

Code final présent tout en bas du document.

### Question 1

```
***** QUESTION 1 *****/
PROC IMPORT DATAFILE=REFFILE
  DBMS=CSV
  OUT=WORK.MARKETING;
  GETNAMES=YES;
  DELIMITER=';';
RUN;
```

### Question 2

```
***** QUESTION 2 *****/
data WORK.MARKETING;
  set WORK.MARKETING;
  rename annee_naissance=date_naiss ID=id_client date_ancienneté=date_anc;
run;

proc print data=WORK.MARKETING;
run;
```

Lignes totales : 2240 Colonnes totales : 27

	id_client	date_naiss	Niveau_educatif	Situation_maritale	Revenu	Nb_enfant	Nb_ado	date_anc
1	5524	1957	Graduation	Single	58138	0	0	04/09/2012
2	2174	1954	Graduation	Single	46344	1	1	08/03/2014
3	4141	1965	Graduation	Together	71613	0	0	21/08/2013
4	6182	1984	Graduation	Together	26646	1	0	10/02/2014

### Question 3

```
***** QUESTION 3 *****/
PROC SQL;
  select distinct * /* Pour chaque colonne de notre table, on affiche les différentes valeurs distinctes, afin de vérifier ainsi l'absence
  from WORK.MARKETING;
quit;
/* 2em Methode doublons et valeurs manquantes*/
PROC SORT DATA=WORK.MARKETING
  nodupkey;
  by _all_;
RUN;
proc print data=WORK.MARKETING;
run;

proc sql;
  describe table WORK.MARKETING; /* afficher la description de notre table*/
run;

/** 2) Vérification des valeurs manquantes de tous les variables étudiées **/
Proc means data = WORK.MARKETING n nmiss;
  var _numeric_;
run;
Proc freq data =WORK.MARKETING;
  tables _character_ / missing;
run;
```

La procédure MEANS

Variable	N	Nbre manquant
id_client	2240	0
date_naiss	2240	0
Revenu	2216	24
Nb_enfant	2240	0
Nb_ado	2240	0
date_anc	2240	0
Nb_jours_achat	2240	0
mt_vin	2240	0
mt_fruit	2240	0
mt_v viande	2240	0
mt_poisson	2240	0
mt_bonbon	2240	0
mt_or	2240	0
nb_achat_discount	2240	0
nb_achat_web	2240	0
nb_achat_catalogue	2240	0
nb_achat_magasin	2240	0
nb_web_visite	2240	0
offre_accepte_1	2240	0
offre_accepte_2	2240	0
offre_accepte_3	2240	0
offre_accepte_4	2240	0
offre_accepte_5	2240	0
Plainte	2240	0
Reponse	2240	0

## Question 4

```

/** QUESTION 4 **/

/* Visualisation de la table de données avec le type de chaque variables spécifié/
proc contents data=WORK.MARKETING;
run;
proc freq data=WORK.MARKETING;
run;

/ Après vérification suite au résultat de la question précédente /
/ On constate donc que tout le type des variables de notre /
/ table est bien le même que celui spécifié dans le tableau 1 /

/ Dans le cas où il y aurait une ou plusieurs variables qui dont /
/ le type ne serait pas celui spécifié dans le tableau 1 on peut /
/ donc exécuter la procédure si dessous /

proc sql;
select name into:liste separated by ';' from WORK.MARKETING;
quit;
*/

```

## Question 5

```

/***** QUESTION 5 *****/

data work.marketing;
set work.marketing;
drop MT_OR;
run;

```

Avant l'exécution du code :

Lignes totales : 2240 Colonne(s) totales : 27

	id_client	date_naiss	Niveau_educatif	Situation_maritale	Revenu	Nb_enfant	Nb_ado	date_anc
1	5524	1957	Graduation	Single	58138	0	0	04/09/2012
2	2174	1954	Graduation	Single	46344	1	1	08/03/2014
3	4141	1965	Graduation	Together	71613	0	0	21/08/2013
4	6182	1984	Graduation	Together	26646	1	0	10/02/2014

Après l'exécution du code :

Lignes totales : 2240 Colones totales : 26

	id_client	date_naiss	Niveau_educatif	Situation_maritale	Revenu	Nb_enfant
1	0	1985	Graduation	Married	70951	0
2	1	1961	Graduation	Single	57091	0

## Question 6

```

                                /***** QUESTION 6 *****/
data work.marketing;
  set work.marketing;
  format reponse_2 $40.;
  length reponse_2 $40 default=4. ;

  if reponse=1 then reponse_2='Le client a réalisé un achat';
  else reponse_2='Le client n'a pas réalisé un achat';
run;

proc print data=work.marketing;
run;
```

