

Université du Québec en Outaouais

Département d'informatique et d'ingénierie

**Examen final du cours d'Analyse statistique et
visualisation de données**

TITRE : Analyse Intégrée des Données de Bornes de Recharge Électrique

Projet final – Examen de fin de session

ÉnerData Canada

INF1833 – Automne 2025

Membres de l'équipe :

1. Apetoh, Komlan Kévin
2. Manfouo Nzoné, Boris
3. Tambat, Tresor Megane
4. Nyami Ndjeukou Djokomeni, Annie Christelle

Professeur : Pr. Daniel N'dah Yapi, Ph.D.

Date : 30 novembre 2025

Table des matières

1	Introduction	4
1.1	Contexte général	4
1.2	Problématique et enjeux	4
1.3	Questions de recherche	4
1.4	Objectifs du projet	5
1.5	Organisation du rapport	5
2	Méthodologie et description du jeu de données	6
2.1	Source des données	6
2.2	Description des variables	6
2.3	Méthodologie	8
2.3.1	3.1 Préparation des données	8
2.3.1.1	3.1.1 Vérification de la complétude et nettoyage	8
2.3.1.2	3.1.2 Création de la série temporelle	8
2.3.1.3	3.1.3 Transformation pour la chaîne de Markov	9
2.3.1.4	3.1.4 Préparation pour l'ANOVA	9
2.3.1.5	3.1.5 Préparation pour l'approche bayésienne	10
2.3.2	3.2 Méthodes statistiques mobilisées	10
2.3.3	3.3 Justification des choix méthodologiques	10
3	Analyse de variance (ANOVA)	11
3.0.1	3.1 Hypothèses statistiques	11
3.0.1.1	3.1.1 Hypothèses des tests	11
3.0.2	3.2 Vérification des conditions de l'ANOVA	11
3.0.2.1	3.2.1 Normalité des résidus	11
3.0.2.2	3.2.2 Homogénéité des variances	12
3.0.2.3	3.2.3 Indépendance	12
3.0.3	3.3 Résultats de l'ANOVA à deux facteurs	12
3.0.3.1	3.3.1 Interprétation des résultats	13
3.0.4	3.4 Tests post-hoc (Tukey HSD)	13
3.0.5	3.5 Implications opérationnelles pour ÉnerData Canada	13
3.0.6	4.6 Conclusion de la section ANOVA	14
4	Processus stochastiques : modélisation de la demande, stabilité et prévisions	15
4.0.1	4.1 Analyse par chaîne de Markov	15
4.0.1.1	Construction des états	15
4.0.1.2	Matrice de transition	15

4.0.1.3	Distribution stationnaire	16
4.0.2	4.2 Modélisation ARIMA	16
4.0.2.1	Stationnarité et prétraitement	17
4.0.2.2	Décomposition temporelle	17
4.0.2.3	Sélection du meilleur modèle ARIMA	17
4.0.2.4	Prévision à 10 jours	17
4.0.2.5	Prévisions segmentées	18
4.0.3	Analyse globale : stabilité, cycles et tendance	18
5	Approche bayésienne : probabilité de défaillance après alerte et rôle des croyances a priori	20
5.0.1	Probabilité qu'une borne soit défectueuse après une alerte	20
5.0.2	Probabilité après deux alertes consécutives	20
5.0.3	Sensibilité aux croyances a priori	21
5.0.4	Implications opérationnelles pour ÉnerData	22
6	Visualisation et communication des résultats	24
6.0.1	Visualisation des effets ANOVA et interactions	24
6.0.2	Visualisation des prévisions temporelles (ARIMA)	25
6.0.3	Comparaison : Demande actuelle vs prévisions	26
6.0.4	Heatmaps Province \times Type de borne	27
6.0.5	Visualisation géospatiale : Carte de satisfaction	28
6.0.6	Interprétation synthétique des visualisations	29
6.0.7	Conclusion de la section	29
7	Conclusion et recommandations	30
7.0.1	Synthèse générale des résultats	30
7.0.2	Détection des anomalies (Bayes)	30
7.0.3	Comparaison des segments (Province \times Type de borne)	31
7.0.4	Recommandations stratégiques	31
7.0.5	Conclusion générale	32

1 Introduction

1.1 Contexte général

L'entreprise ÉnerData Canada gère un réseau de bornes de recharge rapide réparties au Québec (QC), en Ontario (ON) et en Colombie-Britannique (BC). En raison de l'essor des véhicules électriques, les besoins en infrastructures fiables et performantes augmentent rapidement. Depuis 2024, ÉnerData observe toutefois des variations importantes dans la performance de ses bornes selon la province, le type de borne (A : 50 kW, B : 150 kW), les conditions météorologiques et le comportement des utilisateurs. Ces variations influencent directement la durée de recharge, la satisfaction des usagers et la demande énergétique.

La direction souhaite donc mieux comprendre ces tendances afin d'optimiser ses décisions de maintenance, d'investissement et de déploiement futur.

1.2 Problématique et enjeux

Malgré la quantité importante de données collectées quotidiennement, ÉnerData fait face à plusieurs défis :

- Les performances diffèrent significativement selon les régions et les types de bornes, mais l'ampleur réelle de ces différences reste incertaine.
- La demande énergétique varie au fil du temps, et il est difficile de prévoir son évolution sans modèle statistique rigoureux.
- Le système de détection des anomalies génère des alertes capteurs, mais l'entreprise ignore le niveau réel de fiabilité de ces alertes.
- La prise de décision nécessite des visualisations claires, permettant d'identifier rapidement les tendances interrégionales et les zones critiques.

Sans analyse statistique intégrée, les décisions stratégiques risquent d'être basées sur des impressions plutôt que sur des preuves quantitatives.

1.3 Questions de recherche

L'étude cherche à répondre aux questions suivantes :

1. Les différences de performance observées entre provinces et entre types de bornes sont-elles statistiquement significatives ?
2. La demande énergétique quotidienne (*sessions*) suit-elle une dynamique stable, cyclique ou instable ?
3. Un modèle stochastique (chaîne de Markov ou ARIMA) permet-il de prévoir efficacement la demande future ?
4. Quelle est la fiabilité d'une alerte « surtension » pour diagnostiquer une borne réellement défectueuse ?

5. Quelles visualisations permettent le mieux de comparer les régions et d'appuyer les décisions stratégiques ?

1.4 Objectifs du projet

Ce projet vise à produire une analyse statistique complète permettant à ÉnerData :

- d'évaluer l'impact de la province et du type de borne sur la durée moyenne de recharge (ANOVA à deux facteurs) ;
- de modéliser la dynamique temporelle de la demande soit par chaîne de Markov, soit par modèle ARIMA, selon la structure de la série ;
- d'estimer la probabilité qu'une borne soit défectueuse après une ou plusieurs alertes (approche bayésienne) ;
- de visualiser clairement les performances interrégionales et les zones prioritaires ;
- de formuler des recommandations stratégiques appuyées par des résultats quantitatifs.

1.5 Organisation du rapport

Le rapport est structuré comme suit :

- **Section 1** : Introduction, problématique, questions de recherche et objectifs.
- **Section 2** : Méthodologie et description des données (source, variables et préparation).
- **Section 3** : Analyse de variance (ANOVA) sur le temps de recharge en fonction de la province et du type de borne.
- **Section 4** : Processus stochastiques (chaîne de Markov et modèle ARIMA) pour la modélisation de la demande.
- **Section 5** : Approche bayésienne pour l'évaluation de la fiabilité des alertes « sur-tension ».
- **Section 6** : Visualisation et discussion critique des résultats.
- **Section 7** : Conclusion générale et recommandations pour ÉnerData Canada.

