STRUCTURED DATA ASSIGNMENT

PROBLEM STATEMENT-2

Data Pre-processing

1.Data Jar

1.1 Data Cleaning

- Imported necessary libraries
- Load the dataset
- Checked data types and Structure of data
- Removed duplicate values
- Checked missing values

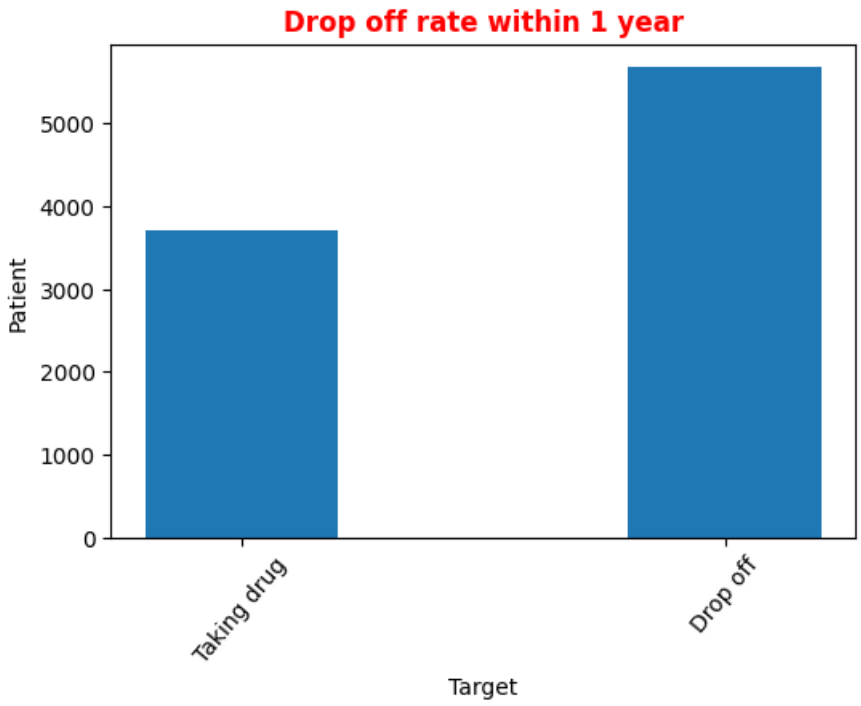
1.2 Restructuring Data

- Taken data based on the patient how taken the target drug and formed new dataset
- Added temp date column based on the first target drug date taken by each patient
- Then added a new no of month column to find how many months the patient taken the drug
- Created target column if the patient stop using the drug within one-year target value will be 1 if not 0

2.EDA

1.1 How many patients drop off within 1 year:

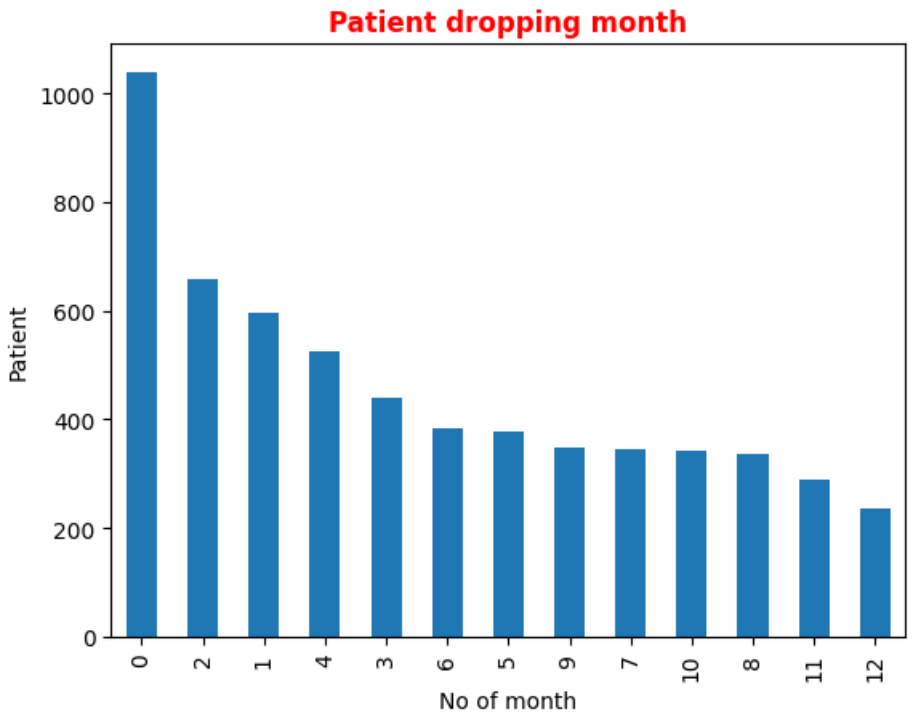
- Create a temp dataset with patient id and target column
- Based on the target column created the plot



- As we can see 9374 people taking the Target Drug in that 5675 people stop using the Target drug within one year.
- More than 60% of the people stop using the Target Drug

1.2 Each Month how many patient dropping off:

- Create a temp dataset based on the no of month column how is taking drug below 12 months
- Based on this plot is created



- As we can see more than 1000 people stop using the drug by first dose itself.
- By first dose, second month, first month majority of people stop taking drug
- More than 38% of the people drop off in this three month

Conclusion

- Based on the insights more than 60% of people dropping off the medication within 1 year.
- Majority of people drop off by first dose itself more than 17% of people stop taking drug.