

BMDec, des enquêtes locales et longitudinales sur la mobilité domicile-travail périurbaine

Marius Garénaux-Gruau*, François Bodin†, Laurent Morin‡- IRISA

Table des matières

1	Introduction	2
1.1	Présentation des enquêtes	2
1.2	Objectifs des enquêtes	2
2	Collecte des données	3
2.1	Construction du panel	3
2.2	Collecte des données	4
2.3	Impact sur les panellistes	4
2.4	Participation	5
2.5	Curation	6
2.6	Structure des données	6
3	Autres enquêtes sur la mobilité	7
3.1	Enquête Mobilité Certifiée Cerema	7
3.2	Enquête Mobilité des Personnes (EMP)	7
4	Analyse des données socio-économiques du panel	7
4.1	Genre, âge	8
4.2	Niveau de revenu	8
4.3	Distance domicile-travail	9
5	Modes de transport utilisés par le panel	9
5.1	Multimodalité	9
5.2	Évolution temporelle des proportions des modes de transports	10
6	Modes de transport des panellistes	11
6.1	Mode principal	11
6.2	Représentations des couples de variables et dépendances	13
6.2.1	Emploi de catégorie "cadre" et mode principal	13
6.2.2	Genre et mode principal	14
6.2.3	Aparté sur le test du chi2	14
7	Émissions de gaz à effet de serre	15
7.1	Distances et émissions de CO_2 par trajet	15
7.2	Émissions de CO_2 et distances parcourues totales, par mode	16

*marius.garenaux-gruau@irisa.fr

†Francois.Bodin@irisa.fr

‡laurent.morin@irisa.fr

8	Modélisation du comportement des panellistes	17
8.1	Arbre de décision	18
8.2	Modèle à choix discrets	19
9	Diffusion des données	20
9.1	Diffusion académique	20
9.2	Modèles	21
10	Conclusion	21
11	Bibliographie	21

Les graphiques de ce document ont été produits avec l'interface Observable Framework (1)

1 Introduction

Ce document présente les enquêtes **BMDec (Breizh Mobilité Décarbonnée)** réalisées à Rennes et Brest en 2023 et 2024 respectivement.

1.1 Présentation des enquêtes

Deux enquêtes ont été menées dans le cadre du projet BMDec¹. Ces enquêtes, réalisées à un an d'intervalle à Rennes et Brest (en 2023 et 2024), ont pour objectif d'améliorer la connaissance des mécanismes déterminant les comportements de mobilité périurbaine.

Pour chacune des deux enquêtes, un panel d'environ 300 personnes a été constitué à partir de **volontaires**. Les enquêtes se sont déroulées sur **3 mois**. Chaque jour, les participants et participantes remplissent un formulaire sur les modes de transports utilisés pour leurs trajets domicile-travail (aller et retour).

Les enquêtes BMDec présentent l'**évolution sur plusieurs mois** des comportements de mobilité domicile-travail **périurbaines** de deux panels.

Le projet BMDec est porté par un consortium de quatre partenaires : le cabinet SETUR², RatpDev³ et deux laboratoires de recherche de l'Université Rennes 1 : le CREM⁴ et l'IRISA⁵. À l'IRISA, la collecte des données a été gérée au sein de l'équipe de recherche **Logica**; par François Bodin et Laurent Morin. L'étude sur les mécanismes déterminant les comportements de mobilité urbaines a été menée au CREM par Laurent Denant-Boemont et Sabrina Hammiche (2).

1.2 Objectifs des enquêtes

Les enquêtes BMDec présentent plusieurs objectifs.

Premièrement, le **CREM** a mené une étude sur "l'efficacité des incitations non monétaires dans la diminution de la part modale de la voiture particulière dans les déplacements domicile-travail" (2). Concrètement, il s'agit de déterminer si la présence d'informations additionnelles (par exemple les émissions moyennes de CO_2) peut influencer le choix de mode de transport d'une personne.

Ensuite, les enquêtes ont permis de récolter des données de mobilités **longitudinales** : des informations quotidiennes sur la mobilité d'un groupe de personnes. Ce type d'enquête est plutôt rare; les enquêtes habituelles de mobilités étant plutôt instantanées (on mesure le comportement de mobilité d'une population

1. BMDec : Breizh Mobilité Décarbonnée, <https://project.inria.fr/bmdec/fr/>

2. SETUR : <https://setur.eu/>

3. RATP Dev : <https://www.ratpdev.com/fr/references/france-brest-tramway>

4. CREM : Centre de Recherche en Economie & Management, <https://crem.univ-rennes.fr/>

5. IRISA : Institut de Recherche en Informatique et Systèmes Aléatoires <https://www.irisa.fr/>

à un instant donné). Deux exemples d'enquêtes françaises *instantanées* sont : les **Enquêtes Mobilité des Personnes** et les **Enquêtes Ménages Déplacement**.

Ces enquêtes *-instantanées-* présentent comme défaut de ne pas bien capturer une évolution dans le temps du comportement des usagers et usagères (3) (1.4.3 *Cross-section and Time Series*). Les enquêtes longitudinales sont plus appropriées pour capturer ces évolutions. Cependant, la collecte de données temporelles vient avec des difficultés techniques :

- mise en place d'une plateforme de collecte de **données personnelles soumises au RGPD**⁶,
- **maintenance** sur plusieurs mois,
- **rémunération** des panellistes,
- **calcul automatique des distances** pour chaque mode et chaque trajet - dans notre cas avec Navita (4) et OSM (5),
- appel quotidien à l'**API Impact CO₂** de l'ADEME pour estimer les émissions des trajets (6).

Le projet BMDec a également pour objectif d'explorer ces difficultés et d'y apporter des solutions. De plus, certaines *meta*-informations comme la durée d'enquête ou le taux de participation sont intéressantes pour guider de futures enquêtes du même type.

Enfin, l'analyse des données ainsi que la production de modèles synthétisant ces dernières favorisent une future **réutilisation des données**, tout en respectant la réglementation (**RGPD**).

2 Collecte des données

Un des points centraux des enquêtes BMDec est la mise en place d'une plateforme permettant de recueillir **quotidiennement** des données de mobilités. On décrit ici comment le panel a été sélectionné, ainsi que la collecte des données.

2.1 Construction du panel

Les individus participant au panel sont **volontaires** et font partie de la **population active** (l'objectif est d'étudier les mobilités domicile-travail). Seule exception est faite aux étudiants et étudiantes, qui ont l'autorisation de participer. Une méthode de **sondage stratifié** a été mise en place pour choisir les panellistes parmi les volontaires. Au final, on trouve environ 300 panellistes dans chaque enquête.

La méthode de sondage stratifié consiste à choisir les individus possiblement **non-uniformément** dans la population. Plus précisément, on sélectionne en amont certaines catégories de la population échantillonnée (les "strates"; par exemple : femmes / hommes, moins de 35 ans / plus de 35 ans, ...), et on choisit une proportion de l'échantillon selon des quotas préalablement fixés dans chacune des catégories (par exemple on choisit 25 % de femmes). Le choix des strates varie selon l'expérience que l'on mène. Voir le livre "Manuel de sondages" (7) pour plus de détails.

Dans les enquêtes BMDec, les strates ont été choisies selon les critères suivants :

- l'âge : deux catégories sont possibles, les 30-60 ans et les autres (moins de 30 ou plus de 60 ans),
- le genre : homme ou femme,
- la distance domicile-travail : courte (moins de 5 km), longue (strictement plus de 5 km).

Les strates sont les 8 (2³) choix possibles parmi ces 3 catégories. Par exemple, une strate est : "homme de 30-60 ans vivant à moins de 5km de son lieu de travail". Les proportions pour chaque strate ont été déterminées à l'aide d'enquêtes locales de mobilités déjà existantes. On se référera à l'article du CREM (2) pour plus d'informations sur la procédure de sélection des panellistes.

6. RGPD : Règlement Général sur la Protection des Données : <https://www.cnil.fr/fr/comprendre-le-rgpd>.

2.2 Collecte des données

Après avoir rempli un formulaire contenant des données socio-économiques et de mobilités, l'expérimentation commence. Les données sont collectées quotidiennement pendant 3 mois. Chaque jour, chacune des personnes participantes reçoit un SMS lui rappelant de remplir en ligne le formulaire. Pour les besoins de l'enquête menée par le CREM, certains des formulaires sont *nudgés*.

Le *nudge* ("coup de coude") est une information supplémentaire donnée à certains panellistes, par exemple l'émission de CO_2 de leur dernier trajet. L'objectif de l'enquête menée par le CREM est de mesurer l'impact de ces *nudges* sur les choix des panellistes.

Les informations additionnelles ont parfois été très mal reçues par les panellistes. On présente rapidement ce sujet dans la suite (2.3)

Le choix des formulaires *nudgés* est fait selon les besoins de l'enquête du CREM afin de garantir des résultats statistiques cohérents. Seule une partie des panellistes est *nudgée* ; et les panellistes avec *nudge* ne le sont pas sur tous leurs formulaires (l'enquête est divisée en 3 parties, une première sans *nudge*, une seconde avec des *nudges* et une dernière sans). Les participantes et participants ne sont pas au courant de la procédure d'enquête.

Deux types de nudges peuvent être présentés aux panellistes :

- le *nudge 1* - appelé "*nudge* par défaut" - consiste, lors du remplissage du mode de transport utilisé, à masquer les modes de transport carbonés (voiture, bus, ...). L'utilisateur ou l'utilisatrice doit alors effectuer un "clic" supplémentaire pour afficher ces modes.
- le *nudge 2* - appelé "*nudge CO_2* " - consiste à afficher des informations supplémentaires. Plus spécifiquement, sont affichées :
 - Une estimation de l'émission de CO_2 du dernier trajet de la personne. Le calcul de l'émission de CO_2 est fait avec l'API Impact CO_2 de l'ADEME (6).
 - Une image montrant un ou plusieurs smileys. Ces smileys peuvent être "contents" ou "pas contents". Le nombre de smileys et leur état (content/ pas content) est calculé en fonction de l'émission de CO_2 du dernier trajet de l'utilisateur ou utilisatrice, par rapport à la moyenne des autres panellistes.
 - Une estimation de l'émission de CO_2 moyenne des autres panellistes ayant utilisé le même mode de transport sur leur dernier trajet.

