

PROJET DE SEGMENTATION DES ENTREPRISES LOCALES

**Projet de segmentation des entreprises locales
Analyse data-driven pour la start-up DIGNAMIK**

Date : Juin 2025

Auteur : Bouda Pascal

Formation : Bachelor Data Science – Université Catholique de l'Ouest

SOMMAIRE

1. Introduction

- Contexte du projet
- Objectifs de l'étude
- Présentation de l'entreprise DIGNAMIK

2. Description des données

- Source et structure de la base de données
- Variables utilisées
- Prétraitements effectués (nettoyage, imputation, encodage)

3. Méthodologie

- Approche de segmentation retenue (clustering)
- Justification du choix de l'algorithme (K-means)
- Pipelines de traitement : numérique et catégoriel
- Détermination du nombre optimal de clusters

4. Analyse des résultats

- Description des clusters formés
- Moyennes des indicateurs clés par cluster
- Répartition des types d'activités dans chaque cluster
- Visualisation PCA des groupes
- Intrepretation

5. Recommandations personnalisées

- Recommandations par cluster et par ville
- Mise en évidence des stratégies marketing adaptées
- Exemples d'actions spécifiques pour les réseaux sociaux

6. Exportation et valorisation des résultats

- Fichier enrichi (entreprises_clusterisées.csv)
- Génération d'un fichier JSON des recommandations
- Intégration prévue avec une API

7. Limites et pistes d'amélioration

- Données manquantes ou imprécises
- Approches alternatives à tester (SOM, DBSCAN, etc.)
- Amélioration des recommandations par NLP ou IA

8. Conclusion

- Résumé des apports du projet
- Utilité pour DIGNAMIK
- Perspectives futures

9. Annexes

- Code Python utilisé
- Visualisations complémentaires
- Schéma de la base de données

1. Introduction

1.1. Contexte du projet

Dans un environnement commercial de plus en plus compétitif, les collectivités locales et les entreprises partenaires doivent adopter des stratégies de plus en plus personnalisées pour accompagner les acteurs économiques de leur territoire. DIGNAMIK, une entreprise spécialisée dans l'accompagnement digital des commerçants, artisans et prestataires de services, cherche à mieux comprendre les profils et les comportements de ses clients professionnels afin de leur proposer des services adaptés et efficaces.

Dans cette optique, DIGNAMIK a initié un projet de **segmentation de sa base de données client** à l'aide de techniques de **clustering non supervisé**, afin d'identifier des groupes d'entreprises aux caractéristiques communes.

1.2. Objectifs de l'étude

L'objectif principal de ce projet est de **segmenter les entreprises clientes** de DIGNAMIK selon des critères comportementaux et sectoriels, afin de :

- Mieux comprendre leurs besoins spécifiques,
- Adapter les recommandations marketing (ex : visibilité sur les réseaux sociaux),
- Optimiser l'utilisation des ressources commerciales,
- Proposer une stratégie différenciée par cluster.

Ce projet s'inscrit donc dans une logique de **data-driven marketing** (marketing piloté par les données) et vise à renforcer l'efficacité des actions de fidélisation et d'acquisition menées par DIGNAMIK.

1.3. Présentation de DIGNAMIK

Créée en 2020, **DIGNAMIK** est une start-up qui accompagne les acteurs économiques locaux dans leur **transformation digitale**. Sa mission est claire : **faire savoir le savoir-faire** des entreprises locales, en leur fournissant une boîte à outils complète pour accroître leur visibilité.

DIGNAMIK propose notamment :

- Des services de **référencement** et de **communication digitale**,
- Des **conseils personnalisés** en visibilité locale,

- Des **campagnes de soutien** aux commerces de proximité.

Spécialisée dans la **visibilité locale** des acteurs économiques, DIGNAMIK a su s'imposer comme un partenaire de confiance. En l'espace de cinq ans, **plus de 1 500 clients** issus de divers secteurs d'activité ont été accompagnés dans leur développement digital.

