

2018  
2019

# Analyses et cartographie des données OpenAgenda

MASTER GEOGRAPHIE NUMERIQUE  
MODULE 1F2 : PROJET GEOMATIQUE

Réalisé par :  
*Bénédicte Thomas*  
*Salomé de Foville*  
*Thomas Mabire*  
*Axel Fourneyron*



# Table des matières

|  |    |
|--|----|
| Introduction.....                              | 2  |
| 1 Structuration des données.....               | 2  |
| 1.1 Maîtrise de l'API.....                     | 3  |
| 1.2 Export.....                                | 5  |
| 1.3 Nettoyer les données.....                  | 6  |
| 2 Méthode et pistes exploratoires.....         | 8  |
| 2.1 Définition pistes d'étude.....             | 8  |
| 2.2 Données croisées.....                      | 8  |
| 3 Résultats.....                               | 10 |
| 3.1 Présentation des résultats.....            | 10 |
| 3.1.1 Des statistiques.....                    | 10 |
| 3.1.2 Des cartes interactives.....             | 12 |
| 3.1.3 Des analyses spatiales.....              | 13 |
| 3.1.4 Des cartes.....                          | 15 |
| 3.1.5 Des cartes générées automatiquement..... | 16 |
| 3.2 Proposition d'amélioration.....            | 19 |
| 3.3 Proposition d'outils.....                  | 21 |
| 3.3.1 Logiciels SIG classiques.....            | 21 |
| 3.3.2 Webmapping.....                          | 22 |
| 3.3.3 Logiciel Textométrique :.....            | 23 |
| Conclusion.....                                | 23 |
| ANNEXE page 24 - 52                            |    |

## Introduction

OpenAgenda est une start-up créée en 2011, conçue pour centraliser et partager des événements dans le monde entier. Cette plateforme en ligne stocke les événements à venir et sous forme d'archive les événements passés qui ont été recensés sur le site OpenAgenda. Cette base de données est accessible par tout utilisateur après s'être au préalable créé un compte qui délivrera une clé unique, demandée lors de l'extraction de données sur l'API mise à disposition sur leur site. OpenAgenda stocke les données sous forme de base de données spatiale : les données sont spatialisées dans l'espace et dans le temps. Cette base ne cesse de s'accroître avec plus de 16 000 agendas intégrés et 428 000 événements enregistrés. Les fondateurs d'OpenAgenda nous ont donc ouvert leur base avec un objectif en tête : Faire parler leur base de données. Actuellement, aucune analyse statistique, aucune exploitation de la base de données n'avaient été réalisées.

OpenAgenda a donc fait appel à nous pour faire parler ses données aux vues du nombre d'information qui deviennent de plus en plus importante (420 000). Le but est ici, non pas de s'intéresser au profil des utilisateurs d'OpenAgenda mais bien d'exploiter au maximum la base de données créée avec les informations stockées depuis 2011 pour démontrer les bénéfices que les acteurs publics peuvent avoir à travailler avec cette base de données pour répondre à diverses problématiques socio-économiques, d'aménagement, de protection civile et d'accessibilité par exemple. Nous avons défini, pour répondre à la commande de notre commanditaire une zone d'étude qui est Paris. Notre projet s'est conçu autour de la problématique suivante :

---

### *« En quoi les acteurs publics peuvent tirer bénéfice de la base de données OpenAgenda ? »*

---

Dans un premier temps, nous étudierons la question de l'acquisition et structuration des données issues de l'API d'OpenAgenda puis, nous présenterons nos pistes exploratoires retenues dans ce projet avec l'explication des méthodes utilisées pour les appliquer et enfin nous détaillerons les résultats obtenus et proposerons divers outils pour permettre à cette start up de continuer l'analyse statistiques en parallèle du développement de la base.

Tout d'abord définissons les termes du sujet : Dans notre étude est considéré comme événement tout genre de rendez-vous publics.

Représenter les données dans un système d'informations géographique permet de faire un premier état des lieux de la spatialisation des événements selon certains phénomènes. Avec les différents logiciels qui nous ont permis de visualiser ces données cela permettent d'obtenir un premier coup d'œil de la spatialisation des événements et de représenter plusieurs autres variables pour une analyse croisée de l'information. Il s'agit ici de rechercher des relations spatiales entre les événements. Les SIG ne sont pas encore exploités par OpenAgenda mais il s'agit d'un outil puissant qui va permettre d'analyser et faire parler leur base de données spatiale.

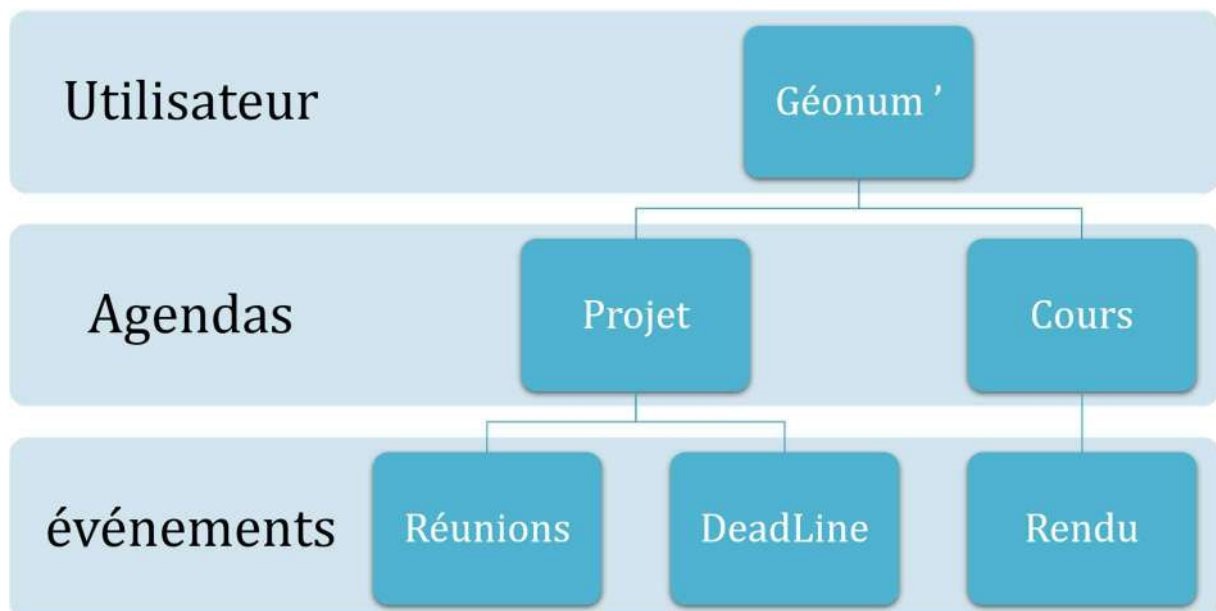
Il a fallu définir les résultats attendus, les données mobilisées, les traitements, logiciels à mobiliser, décomposer en étapes notre travail s'assurer de la faisabilité de nos étapes et de nos temps de travail pour ne pas être pris de cours lors des difficultés qui pourraient se présenter. Nous avons décomposé et répartis nos tâches de manière visuelle sur le site gant et Excel cf ( Annexe :1)

## 1 Structuration des données

### 1.1 Maîtrise de l'API

OpenAgenda partage des événements dans le monde entier grâce à son site internet. Afin, de diffuser les événements le site web d'OpenAgenda stocke les données et les rend accessibles à tout utilisateur ayant possession d'un compte gratuit OpenAgenda, grâce à un API. Pour mieux comprendre comment ces données sont partagées et transmises par l'API nous avons d'abord cherché à comprendre le fonctionnement général du site web, puis avons trouvé des méthodes d'exportation des données.

Pour cela nous avons créé un organigramme qui permet de restituer simplement le fonctionnement général de la plateforme :



Différentes données sont créées sur le site web. Il faut d'abord des données utilisateur qui vont permettre la création d'agendas. Ils peuvent ainsi ajouter à la base de données un ou plusieurs agendas. Une fois l'agenda créé par un utilisateur il sera possible d'y ajouter différents événements au sein de l'agenda. Il est également possible pour un utilisateur d'ajouter un événement issu d'un autre agenda d'un autre contributeur dans son propre agenda.

Ainsi se résume le fonctionnement du site web d'OpenAgenda en ce qui concerne l'ajout de données agendas/événements. Voyons maintenant comment est défini un événement. Quels attributs les définissent ? Nous avons réalisé un schéma recensant les attributs des événements :

# événements

## ID

- Identifiant événements et agenda
- Origine

## Textes descriptifs

- Titres
- Descriptions (Courtes, Longues)
- Détails des conditions (Prix ou restrictions)
- Descriptions HTML
- Mots clés, catégorie, Tags, Propositions
- Descriptifs lieux

## Horaire

- Last Update
- Horaires détaillés et ISO - Résumé horaire
- Première/dernière date - Première/Dernière ouverture/fermeture

## Liens extérieurs

- Lien événement
- Image – aperçus images – images sources

## Données Géographiques

- Pays, Région, Département, Arrondissement, Ville, Code postal, Adresse, Nom du lieu,
- Latitude et Longitude (en WGS)

Les métadonnées pour décrire un événement sont approfondies et complètes :

- Des champs uniques pour identifier, les événements et leurs agendas,
- Des champs de type texte pour exposer les événements de multiples façons,
- Des champs de type dates, pour tout connaître de la durée, du début, de la fermeture, des répétitions, ...
- Des champs URL, pour permettre la redirection vers diverses adresses URL : un site web, images ou bien encore un lien d'un réseau social par exemple.
- Des champs géographiques, pour situer précisément l'événements.

L'ensemble des informations permettant de décrire un événement sont possibles, cependant nous avons remarqué quelques problèmes concernant la qualité des données brutes récupérées, qui peut créer des soucis à Open Agenda pour l'exploitation de leurs données.

Les principales problématiques sont liées au fait que l'utilisateur lors de la création d'événements est confronté à des champs qui ne sont pas bornés sous forme de cases à cocher ou listes déroulantes mais sous forme de zone de texte ou tout le monde décrit son événement de manière non homogène. Il n'y pas d'uniformisation des données. Par exemple les prix pourraient être deux champs distincts avec heure de début et heure de fin, ou bien pareil pour le tarif mettre une liste déroulante « Evènement payant » et évènement libre » renvoyant vers le site web de l'évènement pour les détails des tarifs si tel est le cas. Parfois la géolocalisation indique le même emplacement (longitude, latitude) à différente adresse (manque de précision de l'utilisateur) pour des adresses différentes... De plus la localisation n'est pas toujours renseignée ou mal fait...

Pour les Mots-clés, catégories, tags, propositions : L'ensemble de ces champs ne sont pas dynamiques, il ne dépend d'aucune liste et tout le monde est libre d'ajouter ce qu'il veut. Ce qui crée d'importantes difficultés à différencier les événements. La catégorisation des événements est pourtant un point intéressant dans l'analyse, notamment sur l'organisation spatiale de certain type d'événements.

Ce sont les principaux problèmes liés à la qualité de la donnée qui peut perturber une analyse car c'est un point essentiel pour les géomaticiens afin de mener une analyse spatiale et temporelle pour comprendre un phénomène sur un territoire. Nous avons essayé de proposer des solutions simples pour améliorer ces soucis (Partie 3.b. Proposition d'amélioration et 3.C Proposition d'outils). En effet, pour faire parler les données et les représenter spatialement ou même les faire parler statistiquement il est nécessaire de centraliser les données dans une base de données structurée pour permettre de « la gérer, interroger, analyser, visualiser et transmettre les données » (Cours T. Joliveau « Théorie et pratique de la Géomatique cours 1 » dans « SIG premier tour d'horizon »). Les attributs de la base de données sont à la fois des données numériques pour certaines et pour d'autres des données dites chaîne de caractères mais doivent être atomiques, exhaustives et non redondantes.

Pour réaliser nos différentes analyses afin de révéler le potentiel des données d'OpenAgenda, nous avons, exporté un échantillon des données, puis l'avons nettoyé pour montrer le potentiel d'analyse avec des données de meilleures qualités.

## 1.2 Export

Open Agenda met à disposition un API qui permet de télécharger ses données. Il y a 2 versions de l'API, ayant un fonctionnement différent.



### L'ancienne version :

Elle passe uniquement par des requêtes URL qui interrogent l'ensemble des événements dans la base, et qui renvoie les 100 premiers événements correspondant à certains critères à travers un fichier .JSON.

Cette méthode est avantageuse pour parcourir et récupérer l'ensemble des événements souhaités. Mais nous n'avons pas retenu cette méthode, car les critères se limitent seulement en une gestion de date et de géographie peu précise. (Radian par rapport à des coordonnées X,Y)

```
https://api.openagenda.com/v1/events?key=14b3247deebb42628cca0ca9ee9b036f
&oaq[category]=concerts
&oaq[what]=rock
&oaq[passed]=1
```

## Nouvelle version :

Celle-ci permet à travers un agenda de télécharger l'ensemble ou seulement une partie des événements qu'il contient. Possibilité de télécharger seulement les événements à venir ou la totalité, et possibilité de filtrer les événements grâce à un mot-clé. Exemple : Au sein de l'Agenda fête de la musique, chercher les événements « Rock ».



De plus cette version permet un téléchargement des données à différents formats : google calendar, Ical, ICS, RSS, CSV, xlsx, PDF, JSON.

Dans l'objectif de récupérer un jeu de donnée suffisamment représentatif des données OpenAgenda nous nous sommes d'abord concentrés sur la région Île-de-France qui recense plus de 50 000 événements (soit 1/10ème des données). Une part importante des données de cet agenda est centrée sur Paris, nous avons focalisé la majeure partie de nos analyses sur cette ville.

Cependant il faut avoir en tête que Paris s'ancre dans un contexte bien particulier qui n'est pas similaire aux autres territoires, il faut donc avoir conscience de cette différence avant de reproduire ces analyses sur un autre espace.

### 1.3 Nettoyer les données

Pour des analyses pertinentes et cohérentes nous avons souhaité nettoyer le jeu de données de l'Île de France. La quantité de données rend la tâche difficile à automatiser.

Dans un premier temps nous avons voulu essayer de créer des catégories parlantes par rapport aux différentes données dans l'agenda, nous avons pensé à un code python qui permettrait à partir de mots clés trouvés dans la colonne description d'associer une catégorie d'événement pour le champ « Type d'événement » mais comme chaque utilisateur définit un même type d'événement de manière très différente avec de multiples mots clés cette tâche est extrêmement difficile. (Par exemple un concert peut être décrit par un mot clé « concert », « musique », « dj », « fête musicale »).

Nous avons donc rattaché différents mots clés que nous recherchions sur l'agenda qui recensait les événements parisiens pour recueillir toutes les bases de données pouvant faire partie d'une catégorie. L'ensemble des mots-clés rattachés à leurs futurs types d'événements figurent ci-dessous :

## Spectacles

- Théâtre(s), danse(s), feu d'artifice(s), bal(s), karaoké(s), carnaval(s), opéra(s)

## Expositions

- Vernissage(s), exposition temporaire(s), musée(s), patrimoine(s), visite guidée(s)

## Concert

- Music(s), musique(s), chant(s), fanfare(s), festival(s), Garelive(s)

## Religion

- Messe(s), procession(s), communion(s), baptême(s), pèlerinage(s), catéchisme(s), événements diocésain(s), nativité, sanctuaire(s)

## Divers

- Autres événements

Méthode :

1. Aller sur l'agenda Ile de France
2. Inclure tous les événements passés
3. Filtrer avec des mots-clés
4. Télécharger le fichier au format CSV
5. Ouvrir le fichier
6. Créer la colonne catégorie
7. Puis appliquer la catégorie liée à la recherche pour tous les événements
8. Combiner l'ensemble des fichiers

Grâce à cette méthode nous avons pu à la fois récupérer l'intégralité des événements de Paris, et définir leur type. Cependant nous avons créé de multiples doublons. Il a donc fallu procéder à un nettoyage de cette base de données grâce à une suppression des doublons par rapport au champ ID de l'événement. Cette méthode nous a permis d'améliorer un peu la qualité des données, mais surtout de réaliser des analyses en fonction du type d'événement.

Dans un deuxième temps, toujours dans l'objectif d'améliorer les données nous avons souhaité reformater le champ "prix" afin d'obtenir un champ seulement numérique. Cette fois nous avons utilisé plusieurs formules Calc (Excel marche aussi).

Une formule utilisant la fonction "STXT" qui permet d'observer une chaîne de caractères et d'extraire de cette dernière seulement le nombre présent. Cette méthode permet facilement de ressortir les prix des événements mais elle n'est pas infaillible, certaines chaînes de caractères comportent des données erronées, par exemple "interdit aux moins de 18" ressort avec un prix de 18 euros. Cependant après une vérification seulement quelques erreurs de ce type ont été trouvées et corrigées.

Une formule utilisant la fonction CHERCHE qui permet de savoir si un mot est présent dans une chaîne de caractères. L'objectif était de détecter les événements "Gratuit" ou "Prix libre".

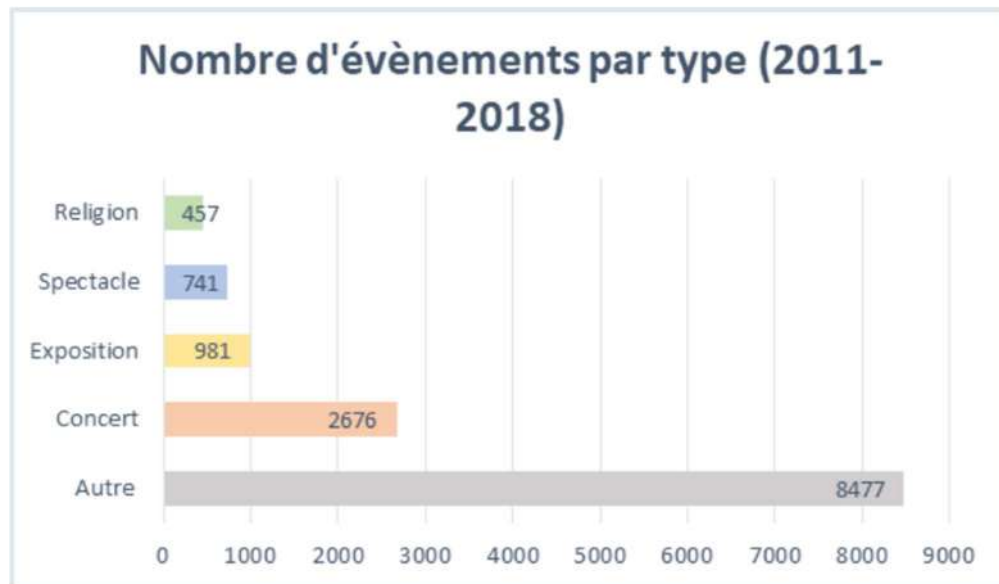
Enfin pour les champs ne contenant pas d'indications de prix nous avons attribué le mot ND (Non Dévoilé) afin de ne pas biaiser les données et l'analyse.

Dans un troisième temps pour nettoyer l'intégralité de la base de données nous avons aussi :

Trier par département et ville afin d'avoir uniquement les événements de Paris et supprimer les événements qui s'y trouvent en dehors. Attribuer correctement l'arrondissement en fonction du code postal (Fonction DROITE).



Ainsi grâce à ces étapes qui ont nécessité une grande partie de notre temps consacré au projet, nous avons pu observer la dispersion suivante entre nos cinq catégories :



Le type « Autre » est malheureusement celui qui concentre le plus d'évènements malgré notre volonté de catégorisation des évènements.

## 2 Méthode et piste exploratoires

### 2.1 Définition pistes d'études

Après avoir restructuré notre base de données, il nous fallait donc définir nos différentes pistes d'études. Un grand nombre d'analyses statistiques sont possibles, mais il nous fallait évidemment choisir et définir celles que nous voulions explorer. Nous avons donc choisi de faire des analyses chronologiques, saisonnière et journalière, mais aussi selon la typologie et l'intensité événementielle. Nous avons également choisi comme piste d'étude différentes analyses basées sur la géolocalisation des évènements, ainsi qu'une analyse basée sur leurs prix.

Nous avons choisi ces différentes pistes d'études de notre base de données afin de présenter différents aspects de celle-ci, en nous focalisant sur les différentes données comprises dans celle-ci, mais également selon les représentations spatiales possible, que cela soit selon la chronologie ou la localisation des évènements.

### 2.2 Données croisées

Afin de pousser nos analyses, il était important de croiser notre base de données avec des données extérieures. En particulier, les analyses basées sur la géolocalisation des évènements nécessitaient des données extérieures afin d'aboutir. Nous avons donc exploité des données descriptives de la population, tel que le revenu médian et la densité de population, des données sur les services publics, tel que les transports en communs et l'éclairage public, ainsi que des données sur les espaces, nous avons eu la chance de prendre Paris comme zone d'études car de nombreuses données sont libre d'accès grâce aux sites :

<https://opendata.paris.fr/explore/?sort=modified>

<https://data.iledefrance.fr/explore/?sort=modified>

Nos différentes pistes d'étude sont donc :

- Analyse chronologique : saisonnière et journalière
- Typologie et intensité événementielle étude chronologique (nombre d'événements selon les saisons)
- Correspondances entre la géolocalisation des événements et type d'événements
- Correspondances entre la géolocalisation des événements nocturnes et l'éclairage publics
- Correspondances entre la géolocalisation des événements et transports en communs (les événements se situent-ils à moins de 250-500m et 1km des stations de métros et RER)
- Taxonomie des événements et prix du m<sup>2</sup> (événements selon prix aux mètres carrés)

Pour répondre au commanditaire, nous manipulons des données spatiales au format vecteur car il s'agit du mode qui nous permettra de représenter facilement une information pour la communiquer sous forme de carte au grand public et acteurs publics de la mairie de Paris.

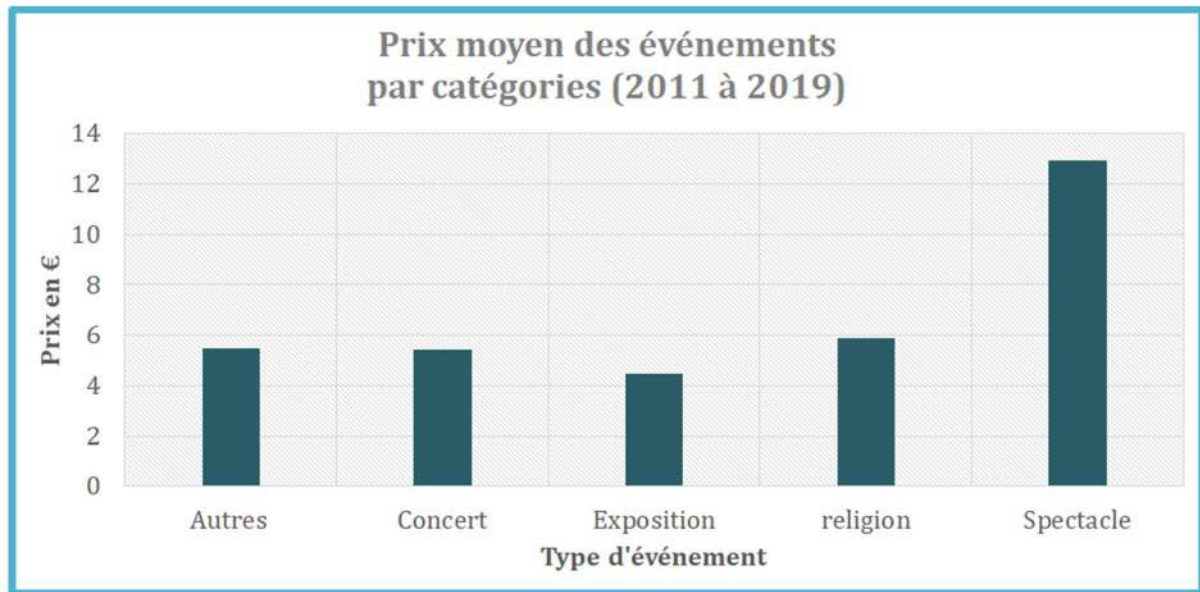
### 3 Résultats et proposition

#### 3.1 Présentation des résultats.

Nous avons effectué diverses analyses afin de faire parler aux mieux nos données concernant les évènements situés à Paris proposé par OpenAgenda, en optant sur plusieurs angles d'approches.

##### 3.1.1 Des statistiques.

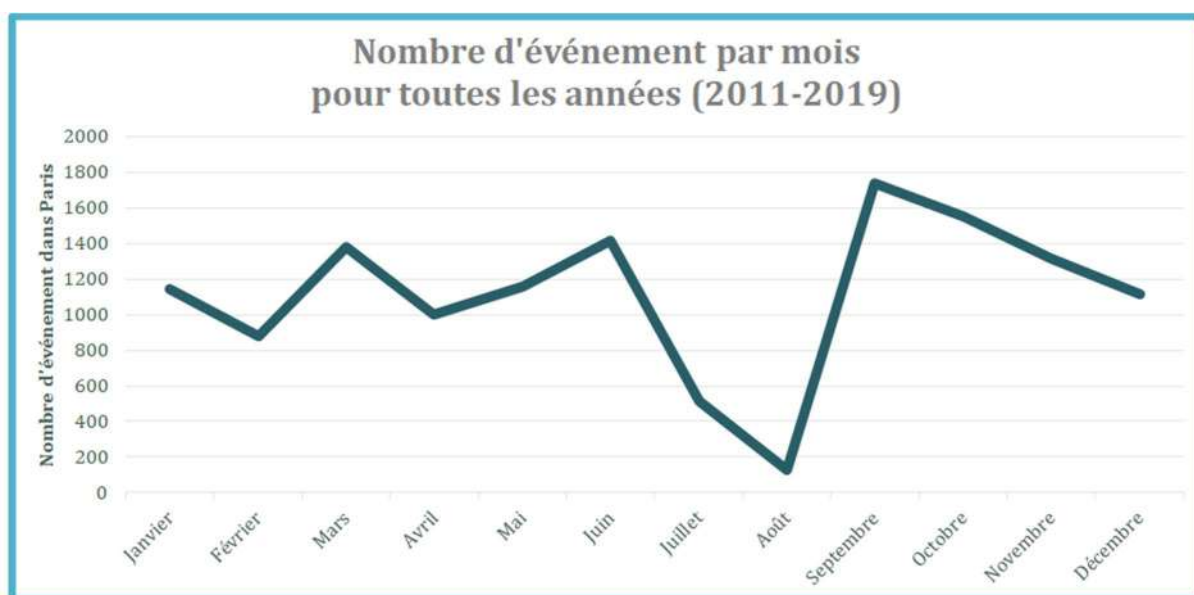
Nous avons fait des statistiques descriptives des évènements proposé pour observer comment se comportent les données :



Deux choses notables ressortent du graphique. Le prix moyen des spectacles est d'environ le double de nos autres catégories. Cela est certainement dû en partie au fait que dans spectacle il y a aussi des formations qui se sont glissées, les formations aux arts du théâtre par exemple. La seconde chose est que les évènements religieux ont un prix, deux choses peuvent l'expliquer des erreurs d'encodages et vu que nous les avons récupérése par une procédure automatisée, nous ne l'avons pas détecté et aussi certains évènements religieux se base sur un système d'offrande pour certaines manifestations religieuses.



Ce graphique sur le nombre d'événement annuel, montre clairement une augmentation du nombre de contributeurs sur OpenAgenda organisant des évènements à Paris et cela à partir de 2014, il est évident que cela ne reflète pas une évolution du nombre d'évènements dans cette ville mais une expansion d'OpenAgenda à Paris.



Ce dernier graphique permet d'observer lui un creux dans la période estivale, les contributeurs d'OpenAgenda réalise donc leurs évènements en dehors de cette période particulière de l'année, où Paris voit un nombre conséquent de ces habitants partir en vacances. Notons un pic d'évènements lors du mois de septembre, soit la rentrée scolaire et la fin des vacances estivales. Les évènements de proximité sont donc de mises.

### 3.1.2 Des cartes interactives

Nous avons réalisé plusieurs géovisualisations avec CARTO via des approches multiples. En effet, notre commanditaire était particulièrement intéressé par ce type rendu et surtout la première carte dynamique présentée, qui est très proche de sa demande initiale de webmapping.

Nous avons créé une carte interactive sur l'emplacement événements accompagné de choix sélectionnables pour l'apparition des événements sur la carte. En effet il est possible de choisir la date (mois et année dissocié), les lieux des événements, leurs catégories et leurs tranches de prix. Il est donc possible d'observer la répartition des événements avec une approche plus ou moins ciblé en fonction des critères sélectionnés.

<https://fourneyrona.carto.com/builder/769f84ca-5b46-42cb-8de5-03822b7e1f08/embed>

De plus, nous avons croisé les données d'OpenAgenda avec d'autres afin de détecter d'éventuelles corrélations. Cette carte dynamique concerne les événements parisiens catégorisés par leurs types croisés avec comme fond de carte la densité de population des IRIS, en 2015. Afin d'observer si la densité joue sur la présence d'événements et de voir si fonction de leurs catégories cela a plus ou moins d'impact.

<https://thomasm69.carto.com/builder/ef466784-75c6-43ac-a685-49551c5a4f8d/embed>

Une autre carte concernant cette fois un événement précis répertorié sur OpenAgenda à savoir la Fête de la Musique en 2015 que l'on a croisé avec le prix au mètre carré de l'immobilier en 2015, afin d'observer si le prix de l'immobilier avait un impact sur l'emplacement des concerts d'un événement d'ampleur, tel la Fête de la Musique.

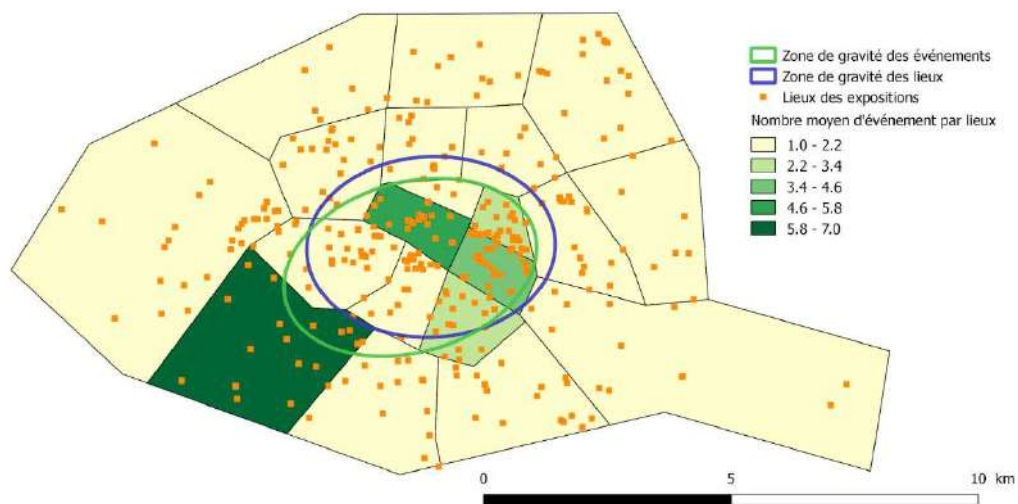
<https://thomasm69.carto.com/builder/3882d5e5-c392-4563-9ae6-2dfa7f05ed90/embed>

### 3.1.3 Des analyses spatiales

Elles ont pour but de mieux appréhender la dispersion des événements en fonction de leurs de catégories. Nous avons mené cette étude à la fois sur les lieux et le nombre d'évènements en fonction de leurs catégories. Pour cela nous avons calculé leurs ellipses de dispersion réciproques et avons ajouté en fond de carte le nombre d'évènements par lieux en moyenne dans chaque arrondissement.

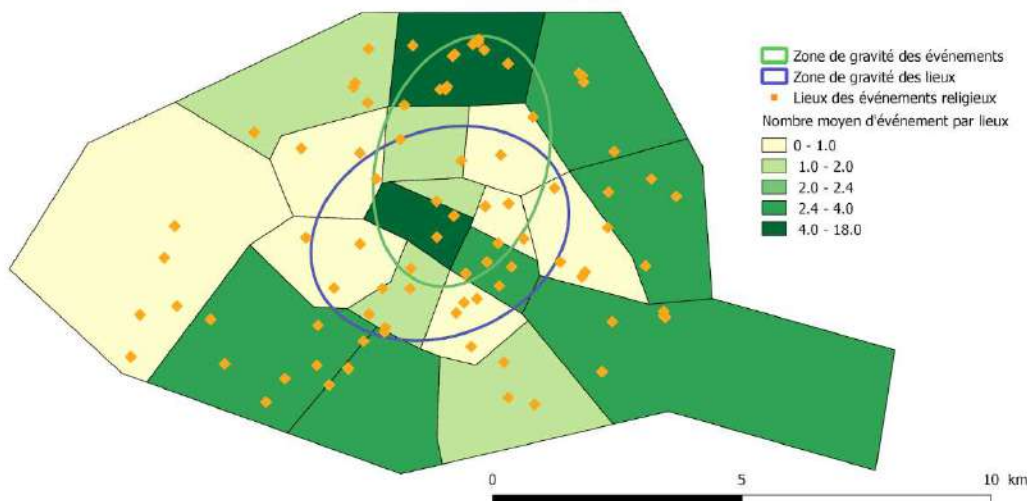
Ces ellipses correspondent à l'aire d'attraction des événements.

## Description de la dispersion des événements et des lieux accueillants les expositions



Pour la catégorie des expositions, nous observons que celles-ci sont concentrés dans le centre de Paris ce qui explique la position centrale des deux ellipses. Mais il y a une exception, le XV arrondissement au Sud-Ouest, qui a le nombre moyen d'événements par lieux le plus conséquent, ce qui explique l'ellipse de dispersion du nombre des événements soit plus orienté et décalé vers cet arrondissement.

## Description de la dispersion des événements et des lieux accueillants les événements religieux



Pour la catégorie religion, cela nous permet de comprendre rapidement que les nombres événements religieux par lieux et leurs localisations ont un emplacement différent de leurs ellipses de dispersion réciproques. En effet celle des lieux est plus centrale que celle des événements qui est bien plus au Nord dû à un nombre important d'événements religieux dans quelques lieux de notre zone d'étude, comme on peut le voir grâce à l'aplatissement de couleur du XVIII arrondissement dans la partie nord de Paris.

