

PORTFOLIO

Diallo Thierno



1. Présentation personnelle.....	2
2. Pourquoi la science des données ?.....	3
3. Focus sur une spécialité.....	3
4. Bilan des compétence à partir des SAE.....	4
A. SAE 101 : Création de reporting à partir de données stockées dans une SGBD relationnel.....	4
B. SAE 102 : Lecture et écriture de fichier de données.....	5
C. SAE 103 : Préparation et synthèse d'un tableau de donnée en vue d'une analyse exploratoire simple.....	5
D. SAE 104 : Apprendre en situation la production de données en entreprise.....	6
E. SAE 105 : Présentation en anglais d'un territoire économique et culturel.....	6
F. SAE 106 et 107 : Mise en œuvre d'une enquête.....	7
G. SAE 201 : Conception et implémentation d'une base de données.....	7
H. SAE 202 : Estimation par sondage simple.....	8
I. SAE 203 : Régression sur données réelles.....	8
J. SAE 204 : DATAVIZ.....	8
K. SAE 205 : Construction et présentation d'indicateurs de performance.....	9
L. SAE 206 : Analyse de données, reporting et datavisualisation.....	10
5. Conseil d'un professionnel.....	10
6. Objectifs.....	11

1.Présentation personnelle

Je suis Thierno Diallo, étudiant en première année de BUT SD(ex statistique et informatique décisionnel) à IUT de villetaneuse. Dans cette formation nous avons abordé plusieurs notions allant des mathématiques à l'informatique passant par la communication qui m'a permis de développer des compétences comme le traitement, l'analyse et la valorisation des données.

J'aime particulièrement l'aspect analyse de cette formation, car elle fait appel à des notions de mathématique et d'informatique qui sont deux domaines qui me passionnent, mais également à l'esprit critique dont il faut faire preuve vis-a-vis de ce que les données racontent et de comment on leurs donnent une signification.

Je souhaite devenir data scientist. Pour atteindre cet objectif, j'ai commencé par un BUT SD afin de développer mes compétences en traitement, analyse et valorisation de données. Je travaille également de mon côté en apprenant des outils tels que "scikit-learn", "scipy" ou "plotly" qui ne sont pas au programme des cours, mais qui sont très utiles dans le monde de l'analyse de données. Je compte également poursuivre par un diplôme d'ingénieur dans le domaine des mathématiques appliquées afin d'acquérir toutes les connaissances nécessaires pour atteindre mon objectif.

- Mes compétences :

Cette formation m'a permis d'acquérir, d'une part des compétences en traitement et analyse de données grâce au différentes ressources comme "mathématique", "statistique descriptive", "base de la programmation", "SQL" et bien d'autre qui nous permettent de tirer des informations dans une grande masse de données. D'autre part, j'ai développé aussi des soft skill tels que le sens du collectif et de mon écoute afin de comprendre les besoins de chaque membre de mon équipe, ce qui a renforcé leur confiance en moi. Ma confiance en moi et ma capacité de gestion du stress me permettent de travailler dans un environnement sous pression tout en gérant efficacement mon temps, ce qui m'évite de perdre mes moyens. j'ai un bon esprit critique et le sens des responsabilités. Ainsi toutes ces capacités de leader et de pensée stratégique, me permettent de m'adapter et de guider mon équipe au terme de nos projets.

- Mes qualités :

Tout d'abord, je suis sérieux, car lorsque j'ai une tâche à accomplir, je m'implique pleinement et mène le projet à terme. Cela s'illustre dans différents projets que j'ai eu à réaliser, comme la SAE de programmation, qui demandait beaucoup de préparation en amont.

Ensuite, je me définirais aussi comme quelqu'un de calme et posée, cela se reflète dans les différentes relations que j'ai eu, que ça soit de travail ou amicale, j'arrive à gérer les différends qui émane des groupes dans lesquels je suis et à éviter les conflits.

Puis, je suis une personne persévérante parce que lorsque j'ai une tâche à accomplir, peu importe la difficulté, je la mène à terme et comme illustration, pendant mes années de lycée, j'étais dans une série qui demandait énormément de travail, mon objectif était d'être parmi les meilleurs.

et j'ai tout fais pour y arriver, ce qui me poussait à bosser parfois jusqu'à 1h du matin

Enfin, je suis une personne qui a une grande capacité d'abstraction, un esprit d'analyse, et travaille de manière méthodique. j'ai acquis c'est qualité grâce au étude scientifique que j'ai mené. D'abord au collège, j'étais très à l'aise avec les mathématiques, physique et chimie. Au lycée, j'ai continué avec la série C(math, pc) et aussi au but. j'ai donc acquis cette rigueur et méthode dont il faut faire preuve dans ces domaines. Mon plus grand défaut concerne la gestion du stress face à des situations inédites, ce qui me pousse à anticiper les choses, et parfois un peu trop. Je suis également réservée et fidèle à mon entourage avec qui j'aime partager des moments de joie.

2. Pourquoi la science des données ?

Nous vivons dans un monde où nous générons une quantité énorme de données, estimée à environ 29000 Go/s. Toutes ces données sont des mines d'or pour les entreprises, grâce à ces données, ils peuvent concocter de meilleur produit/service pour leurs clients, faire plus de bénéfice et j'en passe. Et ce, peu importe le secteur, tous les secteurs allant de la médecine au commerce, ont besoin des informations contenues dans ces données. Cependant ces données ont besoin d'être traitées avant d'être utilisées. Or, pour traiter ces données, ils ont besoin de personne avec des connaissances en mathématiques(statistique, probabilité, etc...) mais également en informatique car il est impossible de faire ces analyses à la main. Étant curieux et passionné des mathématiques, de l'informatique et des nouvelles technologies, travailler dans ce secteur est pour moi une évidence, car il me permet d'allier toutes mes passions et de vivre de ces dernières.

