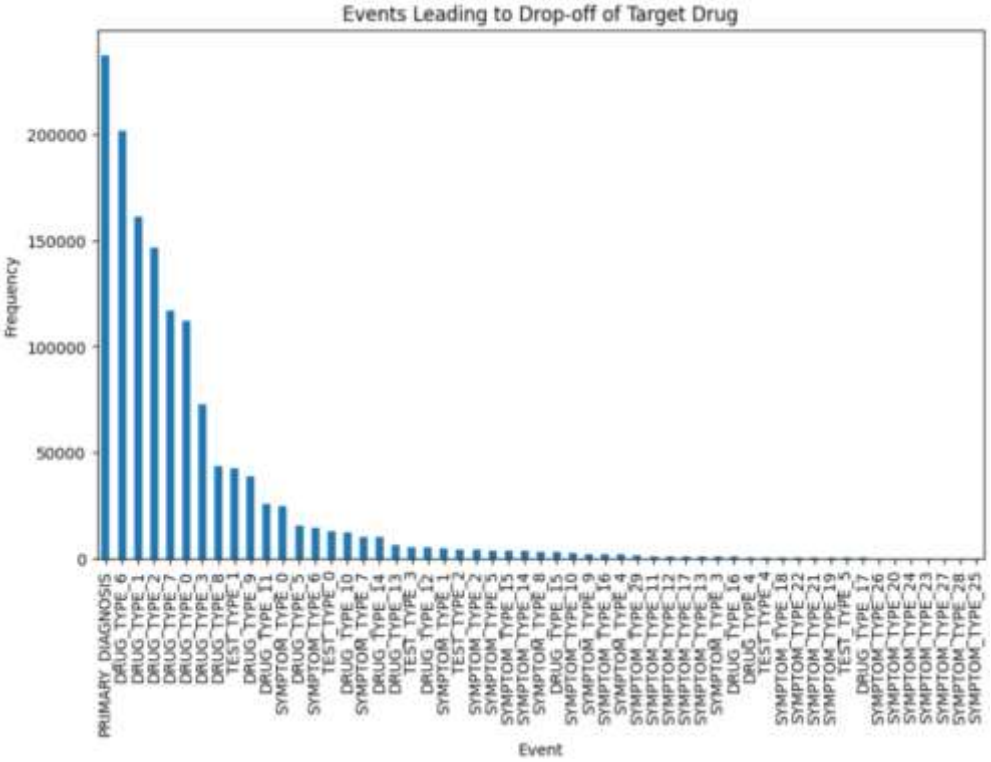
STRUCTURED DATA- PROBLEM STATEMENT-2

The following process are followed to solve this problem:

- Load the Dataset train.parquet file
- Split the data into TARGET DRUG as eligible, other than Target Drug is not eligible.
- Preprocess the date columns
- Extract the Month from the date column.
- Calculate the No.of.Days between the Incident for all unique patients.
- It shows the Days taken by the patients for the next drug
- Calculate the Dropoff-rate by Considering the Days below a (365days) year.
- Visualize the Dropoff –Rates in different graphs to get an better understanding.
- A line graph shows the Dropoff rates over the period of time.
- In addition to that, I have done bar graph to get more clear insights from the drop-off rates.

1. **frequency count of incident**
    - with the help of bar graph I have created the frequent of each incident

