



# Défis en design d'expérience dans l'industrie du commerce électronique

**Isabelle Lefebvre**  
**Staff Data Scientist, Shopify**



Présenté le

15 Mars 2024

## Agenda

**01**

**Mon parcours**

**02**

**Shopify**

**03**

**Défis en design d'expérience**

**04**

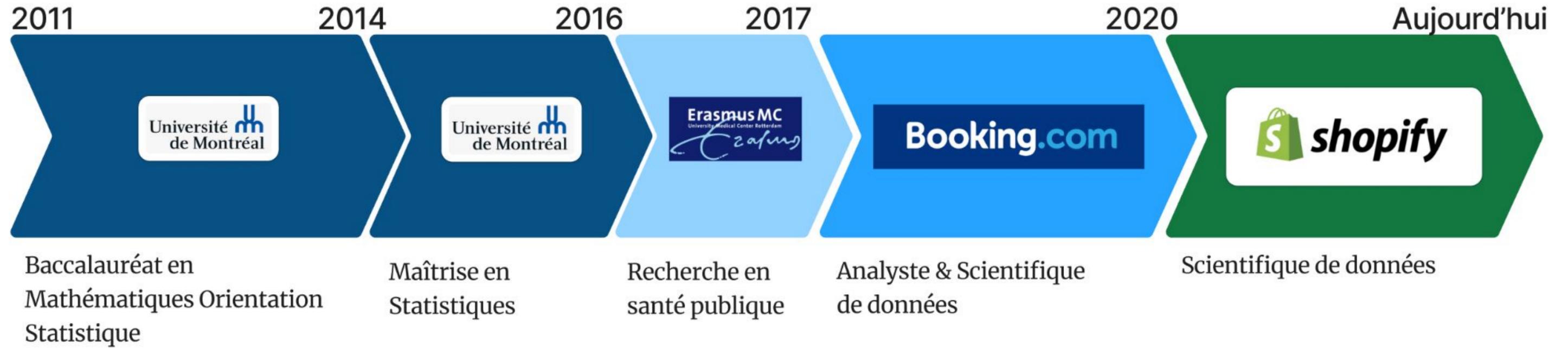
**Questions?**



# Mon Parcours



# Mon parcours



# Shopify

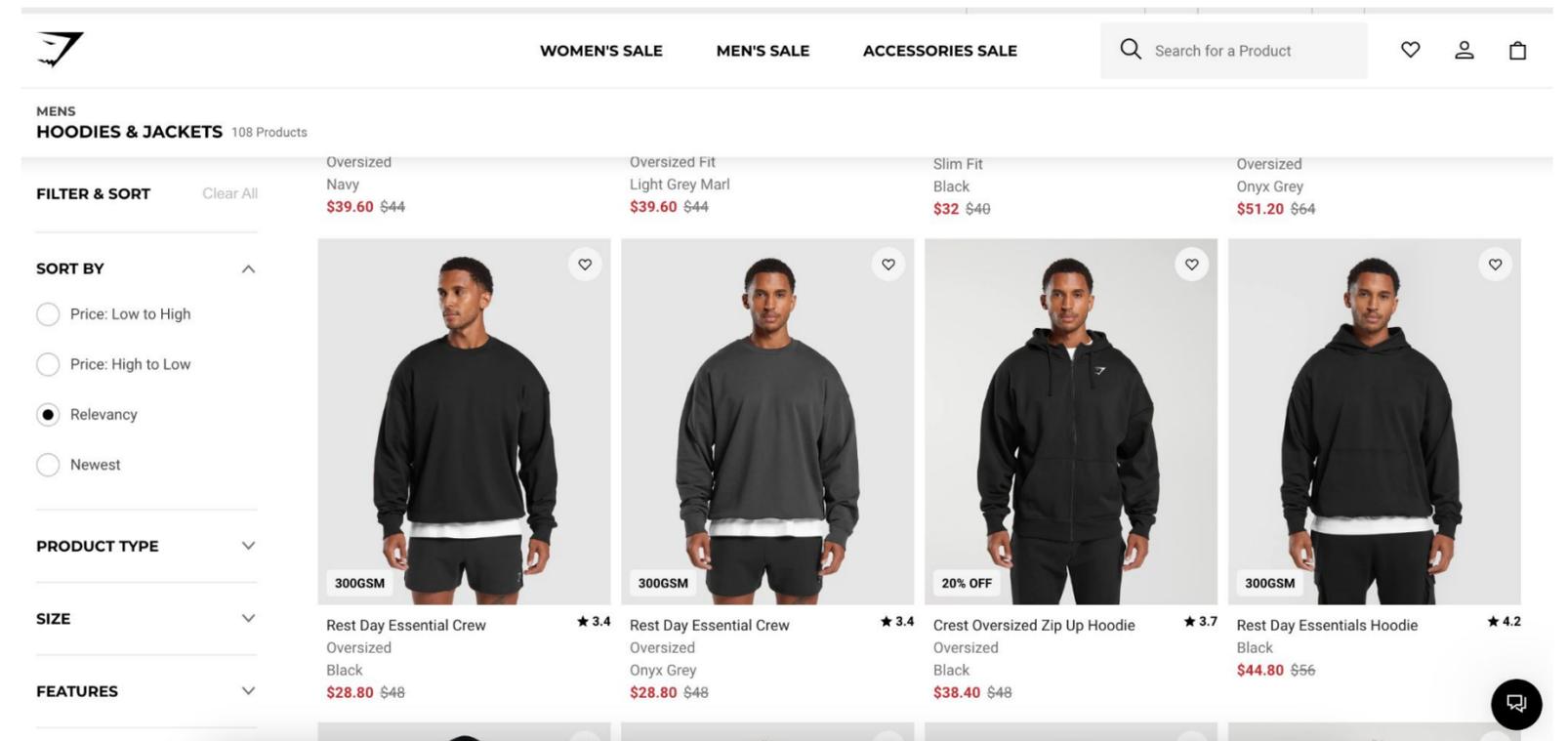
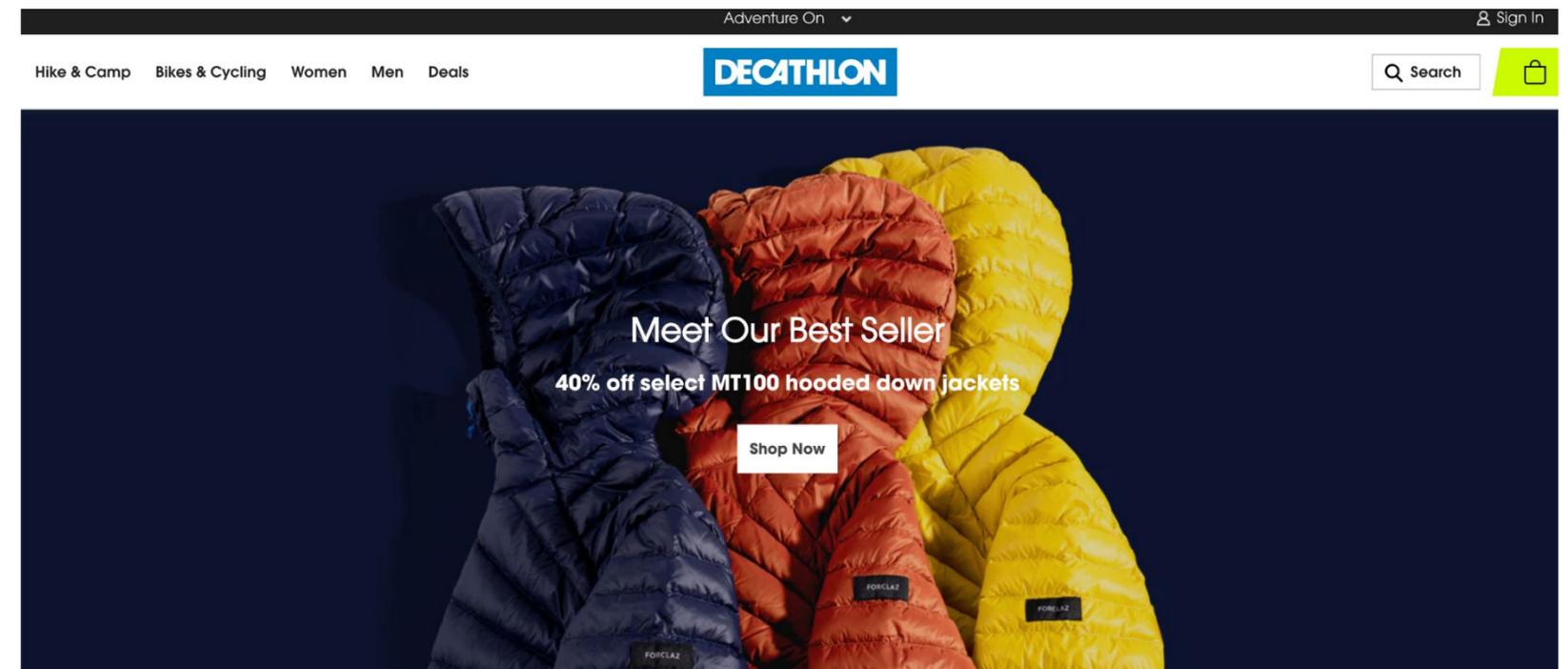


