

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from scipy.stats import skew, kurtosis

# To display plots inline in the notebook
%matplotlib inline

# Set plot styles
sns.set(style="whitegrid")

#Loading data into a DataFrame
data=pd.read_csv('drive/MyDrive/Dataset.csv')
df = pd.DataFrame(data)
df.drop(['Event', 'Time','file_number','event_number'] , axis=1, inplace=True)
#df.drop('Age' , axis=1, inplace=True)

#df.drop(['Event', 'Time','file_number','event_number'] , axis=1)
df.info()

→ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 265627 entries, 0 to 265626
Data columns (total 38 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Cell_length  265627 non-null   int64  
 1   DNA1         265627 non-null   float64 
 2   DNA2         265627 non-null   float64 
 3   CD45RA       265627 non-null   float64 
 4   CD133        265627 non-null   float64 
 5   CD19         265627 non-null   float64 
 6   CD22         265627 non-null   float64 
 7   CD11b        265627 non-null   float64 
 8   CD4          265627 non-null   float64 
 9   CD8          265627 non-null   float64 
 10  CD34         265627 non-null   float64 
 11  Flt3         265627 non-null   float64 
 12  CD20         265627 non-null   float64 
 13  CXCR4        265627 non-null   float64 
 14  CD235ab      265627 non-null   float64 
 15  CD45         265627 non-null   float64 
 16  CD123        265627 non-null   float64 
 17  CD321        265627 non-null   float64 
 18  CD14         265627 non-null   float64 
 19  CD33         265627 non-null   float64 
 20  CD47         265627 non-null   float64 
 21  CD11c        265627 non-null   float64 
 22  CD7          265627 non-null   float64 
 23  CD15         265627 non-null   float64 
 24  CD16         265627 non-null   float64 
 25  CD44         265627 non-null   float64 
 26  CD38         265627 non-null   float64 
 27  CD13         265627 non-null   float64 
 28  CD3          265627 non-null   float64 
 29  CD61         265627 non-null   float64 
 30  CD117        265627 non-null   float64 
 31  CD49d        265627 non-null   float64 
 32  HLA-DR       265627 non-null   float64 
 33  CD64         265627 non-null   float64 
 34  CD41         265627 non-null   float64 
 35  Viability    265627 non-null   float64 
 36  label         104184 non-null   float64 
 37  individual    265627 non-null   int64  
dtypes: float64(36), int64(2)
memory usage: 77.0 MB

print(df.shape)

→ (265627, 38)

print(df.describe())
```

	Cell_length	DNA1	DNA2	CD45RA	\	
count	265627.000000	265627.000000	265627.000000	265627.000000		
mean	34.450572	4.606956	5.198308	0.688127		
std	11.446694	1.312831	1.150357	0.609105		
min	10.000000	2.786488	2.236450	-0.057305		
25%	26.000000	3.700023	4.407822	0.204625		
50%	33.000000	4.022127	4.698415	0.549387		
75%	41.000000	6.353313	6.766268	1.031198		
max	65.000000	7.001489	7.472308	6.691197		
	CD133	CD19	CD22	CD11b	\	
count	265627.000000	265627.000000	265627.000000	265627.000000		
mean	0.145960	0.509301	0.397323	0.710319		
std	0.259267	0.857462	0.762126	1.011434		
min	-0.058081	-0.058089	-0.057342	-0.058236		
25%	-0.022935	-0.018838	-0.020689	-0.000294		
50%	0.025353	0.075210	0.058790	0.257923		
75%	0.224299	0.548386	0.386481	0.923517		
max	5.527494	4.990085	5.160477	5.260789		
	CD4	CD8	...	CD3	CD61	\
count	265627.000000	265627.000000	...	265627.000000	265627.000000	
mean	0.367597	0.565222	...	2.165756	0.344455	
std	0.517376	1.003697	...	2.268698	0.724138	
min	-0.057751	-0.058003	...	-0.058241	-0.057642	
25%	-0.012590	-0.017322	...	0.084946	-0.012853	
50%	0.131218	0.073634	...	0.603755	0.095691	
75%	0.578119	0.486421	...	4.665223	0.415791	
max	6.581762	4.693694	...	6.748362	7.748498	
	CD117	CD49d	HLA-DR	CD64	\	
count	265627.000000	265627.000000	265627.000000	265627.000000		
mean	0.131199	0.794938	1.521812	0.551512		
std	0.313208	0.627619	1.694211	0.888739		
min	-0.057668	-0.058064	-0.057974	-0.058199		
25%	-0.023957	0.283013	0.057709	-0.010582		
50%	-0.000410	0.677212	0.611335	0.122493		
75%	0.154736	1.190787	2.888240	0.604131		
max	5.502125	5.153438	7.052507	4.517843		
	CD41	Viability	label	individual		
count	265627.000000	265627.000000	104184.000000	265627.000000		
mean	0.261754	0.570037	8.116102	1.279625		
std	0.617065	0.589738	2.457486	0.448816		
min	-0.058244	-0.057979	1.000000	1.000000		
25%	-0.020166	0.065523	7.000000	1.000000		
50%	0.052229	0.398230	8.000000	1.000000		
75%	0.305591	0.931058	10.000000	2.000000		
max	7.718288	2.433031	14.000000	2.000000		

[8 rows x 38 columns]

Start coding or [generate](#) with AI.

```
null_counts = df.isnull().sum()
non_null_counts = df.notnull().sum()

# Create a new DataFrame for plotting
counts_df = pd.DataFrame({
    'Null': null_counts,
    'Non-Null': non_null_counts
})

# Plot the stacked bar chart
counts_df.plot(kind='bar', stacked=True)

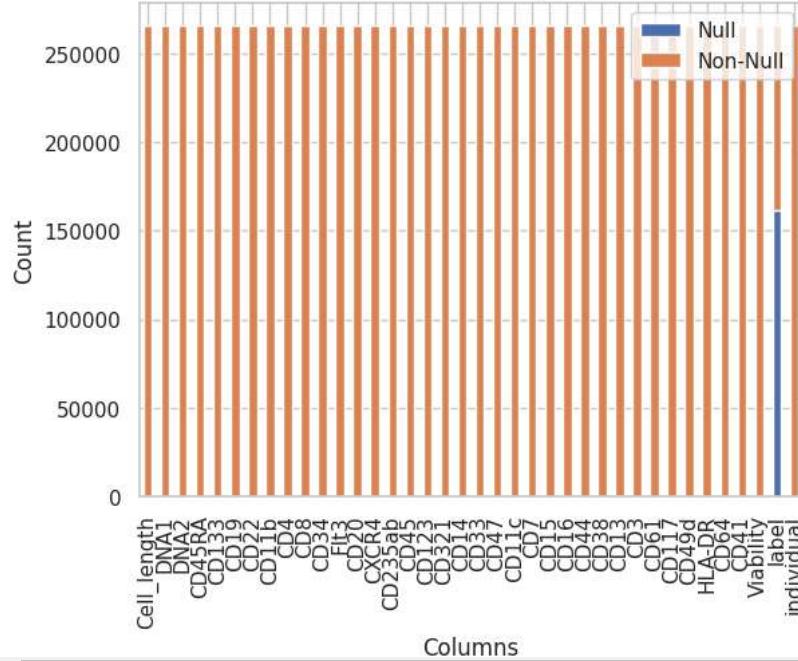
# Add title and labels
plt.title('Null vs Non-Null Values per Column')
plt.ylabel('Count')
plt.xlabel('Columns')

# Rotate the x-axis labels for better readability
plt.xticks(rotation=90)

# Show the plot
plt.show()
```

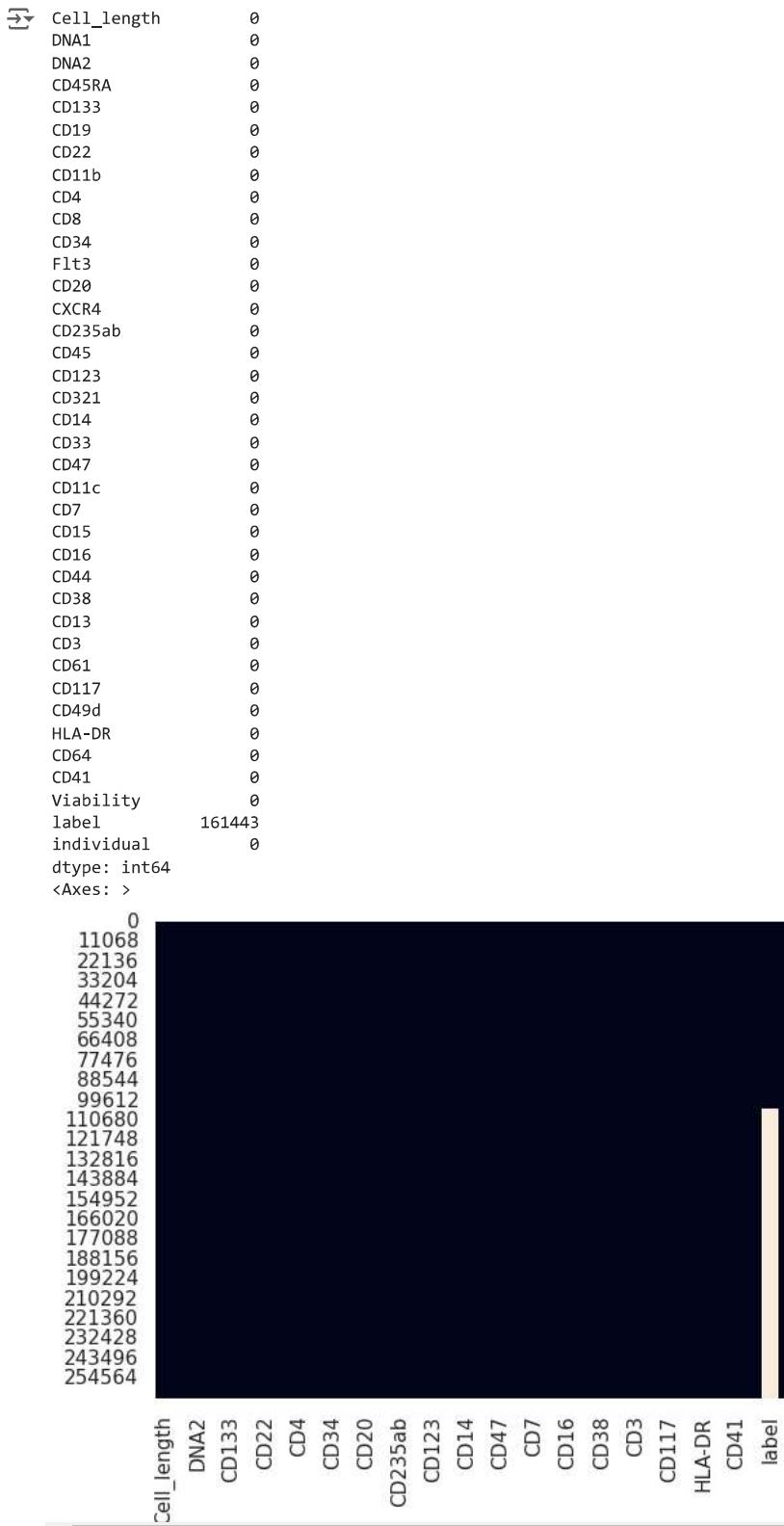


Null vs Non-Null Values per Column



```
print(df.isnull().sum())
```

```
sns.heatmap(df.isnull(), cbar=False)
```



Start coding or generate with AI.

```
# Cell 4: Range of each numerical feature (min and max values)
numerical_columns = df.select_dtypes(include=[np.number]).columns

# Calculate range (max - min)
feature_ranges = pd.DataFrame({
    'Min': df[numerical_columns].min(),
    'Max': df[numerical_columns].max(),
    'Range': df[numerical_columns].max() - df[numerical_columns].min()
})

print("Range of Numerical Features:")
print(feature_ranges)

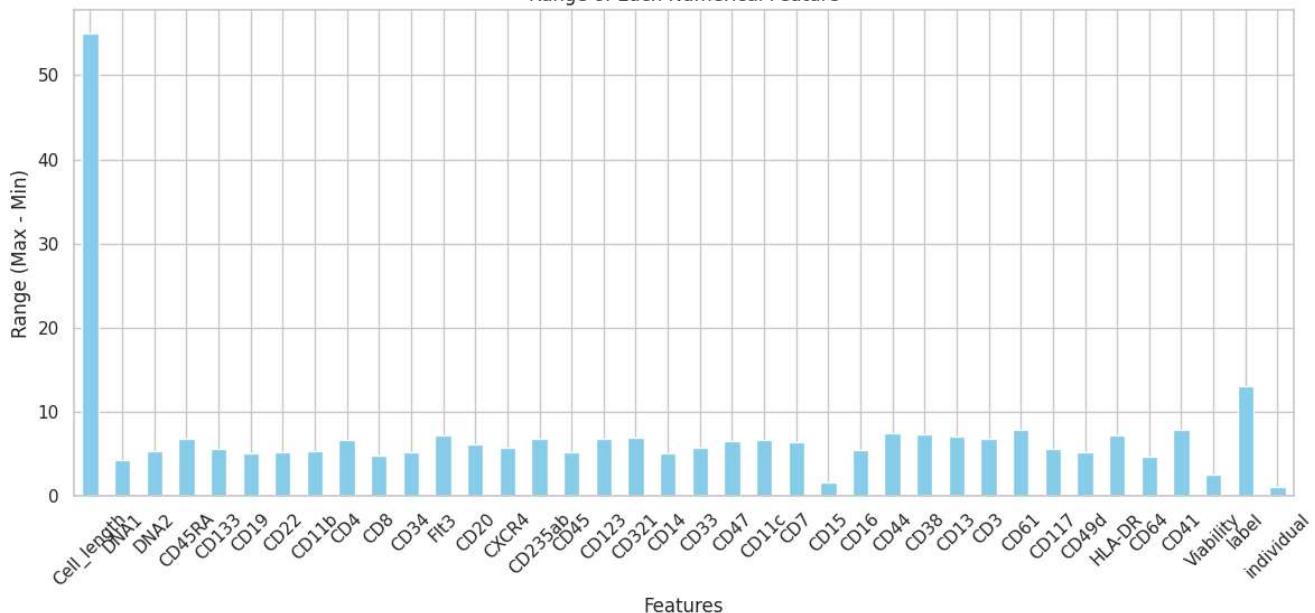
# Optionally plot the range as a bar chart for better visualization
feature_ranges['Range'].plot(kind='bar', figsize=(12, 6), color='skyblue')
```

