



La Grande Ecole de l'IA & de la Data  
Paris Île-de-France · Nice Côte d'Azur

## Analyse de données agroalimentaires pour une expansion internationale pour une start-up AIGriTech

Clinique de l'IA Programme Grande Ecole  
Année universitaire 2025 2026

Théo Szappanyos | Paul Ferran | Amin Addal | Camil Zerhouni

# Table des matières

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
<b>1.1</b>	<b>Contexte business .....</b>	<b>1</b>
<b>1.2</b>	<b>Problématiques et enjeux .....</b>	<b>1</b>
<b>2</b>	<b>Exploratory Data Analysis .....</b>	<b>2</b>
<b>2.1</b>	<b>Présentation Générale .....</b>	<b>2</b>
<b>2.2</b>	<b>Qualité des données .....</b>	<b>2</b>
<b>2.3</b>	<b>Cohérence des données .....</b>	<b>2</b>
<b>2.4</b>	<b>Conclusion de l'EDA .....</b>	<b>2</b>
<b>3</b>	<b>Analyse économique   étude de marché .....</b>	<b>3</b>
<b>3.1</b>	<b>Identification des indicateurs de base .....</b>	<b>3</b>
<b>3.2</b>	<b>Définition des critères d'analyse du marché .....</b>	<b>3</b>
<b>3.3</b>	<b>Clustering par profils économiques similaires .....</b>	<b>4</b>
<b>3.4</b>	<b>Calcul du score de marché .....</b>	<b>6</b>
<b>4</b>	<b>Analyse démographique   étude prédictive .....</b>	<b>7</b>
<b>4.1</b>	<b>Régression linéaire .....</b>	<b>7</b>
<b>4.2</b>	<b>Régression polynomiale de degré 2 .....</b>	<b>8</b>
<b>4.3</b>	<b>Comparaison et conclusion partielle .....</b>	<b>9</b>
<b>4.4</b>	<b>Calcul de critères de prédiction .....</b>	<b>9</b>
<b>5</b>	<b>Modélisation &amp; Interprétation des analyses .....</b>	<b>10</b>
<b>5.1</b>	<b>Diagrammes .....</b>	<b>10</b>
<b>5.2</b>	<b>Matrice Stratégique .....</b>	<b>12</b>
<b>5.3</b>	<b>Recommandations business .....</b>	<b>13</b>

## 1.Introduction

### 1.1 Contexte Business

La société AIGriTech est spécialisée dans l'agriculture automatisée de céréales, grâce à l'intelligence artificielle. En particulier le blé, l'orge et le maïs, qui sont des céréales essentielles à tous les pays et qui sont la base de l'alimentation globale. Les engins agricoles sont pilotés par des algorithmes, réduisant le nombre d'employés et les coûts de production. Simultanément, l'agriculture de ces céréales est peu complexe, ce qui nous a permis de rapidement augmenter l'échelle de la production, en minimisant le coût de production marginal et maximisant les quantités produites. Cette configuration est optimale pour l'exportation en masse à l'internationale : ces produits sont achetés en masse par des distributeurs locaux, et nos prix bas permettent une implantation agressive en coupant les prix.

La mission de ce projet est d'utiliser les bases de données fournies sur le marché agroalimentaire mondial, déterminer si elles sont utilisables et si oui, fournir des analyses stratégiques pour les directeurs de l'entreprise, des recommandations business qui vont aider aux décisions de ce projet d'expansion internationale.

### 1.2 Problématiques et enjeux

Les enjeux et problématiques de cette analyse sont nombreux. Les jeux de données utilisés doivent être scientifiquement explorés, pour déterminer leur utilisabilité, pertinence, repérer les possibles biais et disparités sociales, en fonction des zones géographiques. L'identification des indicateurs à utiliser doit être judicieuse, et l'analyse doit être pluridimensionnelle : économique (quels types de marché), géographique (filtrer les pays cibles), et temporelle (expansion viable dans le futur). On doit également prendre en compte des problématiques extérieures à ces données, comme la logistique d'acheminement par voie maritime des produits, mais aussi les enjeux et risques géopolitiques.

Les livrables de ce projet sont :

- Une préanalyse exploratoire des jeux de données, où EDA (Exploratory Data Analysis), qui a pour but de décrire, comprendre les données, mais également de valider leur qualité et repérer les potentielles valeurs manquantes ou aberrantes et les doublons.
- Un rapport détaillé qui montre les étapes du projet, la méthodologie utilisée, les analyses précises, dans un format professionnel.
- Des graphiques clairs modélisant les analyses des données pour qu'elles soient facilement interprétable par le comité direction.
- Des recommandation business pour aider les décisions sur l'expansion, avec différentes nuances et point de vue (géographiques, temporelles, économiques).
- Un notebook Jupyter Lab, contenant tout l'algorithme/code source en python, documenté et organisé de manière professionnelle.

## 2. Exploratory Data Analysis (EDA)

L'enjeu de cette section est essentiel, car si l'exploration préliminaire des datasets n'est pas satisfaisante, on ne peut pas procéder à l'analyse. On utilise le module pandas dans python.

### 2.1. Présentation Générale

Les deux jeux de données sont ‘DisponibiliteAlimentaire\_2017.csv’ et ‘Population\_2000\_2018.csv’ et sont fournis par la FAO (Food and Agriculture Organization). Cette organisation suit une méthodologie de collecte de données éthique et responsable et garantit un standard de qualité des données. Nos fichiers csv contiennent 176600 lignes & 14 colonnes pour DisponibiliteAlimentaire\_2017.csv ; et 4411 lignes, 15 colonnes pour Population\_2000\_2018.csv. Ils ont exactement les mêmes colonnes, parmi lesquelles ‘Zone’ donne le pays, ‘Année’ donne l’année de la statistique, ‘Elément’ donne la nature de la statistique (imports, production, disponibilité intérieure, etc.), ‘Produit’ le produit en question et ‘Valeur’ donne la valeur de l’observation. Le dataset avec la population contient une colonne supplémentaire ‘Note’ qui apporte occasionnellement des annotations aux lignes.

### 2.2. Qualité

Ici, on vérifie la complétude des données, qui peut fortement compromettre notre analyse et réduire sa fiabilité. On trouve 0 valeurs manquantes et 0 doublons sauf pour 4153 valeurs manquantes dans la colonne ‘Note’, ce qui est logique car il n’y en a pas à chaque ligne. Cette colonne n’est pas nécessaire à notre analyse donc on la supprimera.

