re-edited-happiness

November 21, 2023

[5]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

%matplotlib inline
```

[6]:
```
df = pd.read_csv(r'C:\\Users\\HP\\Desktop\\2015.csv')
```

[7]:
```
df.shape
```

[7]:
```
(158, 12)
```

[8]:
```
df.head(4)
```

[8]:
```
       Country          Region  Happiness Rank  Happiness Score  \
0  Switzerland  Western Europe               1            7.587
1      Iceland  Western Europe               2            7.561
2      Denmark  Western Europe               3            7.527
3       Norway  Western Europe               4            7.522

   Standard Error  Economy (GDP per Capita)   Family  \
0         0.03411                   1.39651  1.34951
1         0.04884                   1.30232  1.40223
2         0.03328                   1.32548  1.36058
3         0.03880                   1.45900  1.33095

   Health (Life Expectancy)  Freedom  Trust (Government Corruption)  \
0                   0.94143  0.66557                        0.41978
1                   0.94784  0.62877                        0.14145
2                   0.87464  0.64938                        0.48357
3                   0.88521  0.66973                        0.36503

   Generosity  Dystopia Residual
0     0.29678            2.51738
1     0.43630            2.70201
2     0.34139            2.49204
3     0.34699            2.46531
```

1 DATA CLEANING

[9]:
```
# Printing out our columns
print(df.columns)
```

```
Index(['Country', 'Region', 'Happiness Rank', 'Happiness Score',
       'Standard Error', 'Economy (GDP per Capita)', 'Family',
       'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)',
       'Generosity', 'Dystopia Residual'],
      dtype='object')
```

[10]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 158 entries, 0 to 157
Data columns (total 12 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   Country                        158 non-null    object
 1   Region                         158 non-null    object
 2   Happiness Rank                 158 non-null    int64
 3   Happiness Score                158 non-null    float64
 4   Standard Error                 158 non-null    float64
 5   Economy (GDP per Capita)       158 non-null    float64
 6   Family                         158 non-null    float64
 7   Health (Life Expectancy)       158 non-null    float64
 8   Freedom                        158 non-null    float64
 9   Trust (Government Corruption)  158 non-null    float64
 10  Generosity                     158 non-null    float64
 11  Dystopia Residual              158 non-null    float64
dtypes: float64(9), int64(1), object(2)
memory usage: 14.9+ KB
```

1.0.1 The above information about the dataset shows their is no missing or null values

[11]:
```
# Lets still try and drop missing value with this line of code
df.dropna(axis=0, inplace = True)
```

[12]:
```
df.shape
```

[12]:
```
(158, 12)
```

[13]:
```
df = df.drop_duplicates(keep='first')
```

[14]:
```
df.shape
```

[14]:
```
(158, 12)
```

1.0.2 The above line of code is used to drop duplicate values in our dataset

[15]:
```
df.describe()
```

[15]:
```
       Happiness Rank  Happiness Score  Standard Error  \
count      158.000000       158.000000      158.000000
mean        79.493671         5.375734        0.047885
std         45.754363         1.145010        0.017146
min          1.000000         2.839000        0.018480
25%         40.250000         4.526000        0.037268
50%         79.500000         5.232500        0.043940
75%        118.750000         6.243750        0.052300
max        158.000000         7.587000        0.136930

       Economy (GDP per Capita)      Family  Health (Life Expectancy)  \
count                158.000000  158.000000                158.000000
mean                   0.846137    0.991046                  0.630259
std                    0.403121    0.272369                  0.247078
min                    0.000000    0.000000                  0.000000
25%                    0.545808    0.856823                  0.439185
50%                    0.910245    1.029510                  0.696705
75%                    1.158448    1.214405                  0.811013
max                    1.690420    1.402230                  1.025250

          Freedom  Trust (Government Corruption)  Generosity  \
count  158.000000                     158.000000  158.000000
mean     0.428615                       0.143422    0.237296
std      0.150693                       0.120034    0.126685
min      0.000000                       0.000000    0.000000
25%      0.328330                       0.061675    0.150553
50%      0.435515                       0.107220    0.216130
75%      0.549092                       0.180255    0.309883
max      0.669730                       0.551910    0.795880

       Dystopia Residual
count         158.000000
mean            2.098977
std             0.553550
min             0.328580
25%             1.759410
50%             2.095415
75%             2.462415
max             3.602140
```