Afin de calculer les distances pour chaque mode, **d'autres données externes viennent quotidiennement enrichir le jeu de données**, à l'aide de Navitia (4) et d'Open Street Map (5). Si des trajets alternatifs sont possibles, les distances et temps pour chaque mode sont également rajoutés aux données.

Le jeu de données brut sortant d'une enquête BMDec est donc constitué des données socio-économiques des panellistes, ainsi que des formulaires de mobilités quotidiennes recueillis sur 3 mois, **enrichis de données externes** d'émission de CO_2 , de distances et de trajets alternatifs. De plus, chaque participant et participante a la possibilité d'ajouter un commentaire pour chaque trajet.

2.3 Impact sur les panellistes

Nous tenons à souligner l'impact qu'a une telle enquête sur les panellistes. Nous avons pu observer, à travers de nombreux échanges de mails et les commentaires que les enquêtés sont très sensibles à leurs émissions de CO_2 . La population échantillonnée est probablement d'autant plus sensible à ces questions qu'elle est constituée de volontaires, qui pour la plupart font déjà des efforts pour limiter leurs émissions quotidiennes (transport en commun, covoiturage, ...).

Pour les besoins de l'enquête (*nudges*), nous avons communiqué à certains de ces panellistes nos estimations de leurs émissions de CO_2 . Du fait de la quantité importante de cas particuliers à gérer (trajets multimodaux, plusieurs véhicules différents pour le même trajet, trajets inhabituels, ...); il est difficile d'estimer quotidiennement, de manière automatique, les émissions de CO_2 d'une personne. Cela a donné lieu

à des erreurs (sous- et surestimations) - corrigées après le sondage. Dans le cas de surestimations, les panellistes nous ont très régulièrement corrigé.

Ces erreurs combinées à l'apparition de *smileys* (mécontent) ; ou de comparaisons par rapport à la moyenne des autres usagères et usagers a parfois déclenché des réactions importantes auprès des panellistes (du commentaire mécontent à l'abandon de l'enquête). Ces impacts doivent être anticipés avec sérieux dans de futures enquêtes du même type ; afin déjà de ne pas miner le moral des panellistes qui font des efforts (là où l'objectif de cette enquête est aussi de sensibiliser aux impacts environnementaux), mais aussi pour que l'enquête se déroule le mieux possible.

2.4 Participation

Les participants et participantes reçoivent une **rémunération**, en fonction de leur assiduité. Un des objectifs de la rémunération est d'améliorer les taux de participation.

Répartition des pourcentages de participation au sein des panels

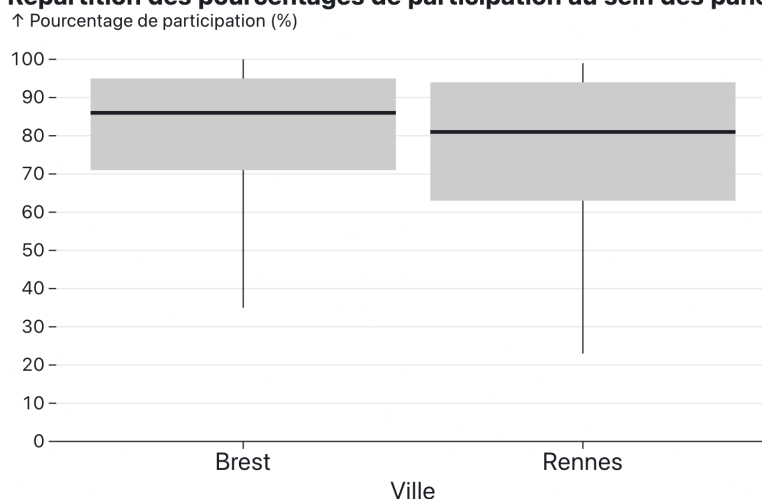


Figure 1 – Taux de participation

Dans les boîtes à moustaches ci-dessus, on a représenté la répartition des pourcentages de participation aux seins des deux panels (Rennes et Brest). Par exemple, on y lit que plus de la moitié des panellistes de Brest ont un taux de participation supérieur à 80 %.

La fonction utilisée pour la rémunération est la suivante. Pour un pourcentage de participation $x \in [0, 100]$,

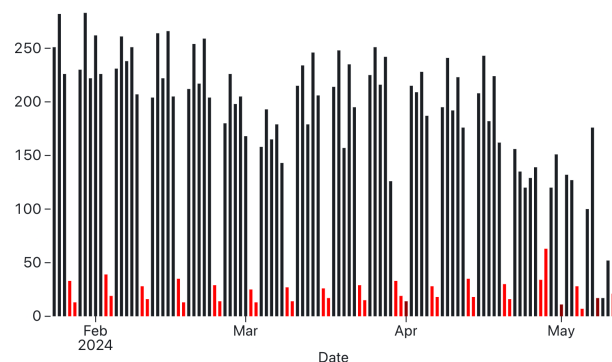
la rémunération est calculée comme suit :

$$f(x) = \begin{cases} 0 & x \leq 20 \\ 0.25x + 25 & 20 < x < 60 \\ 70 & 60 \leq x \end{cases}$$

Enfin, l'objectif de l'enquête étant d'analyser les trajets domicile-travail, ne sont gardés pour l'analyse que des trajets de ce type. Ci-dessous est représenté le nombre de trajets domicile-travail récoltés par jour (un aller-retour ne compte que pour un trajet).

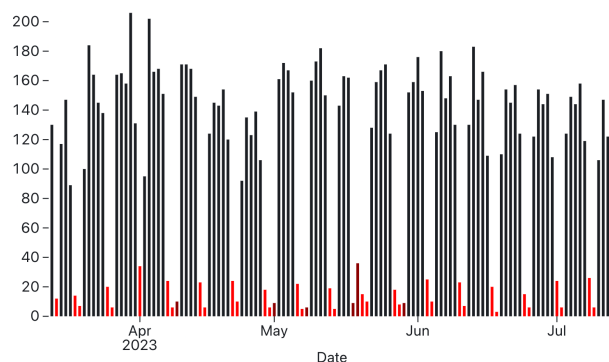
Nombre de trajets domicile-travail collectés, par jour - BMDec Brest

↑ Nombre de réponses aux formulaires



Nombre de trajets domicile-travail collectés, par jour - BMDec Rennes

↑ Nombre de réponses aux formulaires



On constate une nette baisse du nombre de formulaires reçus les week-ends et jours fériés (respectivement en rouges et rouges foncés) ; car sans surprise ce sont pour la plupart des panellistes des jours non travaillés. Certaines statistiques présentées dans la suite ne prennent pas en compte les week-ends et les jours fériés. Lorsque c'est le cas, la mention suivante est indiquée : *“Les données des jours fériés et des week-ends ont été retirées”*.

On constate également une nette baisse de participation en fin de l'enquête de Brest. Ceci s'explique par la période de vacances scolaires (20 avril au 6 mai) ; succédée par le pont (8-9-10 mai). On peut également supposer une lassitude liée à la fin du sondage.

2.5 Curation

À la suite des enquêtes, les données recueillies ont été nettoyées manuellement. Pour cela, un logiciel utilisant l'interface Streamlit (8) a été développé. Cette curation vise à produire un jeu de données plus robuste, en recalculant les distances pour chaque mode et chaque trajet ; ainsi que les émissions de CO_2 . On prend également en compte les différents commentaires des usagers et usagères pour ajuster ces informations.

Enfin, un score de qualité de donnée a été appliqué à chaque panelliste et chaque trajet. Ce score note la qualité de la donnée observée. Il est basé sur la procédure de Sea Data Net de contrôle de la qualité des données (9).

La procédure de curation des données est fastidieuse - autour d'une dizaine de minutes par utilisateur - mais nécessaire pour obtenir un jeu de données plus robuste. Elle permet également d'avoir une vue d'ensemble sur les trajets effectués par les panellistes : lignes de trains / bus / tramway utilisées, trajets “types”, ... Le caractère manuel permet également de mieux détecter les cas rares (trajets avec 3 ou 4 modes de transports), et de les intégrer dans les données.

2.6 Structure des données

Les données collectées dans le cadre des enquêtes BMDec sont des données de panel. C'est-à-dire qu'une série de données est enregistrée pour chaque panelliste (ses trajets quotidiens). Comme ces données évoluent au cours du temps, on parle plus précisément de **séries temporelles** (ou de **données longitudinales**). Les données présentent donc deux dimensions principales :

- une dimension **temporelle** : le temps du sondage,
- une dimension **usager** : les caractéristiques des panellistes.

Ce format d'enquête a l'avantage de permettre la mise en place d'expériences mesurant la réaction des panellistes face à une information (dans notre cas, les *nudges*).

Il est important de garder à l'esprit qu'une partie des panellistes a subi une influence (*via* les *nudges*). Les données peuvent en être affectées.

3 Autres enquêtes sur la mobilité

Il existe déjà de nombreuses enquêtes sur les mobilités. Cependant, les enquêtes portant sur **l'évolution temporelle** des comportements de mobilités sont plus rares.

En se restreignant aux enquêtes françaises, on trouve principalement ces deux séries d'enquêtes :

- les Enquêtes Mobilité Certifiées Cerema (anciennement **Enquêtes Ménages Déplacements**).
- les Enquêtes Mobilité des Personnes (anciennement **Enquêtes Nationales Transports et Déplacements**).

Ces deux séries d'enquêtes diffèrent par leur fréquence de réalisation, mais surtout par leur échelle. La première apporte des informations **locales**, et la seconde **nationale**. Chacune garantit un **échantillon représentatif** de la population observée. De plus, elles font état du comportement de mobilité d'une population à un instant fixé. La principale différence avec les enquêtes BMDec est la **durée de l'enquête**. On présente un peu plus en détail ces enquêtes dans la suite (3.1, 3.2).

Certaines enquêtes étrangères ont déjà eu pour objectif de recueillir des **données temporelles**. Citons par exemple l'enquête **Mobidrive 1999** ((10), (11)). Menée en Allemagne dans les villes de Karlsruhe et Halle pendant 6 semaines, cette enquête avait pour objectif de recueillir et d'analyser des données sur les **habitudes** des panellistes. Elle est très similaire aux enquêtes BMDec.

