

# Structured Data Assignment

## Data Description:

- **Train.parquet** - Dataset to be used for training
- **Test.parquet** - Dataset to be used for testing

The data consists of three primary columns:

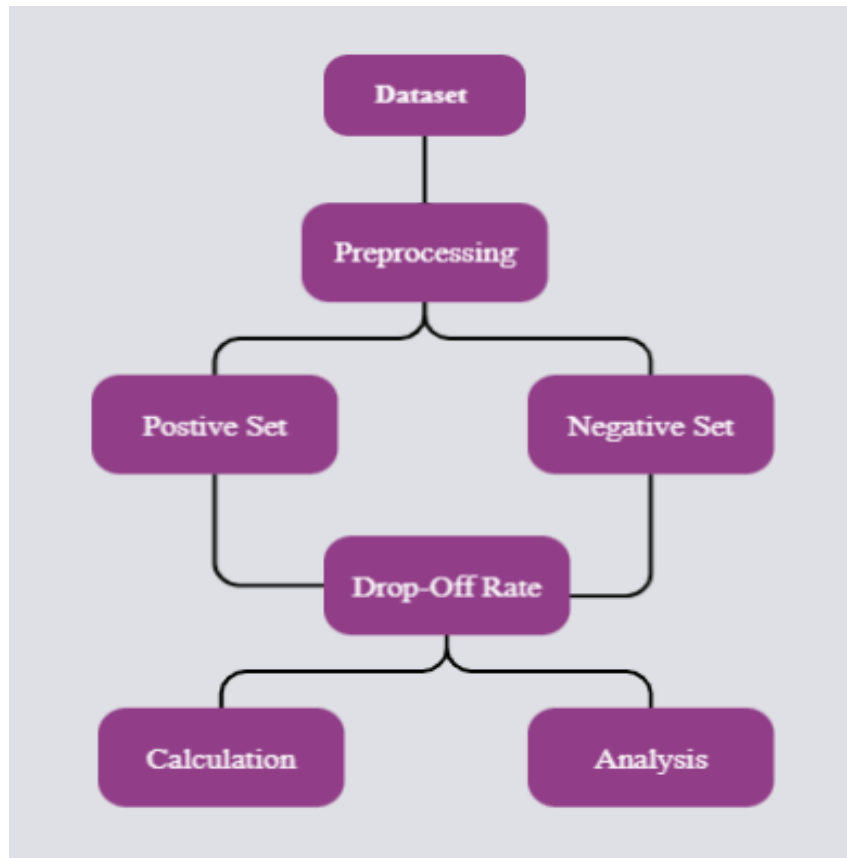
- **Patient-Uid:** A unique alphanumeric identifier assigned to each patient who are taking the medicine.
- **Date:** The specific date when the patient experienced the event.
- **Incident:** This column provides a description of the event that took place on the given date for the patient.

## Problem Statement 2

Drugs are generally administered/prescribed by the physicians for a certain period of time or they are administered at regular intervals, but for various reasons patients might stop taking the treatment. Consider following example for better understanding Let's say you get a throat infection, the physician prescribes you an antibiotic for 10 days, but you stop taking the treatment after 3 days because of some adverse events. In the above example ideal treatment duration is 10 days but patients stopped taking treatment after 3 days due to adverse events. Patients stopping a treatment is called drop-off. We want to study drop-off for "Target Drug", the aim is to generate insights on what events lead to patients stopping on "Target Drug". Assume ideal treatment duration for "Target Drug" is 1 year, come up with analysis showing how drop-off rate is, drop-off rate is defined as number of patients dropping off each month. Then come up

with analysis to generate insights on what events are driving a patient to stop taking “Target Drug”.

### **Process Flow:**



### **Implementation:**

#### **Initial analysis of data:**

- Loading the dataset

	Patient-Uid	Date	Incident
0	a0db1e73-1c7c-11ec-ae39-16262ee38c7f	2019-03-09	PRIMARY_DIAGNOSIS
1	a0dc93f2-1c7c-11ec-9cd2-16262ee38c7f	2015-05-16	PRIMARY_DIAGNOSIS
3	a0dc94c6-1c7c-11ec-a3a0-16262ee38c7f	2018-01-30	SYMPTOM_TYPE_0
4	a0dc950b-1c7c-11ec-b6ec-16262ee38c7f	2015-04-22	DRUG_TYPE_0
8	a0dc9543-1c7c-11ec-bb63-16262ee38c7f	2016-06-18	DRUG_TYPE_1

- Calculating the total number of distinct patients and the total number of distinct incident
- Grouping the data by Patient-Uid and Date

