

ConcatenateData_VF

October 2, 2020

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

0.1 Step 1: Reading the data related to tblClaimItems

```
[2]: # pkl file related to tblClaimItems and all associated tables:  
df_claimitems = pd.read_pickle('openIMIS csv/ClaimItems_Items2020_sel.pkl')  
  
memStats_claimitems = (df_claimitems.memory_usage()/1024/1024).sum()  
shape_claimitems = df_claimitems.shape
```

0.2 Step 2: Reading data related to tblClaimServices

```
[3]: # pkl file related to tblClaimServices and all associated tables:  
df_claimservices = pd.read_pickle('openIMIS csv/ClaimServices_Services2020_sel.  
→pkl')  
  
memStats_claimservices = (df_claimservices.memory_usage()/1024/1024).sum()  
shape_claimservices = df_claimservices.shape
```

0.3 Step 3: Concatenate the previous dataframes

```
[4]: df_claim_final= df_claimitems.append(df_claimservices, ignore_index=True)  
  
memStats_final = (df_claim_final.memory_usage()/1024/1024).sum()  
shape_final = df_claim_final.shape
```

```
[5]: del [[df_claimitems,df_claimservices]]  
df_claimitems = pd.DataFrame()  
df_claimservices = pd.DataFrame()  
gc.collect()
```

[5]: 20

```
[6]: df_claim_final.to_feather('openIMIS csv/AllData.ftr')
      #df_claim_final.to_pickle('openIMIS csv/AllData.pkl')
```

0.4 Summary:

```
[7]: print(f''Summary of the concatenation process:
- tblClaimItems has : {shape_claimitems[0]} rows; {shape_claimitems[1]} columns;
  →\
{round(memStats_claimitems,2)} MB memory consumption;
- tblClaimService has : {shape_claimservices[0]} rows; {shape_claimservices[1]}
  →columns;\
{round(memStats_claimservices,2)} MB memory consumption;
- Concatenation of tblClaimItems and tblClaimServices has : {shape_final[0]}
  →rows;\
{shape_final[1]} columns; {round(memStats_final,2)} MB memory consumption;
''')
```

Summary of the concatenation process:

```
- tblClaimItems has : 12371992 rows; 82 columns;7834.44 MB memory consumption;
- tblClaimService has : 16600273 rows; 82 columns;10511.95 MB memory
consumption;
- Concatenation of tblClaimItems and tblClaimServices has : 28972265 rows;82
columns; 18125.35 MB memory consumption;
```