HF_Locations_VF

October 2, 2020

[1]:
```
# import necessaries modules:
import numpy as np
import datetime
import pandas as pd
import gc
```

0.1 Step 1: Reading data related to tblHFs and tblLocation

0.1.1 Step 1.1 : Reading data related to tblHFs

[2]:
```
# read the csv file and selecting the necessary columns
filename = 'openIMIS csv/health_facilities2020.csv'
cols =␣
 ↪['HfID','HFCode','HFLevel','HFSublevel','LocationId','HFCareType','HfUUID']
df_HF_raw = pd.read_csv(filename,low_memory=False,usecols=cols)
df_HF_raw = df_HF_raw.iloc[:-2,:]

#rename columns
df_HF_raw.rename(columns = {'HfID': 'HFID','HfUUID': 'HFUUID'}, inplace = True)

memStats_HF = (df_HF_raw.memory_usage()/1024/1024).sum()
shape_HF = df_HF_raw.shape
```

0.1.2 Step 1.2 : Reading data related to tblLocations

[3]:
```
filename = 'openIMIS csv/locations2020.csv'
cols = ['LocationId', 'LocationCode','LocationName','LocationUUID']
df_location_raw = pd.read_csv(filename,low_memory=False,usecols=cols)
df_location_raw = df_location_raw.iloc[:-2,:]

df_location_raw['LocationId'] = df_location_raw['LocationId'].astype(int)

memStats_Locs = (df_location_raw.memory_usage()/1024/1024).sum()
shape_Locs = df_location_raw.shape
```

0.2 Step 2 : Concatenate the dataframes related to tblHFs and tblLocations

[4]:
```
df_HF_locations = pd.merge(df_HF_raw,df_location_raw,on='LocationId')

df_HF_locations.rename(columns = {'LocationId': 'HFLocationId',
                                  'LocationName':'HFLocationName',
                                  'LocationType': 'HFLocationType',
                                  'LocationUUID': 'HFLocationUUID',
                                  'HfID': 'HFID'
                                 }, inplace = True)

memStats_concat = (df_HF_locations.memory_usage()/1024/1024).sum()
shape_concat = df_HF_locations.shape
```

[5]:
```
# save data to files
df_HF_locations.to_pickle('openIMIS csv/HF_Locations2020_sel.pkl')
#df_HF_locations.to_csv('openIMIS csv/HF_Locations2020_sel.csv')

df_HF_sel = df_HF_raw[['HFID','HFUUID']]
df_HF_sel.to_pickle('openIMIS csv/HF2020_sel.pkl')
```

0.3 Summary:

[6]:
```
print(f'''
- tblHFs has : {shape_HF[0]} rows; {shape_HF[1]} columns; \
{round(memStats_HF,2)} MB memory consumption;
- tblLocations has : {shape_Locs[0]} rows; {shape_Locs[1]} columns; \
{round(memStats_Locs,2)} MB memory consumption;
- Concatenation of tblHFs and tblLocations has : {shape_concat[0]} rows;␣
 ↪{shape_concat[1]} columns; \
{round(memStats_concat,2)} MB memory consumption;
''')
```

```
- tblHFs has : 780 rows; 7 columns; 0.04 MB memory consumption;
- tblLocations has : 10350 rows; 4 columns; 0.32 MB memory consumption;
- Concatenation of tblHFs and tblLocations has : 780 rows; 10 columns; 0.07 MB
memory consumption;
```