In [3]:

```
root_path = 'C:/Users/rodri/Dropbox/Malawi/SIEG2021 (1)/2023 July'
path_feb23 = 'C:/Users/rodri/Dropbox/Malawi/SIEG2021 (1)/2023 Feb/Data/Clean data/Ph

import warnings
warnings.filterwarnings('ignore')
import numpy as np
import pandas as pd
import os


# Set the working directory
os.chdir(root_path+'/Data/Clean data/Phase 1 - Roster')

## Display  set-up
pd.options.display.float_format = '{:,.2f}'.format
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)


percentiles = [0.05, 0.1, .25, .5, .75, 0.8, 0.9, 0.95, 0.99]

# ========================================================================
# Import data: Data from the field and conversion rates (ISA-LSMS price conversions)
# ========================================================================

roster = pd.read_stata(root_path+"/Data/Raw data/[1]-SIEG-Household Listing- July 20

roster_raw = roster

## Look at duplicates:
dupl_roster = pd.value_counts(roster['hhid'])
print('================================================================')
print('These households are duplicate:')
print('0 households')

print('================================================================')
print('Households interviewed in February 2023')
print(pd.value_counts(roster['hh_inter']))
print("20 'new' households")

print('================================================================')
print('Old households comments regarding interview february 2023')
print(roster.loc[(roster['oldroster_notes']!='None') & (roster['oldroster_notes']!='

print('================================================================')
print('household with new nembers:')
print(pd.value_counts(roster['newmem']))
print('')

print('================================================================')
print('For households already interviewed we recover feb 23 data and add the new mem
print('================================================================')
print('note 1: issues on hh size and other variables are still present in this round
print('Note 2: Suprisingly, there are 20 new households since february but also ther
# new households
new_hh = roster.loc[roster['hh_inter']=='No']

# old households
rosterold =roster.loc[roster['hh_inter']=='Yes']
rosterfeb = pd.read_csv(path_feb23+'/roster_feb23.csv')

rosterold_new = rosterold.loc[rosterold['num_people']>0]
```

```
rosterold_new['new_head']= 0

print('================================================================')
print('Updating household size')
rosterold_new['hh_size_new'] = rosterold_new['npeople'].astype(float)+rosterold_new[



for i in range(1, int(max(rosterold_new['num_people']))):
    print(i)
    print(sum((rosterold_new['rel_hhhead_'+str(i)]==1)))
print('there are 7 new household heads in households that were already interviewed i

# only new heads in 1st iteration. easy to change
rosterold_new['new_head'] = 1*(rosterold_new['rel_hhhead_1']==1)

rosterold_new.loc[rosterold_new['rel_hhhead_2']==1, 'new_head']=1

roster.rename(columns={'background_gpslatitude':'gps_lat_3','background_gpslongitude

roster= roster[['hhid','gps_lat_3','gps_long_3','invillage_feb23','interviewed_feb23
roster['wave'] ='July 2023'

# update roster for households with new size, and new head characteristics (all new
roster = roster.merge(rosterold_new[['hhid','new_head','hh_size_new','age_years_1','

#update size
roster['hh_size'] = np.where(roster['hh_size_new']>0, roster['hh_size_new'],roster['

# update household head age
# roster['head_age'] +=1 done in february. No need to update

# update head characteristics
roster.loc[(roster['new_head']==1) & (rosterold_new['rel_hhhead_1']==1),['head_age',
roster.loc[(roster['new_head']==1) & (rosterold_new['rel_hhhead_2']==1),['head_age',


# now let's do the roster for the new households:
new_hh.rename(columns={'hh_vill':'invillage_feb23','hh_inter':'interviewed_feb23','i
new_hh.replace(['6 Years and Over','Less 6 Years',],0, inplace=True)

max_head = int(max(new_hh['hh_size']))

for i in range(1,max_head):
    new_hh['age_years_'+str(i)] = pd.to_numeric(new_hh['age_years_'+str(i)])

pd.value_counts(new_hh['rel_hhhead_1'])

new_hh[['head_age','head_name','head_gender','head_marital', 'head_nickname']] = new


for i in  range(2,max_head):
    #print(i)
    new_hh[['head_age']] = np.where(new_hh['rel_hhhead_'+str(i)] == 1, new_hh['age_y
    new_hh[['head_nickname']] = np.where(new_hh['rel_hhhead_'+str(i)] == 1, new_hh['
    new_hh[['head_name']] = np.where(new_hh['rel_hhhead_'+str(i)] == 1, new_hh['mem_
    new_hh[['head_gender']] = np.where(new_hh['rel_hhhead_'+str(i)] == 1, new_hh['ge
