

b) Use the Students' Social Network Profile Clustering dataset from below Kaggle link and create an end-to-end project on Jupyter/Colab.

<https://www.kaggle.com/datasets/zabihullah18/students-social-network-profile-clustering/data>

i. Download the dataset from above link and load it into your Python environment. ii. Perform the EDA and do the visualizations. iii. Check the distributions/skewness in the variables and do the transformations if required. iv. Check/Treat the outliers and do the feature scaling if required. v. Create a ML model to segment the students with similar interests, demographic profiling, and trend analysis over time. vi. Try out all the 3 clustering methods (K-Mean, Hierarchical, DBSCAN) and compare their silhouette scores.

Student's Social Network Profile Clustering

Problem Statement: The objective is to use the given student social media dataset to analyse data and use K-Means, Hierarchical and DBSCAN unsupervised machine learning algorithms to create Clusters to understand the data better. Compare the clustering results using the Silhouette Score, which measures the quality of clustering, and identify the most appropriate clustering algorithm for the dataset. This analysis will allow us to better understand student behaviors, interests, and social network trends, providing valuable insights for educational institutions, social platforms, or marketing efforts targeting students.

```
In [3]: #Importing Libraries
import numpy as np
import pandas as pd
import matplotlib
#for visualizations
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [4]: df=pd.read_csv('/content/03_Clustering_Marketing.csv')
```

```
In [5]: df.head()
```

Out[5]:

	gradyear	gender	age	NumberOffriends	basketball	football	soccer	softball	volleyball	swimming	...	blonde	mall	shopping	clothes	hollist
0	2007	NaN	NaN	0	0	0	0	0	0	0	...	0	0	0	0	0
1	2007	F	17.41	49	0	0	1	0	0	1	...	0	0	0	0	0
2	2007	F	17.511	41	0	0	0	0	0	0	...	0	1	0	0	0
3	2006	F	NaN	36	0	0	0	0	0	0	...	0	0	0	0	0
4	2008	F	16.657	1	0	0	0	0	0	1	...	0	0	0	0	3

5 rows × 40 columns



In [7]:

`df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15000 entries, 0 to 14999
Data columns (total 40 columns):
 #   Column            Non-Null Count Dtype  
 ---  -- 
 0   gradyear          15000 non-null  int64  
 1   gender             13663 non-null  object  
 2   age                12504 non-null  object  
 3   NumberOfffriends  15000 non-null  int64  
 4   basketball         15000 non-null  int64  
 5   football           15000 non-null  int64  
 6   soccer              15000 non-null  int64  
 7   softball            15000 non-null  int64  
 8   volleyball          15000 non-null  int64  
 9   swimming            15000 non-null  int64  
 10  cheerleading        15000 non-null  int64  
 11  baseball            15000 non-null  int64  
 12  tennis              15000 non-null  int64  
 13  sports              15000 non-null  int64  
 14  cute                15000 non-null  int64  
 15  sex                 15000 non-null  int64  
 16  sexy                15000 non-null  int64  
 17  hot                 15000 non-null  int64  
 18  kissed               15000 non-null  int64  
 19  dance               15000 non-null  int64  
 20  band                15000 non-null  int64  
 21  marching             15000 non-null  int64  
 22  music               15000 non-null  int64  
 23  rock                15000 non-null  int64  
 24  god                 15000 non-null  int64  
 25  church              15000 non-null  int64  
 26  jesus               15000 non-null  int64  
 27  bible               15000 non-null  int64  
 28  hair                15000 non-null  int64  
 29  dress               15000 non-null  int64  
 30  blonde              15000 non-null  int64  
 31  mall                15000 non-null  int64  
 32  shopping             15000 non-null  int64  
 33  clothes              15000 non-null  int64  
 34  hollister            15000 non-null  int64  
 35  abercrombie          15000 non-null  int64  
 36  die                 15000 non-null  int64  
 37  death                15000 non-null  int64  
 38  drunk               15000 non-null  int64  
 39  drugs               15000 non-null  int64
```

```
dtypes: int64(38), object(2)
memory usage: 4.6+ MB
```

```
In [8]: df.shape
```

```
Out[8]: (15000, 40)
```

```
In [9]: df.columns
```

```
Out[9]: Index(['gradyear', 'gender', 'age', 'NumberOffriends', 'basketball',
   'football', 'soccer', 'softball', 'volleyball', 'swimming',
   'cheerleading', 'baseball', 'tennis', 'sports', 'cute', 'sex', 'sexy',
   'hot', 'kissed', 'dance', 'band', 'marching', 'music', 'rock', 'god',
   'church', 'jesus', 'bible', 'hair', 'dress', 'blonde', 'mall',
   'shopping', 'clothes', 'hollister', 'abercrombie', 'die', 'death',
   'drunk', 'drugs'],
  dtype='object')
```

```
In [10]: #Convert the first letter of each column name to uppercase
new_columns = []
for col in df.columns:
    new_columns.append(col[0].upper() + col[1:])

df.columns = new_columns
df.columns
```

```
Out[10]: Index(['Gradyear', 'Gender', 'Age', 'NumberOffriends', 'Basketball',
   'Football', 'Soccer', 'Softball', 'Volleyball', 'Swimming',
   'Cheerleading', 'Baseball', 'Tennis', 'Sports', 'Cute', 'Sex', 'Sexy',
   'Hot', 'Kissed', 'Dance', 'Band', 'Marching', 'Music', 'Rock', 'God',
   'Church', 'Jesus', 'Bible', 'Hair', 'Dress', 'Blonde', 'Mall',
   'Shopping', 'Clothes', 'Hollister', 'Abercrombie', 'Die', 'Death',
   'Drunk', 'Drugs'],
  dtype='object')
```

```
In [11]: df.isnull().sum()
```

Out[11]:	0
Gradyear	0
Gender	1337
Age	2496
NumberOffriends	0
Basketball	0
Football	0
Soccer	0
Softball	0
Volleyball	0
Swimming	0
Cheerleading	0
Baseball	0
Tennis	0
Sports	0
Cute	0
Sex	0
Sexy	0
Hot	0
Kissed	0
Dance	0
Band	0
Marching	0
Music	0
Rock	0
God	0

	0
Church	0
Jesus	0
Bible	0
Hair	0
Dress	0
Blonde	0
Mall	0
Shopping	0
Clothes	0
Hollister	0
Abercrombie	0
Die	0
Death	0
Drunk	0
Drugs	0

dtype: int64

As we can observe 'Age' and 'Gender' features have missing values lets deal with the missing values

In [12]: `df['Gender'].value_counts(dropna = False)`

```
Out[12]:
```

```
count
```

Gender

F	11057
---	-------

M	2606
---	------

NaN	1337
-----	------

dtype: int64

```
In [13]: df['Gender'].fillna(df['Gender'].mode()[0], inplace=True)
```

<ipython-input-13-90613879d37d>:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.

The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

```
df['Gender'].fillna(df['Gender'].mode()[0], inplace=True)
```

```
In [14]: #Dealing with Age variable
```

```
df['Age'] = df['Age'].astype(str)
df['Age'] = df['Age'].str.extract('(\d+).').astype(float)
df['Age'] = df['Age'].fillna(df['Age'].median())
print(df.isnull().sum())
```

```
Gradyear      0
Gender        0
Age          0
NumberOfffriends 0
Basketball    0
Football      0
Soccer        0
Softball      0
Volleyball    0
Swimming      0
Cheerleading   0
Baseball      0
Tennis         0
Sports         0
Cute          0
Sex           0
Sexy          0
Hot           0
Kissed         0
Dance         0
Band          0
Marching      0
Music         0
Rock          0
God           0
Church         0
Jesus          0
Bible          0
Hair          0
Dress          0
Blonde         0
Mall           0
Shopping       0
Clothes        0
Hollister      0
Abercrombie    0
Die            0
Death          0
Drunk          0
Drugs          0
dtype: int64
```

```
In [15]: df['Age'] = df['Age'].astype(int)
df.Age
```

Out[15]:

Age

0 17

1 17

2 17

3 17

4 16

... ...

14995 16

14996 16

14997 17

14998 17

14999 15

15000 rows × 1 columns

dtype: int64

In [16]: df.describe()

Out[16]:

	Gradyear	Age	NumberOffriends	Basketball	Football	Soccer	Softball	Volleyball	Swimming	Cheerleading
count	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000
mean	2007.496933	17.397933	29.834533	0.267000	0.255467	0.22200	0.159667	0.142933	0.135000	0.105133
std	1.116516	6.976677	35.386649	0.788851	0.702260	0.92042	0.737344	0.638747	0.548691	0.502491
min	2006.000000	4.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000
25%	2006.000000	16.000000	3.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000
50%	2008.000000	17.000000	20.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000
75%	2008.000000	18.000000	44.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000
max	2009.000000	106.000000	605.000000	22.000000	9.000000	22.00000	17.000000	14.000000	31.000000	8.000000

8 rows × 39 columns

In [17]: `df.duplicated().sum()`

Out[17]: 864

In [18]: `df.drop_duplicates(inplace=True)`In [19]: `df.duplicated().sum()`

Out[19]: 0

In [20]:

```
#Categorical data
cat_data=df.select_dtypes(include='object')
cat_data
```

Out[20]:

Gender	
0	F
1	F
2	F
3	F
4	F
...	...
14995	F
14996	F
14997	F
14998	F
14999	F

14136 rows × 1 columns

In [21]:

```
#Numerical data
num=df.select_dtypes(include=np.number)
num
```

Out[21]:

	Gradyear	Age	NumberOffriends	Basketball	Football	Soccer	Softball	Volleyball	Swimming	Cheerleading	...	Blonde	Mall	Shopping	Cloth
0	2007	17	0	0	0	0	0	0	0	0	...	0	0	0	0
1	2007	17	49	0	0	1	0	0	1	0	...	0	0	0	0
2	2007	17	41	0	0	0	0	0	0	0	...	0	1	0	0
3	2006	17	36	0	0	0	0	0	0	0	...	0	0	0	0
4	2008	16	1	0	0	0	0	0	1	0	...	0	0	0	0
...
14995	2008	16	21	0	0	0	0	0	0	0	...	0	0	0	0
14996	2008	16	50	0	0	0	0	0	0	0	...	0	0	0	0
14997	2007	17	32	0	0	0	0	0	0	1	...	0	0	0	0
14998	2007	17	20	0	0	0	0	0	0	0	...	0	0	0	0
14999	2009	15	25	0	0	7	0	0	0	0	...	0	2	1	

14136 rows × 39 columns



In [22]:

`num.corr()`

Out[22]:

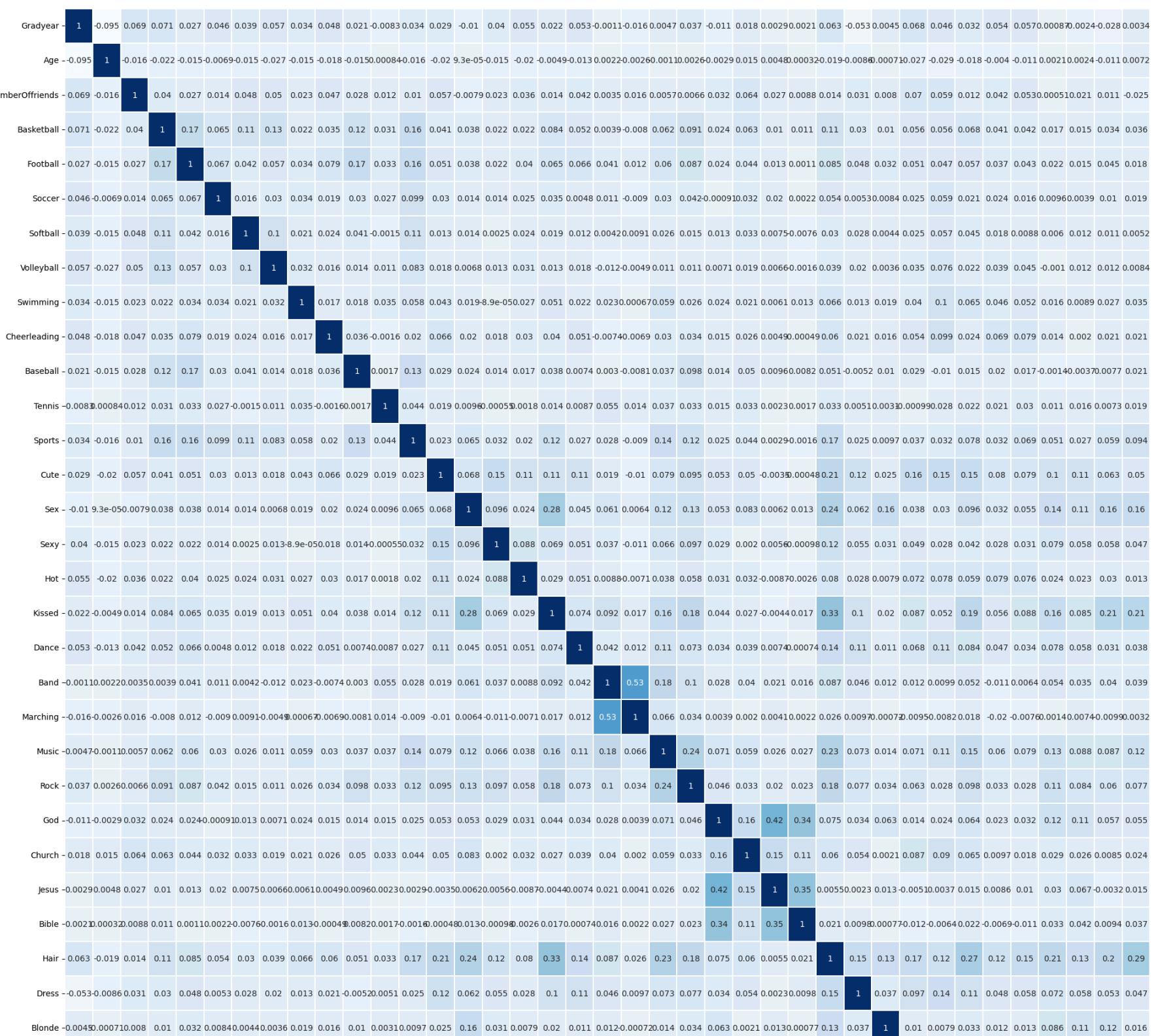
	Gradyear	Age	NumberOffriends	Basketball	Football	Soccer	Softball	Volleyball	Swimming	Cheerleading	...	Blonde	
Gradyear	1.000000	-0.095348		0.069395	0.070621	0.027375	0.045889	0.038977	0.056765	0.033908	0.048348	...	0.004507
Age	-0.095348	1.000000		-0.016129	-0.022067	-0.015034	-0.006861	-0.015153	-0.026807	-0.015089	-0.018098	...	-0.000705
NumberOffriends	0.069395	-0.016129		1.000000	0.039974	0.026639	0.014226	0.048052	0.050473	0.022851	0.046786	...	0.008035
Basketball	0.070621	-0.022067		0.039974	1.000000	0.172628	0.064581	0.111431	0.131003	0.021819	0.035300	...	0.010240
Football	0.027375	-0.015034		0.026639	0.172628	1.000000	0.067421	0.041588	0.056792	0.033654	0.078632	...	0.031657
Soccer	0.045889	-0.006861		0.014226	0.064581	0.067421	1.000000	0.016284	0.030276	0.034468	0.018924	...	0.008374
Softball	0.038977	-0.015153		0.048052	0.111431	0.041588	0.016284	1.000000	0.099944	0.020796	0.023840	...	0.004361
Volleyball	0.056765	-0.026807		0.050473	0.131003	0.056792	0.030276	0.099944	1.000000	0.031977	0.016244	...	0.003643
Swimming	0.033908	-0.015089		0.022851	0.021819	0.033654	0.034468	0.020796	0.031977	1.000000	0.017396	...	0.019472
Cheerleading	0.048348	-0.018098		0.046786	0.035300	0.078632	0.018924	0.023840	0.016244	0.017396	1.000000	...	0.015618
Baseball	0.020782	-0.014702		0.027609	0.119140	0.168284	0.030378	0.041250	0.013833	0.018127	0.036471	...	0.010188
Tennis	-0.008348	0.000844		0.012160	0.031495	0.033236	0.027211	-0.001519	0.011344	0.035164	-0.001602	...	0.003148
Sports	0.033857	-0.016288		0.010224	0.159462	0.156321	0.099328	0.110443	0.083454	0.057628	0.019960	...	0.009680
Cute	0.029054	-0.019881		0.057148	0.040986	0.051185	0.030068	0.012856	0.017918	0.043324	0.066449	...	0.024972
Sex	-0.010351	0.000093		-0.007921	0.038475	0.038304	0.014330	0.014165	0.006758	0.018648	0.020158	...	0.158213
Sexy	0.039610	-0.015439		0.023030	0.021791	0.022247	0.013717	0.002504	0.013244	-0.000089	0.018139	...	0.030568
Hot	0.054725	-0.019901		0.036135	0.022024	0.039964	0.024916	0.024436	0.031225	0.026911	0.030132	...	0.007911
Kissed	0.021624	-0.004867		0.014336	0.083858	0.064688	0.034660	0.018680	0.013174	0.050763	0.039954	...	0.020066
Dance	0.052526	-0.013233		0.041524	0.051551	0.066218	0.004821	0.012038	0.017943	0.022377	0.050813	...	0.010983
Band	-0.001111	0.002176		0.003472	0.003891	0.041128	0.011402	0.004206	-0.011803	0.023288	-0.007352	...	0.012170
Marching	-0.016154	-0.002586		0.015982	-0.007973	0.012425	-0.008997	0.009135	-0.004875	0.000674	-0.006915	...	-0.000716
Music	0.004667	-0.001091		0.005699	0.061949	0.059846	0.030448	0.026098	0.010639	0.059061	0.029925	...	0.013952
Rock	0.037113	0.002643		0.006556	0.091481	0.087156	0.041693	0.015115	0.010821	0.026444	0.033689	...	0.034460
God	-0.011408	-0.002864		0.031897	0.024477	0.024354	-0.000910	0.013026	0.007078	0.023697	0.014745	...	0.063222
Church	0.017579	0.014737		0.063704	0.062520	0.044069	0.031723	0.033210	0.019356	0.021450	0.026002	...	0.002129

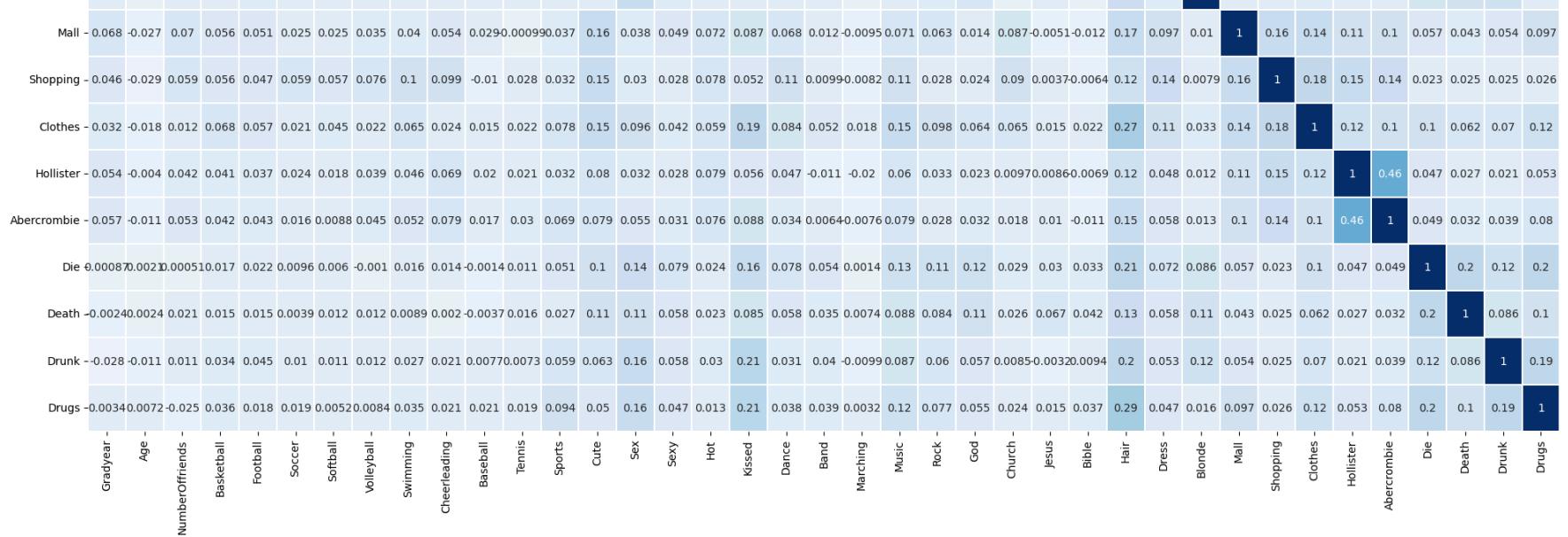
	Gradyear	Age	NumberOffriends	Basketball	Football	Soccer	Softball	Volleyball	Swimming	Cheerleading	...	Blonde	
Jesus	0.002891	0.004815		0.026824	0.010128	0.012743	0.019596	0.007456	0.006596	0.006149	0.004880	...	0.013374
Bible	0.002050	0.000316		0.008791	0.011046	0.001097	0.002228	-0.007603	-0.001634	0.012835	-0.000491	...	0.000771
Hair	0.062683	-0.018773		0.013675	0.111687	0.085003	0.053905	0.029817	0.038795	0.065778	0.060196	...	0.131953
Dress	-0.053151	-0.008609		0.030853	0.029517	0.048297	0.005287	0.027922	0.020191	0.013006	0.021397	...	0.036520
Blonde	0.004507	-0.000705		0.008035	0.010240	0.031657	0.008374	0.004361	0.003643	0.019472	0.015618	...	1.000000
Mall	0.068029	-0.026712		0.069559	0.056061	0.050640	0.024585	0.024624	0.034648	0.039982	0.054305	...	0.010099
Shopping	0.045939	-0.028594		0.058802	0.055657	0.046526	0.058752	0.057119	0.075966	0.101534	0.098632	...	0.007891
Clothes	0.032179	-0.017968		0.011582	0.067625	0.057072	0.021238	0.044712	0.021566	0.065084	0.024484	...	0.032935
Hollister	0.054405	-0.003967		0.042245	0.040547	0.037005	0.024344	0.017714	0.039267	0.045526	0.069431	...	0.012113
Abercrombie	0.057239	-0.011134		0.053161	0.041868	0.043431	0.015833	0.008801	0.045067	0.051941	0.079085	...	0.013042
Die	0.000874	0.002069		0.000509	0.016840	0.022163	0.009560	0.005963	-0.000996	0.016108	0.014416	...	0.086384
Death	-0.002359	0.002414		0.021406	0.014976	0.014606	0.003899	0.011846	0.011669	0.008872	0.001983	...	0.108255
Drunk	-0.028240	-0.011385		0.010933	0.034016	0.044958	0.010186	0.010988	0.012022	0.027333	0.021066	...	0.123853
Drugs	0.003450	0.007196		-0.025218	0.036156	0.018245	0.018992	0.005159	0.008413	0.034579	0.021287	...	0.015552

39 rows × 39 columns

```
In [48]: plt.figure(figsize=(30,30))
sns.heatmap(num.corr(), annot=True, linewidth="0.2", cmap='Blues')
```

```
Out[48]: <Axes: >
```





```
In [49]: corr_matrix = df2.corr()
def categorize_correlation(corr_value):
    if corr_value >= 0.7:
        return 'Strong'
    elif corr_value >= 0.3:
        return 'Moderate'
    else:
        return 'Low'

# Create a categorized correlation matrix
categorized_corr = corr_matrix.applymap(categorize_correlation)

#print("Correlation Matrix:")
#print(corr_matrix)

print("\nCategorized Correlation Matrix:")
#print(categorized_corr)
categorized_corr
```

Categorized Correlation Matrix:

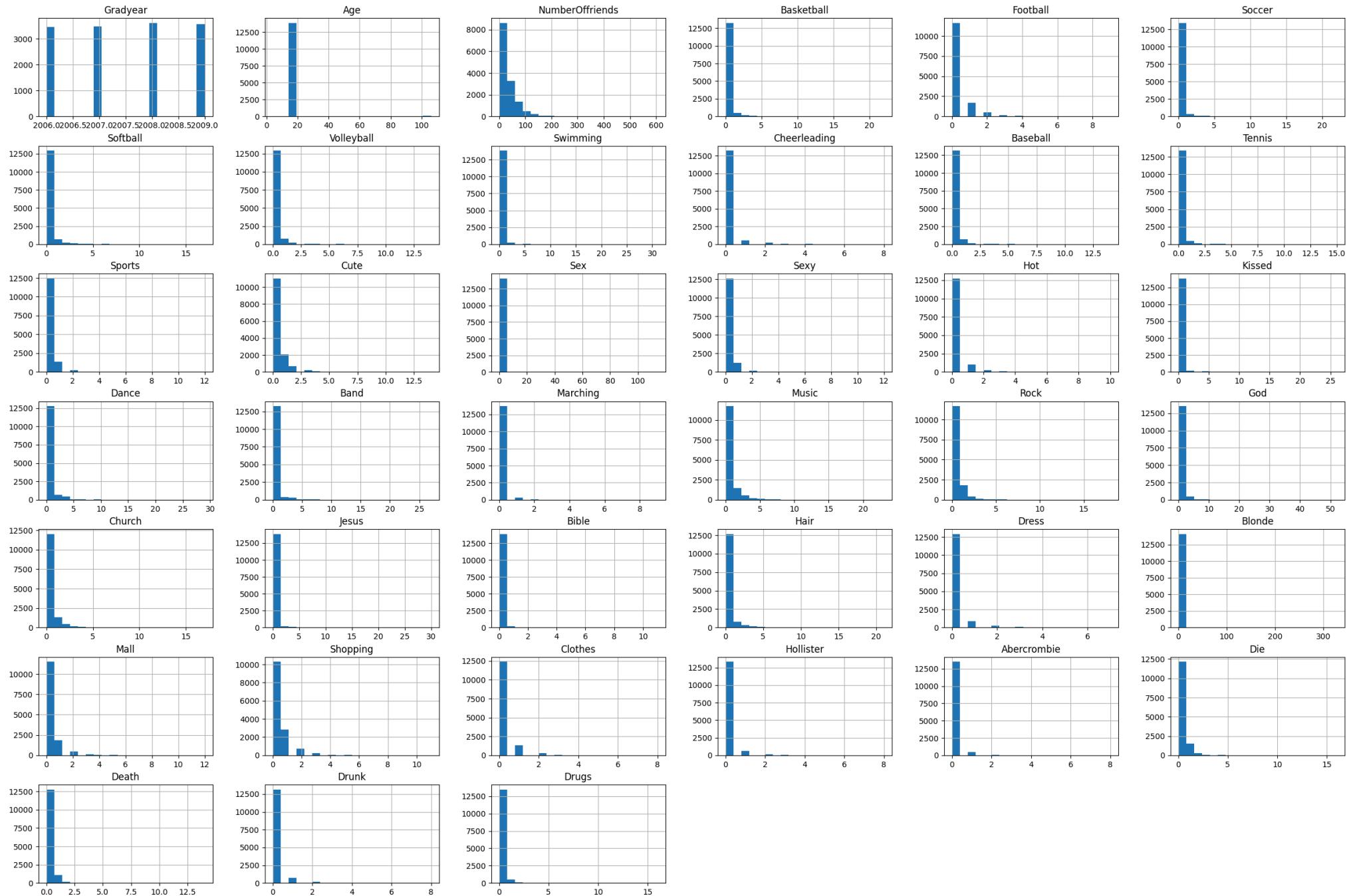
Out[49]:

	Gradyear	Gender	Age	NumberOffriends	Basketball	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	Shoppi
Gradyear	Strong	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Gender	Low	Strong	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Age	Low	Low	Strong	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
NumberOffriends	Low	Low	Low	Strong	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Basketball	Low	Low	Low	Low	Strong	Low	Low	Low	Low	Low	...	Low	Low	Low
Football	Low	Low	Low	Low	Low	Strong	Low	Low	Low	Low	...	Low	Low	Low
Soccer	Low	Low	Low	Low	Low	Low	Strong	Low	Low	Low	...	Low	Low	Low
Softball	Low	Low	Low	Low	Low	Low	Low	Strong	Low	Low	...	Low	Low	Low
Volleyball	Low	Low	Low	Low	Low	Low	Low	Low	Strong	Low	...	Low	Low	Low
Swimming	Low	Low	Low	Low	Low	Low	Low	Low	Low	Strong	...	Low	Low	Low
Cheerleading	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Baseball	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Tennis	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Sports	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Cute	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Sex	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Sexy	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Hot	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Kissed	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Dance	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Band	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Marching	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Music	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
Rock	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low
God	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	...	Low	Low	Low

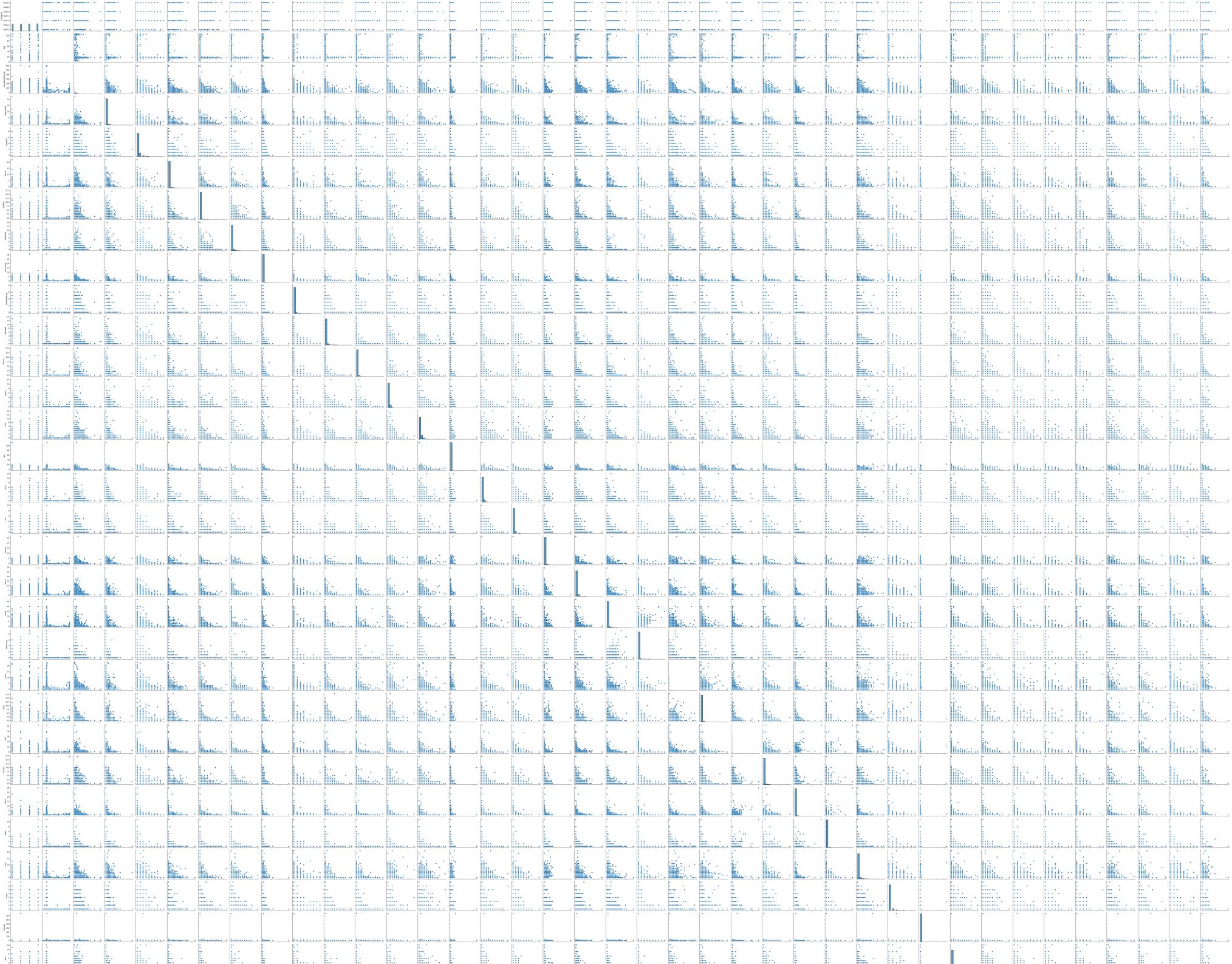
	Gradyear	Gender	Age	NumberOffriends	Basketball	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	Shoppi
Church	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Jesus	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Bible	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Hair	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Dress	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Blonde	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Strong	Low
Mall	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Strong
Shopping	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Strong
Clothes	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Hollister	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Abercrombie	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Die	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Death	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Drunk	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low
Drugs	Low	Low	Low		Low	Low	Low	Low	Low	Low	Low	...	Low	Low

40 rows × 40 columns

```
In [50]: # Distribution of the numerical features
df.hist(bins=20, figsize=(30, 20))
plt.show()
```



```
In [51]: sns.pairplot(df)
plt.show()
```





```
In [52]: #converting categorical data into numerical data
df1=df.copy()
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df1['Gender'] = le.fit_transform(df1['Gender'])
df1.head()
```

```
Out[52]:
```

	Gradyear	Gender	Age	NumberOffriends	Basketball	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	Shopping	Clothes	Hollist
0	2007	0	17	0	0	0	0	0	0	0	...	0	0	0	0	0
1	2007	0	17	49	0	0	1	0	0	1	...	0	0	0	0	0
2	2007	0	17	41	0	0	0	0	0	0	...	0	1	0	0	0
3	2006	0	17	36	0	0	0	0	0	0	...	0	0	0	0	0
4	2008	0	16	1	0	0	0	0	0	1	...	0	0	0	0	3

5 rows × 40 columns



```
In [53]: df1.skew()
```

Out[53]:

0

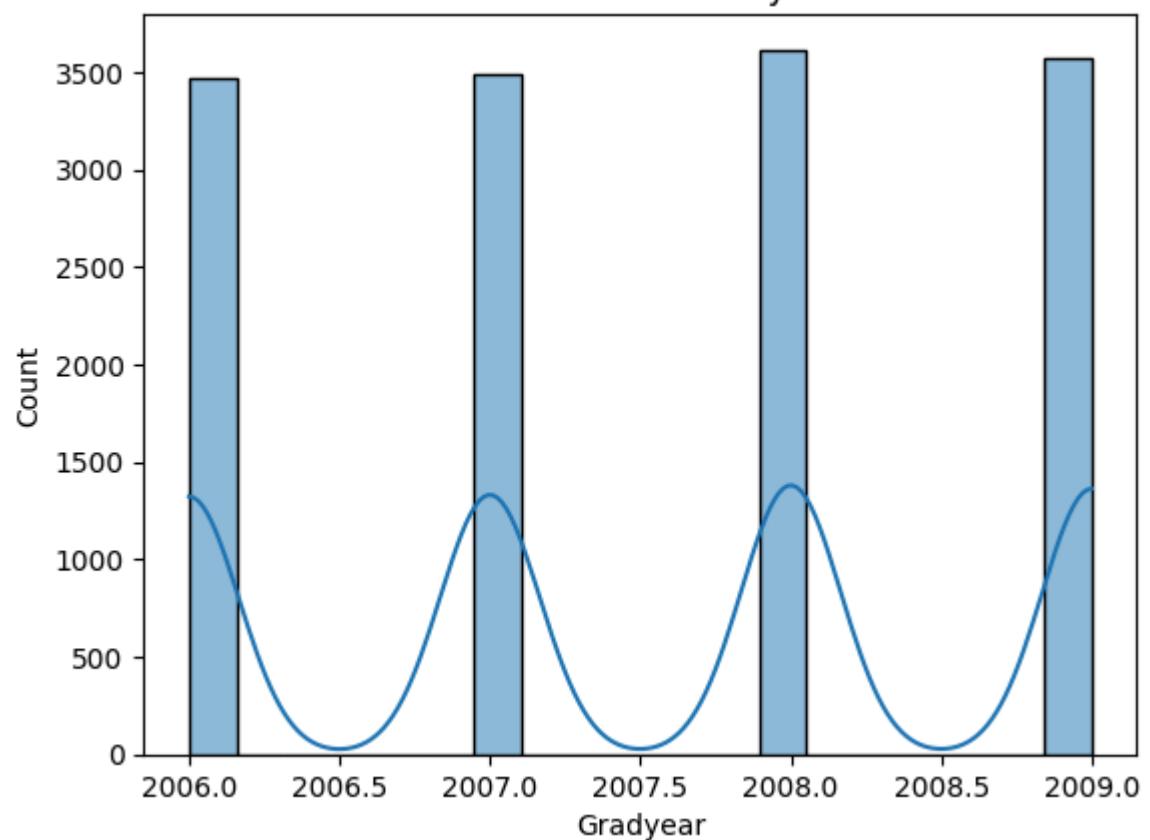
Gradyear	-0.023387
Gender	1.701090
Age	11.266039
NumberOffriends	2.697547
Basketball	5.517250
Football	3.984203
Soccer	7.060783
Softball	7.609135
Volleyball	7.321969
Swimming	15.963560
Cheerleading	6.708042
Baseball	9.087573
Tennis	10.909385
Sports	5.710232
Cute	4.035813
Sex	48.616806
Sexy	6.495363
Hot	5.512999
Kissed	12.503121
Dance	5.575502
Band	7.202506
Marching	11.325304
Music	4.308587
Rock	6.273353
God	11.229820

	0
Church	5.383541
Jesus	18.629321
Bible	20.792683
Hair	4.500486
Dress	5.372009
Blonde	113.836585
Mall	4.557489
Shopping	2.843203
Clothes	4.428123
Hollister	7.553622
Abercrombie	7.923235
Die	6.805591
Death	7.876473
Drunk	7.319200
Drugs	12.146026

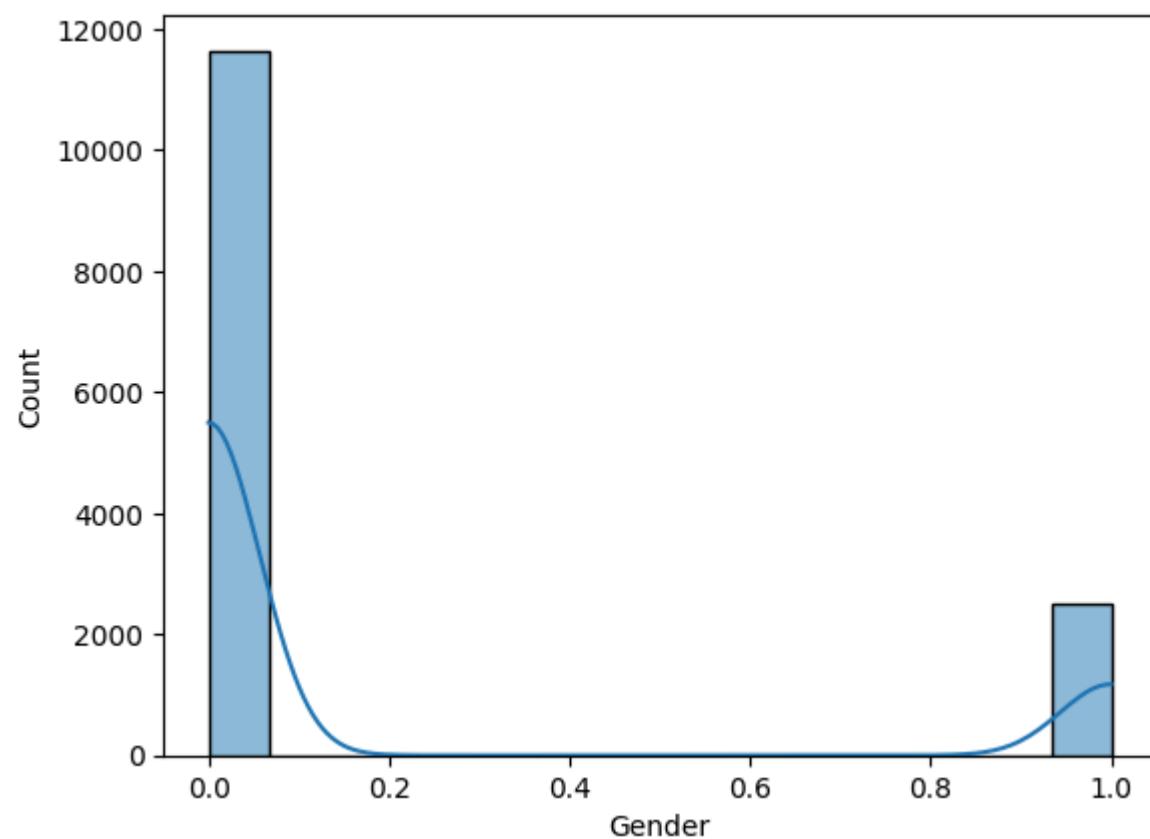
dtype: float64