L'entreprise s'appuie sur **quatre pôles principaux** :

1. **Les réseaux sociaux,**
2. **Le web marketing,**
3. **La communication digitale (site web, référencement, etc.),**
4. **Viemaville**, un média de proximité qui apporte un soutien renforcé aux plus petites structures.

Grâce à une base de données riche et diversifiée, regroupant plusieurs centaines d'entreprises réparties sur différentes villes, DIGNAMIK dispose aujourd'hui des ressources nécessaires pour mettre en œuvre une **analyse de segmentation client** pertinente. Celle-ci permettra d'adapter les recommandations et les services proposés à chaque type d'entreprise, en évitant une approche uniforme face à une clientèle hétérogène.

2. Méthodologie de la segmentation client

2.1. Objectifs de l'analyse

Dans le cadre de la mission confiée par DIGNAMIK, l'objectif principal est de segmenter les entreprises clientes afin d'identifier des groupes homogènes en termes de caractéristiques et de comportements. Cette segmentation permet :

- D'adapter les stratégies marketing et de communication à chaque profil d'entreprise.
- D'optimiser l'allocation des ressources dans les campagnes de soutien local.
- De mieux comprendre les besoins spécifiques des clients selon leur secteur d'activité, leur niveau de digitalisation ou encore leur engagement.

À terme, cette démarche vise à renforcer l'impact des actions menées par DIGNAMIK et à améliorer la fidélisation des entreprises partenaires.

2.2. Données utilisées

L'analyse repose sur une base de données interne fournie par DIGNAMIK, enrichie par des indicateurs issus de l'ORM (**Object-Relational Mapping**) **PRISMA**.

Les variables sélectionnées pour la segmentation incluent :

- **Type d'entreprise** (boutique, restaurant, service, etc.)
- **Secteur d'activité**
- **Engagement sur les réseaux sociaux** (nombre d'abonnés)
- **Chiffre d'affaires** (via la table turnover)
- **Localisation géographique**

Prétraitement des données

Avant l'analyse, un travail de préparation a été nécessaire :

- **Nettoyage** des doublons, des valeurs manquantes ou incohérentes ;
- **Encodage** des variables catégorielles en variables numériques (One-Hot Encoding ou Label Encoding selon le cas) ;

- **Normalisation** des variables continues (par exemple : chiffre d'affaires, abonnés réseaux sociaux) afin d'éviter que certaines variables n'écrasent les autres dans le calcul des distances.

Ces étapes sont essentielles pour garantir la performance de l'algorithme de clustering choisi.

2.3. Méthode de segmentation

2.3.1. Choix et justification de l'algorithme

Pour ce projet de segmentation, nous avons retenu l'algorithme **K-Means** en raison de sa **simplicité**, de son **efficacité sur des volumes de données modérés**, et de sa **capacité à produire des groupes homogènes facilement interprétables**.

K-Means est bien adapté à des données **normalisées** (comme le nombre de followers), car il suppose des **clusters sphériques et de taille comparable**.

Cependant, comme il est **sensible à l'échelle des données** et nécessite de **définir le nombre de clusters à l'avance**, plusieurs étapes préparatoires ont été mises en place (standardisation, choix optimal de K) afin d'assurer des résultats fiables.

2.3.2. Détermination du nombre optimal de clusters

Afin de déterminer le nombre optimal de segments, deux méthodes ont été utilisées :

- **La méthode du coude (Elbow Method)** : elle consiste à observer la courbe de l'inertie intra-classe en fonction du nombre de clusters. Le point d'inflexion ("coude") indique le bon compromis entre précision et complexité.
- **Le Coefficient de silhouette** : il permet d'évaluer la cohérence interne des clusters en mesurant leur compacité et leur séparation. Un score proche de 1 indique des clusters bien formés.