	Patient-Uid	Incident
0	a0db1e73-1c7c-11ec-ae39-16262ee38c7f	DRUG_TYPE_7 SYMPTOM_TYPE_2 DRUG_TYPE_7 SYMPTOM...
1	a0dc93f2-1c7c-11ec-9cd2-16262ee38c7f	DRUG_TYPE_0 DRUG_TYPE_2 DRUG_TYPE_0 PRIMARY_DI...
2	a0dc94c6-1c7c-11ec-a3a0-16262ee38c7f	DRUG_TYPE_0 PRIMARY_DIAGNOSIS DRUG_TYPE_7 DRUG...
3	a0dc950b-1c7c-11ec-b6ec-16262ee38c7f	DRUG_TYPE_0 DRUG_TYPE_7 DRUG_TYPE_2 PRIMARY_DI...
4	a0dc9543-1c7c-11ec-bb63-16262ee38c7f	DRUG_TYPE_1 TEST_TYPE_1 SYMPTOM_TYPE_8 DRUG_TY...
...	...	...
27028	a0f0d4c5-1c7c-11ec-bfec-16262ee38c7f	DRUG_TYPE_6 DRUG_TYPE_0 DRUG_TYPE_6 DRUG_TYPE_...
27029	a0f0d4f4-1c7c-11ec-b144-16262ee38c7f	DRUG_TYPE_6 DRUG_TYPE_8 DRUG_TYPE_1 DRUG_TYPE_...
27030	a0f0d523-1c7c-11ec-89d2-16262ee38c7f	DRUG_TYPE_6 DRUG_TYPE_1 DRUG_TYPE_9 DRUG_TYPE_...
27031	a0f0d553-1c7c-11ec-a70a-16262ee38c7f	DRUG_TYPE_9 SYMPTOM_TYPE_7 DRUG_TYPE_2 DRUG_TY...
27032	a0f0d582-1c7c-11ec-a6c1-16262ee38c7f	DRUG_TYPE_6 DRUG_TYPE_1 DRUG_TYPE_6 DRUG_TYPE_...

27033 rows × 2 columns

## Positive-set and Negative-set segregation:

We have organized each patient's information into a single row, and their entire history is now sorted by timestamps and stored as features. Additionally, the classes have been segregated and stored in the "classes" column 0's and 1's.

	Patient-Uid	Incident classes	
0	a0db1e73-1c7c-11ec-ae39-16262ee38c7f	DRUG_TYPE_7 SYMPTOM_TYPE_2 DRUG_TYPE_7 SYMPTOM...	0
1	a0dc93f2-1c7c-11ec-9cd2-16262ee38c7f	DRUG_TYPE_0 DRUG_TYPE_2 DRUG_TYPE_0 PRIMARY_DI...	0
2	a0dc94c6-1c7c-11ec-a3a0-16262ee38c7f	DRUG_TYPE_0 PRIMARY_DIAGNOSIS DRUG_TYPE_7 DRUG...	0
3	a0dc950b-1c7c-11ec-b6ec-16262ee38c7f	DRUG_TYPE_0 DRUG_TYPE_7 DRUG_TYPE_2 PRIMARY_DI...	0
4	a0dc9543-1c7c-11ec-bb63-16262ee38c7f	DRUG_TYPE_1 TEST_TYPE_1 SYMPTOM_TYPE_8 DRUG_TY...	0
...	...	...	...
27028	a0f0d4c5-1c7c-11ec-bfec-16262ee38c7f	DRUG_TYPE_6 DRUG_TYPE_0 DRUG_TYPE_6 DRUG_TYPE_...	1
27029	a0f0d4f4-1c7c-11ec-b144-16262ee38c7f	DRUG_TYPE_6 DRUG_TYPE_8 DRUG_TYPE_1 DRUG_TYPE_...	1
27030	a0f0d523-1c7c-11ec-89d2-16262ee38c7f	DRUG_TYPE_6 DRUG_TYPE_1 DRUG_TYPE_9 DRUG_TYPE_...	1
27031	a0f0d553-1c7c-11ec-a70a-16262ee38c7f	DRUG_TYPE_9 SYMPTOM_TYPE_7 DRUG_TYPE_2 DRUG_TY...	1
27032	a0f0d582-1c7c-11ec-a6c1-16262ee38c7f	DRUG_TYPE_6 DRUG_TYPE_1 DRUG_TYPE_6 DRUG_TYPE_...	1

## Drop-off rate:

To study drop-off for the "Target Drug" and generate insights into the events leading to patients stopping its usage, the aim is to analyze how the drop-off rate varies as the number of patients discontinuing the drug each month changes. It is important to note that the ideal treatment duration for the "Target Drug" is one year. Next, we perform an analysis aimed at uncovering the factors or events that influence a patient's decision to discontinue the use of the "Target Drug."

