**Plan méthodologique pour répondre à la demande de la mairie de Montpellier sur l'utilisation des parkings voitures et vélos**

**Problématique :**

Vous et votre binôme avez été mandatés par la mairie de Montpellier afin d’étudier l’utilisation des parkings de la ville. Plusieurs aspects intéressent Monsieur le Maire, aussi fait-il appel à votre sagacité concernant les points suivants de sa politique de développement de la ville :

*⊲* Le taux d’occupation des parkings voiture *⊲* Le taux d’occupation des parkings vélos *⊲* Le bon fonctionnement du relais voiture / vélo.

1. **Collecte des données :**

* Récupération des données historiques d'occupation des parkings voitures et vélos sur l'année 2023, toutes les 5 minutes, sur le site open data de la mairie de Montpellier
* Récupération des données d'occupation en temps réel des mêmes parkings

1. **Préparation des données :**

* Nettoyage et vérification de la qualité des données collectées
* Formatage dans des tables structurées pour analyse (dates, horaires, taux d'occupation...)

1. **Analyses descriptives :**

* Calcul des taux d'occupation moyens pour chaque parking, toutes les 5 minutes, heures, jours, mois
* Identification des périodes de saturation (taux > 90%)
* Pour les parkings seule la saturation pose problème, pour les vélos on aura : saturation (impossibilité ou difficultés de déposer un vélo), mais aussi pénurie (absence de vélo à emprunter = 0 vélo dispo)
* Comparaison des évolutions selon les types de parkings et leur localisation

1. **Analyses explicatives**

* Corrélation entre taux d'occupation voitures et vélos selon les lieux et périodes
* Impact du tramway sur les taux d'occupation vélos loin des voitures
* Évolution des usages le WE / semaine

1. **Synthèse et recommandations**

* Rédaction d'une synthèse des résultats sous forme de dashboard interactif Graphiques commentés, tableaux de bord (Jupiter Notebook)
* Recommandations pour dimensionnement des parkings
* Mise à disposition sur Jupiter Notebook
* Stockage des données, codes et synthèse sur GitHub

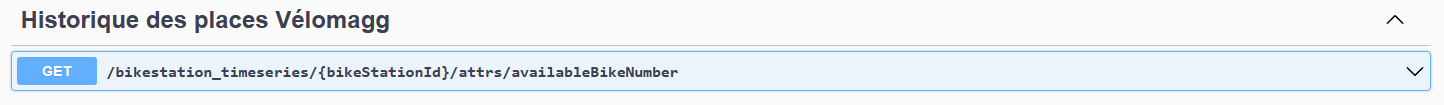
1. **Collecte des données :**

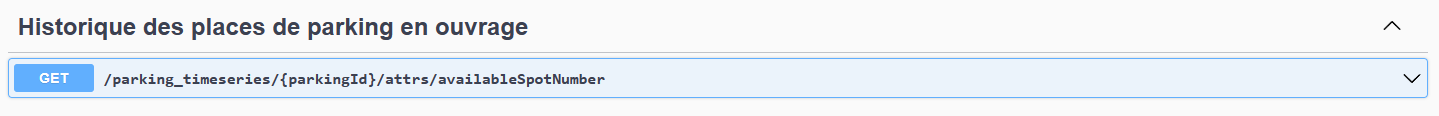
* **Sources :**

**Les donnée sont disponibles sur le site : Open Data Montpellier Méditerranée Métropole**

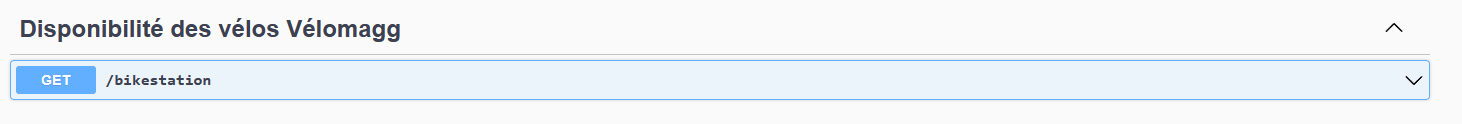
**Plusieurs API sont disponibles :**

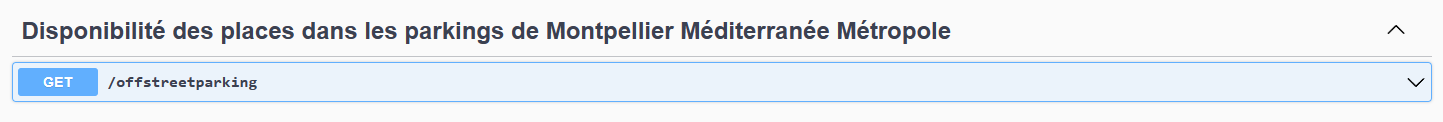
* Récupération des données historiques d'occupation des parkings voitures et vélos sur l'année 2023, toutes les 5 minutes, sur le site open data de la mairie de Montpellier





* Récupération des données d'occupation en temps réel des mêmes parkings

****

****

Compte tenu de l’objectif poursuivi : Analyse des taux d’occupation, les bases qui nous intéressent sont les données d’historique.

Les données en temps réels n’apportent pas d’information pertinente (en dehors de la valeur à un instant ‘t’) sur le taux d’occupation ou la pénurie des vélos au sein de la métropole de Montpellier.

Nous avons décidé de récupérer tous les ‘data’ des Parkings et Parcs à Vélo présents dans les bases de données de ’Open Data Montpellier Méditerranée Métropole’ pour l’année 2023 de façon à avoir une année complète.

Les bases de données ne disposent pas d’historique en deçà de 2023 (First recorded date at 2023-01-04).

On ne pourra donc pas établir des statistiques sur plusieurs années, ni faire une analyse des séries temporelles (étude saisonnalité et tendances).

* **bases Parkings :**

Les bases de données concernent 23 parkings, dont 19 sur la commune de Montpellier et 4 localisés sur les communes limitrophes (St Jean de Vedas (1) et Castelnau le Lez (3).

Sur les 23 parkings, 2 ne contiennent aucun historique sur la période étudiée (année 2023) et ne peuvent être traités :

* Gaumont Est
* Gaumont Ouest
* **bases Vélos :**

Les bases de données concernent 57 parcs à vélo, dont 56 sur la commune de Montpellier et 1 localisé sur une autre commune (Pérols).

* **Format des fichiers :**

Les données stockées sont au format JSON.

JSON est un format qui permet de stocker des informations structurées :

* Les données sont présentées sous forme de paires clé/valeur.
* ':' est placé entre chaque clé et chaque valeur
* Les éléments de données sont séparés par des virgules.
* Les crochets {} désignent les objets.
* Les crochets [] désignent des tableaux.
* {“key”:“value”,“key”:“value”,“key”:“value”.}
* **Structure des fichiers :**

**Chaque parking voiture ou parc vélo dispose d’un code unique, qui se présente de la façon suivante : « urn:ngsi-ld:parking:001 »**

**Chaque fichier parking contient les champs suivants :**

* attrName, str, 19 (nom de variable : availableSpotNumber, nombre de places disponibles)
* entityId, str, 23 (code parking)
* entityType, str, 16 (type : OffStreetParking)
* index, list, 8526 (timestamp)
* values, list, 8526 (nbr de places disponibles)

Timestamps constitue l’horodatage de la donnée (values), cette variable permet d’identifier la date et l’heure, elle se présente de la façon suivante : « 2023-01-04T10:50:51.000+00:00 »

## **Préparation des données :**

## 2.1 Extraction des données :

Collecte des métadonnées (description des variables, unités, période couverte, etc) :

L’extraction et le traitement des données a été réalisée à partir de programmes spécifiques développés sous Python.

* **Base Parking :**
* nombre d'enregistrements attendus : 365 jours x 24 heures x 60 mn / 5 mn (temps de réponse entre 2 enregistrements) = 105.120 enregistrements
* les limites d'extraction sont de 10.000 lignes par extraction, il faut donc au moins 11 extractions pour récupérer une base complète
* le plus simple c'est de générer une boucle avec une extraction par mois, avec un cumul dans une même variable (par data.extend())
* début historique : 2023-01-04T10:50 (toutes les données commencent le 4 janvier à 10h50)
* **Base Parc Velo :** 
  + nombre d'enregistrements attendus : 365 jours x 24 heures x 60 mn / 1 mn (temps de réponse entre 2 enregistrements) = 525.600 enregistrements
  + les limites d'extraction sont de 10.000 lignes par extraction, il faut donc au moins 53 extractions pour récupérer une base complète
  + le plus simple c'est de générer une boucle avec une extraction par jour, avec un cumul dans une même variable (par data.extend())
  + début historique : 2023-01-04T10:50 (toutes les données commencent le 4 janvier à 10h50)
* **extraction des données de l’API :**

url = "https://portail-api-data.montpellier3m.fr/parking\_timeseries/"+parking\_id+"/attrs/availableSpotNumber"

params = {

"fromDate": from\_date,

"toDate": to\_date

}

response = requests.get(url, params=params)

data = response.json()

* **cumul dans une base totale pour chaque parking:**

# Ajoutez les nouvelles valeurs des listes.

data\_cumul["index"].extend(data["index"])

data\_cumul["values"].extend(data["values"])

## Extraction des données :

**Vérification de la qualité des données :**

* Valeurs manquantes ou aberrantes : les disponibilités ne peuvent être > capacité, création d’un filtre :   
  **dataframe = dataframe[(dataframe['values'] <= capacite\_max)]**

Les fichiers par parking contiennent un nombre variable de jours (en moyenne 255 jours).