    new_hh[['head_marital']] = np.where(new_hh['rel_hhhead_'+str(i)] == 1, new_hh['m

# new to update these characteristics
new_hh[['chief_related']] = 1*(new_hh[['chief_related']]=='Yes')
new_hh[['village_born']] = 1*(new_hh[['village_born']]=='Yes')
new_hh['elder_yes'] =  1*(new_hh[['elder_yes']]=='Yes')
new_hh['elders_related'] = 1*(new_hh[['elders_related']]=='Yes')
```

```
#new_hh = new_hh[['hh_size', 'intervieweename','head_age','head_name',  'head_religi


roster.loc[roster['interviewed_feb23']=='No', ['hh_size', 'intervieweename','head_ag



## Gender, marital-status, religion, and educ dummies
pd.value_counts(roster['head_gender'], normalize=True)
roster['head_female'] = 1*(roster['head_gender']==2)




## Marital status head
print('================================================')
print('Summary head marital status')
print(pd.value_counts(roster['head_marital'], normalize=True))
roster['head_married_mono'] = 1*(roster['head_marital']==1)
roster['head_married_poly'] = 1*(roster['head_marital']==2)
roster['head_divorced'] = 1*(roster['head_marital']==4)
roster['head_widowed'] = 1*(roster['head_marital']==5)
roster['head_separated'] = 1*(roster['head_marital']==3)
roster['head_nevermarried'] = 1*(roster['head_marital']==6)


# head characteristics
print('================================================')
print('Summary household size and household head characteristics')
print(roster[['hh_size','head_age','head_female','head_marital','head_divorced','hea



# head education
print('================================================')
print('Summary household head education')
print(pd.value_counts(roster['head_educ'], normalize=True))
roster['head_noeduc'] = 1*(roster['head_educ']=='No education')
roster['head_belowprimary4'] = 1*((roster['head_educ']=='Primary Standard 1')|(roste
roster['head_belowprimary7'] = 1*((roster['head_educ']=='Primary Standard 4')|(roste
roster['head_belowsecond3'] = 1*((roster['head_educ']=='Secondary form 1')|(roster['
roster['head_secondary'] = 1*((roster['head_educ']=='Secondary form 3')|(roster['hea

roster['head_educ_countin'] = 0
roster.loc[ roster['head_educ']=='Primary Standard 1' , 'head_educ_countin'] = 1
roster.loc[ roster['head_educ']=='Primary Standard 2' , 'head_educ_countin'] = 2
roster.loc[ roster['head_educ']=='Primary Standard 3' , 'head_educ_countin'] = 3
roster.loc[ roster['head_educ']=='Primary Standard 4' , 'head_educ_countin'] = 4
roster.loc[ roster['head_educ']=='Primary Standard 5' , 'head_educ_countin'] = 5
roster.loc[ roster['head_educ']=='Primary Standard 6' , 'head_educ_countin'] = 6
roster.loc[ roster['head_educ']=='Primary Standard 7' , 'head_educ_countin'] = 7
roster.loc[ roster['head_educ']=='Primary Standard 8' , 'head_educ_countin'] = 8

roster.loc[ roster['head_educ']=='Secondary form 1' , 'head_educ_countin'] = 9
roster.loc[ roster['head_educ']=='Secondary form 2' , 'head_educ_countin'] = 10
roster.loc[ roster['head_educ']=='Secondary form 3' , 'head_educ_countin'] = 11
roster.loc[ roster['head_educ']=='Secondary form 4' , 'head_educ_countin'] = 12
roster.loc[ roster['head_educ']=='Training college year 2', 'head_educ_countin'] = 1
roster.loc[ roster['head_educ']=='University 4', 'head_educ_countin'] = 12
```

```
# head/family religion, ethnicity

roster[['head_christian']] = 1*(roster[['head_religion']]=='Christian')


print('================================================')
print('Summary religion, village background, chiefs and elders')
print(roster[['head_christian','village_born','village_years','chief_related','chief

print(roster[['village','subvillage','head_religion','ethnic','mlanguage','chief_rel

# other variables


roster = roster[['hhid','wave','invillage_feb23','interviewed_feb23', 'intervieweena
 'head_educ', 'head_religion', 'head_female', 'head_married_mono','head_married_poly
 "spouse_educ", "ethnic",  "mlanguage",'village_born','village_years','chief_related
 'head_belowprimary4', 'head_belowprimary7', 'head_belowsecond3', 'head_secondary',



roster.to_csv(root_path+"/Data/Clean data/Phase 1 - Roster/roster_july23.csv",index=


print('================================================')
print('final dataset saved in clean data/phase 1/roster_july23.csv')
print('================================================')
print("Containts the following variables: 'hhid','wave','invillage_feb23','interview
```

```
================================================================
These households are duplicate:
0 households
================================================================
Households interviewed in February 2023