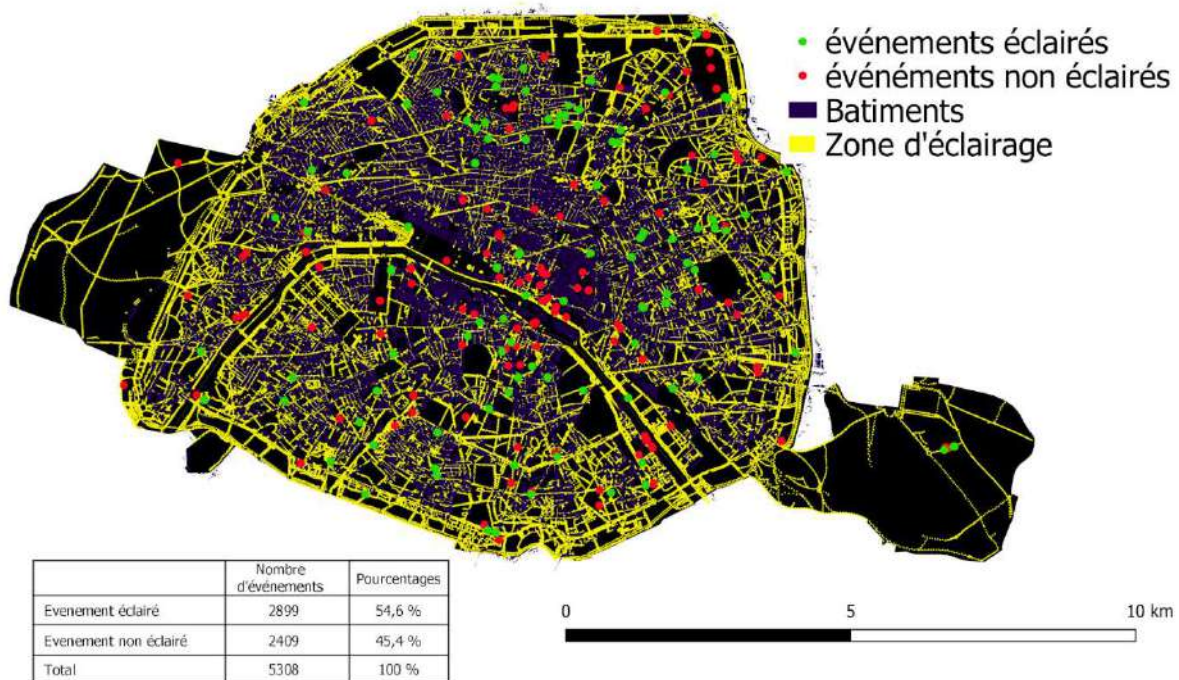
De plus, l'ellipse de dispersion des lieux est de taille assez importante, dû à une répartition de la localisation relativement homogène des lieux des événements religieux sur l'ensemble de la ville.

### 3.1.4 Des cartes

Etant donnée l'objet de notre projet, nous avons pu réaliser des cartes aux approches et thématiques variés :

Les événements selon leurs localisations à proximité ou non de l'éclairage public.

## L'éclairages des événements extérieures après 20h

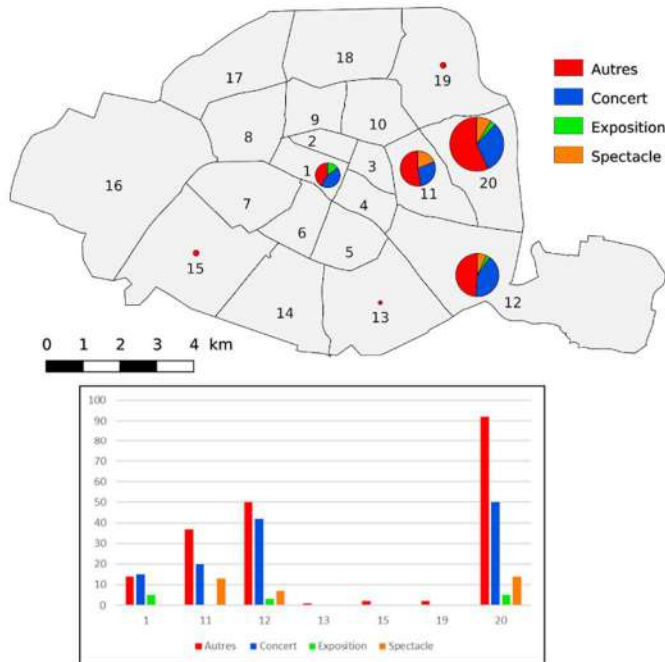


Cela nous permet d'observer que la répartition des événements extérieurs dans une zone d'éclairage public ou non (soit à proximité d'un lampadaire) est presque semblable, cela ne semble donc pas être un critère déterminant pour le choix de l'emplacement des événements nocturnes extérieurs. De plus, on peut remarquer que de nombreux événements situés dans l'hypercentre parisien sont en dehors de la zone d'éclairage public.

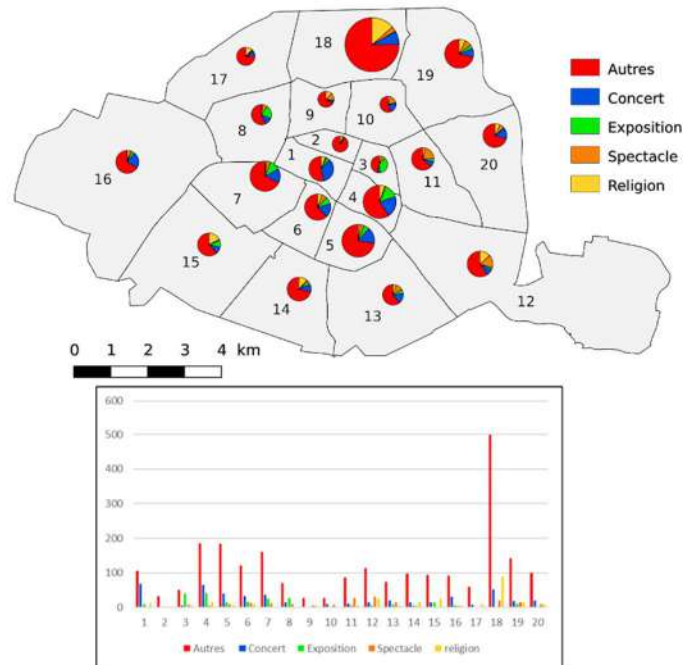


Des cartes accompagnées chacune d'un graphique pour permettre d'observer l'évolution des évènements et leurs catégories en fonction de chaque arrondissement entre 2011 et 2018.

Types et nombres d'évènements à Paris en 2011



Types et nombres d'évènements à Paris en 2018



Cette comparaison temporelle de la base d'OpenAgenda permet de comprendre l'évolution des contributeurs et des offres proposées par ceux-ci sur Paris. Ainsi on peut voir que lors de sa création OpenAgenda avait peu de contributeurs et très peu d'arrondissements qui comportaient des évènements. Le XX arrondissement, représentait une part importante d'évènements parisiens. A l'exception de la catégorie "autres", les concerts sont l'essentiel des évènements présent à Paris.

En 2018, il y a une augmentation des évènements ayant lieu à Paris, de plus ceux-ci sont désormais présents dans l'ensemble des arrondissements. Une diversité dans les différents types d'évènements, dans le centre on peut observer un nombre important d'exposition, les concerts sont relativement dispersés entre tous les arrondissements. Enfin, le XVIII arrondissement, comprend un nombre bien plus conséquent d'évènements que les autres arrondissements et cela est encore plus accentué pour la catégorie des évènements religieux.

### 3.1.5 Des cartes générées automatiquement.

Nous avons programmé deux codes python pour obtenir : (nous remercions YanKiu Mok et Léonie Gonon nos binômes d'algorithme qui se sont associées pour cette partie au projet)

- Une carte générée automatiquement des arrondissements parisiens avec l'emplacement des évènements s'y déroulant catégorisés en fonction de leurs types.

- Une carte générée automatiquement d'un département français, soit en dehors de notre zone d'étude qui suite à l'entrée par l'utilisateur d'un prix maximum permettra d'afficher sur la carte uniquement les événements ayant un prix inférieur à celui indiqué

Les codes sources Python sont disponibles page 46-51 et un exemple de carte générée par ces codes est disponible en annexe page 52.

Par ailleurs, nous avons construit un model graphique sous qgis.

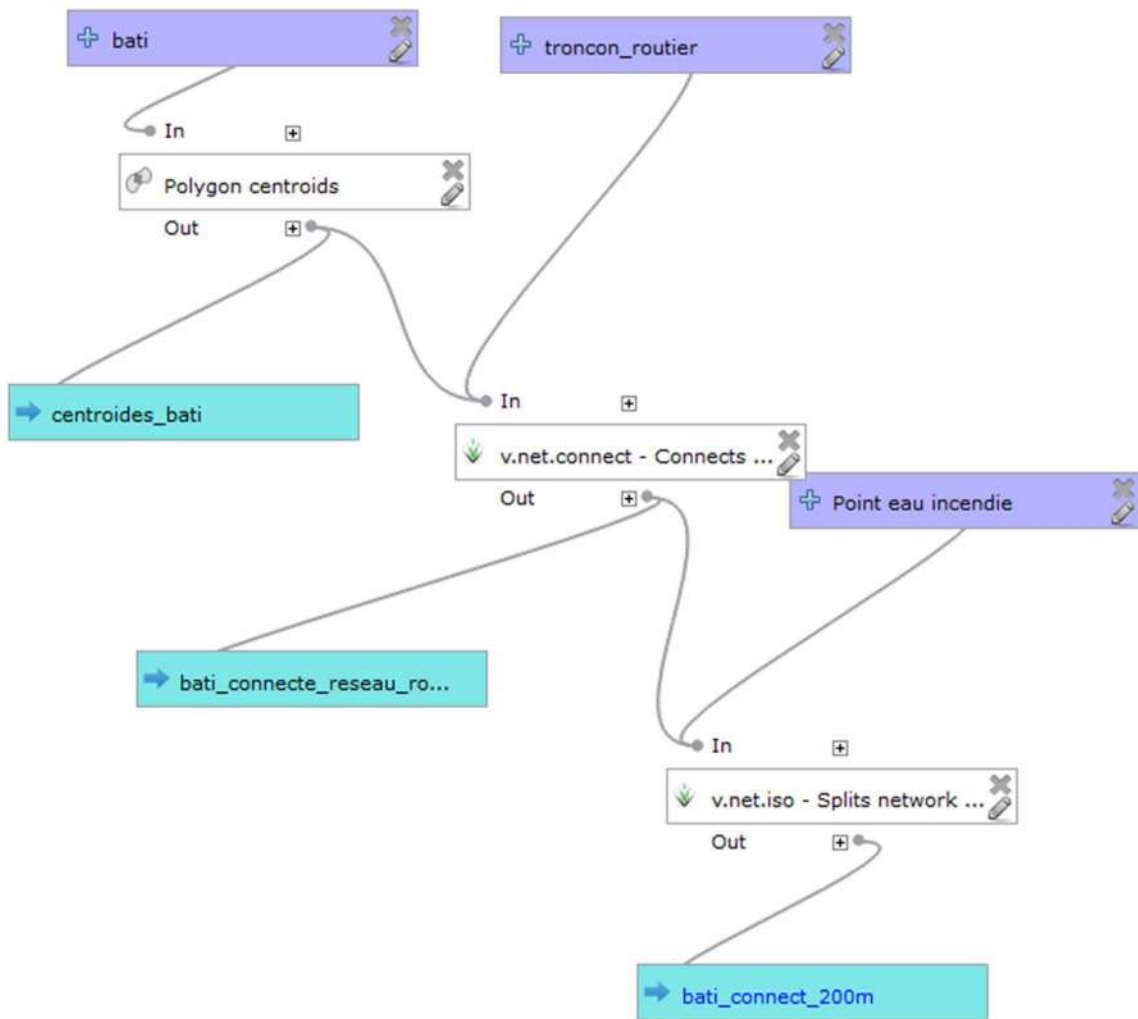
Ici, nous avons observé la notion d'accessibilité des événements par les sapeurs-pompiers de Paris en cas d'incendies sur un des lieux événementiels. En effet, pour les salles de spectacles, ou bâtiments d'habitation, bureau etc. Les établissements recevant du public en général, un point d'eau incendie doit se trouver à moins de 200m de chaque établissement. Ici, cette carte va permettre de montrer que l'on peut savoir si un point d'eau incendie est accessible à 200m des établissements recevant les événements.

Pour ce faire nous utilisons le logiciel libre QGIS pour faire une analyse de l'accessibilité des points d'eau incendie en fonction du référentiel routier Parisien qui est le fruit « des données des capteurs permanents du trafic routier » de la ville de Paris mise en ligne sur le site OpenData de la ville de Paris (<https://opendata.paris.fr/explore/dataset/referentiel-comptages-routiers/information/?dataChart=eyJxdWVyaWVzIjpbeyJjaGFydHMiOlt7InR5cGUiOiJsaW5lIiwuZnVuYyI6IkFWRyIsInlBeGlzIjoib2JqZWNOaWQiLCJzY2IibnRpZmljRGlzcGxheSI6dHJ1ZSwiY29sb3IiOiIjMjYzODkyn1dLCj4QXhpcyl6ImRhdGVfZmluX3ZhbGlkaXRliiwibWF4cG9pbmRzIjoilwidGltZXNjYWxlljoieWVhcilslNvvcnQiOiliLCJjb25maWciOnsiZGF0YXNldCI6InJlZmVyZW50aWVsLWNvbXB0YWdlcy1yb3V0aWVycylslm9wdGlvbniMiOnt9fX1dLCJkaXNwbGF5TG5kIj0cnVlLCJhbGlbnk1vbnR0Ij0cnVlLCJ0aW1lc2NhbgUiOiliifQ%3D%3D&location=14,48.86261,2.36519&basemap=jawg.streets>).

Les hydrants sont également récupérés sur le site de l'OpenData Paris. Le mobilier de distribution d'eau potable et non potable mobilisés en cas d'incendie est caractérisé par une géométrie, un type de mobilier et un libelle du mobilier. Nous utilisons une projection Lambert 93 (code 2154). Qgis présente les modules GRASS qui sont nécessaire et performant pour les traitements concernant l'accessibilité depuis un réseau (module v.net). Dans un premier temps après avoir observé la géométrie des hydrants qui sont modélisés sous forme de multi ligne nous les convertissons en polygones. Puis, nous créons leur centroïde pour les relier au référentiel routier via le module v.net.connect.

Ensuite, en mobilisant v.net.iso qui calcule la distance entre le point d'eau incendie et les événements par le réseau routier nous adaptons la symbologie (Propriétés > Style > Catégorisé) pour analyser les événements qui possèdent bien un point d'eau incendie à moins de 200m pour assurer la sécurité de la population allant sur ces événements.

En vert sont donc représenter les événements à 200mètres d'un poteau incendie. Par ailleurs les tronçons routiers en vert sont ceux qui sont ceux qui relient en 200mètres les Poteaux incendies aux lieux d'événement.



## 3.2 Proposition d'amélioration

Au vue de la structure de la base de données qui présente une redondance d'informations et ne permet pas en tant que tel d'effectuer quelconque traitements statistiques nous proposons ci-dessous un nouveau modèle de base de données. L'objectif du modèle est de présenter une structure robuste, performante, pour pouvoir faire évoluer facilement la BD avec de nouvelles données et assurer leur atomicité.

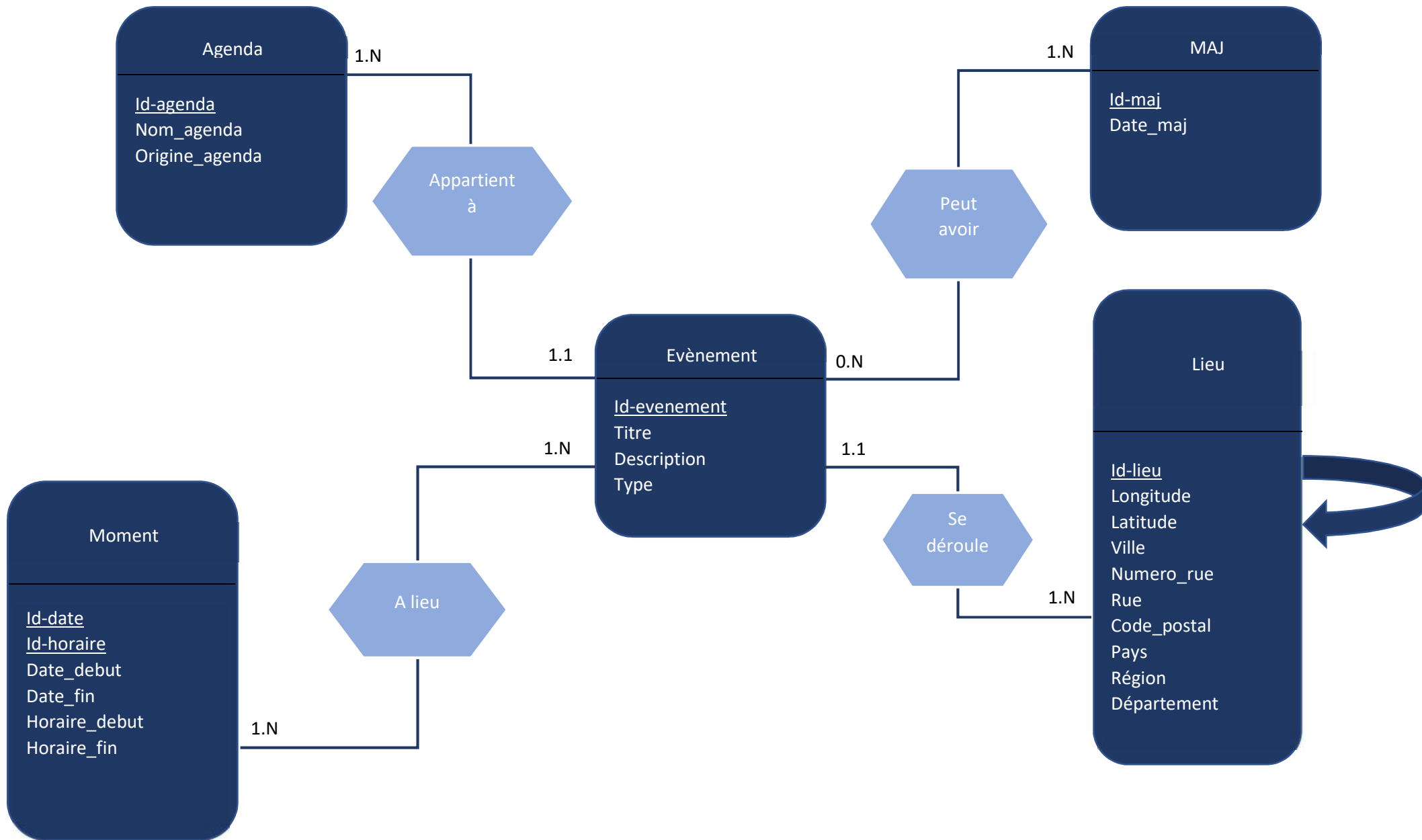
Nous proposons donc un modèle conceptuel dans un premier temps puis un modèle relationnel pour éventuellement impulser ce modèle dans un logiciel de base de données de manière structurée, et pour rendre cohérente et optimale l'analyse des données.

