Synopsis : Création d’une application qui permet d’organiser sa propre collection de rhum. Soit en s’alimentant d’une bdd inhérente à l’application. Soit en entrant directement les informations nécessaires.

Contexte du projet :

Le Client : Mr Collecto

Prestataire : Mr Pèllé

Besoin : Mr Collecto est un grand collectionneur de rhum Caribéen. Il souhaite disposer d’une application qui lui permet de gérer sa collection de rhum. Son but est de créer sa propre liste comprenant l’ensemble de ses rhums. Ils souhaitent pouvoir ajouter et retirer des bouteilles en fonction des arrivées et sorties. Mais pour lui faciliter le travail il souhaite disposer d’une base de données, listant les rhums caraïbéens afin de sélectionner dans cette liste les bouteilles qu’ils disposent déjà.

Grand acheteur, il aimerait pouvoir suivre l’évolution des enchères sur ebay pour être sûr de ne louper aucune vente intéressante.

Solution envisagée :

Utilisation d’un framework pour créer l’application.

Techno: python, postgresql, webscrapping, flask, sqlalchemy

Livrable: une application, un rapport, une doc pour utiliser l’application, les scripts.

Présentation du projet

Objectifs et enjeux :

Construire une base de données facilement exploitable. Le but est d’avoir une première base de données sur laquelle l’utilisateur pourra s’appuyer pour créer sa propre liste.

Les collectionneurs sont toujours à l’affut de bonnes affaires, grâce à l’api ebay, l’utilisateur pourra suivre l’évolution des prix.

Les besoins :

1. Pouvoir interroger la base de données
2. Pouvoir récupérer des données de la bdd principale et la stocker dans sa propre bdd/table
3. Pouvoir ajouter des enregistrements dans sa propre bdd/table
4. Pouvoir supprimer des enregistrements dans sa propre bdd/table
5. Consulter le cours des enchères

Périmètre du projet :

On se concentre juste sur quelques pays : Jamaïque, Barbade, Martinique, Guadeloupe. Ce sont en gal les pays les plus plébiscité par les collectionneurs.

Les jeux de données que nous allons obtenir n’auront pas la granularité que souhaite un collectionneur. Nous allons nous focaliser sur la réalisation technique du projet.

Sources de données :

Site e-commerce spécialisé dans le rhum + APIebay + source personnelle

Description de l’architecture retenue : encours

Description des contraintes :

1. Récupération des données
2. Faire communiquer les tables de façon automatique
3. Affichages des données

Les contraintes de délais : seulement 6 semaines de dev avec du temps pour une autre certif

Les adhérences et/dépendances avec d’autres projets : ???

Organisation du projet : agile, planning sur clickup, point régulier (1-2 fois par semaines avec le formateur pour faire l point sur l’avancement du projet)

La répartition du travail : one man team

Spécifications

Les besoins à couvrir par l’application :

Partie 1

Partie 2

Moyens techniques

Conséquences organisationnelles

Formation

Pour les utilisateurs

Pour la maintenance

Construction de la base de données :

Difficultés : comment classer les données : par catégorie (blanc, ambré, vieux), par origine (Martinique, …), par marque (HSE, …). Vers quelle granularité je peux aller dans les données.

Le choix de la stratégie de construction est très important, car il va conditionner la stratégie de scraping des sites.

On peut aborder le scraping des sites par différents moyen. L’utilisateur peut accéder au produit de différentes manières. Ces manières influent sur l’url et sur le nombre de produits je peux accéder.

Choix du type de base de données sql ou nosql (avantage inconvénient)

Modèle relationnel.

On aura 4 tables de 4 provenances différentes avec une table qui centralise l’origine.

La conception des tables est assez simple, elles suivent tous la même logique.

La table User pose plus de problème. Comment lié cette table aux autres.

Question de la gestion des clés primaires. Soit une clé primaire spécifique à chaque table qui s’auto-incrémente en fonction des entrées. Mais risque de conflit lors des requêtes.

Il faut que la table user est sa propre clé primaire.

Gestion des dépendances et des contraintes, et gestion des transactions.

Modèle nosql :

On aura plusieurs collections. La clé primaire se gérera en fonction des enregistrements.

On évite ainsi les risques de doublons et de conflit.

Comme on ne peut pas rentrer dans une granularité optimale (choix mise en bouteille, à expliquer en amont).

