Zadanie 1

In [3]:
```
import pandas as pd
```

Zadanie 2

In [4]:
```
dict_student = {"Student" : ["A", "B", "C"],
            "Grade" : [5,   4,  4.5]}

df = pd.DataFrame(dict_student)
df
```

Out[4]:

	Student	Grade
0	A	5.0
1	B	4.0
2	C	4.5

Zadanie 3

In [13]:
```
df = pd.read_csv("IHME_DAH_DATABASE_1990_2020_Y2021M09D22.CSV", encoding='utf-8', engine
```

Zadanie 4

In [7]:
```
lists_students = [["A", "B", "C"],
[5,   4,  4.5]]

pd.DataFrame(lists_students)
```

Out[7]:

	0	1	2
0	A	B	C
1	5	4	4.5

Zadanie 5

In [8]:
```
pd.DataFrame(lists_students).T
```

Out[8]:

	0	1
0	A	5
1	B	4
2	C	4.5

Zadanie 6

In [14]:
```
df.head(10)
```

Out[14]:

	year	source	channel	recipient_isocode	recipient_country	gbd_location_id	wb_regioncode	wb_location_id
0	1990	Australia	BIL_AUS	AGO	Angola	168	SSA	242

1	1990	Australia	BIL_AUS	BDI	Burundi	175	SSA	242
2	1990	Australia	BIL_AUS	BEN	Benin	200	SSA	242
3	1990	Australia	BIL_AUS	BFA	Burkina Faso	201	SSA	242
4	1990	Australia	BIL_AUS	BWA	Botswana	193	SSA	242
5	1990	Australia	BIL_AUS	CAF	Central African Republic	169	SSA	242
6	1990	Australia	BIL_AUS	CHN	China	6	EAP	239
7	1990	Australia	BIL_AUS	CIV	Cote d'Ivoire	205	SSA	242
8	1990	Australia	BIL_AUS	CMR	Cameroon	202	SSA	242
9	1990	Australia	BIL_AUS	COD	Democratic Republic of the Congo	171	SSA	242

10 rows × 76 columns

Zadanie 7

In [15]:
```
df.tail(10)
```

Out[15]:

	year	source	channel	recipient_isocode	recipient_country	gbd_location_id	wb_regioncode
384296	2020	United_States	INTLNGO	QZA	Unallocated/Unspecified	44598	NaN
384297	2020	United_States	NGO	QZA	Unallocated/Unspecified	44598	NaN
384298	2020	United_States	PAHO	QZA	Unallocated/Unspecified	44598	NaN
384299	2020	United_States	UNAIDS	QZA	Unallocated/Unspecified	44598	NaN
384300	2020	United_States	UNFPA	QZA	Unallocated/Unspecified	44598	NaN
384301	2020	United_States	UNICEF	QZA	Unallocated/Unspecified	44598	NaN
384302	2020	United_States	UNITAID	QZA	Unallocated/Unspecified	44598	NaN
384303	2020	United_States	UNITAID	QZA	Unallocated/Unspecified	44598	NaN

384304	2020	United_States	WB_IDA	QZA	Unallocated/Unspecified	44598	NaN
384305	2020	United_States	WHO	QZA	Unallocated/Unspecified	44598	NaN

10 rows × 76 columns

Zadanie 8

In [16]:
```
df.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
```

Zadanie 9

In [17]:
```
df.shape
```

Out[17]:
```
(384306, 76)
```

Zadanie 10

In [18]:
```
df.describe()
```

Out[18]:

	year	gbd_location_id	wb_location_id	gbd_region_id	gbd_superregion_id	elim_ch	prel
count	384306.000000	384306.000000	384306.000000	383993.000000	383993.000000	384306.000000	384306.
mean	2008.127521	1765.935533	2240.752439	1745.812671	1733.144388	0.252052	0.
std	6.945191	8325.915434	9204.906147	8328.525983	8330.949734	0.434191	0.
min	1990.000000	1.000000	239.000000	1.000000	1.000000	0.000000	0.
25%	2004.000000	110.000000	241.000000	96.000000	64.000000	0.000000	0.
50%	2009.000000	169.000000	242.000000	159.000000	158.000000	0.000000	0.
75%	2014.000000	200.000000	242.000000	192.000000	166.000000	1.000000	0.
max	2020.000000	44598.000000	44621.000000	44598.000000	44598.000000	1.000000	1.

Zadanie 11

In [19]:
```
df.describe(include = 'all')
```

Out[19]:

	year	source	channel	recipient_isocode	recipient_country	gbd_location_id	wb_regionc

count	384306.000000	384306	384306	384306	383773	384306.000000	370
unique	NaN	32	46	176	174	NaN	
top	NaN	United_Kingdom	GFATM	INKIND	Administrative expenses	NaN	
freq	NaN	20452	57306	8379	8379	NaN	172
mean	2008.127521	NaN	NaN	NaN	NaN	1765.935533	I
std	6.945191	NaN	NaN	NaN	NaN	8325.915434	I
min	1990.000000	NaN	NaN	NaN	NaN	1.000000	I
25%	2004.000000	NaN	NaN	NaN	NaN	110.000000	I
50%	2009.000000	NaN	NaN	NaN	NaN	169.000000	I
75%	2014.000000	NaN	NaN	NaN	NaN	200.000000	I
max	2020.000000	NaN	NaN	NaN	NaN	44598.000000	I

11 rows × 76 columns

Zadanie 12

In [20]:
```
df.dropna(inplace=True)
```

Zadanie 13

In [22]:
```
df["source"]
```

Out[22]:
```
0              Australia
1              Australia
2              Australia
3              Australia
4              Australia
               ...
383233    United_States
383234    United_States
383235    United_States
383236    United_States
383237    United_States
Name: source, Length: 369785, dtype: object
```

In [23]:
```
df.source
```

Out[23]:
```
0              Australia
1              Australia
2              Australia
3              Australia
4              Australia
               ...
383233    United_States
383234    United_States
383235    United_States
383236    United_States
383237    United_States
Name: source, Length: 369785, dtype: object
```

In [24]:
```
df[["year","other_dah_20","ncd_dah_20"]]
```

Out[24]:

	year	other_dah_20	ncd_dah_20

0	1990	0	0
1	1990	0	0
2	1990	0	0
3	1990	0	0
4	1990	0	0
...	...	...	...
383233	2018	10480	22
383234	2018	-	0
383235	2018	36	0
383236	2018	2	0
383237	2018	0	0

369785 rows × 3 columns

In [25]:
```
df.loc[:, "gbd_region_id":"other_dah_20"]
```

Out[25]:

	gbd_region_id	gbd_superregion	gbd_superregion_id	elim_ch	prelim_est	dah_20	rmh_fp_dah_20	rmh_m
0	167.0	Sub-Saharan Africa	166.0	0	0	14	1	
1	174.0	Sub-Saharan Africa	166.0	0	0	12	1	
2	199.0	Sub-Saharan Africa	166.0	0	0	12	1	
3	199.0	Sub-Saharan Africa	166.0	0	0	13	2	
4	192.0	Sub-Saharan Africa	166.0	0	0	25	1	
...	...	...	...	...	...	...	...	
383233	1.0	Global	1.0	1	0	79175	9737	
383234	138.0	North Africa and Middle East	137.0	1	0	310	-	
383235	192.0	Sub-Saharan Africa	166.0	1	0	573	82	
383236	174.0	Sub-Saharan Africa	166.0	1	0	249	10	
383237	192.0	Sub-Saharan Africa	166.0	1	0	192	6	

369785 rows × 58 columns

In [26]:
```
df.loc[1000:1010, "gbd_region_id":"other_dah_20"] #1000 do 1010
```

Out[26]:

	gbd_region_id	gbd_superregion	gbd_superregion_id	elim_ch	prelim_est	dah_20	rmh_fp_dah_20	rmh_mh_
1000	199.0	Sub-Saharan Africa	166.0	0	0	62	-	