```
plt.title('Range of Each Numerical Feature')
plt.ylabel('Range (Max - Min)')
plt.xlabel('Features')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Range of Numerical Features:

	Min	Max	Range
Cell_length	10.000000	65.000000	55.000000
DNA1	2.786488	7.001489	4.215001
DNA2	2.236450	7.472308	5.235858
CD45RA	-0.057305	6.691197	6.748502
CD133	-0.058081	5.527494	5.585575
CD19	-0.058089	4.990085	5.048174
CD22	-0.057342	5.160477	5.217819
CD11b	-0.058236	5.260789	5.319025
CD4	-0.057751	6.581762	6.639513
CD8	-0.058003	4.693694	4.751697
CD34	-0.058008	5.147996	5.206004
Flt3	-0.057884	7.117323	7.175207
CD20	-0.058132	6.051411	6.109543
CXCR4	-0.057042	5.696674	5.753717
CD235ab	-0.057612	6.646699	6.704311
CD45	2.040243	7.238076	5.197833
CD123	-0.058003	6.640626	6.698630
CD321	-0.053552	6.867388	6.920940
CD14	-0.057954	5.006121	5.064075
CD33	-0.058079	5.612469	5.670548
CD47	-0.055887	6.402488	6.457575
CD11c	-0.058053	6.520939	6.578992
CD7	-0.058162	6.319219	6.377381
CD15	-0.058077	1.534151	1.592227
CD16	-0.057780	5.338305	5.396085
CD44	0.026061	7.404564	7.378503
CD38	-0.057194	7.293085	7.350279
CD13	-0.057728	6.981187	7.038915
CD3	-0.058241	6.748362	6.806603
CD61	-0.057642	7.748498	7.806139
CD117	-0.057668	5.502125	5.559793
CD49d	-0.058064	5.153438	5.211502
HLA-DR	-0.057974	7.052507	7.110481
CD64	-0.058199	4.517843	4.576042
CD41	-0.058244	7.718288	7.776532
Viability	-0.057979	2.433031	2.491010
label	1.000000	14.000000	13.000000
individual	1.000000	2.000000	1.000000

Range of Each Numerical Feature



```
# Cell 7: Correlation matrix for numerical columns
print("\nCorrelation Matrix:")
correlation_matrix = df.corr()
print(correlation_matrix)
```

```
# Plot the correlation matrix as a heatmap
```

```
plt.figure(figsize=(24, 22))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
#plt.title('Correlation Heatmap')
#plt.show()
sns.pairplot(df[['Time', 'Cell_length', 'DNA1','DNA2','CD45RA', ]])#CD133 CD19 CD22 CD11b CD4 CD8 CD34 Flt3 CD20 CXCR4 CD235a#
# CD7 CD15 CD16 CD44 CD38 CD13', 'CD3', 'CD61', 'CD117', 'CD49d', 'HLA-DR', 'CD64', 'Ct
```



Correlation Matrix:

	Cell_length	DNA1	DNA2	CD45RA	CD133	CD19	\
Cell_length	1.000000	0.416423	0.431905	0.111393	-0.059958	-0.005861	
DNA1	0.416423	1.000000	0.984877	0.100069	-0.285511	-0.025971	
DNA2	0.431905	0.984877	1.000000	0.104679	-0.277263	-0.026996	
CD45RA	0.111393	0.100069	0.104679	1.000000	0.092387	0.361185	
CD133	-0.059958	-0.285511	-0.277263	0.092387	1.000000	0.094540	
CD19	-0.005861	-0.025971	-0.026996	0.361185	0.094540	1.000000	
CD22	-0.012262	-0.063193	-0.063374	0.359608	0.112904	0.754304	
CD11b	0.008631	-0.000526	0.004155	-0.119707	0.016588	-0.172359	
CD4	0.033517	-0.005337	0.000563	-0.108922	-0.013967	-0.240937	
CD8	0.022215	0.002270	0.006531	0.079826	0.089389	-0.193806	
CD34	-0.014858	-0.050997	-0.051362	-0.059609	0.134702	0.012203	
Flt3	-0.005296	-0.177849	-0.169849	0.085093	0.141308	0.043233	
CD20	-0.063839	-0.167952	-0.168882	0.302138	0.141844	0.677729	
CXCR4	0.125692	0.213119	0.213942	0.100903	-0.021030	0.136801	
CD235ab	0.055211	-0.116682	-0.106010	0.072602	0.098651	-0.067810	
CD45	0.115135	0.013762	0.027237	0.235714	0.076465	-0.027327	
CD123	0.012582	0.018822	0.018855	-0.066718	0.003006	-0.094447	
CD321	0.087060	0.152760	0.158168	-0.113485	0.043251	-0.225180	
CD14	0.174691	0.003549	0.007072	-0.069475	0.001983	-0.068132	
CD33	0.216010	0.257904	0.257590	-0.167203	-0.081139	-0.145424	
CD47	0.234082	0.230418	0.235722	0.034710	-0.032528	0.045457	
CD11c	0.044275	-0.051675	-0.045633	-0.141138	0.018919	-0.182480	
CD7	0.044125	-0.085472	-0.080578	0.098950	0.035054	-0.280029	
CD15	-0.000245	-0.371413	-0.367335	-0.031962	0.149928	0.010467	
CD16	0.036142	-0.094131	-0.092328	0.066533	0.042190	-0.054485	
CD44	0.169501	0.027131	0.044926	-0.018587	0.079590	-0.120735	
CD38	0.128767	0.065785	0.065349	-0.058246	-0.032274	0.179445	
CD13	0.239758	0.039993	0.043661	-0.134795	0.009715	-0.101015	
CD3	0.046857	-0.053835	-0.045767	-0.077166	0.007174	-0.382992	
CD61	0.073699	-0.217570	-0.211408	0.036778	0.201075	-0.027492	
CD117	0.115849	-0.029422	-0.028541	-0.097947	0.063662	-0.159949	
CD49d	0.115718	-0.338205	-0.331882	0.011911	0.141077	0.174649	
HLA-DR	0.067276	0.072721	0.073840	0.121525	0.030683	0.510022	
CD64	0.017584	-0.113142	-0.108537	-0.167486	0.005290	-0.183043	
CD41	0.088435	-0.171759	-0.166552	0.062116	0.193653	0.040460	
Viability	-0.037960	-0.501241	-0.496669	-0.024698	0.185545	0.008273	
label	0.048118	0.028648	0.030758	0.006543	-0.015698	0.403745	
individual	0.462250	0.970535	0.963337	0.099729	-0.299903	-0.027574	

	CD22	CD11b	CD4	CD8	...	CD3	CD61	\
Cell_length	-0.012262	0.008631	0.033517	0.022215	...	0.046857	0.073699	
DNA1	-0.063193	-0.000526	-0.005337	0.002270	...	-0.053835	-0.217570	
DNA2	-0.063374	0.004155	0.000563	0.006531	...	-0.045767	-0.211408	
CD45RA	0.359608	-0.119707	-0.108922	0.079826	...	-0.077166	0.036778	
CD133	0.112904	0.016588	-0.013967	0.089389	...	0.007174	0.201075	
CD19	0.754304	-0.172359	-0.240937	-0.193806	...	-0.382992	-0.027492	
CD22	1.000000	-0.105285	-0.191386	-0.151013	...	-0.313831	-0.002195	
CD11b	-0.105285	1.000000	-0.095881	-0.191399	...	-0.397684	0.191408	
CD4	-0.191386	-0.095881	1.000000	-0.162295	...	0.484001	0.031420	
CD8	-0.151013	-0.191399	-0.162295	1.000000	...	0.471683	-0.020587	
CD34	0.012055	-0.108853	-0.127823	-0.088330	...	-0.202279	0.047972	
Flt3	0.051363	0.019224	0.032149	0.052240	...	0.025606	0.132403	
CD20	0.657568	-0.134545	-0.169660	-0.094283	...	-0.257291	0.003904	
CXCR4	0.069354	-0.110342	0.001769	-0.032875	...	-0.043566	0.028881	
CD235ab	-0.012597	0.000910	0.166719	0.150227	...	0.247799	0.068670	
CD45	0.091451	0.059682	0.292664	0.313013	...	0.479654	0.011057	
CD123	-0.089570	0.104383	-0.108263	-0.165568	...	-0.290665	0.087569	
CD321	-0.244964	0.476634	0.041311	-0.148079	...	-0.215350	0.228276	
CD14	-0.053846	0.388605	-0.026638	-0.100799	...	-0.166823	0.133992	
CD33	-0.125780	0.522986	-0.113557	-0.203918	...	-0.348791	0.061931	
CD47	-0.053796	-0.177834	0.192686	-0.230935	...	0.040753	0.045996	
CD11c	-0.123396	0.752404	-0.103205	-0.231567	...	-0.419199	0.185977	
CD7	-0.230389	-0.209502	0.094930	0.253857	...	0.355685	-0.034703	
CD15	0.022256	0.030678	-0.019730	-0.033721	...	-0.033896	0.155460	
CD16	-0.041339	0.017394	-0.084435	-0.015360	...	-0.148557	0.027295	
CD44	0.019839	0.347100	0.159809	-0.011068	...	0.060048	0.117334	
CD38	-0.054066	0.088286	-0.251735	-0.311973	...	-0.507752	0.035300	
CD13	-0.134976	0.387089	-0.081900	-0.177925	...	-0.266303	0.165532	
CD3	-0.313831	-0.397684	0.484001	0.471683	...	1.000000	-0.037301	
CD61	-0.002195	0.191408	0.031420	-0.020587	...	-0.037301	1.000000	
CD117	-0.131731	-0.173298	0.035361	0.038080	...	0.129453	0.006636	
CD49d	0.129092	0.073541	-0.111404	-0.122583	...	-0.266885	0.113150	
HLA-DR	0.456959	0.151646	-0.325505	-0.374539	...	-0.671096	0.003343	
CD64	-0.131843	0.689275	-0.117753	-0.228732	...	-0.397595	0.192315	
CD41	0.060769	0.206182	-0.017621	-0.070514	...	-0.146346	0.844115	
Viability	0.046020	0.152711	-0.015876	-0.033143	...	-0.063177	0.203623	
label	0.200137	0.254222	-0.172029	-0.074506	...	-0.249788	0.089280	
individual	-0.065132	-0.040419	0.001191	0.014804	...	-0.016063	-0.223054	

	CD117	CD49d	HLA-DR	CD64	CD41	Viability	\
Cell_length	0.115849	0.115718	0.067276	0.017584	0.088435	-0.037960	
DNA1	-0.029422	-0.338205	0.072721	-0.113142	-0.171759	-0.501241	
DNA2	-0.028541	-0.331882	0.073840	-0.108537	-0.166552	-0.496669	
CD45RA	-0.097947	0.011911	0.121525	-0.167486	0.062116	-0.024698	
CD133	0.063662	0.141077	0.030683	0.005290	0.193653	0.185545	
CD19	-0.159949	0.174649	0.510022	-0.183043	0.040460	0.008273	
CD22	-0.131731	0.129092	0.456959	-0.131843	0.060769	0.046020	

