

Task1

October 15, 2024

```
[31]: import pandas as pd
```

```
[32]: data = {'Country': ['USA', 'UK', 'Germany'],  
            'Population': [331, 67, 83],  
            'GDP': [21.43, 2.83, 3.86]}  
df_dict = pd.DataFrame(data)  
print(df_dict)
```

	Country	Population	GDP
0	USA	331	21.43
1	UK	67	2.83
2	Germany	83	3.86

```
[33]: df_dict.to_csv("countries.csv")
```

```
[34]: lists_uni = [ ["AGH", "UBB", "UW"],  
                  ["Kraków", "Bielsko-Biała", "Warszawa"] ]  
  
pd.DataFrame(lists_uni)
```

```
[34]:
```

	0	1	2
0	AGH	UBB	UW
1	Kraków	Bielsko-Biała	Warszawa

```
[35]: pd.DataFrame(lists_uni).T
```

```
[35]:
```

	0	1
0	AGH	Kraków
1	UBB	Bielsko-Biała
2	UW	Warszawa

```
[36]: file_path = 'IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV'  
chunksize = 100000  
df_csv = pd.concat(pd.read_csv(file_path, chunksize=chunksize,  
                                ↪encoding="utf-8", low_memory=False))
```

```
[37]: df_csv.head(10)
```

```

[37]:  year      source  channel recipient_isocode \
0  1990  Australia  BIL_AUS                AGO
1  1990  Australia  BIL_AUS                BDI
2  1990  Australia  BIL_AUS                BEN
3  1990  Australia  BIL_AUS                BFA
4  1990  Australia  BIL_AUS                BWA
5  1990  Australia  BIL_AUS                CAF
6  1990  Australia  BIL_AUS                CHN
7  1990  Australia  BIL_AUS                CIV
8  1990  Australia  BIL_AUS                CMR
9  1990  Australia  BIL_AUS                COD

      recipient_country  gbd_location_id  wb_regioncode \
0                Angola                168            SSA
1              Burundi                175            SSA
2              Benin                200            SSA
3          Burkina Faso                201            SSA
4              Botswana                193            SSA
5    Central African Republic                169            SSA
6                China                  6            EAP
7          Cote d'Ivoire                205            SSA
8              Cameroon                202            SSA
9  Democratic Republic of the Congo                171            SSA

wb_location_id      gbd_region  gbd_region_id  ... \
0          242  Sub-Saharan Africa, Central        167.0  ...
1          242  Sub-Saharan Africa, Eastern        174.0  ...
2          242  Sub-Saharan Africa, Western        199.0  ...
3          242  Sub-Saharan Africa, Western        199.0  ...
4          242  Sub-Saharan Africa, Southern        192.0  ...
5          242  Sub-Saharan Africa, Central        167.0  ...
6          239                Asia, East           5.0  ...
7          242  Sub-Saharan Africa, Western        199.0  ...
8          242  Sub-Saharan Africa, Western        199.0  ...
9          242  Sub-Saharan Africa, Central        167.0  ...

other_dah_20  rmh_dah_20  nch_dah_20  ncd_dah_20  hiv_dah_20  mal_dah_20 \
0              0          5          0          0          7          3
1              0          6          0          0          5          1
2              0          6          0          0          5          2
3              0          5          0          0          7          2
4              0          1          0          0         23          -
5              0          1          0          0          2          -
6             38          7          0          0        367          5
7              0          5          0          0         18          1
8              0          2          0          0          8          1
9              0         14          0          0         18          6

```

	tb_dah_20	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
0	0	0	0	-
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	-
5	0	0	0	-
6	0	130	0	0
7	0	0	0	0
8	0	0	0	-
9	0	0	0	0

[10 rows x 76 columns]

```
[38]: df_csv.tail(10)
```

```
[38]:
```

	year	source	channel	recipient_isocode	\
384296	2020	United_States	INTLNGO	QZA	
384297	2020	United_States	NGO	QZA	
384298	2020	United_States	PAHO	QZA	
384299	2020	United_States	UNAIDS	QZA	
384300	2020	United_States	UNFPA	QZA	
384301	2020	United_States	UNICEF	QZA	
384302	2020	United_States	UNITAID	QZA	
384303	2020	United_States	UNITAID	QZA	
384304	2020	United_States	WB_IDA	QZA	
384305	2020	United_States	WHO	QZA	

	recipient_country	gbd_location_id	wb_regioncode	\
384296	Unallocated/Unspecified	44598	NaN	
384297	Unallocated/Unspecified	44598	NaN	
384298	Unallocated/Unspecified	44598	NaN	
384299	Unallocated/Unspecified	44598	NaN	
384300	Unallocated/Unspecified	44598	NaN	
384301	Unallocated/Unspecified	44598	NaN	
384302	Unallocated/Unspecified	44598	NaN	
384303	Unallocated/Unspecified	44598	NaN	
384304	Unallocated/Unspecified	44598	NaN	
384305	Unallocated/Unspecified	44598	NaN	

	wb_location_id	gbd_region	gbd_region_id	...	\
384296	44621	Unallocated/Unspecified	44598.0	...	
384297	44621	Unallocated/Unspecified	44598.0	...	
384298	44621	Unallocated/Unspecified	44598.0	...	
384299	44621	Unallocated/Unspecified	44598.0	...	
384300	44621	Unallocated/Unspecified	44598.0	...	

384301	44621	Unallocated/Unspecified	44598.0	...
384302	44621	Unallocated/Unspecified	44598.0	...
384303	44621	Unallocated/Unspecified	44598.0	...
384304	44621	Unallocated/Unspecified	44598.0	...
384305	44621	Unallocated/Unspecified	44598.0	...

	other_dah_20	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	\
384296	132174	2351	4047	120	5090	623	
384297	1293825	230205	148340	11652	264680	20224	
384298	13449	0	1121	9526	0	0	
384299	0	0	0	0	57674	0	
384300	0	0	0	0	0	0	
384301	0	17509	265811	0	9398	0	
384302	0	0	0	0	1037	1085	
384303	0	0	0	0	1250	166	
384304	49485	17337	26498	1907	5076	763	
384305	17823	13250	32712	22436	11098	8991	

	tb_dah_20	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
384296	765		25	640
384297	13315		98494	61832
384298	0		58465	21481
384299	0		0	2885
384300	0		0	0
384301	0		0	116214
384302	895		449	407
384303	112		173	217
384304	3424		46292	47569
384305	9457		193266	113937

[10 rows x 76 columns]

```
[39]: df_csv.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 384306 entries, 0 to 384305
Data columns (total 76 columns):
#   Column              Non-Null Count  Dtype
---  -
0   year                384306 non-null  int64
1   source              384306 non-null  object
2   channel             384306 non-null  object
3   recipient_isocode   384306 non-null  object
4   recipient_country   383773 non-null  object
5   gbd_location_id     384306 non-null  int64
6   wb_regioncode       370318 non-null  object
7   wb_location_id      384306 non-null  int64
8   gbd_region          383993 non-null  object
```

9	gbd_region_id	383993	non-null	float64
10	gbd_superregion	383993	non-null	object
11	gbd_superregion_id	383993	non-null	float64
12	elim_ch	384306	non-null	int64
13	prelim_est	384306	non-null	int64
14	dah_20	384306	non-null	object
15	rmh_fp_dah_20	384306	non-null	object
16	rmh_mh_dah_20	384306	non-null	object
17	rmh_hss_other_dah_20	384306	non-null	object
18	rmh_hss_hrh_dah_20	384306	non-null	object
19	rmh_other_dah_20	384306	non-null	object
20	nch_cnn_dah_20	384306	non-null	object
21	nch_cnv_dah_20	384306	non-null	object
22	nch_other_dah_20	384306	non-null	object
23	nch_hss_other_dah_20	384306	non-null	object
24	nch_hss_hrh_dah_20	384306	non-null	object
25	hiv_treat_dah_20	384306	non-null	object
26	hiv_prev_dah_20	384306	non-null	object
27	hiv_pmtct_dah_20	384306	non-null	object
28	hiv_other_dah_20	384306	non-null	object
29	hiv_ct_dah_20	384306	non-null	object
30	hiv_ovc_dah_20	384306	non-null	object
31	hiv_care_dah_20	384306	non-null	object
32	hiv_hss_other_dah_20	384306	non-null	object
33	hiv_hss_hrh_dah_20	384306	non-null	object
34	hiv_amr_dah_20	384306	non-null	object
35	mal_diag_dah_20	384306	non-null	object
36	mal_hss_other_dah_20	384306	non-null	object
37	mal_hss_hrh_dah_20	384306	non-null	object
38	mal_con_nets_dah_20	384306	non-null	object
39	mal_con_irs_dah_20	384306	non-null	object
40	mal_con_oth_dah_20	384306	non-null	object
41	mal_treat_dah_20	384306	non-null	object
42	mal_comm_con_dah_20	384306	non-null	object
43	mal_other_dah_20	384306	non-null	object
44	mal_amr_dah_20	384306	non-null	object
45	tb_other_dah_20	384306	non-null	object
46	tb_treat_dah_20	384306	non-null	object
47	tb_diag_dah_20	384306	non-null	object
48	tb_hss_other_dah_20	384306	non-null	object
49	tb_hss_hrh_dah_20	384306	non-null	object
50	tb_amr_dah_20	384306	non-null	object
51	oid_hss_other_dah_20	384306	non-null	object
52	oid_hss_hrh_dah_20	384306	non-null	object
53	oid_ebz_dah_20	384306	non-null	object
54	oid_zika_dah_20	384306	non-null	object
55	oid_covid_dah_20	384306	non-null	object
56	oid_other_dah_20	384306	non-null	object

```

57  oid_amr_dah_20      384306 non-null object
58  ncd_hss_other_dah_20 384306 non-null object
59  ncd_hss_hrh_dah_20   384306 non-null object
60  ncd_tobac_dah_20     384306 non-null object
61  ncd_mental_dah_20    384306 non-null object
62  ncd_other_dah_20     384306 non-null object
63  swap_hss_other_dah_20 384306 non-null object
64  swap_hss_hrh_dah_20  384306 non-null object
65  swap_hss_pp_dah_20   384306 non-null object
66  other_dah_20         384306 non-null object
67  rmh_dah_20           384306 non-null object
68  nch_dah_20           384306 non-null object
69  ncd_dah_20           384306 non-null object
70  hiv_dah_20           384306 non-null object
71  mal_dah_20           384306 non-null object
72  tb_dah_20            384306 non-null object
73  swap_hss_total_dah_20 384306 non-null object
74  oid_dah_20           384306 non-null object
75  unalloc_dah_20       384306 non-null object
dtypes: float64(2), int64(5), object(69)
memory usage: 222.8+ MB

```

```

[40]: # Here we can see that the columns 14-75 are the 'object' type when in reality
      ↪ they contain numerical valuse, that's why we will change it here:
cols_to_convert = df_csv.columns[14:]
df_csv[cols_to_convert] = df_csv[cols_to_convert].apply(pd.to_numeric,
      ↪ errors='coerce')

```

```

[41]: df_csv.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 384306 entries, 0 to 384305
Data columns (total 76 columns):
#   Column                Non-Null Count  Dtype
---  -
0   year                  384306 non-null  int64
1   source                384306 non-null  object
2   channel               384306 non-null  object
3   recipient_isocode     384306 non-null  object
4   recipient_country     383773 non-null  object
5   gbd_location_id       384306 non-null  int64
6   wb_regioncode         370318 non-null  object
7   wb_location_id        384306 non-null  int64
8   gbd_region            383993 non-null  object
9   gbd_region_id         383993 non-null  float64
10  gbd_superregion       383993 non-null  object
11  gbd_superregion_id    383993 non-null  float64
12  elim_ch               384306 non-null  int64

```

13	prelim_est	384306	non-null	int64
14	dah_20	330001	non-null	float64
15	rmh_fp_dah_20	374481	non-null	float64
16	rmh_mh_dah_20	361849	non-null	float64
17	rmh_hss_other_dah_20	375779	non-null	float64
18	rmh_hss_hrh_dah_20	370646	non-null	float64
19	rmh_other_dah_20	367001	non-null	float64
20	nch_cnn_dah_20	361438	non-null	float64
21	nch_cnv_dah_20	366905	non-null	float64
22	nch_other_dah_20	373534	non-null	float64
23	nch_hss_other_dah_20	371627	non-null	float64
24	nch_hss_hrh_dah_20	364009	non-null	float64
25	hiv_treat_dah_20	362595	non-null	float64
26	hiv_prev_dah_20	360775	non-null	float64
27	hiv_pmtct_dah_20	361860	non-null	float64
28	hiv_other_dah_20	338847	non-null	float64
29	hiv_ct_dah_20	362654	non-null	float64
30	hiv_ovc_dah_20	361165	non-null	float64
31	hiv_care_dah_20	367222	non-null	float64
32	hiv_hss_other_dah_20	362992	non-null	float64
33	hiv_hss_hrh_dah_20	350628	non-null	float64
34	hiv_amr_dah_20	368712	non-null	float64
35	mal_diag_dah_20	375339	non-null	float64
36	mal_hss_other_dah_20	376802	non-null	float64
37	mal_hss_hrh_dah_20	373283	non-null	float64
38	mal_con_nets_dah_20	376273	non-null	float64
39	mal_con_irs_dah_20	375858	non-null	float64
40	mal_con_oth_dah_20	376861	non-null	float64
41	mal_treat_dah_20	375816	non-null	float64
42	mal_comm_con_dah_20	375780	non-null	float64
43	mal_other_dah_20	367092	non-null	float64
44	mal_amr_dah_20	375345	non-null	float64
45	tb_other_dah_20	363532	non-null	float64
46	tb_treat_dah_20	368462	non-null	float64
47	tb_diag_dah_20	365444	non-null	float64
48	tb_hss_other_dah_20	374214	non-null	float64
49	tb_hss_hrh_dah_20	374144	non-null	float64
50	tb_amr_dah_20	371101	non-null	float64
51	oid_hss_other_dah_20	373347	non-null	float64
52	oid_hss_hrh_dah_20	375266	non-null	float64
53	oid_ebz_dah_20	379424	non-null	float64
54	oid_zika_dah_20	384142	non-null	float64
55	oid_covid_dah_20	384300	non-null	float64
56	oid_other_dah_20	356419	non-null	float64
57	oid_amr_dah_20	383970	non-null	float64
58	ncd_hss_other_dah_20	376866	non-null	float64
59	ncd_hss_hrh_dah_20	381501	non-null	float64
60	ncd_tobac_dah_20	383464	non-null	float64

```

61 ncd_mental_dah_20      379001 non-null float64
62 ncd_other_dah_20      376160 non-null float64
63 swap_hss_other_dah_20 361772 non-null float64
64 swap_hss_hrh_dah_20   363913 non-null float64
65 swap_hss_pp_dah_20    380760 non-null float64
66 other_dah_20          342691 non-null float64
67 rmh_dah_20            353243 non-null float64
68 nch_dah_20            355316 non-null float64
69 ncd_dah_20            370024 non-null float64
70 hiv_dah_20            345158 non-null float64
71 mal_dah_20            365527 non-null float64
72 tb_dah_20             364731 non-null float64
73 swap_hss_total_dah_20 352846 non-null float64
74 oid_dah_20            351367 non-null float64
75 unalloc_dah_20        228920 non-null float64
dtypes: float64(64), int64(5), object(7)
memory usage: 222.8+ MB

