

# Reunion Datacup Challenge:

## La bonne adresse

**ESI'Flipside :**

GONTHIER Jade, METAS Toan, RATALE Dimitri, VITRY Benjamin, VITRY Mateo

# Problématique

## Difficultés rencontrées :

- Existence de multiples listes d'adresses, souvent incohérentes.
- Données rarement contrôlées ou mises à jour.
- Dégradation progressive de la qualité des informations collectées.



**Comment garantir la fiabilité des bases d'adresses de tout service ?**

# Objectifs

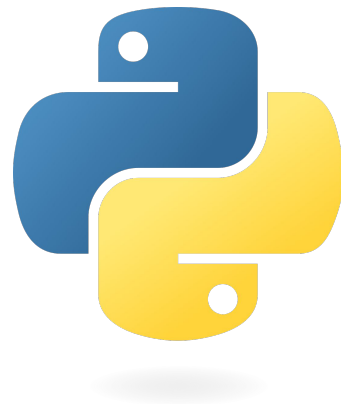


- Aider tout organisme ou collectivité à valoriser leurs bases d'adresses.
- Comparer les données sources à des référentiels fiables comme la **BAN**.
- Identifier les données incorrectes pour améliorer leur qualité.
- Proposer des corrections pertinentes et adaptées aux besoins des utilisateurs.



# Outils et applications

- Données BAN de la commune
- Python
- RapidFuzz
- Pandas
- Flask
- Dash



# Solution



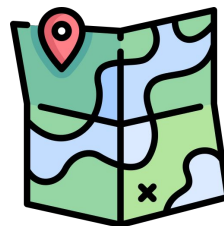
## Identification des erreurs:

- Orthographiques : Erreurs de frappe ou de transcription.
- Structurelles : Formats d'adresses non conformes.
- Géographiques : Localisations incorrectes ou incohérentes.

**Taux de précision élevé** : 85 % des suggestions de corrections sont pertinentes.

## Visualisation:

- Aide à la prise de décision.
- Les utilisateurs peuvent identifier les zones nécessitant des corrections.

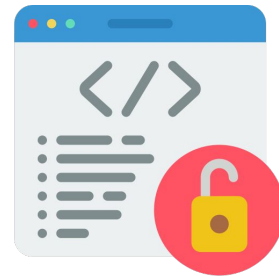


[Solution déployée](#)

# Conclusion

## Pérennisation:

- Réplicables pour toutes les structures utilisant des adresses.
- Outil open source tel que **Python**.



## Adaptabilité:

- Simple et intuitif, accessible à tous, non-spécialistes inclus.
- Peut gérer des volumes de données variés, petits ou massifs.

# Exemples de correction d'erreur

```
correct[correct["nom_voie"]=="CD 9 de la Montée Panon"]
```



0.0s

|     | id                 | id_fantoir | numero | rep | nom_voie                      | code_postal | code_insee | nom_commune           |
|-----|--------------------|------------|--------|-----|-------------------------------|-------------|------------|-----------------------|
| 146 | 97423_akwj91_00003 | NaN        | 3      | NaN | CD 9 de<br>la Montée<br>Panon | 97426       | 97423      | Les Trois-<br>Bassins |
| 147 | 97423_akwj91_00006 | NaN        | 6      | NaN | CD 9 de<br>la Montée<br>Panon | 97426       | 97423      | Les Trois-<br>Bassins |

```

correct["type_voie"]=correct["type_voie"].str.replace("CHEM", "CHEMIN", regex=False)
correct["type_voie"]=correct["type_voie"].str.replace("CD", "CHEMIN DEPARTEMENTAL", regex=False)
correct["type_voie"] = correct["type_voie"].str.replace("CHEMININ", "CHEMIN", regex=False)

```

| libelle_acheminement | type_voie               | nom_afnor                  |
|----------------------|-------------------------|----------------------------|
| LES TROIS BASSINS    | CHEMIN<br>DEPARTEMENTAL | 9 DE LA<br>MONTEE<br>PANON |
| LES TROIS BASSINS    | CHEMIN<br>DEPARTEMENTAL | 9 DE LA<br>MONTEE<br>PANON |
| LES TROIS BASSINS    | CHEMIN<br>DEPARTEMENTAL | 9 DE LA<br>MONTEE<br>PANON |
| LES TROIS BASSINS    | CHEMIN<br>DEPARTEMENTAL | 9 DE LA<br>MONTEE<br>PANON |