1001	199.0	Sub-Saharan Africa	166.0	0	0	52	-
1002	167.0	Sub-Saharan Africa	166.0	0	0	14	0
1003	104.0	Latin America and Caribbean	103.0	0	0	10	0
1004	124.0	Latin America and Caribbean	103.0	0	0	564	0
1005	124.0	Latin America and Caribbean	103.0	0	0	358	0
1006	104.0	Latin America and Caribbean	103.0	0	0	443	0
1007	9.0	Southeast Asia, East Asia, and Oceania	4.0	0	0	-	0
1008	159.0	South Asia	158.0	0	0	1	0
1010	138.0	North Africa and Middle East	137.0	0	0	-	0

10 rows × 58 columns

In [28]:
```
df.iloc[1000:1010, 1:4] #1000 do 1009
```

Out[28]:

	source	channel	recipient_isocode
1067	Finland	BIL_FIN	TUN
1068	Finland	BIL_FIN	TZA
1069	Finland	BIL_FIN	UGA
1070	Finland	BIL_FIN	UZB
1071	Finland	BIL_FIN	VCT
1072	Finland	BIL_FIN	VNM
1073	Finland	BIL_FIN	WLD
1074	Finland	BIL_FIN	YEM
1075	Finland	BIL_FIN	ZAF
1076	Finland	BIL_FIN	ZMB

Zadanie 14

In [29]:
```
df[df["channel"] == "NGO"]
```

Out[29]:

	year	source	channel	recipient_isocode	recipient_country	gbd_location_id	wb_regioncode	wb_loc
3543	1990	United_States	NGO	SSD	South Sudan	435	SSA	
7251	1991	United_States	NGO	SSD	South Sudan	435	SSA	

10402	1992	Private_other	NGO	AFG	Afghanistan	160	SAS
10403	1992	Private_other	NGO	AGO	Angola	168	SSA
10404	1992	Private_other	NGO	ARG	Argentina	97	LAC
...	...	...	...	...	...	...	...
382994	2018	United_States	NGO	ZAF	South Africa	196	SSA
382995	2018	United_States	NGO	ZMB	Zambia	191	SSA
382996	2018	United_States	NGO	ZMB	Zambia	191	SSA
382997	2018	United_States	NGO	ZWE	Zimbabwe	198	SSA
382998	2018	United_States	NGO	ZWE	Zimbabwe	198	SSA

7929 rows × 76 columns

Zadanie 15

In [35]:
```
df[(df["channel"] == "NGO") & (df["recipient_isocode"] == "SSD") & (df["year"] == 1990)]
```

Out[35]:

	year	source	channel	recipient_isocode	recipient_country	gbd_location_id	wb_regioncode	wb_locat
3543	1990	United_States	NGO	SSD	South Sudan	435	SSA	

1 rows × 76 columns

Zadanie 16

In [36]:
```
df[df["gbd_region"].str.contains("America")]
```

Out[36]:

	year	source	channel	recipient_isocode	recipient_country	gbd_location_id	wb_regioncode	wb_loc
92	1990	Australia	WB_IDA	BOL	Bolivia	121	LAC	
115	1990	Australia	WB_IDA	HND	Honduras	129	LAC	
142	1990	Australia	WB_IDA	NIC	Nicaragua	131	LAC	

182	1990	Austria	IDB	ARG	Argentina	97	LAC
183	1990	Austria	IDB	BOL	Bolivia	121	LAC
...	...	...	...	...	...	...	...
383023	2018	United_States	PAHO	SLV	El Salvador	127	LAC
383026	2018	United_States	PAHO	VEN	Venezuela	133	LAC
383063	2018	United_States	WB_IDA	BOL	Bolivia	121	LAC
383086	2018	United_States	WB_IDA	HND	Honduras	129	LAC
383113	2018	United_States	WB_IDA	NIC	Nicaragua	131	LAC

33208 rows × 76 columns

Zadanie 17

In [37]:
```
df[df["gbd_region"].str.contains("America") == False]
```

Out[37]:

	year	source	channel	recipient_isocode	recipient_country	gbd_location_id	wb_regioncode	wb_loc
0	1990	Australia	BIL_AUS	AGO	Angola	168	SSA	
1	1990	Australia	BIL_AUS	BDI	Burundi	175	SSA	
2	1990	Australia	BIL_AUS	BEN	Benin	200	SSA	
3	1990	Australia	BIL_AUS	BFA	Burkina Faso	201	SSA	
4	1990	Australia	BIL_AUS	BWA	Botswana	193	SSA	
...	...	...	...	...	...	...	...	
383233	2018	United_States	WHO	WLD	Global	1	WLD	
383234	2018	United_States	WHO	YEM	Yemen	157	MNA	

383235	2018	United_States	WHO	ZAF	South Africa	196	SSA	
383236	2018	United_States	WHO	ZMB	Zambia	191	SSA	
383237	2018	United_States	WHO	ZWE	Zimbabwe	198	SSA	

336577 rows × 76 columns

Zadanie 18

In [39]:
```
df["gbd_region_id_int"] = df["gbd_region_id"].astype(int)
df.head()
```

Out[39]:

	year	source	channel	recipient_isocode	recipient_country	gbd_location_id	wb_regioncode	wb_location_id
0	1990	Australia	BIL_AUS	AGO	Angola	168	SSA	242
1	1990	Australia	BIL_AUS	BDI	Burundi	175	SSA	242
2	1990	Australia	BIL_AUS	BEN	Benin	200	SSA	242
3	1990	Australia	BIL_AUS	BFA	Burkina Faso	201	SSA	242
4	1990	Australia	BIL_AUS	BWA	Botswana	193	SSA	242

5 rows × 77 columns

Zadanie 19

In [40]:
```
df.drop("rmh_hss_other_dah_20", axis=1, inplace = True)
```

Zadanie 20

In [41]:
```
df.rename(columns = {"recipient_country": "country"}, inplace = True)
```

Zadanie 21

In [42]:
```
df.to_csv("lab1.csv")
```

Zadanie 22

In [51]:
```
df["gbd_superregion_id"].mean()
```

Out[51]:
```
114.09503630487987
```

In [52]:
```
df['gbd_superregion_id'].max()
```

Out[52]:
```
166.0
```

In [53]:
```
df['gbd_superregion_id'].min()
```

Out[53]:
```
1.0
```

Zadanie 23

In [54]:
```
df['year'].count()
```

Out[54]:
```
369785
```

Zadanie 24

In [55]:
```
df['year'].unique()
```

Out[55]:
```
array([1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000,
       2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011,
       2012, 2013, 2014, 2015, 2016, 2017, 2018], dtype=int64)
```

Zadanie 25

In [56]:
```
df['year'].value_counts()
```

Out[56]:
```
2011    21742
2009    21419
2014    21214
2013    20921
2012    20842
2016    20401
2008    20280
2015    19866
2007    19296
2006    19062
2010    18543
2017    17538
2018    16638
2005    14282
2004    13426
2003    11451
2002     9440
2001     8069
2000     7832
1999     6936
1998     6582
1997     5366
1996     4707
1995     4552
1993     4494
1992     4069
1994     4055
1990     3402
1991     3360
Name: year, dtype: int64
```

Zadanie 26

In [57]:
```
df.sort_values(['gbd_location_id'], ascending = True)
```

Out[57]:

	year	source	channel	recipient_isocode	country	gbd_location_id	wb_regioncode	wb_location
170138	2008	Netherlands	GFATM	WLD	Global	1	WLD	446
380554	2018	Sweden	INTLNGO	WLD	Global	1	WLD	446
6859	1991	Sweden	WHO	WLD	Global	1	WLD	446
254662	2012	Luxembourg	WHO	WLD	Global	1	WLD	446
338212	2016	Korea	INTLNGO	WLD	Global	1	WLD	446
...	...	...	...	...	...	...	...	
16102	1993	United_Kingdom	BIL_GBR	AIA	Anguilla	44598	LAC	2
66872	2001	United_Kingdom	BIL_GBR	AIA	Anguilla	44598	LAC	2
50322	1999	United_Kingdom	BIL_GBR	AIA	Anguilla	44598	LAC	2
116620	2005	United_Kingdom	BIL_GBR	AIA	Anguilla	44598	LAC	2
181872	2009	Canada	INTLNGO	AIA	Anguilla	44598	LAC	2