2 Méthodologie et description du jeu de données

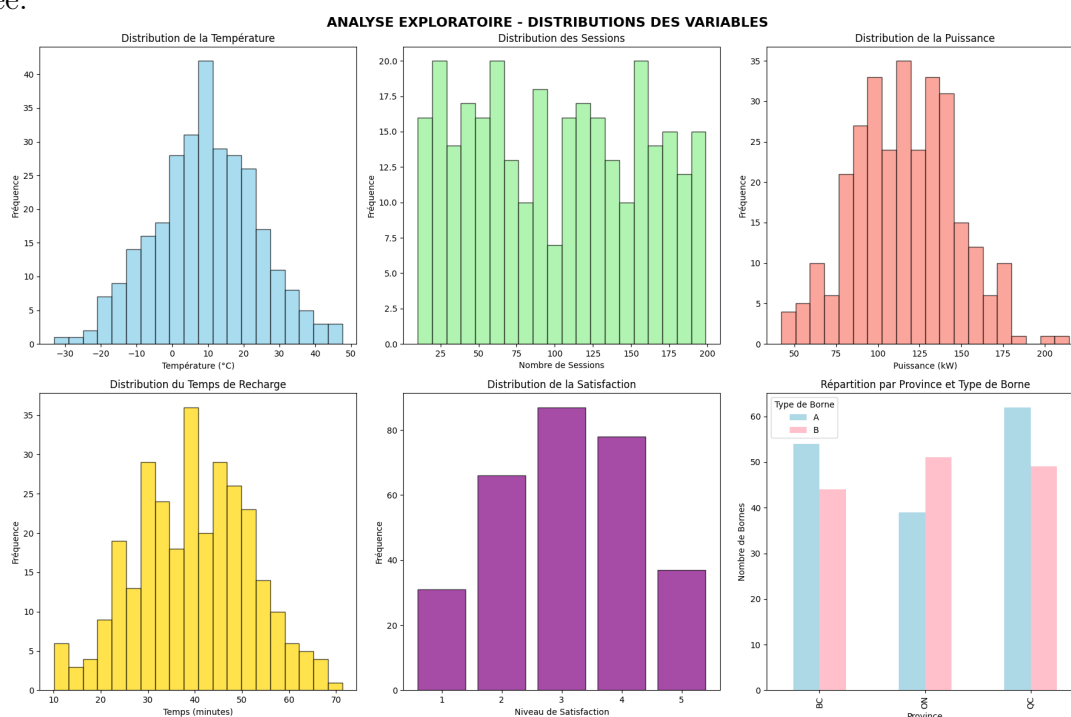
2.1 Source des données

Les données utilisées dans ce projet proviennent d'un jeu fourni dans le cadre du cours INF1833 – *Analyse statistique et visualisation de données*. Elles simulent l'activité quotidienne d'un réseau de bornes de recharge rapide opérées par l'entreprise ÉnerData Canada au cours de l'année 2024.

Le jeu de données inclut des observations provenant de trois provinces canadiennes :

- Québec (QC) ;
- Ontario (ON) ;
- Colombie-Britannique (BC).

Chaque observation représente les mesures agrégées d'une journée pour une borne donnée.

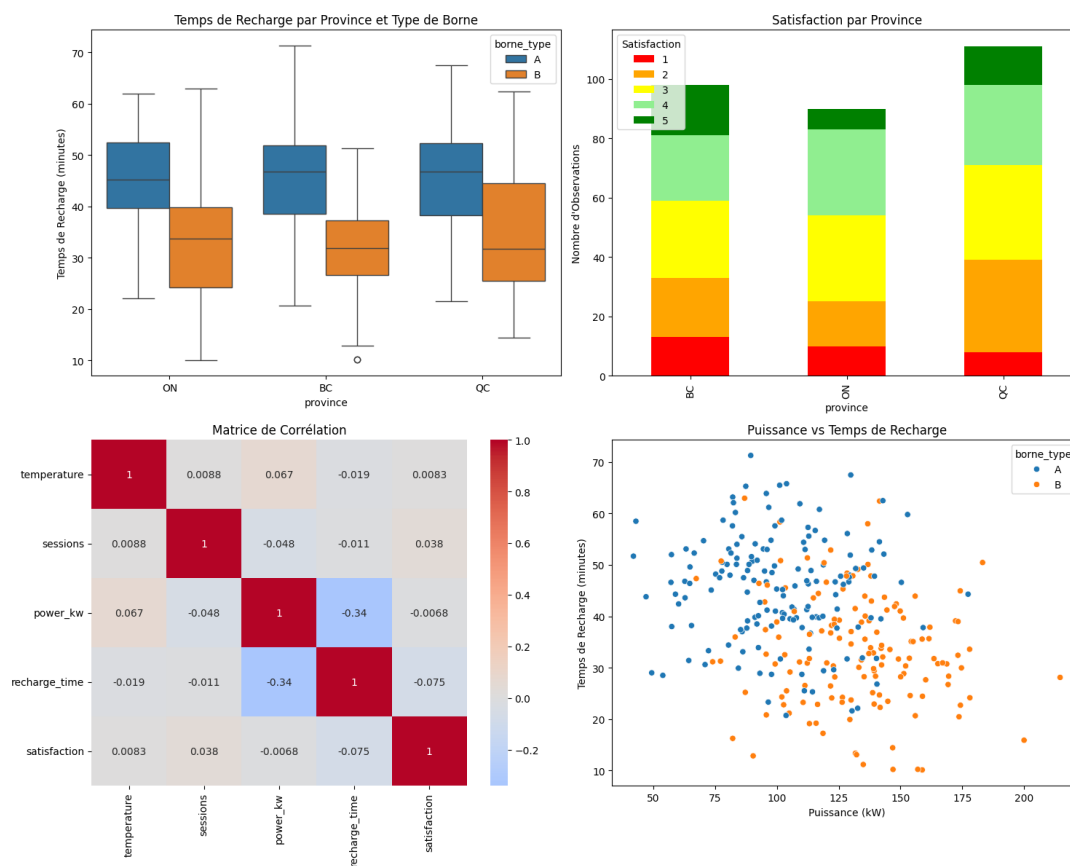


2.2 Description des variables

Le tableau 2.1 présente les variables incluses dans l'analyse.

Variable	Type	Description
province	Catégorielle	Localisation géographique (QC, ON, BC).
borne_type	Catégorielle	Type de borne : A (50 kW) ou B (150 kW).
temperature	Numérique	Température moyenne extérieure (en degrés Celsius).
sessions	Numérique	Nombre moyen de sessions de recharge par jour.
power_kw	Numérique	Puissance moyenne délivrée par la borne (en kW).
recharge_time	Numérique	Durée moyenne d'une recharge (en minutes).
satisfaction	Ordinale (1 à 5)	Niveau moyen de satisfaction des usagers (échelle de 1 = très insatisfait à 5 = très satisfait).

TABLE 2.1 – Description des variables du jeu de données ÉnerData Canada.



2.3 Méthodologie

3.1 Préparation des données

3.1.1 Vérification de la complétude et nettoyage

Les étapes suivantes ont été réalisées :