1.0.3 This code help us confirm if all our numerical values are in floats or int and does not contain any invalid or string

[ ]:
```
# df.head(4)
```

2 EXPLORATORY DATA ANALYSIS

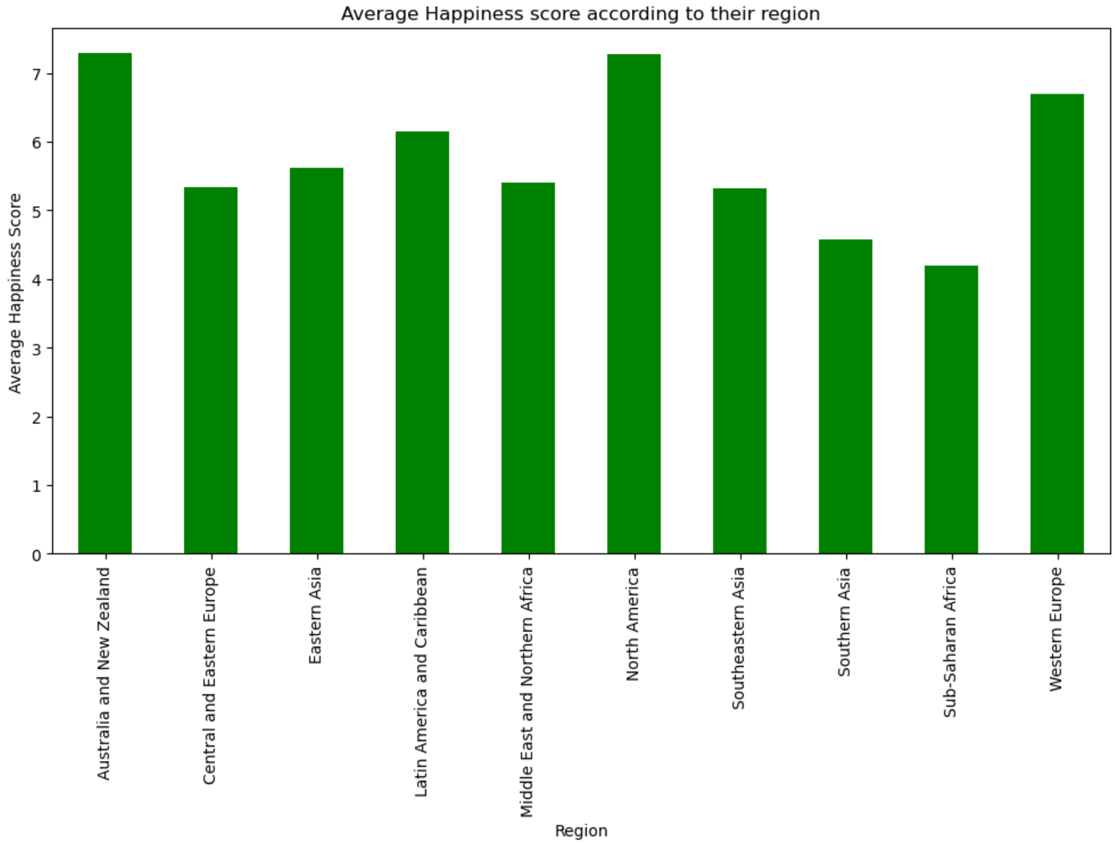
2.0.1 Relationship between Happiness Score and Region

