



PROJET A RENDRE

Présenté par : **MAYENE** Bienvenue et **GUERINECHE** Hamza
(*Binome 4*)

L'objectif de la direction Marketing de l'entreprise "Envies d'Ailleurs" sur l'année à venir est de Déterminer les potentiels de croissance de l'entreprise.

Pour ce faire, il faudra établir des analyses de cibles et proposer des actions commerciales de développement.

Notre objectif dans cette étude sera dans un premier temps de déterminer les critères de ciblage des prochaines **offres de séjour pour célibataire** et ensuite de calculer le nombre de clients à cibler, pour envoi des messages par internet.

Choix de la méthode statistique :

Dans notre étude nous aurons à expliquer une variable binaire à l'aide des variables explicatives continues et qualitatives. Ceci s'effectuera par la méthode de régression logistique.

Etape N°1 : Construction de la VAE et réorganisation de la base de données

Construction de la variable à expliquer :

Etant donné que nous travaillons pour les offres de séjours des célibataires, notre variable à expliquer sera donc basée sur les variables de notre base ayant le même thème. Il s'agit ici des variables *celib_F_1* et *celibat_1*. Nous construirons alors notre VAE de la façon suivante :

***** Creation de notre librairie *****;

LIBNAME PROJET "/home/u62726202/PROJET_FINAL";

/* ETAPE 1 : Importation et organisation de la base de donnée */

***** Importation de la base de données *****;

%if %sysfunc(exist(PROJET.'DATA'n)) %then %do;

proc sql;

drop table PROJET.'DATA'n;

run;

%end;

FILENAME REFFILE '/home/u62726202/PROJET_FINAL/ficetude_voyage_seg3.txt';

PROC IMPORT DATAFILE=REFFILE

```

DBMS=DLM

OUT=PROJET.'DATA'n;

DELIMITER=",";

GETNAMES=YES;

RUN;

PROC CONTENTS DATA=PROJET.'DATA'n; RUN;

file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-
r%C3%A9sultats%20(1).html

DATA PROJET.newdata ;

SET PROJET.'DATA'n;

/* Creation de la VAE */

IF (celib_F_1 = 1 ) or (celibat_1 =1) then VAE = 1;

else VAE = 0;

/* Supression des autres variables qui sont en année -1 */

DROP AGENCE_1 COURRIER_1 ETE_1 HIVER_1 CURES_1 FAMILLE_1 GROUPE_1 HAUTGAMME_1
LASTTRAVEL_1 PLAGES_1 SENIOR_1 TRAVEL_BABY_1 FAMILY_1 grou_adul_1 grou_enf_1 hotel_1
parc_lois_1 sej_CAM_1 sej_CAP_1 senior_HG_1 sport_1 thalasso_1 theme_1 trekking2_1
TREKKING_1 ENF_0_2ANS_1 ENF_11_12ANS_1 ENF_13_14ANS_1 ENF_14_16ANS_1 ENF_3_4ANS_1
ENF_5_6ANS_1 ENF_7_8ANS_1 ENF_9_10ANS_1 ENFANT_1;

RUN;

```

Nous avons décidé par la suite de supprimer toutes les autres variables de notre base de données se situant à l'année -1. En effet nous n'utiliserons pas ces variables dans notre modélisation.

Notons que nous allons considérer que les variables *valca_s1* et *NBACT_HA_S1* sont les variables de l'année -1, année à laquelle nous supposons ne pas connaître et pour laquelle nous souhaitons avoir le profil des clients intéressés par l'offre de séjour pour célibataire.

Réorganisation des variables explicatives

Nous allons combiner les variables explicatives en fonction d'une part des 4 dernières années précédant l'année -1 et d'autre part en fonction des 4 dernières années précédant les 4 dernières années précédant l'année -1 c'est à dire les plus de 5 ans dans le passé.

Nous allons construire des variables qui commenceront par **LFY** sigle défini pour “*Last four year*”. Ces variables seront codées de la façon suivante :

```
%macro transfvr1(var2,var3, var4, var5) ;  
  
if (&var2 = 1) or (&var3=1) or (&var4=1) or (&var5 = 1) then LFY_&var2 = 1;  
  
else if (&var2 = 0) and (&var3=0) and (&var4=0) and (&var5 = 0) then LFY_&var2 = 0;  
  
%mend transfvr1;  
  
DATA PROJET.newdata ;  
  
SET PROJET.newdata;  
  
%transfvr1(AGENCE_2, AGENCE_3, AGENCE_4,AGENCE_5);  
  
%transfvr1(COURRIER_2, COURRIER_3, COURRIER_4,COURRIER_5);  
  
%transfvr1(CURES_2, CURES_3, CURES_4,CURES_5);  
  
%transfvr1(ENFANT_2, ENFANT_3, ENFANT_4,ENFANT_5);  
  
%transfvr1(ENF_0_2ANS_2, ENF_0_2ANS_3, ENF_0_2ANS_4,ENF_0_2ANS_5);  
  
%transfvr1(ENF_11_12ANS_2, ENF_11_12ANS_3, ENF_11_12ANS_4,ENF_11_12ANS_5);  
  
%transfvr1(ENF_13_14ANS_2, ENF_13_14ANS_3, ENF_13_14ANS_4,ENF_13_14ANS_5);  
  
%transfvr1(ENF_14_16ANS_2, ENF_14_16ANS_3, ENF_14_16ANS_4,ENF_14_16ANS_5);  
  
%transfvr1(ENF_3_4ANS_2, ENF_3_4ANS_3, ENF_3_4ANS_4,ENF_11_12ANS_5);  
  
%transfvr1(ENF_5_6ANS_2, ENF_5_6ANS_3, ENF_5_6ANS_4,ENF_5_6ANS_5);  
  
%transfvr1(ENF_7_8ANS_2, ENF_7_8ANS_3, ENF_7_8ANS_4,ENF_7_8ANS_5);  
  
%transfvr1(ENF_9_10ANS_2, ENF_9_10ANS_3, ENF_9_10ANS_4,ENF_9_10ANS_5);  
  
%transfvr1(trekking2_2, trekking2_3, trekking2_4,trekking2_5);  
  
%transfvr1(therme_2, therme_3, therme_4,therme_5);  
  
%transfvr1(thalasso_2, thalasso_3, thalasso_4,thalasso_5);  
  
%transfvr1(sport_2, sport_3, sport_4,sport_5);  
  
%transfvr1(senior_HG_2, senior_HG_3, senior_HG_4,senior_HG_5);  
  
%transfvr1(sej_CAP_2, ssej_CAP_3, sej_CAP_4,sej_CAP_5);  
  
%transfvr1(sej_CAM_2, sej_CAM_3, sej_CAM_4,sej_CAM_5);  
  
%transfvr1(perc_lois_2, perc_lois_3, perc_lois_4,perc_lois_5);  
  
%transfvr1(grou_enfant_2, grou_enfant_3, grou_enfant_4,grou_enfant_5);  
  
%transfvr1(grou_adul_2, grou_adul_3, grou_adul_4,grou_adul_5);  
  
%transfvr1(FAMILY_2, FAMILY_3, FAMILY_4,FAMILY_5);
```

```

%transfvr1(FAMILLE_2, FAMILLE_3, FAMILLE_4,FAMILLE_5);
%transfvr1(GROUPE_2, GROUPE_3, GROUPE_4,GROUPE_5);
%transfvr1(HAUTGAMME_2, HAUTGAMME_3, HAUTGAMME_4,HAUTGAMME_5);
%transfvr1(ETE_2, ETE_3, ETE_4,ETE_5);
%transfvr1(HA_2, HA_3, HA_4,HA_5);
%transfvr1(HIVER_2, HIVER_3, HIVER_4,HIVER_5);
%transfvr1(INTERNET_2, INTERNET_3, INTERNET_4,INTERNET_5);
%transfvr1(LASTTRAVEL_2, LASTTRAVEL_3, LASTTRAVEL_4,LASTTRAVEL_5);
%transfvr1(NBACT_HA_2, NBACT_HA_3, NBACT_HA_4,NBACT_HA_5);
%transfvr1(PLAGE_2, PLAGE_3, PLAGE_4,PLAGE_5);
%transfvr1(SENIOR_2, SENIOR_3, SENIOR_4,SENIOR_5);
%transfvr1(TRAVEL_BABY_2, TRAVEL_BABY_3, TRAVEL_BABY_4,TRAVEL_BABY_5);
%transfvr1(trekking_2, trekking_3, trekking_4,trekking_5);
%transfvr1(celeb_f_2, celeb_f_3, celeb_f_4,celeb_f_5);
%transfvr1(celebat_2, celibat_3, celibat_4,celibat_5);
%transfvr1(TRAVEL_BABY_2, TRAVEL_BABY_3, TRAVEL_BABY_4,TRAVEL_BABY_5);
%transfvr1(hotel_2, hotel_3, hotel_4,hotel_5);

