

Coriolan Gout

Sous la direction de Matthieu Schorung et Laetitia Dablanc

Octobre 2023

Note de recherche

Proposition méthodologique pour l'analyse de la distribution spatiale des entrepôts et du marché locatif de l'immobilier logistique dans les aires métropolitaines de Paris et de New York

Résumé

Cette note de recherche fait une proposition méthodologique pour l'acquisition et le traitement de données issues de bases de données non harmonisées, à partir des deux études de cas de Paris et de New York. Cette étude fait suite à une étude préliminaire réalisée sur le cas de Tokyo (mai 2023) pour construire cette méthodologie innovante. Deux objectifs ont guidé la rédaction de cette note de recherche : (i) comprendre la répartition spatiale des entrepôts dans les agglomérations de Paris et de New York ; (ii) obtenir des informations sur les loyers de l'immobilier logistique dans les agglomérations de Paris et de New York.

La contribution de cette recherche est d'abord méthodologique : développer un cadre avec des données homogènes sur les prix de l'immobilier logistique pour comparer les régions métropolitaines en considérant la structure spatiale des installations logistiques, les prix des loyers de l'immobilier logistique et les facteurs qui expliquent la dynamique des prix de l'immobilier logistique. Dans le cadre de cette étude, deux sites Internet ont été identifiés à partir desquels nous avons pu extraire des informations sur les entrepôts, leurs caractéristiques, leurs loyers et leur localisation pour le cas de Paris (immobilier.cbre.fr ; immobilier.jll.fr) et un pour le cas de New York (Loopnet.com). Pour obtenir ces données, nous utilisons une méthode d'extraction de données par *web-scraping* déjà testée dans Oliveira et al. (2022). Cependant, contrairement aux sites de JLL et CBRE, il n'a pas été possible d'utiliser des méthodes de *web-scraping* (à l'aide du package *rvest*) pour extraire les données. Face à cette difficulté, nous avons opté pour une extraction manuelle des informations, abandonnant l'idée de construire un script et proposant ainsi une méthode systématique.

Cette recherche apporte un cadre innovant de comparaison des aires métropolitaines dans des contextes géographiques différents, basé sur des analyses spatiales et une méthode homogène. Les résultats de ce travail sont reproductibles et peuvent inciter les autorités publiques locales et régionales à développer une politique publique plus efficace en matière d'aménagement logistique du territoire et de planification des transports.

Summary

This research note makes a methodological proposal for acquiring and processing data from non-harmonized databases, based on the two case studies of Paris and New York. This study follows on from a preliminary study carried out on the case of Tokyo (May 2023) to build this innovative methodology. Two objectives guided the writing of this research note: (i) to understand the spatial distribution of warehouses in the Paris and New York metropolitan areas; (ii) to obtain information on logistics real estate rents in the Paris and New York metropolitan areas.

This research's contribution is first methodological: developing a framework with homogeneous data on logistics real estate prices for comparing metropolitan regions considering the spatial pattern of logistics facilities, logistics real estate rental prices and the factors that explain the dynamics of logistics real estate prices. As part of this study, two websites were identified from which we were able to extract information on warehouses, their characteristics, rents and location for the Paris case (immobilier.cbre.fr; immobilier.jll.fr) and one for the New York case (Loopnet.com). To obtain these data, we use a web-scraping data extraction method already tested in Oliveira et al. (2022). However, unlike the JLL and CBRE sites, it was not possible to use web-scraping methods (using the *rvest* package) to extract data. Faced with this difficulty, we opted to extract the information manually, abandoning the idea of building a script and thus proposing a systematic method.

This research brings an innovative framework for comparing metropolitan areas within different geographic contexts based on spatial analyses and an homogenous method. This work's results are reproducible and can induce local and regional public authorities to develop a more effective public policy addressing logistics land use and transportation planning.

Sommaire

1. Le cas de Paris	3
1.1 Introduction : cadre de recherche et objectifs	3
1.2 Présentation de la méthode d'extraction des données	3
1.3 Création du script d'extraction	3
1.4 Analyse de la distribution spatiale des entrepôts.....	6
1.5 Création du modèle	8
1.6 Estimations à partir du modèle.....	11
1.7 Limites.....	14
2. Le cas de New York	15
2.1 Présentation de la méthode d'extraction des données	15
2.2 Extraction manuelle des annonces	15
2.3 Création du modèle	19
2.4 Estimations à partir du modèle.....	22
2.5 Limites.....	24

1. Le cas de Paris

1.1 Introduction : cadre de recherche et objectifs

Cette étude a été réalisée dans le cadre d'une mission de recherche supervisée par Matthieu Schorung et financée par la Chaire Logistics City. Elle fait suite à une [étude préliminaire](#) sur le cas de l'agglomération de Tokyo intitulée : « *Proposition méthodologique pour l'analyse de la distribution spatiale des entrepôts et du marché locatif de l'immobilier logistique dans l'aire métropolitaine de Tokyo* ». Elle s'inscrit dans le thème 1 de la Chaire Logistics City sur l'immobilier logistique urbain, les nouveaux modèles économiques de l'immobilier logistique métropolitain, les stratégies d'implantation des bâtiments logistiques dans les grandes métropoles. Elle contribue aux travaux de validation des hypothèses sur les relations entre localisation des entrepôts et formes urbaines posées par L. Dablanc dans [l'analyse](#) *Locational patterns of warehouses in 74 cities around the world, a comparative meta-analysis*, (Dablanc, Palacios- Argüello, de Oliveira, 2020).

Cette note de recherche fait une **proposition méthodologique pour l'acquisition de données à partir de bases non harmonisées et leur traitement**, à partir des deux cas d'études de Paris et New York.

Deux objectifs ont guidé la rédaction de cette note de recherche :

- Comprendre la distribution spatiale des entrepôts dans les aires métropolitaines de Paris et New York.
- Obtenir des informations sur les loyers de l'immobilier logistique dans les aires métropolitaines de Paris et New York.

1.2 Présentation de la méthode d'extraction des données

Pour obtenir des données sur la localisation des entrepôts et les loyers de l'immobilier logistique dans l'aire métropolitaine de Paris, nous utilisons la méthode présentée dans [l'article](#) (2022) *Changes in warehouse spatial patterns and rental prices: Are they related? Exploring the case of US metropolitan areas*, par R. de Oliveira, L. Dablanc et M. Schorung. L'article présente une méthode **d'extraction de données par web-scraping**, utilisée pour transformer les informations d'un site web en données exploitables.

Dans le cadre de cette étude, ont été identifiés deux sites web desquels nous avons pu extraire des informations sur les entrepôts, leurs caractéristiques, loyers et localisation : *immobilier.cbre.fr* et *immobilier.jll.fr*.

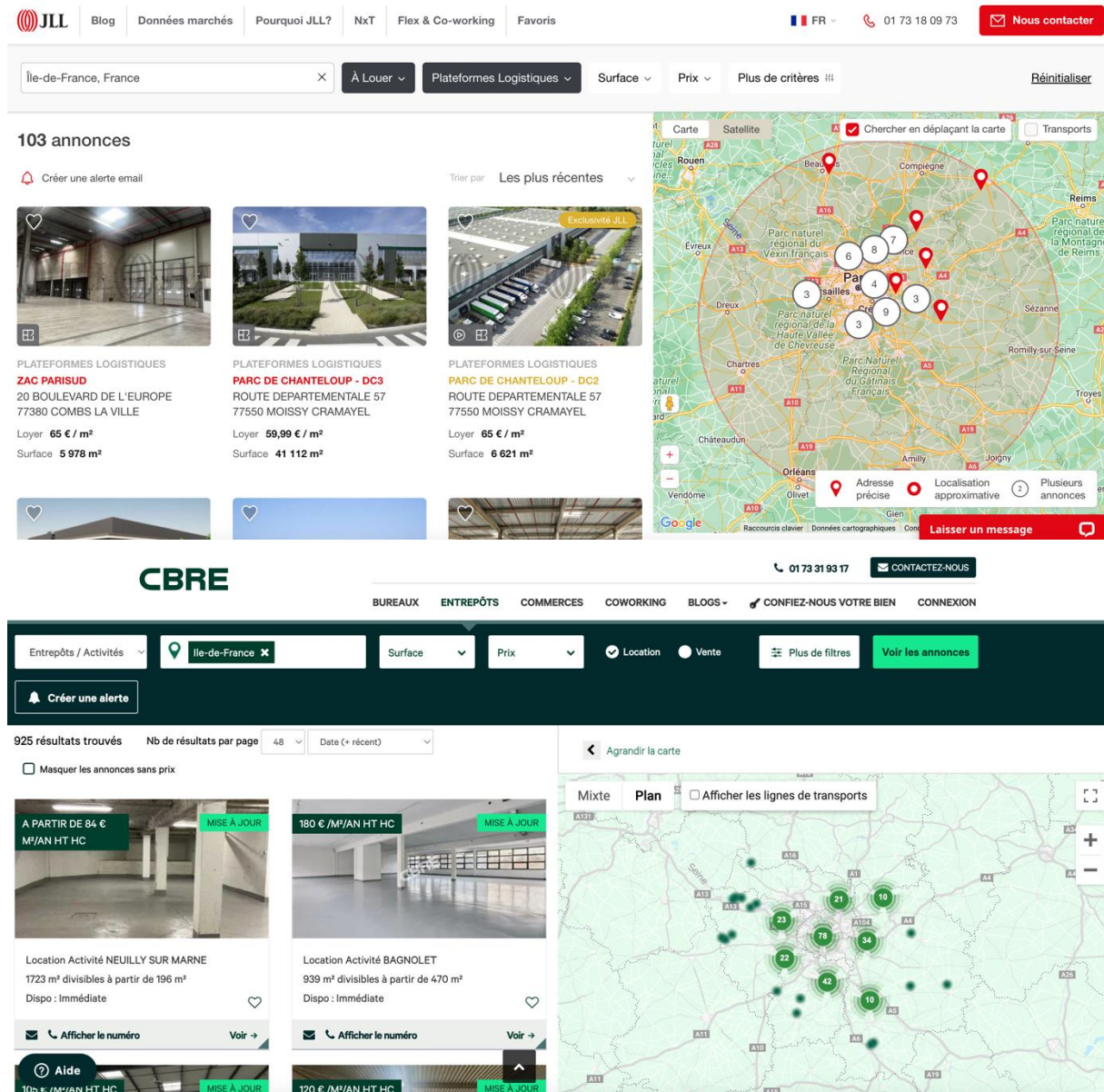
Seul le site de JLL recense les adresses précises des entrepôts, mais le nombre d'annonces présentées est trop faible pour pouvoir mettre en place une méthode d'estimation des loyers. Le site de CBRE, lui, n'indique que les communes dans lesquelles se situent les entrepôts. Nous avons fait le choix d'utiliser les données de ces deux sites pour pouvoir garantir un échantillon d'une taille convenable, malgré la différence de qualité des données.

1.3 Création du script d'extraction

Les deux sites fonctionnent comme des sites d'annonces immobilières classiques. Il est possible d'y chercher des propriétés à louer en fonction de leur localisation, leur surface, leur prix, etc. Dans le cadre de cette étude, nous avons fait le choix d'extraire les entrepôts situés dans un périmètre régional (l'Île-de-France) relativement large.

Voici des exemples d'annonces publiées par les sites :

Figure 1. Captures d'écran de deux sites d'annonce immobilière et des informations disponibles



Le script de *web-scraping*, rédigé en langage R, permet d'extraire le texte du site et de l'organiser dans un tableau. Il se base sur les « nœuds HTML » pour repérer à quoi correspondent les zones de texte. Une majorité de sites possèdent une partie de code en HTML, qui permet d'organiser leurs différents éléments constitutants et leur apparence (texte, image, vidéo...). Au sein de ce code, des « nœuds » permettent d'indiquer à quelle catégorie correspondent les zones de texte et de leur faire hériter les caractéristiques associées à ladite catégorie.

Grâce au package *rvest* et à la fonction `html_nodes`, il est possible d'extraire tous les textes appartenant à un même « nœud ». Ici notre objectif était d'extraire deux informations pour chaque entrepôt :

- Sa localisation (adresse complète ou ville)
- Son loyer/m²/mois

Script pour le site JLL.fr

Dans le code HTML, les adresses sont indiquées par le nœud « .a » et les loyers par le nœud « .PropertyMetric__item »

Une fois extraites et organisées sous forme de tableau, nous géocodons les adresses à l'aide de la fonction geocode() de l'extension ggmmap. Cette fonction utilise l'API de Google Maps pour transformer les adresses en coordonnées. Elle est très performante mais nécessite une clé API pour fonctionner. Il suffit d'effectuer la demande sur le site developers.google.com pour en obtenir une.

A partir des latitudes et longitudes obtenues et grâce à la fonction st_as_sf, nous créons une colonne « géométrie » stockant les informations géospatiales des entrepôts : des points projetés en Lambert 93.

Figure 2. Extrait de la base de données constituée pour la région parisienne (adresse, loyer, coordonnées géographiques)

	adresse_texte	loyer_texte	lon	lat	geom
1	PA LES DOUCETTES2 AVENUE DES MORILLONS95140 ...	150.00	2.400724	48.97178	POINT (656126.5 6874815)
2	Contactez JLL pour plus d'information94460 VALENTON	147.25	2.467475	48.74814	POINT (660845.6 6849912)
3	MESSAGERIE6 RUE EUGENE POTTIER95670 MARLY LA ...	145.00	2.519219	49.08360	POINT (664877.2 6887191)
4	LES PORTES DE VEMARS – DC 6 RUE DE LA TOUR9547...	140.00	2.562319	49.06912	POINT (668016.9 6885562)
5	DC 25 GARONOR OUEST93600 AULNAY SOUS BOIS	135.00	2.463375	48.95919	POINT (660703.6 6873382)
6	DC110 RUE DIDEROT93110 ROSNY SOUS BOIS	135.00	2.485932	48.87566	POINT (662295 6864083)
7	IDF SUD / PORT DE BONNEUIL16 ROUTE DE STAINS94...	135.00	2.495645	48.78145	POINT (662940.4 6853603)
8	C3922 ROUTE DU MOLE 2 ET 392230 GENNEVILLIERS	134.92	2.293275	48.92553	POINT (648214.2 6869736)
9	Contactez JLL pour plus d'information92230 GENNEVI...	130.00	2.293275	48.92553	POINT (648214.2 6869736)
10	Contactez JLL pour plus d'information94000 CRETEIL	123.86	2.455572	48.79037	POINT (660002.8 6854614)
11	VAL ACTI OUEST – PROJET 122 RUE DES OSIERS78310...	120.00	1.916261	48.75028	POINT (620322.3 6850566)
12	BATIMENT 14 RUE ROBERT BREMOND93600 AULNAY ...	120.00	2.458573	48.95508	POINT (660348.9 6872927)
13	50 RUE DES OSIERS78310 COIGNIERES	120.00	1.916261	48.75028	POINT (620322.3 6850566)
14	IDF SUD / POLE DE TRAPPES ELANCOURT10–20 AVEN...	120.00	2.001844	48.77673	POINT (626651.3 6853423)
15	A910 ROUTE OUEST DU MOLE 192230 GENNEVILLIERS	110.00	2.293275	48.92553	POINT (648214.2 6869736)

La majorité des adresses indiquées sur le site sont complètes. Les quelques adresses n'indiquant que la commune dans laquelle se situe l'entrepôt ont été géocodées au niveau du nom de la commune sur Google Maps (par exemple : la ligne 2 de la figure 2 avec un entrepôt situé à Valenton).

Script pour le site CBRE.fr

Dans le code HTML, les loyers sont indiqués par le nœud « .price ». Les adresses quant à elles sont contenues dans un champ de texte indiqué par le nœud « .caption ».

.caption rassemble toute la partie inférieure de l'annonce, avec le type de location (entrepôt, activité), la surface de la location et sa disponibilité.

Toutes les annonces référencées par la catégorie « entrepôts / activités » ne concernent pas des entrepôts. La majorité d'entre-elles sont désignées comme « activités », terme qui rassemble des lots très divers : des bâtiments mixtes, des bureaux et même des entrepôts. Dans le doute,

nous avons fait le choix de ne garder que les annonces indiquées comme « entrepôts » dans la partie .caption.

Nous obtenons un tableau avec deux colonnes : le loyer et la commune de l'entrepôt.

Pour localiser les entrepôts à partir de la commune, la méthode la plus simple était de leur donner les coordonnées des centroïdes de leur commune. Cette méthode permet de placer les entrepôts au centre géographique des communes, qui est parfois assez éloigné du vrai centre-ville mais aussi des zones d'activité. Nous avons alors préféré utiliser le shapefile CHFLIEU_COMMUNE de la base ADMIN_EXPRESS produite par l'IGN. Cette couche vecteur représente la mairie ou l'hôtel de ville de chaque commune par un point. Les points ne se situent pas nécessairement proche des centres géographiques des communes, mais plus souvent au niveau du centre-ville et donc proche des infrastructures de transports, activités, commerces... Voici les étapes pour joindre les loyers à la couche CHFLIEU_COMMUNE :

- Géocoder les communes de la table CBRE
- Effectuer une jointure spatiale pour joindre les loyers à la couche COMMUNE (polygones) de la BDTOPO
- Effectuer une jointure attributaire pour joindre les loyers depuis la couche COMMUNE à la couche CHFLIEU_COMMUNE, en utilisant l'attribut « CODE_INSEE »

On obtient une couche avec un point pour chaque entrepôt, situé au niveau de la mairie ou de l'hôtel de ville de sa commune et auquel est associé son loyer.

Les données utilisées ont été extraites le 3 septembre 2023.

1.4 Analyse de la distribution spatiale des entrepôts

L'échantillon final est composé de 140 entrepôts situés en région Ile-de-France. Nous devons nous contenter de cet échantillon assez réduit ; aucune donnée historique n'est disponible sur les sites CBRE et JLL.

Voici une carte du jeu de données :

Figure 3. Distribution spatiale des entrepôts logistiques de l'échantillon de l'étude en Ile-de-France (septembre 2023)

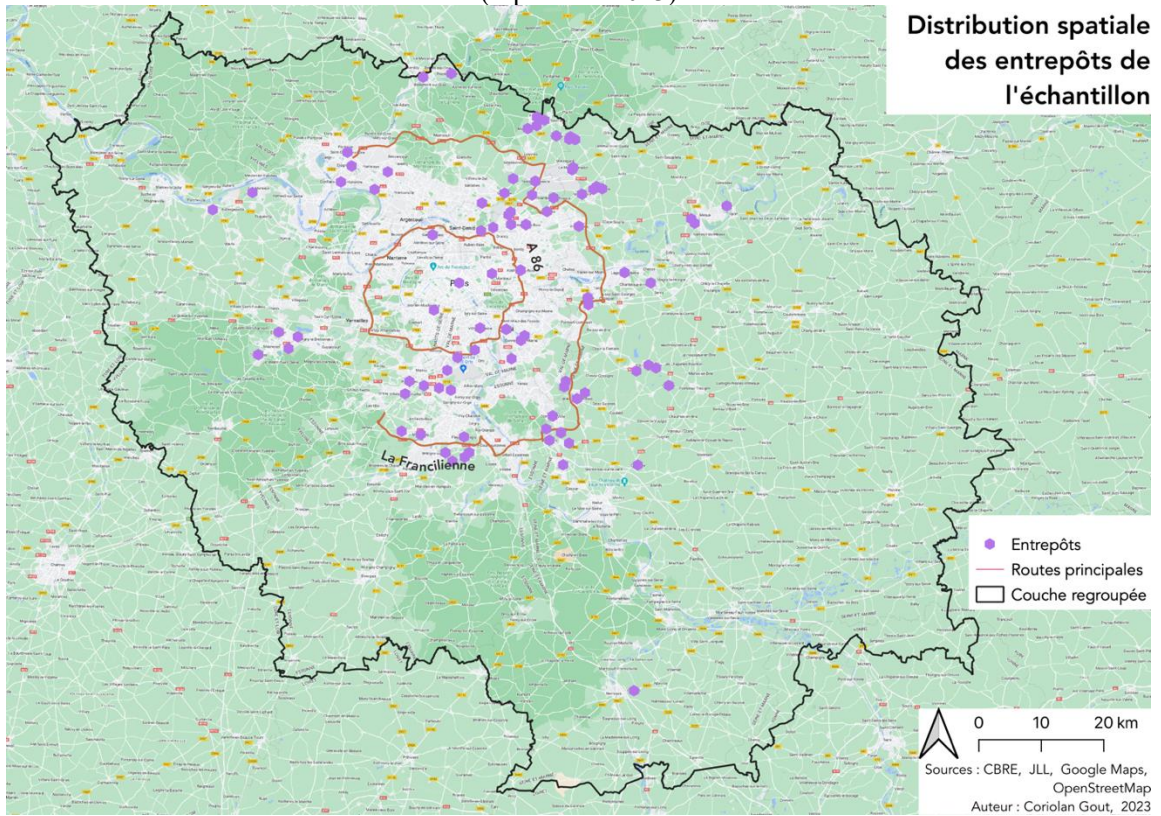
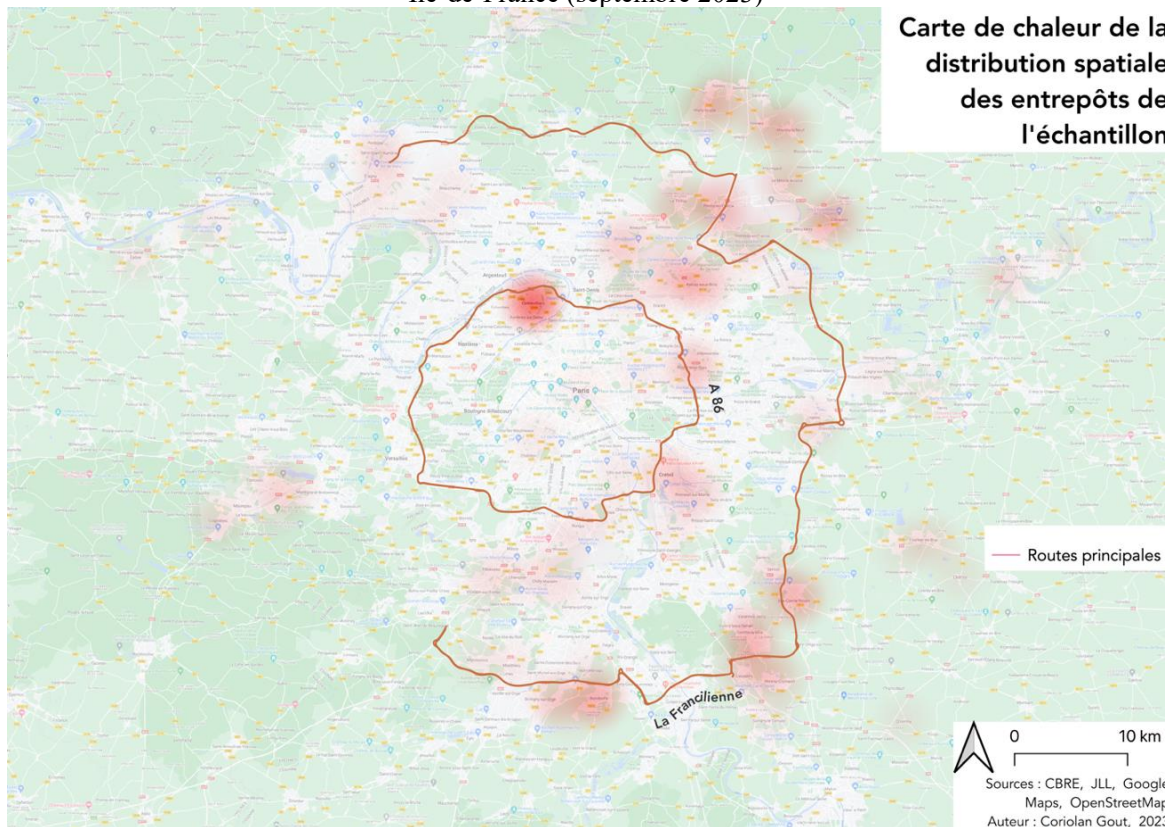