[17]:
```
# Grouping Happiness score by regions
group_data = df.groupby('Region')['Happiness Score'].mean()
# Printing our results
group_data
```

[17]:
```
Region
Australia and New Zealand          7.285000
Central and Eastern Europe         5.332931
Eastern Asia                       5.626167
Latin America and Caribbean        6.144682
Middle East and Northern Africa    5.406900
North America                      7.273000
Southeastern Asia                  5.317444
Southern Asia                      4.580857
Sub-Saharan Africa                 4.202800
Western Europe                     6.689619
Name: Happiness Score, dtype: float64
```

[19]:
```
# Plotting a bar chart for our data
plt.figure(figsize=(12, 6))

group_data.plot(kind='bar', color = 'green')
plt.xlabel('Region')
plt.ylabel('Average Happiness Score')
plt.title('Average Happiness score according to their region')
plt.show()
```



The presented analysis and visualization provide valuable insights into regional happiness rankings. Notably, Australia and New Zealand emerge as the region with the highest average happiness rank, positioning them at the top of the list. In contrast, Sub-Saharan Africa emerges as the region with the least happiness, reflecting the lowest average happiness rank within the dataset for the year 2015. This comparison underscores the diverse spectrum of well-being across regions, with Australia and New Zealand experiencing the highest happiness levels and Sub-Saharan Africa facing greater challenges in achieving happiness