Yes    264
No      20
Name: hh_inter, dtype: int64
20 'new' households
================================================================
Old households comments regarding interview february 2023
     hhid                                    oldroster_notes
2    1306  Both are living in the household 1306 except H...
11   1416       All household members live in the household.
12   1417  Both live in the household( 1417) however, the...
17   1409  All household members still live in the househ...
18   1412  Both live in the household except Alinafe Yaha...
31   1300              They all live in the household (1300)
32   1302                  Both live in the household (1302)
39   1020        They all still live in the household (1020)
40   1119        They all still live in the household (1119)
48   1021  All do live in the household (1021) except the...
50   1522  All household members still live in household ...
54   1503  Hawa Wasili moved out with her son, Abdullah M...
60   1120             All still live in the household( 1120)
63   2016             All still live in the household (2016)
71   1022             All still live in the household (1022)
72   1506        They all still live in the household (1506)
81   1218        They all still live in the household (1218)
83   1521        They all still live in the household (1521)
85   1221        They all still live in the household( 1221)
86   1321        They all still live in the household (1321)
102  1008        They all still live in the household (1008)
103  1009        They all still live in the household (1009)
113  1010        They all still live in the household (1010)
114  1108        They all still live in the household (1108)
120  1107        They all still live in the household (1107)
122  1109        They all still live in the household (1109)
```

```
133  1018        They all still live in the household (1018)
134  1026        They all still live in the household (1026)
135  1030        They all still live in the household (1030)
144  2012         The wife doesnt know the nature of the Job
148  1225        They all still live in the household (1225)
167  1523        They all still live in the household (1523)
169  1524        They all still live in the household (1524)
173  1032          They all still live in the household 1032
174  1038        They all still live in the household (1038)
175  1232        They all still live in the household (1232)
190  1336        They all still live in the household (1336)
193  1337        They all still live in the household (1337)
194  1437        They all still live in the household( 1437)
199  1233                                                 2
209  1033  The spouse and the son still live in the house...
210  1138        They all still live in the household (1138)
211  1235        They all still live in the household (1235)
212  1236        They all still live in the household (1236)
213  1440        They all still live in the household (1440)
214  1448        They all still live in the household (1448)
218  1342                                    They Divorced
221  1449  Agness went to her home village to stay there,...
235  1042        They all still live in the household (1042)
244  1041        They all still live in the household (1041)
246  1043        They all still live in the household (1043)
248  1144        They all still live in the household (1144)
250  1239        They all still live in the household (1239)
252  1443        They all still live in the household (1443)
253  1549        They all still live in the household (1549)
259  1046                                              none
271  1241        They all still live in the household (1241)
276  1544        They all still live in the household (1544)
282  1050        They all still live in the household (1050)
283  1223        They all still live in the household (1223)
================================================================
household with new nembers:
No     210
Yes     54
Name: newmem, dtype: int64

================================================================
For households already interviewed we recover feb 23 data and add the new members
================================================================
note 1: issues on hh size and other variables are still present in this round of the
roster
Note 2: Suprisingly, there are 20 new households since february but also there are a
round 30 hhs that were in february round and not now... could they be the same but a
change in the head?