Figure 4. Distribution spatiale en carte de chaleur des entrepôts logistiques de l'échantillon de l'étude en Ile-de-France (septembre 2023)



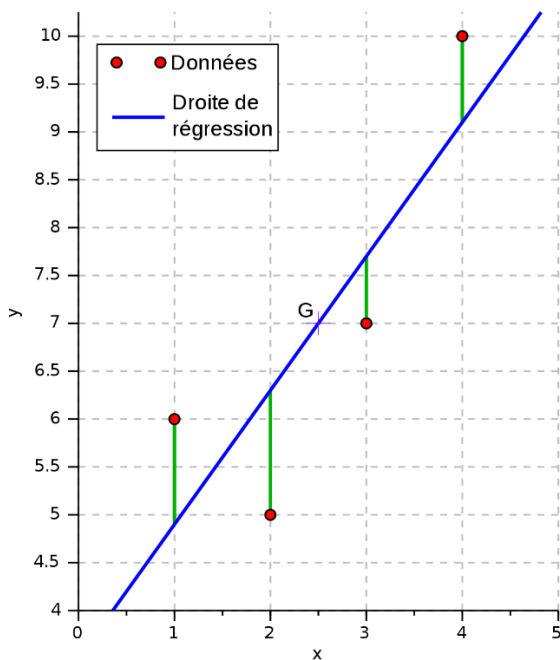
Plusieurs entrepôts peuvent se situer au même point. La carte de chaleur permet ainsi de représenter plus clairement la distribution spatiale des entrepôts. On constate que la majorité des entrepôts de l'échantillon se trouvent dans la moitié Est de la région. Il y a une forte concentration d'entrepôts au Sud-Est et au Nord-Est de la capitale.

En plus de cette distribution différenciée entre Ouest et Est, la répartition des entrepôts semble suivre une logique circulaire. Un premier cercle se dessine « à l'extérieur » de l'A86, concentrant des entrepôts à une douzaine de kilomètres du centre de Paris. Le deuxième cercle suit le tracé de la Francilienne, de 25 à 28 km du centre de Paris.

1.5 Création du modèle

Le deuxième objectif de l'étude est d'estimer les loyers de l'immobilier logistique dans l'aire métropolitaine de Paris. Pour y répondre, nous avons créé un modèle de régression linéaire multiple¹. Un modèle de régression linéaire multiple fonctionne comme un modèle de régression simple, à la différence qu'il utilise plusieurs variables prédictives.

Figure 5. Schéma d'illustration du principe de régression linéaire.



Cette droite de régression peut être mise en équation : $y = ax + b$

- y correspond à la valeur de la variable expliquée
- x correspond à la valeur de la variable explicative
- a est la « pente » de l'équation, le coefficient multiplicateur de la variable explicative
- b est l'« origine » ou « intercept ». C'est une constante indiquant la valeur de y si $x = 0$

¹ Pour rappel. Un modèle de régression linéaire simple cherche à établir une relation linéaire entre une variable, dite expliquée, et une variable, dite explicative. À partir d'un jeu de données, il permet d'estimer une droite de régression la plus fidèle possible à la relation entre les deux variables.

Une fois la droite de régression estimée grâce au modèle, il est possible d'estimer la valeur de y pour toute valeur de x . La régression linéaire multiple fonctionne de la même manière, mais avec autant de coefficients a_1, a_2, \dots qu'il y a de variables explicatives x_1, x_2, \dots .

L'objectif est alors d'estimer les coefficients multiplicateurs de chaque variable explicative, qui permettront d'obtenir des valeurs estimées de la variable expliquée, ici le loyer au m² des entrepôts.

Nous avons créé un jeu de données de plus d'une vingtaine de variables, pour tenter d'expliquer le loyer annuel au m² des entrepôts dans la zone d'étude. Les variables sont observables à l'échelle de la commune. Voici les différentes variables retenues et leur source.

- Densité de Population (INSEE)
- Loyer moyen pour un appartement du parc privé (Observatoire des territoires)
- Occupation du sol (Observatoire des territoires)
- Part des emplois par secteur d'activité (Observatoire des territoires)
- Distances à vol d'oiseau à différentes infrastructures de transport (Calcul personnel à partir de données de la base ADMIN EXPRESS et d'Openstreetmap)

Voici le modèle que nous avons gardé, après avoir testé les différentes variables du jeu de données :

Log(Loyer annuel au m²) ~ Loyer moyen par m² pour un appartement du parc privé + Distance à vol d'oiseau à la Francilienne + Distance à vol d'oiseau à l'Aéroport d'Orly + Distance à vol d'oiseau à l'A86 + Part des emplois dans l'industrie

Ce modèle permet d'estimer le logarithme du loyer annuel au m² des entrepôts d'Ile-de-France. C'est un modèle linéaire avec logarithme.

Dans un modèle linéaire standard (sans logarithme), les coefficients sont interprétés comme le changement dans la variable dépendante pour un changement d'une unité dans la variable indépendante, tout en maintenant les autres variables constantes. Cette approche suppose une relation linéaire directe entre les variables.

En revanche, un modèle linéaire avec logarithme prend en compte des relations non linéaires en prenant le logarithme de certaines variables. Les coefficients dans ce modèle indiquent maintenant le pourcentage de changement (ou le changement multiplicatif) dans la variable dépendante pour un changement d'une unité dans la variable indépendante. Cette approche est utile lorsque la relation entre les variables n'est pas linéaire à l'origine, ce qui peut être le cas dans de nombreuses situations réelles.

Dans notre cas, l'utilisation d'un modèle linéaire avec logarithme rend la relation entre la valeur prédite et les différents facteurs beaucoup plus significative.

Le tableau suivant résume le résultat du modèle.

Tableau 1. Résultats statistiques du modèle linéaire et significativité des relations aux variables.

Variable	Coefficient	Erreur standard	t-value	p-value	Significativité
Intercept	3.863e+00	3.593e-01	10.751	< 2e-16	***
Loyer moyen par m² pour un appartement du parc privé (en €)	4.463e-02	1.972e-02	2.264	0.0252	*
Distance à vol d'oiseau à la Francilienne (en m)	1.042e-05	4.050e-06	2.574	0.0112	*

Distance à vol d'oiseau à l'Aéroport d'Orly (en m)	4.396e-06	2.624e-06	1.675	0.0962	*
Distance à vol d'oiseau à l'A86 (en m)	-2.348e-05	4.520e-06	-5.195	7.57e-07	***
Part des emplois dans l'industrie (en %)	-5.880e-03	3.673e-03	-1.601	0.1118	*

Puisque notre modèle estime le logarithme du loyer annuel au m², l'interprétation de ses résultats est différente de ceux d'un modèle linéaire standard.

En interprétant les coefficients, on peut dire que chaque augmentation d'une unité sur l'échelle logarithmique équivaut à un certain pourcentage de changement dans la variable dépendante. Par exemple, si le coefficient associé à une variable est de 0,1, cela signifie qu'une augmentation d'une unité dans cette variable est associée à une augmentation d'environ 10 % dans la variable dépendante, toutes choses étant égales par ailleurs.

Prenons un cas précis : si le loyer par mètre carré pour un appartement du parc privé augmente d'un euro, la valeur attendue du logarithme du loyer annuel par m² augmentera d'environ 4,463 %. Il est important de noter que cette interprétation est basée sur la supposition que toutes les autres variables restent constantes, comme le spécifie un modèle de régression linéaire.

Résumons donc les différentes relations entre le loyer et les variables explicatives.

Tableau 2. Résumé des relations entre le loyer et les variables.