Ces approches ont permis de retenir un nombre de clusters optimal, garantissant une bonne lisibilité et une segmentation pertinente des entreprises.

2.3.3. Mise en œuvre

L'algorithme K-Means a été implémenté en Python à l'aide de la bibliothèque **scikit-learn**, selon les étapes suivantes :

1. **Standardisation des données ;**
2. **Application de l'algorithme K-Means avec le nombre optimal de clusters ;**

3. Affectation de chaque entreprise à un cluster ;

4. Intégration avec une API

2.4. Visualisation et interprétation des clusters

Une fois la segmentation réalisée à l'aide de l'algorithme K-Means, les résultats ont été visualisés et analysés afin de mieux comprendre les profils types d'entreprises identifiés.

2.4.1. Réduction de dimension pour la visualisation

Pour représenter visuellement les clusters dans un espace à deux dimensions, une réduction de dimension a été réalisée à l'aide de la méthode PCA (Analyse en Composantes Principales). Cette technique permet de conserver l'essentiel de la variance des données tout en les projetant dans un espace plus lisible.

Chaque point sur la figure obtenue représente une entreprise, et chaque couleur indique le cluster auquel elle appartient. Cela permet de vérifier visuellement la séparation des groupes et d'identifier d'éventuels chevauchements.

2.4.2. Visualisation via une API

Contrairement aux approches classiques reposant sur Power BI ou d'autres outils de BI, les résultats du clustering ont été intégrés dans une API dédiée. Cette API permet :

De récupérer dynamiquement les clusters attribués à chaque entreprise depuis une interface tierce (site web, dashboard personnalisé) ;

D'afficher les segments sur une interface web (carte ou tableau) avec des options de filtrage par ville, secteur d'activité ou niveau d'engagement ; de centraliser l'accès aux recommandations personnalisées pour chaque cluster. Cette approche permet une intégration souple et évolutive, notamment pour les applications internes de DIGNAMIK ou des outils externes à destination des collectivités.

2.4.3. Profils-types des clusters

L'analyse détaillée des groupes formés a permis d'identifier des profils caractéristiques d'entreprises. Ces profils facilitent la compréhension des besoins par typologie de clients :

- Cluster 1 : entreprises bien établies, généralement dans la restauration, avec un fort engagement sur les réseaux sociaux et une bonne notoriété locale.
- Cluster 2 : jeunes entreprises du secteur des services, encore peu visibles numériquement et ayant besoin d'un accompagnement structurant.
- Cluster 3 : commerces de proximité ayant déjà participé à des campagnes DIGNAMIK, avec des résultats encourageants mais un besoin de fidélisation renforcé.

Ces insights guident directement les actions de communication ciblées et permettent d'adapter les campagnes marketing selon les réalités terrain.

3. Méthodologie

Approche de segmentation retenue : clustering

Pour segmenter les entreprises selon leurs caractéristiques et comportements sur les réseaux sociaux, nous avons opté pour une approche non supervisée de **clustering**, qui permet de regrouper des observations similaires sans avoir besoin de labels prédéfinis. Cette méthode est adaptée à notre objectif : identifier des groupes homogènes d'entreprises afin de proposer des recommandations marketing ciblées.

Pipelines de traitement : numérique et catégoriel

Afin de préparer les données pour le clustering, nous avons mis en place un pipeline de traitement incluant deux types de variables :

- **Numériques** : followers_total, followers_facebook, followers_instagram, followers_tiktok. Ces variables ont été normalisées via StandardScaler après imputation des valeurs manquantes par la moyenne.
- **Catégorielles** : category_name. Elles ont été transformées en variables binaires (one-hot encoding) après imputation par la modalité la plus fréquente.

