



## CanPath Synthetic Dataset Workshop User Acknowledgement

### Background

The Canadian Partnership for Tomorrow's Health (CanPath) has created a mock dataset, the *CanPath Synthetic Dataset*, designed to mimic the real data collected by CanPath. The Synthetic Dataset offers a valuable resource for learning about population health research and may be used for training and educational purposes.

### Purpose

To provide users with practical experience using a synthetic dataset that closely resembles real-world population-level data.

### Availability

The dataset will be available to the registered hackathon participants through a **secure OneDrive folder** managed under **The Université de Montréal institutional account**, during the duration of the event. Access will be restricted to authorized participants only, and sharing of access credentials is strictly prohibited.

### Content

The synthetic dataset includes over 100,000 observations and 900+ mock variables covering socio-demographic information, lifestyle behaviours, health perceptions, self-reported diseases, and environmental variables across three timepoints.

### Usage

The dataset is strictly for training and educational purposes during the hackathon and should not be used for research or research-related publications.

Beyond this hackathon, the CanPath Synthetic dataset is available at no cost for use in academic courses and for training purposes. Please contact [apply@canpath.ca](mailto:apply@canpath.ca) for more information. Users interested in accessing real-world population health data for research purposes are invited to [apply to access](#) CanPath data.

This **Synthetic Dataset User Acknowledgement** sets out the terms and conditions governing access to and use of the Synthetic Dataset for educational and training purposes during the hackathon. By signing this acknowledgment, the user agrees to adhere to the following terms and conditions:

1. **Ownership and Intellectual Property**: The user shall not have intellectual property rights in any results arising from the Synthetic Dataset.
2. **Liability**: CanPath makes no representations or warranties regarding the accuracy, completeness, or suitability of the Synthetic Dataset for any purpose. The use of the Synthetic Dataset is at the user's own risk, and CanPath shall not be liable for any damages or losses arising from access to or use of the dataset.

3. **Terms of Use:** The dataset is exclusively for training and educational purposes during the hackathon. The Synthetic Dataset and any results arising from analyses of the dataset may not be used for research activities, publication, or presentation in any academic or public forum. Users are permitted to share their code on public platforms like GitHub. However, public distribution, replication, or sharing of the dataset itself is strictly prohibited.
4. **Access:** The Synthetic Dataset will be made accessible through a **secure OneDrive folder** managed by the **Université de Montréal** and restricted to registered participants of the Digital Health Hackathon.
5. **Destruction:** Upon completion of the hackathon, the user shall cease using the dataset. The user will **delete all local copies** and ensure that no portion of the dataset is retained, transferred, or reproduced in any form.

By signing below, I confirm that I have read, understood, and agree to abide by the terms outlined above.

Name:

Signature:

Date:

# **Engagement d'utilisation de l'ensemble de données synthétiques de CanPath**

## **Contexte**

Le Partenariat canadien pour la santé de demain (CanPath) a créé un jeu de données fictif, *l'ensemble de données synthétiques de CanPath*, conçu pour imiter les données réelles recueillies par CanPath.

Ce jeu de données constitue une ressource précieuse pour l'apprentissage et la formation en recherche sur la santé des populations, et peut être utilisé à des fins pédagogiques et de formation.

## **Objectif**

Offrir aux utilisateurs une expérience pratique d'utilisation d'un jeu de données synthétiques reproduisant fidèlement des données réelles à l'échelle populationnelle.

## **Disponibilité**

Le jeu de données sera mis à la disposition des participant·es inscrit·es au hackathon à travers un dossier OneDrive sécurisé géré sous un compte institutionnel de l'Université de Montréal, pendant toute la durée de l'événement.

L'accès sera strictement réservé aux participant·es autorisé·es, et le partage des identifiants d'accès est formellement interdit.

## **Contenu**

Le jeu de données synthétiques comprend plus de 100 000 observations et plus de 900 variables simulées, couvrant des informations sociodémographiques, des comportements de vie, des perceptions de santé, des maladies autodéclarées et des variables environnementales à travers trois périodes de collecte.

## **Utilisation**

Le jeu de données est destiné exclusivement à des fins pédagogiques et de formation dans le cadre du hackathon. Il ne peut en aucun cas être utilisé à des fins de recherche, ni faire l'objet de publications scientifiques, communications, ou présentations publiques.

En dehors de ce hackathon, le jeu de données synthétiques de CanPath est disponible gratuitement pour une utilisation dans le cadre de cours universitaires ou d'activités de formation. Pour plus d'informations, veuillez contacter [apply@canpath.ca](mailto:apply@canpath.ca).

Les utilisateurs souhaitant accéder à des données réelles pour des travaux de recherche sont invités à déposer une demande d'accès officielle auprès de CanPath.

## **Propriété intellectuelle**

L'utilisateur ou l'utilisatrice ne détient **aucun droit de propriété intellectuelle** sur les résultats ou produits dérivés du jeu de données synthétiques.

## **Responsabilité**

CanPath ne fait aucune déclaration ni garantie quant à l'exactitude, l'exhaustivité ou la pertinence du jeu de données synthétiques pour un usage particulier. L'utilisation du jeu de données est faite aux risques de l'utilisateur ou de l'utilisatrice, et CanPath ne sauraient être tenus responsables de tout dommage ou perte découlant de l'accès ou de l'utilisation de ces données.

## **Conditions d'utilisation**

Le jeu de données est réservé exclusivement à la formation et à l'enseignement durant le hackathon. Les analyses et résultats produits à partir du jeu de données ne peuvent être utilisés à des fins de recherche, de publication ou de diffusion académique.

Les participant·es sont autorisé·es à partager leur code ou scripts développés durant le hackathon sur des plateformes publiques telles que GitHub. Cependant, toute distribution publique, reproduction ou partage du jeu de données lui-même est strictement interdit.

## **Accès**

Le jeu de données synthétiques sera accessible via un dossier OneDrive sécurisé, géré par l'Université de Montréal, et réservé aux participant·es inscrit·es au Hackathon en santé numérique. Les utilisateur·rices ne doivent ni télécharger, ni copier, ni transférer le jeu de données hors de cet environnement sécurisé.

## **Destruction**

À la fin du hackathon, les utilisateur·rices doivent cesser toute utilisation du jeu de données.

Ils et elles doivent supprimer toutes les copies locales et s'assurer qu'aucune portion du jeu de données n'est conservée, transférée ou reproduite, sous quelque forme que ce soit.

En signant ci-dessous, je confirme avoir lu, compris et accepté les conditions énoncées ci-dessus.

**Nom :** \_\_\_\_\_

**Signature :** \_\_\_\_\_

**Date :** \_\_\_\_\_

## 6. hackathonD