

PORTFOLIO DE Sébastien GRIGORE

BUT SCIENCE DES DONNEES
1ere année

Poursuite d'étude :
Master informatique

Objectif professionnel :
Data Scientist



Compétences traitées

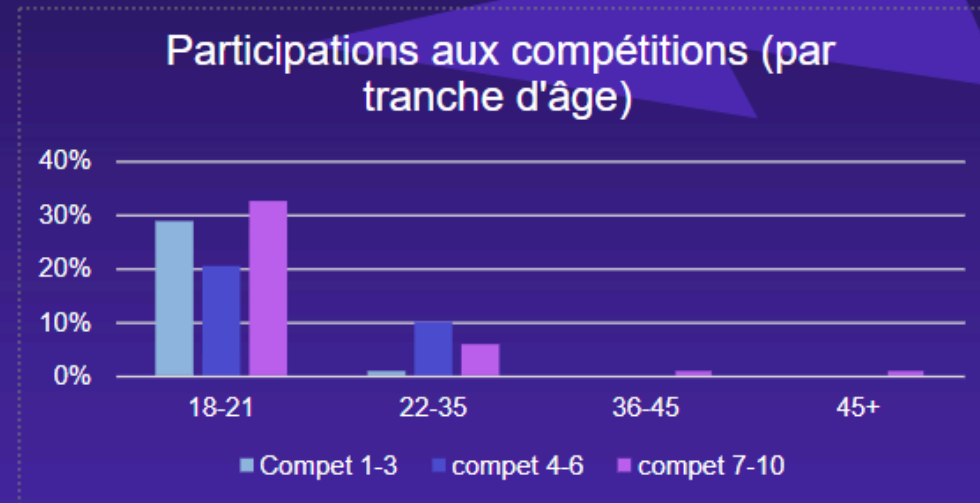
<u>Traiter</u>	Traiter des données à des fins décisionnelles.
<u>Analyser</u>	Analyser statistiquement les données.
<u>Valoriser</u>	Valoriser une production dans un contexte pro.
<u>Modéliser</u>	Modéliser les données dans un cadre statistique.

PROJET 1 : Enquête sur l'Esport

Engagement dans l'Esport

Quelle tranche d'âge participe le plus à des compétitions ?

Ce sont ceux entre 18 et 21 ans qui ont participé le plus à des compétitions. (sur un total de 122 répondants).



Jeux vidéos préférés des personnes sachant ou non définir l'esport



Les participants qui savent définir l'Esport préfèrent-ils certains jeux ?

League of Legends est le jeu qui ressort le plus en tant que jeu préféré (sur un total de 287 répondants).

Contexte : Nous avons réalisé en équipe de 4 une enquête pour connaître l'impact de l'Esport en France auprès de 310 personnes âgées de 18 à 50 ans.

Compétences Mobilisées :

Traiter : Normalisation de données complexes (réponses ouvertes) pour produire des graphiques clairs.

Analyser : Création de statistiques croisées pour identifier les tendances comportementales.

Valoriser : Valorisation du travail à l'aide d'une présentation orale.

Bilan personnel :

Cette expérience a développé mes compétences en recherche d'information et en normalisation de données complexes pour produire des graphiques clairs, malgré un apprentissage encore à renforcer sur l'analyse pure.

