

Université Cergy Paris



CERGY PARIS

UNIVERSITÉ

Diplôme universitaire :
Data Analyst

Examen
Visualisation de données

11 décembre 2022

Haury Fabien

Examen visualisation de données

Haury Fabien

11 décembre 2022

Table des matières

1	Approche descriptive	2
2	Approche hypothético-déductive	5
3	Approche prédictive	8
4	Recherche de patterns et approche exploratoire	11
	Table des figures	13

Chapitre 1

Approche descriptive

Question 1 :

Quel est le nombre total de tremblement de terre ayant une magnitude supérieure à 5,5 sur l'échelle de Richter enregistré entre 1965 et 2016 ?

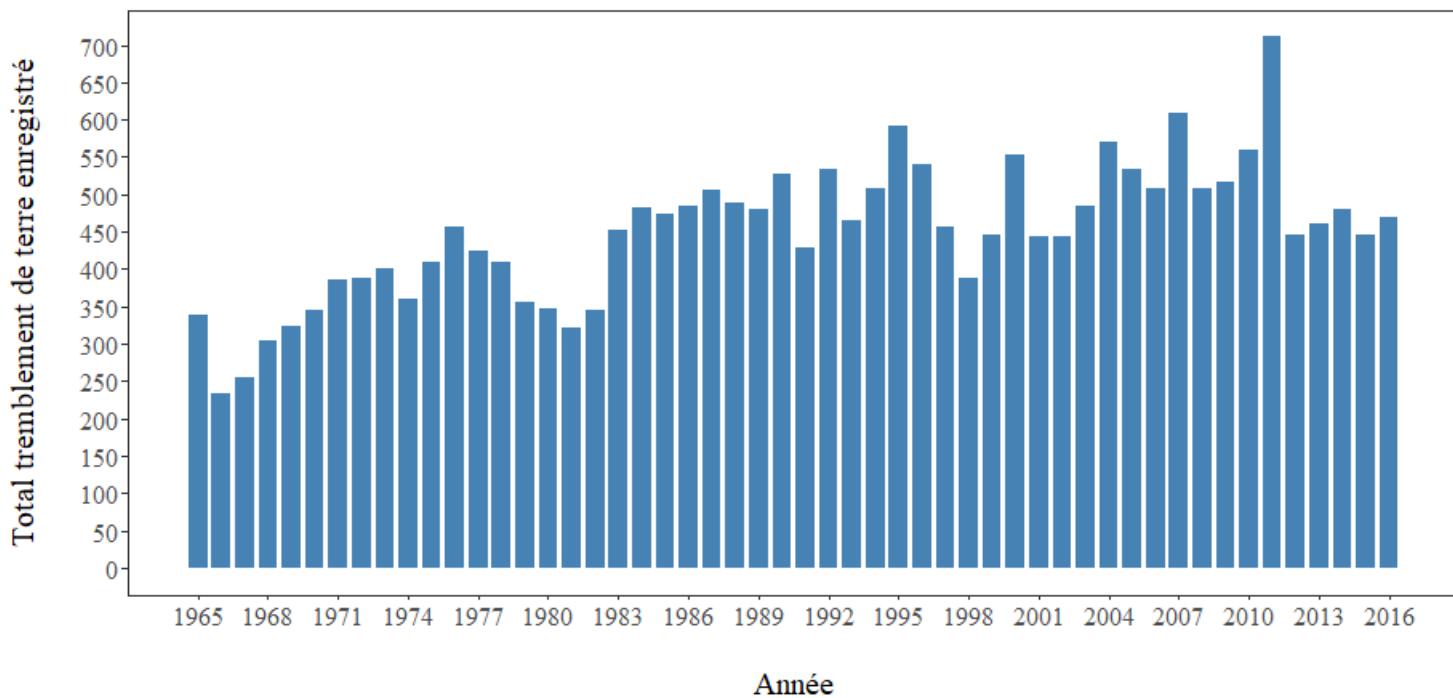


FIGURE 1 – Distribution du total des tremblements de terre pour la période allant de 1965 à 2016, $N = 23406$

La Figure 1 représente le nombre de tremblement de terre enregistré pour la période allant de 1965 à 2016. Il y a en moyenne 450 tremblements de terre par an pour cette période. L'année 2011 est l'année ayant enregistré le plus de tremblement de terre avec un total de 711. L'année 1966 est l'année ayant enregistré le moins de tremblement de terre avec un total de 234.

Question 2 :

Quelle est la représentation spatiale des tremblements de terre recensé en 2016 avec leurs magnitudes respectives ?