```
In [54]: for col in df1.select_dtypes(include=['float64', 'int64']).columns:  
    sns.histplot(df1[col], kde=True)  
    plt.title(f'Distribution of {col}')  
    plt.show()
```

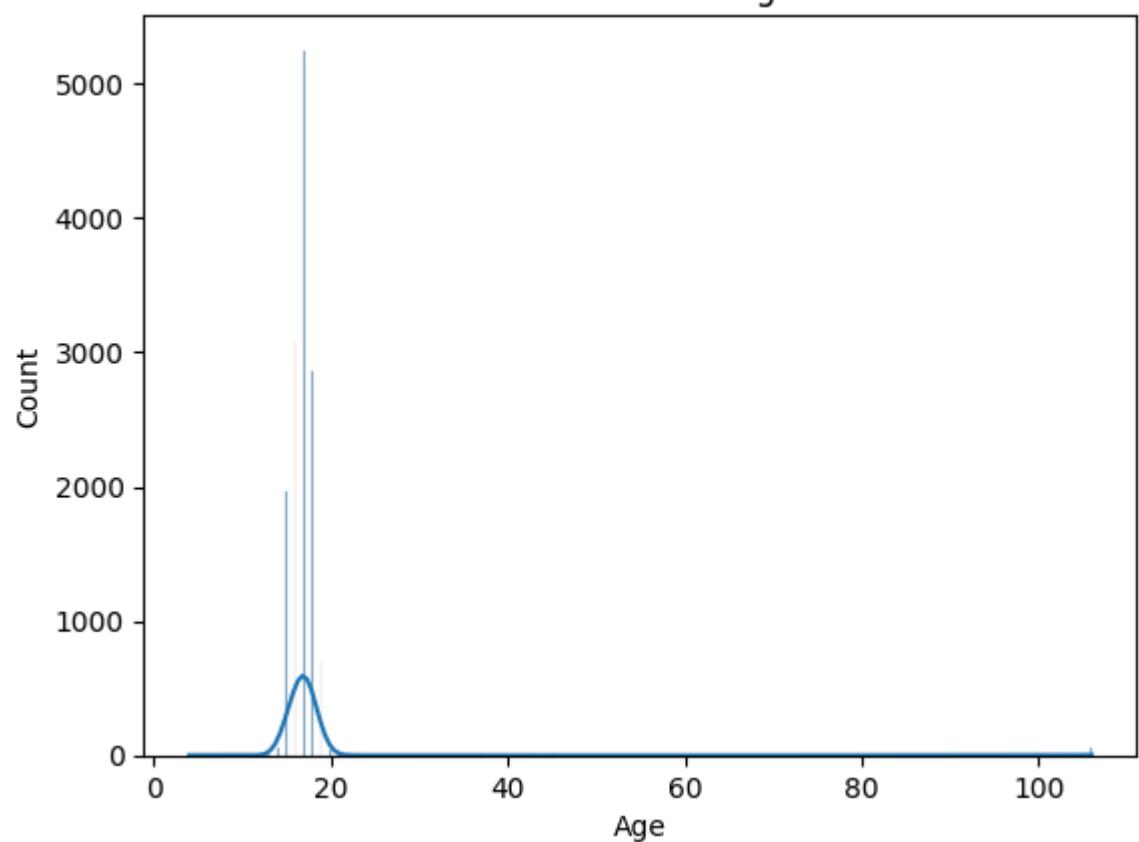
Distribution of Gradyear



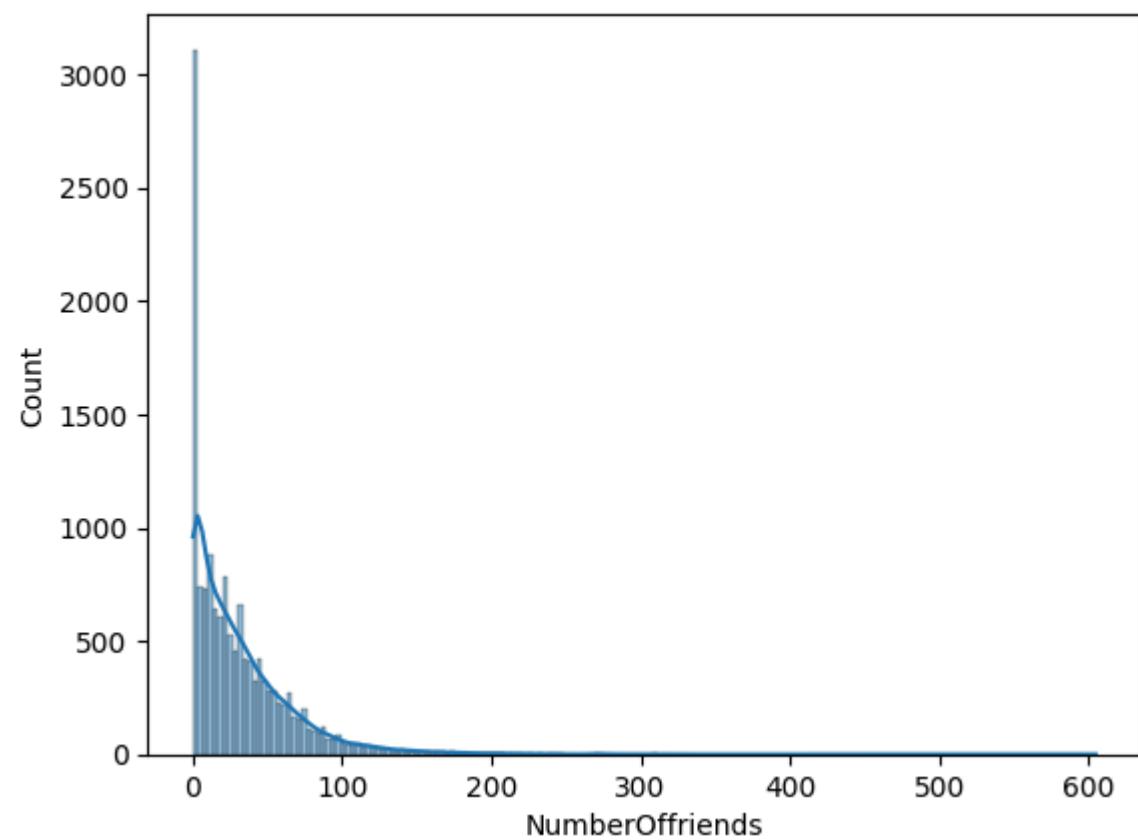
Distribution of Gender

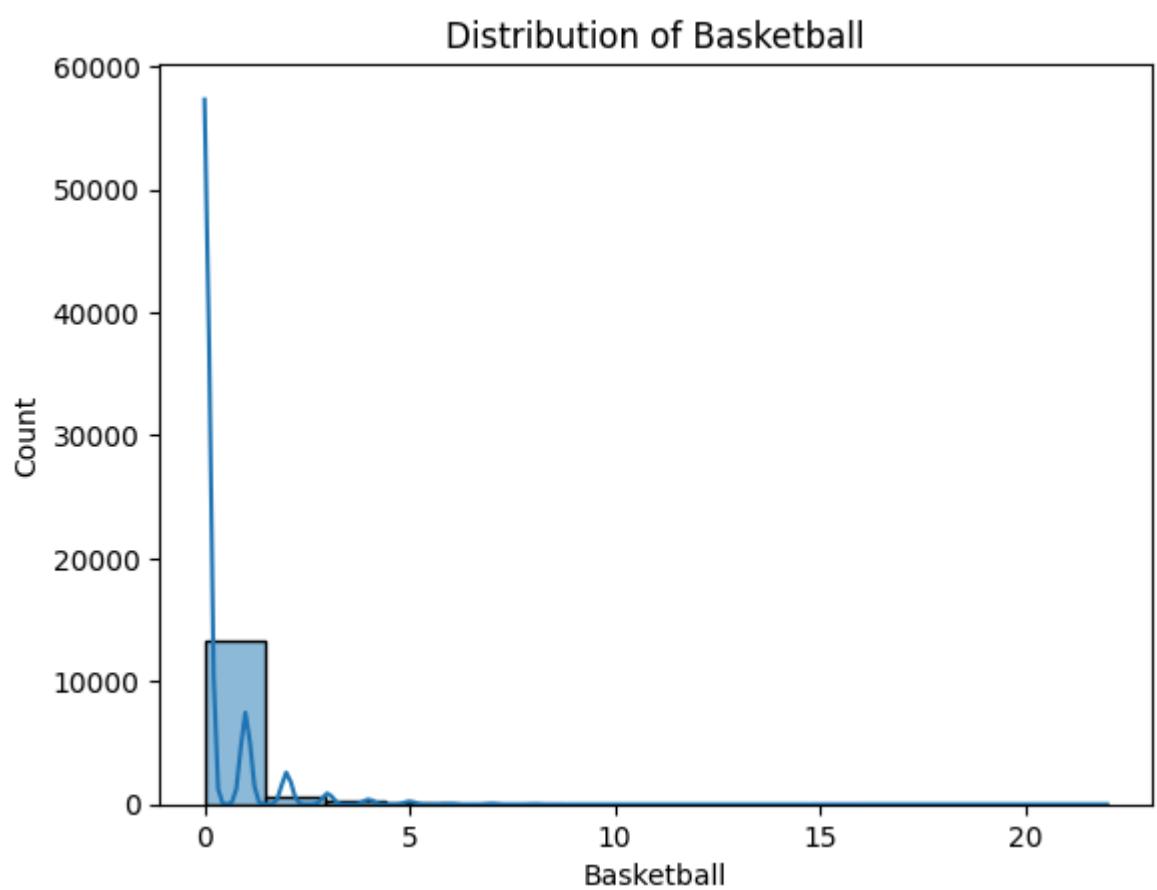


Distribution of Age

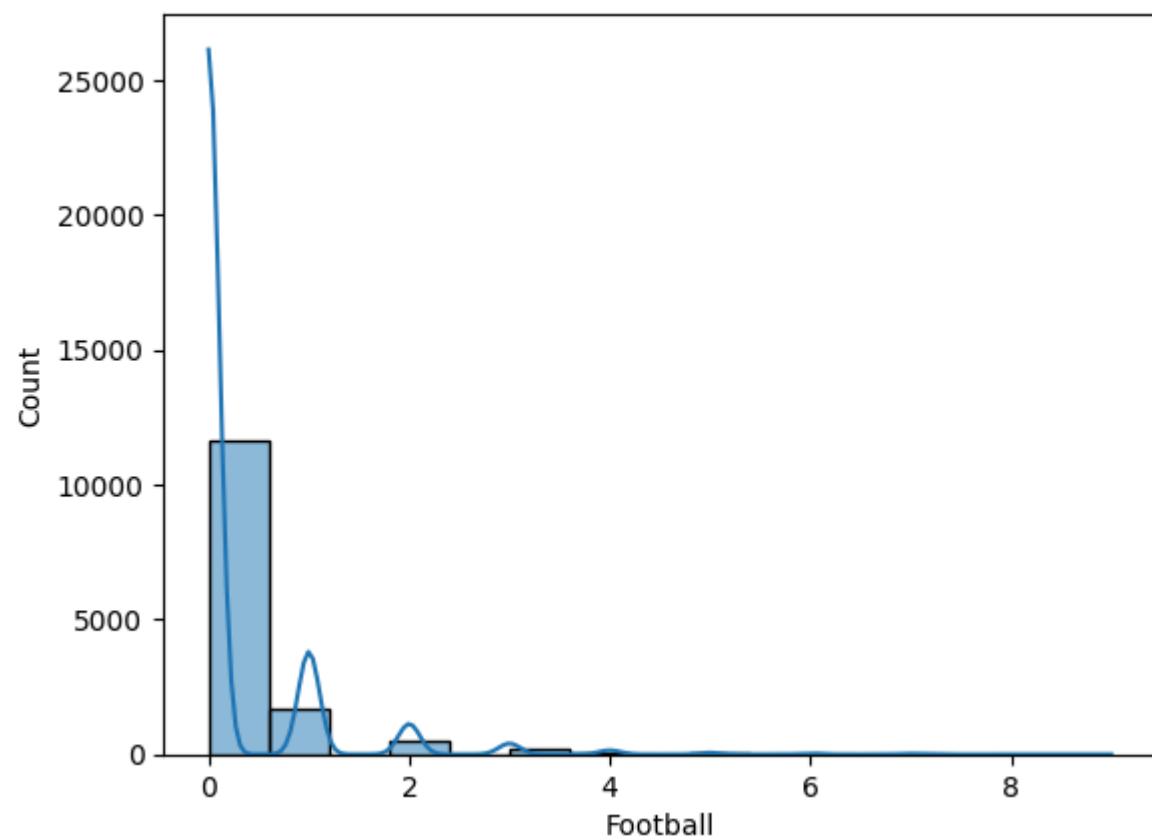


Distribution of NumberOffriends

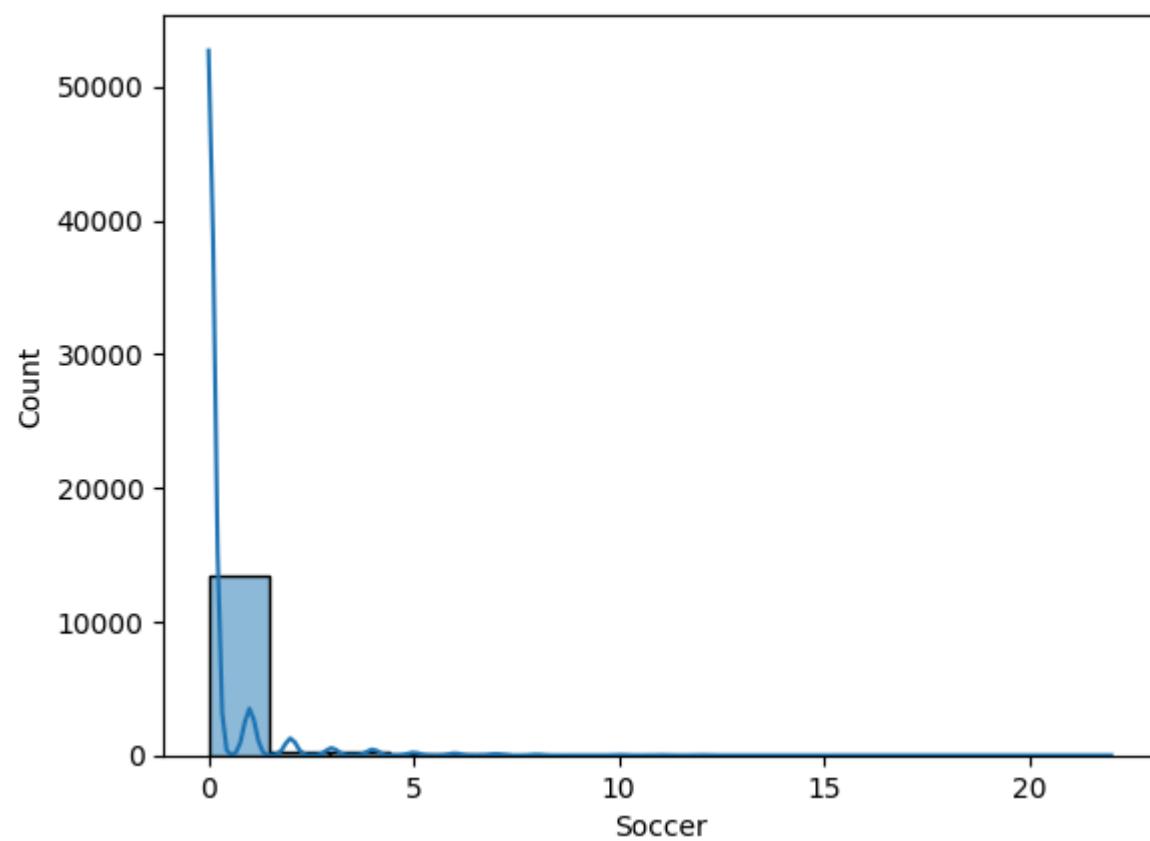




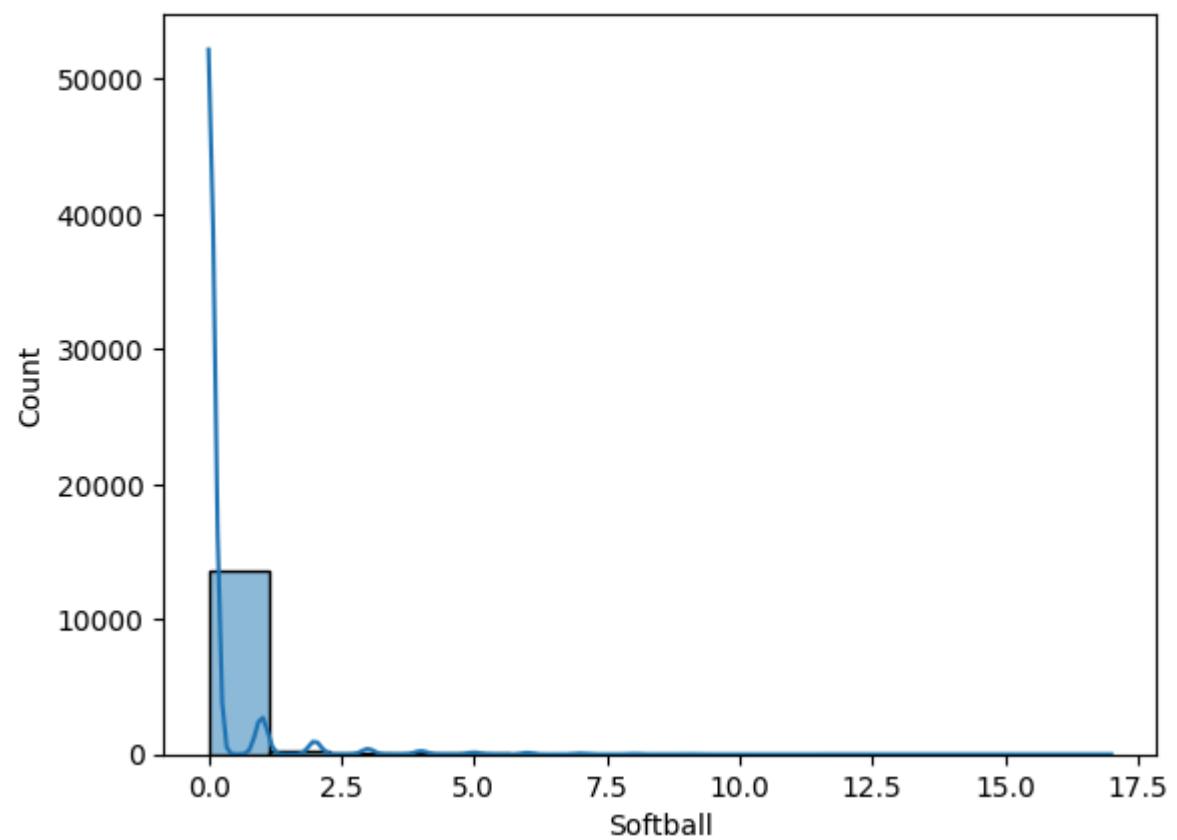
Distribution of Football



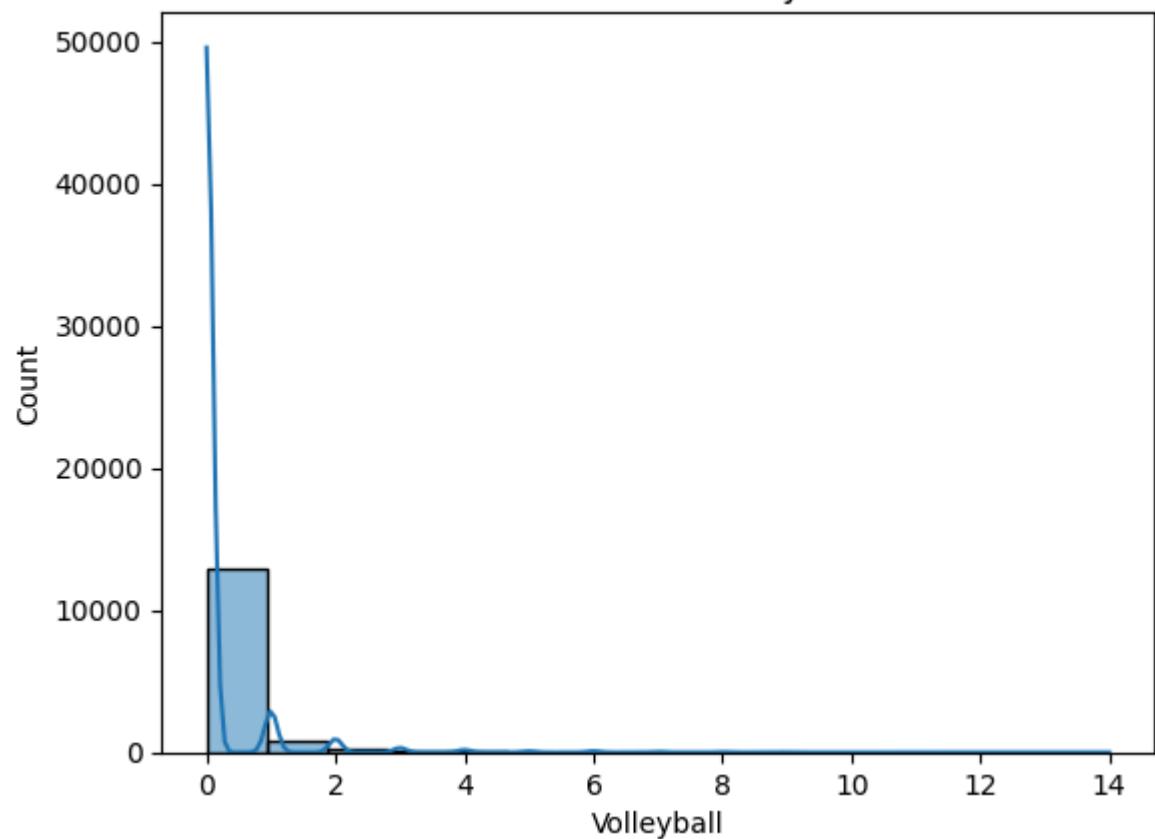
Distribution of Soccer



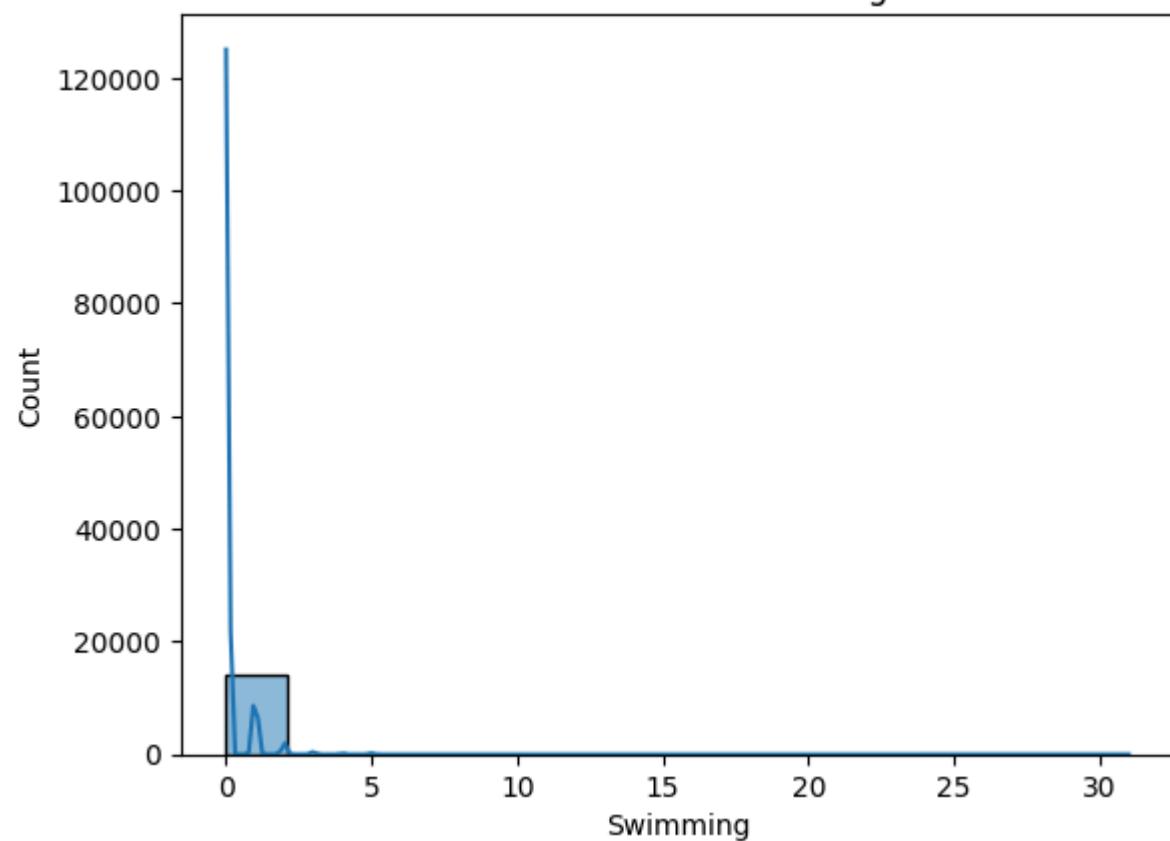
Distribution of Softball



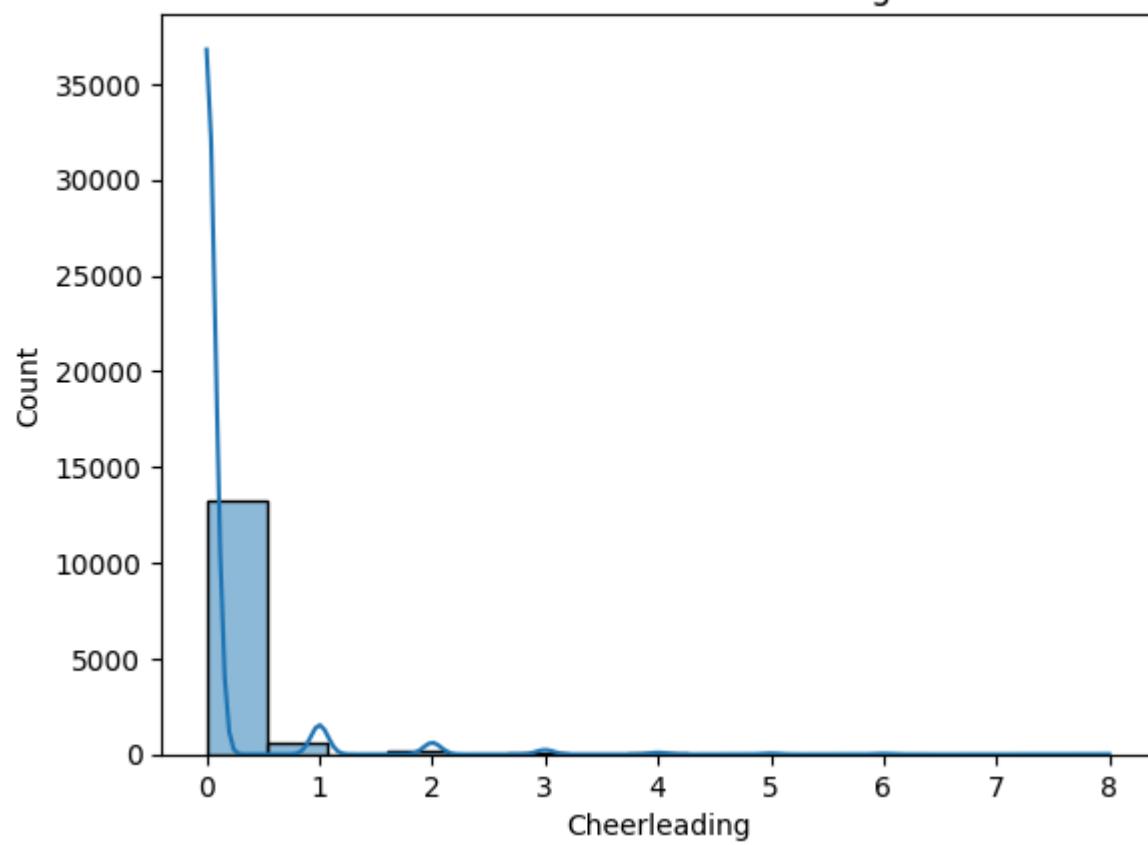
Distribution of Volleyball



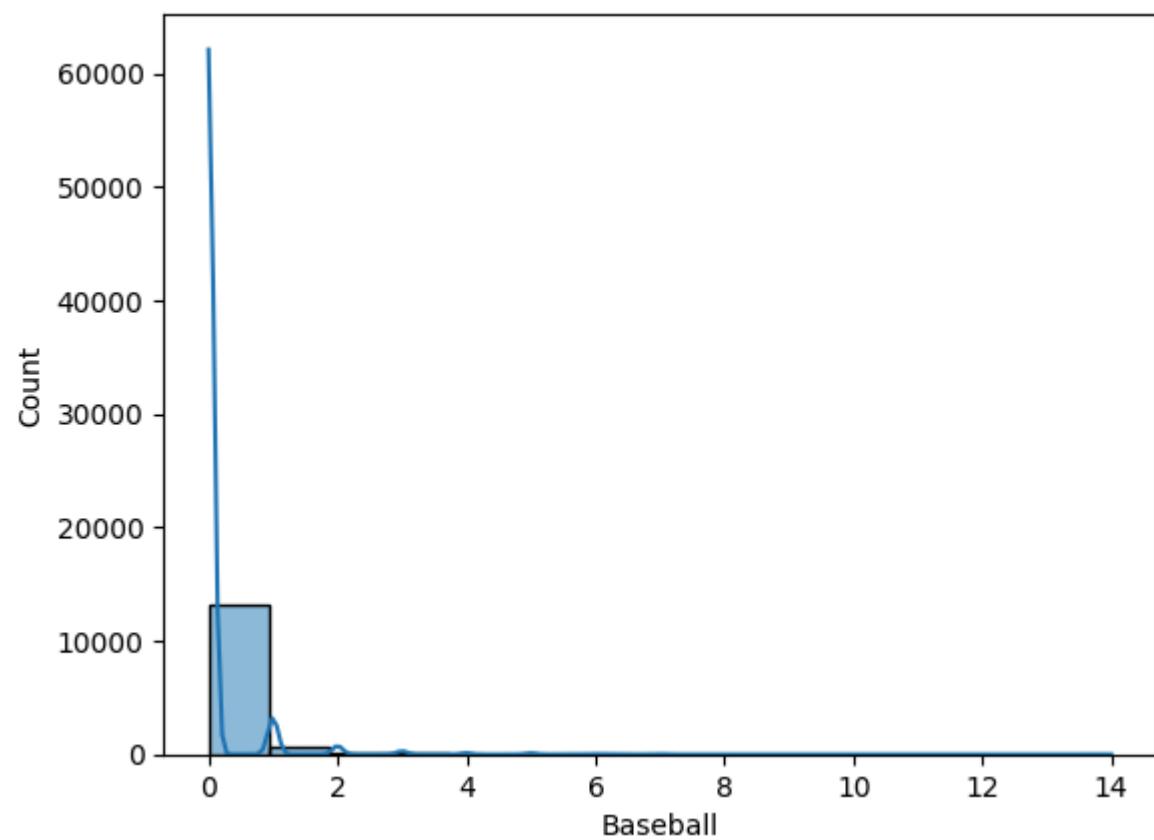
Distribution of Swimming



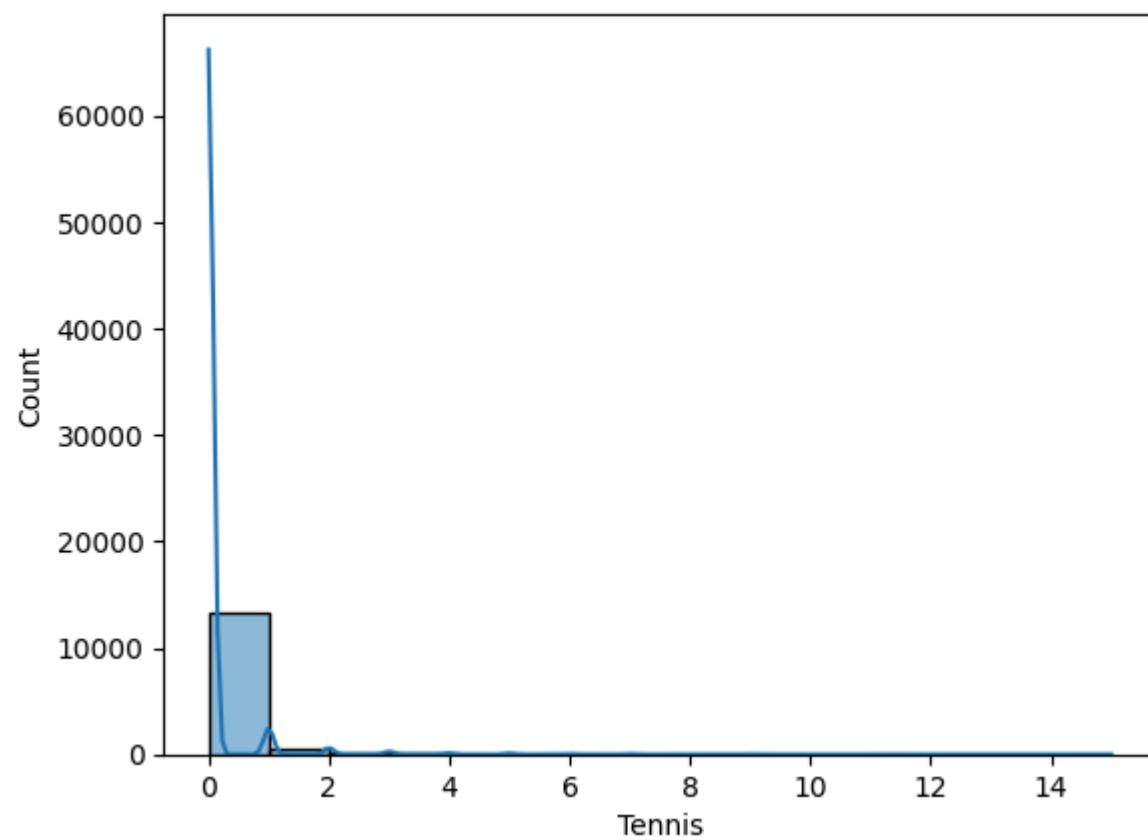
Distribution of Cheerleading



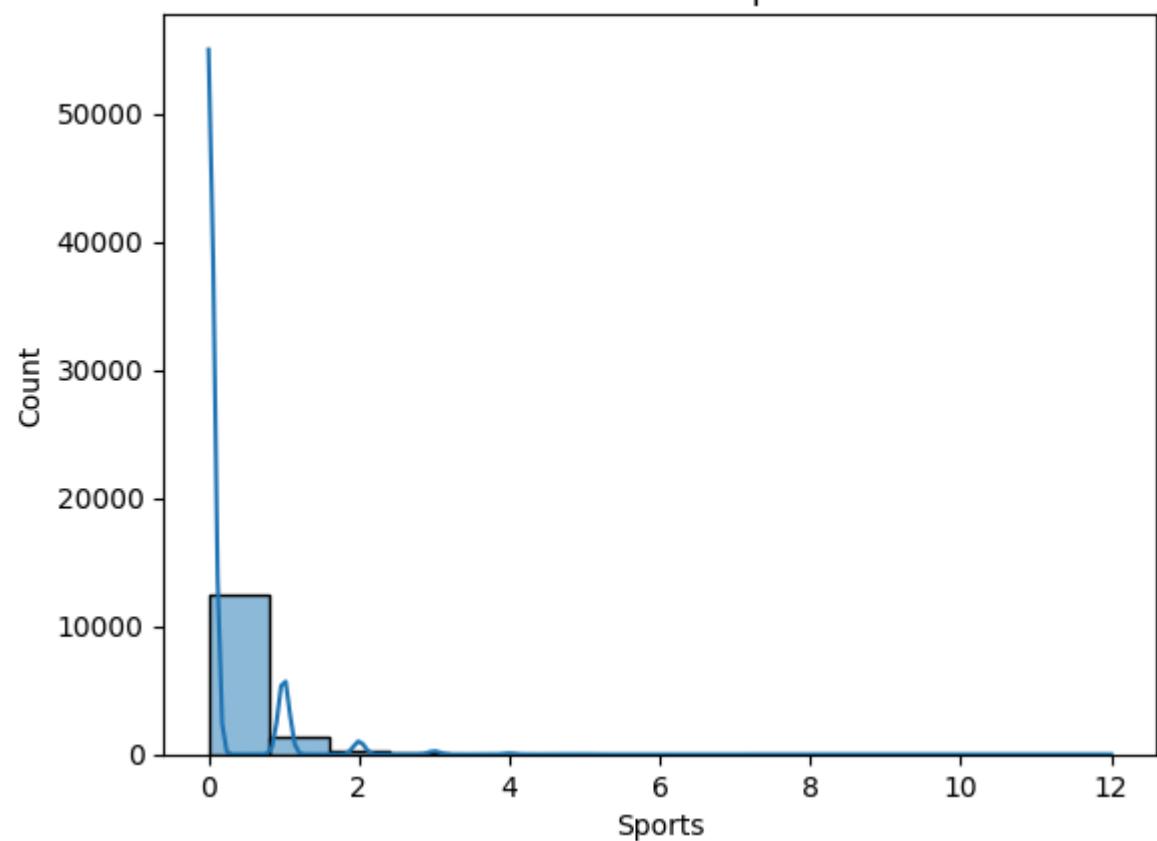
Distribution of Baseball

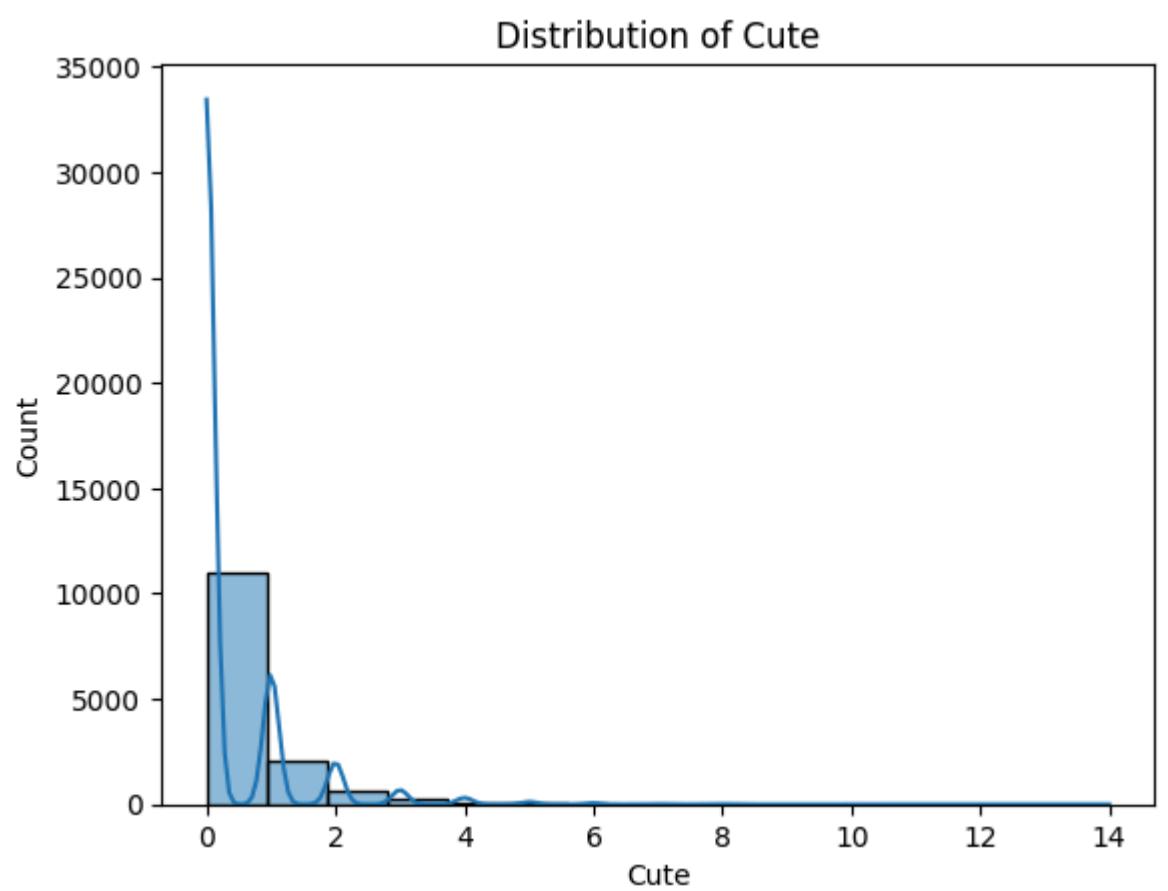


Distribution of Tennis

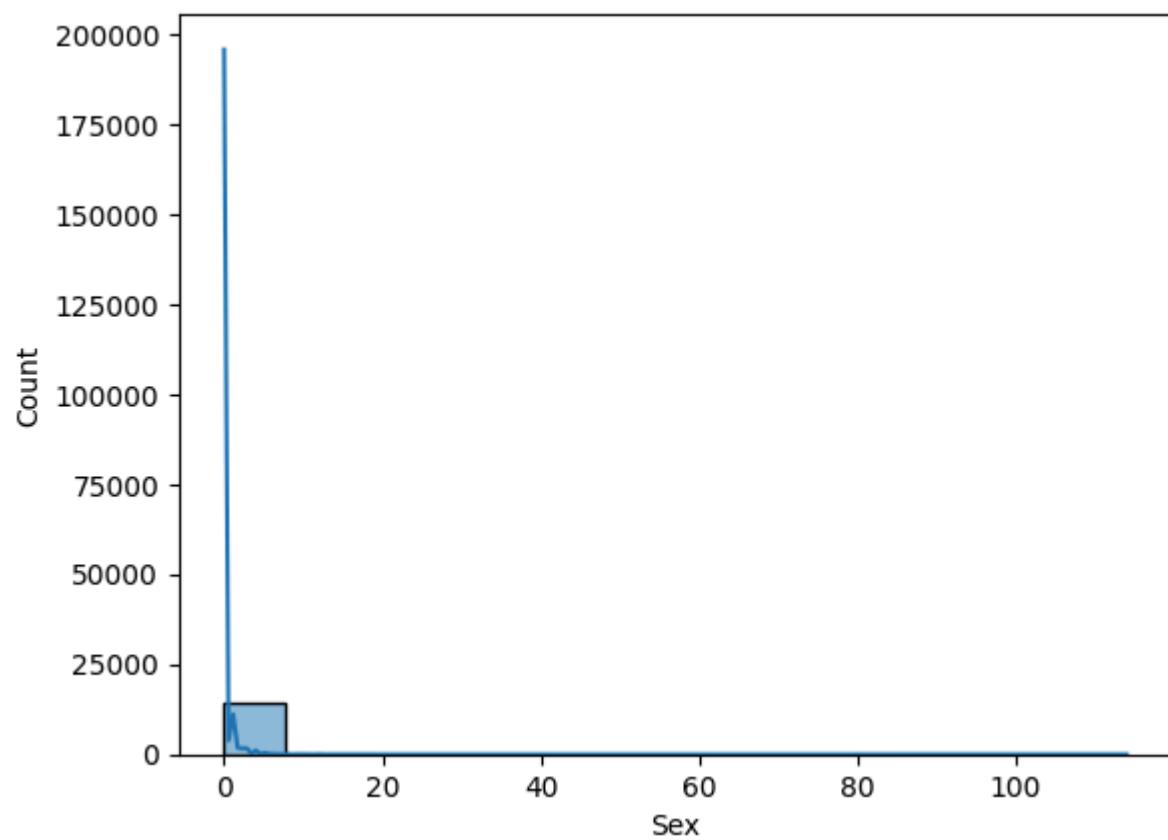


Distribution of Sports

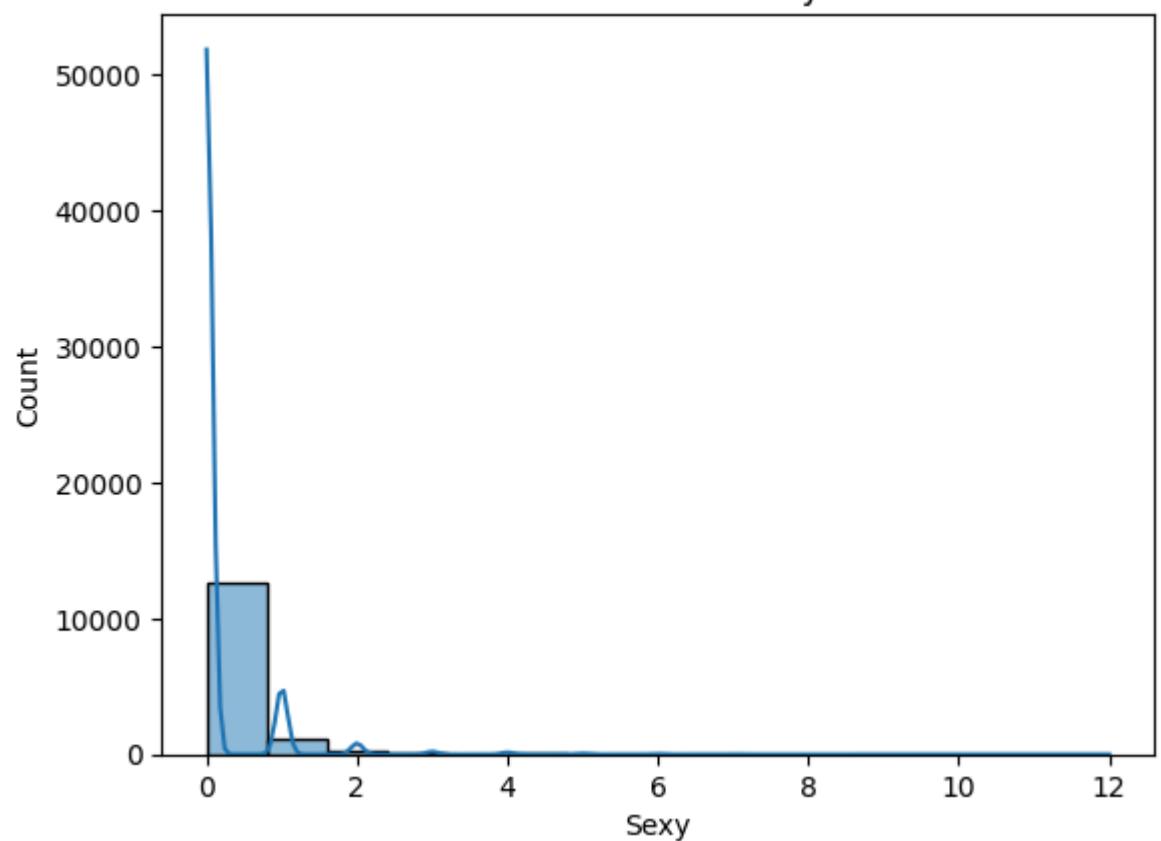




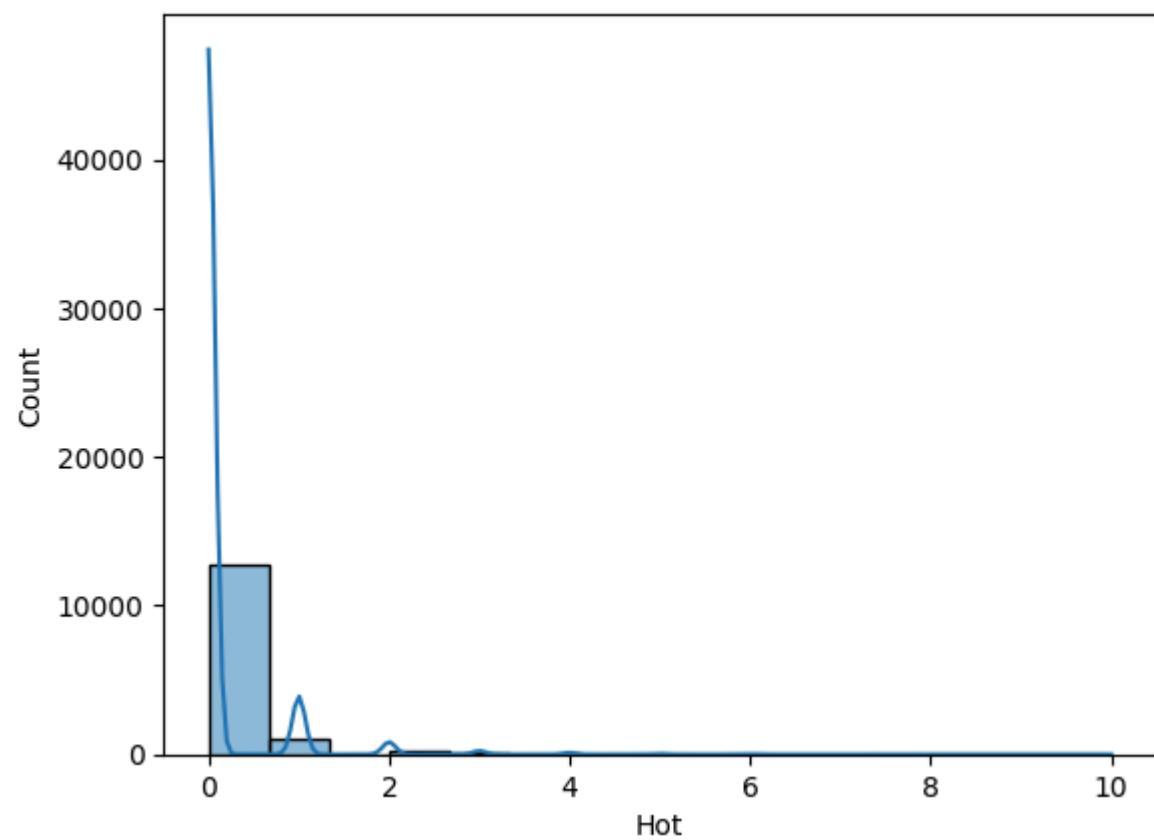
Distribution of Sex

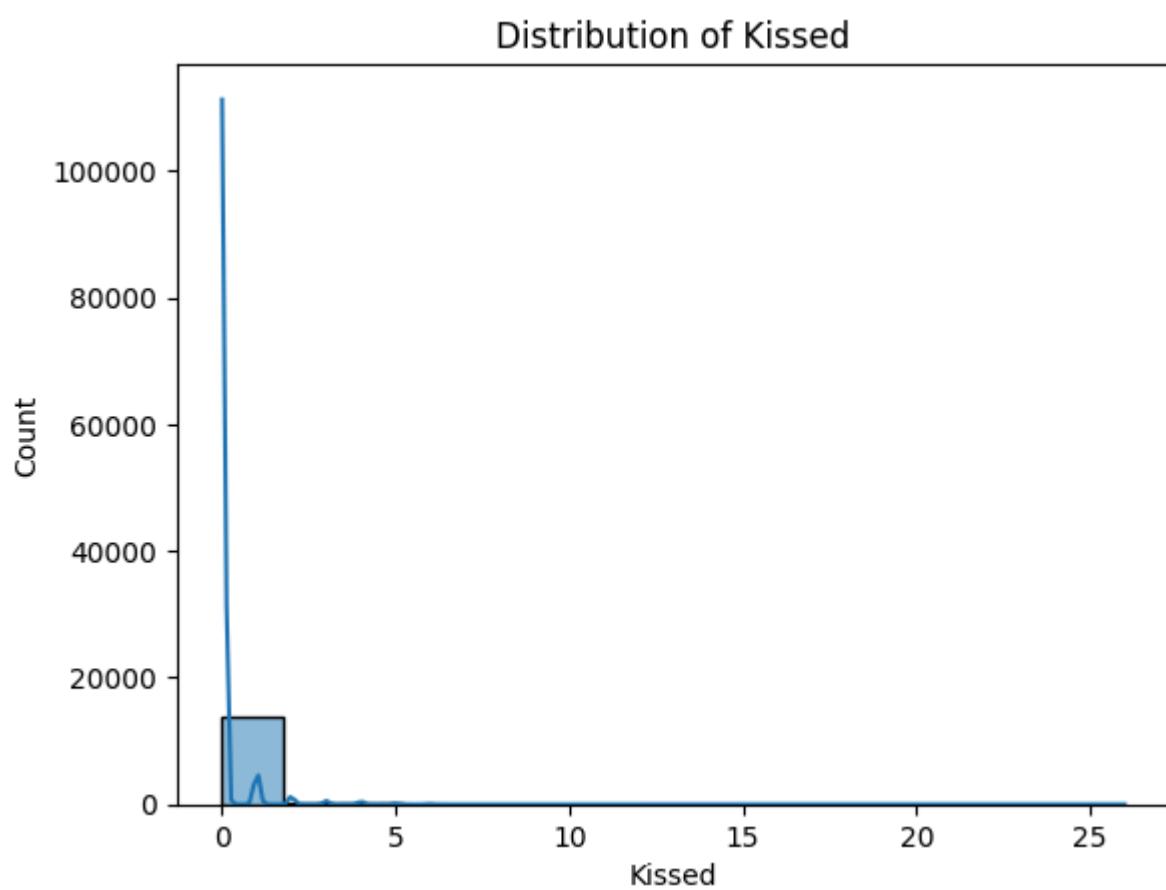


Distribution of Sexy

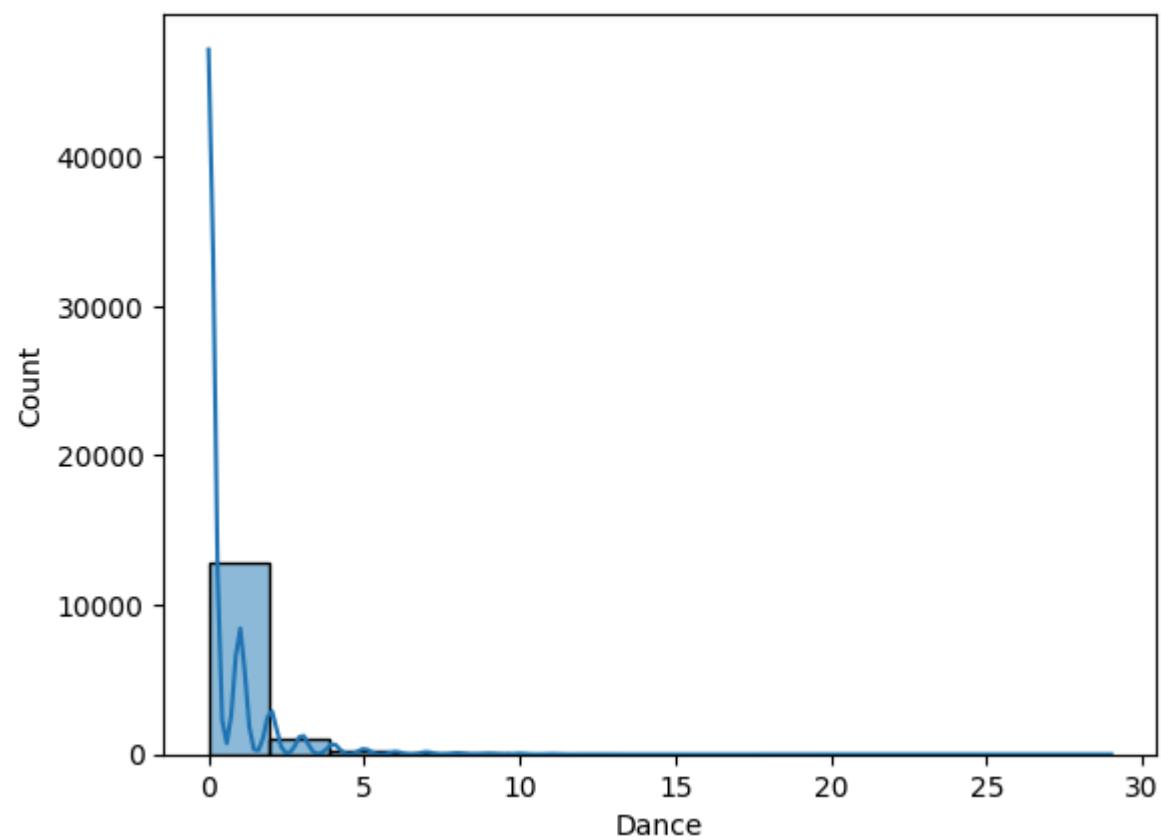


Distribution of Hot

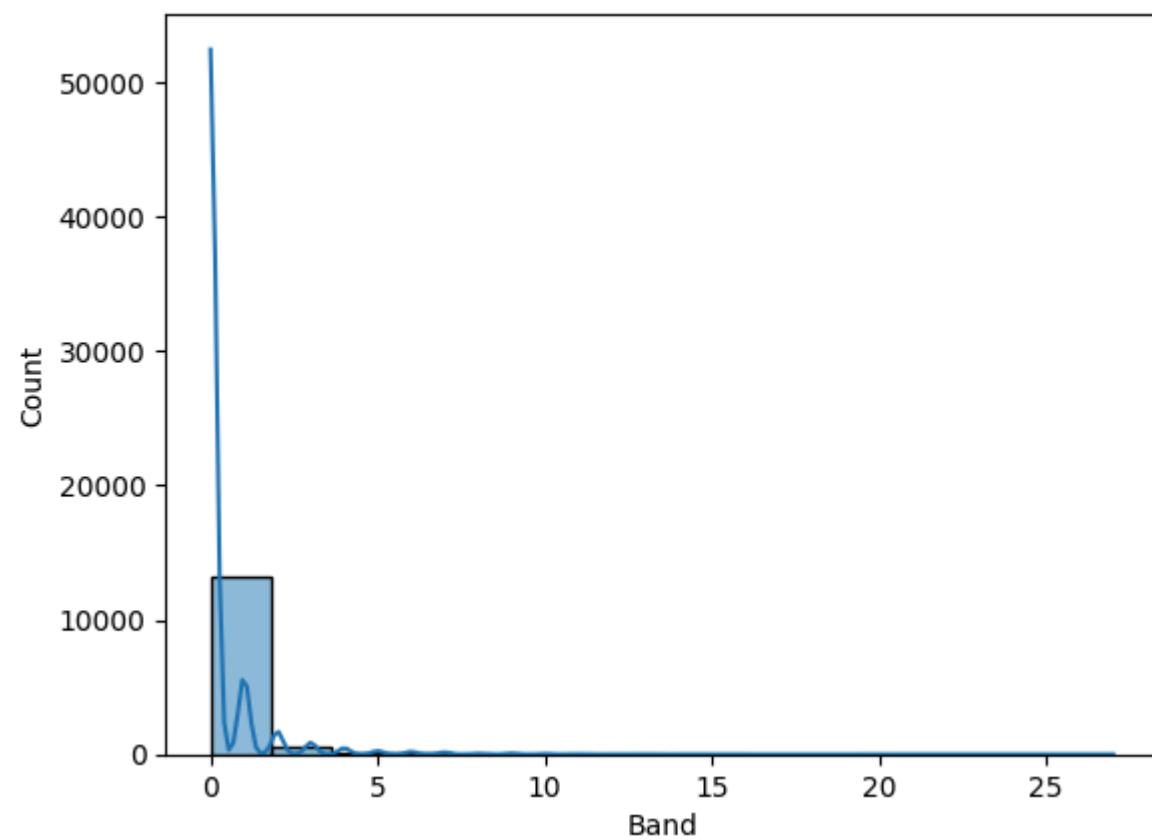




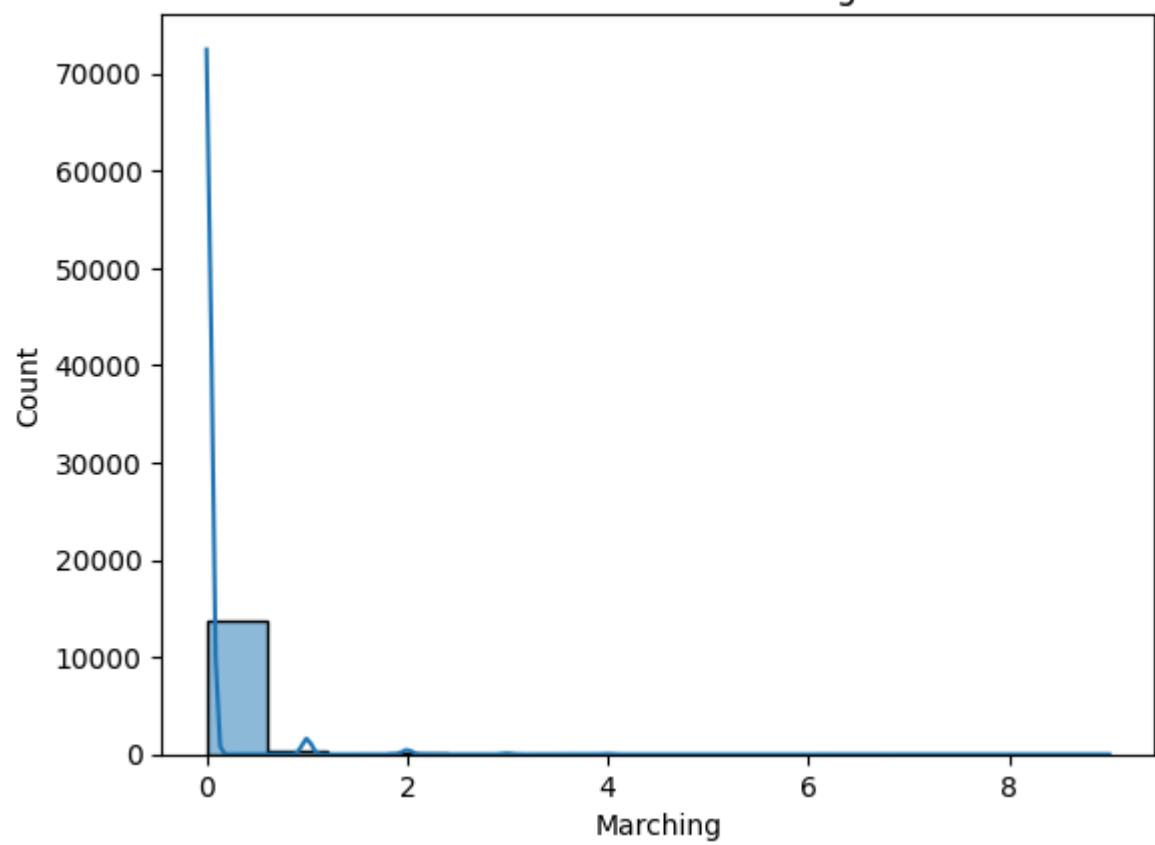
Distribution of Dance



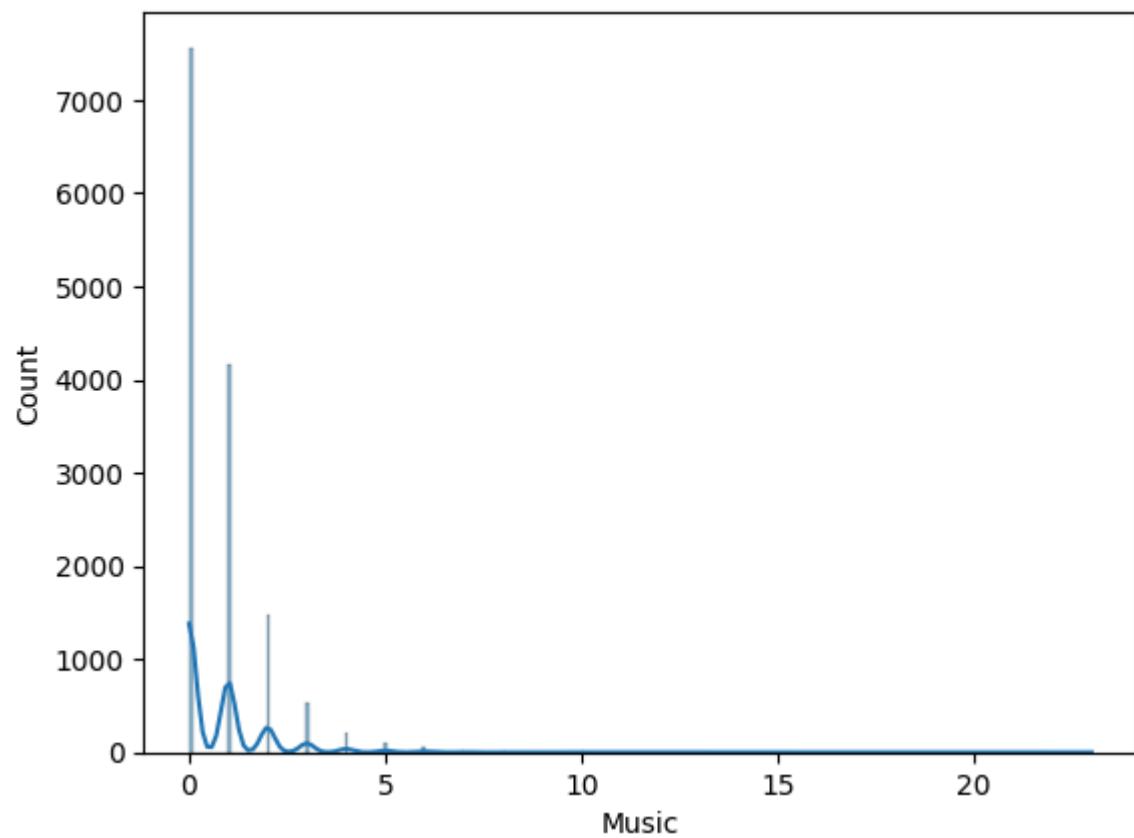
Distribution of Band



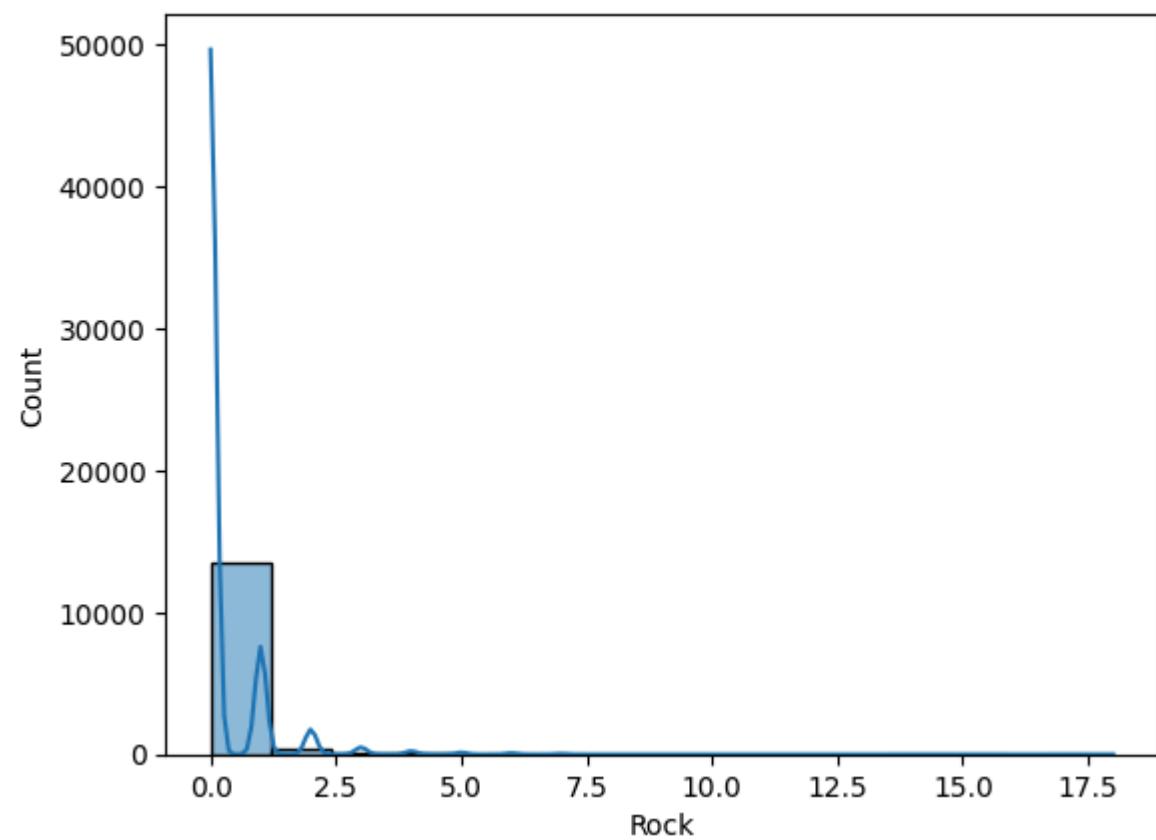
Distribution of Marching



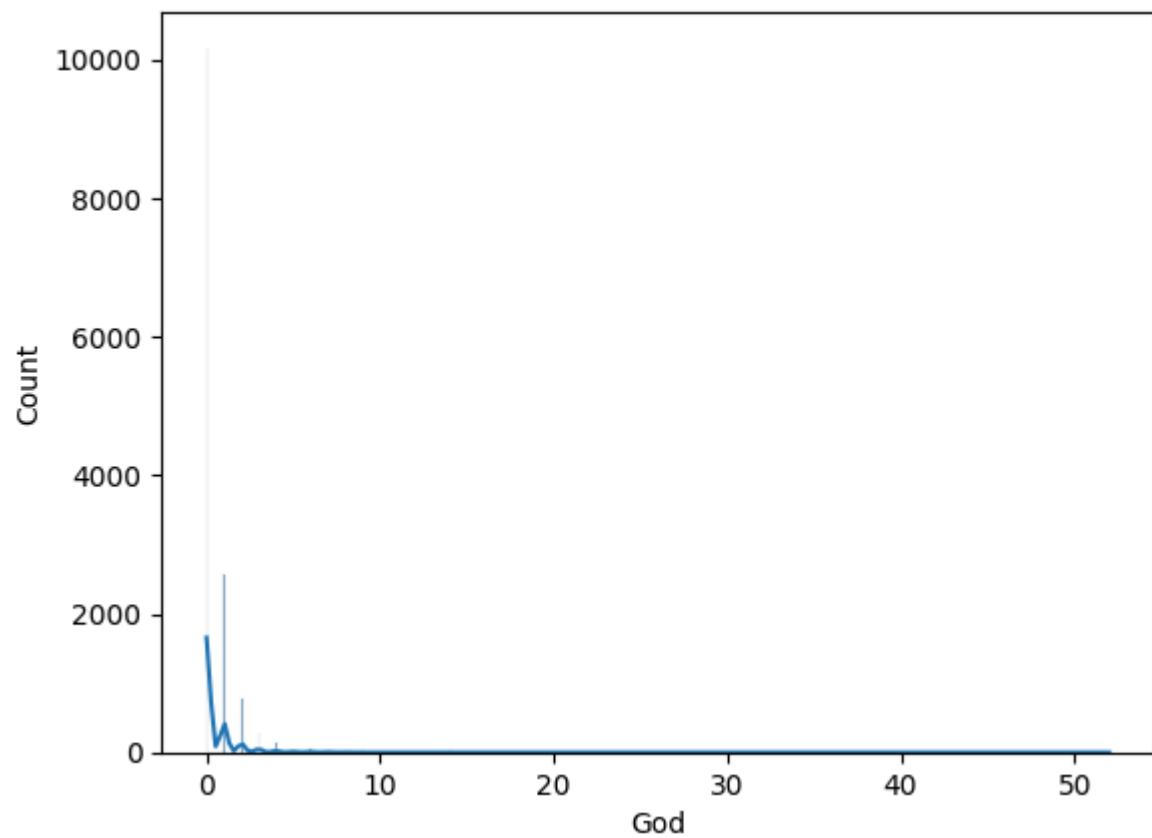
Distribution of Music



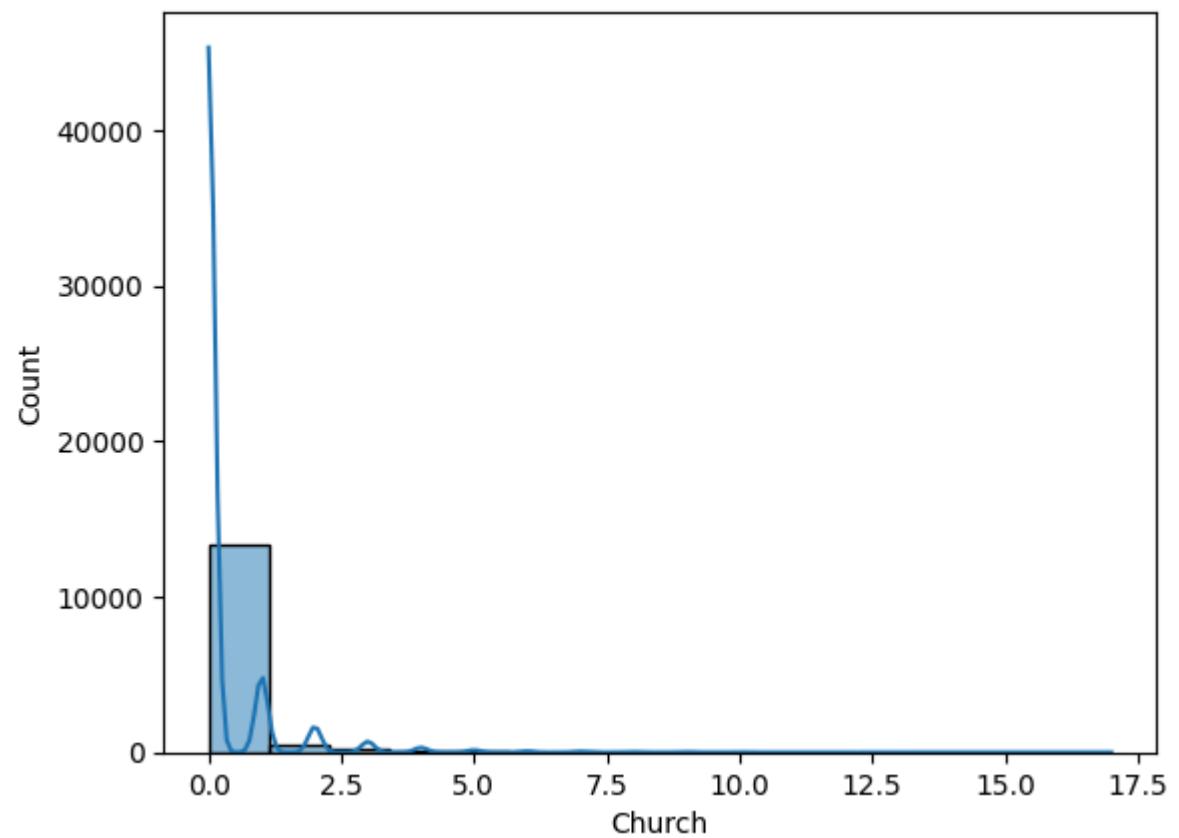
Distribution of Rock



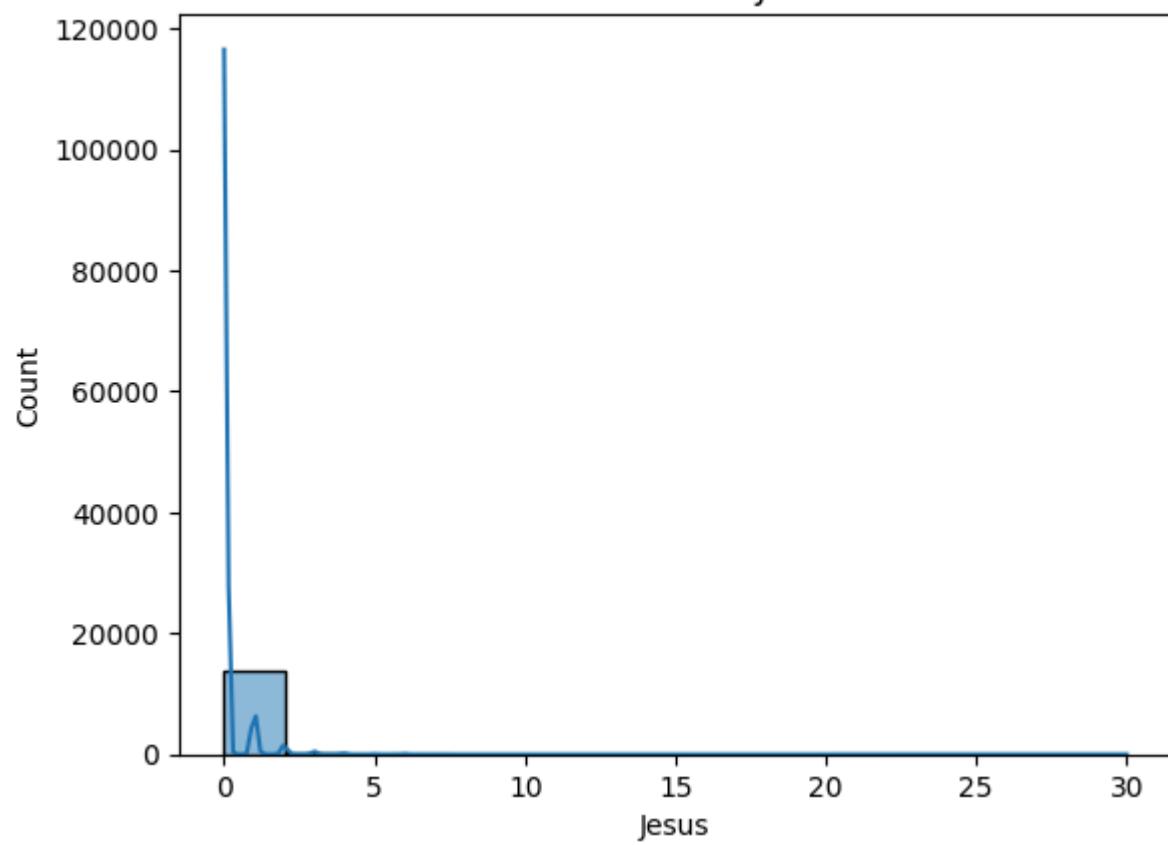
Distribution of God



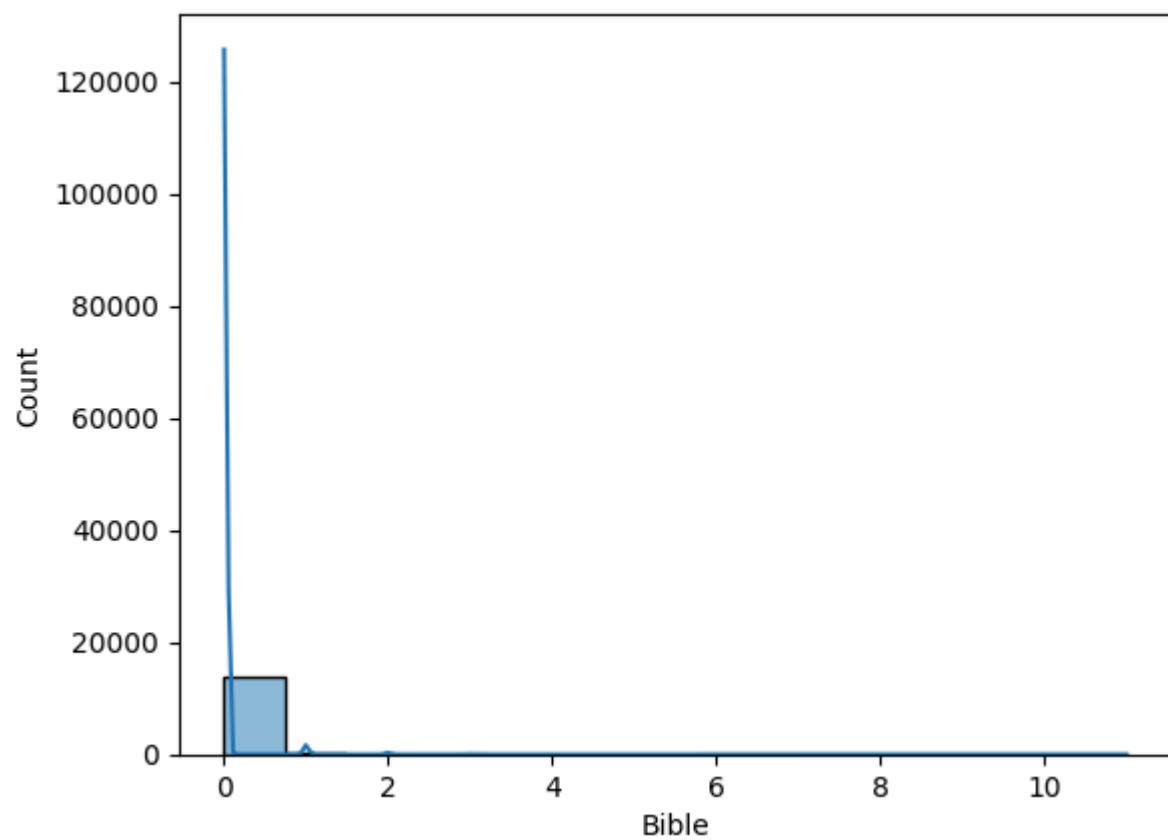
Distribution of Church



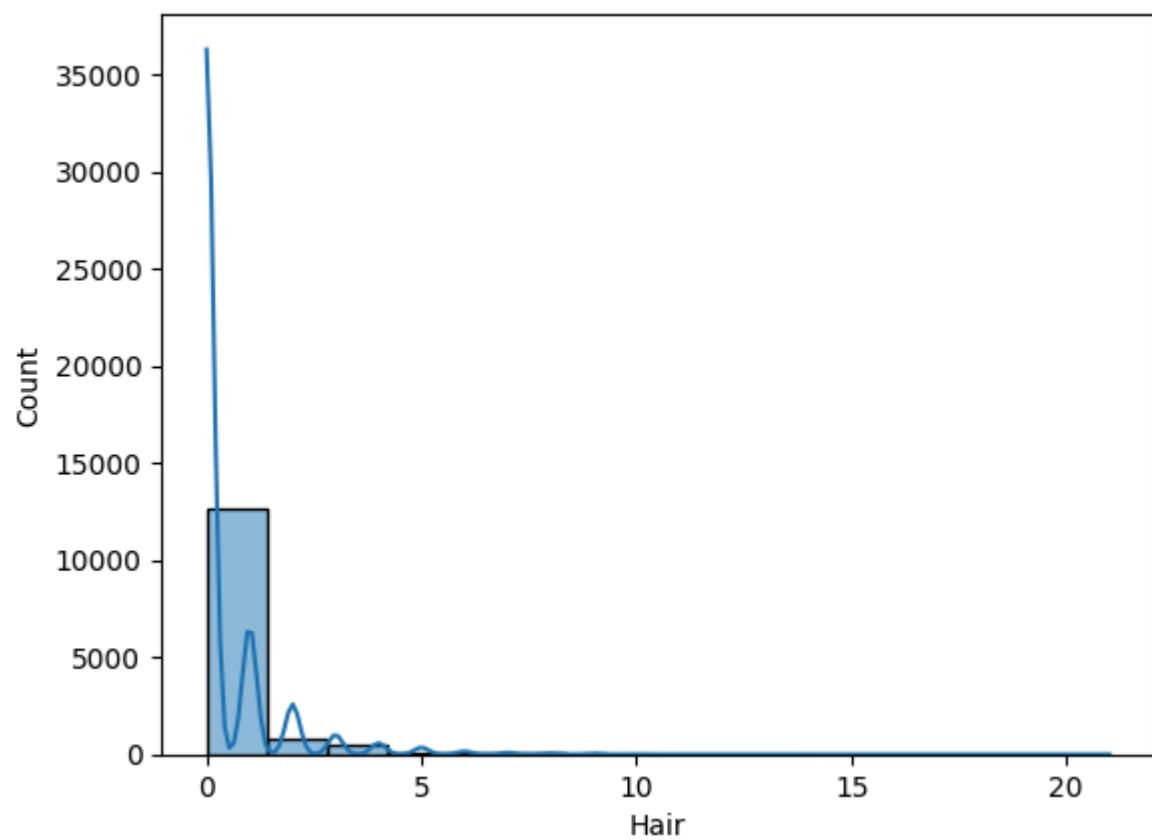
Distribution of Jesus



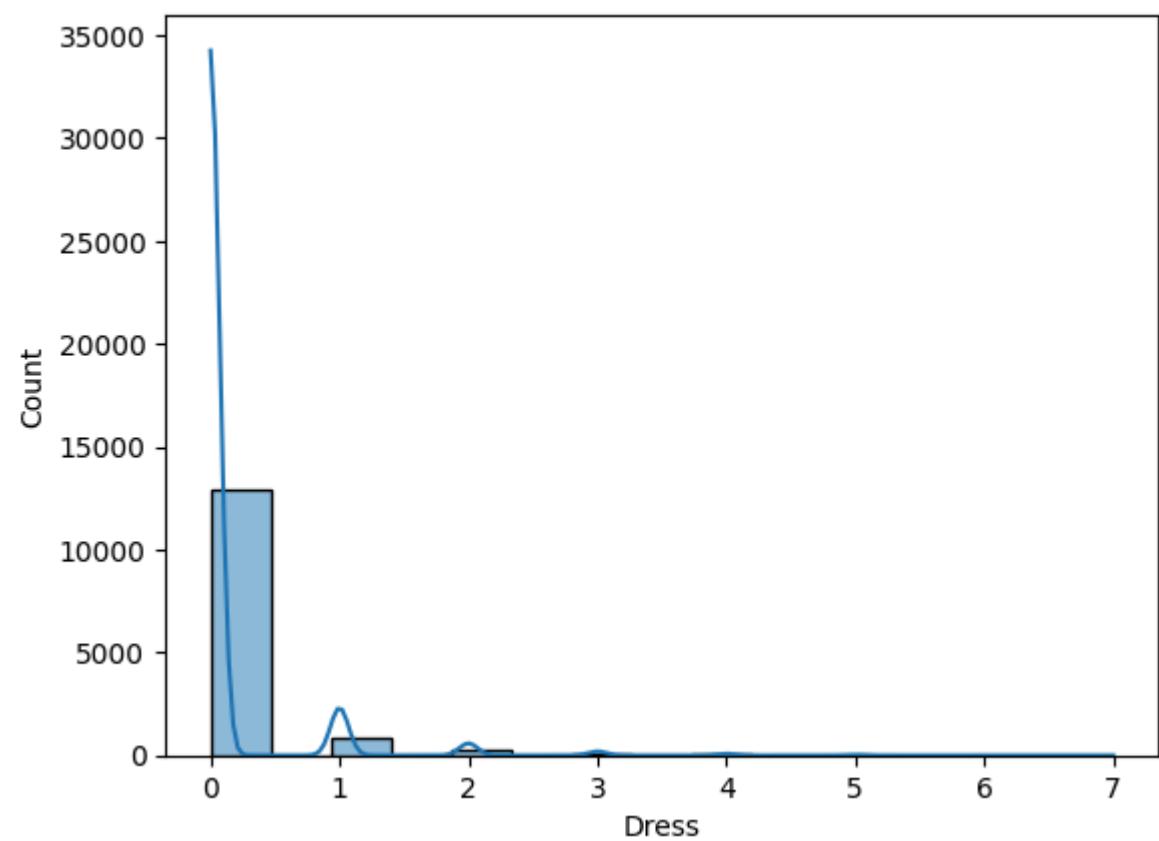
Distribution of Bible



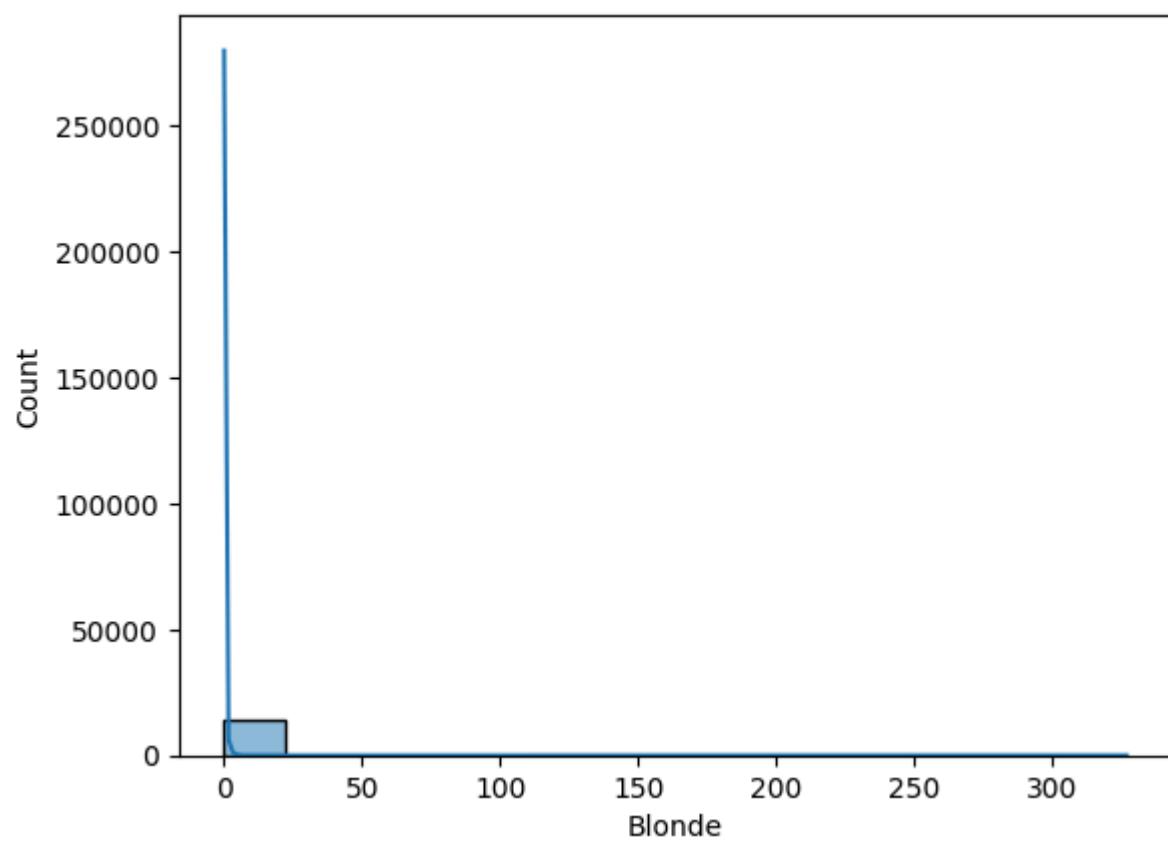
Distribution of Hair



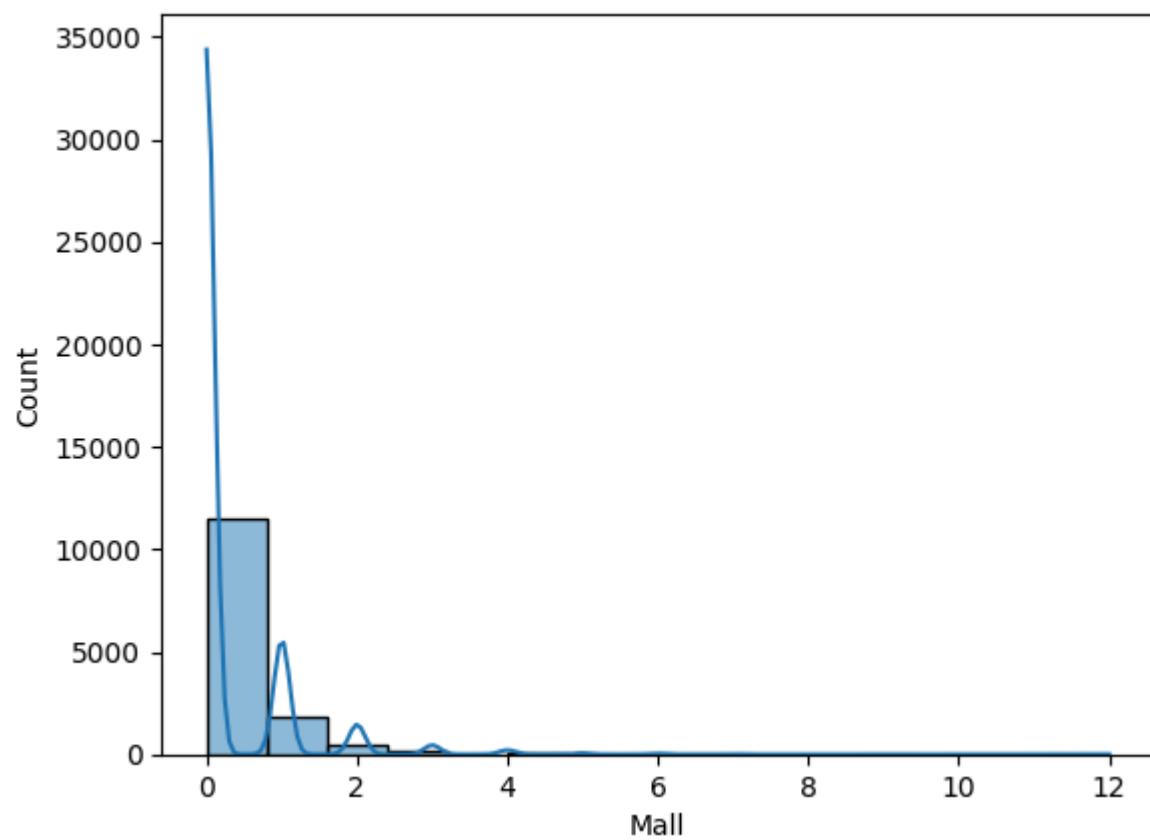
Distribution of Dress



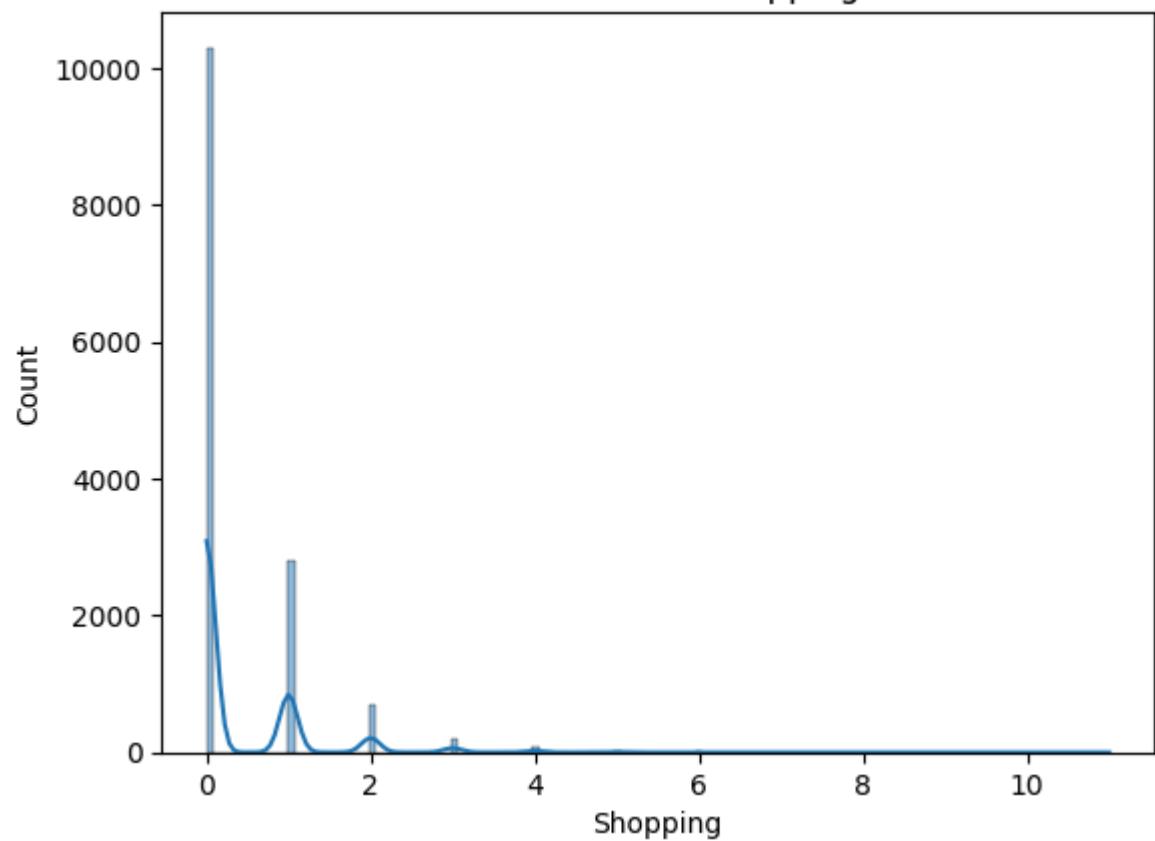
Distribution of Blonde



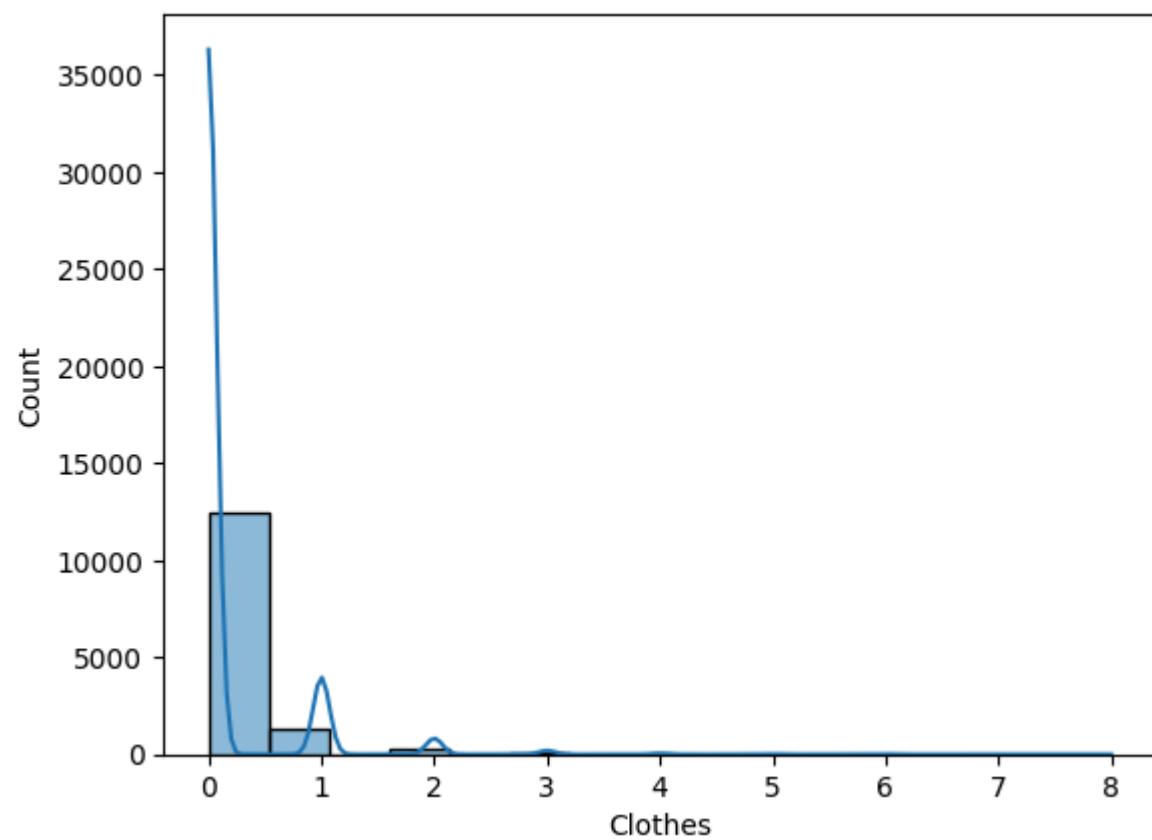
Distribution of Mall



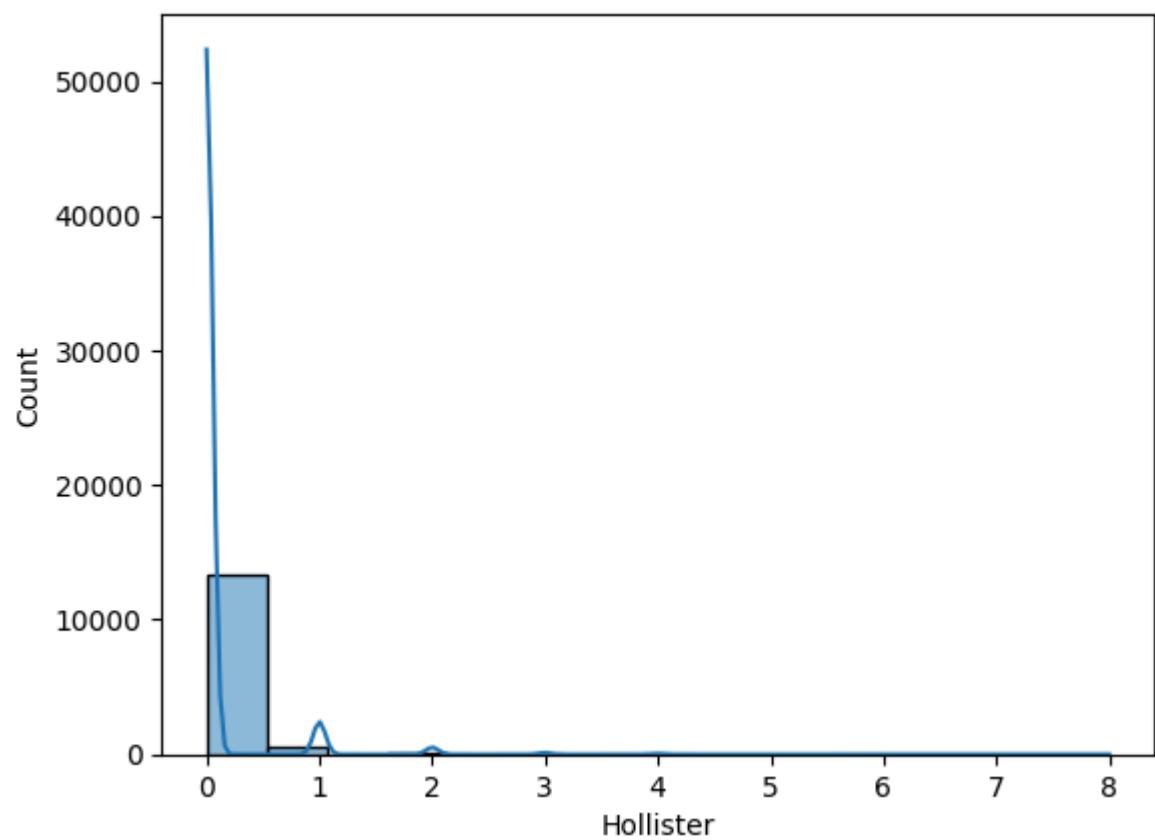
Distribution of Shopping



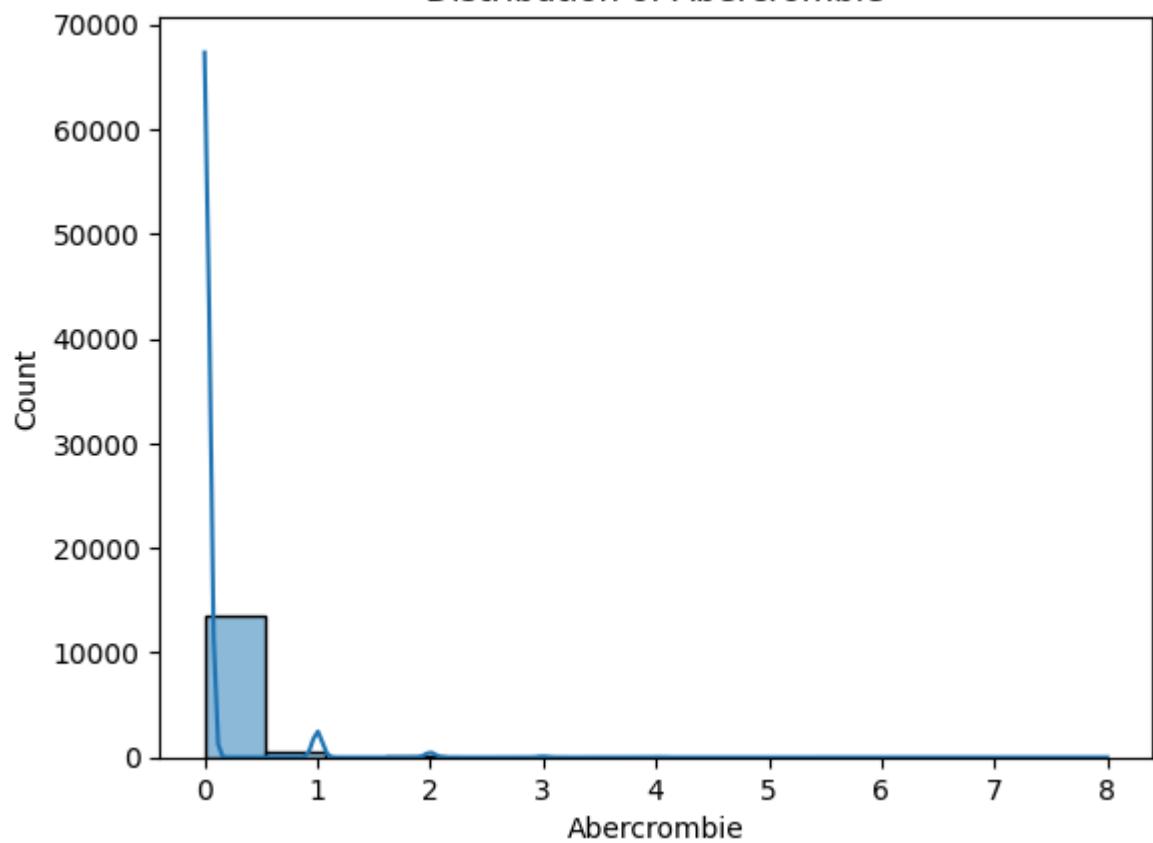
Distribution of Clothes



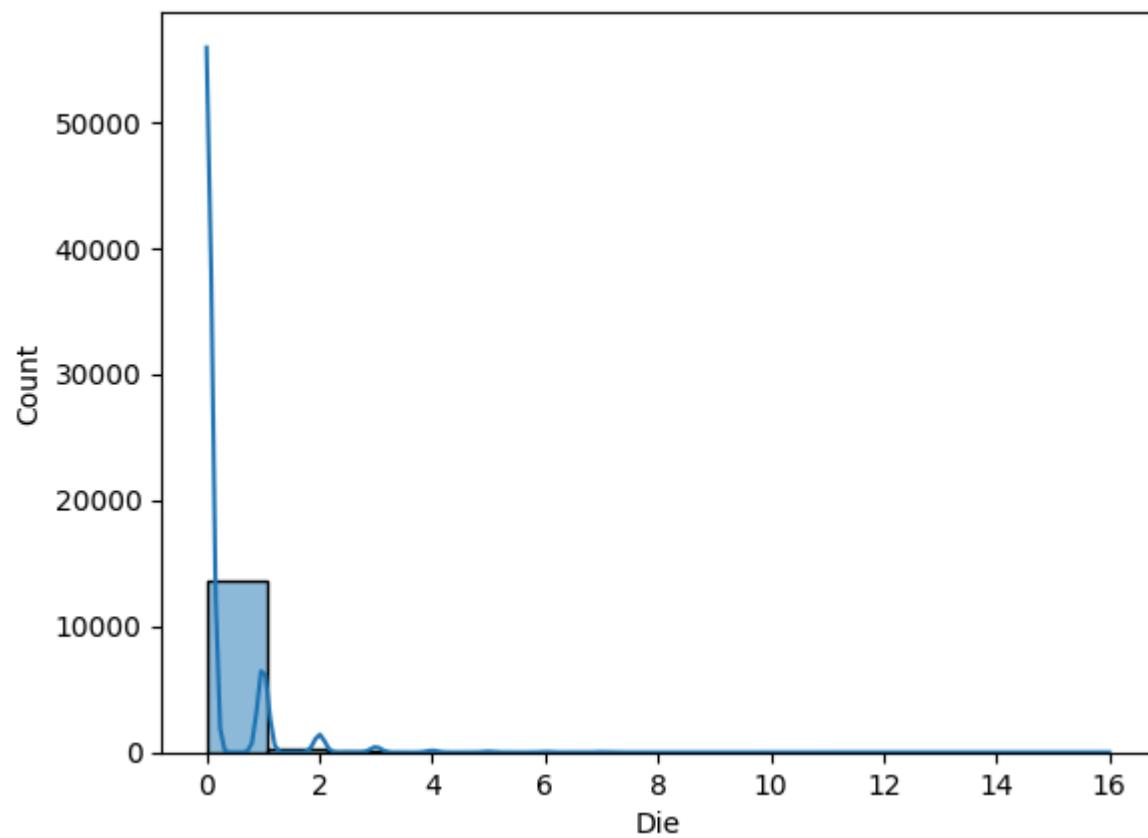
Distribution of Hollister



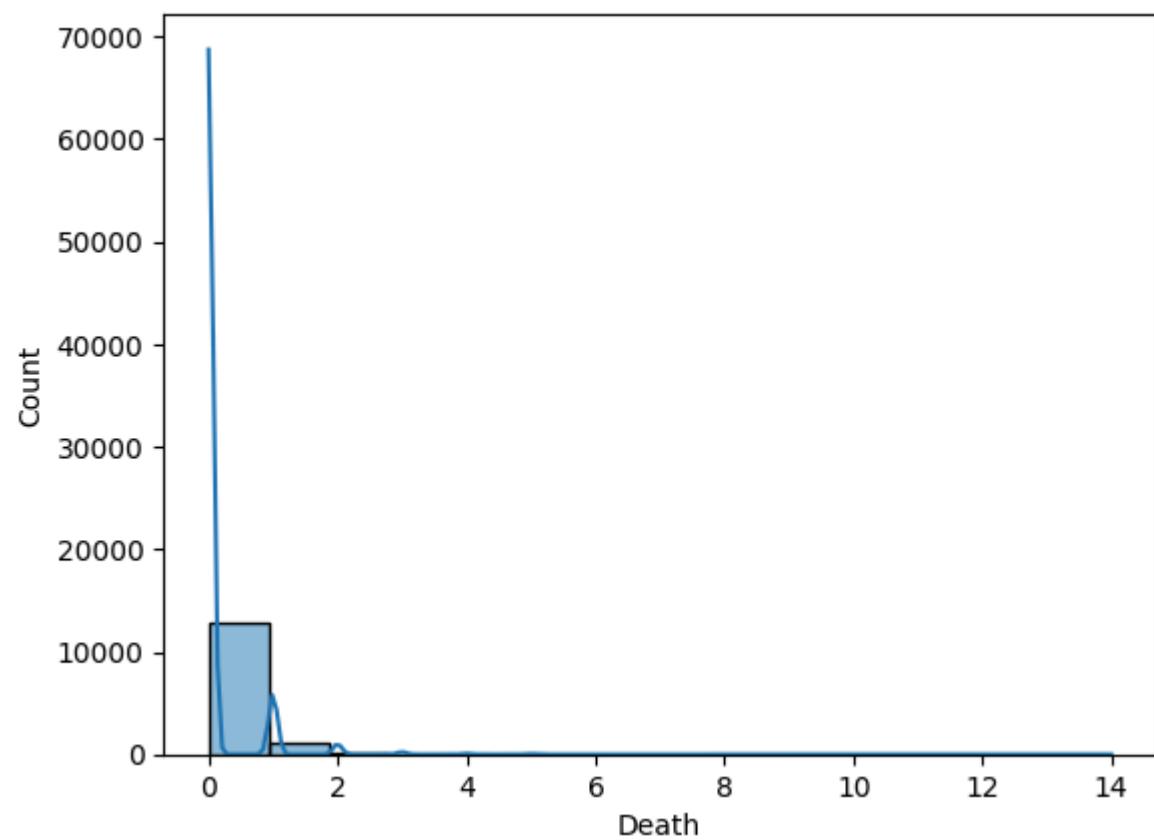
Distribution of Abercrombie



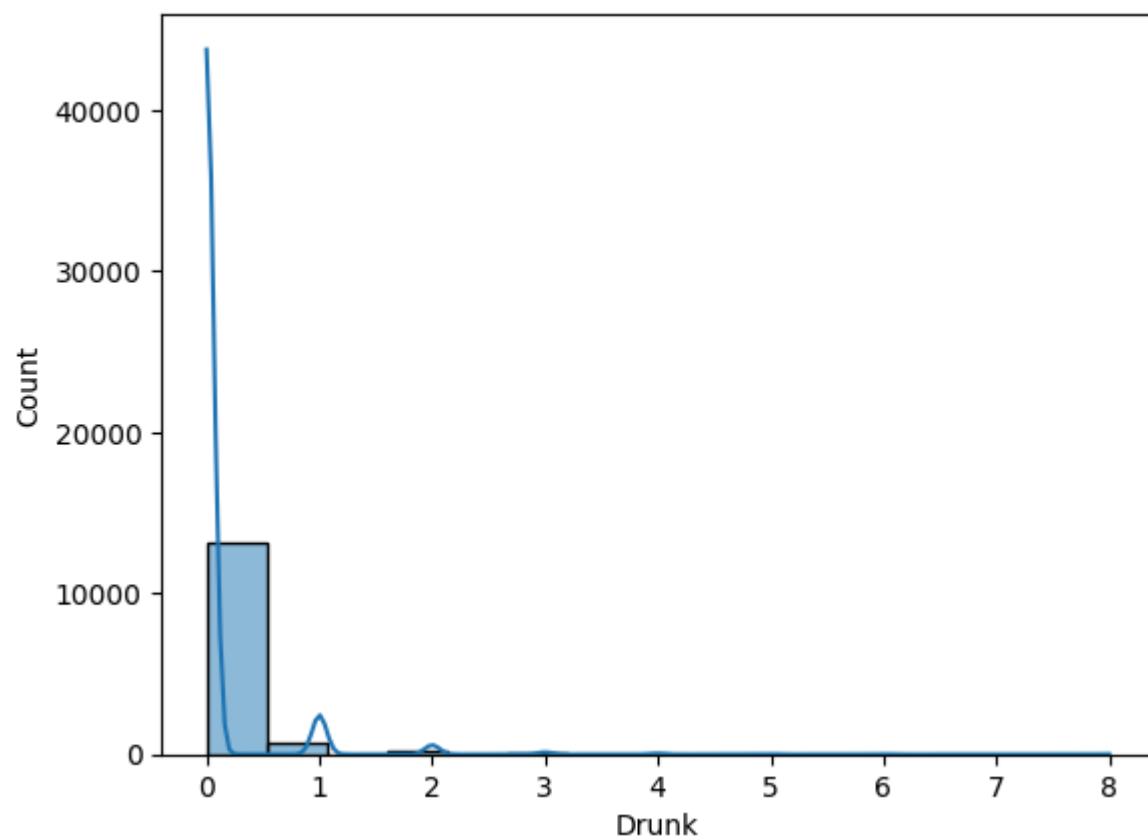
Distribution of Die



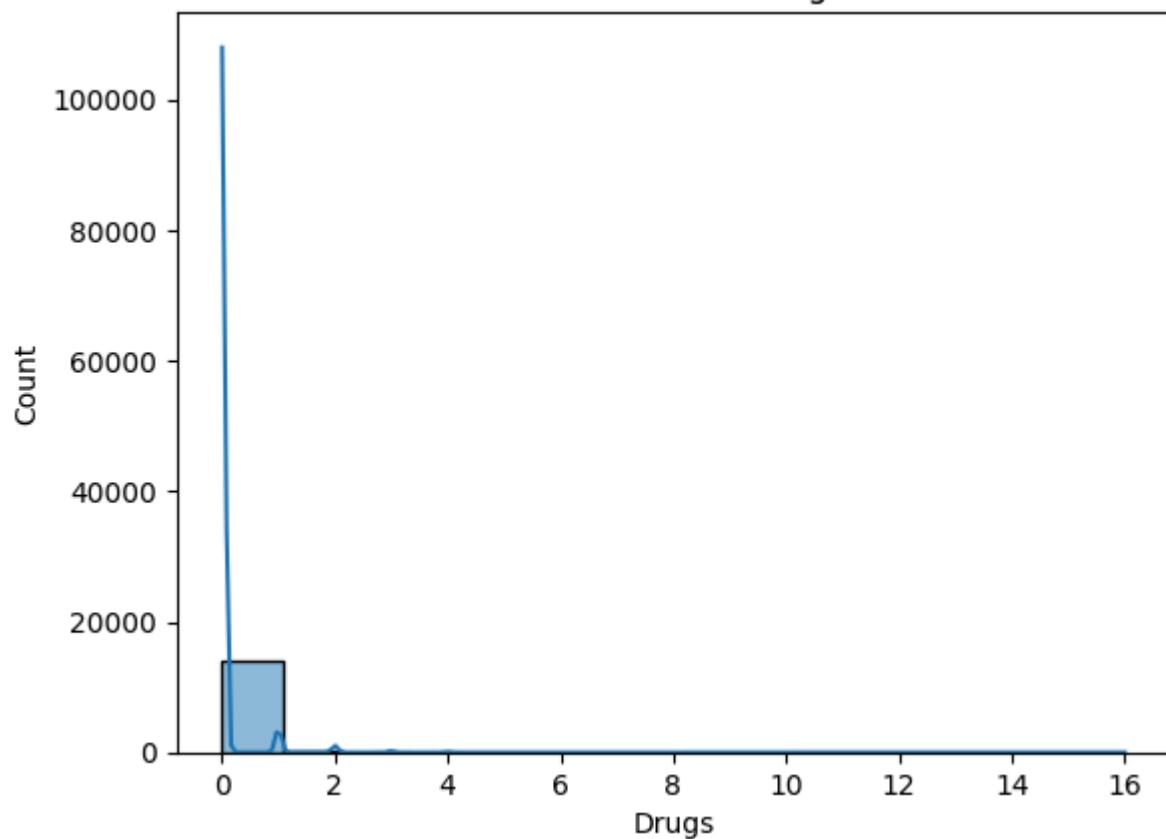
Distribution of Death



Distribution of Drunk



Distribution of Drugs