```

```
[42]: df_csv.shape
```

```
[42]: (384306, 76)
```

```
[43]: df_csv.describe()
```

```

[43]:
count      year  gbd_location_id  wb_location_id  gbd_region_id  \
count  384306.000000    384306.000000    384306.000000    383993.000000
mean    2008.127521      1765.935533      2240.752439      1745.812671
std         6.945191      8325.915434      9204.906147      8328.525983
min     1990.000000         1.000000       239.000000         1.000000
25%     2004.000000        110.000000       241.000000        96.000000
50%     2009.000000        169.000000       242.000000       159.000000
75%     2014.000000        200.000000       242.000000       192.000000
max     2020.000000      44598.000000      44621.000000      44598.000000

      gbd_superregion_id      elim_ch      prelim_est      dah_20  \
count    383993.000000    384306.000000    384306.000000    3.300010e+05
mean         1733.144388         0.252052         0.014358    2.519363e+03
std         8330.949734         0.434191         0.118963    3.669331e+04
min           1.000000         0.000000         0.000000   -1.231000e+03
25%           64.000000         0.000000         0.000000    5.000000e+00
50%          158.000000         0.000000         0.000000    4.400000e+01
75%          166.000000         1.000000         0.000000    3.620000e+02
max         44598.000000         1.000000         1.000000    7.860631e+06

      rmh_fp_dah_20  rmh_mh_dah_20  ...  other_dah_20      rmh_dah_20  \
count  374481.000000  361849.000000  ...  3.426910e+05    353243.000000
mean         99.103615      105.914586  ...  3.966702e+02      319.974462
std        2204.948337      2242.693627  ...  7.124986e+03      5571.734197

```


min	-190.000000	-1180.000000	...	-5.290000e+02	-2215.000000
25%	0.000000	0.000000	...	0.000000e+00	0.000000
50%	0.000000	0.000000	...	0.000000e+00	0.000000
75%	0.000000	0.000000	...	8.000000e+00	1.000000
max	469563.000000	625147.000000	...	1.293825e+06	859482.000000

	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	\
count	355316.000000	370024.000000	3.451580e+05	365527.000000	
mean	393.067827	31.765712	5.588664e+02	110.119873	
std	6890.644755	805.968524	1.651906e+04	2529.367685	
min	-1192.000000	-34.000000	-3.290000e+02	-377.000000	
25%	0.000000	0.000000	0.000000e+00	0.000000	
50%	0.000000	0.000000	0.000000e+00	0.000000	
75%	3.000000	0.000000	1.000000e+01	0.000000	
max	728728.000000	109415.000000	4.906421e+06	584684.000000	

	tb_dah_20	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20	
count	364731.000000	352846.000000	3.513670e+05	228920.000000	
mean	71.515577	311.781548	1.637497e+02	18.532042	
std	1512.909226	5437.156070	5.002576e+03	940.664108	
min	-132.000000	-357.000000	-8.600000e+01	0.000000	
25%	0.000000	0.000000	0.000000e+00	0.000000	
50%	0.000000	0.000000	0.000000e+00	0.000000	
75%	0.000000	1.000000	0.000000e+00	0.000000	
max	352293.000000	625164.000000	1.468522e+06	200761.000000	

[8 rows x 69 columns]

```
[44]: df_csv.describe(include='object')
```

```
[44]:
```

	source	channel	recipient_isocode	recipient_country	\
count	384306	384306	384306	383773	
unique	32	46	176	174	
top	United_Kingdom	GFATM	INKIND	Administrative expenses	
freq	20452	57306	8379	8379	

	wb_regioncode	gbd_region	gbd_superregion	
count	370318	383993	383993	
unique	7	22	9	
top	SSA	Sub-Saharan Africa, Western	Sub-Saharan Africa	
freq	172395	72174	168660	

```
[45]: df_csv.describe(include='all')
```

```
[45]:
```

	year	source	channel	recipient_isocode	\
count	384306.000000	384306	384306	384306	
unique	NaN	32	46	176	

top	NaN	United_Kingdom	GFATM	INKIND
freq	NaN	20452	57306	8379
mean	2008.127521	NaN	NaN	NaN
std	6.945191	NaN	NaN	NaN
min	1990.000000	NaN	NaN	NaN
25%	2004.000000	NaN	NaN	NaN
50%	2009.000000	NaN	NaN	NaN
75%	2014.000000	NaN	NaN	NaN
max	2020.000000	NaN	NaN	NaN

	recipient_country	gbd_location_id	wb_regioncode	\
count	383773	384306.000000	370318	
unique	174	NaN	7	
top	Administrative expenses	NaN	SSA	
freq	8379	NaN	172395	
mean	NaN	1765.935533	NaN	
std	NaN	8325.915434	NaN	
min	NaN	1.000000	NaN	
25%	NaN	110.000000	NaN	
50%	NaN	169.000000	NaN	
75%	NaN	200.000000	NaN	
max	NaN	44598.000000	NaN	

	wb_location_id	gbd_region	gbd_region_id	...	\
count	384306.000000	383993	383993.000000	...	
unique	NaN	22	NaN	...	
top	NaN	Sub-Saharan Africa, Western	NaN	...	
freq	NaN	72174	NaN	...	
mean	2240.752439	NaN	1745.812671	...	
std	9204.906147	NaN	8328.525983	...	
min	239.000000	NaN	1.000000	...	
25%	241.000000	NaN	96.000000	...	
50%	242.000000	NaN	159.000000	...	
75%	242.000000	NaN	192.000000	...	
max	44621.000000	NaN	44598.000000	...	

	other_dah_20	rmh_dah_20	nch_dah_20	ncd_dah_20	\
count	3.426910e+05	353243.000000	355316.000000	370024.000000	
unique	NaN	NaN	NaN	NaN	
top	NaN	NaN	NaN	NaN	
freq	NaN	NaN	NaN	NaN	
mean	3.966702e+02	319.974462	393.067827	31.765712	
std	7.124986e+03	5571.734197	6890.644755	805.968524	
min	-5.290000e+02	-2215.000000	-1192.000000	-34.000000	
25%	0.000000e+00	0.000000	0.000000	0.000000	
50%	0.000000e+00	0.000000	0.000000	0.000000	
75%	8.000000e+00	1.000000	3.000000	0.000000	

max	1.293825e+06	859482.000000	728728.000000	109415.000000
-----	--------------	---------------	---------------	---------------

	hiv_dah_20	mal_dah_20	tb_dah_20	swap_hss_total_dah_20 \
count	3.451580e+05	365527.000000	364731.000000	352846.000000
unique	NaN	NaN	NaN	NaN
top	NaN	NaN	NaN	NaN
freq	NaN	NaN	NaN	NaN
mean	5.588664e+02	110.119873	71.515577	311.781548
std	1.651906e+04	2529.367685	1512.909226	5437.156070
min	-3.290000e+02	-377.000000	-132.000000	-357.000000
25%	0.000000e+00	0.000000	0.000000	0.000000
50%	0.000000e+00	0.000000	0.000000	0.000000
75%	1.000000e+01	0.000000	0.000000	1.000000
max	4.906421e+06	584684.000000	352293.000000	625164.000000

	oid_dah_20	unalloc_dah_20
count	3.513670e+05	228920.000000
unique	NaN	NaN
top	NaN	NaN
freq	NaN	NaN
mean	1.637497e+02	18.532042
std	5.002576e+03	940.664108
min	-8.600000e+01	0.000000
25%	0.000000e+00	0.000000
50%	0.000000e+00	0.000000
75%	0.000000e+00	0.000000
max	1.468522e+06	200761.000000

[11 rows x 76 columns]

```
[46]: df_csv = df_csv.dropna().reset_index(drop=True)
```

```
[48]: df_csv.head(100)
```

```
[48]:
```

	year	source	channel	recipient_isocode \
0	1990	Australia	BIL_AUS	CHN
1	1990	Australia	BIL_AUS	CIV
2	1990	Australia	BIL_AUS	COD
3	1990	Australia	BIL_AUS	GHA
4	1990	Australia	BIL_AUS	KIR
..
95	1990	Australia	WB_IDA	WSM
96	1990	Australia	WB_IDA	ZMB
97	1990	Australia	WB_IDA	ZWE
98	1990	Austria	BIL_AUT	IRQ
99	1990	Austria	BIL_AUT	UGA

	recipient_country	gbd_location_id	wb_regioncode	\
0	China	6	EAP	
1	Cote d'Ivoire	205	SSA	
2	Democratic Republic of the Congo	171	SSA	
3	Ghana	207	SSA	
4	Kiribati	23	EAP	
..	
95	Samoa	27	EAP	
96	Zambia	191	SSA	
97	Zimbabwe	198	SSA	
98	Iraq	143	MNA	
99	Uganda	190	SSA	

	wb_location_id	gbd_region	gbd_region_id	...	\
0	239	Asia, East	5.0	...	
1	242	Sub-Saharan Africa, Western	199.0	...	
2	242	Sub-Saharan Africa, Central	167.0	...	
3	242	Sub-Saharan Africa, Western	199.0	...	
4	239	Oceania	21.0	...	
..	
95	239	Oceania	21.0	...	
96	242	Sub-Saharan Africa, Eastern	174.0	...	
97	242	Sub-Saharan Africa, Southern	192.0	...	
98	243	North Africa/Middle East	138.0	...	
99	242	Sub-Saharan Africa, Eastern	174.0	...	

	other_dah_20	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	\
0	38.0	7.0	0.0	0.0	367.0	5.0	
1	0.0	5.0	0.0	0.0	18.0	1.0	
2	0.0	14.0	0.0	0.0	18.0	6.0	
3	0.0	19.0	0.0	0.0	14.0	3.0	
4	28.0	20.0	56.0	28.0	0.0	0.0	
..	
95	0.0	0.0	0.0	0.0	0.0	0.0	
96	0.0	0.0	0.0	0.0	0.0	0.0	
97	0.0	0.0	0.0	0.0	0.0	0.0	
98	60.0	0.0	0.0	0.0	0.0	0.0	
99	0.0	0.0	0.0	0.0	0.0	0.0	

	tb_dah_20	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
0	0.0	130.0	0.0	0.0
1	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	0.0	102.0	0.0	0.0
..
95	0.0	0.0	0.0	0.0

96	0.0	0.0	0.0	0.0
97	0.0	0.0	0.0	0.0
98	0.0	0.0	0.0	0.0
99	0.0	38906.0	0.0	0.0

[100 rows x 76 columns]

```
[49]: df_csv["recipient_country"]
```

```
[49]: 0          China
      1      Cote d'Ivoire
      2  Democratic Republic of the Congo
      3          Ghana
      4      Kiribati
      ...
134557  Democratic Republic of the Congo
134558          Mali
134559          Niger
134560          Nigeria
134561          Rwanda
Name: recipient_country, Length: 134562, dtype: object
```

```
[50]: df_csv.recipient_country
```

```
[50]: 0          China
      1      Cote d'Ivoire
      2  Democratic Republic of the Congo
      3          Ghana
      4      Kiribati
      ...
134557  Democratic Republic of the Congo
134558          Mali
134559          Niger
134560          Nigeria
134561          Rwanda
Name: recipient_country, Length: 134562, dtype: object
```

```
[51]: df_csv[["recipient_country", "gbd_location_id", "wb_regioncode"]]
```

```
[51]:      recipient_country  gbd_location_id  wb_regioncode
0          China          6          EAP
1      Cote d'Ivoire        205          SSA
2  Democratic Republic of the Congo        171          SSA
3          Ghana        207          SSA
4      Kiribati          23          EAP
...
134557  Democratic Republic of the Congo        171          SSA
```

134558	Mali	211	SSA
134559	Niger	213	SSA
134560	Nigeria	214	SSA
134561	Rwanda	185	SSA

[134562 rows x 3 columns]

```
[52]: df_csv.loc[:, "year":"recipient_country"]
```

```
[52]:
```

	year	source	channel	recipient_isocode	\
0	1990	Australia	BIL_AUS	CHN	
1	1990	Australia	BIL_AUS	CIV	
2	1990	Australia	BIL_AUS	COD	
3	1990	Australia	BIL_AUS	GHA	
4	1990	Australia	BIL_AUS	KIR	
...	
134557	2018	United_States	WHO	COD	
134558	2018	United_States	WHO	MLI	
134559	2018	United_States	WHO	NER	
134560	2018	United_States	WHO	NGA	
134561	2018	United_States	WHO	RWA	

	recipient_country
0	China
1	Cote d'Ivoire
2	Democratic Republic of the Congo
3	Ghana
4	Kiribati
...	...
134557	Democratic Republic of the Congo
134558	Mali
134559	Niger
134560	Nigeria
134561	Rwanda

[134562 rows x 5 columns]

```
[53]: df_csv.loc[1000:1500, "year":"recipient_country"]
```

```
[53]:
```

	year	source	channel	recipient_isocode	recipient_country
1000	1990	Germany	WB_IDA	GRD	Grenada
1001	1990	Germany	WB_IDA	GUY	Guyana
1002	1990	Germany	WB_IDA	HND	Honduras
1003	1990	Germany	WB_IDA	IDN	Indonesia
1004	1990	Germany	WB_IDA	IND	India
...
1496	1990	Netherlands	BIL_NLD	PAN	Panama

1497	1990	Netherlands	BIL_NLD	PER	Peru
1498	1990	Netherlands	BIL_NLD	PRY	Paraguay
1499	1990	Netherlands	BIL_NLD	RWA	Rwanda
1500	1990	Netherlands	BIL_NLD	SEN	Senegal

[501 rows x 5 columns]

```
[55]: df_csv.iloc[1000:1500, 0:4]
```

```
[55]:
```

	year	source	channel	recipient_isocode
1000	1990	Germany	WB_IDA	GRD
1001	1990	Germany	WB_IDA	GUY
1002	1990	Germany	WB_IDA	HND
1003	1990	Germany	WB_IDA	IDN
1004	1990	Germany	WB_IDA	IND
...
1495	1990	Netherlands	BIL_NLD	NPL
1496	1990	Netherlands	BIL_NLD	PAN
1497	1990	Netherlands	BIL_NLD	PER
1498	1990	Netherlands	BIL_NLD	PRY
1499	1990	Netherlands	BIL_NLD	RWA

[500 rows x 4 columns]

```
[56]: df_csv[df_csv["recipient_country"] == "China"]
```

```
[56]:
```

	year	source	channel	recipient_isocode	recipient_country \
0	1990	Australia	BIL_AUS	CHN	China
217	1990	Belgium	WB_IDA	CHN	China
295	1990	Canada	BIL_CAN	CHN	China
505	1990	Debt_repayments	WB_IDA	CHN	China
721	1990	Finland	WB_IDA	CHN	China
...
134022	2018	United_Kingdom	AsDB	CHN	China
134043	2018	United_Kingdom	EC	CHN	China
134264	2018	United_Kingdom	WB_IDA	CHN	China
134342	2018	United_States	AsDB	CHN	China
134498	2018	United_States	WB_IDA	CHN	China

	gbd_location_id	wb_regioncode	wb_location_id	gbd_region \
0	6	EAP	239	Asia, East
217	6	EAP	239	Asia, East
295	6	EAP	239	Asia, East
505	6	EAP	239	Asia, East
721	6	EAP	239	Asia, East
...
134022	6	EAP	239	Asia, East

134043	6	EAP	239	Asia, East
134264	6	EAP	239	Asia, East
134342	6	EAP	239	Asia, East
134498	6	EAP	239	Asia, East

	gbd_region_id	...	other_dah_20	rmh_dah_20	nch_dah_20	ncd_dah_20	\
0	5.0	...	38.0	7.0	0.0	0.0	
217	5.0	...	99.0	47.0	153.0	0.0	
295	5.0	...	673.0	0.0	0.0	0.0	
505	5.0	...	319.0	153.0	492.0	0.0	
721	5.0	...	46.0	22.0	72.0	0.0	
...	
134022	5.0	...	110.0	0.0	0.0	0.0	
134043	5.0	...	0.0	19.0	11.0	0.0	
134264	5.0	...	0.0	0.0	0.0	0.0	
134342	5.0	...	126.0	0.0	0.0	0.0	
134498	5.0	...	0.0	0.0	0.0	0.0	

	hiv_dah_20	mal_dah_20	tb_dah_20	swap_hss_total_dah_20	oid_dah_20	\
0	367.0	5.0	0.0	130.0	0.0	
217	0.0	0.0	79.0	408.0	79.0	
295	0.0	0.0	0.0	0.0	0.0	
505	0.0	0.0	256.0	1314.0	256.0	
721	0.0	0.0	37.0	192.0	37.0	
...	
134022	0.0	0.0	0.0	435.0	0.0	
134043	7.0	0.0	0.0	0.0	7.0	
134264	0.0	0.0	0.0	0.0	0.0	
134342	0.0	0.0	0.0	497.0	0.0	
134498	0.0	0.0	0.0	0.0	0.0	

	unalloc_dah_20
0	0.0
217	0.0
295	0.0
505	0.0
721	0.0
...	...
134022	0.0
134043	0.0
134264	0.0
134342	0.0
134498	0.0

[1403 rows x 76 columns]


```
[57]: df_csv[(df_csv["recipient_country"] == "China") & (df_csv["source"] ==
↳ "United_States")]
```

```
[57]:
```

	year	source	channel	recipient_isocode	recipient_country	\
2220	1990	United_States	BIL_USA	CHN	China	
4748	1991	United_States	BIL_USA	CHN	China	
4845	1991	United_States	WB_IDA	CHN	China	
7279	1992	United_States	BIL_USA	CHN	China	
7357	1992	United_States	WB_IDA	CHN	China	
...	
128626	2017	United_States	INTLNGO	CHN	China	
128680	2017	United_States	WB_IDA	CHN	China	
128732	2017	United_States	WHO	CHN	China	
134342	2018	United_States	AsDB	CHN	China	
134498	2018	United_States	WB_IDA	CHN	China	

	gbd_location_id	wb_regioncode	wb_location_id	gbd_region	\
2220	6	EAP	239	Asia, East	
4748	6	EAP	239	Asia, East	
4845	6	EAP	239	Asia, East	
7279	6	EAP	239	Asia, East	
7357	6	EAP	239	Asia, East	
...	
128626	6	EAP	239	Asia, East	
128680	6	EAP	239	Asia, East	
128732	6	EAP	239	Asia, East	
134342	6	EAP	239	Asia, East	
134498	6	EAP	239	Asia, East	

	gbd_region_id	...	other_dah_20	rmh_dah_20	nch_dah_20	ncd_dah_20	\
2220	5.0	...	180.0	0.0	0.0	0.0	
4748	5.0	...	672.0	0.0	0.0	0.0	
4845	5.0	...	711.0	447.0	1206.0	0.0	
7279	5.0	...	1306.0	0.0	0.0	0.0	
7357	5.0	...	2143.0	1165.0	3284.0	0.0	
...	
128626	5.0	...	0.0	0.0	0.0	0.0	
128680	5.0	...	0.0	0.0	0.0	0.0	
128732	5.0	...	0.0	33.0	7.0	0.0	
134342	5.0	...	126.0	0.0	0.0	0.0	
134498	5.0	...	0.0	0.0	0.0	0.0	

	hiv_dah_20	mal_dah_20	tb_dah_20	swap_hss_total_dah_20	oid_dah_20	\
2220	0.0	0.0	0.0	0.0	0.0	
4748	0.0	0.0	0.0	0.0	0.0	
4845	0.0	0.0	627.0	4003.0	627.0	
7279	0.0	0.0	0.0	0.0	0.0	

7357	0.0	0.0	1158.0		4452.0	3271.0
...
128626	0.0	0.0	0.0		472.0	0.0
128680	0.0	0.0	0.0		0.0	0.0
128732	38.0	98.0	0.0		0.0	0.0
134342	0.0	0.0	0.0		497.0	0.0
134498	0.0	0.0	0.0		0.0	0.0

	unalloc_dah_20
2220	0.0
4748	0.0
4845	0.0
7279	0.0
7357	0.0
...	...
128626	0.0
128680	0.0
128732	0.0
134342	0.0
134498	0.0

[64 rows x 76 columns]

```
[58]: df_csv[df_csv["recipient_country"].str.contains("Republic")]
```

```
[58]:
```

	year	source	channel	recipient_isocode	\
2	1990	Australia	BIL_AUS	COD	
25	1990	Australia	WB_IDA	CAF	
28	1990	Australia	WB_IDA	COD	
113	1990	Austria	WB_IDA	CAF	
116	1990	Austria	WB_IDA	COD	
...	
134455	2018	United_States	NGO	COD	
134469	2018	United_States	PAHO	DOM	
134486	2018	United_States	WB	COD	
134500	2018	United_States	WB_IDA	COD	
134557	2018	United_States	WHO	COD	

	recipient_country	gbd_location_id	wb_regioncode	\
2	Democratic Republic of the Congo	171	SSA	
25	Central African Republic	169	SSA	
28	Democratic Republic of the Congo	171	SSA	
113	Central African Republic	169	SSA	
116	Democratic Republic of the Congo	171	SSA	
...	
134455	Democratic Republic of the Congo	171	SSA	
134469	Dominican Republic	111	LAC	

134486	Democratic Republic of the Congo	171	SSA
134500	Democratic Republic of the Congo	171	SSA
134557	Democratic Republic of the Congo	171	SSA

	wb_location_id	gbd_region	gbd_region_id	...	\
2	242	Sub-Saharan Africa, Central	167.0	...	
25	242	Sub-Saharan Africa, Central	167.0	...	
28	242	Sub-Saharan Africa, Central	167.0	...	
113	242	Sub-Saharan Africa, Central	167.0	...	
116	242	Sub-Saharan Africa, Central	167.0	...	
...	
134455	242	Sub-Saharan Africa, Central	167.0	...	
134469	241	Caribbean	104.0	...	
134486	242	Sub-Saharan Africa, Central	167.0	...	
134500	242	Sub-Saharan Africa, Central	167.0	...	
134557	242	Sub-Saharan Africa, Central	167.0	...	

	other_dah_20	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	\
2	0.0	14.0	0.0	0.0	18.0	
25	0.0	0.0	0.0	0.0	0.0	
28	0.0	0.0	0.0	0.0	60.0	
113	0.0	0.0	0.0	0.0	0.0	
116	0.0	0.0	0.0	0.0	11.0	
...	
134455	491.0	16758.0	2277.0	0.0	4601.0	
134469	30.0	24.0	0.0	1.0	0.0	
134486	0.0	1335.0	694.0	0.0	0.0	
134500	794.0	2443.0	2850.0	814.0	0.0	
134557	0.0	1357.0	1124.0	0.0	0.0	

	mal_dah_20	tb_dah_20	swap_hss_total_dah_20	oid_dah_20	\
2	6.0	0.0		0.0	
25	0.0	0.0		0.0	
28	0.0	0.0		0.0	
113	0.0	0.0		0.0	
116	0.0	0.0		0.0	
...	
134455	2349.0	7911.0		4.0	260.0
134469	156.0	0.0		1.0	0.0
134486	0.0	103.0		0.0	0.0
134500	0.0	0.0		2734.0	0.0
134557	33.0	10.0		0.0	252.0

	unalloc_dah_20
2	0.0
25	0.0
28	0.0

```

113          0.0
116          0.0
...
134455        0.0
134469        0.0
134486        0.0
134500        0.0
134557        0.0

```

[3673 rows x 76 columns]

```
[59]: df_csv[df_csv["recipient_country"].str.contains("Republic")== False]
```

```
[59]:
```

	year	source	channel	recipient_isocode	recipient_country \
0	1990	Australia	BIL_AUS	CHN	China
1	1990	Australia	BIL_AUS	CIV	Cote d'Ivoire
3	1990	Australia	BIL_AUS	GHA	Ghana
4	1990	Australia	BIL_AUS	KIR	Kiribati
5	1990	Australia	BIL_AUS	MOZ	Mozambique
...
134556	2018	United_States	WHO	CMR	Cameroon
134558	2018	United_States	WHO	MLI	Mali
134559	2018	United_States	WHO	NER	Niger
134560	2018	United_States	WHO	NGA	Nigeria
134561	2018	United_States	WHO	RWA	Rwanda

	gbd_location_id	wb_regioncode	wb_location_id \
0	6	EAP	239
1	205	SSA	242
3	207	SSA	242
4	23	EAP	239
5	184	SSA	242
...
134556	202	SSA	242
134558	211	SSA	242
134559	213	SSA	242
134560	214	SSA	242
134561	185	SSA	242

	gbd_region	gbd_region_id	...	other_dah_20 \
0	Asia, East	5.0	...	38.0
1	Sub-Saharan Africa, Western	199.0	...	0.0
3	Sub-Saharan Africa, Western	199.0	...	0.0
4	Oceania	21.0	...	28.0
5	Sub-Saharan Africa, Eastern	174.0	...	0.0
...
134556	Sub-Saharan Africa, Western	199.0	...	0.0

134558	Sub-Saharan Africa, Western	199.0	...	0.0
134559	Sub-Saharan Africa, Western	199.0	...	0.0
134560	Sub-Saharan Africa, Western	199.0	...	0.0
134561	Sub-Saharan Africa, Eastern	174.0	...	0.0

	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
0	7.0	0.0	0.0	367.0	5.0	0.0	
1	5.0	0.0	0.0	18.0	1.0	0.0	
3	19.0	0.0	0.0	14.0	3.0	0.0	
4	20.0	56.0	28.0	0.0	0.0	0.0	
5	46.0	0.0	0.0	57.0	4.0	0.0	
...	
134556	18.0	28.0	0.0	0.0	7.0	1.0	
134558	172.0	49.0	0.0	9.0	56.0	1.0	
134559	34.0	40.0	0.0	0.0	5.0	3.0	
134560	3911.0	2110.0	0.0	0.0	72.0	16.0	
134561	108.0	67.0	0.0	0.0	67.0	6.0	

	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
0	130.0	0.0	0.0
1	0.0	0.0	0.0
3	0.0	0.0	0.0
4	102.0	0.0	0.0
5	0.0	0.0	0.0
...
134556	0.0	79.0	0.0
134558	0.0	17.0	0.0
134559	0.0	3.0	0.0
134560	0.0	62.0	0.0
134561	0.0	7.0	0.0

[130889 rows x 76 columns]

```
[60]: df_csv['total_diag_dah'] = df_csv[['mal_diag_dah_20', 'tb_diag_dah_20']].
      ↪sum(axis=1)
```

```
[61]: df_csv['total_diag_dah']
```

```
[61]: 0      0.0
      1      0.0
      2      0.0
      3      0.0
      4      0.0
      ...
      134557  0.0
      134558  0.0
      134559  0.0
```

```

134560    0.0
134561    0.0
Name: total_diag_dah, Length: 134562, dtype: float64