### 2.3. Cohérence/Typage

Le typage des données est natif et correct pour chaque colonne, il n’y a aucune conversion à faire. Les valeurs qualitatives comme la zone et le produit sont au format ‘object’, mais aussi au format ‘int’ pour le code zone ou l’année, qui reste des valeurs qualitatives. Les valeurs quantitatives sont au format ‘float’ car comporte souvent des décimales. Les unités de mesures sont souvent en millier de tonnes ou en millier de personnes, mais aussi en kg/habitant/an. Enfin, le dataset de disponibilité couvre 174 pays, contre 236 pour le dataset de population.

### 2.4. Conclusion

Comme la FAO a promis, les données sont optimales pour l’analyse. On a deux datasets de qualité, 0 doublons 0 valeurs manquantes car la colonne ‘Note’ dans Population\_2000\_2018.csv est inutile donc on la supprime. Les données n’ont aucune erreur de typage et concordent entre les deux dataframe. Une harmonisation des pays entre les deux dataframes est nécessaire car il y en a 62 de plus dans le fichier csv des populations. On peut procéder au nettoyage.

### 3.Analyse économique | Etude de marché

On analyse DisponibiliteAlimentaire.csv, qu'on renomme df, pour en extraire les statistiques nécessaires à une étude du marché. Il contient les statistiques économiques pour 2017 de 174 pays. On cherche à trouver les pays les plus intéressants pour notre type d'export, c'est-à-dire des quantités massives et des prix bas.

#### 3.1. Identification des indicateurs de base

En observant les 15 statistiques différentes dans la colonne ‘Elément’, on en retire deux intéressantes. D’abord, la disponibilité intérieure, qui selon la FOA, représente la quantité d’un produit disponible à la consommation pour le pays en question (calcul : production + import – export). Concrètement elle représente la consommation, la demande du produit, en quantité (milliers de tonnes), dans ce pays. Et plus elle est élevée, plus le pays nous intéresse.

Elle est également présente sous la forme Disponibilité en kg par habitant par an, ce qui aurait été utile dans le cas où nos produits étaient directement adressés aux populations, dans les magasins locaux, etc. Dans notre cas ces céréales forment la base de toute alimentation, nos clients seraient donc des distributeurs locaux, qui achètent en masse et revendent au détail.

L’autre statistique intéressante, c’est l’import en quantité : on cherche à exporter en masse, donc le volume de l’import est très important, il traduit la taille du marché. Un fort volume d’importations pourrait traduire la présence d’un grand nombre déjà présente dans le marché, mais nos faibles coûts de production nous permettent de couper les prix de ce marché et de s’implémenter efficacement.

On filtre donc df pour ne garder que les lignes dont la colonne ‘Produit’ correspond à nos produits cible : ‘Blé et produits’, ‘Maïs et produits’, ‘Orge et produits’. Ensuite, on garde seulement les lignes où l’élément de la statistique correspond à ‘Disponibilité Intérieure’ et ‘Importations – Quantité’. Enfin on regroupe ces données dans un Dataframe ‘merged’ qui comportera toutes les données nécessaires à l’analyse (Zone, Imports, Disponibilité/Demande) On est prêt pour l’analyse.

#### 3.2. Définition de critères d’analyse de marché

Les indicateurs de base sont essentiels mais on les combine pour former un dernier critère essentiel pour analyser ce marché, c’est la part des imports dans la disponibilité intérieure d’un pays. Il traduit le taux de dépendance d’un pays aux imports. Cette nuance est essentielle, car un pays peut avoir une disponibilité élevée mais ne pas dépendre des imports. Le ratio sera donc bas, et notre startup aura beaucoup de mal de s’implanter sur ce marché. On cherche donc des pays avec un volume d’importation élevée, une demande intérieure forte, et une part des imports dans la disponibilité élevée. Ce taux est calculé comme ceci : imports / demande, et est donc compris entre 0 et 1. Par conséquent on normalise les deux autres critères, et on ajoute ces

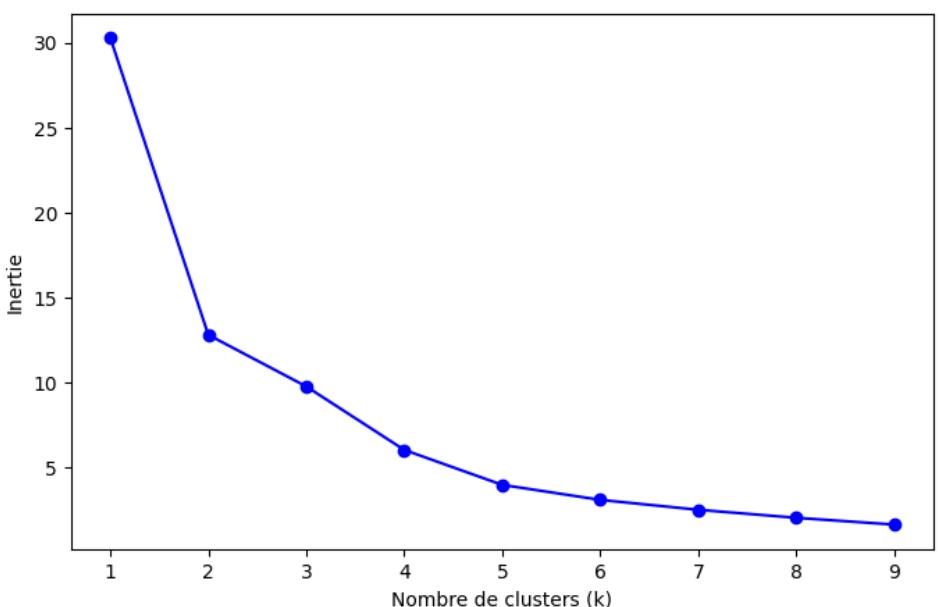
trois nouvelles colonnes (import normalisé, demande normalisée, part des imports) au dataframme ‘merged’. Ces trois critères sont ceux qui vont servir à l’analyse du marché.

