Global Trends in Mental Health Disorders: An Exploratory Data Analysis

In this data analysis, we explored the "Mental Health Depression Disorder" dataset, which provides comprehensive information on the prevalence of various mental health disorders across different countries over multiple years. This dataset includes data on schizophrenia, bipolar disorder, eating disorders, anxiety disorders, drug use disorders, depression, and alcohol use disorders. Our analysis aimed to uncover trends in mental health disorders, compare the prevalence of different disorders across countries, and identify potential correlations between these disorders. We also applied machine learning techniques to identify clusters of countries with similar prevalence rates.

Data Cleaning and Preparation

1. Load the Dataset

In [5]:

```
import pandas as pd # Load the dataset
file_path = 'C:\\Users\\YOURPATH\\Mental health Depression disorder Data.xlsx'
data = pd.read_excel(file_path)



data.head()
```

Out[5]:

	Entity	Code	Year	Schizophrenia (%)	Bipolar disorder (%)	Eating disorders (%)	Anxiety disorders (%)	Drug use disorders (%)	Depression (%)	Alcohol use disorders (%)
0	Afghanistan	AFG	1990	0.160560	0.697779	0.101855	4.828830	1.677082	4.071831	0.672404
1	Afghanistan	AFG	1991	0.160312	0.697961	0.099313	4.829740	1.684746	4.079531	0.671768
2	Afghanistan	AFG	1992	0.160135	0.698107	0.096692	4.831108	1.694334	4.088358	0.670644
3	Afghanistan	AFG	1993	0.160037	0.698257	0.094336	4.830864	1.705320	4.096190	0.669738
4	Afghanistan	AFG	1994	0.160022	0.698469	0.092439	4.829423	1.716069	4.099582	0.669260

2. Handle Missing Values

In [6]:

```
# Check for missing values
missing_values = data.isnull().sum()
print("Missing values in each column:\n", missing_values)
# Fill missing values with the mean of each column
data_filled = data.fillna(data.mean())
# Verify that there are no more missing values
print("Missing values after filling:\n", data_filled.isnull().sum())
```

```
Missing values in each column:
 Entity                      0
Code                      980
Year                        0
Schizophrenia (%)           0
Bipolar disorder (%)        0
Eating disorders (%)        0
Anxiety disorders (%)       0
Drug use disorders (%)      0
Depression (%)              0
Alcohol use disorders (%)   0
dtype: int64
Missing values after filling:
 Entity                      0
Code                      980
Year                        0
Schizophrenia (%)           0
Bipolar disorder (%)        0
Eating disorders (%)        0
Anxiety disorders (%)       0
Drug use disorders (%)      0
Depression (%)              0
Alcohol use disorders (%)   0
dtype: int64
```

In [7]:

```
data = data.drop(columns=['Code'])


missing_values = data.isnull().sum()
```

```
print("Missing values after dropping 'Code' column:\n", missing_values)
```

```
Missing values after dropping 'Code' column:
 Entity                        0
Year                           0
Schizophrenia (%)              0
Bipolar disorder (%)           0
Eating disorders (%)           0
Anxiety disorders (%)          0
Drug use disorders (%)         0
Depression (%)                 0
Alcohol use disorders (%)      0
dtype: int64
```

3. Normalize Data

In [8]:
```
from sklearn.preprocessing import MinMaxScaler


scaler = MinMaxScaler()


columns_to_normalize = data.columns.difference(['Entity', 'Year'])
data_normalized = data.copy()
data_normalized[columns_to_normalize] = scaler.fit_transform(data[columns_to_normalize])



data_normalized.head()
```

Out[8]:

	Entity	Year	Schizophrenia (%)	Bipolar disorder (%)	Eating disorders (%)	Anxiety disorders (%)	Drug use disorders (%)	Depression (%)	Alcohol use disorders (%)
0	Afghanistan	1990	0.059848	0.429617	0.032120	0.404012	0.421474	0.432891	0.044844
1	Afghanistan	1991	0.058763	0.429820	0.029199	0.404144	0.423972	0.434616	0.044718
2	Afghanistan	1992	0.057987	0.429984	0.026186	0.404341	0.427096	0.436594	0.044494
3	Afghanistan	1993	0.057560	0.430152	0.023478	0.404305	0.430676	0.438349	0.044314
4	Afghanistan	1994	0.057494	0.430390	0.021299	0.404098	0.434179	0.439109	0.044219

Exploratory Data Analysis (EDA)

