (none)

Total rows: 1864 Total columns: 13 Rows 1-100

	hhid	custid	proposal_num	policy_num	dob	police
1	HH50001	CC4000001	PP2001	30001	06/01/1975	Lapse
2	HH50004	CC4000004	PP2004	30004	20/07/1980	Paym
3	HH50008	CC4000008	PP2008	30008	12/11/1985	Inforc
4	HH50011	CC4000011	PP2011	30011	20/07/1992	Paym
5	HH50012	CC4000012	PP2012	30012	02/04/1979	Paym

Importing Fourth file

```
FILENAME REFFILE '/home/u62305191/Datasets/Third_Party.csv';  
PROC IMPORT DATAFILE=REFFILE  
    DBMS=CSV  
    OUT=SAS.Third;  
    GETNAMES=YES;  
RUN;  
PROC CONTENTS DATA= SAS.Third; RUN;
```

FILE INFORMATION

SOURCE FILE

File name: Third_Party.csv
Source location: /home/u62305191/Datasets

CODE LOG RESULTS OUTPUT DATA

7
8
9 FILENAME REFFILE '/home/u62305191/Datasets/Third_Party.csv';
10
11 PROC IMPORT DATAFILE=REFFILE
12 DBMS=CSV
13 OUT=WORK.IMPORT3;
14 GETNAMES=YES;
15 RUN;
16
17 PROC CONTENTS DATA=WORK.IMPORT3; RUN;
18

RESULTS OUTPUT DATA

View: Column names Filter: (none)

Total rows: 1726 Total columns: 13 Rows 1-100

	hhid	custid	proposal_num	policy_num	dob	police
1	HH50000	CC4000000	PP2000	30000	06/09/1974	Paym
2	HH50005	CC4000005	PP2005	30005	26/09/1985	Lapse
3	HH50009	CC4000009	PP2009	30009	10/12/1981	Inforc
4	HH50014	CC4000014	PP2014	30014	17/11/1969	Paym
5	HH50015	CC4000015	PP2015	30015	14/11/1971	Paym

Great Learning SAS Assessment

Ishita Sarkar

2. Create one dataset from all the 4 dataset?

Data SAS.Merge;

set Online Third_Party Roll_Agent;

run;

/* sorting data using the agentid also using out statement to avoid overriding of data*/

proc sort data =SAS.Merge out=SAS.Merge1;

by agentid ;

run;

/* sorting the data using the agentid also using out statement to avoid overriding of data */

proc sort data =SAS.Agent out=SAS.Agent1;

by agentid ;

run;

/* merging data using the agentid and usin*/

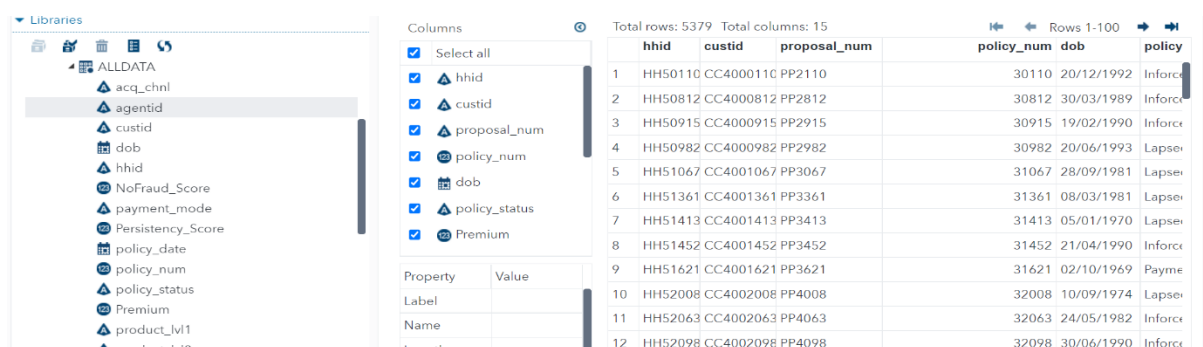
data SAS.MergedData;