### 3.3. Clustering : groupement de pays avec des profils similaires

Grâce à la définition des trois critères d’analyse, on peut regrouper tous les pays selon leur profil, en fonction de ces critères. Pour cela on utilise la méthode du clustering. Cette méthode consiste au partitionnement d’un nuage de données en groupes (clusters) de données qui sont proches dans ce nuage. Ici, on fait des moyennes de ces trois critères pour les caser dans un certain nombre de groupes.

Mais comment identifier le nombre de partition idéales pour 174 pays ? Beaucoup de partition donnera des catégories très homogènes mais il y aura trop de classes pour l’analyse. Peu de partition donnerait des classes trop hétérogènes donc une perte d’intérêt. Pour estimer le nombre  $k$  optimal de partition, on utilise la Elbow Method, soit la méthode du coude. Elle consiste à créer  $k$  partitions, avec  $k$  augmentant de 1 à chaque fois (on s’arrête à  $k=10$  ici), calculer l’inertie intra-classe (écart de valeur des données dans les classes, c’est-à-dire l’homogénéité d’une partition) à chaque fois, puis tracer une courbe des inerties pour chaque itération du clustering. Pour lire cette courbe, on essaye de distinguer un « coude », après lequel l’augmentation du nombre de cluster  $k$  présente un gain marginal d’inertie inférieur au  $k$  précédent.

**Fig 1 :** Courbe Elbow Method

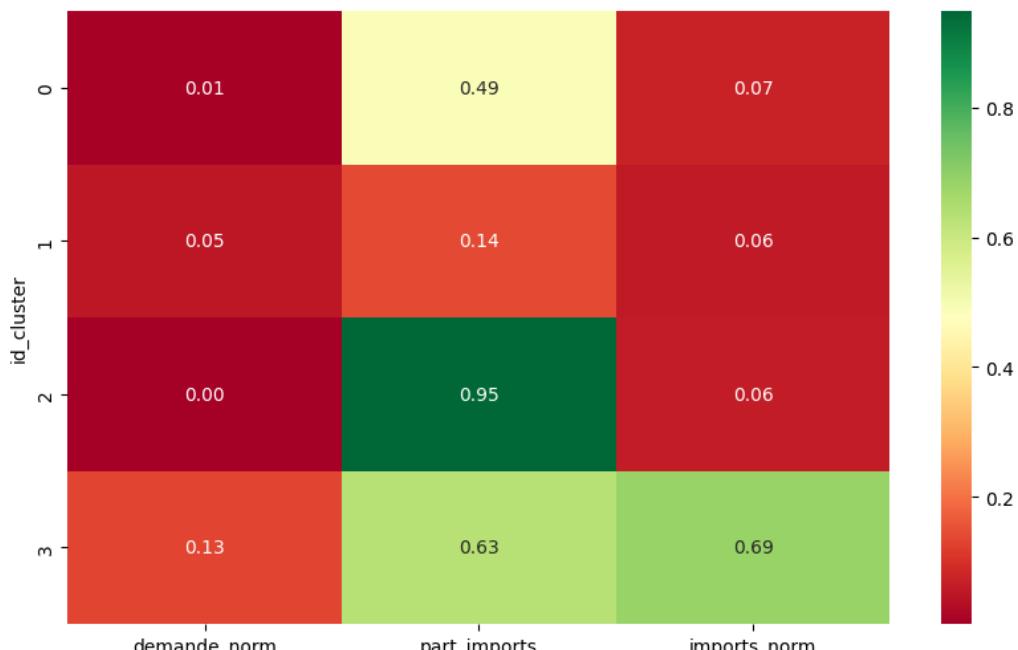


Champs :    colonnes    ‘imports\_norm’, ‘part\_imports’, ‘demande\_norm’  
depuis le data ‘merged’

Dans notre situation, la courbe est atypique. On pourrait observer un coude à  $k=2$ , mais on voit que la baisse d’inertie quand on passe de 3 à 4 clusters (-3.76) est supérieure à la baisse d’inertie en passant de 2 à 3 partitions (-3.07). Notre coude ultime se situe donc à  $k=4$ . On décide donc de partitionner nos données en 4 clusters, donc 4 groupes de pays possédant un profil de marché similaire. On ajoute une colonne ‘id\_cluster’ à merged.

Maintenant, l'intérêt est de savoir lequel de ces profils nous intéresse, pour réduire le champ d'analyse à un petit groupe de pays plutôt que 174. Pour observer les profils, on utilise une heat map ou carte thermique, qui est un tableau avec les moyennes des critères pour chaque groupe. Les cases sont colorées suivant une échelle du rouge au vert, qui montre la valeur moyenne du critère en question pour le cluster en question, par rapport à l'échelle de couleur des critères.

**Fig 2 : Heat Map du clustering en 4 groupes de pays avec des profils similaires.**



*Champs : moyenne par critères d'analyse et par cluster depuis 'merged'*

Dans notre heatmap l'échelle est donc de 0 à 1 puisqu'on a normalisé les critères. Plus une case est verte, plus le critère est idéal pour notre objectif d'expansion. Par conséquent, on élimine par défaut le premier et deuxième groupe, car leurs critères sont très faibles. On remarque une forte présence de vert dans le troisième (cluster 2) et le quatrième (cluster 3) groupe. Dans le troisième, la part des imports est extrêmement élevée, mais le volume et la demande sont très faibles. Ce sont des petits pays, qui sont dépendants des productions extérieures.

En revanche dans le quatrième groupe, la part des imports est légèrement plus faible, mais le volume d'import est très élevé, et la demande normalisée est la plus haute des groupes. Ce sont les « géants importateurs », c'est-à-dire les moteurs de la demande mondiale. Ils combinent un fort volume d'imports/de consommation, et une dépendance aux imports extérieurs, souvent à cause d'une incapacité de production locale pour des raisons climatiques ou autres. Dans ce groupe on retrouve des pays comme le Japon, la Chine, l'Arabie Saoudite, l'Egypte, l'Indonésie, ou encore le Mexique. Ce profil est idéal selon nos critères, par conséquent on va mener la suite de l'étude sur ces 14 pays uniquement, ce sont nos « pays cibles ». On crée df\_final, le dataframe contenant seulement les pays du troisième cluster.