```
In [55]: skewness = df1.skew()
skewness
def Level_of_skewness(s):
    if s > 1 or s < -1:
        return 'Highly Skewed'
    elif s > 0.5 or s<-0.5:
        return 'Moderately Skewed'
    else:
        return 'Lightly Skewed'

skewness_category = skewness.apply(Level_of_skewness)
print(skewness_category)
```

```
Gradyear      Lightly Skewed
Gender        Highly Skewed
Age          Highly Skewed
NumberOfffriends  Highly Skewed
Basketball    Highly Skewed
Football      Highly Skewed
Soccer        Highly Skewed
Softball      Highly Skewed
Volleyball    Highly Skewed
Swimming      Highly Skewed
Cheerleading   Highly Skewed
Baseball      Highly Skewed
Tennis         Highly Skewed
Sports         Highly Skewed
Cute           Highly Skewed
Sex            Highly Skewed
Sexy           Highly Skewed
Hot            Highly Skewed
Kissed          Highly Skewed
Dance          Highly Skewed
Band           Highly Skewed
Marching       Highly Skewed
Music          Highly Skewed
Rock           Highly Skewed
God            Highly Skewed
Church          Highly Skewed
Jesus          Highly Skewed
Bible          Highly Skewed
Hair           Highly Skewed
Dress          Highly Skewed
Blonde         Highly Skewed
Mall           Highly Skewed
Shopping        Highly Skewed
Clothes         Highly Skewed
Hollister       Highly Skewed
Abercrombie    Highly Skewed
Die             Highly Skewed
Death           Highly Skewed
Drunk           Highly Skewed
Drugs           Highly Skewed
dtype: object
```

From above result we can observe that all of the features except 'Gradyear' are highly skewed as we know the data ranges are low as compared to 'Gradyear'

```
In [56]: df2=df1.copy()#Creating copy of data
```

In [57]: #Dealing with skewness

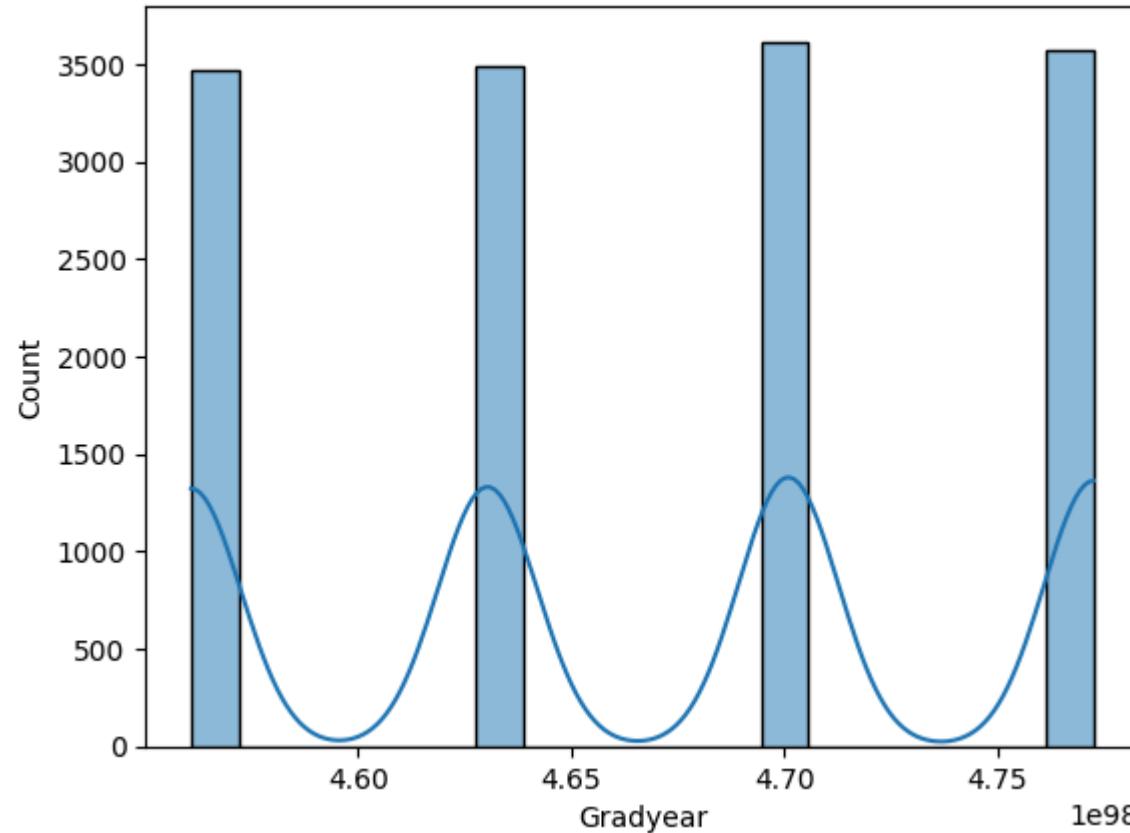
```
from scipy.stats import boxcox

for col in df2.select_dtypes(include=['float64', 'int64']).columns:
    # Check if all values are positive (required for Box-Cox)
    if (df2[col] > 0).all():
        df2[col], _ = boxcox(df2[col])
    else:
        # If negative values present, consider Yeo-Johnson
        from scipy.stats import yeojohnson
        df2[col], _ = yeojohnson(df2[col])