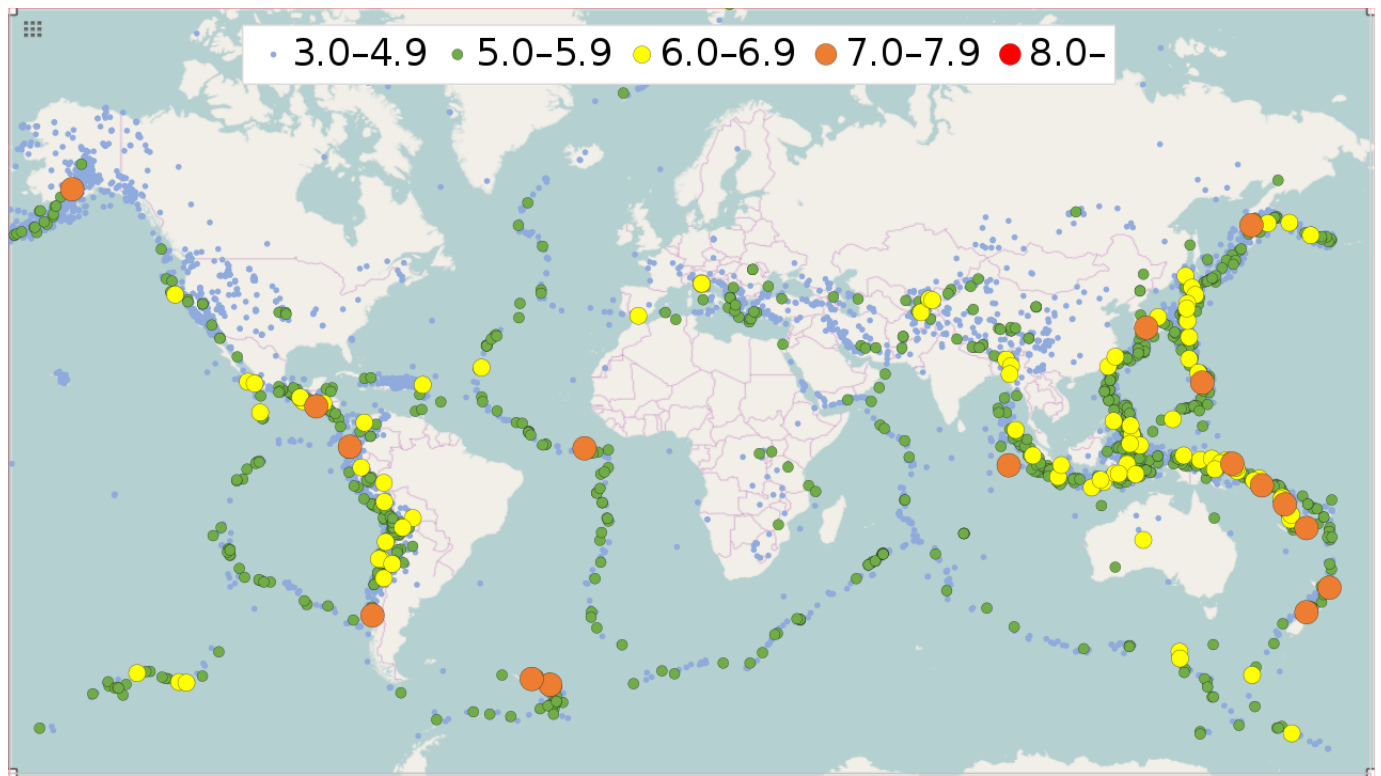


FIGURE 2 – Carte mondiale des tremblements de terre recensé en 2016 avec leurs magnitudes respectives

La Figure 2 représente la carte mondiale des tremblements de terre recensé en 2016 avec leurs magnitudes respectives. Nous pouvons voir que l'extrême majorité de ceux-ci suivent les lignes des plaques tectoniques. Pour ceux ne suivant pas ces lignes, la cause la plus probable est l'activité humaine tel que minage, fracturation hydraulique, etc. Une immense majorité des tremblements de terre sont d'intensité 5,9 ou inférieure. L'ensemble des tremblements de terre d'intensité six ou plus sont le long des lignes des plaques tectoniques et une majorité sont situées en Asie.

Question 3 :

Quelle est la magnitude moyenne des tremblements de terre entre 1965 et 2016 par région ?