### 3.4. Calcul du score de marché

Pour modéliser l'analyse, il est idéal de combiner ces trois critères et en créer une variable « score de marché ». Ces critères n'ont pas tous la même importance donc on ne peut pas en faire une moyenne. On doit multiplier les trois critères par des poids respectifs, puis les additionner. Mais on ne peut pas choisir des poids sans fondement. C'est pour cela qu'on utilise la méthode de l'Analytic Hierarchy Process (AHP), développée par Thomas Saaty dans les années 70. Elle consiste à comparer des choses difficilement comparables et d'une importance différente. Pour le faire on utilise la table de Saaty, qui compare chaque critère aux autres et attribue une note d'importance comparée. L'échelle des notes d'importance est la suivante : 1=les deux se valent ; 3=modéré ; 5=fort ; 7=très élevé ; 9=extrême.

Exemple : si  $A$  a une note de ‘5’ comparé à  $B$ , alors  $A$  est fortement plus important que  $B$ . A l'inverse, quand on compare  $B$  à  $A$ , on inverse la note, donc ici ‘1/5’.

Dans notre situation le volume d'import est le plus important (ce que le pays achète), puisqu'on cherche à exporter en masse de manière agressive. Ensuite, la part des imports est le deuxième plus important, puisqu'il donne l'ouverture du marché. Un pays avec une basse dépendance aux imports est un marché fermé aux importateurs. Enfin la demande ou disponibilité, est ce que consomme un pays en un an. Cet indicateur montre la taille du marché, ce qui est intéressant mais ne vaut rien sans les 2 autres critères. On transpose ces rapports d'importance dans une table ou matrice de Saaty :

**Fig 3 : Matrice de jugement de Saaty**

	imports_norm	part_imports	demande_norm
imports_norm	1	3	5
part_imports	1/3	1	3
demande_norm	1/5	1/3	1

*Champs : Importance des critères : import normalisé, part des imports, demande normalisée*

Une fois cette table dressée, on peut calculer les poids en suivant les 3 étapes suivantes :

- On fait la somme de chaque colonne.
- On normalise la matrice en divisant chaque élément par la somme de sa colonne.
- On calcule la moyenne de chaque ligne, c'est le poids du critère de la ligne en question.

Les poids trouvés sont : 0.63 pour le volume des imports, 0.26 pour la part des imports, 0.11 pour le volume de demande normalisée. Le score final pour chaque pays est donc de la forme :  $\text{score\_2017} = 0.63 * \text{imports\_norm} + 0.26 * \text{part\_imports} + 0.11 * \text{demande\_norm}$ .

On ajoute donc la colonne ‘score\_2017’ au datafram df\_final, qu'on renomme ‘top’ et qu'on trie en fonction du score de marché décroissant. Notre analyse économique du dataset DisponibiliteAlimentaire\_2017.csv est terminée.

## 4.Analyse démographique | étude prédictive

Notre étude de marché est terminée pour 2017, mais AIGriTech a besoin de connaître l'évolution de ces marchés dans le futur, car la logistique d'acheminement par voie maritime est coûteuse et longue à mettre en place : la réorganiser prend plus de 5 ans. On est limité par le premier dataset car il ne porte que sur 2017, donc on utilise Population\_2000\_2018.csv.

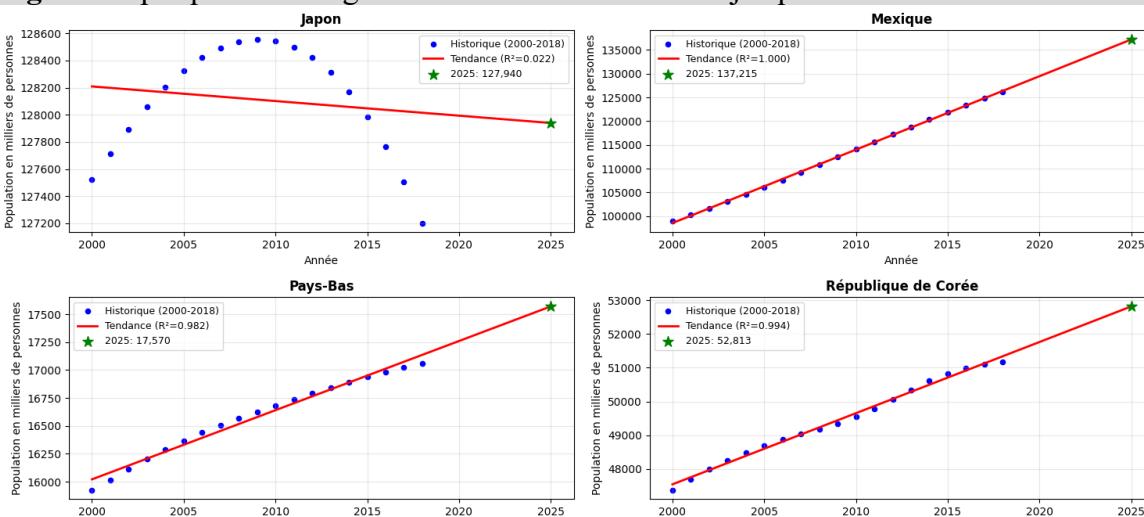
On sait que les biens exportés par notre startup sont essentiels à l'alimentation d'un pays, et que la population entière consomme et dépend de ces céréales. Par conséquent, l'évolution de la démographie de nos pays cible est un indicateur évident de l'intérêt du marché dans le futur. En effet, si la population augmente, la consommation de ces biens augmente proportionnellement et cela est vrai à l'inverse.

De plus ce deuxième dataset nous fournit les données démographiques des pays cibles de 2000 à 2018, nous permettant d'estimer leurs démographies futures, au minimum à l'horizon de 5 ans (temps minimum pour réorganiser la logistique d'acheminement), et maximum à 10 ans, car au-delà les prédictions deviendraient moins précises. On choisit donc de prédire jusqu'en 2025, soit 8 ans dans le futur, grâce à des algorithmes de régression. On filtre df\_pop sur nos 14 pays cibles et on pivote la colonne 'Année' pour les avoir en colonnes individuelles, et on met cela dans le dataframe 'df\_pop\_top'.