* Types et formats des variables : pour traiter plus facilement les données, celles-ci après avoir été chargées sont transférées du format JSON au format « dataframe » de la librairie Pandas sous Python.
* Cohérence des dates/horaires : le timstamp permet une codification cohérente des dates et des heures. Les délais entre 2 enregistrements et d’environ 5 mn pour les parkings et d’une minute pour les stations de vélos.

**Structure des données :**

* Création d'un fichier par parking ou (occupation en temps réel, historique)
* Définition des variables : identifiant du parking, date, heure, taux d'occupation

**Transformation des données :**

* Extraction des variables temporelles :mois, jour, heure
* Calcul de nouvelles variables dérivées : taux moyen par période, écart-type..

**Joindre/assembler les tables :**

* Association des données temps réel et historique
* Fusion éventuelle des tables parkings voitures et vélos
* Export des données nettoyées et structurées dans un format exploitable pour les analyses (dataframe)

## **Analyses descriptives :**

**Calcul des taux d'occupation moyens :**

* Moyenne générale pour chaque parking
* Moyenne par période de 5 minutes
* Moyenne horaire
* Moyenne journalière
* Moyenne mensuelle

**Analyse des variations :**

* Écart-type du taux d'occupation
* Minimum, 1er quartile, médiane, 3ème quartile, maximum

**Identification des périodes de forte affluence :**

* Taux supérieur ou égal à 90%
* Nombre d'heures/jours avec taux élevé

**Comparaison entre parkings :**

* Diagrammes de Tukey pour groupes
* Tests statistiques de comparaison (t-test, ANOVA)

**Analyse selon la localisation :**

* Cartographie du niveau moyen d’occupation par parking (min, moy, max)
* Cartographie des seuils dépassés par parking (nb heures > seuil)
* Corrélation avec facteurs géographiques (population, transports...)

**Évolution temporelle :**

* Séries chronologiques
* Décomposition par tendance, saisonnalité, irrégularité

**Analyse par type de jour :**

* Week-end / semaine
* Jours fériés / ouvrés

## **Analyses explicatives**

**Corrélation taux occupation voitures/vélos :**

* Calcul des coefficients de corrélation
* Analyse par périodes (heures, jours, mois)
* Selon la localisation (proximité parking voiture/vélo)

**Impact du réseau de transport :**

* Comparaison taux vélos proches/éloignés du tramway
* Recherche de ruptures/seuils dans les corrélations

**Modélisation prédictive :**

* Régression linéaire multiple
* Autres algorithmes (forêts aléatoires, etc)
* Variable à prédire : taux vélo
* Prédicteurs : taux voiture, zone, météo

**Analyse des profils d'usage :**

* Classification (k-means, hiérarchique)
* Description des groupes
* Corrélation avec caract. sociodémographiques

**Interprétation des résultats :**

* Mise en perspectives avec la littérature
* Limites et biais des analyses

**Recommandations opérationnelles :**

* Dimensionnement des parkings
* Actions de communication intermodale

## **Synthèse et recommandations**

**Rédaction d'un rapport de synthèse :**

* Description du contexte et objectifs
* Méthodologie employée
* Principaux résultats descriptifs et explicatifs
* Limites des analyses
* Perspectives

**Création d'un dashboard interactif :**

* Interface ergonomique
* Navigation entre jeux de données
* Visualisations dynamiques (cartes, graphiques)

**Mise en exergue des résultats clés :**

* Taux d'occupation moyens et maximaux
* Corrélations entre voitures et vélos
* Impact du tramway sur les vélos

**Formulation de recommandations :**

* Dimensionnement des parkings à revoir
* Communication intermodale à promouvoir
* Pistes d'aménagement des réseaux de transports

**Mise à disposition :**

* Hébergement du rapport et dashboard
* Documentation des codes et jeux de données
* Remise au commanditaire

**Valorisation des travaux :**

* Présentation orale
* Publication d'un article de synthèse