```
import pandas as pd
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))

cols_to_convert = df_csv.columns[14:]
df_csv[cols_to_convert] = df_csv[cols_to_convert].apply(pd.to_numeric,
errors='coerce')

print(df_csv.head())
```

```
   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  ...          0.0
5.0
1   Sub-Saharan Africa, Eastern          174.0  ...          0.0
6.0
2   Sub-Saharan Africa, Western          199.0  ...          0.0
6.0
3   Sub-Saharan Africa, Western          199.0  ...          0.0
5.0
4  Sub-Saharan Africa, Southern          192.0  ...          0.0
1.0

   nch_dah_20  ncd_dah_20  hiv_dah_20  mal_dah_20  tb_dah_20  \
0         0.0         0.0         7.0         3.0        0.0
1         0.0         0.0         5.0         1.0        0.0
2         0.0         0.0         5.0         2.0        0.0
3         0.0         0.0         7.0         2.0        0.0
4         0.0         0.0        23.0         NaN        0.0

   swap_hss_total_dah_20  oid_dah_20  unalloc_dah_20
0                    0.0         0.0             NaN
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             NaN

[5 rows x 76 columns]
```

```
print(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
None

print(df_csv.describe())

                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]
```

```
mean_dah = df_csv["dah_20"].mean()
print(f"Srednia pomoc w latach 1990-2020 wyniosła w k$ {mean_dah}")
```

```
Srednia pomoc w latach 1990-2020 wyniosła 2519.3625352650447
```

```
df_csv.dropna(subset=["dah_20"], inplace=True)
median_dah = df_csv["dah_20"].median()
print(f"Mediana pomocy w obszarze zdrowia wyniosła {median_dah}")
```

```
Mediana pomocy w obszarze zdrowia wyniosła 44.0
```

```
std_dah = df_csv["dah_20"].std()
print(f"Odchylenie standardowe pomocy w obszarze zdrowia wyniosło
{std_dah}")
```

```
Odchylenie standardowe pomocy w obszarze zdrowia wyniosło
36693.31180470732
```

```
missing_values = df_csv.isnull().sum()
print("Brakujące wartości w każdej kolumnie:")
print(missing_values)
```

```
Brakujące wartości w każdej kolumnie:
year                          0
source                        0
channel                       0
recipient_isocode             0
recipient_country           521
                          ...
mal_dah_20                14960
tb_dah_20                 15339
swap_hss_total_dah_20     16049
oid_dah_20                21227
unalloc_dah_20           140458
Length: 76, dtype: int64
```

```
#wypełnianie wartości brakujących wartościami średnimi w przypadku
tego datasetu nie ma sensu
df_csv.dropna(subset=["hiv_dah_20"], inplace=True)

Q1 = df_csv["dah_20"].quantile(0.25)
Q3 = df_csv["dah_20"].quantile(0.75)
IQR = Q3-Q1
outliers = df_csv[(df_csv["dah_20"] < (Q1-1.5*IQR)) |
(df_csv["dah_20"] > (Q3+1.5*IQR))]
print("Wartości odstające")
print(outliers)
```

```
Wartości odstające
        year         source  channel recipient_isocode  \
15      1990      Australia  BIL_AUS               ERI
26      1990      Australia  BIL_AUS            INKIND
49      1990      Australia  BIL_AUS               PNG
65      1990      Australia  BIL_AUS               TLS
70      1990      Australia  BIL_AUS               VNM
...      ...            ...      ...               ...
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  \
15                      Eritrea              178           SSA
26      Administrative expenses            44598           NaN
49             Papua New Guinea               26           EAP
65                  Timor-Leste               19           EAP
70                      Vietnam               20           EAP
...                         ...              ...           ...
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  ...  \
15                 242  Sub-Saharan Africa, Eastern
174.0  ...
26               44621      Administrative expenses
44598.0  ...
49                 239                      Oceania
21.0  ...
65                 239              Asia, Southeast
9.0  ...