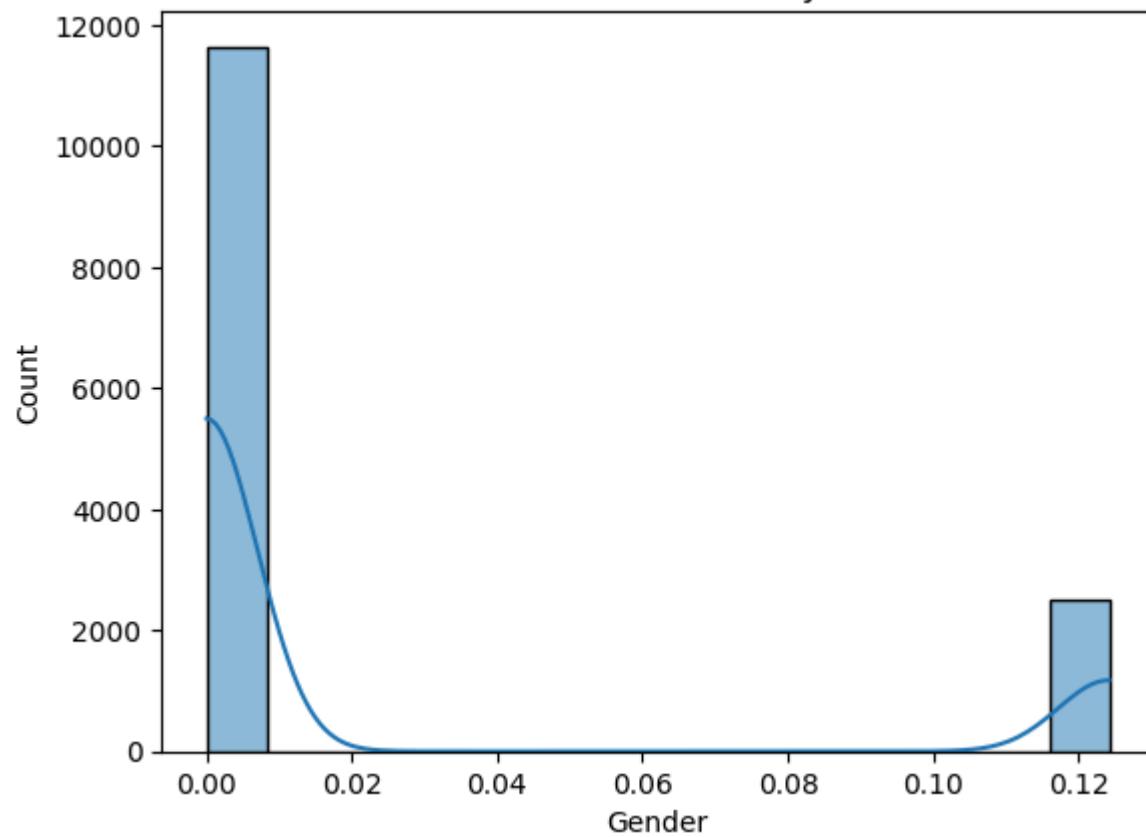
sns.histplot(df2[col], kde=True)
plt.title(f'Distribution of {col} after Box-Cox/Yeo-Johnson transformation')
plt.show()