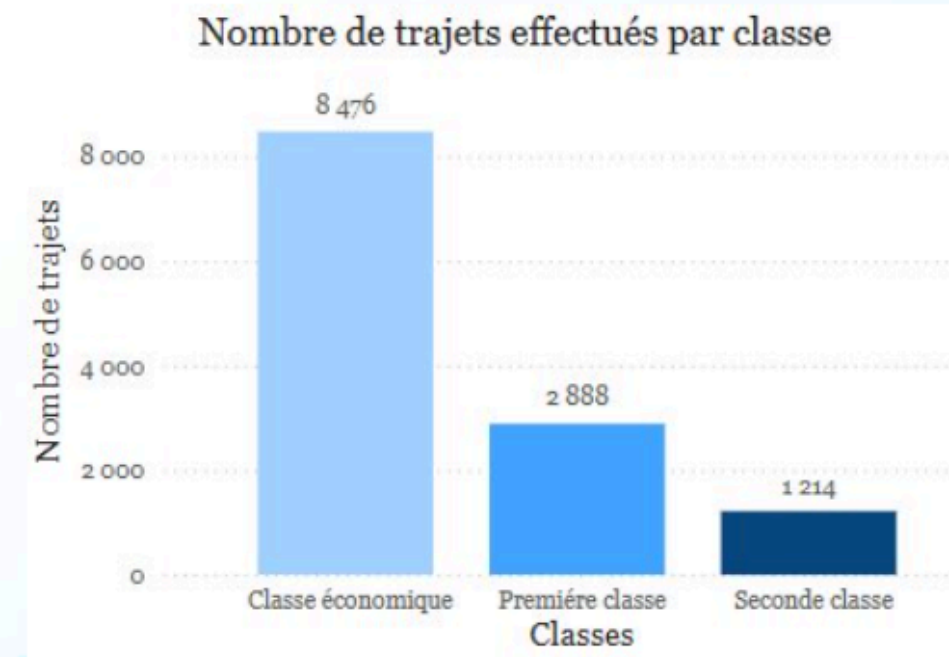
PROJET 2 : Rapport PowerBI pour la SNCF

II. Suivi des trajets effectués

Nombre de trajets effectués par ville

Villes	nb_trajets_pas_annulés
Paris	2 322
Lyon	697
Marseille	627
Rouen	413
Nantes	405
Lille	395
Rennes	389
Bordeaux	363
Cannes	350
Avignon	342
Toulouse	332
Le Havre	327
Montpellier	322
Strasbourg	220
Total	12 578

- Forte centralisation trafic ferroviaire autour grandes villes.



- Classe économique priorisée, suggérant souci d'économie ou accessibilité plus forte de la classe économique.

Contexte : Création d'un tableau de bord décisionnel PowerBI en équipe (3 personnes) pour l'analyse du trafic ferroviaire.

Compétences Mobilisées :

Traiter : Nettoyage de bases de données Excel et structuration des relations entre tables.

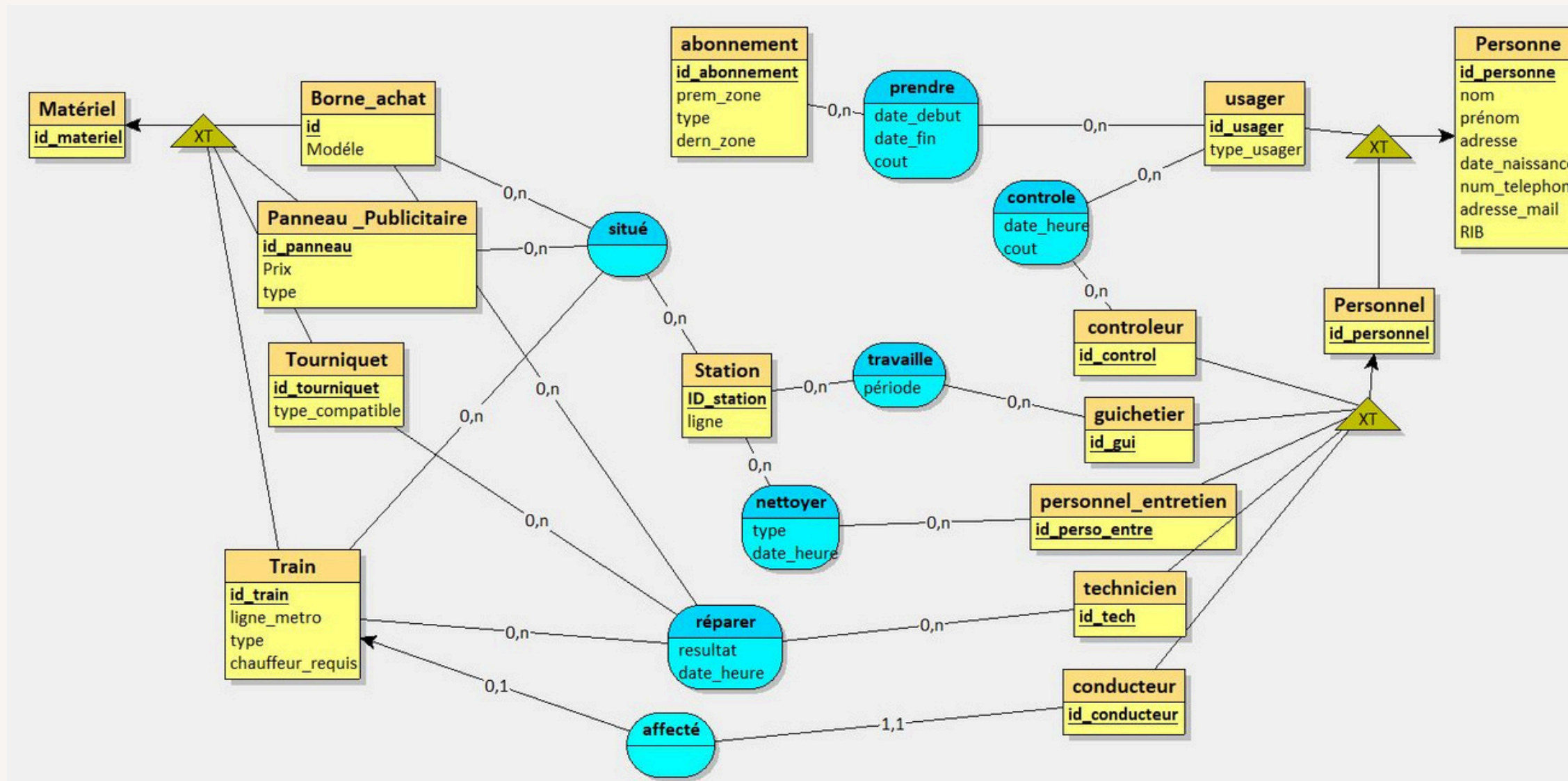
Analyser : Création de visualisations pertinentes et interprétation des tendances de flux.