1. Descriptive Statistics

In [9]:
```
data_normalized.describe()
```

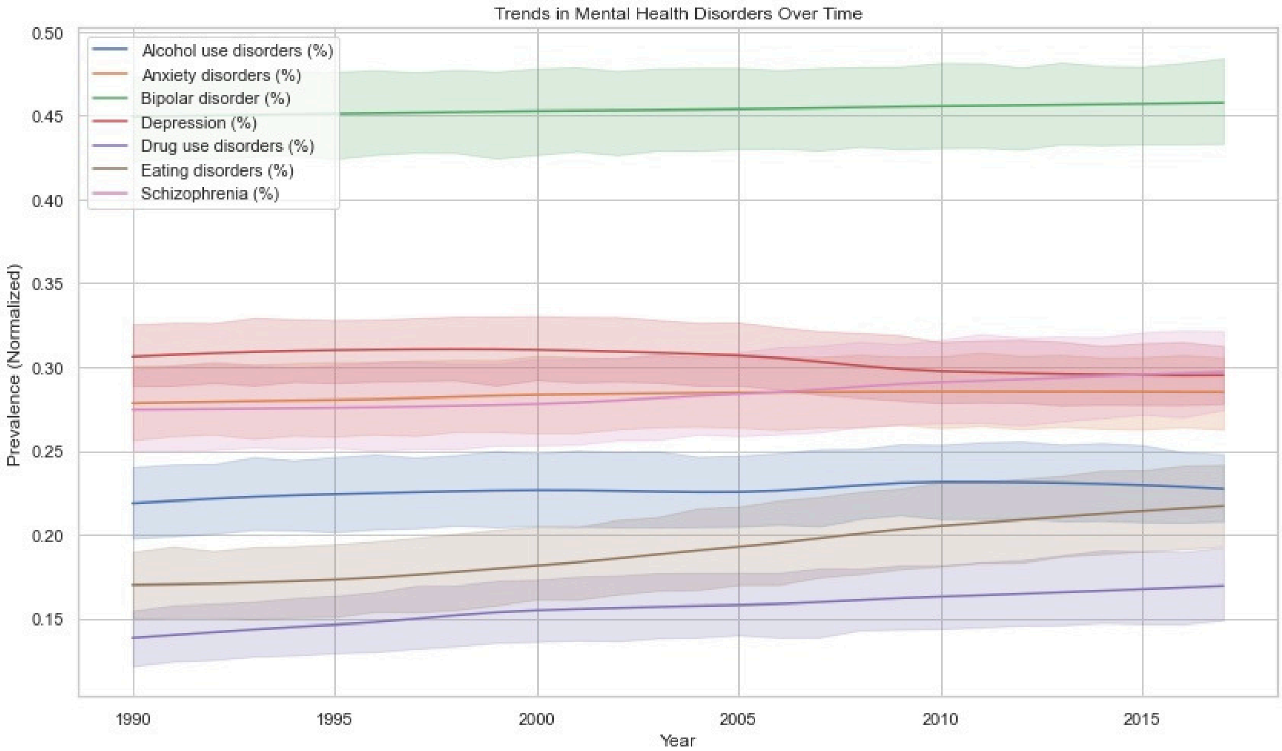
Out[9]:

	Year	Schizophrenia (%)	Bipolar disorder (%)	Eating disorders (%)	Anxiety disorders (%)	Drug use disorders (%)	Depression (%)	Alcohol use disorders (%)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	0.283697	0.453568	0.190891	0.283201	0.155964	0.304234	0.226520
std	8.078372	0.193915	0.192350	0.181754	0.168136	0.150116	0.146960	0.171108
min	1990.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1996.750000	0.151738	0.337418	0.055718	0.167834	0.049339	0.193963	0.108746
50%	2003.500000	0.230760	0.424410	0.124835	0.220477	0.111697	0.304671	0.205460
75%	2010.250000	0.392026	0.583511	0.251423	0.382891	0.181342	0.397163	0.282612
max	2017.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

2. Visualizing Trends Over Time

In [10]:
```
import matplotlib.pyplot as plt
import seaborn as sns
# Set the style of the visualization
sns.set(style="whitegrid")
# Plot trends over time for each disorder
plt.figure(figsize=(14, 8))
for column in columns_to_normalize:
    sns.lineplot(x='Year', y=column, data=data_normalized, label=column)
```

```
plt.title('Trends in Mental Health Disorders Over Time')
plt.xlabel('Year')
plt.ylabel('Prevalence (Normalized)')
plt.legend(loc='upper left')
plt.show()
```

