**Research Assistant work under professor Olena**

Project name on Github: Data Analytics in Health Care – RA

Columns in Data Set

'PATIENT\_ID', 'PROVIDER\_IDS', 'PROVIDER\_TYPES', 'CARE\_SETTING\_CATEGORY',

'ADMIT\_DATE', 'DISCHARGE\_DATE', 'DISCHARGE\_DISPOSITION',

'OVERDOSE\_FLAG', 'OD\_DX\_DATE', 'OD\_NOTE\_DATE', 'OD\_TERMS',

'OD\_DRUG\_TERMS', 'OD\_AR\_FLAG', 'ADVERSE\_OPIOID\_EVENT\_FLAG',

'ADVERSE\_OPIOID\_EVENT\_DRUGS', 'ADVERSE\_OPIOID\_EVENT\_PROBLEMS',

'COMMERCIAL\_FLAG', 'OTHERGOV\_FLAG', 'SELFPAY\_FLAG', 'WORKERSCOMP\_FLAG',

'INSTITUTIONALIZED\_FLAG', 'CHARITY\_FLAG', 'MEDICARE\_FLAG',

'MEDICAID\_FLAG', 'LIFE\_FLAG', 'MDWISE\_FLAG', 'NODATA\_FLAG'],

dtype='object')

**Project AIM**

1. Clean the Encounter table of the data
2. Merge encounter table with our final cohort (already built): crucial to have the dates preserved, so we can assess the temporal nature of events
3. Diagnoses table: clean and merge with our final cohort
4. Referrals table: clean and merge with our final cohort

Task: Perform Analyze the patients encounter with the hospitals based on 'CARE\_SETTING\_CATEGORY'

1. In Patient: Patient who are admitted to hospital normally.
2. Emergency: Patient who are admitted to hospital in emergency ward
3. Outgoing: Patient discharged from the hospital

The numbers for the mean and the range of ED visits (E category) and hospitalizations (I category) per patient

I also want to see the distribution of the ED visits and hospitalizations.

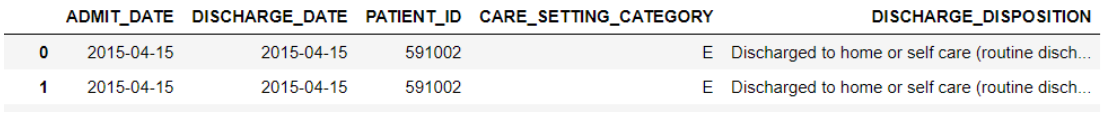
Mini Task

Data Processing

After the discussion conclusion was to use data after the date October 2014

Issues Resolved

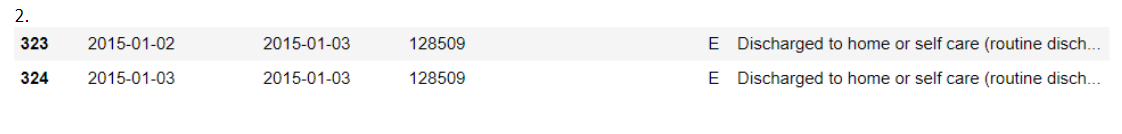
There are two scenarios:

1. Admit date and Discharge date being same for two encounters: 

In this case, we can ignore the one of them, but the more precise decision can be taken if we have time data. It might happen that, the patient was admitted at 9:00am and discharged at 12 pm and again admitted at 3 pm and discharged again at same day.

Solution: remove the duplicates

If there are duplicates with inpatient and emergency patients keep in patient record



In this case, we were considering these both encounters as a single encounter as the admit date were continuous and were ignoring the discharge date, these two are unique encounters as the patient was admitted on 2nd and was discharged on 3 rd and again was admitted on 3rd.

Solution: merge the two rows and final output: 2015-01-02 – 2015-01-03 only one line