```

Ainsi donc ces variables auront des modalités binaires 0 et 1 avec :

Modalité	Libellé
0	<i>Ne concerne aucune des 4 précédentes années</i>
1	<i>Concerne au moins une des 4 précédentes années</i>

Nous allons aussi construire des variables qui commenceront par **MT5Y** sigle défini pour “*More Than 5 Years*”. Ces variables concerneront les évènements qui se sont déroulés plus de 5 ans en arrière et seront transformées de la façon suivante :

```

%macro transfvr1(var2,var3, var4, var5) ;
if (&var2 = 1) or (&var3=1) or (&var4=1) or (&var5 = 1) then MT5Y_&var2 = 1;
else if (&var2 = 0) and (&var3=0) and (&var4=0) and (&var5 = 0) then MT5Y_&var2 = 0;
%mend transfvr2;

```

DATA PROJET.newdata ;

SET PROJET.newdata;

%transfvr1(AGENCE_6, AGENCE_7, AGENCE_8,AGENCE_9);

%transfvr1(COURRIER_6, COURRIER_7, COURRIER_8,COURRIER_9);

%transfvr1(CURES_6, CURES_7, CURES_8,CURES_9);

%transfvr1(ENFANT_6, ENFANT_7, ENFANT_8,ENFANT_9);

%transfvr1(ENF_0_2ANS_6, ENF_0_2ANS_7 ENF_0_2ANS_8,ENF_0_2ANS_9);

%transfvr1(ENF_11_12ANS_6, ENF_11_12ANS_7, ENF_11_12ANS_8,ENF_11_12ANS_9);

%transfvr1(ENF_13_14ANS_6, ENF_13_14ANS_7, ENF_13_14ANS_8,ENF_13_14ANS_9);

%transfvr1(ENF_14_16ANS_6, ENF_14_16ANS_7, ENF_14_16ANS_8,ENF_14_16ANS_9);

%transfvr1(ENF_3_4ANS_6, ENF_3_4ANS_7, ENF_3_4ANS_8,ENF_11_12ANS_9);

%transfvr1(ENF_5_6ANS_6, ENF_5_6ANS_7, ENF_5_6ANS_8,ENF_5_6ANS_9);

%transfvr1(ENF_7_8ANS_6, ENF_7_8ANS_7, ENF_7_8ANS_8,ENF_7_8ANS_9);

%transfvr1(ENF_9_10ANS_6, ENF_9_10ANS_7, ENF_9_10ANS_8,ENF_9_10ANS_9);

%transfvr1(trekking2_6, trekking2_7, trekking2_8,trekking2_9);

%transfvr1(therme_6, therme_7, therme_8,therme_9);

%transfvr1(thalasso_6, thalasso_7, thalasso_8,thalasso_9);

%transfvr1(sport_6, sport_7, sport_8,sport_9);

%transfvr1(senior_HG_6, senior_HG_7, senior_HG_8,senior_HG_9);

%transfvr1(sej_CAP_6, sej_CAP_7, sej_CAP_8,sej_CAP_9);

%transfvr1(sej_CAM_6, sej_CAM_7, sej_CAM_8,sej_CAM_9);

%transfvr1(perc_lois_6, perc_lois_7, perc_lois_8,perc_lois_9);

%transfvr1(grou_enfant_6, grou_enfant_7, grou_enfant_8,grou_enfant_9);

%transfvr1(grou_adul_6, grou_adul_7, grou_adul_8,grou_adul_9);

%transfvr1(FAMILY_6, FAMILY_7, FAMILY_8,FAMILY_9);

%transfvr1(FAMILLE_6, FAMILLE_7, FAMILLE_8,FAMILLE_9);

%transfvr1(GROUPE_6, GROUPE_7, GROUPE_8,GROUPE_9);

%transfvr1(HAUTGAMME_6, HAUTGAMME_7, HAUTGAMME_8,HAUTGAMME_9);