Notre modèle conceptuel se compose de cinq entités elles-mêmes décrites par des attributs propres à chacune d'entre elles. Les entités sont liées les unes entre elles par des associations.

L'entité « événement » est l'entité centrale du modèle. Elle est décrite par plusieurs attributs « Titre, description et type d'événements, et est identifier de manière unique par une clé primaire ici, id-evenement.

L'entité « Agenda » recensé par son identifiant (clé primaire), son nom et son origine est liée à Evenement via l'association « Appartient à ». Un agenda peut recenser un ou plusieurs évènements d'où la relation (1.N).

Par ailleurs, un événement peut être une ou plusieurs fois mis à jour (relation N.N) tout comme il peut se dérouler sur plusieurs horaires (cf festival sur plusieurs jours ou séminaires). Et enfin un lieu peut accueillir un ou plusieurs évènements (relation N.N) et est caractérisé par une clé primaire (id-lieu), une longitude, latitude, un numéro de rue, rue, un code postal, , un pays, une région, et un département.



### 3.3 Proposition d'outils

Tout du long de notre projet, nous avons réalisé différentes analyses spatiales grâce un panel d'outils, l'objectif des réalisations été d'intriguer les différents acteurs publics sur les bénéfices d'utilisation de la base de données.

En ce qui concerne le panel d'outils, il peut s'adresser aux acteurs publics qui souhaitent aller plus loin dans la démarche, ainsi qu'à notre commanditaire s'il désire réaliser des analyses supplémentaires.

Nous avons donc trois typologies d'outils différents webmapping, logiciels SIG, logiciel textométrique, pour chacune des typologies nous avons essayé d'utiliser des logiciels libres et sous licences payantes afin de diversifier les solutions pour réaliser nos analyses.

#### 3.3.1 Logiciels SIG classiques

Les outils SIG, sont principalement utiliser pour analyser différentes données et d'en représenter différents phénomènes géographiques liés à un territoire. Ses représentations peuvent nous aider à percevoir des corrélations, cela peut intéresser les acteurs publics s'ils souhaitent observer les différents phénomènes liés aux événements.

Nous nous sommes donc principalement intéressés à deux logiciels :

- QGIS :

Qgis est un logiciel open-source, qui permet cartographie de données, de nombreux acteurs utilisent cet outil. Sa gratuité ainsi que sa communauté sont un réel atout pour le développement de ce logiciel qui est fréquemment amélioré. Mais la prise en main du logiciel n'est pas aisée pour une utilisation approfondie, comme tous les logiciels de SIG, il nécessite souvent une compréhension et une maîtrise des données géographique.

<https://www.qgis.org/fr/site/>

- ArcGIS :

Arcgis est un logiciel payant, qui permet lui aussi de faire de la cartographie de données. Également il propose différents outils de cartographie, comme des applications mobiles, des représentations en ligne, qui permet d'ajouter un plus à la cartographie. Ainsi que de la récolte de données, l'automatisation des traitements des données et le stockage des données en ligne. Cependant la tarification mensuelle reste très coûteuse, mais adapté en fonction des compétences des différentes utilisations.

<https://www.esri.com/en-us/arcgis/products/index>

### 3.3.2 Webmapping

Les outils de webmapping, comme expliqué auparavant, sont des outils de représentation cartographique en ligne. Dans notre cas, cette possibilité de visionner une représentation d'un territoire à travers un site internet rend l'outil attractive, renforcé par le côté dynamique des cartes ainsi que parfois leurs statistiques en fonction de ce qui est visible à l'écran. Nous nous sommes donc principalement intéressés à trois logiciels :

- LIZMAP :

Logiciel open-source, qui permet à travers le logiciel SIG QGIS de créer des applications Web cartographique. Mais il faut une bonne compétence pour la mise en place de cet outil et nécessite un serveur.

*Compétence : maîtrise des Outils SIG QGIS + LizMAP, maîtrise de représentation cartographique, gestion des données géographiques, gestion d'un serveur.*

<https://www.3liz.com/lizmap.html>

- CARTO :

Logiciel qui permet de mettre en ligne des cartes interactives avec divers représentations, données et échelles. Carto met à disposition son code source gratuitement sur github, mais a aussi développé un service payant afin d'aider la mise en place et l'entretien des serveurs. Pour les deux cas des compétences sont nécessaires.

*Compétence : maîtrise de représentation cartographique, gestion des données géographiques, gestion d'un serveur, mise en place d'un code source.*

<https://carto.com/>

- ArcGIS online :

ArcGIS online, qui est un service supplémentaire de ArcGIS qui permet la vision en ligne de cartes, avec possibilité de mettre en places des "applications" afin de créer une interactivité au sein de la carte.

*Compétence : peut être mis en place par une prestation de service.*

<https://www.arcgis.com/home/index.html>

### 3.3.3 Logiciel Textométrie :

Nous avons eu des difficultés à mettre en place des types de catégories pour les événements. Au cours de notre année, nous avons découvert un logiciel de textométrie TXM qui offre la possibilité d'analyser les mots d'un texte. Des observations sur l'occurrence des mots sont possible, ce qui peut permettre la création de catégories et de mots-clés.

L'utilisation de ce logiciel libre d'utilisation ne nécessite pas de compétence particulière, sauf pour des recherches avancées qui sont possible via un langage de programmation particulier au logiciel.

<http://textometrie.ens-lyon.fr/>

## Conclusion

Ce projet montre la puissance des SIG dans un domaine qu'est l'événementiel pour analyser les données et les visualiser. Cette commande nous a montré qu'il était nécessaire d'avoir des données de qualité pour passer directement à l'analyse et la visualisation des données. Nous avons passé une grande partie de notre projet à réfléchir comment concevoir une nouvelle structuration de base de données pour que l'export de la base de données depuis l'API soit la plus exploitable possible directement sans passer par une phase de Re-nettoyage/ réorganisation des données.

La puissance des SIG se réalise lorsque les données sont structurées pour que nous puissions les analyser et présenter les résultats de nos pistes exploratoires au grand public.

Après avoir pérennisé la base, une méthode d'analyse a été mise en place en définissant des pistes exploratoires qui permettent d'affirmer qu'OpenAgenda dispose d'une base de données, très riche en informations. Nous avons impulsé les capacités d'analyse possibles via l'utilisation de la base de données, pour les collectivités territoriales/acteurs publics en général. (Analyse spatiale selon type d'événement, suivi temporel/étude diachronique).

Derrière la dimension technique, organisationnelle, et partenariale qu'à demander ce projet, en retour sur notre travail n'est qu'il t pourrait encore plus être tourné vers l'automatisation de tâches si OpenAgenda souhaite approfondir ces analyses et les rendre plus exploitables.

Par ailleurs, d'autres pistes exploratoires pourraient être prises en compte mais il s'agissait ici, d'impulser une méthode d'analyse pour montrer le potentiel de la base de données OpenAgenda. Bien sûr notre analyse repose sur une zone d'étude bien précise et particulière. La capitale de France n'est pas représentative de toutes les villes situées en provinces. Les logiques spatiales ne sont donc pas forcément les mêmes et la disponibilité des données extérieures croisées ne se révèlent pas toutes aussi disponibles que pour Paris. En termes de perspective, des outils OpenSource pourraient être plus exploré pour limiter le coût tout en assurant le rendement des opérations et en promouvant ces multiples informations.

Avec ce projet, nous avons appris à construire une problématique, adapter la méthodologie, proposer des solutions et rédiger une analyse rigoureuse pour présenter les résultats. Cette recherche nous a également permis de croiser nos différents profils pour mener à bien ce projet, en mobilisant nos connaissances et affirmé que la géomatique est un domaine pluridisciplinaire.



## 4 ANNEXES

### 4.1 Document intermédiaire rendu en décembre à notre commanditaire

#### 1 PRÉSENTATION

OpenAgenda, créé en 2011, permet de centraliser et diffuser des événements. Est considéré comme événement, dans notre étude, tout genre de rendez-vous public.

La base de données de cette entreprise est très peu exploitée de façon statistique, là est donc toute notre mission : valoriser la base de données.

Les objectifs sont donc de faire parler les données et de leur associer une représentation cartographique adaptée.

Pour répondre au mieux à cette commande nous avons pensé à utiliser, pour la visualisation de données, le logiciel Carto.com qui propose une application serveur, une application cliente (navigateur web) et offre la possibilité d'importer une base de données. Un utilisateur peut alors demander à un serveur, grâce à un navigateur web et des services, d'afficher des informations géographiques sur une carte.

Par ailleurs, nous souhaiterions associer des traitements statistiques à la visualisation des données (pop-up qui s'affichent quand on clique sur un arrondissement par exemple et qui présente les statistiques associées à cet arrondissement). Nous utiliserons pour cette partie des logiciels comme RCommander, Geoda, ou ArcGisPro qui proposent de nombreux traitements.

#### 2 LOGICIELS UTILISES

Pour réaliser le projet nous utiliserons, trois logiciels : CARTO, ArcGis et QGIS.

Carto est un logiciel qui propose de la visualisation avancée en cartographie. A partir du système *software as a service*, il est possible de créer des cartes dynamiques en stockant ces données directement en ligne.

##### INTÉGRATION DES DONNÉES

|                      | Formats de données  | Fonctions  | Installation / Accessibilité  |
|----------------------|---|--|---|
| <b>CARTO</b>         | .csv,<br>.shp,<br>.kml,<br>.geojson,<br>.svg,<br>.gpkg,<br>.osm | Cartographie dynamique, visualisation de données de manière interactive.   | Téléchargeable rapidement et en ligne.<br>150€/mois pour l'application et un devis doit être envoyé à l'équipe de Carto pour quantifier le coût des licences professionnelles. Il est nécessaire également de se créer un compte sur Carto.com. Pour l'achat de licence professionnelle, celle-ci offre la possibilité d'avoir accès à un cloud pour stocker directement ces données et gérer les propres cartes produites. |
| <b>ArcGis / QGIS</b> | .csv,<br>.shp,<br>.txt,<br>.gpx,<br>.geojson                    | ArcGis et QGIS sont des logiciels téléchargeables très rapidement pour faire du traitement de données spatiales. | ArcGis plusieurs milliers d'euros par an pour des licences professionnelles. Logiciel très complet et ergonomique.<br>QGIS est un logiciel <b>libre</b> proposant les mêmes fonctionnalités qu'ArcGis. Il n'est pas possible d'effectuer de statistiques poussées sur ce logiciel contrairement à ArcGis.   |

## INTERACTIVITÉ SUR CARTO

Carto rend les cartes interactives. Nous pouvons définir ce degré d'interactivité avant de publier la carte en modifiant les paramètres : changement de zoom, et en choisissant les couches à visualiser.

Les widgets permettent une grande interactivité données/utilisateur. Ce que l'on sélectionne sur le widget s'affiche directement sur la carte. Par exemple, si nous souhaitons visualiser seulement les événements de 2015, en cliquant sur l'année souhaitée (2015), seuls les événements de cette année-là apparaissent sur la carte.

Par ailleurs, les pop-ups fonctionnent de la même manière que les widgets. Si nous cliquons sur une entité alors les informations de cette seule entité apparaissent.

### 3 METHODES

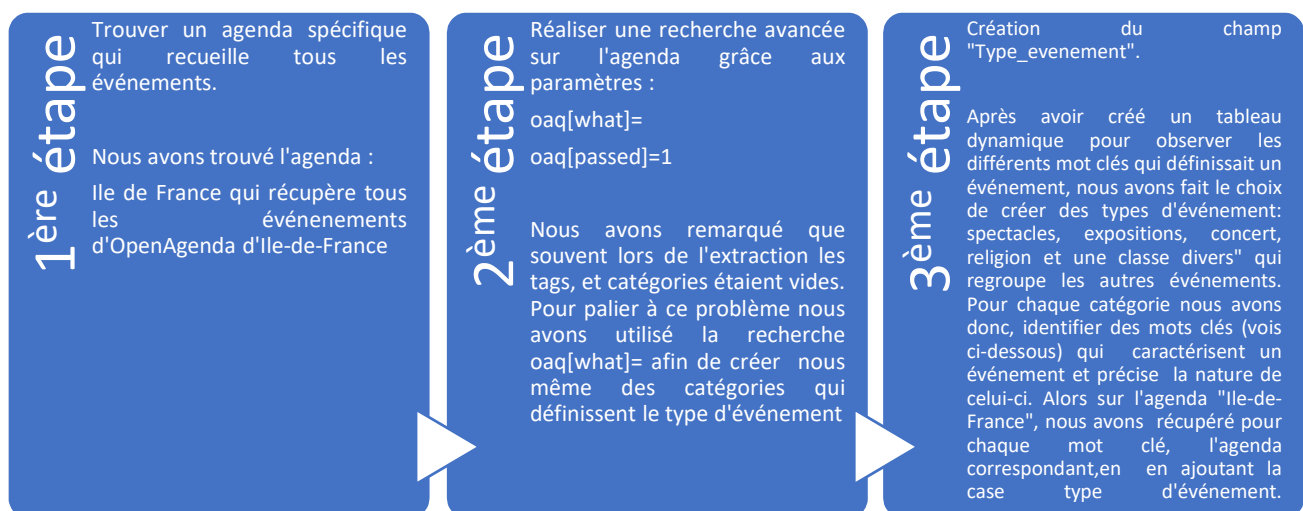
Dans le but d'analyser la base de données d'OpenAgenda, nous avons réfléchi à différentes fonctionnalités à mettre en place pour présenter simplement et efficacement les données. Pour ce faire, à travers des solutions web ou des logiciels d'analyse et de géo-traitement plus avancés, nous vous proposons des premiers résultats.

Il s'agit, ici, d'expliquer nos différentes méthodes d'analyses spatio-temporelles réalisées sur vos données, pour vous expliquer notre démarche et quelques exemples de fonctionnalités de logiciels utilisés.

Pour cela vous trouverais ci-après, nos différents schémas de traitement qui aboutissent à l'extraction des données et à leur implémentation.

#### RECUPERATION DES DONNÉES

Nous avons rencontré de nombreuses difficultés lors de la récupération des données. Les requêtes que nous effectuons sur la ville de Paris ne nous extrayaient pas les bonnes données. Pour ce faire, nous avons fouillé les agendas proposés par OpenAgenda et avons téléchargé celui prénommé agenda « Ile-de-France ».



## Spectacles

- Théâtre(s), danse(s), feu d'artifice(s), bal(s), karaoké(s), carnaval(s), opéra(s)

## Expositions

- Vernissage(s), exposition temporaire(s), musée(s), patrimoine(s), visite guidée(s)

## Concert

- Music(s), musique(s), chant(s), fanfare(s), festivale(s), Garelive(s)

## Religion

- Messe(s), procession(s), communion(s), baptême(s), pèlerinage(s), catéchisme(s), événements diocésain(s), nativité, sanctuaire(s)

## Divers

- Autres événements

Par cette méthode nous avons récupéré l'intégralité des événements d'Ile de France en leur attribuant un type. Cette méthode présente un avantage : lorsque nous extrayons un agenda nous savons que tous les événements extraits appartiennent à la même catégorie. Ensuite, nous avons juste à fusionner tous les tableaux extraits pour reformer la base de données complète. Cependant par cette méthode nous avons créé de multiples doublons. Il a donc fallu procéder à un nettoyage en retirant donc les doublons, via une fonction Excel et en effectuant un tri pour ne conserver les événements compris dans la petite couronne de Paris. Les traitements de nettoyage sont spécifiés ci-dessous.

### *NETTOYAGE DE LA BASE DE DONNEES EXTRAITE*

Afin de supprimer les doublons, et conserver seulement les événements nécessaires à notre étude, il a fallu dans un premier temps :

- Trier par département et ville afin d'avoir uniquement les événements de Paris et supprimer les événements qui s'y trouvent en dehors.
- Attribuer correctement l'arrondissement en fonction du code postal (Fonction DROITE)
- Recoder le champ prix pour attribuer « ND = non dévoilé » aux entités ne spécifiant cette information et uniformiser l'information
- Ajouter un champ « Type\_evenement »
- Fusionner les différents fichiers corrigés
- Supprimer les doublons entre fichiers
- Gérer les dernières erreurs d'encodage

Ce nettoyage a été réalisé sur chaque fichier correspondant à un mot clé. Ils ont été traités séparément avant fusion car les fichiers étant particulièrement lourds, il est difficile pour Excel (et encore plus calc) de gérer toutes ces informations.