Valoriser : Soutenance orale et rédaction d'une documentation technique (Guide d'utilisation).

Bilan personnel :

Ce projet a renforcé ma capacité à collaborer en équipe et à faire preuve de rigueur analytique. J'ai su valoriser des points d'ombre de l'analyse lors de la présentation finale, prouvant ma maîtrise du sujet.

PROJET 3 : Base de données pour la RATP



Contexte : Réalisation en équipe de 3 d'une base de données accompagnée d'un rapport explicatif, d'un cahier des charges et d'une présentation orale.

Compétences Mobilisées :

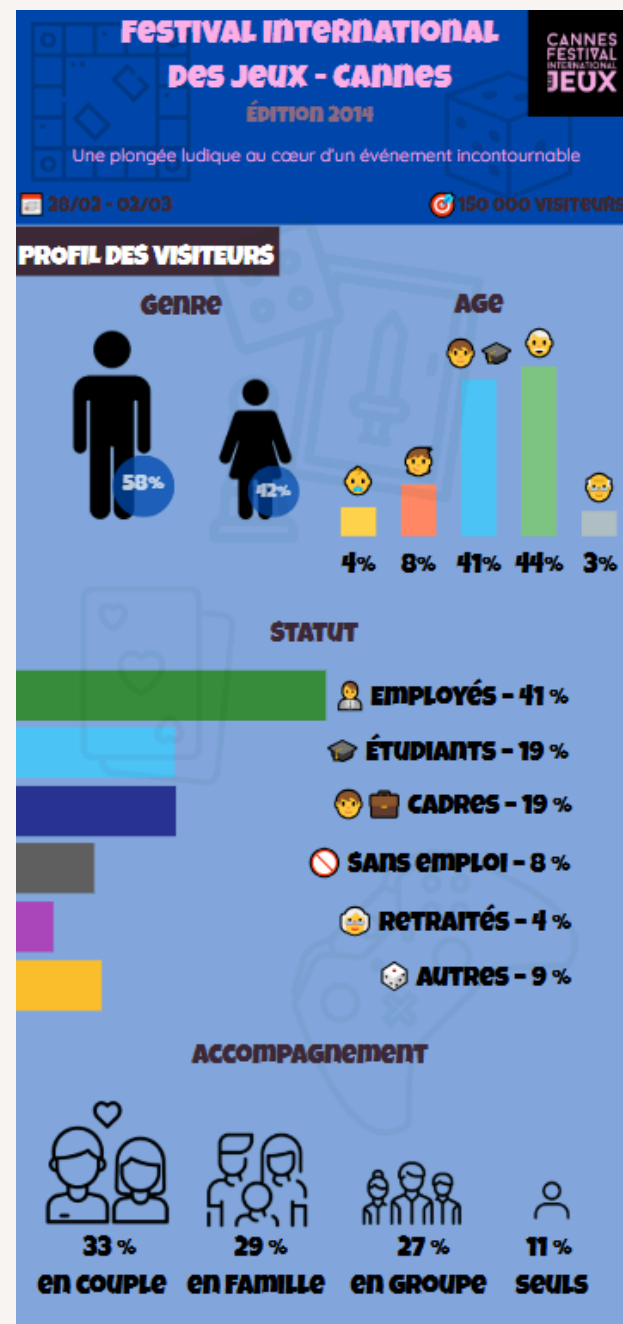
Traiter : Exploitation des informations fournies pour structurer techniquement la base de données.

Valoriser : Présentation du projet à l'oral et rédaction de la documentation technique.

Bilan personnel :

Ce projet fut difficile en raison de la complexité du schéma MCD (notion d'héritage), mais la valorisation finale a permis de compenser les lacunes initiales de compréhension.

PROJET 4 : Infographie Festival des Jeux



Contexte :

Création d'une infographie synthétisant un sondage réalisé en 2014 sur les visiteurs du Festival International des Jeux à Cannes, suivie d'une présentation orale.

Compétences Mobilisées :

Traiter : Analyse et tri des informations du questionnaire pour alimenter le support visuel.

Analyser : Réalisation de graphiques à partir des données traitées.

Valoriser : Conception d'un visuel structuré avec un code couleur cohérent pour une présentation efficace.

Bilan personnel :

Ce projet m'a appris à structurer graphiquement des données parfois confuses. C'est un travail globalement réussi pour une première infographie malgré quelques manques de clarté dans les nomenclatures initiales.

PROJET 5 : Sondage sur l'impact du numérique sur l'apprentissage

Contexte de l'étude

Taux de redressement

Taux de redressement (Tranche d'âge par genre)							
Tranche d'âge	% Cible (H)	% Cible (F)	% Sondage (H)	% Sondage (F)	Coefficient de redressement (H)	Coefficient de redressement (F)	Analyse
- 15 ans	51,20%	48,80%	50,00%	50,00%	1,02	0,98	Les femmes sont légèrement sur-représentées.
15 - 18 ans	51,64%	48,36%	51,72%	48,28%	1,00	1,00	Pas de différence.
19 - 20 ans	51,06%	48,94%	76,60%	23,40%	0,67	2,09	Les femmes sont sous-représentées.
21 - 23 ans	51,06%	48,94%	50,00%	50,00%	1,02	0,98	Les hommes sont légèrement sur-représentés.
24 - 26 ans	49,97%	50,03%	50,00%	50,00%	1,00	1,00	Pas de différence.
27 - 30 ans	49,97%	50,03%	66,67%	33,33%	0,75	1,50	Les hommes sont sous-représentés.

Taux de redressement (Genre)						
Population	% Cible (H)	% Cible (F)	% Sondage (H)	% Sondage (F)	Coefficient de redressement (H)	Coefficient de redressement (F)
%	50,85%	49,15%	57,50%	42,50%	0,88	1,16

5

Contexte : Nous avons réalisé en équipe un sondage auprès de 125 étudiants, suivi d'un rapport statistique et d'une présentation orale pour analyser l'impact des outils numériques sur leurs habitudes d'apprentissage.

Compétences Mobilisées :

Traiter : Nettoyage et préparation des données issues des 125 réponses obtenues.

Analyser : Étude approfondie des comportements d'apprentissage et des tendances numériques.