2.0.2 Analysis on how Freedom affect Happiness Score across different Region

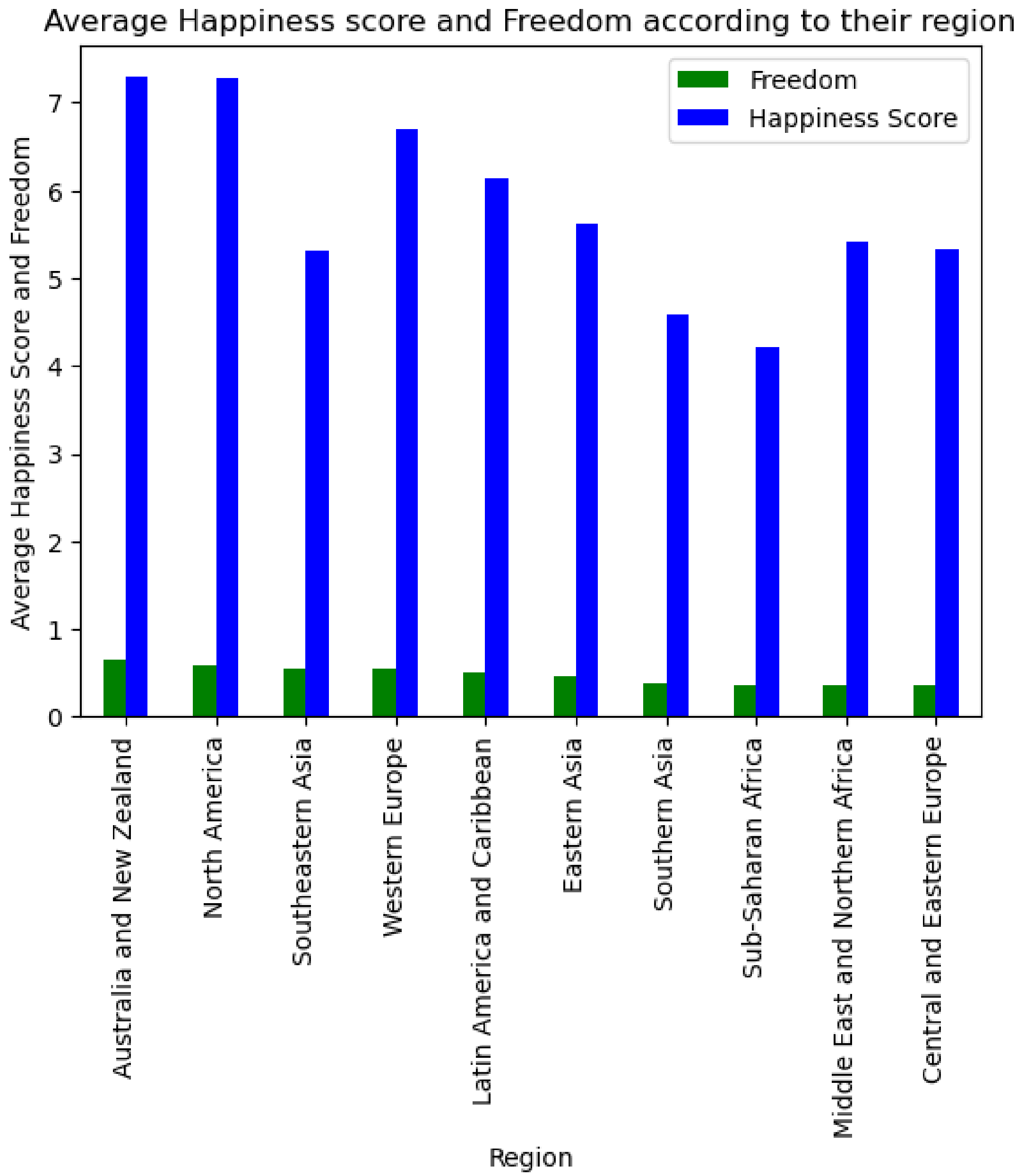
[22]:
```
# Group data by region
grouped = df.groupby('Region')
# Calculate the average freedom score for each region and their happiness score
region_stats = grouped[['Freedom', 'Happiness Score']].mean()
# Sorting the region based on their freedom score in descenging order
sorted_regions = region_stats.sort_values('Freedom', ascending=False)
#Print result
print(sorted_regions)
```

```
                               Freedom  Happiness Score
Region
Australia and New Zealand     0.645310         7.285000
North America                 0.589505         7.273000
Southeastern Asia             0.557104         5.317444
Western Europe                0.549926         6.689619
Latin America and Caribbean   0.501740         6.144682
Eastern Asia                  0.462490         5.626167
Southern Asia                 0.373337         4.580857
Sub-Saharan Africa            0.365944         4.202800
Middle East and Northern Africa  0.361751      5.406900
Central and Eastern Europe    0.358269         5.332931
```