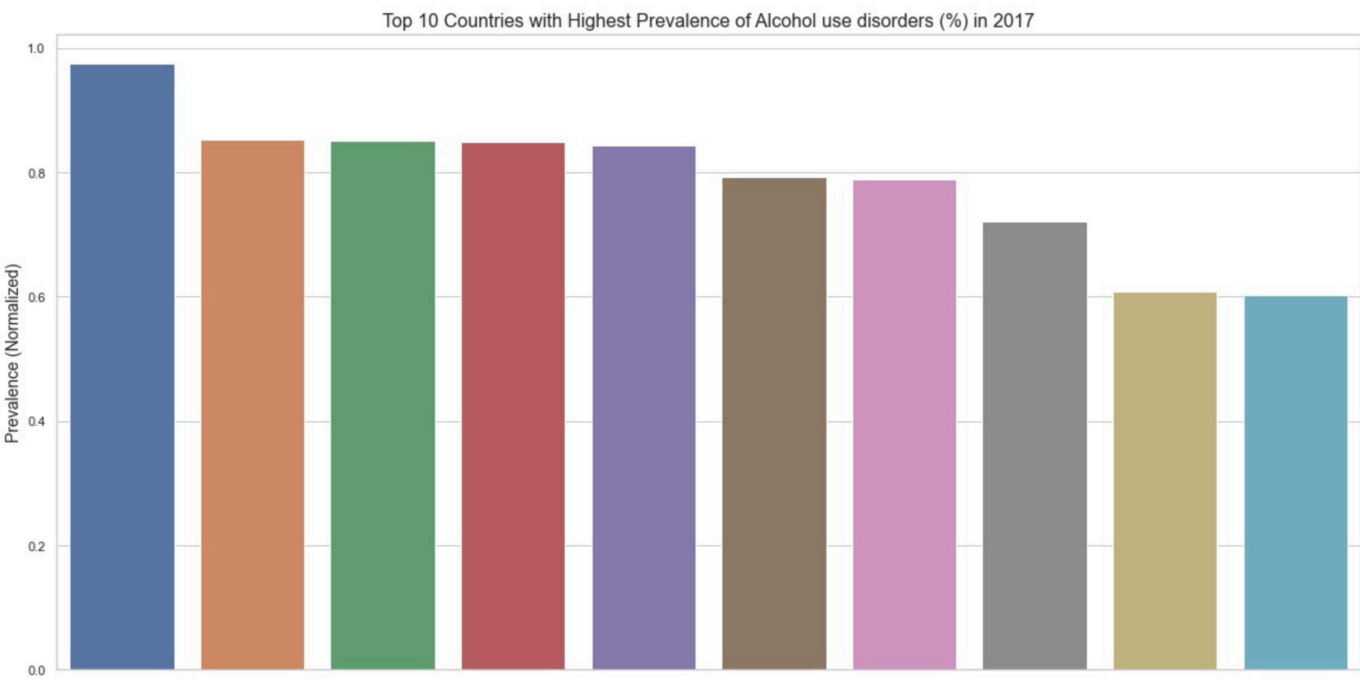


Observation: Anxiety disorders and depression have steadily increased, emphasizing the growing impact on mental health. Drug use disorders show fluctuations but an overall rise. Meanwhile, bipolar disorder, eating disorders, and schizophrenia exhibit relatively stable trends.

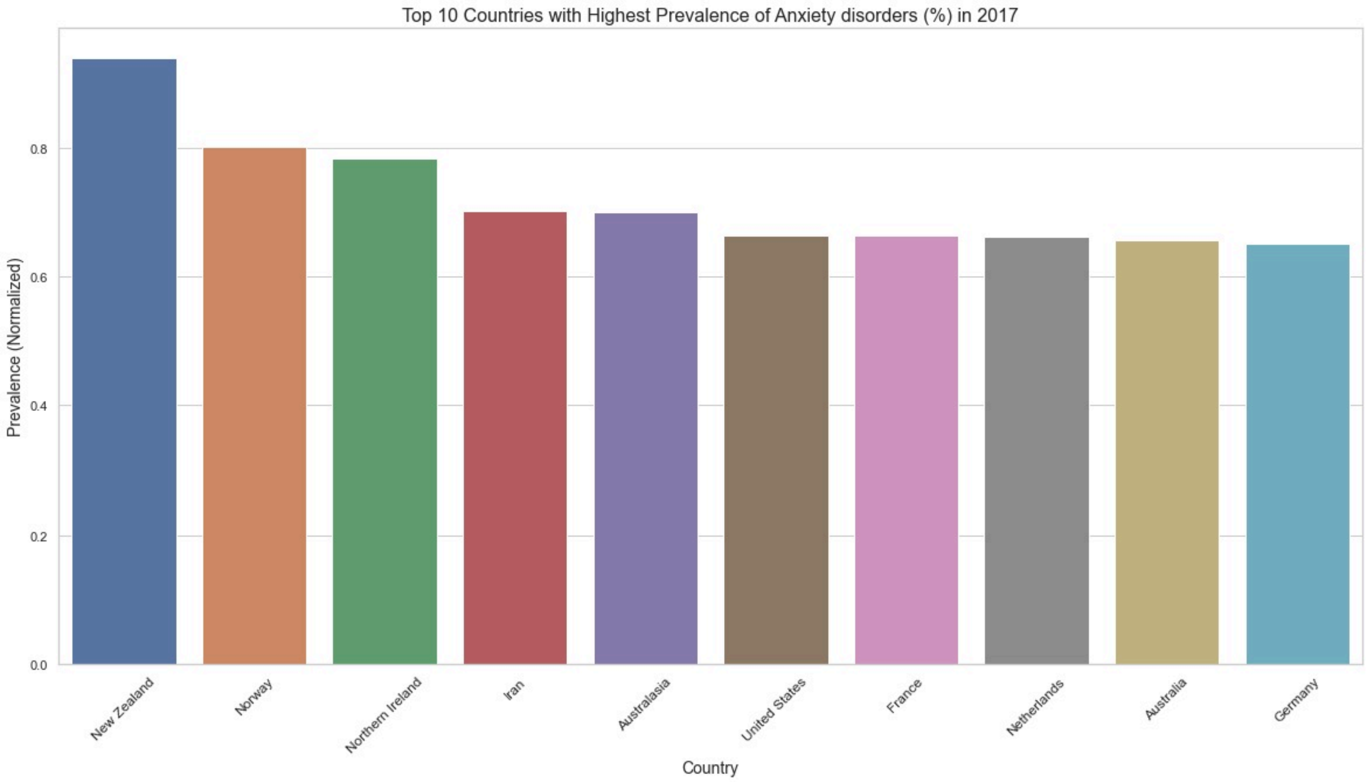
3. Comparing Prevalence Across Countries