```

```
[62]: df_csv.drop("swap_hss_total_dah_20", axis=1, inplace = True)
```

```
[63]: df_csv.rename(columns = {"source": "source_country"}, inplace = True)
df_csv
```

```
[63]:
```

	year	source_country	channel	recipient_isocode	\
0	1990	Australia	BIL_AUS	CHN	
1	1990	Australia	BIL_AUS	CIV	
2	1990	Australia	BIL_AUS	COD	
3	1990	Australia	BIL_AUS	GHA	
4	1990	Australia	BIL_AUS	KIR	
...
134557	2018	United_States	WHO	COD	
134558	2018	United_States	WHO	MLI	
134559	2018	United_States	WHO	NER	
134560	2018	United_States	WHO	NGA	
134561	2018	United_States	WHO	RWA	

		recipient_country	gbd_location_id	wb_regioncode	\
0		China	6	EAP	
1		Cote d'Ivoire	205	SSA	
2	Democratic Republic of the Congo		171	SSA	
3		Ghana	207	SSA	
4		Kiribati	23	EAP	
...
134557	Democratic Republic of the Congo		171	SSA	
134558		Mali	211	SSA	
134559		Niger	213	SSA	
134560		Nigeria	214	SSA	
134561		Rwanda	185	SSA	

	wb_location_id		gbd_region	gbd_region_id	...	\
0	239		Asia, East	5.0	...	
1	242	Sub-Saharan Africa, Western		199.0	...	
2	242	Sub-Saharan Africa, Central		167.0	...	
3	242	Sub-Saharan Africa, Western		199.0	...	
4	239		Oceania	21.0	...	
...
134557	242	Sub-Saharan Africa, Central		167.0	...	
134558	242	Sub-Saharan Africa, Western		199.0	...	
134559	242	Sub-Saharan Africa, Western		199.0	...	
134560	242	Sub-Saharan Africa, Western		199.0	...	
134561	242	Sub-Saharan Africa, Eastern		174.0	...	

	other_dah_20	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	\
0	38.0	7.0	0.0	0.0	367.0	
1	0.0	5.0	0.0	0.0	18.0	
2	0.0	14.0	0.0	0.0	18.0	
3	0.0	19.0	0.0	0.0	14.0	
4	28.0	20.0	56.0	28.0	0.0	
...	
134557	0.0	1357.0	1124.0	0.0	0.0	
134558	0.0	172.0	49.0	0.0	9.0	
134559	0.0	34.0	40.0	0.0	0.0	
134560	0.0	3911.0	2110.0	0.0	0.0	
134561	0.0	108.0	67.0	0.0	0.0	

	mal_dah_20	tb_dah_20	oid_dah_20	unalloc_dah_20	total_diag_dah
0	5.0	0.0	0.0	0.0	0.0
1	1.0	0.0	0.0	0.0	0.0
2	6.0	0.0	0.0	0.0	0.0
3	3.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0
...
134557	33.0	10.0	252.0	0.0	0.0
134558	56.0	1.0	17.0	0.0	0.0
134559	5.0	3.0	3.0	0.0	0.0
134560	72.0	16.0	62.0	0.0	0.0
134561	67.0	6.0	7.0	0.0	0.0

[134562 rows x 76 columns]

```
[64]: df_csv.to_csv("ihme_dah_database.csv")
```

```
[65]: df_csv["dah_20"].mean()
```

```
[65]: 1129.0374548535249
```

```
[66]: df_csv["dah_20"].max()
```

```
[66]: 2171845.0
```

```
[67]: df_csv["dah_20"].min()
```

```
[67]: -1171.0
```

```
[68]: df_csv.shape[0]
```

```
[68]: 134562
```

```
[69]: df_csv['source_country'].unique()
```

```
[69]: array(['Australia', 'Austria', 'Belgium', 'Canada', 'Debt_repayments',
        'Denmark', 'Finland', 'France', 'Germany', 'Greece', 'Ireland',
        'Italy', 'Japan', 'Korea', 'Luxembourg', 'Netherlands',
        'New_Zealand', 'Non_OECD_DAC_countries', 'Norway', 'Other',
        'Portugal', 'Private_other', 'Spain', 'Sweden', 'Switzerland',
        'United_Kingdom', 'United_States', 'China',
        'Other_OECD_DAC_countries', 'BMGF'], dtype=object)
```

```
[70]: df_csv['source_country'].value_counts()
```

```
[70]: source_country
United_Kingdom      7844
France              7553
Italy              7151
Germany            6739
United_States      6723
Netherlands        6713
Sweden             6385
Norway             6057
Spain              5880
Denmark            5496
Canada             5408
Japan              5252
BMGF               5107
Belgium            4963
Australia          4826
Ireland            4672
Finland            4309
Austria            4305
Luxembourg         3619
Debt_repayments    3522
Switzerland        3146
Private_other      3094
Korea              2831
Portugal           2490
New_Zealand        2490
Non_OECD_DAC_countries 2285
Greece             2146
Other              1776
China              1006
Other_OECD_DAC_countries 774
Name: count, dtype: int64
```

```
[71]: df_csv.sort_values(['year'], ascending = True)
```

```
[71]:   year source_country channel recipient_isocode recipient_country \
0    1990      Australia  BIL_AUS              CHN              China
```


1620	1990	Netherlands	WB_IDA		ZMB	Zambia
1621	1990	Netherlands	WB_IDA		ZWE	Zimbabwe
1622	1990	New_Zealand	BIL_NZL		WLD	Global
1623	1990	New_Zealand	WB_IDA		AFG	Afghanistan
...
130678	2018	France	UNFPA		KEN	Kenya
130677	2018	France	UNFPA		GNB	Guinea-Bissau
130676	2018	France	UNFPA		GIN	Guinea
130674	2018	France	UNFPA		ERI	Eritrea
134561	2018	United_States	WHO		RWA	Rwanda

	gbd_location_id	wb_regioncode	wb_location_id	\
0	6	EAP	239	
1620	191	SSA	242	
1621	198	SSA	242	
1622	1	WLD	44620	
1623	160	SAS	244	
...	
130678	180	SSA	242	
130677	209	SSA	242	
130676	208	SSA	242	
130674	178	SSA	242	
134561	185	SSA	242	

	gbd_region	gbd_region_id	...	other_dah_20	\
0	Asia, East	5.0	...	38.0	
1620	Sub-Saharan Africa, Eastern	174.0	...	0.0	
1621	Sub-Saharan Africa, Southern	192.0	...	0.0	
1622	Global	1.0	...	0.0	
1623	North Africa/Middle East	138.0	...	0.0	
...	
130678	Sub-Saharan Africa, Eastern	174.0	...	0.0	
130677	Sub-Saharan Africa, Western	199.0	...	0.0	
130676	Sub-Saharan Africa, Western	199.0	...	0.0	
130674	Sub-Saharan Africa, Eastern	174.0	...	0.0	
134561	Sub-Saharan Africa, Eastern	174.0	...	0.0	

	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
0	7.0	0.0	0.0	367.0	5.0	0.0	
1620	0.0	0.0	0.0	0.0	0.0	0.0	
1621	0.0	0.0	0.0	0.0	0.0	0.0	
1622	0.0	0.0	0.0	0.0	0.0	0.0	
1623	0.0	0.0	0.0	0.0	0.0	0.0	
...	
130678	27.0	11.0	0.0	0.0	0.0	0.0	
130677	1.0	3.0	0.0	0.0	0.0	0.0	
130676	264.0	174.0	0.0	0.0	0.0	0.0	

130674	1.0	1.0	0.0	0.0	0.0	0.0
134561	108.0	67.0	0.0	0.0	67.0	6.0

	oid_dah_20	unalloc_dah_20	total_diag_dah
0	0.0	0.0	0.0
1620	0.0	0.0	0.0
1621	0.0	0.0	0.0
1622	0.0	0.0	0.0
1623	0.0	0.0	0.0
...
130678	0.0	0.0	0.0
130677	0.0	0.0	0.0
130676	0.0	0.0	0.0
130674	0.0	0.0	0.0
134561	7.0	0.0	0.0

[134562 rows x 76 columns]

```
[72]: df_csv.sort_values(['year'], ascending = False)
```

```
[72]:
```

	year	source_country	channel	recipient_isocode	recipient_country	\
134561	2018	United_States	WHO	RWA	Rwanda	
130675	2018	France	UNFPA	ETH	Ethiopia	
130677	2018	France	UNFPA	GNB	Guinea-Bissau	
130678	2018	France	UNFPA	KEN	Kenya	
130679	2018	France	UNFPA	LBR	Liberia	
...	
1624	1990	New_Zealand	WB_IDA	AGO	Angola	
1623	1990	New_Zealand	WB_IDA	AFG	Afghanistan	
1622	1990	New_Zealand	BIL_NZL	WLD	Global	
1621	1990	Netherlands	WB_IDA	ZWE	Zimbabwe	
0	1990	Australia	BIL_AUS	CHN	China	

	gbd_location_id	wb_regioncode	wb_location_id	\
134561	185	SSA	242	
130675	179	SSA	242	
130677	209	SSA	242	
130678	180	SSA	242	
130679	210	SSA	242	
...	
1624	168	SSA	242	
1623	160	SAS	244	
1622	1	WLD	44620	
1621	198	SSA	242	
0	6	EAP	239	

	gbd_region	gbd_region_id	...	other_dah_20	\
--	------------	---------------	-----	--------------	---

134561	Sub-Saharan Africa, Eastern	174.0	...	0.0
130675	Sub-Saharan Africa, Eastern	174.0	...	0.0
130677	Sub-Saharan Africa, Western	199.0	...	0.0
130678	Sub-Saharan Africa, Eastern	174.0	...	0.0
130679	Sub-Saharan Africa, Western	199.0	...	0.0
...
1624	Sub-Saharan Africa, Central	167.0	...	0.0
1623	North Africa/Middle East	138.0	...	0.0
1622	Global	1.0	...	0.0
1621	Sub-Saharan Africa, Southern	192.0	...	0.0
0	Asia, East	5.0	...	38.0

	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
134561	108.0	67.0	0.0	0.0	67.0	6.0	
130675	51.0	29.0	0.0	0.0	0.0	0.0	
130677	1.0	3.0	0.0	0.0	0.0	0.0	
130678	27.0	11.0	0.0	0.0	0.0	0.0	
130679	7.0	3.0	0.0	0.0	0.0	0.0	
...	
1624	0.0	0.0	0.0	0.0	0.0	0.0	
1623	0.0	0.0	0.0	0.0	0.0	0.0	
1622	0.0	0.0	0.0	0.0	0.0	0.0	
1621	0.0	0.0	0.0	0.0	0.0	0.0	
0	7.0	0.0	0.0	367.0	5.0	0.0	

	oid_dah_20	unalloc_dah_20	total_diag_dah
134561	7.0	0.0	0.0
130675	0.0	0.0	0.0
130677	0.0	0.0	0.0
130678	0.0	0.0	0.0
130679	0.0	0.0	0.0
...
1624	0.0	0.0	0.0
1623	0.0	0.0	0.0
1622	0.0	0.0	0.0
1621	0.0	0.0	0.0
0	0.0	0.0	0.0

[134562 rows x 76 columns]

```
[73]: df_csv.nlargest(10, 'dah_20')
```

	year	source_country	channel	recipient_isocode	recipient_country	\
134361	2018	United_States	BIL_USA	WLD	Global	
43140	2003	United_States	BIL_USA	WLD	Global	
92210	2011	United_States	BIL_USA	WLD	Global	
111285	2014	United_States	BIL_USA	WLD	Global	

66594	2007	United_States	BIL_USA	WLD	Global
117145	2015	United_States	BIL_USA	WLD	Global
92211	2011	United_States	BIL_USA	WLD	Global
111399	2014	United_States	NGO	WLD	Global
48421	2004	United_States	BIL_USA	WLD	Global
98755	2012	United_States	NGO	WLD	Global

	gbd_location_id	wb_regioncode	wb_location_id	gbd_region	\
134361	1	WLD	44620	Global	
43140	1	WLD	44620	Global	
92210	1	WLD	44620	Global	
111285	1	WLD	44620	Global	
66594	1	WLD	44620	Global	
117145	1	WLD	44620	Global	
92211	1	WLD	44620	Global	
111399	1	WLD	44620	Global	
48421	1	WLD	44620	Global	
98755	1	WLD	44620	Global	

	gbd_region_id	...	other_dah_20	rmh_dah_20	nch_dah_20	ncd_dah_20	\
134361	1.0	...	268067.0	257265.0	29398.0	817.0	
43140	1.0	...	374666.0	150730.0	0.0	0.0	
92210	1.0	...	219937.0	78056.0	13096.0	15125.0	
111285	1.0	...	74911.0	83748.0	10647.0	688.0	
66594	1.0	...	97304.0	161046.0	34546.0	1451.0	
117145	1.0	...	52544.0	75226.0	11077.0	1169.0	
92211	1.0	...	0.0	0.0	0.0	0.0	
111399	1.0	...	14120.0	92113.0	10034.0	698.0	
48421	1.0	...	349089.0	5848.0	0.0	0.0	
98755	1.0	...	6662.0	97779.0	43140.0	66.0	

	hiv_dah_20	mal_dah_20	tb_dah_20	oid_dah_20	unalloc_dah_20	\
134361	1384086.0	114445.0	36575.0	72572.0	0.0	
43140	524652.0	0.0	0.0	133643.0	0.0	
92210	486524.0	35805.0	9863.0	50127.0	0.0	
111285	574420.0	35043.0	9012.0	120692.0	0.0	
66594	605696.0	5822.0	1340.0	41942.0	0.0	
117145	537597.0	31401.0	15909.0	94336.0	0.0	
92211	817428.0	0.0	0.0	0.0	0.0	
111399	643079.0	8694.0	11130.0	4414.0	0.0	
48421	214019.0	0.0	0.0	223321.0	0.0	
98755	598883.0	10347.0	8797.0	3233.0	0.0	

	total_diag_dah
134361	10.0
43140	0.0
92210	0.0

```

111285      0.0
66594      0.0
117145     0.0
92211      0.0
111399     1502.0
48421      0.0
98755      0.0