Variable	Relation	Explication
Loyer moyen par m² pour un appartement du parc privé (en €)	Positive	Le loyer augmente quand le loyer moyen pour un appartements du parc privé augmente
Distance à vol d'oiseau à la Francilienne (en m)	Positive	Le loyer augmente plus on s'éloigne de la Francilienne. Il diminue quand on s'en rapproche
Distance à vol d'oiseau à l'Aéroport d'Orly (en m)	Positive	Le loyer augmente plus on s'éloigne de l'Aéroport d'Orly. Il diminue quand on s'en rapproche
Distance à vol d'oiseau à l'A86 (en m)	Négative	Le loyer diminue plus on s'éloigne de l'Aéroport d'Orly. Il augmente quand on s'en rapproche
Part des emplois dans l'industrie (en %)	Négative	Le loyer diminue quand la part des emplois dans l'industrie augmente

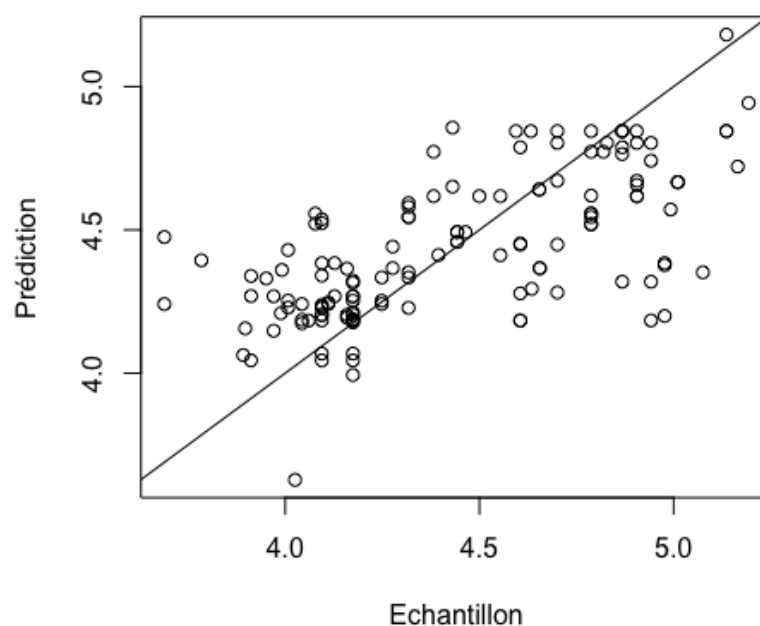
Le tableau donne également des informations sur la fiabilité du modèle. La p-value permet de quantifier la significativité de la relation entre la variable explicative et la variable expliquée, dans le cadre d'un modèle spécifique. Plus elle est faible, plus la probabilité que la relation observée soit le fruit du hasard est faible. Pour la variable « part des emplois dans l'industrie », la p-value est de 0,11. On peut affirmer qu'il y a 11 % de probabilité que la relation entre les deux variables soit due au hasard, 89 % pour la réciproque. On ne peut pas dire que la relation

est strictement significative, mais elle l'est assez, au vu de la taille réduite de l'échantillon observé. A contrario, la relation entre le loyer annuel au m² et la distance à vol d'oiseau à l'A86 est strictement significative.

Le R², aussi appelé « coefficient de détermination », permet de juger la qualité de la régression. Plus le R² est élevé, plus on peut considérer que le résultat du modèle est fiable. Dans le cas de ce modèle, il est de 0,44. Il peut se traduire ainsi : le résultat du modèle correspond à la réalité observée dans l'échantillon dans 44 % des cas.

Le nuage de points suivant présente la relation entre l'échantillon et la prédiction. Plus les points se situent proches de la ligne (x=y), plus la prédiction est bonne et le R² élevé. Dans notre cas, on observe que le modèle a tendance à surestimer les loyers, en particulier les loyers les plus faibles.

Figure 6. Coefficient de détermination visant à déterminer la solidité du modèle.



1.6 Estimations à partir du modèle

En appliquant le modèle à l'échelle de la zone d'étude, on obtient les cartes suivantes :

Tableau 3. Estimation des loyers (en euros/m²/an) des entrepôts logistiques dans la région Ile-de-France.

Zone	Moyenne	Médiane	Minimum	Maximum
<i>Ile-de-France</i>	69	63	19	202
Seine et Marne	50	48	19	92
Yvelines	80	73	45	150
Essonne	65	62	36	114
Hauts de Seine	145	143	118	202
Seine-Saint-Denis	110	111	83	136
Val de Marne	114	116	70	168
Val d'Oise	75	72	41	127

Figure 7. Estimation des loyers (en euros/m²/an) des entrepôts logistiques dans la région Ile-de-France.

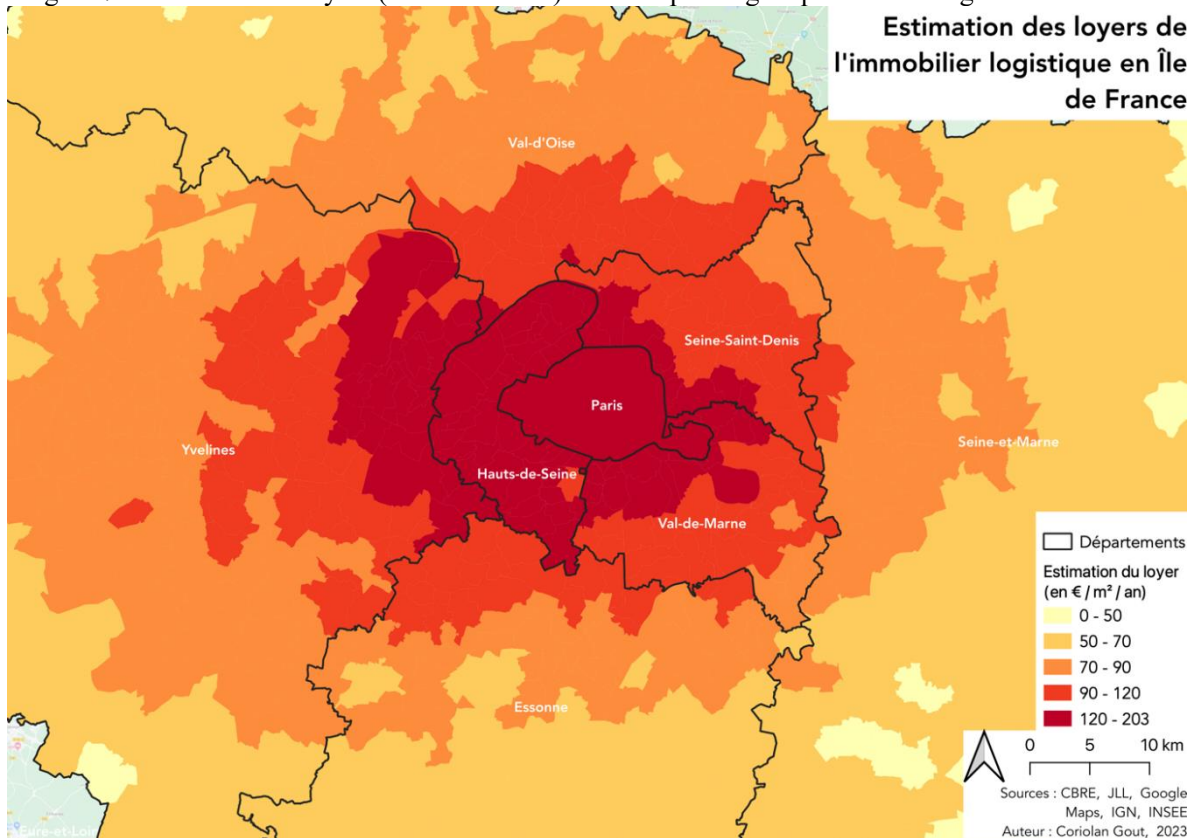
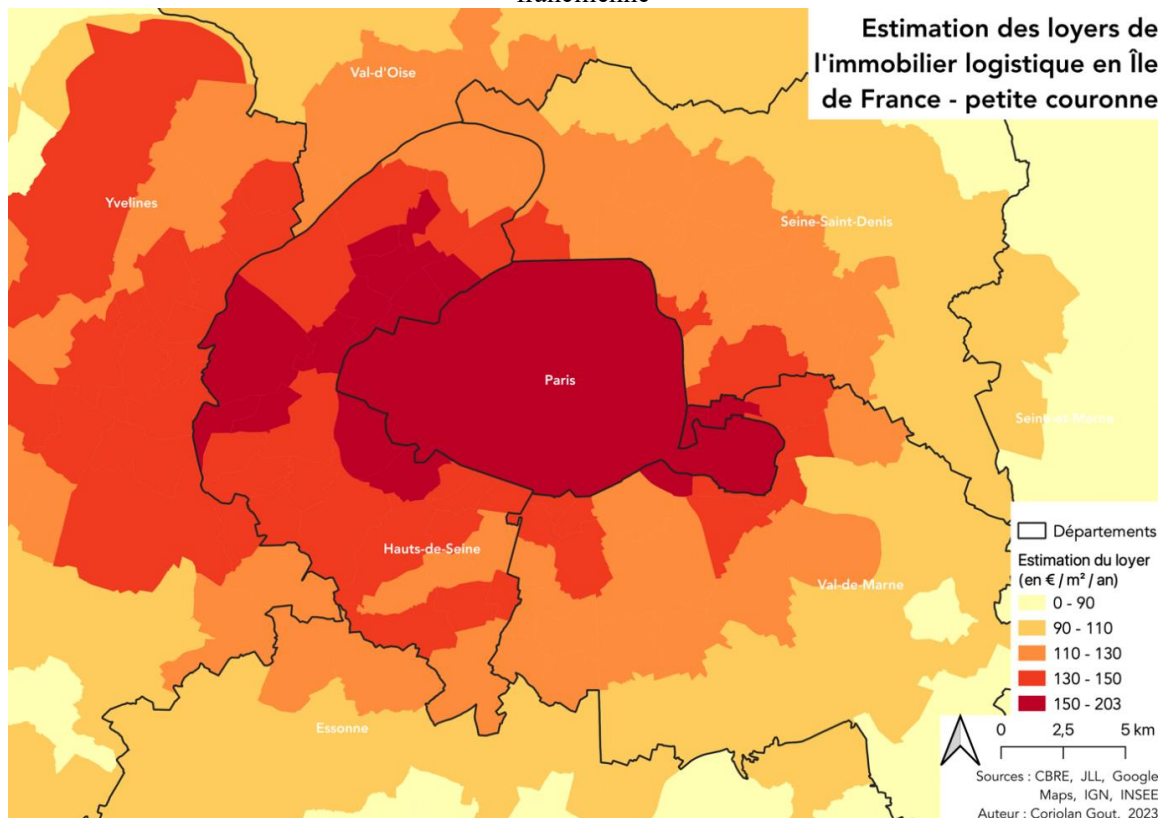


Figure 8. Estimation des loyers (en euros/m²/an) des entrepôts logistiques dans la petite couronne francilienne



Dès lors, on peut calculer des statistiques simples.

Tableau 4. Estimation des loyers (en euros/m²/an) des entrepôts logistiques par CBRE dans les départements de la région Ile-de-France (source : site internet de CBRE).

Zone	Moyenne	Médiane	Minimum	Maximum
<i>Ile-de-France</i>	69	63	19	202
Seine et Marne	50	48	19	92
Yvelines	80	73	45	150
Essonne	65	62	36	114
Hauts de Seine	145	143	118	202
Seine-Saint-Denis	110	111	83	136
Val de Marne	114	116	70	168
Val d'Oise	75	72	41	127

	Produits neufs		Produits 2nde main	
	Min	Max	Min	Max
Seine et Marne	80 €	145 €	70 €	110 €
Yvelines	95 €	145 €	70 €	120 €
Essonne	95 €	145 €	80 €	100 €
Hauts-de-Seine	140 €	250 €	110 €	185 €
Seine-saint-Denis	80 €	180 €	80 €	180 €
Val de Marne	95 €	250 €	85 €	185 €
Val d'Oise	80 €	140 €	70 €	120 €

Mis à jour le 09/05/2023

On observe que le modèle sous-estime assez largement le loyer pour tous les départements d'Île de France, en particulier ceux de la grande couronne. Le loyer étant expliqué en grande partie par des distances à des infrastructures (A86, Francilienne, Orly), les communes les plus excentrées de la zone d'étude se retrouvent mécaniquement avec des estimations très faibles.