### 4.1 Algorithme de régression linéaire

L'algorithme de régression linéaire prend les données historiques pour déterminer la tendance future. Concrètement, il cherche la droite de la forme  $Y = aX + b$ , qui passe le plus près possible de nos points historiques. Il établit aussi un coefficient de détermination  $R^2$  inclus dans [0 ; 1], qui est le « score de fiabilité » du modèle. Plus il est proche de 0, moins il y a de corrélation entre la droite trouvée et les données historiques, et plus il est proche de 1, plus la droite est « alignée » avec les données. On a appliqué ce modèle aux pays cibles, en voici quelques uns :

**Fig 4 : Graphiques de la régression linéaire : estimations jusqu'à 2025**



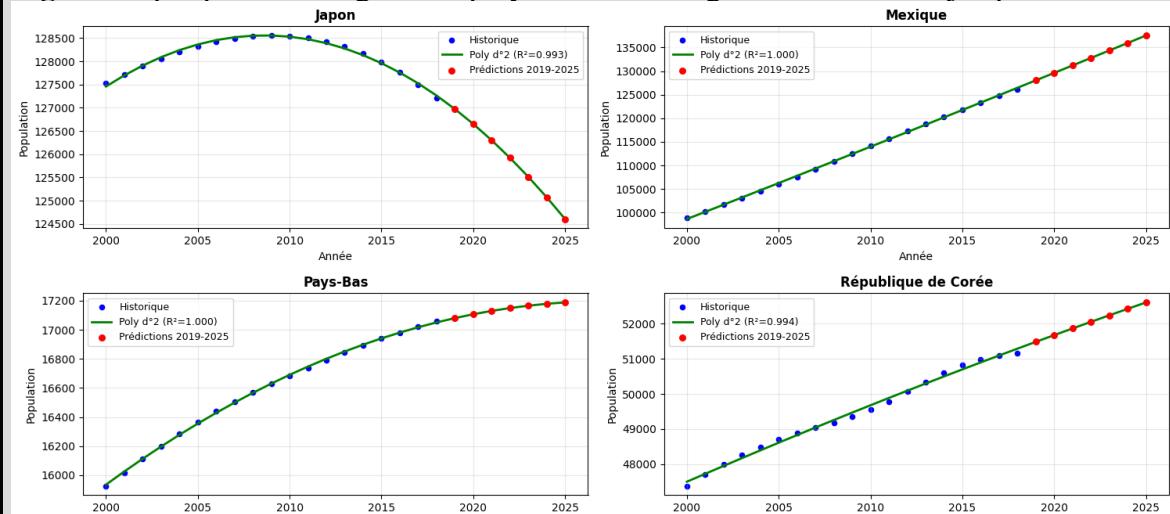
Champs : Données de la population des 14 pays cibles de 2000 à 2018, depuis 'df pop\_top'

On observe que la droite tracée indique la direction générale de l'évolution de la démographie. Les scores  $R^2$  sont variés : le Mexique est à 1.00 car l'évolution de sa population est très linéaire, ses données historiques sont alignées. En revanche, le Japon a un score  $R^2$  médiocre de 0.022, car l'évolution de sa population suit une parabole. Cela montre parfaitement, les limites de la régression linéaire. Le modèle échoue à capturer les effets d'accélération ou de stagnation, créant des prédictions potentiellement fausses.

#### 4.2 Algorithme de régression polynomiale de degré 2

Pour prédire la population future, on peut aussi utiliser un modèle de régression polynomiale, de degré 2. Ce modèle est identique au précédent dans le concept : il utilise les données passées pour prédire les données futures. Mais il ne trace pas une droite, mais une courbe qui épouse la forme des points, grâce un polynôme de degré  $n$ . Ici, on choisit  $n=2$ , car un polynôme de ce degré à une courbe en forme de parabole avec une seule courbure adaptée aux dynamiques démographiques de nos pays. Un degré supérieur pourrait créer un phénomène d'« over fitting », c'est-à-dire que le modèle sera une courbe avec plusieurs points d'inflexion, et qui en cherchant à épouser impeccablement les données, va inclure des variations parasites et potentiellement fausser les prédictions. La courbe aura comme équation :  $Y = aX^2 + bX + c$ . Le coefficient  $R^2$  inclut dans [0 ; 1], respecte les mêmes règles. Voici quelques prédictions :

**Fig 5 : Graphiques de la régression polynomiale de degré 2 : estimations jusqu'à 2025**



Champs : Données de la population des 14 pays cibles de 2000 à 2018, depuis 'df\_pop\_top'

On observe que la courbe tracée permet de prédire la démographie de ces pays jusqu'en 2025 de manière fiable, puisqu'elle prend en compte la forme, la dynamique d'évolution de la démographie. On observe les bien des phénomènes de stagnation (Pays-Bas) et d'accélération (Japon, accélération négative ici). Les scores  $R^2$  sont tous très proche de 1, ce qui montre que le modèle réussit à bien trouver la corrélation des données, et que les prédictions sont beaucoup plus fiables : le score du Japon passe de 0.022 à 0.993 grâce à ce modèle.

### 4.3 Comparaison et conclusion

Comparons les deux modèles : la régression linéaire obtient un coefficient de détermination moyen  $R^2 = 0.85$  pour les 14 pays cibles, alors que la régression polynomiale de degré 2 obtient  $R^2 = 0.98$ , un score proche de la perfection, synonyme d'un risque très faible et des prédictions fiables. On choisit donc de garder les prédictions du second modèle, et on les ajoute dans le dataframe des populations par pays de 2017 à 2025 : ‘pred\_top’. Cela prouve au comité direction que nous ne parions pas sur des marchés instables mais sur des tendances mathématiquement démontrées.

### 4.4 Critères d'analyse des prédictions

Grâce à ces nouvelles prédictions, on peut calculer des nouveaux critères, pour faire une analyse dans le temps de l'évolution du marché des pays cibles, ce qui permet de nuancer les résultats obtenus dans l'analyse économique du marché, qui est seulement en 2017.

Le premier critère est le taux de croissance annuel moyen (TCAM). Il traduit le dynamisme de l'évolution de la démographie d'un pays jusqu'à 2025. Il se calcule grâce à la formule suivante :  $TCAM = (\text{Population 2025} / \text{Population 2017})^{1/8} - 1$ . Il permet de montrer la vitesse de croissance d'un marché, ici dans le futur, mais ne permet pas de quantifier cette croissance.

Le deuxième est le volume brut de la variation de la population en millier de personnes ( $\text{Pop 2025} - \text{Pop 2017}$ ). Ce critère est évidemment essentiel pour notre logique d'export de masse.