```

```
[10 rows x 76 columns]
```

```
[74]: df_csv.nsmallest(10, 'dah_20')
```

```

[74]:      year source_country channel recipient_isocode recipient_country \
84084  2010 Private_other   GAVI                SDN             Sudan
52702  2005 Private_other   GAVI                SDN             Sudan
65185  2007 Private_other   GAVI                SDN             Sudan
75850  2009      Ireland BIL_IRL                SDN             Sudan
47200  2004 Private_other   GAVI                SDN             Sudan
42174  2003 Private_other   GAVI                SDN             Sudan
18311  1997      Canada BIL_CAN                IDN             Indonesia
32863  2001      Ireland BIL_IRL                SRB             Serbia
15556  1996      Canada BIL_CAN                IDN             Indonesia
21476  1998      Canada BIL_CAN                IDN             Indonesia

      gbd_location_id wb_regioncode  wb_location_id \
84084              522          SSA              242
52702              522          SSA              242
65185              522          SSA              242
75850              522          SSA              242
47200              522          SSA              242
42174              522          SSA              242
18311              11          EAP              239
32863              53          ECA              240
15556              11          EAP              239
21476              11          EAP              239

      gbd_region  gbd_region_id  ... other_dah_20  rmh_dah_20 \
84084 North Africa/Middle East    138.0  ...      0.0      0.0
52702 North Africa/Middle East    138.0  ...      0.0      0.0
65185 North Africa/Middle East    138.0  ...      0.0      0.0
75850 North Africa/Middle East    138.0  ...    -210.0      0.0
47200 North Africa/Middle East    138.0  ...      0.0      0.0
42174 North Africa/Middle East    138.0  ...      0.0      0.0
18311      Asia, Southeast        9.0  ...      0.0     -8.0
32863  Europe, Central         42.0  ...     -8.0      0.0
15556      Asia, Southeast        9.0  ...     -4.0      0.0
21476      Asia, Southeast        9.0  ...      0.0     -3.0

```

	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	oid_dah_20	\
84084	-1148.0	0.0	0.0	0.0	0.0	0.0	
52702	-631.0	0.0	0.0	0.0	0.0	0.0	
65185	-408.0	0.0	0.0	0.0	0.0	0.0	
75850	-2.0	0.0	-6.0	0.0	0.0	0.0	
47200	-38.0	0.0	0.0	0.0	0.0	0.0	
42174	-20.0	0.0	0.0	0.0	0.0	0.0	
18311	0.0	0.0	0.0	0.0	0.0	0.0	
32863	0.0	0.0	0.0	0.0	0.0	0.0	
15556	0.0	0.0	0.0	0.0	0.0	0.0	
21476	0.0	0.0	0.0	0.0	0.0	0.0	

	unalloc_dah_20	total_diag_dah
84084	0.0	0.0
52702	0.0	0.0
65185	0.0	0.0
75850	0.0	0.0
47200	0.0	0.0
42174	0.0	0.0
18311	0.0	0.0
32863	0.0	0.0
15556	0.0	0.0
21476	0.0	0.0

[10 rows x 76 columns]

```
[75]: df_csv[df_csv['year'] == 2009].nlargest(10, 'dah_20')
```

	year	source_country	channel	recipient_isocode	recipient_country	\
73786	2009	BMGF	GFATM	WLD	Global	
79583	2009	United_States	BIL_USA	NGA	Nigeria	
75213	2009	France	UNITAID	WLD	Global	
74405	2009	Debt_repayments	WB_IBRD	COL	Colombia	
79734	2009	United_States	NGO	WLD	Global	
74396	2009	Debt_repayments	WB_IBRD	BRA	Brazil	
73767	2009	BMGF	GAVI	WLD	Global	
77838	2009	Other	AfDB	MAR	Morocco	
79827	2009	United_States	WHO	WLD	Global	
74407	2009	Debt_repayments	WB_IBRD	DOM	Dominican Republic	

	gbd_location_id	wb_regioncode	wb_location_id	\
73786	1	WLD	44620	
79583	214	SSA	242	
75213	1	WLD	44620	
74405	125	LAC	241	
79734	1	WLD	44620	

74396	135	LAC	241
73767	1	WLD	44620
77838	148	MNA	243
79827	1	WLD	44620
74407	111	LAC	241

	gbd_region	gbd_region_id	...	other_dah_20	\
73786	Global	1.0	...	0.0	
79583	Sub-Saharan Africa, Western	199.0	...	4496.0	
75213	Global	1.0	...	198822.0	
74405	Latin America, Central	124.0	...	95076.0	
79734	Global	1.0	...	16147.0	
74396	Latin America, Tropical	134.0	...	892.0	
73767	Global	1.0	...	89878.0	
77838	North Africa/Middle East	138.0	...	84878.0	
79827	Global	1.0	...	5279.0	
74407	Caribbean	104.0	...	23368.0	

	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
73786	0.0	0.0	0.0	119837.0	71340.0	59919.0	
79583	6140.0	4491.0	7.0	185908.0	966.0	67.0	
75213	0.0	0.0	0.0	0.0	0.0	0.0	
74405	0.0	60445.0	0.0	0.0	0.0	0.0	
79734	42032.0	21643.0	0.0	68535.0	216.0	0.0	
74396	98.0	1457.0	2291.0	0.0	0.0	0.0	
73767	0.0	0.0	0.0	0.0	0.0	0.0	
77838	0.0	0.0	0.0	0.0	0.0	0.0	
79827	18249.0	5986.0	670.0	20103.0	2539.0	7524.0	
74407	0.0	0.0	0.0	0.0	0.0	0.0	

	oid_dah_20	unalloc_dah_20	total_diag_dah
73786	0.0	0.0	0.0
79583	879.0	0.0	0.0
75213	0.0	0.0	0.0
74405	0.0	0.0	0.0
79734	2905.0	0.0	0.0
74396	2386.0	0.0	0.0
73767	0.0	0.0	0.0
77838	0.0	0.0	0.0
79827	23638.0	0.0	0.0
74407	0.0	0.0	0.0

[10 rows x 76 columns]

```
[76]: df_csv.groupby('recipient_country').mean(numeric_only=True)
```

[76]:

	year	gbd_location_id	wb_location_id	\
recipient_country				
Afghanistan	2006.802297	160.0	244.0	
Albania	2006.684882	43.0	240.0	
Algeria	2007.213768	139.0	243.0	
Angola	2006.135774	168.0	242.0	
Anguilla	2000.666667	44598.0	241.0	
...	
Vietnam	2006.965581	20.0	239.0	
Wallis and Futuna Islands	2008.988235	423.0	239.0	
Yemen	2007.769365	157.0	243.0	
Zambia	2006.930862	191.0	242.0	
Zimbabwe	2006.381671	198.0	242.0	

	gbd_region_id	gbd_superregion_id	elim_ch	\
recipient_country				
Afghanistan	138.0	137.0	0.272672	
Albania	42.0	31.0	0.129326	
Algeria	138.0	137.0	0.202899	
Angola	167.0	166.0	0.280702	
Anguilla	104.0	103.0	0.083333	
...	
Vietnam	9.0	4.0	0.245851	
Wallis and Futuna Islands	21.0	4.0	0.282353	
Yemen	138.0	137.0	0.247171	
Zambia	174.0	166.0	0.317177	
Zimbabwe	192.0	166.0	0.271462	

	prelim_est	dah_20	rmh_fp_dah_20	\
recipient_country				
Afghanistan	0.012696	1181.133011	22.415357	
Albania	0.011840	275.766849	11.757741	
Algeria	0.018116	135.057971	1.072464	
Angola	0.009916	389.197559	0.467582	
Anguilla	0.000000	334.250000	0.166667	
...	
Vietnam	0.013522	1174.357099	51.279041	
Wallis and Futuna Islands	0.035294	837.517647	0.400000	
Yemen	0.013925	820.404700	30.387293	
Zambia	0.006415	549.763364	5.903778	
Zimbabwe	0.009281	717.694316	31.060325	

	rmh_mh_dah_20	...	other_dah_20	rmh_dah_20	\
recipient_country		...			
Afghanistan	112.497582	...	329.749698	177.455865	
Albania	16.215847	...	117.440801	35.722222	
Algeria	1.061594	...	83.974638	2.913043	

Angola	10.038139	...	101.154081	24.247140
Anguilla	0.000000	...	264.416667	0.416667
...
Vietnam	19.183159	...	328.055931	110.143208
Wallis and Futuna Islands	0.023529	...	685.200000	0.447059
Yemen	54.275022	...	315.196693	149.874674
Zambia	16.317177	...	93.397719	36.327156
Zimbabwe	76.727378	...	69.734919	125.693735

	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	\
recipient_country					
Afghanistan	366.124547	7.399637	4.737606	3.371826	
Albania	8.090164	11.434426	8.653005	0.000000	
Algeria	1.054348	0.644928	18.630435	0.000000	
Angola	130.674294	0.517162	27.753623	17.832952	
Anguilla	0.166667	0.000000	6.583333	0.000000	
...	
Vietnam	150.271666	17.904733	137.866626	18.077443	
Wallis and Futuna Islands	53.764706	0.341176	0.000000	0.000000	
Yemen	213.559617	5.422977	4.751958	3.330722	
Zambia	177.012117	1.371347	112.853172	22.642195	
Zimbabwe	176.654872	1.385731	233.962877	7.684455	

	tb_dah_20	oid_dah_20	unalloc_dah_20	\
recipient_country				
Afghanistan	26.067715	55.429262	0.0	
Albania	1.281421	8.136612	0.0	
Algeria	0.137681	0.235507	0.0	
Angola	1.578947	21.665904	0.0	
Anguilla	0.000000	0.000000	0.0	
...	
Vietnam	20.979717	49.065765	0.0	
Wallis and Futuna Islands	0.000000	1.070588	0.0	
Yemen	6.857267	28.463011	0.0	
Zambia	10.794726	3.725588	0.0	
Zimbabwe	6.587587	7.232019	0.0	

	total_diag_dah
recipient_country	
Afghanistan	0.041112
Albania	0.000000
Algeria	0.000000
Angola	0.007628
Anguilla	0.000000
...	...
Vietnam	0.229871
Wallis and Futuna Islands	0.000000

Yemen	0.015666
Zambia	0.336422
Zimbabwe	0.021462

[171 rows x 69 columns]

```
[77]: df_country = df_csv.groupby('recipient_country').agg({'year': ['count'],
                  'dah_20': ['mean', 'median']})
```

```
[78]: df_country
```

```
[78]:
```

	year	dah_20	
recipient_country	count	mean	median
Afghanistan	1654	1181.133011	36.0
Albania	1098	275.766849	12.0
Algeria	276	135.057971	3.0
Angola	1311	389.197559	33.0
Anguilla	12	334.250000	263.0
...
Vietnam	1627	1174.357099	142.0
Wallis and Futuna Islands	85	837.517647	7.0
Yemen	1149	820.404700	63.0
Zambia	1403	549.763364	53.0
Zimbabwe	1724	717.694316	18.0

[171 rows x 3 columns]

```
[79]: df_country.columns
```

```
[79]: MultiIndex([( 'year', 'count'),
                  ('dah_20', 'mean'),
                  ('dah_20', 'median')],
                )
```

```
[80]: df_country['dah_20']['mean'].sort_values(ascending = False)
```

```
[80]: recipient_country
```

Global	21862.702868
Mayotte	11525.857143
Russia	8324.692308
Romania	7583.630952
Argentina	6779.763158
...	...
Niue	18.423529
Antigua and Barbuda	16.750000
Tokelau	13.406250

```
Dominica                                10.255443
Saint Vincent and the Grenadines        7.303254
Name: mean, Length: 171, dtype: float64
```

```
[81]: df_pivot = df_csv.pivot_table(values='dah_20', index='recipient_country',
    columns='source_country', aggfunc='mean',
    margins=False, dropna=True, fill_value=None)
df_pivot
```

```
[81]: source_country      Australia      Austria      BMGF      Belgium \
recipient_country
Afghanistan      430.246154      141.616667      2485.985294      126.036364
Albania           11.875000      326.661290      290.888889      33.516129
Algeria           1.500000      44.000000       8.888889      284.187500
Angola           144.642857      55.096774      745.813559      66.217391
Anguilla          NaN          NaN          NaN          NaN
...
Vietnam          971.573333      1681.618182      724.819672      345.825397
Wallis and Futuna Islands      46.777778          NaN          NaN          NaN
Yemen            174.205882      254.371429      915.293103      188.645161
Zambia           123.244898      69.219512      1114.866667      78.862069
Zimbabwe         388.241935      64.500000      787.855072      109.707692
```

```
source_country      Canada      China      Debt_repayments \
recipient_country
Afghanistan      863.969231      98.833333      202.227273
Albania          82.812500      3.100000      489.553191
Algeria          7.833333      0.000000      0.000000
Angola          211.340000      54.875000      229.953488
Anguilla         66.000000          NaN          NaN
...
Vietnam          511.811321      423.533333      6703.875000
Wallis and Futuna Islands      15.666667          NaN          NaN
Yemen           429.000000      829.777778      7618.833333
Zambia           427.838235      19.750000      602.000000
Zimbabwe         180.694444      1.785714      653.237288
```

```
source_country      Denmark      Finland      France ... \
recipient_country
Afghanistan      328.634146      145.704918      898.333333 ...
Albania          25.170732      14.173913      206.000000 ...
Algeria          30.545455      18.000000      332.148148 ...
Angola          68.912281      109.806452      790.051282 ...
Anguilla         NaN          NaN          NaN ...
...
Vietnam          214.200000      427.971429      1519.870588 ...
Wallis and Futuna Islands      NaN      3.000000      4638.866667 ...
```

Yemen	143.789474	42.527778	524.383333	...
Zambia	854.725806	53.723404	423.146667	...
Zimbabwe	585.739726	55.781818	793.719512	...