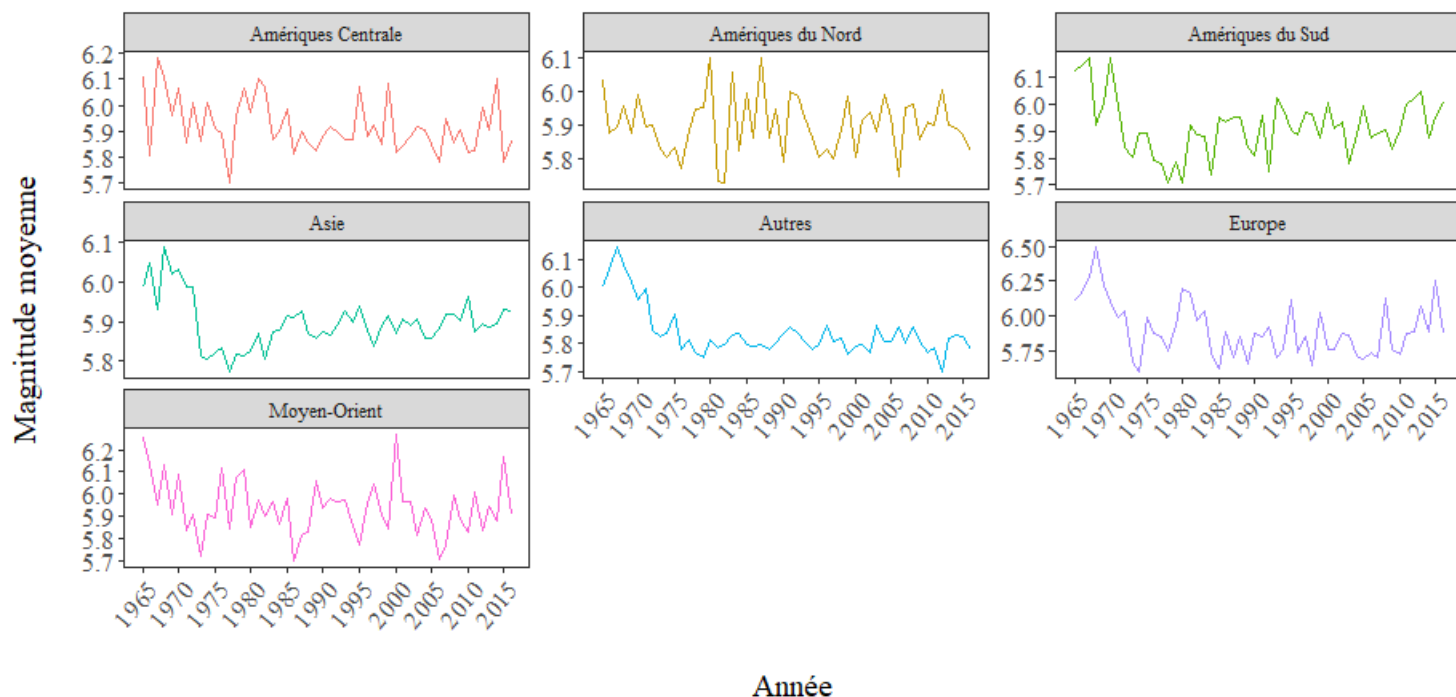


FIGURE 3 – Évolution des moyennes des magnitudes des tremblements de terre par région, période de 1965 à 2016

La figure 3 représente l'évolution des moyennes des magnitudes des tremblements de terre par région pour la période allant de 1965 à 2016. Pour la période de 1965 à 1975, nous pouvons voir que l'Europe possède le plus haut pic avec 6,5 en 1968. Elle possède le plus bas niveau en 1974 avec 5,6. Pour la période de 1976 à 1985, l'Europe possède le plus haut pic avec 6,2 en 1980. Elle possède le plus bas niveau avec 5,62 en 1985. Pour la période de 1986 à 1995, l'Europe possède le plus haut pic avec 6,2 en 1995. Le Moyen-Orient et l'Europe possèdent les plus bas niveaux avec 5,7 en 1986 et 1995 respectivement. Pour la période de 1996 à 2005, le Moyen-Orient possède le plus haut pic avec 6,2 en 2000. L'Europe possède le plus bas niveau avec 5,6 en 1998. Enfin, la période de 2006 à 2016, l'Europe possède le plus haut pic avec 6,3 en 2015. L'Europe et Autres possèdent les plus bas niveaux avec 5,7 en 2007 et 2012 respectivement.

Chapitre 2

Approche hypothético-déductive

Question n°1 :

Pourquoi certaines cours en ligne sont plus achetées que d'autres ?

Hypothèse n°1 :

La note (rating) influence le choix d'achat par le client.