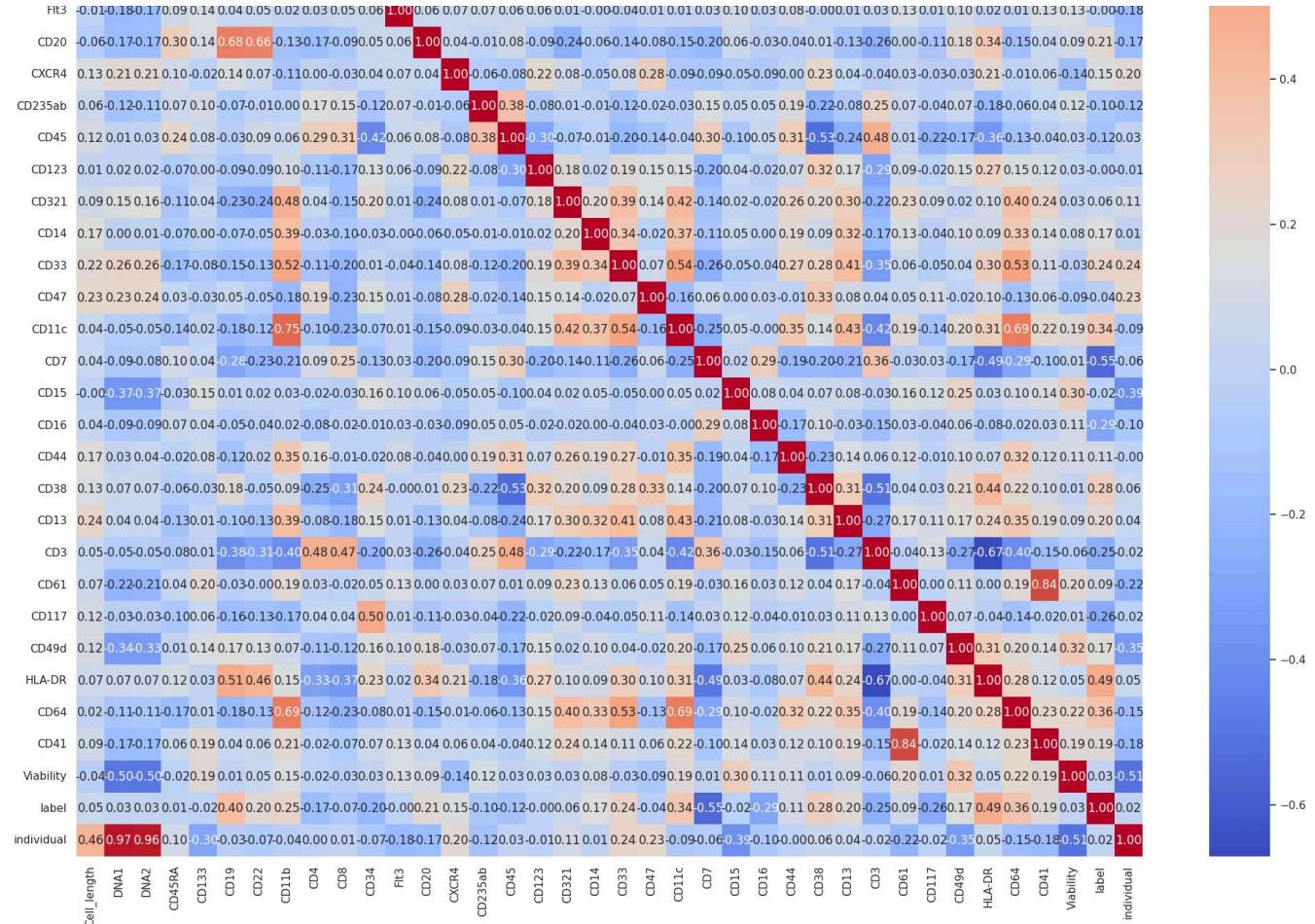
CD11b	-0.173298	0.073541	0.151646	0.689275	0.206182	0.152711
CD4	0.035361	-0.111404	-0.325505	-0.117753	-0.017621	-0.015076
CD8	0.038080	-0.122583	-0.374539	-0.228732	-0.070514	-0.033143
CD34	0.495477	0.156996	0.232057	-0.080071	0.070400	0.032199
Flt3	0.011228	0.103909	0.022111	0.013337	0.127140	0.125719
CD20	-0.109846	0.184038	0.339158	-0.146238	0.043444	0.093348
CXCR4	-0.032972	-0.029809	0.213350	-0.011757	0.061844	-0.144180
CD235ab	-0.042192	0.065299	-0.184196	-0.057670	0.036325	0.123578
CD45	-0.222683	-0.171252	-0.357085	-0.127649	-0.039397	0.028320
CD123	-0.021723	0.151016	0.272931	0.153761	0.123819	0.034611
CD321	0.085988	0.023122	0.104760	0.403274	0.243086	0.030446
CD14	-0.041438	0.098319	0.091239	0.328821	0.143476	0.081498
CD33	-0.045993	0.038415	0.297926	0.532189	0.107699	-0.029332
CD47	0.110119	-0.016422	0.101504	-0.128859	0.057525	-0.091431
CD11c	-0.139600	0.196323	0.305424	0.686101	0.216689	0.190505
CD7	0.029385	-0.167813	-0.489190	-0.289364	-0.095291	0.006370
CD15	0.120870	0.250531	0.032571	0.100006	0.138293	0.300861
CD16	-0.036560	0.059456	-0.082477	-0.021666	0.027416	0.112051
CD44	-0.011624	0.100420	0.070566	0.316340	0.115712	0.105561
CD38	0.026857	0.206279	0.436875	0.217328	0.100041	0.014284
CD13	0.105276	0.171125	0.241906	0.345855	0.194092	0.086075
CD3	0.129453	-0.266885	-0.671096	-0.397595	-0.146346	-0.063177
CD61	0.000636	0.113150	0.003343	0.192315	0.844115	0.203623
CD117	1.000000	0.071585	-0.044360	-0.141553	-0.024822	0.008143
CD49d	0.071585	1.000000	0.308961	0.195341	0.137746	0.324575
HLA-DR	-0.044360	0.308961	1.000000	0.278228	0.122535	0.045084
CD64	-0.141553	0.195341	0.278228	1.000000	0.228247	0.224292
CD41	-0.024822	0.137746	0.122535	0.228247	1.000000	0.188394
Viability	0.008143	0.324575	0.045084	0.224292	0.188394	1.000000
label	-0.260230	0.172392	0.492008	0.362429	0.187400	0.031471
individual	-0.021610	-0.350099	0.049476	-0.149619	-0.177829	-0.511000

	label	individual
Cell_length	0.048118	0.462250
DNA1	0.028648	0.970535
DNA2	0.030758	0.963337
CD45RA	0.006543	0.099729
CD133	-0.015698	-0.299903
CD19	0.403745	-0.027574
CD22	0.200137	-0.065132
CD11b	0.254222	-0.040419
CD4	-0.172029	0.001191
CD8	-0.074506	0.014804
CD34	-0.199478	-0.073039
Flt3	-0.004128	-0.184494
CD20	0.213955	-0.167543
CXCR4	0.148705	0.204654
CD235ab	-0.103473	-0.124068
CD45	-0.119191	0.026696
CD123	-0.004536	-0.006421
CD321	0.058874	0.106941
CD14	0.170154	0.005152
CD33	0.239835	0.243979
CD47	-0.040625	0.231153
CD11c	0.337001	-0.085888
CD7	-0.550401	-0.063152
CD15	-0.019566	-0.385950
CD16	-0.285620	-0.095535
CD44	0.105486	-0.000064
CD38	0.282430	0.056586
CD13	0.195687	0.041072
CD3	-0.249788	-0.016063
CD61	0.089280	-0.223054
CD117	-0.260230	-0.021610
CD49d	0.172392	-0.350099
HLA-DR	0.492008	0.049476
CD64	0.362429	-0.149619
CD41	0.187400	-0.177829
Viability	0.031471	-0.511000
label	1.000000	0.019896
individual	0.019896	1.000000

[38 rows x 38 columns]

<Axes: >





```
from scipy.stats import skew

skewness = df.apply(skew)

# Function to categorize skewness
def categorize_skewness(value):
    if value > 0.5:
        return 'Right-skewed'
    elif value < -0.5:
        return 'Left-skewed'
    else:
        return 'Approximately symmetrical'

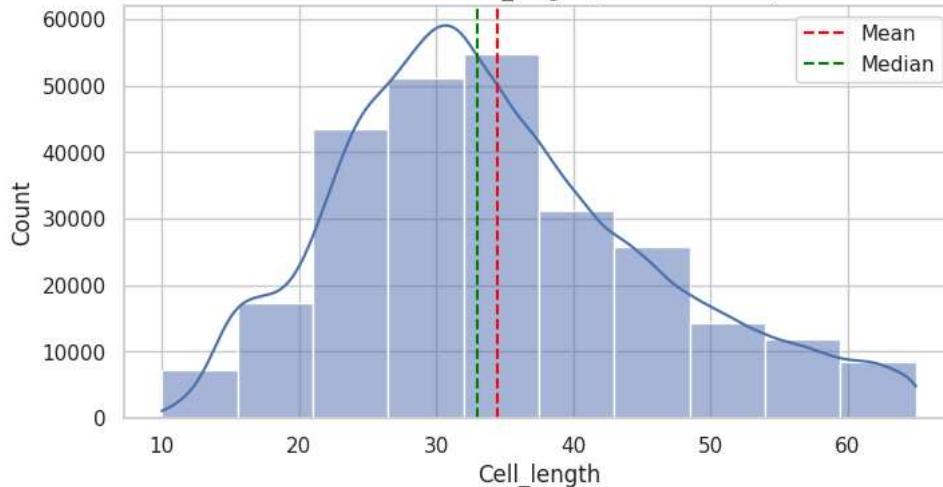
# Apply the categorization
skewness_category = skewness.apply(categorize_skewness)

# Display skewness and its categorization
skewness_df = pd.DataFrame({'Skewness': skewness, 'Category': skewness_category})
print(skewness_df)

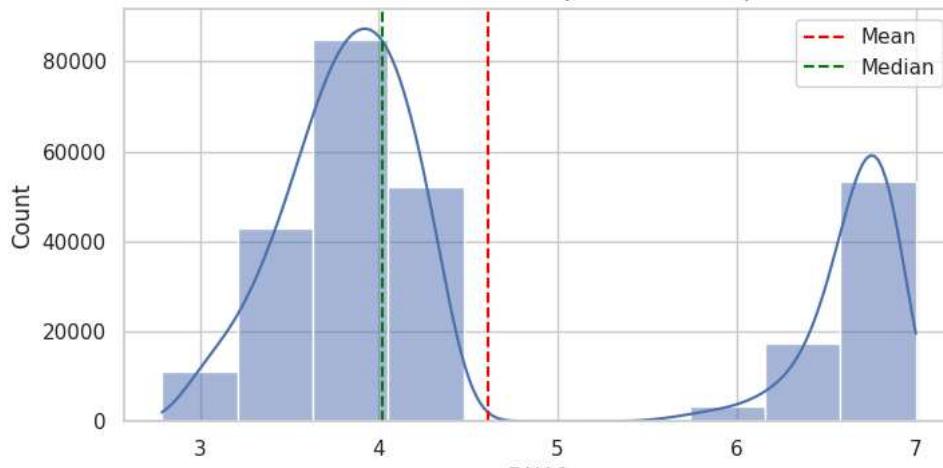
# Plot histograms for each numerical column
for col in df:
    plt.figure(figsize=(8, 4))
    sns.histplot(df[col], bins=10, kde=True)
    plt.title(f'Distribution of {col} (Skewness: {skewness[col]:.2f})')
    plt.axvline(df[col].mean(), color='red', linestyle='--', label='Mean')
    plt.axvline(df[col].median(), color='green', linestyle='--', label='Median')
    plt.legend()
    plt.show()
```

	Skewness	Category
Cell_length	0.527832	Right-skewed
DNA1	0.845010	Right-skewed
DNA2	0.779167	Right-skewed
CD45RA	1.191595	Right-skewed
CD133	2.141953	Right-skewed
CD19	1.682609	Right-skewed
CD22	2.283181	Right-skewed
CD11b	1.679089	Right-skewed
CD4	1.622044	Right-skewed
CD8	1.775713	Right-skewed
CD34	3.492437	Right-skewed
Flt3	7.098151	Right-skewed
CD20	2.754699	Right-skewed
CXCR4	0.955342	Right-skewed
CD235ab	2.001479	Right-skewed
CD45	-1.484824	Left-skewed
CD123	3.648890	Right-skewed
CD321	0.247097	Approximately symmetrical
CD14	3.609006	Right-skewed
CD33	2.724977	Right-skewed
CD47	-0.250323	Approximately symmetrical
CD11c	1.733888	Right-skewed
CD7	1.606528	Right-skewed
CD15	1.445147	Right-skewed
CD16	5.733203	Right-skewed
CD44	-0.431589	Approximately symmetrical
CD38	1.141482	Right-skewed
CD13	2.234311	Right-skewed
CD3	0.342239	Approximately symmetrical
CD61	4.894707	Right-skewed
CD117	4.097508	Right-skewed
CD49d	0.856805	Right-skewed
HLA-DR	0.795359	Right-skewed
CD64	1.743733	Right-skewed
CD41	5.366314	Right-skewed
Viability	0.985417	Right-skewed
label	Nan	Approximately symmetrical
individual	0.982030	Right-skewed

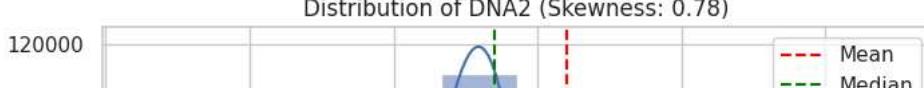
Distribution of Cell_length (Skewness: 0.53)

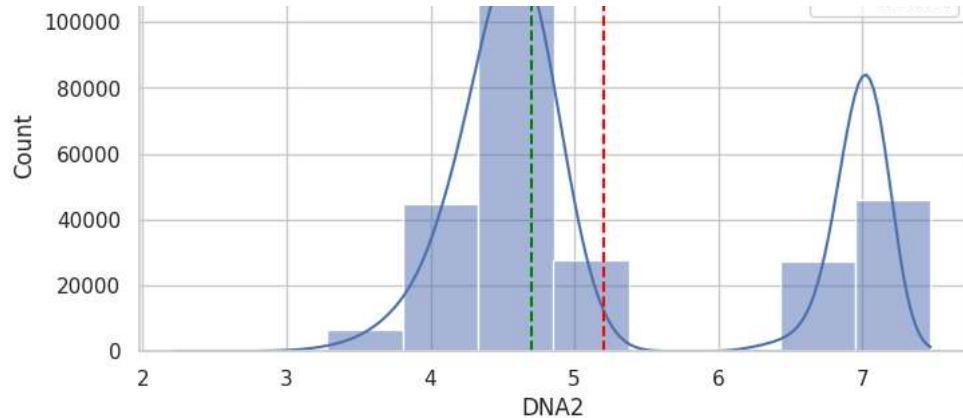


Distribution of DNA1 (Skewness: 0.85)

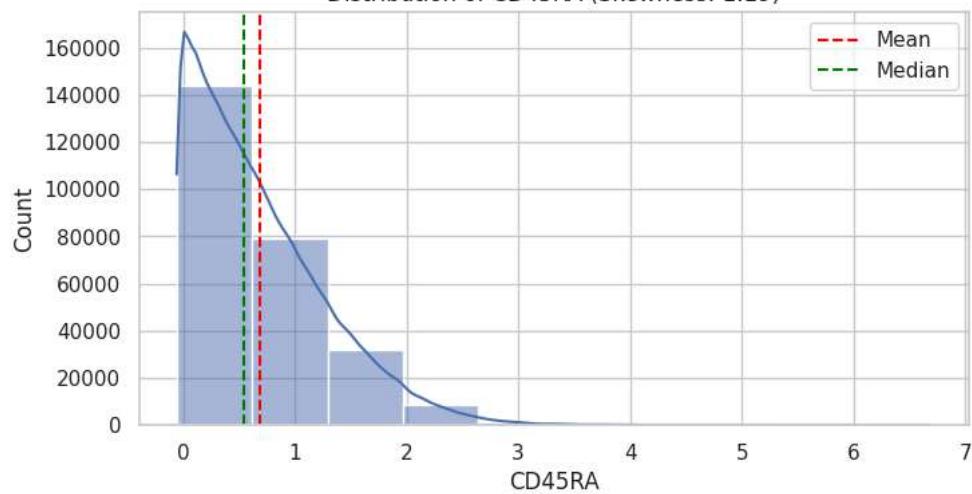


Distribution of DNA2 (Skewness: 0.78)