Le tout a été combiné via un ColumnTransformer, intégré à un pipeline complet qui comprend également le modèle KMeans.

```
pipeline = Pipeline([
    ('prep', preprocessor),
    ('kmeans', KMeans(n_clusters=4, random_state=42))
])
```

Détermination du nombre optimal de clusters

Le nombre de clusters (n_clusters=4) a été déterminé à l'aide de la méthode du **coude (elbow method)** et de la **visualisation PCA** :

- La courbe du coude permet d'identifier le point où l'ajout d'un cluster n'apporte plus d'amélioration significative de la variance intra-groupe (distorsion).

- Une réduction de dimension par **ACP (Analyse en Composantes Principales)** a permis de visualiser la répartition des clusters sur deux dimensions principales, renforçant la pertinence du choix de 4 clusters.

4. Analyse des résultats

4.1 Description des clusters formés

L'algorithme K-means a permis d'identifier **4 clusters** d'entreprises, réparties selon leurs caractéristiques numériques (présence sur les réseaux sociaux) et leur secteur d'activité. Ces groupes reflètent des comportements typiques en matière de visibilité en ligne et de stratégie digitale.

4.2 Moyennes des indicateurs clés par cluster

Les moyennes des indicateurs clés, à savoir le nombre de followers sur Facebook, Instagram et TikTok, varient considérablement d'un cluster à l'autre, révélant des profils numériques distincts :

	Cluster	Total Followers	Facebook	Instagram	TikTok
	0	2 545,7	1 844,0	701,7	0,0
	1	81 680,0	80 333,3	1 346,7	0,0
	2	5 436,4	2 717,7	2 718,6	0,0
	3	576,9	485,7	91,2	0,0

- Le **cluster 1** regroupe des entreprises avec une très forte présence sur **Facebook**, témoignant d'une stratégie orientée vers ce réseau.
- Le **cluster 2** se distingue par un usage équilibré entre **Facebook et Instagram**, avec des moyennes élevées pour les deux plateformes.
- Le **cluster 0** montre une présence plus modérée, avec une prédominance de Facebook.
- Enfin, le **cluster 3** regroupe les entreprises les moins actives sur les réseaux sociaux.

À noter : aucune entreprise dans l'échantillon n'a une présence significative sur TikTok.