Tout comme pour la distribution spatiale des entrepôts, la distribution spatiale des loyers semble suivre deux logiques :

- Une logique Ouest-Est. Les loyers sont considérablement plus élevés dans l'ouest parisien, et ce sur une plus grande distance qu'à l'Est. On trouve des loyers supérieurs à 120€ / m² jusqu'à 22 km à l'ouest du centre de Paris. Ce chiffre est de 12 km à l'Est.
- Une logique concentrique. Il semble exister un gradient centre-périphérie du loyer, ce dernier diminuant lorsqu'on s'éloigne du centre de l'agglomération parisienne.

1.7 Limites

Les résultats et analyses présentés dans cette note doivent être pris avec précaution. L'intention n'est pas de proposer une lecture exhaustive et finie de la distribution spatiale des entrepôts, ni de leurs loyers.

Pour des raisons de simplification de l'extraction et du traitement des données, nous avons choisi de limiter notre zone d'étude à l'Ile-de-France. Pour étudier Paris plus finement, il aurait été possible de limiter la zone d'étude à l'aire urbaine de Paris. Cette approche aurait été plus fonctionnelle, dans le sens où elle aurait dépendu de facteurs observables et non arbitraires, contrairement au découpage régional. Cependant, cela n'aurait pas permis d'ajouter un nombre significatif d'entrepôts à l'échantillon (moins de dix) et aurait largement rallongé toutes les démarches. Il aurait fallu que nous téléchargeions les données sur le logement, la démographie, l'immobilier, l'emploi, pour six régions plutôt que pour une seule.

La localisation des entrepôts est souvent imprécise. Comme indiqué précédemment, tous les entrepôts issus du site de CBRE ont été localisés au niveau de la mairie ou de l'hôtel de ville de la commune dans laquelle ils se situent. De plus, le géocodage des adresses des entrepôts issus du site de JLL peut être également imprécis. Avec la localisation des entrepôts, ce sont plusieurs variables explicatives qui se retrouvent touchées. Les distances à vol d'oiseau peuvent être erronées de plusieurs kilomètres. Nous n'avons pas réussi à calculer les distances-temps. Nous comptions utiliser le package *dodgr*, qui permet d'utiliser un réseau Openstreetmap pour calculer les distances-temps entre plusieurs points, mais il semble que le réseau soit trop vaste, trop lourd pour que cela puisse fonctionner.

Il aurait été préférable d'utiliser des distances-temps plutôt que des distances à vol d'oiseau. Le temps de transport et l'accessibilité sont des éléments capitaux dans le transport de marchandises. En utilisant des distances à vol d'oiseau, le modèle renforce une logique concentrique et uniforme de la distribution spatiale des loyers et efface dans le même temps des lieux singuliers, des nœuds logistiques particulièrement accessibles.

Le découpage interne choisi pour créer le modèle et effectuer des estimations est celui de la commune. Ce type de découpage est tout à fait utilisable, à condition d'utiliser des ratios comme variables (loyer/m², densité de population...). Cependant, il varie très fortement en taille et en forme. Cela peut être un frein à la comparaison.

Les résultats présentés sont des estimations. Le modèle n'est pas très robuste (R^2 moyen, bonnes p-values) et l'échantillon assez faible. Les cartes créées à partir du modèle donnent un ordre d'idée des loyers de l'immobilier logistique à l'échelle de l'aire métropolitaine mais ne peuvent pas être utilisées pour produire des analyses à plus petite échelle.

Cette étude pourrait être renforcée par l'utilisation d'un échantillon plus varié, conséquent et précis. La solution la plus immédiate serait d'extraire les données d'autres sites d'annonces, mais il n'en existe pas d'aussi prolifiques que CBRE et JLL. Il serait aussi possible de constituer une base de données plus fournie au fil du temps, en effectuant des extractions régulières sur les deux sites, en particulier sur celui de JLL. Cela permettrait en prime d'observer l'évolution des annonces publiées, et ainsi du marché locatif de l'immobilier logistique en Ile-de-France.

2. Le cas de New York

2.1 Présentation de la méthode d'extraction des données

Pour le cas new-yorkais, la méthode d'extraction des données sur les loyers des entrepôts s'est révélée être plus ardue que pour le cas parisien.

Nous avons identifié un unique site répertoriant un nombre conséquent d'annonces : Loopnet.com. Cependant, contrairement aux sites de JLL et CBRE, il n'est pas possible d'utiliser des méthodes de *web-scraping* (en utilisant le package *rvest*) pour extraire des données, difficulté liée a priori à des mesures de protection prises par les administrateurs du site. Face à cette difficulté, nous avons fait le choix d'extraire les informations manuellement, abandonnant l'idée de construire un script et ainsi de proposer une méthode systématique. Pour la zone d'étude de New York, le choix a été fait de choisir la CMSA (Consolidated Metropolitan Statistical Area) de New York comme base géographique. Le découpage territorial de la zone d'étude est celui du Zipcode (équivalent américain du code postal qui sert également d'unité statistique).

2.2 Extraction manuelle des annonces

Nous avons extrait les annonces en reportant leurs informations dans un tableur Excel.

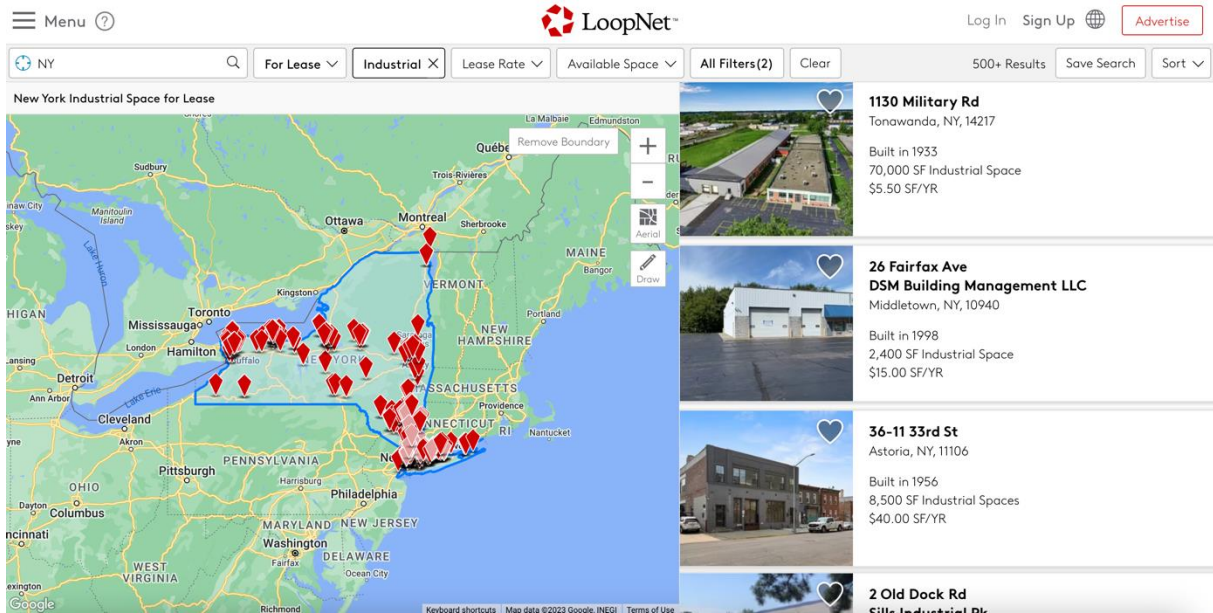
Tableau 5. Extrait de la base de données constituée pour la région de New York (adresse, loyer, surface).

Address	Price	Surface	Date
777 E 91st Brooklyn NY 11236	21,7	2600	1959
295-297 Georgia Ave Brooklyn NY 11207	23	20000	1916
1 Martin Pl Port Chester NY 10573	19	6600	1960
5326 37th St, Long Island City NY	30	5400	1959
677 Nepperhan Ave Yonkers NY 10703	16	14000	1960
110 Emjay Blvd Brentwood NY 11717	15	56700	1961
1943 Pitkin Ave Brooklyn NY 11207	14,74	3500	1930
22-11 38th Ave Long Island City NY 11101	34	57900	2015
280 Adams Blvd Farmingdale NY 11735	18,95	20000	1973
1369 Blondell Ave Bronx NY 10461	24	22000	1956
50 Wireless Blvd Hauppauge NY 11788	22,5	15000	1985
147-151 43rd St Brooklyn Ny 11232	30	5000	1989
33-33 11th St Long Island City NY 11101	33	7000	1930
185 Adams Ave Hauppauge NY 11788	14,5	51000	1982
315 E Third St Mount Vernon NY 10553	14,5	22250	1945

À partir des adresses et du package *ggmap*, nous avons pu obtenir les coordonnées des entrepôts et ainsi transformer les annonces en données géographiques.

Extraire les annonces manuellement nous a permis de les consulter plus précisément et de nous assurer de n'inclure que des entrepôts dans l'échantillon. Nous avons paramétré le moteur de recherche de Loopnet de sorte à ce qu'il présente toutes les annonces correspondant à la catégorie « Industrial » et au mot clé « warehouse ».

Figure 9. Capture d'écran du site d'annonce immobilière LoopNet et des informations disponibles.




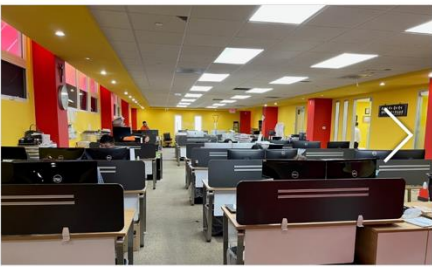
Nous avons consulté toutes les annonces et retenu celles qui correspondaient à ces critères :

- Une surface minimale de 2 200 pieds carrés (environ 200 m²)
- Présence de quais de chargements
- La surface doit être majoritairement plane et ouverte (pas de cloison)
- Mention dans l'annonce que le bâtiment peut être utilisé pour des activités d'entreposage


Figure 10. Capture d'écran d'une annonce immobilière sur LoopNet et des informations disponibles.