On ne va pas faire de requête complexe, le but étant de trouver rapidement une bouteille.

Il faut bien réfléchir à sa structure des données, car cela va conditionner la qualité de nos requêtes.

Pas besoin de faire d’analyse poussée.

Gère maintenant les transactions

Gestion des doublons, elle est difficile mais primordiale

SOMMAIRE

Introduction

* Enonciation du projet
* Genèse du projet
* Annonce du plan

Dev

1. Préparation du projet
   1. Phase de planification
      1. Phase macro
      2. Kanban
   2. Expression du besoin
      1. Etat de l’art
         1. Le marché du rhum
         2. Les collectionneurs
         3. Les applications existantes
      2. Le périmètre du projet

* Le choix du rhum pour le moment
* La granularité des données, pour trouver un maximum de données

1. Conception du projet
   1. L’application
      1. Schéma prévisionnel de l’app
      2. Schéma fonctionnel
      3. Choix des technos
   2. Sources des données
      1. Liste
      2. Collecte des données

* Fichier csv
* Scrapping
* Api ebay
  + 1. Rgpd
  1. MCD – MLD
  2. Création de la BDD
  3. Préparation des données
     1. Pandas
     2. Regex
  4. Insertion des données
  5. Optimisation/ gestion des doublons

1. Exploitation des données
   1. Requête sql interrogation de la bdd
   2. Dash
2. Conclusion

# INTRODUCTION

Nous allons aborder dans ce rapport la thématique de la gestion d’une collection de spiritueux appliquée à un travail technique lié à la data. Le but est de concevoir une base de données permettant à un utilisateur de créé rapidement l’inventaire de sa collection.

L’idée du projet vient de mon histoire personnelle. J’ai travaillé pendant trois ans pour un producteur de rhum en Martinique appelé HSE. Cela m’a permis d’apprendre énormément sur l’univers des spiritueux. Et ce qui était d’abord un travail, c’est transformé petit à petit, de dégustation en dégustation, en passion. Au fil du temps j’ai commencé à acquérir des bouteilles de plus en plus prestigieuses et ainsi constituer une petite collection d’une centaine de références. Lorsqu’il a fallu déménager en France, j’ai dû établir « à la main » un inventaire de ma collection à l’aide d’un fichier excel. La tâche c’est révélé plus ardu qu’anticipé. C’est à partir de cet événement qu’à germé l’idée de créer une application pour faciliter ce travail. Je profite de cette formation pour me lancer dans ce projet.

Pour ce projet, nous allons essayer de répondre à des problématiques tournant autour de la gestion d’une collection. Nous allons nous mettre dans la peau d’un collectionneur de rhum pour ce rapport.

Mais dans un premier temps, il se doit de définir ce qu’est le rhum. Wikipédia défini Le rhum (anglais : rum, espagnol : ron) comme une eau-de-vie originaire des Amériques, produite soit par distillation de sous-produits fermentés de l'industrie sucrière, aussi appelé mélasse : le rhum industriel ou traditionnel, soit à partir du jus de canne à sucre fermenté : le rhum agricole. On distingue quatre catégories de rhum :

* Les rhums blancs (directement issue de la distillation)
* Les rhums ambrés (élevé en fût moins de trois ans)
* Les rhums « vieux » (vieillie plus de trois ans en fût)
* Les rhums épicés (ou rhum arrangé)

Grâce à sa diversité, sa qualité et sa richesse aromatique, le rhum a su séduire les amateurs.

Souvent regardé comme un spiritueux bas gamme. Il a su au fil du temps, notamment grâce aux rhums caribéens et leurs rhums vieillis, acquérir ses lettres de noblesse.

Notamment les rhums vieux caribéens, car il bénéficie d’un climat tropical qui accélèrent le vieillisement.

Différence de prix avec le reste du monde. Prix augmentent mais encore très accessible par rapport aux whiskys

Le monde du rhum réunit une forte communauté. Par exemple, la Confrérie du rhum, le plus grand groupe francophone, compte plus de 40 000 membres.

Il y a un véritable engouement autour du rhum, les clients veulent en savoir plus. Ils existent une très grande diversité dans la production des rhums.

Périmètre du projet :

Le but est de créer, dans un premier temps une base de données avec lequel, un collectionneur pourra facilement créer sa collection.

Même si ce projet tourne autour du rhum, le but du projet est de pouvoir intégrer d’autres style de spiritueux (whisky, gin, vodka, …).