print(df2.skew())
```

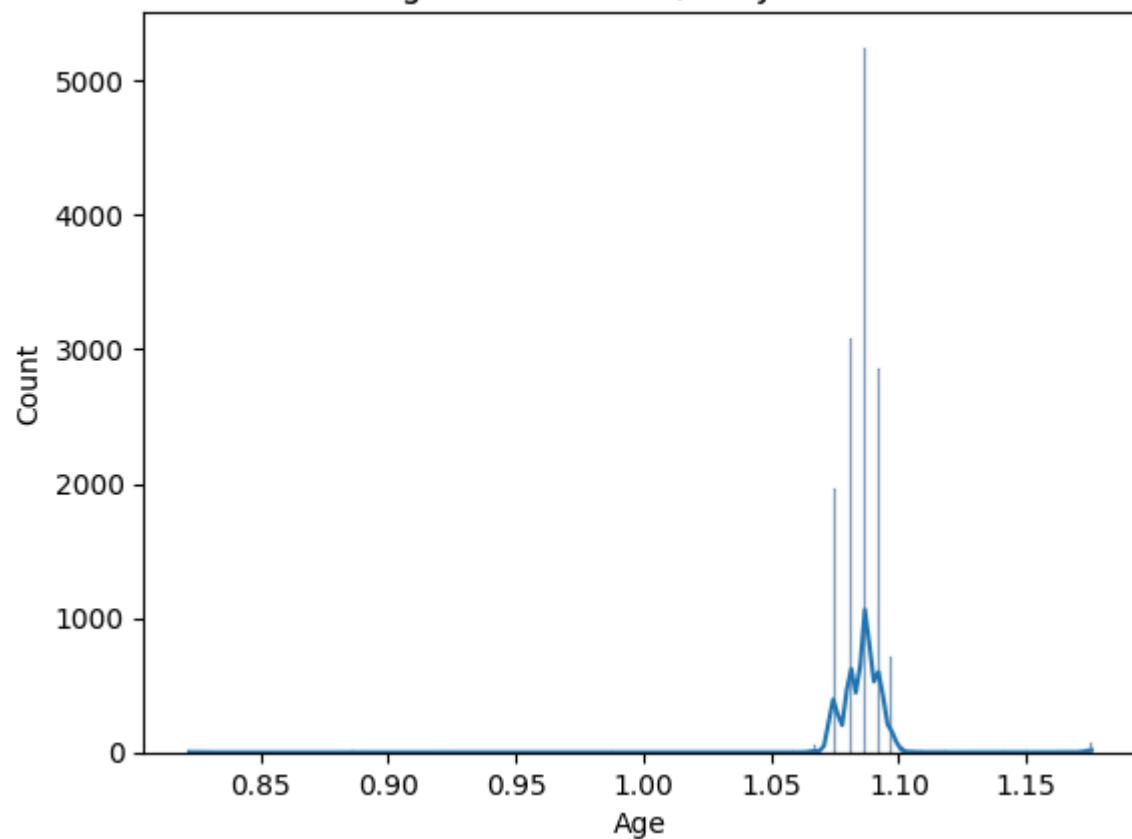
Distribution of Gradyear after Box-Cox/Yeo-Johnson transformation



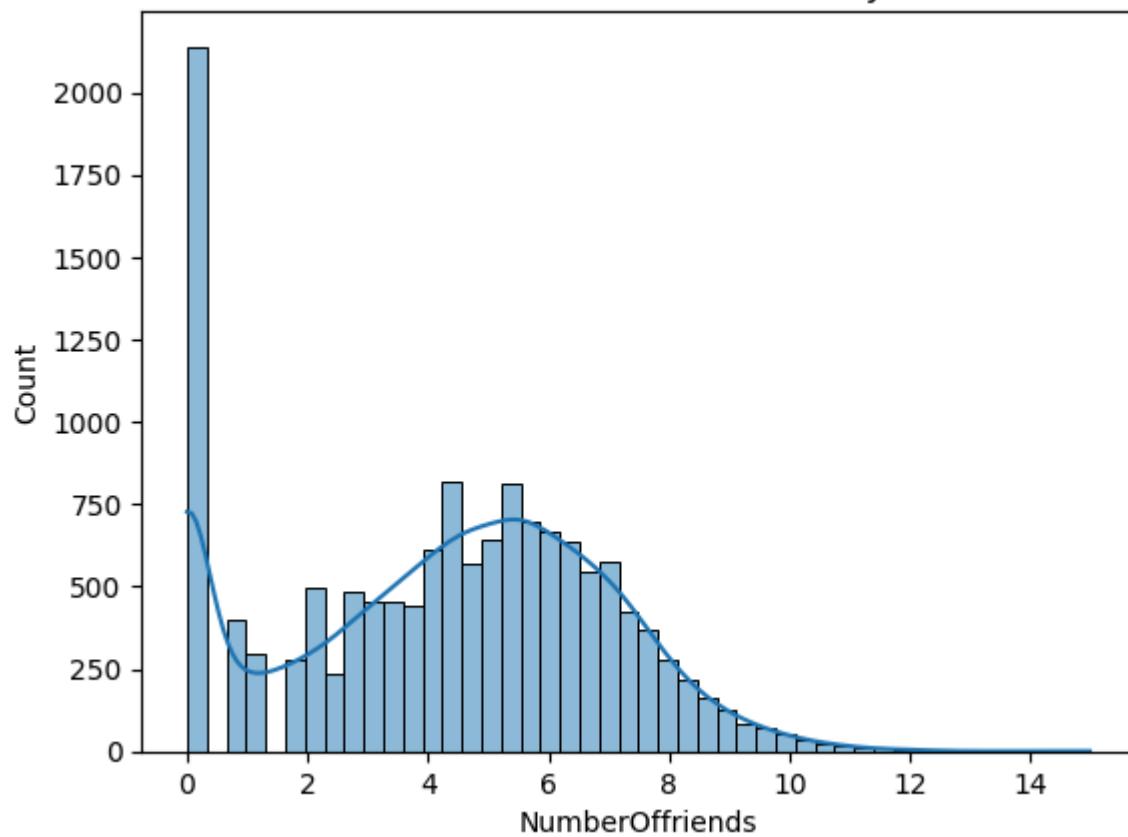
Distribution of Gender after Box-Cox/Yeo-Johnson transformation



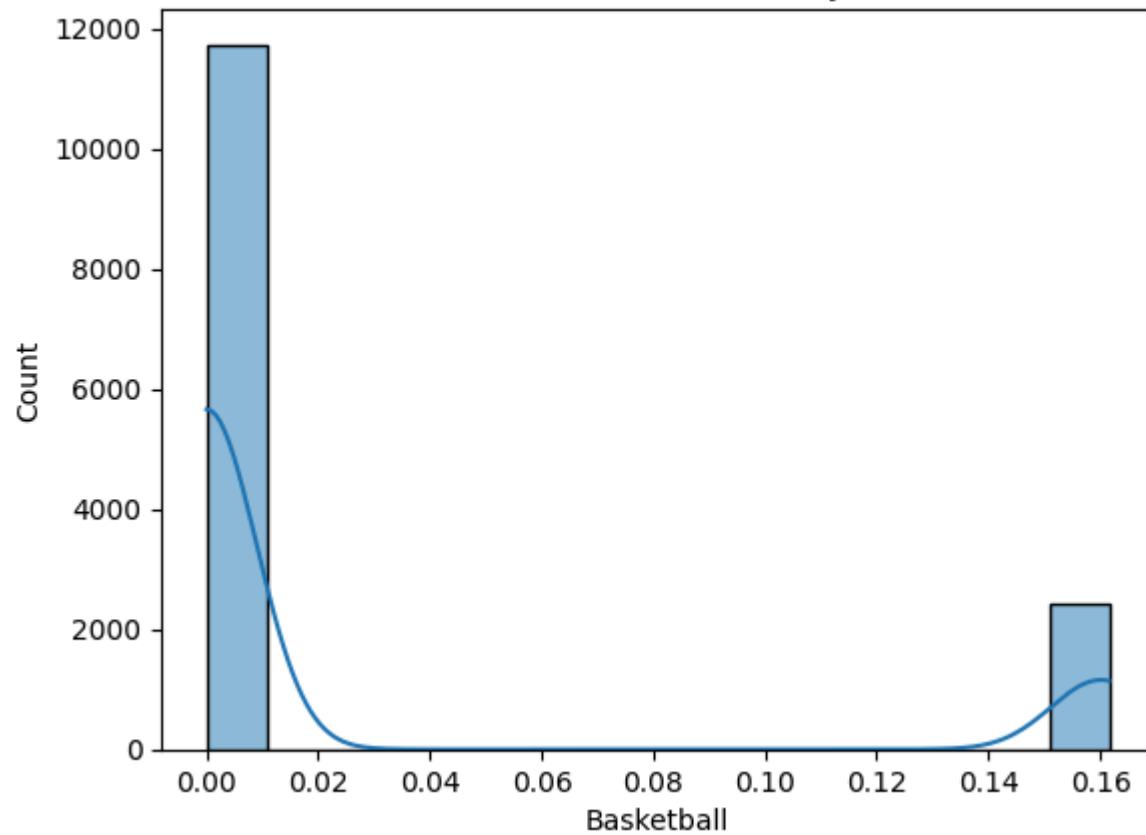
Distribution of Age after Box-Cox/Yeo-Johnson transformation

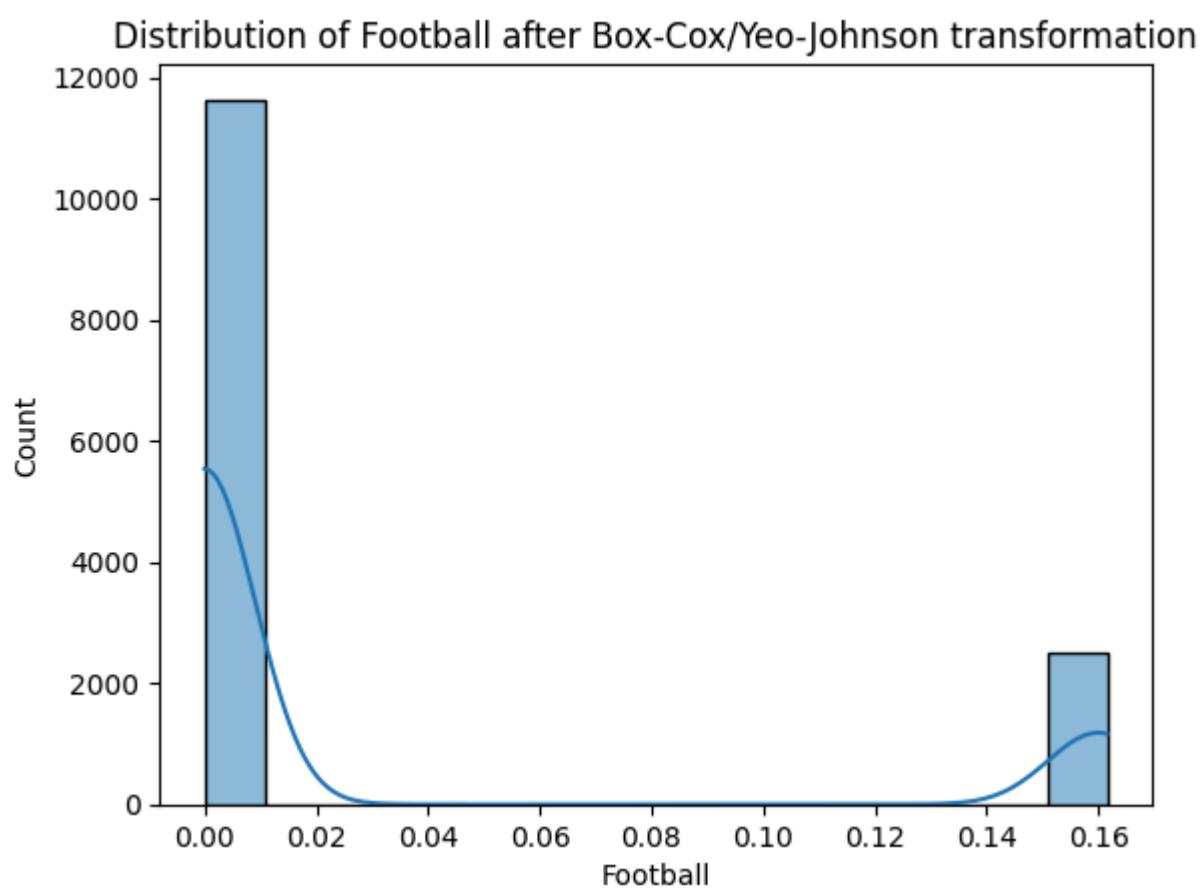


Distribution of NumberOffriends after Box-Cox/Yeo-Johnson transformation

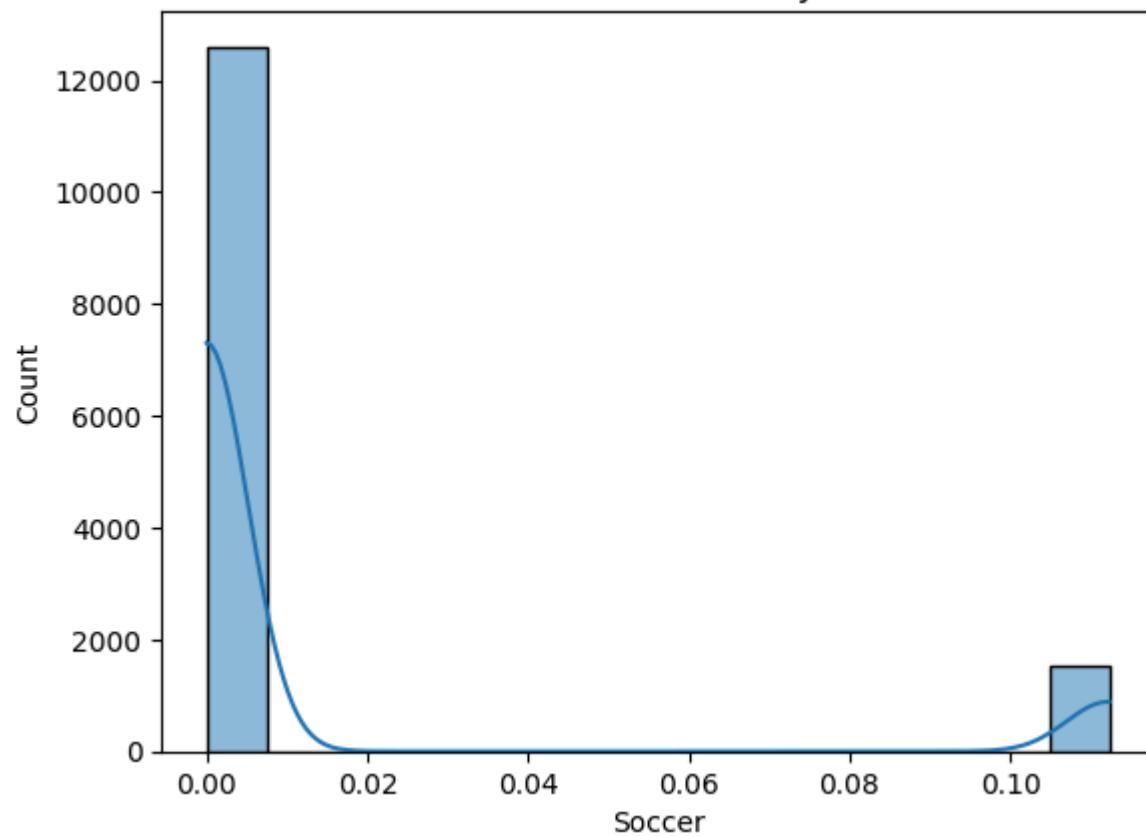


Distribution of Basketball after Box-Cox/Yeo-Johnson transformation

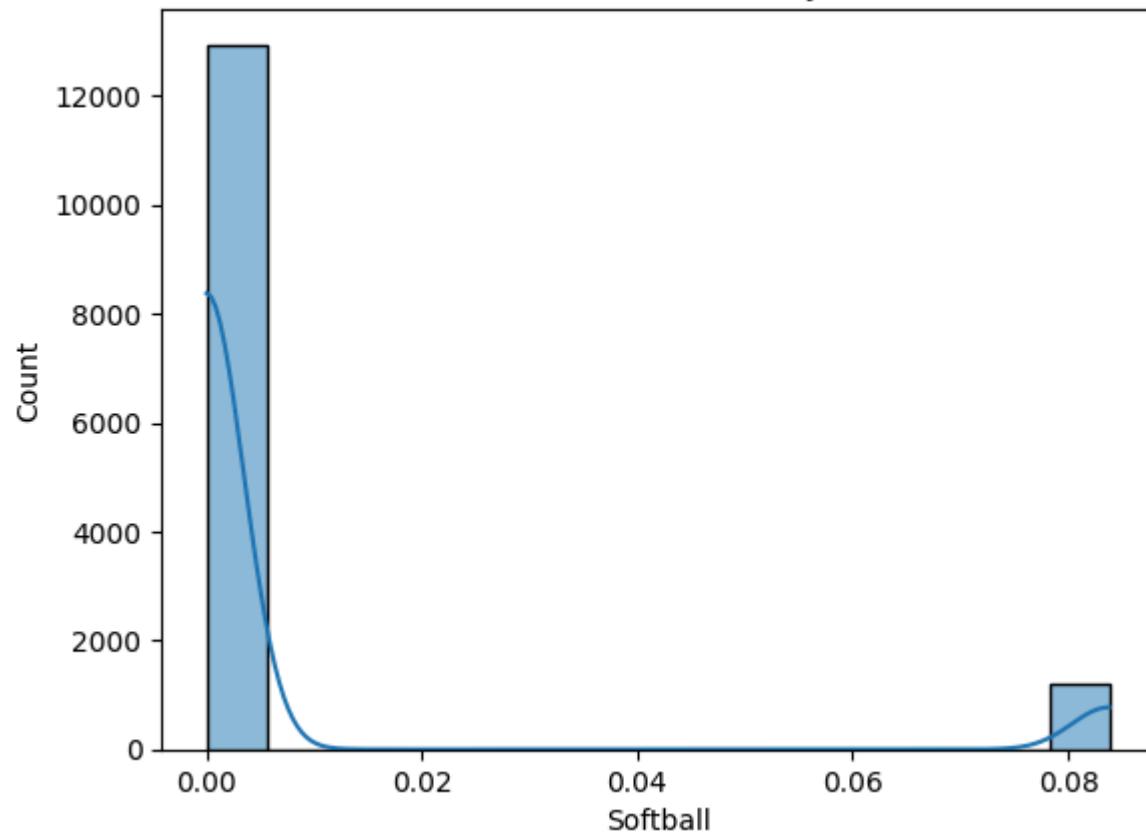




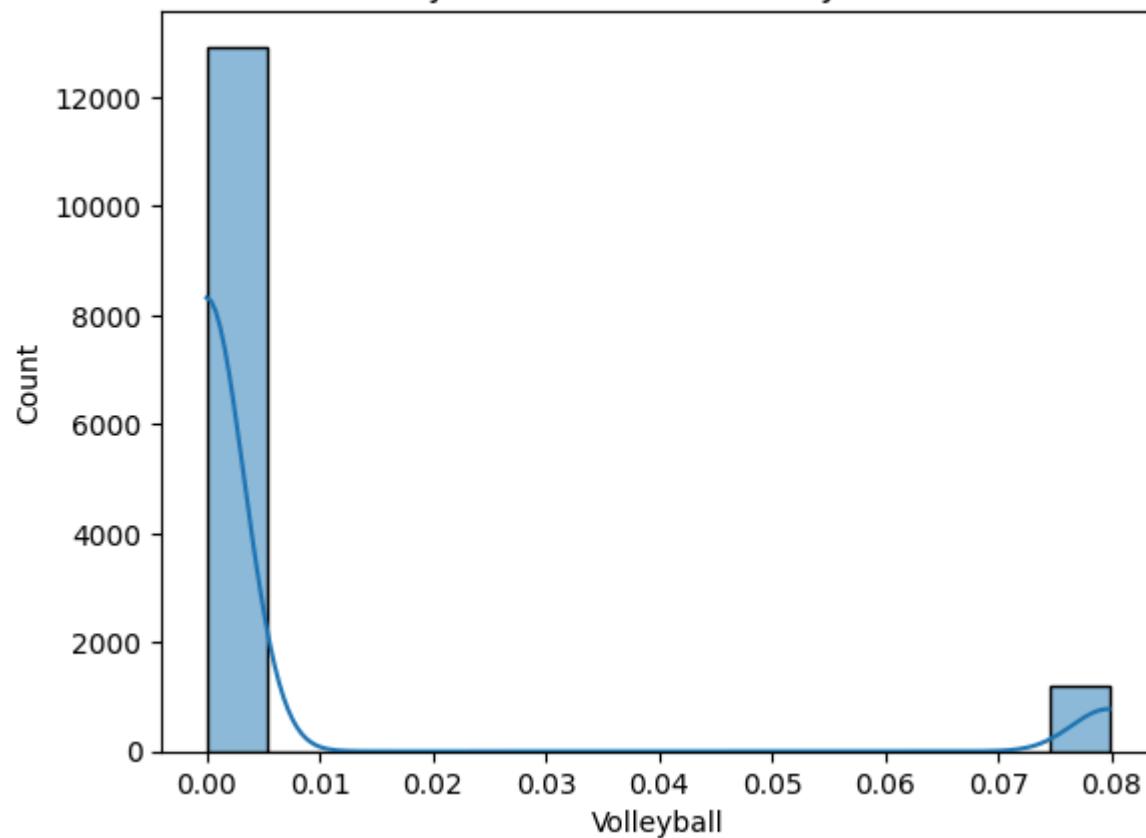
Distribution of Soccer after Box-Cox/Yeo-Johnson transformation



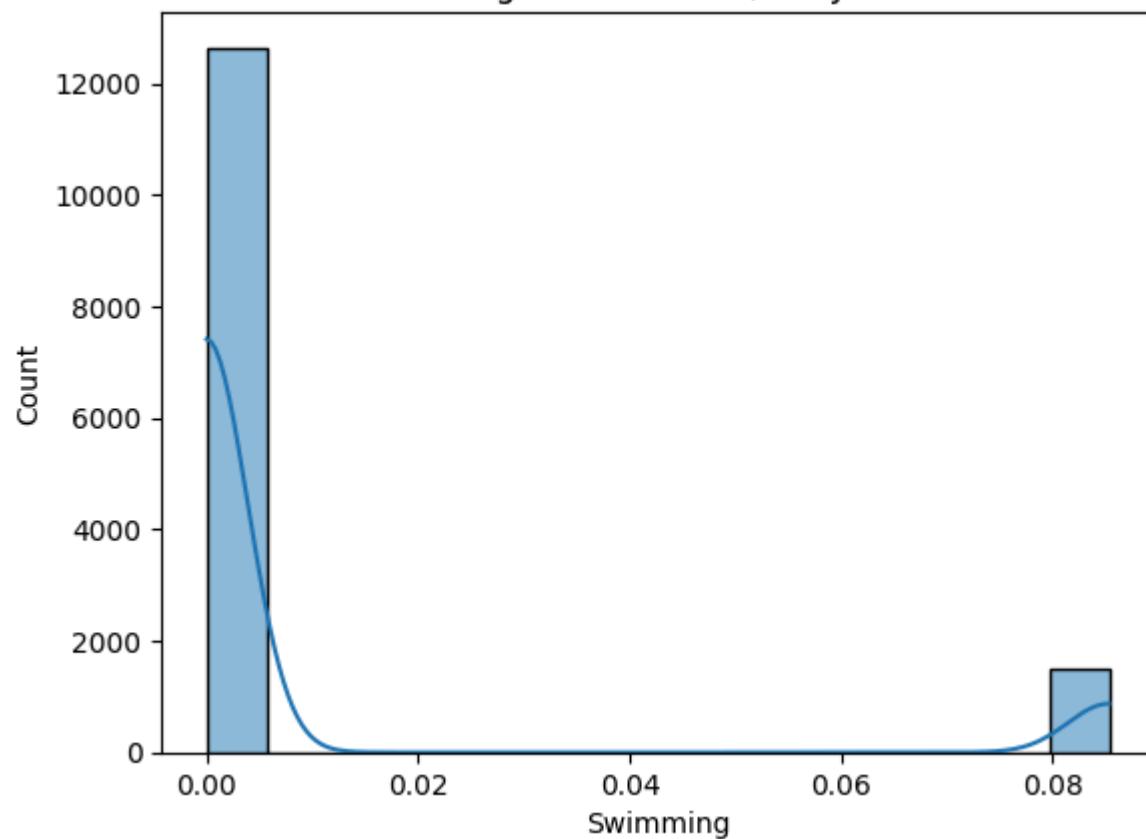
Distribution of Softball after Box-Cox/Yeo-Johnson transformation



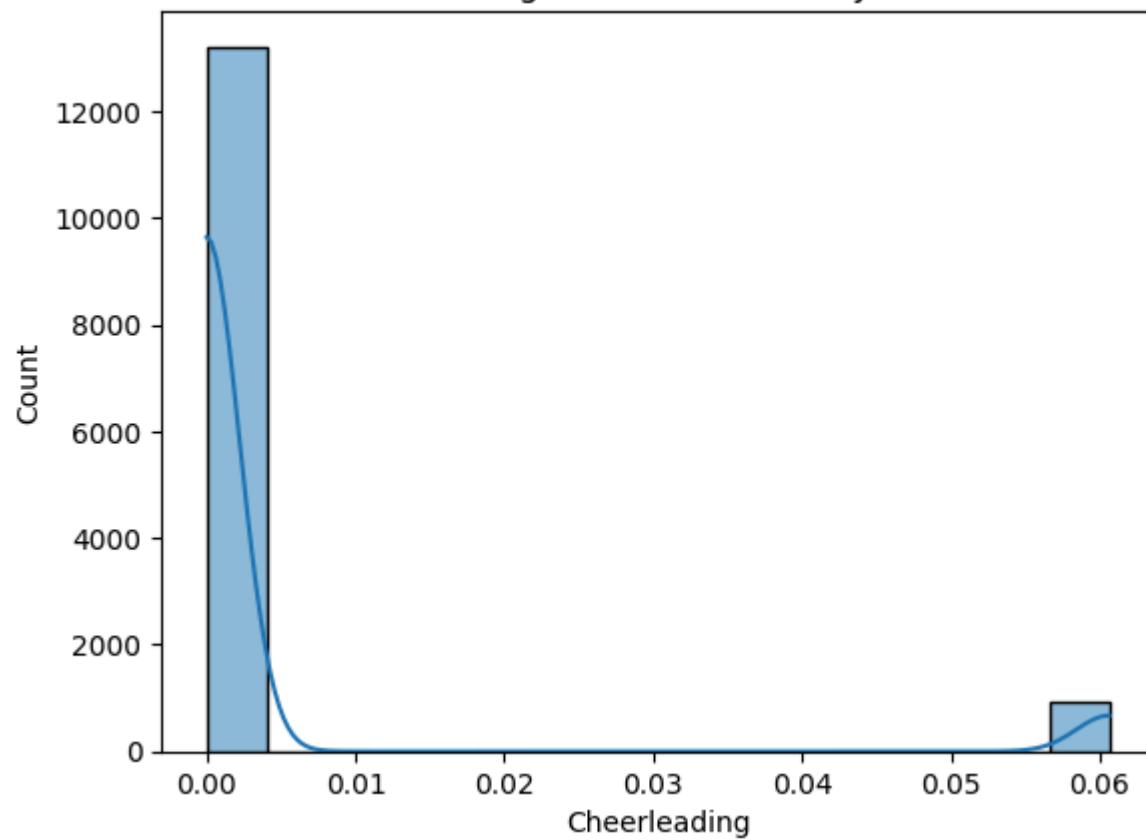
Distribution of Volleyball after Box-Cox/Yeo-Johnson transformation



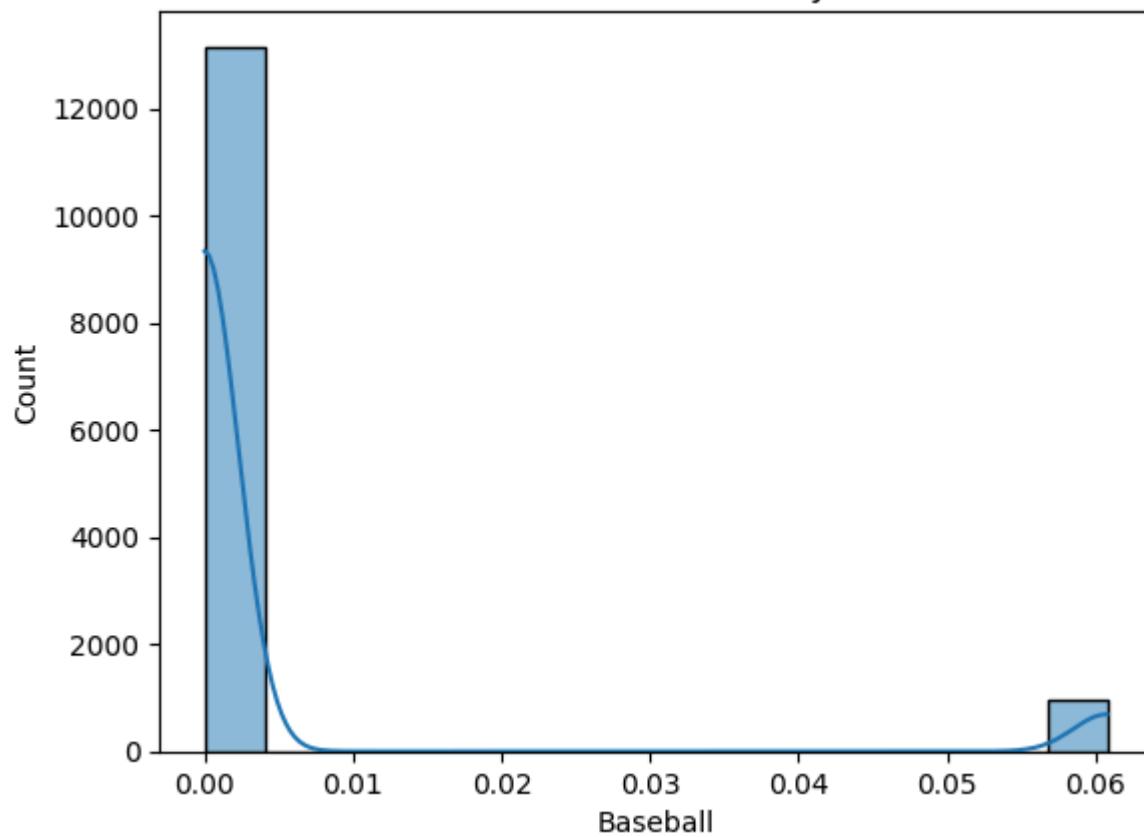
Distribution of Swimming after Box-Cox/Yeo-Johnson transformation

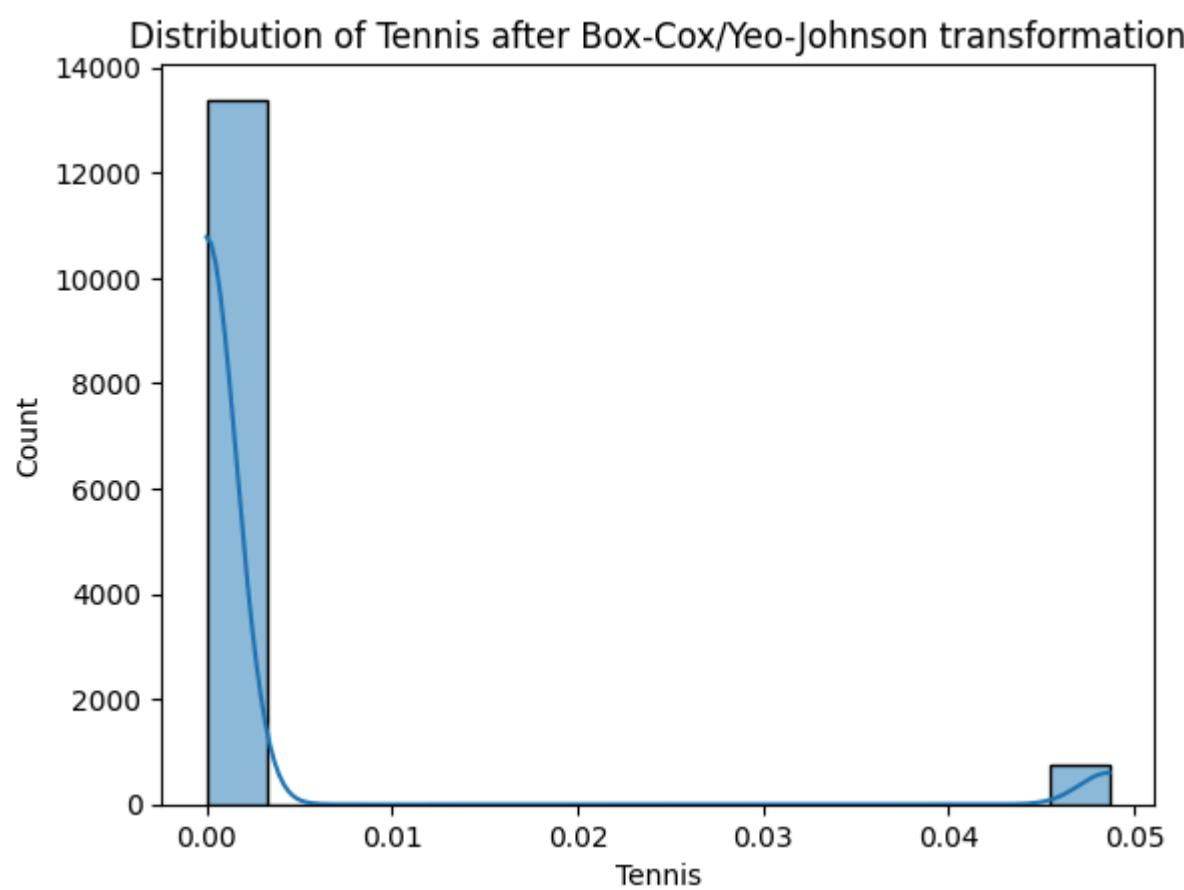


Distribution of Cheerleading after Box-Cox/Yeo-Johnson transformation

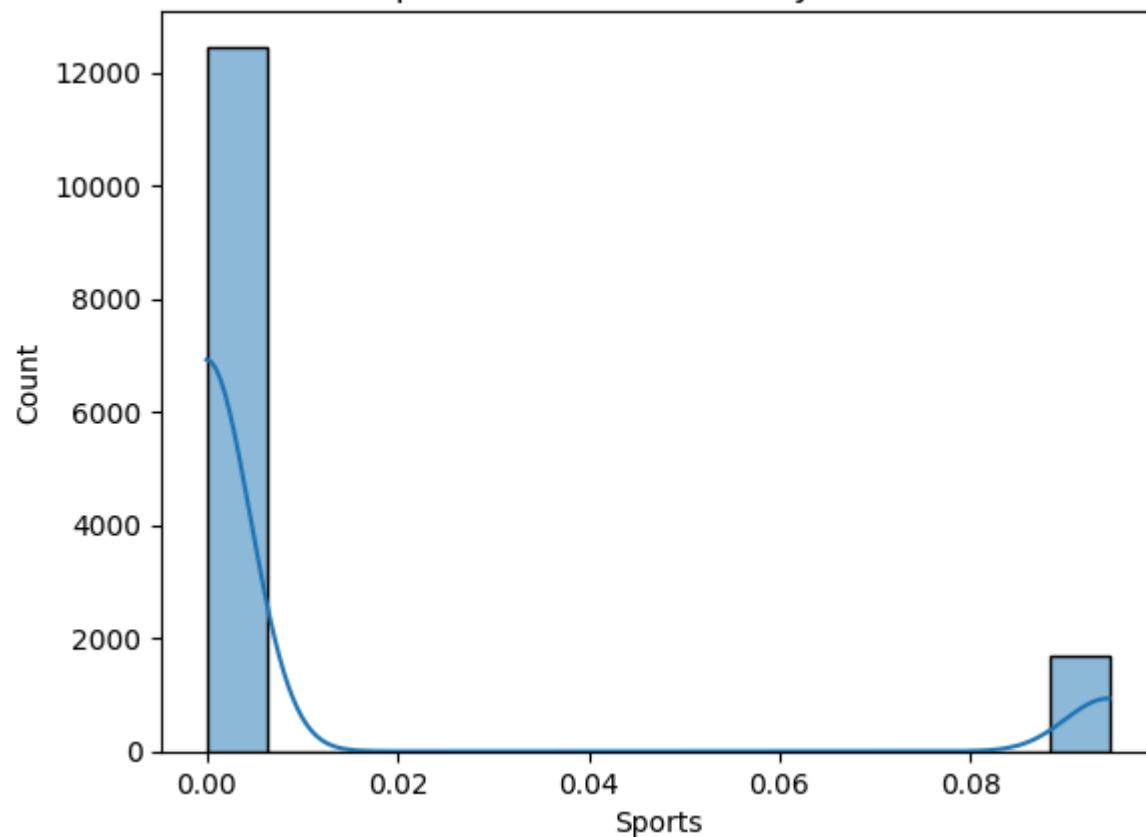


Distribution of Baseball after Box-Cox/Yeo-Johnson transformation

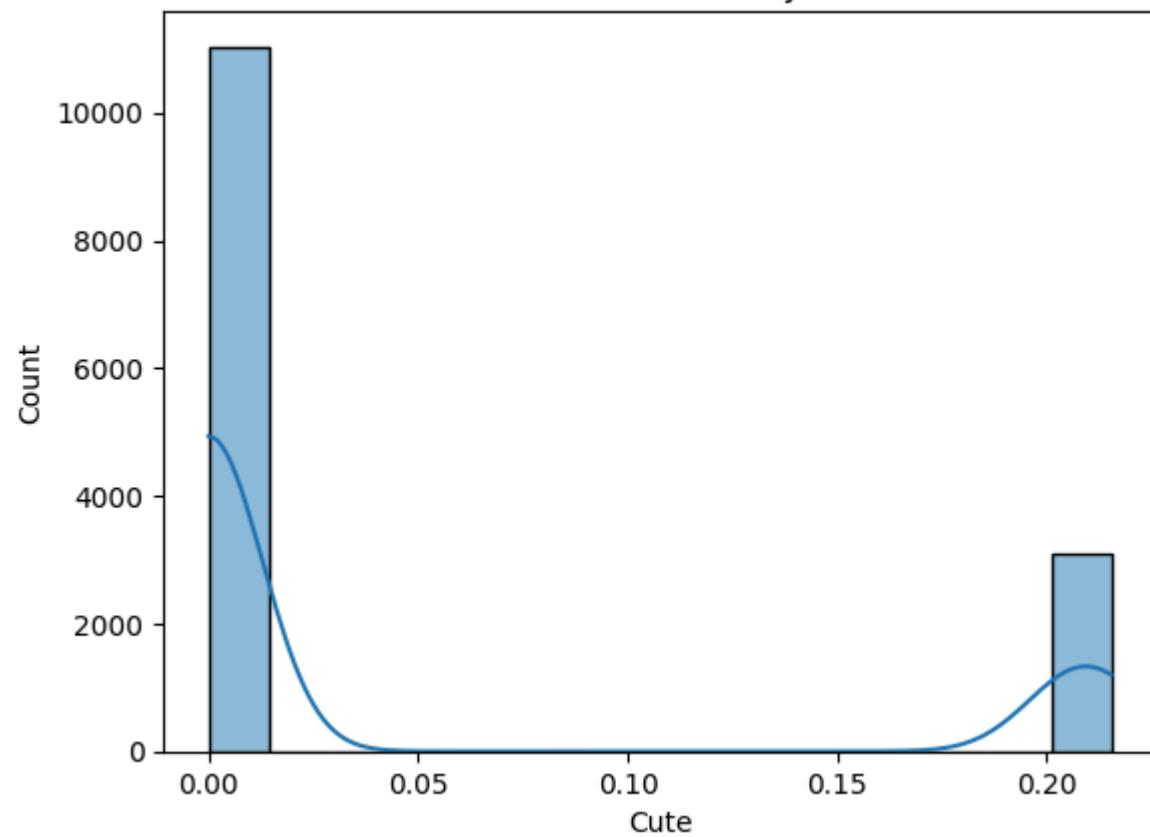




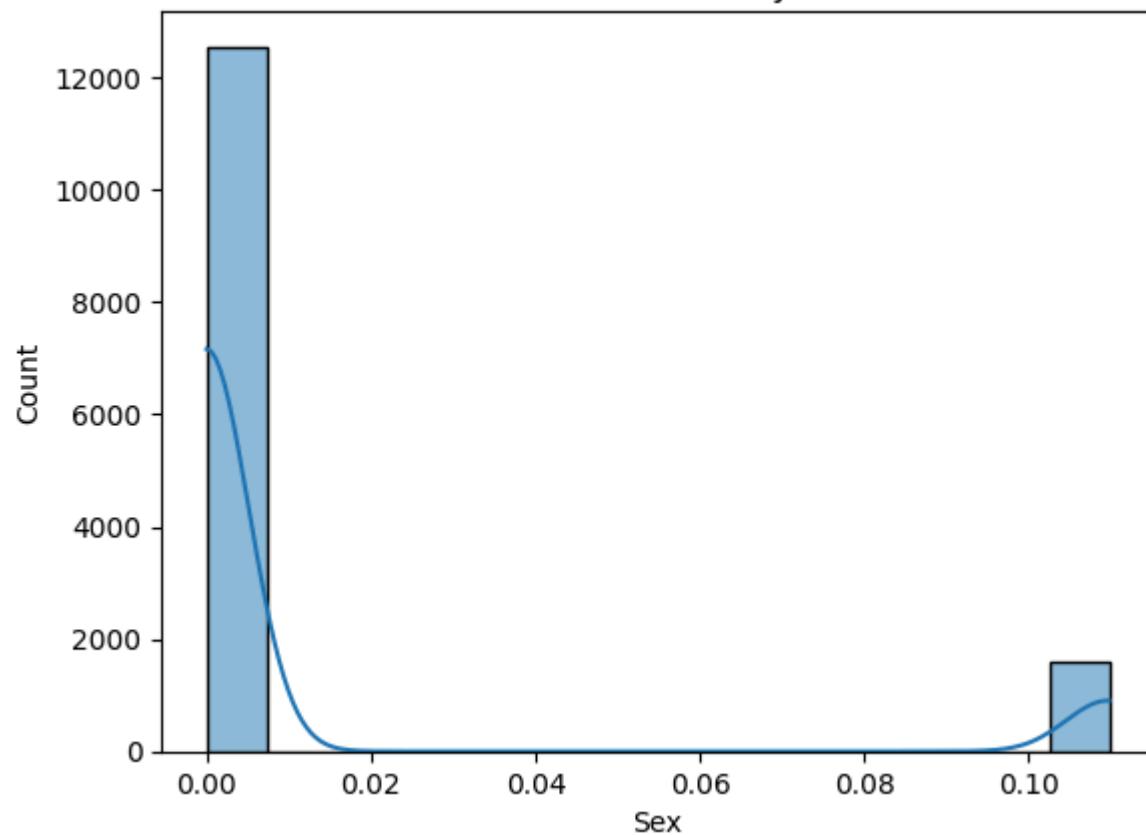
Distribution of Sports after Box-Cox/Yeo-Johnson transformation



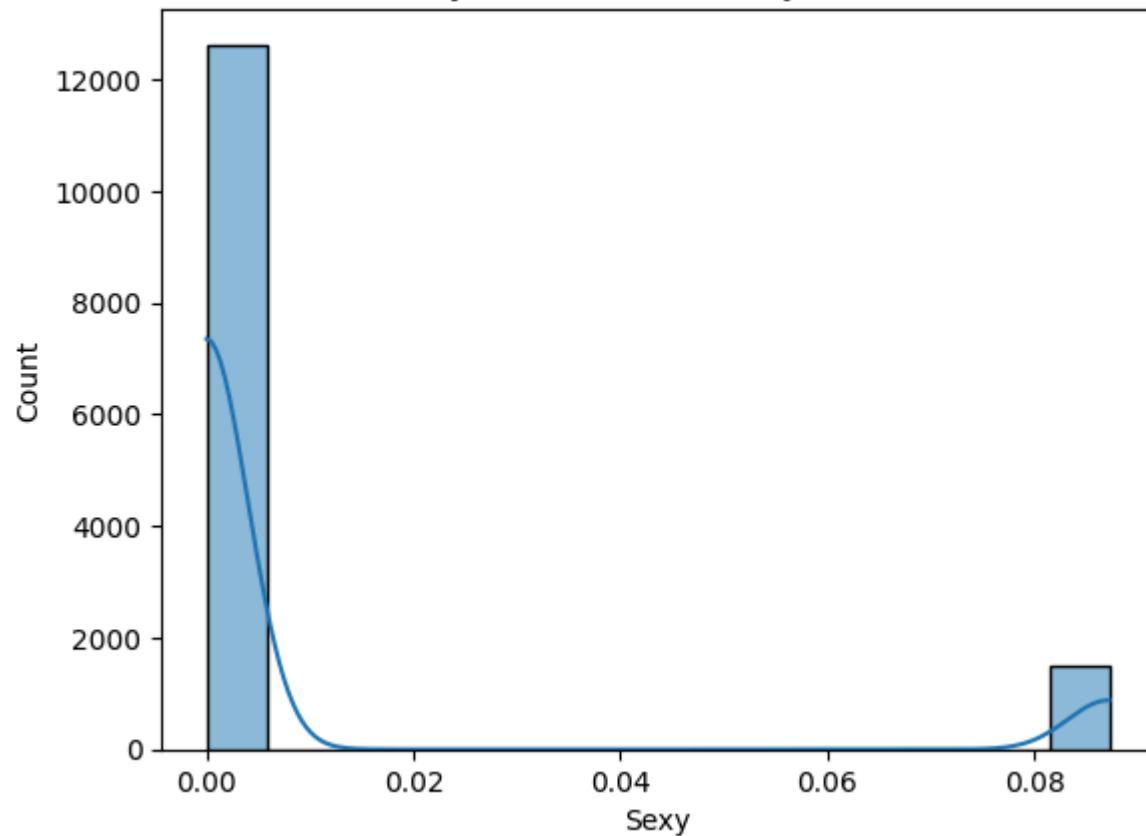
Distribution of Cute after Box-Cox/Yeo-Johnson transformation



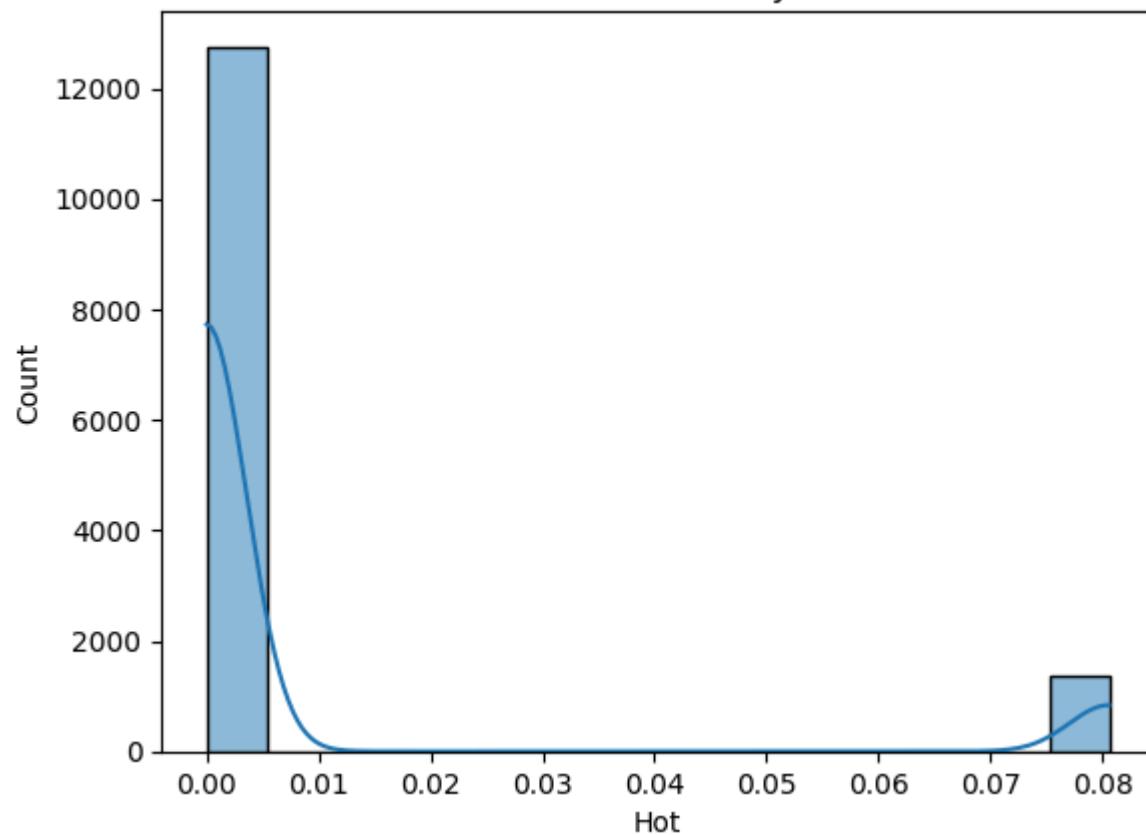
Distribution of Sex after Box-Cox/Yeo-Johnson transformation



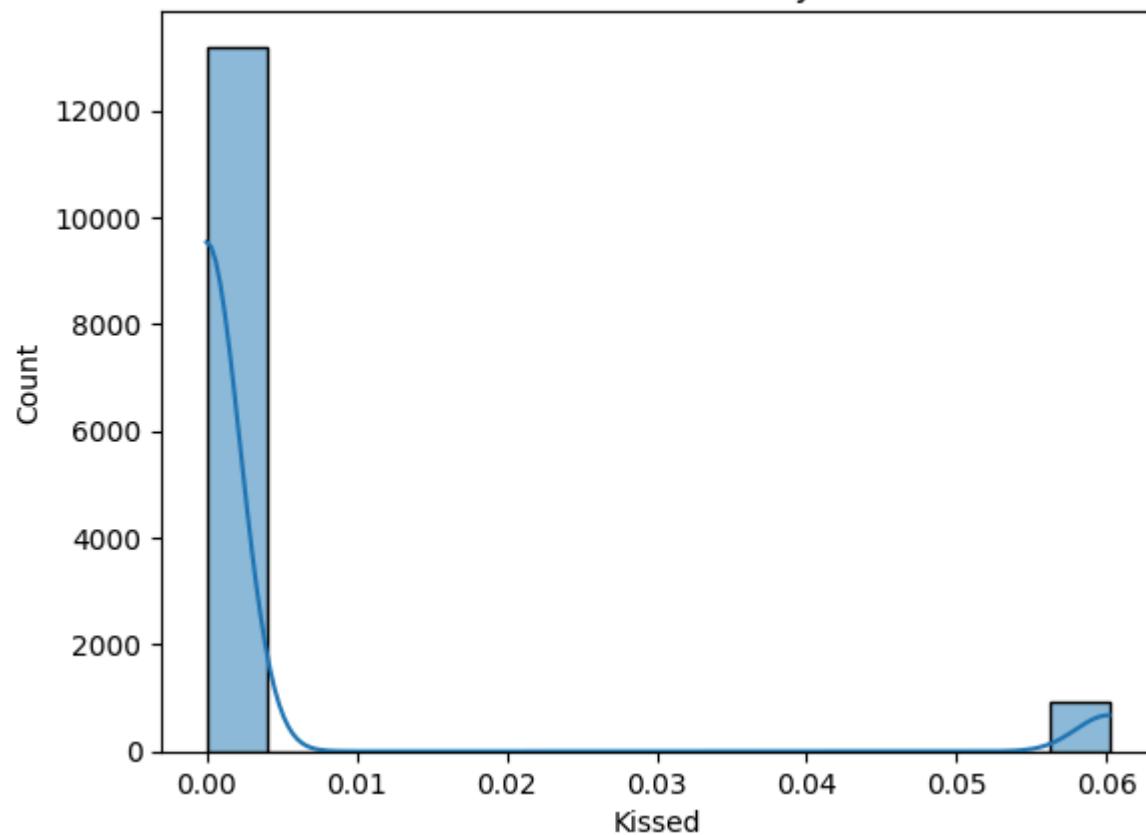
Distribution of Sexy after Box-Cox/Yeo-Johnson transformation



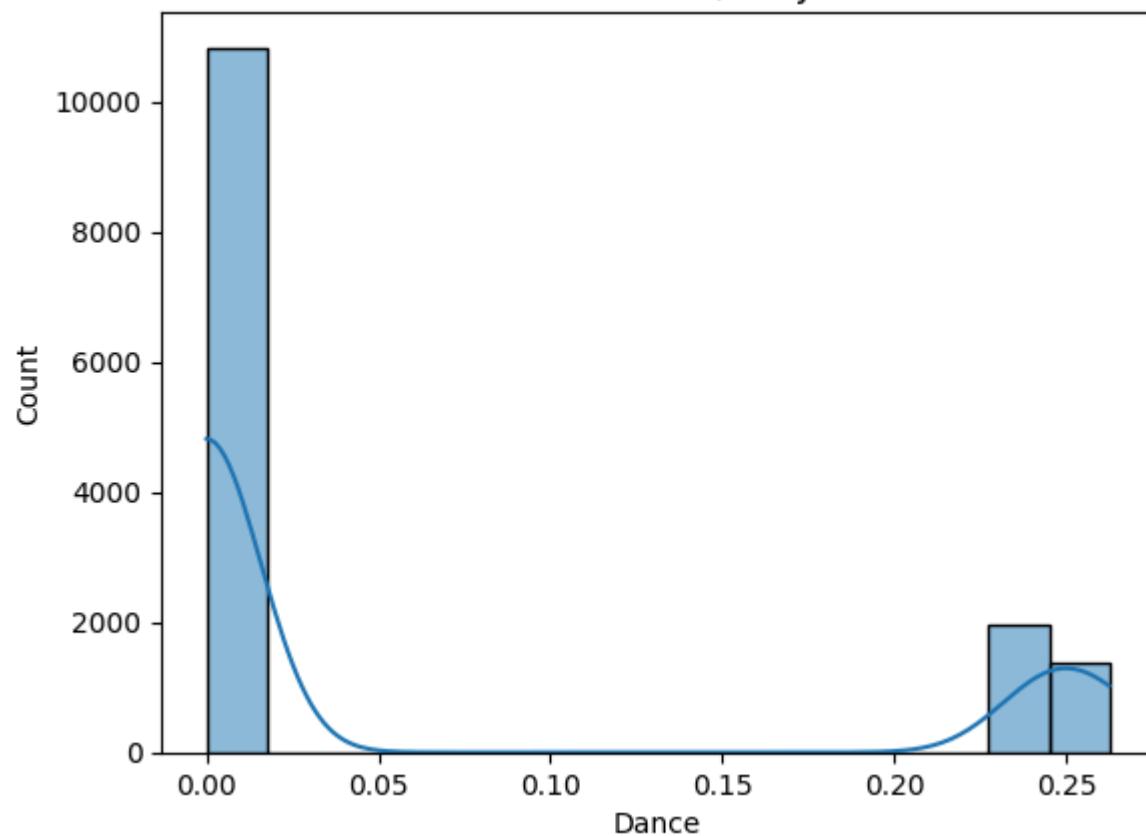
Distribution of Hot after Box-Cox/Yeo-Johnson transformation



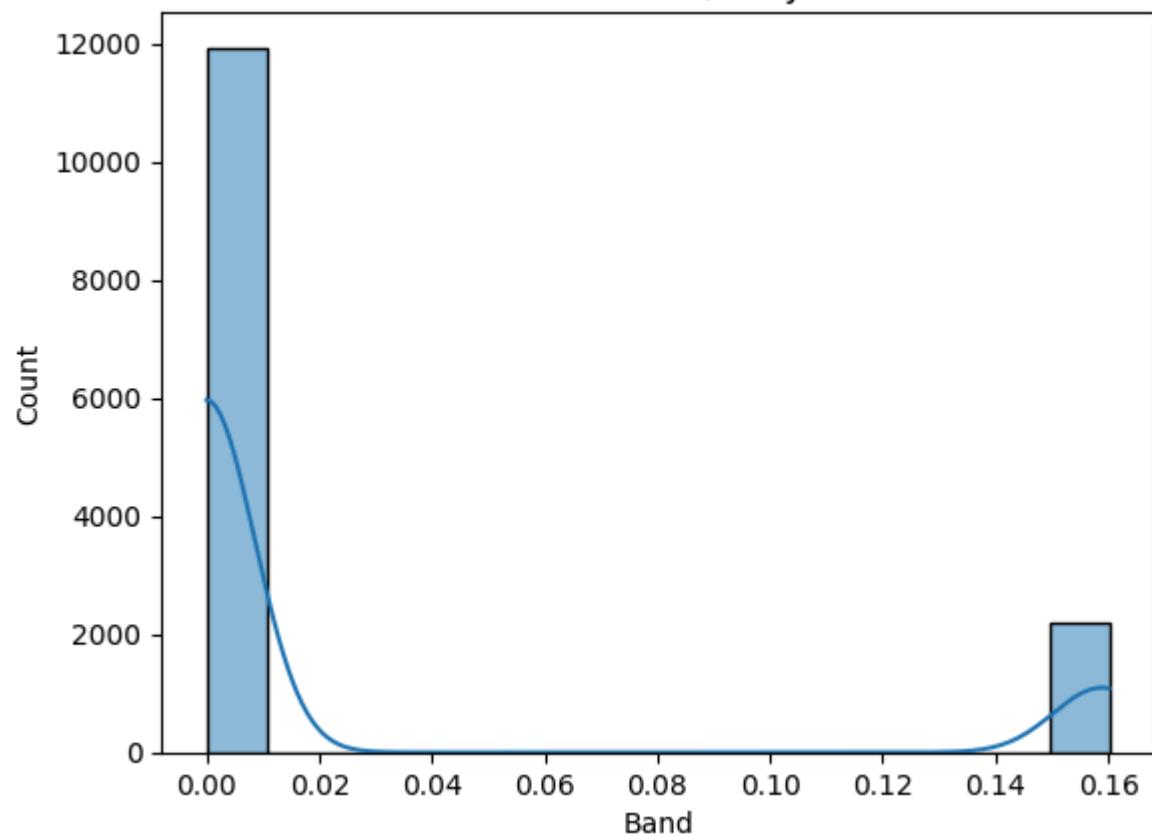
Distribution of Kissed after Box-Cox/Yeo-Johnson transformation



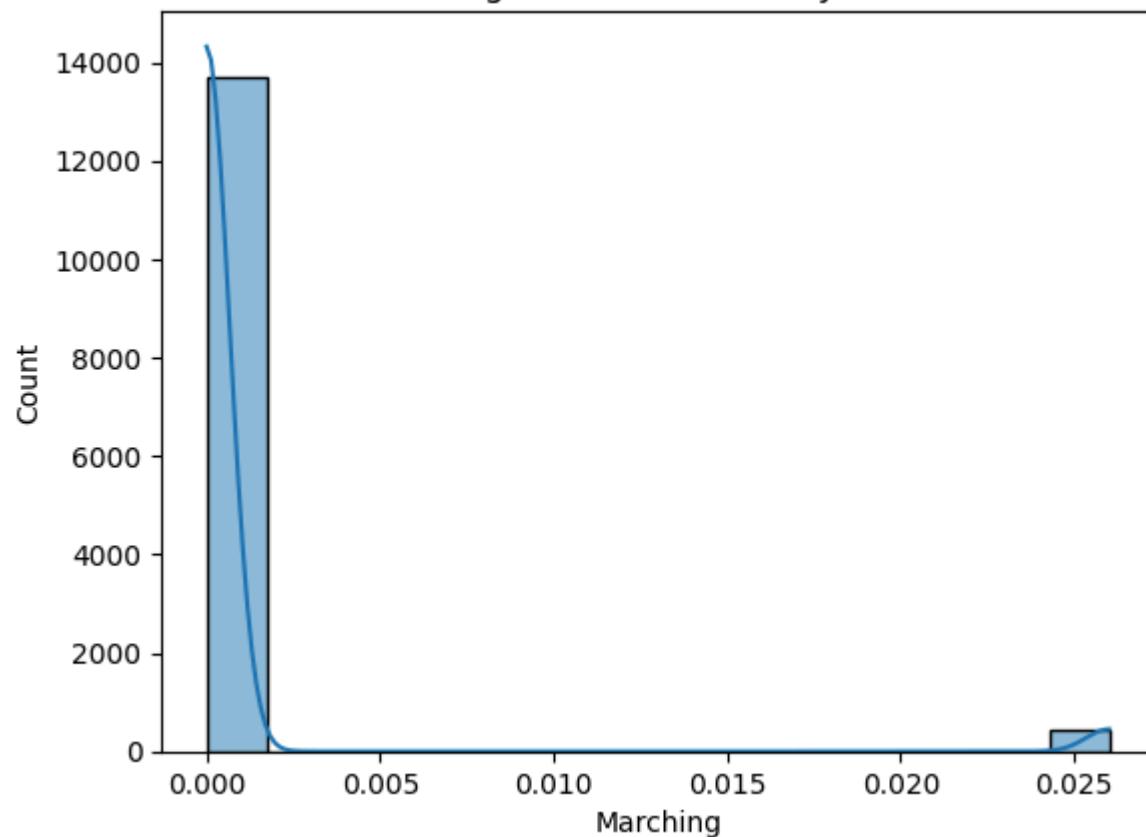
Distribution of Dance after Box-Cox/Yeo-Johnson transformation



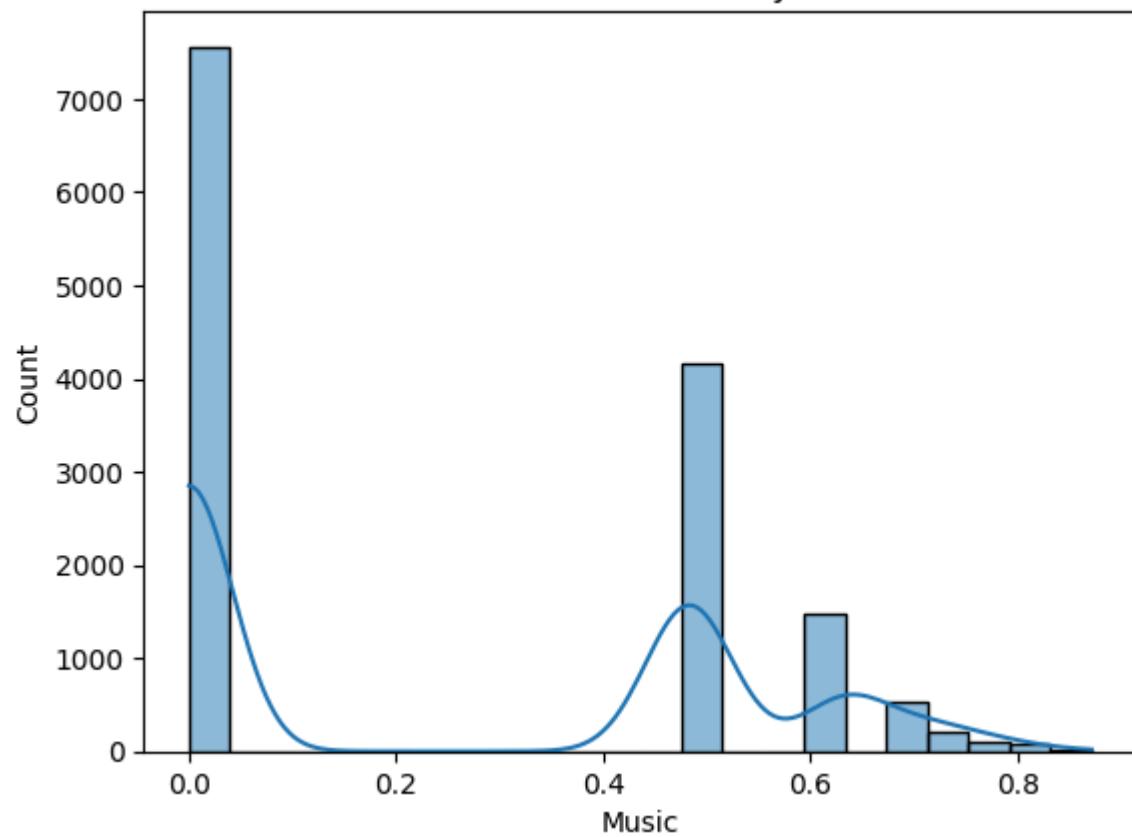
Distribution of Band after Box-Cox/Yeo-Johnson transformation



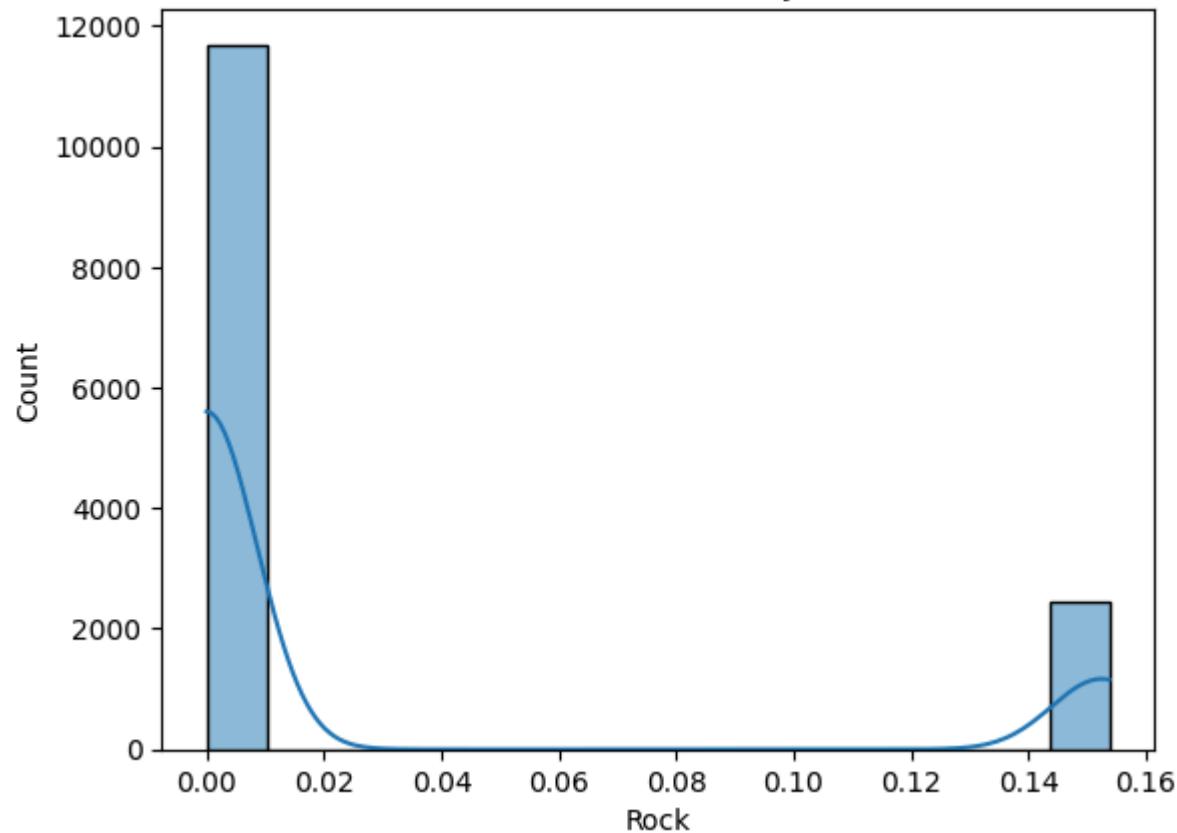
Distribution of Marching after Box-Cox/Yeo-Johnson transformation



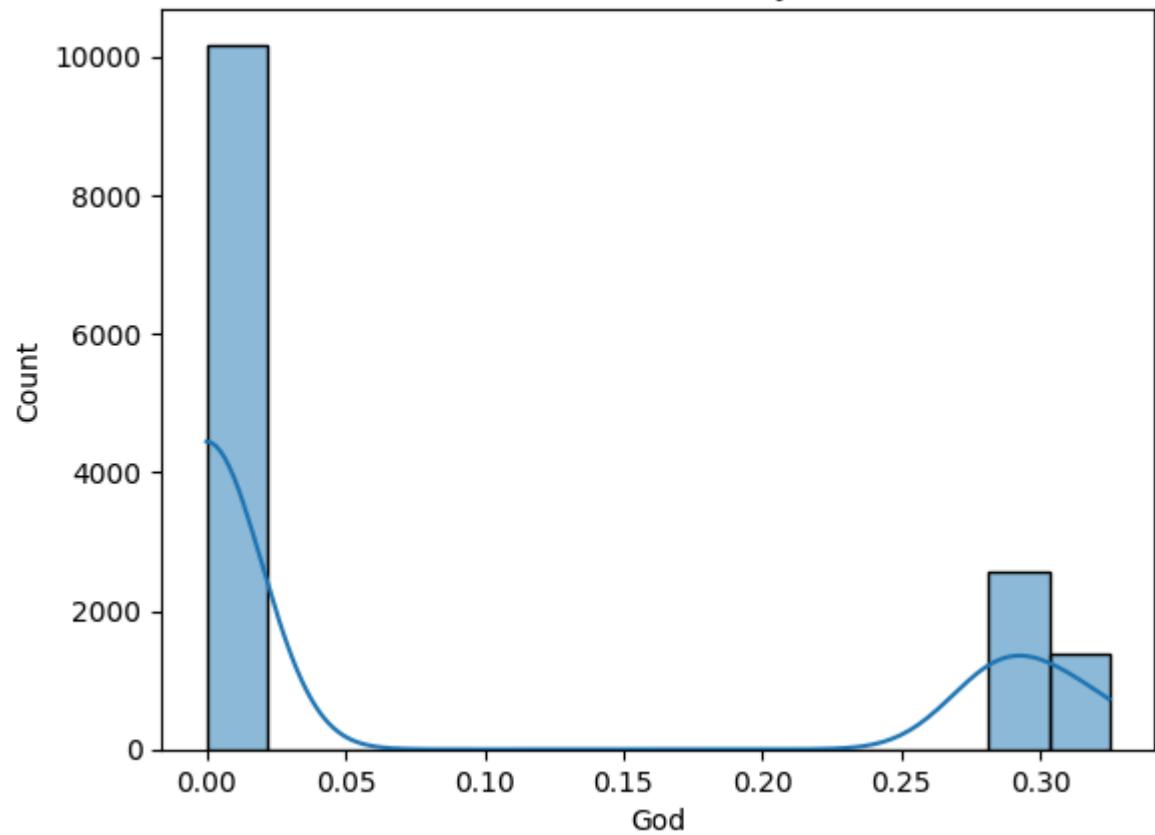
Distribution of Music after Box-Cox/Yeo-Johnson transformation



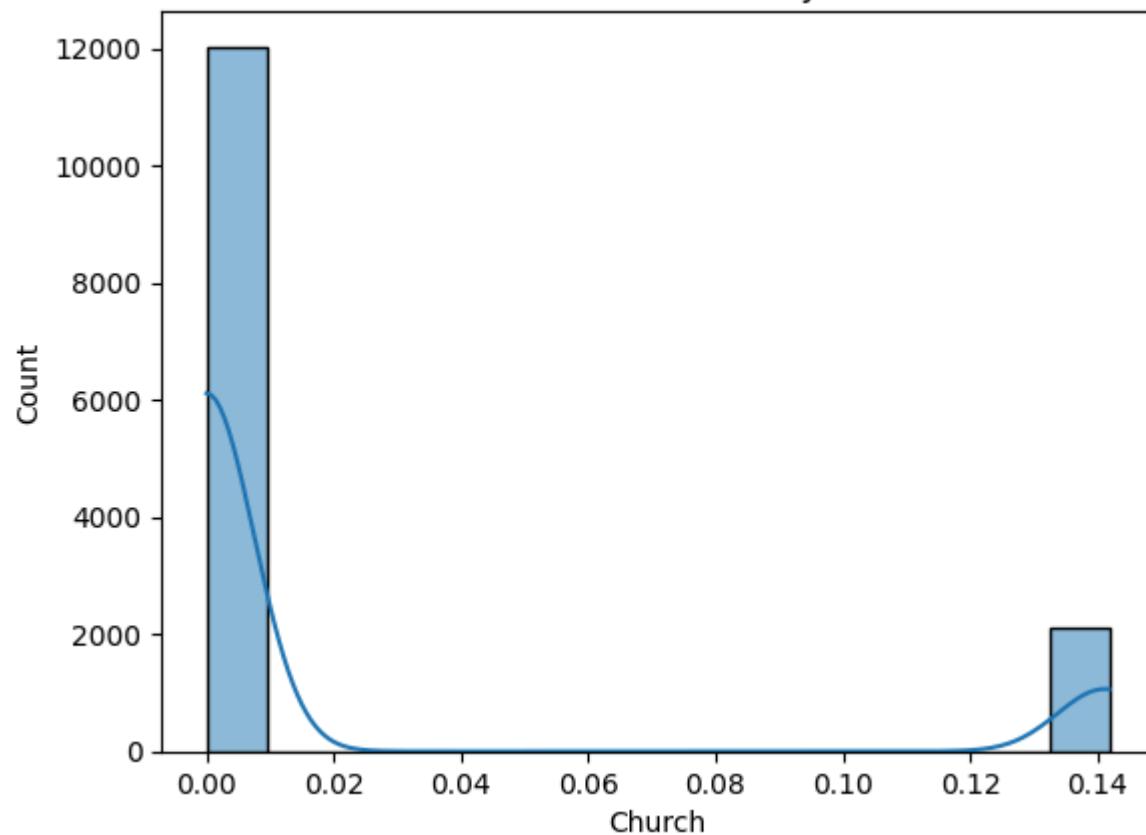
Distribution of Rock after Box-Cox/Yeo-Johnson transformation



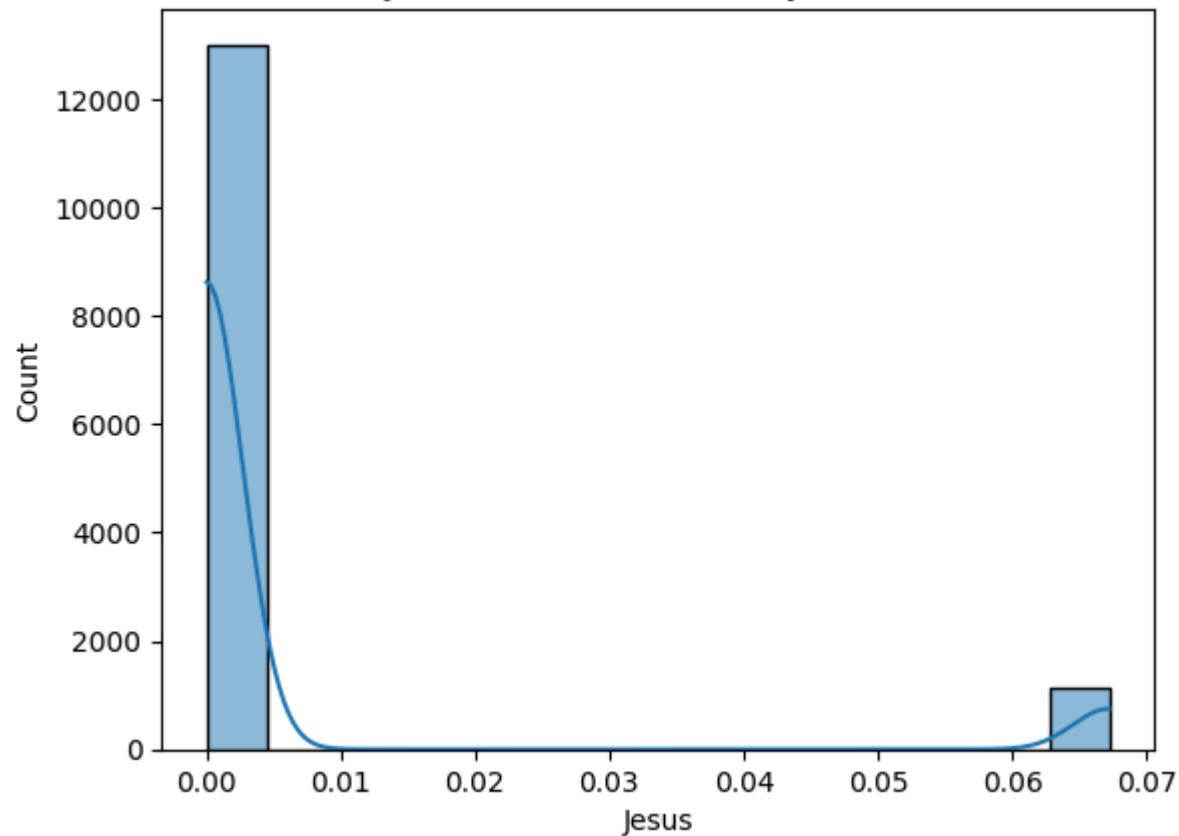
Distribution of God after Box-Cox/Yeo-Johnson transformation



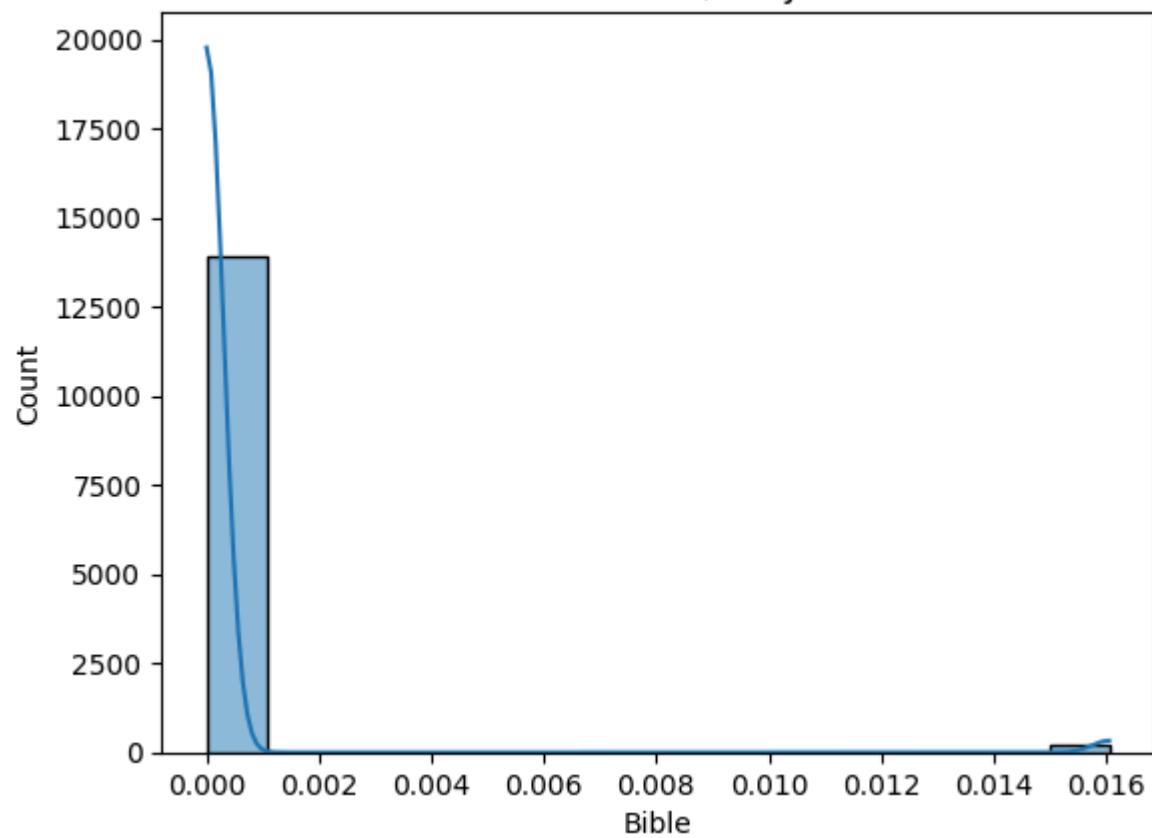
Distribution of Church after Box-Cox/Yeo-Johnson transformation



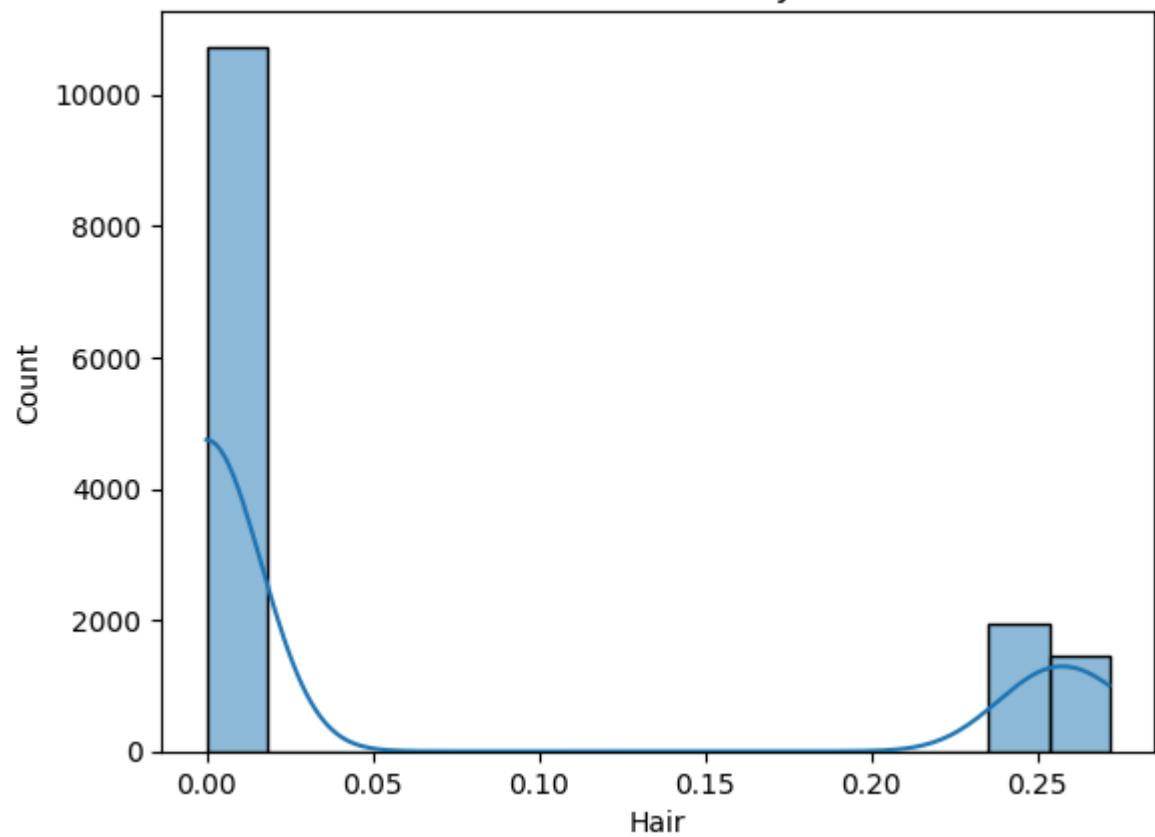
Distribution of Jesus after Box-Cox/Yeo-Johnson transformation



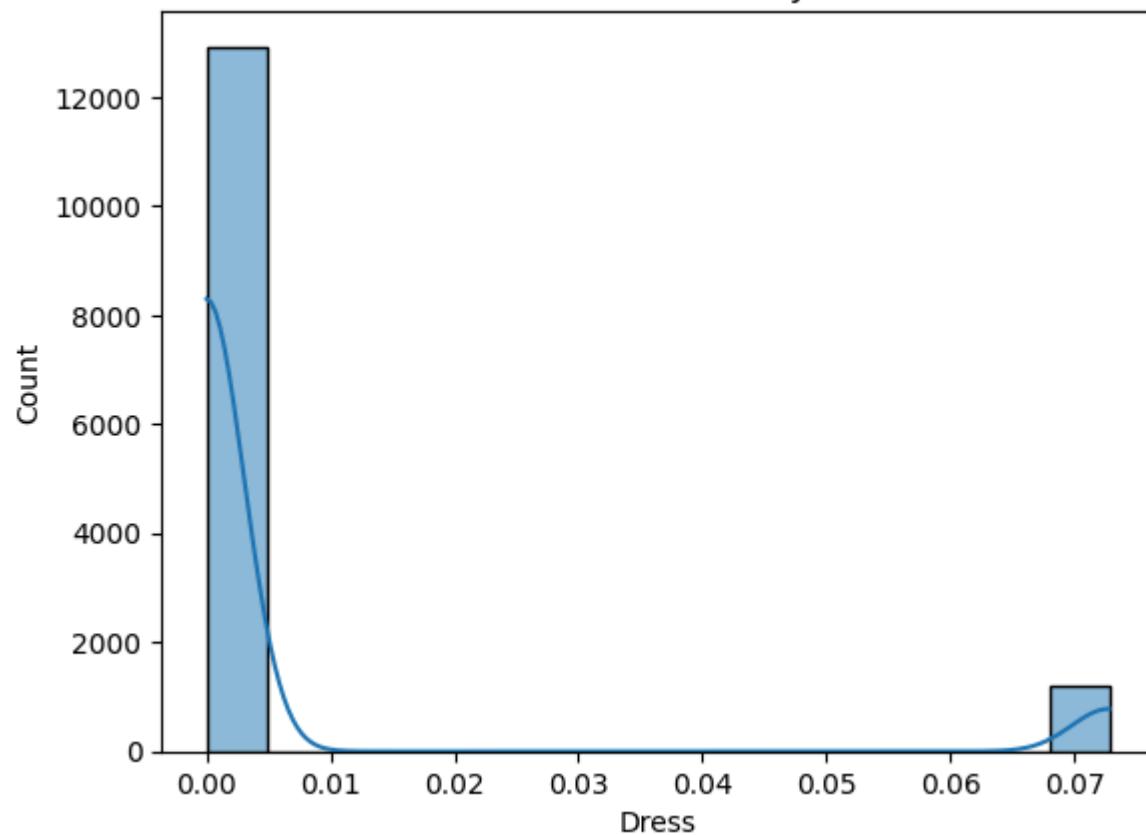
Distribution of Bible after Box-Cox/Yeo-Johnson transformation



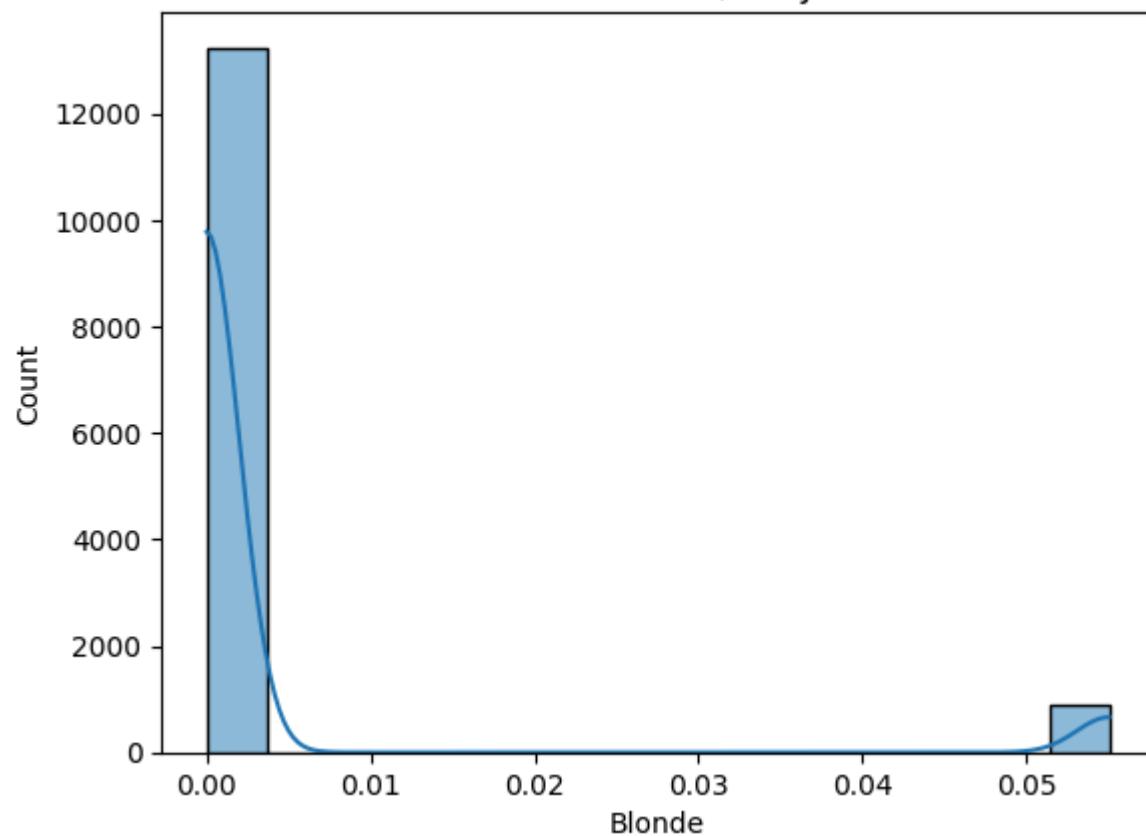
Distribution of Hair after Box-Cox/Yeo-Johnson transformation

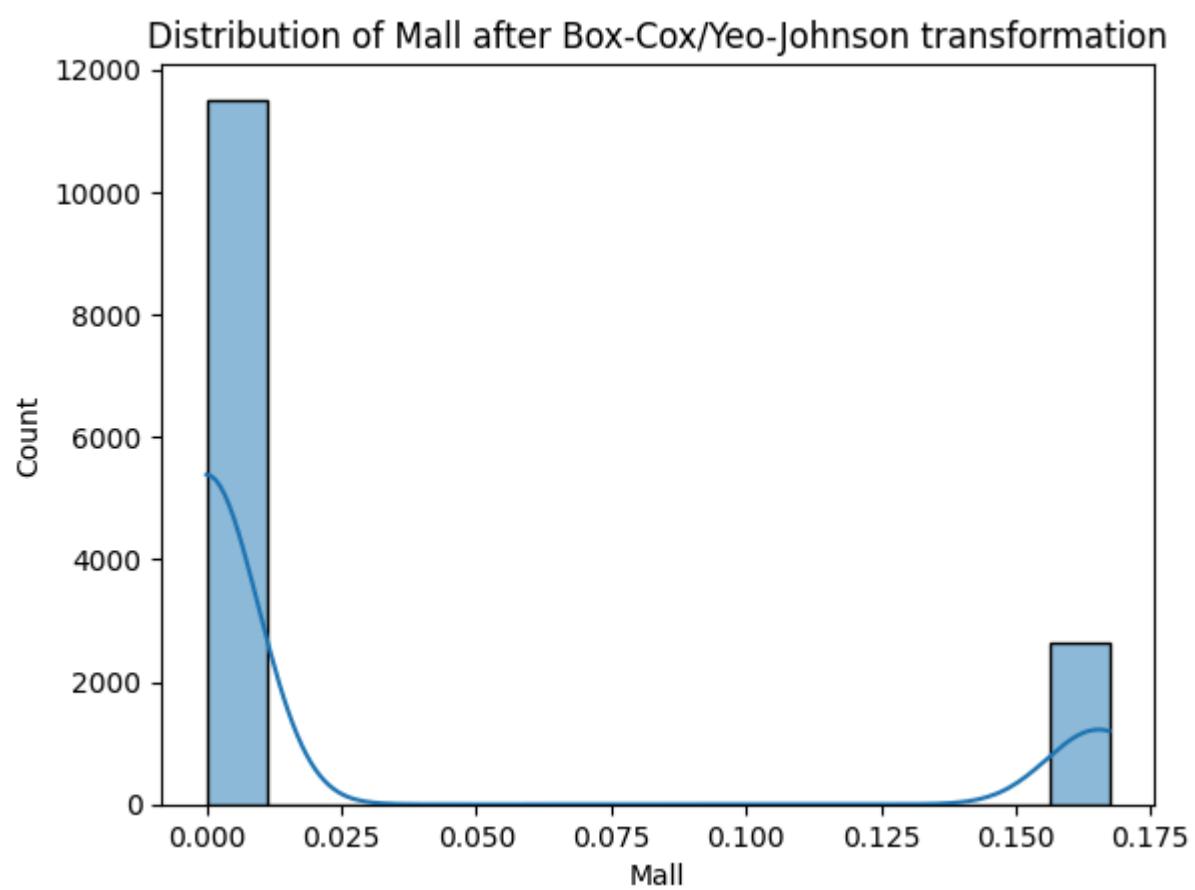


Distribution of Dress after Box-Cox/Yeo-Johnson transformation

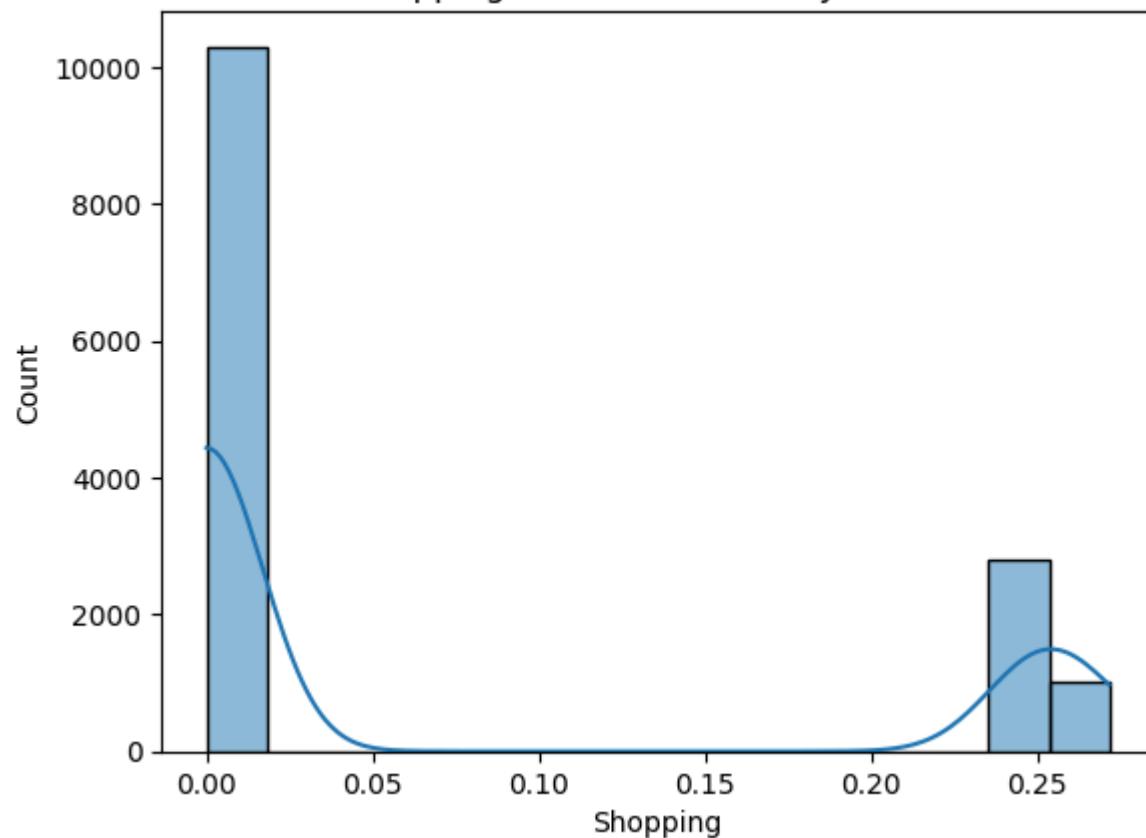


Distribution of Blonde after Box-Cox/Yeo-Johnson transformation

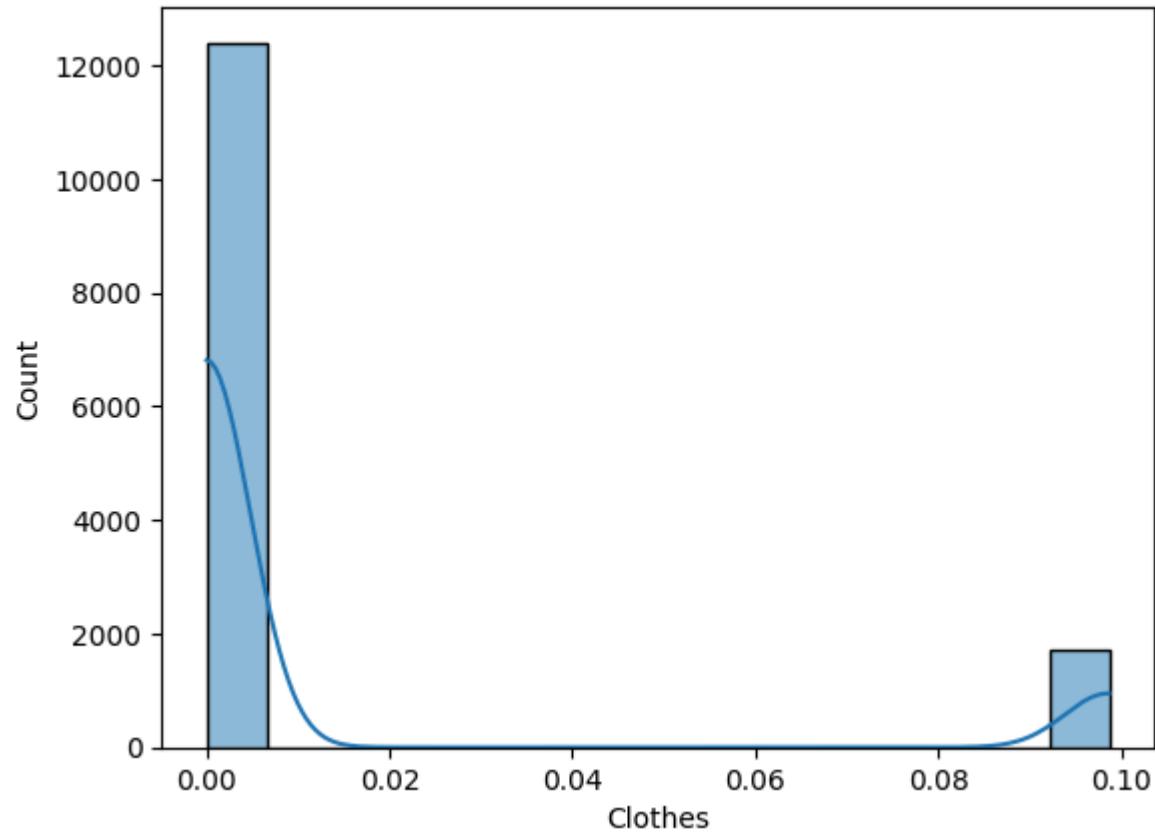


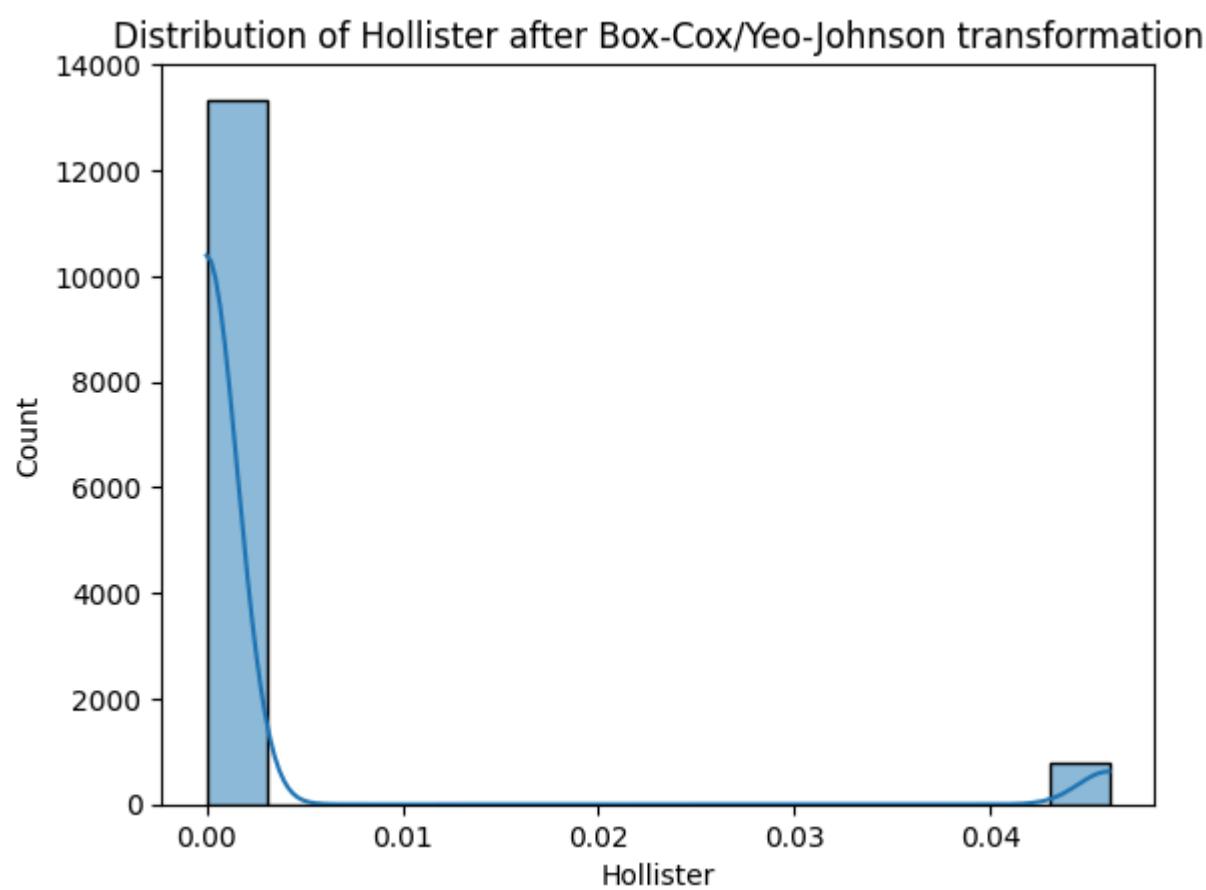


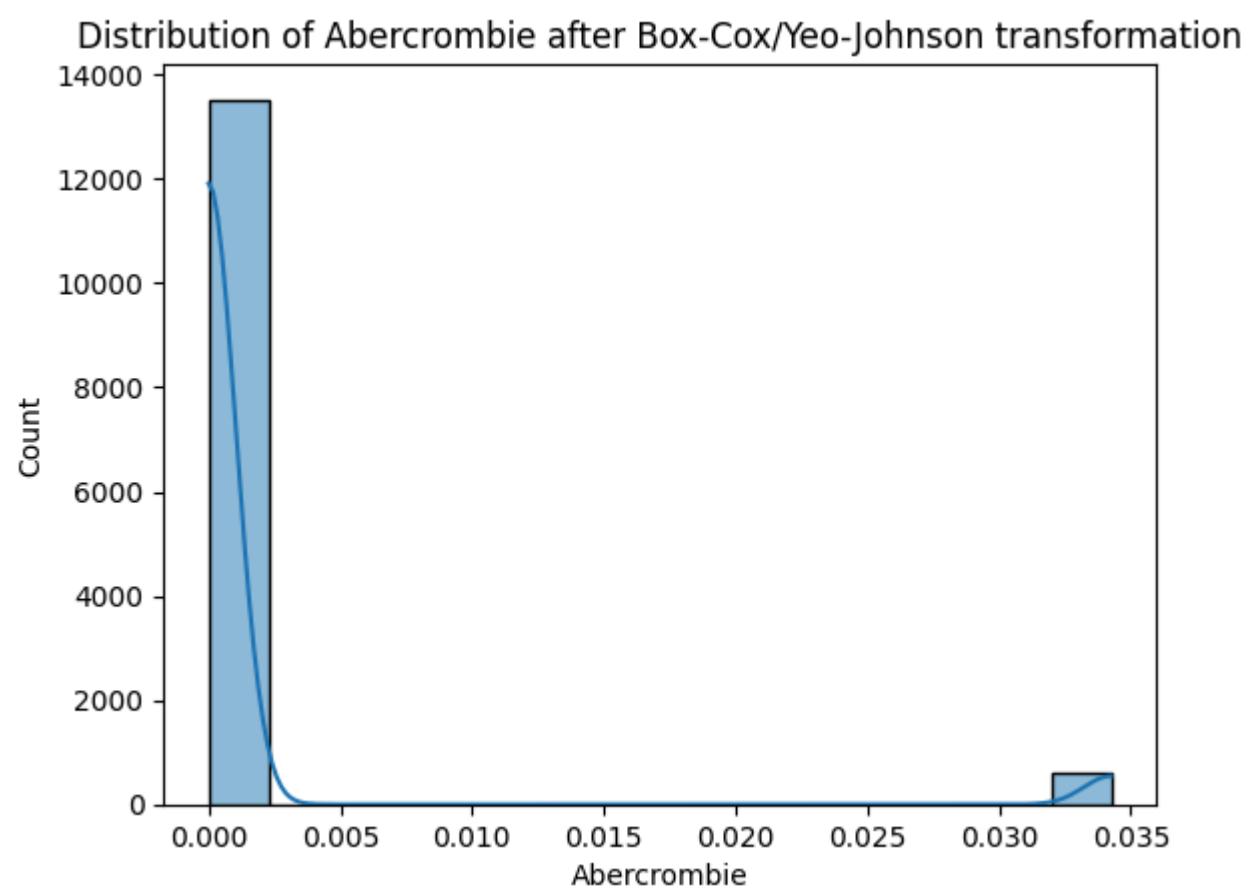
Distribution of Shopping after Box-Cox/Yeo-Johnson transformation



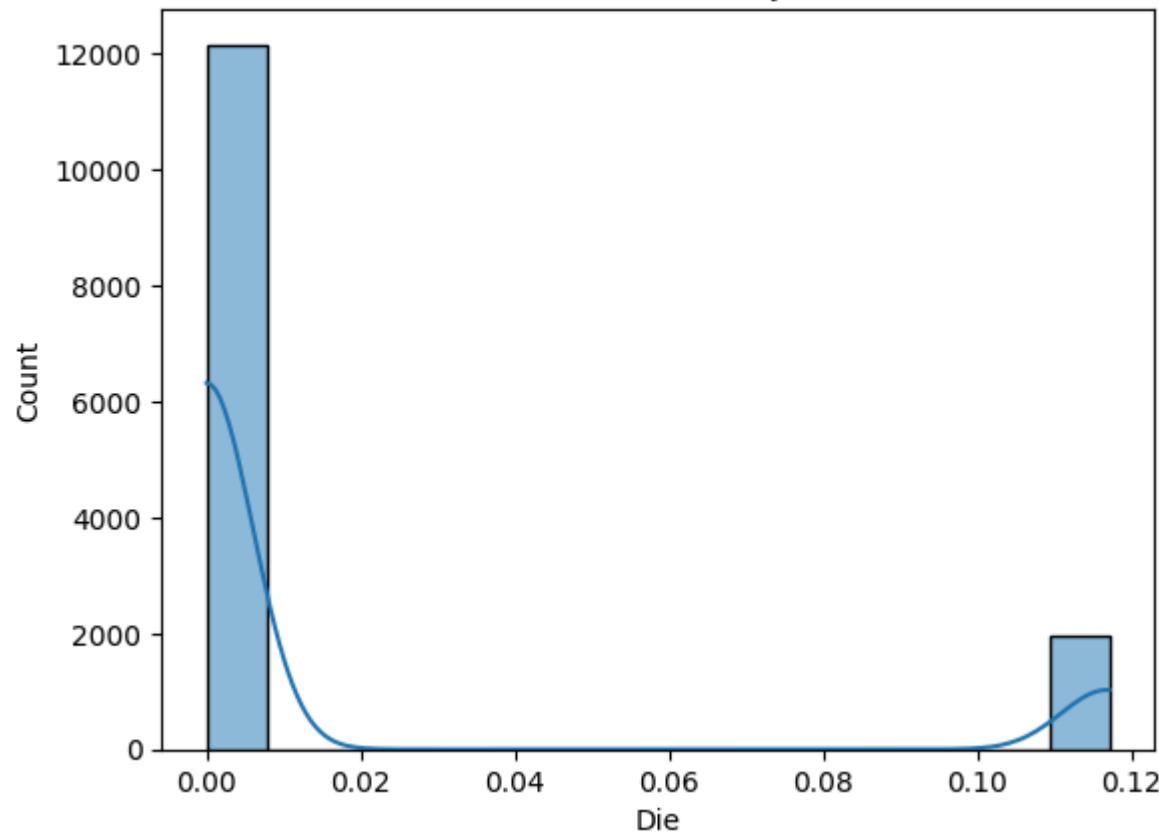
Distribution of Clothes after Box-Cox/Yeo-Johnson transformation



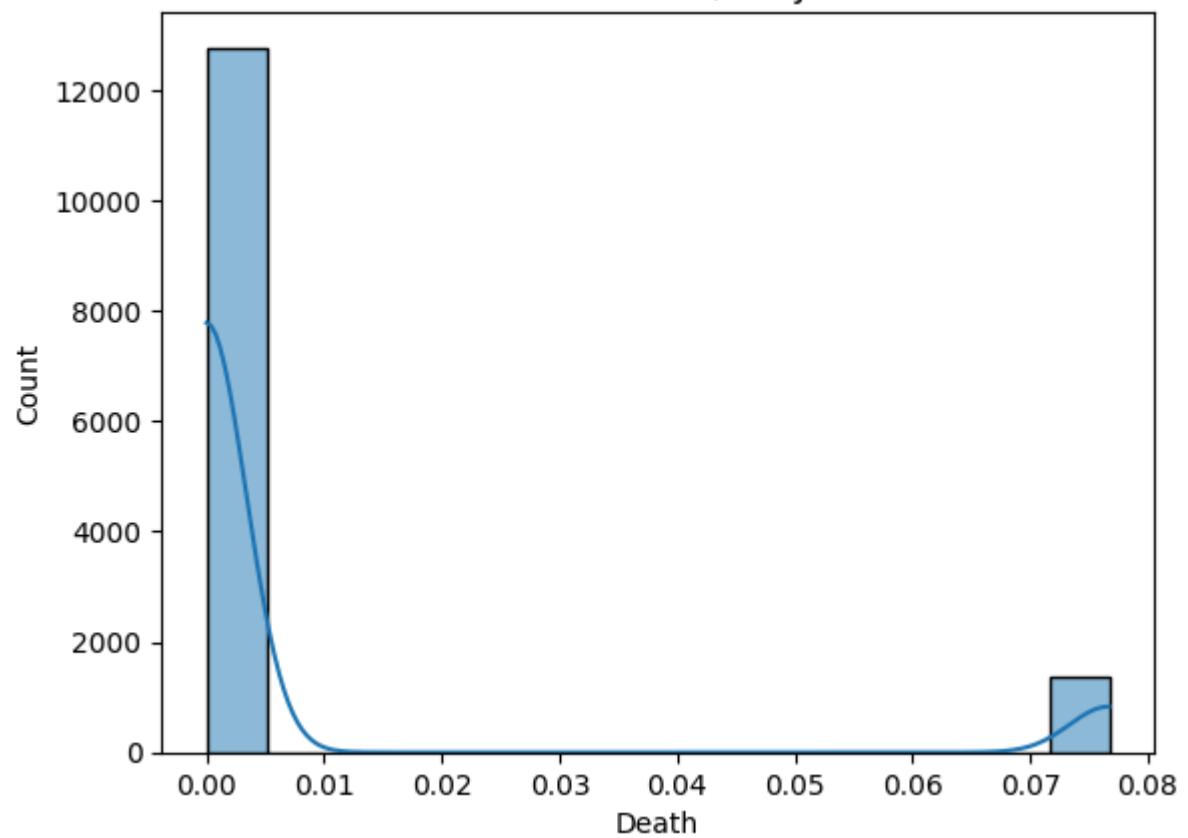




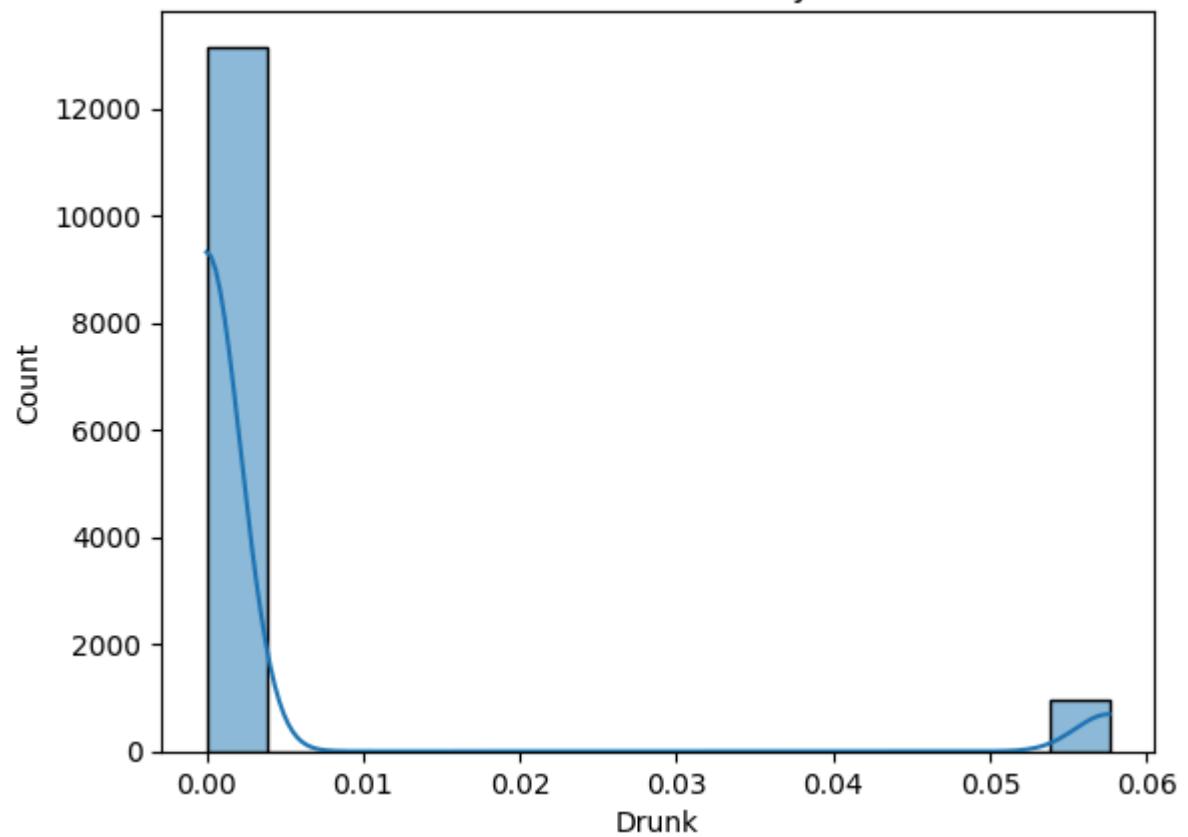
Distribution of Die after Box-Cox/Yeo-Johnson transformation

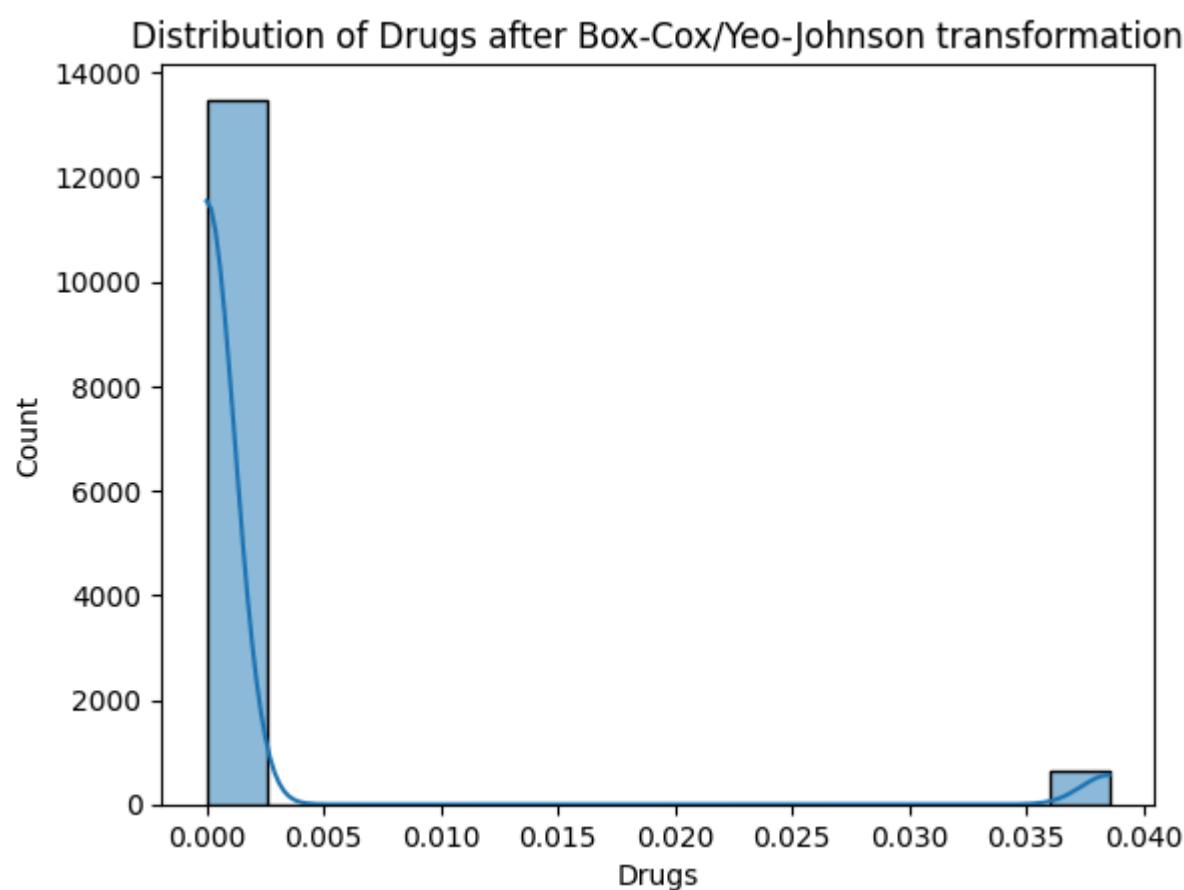


Distribution of Death after Box-Cox/Yeo-Johnson transformation



Distribution of Drunk after Box-Cox/Yeo-Johnson transformation