3.1 Enquête Mobilité Certifiée Cerema

Anciennement Enquêtes Ménages Déplacement, ces enquêtes sont menées tous les ans dans des territoires différents. Ce sont "des outils de connaissance de la mobilité quotidienne d'une population habitant dans un périmètre". Elles sont menées à une **échelle locale** (centrées sur une ville), et font état de la mobilité **à un instant donné** des habitants et habitantes d'un territoire. Ces enquêtes traitent tous les déplacements, et non uniquement les déplacements domicile-travail.

Depuis 1976, près de 200 enquêtes ont été réalisées sur plus de 100 territoires (http://www.progedo-adisp.fr/serie_emd.php). On pourra consulter les synthèses des enquêtes de Rennes (12) et Brest (13) en 2018 pour avoir des quantités de référence.

3.2 Enquête Mobilité des Personnes (EMP)

Ces enquêtes sont menées tous les 10 ans depuis 2007-2008, à l'**échelle nationale** (on recense à ce jour deux enquêtes, celle de 2008 et celle de 2019). Elles ont pour but d'apporter des éclairages sur le parc de véhicules à disposition des ménages, et les mobilités "locales" et "longue distances" des Français. Plusieurs articles ont été publiés autour de ces enquêtes, synthétisant des résultats sur des thèmes différents (trajets domicile-travail (14), émissions de CO_2 , ...). On pourra consulter le **site suivant** pour accéder à différentes ressources.

4 Analyse des données socio-économiques du panel

Comme vu précédemment, le choix des panellistes s'est fait sur la base du volontariat; en respectant des quotas dans des strates (âge, genre, distance domicile-travail).

On présente ici quelques analyses classiques sur les données socio-économiques des panellistes.

4.1 Genre, âge

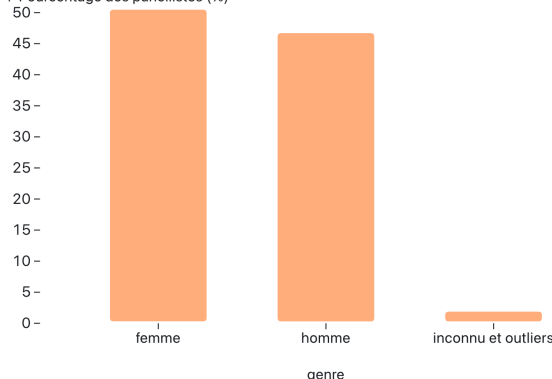
Genre des panellistes - BMDec Brest

↑ Pourcentage des panellistes (%)



Genre des panellistes - BMDec Rennes

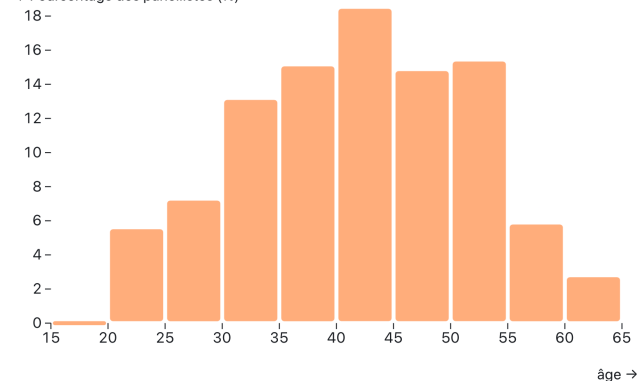
↑ Pourcentage des panellistes (%)



Les genres sont globalement équilibrés, malgré une légère sur-représentation des femmes dans le panel de Brest. En revanche, la distribution de l'âge des panellistes est nettement différente entre les deux panels. Le panel de Rennes contient une part très importante de personnes ayant entre 18 et 22 ans (plus de 20%).

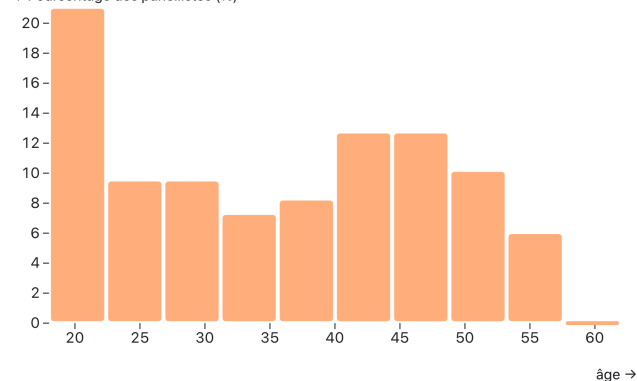
Répartition des âges - BMDec Brest

↑ Pourcentage des panellistes (%)



Répartition des âges - BMDec Rennes

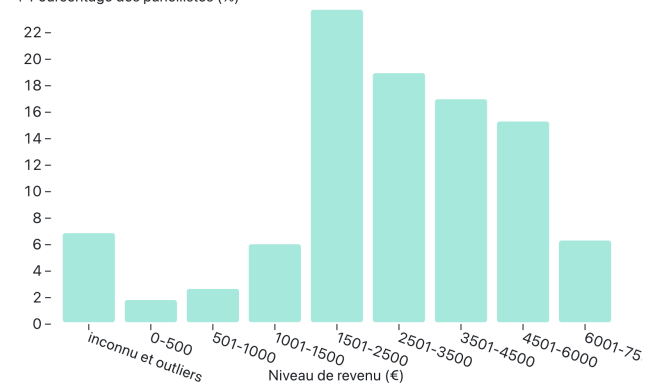
↑ Pourcentage des panellistes (%)



4.2 Niveau de revenu

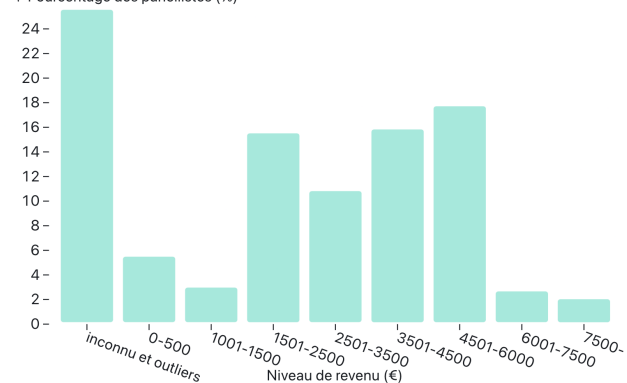
Répartition du niveau de revenu des panellistes - BMDec Brest

↑ Pourcentage des panellistes (%)



Répartition du niveau de revenu des panellistes - BMDec Rennes

↑ Pourcentage des panellistes (%)

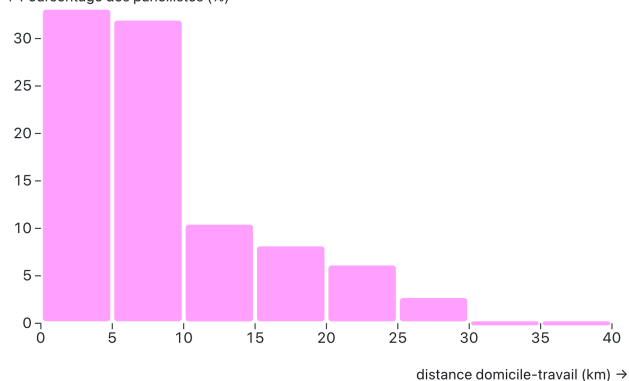


Les données sur le niveau de revenu des panellistes ont beaucoup moins été renseignées à Rennes qu'à Brest, avec une part non négligeable de valeurs inconnues (~25 %). Il est donc difficile de faire des conclusions sur les revenus du panel de Rennes.

4.3 Distance domicile-travail

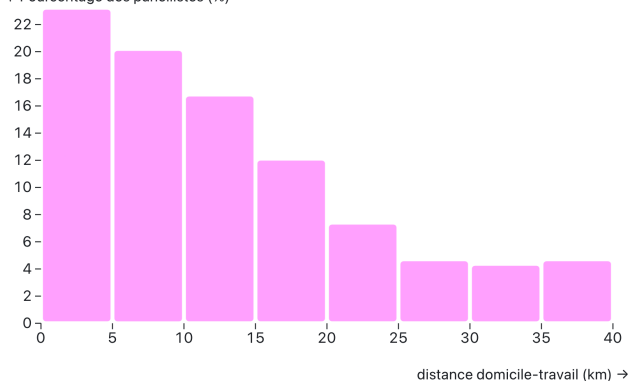
Répartition des distances domicile-travail - BMDec Brest

↑ Pourcentage des panellistes (%)



Répartition des distances domicile-travail - BMDec Rennes

↑ Pourcentage des panellistes (%)



Les membres du panel de Brest habitent plus près de leur lieu de travail que celles et ceux de Rennes : 65% des panellistes brestois habitent à moins de 10km de leur lieu de travail, contre 43% des panellistes rennais.

5 Modes de transport utilisés par le panel

Cette partie présente les données de mobilité du **panel** : la proportion de trajets multimodaux ainsi que l'évolution au cours du sondage des modes choisis par le panel. **Les statistiques présentées ici sont plus influencées par les données des panellistes ayant participé plus fréquemment au sondage.**

5.1 Multimodalité

Dans un premier temps, on traite les trajets **multimodaux** (c'est-à-dire des trajets utilisant plus d'un mode de transport, par exemple vélo + voiture). La plupart des trajets recensés n'ont qu'un seul mode déclaré : on affiche ci-dessous les proportions du nombre de trajets utilisant 1, 2 ou 3 modes de transports différents, parmi tous les trajets. Les modes proposés aux panellistes sont présentés dans la suite.