369785 rows × 76 columns

In [58]:
```
df.sort_values(['gbd_location_id'], ascending = False)
```

Out[58]:

	year	source	channel	recipient_isocode	country	gbd_location_id	wb_regioncode	wb_location
25444	1995	United_Kingdom	BIL_GBR	AIA	Anguilla	44598	LAC	2
50322	1999	United_Kingdom	BIL_GBR	AIA	Anguilla	44598	LAC	2
181872	2009	Canada	INTLNGO	AIA	Anguilla	44598	LAC	2
58411	2000	United_Kingdom	BIL_GBR	AIA	Anguilla	44598	LAC	2
282618	2013	United_Kingdom	BIL_GBR	AIA	Anguilla	44598	LAC	2
...	...	...	...	...	...	...	...	
269222	2013	Finland	GFATM	WLD	Global	1	WLD	446
150000	2007	Netherlands	WHO	WLD	Global	1	WLD	446
347740	2016	United_States	UNICEF	WLD	Global	1	WLD	446
206113	2010	Finland	WHO	WLD	Global	1	WLD	446
290759	2014	Finland	UNICEF	WLD	Global	1	WLD	446

369785 rows × 76 columns

Zadanie 27

In [59]:
```
df.nlargest(10,'year')
```

Out[59]:

	year	source	channel	recipient_isocode	country	gbd_location_id	wb_regioncode	wb_location_id	g
365998	2018	Australia	AsDB	BGD	Bangladesh	161	SAS	244	
365999	2018	Australia	AsDB	CHN	China	6	EAP	239	
366000	2018	Australia	AsDB	IND	India	163	SAS	244	

366002	2018	Australia	AsDB	KHM	Cambodia	10	EAP	239	
366003	2018	Australia	AsDB	LAO	Laos	12	EAP	239	
366004	2018	Australia	AsDB	MMR	Myanmar	15	EAP	239	
366005	2018	Australia	AsDB	MNG	Mongolia	38	EAP	239	
366006	2018	Australia	AsDB	NRU	Nauru	369	EAP	239	
366007	2018	Australia	AsDB	PAK	Pakistan	165	SAS	244	
366008	2018	Australia	AsDB	PHL	Philippines	16	EAP	239	

10 rows × 76 columns

In [60]:
```
df.nsmallest(10, 'year')
```

Out[60]:

	year	source	channel	recipient_isocode	country	gbd_location_id	wb_regioncode	wb_location_id	gbd_re
0	1990	Australia	BIL_AUS	AGO	Angola	168	SSA	242	Sah A Ce
1	1990	Australia	BIL_AUS	BDI	Burundi	175	SSA	242	Sah A Eas
2	1990	Australia	BIL_AUS	BEN	Benin	200	SSA	242	Sah A Wes
3	1990	Australia	BIL_AUS	BFA	Burkina Faso	201	SSA	242	Sah A Wes
4	1990	Australia	BIL_AUS	BWA	Botswana	193	SSA	242	Sah A Sout
5	1990	Australia	BIL_AUS	CAF	Central African Republic	169	SSA	242	Sah A Ce
6	1990	Australia	BIL_AUS	CHN	China	6	EAP	239	Asia,
7	1990	Australia	BIL_AUS	CIV	Cote d'Ivoire	205	SSA	242	Sah A Wes
8	1990	Australia	BIL_AUS	CMR	Cameroon	202	SSA	242	Sah A Wes
9	1990	Australia	BIL_AUS	COD	Democratic	171	SSA	242	

					Republic of the Congo				Sah A Ce

10 rows × 76 columns

Zadanie 28

In [61]:
```
df[df['year'] == 2015].nlargest(10,'wb_location_id')
```

Out[61]:

	year	source	channel	recipient_isocode	country	gbd_location_id	wb_regioncode	wb_location_id	gb
306663	2015	Australia	BIL_AUS	WLD	Global	1	WLD	44620	
306740	2015	Australia	GAVI	WLD	Global	1	WLD	44620	
307009	2015	Australia	INTLNGO	WLD	Global	1	WLD	44620	
307010	2015	Australia	INTLNGO	WLD	Global	1	WLD	44620	
307027	2015	Australia	UNICEF	WLD	Global	1	WLD	44620	
307168	2015	Australia	WHO	WLD	Global	1	WLD	44620	
307294	2015	Austria	BIL_AUT	WLD	Global	1	WLD	44620	
307415	2015	Austria	EC	WLD	Global	1	WLD	44620	
307755	2015	Austria	WHO	WLD	Global	1	WLD	44620	
307864	2015	BMGF	BMGF	WLD	Global	1	WLD	44620	

10 rows × 76 columns

Zadanie 29

In [62]:
```
df.groupby('channel').agg('mean')
```

Out[62]:

channel	year	gbd_location_id	wb_location_id	gbd_region_id	gbd_superregion_id	elim_ch	prelim_est
AfDB	2009.026676	203.219108	282.913043	173.146011	156.135279	0.030171	0.000000
AsDB	2006.937070	54.380502	248.232290	48.646492	43.640475	0.063446	0.000000
BIL_ARE	2008.665823	154.698734	3500.531646	131.878481	123.696203	0.000000	0.000000
BIL_AUS	2004.346463	150.769025	733.268069	107.667304	95.047801	0.000000	0.000000
BIL_AUT	2007.982321	157.263643	1230.833974	136.898540	122.703305	0.000000	0.000000
BIL_BEL	2004.487395	164.591597	962.705882	144.570308	129.839776	0.000000	0.000000
BIL_CAN	2005.111483	146.255254	633.529698	125.674688	113.365519	0.000000	0.000000
BIL_CHE	2005.807892	164.519730	909.924714	141.181205	127.087227	0.000000	0.000000
BIL_DEU	2006.735355	142.254652	684.906616	121.916609	109.638181	0.000000	0.000000
BIL_DNK	2004.664112	164.364623	1063.686462	144.613027	131.662197	0.000000	0.000000
BIL_ESP	2007.383581	151.259740	838.487477	133.176252	119.706401	0.000000	0.000000
BIL_FIN	2000.898403	152.629898	864.264151	129.507499	117.227866	0.000000	0.000000
BIL_FRA	2005.174430	154.334404	654.808545	125.162223	111.621266	0.000000	0.000000