%transfvr1(ETE_6, ETE_7, ETE_8,ETE_9);

```

%transfvr1(HA_6, HA_7, HA_8,HA_9);
%transfvr1(HIVER_6, HIVER_7, HIVER_8,HIVER_9);
%transfvr1(INTERNET_6, INTERNET_7, INTERNET_8,INTERNET_9);
%transfvr1(LASTTRAVEL_6, LASTTRAVEL_7, LASTTRAVEL_8,LASTTRAVEL_9);
%transfvr1(NBACT_HA_6, NBACT_HA_7, NBACT_HA_8,NBACT_HA_9);
%transfvr1(PLAGE_6, PLAGE_7, PLAGE_8,PLAGE_9);
%transfvr1(SENIOR_6, SENIOR_7, SENIOR_8,SENIOR_9);
%transfvr1(TRAVEL_BABY_6, TRAVEL_BABY_7, TRAVEL_BABY_8,TRAVEL_BABY_9);
%transfvr1(trekking_6, trekking_7, trekking_8,trekking_9);
%transfvr1(celeb_f_6, celeb_f_7, celeb_f_8,celeb_f_9);
%transfvr1(ceibat_6, ceibat_7, ceibat_8,ceibat_9);
%transfvr1(TRAVEL_BABY_6, TRAVEL_BABY_7, TRAVEL_BABY_8,TRAVEL_BABY_9);
%transfvr1(hotel_6, hotel_7, hotel_8,hotel_9);

```

Ainsi donc ces variables auront aussi des modalités binaires 0 et 1 avec:

Modalité	Libellé
0	<i>Ne concerne aucune des années d'il y'a plus de 5 ans.</i>
1	<i>Concerne au moins une des années d'il y'a plus de 5 ans</i>

Nous allons maintenant créer une nouvelle base de données dans laquelle nous supprimerons des variables et nous transformerons nos variables quantitatives en variables qualitatives .

```
/***** Base de données finale *****/
```

```
DATA PROJET.BASE_FINALE;
```

```
set PROJET.newdata;
```

```
DROP AGENCE_2 AGENCE_3 AGENCE_4 AGENCE_5
```

```
COURRIER_2 COURRIER_3 COURRIER_4 COURRIER_5
```

```
CURES_2 CURES_3 CURES_4 CURES_5
```

```
ENFANT_2 ENFANT_3 ENFANT_4 ENFANT_5 ENF_0_2ANS_2 ENF_0_2ANS_3
```

```
ENF_0_2ANS_4 ENF_0_2ANS_5 ENF_11_12ANS_2 ENF_11_12ANS_3
```

ENF_11_12ANS_4 ENF_11_12ANS_5 ENF_13_14ANS_2 ENF_13_14ANS_3
ENF_13_14ANS_4 ENF_13_14ANS_5 ENF_14_16ANS_2 ENF_14_16ANS_3 ENF_14_16ANS_4
ENF_14_16ANS_5
ENF_3_4ANS_2 ENF_3_4ANS_3 ENF_3_4ANS_4 ENF_3_4ANS_5
ENF_5_6ANS_2 ENF_5_6ANS_3 ENF_5_6ANS_4 ENF_5_6ANS_5
ENF_7_8ANS_2 ENF_7_8ANS_3 ENF_7_8ANS_4 ENF_7_8ANS_5
ENF_9_10ANS_2 ENF_9_10ANS_3 ENF_9_10ANS_4 ENF_9_10ANS_5
trekking2_2 trekking2_3 trekking2_4 trekking2_5
therme_2 therme_3 therme_4 therme_5
thalasso_2 thalasso_3 thalasso_4 thalasso_5
sport_2 sport_3 sport_4 sport_5
senior_HG_2 senior_HG_3 senior_HG_4 senior_HG_5
sej_CAP_2 sej_CAP_3 sej_CAP_4 sej_CAP_5
sej_CAM_2 sej_CAM_3 sej_CAM_4 sej_CAM_5
parc_lois_2 parc_lois_3 parc_lois_4 parc_lois_5
grou_enf_2 grou_enf_3 grou_enf_4 grou_enf_5
grou_adul_2 grou_adul_3 grou_adul_4 grou_adul_5
FAMILY_2 FAMILY_3 FAMILY_4 FAMILY_5
FAMILLE_2 FAMILLE_3 FAMILLE_4 FAMILLE_5
GROUPE_2 GROUPE_3 GROUPE_4 GROUPE_5
HAUTGAMME_2 HAUTGAMME_3 HAUTGAMME_4 HAUTGAMME_5
ETE_2 ETE_3 ETE_4 ETE_5 HA_2 HA_3 HA_4 HA_5
HIVER_2 HIVER_3 HIVER_4 HIVER_5
INTERNET_2 INTERNET_3 INTERNET_4 INTERNET_5
LASTTRAVEL_2 LASTTRAVEL_3 LASTTRAVEL_4 LASTTRAVEL_5
NBACT_HA_2 NBACT_HA_3 NBACT_HA_4 NBACT_HA_5
PLAGE_2 PLAGE_3 PLAGE_4 PLAGE_5
SENIOR_2 SENIOR_3 SENIOR_4 SENIOR_5
TRAVEL_BABY_2 TRAVEL_BABY_3 TRAVEL_BABY_4 TRAVEL_BABY_5
trekking_2 trekking_3 trekking_4 trekking_5
celib_f_2 celib_f_3 celib_f_4 celib_f_5

