

Réalisation d’une application web de création de recette et d’analyse des prix du marché

**RAPPORT DE PROJET**

# Table des matières

[**Table des matières** 2](#_Toc26968509)

[INTRODUCTION 4](#_Toc26968510)

[2. PRESENTATION DE L’ASSOCIATION 5](#_Toc26968511)

[2.1. Besoins client 5](#_Toc26968512)

[2.2. Les objectifs 5](#_Toc26968513)

[2.3. Les outils utilisés 5](#_Toc26968514)

[3. Organisation et planification du projet 7](#_Toc26968515)

[3.1. Diagramme de GANTT prévisionnel 7](#_Toc26968516)

[3.2. Diagramme de GANTT Réel 8](#_Toc26968517)

[4. Diagramme de conception 9](#_Toc26968518)

[4.1. Modèle Conceptuel de Données 9](#_Toc26968519)

[5. Base de données 10](#_Toc26968520)

[5.1. Dictionnaire de données 10](#_Toc26968521)

[5.2. Relation entre les entitées 11](#_Toc26968522)

[5.3. Processus de récolte (webscraping) 12](#_Toc26968523)

[5.4. Import dans la base de données 13](#_Toc26968524)

[6. Application web 15](#_Toc26968525)

[6.1. Section analyse du marché 15](#_Toc26968526)

[6.1.1. Support de visualisation 15](#_Toc26968527)

[6.1.2. Outil de diffusion 15](#_Toc26968528)

[6.2. Section création de recette 16](#_Toc26968529)

[6.2.1. Système CRUD 16](#_Toc26968530)

[6.2.2. Rapport d'analyse de la valeur nutritionnelle 16](#_Toc26968531)

[7. Bilan technique 19](#_Toc26968532)

[8. Conclusion 20](#_Toc26968533)

REMERCIEMENTS

Remerciements aux formateurs M. Eric Faure et M. Khalil Laghmari qui nous ont efficacement enseigné le métier de développeur data et supporté durant tout le long de la formation. Remerciements également à Mme. Emilie Diallo, chargée de médiation et d’insertion au sein de DIGIFAB, et M. Diogo Barry, directeur du centre de formation.

Je tiens aussi à remercier les apprenants de la formation qui ont été d’une grande aide.

# INTRODUCTION

Le projet chef-d’œuvre s’effectue dans le cadre de la formation « Développeur Data » de Simplon afin de mettre en application toutes les compétences acquises lors de la formation.

L’objectif de ce projet est de valider tous les éléments de la certification vu au cours de cette formation : développement d’une base de données, exploitation des données en passant par la gestion de projet. Le projet peut correspondre à une réelle demande ou être fictif.

Il s’agit ici d’une demande fictive d’une association ayant besoins d’une interface pour effectuer des recettes de soupe. Les détails de l’association et de sa demande sera expliqué plus en détail dans le rapport.

Dans un premier temps le fonctionnement de l’association sera présenté. Ensuite, de manière synthétique, la gestion des données en présentant leur structure et leurs importations. Et, pour finir, chaque partie de l’application sera développée pour terminer sur une conclusion.

# 2. PRESENTATION DE L’ASSOCIATION

L’association « Soupe Chaude » est une association préparant des soupes pour d’autres associations assurant des missions sociales tels que « Les Restos du Cœur » ou « Secours Catholique ». Les ingrédients principaux pour effectuer les recettes de soupes sont les légumes. L’association se fourni principalement à Leader Price ou commande directement sur internet sur le site « rungischezvous.com ».

## 2.1. Besoins client

L’association souhaite avoir une interface qui permet de savoir où acheter les légumes pour avoir des soupes au prix le plus bas. L’association souhaite aussi pouvoir composer des recettes de soupe directement depuis l’interface et gérer les ventes des soupes aux autres associations. Deux niveaux de droit être établis :

-Les membres : ils peuvent visualiser le prix des ingrédients et gérer les recettes.

-L’administrateur : en plus de tout pouvoir visualiser, il peut également gérer les commandes les recettes.

## 2.2. Les objectifs

L’objectif de cette application est donc de permettre à l’association de prendre une décision plus rapide sur le choix d’achat des ingrédients grâce à l’interface de comparaison des prix. L’application permettra aussi d’obtenir un historique des dépenses avec une section mis à disposition pour les commandes.