[33]:
```
plt.figure(figsize=(12, 6))

sorted_regions.plot(kind='bar', color = ['green', 'blue'])
plt.xlabel('Region')
plt.ylabel('Average Happiness Score and Freedom')
plt.title('Average Happiness score and Freedom according to their region')
plt.show()
```

```
<Figure size 1200x600 with 0 Axes>
```



The analysis and visualization unveil a compelling pattern, underscoring the significant role of freedom in shaping the happiness scores across various regions. Noteworthy examples include Australia and New Zealand, along with North America, which not only exhibit the highest levels of freedom but also boast the most substantial happiness scores. This correlation implies that a greater degree of freedom positively contributes to higher happiness scores in these regions.Conversely, regions such as Sub-Saharan Africa, Middle East and Northern Africa, and Central and Eastern Europe portray a contrasting picture. These areas experience both limited freedom and some of the lowest happiness scores. This observation suggests that the lack of freedom can potentially lead to lower happiness scores, emphasizing the intricate connection between societal liberties and overall well-being in different regions.

2.0.3 Analysis on how Happiness Score affect Health (Life Expectancy) across different Region

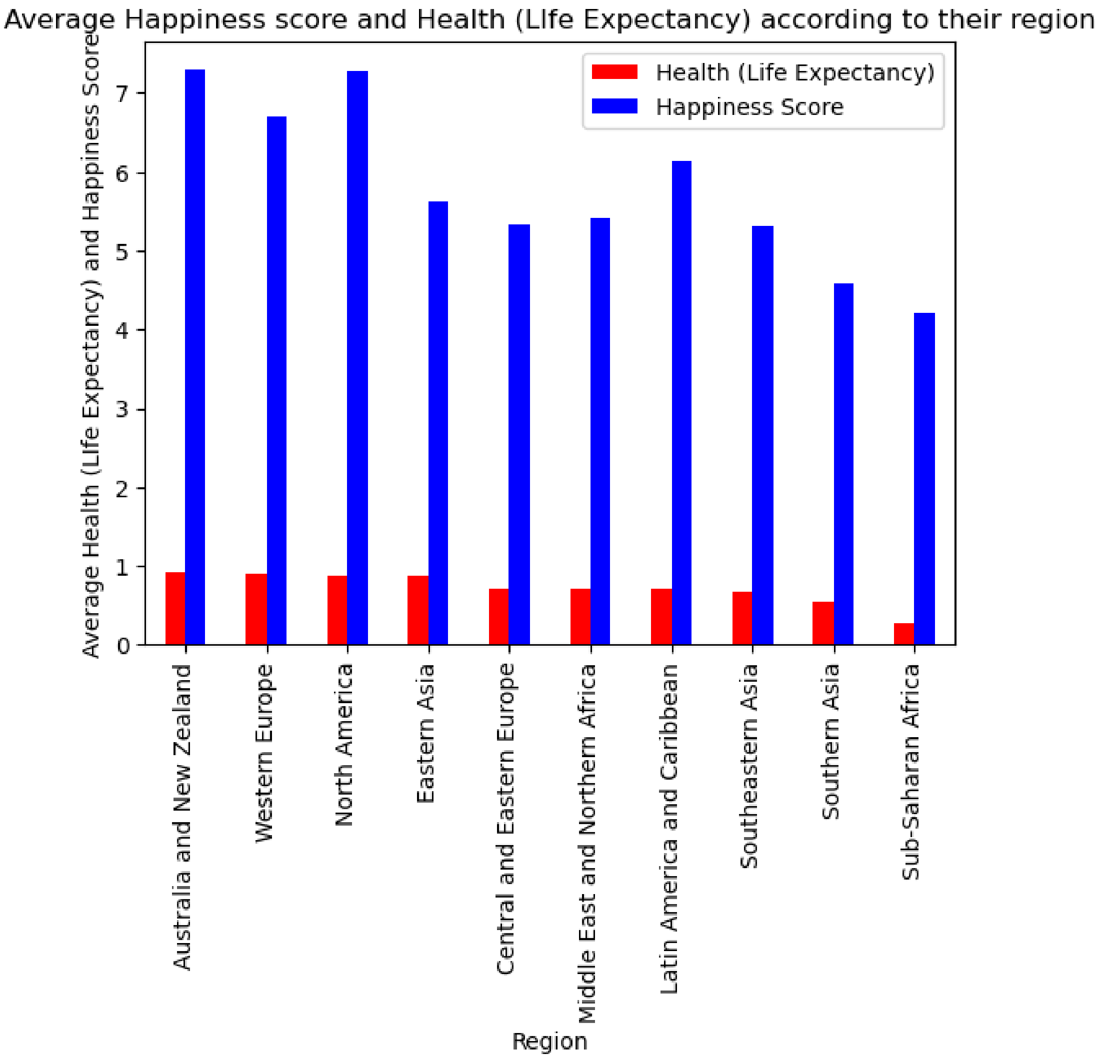
[34]:
```
# Group data by region
grouped = df.groupby('Region')
# Calculate the average Health (Life Expectancy) score for each region and␣
↪their happiness score
region_stats = grouped[['Health (Life Expectancy)', 'Happiness Score']].mean()
# Sorting the region based on their freedom score in descenging order
sorted_regions = region_stats.sort_values('Health (Life Expectancy)',␣
↪ascending=False)
#Print result
print(sorted_regions)
```

```
                                  Health (Life Expectancy)  Happiness Score
Region
Australia and New Zealand                         0.919965         7.285000
Western Europe                                    0.909148         6.689619
North America                                     0.883710         7.273000
Eastern Asia                                      0.877388         5.626167
Central and Eastern Europe                        0.718774         5.332931
Middle East and Northern Africa                   0.705615         5.406900
Latin America and Caribbean                       0.703870         6.144682
Southeastern Asia                                 0.677357         5.317444
Southern Asia                                     0.540830         4.580857
Sub-Saharan Africa                                0.282332         4.202800
```

[35]:
```
plt.figure(figsize=(14, 8))

sorted_regions.plot(kind='bar', color = ['red', 'blue'])
plt.xlabel('Region')
plt.ylabel('Average Health (LIfe Expectancy) and Happiness Score')
plt.title('Average Happiness score and Health (LIfe Expectancy) according to␣
↪their region')
plt.show()
```

```
<Figure size 1400x800 with 0 Axes>
```



Based on the analysis and visualization, a clear pattern emerges in the relationship between happiness scores and health (life expectancy). The results from the visualization and analysis table demonstrate a notable correlation: regions with lower happiness scores tend to exhibit lower life expectancies.For instance, Southern Asia and Sub-Saharan Africa, characterized by the lowest happiness scores, also showcase lower life expectancies. Conversely, regions like Australia and New Zealand, Western Europe, and North America, which boast higher average happiness scores, also demonstrate higher life expectancies.This observation suggests a meaningful connection between happiness and health outcomes, with higher happiness scores aligning with better life expectancies. This correlation underscores the intricate interplay between subjective well-being and objective health metrics, highlighting the potential impact of happiness on overall population health.

2.0.4 Analyzing Trust (Government Corruption) and Happiness Score across Regions

[37]:
```
# Group data by region
grouped = df.groupby('Region')
# Calculate the average trust score for each region and their happiness score
region_stats = grouped[['Trust (Government Corruption)', 'Happiness Score']].
 ↪mean()
# Sorting the region based on their freedom score in descenging order
sorted_regions = region_stats.sort_values('Trust (Government Corruption)',␣
 ↪ascending=False)
#Print result
print(sorted_regions)
```