source_country	Norway	Other	Other_OECD_DAC_countries	\
recipient_country				
Afghanistan	590.038961	518.722222		1.111111
Albania	154.037037	22.571429		0.375000
Algeria	42.454545	0.000000		0.000000
Angola	494.910448	562.642857		3.333333
Anguilla	NaN	NaN		NaN
...
Vietnam	379.112903	5083.291667		19.400000
Wallis and Futuna Islands	35.500000	NaN		NaN
Yemen	540.500000	727.444444		38.500000
Zambia	613.027027	791.333333		5.333333
Zimbabwe	837.510638	237.400000		2.000000

source_country	Portugal	Private_other	Spain	Sweden	\
recipient_country					
Afghanistan	44.177778	110.526316	587.000000	415.825581	
Albania	4.437500	86.973684	102.829268	154.523810	
Algeria	7.333333	4.833333	98.416667	70.333333	
Angola	309.000000	50.966667	281.543860	417.203704	
Anguilla	NaN	NaN	NaN	NaN	
...	
Vietnam	29.454545	188.594595	550.471264	814.244186	
Wallis and Futuna Islands	NaN	6.333333	NaN	NaN	
Yemen	22.521739	89.137931	178.166667	146.209677	
Zambia	9.142857	228.620690	137.250000	1073.984848	
Zimbabwe	24.333333	31.739130	345.746032	741.618421	

source_country	Switzerland	United_Kingdom	United_States
recipient_country			
Afghanistan	155.878788	1360.367089	7809.011494
Albania	612.153846	191.846154	550.483333
Algeria	33.166667	51.400000	119.428571
Angola	99.771429	506.625000	1151.448276
Anguilla	NaN	358.636364	NaN
...
Vietnam	404.000000	1139.212500	2343.901639
Wallis and Futuna Islands	NaN	1.000000	30.000000
Yemen	150.590909	1555.478261	1990.895833
Zambia	65.864865	778.510638	1569.981132
Zimbabwe	325.465116	3076.683168	2624.721519

[171 rows x 30 columns]

```
[82]: df_pivot.index
```

```
[82]: Index(['Afghanistan', 'Albania', 'Algeria', 'Angola', 'Anguilla',
          'Antigua and Barbuda', 'Argentina', 'Armenia', 'Azerbaijan', 'Bahrain',
          ...,
          'Ukraine', 'Uruguay', 'Uzbekistan', 'Vanuatu', 'Venezuela', 'Vietnam',
          'Wallis and Futuna Islands', 'Yemen', 'Zambia', 'Zimbabwe'],
          dtype='object', name='recipient_country', length=171)
```

```
[83]: df_pivot.columns
```

```
[83]: Index(['Australia', 'Austria', 'BMGF', 'Belgium', 'Canada', 'China',
          'Debt_repayments', 'Denmark', 'Finland', 'France', 'Germany', 'Greece',
          'Ireland', 'Italy', 'Japan', 'Korea', 'Luxembourg', 'Netherlands',
          'New_Zealand', 'Non_OECD_DAC_countries', 'Norway', 'Other',
          'Other_OECD_DAC_countries', 'Portugal', 'Private_other', 'Spain',
          'Sweden', 'Switzerland', 'United_Kingdom', 'United_States'],
          dtype='object', name='source_country')
```

```
[84]: df_pivot = df_csv.pivot_table(values='dah_20', index=['year',
          ↪ 'recipient_country'], columns='source_country', aggfunc='mean',
          margins=False, dropna=True, fill_value=None)
```

```
[85]: df_pivot
```

```
[85]: source_country      Australia  Austria  BMGF  Belgium \
year recipient_country
1990 Afghanistan      0.000000      0.0    NaN  0.000000
      Albania          0.000000      0.0    NaN  0.000000
      Algeria          NaN        NaN    NaN      NaN
      Angola          0.000000      0.0    NaN  0.000000
      Argentina        NaN        NaN    NaN      NaN
...
2018 Vietnam          185.666667      NaN  1080.0  369.500000
      Wallis and Futuna Islands      NaN      NaN    NaN      NaN
      Yemen          362.333333    434.0  3825.0  657.666667
      Zambia          76.000000     30.0  2819.6  29.500000
      Zimbabwe         22.666667    159.5   775.0  24.500000

source_country      Canada  China  Debt_repayments \
year recipient_country
1990 Afghanistan      0.000000      NaN      0.0
      Albania          0.000000      NaN      0.0
      Algeria          NaN        NaN      0.0
      Angola          30.500000      NaN      0.0
      Argentina        NaN        NaN      0.0
...
```

2018	Vietnam	373.833333	305.5	NaN
	Wallis and Futuna Islands	45.000000	NaN	NaN
	Yemen	628.000000	687.5	35365.0
	Zambia	436.200000	NaN	NaN
	Zimbabwe	207.500000	0.0	190.0
source_country				
		Denmark	Finland	France ... \
year recipient_country				
1990	Afghanistan	0.000000	0.000000	0.000000 ...
	Albania	0.000000	0.000000	0.000000 ...
	Algeria	NaN	NaN	NaN ...
	Angola	0.000000	91.500000	475.500000 ...
	Argentina	NaN	NaN	NaN ...
...				
2018	Vietnam	193.000000	163.000000	1324.000000 ...
	Wallis and Futuna Islands	NaN	NaN	NaN ...
	Yemen	142.400000	32.333333	1245.166667 ...
	Zambia	37.500000	NaN	378.250000 ...
	Zimbabwe	169.333333	200.000000	1417.333333 ...
source_country				
		Norway	Other	Other_OECD_DAC_countries \
year recipient_country				
1990	Afghanistan	0.0	NaN	NaN
	Albania	0.0	NaN	NaN
	Algeria	NaN	NaN	NaN
	Angola	0.0	NaN	NaN
	Argentina	NaN	NaN	NaN
...				
2018	Vietnam	265.5	248.0	3.0
	Wallis and Futuna Islands	NaN	NaN	NaN
	Yemen	NaN	NaN	72.5
	Zambia	72.0	NaN	NaN
	Zimbabwe	15.5	NaN	0.0
source_country				
		Portugal	Private_other	Spain \
year recipient_country				
1990	Afghanistan	NaN	NaN	NaN
	Albania	NaN	NaN	NaN
	Algeria	NaN	NaN	NaN
	Angola	NaN	NaN	NaN
	Argentina	NaN	NaN	NaN
...				
2018	Vietnam	0.0	174.75	241.800000
	Wallis and Futuna Islands	NaN	NaN	NaN
	Yemen	27.5	132.00	464.800000
	Zambia	NaN	222.00	50.000000
	Zimbabwe	113.0	9.00	34.333333

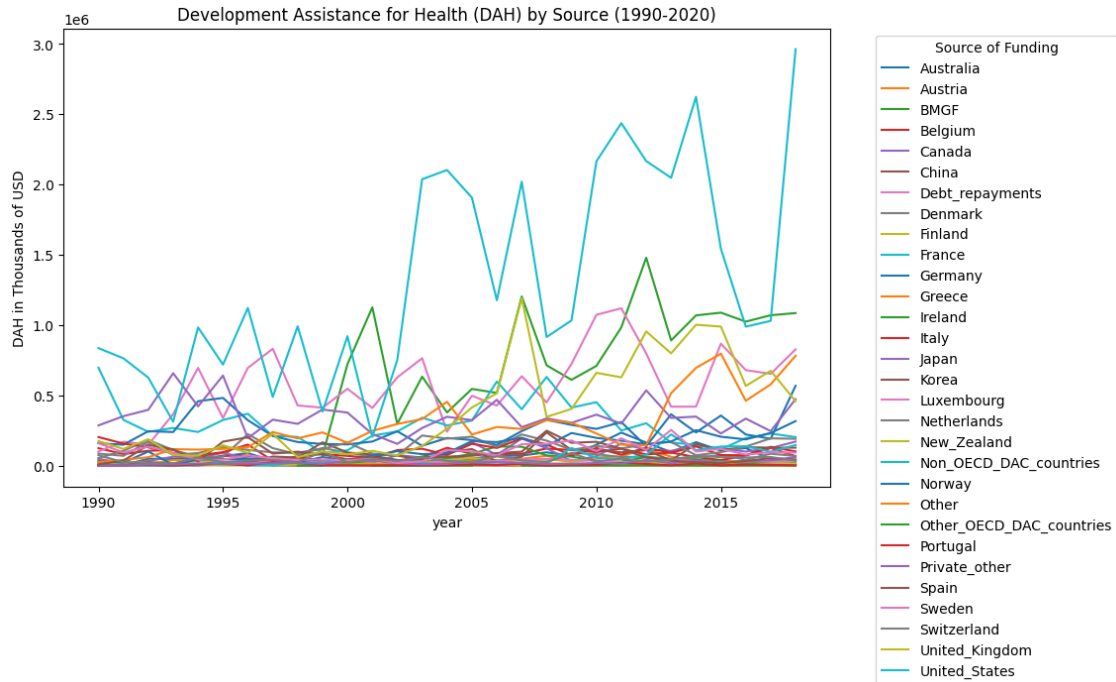
source_country	Sweden	Switzerland	United_Kingdom	\
year recipient_country				
1990 Afghanistan	2640.500000	193.0	0.000000	
Albania	0.000000	NaN	0.000000	
Algeria	NaN	NaN	NaN	
Angola	0.000000	6.0	0.000000	
Argentina	473.500000	NaN	NaN	
...	
2018 Vietnam	526.500000	NaN	699.500000	
Wallis and Futuna Islands	NaN	NaN	NaN	
Yemen	561.000000	120.0	182.600000	
Zambia	191.333333	2.0	289.333333	
Zimbabwe	129.333333	1084.0	233.666667	

source_country	United_States
year recipient_country	
1990 Afghanistan	29265.500000
Albania	0.000000
Algeria	2.000000
Angola	609.000000
Argentina	4821.000000
...	...
2018 Vietnam	2409.666667
Wallis and Futuna Islands	NaN
Yemen	3861.500000
Zambia	3724.000000
Zimbabwe	815.333333

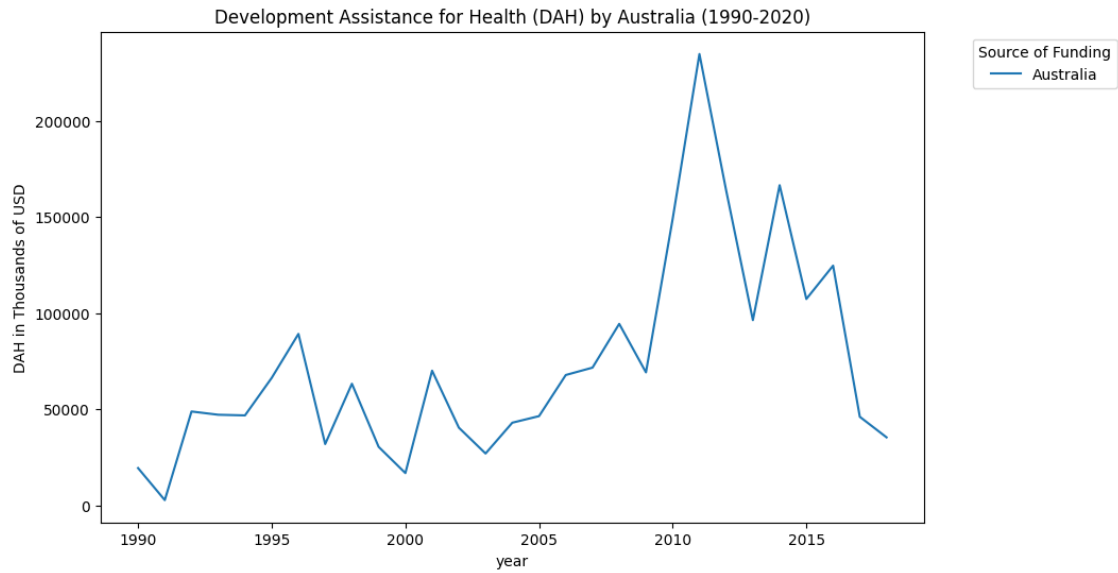
[4456 rows x 30 columns]

```
[86]: import matplotlib.pyplot as plt
      %matplotlib inline
```

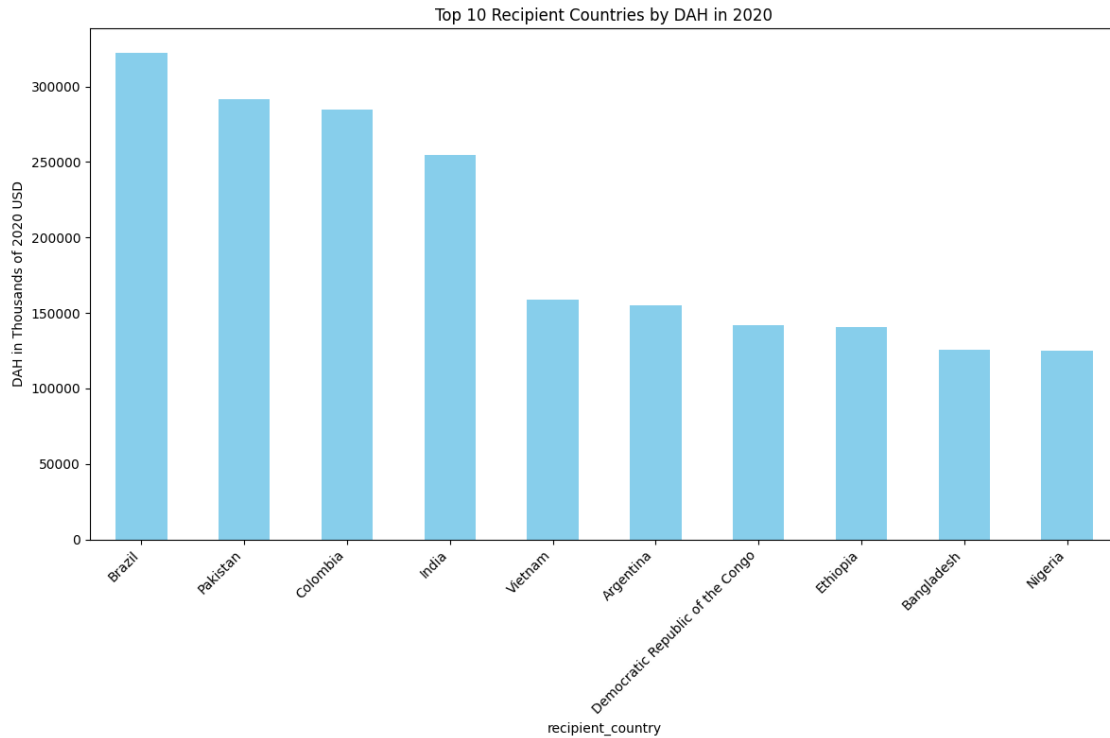
```
[87]: dah_by_source = df_csv.pivot_table(values='dah_20', index='year',
      ↪columns='source_country', aggfunc='sum')
      dah_by_source.plot(kind='line', figsize=(10, 6))
      plt.ylabel('DAH in Thousands of USD')
      plt.title('Development Assistance for Health (DAH) by Source (1990-2020)')
      plt.legend(title='Source of Funding', bbox_to_anchor=(1.05, 1), loc='upper_
      ↪left')
      plt.show()
```



```
[88]: dah_by_source = df_csv[df_csv['source_country'] == 'Australia'].
      ↪pivot_table(values='dah_20', index='year', columns='source_country',
      ↪aggfunc='sum')
      dah_by_source.plot(kind='line', figsize=(10, 6))
      plt.ylabel('DAH in Thousands of USD')
      plt.title('Development Assistance for Health (DAH) by Australia (1990-2020)')
      plt.legend(title='Source of Funding', bbox_to_anchor=(1.05, 1), loc='upper_
      ↪left')
      plt.show()
```

```
[89]: df_2020 = df_csv[df_csv['year'] == 2010]
dah_by_country = df_2020[df_2020['recipient_country'] != 'Global'].
    ↳groupby('recipient_country')['dah_20'].sum()
dah_by_country = dah_by_country.sort_values(ascending=False).head(10)
dah_by_country.plot(kind='bar', figsize=(12, 8), color='skyblue')
plt.ylabel('DAH in Thousands of 2020 USD')
plt.title('Top 10 Recipient Countries by DAH in 2020')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