Répartition du nombre de modes de transports utilisés par trajet - BMDec

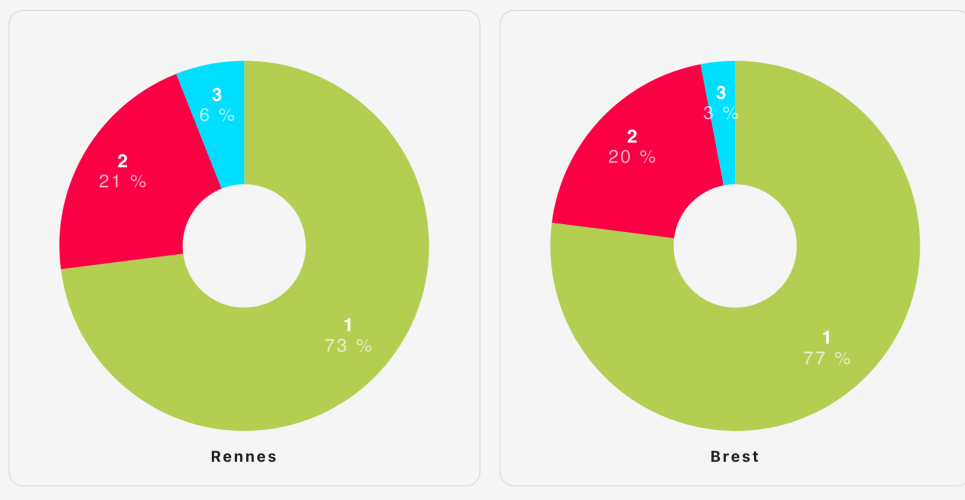


Figure 2 – Trajets multimodaux - Nombre de modes utilisés par trajet

Les trajets multimodaux représentent une part significative des trajets. Afin d'étudier les modes de transports utilisés, nous avons étudié plusieurs stratégies pour déterminer un mode principal :

- garder le mode de transport **le plus lourd** (c'est le choix qui a été fait dans les données de l'EMP 2019),
- garder le mode de transport utilisé sur **la plus longue distance**, et éventuellement regrouper les modes en **catégories** juste avant. Notons que cette stratégie est possible grâce aux données supplémentaires extraites de Navitia (4) et OSM (5).

La stratégie gardée pour ce document est celle qui consiste à garder le mode utilisé sur la plus longue distance.

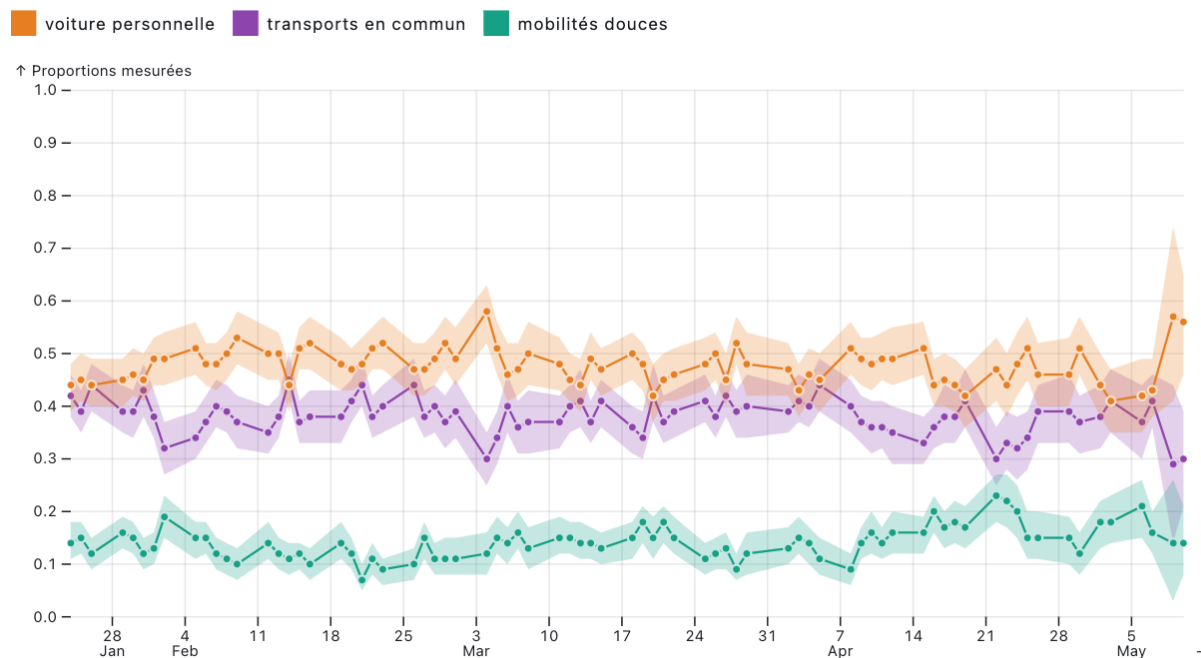
5.2 Évolution temporelle des proportions des modes de transports

Dans cette partie, on étudie les proportions des modes de transports utilisés par le panel. On explore la dimension **temporelle** des données.

On présente dans les graphiques ci-dessous l'évolution de ces proportions.

*Un point sur le graphique représente la part du panel ayant utilisé un mode ce jour-ci. Les intervalles de confiance à 95 % sont affichés autour des lignes.*⁷

Evolution des proportions des modes de transport au sein du panel | BMDec - Brest



Les données des jours fériés et des week-ends ont été retirées; ainsi que les données des modes avec trop peu d'utilisations

Figure 3 – Évolution temporelle des modes de transport - Brest

Les panellistes se voient proposer les modes suivants : autopartage, covoiturage, auto (voiture personnelle), deux ou trois roues thermiques, taxi, bus urbains, métro, train, tramway, téléphérique, autocar, vélo, vélo en libre-service, vélo électrique, engin de déplacement personnel, à pied uniquement.

7. Les intervalles de confiances sont obtenus en supposant que les proportions calculées quotidiennement sont des réalisations de loi Normales. Cette hypothèse est habituellement supposée vraie lorsque la taille de l'échantillon n est supérieure à 30 ; et que la probabilité de succès p vérifie : $\min(np, n(1-p)) > 5$. Sous cette hypothèse, le niveau 95% signifie qu'il y a 5% de risque que la proportion théorique soit en dehors de cet intervalle. On peut citer la vidéo (15) pour une explication des intervalles de confiance.

Evolution des proportions des modes de transport au sein du panel | BMDec - Rennes

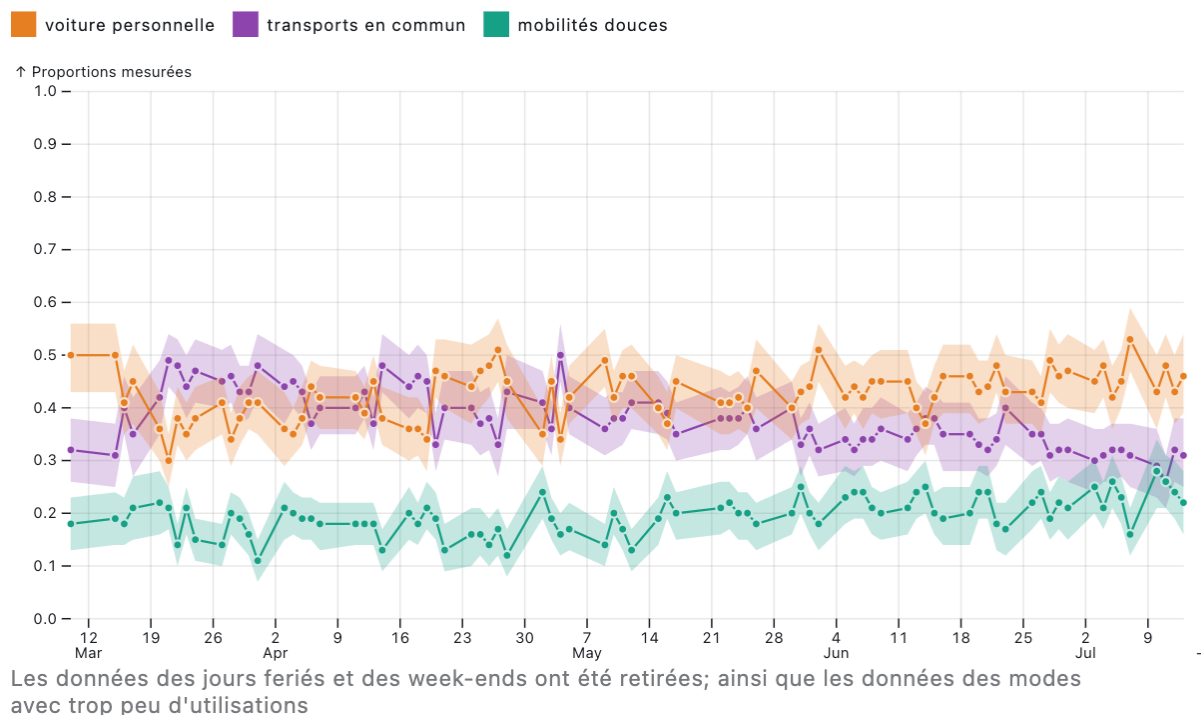


Figure 4 – Évolution temporelle des modes de transport - Rennes

Après avoir calculé chaque distance pour les modes déclarés sur un trajet, on regroupe les modes en catégories :

- véhicule personnel : autopartage, covoiturage, auto (voiture personnelle), deux ou trois roues thermiques, taxi,
- transport en commun : bus urbains, métro, train, tramway, téléphérique, autocar,
- mobilités douces : vélo, vélo en libre-service, vélo électrique, engin de déplacement personnel, à pied uniquement.

Lorsque les trajets sont multimodaux (après le regroupement), on choisit le mode principal en prenant celui utilisé sur la plus longue distance.

On constate au premier abord sur ces graphiques que les proportions ne sont pas constantes au cours du temps. Par exemple, la proportion de panellistes utilisant le mode “transport en commun” varie de 30 à 50%. De plus, les proportions mesurées restent éloignées des résultats des autres enquêtes similaires (voir 6.1).

6 Modes de transport des panellistes

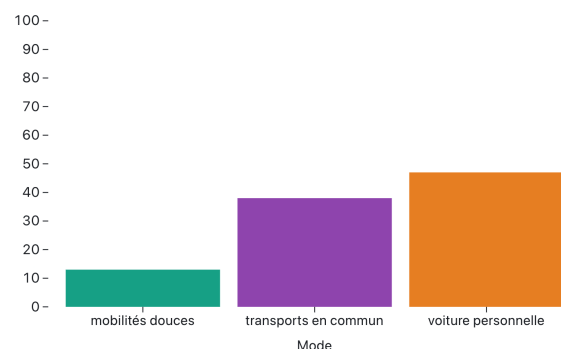
Dans cette partie, on se concentre sur les choix des panellistes. Le but est de représenter les choix de chacun et chacune des panellistes. Chaque personne est représentée à égalité avec les autres, contrairement à la partie précédente.