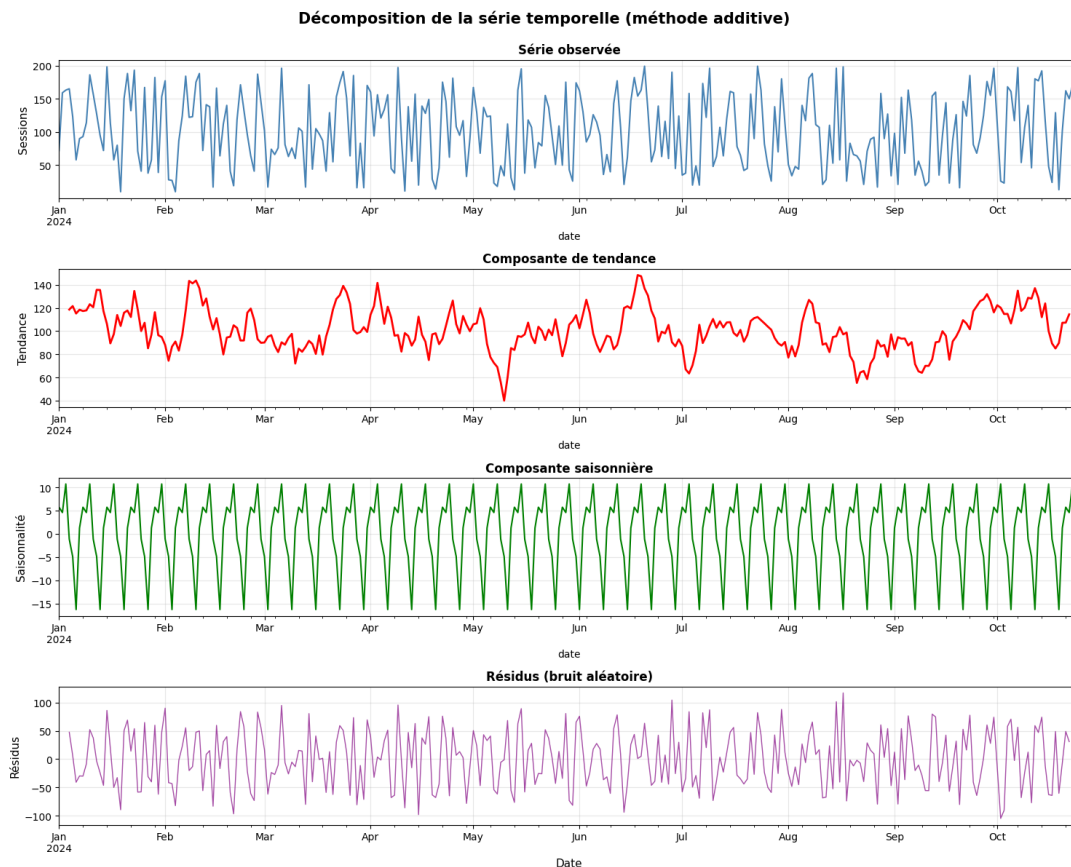
- Vérification de l'absence de valeurs manquantes (`df.isna().sum() = 0`).
- Suppression d'une valeur aberrante identifiée dans `recharge_time`.
- Ajustement des types de variables, notamment la conversion de la date en index temporel pour l'analyse ARIMA.
- Conservation des valeurs extrêmes, car elles représentent des conditions d'utilisation réelles (ex. : températures très basses au Québec).

Statistiques descriptives:					
	temperature	sessions	power_kw	recharge_time	satisfaction
count	299.000000	299.000000	299.000000	299.000000	299.000000
mean	9.347464	101.655518	115.945489	39.615684	3.080268
std	14.286772	55.596688	30.338399	12.182979	1.178774
min	-33.018374	10.000000	41.900000	10.102919	1.000000
25%	0.262451	51.000000	95.035825	30.830628	2.000000
50%	9.400000	102.000000	114.774005	39.656089	3.000000
75%	19.040497	152.500000	137.619414	48.298164	4.000000
max	47.571296	199.000000	214.668522	71.300000	5.000000

3.1.2 Création de la série temporelle

Dans le cadre de la Partie 2, nous avons construit une série temporelle journalière :

- Génération d'une colonne `date` couvrant une séquence de 30 jours.
- Indexation de la variable `sessions` sur cette série afin de permettre l'application des processus stochastiques.



3.1.3 Transformation pour la chaîne de Markov

La modélisation Markovienne nécessite une variable discrète. Ainsi, la série `sessions` a été discrétisée en trois états :

- **faible** : $sessions < 80$
- **moyen** : $80 \leq sessions < 140$
- **élevé** : $sessions \geq 140$

Cette transformation a permis d'estimer la matrice de transition et la distribution stationnaire.

3.1.4 Préparation pour l'ANOVA

Pour l'analyse de variance à deux facteurs (province \times type de borne), les prérequis suivants ont été vérifiés :

- Variable dépendante quantitative : `recharge_time`.
- Facteurs catégoriels : `province` et `borne_type`.
- Normalité des résidus (test de Shapiro–Wilk + QQ-plots).
- Homogénéité des variances (test de Levene).
- Indépendance des observations, garantie par la structure du jeu simulé.

3.1.5 Préparation pour l'approche bayésienne

Cette partie repose exclusivement sur les probabilités fournies dans l'énoncé :

$$P(D) = 0.06, \quad P(A|D) = 0.9, \quad P(A|\neg D) = 0.15.$$

Aucune transformation des données n'était nécessaire.

3.2 Méthodes statistiques mobilisées

Les approches suivantes ont été utilisées :

- **ANOVA à deux facteurs** : effet de la province, du type de borne et de leur interaction.
- **Chaîne de Markov** : estimation de la matrice de transition et de la distribution stationnaire.
- **Modèle ARIMA** : sélection des paramètres (p, d, q) via ACF/PACF et Grid Search, prévision à 10 jours, diagnostic des résidus.
- **Approche bayésienne** : calcul de $P(D|A)$ et $P(D|\neg A)$ pour l'aide à la décision.
- **Visualisation** : boxplots, heatmaps, courbes temporelles, histogrammes et graphiques comparatifs.

3.3 Justification des choix méthodologiques

Les choix méthodologiques sont motivés par :

- Pertinence de l'ANOVA pour une variable dépendante continue et deux facteurs catégoriels.
- Capacité des chaînes de Markov à modéliser des transitions d'états (faible \rightarrow moyen \rightarrow élevé).
- Efficacité d'ARIMA pour modéliser des séries stationnaires et produire des prévisions robustes.
- Utilité de Bayes pour raisonner en contexte d'incertitude (alertes capteurs).
- Importance des visualisations pour soutenir les décisions stratégiques.

3 Analyse de variance (ANOVA)

Cette section étudie l'effet de deux facteurs (la province et le type de borne) sur la durée moyenne de recharge. L'objectif est d'évaluer si les performances observées dans le réseau d'ÉnerData Canada diffèrent selon la localisation ou la technologie utilisée, et d'en tirer des implications opérationnelles.

3.1 Hypothèses statistiques

Deux facteurs explicatifs sont considérés :

- **Province** : QC, ON, BC
- **Type de borne** : A (50 kW), B (150 kW)

La variable dépendante est la durée moyenne de recharge (`recharge_time`, en minutes).

3.1.1 Hypothèses des tests

Effet principal : province

$$H_{0_1} : \mu_{QC} = \mu_{ON} = \mu_{BC}$$

Interprétation : la province n'a aucun effet sur la durée de recharge.

Effet principal : type de borne

$$H_{0_2} : \mu_A = \mu_B$$

Interprétation : le type de borne n'a aucun effet sur la durée de recharge.

Interaction

Il n'existe aucune interaction entre province et type de borne.

Autrement dit, l'effet du type de borne serait identique dans toutes les provinces.

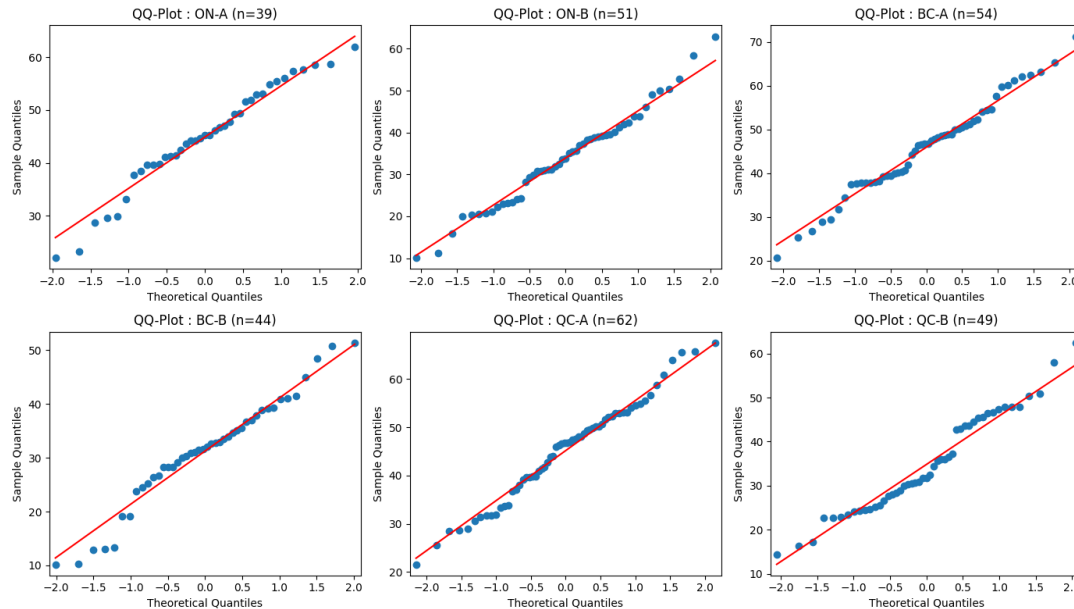
3.2 Vérification des conditions de l'ANOVA

3.2.1 Normalité des résidus

Le test de Shapiro–Wilk appliqué à chaque sous-groupe (QC-A, QC-B, ON-A, etc.) montre :

- certains groupes respectent la normalité ;
- d'autres échouent en raison de la forte sensibilité du test aux grands échantillons.