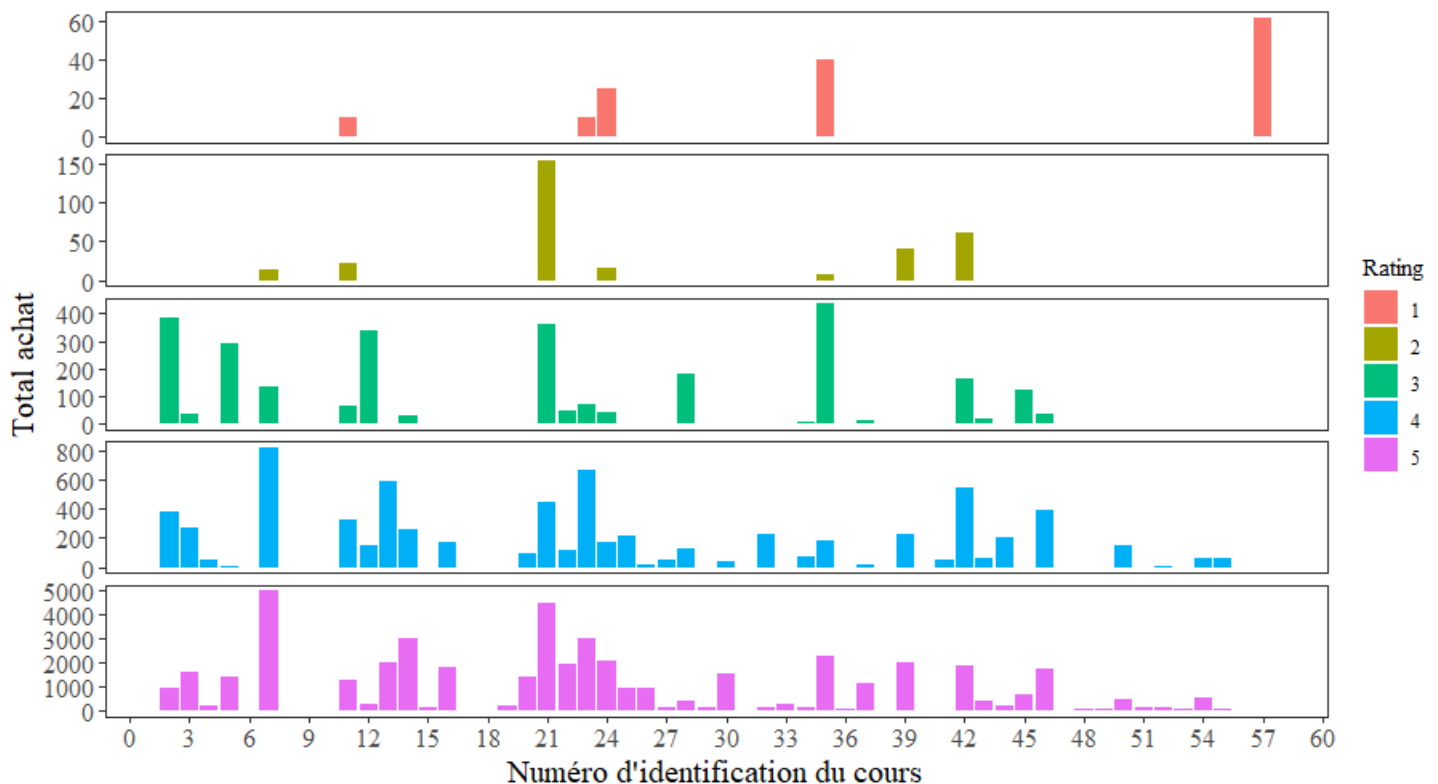


FIGURE 4 – Total des achats pour chaque cours séparé selon ses notes.

La Figure 4 représente le nombre total d'achat pour chaque cours existant selon la note lui correspondant. Nous pouvons voir que un cours avec une noté élevée sera plus acheté qu'un cours ayant une note basse. Les cours ayant une note de cinq sont les cours les plus achetés, avec un maximum de 5000 pour le cours n°7 et un minimum de 24 pour le cours n°24. Il semblerait donc que la note d'un cours influe sur l'achat ou non du cours.

Hypothèse n°2 :

Une période de solde influence le choix d'achat par le client.

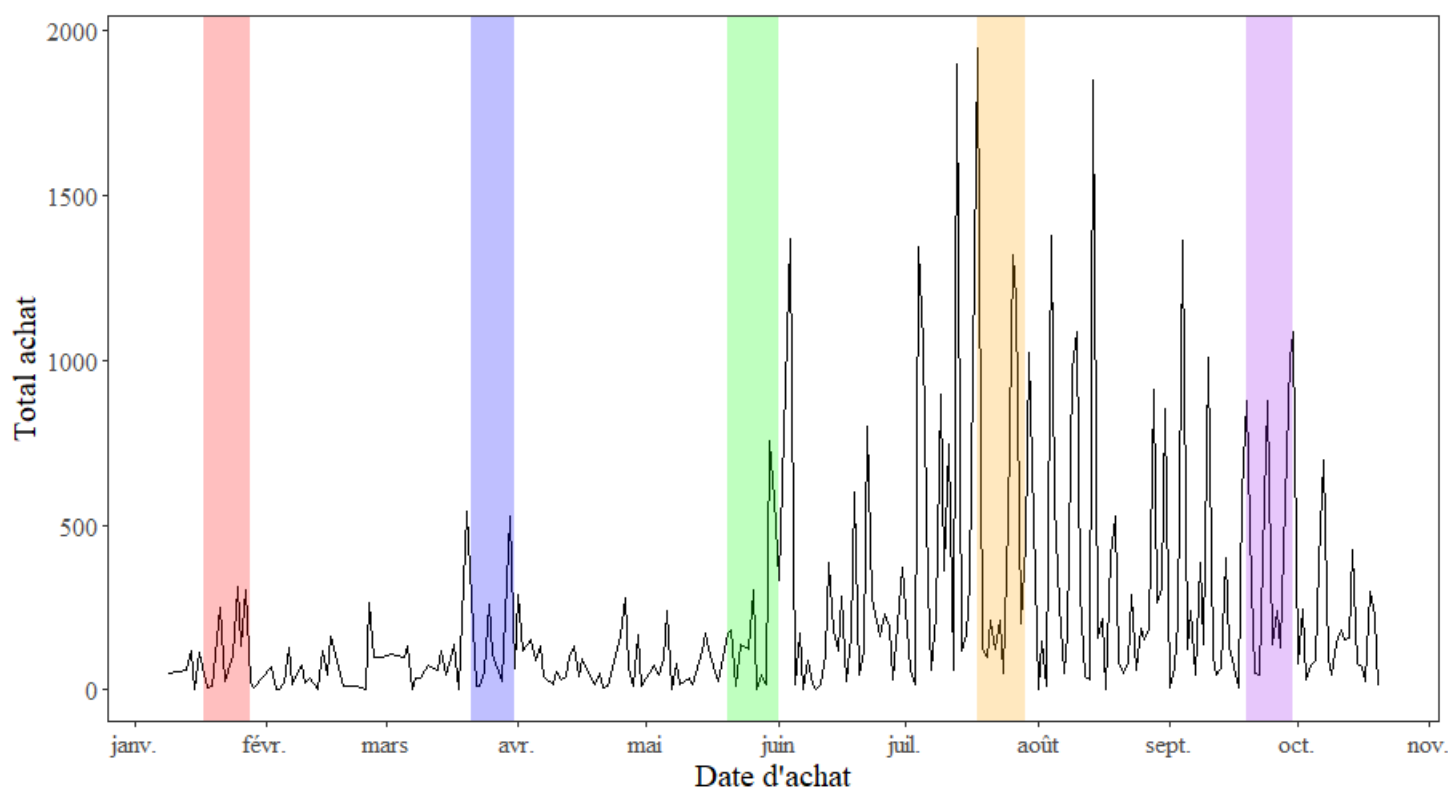


FIGURE 5 – Évolution du total des achats de cours, avec différentes dates de soldes

La Figure 5 représente l'évolution du total d'achat de cours en ligne, avec les périodes de soldes. Nous pouvons voir que le nombre de ventes de cours a augmentée au fil de l'année. Nous pouvons voir que pour chaque période de solde, il existe une hausse des achats de cours, généralement situé vers la fin de la période de solde. Il semblerait donc qu'une période de solde influence modérément le nombre de vente.

