

```

*****
Creation de Librairie
*****
Libname projet "/home/u60064275/Projet_sas";

/*Etape 1 : Importation*/

/*1- Ecrivez un script qui lit les trois fichiers de train_url1, train_url2 et train_url3.*/

/* Importation des donnees de url1 */
Proc IMPORT datafile="/home/u60064275/Projet_sas/train_url1.xlsx"
  DBMS=XLSX
  OUT= projet.Train_url_1;
  Getnames=Yes;
Run;
/* Importation des donnees de url2 */
Proc IMPORT datafile="/home/u60064275/Projet_sas/train_url2.xlsx"
  DBMS=XLSX
  OUT= projet.Train_url_2;
  Getnames=Yes;
Run;
/* Importation des donnees de url3 */
Proc IMPORT datafile="/home/u60064275/Projet_sas/train_url3.xlsx"
  DBMS=XLSX
  OUT= projet.Train_url_3;
  Getnames=Yes;
Run;

/*Concaténation :*/

/*1- Écrivez un script qui vous permet d'effectuer une concaténation entre train_url1 et
train_url2 pour créer un fichier nommé train_url1_2*/

/* On va d'abord trier les bases */
proc sort data=projet.Train_url_1;
by passengerId;
run;
proc sort data=projet.Train_url_2;
by passengerId;
run;
/* on fait la concatenation */
data projet.train_url1_2;
merge projet.Train_url_1 projet.Train_url_2;
by passengerId;
run;
/* Imprimer le resultat */
proc print data=projet.train_url1_2;

/*2- Écrivez un script qui vous permet d'effectuer une nouvelle concaténation entre
train_url1_2 et train_url3 pour créer un fichier nommé train_url*/

data projet.train_url;
set projet.train_url1_2 projet.Train_url_3;
run;

proc print data=projet.train_url;

/*Etape 2 : Analyse descriptive du fichier Train_url*/

/*1- Déterminer le nombre :
a- de survivants et le nombre de non survivant ?*/

Proc Freq data=projet.train_url;
Tables survived ;
run;

/*b- de passagers hommes et femmes ?*/

```

```
Proc Freq data=projet.train_url;
Tables sex;
run;
```

```
PROC GCHART DATA = projet.train_url;
PIE sex; /*Pour un diagramme circulaire*/
title "Distribution de la variable sexe";
GOPTIONS colors=(red blue);
RUN;
QUIT;
```

```
/*2- Combien y a-t-il de passagers :
a- par point d'embarquement ?*/
```

```
Proc sql;
select Embarked,count(DISTINCT PassengerId) as Nombre_Passager
from projet.train_url
group by Embarked
ORDER BY Nombre_Passager ASC
;
Quit;
```

```
/*b- hommes et femmes par classe ?*/
```

```
Proc sql;
select Sex,Pclass,count(DISTINCT PassengerId) as Nombre_Total
from projet.train_url
group by Sex,Pclass
ORDER BY Nombre_Total ASC
;
Quit;
```

```
/*Etape 3 : Analyse descriptive avancée*/
```

```
/*1- Pour la variable âge : :
a- Calculer la moyenne de la variable âge par sexe*/
```

```
data projet.train_url;
set projet.train_url;
age_num = input(age ,8.);
run;
```

```
PROC SQL;
Select sex, AVG(age_num ) as Moyenne_Age
from projet.train_url
group by sex;
QUIT;
```

```
/*b- Faire un histogramme de la variable Age par sexe. Interpréter le graphique*/
```

```
PROC GCHART DATA=projet.train_url;
VBAR age_num/ subgroup= sex;
title "histogramme d'hommes et de femmes par sexe";
GOPTIONS colors=(red blue);
RUN;
QUIT;
```

```
title "histogramme d'hommes et de femmes par sexe";
proc gchart data=projet.train_url;
block sex / sumvar=age_num;
GOPTIONS colors=(green);
run;
quit;
```

```
proc gchart data=projet.train_url;
vbar age_num/ subgroup=sex;
title "histogramme d'hommes et de femmes par sexe";
GOPTIONS colors=(red blue);
```

```
run;
```

```
/*c- Faire une boîte à moustaches de la variable Age par sexe*/