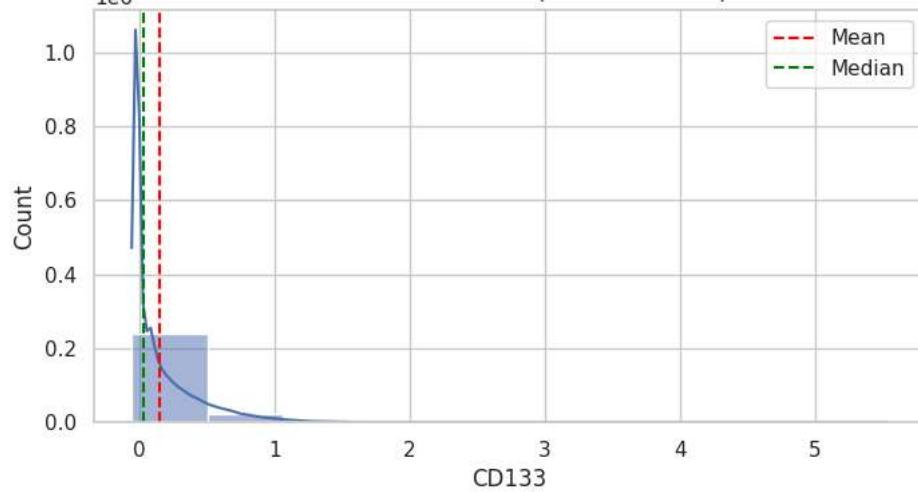




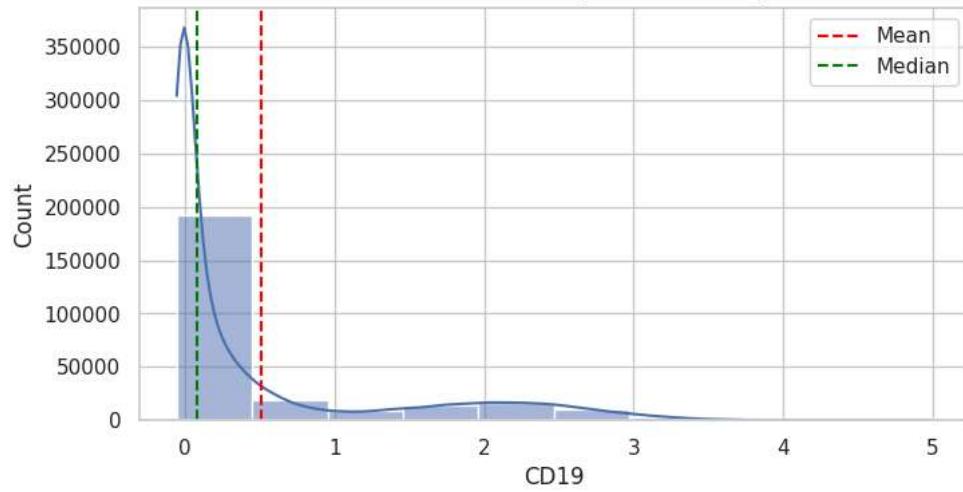
Distribution of CD45RA (Skewness: 1.19)



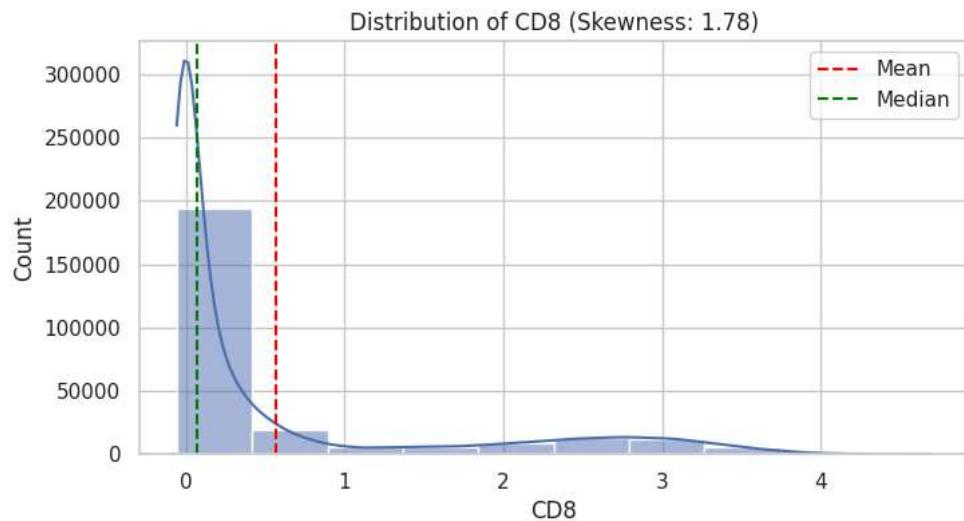
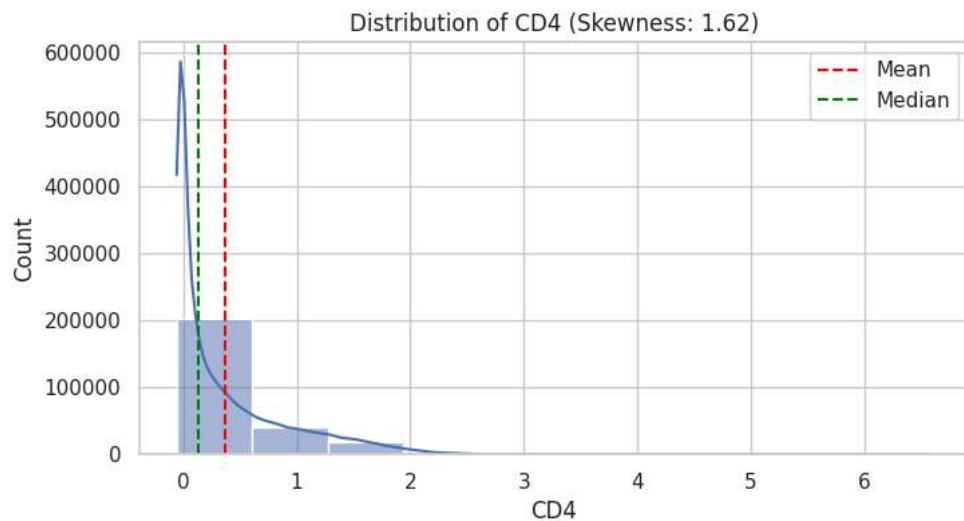
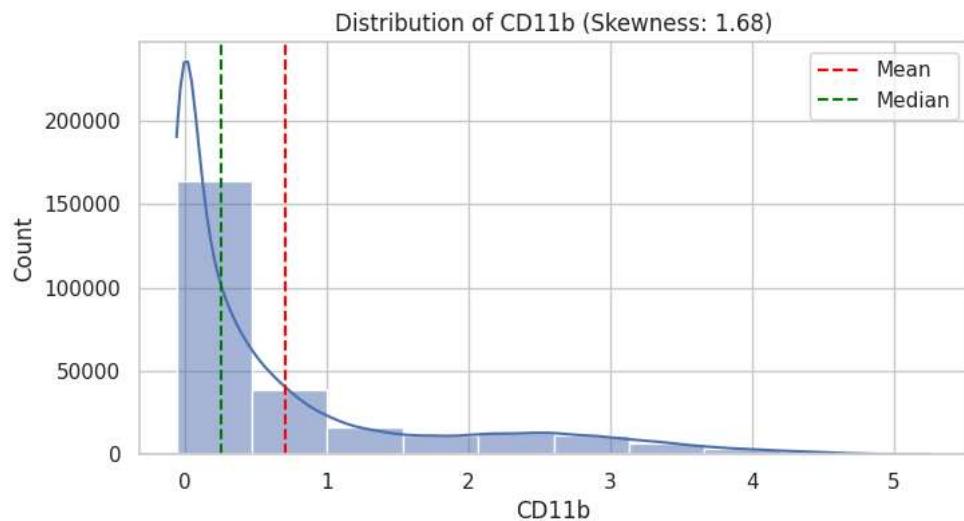
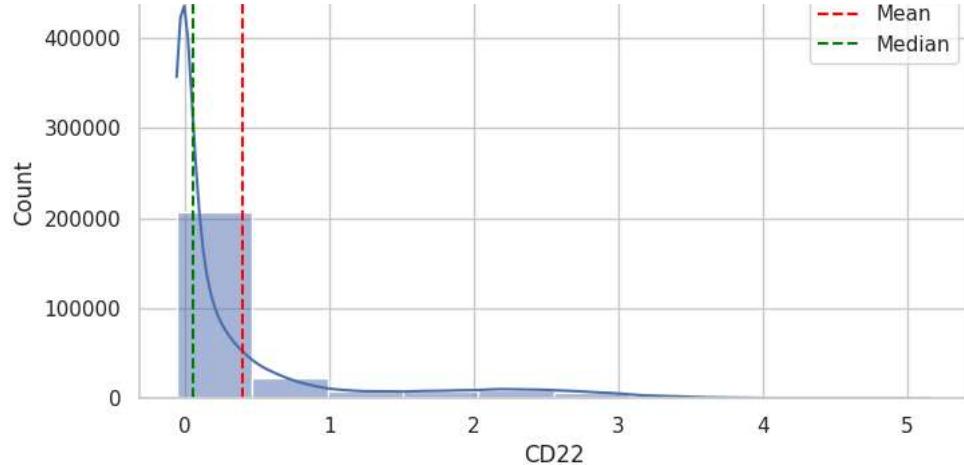
Distribution of CD133 (Skewness: 2.14)



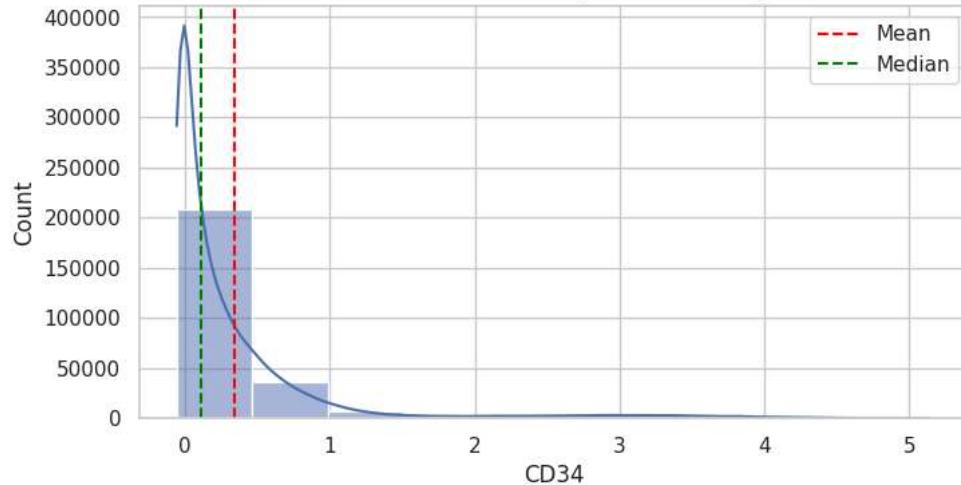
Distribution of CD19 (Skewness: 1.68)



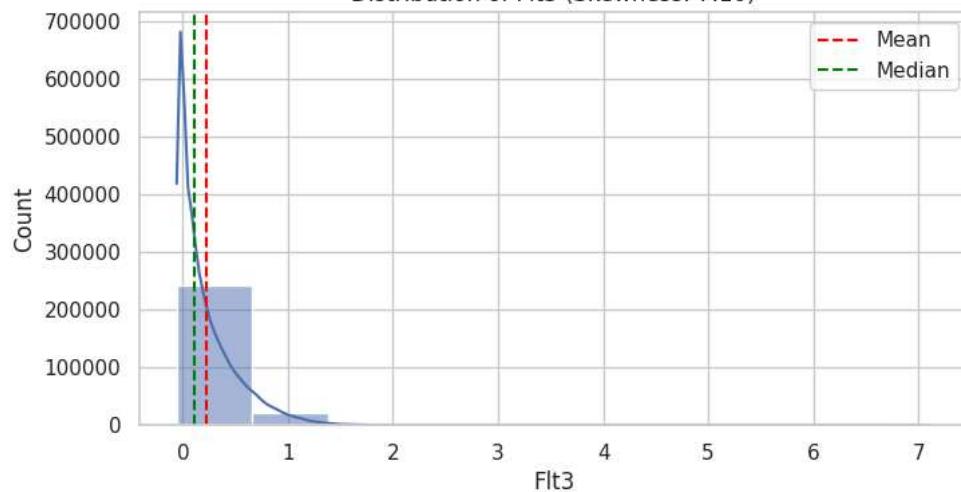
Distribution of CD22 (Skewness: 2.28)



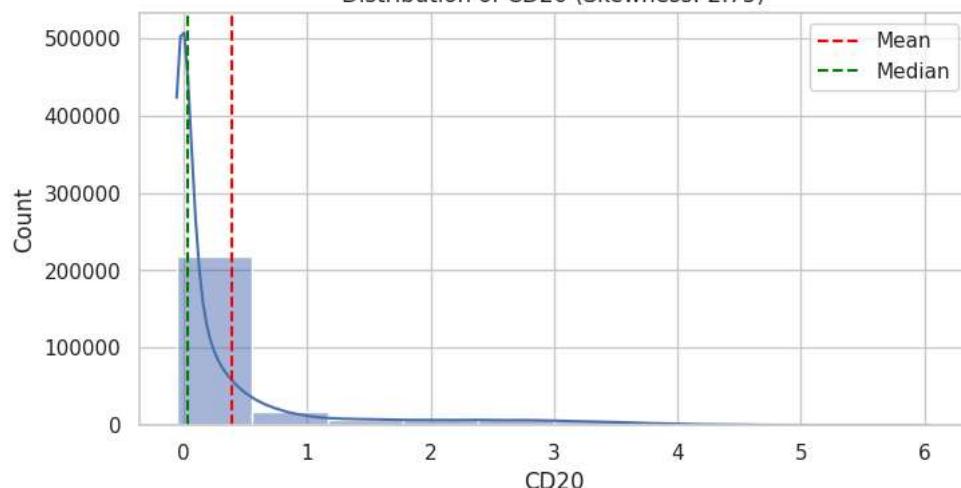
Distribution of CD34 (Skewness: 3.49)