```
[90]: df_csv_2 = pd.read_csv('IHME_DAH_DATABASE_1990_2016_Y2017M04D19.csv',
    ↪ low_memory=False)
```

```
[91]: df_csv_2.head()
```

```
[91]:
```

	year	source	channel	recipient_isocode	recipient_country \
0	1990	Austria	WB_IDA	AFG	Afghanistan
1	1990	United_States	BIL_USA	AFG	Afghanistan
2	1990	Ireland	WB_IDA	AFG	Afghanistan
3	1990	Netherlands	WB_IDA	AFG	Afghanistan
4	1990	United_Kingdom	WB_IDA	AFG	Afghanistan

	gbd_location_id	wb_regioncode	wb_location_id	gbd_region	gbd_region_id \
0	160	SAS	244	Asia, South	138.0
1	160	SAS	244	Asia, South	138.0
2	160	SAS	244	Asia, South	138.0
3	160	SAS	244	Asia, South	138.0
4	160	SAS	244	Asia, South	138.0

	...	other_dah_15	hiv_dah_15	mal_dah_15	mh_dah_15	ch_dah_15	tb_dah_15 \
0	...	0	0	0	0	0	0
1	...	250009	0	0	0	0	0
2	...	0	0	0	0	0	0

3	...	0	0	0	0	0	0
4	...	0	0	0	0	0	0

	oid_dah_15	ncd_dah_15	unalloc_dah_15	prelim_est
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

[5 rows x 58 columns]

```
[92]: df_csv_2.rename(columns = {"source": "source_country"}, inplace = True)
df_csv_2
```

```
[92]:
```

	year	source_country	channel	recipient_isocode	\
0	1990	Austria	WB_IDA	AFG	
1	1990	United_States	BIL_USA	AFG	
2	1990	Ireland	WB_IDA	AFG	
3	1990	Netherlands	WB_IDA	AFG	
4	1990	United_Kingdom	WB_IDA	AFG	
...	
495723	2016	Spain	GFATM	QZA	
495724	2016	Korea	GFATM	QZA	
495725	2016	Other	GFATM	QZA	
495726	2016	Portugal	GFATM	QZA	
495727	2016	Other	GFATM	QZA	

	recipient_country	gbd_location_id	wb_regioncode	\
0	Afghanistan	160	SAS	
1	Afghanistan	160	SAS	
2	Afghanistan	160	SAS	
3	Afghanistan	160	SAS	
4	Afghanistan	160	SAS	
...	
495723	Unallocated/Unspecified	44598	NaN	
495724	Unallocated/Unspecified	44598	NaN	
495725	Unallocated/Unspecified	44598	NaN	
495726	Unallocated/Unspecified	44598	NaN	
495727	Unallocated/Unspecified	44598	NaN	

	wb_location_id	gbd_region	gbd_region_id	...	\
0	244	Asia, South	138.0	...	
1	244	Asia, South	138.0	...	
2	244	Asia, South	138.0	...	
3	244	Asia, South	138.0	...	
4	244	Asia, South	138.0	...	

...
495723	44621	Unallocated/Unspecified			NaN	...	
495724	44621	Unallocated/Unspecified			NaN	...	
495725	44621	Unallocated/Unspecified			NaN	...	
495726	44621	Unallocated/Unspecified			NaN	...	
495727	44621	Unallocated/Unspecified			NaN	...	

	other_dah_15	hiv_dah_15	mal_dah_15	mh_dah_15	ch_dah_15	tb_dah_15	\
0	0	0	0	0	0	0	
1	250009	0	0	0	0	0	
2	0	0	0	0	0	0	
3	0	0	0	0	0	0	
4	0	0	0	0	0	0	

...
495723	0	10881	0	0	0	221	
495724	0	639	0	0	0	370	
495725	0	20732	0	0	0	68831	
495726	0	16069	2243	0	0	1444	
495727	0	7231	0	0	0	6756	

	oid_dah_15	ncd_dah_15	unalloc_dah_15	prelim_est
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

...
495723	0	0	0	1
495724	0	0	0	1
495725	0	0	40423	1
495726	0	0	0	1
495727	0	0	0	1

[495728 rows x 58 columns]

```
[93]: #I had to reduce the datasets, because trying to merge the whole datasets
      ↪resulted in Memory Error
df_csv_red = df_csv[df_csv['year'].isin([2001, 2002, 2003])]
df_csv_2_red = df_csv_2[df_csv_2['year'].isin([2001, 2002, 2003])]

df_all = pd.merge(df_csv_red, df_csv_2_red, on=['year', 'source_country',
      ↪'channel', 'recipient_country', 'gbd_region'], how='inner')
```

```
[94]: df_all.head()
```

```
[94]:   year source_country channel recipient_isocode_x recipient_country \
0  2001      Australia  BIL_AUS                ARM          Armenia
```

1	2001	Australia	BIL_AUS	AZE	Azerbaijan
2	2001	Australia	BIL_AUS	BGD	Bangladesh
3	2001	Australia	BIL_AUS	BTN	Bhutan
4	2001	Australia	BIL_AUS	GEO	Georgia

	gbd_location_id_x	wb_regioncode_x	wb_location_id_x	gbd_region	\
0	33	ECA	240	Asia, Central	
1	34	ECA	240	Asia, Central	
2	161	SAS	244	Asia, South	
3	162	SAS	244	Asia, South	
4	35	ECA	240	Asia, Central	

	gbd_region_id_x	...	other_dah_15	hiv_dah_15	mal_dah_15	mh_dah_15	\
0	32.0	...	0	4457	0	0	
1	32.0	...	0	2231	0	0	
2	159.0	...	0	24510	0	0	
3	159.0	...	0	6	0	0	
4	32.0	...	0	7739	0	0	

	ch_dah_15	tb_dah_15	oid_dah_15	ncd_dah_15	unalloc_dah_15	prelim_est_y
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0

[5 rows x 129 columns]

```
[95]: df_all.shape
```

```
[95]: (12924, 129)
```

```
[96]: df_all_1 = df_all.iloc[:14000,:]
df_all_2 = df_all.iloc[14000:,:]
```

```
[97]: df_all_new = pd.concat([df_all_1, df_all_2], axis = 0)
df_all_new.shape
```

```
[97]: (12924, 129)
```

```
[98]: print(df_csv.columns)
```

```
Index(['year', 'source_country', 'channel', 'recipient_isocode',
      'recipient_country', 'gbd_location_id', 'wb_regioncode',
      'wb_location_id', 'gbd_region', 'gbd_region_id', 'gbd_superregion',
      'gbd_superregion_id', 'elim_ch', 'prelim_est', 'dah_20',
      'rmh_fp_dah_20', 'rmh_mh_dah_20', 'rmh_hss_other_dah_20',
      'rmh_hss_hrh_dah_20', 'rmh_other_dah_20', 'nch_cnn_dah_20',
```

```

'nch_cnv_dah_20', 'nch_other_dah_20', 'nch_hss_other_dah_20',
'nch_hss_hrh_dah_20', 'hiv_treat_dah_20', 'hiv_prev_dah_20',
'hiv_pmtct_dah_20', 'hiv_other_dah_20', 'hiv_ct_dah_20',
'hiv_ovc_dah_20', 'hiv_care_dah_20', 'hiv_hss_other_dah_20',
'hiv_hss_hrh_dah_20', 'hiv_amr_dah_20', 'mal_diag_dah_20',
'mal_hss_other_dah_20', 'mal_hss_hrh_dah_20', 'mal_con_nets_dah_20',
'mal_con_irs_dah_20', 'mal_con_oth_dah_20', 'mal_treat_dah_20',
'mal_comm_con_dah_20', 'mal_other_dah_20', 'mal_amr_dah_20',
'tb_other_dah_20', 'tb_treat_dah_20', 'tb_diag_dah_20',
'tb_hss_other_dah_20', 'tb_hss_hrh_dah_20', 'tb_amr_dah_20',
'oid_hss_other_dah_20', 'oid_hss_hrh_dah_20', 'oid_ebz_dah_20',
'oid_zika_dah_20', 'oid_covid_dah_20', 'oid_other_dah_20',
'oid_amr_dah_20', 'ncd_hss_other_dah_20', 'ncd_hss_hrh_dah_20',
'ncd_tobac_dah_20', 'ncd_mental_dah_20', 'ncd_other_dah_20',
'swap_hss_other_dah_20', 'swap_hss_hrh_dah_20', 'swap_hss_pp_dah_20',
'other_dah_20', 'rmh_dah_20', 'nch_dah_20', 'ncd_dah_20', 'hiv_dah_20',
'mal_dah_20', 'tb_dah_20', 'oid_dah_20', 'unalloc_dah_20',
'total_diag_dah'],
dtype='object')

```

```

[99]: df_csv['other_diseases_dah_20'] = ( df_csv['oid_other_dah_20'] +
↳df_csv['tb_other_dah_20'] + df_csv['mal_other_dah_20'] +
↳df_csv['hiv_other_dah_20'] + df_csv['ncd_other_dah_20'])

```

```

[100]: df_csv["dah_20_round"] = df_csv["dah_20"].round(decimals = 2)

```

```

[101]: Years = [2000, 2005, 2010]

```

```

[102]: df_csv['Years_2000_2005_2010'] = df_csv['year'].apply(lambda x: x in Years)
df_csv[df_csv['Years_2000_2005_2010'] == True]

```

```

[102]:
   year source_country channel recipient_isocode recipient_country \
27809  2000      Australia   AsDB              IDN      Indonesia
27810  2000      Australia   AsDB              IDN      Indonesia
27811  2000      Australia   AsDB              KHM      Cambodia
27812  2000      Australia   AsDB              LAO        Laos
27813  2000      Australia   AsDB              MHL  Marshall Islands
...    ...           ...      ...              ...           ...
85483  2010  United_States   WHO                SUR      Suriname
85484  2010  United_States   WHO                TUR      Turkey
85485  2010  United_States   WHO                UKR      Ukraine
85486  2010  United_States   WHO                VNM      Vietnam
85487  2010  United_States   WHO                WLD      Global

   gbd_location_id wb_regioncode  wb_location_id \
27809              11          EAP             239
27810              11          EAP             239
27811              10          EAP             239

```

27812	12	EAP	239
27813	24	EAP	239
...
85483	118	LAC	241
85484	155	ECA	240
85485	63	ECA	240
85486	20	EAP	239
85487	1	WLD	44620

	gbd_region	gbd_region_id	...	ncd_dah_20	hiv_dah_20	\
27809	Asia, Southeast	9.0	...	0.0	0.0	
27810	Asia, Southeast	9.0	...	0.0	0.0	
27811	Asia, Southeast	9.0	...	0.0	0.0	
27812	Asia, Southeast	9.0	...	0.0	0.0	
27813	Oceania	21.0	...	0.0	0.0	
...	
85483	Caribbean	104.0	...	0.0	0.0	
85484	North Africa/Middle East	138.0	...	0.0	0.0	
85485	Europe, Eastern	56.0	...	0.0	0.0	
85486	Asia, Southeast	9.0	...	0.0	113.0	
85487	Global	1.0	...	0.0	20106.0	

	mal_dah_20	tb_dah_20	oid_dah_20	unalloc_dah_20	total_diag_dah	\
27809	0.0	0.0	179.0	0.0	0.0	
27810	0.0	0.0	0.0	0.0	0.0	
27811	0.0	0.0	0.0	0.0	0.0	
27812	0.0	0.0	0.0	0.0	0.0	
27813	0.0	0.0	0.0	0.0	0.0	
...	
85483	0.0	0.0	1.0	0.0	0.0	
85484	0.0	0.0	204.0	0.0	0.0	
85485	0.0	0.0	217.0	0.0	0.0	
85486	38.0	68.0	1094.0	0.0	0.0	
85487	5158.0	34725.0	2931.0	0.0	0.0	

	other_diseases_dah_20	dah_20_round	Years_2000_2005_2010
27809	179.0	333.0	True
27810	0.0	25.0	True
27811	0.0	106.0	True
27812	0.0	17.0	True
27813	0.0	22.0	True
...
85483	0.0	1.0	True
85484	0.0	204.0	True
85485	0.0	217.0	True
85486	0.0	3154.0	True
85487	31216.0	101993.0	True

[14502 rows x 79 columns]