- Filtering the dataset to include only patients who have used "TARGET\_DRUG."
- Creating a list of incident dates for each patient and incorporating it into the dataset.
- Compiling a list of dates when the "target\_drug" was taken by each patient and adding it to the dataset.

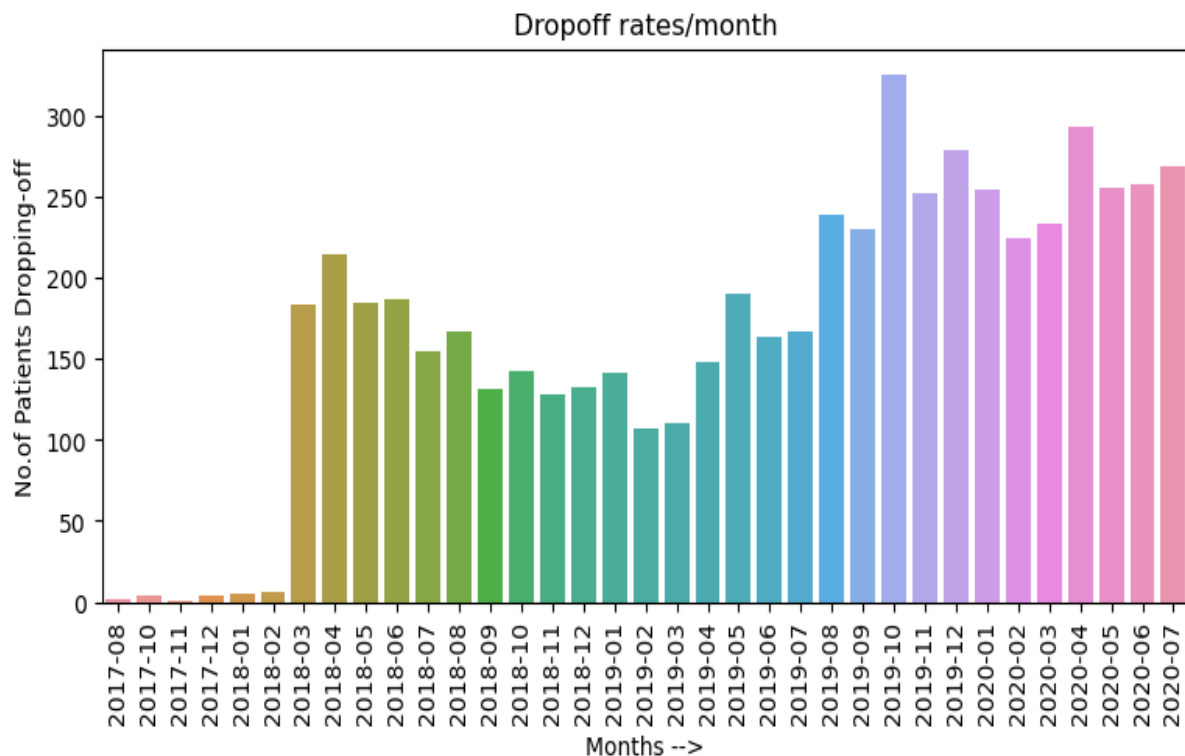
### Calculating drop-off rate:

Given that the recommended timeframe for "TARGET\_DRUG" usage is one year, we can examine when patients commenced and discontinued its usage. We can then verify whether the duration aligns with this one-year timeframe. In cases where the duration deviates from one year, we can consider it as a discontinuation event and make a record of the cessation date for subsequent analysis.

	Months	No.of Patients Dropping-off
0	2017-08	2
1	2017-10	4
2	2017-11	1
3	2017-12	4
4	2018-01	5
5	2018-02	6
6	2018-03	183
7	2018-04	215
8	2018-05	185
9	2018-06	187
10	2018-07	155
11	2018-08	167
12	2018-09	131

## Analysis of drop-off rates:

- To understand why patients stop using the "target\_drug," we gather a list of symptoms before their discontinuation dates.
- Patients who've discontinued the drug have been identified, and we analyze the events leading up to their discontinuation.
- We focus on an average of three incidents that occurred just before a patient's final use of the "target\_drug," and these events are stored in the "**dic**" dictionary, detailing the frequency of each event's occurrence.



Events might have led to the stoppage of the target drug.

