Happiness & Alcohol Consumption

Is alcohol consumption related to a person's happiness? Data from countries across the world is analyzed to find if there is a relationship between happiness and consumption of different types of alcohol.

Contents

1. Importing libraries

```python
import numpy as np
import pandas as pd
import seaborn as sns

import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('ggplot')

from wordcloud import WordCloud

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

2. Loading dataset

```python
df = pd.read_csv(r"G:\EDA - Kaggle\Happiness and Alcohol
Consumption.csv")
df.head()
```

```
       Country          Region Hemisphere  HappinessScore  HDI
GDP_PerCapita  \
0      Denmark  Western Europe      north           7.526  928
53.579
1  Switzerland  Western Europe      north           7.509  943
79.866
2      Iceland  Western Europe      north           7.501  933
60.530
3       Norway  Western Europe      north           7.498  951
```

```
70.890
4      Finland  Western Europe       north            7.413  918
43.433

   Beer_PerCapita  Spirit_PerCapita  Wine_PerCapita
0             224                81             278
1             185               100             280
2             233                61              78
3             169                71             129
4             263               133              97
```

3. Data Dictionary

1. **Country** - Name of the Country
2. **Region** - Region the Country belongs to
3. **Hemisphere** - hemisphere of the Country
4. **Happiness Score** - A metric measured in 2016 by asking the sampled people the question: "How would you rate your happiness on a scale of (1-10)
5. **HDI** - Human Development Index by United Nations Development Programme
6. **GDP_PerCapita** - Gross Domestic Product index
7. **Beer_PerCapita** - Liters (per capita) of beer consumption
8. **Spirit_PerCapita** - Consumption of spirits drink (per capita)
9. **Wine_PerCapita** - Wine consumption

4. Basics - null/ duplicate values, cleaning

```
# checking basic information
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 122 entries, 0 to 121
Data columns (total 9 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Country           122 non-null    object
 1   Region            122 non-null    object
 2   Hemisphere        122 non-null    object
 3   HappinessScore    122 non-null    float64
 4   HDI               122 non-null    int64
 5   GDP_PerCapita     122 non-null    float64
 6   Beer_PerCapita    122 non-null    int64
 7   Spirit_PerCapita  122 non-null    int64
 8   Wine_PerCapita    122 non-null    int64
dtypes: float64(2), int64(4), object(3)
memory usage: 8.7+ KB
```

```
print(df.isnull().sum())
print('\n')
print(df.duplicated().any())
```

```
Country              0
Region               0
Hemisphere           0
HappinessScore       0
HDI                  0
GDP_PerCapita        0
Beer_PerCapita       0
Spirit_PerCapita     0
Wine_PerCapita       0
dtype: int64
```

```
False
```

- There are no null or duplicate values present in our dataset.

```
# converting columns from upper to lowercase
df = df.rename(columns = dict(zip(df.columns, [c.lower() for c in
df.columns])))

# unique values for hemisphere
df.hemisphere.unique()

# changing 'noth' to 'north'
df.loc[df.hemisphere == 'noth', 'hemisphere'] = 'north'
```

5. Exploratory Data Analysis

1. List of countireis grouped by region.
```
grouped_region = df.groupby(['region']).sum()

# highlighting max and min values for each attribute
grouped_region.style\
        .highlight_max(color = 'lightgreen')\
        .highlight_min(color = 'coral')
```

```
<pandas.io.formats.style.Styler at 0x1d09ba788b0>
```

Conclusoions :

1. Happiness_score and HDI is highest in Central and Eastern Europe while North America has the lowest score in these attributes.
2. From the above table we can see that Sub-Saharan Africa has the highest gdp_percapita. This might be because majority of the countries in our dataset belongs to this region.

```
df.region.value_counts()
```

```
Sub-Saharan Africa               28
Central and Eastern Europe       27
Latin America and Caribbean      23
Western Europe                   20
Middle East and Northern Africa  11
Southeastern Asia                 5
Eastern Asia                      4
North America                     2
Australia and New Zealand         2
Name: region, dtype: int64
```

```
df.loc[(df.region == 'North America')]
```

```
        country         region hemisphere  happinessscore  hdi  \
5        Canada  North America      north           7.404  922
12  United States  North America    north           7.104  922

    gdp_percapita  beer_percapita  spirit_percapita  wine_percapita
5          42.349             240               122             100
12         57.589             249               158              84
```

- We have only 2 countries of North American region.
- Clearly our dataset is not evenly distributed. Anyway we'll continue with this as we don't have any other option.

2. Checking the Happiness Score of different Regions
```
import matplotlib.patches as mpatches

# plotting
fig, ax = plt.subplots(figsize = (16,6), dpi = 100)
ax = sns.barplot(x = 'happinessscore',
                 y = 'region',
                 data = df)
# plotting mean
happiness_mean = round(df.happinessscore.mean(), 4)
plt.axvline(df.happinessscore.mean(),
            color = 'r',
            linestyle = 'dashed',
            label = "mean = " + str(happiness_mean))

# labelling
plt.legend(fontsize = 15)
plt.xlabel('Happiness score')
plt.title('Happiness Score of different Regions')
plt.show()
```
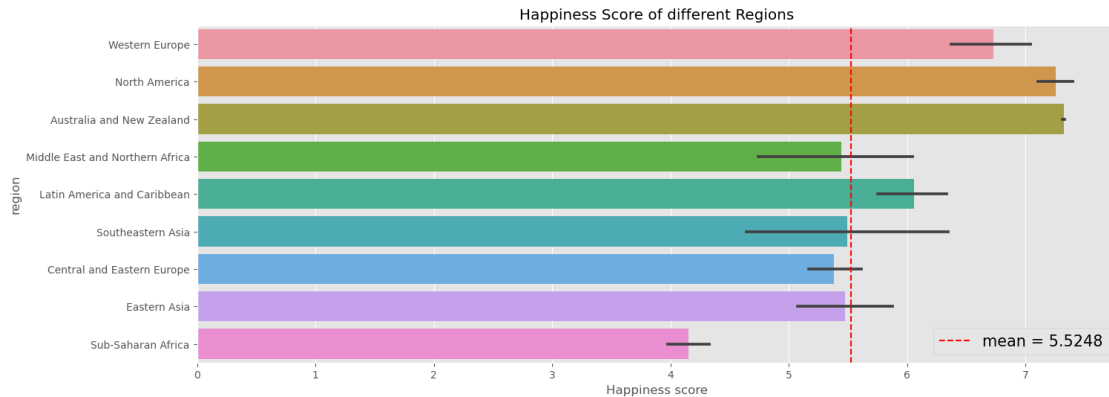
Happiness Score of different Regions

- • Australians, New Zealanders and North Americans are the happiest people, while Sub-Saharan Africans along with Central and Eastern Europeans are the least happy people.

3. From Happiest to least happy
```
# plt.figure(figsize = (30,8), dpi = 100)
df.plot(x = 'country',
        y = 'happinessscore',
        kind = 'bar',
        figsize = (30,8))

# plotting mean
happiness_mean = round(df.happinessscore.mean(), 4)
plt.axhline(df.happinessscore.mean(),
            color = 'k',
            linestyle = 'dashed',
            label = "mean = " + str(happiness_mean))

plt.legend(fontsize = 20)
plt.xlabel('Countries', fontsize = 20)
plt.ylabel('Happiness Score', fontsize = 20)
plt.title('Happiness Score of all Countries', fontsize = 25)
plt.show()
```
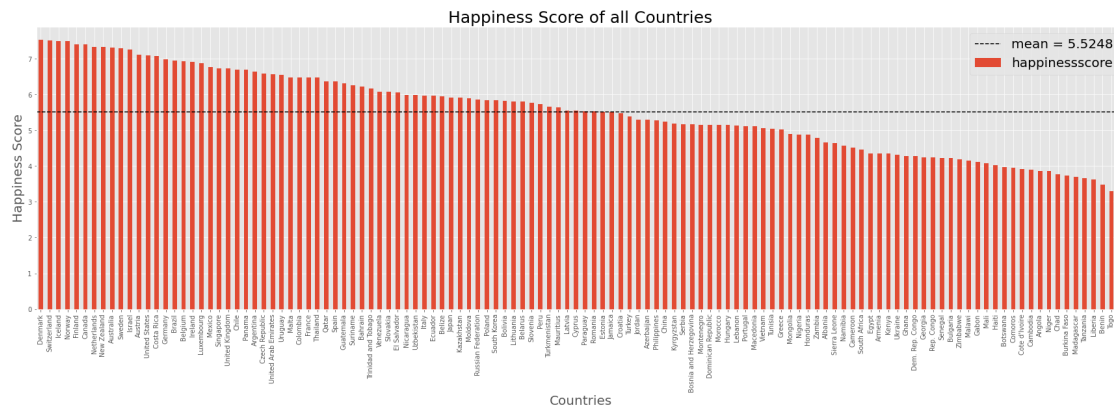

Happiness Score of all Countries

```python
fig, ax2 = plt.subplots(figsize = (16,6), dpi = 100)
ax2 = sns.barplot(x = 'happinessscore',
                  y = 'country',
                  data = df.nlargest(20, 'happinessscore'),
                  palette = 'rocket')

# labelling bar values
for values in ax2.containers:
    ax2.bar_label(values)

plt.xlabel('Happiness Score')
plt.ylabel('Countries')
plt.title('Top 10 Happiest Countries')
plt.show()
```


Top 10 Happiest Countries
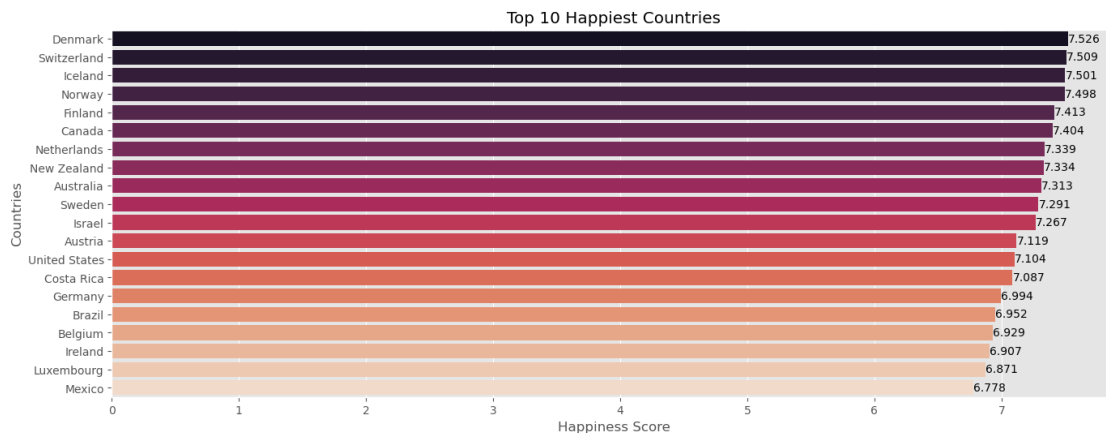
Conclusions :

1. Denmark is the happiest Country according to our dataset.
2. It can be seen that majority of the countries are either European or North American.
3. Also there are no African Countries in this list.

5. Least happy Countries

```python
least_happy = df.sort_values(by = ['happinessscore'], ascending = False).tail(5)

fig, ax3 = plt.subplots(figsize = (16,4), dpi = 100)
ax2 = sns.barplot(x = 'country',
                  y = 'happinessscore',
                  data = least_happy,
                  palette = 'rocket')

# labelling bar values
for values in ax3.containers:
    ax3.bar_label(values)
```

```
plt.xlabel('Happiness Score')
plt.ylabel('Countries')
plt.title('Least happy Countries')
plt.show()
```
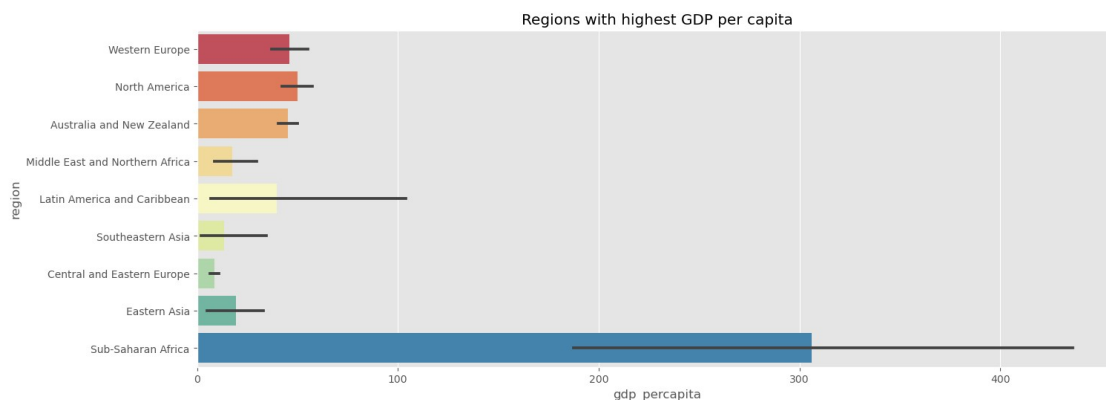

Least happy Countries
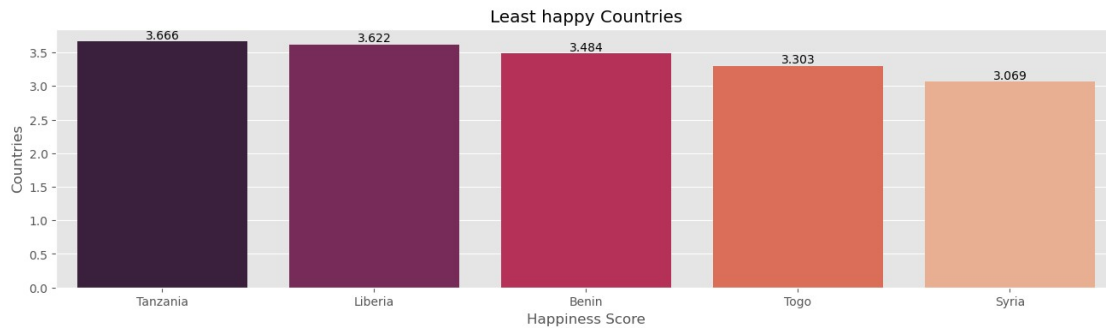
- Syria seems to be the least happy country according to our dataset.

6. Analysing Happiness_Score with respect to GDP_per_capita

```
# plotting
plt.figure(figsize = (16,6), dpi = 100)
sns.barplot(x = 'gdp_percapita',
            y = 'region',
            data = df,
            palette = 'Spectral')
plt.title('Regions with highest GDP per capita')
plt.show()
```


Regions with highest GDP per capita

Important :

There is something wrong with the GDP_per_capita column. It is not possible that African Countries have a higher GDP_per_capita than European and North American Countries.

- Therefore it makes no sense to look too much into the GDP Per Capita variable.

The HDI is a summary measure of human development. How is it defined? The HDI is a summary composite measure of a country's average achievements in three basic aspects of human development: health, knowledge and standard of living.

```
plt.figure(figsize = (16,6), dpi = 100)
sns.barplot(x = 'hdi',
            y = 'region',
            data = df,
            palette = 'mako')

# plotting mean
hdi_mean = round(df.hdi.mean(), 4)
plt.axvline(df.hdi.mean(),
            color = 'r',
            linestyle = 'dashed',
            label = "mean = " + str(hdi_mean))

