Structured Data Assignment – Akaike
By Kavibarathi K

Problem Statement:

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 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".

Necessary Libraries:

Python libraries that are necessary for performing desired functions.

```
# import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
```

Reading the dataset:

```
#  loading the dataset
data =pd.read_parquet(r"D:\certificates\Structured_Data_Assignment\Structured_Data_Assignment\train.parquet")
data
```

	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
...	...	...	...
29080886	a0ee9f75-1c7c-11ec-94c7-16262ee38c7f	2018-07-06	DRUG_TYPE_6
29080897	a0ee1284-1c7c-11ec-a3d5-16262ee38c7f	2017-12-29	DRUG_TYPE_6
29080900	a0ee9b26-1c7c-11ec-8a40-16262ee38c7f	2018-10-18	DRUG_TYPE_10
29080903	a0ee1a92-1c7c-11ec-8341-16262ee38c7f	2015-09-18	DRUG_TYPE_6
29080911	a0ee146e-1c7c-11ec-baee-16262ee38c7f	2018-10-05	DRUG_TYPE_1

3220868 rows × 3 columns

Finding out number of drop-offs per month for one year period:

This will help us to identify the month where patients are stopped taking their prescriptions.

finding number of dropoff's per month for a period of one year

```
# creating dropoff object and getting the unique values from it
dropoff_rates = target.groupby('Month')['Patient-Uid'].nunique()
dropoff_rates
```

```
Month
1     3882
2     3681
3     3975
4     4442
5     4834
6     4887
7     5171
8     5050
9     3457
10    3719
11    3590
12    3662
Name: Patient-Uid, dtype: int64
```

```
# visualizing the dropoff rates for better understanding
plt.figure(figsize = (8, 5))
dropoff.plot()
plt.xlabel("Month")
plt.ylabel("Dropoff_numbers")
plt.title("DROP-OFF RATES FOR EVERY MONTH FOR TARGET DRUG")
plt.show
```
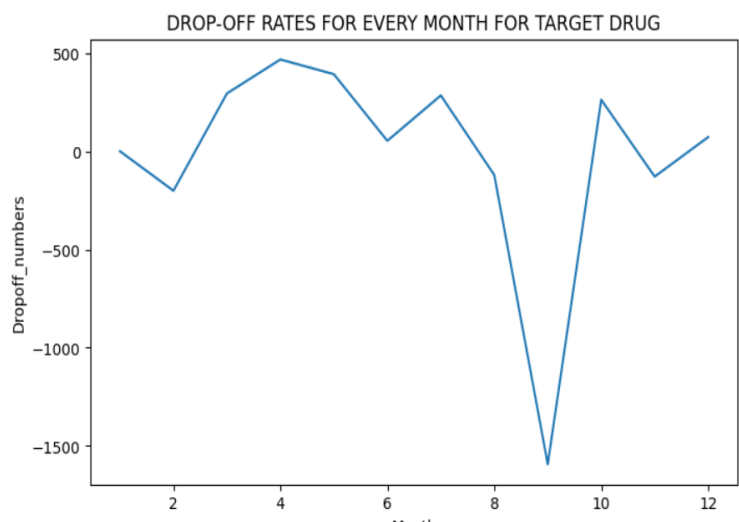

Reason for drop-off's:

This particular step's is very helpful in finding out that what event or incident that leads to the dropping off the prescribed drugs.

```
# reasons that lead to drop off
drop_reasons = drop_reasons[drop_reasons['Incident'] != 'TARGET DRUG']
drop_reasons
```

	Patient-Uid	Date	Incident
8	a0e9c384-1c7c-11ec-81a0-16262ee38c7f	2018-02-22	SYMPTOM_TYPE_6
22	a0e9c3b3-1c7c-11ec-ae8e-16262ee38c7f	2018-02-21	SYMPTOM_TYPE_6
23	a0e9c3e3-1c7c-11ec-a8b9-16262ee38c7f	2017-05-11	SYMPTOM_TYPE_10
29	a0e9c414-1c7c-11ec-889a-16262ee38c7f	2019-11-22	PRIMARY_DIAGNOSIS
32	a0e9c443-1c7c-11ec-9eb0-16262ee38c7f	2020-01-28	PRIMARY_DIAGNOSIS
...	...	...	...
29080886	a0ee9f75-1c7c-11ec-94c7-16262ee38c7f	2018-07-06	DRUG_TYPE_6
29080897	a0ee1284-1c7c-11ec-a3d5-16262ee38c7f	2017-12-29	DRUG_TYPE_6
29080900	a0ee9b26-1c7c-11ec-8a40-16262ee38c7f	2018-10-18	DRUG_TYPE_10
29080903	a0ee1a92-1c7c-11ec-8341-16262ee38c7f	2015-09-18	DRUG_TYPE_6
29080911	a0ee146e-1c7c-11ec-baee-16262ee38c7f	2018-10-05	DRUG_TYPE_1

1358355 rows × 3 columns

Frequency of symptoms:

The frequency of symptoms played a major role in dropping off the prescribed drugs.

```
# to calculate the frequency of each event leading to drop-off
event_frequency = drop_reasons['Incident'].value_counts()
event_frequency
```

```
Incident
PRIMARY_DIAGNOSIS   237455
DRUG_TYPE_6         201366
DRUG_TYPE_1         160903
DRUG_TYPE_2         146560
DRUG_TYPE_7         116819
DRUG_TYPE_0         111912
DRUG_TYPE_3          72745
DRUG_TYPE_8          43289
TEST_TYPE_1          42246
DRUG_TYPE_9          38383
DRUG_TYPE_11         25627
SYMPTOM_TYPE_0       24475
DRUG_TYPE_5          15196
SYMPTOM_TYPE_6       14507
TEST_TYPE_0          12642
DRUG_TYPE_10         12349
SYMPTOM_TYPE_7       10098
DRUG_TYPE_14          9901
DRUG_TYPE_13          6344
TEST_TYPE_3           5359
DRUG_TYPE_12          4958
SYMPTOM_TYPE_1        4511
TEST_TYPE_2           4199
SYMPTOM_TYPE_2        3887
SYMPTOM_TYPE_5        3602
SYMPTOM_TYPE_15       3515
SYMPTOM_TYPE_14       3302
SYMPTOM_TYPE_8        2886
```

Visualization of frequency of drop-off's: