



# RAPPORT OFFICIEL

ÉQUIPE JAUNE : COMPÉTITION I3C

**FÉLIX POIRIER**

*felix.poirier@hec.ca*

**ALEXANDRO TOMASSI**

*alexandro.tomassi@hec.ca*

**ISLAM ABDELATIF**

*islam.abdelatif@hec.ca*

# TABLE DES MATIÈRES

<b>Table des Matières</b>	<b>2</b>
<b>Sommaire Exécutif</b>	<b>4</b>
<b>Introduction</b>	<b>5</b>
<b>Méthodologie</b>	<b>6</b>
Vue d'ensemble	6
La catégorie géographique	6
La catégorie locale	8
La catégorie macroéconomique	8
Table des variables utilisées	9
Le modèle	10
<b>Analyse des résultats</b>	<b>12</b>
L'entraînement	12
Le test	13
La prédiction	13
Le cas de Honolulu	14
<b>Recommandations</b>	<b>15</b>
Long Island & New York	15
Sacramento & San Francisco	16
Kansas City	16
Les pires villes	16
<b>ESG</b>	<b>18</b>
<b>Conclusion</b>	<b>20</b>
<b>Annexe</b>	<b>21</b>
Annexe 1	21
Annexe 2	21
Annexe 3	22
Annexe 4	22
Annexe 5	22
Annexe 6	23
Annexe 7	23
<b>Bibliographie</b>	<b>24</b>





## SOMMAIRE EXÉCUTIF

Ce rapport vise à répondre à la question, comment identifier les villes et les lieux de demain? Dans un contexte d'investissement immobilier. Nous avons utilisé une méthodologie basée sur la prédiction de la croissance des flux de trésorerie nets des propriétés industrielles dans une région, en utilisant des données géographiques, locales et macroéconomiques.

Les résultats montrent que les villes de New York, Long Island, Sacramento San Francisco et Kansas City sont les plus performantes en termes de croissance des flux de trésorerie nets. Le modèle géographique utilisé a démontré sa supériorité par rapport à un modèle de contrôle, en prédisant une tendance de croissance plutôt que la simple moyenne. Les données géographiques ont joué un rôle clé dans l'identification de ces tendances, en montrant un lien entre la densité de population, la présence de ports et de bornes de recharge électrique avec la performance économique des villes.

En se basant sur les résultats obtenus, nous recommandons de considérer les villes mentionnée ci-haut pour les investissements immobiliers industriels. San Francisco est particulièrement intéressante, car elle présente un potentiel de croissance de 8% supérieur aux anticipations du marché. Il est important de noter que ces recommandations sont à titre indicatif et ne doivent pas être utilisées uniquement sur la base des résultats de ce modèle. Il est nécessaire de prendre en compte d'autres facteurs tels que les politiques réglementaires, les tendances de l'industrie et les facteurs macroéconomiques avant de prendre une décision d'investissement. Il est également important de surveiller les tendances à long terme pour évaluer la viabilité à long terme de l'investissement.

En conclusion, ce rapport a examiné une méthode pour repérer les lieux porteurs d'investissement dans le secteur immobilier industriel en utilisant la prédiction de la croissance future des flux monétaires nets des actifs industriels d'une région. Les données utilisées incluent des informations géographiques, locales et macroéconomiques, qui ont été combinées pour créer un modèle de prédiction. Les résultats montrent que le modèle géographique est capable de détecter une tendance entre la variable dépendante et les variables indépendantes, offrant une meilleure prédiction que le modèle de contrôle. Les recommandations données sont basées sur les résultats obtenus par le modèle et devraient être considérées avec prudence en combinaison avec d'autres facteurs tels que les politiques réglementaires, les tendances de l'industrie et les facteurs macroéconomiques avant de prendre une décision d'investissement. Il est important de surveiller les tendances à long terme pour évaluer la viabilité à long terme de l'investissement.

---

# INTRODUCTION

Ce présent rapport tente de répondre à la question : comment identifier les villes et les lieux de demain? Cette question laisse sans doute place à un grand niveau de créativité, une liberté dont nous avons amplement profité. Tel Ivanhoé Cambridge, l'équipe fut abasourdie par l'engouement que les centres de distribution, propulsés par une vague de consommation en ligne, pouvaient générer. Ces centres sont dépendants d'une infrastructure généralement déjà bien établie qui permet entre autres d'avoir accès à de la main-d'œuvre ou à des nœuds du réseau de transport.

Passionnés par la géographie, nous avons créé des procédures permettant la simplification de ces données. Les procédures novatrices présentées dans ce rapport ne prétendent pas être assez soignées pour pouvoir être utilisées dans la pratique rigoureuse de l'investissement. Somme toute, elles tentent de démontrer qu'elles sont viables, et qu'éventuellement une approche qui s'en inspire pourrait porter fruit pour Ivanhoé Cambridge.

Nous sommes conscients que la reproductibilité de nos résultats est une considération très importante dans le contexte de l'évaluation des équipes. De ce fait, pour favoriser votre utilisation de nos programmes, nous avons créé une table avec une description courte de la raison d'être des fichiers les plus utiles et de la structure du projet sur DataBricks. (voir Annexe 7)

# MÉTHODOLOGIE

## Vue d'ensemble

L'objectif primaire de ce rapport est d'arriver à une conclusion quant à une méthode pour repérer les lieux porteurs d'investissement dans le secteur immobilier industriel. Pour ce faire, nous tentons de prédire la croissance future des flux monétaires nets des actifs industriels d'une région (net cash-flow growth). Cette variable se démarque puisqu'elle prend en considération non seulement les revenus d'une propriété, mais aussi les multiples coûts. Intégrer ces données sous-tend deux hypothèses: les données fournies par GreenStreet, représentent de manière holistique la région étudiée notamment l'anticipation des marchés et les méthodes de mesure utilisées par GreenStreet sont constantes à travers le temps. Ces hypothèses nous permettent d'arriver à des conclusions utiles.

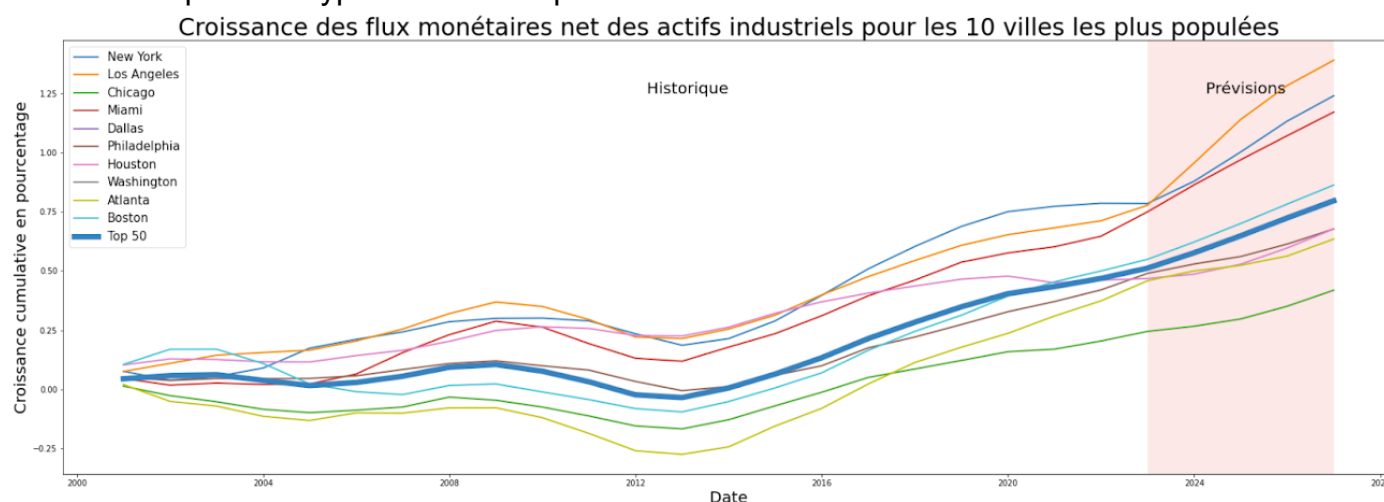


Figure 1 Croissance des flux de capitaux nets d'une propriété selon GreenStreet

Notre objectif est de prédire la croissance de cette métrique pour la prochaine période, dans le contexte de ce rapport, une période représente un an. De ce fait, nous tentons, à l'aide de données accessibles aujourd'hui, de prédire quelle sera la croissance réalisée des flux financiers d'une propriété en un an. Cette métrique émane de plusieurs caractéristiques intéressantes, d'une part les valeurs semblent corrélées, ce qui implique que les marchés américains sont tous influencés par des facteurs macroéconomiques. On observe aussi que les marchés matures des 10 plus grandes villes américaines ont tendance à rester stables, ce qui réduit la probabilité de générer de majeures erreurs lors d'une éventuelle adoption du modèle.

À la suite d'une analyse rigoureuse de toutes les données disponibles, nous avons divisé les données en trois catégories distinctes, soit les données géographiques (géo), les données locales ainsi que les données macroéconomiques (macro).

## La catégorie géographique

Les données géographiques, présentées ci-haut, sont représentées par des matrices. Ce modèle est inédit, et a été créé dans son entièreté par les membres de l'équipe. Ce modèle répertorie des données qui seraient autrement seulement capables d'être organisées dans des fichiers KML, ou geojson, dans des matrices flexibles pouvant être sauvegardées dans des fichiers beaucoup moins volumineux et prêts pour différents modèles d'apprentissage machine. Cette structure de donnée se démarque d'une

approche plus traditionnelle, où l'information serait présentée sous forme de table, puisqu'elle représente l'information de manière physique ce qui permet entre autres, de représenter le réseau infrastructurel d'une région de manière fiable. Prenons Montréal comme exemple, disons que nous nous situons au centre de la ville, à l'intersection de l'autoroute 15 Nord et l'autoroute 40. De notre position, le Port de Montréal ainsi que l'aéroport YUL sont à des distances équivalentes. Si ces distances sont enregistrées dans une table, puisque leur distance est similaire, on pourrait faire l'erreur de croire que ces deux lieux sont proches, quand pourtant en réalité il faut traverser Montréal pour se rendre de l'un à l'autre. Le modèle géographique permet d'avoir une vision plus représentative de cette réalité. Ceci a bien évidemment des répercussions sur les différentes installations industrielles de la région.

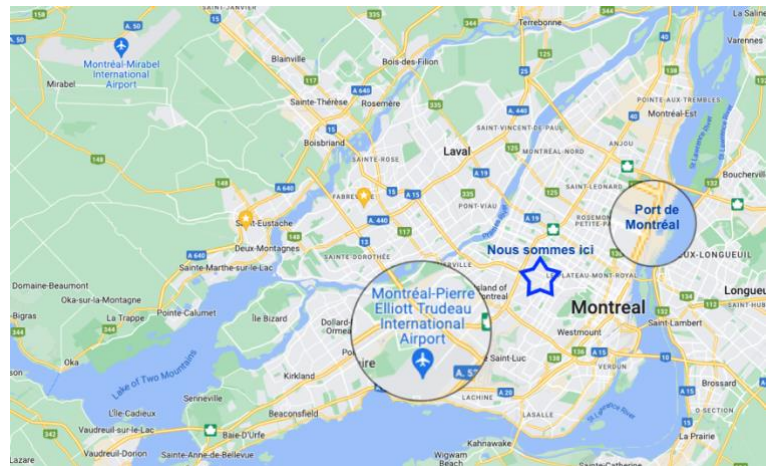


Figure 2 Carte de l'île de Montréal extraite de Google Maps

Pour construire un tel modèle, nous devons tout d'abord bâtir une matrice de référence (appelée xMap). La manière la plus efficace de bâtir un tel modèle est vraisemblablement avec la latitude et la longitude géographique associée à chacun des lieux étudiés. Dès lors nous devons bâtir cette matrice de référence et y répertorier ces deux valeurs quintessentiels au modèle. Plus précisément cette matrice est de trois dimensions, la première dimension représentant l'axe des x, la deuxième l'axe des y, et puis la dernière répertoriant la latitude et la longitude. Les bornes de la matrice doivent être prédéfinies par l'utilisateur, dans le cadre de cette analyse, nous avons choisi le territoire continental américain. Cette matrice est définie par une certaine granularité. Plus la granularité est petite, plus la matrice est grande et donc plus précise. Dans le contexte de cette analyse, nous avons décidé d'utiliser une grande granularité puisque nous cherchons seulement à déterminer si ce modèle est viable.

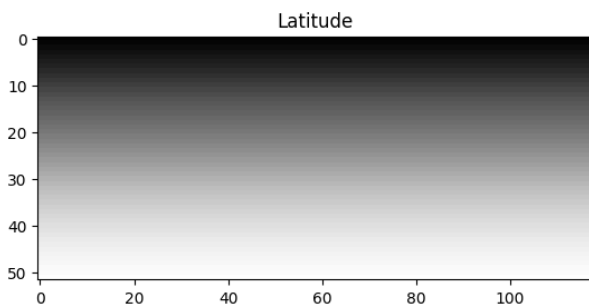


Figure 3 Premier élément de la matrice xMap

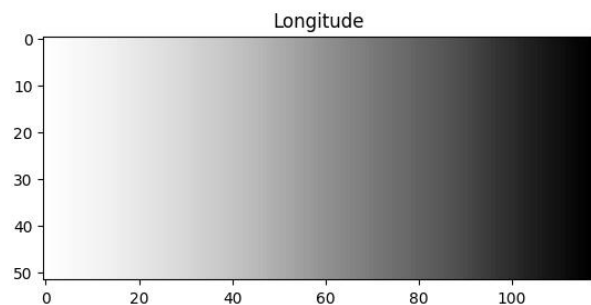


Figure 4 Deuxième élément de la matrice xMap

Après avoir bâti la matrice de référence, nous devons définir une matrice pour chacune des données que nous voulons étudier; il y aura une matrice pour décrire les arrêts de transports en commun, une matrice pour décrire les quais, etc. De manière globale nous avons généré 14 matrices, dont 3 répertoriant les différentes juridictions présente sur notre carte (États-Unis, Canada et Mexique), ainsi que plusieurs données présentées ci-haut. Généralement, lorsque nous transformons des fichiers géographiques complexes, ces fichiers ne contiennent pas qu'un nom et un lieu, très souvent on y retrouve des données plus pertinentes. Certaines stations-service ont des capacités maximales. Généralement nous utiliserons ces données et nous ne ferons qu'une somme de ces valeurs pour chacun des points, autrement nous ne ferons qu'un compte.



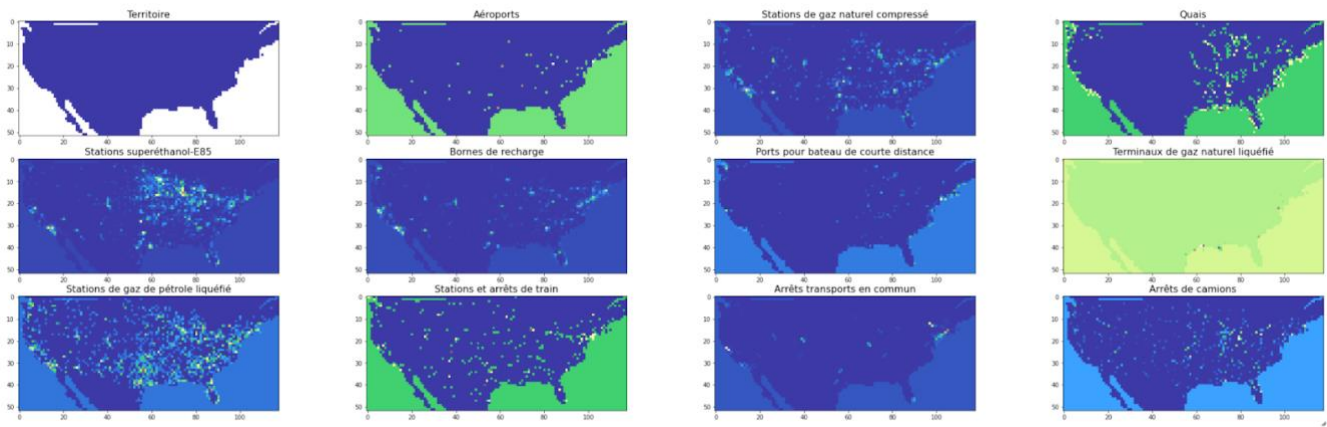


Figure 5 Représentation graphiques des cartes utilisées dans le modèle

Pour intégrer cette structure de données dans un modèle d'apprentissage machine, nous devons nous assurer que le modèle sait où se trouve la ville qu'il étudie. Il y a visiblement plusieurs manières d'y parvenir, notre approche est d'utiliser un "zoom". En d'autres mots, nous faisons une sélection par proximité, on choisit une sous-matrice entourant le point d'intérêt selon un rayon. Ce rayon peut être ajusté très facilement (voir Annexe 1 pour démonstration visuelle). Dans le contexte de cette analyse, nous avons choisi d'utiliser un zoom de 15, soit 15 pixels depuis le centre, ce qui équivaut à un diamètre de 30 pixels, soit environ 1250 km (de gauche à droite), ce qui représente une superficie d'environ 1, 562, 500 km. Après avoir pris en compte toutes ces considérations, prédire à l'aide des matrices géographiques devient très similaire à l'apprentissage à l'aide d'images, une structure de données en pratique très similaire à MNIST.

## La catégorie locale

La catégorie locale, telle notre variable dépendante (*net cash flow growth*) est de nature longitudinale. C'est-à-dire que ces données changent selon la période et l'observation. Toutes les valeurs de catégories locales sont donc uniques. En pratique, les données locales pourraient bel et bien être insérées dans le modèle géographique, nous les gardons dans leur état original puisque sous forme géographique ces données n'apporteraient pas beaucoup plus d'information.

## La catégorie macroéconomique

Comme son nom sous-tend, cette catégorie inclut des données générales ayant un impact sur le territoire américain dans son ensemble. Il est indéniable que certaines données ont un impact national et parfois même global. Ceci permet entre autres d'ajuster pour les facteurs provocants de la corrélation entre les régions observées dans le graphique ci-haut.



### Géographique

Représente les ressources intangibles environnant le lieu étudié



### Locale

Représente les caractéristiques spécifiques au lieu étudié



### Macroéconomique

Représente les caractéristiques globales et nationales ayant des effets symétriques



## Table des variables utilisées

<b>Nom de la variable</b>	<b>Étiquette</b>	<b>Catégorie</b>	<b>Contexte et justification</b>
<i>Aéroports</i>	airports	Géo	Liste de tous les aéroports internationaux sur le territoire américain. Sommée selon le nombre de pistes d'atterrissage. Utile pour le transport rapide de cargaison.
<i>Age médian</i>	age_median	Locale	Âge médian de la population du lieu étudié. Utile pour l'accès à la main-d'œuvre.
<i>Arrêts de camion</i>	truck_stops	Géo	Liste de tous les arrêts de camions sur le territoire américain. Utile pour comprendre l'accès à de l'énergie et du logement pour les conducteurs de poids lourds.
<i>Arrêts de transports en commun</i>	transit_stops	Géo	Liste de tous les arrêts de transport en commun, dont les autobus, trains de banlieue, tramway, etc. Utile pour comprendre la proximité aux bassins de population.
<i>Balance commerciale</i>	netexp	Macro	Imports moins exports pour l'entièreté de États-Unis. Utile pour comprendre l'évolution du secteur manufacturier américain.
<i>Bornes de recharge</i>	elec_stations	Géo	Liste de toutes les bornes de recharge électrique. Utile pour comprendre l'accès à des énergies moins polluantes.
<i>Croissance du revenu du marché par pied disponible</i>	mrevpaf_growth_yoy_credit	Locale	Revenu de ventes aux consommateurs selon par mètre de pieds carrés. Utile pour comprendre la croissance de la lucrativité du secteur de la consommation dans une région.
<i>Distance de l'autoroute</i>	interstate_distance	Locale	Distance moyenne dans une ville à une autoroute majeure. Utile pour comprendre l'accès au vaste réseau automobile américain.
<i>Longueur de l'autoroute</i>	interstate_miles	Locale	Distance moyenne parcourue par les autoroutes présentes dans la ville. Utile pour comprendre la portance d'une ville sur le territoire américain.
<i>Momentum de la valeur des actifs</i>	asset_value_momentum	Locale	Le Momentum est un indicateur communément utilisé en finance pour déterminer l'intensité du mouvement d'un actif. Utile sous l'hypothèse que les marchés immobiliers sont tendanciels.
<i>Population</i>	population_500mi	Locale	Taille de la population dans un rayon de 500 milles à la ville étudiée. Utile pour comprendre la population pouvant être desservie par un centre de distribution.
<i>Ports pour bateau de courte distance</i>	ferry_terminals	Géo	Liste de tous les ports pour bateau de courte distance sur le territoire américain.
<i>Quais</i>	docks	Géo	Liste de tous les quais sur le territoire américain pour héberger des bateaux de grande taille. Utile pour comprendre l'accès aux systèmes maritimes.

Quintile de désirabilité	desirability_quintile	Locale	Quintile du niveau de désirabilité des installations industrielles dans une région, selon le recensement de Green Street. Utile pour comprendre le sentiment des marchés.
Quintile de santé fiscale	fiscal_health_tax_Quintile	Locale	Quintile de l'efficacité des impôts dans une région. Utile pour déterminer si les réinvestissements d'une ville sont porteurs.
Stations et arrêts de train	rlrd_nodes	Géo	Liste de toutes les stations et arrêts de train sur le territoire américain. Utile pour comprendre l'accès au système ferroviaire.
Stations de gaz de pétrole liquéfié	lpg_stations	Géo	Liste de toutes les stations de gaz naturel liquéfié. Utile pour comprendre l'accès à des énergies moins polluantes.
Station de gaz naturel compressé	cng_stations	Géo	Liste de toutes les stations de gaz naturel compressé (propane). Utile pour comprendre l'accès à des énergies moins polluantes.
Stations de superéthanol-E85	e85_stations	Géo	Liste de toutes les stations de superéthanol-E85. Utile pour comprendre l'accès à des énergies moins polluantes.
Taux de change	dexmeus, dexuseu & dexcaus	Macro	Taux de change mexicain, canadien et européen. Utile pour comprendre l'évolution des prix des biens américains.
Taux d'intérêt	dgs5	Macro	Le taux d'intérêt sur obligations fédérales à maturité constante de 5 ans. Utile pour comprendre le coût du financement pour un investisseur en immobilier industriel.
Taux d'occupation	occupancy	Locale	Taux d'installations industrielles occupées dans la région étudiée. Utile pour comprendre la santé du marché immobilier dans une région.
Volume d'aéroports	airport_volume	Locale	Volume de personnes pouvant se retrouver dans le(s) aéroport(s) du lieu étudié. Utile pour déterminer la capacité d'une région.
NCF growth de la dernière période	ncf_growth_1, ncf_growth_2 & ncf_growth_3	Non applicable	Dernière valeur de la variable dépendante. Méthode communément utilisée pour réduire l'autocorrélation.

## Le modèle

Dans le cadre de ce rapport, nous avons décidé d'utiliser un modèle neuronal, assez particulier. En effet, la nature matricielle des données géographiques nous empêche d'insérer ces données dans un modèle plus simple. Puisque les images, en guise de simplicité, peuvent être considérées comme des images, il est important qu'elles soient bien interprétées, il n'y a rien de mieux qu'un réseau neuronal pour accomplir cette tâche.

Ce réseau comporte deux branches, qui éventuellement convergent en une seule branche s'occupant de générer une prédiction trouvée grâce à l'optimisation par l'algorithme du gradient stochastique. La première branche intitulée "geo\_input" prend comme argument chacune des 14 matrices, ceci nécessite 1260 neurones soit un neurone par pixels pour chacune des matrices. Cette branche traite

indépendamment les données géographiques avec quatre couches distinctes. La première couche aplatit les données en un vecteur d'une longueur de 1260 lignes, les trois prochaines couches permettent au modèle d'en apprendre plus sur la réalité géographique d'un lieu. Chacune de ces couches utilise une fonction d'activation "ReLU" notamment pour son efficacité computationnelle, mais surtout puisque c'est elle qui génère les meilleurs résultats.

La deuxième branche intitulée "local\_macro\_input" est chargée de traiter les données locales et macroéconomiques. Cette branche est composée de deux couches, avec des fonctions d'activations "ReLU".

Ces deux branches convergent éventuellement, et sont concaténées en une seule couche. La fin du modèle neuronal s'occupe de trouver des relations entre les données géographiques et les données locales/macroéconomiques. Après la couche concaténée, le modèle comporte deux couches dédiées à cette tâche précise, encore avec des fonctions d'activation "ReLU". (Voir Annexe 2 pour une représentation graphique) Ultimement, la forme de ce présent modèle n'est fort probablement pas la forme la plus optimale, toutefois sa structure est indéniablement intéressante, celle-ci permet au modèle d'apprendre la nature géographique du lieu indépendamment des données locales et macroéconomiques, ceci réserve de la computation au traitement des cartes géographiques. Les poids résultant de l'entraînement ne seront peut-être pas optimaux pour un autre secteur, mais devraient toutefois véhiculer des informations intéressantes pour celui-ci. Les poids ne sont probablement même pas optimaux pour les installations industrielles, d'une part par la faible quantité d'observation et par le processus aléatoire de la détermination des poids des neurones. En somme, la finalité de cette démonstration demeure indicative, évidemment ce modèle n'est pas prêt à être déployé. Ce que le lecteur doit retirer de ce rapport c'est la question à laquelle ce rapport tente de répondre : la géographie est-elle une facette primordiale qui doit être prise en compte dans l'investissement immobilier et la manière dont on l'apprête peut-elle être améliorée par la science des données?

***La géographie est-elle une facette primordiale qui doit être prise en compte dans l'investissement immobilier et la manière dont on l'apprête peut-elle être améliorée par la science des données?***

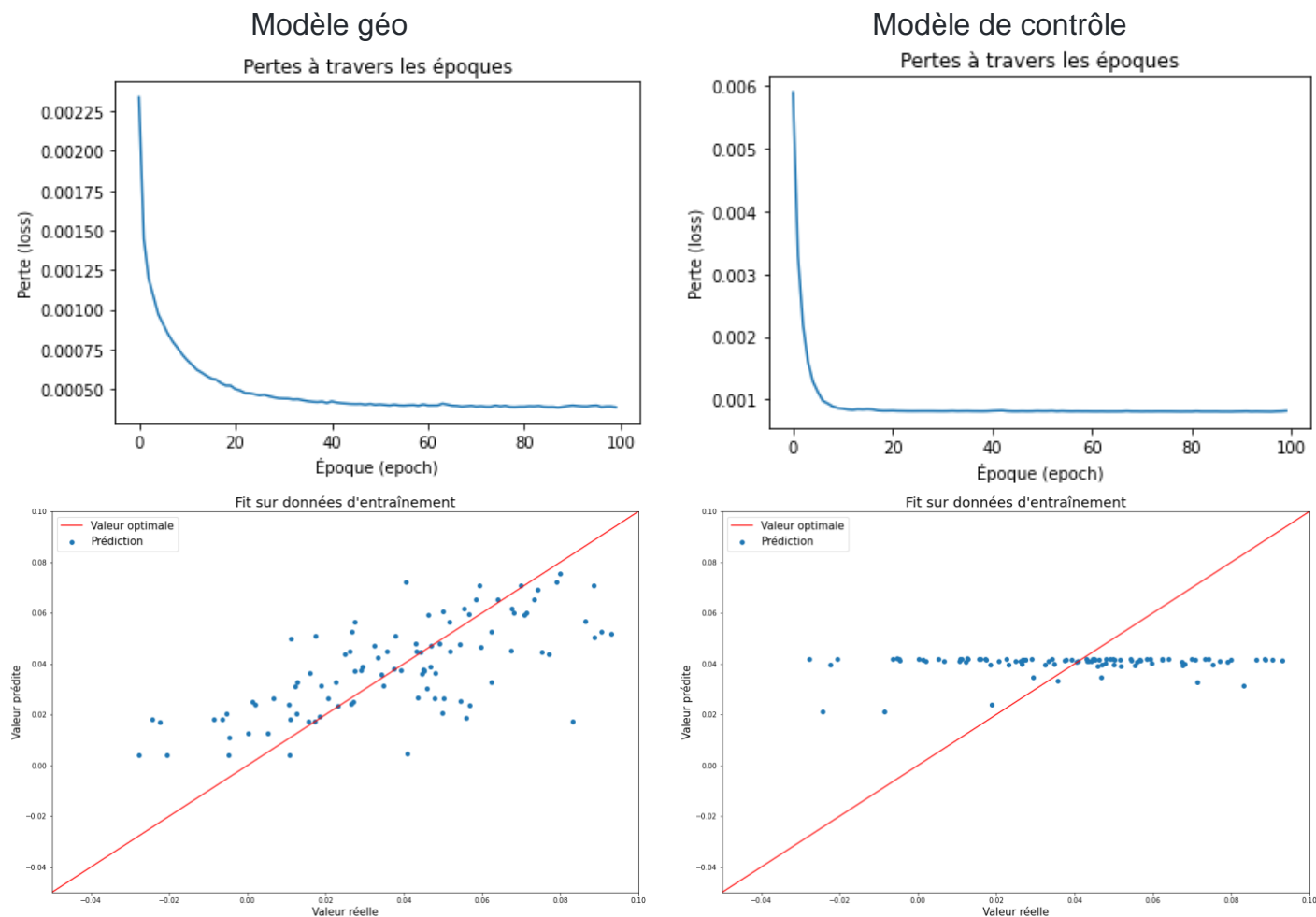
# ANALYSE DES RÉSULTATS

Pour tester la validité de la composante géographique, nous avons procédé à un test assez simple, mais qui devrait donner des résultats assez concluants. Ce test consiste à comparer le modèle doué de la composante géographique au modèle avec seulement la troisième branche. (Revoir Annexe 2 pour contexte) Puisque les données locales et macroéconomiques sont relativement simples, ignorer la deuxième branche n'a pas d'effet, cette branche existait en grande partie pour préparer les données à l'étape de concaténation avec les données géographiques. Le modèle avec la composante géographique sera nommé "modèle géo" et le deuxième modèle sera nommé "modèle de contrôle".

Les graphiques qui suivent comparent les données prédites à des données de prédictions, les lignes rouges servent de référence, si un point est sur ces lignes, cela veut dire que le modèle fait une bonne prédiction. On recherche à avoir des points rapprochés de ces lignes rouges.

## L'entraînement

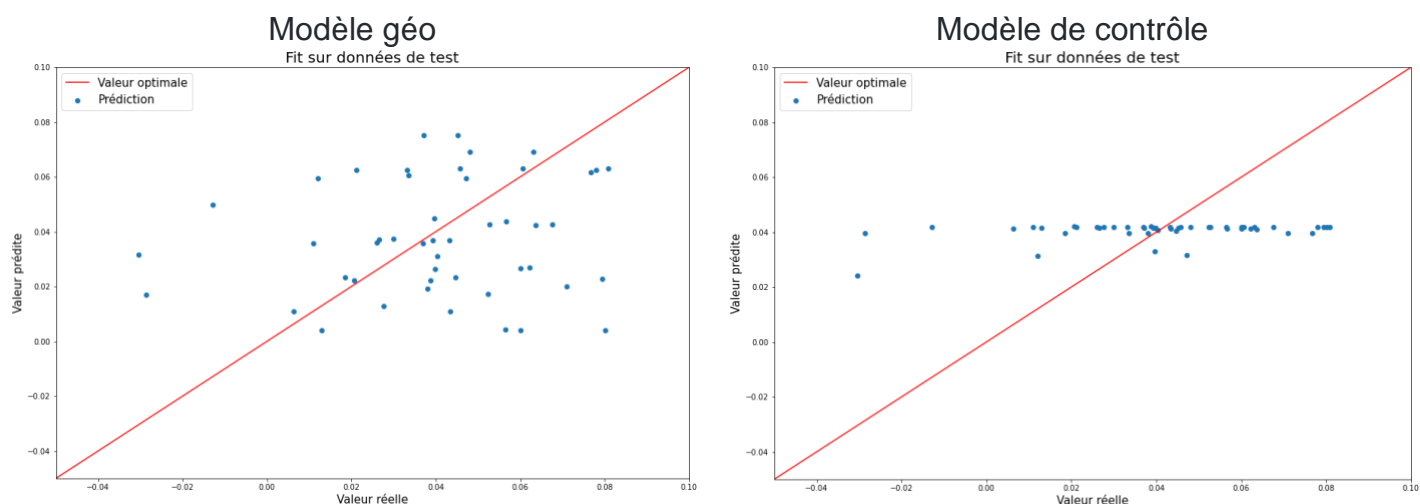
L'entraînement se fait sur les données d'entraînement, échantillonnées par une pige aléatoire parmi les 150 observations. Le deux tiers des données sont réservés à l'entraînement, soit 100 observations. Ceci implique que 50 observations sont réservées à la phase test, la prochaine étape qui nous permettra de déterminer si notre modèle fait de bonnes prédictions.



Le processus d'entraînement est très différent pour les deux modèles, on aperçoit que le modèle de contrôle est optimisé beaucoup plus rapidement. Ceci n'est pas surprenant puisque le modèle de contrôle comporte moins de composantes et d'intrants que le modèle géo. Toutefois, ce qui est frappant c'est l'apprentissage réalisé par les modèles, le modèle géo est visiblement supérieur à son modèle de contrôle. Le modèle de contrôle prédit la moyenne avec quelques variations minimales. Ceci implique que le modèle est incapable de trouver une solution viable. Le modèle géo suit essentiellement la ligne diagonale, ce qui implique que le modèle détecte une tendance. Cette première analyse démontre clairement que le modèle géographique détecte un lien entre la variable dépendante et les variables indépendantes.

## Le test

L'étape du test consiste à procéder à une prédiction sur un échantillon sur lequel nous connaissons la réelle réponse. Si nos prédictions divergent de manière substantielle des valeurs réelles observées, nous pouvons déduire que notre modèle est mauvais. La nature même des données implique plusieurs risques de biais, il faut s'assurer que les villes sur lesquelles les modèles se sont entraînés ne se retrouvent pas dans notre échantillon de test. Puisque les données sont longitudinales, elles ont une composante temporelle, dès lors nous devons aussi nous assurer que les données ne sont pas sélectionnées de manière aléatoire.

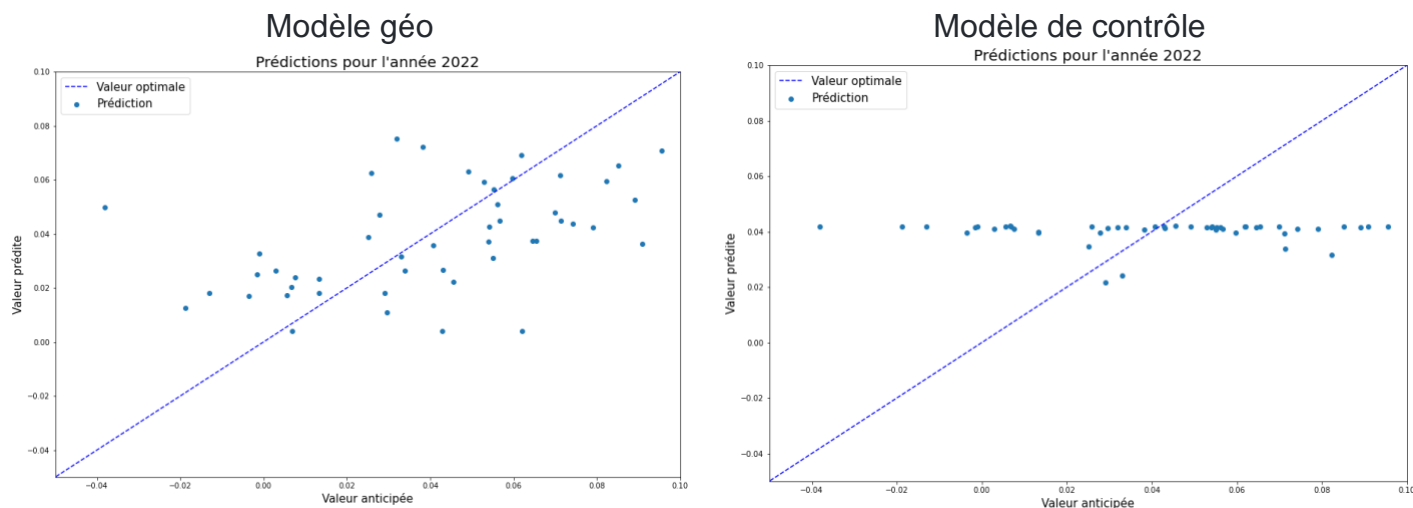


On voit clairement que le modèle géographique est moins précis sur ces données de test, ce n'est pas surprenant. Sur ce test le modèle émane encore d'un certain pouvoir prédictif, toutefois il est grandement amputé par les valeurs extrêmes. Le modèle de contrôle demeure mauvais, et évidemment il le demeurera. Les résultats du test du modèle géographique peuvent être expliqués par plusieurs facteurs, soit le manque d'observation, les métriques utilisées et la structure du réseau neuronal. En somme, après avoir créé un modèle simple, on aperçoit que le modèle contient encore un certain niveau de précision, ce qui est encourageant.

## La prédiction

Nous avons procédé à une prédiction, qui est à prendre à la légère puisque le modèle demeure embryonnaire. Ces prédictions sont pour l'année 2022, elles prennent en compte les données les plus récentes (début 2022) et tentent de prédire quels vont être les flux nets monétaires des installations industrielles d'ici la fin de l'année. Les graphiques suivants doivent être interprétés différemment, si les points sont sur la diagonale, la prédiction est la même que l'anticipation du marché. Autrement, un

point à la gauche de la diagonale, signifie que nous prédisons que ce marché va surperformer les anticipations tandis qu'un point à la droite signifie que nous prédisons le contraire.



Nous observons que les prédictions ont tendance à suivre une différente droite, un peu plus horizontale que la diagonale. Ceci démontre que le modèle géo est plus conservateur que les marchés. Toutefois il est évident que ces deux valeurs sont corrélées, ceci est une autre indication que le modèle fonctionne et qu'il est cohérent. Rendu ici, il n'y a pas grand-chose à dire sur le modèle de contrôle, celui-ci a clairement démontré qu'il est incapable de proposer une prédiction intéressante.

## Le cas de Honolulu

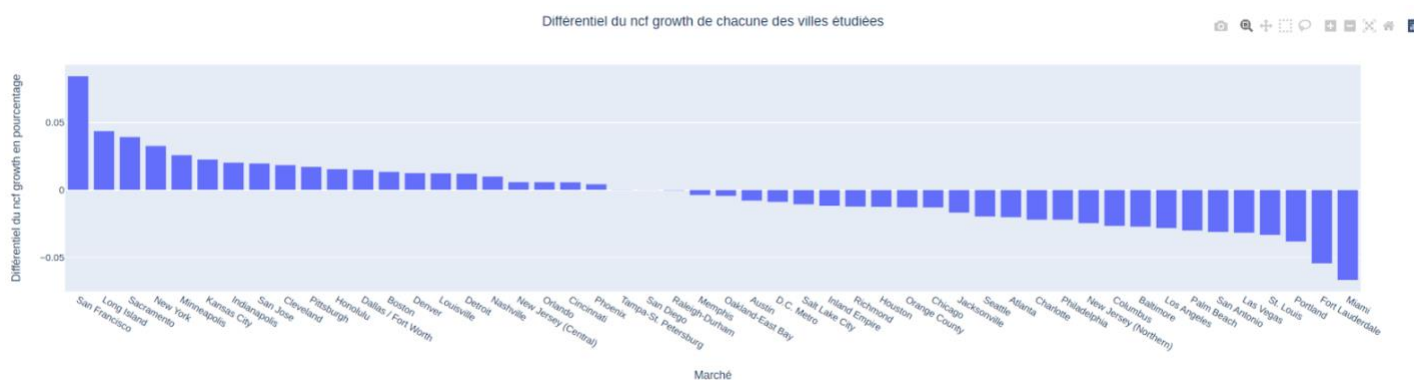
Pour vérifier la validité de la composante géographique, nous avons décidé de garder la ville de Honolulu, située dans l'État d'Hawaii à un peu moins de 4000 km du territoire continental américain. Cette ville n'est pas douée d'une composante géographique (les matrices sont vides), elle sert dès lors d'un très bon second test pour notre modèle géographique. Nous avons procédé à plusieurs tentatives d'entraînement, avec lesquelles nous procédions à une prédiction. Ces prédictions montraient des résultats similaires, généralement avec San Francisco, Sacramento, Kansas City, New York et Long Island en tête. En contrepartie, la ville de Honolulu pouvait se situer en première place comme elle pouvait se situer en plein milieu du lot, sans réelle constance. Par induction, ce résultat erratique nous permet d'arriver à une conclusion très utile; la composante géographique est utilisée et sa présence permet bel et bien au modèle de mieux comprendre l'apport de la géographie quant à la viabilité économique d'une installation industrielle.



# RECOMMANDATIONS

Les recommandations qui vont suivre sont à titre indicatif. Les trois agglomérations les plus performantes trouvées bien qu'intéressantes ne peuvent être sélectionnées uniquement par l'utilisation de ce modèle. (Voir annexe 3 pour table des données) Par le proxy de ces villes nous allons tenter d'en ressortir des tendances quant aux installations présentes dans le modèle géographique et leur relation avec le succès des villes. La nature mystérieuse d'un réseau neuronal peut parfois apporter les parties prenantes à prendre de très mauvaises décisions, de ce fait nous allons tenter de démystifier le tout.

Il est indéniable qu'une prédiction pour l'année 2022 sur la croissance des revenus nets d'une propriété ne permet pas individuellement de prédire la viabilité à long terme d'une installation industrielle. Notre recommandation repose sur l'hypothèse que ces revenus demeureront quelque peu constants, une hypothèse très similaire à celle faite par les marchés recensés par GreenStreet. (voir figure 1) Idéalement, le modèle peut aussi être refaçonné pour prédire le prix d'une propriété, une donnée à laquelle nous n'avons pas accès.



## Long Island & New York

New York est la plus grande ville des États-Unis et est considérée comme l'un des principaux centres financiers, commerciaux et culturels du monde. Long Island est une ville très proche de New York qui tend à être plus industrielle que la grande pomme. L'immobilier commercial et industriel à New York et Long Island est très demandé en raison de la forte concentration de grandes entreprises et de l'attractivité de la ville en tant que lieu de travail et de vie. Les prix de l'immobilier commercial et industriel à New York sont parmi les plus élevés du pays, en particulier dans les quartiers du centre-ville et de Lower Manhattan. Les espaces de bureaux sont particulièrement recherchés à New York, avec de nombreux grands immeubles de bureaux dans les quartiers du centre-ville et de Midtown Manhattan. Il y a également une forte demande pour des espaces de stockage et de production dans les quartiers périphériques de la ville, tels que Brooklyn, Long Island et Queens. Les réglementations locales et les politiques d'urbanisme peuvent avoir une influence sur l'immobilier commercial et industriel à New York. Par exemple, les villes ont récemment adopté des politiques visant à encourager le développement de projets immobiliers durables et à favoriser la création de logements abordables. New York et Long Island ressortent parmi nos observations comme une des villes ayant le plus grand bassin de population, le plus de ports et la plus grande densité de bornes de recharge électrique. (voir annexe 4) New York est historiquement une des villes les plus performantes de notre échantillon, le modèle semble suggérer que ceci devrait persister, et que cette performance devrait



surpasser les anticipations du marché. Plus précisément, le modèle anticipe que la ville de New York devrait surpasser les anticipations par 3%.

## **Sacramento & San Francisco**

Sacramento est la capitale de l'État de Californie et est située en Californie du Nord, dans la région de la vallée de Sacramento. La ville de San Francisco est très proche de Sacramento et elle bénéficie d'environ tous les mêmes bénéfices. Les villes ont connu une croissance rapide au cours des dernières décennies, en partie grâce à l'expansion de l'industrie de la technologie et de l'agriculture. L'immobilier commercial et industriel à Sacramento a suivi cette croissance, avec de nombreux projets de développement immobiliers récents ou en cours dans la ville. Les prix de l'immobilier commercial et industriel à Sacramento ont tendance à être inférieurs à ceux de la région de la baie de San Francisco, mais ils ont tout de même augmenté au cours des dernières années en raison de la demande croissante. Il y a eu une forte demande pour des espaces de bureaux modernes et pour des installations industrielles et logistiques dans les villes, en particulier dans les quartiers du centre-ville et du sud de Sacramento. Les réglementations locales et les politiques d'urbanisme peuvent avoir une influence sur l'immobilier commercial et industriel à Sacramento et à San Francisco. Par exemple, les villes ont récemment adopté des politiques visant à encourager le développement de projets immobiliers durables et à favoriser la création de logements abordables. La ville de Sacramento coïncide avec la jonction de multiples routes artérielles desservant les lieux environnants.

La ville de Sacramento se démarque par son accès au système de transport californien, à sa proximité à l'océan pacifique et à sa densité de bornes de recharge électrique. (voir annexe 5) Notre modèle estime que la ville de Sacramento devrait surpasser les anticipations par 3%, et la ville de San Francisco par 8,5%.

## **Kansas City**

Kansas City est une ville située dans l'État du Missouri, dans la région des Prairies du Midwest des États-Unis. La ville est un important centre commercial et industriel, avec de nombreuses entreprises de divers secteurs d'activité, notamment l'agriculture, la logistique et la fabrication. L'immobilier commercial et industriel à Kansas City est en demande en raison de sa proximité avec d'autres grandes villes du Midwest, telles que Saint-Louis et Omaha, et de sa position stratégique le long de la rivière Missouri, qui converge éventuellement avec la fameuse rivière du Mississippi. Les prix de l'immobilier commercial et industriel à Kansas City sont généralement inférieurs à ceux des grandes villes côtières, mais ils ont tout de même augmenté au cours des dernières années en raison de la demande croissante. Il y a eu une forte demande pour des espaces de bureaux modernes et pour des installations industrielles et logistiques dans la ville, en particulier dans les quartiers du centre-ville et de l'aéroport international de Kansas City. Les réglementations locales et les politiques d'urbanisme peuvent avoir une influence sur l'immobilier commercial et industriel à Kansas City. Par exemple, la ville a récemment adopté des politiques visant à encourager le développement de projets immobiliers durables et à favoriser la création de logements abordables.

Kansas City se démarque notamment par son accès à de multiples quais longeant la rivière du Mississippi et la rivière du Missouri, par sa densité de stations de gaz de pétrole liquéfié et sa densité de stations de superéthanol-E85. (voir annexe 6) Notre modèle estime que Kansas City devrait surpasser les anticipations par 5,5%.

## **Les pires villes**

Les villes de Fort Lauderdale, Miami et Portland sont les villes qui évoquent le moins d'optimisme. Fort Lauderdale et Miami forment essentiellement une seule ville, l'équivalent de Mississauga et Toronto. Ces régions sont souvent reconnues pour leur secteur touristique, toutefois leur secteur industriel

manque de carrure, et bien que ces régions ont enregistré une croissance démographique importante et une croissance soutenue du secteur industriel, le modèle prévoit que ceci va très vite s'arrêter. La ville de Portland située dans l'état de l'Oregon est à l'intersection entre le sud et le nord de l'Ouest américain. Bien que cette ville pourrait sembler importante, la nature même de l'Ouest américain, surtout reconnu pour son commerce avec les marchés asiatiques via le Pacifique, ne nécessite pas réellement d'une grande capacité de transport tout au long de sa côte. Tout de même la ville de Portland n'est pas sur le point de disparaître et pourrait très bien surprendre.

# ESG

L'ESG est une composante de plus en plus prise en compte par les investisseurs immobiliers qui cherchent à évaluer la durabilité et la responsabilité de leurs actifs immobiliers, notamment l'immobilier commercial et industriel. En effet, le secteur de l'énergie, propulsant le secteur industriel, contribue à environ 73% des gaz à effet de serre (source : <https://ourworldindata.org/emissions-by-sector>). En pratique, les préoccupations environnementales sont de plus en plus présentes dans l'esprit des consommateurs et des investisseurs. Les propriétés qui ont une empreinte carbone réduite et qui adoptent des pratiques durables sont de plus en plus prisées. De nombreux investisseurs et consommateurs sont de plus en plus conscients de l'impact de leurs décisions sur l'environnement. Ils cherchent donc des propriétés qui ont une empreinte carbone réduite et qui adoptent des pratiques durables.

***Le secteur de l'énergie, qui propulse le secteur industriel, contribue à environ 73% des gaz à effet de serre.***

En outre, les lois et réglementations en matière d'environnement sont de plus en plus strictes aux États-Unis, ce qui signifie que les propriétés qui ne respectent pas ces réglementations peuvent faire face à des coûts de conformité élevés. Aux États-Unis, il existe de nombreuses lois et réglementations visant à protéger l'environnement. Les propriétés qui ne respectent pas ces réglementations peuvent faire face à des coûts de conformité élevés, ce qui peut affecter la valeur de la propriété. Par exemple, une propriété qui génère des déchets dangereux ou qui rejette des substances polluantes dans l'air ou l'eau peut être moins attrayante pour les investisseurs en raison des coûts potentiels associés à la conformité réglementaire.

Enfin, les préoccupations sociales et de gouvernance ont également un impact sur la valeur de l'immobilier commercial et industriel. Par exemple, les propriétés situées dans des quartiers à faible revenu peuvent être moins attrayantes pour certains investisseurs en raison des problèmes sociaux qui y sont associés. Les préoccupations sociales et de gouvernance peuvent également affecter la valeur de l'immobilier commercial et industriel. Par exemple, une propriété située dans un quartier à faible revenu peut être moins attrayante pour certains investisseurs en raison des problèmes sociaux qui y sont associés. De même, une propriété gérée de manière peu transparente ou responsable peut être moins attrayante pour les investisseurs soucieux de la gouvernance.

Dans ce sens, nous avons énuméré dans ce rapport des infrastructures qui sont pertinentes à prendre en considération lors de la recherche d'un lieu idéal pour investir dans l'immobilier commercial et industriel, plus particulièrement l'accès à des sources d'énergie moins polluante dans l'analyse des lieux industriels, telle que les stations de gaz naturel comprimé, les stations superéthanol-E85, les bornes de recharge, ainsi que les terminaux de gaz naturel liquéfié.

En ce qui concerne la présence de stations de gaz naturel comprimé, elle peut être une métrique utile pour évaluer l'accès à l'énergie dans une région donnée. Le gaz naturel comprimé est un combustible

alternatif qui peut être utilisé comme carburant pour les véhicules à moteur. Les bornes peuvent être installées dans différents endroits, telles que sur les parkings de centres commerciaux, le long des routes, etc. L'accès à ces stations peut être une métrique utile pour évaluer l'accès à l'énergie dans une région donnée, car elle peut indiquer la disponibilité du gaz naturel comprimé comme carburant alternatif. Cette métrique peut être utilisée pour orienter les décisions de développement et de planification en matière d'infrastructure de transport et pour évaluer les opportunités de transition vers des carburants alternatifs.

En outre, le superéthanol-E85 (E85) est un carburant alternatif qui est produit à partir de mélanges de biocarburants (principalement de l'éthanol) et de carburants fossiles (principalement de l'essence). Il peut être utilisé comme carburant pour certains types de véhicules à moteur. Les stations de E85 peuvent être installées dans différents endroits, telles que sur les parkings de centres commerciaux, le long des routes, etc. L'accès à ces stations peut être une métrique utile pour évaluer les perspectives de développement durable dans une région donnée, car elle peut indiquer la disponibilité du E85 comme carburant alternatif. Cette métrique peut être utilisée pour orienter les décisions de développement et de planification en matière d'infrastructure de transport et pour évaluer les opportunités de transition vers des carburants alternatifs à faibles émissions de carbone.

De plus, la présence de bornes de recharge permette d'évaluer les perspectives de développement durable dans une région donnée. Les bornes de recharge sont des points de recharge pour les véhicules électriques. Elles permettent aux propriétaires de véhicules électriques de recharger leurs véhicules lorsqu'ils sont garés et de rouler avec une autonomie étendue. Cette métrique peut indiquer la disponibilité de l'infrastructure de recharge pour les véhicules électriques. Elle est également utilisée pour orienter les décisions de développement et de planification en matière d'infrastructure de transport et pour évaluer les opportunités de transition vers des véhicules à faibles émissions de carbone.

