

TEAM MEMBERS DETAILS

GROUP NAME: PACHACUTEAM

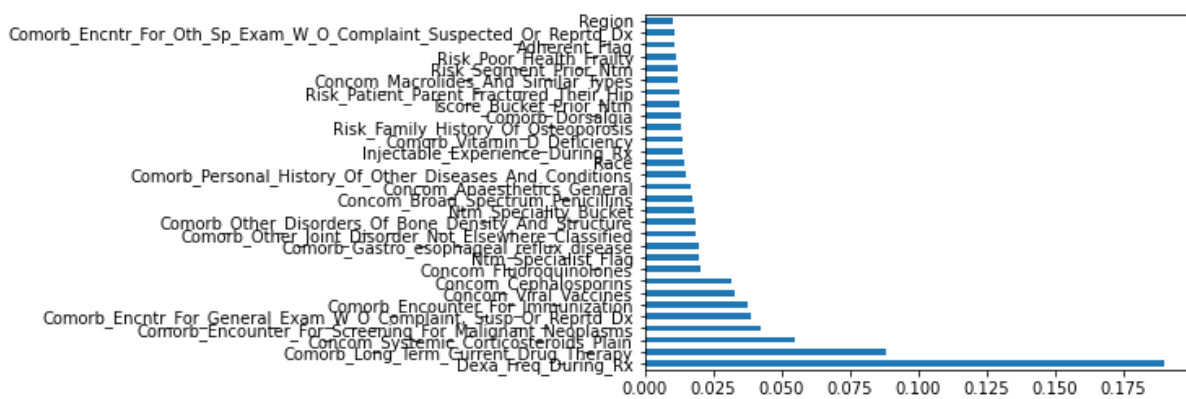
SPECIALIZATION: Data Science

| Name | Email | Country | College/Company |
|------------------|--|---------|-----------------|
| Andersson Romero | andromdez@gmail.com | Perú | Freelancer |

Categorical Data Analysis

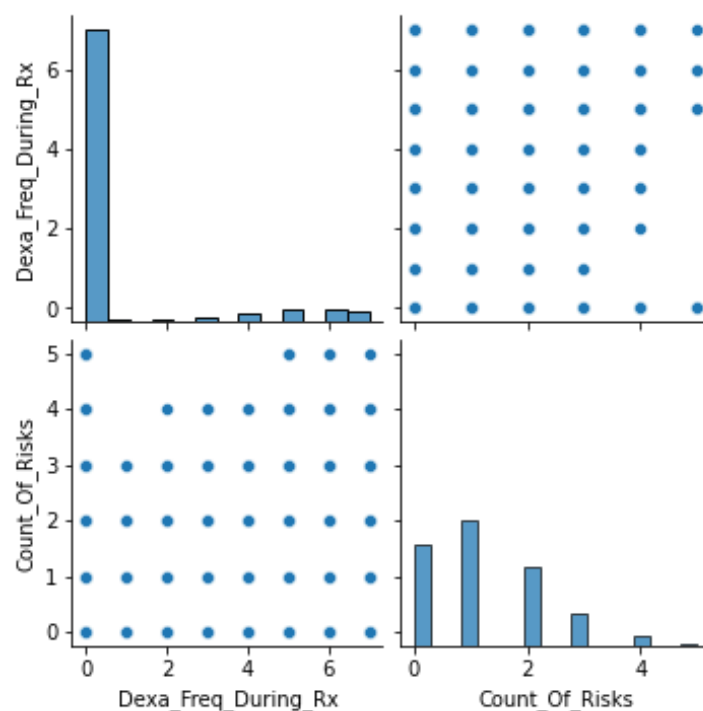
- Feature Relevance

Using xboosting helps us define the most relevant variables. In our case, 'Dexa_Freq_During_Rx' and the first 20 have very high relevance values with respect to 'Persistency_Flag'.