## POSSIBILITES DE CROISER LES DONNEES AVEC D'AUTRES INFORMATIONS

L'analyse spatiale, temporelle ou événementielle dépend parfois de facteur externe à la base d'OpenAgenda. Il est donc intéressant de pouvoir croiser les données, avec d'autres variables afin de mieux comprendre, interpréter ou analyser un phénomène.

Pour cela, l'Open Data d'aujourd'hui nous offre de nombreuses possibilités notamment grâce à certaines données officielles que l'on peut retrouver sur : [Data.Gouv](#), [INSEE](#), [Data.SNCF](#) ou encore les Open Data des grandes villes tels que [Opendata.Paris](#), [data.iledefrance](#), [data.grandlyon](#).

Cependant, il faut faire attention aux données libres de service, elles sont parfois construites de manière maladroites et peuvent induire en erreur. Il faut donc être vigilant lors de l'utilisation de ces données.

Listes des différentes données récupérables :

1. Flux des transports
2. Revenu par communes
3. Subvention culturelle
4. Nombres d'associations par communes
5. Nombre d'habitant
6. Catégorie socioprofessionnelle
7. Occupation du sol
8. Relief
9. Prix des locations des lieux culturels public

## IMPLEMENTATION DANS LES LOGICIELS et TRAITEMENTS STATISTIQUES

Une fois la base nettoyée, recodée, nous effectuons des analyses spatiales (comment se répartissent les événements au sein des arrondissements de Paris) et statistiques (y'a-t-il un arrondissement qui accueille plus de concert ? de vernissage ?). Enfin, nous souhaitons croiser ces résultats avec des variables extérieures comme le prix au mètre carré des arrondissements, la densité de population et le revenu médian des arrondissements.

Il s'agira dans un premier temps d'effectuer de simples statistiques descriptives pour faire parler les données au premier abord puis d'enrichir les analyses avec des statistiques plus poussées.

Pour réaliser cette étape, nous souhaitons dans un premier temps, terminer le nettoyage des données pour effectuer les traitements statistiques. Les résultats ne seront donc pas présentés dans ce document. En revanche, nous pouvons d'ores et déjà présenter des représentations spatiales de la répartition des événements.

### 4 VISUALISATION DES DONNÉES

---

#### PREMIERS RESULTATS

Voici, nos premiers résultats obtenus après extraction de la base de données en .csv :

<https://fourneyrona.carto.com/builder/769f84ca-5b46-42cb-8de5-03822b7e1f08/embed>

Le lien dirige vers une carte interactive représente le nombre d'événements du département de Paris entre 2008 et 2017.

Grâce à cette carte, il est possible d'analyser spatialement la répartition des événements en fonction de plusieurs facteurs :

- Années (2008-2017)
- Types d'événement
- Mois
- Prix de l'événement
- Arrondissement
- Nom du lieu réceptionnant l'événement

L'observation des différents histogrammes, permet en un coup d'œil d'observer l'évolution du nombre d'événement ajouter sur OpenAgenda. Par ailleurs, il est possible d'apprécier les arrondissements qui accueillent le plus d'événements ainsi que les périodes (par mois) les plus représentées.

Bien d'autres analyses seraient possibles mais attention aux résultats, il faut garder en tête que OpenAgenda n'est pas utilisé par tous les acteurs porteurs d'événements.

Par ailleurs, voici une carte réalisée sous le logiciel QGIS qui représente une étude diachronique en 2011, date de création d'OpenAgenda et 2018, pour visualiser la nature et le nombre d'événement par arrondissement parisien :

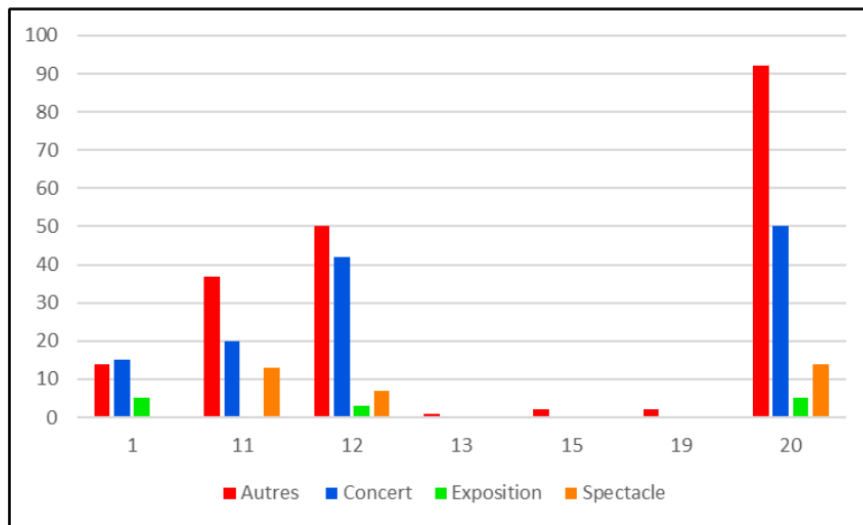


Figure 1: Nature et nombre d'événements par arrondissement en 2011

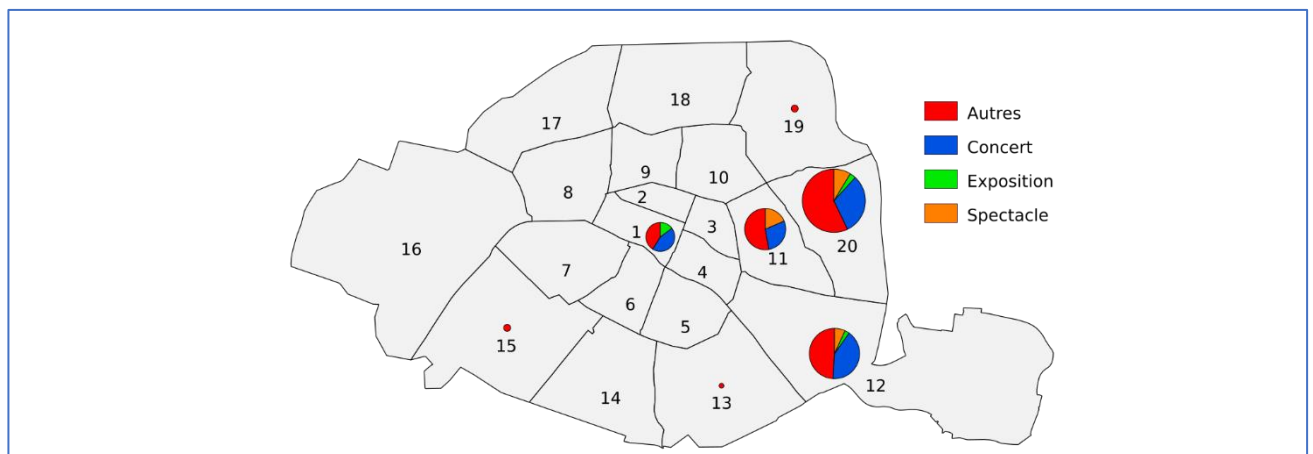


Figure 2: Nature des événements par arrondissement en 2011

Le XX<sup>ème</sup>, XII<sup>ème</sup> et XI<sup>ème</sup> sont des arrondissements qui accueillent le plus de concert en 2011, selon les agendas rentrés dans OpenAgenda tandis que le I<sup>er</sup> arrondissement, quartier touristique est celui qui regroupait le plus d'expositions.

Observons alors en 2018, si ces arrondissements suivent la même tendance. Tout d'abord, il faut savoir que beaucoup plus d'agenda ont été intégrés en 2018 sur OpenAgenda par rapport à 2011.

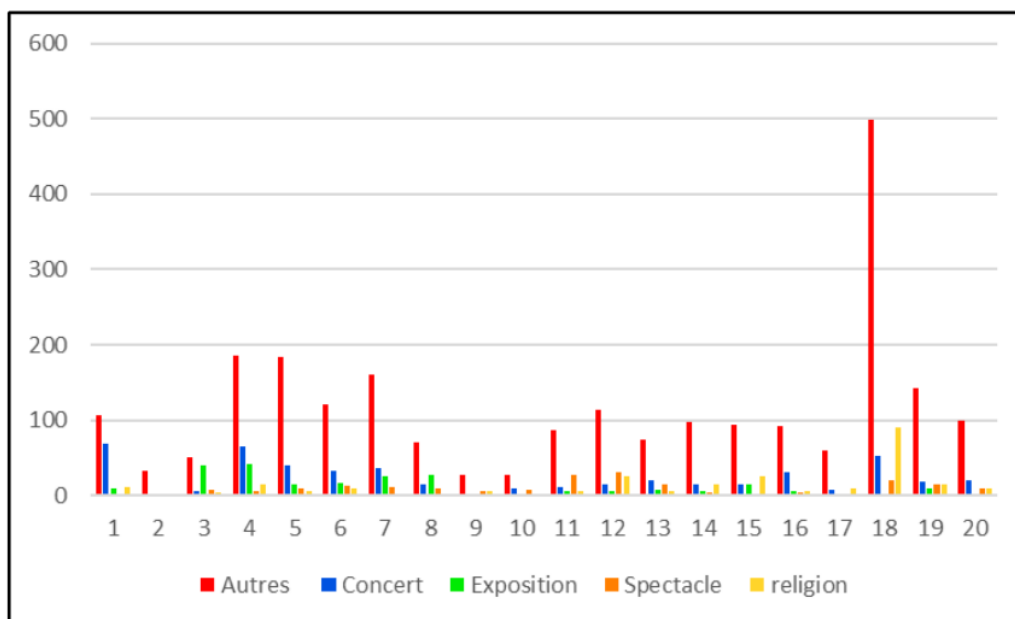


Figure 3: Type et nombre d'événement, par arrondissement en 2018

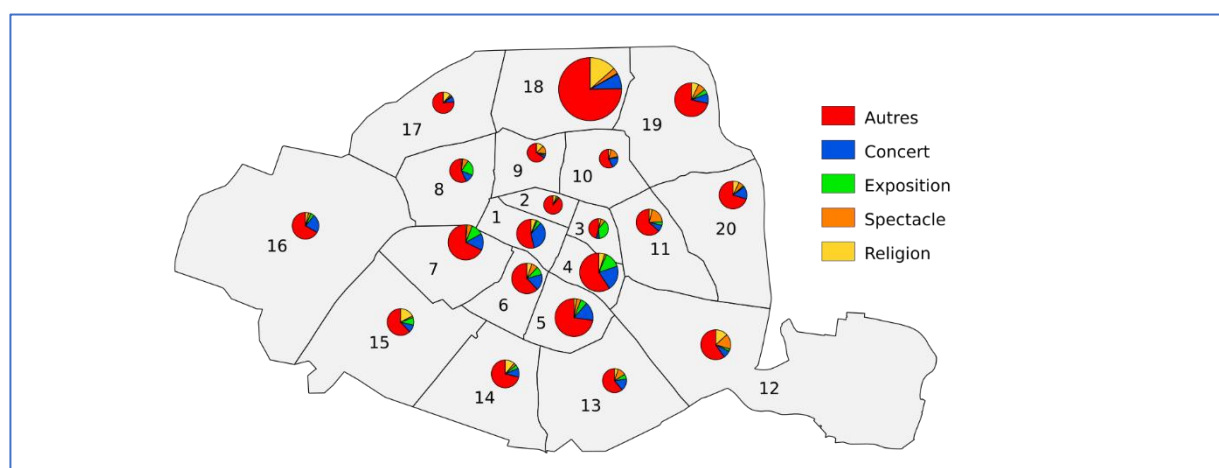


Figure 4: Nature des événements par arrondissement en 2018

Nous observons qu'en 2018, un nombre plus important d'agenda a été implantés dans la base de données OpenAgenda. Des caractéristiques nouvelles apparaissent comme par exemple le XVIII<sup>ème</sup> arrondissement avec de nombreux événements de nature religieuse ou l'hypercentre de Paris qui regorgent d'expositions.

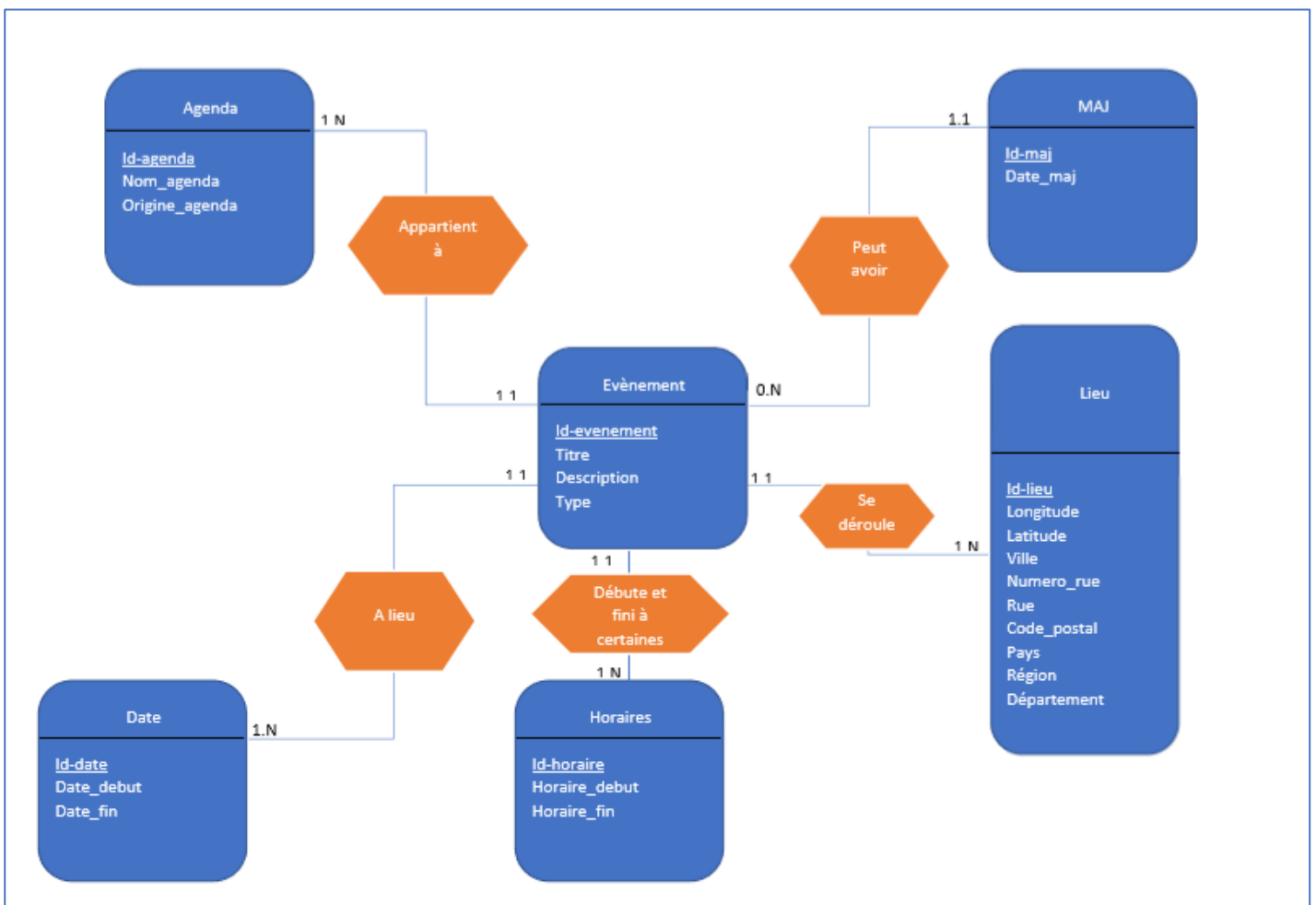
Un constat en 2018, le centre de Paris, accueillent de nombreuses expositions tandis que la périphérie, organise des événements plus populaires tels que des concerts ou des spectacles.

## 5 PROPOSITIONS D'AMELIORATION DE LA BASE DE DONNEES POUR LA RENDRE PLUS EXPLOITABLE

Pour minimiser au mieux le temps de nettoyage de la base de données, nous suggérons peut-être de repenser le formulaire de publication d'agenda sur la plateforme. Cette étape permettra de guider les utilisateurs à poster leurs agendas en les cadrant dans les saisies des champs, à fournir dans le formulaire de publication. Effectivement, en exportant, les données en .csv à l'heure actuelle, une base de données nous est fournie mais celle-ci est difficilement exploitable. En effet, il n'y a pas de champ spécifique pour préciser la nature de l'événement (concert, exposition, messe, vernissage, spectacle). Cette étape limiterait une partie des erreurs d'encodage. Les utilisateurs utilisent à l'heure actuelle des manières très différentes pour qualifier

leurs événements. Soit par des textes assez denses, soit par des mots clés, les horaires sont soit spécifiées via une phrase ou de la forme 8h ou 8 :00. Toutes ces manières différentes d'encodage peuvent paraître absurdes mais il est nécessaire d'avoir le même format pour une exploitation optimale de la base de données sans perdre de temps à nettoyer celle-ci. Enfin, le champ adresse n'est pas correctement intégré à la base. Les données ne sont pas atomiques ce qui limite la puissance de la base. Ici, par exemple, le champ adresse comporte plusieurs informations (numéro, rue). Il faudrait tout décomposer pour une base de données puissante et exploitable.