. Mais le but est de s’intéresser, dans le futur, a d’autre spiritueux.

Le but c’est de pouvoir donné la possibilité à l’utilisateur d’avoir un aperçu rapide de sa collection. D’avoir des informations intéressantes à propos de sa collection. Le nombre de bouteilles, …

Qu’elles sont les besoins d’un collectionneur. Le challenge est de créer une base de données faciles à exploiter.

Le premier challenge a été de récolter des données pour

1. L’état de l’art
2. Le marché du rhum

Le secteur du rhum a beaucoup évolué ces dix dernières années. Le marché des spiritueux, en France, est dominé par le whisky. Les ventes de rhum ne représentent que 10% du marché. Mais c’est le seul spiritueux en progression. Les autres alcools sont en stagnation voir en retrait (boisson anisé, cognac, …).

Rajouté des chiffres. La consommation de rhum en France. Graphique ?

Les principaux producteurs de rhum sont les ……

Avant tout porté par la consommation de rhum blanc (surtout pour les cocktails). On constate depuis quelques années que la consommation de rhum vieux augmente fortement.

1. Applications existantes

Ils existent des dizaines d’applications surout pour le whyski et le vin. Vivino pour le vin

Il existe actuellement deux applications autour de la gestion de collection spécifique au rhum et elles sont françaises.

Wikirhum qui est une base de données très complète sur le rhum.

RumX.

1. Expression du besoin

Les collectionneurs de spiritueux sont en général des personnes assez exigeantes sur leur achat. Pour l’acquisition d’une nouvelle bouteille, ils vont faire attention à un certains nombre de détails et chercher à obtenir un maximum d’informations possible.

De nombreux facteurs jouent : la rareté, la réputation de la maison, le style du flacon, le numéro de la bouteille, l’étiquette, la date de mise en bouteille, la date de mise en fût, …

La granularité des informations peut être importante. Mais si on pousse le souci du détails on risque de se perdre en information et de créer une application indigeste.

Même si on se concentre sur le rhum pour lancer le projet, le but est d’incorporer d’autres alcools dans l’application. Collecté autant d’information ne permettrai pas cette flexibilité.

Pour ce projet, nous allons nous concentrer sur les informations générales.

Le but de ce projet est de créer une base de données. Cette base de données servira de point de départ pour qu’un utilisateur puisse créer son propre listing

# CONCEPTION DU PROJET

1. La gestion du projet

Notre projet de certification est un projet individuel dont le sujet est libre. La première étape a donc été de faire des recherches préliminaires pour déterminer le sujet et sa faisabilité.

Après la validation du sujet par nos formateurs, le premier challenge a été de déterminer le planning prévisionnel du projet. En tant que novice, il n’est pas évident d’évaluer le temps à consacrer à chaque tâche. J’ai utilisé l’application Clickup pour créer mon planning. Elle dispose de nombreuses fonctionnalités permettant au mieux de gerer la plannifaction de faire un graphique de gant, des todolist.

J’ai d’abor défini les grandes étapes du projet en trois grandes phases : la phase de recherche, celle de développement et enfin celle de rédaction.

Capture d’écran

Dans un second temps, j’ai essayé de définir des tâches plus spécifiques à accomplir quotidiennement. Je me suis aussi inspiré de Kanban, pour la gestion quotidienne de mon projet.

Capture d’écran

Je contacté deux fois par semaine mes formateurs, afin de leur faire un retour sur l’avancé de mon projet, les difficultés que je rencontrais et les solutions que je trouvais.

1. Les sources de données

Trouver un jeu de données sur le rhum c’est avéré très ardus. Pour constituer la base de données initial, nous avons quatre jeux de données :

* Inventaire.xls : l’inventaire de ma collection dans un fichier excel.
* Rum\_data.csv : jeu de données en format csv trouvé sur kaggle (site qui contient des dataset opensource). Il contient une liste de 6000 rhums.
* Test.json : jeux de données obtenu par scrapping en utilisant le module scrappy
* Liste\_wikirum : fichier csv obtenu grâce à l’aimable participation du site wikirum.
* Api\_ebay

Ayant des données obtenues de sources différentes, j’ai sélectionné les informations les plus pertinentes à sélectionnées.

J’ai choisi des informations assez générales, je me suis concentré sur les données suivantes :

Le nom, la marque, l’origine, le volume et le degré