Lignes totales : 2240 Colones totales : 27

⏪ ⏩ Lignes 1-100 ➡ ⏴

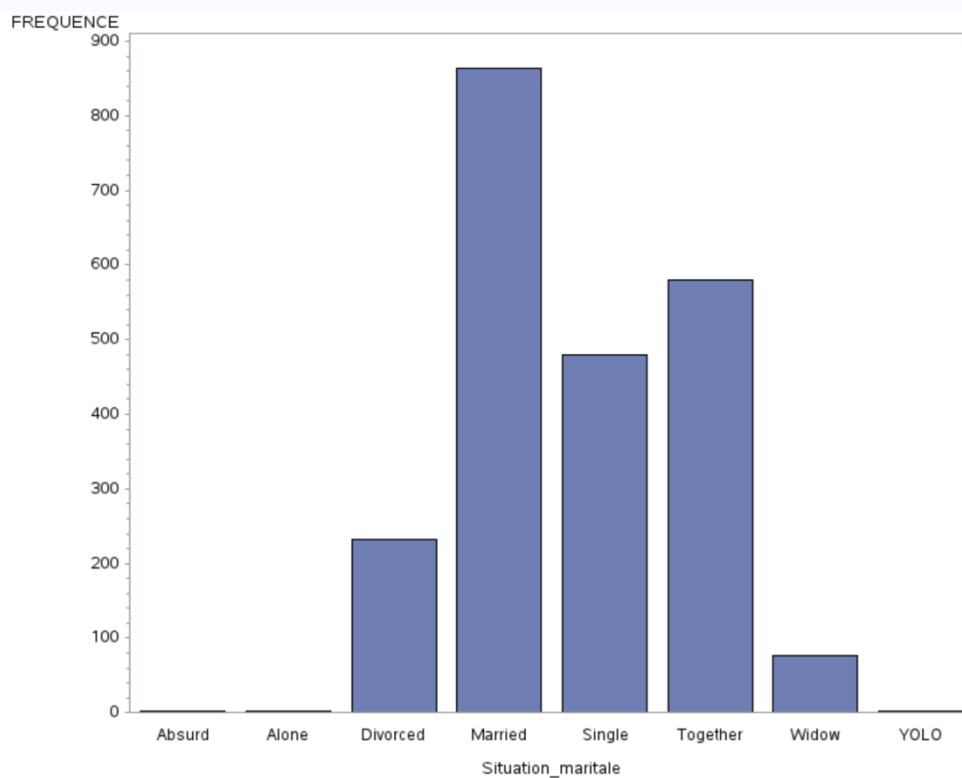
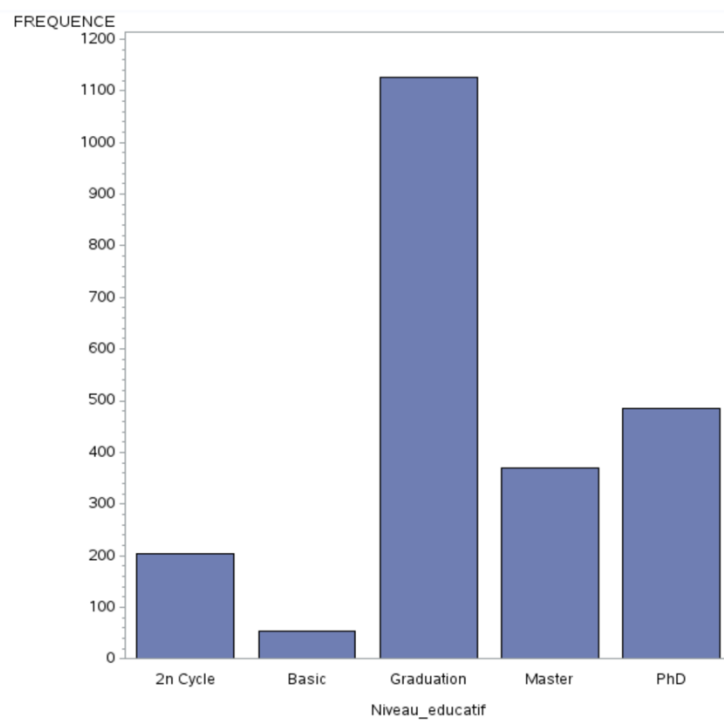
3	offre_accepte_4	offre_accepte_5	Plainte	Reponse	reponse_2
0	0	0	0	0	Le client n'a pas réalisé un achat
0	0	0	0	1	Le client a réalisé un achat
0	0	0	0	0	Le client n'a pas réalisé un achat
0	0	0	0	0	Le client n'a pas réalisé un achat
0	0	0	0	0	Le client n'a pas réalisé un achat
0	0	0	0	0	Le client n'a pas réalisé un achat

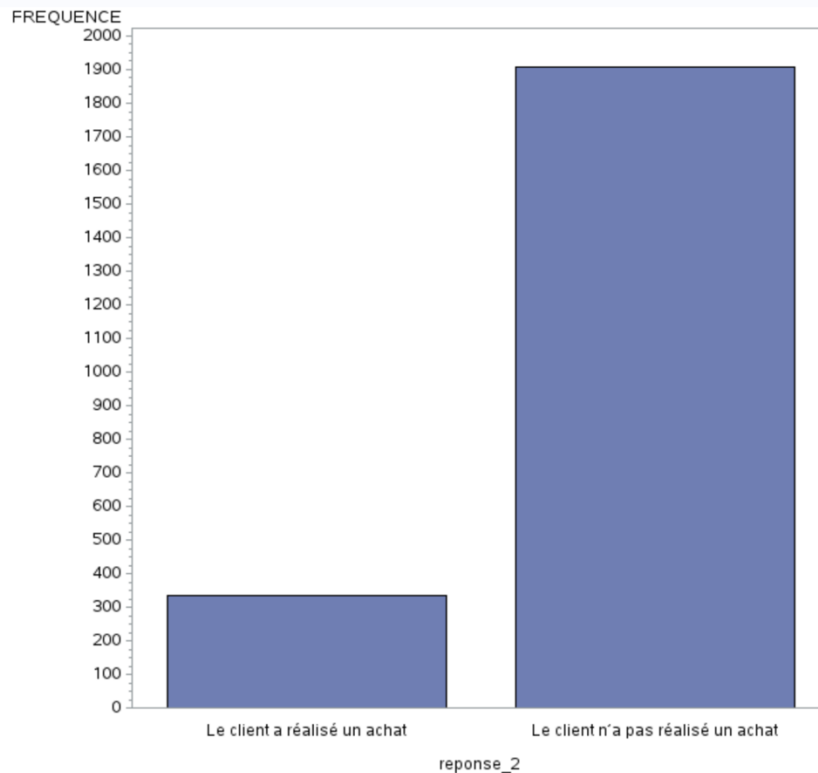
## Partie B : Statistiques descriptives

## Question 7

```

                                /***** QUESTION 7 *****/
PROC GCHART DATA = WORK.MARKETING;
  vbar Niveau_educatif Situation_maritale reponse_2 ;
run; quit;
```





