

03_validation

September 14, 2025

```
[2]: %load_ext autoreload
      %autoreload 2
```

```
[3]: import pandas as pd
      import pickle
      import re
      from dime_take_home.utils import get_locations_from_text
```

```
2025-09-13 18:16:50.070 | INFO      |
dime_take_home.config:<module>:11 - PROJ_ROOT
path is: /Users/alexisrutherford/dime_take_home
```

0.1 Take a quick look at some [HDX food insecurity data on Syria](#)

```
[8]: df = pd.read_excel('../data/external/syr-food-sector-pin_severity_2024.xlsx',
      ↪sheet_name='Food_PiN_by_severity_HNO_2024', header=1)
```

```
[9]: df.head()
```

```
[9]:  #adm1 +name #adm1 +code #adm2 +name #adm2 +code #adm3 +name #adm3 +code \
0      Damascus      SY01      Damascus      SY0100      Damascus      SY010000
1      Aleppo      SY02      Ain Al Arab      SY0206      Lower Shyookh      SY020601
2      Aleppo      SY02      Al Bab      SY0202      Dayr Hafir      SY020202
3      Aleppo      SY02      Ain Al Arab      SY0206      Ain al Arab      SY020600
4      Aleppo      SY02      A'zaz      SY0204      Tall Refaat      SY020402

      #Indicator +num      #inneed +ind
0              3  898967.999742
1              3  12597.570000
2              3  20229.982609
3              3  60936.990000
4              3  4928.040000
```

```
[10]: with open("../data/processed/en_locations_reversed.pkl", "rb") as f:
      locations_dict_en_reversed = pickle.load(f)
```

```
[13]: locations_pattern_en = re.compile(r"\b(?:%s)\b" % "|".join(map(re.escape,
↳ locations_dict_en_reversed.keys())) , re.IGNORECASE)
```

```
[21]: get_locations_from_text('al bab', locations_lookup=locations_dict_en_reversed,
↳ locations_pattern=locations_pattern_en)
```

```
[21]: ['sy_hl_3']
```

1 Try Palestine Food Insecurity Index

```
[21]: df = pd.read_csv('../data/external/ipc_pse_area_long.csv', header=0)
```

```
[22]: df.head()
```

```
[22]:
```

	Date of analysis	Country	Total country population	Level 1 \
0	#date+analysis	#country+code	#population	#loc+level1
1	Jul 2025	PSE	2100719	NaN
2	Jul 2025	PSE	2100719	NaN
3	Jul 2025	PSE	2100719	NaN
4	Jul 2025	PSE	2100719	NaN

	Area Validity period	From	To	Phase \	
0	#loc+area	#period+v_ipc	#date+start	#date+end	#severity+v_ipc
1	Deir al-Balah	current	2025-07-01	2025-08-31	all
2	Deir al-Balah	current	2025-07-01	2025-08-31	3+
3	Deir al-Balah	current	2025-07-01	2025-08-31	1
4	Deir al-Balah	current	2025-07-01	2025-08-31	2

	Number	Percentage
0	#affected+num	#affected+pct
1	482315	1.0
2	482315	1
3	0	0
4	0	0

```
[23]: df = df.iloc[1:,:]
# Drop the extra header ro
```

```
[24]: df.groupby(['From', 'To']).size()
```

```
[24]:
```

From	To	
2023-11-01	2023-12-31	21
2023-12-01	2024-02-29	21
2024-02-01	2024-03-31	21
2024-03-01	2024-07-31	21

```

2024-05-01 2024-06-30 21
2024-06-01 2024-09-30 21
2024-09-01 2024-10-31 21
2024-11-01 2025-04-30 21
2025-04-01 2025-05-31 35
2025-05-01 2025-09-30 35
2025-07-01 2025-08-31 21
2025-08-01 2025-09-30 21
dtype: int64