merge SAS.Merge1(in=a) SAS.Agent1(in=b);

by agentid;

if a;

run;



	hhid	custid	proposal_num	policy_num	dob	policy
1	HH50110	CC4000110	PP2110	30110	20/12/1992	Inforce
2	HH50812	CC4000812	PP2812	30812	30/03/1989	Inforce
3	HH50915	CC4000915	PP2915	30915	19/02/1990	Inforce
4	HH50982	CC4000982	PP2982	30982	20/06/1993	Lapsed
5	HH51067	CC4001067	PP3067	31067	28/09/1981	Lapsed
6	HH51361	CC4001361	PP3361	31361	08/03/1981	Lapsed
7	HH51413	CC4001413	PP3413	31413	05/01/1970	Lapsed
8	HH51452	CC4001452	PP3452	31452	21/04/1990	Inforce
9	HH51621	CC4001621	PP3621	31621	02/10/1969	Payment
10	HH52008	CC4002008	PP4008	32008	10/09/1974	Lapsed
11	HH52063	CC4002063	PP4063	32063	24/05/1982	Inforce
12	HH52098	CC4002098	PP4098	32098	30/06/1990	Inforce

Great Learning SAS Assessment

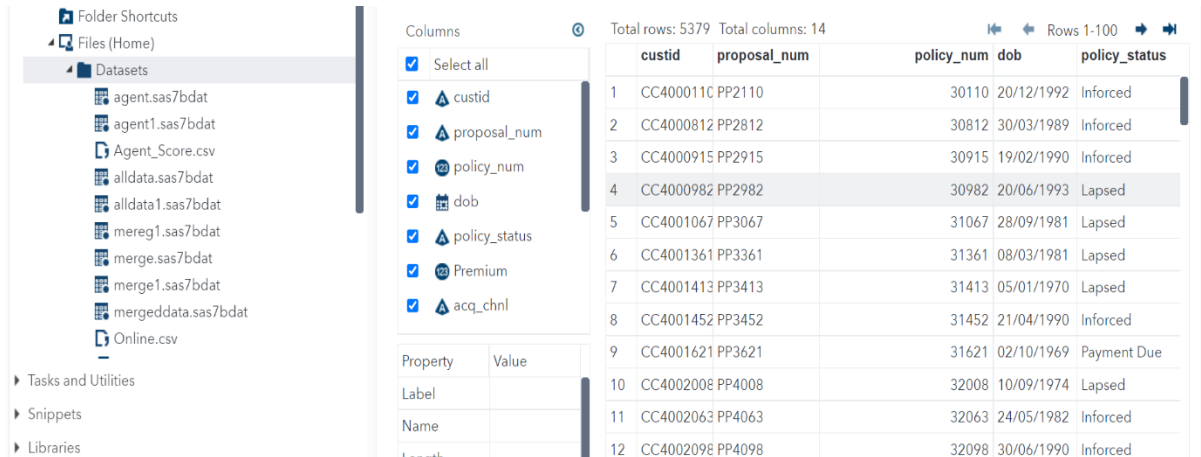
Ishita Sarkar

3. Remove all unwanted ID variables?

Data SAS.MergedData (drop=hhid);

set SAS.MergedData;

run;



The screenshot shows the SAS Studio interface. On the left, the 'Files (Home)' pane displays a list of datasets under 'Datasets', including 'agent.sas7bdat', 'agent1.sas7bdat', 'Agent_Score.csv', 'alldata.sas7bdat', 'alldata1.sas7bdat', 'merreg1.sas7bdat', 'merge.sas7bdat', 'merge1.sas7bdat', 'mergeddata.sas7bdat', and 'Online.csv'. The 'Columns' panel in the center shows a list of variables: 'Select all', 'custid', 'proposal_num', 'policy_num', 'dob', 'policy_status', 'Premium', and 'acq_chnl'. The 'Property' and 'Value' columns are empty. The main data table on the right displays 12 rows of data with columns: 'custid', 'proposal_num', 'policy_num', 'dob', and 'policy_status'.