Enfin, la présence de terminaux de gaz naturel liquéfié (GNL) peut être une métrique utile pour évaluer l'accès à l'énergie à proximité dans une région donnée. Le gaz naturel liquéfié est un combustible alternatif qui peut être utilisé comme carburant pour les véhicules à moteur ou comme source d'énergie pour le chauffage, la cuisson et la production d'électricité. Les terminaux de GNL sont des installations qui sont utilisées pour stocker et distribuer le GNL. Ils peuvent être installés dans différents endroits, tels que près de ports ou de terminaux de gazoducs, et peuvent être utilisés pour acheminer le GNL vers les consommateurs finaux, tels que les centrales électriques, les usines de traitement de gaz, etc.

En d'autres mots, la présence des stations de superéthanol et les bornes de recharge favorisent la transition vers des véhicules à faibles émissions de carbone. En ce qui concerne le superéthanol, il s'agit d'un biocarburant qui est utilisé de plus en plus dans les véhicules de tourisme, mais également dans certains camions de transport. Bien entendu, pour ce qui est des bornes de recharge, leur présence élevée à certains endroits a tendance à indiquer une présence plus élevée de véhicules électriques. Par conséquent, dans le cadre d'une approche ESG, il est plus cohérent de favoriser l'investissement dans des villes et lieux où les véhicules utilisant des sources d'énergie renouvelable sont les plus présents. Ces éléments ont été considérés comme des éléments positifs dans l'élaboration de notre modèle.

Finalement, notons également que nous avons également pris en compte les stations de trains et les arrêts de transport en commun qui sont des éléments importants en matière d'immobilier résidentiel. En effet, il est plus conforme aux normes ESG, plus précisément la composante environnementale, d'investir dans des villes et lieux qui favorisent le transport en commun et mettent à la disposition de leurs habitants des moyens de transport en commun ce qui contribue à réduire le nombre d'automobiles sur les routes, et donc, les émissions de gaz à effet de serre.

---

# CONCLUSION

En résumé, le but du présent rapport est de répondre à la question “comment identifier les villes et lieux de demain?” Nous tentons de prédire la croissance future des flux monétaires nets des actifs industriels d'une région (net cash-flow growth). Pour ce faire, les données fournies par GreenStreet sont utilisées en utilisant deux hypothèses: que les données fournies par GreenStreet représentent de manière holistique la région étudiée et que les méthodes de mesure utilisées par GreenStreet sont constantes à travers le temps. De plus nous soulevons l'hypothèse que les anticipations du marché pour l'année 2022 demeureront constantes dans les années futures. Les données sont ensuite divisées en trois catégories distinctes: les données géographiques, les données locales et les données macroéconomiques. La catégorie géographique utilise un modèle inédit qui répertorie les données dans des matrices flexibles pouvant être utilisées pour différents modèles d'apprentissage automatique, cela permet de représenter le réseau infrastructurel d'une région de manière fiable. Le texte présente une analyse des résultats d'un test visant à valider l'utilité de la composante géographique dans un modèle pour prédire la croissance future des flux monétaires nets d'actifs industriels d'une région. Le test consiste à comparer un modèle comprenant cette composante géographique à un modèle ne la comprenant pas, et montre que le modèle comprenant cette composante est supérieur en prédisant la tendance de croissance. La conclusion est que cette composante géographique détecte un lien entre la variable dépendante et les variables indépendantes. Le texte présente des recommandations basées sur l'utilisation d'un modèle pour prédire la croissance future des flux monétaires nets d'actifs industriels dans différentes villes. Il est précisé que ces recommandations ne peuvent pas être utilisées seules pour prendre des décisions d'investissement et qu'il est nécessaire de considérer d'autres facteurs. En particulier, des tendances ont été identifiées pour les trois villes les plus performantes, New York, San Francisco et Kansas City, qui montrent que ces villes ont un grand bassin de population, de ports et de bornes de recharge électrique. Le modèle prévoit également que San Francisco devrait surpasser les attentes de croissance du marché de 8%.

# ANNEXE

## Annexe 1



Figure 6 Phénomène du "zoom" observé pour la ville de New York, présent sur le fichier « ML » avec la fonction « get\_zone() »

## Annexe 2

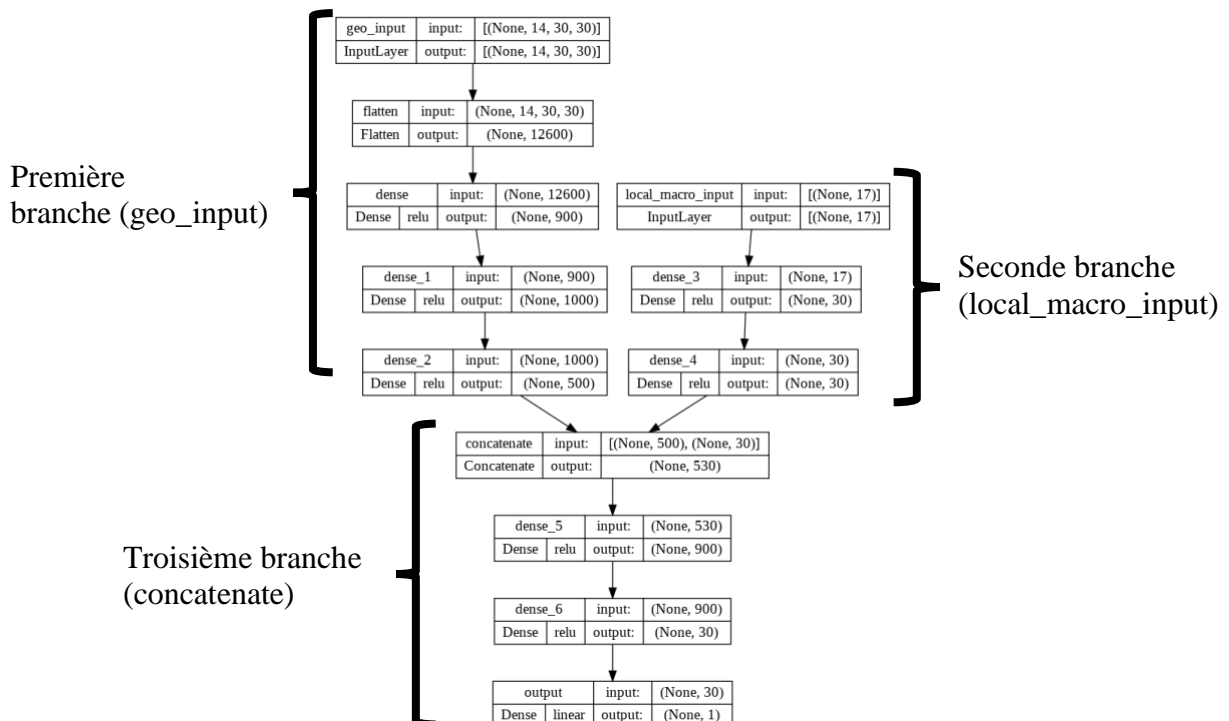


Figure 7 Réseau neuronal utilisé dans le cadre de l'analyse, présent sur le fichier « ML » mais illustration généré via google colab

## Annexe 3

market_publish	date	age_median	airport_volume	asset_value_momentum	desirability Quintile	fiscal_health_tax Quintile	interstate_distance	interstate_miles	mrevpaf_growth_yoy_credit	occupancy	population_500mi	latitude	longitude	DCOILWTIC0	DEMXJUS	DEXUSEU	DEXCAUS	NETEXP
New York	2019	36.581302	2.387905e+09	0.143862	2.0	2.0	1.383107	128.178417	0.0664	0.874961	46174358.0	40.750633	-73.997170	45.15	19.635	1.1456	1.3644	-599.911
	2020	37.038044	4.411078e+09	0.105195	1.0	2.0	1.608208	141.647176	0.0615	0.899010	73095917.0	40.750633	-73.997170	61.14	18.860	1.1227	1.2962	-522.725
	2021	37.474197	4.411078e+09	0.125658	1.0	1.0	1.608208	141.647176	0.0664	0.897000	72954917.0	40.750633	-73.997170	48.35	19.892	1.2230	1.2753	-808.606
	2022	37.856410	6.329436e+09	0.339700	2.0	1.0	1.526853	153.433923	0.1242	0.920082	73095917.0	40.750633	-73.997170	75.99	20.570	1.1290	1.2757	-914.708
Sacramento	2019	36.371471	2.389111e+08	0.234433	4.0	2.0	2.634376	105.769666	0.0664	0.914961	46174358.0	38.580568	-121.495590	45.15	19.635	1.1456	1.3644	-599.911
	2020	36.515148	2.884461e+08	0.133502	4.0	2.0	3.093250	104.417115	0.0713	0.939010	36743484.0	38.580568	-121.495590	61.14	18.860	1.1227	1.2962	-522.725
	2021	36.495360	2.884461e+08	0.053841	3.0	2.0	3.093250	104.417115	0.0664	0.934000	36908036.0	38.580568	-121.495590	48.35	19.892	1.2230	1.2753	-808.606
	2022	36.321503	2.937877e+08	0.522611	3.0	2.0	3.233346	105.769666	0.1204	0.942000	36743484.0	38.580568	-121.495590	75.99	20.570	1.1290	1.2757	-914.708
Kansas City	2019	37.215679	5.885946e+08	0.054550	2.0	2.0	1.867944	237.417808	0.0664	0.910961	46174358.0	39.104839	-94.572176	45.15	19.635	1.1456	1.3644	-599.911
	2020	37.379758	6.754297e+08	0.105454	2.0	2.0	2.090132	237.566689	0.0442	0.926010	29400644.0	39.104839	-94.572176	61.14	18.860	1.1227	1.2962	-522.725
	2021	37.407494	6.754297e+08	0.099610	2.0	3.0	2.090132	237.566689	0.0664	0.940000	29476460.0	39.104839	-94.572176	48.35	19.892	1.2230	1.2753	-808.606
	2022	37.289322	6.754297e+08	0.413577	1.0	3.0	2.332135	237.566689	0.1040	0.952000	29400644.0	39.104839	-94.572176	75.99	20.570	1.1290	1.2757	-914.708