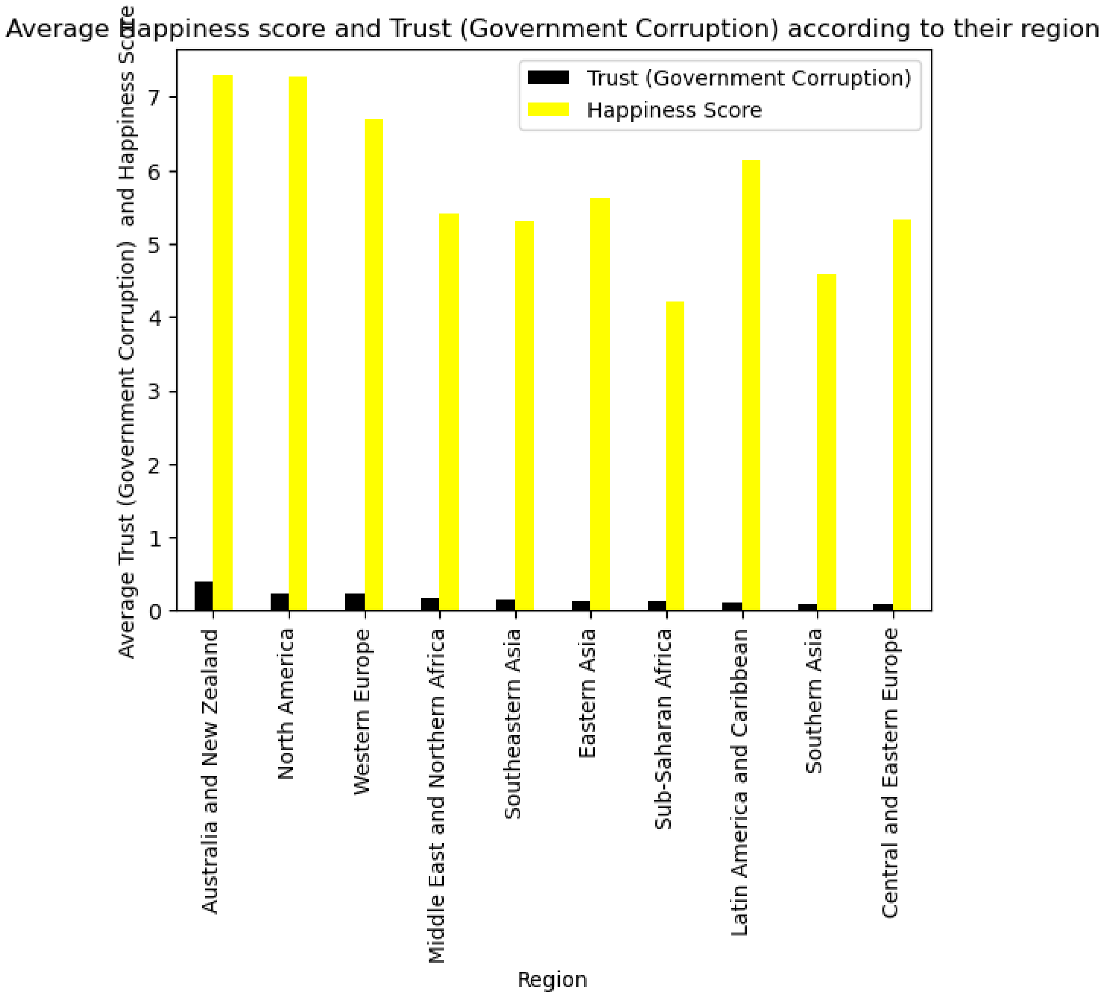
```
                                 Trust (Government Corruption)  \
Region
Australia and New Zealand                             0.392795
North America                                         0.244235
Western Europe                                        0.231463
Middle East and Northern Africa                       0.181702
Southeastern Asia                                     0.151276
Eastern Asia                                          0.127695
Sub-Saharan Africa                                    0.123878
Latin America and Caribbean                           0.117172
Southern Asia                                         0.102536
Central and Eastern Europe                            0.086674

                                 Happiness Score
Region
Australia and New Zealand               7.285000
North America                           7.273000
Western Europe                          6.689619
Middle East and Northern Africa         5.406900
Southeastern Asia                       5.317444
Eastern Asia                            5.626167
Sub-Saharan Africa                      4.202800
Latin America and Caribbean             6.144682
Southern Asia                           4.580857
Central and Eastern Europe              5.332931
```

[38]:
```
plt.figure(figsize=(16, 10))

sorted_regions.plot(kind='bar', color = ['black', 'yellow'])
plt.xlabel('Region')
plt.ylabel('Average Trust (Government Corruption)  and Happiness Score')
plt.title('Average Happiness score and Trust (Government Corruption) according␣
 ↪to their region')
plt.show()
```

<Figure size 1600x1000 with 0 Axes>



This above analysis reveals intriguing patterns in the relationship between trust (government corruption) and happiness scores across different regions. Regions such as Australia and New Zealand, North America, and Western Europe, characterized by higher levels of trust, also exhibit elevated happiness scores, suggesting a positive correlation between trust and well-being.Conversely, regions with lower trust levels, such as Sub-Saharan Africa, tend to have lower happiness scores. This underscores the potential impact of governmental trust on the overall happiness of a region's inhabitants.The overall findings highlight the intricate interplay between trust in government institutions and the subjective well-being of populations across diverse regions.

```
[ ]:
```