Maintenant qu'on a rassemblé tous nos critères d'analyse, on peut créer un dataframe final contenant les données que nous allons utiliser pour prendre les décisions ultimes. On merge le data frame ‘top’ et ‘pred\_top’ pour créer ‘top\_final’. Ce dataframe comporte pour chacun de nos pays cible : le score de marché en 2017, le TCAM 2017-2025, et le volume de croissance.

**Fig 6 :** Dataframe ‘top\_final’

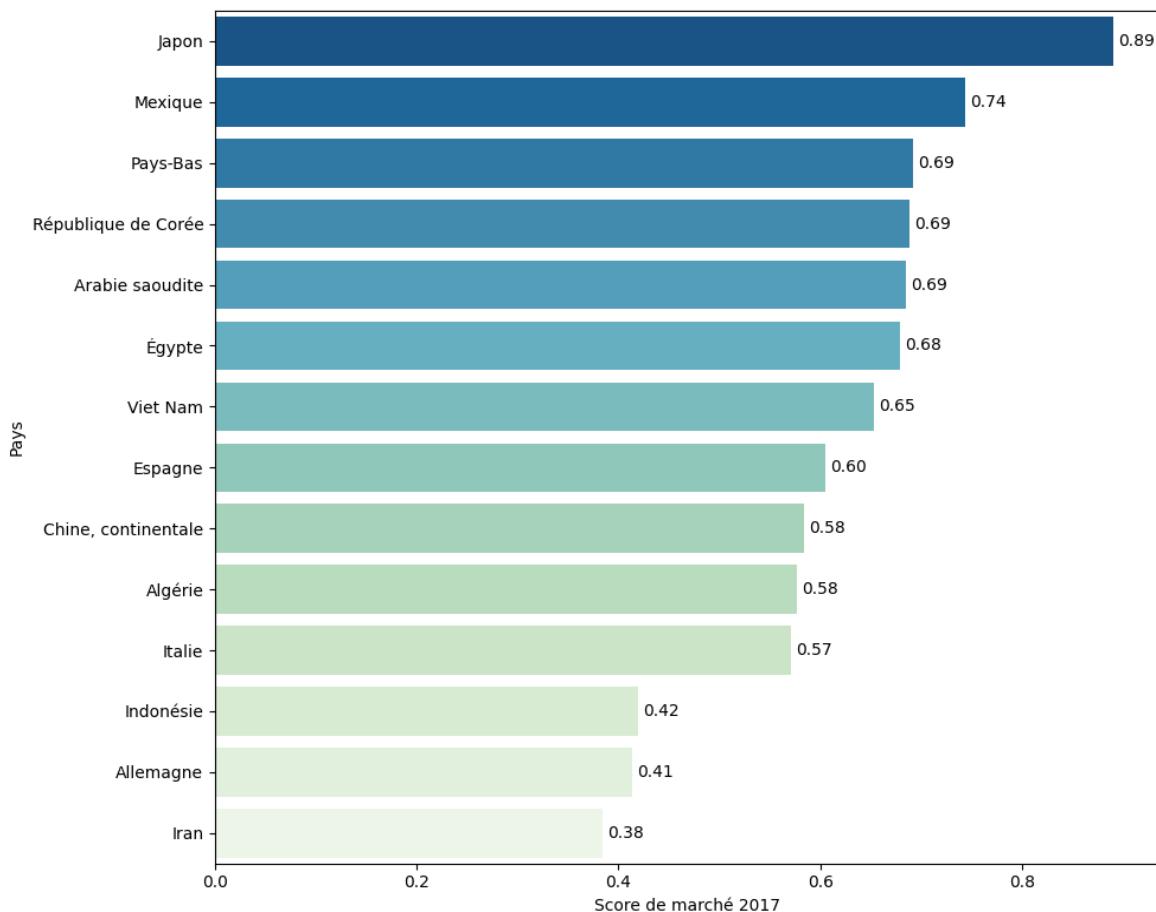
	Zone	TCAM	volume_croissance	score_2017
0	Japon	-0.287140	-2899.623	0.890748
1	Mexique	1.226869	12785.819	0.743526
2	Pays-Bas	0.121649	166.358	0.691542
3	République de Corée	0.365320	1512.556	0.688042
4	Arabie saoudite	2.538005	7349.180	0.685365
5	Égypte	2.265275	18927.810	0.679583
6	Viet Nam	1.079973	8488.996	0.653033
7	Espagne	-1.124393	-4034.535	0.604659
8	Chine, continentale	0.487955	56428.294	0.584299
9	Algérie	2.260565	8104.794	0.576779
10	Italie	0.136449	665.480	0.570987
11	Indonésie	1.249109	27631.761	0.419066
12	Allemagne	0.480317	3230.091	0.413501
13	Iran	1.297396	8763.529	0.384016

## 5.Modélisation & Interprétation Business des analyses

Pour aider le comité direction à visualiser ce dataframe ‘top\_final’, nous avons créé des visualisations et leur interprétation. Enfin on propose des business insights tirés de l’analyse des datasets et enrichis par la prise en compte du contexte réel mondial, pour aider le comité direction à prendre la décision.

### 5.1 Diagrammes

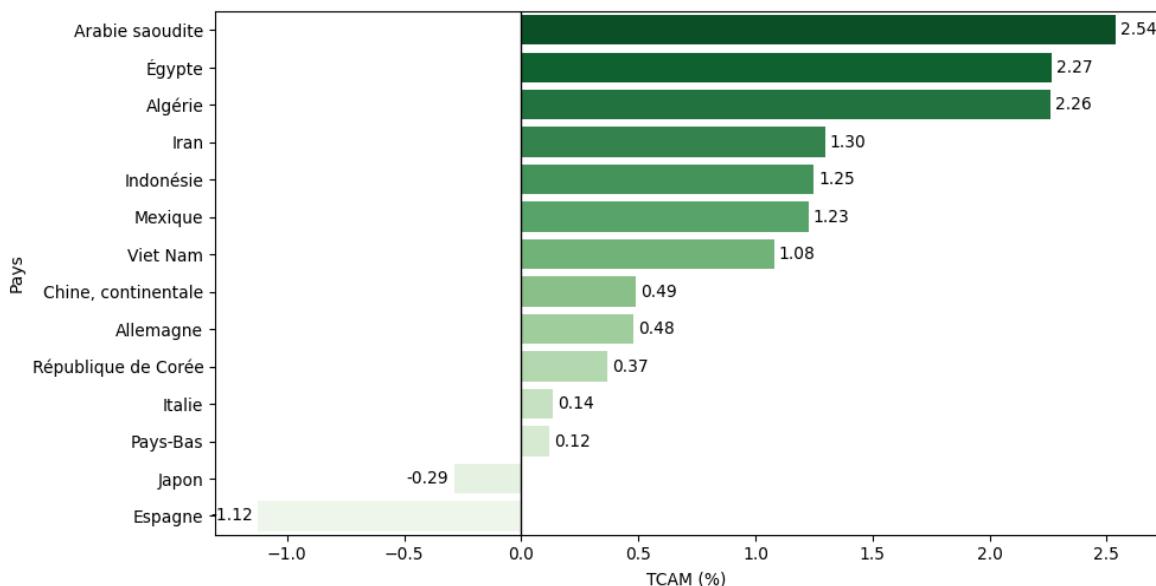
**Fig 6 : Classement des pays cibles selon le score de marché (2017)**