```
Gradyear      -0.007613
Gender        1.701090
Age          -3.221129
NumberOfffriends -0.110822
Basketball     1.753496
Football       1.695520
Soccer         2.522294
Softball        2.990517
Volleyball      2.967643
Swimming        2.572752
Cheerleading    3.529038
Baseball        3.402205
Tennis          4.004190
Sports          2.353830
Cute            1.358382
Sex             2.460029
Sexy            2.544887
Hot             2.710330
Kissed          3.484826
Dance           1.260589
Band            1.897368
Marching        5.455278
Music           0.319238
Rock            1.729009
God             0.989367
Church          1.976609
Jesus            3.096397
Bible           7.647864
Hair            1.217837
Dress           2.956341
Blonde          3.586745
Mall            1.615125
Shopping         1.034466
Clothes          2.305713
Hollister        3.841244
Abercrombie     4.471896
Die              2.079507
Death            2.734874
Drunk            3.393535
Drugs            4.320550
dtype: float64
```

Checking for **Outliers**

```
In [58]: def check_outliers(df2):
    outliers = {}
```

```
for column in df.select_dtypes(include=['number']).columns:
    Q1 = df2[column].quantile(0.25)
    Q3 = df2[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers[column] = df2[(df2[column] < lower_bound) | (df2[column] > upper_bound)]
    #data['Column'] = np.where(data['column'] < lower_bound, lower_bound, data['column'])
return outliers
outliers = check_outliers(df2)
print(outliers)
```

{'Gradyear': Empty DataFrame

Columns: [Gradyear, Gender, Age, NumberOffriends, Basketball, Football, Soccer, Softball, Volleyball, Swimming, Cheerleading, Baseball, Tennis, Sports, Cute, Sex, Sexy, Hot, Kissed, Dance, Band, Marching, Music, Rock, God, Church, Jesus, Bible, Hair, Dress, Blonde, Mall, Shopping, Clothes, Hollister, Abercrombie, Die, Death, Drunk, Drugs]

Index: []

[0 rows x 40 columns], 'Age': Gradyear Gender Age NumberOffriends Basketball \

112	4.560946e+98	0.000000	0.987975	0.000000	0.000000	
117	4.630399e+98	0.000000	1.175558	4.251580	0.000000	
253	4.772387e+98	0.000000	1.175558	3.914234	0.161589	
264	4.772387e+98	0.124251	1.175363	6.326553	0.000000	
450	4.700874e+98	0.000000	1.149878	5.633260	0.000000	
...	
14345	4.560946e+98	0.000000	1.170286	6.220999	0.000000	
14351	4.700874e+98	0.000000	1.174961	5.633260	0.000000	
14890	4.700874e+98	0.124251	1.048818	0.000000	0.000000	
14899	4.560946e+98	0.000000	0.987975	2.469581	0.000000	
14989	4.630399e+98	0.000000	1.175558	6.055888	0.000000	

Football Soccer Softball Volleyball Swimming ... Blonde Mall \

112	0.0	0.000000	0.0	0.0	0.08534	...	0.0	0.0	
117	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.0	
253	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.0	
264	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.0	
450	0.0	0.000000	0.0	0.0	0.08534	...	0.0	0.0	
...	
14345	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.0	
14351	0.0	0.112332	0.0	0.0	0.00000	...	0.0	0.0	
14890	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.0	
14899	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.0	
14989	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.0	

Shopping Clothes Hollister Abercrombie Die Death Drunk \

112	0.00000	0.098613	0.000000	0.0	0.000000	0.0	0.000000	
117	0.25053	0.098613	0.046088	0.0	0.117151	0.0	0.057613	
253	0.00000	0.098613	0.000000	0.0	0.000000	0.0	0.000000	
264	0.25053	0.000000	0.000000	0.0	0.000000	0.0	0.000000	
450	0.00000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	
...	
14345	0.00000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	
14351	0.00000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	
14890	0.00000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	
14899	0.00000	0.000000	0.000000	0.0	0.000000	0.0	0.057613	
14989	0.00000	0.000000	0.000000	0.0	0.116845	0.0	0.000000	

Drugs

```
112  0.000000
117  0.038533
253  0.000000
264  0.000000
450  0.038533
...
14345 0.000000
14351 0.000000
14890 0.000000
14899 0.000000
14989 0.000000
```

```
[166 rows x 40 columns], 'NumberOffriends':           Gradyear   Gender    Age NumberOfffriends Basketball \
297  4.700874e+98  0.124251  1.081084      14.850032  0.000000
3398 4.772387e+98  0.000000  1.074503      12.229655  0.000000
5476 4.772387e+98  0.000000  1.081084      13.750083  0.000000
6560 4.700874e+98  0.000000  1.081084      12.241975  0.000000
12409 4.560946e+98  0.000000  1.092219      12.778963  0.000000
14056 4.630399e+98  0.124251  1.086951      14.978608  0.159402
```

```
Football Soccer Softball Volleyball Swimming ... Blonde Mall \
297      0.0     0.0  0.000000  0.000000  0.000000 ... 0.055048  0.0
3398     0.0     0.0  0.000000  0.000000  0.08534 ... 0.000000  0.0
5476     0.0     0.0  0.083794  0.079812  0.000000 ... 0.000000  0.0
6560     0.0     0.0  0.000000  0.000000  0.000000 ... 0.000000  0.0
12409    0.0     0.0  0.000000  0.079798  0.000000 ... 0.000000  0.0
14056    0.0     0.0  0.000000  0.000000  0.000000 ... 0.000000  0.0
```

```
Shopping Clothes Hollister Abercrombie Die Death \
297  0.00000  0.0     0.0     0.0  0.117151  0.076709
3398 0.00000  0.0     0.0     0.0  0.116845  0.000000
5476 0.25053  0.0     0.0     0.0  0.000000  0.076709
6560 0.00000  0.0     0.0     0.0  0.000000  0.000000
12409 0.00000  0.0     0.0     0.0  0.000000  0.000000
14056 0.00000  0.0     0.0     0.0  0.000000  0.000000
```

```
Drunk Drugs
297  0.057613  0.0
3398 0.000000  0.0
5476 0.000000  0.0
6560 0.000000  0.0
12409 0.000000  0.0
14056 0.000000  0.0
```

```
[6 rows x 40 columns], 'Basketball':           Gradyear   Gender    Age NumberOfffriends Basketball \
23   4.630399e+98  0.000000  1.092219      8.141973  0.159402
```

38	4.560946e+98	0.000000	1.092219	9.858907	0.159402
42	4.772387e+98	0.000000	1.081084	5.142291	0.159402
43	4.630399e+98	0.000000	1.086951	6.763717	0.161439
53	4.560946e+98	0.000000	1.092219	9.338978	0.161439
...
14977	4.700874e+98	0.000000	1.081084	6.941186	0.159402
14978	4.772387e+98	0.000000	1.074503	0.753045	0.159402
14982	4.772387e+98	0.000000	1.074503	4.550377	0.159402
14984	4.700874e+98	0.000000	1.081084	5.142291	0.159402
14990	4.772387e+98	0.124251	1.081084	4.354949	0.159402

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
23	0.000000	0.000000	0.000000	0.079812	0.0	...	0.000000	
38	0.000000	0.000000	0.000000	0.000000	0.0	...	0.000000	
42	0.000000	0.112097	0.000000	0.000000	0.0	...	0.000000	
43	0.000000	0.000000	0.000000	0.000000	0.0	...	0.000000	
53	0.000000	0.000000	0.083815	0.000000	0.0	...	0.000000	
...
14977	0.161396	0.112331	0.083794	0.000000	0.0	...	0.000000	
14978	0.000000	0.000000	0.000000	0.079798	0.0	...	0.000000	
14982	0.000000	0.000000	0.000000	0.079798	0.0	...	0.055048	
14984	0.000000	0.000000	0.000000	0.000000	0.0	...	0.000000	
14990	0.000000	0.000000	0.000000	0.000000	0.0	...	0.000000	

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
23	0.000000	0.00000	0.000000	0.0	0.034255	0.000000	
38	0.164666	0.25053	0.098613	0.0	0.000000	0.000000	
42	0.000000	0.25053	0.000000	0.0	0.000000	0.000000	
43	0.000000	0.00000	0.000000	0.0	0.000000	0.000000	
53	0.000000	0.25053	0.098613	0.0	0.000000	0.000000	
...
14977	0.000000	0.00000	0.098613	0.0	0.000000	0.117151	
14978	0.000000	0.00000	0.000000	0.0	0.000000	0.000000	
14982	0.000000	0.00000	0.000000	0.0	0.000000	0.000000	
14984	0.000000	0.00000	0.000000	0.0	0.000000	0.116845	
14990	0.000000	0.00000	0.000000	0.0	0.000000	0.000000	

	Death	Drunk	Drugs
23	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000000
42	0.000000	0.000000	0.000000
43	0.000000	0.000000	0.000000
53	0.000000	0.000000	0.000000
...
14977	0.076709	0.057613	0.038533
14978	0.000000	0.000000	0.000000

14982	0.000000	0.000000	0.000000
14984	0.000000	0.000000	0.000000
14990	0.000000	0.000000	0.000000

[2409 rows x 40 columns], 'Football':

	Gradyear	Gender	Age	NumberOffriends	Basketball	\
5	4.700874e+98	0.124251	1.092219	5.432670	0.000000	
15	4.630399e+98	0.000000	1.086951	5.759993	0.000000	
21	4.560946e+98	0.000000	1.092219	6.378075	0.000000	
32	4.700874e+98	0.000000	1.081084	3.914234	0.000000	
34	4.630399e+98	0.000000	1.086951	5.881727	0.000000	
...	
14966	4.700874e+98	0.000000	1.081084	4.643036	0.000000	
14977	4.700874e+98	0.000000	1.081084	6.941186	0.159402	
14981	4.630399e+98	0.000000	1.086951	6.220999	0.000000	
14985	4.560946e+98	0.000000	1.092219	5.940858	0.000000	
14988	4.772387e+98	0.000000	1.074503	8.798106	0.000000	

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
5	0.161574	0.000000	0.000000		0.0	0.0	...	0.0
15	0.159361	0.000000	0.083816		0.0	0.0	...	0.0
21	0.159361	0.000000	0.000000		0.0	0.0	...	0.0
32	0.161396	0.000000	0.083816		0.0	0.0	...	0.0
34	0.161396	0.000000	0.000000		0.0	0.0	...	0.0
...	
14966	0.159361	0.000000	0.000000		0.0	0.0	...	0.0
14977	0.161396	0.112331	0.083794		0.0	0.0	...	0.0
14981	0.159361	0.000000	0.000000		0.0	0.0	...	0.0
14985	0.159361	0.000000	0.000000		0.0	0.0	...	0.0
14988	0.161396	0.000000	0.000000		0.0	0.0	...	0.0

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
5	0.00000	0.00000	0.000000		0.0	0.0	0.000000
15	0.00000	0.00000	0.000000		0.0	0.0	0.000000
21	0.00000	0.00000	0.000000		0.0	0.0	0.116845
32	0.00000	0.00000	0.000000		0.0	0.0	0.000000
34	0.00000	0.00000	0.000000		0.0	0.0	0.000000
...	
14966	0.00000	0.00000	0.000000		0.0	0.0	0.117160
14977	0.00000	0.00000	0.098613		0.0	0.0	0.117151
14981	0.00000	0.00000	0.098613		0.0	0.0	0.000000
14985	0.00000	0.25053	0.000000		0.0	0.0	0.000000
14988	0.16732	0.00000	0.000000		0.0	0.0	0.000000

	Death	Drunk	Drugs
5	0.000000	0.000000	0.000000
15	0.076709	0.000000	0.000000

21	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000
...
14966	0.000000	0.000000	0.000000
14977	0.076709	0.057613	0.038533
14981	0.000000	0.057613	0.000000
14985	0.000000	0.000000	0.000000
14988	0.000000	0.000000	0.000000

[2498 rows x 40 columns], 'Soccer':

	Gradyear	Gender	Age	NumberOffriends	Basketball	\
1	4.630399e+98	0.000000	1.086951	6.428800	0.000000	
40	4.630399e+98	0.124251	1.092219	7.348854	0.000000	
42	4.772387e+98	0.000000	1.081084	5.142291	0.159402	
67	4.560946e+98	0.000000	1.092219	7.572774	0.000000	
73	4.630399e+98	0.000000	1.086951	3.914234	0.159402	
...	
14926	4.630399e+98	0.000000	1.086951	6.853601	0.000000	
14947	4.560946e+98	0.124251	1.092219	4.550377	0.159402	
14977	4.700874e+98	0.000000	1.081084	6.941186	0.159402	
14980	4.772387e+98	0.000000	1.086951	3.224451	0.000000	
14999	4.772387e+98	0.000000	1.074503	4.903694	0.000000	

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
1	0.000000	0.112097	0.000000	0.0	0.08534	...	0.000000	
40	0.000000	0.112326	0.000000	0.0	0.00000	...	0.000000	
42	0.000000	0.112097	0.000000	0.0	0.00000	...	0.000000	
67	0.000000	0.112332	0.000000	0.0	0.00000	...	0.000000	
73	0.159361	0.112332	0.000000	0.0	0.00000	...	0.055048	
...	
14926	0.159361	0.112331	0.000000	0.0	0.00000	...	0.000000	
14947	0.159361	0.112097	0.000000	0.0	0.00000	...	0.000000	
14977	0.161396	0.112331	0.083794	0.0	0.00000	...	0.000000	
14980	0.000000	0.112097	0.000000	0.0	0.00000	...	0.000000	
14999	0.000000	0.112332	0.000000	0.0	0.00000	...	0.000000	

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
1	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
40	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
42	0.000000	0.25053	0.000000	0.0	0.0	0.000000	
67	0.164666	0.00000	0.000000	0.0	0.0	0.000000	
73	0.164666	0.00000	0.098613	0.0	0.0	0.000000	
...	
14926	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
14947	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
14977	0.000000	0.00000	0.098613	0.0	0.0	0.117151	

14980	0.000000	0.25053	0.000000	0.0	0.0	0.000000
14999	0.167088	0.25053	0.098613	0.0	0.0	0.000000

	Death	Drunk	Drugs
1	0.000000	0.057613	0.000000
40	0.000000	0.000000	0.000000
42	0.000000	0.000000	0.000000
67	0.000000	0.000000	0.000000
73	0.000000	0.000000	0.038533
...
14926	0.000000	0.000000	0.000000
14947	0.000000	0.000000	0.000000
14977	0.076709	0.057613	0.038533
14980	0.000000	0.000000	0.000000
14999	0.000000	0.000000	0.038533

[1530 rows x 40 columns], 'Softball':

		Gradyear	Gender	Age	NumberOffriends	Basketball	Football	\
15	4.630399e+98	0.0	1.086951	5.759993	0.000000	0.159361		
32	4.700874e+98	0.0	1.081084	3.914234	0.000000	0.161396		
45	4.560946e+98	0.0	1.096975	5.142291	0.000000	0.000000		
53	4.560946e+98	0.0	1.092219	9.338978	0.161439	0.000000		
75	4.700874e+98	0.0	1.081084	3.058506	0.000000	0.000000		
...		
14931	4.772387e+98	0.0	1.074503	6.671390	0.000000	0.000000		
14943	4.772387e+98	0.0	1.086951	7.150925	0.159402	0.000000		
14962	4.630399e+98	0.0	1.086951	6.671390	0.000000	0.159361		
14971	4.560946e+98	0.0	1.092219	1.957650	0.000000	0.000000		
14977	4.700874e+98	0.0	1.081084	6.941186	0.159402	0.161396		

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	Shopping	\
15	0.000000	0.083816		0.0	0.0	...	0.0	0.0	0.000000
32	0.000000	0.083816		0.0	0.0	...	0.0	0.0	0.000000
45	0.000000	0.083794		0.0	0.0	...	0.0	0.0	0.000000
53	0.000000	0.083815		0.0	0.0	...	0.0	0.0	0.250530
75	0.112097	0.083794		0.0	0.0	...	0.0	0.0	0.000000
...
14931	0.000000	0.083794		0.0	0.0	...	0.0	0.0	0.270074
14943	0.000000	0.083815		0.0	0.0	...	0.0	0.0	0.000000
14962	0.000000	0.083816		0.0	0.0	...	0.0	0.0	0.266961
14971	0.000000	0.083794		0.0	0.0	...	0.0	0.0	0.000000
14977	0.112331	0.083794		0.0	0.0	...	0.0	0.0	0.000000

	Clothes	Hollister	Abercrombie	Die	Death	Drunk	\
15	0.000000	0.000000		0.0	0.000000	0.076709	0.000000
32	0.000000	0.000000		0.0	0.000000	0.000000	0.000000
45	0.000000	0.046088		0.0	0.000000	0.076709	0.000000

53	0.098613	0.000000	0.0	0.000000	0.000000	0.000000
75	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
...
14931	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
14943	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
14962	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
14971	0.000000	0.000000	0.0	0.116845	0.000000	0.000000
14977	0.098613	0.000000	0.0	0.117151	0.076709	0.057613

Drugs

15	0.000000
32	0.000000
45	0.000000
53	0.000000
75	0.000000
...	...
14931	0.000000
14943	0.000000
14962	0.000000
14971	0.000000
14977	0.038533

[1193 rows x 40 columns], 'Volleyball':

			Gradyear	Gender	Age	NumberOffriends	Basketball	Football	\
20	4.700874e+98	0.0	1.086951	0.000000	0.000000	0.000000			
23	4.630399e+98	0.0	1.092219	8.141973	0.159402	0.000000			
47	4.700874e+98	0.0	1.081084	3.791035	0.000000	0.159361			
62	4.630399e+98	0.0	1.092219	5.432670	0.000000	0.159361			
82	4.772387e+98	0.0	1.074503	5.881727	0.000000	0.000000			
...			
14919	4.772387e+98	0.0	1.074503	1.639881	0.000000	0.000000			
14922	4.700874e+98	0.0	1.081084	7.387153	0.000000	0.159361			
14954	4.630399e+98	0.0	1.081084	4.143966	0.000000	0.000000			
14978	4.772387e+98	0.0	1.074503	0.753045	0.159402	0.000000			
14982	4.772387e+98	0.0	1.074503	4.550377	0.159402	0.000000			

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
20	0.000000	0.000000	0.079812	0.085365	...	0.055048	0.164666	
23	0.000000	0.000000	0.079812	0.000000	...	0.000000	0.000000	
47	0.000000	0.000000	0.079798	0.000000	...	0.000000	0.000000	
62	0.000000	0.000000	0.079798	0.000000	...	0.000000	0.000000	
82	0.112332	0.000000	0.079798	0.000000	...	0.055048	0.000000	
...	
14919	0.000000	0.083794	0.079812	0.000000	...	0.000000	0.000000	
14922	0.000000	0.000000	0.079798	0.085340	...	0.000000	0.000000	
14954	0.000000	0.000000	0.079798	0.000000	...	0.000000	0.000000	
14978	0.000000	0.000000	0.079798	0.000000	...	0.000000	0.000000	

14982	0.000000	0.000000	0.079798	0.000000	...	0.055048	0.000000	
	Shopping	Clothes	Hollister	Abercrombie	Die	Death	Drunk	\
20	0.000000	0.098700	0.0	0.000000	0.116845	0.0	0.000000	
23	0.000000	0.000000	0.0	0.034255	0.000000	0.0	0.000000	
47	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.057613	
62	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	
82	0.266961	0.000000	0.0	0.000000	0.000000	0.0	0.000000	
...
14919	0.000000	0.098613	0.0	0.000000	0.000000	0.0	0.000000	
14922	0.250530	0.098613	0.0	0.000000	0.000000	0.0	0.000000	
14954	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	
14978	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	
14982	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	

Drugs

20	0.0
23	0.0
47	0.0
62	0.0
82	0.0
...	...
14919	0.0
14922	0.0
14954	0.0
14978	0.0
14982	0.0

[1207 rows x 40 columns], 'Swimming':

			Gradyear	Gender	Age	NumberOffriends	Basketball	Football	\
1	4.630399e+98	0.0	1.086951	6.428800	0.000000	0.000000			
4	4.700874e+98	0.0	1.081084	0.753045	0.000000	0.000000			
8	4.560946e+98	0.0	1.096975	0.000000	0.000000	0.000000			
20	4.700874e+98	0.0	1.086951	0.000000	0.000000	0.000000			
29	4.772387e+98	0.0	1.074503	4.031683	0.000000	0.000000			
...			
14916	4.700874e+98	0.0	1.086951	0.000000	0.000000	0.000000			
14920	4.560946e+98	0.0	1.092219	4.819488	0.159402	0.159361			
14922	4.700874e+98	0.0	1.081084	7.387153	0.000000	0.159361			
14936	4.560946e+98	0.0	1.096975	0.000000	0.000000	0.159361			
14940	4.700874e+98	0.0	1.081084	0.000000	0.000000	0.000000			

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
1	0.112097	0.0	0.000000	0.085340	...	0.000000	0.000000	
4	0.000000	0.0	0.000000	0.085340	...	0.000000	0.000000	
8	0.000000	0.0	0.000000	0.085340	...	0.000000	0.000000	
20	0.000000	0.0	0.079812	0.085365	...	0.055048	0.164666	

29	0.000000	0.0	0.000000	0.085340	...	0.000000	0.000000
...
14916	0.000000	0.0	0.000000	0.085340	...	0.000000	0.000000
14920	0.000000	0.0	0.000000	0.085340	...	0.000000	0.000000
14922	0.000000	0.0	0.079798	0.085340	...	0.000000	0.000000
14936	0.000000	0.0	0.000000	0.085340	...	0.000000	0.000000
14940	0.000000	0.0	0.000000	0.085340	...	0.000000	0.000000

	Shopping	Clothes	Hollister	Abercrombie	Die	Death	Drunk	\
1	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.057613	
4	0.000000	0.098701	0.0	0.0	0.000000	0.0	0.000000	
8	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.000000	
20	0.000000	0.098700	0.0	0.0	0.116845	0.0	0.000000	
29	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.000000	
...	
14916	0.266961	0.000000	0.0	0.0	0.000000	0.0	0.000000	
14920	0.250530	0.000000	0.0	0.0	0.000000	0.0	0.000000	
14922	0.250530	0.098613	0.0	0.0	0.000000	0.0	0.000000	
14936	0.250530	0.000000	0.0	0.0	0.000000	0.0	0.000000	
14940	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.000000	

	Drugs
1	0.0
4	0.0
8	0.0
20	0.0
29	0.0
...	...
14916	0.0
14920	0.0
14922	0.0
14936	0.0
14940	0.0

[1488 rows x 40 columns], 'Cheerleading':					Gradyear	Gender	Age	NumberOffriends	Basketball	Football	\
23	4.630399e+98	0.0	1.092219		8.141973	0.159402	0.000000				
44	4.772387e+98	0.0	1.074503		3.791035	0.000000	0.000000				
75	4.700874e+98	0.0	1.081084		3.058506	0.000000	0.000000				
90	4.772387e+98	0.0	1.074503		9.077433	0.000000	0.000000				
102	4.630399e+98	0.0	1.086951		3.914234	0.000000	0.159361				
...				
14915	4.772387e+98	0.0	1.074503		5.940858	0.159402	0.159361				
14931	4.772387e+98	0.0	1.074503		6.671390	0.000000	0.000000				
14979	4.700874e+98	0.0	1.086951		7.749401	0.000000	0.000000				
14982	4.772387e+98	0.0	1.074503		4.550377	0.159402	0.000000				
14997	4.630399e+98	0.0	1.086951		5.432670	0.000000	0.000000				

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
23	0.000000	0.000000	0.079812	0.0	...	0.000000	0.0	
44	0.000000	0.000000	0.000000	0.0	...	0.000000	0.0	
75	0.112097	0.083794	0.000000	0.0	...	0.000000	0.0	
90	0.000000	0.000000	0.000000	0.0	...	0.000000	0.0	
102	0.000000	0.000000	0.000000	0.0	...	0.000000	0.0	
...	
14915	0.000000	0.000000	0.000000	0.0	...	0.000000	0.0	
14931	0.000000	0.083794	0.000000	0.0	...	0.000000	0.0	
14979	0.000000	0.000000	0.000000	0.0	...	0.055048	0.0	
14982	0.000000	0.000000	0.079798	0.0	...	0.055048	0.0	
14997	0.000000	0.000000	0.000000	0.0	...	0.000000	0.0	

	Shopping	Clothes	Hollister	Abercrombie	Die	Death	Drunk	Drugs
23	0.000000	0.000000	0.000000	0.034255	0.0	0.0	0.0	0.0
44	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0
75	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0
90	0.000000	0.000000	0.000000	0.034255	0.0	0.0	0.0	0.0
102	0.250530	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0
...
14915	0.250530	0.098613	0.000000	0.000000	0.0	0.0	0.0	0.0
14931	0.270074	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0
14979	0.266961	0.000000	0.046088	0.034255	0.0	0.0	0.0	0.0
14982	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0
14997	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0

	[919 rows x 40 columns], 'Baseball':	Gradyear	Gender	Age	NumberOffriends	Basketball	\
2	4.630399e+98 0.000000 1.086951	5.998896	0.000000				
16	4.700874e+98 0.000000 1.086951	8.015605	0.000000				
21	4.560946e+98 0.000000 1.092219	6.378075	0.000000				
32	4.700874e+98 0.000000 1.081084	3.914234	0.000000				
43	4.630399e+98 0.000000 1.086951	6.763717	0.161439				
...	
14956	4.700874e+98 0.124251 1.081084	7.271002	0.159402				
14967	4.560946e+98 0.124251 1.086951	4.354949	0.000000				
14977	4.700874e+98 0.000000 1.081084	6.941186	0.159402				
14979	4.700874e+98 0.000000 1.086951	7.749401	0.000000				
14996	4.700874e+98 0.000000 1.081084	6.478755	0.000000				

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
2	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	
16	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	
21	0.159361	0.000000	0.000000	0.0	0.0	...	0.000000	
32	0.161396	0.000000	0.083816	0.0	0.0	...	0.000000	
43	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	

...
14956 0.000000 0.000000 0.000000 0.0 0.0 ... 0.000000
14967 0.000000 0.000000 0.000000 0.0 0.0 ... 0.000000
14977 0.161396 0.112331 0.083794 0.0 0.0 ... 0.000000
14979 0.000000 0.000000 0.000000 0.0 0.0 ... 0.055048
14996 0.000000 0.000000 0.000000 0.0 0.0 ... 0.000000

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
2	0.164666	0.000000	0.000000	0.000000	0.000000	0.000000	
16	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
21	0.000000	0.000000	0.000000	0.000000	0.000000	0.116845	
32	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
43	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
...	
14956	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
14967	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
14977	0.000000	0.000000	0.098613	0.000000	0.000000	0.117151	
14979	0.000000	0.266961	0.000000	0.046088	0.034255	0.000000	
14996	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

	Death	Drunk	Drugs
2	0.000000	0.057613	0.038533
16	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000
43	0.000000	0.000000	0.000000
...
14956	0.000000	0.000000	0.000000
14967	0.000000	0.057613	0.000000
14977	0.076709	0.057613	0.038533
14979	0.000000	0.000000	0.000000
14996	0.076709	0.000000	0.000000

[975 rows x 40 columns], 'Tennis':

	Gradyear	Gender	Age	NumberOffriends	Basketball	Football	\
4	4.700874e+98	0.0	1.081084	0.753045	0.000000	0.000000	
9	4.630399e+98	0.0	1.086951	4.550377	0.000000	0.000000	
21	4.560946e+98	0.0	1.092219	6.378075	0.000000	0.159361	
81	4.772387e+98	0.0	1.081084	7.231427	0.000000	0.161546	
135	4.772387e+98	0.0	1.086951	5.432670	0.159402	0.159361	
...	
14898	4.560946e+98	0.0	1.092219	8.353561	0.000000	0.159361	
14930	4.630399e+98	0.0	1.086951	6.527967	0.000000	0.000000	
14977	4.700874e+98	0.0	1.081084	6.941186	0.159402	0.161396	
14983	4.772387e+98	0.0	1.086951	0.000000	0.000000	0.000000	
14984	4.700874e+98	0.0	1.081084	5.142291	0.159402	0.000000	

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	Shopping	\
4	0.000000	0.000000	0.000000	0.08534	...	0.0	0.0	0.000000	
9	0.000000	0.000000	0.000000	0.00000	...	0.0	0.0	0.250530	
21	0.000000	0.000000	0.000000	0.00000	...	0.0	0.0	0.000000	
81	0.000000	0.000000	0.000000	0.00000	...	0.0	0.0	0.000000	
135	0.112097	0.000000	0.079798	0.00000	...	0.0	0.0	0.000000	
...	
14898	0.000000	0.000000	0.000000	0.00000	...	0.0	0.0	0.270074	
14930	0.000000	0.000000	0.000000	0.00000	...	0.0	0.0	0.000000	
14977	0.112331	0.083794	0.000000	0.00000	...	0.0	0.0	0.000000	
14983	0.000000	0.000000	0.000000	0.00000	...	0.0	0.0	0.000000	
14984	0.000000	0.000000	0.000000	0.00000	...	0.0	0.0	0.000000	

	Clothes	Hollister	Abercrombie	Die	Death	Drunk	\
4	0.098701	0.0	0.0	0.000000	0.000000	0.000000	
9	0.098613	0.0	0.0	0.000000	0.000000	0.000000	
21	0.000000	0.0	0.0	0.116845	0.000000	0.000000	
81	0.000000	0.0	0.0	0.000000	0.000000	0.000000	
135	0.000000	0.0	0.0	0.000000	0.000000	0.000000	
...	
14898	0.000000	0.0	0.0	0.000000	0.000000	0.000000	
14930	0.000000	0.0	0.0	0.000000	0.000000	0.000000	
14977	0.098613	0.0	0.0	0.117151	0.076709	0.057613	
14983	0.000000	0.0	0.0	0.000000	0.000000	0.000000	
14984	0.000000	0.0	0.0	0.116845	0.000000	0.000000	

Drugs

4	0.000000
9	0.000000
21	0.000000
81	0.000000
135	0.000000
...	...
14898	0.000000
14930	0.000000
14977	0.038533
14983	0.000000
14984	0.000000

[745 rows x 40 columns], 'Sports':

	Gradyear	Gender	Age	NumberOffriends	Basketball	\
15	4.630399e+98	0.000000	1.086951	5.759993	0.000000	
16	4.700874e+98	0.000000	1.086951	8.015605	0.000000	
20	4.700874e+98	0.000000	1.086951	0.000000	0.000000	
23	4.630399e+98	0.000000	1.092219	8.141973	0.159402	
33	4.772387e+98	0.124251	1.074503	4.143966	0.000000	
...	

14980	4.772387e+98	0.000000	1.086951	3.224451	0.000000
14981	4.630399e+98	0.000000	1.086951	6.220999	0.000000
14982	4.772387e+98	0.000000	1.074503	4.550377	0.159402
14983	4.772387e+98	0.000000	1.086951	0.000000	0.000000
14984	4.700874e+98	0.000000	1.081084	5.142291	0.159402

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
15	0.159361	0.000000	0.083816	0.000000	0.000000	...	0.000000	
16	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
20	0.000000	0.000000	0.000000	0.079812	0.085365	...	0.055048	
23	0.000000	0.000000	0.000000	0.079812	0.000000	...	0.000000	
33	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
...
14980	0.000000	0.112097	0.000000	0.000000	0.000000	...	0.000000	
14981	0.159361	0.000000	0.000000	0.000000	0.000000	...	0.000000	
14982	0.000000	0.000000	0.000000	0.079798	0.000000	...	0.055048	
14983	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
14984	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
15	0.000000	0.00000	0.000000	0.0	0.000000	0.000000	
16	0.000000	0.00000	0.000000	0.0	0.000000	0.000000	
20	0.164666	0.00000	0.098700	0.0	0.000000	0.116845	
23	0.000000	0.00000	0.000000	0.0	0.034255	0.000000	
33	0.000000	0.00000	0.000000	0.0	0.000000	0.000000	
...
14980	0.000000	0.25053	0.000000	0.0	0.000000	0.000000	
14981	0.000000	0.00000	0.098613	0.0	0.000000	0.000000	
14982	0.000000	0.00000	0.000000	0.0	0.000000	0.000000	
14983	0.000000	0.00000	0.000000	0.0	0.000000	0.000000	
14984	0.000000	0.00000	0.000000	0.0	0.000000	0.116845	

	Death	Drunk	Drugs
15	0.076709	0.000000	0.0
16	0.000000	0.000000	0.0
20	0.000000	0.000000	0.0
23	0.000000	0.000000	0.0
33	0.000000	0.000000	0.0
...
14980	0.000000	0.000000	0.0
14981	0.000000	0.057613	0.0
14982	0.000000	0.000000	0.0
14983	0.000000	0.000000	0.0
14984	0.000000	0.000000	0.0

[1682 rows x 40 columns], 'Cute':

Gradyear Gender Age NumberOffriends Basketball Football \

4	4.700874e+98	0.0	1.081084	0.753045	0.000000	0.000000
8	4.560946e+98	0.0	1.096975	0.000000	0.000000	0.000000
12	4.560946e+98	0.0	1.092219	5.217624	0.000000	0.000000
19	4.700874e+98	0.0	1.081084	0.000000	0.000000	0.000000
20	4.700874e+98	0.0	1.086951	0.000000	0.000000	0.000000
...
14979	4.700874e+98	0.0	1.086951	7.749401	0.000000	0.000000
14981	4.630399e+98	0.0	1.086951	6.220999	0.000000	0.159361
14984	4.700874e+98	0.0	1.081084	5.142291	0.159402	0.000000
14991	4.700874e+98	0.0	1.081084	3.058506	0.000000	0.000000
14997	4.630399e+98	0.0	1.086951	5.432670	0.000000	0.000000

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
4	0.0	0.0	0.000000	0.085340	...	0.000000	0.000000	
8	0.0	0.0	0.000000	0.085340	...	0.000000	0.000000	
12	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
19	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
20	0.0	0.0	0.079812	0.085365	...	0.055048	0.164666	
...	
14979	0.0	0.0	0.000000	0.000000	...	0.055048	0.000000	
14981	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
14984	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
14991	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
14997	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	

	Shopping	Clothes	Hollister	Abercrombie	Die	Death	\
4	0.000000	0.098701	0.000000	0.000000	0.000000	0.000000	
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
12	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
19	0.000000	0.000000	0.046088	0.000000	0.000000	0.000000	
20	0.000000	0.098700	0.000000	0.000000	0.116845	0.000000	
...	
14979	0.266961	0.000000	0.046088	0.034255	0.000000	0.000000	
14981	0.000000	0.098613	0.000000	0.000000	0.000000	0.000000	
14984	0.000000	0.000000	0.000000	0.000000	0.116845	0.000000	
14991	0.250530	0.098613	0.000000	0.000000	0.000000	0.076709	
14997	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

	Drunk	Drugs
4	0.000000	0.0
8	0.000000	0.0
12	0.000000	0.0
19	0.000000	0.0
20	0.000000	0.0
...
14979	0.000000	0.0

14981	0.057613	0.0
14984	0.000000	0.0
14991	0.057613	0.0
14997	0.000000	0.0

[3100 rows x 40 columns], 'Sex':

	Gradyear	Gender	Age	NumberOffriends	Basketball	Football	\	
2	4.630399e+98	0.0	1.086951	5.998896	0.000000	0.000000		
8	4.560946e+98	0.0	1.096975	0.000000	0.000000	0.000000		
13	4.560946e+98	0.0	1.086951	0.000000	0.000000	0.000000		
29	4.772387e+98	0.0	1.074503	4.031683	0.000000	0.000000		
36	4.700874e+98	0.0	1.086951	6.428800	0.000000	0.000000		
...		
14970	4.772387e+98	0.0	1.074503	1.957650	0.000000	0.000000		
14971	4.560946e+98	0.0	1.092219	1.957650	0.000000	0.000000		
14977	4.700874e+98	0.0	1.081084	6.941186	0.159402	0.161396		
14984	4.700874e+98	0.0	1.081084	5.142291	0.159402	0.000000		
14991	4.700874e+98	0.0	1.081084	3.058506	0.000000	0.000000		
	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
2	0.000000	0.000000		0.0	0.00000	...	0.0	0.164666
8	0.000000	0.000000		0.0	0.08534	...	0.0	0.000000
13	0.000000	0.000000		0.0	0.00000	...	0.0	0.000000
29	0.000000	0.000000		0.0	0.08534	...	0.0	0.000000
36	0.000000	0.000000		0.0	0.00000	...	0.0	0.164666
...	
14970	0.000000	0.000000		0.0	0.00000	...	0.0	0.000000
14971	0.000000	0.083794		0.0	0.00000	...	0.0	0.000000
14977	0.112331	0.083794		0.0	0.00000	...	0.0	0.000000
14984	0.000000	0.000000		0.0	0.00000	...	0.0	0.000000
14991	0.000000	0.000000		0.0	0.00000	...	0.0	0.000000
	Shopping	Clothes	Hollister	Abercrombie	Die	Death	\	
2	0.00000	0.000000		0.0	0.0	0.000000	0.000000	
8	0.00000	0.000000		0.0	0.0	0.000000	0.000000	
13	0.00000	0.000000		0.0	0.0	0.000000	0.000000	
29	0.00000	0.000000		0.0	0.0	0.000000	0.000000	
36	0.00000	0.098613		0.0	0.0	0.000000	0.000000	
...	
14970	0.00000	0.000000		0.0	0.0	0.000000	0.000000	
14971	0.00000	0.000000		0.0	0.0	0.116845	0.000000	
14977	0.00000	0.098613		0.0	0.0	0.117151	0.076709	
14984	0.00000	0.000000		0.0	0.0	0.116845	0.000000	
14991	0.25053	0.098613		0.0	0.0	0.000000	0.076709	
	Drunk	Drugs						
2	0.057613	0.038533						