BIL_GBR	2005.559356	340.680751	673.016432	120.650235	108.700201	0.000000	0.000000
BIL_GRC	2007.309829	142.881410	1616.403846	119.112179	107.007479	0.000000	0.000000
BIL_IRL	2007.344692	170.403670	1928.463958	145.281782	130.756225	0.000000	0.000000
BIL_ITA	2004.787816	156.160796	775.005387	136.703688	122.925818	0.000000	0.000000
BIL_JPN	2008.004693	140.312996	705.802527	109.283032	97.303610	0.000000	0.000000
BIL_KOR	2011.274121	141.247069	1401.835888	113.834986	102.304779	0.000000	0.000000
BIL_LUX	2008.682716	156.004938	1830.397531	135.053086	120.585185	0.000000	0.000000
BIL_NLD	2002.602684	155.808796	721.320164	134.595602	121.657100	0.000000	0.000000
BIL_NOR	2003.740378	146.883135	691.885934	127.493002	115.365640	0.000000	0.000000
BIL_NZL	2008.020566	164.722365	1894.906170	92.899743	78.350900	0.000000	0.000000
BIL_PRT	2008.417570	189.573753	1637.725597	162.939262	145.732104	0.000000	0.000000
BIL_SWE	2003.188500	150.187847	662.010127	129.484809	116.843515	0.000000	0.000000
BIL_USA	2004.285365	140.716807	759.495520	118.452892	106.246810	0.005973	0.000000
BMGF	2008.047399	141.671980	613.785654	122.541107	110.457215	0.000000	0.000000
CEPI	2017.600000	1.000000	44620.000000	1.000000	1.000000	0.000000	0.000000
EC	2007.277381	145.419498	696.604122	122.818951	110.593816	0.002959	0.000000
EEA	2011.235294	51.602941	240.000000	46.323529	31.000000	0.000000	0.000000
GAVI	2011.306487	156.852421	861.399984	138.794382	124.716985	0.186879	0.000000
GFATM	2010.909216	137.919538	331.205242	118.160255	105.157225	0.143240	0.000000
IDB	2002.768485	120.464755	248.239895	115.164476	98.465495	0.022839	0.000000
INTLNGO	2011.836618	162.319908	922.027115	135.409902	121.890125	0.387162	0.139678
NGO	2008.259301	145.553916	666.876403	125.910455	113.459957	0.817001	0.105562
PAHO	2006.372283	123.566576	1225.861866	110.254982	97.425725	1.000000	0.000000
UNAIDS	2008.044602	155.615385	3827.583064	133.339367	120.585650	0.999354	0.000000
UNFPA	2010.039476	175.593289	1491.188760	149.669021	134.443071	0.998338	0.000000
UNICEF	2010.880682	160.296171	786.541775	138.520672	125.480125	0.999668	0.000000
UNITAID	2009.176471	1.000000	44620.000000	1.000000	1.000000	1.000000	0.000000
US_FOUND	2006.348972	151.423816	568.442583	113.302726	100.889410	0.130474	0.029044
WB	2009.480630	162.776466	545.129066	139.910503	126.780609	1.000000	0.000000
WB_IBRD	2007.752220	136.036505	460.351656	119.343058	107.603101	0.707681	0.000000
WB_IDA	2004.482014	133.958553	242.919052	119.328811	106.670662	0.019187	0.000000
WHO	2010.128569	159.082439	974.079304	135.226110	122.149158	0.999794	0.000000

Zadanie 30

In [66]:
```
df.groupby('channel').agg({'year': ['count'], 'gbd_region_id': ['mean', 'median']})
```

Out[66]:

	year	gbd_region_id	
	count	mean	median

channel			
AfDB	16307	173.146011	174.0
AsDB	5816	48.646492	9.0
BIL_ARE	395	131.878481	138.0
BIL_AUS	2615	107.667304	138.0
BIL_AUT	1301	136.898540	167.0
BIL_BEL	1785	144.570308	174.0
BIL_CAN	3283	125.674688	138.0
BIL_CHE	1926	141.181205	174.0
BIL_DEU	2902	121.916609	138.0
BIL_DNK	1566	144.613027	174.0
BIL_ESP	2156	133.176252	138.0
BIL_FIN	2067	129.507499	159.0
BIL_FRA	3113	125.162223	138.0
BIL_GBR	2982	120.650235	138.0
BIL_GRC	936	119.112179	138.0
BIL_IRL	763	145.281782	174.0
BIL_ITA	2413	136.703688	138.0
BIL_JPN	2770	109.283032	124.0
BIL_KOR	1109	113.834986	138.0
BIL_LUX	810	135.053086	170.5
BIL_NLD	2683	134.595602	159.0
BIL_NOR	2858	127.493002	138.0
BIL_NZL	778	92.899743	32.0
BIL_PRT	922	162.939262	174.0
BIL_SWE	3061	129.484809	138.0
BIL_USA	3683	118.452892	134.0
BMGF	2384	122.541107	138.0
CEPI	15	1.000000	1.0
EC	40558	122.818951	138.0
EEA	68	46.323529	42.0
GAVI	25776	138.794382	174.0
GFATM	56772	118.160255	138.0
IDB	12172	115.164476	120.0
INTLNGO	29024	135.409902	167.0
NGO	7929	125.910455	138.0
PAHO	2208	110.254982	104.0
UNAIDS	1547	133.339367	174.0

UNFPA	9626	149.669021	174.0
UNICEF	15069	138.520672	174.0
UNITAID	17	1.000000	1.0
US_FOUND	4476	113.302726	124.0
WB	4827	139.910503	174.0
WB_IBRD	7095	119.343058	138.0
WB_IDA	54986	119.328811	159.0
WHO	24236	135.226110	167.0

Zadanie 31

In [70]:
```
df2 = df.groupby('channel').agg({'year': ['count'], 'gbd_region_id': ['mean', 'median']})
df2.columns
```

Out[70]:
```
MultiIndex([(         'year',   'count'),
            ('gbd_region_id',    'mean'),
            ('gbd_region_id', 'median')],
           )
```

Zadanie 32

In [71]:
```
df2['year']['count'].sort_values(ascending = False)
```

Out[71]:
```
channel
GFATM        56772
WB_IDA       54986
EC           40558
INTLNGO      29024
GAVI         25776
WHO          24236
AfDB         16307
UNICEF       15069
IDB          12172
UNFPA         9626
NGO           7929
WB_IBRD       7095
AsDB          5816
WB            4827
US_FOUND      4476
BIL_USA       3683
BIL_CAN       3283
BIL_FRA       3113
BIL_SWE       3061
BIL_GBR       2982
BIL_DEU       2902
BIL_NOR       2858
BIL_JPN       2770
BIL_NLD       2683
BIL_AUS       2615
BIL_ITA       2413
BMGF          2384
PAHO          2208
BIL_ESP       2156
BIL_FIN       2067
BIL_CHE       1926
BIL_BEL       1785
BIL_DNK       1566
UNAIDS        1547
```

```
BIL_AUT      1301
BIL_KOR      1109
BIL_GRC       936
BIL_PRT       922
BIL_LUX       810
BIL_NZL       778
BIL_IRL       763
BIL_ARE       395
EEA            68
UNITAID        17
CEPI           15
Name: count, dtype: int64
```

Zadanie 33

In [77]:
```
p_table = df.pivot_table(values='gbd_superregion_id', index='channel', columns='gbd_regi
                         margins=False, dropna=True, fill_value=None)
p_table
```

Out[77]:

gbd_region_id	1.0	5.0	9.0	21.0	32.0	42.0	56.0	65.0	73.0	96.0	104.0	120.0	124.0	134.0	138.0	15
channel																
AfDB	1.0	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	103.0	NaN	NaN	NaN	137.0	N
AsDB	1.0	4.0	4.0	4.0	31.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	137.0	15
BIL_ARE	1.0	NaN	4.0	NaN	31.0	31.0	31.0	NaN	NaN	64.0	103.0	NaN	103.0	NaN	137.0	15
BIL_AUS	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_AUT	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_BEL	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	64.0	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_CAN	1.0	4.0	4.0	4.0	31.0	31.0	31.0	64.0	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_CHE	1.0	4.0	4.0	NaN	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_DEU	1.0	4.0	4.0	4.0	31.0	31.0	31.0	64.0	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_DNK	1.0	4.0	4.0	NaN	31.0	31.0	NaN	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_ESP	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_FIN	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_FRA	1.0	4.0	4.0	4.0	31.0	31.0	31.0	64.0	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_GBR	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_GRC	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_IRL	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_ITA	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_JPN	1.0	4.0	4.0	4.0	31.0	31.0	31.0	64.0	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_KOR	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	NaN	103.0	103.0	103.0	103.0	137.0	15
BIL_LUX	1.0	4.0	4.0	NaN	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_NLD	1.0	4.0	4.0	4.0	31.0	31.0	31.0	64.0	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_NOR	1.0	4.0	4.0	NaN	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_NZL	1.0	4.0	4.0	4.0	31.0	NaN	NaN	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_PRT	1.0	NaN	4.0	NaN	31.0	31.0	NaN	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15