Question n°2 :

Pourquoi une formation en ligne n'est-elle pas aussi lucrative que d'autres formations en ligne ?

Hypothèse n°1 :

La durée du cours est soit trop courte, soit trop longue.

La Figure 6 représente le revenu moyen d'un cours en ligne relativement au nombre d'heures de celui-ci. Nous pouvons voir que le nombres d'heures rapportant le plus est compris entre 10 et 25 heures au total.



FIGURE 6 – Revenu moyen relatif au nombre total d’heures consacrées aux cours en ligne

Chapitre 3

Approche prédictive

Question 1 :

Quelles sont les prédictions des prix de l'électricité, du gaz, du système de quotas d'émission et du baril Brent pour la période de 2022 à 2024 en se basant sur la période de 2018 à 2021 ?

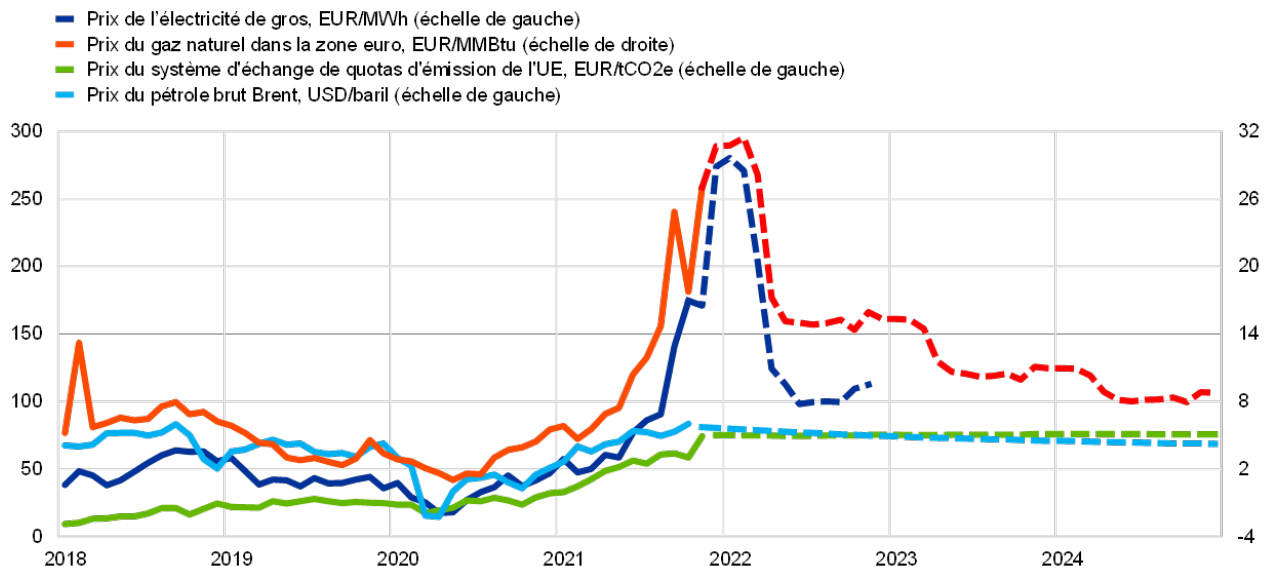


FIGURE 7 – Evolution du prix de l'électricité, du gaz, du système d'échange de quotas d'émissions de CO2 et du pétrole brut Brent pour la période 2018 à 2021 avec leurs prévisions respectives sur trois années

La Figure 7 représente l'évolution du prix de l'électricité, du gaz, du système d'échange de quotas d'émissions de CO2 et du pétrole brut Brent pour la période 2018 à 2021 avec leurs prévisions respectives. Nous pouvons voir que les prévisions pour l'électricité et le gaz montrent un pic pour l'hiver 2021-2022 et décroît sur les trois prochaines années. Nous pouvons voir que les prévisions du prix du système d'échange de quotas d'émissions de CO2 et du prix du baril Brent sont plats.

Question 2 :

Quelle va être l'évolution du nombre de naissances mensuelles à New York pour la période de 1960 à 1970 en se basant sur la période de 1946 à 1959 ?

La Figure 8 représente l'évolution du nombre de naissances mensuelles à New York pour la période 1946 à 1959, avec prédictions sur dix ans. Nous pouvons voir que le nombre de ventes est quasi-nulle et suit un

pattern avec un pic chaque années. Nous pouvons voir une agmentation constante des ventes chaque années. La période prédite après 1993 montre une continuité de cette augmentation et le même pattern.