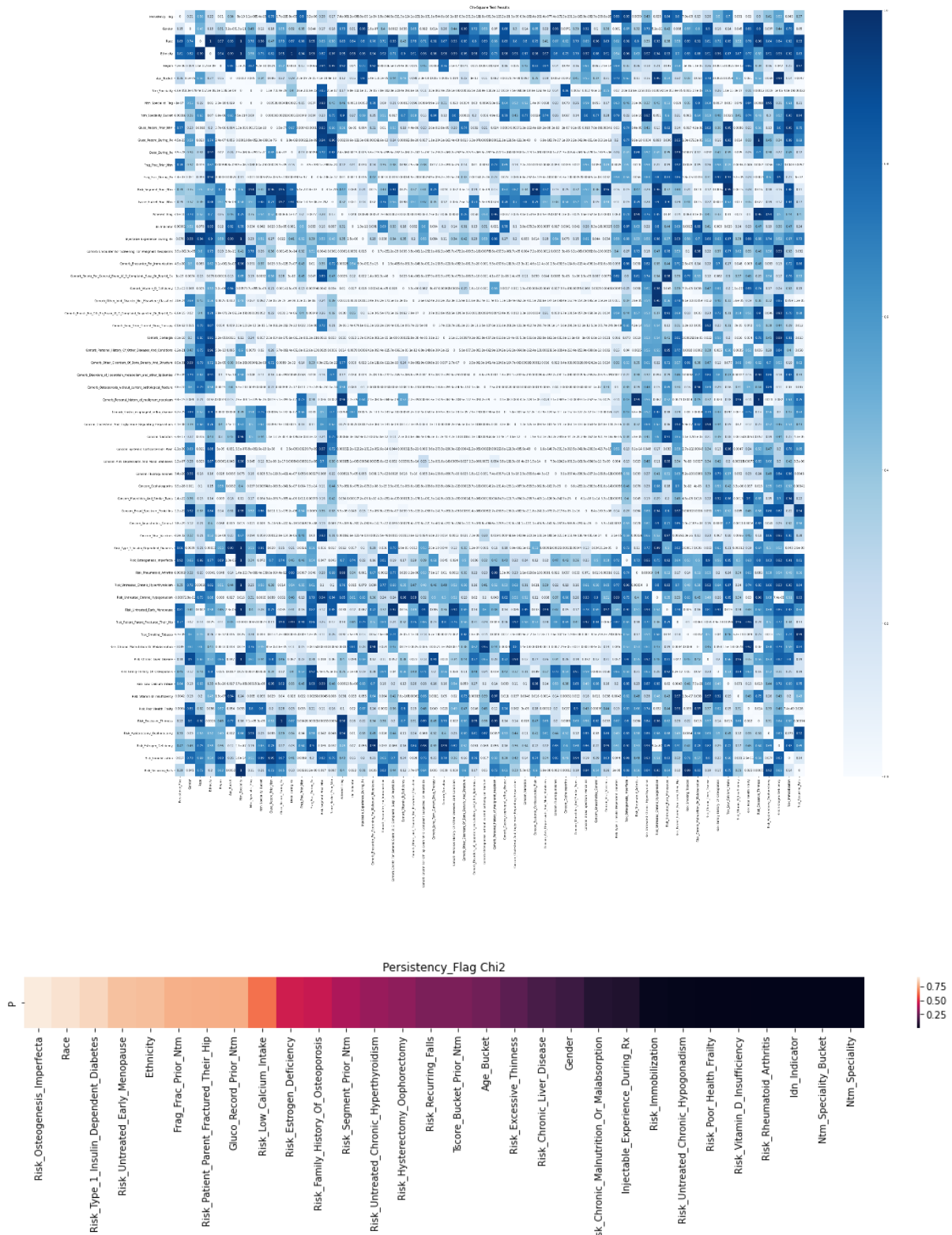
- Pariplot

Create numerical relationships.