```

8    0.000000  0.000000
13   0.000000  0.000000
29    0.000000  0.000000
36    0.000000  0.000000
...
...
14970  0.000000  0.000000
14971  0.000000  0.000000
14977  0.057613  0.038533
14984  0.000000  0.000000
14991  0.057613  0.000000

```

[1584 rows x 40 columns], 'Sexy':

	Gradyear	Gender	Age	NumberOffriends	Basketball	\
4	4.700874e+98	0.000000	1.081084	0.753045	0.000000	
6	4.560946e+98	0.124251	1.092219	4.251580	0.000000	
10	4.772387e+98	0.000000	1.086951	0.000000	0.000000	
20	4.700874e+98	0.000000	1.086951	0.000000	0.000000	
25	4.700874e+98	0.000000	1.081084	7.109973	0.000000	
...	
14951	4.630399e+98	0.000000	1.086951	0.000000	0.000000	
14952	4.700874e+98	0.000000	1.081084	0.000000	0.000000	
14970	4.772387e+98	0.000000	1.074503	1.957650	0.000000	
14976	4.630399e+98	0.000000	1.086951	5.064946	0.000000	
14984	4.700874e+98	0.000000	1.081084	5.142291	0.159402	

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
4	0.0	0.0	0.0	0.000000	0.085340	...	0.000000	
6	0.0	0.0	0.0	0.000000	0.000000	...	0.000000	
10	0.0	0.0	0.0	0.000000	0.000000	...	0.000000	
20	0.0	0.0	0.0	0.079812	0.085365	...	0.055048	
25	0.0	0.0	0.0	0.000000	0.000000	...	0.000000	
...	
14951	0.0	0.0	0.0	0.000000	0.000000	...	0.000000	
14952	0.0	0.0	0.0	0.000000	0.000000	...	0.055048	
14970	0.0	0.0	0.0	0.000000	0.000000	...	0.000000	
14976	0.0	0.0	0.0	0.000000	0.000000	...	0.000000	
14984	0.0	0.0	0.0	0.000000	0.000000	...	0.000000	

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
4	0.000000	0.00000	0.098701	0.0	0.0	0.000000	
6	0.000000	0.00000	0.098613	0.0	0.0	0.000000	
10	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
20	0.164666	0.00000	0.098700	0.0	0.0	0.116845	
25	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
...	
14951	0.000000	0.25053	0.000000	0.0	0.0	0.000000	
14952	0.167088	0.00000	0.098613	0.0	0.0	0.116845	

14970	0.000000	0.00000	0.000000	0.0	0.0	0.000000
14976	0.164666	0.00000	0.000000	0.0	0.0	0.116845
14984	0.000000	0.00000	0.000000	0.0	0.0	0.116845

	Death	Drunk	Drugs
4	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000
25	0.000000	0.057613	0.000000
...
14951	0.076709	0.000000	0.000000
14952	0.000000	0.057613	0.038533
14970	0.000000	0.000000	0.000000
14976	0.000000	0.000000	0.000000
14984	0.000000	0.000000	0.000000

[1511 rows x 40 columns], 'Hot':

	Gradyear	Gender	Age	NumberOffriends	Basketball	\
19	4.700874e+98	0.000000	1.081084	0.000000	0.000000	
21	4.560946e+98	0.000000	1.092219	6.378075	0.000000	
23	4.630399e+98	0.000000	1.092219	8.141973	0.159402	
29	4.772387e+98	0.000000	1.074503	4.031683	0.000000	
33	4.772387e+98	0.124251	1.074503	4.143966	0.000000	
...	
14946	4.630399e+98	0.000000	1.086951	4.454441	0.000000	
14952	4.700874e+98	0.000000	1.081084	0.000000	0.000000	
14957	4.700874e+98	0.000000	1.081084	8.554224	0.000000	
14959	4.700874e+98	0.124251	1.081084	6.326553	0.000000	
14966	4.700874e+98	0.000000	1.081084	4.643036	0.000000	

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
19	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	
21	0.159361	0.0	0.0	0.000000	0.00000	...	0.000000	
23	0.000000	0.0	0.0	0.079812	0.00000	...	0.000000	
29	0.000000	0.0	0.0	0.000000	0.08534	...	0.000000	
33	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	
...	
14946	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	
14952	0.000000	0.0	0.0	0.000000	0.00000	...	0.055048	
14957	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	
14959	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	
14966	0.159361	0.0	0.0	0.000000	0.00000	...	0.000000	

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	Death	\
19	0.000000	0.0	0.000000	0.046088	0.000000	0.000000	0.0	
21	0.000000	0.0	0.000000	0.000000	0.000000	0.116845	0.0	

23	0.000000	0.0	0.000000	0.000000	0.034255	0.000000	0.0
29	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0
33	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0
...
14946	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0
14952	0.167088	0.0	0.098613	0.000000	0.000000	0.116845	0.0
14957	0.000000	0.0	0.000000	0.046088	0.034255	0.000000	0.0
14959	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0
14966	0.000000	0.0	0.000000	0.000000	0.000000	0.117160	0.0

	Drunk	Drugs
19	0.000000	0.000000
21	0.000000	0.000000
23	0.000000	0.000000
29	0.000000	0.000000
33	0.000000	0.000000
...
14946	0.000000	0.000000
14952	0.057613	0.038533
14957	0.000000	0.000000
14959	0.000000	0.000000
14966	0.000000	0.000000

[1381 rows x 40 columns], 'Kissed':

	Gradyear	Gender	Age	NumberOffriends	Basketball	\
13	4.560946e+98	0.000000	1.086951	0.000000	0.000000	
38	4.560946e+98	0.000000	1.092219	9.858907	0.159402	
41	4.560946e+98	0.000000	1.092219	4.550377	0.000000	
73	4.630399e+98	0.000000	1.086951	3.914234	0.159402	
83	4.560946e+98	0.124251	1.096975	3.524466	0.000000	
...	
14909	4.630399e+98	0.000000	1.086951	6.527967	0.000000	
14911	4.700874e+98	0.000000	1.081084	4.643036	0.000000	
14915	4.772387e+98	0.000000	1.074503	5.940858	0.159402	
14973	4.700874e+98	0.000000	1.081084	3.791035	0.000000	
14977	4.700874e+98	0.000000	1.081084	6.941186	0.159402	

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
13	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	
38	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	
41	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	
73	0.159361	0.112332	0.000000	0.0	0.0	...	0.055048	
83	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	
...	
14909	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	
14911	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000	
14915	0.159361	0.000000	0.000000	0.0	0.0	...	0.000000	

14973	0.000000	0.000000	0.000000	0.0	0.0	...	0.000000
14977	0.161396	0.112331	0.083794	0.0	0.0	...	0.000000

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
13	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
38	0.164666	0.25053	0.098613	0.0	0.0	0.000000	
41	0.000000	0.00000	0.098613	0.0	0.0	0.000000	
73	0.164666	0.00000	0.098613	0.0	0.0	0.000000	
83	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
...	
14909	0.000000	0.00000	0.098700	0.0	0.0	0.116845	
14911	0.000000	0.00000	0.000000	0.0	0.0	0.117151	
14915	0.000000	0.25053	0.098613	0.0	0.0	0.000000	
14973	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
14977	0.000000	0.00000	0.098613	0.0	0.0	0.117151	

	Death	Drunk	Drugs
13	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000000
41	0.076709	0.000000	0.000000
73	0.000000	0.000000	0.038533
83	0.000000	0.000000	0.000000
...
14909	0.000000	0.000000	0.000000
14911	0.076709	0.000000	0.000000
14915	0.000000	0.000000	0.000000
14973	0.000000	0.000000	0.000000
14977	0.076709	0.057613	0.038533

[938 rows x 40 columns], 'Dance':	Gradyear	Gender	Age	NumberOffriends	Basketball	Football	\
8	4.560946e+98	0.0	1.096975	0.000000	0.000000	0.000000	
10	4.772387e+98	0.0	1.086951	0.000000	0.000000	0.000000	
16	4.700874e+98	0.0	1.086951	8.015605	0.000000	0.000000	
20	4.700874e+98	0.0	1.086951	0.000000	0.000000	0.000000	
26	4.560946e+98	0.0	1.086951	0.000000	0.000000	0.000000	
...	
14957	4.700874e+98	0.0	1.081084	8.554224	0.000000	0.000000	
14975	4.630399e+98	0.0	1.086951	5.064946	0.000000	0.000000	
14984	4.700874e+98	0.0	1.081084	5.142291	0.159402	0.000000	
14985	4.560946e+98	0.0	1.092219	5.940858	0.000000	0.159361	
14996	4.700874e+98	0.0	1.081084	6.478755	0.000000	0.000000	

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
8	0.0	0.0	0.000000	0.085340	...	0.000000	0.000000	
10	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
16	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	

20	0.0	0.0	0.079812	0.085365	...	0.055048	0.164666
26	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
...
14957	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
14975	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
14984	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
14985	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
14996	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000

	Shopping	Clothes	Hollister	Abercrombie	Die	Death	Drunk	\
8	0.00000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
10	0.00000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
16	0.00000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
20	0.00000	0.0987	0.000000	0.000000	0.116845	0.000000	0.000000	0.0
26	0.00000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
...
14957	0.00000	0.0000	0.046088	0.034255	0.000000	0.000000	0.000000	0.0
14975	0.00000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
14984	0.00000	0.0000	0.000000	0.000000	0.116845	0.000000	0.000000	0.0
14985	0.25053	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
14996	0.00000	0.0000	0.000000	0.000000	0.000000	0.076709	0.000000	0.0

	Drugs
8	0.0
10	0.0
16	0.0
20	0.0
26	0.0
...	...
14957	0.0
14975	0.0
14984	0.0
14985	0.0
14996	0.0

[3310 rows x 40 columns], 'Band':	Gradyear	Gender	Age	NumberOfffriends	Basketball	\
5	4.700874e+98	0.124251	1.092219	5.432670	0.000000	
21	4.560946e+98	0.000000	1.092219	6.378075	0.000000	
25	4.700874e+98	0.000000	1.081084	7.109973	0.000000	
34	4.630399e+98	0.000000	1.086951	5.881727	0.000000	
38	4.560946e+98	0.000000	1.092219	9.858907	0.159402	
...	
14949	4.772387e+98	0.000000	1.074503	0.753045	0.161439	
14954	4.630399e+98	0.000000	1.081084	4.143966	0.000000	
14977	4.700874e+98	0.000000	1.081084	6.941186	0.159402	
14990	4.772387e+98	0.124251	1.081084	4.354949	0.159402	

14999	4.772387e+98	0.000000	1.074503		4.903694	0.000000		
	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
5	0.161574	0.000000	0.000000	0.000000	0.0	...	0.0	
21	0.159361	0.000000	0.000000	0.000000	0.0	...	0.0	
25	0.000000	0.000000	0.000000	0.000000	0.0	...	0.0	
34	0.161396	0.000000	0.000000	0.000000	0.0	...	0.0	
38	0.000000	0.000000	0.000000	0.000000	0.0	...	0.0	
...	
14949	0.000000	0.000000	0.000000	0.000000	0.0	...	0.0	
14954	0.000000	0.000000	0.000000	0.079798	0.0	...	0.0	
14977	0.161396	0.112331	0.083794	0.000000	0.0	...	0.0	
14990	0.000000	0.000000	0.000000	0.000000	0.0	...	0.0	
14999	0.000000	0.112332	0.000000	0.000000	0.0	...	0.0	

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	V
5	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
21	0.000000	0.00000	0.000000	0.0	0.0	0.116845	
25	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
34	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
38	0.164666	0.25053	0.098613	0.0	0.0	0.000000	
...
14949	0.164666	0.00000	0.000000	0.0	0.0	0.000000	
14954	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
14977	0.000000	0.00000	0.098613	0.0	0.0	0.117151	
14990	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
14999	0.167088	0.25053	0.098613	0.0	0.0	0.000000	

	Death	Drunk	Drugs
5	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000
25	0.000000	0.057613	0.000000
34	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000000
...
14949	0.000000	0.000000	0.000000
14954	0.000000	0.000000	0.000000
14977	0.076709	0.057613	0.038533
14990	0.000000	0.000000	0.000000
14999	0.000000	0.000000	0.038533

```
[2204 rows x 40 columns], 'Marching':           Gradyear   Gender      Age NumberOffriends Basketball
25    4.700874e+98  0.000000  1.081084    7.109973  0.000000
90    4.772387e+98  0.000000  1.074503    9.077433  0.000000
138   4.630399e+98  0.000000  1.086951    3.379240  0.000000
149   4.630399e+98  0.000000  1.086951    4.732667  0.000000
```

168	4.630399e+98	0.124251	1.086951	0.000000	0.000000
...
14840	4.560946e+98	0.000000	1.086951	5.759993	0.000000
14876	4.630399e+98	0.124251	1.086951	0.000000	0.000000
14918	4.560946e+98	0.000000	1.096975	7.851461	0.000000
14944	4.560946e+98	0.000000	1.092219	3.524466	0.159402
14954	4.630399e+98	0.000000	1.081084	4.143966	0.000000

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
25	0.000000	0.0	0.0	0.000000	0.0	...	0.0	
90	0.000000	0.0	0.0	0.000000	0.0	...	0.0	
138	0.000000	0.0	0.0	0.000000	0.0	...	0.0	
149	0.000000	0.0	0.0	0.000000	0.0	...	0.0	
168	0.000000	0.0	0.0	0.000000	0.0	...	0.0	
...	
14840	0.161546	0.0	0.0	0.000000	0.0	...	0.0	
14876	0.000000	0.0	0.0	0.000000	0.0	...	0.0	
14918	0.000000	0.0	0.0	0.000000	0.0	...	0.0	
14944	0.000000	0.0	0.0	0.000000	0.0	...	0.0	
14954	0.000000	0.0	0.0	0.079798	0.0	...	0.0	

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	Death	\
25	0.000000	0.00000	0.0	0.0	0.000000	0.0	0.0	
90	0.000000	0.00000	0.0	0.0	0.034255	0.0	0.0	
138	0.000000	0.00000	0.0	0.0	0.000000	0.0	0.0	
149	0.000000	0.00000	0.0	0.0	0.000000	0.0	0.0	
168	0.000000	0.25053	0.0	0.0	0.000000	0.0	0.0	
...	
14840	0.164666	0.00000	0.0	0.0	0.000000	0.0	0.0	
14876	0.000000	0.00000	0.0	0.0	0.000000	0.0	0.0	
14918	0.000000	0.00000	0.0	0.0	0.000000	0.0	0.0	
14944	0.000000	0.00000	0.0	0.0	0.000000	0.0	0.0	
14954	0.000000	0.00000	0.0	0.0	0.000000	0.0	0.0	

	Drunk	Drugs
25	0.057613	0.0
90	0.000000	0.0
138	0.000000	0.0
149	0.000000	0.0
168	0.000000	0.0
...
14840	0.000000	0.0
14876	0.000000	0.0
14918	0.000000	0.0
14944	0.000000	0.0
14954	0.000000	0.0

[432 rows x 40 columns], 'Music': Empty DataFrame

Columns: [Gradyear, Gender, Age, NumberOffriends, Basketball, Football, Soccer, Softball, Volleyball, Swimming, Cheerleading, Baseball, Tennis, Sports, Cute, Sex, Sexy, Hot, Kissed, Dance, Band, Marching, Music, Rock, God, Church, Jesus, Bible, Hair, Dress, Blonde, Mall, Shopping, Clothes, Hollister, Abercrombie, Die, Death, Drunk, Drugs]

Index: []

	'Rock':							
	Gradyear	Gender	Age	NumberOffriends	Basketball	Volleyball	Swimming	Cheerleading
2	4.630399e+98	0.000000	1.086951	5.998896	0.000000	0.000000	0.000000	0.000000
5	4.700874e+98	0.124251	1.092219	5.432670	0.000000	0.000000	0.000000	0.000000
19	4.700874e+98	0.000000	1.081084	0.000000	0.000000	0.000000	0.000000	0.000000
20	4.700874e+98	0.000000	1.086951	0.000000	0.000000	0.000000	0.000000	0.000000
23	4.630399e+98	0.000000	1.092219	8.141973	0.159402	0.000000	0.000000	0.000000
...
14984	4.700874e+98	0.000000	1.081084	5.142291	0.159402	0.000000	0.000000	0.000000
14985	4.560946e+98	0.000000	1.092219	5.940858	0.000000	0.000000	0.000000	0.000000
14987	4.560946e+98	0.124251	1.086951	2.879353	0.000000	0.000000	0.000000	0.000000
14990	4.772387e+98	0.124251	1.081084	4.354949	0.159402	0.000000	0.000000	0.000000
14995	4.700874e+98	0.000000	1.081084	4.550377	0.000000	0.000000	0.000000	0.000000
	Football Soccer Softball Volleyball Swimming ... Blonde \							
2	0.000000	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
5	0.161574	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
19	0.000000	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
20	0.000000	0.0	0.0	0.079812	0.085365	...	0.055048	0.000000
23	0.000000	0.0	0.0	0.079812	0.000000	...	0.000000	0.000000
...
14984	0.000000	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
14985	0.159361	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
14987	0.000000	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
14990	0.000000	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
14995	0.000000	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000
	Mall Shopping Clothes Hollister Abercrombie Die \							
2	0.164666	0.00000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.00000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.00000	0.0000	0.046088	0.000000	0.000000	0.000000	0.000000
20	0.164666	0.00000	0.0987	0.000000	0.000000	0.000000	0.116845	0.000000
23	0.000000	0.00000	0.0000	0.000000	0.034255	0.000000	0.000000	0.000000
...
14984	0.000000	0.00000	0.0000	0.000000	0.000000	0.000000	0.116845	0.000000
14985	0.000000	0.25053	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
14987	0.000000	0.00000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
14990	0.000000	0.00000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
14995	0.000000	0.00000	0.0987	0.000000	0.000000	0.000000	0.000000	0.000000

	Death	Drunk	Drugs
2	0.000000	0.057613	0.038533
5	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000
...
14984	0.000000	0.000000	0.000000
14985	0.000000	0.000000	0.000000
14987	0.000000	0.000000	0.000000
14990	0.000000	0.000000	0.000000
14995	0.076709	0.000000	0.000000

[2446 rows x 40 columns], 'God': Empty DataFrame

Columns: [Gradyear, Gender, Age, NumberOffriends, Basketball, Football, Soccer, Softball, Volleyball, Swimming, Cheerleading, Baseball, Tennis, Sports, Cute, Sex, Sexy, Hot, Kissed, Dance, Band, Marching, Music, Rock, God, Church, Jesus, Bible, Hair, Dress, Blonde, Mall, Shopping, Clothes, Hollister, Abercrombie, Die, Death, Drunk, Drugs]

Index: []

	Gradyear Gender Age NumberOfffriends Basketball \						
2	4.630399e+98	0.000000	1.086951	5.998896	0.000000		
4	4.700874e+98	0.000000	1.081084	0.753045	0.000000		
6	4.560946e+98	0.124251	1.092219	4.251580	0.000000		
8	4.560946e+98	0.000000	1.096975	0.000000	0.000000		
16	4.700874e+98	0.000000	1.086951	8.015605	0.000000		
...		
14966	4.700874e+98	0.000000	1.081084	4.643036	0.000000		
14977	4.700874e+98	0.000000	1.081084	6.941186	0.159402		
14983	4.772387e+98	0.000000	1.086951	0.000000	0.000000		
14990	4.772387e+98	0.124251	1.081084	4.354949	0.159402		
14996	4.700874e+98	0.000000	1.081084	6.478755	0.000000		

	Football Soccer Softball Volleyball Swimming ... Blonde \						
2	0.000000	0.000000	0.000000	0.0	0.00000	...	0.0
4	0.000000	0.000000	0.000000	0.0	0.08534	...	0.0
6	0.000000	0.000000	0.000000	0.0	0.00000	...	0.0
8	0.000000	0.000000	0.000000	0.0	0.08534	...	0.0
16	0.000000	0.000000	0.000000	0.0	0.00000	...	0.0
...
14966	0.159361	0.000000	0.000000	0.0	0.00000	...	0.0
14977	0.161396	0.112331	0.083794	0.0	0.00000	...	0.0
14983	0.000000	0.000000	0.000000	0.0	0.00000	...	0.0
14990	0.000000	0.000000	0.000000	0.0	0.00000	...	0.0
14996	0.000000	0.000000	0.000000	0.0	0.00000	...	0.0

Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
------	----------	---------	-----------	-------------	-----	---

2	0.164666	0.0	0.000000	0.0	0.0	0.000000
4	0.000000	0.0	0.098701	0.0	0.0	0.000000
6	0.000000	0.0	0.098613	0.0	0.0	0.000000
8	0.000000	0.0	0.000000	0.0	0.0	0.000000
16	0.000000	0.0	0.000000	0.0	0.0	0.000000
...
14966	0.000000	0.0	0.000000	0.0	0.0	0.117160
14977	0.000000	0.0	0.098613	0.0	0.0	0.117151
14983	0.000000	0.0	0.000000	0.0	0.0	0.000000
14990	0.000000	0.0	0.000000	0.0	0.0	0.000000
14996	0.000000	0.0	0.000000	0.0	0.0	0.000000

	Death	Drunk	Drugs
2	0.000000	0.057613	0.038533
4	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000
...
14966	0.000000	0.000000	0.000000
14977	0.076709	0.057613	0.038533
14983	0.000000	0.000000	0.000000
14990	0.000000	0.000000	0.000000
14996	0.076709	0.000000	0.000000

[2100 rows x 40 columns], 'Jesus':

	Gradyear	Gender	Age	NumberOffriends	Basketball	\
18	4.700874e+98	0.000000	1.081084	3.224451	0.000000	
21	4.560946e+98	0.000000	1.092219	6.378075	0.000000	
57	4.772387e+98	0.000000	1.074503	4.031683	0.000000	
58	4.630399e+98	0.000000	1.086951	7.425048	0.159402	
60	4.560946e+98	0.000000	1.092219	4.819488	0.000000	
...	
14929	4.560946e+98	0.000000	1.092219	3.914234	0.000000	
14943	4.772387e+98	0.000000	1.086951	7.150925	0.159402	
14950	4.700874e+98	0.000000	1.081084	0.000000	0.000000	
14968	4.630399e+98	0.124251	1.086951	6.624262	0.000000	
14995	4.700874e+98	0.000000	1.081084	4.550377	0.000000	

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
18	0.000000	0.0	0.000000	0.0	0.0	...	0.0	
21	0.159361	0.0	0.000000	0.0	0.0	...	0.0	
57	0.000000	0.0	0.000000	0.0	0.0	...	0.0	
58	0.000000	0.0	0.000000	0.0	0.0	...	0.0	
60	0.000000	0.0	0.000000	0.0	0.0	...	0.0	
...	
14929	0.000000	0.0	0.000000	0.0	0.0	...	0.0	

14943	0.000000	0.0	0.083815	0.0	0.0	...	0.0
14950	0.000000	0.0	0.000000	0.0	0.0	...	0.0
14968	0.000000	0.0	0.000000	0.0	0.0	...	0.0
14995	0.000000	0.0	0.000000	0.0	0.0	...	0.0

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
18	0.000000	0.25053	0.000000	0.0	0.0	0.000000	
21	0.000000	0.00000	0.000000	0.0	0.0	0.116845	
57	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
58	0.000000	0.25053	0.000000	0.0	0.0	0.000000	
60	0.000000	0.25053	0.098613	0.0	0.0	0.000000	
...	
14929	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
14943	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
14950	0.000000	0.25053	0.000000	0.0	0.0	0.000000	
14968	0.167088	0.25053	0.000000	0.0	0.0	0.000000	
14995	0.000000	0.00000	0.098700	0.0	0.0	0.000000	

	Death	Drunk	Drugs
18	0.000000	0.0	0.0
21	0.000000	0.0	0.0
57	0.000000	0.0	0.0
58	0.000000	0.0	0.0
60	0.000000	0.0	0.0
...
14929	0.000000	0.0	0.0
14943	0.000000	0.0	0.0
14950	0.000000	0.0	0.0
14968	0.076709	0.0	0.0
14995	0.076709	0.0	0.0

[1131 rows x 40 columns], 'Bible':	Gradyear	Gender	Age	NumberOffriends	Basketball	\
297	4.700874e+98	0.124251	1.081084	14.850032	0.000000	
321	4.630399e+98	0.000000	1.086951	5.881727	0.000000	
358	4.630399e+98	0.000000	1.086951	6.717868	0.000000	
402	4.630399e+98	0.124251	1.092219	6.428800	0.000000	
499	4.700874e+98	0.000000	1.081084	0.000000	0.159402	
...	
14596	4.772387e+98	0.124251	1.081084	4.985461	0.161619	
14605	4.560946e+98	0.124251	1.092219	5.881727	0.000000	
14651	4.560946e+98	0.124251	1.092219	4.903694	0.000000	
14771	4.630399e+98	0.000000	1.081084	2.684279	0.000000	
14774	4.630399e+98	0.124251	1.086951	0.000000	0.000000	

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
297	0.000000	0.000000	0.000000	0.0	0.00000	...	0.055048	

321	0.000000	0.000000	0.000000	0.0	0.00000	...	0.000000
358	0.000000	0.000000	0.000000	0.0	0.08534	...	0.000000
402	0.000000	0.000000	0.000000	0.0	0.00000	...	0.000000
499	0.000000	0.000000	0.083794	0.0	0.00000	...	0.000000
...
14596	0.159361	0.000000	0.000000	0.0	0.00000	...	0.000000
14605	0.000000	0.000000	0.000000	0.0	0.00000	...	0.000000
14651	0.159361	0.112331	0.000000	0.0	0.00000	...	0.000000
14771	0.000000	0.000000	0.000000	0.0	0.00000	...	0.000000
14774	0.000000	0.000000	0.000000	0.0	0.00000	...	0.000000

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	Death	\
297	0.0	0.00000	0.000000	0.0	0.0	0.117151	0.076709	
321	0.0	0.25053	0.000000	0.0	0.0	0.000000	0.000000	
358	0.0	0.25053	0.098613	0.0	0.0	0.000000	0.000000	
402	0.0	0.00000	0.000000	0.0	0.0	0.000000	0.000000	
499	0.0	0.00000	0.000000	0.0	0.0	0.000000	0.000000	
...
14596	0.0	0.00000	0.098613	0.0	0.0	0.000000	0.076709	
14605	0.0	0.00000	0.000000	0.0	0.0	0.000000	0.000000	
14651	0.0	0.00000	0.000000	0.0	0.0	0.000000	0.000000	
14771	0.0	0.00000	0.000000	0.0	0.0	0.000000	0.000000	
14774	0.0	0.00000	0.000000	0.0	0.0	0.000000	0.000000	

	Drunk	Drugs
297	0.057613	0.000000
321	0.000000	0.038533
358	0.000000	0.000000
402	0.000000	0.000000
499	0.000000	0.000000
...
14596	0.000000	0.000000
14605	0.000000	0.000000
14651	0.000000	0.000000
14771	0.000000	0.000000
14774	0.000000	0.000000

[230 rows x 40 columns], 'Hair':

2	4.630399e+98	0.0	1.086951	
4	4.700874e+98	0.0	1.081084	
10	4.772387e+98	0.0	1.086951	
20	4.700874e+98	0.0	1.086951	
21	4.560946e+98	0.0	1.092219	
...
14982	4.772387e+98	0.0	1.074503	
14984	4.700874e+98	0.0	1.081084	

Gradyear Gender Age NumberOffriends Basketball Football \

5.998896	0.000000	0.000000
0.753045	0.000000	0.000000
0.000000	0.000000	0.000000
0.000000	0.000000	0.000000
6.378075	0.000000	0.159361
...
4.550377	0.159402	0.000000
5.142291	0.159402	0.000000

14988	4.772387e+98	0.0	1.074503	8.798106	0.000000	0.161396
14991	4.700874e+98	0.0	1.081084	3.058506	0.000000	0.000000
14995	4.700874e+98	0.0	1.081084	4.550377	0.000000	0.000000

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
2	0.0	0.0	0.000000	0.000000	...	0.000000	0.164666	
4	0.0	0.0	0.000000	0.085340	...	0.000000	0.000000	
10	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
20	0.0	0.0	0.079812	0.085365	...	0.055048	0.164666	
21	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
...	
14982	0.0	0.0	0.079798	0.000000	...	0.055048	0.000000	
14984	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
14988	0.0	0.0	0.000000	0.000000	...	0.000000	0.167320	
14991	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	
14995	0.0	0.0	0.000000	0.000000	...	0.000000	0.000000	

	Shopping	Clothes	Hollister	Abercrombie	Die	Death	\
2	0.00000	0.000000	0.0	0.0	0.000000	0.000000	
4	0.00000	0.098701	0.0	0.0	0.000000	0.000000	
10	0.00000	0.000000	0.0	0.0	0.000000	0.000000	
20	0.00000	0.098700	0.0	0.0	0.116845	0.000000	
21	0.00000	0.000000	0.0	0.0	0.116845	0.000000	
...	
14982	0.00000	0.000000	0.0	0.0	0.000000	0.000000	
14984	0.00000	0.000000	0.0	0.0	0.116845	0.000000	
14988	0.00000	0.000000	0.0	0.0	0.000000	0.000000	
14991	0.25053	0.098613	0.0	0.0	0.000000	0.076709	
14995	0.00000	0.098700	0.0	0.0	0.000000	0.076709	

	Drunk	Drugs
2	0.057613	0.038533
4	0.000000	0.000000
10	0.000000	0.000000
20	0.000000	0.000000
21	0.000000	0.000000
...
14982	0.000000	0.000000
14984	0.000000	0.000000
14988	0.000000	0.000000
14991	0.057613	0.000000
14995	0.000000	0.000000

[3405 rows x 40 columns], 'Dress':	Gradyear	Gender	Age	NumberOfffriends	Basketball	Football	\
22	4.560946e+98	0.0	1.092219	7.917981	0.000000	0.000000	
43	4.630399e+98	0.0	1.086951	6.763717	0.161439	0.000000	

53	4.560946e+98	0.0	1.092219	9.338978	0.161439	0.000000
68	4.560946e+98	0.0	1.092219	5.142291	0.000000	0.159361
77	4.630399e+98	0.0	1.086951	5.362712	0.000000	0.161546
...
14935	4.560946e+98	0.0	1.092219	5.633260	0.000000	0.000000
14941	4.772387e+98	0.0	1.074503	7.068537	0.000000	0.161396
14951	4.630399e+98	0.0	1.086951	0.000000	0.000000	0.000000
14979	4.700874e+98	0.0	1.086951	7.749401	0.000000	0.000000
14993	4.630399e+98	0.0	1.081084	0.000000	0.000000	0.000000

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
22	0.0	0.000000		0.0	0.0	...	0.000000	0.164666
43	0.0	0.000000		0.0	0.0	...	0.000000	0.000000
53	0.0	0.083815		0.0	0.0	...	0.000000	0.000000
68	0.0	0.000000		0.0	0.0	...	0.000000	0.000000
77	0.0	0.000000		0.0	0.0	...	0.000000	0.000000
...
14935	0.0	0.000000		0.0	0.0	...	0.000000	0.000000
14941	0.0	0.000000		0.0	0.0	...	0.000000	0.164666
14951	0.0	0.000000		0.0	0.0	...	0.000000	0.000000
14979	0.0	0.000000		0.0	0.0	...	0.055048	0.000000
14993	0.0	0.000000		0.0	0.0	...	0.000000	0.000000

	Shopping	Clothes	Hollister	Abercrombie	Die	Death	\
22	0.266961	0.098613	0.000000	0.000000	0.117151	0.000000	
43	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
53	0.250530	0.098613	0.000000	0.000000	0.000000	0.000000	
68	0.266961	0.000000	0.000000	0.000000	0.000000	0.000000	
77	0.250530	0.000000	0.000000	0.000000	0.000000	0.000000	
...
14935	0.000000	0.000000	0.000000	0.034255	0.000000	0.000000	
14941	0.266961	0.000000	0.000000	0.000000	0.116845	0.000000	
14951	0.250530	0.000000	0.000000	0.000000	0.000000	0.076709	
14979	0.266961	0.000000	0.046088	0.034255	0.000000	0.000000	
14993	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

	Drunk	Drugs
22	0.000000	0.0
43	0.000000	0.0
53	0.000000	0.0
68	0.000000	0.0
77	0.000000	0.0
...
14935	0.000000	0.0
14941	0.057613	0.0
14951	0.000000	0.0

14979	0.000000	0.0
14993	0.000000	0.0

[1214 rows x 40 columns], 'Blonde':				Gradyear	Gender	Age	NumberOffriends	Basketball	\
20	4.700874e+98	0.000000	1.086951	0.000000	0.000000				
73	4.630399e+98	0.000000	1.086951	3.914234	0.159402				
82	4.772387e+98	0.000000	1.074503	5.881727	0.000000				
99	4.700874e+98	0.000000	1.086951	4.732667	0.161612				
104	4.700874e+98	0.000000	1.081084	0.000000	0.000000				
...				
14904	4.560946e+98	0.000000	1.096975	3.224451	0.000000				
14939	4.700874e+98	0.124251	1.081084	3.058506	0.000000				
14952	4.700874e+98	0.000000	1.081084	0.000000	0.000000				
14979	4.700874e+98	0.000000	1.086951	7.749401	0.000000				
14982	4.772387e+98	0.000000	1.074503	4.550377	0.159402				
Football Soccer Softball Volleyball Swimming ... Blonde \									
20	0.000000	0.000000	0.0	0.079812	0.085365	...	0.055048		
73	0.159361	0.112332	0.0	0.000000	0.000000	...	0.055048		
82	0.000000	0.112332	0.0	0.079798	0.000000	...	0.055048		
99	0.000000	0.000000	0.0	0.000000	0.000000	...	0.055048		
104	0.000000	0.112326	0.0	0.079812	0.000000	...	0.055048		
...		
14904	0.000000	0.000000	0.0	0.000000	0.000000	...	0.055048		
14939	0.000000	0.000000	0.0	0.000000	0.000000	...	0.055048		
14952	0.000000	0.000000	0.0	0.000000	0.000000	...	0.055048		
14979	0.000000	0.000000	0.0	0.000000	0.000000	...	0.055048		
14982	0.000000	0.000000	0.0	0.079798	0.000000	...	0.055048		
Mall Shopping Clothes Hollister Abercrombie Die Death \									
20	0.164666	0.000000	0.098700	0.000000	0.000000	0.116845	0.0		
73	0.164666	0.000000	0.098613	0.000000	0.000000	0.000000	0.0		
82	0.000000	0.266961	0.000000	0.000000	0.000000	0.000000	0.000000		
99	0.164666	0.000000	0.000000	0.000000	0.000000	0.117160	0.0		
104	0.000000	0.000000	0.098613	0.000000	0.000000	0.000000	0.000000		
...		
14904	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000		
14939	0.000000	0.000000	0.000000	0.046088	0.034255	0.000000	0.0		
14952	0.167088	0.000000	0.098613	0.000000	0.000000	0.116845	0.0		
14979	0.000000	0.266961	0.000000	0.046088	0.034255	0.000000	0.0		
14982	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000		

Drunk	Drugs	
20	0.000000	0.000000
73	0.000000	0.038533
82	0.000000	0.000000

99	0.057613	0.000000
104	0.000000	0.000000
...
14904	0.000000	0.000000
14939	0.000000	0.000000
14952	0.057613	0.038533
14979	0.000000	0.000000
14982	0.000000	0.000000

[895 rows x 40 columns], 'Mall':

		Gradyear	Gender	Age	NumberOffriends	Basketball	\
2	4.630399e+98	0.000000	1.086951	5.998896	0.0		
14	4.700874e+98	0.124251	1.081084	8.015605	0.0		
20	4.700874e+98	0.000000	1.086951	0.000000	0.0		
22	4.560946e+98	0.000000	1.092219	7.917981	0.0		
36	4.700874e+98	0.000000	1.086951	6.428800	0.0		
...		
14968	4.630399e+98	0.124251	1.086951	6.624262	0.0		
14976	4.630399e+98	0.000000	1.086951	5.064946	0.0		
14988	4.772387e+98	0.000000	1.074503	8.798106	0.0		
14994	4.700874e+98	0.000000	1.086951	6.378075	0.0		
14999	4.772387e+98	0.000000	1.074503	4.903694	0.0		

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
2	0.000000	0.000000		0.0	0.000000	0.000000	...	0.000000
14	0.000000	0.000000		0.0	0.000000	0.000000	...	0.000000
20	0.000000	0.000000		0.0	0.079812	0.085365	...	0.055048
22	0.000000	0.000000		0.0	0.000000	0.000000	...	0.000000
36	0.000000	0.000000		0.0	0.000000	0.000000	...	0.000000
...
14968	0.000000	0.000000		0.0	0.000000	0.000000	...	0.000000
14976	0.000000	0.000000		0.0	0.000000	0.000000	...	0.000000
14988	0.161396	0.000000		0.0	0.000000	0.000000	...	0.000000
14994	0.000000	0.000000		0.0	0.000000	0.000000	...	0.000000
14999	0.000000	0.112332		0.0	0.000000	0.000000	...	0.000000

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
2	0.164666	0.000000	0.000000	0.0	0.0	0.000000	
14	0.164666	0.000000	0.000000	0.0	0.0	0.000000	
20	0.164666	0.000000	0.098700	0.0	0.0	0.116845	
22	0.164666	0.266961	0.098613	0.0	0.0	0.117151	
36	0.164666	0.000000	0.098613	0.0	0.0	0.000000	
...	
14968	0.167088	0.250530	0.000000	0.0	0.0	0.000000	
14976	0.164666	0.000000	0.000000	0.0	0.0	0.116845	
14988	0.167320	0.000000	0.000000	0.0	0.0	0.000000	
14994	0.164666	0.000000	0.000000	0.0	0.0	0.000000	

	Death	Drunk	Drugs			
2	0.000000	0.057613	0.038533			
14	0.000000	0.000000	0.000000			
20	0.000000	0.000000	0.000000			
22	0.000000	0.000000	0.000000			
36	0.000000	0.000000	0.000000			
...			
14968	0.076709	0.000000	0.000000			
14976	0.000000	0.000000	0.000000			
14988	0.000000	0.000000	0.000000			
14994	0.000000	0.000000	0.038533			
14999	0.000000	0.000000	0.038533			

[2628 rows x 40 columns], 'Shopping': Empty DataFrame

Columns: [Gradyear, Gender, Age, NumberOffriends, Basketball, Football, Soccer, Softball, Volleyball, Swimming, Cheerleading, Baseball, Tennis, Sports, Cute, Sex, Sexy, Hot, Kissed, Dance, Band, Marching, Music, Rock, God, Church, Jesus, Bible, Hair, Dress, Blonde, Mall, Shopping, Clothes, Hollister, Abercrombie, Die, Death, Drunk, Drugs]

Index: []

[0 rows x 40 columns], 'Clothes':				Gradyear	Gender	Age	NumberOffriends	Basketball
4	4.700874e+98	0.000000	1.081084	0.753045	0.000000			
6	4.560946e+98	0.124251	1.092219	4.251580	0.000000			
9	4.630399e+98	0.000000	1.086951	4.550377	0.000000			
20	4.700874e+98	0.000000	1.086951	0.000000	0.000000			
22	4.560946e+98	0.000000	1.092219	7.917981	0.000000			
...
14977	4.700874e+98	0.000000	1.081084	6.941186	0.159402			
14981	4.630399e+98	0.000000	1.086951	6.220999	0.000000			
14991	4.700874e+98	0.000000	1.081084	3.058506	0.000000			
14995	4.700874e+98	0.000000	1.081084	4.550377	0.000000			
14999	4.772387e+98	0.000000	1.074503	4.903694	0.000000			
Football Soccer Softball Volleyball Swimming ... Blonde \\								
4	0.000000	0.000000	0.000000	0.000000	0.085340	...	0.000000	
6	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
9	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
20	0.000000	0.000000	0.000000	0.079812	0.085365	...	0.055048	
22	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
...
14977	0.161396	0.112331	0.083794	0.000000	0.000000	...	0.000000	
14981	0.159361	0.000000	0.000000	0.000000	0.000000	...	0.000000	
14991	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
14995	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
14999	0.000000	0.112332	0.000000	0.000000	0.000000	...	0.000000	

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
4	0.000000	0.000000	0.098701	0.0	0.0	0.000000	
6	0.000000	0.000000	0.098613	0.0	0.0	0.000000	
9	0.000000	0.250530	0.098613	0.0	0.0	0.000000	
20	0.164666	0.000000	0.098700	0.0	0.0	0.116845	
22	0.164666	0.266961	0.098613	0.0	0.0	0.117151	
...	
14977	0.000000	0.000000	0.098613	0.0	0.0	0.117151	
14981	0.000000	0.000000	0.098613	0.0	0.0	0.000000	
14991	0.000000	0.250530	0.098613	0.0	0.0	0.000000	
14995	0.000000	0.000000	0.098700	0.0	0.0	0.000000	
14999	0.167088	0.250530	0.098613	0.0	0.0	0.000000	

	Death	Drunk	Drugs
4	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000
...
14977	0.076709	0.057613	0.038533
14981	0.000000	0.057613	0.000000
14991	0.076709	0.057613	0.000000
14995	0.076709	0.000000	0.000000
14999	0.000000	0.000000	0.038533

	Gradyear	Gender	Age	NumberOffriends	Basketball	\
[1729 rows x 40 columns], 'Hollister':						
19	4.700874e+98	0.000000	1.081084	0.000000	0.000000	
45	4.560946e+98	0.000000	1.096975	5.142291	0.000000	
117	4.630399e+98	0.000000	1.175558	4.251580	0.000000	
129	4.630399e+98	0.000000	1.092219	4.903694	0.000000	
175	4.630399e+98	0.000000	1.086951	9.475029	0.000000	
...	
14877	4.700874e+98	0.000000	1.081084	5.821456	0.159402	
14888	4.772387e+98	0.000000	1.074503	5.567866	0.159402	
14939	4.700874e+98	0.124251	1.081084	3.058506	0.000000	
14957	4.700874e+98	0.000000	1.081084	8.554224	0.000000	
14979	4.700874e+98	0.000000	1.086951	7.749401	0.000000	

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
19	0.000000	0.000000	0.000000	0.000000	0.00000	...	0.000000	
45	0.000000	0.000000	0.083794	0.000000	0.00000	...	0.000000	
117	0.000000	0.000000	0.000000	0.000000	0.00000	...	0.000000	
129	0.000000	0.112097	0.083815	0.000000	0.00000	...	0.055048	
175	0.000000	0.000000	0.000000	0.000000	0.00000	...	0.000000	

...
14877 0.159361 0.112332 0.000000 0.000000 0.08534 ... 0.000000
14888 0.159361 0.112097 0.083794 0.079798 0.00000 ... 0.000000
14939 0.000000 0.000000 0.000000 0.000000 0.00000 ... 0.055048
14957 0.000000 0.000000 0.000000 0.000000 0.00000 ... 0.000000
14979 0.000000 0.000000 0.000000 0.000000 0.00000 ... 0.055048

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
19	0.000000	0.000000	0.000000	0.046088	0.000000	0.000000	
45	0.000000	0.000000	0.000000	0.046088	0.000000	0.000000	
117	0.000000	0.250530	0.098613	0.046088	0.000000	0.117151	
129	0.000000	0.250530	0.000000	0.046088	0.000000	0.000000	
175	0.164666	0.266961	0.000000	0.046088	0.000000	0.000000	
...
14877	0.000000	0.266961	0.000000	0.046088	0.034255	0.000000	
14888	0.167088	0.000000	0.000000	0.046088	0.000000	0.000000	
14939	0.000000	0.000000	0.000000	0.046088	0.034255	0.000000	
14957	0.000000	0.000000	0.000000	0.046088	0.034255	0.000000	
14979	0.000000	0.266961	0.000000	0.046088	0.034255	0.000000	

	Death	Drunk	Drugs
19	0.000000	0.000000	0.000000
45	0.076709	0.000000	0.000000
117	0.000000	0.057613	0.038533
129	0.000000	0.000000	0.000000
175	0.076709	0.000000	0.000000
...
14877	0.000000	0.057613	0.000000
14888	0.000000	0.000000	0.000000
14939	0.000000	0.000000	0.000000
14957	0.000000	0.000000	0.000000
14979	0.000000	0.000000	0.000000

[799 rows x 40 columns], 'Abercrombie':
Gradyear Gender Age NumberOffriends Basketball \
23 4.630399e+98 0.000000 1.092219 8.141973 0.159402
31 4.630399e+98 0.000000 1.086951 7.191406 0.000000
90 4.772387e+98 0.000000 1.074503 9.077433 0.000000
111 4.630399e+98 0.000000 1.092219 6.808955 0.000000
159 4.700874e+98 0.000000 1.081084 4.251580 0.000000
...
14935 4.560946e+98 0.000000 1.092219 5.633260 0.000000
14939 4.700874e+98 0.124251 1.081084 3.058506 0.000000
14942 4.560946e+98 0.000000 1.092219 5.998896 0.000000
14957 4.700874e+98 0.000000 1.081084 8.554224 0.000000
14979 4.700874e+98 0.000000 1.086951 7.749401 0.000000

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
23	0.000000	0.0	0.0	0.079812	0.00000	...	0.000000	0.0	
31	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	0.0	
90	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	0.0	
111	0.000000	0.0	0.0	0.000000	0.08534	...	0.000000	0.0	
159	0.161396	0.0	0.0	0.079812	0.00000	...	0.000000	0.0	
...	
14935	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	0.0	
14939	0.000000	0.0	0.0	0.000000	0.00000	...	0.055048	0.0	
14942	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	0.0	
14957	0.000000	0.0	0.0	0.000000	0.00000	...	0.000000	0.0	
14979	0.000000	0.0	0.0	0.000000	0.00000	...	0.055048	0.0	

	Shopping	Clothes	Hollister	Abercrombie	Die	Death	Drunk	Drugs
23	0.000000	0.0	0.000000	0.034255	0.0	0.0	0.0	0.0
31	0.250530	0.0	0.000000	0.034255	0.0	0.0	0.0	0.0
90	0.000000	0.0	0.000000	0.034255	0.0	0.0	0.0	0.0
111	0.000000	0.0	0.000000	0.034255	0.0	0.0	0.0	0.0
159	0.270074	0.0	0.000000	0.034255	0.0	0.0	0.0	0.0
...
14935	0.000000	0.0	0.000000	0.034255	0.0	0.0	0.0	0.0
14939	0.000000	0.0	0.046088	0.034255	0.0	0.0	0.0	0.0
14942	0.000000	0.0	0.000000	0.034255	0.0	0.0	0.0	0.0
14957	0.000000	0.0	0.046088	0.034255	0.0	0.0	0.0	0.0
14979	0.266961	0.0	0.046088	0.034255	0.0	0.0	0.0	0.0

[616 rows x 40 columns], 'Die':				Gradyear	Gender	Age	NumberOffriends	Basketball
20	4.700874e+98	0.000000	1.086951		0.000000	0.000000		
21	4.560946e+98	0.000000	1.092219		6.378075	0.000000		
22	4.560946e+98	0.000000	1.092219		7.917981	0.000000		
37	4.560946e+98	0.124251	1.086951		8.110772	0.000000		
49	4.630399e+98	0.124251	1.086951		0.753045	0.000000		
...		
14971	4.560946e+98	0.000000	1.092219		1.957650	0.000000		
14976	4.630399e+98	0.000000	1.086951		5.064946	0.000000		
14977	4.700874e+98	0.000000	1.081084		6.941186	0.159402		
14984	4.700874e+98	0.000000	1.081084		5.142291	0.159402		
14989	4.630399e+98	0.000000	1.175558		6.055888	0.000000		

14971	0.000000	0.000000	0.083794	0.000000	0.000000	...	0.000000
14976	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
14977	0.161396	0.112331	0.083794	0.000000	0.000000	...	0.000000
14984	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
14989	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
20	0.164666	0.000000	0.098700	0.0	0.0	0.116845	
21	0.000000	0.000000	0.000000	0.0	0.0	0.116845	
22	0.164666	0.266961	0.098613	0.0	0.0	0.117151	
37	0.000000	0.000000	0.000000	0.0	0.0	0.116845	
49	0.000000	0.000000	0.000000	0.0	0.0	0.116845	
...	
14971	0.000000	0.000000	0.000000	0.0	0.0	0.116845	
14976	0.164666	0.000000	0.000000	0.0	0.0	0.116845	
14977	0.000000	0.000000	0.098613	0.0	0.0	0.117151	
14984	0.000000	0.000000	0.000000	0.0	0.0	0.116845	
14989	0.000000	0.000000	0.000000	0.0	0.0	0.116845	

	Death	Drunk	Drugs
20	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000
37	0.000000	0.000000	0.000000
49	0.000000	0.000000	0.000000
...
14971	0.000000	0.000000	0.000000
14976	0.000000	0.000000	0.000000
14977	0.076709	0.057613	0.038533
14984	0.000000	0.000000	0.000000
14989	0.000000	0.000000	0.000000

[1974 rows x 40 columns], 'Death':

	Gradyear	Gender	Age	NumberOfffriends	Basketball	\
15	4.630399e+98	0.000000	1.086951	5.759993	0.000000	
41	4.560946e+98	0.000000	1.092219	4.550377	0.000000	
45	4.560946e+98	0.000000	1.096975	5.142291	0.000000	
85	4.630399e+98	0.000000	1.092219	9.452657	0.161439	
87	4.560946e+98	0.000000	1.092219	5.217624	0.000000	
...	
14968	4.630399e+98	0.124251	1.086951	6.624262	0.000000	
14977	4.700874e+98	0.000000	1.081084	6.941186	0.159402	
14991	4.700874e+98	0.000000	1.081084	3.058506	0.000000	
14995	4.700874e+98	0.000000	1.081084	4.550377	0.000000	
14996	4.700874e+98	0.000000	1.081084	6.478755	0.000000	

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
--	----------	--------	----------	------------	----------	-----	--------	---

15	0.159361	0.000000	0.083816	0.0	0.0	...	0.0
41	0.000000	0.000000	0.000000	0.0	0.0	...	0.0
45	0.000000	0.000000	0.083794	0.0	0.0	...	0.0
85	0.000000	0.000000	0.000000	0.0	0.0	...	0.0
87	0.000000	0.000000	0.083816	0.0	0.0	...	0.0
...
14968	0.000000	0.000000	0.000000	0.0	0.0	...	0.0
14977	0.161396	0.112331	0.083794	0.0	0.0	...	0.0
14991	0.000000	0.000000	0.000000	0.0	0.0	...	0.0
14995	0.000000	0.000000	0.000000	0.0	0.0	...	0.0
14996	0.000000	0.000000	0.000000	0.0	0.0	...	0.0

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
15	0.000000	0.00000	0.000000	0.000000	0.0	0.000000	
41	0.000000	0.00000	0.098613	0.000000	0.0	0.000000	
45	0.000000	0.00000	0.000000	0.046088	0.0	0.000000	
85	0.164666	0.25053	0.098613	0.000000	0.0	0.000000	
87	0.000000	0.25053	0.000000	0.000000	0.0	0.116845	
...	
14968	0.167088	0.25053	0.000000	0.000000	0.0	0.000000	
14977	0.000000	0.00000	0.098613	0.000000	0.0	0.117151	
14991	0.000000	0.25053	0.098613	0.000000	0.0	0.000000	
14995	0.000000	0.00000	0.098700	0.000000	0.0	0.000000	
14996	0.000000	0.00000	0.000000	0.000000	0.0	0.000000	

	Death	Drunk	Drugs
15	0.076709	0.000000	0.000000
41	0.076709	0.000000	0.000000
45	0.076709	0.000000	0.000000
85	0.076709	0.000000	0.000000
87	0.076709	0.000000	0.000000
...
14968	0.076709	0.000000	0.000000
14977	0.076709	0.057613	0.038533
14991	0.076709	0.057613	0.000000
14995	0.076709	0.000000	0.000000
14996	0.076709	0.000000	0.000000

[1363 rows x 40 columns], 'Drunk':	Gradyear	Gender	Age	NumberOfffriends	Basketball	\
1	4.630399e+98	0.000000	1.086951	6.428800	0.000000	
2	4.630399e+98	0.000000	1.086951	5.998896	0.000000	
25	4.700874e+98	0.000000	1.081084	7.109973	0.000000	
47	4.700874e+98	0.000000	1.081084	3.791035	0.000000	
70	4.630399e+98	0.000000	1.092219	4.251580	0.000000	
...	
14952	4.700874e+98	0.000000	1.081084	0.000000	0.000000	

14967	4.560946e+98	0.124251	1.086951		4.354949	0.000000
14977	4.700874e+98	0.000000	1.081084		6.941186	0.159402
14981	4.630399e+98	0.000000	1.086951		6.220999	0.000000
14991	4.700874e+98	0.000000	1.081084		3.058506	0.000000

	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	\
1	0.000000	0.112097	0.000000	0.000000	0.08534	...	0.000000	
2	0.000000	0.000000	0.000000	0.000000	0.00000	...	0.000000	
25	0.000000	0.000000	0.000000	0.000000	0.00000	...	0.000000	
47	0.159361	0.000000	0.000000	0.079798	0.00000	...	0.000000	
70	0.000000	0.000000	0.000000	0.000000	0.00000	...	0.000000	
...	
14952	0.000000	0.000000	0.000000	0.000000	0.00000	...	0.055048	
14967	0.000000	0.000000	0.000000	0.000000	0.00000	...	0.000000	
14977	0.161396	0.112331	0.083794	0.000000	0.00000	...	0.000000	
14981	0.159361	0.000000	0.000000	0.000000	0.00000	...	0.000000	
14991	0.000000	0.000000	0.000000	0.000000	0.00000	...	0.000000	

	Mall	Shopping	Clothes	Hollister	Abercrombie	Die	\
1	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
2	0.164666	0.00000	0.000000	0.0	0.0	0.000000	
25	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
47	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
70	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
...	
14952	0.167088	0.00000	0.098613	0.0	0.0	0.116845	
14967	0.000000	0.00000	0.000000	0.0	0.0	0.000000	
14977	0.000000	0.00000	0.098613	0.0	0.0	0.117151	
14981	0.000000	0.00000	0.098613	0.0	0.0	0.000000	
14991	0.000000	0.25053	0.098613	0.0	0.0	0.000000	

	Death	Drunk	Drugs
1	0.000000	0.057613	0.000000
2	0.000000	0.057613	0.038533
25	0.000000	0.057613	0.000000
47	0.000000	0.057613	0.000000
70	0.000000	0.057613	0.000000
...
14952	0.000000	0.057613	0.038533
14967	0.000000	0.057613	0.000000
14977	0.076709	0.057613	0.038533
14981	0.000000	0.057613	0.000000
14991	0.076709	0.057613	0.000000

[979 rows x 40 columns], 'Drugs':

2	4.630399e+98	0.0	1.086951
---	--------------	-----	----------

Gradyear Gender Age NumberOfffriends Basketball Football \

2	5.998896	0.000000	0.000000
---	----------	----------	----------

73	4.630399e+98	0.0	1.086951	3.914234	0.159402	0.159361
96	4.772387e+98	0.0	1.081084	5.432670	0.000000	0.000000
105	4.772387e+98	0.0	1.081084	4.354949	0.161589	0.000000
117	4.630399e+98	0.0	1.175558	4.251580	0.000000	0.000000
...
14887	4.700874e+98	0.0	1.086951	3.524466	0.000000	0.000000
14952	4.700874e+98	0.0	1.081084	0.000000	0.000000	0.000000
14977	4.700874e+98	0.0	1.081084	6.941186	0.159402	0.161396
14994	4.700874e+98	0.0	1.086951	6.378075	0.000000	0.000000
14999	4.772387e+98	0.0	1.074503	4.903694	0.000000	0.000000

	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	\
2	0.000000	0.000000		0.0	0.0	...	0.000000	0.164666
73	0.112332	0.000000		0.0	0.0	...	0.055048	0.164666
96	0.000000	0.000000		0.0	0.0	...	0.000000	0.000000
105	0.112332	0.000000		0.0	0.0	...	0.000000	0.164666
117	0.000000	0.000000		0.0	0.0	...	0.000000	0.000000
...
14887	0.000000	0.000000		0.0	0.0	...	0.000000	0.164666
14952	0.000000	0.000000		0.0	0.0	...	0.055048	0.167088
14977	0.112331	0.083794		0.0	0.0	...	0.000000	0.000000
14994	0.000000	0.000000		0.0	0.0	...	0.000000	0.164666
14999	0.112332	0.000000		0.0	0.0	...	0.000000	0.167088

	Shopping	Clothes	Hollister	Abercrombie	Die	Death	\
2	0.00000	0.000000	0.000000		0.0	0.000000	0.000000
73	0.00000	0.098613	0.000000		0.0	0.000000	0.000000
96	0.00000	0.000000	0.000000		0.0	0.000000	0.000000
105	0.25053	0.098613	0.000000		0.0	0.000000	0.000000
117	0.25053	0.098613	0.046088		0.0	0.117151	0.000000
...
14887	0.25053	0.000000	0.000000		0.0	0.000000	0.000000
14952	0.00000	0.098613	0.000000		0.0	0.116845	0.000000
14977	0.00000	0.098613	0.000000		0.0	0.117151	0.076709
14994	0.00000	0.000000	0.000000		0.0	0.000000	0.000000
14999	0.25053	0.098613	0.000000		0.0	0.000000	0.000000

	Drunk	Drugs
2	0.057613	0.038533
73	0.000000	0.038533
96	0.000000	0.038533
105	0.000000	0.038533
117	0.057613	0.038533
...
14887	0.000000	0.038533
14952	0.057613	0.038533

```
14977  0.057613  0.038533  
14994  0.000000  0.038533  
14999  0.000000  0.038533
```

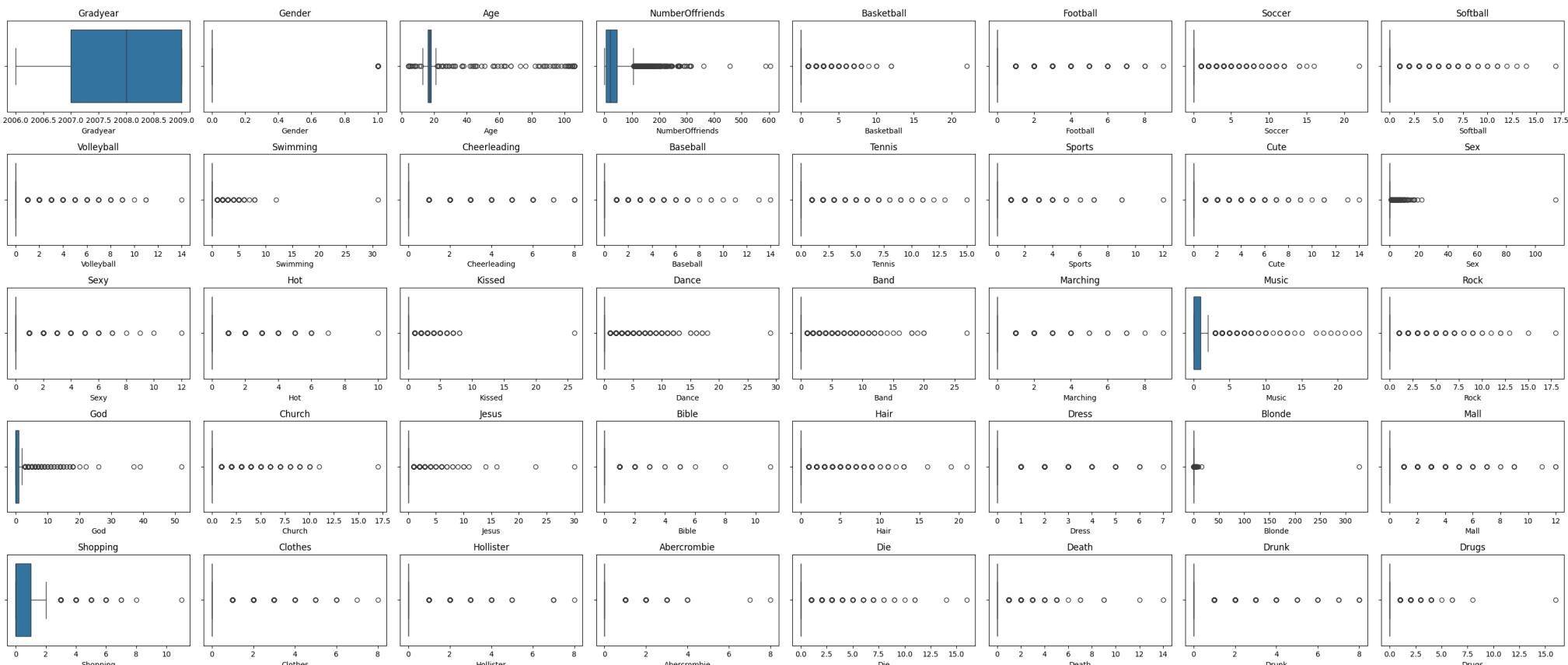
[654 rows x 40 columns]}

Plotting data without removing skewness shows outliers