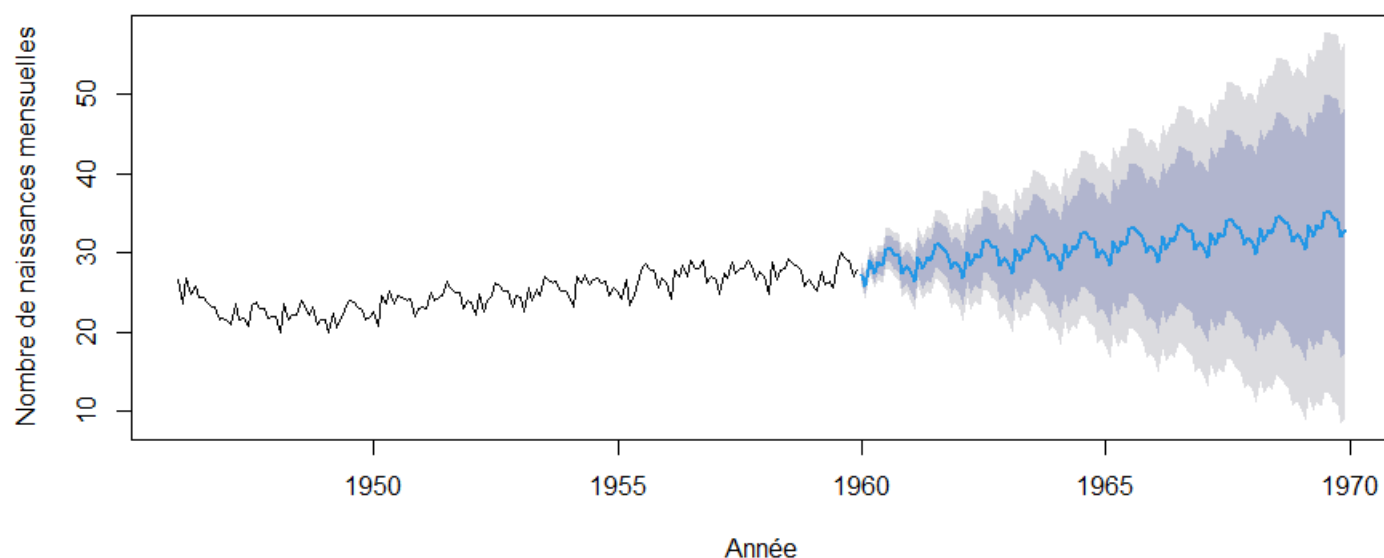


FIGURE 8 – Évolution du nombre de naissances mensuelles à New York pour la période 1946 à 1959, avec prédictions sur dix ans

Question 3 :

Quelle va être l'évolution du nombre de ventes d'un magasin de souvenirs à beach resort town, Queensland, Australie pour la période de 1993 à 2013 en se basant sur la période de 1987 à 1993 ?

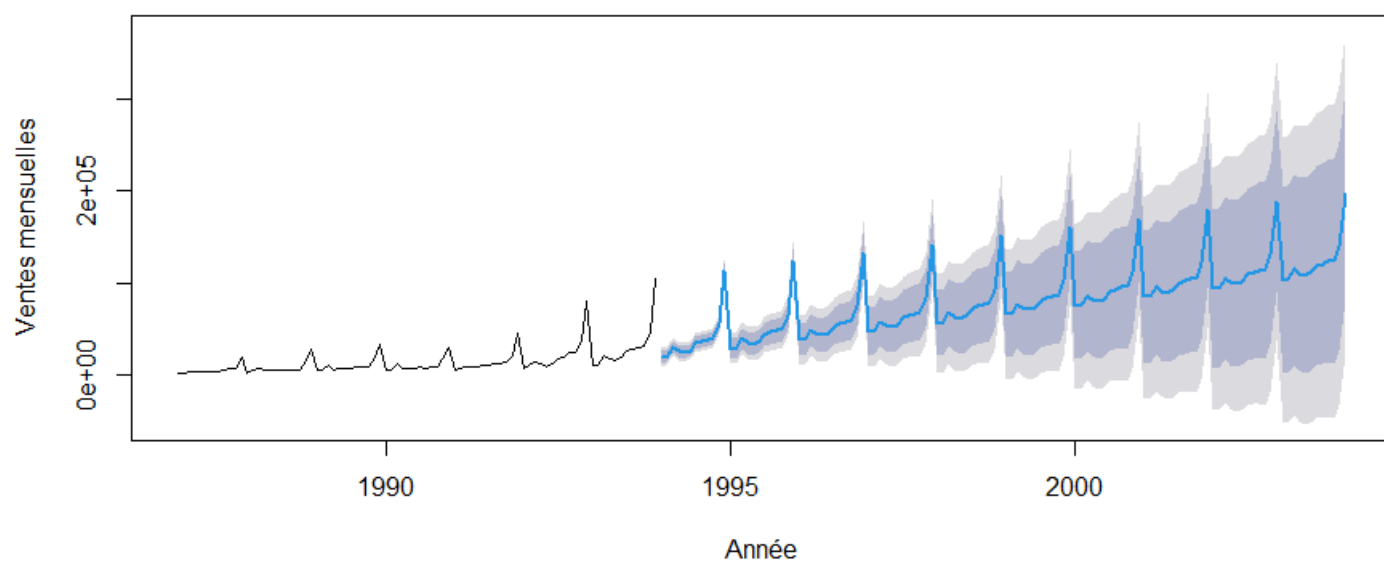


FIGURE 9 – Évolution du nombre de ventes d'un magasin de souvenirs à beach resort town, Queensland, Australie pour la période de 1987 à 1993 avec prédictions sur dix ans

La Figure 9 représente l'évolution du nombre de ventes d'un magasin de souvenirs à beach resort town, Queensland, Australie pour la période de 1987 à 1993 avec prédictions sur dix ans. Nous pouvons voir que le nombre de naissance mensuelles est en légère augmentation constante avec 22 naissances en moyenne pour 1946 et 26 naissances en moyenne pour 1959. La période prédite après 1960 montre une continuité de cette augmentation.