# C'est quoi Shopify?

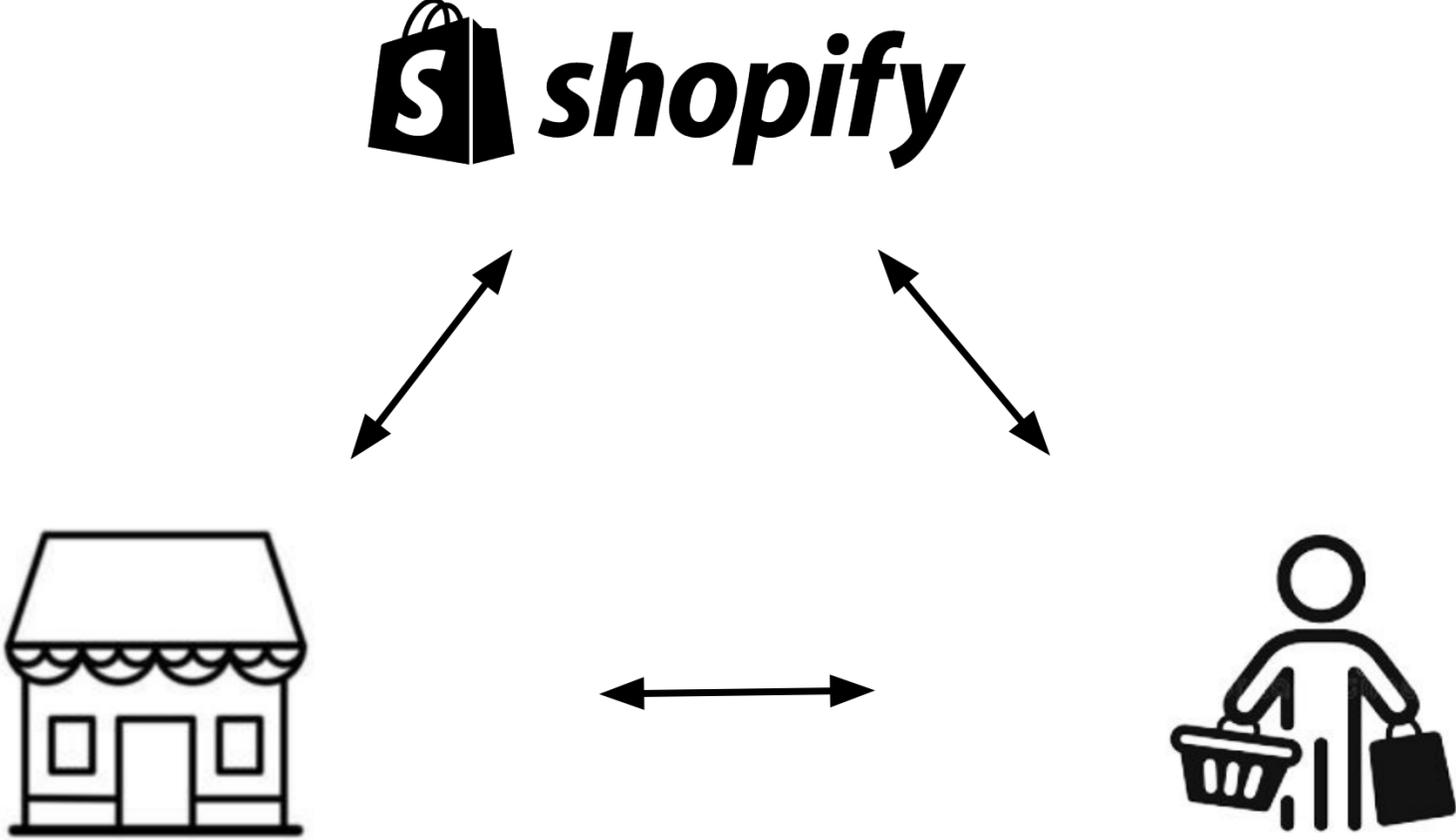
Une plateforme par abonnement qui permet à tous les entrepreneurs de créer une boutique en ligne et de vendre leurs produits et services.

- L'entrepreneur n'a pas besoin de savoir programmer
- Solutions disponibles pour tous les essentiels tel que paiements et livraison.

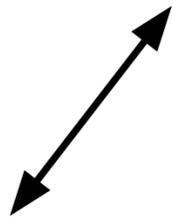
Quelques exemples de boutiques en ligne créées avec Shopify: GymShark, Decathlon, Fashion Nova, Altitude Sports, Pajar



# C'est quoi Shopify?



# C'est quoi Shopify?



A screenshot of the Shopify admin dashboard. The interface is dark-themed with a light grey sidebar on the left containing navigation menus for Home, Orders, Products, Customers, Content, Analytics, Marketing, Discounts, Sales channels, Online Store, and Apps. The main content area shows a top navigation bar with a search bar and user profile. Below this, there are filters for 'Last 30 days' and 'All channels', and a 'Next payout: \$0.00' indicator. A central analytics card displays four key metrics: Online store sessions (1), Total sales (\$0.00), Total orders (0), and Conversion rate (0%). Below the metrics is a line chart showing data from Feb 12 to Mar 10, 2024. At the bottom, a 'Things to do next' section includes a 'Setup guide' with a progress indicator '2 / 3 completed' and a list of tasks: 'Add your first product' (checked), 'Customize your online store' (checked), and 'Add a custom domain' (unchecked).

