```
/*Question1: Importing file*/
proc import datafile = 'C:/Users/rishabh.pandey/Desktop/SAS/Week 2/Practice/Medical Insurance
Dataset.csv'
out=medical_insurance_data dbms=CSV;
run;

/*Question2: Content of the dataset*/
proc contents data=medical_insurance_data varnum;
run;

/*Question3: Outlier checking*/
proc univariate data=medical_insurance_data;
var age Overall_Satisfaction_Score premium Network_hospital_nearby not_passed_percent_claim;
run;

/*we have some outlier and below is the flooring and cappping for those variables*/

data medical_insurance_data;
set medical_insurance_data;
if premium < 16904 then premium = 16904;
if Network_hospital_nearby > 10 then Network_hospital_nearby = 10;
run;

/*checking distribution after flooring and capping*/
proc univariate data=medical_insurance_data;
var age Overall_Satisfaction_Score premium Network_hospital_nearby not_passed_percent_claim;
run;

/*Question4: Checking for missing values*/
proc means data=medical_insurance_data nmiss;
run;

/*Question5: percentile distribution for churners and no churner*/
proc means data=medical_insurance_data n nmiss min p1 p5 p10 p25 p50 p75 p90 p95 p99 max;
var age Overall_Satisfaction_Score premium Network_hospital_nearby not_passed_percent_claim;
run;

/*Questoion6: Check the percentage of class in respective variables*/
proc freq data=medical_insurance_data;
table City_Tier riders Education Gender Payment_freq MaritalStatus Claim_raised_last_year/ nocum;
run;

/*Question7: Create Macro*/
/*Created Marcro*/
%MACRO policy_info();
DATA output (keep = Mobile_num policy_num premium);
SET medical_insurance_data;
where policy_num in (&policy_num.);
RUN;

proc print data=output;
run;
%MEND;

/*Provided input mobile number*/
%let policy_num = 10011,10013;

/*run macro for output*/
%policy_info;

/*Question8: Correlation check*/
proc corr data=medical_insurance_data NOPROB;
var age Overall_Satisfaction_Score premium Network_hospital_nearby not_passed_percent_claim;
run;
```

```
/*Question9: Create train and test sample*/
proc freq data=medical_insurance_data;
table Renewal /nocum;
run;

proc surveyselect data= medical_insurance_data method = srs rep=1
samplesize=600 seed = 1234 out =test;
RUN;

proc contents data=test varnum;
run;

proc freq data=test;
table Renewal /nocum;
run;

proc sql;
create table train as select t1.* from medical_insurance_data as t1
where policy_num not in (select policy_num from test);
quit;

proc freq data=train;
table Renewal /nocum;
run;

/*Question11: Creating base model on numeric variables only*/
%let var =  age Overall_Satisfaction_Score premium Network_hospital_nearby not_passed_percent_claim;

proc logistic data=train descending outmodel=model;
model Renewal = &var / lackfit;
output out = train_output xbeta = coeff stdxbeta = stdcoeff predicted = prob;
run;

/*Question12: Predict test dataset*/

/*Predicting by equation, you can use score statment, in my version of SAS score function is not
present*/
data test;
set test;
prob = -0.0226-0.0398*Age+0.4174*Overall_Satisfaction_Score
-0.00009*Premium+0.0930*Network_hospital_nearby+0.0289*not_passed_percent_claim;
score = exp(prob)/(1+exp(prob));
run;

/*Question13: Creating confusion matrix*/

data train_output;
set train_output;
if prob>0.20 then Renewal_pred = 1;
else Renewal_pred = 0;
run;

data test;
set test;
if score>0.20 then Renewal_pred = 1;
else Renewal_pred = 0;
run;

proc freq data= train_output;
table renewal*renewal_pred / nocol norow nopercent;
run;

proc freq data= test;
```

```
table renewal*renewal_pred / nocol norow nopercent;
run;
```