6.1 Mode principal

On représente ici le mode principal de chaque panelliste. Il est calculé de la manière suivante : pour chaque trajet, on ne garde que le mode utilisé sur la plus longue distance. Sur tous les trajets de l'utilisateur ou utilisatrice, on regarde ensuite quel mode a été utilisé le plus fréquemment.

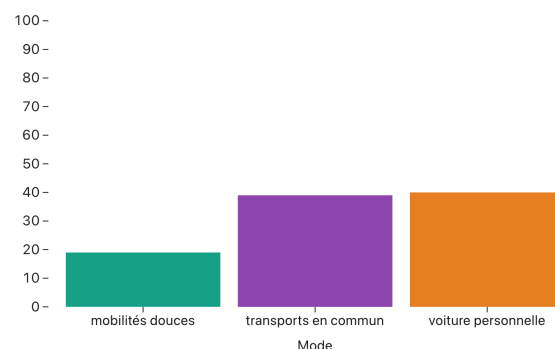
Proportions d'utilisation des modes de transport | BMDec - Brest

↑ Pourcentage



Proportions d'utilisation des modes de transport | BMDec - Rennes

↑ Pourcentage



On observe que le mode “véhicule personnel” reste globalement peu utilisé (autour de 50% des panellistes), par rapport aux proportions habituelles sur ce genre d’enquête.

En effet, si on compare avec l’enquête Ménage Déplacement de Rennes (2018) (12) (Figure 5) ; on observe que la part de la voiture dans les déplacements domicile-travail est autour de 70%.

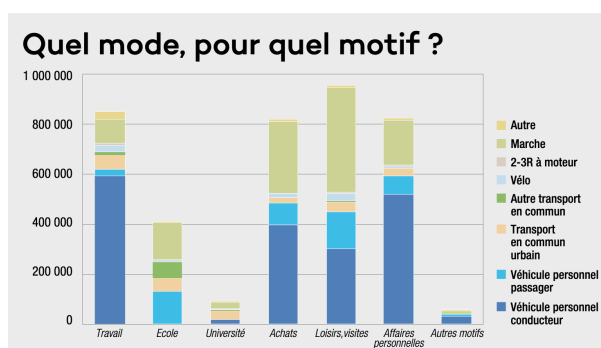


Figure 5 – Graphique issu de la synthèse de l'EMD Rennes 2018 (12)

Pour Brest, la synthèse de l'EMD Brest (2018) ne présente pas de proportion pour les trajets domicile-travail uniquement, mais il y a tout de même une nette différence sur la proportion du mode “voiture” (Figure 6).

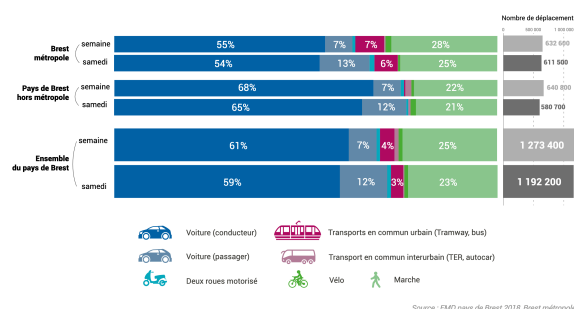
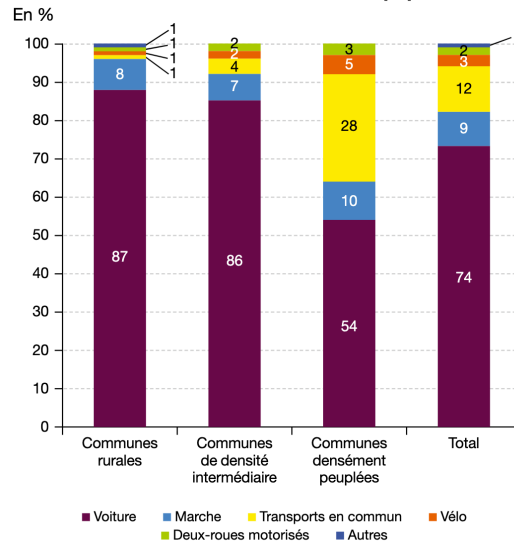


Figure 6 – Graphique issu de la synthèse de l'EMD Brest 2018 (13)

Enfin, l'Enquête Mobilité des Personnes fait plutôt état d'une proportion de voiture de l'ordre de 75% sur les trajets domicile-travail (14) (Figure 7).

Graphique 1 : modes de transport utilisés pour les trajets vers ou depuis un lieu de travail habituel en fonction de la densité communale de population



*Note de lecture : 74 % de l'ensemble des trajets vers ou depuis un lieu de travail habituel sont réalisés en voiture.
 Champ : France métropolitaine, individus de 15 ans et plus en emploi, trajets en mobilité locale semaine entière.
 Source : SDES-Insee, enquête mobilité des personnes 2018-2019*

Figure 7 – Graphique issu d'une synthèse de l'EMP, sur les déplacements domicile-travail (14)

6.2 Représentations des couples de variables et dépendances

On présente ici différentes relations entre le mode de transport principal des usagers et usagères, et certaines caractéristiques (possession d'une carte de transport en commun, genre, ...). Pour cela on affiche des tableaux dans lesquels chaque valeur correspond au nombre de panellistes appartenant à une catégorie.

Ici, le mode principal d'un utilisateur ou d'une utilisatrice est celui utilisé sur une majorité de trajets.

6.2.1 Emploi de catégorie "cadre" et mode principal

Sur le panel de Brest, les panellistes "cadres" ont plus tendance à prendre la voiture face aux transports en commun. À Rennes, en revanche, ce phénomène n'est pas observé.

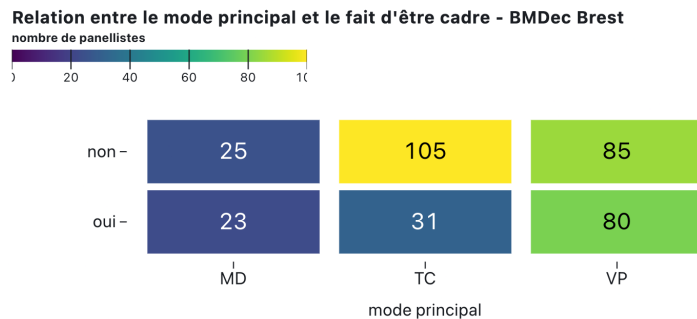
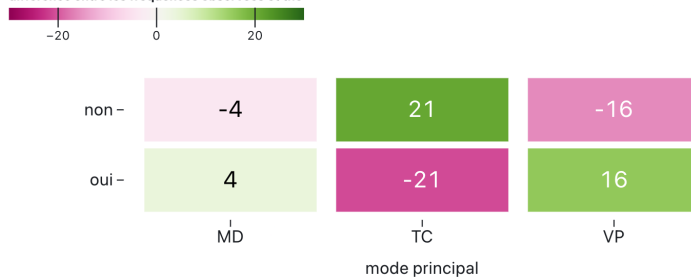


Figure 8 – Mode principal et emploi de catégorie "cadre" - Brest

Comparaison entre le mode principal et le fait d'être cadre - Différence avec les fréquences attendues, p-valeur : 0.0000105 - BMDec Brest



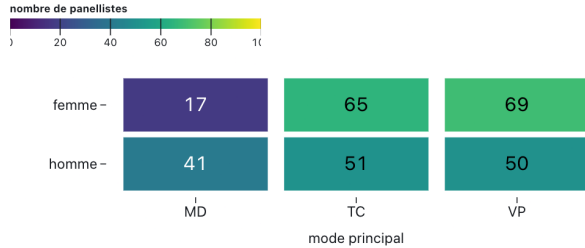
On affiche ici la différence case par case entre les valeurs mesurées et celles que l'on aurait eu si le mode principal et le fait d'être cadre étaient indépendants. Une valeur **positive** signifie que les observations sont **plus élevées** que les valeurs théoriques (sous l'hypothèse d'indépendance), et une valeur **négative** signifie que les observations sont **en dessous** des valeurs théoriques (là encore sous l'hypothèse d'indépendance). Voir l'encart ci-dessous pour plus d'explications sur le test.

Figure 9 – Mode principal et emploi de catégorie “cadre” - Différences de fréquences - Brest

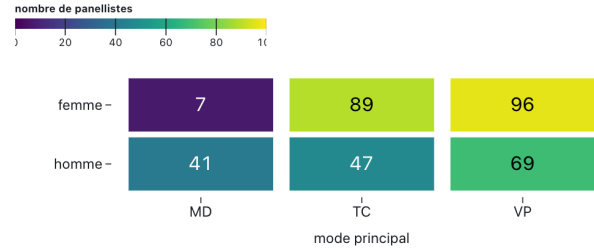
6.2.2 Genre et mode principal

On observe également que les femmes utilisent moins les mobilités douces que les hommes. Ce phénomène est plus marqué à Brest qu'à Rennes.

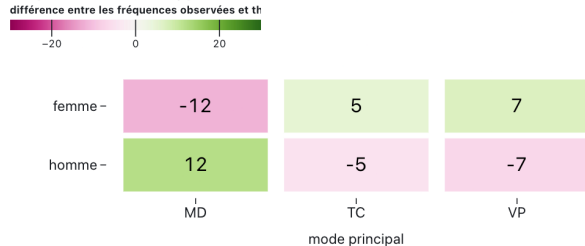
Relation entre le mode principal et le genre - BMDec Rennes



Relation entre le mode principal et le genre - BMDec Brest

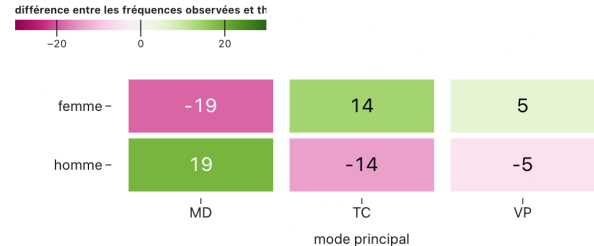


Comparaison entre le mode principal et le genre - Différence avec les fréquences attendues, p-valeur : 0.00075 - BMDec Rennes



On affiche ici la différence case par case entre les valeurs mesurées et celles que l'on aurait eu si le mode principal et le genre étaient indépendants. Une valeur **positive** signifie que les observations sont **plus élevées** que les valeurs théoriques (sous l'hypothèse d'indépendance), et une valeur **négative** signifie que les observations sont **en dessous** des valeurs théoriques (là encore sous l'hypothèse d'indépendance). Voir l'encart ci-dessous pour plus d'explications sur le test.