# C'est quoi Shopify?

**PAJAR** CANADA  
Cart > Information > Shipping > Payment

**Contact** Have an account? Log in

Email

Email me with news and offers

**Shipping address**

Country/region  
Canada

First name  Last name

Address

Apartment, suite, etc. (optional)

City  Province  Postal code

Phone

[Return to cart](#)

[Refund policy](#) [Shipping policy](#) [Privacy policy](#) [Terms of service](#)

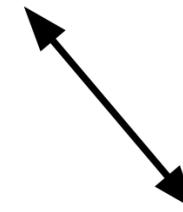
**Order Summary:**

- 1 Carry F Women's Heritage Knee-High Boot 5 / BLACK \$391.00

Discount code

Subtotal \$391.00  
Shipping  Calculated at next step

Total CAD \$391.00



# La science de données chez Shopify



## Data Engineer

Transforme les données massives pour obtenir des jeux de données faciles d'utilisation.

Les 2 autres rôles consomment les jeux de données produits.



## Product Data Scientist

Met à profit les données massives pour aider l'entreprise à prendre les meilleures décisions quant aux produits offerts.



## Machine Learning Engineer

Crée et maintient des modèles d'apprentissage automatiques pour offrir les meilleures solutions aux marchands.

# La science de données chez Shopify

## Data Engineer

Transforme les données massives pour obtenir des jeux de données faciles d'utilisation.

Les 2 autres rôles consomment les jeux de données produits.

## Product Data Scientist

Met à profit les données massives pour aider l'entreprise à prendre les meilleures décisions quant aux produits offerts.



## Machine Learning Engineer

Crée et maintient des modèles d'apprentissage automatiques pour offrir les meilleures solutions aux marchands.

# Au quotidien

- Discussions sur projets avec des collaborateurs multidisciplinaires: produit, marketing, développeurs, design
- Mesurer l'impact d'une intervention (future ou passée)
- Aider l'équipe à définir la stratégie des prochains mois
- Démocratisation des données en construisant des tableaux de bords
- Mentorat des membres de l'équipe qui sont plus junior
- Programmation en SQL, Python, pySpark

# Défis en design d'expérience



# Ratio de variables aléatoires

Intervention

On veut permettre à l'utilisateur de payer dans sa propre devise.

Hypothèse

On s'attend à:

- Plus de réservations
- Mais aussi plus de coûts

Mesure

$$\text{Retour sur l'investissement} = \frac{\text{Revenue}}{\text{Coûts}}$$

Design

Randomisation des visiteurs de la population cible

Problème

- C'est un ratio de variables aléatoires
- Distribution inconnue

**Montant total CAD 231,46**

Taxes et frais compris

Dans la devise de l'établissement: € 157,05

**Informations sur le tarif**

Taxes et frais de CAD 44,47 compris  
 10 % de TVA CAD 20,50  
 taxe de séjour CAD 23,96

Dépôt de garantie (Entièrement remboursable) CAD 74

Le montant est converti en CAD pour vous indiquer le prix approximatif. Vous paierez en €. Le taux de change peut varier avant votre paiement.

N'oubliez pas que l'émetteur de votre carte peut vous facturer des frais sur les paiements en devises étrangères.

Masquer les détails

Contrôle

**Montant total CAD 231,46**

Taxes et frais compris

~~Dans la devise de l'établissement: € 157,05~~

**Informations sur le tarif**

Taxes et frais de CAD 44,47 compris  
 10 % de TVA CAD 20,50  
 taxe de séjour CAD 23,96

Dépôt de garantie (Entièrement remboursable) CAD 74

~~Le montant est converti en CAD pour vous indiquer le prix approximatif. Vous paierez en €. Le taux de change peut varier avant votre paiement.~~

N'oubliez pas que l'émetteur de votre carte peut vous facturer des frais sur les paiements en devises étrangères.

Masquer les détails

Test

# Ratio de variables aléatoires

## Solution

On utilise le bootstrap pour construire nos intervalles de confiance

1. On fait un échantillonnage avec remise dans notre groupe contrôle et notre groupe test.
2. On calcule nos deux mesure d'intérêt  $ROI_C$  et  $ROI_T$  à partir de l'échantillon obtenu.
3. On calcule l'effet de l'intervention qu'on dénote:

$$Effet_i = \frac{ROI_T - ROI_C}{ROI_C}$$

4. On répète les étapes 1 à 3, un très grand nombre de fois pour obtenir la distribution de l'effet.
5. On utilise le 5<sup>e</sup> et 95<sup>e</sup> percentile de la distribution de l'effet pour construire l'intervalle de confiance.