Pour remédier à ces problèmes il faudrait proposer sous forme de liste déroulante un choix de nature d'événement, d'horaire, etc, pour uniformiser la saisie des informations. Si un utilisateur ajoute un événement d'un nouveau type, nous pourrions imaginer un bouton permettant l'ajout automatique d'un nouveau type d'événement dans la base de données. Par ailleurs, le champ description serait conservé pour décrire l'événement (concert rock, classique, jazz etc...). Ainsi, lors de l'extraction de la base de données, les champs seront remplis tous de la même manière et offrirait une base plus propre prête à être exploitée. Voici donc, une proposition de modèle conceptuel et relationnel pour améliorer la base de données :



Agenda(Id-agenda, Nom\_agenda, Origine\_agenda)

Evenement(Id-evenement, Titre, Description, Type, #Agenda, #id\_horraire,#id\_lieu,#id\_date)

MAJ( id-maj, date\_maj,#id\_evenement)

Lieu(Id-lieu, Longitude, Latitude, Ville, Numero\_rue, Rue, Code\_postal, Département, Région, Pays)



Date(id-date, date\_debut, date\_fin)

Horaire( id\_horaire,horaire\_debut, horaire\_fin)

Le modèle conceptuel doit représenter l'ensemble des entités avec leurs attributs. Un identifiant doit être présent pour qualifier chaque entité de manière unique. Les cardinalités sont ici, pour faciliter la lecture et permettre le passage au modèle relationnel. Par exemple, nous lisons de la sorte un événement fait parti d'un seul agenda tandis qu'un agendan peut présenter plusieurs événements. Dans le modèle relationnel, les # font référence aux clés étrangères. C'est une règles appliquée à partir des cardinalités, pour créer un modèle efficace lors d e l'implémentation dans le logiciel de base de données.

Après ces règles de transformation effectuées, il faudrait étudier les formes normales des entités pour que la base de données soit bien conçue. Il s'agit de règles de transformation pour que la base de données soit performante. Par ailleurs, pour que la base de données soit exploitable et bien conçue les données doivent être cohérentes, non redondantes et structurées. Les règles de normalisation sont là pour vérifier ces trois points.

Vient alors l'implémentation dans une base de données sous forme de script SQL. PostGresSQL est un logiciel de base de données qui est très efficace dans le domaine des bases de données.

#### 4.1 Document de gestion de projet réalisé au cours de l'année

### Gestion de projet : Open Agenda (1<sup>er</sup> TD)

OpenAgenda, créé en 2011 permet de centraliser et diffuser des événements. Est considéré comme événement, dans notre étude, tout genre de rendez-vous public. L'objet du projet est de mettre en ligne une cartographie interactive des événements de la base de données d'OpenAgenda qui contiendra également des statistiques associées. Celles-ci doivent être pertinentes puisqu'il s'agit d'attirer des acteurs publics afin qu'ils travaillent sur ces statistiques. Par exemple, si nos résultats révèlent qu'un quartier n'a aucun événement et qu'il y a un projet de réhabilitation, alors les acteurs vont mettre en place des événements pour dynamiser ce quartier.

#### Résultat attendu :

La base de données d'OpenAgenda est très peu exploitée de façon statistique, là est toute notre mission : Valoriser la base de données pour attirer les acteurs publics afin qu'ils exploitent eux-mêmes cette base de données.

#### Objectifs :

Le commanditaire attend de nous de faire parler les données en leur associant une représentation cartographique adaptée. Il souhaite un rendu sous forme de Webmapping. D'après la définition issue du wiki SIG du master géographies numériques, il s'agit « d'un processus qui permet l'utilisation et l'analyse de cartes sur le réseau Internet. Cela nécessite une application serveur, une application cliente (navigateur web) et une base de données. Un utilisateur peut alors demander à un serveur, grâce à un navigateur web et des services, d'afficher de l'information géographique sur une carte ».

Il s'agira donc pour nous d'associer traitements statistiques et cartographie. Pour ce faire nous avons défini une zone d'étude avec le commanditaire. Il s'agit de Paris et sa grande couronne.

En effet, dans le temps imparti nous ne pouvons pas traiter et exploiter les 275 000 milles données mises à disposition pour la France entière.

Notre application devra donner la possibilité à l'utilisateur de sélectionner un territoire (Ville, par exemple) pour obtenir les statistiques liées aux événements. Par exemple : quel quartier a le plus d'événements l'été ? De quels types sont-ils ? Vernissages, concerts, dégustations de vins, etc. Un bouton recherche par type d'événements devra également être établi. Il s'agit de faire une carte interactive qui lie nos statistiques avec de la cartographie.

## Gestion et récupération des données :

Les données d'open Agenda sont mises à disposition par un API qui permet d'extraire un .JSON ou CSV, il y a plus de 314 500 événements enregistrés. Les données sont publiques et disponibles sur le site : <https://openagenda.zendesk.com/hc/fr/articles/115003170853--v1-events-R%C3%A9cup%C3%A9rer-des-%C3%A9v%C3%A9nements>

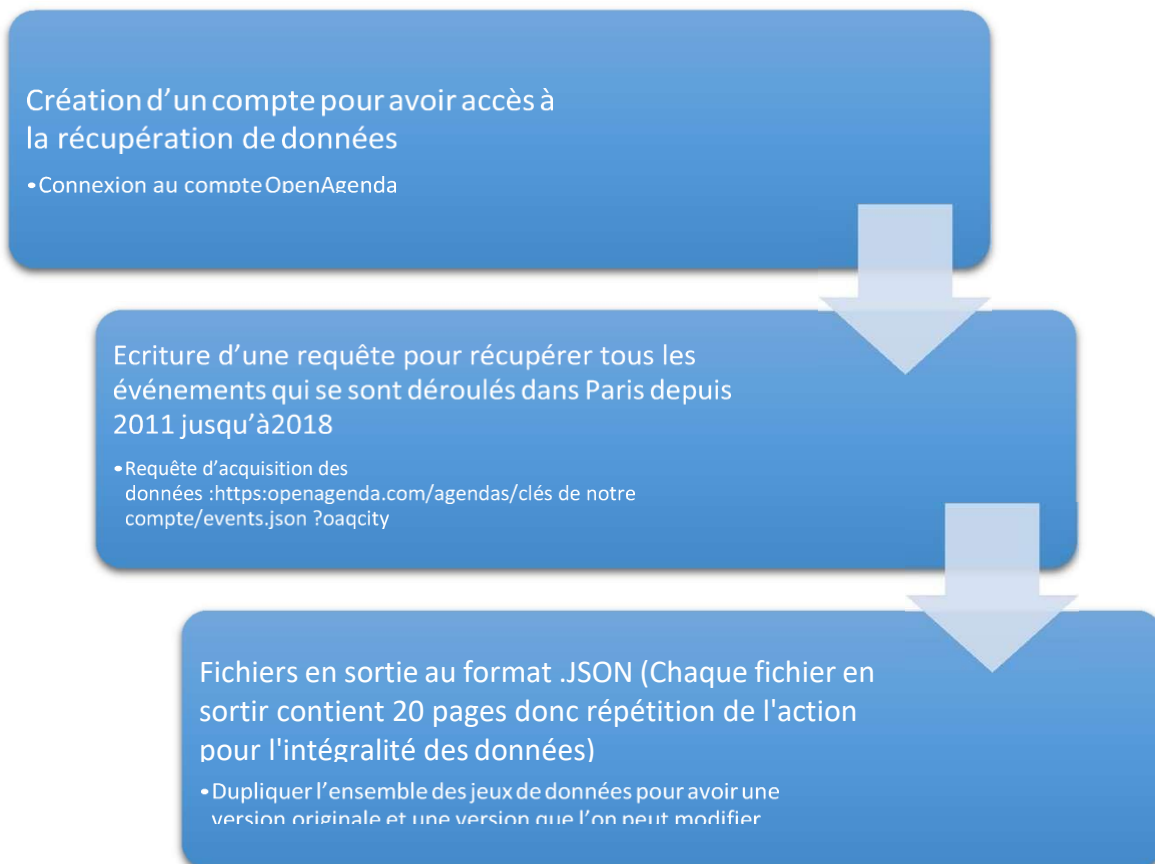
Nous pouvons extraire un maximum de 300 données à la fois. Il s'agit donc d'une étape fastidieuse puisque notre zone d'étude contient plus de 75 000 données. Les données d'Open Agenda se caractérisent de cette manière : les données « Agenda », « Evènements », « Lieux » et « Images » qui sont caractérisées par des attributs et un identifiant unique.

| Agenda   |
|--|
| <ul style="list-style-type: none"><li>• <u>Identifiant</u></li><li>• Avatar (optionnel)</li><li>• Nom (255 caractères maximum)</li><li>• Description (optionnelle, 160 caractères maximum)</li><li>• Tags(optionnels): le propriétaire d'un agenda définit ses propres tags, afin d'organiser ses événements</li></ul> |
| Lieux  |
| <ul style="list-style-type: none"><li>• Nom du lieu</li><li>• Adresse (texte saisi par l'utilisateur)</li><li>• Latitude &amp; longitude</li><li>• Département</li><li>• Région</li><li>• Ville</li><li>• Pays</li></ul>   |
| Images   |
| <ul style="list-style-type: none"><li>• Image</li><li>• Description multilingue</li><li>• Téléphone</li><li>• URL(s)</li><li>• Mots-clés</li></ul>   |

| Evènements   |
|--|
| <ul style="list-style-type: none"><li>• <u>Identifiant</u></li><li>• URL: URL de la page événement sur openagenda</li><li>• Image : optionnelle. Formats acceptés : png, jpg, bmp.</li><li>• Langue : un événement peut être multilingue</li><li>• Titre</li><li>• Description courte : facultatif</li><li>• Mots-clés : Champ facultatif</li><li>• Conditions de participation : texte libre, facultatif.</li><li>• Moyens d'inscription : champ facultatif, il permet de saisir email, URL ou numéro de</li><li>• Horaires : horaires de début et de fin, obligatoire, plusieurs plages horaires possibles. Les exports donnent une période (résumé des plages horaires), les horaires en FR/EN et ISO. Accessibilité – handicap : facultatif</li><li>• Accessibilité – âge : deux champs décimaux facultatifs (âge minimum, âge maximum)</li><li>• Date de publication ou de dernière date de mise à jour : Date au format ISO.</li></ul> |

Page 2 sur 6

Voici un schéma de traitement concernant la récupération des données :



Une grande partie des données publiques quant à elles sont disponibles sur le site :

<https://opendata.paris.fr/explore/?sort=modified>

## Outils d'analyse et de représentation

Nous n'avons pas d'outils spécifiques mis à disposition par le commanditaire. Ses données sont directement accessibles sur son site internet et les logiciels utilisés pour les statistiques sont ceux que nous avons abordés en cours et téléchargeables gratuitement. Nous utiliserons RCommander, Rstudio, ExplorR, GeoDA qui sont quatre logiciels complémentaires pour le traitement de données.

Pour les logiciels de représentation spatiale, nous pensons utiliser les logiciels : Data catalogue, Flourish, Pixelmap, GéoDA, et carto. Ce dernier une solution open source accessible en version gratuite ou payante (plusieurs limitations dans sa version gratuite, par exemple 250 Mb seulement de stockage) et permet de réaliser des cartographies à partir de bases de données. Il de mettre en forme des données de façon interactive et visuelle puis de les partager. Il nous semble donc être le logiciel le plus approprié pour répondre à la commande notre commanditaire.

## Méthodologie

Afin de réaliser notre projet nous avons élaboré une méthodologie de travail pour nous aider à effectuer les différentes tâches (voir page 6 schéma de répartition des tâches). Cette planification, nous permet d'avoir des dates clés réaliser pour avancer le travail avec les exigences du commanditaire.

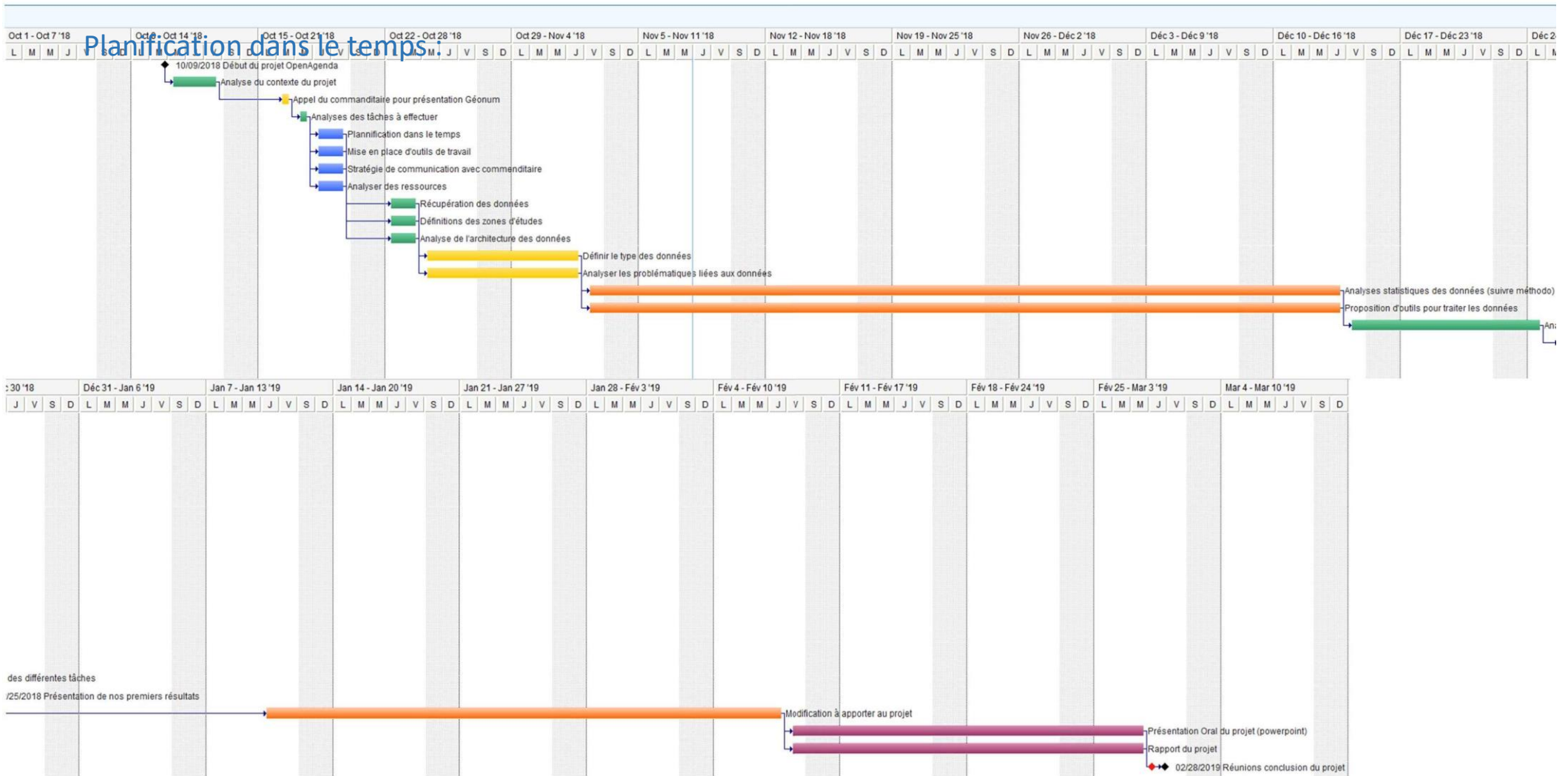
Elle nous permet aussi de travailler en parallèle sur différentes tâches nécessaires à la réalisation du projet afin d'être plus efficaces et ne pas perdre de temps.

Voici notre méthodologie :

1. Récupération des données
2. Définition zone étude
3. Définition des pistes exploratoires avec des données extérieures à la base de données OpenAgenda (ex : revenu médian, densité, ...). Que voulons-nous montrer à travers les statistiques ? (Exemples : Est-ce qu'un quartier qui a un prix élevé aux mètres carrés a tendance à accueillir un certain type d'événement ? Il y a-t-il des quartiers plus animés que d'autres ? En quelle période ?)
4. Définition des traitements nécessaires pour chaque point que nous voulons montrer avec un schéma de traitement associé à chaque traitement effectué
5. Calculs des statistiques (test Khi2, ANOVA, ACM).
6. Représentation des données sur le logiciel de cartographie.

## Conclusion :

Nous sommes très libres dans le choix de ce que l'on souhaite étudier et montrer tant que cela est utile pour attirer les acteurs publics. Ceux-ci pourront l'exploiter et mettre en place des systèmes de réaménagements, des plans d'urbanisme ou renforcer des lignes de transports par exemple (si un quartier accueille beaucoup d'événements et qu'il est mal desservi par exemple).



