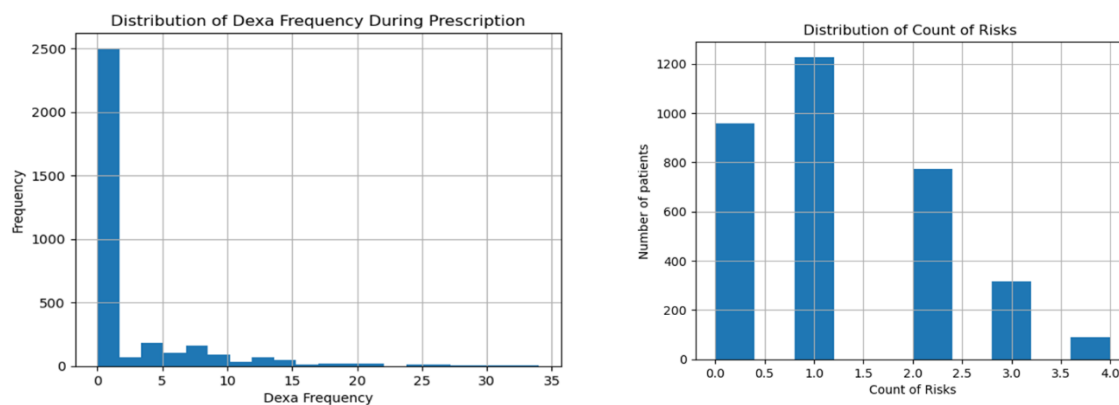# Week 10 Deliverables

Name: Data Minders

Problem Description

After did the data clean (check missing value, check duplicate, and check outlier, rename the features), we created frequency table of all variables, checked quantile for numerical variables, build cross table between persistency_flag VS all features.

The following results are from the histogram

It is necessary to do the log transformation to the feature of dexa_ferquency_during _prescription and log transformation to the feature of count_of_risks. The problem is after the log transformation of the 2 features still including 0. The new variable of log_ dexa_frequency_during_prescription + 1 and log_count_of_risks + 1 is needed.

Based on the cross table we found some features have no impact on the target variable, we have to drop them.

Recommendation for Tech Team

- The following variables have no impact on drug persistency,

  ptid , gender, frag_frac_prior_ntm, gluco_record_prior_ntm,

  risk_patient_parent_fractured_their_hip, risk_segment_prior_ntm,

  risk_untreated_chronic_hyperthyroidism, risk_estrogen_deficiency,

  risk_family_history_of_osteoporosis

- The top 5 features which can impact the drug persistency are,

  total_comorbidity, dexa_freq_during_rx_p1_log,

dexa_freq_during_rx, total_concomitancy,

dexa_freq_during_rx_p1