

Rapport de stage

Pour le diplôme d'ingénieur de l'Ecole des Ingénieurs de la Ville de Paris

Année 2025-2026

Coordonner l'Europe, apaiser la ville

Du pilotage du projet REALLOCATE à l'étude des perceptions de la ZTL Paris Centre

GOPALAPILLAI Gokul

Promotion 65

Février à Juillet 2026

Sous la direction de :

SONET Ungku Norani

SNITER Valérie

Notice bibliographique

AUTEUR du mémoire	
NOM	GOPALAPILLAI
Prénom	Gokul
ORGANISME de stage	
NOM organisme	University College Dublin, Spatial Dynamics Lab
NOM, Prénom du maître de stage	SONET Ungku Norani
NOM, Prénom du tuteur EIVP	SNITER Valérie
ANALYSE	
TITRE du mémoire	Coordonner l'Europe, apaiser la ville : du pilotage du projet REALLOCATE à l'étude des perceptions de la ZTL Paris Centre
TITLE	Coordinating Europe, Calming the City: From Managing the REALLOCATE Project to Studying Perceptions of the Paris Centre Low Traffic Zone
RÉSUMÉ (15 à 20 lignes)	Ce rapport présente les six mois de stage que j'ai réalisés au Spatial Dynamics Lab de l'University College Dublin, qui coordonne le projet européen REALLOCATE. Il analyse d'abord les réalités de la coordination d'un tel projet, à travers ma contribution au Work Package 3 dédiée à l'aménagement urbain, en mettant en évidence les écarts entre le cadre contractuel et la mise en œuvre sur le terrain, ainsi que la nécessité de composer avec des contraintes scientifiques, opérationnelles et politiques propres à chaque territoire. Le rapport développe ensuite un volet de recherche consacré à l'acceptabilité sociale des politiques d'apaisement du trafic, à travers le cas de la Zone à Trafic Limité de Paris Centre, plus d'un an après son entrée en vigueur. Un premier terrain exploratoire mené à Setia Alam, en Malaisie, m'a ensuite permis de tester et d'ajuster une première approche méthodologique. L'étude repose alors sur le croisement de trois types de données : les perceptions exprimées par les citoyens sur les réseaux sociaux, analysées à l'aide du NLP, la perception visuelle de l'espace public évaluée par un modèle d'intelligence artificielle (VLM), et des indicateurs physiques du territoire tels que le bruit, la qualité de l'air et les flux. L'objectif est de comparer la réalité des transformations du territoire à la façon dont les usagers les ressentent.
ABSTRACT	This report presents a six-month internship carried out at the Spatial Dynamics Lab of University College Dublin, which coordinates the European project REALLOCATE (Horizon Europe), aiming to reallocate public road space in favour of active mobility and public transport. It first analyses the realities of coordinating such a project, through the contribution to Work Package 3, dedicated to urban planning, highlighting the gaps between the contractual framework and the implementation on the ground, as well as the need to deal with scientific, operational and political constraints specific to each territory. The report then develops a research part dedicated to the social acceptability of traffic calming policies, through the case of the Limited Traffic Zone of Paris Centre, more than a year after its entry into force. An initial exploratory fieldwork conducted in Setia Alam, Malaysia, allowed me to test and adjust a first methodological approach. The study then draws on the intersection of three types of data: citizens' perceptions expressed on social media, analyzed using NLP; the visual perception of public space assessed by an artificial intelligence

	model (VLM); and physical indicators of the territory such as noise, air quality, and traffic flows in order to understand how the actual transformation of the is perceived by its users.		
Mots-clés du thésaurus Archirès	acceptabilité sociale ; aménagement urbain ; analyse des réseaux sociaux ; coordination multi-acteurs ; mobilité durable ; perception urbaine ; projet européen ; zone de trafic limité ;		
Keyword thesaurus (anglais, 8 maxi)	european project; multi-actor coordination; limited traffic zone; social acceptability; social media analysis; sustainable mobility; urban perception; urban planning;		
Références			
	Nb de pages du mémoire	Annexes (nombre de p.)	Bibliogr. : nb de références
	37	17	31

Remerciements

Je tiens à exprimer ma profonde gratitude à Monsieur Pilla, co-directeur du laboratoire *Spatial Dynamics* à University College Dublin (UCD), et Madame Sonet, professeure et chercheuse, pour m'avoir offert l'opportunité d'effectuer mon stage TFE dans leur laboratoire.

Je me souviendrai de ces six mois passés à Dublin comme étant une expérience enrichissante tant sur le plan professionnel que personnel. De plus, je reconnais la chance de travailler au sein d'un environnement de recherche international bienveillant et stimulant. Cela m'a permis de développer mon autonomie et d'acquérir une ouverture d'esprit indispensable pour la suite de ma carrière dans la fonction publique à la ville de Paris.

Enfin, j'adresse des remerciements à l'ensemble de l'équipe du laboratoire *Spatial Dynamics* mais aussi à toutes les personnes rencontrées lors de ce séjour en Irlande. Leur accueil, leur bienveillance et les nombreux échanges que j'ai pu avoir avec eux ont largement contribué à rendre cette période plaisante à vivre.

Sommaire

Remerciements	1
Liste des illustrations.....	4
Liste des tableaux	4
Introduction	6
1. Organisme d'accueil (Spatial Dynamics Lab)	8
2. REALLOCATE : fonctionnement du projet et retour d'expérience	9
2.1. Présentation du projet REALLOCATE	9
2.1.1. Contexte, objectifs et inscription dans la Mission Cities 2030	9
2.1.2. Organisation du consortium et stratégie de déploiement.....	9
2.1.3. Financement, cadre contractuel et rôle du coordinateur	10
2.2. Organisation et pilotage du projet	10
2.2.1. Architecture en Work Packages et rôle du WP3.....	10
2.2.2. Gouvernance et coordination du consortium	11
2.2.3. Validation des livrables, suivi budgétaire et indicateurs.....	11
2.3. Mes missions au sein du WP3	12
2.4. Retour d'expérience : entre cadre européen et réalités du terrain	13
2.4.1 Retour d'expérience : entre cadre européen et réalités du terrain.....	13
2.4.2 Cultures professionnelles, jeux d'acteurs et évolution vers le volet recherche	14
3. Phase exploratoire et enseignements méthodologiques : le cas d'étude de Setia Alam (Malaisie)	17
3.1. Analyse des sentiments sur Twitter/X : écouter la voix des habitants	17
3.1.1. Pourquoi Twitter/X ?	17
3.1.2. Protocole en trois étapes	17
3.1.3. Résultats	18
3.1.4. Limites et enseignements de l'approche	19
3.2. Évaluation perceptive du cadre bâti par vision par ordinateur (Google Street View)...	20
3.2.1. Principe et justification de la vision par ordinateur	20
3.2.2. Protocole de collecte	20
3.2.3. Résultats : tendances et limites	21
3.3. Bilan et réorientation méthodologique	22
4. Élaboration du protocole méthodologique final : le cas de la ZTL de Paris Centre	23
4.1. Présentation du cas d'étude et justification du choix de Twitter/X en France	23
4.1.1. La ZTL de Paris Centre : périmètre, objectifs et calendrier.....	23
4.1.2. Justification du terrain au regard des leçons de Setia Alam	24
4.1.3. Justification du choix de Twitter/X comme source de données.....	24

4.2. Correction du volet perception visuelle : passage au relevé photographique in situ.....	25
4.2.1. Correction du mode d'acquisition	25
4.2.2 Première lecture des scores perçus	26
4.2.3. Contreparties et limites résiduelles	27
4.3. Ingénierie de collecte textuelle : stratification, apprentissage actif et filtrage	27
4.3.1. Contexte et principes méthodologiques	27
4.3.3. Collecte finale et pipeline de nettoyage	28
4.4. Architecture d'analyse NLP : triangulation multi-modèles et détection de l'ironie.....	29
4.4.1. Prétraitement du texte	29
4.4.3. Modèles mobilisés (triangulation)	29
4.4.3. Détection heuristique de l'ironie	29
4.5 Programme de visualisation et exploitation des séries spatio-temporelles.....	30
4.5.1. Dynamique temporelle des perceptions	30
4.5.2. Structure thématique du débat	31
4.5.3. Dimension spatiale : le sentiment par lieu	32
4.5.4. Robustesse et contrôles de qualité	33
4.6. Protocole de confrontation entre données perçues et mesures physiques objectives	33
4.6.1. Justification de la démarche	33
4.6.2. Harmonisation et construction du panneau	34
4.6.3. Analyses prévues	34
4.7. Bilan critique, limites et reproductibilité de la démarche.....	35
Conclusion.....	36
Bibliographie et Sitographie.....	37
Documentation officielle du projet européen.....	37
Annexes	39

Liste des illustrations

Figure 1 - Cartographie de la stratégie de jumelage et des groupes d'apprentissage entre villes européennes traduite en français	9
Figure 2 - Architecture globale des Work Packages traduite en français Source : Reallocalemobility.eu.....	11
Figure 3 - Représentation cartographique des trois échelles d'analyse spatiale des tweets (C1, C2, C3).....	18
Figure 4 - Distribution du sentiment par cercle.....	18
Figure 5 - Exemples représentatifs des photographies capturées illustrant le cadre bâti de Setia Alam.	21
Figure 6 - Carte de localisation des points de capture Google Street View à Setia Alam.	21
Figure 7 - Carte du périmètre de la ZTL Paris Centre Source : mairiepariscentre.paris.fr	24
Figure 8 - Carte des points de prise de vue à Paris et itinéraires parcourus.	26
Figure 9 - Évolution du sentiment moyen (ZTL Paris Centre), janvier 2023 – juin 2026. Courbe hebdomadaire lissée (moyenne mobile sur 4 semaines), annotée des dates clés : enquête publique (avril 2024), ouverture des JOP (26/07/2024), arrêté ZTL (31/10/2024), ent	30
Figure 10 - Étude d'événement : sentiment moyen ± 30 jours autour des dates clés, avec intervalles de confiance à 95 %.	31
Figure 11 - Distribution du sentiment par thème (boxplots).	31
Figure 12 - Sentiment moyen par lieu du périmètre ZTL	32

Liste des tableaux

Tableau 1 -Étapes du pipeline de nettoyage et justification scientifique de chaque filtrage. ..	28
---	----

Liste des annexes

Annexe I: Tableau de présentation des différents Labs de chaque ville pour le projet Reallocate traduite en français Source : Reallocalemobility.eu	39
Annexe II: Cartographie des domaines d'expertise et répartition des partenaires auprès des villes du projet REALLOCATE traduite en français Source : Reallocalemobility.eu.....	39
Annexe III: Captures d'écran illustratives de la Storymaps principale (Des Storymaps complémentaires ont été créées pour chaque partie afin d'enrichir le propos sans la surcharger)	40
Annexe IV: Résultats complémentaires de l'analyse des sentiments et des émotions sur Twitter/X pour l'étude de Setia Alam (avec la répartition des émotions, l'évolution temporelle du sentiment moyen et la répartition du sentiment par thème et sous-thème climatique)	45
Annexe V: Carte des scores de beauté et d'ambiance urbaine perçue (Beautiful et Lively) le long des axes couverts à Setia Alam sur Google Streetview	47
Annexe VI: Les histogrammes détaillés des scores pour les six adjectifs des captures d'écran prises sur Google Streetview pour le corpus de Setia Alam	48
Annexe VII: Histogrammes des six scores (Beautiful, Boring, Depressing, Lively, Safety, Wealthy) pour le corpus parisien	48
Annexe VIII: Cartes de scores (Beautiful, Lively et Safety) produites sous ArcGIS Pro	49

Annexe IX: Programme Python de collecte et de filtrage de tweets utilisés pour l'étude de la ZTL de Paris.....	50
Annexe X : Streamgraph des émotions en parts normalisées	54
Annexe XI: Régimes émotionnels dans le temps	54
Annexe XII: Distribution du sentiment : Avec vs Sans Comptes Suspects	55
Annexe XIII: Matrice de corrélations de Spearman	55
Annexe XIV: Heatmap jour x heure.....	56

Introduction

Les enjeux de mobilités durables sont devenus une préoccupation dans le développement de la ville de demain. Dans le contexte mondial d'urgence climatique et de transition écologique. Portées par les orientations du Pacte Vert européen et de l'objectif de la neutralité carbone d'ici 2050, la transformation des métropoles est devenue une priorité politique majeure. C'est pourquoi, cela passe par réduire la prépondérance de la voiture individuelle dans les villes. C'est dans cette dynamique que s'inscrit mon stage de six mois effectués au sein du laboratoire Spatial Dynamics de l'University College Dublin (UCD). Ce laboratoire de recherche est spécialisé dans l'analyse spatiale, la modélisation géographique et l'étude des transformations urbaines. Sous la supervision de Madame Sonet, professeur-chercheuse dans ce laboratoire, j'ai réalisé la majeure partie de mes missions autour du projet européen Reallocate EU. En effet, c'est un projet, financé par l'Union européenne, réunissant un réseau de villes et de partenaires techniques visant à concevoir, tester et évaluer des stratégies de réallocation de l'espace public routier au profit des mobilités douces, des transports collectifs et des espaces apaisés. La mise en commun des données et leurs retours d'expérience, les partenaires du projet posent les bases de méthodologies transposables à l'ensemble du territoire européen.

Au sein de ce programme, mes missions initiales ont consisté à apporter un soutien organisationnel et technique au Work Package 3, et plus particulièrement dans le cadre de la Task 3.1, dédiée à la définition du cadre d'analyse spatiale et à la collecte de données de diagnostic initial pour la réallocation de l'espace public. Elles ont notamment porté sur la participation aux réunions de suivi entre les différents partenaires, la conception de supports de présentation (PowerPoint) ainsi que le développement d'outils géospatiaux interactifs via la suite ArcGIS, notamment des cartes narratives (*StoryMaps*) et des formulaires d'enquête (*Survey*).

Si ces tâches ont permis d'appréhender le fonctionnement d'un grand projet européen, il nécessitait de les compléter afin de répondre pleinement aux exigences demandées pour le stage TFE. Il m'a notamment été proposé de contribuer directement à la rédaction d'articles scientifiques en étudiant un cas à Setia Alam (Malaisie) mais surtout l'acceptation sociale de la Zone à Trafic Limité (ZTL) du centre de Paris (1er au 4e arrondissements), plus d'un an après son entrée en vigueur.

Les problématiques de ce mémoire se s'articulent ainsi autour des questions suivantes : Comment un projet européen d'aménagement urbain concilie-t-il la rigidité de son cadre contractuel avec les temporalités et contraintes propres à chaque territoire ? Dans quelle mesure le croisement entre perception citoyenne (via Twitter) et données physiques de terrain permet-il d'évaluer l'acceptabilité et les impacts réels de la ZTL de Paris ?

Pour répondre à ces deux problématiques, le mémoire suit le fil de mon stage, de sa dimension professionnelle jusqu'à la construction de ma démarche de recherche. Je commence par présenter le Spatial Dynamics Lab de l'University College Dublin, le laboratoire qui m'a accueillie, afin de situer le cadre dans lequel j'ai évolué et de présenter les activités qui ont constitué mon environnement de travail. Je reviens ensuite sur le projet européen REALLOCATE, dont le laboratoire assure la coordination. Cette partie permet de montrer concrètement ce qu'implique le pilotage du projet européen au quotidien, tout en revenant sur les missions que j'ai réalisées au sein du WP3 et sur les enseignements que j'en ai tirés.

Dans un troisième temps, je m'appuie sur le terrain exploratoire réalisé à Setia Alam, en Malaisie, pour présenter les premières pistes méthodologiques que j'ai testées et les ajustements auxquels elles m'ont conduite. Cette expérience a joué un rôle important dans la construction

progressive du protocole de recherche, avant son application au cas d'étude parisien. Enfin, je présente le protocole méthodologique retenu pour étudier la Zone à Trafic Limité de Paris Centre. Celui-ci croise les perceptions exprimées sur les réseaux sociaux, la perception visuelle de l'espace public et différents indicateurs physiques du territoire. Cette dernière partie vient ainsi prolonger les réflexions développées au cours du stage et permet de mettre en regard la perception de la ZTL avec les transformations et les effets observables sur le terrain.

1. Organisme d'accueil (Spatial Dynamics Lab)

Le Spatial Dynamics Lab (SDL) est un centre de recherche rattaché à l'University College Dublin (UCD) en Irlande. Ses travaux s'articulent autour de plusieurs grands axes tels que l'économie évolutionniste, le changement technologique, les politiques environnementales, la planification urbaine, ou encore les villes intelligentes et la croissance durable. La philosophie du laboratoire repose fondamentalement sur l'échange d'idées entre des domaines scientifiques différents mais complémentaires. En mêlant différents domaines de recherche, cette approche multidisciplinaire fait dialoguer les expertises dans le but de faire émerger des idées neuves, de générer de nouvelles découvertes et d'explorer de nouvelles voies de recherches.

Pour mener à bien ses projets, le laboratoire s'appuie sur une équipe internationale de plus de trente membres, réunissant des professeurs-chercheurs, des chercheurs post-doctoraux et des doctorants. Le SDL travaille sur des problématiques variées allant de problématiques locales en Irlande à des enjeux internationaux. Il participe notamment aux programmes TechEvo, consacré à l'analyse des dynamiques d'innovation et de technologie à l'échelle continentale, et iSCAPE, dédié à la lutte contre la pollution de l'air et à l'aménagement durable des villes européennes.

Le laboratoire est co-dirigé par M.Kogler et M.Pilla. C'est d'ailleurs ce dernier qui a reçu ma candidature et qui m'a proposé d'intégrer l'équipe sous la supervision de Mme Sonet, enseignante-chercheuse. Si elle s'implique dans de multiples travaux au sein du laboratoire, mes missions ont consisté à l'assister plus spécifiquement dans le cadre du projet européen Reallocate EU.

2. REALLOCATE : fonctionnement du projet et retour d'expérience

2.1. Présentation du projet REALLOCATE

2.1.1. Contexte, objectifs et inscription dans la Mission Cities 2030

Ce projet est coordonné par l'University College Dublin et s'étend sur quatre ans, de mai 2023 à avril 2027. Il a pour ambition de faire évoluer dans toute l'Europe les pratiques de mobilité en transformant les rues en espaces urbains inclusifs, écologiques, sûrs et adaptés aux enjeux de mobilité de demain. Cette démarche rejoint alors les objectifs de neutralité climatique de la « Mission Cities 2030 », qui vise à accompagner les villes vers le zéro carbone et à accélérer le développement de solutions de mobilité urbaine durables.

Pour atteindre ces objectifs, REALLOCATE s'appuie sur un consortium européen composé de trente-sept organisations réparties dans douze pays. Les partenaires sont issus de milieux différents : collectivités locales, universités, consultants spécialisés dans la mobilité et l'urbanisme, ainsi qu'organisations à but non lucratif. Cette diversité permet de mettre en commun des compétences et des expériences complémentaires.

2.1.2. Organisation du consortium et stratégie de déploiement

La stratégie du projet repose sur un réseau de villes collaborant de manière structurée. Dans un premier temps, dix villes sont rep en cinq binômes. Chaque binôme est composé d'une ville pilote (Lead City) et d'une ville jumelle (Twin City). Les cinq binômes sont les suivants : Barcelone-Bologne, Budapest-Zagreb, Göteborg-Tampere, Heidelberg-Utrecht et Lyon-Varsovie.

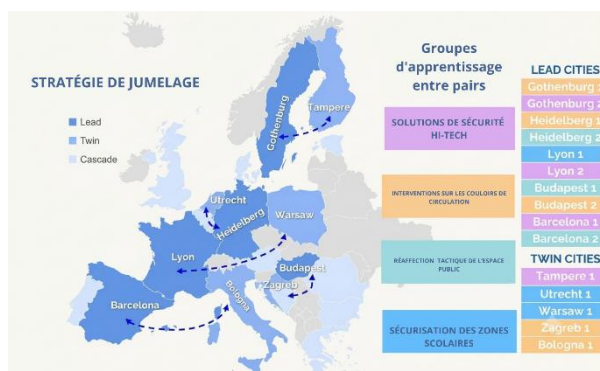


Figure 1 - Cartographie de la stratégie de jumelage et des groupes d'apprentissage entre villes européennes traduite en français Source : Reallocatemobility.eu

Ensemble, ils travaillent en partenariat avec des spécialistes (conseil, bureau d'études et organismes de recherche) pour développer et de mettre en place des solutions encourageant une mobilité plus active, inclusive et sans émissions. À terme, l'ensemble des résultats du projet sera rendu public, afin que d'autres villes puissent s'en inspirer. En effet, le dispositif a également été pensé pour que les résultats ne restent pas limités aux dix villes directement engagées dans les expérimentations. C'est pourquoi, le projet implique aussi des *Cascades Cities*. Ces villes participent à des activités de renforcement des compétences et reçoivent de

ressources et de recommandations pour pouvoir, à leur tour, mettre en place sur leur territoire au moins une des solutions testées par les villes pilotes. De plus, la liste des Cascade Cities a vocation à s'élargir tout au long du projet et compte à présent onze villes. Les solutions testées peuvent être regroupées autour de quatre grands domaines : le design urbain innovant, le nudging, qui consiste à encourager certains comportements sans imposer de contrainte, les technologies dites intelligentes et les méthodes fondées sur l'utilisation des données. Au-delà des solutions elles-mêmes, ces expérimentations permettent aux équipes municipales de comparer leurs approches, de partager leurs difficultés et de tirer des enseignements des expériences menées dans les autres villes. Pour plus d'informations sur les projets menés dans chaque ville, voir l'Annexe I.

2.1.3. Financement, cadre contractuel et rôle du coordinateur

Le financement de REALLOCATE provient du programme Horizon Europe, le principal programme de l'Union européenne consacré à la recherche et à l'innovation. C'est dans le cadre de la Mission Villes, intitulée « 100 Climate-Neutral and Smart Cities by 2030 ». Le cadre contractuel, défini dans le Grant Agreement, non seulement fixe le financement du projet, mais aussi les responsabilités de chaque acteur, le calendrier des livrables et les règles relatives à l'utilisation du budget. Le Spatial Dynamics Lab coordonne ce projet, dont le budget total s'élève à 12 millions d'euros. Un rôle est attribué surtout pour ses activités de recherche du laboratoire, notamment dans les domaines de l'aménagement urbain, des smart cities et de la croissance durable. La coordination d'un projet européen de cette envergure représente ainsi une occasion de mettre en pratique leurs méthodes développées et de faire le lien entre la recherche, les décideurs publics et les acteurs opérationnels. À travers REALLOCATE, ces différentes dimensions sont réunies autour d'un même objectif : repenser les espaces urbains pour les rendre plus durables.