Comparaison entre le mode principal et le genre - Différence avec les fréquences attendues, p-valeur : 4.71e-9 - BMDec Brest



On affiche ici la différence case par case entre les valeurs mesurées et celles que l'on aurait eu si le mode principal et le genre étaient indépendants. Une valeur **positive** signifie que les observations sont **plus élevées** que les valeurs théoriques (sous l'hypothèse d'indépendance), et une valeur **négative** signifie que les observations sont **en dessous** des valeurs théoriques (là encore sous l'hypothèse d'indépendance). Voir l'encart ci-dessous pour plus d'explications sur le test.

6.2.3 Aparté sur le test du chi2

Le test du chi2 est un test statistique sur l'indépendance de deux variables aléatoires. Ici, on observe le mode principal et d'autres caractéristiques, comme le fait de posséder ou non une carte de transport en commun. Intuitivement, l'indépendance de deux variables aléatoires représente le fait que la connaissance de l'une n'impacte pas la distribution de l'autre. Par exemple, dire que le mode principal est indépendant de la possession d'une carte de transport en commun revient à dire que le fait de savoir ou non qu'une personne possède une carte n'a aucun impact sur le fait qu'elle utilise les transports en commun. C'est très probablement faux : les personnes ayant une carte de transport en commun utilisent plus les transports en

commun que les autres.

En pratique, le test fonctionne de la manière suivante :

- on estime les distributions de **chacune** des variables aléatoires : ici cela correspond à calculer les proportions des panellistes utilisant les transports en commun, la voiture et les mobilités douces, ainsi que la proportion des utilisateurs et utilisatrices possédant une carte de transport en commun.
- **si les variables étaient indépendantes**, cela suffirait à calculer la distribution du couple. On pourrait donc par exemple estimer le nombre de personnes ayant une carte de transport en commun **et** prenant la voiture. Cela nous fournit un second tableau de même taille que celui qu'on observe.
- on fait enfin la différence entre ces deux tableaux : si les écarts sont trop grands, on conclut que les variables ne sont pas indépendantes. Afin de déterminer si ces écarts sont effectivement trop grands, on utilise une *garantie théorique (la p-valeur)*. Elle représente la probabilité que ces écarts ont d'apparaître si les variables étaient indépendantes. Si cette probabilité est trop faible (dans notre cas, inférieure à **0.05**), on conclut que les variables ne sont pas indépendantes.

7 Émissions de gaz à effet de serre

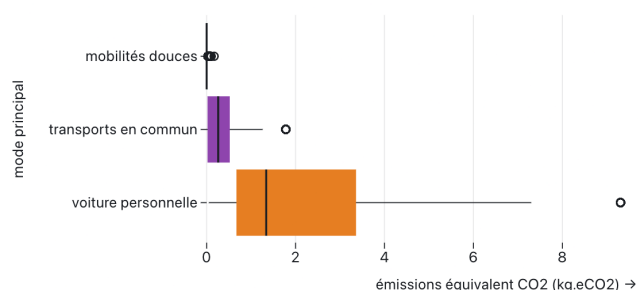
Les émissions de GES (gaz à effet de serre) sont calculées quotidiennement avec l'API Impact CO_2 de l'ADEME (6), une fois les distances pour chaque mode inférées à l'aide de Navitia (4) et d'Open Street Map (5). Le calcul des émissions prend en compte la multimodalité, ainsi que les différents types de moteurs (diesel, essence, électrique).

7.1 Distances et émissions de CO_2 par trajet

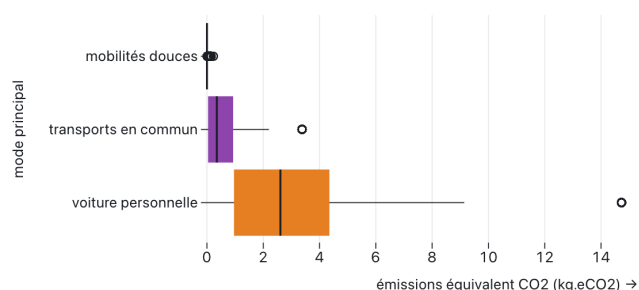
Ci-dessous, on présente des boîtes à moustaches illustrant la répartition des émissions de GES - mesurées en kg équivalent CO_2 - et des distances, pour chaque trajet et chaque mode. Le mode principal est ici le mode utilisé sur la plus longue distance.

Le mode principal affiché n'est pas exactement celui utilisé dans le calcul des émissions de GES. Par exemple, deux trajets en transport en commun de même distance peuvent ne pas avoir la même émission, si l'un a été effectué en bus et l'autre en métro. Ceci explique aussi pourquoi certains trajets avec le mode "mobilités douces" ont une émission non-nulle.

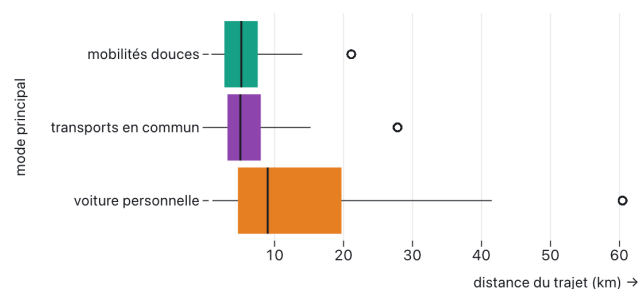
Répartition des émissions de CO_2 des trajets selon les modes - BMDec Brest



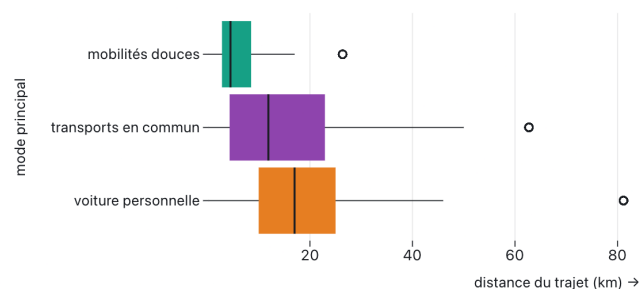
Répartition des émissions de CO_2 des trajets selon les modes - BMDec Rennes



Répartition des distances des trajets selon les modes - BMDec Brest



Répartition des distances des trajets selon les modes - BMDec Rennes



Les données des *outliers* ont été agrégées. Plus précisément, les *outliers* sont ici les trajets dont la valeur est supérieure à : $Q3 + 1.5 \cdot (Q3 - Q1)$; où Q3 représente la fin de la boîte, et Q1 le début (Q3-Q1 étant alors la largeur de la boîte). Leur valeur est remplacée par la moyenne des valeurs des *outliers*, et représentée par un cercle sur le diagramme.

Sur Brest, les distributions des distances pour les modes “mobilités douces” et “transports en commun” sont toutes deux localisées entre 0 et 10km. Le véhicule personnel est quasi systématiquement utilisé pour des distances supérieures. À Rennes en revanche, les transports en communs sont utilisés sur des distances plus longues. Une raison pouvant expliquer cela est l'utilisation du **train** pour les mobilités domicile-travail. À Rennes, on compte environ 3000 trajets et 85000 km en train, contre 1000 trajets et 25000 km à Brest (voir 7.2). Rien de surprenant, on compte une ligne de TER arrivant à Brest (passant par Landernau), contre 5 à Rennes. Il faut également prendre en compte que les panellistes brestois habitent en moyenne plus près de leur lieu de travail que les rennais : 65% des panellistes brestois habitent à moins de 10km de leur lieu de travail, contre 43% des panellistes rennais ((4.3)).

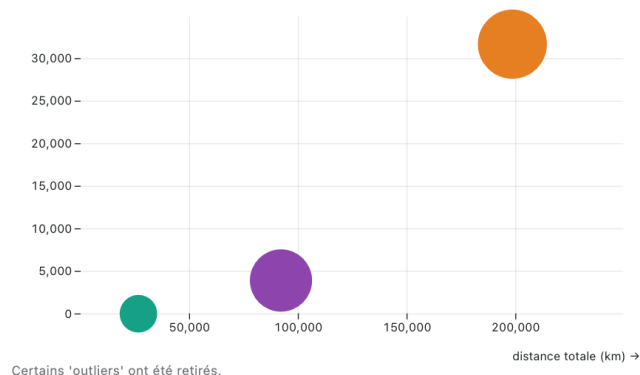
7.2 Émissions de CO_2 et distances parcourues totales, par mode

On représente en abscisse la distance totale (sur toute la durée de l'enquête), et en ordonnée les émissions d'équivalent CO_2 totales. Chaque disque représente un mode de transport. Plus un disque est grand, plus le mode a été utilisé pendant l'enquête.

Émissions de GES totales par mode | BMDec - Brest

voiture personnelle transports en commun mobilités douces

↑ co2 total (kg.eCO2)

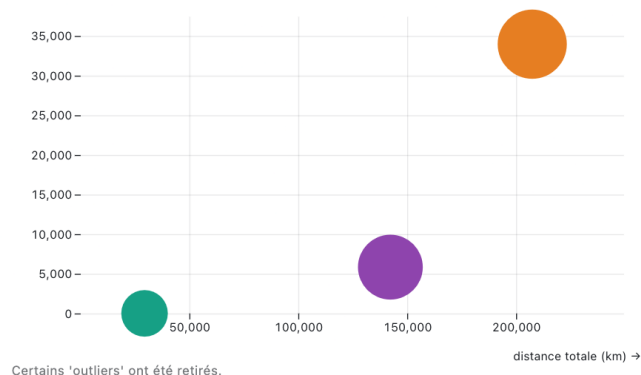


Certains 'outliers' ont été retirés.

Émissions de GES totales par mode | BMDec - Rennes

voiture personnelle transports en commun mobilités douces

↑ co2 total (kg.eCO2)



Certains 'outliers' ont été retirés.

