

SAE 1.04 Création d'une base de données

Phase 1:

titanic_train												
PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S	
2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C	
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925		S	
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S	
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S	
6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q	
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S	
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S	
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S	
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708		C	
11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	16.7	G6	S	
12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103	S	
13	0	3	Saunderscock, Mr. William Henry	male	20	0	0	A/5. 2151	8.05		S	
14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082	31.275		S	
15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406	7.8542		S	
16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706	16		S	
17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29.125		Q	
18	1	2	Williams, Mr. Charles Eugene	male		0	0	244373	13		S	
19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)	female	31	1	0	345763	18		S	
20	1	3	Masselmann, Mrs. Fatima	female		0	0	2649	7.225		C	
21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865	26		S	
22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698	13	D56	S	
23	1	3	McGowan, Miss. Anna "Annie"	female	15	0	0	330923	8.0292		Q	
24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788	35.5	A6	S	
25	0	3	Palsson, Miss. Torborg Danira	female	8	3	1	349909	21.075		S	
26	1	3	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson)	female	38	1	5	347077	31.3875		S	
27	0	3	Emir, Mr. Farred Chehab	male		0	0	2631	7.225		C	
28	0	1	Fortune, Mr. Charles Alexander	male	19	3	2	19950	263	C23 C25 C27	S	
29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female		0	0	330959	7.8792		Q	
30	0	3	Todoroff, Mr. Lallo	male		0	0	349216	7.8958		S	
31	0	1	Uruchurtu, Don. Manuel E	male	40	0	0	PC 17601	27.7208		C	
32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	female		1	0	PC 17569	146.5208	B78	C	
33	1	3	Glynn, Miss. Mary Agatha	female		0	0	335677	7.75		Q	
34	0	2	Wheadon, Mr. Edward H	male	66	0	0	C.A. 24579	10.5		S	
35	0	1	Meyer, Mr. Edgar Joseph	male	28	1	0	PC 17604	82.1708		C	
36	0	1	Holverson, Mr. Alexander Oskar	male	42	1	0	113789	52		S	
37	1	3	Mamee, Mr. Hanna	male		0	0	2677	7.2292		C	
38	0	3	Cann, Mr. Ernest Charles	male	21	0	0	A/5. 2152	8.05		S	
39	0	3	Vander Planke, Miss. Augusta Maria	female	18	2	0	345764	18		S	
40	1	3	Nicola-Yarred, Miss. Jamila	female	14	1	0	2651	11.2417		C	
41	0	3	Ahlin, Mrs. Johan (Johanna Persdotter Larsson)	female	40	1	0	7546	9.475		S	
42	0	2	Turpin, Mrs. William John Robert (Dorothy Ann Wonnacott)	female	27	1	0	11668	21		S	
43	0	3	Kraeff, Mr. Theodor	male		0	0	349253	7.8958		C	
44	1	2	Laroche, Miss. Simonne Marie Anne Andree	female	3	1	2	SC/Paris 2123	41.5792		C	
45	1	3	Devaney, Miss. Margaret Delia	female	19	0	0	330958	7.8792		Q	
46	0	3	Rogers, Mr. William John	male		0	0	S.C./A.4. 23567	8.05		S	
47	0	3	Lennon, Mr. Denis	male		1	0	370371	15.5		Q	

Cardinalité :891 tuples

Degrés :12

Attribut PassengerId= Numéro de passager va de 1 à 891

Attribut Survived {0,1}(0 = mort / 1 = vivant)

Il y a eu 549 morts et 342 survivant

Attribut PClasse {1,2,3}

il y a eu 216 tickets de 1ere classe vendue, 184 pour la 2eme classe et 491 tickets pour la 3eme classe

Attribut Name = Nom et Prénom

Attribut sex {male, female}

65% étaient des hommes contre 35% des femmes

Attribut Age= age des passager (entier positif or null)

Attribut SibSp= nb de frère et soeur/ conjoint à bord (entier positif)

Attribut Parch = nb de parent/ enfant à bord (entier)

Attribut Ticket = Numéro de ticket= différent les une et les autres

Attribut Fare = tarif des passager (entier positif)

Attribut Cabin = Numéro de Cabine, beaucoup de case sont vides, donc il faut considérer qu'il sont null