Champs : colonne ‘score\_2017’ dans le dataframe ‘top\_final’

Ici, les pays cibles sont classés selon leur score de marché. Ce score prend en compte les trois critères économiques établis dans l’analyse de ‘DisponibiliteAlimentaire\_2017.csv’ et les multiplie par leur poids (importance) attribués suivant la méthode AHP. Le score de marché traduit l’intérêt actuel du marché. Le Japon mène largement avec un score de 0.89, ce qui veut dire qu’il importe un volume énorme, qu’il est très dépendant des imports, et que la consommation de céréales y est très élevée.

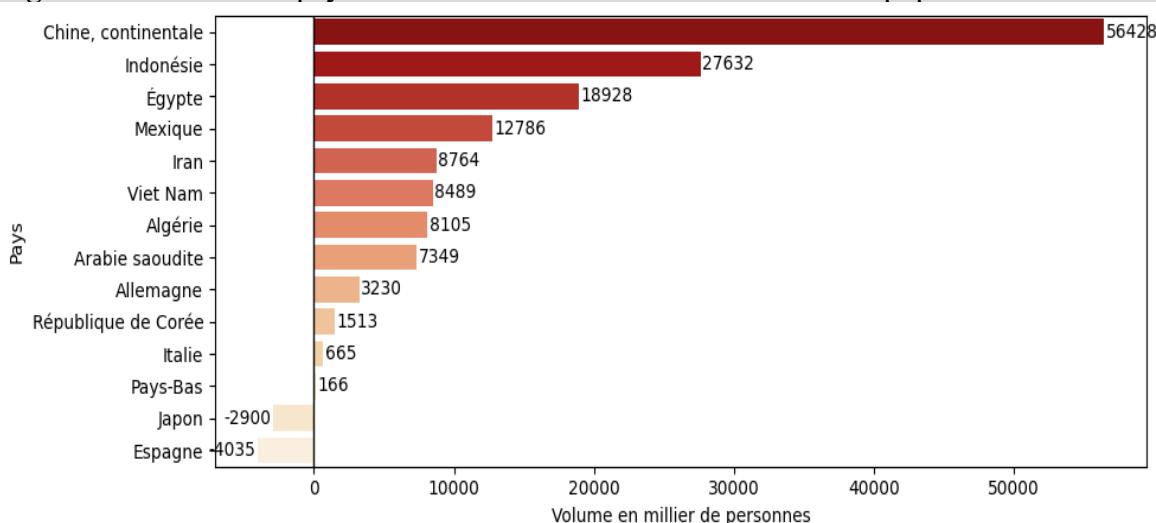
**Fig 7 : Classement des pays cibles selon le TCAM de 2017 à 2025**



*Champs : colonne ‘TCAM dans le dataframe ‘top\_final’*

Ici, les pays sont classés selon leur taux de croissance annuel moyen, de 2017 à 2025, calculé grâce aux prédictions du modèle de régression polynomiale de degré 2. Il montre la vitesse, le dynamisme de l'évolution de la démographie. On peut donc voir que l'Arabie Saoudite est le pays dont la population va augmenter le plus rapidement, avec un TCAM de 2.54%. D'autres pays comme le Japon ou l'Espagne, ont un TCAM négatif, ce qui montre que leur population va baisser, et par conséquent l'intérêt de leur marché aussi.

**Fig 8 : Classement des pays cibles selon le volume de variation de la population 2017-2025**

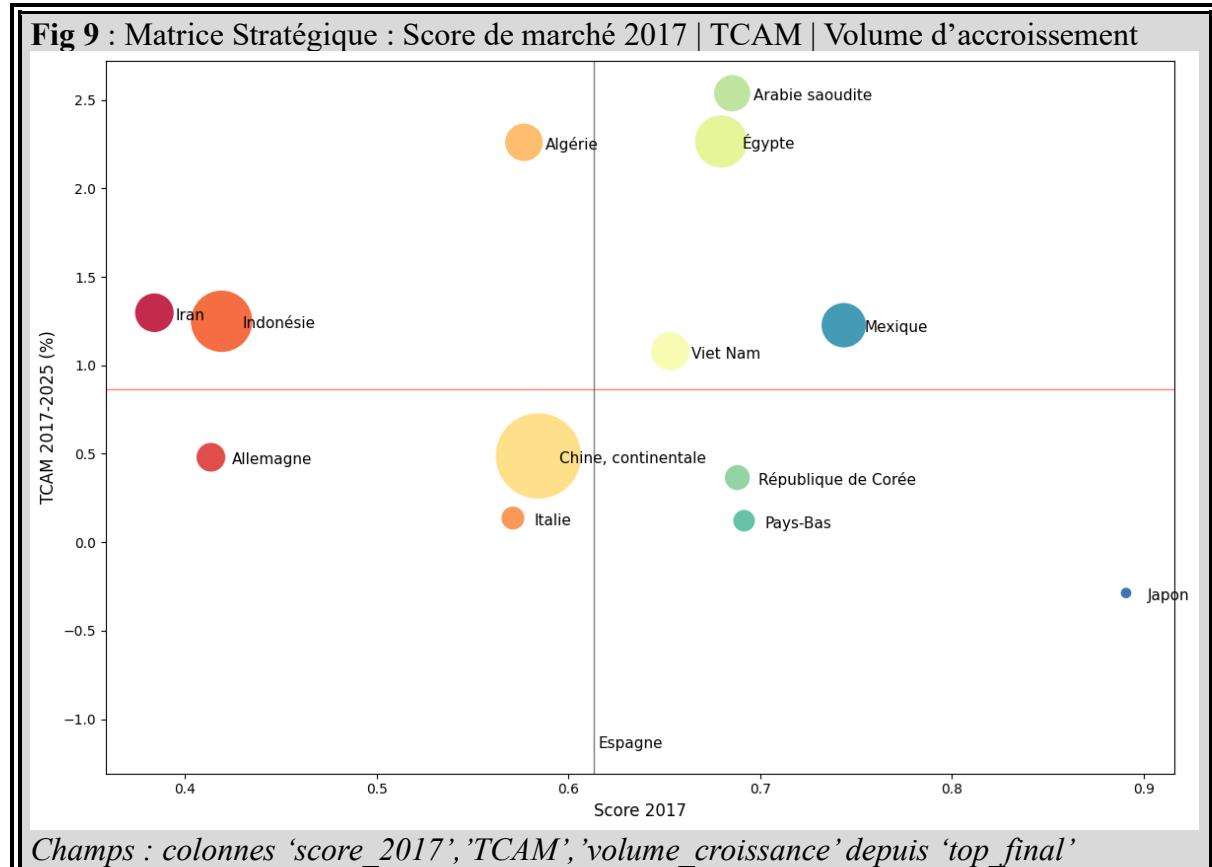


*Champs : colonne ‘volume croissance’ dans le dataframe ‘top\_final’*

Enfin, ce dernier diagramme classe les pays selon le volume brut de la variation de leur population, entre 2017 à 2025, grâce à nos prédictions. Cela quantifie l'évolution de la population, et donc l'intérêt du marché en termes de volume. On observe que la Chine continentale est le pays dont la quantité de population va le plus augmenter d'ici 2025, d'un extrême 56 millions d'habitants. Cette augmentation de population n'est pas négligeable pour notre stratégie qui repose sur l'export en masse.

## 5.2 Matrice stratégique

On combine ces 3 diagrammes en une matrice stratégique.



Cette matrice stratégique est un scatter plot, ou nuage de point, qui place chacun de nos pays cibles dans le graphique en prenant pour abscisse le score de marché, en ordonnée le TCAM, et la taille du point représente le volume d'accroissement en millier de personnes. L'axe noir vertical est la moyenne du score de marché en 2017 pour ces pays, l'axe rouge horizontal est la moyenne du TCAM. Le pays idéal pour notre expansion se trouverais dans le coin supérieur droit. Dans notre cas, le pays qui s'en rapproche le plus est difficile à déterminer donc on considère que les pays dans le cadran supérieur droit, dessiné par les axes, sont les meilleurs (ex : Arabie Saoudite). Ensuite le cadran supérieur gauche : ce sont les pays qui actuellement n'ont pas le meilleur score de marché, mais qui ont un potentiel futur important, souvent des pays en développement avec une forte et volumineuse dynamique démographique (ex : Indonésie). Ensuite le cadran inférieur gauche regroupe les pays avec un profil de marché très attrayant actuellement, mais qui a un potentiel futur très faible voir négatif (ex : Japon). Enfin le cadran inférieur droit regroupe les pays les moins intéressants.

### **5.3 Recommandations business**

Pour concrétiser au maximum l'analyse, il est nécessaire de prendre en compte la géographie, car le transport maritime de céréales est très couteux. Il se fait grâce à des bulk carrier ships, soit des navires de transport de vrac et nécessite une lourde infrastructure portuaire. Notre entreprise basée en France, a accès à deux fronts de mers, la Méditerranée et l'Atlantique. La nuance apportée par la logistique de transport est cruciale et permet d'affiner l'analyse et fournir les recommandations business suivantes :

D'abord, l'analyse déconseille fortement les pays d'Europe mais aussi le Japon et la Corée. Bien que les scores de marché de ces pays soient bons actuellement, leurs dynamiques futures sont peu idéales, TCAM faibles voire négatifs pour l'Espagne et le Japon, et la taille de ces marchés est minimes comparés à leurs concurrents. AIGriTech pourrait envisager d'y exporter maintenant puis se retirer plus tard, mais les coûts logistiques trop élevés ne le permettent pas.

Nous déconseillons également le marché chinois, mais pour d'autres raisons. En effet, selon les prédictions, la Chine témoigne d'une énorme augmentation en volume brut de sa population pour 2025 (56 millions d'habitants). Cela pourrait paraître idéal mais le marché est saturé : la part des imports est faible (4% de la demande), et le développement local de l'agriculture crée une compétition à notre désavantage. Les coûts logistiques sont également trop élevés.

De plus l'Iran est fortement déconseillé. En théorie la dynamique prédictive et le volume d'augmentation sont bons. En réalité ce pays est sous sanctions internationales lourdes, les banques sont coupées du système SWIFT, la juridiction est très compliquée, et la réputation est mauvaise. AIGriTech ne peut pas se permettre de prendre de tels risques.

L'analyse conseille donc dans un premier temps (2017), d'attaquer la zone MENA (Moyen-Orient & Afrique du Nord). Cette zone contient l'Algérie, l'Egypte et l'Arabie Saoudite, des pays avec un profil de marché attrayant (score en 2017 : 0.6 à 0.7), un potentiel futur énorme (TCAM prédit : 2 à 2.5 % | Augmentation prédictive : +30 millions de personnes). Mais surtout la face maritime méditerranéenne de la France permet de mettre en place rapidement, facilement, et à faible coût la logistique d'acheminement.

Dans un second temps (2025), AIGriTech pourrait viser le Mexique, le Vietnam et l'Indonésie. Ces pays sont déjà très intéressants, combinant attractivité économique (scores élevés voire très élevés : 0.74 pour le Mexique), dynamique future très forte (TCAM prédit : 1 à 1.5 %), mais surtout un volume brut d'accroissement de la population énorme (Augmentation prédictive : +50 millions de personnes). Ce sont des pays émergents, qui dépendent énormément de l'importation de denrées alimentaires essentielles comme les céréales. C'est leur distance lointaine de la France qui nous retarde pour entrer sur ces marchés, mais les recettes d'ici 2025 devraient être suffisantes pour mettre en place la logistique de transport.