Problem Statement

- Domain: Insurance
- Business Context - A marketing officer in XYZ insurance company wants to know little important KPI distribution across acquisition channel, payment mode products and policy status. So they can design New financial Year marketing plan and assign expense to the designed plan.
- Dataset description:

On roll agent, Online and Third Party	
Variable Name	Variable Detail
Hhid	Household number assign against policy number
Custid	Customer who have taken policy
proposal_num	Proposal number of the policy
policy_num	Policy num (Unique ID for our case study)
Dob	Date of birth of the policy holder
policy_status	Current policy Status
Premium	Premium amount paid by policy holder on every due date. It is not annual premium.
acq_chnl	Acquisition channel of policy
product_lvl1	First level segment of the product name
product_lvl2	Actual product name with first four letter as a code of product
Agented	Agent who sell the policy to the customer
payment_mode	Payment mode for policy, use this information to calculate annual premium
policy_date	Date of the policy login in the system

Agent Score	
Variable Name	Variable Detail
Agented	Unique ID related with agent
Persistency_Score	If a agent sell 100 policy in last year and out of them 80 are inforced, then persistency score is 0.80
NoFraud_Score	If a agent sell 100 policy in last year and out of them 5 are fraud, then no fraud score is 0.95

- You have found out the all channels owners (Online, OnRoll Agent and third party) store their data in different data source. And training team save agents score on different share drive. So you first job is to collate the data from different sources, clean it and make it ready for analysis.

  Below are the required views by marketing officer
  1. The distribution of customers across product and policy status
  2. Average annual premium for different payment mode
  3. Persistency score, no fraud score and tenure of customers across product and policy status
  4. Average age of customer across acquisition channel and policy status.

- Learning Steps to perform:

Data Preparation Step
1. Import all the 4 files in SAS data environment
2. Append the 3 files except agent score
3. Do a left join on the appended dataset with agent score to get persistency score and no fraud score
4. Remove all the unwanted ID variables
5. Calculate annual premium for all policy
6. Calculate age in year from DOB
7. Calculate tenure in month from policy date
8. Remove product code from product level 2 to extract product name
9. Concat product level 1 and extracted product name to create final product name

Question 1: Import all the 4 files in SAS data environment 8

```
*First file Agent_Score ;
PROC IMPORT
datafile='/home/u48688022/ProjectWeek3/Agent_Score.csv'
out=finance.Agent_Score replace;
delimiter=",";
GETNAMES=yes;
GUESSINGROWS=1000;
run;
*Second File Online.csv;
PROC IMPORT
datafile='/home/u48688022/ProjectWeek3/Online.csv'
out=finance.Online replace;
delimiter=",";
GETNAMES=yes;
GUESSINGROWS=1000;
run;
*Third file Roll_Agent.csv;
PROC IMPORT
datafile='/home/u48688022/ProjectWeek3/Roll_Agent.csv'
out=finance.Roll_Agent replace;
delimiter=",";
GETNAMES=yes;
GUESSINGROWS=1000;
run;
*Fourth File Third_Party.csv;
PROC IMPORT
datafile='/home/u48688022/ProjectWeek3/Roll_Agent.csv'
out=finance.Third_Party replace;
delimiter=",";
GETNAMES=yes;
GUESSINGROWS=1000;
run;
```

Question 2: Create one dataset from all the 4 dataset? (8 Mark)

```
data finance.insurance;
set finance.Online finance.Roll_Agent finance.Third_Party;
run;
Proc Print data=finance.insurance (obs=100);
run;
```

Question 3: Remove all unwanted ID variables? (2 Mark)

```
data finance.insurance (drop = hhid custmid);
set finance.insurance;
run;
```

Question 4: Calculate annual premium for all customers? (4 Mark)

```
data finance.Annual_Premium;
set finance.insurance;
Annual_Premium = 12 * Premium  ;
run;
****************** using proc SQL ***************************;
proc sql ;
select sum(annual_premium) as Annual_Premium from finance.annual_Premium;
quit;
*************************using sum ************************************;
proc print data = finance.Annual_Premium;
sum annual_premium ;
run;
```

Question 5: Calculate age and tenure as of 31 July 2020 for all customers? (4 Marks)

```
data finance.insurance ;
set finance.insurance;
customer_age = intck('year',dob,'31jul2020'D);
customer_tenure= intck('year',policy_date,'31jul2020'D);
run;
```

Question 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? (4 Marks)

```
data finance.insurance;
set finance.insurance;
product_name= compbl(cat(substr(product_lvl2,6,length(product_lvl2)),' ',product_lvl1));
run;
data finance.insurance;
lenght product_name $25;
set finance.insurance;
run;
```

Question 7: After doing clean up in your data, you have to calculate the distribution of customers across product and policy status and interpret the result (4+1 Marks)

```
proc sql;
select count(policy_num) as no_of_customers,Final_product,policy_status
from finance.insurance
group by policy_status,Final_product;
quit;

proc sql;
select count(policy_num) as no_of_customers,Final_product
from finance.insurance
group by Final_product;
quit;

proc sql;
select count(policy_num) as no_of_customers,policy_status
from finance.insurance
group by policy_status;
quit;
```

Most Custmers have Policy Status - Payment Due Most Customer buy policy of Term Kishan

Question 8: Calculate Average annual premium for different payment mode and interpret the result? (4+1 Marks)

```
 proc sql;
select payment_mode, avg(premium)
from finance.insurance
group by payment_mode;
quit;
```

Monthly payment mode customers have heighest annual premium where as Annual payment mode have least annual premium

Question 9: Calculate Average persistency score, no fraud score and tenure of customers across product and policy status, and interpret the result? (4+1 Marks)

```
proc sql;
select avg(Presistency_Score) as Avg_Presistency_score,
    avg(NoFraud_Score) as Avg_NoFraud_Score ,
    avg(Customer_tenure) as Avg_Tenure
from finance.insurance
group by product, policy_status;
quit;
```

Average Presistency score, Average No Fraud Score and Average Tenure across product and policy status

is 0.8, 0.82, 17 months respectively

Question 10: Calculate Average age of customer across acquisition channel and policy status, and interpret the result?(4+1Marks)

```
proc sql;
select acq_chnl,policy_status,avg(Customer_age) as average_customer_age
from finance.insurance
group by acq_chnl, policy_status;
quit;
```

We can interpret that average age of customer across acquisition channel and policy status is 38 years.