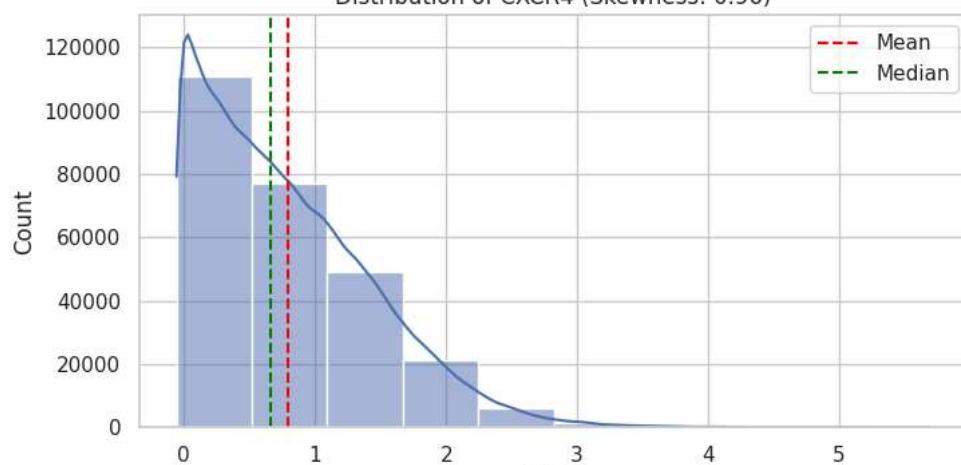
Distribution of Flt3 (Skewness: 7.10)



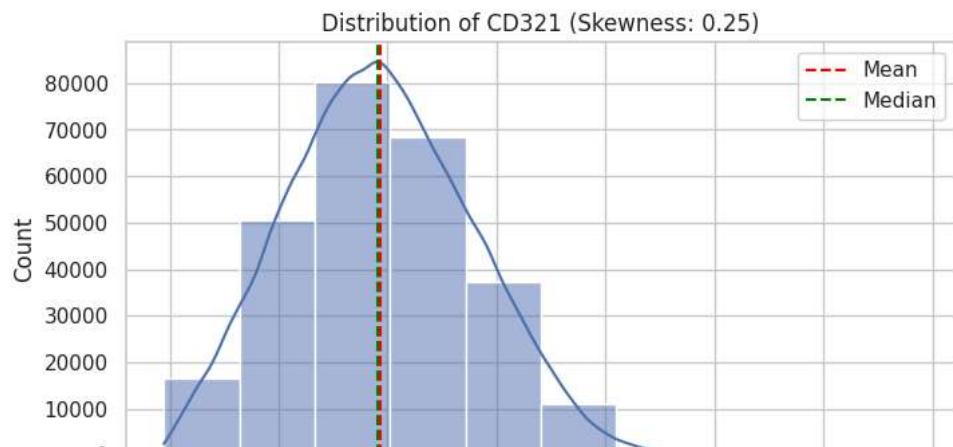
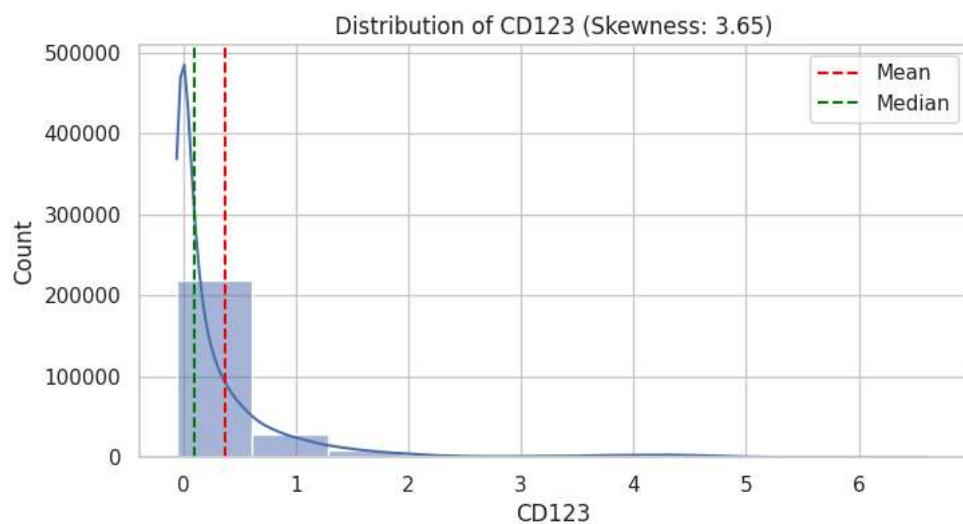
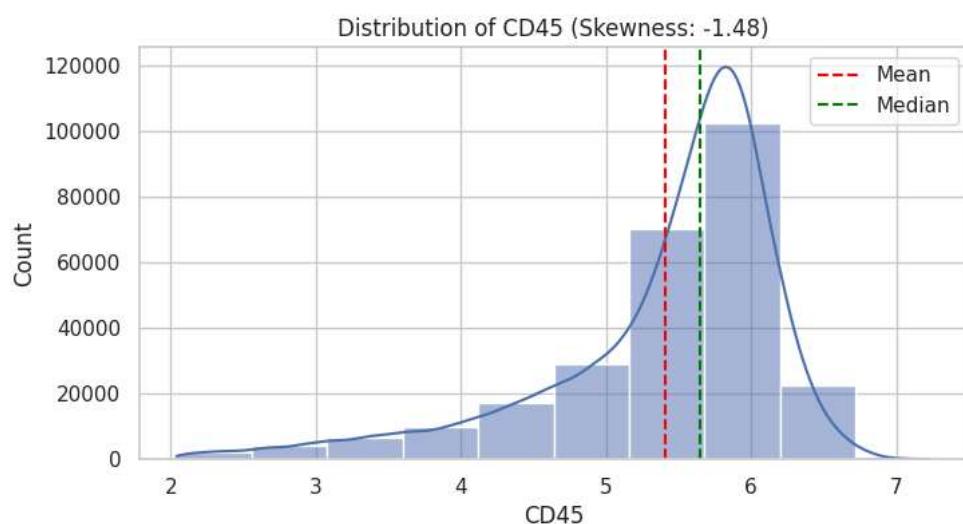
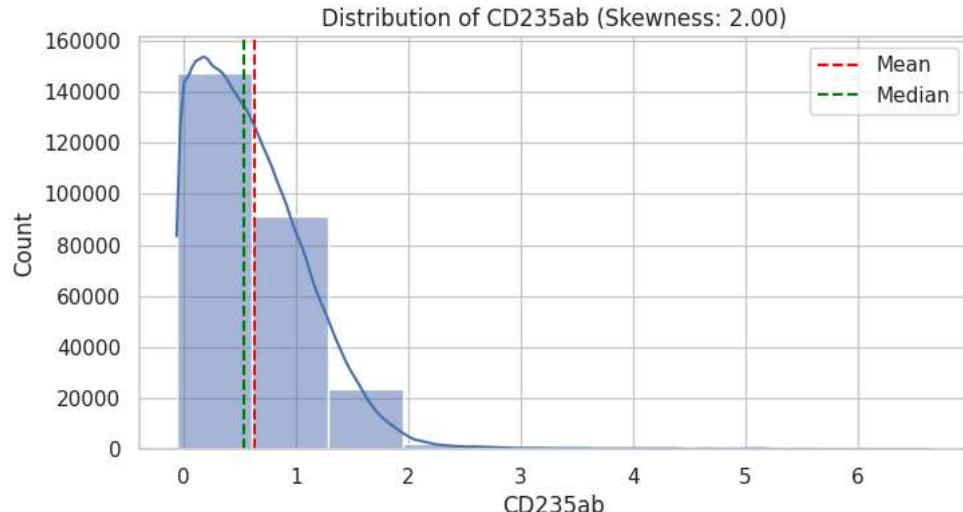
Distribution of CD20 (Skewness: 2.75)

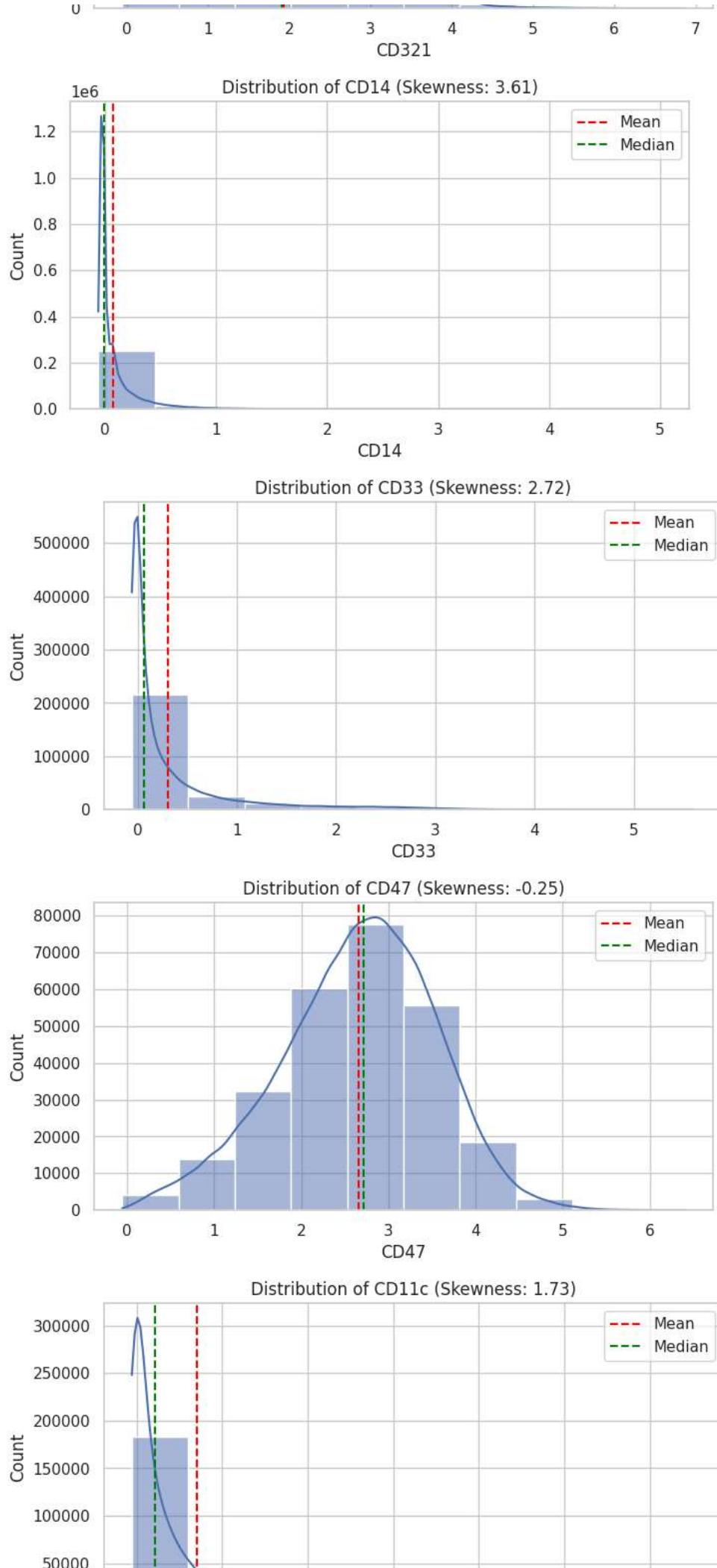


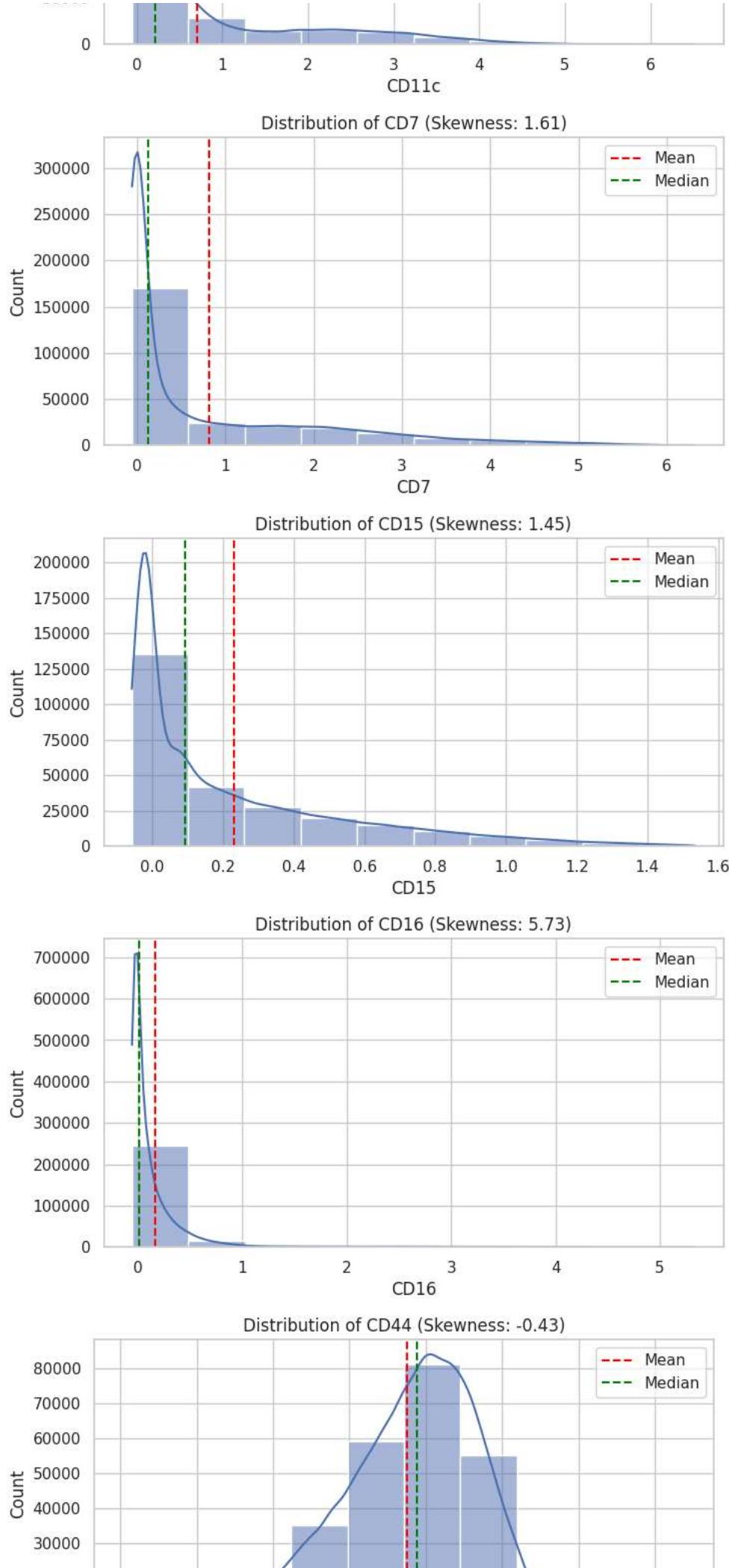
Distribution of CXCR4 (Skewness: 0.96)

