**420-A53-SF - Projet de session**

**COLLECTE ET STOCKAGE DES DONNÉES groupe 10128**

Sujet

**Analyse des données collectés et stockes par technique de web a scraper**

**Étudiant:**

Mohamed Toros 2097154

**Professeur encadrant**

**Amor Imami**

Janvier 2020

TABLE DES MATIÈRES

[Description de l’entreprise 4](#_Toc62207079)

[Besoin de grattage 4](#_Toc62207080)

[L’importation des modules nécessaire pour extraction de données: 5](#_Toc62207081)

[Méthode pour récupérer un le contenu de la page gratter: 5](#_Toc62207082)

[Méthode pour récupérer un objet DataFrame : 6](#_Toc62207083)

[Représentation analytique les cinq premiers spéciaux : 6](#_Toc62207084)

[Représentation graphique les dix premiers spéciaux : 7](#_Toc62207085)

[Stockage des données dans Base des données MongoDB : 7](#_Toc62207086)

# Description de l’entreprise

L’entreprise sujet de notre travail est spécialisé dans la vente et livraison de fleurs, qui propose à ses clients un très grand choix de fleurs en vente en ligne sur le site **:  <https://www.fleuriste.ca/francais/>** considéré le plus populaire au Québec.

# Besoin de grattage

Être en mesure de rendre service aux clients qui désirent acheter des fleurs, surtout dans les périodes ou il y a beaucoup d’échange des cadeaux entre les familles et les amis, a titre d’exemple les fêtes de Noel. Les clients en question dans ce genre de situations ou le nombre des achats est considéré élevé quand même. Et dans le cadre de la gestion de leur budget, Ils cherchent des promotions offertes par les vendeurs en ligne. Dans cet perspectif on offre aux clients les informations qui concerne les prix de la promotion et le montant qui économisé à une transaction ou achat donné. Comme ça le client va avoir une idée sur le montant total et le montant sauvé.



## L’importation des modules nécessaire pour extraction de données:

Graphical user interface, text

Description automatically generated

## Méthode pour récupérer un le contenu de la page gratter:

En passant en paramètre le lien du site en question, on obtient le contenu de la page html via la méthode ci-dessous :

Graphical user interface, text, application

Description automatically generated

## Méthode pour récupérer un objet DataFrame :

Dans le but de séparer les taches de collecte et stockage, on a créé une autre méthode dont son rôle réside à extraire et nettoyer les données spécifiques à notre futur analyse à l’étape suivant. Dont la définition la méthode en question :

Text

Description automatically generated

Donc, on trouve ici deux champs l’un pour stocker les anciens prix des fleurs(flowers\_old\_price) et un autre pour stocker les nouvelles valeurs sujet du promotion (flowers\_special\_price). Par la suite on a ajouté un autre champ calculé du montant qui va être économise si le client veut passer cette transaction, et d’ailleurs le client peut aussi consulter le lien pour savoir le nom et l’image de type de bouquet des fleurs choisi.

## Représentation analytique les cinq premiers spéciaux :

A ce niveau on offre aux clients les informations, à titre d’exemple on illustre ici juste les cinq premiers spéciaux, comme ça les clients peuvent prendre une idée sur les 5 top promotions à ce moment avec les liens des images des fleurs sujet de cette promotion.

Table

Description automatically generated

## Représentation graphique les dix premiers spéciaux :

A ce niveau comme montre l’image ci-dessous, on offre aux clients un aperçu rapide sur les dix premiers spéciaux et les noms fleurs sujet de la promotion actuelle. Sachant qu’il y a des clients qui préfèrent l’aspect visuel que celui tabulaire pour prendre une décision du choix.

Chart, bar chart

Description automatically generated

## Stockage des données dans Base des données MongoDB :

Finalement, on sauvegarde nos données dans une base de type NoSql, comme le cas ici MongoDB , en passant un objet de type DataFrame a note méthode définit par :

Graphical user interface, text

Description automatically generated