Chapitre 4

Recherche de patterns et approche exploratoire

Question 1 :

Quels facteurs de risque peut-on utiliser pour établir un diagnostic de diabète ?

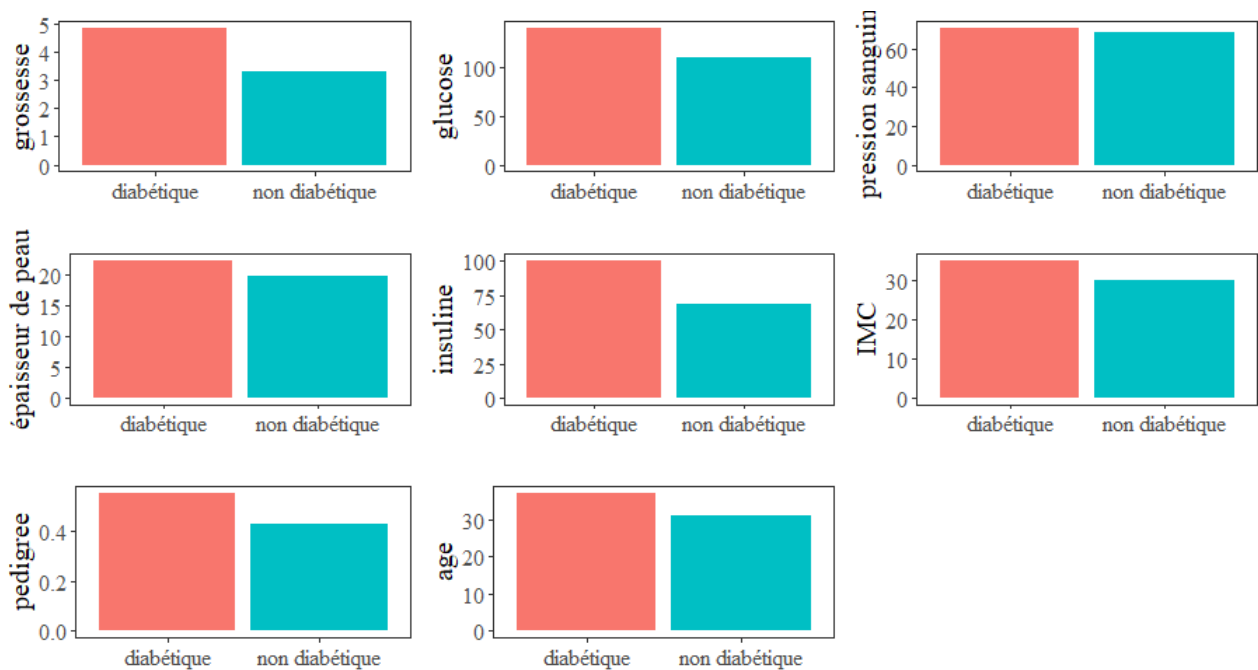


FIGURE 10 – Distribution des moyennes pour différents facteurs de risque en fonction des personnes ayant ou non un diabète

La Figure 10 représente différents facteurs de risque pouvant potentiellement permettre d'établir un diagnostic de diabète.

Question 2 :

Quels facteurs de risque peut-on utiliser pour établir un diagnostic d'un cancer des poumons ?