## 2.3. Les outils utilisés

-Python :

Python est un langage de programmation interprété. Ce langage s’est révélé très efficace pour la récolte des données ainsi que pour le nettoyage et l’importation des données à la base de données. Cela a été possible grâce à la librairie BeautifulSoup4 qui permet d’extraire les éléments d’une page HTML. La librairie Pandas a aussi été très utile pour le nettoyage.

-Planificateur des tâches Windows :

Le planificateur des tâches est une fonctionnalité présente par défaut avec Windows. Elle permet de programmer des exécutions de script, notamment des scripts de mise à jour de base de données (pour les prix des ingrédients).

-HTML/CSS :

Le HTML (HyperText Mark-Up Langage) est un langage dit de balise permettant la conception de page web avec des balises de formatage afin de structurer le contenu de la page.

Le CSS (Cascading Style Sheets) qui vient compléter le HTML. Quand le HTML va structurer la page, le CSS va, quant à lui, mettre en forme en apportant du style aux éléments de la page.

Pour faciliter la création de l’interface, le Framework CSS Bootstrap a été employé. Pour ce faire, celui-ci propose une collection de CSS pré-générés et prêts à l’utilisation.

-Javascript :

Le JavaScript est un langage de script incorporé dans un document HTML. Ce langage permet d'apporter des améliorations au langage HTML en exécutant des commandes du côté client. Le JavaScript a essentiellement permis de dynamiser les pages (Graphiques, Diagrammes, …).

Ici aussi, des Frameworks ont été utilisés : Highcharts et CanvasJS.

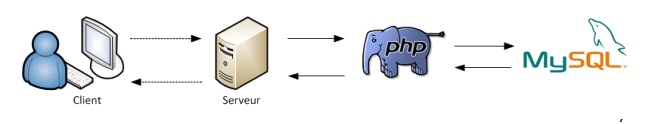
-SQLBackupAndFTP :

C’est un logiciel permettant d’automatiser les backups de la base de données.

-PHP/MySQL :

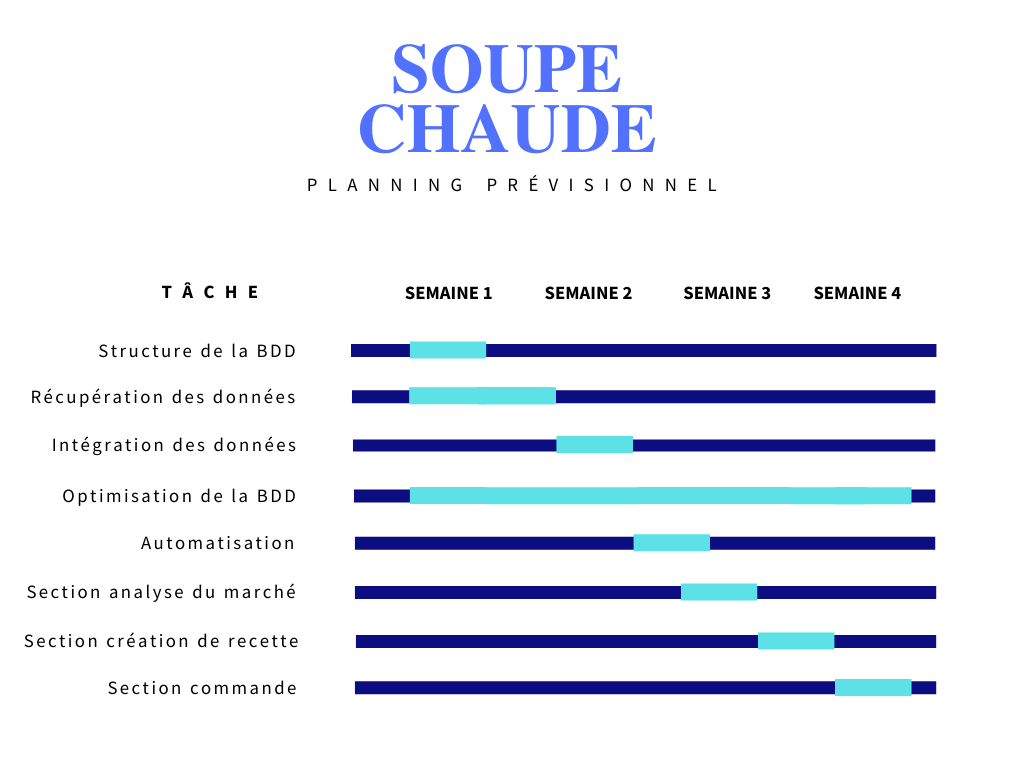
Le PHP est un langage de script exécuté du côté serveur. C’est un langage simple pour gérer la liaison entre le HTML et une base de données.