2.2. Organisation et pilotage du projet

2.2.1. Architecture en Work Packages et rôle du WP3

Le projet REALLOCATE s'articule autour d'une organisation en *Work Packages* (WP), divisés en plusieurs tâches (Task), avec des livrables à produire à différentes étapes du projet. Le Spatial Dynamics Lab intervient dans l'ensemble de ces activités, mais il est directement responsable du WP1 (*Project coordination*) et du WP3 (*Urban design, behavioural, and technological solutions*). Les autres *Work Packages* ont des fonctions différentes. Le WP2 concerne les laboratoires vivants consacrés aux questions de mobilité. Le WP4 porte sur les échanges entre les villes et le renforcement des compétences. Le WP5 est consacré au suivi et à l'évaluation du projet, tandis que le WP6 regroupe les activités liées à la communication, à la diffusion et au partage des résultats.

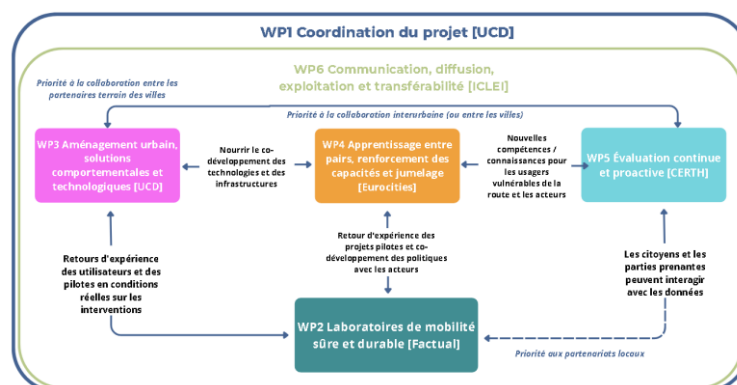


Figure 2 - Architecture globale des Work Packages traduite en français Source :
Reallocatemobility.eu

Le Work Package 3 (WP 3), au sein duquel j'ai apporté ma contribution, joue un rôle charnière dans l'organisation globale. Il apporte un appui scientifique et technique transversal à l'ensemble des projets d'aménagement urbain, désignés sous le terme de *Labs*, en référence à leur rôle de « laboratoires vivants ». Ces derniers expérimentent différentes solutions sur 15 sites pilotes situés dans 10 villes européennes. L'objectif est de trouver des solutions permettant de rendre les déplacements plus sûrs, plus accessibles et moins polluants. Plusieurs domaines sont étudiés dans ce WP. Certains travaux portent par exemple sur l'utilisation de l'intelligence artificielle pour la modélisation spatiale ou sur la conception de projets en 3D. D'autres concernent la planification urbaine en lien avec la neutralité climatique ou encore la Vision Zéro dans le domaine de la sécurité routière. Le WP3 s'intéresse également aux outils numériques, aux besoins des usagers et à de nouvelles formes de gouvernance et de financement dans le secteur des transports.

2.2.2. Gouvernance et coordination du consortium

La coordination du projet, qui réunit trente-sept organisations dans douze pays, implique de nombreux échanges afin d'offrir aux villes du projet REALLOCATE un accompagnement d'experts dans des domaines clés pour la conception de leur projet (voir Annexe II). Plusieurs types de réunions sont donc organisés régulièrement et prévues dès le début du projet et figurent dans le Grant Agreement. Six instances régulières structurent le calendrier du consortium.

Les « consortium meetings » ont lieu tous les six mois et réunissent physiquement les représentants des différents partenaires. Ils sont organisés à tour de rôle dans les villes participantes, ce qui permet à chaque réunion de se dérouler dans un contexte urbain différent. Ces rencontres sont également l'occasion de découvrir directement certains sites d'expérimentation et de mieux comprendre les projets développés localement. Chaque mois, un « general meeting » rassemble aussi l'ensemble du consortium mais virtuellement. Les villes peuvent y présenter l'avancement de leur projet et obtenir des retours des autres partenaires. À cette organisation s'ajoutent les réunions mensuelles propres pour chaque ville mais aussi pour chaque Work Package ainsi que les « management meetings », qui réunissent les responsables des différents WP autour du coordinateur de l'University College Dublin. Des points bimensuels sont également organisés avec chaque ville, en présence du coordinateur et des consultants, afin de suivre plus précisément l'avancement des projets pilotes. Depuis le lancement du projet, aucune des réunions prévues n'a été reportée. Cette régularité facilite le suivi des différents projets et permet également de réagir rapidement lorsqu'une difficulté apparaît.

Les Cascade Cities disposent quant à elles de réunions dédiées. Leur fréquence a progressivement augmenté avec l'arrivée de nouvelles villes souhaitant découvrir les méthodes développées dans REALLOCATE. Enfin, des mitigation meetings peuvent être organisées lorsqu'une ville rencontre une difficulté importante et demande l'aide du consortium. C'est notamment dans ce cadre que j'ai participé à la conception d'une enquête pour Heidelberg, liée à leur système de vélos en libre-service.

2.2.3. Validation des livrables, suivi budgétaire et indicateurs

La préparation d'un livrable suit plusieurs étapes avant sa transmission à la Commission européenne. Le document est premièrement vérifié par le coordinateur afin de s'assurer que les

objectifs définis et les dépenses associées aux projets restent conformes à ce qui était prévu. La modification des conditions entraîne un coût supplémentaire, par exemple en raison d'un aménagement supplémentaire imprévu de l'espace public, et dans ce cas l'accord du Project Officer est nécessaire pour réévaluer le financement.

Ce type d'évènement s'est présenté à moi dans le cadre d'un projet de nudging. Les règles d'aménagement urbain de cette ville imposent l'installation obligatoire d'une signalisation lumineuse pour un type de voirie construite. Cela n'avait pas été pris en compte dans le budget initial. Malheureusement, le coût supplémentaire n'ayant pas été accepté par le Project Officer, la municipalité a finalement dû le prendre en charge sur ses propres ressources. Cet exemple montre que les budgets des projets pilotes sont relativement contraints. Les montants définis dans le Grant Agreement ne peuvent pas être modifiés librement. Les villes disposent néanmoins d'une certaine marge de manœuvre dans la manière de mettre en œuvre leur projet, à condition de rester dans le cadre fixé par le consortium.

Une fois préparé, le livrable est relu par des examinateurs issus d'autres organisations du consortium et n'ayant pas participé à sa rédaction. Après son envoi, le Project Officer et des experts indépendants disposent de 60 jours pour l'évaluer. Le livrable peut alors être accepté, faire l'objet de demandes de modification ou nécessiter des informations complémentaires.

Le consortium doit également produire des rapports périodiques afin de rendre compte de l'avancement technique et financier du projet. Les partenaires y renseignent notamment les mois de travail réalisés, les frais de déplacement et les dépenses liées aux équipements. La validation de ces rapports conditionne le versement des différentes tranches de financement. Le projet repose aussi sur des Milestones, qui correspondent à des points de contrôles ou des étapes clés de son avancement, ainsi que sur des KPI (Key Performance Indicator) permettant de mesurer les résultats obtenus. Ces indicateurs peuvent par exemple porter sur la surface de voirie réaménagée, la réduction estimée des émissions de CO₂, la participation des citoyens ou encore le nombre de solutions reprises par les Twin Cities.

La charge de travail varie fortement d'un partenaire à l'autre. En tant que coordinateur, l'University College Dublin, à travers le Spatial Dynamics Lab, est logiquement l'un des partenaires qui mobilise le plus de person-months. Son rôle implique en effet un suivi continu du projet, depuis la préparation et la validation des livrables jusqu'aux échanges avec le Project Officer et à l'organisation des différentes réunions. À l'inverse, certains partenaires interviennent principalement sur des tâches ou des périodes précises. La répartition du travail dépend donc directement du rôle de chaque organisation au sein de REALLOCATE.

2.3. Mes missions au sein du WP3

J'ai rejoint le projet REALLOCATE au cours de la troisième année du projet, environ 30 mois après son lancement. Le consortium se trouvait alors dans une phase particulièrement opérationnelle : les bases méthodologiques avaient été définies, les outils d'IA et de modélisation étaient déjà déployés et les villes pilotes commençaient à mettre en œuvre leurs plans d'action.

Au sein du WP3, j'ai principalement travaillé sur la task 3.1, pilotée par l'UCD en collaboration avec DEKRA et Demos. Cette tâche vise à réfléchir à la manière de concevoir des espaces publics plus inclusifs, tout en limitant leur impact sur les ressources. Les travaux portent sur les Solutions Fondées sur la Nature, comme les noues paysagères, les murs végétalisés ou encore les systèmes de récupération d'eaux pluviales. La question des matériaux est également prise en compte à travers leur empreinte carbone, leur recyclabilité et leur potentiel de

démontage ou de réutilisation. Les analyses de cycle de vie permettent ainsi d'évaluer plus précisément leur impact.

L'enjeu est de concevoir des aménagements capables de mieux répondre aux effets du changement climatique tout en restant accessibles au plus grand nombre, notamment aux personnes en situation de handicap.

C'est dans ce contexte que j'ai été chargé de développer le StoryMaps sur ArcGIS StoryMaps. L'objectif était de centraliser et de présenter clairement aussi bien aux acteurs du projet qu'au grand public, les interventions prévues dans six villes pilotes : Barcelone, Bologne, Budapest, Heidelberg, Utrecht et Varsovie. Pour comprendre les spécificités de chaque territoire, j'ai notamment participé aux réunions de coordination consacrées à la finalisation des designs. Cela m'a permis de suivre plus directement l'évolution des projets, notamment autour des aménagements prévus à proximité des écoles, dont certains chantiers devaient débiter avant la fin de mon stage.

La StoryMaps a été conçu comme un outil permettant de rassembler plusieurs dimensions du projet. Il présente notamment les démarches de diagnostic et de conception, les possibilités de Green Mapping identifiées en amont ainsi que les résultats d'un concours de design organisé par le laboratoire auprès des étudiants en architecture et en génie civil de l'UCD. Ce concours portait notamment sur des concepts innovants autour de l'Educology Planter Box, avec la possibilité de les intégrer ensuite à REALLOCATE.

La plateforme comprend également un benchmark mondial de Solutions Fondées sur la Nature, indépendant du projet, destiné à fournir des exemples et des pistes d'inspiration. J'y ai également intégré les principales données issues des rapports d'expertise de DEKRA sur l'Analyse du Cycle de Vie et de Demos sur la vision stratégique à long terme. L'objectif était de proposer aux décideurs politiques, aux ingénieurs du consortium et aux Cascade Cities un support à la fois clair, visuel et suffisamment rigoureux pour faciliter la compréhension et la réplique des solutions étudiées.

En parallèle du développement de la StoryMaps, j'ai apporté un soutien régulier au consortium, notamment dans la préparation des réunions et des supports de présentation. J'ai également participé aux activités de la Task 3.5, en contribuant à la construction d'une enquête destinée à étudier les usages du système de vélos en libre-service VRNnextbike à Heidelberg.

Rétrospectivement, la StoryMaps que j'ai développé peut être vu à la fois comme un outil de coordination transversale et comme un support de valorisation. Il joue d'abord un rôle d'interface entre technique et politique : il transforme des données géospatiales complexes en un récit visuel accessible aux décideurs locaux et aux non-spécialistes. Il assure aussi une fonction de liaison entre les différents Work Packages, en faisant le lien entre les équipes qui collectent et traitent les données et celles qui gèrent la communication et la réplique vers les villes suiveuses. Enfin, il sert de support pour objectiver l'avancement du projet : en présentant de manière unifiée l'état des réaménagements dans chaque ville, il permet des comparaisons qui n'étaient pas possibles avant, et nourrit indirectement les échanges entre partenaires (cf. Annexe 3 pour les captures d'écran des StoryMaps présentant le projet de Varsovie).

2.4. Retour d'expérience : entre cadre européen et réalités du terrain

2.4.1 Constats dans le projet

Le premier constat dans ce stage c'est de voir à quel il est courant d'observer un écart entre les prévisions initiales et le déroulement effectif du projet. En effet, depuis le lancement de REALLOCATE, plusieurs projets pilotes ont été modifiés, parfois de manière importante, et

certains ont même été abandonnés. Par exemple, un aménagement prévu aux abords d'un établissement scolaire a finalement été supprimé après le refus de l'établissement concerné. Les projets pilotes ont ainsi fait l'objet de nombreuses révisions au cours du stage.

Cette évolution ne traduit pas nécessairement un dysfonctionnement. Elle montre plutôt la difficulté de faire correspondre les temporalités d'un projet européen avec celles de projets d'aménagement urbain, qui dépendent de contraintes locales et de décisions prises au fil de leur réalisation. Néanmoins cela a eu une conséquence directe sur mon travail, car la StoryMaps devait rendre compte de l'état d'avancement des différents pilotes, il fallait régulièrement modifier les contenus cartographiques et éditoriaux pour tenir compte des dernières évolutions.

C'est pourquoi, la grande difficulté rencontrée dans la rédaction de la StoryMaps a finalement été l'échange des données par les municipalités. Au fil du projet, j'ai pu constater que ce problème était dû à des contraintes qui se cumulaient et pas simplement simplement d'un manque de réactivité.

La première était d'ordre administratif. Les municipalités doivent plusieurs validations avant de pouvoir transmettre certaines données, notamment lorsqu'elles concernent des informations soumises au RGPD. Les services techniques disposent par ailleurs de ressources limitées et n'ont pas nécessairement une personne dédiée au projet. Leur rythme de traitement des demandes se diffère totalement de celui d'un cabinet de conseil ou d'un laboratoire.

La méfiance entre les différents acteurs est une autre difficulté que j'ai découvert. Certaines villes se montraient réticentes à échanger leurs données, même lorsque leur partage avait été prévu dès le début du projet. Parfois, les informations demandées n'étaient pas détenues directement par la municipalité, mais par des partenaires locaux ou des opérateurs régionaux extérieurs au consortium. Il fallait alors mettre en place des conventions spécifiques pour pouvoir y accéder.

La qualité des données constituait une troisième difficulté. Certaines données collectées ne pouvaient finalement pas être utilisées, comme une enquête réalisée auprès d'élèves qui ne répondait pas aux exigences méthodologiques nécessaires. À cela s'ajoutent une forte hétérogénéité des formats utilisés par les différentes villes, ce qui nécessitait un travail de nettoyage et d'harmonisation avant de pouvoir exploiter les informations.

Enfin, j'ai également rencontré une difficulté plus technique liée aux outils de partage utilisés par le consortium. Les documents sont principalement stockés sur Google Drive, notamment parce que l'University College Dublin utilise des comptes universitaires Google. Or certains partenaires, notamment les collectivités françaises comme la Métropole de Lyon, ne disposent pas de comptes Google et ne peuvent donc pas utiliser l'outil de la même manière. Un choix technique qui semblait relativement neutre au départ pouvait ainsi devenir une contrainte supplémentaire pour certains partenaires.

Le consortium dispose pourtant de templates et d'un plan de gestion des données définis dans le Grant Agreement. Leur existence ne suffit donc pas toujours à garantir une circulation fluide des données. Lorsqu'un blocage apparaît, les partenaires organisent généralement une réunion dédiée afin d'en comprendre l'origine et de chercher une solution commune. Dans la pratique, le dialogue et la recherche de compromis constituent donc les principaux moyens utilisés pour résoudre ces difficultés.

2.4.2 Cultures professionnelles, jeux d'acteurs et évolution vers le volet recherche

Au-delà des difficultés liées aux données, mon stage m'a également permis d'observer les différences de fonctionnement entre les acteurs du consortium. La question des temporalités est probablement celle qui m'a le plus marqué. Chaque type d'organisation possède ses propres contraintes et ses propres délais de validation, et ces rythmes ne sont pas toujours compatibles avec le calendrier fixé à l'échelle européenne.

La langue de travail est un autre élément qui peut parfois compliquer les échanges car tout le monde ne le maîtrise pas de la même manière. Les chercheurs et les consultants, habitués aux projets internationaux sont généralement à l'aise, tandis que certains agents territoriaux peuvent avoir plus de difficultés à s'exprimer. Cela se ressent notamment dans les échanges informels, qui sont parfois moins spontanés alors qu'ils contribuent pourtant beaucoup à créer une véritable dynamique commune au sein du projet.

J'ai aussi pu observer des différences dans les façons de travailler selon les cultures professionnelles. Certains partenaires accordent davantage d'importance à la recherche de consensus, tandis que d'autres fonctionnent de manière plus hiérarchique. Les différences de périodes de congés entre les pays peuvent également avoir un impact sur les calendriers. Ces aspects peuvent paraître assez secondaires au premier abord, mais ils ont finalement des conséquences concrètes sur l'organisation du projet et sur le respect des différentes échéances.

Le fonctionnement entre les Lead City et les Twin City met aussi en évidence des différences assez importantes d'une ville à l'autre. Les Lead Cities disposent généralement d'infrastructures de données plus développées, d'une certaine expérience des living labs et d'un soutien politique qui leur permet de mettre plus rapidement en place les solutions étudiées. Les Twin Cities peuvent, de leur côté, être confrontées à davantage de difficultés, qu'elles soient techniques, financières ou méthodologiques. Leur rôle consiste alors davantage à s'approprier les solutions développées et à favoriser leur transfert et leur réplication dans leur propre contexte.

L'un des enjeux du projet est donc de maintenir leur implication malgré les contraintes de calendrier imposées par les villes pilotes. Les attentes ne sont pas non plus les mêmes selon les acteurs. Les partenaires académiques accordent une grande importance à la rigueur scientifique, à la qualité des données et à la valorisation des résultats par la publication. Contrairement aux collectivités qui attendent davantage des résultats concrets et relativement rapides, qu'elles puissent présenter à leurs administrés. Quant aux consultants, le plus important est de respecter les objectifs qui leur sont confiés tout en maîtrisant le budget et le nombre de person-months prévus. Les villes doivent enfin composer avec des contraintes politiques et réglementaires qui peuvent modifier le calendrier à tout moment.

Pour les collectivités, la commande publique et la consultation des habitants jouent notamment un rôle important. Cette dernière peut conduire à modifier fortement un projet, voire à l'abandonner. L'installation de bacs végétalisés dans une ville a par exemple nécessité plusieurs mois d'échanges avec les habitants avant de déterminer leur emplacement définitif. Ce type de processus est difficile à accélérer et ne correspond pas toujours au calendrier relativement rigide d'un projet européen.

J'en retiens que la coordination d'un projet comme REALLOCATE consiste en grande partie à trouver un équilibre entre ce qui serait idéal d'un point de vue scientifique et ce qui est réellement réalisable sur le terrain. La réussite du projet dépend donc autant de la qualité des méthodes employées que de la capacité des différents acteurs à comprendre leurs contraintes respectives et à construire des compromis.

Cette immersion m'a apporté une vision beaucoup plus concrète de la gestion d'un projet européen, aussi bien sur le plan méthodologique que sur le plan humain. Le travail autour du

StoryMaps restait toutefois principalement descriptif et me donnait encore peu d'occasions d'intervenir directement sur la conception technique des projets. De plus, l'avancement du projet dépendait aussi beaucoup du rythme des collectivités et de leur capacité à transmettre les données nécessaires, qu'il s'agisse de données d'impact, de plans ou encore d'informations sur les choix de végétaux.

J'ai donc échangé avec ma responsable à ce sujet afin de faire évoluer mes missions d'assistance à la coordination du WP3. Nous avons finalement décidé d'ajouter une dimension davantage orientée vers la recherche à mon plan de charge, en participant à la rédaction d'articles scientifiques.

3. Phase exploratoire et enseignements méthodologiques : le cas d'étude de Setia Alam (Malaisie)

Avant d'aborder le terrain parisien, une première phase exploratoire a été menée à Setia Alam, un quartier résidentiel situé dans l'État de Selangor, en Malaisie. Cette étape avait le but de se familiariser avec les outils et la méthodologie dans un contexte peu complexe et surtout de produire des données complémentaires à une thèse en cours consacrée aux stratégies de végétalisation visant à réduire les îlots de chaleur urbains en Malaisie.

Deux approches ont alors été testées. La première consistait à recueillir des habitants face à la chaleur, aux inondations et, plus généralement, à leur environnement climatique quotidien grâce aux réseaux sociaux, ici Twitter/X. La seconde reposait sur des captures d'écran provenant de Google Street View grâce à un modèle d'intelligence artificielle et de les analyser. Elle permet d'observer à grande échelle certaines caractéristiques de l'espace public des lieux. L'objectif était ensuite de comparer ces données avec des mesures plus physiques du territoire, notamment celles développées dans la thèse. Il s'agissait de voir si les résultats obtenus à partir des réseaux sociaux et des images correspondaient réellement à ce que les habitants vivaient et ressentaient au quotidien. Cette phase exploratoire permet de présenter le protocole mis en place et les résultats obtenus, mais surtout de revenir sur les limites rencontrées et sur les raisons qui ont conduit à faire évoluer la méthode pour la suite du stage.

3.1. Analyse des sentiments sur Twitter/X : écouter la voix des habitants

3.1.1. Pourquoi Twitter/X ?

Adapter une ville à la chaleur ne se résume pas à planter des arbres ou à modifier les matériaux : encore faut-il que les habitants acceptent ces changements, les entretiennent et se les approprient (Lin et al., 2019 ; Kabisch et al., 2017). Autrement dit, la performance technique d'un aménagement ne suffit pas ; il faut aussi tenir compte du ressenti de celles et ceux qui vivent au quotidien dans ces espaces (Adger et al., 2009 ; Wamsler et al., 2020).

Pour capter ce ressenti, Twitter (aujourd'hui X) a été choisi comme source de données. C'est une plateforme très utilisée en Malaisie pour discuter des questions environnementales en temps réel (Tan et al., 2021), déjà éprouvée dans la recherche sur le climat (Cody et al., 2015 ; Kirilenko & Stepchenkova, 2014). Son caractère public, bilingue (anglais et malais) et horodaté permet de saisir les réactions spontanées de la population, ce que des réseaux plus fermés (Facebook, WhatsApp) ne permettent pas.