# Design idéal n'est pas réalisable

## Intervention

On veut lancer Shopify Paiement en France

## Hypothèse

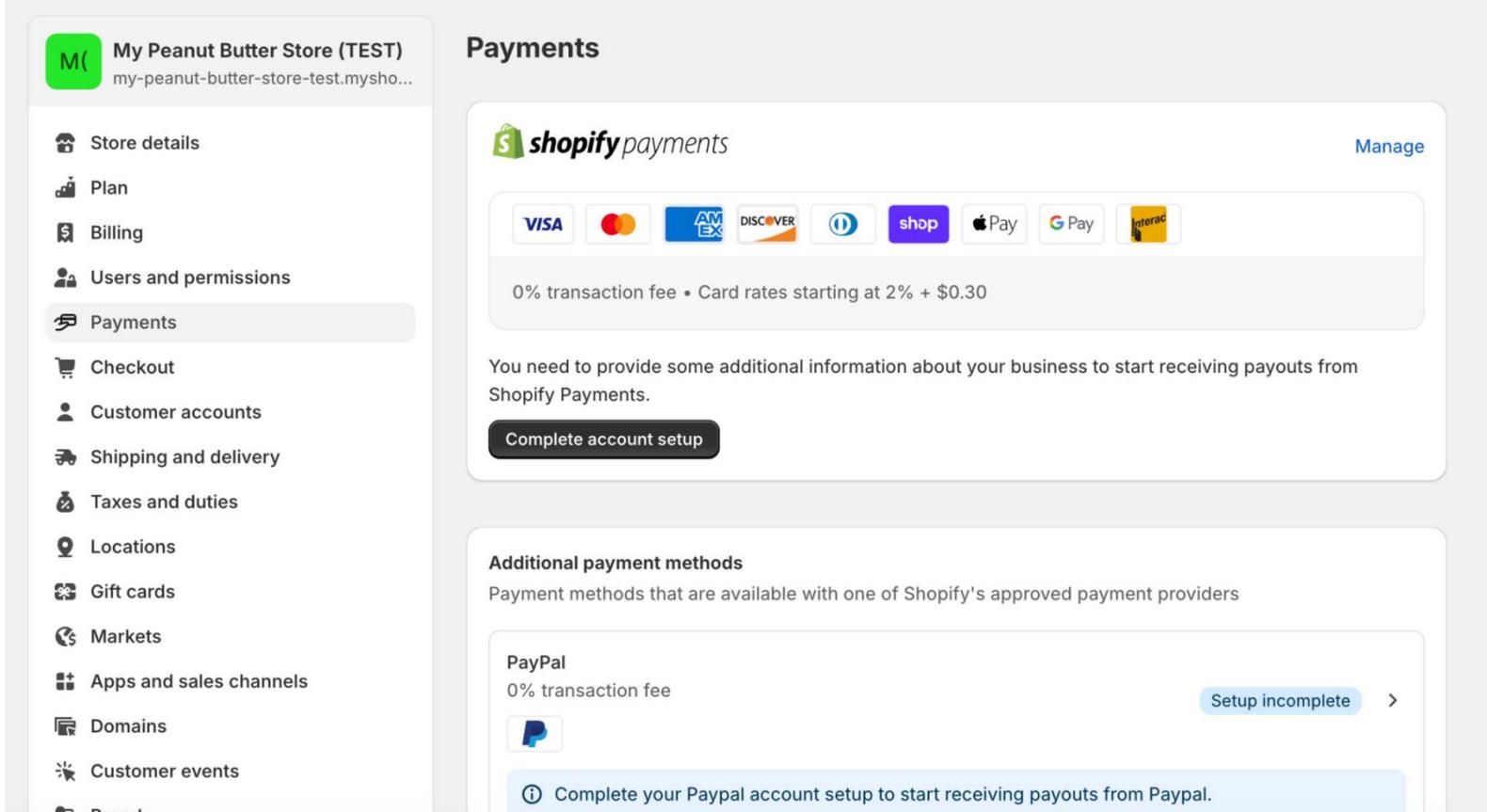
On s'attend à ce que nos marchands Français créent un site web fonctionnel plus rapidement.

