clean\_claims(df) – takes dataframe and removes outpatient claims associated with emergency department visits (SERVICE\_TYPE). Also removes any duplicate CLAIM\_NUMs and only keeps diagnosis codes 1-7. Saves output to ‘Data/claimsCleanSmall.RDS’. Note this is a rds file but function can be edited to output to csv or sql table. File contains all required columns below.

Columns required: MRN\_ALIAS, MEMBER\_SEX, MEMBER\_AGE, CLAIM\_SEQ, EPISODE\_SEQ, YEAR, SERVICE\_TYPE, PLACE\_OF\_SERVICE\_DESC, diagnosis codes as CODE\_1, CODE\_2, etc., ED\_DISCHARGE\_DX\_DESC, PREVENTABILITY, ED\_NOT\_NEEDED\_PROP, NCLASSIFIED\_ED, PCP\_ID, VENDOR\_PROV\_ID, CLAIM\_TYPE, TOB\_CATEGORY, APPROVED\_DAYS, APPROVED\_AMT, CLAIM\_NUM

icd9\_vector(df) - takes a dataframe of diagnosis codes and determines if it is ICD version 9 and places those codes into a vector. This vector can be used to determine the version of a column of ICD codes.

Columns required: diagnosis codes labeled as “CODE\_1”, “CODE\_2, etc.

find\_chornic\_diagnosis(df) – takes a dataframe with the column “DIAG\_CODE” and creates a new column called CHRONIC\_CONDITION. You can edit this function to label any number of diagnosis codes (with the punctuation removed) as an indicator of a chronic condition

Columns required: one column containing diagnosis codes labeled “DIAG\_CODE”

get\_chronic\_condition(df, splits = 5) - takes claims data and determines if a member has been diagnosed with a chronic condition in the past. splits controls the number of times to split the data. The default number of splits is 5. The is useful for analyzing larger datasets on a single computer. Saves output the ‘Data/claims\_chronic\_condition.csv’ in working directory. File contains MRN\_ALIAS, and a column for the chronic conditions in find\_chronic\_diagnosis() function. Import and join on claims data.

Columns required: diagnosis codes labeled as ‘CODE\_1”, “CODE\_2, etc. MRN\_ALIAS, YEAR, CLAIM\_NUM, CLAIM\_SEQ

impute\_ccs(df, splits = 5) - takes a dataframe and replaces the diagnosis codes with ccs category code. Splits controls the number of time to split the data. The default number of splits is 5. The is useful for analyzing larger datasets on a single computer. Saves output the ‘Data/claims\_claim\_ccs\_category.csv’ in working directory. File contains CLAIM\_NUM, ccs code as “CODE\_1”, “CODE\_2”, etc.

Columns required: diagnosis codes labeled as ‘CODE\_1”, “CODE\_2, etc. MRN\_ALIAS, YEAR, CLAIM\_NUM, CLAIM\_SEQ

get\_upcoming\_visits(df, save = TRUE) - takes a dataframe and creates a column if a member goes to the emergency department in the next 3 visits. This is used as the dependent variable in the model. You can choose to save to file as ‘Data/claims\_ed\_next\_3 or set save to FALSE and save to an environment variable

Columns required: CLAIM\_NUM, MRN\_ALIAS, SERVICE\_TYPE, YEAR, CLAIM\_SEQ

get\_age\_group – takes a dataframe and turns MEMBER\_AGE into a age ranger represented by common grouping names (Baby, PreSchool, Childe, etc.)

Columns required: MEMBER\_AGE

import\_lib – works the same as the base function ‘library()’ if the package is already downloaded. If the package is not downloaded, it will download the package and imports it to the environment.