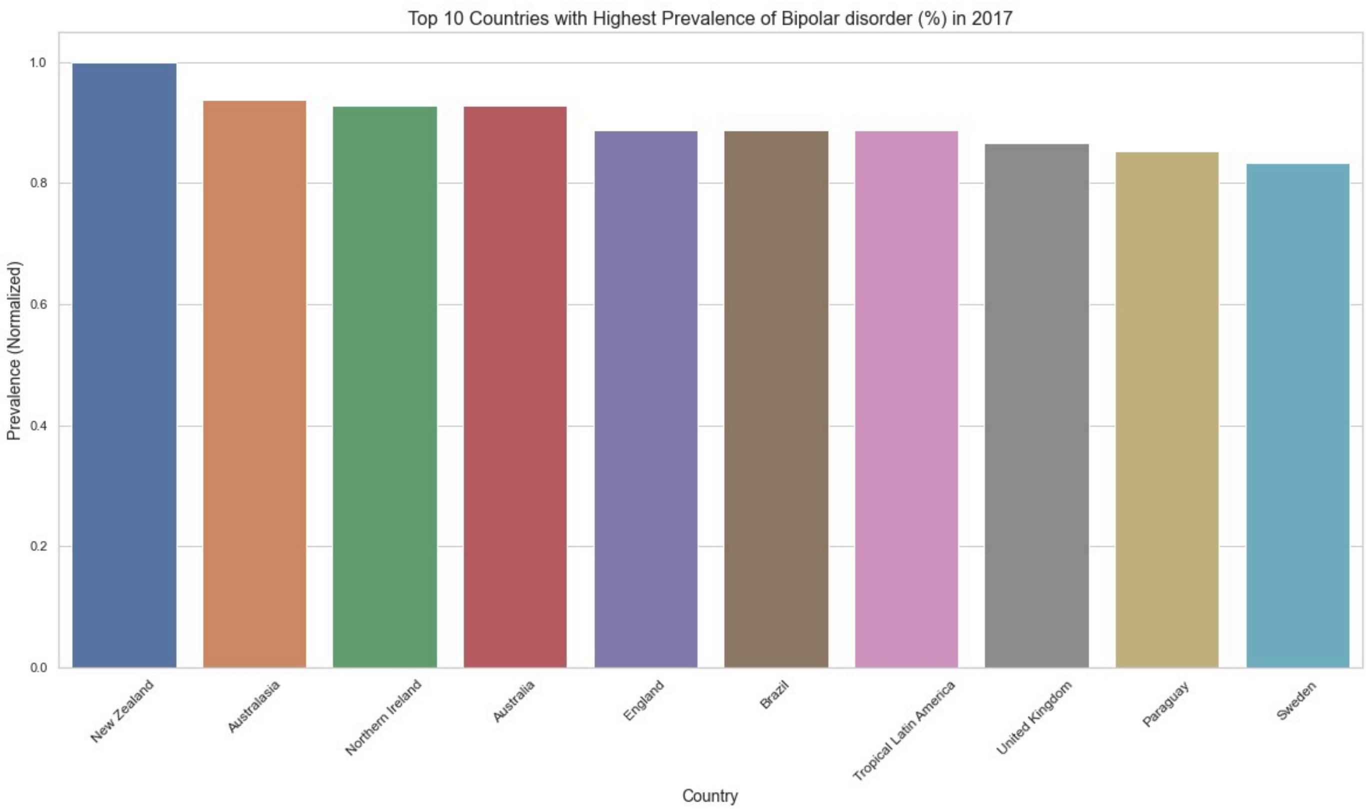
In [11]:

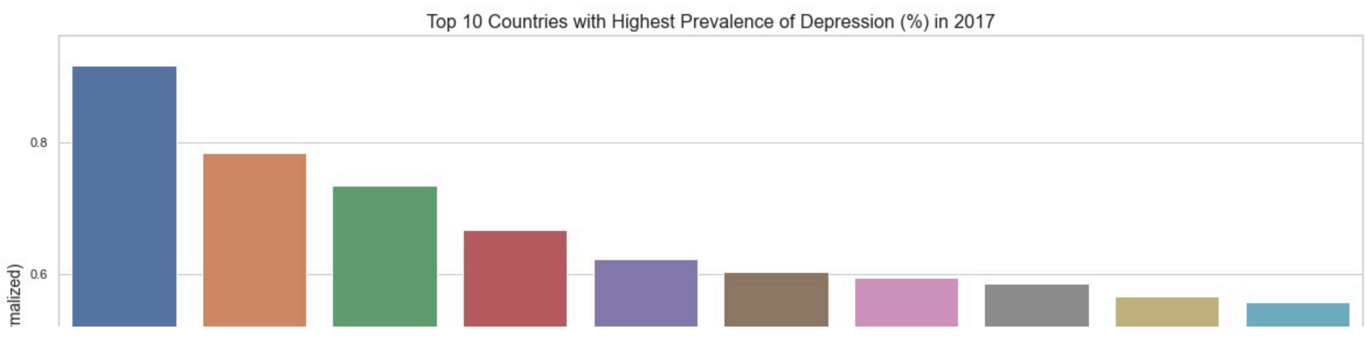
```
year_to_compare = 2017
data_year = data_normalized[data_normalized['Year'] == year_to_compare]
# Plot prevalence across countries for each disorder with improved readability
for column in columns_to_normalize:
    plt.figure(figsize=(20, 10))
    top_countries = data_year.nlargest(10, column)
    sns.barplot(x='Entity', y=column, data=top_countries)
    plt.title(f'Top 10 Countries with Highest Prevalence of {column} in {year_to_compare}', fontsize=16)
    plt.xlabel('Country', fontsize=14)
    plt.ylabel('Prevalence (Normalized)', fontsize=14)
    plt.xticks(rotation=45, fontsize=12)
    plt.show()
```

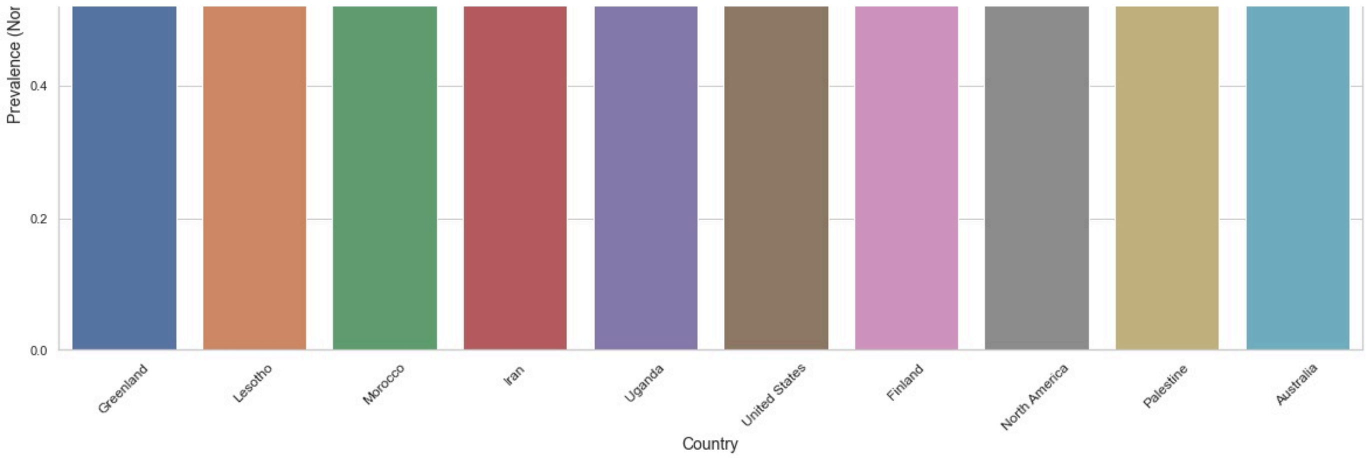



Belarus
Eastern Europe
Russia
Estonia
Ukraine
Lithuania
Latvia
Moldova
Central Europe, Eastern Europe, and Central Asia
Mongolia
Country