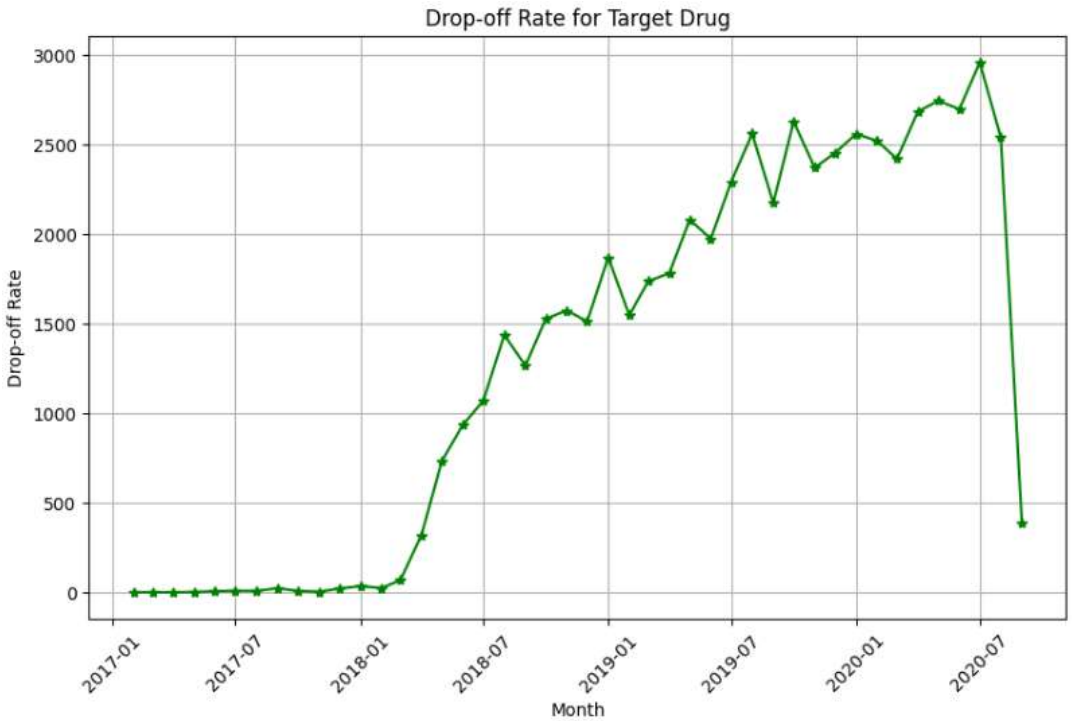


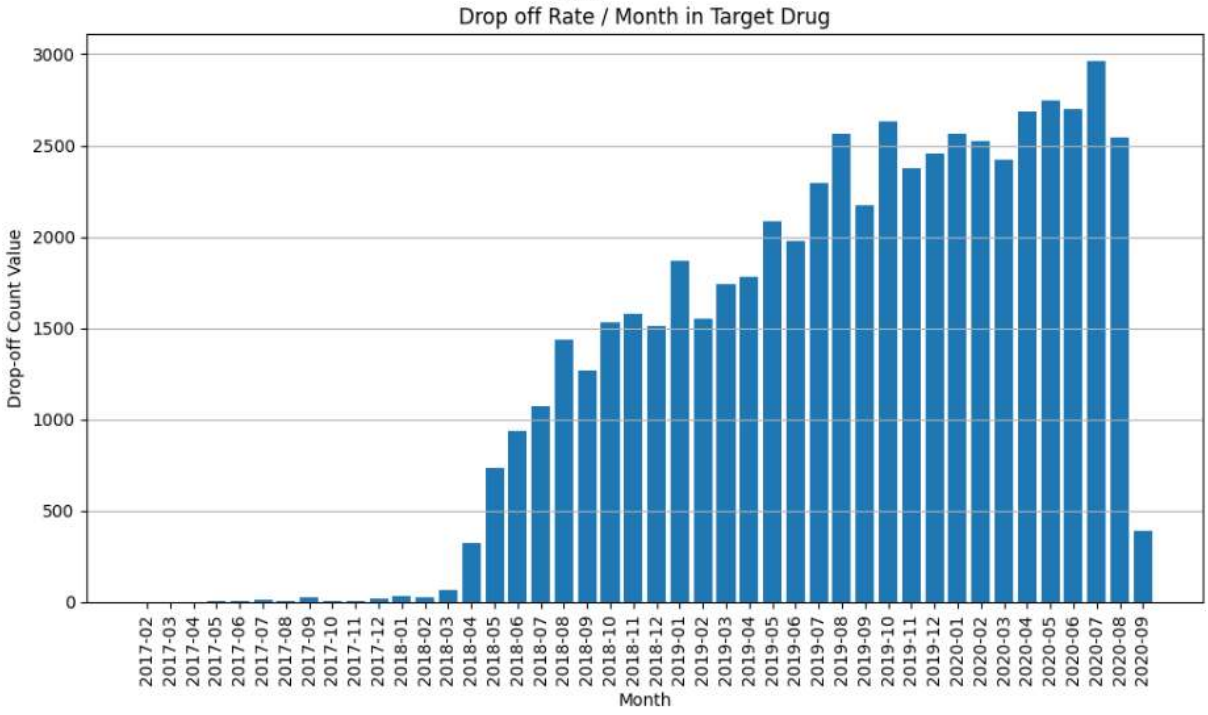
Interpretation:

The following medical events are more frequent by comparing other events:

1. Primary Diagnosis
2. Drug_Type_6
3. Drug_Type_1
4. Drug_Type

2.Drop rate for Target Drug





Interpretation:

- From above result we can clearly understand the usage of target drug from 2018 jan onwards gradually increase up to 2020 aug
- And then the sudden down fall happen after the month of august 2020.
- The highest peak in drop-off rate in July 2020 around 3000 per month