# Exploratory Data Analysis Project

## Project Objective

This project aims to visualize the count of medications in Canada by their schedule, status, and its anatomical therapeutic code using Python and Tableau.

## Introduction

The pharmaceutical industry is a critical sector that has various impact on our day-to-day life. In Canada, all the drug products are strictly regulated by Health Canada. The drug products are classified into 11 categories by Health Canada. However, for the patients, the most important feature of a drug is probably whether its prescription drug or over-the-counter drug, as the in order to obtain a prescription drug, a prescription from a healthcare practitioner is mandatory. Therefore, the final delivery mainly focuses on the comparison between the number of prescription drug and OTC drug by different anatomical therapeutic group.

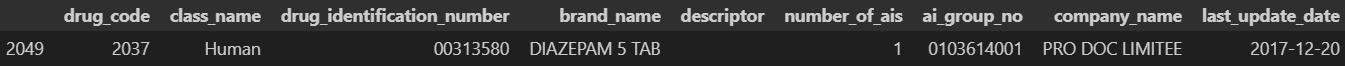
## Data Description

The data were collected from [Drug Product Database (DPD) API](https://health-products.canada.ca/api/documentation/dpd-documentation-en.html#a2). Four datasets obtained from the website will be cleaned and merged: Drug Product, Schedule, Product Status and Therapeutic Class. The analysis should primarily focus on Product Schedule and Therapeutic Class. Product Status table can be also used as it indicates since when the drug has been on market and whether the drug continued to be on market.

1. Drug Product (key: drug\_code)

A close-up of a product

Description automatically generated

Example:

Note that the ai\_group\_no ([AIG number](https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/drug-product-database/terminology.html)) will indicate the active compounds in the medication

First 2 digits indicates the number of active ingredients, the next 5 digits identifies the unique groups of active ingredients, and finally the last 3 digits indicates the active ingredient group strength.

1. Schedule (key: drug\_code)

A close-up of a text

Description automatically generated

Example:

A black and white text

Description automatically generated

Below are all the drug schedules (FDA standards) which can be grouped into 7 larger Categories

1. Prescription

* Prescription (prescription drugs included in the Prescription Drug List)
* Prescription Recommended (drugs that are recommended to be listed on the Prescription Drug List)

1. Controlled Drugs

* Schedule G (control drugs)
* **Schedule G (*Controlled Drugs and Substances Act* [CDSA] III)**
* **Schedule G (CDSA IV)**

1. Special Schedule

* Schedule C (drugs listed in Schedule C of the *Food and Drugs Act*, for example, radiopharmaceutical drugs)
* Schedule D (drugs listed in Schedule D of the *Food and Drugs Act*, ie. biological products)

1. Narcotic

* Narcotic (*Narcotic Control Act*)
* **Narcotic (CDSA I)**
* **Narcotic (CDSA II)**

1. Targeted

* **Targeted (CDSA IV)**

1. Undetermined (in process)

* CDSA Recommended- Undergoing Regulatory Amendment to add this new substance to Controlled Drugs Substances Act

1. OTC and Ethical

* OTC (over the counter drugs that do not appear on a schedule or are not recommended to appear on any schedule)
* Ethical: a drug that, in accordance with Federal Legislation, does not require a prescription, but that is generally prescribed by a medical practitioner. Ethical products are unscheduled non-prescription professional use products (e.g. MRI contrast agents, hemodialysis solutions) and a few emergency use products (e.g. nitroglycerine)

1. Product Status (key: drug\_code)

A screenshot of a product status

Description automatically generated

Example:

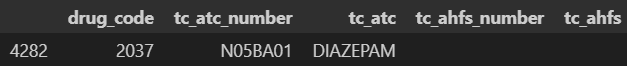


1. Therapeutic Class (key: drug\_code)

A close-up of a text

Description automatically generated

Example



Note that tc\_atc\_number indicates therapeutic uses. The meaning of the ACT code to its 2nd level is scraped from [Norwegian Institute of Public Health](https://atcddd.fhi.no/atc_ddd_index/)

## Methodology

Dashboards will be used for visualization and there will be interactive components:

* 1. Barplot: Distribution of drugs among different ATC code to first and second level.
  2. Time series lineplot: To visualize trend of drugs on market based on time

## Expected Deliverables

It should be a Tableau Dashboard, or even a Dashboard which can link to another Dashboard. Something like:

A screenshot of a computer screen

Description automatically generated

## Timeline and Task

8/19 – 8/23 = EDA + Terminology Check + Finalize Dataset + Finalize Proposal

8/26 – 8/30 = Learn Tableau + Make first Interactive Components

9/1 – 9/4 = Try to make the external link part work.

## Potential Challenge

1. Difficulty Using Tableau… - Learn Tableau
2. Difficulty understanding terminology and potential to give misleading explanations

Consult official website and professionals

1. Too many different categories in one variable according to different regulatory agencies.

Consult official website and professionals to come up with supergroups, or group smaller groups as others.