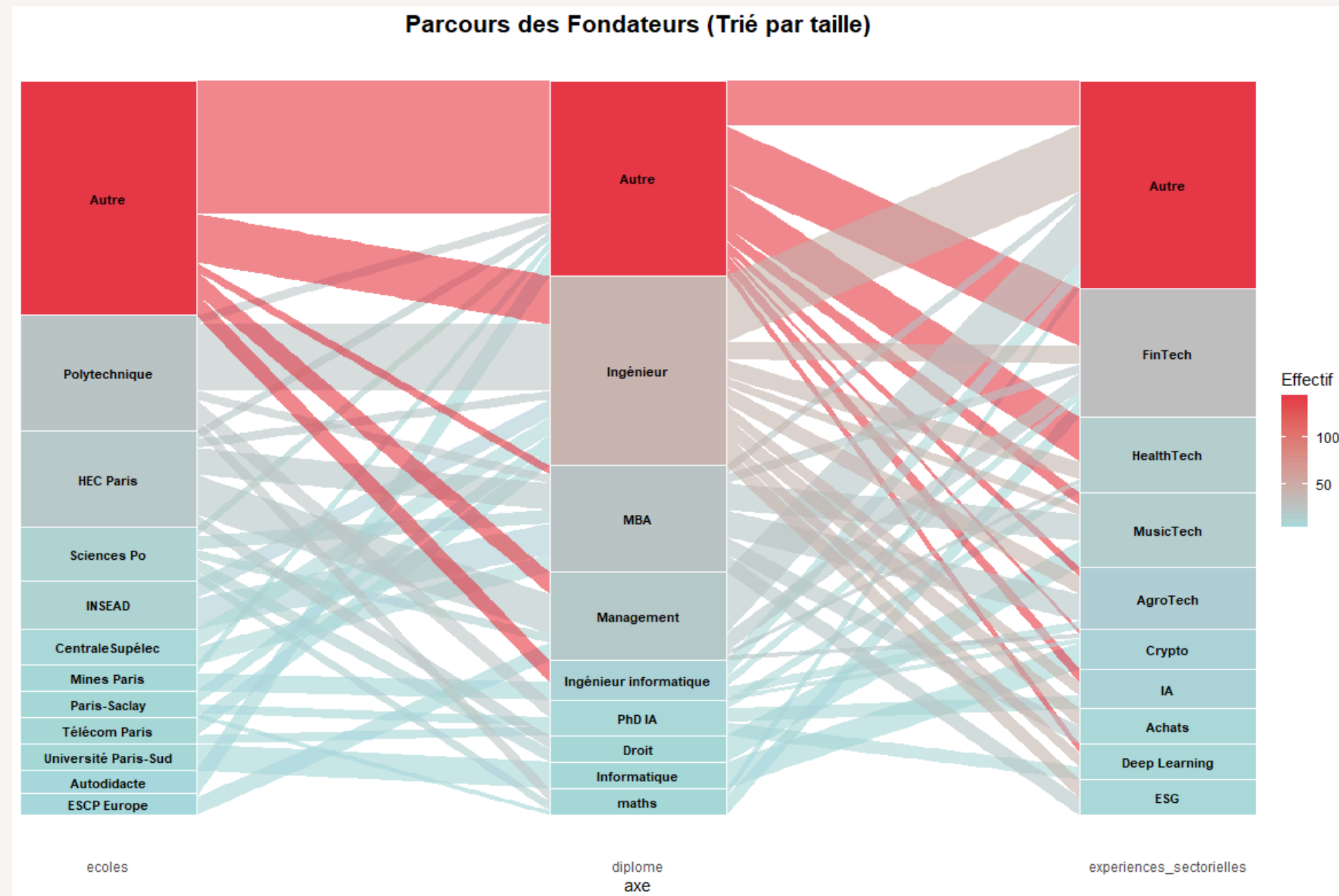
Valoriser : Restitution des résultats via un rapport écrit et une présentation orale pour communiquer nos conclusions.

Modéliser : Utilisation de techniques de sondage avancées, comme le calcul du taux de redressement pour corriger les sous-représentations (ex: femmes à 2,09 sur la tranche 19-20 ans).

Bilan personnel :

Ce projet a renforcé ma rigueur mathématique, notamment grâce à la mise en pratique du redressement de données pour fiabiliser nos conclusions. La présentation orale a été l'occasion de démontrer ma capacité à vulgariser des concepts techniques complexes tout en défendant nos analyses statistiques de manière professionnelle.

PROJET 6 : Les fondateurs de Licornes Françaises



Contexte : Nous avons créé une base de données inédite via des recherches web pour analyser les parcours types des fondateurs de licornes françaises et identifier des corrélations entre leurs profils, le tout restitué lors d'une présentation orale.

Compétences Mobilisées :

Traiter : Collecte de données brutes sur Internet et structuration d'un jeu de données complet sur Excel avec des variables spécifiques (études, âge, secteur, etc.).

Analyser : Identification de corrélations entre les variables et création de graphiques pertinents pour répondre à la problématique des "parcours types".

Valoriser : Synthèse des résultats et présentation orale des conclusions sur l'écosystème entrepreneurial français.

Bilan personnel : Ce projet m'a permis de prendre conscience des défis liés à la création de données : de la recherche d'informations à la gestion des limites de l'analyse. Bien que nous ayons identifié des tendances fortes, l'étude a révélé une réelle diversité de parcours possibles. Ce travail a affiné ma capacité à transformer des données éparses en une analyse d'aide à la décision cohérente.