celibat_2 celibat_3 celibat_4 celibat_5
TRAVEL_BABY_2 TRAVEL_BABY_3 TRAVEL_BABY_4 TRAVEL_BABY_5
hotel_2 hotel_3 hotel_4 hotel_5
AGENCE_6 AGENCE_7 AGENCE_8 AGENCE_9
COURRIER_6 COURRIER_7 COURRIER_8 COURRIER_9
CURES_6 CURES_7 CURES_8 CURES_9
ENFANT_6 ENFANT_7 ENFANT_8 ENFANT_9
ENF_0_2ANS_6 ENF_0_2ANS_7 ENF_0_2ANS_8 ENF_0_2ANS_9
ENF_11_12ANS_6 ENF_11_12ANS_7 ENF_11_12ANS_8 ENF_11_12ANS_9
ENF_13_14ANS_6 ENF_13_14ANS_7 ENF_13_14ANS_8 ENF_13_14ANS_9
ENF_14_16ANS_6 ENF_14_16ANS_7 ENF_14_16ANS_8 ENF_14_16ANS_9
ENF_3_4ANS_6 ENF_3_4ANS_7 ENF_3_4ANS_8 ENF_3_4ANS_9
ENF_5_6ANS_6 ENF_5_6ANS_7 ENF_5_6ANS_8 ENF_5_6ANS_9
ENF_7_8ANS_6 ENF_7_8ANS_7 ENF_7_8ANS_8 ENF_7_8ANS_9
ENF_9_10ANS_6 ENF_9_10ANS_7 ENF_9_10ANS_8 ENF_9_10ANS_9
trekking2_6 trekking2_7 trekking2_8 trekking2_9
therme_6 therme_7 therme_8 therme_9
thalasso_6 thalasso_7 thalasso_8 thalasso_9
sport_6 sport_7 sport_8 sport_9
senior_HG_6 senior_HG_7 senior_HG_8 senior_HG_9
sej_CAP_6 sej_CAP_7 sej_CAP_8 sej_CAP_9
sej_CAM_6 sej_CAM_7 sej_CAM_8 sej_CAM_9
parc_lois_6 parc_lois_7 parc_lois_8 parc_lois_9
grou_enf_6 grou_enf_7 grou_enf_8 grou_enf_9
grou_adul_6 grou_adul_7 grou_adul_8 grou_adul_9
FAMILY_6 FAMILY_7 FAMILY_8 FAMILY_9
FAMILLE_6 FAMILLE_7 FAMILLE_8 FAMILLE_9
GROUPE_6 GROUPE_7 GROUPE_8 GROUPE_9
HAUTGAMME_6 HAUTGAMME_7 HAUTGAMME_8 HAUTGAMME_9
ETE_6 ETE_7 ETE_8 ETE_9
HA_6 HA_7 HA_8 HA_9

```

HIVER_6 HIVER_7 HIVER_8 HIVER_9
INTERNET_6 INTERNET_7 INTERNET_8 INTERNET_9
LASTTRAVEL_6 LASTTRAVEL_7 LASTTRAVEL_8 LASTTRAVEL_9
NBACT_HA_6 NBACT_HA_7 NBACT_HA_8 NBACT_HA_9
PLAGE_6 PLAGE_7 PLAGE_8 PLAGE_9
SENIOR_6 SENIOR_7 SENIOR_8 SENIOR_9
TRAVEL_BABY_6 TRAVEL_BABY_7 TRAVEL_BABY_8 TRAVEL_BABY_9
trekking_6 trekking_7 trekking_8 trekking_9
celib_f_6 celib_f_7 celib_f_8 celib_f_9
celibat_6 celibat_7 celibat_8 celibat_9
TRAVEL_BABY_6 TRAVEL_BABY_7 TRAVEL_BABY_8 TRAVEL_BABY_9
hotel_6 hotel_7 hotel_8 hotel_9;
RUN;
PROC CONTENTS DATA= PROJET.BASE_FINALE; RUN;
/***** Transformation de la variable agereel *****/
data PROJET.BASE_FINALE;
set PROJET.BASE_FINALE;
If agereel < 20 then AGE = "Moins de 20 ans";
else If agereel >= 20 and agereel < 30 then AGE = "[20;30[";
else If agereel >= 30 and agereel < 40 then AGE = "[30;40[";
else If agereel >= 40 and agereel < 50 then AGE = "[40;50[";
else If agereel >= 50 and agereel < 60 then AGE = "[50;60[";
else If agereel >= 60 and agereel < 70 then AGE = "[60;70[";
else If agereel >= 70 and agereel < 80 then AGE = "[70;80[";
else If agereel >= 80 and agereel < 90 then AGE = "[80;90[";
else If agereel >= 90 and agereel < 100 then AGE = "[90;100[";
else If agereel >= 90 and agereel < 120 then AGE = "Plus de 100 ans";
else If agereel >= 2004 then AGE = "Non renseigné";
Drop agereel;
if civilite = "Mo" then civilite = "Homme";
else if civilite = "Ma" then civilite = "Femme";

```

```

if tranche_urbaine = "Anomalies" then tranche_urbaine = "Non affecté";

run;

/***** Création et Transformation de la variable LFY_valca *****/

data PROJET.BASE_FINALE2;

set PROJET.BASE_FINALE;

LFY_valca = valca_s2 + valca_s3 + valca_s4 + valca_s5;

run;

proc means data=PROJET.BASE_FINALE2;

var LFY_valca;

ods output Summary=proc_means;

run;

file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-
r%C3%A9sultats%20(2).html

data PROJET.BASE_FINALE2;

set PROJET.BASE_FINALE2;

If LFY_valca < 5000 then LFY_valca2 = "<5K";

else If LFY_valca >= 5000 and LFY_valca < 10000 then LFY_valca2 = "5K-10K";

else If LFY_valca >= 10000 and LFY_valca < 15000 then LFY_valca2 = "10K-15K";

else If LFY_valca >= 15000 and LFY_valca < 20000 then LFY_valca2 = "15K-20K";

else If LFY_valca >= 20000 and LFY_valca < 25000 then LFY_valca2 = "20K-25K";

else If LFY_valca >= 25000 and LFY_valca < 30000 then LFY_valca2 = "25K-30K";

else If LFY_valca >= 30000 then LFY_valca2 = ">30K" ;

Drop valca_s2 valca_s3 valca_s4 valca_s5 LFY_valca;

run;

proc freq data = projet.base_finale2;

tables LFY_valca2;

run;

file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-
r%C3%A9sultats%20(3).html

PROC CONTENTS DATA= PROJET.BASE_FINALE2; RUN;

file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-
r%C3%A9sultats%20(4).html

```

Etape N°2 : Exploration des données

****** ETAPE 2 : Exploration des donnees ***** ;*

***** analyse univariée des variables qualitatives**** ;*

proc template;

define statgraph SASStudio.Pie;

begingraph;

entrytitle "Répartition de la souscription à une offre de séjour pour célibataires durant les 4 dernières années";

layout region;

piechart category= civileite /;

endlayout;

endgraph;

end;

