PROJET DATA

Groupe INFINITY

- KONE Khalled
- ROTSEN Zoé
- NABAJOTH Jérémy



Sommaire

Introduction

- Contexte du projet
- Objectifs globaux

Sources de données

- Fichiers utilisés
- Origine et fiabilité des données
- Sources complémentaires

Nettoyage et structuration des données

- Table Personnes
- Table Associations
- Tables des Établissements Privés et Publics
- Schéma de base de données

• Typologie des entreprises et comparatif

- Répartition des entreprises par taille
- Classification par secteur d'activité
- Analyse

• Dynamique de création d'entreprises

- Dynamiques de création du Gosier
- Dynamiques de création des Abymes
- Dynamiques de création de Baie-Mahault

• Étude du secteur touristique au Gosier

- Typologie et spécialités
- Couverture des métiers du tourisme

Cartographie des entreprises et associations

Outils utilisés

Analyse de l'emploi local

 Etude de l'emploi des Gosériens dans les entreprises locales

• Traitement des données personnelles

- Rappel RGPD
- Usages commercial des données personnelles
- Précautions à prendre
- Usage politiques des données personnelles

Introduction

Présentation du projet

Ce projet a pour objectif d'exploiter les données locales de la commune du Gosier, en se concentrant notamment sur les entreprises, les associations et la population. Il vise également à analyser le fonctionnement de la commune ainsi que ses besoins socio-économiques.

Objectifs du projet

- Nettoyer, consolider et structurer les données disponibles
- Typologiser les entreprises du Gosier selon leur taille, leur activité et leur nature juridique.
- Analyser la dynamique entrepreneuriale sur la période 2018–2024
- Géolocaliser les entreprises
- Étudier spécifiquement le secteur touristique
- Soulever les enjeux éthiques et juridiques liés à l'usage des données personnelles

Sources de données

Source des données

Fichiers fournis:

- Fichiers Annuaires
- Fichiers Entreprises
- Fichiers Associations

Fichiers trouvés:

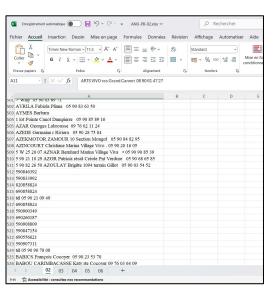
Les données ont été récupérées sur divers site tels que l'INSEE, la Région Guadeloupe, et les sites du gouvernement.

Liens

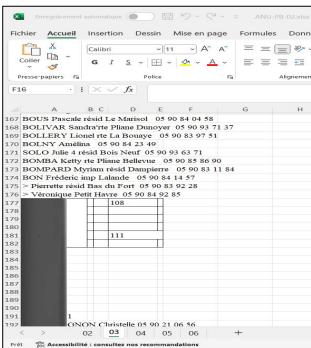
- https://www.journal-officiel.gouv.fr/pages/associations-recherche/?disjunctive.sour ce&sort=cronosort&q.titre=&q.assotype=association_type:assoLoi1901&refine.loc alisation_facette=Guadeloupe&refine.annonce_type_facette=Associations%20loi %20du%201er%20juillet%201901#resultarea
- https://regionguadeloupe.opendatasoft.com/explore/dataset/repertoire-national-des-associations-france/api/
- https://www.annuaire-mairie.fr/entreprise-le-gosier.html
- https://annuaire-entreprises.data.gouv.fr/entreprise/commune-du-gosier-21971113
 2
- https://regionguadeloupe.opendatasoft.com/explore/dataset/sirene-v3-consolidee-guadeloupe/table/?disjunctive.libellecommuneetablissement&disjunctive.etatadministratifetablissement&disjunctive.naturejuridiqueunitelegale&disjunctive.sectionetablissement&disjunctive.divisionetablissement&disjunctive.divisionetablissement&disjunctive.divisionetablissement&disjunctive.classeetablissement&disjunctive.sectionunitelegale&disjunctive.sectionunitelegale&disjunctive.classeunitelegale&disjunctive.activiteprincipaleetablissement
- https://www.sirene.fr/sirene/public/creation-fichier#nature-juridique