10915 178th St
51,324 SF of Industrial Space Available in Jamaica, NY

SPACE	SIZE	TERM	RENTAL RATE	SPACE USE	CONDITION	AVAILABLE
1st Floor	51,324 SF	Negotiable	\$22.00 /SF/YR	Industrial	-	July 01, 2024


Free standing Class A industrial space with high ceilings and very nice offices Located +/- 15 minutes from JFK airport 2,418 pallet positions, heavy power, new roof membrane, LED lighting, Floor drains WAREHOUSE PALLET RACKING AND OFFICE FURNITURE AVAILABLE There is 4,500/sf of 2nd floor office space and +/-1,509/sf 1st floor office space. WAREHOUSE PALLET RACKING AND OFFICE FURNITURE AVAILABLE 1st floor: 46,424sf and 2nd floor offices (4,500sf)



SCHACKER REALTY
www.schackerrealty.com
435.293.3100

Call

Message



28-10 Whitestone Expy - Colleg...
240,268 SF Available

À titre d'exemple, nous aurions retenu cette annonce. La surface est conséquente, on observe des quais de chargement et il est clairement indiqué que le bâtiment peut être utilisé pour des activités d'entreposage.

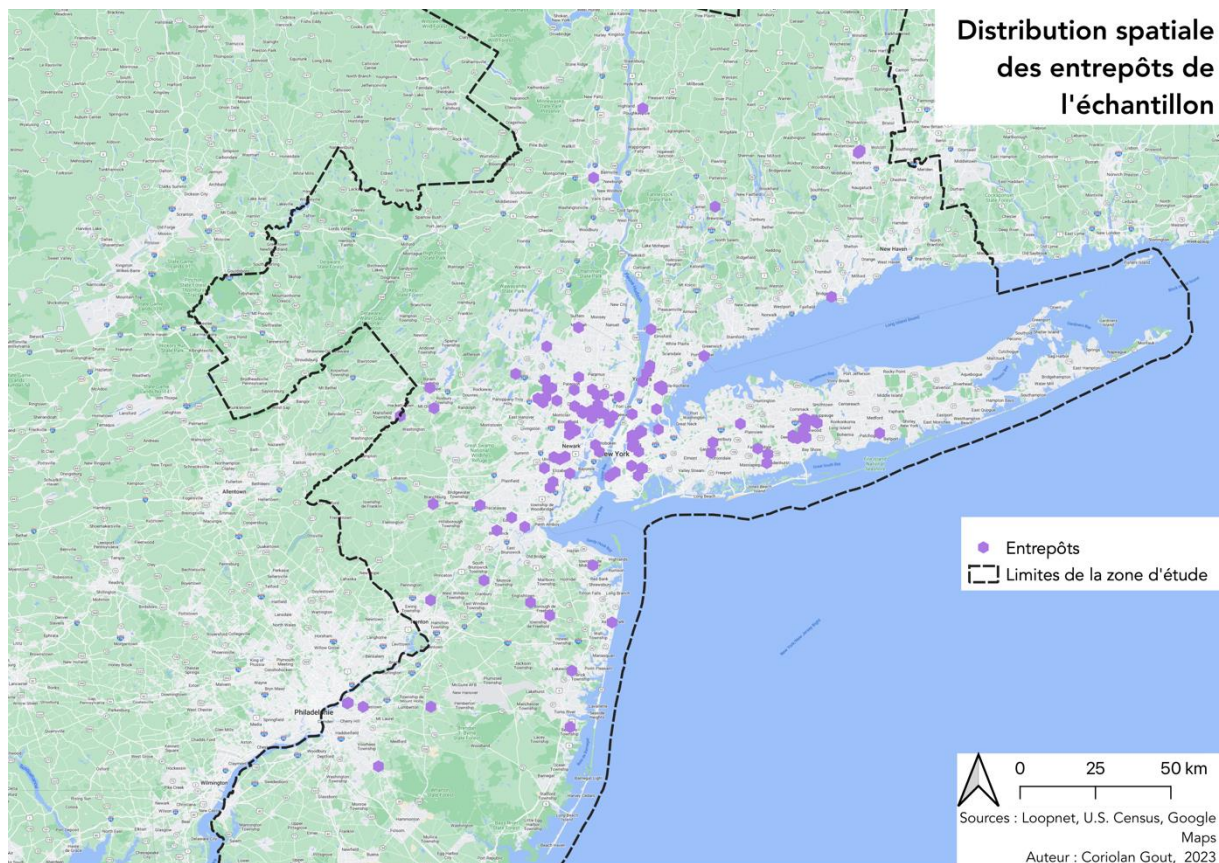
Ici, le lot est loué avec 4 500 pieds carrés de bureaux (420 m² environ). Dans ce genre de cas, nous avons gardé la surface totale du lot, bureau compris. Si plusieurs lots sont présentés sur une même annonce, avec le même loyer, nous additionnons leurs surfaces et les intégrons au tableur comme une seule annonce.

Présentation de l'échantillon

Notre échantillon a été constitué entre le 13 et le 15 septembre 2023. Il est composé de 111 annonces. Parmi elles, 48 sont situées dans l'Etat de New York, 59 dans l'Etat du New Jersey et 4 dans l'Etat du Connecticut. Cette différence de représentation entre Etats peut s'expliquer par deux raisons. L'Etat du Connecticut ne présente que peu d'annonces valides (d'après notre méthode de sélection) dans l'aire de la CMSA de New York. Pour les Etats de New York et du New Jersey, l'idée était de conserver une représentation proche du réel. Dans le Zipcode Business Pattern (ZBP) (enquête fédérale sur les établissements économiques) de 2018, il est indiqué que 560 entreprises ayant pour activité principale « *Warehousing and storage* »² se trouvaient dans l'Etat de New York. Ce chiffre était de 717 pour le New-Jersey, ce qui équivalait à un rapport de 128 pour 100³. Dans notre échantillon, le rapport est de 122 pour 100.

Voici une carte de la distribution spatiale des entrepôts de l'échantillon.

Figure 11. Distribution spatiale des entrepôts logistiques de l'échantillon de l'étude dans la CMSA de New York (octobre 2023)

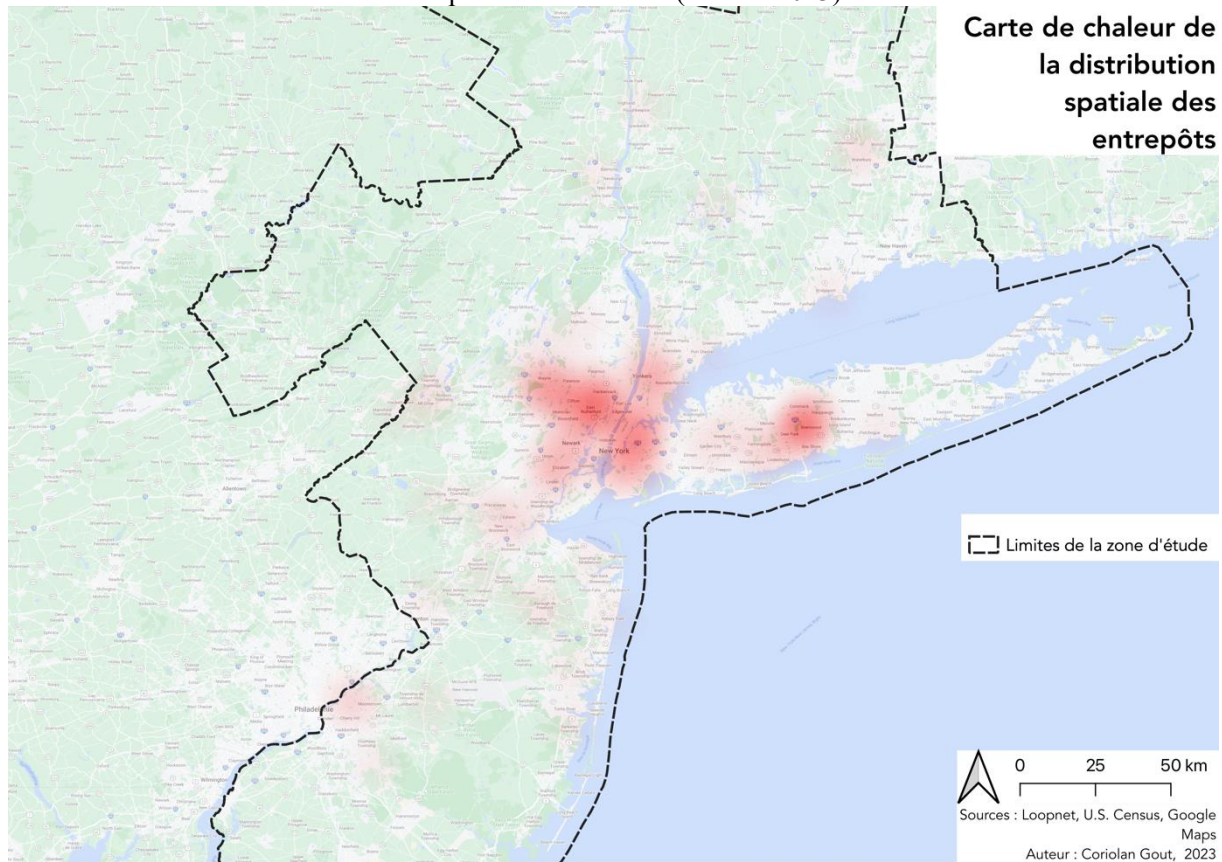


² Intitulé du code NAICS 493.

³ S'il y a 100 entreprises dans l'Etat de New York, il y en aurait 128 dans le New Jersey.

Pour plus de clarté, voici une carte de chaleur.

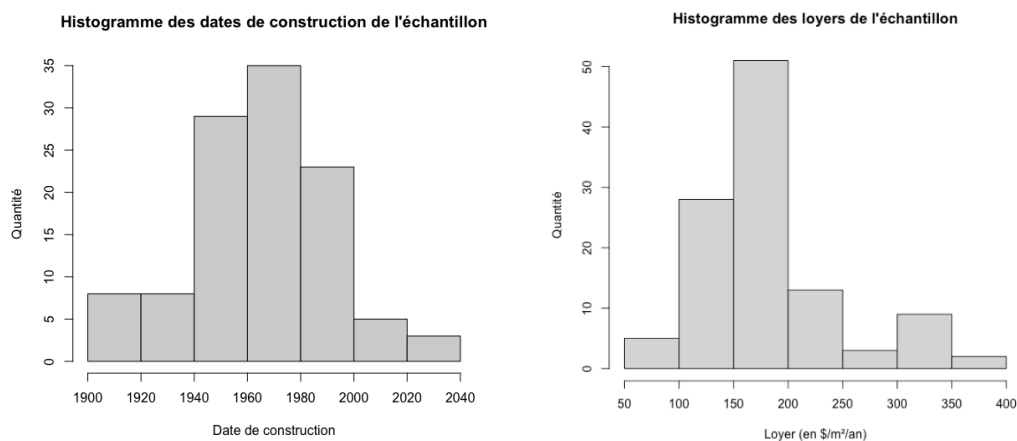
Figure 12. Distribution spatiale des entrepôts logistiques de l'échantillon de l'étude dans la CMSA de New York par carte de chaleur (octobre 2023)

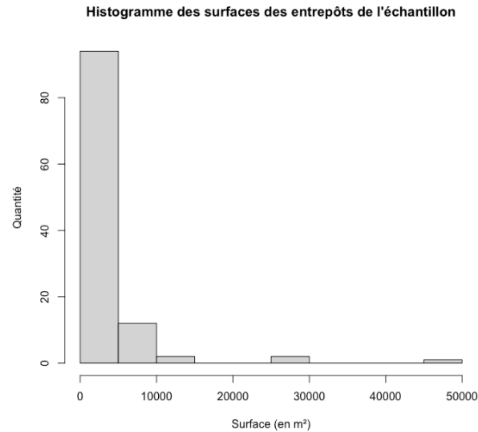


On observe que les entrepôts de l'échantillon se concentrent majoritairement dans quatre zones :

- Dans l'Etat du New Jersey, à l'ouest de l'île de Manhattan et à proximité de Newark.
- Dans le quartier de Brooklyn
- Au nord du quartier du Bronx
- Au centre de l'île de Long Island

Figure 13. Histogrammes relatifs au jeu de données pour le cas de la CMSA de New York.