70                 239              Asia, Southeast
9.0  ...
...                ...                          ...            ...  ..
.
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  \
15                0.0         NaN         0.0         0.0         5.0
26              302.0       799.0       347.0        46.0       384.0
49              415.0       769.0        47.0        47.0       478.0
65                4.0         7.0         0.0         0.0         8.0
70               42.0        21.0      2023.0         0.0       334.0
...               ...         ...         ...         ...         ...
384301            0.0     17509.0    265811.0         0.0      9398.0
384302            0.0         0.0         0.0         0.0      1037.0
384303            0.0         0.0         0.0         0.0      1250.0
384304        49485.0     17337.0     26498.0      1907.0      5076.0
384305        17823.0     13250.0     32712.0     22436.0     11098.0

        mal_dah_20  tb_dah_20  swap_hss_total_dah_20  oid_dah_20  \
15             NaN        0.0                    0.0         0.0
26            15.0        0.0                   71.0         0.0
49             0.0        0.0                    2.0         0.0
65             1.0        0.0                    1.0         0.0
70             5.0        0.0                   88.0         0.0
...            ...        ...                    ...         ...
384301         0.0        0.0                    0.0    116214.0
```

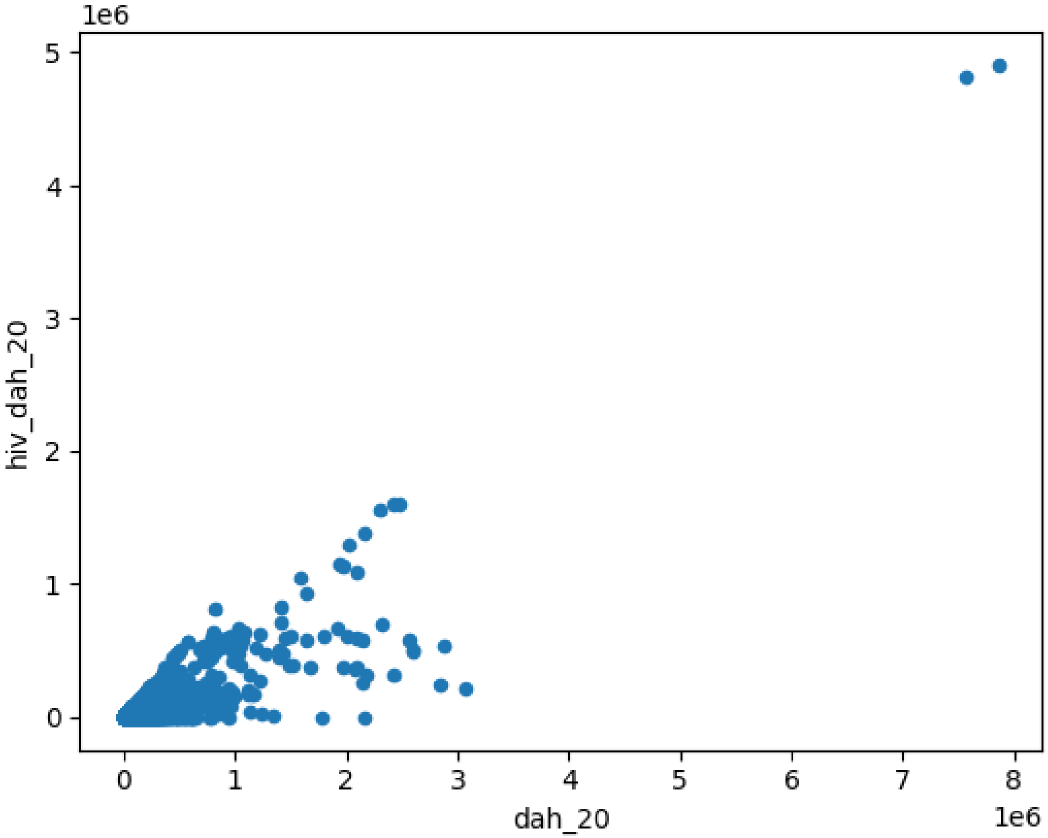
```
384302      1085.0      895.0                  449.0       407.0
384303       166.0      112.0                  173.0       217.0
384304       763.0     3424.0                46292.0     47569.0
384305      8991.0     9457.0               193266.0    113937.0

        unalloc_dah_20
15              1258.0
26              1384.0
49                 0.0
65              9916.0
70                 0.0
...                ...