Le S.G.B.D. (Système de gestion de base de données) qui a été choisi est MySQL. Ce choix s’est fait simplement ; il s’agit du S.G.B.D gratuit le plus utilisé, disponible sur de nombreuses plateformes. La documentation à propos de ce S.G.B.D est très riche sur internet et permet donc de corriger très rapidement les erreurs.



# 3. Organisation et planification du projet

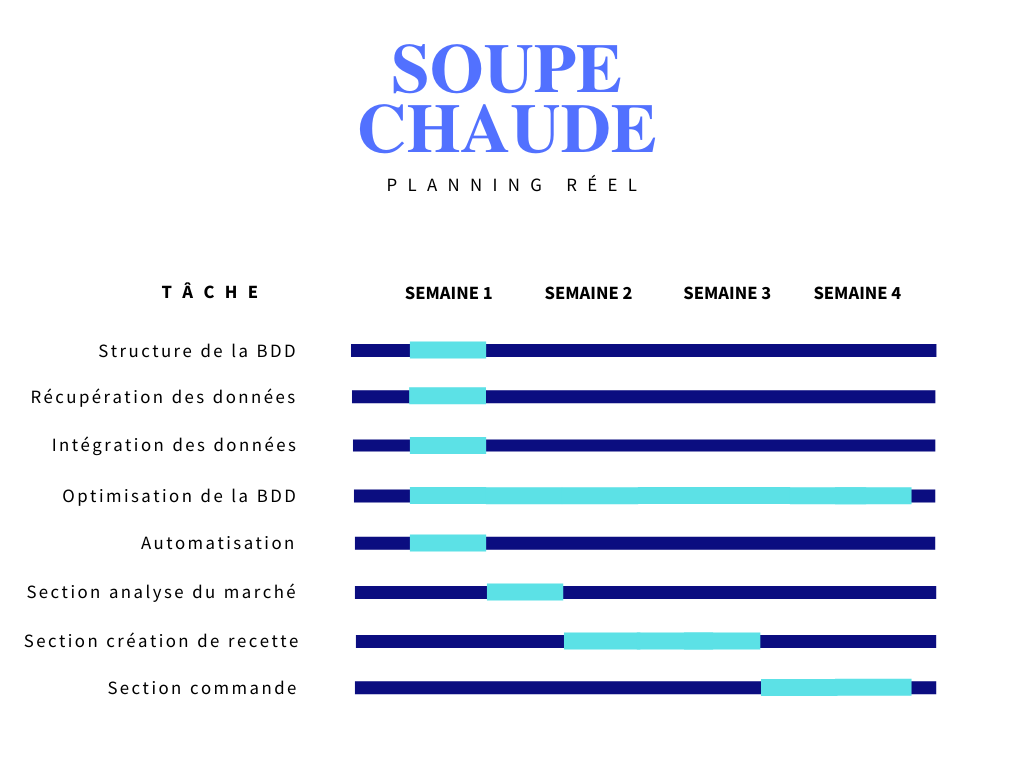
## 3.1. Diagramme de GANTT prévisionnel



Les deux premières semaines ont été dédiées au développement de la base de données ainsi qu’à l’automatisation de sa sauvegarde et de son alimentation. Ensuite, les deux dernières semaines ont été dédiées

à la visualisation.

## 3.2. Diagramme de GANTT Réel



Comme on le voit sur le diagramme réel, la partie dédiée au développement de la base à données à été effectuée en moins d’une semaine. Les librairies python ont été facile à comprendre et à assimiler et il y a énormément de documentations disponibles sur ce langage, ce qui a permis de terminer très rapidement cette partie. En revanche, la partie dédiée à la visualisation s’est révélée beaucoup plus longue que prévu : l’utilisation de Javascript a été un vrai challenge (notamment pour le système de CRUD pour les recettes) car il n’a pas été anticipé que le système de CRUD (Creation Read Update Delete) aurait besoin de ce langage pour fonctionner.

Étant la première fois que j’utilise ce langage, cela explique le retard de la deuxième partie.