plt.legend(fontsize = 10)
plt.xlabel('HDI')
plt.ylabel('Region')
plt.title('Countries with highest HDI')
plt.show()
```
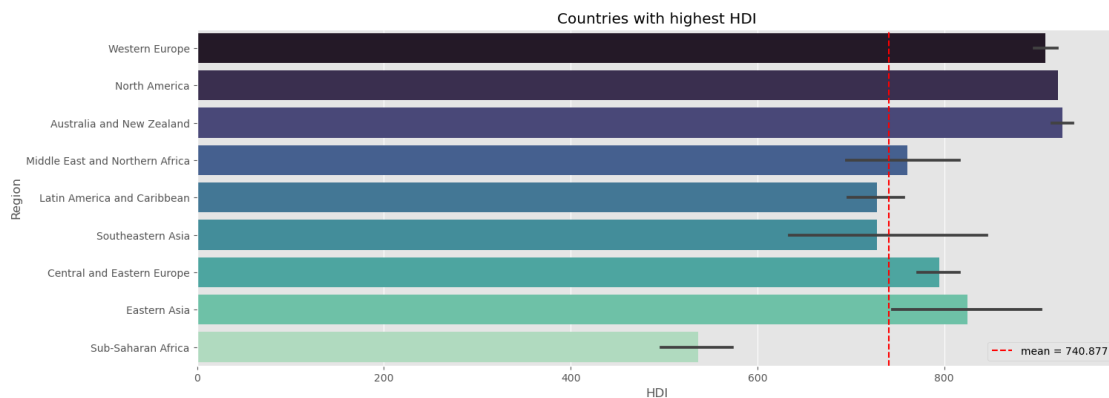

- This plot is very similar to the Happiness Score plot that we saw earlier with Western Europe being at the top and Sub-Saharan Africa at bottom.
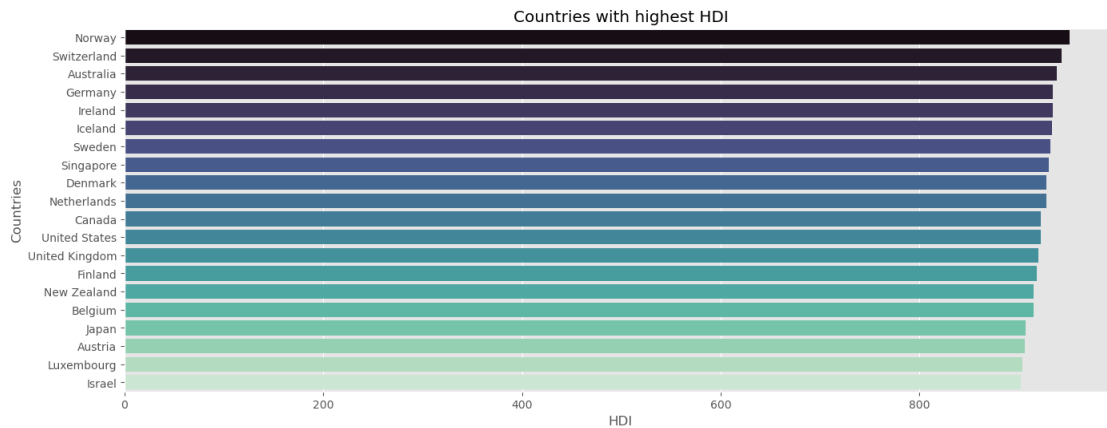
Lets bifurcate regions into countries now.

8. Countries with highest HDI
```
plt.figure(figsize = (16,6), dpi = 100)
sns.barplot(x = 'hdi',
            y = 'country',
            data = df.nlargest(20, 'hdi'),
            palette = 'mako')
plt.xlabel('HDI')
plt.ylabel('Countries')
```

```python
plt.title('Countries with highest HDI')
plt.show()
```
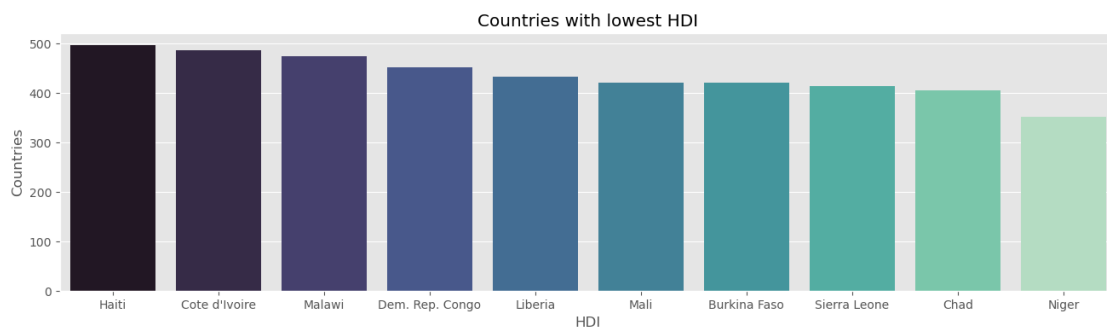

Countries with highest HDI

9. Countries with lowest HDI

```python
lowest_HDI = df.sort_values(by = ['hdi'], ascending = False).tail(10)

# plotting
plt.figure(figsize = (16,4), dpi = 100)
sns.barplot(x = 'country',
            y = 'hdi',
            data = lowest_HDI,
            palette = 'mako')

plt.xlabel('HDI')
plt.ylabel('Countries')
plt.title('Countries with lowest HDI')
plt.show()
```


Countries with lowest HDI

Conclusions :

1. Norway sits at the top of the list.
2. There are no South American or African Countries in the the top 20.
3. Niger sits at the bottom. Norway is almost 3 times the Niger value.

10. Beer per capita by Region

```
plt.figure(figsize=(16,6), dpi = 100)

sns.barplot(x = 'beer_percapita',
            y = 'region',
            data = df,
            palette = 'YlOrBr')
plt.xlabel('Beer Per Capita')
plt.ylabel('Region')
plt.title('Region with highest beer_per_capita')
plt.show()
```
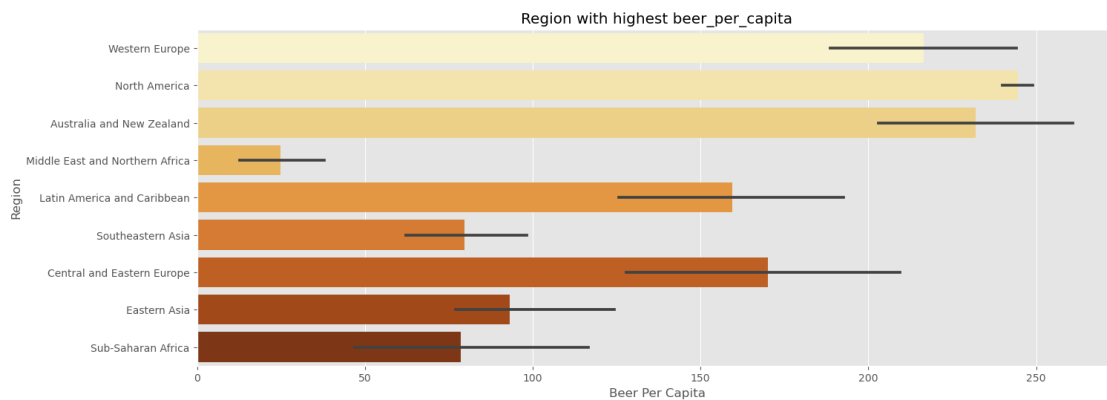

- North America, Australia and New Zealand are yet again at the top for this category too. There is a fairly big drop in rest of the regions for this category. Middle East and Northern Africa have the least consumption of Beer.

11. Beer consumption by different Countries