## Question 8

```

/***** QUESTION 8 *****/
PROC FREQ DATA = work.marketing;
  table Situation_maritale*reponse_2 Niveau_educatif*reponse_2 / nopercnt norow nocol;
run;

```

La procédure FREQ

Fréquence	Table de Situation_maritale par reponse_2			
	reponse_2			Total
	Situation_maritale	Le client a réalisé un achat	Le client n'a pas réalisé un achat	
	Absurd	1	1	2
	Alone	1	2	3
	Divorced	48	184	232
	Married	98	766	864
	Single	106	374	480
	Together	60	520	580
	Widow	19	58	77
	YOLO	1	1	2
	Total	334	1906	2240

Fréquence	Table de Niveau_educatif par reponse_2			
	reponse_2			Total
	Niveau_educatif	Le client a réalisé un achat	Le client n'a pas réalisé un achat	
	2n Cycle	22	181	203
	Basic	2	52	54
	Graduation	152	975	1127
	Master	57	313	370
	PhD	101	385	486
	Total	334	1906	2240

**Constat** : Les différents constats observés sont les suivants :

- La majorité des clients sont variés et diplômés
- La majorité des personnes n'ayant pas réalisé d'achat sont mariées
- La plupart des clients ayant réalisé un achat sont célibataire et diplômés
- Le nombre total de clients n'ayant pas réalisé d'achats est plus élevé que le nombre de clients ayant réalisés au moins un achat

## Question 9

```
/****** QUESTION 9 *****/

proc means data=work.marketing;
  var nb_achat_web nb_achat_catalogue nb_achat_magasin revenu;
run;

proc sort data=work.marketing; by plainte; run;

proc means data=work.marketing;
  var mt_fruit mt_viande mt_poisson mt_bonbon mt_vin;
  by plainte;
run;

proc sort data=work.marketing; by reponse; run;

proc GCHART data=work.marketing;
  pie reponse;
run;

proc sql;
  select sum(offre_accepte_1) "offre_1", sum(offre_accepte_2) "offre_2", sum(offre_accepte_3) "offre_3", sum(offre_accepte_4) "offre_4",
  sum(offre_accepte_5) "offre_5" from work.marketing;
run;

proc sql;
  select nb_ado, sum(mt_vin) "vin", sum(mt_fruit) "fruit", sum(mt_viande) "viande", sum(mt_poisson) "poisson",
  sum(mt_bonbon) "bonbon" from work.marketing
  group by nb_ado;
quit;

proc corr data=work.marketing;
  var nb_achat_magasin nb_achat_web;
run;
```

La procédure MEANS

Variable	N	Moyenne	Ec-type	Minimum	Maximum
nb_achat_web	2240	4.0848214	2.7787141	0	27.0000000
nb_achat_catalogue	2240	2.6620536	2.9231007	0	28.0000000
nb_achat_magasin	2240	5.7901786	3.2509581	0	13.0000000
Revenu	2216	52247.25	25173.08	1730.00	666666.00