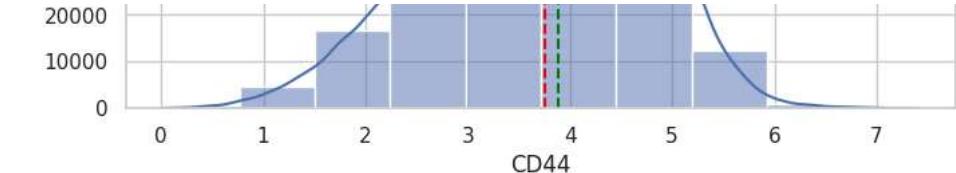


CXCR4

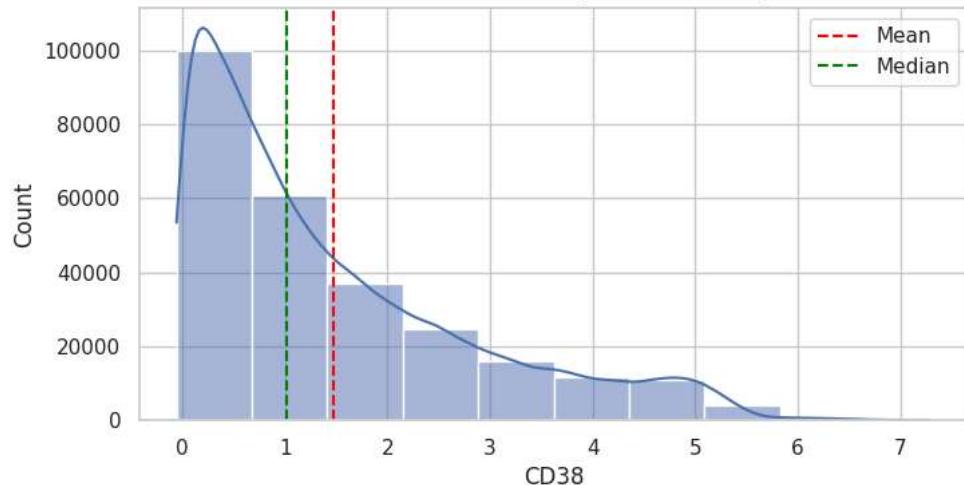




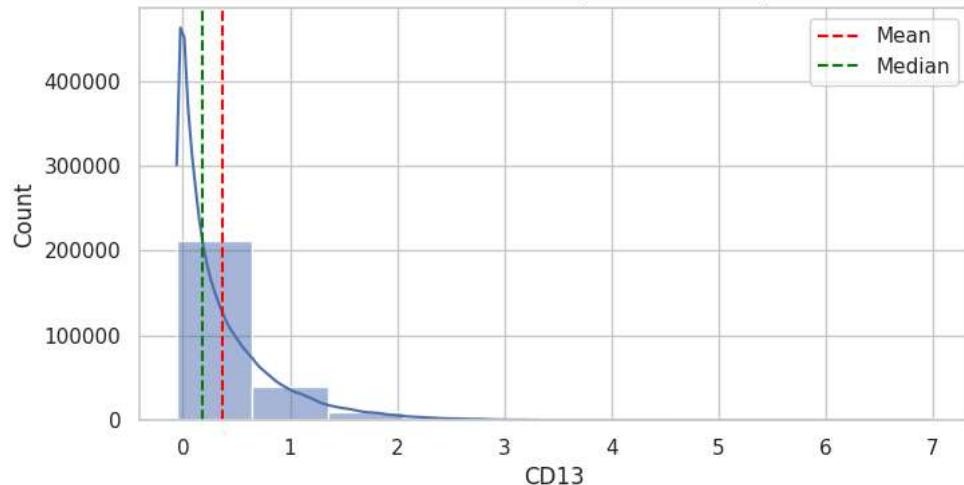




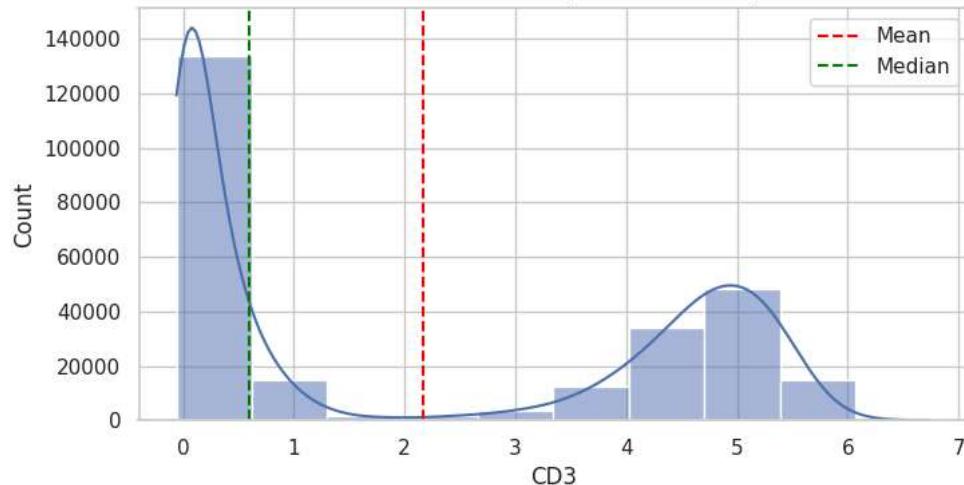
Distribution of CD38 (Skewness: 1.14)



Distribution of CD13 (Skewness: 2.23)

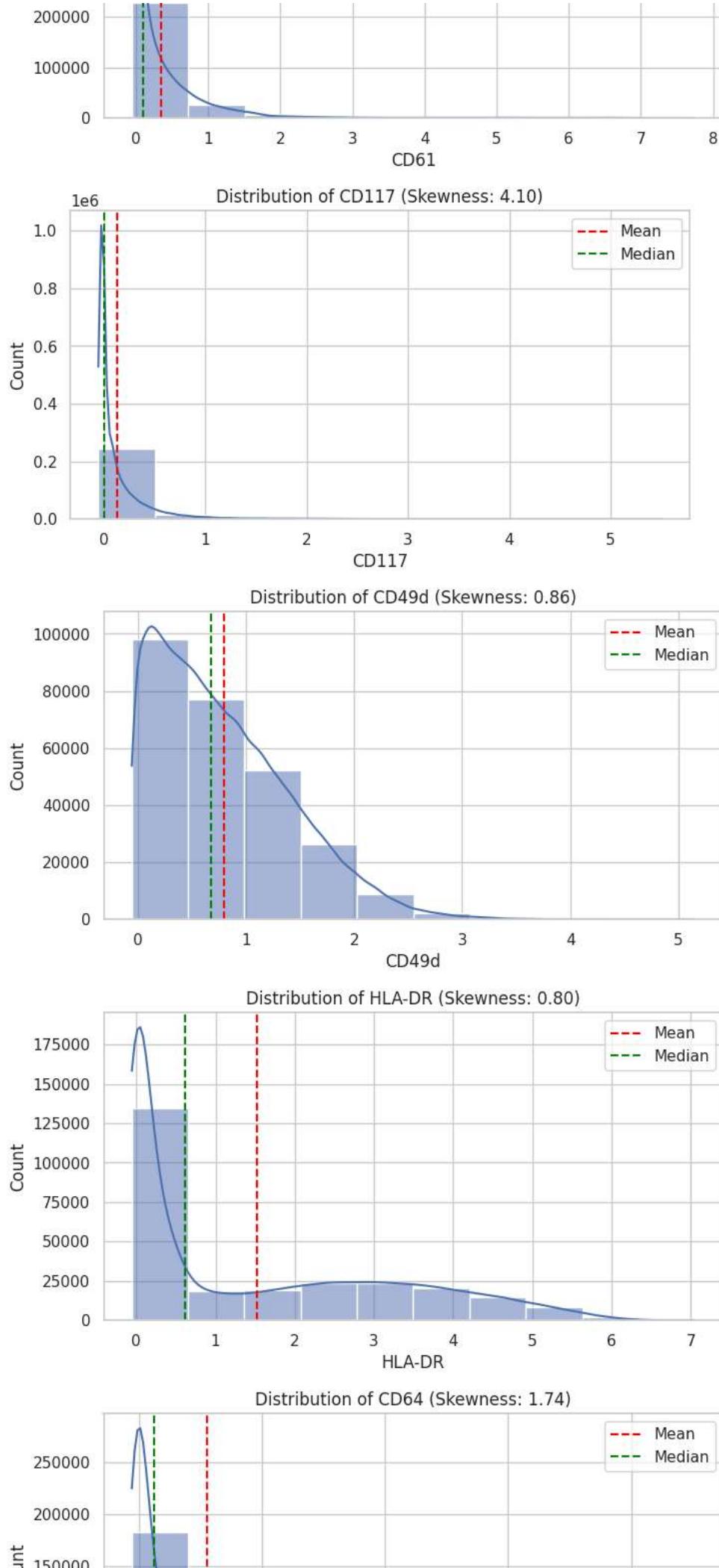


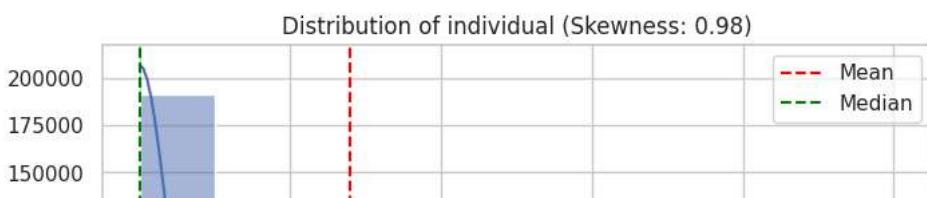
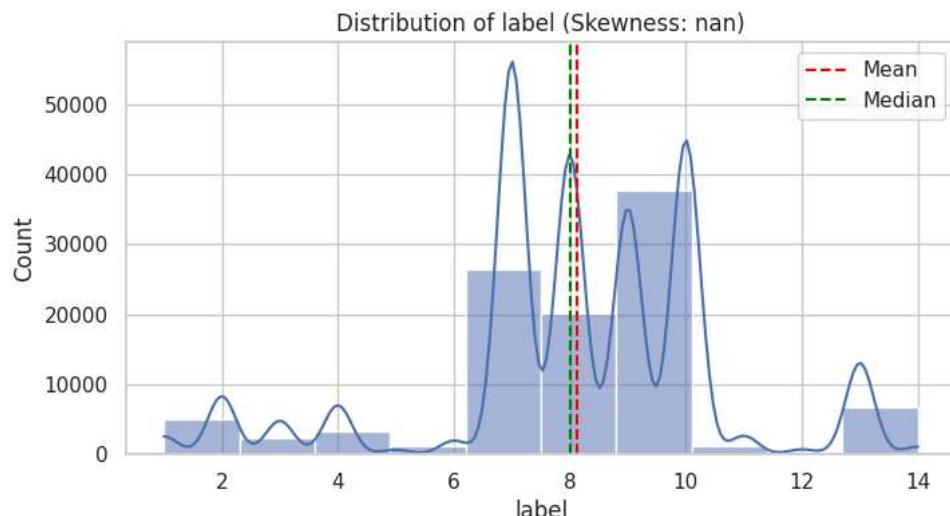
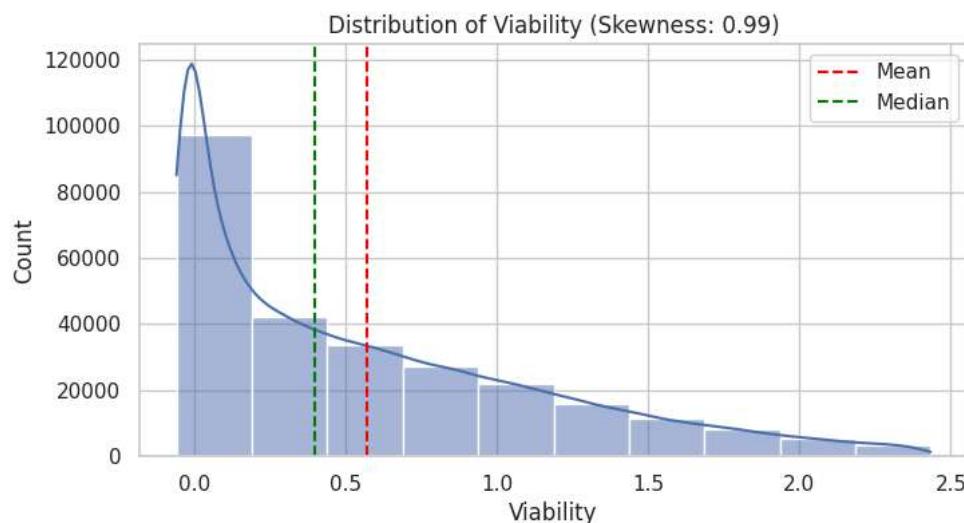
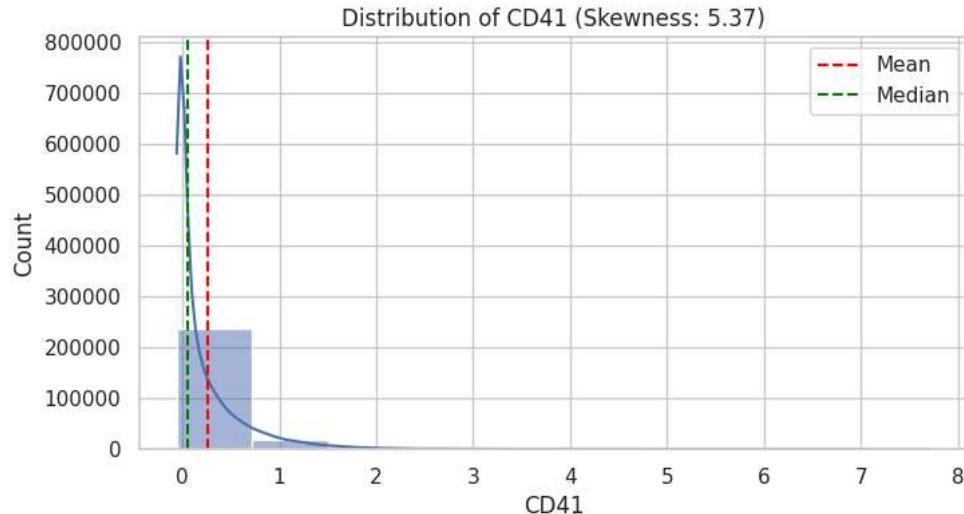
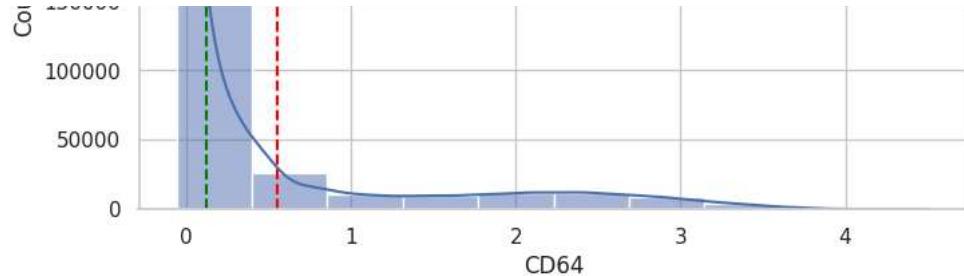
Distribution of CD3 (Skewness: 0.34)