## Mesure

Ventes moyennes des marchands Français

## Problème

La stratégie marketing ne permet pas de faire une expérience randomisée



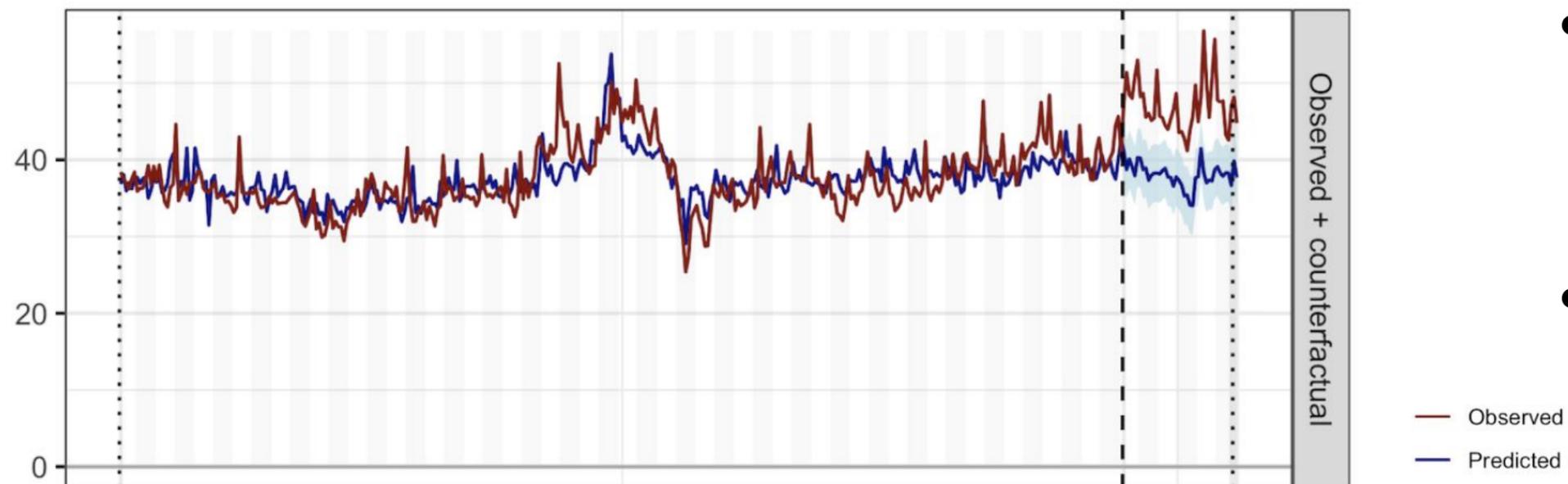
The screenshot shows the Shopify admin interface for a test store named "My Peanut Butter Store (TEST)". The left sidebar contains a navigation menu with options like "Store details", "Plan", "Billing", "Payments", "Checkout", etc. The main content area is titled "Payments" and displays the "shopify payments" section. It shows supported payment methods (VISA, Mastercard, AMEX, DISCOVER, shop, Apple Pay, G Pay, Interac) and a "0% transaction fee" offer. A message indicates that additional business information is needed to start receiving payouts, with a "Complete account setup" button. Below this, the "Additional payment methods" section shows "PayPal" with a "Setup incomplete" status and a message to complete the setup.

# Design idéal n'est pas réalisable

## Solution

On utilise l'inférence causale pour mesurer l'effet du lancement

L'idée est de prédire comment les ventes moyennes auraient évoluées en France en l'absence du lancement de Shopify Paiement, c'est-à-dire le contrefactuel.

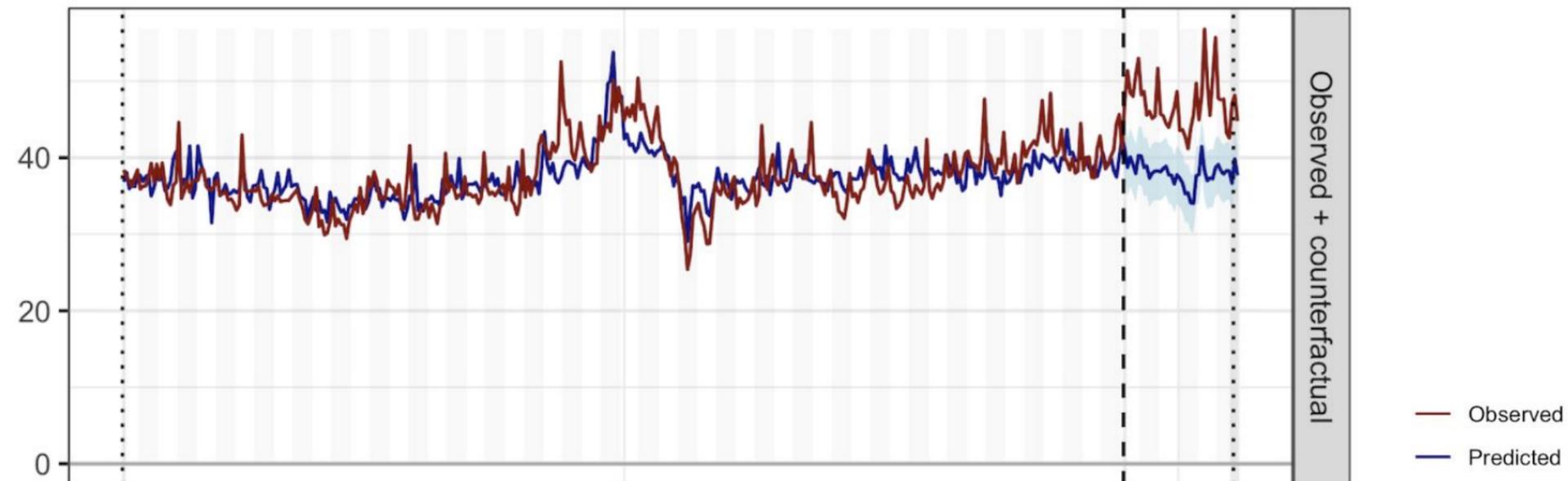


- On utilise les observations d'autres pays qui sont fortement corrélés avec la France.
- L'approche est celle de [Kerman et al 2017](#). "Time-based regression" qui utilise l'inférence bayésienne pour prédire le contrefactuel.
- Hypothèse que la série chronologique du groupe contrôle est suffisante pour prédire les résultats du groupe test avant et après l'intervention.
- Approche utilisée par Google pour mesurer l'efficacité de leur annonces