```
[103]: for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
                                     chunksize = 50000):
        print("CHUNK DF")
        print(chunk_df.head())
```

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\554763948.py:1: DtypeWarning: Columns (29,32,34,35,36,37,38,40,41,42,44,46,47,48,49,50,51,52,60,65) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

CHUNK DF

	year	source	channel	recipient_isocode	recipient_country	\
0	1990	Australia	BIL_AUS	AGO	Angola	
1	1990	Australia	BIL_AUS	BDI	Burundi	
2	1990	Australia	BIL_AUS	BEN	Benin	
3	1990	Australia	BIL_AUS	BFA	Burkina Faso	
4	1990	Australia	BIL_AUS	BWA	Botswana	

	gbd_location_id	wb_regioncode	wb_location_id	\
0	168	SSA	242	
1	175	SSA	242	
2	200	SSA	242	
3	201	SSA	242	
4	193	SSA	242	

	gbd_region	gbd_region_id	...	other_dah_20	rmh_dah_20	\
0	Sub-Saharan Africa, Central	167	...	0	5	
1	Sub-Saharan Africa, Eastern	174	...	0	6	
2	Sub-Saharan Africa, Western	199	...	0	6	
3	Sub-Saharan Africa, Western	199	...	0	5	
4	Sub-Saharan Africa, Southern	192	...	0	1	

	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
0	0	0	7	3	0	
1	0	0	5	1	0	
2	0	0	5	2	0	
3	0	0	7	2	0	
4	0	0	23	-	0	

	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20	
0	0	0	-	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	-	

[5 rows x 76 columns]

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\554763948.py:1: DtypeWarning: Columns (35,38,39,40,42,44,46,47,51,53,57,65) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

CHUNK DF

	year	source	channel	recipient_isocode	recipient_country	\
50000	1999	Sweden	UNFPA	STP	Sao Tome and Principe	
50001	1999	Sweden	UNFPA	SWZ	Swaziland	
50002	1999	Sweden	UNFPA	SYC	Seychelles	
50003	1999	Sweden	UNFPA	TCD	Chad	
50004	1999	Sweden	UNFPA	TGO	Togo	

	gbd_location_id	wb_regioncode	wb_location_id	\
50000	215	SSA	242	
50001	197	SSA	242	
50002	186	SSA	242	
50003	204	SSA	242	
50004	218	SSA	242	

	gbd_region	gbd_region_id	...	other_dah_20	\
50000	Sub-Saharan Africa, Western	199.0	...	0	
50001	Sub-Saharan Africa, Southern	192.0	...	0	
50002	Asia, Southeast	9.0	...	0	
50003	Sub-Saharan Africa, Western	199.0	...	0	
50004	Sub-Saharan Africa, Western	199.0	...	0	

	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
50000	-	0	0	-	0	0	
50001	-	0	0	1	0	0	
50002	0	0	0	-	0	0	
50003	2	0	0	1	0	0	
50004	-	0	0	-	0	0	

	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
50000	0	0	-
50001	0	0	0
50002	0	0	0
50003	0	0	0
50004	0	0	0

[5 rows x 76 columns]

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\554763948.py:1: DtypeWarning: Columns (15,53,57,59,61) have mixed types. Specify dtype option on import or set low_memory=False.

```

for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
CHUNK DF
      year      source channel recipient_isocode recipient_country \
100000  2004  Private_other      NGO              LBY              Libya
100001  2004  Private_other      NGO              LCA              Saint Lucia
100002  2004  Private_other      NGO              LKA              Sri Lanka
100003  2004  Private_other      NGO              LSO              Lesotho
100004  2004  Private_other      NGO              MAR              Morocco

      gbd_location_id wb_regioncode  wb_location_id \
100000              147            MNA            243
100001              116            LAC            241
100002               17            SAS            244
100003              194            SSA            242
100004              148            MNA            243

      gbd_region  gbd_region_id  ... other_dah_20 \
100000  North Africa/Middle East      138.0  ...      -
100001              Caribbean      104.0  ...      1
100002      Asia, Southeast         9.0  ...      8
100003  Sub-Saharan Africa, Southern      192.0  ...      1
100004  North Africa/Middle East      138.0  ...     10

      rmh_dah_20  nch_dah_20  ncd_dah_20  hiv_dah_20  mal_dah_20  tb_dah_20 \
100000          -          -           0           -           0           0
100001          0          0           0           0           0           0
100002          -          1           0           2           0           0
100003          1          -           0           3           0           0
100004          8          1           0           -           0           0

      swap_hss_total_dah_20  oid_dah_20  unalloc_dah_20
100000                    0           0           -
100001                    0           0           0
100002                    0           0           -
100003                    3           0           0
100004                    0           0           0

```

[5 rows x 76 columns]

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\554763948.py:1: DtypeWarning: Columns (35,48,53,57,60,61) have mixed types. Specify dtype option on import or set low_memory=False.

```

for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
CHUNK DF
      year      source channel recipient_isocode recipient_country \
150000  2007  Netherlands      WHO              WLD              Global
150001  2007  Netherlands      WHO              YEM              Yemen

```

150002	2007	Netherlands	WHO	ZAF	South Africa
150003	2007	Netherlands	WHO	ZMB	Zambia
150004	2007	Netherlands	WHO	ZWE	Zimbabwe

	gbd_location_id	wb_regioncode	wb_location_id	\
150000	1	WLD	44620	
150001	157	MNA	243	
150002	196	SSA	242	
150003	191	SSA	242	
150004	198	SSA	242	

	gbd_region	gbd_region_id	...	other_dah_20	\
150000	Global	1.0	...	214	
150001	North Africa/Middle East	138.0	...	0	
150002	Sub-Saharan Africa, Southern	192.0	...	0	
150003	Sub-Saharan Africa, Eastern	174.0	...	0	
150004	Sub-Saharan Africa, Southern	192.0	...	0	

	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
150000	0	0	0	640	573	573	
150001	0	0	0	0	0	0	
150002	0	0	0	0	-	0	
150003	0	0	0	0	3	0	
150004	0	0	0	0	2	0	

	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
150000	0	0	-
150001	-	0	0
150002	1	0	0
150003	-	0	0
150004	-	0	0

[5 rows x 76 columns]

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\554763948.py:1: DtypeWarning: Columns (36,53,57,60) have mixed types. Specify dtype option on import or set low_memory=False.

for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',

CHUNK DF

	year	source	channel	recipient_isocode	\
200000	2009	United_States	INTLNGO	VCT	
200001	2009	United_States	INTLNGO	VEN	
200002	2009	United_States	INTLNGO	VEN	
200003	2009	United_States	INTLNGO	VNM	
200004	2009	United_States	INTLNGO	VNM	

	recipient_country	gbd_location_id	wb_regioncode	\
200000	Saint Vincent and the Grenadines	117	LAC	

200001	Venezuela	133	LAC
200002	Venezuela	133	LAC
200003	Vietnam	20	EAP
200004	Vietnam	20	EAP

	wb_location_id	gbd_region	gbd_region_id	...	\
200000	241	Caribbean	104.0	...	
200001	241	Latin America, Central	124.0	...	
200002	241	Latin America, Central	124.0	...	
200003	239	Asia, Southeast	9.0	...	
200004	239	Asia, Southeast	9.0	...	

	other_dah_20	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	\
200000	0	-	-	0	2	0	
200001	0	0	0	-	0	7	
200002	0	0	204	0	0	0	
200003	0	0	0	0	9	13	
200004	0	4	3	0	42	0	

	tb_dah_20	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
200000	0		0	-
200001	0		6	-
200002	0		0	-
200003	7		0	-
200004	7		0	0

[5 rows x 76 columns]

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\554763948.py:1: DtypeWarning: Columns (59,61) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

CHUNK DF

	year	source	channel	recipient_isocode	recipient_country	\
250000	2012	Germany	GFATM	GIN	Guinea	
250001	2012	Germany	GFATM	GMB	The Gambia	
250002	2012	Germany	GFATM	GMB	The Gambia	
250003	2012	Germany	GFATM	GNB	Guinea-Bissau	
250004	2012	Germany	GFATM	GRD	Grenada	

	gbd_location_id	wb_regioncode	wb_location_id	\
250000	208	SSA	242	
250001	206	SSA	242	
250002	206	SSA	242	
250003	209	SSA	242	
250004	112	LAC	241	

	gbd_region	gbd_region_id	...	other_dah_20	\
--	------------	---------------	-----	--------------	---

250000	Sub-Saharan Africa, Western	199.0	...	0
250001	Sub-Saharan Africa, Western	199.0	...	0
250002	Sub-Saharan Africa, Western	199.0	...	0
250003	Sub-Saharan Africa, Western	199.0	...	0
250004	Caribbean	104.0	...	0

	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
250000	0	0	0	70	1282	0	
250001	0	0	0	326	279	82	
250002	0	0	0	0	65	0	
250003	0	0	0	93	18	57	
250004	0	0	0	32	0	0	

	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
250000	0	0	-
250001	0	0	-
250002	0	0	-
250003	0	0	-
250004	0	0	-

[5 rows x 76 columns]

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\554763948.py:1: DtypeWarning: Columns (15,19,54,59) have mixed types. Specify dtype option on import or set low_memory=False.

for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',

CHUNK DF

	year	source	channel	recipient_isocode	recipient_country	\
300000	2014	Norway	WHO	SSD	South Sudan	
300001	2014	Norway	WHO	STP	Sao Tome and Principe	
300002	2014	Norway	WHO	SWZ	Swaziland	
300003	2014	Norway	WHO	SYR	Syria	
300004	2014	Norway	WHO	TCD	Chad	

	gbd_location_id	wb_regioncode	wb_location_id	\
300000	435	SSA	242	
300001	215	SSA	242	
300002	197	SSA	242	
300003	153	MNA	243	
300004	204	SSA	242	

	gbd_region	gbd_region_id	...	other_dah_20	\
300000	Sub-Saharan Africa, Eastern	174	...	0	
300001	Sub-Saharan Africa, Western	199	...	0	
300002	Sub-Saharan Africa, Southern	192	...	0	
300003	North Africa/Middle East	138	...	4	
300004	Sub-Saharan Africa, Western	199	...	0	

	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
300000	0	0	0	0	0	0	
300001	0	0	0	0	0	0	
300002	0	0	0	0	0	0	
300003	0	0	0	0	0	0	
300004	0	0	0	0	0	0	

	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
300000	18	3	-
300001	-	-	0
300002	0	-	0
300003	0	0	0
300004	0	5	0

[5 rows x 76 columns]

CHUNK DF

	year	source	channel	recipient_isocode	recipient_country	\
350000	2017	Belgium	GAVI	ZWE	Zimbabwe	
350001	2017	Belgium	GFATM	AFG	Afghanistan	
350002	2017	Belgium	GFATM	AFG	Afghanistan	
350003	2017	Belgium	GFATM	AGO	Angola	
350004	2017	Belgium	GFATM	ALB	Albania	

	gbd_location_id	wb_regioncode	wb_location_id	\
350000	198	SSA	242	
350001	160	SAS	244	
350002	160	SAS	244	
350003	168	SSA	242	
350004	43	ECA	240	

	gbd_region	gbd_region_id	...	other_dah_20	\
350000	Sub-Saharan Africa, Southern	192.0	...	0	
350001	North Africa/Middle East	138.0	...	0	
350002	North Africa/Middle East	138.0	...	0	
350003	Sub-Saharan Africa, Central	167.0	...	0	
350004	Europe, Central	42.0	...	0	

	rmh_dah_20	nch_dah_20	ncd_dah_20	hiv_dah_20	mal_dah_20	tb_dah_20	\
350000	0	14	0	0	0	0	
350001	0	0	0	27	37	34	
350002	0	0	0	2	0	0	
350003	0	0	0	29	82	46	
350004	0	0	0	-	0	-	

	swap_hss_total_dah_20	oid_dah_20	unalloc_dah_20
350000	0	0	-
350001	24	0	-
350002	0	0	0

350003	0	0	-
350004	0	0	0

[5 rows x 76 columns]

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\554763948.py:1: DtypeWarning: Columns (55) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

```
[106]: new_df = pd.DataFrame()
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
                             chunksize = 50000):
    chunk_df['dah_20'] = pd.to_numeric(chunk_df['dah_20'], errors='coerce')
    result = chunk_df.groupby('recipient_country').agg({'year': ['count'],
                                                         'dah_20': ['mean', 'median']})
    new_df = pd.concat([new_df, result])
```

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\3003579497.py:2:

DtypeWarning: Columns

(29,32,34,35,36,37,38,40,41,42,44,46,47,48,49,50,51,52,60,65) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\3003579497.py:2:

DtypeWarning: Columns (35,38,39,40,42,44,46,47,51,53,57,65) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\3003579497.py:2:

DtypeWarning: Columns (15,53,57,59,61) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\3003579497.py:2:

DtypeWarning: Columns (35,48,53,57,60,61) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\3003579497.py:2:

DtypeWarning: Columns (36,53,57,60) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\3003579497.py:2:

DtypeWarning: Columns (59,61) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\3003579497.py:2:

DtypeWarning: Columns (15,19,54,59) have mixed types. Specify dtype option on import or set low_memory=False.

```
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

C:\Users\121731\AppData\Local\Temp\ipykernel_11696\3003579497.py:2:

DtypeWarning: Columns (55) have mixed types. Specify dtype option on import or

```
set low_memory=False.
for chunk_df in pd.read_csv('IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV',
```

```
[107]: new_df
```

```
[107]:
```

	year	dah_20	
	count	mean	median
recipient_country			
Administrative expenses	2189	5542.780083	581.5
Afghanistan	291	613.273356	0.0
Albania	252	386.020000	5.5
Algeria	157	145.029412	3.0
Angola	459	807.973333	37.5
...
Vietnam	291	1689.396552	210.0
Wallis and Futuna Islands	19	188.187500	13.5
Yemen	278	2114.059041	119.0
Zambia	335	3792.509375	248.0
Zimbabwe	354	2622.639053	180.5

```
[1261 rows x 3 columns]
```

```
[ ]:
```