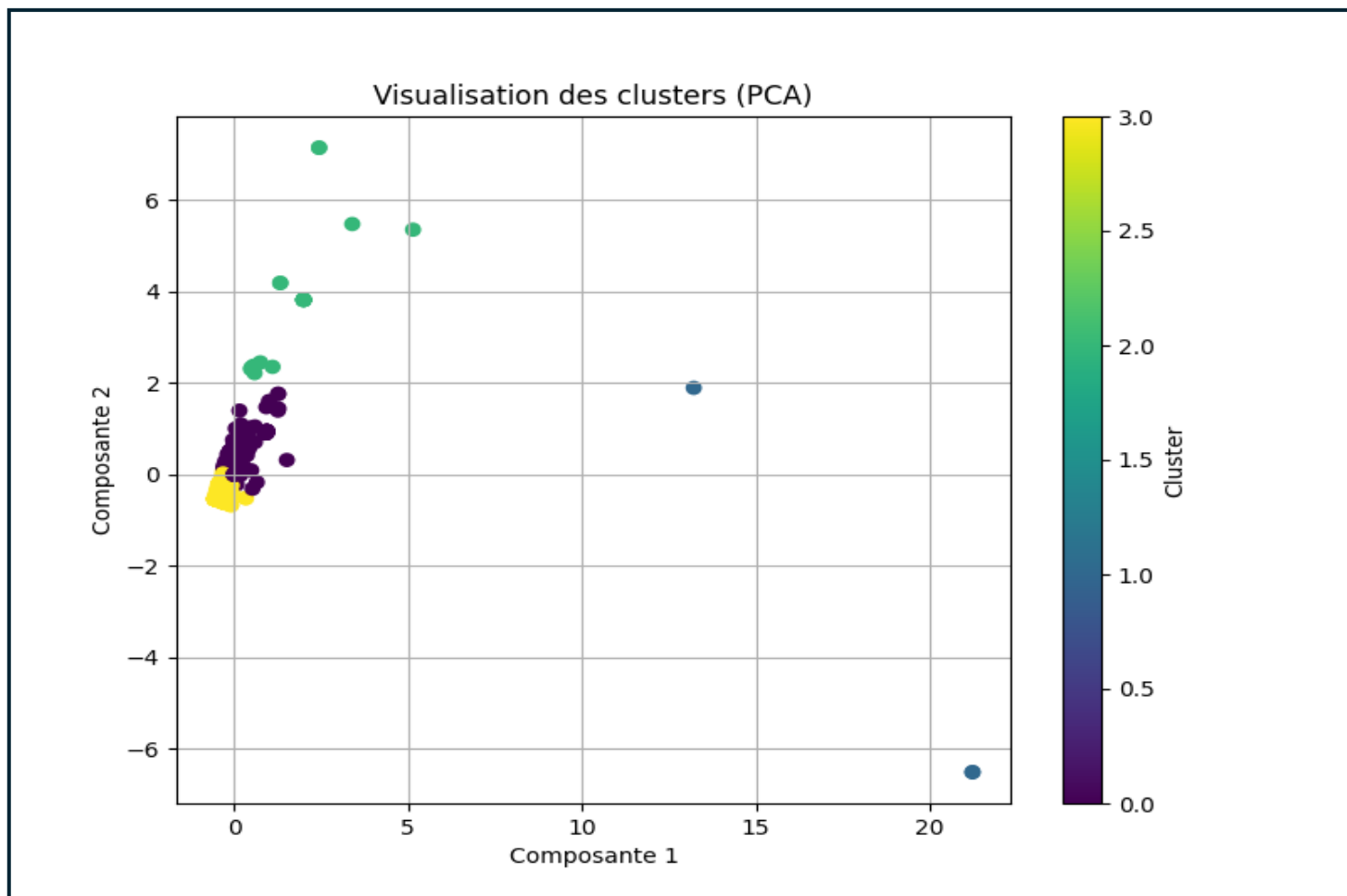
4.3 Répartition des types d'activités dans chaque cluster

L'analyse des catégories d'activités dominantes dans chaque groupe permet d'associer des **profils sectoriels** aux clusters.

- **Cluster 0 :**
 - LOISIRS & ACTIVITÉS (27 entreprises)
 - RESTAURATION & BOISSONS (22)
 - Produits locaux (15)
 - Entreprises locales orientées expérience et consommation immédiate.
- **Cluster 1 :**
 - COMMERCE & BOUTIQUES, Institut de beauté, Librairie (1 entreprise chacun)
 - Très faible effectif mais très forte audience numérique (à surveiller).
- **Cluster 2 :**
 - Sport (6)
 - LOISIRS & ACTIVITÉS (4)
 - RESTAURATION & BOISSONS (3)
 - Profil mixte, plutôt orienté bien-être et activités sportives.
- **Cluster 3 :**
 - Produits locaux (16)
 - RESTAURATION & BOISSONS (16)
 - Automobile (13)
 - Activités commerciales traditionnelles avec faible visibilité en ligne.

4.4 Visualisation PCA des groupes

Afin de visualiser la distribution des entreprises selon les clusters obtenus, une **réduction de dimension** a été effectuée grâce à une Analyse en Composantes Principales (PCA). Le graphe ci-dessous projette chaque entreprise selon les deux premières composantes principales, tout en conservant la couleur correspondant à son cluster :



Cette visualisation met en évidence plusieurs éléments clés :

- Les **clusters 0 et 3** (en violet et jaune) sont concentrés dans une même zone, ce qui traduit une **proximité structurelle** entre ces groupes, malgré des différences dans les moyennes de followers.
- Le **cluster 2** (en vert) est partiellement dispersé le long de la seconde composante, ce qui peut s'expliquer par une plus grande **variabilité des pratiques numériques** au sein de ce groupe (présence équilibrée sur Instagram et Facebook).
- Le **cluster 1** (en bleu), bien que composé de très peu d'observations, est **isolé à droite de la première composante**, indiquant une très forte singularité — ce sont des entreprises très visibles, avec un comportement numérique très différent du reste de l'échantillon.