	custid	proposal_num	policy_num	dob	policy_status
1	CC400011C	PP2110	30110	20/12/1992	Inforced
2	CC4000812	PP2812	30812	30/03/1989	Inforced
3	CC4000915	PP2915	30915	19/02/1990	Inforced
4	CC4000982	PP2982	30982	20/06/1993	Lapsed
5	CC4001067	PP3067	31067	28/09/1981	Lapsed
6	CC4001361	PP3361	31361	08/03/1981	Lapsed
7	CC4001413	PP3413	31413	05/01/1970	Lapsed
8	CC4001452	PP3452	31452	21/04/1990	Inforced
9	CC4001621	PP3621	31621	02/10/1969	Payment Due
10	CC4002008	PP4008	32008	10/09/1974	Lapsed
11	CC4002063	PP4063	32063	24/05/1982	Inforced
12	CC4002098	PP4098	32098	30/06/1990	Inforced

4. Calculate annual premium for all customers?

Data SAS.MergedData;

set SAS.MergedData;

if payment_mode = "Annual" then annual_premium=premium;

else if payment_mode="Quaterly" then annual_premium=premium*4;

else if payment_mode="Semi Annual" then annual_premium=premium*2;

else if payment_mode="Monthly" then annual_premium=premium*12;

run;

proc print data=SAS.MergedData;



The screenshot shows the SAS Studio interface. On the left, the 'Files (Home)' pane displays a list of datasets under 'Datasets', including 'agent.sas7bdat', 'agent1.sas7bdat', 'Agent_Score.csv', 'alldata.sas7bdat', 'alldata1.sas7bdat', 'merreg1.sas7bdat', 'merge.sas7bdat', 'merge1.sas7bdat', 'mergeddata.sas7bdat', and 'Online.csv'. The 'Columns' panel in the center shows a list of variables: 'Select all', 'custid', 'proposal_num', 'policy_num', 'dob', 'policy_status', 'Premium', and 'acq_chnl'. The 'Property' and 'Value' columns are empty. The main data table on the right displays 12 rows of data with columns: 'lob', 'policy_status', 'Premium', 'acq_chnl', 'product_lv1', 'product_lv2', 'agentid', 'payment_mode', 'policy_date', 'Persistency_Score', 'NoFraud_Score', and 'annual_premium'.