3.1.2. Protocole en trois étapes

La démarche s'organise en trois phases successives :

Phase 1 - Collecte et prétraitement. Les tweets géoréférencés de l'État de Selangor ont été collectés par mots-clés thématiques (chaleur, inondation, brume sèche, pluies et orages, changement climatique) sur la période de janvier 2021 à mai 2026. Chaque tweet a été ensuite classé dans l'une des cinq catégories thématiques retenues : chaleur, inondation, brume sèche, pluies et orages, changement climatique. Pour cela, un dictionnaire bilingue anglais-malais a été spécialement créé pour l'étude. Il permettait de relier les mots-clés utilisés pour la collecte des tweets à leurs équivalents dans les deux langues. La période étudiée couvre notamment les années qui ont suivi la pandémie, une période durant laquelle les discussions liées aux questions environnementales ont pris une place importante sur les réseaux sociaux. Elle comprend également les importantes inondations survenues en décembre 2021, qui constituent un

événement particulièrement intéressant pour analyser les réactions des habitants face aux risques environnementaux. La collecte des données a été réalisée à l'aide du service tiers twitterapi.io, l'accès à l'API officielle de X étant devenu particulièrement coûteux depuis 2023. Afin de conserver un échantillonnage équilibré, le nombre de tweets a été limité à 1 000 pour chaque combinaison entre un mot-clé, un mois et une zone. Seuls les messages rendus publics ont été collectés et aucune information personnelle n'a été conservée, conformément aux principes éthiques définis dans la littérature (Williams, Burnap & Sloan, 2017 ; Townsend & Wallace, 2016).

Phase 2 - Classification automatique. Deux modèles informatiques ont ensuite été utilisés pour analyser les tweets. Le premier, twitter-xlm-roberta-base-sentiment (Barbieri et al., 2022), est un modèle multilingue qui permet d'attribuer à chaque message une polarité, positive, neutre ou négative. Le second, emotion-english-distilroberta-base (Hartmann, 2022), permet d'identifier six émotions de base : la colère, le dégoût, la peur, la joie, la tristesse et la surprise. Les résultats ont également été comparés à ceux obtenus avec le lexique VADER afin de disposer d'une vérification supplémentaire. Un seuil de confiance de 0,5 a été retenu. Lorsque le modèle obtenait un score inférieur à ce seuil, la classification était considérée comme trop incertaine et le tweet n'était pas pris en compte dans les analyses émotionnelles.

Phase 3 - Analyse spatiale. Les tweets ont enfin été répartis selon trois cercles concentriques autour de Setia Alam :

C1 - Setia Alam même (échelle du quartier) ;

C2 - Shah Alam et communes adjacentes (échelle urbaine) ;

C3 - État de Selangor dans son ensemble (échelle régionale).

Chaque tweet est attribué au cercle le plus restreint auquel il appartient, ce qui évite les doublons entre les trois niveaux.

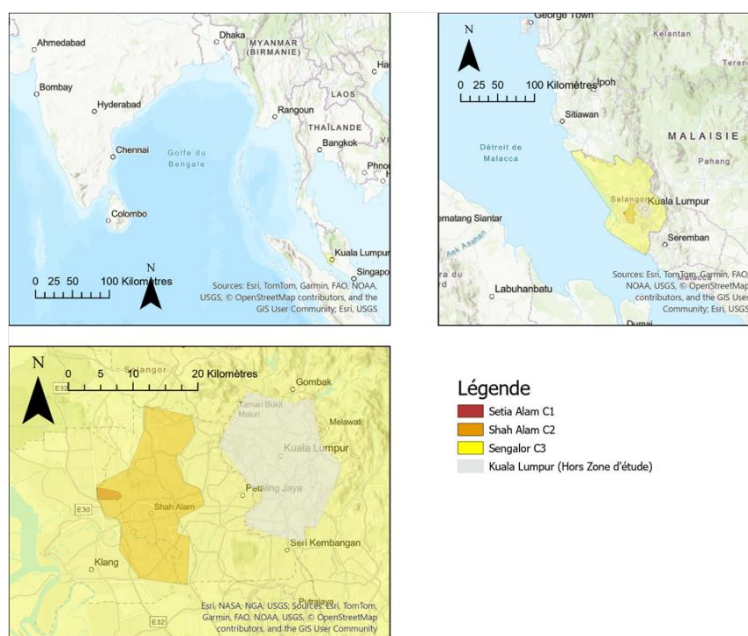


Figure 3 - Représentation cartographique des trois échelles d'analyse spatiale des tweets (C1, C2, C3) Source : GOPALAPILLAI Gokul

3.1.3. Résultats

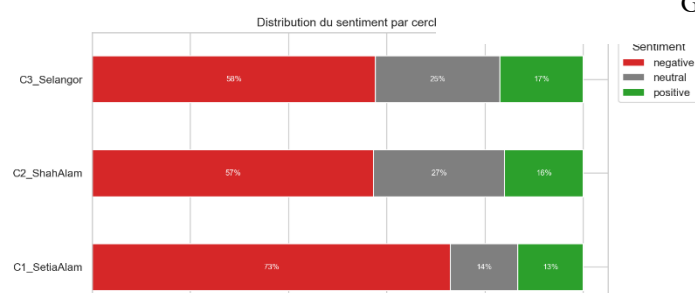


Figure 4 - Distribution du sentiment par cercle Source : GOPALAPILLAI Gokul

Au total, 11 193 tweets uniques ont été obtenus après déduplication. Leur répartition entre les trois cercles étudiés est toutefois très inégale : 385 tweets pour Setia Alam (C1), soit environ 6 tweets par mois, 3 956 pour Shah Alam et les communes adjacentes (C2), soit environ 61 tweets par mois, et 6 852 à l'échelle de l'État de Selangor (C3), soit

environ 105 tweets par mois sur les 65 mois étudiés. D'après la figure 4, les trois cercles montrent une prédominance générale des messages négatifs, bien que les différences observées entre les échelles géographiques doivent être interprétées avec prudence. Ces résultats doivent être interprétés par prudence, car le corpus est très déséquilibré. Car seuls 385 tweets concernent directement Setia Alam (C1), contre 3 956 pour Shah Alam et ses environs (C2) et 6 852 pour l'ensemble du Selangor (C3). Les analyses complémentaires présentées en Annexe 4 vont dans le même sens. La classification des émotions produit principalement des résultats neutres ou trop incertains pour pouvoir être réellement interprétés, tandis que l'analyse de leur évolution dans le temps ne fait apparaître aucune tendance claire. Le sentiment varie toutefois selon les thèmes abordés : les tweets liés aux inondations, à la pluie et aux orages présentent par exemple une tonalité plus négative. Ces résultats renseignent néanmoins davantage sur la façon dont ces événements sont évoqués sur Twitter que sur l'inconfort thermique réellement ressenti par les habitants. Enfin, XLM-R et VADER attribuent la même polarité à seulement 33,3 % des tweets, ce qui montre que les résultats dépendent beaucoup de l'outil utilisé, notamment parce que ces modèles sont encore mal adaptés aux publications en malais. Le faible nombre de tweets à l'échelle du quartier, les incertitudes de classification et l'absence de tendances temporelles claires ont conduit à réorienter l'étude vers d'autres sources de données.

3.1.4. Limites et enseignements de l'approche

L'analyse a surtout permis de mettre en évidence les difficultés liées à l'utilisation des réseaux sociaux pour étudier l'inconfort thermique à l'échelle de Setia Alam. La première difficulté concerne les outils utilisés pour analyser les tweets. Le modèle de classification des émotions ayant été entraîné à partir de données en anglais, il était peu adapté aux tweets en malais. Une grande partie des résultats était classée comme neutre ou obtenait des scores trop faibles pour permettre une interprétation fiable. Le lexique VADER, aussi conçu pour l'anglais, rencontrait des difficultés similaires. Certains mots malais étaient par exemple considérés comme neutres, alors que le modèle multilingue leur attribuait une polarité négative. Cette différence permet notamment d'expliquer la faible concordance observée entre les deux méthodes.

Le volume de données constitue une autre limite importante. Le plafond de 1 000 tweets fixé pour chaque combinaison n'a en effet jamais été atteint, confirmant que la rareté des données est une caractéristique propre au terrain. Avec moins de six tweets par mois en moyenne, le corpus est limité pour une analyse temporelle fiable. L'exemple des inondations de décembre 2021 illustre cette limite. Cet événement a pourtant fortement touché le Selangor et suscité de nombreuses réactions en ligne. Malgré cela, aucune augmentation clairement identifiable du nombre de tweets, ni évolution nette de leur polarité, n'apparaît dans le corpus analysé. Si un événement aussi important n'est pas suffisamment visible dans les données, il devient difficile d'utiliser cette méthode pour repérer des variations plus quotidiennes de l'inconfort thermique.

À cela s'ajoutent deux biais qui restent difficiles à éviter. Il est assez aisé de peut produire des faux positifs, car certains termes liés à la chaleur peuvent être utilisés en malais dans un sens figuré pour exprimer de l'irritation. De plus, le fait qu'un tweet mentionne Setia Alam ou un autre lieu ne signifie pas nécessairement que son auteur s'y trouvait au moment de la publication.

Ces constats montrent finalement que les difficultés importantes ne sont pas seulement liées aux réglages des outils, mais surtout à l'adéquation entre la méthode et les données disponibles. Pour que cette approche soit réellement exploitable, il faudrait disposer d'un modèle mieux adapté à la langue utilisée sur le terrain, d'un volume de données plus important à l'échelle du quartier et d'une vérification manuelle d'une partie du corpus par une personne maîtrisant le malais. Malgré ses limites, cette première expérience a donc permis de mieux cerner les

conditions nécessaires à ce type d'analyse et a servi de base à la réorientation méthodologique du stage.

3.2. Évaluation perceptive du cadre bâti par vision par ordinateur (Google Street View)

3.2.1. Principe et justification de la vision par ordinateur

Depuis une dizaine d'années, l'imagerie de rue (Google Street View, Mapillary) permet d'évaluer automatiquement la qualité des espaces urbains. Cela permet de créer qui sont entraînés sur le jeu de données MIT Place Pulse 2.0, qui regroupe 1,1 million de comparaisons d'images provenant de 56 villes et réalisées par 81 000 participants. Ces outils permettent d'estimer ce que ressentent les personnes face à un paysage les six critères : *Beautiful, Boring, Depressing, Lively, Safety et Wealthy*. Dans leur étude de référence, Dubey et al. (2016) (*Deep Learning the City: Quantifying Urban Perception at a Global Scale*) montrent qu'une intelligence artificielle (CNN) peut reproduire ce ressenti à grande échelle. Les notes obtenus ne mesurent donc pas objectivement la qualité urbaine, mais plutôt une perception humaine moyenne. Le modèle utilisé ici présente aussi l'avantage de s'appuyer sur un jeu de données bien documenté, contrairement à certains modèles de langage-vision (VLM) plus récents, dont les données d'entraînement sont souvent moins transparentes.

L'évaluation du cadre bâti par vision par ordinateur vient compléter l'analyse des tweets, car les deux approches apportent des informations différentes. Les tweets permettent de voir comment les usagers parlent de leur environnement et ce qu'ils en ressentent. Cependant, à l'échelle de Setia Alam, le nombre de tweets disponibles reste limité et leur analyse est compliquée par les différences linguistiques du corpus.

Les images de Google Street View apportent un autre regard, en observant directement les caractéristiques visuelles du cadre bâti et en évaluant des éléments comme sa beauté ou sa vitalité perçue. L'idée est alors de croiser ce que les usagers pensent sur leur environnement avec ce que l'on peut observer directement de ce même environnement. Cela permet d'avoir une lecture plus complète de la perception du cadre urbain, sans dépendre uniquement des données issues de Twitter.

3.2.2. Protocole de collecte

Une campagne de capture manuelle a été réalisée sur Google Street View dans la zone de Setia Alam. Les points de prise de vue ont été sélectionnés aux endroits où le paysage urbain change nettement le long des itinéraires disponibles, afin d'obtenir une lecture fine du cadre bâti. Au total, 1 489 images ont été capturées. Pour chaque image, les coordonnées géographiques (latitude, longitude) et l'orientation de la caméra ont été extraites de l'URL Street View, structurées dans un fichier CSV, puis importées dans ArcGIS Pro pour produire une carte géoréférencée des points d'observation. Chaque image a ensuite été soumise au modèle pré-entraîné, qui lui a attribué une note sur chacun des six critères.



Figure 5 - Exemples représentatifs des photographies capturées illustrant le cadre bâti de Setia Alam. Source : Google Street View

3.2.3. Résultats : tendances et limites



Figure 6 - Carte de localisation des points de capture Google Street View à Setia Alam. Source : GOPALAPILLAI Gokul

La répartition des 1 489 points de capture met en évidence un biais spatial important. Les images disponibles se concentrent surtout le long des grands axes routiers, comme Persiaran Setia Prima, Persiaran Setia Indah et Jalan Meru Link, ainsi que dans quelques rues commerçantes. À l'inverse, une grande partie des quartiers résidentiels reste absente, alors qu'elle occupe une place importante dans le tissu urbain de Setia Alam. Cela est dû aux limites de Google Street View, car beaucoup de voies sont privées ou inaccessibles aux véhicules chargés de prendre les

images. Par ailleurs, de nombreux quartiers résidentiels sont également fermés et leur accès est contrôlé par des barrières ou des postes de garde. À cela s'ajoute le fait que certaines images sont anciennes, incomplètes ou simplement indisponibles.

Ces limites expliquent alors pourquoi les scores obtenus ne donnent pas une vision complète de Setia Alam, mais reflètent surtout la situation des axes de circulation couverts par Street View. Ils ne peuvent par conséquent pas être généralisés à l'ensemble du quartier résidentiel.

Les résultats montrent une tendance globalement positive pour les critères Beautiful et Lively. Les scores élevés sont majoritaires, avec notamment 384 images classées entre 8 et 10 pour Beautiful et près de 470 dans cette même classe pour Lively. Les cartes détaillées de la répartition spatiale de ces deux critères sont présentées en Annexe V.

À l'inverse, les scores de Boring sont principalement faibles à intermédiaires, tandis que Depressing et Wealthy se concentrent davantage autour des valeurs intermédiaires. La distribution des scores pour les six critères est présentée sous forme d'histogrammes en Annexe VI. Dans l'ensemble, les espaces observés sont donc plutôt perçus comme esthétiques et vivants, sans être systématiquement associés à des niveaux élevés d'ennui ou de caractère déprimant.

La répartition des scores Beautiful et Lively varie selon les axes étudiés, avec des valeurs particulièrement élevées dans certains secteurs, notamment autour de Setia Prima, Setia Avenue et de plusieurs axes principaux. Ces résultats montrent que la ressenti face à l'environnement change d'un lieu à l'autre, une analyse qui reste générale.

3.3. Bilan et réorientation méthodologique

Les deux approches testées à Setia Alam n'ont finalement pas permis d'obtenir des résultats suffisamment robustes pour répondre aux objectifs initiaux. Cette phase exploratoire a néanmoins permis d'identifier plusieurs difficultés qui ont influencé la suite du stage. Elle a notamment montré l'importance d'adapter les outils au contexte local : le modèle multilingue était encore peu performant en malais, et son amélioration dépassait le cadre du stage. Des données insuffisantes ou mal réparties fragilisent également les résultats. Ce qui oblige ne vérification humaine reste surtout pour les contenus plurilingues ou ambigus.

Ces constats ont conduit à repenser la méthodologie de l'analyse des réseaux sociaux de Twitter/X et à revoir le rôle des images Street View dans la méthodologie. La suite du stage s'est ainsi recentrée sur la ZTL de Paris, en adaptant les méthodes de travail en fonction des difficultés rencontrées durant cette première phase.

4. Élaboration du protocole méthodologique final : le cas de la ZTL de Paris Centre

Les enseignements de la phase exploratoire sont ici repris sur un nouveau terrain : la Zone à Trafic Limité (ZTL) de Paris Centre. Sauf qu'ici les trois contraintes rencontrées en Malaisie peuvent être contournées : le français est bien pris en charge par les modèles, les débats autour de la ZTL génèrent suffisamment de données en ligne et le périmètre des quatre premiers arrondissements reste facilement accessible à pied pour la collecte photographique.

Ce volet ayant commencé seulement durant les deux derniers mois du stage, il n'est pas encore totalement finalisé. Réalisé en autonomie, il continue à être révisé par ma tutrice en vue d'une possible valorisation sous forme d'article scientifique. Des ajustements pourront donc encore être apportés à la méthodologie ou à l'interprétation des résultats d'ici la soutenance.

Cette partie présente alors successivement le cas d'étude, l'adaptation du volet de perception visuelle, la collecte et le nettoyage des données textuelles, l'analyse par traitement automatique du langage, la visualisation des résultats, puis la confrontation entre les perceptions exprimées en ligne et les mesures physiques disponibles. Elle se termine par une discussion des limites et de la reproductibilité de la démarche.

4.1. Présentation du cas d'étude et justification du choix de Twitter/X en France

4.1.1. La ZTL de Paris Centre : périmètre, objectifs et calendrier

La Zone à Trafic Limité de Paris Centre constitue un cas d'étude intéressant pour analyser les politiques publiques visant à réduire le trafic et à apaiser les espaces urbains. Mise en place dans les 1er, 2e, 3e et 4e arrondissements, à l'exception des îles de la Cité et Saint-Louis, elle vise notamment à réduire la pollution et les nuisances sonores en limitant le trafic de transit, c'est-à-dire les véhicules qui traversent le secteur sans s'y arrêter. L'accès est réservé aux personnes qui ont besoin de se déplacer dans le quartier comme les résidents, les professionnels, les livreurs et les services de secours. Le périmètre s'étend sur plusieurs secteurs centraux de Paris (les Halles, Châtelet, le Marais, autour du Louvre et du Centre Pompidou), et comprend des axes importants comme la rue de Rivoli et le boulevard de Sébastopol.

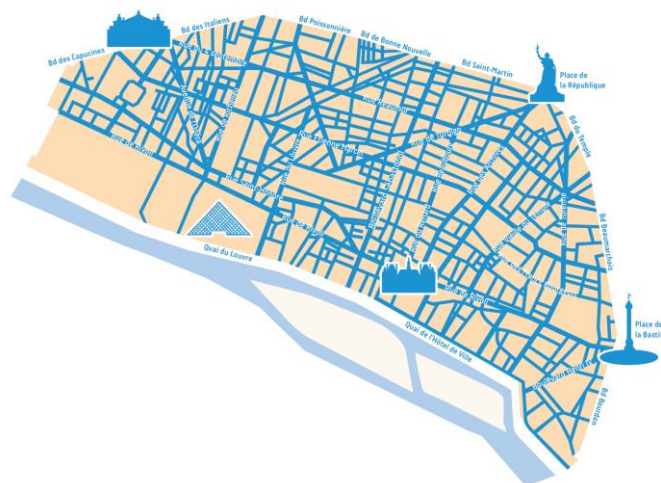


Figure 7 – Carte du périmètre de la ZTL Paris Centre Source : mairiepariscentre.paris.fr

Précédé par l'enquête publique du printemps 2024, l'ouverture des JOP (26 juillet) et l'arrêté conjoint maire-préfet (31 octobre), le dispositif entre en vigueur le 4 novembre 2024, avant une phase de pédagogie qui doit s'achever avec le début de la verbalisation, en janvier 2026. Autant de repères pour l'analyse longitudinale. Ce calendrier fournit autant de points de repère pour l'analyse longitudinale des perceptions.

4.1.2. Justification du terrain au regard des leçons de Setia Alam

Le choix de la ZTL de Paris Centre comme terrain d'application s'imposait naturellement au regard des enseignements de la phase exploratoire malaisienne. En plus de ma bonne connaissance de la langue, de la ville et des usages du réseau social étudié, la zone étudiée présente plusieurs avantages :

- Des modèles adaptés et entraînés sur de grands corpus francophones
- Un débat public très actif : depuis 2023, la ZTL fait régulièrement réagir beaucoup de monde que ce soit les élus, les riverains, les commerçants ou les usagers, ce qui nous fournit un volume non négligeable de données sur la durée.
- Un périmètre clairement défini : ses limites facilitent à la fois la collecte des données en ligne mais aussi la couverture photographique du terrain.
- Des données environnementales disponibles : les comptages de trafic, les mesures d'Airparif et les données de Bruitparif permettent de comparer les perceptions des usagers avec des indicateurs plus objectifs.

4.1.3. Justification du choix de Twitter/X comme source de données

Tout d'abord, la plateforme reste particulièrement adaptée à une collecte programmatique de discours publics. Contrairement à Facebook, dont l'accès aux données est devenu beaucoup plus limité, ou à Instagram et TikTok, dont les données sont plus difficiles à exploiter dans un cadre de recherche, Twitter/X permet encore de travailler à partir de contenus publics accessibles par API (González-Bailón et al., 2014 ; Jungherr, 2015). Cette accessibilité constitue un avantage important pour l'étude des débats urbains contemporains, même si la plateforme présente des biais de représentation qu'il faut prendre en compte (Boyd & Crawford, 2012).

Twitter/X présente également l'intérêt de donner accès à des réactions formulées directement par les utilisateurs. Contrairement aux sondages, qui reposent sur des questions préparées par le chercheur, ou aux médias traditionnels, qui sélectionnent et mettent en forme l'information, les publications permettent d'observer des réactions plus spontanées. Cette caractéristique est particulièrement intéressante pour une politique urbaine comme la ZTL, qui suscite des réactions immédiates de la part des habitants, commerçants et usagers.

La plateforme permet par ailleurs d'accéder à un vocabulaire plus proche des usages quotidiens : expressions familières, argot, ironie ou métaphores. Des termes comme « rideau baissé » ou « velotaf » peuvent ainsi apparaître dans les publications, alors qu'ils sont absents du vocabulaire institutionnel utilisé pour présenter la ZTL (« zone apaisée », « mobilité douce »). Ce langage vernaculaire constitue donc une source d'information intéressante pour comprendre la manière dont la zone est réellement décrite dans le débat public (Evans & Aceves, 2016).

Un autre avantage important concerne la dimension temporelle. Les données peuvent être observées de manière continue sur 42 mois, avec une granularité mensuelle. Il devient ainsi possible de mettre en relation l'évolution du discours avec différents événements, comme les Jeux olympiques de 2024, les épisodes de canicule ou les annonces politiques (Mellon, 2013).

Ces avantages ne signifient toutefois pas que Twitter/X soit représentatif de l'ensemble de la population. La plateforme rassemble notamment davantage de jeunes urbains, de catégories favorisées et de personnes déjà très impliquées dans le débat public. Les résultats doivent donc être lus comme le reflet des échanges en ligne autour de la ZTL, et non comme une mesure de l'opinion de tous les habitants. Les données présentent aussi un intérêt en termes de traçabilité et de reproductibilité. Chaque publication possède un identifiant, une date et un auteur, tandis que le pipeline peut être réexécuté. Cette démarche respecte les principes de la science ouverte et des données FAIR (Wilkinson et al., 2016).