## Répartition du travail :



## Gestion de projet : Open Agenda (2<sup>nd</sup> TD)

| Nature du problème                | Difficultés   | Solutions envisageables  |
|-----------------------------------|---|--|
| <p>Récupération des données :</p> | <p>La récupération de données s'effectue via une API. Dans notre étude nous souhaitons effectuer une requête qui extrait tous les événements entre 2011 et 2018 <u>et</u> qui ont eu lieu dans la petite couronne de Paris. Or, lorsque nous effectuons cette requête simple, en résultat nous obtenons des événements qui se sont déroulés en Gironde, Alsace, PACA etc... Nous avons donc poussé nos recherches sur le site internet d'OpenAgenda où est mis à disposition des exemples de requêtes sur certains agendas. En les testant celles-ci ne fonctionnent pas non plus... Il est donc impossible de récupérer des données selon certains champs de la base de données malgré ce qui est indiqué sur le site internet du commanditaire.</p> | <p>Nous avons donc testé l'extraction des événements parisiens entre 2011 et 2018 en fonction d'autres champs qui eux fonctionnent lors de l'exécution de la requête : Nous les avons filtrées par date, puis à partir de la longitude et latitude.</p> <p>En fouillant la base de données de l'API, nous avons trouvé un dossier où sont recensés tous les événements de Paris depuis 2011 jusqu'à aujourd'hui. Nous l'avons donc récupéré ce qui nous a évité d'extraire les données via l'API qui ne fonctionne pas correctement...</p> |



|                                    |  |   |
|------------------------------------|--|---|
| <p>Mise en forme des données :</p> | <p>Le fichier .csv recensant les données parisiennes n'est pas nettoyé.</p> <p>Lorsqu'un utilisateur souhaite publier un événement, le formulaire pour décrire l'événement avant de le publier ne propose pas un choix d'options, via une liste déroulante qui permettrait en premier lieu d'avoir une base de données un peu plus lisible et interprétable. Au lieu de cette manière, chaque utilisateur peut rentrer à la main ce qu'il veut dans chaque champs... chacun rentre les informations comme il l'entend. Il n'y pas de restriction pour remplir les champs par les utilisateurs qui crée des agendas et donc pas d'harmonisation entre les données.</p> <p>La forme des données récupérées n'est pas adaptée aux traitements de données, et leur exploitation est donc impossible en l'état.</p> <p>(Nous aurions souhaité représente les événements selon leur type (concert, vernissage, théâtre, don du sang etc) au cours du temps, pour observer si un type d'événements c'est plus développé plutôt qu'un autre, mais il n'y a pas de champs type... Les 10 000 données sont décrites via un champ « description » mais il est inexploitable. Les données ne sont pas atomiques, il est donc à nous de créer le champ « Type événement » pour pouvoir effectuer des traitements statistiques.)</p> | <p>Dans cette étape-ci, il s'agit de nettoyer la base de données pour pouvoir l'exploiter.</p> <p>Cependant, la quantité de données rend la tâche difficile à automatiser. Nous avons pensé à un code python qui permettrait à parti de mots clés trouvé dans la colonne description d'associer une catégorie d'événement pour le champ « Type d'événement » mais comme chaque utilisateur défini un même type d'événement de manière très différente avec de multiples mots clés cette tâche est extrêmement difficile. (Par exemple un concert peut être décrit par un mot clé « concert », « musique », « dj », « fête musicale »).</p> <p>Notre solution serait de faire une proposition, aux responsables du site d'OpenAgenda, quant à la mise en place d'une restriction dans le remplissage des champs lors de la publication d'un événement.</p> |
|------------------------------------|--|---|

|   |  |   |
|---|--|---|
| <p>La nature du rendu du projet :</p>     | <p>Suite aux différents obstacles rencontrés lors de l'acquisition des données, nous avons dû remodeler notre rendu et notre projet afin de satisfaire notre commanditaire.</p>  | <p>Créer un document expliquant que les données doivent être atomiques pour pouvoir les traiter et expliquant notre nouvelle stratégie pour pouvoir répondre à la commande.</p> <p>Nous allons utiliser le logiciel carto.com. Le petit document que nous lui remettrons le 17 décembre, expliquera les fonctionnalités que nous allons utiliser avec en exemple des données que nous avons récupéré et trier (il s'agit d'un petit agenda avec peu d'événement sur la fête de la musique en 2015 que nous avons pu efficacement remettre en forme).</p> <p>Faire le travail sur un petit échantillon de données nous permet d'indiquer à notre commanditaire comment nous voyons les choses et ce que nous pourrions faire si la base de données était mieux conçue.</p> |
| <p>Retard dans la gestion de projet :</p> | <p>L'analyse statistique de ses données, les corrélations avec d'autres variables devient complexes du fait que la base de données n'a pas été bien pensée...</p>  | <p>Réaménager avant le 10 décembre notre rendu et nos objectifs. En faire part à notre commanditaire avant le 12 décembre.</p> <p>Notre planning de tâches est modifié.</p>   |
| <p>Attentes du commanditaire:</p>         | <p>Notre commanditaire n'a jamais travaillé auparavant avec des géomaticiens. Il entend ce projet comme une réelle opportunité pour nous, étudiants, de se créer un réseau, de travail en mairie avec des collectivités locales. Ses demandes ne sont pas très précises et ses propositions de commande évoluent assez fréquemment. La difficulté ici, est que nous nous sommes rendu compte qu'il ne sait pas, lui-même, exactement ce qu'il attend de nous hormis que nous développons notre réseau via la création de compte Twitter pour partager l'avancée du projet...</p> | <p>Définir avec lui exactement ce que nous n'allons pas faire.</p>  |

Voici donc la nouvelle répartition des tâches dans le temps puis la répartition des tâches au sein du groupe :



|   | Axel | Bénédicte | Salomé | Thomas |
|---|------|-----------|--------|--------|
| Document texte présentant les analyses possibles / comment refaire la base de données pour plus d'efficacité par la suite pour OpenAgenda |      | X         | X      |        |
| Exemple de représentation d'analyse via carto.com   | X    |           |        | X      |
| Nettoyer la base de données pour créer un champs "Type d'événements"  | X    | X         | X      | X      |
| Effectuer les traitements statistiques sur la base de données nettoyée  | X    | X         | X      | X      |

## Etude du résultat : Open Agenda (3eme TD)

### Comparer le produit final à la commande initiale

Rappelons tout d'abord nos objectifs initiaux : OpenAgenda spécialisé dans l'évènementiel possède une grande base de données constituée d'informations relatives des événements implémentés sur le site (nom de l'événement, lieu, date, horaire, tarif etc). Notre mission est de réaliser une analyse spatiale pour valoriser leur base de données et montrer qu'elle est riche en information. Pour nous guider dans ce projet notre commanditaire nous avait soumis certaines pistes d'analyses à explorer. Par exemple, observer la spatialité des événements par rapport au réseau de bus, au prix aux mètres carrés des arrondissements. En réalité notre travail consiste à attirer les acteurs publics afin qu'ils exploitent la base de données OpenAgenda pour réaliser leur projet et qu'ils n'aillent pas récupérer des bases issues d'autres organismes.

Initialement, nos objectifs étaient de façonner un rendu sous forme de Webmapping. Dans l'idéal cette application devait offrir la possibilité à l'utilisateur de sélectionner un territoire (ville, par exemple) pour obtenir les statistiques liées aux événements. Par exemple : quel quartier a le plus d'événements l'été ? De quels types sont-ils ? Vernissages, concerts, dégustations de vins, etc. Un bouton recherche par type d'événements devra également être établi. Il s'agit de faire une carte interactive qui lie nos statistiques avec de la cartographie en sélectionnant une zone d'étude.

Dans le temps imparti, nous avons pu développer 3 cartes interactives : Une qui propose à l'utilisateur de trier les événements par catégorie (concert, religion, spectacle, exposition) par prix, par mois par années, par lieux avec visualisation du nombre d'événement par quartier. Ensuite, une seconde carte permet de visualiser la spatialité des événements par rapport au prix aux mètres carrés des arrondissements parisiens. Enfin, une dernière carte interactive permet d'analyser spatialement les événements par rapport à la densité de la population en fonction du temps et du type d'événements.

Par ailleurs, nous avons poussé l'étude en réalisant des cartes statiques comportant d'autres pistes d'études que nous n'avions pas définies dans nos objectifs au début du projet. Par exemple, nous avons étudié l'accessibilité des événements par rapport au réseau de transport Parisiens (sélection seulement des transports Métros et RER). En outre, nous avons également croisé la spatialité des événements qui se déroulent à partir de 21h, à l'extérieur avec l'éclairage public et analyser l'accessibilité des poteaux incendies par rapport aux événements pour par exemple, aider les organismes de sécurité des événements à savoir si la population fréquentant les événements est en sécurité ou non.

En outre, nous avons fait le choix de nous intéresser aux fonctionnalités proposées par ArcGisPro qui présente notamment un module pour automatiser la création de carte à l'aide de codes Python.

Nous avons établi deux codes avec Arc Python que nous n'avions pas prévu de réaliser auparavant. L'un d'eux permettent de créer la carte de chaque arrondissement avec le compte des évènements s'y déroulant et leurs catégories. Le second permet de créer une carte par département français avec la catégorisation des prix des évènements.

Notre produit final est donc plus complet que nous l'aurions pensé même si celui-ci présente quelques biais. Par exemple, lors du nettoyage de la base de données nous avons décidé restreindre encore la zone d'étude. Auparavant elle était de Paris et sa grande couronne, finalement nous nous sommes limités à Paris. Par ailleurs, nous avons dû repenser le modèle de base données, face à une mauvaise conception de la base de données, ce qui n'était pas prévu dans le rendu initial. Ainsi, nous avons créé cinq catégories d'évènements. Cependant, cette limite catégorie restreint l'analyse car en réalité il y a bien plus de catégories d'évènements. Nous avons créé ces catégories car il s'agit de celles qui regroupent le plus d'évènements.

En revanche, une limite de notre prévision initiale concerne les statistiques. Nous pensions avoir le temps de réaliser des statistiques inférentielles or nous sommes restés dans des statistiques purement descriptives. Certes, cela donne une idée de comment se comporte les données mais elles auraient pu être plus poussées. Ainsi notre produit final reste assez proche de la commande initiale, avec des améliorations que nous avons pu apporter.

Comparer les étapes et charges initiales avec les étapes effectivement réalisées et avec la charge effective et analyse des écarts par rapport au début du projet.

Tout au long de ce projet nous avons su respecter notre tableau de bord tout en le faisant évoluer suite aux difficultés rencontrées lors de la récupération de la base de données. Ces contre temps nous ont poussé à avoir un nouveau regard sur le projet. Il a fallu repenser les étapes de traitement et prévoir du temps pour les résoudre et réorganiser les tâches à effectuer.

Par exemple, pour la récupération de données, cette étape a bloqué le projet pendant 3 semaines. Nous avons par conséquent décidé de passer plus temps à analyser l'architecture des données et repenser le modèle plutôt que de faire des statistiques poussées sur les données. En effet, il nous a semblé plus important d'accorder plus de temps à cette étape que prévu, car pour OpenAgenda s'ils souhaitent exploiter la base sur le long terme mieux vaut avoir un modèle de base de données cohérent et performant.

Enfin, l'analyse spatiale des données a débuté beaucoup plus en retard que prévu, les cours sur cette thématique débutant trois semaines avant le rendu final. Mais cela nous a permis de découvrir de nouvelles méthodes et de nouveaux axes de recherche dans le cadre de notre projet OpenAgenda. Par exemple l'aire d'influence des transports en communs (métro et RER) sur l'organisation spatiale des évènements.

Les tâches effectuées ont été plus approfondies et détaillées au fur et à mesure du projet. Les difficultés rencontrées nous ont aussi permis de cibler certaines problématiques au sein du projet et ainsi permettre de rendre un produit final plus abouti.

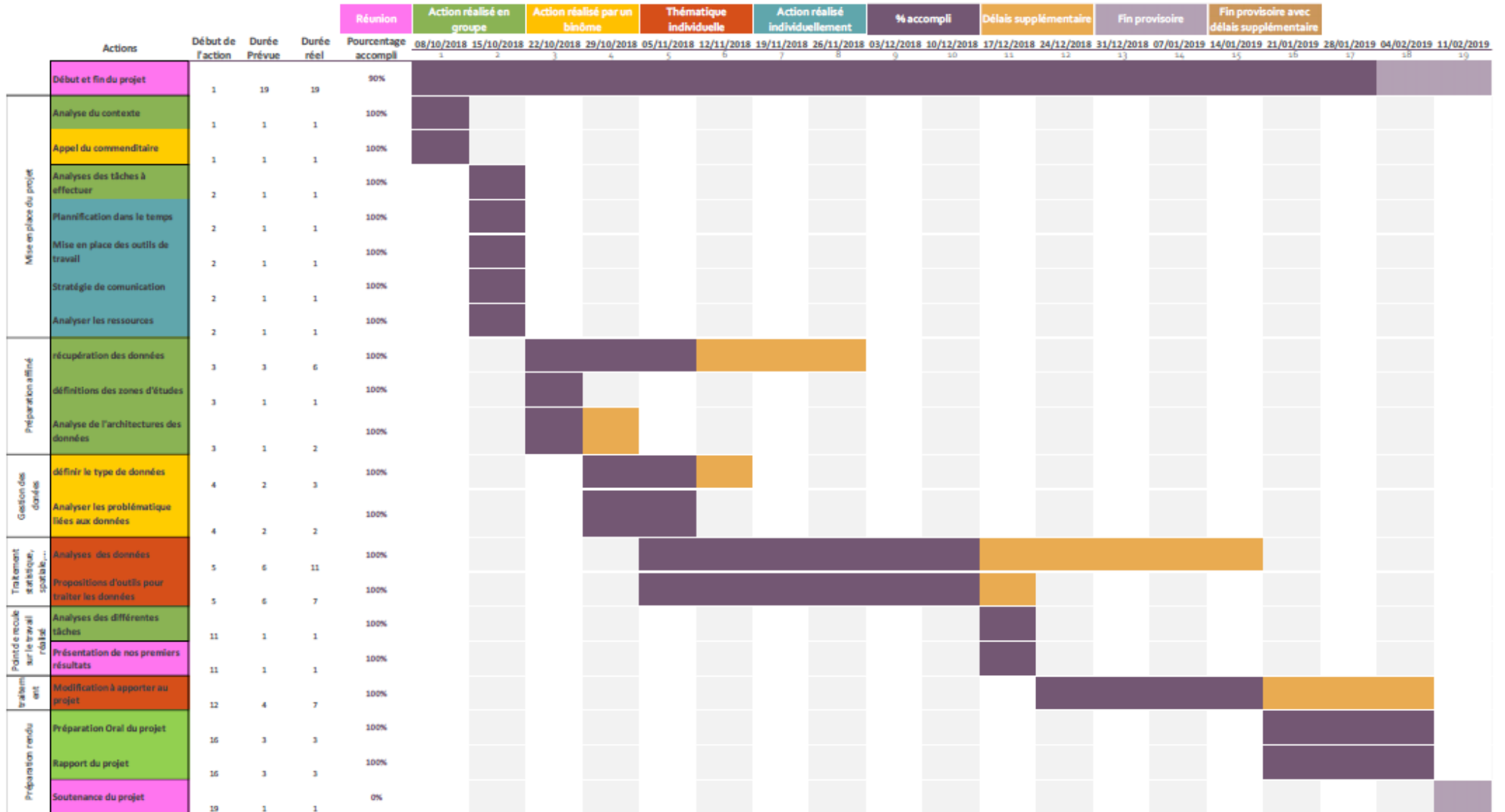
## Proposition de solutions pour une expérience similaire

Si nous devions recommencer le projet, il aurait fallu être moins ambitieux sur la gestion des tâches, et rajouter une tâche qui concernerait une capitalisation d'expériences sur des projets similaires pour nous aider à analyser et exploiter une base de données similaire. Effectivement, nous avons pensé notre modèle de répartition des tâches sans tenir compte que nous pourrions rencontrer des difficultés qui retarderaient et mettraient entre parenthèse l'avancée du projet. Par ailleurs, nous planifierions directement des rendez-vous par vidéo-conférence ou présentiel avec notre commanditaire. Nous n'avons échangé que par mail avec celui-ci avec une visioconférence en début de projet et parler de vive voix avec son commanditaire permet de clarifier certaines choses et plus enrichissant pour les suivis de projet.

Malgré tout, nous avons su faire face aux obstacles qui sont apparues au cours de notre projet. L'équipe est restée soudée dans l'adversité. Ces différentes difficultés ont en réalité permis de repenser le projet en proposant un produit fini conforme aux attentes du commanditaire et plus abouti que le produit initial.