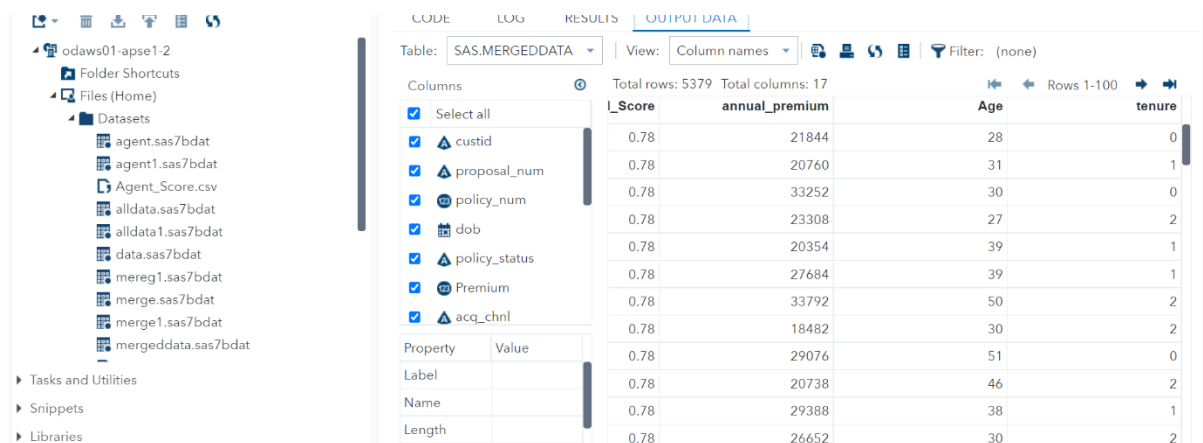
lob	policy_status	Premium	acq_chnl	product_lv1	product_lv2	agentid	payment_mode	policy_date	Persistency_Score	NoFraud_Score	annual_premium
992	Inforced	5461	Online	Return	P003 Suracha	AA1	Quaterly	23/05/2020	0.83	0.78	21844
989	Inforced	5190	Online	Return	P002 Jawaan	AA1	Quaterly	04/05/2019	0.83	0.78	20760
990	Inforced	2771	Online	Return	P005 5D plan	AA1	Monthly	20/01/2020	0.83	0.78	33252
993	Lapsed	11654	Online	Return	P006 Saath	AA1	Semi Annual	09/07/2018	0.83	0.78	23308
981	Lapsed	10177	Online	Term	P006 Saath	AA1	Semi Annual	29/01/2019	0.83	0.78	20354
981	Lapsed	2307	Online	Return	P006 Saath	AA1	Monthly	31/01/2019	0.83	0.78	27684
970	Lapsed	2816	Online	Term	P005 5D plan	AA1	Monthly	15/04/2018	0.83	0.78	33792
990	Inforced	18482	Online	Term	P005 5D plan	AA1	Annual	11/10/2018	0.83	0.78	18482
969	Payment Due	2423	Online	Return	P004 Protection	AA1	Monthly	02/03/2020	0.83	0.78	29076
974	Lapsed	10369	Online	Return	P001 Kishan	AA1	Semi Annual	11/05/2018	0.83	0.78	20738
982	Inforced	2449	Online	Term	P002 Jawaan	AA1	Monthly	15/01/2019	0.83	0.78	29388
990	Inforced	2221	Online	Return	P001 Kishan	AA1	Monthly	16/05/2018	0.83	0.78	26652
994	Inforced	2725	Online	Return	P006 Saath	AA1	Monthly	13/02/2019	0.83	0.78	32700
978	Lapsed	11993	Online	Term	P001 Kishan	AA1	Semi Annual	13/06/2020	0.83	0.78	23986

Great Learning SAS Assessment

Ishita Sarkar

5. Calculate age and tenure as of 31 July 2020 for all customers?

```
Data SAS.MergedData (drop=temp);  
set SAS.MergedData;  
temp='31jul2020'd;  
format temp ddmmyy10.;  
Age=intck('year',dob,temp);  
tenure=intck('year',policy_date,temp);  
run;
```



l_score	annual_premium	Age	tenure
0.78	21844	28	0
0.78	20760	31	1
0.78	33252	30	0
0.78	23308	27	2
0.78	20354	39	1
0.78	27684	39	1
0.78	33792	50	2
0.78	18482	30	2
0.78	29076	51	0
0.78	20738	46	2
0.78	29388	38	1
0.78	26652	30	2

6. Create a product name by using both level of product information. And product name should be representable i.e. no code should be present in final product name?

```
Data SAS.MergedData ;  
set SAS.MergedData;  
length ProductName $ 32;  
ProductName=Compress(cat(product_lvl1,product_lvl2));  
put ProductName=;  
run;
```

Great Learning SAS Assessment

Ishita Sarkar

Age	tenure	ProductName
28	0	ReturnP003Suracha
31	1	ReturnP002Jawaan
30	0	ReturnP0055Dplan
27	2	ReturnP006Saath
39	1	TermP006Saath
39	1	ReturnP006Saath
50	2	TermP0055Dplan
30	2	TermP0055Dplan
51	0	ReturnP004Protection
46	2	ReturnP001Kishan
38	1	TermP002Jawaan
30	2	ReturnP001Kishan

- After doing clean up in your data, you have to calculate the distribution of customers across product and policy status and interpret the result

```
proc sort data=SAS.MergedData out=SAS.MergedData1;
by ProductName policy_Status;
run;

proc freq data=SAS.MergedData1;
table ProductName*policy_Status;
run;
```

The FREQ Procedure

Table of ProductName by policy_status				
ProductName	policy_status			
	Inforced	Lapsed	Payment Due	Total
ReturnP001Kishan	142	163	144	449
	2.64	3.03	2.68	8.35
	31.63	36.30	32.07	
	8.04	9.09	7.92	
ReturnP002Jawaan	159	154	150	463
	2.96	2.86	2.79	8.61
	34.34	33.26	32.40	
	9.00	8.58	8.25	
ReturnP003Suracha	142	137	159	438
	2.64	2.55	2.96	8.14
	32.42	31.28	36.30	
	8.04	7.64	8.75	
ReturnP004Protection	151	145	152	448
	2.81	2.70	2.83	8.33
	33.71	32.37	33.93	
	8.55	8.08	8.36	
ReturnP0055Dplan	148	150	152	450
	2.75	2.79	2.83	8.37
	32.89	33.33	33.78	
	8.38	8.36	8.36	
ReturnP006Saath	143	147	164	454
	2.66	2.73	3.05	8.44

Great Learning SAS Assessment

Ishita Sarkar

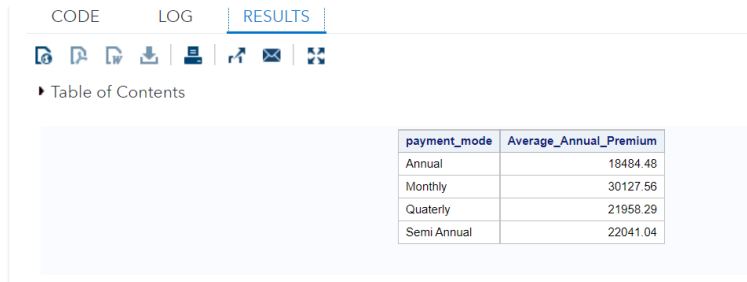
8. Calculate Average annual premium for different payment mode and interpret the result?

```
proc sql;
```

```
select payment_mode,avg(annual_premium)As Average_Annual_Premium  
from SAS.MergedData1
```

```
group by payment_mode;
```

```
run;
```



payment_mode	Average_Annual_Premium
Annual	18484.48
Monthly	30127.56
Quarterly	21958.29
Semi Annual	22041.04

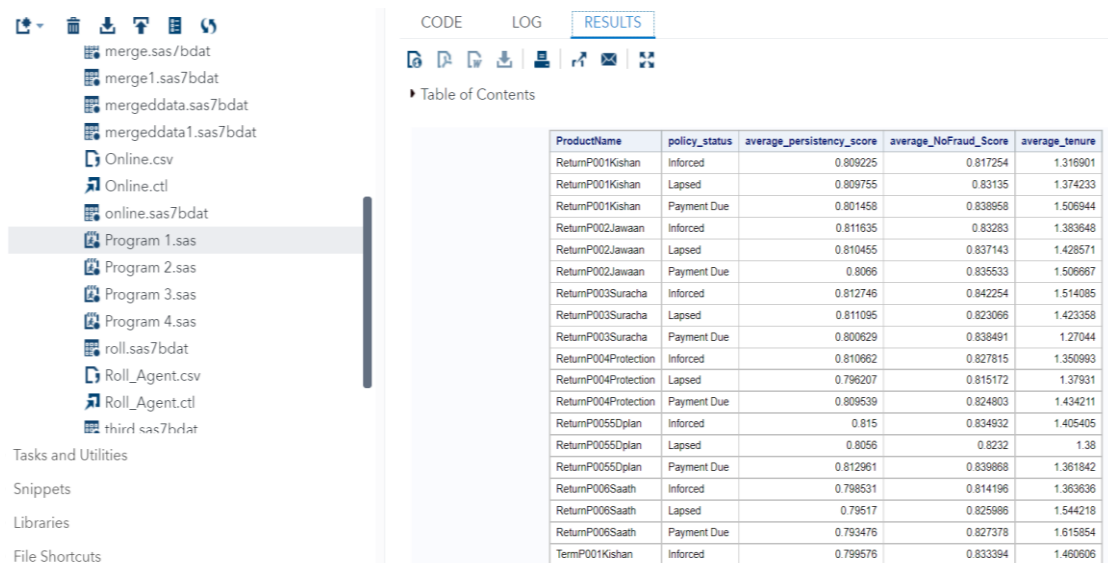
9. Calculate Average persistency score, no fraud score and tenure of customers across product and policy status, and interpret the result?

```
proc sql ;
```

```
select productName,policy_status,avg(persistency_score)as average_persistency_score,  
avg(NoFraud_Score)as average_NoFraud_Score,avg(tenure) as average_tenure  
from SAS.MergedData1
```

```
group by productName,policy_status;
```

```
run;
```