================================================================
Updating household size
1
6
2
1
3
0
there are 7 new household heads in households that were already interviewed in febru
ary!
============================================
Summary head marital status
1.00   0.63
4.00   0.17
2.00   0.09
6.00   0.06
3.00   0.03
5.00   0.02
Name: head_marital, dtype: float64
============================================
Summary household size and household head characteristics
```

```
       hh_size  head_age  head_female  head_marital  head_divorced  \
count   283.00    283.00       284.00        283.00         284.00
mean      4.84     43.64         0.31          2.02           0.17
std       1.99     17.69         0.46          1.56           0.38
min       1.00     11.00         0.00          1.00           0.00
25%       3.00     28.00         0.00          1.00           0.00
50%       4.00     41.00         0.00          1.00           0.00
75%       6.00     53.50         1.00          3.00           0.00
max      11.00     91.00         1.00          6.00           1.00

       head_widowed  head_separated
count        284.00          284.00
mean           0.02            0.03
std            0.14            0.17
min            0.00            0.00
25%            0.00            0.00
50%            0.00            0.00
75%            0.00            0.00
max            1.00            1.00
================================================
Summary household head education
Primary Standard 8        0.15
Primary Standard 7        0.14
Primary Standard 5        0.12
No education              0.10
Primary Standard 4        0.09
Primary Standard 3        0.08
Primary Standard 6        0.07
Primary Standard 1        0.06
Secondary form 4          0.05
Primary Standard 2        0.05
Secondary form 2          0.05
Secondary form 1          0.03
Secondary form 3          0.02
Training college year 3   0.00
Training college year 2   0.00
Name: head_educ, dtype: float64
================================================
Summary religion, village background, chiefs and elders
       head_christian  village_born  village_years  chief_related  elder_yes  \
count          284.00        283.00         283.00         283.00     283.00
mean             0.16          0.70          14.18           0.59       0.13
std              0.37          0.46          14.20           0.49       0.34
min              0.00          0.00           0.00           0.00       0.00
25%              0.00          0.00           2.00           0.00       0.00
50%              0.00          1.00          10.00           1.00       0.00
75%              0.00          1.00          21.00           1.00       0.00
max              1.00          1.00          56.00           1.00       1.00

       elders_related
count          283.00
mean             0.45
std              0.50
min              0.00
25%              0.00
50%              0.00
75%              1.00
max              1.00
                                village subvillage head_religion ethnic  \
count                                247        114           283    283
unique                                 5          7             3      7
top     Geradi (different sub-villages).     Geradi        Muslim    Yao
freq                                 114         32           235    224

       mlanguage chief_relation      elders_relation
count        283            166                  126
unique         5             15                   16
top          Yao    Grandparent  Maternal aunt/uncle
freq         180             41                   26
```

```
================================================
final dataset saved in clean data/phase 1/roster_july23.csv
================================================
Containts the following variables: 'hhid','wave','invillage_feb23','interviewed_feb2
3', 'intervieweename', 'head_name', 'village', 'subvillage', 'key_landmark','mosque_
church','hh_size', 'hh_phone', 'head_gender', 'head_marital', 'head_age', 'head_nick
name', 'head_educ', 'head_religion', 'head_female', 'head_married_mono','head_marrie
d_poly','head_nevermarried', 'head_divorced', 'head_widowed', 'head_separated', 'hea
d_christian', 'head_noeduc', 'spouse_educ', 'ethnic',  'mlanguage','village_born','v
illage_years','chief_related','chief_relation','elder_yes','elders_related','elders_
relation','head_belowprimary4', 'head_belowprimary7', 'head_belowsecond3', 'head_sec
ondary', 'head_educ_countin',  'gps_lat_3', 'gps_long_3'
```

In [ ]: