Meeting #6

10/21/21

Omar Luna

Deliverables

- Continue parsing livedata.json file

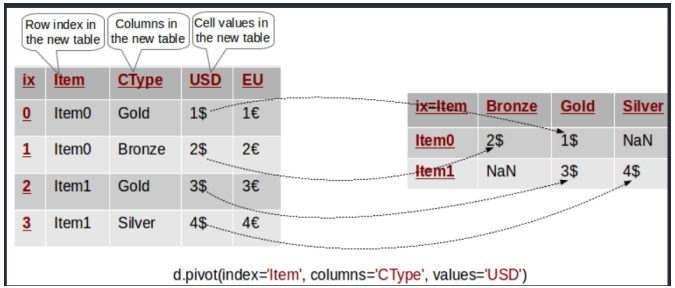
Methodology and Learnings

- DataFrame.pivot
  - Reshapes DataFrame
- DataFrame.stack
  - Takes inner-most column and makes inner-most row
- Consolidated parsing function
  - Too inefficient
  - Rethink parsing method / reduce data
- Used loc to add new rows to DataFrame

Learn

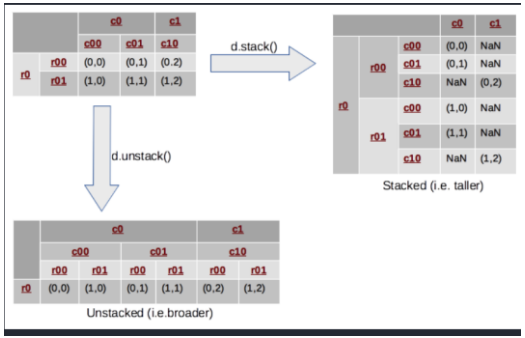
- pivot
- stack

Results



ix	Item	CType	USD	EU
0	Item0	Gold	1$	1€
1	Item0	Bronze	2$	2€
2	Item1	Gold	3$	3€
3	Item1	Silver	4$	4€

ix=Item	Bronze	Gold	Silver
Item0	2$	1$	NaN
Item1	NaN	3$	4$




```
def data_parsing(s):
  if "gidx" in s:
    if "gd" in s:
      data = [s["ts"], s["s"], s["gidx"], s["gd"], s["eye"] ]
      return 1, data
    elif "pc" in s:
      data = [s["ts"], s["s"], s["gidx"], s["pc"], s["eye"] ]
      return 2, data
    elif "l" in s:
      data = [s["ts"], s["s"], s["gidx"], s["l"], s["gp"] ]
      return 3, data
    elif "pd" in s:
      data = [s["ts"], s["s"], s["gidx"], s["pd"], s["eye"] ]
      return 4, data
    elif "gp3" in s:
      data = [s["ts"], s["s"], s["gidx"], s["gp3"]]
      return 5, data
  elif "ac" in s:
    data = [s["ts"], s["s"], s["ac"]]
    return 6, data
    # parsing ac function
  elif "gy" in s:
    data = [s["ts"], s["s"], s["gy"]]
    return 7, data
    # parsing gy function
  elif "vts" in s:
    data = [s["ts"], s["s"], s["vts"]]
    return 8, data
```

```
import json as js
import pandas as pd
import numpy as np


df_gd_gidx = pd.DataFrame(columns= ['ts','s', 'gidx', 'gd' ,'eye'])
df_pc_gidx = pd.DataFrame(columns= ['ts','s', 'gidx', 'pc' ,'eye'])
df_l_gidx = pd.DataFrame(columns= ['ts','s', 'gidx', 'l', 'gp'])
df_pd_gidx = pd.DataFrame(columns= ['ts','s', 'gidx', 'pd' ,'eye'])
df_gp3_gidx = pd.DataFrame(columns= ['ts','s', 'gidx', 'gp3'])

# dataframe for ac data
df_ac = pd.DataFrame(columns= ['ts','s', 'ac'])
# dataframe for ac data
df_gy = pd.DataFrame(columns= ['ts','s', 'gy'])
# dataframe for ac data
df_vts = pd.DataFrame(columns= ['ts','s', 'vts'])
# dataframe for ac data
df_pts = pd.DataFrame(columns= ['ts','s', 'pts'])
#dataframe for evts data
df_evts = pd.DataFrame(columns= ['ts','s', 'evts'])
#dataframe for dir data
df_dir = pd.DataFrame(columns= ['ts','s', 'dir', 'sig'])
```

```
f = 0
with open('livedata_small.json') as f_livedata:
    for i, line in enumerate(f_livedata):
        #print(line)
        s = js.loads(line)
        f, l_data = data_parsing(s)
        #print(f)
        if f == 1:
          df_gd_gidx.loc[len(df_gd_gidx.index)] = l_data
          #print(l_data)
        elif f ==2 :
          df_pc_gidx.loc[len(df_pc_gidx.index)] = l_data
        elif f ==3 :
          df_l_gidx.loc[len(df_l_gidx.index)] = l_data
        elif f ==4 :
          df_pd_gidx.loc[len(df_pd_gidx.index)] = l_data
        elif f ==5 :
          df_gp3_gidx.loc[len(df_gp3_gidx.index)] = l_data
```

```
df_gd = df_gd_gidx.pivot(index=['ts'], columns= ['eye']).stack()
print("gd DataFrame\n",df_gd)

df_pc = df_pc_gidx.pivot(index=['ts'], columns= ['eye']).stack()
print("pc DataFrame\n",df_pc)


df_pd = df_pd_gidx.pivot(index=['ts'], columns= ['eye']).stack()
print("pd DataFrame\n",df_pd)

print("gp3 Dataframe\n", df_gp3_gidx.set_index('ts'))
print("ac Dataframe \n", df_ac.set_index('ts'))
```

```
gd DataFrame
                        s    gidx                          gd
ts         eye
2225159387 left     4  154857              [0.0, 0.0, 0.0]
           right    0  154857    [-0.053, 0.0201, 0.9984]
2225169378 left     0  154858   [-0.2212, 0.0316, 0.9747]
           right    0  154858   [-0.0574, 0.0179, 0.9982]
2225179369 left     4  154859              [0.0, 0.0, 0.0]
           right    0  154859    [-0.0596, 0.018, 0.9981]
    pc DataFrame
                        s    gidx                         pc
    ts         eye
    2225159387 left     1  154857           [0.0, 0.0, 0.0]
               right    0  154857    [-31.65, -24.3, -42.0]
    2225169378 left     0  154858    [28.94, -29.2, -41.14]
               right    0  154858  [-31.89, -24.05, -42.01]
    2225179369 left     1  154859           [0.0, 0.0, 0.0]
               right    0  154859   [-31.62, -24.3, -42.01]
```

gy Dataframe

ts	s	gy
2225119438	0	[1.12, 4.692, 2.818]
2225130382	0	[0.98, 5.183, 2.051]
2225141326	0	[0.49, 5.203, 1.131]
2225152270	0	[0.28, 4.923, 0.299]
2225163214	0	[-0.42, 4.708, -0.094]

```
index_list = df_gd.index.values
x = index_list[0]
print(x)
y = 2225159387, 'right'
print(df_gd.loc[y])
```

```
(2225159387, 'left')
s                                0
gidx                        154857
gd       [-0.053, 0.0201, 0.9984]
Name: (2225159387, right), dtype: object
```