

```
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))
```

```
df_csv.head(10)
```

	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
2	0		0	0
3	0		0	0
4	0		0	-
5	0		0	-
6	0	130	0	0
7	0		0	0
8	0		0	-
9	0		0	0

[10 rows x 76 columns]

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

```
import numpy as np
```

```
mean_val = np.mean(df_csv["dah_20"])
```

```
print(mean_val)
```

2519.3625352650447

```
df_csv.dropna(subset=["dah_20"], inplace=True)
```

```
median_val = np.median(df_csv["dah_20"])
```

```
print(median_val)
```

44.0

```
std_dev = np.std(df_csv["dah_20"])
```

```
print(std_dev)
```

36693.2562089067

```
variance = np.var(df_csv["dah_20"])  
print(variance)
```

1346395051.2124703

```
df_csv.dropna(subset=["other_dah_20"], inplace=True)  
data_1 = df_csv["other_dah_20"]  
data_2 = df_csv["dah_20"]  
correlation = np.corrcoef(data_1, data_2)[0, 1]  
print(correlation)
```

0.753662974820274

```
covariance = np.cov(data_1, data_2)[0, 1]  
print(covariance)
```

213078437.0006457