

Sadjia Khadir

ADN Tourisme

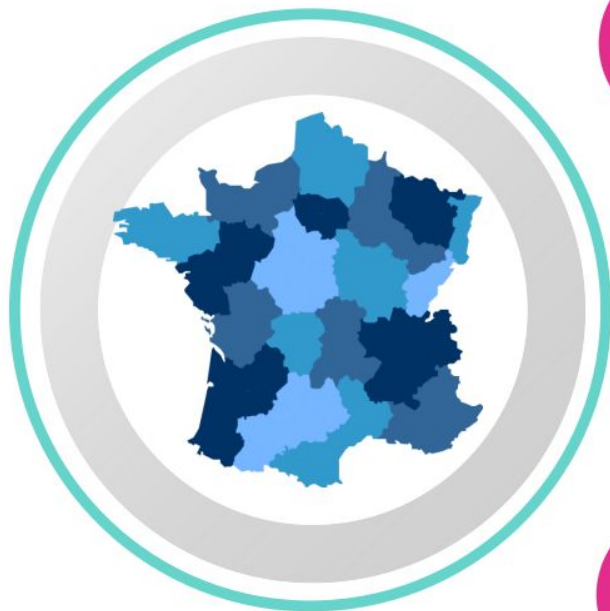
Projet 4 - Wild Code School

Projet réalisé et présenté par:



App
Streamlit

Sommaire



1

Contexte & Objectifs

2

Méthodologie & Outils

3

Nos Conclusions

1 Le Contexte



ADN Tourisme est née le 11 mars 2020 du **regroupement des trois fédérations historiques des acteurs institutionnels du tourisme** : Offices de Tourisme de France, Tourisme & Territoires et Destination Régions



ADN Tourisme représente, au niveau national, les forces conjuguées de **1200 structures** et **13 500 salariés**.

1 Les Délivrables et Objectifs



Un Prototype d'interface web



Comparer son offre

Un tableau de bord géographique permettant d'accéder aux informations touristiques



Montrer les tendances touristiques

Prédire la catégorie en fonction de la description de l'établissement



Harmonisation automatique de la catégorie

2 La Méthodologie



1



Récupération des données (450 000 POI) via Open data au format JSON.

2



Création d'une boucle pour ouvrir, aplatir et concaténer tous les fichiers

3



Nettoyage des données et sélection des colonnes

4



Prototype Web Dashboard ML

Choose your region

Normandie

Choose your department

Calvados

Select your interest point

" Les crins de verdure "

Select your information

Choose an option



Welcome to Appli_ADN_Tourisme



HOME



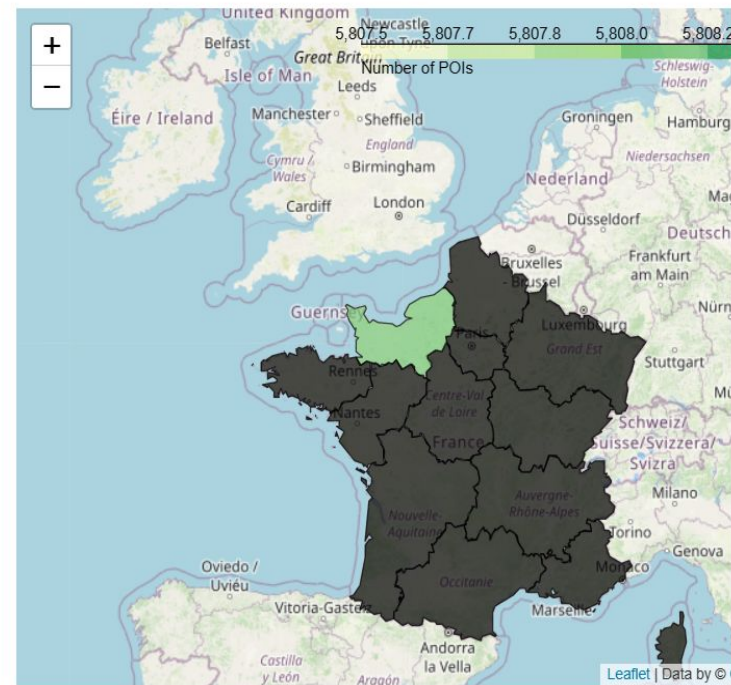
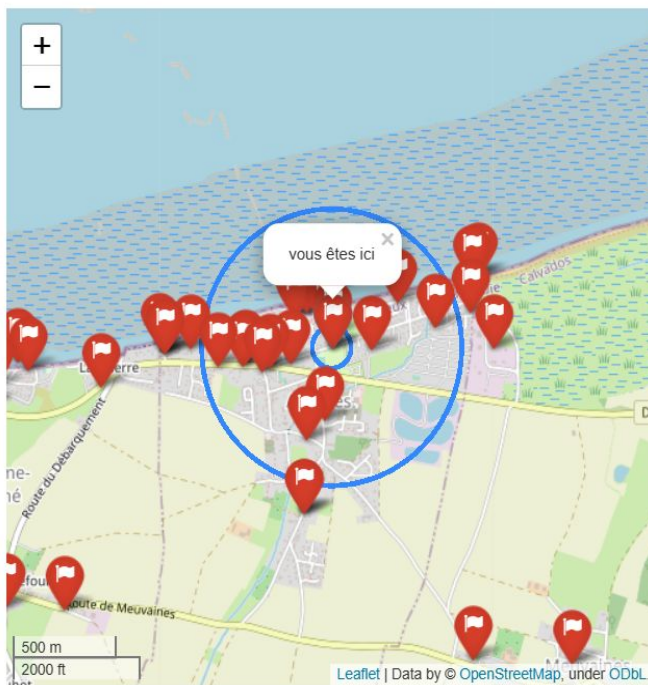
POI_Comparator



Dashboard



Category_Predictor



Choose your region

Normandie



Choose your department

Calvados

Select your interest point

Gold Beach Hotel

Select your information

POI_label x description_fr x

address x postcode x

departement_la... x

region_label x telephone x

email x



Welcome to Appli_ADN_Tourisme



HOME



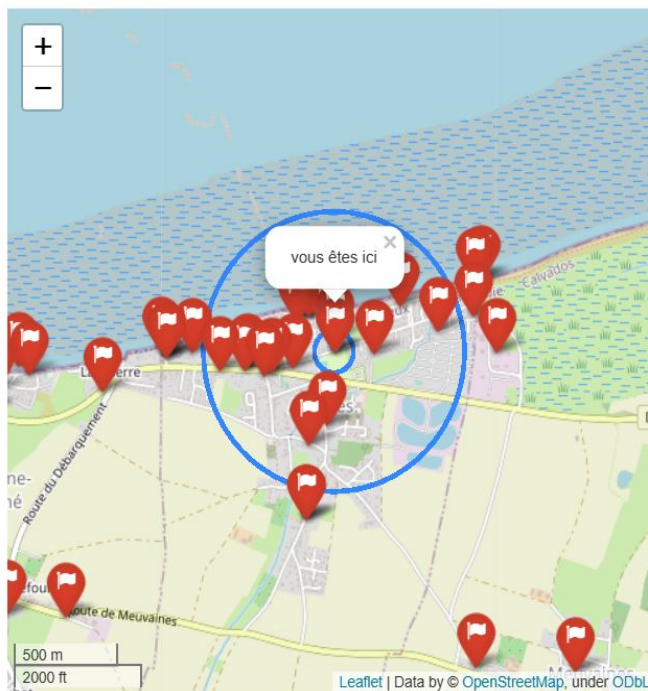
POI_Comparator



Dashboard



Category_Predictor



```
{
  "POI_label" : "Gold Beach Hotel"
  "description_fr" :
    "La Résidence Gold Beach vous propose 46 appartements de
    qualité pouvant accueillir de 1 à 5 personnes. Elle se situe
    à quelques mètres de la plage de sable fin d'Asnelles et à
    proximité de tous les commerces. Point de départ idéal pour
    découvrir..."
  "email" : "reception@gold-beach-hotel.com"
  "telephone" : "+33 2 31 51 11 10"
  "postcode" : "14960"
  "address" : "1 rue the Devonshire Régiment"
  "departement_label" : "Calvados"
  "region_label" : "Normandie"
}
```



HOME



POI_Comparator

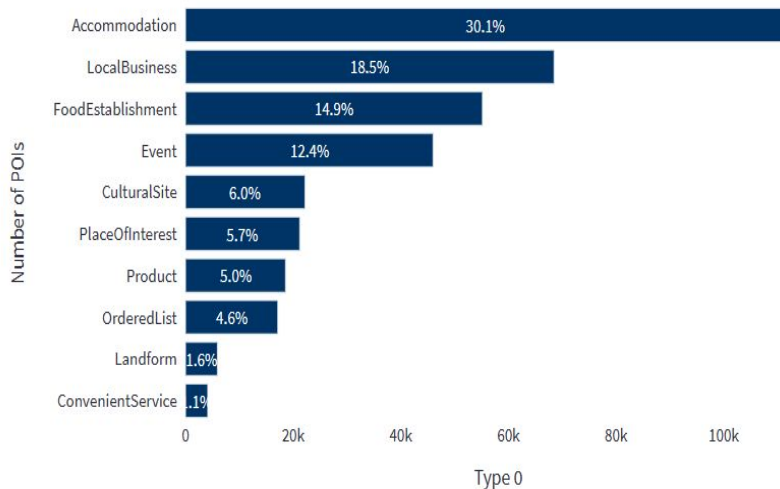


Dashboard

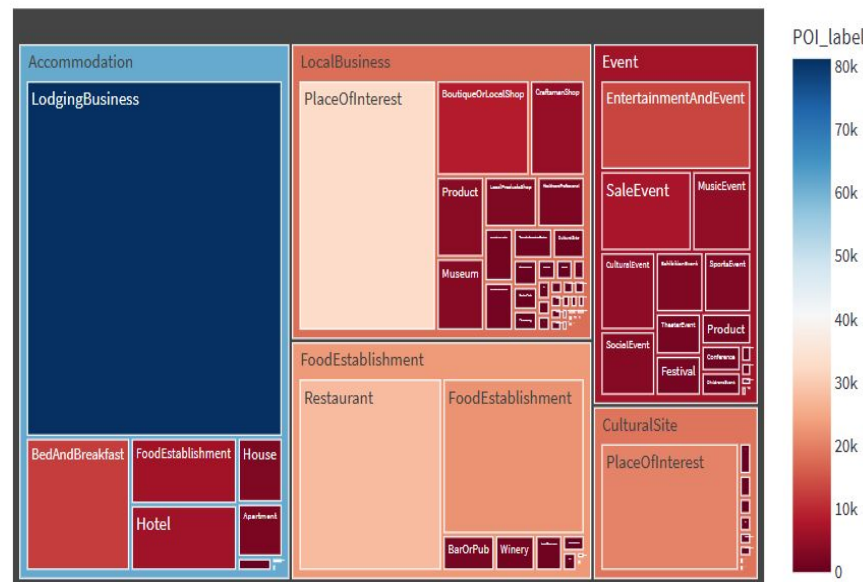


Category_Predictor

Top 10 by category



Points of interest by category



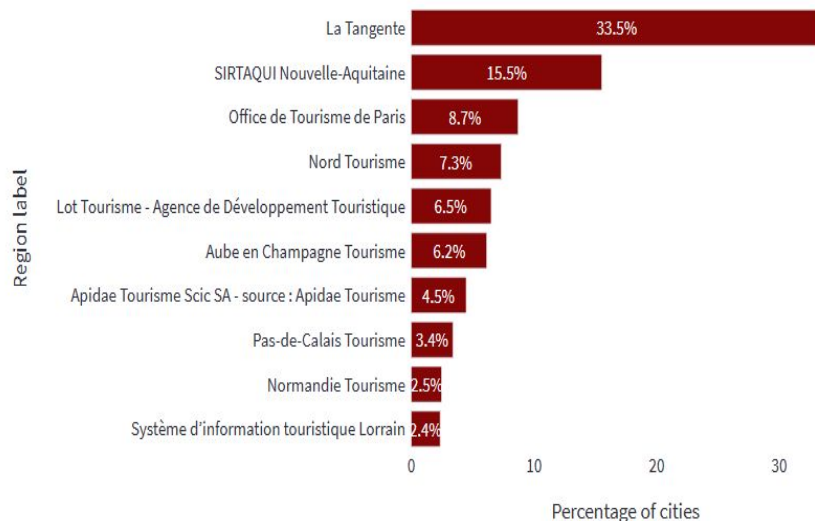

[HOME](#)

[POI_Comparator](#)

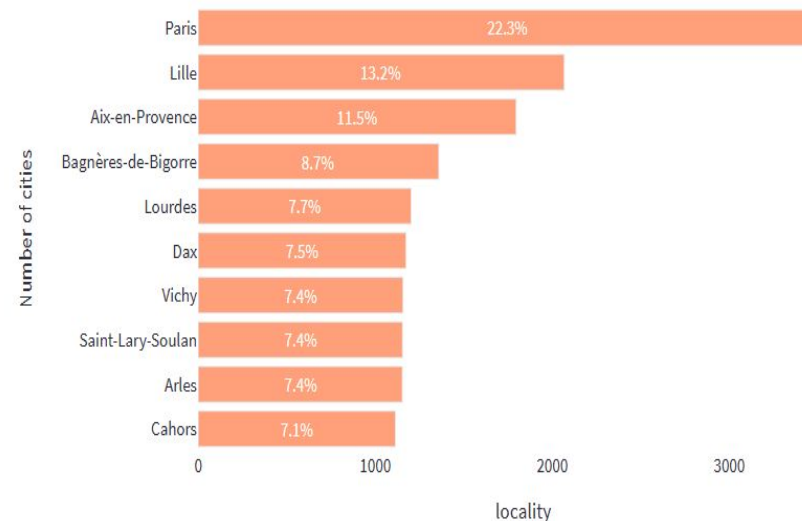
[Dashboard](#)

[Category_Predictor](#)

Absence of description by supplier



Top 10 by cities



The objective of this machine learning model is to verify the category and subcategories of a point of interest according to DATAtourism Ontology

The category is predicted using the establishment's name and description as input features for a supervised machine learning model (Logistic Regression)

This prototype has been deployed for accommodation and foodestablishment categories

 HOME POI_Comparator Dashboard**Category_Predictor**

- L'intention de l'analyse ML est de prédire une catégorie de points d'intérêt
- 4 modèles de classification supervisée sont évalués
- Analyse basée sur les fonctionnalités NLP
- Model retenu : Logistic Regression

Model	Train/test score	Accuracy/R ²
Logistic Regression	96.2/93.3	93.3
KNN Classification	93.9/91.4	91.4
Decision Tree Classifier	99.9/91.1	91.1
Random Forest Classifier	99.9/93.9	94.0

The objective of this machine learning model is to verify the category and subcategories of a point of interest according to DATAtourism Ontology

The category is predicted using the establishment's name and description as input features for a supervised machine learning model (Logistic Regression)

This prototype has been deployed for accommadation and foodestablishment categories



HOME



POI_Comparator



Dashboard



Category_Predictor



Here, you need to provide!

Name of your
establishment



Description of
your
establishment



Please, Enter the name and description of your establishment in French

Name:

haxo sushi

Description :

Click to see the category



The main category of the POI is:

['FoodEstablishment']

3 Conclusion

Nos Conclusions

- **30%** des POI appartiennent à la catégorie **hébergement**
- La Région la mieux référencée est la **Nouvelle Aquitaine**
- La Région la moins référencée est la **Guadeloupe**
- **34 086 POI (8%)** n'ont aucune description (parmi eux, la Tangente*, agence des Hauts de France représente 33%)



Nos Recommandations

- Harmoniser les champs à remplir dans chacune des offices de tourisme
- Collecter certaines données touristiques clefs (**labels et classifications**):
 - ✓ Prix
 - ✓ Accès Handicapés
 - ✓ Location Vélos
 - ✓ Villages fleuries/Villages de caractères...
- Systématiser l'utilisation du Machine Learning pour remplir les catégories