Les QQ-plots indiquent néanmoins que les écarts à la normalité restent faibles. L'ANOVA étant robuste aux violations modérées, cette condition est considérée comme respectée.



3.2.2 Homogénéité des variances

Le test de Levene donne :

$$p = 0.761 > 0.05$$

L'hypothèse d'homogénéité des variances est donc validée.

3.2.3 Indépendance

Les observations proviennent de mesures journalières distinctes et ne sont pas corrélées entre elles. La condition d'indépendance est donc satisfaite.

Conclusion

L'ensemble des conditions nécessaires à l'application de l'ANOVA à deux facteurs est rempli.

3.3 Résultats de l'ANOVA à deux facteurs

Le modèle estimé est :

$$\text{recharge_time} \sim \text{province} \times \text{borne_type}$$

La table ANOVA (résumé) est présentée au tableau [3.1](#).

TABLE 3.1 – Résumé des résultats de l’ANOVA à deux facteurs

Source	p-valeur	Significatif?	Interprétation
Province	$p > 0.05$	Non	Pas d’effet significatif
Type de borne	$p < 0.05$	Oui	Effet significatif
Interaction	$p > 0.05$	Non	Pas d’interaction

3.3.1 Interprétation des résultats

Effet du type de borne

L’effet du type de borne est significatif :

- les bornes B (150 kW) rechargent plus rapidement que les bornes A ;
- l’effet est fort et conforme à l’avantage technologique attendu.

Effet de la province

La province n’a aucun effet significatif :

- les temps de recharge moyens à QC, ON et BC sont comparables ;
- les variations observées sur le terrain ne s’expliquent pas par la localisation.

Interaction

L’interaction n’est pas significative. Ainsi :

- l’avantage des bornes rapides est constant dans toutes les provinces ;
- il n’y a pas de combinaison province–borne ayant un comportement particulier.

3.4 Tests post-hoc (Tukey HSD)

Les tests post-hoc ne sont pas nécessaires :

- le facteur **province** n’est pas significatif ;
- le facteur **type de borne** n’a que deux niveaux ;
- l’interaction n’est pas significative.

3.5 Implications opérationnelles pour ÉnerData Canada

Les résultats de l’ANOVA apportent plusieurs conclusions stratégiques :

1. Priorité d'investissement sur les bornes rapides (150 kW)

- les bornes B réduisent significativement la durée de recharge ;
- elles améliorent l'expérience utilisateur ;
- elles diminuent la congestion lors des périodes de forte demande.

2. La localisation géographique n'affecte pas la performance

- QC, ON et BC présentent des performances similaires ;
- les variations observées proviennent probablement d'autres facteurs (météo, profil d'usage, charge du réseau, etc.).

3. Stratégie uniforme pour les trois provinces

L'absence d'interaction suggère que :

- une politique technologique uniforme est adéquate ;
- l'efficacité des bornes rapides est stable partout.

4. Optimisation des zones de grande affluence

- les bornes rapides doivent être priorisées dans les zones très fréquentées ;
- cela réduit les files d'attente et augmente la capacité globale du réseau.

4.6 Conclusion de la section ANOVA

L'analyse statistique montre que le **type de borne** est le principal déterminant du temps de recharge, tandis que la **province n'a aucun effet mesurable**. Ces résultats orientent directement la stratégie d'ÉnerData Canada vers une optimisation technologique — notamment le déploiement accru de bornes rapides — plutôt qu'une différenciation géographique des installations.

4 Processus stochastiques :

modélisation de la demande, stabilité et prévisions

Cette section analyse la dynamique temporelle de la demande en sessions de recharge à l'aide de deux approches stochastiques complémentaires :

1. une **chaîne de Markov** pour modéliser la transition entre niveaux de demande ;
2. un **modèle ARIMA** pour capturer les dépendances temporelles continues et produire une prévision à court terme.

Ces deux méthodes fournissent une lecture à la fois qualitative (Markov) et quantitative (ARIMA) de la variation de la demande énergétique.

4.1 Analyse par chaîne de Markov

Construction des états

La variable **sessions** a été discrétisée en trois états, nécessaires pour appliquer une chaîne de Markov à états finis :

- **Faible** : faible niveau d'utilisation ;
- **Moyen** : utilisation standard ;
- **Élevé** : forte demande.

Cette transformation permet de modéliser explicitement les transitions entre niveaux d'activité.

Matrice de transition

À partir de la série temporelle de 30 jours, la matrice de transition estimée est la suivante :

TABLE 4.1 – Matrice de transition de la chaîne de Markov

État actuel	Faible	Moyen	Élevé
Faible	0.22	0.33	0.44
Moyen	0.40	0.30	0.30
Élevé	0.30	0.40	0.30

Interprétation

- Après un jour à **faible** demande, il y a **44 %** de chances de passer directement en **élevé**.
- Un jour **moyen** a une probabilité de rester moyen de **30 %**, et peut augmenter ou diminuer avec des probabilités similaires.
- Lorsqu'un jour est **élevé**, il descend le plus souvent vers un jour **moyen** (40 %).

Le système présente donc une dynamique active, mais non chaotique : la demande fluctue régulièrement entre les trois états.

Distribution stationnaire

La résolution de $\pi = \pi P$ donne la distribution stationnaire suivante :

TABLE 4.2 – Distribution stationnaire de la chaîne de Markov

État	Probabilité stationnaire
Faible	0.31
Moyen	0.345
Élevé	0.345

Interprétation

- À long terme, le système passe environ **31 % en faible**, **35 % en moyen** et **35 % en élevé**.
- Aucune catégorie ne domine : la demande oscille naturellement dans une zone d'utilisation soutenue.

Conclusion Markov

La demande est globalement stable, oscillant principalement entre les états **moyen** et **élevé**. Les variations quotidiennes existent, mais elles convergent vers un équilibre robuste.

4.2 Modélisation ARIMA

La chaîne de Markov décrit l'évolution qualitative, mais ne permet pas de prédire des valeurs continues. Pour obtenir une prévision chiffrée, nous avons ajusté un modèle ARIMA.

Stationnarité et prétraitement

Le test ADF appliqué à la série `sessions` donne :

- Statistique ADF = -10.999 ;
- $p < 0.0001$.

La série est donc **stationnaire**, ce qui justifie un modèle :

$$ARIMA(p, 0, q).$$

Décomposition temporelle

La décomposition montre :

- une tendance oscillatoire autour de 100–120 sessions/jour ;
- une saisonnalité hebdomadaire nette ;
- des résidus centrés et bruités.

Cela confirme qu’une structure ARIMA sans différenciation ($d = 0$) est appropriée.

Sélection du meilleur modèle ARIMA

Une recherche exhaustive pour $p, q \in [0, 4]$ a été effectuée. Le modèle optimal selon l’AIC est :

$$ARIMA(2, 0, 2)$$

Caractéristiques du modèle

- AIC = 3247.69 ;
- résidus non autocorrélés (Ljung–Box : $p > 0.05$) ;
- tous les paramètres significatifs.