PROJET 7 : Programmation Java – Manipulateur de Polynômes

```
4 //représente un polynôme et contient les opérations algébriques
5 class Polynome implements IPolynomeOperations {
6     //initialisation
7     //stocke le polynome sous forme de liste
8     List<Monome> monomes;
9
10    public Polynome() {
11        monomes = new ArrayList<Monome>();
12    }
13
14    public void ajouterMonome(Monome m) {
15        monomes.add(m);
16    }
17
18    //convertie "3x^2 - 2x" en objet Polynome
19    public static Polynome depuisChaine(String texte) {
20
21        Polynome p = new Polynome();
22        texte = texte.toLowerCase();
23
24        //remplacement " - " par "+-" pour pouvoir découper la chaîne ((Ex: "x - 2" devient "x + -2")
25        texte = texte.replace("-", "+-");
26        String[] parties = texte.split("\\+");
27
28        for (int i = 0; i < parties.length; i++) {
29            String part = parties[i].trim();
30
31            if (part.equals("")) continue; // sécurité si double espace
32
33            double coef = 1;
34            int pos = 0;
35
36            //analyse du Coefficient
37            if (part.charAt(0) == '-') {
38                coef = -1; //cas du "-x"
39                pos = 1;
40            } else if (Character.isDigit(part.charAt(0))) {
41
42                //si ça commence par un chiffre, on lit tout le nombre
43                int fin = 0;
44                while (fin < part.length() && Character.isDigit(part.charAt(fin))) {
45                    fin++;
46                }
47                coef = Double.parseDouble(part.substring(0, fin));
48                pos = fin;
49            }
50        }
51    }
52}
```

Contexte : Conception et implémentation d'un programme en Java pour manipuler des polynômes complexes (ex: $2x^2y^3z + 2xy$), incluant une analyse algorithmique préalable et une présentation orale finale détaillant le fonctionnement et les limites du système.

Compétences Mobilisées :

Traiter : Analyse descendante (Top-Down) pour décomposer le problème et choix stratégique des structures de données.

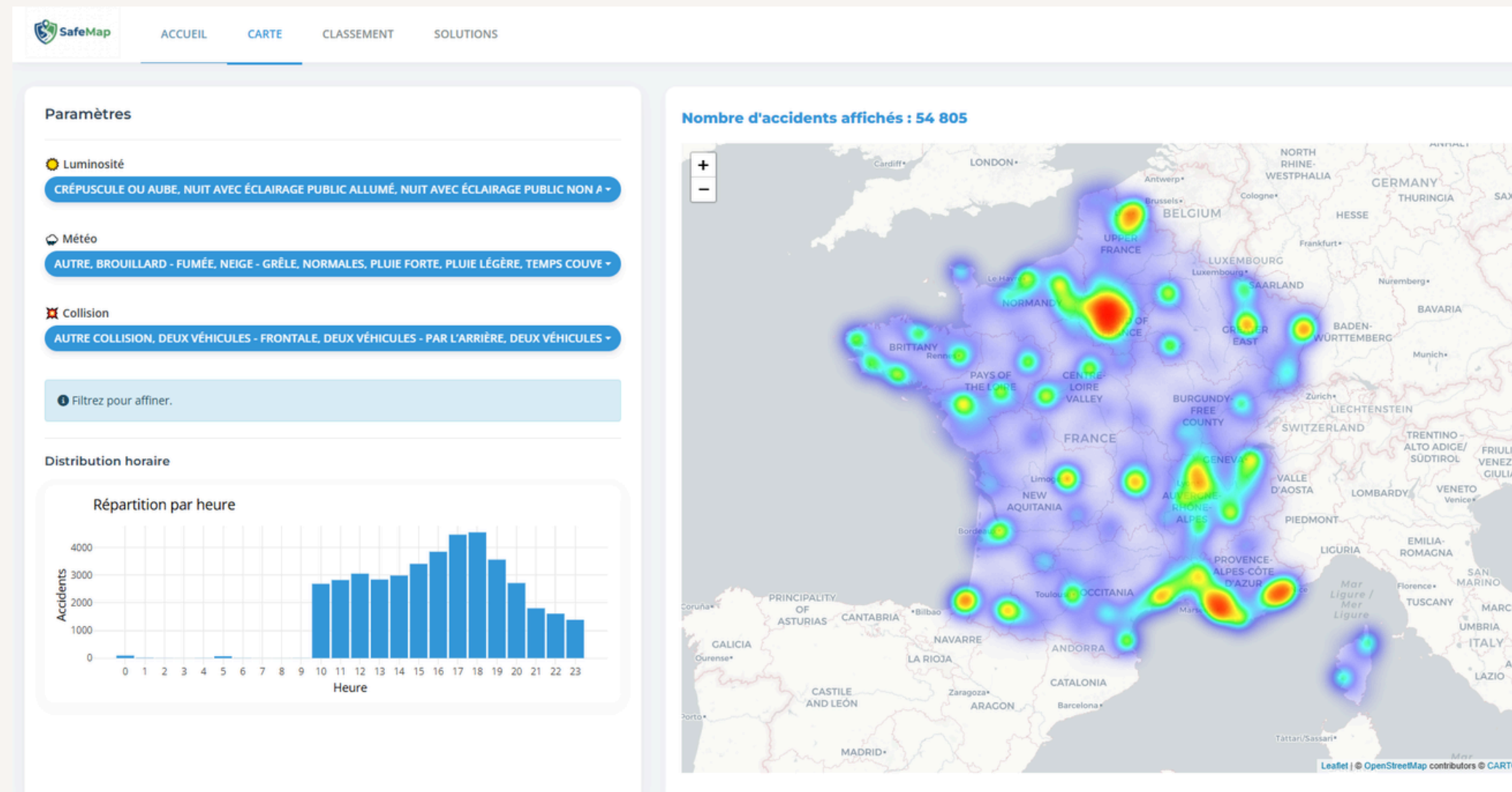
Analyser : Élaboration de solutions algorithmiques pour la saisie, le calcul et l'affichage de polynômes sous forme réduite.