# 4. Diagramme de conception

## 4.1. Modèle Conceptuel de Données



# 5. Base de données

## 5.1. Dictionnaire de données

|  |  |  |  |
| --- | --- | --- | --- |
| Nom | Code | Type de données | Longueur |
| date | DATE | Date |  |
| glucide | GLUCIDE | Réel |  |
| id\_commande | ID\_COMMANDE | Séquentiel |  |
| id\_ingredient | ID\_INGREDIENT | Séquentiel |  |
| id\_recette | ID\_RECETTE | Séquentiel |  |
| id\_unite | ID\_UNITE | Séquentiel |  |
| id\_users | ID\_USERS | Séquentiel |  |
| lipide | LIPIDE | Réel |  |
| login | LOGIN | Caractère variable (100) | 100 |
| nom\_ingredient | NOM\_INGREDIENT | Caractère variable (50) | 50 |
| nom\_recette | NOM\_RECETTE | Caractère variable (100) | 100 |
| password | PASSWORD | Caractère variable (255) | 255 |
| prix\_leader | PRIX\_LEADER | Réel |  |
| prix\_rungis | PRIX\_RUNGIS | Réel |  |
| prixr\_leader | PRIXR\_LEADER | Réel |  |
| prixr\_rungis | PRIXR\_RUNGIS | Réel |  |
| proteine | PROTEINE | Réel |  |
| quantite | QUANTITE | Entier |  |
| unite\_masse | UNITE\_MASSE | Caractère variable (10) | 10 |
| unite\_monnaie | UNITE\_MONNAIE | Caractère variable (10) | 10 |
| valeurcal | VALEURCAL | Réel |  |

## 5.2. Relation entre les entitées

|  |  |  |  |
| --- | --- | --- | --- |
| Association | Entité | Identifiant | Cardinalité du rôle |
| genere | Commande |  | 1,1 |
| compose | Ingredient |  | 0,n |
| compose | Recette |  | 1,n |
| genere | Recette |  | 1,n |
| compose | Unite |  | 1,1 |

Une recette peut avoir plusieurs ingrédients et un ingrédient peut être présent dans une ou plusieurs recette, on a donc une table intermédiaire nommé "compose" où on retrouve les id des tables "Ingrédient" et "Recette". Lorsqu’on compose une recette, un ingrédient ne peut avoir qu’une et une seule unité.

Une commande est générée à partir d’une recette et une recette peut se trouver dans plusieurs commande.

La table "users" n’a de relation avec aucune entité. Cette table ne sert qu’à stocker les paramètres de connexion des utilisateurs

La table compose est un élément important du système de création de recette : c’est dans celle-ci qu’on récupère l’id de la recette créée avec les id des ingrédients qui lui sont associés ainsi que l’id de l’unité qui est associé à l’ingredient.

## 5.3. Processus de récolte (webscraping)

Le processus de récolte des données comporte 3 étapes : l’extraction du code html de la page ciblée, le nettoyage des données extraites, la suppression des doublons et stockage des données dans un fichier « propre ». C’est le fichier ‘’scrapbs4.py’’ qui s’occupe du scraping, codé en python.

-Extraction du code html :

L’extraction s’effectue seulement en quelques lignes après avoir définie l’url cible :

    response = requests.get(url)

    content = BeautifulSoup(response.content, 'lxml')

    # sauvegarde des données à l'état brute

    df = pd.DataFrame(data=content)

    df.to\_csv(r'LeaderSale1.txt', mode='a', index=False)

La librairie Pandas permet de stocker les donées dans une DataFrame et de ainsi pouvoir les stocker dans un fichier. Ici, les données sont stockées à l’état brute dans un fichier texte (.txt).

-Nettoyage des données :

    ###

    #Nettoyage

    ###

    # selection de la liste pour un ingredient

    prix = content.find\_all("div", attrs={"class": "product-price\_\_number"})

    df = pd.DataFrame(data=prix)

    # sauvegarde du deuxième état brute

    df.to\_csv(r'LeaderSale2.txt', mode='a', index=False)

    # traitement des prix de la liste

    # s'il y a plus d'un élément sur la liste

    if len(prix) > 1:

        # traitement du format du prix

        prix1 = prix[0]

        prixint = prix1.find("span", attrs={"class" : "product-price\_\_integer"}).text

        prixfloat = prix1.find("span", attrs={"class" : "product-price\_\_decimal"}).text

        prixfinal = float(prixint+'.'+prixfloat)

        prix2 = prix[1]

        prixint = prix2.find("span", attrs={"class" : "product-price\_\_integer"}).text

        prixfloat = prix2.find("span", attrs={"class" : "product-price\_\_decimal"}).text

        prixfinal2 = float(prixint+'.'+prixfloat)

        # sauvegarde du troisième état brute

        df = pd.DataFrame(data=[prix1, prix2])

        df.to\_csv(r'LeaderSale3.txt', mode='a', index=False)

Dans le code ci-dessus on récupère l’élément du code HTML qui nous intérèsse : le prix. Une fois le prix isolé du code, il est converti en type « flottant » car il est de type « caractère » par défaut.