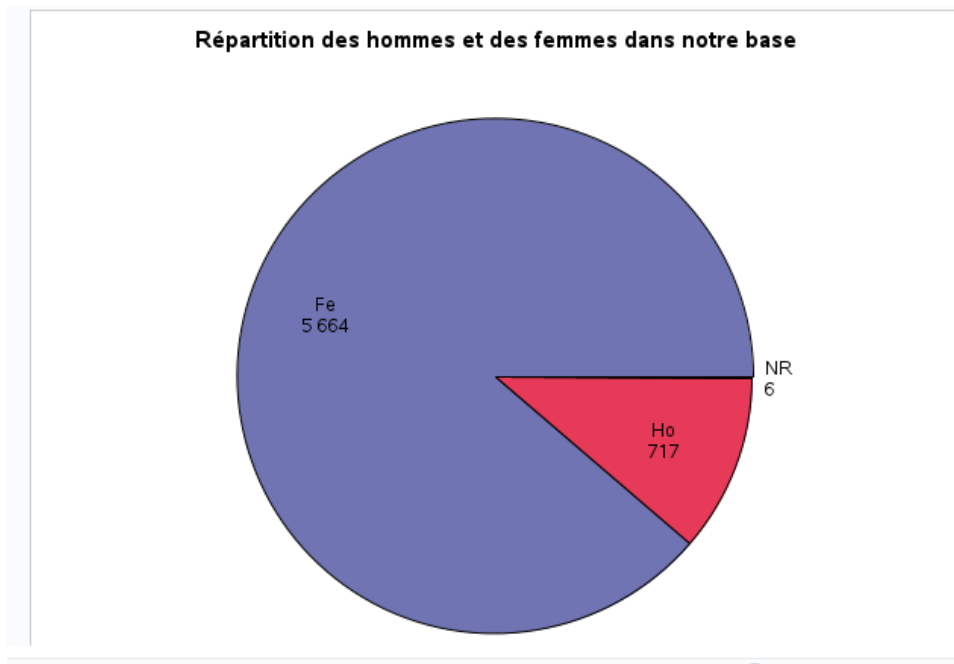
run;

ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgrender template=SASStudio.Pie data=PROJET.BASE_FINALE;

run;

ods graphics / reset;



D'après le graphique ci-dessus, nous constatons que nous avons beaucoup plus de femmes que d'hommes. En effet, le nombre de femme s'élève à 5664 soit **88,68%** et le nombre d'homme s'élève à 717 soit **11,22%**.

```
proc freq data = projet.base_finale2;
```

```
tables LFY_celib_f_2 MT5Y_celib_f_6 lfy_valca_2 Age Region ;
```

```
run;
```

file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20(5).html

D'après les tableaux obtenus en cliquant sur le lien ci-dessus, nous pouvons voir que ce soit il y'a 4 ans ou plus, peu de personnes ont pris l'offre de séjour pour célibataires pour femmes.

Le montant le plus dépensé par les clients est inférieur à 5000.

Sur les âges renseignés nous remarquons que ce sont les personnes âgées entre 30 et 40 ans qui souscrivent le plus aux offres de séjour de l'entreprise. Ces personnes résident majoritairement en Ile-De-France.

****** analyse bivariable des variables qualitatives**** ;**

```
proc corr data= projet.base_finale2;
```

```
run;
```

Lien pour l'observation des résultats :

file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20(6).html

Etape N°3 : Relation entre la VAE et les variables explicatives

L'étape suivante permettant la réalisation d'un Scoring, consiste à construire des tris croisés entre la VAE (Variable A Expliquer) et les variables explicatives qualitatives (les variables quantitatives ayant été transformées) à l'aide du test du khi-deux afin de déterminer s'il existe une liaison ou non entre les variables.

Si toutefois la liaison entre deux variables était vérifiée, l'intensité en serait mesurée avec la statistique V de Cramer.

Principe du test du khi-deux :

On cherche à vérifier l'indépendance ou non entre deux variables :

Hypothèses du test :

H_0 : X et Y sont indépendants vs H_1 : X et Y ne sont pas indépendants

Le seuil étant fixé à 5%

Règle de décision : Si $p < 0.05$ --> Rejet de l'hypothèse nulle d'indépendance.

Dans ce cas nous nous intéresserons alors à la valeur du V de Cramer, afin de mesurer l'intensité de la relation.

***** ETAPE 3 : Relation entre la VAE et les variables explicatives ***** ;

/***** Test du chi-2 *****/

%macro frq_chisq(var);

*ods trace on;

ods output ChiSq=sample_resultats;

PROC FREQ DATA=projet.base_finale2;

TABLES VAE* &var/ CHISQ noprint; RUN;

data val_chisq;

```

set sample_resultats;

if _n_=1; keep table prob;run;

data vcramer;

set sample_resultats;

if _n_=6; keep table value;run;

data sample_resultats;

merge val_chisq vcramer;run;

proc append base = chisq_quali

data=sample_resultats force; run;

%mend frq_chisq;

%frq_chisq(region);

%frq_chisq(civilite);

%frq_chisq(type_habitat);

%frq_chisq(tranche_urbaine);

%frq_chisq(LFY_AGENCE_2);

%frq_chisq(LFY_COURRIER_2);

%frq_chisq(LFY_CURES_2);

%frq_chisq(LFY_ENFANT_2);

%frq_chisq(LFY_ENF_0_2ANS_2);

%frq_chisq(LFY_ENF_11_12ANS_2);

%frq_chisq(LFY_ENF_13_14ANS_2);

%frq_chisq(LFY_ENF_14_16ANS_2);

%frq_chisq(LFY_ENF_3_4ANS_2);

%frq_chisq(LFY_ENF_5_6ANS_2);

%frq_chisq(LFY_ENF_7_8ANS_2);

```

%frq_chisq(LFY_ENF_9_10ANS_2);

%frq_chisq(LFY_trekking2_2);

%frq_chisq(LFY_therme_2);

%frq_chisq(LFY_thalasso_2);

%frq_chisq(LFY_sport_2);

%frq_chisq(LFY_senior_HG_2);

%frq_chisq(LFY_sej_CAP_2);

%frq_chisq(LFY_sej_CAM_2);

%frq_chisq(LFY_parc_lois_2);

%frq_chisq(LFY_grou_enf_2);

%frq_chisq(LFY_grou_adul_2);

%frq_chisq(LFY_FAMILY_2);

%frq_chisq(LFY_FAMILLE_2);

%frq_chisq(LFY_GROUPE_2);

%frq_chisq(LFY_HAUTGAMME_2);

%frq_chisq(LFY_ETE_2);

%frq_chisq(LFY_HA_2);

%frq_chisq(LFY_HIVER_2);

%frq_chisq(LFY_INTERNET_2);

%frq_chisq(LFY_LASTTRAVEL_2);

%frq_chisq(LFY_NBACT_HA_2);

%frq_chisq(LFY_PLAGE_2);

%frq_chisq(LFY_SENIOR_2);