384301             NaN
384302             NaN
384303             0.0
384304             NaN
384305             0.0

[49469 rows x 76 columns]
```

```
correlation = df_csv['dah_20'].corr(df_csv['hiv_dah_20'])
print("Współczynnik korelacji:", correlation)
```

```
Współczynnik korelacji: 0.8308245442028445
```

```
df_csv.plot.scatter(x = 'dah_20', y = 'hiv_dah_20')
```

```
<Axes: xlabel='dah_20', ylabel='hiv_dah_20'>
```



```python
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'])

grouped = df_csv.groupby('recipient_country')
['dah_20'].mean(numeric_only=True)
print('srednia pomoc dla danego kraju')
print(grouped)
```

```
srednia pomoc dla danego kraju
recipient_country
Administrative expenses      11703.161571
Afghanistan                   1736.564483
Albania                        281.901707
Algeria                         82.570909
Angola                         664.607132
                                 ...
Vietnam                       1669.274224
Wallis and Futuna Islands      456.464286
Yemen                          859.378356
Zambia                        2222.785955
Zimbabwe                      1506.333514
Name: dah_20, Length: 174, dtype: float64
```

```python
df_sorted = df_csv.sort_values(by = 'dah_20', ascending=False)
print("Dane posortowane według wysokości pomocy")
print(df_sorted.head())
```

```
Dane posortowane według wysokości pomocy
        year         source  channel recipient_isocode  \
384290  2020  United_States  BIL_USA               QZA
383776  2019  United_States  BIL_USA               QZA
383827  2020           BMGF     BMGF               QZA
361639  2017  Private_other      NGO               QZA
383276  2019           BMGF     BMGF               QZA

              recipient_country  gbd_location_id wb_regioncode  \
384290  Unallocated/Unspecified            44598           NaN
383776  Unallocated/Unspecified            44598           NaN
383827  Unallocated/Unspecified            44598           NaN
361639  Unallocated/Unspecified            44598           NaN
383276  Unallocated/Unspecified            44598           NaN

        wb_location_id               gbd_region  gbd_region_id  ...  \
384290           44621  Unallocated/Unspecified        44598.0  ...
383776           44621  Unallocated/Unspecified        44598.0  ...
383827           44621  Unallocated/Unspecified        44598.0  ...
361639           44621  Unallocated/Unspecified        44598.0  ...
383276           44621  Unallocated/Unspecified        44598.0  ...

       rmh_dah_20  nch_dah_20  ncd_dah_20  hiv_dah_20  mal_dah_20
tb_dah_20  \
384290   859482.0    229677.0      8207.0   4906421.0    584684.0
167017.0
383776   853505.0    222341.0      8021.0   4809268.0    577341.0
163227.0
383827   307506.0    728728.0     55599.0    220204.0    256736.0
120803.0
361639   363802.0    701796.0    109415.0    536386.0    102369.0
44925.0
383276   334782.0    718327.0     51075.0    245602.0    241744.0
123794.0

        swap_hss_total_dah_20  oid_dah_20  unalloc_dah_20  \
384290                76960.0    431096.0             0.0
383776                68938.0    259510.0             NaN
383827               194442.0    584443.0             NaN
361639               185663.0    109757.0             NaN
383276               185972.0    278640.0             NaN

        other_diseases_dah_20
384290               263288.0
383776               243166.0
383827               614706.0
```

```
361639               151777.0
383276               604162.0

[5 rows x 77 columns]
```