Figure 8 Table de donnée de 3 villes les plus performantes. Les données de 2019 à 2021 sont utilisées pour l'entraînement et le test, tandis que les données de 2022 sont utilisées pour faire une prédiction

## Annexe 4

Données géographiques de New York

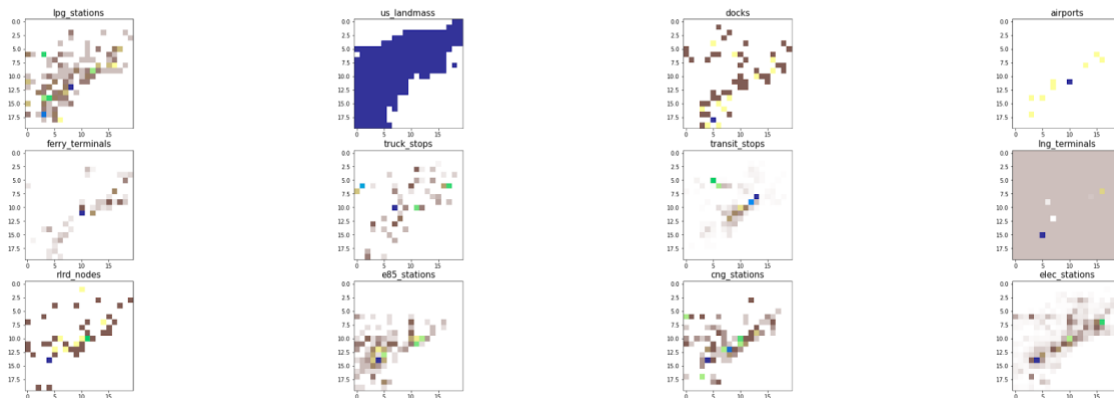


Figure 9 Matrice géographiques utilisées pour la ville de New York, présent sur le fichier « ML » vers la fin

## Annexe 5

Données géographiques de Sacramento

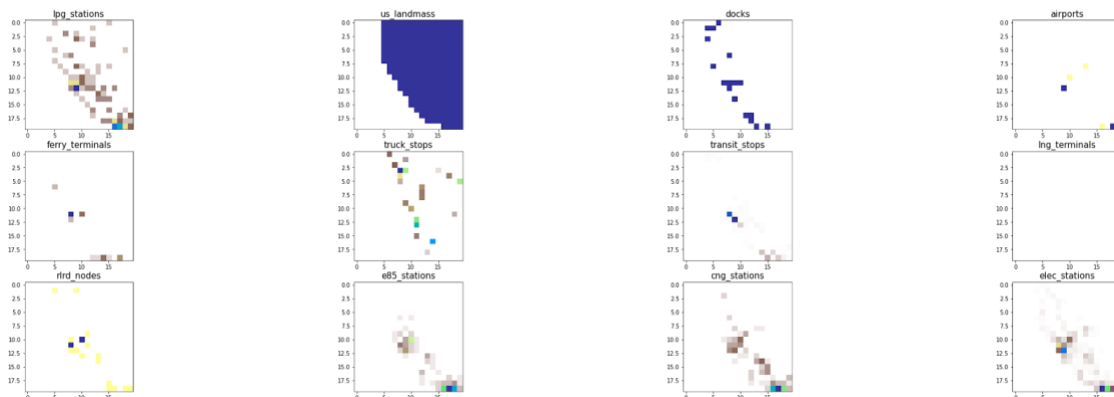


Figure 10 Matrice géographique utilisées pour la ville de Sacramento, présent sur le fichier « ML » vers la fin



## Annexe 6

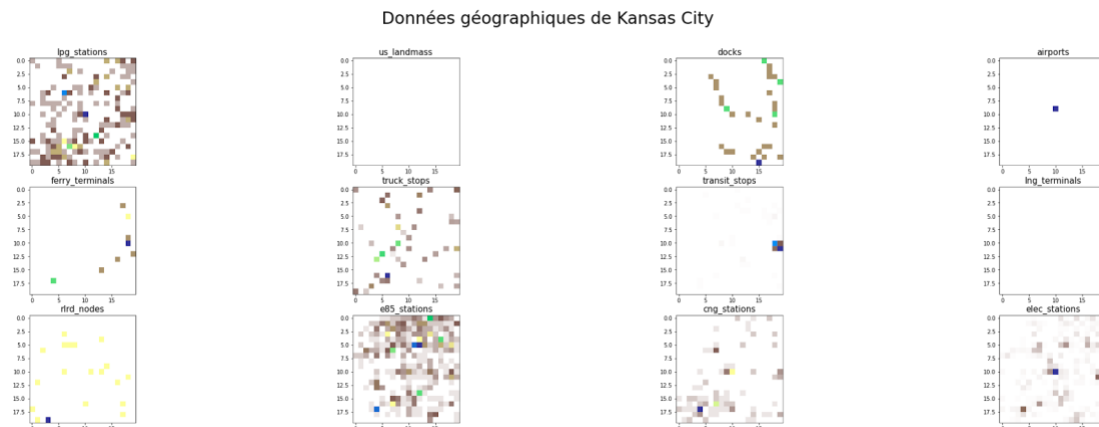


Figure 11 Matrices géographiques utilisées pour la ville de Kansas City, présent sur le fichier « ML » vers la fin

## Annexe 7

### Nom du/des fichier(s) Description

<i>classification_lieux.py</i>	Fichier sur lequel nous avons procédé au nettoyage de données, et l'extraction de leurs coordonnées.
<i>classification_site.py</i>	Fichier pour l'exploration de la variable dépendante.
<i>coordinates.py</i>	Algorithme connecté à l'api OpenStreetMap nous permettant de retirer les coordonnées d'une ville avec son nom.
<i>generate_matrix.py</i>	Fichier comportant tous les algorithmes nécessaires pour la création et l'utilisation du modèle géographique.
<i>layers.ipynb</i>	Fichier sur lequel contient le processus de population de tous les matrices géographiques du rapport.
<i>ML.py</i>	Fichier sur lequel contient le nettoyage final de nos données, la création et l'entraînement du modèle neuronal et la plupart des visualisations du rapport.
<i>real_estate.py</i>	Fichier sur lequel contient le reste des visualisations utilisées dans ce rapport.

### Nom du dossier Description

<i>data/</i>	Tous les données nettoyées utilisées pour populer les matrices géographiques.
<i>ml/data</i>	Tous les matrices géographiques, enregistrées sous format .npy.
<i>img/</i>	Certaines des images utilisées pour le rapport.

Lien vers le projet sur GitHub : <https://github.com/felixpoirier1/I3C>

---

# BIBLIOGRAPHIE

Sources utilisées pour les matrices géographiques :

*Bureau of Transportation Statistics*, <https://geodata.bts.gov/>

Global Energy Monitor (terminaux LNG), <https://globalenergymonitor.org/projects/global-gas-infrastructure-tracker/download-data/>

Source utilisée pour les données locales :

*GreenStreet*, market\_sectors\_\_historical\_market

Source utilisée pour les données macroéconomiques :

St. Louis Fed Economic Data, <https://fred.stlouisfed.org/>

Source utilisée pour les données dépendantes :

*GreenStreet*, forecasts\_\_historical\_baseline

Autres sources :

Nominatim, <https://nominatim.org/>

Base de donnée avec latitude et longitude pour plusieurs villes,  
<https://gist.github.com/dikaio/0ce2a7e9f7088918f8c6ff24436fd035>