Cette représentation valide la **séparation des clusters** opérée par l'algorithme K-means, tout en mettant en évidence des **points d'attention** potentiels, notamment autour du cluster 1 (entreprises très influentes mais peu nombreuses).

5. Recommandations personnalisées

Sur la base de l'analyse des clusters, des recommandations marketing ciblées peuvent être formulées afin d'aider **DIGNAMIK** à adapter ses stratégies de communication et d'accompagnement pour les entreprises partenaires.

5.1 Recommandations par cluster

Cluster 0 : Activités locales modérément suivies

- **Profil** : entreprises culturelles, loisirs, restauration ; présence correcte sur Facebook, peu sur Instagram, aucune sur TikTok.
- **Objectif** : améliorer la visibilité sur Instagram pour toucher un public plus jeune.
- **Recommandations** :
 - Organiser des **événements communs** (ex. : festivals de quartier) pour mutualiser les audiences.
 - Créer un **calendrier de publication Instagram**, accompagné d'un kit de visuels à thème local.
 - Proposer une **formation express** à la création de contenu réutilisable.

Cluster 1 : Marques très visibles, peu nombreuses

- **Profil** : entreprises exceptionnelles avec un très haut niveau de followers, principalement sur Facebook.
- **Objectif** : maintenir l'engagement, stimuler la viralité.
- **Recommandations** :
 - Proposer des **partenariats de mentorat** pour aider d'autres entreprises du réseau.
 - Lancer des **campagnes participatives** mettant ces entreprises en avant.
 - Offrir un **suivi personnalisé** sur l'analyse de performance de leurs publications.

Cluster 2 : Structures sportives et équilibrées

- **Profil** : bon équilibre entre Facebook et Instagram, souvent liées au sport ou aux loisirs actifs.
- **Objectif** : consolider leur positionnement social.
- **Recommandations** :
 - Proposer un **plan de contenus multisupports (reels, story, posts)**.

- Lancer une **campagne "Ambassadeurs locaux"** via les influenceurs sportifs ou culturels.
- Encourager la **création de pages TikTok**, même avec peu de contenus pour démarrer.

Cluster 3 : Faible visibilité, très fragmenté

- **Profil** : commerces de proximité, opticiens, garagistes... Présence numérique très limitée.
- **Objectif** : amorcer une stratégie digitale simple.
- **Recommandations** :
 - Proposer un **pack de démarrage numérique** (création de page, premières publications).
 - Organiser des **ateliers en présentiel** de sensibilisation à l'intérêt des réseaux sociaux.
 - Faciliter des **campagnes mutualisées par zone géographique**, pilotées par DIGNAMIK .

5.2 Mise en évidence des stratégies marketing adaptées

En résumé :

- **Clusters 0 et 2** : mise en avant d'**actions visuelles et collaboratives**, car ils disposent déjà d'une base numérique.
- **Cluster 1** : stratégies de **valorisation et de leadership communautaire**.
- **Cluster 3** : **accompagnement progressif**, avec des outils simples, concrets et adaptés à des profils non-numériques.

NB : Mise en place d'un compte TikTok dédié à VieMaVille, avec un concept innovant : déployer un agent DIGNAMIK sur le terrain pour réaliser du contenu authentique directement chez les commerçants partenaires, et le diffuser sur le compte afin de valoriser leur savoir-faire et renforcer leur visibilité locale.