3. Focus sur une spécialité

La modélisation prédictive est une partie de la Data Science qui utilise les statistiques (ex : calcul de corrélation) et l'informatique (des algorithmes de d'IA) pour prédire des événements. Les étapes pour faire de l'analyse prédictive sont : définition de la problématique, la collecte de données, le nettoyage et la transformation de ces données, la sélection du modèle, l'entraînement du modèle, et l'évaluation du modèle avant d'être déployé.

Les techniques les plus utilisées dans l'analyse prédictive incluent les régressions linéaires (prédire des valeurs quantitative : salaire, prix immobilier) et logistiques (pour une classification binaire), les arbres de décision, le random forest, le KNN, les réseaux de neurones et les SVM.

L'analyse prédictive a de nombreux domaines d'application, tels que :

- Marketing : Prédiction des comportements d'achat et segmentation de la clientèle.
- Finance : Détection de fraudes, gestion des risques et prédiction des tendances boursières.
- Santé : Prédiction des épidémies, personnalisation des traitements et diagnostic précoce des maladies (tumeur, cancer).
- Transport : Optimisation des itinéraires et gestion prédictive de la maintenance des véhicules.
- Industrie : Prédiction des pannes d'équipement et gestion de la chaîne d'approvisionnement.

Par exemple, dans le domaine de la santé, les modèles prédictifs peuvent être utilisés pour savoir si un patient a une tumeur maligne ou bénigne, en fonction de quelque caractéristique (rayon moyen, texture, etc ...), permettant ainsi de lui attribuer un meilleur traitement.

L'analyse prédictive m'intéresse beaucoup, d'une part, car elle nous permet de résoudre des problèmes concrets auquel on est confronté chaque jour, d'autre part, elle a un vaste champ d'application, en partant de la finance à la médecine, l'analyse prédictive peut apporter du progrès en aidant à la prise de décision

4. Bilan des compétence à partir des SAE

A. SAE 101 : Création de reporting à partir de données stockées dans une SGBD relationnel

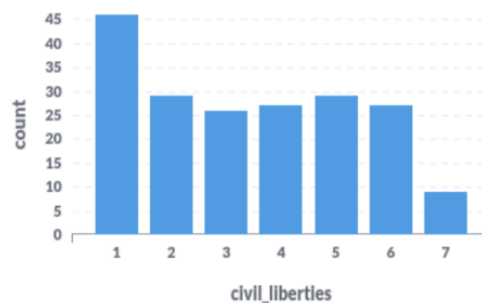
2. Requête SQL permettant de calculer la moyenne de deux indicateurs de liberté en 2020 pour chacune des régions.

Requête : « `SELECT AVG(civil_liberties) AS moy_civil_liberties, AVG(political_rights) AS moy_political_rights, region.name FROM freedom JOIN country USING(id_country) JOIN region USING(region_code) WHERE freedom.year = 2020 GROUP BY region.name;` »

3. Diagrammes en barre pour observer la répartition des pays en fonction des indicateurs.

a) Civil liberties

Requête : « `SELECT COUNT(civil_liberties), civil_liberties FROM freedom WHERE year = 2020 GROUP BY civil_liberties;` ». Ce qui nous donne :



Commentaire :

Nous constatons que environ 25% des pays ont reçu une note de 1. Ensuite, le nombre de pays ayant obtenu une note de 2 à 6 est environ le même, soit 25 pays par modalité. Et enfin, nous avons environ 10 pays qui ont obtenu une note de 7. Nous pouvons dire ainsi que qu'il y'a plus de la moitié

des pays qui ont un degré de liberté civil acceptable (soit moins de 3,5).

Cette SAE avait pour objectif la collecte, analyse et la synthèse des données enregistrées dans une base de données Postgresql. Pour sa réalisation, elle demandait des prérequis, notamment des connaissances en SGBDR, SQL et analyse de graphique. Cette SAE était importante à mes yeux, car elle est très similaire au situation auxquels on peut être confronté en entreprise, c'est une compétence très demandée. certe j'avais déjà quelques connaissances en SQL, mais grâce à cette SAE j'ai pu renforcer ces connaissance et aussi appris de nouvelle chose, tel que les sous requête. mais le point positif le plus important est d'avoir pu mettre mes connaissances en application dans une situation réelle. et je compte continuer sur cette lancer en apprenant de nouvelles choses (normalisation de base de données, les transactions sql, et bien d'autres). et pour cela je compte être encore plus impliqué dans ma formation à l'IUT, mais également faire des recherches personnelles (google, youtube).

B. SAE 102 : Lecture et écriture de fichier de données.

reponse

```
[5]: for transaction in transactions:

    #recuperation des information utile a la fracture
    date_commande = transaction[2]
    numero_fracture = transaction[0]

    clientID = transaction[1]
    client_name = clients[clientID][0]
    client_adresse = clients[clientID][1]

    # [nameproduct, quantite, prix_unitaire, prix_total, fournisseur](format des tableau contenu dans product)
    # recuperation des info pour chaque produit
    products = []
    total_ht = 0
    i = 3
    while i < len(transaction):
        refID = transaction[i][0]
        quantity = transaction[i][1]
        nom_produit = references[refID][2]
        prix_unitaire = references[refID][3]
        prix_total = prix_unitaire * quantity
        total_ht += prix_total

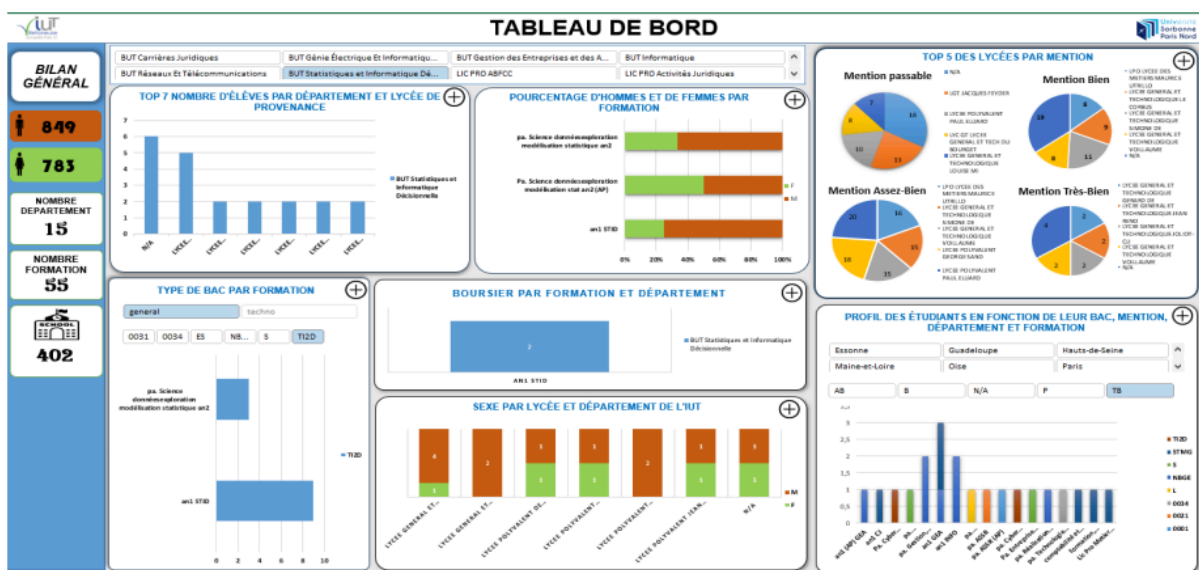
    product = [nom_produit, quantity, prix_unitaire, prix_total]
    products.append(product)
```

Cette SAE avait pour but la collecte, l'examen et la synthèse de données enregistrées dans des fichiers de csv. Pour sa réalisation, elle demandait des prérequis, notamment des connaissances en manipulation de fichiers, en programmation Python. Cette SAE était primordiale à mes yeux, car elle est très similaire aux situations auxquelles on peut être confronté en entreprise; c'est une aptitude très recherchée.

Certes, j'avais déjà quelques connaissances en manipulation de fichiers et en programmation, mais grâce à cette SAE, j'ai pu consolider ces connaissances et aussi apprendre de nouvelles techniques et astuces pour manipuler les fichiers de données avec python. L'aspect positif le plus important est d'avoir pu mettre mes connaissances en application dans une situation réelle.

Je compte continuer sur cette lancée en apprenant de nouvelles choses, comme la gestion des fichiers JSON et XML, les différentes bibliothèques (pandas, numpy, plotly), l'optimisation de la lecture et de l'écriture de fichiers volumineux, et les meilleures pratiques de manipulation de données. Pour cela, je compte être encore plus impliqué dans ma formation à l'IUT, mais également faire des recherches personnelles sur des plateformes comme Google, YouTube et Kaggle.

C.SAE 103 : Préparation et synthèse d'un tableau de donnée en vue d'une analyse exploratoire simple



Cette SAE avait pour objectif la préparation et la synthèse d'un tableau de données en vue d'une analyse exploratoire simple et la création d'un dashboard avec Excel. Pour sa réalisation, elle exigeait des prérequis, notamment des connaissances en manipulation de données, en utilisation d'Excel, et en analyse de données. Cette SAE a été pour moi une expérience déterminante, car elle m'a plongé dans un environnement de travail réaliste et m'a permis d'acquérir des compétences essentielles et très demandées par les entreprises.

Certes, j'avais déjà quelques notions en manipulation de données et en utilisation d'Excel, mais grâce à cette SAE, j'ai pu renforcer ces compétences et aussi apprendre de nouvelles techniques, telles que l'utilisation de formules avancées, de tableaux croisés dynamiques et de graphiques. L'aspect positif qui m'a le plus marquée est d'avoir pu mettre mes connaissances en application dans une situation concrète.

Je compte continuer sur cette lancée en apprenant de nouvelles choses, comme l'automatisation des tâches avec VBA, l'optimisation des tableaux de données volumineux, et les meilleures pratiques en visualisation de données. Pour cela, je compte être encore plus impliqué dans ma formation à l'IUT, mais également faire des recherches personnelles sur des plateformes comme Google et YouTube.