Patient with maximum number of Visits in E category

### Patient with Maximum visit in E Category--- 303 times

## Mean number of visits for E Category--3.635

## **Median number of visits for E Category-- 2.0**

Standard Deviation--6.555384393354021

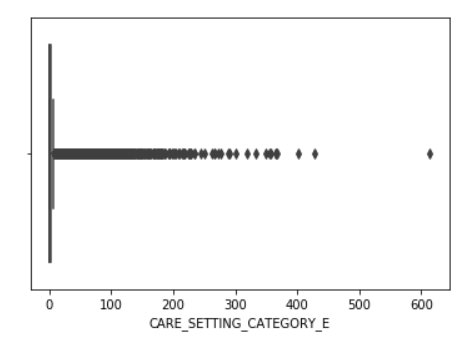
**Distribution in E AND I Category**

E 1504620

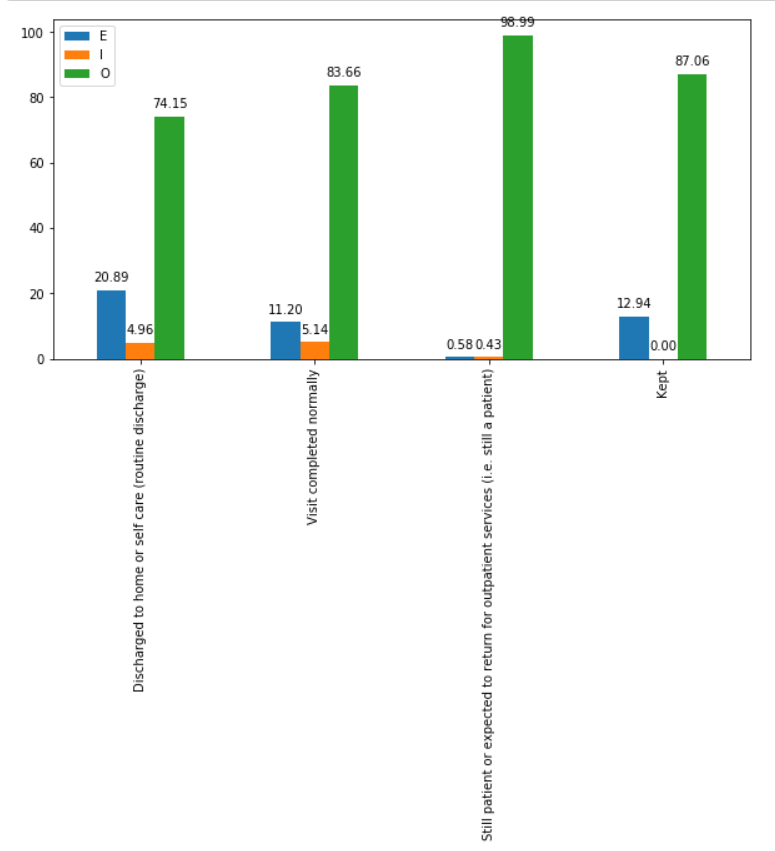
I 541672

Boxplot for Emergency category. This plot was made to check the outliers in the data.

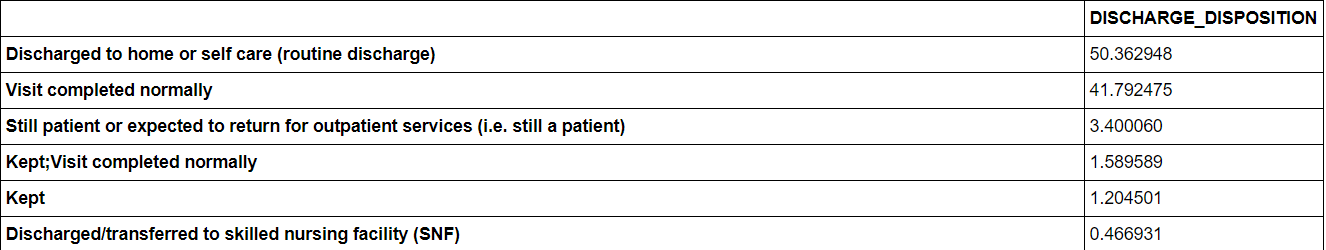
### Patient\_id outliers in E Category--count--6456



Distribution of Discharge Disposition vs Care Setting Category



Distribution of Discharge Disposition

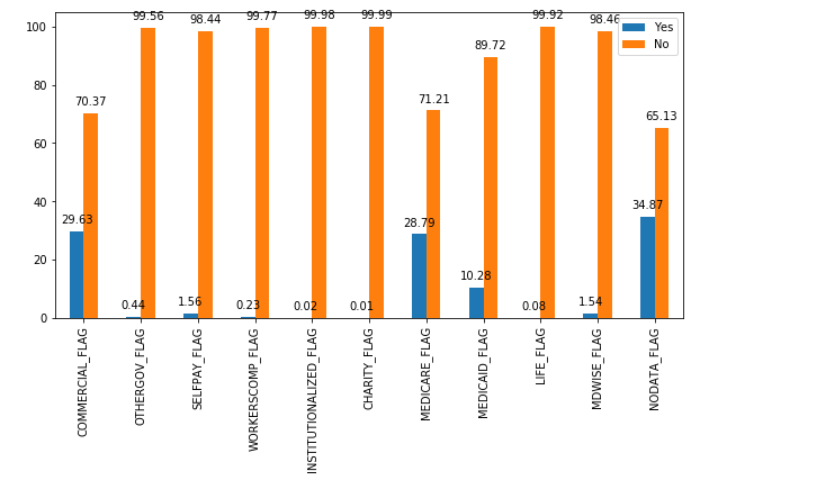


Dealing with Insurance Data

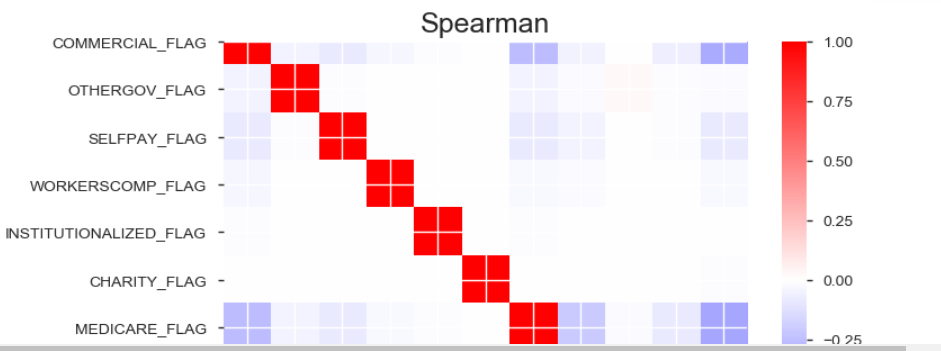
['COMMERCIAL\_FLAG', 'OTHERGOV\_FLAG', 'SELFPAY\_FLAG', 'WORKERSCOMP\_FLAG',

'INSTITUTIONALIZED\_FLAG', 'CHARITY\_FLAG', 'MEDICARE\_FLAG',

'MEDICAID\_FLAG', 'LIFE\_FLAG', 'MDWISE\_FLAG', 'NODATA\_FLAG'])



Spearman Correlation between insurance



## Explaination of ADVERSE\_OPIOID\_EVENT\_DRUGS

Percentage each Adverse opioid event drugs