-Suppression des doublons et stockage des données dans un fichier « propre » :

        #Suppression des doublons

        ###

        if prixfinal < prixfinal2 :

            d = {'nom\_ingredient' : legume ,'prix\_leader' : [prixfinal]}

            df = pd.DataFrame(data=d)

            # sauvegarde des données propres dans un csv

            df.to\_csv('Leader.csv', mode='a', header=False, index=False)

            print(legume+'Leader'+' done')

        else :

            d = {'nom\_ingredient' : legume ,'prix\_leader' : [prixfinal2]}

            df = pd.DataFrame(data=d)

            df.to\_csv('Leader.csv', mode='a', header=False, index=False)

            print(legume+'Leader'+' done')

Le critère de suppression du doublon est le prix : le prix le moins cher est enregistré dans le fichier. C’est la conversion du prix en flottant, dans l’étape précédente, qui a permis que le prix puisse être comparé.

Une fois la comparaison faites, le prix est enregistré, ainsi que le nom du legume, dans un fichier CSV.

## 5.4. Import dans la base de données

L’import dans la base de données s’effectue en python avec le fichier ‘’importbdd.py’’ avec la librairie mysql.connector.

Il s’agit dans un premier temps d’appeler la fonction de scraping présent dans le fichier ‘’scrapbs4.py’’ avec comme argument le nom de l’ingrédient voulu :

def importtobdd():

    print('Récupération des données \n')

    scrap('carotte')

    scrap('courgette')

    scrap('ail')

Dans un second temps, les données récoltées sont importées à la base de données avec le code suivant :

 print('Importation des données dans la base\n')

    for row in csv\_data:

        print('row', row)

        sql = 'INSERT INTO ingredient(nom\_ingredient, prix\_leader) VALUES("{}", "{}")'.format(row[0], row[1])

        print(sql)

        mycursor.execute(sql)

    for row in csv\_data2:

        print('row', row)

        sql = "UPDATE ingredient SET prix\_rungis = %s WHERE nom\_ingredient = %s"

        val = (row[1], row[0])

        print(sql)

        mycursor.execute(sql, val)

Dans la deuxième requête on effectue un ‘’update’’ car la colonne prix\_rungis et prix\_leader se situe dans la même table.

# 6. Application web

## 6.1. Section analyse du marché

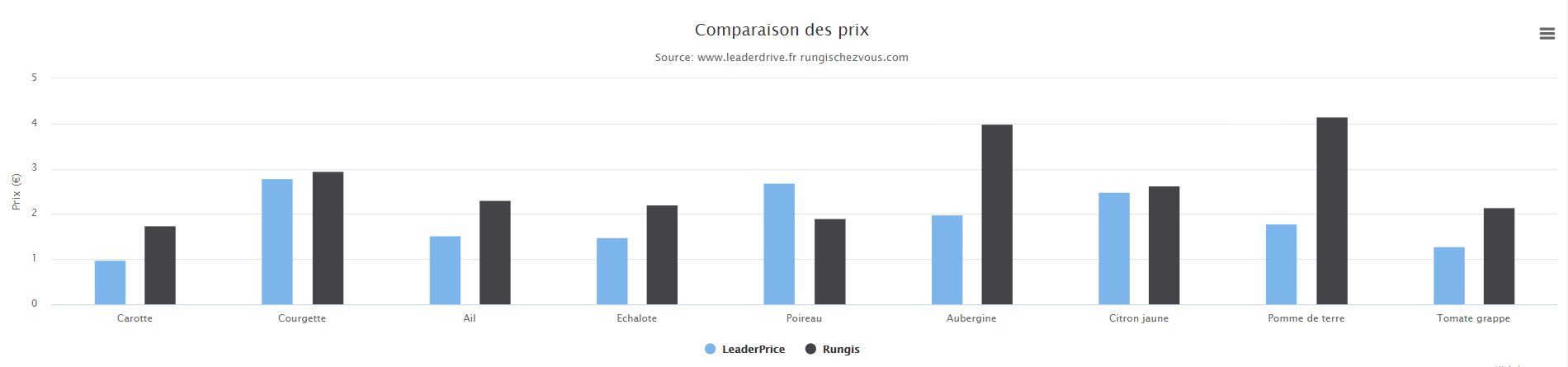
### 6.1.1. Support de visualisation

Deux supports de visualisation on été choisis pour l’analyse des prix des articles : un tableau regroupant les prix des légumes selon les sources ainsi qu’un diagramme à barres permettant une comparaison plus intélligible des prix.