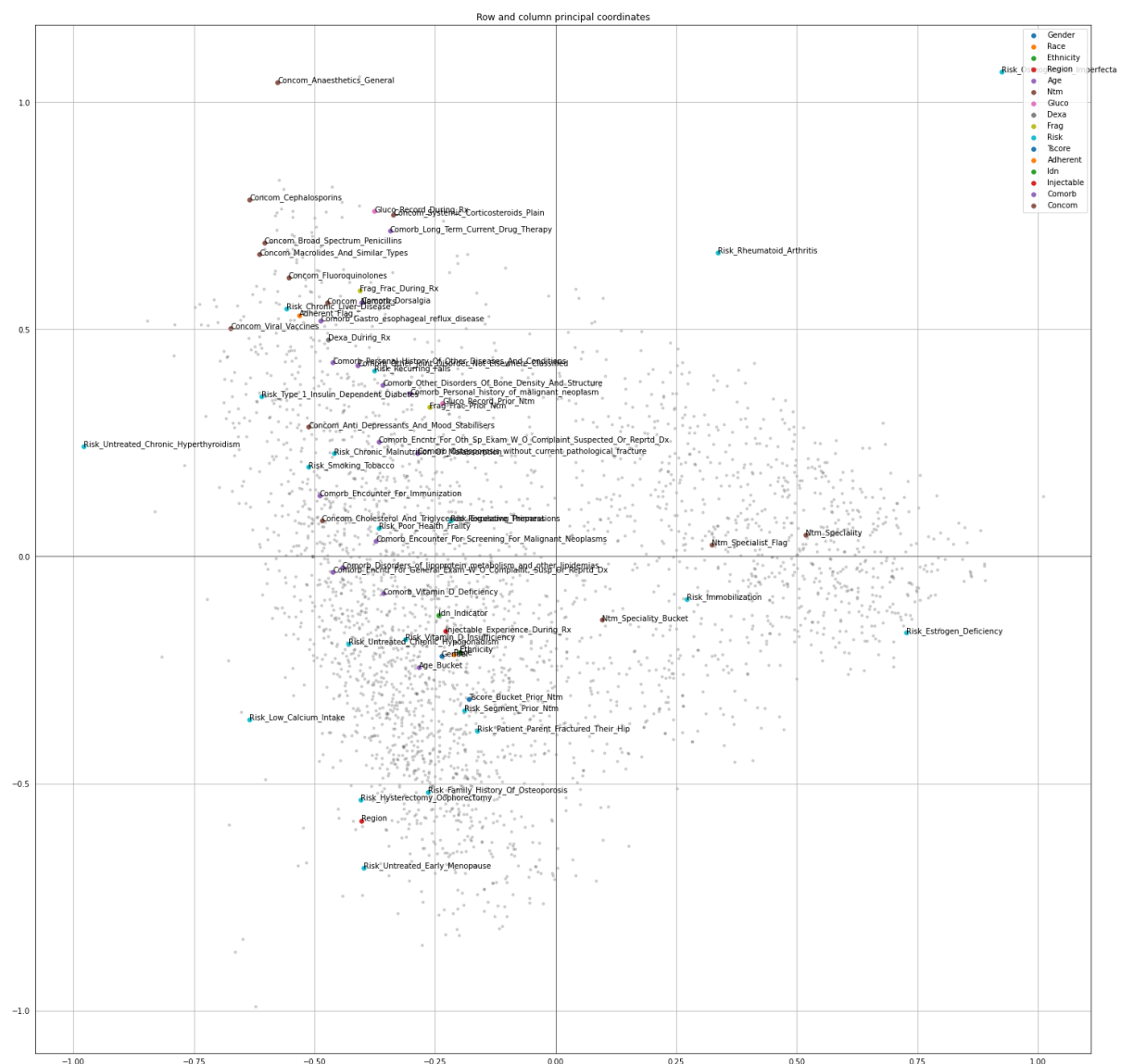


- Chi Square

The Chi-Square Test of Independence determines whether there is an association between categorical variables.