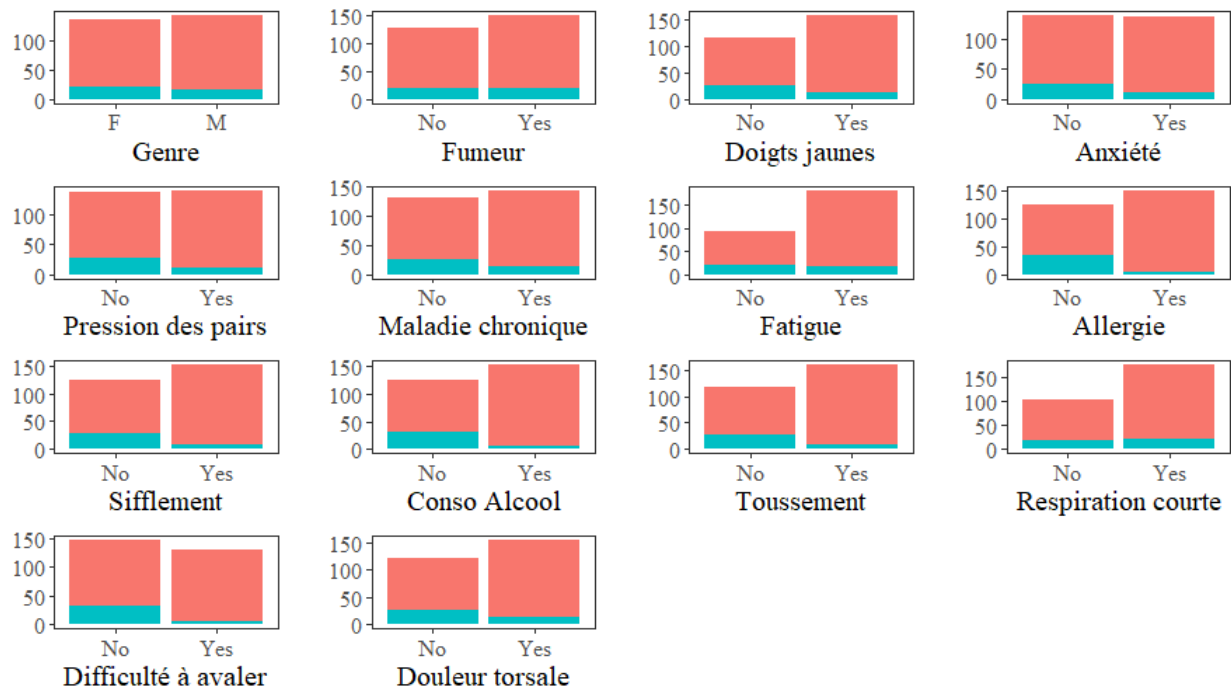


FIGURE 11 – Distribution du total de différents facteurs de risque en fonction des personnes ayant ou non un cancer des poumons , orange : cancer, bleu : pas de cancer

La Figure 11 représente différents facteurs de risque pouvant potentiellement permettre d'établir un diagnostic de cancer des poumons.

Question 3 :

Quels sont les facteurs de risque pouvant déclencher un AVC ?

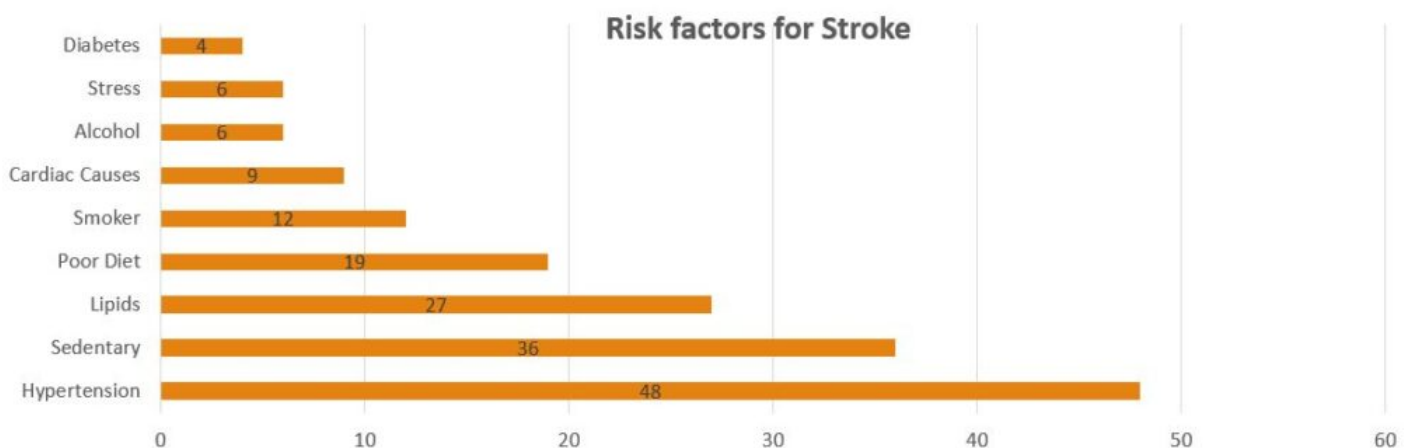


FIGURE 12 – Distribution de différents facteurs de risques pouvant entraîner un AVC

La Figure 12 représente différents facteurs de risques pouvant entraîner un AVC. Nous pouvons voir que souffrir d'hypertension est le facteur principale entraînant un AVC. Souffrir de diabète n'est pas un facteur majeur pouvant amener à un AVC.

Table des figures

1	Distribution du total des tremblements de terre pour la période allant de 1965 à 2016, N = 23406	2
2	Carte mondiale des tremblements de terre recensé en 2016 avec leurs magnitudes respectives .	3
3	Évolution des moyennes des magnitudes des tremblements de terre par région, période de 1965 à 2016	4
4	Total des achats pour chaque cours séparé selon ses notes.	5
5	Évolution du total des achats de cours, avec différentes dates de soldes	6
6	Revenu moyen relatif au nombre total d'heures consacrées aux cours en ligne	7
7	Evolution du prix de l'électricité, du gaz, du système d'échange de quotas d'émissions de CO2 et du pétrole brut Brent pour la période 2018 à 2021 avec leurs prévisions respectives sur trois années	8
8	Évolution du nombre de naissances mensuelles à New York pour la période 1946 à 1959, avec prédictions sur dix ans	9
9	Évolution du nombre de ventes d'un magasin de souvenirs à beach resort town, Queensland, Australie pour la période de 1987 à 1993 avec prédictions sur dix ans	9
10	Distribution des moyennes pour différents facteurs de risque en fonction des personnes ayant ou non un diabète	11
11	Distribution du total de différents facteurs de risque en fonction des personnes ayant ou non un cancer des poumons , orange : cancer, bleu : pas de cancer	12
12	Distribution de différents facteurs de risques pouvant entraîner un AVC	12