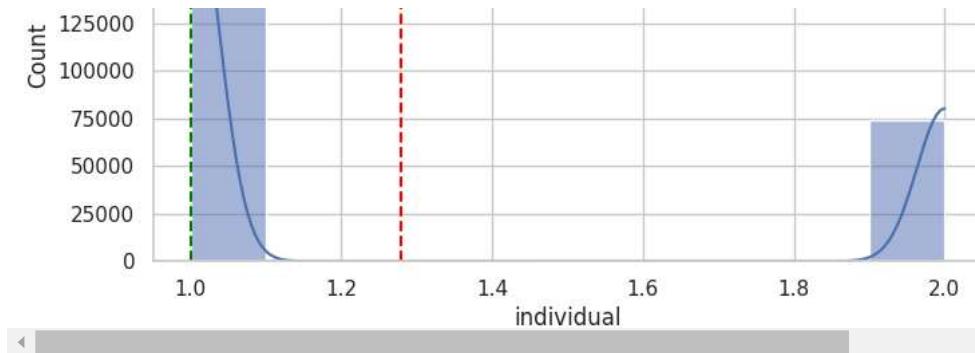


Distribution of CD61 (Skewness: 4.89)









```
from scipy.stats import kurtosis
# Calculate kurtosis for each column
kurtosis_values = df.apply(kurtosis, fisher=False) # Fisher=False gives Pearson kurtosis (normal kurtosis = 3)

# Create a DataFrame with kurtosis values
kurtosis_df = pd.DataFrame({'Column': df.columns, 'Kurtosis': kurtosis_values})

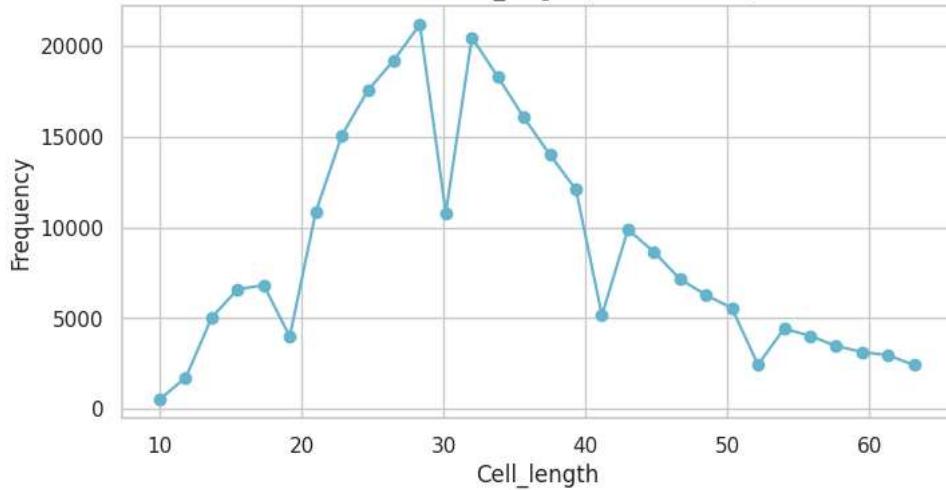
# Categorize the kurtosis values (Leptokurtic, Mesokurtic, Platykurtic)
def categorize_kurtosis(value):
    if value > 3:
        return 'Leptokurtic (heavy tails)'
    elif value < 3:
        return 'Platykurtic (light tails)'
    else:
        return 'Mesokurtic (normal tails)'

kurtosis_df['Category'] = kurtosis_df['Kurtosis'].apply(categorize_kurtosis)

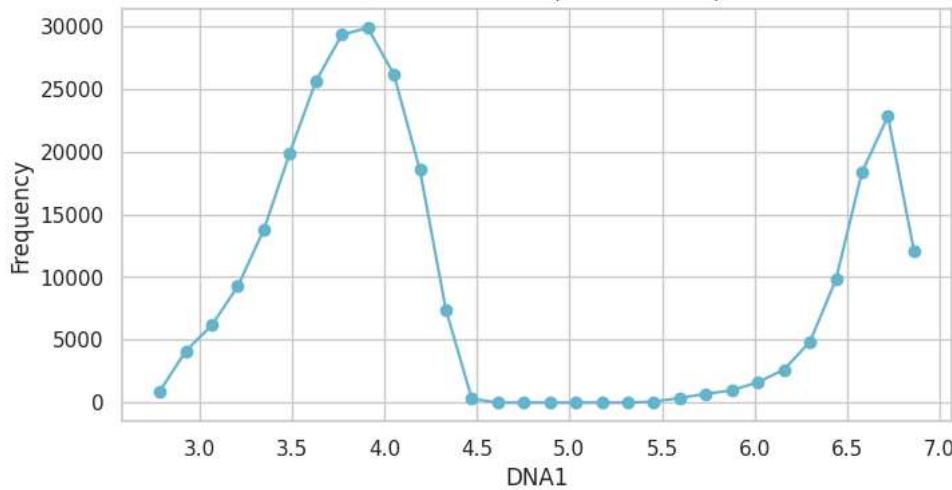
# Print the kurtosis values and their categories
print(kurtosis_df)  
# 4.1 Histogram for each column  
'''  
  
for column in df.columns:  
    plt.figure(figsize=(8, 4))  
    plt.hist(df[column].dropna(), bins=30, color='c', edgecolor='black', alpha=0.7)  
    plt.title(f'Histogram of {column} (Kurtosis: {kurtosis_df.loc[kurtosis_df["Column"] == column, "Kurtosis"].values[0]:.2f})')  
    plt.xlabel(column)  
    plt.ylabel('Frequency')  
    plt.grid(True)  
    plt.show()'''  
  
# 4.1 Line graph for each column (without KDE)  
for column in df.columns:  
    plt.figure(figsize=(8, 4))  
  
    # Calculate histogram values (frequency and bin edges)  
    counts, bin_edges = np.histogram(df[column].dropna(), bins=30)  
  
    # Plot as a line graph  
    plt.plot(bin_edges[:-1], counts, color='c', marker='o', linestyle='-', label=f'{column} Distribution')  
  
    plt.title(f'Line Plot of {column} (Kurtosis: {kurtosis_df.loc[kurtosis_df["Column"] == column, "Kurtosis"].values[0]:.2f})')  
    plt.xlabel(column)  
    plt.ylabel('Frequency')  
    plt.grid(True)  
    plt.show()
```

	Column	Kurtosis	Category
Cell_length	Cell_length	2.834033	Platykurtic (light tails)
DNA1	DNA1	1.994037	Platykurtic (light tails)
DNA2	DNA2	1.975021	Platykurtic (light tails)
CD45RA	CD45RA	4.964272	Leptokurtic (heavy tails)
CD133	CD133	9.190066	Leptokurtic (heavy tails)
CD19	CD19	4.590887	Leptokurtic (heavy tails)
CD22	CD22	7.500223	Leptokurtic (heavy tails)
CD11b	CD11b	4.964495	Leptokurtic (heavy tails)
CD4	CD4	5.844261	Leptokurtic (heavy tails)
CD8	CD8	4.745776	Leptokurtic (heavy tails)
CD34	CD34	16.596416	Leptokurtic (heavy tails)
Flt3	Flt3	85.583534	Leptokurtic (heavy tails)
CD20	CD20	10.435449	Leptokurtic (heavy tails)
CXCR4	CXCR4	3.936307	Leptokurtic (heavy tails)
CD235ab	CD235ab	13.440586	Leptokurtic (heavy tails)
CD45	CD45	5.246770	Leptokurtic (heavy tails)
CD123	CD123	18.361217	Leptokurtic (heavy tails)
CD321	CD321	2.914593	Platykurtic (light tails)
CD14	CD14	23.062535	Leptokurtic (heavy tails)
CD33	CD33	10.967536	Leptokurtic (heavy tails)
CD47	CD47	2.943834	Platykurtic (light tails)
CD11c	CD11c	5.117156	Leptokurtic (heavy tails)
CD7	CD7	4.885115	Leptokurtic (heavy tails)
CD15	CD15	4.504387	Leptokurtic (heavy tails)
CD16	CD16	42.287749	Leptokurtic (heavy tails)
CD44	CD44	2.918792	Platykurtic (light tails)
CD38	CD38	3.521190	Leptokurtic (heavy tails)
CD13	CD13	10.637564	Leptokurtic (heavy tails)
CD3	CD3	1.264612	Platykurtic (light tails)
CD61	CD61	34.878020	Leptokurtic (heavy tails)
CD117	CD117	26.375108	Leptokurtic (heavy tails)
CD49d	CD49d	3.468119	Leptokurtic (heavy tails)
HLA-DR	HLA-DR	2.309924	Platykurtic (light tails)
CD64	CD64	4.910631	Leptokurtic (heavy tails)
CD41	CD41	41.521113	Leptokurtic (heavy tails)
Viability	Viability	3.156935	Leptokurtic (heavy tails)
label	label	NaN	Mesokurtic (normal tails)
individual	individual	1.964382	Platykurtic (light tails)

Line Plot of Cell_length (Kurtosis: 2.83)

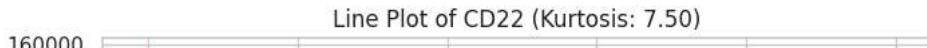
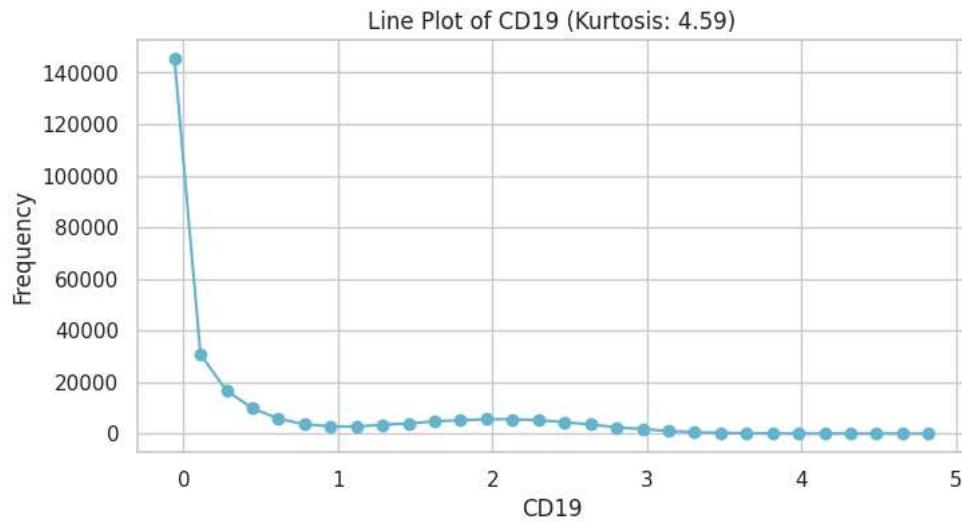
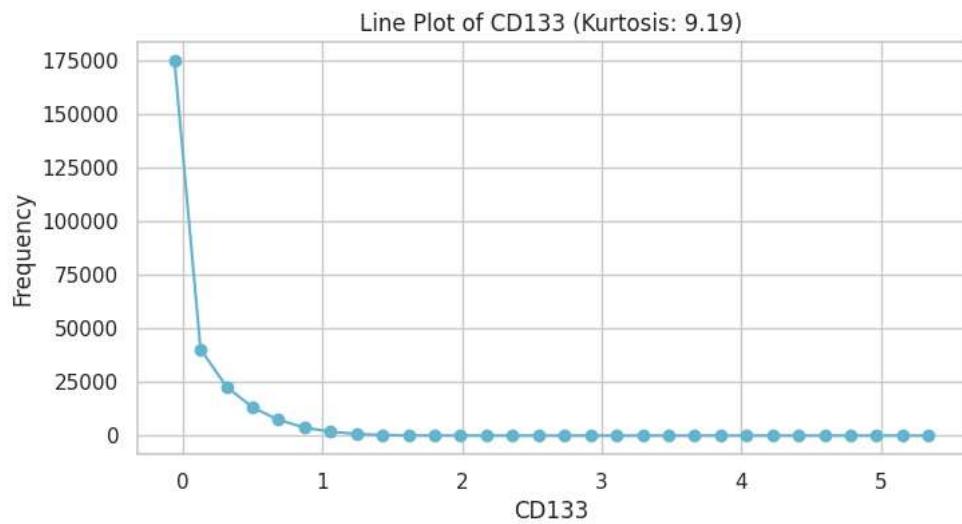
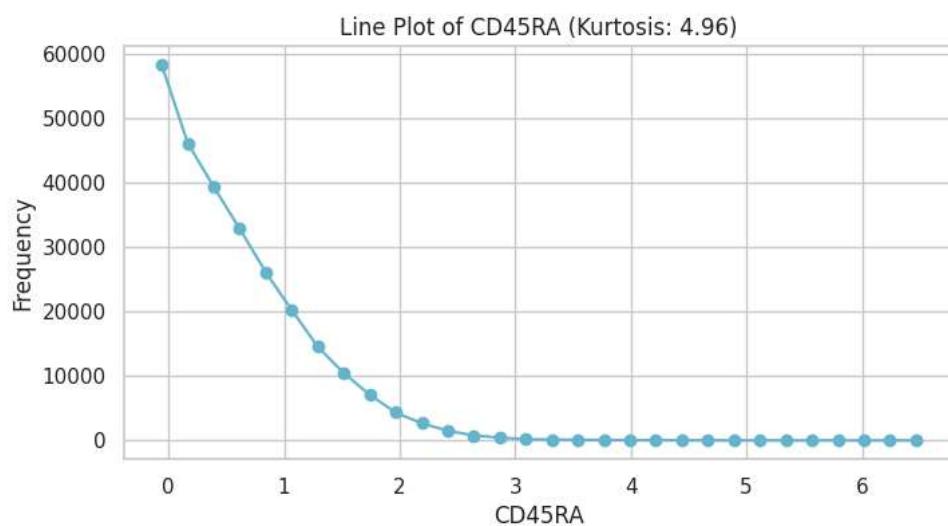
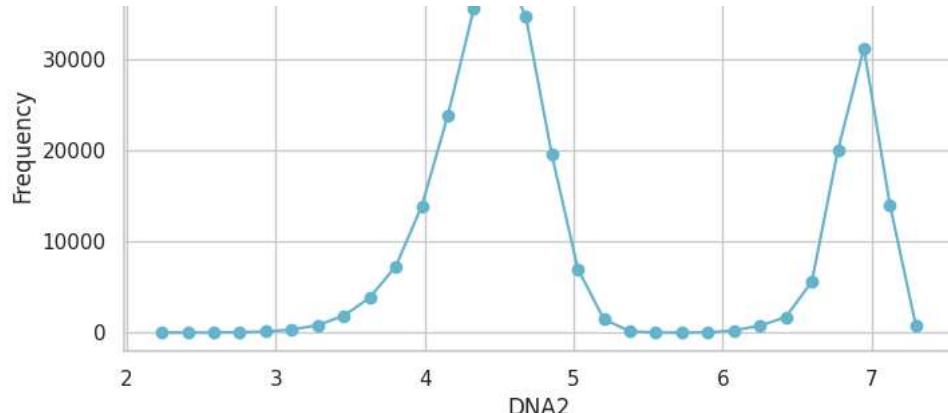