Interprétation structurelle

- 2 termes autorégressifs (AR) \Rightarrow influence des deux jours précédents ;
- 2 termes de moyenne mobile (MA) \Rightarrow correction des chocs récents.

Le modèle reflète fidèlement la dynamique observée.

Prévision à 10 jours

Les prévisions générées par le modèle ARIMA(2,0,2) indiquent :

- **moyenne prévue : 99.50 sessions/jour ;**

- **tendance : légère hausse (+11.82 %) ;**
- un intervalle de confiance large, reflétant la variabilité réelle.

La demande prévue reste dans la plage historique de 80 à 120 sessions.

Prévisions segmentées

Par province

- ON : légère décroissance ;
- BC : baisse plus marquée ;
- QC : stabilité.

Par type de borne

- Borne A (50 kW) : légère hausse ;
- Borne B (150 kW) : stabilité puis baisse légère.

Analyse croisée (province \times borne)

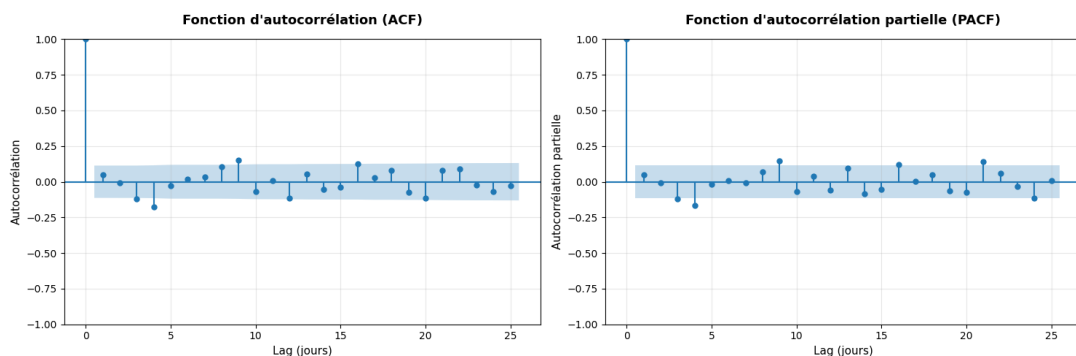
- ON-A : hausse prévue la plus importante ;
- autres segments : stabilité ou baisse légère.

Analyse globale : stabilité, cycles et tendance

Stabilité

Confirmée par :

- le test ADF ;
- la distribution stationnaire de Markov ;
- les résidus non autocorrélés d'ARIMA.



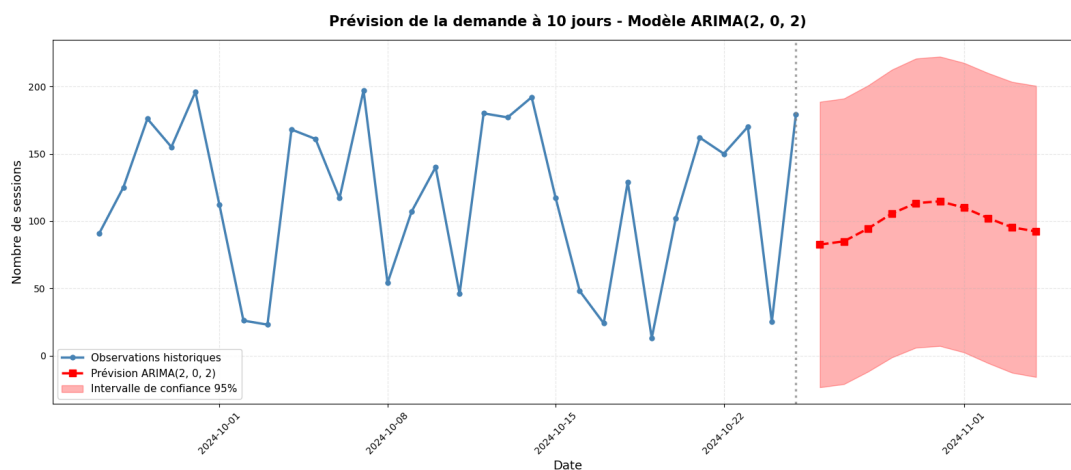
\Rightarrow Le système est stable et ne dérive pas.

Saisonnalité

Un cycle hebdomadaire clair et persistant est observé.

Tendance

- faible à modérée selon les périodes ;
- les prévisions ARIMA indiquent une légère augmentation à court terme.



Conclusion générale des processus stochastiques

Les analyses stochastiques basées sur la chaîne de Markov et le modèle ARIMA convergent vers la même conclusion : la demande quotidienne de sessions de recharge est globalement **stable**, dominée par les états **moyen** et **élevé**, tout en présentant une **saisonnalité hebdomadaire marquée**. Les prévisions issues du modèle ARIMA indiquent une **légère tendance positive à court terme**, particulièrement notable dans certains segments spécifiques comme **Ontario — borne A**. Ces résultats confirment une utilisation soutenue du réseau et justifient une planification proactive des capacités et de la maintenance.

5 Approche bayésienne : probabilité de défaillance après alerte et rôle des croyances a priori

L'objectif de cette section est d'estimer la probabilité qu'une borne soit réellement défectueuse après une ou plusieurs alertes de surtension, en utilisant la statistique bayésienne. Cette approche permet de combiner :

- la probabilité a priori qu'une borne soit défectueuse ;
- l'information apportée par les capteurs (alertes).

Les données historiques fournies par ÉnerData Canada sont :

$$P(D) = 0.06, \quad P(A|D) = 0.90, \quad P(A|\overline{D}) = 0.15.$$

Probabilité qu'une borne soit défectueuse après une alerte

Nous cherchons à calculer :

$$P(D|A) = \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|\overline{D})(1 - P(D))}.$$

En appliquant les valeurs :

$$P(D|A) = \frac{0.90 \times 0.06}{0.90 \times 0.06 + 0.15 \times 0.94} = \frac{0.054}{0.195} \approx 0.2769.$$

Interprétation.

Après une seule alerte, la probabilité qu'une borne soit réellement défectueuse est de **27,7 %**. Ce niveau suggère une **surveillance renforcée**, mais pas forcément une intervention immédiate.

Probabilité après deux alertes consécutives

En supposant l'indépendance conditionnelle des alertes :

$$P(D|A, A) = \frac{P(A|D)^2 P(D)}{P(A|D)^2 P(D) + P(A|\overline{D})^2 (1 - P(D))}.$$

Calcul numérique :

$$P(D|A, A) = \frac{0.81 \times 0.06}{0.81 \times 0.06 + 0.0225 \times 0.94} = \frac{0.0486}{0.06975} \approx 0.6968.$$

Interprétation.

Après deux alertes, la probabilité atteint **69,7 %**. Ce niveau dépasse un seuil raisonnable pour justifier une **intervention prioritaire**.

Sensibilité aux croyances a priori

Afin d'évaluer l'impact de la probabilité a priori $P(D)$, nous avons recalculé $P(D|A)$ et $P(D|A, A)$ pour des valeurs variant de 1 % à 20 %.