BIL_SWE	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BIL_USA	1.0	4.0	4.0	4.0	31.0	31.0	31.0	64.0	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
BMGF	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
CEPI	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
EC	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
EEA	NaN	NaN	NaN	NaN	NaN	31.0	31.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
GAVI	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	NaN	103.0	103.0	103.0	NaN	137.0	15
GFATM	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
IDB	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	64.0	103.0	103.0	103.0	103.0	NaN	N
INTLNGO	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
NGO	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
PAHO	1.0	4.0	4.0	NaN	31.0	NaN	NaN	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
UNAIDS	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
UNFPA	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
UNICEF	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
UNITAID	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
US_FOUND	1.0	4.0	4.0	4.0	31.0	31.0	31.0	64.0	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
WB	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
WB_IBRD	1.0	4.0	4.0	4.0	31.0	31.0	31.0	64.0	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15
WB_IDA	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	NaN	103.0	103.0	103.0	NaN	137.0	15
WHO	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0	137.0	15

Zadanie 34

In [80]:
```
p_table.index
```

Out[80]:
```
Index(['AfDB', 'AsDB', 'BIL_ARE', 'BIL_AUS', 'BIL_AUT', 'BIL_BEL', 'BIL_CAN',
       'BIL_CHE', 'BIL_DEU', 'BIL_DNK', 'BIL_ESP', 'BIL_FIN', 'BIL_FRA',
       'BIL_GBR', 'BIL_GRC', 'BIL_IRL', 'BIL_ITA', 'BIL_JPN', 'BIL_KOR',
       'BIL_LUX', 'BIL_NLD', 'BIL_NOR', 'BIL_NZL', 'BIL_PRT', 'BIL_SWE',
       'BIL_USA', 'BMGF', 'CEPI', 'EC', 'EEA', 'GAVI', 'GFATM', 'IDB',
       'INTLNGO', 'NGO', 'PAHO', 'UNAIDS', 'UNFPA', 'UNICEF', 'UNITAID',
       'US_FOUND', 'WB', 'WB_IBRD', 'WB_IDA', 'WHO'],
      dtype='object', name='channel')
```

In [81]:
```
p_table.columns
```

Out[81]:
```
Float64Index([  1.0,   5.0,   9.0,  21.0,  32.0,  42.0,  56.0,  65.0,  73.0,
               96.0, 104.0, 120.0, 124.0, 134.0, 138.0, 159.0, 167.0, 174.0,
              192.0, 199.0],
             dtype='float64', name='gbd_region_id')
```

Zadanie 35

In [82]:
```
p_table = df.pivot_table(values='gbd_superregion_id', index=['channel', 'source'], colum
                         margins=False, dropna=True, fill_value=None)
p_table
```

Out[82]:

	gbd_region_id	1.0	5.0	9.0	21.0	32.0	42.0	56.0	65.0	73.0	96.0	104.0	120.0	124.0	134.0

channel	source														
AfDB	**Austria**	1.0	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	103.0	NaN	NaN	NaN
	BMGF	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	103.0	NaN	NaN	NaN
	Belgium	1.0	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	103.0	NaN	NaN	NaN
	Canada	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	103.0	NaN	NaN	NaN
	China	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	103.0	NaN	NaN	NaN
...	**...**	...	...	...	...	...	...	...	...	...	...	...	...	...	...
WHO	**Spain**	1.0	4.0	4.0	NaN	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0
	Sweden	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	NaN	103.0	NaN
	Switzerland	1.0	NaN	4.0	NaN	31.0	31.0	31.0	NaN	NaN	NaN	103.0	NaN	NaN	NaN
	United_Kingdom	1.0	4.0	4.0	NaN	31.0	31.0	31.0	NaN	NaN	NaN	103.0	NaN	103.0	NaN
	United_States	1.0	4.0	4.0	4.0	31.0	31.0	31.0	NaN	NaN	64.0	103.0	103.0	103.0	103.0

408 rows × 20 columns

In [83]:
```
p_table.index
```

Out[83]:
```
MultiIndex([('AfDB',                 'Austria'),
            ('AfDB',                    'BMGF'),
            ('AfDB',                 'Belgium'),
            ('AfDB',                  'Canada'),
            ('AfDB',                   'China'),
            ('AfDB',                 'Denmark'),
            ('AfDB',                 'Finland'),
            ('AfDB',                  'France'),
            ('AfDB',                 'Germany'),
            ('AfDB',                  'Greece'),
            ...
            ( 'WHO',             'New_Zealand'),
            ( 'WHO', 'Non_OECD_DAC_countries'),
            ( 'WHO',                  'Norway'),
            ( 'WHO',                'Portugal'),
            ( 'WHO',           'Private_other'),
            ( 'WHO',                   'Spain'),
            ( 'WHO',                  'Sweden'),
            ( 'WHO',             'Switzerland'),
            ( 'WHO',          'United_Kingdom'),
            ( 'WHO',           'United_States')],
           names=['channel', 'source'], length=408)
```

Zadanie 36

In [86]:
```
import matplotlib.pyplot as plt
```

Zadanie 37

In [87]:
```
%matplotlib inline
```

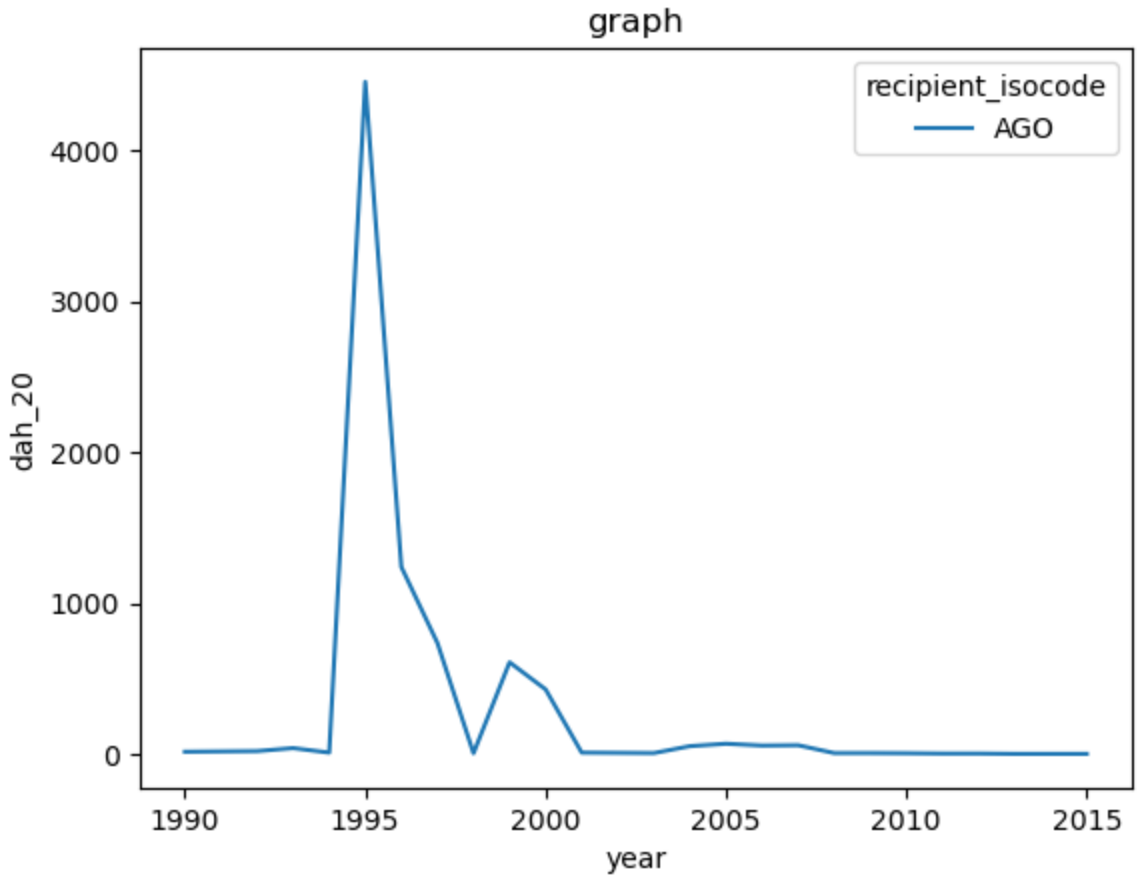
Zadanie 38

In [96]:
```
df['dah_20'] = df['dah_20'].replace('-', '0')
df['dah_20'] = df['dah_20'].astype(float)
```

```
df['rmh_fp_dah_20'] = df['rmh_fp_dah_20'].replace('-', '0')
df['rmh_fp_dah_20'] = df['rmh_fp_dah_20'].astype(float)
```