# Calcul du plus petit effet détectable

**Problème** Quel est l'effet le plus petit effet qu'on pourra détecter avec cette approche?

**Solution** Étude par simulation!



# Détecter des effets plus petits avec la même taille d'échantillon

Intervention

On ajoute Interac à Shopify Paiement au Canada

Hypothèse

On s'attend à ce que cette méthode augmente les ventes des marchands

Mesure

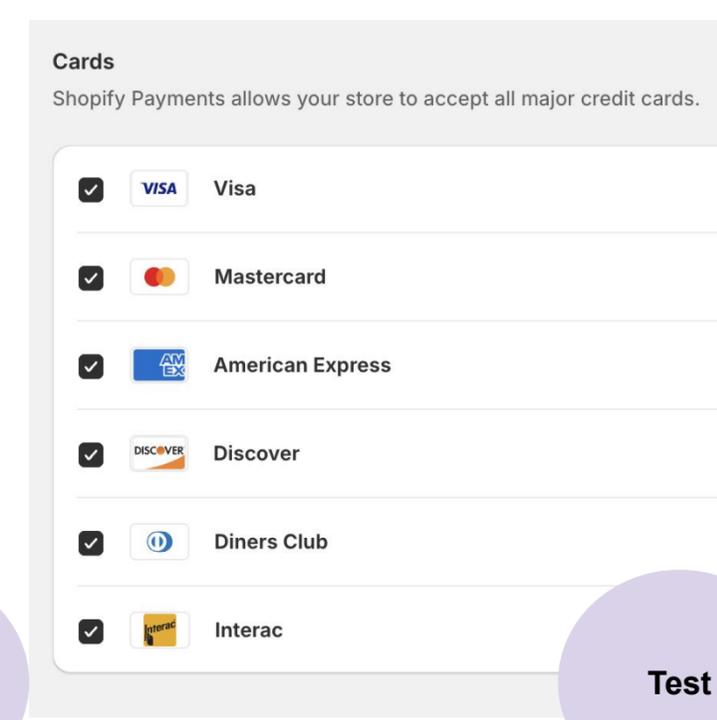
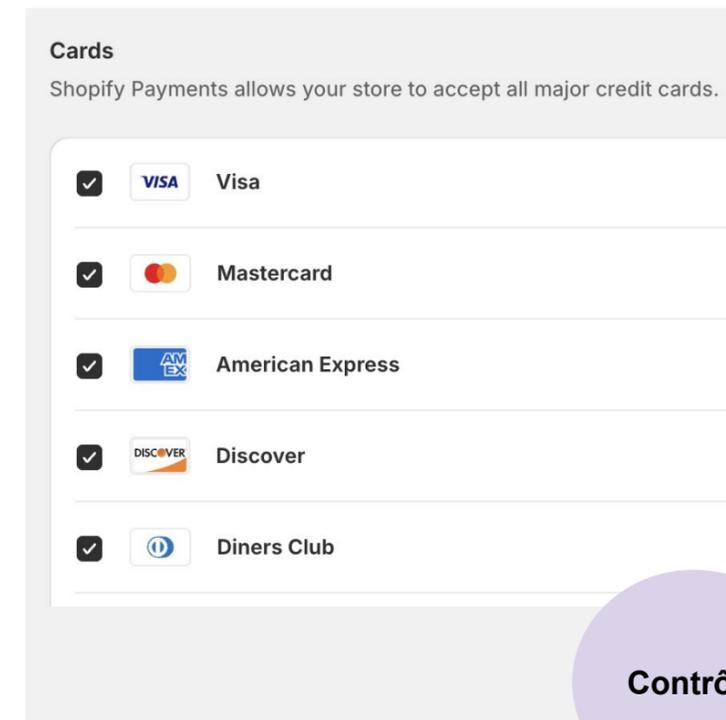
Ventes moyennes des marchands