Nettoyage et structuration des données

Table Personnes



Visuel de l'annuaire



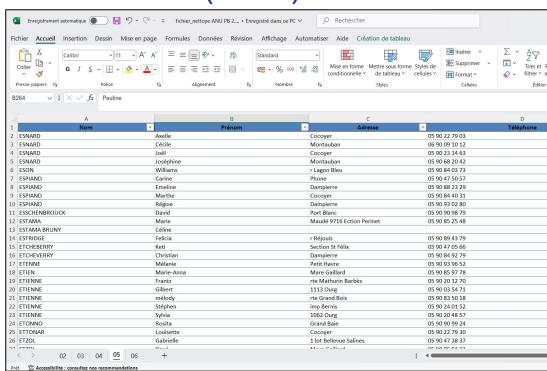
Les étapes du nettoyage :

- Analyse des données
- Production d'un script
 Python pour structurer les données
- Intervention manuelle pour ajuster les données
- Intégration des données des autres groupes via un code **Python**
- Suppression des doublons

Table Personnes (suite)

```
Donnée Etude > Traitement de l'annuaire > 🕏 code 1.py > .
     mport pandas as pd
    import re
   def extraire_infos(texte):
       texte = str(texte)
       if not texte or texte.lower() == 'nan':
           return pd.Series(['', '', '', ''])
       # Extraire le numéro de téléphone
       match tel = re.search(
           r'(?:[ .]?\d{2}){3,4})',
           texte.
           re.IGNORECASE
       numero = match tel.group(1) if match tel else ''
       texte_sans_tel = texte.replace(numero, '').strip()
       # Séparer le texte
       parties = texte sans tel.split()
       if not parties:
           return pd.Series(['', '', '', numero])
       # Cas où la ligne commence par un chiffre (donc adresse)
       if re.match(r'^\d', parties[0]):
```

Script pour le nettoyage



Résultat obtenu

Table Personnes (suite)

Qualité des données

Attributs	Nom	Prénom	Adresse Mail	Rue	Ville	Code Postal	Pays	Téléphone
Nombres de champs vides	30/2413	33/2413	2413/2413	197/2413	0/2413	0/2413	0/2413	297/2413
Pourcentage	1.24%	1.36%	100%	8.16%	0%	0%	0%	12.30%

Table des Associations

ETL (Extract Transform Load)

	ļ	
idea .idea	20/06/2025 09:53	Dossier de fichiers
.venv	10/06/2025 19:12	Dossier de fichiers
DATA	18/06/2025 17:30	Dossier de fichiers
EXTRACTED_DATA	18/06/2025 17:30	Dossier de fichiers
SCRIPTS	20/06/2025 08:01	Dossier de fichiers

- EXTRACTED_DATA
 - 2 associations_traitees_jo.csv
 - ETABLISSEMENTS_PRIVES_GOSIER.csv
 - Pichier_RNA-Nomenclature_Objet_Social.xlsx
 - ? jo_associations(Guadeloupe)_Gosier.csv
 - i joafe_domaine_activite.csv
 - Liste-Association-Gosier_V1.xlsx
 - ? Nomenclature_objet_social.xlsx
 - ? repertoire-national-des-associations-france.csv
 - rna_import_20250601_dpt_971.csv
 - ? rna_waldec_20250601_dpt_971.csv
 - ? sirene-v3-consolidee-guadeloupe (1).csv
 - 2 version 1.xlsx

EXTRACT

- ✓ □ SCRIPTS
 - <> carte_entreprises_publics_actives.html
 - domaine_activite.py
 - ETABLISSEMENTS_PUBLICS.csv
 - Pusion_des_fichiers(merge-final).py
 - Geoloc(ENT_PRIVEE).py
 - Geoloc(ENT_PUBLIC).py
 - 🥏 Geolocalisation(assos).py
 - e main.py
 - Qualite_donees.py
 - Traitement(jo_association).py
 - Traitement(Liste-assos_xlsx).py
 - Traitement(rep_natio_assos).py
 - Traitement(rna_import).py
 - Traitement(rna_waldec).py
 - Traitement_csv.py

TRANSFORM

- ✓ □ DATA
 - associations_fusionnees_final.csv
 - associations_guadeloupe_map.html
 - associations_merge_final.csv
 - associations_merge_final Copie.csv
 - associations_traitees_gosier_liste.csv
 - associations_traitees_jo.csv
 - 2 associations_traitees_repertoire.csv
 - associations_traitees_rna_import.csv
 - associations_traitees_rna_waldec.csv
 - <> carte_entreprises_actives.html
 - insert_associations.sql