In [109…
```
df[(df['channel'] == 'BIL_AUS')
   & (df['source'] == 'Australia')
   & (df['recipient_isocode'] == 'AGO')].pivot_table(values='dah_20', index='year', colu
                                                     aggfunc='mean',
                                                fill_value=None, margins=False, dropna=True).p
plt.ylabel('dah_20')
plt.title('graph')
```

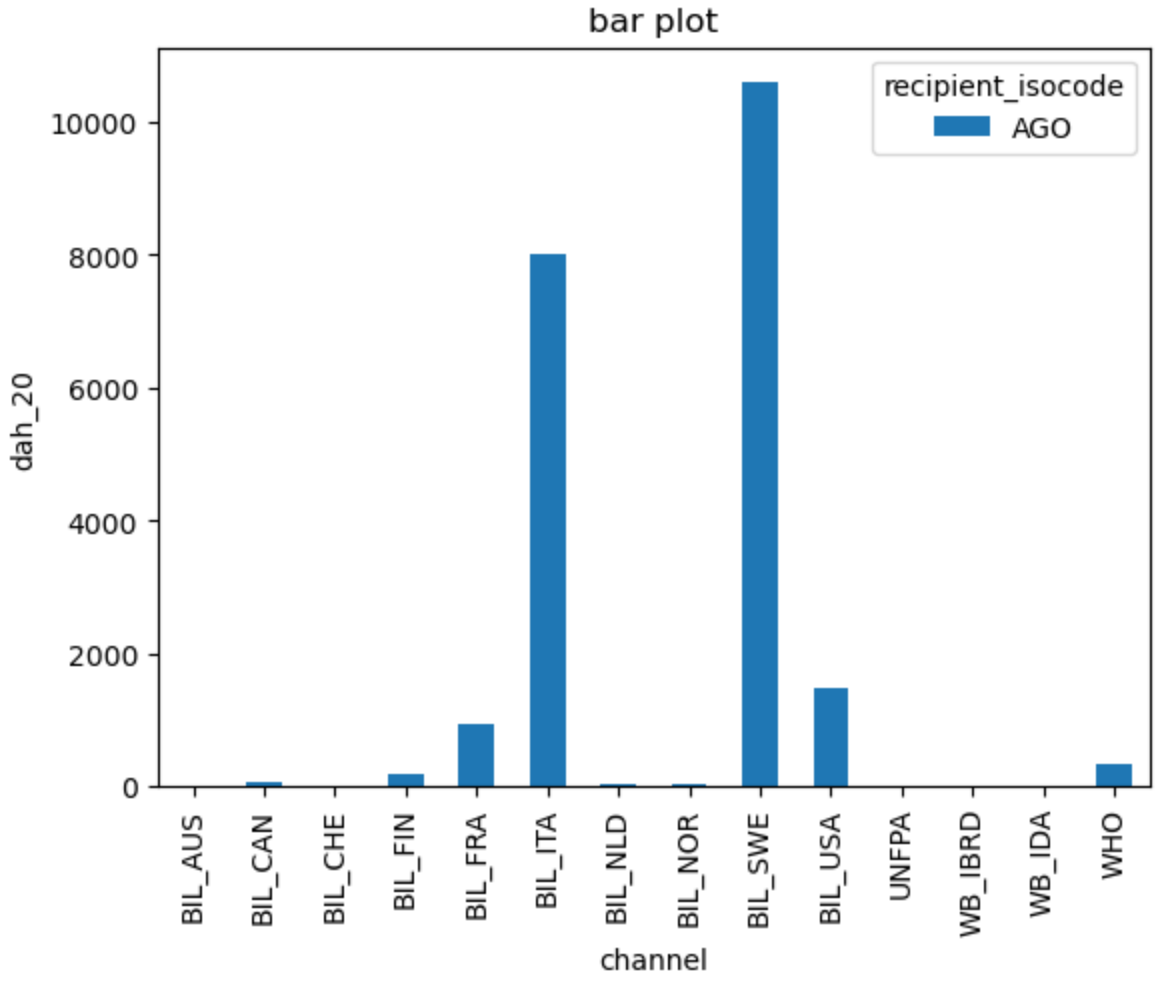
Out[109]:
```
Text(0.5, 1.0, 'graph')
```



Zadanie 39

In [113…
```
df_bar = df[(df['recipient_isocode'].isin(['AGO'])) & (df['year'] == 1990)].pivot_table(
                index='channel', columns='recipient_isocode', aggfunc='mean',
                      fill_value=None, margins=False, dropna=True)
df_bar.plot(kind = 'bar')
plt.ylabel('dah_20')
plt.title('bar plot')
```

Out[113]:
```
Text(0.5, 1.0, 'bar plot')
```



Zadanie 40

In [118…
```
df1 = pd.read_csv("IHME-GBD_2019_DATA-15798851-2.csv", encoding='utf-8', engine='python'
df2 = pd.read_csv("IHME-GBD_2019_DATA-ff08d9bc-1.csv", encoding='utf-8', engine='python'
```

In [119…
```
df1.head()
```

Out[119]:

	measure	location	sex	age	cause	metric	year	val	upper	lower
0	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Maternal and neonatal disorders	Rate	2012	7475.212700	9104.773541	6157.428603
1	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Maternal and neonatal disorders	Rate	2012	7814.344518	9667.960848	6289.146375
2	DALYs (Disability-Adjusted Life Years)	Gambia	Male	All Ages	Substance use disorders	Number	2012	1659.038707	2126.829521	1239.172699
3	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Substance use disorders	Number	2012	874.432466	1186.560596	618.271780
4	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Substance use disorders	Number	2012	2533.471173	3231.220866	1868.204609

In [120…
```
df2.head()
```

Out[120]:

	measure	location	sex	age	cause	metric	year	val	upper	lower
0	Deaths	Samoa	Male	All Ages	Chronic respiratory diseases	Rate	2000	64.470214	81.808307	53.476793
1	Deaths	Samoa	Female	All Ages	Chronic respiratory diseases	Rate	2000	55.234399	77.883497	39.978647
2	Deaths	Samoa	Both	All Ages	Chronic respiratory diseases	Rate	2000	60.039961	76.013089	49.941986
3	Deaths	Samoa	Male	All Ages	Skin and subcutaneous diseases	Rate	2000	2.246741	3.467454	1.438979
4	Deaths	Samoa	Female	All Ages	Skin and subcutaneous diseases	Rate	2000	1.368385	1.945448	0.866099

In [122…
```
df1.rename(columns = {'val': 'val1', 'upper':'upper1', 'lower':'lower1'}, inplace = True
df2.rename(columns = {'val': 'val2', 'upper': 'upper2', 'lower':'lower2'}, inplace = Tru
df_both = pd.merge(df1, df2, on = ['location', 'sex', 'age', 'cause', 'metric', 'year'],
df_both.head()
```

Out[122]:

	measure_x	location	sex	age	cause	metric	year	val1	upper1	lower1	measure_y
0	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Maternal and neonatal disorders	Rate	2012	7475.212700	9104.773541	6157.428603	Death
1	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Maternal and neonatal disorders	Rate	2012	7814.344518	9667.960848	6289.146375	Death
2	DALYs (Disability-Adjusted Life Years)	Gambia	Male	All Ages	Substance use disorders	Rate	2012	179.493660	230.104586	134.067784	Death
3	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Substance use disorders	Rate	2012	91.400543	124.025911	64.625204	Death
4	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Substance use disorders	Rate	2012	134.688036	171.782808	99.320179	Death