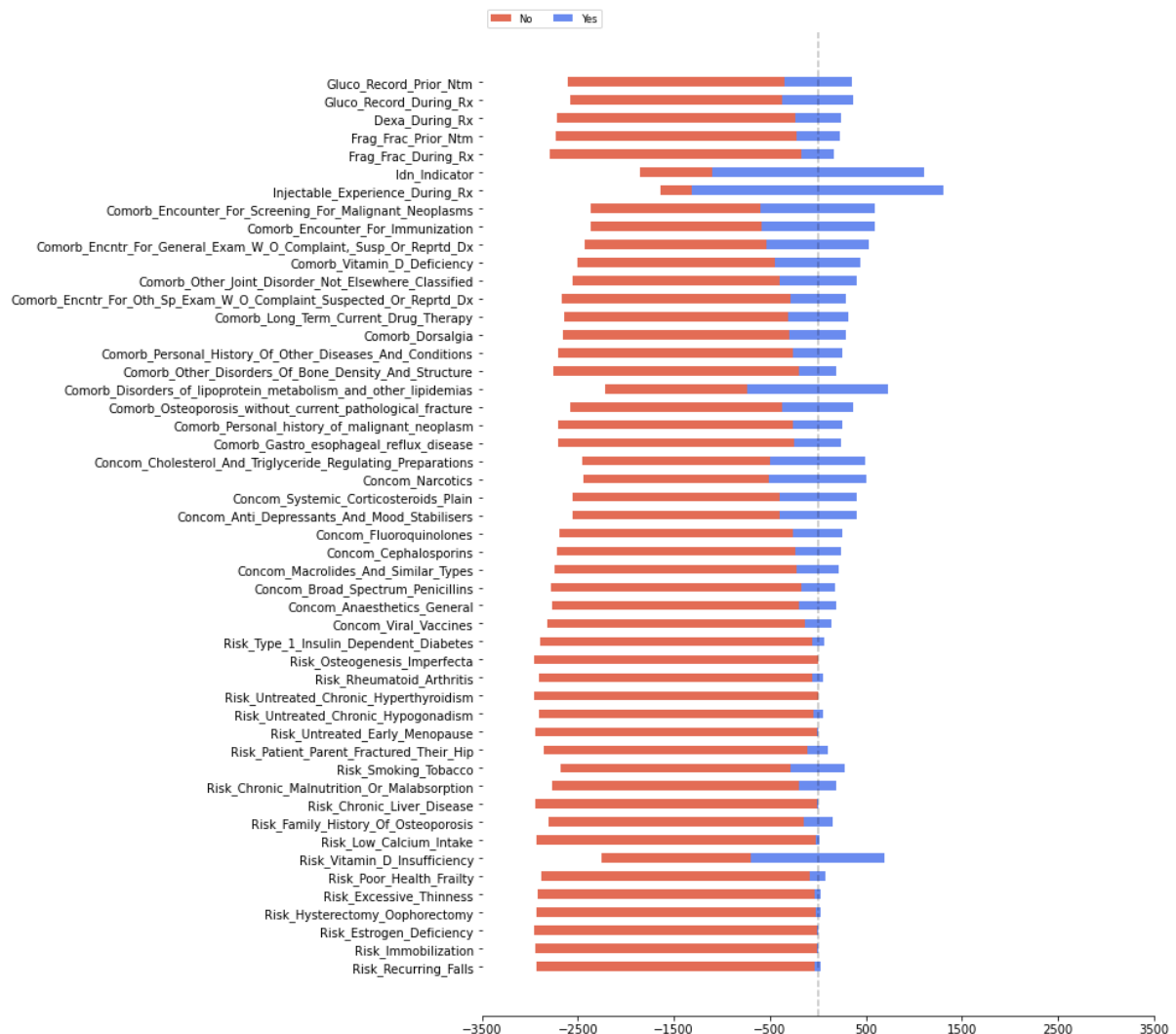


Correspondence analysis, also called reciprocal averaging, is a useful data science visualization technique for finding out and displaying the relationship between categories.