Attribut Embarked {Q, S, C}(Q = Queenstone / S = Southampton / C = Cherbourg) = Le port où l'embarquement à été fait

72% des passagers ont embarqué à Southampton, 19% ont embarqué à Cherbourg et 9% ont embarqué à Queenstone

Phase 2:

script pour créer les tables:

%%sql

```
create table titanic(  
  PassengerId integer,  
  Survived integer,  
  Pclass integer,  
  Name varchar(100),  
  Sex varchar(10),  
  Age float,  
  SibSp integer,  
  Parch integer,  
  numTicket varchar(20),  
  Fare float,  
  Cabin varchar(20),  
  Embarked varchar(1));
```

```
create table Passagers(  
  PassengerId integer primary key,  
  Name varchar(100),  
  Sex varchar(10),  
  Age float,  
  SibSp integer,  
  Parch integer);
```

```
create table Tickets(  
  Pclass integer,  
  Name varchar(100) primary key,  
  numTicket varchar(20),  
  Fare float,  
  Age float,  
  Cabin varchar(20),  
  Embarked varchar(1));
```

```
create table Survivants(  
  PassengerId integer,  
  Survived integer,  
  Name varchar (100),  
  primary key (PassengerId, Name),  
  foreign key (PassengerId) references Passagers(PassengerId),  
  foreign key (Name) references Tickets(Name));
```

Rapport:

Nous avons commencé à créer la table titanic qui regroupe tous les attributs qui sont présents dans le fichier titanic_train.csv. Cette table va nous permettre d'alimenter les autres tables qui sont : passenger, ticket, famille grâce à la commande insert into (...,les attributs) select from titanic;

Description des tables:

- table Passagers: les attributs sont : PassengersId integer (identification du passager), Name varchar(100), Sex varchar(10), Age float, Sibsp integer: le nombre de frères et soeurs ou conjoint de passagers, Parch integer :le nombre de parents ou d'enfants du passager. cette table sert donc à récupérer les informations sur les passagers du bateau,nous avons choisi de mettre en clé primaire PassengersId

- Table Tickets : Les attributs sont : Pclass integer, Name varchar(100), Age float, numTicket varchar(20), Fare float, Cabin varchar(20), Embarked varchar(1). cette table permet de récupérer les informations par rapport au ticket donc le numéro de ticket, la classe, le lieu d'embarquement, le prix, Age pour voir si le prix du ticket dépend de l'âge et le numéro de cabine nous avons décidé de mettre Name en clé primaire

- Table Survivants: Les attributs sont: PassengersId integer, Survived integer, Name varchar (100), cette table permet de nous fournir les informations sur les survivants ,en fonction de son identifiant PassengersId et Name qui sont les clés primaire,et nous avons aussi des clés étrangères: foreign key (PassengersId) references Passagers (PassengersId) foreign key (Name) references Tickets(Name) donc il faut créer la table survivants à la fin

Rapport phase 3

Nous avons réussi à créer toutes les tables que nous avons besoin avec les clés primaires et étrangères pour vérifier cela nous avons créé des requêtes pour pouvoir voir les tables grâce a dataframe. Nous avons ensuite essayé quelques requêtes pour voir si les commandes sql fonctionne correctement et afin d'interroger la base de données et visualiser les résultats. Enfin nous allons maintenant répondre aux requêtes afin de présenter une démonstration de la base de données

Questions 1 : Combien de classes de passagers différentes y avait-il à bord du Titanic ?

```
%%sql
```

```
select distinct pclass from Tickets;
```

```
pclass
```

```
1
```

```
2
```

```
3
```

```
%%sql
```

```
select count(distinct pclass) from tickets;
```

ici on fait un select distinct comme ca il n'affiche pas les doublons

Donc il y a au total 3 classes de passagers différentes à bord du Titanic

Questions 2 : Combien de passagers y avait-il dans chaque classe ?

```
%%sql
```

```
select count(passengerid), pclass from passengers left outer join tickets  
on passengers.name=tickets.name group by pclass order by pclass;
```

```
count    pclass
```

```
216      1
```

```
184      2
```

```
491      3
```

ici on a choisi de faire un select count(),

Donc il y a 216 passagers dans la classe 1, 184 passagers dans la classe 2, 491 passagers dans la classe 3.