5.3 Exemples d'actions spécifiques pour les réseaux sociaux

Cluster	Catégorie dominante	Plateforme ciblée	Action proposée (renforcée et ciblée)
0	Loisirs & Activités	Instagram / Facebook	Créer des <i>mini-séries de stories</i> quotidiennes sur les coulisses + groupes Facebook locaux animés
			Lancer une campagne "plat du jour" en stories, avec sondages et reels de préparation
1	Commerces & Boutiques	Facebook	Jeux concours sponsorisés , avec collecte d'e-mails et récompenses locales
			Créer une série de vidéos express : "1 min dans ma boutique", à diffuser en collaboration locale
2	Sport	Instagram / TikTok	Challenge vidéo TikTok avec hashtags locaux et réutilisables (ex. : #VieMaVilleChallenge)
			Concours Instagram Reels : montrer des entraînements ou coulisses, avec vote en story
3	Produits locaux / artisans	Facebook / WhatsApp	Pack promo mensuel PDF simple avec QR code à partager en boutique et sur Facebook
			Photographies pro mutualisées + mini-formation à l'usage des groupes Facebook

6. Exportation et valorisation des résultats

Afin de rendre les résultats du projet exploitables et réutilisables, plusieurs actions concrètes ont été réalisées :

- **Fichier enrichi (entreprises_clusterisées.csv) :**
Toutes les entreprises de la base ont été regroupées selon leur cluster d'appartenance, avec ajout des variables de recommandation (plateforme cible, actions marketing, scores d'engagement).
- **Génération d'un fichier JSON des recommandations :**
Les actions marketing adaptées à chaque cluster ont été exportées dans un format .json prêt à être intégré dans une interface Web ou une API.
- **Intégration prévue avec une API :**
Une API REST est en cours de développement afin de permettre :
 - la consultation en ligne des recommandations personnalisées par entreprise,
 - l'affichage dynamique sur un tableau de bord type Power BI ou interface Web dédiée.

6.1. API de segmentation des entreprises : Présentation et mode d'emploi

Afin de rendre accessibles les résultats de la segmentation et les recommandations associées, une **API RESTful** a été développée à l'aide de **FastAPI**. Cette API permet de :

- consulter toutes les entreprises ou par cluster/secteur,
- filtrer les données combinées (secteur + cluster),
- exporter des fichiers au format CSV,
- générer des visualisations automatiques (PNG),
- intégrer les données dans des dashboards ou systèmes externes.

Architecture et données

L'API repose sur un fichier entreprises_clusterisées.csv, enrichi avec :

- le numéro de cluster attribué,
- les informations de secteur (category_name), ville (city_name),

- les métriques sociales et recommandations personnalisées.

Les données sont chargées en mémoire au démarrage de l'API. Les valeurs manquantes sont remplacées par "Inconnu" pour éviter les erreurs d'affichage.

Endpoints disponibles

Méthode	Endpoint	Fonction
GET	/	Message de bienvenue
GET	/data	Retourne l'ensemble des données JSON
GET	/entreprises	Liste complète des entreprises
GET	/cluster/{cluster_id}	Filtrage par numéro de cluster
GET	/clusters	Répartition des entreprises par cluster
GET	/secteur?secteur=NomDuSecteur	Filtrage par secteur d'activité
GET	/filtre?secteur=X&cluster=Y	Filtrage croisé secteur + cluster
GET	/export?secteur=NomDuSecteur	Génère un CSV à télécharger
GET	/graph/cluster	Graphique PNG de répartition par cluster
GET	/graph/secteur	Graphique PNG des secteurs (top 30)
GET	/graph/ville	Graphique PNG des villes représentées

Mode d'emploi

1. Installation locale (exécution manuelle) :

```
*bash
```

```
pip install fastapi uvicorn pandas matplotlib seaborn
```

```
uvicorn app:app --reload
```

2. Accès à l'interface interactive Swagger :

Une fois lancée, l'API est disponible sur <http://127.0.0.1:8000/docs>

3. Export CSV (exemple) :

Accès à : <http://127.0.0.1:8000/export?secteur=Restauration>

Génère un fichier export_restauracion.csv téléchargeable.