4.2. Correction du volet perception visuelle : passage au relevé photographique in situ

4.2.1. Correction du mode d'acquisition

Le modèle de vision et les six critères issus de MIT Place Pulse 2.0 (*Beautiful, Boring, Depressing, Lively, Safety, Wealthy*) sont reconduits à l'identique, afin de conserver la comparabilité de l'instrument entre les deux terrains. La méthode de notation reste inchangée : chaque photo est soumise au même modèle pré-entraîné, qui attribue une note continue sur chacun des critères.

La principale modification concerne comment ces images ont été récupérées. Contrairement à Setia Alam où les images avaient été obtenues à distance à partir de captures de Google Street View qui biaisait l'attitude car une grande partie des espaces résidentiels et des cheminements piétons restait peu documentée. Ici pour l'étude parisienne, les photographies ont donc été prises directement sur le terrain, à l'aide d'un smartphone, lors de parcours à pied réalisés dans le périmètre de la ZTL en juillet et août 2026. Cette méthode permet de mieux couvrir les espaces directement concernés par la limitation du trafic, notamment les rues étroites, les places, les passages et les autres espaces piétons. Elle permet également de mieux contrôler les conditions d'acquisition. Les photographies ont été prises sur une période relativement courte et dans des conditions de lumière et de saison comparables. Les coordonnées géographiques ont par ailleurs été récupérées à partir des métadonnées EXIF des photographies, notamment la latitude, la longitude et l'orientation. Environ 800 photographies ont ainsi été prises au cours

des campagnes de terrain et choisies en fonction des changements observés dans le paysage urbain le long des itinéraires parcourus, afin de rendre compte des variations du cadre bâti.

Trois critères ont été retenus pour l'analyse cartographique : *Lively*, *Safety* et *Beautiful*. Ils correspondent directement aux principaux enjeux de la ZTL. *Lively* correspond à l'animation urbaine, ce qu'un apaisement du trafic cherche justement à favoriser. *Safety* correspond au sentiment de sécurité dans l'espace public, tandis que *Beautiful* permet d'évaluer la perception des aménagements et du cadre urbain. Dans les données de MIT Place Pulse 2.0, les différents critères sont en effet fortement liés : *boring* et *depressing* sont notamment négativement corrélés à *lively* et *beautiful*, tandis que *wealthy* est associé à *beautiful*. Cartographier l'ensemble des critères produit des résultats largement redondants. Les cartes des critères secondaires, ainsi que les histogrammes présentant les six scores, sont néanmoins disponibles en annexe afin de vérifier la robustesse de cette sélection.



Figure 8 - Carte des points de prise de vue à Paris et itinéraires parcourus. Source : GOPALAPILLAI Gokul

4.2.2 Première lecture des scores perçus

La lecture des distributions fait apparaître des résultats assez contrastés (voir l'Annexe VII et VIII pour les histogrammes et les cartes mentionnées par la suite). *Lively* est très largement orienté vers les classes élevées, avec plus de 80 % des images situées entre 8 et 10. *Wealthy* suit la même tendance, avec environ 80 % des images entre 6 et 10, tandis que *Beautiful* est plus nuancé, avec près de 60 % des images dans ces mêmes classes. À l'inverse, *Boring* et *Depressing* reste dans des pourcentages bas. Pour *Boring*, près de trois quarts des images sont en dessous de 4, ce qui confirme que c'est un centre animé et bien entretenu. *Safety* se distingue des autres critères, puisqu'environ 70 % des images se situent entre 2 et 6 et plus de 14 % sous 2. C'est donc le critère le moins favorable du corpus. La cartographie confirme globalement ces tendances, tout en montrant quelques différences selon les secteurs. Le score de *Lively* reste élevé partout dans la zone étudiée, sans véritable zone de désertion urbaine, y compris dans les espaces piétonnisés. *Beautiful* présente une répartition plus mitigée. En effet, les secteurs patrimoniaux, notamment le Marais, les quais de Seine et la partie ouest du 1er arrondissement, obtiennent les meilleurs scores, à l'inverse des zones autour des espaces très minéralisés des Halles et de quelques secteurs en chantier. La carte de *Safety* est la plus contrastée. Les points faibles sont dispersés avec une présence peu marquée dans les secteurs centraux et denses, notamment autour des Halles, de Châtelet et dans le 2e arrondissement, sans pour autant faire apparaître un véritable gradient spatial.

En résumé, le centre historique est considéré comme un espace très vivant et plutôt valorisé sur le plan esthétique, mais avec un sentiment de sécurité qui semble moyen. Cela peut s'expliquer par les caractéristiques propres aux centres historiques denses, c'est-à-dire une forte activité et

une importante qualité patrimoniale, mais aussi un certain encombrement visuel lié au mobilier urbain, aux livraisons, aux stationnements des vélos ou encore aux occupations temporaires. Ces éléments peuvent être interprétés par le modèle comme des signes de désordre. Il faut toutefois rester prudent avec le score de *Safety*, car il ne révèle pas complètement sur la sécurité réelle. Il traduit uniquement une perception fondée sur les critères visuels utilisés pour calibrer MIT Place Pulse 2.0. Il s'agit donc davantage d'un indicateur de l'image urbaine que d'un véritable diagnostic de sûreté.

4.2.3. Contreparties et limites résiduelles

Le relevé photographique de terrain demande plus de temps et limite forcément le volume du corpus et son étendue géographique, puisqu'il faut se rendre physiquement sur les différents sites à pied. La prise de vue dépend en outre des conditions rencontrées sur place, notamment de la météo, les contraintes physiques de déplacements et à la marche, les difficultés d'accès à certains espaces, comme les propriétés privées, les zones fermées ou les lieux en travaux, que l'on retrouve également avec Google Street View. et du jugement de l'opérateur dans le choix des lieux et des angles, source de subjectivité dans l'échantillonnage. Reste enfin la limite de fond déjà signalée pour Setia Alam, entière : les notes produites ne sont pas une vérité terrain, mais des prédictions issues d'une perception moyenne mondiale, calibrée sur des jugements qui ne recoupent pas forcément les codes esthétiques et culturels propres au contexte parisien.

4.3. Ingénierie de collecte textuelle : stratification, apprentissage actif et filtrage

4.3.1. Contexte et principes méthodologiques

L'objectif de la collecte est de constituer un corpus suffisamment précis pour suivre l'évolution des perceptions de la ZTL Paris Centre sur 42 mois, de janvier 2023 à juin 2026. Cette période s'étend sur environ deux ans avant la mise en service de la zone et un an et demi après. De plus comme l'étude malaisienne, le fournisseur de tweets utilisé est TwitterAPI.io

La constitution du corpus repose d'abord sur la traçabilité des données. Pour chaque tweet récolté, nous gardons l'identifiant de la requête, la strate associée, la période de collecte, ainsi que l'ensemble des informations du compte (auteur, date de création du compte, nombre d'abonnés, biographie). Ces indices aident à détecter d'éventuels bots. Afin de mesurer leur influence sur les résultats, l'analyse sera d'abord portée sur la totalité du corpus, puis sur un corpus excluant les comptes suspects.

Les requêtes sont classées en trois niveaux :

- **La strate S (mots-clés directs)**, qui regroupe les termes explicitement liés au sujet (exemples : « ZTL », « zone à trafic limité », « piétonnisation »).
- **La strate A (mots-clés contextuels)**, qui rassemble des termes plus larges associable à Paris Centre (exemples : « piste cyclable », « livraison », « commerçant »).
- **La strate B (mots-clés géolocalisés)**, qui inclut des termes ambigus (exemples : « vélo », « fermé », « bloqué ») impérativement combinés à un lieu précis du périmètre (exemples : Rivoli, Châtelet, les Halles, le Marais).

Le nettoyage des données s'effectue en deux temps, au moment de la collecte grâce la requête initiale et puis lors du traitement local. Les premiers filtres éliminent le bruit de fond lié aux sujets connexes mais hors périmètre (transports en commun, sport, politique nationale, faits

divers), tandis que les filtres locaux affinent le corpus au plus près de notre problématique. Cette organisation permet de les ajuster facilement sans avoir à relancer la collecte et, par conséquent, sans engendrer de coûts supplémentaires.

Enfin, les requêtes ont été progressivement ajusté en analysant les fréquences et les expressions récurrentes. Ce travail a permis de repérer les formulations réellement employées par les internautes, mais aussi les principales sources de bruit. Certains mots ont ainsi été retirés parce qu'il ramenait trop de tweets faux positifs, tandis que d'autres ont été ajoutés afin de mieux cibler les lieux et les thématiques associés à la ZTL. Ces ajustements se sont poursuivis jusqu'à ce que l'analyse ne fasse plus apparaître de nouvelle catégorie significative de bruit ni de vocabulaire pertinent.

4.3.3. Collecte finale et pipeline de nettoyage

La collecte finale applique les strates S, A et B enrichies sur 42 mois, produisant 58 207 tweets bruts. Le corpus brut est ensuite traité par un pipeline reproductible en cinq étapes.

Tableau 1 -Étapes du pipeline de nettoyage et justification scientifique de chaque filtrage
Source : GOPALAPILLAI Gokul

Étape	Action	Volume éliminé
1) Déduplication	Suppression des doublons issus des chevauchements entre différents strates	-1 969
2) Retweets	Exclusion des republications pour garder les opinions originales (Boyd, Golder & Lotan, 2010)	-100
3) Réponses courtes	Élimination du bruit conversationnel (réponses de moins de 100 caractères débutant par @)	-3 504
4) Exclusions thématiques	Nettoyage des contenus hors-sujet identifiés par les dictionnaires de bruit créés par les sondes	-19 645
5) Filtrage géographique	Conservation des seuls tweets référant à des lieux cohérents avec le périmètre de la ZTL	-1 275
Corpus après filtration	31 714 tweets gardés (soit 54,6 % du corpus initial)	

Les comptes suspectés d'être des bots restent signalés mais sont conservés à ce stade, afin de mesurer leur impact par comparaison ultérieure. Un taggage thématique multi-label est ensuite

appliqué aux tweets afin de préparer les analyses de sentiment et d'émotions (voir l'annexe IX pour le détail du programme Python pour la collecte).

4.4. Architecture d'analyse NLP : triangulation multi-modèles et détection de l'ironie

Cette phase d'analyse vise à transformer le corpus final de 31 714 tweets en un instrument de mesure longitudinale des perceptions de la ZTL Paris Centre. Elle commence par une étude statistique descriptive (volumes par strate, répartition thématique, structure des comptes suspects), qui caractérise le corpus avant toute interprétation des scores. Puis continue par une analyse puis une analyse NLP (sentiment, émotions, ironie) appliquée à l'ensemble sans exclusion préalable. Ce séquençage répond à un principe de prudence tiré de l'échec malaisien : on ne commente pas des scores de sentiment sans connaître la structure du corpus qui les produit.

4.4.1. Prétraitement du texte

Trois normalisations sont appliquées pour tenir compte des contraintes des modèles de langue, notamment leur vocabulaire fixe et leur fenêtre d'attention limitée à 512 tokens. Les adresses URL sont remplacées par le token <URL> puisqu'elles n'apportent pas d'information utile et prennent inutilement de l'espace. Les mentions de compte sont remplacées par <MENTION> pour éviter que la polarité d'un tweet soit associée à l'élus ou à l'institution mentionnée plutôt qu'au sujet étudié, ici la ZTL. Enfin, les emojis sont convertis en termes décrivant leur sens (« 🙄 » → « visage_en_colère »), car ils jouent un rôle clé dans l'expression des émotions sur Twitter. Les modèles réellement multimodaux étant trop coûteux pour une utilisation sur un processeur standard, cette conversion lexicale constitue un compromis simple et efficace. Comme l'a montré la collecte, aucun nettoyage susceptible d'altérer le contenu n'est effectué : les fautes, les abréviations et le langage familier sont conservés.

4.4.3. Modèles mobilisés (triangulation)

Le sentiment est analysé avec *distilcamembert-base-sentiment*, une version allégée de CamemBERT conçue pour fonctionner efficacement sur CPU. Le modèle donne à chaque texte une polarité et un score de confiance. Cette polarité est ensuite convertie sur une échelle de -1 à +1, ce qui permet de calculer des moyennes et de suivre leur évolution dans le temps. Pour les émotions, le modèle *french_emotion_camembert*, adapté au français, a été utilisé. Il permet de mieux comprendre les expressions familières présentes dans le corpus, comme « velotaf » ou « galère », bien mieux qu'un modèle multilingue.

Étant donné le caractère particulièrement clivant du sujet, s'appuyer sur un seul modèle aurait toutefois été risqué. Le croisement des deux analyses permet de repérer les cas où les résultats divergent, notamment pour identifier de possibles formes d'ironie, et de vérifier dans quelle mesure les conclusions dépendent du modèle utilisé. Cela répond directement à l'un des principaux enseignements tirés du cas malaisien. Enfin, l'inférence est effectuée par lots de 32 tweets, avec une troncature à 512 tokens, sur un processeur standard. Le recours à des modèles distillés permet ainsi de garder un temps de calcul raisonnable.

4.4.3. Détection heuristique de l'ironie

L'ironie est repérée à l'aide d'un indicateur heuristique, *is_irony*, plutôt que corrigée automatiquement. Un tweet est signalé lorsque deux éléments se retrouvent ensemble : une polarité fortement positive (score > 0,6) et des marqueurs négatifs comme « honte », « bouchon

» ou « cauchemar », ou encore certains emojis associés au sarcasme (🙄). L'idée est de repérer un décalage entre sentiment exprimé et vocabulaire réellement utilisé, comme dans « encore une super idée de bloquer Rivoli 🙄 ».

Dans le fonctionnement du processus, les données sont simplement signalées et non supprimées. Le sentiment est donc calculé sur l'ensemble du corpus, y compris les comptes considérés comme suspects, qui pourront ensuite être analysés à part dans les tests de robustesse (annexe C). Les nouvelles informations sont directement ajoutées sous forme de colonnes, sans modifier les données d'origine. Les tableaux croisés sont également exportés avec les figures, afin de pouvoir vérifier chaque résultat directement à partir des données. Enfin, le corpus enrichi est sauvegardé en fin de traitement, ce qui permet de reprendre l'analyse ou de vérifier les résultats sans avoir à relancer toute l'inférence.

4.5 Programme de visualisation et exploitation des séries spatio-temporelles

4.5.1. Dynamique temporelle des perceptions

L'atout de Twitter face aux sondages, ponctuels et coûteux, ou aux entretiens, longitudinaux mais menés sur de petits échantillons, réside dans la possibilité d'une observation continue sur 42 mois à granularité hebdomadaire. Cette résolution temporelle fine est indispensable pour étudier l'impact d'événements exogènes (Jeux olympiques de 2024, épisodes de canicule), les cycles de l'opinion et les effets de la communication politique, notamment les annonces municipales (Mellon, 2013 ; Barberá & Zelizer, 2018).

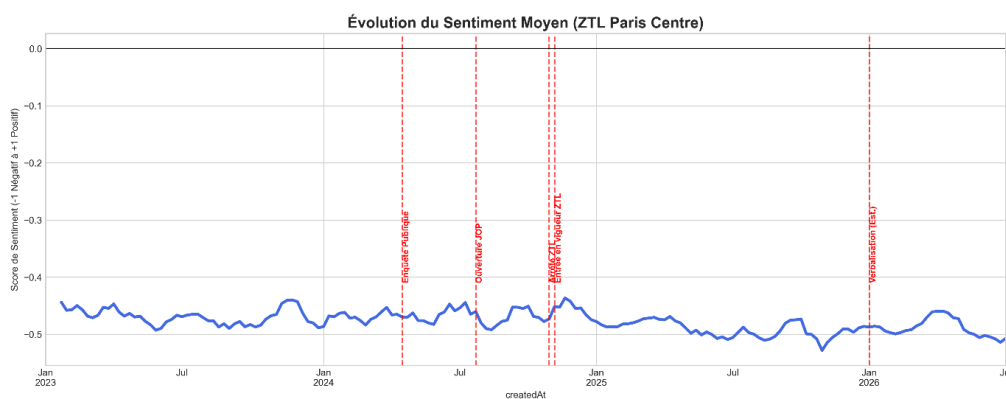


Figure 9 - Évolution du sentiment moyen (ZTL Paris Centre), janvier 2023 – juin 2026.

Courbe hebdomadaire lissée (moyenne mobile sur 4 semaines), annotée des dates clés : enquête publique (avril 2024), ouverture des JOP (26/07/2024), arrêté ZTL (31/10/2024),

Source : GOPALAPILLAI Gokul

D'après la figure 9, le constat principal est sans appel : sur les 42 mois étudiés, le sentiment moyen reste négatif, oscillant entre $-0,3$ et $-0,5$, sans que l'entrée en vigueur de la ZTL ni le début de la verbalisation n'apportent d'amélioration durable. La courbe montre surtout que le mécontentement reste plutôt stable, avec quelques pics autour des événements pointés dans la figure. La contestation est donc présente dès les premières discussions et ne laisse pas vraiment place à une véritable phase d'acceptation. Pour vérifier qu'il s'agit simplement de périodes où le volume de publications augmente, une étude d'événement a été réalisée sur une fenêtre de ± 30 jours autour de chaque date clé. Au-delà de la polarité moyenne, la dynamique fine des

émotions est présentée en Annexe X. Cette visualisation met en avant les alternances entre phases de fort rejet et phases d'adhésion, ce que la courbe de sentiment moyen tend à masquer.

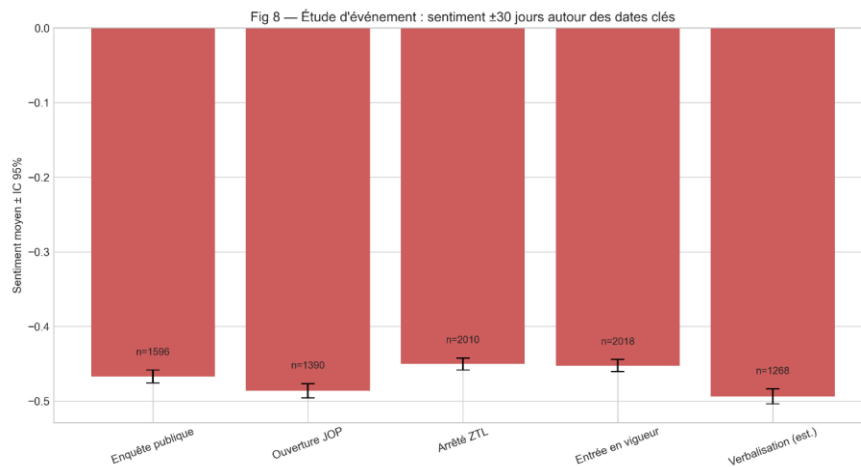


Figure 10 - Étude d'événement : sentiment moyen ± 30 jours autour des dates clés, avec intervalles de confiance à 95 %. Source : GOPALAPILLAI Gokul

La figure 10 confirme cette évolution dans le temps. Les cinq événements étudiés (enquête publique, ouverture des JOP, arrêté ZTL, entrée en vigueur, verbalisation estimée) présentent tous un sentiment moyen nettement négatif, compris entre $-0,3$ et $-0,4$. Les intervalles de confiance à 95 % restent également du côté négatif. Avec 1 268 à 2 018 tweets selon les périodes, les résultats reposent sur des échantillons suffisamment importants. Aucun événement ne fait donc basculer le discours vers une tonalité positive, y compris l'ouverture des JOP, qui n'a pas suffi à apaiser les échanges autour de la ZTL.

4.5.2. Structure thématique du débat

La négativité générale du corpus ne veut pas dire que tous les sujets suscitent le même rejet. L'analyse par thème permet justement de repérer ce qui alimente la contestation et, à l'inverse, ce qui suscite l'adhésion.

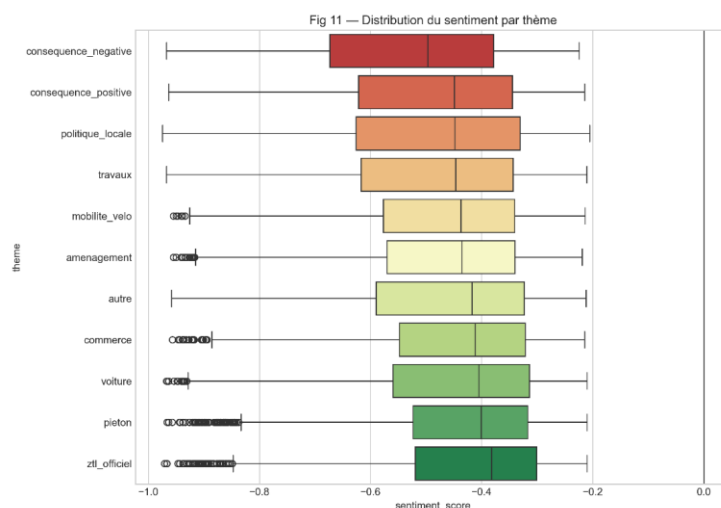


Figure 11 - Distribution du sentiment par thème (boxplots). Source : GOPALAPILLAI Gokul

La figure 11 montre une nette polarisation entre les thèmes. « Consequence_negative » (galère, enfer, mort du commerce) et « voiture » (bouchon, embouteillage, stationnement) concentrent les avis les plus négatifs, avec des médianes sous $-0,5$. À l'opposé, les sujets « consequence_positive » (calme, qualité de vie, respirer) et « mobilite_velo » (piste cyclable, velotaf) obtiennent des scores nettement positifs. Quant aux thèmes « Politique_locale » (Hidalgo, Belliard, mairie de Paris) et « travaux » se situent dans une zone intermédiaire, mais restent principalement négatifs, signe d'un rejet qui vise autant la façon de faire que le projet en lui-même. Cette structure confirme que le débat a deux visages ; il oppose un pôle de contestation (voiture, commerce, travaux) à un pôle d'adhésion (piétonnisation, vélo, qualité de vie), avec un registre politique qui catalyse cette polarisation. Pour visualiser la structure globale du corpus, la carte proportionnelle détaillée (treemap) illustrant le poids relatif de l'ensemble des thématiques est fournie en Annexe XI. Elle confirme la prépondérance des débats institutionnels et des nuisances liées aux chantiers, qui occupent la majeure partie de l'espace discursif.