2.3 Création du modèle

Nous avons créé un jeu de données de plus d'une vingtaine de variables, pour tenter d'expliquer le loyer annuel au m² des entrepôts dans la zone d'étude. Les variables sont observables à l'échelle du Zipcode. Voici les différentes variables retenues et leur source :

- Densité de population (U.S. Census)
- Nombre d'établissements par secteur (Zipcode Business Pattern)
- Distance à vol d'oiseau à différentes infrastructures de transport (calcul à partir d'entités digitalisées)

Voici le modèle que nous avons gardé, après avoir testé les différentes variables du jeu de données :

Loyer annuel au m²) ~ Distance à vol d'oiseau au centre de New-York + Distance à vol d'oiseau à l'aéroport de Newark + Distance à vol d'oiseau à l'aéroport de JFK + Distance à vol d'oiseau du port le plus proche + log(densité de population) + Nombre d'établissements ayant pour activité principale « Transportation and Warehousing »

Pour constituer la liste des ports, nous avons utilisé la liste des containers terminal fournie par la Port Authority of New York and New Jersey :

- Port Newark Container Terminal
- Maher Terminals
- APM Terminals
- GCT New York Terminal
- GCT Bayonne Terminal
- Red Hook Container Terminal

La variable Nombre d'établissements ayant pour activité principale « Transportation and Warehousing » correspond aux établissements du Zipcode Business Pattern identifiés par le code NAICS 48-49.

Le tableau suivant résume le résultat du modèle.

Tableau 6. Résultats statistiques du modèle linéaire et significativité des relations aux variables.

Variable	Coefficient	Erreur standard	t-value	p-value	Significativité
Intercept	1.840e+02	3.813e+01	4.826	4.80e-06	***
Distance à vol d'oiseau au centre de New-York	-2.770e-03	7.938e-04	-3.490	0.00071	***
Distance à vol d'oiseau à l'aéroport de Newark	1.047e-02	1.984e-03	5.280	7.13e-07	***
Distance à vol d'oiseau à l'aéroport de JFK	2.942e-03	8.815e-04	3.337	0.00118	**
Distance à vol d'oiseau du port le plus proche	-1.136e-02	2.220e-03	-5.119	1.42e-06	***
log(densité de population)	-1.081e+01	4.924e+00	-2.195	0.03042	*
Nombre d'établissements ayant pour activité principale « Transportation and Warehousing »	4.023e+00	1.316e+00	3.057	0.00284	**

Les différentes variables du modèle sont toutes significatives à plus de 95 %. Au vu de la faible taille de l'échantillon ce résultat est surprenant, particulièrement si on le compare à ceux des études sur Paris et New-York.

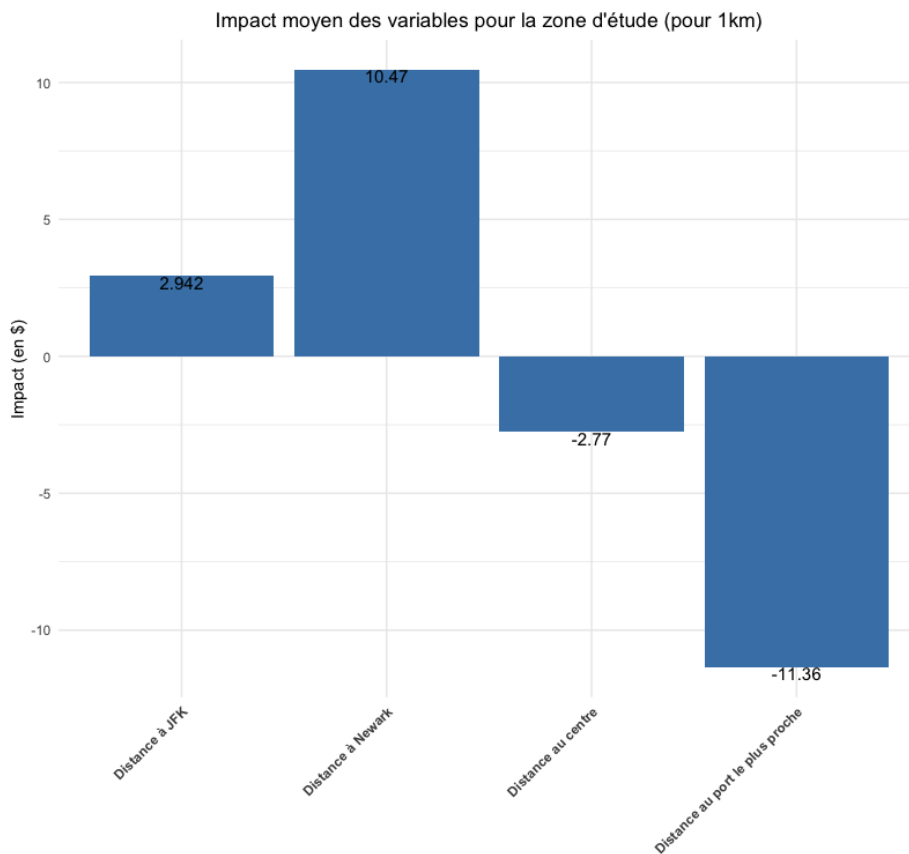
Résumons les différentes relations entre le loyer et les variables explicatives.

Tableau 7. Résumé des relations entre le loyer et les variables.

Variable	Relation	Explication
Distance à vol d'oiseau au centre de New-York	Négative	Le loyer diminue plus on s'éloigne du centre de New-York. Il augmente quand on s'en approche.
Distance à vol d'oiseau à l'aéroport de Newark	Positive	Le loyer augmente plus on s'éloigne de l'aéroport de Newark. Il diminue quand on s'en rapproche.
Distance à vol d'oiseau à l'aéroport de JFK	Positive	Le loyer augmente plus on s'éloigne de l'Aéroport de JFK. Il diminue quand on s'en rapproche.
Distance à vol d'oiseau du port le plus proche	Négative	Le loyer diminue plus on s'éloigne du port le plus proche. Il augmente quand on s'en rapproche.
log(densité de population)	Négative	Le loyer diminue lorsque le logarithme de la densité de population augmente.

Nombre d'établissements ayant pour activité principale « Transportation and Warehousing »	Positive	Le loyer augmente lorsque le nombre d'établissements ayant pour activité principale « Transportation and Warehousing » augmente.
--	----------	--

Figure 14. Histogramme de l'impact de chaque variable sur le loyer mensuel



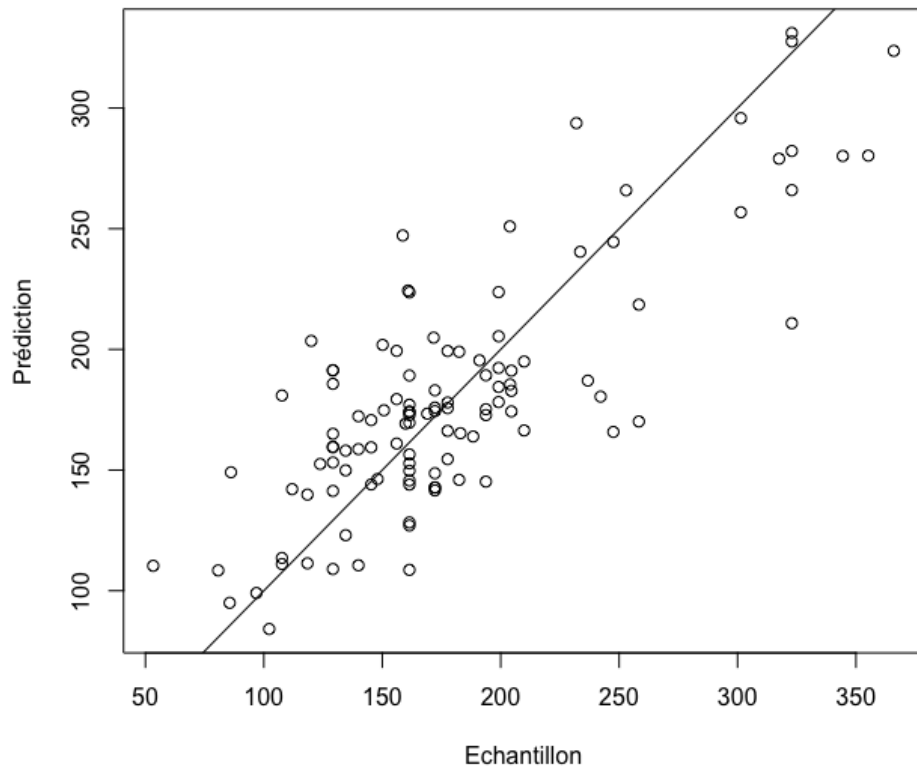
On constate que la proximité aux deux aéroports a un impact négatif sur le loyer. Les loyers au m² augmentent de plus de dix dollars chaque fois que l'on s'éloigne d'1 km de Newark. A contrario, la proximité au centre de New-York (Broadway – 42^e rue) et la proximité au port le plus proche ont un impact positif sur le loyer).

Dans ce modèle, la relation entre loyer et logarithme de la densité de population est négative. Lorsqu'une variable est transformée avec le logarithme, des relations non linéaires peuvent inverser la direction de l'association entre les variables, conduisant à des coefficients de régression qui peuvent être positifs dans l'espace d'origine, mais négatifs dans l'espace logarithmique. Ainsi, il n'est pas possible d'affirmer qu'il y a une relation négative entre loyer et densité de population.

Le R² de ce modèle est de 0,65. Il peut se traduire ainsi : le résultat du modèle correspond à la réalité observée dans l'échantillon dans 65% des cas. Au vu de la taille de l'échantillon, le R² est satisfaisant

Le nuage de points suivant présente la relation entre l'échantillon et la prédiction. Plus les points se situent proches de la ligne (x=y), plus la prédiction est bonne et le R² élevé.

Figure 15. Coefficient de détermination visant à déterminer la solidité du modèle.



2.4 Estimations à partir du modèle

En appliquant le modèle à l'échelle de la zone d'étude, on obtient les cartes suivantes.

Figure 16. Estimation des loyers (en dollars/m²/an) des entrepôts logistiques dans la CMSA de New York.

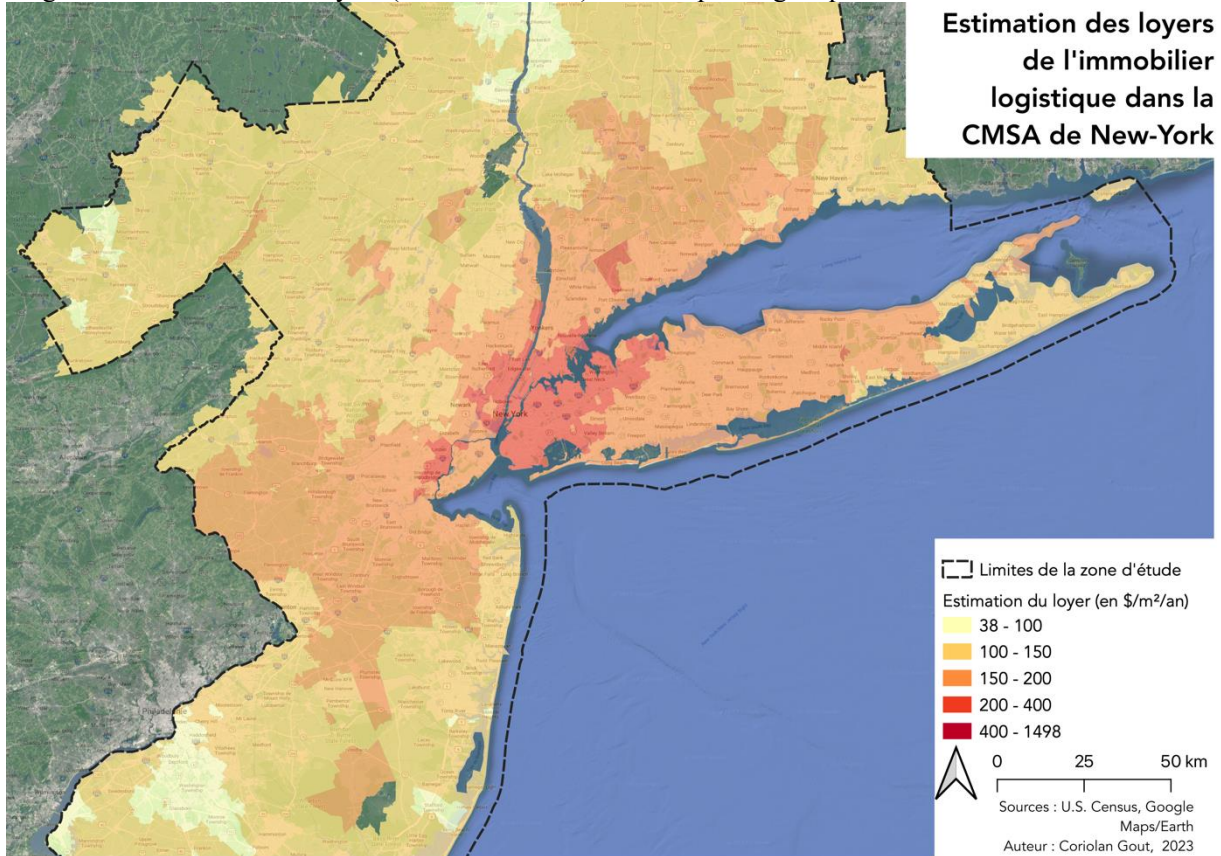
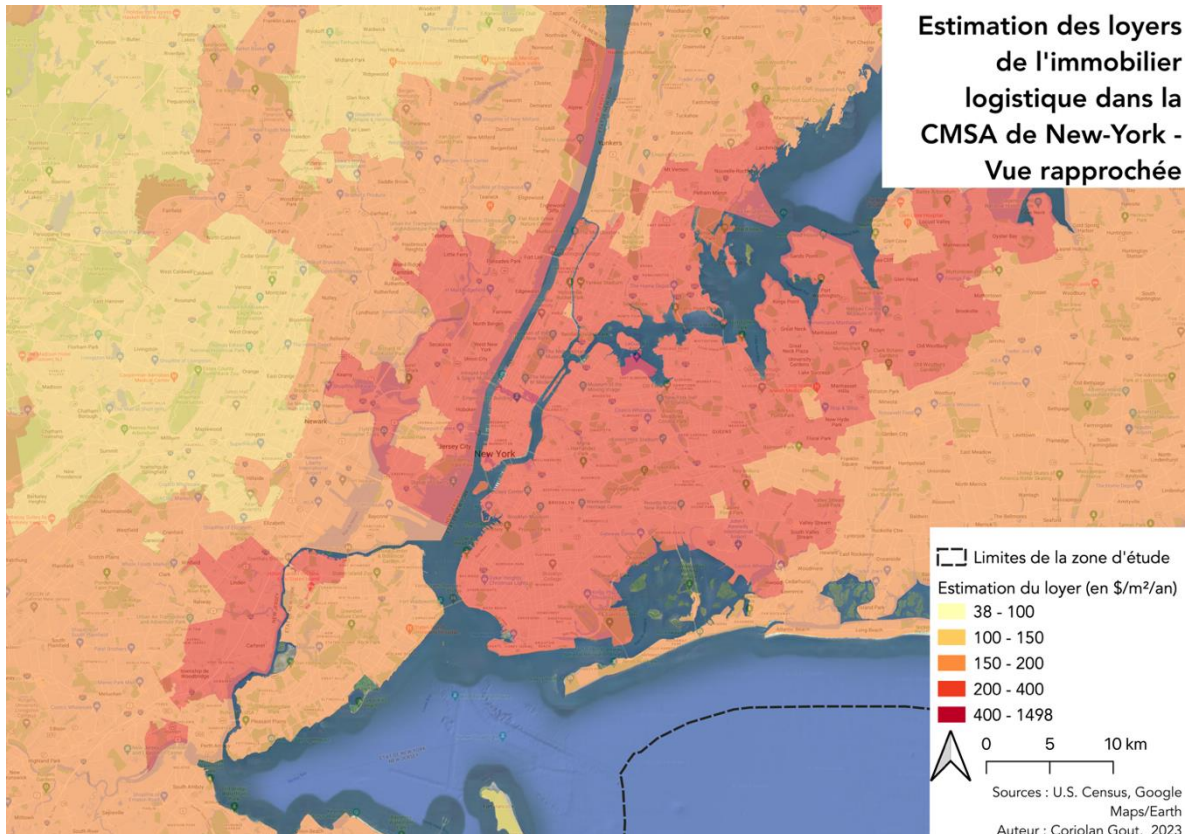


Figure 17. Estimation des loyers (en dollars/m²/an) des entrepôts logistiques dans la partie centrale de la CMSA de New York.



La distribution spatiale des loyers semble suivre ici deux logiques :

- Une logique centre-périphérie. Les loyers sont particulièrement élevés dans le centre de la métropole. Contrairement à Paris, ils ne forment pas de structures concentriques autour du centre.
- Une logique radiale. Les loyers allant de 150 \$ à 200 \$ suivent trois axes :
 - Un axe nord-est longeant la côte Atlantique et partant du Bronx
 - Long Island
 - Un axe sud-ouest

2.5 Limites

Les localisations des entrepôts sont très précises, car toutes à l'adresse. Les calculs de distance ont pu être effectués directement depuis les entrepôts, sans avoir à utiliser des centroïdes. Le modèle est particulièrement robuste. Les variables sont toutes significatives et le R^2 est élevé pour un échantillon si réduit.

Les résultats et analyses présentés dans cette note doivent être pris avec précaution. L'intention n'est pas de proposer une lecture exhaustive et finie de la distribution spatiale des entrepôts, ni de leurs loyers.

La méthode d'extraction et de sélection des annonces immobilières utilisée ici n'est pas systématique. Bien que reposant sur des critères précis, elle n'est pas aussi extensive que pour Paris ou Tokyo. Comme pour le cas de Paris, les distances aux infrastructures sont des distances à vol d'oiseau.

Table des illustrations

Figure 1. Captures d'écran de deux sites d'annonce immobilière et des informations disponibles	4
Figure 2. Extrait de la base de données constituée pour la région parisienne (adresse, loyer, coordonnées géographiques)	5
Figure 3. Distribution spatiale des entrepôts logistiques de l'échantillon de l'étude en Ile-de-France (septembre 2023).....	7
Figure 4. Distribution spatiale en carte de chaleur des entrepôts logistiques de l'échantillon de l'étude en Ile-de-France (septembre 2023).....	7
Figure 5. Schéma d'illustration du principe de régression linéaire.	8
Tableau 1. Résultats statistiques du modèle linéaire et significativité des relations aux variables.	9
Tableau 2. Résumé des relations entre le loyer et les variables.....	10
Figure 6. Coefficient de détermination visant à déterminer la solidité du modèle.....	11
Tableau 3. Estimation des loyers (en euros/m ² /an) des entrepôts logistiques dans la région Ile-de-France.	11
Figure 7. Estimation des loyers (en euros/m ² /an) des entrepôts logistiques dans la région Ile-de-France.	12
Figure 8. Estimation des loyers (en euros/m ² /an) des entrepôts logistiques dans la petite couronne francilienne.....	12
Tableau 4. Estimation des loyers (en euros/m ² /an) des entrepôts logistiques par CBRE dans les départements de la région Ile-de-France (source : site internet de CBRE).	13
Tableau 5. Extrait de la base de données constituée pour la région de New York (adresse, loyer, surface).	15
Figure 9. Capture d'écran du site d'annonce immobilière LoopNet et des informations disponibles. .	16
Figure 10. Capture d'écran d'une annonce immobilière sur LoopNet et des informations disponibles.	16
Figure 11. Distribution spatiale des entrepôts logistiques de l'échantillon de l'étude dans la CMSA de New York (octobre 2023).....	17
Figure 12. Distribution spatiale des entrepôts logistiques de l'échantillon de l'étude dans la CMSA de New York par carte de chaleur (octobre 2023).....	18
Figure 13. Histogrammes relatifs au jeu de données pour le cas de la CMSA de New York.	18
Tableau 6. Résultats statistiques du modèle linéaire et significativité des relations aux variables.	20
Tableau 7. Résumé des relations entre le loyer et les variables.....	20
Figure 14. Histogramme de l'impact de chaque variable sur le loyer mensuel.....	21
Figure 15. Coefficient de détermination visant à déterminer la solidité du modèle.....	22
Figure 16. Estimation des loyers (en dollars/m ² /an) des entrepôts logistiques dans la CMSA de New York.	23
Figure 17. Estimation des loyers (en dollars/m ² /an) des entrepôts logistiques dans la partie centrale de la CMSA de New York.	23