

## Rapport Phase 1

---

Nous avons réalisé l'étude d'un ensemble de données récoltant plusieurs diverses informations sur les passagers du paquebot *Titanic* qui a fait naufrage en 1912.

Il permet donc de recenser et d'identifier les survivants du naufrage, ainsi de connaître les différents groupes de voyageurs selon leur classe de voyage et le tarif payé, informations qui peuvent éventuellement être utilisées à des fins historiques.

Il est possible de distinguer 2 catégories d'informations:

- l'identité de chaque passager  
→ (PassengerId, Name, Sex, Age)
- les informations sur leur séjour sur le bateau  
→ (PassengerId, Survived, Pclass, SibSp, Parch, Ticket, Fare, Cabin, Embarked)

L'attribut "PassengerId" peut agir comme une clé primaire de ce modèle de données.

La cardinalité de cette table est de card(891), ce qui signifie qu'il y a 891 tuples au total, et donc 891 passagers recensés et identifiés dans ce modèle de données.

Le degré dans ce modèle est de deg(12), indiquant qu'il y a 12 attributs différents dans cette relation.

Ces attributs sont listés ainsi:

n°	Attribut	Type d'attribut	Description
1	PassengerId	integer	C'est une valeur attribuée à chaque passager pour les identifier parmi tous les passagers. Cet attribut peut être une clé primaire.
2	Survived	integer	Cette valeur prend soit 0 ou 1 et elle indique si le passager a survécu (1) ou non (0) après le naufrage du <i>Titanic</i> .
3	Pclass	integer	Cet attribut allant de 1 à 3 qualifie la classe de voyage du passager, sachant qu'un voyage sur un tel bateau possède plusieurs classes de voyage selon le budget.
4	Name	string	"Name" désigne le nom et prénom du passager.
5	Sex	string	Cet attribut indique le sexe du passager: "male" pour un homme et "female" pour une femme.
6	Age	float	L'âge de l'individu lors du recensement est donné ici, s'il est connu. En effet, l'âge n'est parfois pas renseigné pour certains passagers.
7	SibSp	integer	
8	Parch	integer	

9	Ticket	string	Le numéro de ticket du passager est renseigné ici
10	Fare	float	L'attribut "Fare" (tarif en anglais) désigne le tarif payé par le voyageur ou la voyageuse pour le séjour sur le <i>Titanic</i> .
11	Cabin	string	Le numéro de cabine d'une personne figure ici, commençant par une lettre puis suivie de 2 voire 3 chiffres. Un passager peut avoir plusieurs cabines enregistrées à son nom. L'information semble également peu disponible, ne figurant que pour quelques passagers.
12	Embarked	string	Cet attribut prend la lettre S, Q ou C.

Nous avons donc défini les tâches à réaliser:

	Date	Tâche	Plateforme
<b>Phase 1</b>	18/09	Analyse de la structure de l'ensemble de données	En ligne (Discord)
		Définition des tâches à réaliser, planification	
	21/09	Rédaction du rapport de la phase 1	
<b>Phase 2</b>	29/09 au 04/10 au 16/10	Création des tables en SQL	En séance SAE  et/ou  En ligne
		Mini rapport sur la description des tables créées	
		Réalisation de script pour créer les tables	
	04/10 au 16/10	Réalisation d'un rapport et d'un script sur notebook	
		Alimentation des tables avec les données du fichier <i>titanic_train.csv</i>	
		Rédaction du rapport de la phase 2, mise en commun des autres rapports	
<b>Phase 3</b>	21/10 au 01/11	Création d'un script en notebook et des visualisations	En séance SAE  et/ou  En ligne
		Création d'une présentation	
	1/11 au 18/11	Création d'une démonstration	
		Traitement des requêtes pour la démonstration	
		Rédaction du rapport de la phase 3	