4.5.3. Dimension spatiale : le sentiment par lieu

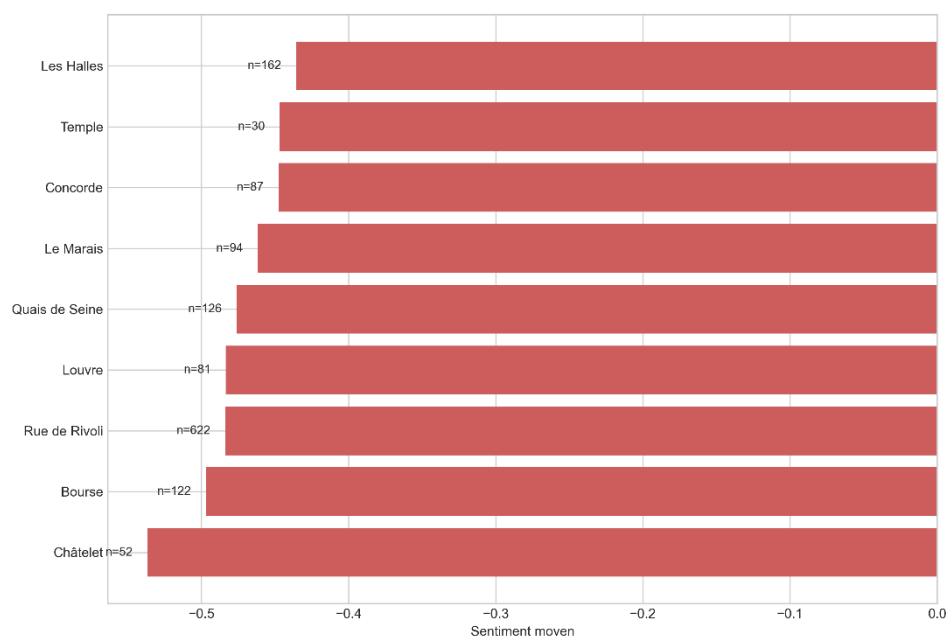


Figure 12 - Sentiment moyen par lieu du périmètre ZTL Source : GOPALAPILLAI Gokul

La figure 12 constitue le pivot spatial de l'analyse Twitter : c'est la seule figure du volet textuel géolocalisée, ce qui permettra de la confronter aux données physiques par axe (voir section 4.6). Ce qu'il faut surtout en retenir, c'est la nette différence selon les secteurs. La rue de Rivoli concentre de loin le mécontentement le plus fort (sentiment moyen autour de $-0,5$, $n = 622$), suivie de près par le Louvre et Châtelet ($\approx -0,4$). Les Halles s'en sortent mieux avec un sentiment nettement moins négatif ($\approx -0,15$, $n = 162$), alors que le Temple, la Concorde, le Marais, les Quais de Seine et la Bourse se situent dans une moyenne intermédiaire ($-0,3$ à $-0,35$). Cette géographie du mécontentement s'explique par le type des voies. Autrefois une véritable artère de transit désormais fermée au trafic de passage, la rue de Rivoli regroupe les tensions autour des reports de circulation et des livraisons. À l'inverse, le secteur des Halles,

largement piétonnisé avant l'arrivée de la ZTL, suscite beaucoup moins de rejet. Le volume de tweets par lieu (de 30 pour le Temple à 622 pour Rivoli) montre aussi l'intensité avec laquelle le débat se cristallise selon les territoires. Cette figure servira de base à la confrontation avec les mesures physiques (réduction sonore, évolution des comptages) menée en section 4.6.

4.5.4. Robustesse et contrôles de qualité

La validité des résultats présentés ci-dessus dépend de leur résistance aux biais structurels de la plateforme. Pour s'en assurer, trois contrôles de robustesse sont détaillés en annexe. L'Annexe XII compare la distribution des sentiments sur l'ensemble du corpus avec celle obtenue après filtrage des comptes suspects : les deux courbes se superposent presque parfaitement, prouvant que cette négativité n'est pas le fruit d'une manipulation ou d'un biais d'automatisation. L'Annexe XIII enfonce le clou : aucune corrélation significative n'apparaît entre le score de « bot » et le sentiment exprimé ($\rho \approx 0,10$). Enfin, l'Annexe XIV apporte un dernier indice d'usage organique : les publications se concentrent aux heures de pointe, le matin (7h-9h) et le soir (18h-21h), un profil typique d'utilisateurs réels plutôt que de comptes automatisés.

Il convient d'adopter un regard critique sur ces premiers résultats. L'analyse du corpus Twitter, dresse un tableau assez uniforme : une négativité structurelle qui peine à révéler de vraies inflexions temporelles. Toutefois, Ces données ne mesurent pas l'efficacité réelle de la ZTL mais traduisent avant tout la polarisation d'un débat numérique, avec ses biais de représentativité et son inclination naturelle à amplifier l'outrage. C'est précisément là que la confrontation aux données physiques (section 4.6) prend tout son sens. Seule, la perception issue des réseaux sociaux reste brute, parfois caricaturale et croisée avec des indicateurs objectifs comme le bruit, la qualité de l'air ou le débit de trafic, elle gagne enfin en nuance et en profondeur. La véritable plus-value de cette étude ne tient donc pas à l'analyse de sentiment prise isolément, mais à cette confrontation entre subjectivité numérique et réalité mesurable du territoire.

4.6. Protocole de confrontation entre données perçues et mesures physiques objectives

4.6.1. Justification de la démarche

L'analyse des sentiments et des émotions produites à partir du corpus de tweets permet de mesurer la réception sociale de la ZTL, puisqu' elle renseigne la saillance du débat public et sa valence, sans pour autant être un instrument de mesure de l'état physique ou économique du territoire. A l'inverse, la seule mesure physique ne suffit pas pour évaluer une politique publique d'apaisement de circulation. En effet, une politique techniquement efficace mais perçue comme un échec sera contestée, révisée ou abandonnée.

Afin d'interpréter correctement les perceptions, il faut confronter les résultats à un second plan de mesure, objectif, composé de quatre familles de données :

- **Qualité de l'air** : mesures journalières PM_{2,5}, PM₁₀, O₃ et NO₂ (Airparif, station Paris 1er, de janvier 2023 à juin 2026) ;
- **Qualité sonore** : écarts en décibels par axe structurant et par heure (Bruitparif), entre les périodes 2022, 2023/2024 et 2025/2026, avec décomposition nuit / pointes / journée et réduction d'énergie sonore ;

- **Comptages routiers et cyclables** (Open Data Ville de Paris) ;
- **Données d'activité commerciale**

L'entrée en vigueur de la ZTL (4 novembre 2024) constitue à cet égard une expérience naturelle : la même logique avant/après est appliquée aux deux plans de mesure, ce qui permet de poser la question centrale de l'étude tout en situant Twitter comme mesure du débat public et non comme succédané d'enquête de satisfaction : les évolutions mesurées du territoire sont-elles perçues, et dans quel sens, par les habitants et usagers ?

4.6.2. Harmonisation et construction du panneau

Pour garantir la comparabilité des données, l'ensemble des séries est aligné sur une échelle temporelle commune (un panneau mensuel de 42 mois) ainsi que sur un découplage spatial partagé qui distingue le périmètre même de la ZTL des axes de mesure environnants. Cette harmonisation s'articule autour de quatre grands choix méthodologiques.

La saisonnalité doit d'abord être prise en compte. La qualité de l'air et la fréquentation commerciale peuvent en effet varier fortement selon les périodes de l'année. Nous travaillons donc soit sur des anomalies, calculées par rapport à la moyenne du mois concerné, soit sur des évolutions annuelles comparées au même mois de l'année précédente. Cette méthode permet d'éviter d'interpréter comme de véritables relations des corrélations qui seraient simplement liées aux variations saisonnières.

Une attention particulière est ensuite accordée à l'interprétation des différents indicateurs. Pour la qualité de l'air, le dioxyde d'azote (NO₂) est utilisé comme principal marqueur du trafic local. Les particules PM_{2.5} et PM₁₀, (surtout liées au chauffage et au contexte régional), ainsi que l'ozone (O₃), sont analysés avec prudence. Pour le bruit, les décibels étant exprimés sur une échelle logarithmique, les données sont calculées en énergie sonore, en distinguant notamment la nuit (22h-6h) et les heures de pointe (7h-10h et 17h-20h). Enfin, pour le trafic, la part du vélo correspond au rapport entre le nombre de cyclistes et celui des véhicules motorisés.

À l'échelle spatiale, les mesures de bruit permettent de comparer les axes situés dans la ZTL avec ceux susceptibles de subir un report de trafic. L'objectif est de voir si les nuisances sonores ont réellement diminué ou si elles se sont simplement déplacées vers d'autres secteurs.

Enfin, le corpus de tweets a été complété par une recherche ciblée sur deux thèmes qui n'étaient pas présents dans le premier taggage : la qualité de l'air et le bruit. Cette extraction repose sur des dictionnaires de mots-clés appliqués selon la même méthode que pour l'identification des lieux, permettant de construire des séries de sentiment directement comparables à chacune des mesures physiques recueillies.

4.6.3. Analyses prévues

Sur ce panneau, sept analyses sont menées : comparaison, par dimension (air, bruit, trafic, commerce), entre évolutions mesurées et perçues ; quadrant croisant changement mesuré et changement perçu ; carte de chaleur des écarts sonores par axe et par heure, complétée par un rapprochement spatial entre baisse du bruit et sentiment exprimé ; analyse des décalages temporels par corrélations croisées (-3 à +3 mois) ; corrélations de Spearman sur séries désaisonnalisées ; et test de cohérence entre bruit et comptages.

Un quadrant sera présenté pour comparer les perceptions des usagers aux évolutions mesurées sur le terrain, pour repérer par exemple les améliorations réelles mais peu perçues. Les résultats sont ensuite vérifiés à travers plusieurs tests de robustesse, notamment en tenant compte des données suspectes et de la période des Jeux olympiques.

Ce petit rapport fixe le cadre, les traitements et les choix méthodologiques de cette analyse comparative, de manière à en garantir la traçabilité et la reproductibilité. Les résultats de cette confrontation ne sont pas développés ici : ils seront présentés, commentés et discutés à l'oral, les figures associées servant alors de support à la discussion des écarts entre trajectoires mesurées du territoire et trajectoires perçues par ses habitants et usagers.

4.7. Bilan critique, limites et reproductibilité de la démarche

Plusieurs limites doivent néanmoins être prises en compte dans l'interprétation des résultats.

La première concerne l'absence de corpus de référence annoté manuellement. Il n'est donc pas possible de mesurer directement la validité des scores produits par les modèles. La démarche repose à la fois sur la comparaison de plusieurs modèles et sur des vérifications lexicales à partir des termes les plus fréquemment associés aux différentes émotions.

La détection de l'ironie présente également une limite importante. Le marqueur utilisé permet d'identifier des publications pouvant relever de l'ironie, mais il ne permet pas de l'établir avec certitude. Cette précaution est d'autant plus importante pour la ZTL, où l'ironie et le sarcasme sont fréquemment utilisés dans les discussions.

Classer par plusieurs étiquette (tagging multi-label) doit lui aussi être interprété avec prudence, car un même tweet peut appartenir à plusieurs catégories, or les proportions thématiques ne pas être additionnées pour représenter la totalité du corpus. De même, la traduction des emojis en expressions textuelles garde une partie de la valeur émotionnelle des messages, mais ne constitue pas une véritable analyse multimodale. De plus, employer des modèles fractionnés permet l'exécution sur un processeur standard, au prix d'une légère perte de précision.

La partie sur la perception visuelle montre la même limite identifiée à Setia Alam. Plutôt que de mesurer objectivement la qualité des espaces photographiés, les notes produites correspondent à des prédictions issues d'une perception moyenne mondiale, qui peut ne pas refléter les codes esthétiques et culturels propres au contexte parisien. La sélection des lieux photographiés comporte également une part de subjectivité.

Par ailleurs, le corpus Twitter manque de représentativité. Car les utilisateurs de la plateforme ne correspondent pas forcément aux habitants ou aux usagers de la ZTL. La population présente sur Twitter est plus jeune, plus urbaine et plus politisée que la moyenne, et rien ne garantit que les personnes qui s'expriment sur la ZTL résident dans le périmètre ou bien l'empruntent régulièrement. Le corpus doit donc être considéré comme une représentation du discours public en ligne autour de la ZTL, et non comme une mesure directe de l'expérience de ses usagers.

Malgré ces limites, l'ensemble de la chaîne de traitement reste reproductible. La collecte, le nettoyage, l'analyse et la visualisation sont scriptés en Python. Les différentes sorties (le corpus de tweets, les tableaux croisés et les figures d'analyse) peuvent ainsi être générées à partir du corpus final sans intervention manuelle supplémentaire. Cette organisation permet de conserver une trace des choix méthodologiques et de faciliter la vérification des traitements effectués.

Conclusion

En somme, ce stage a été une expérience très enrichissante, aussi bien sur le plan professionnel que personnel.

Sur le plan professionnel, travailler au sein d'un consortium international m'a demandé de m'adapter à différentes façons de travailler, d'améliorer mon anglais et de faire évoluer ma manière de communiquer. J'ai aussi appris à travailler de manière rigoureuse surtout pour la partie recherche. Cette expérience m'a permis de gagner en autonomie grâce à la confiance donnée par mon maître de stage, mais également de mieux comprendre comment le monde de la recherche fonctionne et son pouvoir sur les décisions en matière de politiques publiques.

Ces compétences me seront utiles pour la suite de mon parcours lorsque je serai en poste à la Ville de Paris. Pouvoir rejoindre une collectivité aussi engagée dans les transformations urbaines serait pour moi l'occasion de mettre à profit cette double compétence, à la fois dans l'analyse des données et dans la compréhension des réalités du terrain, au service de projets concrets visant à améliorer le quotidien des Parisiens.

Sur le plan personnel, cette période a aussi été importante avec la découverte de la vie en autonomie. Vivre seul m'a beaucoup appris et m'a permis de mieux me connaître. Gérer son quotidien demande de prendre ses responsabilités, de s'adapter et de savoir prendre du recul face aux situations difficiles. C'est quelque chose qui n'a jamais quitté le cocon familial. Cette expérience m'a ainsi permis de gagner en confiance et en maturité, aussi bien dans mon travail que dans ma vie quotidienne.

Il m'a également donné une vision plus concrète de ce que je souhaite pour la suite de ma carrière et de la manière dont je pourrais mettre mes compétences au service de projets ayant un impact concret sur le territoire. Je termine cette expérience avec davantage de confiance et l'envie de poursuivre dans cette direction, avec pour objectif, à terme, de rejoindre la Ville de Paris.

Bibliographie et Sitographie

Documentation officielle du projet européen REALLOCATE

European Commission. (n.d.). *REALLOCATE: Reporting*. CORDIS. <https://cordis.europa.eu/project/id/101103924/reporting>

REALLOCATE. (n.d.). *REALLOCATE: Reallocation of public space for sustainable urban mobility*. <https://reallocatemobility.eu/>

University College Dublin. (n.d.). *Spatial Dynamics Lab*. <https://www.ucd.ie/sdl/>

Articles et publications scientifiques

Adger, W. N., Dessai, S., Goulden, M., Hulme, M., Lorenzoni, I., Nelson, D. R., Naess, L. O., Wolf, J., & Wreford, A. (2009). Are there social limits to adaptation to climate change? *Climatic Change*, 93(3–4), 335–354. <https://doi.org/10.1007/s10584-008-9520-z>

Barbieri, F., Espinosa Anke, L., & Camacho-Collados, J. (2022). XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 258–266.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679.

Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. *2010 43rd Hawaii International Conference on System Sciences*, 1–10.

Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. *PLOS ONE*, 10(8), e0136092.

Dubey et al. (2016). *Deep Learning the City: Quantifying Urban Perception at a Global Scale*

Evans, J. A., & Aceves, P. (2016). Machine translation: Mining text for social theory. *Annual Review of Sociology*, 42, 21–50.

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks*, 38, 16–27.

Kabisch, N., Korn, H., Stadler, J., & Bonn, A. (Eds.). (2017). *Nature-based solutions to climate change adaptation in urban areas: Linkages between science, policy and practice*. Springer. <https://doi.org/10.1007/978-3-319-56091-5>

Kirilenko, A. P., & Stepchenkova, S. O. (2014). Public microblogging on climate change: One year of Twitter worldwide. *Global Environmental Change*, 26, 171–182. <https://doi.org/10.1016/j.gloenvcha.2014.02.008>

Lin, B. B., et al. (2019). Social acceptance and perception of urban green infrastructure. *Urban Forestry & Urban Greening*, 40, 42–50.

Loureiro, M. L., & Alló, M. (2020). Sensing climate change perceptions using digital and social media data. *Current Opinion in Environmental Sustainability*, 42, 44–50.

Tan, L. H., et al. (2021). Social media responses to extreme climate and environmental events in Malaysia. *Journal of Environmental Management*, 285, 112100.

Townsend, L., & Wallace, C. (2016). *Social media research: A guide to ethics*. University of Aberdeen.

Wamsler, C., et al. (2020). Environmental governance and climate change adaptation: Integrating social acceptance and ecological performance. *Environmental Science & Policy*, 108, 79–88.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

Modèles et ressources techniques

Hartmann, J. (2022). *Emotion English DistilRoBERTa-base* [Modèle de langage]. Hugging Face. <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base>

Normes techniques et cadres réglementaires

International Organization for Standardization. (2006). ISO 14044:2006: Environmental management—Life cycle assessment—Requirements and guidelines. ISO.

Department of Standards Malaysia. (n.d.). MS 2680: Energy efficiency and renewable energy in non-residential buildings—Code of practice. Department of Standards Malaysia.

Green Building Index. (n.d.). *Green Building Index Malaysia*. <https://www.greenbuildingindex.org/>

Ville de Paris. (2024). Arrêté portant création de la zone à trafic limité Paris Centre. Ville de Paris.

Sources institutionnelles et données

Airparif. (n.d.). *Airparif: Surveillance de la qualité de l'air en Île-de-France*. <https://www.airparif.asso.fr/>

Bruitparif. (n.d.). *Bruitparif: Observatoire du bruit en Île-de-France*. <https://www.bruitparif.fr/>

Ville de Paris. (n.d.). *Open Data Paris*. <https://opendata.paris.fr/>

Outils logiciels et plateformes

Esri. (n.d.). *ArcGIS StoryMaps*. <https://storymaps.arcgis.com/>

Google. (n.d.). *Google Street View*. Google Maps. <https://www.google.com/streetview/>

Python Software Foundation. (n.d.). *Python*. <https://www.python.org/>

Annexes

Annexe I: Tableau de présentation des différents Labs de chaque ville pour le projet Reallocate traduite en français Source : Reallocatemobility.eu

Binôme	Lead City (pays)	Pilote(s) de la Lead City	Twin City (pays)	Pilote(s) de la Twin City
1	Barcelone (Espagne)	LAB 1 : Diagnostic des conflits d'usage et de l'accidentologie entre piétons, cyclistes et engins de micromobilité dans les espaces partagés. Mise en œuvre d'audits, de sensibilisation et de solutions de réduction des conflits. LAB 2 : Amélioration de l'accessibilité des transports publics pour les personnes en situation de handicap via un service public de véhicules partagés en porte-à-porte afin d'optimiser la flotte et d'accroître les impacts sociaux et environnementaux.	Bologne (Italie)	LAB unique : Création d'un quartier scolaire sûr et durable le long de la <i>Via della Conoscenza (Quartiere Navile)</i> , co-conçu avec les élèves et acteurs locaux (stationnement vélo, recharge d'e-bikes, assises, fontaines et audits de sécurité intégrant la végétation).
2	Budapest (Hongrie)	LAB 1 — Déploiement de « Healthy Superblocks » dans le Ville arrondissement combinant réorganisation de la circulation, réduction de la pollution, désimperméabilisation et sécurité des usagers vulnérables. LAB 2 — Sécurité routière périurbaine via la planification et le monitoring (vitesse, caméras, air/bruit) suivis d'interventions sur les vitesses, voies bus/vélo, feux intelligents et comportements.	Zagreb (Croatie)	LAB unique : Aménagement du corridor central (<i>Selska – Horvatska</i>) s'appuyant sur 700 caméras/radars, des feux intelligents, le redimensionnement des carrefours/pistes cyclables et la priorité aux transports publics et modes actifs.
3	Göteborg (Suède)	LAB 1 — Approche « Safe System » pour la mobilité active des enfants en zone périurbaine (<i>Bergum Gunnlise</i>), incluant un hub virtuel de covoiturage, l'éclairage, la régulation des vitesses et la transition vers un « village du quart d'heure ». LAB 2 — Gestion de la mobilité pendant le chantier de <i>Västlänken</i> à <i>Korsvägen</i> grâce à un jumeau numérique, des visualisations 3D, le réaménagement du stationnement/dépose-minute et le nudging.	Tampere (Finlande)	LAB unique : Emploi de l'IA et d'une plateforme IoT (<i>Vuores</i> et carrefours scolaires) pour identifier les zones à risques, analyser les données de trafic via caméras IA, engager les citoyens et réallouer l'espace public.
4	Heidelberg (Allemagne)	LAB 1 : Elaboration de plans de mobilité domicile-travail régionaux dotés de hubs adaptatifs en partenariat avec les communes limitrophes pour réduire les émissions générées par les navetteurs. LAB 2 : Réallocation tactique de l'espace dans quatre rues pour créer des zones apaisées en convertissant le stationnement et en protégeant les piétons/cyclistes (notamment enfants et	Utrecht (Pays-Bas)	LAB unique : Sécurisation participative des abords d'écoles dans les quartiers vulnérables en combinant outils immersifs (VR, 3D, jumeau numérique), réduction du stationnement, zones 30 km/h, solutions fondées sur la nature et accompagnement comportemental.