Valoriser : Présentation orale du programme et démonstration des opérations algébriques réalisées par l'interface.

Modéliser : Application rigoureuse des principes de la Programmation Orientée Objet (POO) pour garantir un code structuré et évolutif.

Bilan personnel : Ce projet m'a permis de me plonger concrètement dans la logique de programmation et de comprendre l'importance d'une structure de code propre. J'ai appris à anticiper les éventuelles failles du programme et à les exposer avec clarté lors de la présentation. La manipulation de polynômes multi-variables a renforcé ma capacité à concevoir des algorithmes performants tout en respectant une architecture logicielle rigoureuse.

PROJET 8 : Application R-Shiny "SafeMap" – Sécurité Routière



Contexte : Conception en équipe de 4 d'une application interactive sous R-Shiny permettant de visualiser les zones et facteurs d'accidents en France métropolitaine, le tout géré via la méthode Agile (Jira) et présenté intégralement en anglais.

Compétences Mobilisées :

Traiter : Manipulation et nettoyage d'une base de données volumineuse sur la sécurité routière pour en extraire des variables temporelles et géographiques.

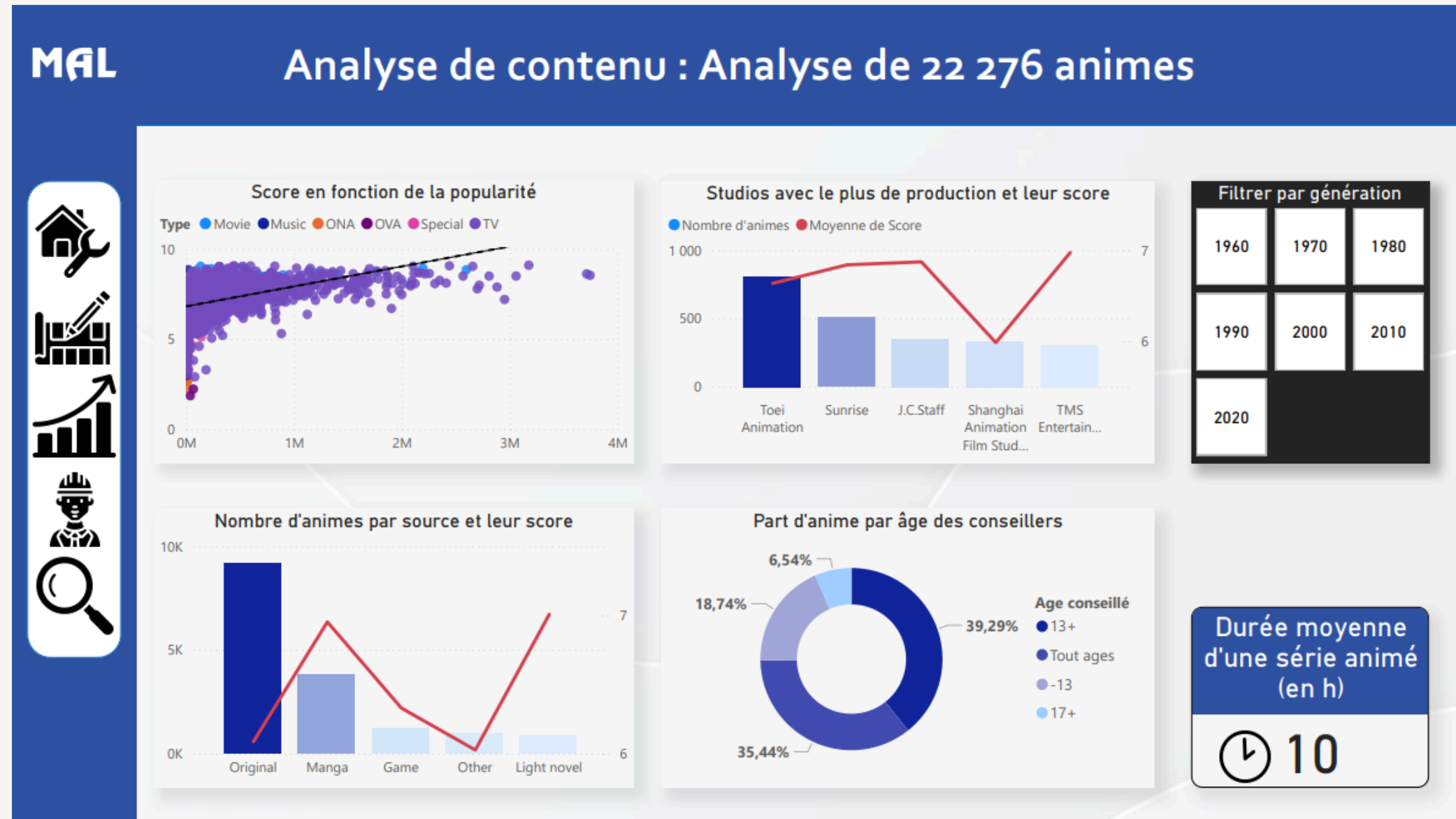
Analyser : Identification des corrélations entre les conditions (météo, luminosité, type de collision) et la fréquence des accidents.

