

# Analyse du Tourisme au Maroc via Reddit

## Projet NLP avec MongoDB et Streamlit

**Présenté par:** Ayman Naaimi  
**Encadré par:** Dr. Bouchra Frikh, PhD

ENSAM Meknès

# Objectif du Projet

Extraction et analyse des discussions touristiques sur Reddit

Extraire, nettoyer et analyser les discussions Reddit liées au **tourisme au Maroc**, pour comprendre :

- Les destinations les plus populaires
- Les sentiments des visiteurs
- Les thèmes touristiques évoqués (hébergement, gastronomie, activités...)

# Technologies utilisées

Stack technique complète

Composant	Outil utilisé
Scraping Reddit	PRAW (API Reddit)
Stockage	MongoDB
Nettoyage/NLP	pandas, TextBlob
Visualisation	Streamlit, Seaborn, Folium

# Étape 1 – Scraping Reddit

Collecte des données avec PRAW

- Utilisation de la bibliothèque praw (Reddit API)
- Génération de requêtes ciblées :
  - "travel to Marrakech"
  - "is Fes safe for tourists"
  - etc.
- Extraction **des posts ET des commentaires**
- Sauvegarde dans MongoDB avec un `_id` unique (évite les doublons)

## Résultat

Plus de 49 villes et villages analysés, avec 10 templates de requêtes chacun

# Étape 2 – Nettoyage des données

## Processus de filtrage et validation

- ❶ **Pré-nettoyage** : Suppression des doublons, fusion title + text → content
- ❷ **Détection de spam** (contains\_spam) :
  - Mots commerciaux (.com, buy now, http)
  - Textes trop courts
- ❸ **Filtrage final** :
  - Doit contenir au moins deux mot-clés touristique
  - Ex : desert, camel, medina, hotel, food

# Résultats du nettoyage

Réduction progressive du dataset

Étape	Nombre de lignes
Dataset brut	121866
Après spam filtering	38497
Après pertinence tourisme	18562

Réduction de 85% du volume initial, mais données plus pertinentes

# Étape 3 – Enrichissement des données

## Ajout de métadonnées et analyses

- **Sentiment analysis** avec TextBlob (Positif, Neutre, Négatif)
- **Classification thématique** :
  - Hébergement, Sécurité, Transport, etc.
- **Géolocalisation** : coordonnées par ville
- **Typologie** : Ville ou Village

### Exemple

Un commentaire sur "les magnifiques riads de Marrakech" serait classé :

- Sentiment : Positif
- Thème : Hébergement
- Localisation : Marrakech (31.6295°N, 7.9811°W)

# Étape 4 – Visualisation avec Streamlit

## Dashboard interactif

### Fonctionnalités :

- Filtres dynamiques :
  - Type de lieu, ville, sentiment, thème
- Visualisations interactives
- Carte folium

Interface intuitive pour explorer les données touristiques



# Carte interactive Folium

Visualisation géographique des discussions

- Regroupement des messages par ville
- Compteur de mentions avec latitude/longitude
- Identification des zones les plus actives
- Marqueurs proportionnels au volume de discussions

# Conclusion

- Nous avons collecté des données Reddit sur le tourisme au Maroc.
- Les données ont été nettoyées pour retirer le spam et les contenus non pertinents.
- Nous avons analysé les thèmes touristiques et les sentiments exprimés.
- Une application Streamlit a été développée pour visualiser les résultats.

**Ce projet montre comment les données sociales peuvent enrichir l'analyse touristique.**