

# RAPPORT A/B TESTING

RetailRocket - Dataset Events

## 1. CONTEXTE DU TEST

**Ce que j'ai fait :** J'ai analysé les données RetailRocket pour tester si une nouvelle variante (Groupe B) améliore le taux d'ajout au panier par rapport à la version actuelle (Groupe A).

**Ma métrique :** J'ai mesuré le taux d'ajout au panier = (Nombre d'ajouts) / (Nombre de vues)

**Données utilisées :**

- Dataset : RetailRocket avec 2,756,101 événements
- Visiteurs uniques : 1,407,500
- Distribution : 50% Groupe A / 50% Groupe B

J'ai séparé les visiteurs de manière aléatoire pour éviter les biais. Chaque visiteur a été assigné une seule fois à un groupe et a conservé cette assignation pendant toute l'analyse.

## 2. MÉTHODOLOGIE

**Comment j'ai procédé :**

Premièrement, j'ai nettoyé les données pour garder uniquement les événements pertinents : les "view" (vues) et les "addtocart" (ajouts au panier). J'ai ensuite assigné aléatoirement chaque visiteur unique soit au groupe A soit au groupe B (50/50).

Ensuite, j'ai utilisé un test statistique (Z-test de proportion) pour comparer les deux groupes. C'est une méthode standard pour tester si deux pourcentages sont vraiment différents ou si c'est juste du hasard.

**Les hypothèses que j'ai posées :**

- H0 (Hypothèse nulle) : Les deux groupes ont le même taux de conversion
- H1 (Hypothèse alternative) : Les deux groupes sont différents
- Seuil accepté : 5% (c'est l'erreur que je suis prêt à accepter)

## 3. MES RÉSULTATS

Métrique	Groupe A	Groupe B	Différence
Nombre de vues	1,329,582	1,334,730	+5,148
Ajouts au panier	34,496	34,836	+340
Taux (%)	2.5945%	2.6100%	+0.0155%
Lift (%)	-	-	+0.60%

**Ce que j'observe :** Le Groupe B a un taux légèrement plus élevé : 2.61% contre 2.59% pour le Groupe A. Cela représente seulement 340 ajouts au panier supplémentaires sur plus d'un million d'événements.

En termes simples : sur 100 visiteurs qui voient le produit, 2.59 l'ajoutent au Groupe A, et 2.61 l'ajoutent au Groupe B. La différence est minuscule.

**Interprétation :** Le Groupe B montre une augmentation légère du taux d'ajout au panier de +0.0155 points de pourcentage (+0.60% relatif). Cette différence est cependant **très petite** et ne représente que 340 ajouts supplémentaires sur 1,3M d'événements.

## 4. LE TEST STATISTIQUE

Paramètre	Valeur
Proportion A	2.5945%
Proportion B	2.6100%
Proportion moyenne (pooled)	2.6022%
Z-score calculé	-0.7929
p-valeur obtenue	0.4279
Seuil accepté	0.05
Résultat	NON SIGNIFICATIF

**Explication simple :** La p-valeur est de 0.4279, ce qui signifie qu'il y a 42.79% de chance que cette différence soit due au hasard. C'est bien au-dessus du seuil de 5% que j'ai fixé.

En d'autres termes, si je faisais ce test 100 fois sur un produit où les deux versions sont vraiment identiques, je verrais une différence aussi grande que celle-ci environ 43 fois. C'est très fréquent. Donc je ne peux pas affirmer que le Groupe B est vraiment meilleur.

## 5. MON ANALYSE

### 1) Taille de la différence :

La différence observée est de 0.0155 points de pourcentage. C'est extrêmement petit. En pratique, cela signifie que sur un million de visiteurs, on aurait seulement 340 ajouts supplémentaires. C'est négligeable.

### 2) Est-ce réel ou du hasard ? :

Avec une p-valeur de 0.4279, il y a 42.79% de chance que cette petite différence soit simplement du hasard. C'est beaucoup trop élevé pour affirmer que le Groupe B est vraiment meilleur. Je ne peux pas la considérer comme une différence réelle.

### 3) Risques et limitations :

Si je déployais le Groupe B et que la différence est due au hasard, j'aurais changé mon système pour rien. C'est un risque inutile. D'autre part, je sais qu'il existe toujours une petite chance (5%) que je me trompe en disant qu'il n'y a pas de différence. Mais vu la petite taille de la différence observée, ce risque est très faible.

### 4) Bénéfice vs Coût :

Même si le Groupe B était meilleur, le bénéfice serait minime (340 ajouts). Le coût de déployer une nouvelle version (infrastructure, tests, déploiement, maintenance) serait probablement plus grand que le bénéfice attendu.

## 6. MA DÉCISION

### J'ai décidé de garder le Groupe A en production

Voici pourquoi :

1. **Pas de preuve de supériorité :** Avec une p-valeur de 0.4279, je n'ai pas de preuve que le Groupe B soit vraiment meilleur. C'est probablement du hasard.

2. **Différence minuscule :** Même si le Groupe B avait un léger avantage, 340 ajouts supplémentaires sur un million d'événements, c'est un changement trop petit pour justifier un déploiement.

3. **Pas de bénéfice prouvé** : Je ne risque pas ma version actuelle pour une amélioration qui n'existe probablement pas.

4. **Prudence** : En l'absence de preuve claire, la meilleure décision est de conserver ce qui fonctionne actuellement.

**Coûts vs Bénéfices :**

- Bénéfice potentiel : 340 ajouts = gain très faible
- Coût de déploiement : Tests, déploiement, monitoring = coût réel
- Conclusion : Les coûts dépassent largement les bénéfices

## 7. CE QUE JE VAIS FAIRE ENSUITE

**Immédiatement :**

- Je garde la version actuelle (Groupe A) en production
- J'archive tous les résultats du test pour la traçabilité
- Je documente cette analyse pour les équipes

**Prochaines étapes (1-2 semaines) :**

- Je vais recueillir les retours des utilisateurs pour comprendre comment améliorer le produit
- J'analyserai les insights des utilisateurs pour identifier de meilleures idées
- J'esquisserai une variante C avec des changements plus importants

**Pour les tests futurs :**

- Je vais attendre d'avoir une idée d'amélioration plus significative
- Je viserai une différence d'au moins 0.5% pour que cela vaille la peine de tester
- Si nécessaire, j'augmenterai la durée du test pour avoir plus de données

La clé est de ne pas déployer quelque chose sans preuve solide d'amélioration.

## 8. CONCLUSION

J'ai réalisé un test A/B complet sur les données RetailRocket pour évaluer si une nouvelle variante (Groupe B) améliorerait le taux d'ajout au panier. Après avoir analysé plus d'un million d'événements provenant de 1.4 millions de visiteurs uniques, les résultats sont clairs :

Il n'y a pas de différence statistiquement significative entre les deux groupes. La p-valeur de 0.4279 indique que la petite différence observée (0.0155%) est très probablement due au hasard, et non à une véritable amélioration.

Pour cette raison, j'ai décidé de **conserver la version actuelle (Groupe A) en production**. C'est la décision la plus raisonnable : ne pas risquer un changement sans preuve solide d'amélioration.

Je vais maintenant me concentrer sur l'identification de véritables améliorations du produit en écoutant les utilisateurs et en testant des variantes plus significativement différentes à l'avenir.

**Date du rapport** : 05/12/2025

**Analyst** : Kouame Adamou Kevin

**Source** : RetailRocket Events Dataset