Valoriser : Soutenance du projet en anglais et gestion de la collaboration d'équipe avec les rôles Scrum (Product Owner, Scrum Master, Developers) sur Jira.

Modéliser : Développement de l'interface utilisateur et du serveur sous R-Shiny pour permettre un filtrage dynamique des données en temps réel.

Bilan personnel : Ce projet a été ma première immersion dans le développement d'applications de data-visualisation. J'ai appris à structurer mon travail au sein d'un cadre Agile, ce qui a rendu la répartition des tâches très efficace. Relever le défi de présenter des solutions techniques complexes en anglais a également renforcé ma capacité à communiquer dans un contexte international.

PROJET 9 : Analyse de données – Facteurs de succès de l'animation japonaise



Contexte : Réalisation en binôme d'un rapport Power BI explorant la base de données exhaustive de MyAnimeList (1940–2023) pour identifier les facteurs de succès et de fidélité des spectateurs à travers l'analyse de plus de 22 000 titres, conclue par une présentation orale.

Compétences Mobilisées :

Traiter : Importation et nettoyage d'une base de données massive (plus de 22 000 entrées) pour garantir la fiabilité des indicateurs temporels et qualitatifs.

Analyser : Croisement de variables clés (score, popularité, studio, source) pour définir scientifiquement ce qui constitue une "pépité".

Valoriser : Conception d'un dashboard interactif et soutenance orale des résultats pour démontrer la pertinence de l'analyse de données dans le secteur culturel.

Bilan personnel : Ce projet a été l'occasion parfaite de lier ma passion pour les animés à mon expertise technique en Data Analyse. J'ai particulièrement apprécié la recherche de corrélations pour répondre à notre problématique, tout en renforçant mon aisance à l'oral pour expliquer des graphiques complexes. Ce travail a confirmé ma capacité à transformer un centre d'intérêt personnel en une étude statistique rigoureuse et professionnelle.

PROJET 10 : Communication Digitale – Roadmap du BUT SD



Contexte : Conception d'une roadmap stratégique pour la page Instagram du BUT Science des Données de Sophia Antipolis, visant à illustrer le parcours type d'un étudiant pour devenir Data Scientist.

Compétences Mobilisées :

Traiter : Sélection et hiérarchisation des compétences clés (techniques et soft skills) acquises durant la formation pour créer un fil conducteur logique.

Analyser : Identification des besoins des futurs étudiants pour proposer un contenu visuel pédagogique et attractif.

Valoriser : Création d'une identité visuelle forte sous forme d'escalier de progression, intégrant les outils (Power BI, R, Python) et les domaines (Statistiques, Économie, Anglais).

Bilan personnel : Ce projet de communication m'a permis de prendre du recul sur mon propre parcours et de synthétiser la richesse de ma formation. J'ai appris à adapter un discours technique pour un réseau social, tout en veillant à ce que l'infographie soit à la fois esthétique et porteuse d'un message clair. C'est un travail qui a renforcé ma capacité à valoriser des compétences variées auprès d'une audience cible.