```
In [73]: plt.figure(figsize=(30, 20))

for col in df1.columns:
    plt.subplot(8, 8, df1.columns.get_loc(col) + 1)
    sns.boxplot(x=df1[col])
    plt.title(col)
    plt.tight_layout()

plt.show()
```



The variables 'Age' and 'Numberoffriends' has more outliers

```
In [ ]: # Alternatively, handle outliers using IQR method
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1
df_clean = df[~((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).any(axis=1)]
```

```
In [63]: df4=df2.copy()

def cap_outliers(df4, column):
    Q1 = df4[column].quantile(0.25)
    Q3 = df4[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    df4[column] = np.where(df4[column] < lower_bound, lower_bound, df4[column])
    df4[column] = np.where(df4[column] > upper_bound, upper_bound, df4[column])
    return df4

for col in df4.select_dtypes(include=['number']).columns:
    df4 = cap_outliers(df4, col)
```

```
In [62]: df2 #data after removing skewness but without removing outliers
```

Out[62]:

	Gradyear	Gender	Age	NumberOffriends	Basketball	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	Shoppin
0	4.630399e+98	0.0	1.086951	0.000000	0.0	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.000000	0.0000
1	4.630399e+98	0.0	1.086951	6.428800	0.0	0.0	0.112097	0.0	0.0	0.08534	...	0.0	0.000000	0.0000
2	4.630399e+98	0.0	1.086951	5.998896	0.0	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.164666	0.0000
3	4.560946e+98	0.0	1.086951	5.697281	0.0	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.000000	0.0000
4	4.700874e+98	0.0	1.081084	0.753045	0.0	0.0	0.000000	0.0	0.0	0.08534	...	0.0	0.000000	0.0000
...
14995	4.700874e+98	0.0	1.081084	4.550377	0.0	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.000000	0.0000
14996	4.700874e+98	0.0	1.081084	6.478755	0.0	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.000000	0.0000
14997	4.630399e+98	0.0	1.086951	5.432670	0.0	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.000000	0.0000
14998	4.630399e+98	0.0	1.086951	4.454441	0.0	0.0	0.000000	0.0	0.0	0.00000	...	0.0	0.000000	0.0000
14999	4.772387e+98	0.0	1.074503	4.903694	0.0	0.0	0.112332	0.0	0.0	0.00000	...	0.0	0.167088	0.2505

14136 rows × 40 columns



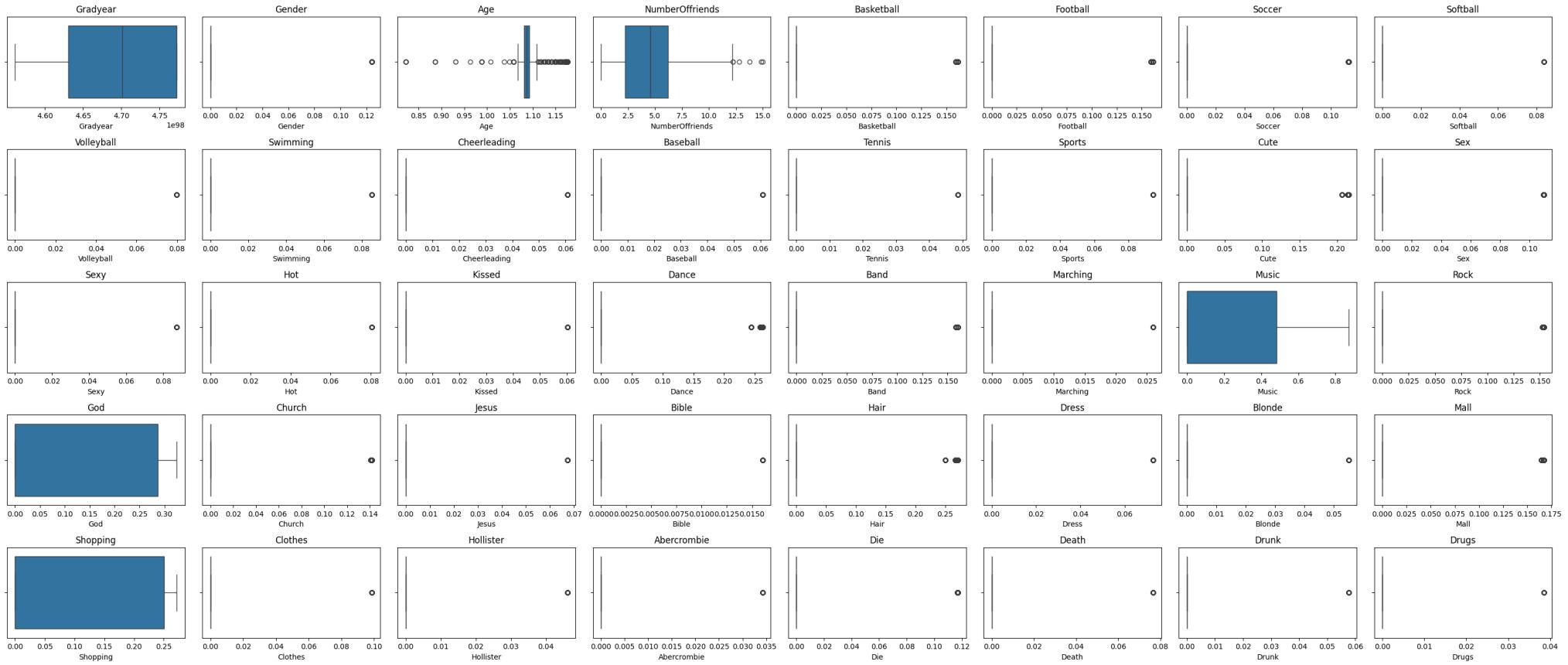
Plotting data (df2) after removing skewnes to show outliers:

In [69]:

```
#df2 without removal of outliers
plt.figure(figsize=(30, 20))

for col in df2.columns:
    plt.subplot(8, 8, df2.columns.get_loc(col) + 1)
    sns.boxplot(x=df2[col])
    plt.title(col)

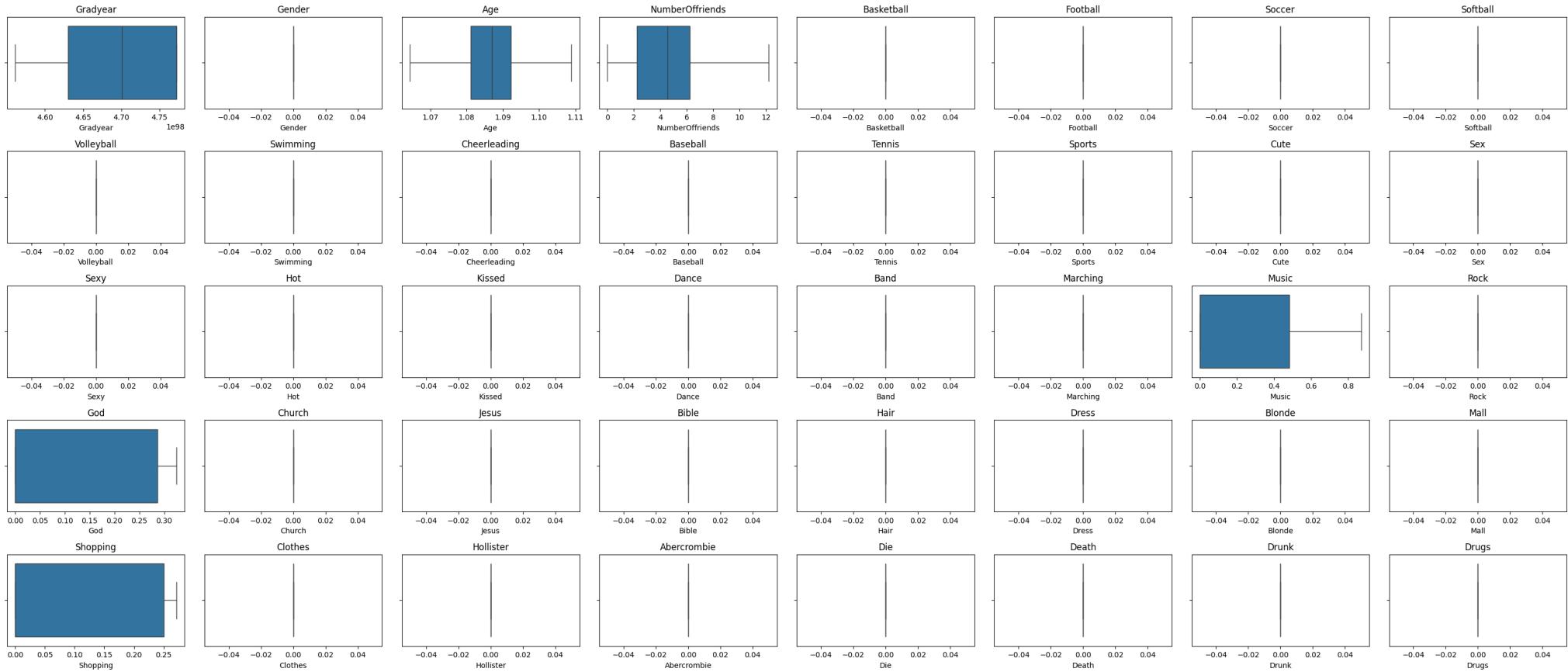
plt.tight_layout()
plt.show()
```



Plotting Data (df4) after removing outliers:

```
In [66]: #After removing outliers
plt.figure(figsize=(30, 20))
for col in df4.columns:
    plt.subplot(8, 8, df4.columns.get_loc(col) + 1)
    sns.boxplot(x=df4[col])
    plt.title(col)

plt.tight_layout()
plt.show()
```



In [68]: `df4.head()`

	Gradyear	Gender	Age	NumberOffriends	Basketball	Football	Soccer	Softball	Volleyball	Swimming	...	Blonde	Mall	Shopping	Clothes
0	4.630399e+98	0.0	1.086951	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
1	4.630399e+98	0.0	1.086951	6.428800	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
2	4.630399e+98	0.0	1.086951	5.998896	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
3	4.560946e+98	0.0	1.086951	5.697281	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
4	4.700874e+98	0.0	1.081084	0.753045	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0

5 rows × 40 columns

We are considering df2 that is the dataset after removing skewness and then performing feature scaling

Feature Scaling

```
In [74]: from sklearn.preprocessing import StandardScaler
```

```
# Standardizing the data (for clustering)
scaler = StandardScaler()
scaled_data = scaler.fit_transform(df2.select_dtypes(include=['float64', 'int64']))

# Convert back to a DataFrame
df_scaled = pd.DataFrame(scaled_data, columns=df2.select_dtypes(include=['float64', 'int64']).columns)
```

```
In [75]: from sklearn.decomposition import PCA
```

```
pca = PCA(n_components=2)
pca_components = pca.fit_transform(df_scaled)

# DataFrame with the PCA components
df_pca = pd.DataFrame(pca_components, columns=['PC1', 'PC2'])

print(f'Explained Variance Ratio: {pca.explained_variance_ratio_}' )
```

```
Explained Variance Ratio: [0.08313143 0.04460879]
```

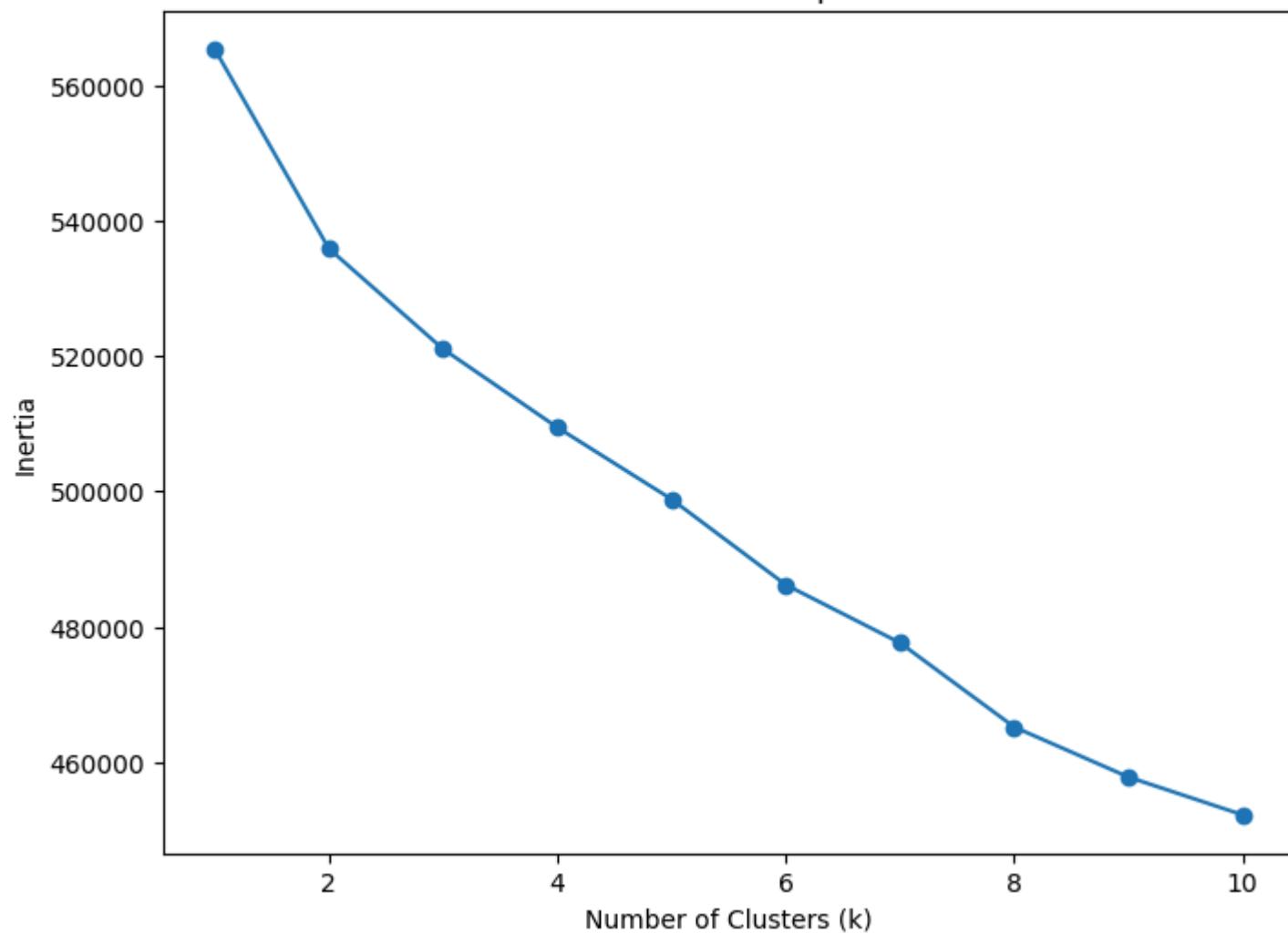
```
In [91]: #Elboe plot
```

```
from sklearn.cluster import KMeans

inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(df_scaled)
    inertia.append(kmeans.inertia_)

plt.figure(figsize=(8, 6))
plt.plot(range(1, 11), inertia, marker='o')
plt.title('Elbow Method')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Inertia')
plt.show()
```

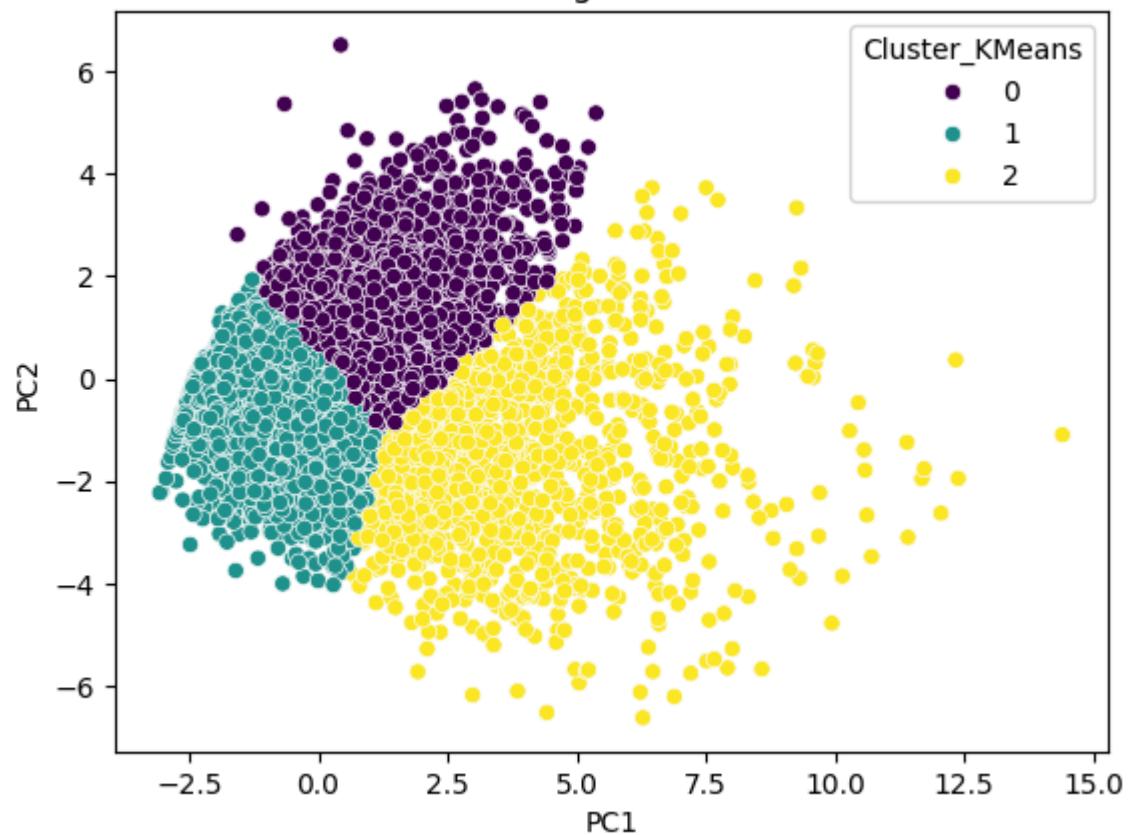
Elbow Method for Optimal k



```
In [76]: # KMeans on PCA components  
#From above elbow plot we take number of clusters as k= 3  
kmeans = KMeans(n_clusters=3)  
df_pca['Cluster_KMeans'] = kmeans.fit_predict(df_pca)
```

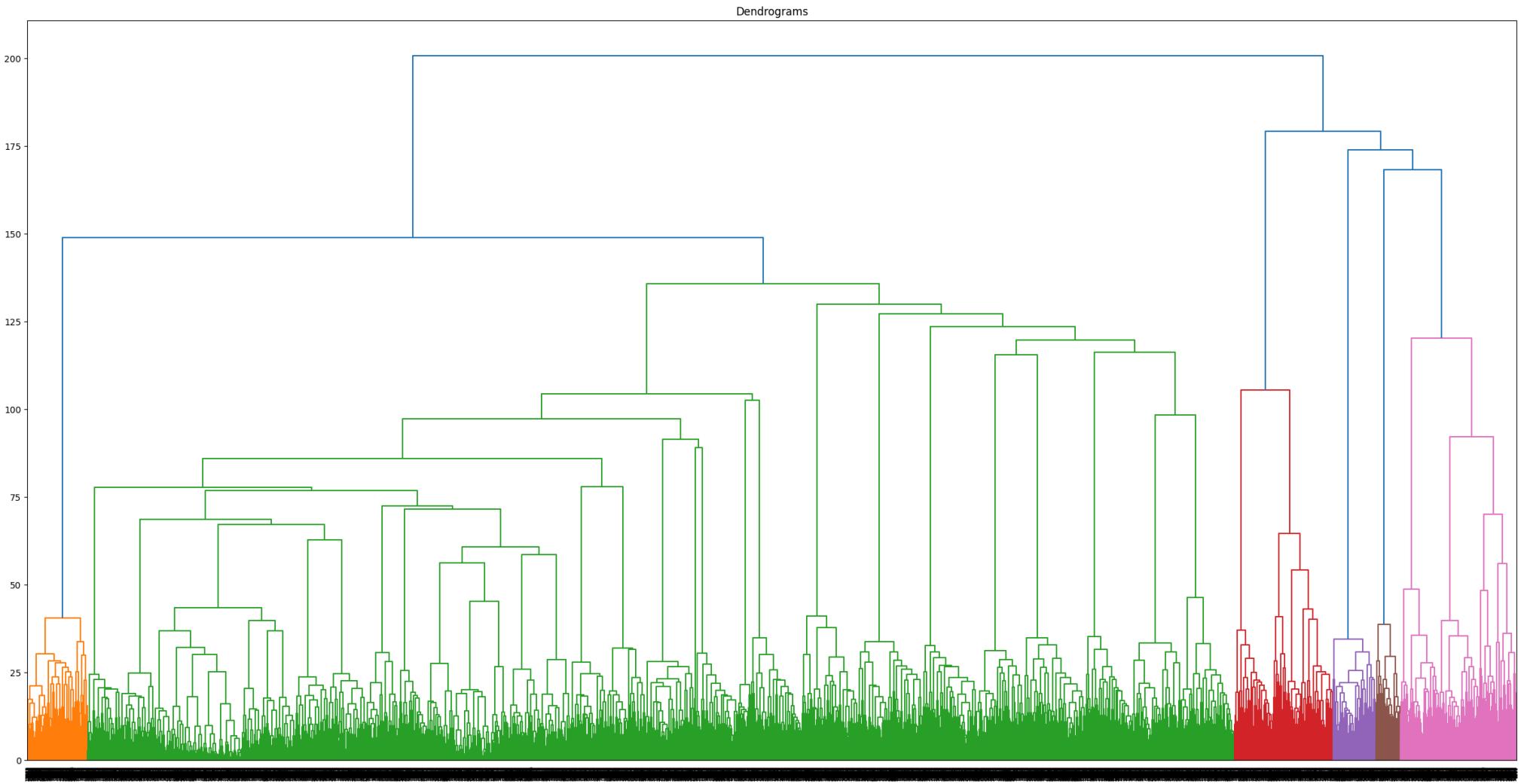
```
sns.scatterplot(x='PC1', y='PC2', hue='Cluster_KMeans', data=df_pca, palette='viridis')  
plt.title('KMeans Clustering (PCA Reduced Data)')  
plt.show()
```

KMeans Clustering (PCA Reduced Data)



```
In [94]: from scipy.cluster.hierarchy import dendrogram  
from scipy.cluster import hierarchy
```

```
In [93]: import scipy.cluster.hierarchy as sch  
plt.figure(figsize=(30, 15))  
plt.title("Dendrograms")  
dend = sch.dendrogram(sch.linkage(df_scaled, method='ward'))  
plt.show()
```



```
In [95]: model = AgglomerativeClustering(n_clusters=None, distance_threshold=0)
```

```
In [96]: cluster_labels = model.fit_predict(df1)
cluster_labels
```

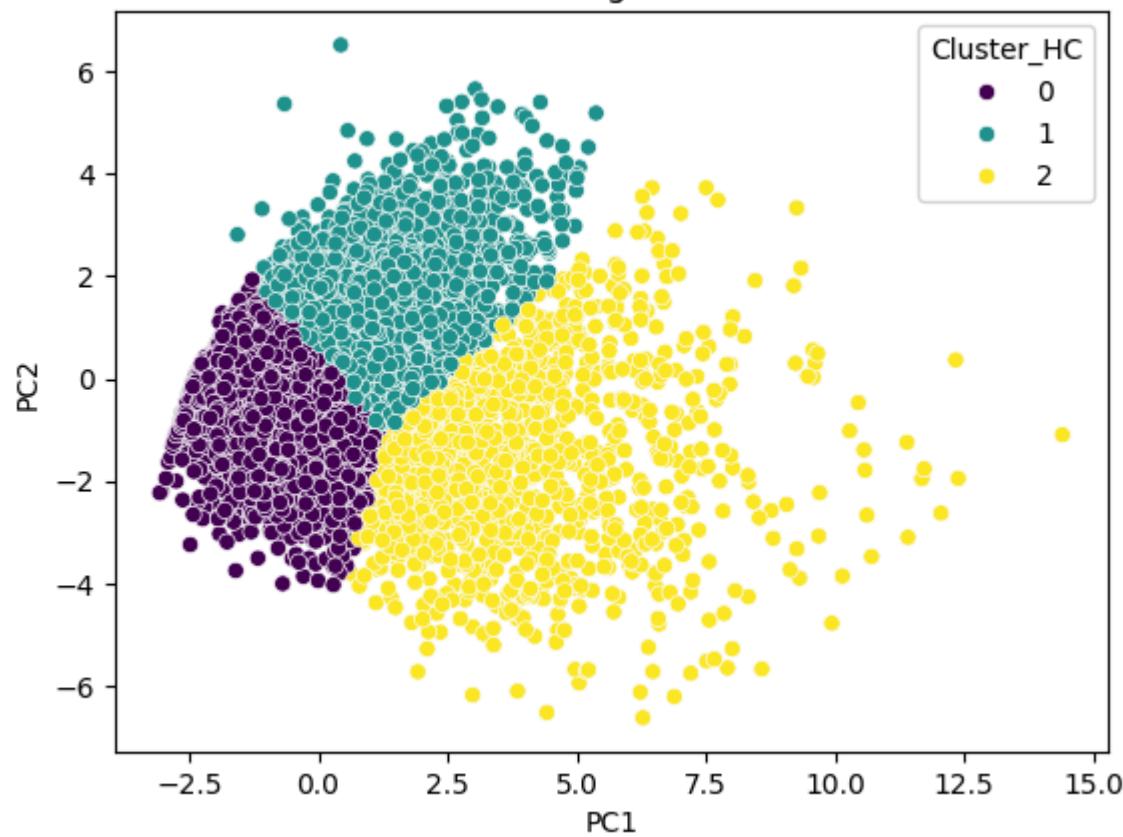
```
Out[96]: array([14135, 10127, 9641, ..., 2, 1, 0])
```

```
In [77]: # Agglomerative Clustering on PCA components
from sklearn.cluster import AgglomerativeClustering

hc = AgglomerativeClustering(n_clusters=3)
df_pca['Cluster_HC'] = hc.fit_predict(df_pca)
```

```
sns.scatterplot(x='PC1', y='PC2', hue='Cluster_HC', data=df_pca, palette='viridis')
plt.title('Hierarchical Clustering (PCA Reduced Data)')
plt.show()
```

Hierarchical Clustering (PCA Reduced Data)

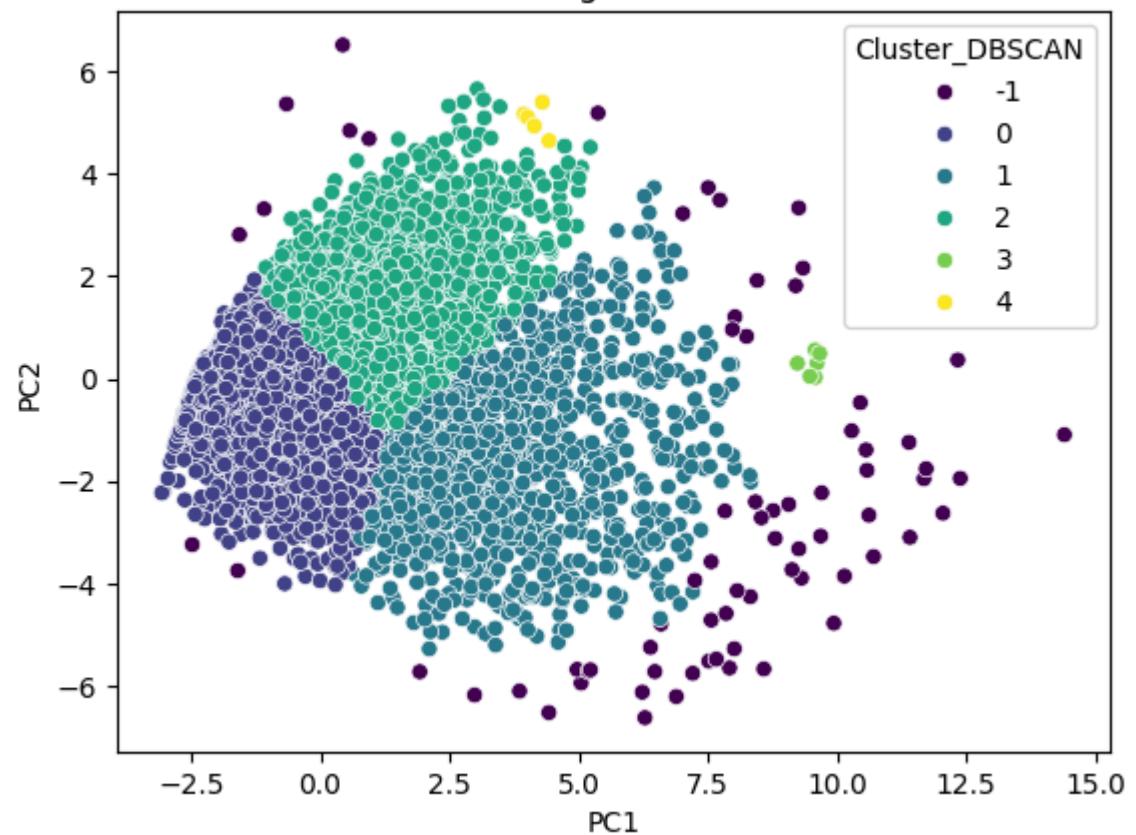


```
In [78]: # Apply DBSCAN on PCA components
from sklearn.cluster import DBSCAN

dbscan = DBSCAN(eps=0.5, min_samples=5)
df_pca['Cluster_DBSCAN'] = dbscan.fit_predict(df_pca)

sns.scatterplot(x='PC1', y='PC2', hue='Cluster_DBSCAN', data=df_pca, palette='viridis')
plt.title('DBSCAN Clustering (PCA Reduced Data)')
plt.show()
```

DBSCAN Clustering (PCA Reduced Data)

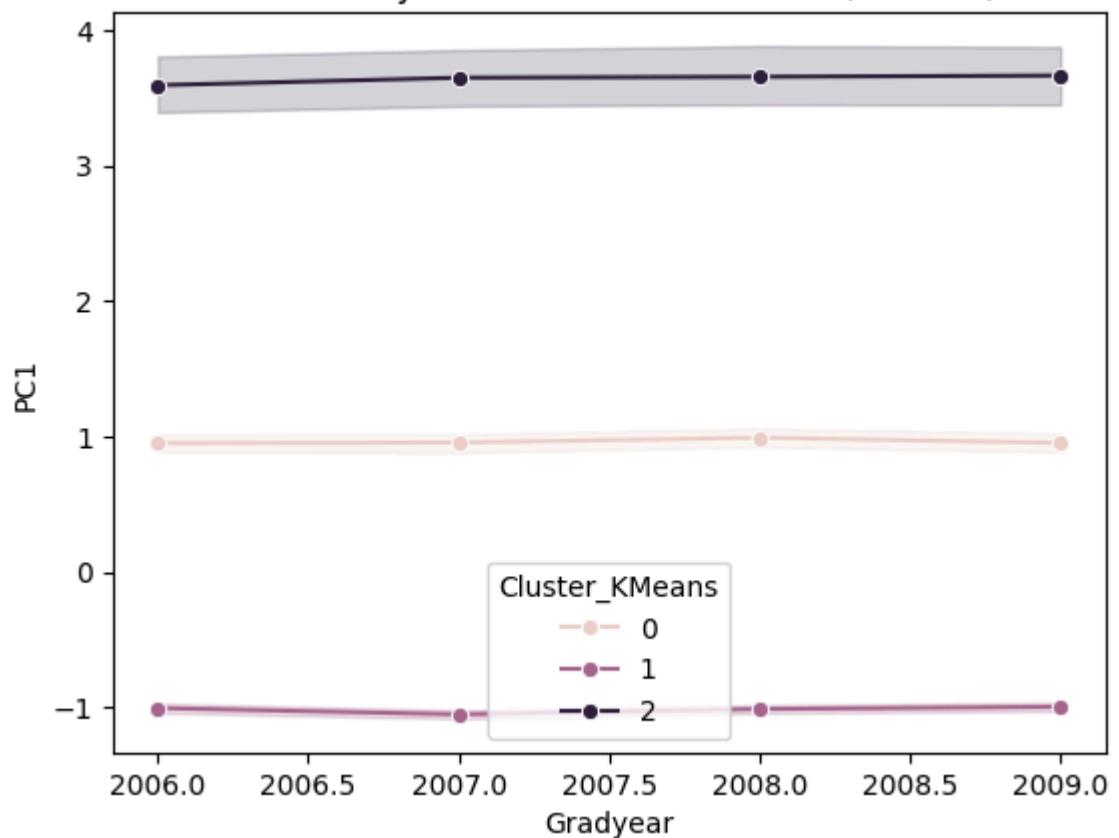


In [80]:

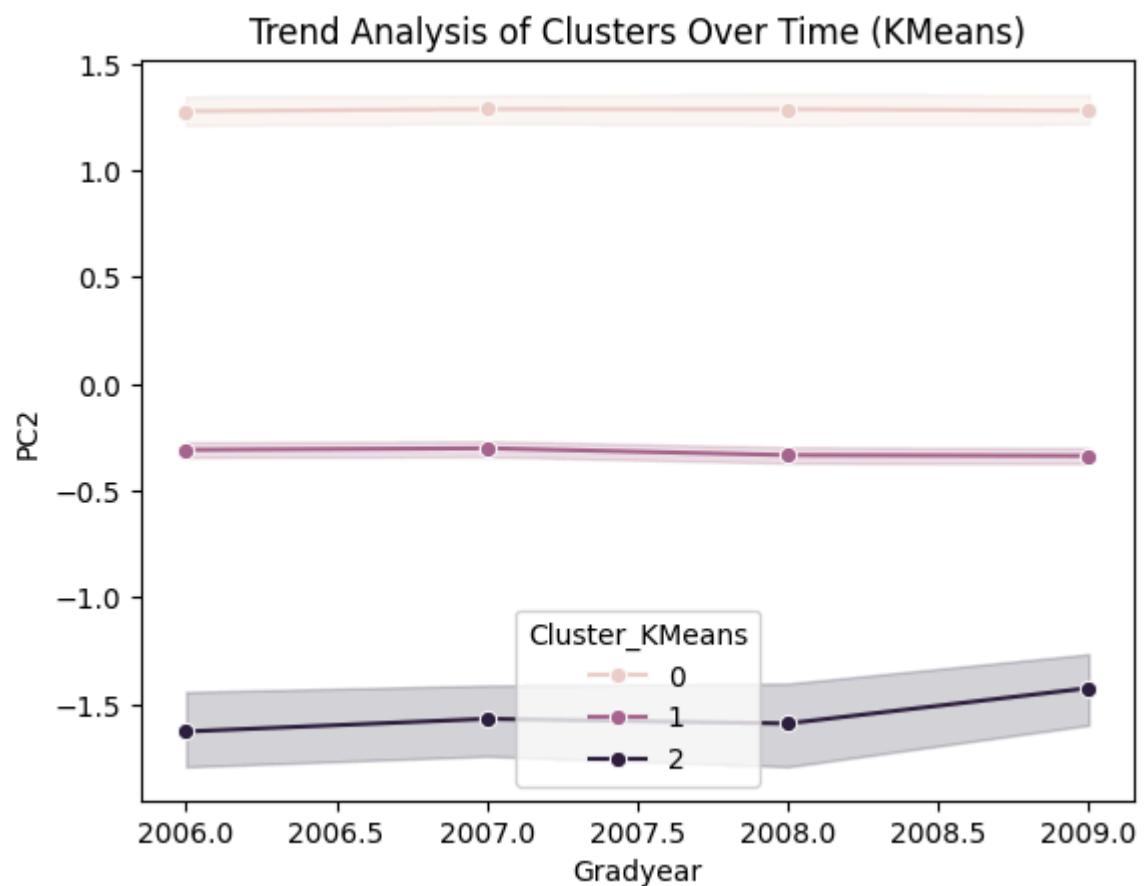
```
#Analysis
df_pca['Gradyear'] = df['Gradyear']

sns.lineplot(data=df_pca, x='Gradyear', y='PC1', hue='Cluster_KMeans', marker='o')
plt.title('Trend Analysis of Clusters Over Time (KMeans)')
plt.show()
```

Trend Analysis of Clusters Over Time (KMeans)



```
In [81]: sns.lineplot(data=df_pca, x='Gradyear', y='PC2', hue='Cluster_KMeans', marker='o')
plt.title('Trend Analysis of Clusters Over Time (KMeans)')
plt.show()
```



```
In [84]: from sklearn.metrics import silhouette_score

# Silhouette scores for each clustering method
silhouette_kmeans = silhouette_score(df_pca[['PC1', 'PC2']], df_pca['Cluster_KMeans'])
silhouette_hc = silhouette_score(df_pca[['PC1', 'PC2']], df_pca['Cluster_HC'])
silhouette_dbSCAN = silhouette_score(df_pca[['PC1', 'PC2']], df_pca['Cluster_DBSCAN'])

print(f"KMeans Silhouette Score: {silhouette_kmeans}")
print(f"Hierarchical Clustering Silhouette Score: {silhouette_hc}")
print(f"DBSCAN Silhouette Score: {silhouette_dbSCAN}")
```

KMeans Silhouette Score: 0.4115131680794632
 Hierarchical Clustering Silhouette Score: 0.4115131680794632
 DBSCAN Silhouette Score: 0.39229084912116113

Silhouette score for KMeans and Hierarchical Clustering is almost same but for DBSCAN Clustering Silhouette score is lower in comparison to other clustering:

KMeans Silhouette Score: 0.4115131680794632

Hierarchical Clustering Silhouette Score: 0.4115131680794632

DBSCAN Silhouette Score: 0.39229084912116113

In [88]:

```
#Cluster wise analysis:  
df['Cluster_KMeans'] = df_pca['Cluster_KMeans']  
df['Cluster_HC'] = df_pca['Cluster_HC']  
df['Cluster_DBSCAN'] = df_pca['Cluster_DBSCAN']  
  
# KMeans  
cluster_demographics_kmeans = df.groupby('Cluster_KMeans').agg({  
    'Age': 'mean',  
    'Gender': lambda x: x.mode()[0], # Most frequent gender  
    'Gradyear': lambda x: x.mode()[0] # Most frequent graduation year  
})  
  
print("Demographic Profile for KMeans Clusters:")  
print(cluster_demographics_kmeans)  
  
# Hierarchical Clusters and DBSCAN Clusters  
cluster_demographics_hc = df.groupby('Cluster_HC').agg({  
    'Age': 'mean',  
    'Gender': lambda x: x.mode()[0],  
    'Gradyear': lambda x: x.mode()[0]  
})  
  
print("Demographic Profile for Hierarchical Clusters:")  
print(cluster_demographics_hc)  
  
cluster_demographics_dbSCAN = df.groupby('Cluster_DBSCAN').agg({  
    'Age': 'mean',  
    'Gender': lambda x: x.mode()[0],  
    'Gradyear': lambda x: x.mode()[0]  
})  
  
print("Demographic Profile for DBSCAN Clusters:")  
print(cluster_demographics_dbSCAN)
```

Demographic Profile for KMeans Clusters:

	Age	Gender	Gradyear
Cluster_KMeans			
0.0	17.486167	F	2009
1.0	17.435254	F	2008
2.0	17.253012	F	2009

Demographic Profile for Hierarchical Clusters:

	Age	Gender	Gradyear
Cluster_HC			
0.0	17.435254	F	2008
1.0	17.486167	F	2009
2.0	17.253012	F	2009

Demographic Profile for DBSCAN Clusters:

	Age	Gender	Gradyear
Cluster_DBSCAN			
-1.0	16.776119	F	2009
0.0	17.435239	F	2008
1.0	17.281300	F	2007
2.0	17.489092	F	2009
3.0	16.333333	F	2008
4.0	16.000000	F	2009

```
In [101]: df['Cluster_KMeans'] = df_pca['Cluster_KMeans']
df['Cluster_HC'] = df_pca['Cluster_HC']
df['Cluster_DBSCAN'] = df_pca['Cluster_DBSCAN']
```

```
demographic_profile = df.groupby('Cluster_KMeans').agg({
    'Age': 'mean', # Average age
    'Gender': lambda x: x.mode()[0], # Most frequent gender
    'Gradyear': lambda x: x.mode()[0], # Most frequent graduation year
    'Basketball': 'mean', # Sports-related activities (average participation)
    'Football': 'mean',
    'Soccer': 'mean',
    'Softball': 'mean',
    'Volleyball': 'mean',
    'Swimming': 'mean',
    'Cheerleading': 'mean',
    'Shopping': 'mean', # Fashion-related activities
    'Mall': 'mean',
    'Hollister': 'mean',
    'Abercrombie': 'mean',
    'Rock': 'mean', # Music-related interests
    'Band': 'mean',
    'Marching': 'mean',
})
```

```
print("Demographic Profile for KMeans Clusters:")
#print(demographic_profile)
demographic_profile
```

Demographic Profile for KMeans Clusters:

Out[101...]

	Age	Gender	Gradyear	Basketball	Football	Soccer	Softball	Volleyball	Swimming	Cheerleading	Shopping	Mall	Hollister
--	-----	--------	----------	------------	----------	--------	----------	------------	----------	--------------	----------	------	-----------

Cluster_KMeans

0.0	17.486167	F	2009	0.300564	0.277196	0.257051	0.170561	0.146387	0.129197	0.118184	0.364222	0.273973	0.0735
1.0	17.435254	F	2008	0.277671	0.276224	0.230408	0.173981	0.141548	0.148662	0.108874	0.384254	0.277068	0.0777
2.0	17.253012	F	2009	0.289910	0.225904	0.210090	0.149096	0.206325	0.139307	0.106928	0.363705	0.277861	0.0820

Interpretation:

Cluster 0: May represent students who are highly interested in sports, with a strong presence in basketball, football, and soccer. These students might be mostly from a specific graduation year and of a certain age group. A sports club or physical activity-based program would be ideal for this group.

Cluster 1: May represent students with an interest in fashion, as evidenced by their participation in shopping, mall activities, and affinity for brands like Hollister and Abercrombie. This group could be targeted for fashion-related events or marketing programs from these brands.

Cluster 2: Could represent students with a strong interest in music (especially rock music), cheerleading, or performing arts. A program in performing arts or a music event could be ideal for them.

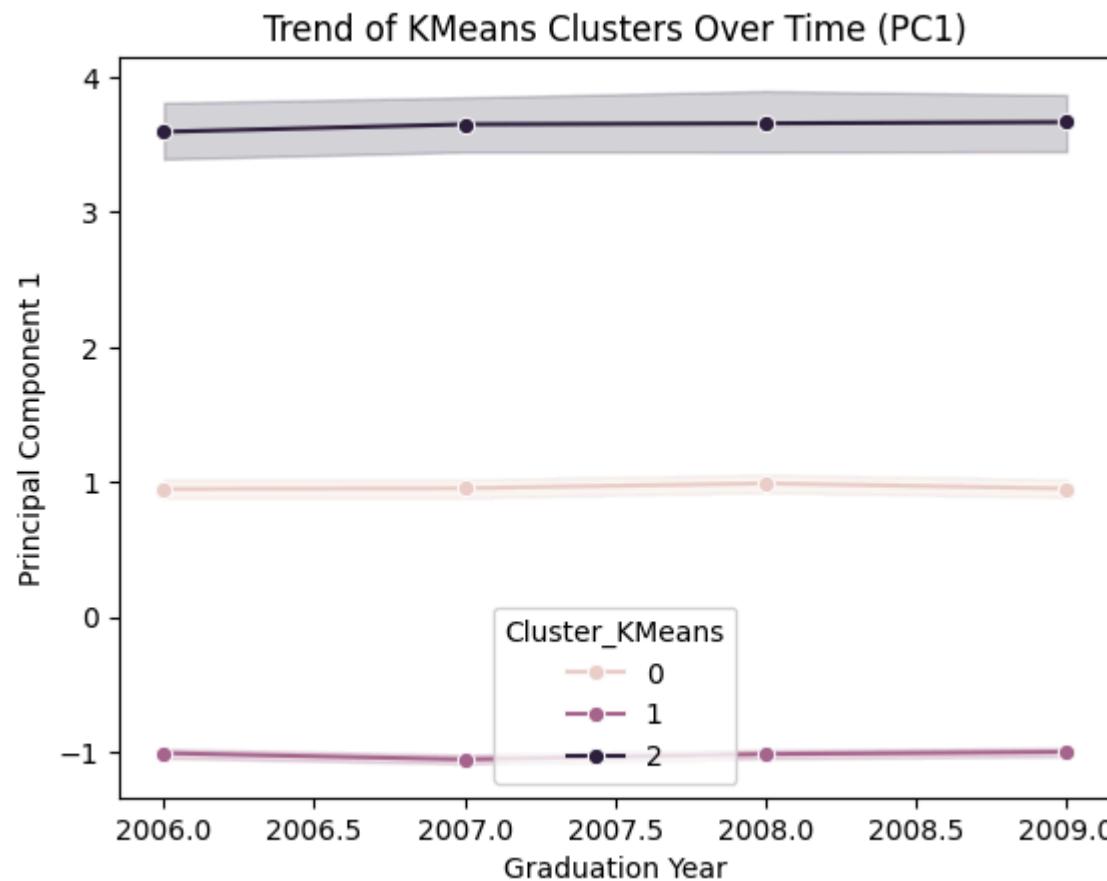
In [103...]

```
#Analysis
df_pca['Gradyear'] = df['Gradyear']

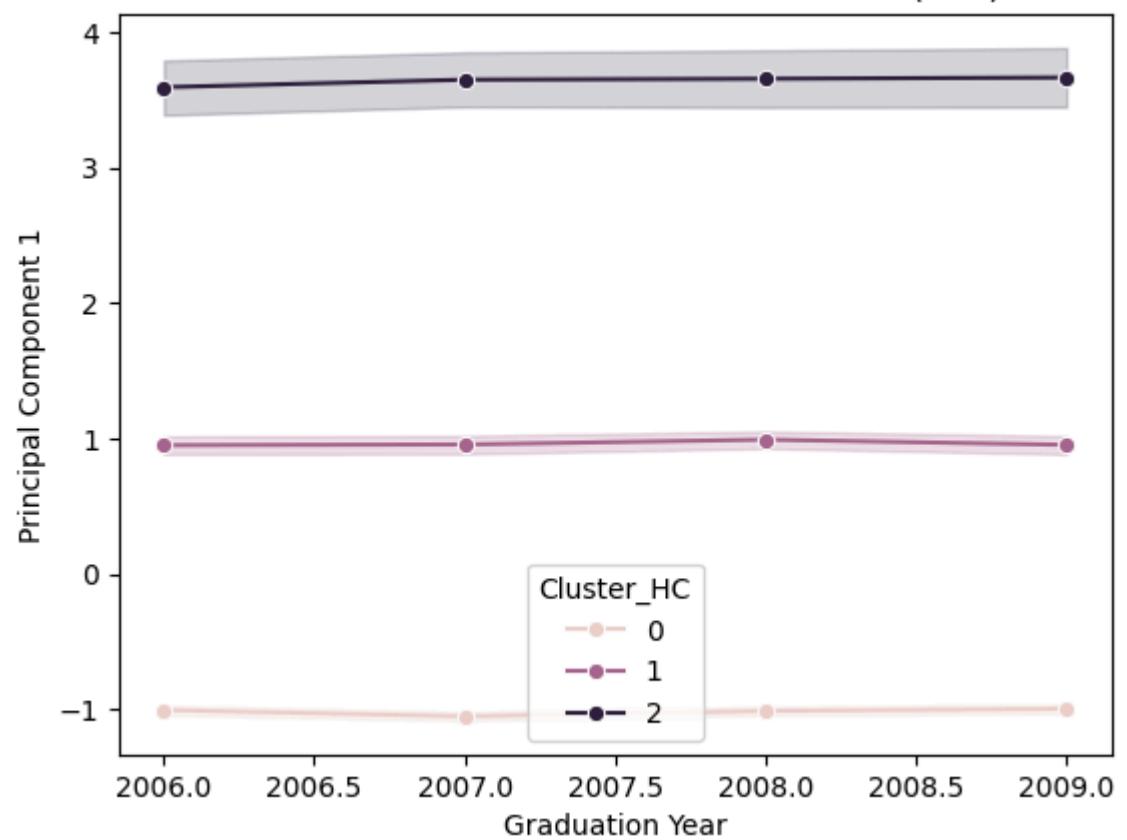
# APlotting of KMeans clusters over time (gradyear)
sns.lineplot(data=df_pca, x='Gradyear', y='PC1', hue='Cluster_KMeans', marker='o')
plt.title('Trend of KMeans Clusters Over Time (PC1)')
plt.xlabel('Graduation Year')
plt.ylabel('Principal Component 1')
plt.show()

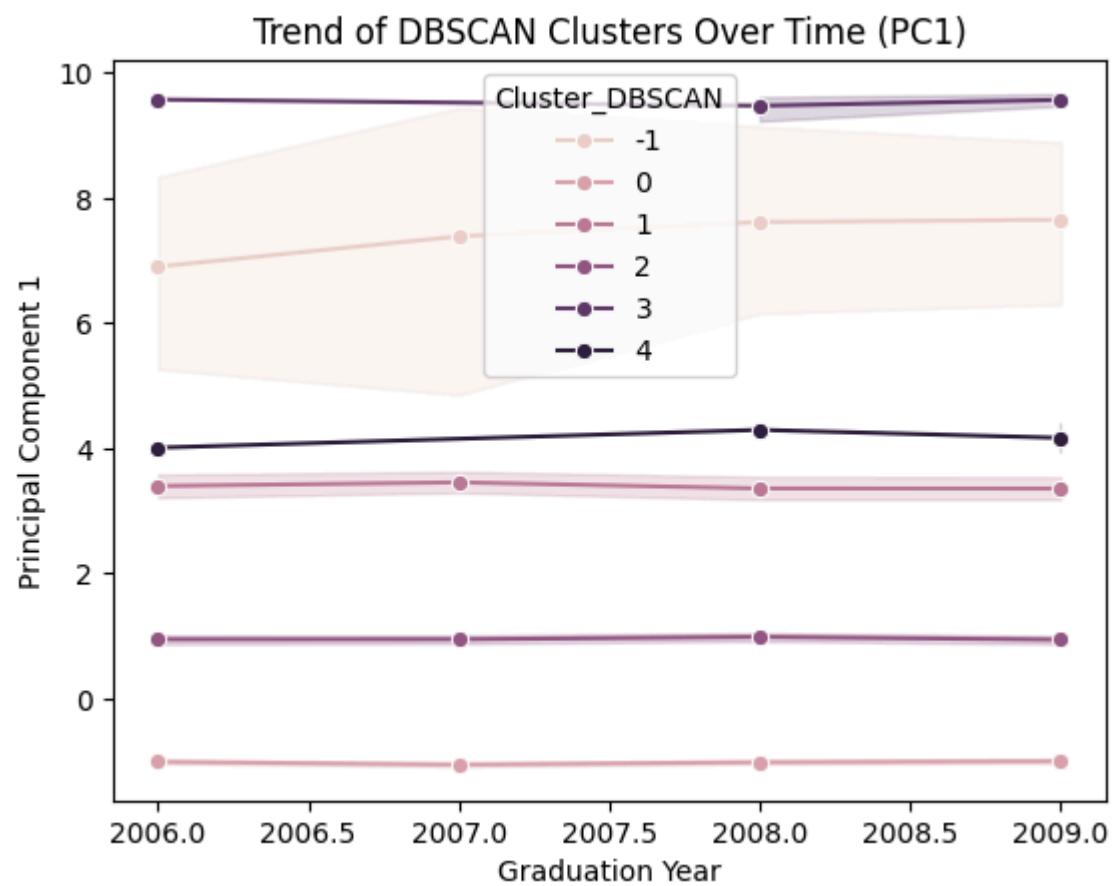
# Similarly for Hierarchical Clusters
sns.lineplot(data=df_pca, x='Gradyear', y='PC1', hue='Cluster_HC', marker='o')
plt.title('Trend of Hierarchical Clusters Over Time (PC1)')
plt.xlabel('Graduation Year')
plt.ylabel('Principal Component 1')
plt.show()
```

```
# Similarly for DBSCAN clusters
sns.lineplot(data=df_pca, x='Gradyear', y='PC1', hue='Cluster_DBSCAN', marker='o')
plt.title('Trend of DBSCAN Clusters Over Time (PC1)')
plt.xlabel('Graduation Year')
plt.ylabel('Principal Component 1')
plt.show()
```



Trend of Hierarchical Clusters Over Time (PC1)





Observation/Inference:

Cluster Evolution: The trend plot shows that one group of students (Cluster 0) was dominant in earlier years and then gradually shrank in more recent years, while another group (Cluster 1) grew in size.

Shifts in Interests: For instance, we might find that in earlier years, students showed more interest in sports-related activities, while in more recent years, there's a noticeable increase in interest in shopping and fashion-related activities.

Demographic Shifts: Over time, we may observe how the demographic profile of each cluster changes. For example, the average age of students in Cluster 0 might decrease over the years, while Cluster 1 could see a rise in younger students.

Cluster based Analysis:

Group 1: Sports Enthusiasts (Cluster 0)

Targeted Program: A sports program, such as intramural sports leagues, fitness challenges, or sports-related events (e.g., tournaments or fan clubs for teams). Marketing Opportunity: Promote school sporting events or partner with athletic brands. Outreach: Encourage these students to join the school's athletic teams or participate in physical fitness programs.

Group 2: Fashion and Shopping Aficionados (Cluster 1)

Targeted Program: Fashion-related clubs, such as a fashion design club or mall-based field trips. Organize fashion shows or events related to popular clothing brands (like Hollister, Abercrombie). Marketing Opportunity: Collaborate with fashion brands for brand ambassadors or discounts. Outreach: Engage them through student fashion showcases or partnerships with fashion-related businesses for school events.

Group 3: Music Lovers & Performers (Cluster 2)

Targeted Program: Music programs, band, choir, or performing arts clubs. Organize concerts or music festivals. Marketing Opportunity: Collaborate with local music venues or brands targeting musical events and gear. Outreach: Encourage them to participate in the school's marching band, drama clubs, or musical productions.

In []: