

# Cahier des charges

réalisé par Virginie Cicéron, Ewen Laxalde et Quentin Guichoux

# **1 Introduction**

## **1.1 Motivations**

### **1.1.1 Le client**

Le Central Statistical Office of Breizhmeiz, centre d'analyse statistique en tout genre

### **1.1.2 Le problème**

De nouvelles données sur les paroles des chansons du rappeur américain EMINEM viennent d'arriver, une première analyse est requise avant toute future utilisation de ces données

### **1.1.3 L'existant**

Un dataset regroupant les paroles des chansons d'EMINEM, le titre de la chanson, le nom de l'album dans lequel elle se trouve, son nombre de vue sur le site d'analyse de parole genius et sa date de sortie.

### **1.1.4 Le besoin non satisfait**

Ce sont des données brutes, aucune analyse statistique n'a été menée dessus

### **1.1.5 Les objectifs**

- Réaliser le MCD de la base correspondante
- Mettre en place la base de données
- Charger le dataset dans la base à l'aide d'un script Python
- Créer une vue et une procédure stockée
- Rédiger un Notebook contenant le MCD, une description du travail de création des bases et le code SQL de la vue et de la procédure stockée
- Analyser les données en Python sur un Notebook
- Visualiser les données via les bibliothèques Python adaptées sur le même Notebook

## **1.2 Contexte**

### **1.2.1 Précisions sur le client**

Le travail à effectuer est une première analyse des données afin de fournir une première expertise.

### **1.2.2 Marché**

L'environnement économique et les enjeux commerciaux.

# **2 Documentation**

## **2.1 Terminologie métier**

- Dataset : jeu de données
- notebook : format de fichier permettant une présentation de code sous forme de cellule en ayant la possibilité d'y insérer des commentaires
- MCD : Modèle Conceptuel de Données, représentation graphique de l'ensemble des données sous forme de schémas afin de permettre une meilleure compréhension de celles-ci
- Python : langage utilisé pour écrire les différentes parties du projet
- SQL : langage utilisé pour la création de la base de données regroupant les données du dataset pour faciliter leur exploitation

## **2.2 Bibliographie succincte**

## **3 Profil des utilisateurs finaux**

### **3.1 En utilisation**

Le système est à destination de musicologues, non formé à l'utilisation des outils utilisés pour son fonctionnement. Une formation sera assurée par nos équipes.

### **3.2 En exploitation**

Le système est à destination de musicologues, non formé à l'utilisation des outils utilisés pour son fonctionnement. Une formation sera assurée par nos équipes.

### **3.3 En maintenance**

La maintenance du systèmes sera assurée par nos équipe, elle sera facturée.

### **3.4 Environnement d'utilisation**

L'entreprise utilise des ordinateurs fonctionnant sous Microsoft Windows 10

## **4 Fonctions à réaliser**

## 4.1 Ce que le système doit faire

- Priorité Fonction Description Critère de performance
- 1 Calculs Calculs les moyennes de toutes les colonnes du dataset.
- Une moyenne par colonne calculée en moins de 30 min sur un ordi ayant la configuration minimum requise
- 2 Graphique Rendu des moyennes générés par le code à partir du dataset sur un graph lisible.
- Les données du graph seront affichées dans une police 12, et les moyennes seront lisibles sur les abscisses et les valeurs sur les ordonnées

## 4.2 Ce que le système ne doit pas faire

Le système ne permet pas l'ajout de données ou la modification des données existantes.

## 5 Contrainte du système

### 5.1 Contraintes matérielles

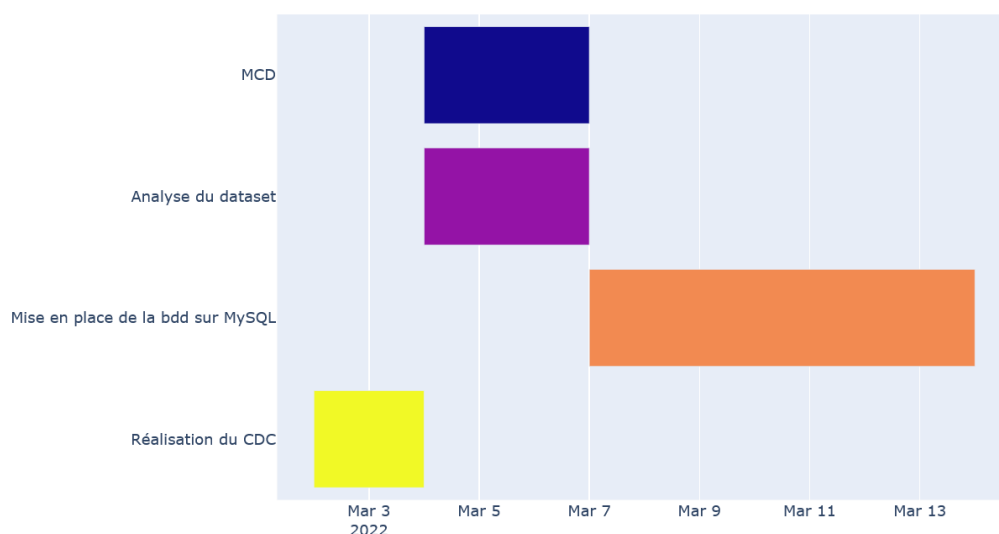
Le client utilise des ordinateurs DELL milieu de gamme.

### 5.2 Contraintes logicielles

Les machines du client fonctionnent sous Windows 10. Le système doit être développé en Python et SQL.

### 5.3 Contraintes fonctionnelles

Le système sera développé par une équipe de 3 personnes, sur 3 machines différentes.



## **5.4 Contraintes d'ergonomie**

Le système sera exploitable via MariaDB et PHPmyadmin. Une formation à l'utilisation de ces logiciels sera proposée au client.

## **5.5 Autres facteurs de qualité exigés par le client**

Fiabilité, intégration, portabilité, maintenabilité...Projet CDC

# **6 Aspects contractuels**

## **6.1 Délais et protocole de livraison**

Le système doit être livré lundi 07/03/2022

Livrable :

5 fichiers seront livrés au client :

- Un notebook de présentation de la base (qui doit contenir le MCD du dataset)
- le fichier SQL permettant de recréer la base
- un fichier python contenant le code de l'exploitation de la base
- Un notebook d'analyse statistique
- Le CDC en pdf.

## **6.2 Estimation des coûts**

Le coût total du système est évalué à 4800€

## **6.3 Aspects juridiques et Commerciaux**

Le système est développé à destination du client uniquement et ne doit pas être diffusé. La diffusion ou la réplication du code n'est pas autorisée.