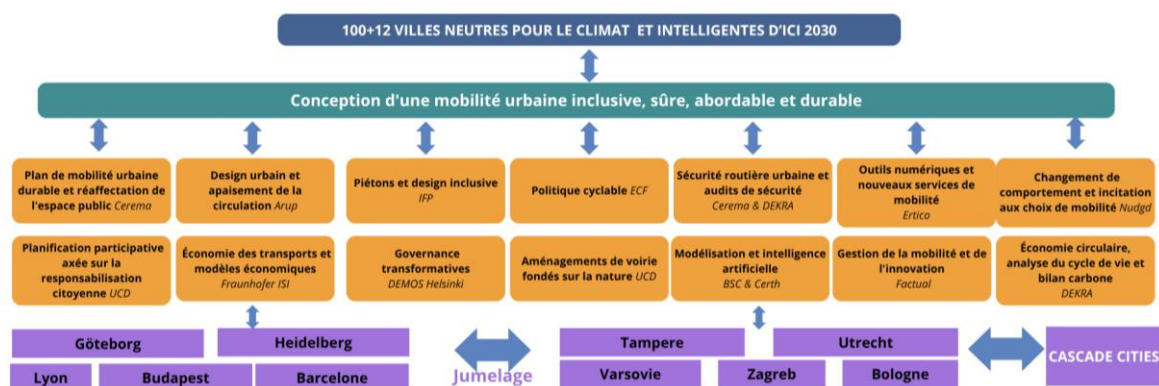
Binôme	Lead City (pays)	Pilote(s) de la Lead City	Twin City (pays)	Pilote(s) de la Twin City
5	Lyon (France)	personnes âgées). LAB 1 : Action « Vision Zéro » à hauteur d'enfant sur 15 sites scolaires (plus de 4 000 enfants impliqués) : mesure des flux/pollutions, apaisement des abords, zones piétonnes, préfiguration de <i>superblocks</i> et interventions artistiques/végétales. LAB 2 : Exploitation des données de mobilité et jumeau numérique, lecture automatisée des plaques, tarification du stationnement modulée selon le poids/moteur et déploiement de stationnements vélo sécurisés.	Varsovie (Pologne)	LAB unique : <i>Projet Warsaw's green & safe road to school</i> autour d'une école du centre-ville, réunissant audits des passages piétons, capteurs participatifs (climatic/trafic), enquêtes de mobilité par GPS/marche, modélisation 2D/3D gamifiée et réallocation de voirie avec éclairage LED

Les Cascade Cities

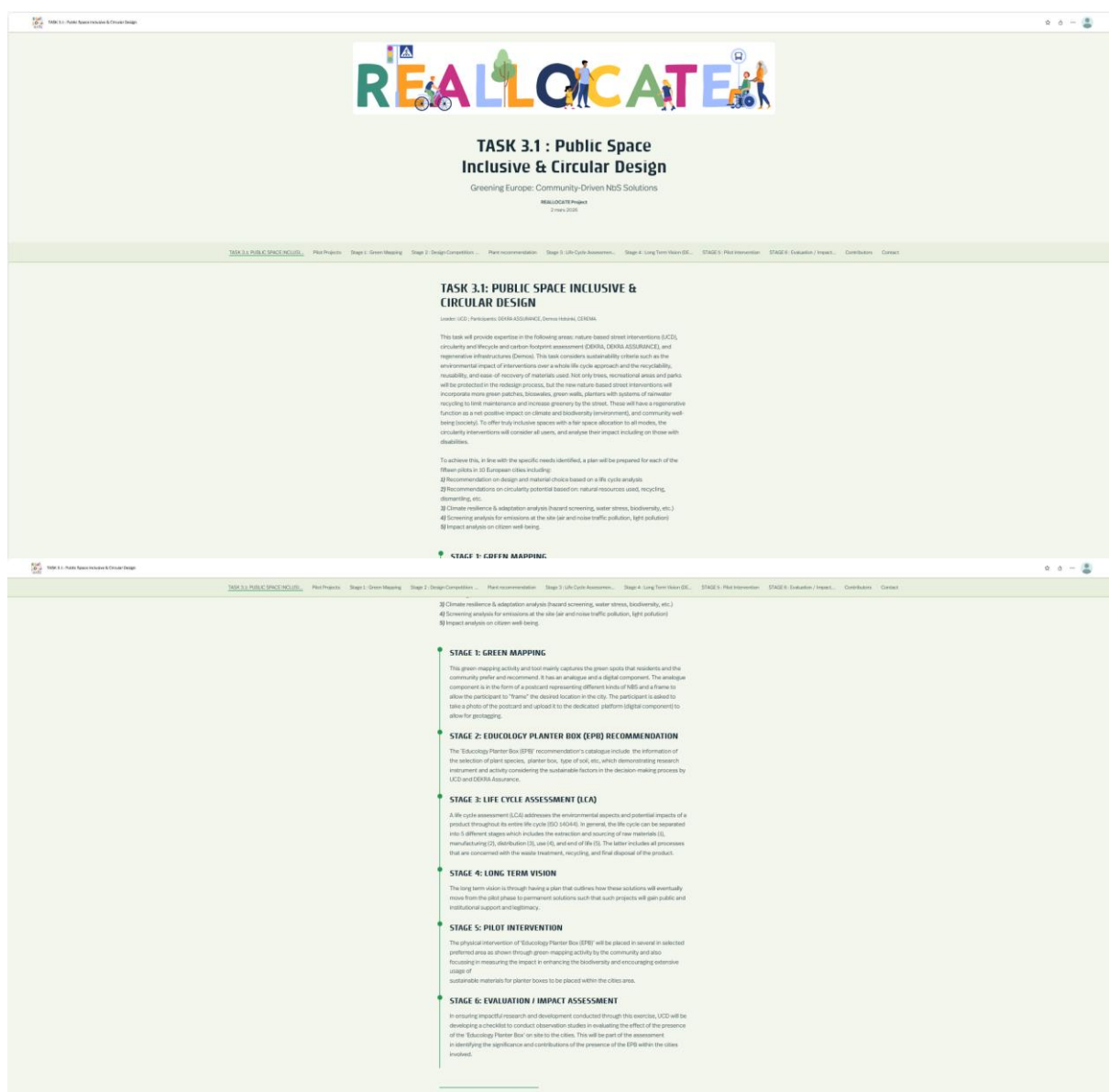
Villes associées bénéficiant du transfert de méthodes et de connaissances développées par le consortium.

Braga (Portugal)	Braşov (Roumanie)	Gand (Belgique)
Région urbaine de Ljubljana (Slovénie)	Lviv (Ukraine)	Rhodes (Grèce)
Canton de Sarajevo (Bosnie-Herzégovine)	Străşeni (Moldavie)	Tbilissi (Géorgie)
Turku (Finlande)	Varna (Bulgarie)	

Annexe II: Cartographie des domaines d'expertise et répartition des partenaires auprès des villes du projet REALLOCATE traduite en français Source : Reallocatemobility.eu



Annexe III: Captures d'écran illustratives de la Storymaps principale (Des Storymaps complémentaires ont été créées pour chaque partie afin d'enrichir le propos sans la surcharger) Source : GOPALAPILLAI Gokul



[TASK 1.1: PUBLIC SPACE INCLUSION](#)
[Pilot Projects](#)
[Stage 1: Green Mapping](#)
[Stage 2: Design Competition](#)
[Pilot Recommendation](#)
[Stage 3: Life Cycle Assessment](#)
[Stage 4: Long Term Vision \(LTC\)](#)
[STAGE 5: Pilot Intervention](#)
[STAGE 6: Evaluation / Impact](#)
[Contributors](#)
[Contact](#)

Pilot Projects

This part focuses on six pilot projects in the Resilocate project: Barcelona, Bologna, Budapest, Heidelberg, Utrecht and Warsaw. Each site is briefly presented in the following map and are provided by links for more details regarding their specific development.

Ajuntament de Barcelona
Barcelona, Spain

Comune di Bologna
Bologna, Italy

BUDAPEST
Budapest, Hungary

Heidelberg
Heidelberg, Germany

Universiteit Utrecht
Utrecht, Netherlands

Warsaw, Poland

[TASK 1.1: PUBLIC SPACE INCLUSION](#)
[Pilot Projects](#)
[Stage 1: Green Mapping](#)
[Stage 2: Design Competition](#)
[Pilot Recommendation](#)
[Stage 3: Life Cycle Assessment](#)
[Stage 4: Long Term Vision \(LTC\)](#)
[STAGE 5: Pilot Intervention](#)
[STAGE 6: Evaluation / Impact](#)
[Contributors](#)
[Contact](#)

Stage 1 : Green Mapping

There are three types of green mapping activity approaches that we recommended to the city to adopt (you may choose one or two) in conducting mapping activity by the local community in collecting data of their most preferred spot for the placement of the Educology Planter Box (EPB).

Including:

- 1) Survey Mapping
- 2) Postcard
- 3) Green Image Survey
- 4) Co-creation Green Mapping Workshops

The full description of each method is described in the Storymaps below.

Green Mapping

Recommended green mapping activity approaches

1) Survey Mapping

Stage 2 : Design Competition: EDUCOLOGY PLANTER BOX (EPB)

The EPB prototype is actually an M0 prototype of a planter box, which consists of sustainable factors in decision-making for multiple factors and variables, including type of species, material of the planter box, type of soil, etc. The EPB prototype will also include a guideline for sustainably maintaining the physical prototype itself. This activity will also be furnished with an M0 dashboard dedicated to providing a relevant catalogue for the usage of the municipalities (decision-makers) and the communities.

Catalogue Plant Species to Pollinator

The plant species associated with pollinators and pollination in Europe exhibit a range of adaptations to their environments, reflecting the complex interactions between plants and their pollinators. These adaptations are crucial for ensuring successful reproduction and maintaining ecosystem health. European plant species have developed various strategies to attract and utilize pollinators, which include insects, birds, and bats. These strategies are often shaped by the specific ecological and evolutionary contexts of their habitats.

Example of an Educology Planter Box made with cardboard

So, a design competition was organized for UCD students to develop the a prototype of an EPB.

Stage 2 : Design Competition: EDUCOLOGY PLANTER BOX (EPB)

The EPB prototype is actually an H&M prototype of a planter box, which consists of various factors in decision-making for multiple factors and outcomes, including type of species, material of the planter box, type of soil, etc. The EPB prototype will also include a guideline for sustainably maintaining the physical prototype itself. The activity will also be furnished with an H&M dashboard dedicated to providing a relevant catalogue for the usage of the municipality (decision-making) another community.

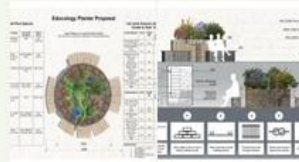


Example of an Educology Planter Box made with Ceramic II

Catalogue Plant Species to Pollinator

The plant species associated with pollination and pollination in Europe exhibit a range of adaptability to their environments, reflecting the complex interactions between plants and their pollinators. These adaptations are crucial for ensuring successful reproduction and maintaining ecosystem health. European plant species have developed various strategies to attract and utilize pollinators, which include insects, birds, and bats. These strategies are often shaped by the specific ecological and evolutionary contexts of their habitats.

As a design competition was organized for UCL students to develop the prototype of an EPB.



Planter Box Design: Educology Planter Box Project
Proposed by Gabe, Paloma & Lyndee Bold-Plant

Description from the creators: Our planter is made from low carbon and recycled materials. It is circular in design and is constructed primarily from upcycled clay pots and recycled pallet timber. The diameter shown in the drawings is 3.5meters, however it can be exponentially scalable to accommodate any public space. Embedded carbon for LCA ranged at 1. All was calculated to be 6.4 kg of CO2 equivalent. The planter was featured in a museum made from pallet timber and recycled pallet timber. Embedded carbon for the bench was calculated to be 28 kg CO2e. All of the materials used in both the planter and its seating are either biodegradable or recyclable.

The planter boxes display into the ground, preventing rainwater from entering the municipal sewer system. It may also be placed under existing downpipes to collect runoff. The plants we have proposed are all native, drought tolerant perennials with bloom periods ranging from March to November, ensuring a continuous food supply for a variety of pollination. The design of our planter is also intended to foster collaboration and community participation. It can be easily associated with related traits and construction knowledge, and local residents should be encouraged to participate in the assembly and planting, as well as the maintenance of the vegetation. The waste of the planter are intentionally left loose, intended to serve as a canvas for future works of public art.

The projects listed below expand the second, third, and fourth-phase concepts in the design context.



Illustration of various plant species and their interactions with pollinators.

For more details on the design competition, please follow the storymap link below for full descriptions of each project.

Design Competition: Educology Planter Box (EPB)

Design Competition: EPB

2021-2022



Revisiting the urban space for beyond adding planter boxes or removing cars, it actually transforms barren urban spaces into living ecosystems. Explore the Storypages below for examples of Nature-based Solutions (NbS) in action.

Nature-based Solution

Nature-based Solution

2021-2022



Stage 3 : Life Cycle Assessment (DEKRA)

A life cycle assessment (LCA) addresses the environmental aspects and potential impacts of a product throughout its entire life cycle (ISO 14046). In general, the life cycle can be separated into different stages which include the extraction and sourcing of raw materials (1), manufacturing (2), distribution (3), use (4), and end of life (5). The latter includes all processes that are concerned with the waste treatment, recycling, and final disposal of the product.

To measure these effects, mass and energy flows are related to a database to compute Global Warming Potential (GWP) in units of kg CO2e. For the Realizate project, the Carbon Footprint (CFP) acts as a key tool to enable sustainable decision-making in the context of the Heidelberg pavilion, the scope of this assessment is set to a 'cradle-to-gate + End of Life' approach. Beyond greenhouse gas emissions, this project examines circularity through the percentage of secondary materials and recyclability of each type. This approach guarantees that urban interventions focus on resource efficiency and nature-based solutions in procurement processes.

The following summarizes the LCA methodologies provided by DEKRA assurance.

STAGE 3: LIFE CYCLE ASSESSMENT

Life cycle assessment (LCA) addresses the environmental aspects and potential impacts of a product throughout its life cycle.

from: [https://www.dekra.de](#)

	Product	Product
1	100	100
2	100	100
3	100	100
4	100	100
5	100	100
6	100	100
7	100	100
8	100	100
9	100	100
10	100	100
11	100	100
12	100	100
13	100	100
14	100	100
15	100	100
16	100	100
17	100	100
18	100	100
19	100	100
20	100	100
21	100	100
22	100	100
23	100	100
24	100	100
25	100	100
26	100	100
27	100	100
28	100	100
29	100	100
30	100	100
31	100	100
32	100	100
33	100	100
34	100	100
35	100	100
36	100	100
37	100	100
38	100	100
39	100	100
40	100	100
41	100	100
42	100	100
43	100	100
44	100	100
45	100	100
46	100	100
47	100	100
48	100	100
49	100	100
50	100	100
51	100	100
52	100	100
53	100	100
54	100	100
55	100	100
56	100	100
57	100	100
58	100	100
59	100	100
60	100	100
61	100	100
62	100	100
63	100	100
64	100	100
65	100	100
66	100	100
67	100	100
68	100	100
69	100	100
70	100	100
71	100	100
72	100	100
73	100	100
74	100	100
75	100	100
76	100	100
77	100	100
78	100	100
79	100	100
80	100	100
81	100	100
82	100	100
83	100	100
84	100	100
85	100	100
86	100	100
87	100	100
88	100	100
89	100	100
90	100	100
91	100	100
92	100	100
93	100	100
94	100	100
95	100	100
96	100	100
97	100	100
98	100	100
99	100	100
100	100	100

This is the excel sheet of Life Cycle Assessment (LCA) mathematical equation, in assisting the data in calculating the LCA.

Calculation Sheet Heidelberg

Stage 4 : Long Term Vision (DEMOS)

It is indispensable for REALLOCATE projects to have a long-term vision. They are to transcend short-lived solutions and create pedestrian-focused and sustainable urban spaces. It is through having a plan that outlines how these solutions will eventually move from the pilot phase to permanent solutions that such projects will gain public and institutional support and legitimacy. It becomes also possible to increase the resilience and sustainability of such solutions. Anticipating risks (e.g. declining community commitment, regulatory barriers, or funding shortages) enables proactive mitigation and strategic resource allocation. As a result, this will produce long-lasting solutions that improve residents' physical and mental health.

This long-term approach is exemplified by the planter box intervention discussed in the Storymaps below. While it was initially created as a temporary intervention aimed at testing innovation and public commitment and support, it was created in such a way that it was scalable and permanent.



The Integrated Green Framework: Vision, Pathways, and Implementation

LONG TERM VISION

Building Resilient Long-Term Vision for Urban Green Infrastructure (NBS) via Planter Box

2021-2022



[illegible]

Recommendation (Option 3)

DECISION MATRIX: Namysłowska Street Intervention

Comparison of the proposed V1 (Recommended) and Proposed V2 (Disallowed)
Summary of Mitigation Requirements
What must be done to make the selected option viable (See Table)

V2 (Recommended)

The Daylighting Scheme: The first 5 metres of furniture near the pedestrian crossing MUST be low planters (300mm) to ensure the visibility aids.
Drainage Offset: Ensure a 200mm gap from the kerb for water flow.

V1 (Not Recommended)

Structural Anchoring: Every high planter must be bolted to the ground or a trench.
Insulation: Benches must be lined with foam VPE, reducing planting width to 300mm.
Plant List: Only plants: 100% DCL, all perennials. No plants with cold hardy grasses (Miscanthus), Redwood only.

Criteria	Proposal V1 (300cm width)	Proposal V2 (300cm width)	UCD Assessment
1. SAFETY: Stability	High Risk. A 100cm tall box with a 60cm base is prone to tipping if climbed on. Mandatory chemical anchoring to asphalt is required.	Low Risk. A wider base offers natural gravity stability. May not require invasive bolting (depending on wind load ratio).	V2 Wins (Safer for School Zones).
2. SAFETY: Visibility	Medium Risk. A narrower profile improves sightlines slightly, but the height (100cm) still obscures children.	High Risk. A wider profile creates a larger visual blind spot near crosswalks.	Tie. Both require the Daylighting sweep (low planters near zebra).
3. BIODIVERSITY	Critical Failure. 100cm width = 100% root issues in Warsaw winters. Shades will not survive. Limited to grasses/perennials.	Success. 300cm width provides a thermal buffer. Supports mostly shrubs (Rosa, Dogrose), mandated by the Biodiversity Protocol.	V2 Wins (only viable option for "Green" goals).
4. PLACEMAKING	Paper Wall Effect. Thin barrier feels flimsy to pedestrians. Offers little psychological protection from passing cars.	Fortress Effect. Creates a robust physical shield. Pedestrians feel safe sitting with their back to traffic.	V2 Wins (higher comfort, speed time).
5. LCA (Carbon)	Wasted Investment. Lower initial carbon, but high risk of plant death = zero ecosystem service return.	Positive ROI. Higher initial carbon (more steel), but plant longevity ensures carbon capture & cooling over 10+ years.	V2 Wins (better long-term value).

Plant Catalogue

The planting catalogue for the Namysłowska Street plan draws from traditional urban landscaping by whitening a Functional Biodiversity approach. Unlike standard decorative planting, which often relies on annuals with high water demands and/or irregular size, this palette is engineered to function as a self-sustaining ecosystem within the constraints of the Zero-modular infrastructure. Species were selected based on a rigorous framework to ensure compliance with REKULOGIC Task 3.0 objectives:

A. Climatic Resilience (The Steel Box Factor): Plants must survive in isolated steel containers which freeze rapidly in winter and overheat in summer. Selected species are rated for Zone 9-10, ensuring survival during Warsaw's continental winters without requiring chemical/irrigation.

B. Native Provenance: To prevent biotic homogenization, 100% of the structural biomass consists of species native to the Macoian region or historical naturalized relicts (archophytes).

C. Trophic Support: The palette prioritizes High Nectar Value species that bridge the Hunger Gap, providing food sources from early May (Spring) to late October (Autumn).

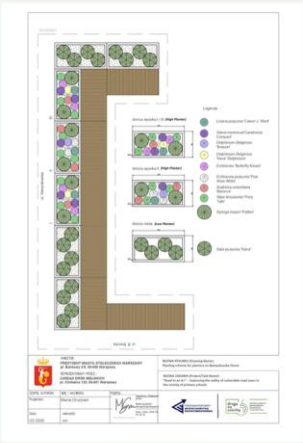
The Warsaw Box Biodiversity Palette

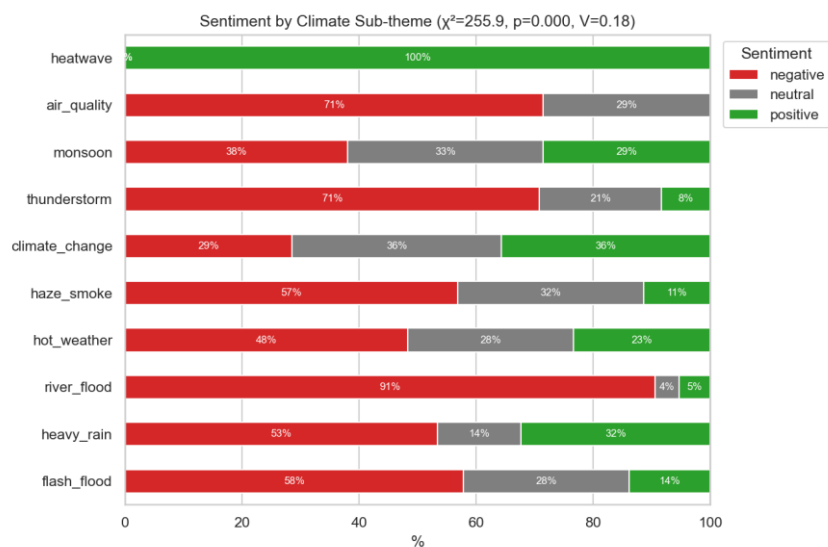
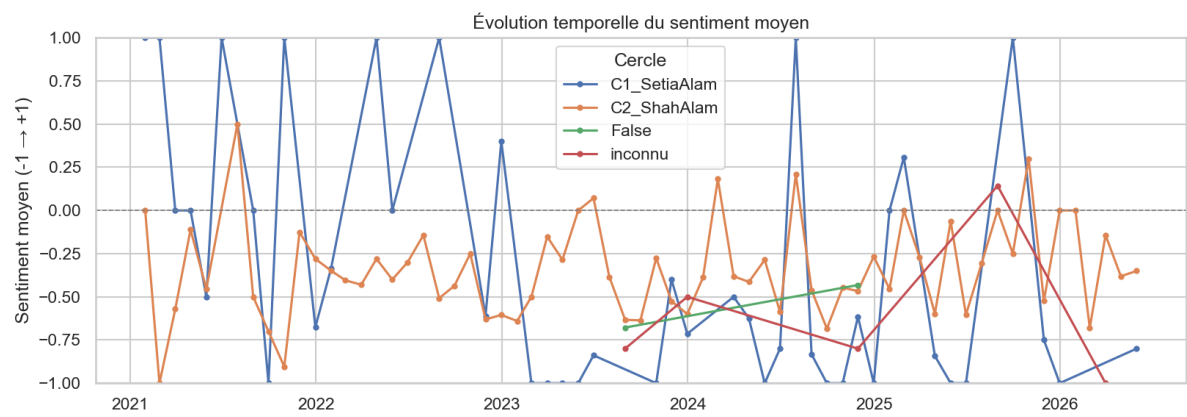
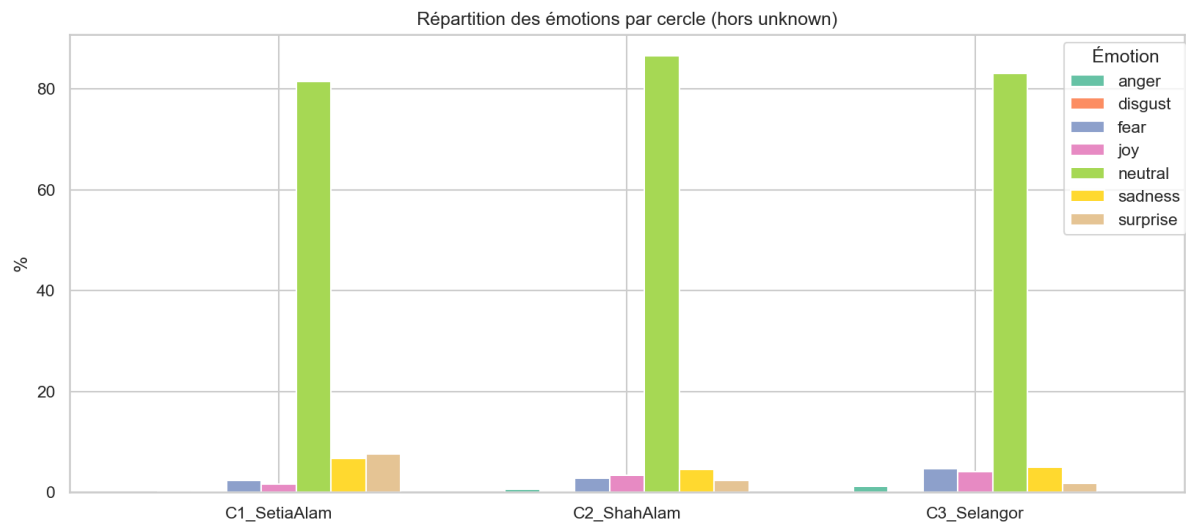
For High Planters (20 x 30 x 100 cm)

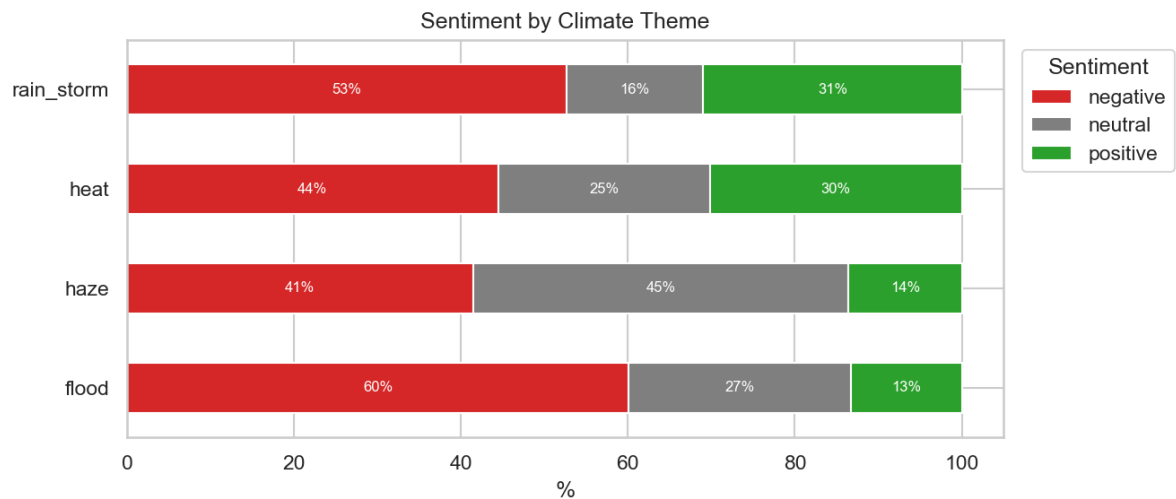
Conditions: Dense soil, potential for root freezing at edges. Needs structural shrubs.

Plant Species (Latin / Polish)	Native Status	Bloom Time	Target Pollinator	Specific Ecological Gains (Task 3.0)
Desfontainea frutescens (Rękoszka kieszonkowa)	Native (Relict)	May - Oct (Continuous)	Hoverflies & Solitary Bees	The "Workhorse": Extremely cold hardy (Zone 9). Provides stable structure and Nectar Class A for 8 months.
Rosa canina (Róża dzika)	Native	June - July	Bumblebees (Bombus)	The "Barrier": Thorny structure prevents climbing. Open yellow-rich flowers feed long-tongued bumblebees. Red hips feed birds in winter.
Spiraea salicifolia (Tawuła wierzbołowa)	Native	June - Aug	Honeybees & Bees	The "Fuzzy" Bloomer: Dense clusters of tiny flowers accessible to short-tongued insects (Bee, bumblebee) that cannot open complex flowers.
Lonicera xylosteum (Włocławek suchobrzewi)	Native	May - June	Moths & Night-fliers	The "Night Shift": Fragrant flowers attract evening pollinators (moths), increasing biodiversity beyond daylight hours.

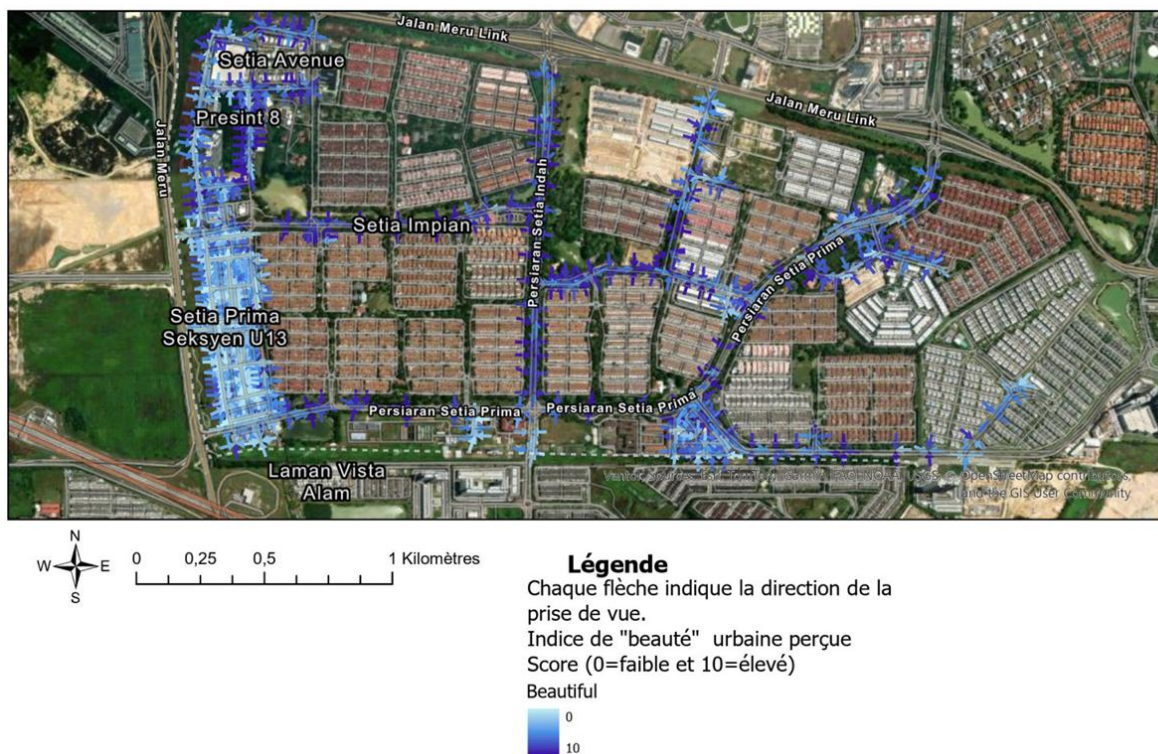
The following presents the translated detailed quantities and specifications for the greening of the Namysłowska Street planters, followed by the corresponding schematic planting diagram.

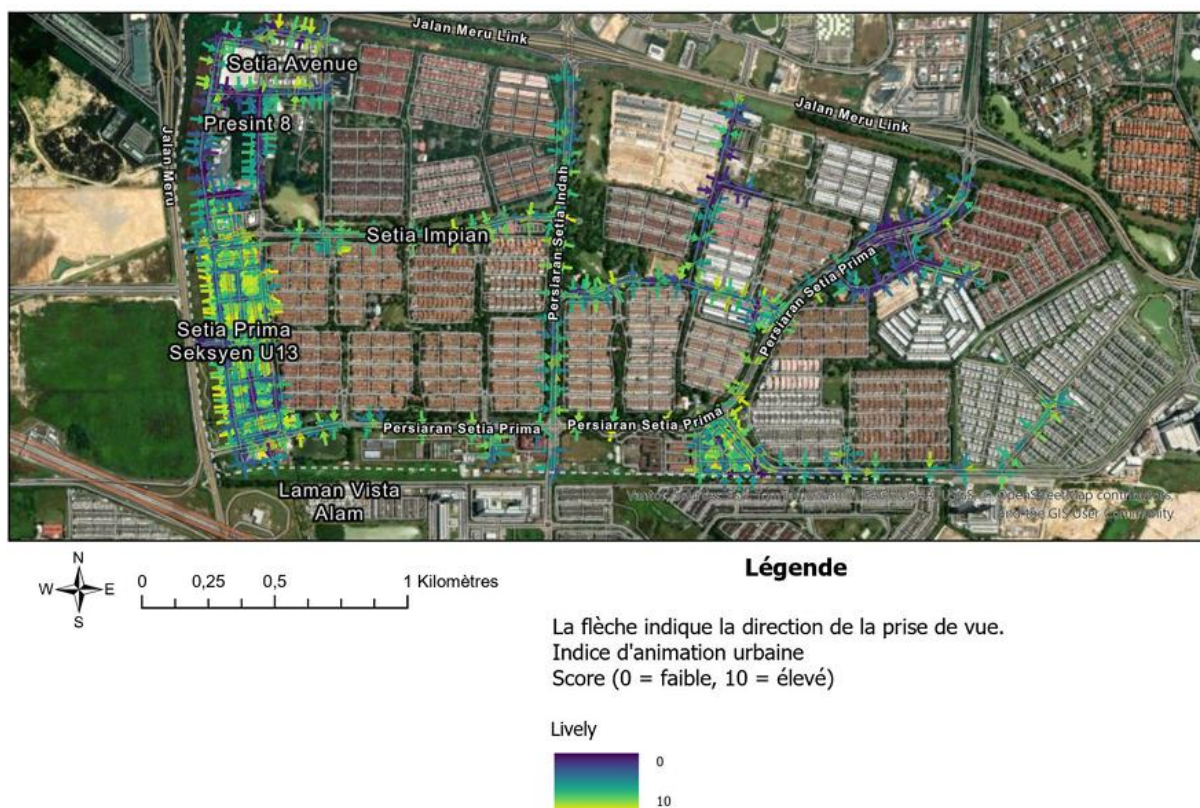




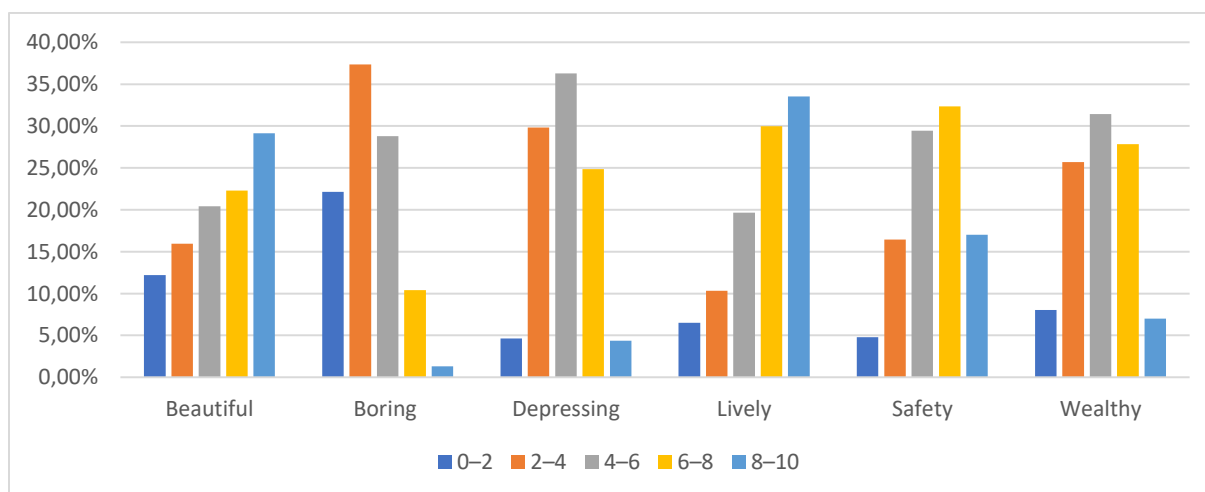


Annexe V: Carte des scores de beauté et d'ambiance urbaine perçue (Beautiful et Lively) le long des axes couverts à Setia Alam sur Google Streetview Source : GOPALAPILLAI Gokul

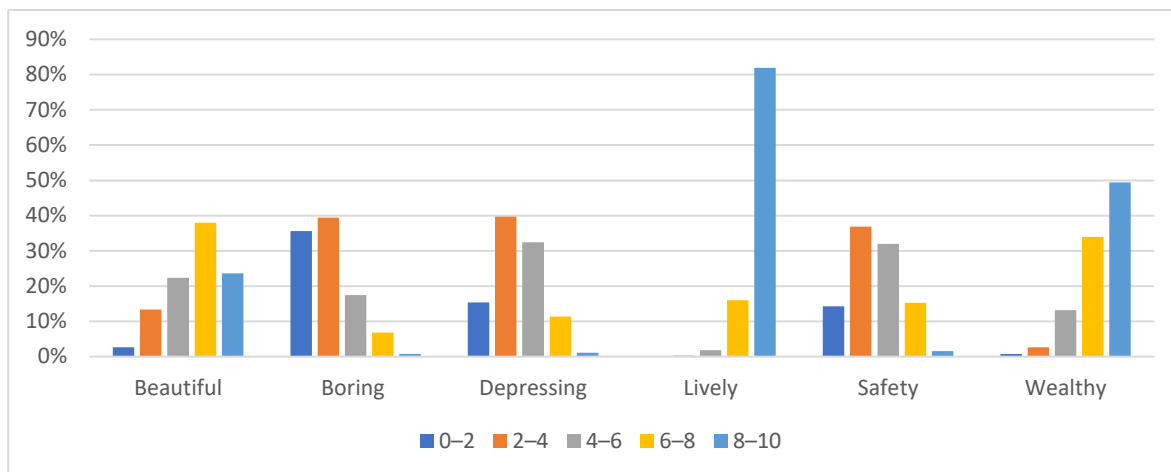




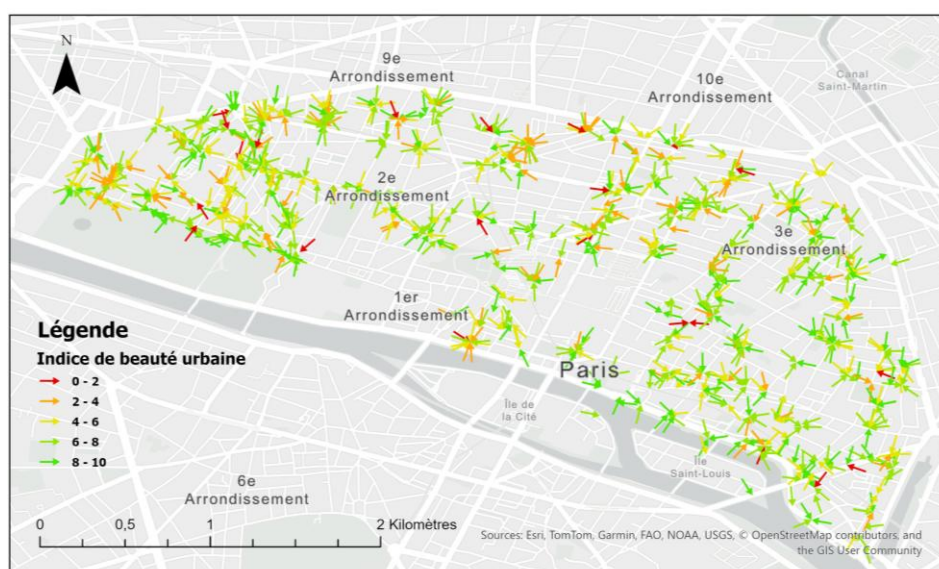
Annexe VI: Les histogrammes détaillés des scores pour les six adjectifs des captures d'écran prises sur Google Streetview pour le corpus de Setia Alam Source : GOPALAPILLAI Gokul

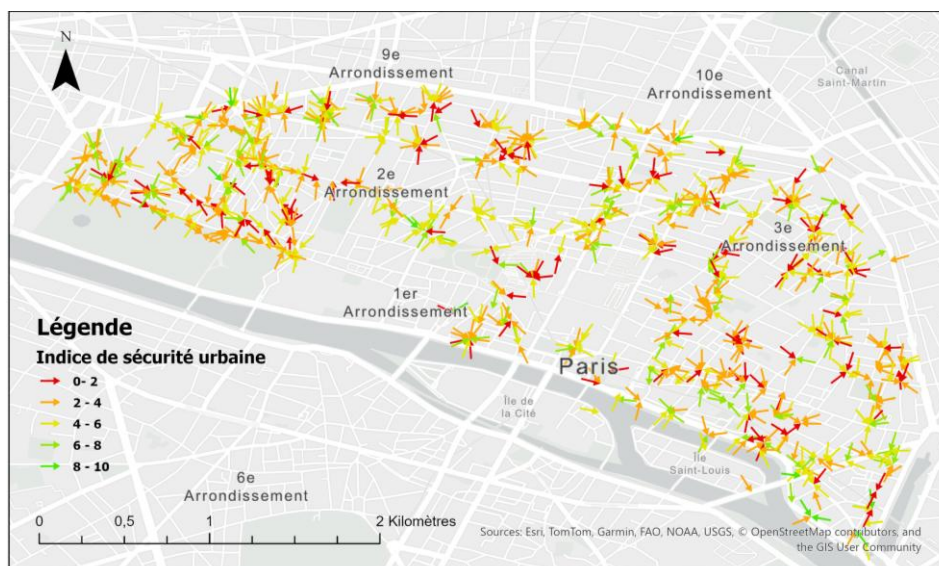
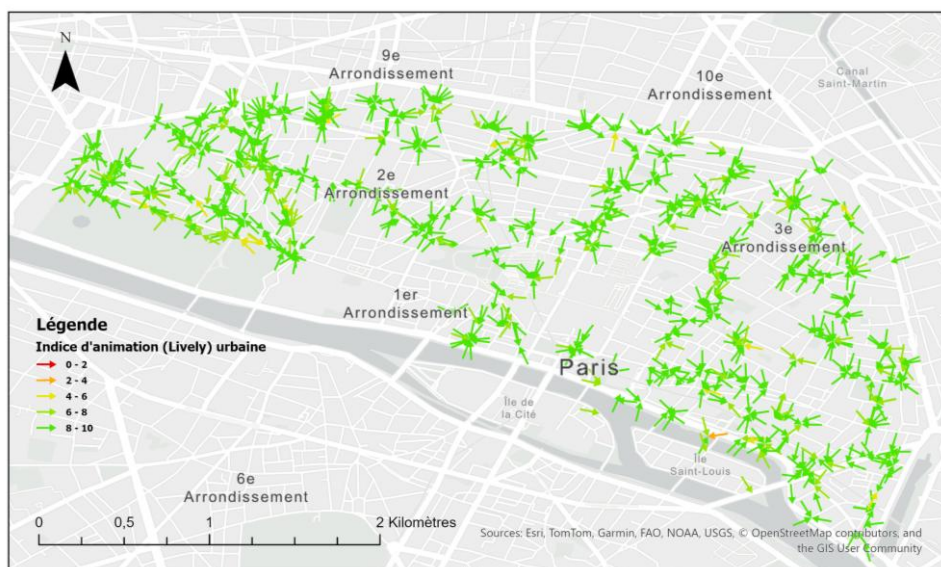


Annexe VII: Histogrammes des six scores (Beautiful, Boring, Depressing, Lively, Safety, Wealthy) pour le corpus parisien Source : GOPALAPILLAI Gokul



Annexe VIII: Cartes de scores (Beautiful, Lively et Safety) produites sous ArcGIS Pro
Source : GOPALAPILLAI Gokul





Annexe IX: Programme Python de collecte et de filtrage de tweets utilisés pour l'étude de la ZTL de Paris Source : GOPALAPILLAI Gokul

```

1 }

'''
Programme de collecte

import requests
import time
import csv
import json
from datetime import datetime, timedelta
from itertools import product

# =====
# CONFIGURATION
# =====

API_KEY = "newl_f8bfa66220b2ff8de0b1ac9d0c89"
ENDPOINT = "https://api.twitter.p1.io/twitter/tweet/advanced_search"
HEADERS = {'X-API-key': API_KEY}

# Période complète
DATE_DEBUT = datetime(2023, 1, 1)
DATE_FIN = datetime(2026, 6, 30)

# Limites
MAX_TWEETS_PAR_REQUETE = 1000
PAUSE_API = 1.5 # secondes

# Fichier de sortie
OUTPUT_FILE = f'txt_paris_2023_2026_complet_{datetime.now().strftime("%Y%m%d_%H%M%S")}.csv'

# =====
# DICTIONNAIRES DE REQUÊTES (Optimisés suite aux sondes)
# =====

# STRATE S : Termes auto-suffisants
STRATE_S = [
    'ZIL',
    'zone à trafic limité',
    'piétonnement',
]

# STRATE A : Termes + autre Paris
STRATE_A_THEMES = [
    "pisto cyclable",
    "livraison",
    "(commerçant OR \"chiffre d'affaires\" OR clientèle)",
    "(désigne OR boillard OR \"maire de Paris\")",
]

ARCHES_PARIS = {
    "Paris OR \"PARIS\" OR \"PARIS\" OR \"PARIS\" OR \"PARIS\" OR \"1er arrondissement\" OR \"2ème arrondissement\" OR \"3ème arrondissement\" OR \"4ème",
}

# STRATE B : Termes + lieux précis (optionnel, plus brutes)
STRATE_B_LIEUX = [
    'rive',
    "rue de rive",
    'château',
    'hall',
    "les halles",
    "saint-denis",
    "saint-martin",
    'marais',
    'temple',
    'républicain',
]

STRATE_B_THEMES = [
    "(via OR vélo)",
    "(ferme OR bloqué OR pignon)",
]

}