In [125…
```
#concat
df_part1 = df_both.iloc[:50000,:]
df_part2 = df_both.iloc[50000:,:]
```

In [128…
```
df_both_2 = pd.concat([df_part1, df_part2], axis = 0)
df_both_2.head()
```

Out[128]:

	measure_x	location	sex	age	cause	metric	year	val1	upper1	lower1	measure_y
0	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Maternal and neonatal disorders	Rate	2012	7475.212700	9104.773541	6157.428603	Death

1	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Maternal and neonatal disorders	Rate	2012	7814.344518	9667.960848	6289.146375	Death
2	DALYs (Disability-Adjusted Life Years)	Gambia	Male	All Ages	Substance use disorders	Rate	2012	179.493660	230.104586	134.067784	Death
3	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Substance use disorders	Rate	2012	91.400543	124.025911	64.625204	Death
4	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Substance use disorders	Rate	2012	134.688036	171.782808	99.320179	Death

Zadanie 41

In [130…
```
#nowa kolumna będącą zaokrągleniem innej
df_both["val2_round"] = df_both["val2"].round(decimals = 1)
df_both.head()
```

Out[130]:

	measure_x	location	sex	age	cause	metric	year	val1	upper1	lower1	measure_y
0	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Maternal and neonatal disorders	Rate	2012	7475.212700	9104.773541	6157.428603	Death
1	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Maternal and neonatal disorders	Rate	2012	7814.344518	9667.960848	6289.146375	Death
2	DALYs (Disability-Adjusted Life Years)	Gambia	Male	All Ages	Substance use disorders	Rate	2012	179.493660	230.104586	134.067784	Death
3	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Substance use disorders	Rate	2012	91.400543	124.025911	64.625204	Death
4	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Substance use disorders	Rate	2012	134.688036	171.782808	99.320179	Death

In [131…
```
# nowa kolumna będąca sumą innych kolumn
df_both["sum"] = df_both["val2"] + df_both["upper2"] + df_both["upper1"]
df_both.head()
```

Out[131]:

	measure_x	location	sex	age	cause	metric	year	val1	upper1	lower1	measure_y
0	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Maternal and neonatal disorders	Rate	2012	7475.212700	9104.773541	6157.428603	Death

1	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Maternal and neonatal disorders	Rate	2012	7814.344518	9667.960848	6289.146375	Death
2	DALYs (Disability-Adjusted Life Years)	Gambia	Male	All Ages	Substance use disorders	Rate	2012	179.493660	230.104586	134.067784	Death
3	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Substance use disorders	Rate	2012	91.400543	124.025911	64.625204	Death
4	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Substance use disorders	Rate	2012	134.688036	171.782808	99.320179	Death

In [132…

```
# nowa kolumna poprzez sumowanie wartości kilku innych kolumn metodą loc
df_both["sum2"] = df_both.loc[:, "val2":"lower2"].sum(axis = 1)
df_both.head()
```

Out[132]:

	measure_x	location	sex	age	cause	metric	year	val1	upper1	lower1	measure_y
0	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Maternal and neonatal disorders	Rate	2012	7475.212700	9104.773541	6157.428603	Death
1	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Maternal and neonatal disorders	Rate	2012	7814.344518	9667.960848	6289.146375	Death
2	DALYs (Disability-Adjusted Life Years)	Gambia	Male	All Ages	Substance use disorders	Rate	2012	179.493660	230.104586	134.067784	Death
3	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Substance use disorders	Rate	2012	91.400543	124.025911	64.625204	Death
4	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Substance use disorders	Rate	2012	134.688036	171.782808	99.320179	Death

In [133…

```
# nowa kolumna poprzez sumowanie wartości kilku innych kolumn metodą loc
# od 11 do 13
df_both["sum2_2"] = df_both.iloc[:, 11:14].sum(axis = 1)
df_both.head()
```

Out[133]:

	measure_x	location	sex	age	cause	metric	year	val1	upper1	lower1	measure_y
0	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Maternal and neonatal disorders	Rate	2012	7475.212700	9104.773541	6157.428603	Death
1	DALYs (Disability-	Gambia	Both	All Ages	Maternal and	Rate	2012	7814.344518	9667.960848	6289.146375	Death

	Adjusted Life Years)				neonatal disorders						
2	DALYs (Disability-Adjusted Life Years)	Gambia	Male	All Ages	Substance use disorders	Rate	2012	179.493660	230.104586	134.067784	Death
3	DALYs (Disability-Adjusted Life Years)	Gambia	Female	All Ages	Substance use disorders	Rate	2012	91.400543	124.025911	64.625204	Death
4	DALYs (Disability-Adjusted Life Years)	Gambia	Both	All Ages	Substance use disorders	Rate	2012	134.688036	171.782808	99.320179	Death

In [134…
```
df_both.drop(['val2_round','sum', 'sum2','sum2_2'], axis = 1, inplace = True)
```

Zadanie 42

In [135…
```
c = ['Poland', 'Czech Republic', 'Italy', 'Sweden', 'Norway', 'Spain', 'Romania']
```

In [138…
```
# nowa kolumna c
# jeżeli kraj istnieje w liście c to true, inaczej false
df_both['c'] = df_both['location'].apply(lambda x: True if x in c else False )
df_both[df_both['c'] == True]
```

Out[138]:

	measure_x	location	sex	age	cause	metric	year	val1	upper1	lower1
29	DALYs (Disability-Adjusted Life Years)	Sweden	Male	All Ages	Skin and subcutaneous diseases	Rate	2013	596.675878	854.436296	416.110131
30	DALYs (Disability-Adjusted Life Years)	Sweden	Female	All Ages	Skin and subcutaneous diseases	Rate	2013	792.975647	1141.436752	545.865970
31	DALYs (Disability-Adjusted Life Years)	Sweden	Both	All Ages	Skin and subcutaneous diseases	Rate	2013	694.964234	998.138978	480.145444
89	DALYs (Disability-Adjusted Life Years)	Sweden	Male	All Ages	Musculoskeletal disorders	Rate	2013	2265.334283	2989.698760	1628.363261
90	DALYs (Disability-Adjusted Life Years)	Sweden	Female	All Ages	Musculoskeletal disorders	Rate	2013	3216.084604	4226.056542	2337.632460
...	...	...	...	...	...	...	...	...	...	...
93741	DALYs (Disability-Adjusted Life Years)	Romania	Female	All Ages	Substance use disorders	Rate	2019	295.064646	389.845097	216.597062
93742	DALYs (Disability-	Romania	Both	All Ages	Substance use disorders	Rate	2019	418.790624	531.482484	325.675656

	Adjusted Life Years)									
93911	DALYs (Disability-Adjusted Life Years)	Poland	Male	All Ages	Neglected tropical diseases and malaria	Rate	2019	56.462532	87.853853	34.154091
93912	DALYs (Disability-Adjusted Life Years)	Poland	Female	All Ages	Neglected tropical diseases and malaria	Rate	2019	62.473695	91.811053	38.948680
93913	DALYs (Disability-Adjusted Life Years)	Poland	Both	All Ages	Neglected tropical diseases and malaria	Rate	2019	59.563301	89.349730	36.535688

2856 rows × 15 columns

In [139…
```
# krótszy zapis
df_both['c'] = df_both['location'].apply(lambda x: x in c)
```

In [140…
```
# wybranie wierszy pasujących do warunku
df_both[df_both['c'] == True]
```