```
plt.figure(figsize=(16,6), dpi = 100)

sns.barplot(x = 'beer_percapita',
            y = 'country',
            data = df.nlargest(20, 'beer_percapita'),
            palette = 'YlOrBr')
plt.xlabel('Beer Per Capita')
plt.ylabel('Countries')
plt.title('Countries with highest beer_per_capita')
plt.show()
```

Countries with highest beer_per_capita
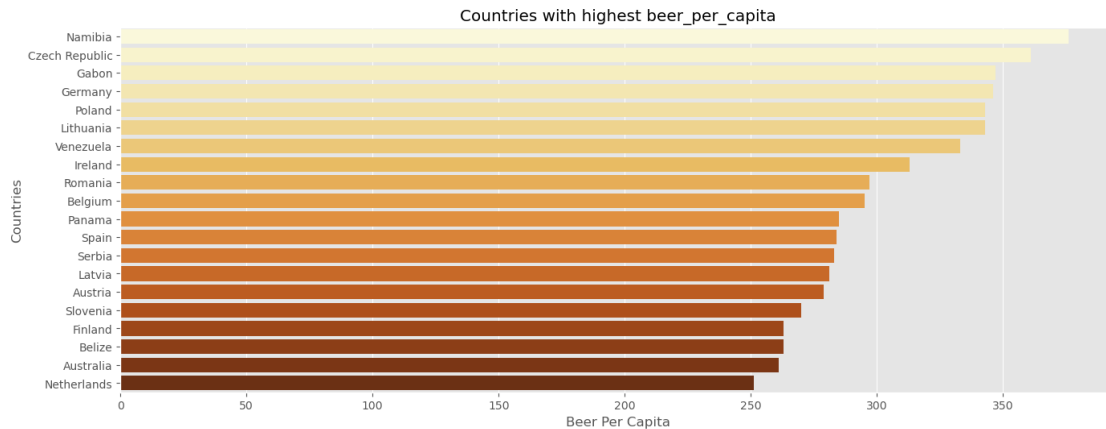
Important :-

- Again this is showing wrong information. The country that drinks the most beer per capita is the Czech Republic. It has topped the list for almost 30 years straight, since 1993. The Czechs drink on average 181.7L of beer per year per person.

12. Countries with lowest Beer_per_capita

```
lowest_beer = df.sort_values(by = ['beer_percapita'], ascending =
False).tail(10)

# plotting
plt.figure(figsize=(16,4), dpi = 100)

sns.barplot(x = 'country',
            y = 'beer_percapita',
            data = lowest_beer,
            palette = 'YlOrBr')
plt.xlabel('Beer Per Capita')
plt.ylabel('Countries')
plt.title('Countries with lowest beer_per_capita')
plt.show()
```
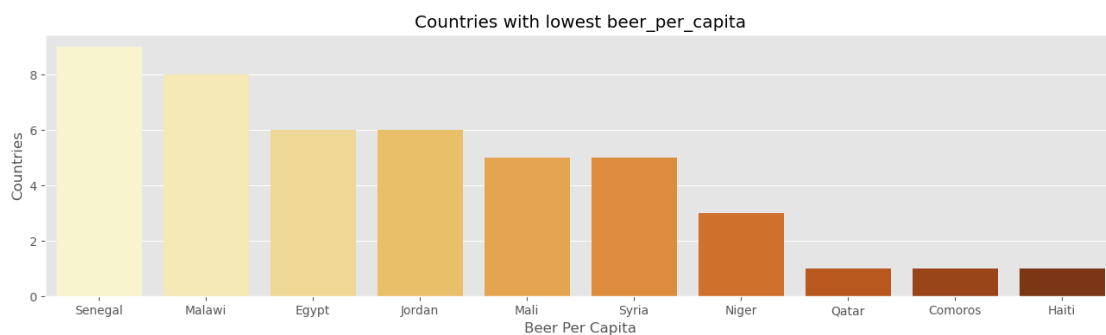

Countries with lowest beer_per_capita

- Qatar, Haiti and Comoros have the lowest beer_per_capita. Most of the countries here belongs to the Middle East region.

14. Spirit per capita in different Regions

```python
plt.figure(figsize=(16,6), dpi = 100)

sns.barplot(x = 'spirit_percapita',
            y = 'region',
            data = df,
            palette = 'icefire')
plt.xlabel('Spirit Per Capita')
plt.ylabel('Region')
plt.title('Region with highest spirit_per_capita')
plt.show()
```
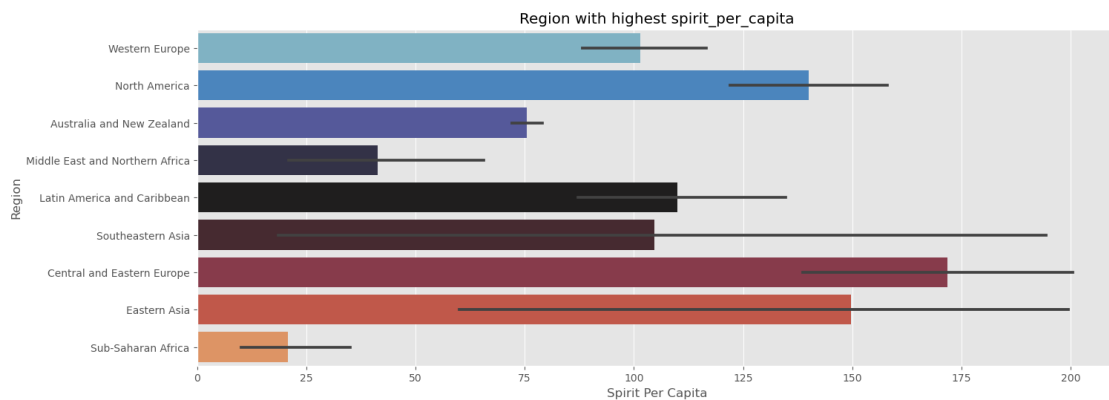

- Central and Eastern Europe sits at the top with Sub-Saharan Africa at the bottom.

15. Countries with highest Spirit per capita

```python
plt.figure(figsize=(16,6), dpi = 100)

sns.barplot(x = 'spirit_percapita',
            y = 'country',
            data = df.nlargest(20, 'spirit_percapita'),
            palette = 'icefire')
plt.xlabel('Spirit Per Capita')
plt.ylabel('Countries')
plt.title('Countries with highest Spirit per capita')
plt.show()
```

Countries with highest Spirit per capita
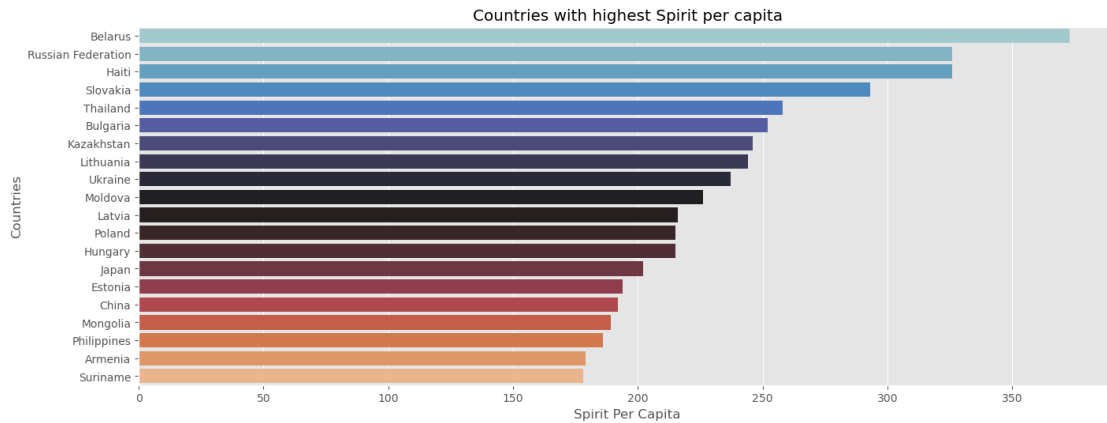
- Belarus is way ahead of every other country by a long shot when it comes to Spirit Per Capita. Russia and Haiti have tied for second place.

16. Countries with lowest Spirit per capita

```python
lowest_spirit = df.sort_values(by = ['spirit_percapita'], ascending = False).tail(10)

# plotting
plt.figure(figsize=(16,4), dpi = 100)

sns.barplot(x = 'country',
            y = 'spirit_percapita',
            data = lowest_spirit,
            palette = 'icefire')
plt.xlabel('Spirit Per Capita')
plt.ylabel('Countries')
plt.title('Countries with lowest spirit_per_capita')
plt.show()
```


Countries with lowest spirit_per_capita
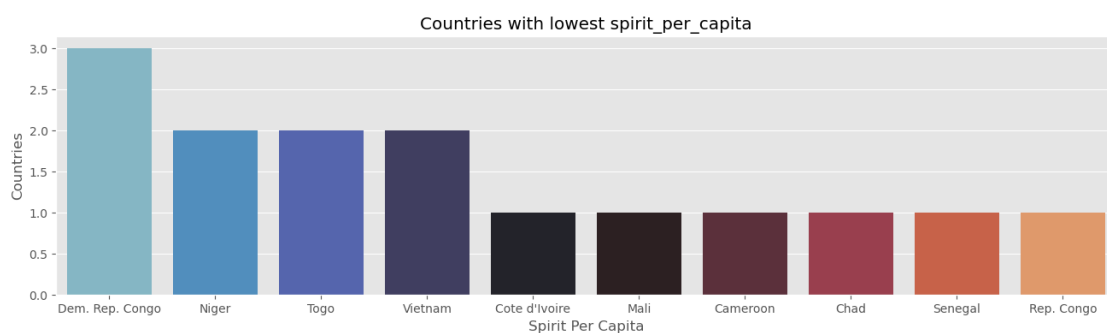
- Republic of Congo sits at the bottom. Looking at the bottom 10 countries it seems as if Spirits are not very popular in Sub-Saharan Africa.

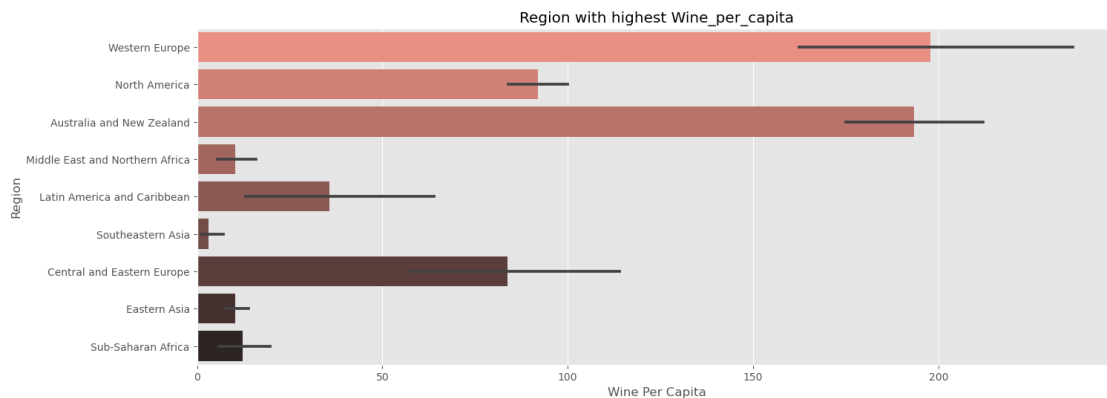## 17. Wine per capita in different Regions

```python
plt.figure(figsize=(16,6), dpi = 100)

sns.barplot(x = 'wine_percapita',
```

```
            y = 'region',
            data = df,
            palette = 'dark:salmon_r')
plt.xlabel('Wine Per Capita')
plt.ylabel('Region')
plt.title('Region with highest Wine_per_capita')
plt.show()
```
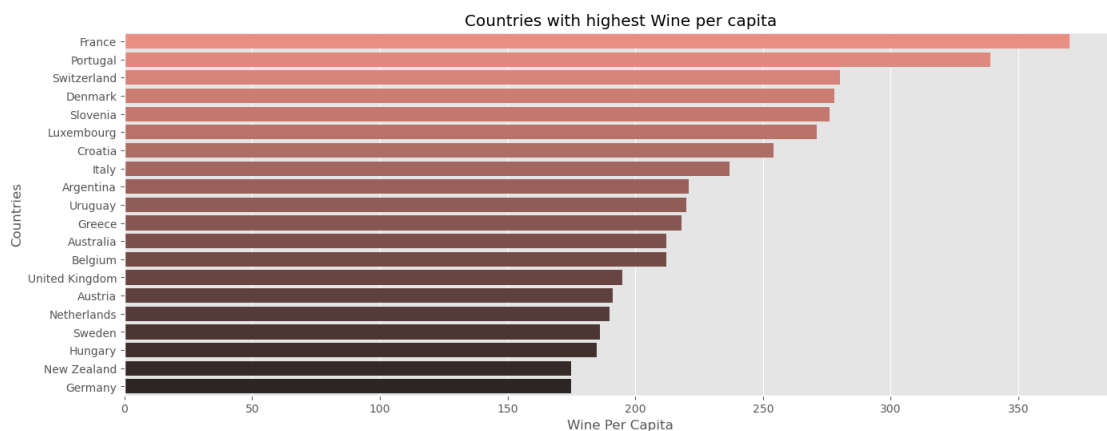

18. Countries with highest Wine per capita

```
plt.figure(figsize=(16,6), dpi = 100)

sns.barplot(x = 'wine_percapita',
            y = 'country',
            data = df.nlargest(20, 'wine_percapita'),
            palette = 'dark:salmon_r')
plt.xlabel('Wine Per Capita')
plt.ylabel('Countries')
plt.title('Countries with highest Wine per capita')
plt.show()
```


- • Wines in Europe represent all that is traditional in winemaking. In no other place a wine reflects the culture of the people who created it and the flavors of the land in which it grew as it does in Europe.

```
lowest_wine = df.sort_values(by = ['wine_percapita'], ascending =
False).tail(25)

# plotting
plt.figure(figsize=(16,4), dpi = 100)

sns.barplot(x = 'wine_percapita',
            y = 'country',
            data = lowest_wine,
            palette = 'dark:salmon_r')
plt.xlabel('Wine Per Capita')
plt.ylabel('Countries')
plt.title('Countries with lowest wine_per_capita')
plt.show()
```
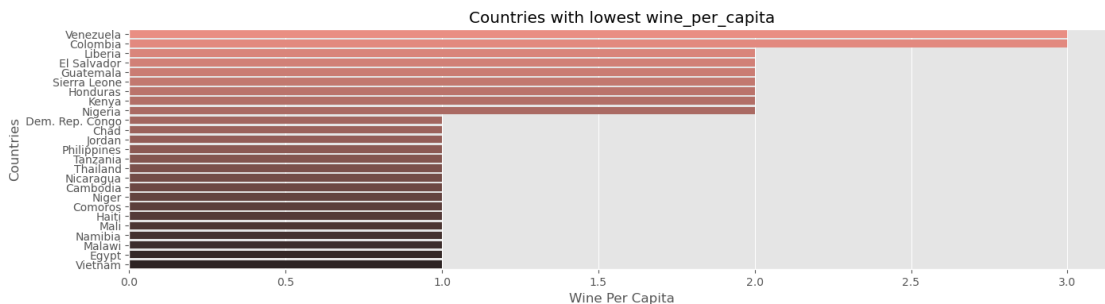

- The bottom 16 countries have the same level of consumption and are a mix of Asian, Latin American and Carribean, African countries, just as we saw in the Regions graph.

20. At last, lets check the Happiness score of different Countries based on their Beer, Spirit and Wine consumption
```
df_beer = df.sort_values(by = 'beer_percapita', ascending =
False).head(20)

# plotting
plt.figure(figsize=(16,4), dpi = 100)
sns.barplot(x = 'country',
            y = 'happinessscore',
            data = df_beer,
            palette = 'YlOrBr')
plt.xlabel('Countries')
plt.xticks(rotation = 90)
plt.ylabel('Happiness Score')
plt.title('Happiness based on Beer per capita')
plt.show()
```
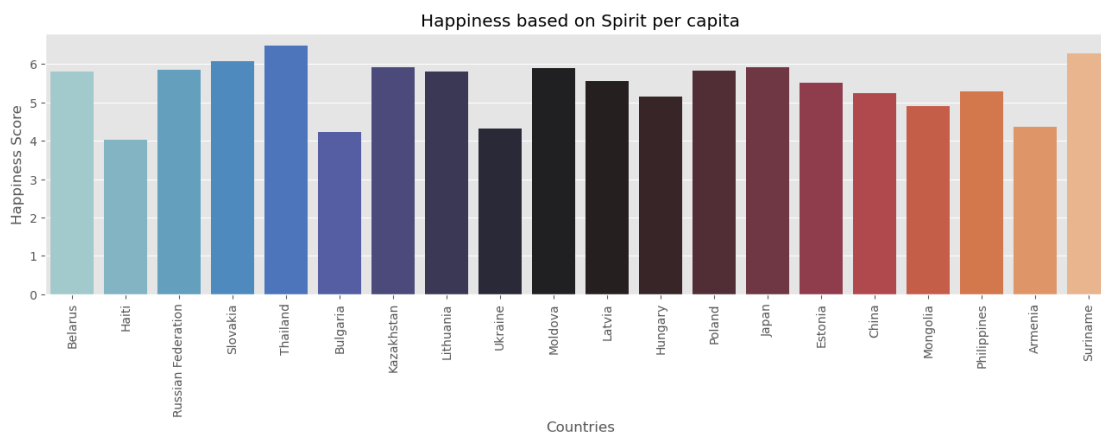
Happiness based on Beer per capita
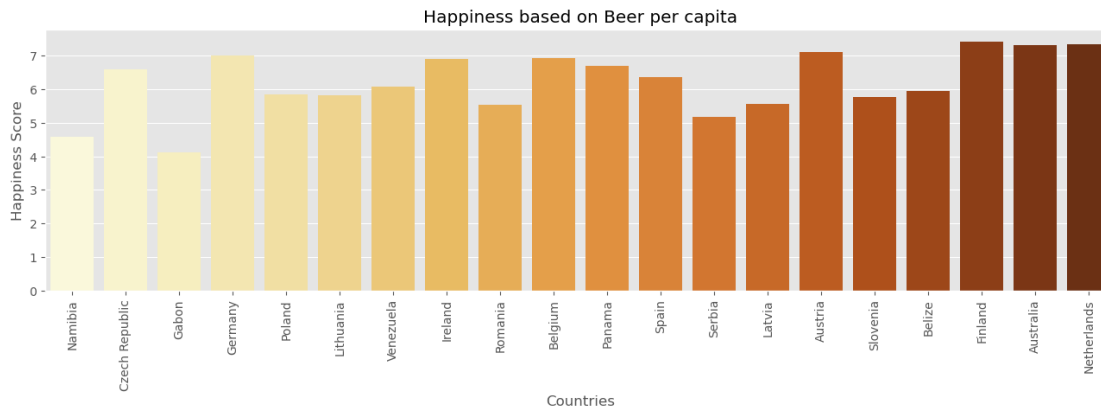

```
df_spirit = df.sort_values(by = 'spirit_percapita', ascending =
False).head(20)

# plotting
plt.figure(figsize=(16,4), dpi = 100)
sns.barplot(x = 'country',
            y = 'happinessscore',
            data = df_spirit,
            palette = 'icefire')
plt.xlabel('Countries')
plt.xticks(rotation = 90)
plt.ylabel('Happiness Score')
plt.title('Happiness based on Spirit per capita')
plt.show()
```

Happiness based on Spirit per capita


```
df_wine = df.sort_values(by = 'wine_percapita', ascending =
False).head(20)

# plotting
plt.figure(figsize=(16,4), dpi = 100)
sns.barplot(x = 'country',dark:salmon_r
            y = 'happinessscore',
            data = df_wine,
            palette = 'dark:salmon_r')
```

```
plt.xlabel('Countries')
plt.xticks(rotation = 90)
plt.ylabel('Happiness Score')
plt.title('Happiness based on Wine per capita')
plt.show()
```


Happiness based on Wine per capita
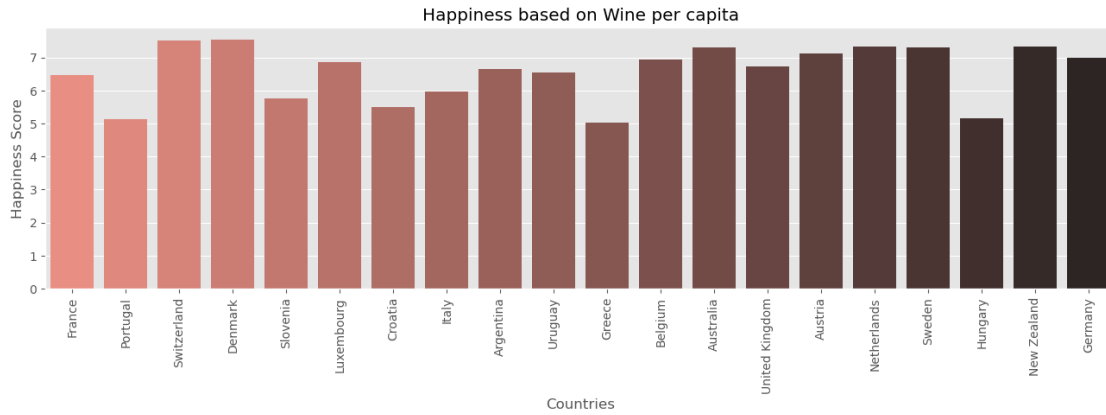
Important :-

- • It can be seen that there is no significant relationship between Happiness and Consumption of Alcohol. Countries with high beer/spirit/wine consumption per capita are not necessarily the happiest countries.

Goto top