Design

Randomisation des marchands entre le groupe contrôle et test

Problème

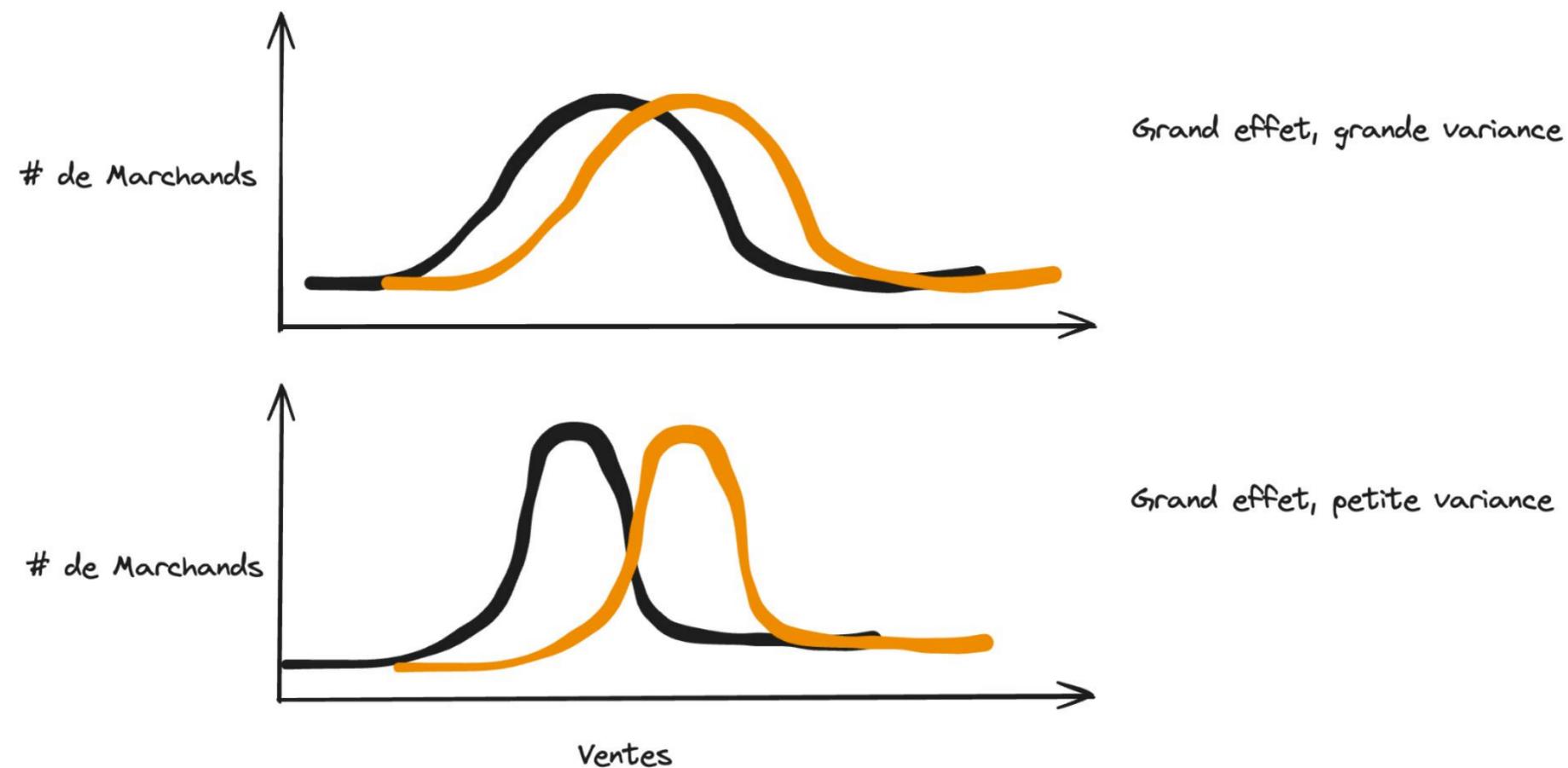
Peu de marchands au Canada et les ventes moyennes ont une très grande variance. Peu de puissance.



# Détecter des effets plus petits avec la même taille d'échantillon

Solution

Essayons de réduire la variance



# Détecter des effets plus petits avec la même taille d'échantillon

## Approche

- Controlled-experiment Using Pre-Experiment Data (CUPED) [Deng et al 2013](#)
- Seulement avec une variable continue
- Nécessite des données dans la période pré-expérimentale (catégorielle ou continue)

## Exemple

Observation	Ventes (pendant l'expérience)	Ventes (14 jours avant l'expérience)	Visiteurs (14 jours avant l'expérience)	Commandes (14 jours avant l'expérience)
1	200\$	190\$	30	5
2	1200\$	1000\$	250	8
3	50\$	10\$	5	1
...	...	...	...	...

# Détecter des effets plus petits avec la même taille d'échantillon

## Approche

- Controlled-experiment Using Pre-Experiment Data (CUPED) [Deng et al.](#)
- Seulement avec une variable continue
- Nécessite des données dans la période pré-expérimentale (catégorielle ou continue)

## Exemple

Observation	Ventes (pendant l'expérience)	Ventes (14 jours avant l'expérience)	Visiteurs (14 jours avant l'expérience)	Commandes (14 jours avant l'expérience)
1	200\$	190\$	30	5
2	1200\$	1000\$	250	8
3	50\$	10\$	5	1
...	...	...	...	...

# Détecter des effets plus petits avec la même taille d'échantillon

$$Ventes_{ajusté} = Ventes - \theta (Ventes_{pré} - Moyenne(Ventes_{pré}))$$

$$\text{où } \theta = Cov(Ventes, Vente_{pré}) / Var(Ventes_{pré})$$

- Estimateur sans biais
- Réduction de la variance par un facteur  $\rho^2$  où  $\rho$  est la corrélation entre la variable d'intérêt et la variable pré expérimentale.
- [Blogue d'un ancien collègue](#)

# En bref

- Les simulations sont un outil en or à ne pas négliger
- Innovation requiert de comprendre des articles scientifiques, blogues, contenu des conférences, et l'appliquer à notre vie de tous les jours. Ça prend une bonne base en statistique.
- Vous passez aussi les compagnies en entrevue.
- Rechercher l'inconfort ça veut dire qu'on apprend!



# Questions?