Voici le détail des nombres de trajets (les valeurs sont arrondies à la centaine et les outliers retirés) :

Rennes :

	Véhicule personnel	Transports en communs	Mobilités douces
Distance totale (km)	207 200	142 000	29 200
Émissions de CO_2 totales (kg eCO2)	34 000	5 900	100
Nombre de trajets	10 400	9 100	4 700

Brest :

	Véhicule personnel	Transports en communs	Mobilités douces
Distance totale (km)	198 400	92 000	26 500
Émissions de CO_2 totales (kg eCO2)	31 700	3 900	< 50
Nombre de trajets	14 400	11 700	4 200

Les modes “transports en commun” et “véhicule personnel” sont utilisés sur des nombres de trajets comparables, dans les deux enquêtes. En revanche, les transports en commun ont été utilisés en tout sur des distances différentes (donc en moyenne sur des trajets plus courts) - probablement lié au TER.

Dans les deux villes, on note que la voiture est utilisée en moyenne sur des trajets **plus longs**, et **émet plus de CO_2 par km**. Ces deux facteurs entraînent que la majorité des émissions de CO_2 des enquêtes sont liées aux véhicules personnels.

8 Modélisation du comportement des panellistes

Afin de **synthétiser** le comportement de mobilité des panellistes, nous avons entraîné quelques modèles classiques de *Machine Learning* sur les données collectées. Ces modèles offrent un point de vue additionnel sur les données.

Ils apprennent à prédire le mode de transport utilisé sur un trajet par un ou une panelliste. L'objectif est de mettre en lumière d'éventuels liens entre des informations sur les panellistes (possession d'une carte de transport, âge, genre, distance domicile-travail) et leur choix de mode de transport.

Comme tous les modèles, ils nous donnent une **approximation de la réalité**, *via* les données que l'on a mesurées. Cette approximation est à interpréter avec du recul, et la capacité des modèles à généraliser sur d'autres populations doit être étudiée avant toute utilisation (il ne faut pas perdre de vue que les panellistes des enquêtes BMDec sont des volontaires).

Les modèles peuvent également être un moyen de diffuser des informations sur les données sans altérer l'anonymat des panellistes, les données étant **sensibles** et soumises au **RGPD**⁸. Cependant, des précautions doivent être prises afin de s'assurer que la diffusion du modèle ne brise pas l'anonymat des panellistes (voir 9).

On recense dans la littérature différentes méthodes pour modéliser les déplacements domicile-travail. Le livre *Modelling Transport* (3) présente différents modèles. La structure longitudinale des données rend la modélisation complexe. Dans un premier temps, nous avons agrégé les données par utilisateur, en calculant les proportions de chaque mode de transport utilisé pendant le sondage. Ceci nous permet de nous ramener à un jeu de données uni-dimensionnel. **Une étude plus approfondie mériterait d'être menée pour explorer les modèles longitudinaux.**

Le jeu de données simplifié contient une ligne par panelliste. Chaque ligne contient des informations socio-économiques ; et le mode de transport favori.

Les données utilisées en entrée des modèles sont :

8. RGPD : Règlement Général sur la Protection des Données : <https://www.cnil.fr/fr/comprendre-le-rgpd>.

- l'âge,
- le genre,
- la possession ou non d'une carte de transport en commun,
- la distance domicile-travail.

Les variables binaires sont encodées avec les valeurs :

- 1 : True
- 0 : False.

Les lignes avec des champs manquants ont été supprimées.

La sortie de ces modèles est le mode de transport utilisé le plus fréquemment pendant l'enquête (parmi : Mobilités Douces, Transports en Commun, Véhicule Personnel).

Nous avons entraîné 2 modèles sur ces données :

- un arbre de décision,
- une régression logistique (aussi appelée dans ce cadre : modèle à choix discrets).

La répartition de la classe à prédire (le mode de transport favori) est déséquilibrée (voir 6.1); avec entre 10 et 20 % seulement du mode "mobilité douces" (selon les villes).

8.1 Arbre de décision

Les arbres de décisions sont entraînés avec la bibliothèque python **scikit-learn**. Pour éviter le sur-apprentissage, la profondeur des arbres est bornée à 3. De plus, un hyper-paramètre de *pruning* a été ajusté manuellement (*ccp_alpha* = 0.05 pour Rennes et 0.01 pour Brest).

Les variables binaires sont encodées avec 0 et 1 : dire que la variable `carte_tc` est supérieure à 0.5 signifie que l'utilisateur possède une carte. Le genre est encodé de la manière suivante : 0 = Homme, 1 = Femme.

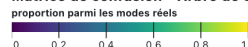
Arbre de décision Rennes :

```
|--- carte_tc <= 0.50
|   |--- dist_dom_travail <= 9.55
|   |   |--- class: MD
|   |--- dist_dom_travail > 9.55
|   |   |--- class: VP
|--- carte_tc > 0.50
|   |--- class: TC
```

Arbre de décision Brest :

```
|--- carte_tc <= 0.50
|   |--- dist_dom_travail <= 8.35
|   |   |--- genre <= 0.50
|   |   |   |--- class: MD
|   |   |--- genre > 0.50
|   |   |   |--- class: VP
|   |--- dist_dom_travail > 8.35
|   |   |--- class: VP
|--- carte_tc > 0.50
|   |--- class: TC
```

Matrice de confusion - Arbre de décision - BMDec Rennes



MD -	0.65	0.21	0.13
TC -	0.04	0.95	0.01
VP -	0.15	0.21	0.64
	MD	TC	VP

Les lignes correspondent aux valeurs réelles et les colonnes aux prédictions. Par exemple, le nombre à la colonne 'TC' et à la ligne 'VP' correspond au nombre de fois où le modèle a prédit 'TC' pour un panelliste utilisant majoritairement le mode 'VP'; divisé par le nombre de panellistes utilisant le mode 'VP'.

Matrice de confusion - Arbre de décision - BMDec Brest



MD -	0.58	0.23	0.19
TC -	0.01	0.9	0.09
VP -	0.15	0.21	0.64
	MD	TC	VP

Les lignes correspondent aux valeurs réelles et les colonnes aux prédictions. Par exemple, le nombre à la colonne 'TC' et à la ligne 'VP' correspond au nombre de fois où le modèle a prédit 'TC' pour un panelliste utilisant majoritairement le mode 'VP'; divisé par le nombre de panellistes utilisant le mode 'VP'.

Proportion de réponses correctes (*accuracy*), modèles croisés - Arbres de décisions

Modèle de	sur les données de Rennes	sur les données de Brest
Rennes	0.77	0.64
Brest	0.74	0.73

Sans surprise, l'arbre de décision capture très bien l'utilisation des transports en commun à l'aide de la variable "possession d'une carte de transport en commun". L'usage de la voiture est discriminé des mobilités douces avec la distance domicile travail. Dans le cas de Brest, le genre permet de discriminer l'usage des mobilités douces de la voiture (les femmes du panel ayant moins tendance que les hommes à utiliser les mobilités douces).

Avec ces 2 ou 3 règles simples, les modèles arrivent à reconstruire une part importante (supérieure à 75% pour Rennes) des choix de mobilité des panellistes (celles et ceux avec des informations manquantes ont été retirées). Le modèle semble s'appliquer moins bien aux données de Brest (le modèle de Brest est même meilleur sur les données de Rennes!).

8.2 Modèle à choix discrets

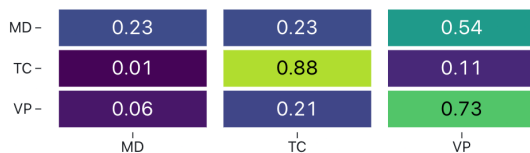
Le modèle à choix discret est issu de l'économétrie (16). Il vise à modéliser les choix des usagers et usagères, et a été utilisé dans la littérature sur les choix de modes de transports (17), (18). Sa construction est basée sur l'hypothèse qu'un usager ou une usagère fait son choix dans le but d'optimiser "l'utilité" de chacune des options. L'utilité est supposée construite de deux composantes : une aléatoire modélisant les goûts individuels de la personne, et une non-aléatoire modélisant les comportements généraux de la population. Le choix de la distribution de la composante aléatoire mène à différents modèles. Par exemple, en modélisant cette composante aléatoire par une loi de Gumbel⁹, on obtient une manière élégante de construire une **régression logistique** (16) (Lemme 1.1). C'est ce modèle qui est utilisé dans l'enquête du CREM (2) pour estimer l'impact des *nudges* sur les choix des panellistes.

Le modèle utilisé dans notre cas est alors une **régression logistique multinomiale**.

9. La loi de Gumbel est une distribution de probabilité sur \mathbb{R} , de fonction de répartition $F(\epsilon) = \exp(-\exp(\frac{\mu-\epsilon}{\beta}))$. https://fr.wikipedia.org/wiki/Loi_de_Gumbel. C'est aussi la Loi d'extremum généralisée de type 1.

Matrice de confusion - Modèle à choix discrets - BMDec Brest

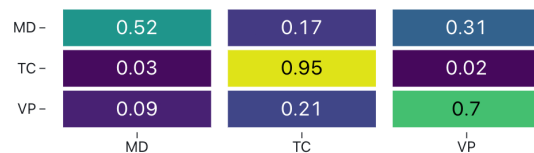
proportion parmi les modes réels



Les lignes correspondent aux valeurs réelles et les colonnes aux prédictions. Par exemple, le nombre à la colonne 'TC' et à la ligne 'VP' correspond au nombre de fois où le modèle a prédit 'TC' pour un panelliste utilisant majoritairement le mode 'VP'; divisé par le nombre de panellistes utilisant le mode 'VP'.

Matrice de confusion - Modèle à choix discrets - BMDec Rennes

proportion parmi les modes réels



Les lignes correspondent aux valeurs réelles et les colonnes aux prédictions. Par exemple, le nombre à la colonne 'TC' et à la ligne 'VP' correspond au nombre de fois où le modèle a prédit 'TC' pour un panelliste utilisant majoritairement le mode 'VP'; divisé par le nombre de panellistes utilisant le mode 'VP'.

Proportion de réponses correctes (*accuracy*), modèles croisés - Modèles à choix discrets

Modèle de	sur les données de Rennes	sur les données de Brest
Rennes	0.76	0.68
Brest	0.7	0.72

Encore une fois, le paramètre 'possession d'une carte de transport en commun' permet au modèle d'être très précis dans cette catégorie. Cependant, on perd beaucoup en précision sur les mobilités douces.