%frq_chisq(LFY_TRAVEL_BABY_2);

%frq_chisq(LFY_trekking_2);

%frq_chisq(LFY_celib_f_2);
%frq_chisq(LFY_celibat_2);
%frq_chisq(LFY_hotel_2);
%frq_chisq(LFY_nbact_ha_s2);
%frq_chisq(MT5Y_AGENCE_6);
%frq_chisq(MT5Y_COURRIER_6);
%frq_chisq(MT5Y_CURES_6);
%frq_chisq(MT5Y_ENFANT_6);
%frq_chisq(MT5Y_ENF_0_2ANS_6);
%frq_chisq(MT5Y_ENF_11_12ANS_6);
%frq_chisq(MT5Y_ENF_13_14ANS_6);
%frq_chisq(MT5Y_ENF_14_16ANS_6);
%frq_chisq(MT5Y_ENF_3_4ANS_6);
%frq_chisq(MT5Y_ENF_5_6ANS_6);
%frq_chisq(MT5Y_ENF_7_8ANS_6);
%frq_chisq(MT5Y_ENF_9_10ANS_6);
%frq_chisq(MT5Y_trekking2_6);
%frq_chisq(MT5Y_therme_6);
%frq_chisq(MT5Y_thalasso_6);
%frq_chisq(MT5Y_sport_6);
%frq_chisq(MT5Y_senior_HG_6);
%frq_chisq(MT5Y_sej_CAP_6);
%frq_chisq(MT5Y_sej_CAM_6);
%frq_chisq(MT5Y_parc_lois_6);
%frq_chisq(MT5Y_grou_enf_6);

```

%frq_chisq(MT5Y_grou_adul_6);

%frq_chisq(MT5Y_FAMILY_6);

%frq_chisq(MT5Y_FAMILLE_6);

%frq_chisq(MT5Y_GROUPE_6);

%frq_chisq(MT5Y_HAUTGAMME_6);

%frq_chisq(MT5Y_ETE_6);

%frq_chisq(MT5Y_HA_6);

%frq_chisq(MT5Y_HIVER_6);

%frq_chisq(MT5Y_INTERNET_6);

%frq_chisq(MT5Y_LASTTRAVEL_6);

%frq_chisq(MT5Y_NBACT_HA_6);

%frq_chisq(MT5Y_PLAGE_6);

%frq_chisq(MT5Y_SENIOR_6);

%frq_chisq(MT5Y_TRAVEL_BABY_6);

%frq_chisq(MT5Y_trekking_6);

%frq_chisq(MT5Y_celib_f_6);

%frq_chisq(MT5Y_celibat_6);

%frq_chisq(MT5Y_hotel_6);

PROC PRINT DATA =chisq_quali;run;

```

Lien pour l'observation des résultats :

file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20(7).html

file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/Etapes_Scoring-r%C3%A9sultats%20(5).html

D'après les résultats obtenus, nous constatons que seul les p-values des variables suivantes sont inférieures au seuil. Il s'agit des variables :

Nom des variables	P-value	V de cramer
-------------------	---------	-------------

civilite	<.0001	0.1073
LFY_ENF_14_16ANS_2	0.0226	0.0285
LFY_grou_enf_2	<.0001	0.0863
LFY_FAMILY_2	0.0412	0.0255
LFY_celib_f_2	0.0001	0.0527
LFY_celibat_2	<.0001	0.1458
LFY_hotel_2	0.0015	0.0396
MT5Y_ENF_14_16ANS_6	0.0478	0.0248
MT5Y_celibat_6	<.0001	0.0859
MT5Y_celib_f_6	0.0258	0.0279
LFY_sej_CAP_2	0.0036	0.0364
MT5Y_AGENCE_6	0.0357	0.0263
MT5Y_trekking2_6	0.0385	0.0259
MT5Y_therme_6	0.0268	0.0277
MT5Y_sej_CAP_6	<.0001	0.0493

D'après le tableau ci-dessus nous pouvons dire que les variables les plus liées à notre VAE sont les variables *civilite*, *LFY_grou_enf_2*, *LFY_celib_f_2*, *LFY_celibat_2*

et *MT5Y_celibat_6* car elles ont un V de Cramer le plus élevé.

Les variables les moins liées à notre VAE sont les variables *LFY_ENF_14_16ANS_2*, *LFY_ENF_14_16ANS_2*, *MT5Y_sej_CAP_6*, *MT5Y_therme_6*, *MT5Y_trekking2_6*, *MT5Y_AGENCE_6*, *LFY_sej_CAP_2*, *MT5Y_celib_f_6*, *LFY_hotel_2* et *LFY_FAMILY_2* car elles ont un V de Cramer le plus faible.

Etape N°4 : Analyse du profil des souscripteurs à l'offre de séjour pour célibataires en année -1

/***** ETAPE 4: Analyse du profil des répondants *****/

```
%macro graphique1(var) ;
```

```
PROC SGPLOT DATA = projet.base_finale2 ;
```

```
VBAR &var / GROUP = VAE groupdisplay= cluster ;
```

```
RUN ;
```

```
%graphique1(civilite);
```

```
%graphique1(LFY_FAMILY_2);
```

```
%graphique1(LFY_celib_f_2);
```

%graphique1(LFY_celibat_2);
 %graphique1(LFY_grou_enf_2);
 %graphique1(LFY_hotel_2);
 %graphique1(LFY_ENF_14_16ANS_2);
 %graphique1(MT5Y_ENF_14_16ANS_6);
 %graphique1(MT5Y_celib_f_6);
 %graphique1(MT5Y_celibat_6);
 %graphique1(LFY_sej_CAP_2);
 %graphique1(MT5Y_AGENCE_6);
 %graphique1(MT5Y_trekking2_6);
 %graphique1(MT5Y_therme_6);
 %graphique1(MT5Y_sej_CAP_6);

Lien vers les résultats :

[file:///home/users/etudiant/T%C3%A9%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20\(8\).html](file:///home/users/etudiant/T%C3%A9%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20(8).html)

Nous avons utilisé ces graphiques en **comparant les modalités des variables représentées en diagramme selon la couleur rouge** qui représente le fait qu'on ait opté pour l'offre de séjour pour célibataire en année -1 c'est à dire VAE = 1.

D'après les graphiques nous pouvons voir que les clients ayant répondu à l'offre de séjour pour célibataire en année -1 sont majoritairement :

- Des femmes
- Ceux qui n'ont pas pris de type de voyage "famille" durant les 4 dernières années.
- Ceux qui en grande partie n'ont pas pris de type de voyage "célibataire femme " durant les 4 dernières années
- Ceux qui n'ont pas pris de type de voyage "célibataire" durant les 4 dernières années
- Ceux qui n'ont pas pris de type de voyage "groupe enfant" durant les 4 dernières années
- Ceux qui n'ont pas pris de type de voyage "hotel" durant les 4 dernières années
- Ceux qui n'ont pas pris de type de voyage "enfant 14-16 ans" durant les 4 dernières années
- Ceux qui n'ont pas pris de type de voyage "enfant 14-16" plus de 5 ans en arrière.
- Ceux qui n'ont pas pris de type de voyage "célibataire" plus de 5 ans en arrière.
- Ceux qui n'ont pas pris de type de voyage "célibataire femme" plus de 5 ans en arrière.
- Ceux qui n'ont pas pris de type de voyage "séjour à la capitale" durant les 4 dernières années

- Ceux qui n'ont pas utilisé l'agence comme canal d'achat plus de 5 ans en arrière.
- Ceux qui n'ont pas pris de type de voyage "trekking2", "therme" et "séjour à la capitale" plus de 5 ans en arrière.

Nous voyons malheureusement que mis à part le critère femme, les autres critères ne sont pas très clairs et ne nous aident pas vraiment dans notre ciblage.

Etape N°5 : Modélisation

Pour effectuer notre modélisation nous allons :

- Diviser notre fichier de travail en deux afin d'obtenir un fichier d'apprentissage et un fichier de validation.
- Générer un premier modèle avec toutes les variables.
- Supprimer les variables non significatives dans le modèle.
- Créer un meilleur modèle et analyser les résultats.

Division du fichier

Cette étape consiste à créer deux fichiers à partir de notre fichier de travail :

- Un fichier d'apprentissage qui contient les 2/3 du fichier total. C'est à partir de ce fichier d'apprentissage que nous allons générer un modèle.
- Un fichier de validation qui contient le tiers restant et qui servira à vérifier la robustesse du modèle qui a été généré dans la phase d'apprentissage.

Ces deux fichiers sont créés de manière aléatoire. On obtient un fichier d'apprentissage contenant 4454 individus et un fichier de validation contenant 1933 individus.

```
/***** ETAPE 5: Modélisation *****/
```

```
Data Projet.apprentissage Projet.test ;
```

```
set projet.base_finale2 ;
```

```
If uniform(0)<=0.7 then output projet.apprentissage ;
```

```
Else output Projet.test ;
```

```
Run ;
```

```
proc logistic data=Projet.apprentissage descending outmodel= projet.modelV1;
```

```
class civilite LFY_grou_enf_2 LFY_celib_f_2 LFY_celibat_2 MT5Y_celibat_6
```

LFY_ENF_14_16ANS_2 MT5Y_sej_CAP_6 MT5Y_therme_6 MT5Y_trekking2_6 MT5Y_AGENCE_6
LFY_sej_CAP_2

MT5Y_celib_f_6 LFY_hotel_2 LFY_FAMILY_2 MT5Y_ENF_14_16ANS_6 /param=ref;

model VAE = civilite LFY_grou_enf_2 LFY_celib_f_2 LFY_celibat_2 MT5Y_celibat_6

LFY_ENF_14_16ANS_2 MT5Y_sej_CAP_6 MT5Y_therme_6 MT5Y_trekking2_6 MT5Y_AGENCE_6
LFY_sej_CAP_2

MT5Y_celib_f_6 LFY_hotel_2 LFY_FAMILY_2 MT5Y_ENF_14_16ANS_6 ;

output out = Projet.apprentissage_V1 predprobs = (individual);

run ;

[file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20\(10\).html](file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20(10).html)

Test de l'hypothèse nulle globale : BETA=0			
Test	khi-2	DDL	Pr > khi-2
Rapport de vrais	202.3514	16	<.0001
Score	231.2917	16	<.0001
Wald	203.1927	16	<.0001

Analyse des effets Type 3			
Effet	DDL	Khi-2 de Wald	Pr > khi-2
civilite	2	54.1561	<.0001
LFY_grou_enf_2	1	12.2924	0.0005
LFY_celib_f_2	1	7.1387	0.0075
LFY_celibat_2	1	47.7141	<.0001
MT5Y_celibat_6	1	21.3357	<.0001
LFY_ENF_14_16ANS_2	1	1.6094	0.2046
MT5Y_sej_CAP_6	1	7.6546	0.0057
MT5Y_therme_6	1	1.2520	0.2632
MT5Y_trekking2_6	1	1.0219	0.3121
MT5Y_AGENCE_6	1	2.5223	0.1122
LFY_sej_CAP_2	1	1.1924	0.2749
MT5Y_celib_f_6	1	2.6782	0.1017
LFY_hotel_2	1	2.5713	0.1088
LFY_FAMILY_2	1	1.3184	0.2509
MT5Y_ENF_14_16ANS_6	1	2.5756	0.1085

Nous allons d'abord vérifier la validation de notre modèle grâce au test du ratio de vraisemblance (Likelihood Ratio) :

Ce test révèle l'acceptation du modèle, en effet au niveau du premier tableau, nous pouvons voir ***Pr> khi-2 <0.001 pour le rapport de vraisemblance.***

Cependant nous constatons que certaines variables ne sont pas significatives dans notre modèle. En effet, le test du khi-deux, décrit précédemment, nous permet d'affirmer que les variables *LFY_FAMILY_2*, *MT5Y_therme_6* *MT5Y_trekking2_6* *MT5Y_AGENCE_6* *LFY_sej_CAP_2*, *LFY_ENF_14_16ANS_2*, *LFY_hotel_2* et *MT5Y_celib_f_6* n'ont pas d'importance dans la relation que nous modélisons.

Le test du khi-deux nous donne des résultats très largement supérieurs à 0.05.

L'hypothèse d'indépendance entre ces variables et notre variable à expliquer ne peut pas être rejetée.

Nous allons donc générer un nouveau modèle avec toutes les autres variables sauf celles-ci.

*/** Generons un autre modèle *****/*

```
proc logistic data=projet.apprentissage plot(maxpoints=none)=all
```

```
descending outmodel= projet.modelV;
```

```
class civilite LFY_grou_enf_2 LFY_celibat_2 MT5Y_celibat_6
```

```
MT5Y_sej_CAP_6 LFY_hotel_2 /param=ref ;
```

```
model VAE = civilite LFY_grou_enf_2 LFY_celibat_2 MT5Y_celibat_6
```

```
MT5Y_sej_CAP_6 LFY_hotel_2 ;
```

```
output out = projet.apprentissage_V1 predprobs = (individual);
```

```
run ;
```

[file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20\(11\).html](file:///home/users/etudiant/T%C3%A9l%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20(11).html)

Statistique d'ajustement du modèle		
Critère	Constante uniquement	Constante et Covariables
AIC	3391.116	3220.696
SC	3397.512	3271.864
-2 Log L	3389.116	3204.696

Test de l'hypothèse nulle globale : BETA=0			
Test	khi-2	DDL	Pr > khi-2
Rapport de vrais	184.4201	7	<.0001
Score	214.0820	7	<.0001
Wald	188.5723	7	<.0001

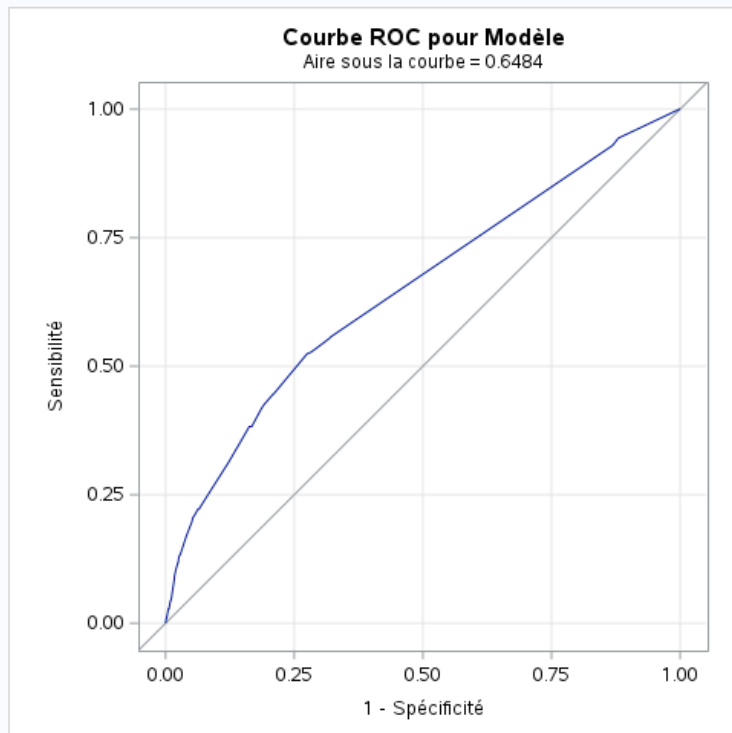
Analyse des effets Type 3			
Effet	DDL	Khi-2 de Wald	Pr > khi-2
civilite	2	57.6391	<.0001
LFY_grou_enf_2	1	12.4611	0.0004
LFY_celibat_2	1	48.9570	<.0001
MT5Y_celibat_6	1	21.1325	<.0001
MT5Y_sej_CAP_6	1	15.7489	<.0001
LFY_celib_f_2	1	7.3589	0.0067

Nous validons notre modèle grâce au test du ratio de vraisemblance (Likelihood Ratio) :

Ce test révèle l'acceptation du modèle, en effet au niveau du premier tableau, nous pouvons voir ***Pr> khi-2 <0.001 pour le rapport de vraisemblance.***

Nous constatons dans ce modèle que toutes nos variables sont significatives et que l'AIC de ce modèle est inférieure à l'AIC du premier modèle.

Nous avons ***AIC_mod2 < AIC_mod1*** (3299.251 < 3303.296)



L'analyse de la Courbe ROC (Receiver Operating Characteristic) permet d'évaluer l'exactitude des prévisions de notre modèle en traçant la sensibilité par rapport au taux de faux positifs (1-spécificité) d'un test de classification

Calcul de la note de score

Analyse des valeurs estimées du maximum de vraisemblance						
Paramètre		DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > khi-2
Intercept		1	-11.1642	318.5	0.0012	0.9720
civilite	Fe	1	10.8576	318.5	0.0012	0.9728
civilite	Ho	1	11.7446	318.5	0.0014	0.9706
LFY_grou_enf_2	0	1	-0.4193	0.1188	12.4611	0.0004
LFY_celibat_2	0	1	-0.8403	0.1201	48.9570	<.0001
MT5Y_celibat_6	0	1	-0.6769	0.1472	21.1325	<.0001
MT5Y_sej_CAP_6	0	1	0.5387	0.1357	15.7489	<.0001
LFY_celib_f_2	0	1	-0.5517	0.2034	7.3589	0.0067

A partir des résultats de la proc logistic, le score de chaque individu va être calculé de la manière suivante :

```
/***** Calcul du SCore *****/
```

```
data PROJET.BASE_FINALE2;
```

```
set PROJET.BASE_FINALE2;
```

```
SCORE = -11.1642 + 10.8576*(civilite = "Fe") + 11.7446*(civilite = "Ho") -  
0.4193*LFY_grou_enf_2 -0.8403*LFY_celibat_2-0.6769*MT5Y_celibat_6 +  
0.5387*MT5Y_sej_CAP_6 -0.5517*LFY_celib_f_2;
```

```
Run;
```

```
proc means data=PROJET.BASE_FINALE2;
```

```
var SCORE;
```

```
ods output Summary=proc_means;
```

```
run;
```

```
proc univariate data= PROJET.BASE_FINALE2;
```

```
var SCORE;
```

```
run;
```

[file:///home/users/etudiant/T%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20\(12\).html](file:///home/users/etudiant/T%C3%A9chargements/SCRIPT_Projet_final-r%C3%A9sultats%20(12).html)

Nous allons créer une base de données appelée cible. Cette base sera la base où se trouveront les individus que nous allons cibler c'est à dire ceux ayant un score supérieur à -0.3066 qui est notre seuil.

Nous avons choisi ce seuil au hasard et celui-ci est le 3e quartile de la variable SCORE. Nous obtenons une base de 4816 individus.

Data Projet.cible;

```
set Projet.base_finale2 (where= (SCORE > - 0.3066)) ;
```

```
run;
```

Nous avons ensuite sélectionné les individus dans la base "cible" qui achètent des offres des produits de l'entreprise via internet.

Data Projet.cible;

```
set Projet.cible (where= (internet_1 =1));
```

```
run;
```

Nous obtenons une base de 3001 individus.