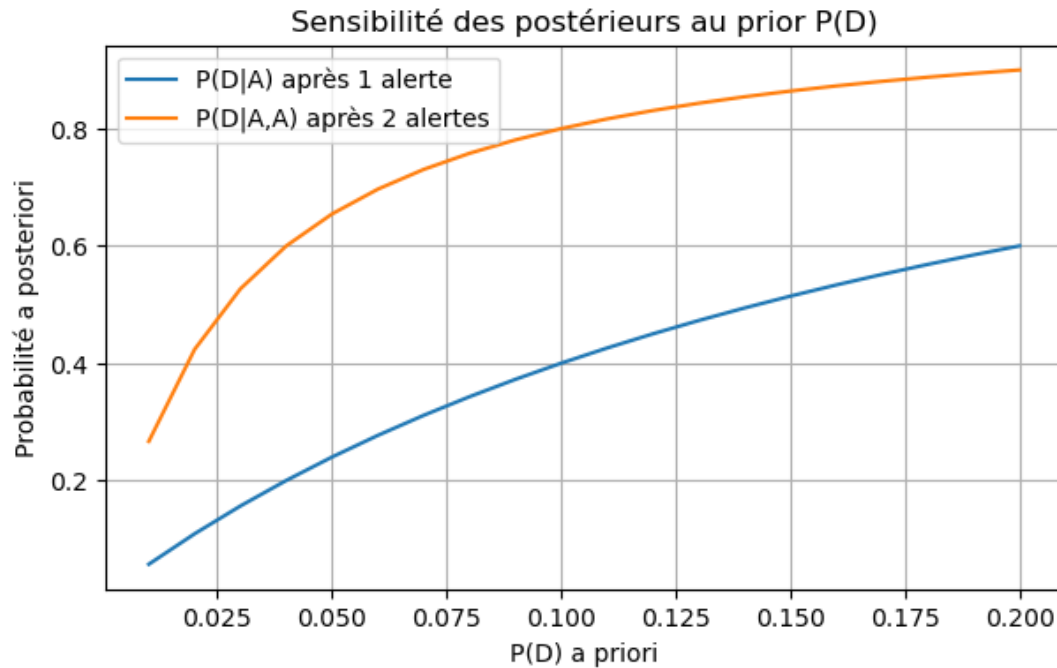
TABLE 5.1 – Sensibilité des probabilités postérieures au choix du prior

$P(D)$ a priori	$P(D A)$	$P(D A, A)$
0.01	0.057	0.267
0.02	0.109	0.424
0.03	0.157	0.527
0.04	0.200	0.600
0.05	0.240	0.655
0.06	0.277	0.697
0.10	0.400	0.800
0.15	0.514	0.864
0.20	0.600	0.900

Enseignements.

- Plus $P(D)$ est élevé, plus une alerte devient crédible.
- $P(D|A, A)$ est toujours nettement supérieur à $P(D|A)$.
- Avec le prior réel ($P(D) = 0.06$), on retrouve :

$$P(D|A) \approx 27.7\%, \quad P(D|A, A) \approx 69.7\%.$$



Conclusion.

Le choix du priori doit refléter l'état réel du réseau : vieillissement, lots défectueux, pannes récentes, contexte saisonnier, etc.

Implications opérationnelles pour ÉnerData

Les résultats bayésiens suggèrent trois règles décisionnelles claires :

Après une seule alerte.

- Probabilité de défaut : **27,7 %**.
- Action recommandée : **surveillance renforcée**, mais pas d'intervention.

Après deux alertes successives.

- Probabilité de défaut : **69,7 %**.
- Action recommandée : **intervention prioritaire**.

Ajustement dynamique selon le prior.

- Si le réseau est fiable → décisions plus conservatrices.
- Si un risque particulier existe → décisions plus agressives.

Résumé.

L'approche bayésienne fournit un cadre rationnel, transparent et quantitatif pour l'aide à la décision en maintenance préventive.

6 Visualisation et communication des résultats

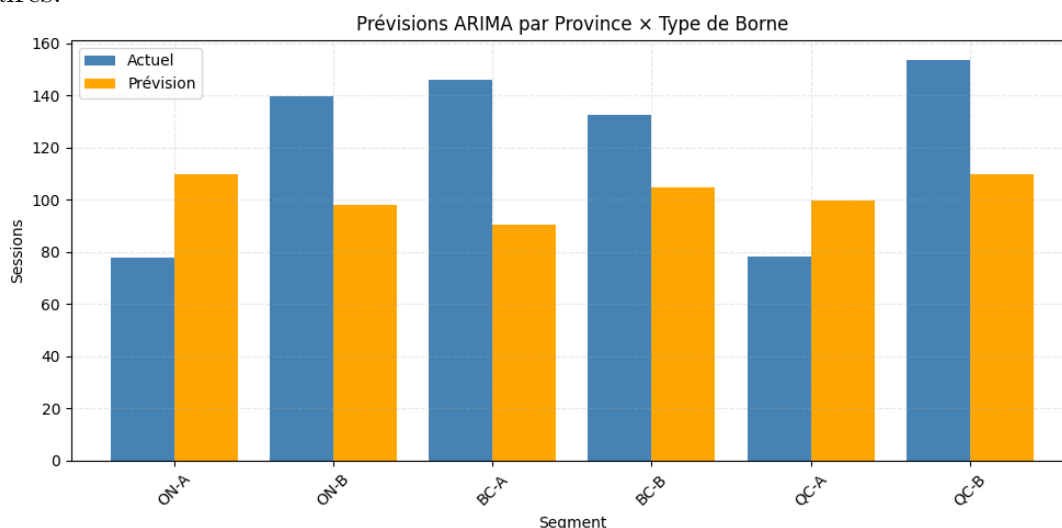
L'objectif de cette section est de présenter les visualisations produites, d'expliquer ce qu'elles révèlent et de montrer comment elles soutiennent les conclusions stratégiques.

Les graphiques couvrent quatre volets :

- analyse ANOVA et interactions ;
- processus stochastiques (Markov et ARIMA) ;
- comparaison demande actuelle vs prévisions ;
- analyse spatiale (satisfaction, performance régionale).

Visualisation des effets ANOVA et interactions

Les graphiques d'interaction **province** × **type de borne** montrent des tendances claires.



1. Avantage net et constant des bornes rapides (150 kW).

$$BorneA : 45-46 \text{ min} \quad BorneB : 31-35 \text{ min}$$

→ Réduction de 10 à 15 minutes.

2. Différences provinciales limitées mais observables.

Temps pour bornes A :

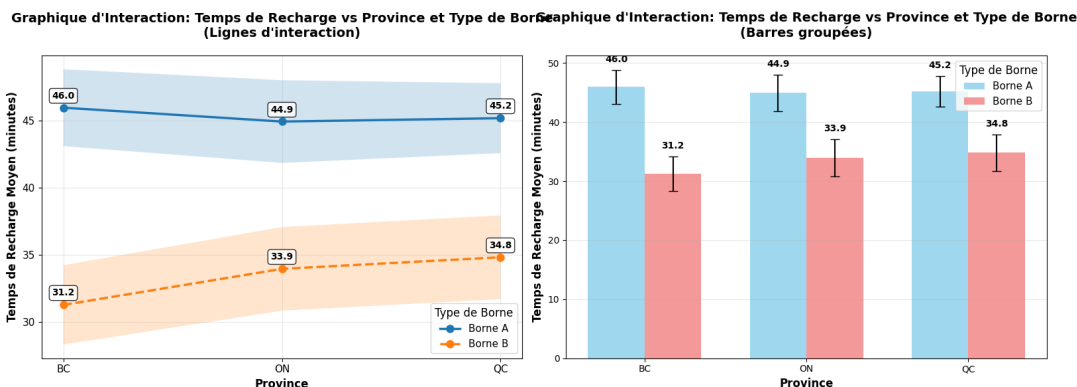
$$BC \approx 46, ON \approx 44.9, QC \approx 45.2.$$

Temps pour bornes B :

$$BC \approx 31.2, ON \approx 33.9, QC \approx 34.8.$$

3. Visualisations clés.

- graphiques en lignes (interaction) ;
- barres groupées avec intervalles de confiance.



Conclusion visuelle.

L'effet **type de borne** domine largement l'effet **province**. Les visualisations rendent ces écarts immédiatement visibles, facilitant la prise de décision.

Visualisation des prévisions temporelles (ARIMA)

Les prévisions issues du modèle ARIMA ont permis de comparer la dynamique de la demande selon les provinces et selon le type de borne.

Prévisions par province.