```

```
[26]: df = df[df['From'].isin(['2024-06-01', '2024-07-01'])]
```

```
[28]: df.shape
```

```
[28]: (21, 11)
```

```
[ ]: df
```

```

[ ]:      Date of analysis Country Total country population Level 1      Area \
162      May 2024      PSE      2251457      NaN      North Gaza
163      May 2024      PSE      2251457      NaN      North Gaza
164      May 2024      PSE      2251457      NaN      North Gaza
165      May 2024      PSE      2251457      NaN      North Gaza
166      May 2024      PSE      2251457      NaN      North Gaza
167      May 2024      PSE      2251457      NaN      North Gaza
168      May 2024      PSE      2251457      NaN      North Gaza
176      May 2024      PSE      2251457      NaN      Rafah
177      May 2024      PSE      2251457      NaN      Rafah
178      May 2024      PSE      2251457      NaN      Rafah
179      May 2024      PSE      2251457      NaN      Rafah
180      May 2024      PSE      2251457      NaN      Rafah
181      May 2024      PSE      2251457      NaN      Rafah
182      May 2024      PSE      2251457      NaN      Rafah
190      May 2024      PSE      2251457      NaN      Deir Al-Balah
191      May 2024      PSE      2251457      NaN      Deir Al-Balah
192      May 2024      PSE      2251457      NaN      Deir Al-Balah
193      May 2024      PSE      2251457      NaN      Deir Al-Balah
194      May 2024      PSE      2251457      NaN      Deir Al-Balah
195      May 2024      PSE      2251457      NaN      Deir Al-Balah
196      May 2024      PSE      2251457      NaN      Deir Al-Balah

```

```

      Validity period      From      To Phase      Number Percentage
162 first projection 2024-06-01 2024-09-30  all      300000          1.0
163 first projection 2024-06-01 2024-09-30   3+      300000           1
164 first projection 2024-06-01 2024-09-30    1           0           0
165 first projection 2024-06-01 2024-09-30    2           0           0
166 first projection 2024-06-01 2024-09-30    3       75000         0.25

```

167	first projection	2024-06-01	2024-09-30	4	150000	0.5
168	first projection	2024-06-01	2024-09-30	5	75000	0.25
176	first projection	2024-06-01	2024-09-30	all	200000	1.0
177	first projection	2024-06-01	2024-09-30	3+	190000	0.95
178	first projection	2024-06-01	2024-09-30	1	0	0
179	first projection	2024-06-01	2024-09-30	2	10000	0.05
180	first projection	2024-06-01	2024-09-30	3	50000	0.25
181	first projection	2024-06-01	2024-09-30	4	70000	0.35
182	first projection	2024-06-01	2024-09-30	5	70000	0.35
190	first projection	2024-06-01	2024-09-30	all	1751457	1.0
191	first projection	2024-06-01	2024-09-30	3+	1663884	0.95
192	first projection	2024-06-01	2024-09-30	1	0	0
193	first projection	2024-06-01	2024-09-30	2	87573	0.05
194	first projection	2024-06-01	2024-09-30	3	788156	0.45
195	first projection	2024-06-01	2024-09-30	4	525437	0.3
196	first projection	2024-06-01	2024-09-30	5	350291	0.2

2 Now try WB FEWS data

```
[34]: df = pd.read_csv('../data/external/
↳FEWS_October_2024_Update_TrueBoundaries_12-09-24.csv')
```

```
[37]: df['country'].unique()
```

```
[37]: array(['Afghanistan', 'Burkina Faso', 'Burundi', 'Cameroon', 'Chad',
'Democratic Republic of the Congo', 'Ethiopia', 'Guatemala',
'Haiti', 'Kenya', 'Madagascar', 'Malawi', 'Mali', 'Mozambique',
'Niger', 'Nigeria', 'Somalia', 'South Sudan', 'Sudan', 'Uganda',
'Yemen', 'Zimbabwe', 'System.IO.MemoryStream'], dtype=object)
```

```
[36]: df[df['country']=='Syria']
```

```
[36]: Empty DataFrame
Columns: [country, admin_code, year_month, year, month, fews_ipc, fews_ha,
fews_proj_near, fews_proj_near_ha, fews_proj_med, fews_proj_med_ha, pop,
pop_source, fews_ipc_adjusted, fews_proj_med_adjusted, admin_name]
Index: []
```

```
[ ]:
```