```

```
PROC SGPLOT DATA = projet.train_url;
  VBOX age_num / category = sex;
RUN;
```

```
/*2- Pour la variable âge : :
    a- Calculer la moyenne de la variables Fare par classe*/
```

```
data projet.train_url;
  set projet.train_url;
  Fare_num = input(Fare ,8.);
run;
```

```
PROC SQL;
  Select Pclass, AVG(Fare_num) as Moyenne_Fare
  from projet.train_url
  group by Pclass;
QUIT;
```

```
/*b- Faire un histogramme de la variable Fare par classe. Interpréter le graphique*/
```

```
proc gchart data=projet.train_url;
vbar Fare_num/ subgroup=Pclass;
title "histogramme de la variable Fare par classe";
GOPTIONS colors=(blue lightblue lightcyan);
run;
```

```
/* c- Faire une boîte à moustaches de la variable Fare par classe. Interpréter le graphique*/
```

```
PROC SGPLOT DATA = projet.train_url;
  VBOX Fare_num / category = Pclass;
RUN;
```

```
proc sort data= projet.train_url;

  By Pclass;
run;
```

```
PROC BOXPLOT data= projet.train_url;
  plot Fare_num * Pclass /
  caxes=black ctext= black
  Cboxes=black boxstyle= schematic
  idcolor=black idsymbol= dot;
```

```
run;
PROC BOXPLOT data= projet.train_url;
  plot Fare_num * Pclass /
  caxes=black ctext= black
  Cboxes=black boxstyle= schematic
  idcolor=black idsymbol= dot;
  INSET MIN MEAN MAX STDDEV/
  HEADER = 'Overall Statistics' POSITION= TM;
  insetgroup N min max NHIGH NLOQ NOUT/
  HEADER='Extreme par Classe';
run;
```

```
/*3- */
/*a- Le nombre d'hommes et de femmes par classe*/
```

```
PROC Freq data =projet.train_url;
  Tables Pclass *sex;
run;
```

```
/* b. La moyenne de la variable Age par classe et par sexe. */
```

```
Proc Sql;
  Select Pclass, avg(age_num) as Moyenne_Age
  from projet.train_url
```

```
group by Pclass;
quit;

Proc Sql;
  Select Sex, avg(age_num) as Moyenne_Age
  from projet.train_url
  group by Sex;
quit;

/*4- */
/*a- Les pourcentages de survivants et non survivants par sexe du fichier train_url*/

Proc freq Data = projet.train_url ;
  Tables Survived * Sex;
Run;

/*b- Les pourcentages de survivants par point d'embarquement*/

Proc freq data =projet.train_url;
  Tables Survived * Embarked ;
run;

/*c- Calculer les pourcentages de survivants par sexe et par classe*/

Proc freq data =projet.train_url;
  Tables Survived * Sex * Pclass;
run;

/*5. Créer une variable pour différencier un adulte d'un enfant:*/

data projet.train_url_bis;
set projet.train_url;
  if age<18 then age_Categorie="Enfant";
  else age_Categorie="Adulte";
run;

/*a- Le nombre de passagers adultes et le nombre d'enfants passagers*/

Proc sql;
  Select age_Categorie, count(age_Categorie) as Nbre_passager
  from projet.train_url_bis
  group by age_Categorie;
quit;

/*b- Le nombre d'enfants et d'adultes qui ont survécu par classe ?*/

Proc sql;
  Select age_Categorie, count(Survived) as Nbre_passager_1
  from projet.train_url_bis
  where Survived=1
  group by age_Categorie;
quit;

Proc sql;
  Select age_Categorie, count(Survived) as Nbre_passager_0
  from projet.train_url_bis
  where Survived=0
  group by age_Categorie;
quit;

Proc freq data =projet.train_url_bis;
  Tables age_Categorie * Survived ;
run;

/*c- Le nombre d'enfants et d'adultes qui ont survécu par classe et par point d'embarcation ?*/

Proc freq data =projet.train_url_bis;
  Tables age_Categorie * Survived* Embarked ;
run;
```