D.SAE 104 : Apprendre en situation la production de données en entreprise



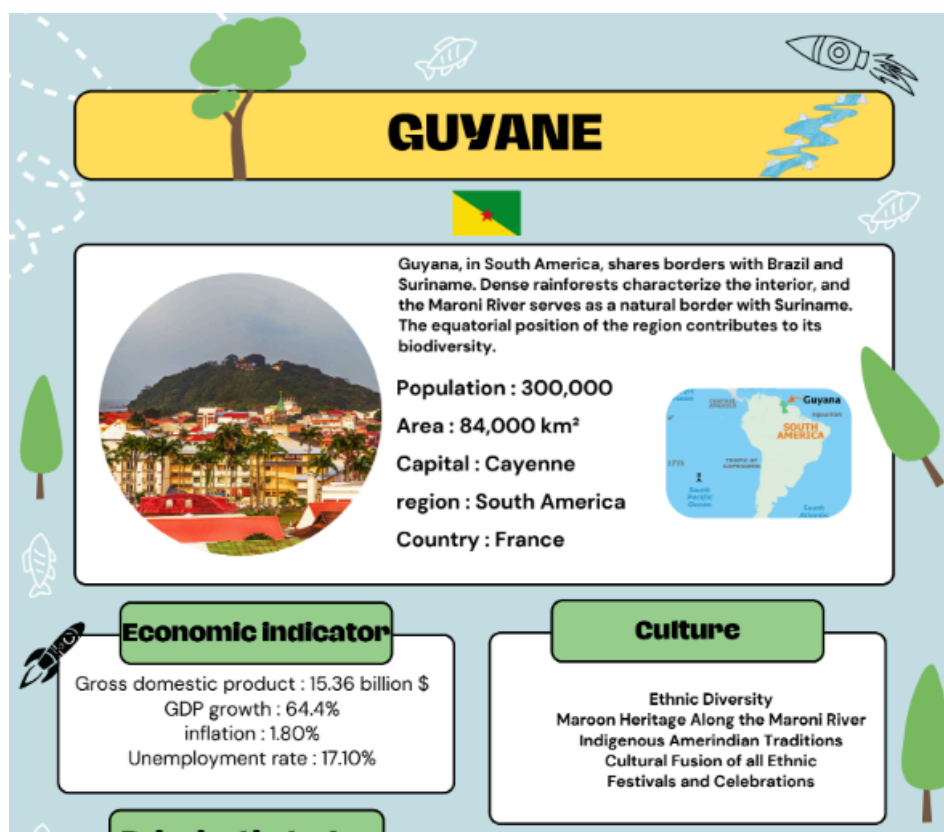
Présentation de la concurrence	
Orange :	Bouygues Telecom :
Chiffre d'affaire : 43,5 milliards	Chiffre d'affaire : 44,3 milliards
Part de marché : 35,5%	Part de marché : 19,1%
Nombre de pays où il opère : 26 pays	Nombre de pays où il opère : 1 (en tant que réseau de télécommunication. Sinon Bouygues est présent dans plus de 80 pays dans d'autre secteur d'activité
Année de création : 1988	Année de création : 1952
SFR :	Indirecte :
Chiffre d'affaire : 7,24 milliards	La Poste
Part de marché : 22,2%	Sosh
Année de création : 1987	

Cette SAE avait pour but de présenter une entreprise en analysant son micro et macro environnement, son secteur d'activité, et ses concurrents. Pour sa réalisation, elle nécessitait des prérequis, notamment des connaissances en économie, en analyse de marché, et en stratégie d'entreprise. Cette SAE était essentielle à mes yeux, car nous avons appris à comprendre l'environnement économique d'une entreprise.

Cette SAE m'a permis d'apprendre de nouvelles choses, telles que l'analyse PESTEL, et l'évaluation des parts de marché des concurrents. Le point positif le plus important est d'avoir pu mettre mes connaissances apprises en cours d'économie en application dans une situation réelle.

Je compte continuer sur cette lancée en approfondissant mes compétences, comme l'analyse SWOT, la modélisation des scénarios économiques, et les techniques avancées de veille concurrentielle. Pour cela, je compte m'investir encore davantage dans ma formation à l'IUT, tout en poursuivant des recherches personnelles sur des plateformes comme Google et YouTube.

E. SAE 105 : Présentation en anglais d'un territoire économique et culturel

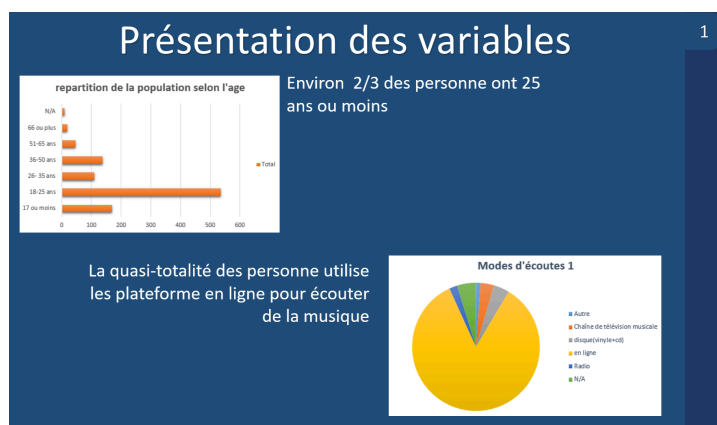


Cette SAE avait pour but la présentation en anglais d'un territoire économique et culturel. Pour sa réalisation, elle exigeait des prérequis, notamment des compétences en langue anglaise, en analyse économique, et en compréhension des aspects culturels. Cette SAE était capitale à mes yeux, car elle m'a permis de pratiquer mon anglais dans un contexte professionnel, en présentant des graphiques, des concepts, etc....

Bien que j'avais déjà quelques bases en anglais et en analyse économique, cette SAE m'a permis de consolider ces compétences et d'apprendre de nouvelles choses, telles que l'élaboration de présentations en anglais, l'analyse des atouts économiques d'une région, et la mise en valeur de son patrimoine culturel.

Je compte continuer sur cette lancée en améliorant mes compétences, comme la maîtrise des présentations orales en anglais, l'analyse approfondie des données économiques régionales, et les techniques de valorisation du patrimoine culturel. Pour cela, je compte m'investir encore davantage dans ma formation à l'IUT.