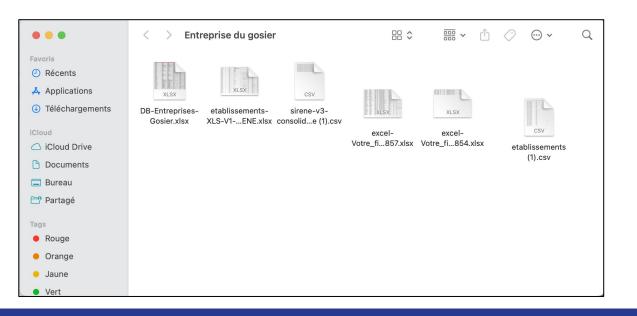
LOAD

VALEURS MANQUANTES (FICHIER_FINAL)

ID_RNA	1
NOM_COURT	0
NOM_COMPLET	0
STATUT_ASSOCIATION	190
TYPE_ACTIVITE	729
SECTEUR_INTERVENTION	729
OBJET	47
STATUT_JURIDIQUE	0
CODE_VILLE	0
ADRESSE_RUE	5
ADRESSE_CODE_POSTAL	0
ADRESSE_VILLE	0
ADRESSE_PAYS	0
DATE_CREATION	0
NB_ADHERENTS	1558
NOM_RESPONSABLE	1557
EMAIL_ASSOCIATION	1558
EMAIL_RESPONSABLE	1557
TELEPHONE_ASSOCIATION	1558
TELEPHONE_RESPONSABLE	1557
SITE_WEB	1523
DATE_DERNIERE_MISE_A_JOUR	190
DATE_FERMETURE	1499
LIEU_DECLARATION_PREFECTURE	729
GEOLOCALISATION	27
SIREN	1528
SIRET	1558
ETAB_PUBLIC	1558
ETAB_PRIVE	1558
SOMME	21216

Table des Établissements Privés et Publics

Données utilisées : Au total 6 fichiers d'un Total de 290 000 lignes ont été utilisé



Nettoyage et traitement des données

Les données ont été d'abord récolté dans les divers fichiers : le fichier principal qui a servi de support est le fichiers de sirène.

Une table UML a été mise en place pour sélectionner ce que l'on attendait dans les établissements publics et les établissement privés.

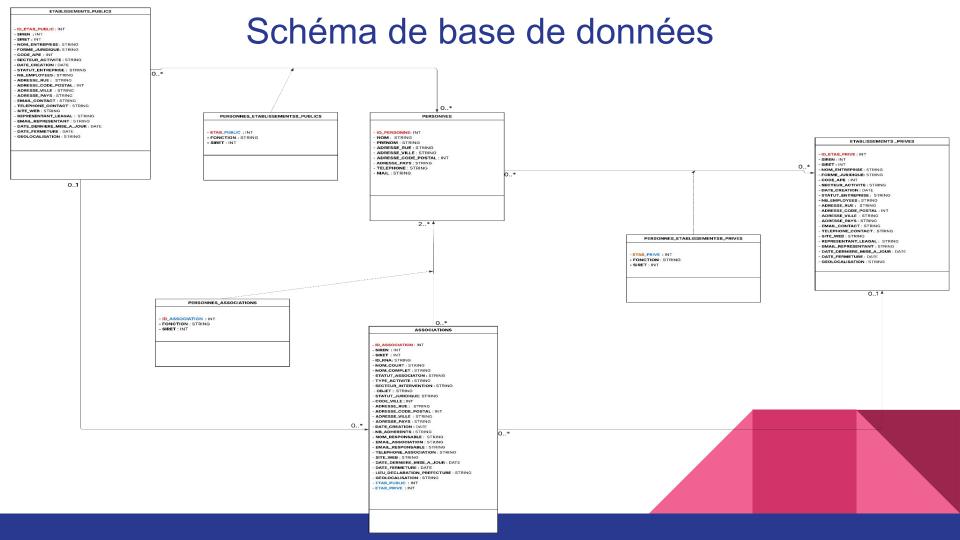
Le traitement des données des entreprise a été fait sur Excel.

Pour séparer les entreprises privées des publiques nous avons utilisé les codes de catégorisation juridique, puis avons cherché tous les codes de catégorisation juridique correspondant aux établissements publics.