Out[140]:

	measure_x	location	sex	age	cause	metric	year	val1	upper1	lower1
29	DALYs (Disability-Adjusted Life Years)	Sweden	Male	All Ages	Skin and subcutaneous diseases	Rate	2013	596.675878	854.436296	416.110131
30	DALYs (Disability-Adjusted Life Years)	Sweden	Female	All Ages	Skin and subcutaneous diseases	Rate	2013	792.975647	1141.436752	545.865970
31	DALYs (Disability-Adjusted Life Years)	Sweden	Both	All Ages	Skin and subcutaneous diseases	Rate	2013	694.964234	998.138978	480.145444
89	DALYs (Disability-Adjusted Life Years)	Sweden	Male	All Ages	Musculoskeletal disorders	Rate	2013	2265.334283	2989.698760	1628.363261
90	DALYs (Disability-Adjusted Life Years)	Sweden	Female	All Ages	Musculoskeletal disorders	Rate	2013	3216.084604	4226.056542	2337.632460
...	...	...	...	...	...	...	...	...	...	...
93741	DALYs (Disability-Adjusted Life Years)	Romania	Female	All Ages	Substance use disorders	Rate	2019	295.064646	389.845097	216.597062
93742	DALYs (Disability-Adjusted Life Years)	Romania	Both	All Ages	Substance use disorders	Rate	2019	418.790624	531.482484	325.675656
93911	DALYs (Disability-	Poland	Male	All Ages	Neglected tropical	Rate	2019	56.462532	87.853853	34.154091

	Adjusted Life Years)				diseases and malaria					
93912	DALYs (Disability-Adjusted Life Years)	Poland	Female	All Ages	Neglected tropical diseases and malaria	Rate	2019	62.473695	91.811053	38.948680
93913	DALYs (Disability-Adjusted Life Years)	Poland	Both	All Ages	Neglected tropical diseases and malaria	Rate	2019	59.563301	89.349730	36.535688

2856 rows × 15 columns

In [141…
```
# warunek jako funkcja
def if_c(x):
    if x in c:
        return True
    else:
        False
```

In [142…
```
df_both['c'] = df_both['location'].apply(if_c)
```

In [143…
```
df_both[df_both['c'] == True]
```

Out[143]:

	measure_x	location	sex	age	cause	metric	year	val1	upper1	lower1
29	DALYs (Disability-Adjusted Life Years)	Sweden	Male	All Ages	Skin and subcutaneous diseases	Rate	2013	596.675878	854.436296	416.110131
30	DALYs (Disability-Adjusted Life Years)	Sweden	Female	All Ages	Skin and subcutaneous diseases	Rate	2013	792.975647	1141.436752	545.865970
31	DALYs (Disability-Adjusted Life Years)	Sweden	Both	All Ages	Skin and subcutaneous diseases	Rate	2013	694.964234	998.138978	480.145444
89	DALYs (Disability-Adjusted Life Years)	Sweden	Male	All Ages	Musculoskeletal disorders	Rate	2013	2265.334283	2989.698760	1628.363261
90	DALYs (Disability-Adjusted Life Years)	Sweden	Female	All Ages	Musculoskeletal disorders	Rate	2013	3216.084604	4226.056542	2337.632460
...	...	...	...	...	...	...	...	...	...	...
93741	DALYs (Disability-Adjusted Life Years)	Romania	Female	All Ages	Substance use disorders	Rate	2019	295.064646	389.845097	216.597062
93742	DALYs (Disability-Adjusted Life Years)	Romania	Both	All Ages	Substance use disorders	Rate	2019	418.790624	531.482484	325.675656
93911	DALYs	Poland	Male	All	Neglected	Rate	2019	56.462532	87.853853	34.154091

	(Disability-Adjusted Life Years)			Ages	tropical diseases and malaria					
93912	DALYs (Disability-Adjusted Life Years)	Poland	Female	All Ages	Neglected tropical diseases and malaria	Rate	2019	62.473695	91.811053	38.948680
93913	DALYs (Disability-Adjusted Life Years)	Poland	Both	All Ages	Neglected tropical diseases and malaria	Rate	2019	59.563301	89.349730	36.535688

2856 rows × 15 columns

Zadanie 43

In [144…

```
df_both.to_csv('df_both.csv')
# podział dużego pliku przy użycio chunksize

for chunk_df in pd.read_csv('df_both.csv',
                        chunksize = 50000):
    print("CHUNK DF")
    print(chunk_df.head())
```

```
CHUNK DF
   Unnamed: 0                                 measure_x location     sex  \
0           0  DALYs (Disability-Adjusted Life Years)   Gambia  Female
1           1  DALYs (Disability-Adjusted Life Years)   Gambia    Both
2           2  DALYs (Disability-Adjusted Life Years)   Gambia    Male
3           3  DALYs (Disability-Adjusted Life Years)   Gambia  Female
4           4  DALYs (Disability-Adjusted Life Years)   Gambia    Both

        age                              cause metric  year         val1  \
0  All Ages  Maternal and neonatal disorders   Rate  2012  7475.212700
1  All Ages  Maternal and neonatal disorders   Rate  2012  7814.344518
2  All Ages          Substance use disorders   Rate  2012   179.493660
3  All Ages          Substance use disorders   Rate  2012    91.400543
4  All Ages          Substance use disorders   Rate  2012   134.688036

        upper1       lower1 measure_y       val2      upper2     lower2    c
0  9104.773541  6157.428603    Deaths  89.867238  110.176286  74.001324  NaN
1  9667.960848  6289.146375    Deaths  89.365457  111.497307  71.539316  NaN
2   230.104586   134.067784    Deaths   1.101136    1.545407   0.786638  NaN
3   124.025911    64.625204    Deaths   0.227136    0.296917   0.173921  NaN
4   171.782808    99.320179    Deaths   0.656605    0.896138   0.490594  NaN
CHUNK DF
       Unnamed: 0                                 measure_x location     sex  \
50000       50000  DALYs (Disability-Adjusted Life Years)    India    Both
50001       50001  DALYs (Disability-Adjusted Life Years)    India    Male
50002       50002  DALYs (Disability-Adjusted Life Years)    India  Female
50003       50003  DALYs (Disability-Adjusted Life Years)    India    Both
50004       50004  DALYs (Disability-Adjusted Life Years)   Turkey    Male

           age                              cause metric  year         val1  \
50000  All Ages        Other infectious diseases   Rate  2016   958.958618
50001  All Ages  Maternal and neonatal disorders   Rate  2016  3857.092523
50002  All Ages  Maternal and neonatal disorders   Rate  2016  3936.244203
50003  All Ages  Maternal and neonatal disorders   Rate  2016  3895.617080
50004  All Ages              Transport injuries   Rate  2016   906.078397

           upper1       lower1 measure_y       val2     upper2     lower2  \
50000  1253.212194   758.548648    Deaths  14.064804  17.711690  11.444635
```

```
50001  4559.365099  3248.746140    Deaths  39.804065  47.428130  33.085022
50002  4575.786077  3357.759508    Deaths  42.591026  49.826017  36.112185
50003  4578.291899  3318.975051    Deaths  41.160530  48.311924  34.780916
50004  1044.082153   685.161081    Deaths  16.515202  19.376146  11.913822

         c
50000  NaN
50001  NaN
50002  NaN
50003  NaN
50004  NaN
```

In [147…

```
# zastosowanie metody groupby oddzielnie do każdej części, a następnie połączenie wyniku
new_df = pd.DataFrame()
for chunk_df in pd.read_csv('df_both.csv',
                            chunksize = 50000):
    result = chunk_df.groupby(['cause', 'year']).agg({'val2': 'mean',
                                                      'val1': 'mean'})
    new_df = pd.concat([new_df,result])
new_df
```

Out[147]:

		val2	val1
cause	**year**		
Cardiovascular diseases	**2011**	112.889873	2945.949359
	2012	238.128059	5118.690134
	2013	236.486609	5005.051157
	2014	242.057921	5088.816632
	2015	253.292782	5250.955897
...	**...**	...	...
Unintentional injuries	**2015**	24.622137	1575.325187
	2016	21.095357	1389.698018
	2017	22.527506	1467.601676
	2018	21.578766	1428.203221
	2019	21.508796	1417.611000

236 rows × 2 columns