Le tableau :

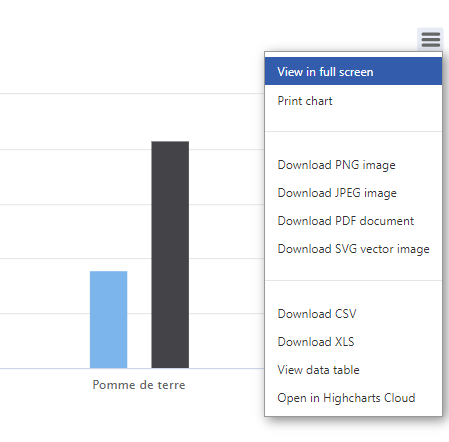


Le diagramme à barres :



### 6.1.2. Outil de diffusion

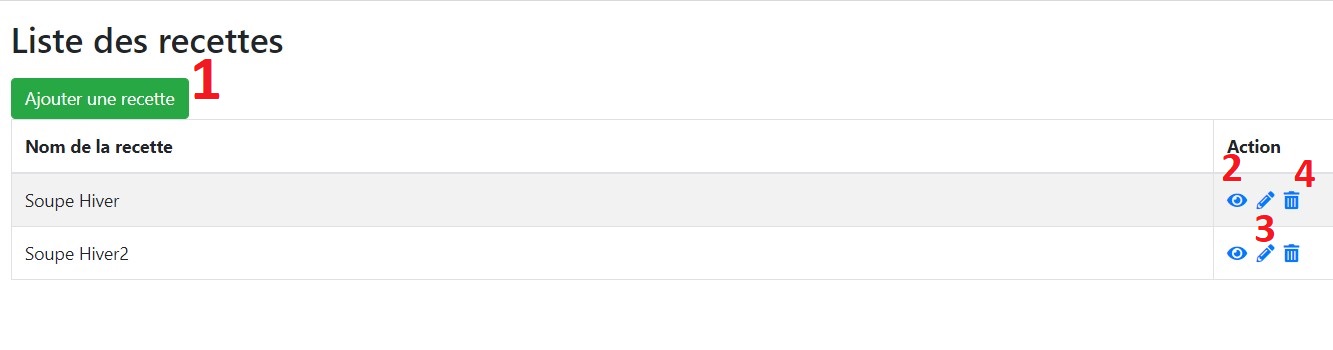
Il est possible d’extraire les données à partir d’un module javascript qui extrait les données sous plusieurs format possible : CSV, XLS, PNG…



## 6.2. Section création de recette

### 6.2.1. Système CRUD

Il est possible pour une recette de pouvoir en créer une nouvelle, de lire les détails de la recette, de la modifier ou de la supprimer. C’est ce qu’on appelle le CRUD (Create, Read, Update, Delete).



Comme on le voit sur l’image, il y a sur la page de la liste des recettes un bouton pour chaque action du CRUD :

-1 : Ajouter une recette

-2 : Voir les détails de la recette

-3 : Modifier la recette

-4 : Supprimer la recette

### 6.2.2. Rapport d'analyse de la valeur nutritionnelle

Dans la page d’affichage des détails d’une recette, il y a les détails liés à la valeur nutritionnelle liée à celle-ci. Il y a un tableau et un diagramme circulaire permettant de voir le pourcentage de lipide, glucide et protéine pour la recette.

Tableau :