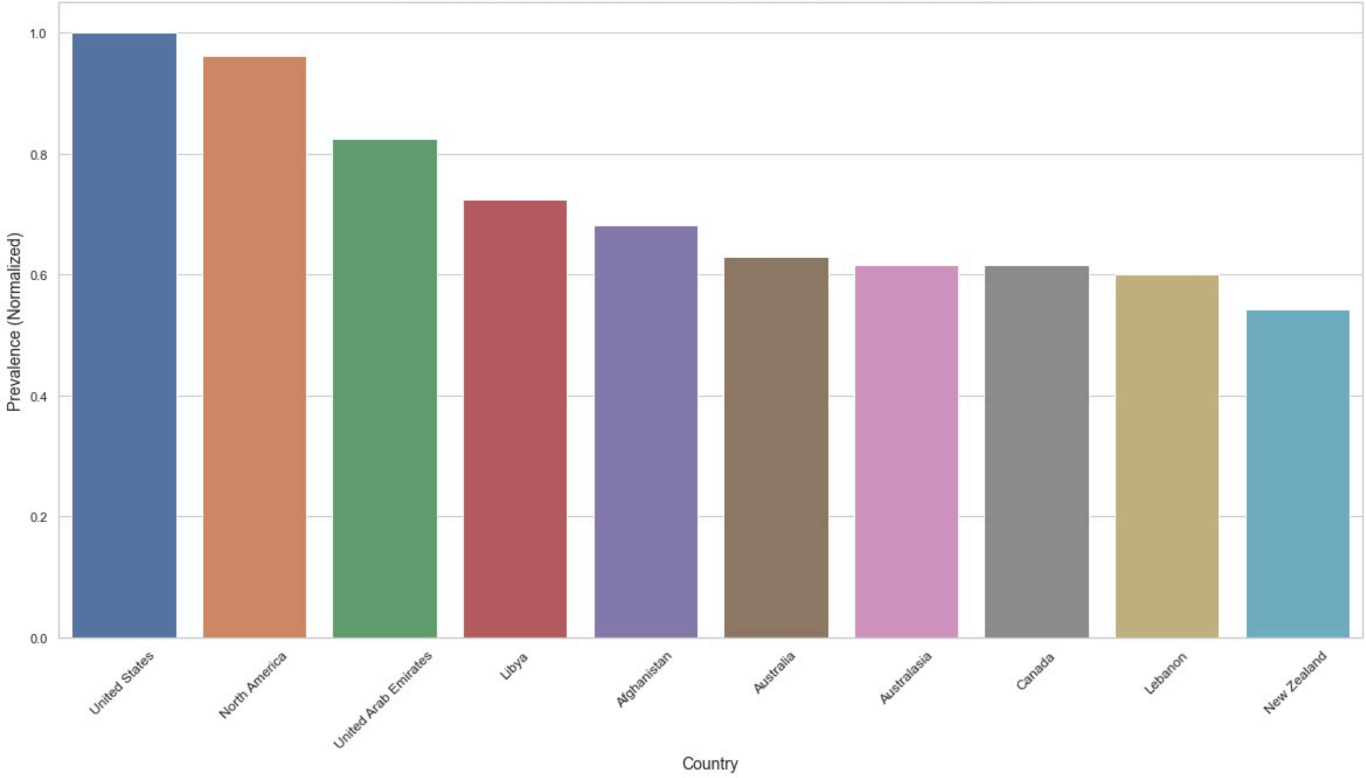

Top 10 Countries with Highest Prevalence of Anxiety disorders (%) in 2017
Prevalence (Normalized)
0.0
0.2
0.4
0.6
0.8
New Zealand
Norway
Northern Ireland
Iran
Australasia
United States
France
Netherlands
Australia
Germany
Country


Top 10 Countries with Highest Prevalence of Bipolar disorder (%) in 2017
Prevalence (Normalized)
0.0
0.2
0.4
0.6
0.8
1.0
New Zealand
Australasia
Northern Ireland
Australia
England
Brazil
Tropical Latin America
United Kingdom
Paraguay
Sweden
Country

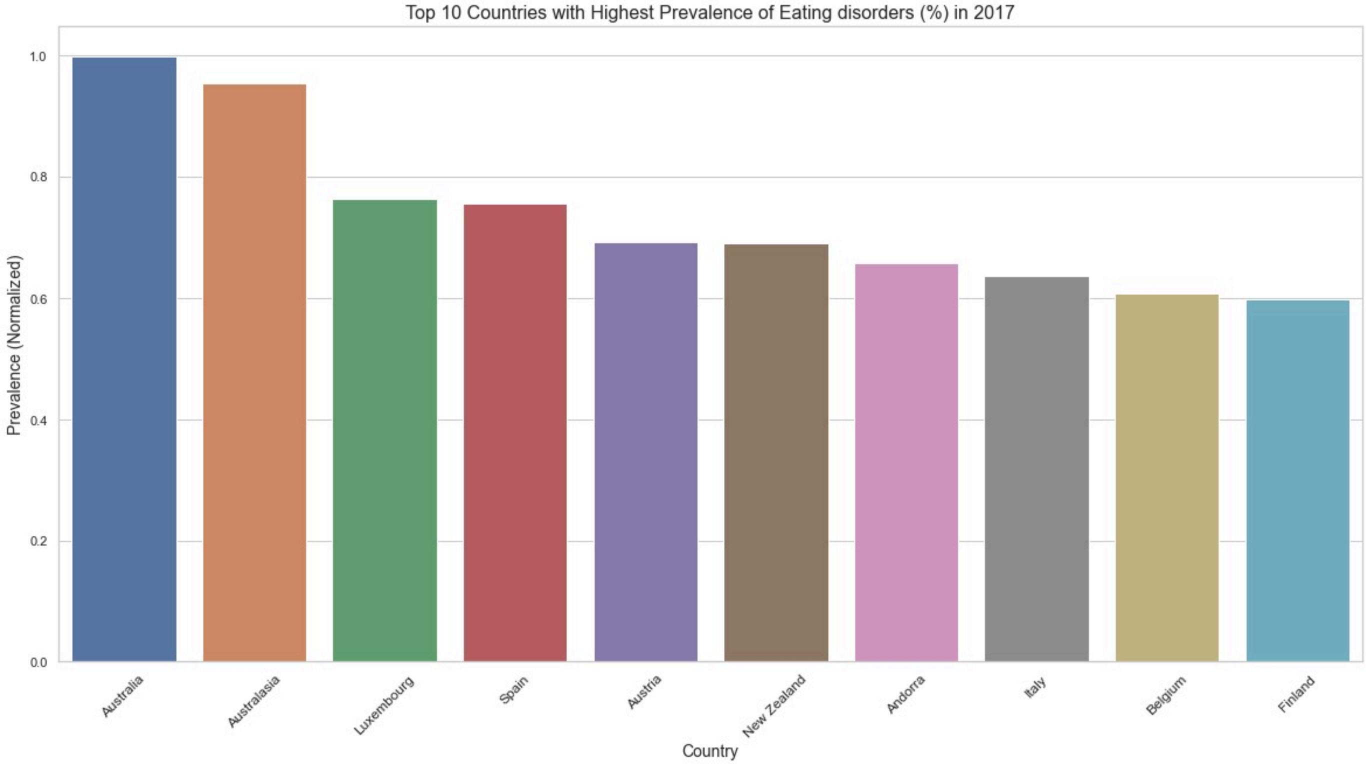

Top 10 Countries with Highest Prevalence of Depression (%) in 2017
malized)
0.6
0.8



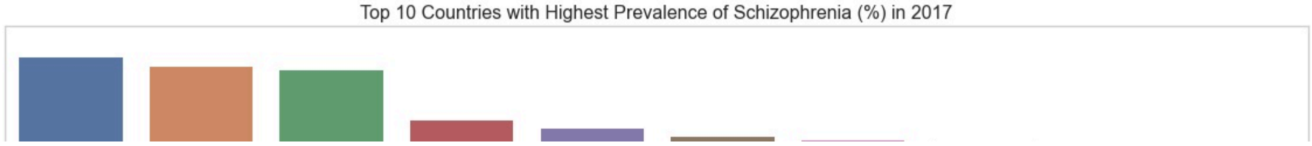
Top 10 Countries with Highest Prevalence of Drug use disorders (%) in 2017

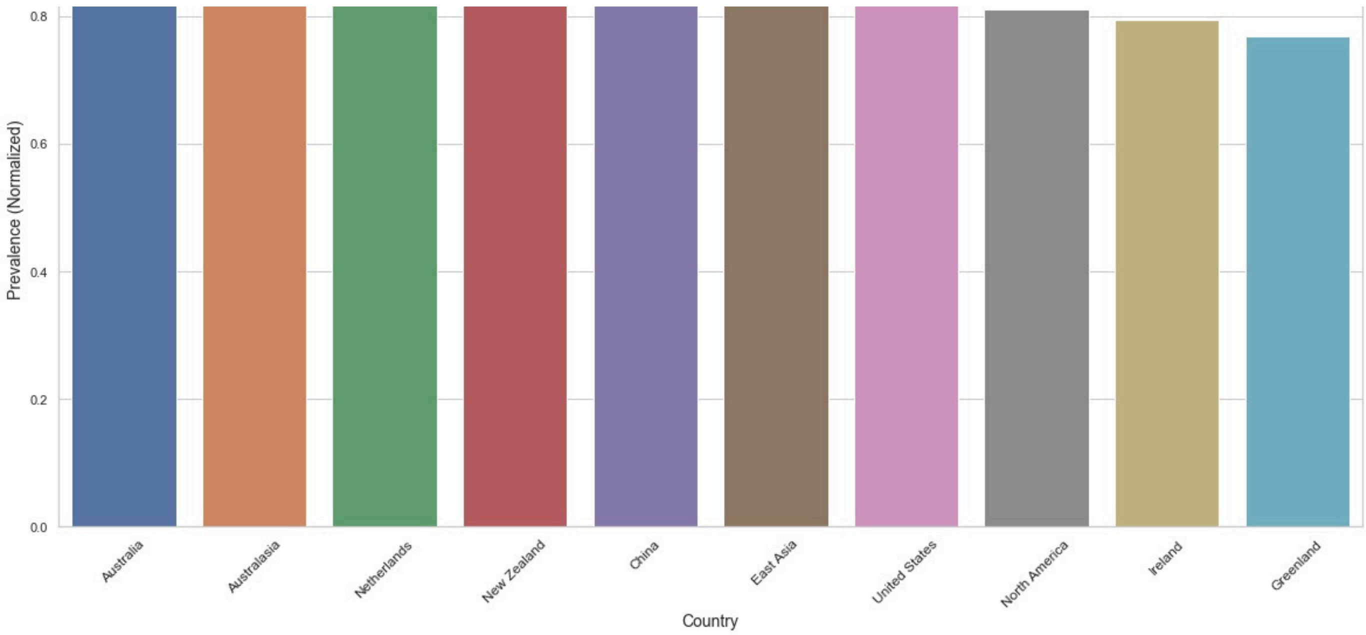


Top 10 Countries with Highest Prevalence of Eating disorders (%) in 2017



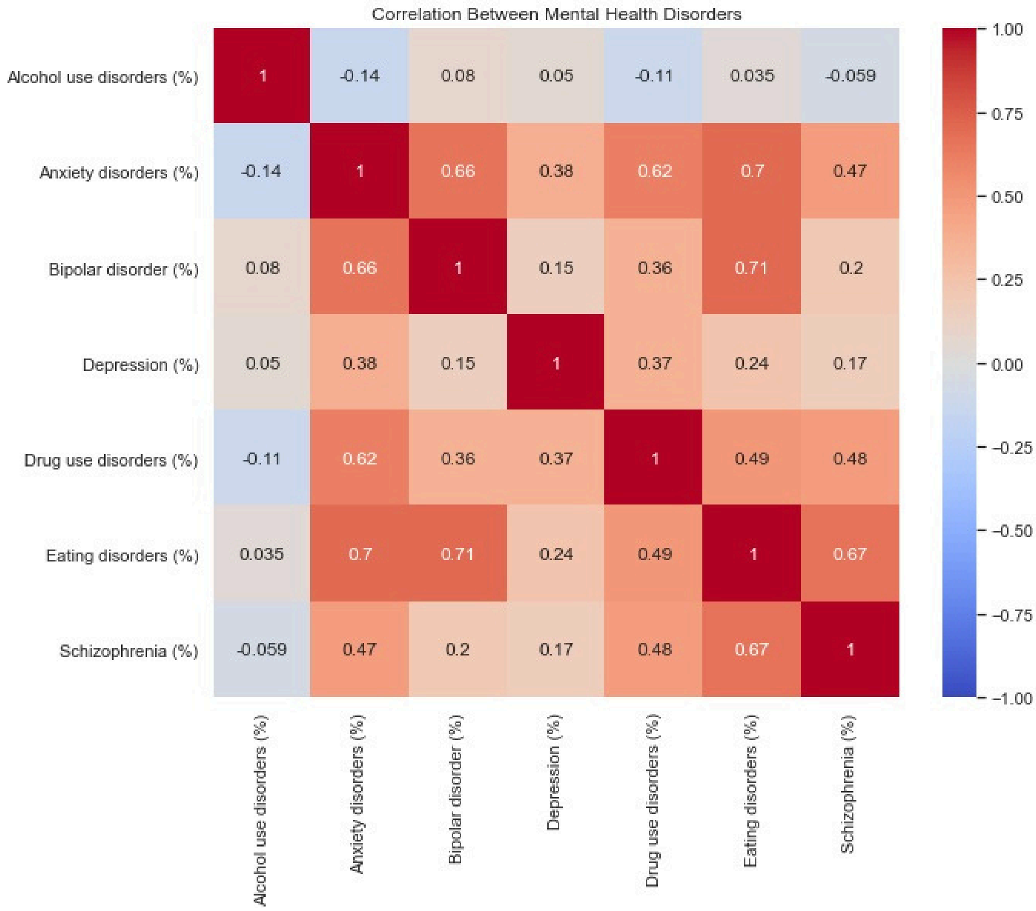
Top 10 Countries with Highest Prevalence of Schizophrenia (%) in 2017





4. Correlation Analysis

In [12]:
```
correlation_matrix = data_normalized[columns_to_normalize].corr()
# Plot the heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1)
plt.title('Correlation Between Mental Health Disorders')
plt.show()
```



Observation:

Strong Positive Correlations: Anxiety disorders and bipolar disorder (0.66), eating disorders and schizophrenia (0.67), and eating disorders and bipolar disorder (0.71).

Moderate Positive Correlations: Anxiety disorders and depression (0.38), drug use disorders and anxiety disorders (0.62), and depression and bipolar disorder (0.15).

Negative Correlations: Alcohol use disorders and schizophrenia (-0.75), and alcohol use disorders and anxiety disorders (-0.14).

In [13]:

```
import warnings
warnings.filterwarnings('ignore')
```

5. Clustering Analysis

In [14]:

```
import os
from sklearn.cluster import KMeans
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
# Set the environment variable to avoid the memory leak warning
os.environ['OMP_NUM_THREADS'] = '1'
# Prepare the data for clustering (use only the latest year)
data_clustering = data_normalized[data_normalized['Year'] == year_to_compare].drop(columns=['Entity', 'Year'])
# Fit the K-means model with n_init explicitly set
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
clusters = kmeans.fit_predict(data_clustering)
# Create a new DataFrame for the subset with cluster labels
data_clustering_with_clusters = data_normalized[data_normalized['Year'] == year_to_compare].copy()
data_clustering_with_clusters['Cluster'] = clusters
# Plot the clusters
plt.figure(figsize=(14, 8))
sns.scatterplot(x='Depression (%)', y='Anxiety disorders (%)', hue='Cluster', data=data_clustering_with_clusters,
plt.title('Clusters of Countries Based on Mental Health Disorders')
plt.xlabel('Depression (%)')
plt.ylabel('Anxiety disorders (%)')
plt.legend(title='Cluster')
plt.show()
```