

```

proc import datafile="/home/u63687835/Projet .Taiwan/default of credit card clients.xls"
    dbms=xls
    out=Work.Taiwan;
    getnames=yes;
    run;

proc contents data=work.taiwan;
run;

/*****Les statistiques descriptives *****/

proc means data=work.taiwan N MEAN STD MIN Q1 MEDIAN Q3 MAX ;
var age BILL_AMT1 BILL_AMT2 BILL_AMT3 BILL_AMT4 BILL_AMT5 BILL_AMT6 LIMIT_BAL;
title"Répartition des clients ";
footnote"Source : Defaut de paiement des clients Taiwanais ";
run;

proc sgplot data=work.taiwan;
    scatter x=Age y=Limit_bal/ markerattrs=(symbol=circlefilled);
    xaxis label="Age ";
    yaxis label="Limit_bal";
    title "Nuage de points pour la relation entre les variables quantitatives";
run;

proc freq data=work.taiwan;
table defaultpayment sex education marriage ;
footnote"Source : Defaut de paiement des clients Taiwanais ";
run;

proc univariate data=work.taiwan;
var defaultpayment ;
title"Statistiques sur le defaut de paiement ";
run;

proc sgplot data=work.taiwan;
vbar defaultpayment/datalabel fillattrs= (color = PINK) categoryorder=respasc;
title "Repartition de clients selon la variable defaut de paiement ";
footnote "Données : défauts de paiement des clients à Taïwan ";
yaxis label="Effectifs";
run;

proc sgplot data=work.taiwan;
vbar Sex/datalabel fillattrs= (color = Red) categoryorder=respasc;
title "Repartition de clients selon la variable defaut de paiement ";
footnote "Données : défauts de paiement des clients à Taïwan ";
yaxis label="Effectifs";
run;

proc sgplot data=work.taiwan;
vbar Marriage/datalabel fillattrs= (color = PINK) categoryorder=respasc;
title "Repartition de clients selon la variable defaut de paiement ";
footnote "Données : défauts de paiement des clients à Taïwan ";
yaxis label="Effectifs";
run;

proc sgplot data=work.taiwan;
vbar education/datalabel fillattrs= (color = PINK) categoryorder=respasc;
title "Repartition de clients selon la variable defaut de paiement ";
footnote "Données : défauts de paiement des clients à Taïwan ";
yaxis label="Effectifs";
run;

proc univariate data=work.taiwan;
histogram Age ;
run;

proc sgplot data=work.taiwan;
    histogram Age/ scale=count datalabel=count binstart=20 binwidth=5 showbins fillattrs=(color="vligb");
    density Age;
    density Age / type=kernel;
    title "Répartition des clients selon l'Âge";
run;

proc sgplot data=work.taiwan;
    vbox limit_bal / category=Age;
    xaxis label="Âge";
    yaxis label="Montant maximum accordé";
    title "Répartition du Montant Maximum Accordé par Tranche d'Âge";
run;

/*****Les corrélations *****/

proc corr data=work.taiwan pearson;
    var _numeric_;
run;

```

```
/*****Autocorrélation des résidus *****/
```

```
GOPTIONS reset=all;  
PROC REG data=work.taiwan;  
MODEL defaultpayment= Age Sex Education/dw ;  
OUTPUT out=file1 p=yhat1 r=res1;  
RUN;
```

```
/*****Statistiques bivariées *****/  
/*****Les Macros *****/
```

```
%macro boxplot(Age, defaultpayment);  
proc sgplot data=work.taiwan;  
vbox &Age / category=&defaultpayment group=&defaultpayment;  
run;  
%mend;  
/*****Application des Macros *****/  
%boxplot(Age,Defaultpayment);  
%boxplot(limit_bal, Defaultpayment);
```

```
/***** Les tests statistiques *****/  
/*****Analyse bivariées *****/
```

```
%macro tableaux_croises(Variable_interet, variable_explicative);  
proc tabulate data=work.taiwan;  
title "Tableau croisé des variables";  
class &variable_explicative &variable_interet;  
table &variable_explicative, &variable_interet * (N Rowpctn);  
footnote "Données : défauts de paiement des clients à Taïwan";  
run;  
%mend;
```

```
%tableaux_croises(Defaultpayment, sex);  
%tableaux_croises(Defaultpayment, Marriage);  
%tableaux_croises(Defaultpayment, Age);  
%tableaux_croises(Defaultpayment, Education);  
/**** Test Khi-deux *****/
```

```
%macro table_freq_chi2(var1, var2);  
proc freq data=work.taiwan;  
tables &var1 * &var2 / chisq;  
run;  
%mend;
```

```
%table_freq_chi2(Sex, Defaultpayment);  
%table_freq_chi2(Age, Defaultpayment);  
%table_freq_chi2(education, Defaultpayment);  
%table_freq_chi2(Marriage, Defaultpayment);
```

```
/*****Test shapiro-wilks : distribution normale *****/
```

```
%macro shapirowilks(var_a_tester);  
proc univariate data =work.taiwan NORMAL;  
var &Var_a_tester;  
run;  
%mend ;
```

```
%shapirowilks(Age);  
%shapirowilks(limit_bal);
```

```
/*****Test de student *****/
```

```
%macro Student_test(var_interet, var_explicative);  
proc ttest data = work.taiwan;  
var &var_explicative;  
run;  
%mend;
```

```
%student_test(Defaultpayment, Age);
```

```
/*****Machine Learning *****/
```

```
/*****Séparation du modèle *****/
```

```
/*****Modèle de base d'apprentissage *****/
```

```
proc surveyselect data=Work.taiwan method= srs seed=2 outall samprate=0.8 out=Work.Taiwan2;  
run;
```

```
/***** Création de la base d'apprentissage *****/
```

```
Data work.Train;  
set work.taiwan2;  
if selected=1;  
run;  
options obs=500;  
Proc print data=Work.Train;
```

```

run;
/***** Création de la base test *****/
Data work.Test;
set work.taiwan2;
if selected=0;
run;
options obs=500;
Proc print data=Work.Test;
run;

options obs=500;
Proc print data=Work.Test;
run;
/*****Logit *****/
proc logistic data=work.train plots=(oddsratio(cldisplay= serifarrow)roc);
class Sex Education Marriage PAY_0 PAY_2 PAY_3 PAY_4 PAY_5 PAY_6 / param=glm;
model defaultpayment(event='1') = Sex Education Marriage PAY_0 PAY_2 PAY_3 PAY_4 PAY_5 PAY_6
Limit_Bal Age
BILL_AMT1 BILL_AMT2 BILL_AMT3 BILL_AMT4 BILL_AMT5 BILL_AMT6
PAY_AMT1 PAY_AMT2 PAY_AMT3 PAY_AMT4 PAY_AMT5 PAY_AMT6
/ link=logit lackfit
selection=backward slstay=0.05
technique=fisher;
score data=work.test out=Work.Prédictions;
run;
/*****Formatage du tab contenant le sprédictions *****/

data work.prédictions(drop= selected F_defaultpayment I_defaultpayment);
set work.prédictions;
length Prediction $10.;

if not missing(P_1) and not missing(P_0) then do;
if P_1 > 0.5 and P_0 < 0.5 then Prediction = "Oui";
else if P_1 <= 0.5 and P_0 >= 0.5 then Prediction = "Non";
else Prediction = "Indéterminé"; /* Gérer les cas non couverts */
end;
else Prediction = "Indéterminé"; /* Gérer les valeurs manquantes */
run;

/*****Matrice de confusion *****/
proc tabulate data=work.prédictions;
title"Matrice de confusion";
class Defaultpayment Prediction;
Table(Defaultpayment),(Prediction)*(N);
run; /*((71+6)/(71+6+12+6))*100 =84,21 % de performance de notre modèle de machine learning */
/*****Test de Hosmer et Lemeshow*****/
H0: L'ajustement du modèle aux données est bon ( si Pvalue>0,05)
H1: L'ajustement du modèle aux données est mauvais (si pvalue <0,05)

On constate que la pvalue est inférieur à 0,05 (0,0001) de ce fait on refuse H0 , alors notre modèle de machine learning r

/*****Affichage *****/
options obs=500;
Proc print data=Work.prédictions;
run;

```