F. SAE 106 et 107 : Mise en œuvre d'une enquête



4. Tentative d'analyse quanti-quali

Justification : dans notre questionnaire, l'âge a été encodé comme une variable qualitative. Or elle pouvait être encodée comme une variable quantitative continue. Afin de mieux analyser notre hypothèse, nous avons recoder notre variable âge comme une variable quantitative continue.

Recodage :

variable age	classe	centre des classe
17 ou moins	[0; 17]	8,5
18-25 ans	[18;25]	21,5
26-35 ans	[26;35]	30,5
36-50 ans	[36;50]	43
51-65 ans	[51;65]	58
66 ou plus	[66;80]	73

Nb : les personnes n'ayant pas précisé leur âge ont été ignoré à cette étape. Ainsi, l'analyse a été effectuée sur une population de 907 personnes au lieu de 1026.

Analyse :

k-groupe	moyenne de classe	variance de classe	effectif de classe	moyenne global	
Autre	22,45	150,93	11		25,28
Chaîne de télévision musicale	36,73	379,46	30	variance intraclasse (W)	177,97
disque(vinyle+cd)	40,82	342,38	36	variance interclasse (B)	31,07
en ligne	23,67	160,47	765	variable global	209,04
N/A	22,48	190,80	47	rapport de corrélation (n²)	0,15
Radio	52,28	240,28	18		

Nous observons que le rapport de corrélation est égal à 0,15, celui-ci étant très proche de 0, il ne nous semble donc pas avoir de lien entre les 2 variables.

Ce projet s'articule autour de 3 points principaux. Premièrement, le traitement des données, deuxièmement, l'analyse des données et enfin la valorisation de ces données. L'objectif de cette SAE était de mettre en application toutes les ressources que nous avons vu au semestre 1 dans un seul projet, ainsi, pour le réaliser, nous devons avoir de nombreux prérequis comme des connaissances en statistique, en excel, anglais, datavisualisation etc.... C'est un projet que j'ai beaucoup aimé car il faisait intervenir tous ce que nous avons appris, et en même temps nous a poussé à faire des recherche externe pour apprendre de nouvelle chose(fonction sur excel, mise en place de questionnaire)

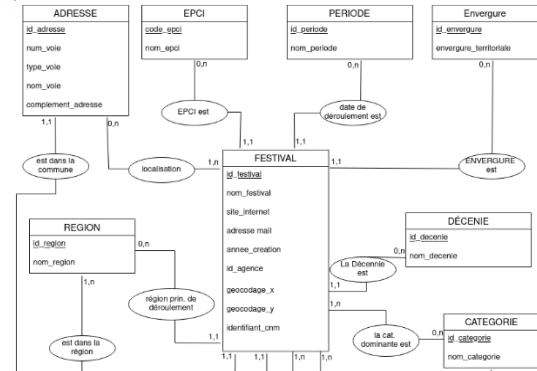
Pour mener ce projet a terme, nous avons dû d'abord, faire de nombreuse recherche afin de mieux comprendre la problématique a la quel nous devons faire face (lien entre caractère sociologique et écoute de musique), grâce à nos cours de communication, nous avons creer un questionnaire que nous avons administré à plus de 1000 personnes. Puis, après la récolte des données, nous sommes passées au traitement de ces données, nous avons utilisé plusieurs logiciels à savoir Sphinx et Excel. Cette étape nous a permis de rendre les données récoltées exploitables. Ensuite, pour l'analyse, nous avons d'abord fait une analyse exploratoire simple(moyenne, médiane, mode, variance, quartile) puis nous sommes intéressés au lien entre les variables (analyse de corrélation, test du Khi-2), ce qui nous a permis de vérifier nos hypothèse, pour effectuer ces calculé, nous avons utilisé microsoft Excel. Pour terminer cette étape, nous avons réalisé des visualisations pour nos données, histoire de préparer l'étape. Enfin pour cette étape, nous avons préparé une présentation en amphi, pour expliquer autre nos résultats.

Grâce à cette SAE j'ai fait beaucoup de progrès dans toutes les compétences (traiter, analyser et valoriser). J'ai aussi amélioré mes compétences en gestion d'équipe car j'étais le chef de mon équipe. Cependant je dois continuer à m'améliorer, surtout sur les présentations en amphi.

G.SAE 201 : Conception et implémentation d'une base de données

```
CREATE TABLE SOUS_CATEGORIE_ACTIVITE(  
  id_sous_categorie INTEGER REFERENCES  
  SOUS_CATEGORIE(id_sous_categorie) ON DELETE CASCADE,  
  id_activite INTEGER REFERENCES ACTIVITE(id_activite) ON DELETE  
  CASCADE,  
  PRIMARY KEY (id_sous_categorie, id_activite)  
);  
CREATE TABLE CATEGORIE_SOUS_CATEGORIE(  
  id_sous_categorie INTEGER REFERENCES  
  SOUS_CATEGORIE(id_sous_categorie) ON DELETE CASCADE,  
  id_categorie INTEGER REFERENCES CATEGORIE(id_categorie) ON DELETE  
  CASCADE,  
  PRIMARY KEY (id_categorie, id_sous_categorie)  
);
```

1. Modèle entités-associations



Cette SAE avait pour objectif, la création, normalisation et le peuplement d'une base de données à partir d'un fichier csv issue du site datagouv au sujet des festivals en France. Pour sa réalisation, cette SAE demandait trois principaux prérequis à savoir : la modélisation et normalisation de base de données, en langage SQL et en programmation, nous avons acquis ces compétences grâce au ressources et au sae du S1. Cette SAE à ouvert mon esprit sur les attentes que les employés ont envers leur employeur au niveau des différentes compétences mentionnées ci dessus, elle est très proche d'un travail qu'un data ingénieur doit accomplir en entreprise. j'ai pu découvrir grâce à ce projet, j'ai découvert que j'avais une bonne maîtrise du langage SQL, et de la programmation, également j'ai fait des progrès dans ma manière de gérer les projets et mon organisation, mais il me reste des progrès à faire. Pour cela je compte faire d'une part, des recherches personnelles grâce à des plateformes comme google et youtube, mais également de m'investir dans les cours à l'IUT.

H. SAE 202 : Estimation par sondage simple

2.5 Naît-il plus de filles ou de garçons ?

Servez-vous de ce théorème pour déterminer un intervalle de confiance de la probabilité qu'un nouveau né soit de sexe masculin. Faites varier le niveau de confiance: $\alpha = 0.95, 0.99, 0.999 \dots$ Quelles sont vos conclusions ?

```
data_sexe = data$SEXE == 1  
  
mean_data_sexe = mean(data_sexe)  
var_data_sex = var(data_sexe)  
  
val = c(0.95, 0.99, 0.999)  
  
for(i in val){  
  level = i  
  q = qnorm((1 + level) / 2)  
  born_inf = mean_data_sexe - q * sqrt(var_data_sex / length((data_sexe)))  
  born_sup = mean_data_sexe + q * sqrt(var_data_sex / length((data_sexe)))  
  
  b = c(born_inf, born_sup)  
  print(paste("intervall de confiance de niveau ", i, " : [", b[1], ", ", b[2], "]"))  
}  
  
## [1] "intervall de confiance de niveau 0.95 : [ 0.509949928573241 , 0.512207460165289 ]"  
## [1] "intervall de confiance de niveau 0.99 : [ 0.509595244654282 , 0.512562144084248 ]"  
## [1] "intervall de confiance de niveau 0.999 : [ 0.509183642187333 , 0.512973746551197 ]"
```

D'après les intervalles de confiances que nous avons eu nous pouvons en conclure qu'il y a une légère différence entre la proportion d'hommes et de femmes en faveur des hommes. Environ 51% d'hommes pour logement, environ 49% de femmes. Cela rejoint la problématique de la question qui émettait que les bébés humains sont plus souvent masculin, en revanche nous n'avons aucune preuve qu'il s'agisse d'un facteur biologique il est beaucoup plus probable qu'il s'agisse d'un facteur socio-économique.

Cette SAE avait pour objectif de nous faire comprendre comment était fait les intervalles de confiance, pour cela, nous avons eu besoin d'avoir de solide base en statistique inférentielle, statistique descriptive, en probabilité et en programmation R. Ainsi j'ai pu appliquer tout ce que j'avais vu en statistique inférentielle sur les intervalles de confiance, différente loi étudiée (Student, normal, etc...) et le TCL (théorème centrale limite) dans un contexte réel. Je constate que je maîtrise bien ces sujets, cependant il faut que je continue à me former en statistique inférentielle, car c'est ce que l'on utilise le plus en entreprise (pour faire des estimations), pour cela, je ferai des recherches approfondies sur ce sujet et ferai également des exercices afin de m'améliorer.

I. SAE 203 : Régression sur données réelles

```
reg = summary(lm(log(area_mean) ~ perimeter_mean, data))
a_e = reg$coefficients['perimeter_mean', 'Estimate']
b_e = reg$coefficients['(Intercept)', 'Estimate']
r2_e = reg$r.squared

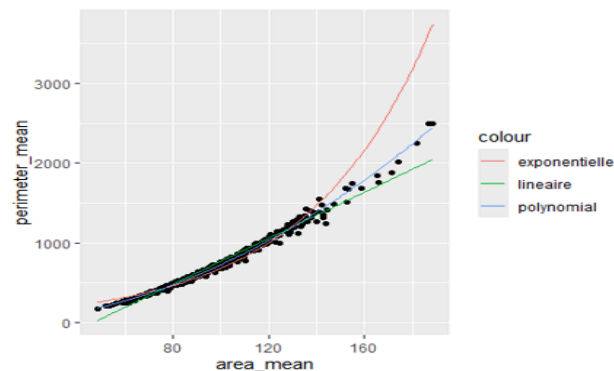
func_e = function(x){
  exp(b_e) * exp(a_e*x)
}

reg = summary(lm(log(area_mean) ~ log(perimeter_mean), data))
a_p = reg$coefficients['log(perimeter_mean)', 'Estimate']
b_p = reg$coefficients['(Intercept)', 'Estimate']
r2_p = reg$r.squared

func_p = function(x){
  exp(b_p) * x^a_p
}

fig = ggplot(data, aes(x = perimeter_mean, y = area_mean)) +
  geom_point() +
  labs(x = "area_mean", y = "perimeter_mean") +
  geom_function(fun = func_l, aes(colour = 'lineaire')) +
  geom_function(fun = func_e, aes(colour = 'exponentielle')) +
  geom_function(fun = func_p, aes(colour = 'polynomial'))

print(fig)
```