9 Diffusion des données

L'objectif premier de l'enquête est la collecte de données pour l'enquête du CREM (2). Cependant, les données collectées pourraient tout à fait être utilisées dans d'autres projets. Il n'est pas envisageable de diffuser telles quelles les données; du fait de leur caractère personnel. On pourra consulter la "politique de protection et d'utilisation des données personnelles du projet BMDec", sur le **site du projet**.

Le RGPD ¹⁰ prévoit qu'avant toute diffusion, les données doivent être **anonymisées**. Selon la CNIL ¹¹, des données sont anonymisées lorsqu'il est impossible d'identifier une personne dans le jeu de données. Certaines informations sont très fortement identifiantes (adresses, comportement de mobilité quotidien, ...). La combinaison de plusieurs informations peut également conduire à identifier des personnes. Par exemple, si deux personnes du panel ont exactement 35 ans, le fait de connaître leur genre permet de les différencier.

La notion de *Differential Privacy* (DP) (19); apporte des garanties de confidentialité; mais est complexe à mettre en oeuvre dans le cas de séries temporelles (20); et fait l'objet de recherches actuelles; dont les implémentations sont coûteuses à mettre en place.

9.1 Diffusion académique

Les données brutes ne sont pas diffusables au grand public, même pseudonymisées (en remplaçant les noms par des identifiants uniques). Elles contiennent trop d'informations sensibles pouvant être identifiantes. Pour une réutilisation académique encadrée, les données peuvent être diffusées pseudonymisées; après la signature d'un engagement de confidentialité. Les informations géographiques de ces données sont agrégées au niveau des IRIS ¹². De plus, les informations directement identifiantes comme l'adresse mail, le numéro de téléphone, les informations bancaires ou certains commentaires / mails sont bien évidemment supprimées.

10. RGPD : Règlement Général sur la Protection des Données : <https://www.cnil.fr/fr/comprendre-le-rgpd>.

11. <https://www.cnil.fr/fr/technologies/anonymisation-de-donnees-personnelles>

12. <https://www.insee.fr/fr/metadonnees/definition/c1523>

9.2 Modèles

Les modèles (arbres de décision, modèles à choix discret) apportent eux aussi des informations sur les données, et il convient de traiter les **risques** associés à l'usage et à la diffusion de ces modèles.

La première méthode de diffusion est de donner un accès au modèle en “boîte noire” (l'utilisateur ou l'utilisatrice du modèle n'accède qu'à l'entrée et à la sortie). Ceci limite les risques de fuite de données. Les risques étant en réalité inconnus, la solution retenue pour ce type de diffusion est de faire signer une clause sur “l'utilisation du modèle *via API*” par ses utilisateurs, s'engageant à ne pas essayer de réidentifier des utilisateurs avec le modèle. Sur des modèles plus complexes (notamment *longitudinaux*) que ceux présentés ci-dessus, cette solution est à envisager car les risques d'attaques devraient être plus importants. Néanmoins, la question de l'utilisation d'un tel modèle doit être posée : quel est l'intérêt d'utiliser un modèle dont on **ne peut pas** comprendre le fonctionnement ? Si l'intérêt est fondé, alors l'utilisation doit être encadrée, et les capacités de généralisation et limites du modèle doivent être étudiées (*quelle population notre modèle représente-il ?*).

La seconde est de diffuser le modèle en “boîte blanche”. Dans ce cas, et selon les modèles, plus d'informations sont données aux utilisateurs pour leur permettre d'utiliser le modèle localement. Cette méthode est plus risquée que la première, car plus d'informations sont données à de potentiels “attaquants”. Néanmoins, l'usage d'un modèle en “boîte blanche” permet d'apporter un point de vue additionnel sur les données (seuils de l'arbre de décision, poids du modèle à choix discrets). Pour assurer les garanties du RGPD, d'autres précautions doivent être prises. Pour l'arbre de décision ci-dessus, les données d'entraînement sont agrégées (on n'a plus de séries temporelles), et ne sont gardés que les seuils permettant de choisir entre l'une ou l'autre des catégories. Le nombre de panellistes correctement classé à chaque étape n'est pas diffusé, car il peut être utilisé dans certaines attaques de reconstruction (21) (22). De plus, des hyper-paramètres (*pruning*, *profondeur de l'arbre*) sont ajustés manuellement, pour éviter le sur-apprentissage, et donc de potentielles reconstructions du jeu de données d'entraînement. Des attaques de reconstruction ont été testées avec différentes valeurs de ces hyper-paramètres (par Julien Ferry¹³), et il en est ressorti qu'effectivement moins de sur-apprentissage diminue le risque de reconstruction.

10 Conclusion

Une part importante du projet BMDec a consisté à gérer des flux de données ; des utilisateurs et utilisatrices remplissant quotidiennement leurs informations de mobilité jusqu'aux personnes analysant les données.

La gestion de ces flux est complexe, notamment car les données sont :

- **sensibles** : ce sont des données personnelles, avec des informations géographiques,
- **longitudinales** : elles présentent une évolution sur plusieurs mois des comportements de mobilité.

Bien que les panellistes représentent une population sensibilisée aux enjeux d'émissions de CO_2 , l'étude de l'évolution de leur comportement au cours du temps permet de quantifier l'impact des *nudges* (2).

De plus, les enquêtes BMDec apportent une réalité de terrain, et donnent une base pour de futurs travaux. Ce type d'étude est nécessaire pour bien comprendre les leviers qui poussent les personnes à prendre des décisions de mobilité, ainsi que les différentes contraintes motivant ces dernières.

11 Bibliographie

- 1) *Observable Framework*. <https://observablehq.com/framework/>. Accessed 28 Feb. 2025.
- 2) Denant-Boemont, Laurent, et al. *L'efficacité Des Incitations Non Monétaires Dans La Diminution de La Part Modale de La Voiture Particulière Dans Les Déplacements Domicile-Travail : Une Expérimentation de Terrain*. 2024.

13. <https://fr.linkedin.com/in/julien-ferry-9435341a7>

- 3) Dios Ortúzar, Juan de, et al. *Modelling Transport*. John Wiley & Sons, Ltd, 2011, <https://doi.org/https://doi.org/10.1002/9781119993308>.
- 4) Navitia. <https://navitia.com/>. Accessed 28 Feb. 2025.
- 5) Open Street Map. <https://www.openstreetmap.org/>. Accessed 28 Feb. 2025.
- 6) ADEME, *Impact CO2*. <https://impactco2.fr/outils/transport>. Accessed 28 Feb. 2025.
- 7) Clairin, Rémy, et al. *Manuel de Sondages : Applications Aux Pays En Développement / Rémy Clairin, Philippe Brion ; Préf. De Jacques Vallin ; [Publ. Par Le] CEPED [Et l'] Institut National de La Statistique Et Des Études Économiques*. Centre français sur la population et le développement (Paris), 1996.
- 8) Streamlit. <https://streamlit.io/>. Accessed 28 Feb. 2025.
- 9) Sea Data Net ; Data Quality. <https://vocab.nerc.ac.uk/collection/L20/current/>. Accessed 28 Feb. 2025.
- 10) Axhausen, Kay W., et al. *Observing the Rhythms of Daily Life : A Six-Week Travel Diary*. 2002, <https://doi.org/10.1023/A:1014247822322>.
- 11) Axhausen, K. W., et al. *Fatigue in Long-Duration Travel Diaries*. 2007, <https://doi.org/10.1007/s11116-006-9106-4>.
- 12) Audiar-Rennes. *Enquête Ménages-Déplacements En Ile-Et-Vilaine 2018 - Synthèse*. 2018, <https://www.audiar.org/publication/mobilites/deplacements/enquete-menages-deplacements-en-ille-et-vilaine-2018/>.
- 13) ADEUPa. *Les Résultats Principaux de l'enquête Ménages Déplacements Du Pays de Brest*. 2018, <https://adeupa-brest.fr/nos-publications/les-resultats-principaux-de-lenquete-menages-deplacements-du-pays-de-brest>.
- 14) Rodolphe Charrier, Fabien Perez. *Comment Les Français Se Déplacent-Ils Pour Aller Travailler ?* 2024, <https://www.statistiques.developpement-durable.gouv.fr/comment-les-francais-se-deplacent-ils-pour-aller-travailler-0>.
- 15) Chat-Sceptique. *Les Intervalles de Confiance Expliqués à Mon Chat*. <https://www.youtube.com/watch?v=WEF2SgS95N8>. Accessed 28 Feb. 2025.
- 16) McFadden, Daniel. *Conditional Logit Analysis of Qualitative Choice Behavior*. 1972, <https://api.semanticscholar.org/CorpusID:153296073>.
- 17) Paredes, Miguel, et al. "Machine Learning or Discrete Choice Models for Car Ownership Demand Estimation and Prediction ?" *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 2017, pp. 780–85, <https://doi.org/10.1109/MTITS.2017.8005618>.
- 18) Hagenauer, Julian, et al. "A Comparative Study of Machine Learning Classifiers for Modeling Travel Mode Choice." *Expert Systems with Applications*, vol. 78, 2017, pp. 273–82, <https://doi.org/https://doi.org/10.1016/j.eswa.2017.01.057>.
- 19) Dwork, Cynthia, et al. "Calibrating Noise to Sensitivity in Private Data Analysis." *Journal of Privacy and Confidentiality*, vol. 7, no. 3, May 2017, pp. 17–51, <https://doi.org/10.29012/jpc.v7i3.405>.
- 20) Allard, Tristan, et al. "Analyzing and explaining privacy risks on time series data : ongoing work and challenges." *SIGKDD Explorations : Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining*, vol. 26, no. 1, 2024, pp. 1–10, <https://doi.org/10.1145/3682112.3682118>.
- 21) Gambs, Sébastien, et al. "Reconstruction Attack Through Classifier Analysis." *Data and Applications Security and Privacy XXVI*, edited by Nora Cuppens-Boulahia et al., Springer Berlin Heidelberg, 2012, pp. 274–81.
- 22) Ferry, Julien, et al. "Probabilistic Dataset Reconstruction from Interpretable Models." *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, 2024, pp. 1–17, <https://doi.org/10.1109/SaTML59370.2024.00009>.