```

```

# STRATE A
for theme in STRATE_A_THEMES:
    query = f'({theme}) (ARCHES_PARIS) (EXCLUSIONS_API) since:(since) until:(until) lang:fr'
    requests.append({
        'id': request_id,
        'strat': 'A',
        'mois': mois,
        'query': query,
        'caractères': len(query)
    })
    request_id += 1

# STRATE B (initié pour éviter explosion)
for lieu in STRATE_B_LIEUX:
    # A seulement 5 lieux principaux
    for theme in STRATE_B_THEMES:
        query = f'({theme}) (lieu) (EXCLUSIONS_API) since:(since) until:(until) lang:fr'
        requests.append({
            'id': request_id,
            'strat': 'B',
            'mois': mois,
            'query': query,
            'caractères': len(query)
        })
        request_id += 1

return requests

# =====
# COLLECTE DES TWEETS
# =====

def collecter_toutes_requetes(requests):
    """Exécute toutes les requêtes et sauvegarde les résultats"""
    tous_les_tweets = []

    print(f"~*~*~ COLLECTE COMPLÈTE - ZIL PARIS CENTRE (2023-2026) ~*~*~")
    print(f"~*~*~")
    print(f"~*~*~ Nombre total de requêtes : {len(requests)} ~*~*~")
    print(f"~*~*~ Fichier de sortie : {OUTPUT_FILE} ~*~*~")
    print(f"~*~*~ Durée estimée : {len(requests) * 2 / 60:.1f} minutes ~*~*~")

    # Sauvegarde des requêtes pour récupération
    with open('requetes_x13_2023_2026.json', 'w', encoding='utf-8') as f:
        json.dump(requests, f, ensure_ascii=False, indent=2)
    print(f"~*~*~ Liste des requêtes sauvegardée : requetes_x13_2023_2026.json ~*~*~")

    # Boucle sur chaque requête
    for i, req in enumerate(requests, 1):
        query = req['query']
        request_id = req['id']
        strat = req['strat']
        mois = req['mois']

        print(f"[{i}/{len(requests)}] Strat: {strat} | Mois: {mois} | ID: {request_id}")
        print(f"~*~*~ Requête : {query} ~*~*~")

        tweets_requete = []
        cursor = None
        params = {'query': query}

        # Pagination pour cette requête
        while len(tweets_requete) < MAX_TWEETS_PAR_REQUETE:
            if cursor:
                params['cursor'] = cursor

            try:
                response = requests.get(ENDPOINT, headers=HEADERS, params=params, timeout=10)
            except Exception as e:
                print(f"~*~*~ Erreur réseau : {e} ~*~*~")
                break

            tweets_requete += response.json()
            cursor = response.get('next_cursor')

```

```

# EXCLUSIONS API (uniquement car ils reviennent trop souvent)
EXCLUSIONS_API = [
    'RER - RER - SNCF - RATP - Transilien - TER',
    'Rég - Rég - Chèques - Mairie',
    'drapeau - drapeau - drapeau - drapeau'
]

# =====
# GÉNÉRATION DES PÉRIODES MOISIELLES
# =====

def generer_periodes_mensuelles(date_debut, date_fin):
    """Génère la liste de toutes les périodes mensuelles"""
    periodes = []
    current = date_debut

    while current <= date_fin:
        since = current.strftime('%Y-%m-%d')

        if current.month == 12:
            next_month = current.replace(year=current.year + 1, month=1, day=1)
        else:
            next_month = current.replace(month=current.month + 1, day=1)

        until = next_month.strftime('%Y-%m-%d')

        periodes.append({
            'since': since,
            'until': until,
            'mois': current.strftime('%Y-%m'),
            'nom': current.strftime('%A %d %Y')
        })

        current = next_month

    return periodes

# =====
# GÉNÉRATION DES REQUÊTES
# =====

def generer_toutes_requetes(periodes):
    """Génère toutes les requêtes pour toutes les périodes"""
    requests = []
    request_id = 1

    for periode in periodes:
        since = periode['since']
        until = periode['until']
        mois = periode['mois']

        # STRATE S
        for theme in STRATE_S:
            query = f'({theme}) since:(since) until:(until) lang:fr'
            requests.append({
                'id': request_id,
                'strat': 'S',
                'mois': mois,
                'query': query,
                'caractères': len(query)
            })
            request_id += 1

```

```

if response.status_code != 200:
    print(f"~*~*~ Erreur HTTP (response.status_code) ~*~*~")
    if response.status_code == 401:
        print(f"~*~*~ Vérifiez le ID API et les crédits ~*~*~")
    elif response.status_code == 429:
        print(f"~*~*~ Rate limit atteint, pause 60s... ~*~*~")
        time.sleep(60)
    continue

data = response.json()
tweets = data.get('tweets', [])

# Extraire les données de chaque tweet
for tweet in tweets:
    author = tweet.get('author', {})
    if isinstance(author, str):
        try:
            author = json.loads(author)
        except:
            author = {}

    tweet_data = {
        'id': tweet.get('id'),
        'text': tweet.get('text', ''),
        'created_at': tweet.get('created_at', ''),
        'lang': tweet.get('lang', ''),
        'retweet_count': tweet.get('retweet_count', 0),
        'like_count': tweet.get('like_count', 0),
        'is_retweeted': tweet.get('retweeted', False) and tweet.get('retweeted_tweet') is not None,
        'request_id': request_id,
        'strat': strat,
        'mois': mois,
        'author_name': author.get('name', ''),
        'author_username': author.get('username', ''),
        'author_followers': author.get('followers', 0),
        'author_following': author.get('following', 0),
        'author_created_at': author.get('created_at', ''),
        'author_status_count': author.get('status_count', 0),
        'author_description': author.get('description', '')
    }

    tweets_requete.append(tweet_data)

cursor = data.get('next_cursor')
if not tweets or not cursor:
    break

time.sleep(PAUSE_API)

# Ajoute les tweets de cette requête au total
tous_les_tweets.extend(tweets_requete)
print(f"~*~*~ {len(tweets_requete)} tweets (total: {len(tous_les_tweets)}) ~*~*~")
print(f"~*~*~ Coût cumulé : {len(tous_les_tweets) * 0.00015:.2f} $ ~*~*~")

# Sauvegarde intermédiaire tous les 50 requêtes
if i % 50 == 0 and tous_les_tweets:
    print(f"~*~*~ Sauvegarde intermédiaire... ~*~*~")
    with open(OUTPUT_FILE, 'w', encoding='utf-8', newline='') as f:
        writer = csv.DictWriter(f, fieldnames=tous_les_tweets[0].keys())
        writer.writeheader()
        writer.writerows(tous_les_tweets)
    print(f"~*~*~ {len(tous_les_tweets)} tweets sauvegardés ~*~*~")

# Pause entre les requêtes
time.sleep(PAUSE_API)

```

```

# Sauvegarde finale
if tous_les_tweets:
    with open(OUTPUT_FILE, 'w', encoding='utf-8', newline='') as f:
        writer = csv.DictWriter(f, fieldnames=tous_les_tweets[0].keys())
        writer.writeheader()
        writer.writerows(tous_les_tweets)

    print("-"*70)
    print(" COLLECTE TERMINÉE ")
    print("-"*70)
    print(f"📄 Fichier final : {OUTPUT_FILE}")
    print(f"📊 Tweets collectés : {len(tous_les_tweets)}")
    print(f"📊 Coût total : {len(tous_les_tweets) * 0.00051:.2f} $")
    print(f"📌 Prochaine étape : Lance le Pipeline de Nettoyage v2 sur ce fichier !")
else:
    print(f"🚫 Aucun tweet n'a été collecté.")

return tous_les_tweets

# =====
# LANCEMENT DU SCRIPT
# =====

if __name__ == "__main__":
    print("-"*70)
    print(" GÉNÉRATION DES REQUÊTES ")
    print("-"*70)

    périodes = generer_périodes_annuelles(DATE_DÉBUT, DATE_FIN)
    print(f"📅 Période : {DATE_DÉBUT.strftime('%Y-%m-%d')} à {DATE_FIN.strftime('%Y-%m-%d')}")
    print(f"📅 Nombre de mois : {len(périodes)}\n")

    requetes = generer_toutes_requetes(périodes)

    nb_s = sum(1 for r in requetes if r['strate'] == 's')
    nb_a = sum(1 for r in requetes if r['strate'] == 'a')
    nb_b = sum(1 for r in requetes if r['strate'] == 'b')

    print(f"📊 Requetes générées : {len(requetes)}")
    print(f"📊 [- strate s : {nb_s}]")
    print(f"📊 [- strate a : {nb_a}]")
    print(f"📊 [- strate b : {nb_b}]")

    # Estimation du coût
    tweets_estimes = len(périodes) * 500 # Basé sur les sondes
    cout_estime = tweets_estimes * 0.00051

    print(f"📊 Estimation :")
    print(f"📊 Tweets attendus : {(tweets_estimes):,}")
    print(f"💰 Coût estimé : {(cout_estime:.2f)} $")

    # Confirmation
    reponse = input("Lancer la collecte ? (oui/non) : ")
    if reponse.lower() in ['oui', 'o', 'yes', 'y']:
        print(f"🚀 Démarrage de la collecte...\n")
        tweets = collecter_toutes_requetes(requetes)
    else:
        print(f"🚫 Collecte annulée.")

```

```

"""
Pipeline de Nettoyage Final
"""

import pandas as pd
import json
import re
from datetime import datetime
from collections import Counter

# =====
# CONFIGURATION
# =====

INPUT_FILE = "rsl_paris_2023_2026_complet_20230808T_152049.csv" # À adapter
OUTPUT_FILE = "rsl_paris_2023_2026_complet_20230808T_152049 propre.csv"

SEUIL_REPONSE_COURTE = 100 # caractères

# =====
# DICTIONNAIRES D'EXCLUSION
# =====

EXCLUSIONS_TRANSPORTS = [
    'RER', 'RERB', 'RER A', 'RER C', 'RER D', 'RER E',
    'DASH', 'RER', 'Metro', 'Metro', 'Transilien', 'TGV',
    'intercirculation', 'aéroport', 'Gare', 'Métro', 'Claye', 'Région',
    'Ligne 1', 'Ligne 4', 'Ligne 7', 'Ligne 11', 'Ligne 14'
]

EXCLUSIONS_SPORT = [
    'Ligue champion', 'lig', 'Paris saint germain', 'saint germain',
    'vitesse pug', 'Équipe ligas', 'tour de france', 'rallye garmes'
]

EXCLUSIONS_POLITIQUE_NATIONALE = [
    'Emmanuel macron', 'macron', 'assemblée nationale', 'étatés unis',
    'new york', 'donald trump', 'gouvernement', 'assemblée',
    'ministre', 'affaires étrangères'
]

EXCLUSIONS_JUSTICE = [
    'une prison', 'bracelet électronique', 'agressions sexuelles', 'viols',
    'cour d'assises', 'garde à vue', 'tribunal'
]

EXCLUSIONS_AUTRES_FIL = [
    'ligue', 'jemma', 'pennsila', 'première ligue', 'grenette', 'rue grenette',
    'bellemeur', 'terreux', 'crisis-musée', 'part-dire',
    'grenoble', 'grenoblois', 'libre', 'strasbourg', 'strasbourgeois',
    'nantes', 'nanta', 'Monspellier', 'Monspellien', 'Nordme', 'Nordclais',
    'Marseille', 'Marseilleais', 'Toulon', 'toulonnais', 'Nice', 'niçois',
    'Ames', 'Lille', 'lillais'
]

EXCLUSIONS_SPAM_PROMUE = [
    'drague', 'deal', 'shit', 'shit', 'weed', 'cocaïne', 'cannabis', 'hard', 'horne',
    'livraison russe', 'DM pour commander', 'WhatsApp', 'discret', 'télégram'
]

EXCLUSIONS_SCOMPENSE = [
    'livraison en ligne', 'e-commerce', 'commerce', 'livraison gratuite',
    'code promo', 'réduction', 'soldes', 'black friday', 'amazon', 'aldisexpress'
]

```



```
def check_arrondissement(row):
    excl, notif = nb_hors_ztl_arrondissement(row['text'], EXCLUSIOM_HORS_ZTL_ARRONDISSEMENTS)
    if excl:
        arrondissements_trouves.append(notif)
    return excl

nb_hors_ztl = df.apply(check_arrondissement, axis=1)
nb_hors_ztl = nb_hors_ztl.sum()
print("Nb hors ztl : ", nb_hors_ztl)

if arrondissements_trouves:
    top_arr = Counter(arrondissements_trouves).most_common(10)
    print("Top 10 arrondissements exclis :")
    for arr, count in top_arr:
        print(" - [arr]: (count)")

df = df[nb_hors_ztl]
print("Nb Twets restants : (len(df))")

# Etape 7 : TAGGING THEMATIQUE
print("====")
print(" Etape 7 : TAGGING THEMATIQUE")
print("====")
df['themes'] = df['text'].apply(lambda x: tagger_themes(x, THEMES_ZTL))

themes_counts = df['themes'].value_counts()
print("Repartition des themes :")
for theme, count in themes_counts.head(15).items():
    print(" - (theme): (count)")

nb_autre = df['themes'] == 'autre'.sum()
pct_autre = nb_autre / len(df) * 100 if len(df) > 0 else 0
print("Nb Twets 'autre' : (nb_autre) (pct_autre: %f)" % pct_autre)

# Etape 8 : SAUVEGARDE
print("====")
print(" Etape 8 : SAUVEGARDE")
print("====")
df.to_csv(OUTPUT_FILE, index=False, encoding='utf-8')
print(" Corpus propre sauvegarde : (OUTPUT_FILE)")
print(" Nb Twets finaux : (len(df))")
taux_conservation = len(df)/total_initial*100 if total_initial > 0 else 0
print(" Taux de conservation : (taux_conservation: %f)" % taux_conservation)

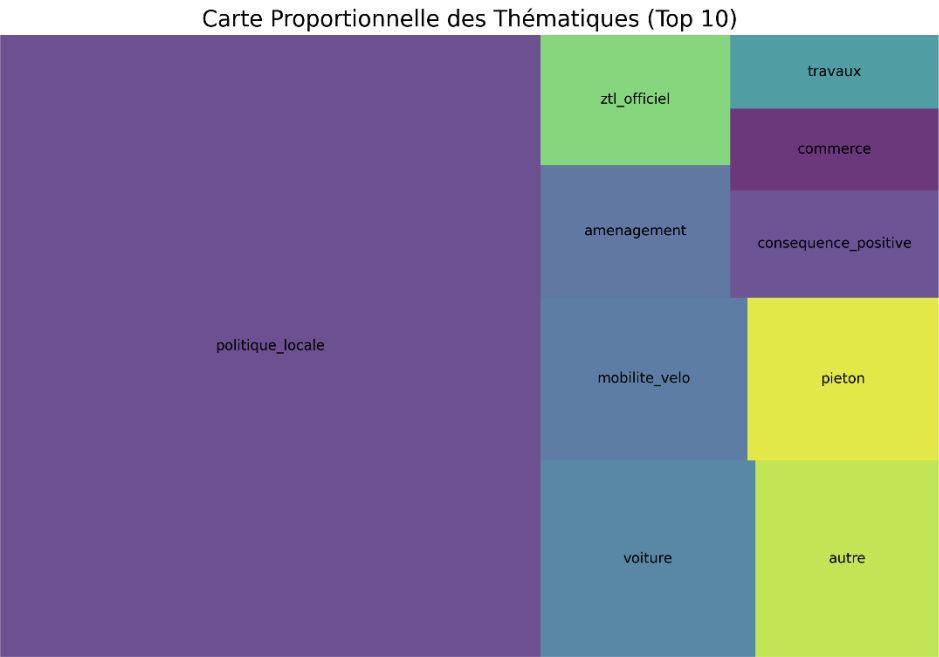
# Etape 9 : COMPTILLEN QUALITEZ
print("====")
print(" COMPTILLEN DE 15 TWETS PROPOSÉS")
print("====")
sample = df.sample(len(df), random_state=42)
for i, (_, row) in enumerate(sample.iterrows(), 1):
    print("Tw[1] Mot: (row.get('bot_score', 'N/A')) | Themes: (row.get('themes', 'N/A'))")
    print(" - (row.get('text', '')) (200) (...)")

# RESUME FINAL
print("====")
print(" RESUME DU PIVILLAGE")
print("====")
print(" Nb Twets initiaux : (total_initial)")
print(" - Après dedup : (nb_dedup)")
print(" - Après M : (rt_count)")
print(" - Après réponses courtes : (nb_reponses_courtes)")
print(" - Après exclusions : (nb_exclay)")
print(" - Après hors ztl : (nb_hors_ztl)")
print(" / CORPUS FINAL : (len(df)) Twets")
print(" / Taux de conservation : (taux_conservation: %f)" % taux_conservation)
print("====")

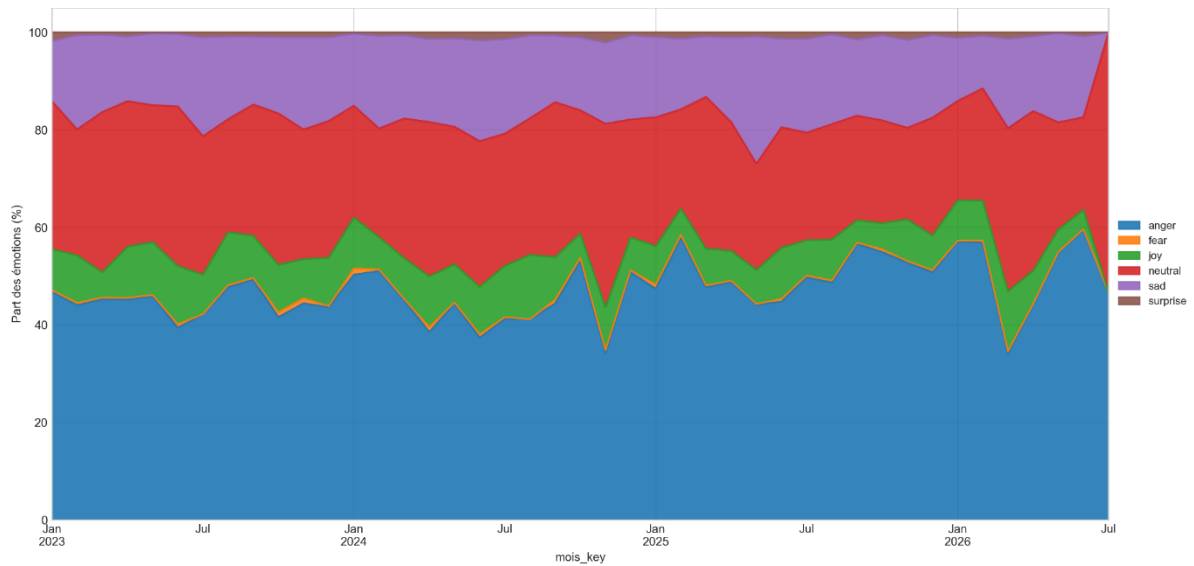
return df

if __name__ == "__main__":
    df_propre = nettoyer_corpus()
```

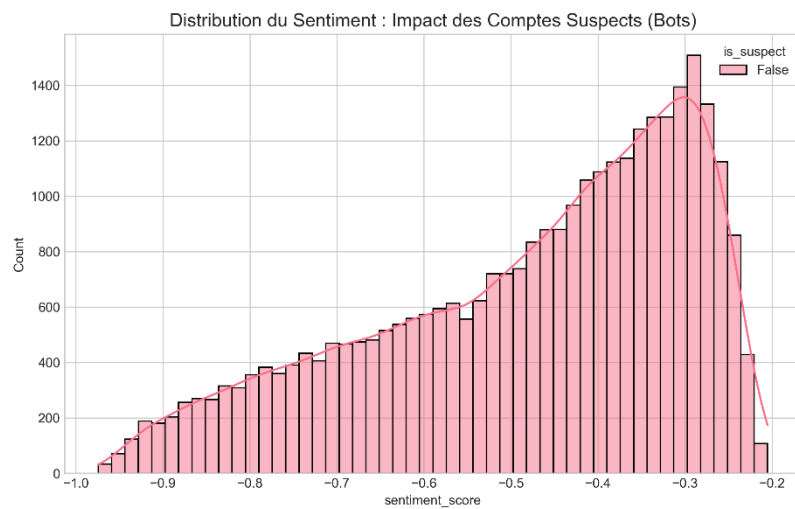
Annexe X : Streamgraph des émotions en parts normalisées Source : GOPALAPILLAI Gokul



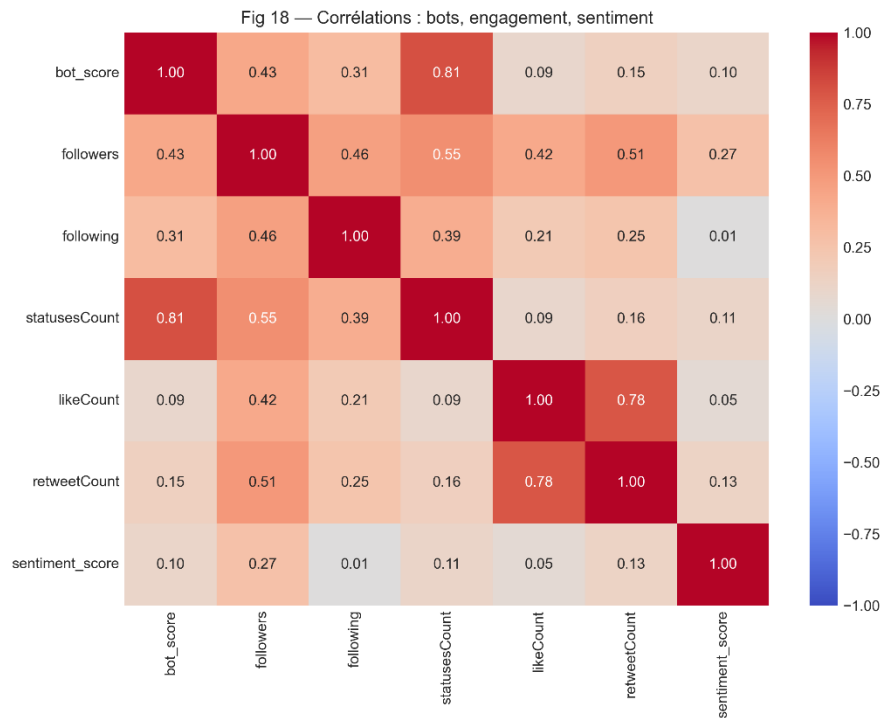
Annexe XI: Régimes émotionnels dans le temps Source : GOPALAPILLAI Gokul



Annexe XII: Distribution du sentiment : Avec vs Sans Comptes Suspects Source : GOPALAPILLAI Gokul



Annexe XIII: Matrice de corrélations de Spearman Source : GOPALAPILLAI Gokul



Annexe XIVV: Heatmap jour x heure Source : GOPALAPILLAI Gokul