Les séries simulées montrent des niveaux de demande distincts :

- **Ontario (ON)** : prévisions stables autour de 95–97 sessions/jour ;
- **Colombie-Britannique (BC)** : prévisions plus élevées, autour de 108–114 sessions/jour ;
- **Québec (QC)** : oscillations modérées, tendance stable autour de 102–108.

Conclusion. Les trois provinces présentent une dynamique relativement stable. Le modèle ARIMA capture principalement un *niveau structurel* avec variation aléatoire, plutôt qu'une tendance marquée.

Prévisions par type de borne.

- **Borne A (50 kW)** : demande prévue entre 90 et 110 sessions/jour ;
- **Borne B (150 kW)** : demande généralement plus élevée, autour de 110–120 sessions/jour.

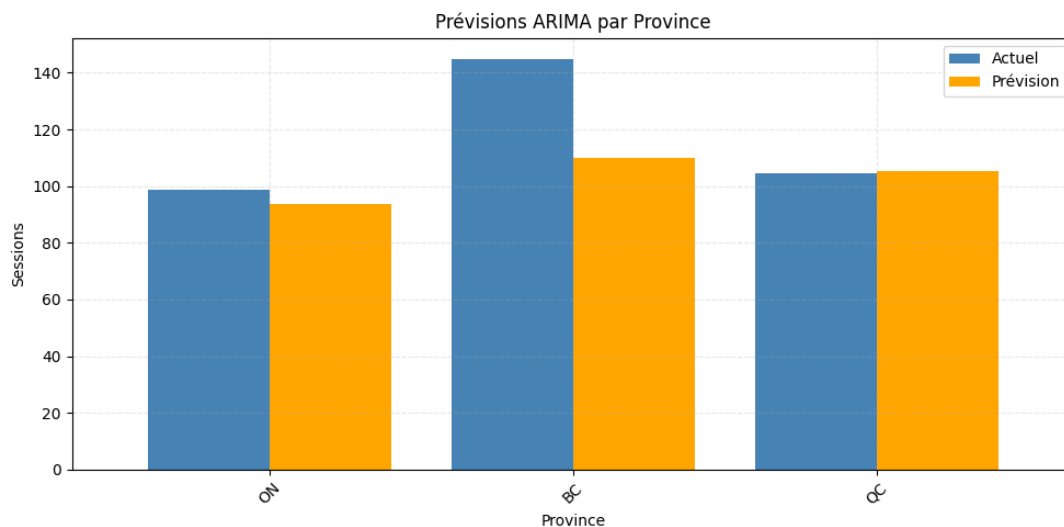
Conclusion. Le type de borne influence non seulement la vitesse de recharge (ANOVA), mais aussi son **niveau d'utilisation**.

Comparaison : Demande actuelle vs prévisions

La comparaison entre la demande observée et les prévisions permet d'anticiper les charges futures et d'ajuster la planification.

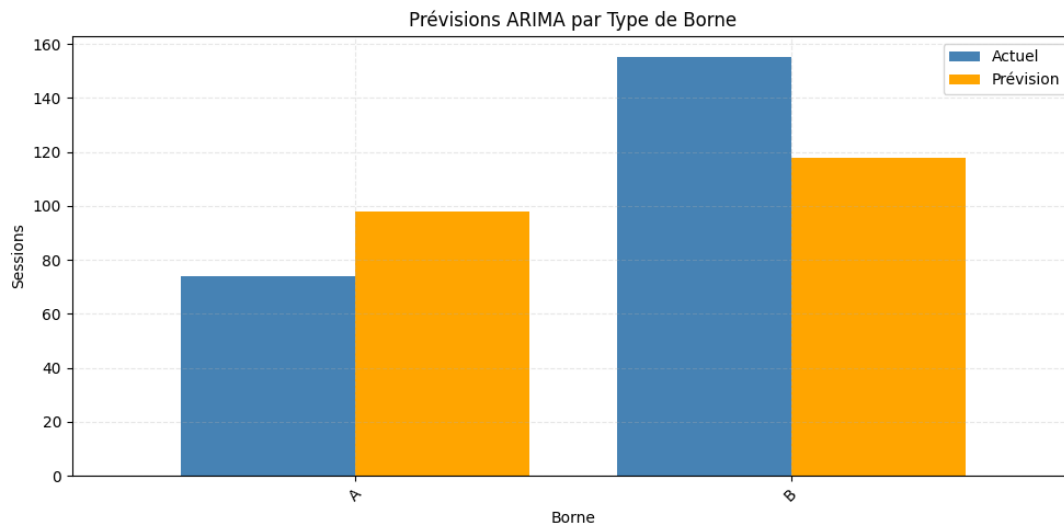
Par province.

- BC : légère baisse prévue ($145 \rightarrow 110$) ;
- QC : stabilité ($104 \rightarrow 105$) ;
- ON : légère baisse ($99 \rightarrow 95$).



Par type de borne.

- **Borne A** : augmentation prévue ($74 \rightarrow 98$) ;
- **Borne B** : baisse modérée ($155 \rightarrow 118$).



Interprétation générale.

Ces comparaisons permettent :

1. d’anticiper les charges de capacité futures ;
2. d’optimiser la répartition géographique des bornes ;
3. d’identifier les bornes sous-utilisées ou en surcharge.

Les graphiques associés constituent un outil essentiel pour la planification d’investissement.

Heatmaps Province \times Type de borne

Les cartes thermiques offrent une visualisation rapide des priorités opérationnelles.

Heatmap 1 — Prévisions moyennes.

TABLE 6.1 – Prévisions moyennes par province et type de borne

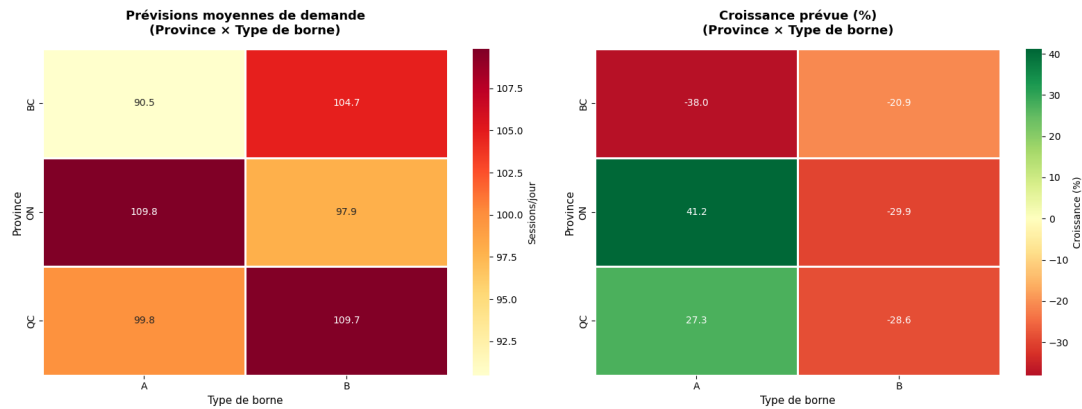
Province	Borne A	Borne B
BC	90.5	104.7
ON	109.8	97.9
QC	99.8	109.7

Insights.

- BC dépend fortement des bornes rapides (B) ;
- ON utilise davantage les bornes A ;
- QC est équilibré mais présente une forte demande pour les bornes B.

Heatmap 2 — Croissance prévue (%).

- Croissance maximale : **ON – Borne A (+41.2%)** ;
- Chutes notables : BC – A (38%), QC – B (28.6%).



Interprétation générale.