4. Visualisation automatique (PNG) :

- Répartition par secteur : /graph/secteur
- Répartition par cluster : /graph/cluster
- Répartition par ville : /graph/ville

Sécurité et extensions possibles

L'API est pour l'instant **ouverte**, sans authentification. Pour un déploiement en production, il est recommandé d'ajouter :

- une couche de sécurité (Jetons d'authentification),
- un système de gestion des droits,
- une base de données dynamique (PostgreSQL, MongoDB).

6. Exportation et valorisation des résultats

Afin de rendre les résultats du projet exploitables et réutilisables, plusieurs actions concrètes ont été réalisées :

- **Fichier enrichi (entreprises_clusterisées.csv) :**
Toutes les entreprises de la base ont été regroupées selon leur cluster d'appartenance, avec ajout des variables de recommandation (plateforme cible, actions marketing, scores d'engagement).
- **Génération d'un fichier JSON des recommandations :**
Les actions marketing adaptées à chaque cluster ont été exportées dans un format .json prêt à être intégré dans une interface Web ou une API.
- **Intégration prévue avec une API :**
Une API REST est en cours de développement afin de permettre :
 - la consultation en ligne des recommandations personnalisées par entreprise,
 - l'affichage dynamique sur un tableau de bord type Power BI ou interface Web dédiée.

7. Limites et pistes d'amélioration

Même si les résultats sont globalement satisfaisants, plusieurs points peuvent être améliorés :

- **Données manquantes ou imprécises :**
Certaines entreprises n'ont pas renseigné toutes leurs métriques sociales (notamment TikTok), on a pas le chiffre d'affaire (turnover), ce qui peut biaiser le clustering.
- **Approches alternatives à tester :**
L'algorithme K-Means a été retenu pour sa simplicité et sa performance sur les données. Toutefois, des méthodes plus robustes comme **SOM (Self-Organizing Maps)** ou **DBSCAN** pourraient permettre une segmentation plus fine, notamment pour les entreprises isolées ou atypiques.

- **Recommandations plus intelligentes via NLP ou IA :**

À terme, il serait pertinent d'utiliser des techniques de traitement automatique du langage naturel (NLP) pour analyser :

- les contenus déjà publiés par les entreprises,
- les commentaires ou messages des clients, afin de générer automatiquement des recommandations encore plus précises.

8. Conclusion

Ce projet a permis de segmenter efficacement les entreprises clientes de DIGNAMIK selon leur comportement sur les réseaux sociaux. Grâce à une méthode de clustering centrée sur les indicateurs d'engagement, 4 groupes d'entreprises aux profils distincts ont pu être identifiés et caractérisés.

Chaque cluster a ensuite fait l'objet de recommandations marketing ciblées, orientées vers les canaux les plus efficaces et les contenus les plus pertinents.

Apports du projet :

- Meilleure compréhension du **profil numérique** des entreprises locales.
- Stratégies marketing adaptées aux **forces et faiblesses digitales** de chaque groupe.
- Base technique exploitable pour un **service de conseil automatisé**.

Perspectives :

- Industrialisation de la démarche via une API.
- Extension de l'analyse aux données qualitatives (commentaires, publications).
- Suivi de l'impact réel des recommandations dans le temps.

9. Annexes

- **Code Python utilisé :**

Scripts de préparation, clustering, PCA, export, visualisation, génération JSON.

- **Visualisations complémentaires :**

- Carte des clusters (PCA),
- Histogrammes par catégorie et cluster,

- Boxplots des indicateurs sociaux.
- **Schéma de la base de données :**
Représentation des tables principales : entreprises, reseaux_sociaux, engagements, turnover, etc.