ProductName	policy_status	average_persistency_score	average_NoFraud_Score	average_tenure
ReturnP001Kishan	Forced	0.809225	0.817254	1.316901
ReturnP001Kishan	Lapsed	0.809755	0.83135	1.374233
ReturnP001Kishan	Payment Due	0.801458	0.838958	1.506944
ReturnP002Jawaan	Forced	0.811635	0.83283	1.383648
ReturnP002Jawaan	Lapsed	0.810455	0.837143	1.428571
ReturnP002Jawaan	Payment Due	0.8066	0.835533	1.506667
ReturnP003Suracha	Forced	0.812746	0.842254	1.514085
ReturnP003Suracha	Lapsed	0.811095	0.823066	1.423358
ReturnP003Suracha	Payment Due	0.800629	0.838491	1.27044
ReturnP004Protection	Forced	0.810662	0.827815	1.350993
ReturnP004Protection	Lapsed	0.796207	0.815172	1.37931
ReturnP004Protection	Payment Due	0.809539	0.824803	1.434211
ReturnP0055Diplan	Forced	0.815	0.834932	1.405405
ReturnP0055Diplan	Lapsed	0.8056	0.8232	1.38
ReturnP0055Diplan	Payment Due	0.812961	0.839868	1.361842
ReturnP006Saath	Forced	0.798531	0.814196	1.363636
ReturnP006Saath	Lapsed	0.79517	0.825986	1.544218
ReturnP006Saath	Payment Due	0.793476	0.827378	1.615854
TermP001Kishan	Forced	0.799576	0.833394	1.460606

Great Learning SAS Assessment

Ishita Sarkar

10. Calculate Average age of customer across acquisition channel and policy status, and interpret the result?

```
roc sql;
```

```
select acq_chnl,policy_Status,avg(age) as average_Age
```

```
from SAS.MergedData1
```

```
group by acq_chnl ,policy_status;
```










```
run;
```

```
proc print data=SAS.MergedData1;
```

CODE

LOG

RESULTS

▶ Table of Contents

acq_chnl	policy_status	average_Age
On rol	Inforced	38.06218
On rol	Lapsed	38.42188
On rol	Payment Due	38.23411
Online	Inforced	38.25748
Online	Lapsed	38.20603
Online	Payment Due	38.51695
Third	Inforced	38.99659
Third	Lapsed	38.25135
Third	Payment Due	37.90566

Obs	custid	proposal_num	policy_num	dob	policy_status	Premium	acq_chnl	product_lv1	product_lv2	agentid	payment_mode	policy_d:
1	CC4002098	PP4098	32098	30/06/1990	Inforced	2221	Online	Return	P001 Kishan	AA1	Monthly	16/05/20
2	CC4003609	PP5609	33609	01/10/1969	Inforced	10088	Online	Return	P001 Kishan	AA1	Semi Annual	01/01/20
3	CC4003699	PP5699	33699	27/11/1990	Inforced	10473	Online	Return	P001 Kishan	AA1	Semi Annual	03/01/20
4	CC4004558	PP6558	34558	14/04/1984	Inforced	2042	Online	Return	P001 Kishan	AA1	Monthly	15/02/20
5	CC4001582	PP3582	31582	22/02/1970	Inforced	2484	On rol	Return	P001 Kishan	AA1	Monthly	07/11/20
6	CC4003061	PP5061	33061	28/07/1979	Inforced	10173	On rol	Return	P001 Kishan	AA1	Semi Annual	11/03/20