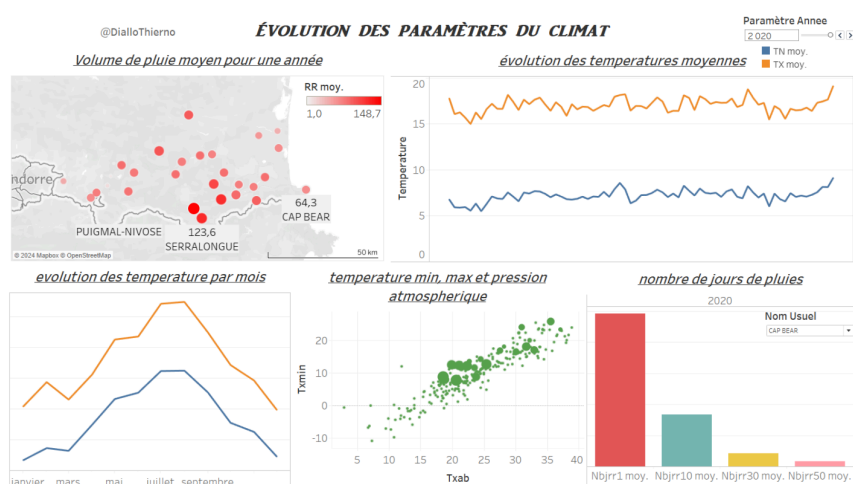


Cette SAE avait pour objectif de déterminer si une tumeur est maligne ou bénigne en cherchant des variables discriminantes dans le jeu de données puis de proposer un modèle. Pour sa réalisation, cette SAE demandait trois principaux prérequis à savoir :

- régression linéaire et exponentielle
- statistique descriptive
- en programmation,

Nous avons acquis ces compétences grâce aux ressources et au sae du S1. Cette SAE a ouvert mon esprit sur les attentes que les employés ont envers leur employeur au niveau des différentes compétences mentionnées ci dessus. Elle est très proche d'un travail qu'un data scientifique doit accomplir en entreprise, c'est-à-dire résoudre des problèmes de classification, mais également de régression aussi. j'ai pu découvrir grâce à ce projet, que j'avais une bonne maîtrise du langage R, et mes connaissances en statistique était aussi solide, mais il me reste des progrès à faire. Pour cela je compte faire d'une part, des recherches personnelles grâce à des plateformes comme google et youtube(machine-learnia, StatQuest With Josh Starmer), mais également de m'investir dans les cours à l'IUT.

J. SAE 204 : DATAVIZ



Dans l'optique de la SAE 204 et du concours Challenge SD 2024 (dataviz), nous avons été amenés à réaliser un dashboard avec le logiciel Tableau Desktop. Afin de mener ce projet à terme, nous avons dû faire appel à de nombreuses compétences techniques telles que la compréhension des données(excel, programmation) et la réalisation du rendu (tableau desktop).

Nous avons dû choisir en premier lieux une problématique et un département sur lequel nous allions porter nos analyses, cela c'est fait en mettant en commun nos connaissances et envies afin de décider. Après avoir choisi notre problématique, nous avons synthétisé l'information contenu dans le dataset en faisant des calculs de moyenne grâce à nos connaissances en statistiques. Enfin avec Tableau Desktop, nous avons réalisé 5 graphiques. Pour ma part, j'ai réalisé une carte sur laquelle nous pouvons visualiser la quantité de pluie moyenne par année, la taille de chaque point était proportionnelle à la quantité de pluie, et lorsque nous survolons un point avec la souris, vous pouvez voir les différents détails sur la région en question, cela est possible grâce aux repères de Tableau Desktop, et nous pouvons choisir l'année avec un paramètre. Enfin avec nos cours de communication et tout

ce que nous avons appris sur la présentation orale, nous avons présenté notre Dashboard à l'autre. Ce projet nous a permis de mettre en application toutes ces différentes connaissances, ce qui nous a permis de les renforcer. Cependant j'ai un point à perfectionner, c'est datavisualisation, c'est-à-dire le fait de réaliser des graphiques facilement compréhensibles par tous. Également, lors du projet de dataviz, j'ai fait usage de nombreux soft skills afin de mener à terme le projet. j'étais responsable de la coordination de l'équipe, pour réussir cela, j'ai dû user de mon sens du collectif et de mon écoute afin de comprendre les besoins de chaque membre de l'équipe, ce qui a renforcé leur confiance en moi. Ma confiance en moi et ma capacité de gestion du stress me permettent de travailler dans un environnement sous pression tout en gérant efficacement mon temps, ce qui m'évite de perdre mes moyens. mes compagnons également assigner le rôle de décideur car il juge que j'ai un bon esprit critique le sens des responsabilités. Ainsi toutes ces capacités de leader et de pensée stratégique, j'ai réussi à guider mon équipe au terme de ce projet. Cette expérience a renforcé ma capacité à diriger efficacement, à collaborer et à garantir des résultats de haute qualité dans des conditions exigeantes.

K. SAE 205 : Construction et présentation d'indicateurs de performance

=	résultat exceptionnel	-879	-303	190,10%
	résultat courant avant impôt	132214	90683	45,80%
+	résultat exceptionnel	-879	-303	190,10%
-	participation des salariés	0	0	0,00%
-	impôt sur les bénéfices	33712	22036	52,99%
=	résultat de l'exercice (bénéfice ou perte)	97623	68344	42,84%

2. interprétation des indicateurs d'activités

Au cours de ces deux dernières années, l'entreprise SOGEMAT a été rentable au niveau des ventes et des achats de marchandises, on le remarque grâce à la marge commerciale qui était de 630 838€ en 2023 et de 603 674€ en 2022. Donc une évolution de 4,5 %.