Questions 3 : Combien de femmes et d'hommes y avait-il dans chaque classe ?

```
%%sql
```

```
select pclass, count(sex), sex from passengers left outer join tickets on  
passengers.name=tickets.name group by sex, pclass order by pclass;
```

```
pclass    count    sex
```

1	94	female
1	122	male
2	108	male
2	76	female
3	144	female
3	347	male

il y a 122 hommes et 94 femmes dans la classe 1

il y a 108 hommes et 76 femmes dans la classe 2

il y a 347 hommes et 144 femmes dans la classe 3

Questions 4 : Comptez le nombre et le pourcentage de survivants et de passagers morts.

%%sql

```
select count(survived)as nombre, count(survived)*100 / (select count(*)
from survivors) as pourcentage from survivors where survived=1;
```

nombre	pourcentage
342	38

Donc il y a 342 survivants soit 38% des passagers

ici nous avons utilisé count(survived)*100 / (select count(*) from survivors) pour calculer le pourcentage de survivants

on fait pareil pour les morts

%%sql

```
select count(survived)as nombre, count(survived)*100 / (select count(*)
from survivors) as pourcentage from survivors where survived=0;
```

Questions 5 : Visualiser la répartition des passagers survivants et morts par classe

%%sql

```
select pclass, count(survived),survived from tickets left outer join
survivants on survivants.name=tickets.name group by pclass,survived
order by pclass;
```

pclass	count	survived
1	80	0
1	136	1
2	87	1
2	97	0
3	372	0
3	119	1

**create view vivantclass1 as select PassengerId,Name,survived from
survivants natural join tickets where survived=1 and pclass=1 order by
passengerid;**

**nous avons décider de crée des vues dynamique pour bien visualiser les
répartition des passager survivant par classe
nous avons crée plusieurs vue pour chaque class**

**Questions 6 : Visualiser la répartition des passagers survivants et des
passagers décédés selon le sexe.**

%%sql

**select sex, count(survived),survived from passagers left outer join
survivants on survivants.passengerid=passagers.passengerid group by
sex,survived order by survived;**

sex	count	survived
female	81	0
male	468	0
female	233	1
male	109	1

%%sql

**create view vivantmale as select Passengerid,name,sex,survived from
survivants natural join passagers where survived=1 and sex='male'
order by passengerid;**

**nous avons décider de crée des vues dynamique pour bien visualiser les
répartition des passager survivant par sexe
nous avons crée plusieurs vue pour chaque sexe**

**pour ces commandes nous avons utilisé les modules
panda,sqlalchemy,plotly pour permettre de crée des histogrammes à
partir des requetes sql**

import pandas

from sqlalchemy import create_engine

import plotly.express as px

engine =

**create_engine("postgresql://postgres:0000@localhost/postgres") grace
a cette commande on se connecte sur notre base de données**

**df1 = pandas.read_sql("select distinct pclass from tickets;", engine) la
requete sql**

**fig = px.bar(df1, y="pclass",color="pclass") parametrage de
histogramme**

fig.show() sert à afficher

Tableau de Tâche:

Phase 1	1= Fait et 0= à faire
intégration de Postgresql	1
Ouvrir la base de donnée	1
Analyse de la structure	1
Trouver la cardinalité	1
Trouver le degrés	1
Chercher le domaine des attributs	1
Faire le tableau de tâche 1	1
Faire le tableau de tâche 2	1
Faire le tableau de tâche 3	1

Phase 2	1= Fait et 0= à faire
Créer les tables nécessaire	1
½ page de compte rendu	1
Mini script pour la création table	1
Fichier en notebook	1
Alimenter les tables	1
Préparer une démo	1

Phase 3	1= Fait et 0= à faire
S'amuser en créant des requêtes	1
un mini rapport avec visualisation et script en mode notebook	1
Démo / Présentation	1
Combien de classes différentes ?	1
Combien de passagers par classe	1
Combien de femmes et d'hommes y avait-il dans chaque classe ?	1
Comptez le nombre et le pourcentage de survivants et de passagers mort	1
Visualiser la répartition des passagers survivants et morts par classe	1
Visualiser la répartition des passagers survivants et des passagers décédés selon le sexe	1