Nombre de données manquante

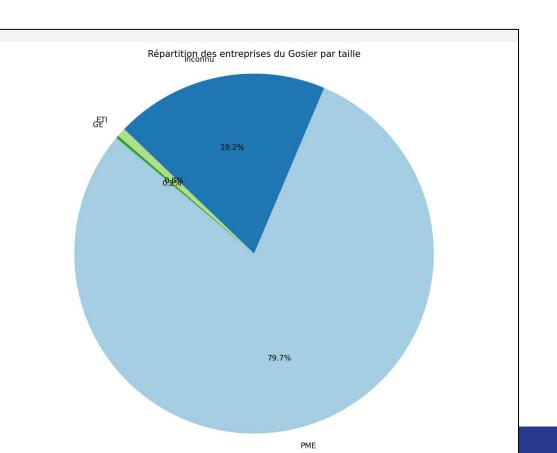
Colonne	Manquants
ID_ETAB_PRIVE	0
SIREN	0
SIRET	0
NOM_ENTREPRISE	9952
FORME_JURIDIQUE	0
TAILLE_ENTREPRISE	11717
CODE_APE	36
SECTEUR_ACTIVITE	40
DATE_CREATION	227
STATUT_ENTREPRISE	0
NB_EMPLOYES	23497
ADRESSE_VILLE	0
ADRESSE_RUE	1904
ADRESSE_PAYS	0
ADRESSE_CODE_POSTAL	0
EMAIL_CONTACT	23497
TELEPHONE_CONTACT	23497
SITE_WEB	23497
REPRESENTANT_LEGAL	11035
EMAIL_REPRESENTANT	23497
DATE_DERNIERE_MAJ	0
DATE_FERMETURE	11475
LATITUDE	275
LONGITUDE	275

Colonne	Manquants
ID_ETAB_PUBLIC	0
SIREN	0
SIRET	0
NOM_ENTREPRISE	0
FORME_JURIDIQUE	37
TAILLE_ENTREPRISE	2
CODE_APE	0
SECTEUR_ACTIVITE	0
DATE_CREATION	37
STATUT_ENTREPRISE	37
NB_EMPLOYES	5
ADRESSE_VILLE	0
ADRESSE_RUE	1
ADRESSE_PAYS	0
ADRESSE_CODE_POSTAL	0
EMAIL_CONTACT	37
TELEPHONE_CONTACT	37
SITE_WEB	37
REPRESENTANT_LEGAL	37
EMAIL_REPRESENTANT	37
DATE_DERNIERE_MAJ	20
DATE_FERMETURE	37
LATITUDE	0
LONGITUDE	0



Typologie des entreprises et comparatif

Répartition des entreprises par taille (Gosier)



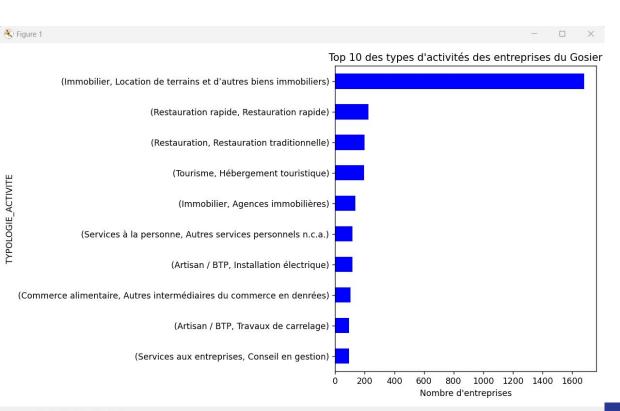
• **PME**: 79.7%

ETI: 0.8%

GE: 0.2%

• Inconnu : 19.2%

Classification par secteur d'activité (Gosier)



L'immobilier (gestion locative) est de loin le premier employeur d'entreprises, suivi du secteur de la restauration et du tourisme, logique pour une commune très touristique.



Analyse

PME (0-249 salariés) : 79,7 %

• Économie largement tournée vers le tourisme (hôtellerie, restauration, activités de loisirs) et les services aux particuliers, où les structures sont majoritairement de petite taille.