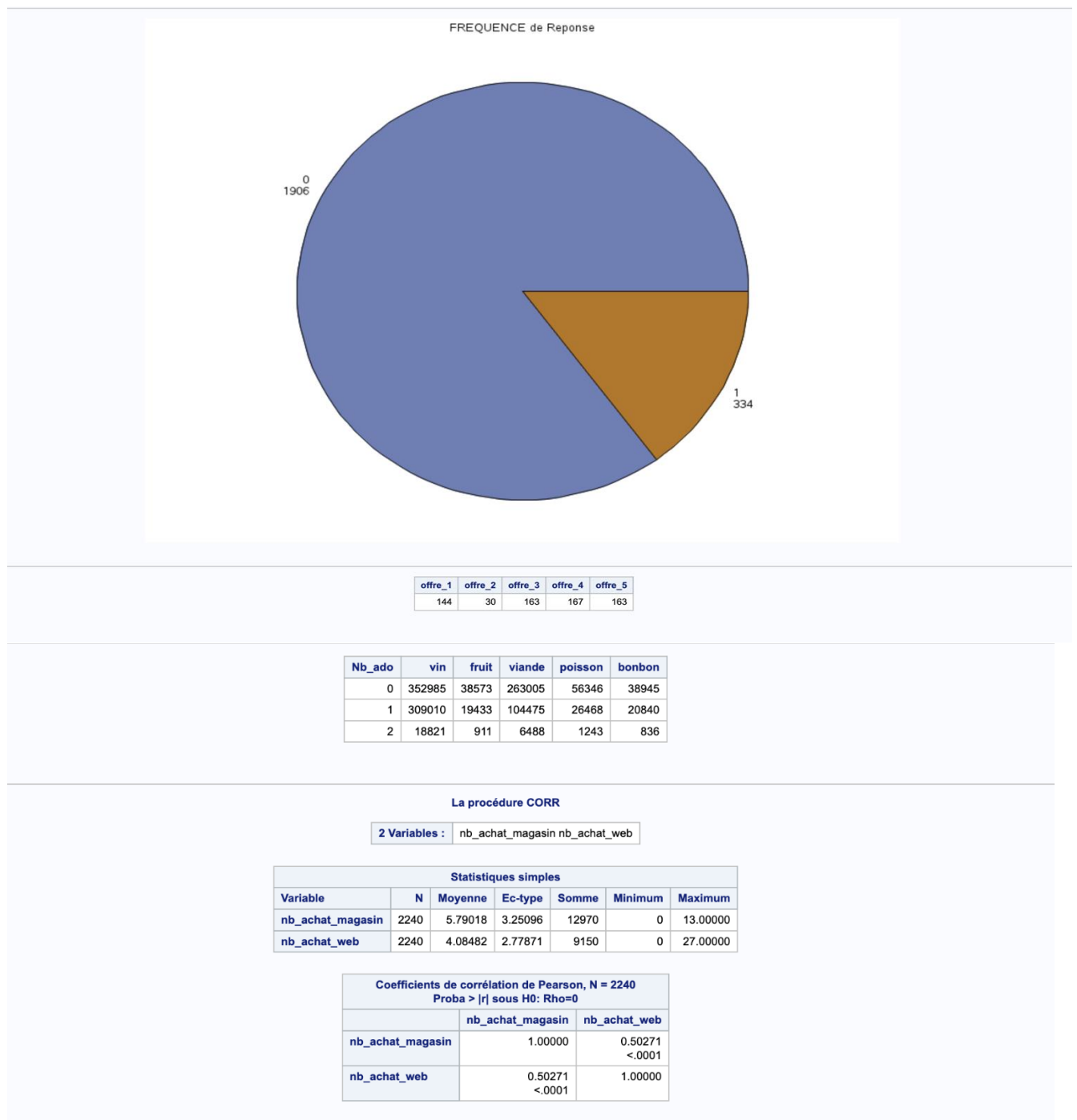
La procédure MEANS

PlaInte=0

Variable	N	Moyenne	Ec-type	Minimum	Maximum
mt_fruit	2219	26.3222172	39.7943123	0	199.0000000
mt_viande	2219	167.4655250	226.2096649	0	1725.00
mt_poisson	2219	37.6367733	54.7562071	0	259.0000000
mt_bonbon	2219	27.1532222	41.3621860	0	263.0000000
mt_vin	2219	305.2127084	337.3477051	0	1493.00

PlaInte=1

Variable	N	Moyenne	Ec-type	Minimum	Maximum
mt_fruit	21	24.1904762	38.3674589	0	137.0000000
mt_viande	21	112.4761905	159.9261139	1.0000000	590.0000000
mt_poisson	21	25.7619048	37.9932951	0	104.0000000
mt_bonbon	21	17.5238095	30.7272177	0	107.0000000
mt_vin	21	169.0000000	208.7702086	1.0000000	629.0000000



**Constat :** Les différents constats observés sont les suivants.

- Nous avons une vraie corrélation entre le nombre d'achats réalisés en ligne et en magasin avec un coeff de corrélation  $\approx 0.50271$  et une estimation de  $0.0001$
- En moyenne le revenu des clients est de 52247
- Il y'a en moyenne 4 commandes passées en ligne et 5 en magasin et 2 via le catalogue
- Le produit ayant reçu plus de plaintes est le vin et celui ayant reçu moins est le poisson.
- Le nombre total de personnes ayant pas accepté l'offre à la dernière campagne est de 1906 contre seulement 334 pour ceux qui ont accepté
- l'offre ayant reçu plus de réponses positives et l'offre 4 et la moins acceptée est l'offre 2 (seulement 30)
- La majorité des clients qui achètent du poisson, fruit, bonbon, viande ont aucun adolescent

## Question 10

```
proc format;
value vin_price low<10='moins de 10 €'
                10-50='entre 10 et 50 €'
                50-100='entre 50 et 100 €'
                100-150='entre 100 et 150 €'
                150-200='entre 150 et 200 €'
                200-250='entre 200 et 250 €'
                250-300='entre 250 et 300 €'
                300-350='entre 300 et 350 €'
                350-400='entre 350 et 400 €'
                500<-high='plus de 400 €';
value fruit_price low<10='moins de 10 €'
                 10-50='entre 10 et 50 €'
                 50-100='entre 50 et 100 €'
                 100-150='entre 100 et 150 €'
                 150-200='entre 150 et 200 €'
                 200-250='entre 200 et 250 €'
                 250-300='entre 250 et 300 €'
                 300-350='entre 300 et 350 €'
                 350-400='entre 350 et 400 €'
                 500<-high='plus de 400 €';
value viande_price low<10='moins de 10 €'
                  10-50='entre 10 et 50 €'
                  50-100='entre 50 et 100 €'
                  100-150='entre 100 et 150 €'
                  150-200='entre 150 et 200 €'
                  200-250='entre 200 et 250 €'
                  250-300='entre 250 et 300 €'
                  300-350='entre 300 et 350 €'
                  350-400='entre 350 et 400 €'
                  500<-high='plus de 400 €';
value poisson_price low<10='moins de 10 €'
                   10-50='entre 10 et 50 €'
                   50-100='entre 50 et 100 €'
                   100-150='entre 100 et 150 €'
                   150-200='entre 150 et 200 €'
                   200-250='entre 200 et 250 €'
                   250-300='entre 250 et 300 €'
                   300-350='entre 300 et 350 €'
                   350-400='entre 350 et 400 €'
                   500<-high='plus de 400 €';
value bonbon_price low<10='moins de 10 €'
                  10-50='entre 10 et 50 €'
                  50-100='entre 50 et 100 €'
                  100-150='entre 100 et 150 €'
                  150-200='entre 150 et 200 €'
                  200-250='entre 200 et 250 €'
                  250-300='entre 250 et 300 €'
                  300-350='entre 300 et 350 €'
                  350-400='entre 350 et 400 €'
                  500<-high='plus de 400 €';
run;

data work.marketing;
set work.marketing;
format prix_vin $20. prix_fruit $20. prix_viability $20. prix_poisson $20. prix_bonbon $20.;
prix_vin=put(mt_vin, vin_price.);
prix_fruit=put(mt_fruit, fruit_price.);
prix_viability=put(mt_viability, viande_price.);
prix_poisson=put(mt_poisson, poisson_price.);
prix_bonbon=put(mt_bonbon, bonbon_price.);
run;
```

## OUTPUT :

Lignes totales : 2240    Colonnes totales : 32

mt_fruit    prix_fruit		mt_viability    prix_viability		mt_poisson    prix_poisson		mt_bonbon    prix_bonbon	
88	entre 50 et 100 €	546	plus de 500 €	172	entre 150 et 200 €	88	entre 50 et 100 €
1	moins de 10 €	6	moins de 10 €	2	moins de 10 €	1	moins de 10 €
49	entre 10 et 50 €	127	entre 100 et 150 €	111	entre 100 et 150 €	21	entre 10 et 50 €
4	moins de 10 €	20	entre 10 et 50 €	10	entre 10 et 50 €	3	moins de 10 €
43	entre 10 et 50 €	118	entre 100 et 150 €	46	entre 10 et 50 €	27	entre 10 et 50 €
42	entre 10 et 50 €	98	entre 50 et 100 €	0	moins de 10 €	42	entre 10 et 50 €



## Question 11

```

/***** QUESTION 11 *****/
proc logistic data= work.marketing plots=roc;
  model Reponse = id_client date_naiss
    Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
    mt_fruit mt_viande mt_poisson mt_bonbon nb_achat_discount
    nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
    offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
    offre_accepte_5 Plainte;
run;

```

### La procédure LOGISTIC

Informations sur le modèle	
Table	WORK.MARKETING
Variable de réponse	Reponse
Nombre de niveaux de réponse	2
Modèle	logit binaire
Technique d'optimisation	Score de Fisher

Nb d'observations lues	2240
Nb d'observations utilisées	2216

Profil de réponse		
Valeur ordonnée	Reponse	Fréquence totale
1	0	1883
2	1	333

La probabilité modélisée est Reponse='0'.

Note: 24 observations supprimées en raison de valeurs manquantes pour les variables de réponse ou explicatives.

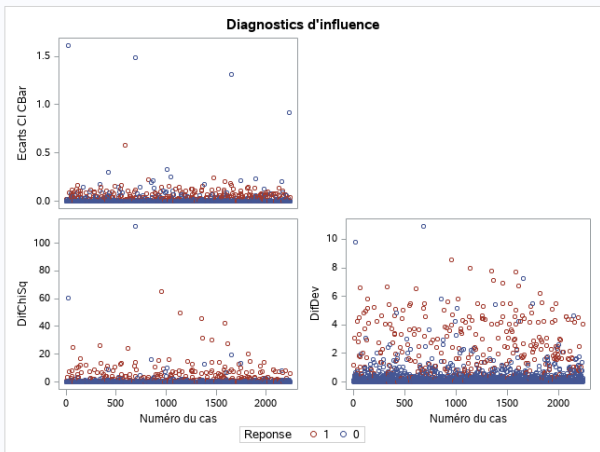
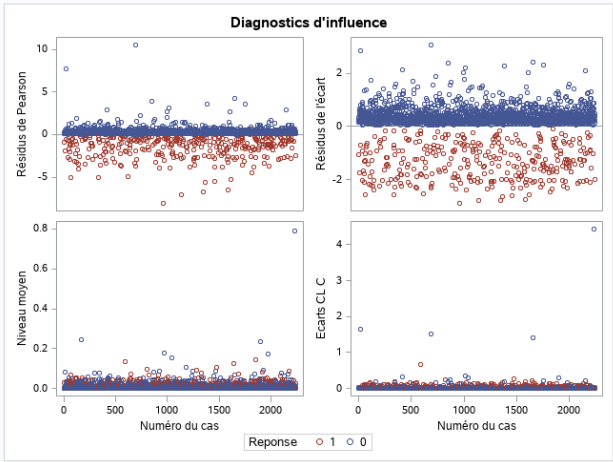
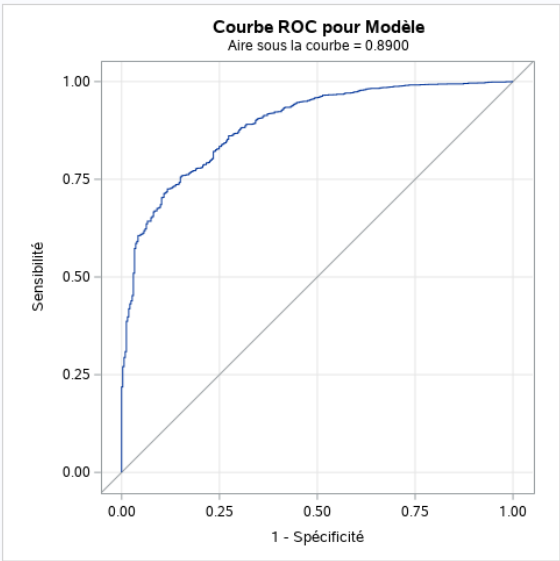
Etat de convergence du modèle	
Critère de convergence (GCONV=1E-8) respecté.	

Statistique d'ajustement du modèle		
Critère	Constante uniquement	Constante et Covariables
AIC	1877.527	1264.077
SC	1883.230	1400.960
-2 Log L	1875.527	1216.077

Test de l'hypothèse nulle globale : BETA=0			
Test	khi-2	DDL	Pr > khi-2
Rapport de vrais	959.4501	23	<.0001
Score	887.8737	23	<.0001
Wald	387.1097	23	<.0001

Analyse des valeurs estimées du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > khi-2
Intercept	1	-.95.3275	16.3095	34.8833	<.0001
id_client	1	0.000014	0.000023	0.3670	0.5446
date_naiss	1	0.0118	0.00963	3.1900	0.0741
Revenu	1	-2.01E-7	3.57E-6	0.0032	0.9551
Nb_enfant	1	-0.1761	0.2047	0.7405	0.3895
Nb_ado	1	0.8656	0.1903	20.6781	<.0001
date_anc	1	0.00383	0.000447	73.3652	<.0001
NB_jours_achat	1	0.0292	0.00287	103.0987	<.0001
mt_vin	1	-0.00008	0.000353	0.0543	0.8158
mt_fruit	1	-0.00226	0.00232	0.9450	0.3310
mt_viande	1	-0.00207	0.000496	17.3472	<.0001
mt_poisson	1	0.00223	0.00178	1.5637	0.2111
mt_bonbon	1	0.000277	0.00226	0.0150	0.9026
nb_achat_discount	1	-0.0977	0.0456	4.5254	0.0334
nb_achat_web	1	-0.1004	0.0316	10.0924	0.0015
nb_achat_catalogue	1	-0.0456	0.0423	1.1602	0.2814
nb_achat_magasin	1	0.1614	0.0345	21.9084	<.0001
nb_web_visite	1	-0.0810	0.0479	2.8576	0.0909
offre_accepte_1	1	-1.4489	0.2746	27.8342	<.0001
offre_accepte_2	1	-1.4690	0.5310	7.6349	0.0057
offre_accepte_3	1	-1.9792	0.2265	76.1492	<.0001
offre_accepte_4	1	-1.1488	0.2805	16.7770	<.0001
offre_accepte_5	1	-1.5188	0.2788	29.6816	<.0001

Association des probabilités prédites et des réponses observées			
Pourcentage concordant	89.0	D de Somers	0.780
Pourcentage discordant	11.0	Gamma	0.780
Pourcentage lié	0.0	Tau-a	0.199
Paires	627039	c	0.890



### Constats :

Les variables significatives de la régression logistique sont les suivantes :Nb\_ado, date\_anc, NB\_jours\_achat, mt\_vieande, nb\_achat\_magasin, offre\_accepte\_1, offre\_accepte\_3, offre\_accepte\_4, offre\_accepte\_5.

Au total on a 9 variables conservées par le modèle.

### Question 12

```
proc logistic data= work.marketing plots=ROC;
  model Reponse = id_client date_naiss
    Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
    mt_fruit mt_vieande mt_poisson mt_bonbon nb_achat_discount
    nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
    offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
    offre_accepte_5 Plainte / SELECTION=BACKWARD;
run;
```

### Constat :

Les variables que le modèle conserve sont tout simplement les variables significatives énoncées lors de la première régression logistique sans spécification des méthodes backward, forward ou stepwise ajouté à cela de nouvelles variables suivantes : nb\_achat\_discount, nb\_achat\_web et enfin offre\_accepte\_2.

Au total on a 12 variables conservées par le modèle.

### Question 13

```

/***** QUESTION 13 *****/
proc logistic data= work.marketing plots=ROC;
  model Reponse = id_client date_naiss
    Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
    mt_fruit mt_vieande mt_poisson mt_bonbon nb_achat_discount
    nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
    offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
    offre_accepte_5 Plainte / SELECTION=STEPWISE;
run;
```

### Constat :

Les variables sont conservées sont les même que celui du modèle avec la SELECTION=BACKWARD et la sélection SELECTION=FORWARD.

## Question 14

```

/***** QUESTION 14 *****/
%macro Proc_log(BACKWARD, FORWARD, STEPWISE);
  proc logistic data= work.marketing plots=ROC;
    model Reponse = id_client date_naiss
      Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
      mt_fruit mt_viande mt_poisson mt_bonbon nb_achat_discount
      nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
      offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
      offre_accepte_5 Plainte / SELECTION=FORWARD SELECTION=BACKWARD SELECTION=STEPWISE;

    run;
    quit;

%mend;
%Proc_log(BACKWARD, FORWARD, STEPWISE);

```

## Question 15

```

/***** QUESTION 15 *****/
proc logistic data= work.marketing plots=ROC;
  OUTPUT OUT=SORTIE PREDPROBS = I;
  model Reponse = id_client date_naiss
    Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
    mt_fruit mt_viande mt_poisson mt_bonbon nb_achat_discount
    nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
    offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
    offre_accepte_5 Plainte / SELECTION=STEPWISE;

run;
Proc Print data=work.marketing;
RUN;

```

## Question 16

```

/***** QUESTION 16 *****/
proc freq data=work.sortie;
  table _FROM_*_INTO_;
run;

```

La procédure FREQ

Fréquence Pourcentage Pct de ligne Pct de col.	Table de _FROM_ par _INTO_		
	_FROM_(Valeur formatée de la réponse observée)	_INTO_(Valeur formatée de la réponse prédite)	
		0	1
			Total
	0	1851 82.63 97.11 90.47	55 2.46 2.89 28.35
	1	195 8.71 58.38 9.53	139 6.21 41.62 71.65
	Total	2046 91.34	194 8.66
			2240 100.00

## Question 17

```

/***** QUESTION 17 *****/
Proc Export data = Work.marketing
  dbms=xls
  Outfile='/Projet_SAS/Données_devoir_SAS.csv';
  Delimiter=";";
Run;
```

# CODE

```

/*****
/***** Projet SAS *****/
*****/

/*importation du fichier Données_devoir_SAS sous sas*/
filename REFFILE '/home/u60864178/projet';
filename REFFILE '/home/u60864178/projet/Données_devoir_SAS.csv';

/***** PARTIE A *****/

/***** QUESTION 1 *****/
PROC IMPORT DATAFILE=REFFILE
    DBMS=CSV /*extension du fichier importé, ici CSV */
    OUT=WORK.MARKETING; /* nom de la table*/
    GETNAMES=YES; /*garder le nom des colonnes */
    DELIMITER=';'; /*spécifier le délimiteur qui sépare les colonnes de données dans le
fichier d'entrée*/
RUN;

/* procédure permettant de voir le contenu d'une table */
PROC CONTENTS DATA=WORK.MARKETING
    varnum /* afficher les variables par ordre d'apparition de la table*/
    short; /* afficher seulement le nom des variables */
RUN;

/* Afficher le contenu d'une table */
proc print data=WORK.MARKETING;
run;

/***** QUESTION 2 *****/
data WORK.MARKETING;
    set WORK.MARKETING; /* reprendre notre table marketing*/
    rename annee_naissance=date_naiss ID=id_client date_ancienneté=date_anc; /*
Spécifier les colonnes qu'on veut renommer*/
run;

proc print data=WORK.MARKETING;
run;

/***** QUESTION 3 *****/
PROC SQL;
    select distinct * /* Pour chaque colonne de notre table, on affiche les différentes
valeurs distinctes, afin de vérifier ainsi l'absence de doublons*/
```

```

        from WORK.MARKETING;
quit;
        /* 2em Methode doublons et valeurs manquantes*/
PROC SORT DATA=WORK.MARKETING
        nodupkey;
        by _all_;
RUN;
proc print data=WORK.MARKETING;
run;

proc sql;
        describe table WORK.MARKETING; /* afficher la description de notre table*/
run;

/** 2) Vérification des valeurs manquantes de tous les variables étudiées **/
Proc means data = WORK.MARKETING n nmiss;
        var _numeric_;
run;
Proc freq data =WORK.MARKETING;
        tables _character_ / missing;
run;
/***** QUESTION 4 *****/

/* Après verification suite au résultat de la question précédente */
/* On constate donc que toutes les types des variables de notre */
/* table est bien le même que celui spécifié dans le tableau 1 */

/* Dans le cas où il y aurait une ou plusieurs variables qui dont */
/* le type ne serait pas celui spécifié dans le tableau 1 on peut */
/* donc exécuter la procédure ci-dessous */

/*proc sql;
        select name into:liste separated by ';' from WORK.MARKETING;
quit;*/

/***** QUESTION 5 *****/
data work.marketing;
        set work.marketing;
        drop MT_OR; /* Suppression de la variable mt_or avec la fonction drop*/
run;

/***** QUESTION 6 *****/
data work.marketing;
        set work.marketing;
        format reponse_2 $40.; /* on déclare notre nouvelle variable en initialisant sa taille et son
type (ici chaîne de caractère de taille 40 max)*/

```

```

length reponse_2 $40 default=4. ;

if reponse=1 then reponse_2='Le client a réalisé un achat';
else reponse_2='Le client n'a pas réalisé un achat';
run;

proc print data=work.marketing;
run;

```

```

/***** PARTIE B *****/

```

```

/***** QUESTION 7 *****/

```

```

PROC GCHART DATA = WORK.MARKETING;
  vbar Niveau_educatif Situation_maritale reponse_2 ;
run; quit;

```

```

/***** QUESTION 8 *****/

```

```

PROC FREQ DATA = work.marketing;
  table Situation_maritale*reponse_2 Niveau_educatif*reponse_2 / nopercnt norow
  nocol;
run;

```

```

/***** CONSTAT *****/

```

```

/* La majorité des clients sont mariés et diplômés */
/* Nous constatons que la majorité des personnes n'ayant pas réalisé d'achat sont mariées
/* La plupart des clients ayant réalisé un achat sont célibataire et diplômés
* Le nombre total de clients n'ayant pas réalisé d'achats est plus élevé que le nombre de
clients ayant réalisés un achat */

```

```

/***** QUESTION 9 *****/

```

```

proc means data=work.marketing;
  var nb_achat_web nb_achat_catalogue nb_achat_magasin revenu;
run;

```

```

proc sort data=work.marketing; by plainte; run;

```

```

proc means data=work.marketing;
  var mt_fruit mt_vieande mt_poisson mt_bonbon mt_vin;
  by plainte;
run;

```



```
proc sort data=work.marketing; by reponse; run;
```

```
proc GCHART data=work.marketing;  
    pie reponse;  
run;
```

```
proc sql;  
    select sum(offre_accepte_1) "offre_1", sum(offre_accepte_2) "offre_2",  
    sum(offre_accepte_3) "offre_3", sum(offre_accepte_4) "offre_4", sum(offre_accepte_5)  
    "offre_5" from work.marketing;  
run;
```

```
proc sql;  
    select nb_ado, sum(mt_vin) "vin", sum(mt_fruit) "fruit", sum(mt_viande) "viande",  
    sum(mt_poisson) "poisson", sum(mt_bonbon) "bonbon" from work.marketing  
    group by nb_ado;  
quit;
```

```
proc corr data=work.marketing;  
    var nb_achat_magasin nb_achat_web;  
run;
```

/\*\*\*\*\* CONSTAT \*\*\*\*\*/

/\* Nous avons une vraie corrélation entre le nombre d'achats réalisés en ligne et en magasin  
\*/

/\* avec un coeff de corrélation == 0.50271 et une estimation de 0001

En moyenne le revenu des clients est de 52247

Il y'a en moyenne 4 commandes passées en ligne et 5 en magasin et 2 via le catalogue

Le produit ayant reçu plus de plaintes est le vin et celui ayant reçu moins est le poisson.

Le nombre total de personnes ayant pas accepté l'offre à la dernière campagne est de 1906 contre seulement 334 pour ceux qui ont accepté

et l'offre ayant reçu plus de réponses positives est l'offre 4 et la moins acceptée est l'offre 2 (seulement 30)

La majorité des clients qui achètent du poisson, fruit, bonbon, viande ont aucun adolescent

```
*/
```

```
proc corr data=work.marketing;  
    var nb_achat_web nb_web_visite nb_achat_catalogue;  
run;
```

```
/***** QUESTION 10 *****/
```

```
proc format;  
value vin_price low-<10='moins de 10 €'  
                10-50='entre 10 et 50 €'  
                50-100='entre 50 et 100 €'  
                100-150='entre 100 et 150 €'  
                150-200='entre 150 et 200 €'  
                200-250='entre 200 et 250 €'  
                250-300='entre 250 et 300 €'  
                300-350='entre 300 et 350 €'  
                350-500='entre 350 et 400 €'  
                500<-hight='plus de 500 €';  
value fruit_price low-<10='moins de 10 €'  
                  10-50='entre 10 et 50 €'  
                  50-100='entre 50 et 100 €'  
                  100-150='entre 100 et 150 €'  
                  150-200='entre 150 et 200 €'  
                  200-250='entre 200 et 250 €'  
                  250-300='entre 250 et 300 €'  
                  300-350='entre 300 et 350 €'  
                  350-400='entre 350 et 400 €'  
                  500<-hight='plus de 500 €';  
value viande_price low-<10='moins de 10 €'  
                   10-50='entre 10 et 50 €'  
                   50-100='entre 50 et 100 €'  
                   100-150='entre 100 et 150 €'  
                   150-200='entre 150 et 200 €'  
                   200-250='entre 200 et 250 €'  
                   250-300='entre 250 et 300 €'  
                   300-350='entre 300 et 350 €'  
                   350-400='entre 350 et 400 €'  
                   500<-hight='plus de 500 €';  
value poisson_price low-<10='moins de 10 €'  
                    10-50='entre 10 et 50 €'  
                    50-100='entre 50 et 100 €'  
                    100-150='entre 100 et 150 €'  
                    150-200='entre 150 et 200 €'  
                    200-250='entre 200 et 250 €'  
                    250-300='entre 250 et 300 €'
```

```

300-350='entre 300 et 350 €'
350-400='entre 350 et 400 €'
500<-hight='plus de 500 €';
value bonbon_price low-<10='moins de 10 €'
10-50='entre 10 et 50 €'
50-100='entre 50 et 100 €'
100-150='entre 100 et 150 €'
150-200='entre 150 et 200 €'
200-250='entre 200 et 250 €'
250-300='entre 250 et 300 €'
300-350='entre 300 et 350 €'
350-400='entre 350 et 400 €'
500<-hight='plus de 500 €';

run;

```

```

data work.marketing;
    set work.marketing;
    format prix_vin $20. prix_fruit $20. prix_viande $20. prix_poisson $20. prix_bonbon
$20.;
    prix_vin=put(mt_vin, vin_price.);
    prix_fruit=put(mt_fruit, fruit_price.);
    prix_viande=put(mt_viande, viande_price.);
    prix_poisson=put(mt_poisson, poisson_price.);
    prix_bonbon=put(mt_bonbon, bonbon_price.);

run;

```

/\*\*\*\*\*\* Partie C \*\*\*\*\*/

/\*\*\*\*\*\* QUESTION 11 \*\*\*\*\*/

```

proc logistic data= work.marketing plots=roc;
    model Reponse = id_client date_naiss
    Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
    mt_fruit mt_viande mt_poisson mt_bonbon nb_achat_discount
    nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
    offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
    offre_accepte_5 Plainte;

run;

```

```

/* Les variables significatives de la regression logistique */
/* sont les suivantes :Nb_ado, date_anc, NB_jours_achat, mt_viande, nb_achat_magasin,
/* offre_accepte_1, offre_accepte_3, offre_accepte_4, offre_accepte_5 */
/* Au total on a conjecturé 9 variables significatives */

```

/\*\*\*\*\*\* QUESTION 12 \*\*\*\*\*/

```

proc logistic data= work.marketing plots=ROC;
    model Reponse = id_client date_naiss
        Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
        mt_fruit mt_viante mt_poisson mt_bonbon nb_achat_discount
        nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
        offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
        offre_accepte_5 Plainte / SELECTION=BACKWARD;
run;
/* Les variables que le modèle conserve sont tout simplement les variables significatives */
/* éliminées lors de la première régression logistique sans spécification des méthodes
backward, forward ou stepwise */
/* ajouté à cela de nouvelles variables suivantes : nb_achat_discount, nb_achat_web et
enfin offre_accepte_2 */
/* Au total on a 12 variables conservées par le modèle */

```

/\*\*\*\*\* QUESTION 13 \*\*\*\*\*/

```

proc logistic data= work.marketing plots=ROC;
    model Reponse = id_client date_naiss
        Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
        mt_fruit mt_viante mt_poisson mt_bonbon nb_achat_discount
        nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
        offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
        offre_accepte_5 Plainte / SELECTION=STEPWISE;
run;

/* Les variables conservées sont les mêmes que celui du modèle */
/* avec la SELECTION=BACKWARD et la sélection SELECTION=FORWARD */

```

/\*\*\*\*\* QUESTION 14 \*\*\*\*\*/

```

%macro Proc_log(BACKWARD, FORWARD, STEPWISE);
    proc logistic data= work.marketing plots=ROC;
        model Reponse = id_client date_naiss
            Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
            mt_fruit mt_viante mt_poisson mt_bonbon nb_achat_discount
            nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
            offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
            offre_accepte_5 Plainte / SELECTION=FORWARD SELECTION=BACKWARD
SELECTION=STEPWISE;
        run;
        quit;
    %mend;
%Proc_log(BACKWARD, FORWARD, STEPWISE);

```

/\*\*\*\*\* QUESTION 15 \*\*\*\*\*/

```

proc logistic data= work.marketing plots=ROC;

```

```

OUTPUT OUT=SORTIE PREDPROBS = I;
model Reponse = id_client date_naiss
  Revenu Nb_enfant Nb_ado date_anc NB_jours_achat mt_vin
  mt_fruit mt_viante mt_poisson mt_bonbon nb_achat_discount
  nb_achat_web nb_achat_catalogue nb_achat_magasin nb_web_visite
  offre_accepte_1 offre_accepte_2 offre_accepte_3 offre_accepte_4
  offre_accepte_5 Plainte / SELECTION=STEPWISE;
run;
Proc Print data=work.marketing;
RUN;

          /***** QUESTION 16 *****/
proc freq data=work.sortie;
  table _FROM_*_INTO_;
run;

          /***** QUESTION 17 *****/
Proc Export data = Work.marketing
  dbms=xls
  Outfile='/Projet_SAS/Données_devoir_SAS.csv';
  Delimiter=",";
Run;

```