SOGEMAT a également réalisé une marge de 25 896€ sur les activités de transformation en 2023 (évolution de 4,19 % par rapport à 2022), ce qui représente une performance industrielle assez faible. SOGEMAT a pour particularité de ne pas avoir de stock de production, ce qui leur permet de réduire leur dépense. De plus, en ayant une marge commerciale(650k), très supérieure à la production d'exercice(25k), on peut en déduire que l'entreprise priorise la vente de produit plutôt que leur fabrication.

Cette SAE avait pour but de présenter une entreprise en analysant son micro et macro environnement, son secteur d'activité, et des indicateurs de performance. Pour sa réalisation, elle nécessitait des prérequis, notamment des connaissances en économie, en analyse de marché, et en stratégie d'entreprise. Cette SAE était essentielle à mes yeux, car nous avons appris à comprendre l'environnement économique d'une entreprise mais surtout de juger de l'état financier d'une entreprise grâce à différents indicateurs.

Cette SAE m'a permis d'apprendre de nouvelles choses, telles que le calcul et l'interprétation des indicateurs financiers(EBE, RCAI, etc...). Le point positif le plus important est d'avoir pu mettre mes connaissances apprises en cours d'économie en application dans une situation réelle.

Je compte continuer sur cette lancée en approfondissant mes compétences, en apprenant les techniques avancées d'analyse financière. Pour cela, je compte m'investir encore davantage dans ma formation à l'IUT, tout en poursuivant des recherches personnelles sur des plateformes comme Google et YouTube.

L. SAE 206 : Analyse de données, reporting et datavisualisation

lien vers la présentation : <https://www.youtube.com/watch?v=3BGD3hc3c-s>

Dans ce projet, nous devons mener une étude de marché pour un client en nous appuyant sur des données issues de l'insee. Ce projet faisait appel à la majorité des compétences que nous avions pendant notre formation a savoir :

- traiter : programmation, excel
- analyse : statistique descriptive, dataviz
- valoriser : communication

Pour réaliser ce projet, j'ai dû apprendre à utiliser le module python pandas afin de nettoyer les données(gestion des cas où les cellules étaient vides, mauvais encodage, etc...) et je me suis rendu compte de l'importance de ce module dans ce secteur. j'ai aussi appris comment visualiser les données avec plotly, qui est également très important.

Ce projet m'a permis d'améliorer mes compétences en traitement de données, mais également en analyse, nous avons généré une quinzaine de graphiques que nous avons dû interpréter et comprendre la signification. puis nous avons créé un powerpoint afin de présenter nos résultats. En somme, ce projet a fait intervenir toutes les connaissances que nous avons engrangées au cours de cette année universitaire. Etant donné que j'étais le responsable du groupe, j'ai dû gérer une équipe, cela a amélioré mes compétences de leader et ma manière de m'organiser. Cependant il me reste un point à améliorer, la valorisation des données(powerpoint facilement compréhensible et pas trop chargé). Ensuite, j'ai beaucoup aimé ce projet car j'ai pu une fois de plus mettre mes connaissances en application dans un cas concret.

5. Conseil d'un professionnel

Interlocuteur : Alpha Diallo, Lead Analytics, chez Hitachi Solutions Europe (France)

Après plusieurs discussions avec lui au sujet du monde la data, il m'a expliqué comment se déroulait les mission dans son domaine ou les différente situation auquel il a été confronté. Il m'a également donné de nombreux conseil tel que :

- maîtriser la programmation (surtout python) : selon lui avoir une maîtrise de la programmation est essentiel pour pouvoir évoluer dans ce domaine, les quantité de données traiter sont de plus en plus grande et les technique de traitement de ses données évolue sans cesse, il faut donc avoir une maîtrise de cette discipline pour ne pas être largué en cours de route et évoluer.
- avoir de solide base en mathématique : pour lui il est essentiel d'avoir de solide base en mathématique pour comprendre la logique et les opération qui se cache derrière les modèle d'IA
- mettre en place une veille technologique : comme dit précédemment, les technologies sont en constante évolution, pour les professionnels du secteur, il est primordial de s'informer sur ces évolutions afin de s'adapter et d'être plus efficace dans son travail.
- Avoir de bonne qualité de vulgarisation : dans notre discussion, il m'a raconté une anecdote dans laquelle un client a refusé un modèle plus performant au détriment d'un modèle moins performant, mais dont il comprenait le fonctionnement. Pour lui, on doit avoir cette capacité de d'expliquer ce que l'on fait(modèle, outil) au autre qui ne sont pas des professionnels du domaine de tel sorte qu'il comprennent et soit plus ouvert à ces technologies.

Il m'a donné aussi des conseils comme il faut savoir travailler en équipe, être curieux et aimer ce que tu fais.

6. Objectifs

- objectif à court terme : mon objectif pour ces prochaine année est de terminer ma formation en BUT SD. Pour cela je compte m'investir encore plus dans la formation afin d'acquérir le maximum de connaissance dans le domaine de la data. Je compte également m'auto former sur des sujets qui ne sont pas dispensés à l'iut grâce à des plateformes comme google ou youtube.
- objectif à moyen terme : après le BUT, je compte intégrer une école d'ingénieurs pour suivre une formation en math appliqué ou IA (en alternance), ce qui va me permettre de continuer ma formation mais aussi d'entrer dans le monde professionnel afin d'acquérir de l'expérience. En alternance, je compte commencer en tant qu' assistant data scientist ou tout poste proche de celui la. Après cela, je voudrais travailler en tant que data scientist
- objectif à long terme : dans 10-15 ans, je me vois toujours dans le secteur de la data, je me vois travailler dans une entreprise technologique innovante, ou même créer ma propre startup, où je pourrais développer des solutions basées sur l'IA et le big data pour améliorer la prise de décision.