# Déroulement du projet OpenAgenda

informations sur le code couleur



## 4.2 Code python réalisé avec ArcGIS pro

### 4.2.1 Code python : représentation des types d'événements

```
1.
2. # -*- coding: cp1252 -*-
3.
4. #Importation arcpy, l'outil qui permet de rentrer en communication
   avec ArcGisPro
5. import arcpy,os, pathlib
6. from arcpy import env
7. from pathlib import Path
8.
9. #Précision du chemin de notre workspace
10. chemin_entree=
   'C:\\Users\\tb02754t\\Desktop\\projet_python_gonon_thomas\\projet_py
   hon_gonon_thomas'
11. print(chemin_entree)
12. #Chemin du projet arcgisPro
13. projet = chemin_entree + '\\benedicte_leonie.aprx'
14. print(projet)
15.
16. #Emplacement des données mobilisées
17. arcpy.env.workspace = chemin_entree + '/donnees'
18.
19. #Nom des couches mobilisées
20. arcpy.MakeFeatureLayer_management("Paris.shp","evenements.shp")
21.
22. #Remplacement des fichiers existants
23. arcpy.env.overwriteOutput = True
24.
25. #Définition du chemin au projet arcgisPro
26. aprx = arcpy.mp.ArcGISProject(projet)
27.
28. #Fonction de mise en page automatique des événements par
   arrondissement
29. def SortieCarte(code_arrondissement):
30.     #Paramétrage de la mise en page à utiliser
31.     m=aprx.listMaps()[0]
32.     lyr = m.listLayers()
33.     lyt = aprx.listLayouts ("Layout")[0]
34.     #Automatisation du titre en fonction de l'arrondissement
   représenté
35.     aidetitre=str (code_arrondissement)
36.     print(lyr)
37.     #Corps de la fonction
38.     for couche in m.listLayers():
39.         print(couche.name)
40.
41.         if couche.name == 'Limite arrondissement':
42.             print(couche.name)
43.             #Requête pour représenter un seul arrondissement sur la
   mise en page
44.             query = 'INSEE_COM = ' + "'" +
   str(code_arrondissement)+"'"
45.             print(query)
46.             #Paramétrage ensemble de définition
47.             couche.definitionQuery = query
48.             print(couche.definitionQuery)
49.             arrondissement = couche
50.
```



```

51.         elif couche.name == 'evenements':
52.             print(couche.name)
53.             #Requête pour représenter les événements de
l'arrondissement sélectionné auparavant sur la mise en page
54.             query1 = 'Arrondisse = ' + "" +
str(code_arrondissement)+"''
55.             print(query1)
56.             #Paramétrage ensemble de définition
57.             couche.definitionQuery = query1
58.             print(couche.definitionQuery)
59.             arrondissement = couche
60.
61.             #Paramétrage des éléments de la mise en page
62.             mf = lyt.listElements()
63.             for elm in mf:
64.                 print(elm.name)
65.             #Zoom sur l'arrondissement représenté
66.             frame = lyt.listElements("mapframe_element")[0]
67.             frame.camera.setExtent(frame.getLayerExtent(arrondissement,
False, True))
68.             #Titre dynamique en fonction du code insee de l'arrondissement
69.             Titre= lyt.listElements("TEXT_ELEMENT","Titre2")[0]
70.             Titre.text= aidetitre
71.             #Export en jpeg
72.             lyt.exportToJPEG(chemin_entree + (code_arrondissement) + ".jpg")
73.
74. #Automatisation de la manipulation sur les 20 arrondissements
75. def tousCodeesArrondissementParis():
76.     liste =
['75101','75102','75103','75104','75105','75106','75107','75108','751
09','75110','75111','75112','75113','75114','75115','75116','75117','
75118','75119','75120']
77.     for arron in liste:
78.         #Appel de la fonction mise en page
79.         SortieCarte(arron)
80. #Appel de la fonction d'automatisation
81. tousCodeesArrondissementParis()
82.
83. #-----COMMENTAIRE AMELIORATION DU CODE-----
-----
84. # Pour compter le nombre d'événements par catégorie au sein de
chaque événements, nous avons tout d'abord envisagé de mobiliser la
fonction arcpy.SummarizeWithin_analysis.
85. #Puis,nous avons décidé de passer par un cursor qui
comptabiliserait dans une liste la fréquence des types d'événements
par arrondissement.
86. #N'ayant pas réussi, à ce jour, à leur faire fonctionner
convenablement veuillez trouver ci-dessous notre code d'essai.
87. #def CompteCat(req_arrondissemnt):
88.
89.     #cat = ['Religion','Concert','Spectacle','Exposition']
90.     #fc =
'C:\Users\tb02754t\Desktop\projet_python_gonon_thomas\projet_python_g
onon_thomas\donnees\evenements.shp'
91.     #print(fc)
92.     #field="Categorie"
93.     #cursor=arcpy.da.SearchCursor(fc,field)
94.     #cursor = arcpy.(req_arrondissemnt)
95.
96.     #for row in cursor:
97.

```

```

98.         #i0 = 0
99.         #i1 = 0
100.        #i2 = 0
101.        #i3 = 0
102.        #if row[0] == cat[0]:
103.            #i0 += 1
104.        #if row[0] == cat[1]:
105.            #i1 += 1
106.        #if row[0] == cat[2]:
107.            #i2 += 1
108.        #if row[0] == cat[3]:
109.            #i3 += 1
110.
111.        #return i0, i1, i2,i3
112.        #print(i0,i1,i2,i3)
113.
114. #CompteCat()

```

#### 4.2.2 Code python : analyse du prix des événements dans un département

```

1. #CODING:cp1252
2. #PROJT PYTHON ENTIER
3.
4.
5. #DEFINTION DES BIBLIOTHEQUE
6. import arcpy
7. import shutil
8. import os
9. import re
10. import time
11.
12. #DEFINITION DES VARIABLES
13. #DEFINTION de l'emplacement des dossiers du projet
    / ! \ A REFAIRE
14. localisation_dossier_general = "C:/Users/fa02232t/Documents"
15. arcpy.env.workspace =
    localisation_dossier_general+"/PROJET_PYTHON/DOSSIER_PROJET/Default.g
    db"
16.
17. #FICHER DE REFERENCE POUR LES DEPARTEMENTS
18. couche_departement =
    localisation_dossier_general+"/PROJET_PYTHON/COUCHE_DE_BASE/DEPARTEME
    NT3/Departement_france.shp"
19. #conversion du fichier shp en lyr pour pouvoir réaliser les
    géotraitement en python
20. #arcpy.MakeFeatureLayer_conversion(source_couche_departement,"couch
    e_departement")
21.
22.
23.
24. #FICHER TABLEAU AVEC LES EVENEMENTS
25. fichier_evenement =
    localisation_dossier_general+"/PROJET_PYTHON/FICHER_EVENEMENT/evenem
    ent_couche_creeer_L93.shp"
26. #conversion du fichier shp en lyr pour pouvoir réaliser les
    géotraitement en python
27. #arcpy.MakeFeatureLayer_conversion(source_fichier_evenement,"fichie
    r_evenement")

```

```

28.
29.
30.
31. #FICHIER ENREGISTREMENT DES SHAPEFILE CREER
32. dossier_shapefile =
    localisation_dossier_general+"/PROJET_PYTHON/COUCHE_EXPORTER_PAR_PROG
    RAMME/"
33. os.mkdir(dossier_shapefile)
34. #FICHIER ENREGISTREMENT DES CARTES FINALE
35. dossier_carte =
    localisation_dossier_general+"/PROJET_PYTHON/CARTE_EN_JPEG/"
36.
37. #DEFINITION DES VARIABLES LIEE AUX PROGRAMMES PYTHON
38. departement_correct = 0
39.
40.
41. #envoi une demande d'indication du prix à etudier
42. prix_indique = input("Entrer le prix maximum de l'etude :")
43. #Indique le prix inique avant
44.
45. print("vous avez renseigné %s ",prix_indique)
46.
47. #envoi une demande d'indication du département à etudier
48. departement_indique = input("Entrer le nom ou code postale du
    departement à étudier :")
49. #Test si le departement indiquer à un espace afin de
    le remplacer par un tiret (pour correspondre correctement à notre
    table département)
50. print("vous avez renseigné %s ",departement_indique)
51. #if departement_indique.find(' ') != -1:
52. #    departement_indique_sans_tiret =
    departement_indique.replace(' ','-') #on remplace les espace par des
    tirets
53.
54.
55. #FONCTION selection département par rapport au département indiqué
56. def selection_departement_prix (dept,prix):
57.     #VARIABLE EN FONCTION département
58.     #departement_select =
    localisation_dossier_general+"/PROJET_PYTHON/COUCHE_DE_BASE/DEPARTEME
    NT3/"+dept+".shp" A VOIR
59.     Output =
    localisation_dossier_general+"/PROJET_PYTHON/COUCHE_EXPORTER_PAR_PROG
    RAMME/"+dept+".shp"
60.     #intersectDeptOutput =
    localisation_dossier_general+"/PROJET_PYTHON/COUCHE_EXPORTER_PAR_PROG
    RAMME/"+dept+".shp"
61.
62.
63.     #CONDITION EN FONCTION DU dept
64.
65.     conditionsDEPT = '"code_insee" = \''+dept+\'"'
66.
67.     #TRANSFORMATION SHP EN LYR ET SELECTION dept en fonction de la
    variable departement [variable au format LYR]
68.     Selection_departement_lyr = arcpy.MakeFeatureLayer_management
    (couche_departement, 'Selection_departement_lyr' , conditionsDEPT)
69.
70.     #VARIABLE pour indiquer l'emplacement des fichiers

```

```

71.     evenement_prix =
    localisation_dossier_general+"/PROJET_PYTHON/FICHIER_EVENEMENT/"+prix
    +".shp"
72.     intersectPRIXETDEPTOutput =
    localisation_dossier_general+"/PROJET_PYTHON/COUCHE_EXPORTER_PAR_PROG
    RAMME/EVENEMENT_TEMP.shp"
73.
74.     #CONDITION EN FONCTION DU PRIX / ! \ par la suite remplacer le
    Identifian par la colone prix du SHP une fois le traitement de texte
    réussi
75.     conditionsPRIX = '"Prix" <= '+prix
76.
77.     #SELECTION evenement en fonction de la variable prix [variable
    au format LYR]
78.     evenement_lyr = arcpy.MakeFeatureLayer_management
    (fichier_evenement, 'evenement_lyr' , conditionsPRIX)
79.
80.     #Création du tableau de couche (en format lyr) des événement
    et du département sélectionné
81.     inFeatures = [evenement_lyr,Selection_departement_lyr]
82.     #REQUETE INTERSECTION en enregistrement de la couche au format
    SHP
83.     arcpy.Intersect_analysis(inFeatures,
    intersectPRIXETDEPTOutput,"ALL")
84.
85.
86. def Mise_en_page (dept,prix):
87.     conditions = '"code_insee" = '+'+"'+dept+"'"
88.     output_lyr = "lyr_Departement_Select"
89.
90.     #CREATION CARTE
91.     aprx =
    arcpy.mp.ArcGISProject(r"C:/Users/fa02232t/Documents/PROJET_PYTHON/DO
    SSIER_PROJET/PROJET_BASE.aprx")
92.     map = aprx.listMaps("Carte")[0]
93.
94.     mise_en_page = aprx.listLayouts()[0]
95.     carte = mise_en_page.listElements("MAPFRAME_ELEMENT")[0]
96.
97.
98.
99.     lyr = map.listLayers()[0]
100.    lyr.definitionQuery = conditions
101.    carte.camera.setExtent(carte.getLayerExtent(lyr))
102.
103.
104.    for elm in mise_en_page.listElements("TEXT_ELEMENT"):
105.        if elm.text == "Source " :
106.            elm.text = "Réalisation par Yankiu Mok et Axel
    Fourneyron"
107.        if elm.text == "Nom de la carte ":
108.            elm.text = "Evénement dans le département du
    "+dept+" inférieure à "+prix+"€"
109.
110.
111.    #export de la carte en PDF
112.    carte_output =
    dossier_carte+"Carte_du_"+dept+"_avec_prix_de_"+prix+".pdf"
113.    mise_en_page.exportToPDF(carte_output)
114.
115.    #Fin de la mise en page

```

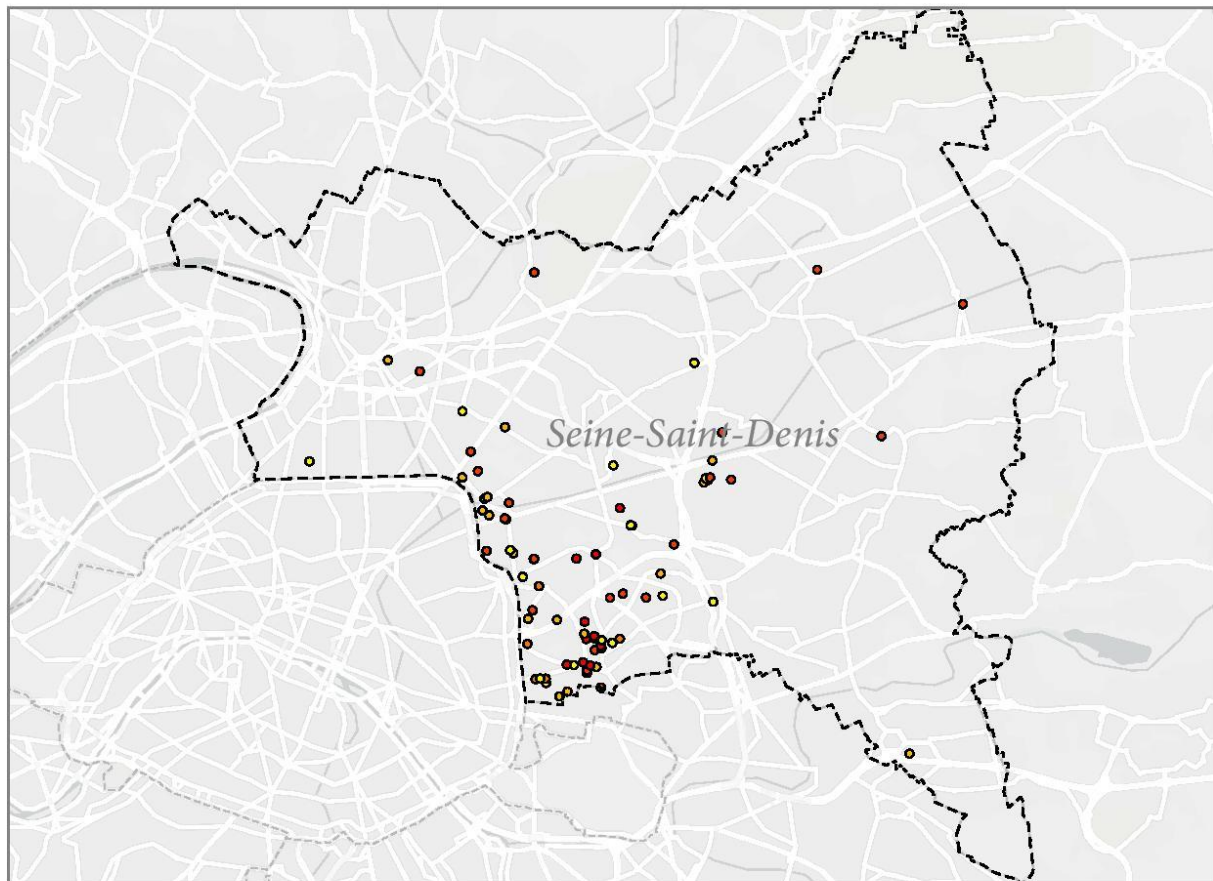
```

116.         print("Fermeture du projet en cours")
117.         arcpy.Delete_management(dossier_shapefile)
118.         print("Suppression des fichiers temporaire")
119.
120.
121. #On vérifie si le département renseigné est bien dans notre table
122. while departement_correct != 1:
123.
124.     #Recherche dans la couche departement : departements-
        20180101.shp
125.     cursor = arcpy.SearchCursor(couche_departement)
126.     row = cursor.next()
127.     #Tant qu'il y a une nouvelle donnée alors (ça tourne en boucle)
128.     while row:
129.         #on test si dans la table département il y a le département
            indiquer
130.         if row.getValue("nom").lower() ==
            departement_indique.lower():
131.             #Si le nom du département indiquer et présent alors on
                passer la variable à 1 pour sortir de la boucle
132.                 departement_correct = 1
133.                 code_insee = row.getValue("code_insee")
134.                 #si le nom du département n'est pas indiquer alors on
                    tester avec son code insee
135.                 elif row.getValue("code_insee") == departement_indique :
136.                     departement_correct = 1
137.                     code_insee = row.getValue("code_insee")
138.                 # on passe à la donnée suivante
139.                 row = cursor.next()
140.
141.         if departement_correct == 0:
142.             print ("nous n'avons pas trouvé %s",departement_indique)
143.             departement_indique = input("Rechercher une autre
                departement :")
144.         else :
145.             print ("nous avons bien trouvé %s",departement_indique)
146.             intersectDeptOutput =
                localisation_dossier_general+"/PROJET_PYTHON/COUCHE_EXPORTER_PAR_PROG
                RAMME/"+departement_indique+".shp"
147.
148. selection_departement_prix(code_insee,prix_indique)
149. Mise_en_page(code_insee,prix_indique)
150. print("Travail terminé merci")
151.

```

#### 4.3 Carte générée par le code

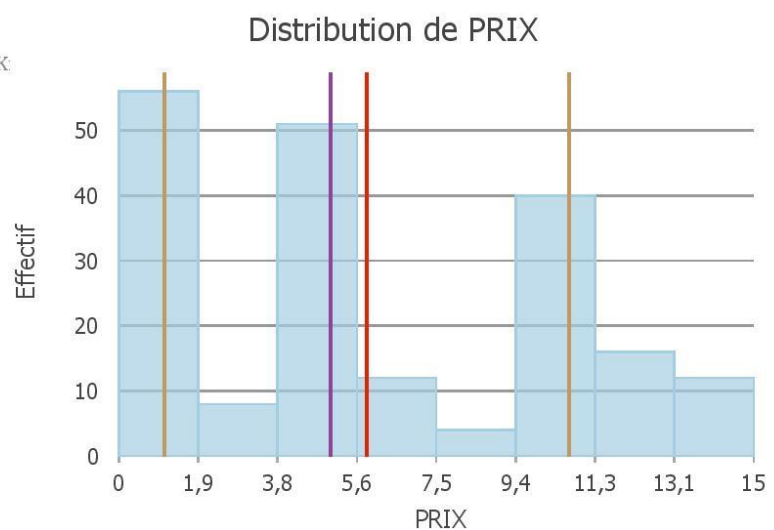
### Événement dans le département du 93 inférieure à 12€



--- Limites départementales

RIX

- ≤0,0
- ≤5,0
- ≤9,0
- ≤12,0
- ≤15,0



— Moyenne : 5,9 — Médiane : 5 — Ecart type : 4,8

Réalisation par Yankiu Mok et Axel Fourneyron