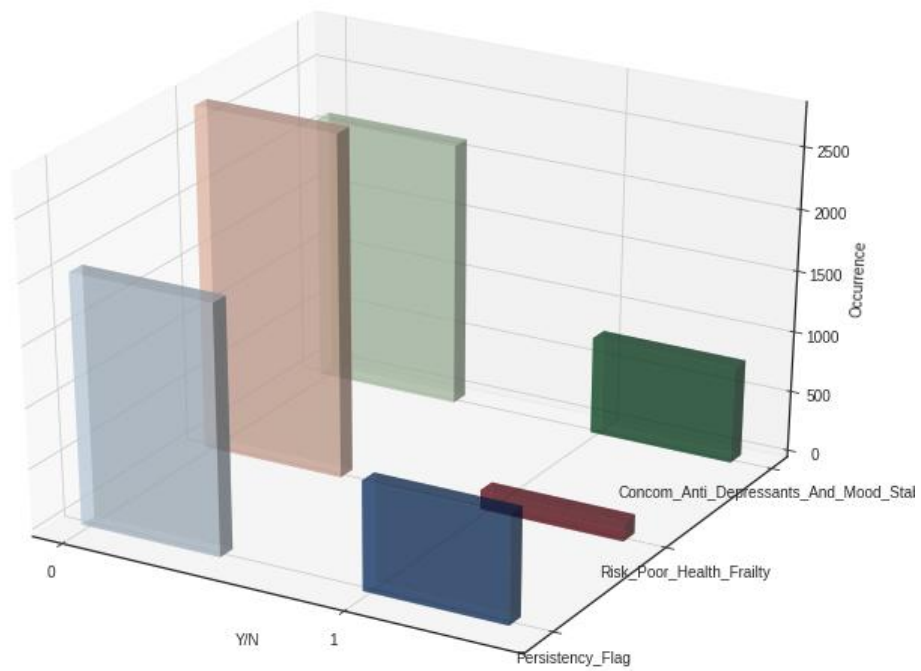
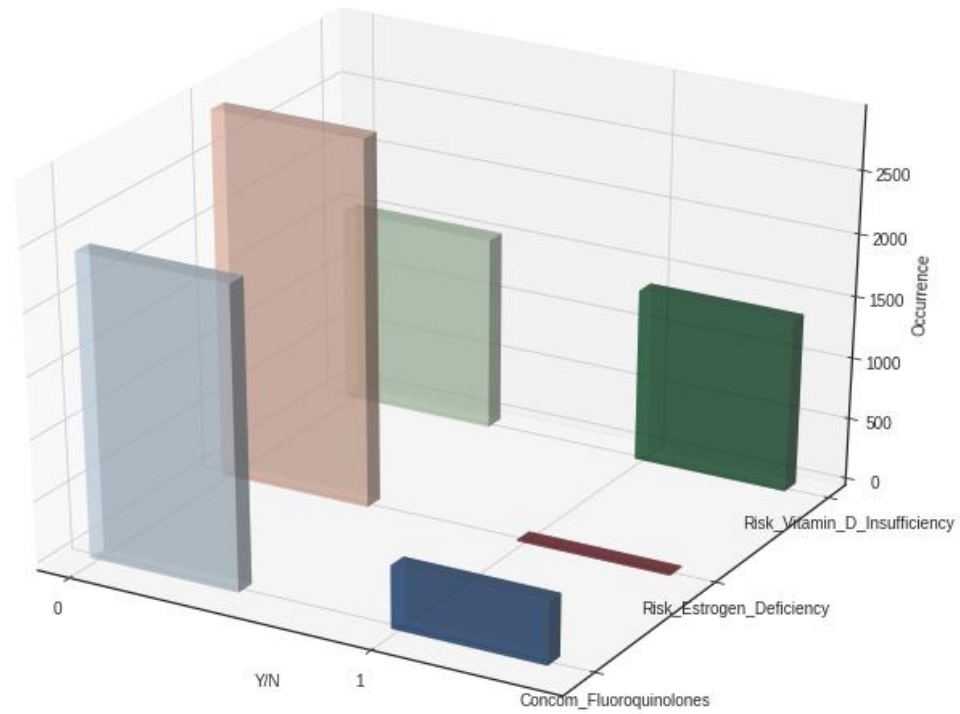
- Tornado chart

Butterfly charts, also called Tornado or Divergent Chart, are essentially bar charts comparing two different metrics at a time.



- Cross tab

Takes a dataframe and at least two variables as input, conducts a crosstabulation of the variables.



DATA INTAKE REPORT

Name: Healthcare project

Report date: 17/04/2022

Internship Batch: LISUM07

Version: 1.0

Data intake by: PACHACUTEAM

Data intake reviewer:<intern who reviewed the report>

Data storage location: GitHub

Tabular data details:

| | |
|-------------------------------------|--------|
| Total number of observations | 3424 |
| Total number of files | 1 |
| Total number of features | 69 |
| Base format of the file | .xlsx |
| Size of the data | 899 KB |

GITHUB LINK: https://github.com/And2300/Healthcare_PersistenceDrug