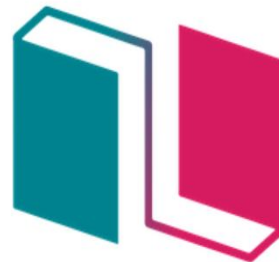

Pipeline ETL et Analyse Marketing

Programme d'extraction de données

- Configuration et utilisation du programme
- Etablir un pipeline ETL
- Démonstration du programme
- Évolutions futures



1. Configuration

→ Environnement virtuel

→ Modules

→ Points de vigilance

```
import requests
from bs4 import BeautifulSoup
import unicodedata
import os
import csv
from PIL import Image
from io import BytesIO
import re
```

```
from product_scraper import ProductScraper
from page_scraper import PageScraper
from category_scraper import CategoryScraper
```

```
class CategoryScraper:
    def __init__(self):
        pass

class PageScraper:
    def __init__(self, folder_csv):
        self.folder_csv = folder_csv

        if not os.path.exists(self.folder_csv):
            os.makedirs(folder_csv)

class ProductScraper:
    def __init__(self, folder_image):
        self.folder_image = folder_image
        if not os.path.exists(folder_image):
            os.makedirs(folder_image)
```

2. Utilisation

→ Classes et méthodes

→ Fonction général

→ Outils de débogage

Etablir un pipeline ETL

- **Extraction**
 - Scrapping / Identification grâce au HTML
 - Requests et BS4
- **Transformation**
 - Adapter le format
 - Convertir les images
- **Chargement**
 - Connaissance des fichiers CSV
 - Organisation des dossiers

Evolutions Futures

- BDD
- Extension du programme
- Suivis offres/demandes