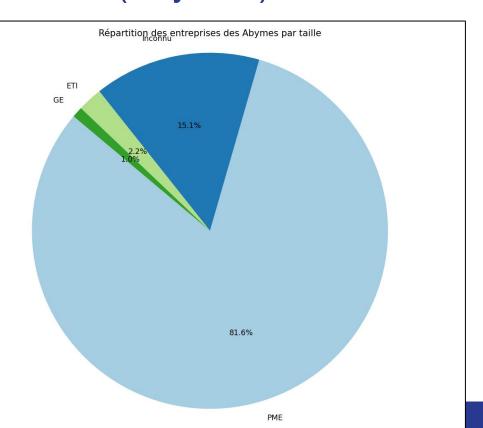
ETI (250-4 999 salariés) : 0,8 %

 Très faible présence d'entreprises de taille intermédiaire : le tissu industriel y est peu développé.

GE (≥ 5 000 salariés) : 0,2 %

 Presque aucune grande entreprise, ce qui reflète l'absence de grands sites industriels ou de sièges sociaux.

Répartition des entreprises par taille (Abymes)



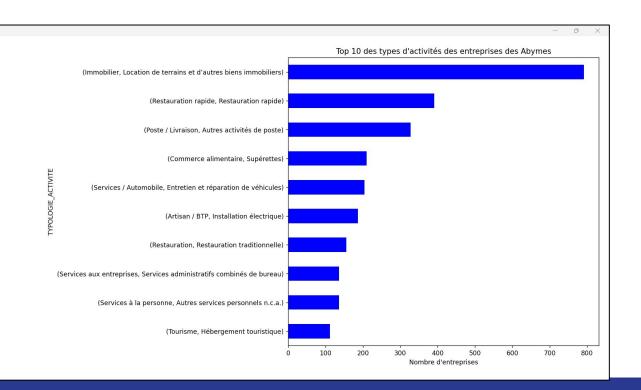
• **PME**: 81.6%

• **ETI**: 2.2%

GE: 1.0%

• Inconnu : 15.1%

Classification par secteur d'activité (Abymes)



Toujours un poids fort de l'immobilier, mais un profil plus « urbain » : logistique (postes, livraison), supérettes de proximité, réparation automobile.

Analyse

PME: 81,6 %

• Tissu majoritairement composé de commerçants, de TPE/PME de logistique et de services (maintenance auto, supérettes, agences de transport).

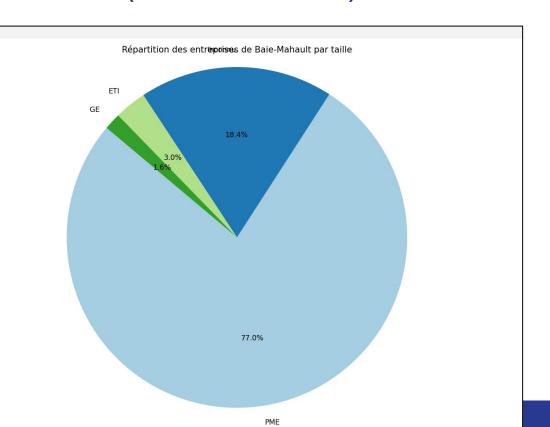
ETI: 2,2 %

 Quelques groupes de taille moyenne, souvent dans la distribution ou la gestion d'infrastructures (Eau, Énergie, Télécoms).

GE: 1,0 %

• Présence limitée de grandes entreprises, liées par exemple à la grande distribution.

Répartition des entreprises par taille (Baie-Mahault)



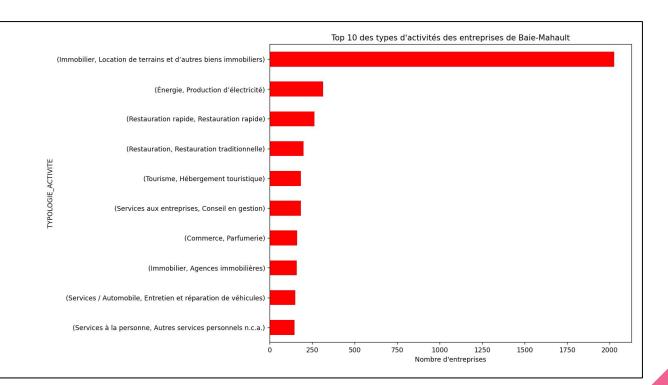
• **PME**: 77.0%

● **ETI**: 3.0%

GE: 1.6%

Inconnu : 18.4%

Classification par secteur d'activité (Baie-Mahault)



En plus de l'immobilier, on note la présence significative d'activités industrielles ou de services liés à l'énergie et à la logistique, en cohérence avec la zone portuaire et industrielle de Baie-Mahault.

Analyse

• PME: 77,0 %

Entreprises diversifiées : BTP, transport, services aux entreprises, négoce, etc., avec un socle solide de PME industrielles et artisanales.

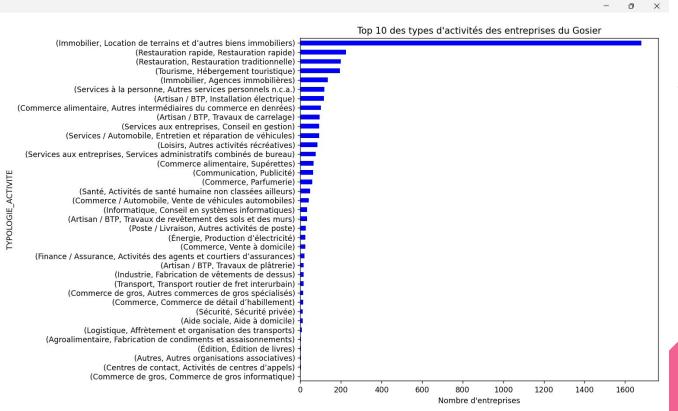
• ETI: 3,0 %

Un peu plus d'ETI qu'ailleurs, liées aux zones industrielles (ZI de Jarry) et aux activités portuaires.

• GE: 1,6 %

Quelques grandes unités industrielles (raffineries, grande logistique) installées dans la zone franche de Jarry.

Types d'entreprises manquantes au Gosier

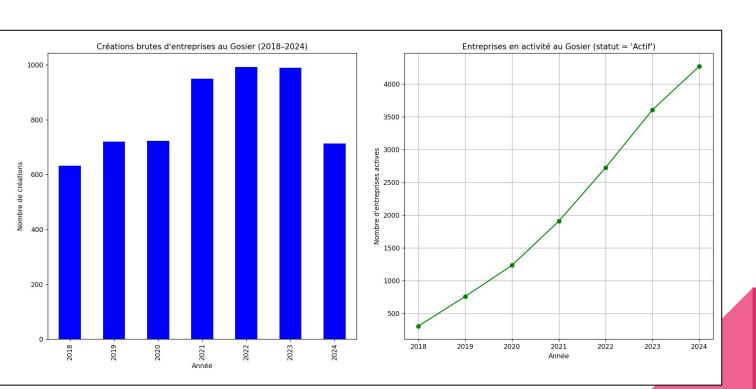


Différents secteurs sont sous-représentés au Gosier :

- Santé
- Petite Enfance
- Transport
- Informatique
- Environnement

Dynamique de création d'entreprises

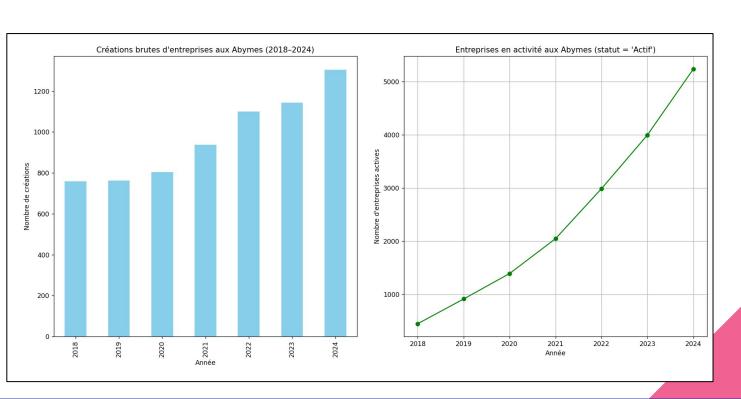
Dynamiques de création d'entreprises (Gosier)



Analyse

- **2019 :** Hausse (+14 %) grâce à de nouveaux dispositifs touristiques et à l'engouement pour l'auto-entrepreneuriat.
- 2020 : Stabilité des créations dû à la pandémie de COVID-19.
- **2021**: Pic de +32 % avec la réouverture progressive du tourisme.
- **2022**: Léger repli de la pente (+5 %) sous l'effet de l'inflation des matières premières dans le BTP et l'hôtellerie.
- 2023 : Quasi-stagnation ; le secteur loisirs commence à ressentir la fin des aides Covid.
- **2024 :** Fort recul (–28 %) : retombée post-relance touristique, hausse drastique des coûts (énergie, assurances) et report de projets d'investissement.

Dynamiques de création d'entreprises (Abymes)



Analyse

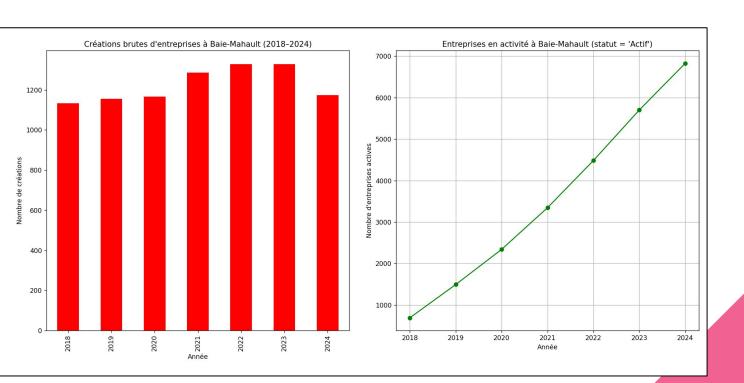
2019–2020 : stagnation, puis léger redémarrage ; la commune, très résidentielle, a pâti du premier confinement.

2021–2022 : Augmentation du nombres de petites entreprises, favorisé par les plans de relance.

2023 : Plateau, sous l'effet de la montée des coûts de production et de l'énergie et la fin des aides du COVID-19.

2024 : Nouveau bond (+13 %) : Forte augmentation liée aux projets de relances et aux projets digitaux

Dynamiques de création d'entreprises (Baie-Mahault)

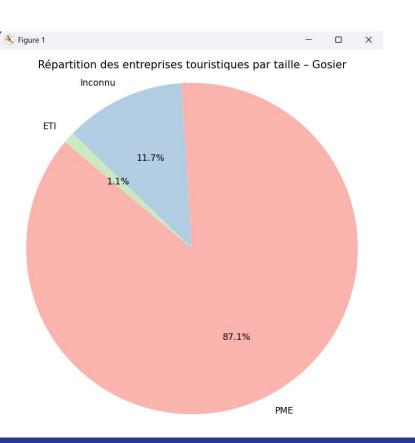


Analyse

- **2018–2020 :** Croissance modérée ; malgré le début de la pandémie fin 2020, les dispositifs d'aide (fonds de solidarité, reports de charges) ont soutenu une légère hausse des créations.
- **2021–2022**: Rebond fort (+9 % puis +3 %): effet de rattrapage post-confinement et boom du statut d'auto-entrepreneur, porté par la digitalisation accélérée.
- **2023**: Stabilisation; malgré l'inflation, les créations restent à un haut niveau grâce aux relances locales.
- 2024 : Recul de 11 % des créations : fin progressive de certains soutiens publics.

Etude touristique du secteur du Gosier

Etude touristique du secteur du Gosier



• **PME** : 87.1%

• ETI: 1.1%

• Inconnu : 11.7%

Etude du diagramme

Dominance des structures légères et autonomes

L'essentiel des acteurs touristiques à Gosier sont des structures à taille humaine, souvent familiales ou individuelles.

Fragilité face aux chocs externes

Covid-19: confinement et restrictions de voyage ont mis au ralenti ces petites entreprises, souvent sans réserves financières suffisantes pour tenir plusieurs mois.

Inflation et remontée des prix : coût du foncier, de l'énergie et des matières premières (restauration, entretien) impactent fortement ces acteurs.

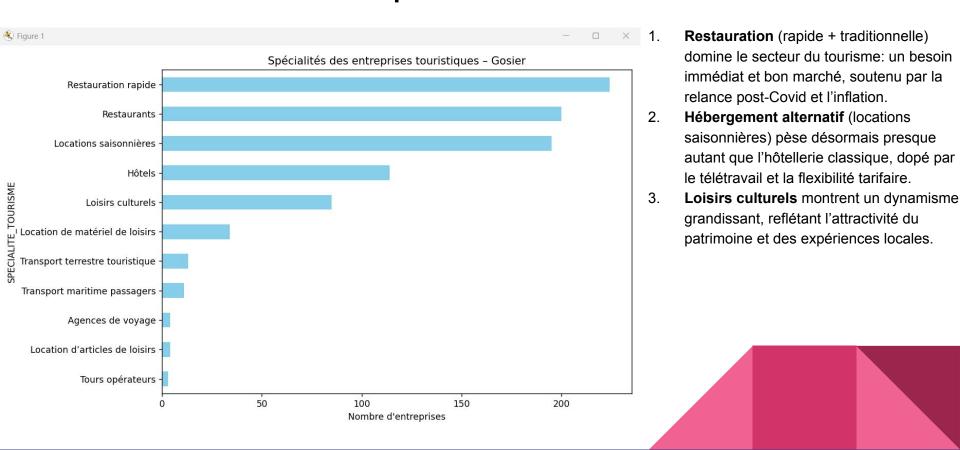
Potentiel de montée en gamme limité

L'absence quasi-totale d'ETI et de GE montre qu'il y a peu de « leaders de filière » capables de porter des projets touristiques d'envergure (zones hôtelières intégrées, parcs loisirs, congrès).

Opportunité pour la structuration

Près de 20 % d'inconnu laisse penser que certaines petites activités sont encore informelles.

Etude touristique du secteur du Gosier



Cartographie des entreprises

Cartographie

La cartographie des entreprises et des associations a été réalisée grâce à un script en python à l'aide des bibliothèques :

- pandas (pour lire le fichier csv)
- Folium (pour générer la carte)
- Folium.plungins (pour créer des clusters)

Les cartographies sont disponibles sur le GitHub

Analyse de l'emploi local

Etude de l'emploi des Gosériens dans les entreprises locales

Pour mener cette étude nous avons réaliser un formulaire :

Lien:

https://docs.google.com/forms/d/e/1FAlpQLSd8NbbpJFOhPJhm9qGdAS7a-4mEKLvi8CMbKBQhBdJOVStGjQ/viewform?usp=sharing&ouid=106284917456853084358

Canaux de diffusion souhaités :

- Réseaux sociaux
- Site web de la mairie
- Affichage avec QR code dans les lieux publics : mairie, médiathèque, commerces
- Relais associatifs: clubs, associations sportives ou culturelles
- Envoi par les écoles ou centres de formation pour toucher les jeunes

Traitement des données personnelles

Rappel RGPD

Données personnelles : Toute information permettant d'identifier directement ou indirectement une personne physique (article 4, paragraphe 1 du Règlement (UE) 2016/679)

Les principes du RGPD

- Le RGPD protège les données personnelles des individus
- Toute utilisation doit reposer sur une base légale (consentement).
- Les personnes doivent être informées et peuvent exercer leurs droits (accès, opposition, suppression).
- Les données doivent être utilisées uniquement pour un but précis, sécurisé, et conservées de façon limitée.
- Même les données issues de sources publiques sont concernées.

Usage commercial des données personnelles

L'usage de ces données personnelles à des fins commercial n'est possible que si la personne a donné son **consentement** clair et explicite.

- Informer la personne de la finalité commerciale du traitement
- Permettre de refuser facilement la prospection
- Tenir à jour une liste d'opposition (ne pas recontacter les personnes qui s'y opposent)

Précautions à prendre

1. Informer les personnes concernées

 Obligation d'informer : identité du responsable, finalité du traitement, droits d'accès, de rectification, d'opposition

2. Respecter les droits des personnes

- Droit d'accès
- Droit de rectification
- Droit d'opposition (notamment au démarchage)
- Droit à l'effacement (sous conditions)

3. Sécuriser les données

 Les données doivent être protégées contre tout accès non autorisé (stockage sécurisé, accès restreint).

4. Limiter l'usage

- Ne pas conserver les données plus longtemps que nécessaire
- Ne pas les réutiliser à d'autres fins non prévues

5. Documenter les traitements

Tenir un registre des traitements

Usage politique des données personnelles

Données sensibles: Les données sensibles sont des informations personnelles qui, si elles sont révélées, peuvent entraîner des discriminations ou des atteintes à la vie privée des personnes concernées

Les opinions politiques sont des données sensibles, et leur traitement est interdit.