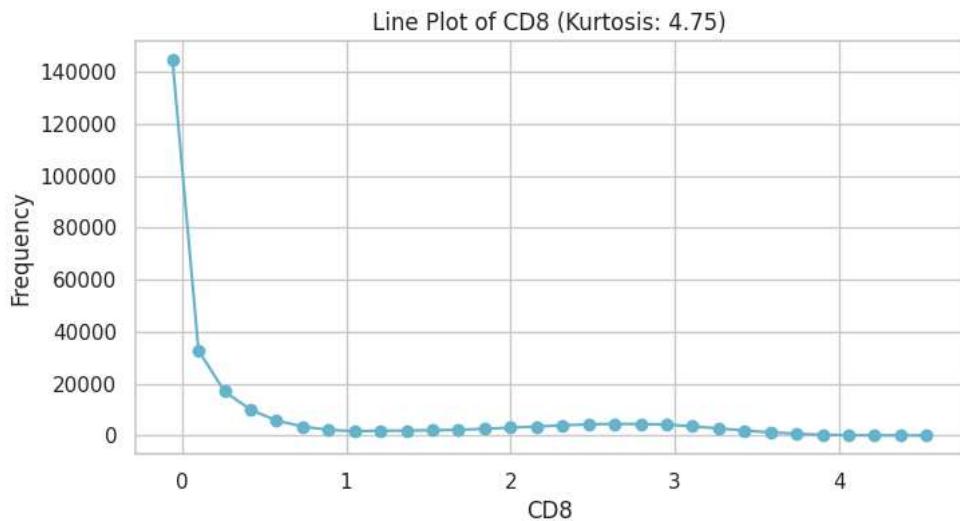
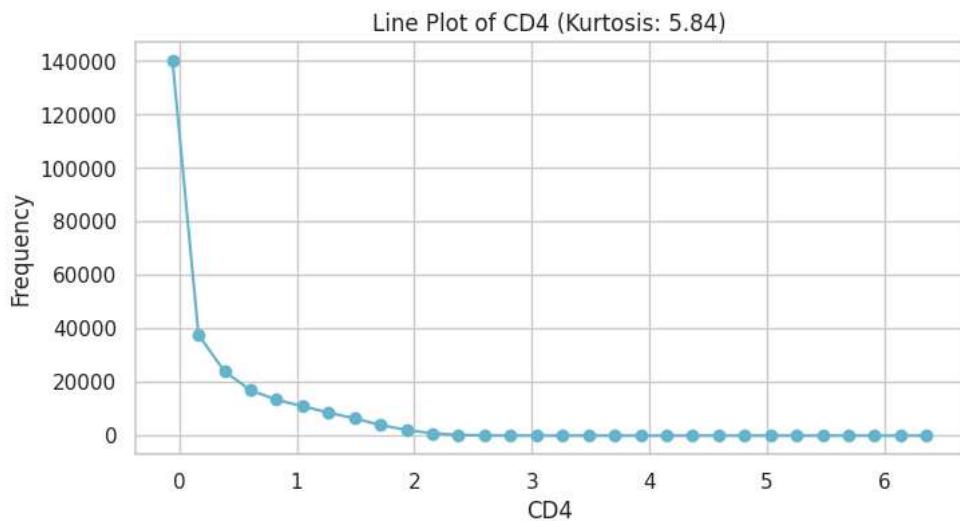
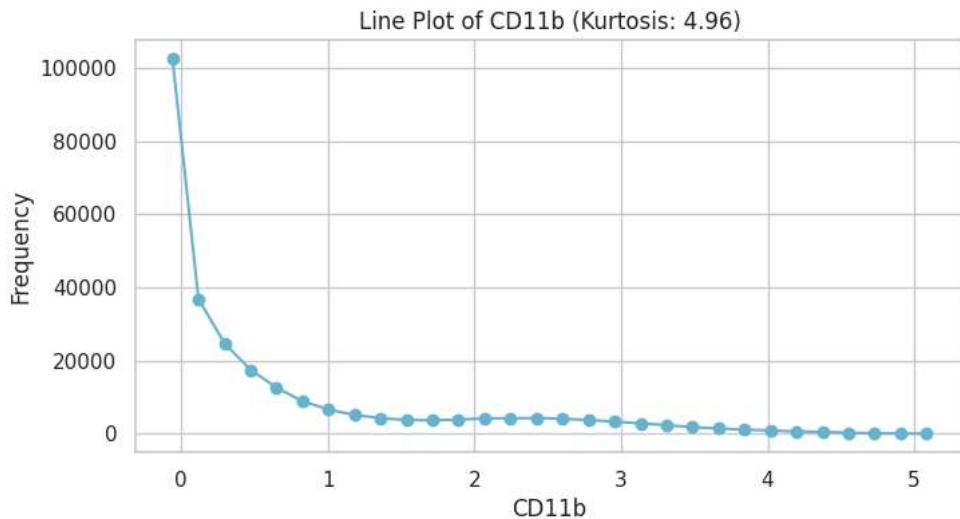
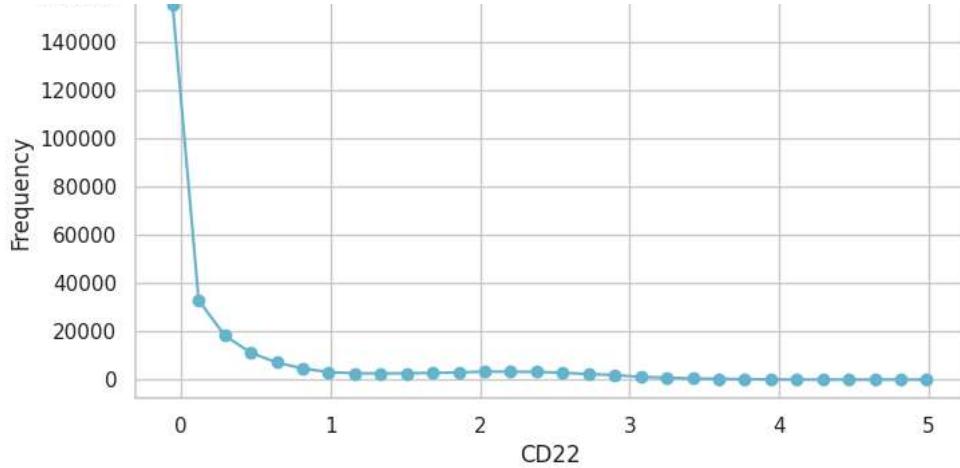
Line Plot of DNA1 (Kurtosis: 1.99)



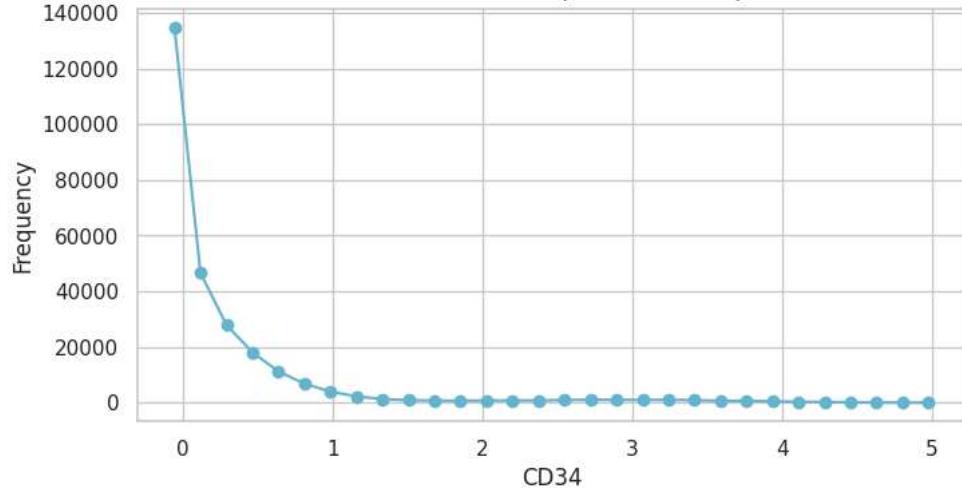
Line Plot of DNA2 (Kurtosis: 1.98)



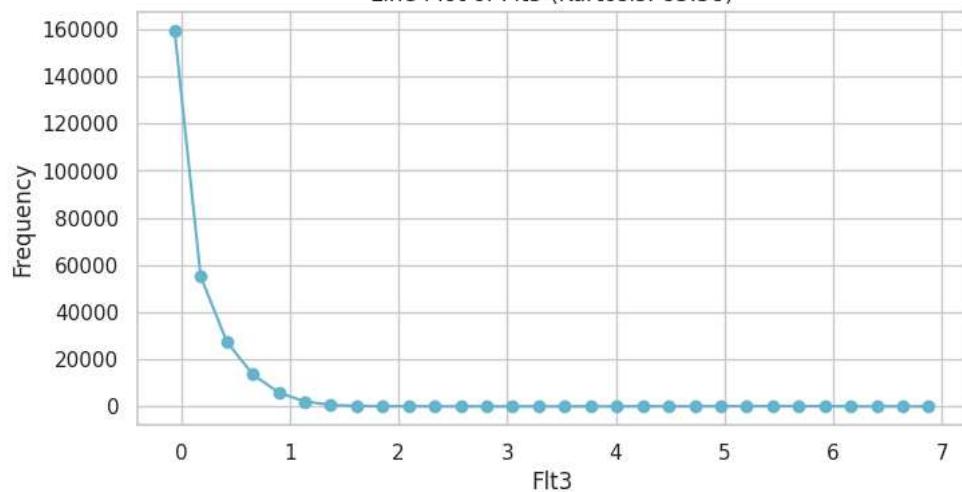




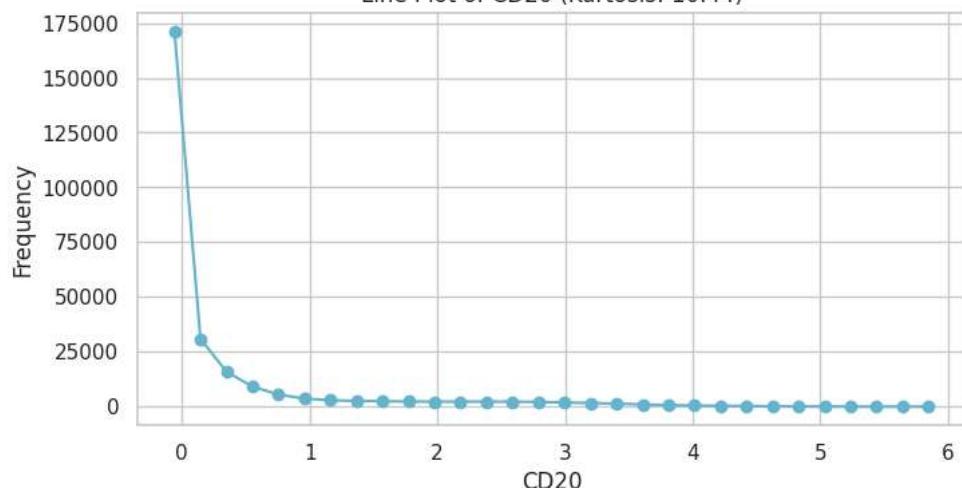
Line Plot of CD34 (Kurtosis: 16.60)



Line Plot of Flt3 (Kurtosis: 85.58)



Line Plot of CD20 (Kurtosis: 10.44)



Line Plot of CXCR4 (Kurtosis: 3.94)