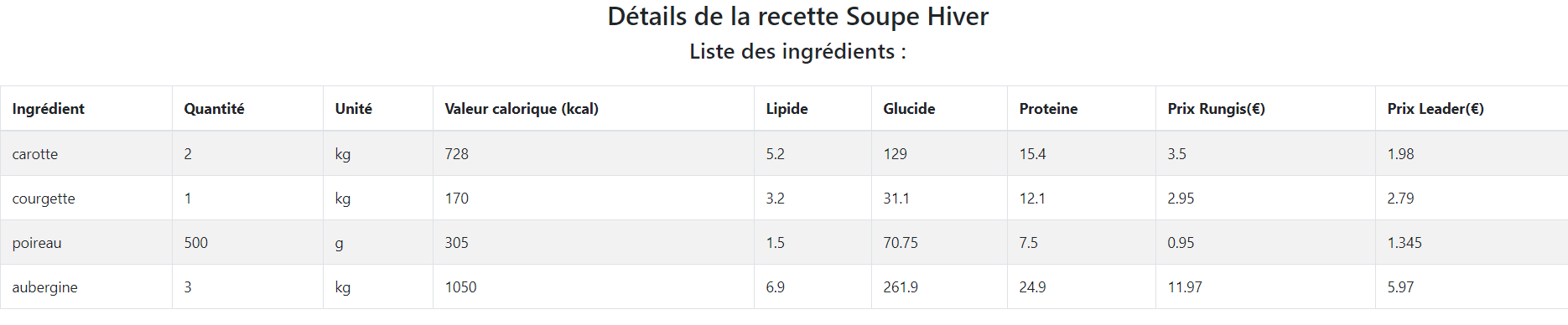
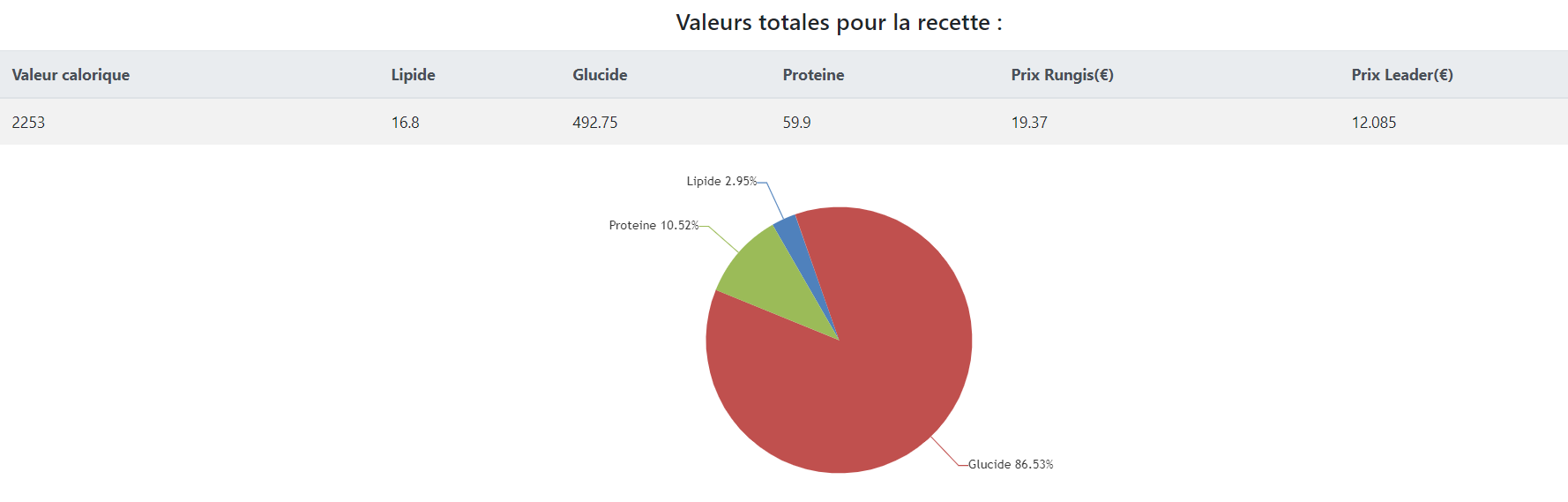


Diagramme circulaire :



Voici les calculs qui ont permis l’affichage des valeurs sur le tableau (en PhP) :

 // Calcul de la valeur sur une référence de la valeur sur 100g et selon la quantité

        $sum1 = ($row['VALEURCAL']\*$row['QUANTITE'])/100;

        $sum2 = ($row['LIPIDE']\*$row['QUANTITE'])/100;

        $sum3 = ($row['GLUCIDE']\*$row['QUANTITE'])/100;

        $sum4 = ($row['PROTEINE']\*$row['QUANTITE'])/100;

        // Conversion de kg en g pour le calcul des valeurs

        if($row['UNITE\_MASSE'] == 'kg'){

          $sum1 = ($row['VALEURCAL']\*1000\*$row['QUANTITE'])/100;

          $sum2 = ($row['LIPIDE']\*1000\*$row['QUANTITE'])/100;

          $sum3 = ($row['GLUCIDE']\*1000\*$row['QUANTITE'])/100;