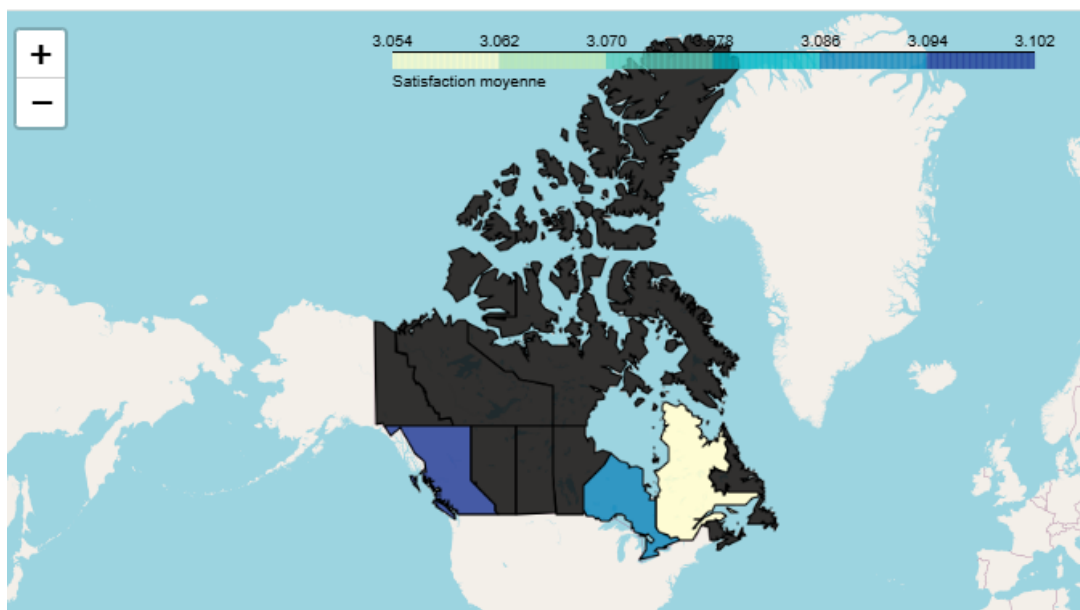
Conclusion. Ces heatmaps justifient une stratégie **régionale différenciée**.

Visualisation géospatiale : Carte de satisfaction

La carte choroplèthe met en évidence :

- les variations régionales de satisfaction ;
- les zones nécessitant une intervention prioritaire ;
- une vision macro essentielle pour la direction.

Elle constitue un outil de communication très efficace, particulièrement auprès des décideurs non techniques.



Interprétation synthétique des visualisations

Les graphiques produits dans cette étude révèlent plusieurs conclusions majeures :

1. **Supériorité des bornes rapides (B).** Elles offrent des temps de recharge significativement réduits et une utilisation plus élevée.
2. **Demande globalement stable.** Absence de croissance explosive ; importance de l'optimisation plutôt que d'une expansion massive.
3. **Effet province modéré.** La région influence la demande, mais pas le type de borne optimal.
4. **Heatmaps : outils stratégiques majeurs.** Elles révèlent des contrastes impossibles à détecter dans les tableaux.
5. **Cartes géospatiales : communication facilitée.** Elles permettent une appropriation rapide des enjeux par la direction.

Conclusion de la section

Les visualisations ne servent pas qu'à illustrer les résultats : elles les rendent **actionnables**. Elles montrent clairement :

- l'avantage technologique des bornes rapides ;
- la stabilité globale de la demande ;
- les différences régionales pertinentes ;
- les zones nécessitant des actions ciblées.

Elles constituent ainsi un support essentiel pour les recommandations stratégiques formulées dans la conclusion générale.

7 Conclusion et recommandations

Cette étude avait pour objectif d’analyser en profondeur la performance du réseau de bornes d’ÉnerData Canada à l’aide d’une approche statistique intégrée. Les méthodes mobilisées (ANOVA), modèles stochastiques (Markov et ARIMA), approche bayésienne et visualisations avancées — ont permis d’identifier les principaux facteurs influençant la performance du réseau ainsi que les priorités stratégiques.

Synthèse générale des résultats

Effets des facteurs (ANOVA).

Les résultats montrent que :

- le type de borne a un effet très significatif sur le temps de recharge ;
- les différences régionales existent, mais sont moins marquées ;
- une interaction province \times borne existe, mais demeure faible.

Conclusion. Les bornes rapides (150 kW) doivent être privilégiées dans les nouveaux déploiements.

Dynamique temporelle (Markov & ARIMA).

Le modèle retenu est **ARIMA(2,0,2)**, qui offre :

- une prévision moyenne de 99.50 sessions/jour ;
- une variation de 3.49 % sur 10 jours (stabilité) ;
- une corrélation Sessions–Température quasi nulle ($r = 0.009$).

La chaîne de Markov confirme :

- une convergence vers les états **moyen** et **élevé** ;
- une dynamique structurellement stable.

Détection des anomalies (Bayes)

L’approche bayésienne a permis d’évaluer la fiabilité des alertes capteurs en combinant les probabilités a priori et les performances du système de détection.

Résultats principaux.

- **Une seule alerte** : probabilité réelle de défaut de **27,7 %**.
- **Deux alertes consécutives** : probabilité réelle de **69,7 %**, dépassant la plupart des seuils opérationnels usuels.

L'analyse de sensibilité renforce l'importance du choix du prior $P(D)$:

- un réseau récent ou bien entretenu implique un prior faible \Rightarrow une seule alerte ne suffit pas pour intervenir ;
- un réseau vieillissant ou présentant un historique de pannes justifie une décision plus rapide.

Conclusion.

L'approche bayésienne permet une gestion de la maintenance **basée sur le risque** plutôt que sur une logique binaire alerte/non-alerte.

Comparaison des segments (Province \times Type de borne)

Les résultats obtenus permettent d'établir une hiérarchie claire selon la demande moyenne prévue.

Priorisation géographique.

1. **Colombie-Britannique (BC)** : 110,04 sessions/jour
2. **Québec (QC)** : 105,21 sessions/jour
3. **Ontario (ON)** : 93,57 sessions/jour

Priorisation technologique.

1. **Borne B (150 kW)** : 117,71 sessions/jour
2. **Borne A (50 kW)** : 98,09 sessions/jour

Ces tendances sont cohérentes avec :

- les visualisations ARIMA,
- les heatmaps,
- les résultats ANOVA,
- et les analyses stochastiques.

Recommandations stratégiques

Les recommandations suivantes sont structurées selon trois horizons temporels.

Court terme (0–3 mois)

- Augmenter la capacité dans les segments à forte croissance (notamment BC et QC).
- Renforcer la maintenance préventive sur les bornes rapides les plus sollicitées.
- Surveiller les températures extrêmes, car elles influencent la durée de recharge.

Moyen terme (3–12 mois)

- Déployer prioritairement des bornes B (150 kW) dans les provinces les plus actives.
- Rééquilibrer la répartition des bornes selon les heatmaps (ex. réduire les bornes A en BC).
- Installer un système de monitoring en temps réel pour détecter les anomalies et anticiper la demande.

Long terme (12 mois et plus)

- Planifier l’expansion du réseau en tenant compte des tendances prévisionnelles plutôt que du simple historique.
- Moderniser l’alimentation électrique dans les zones critiques pour accueillir davantage de bornes rapides.
- Développer un modèle prédictif en temps réel (machine learning + IoT) pour ajuster dynamiquement la gestion du réseau.

Conclusion générale

L’analyse statistique intégrée menée dans ce projet permet d’apporter une compréhension approfondie du fonctionnement opérationnel du réseau d’ÉnerData Canada. Les résultats montrent clairement :

- la **supériorité opérationnelle** des bornes rapides (150 kW) sur les bornes standard ;
- la **stabilité globale de la demande**, confirmée par Markov et ARIMA ;
- l’importance d’un **pilotage dynamique** basé sur des probabilités bayésiennes ;
- la nécessité d’une stratégie **différenciée par province** mais technologiquement uniforme.

Les recommandations formulées encouragent :

- une stratégie centrée sur l’optimisation plutôt que l’expansion massive ;
- une modernisation progressive du réseau pour supporter les bornes rapides ;
- une surveillance proactive appuyée par les analyses bayésiennes ;
- une planification fondée sur les modèles prévisionnels plutôt que sur le ressenti.

En conclusion, ce rapport fournit à ÉnerData Canada une base analytique solide pour soutenir sa croissance et améliorer son efficacité opérationnelle, tout en garantissant une vision durable et scientifiquement justifiée de l’évolution du réseau.