          $sum4 = ($row['PROTEINE']\*1000\*$row['QUANTITE'])/100;

        }

        // Calcul du prix sur une référence du prix au kilo et selon la quantité

        $total1 = $row['PRIX\_RUNGIS']\*$row['QUANTITE'];

        $total2 = $row['PRIX\_LEADER']\*$row['QUANTITE'];

        // Conversion de g en kg pour le calcul du prix

        if($row['UNITE\_MASSE'] == 'g'){

          $total1 = ($row['PRIX\_RUNGIS']\*$row['QUANTITE']/1000);

          $total2 = ($row['PRIX\_LEADER']\*$row['QUANTITE']/1000);

        }

        // Calcul des valeurs totales

        $sumcal += $sum1;

        $sumlip += $sum2;

        $sumglu += $sum3;

        $sumprot += $sum4;

        $sumrungis += $total1;

        $sumleader += $total2;

Pour le diagramme circulaire :

      // Calcul pour le diagramme, résultat en pourcentage (sur 100%)

      $sum = $sumlip+$sumglu+$sumprot;

      $sumlipr= $sumlip\*100/$sum;

      $sumglur= $sumglu\*100/$sum;

      $sumprotr= $sumprot\*100/$sum;

Les données qui ont permis ces calculs ont été récupérées grâce à cette requête :

    $sql = "SELECT \* FROM ingredient INNER JOIN compose ON ingredient.ID\_INGREDIENT = compose.ID\_INGREDIENT INNER JOIN unite ON unite.ID\_UNITE = compose.ID\_UNITE WHERE ID\_RECETTE=".$\_GET['id'];

# 7. Bilan technique

L’application est maintenant terminée. Elle permet de faire tous ce qui a été spécifié dans le cahier des charges de départ. L’interface de visualisation des prix est automatiquement mise à jour grâce à un script. La page de recette est totalement fonctionnelle et permet de manipuler les recettes à souhait. La section réservée aux commandes reste néanmoins en cours de développement. On a également pu voir dans ce rapport qu’un calcul des prix et des valeurs nutritionnelles ont été ajoutés afin d’avoir davantage d’informations sur les recettes. Le principal objectif était de rendre cette application indépendante et maintenable par des personnes non-qualifiées pour cela. C’est pour cela que l’entité recette a été rendue dynamique, flexible afin que n’importe qui puisse procéder à sa modification, sa suppression…

Malgré le fait que cette application soit considérée comme terminée, des améliorations peuvent y être apportées. Tout d’abord, le nombre d’ingrédients disponibles : le scraping de données d’un site web est souvent soumis à une protection anti bot qui réduit grandement la diversification des sources. Le site LeaderDrive ne possède pas une large gamme de légume contrairement à Rungischezvous, ce qui explique le nombre limité d’ingrédient disponible. La contrainte du code est qu’il doit y avoir le même ingrédient disponible dans les deux sites. Une autre contrainte est que le résultat de la liste n’est plus pertinent à partir du troisième résultat, c’est pour cela que la comparaison des prix sur une liste s’effectue sur les deux premiers éléments.

La solution serait de de trouver d’autres sources de données dans un premier temps. Dans un second temps, rendre le code de scraping plus intelligent en identifiant si le produit recherché est disponible ou non dans une des sources et d’ajouter le prix en conséquence. Pour finir, il faudrait que le code soit assez autonome pour déterminer si un élément de la liste est pertinent ou non, pour permettre une analyse plus complète et que l’extraction des données se fasse non pas à partir d’une requête mais à partir de la section légume du site en question.

Or, cette solution n’était pas envisageable sur une durée de 1 mois.

# 8. Conclusion

Ce projet de fin de formation m’a permi de rendre compte sur mes capacités d’apprentissage et d’adaptation sur un temps limité et a été très enrichissant sur plusieurs points.

Tout d’abord, sur la méthode de recherche face à un problème : au fur et à mesure des problèmes rencontrés, les recherches devenaient de plus en plus pertinentes. La résolution des problèmes se faisait donc de plus en plus rapidement.

Ensuite, le développement de cette application a fait intervenir plusieurs langages. J’ai donc dû me former et surtout me documenter et trouver des solutions très rapidement afin de lever certaines contraintes. J’ai donc pu acquérir de solides compétences dans les langages utilisés.

Enfin, ce projet m’a appris à développer la gestion de projet en étant confronté à un planning avec un client et leurs besoins.