Individual Report for COMP7507 Course Project

Group 25
Student Name: LI Dayu
Student UID: 3036411152

1. Overview

Focused problem: Focus on the relationship between social support and happiness score, conduct exploratory visual analysis of the data set, and find out the impact of social support on happiness score and potential data features through visualization and data processing.
Data preparation: Clean the data of recent ten years, remove the empty rows of social support and happiness, unify the column names of the data, facilitate unified processing, and unify the name string inconsistency of the same region in different tables, so as to prevent different regions from being identified in visualization, and then average the data to get a new table. It reflects the average level of social support, happiness score and other attributes in different regions.
Visualization methods: Use tableau for the main visual exploration, aided by python

2. Details of tasks completed

In this part, I take social support as the main entry point and analyze its impact on happiness score combined with other related attributes. And some potential data characteristics of the adopted world happiness report dataset. Visualize each attribute and happiness sore on the world map: The darker the color, the larger the attribute value, and the higher the happiness score, the larger the circle. And we try to find attributes similar to the distribution trend of social support (red in figure 2.1) and combine them to analyze the impact on happiness score. It is found that the distribution trend of the influence of GDP on happiness (green in figure 2.1) is similar to that of social support: the larger the dot, the darker the color.

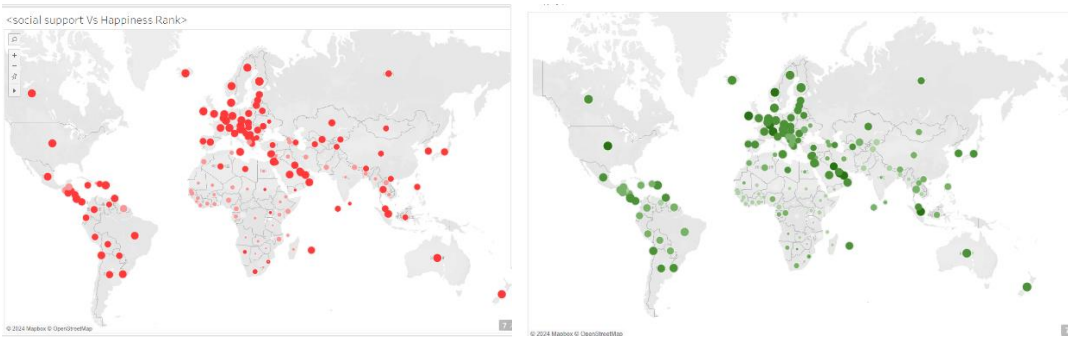

Figure 2.1 Happiness score world map

According to the above visual exploration, I assume that both social support and GDP have a strong positive correlation with happiness score, and I use linear regression to

calculate its correlation coefficient to verify our conjecture:

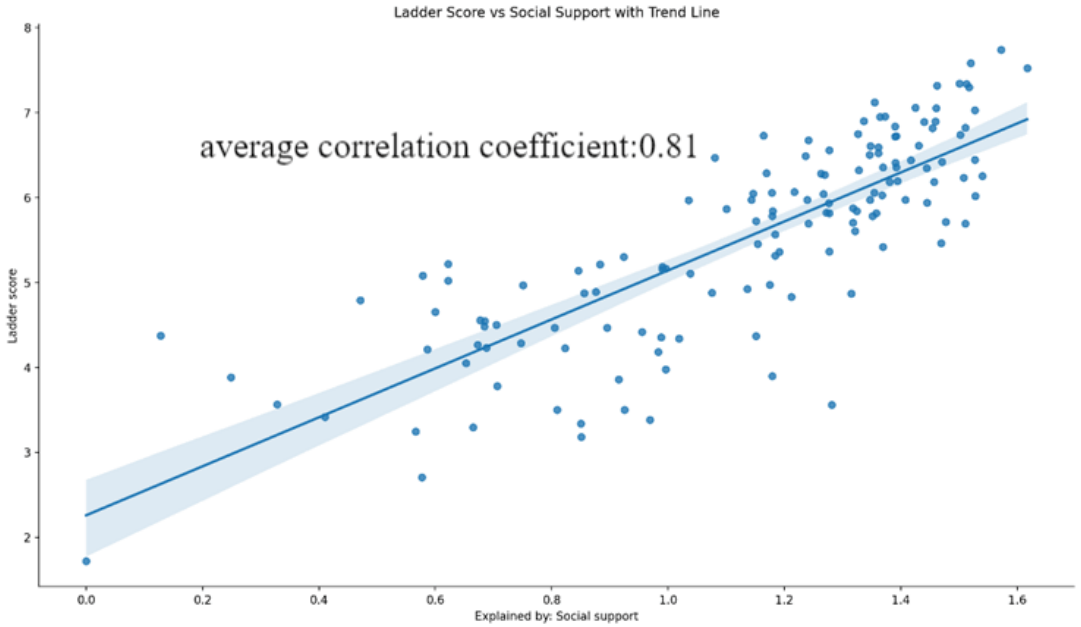

Figure 2.2 Social suppor&happiness score linear regression

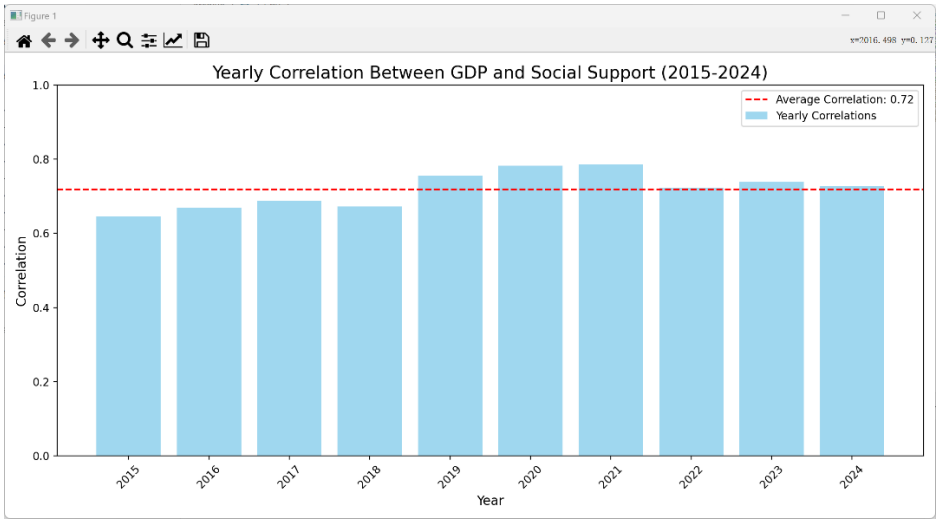

Figure 2.3 Correlation coefficient between social support and GDP

It can be seen that social support has a strong positive correlation with GDP and happiness score. Therefore, it can be concluded that regions with higher GDP generally have higher investment in social support, which will improve the happiness level of the region.

Combined with the visualization of the world map, it is found that the happiness score of several former British colonies is obviously higher than that of former French colonies. Therefore, it is assumed that this pattern is also applicable worldwide. The data was processed to add British or French attributes to some areas (those that were colonized by both countries or had a history that was heavily influenced by both countries).

Here comes a difficulty: Because the higher value of happiness score in the dataset is

about 7 and the lower value is about 5, there is no significant difference between the two numerically, and it is difficult to distinguish them directly for visualization. Even though happiness in the British and French regions has distinguishable distribution features, I tried PCA to extract the features here. However, the effect is not obvious (figure 2.4):

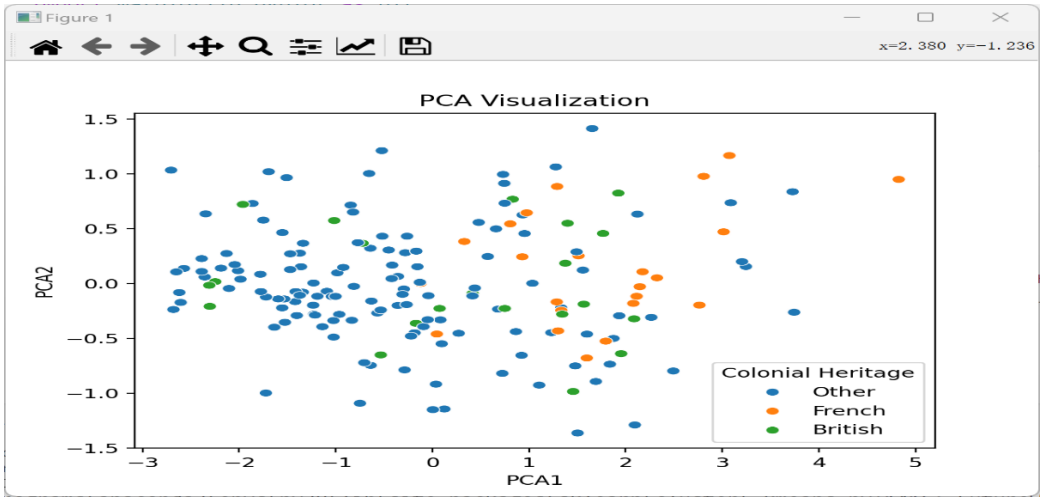

Figure 2.4 PCA

Then I use the density distribution map for observation, and I can can see that the happiness score in the British area has a better distribution trend than that in the French area, and it also has better GDP and social support.

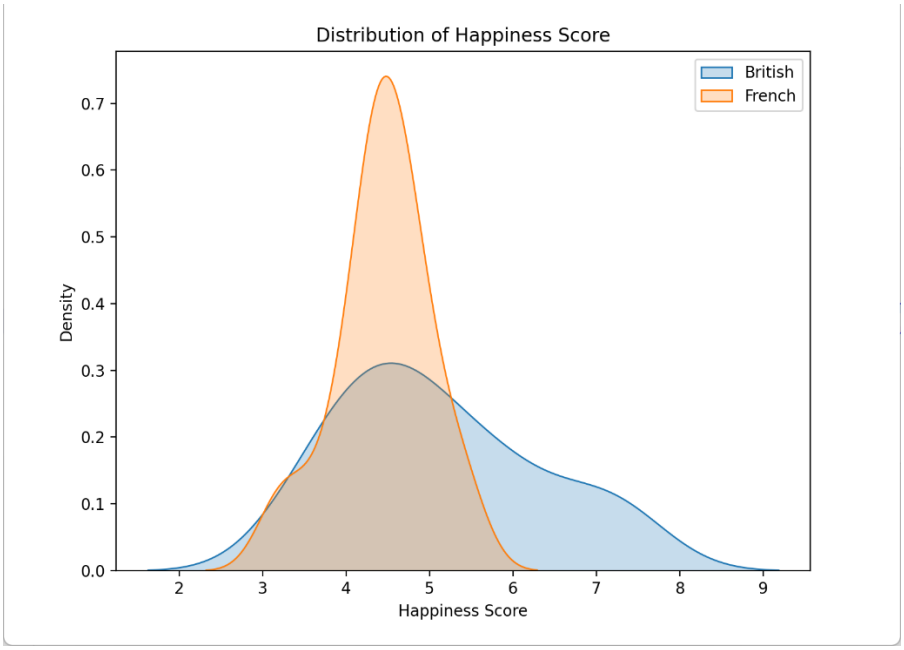

Figure 2.5 Distribution of happiness score

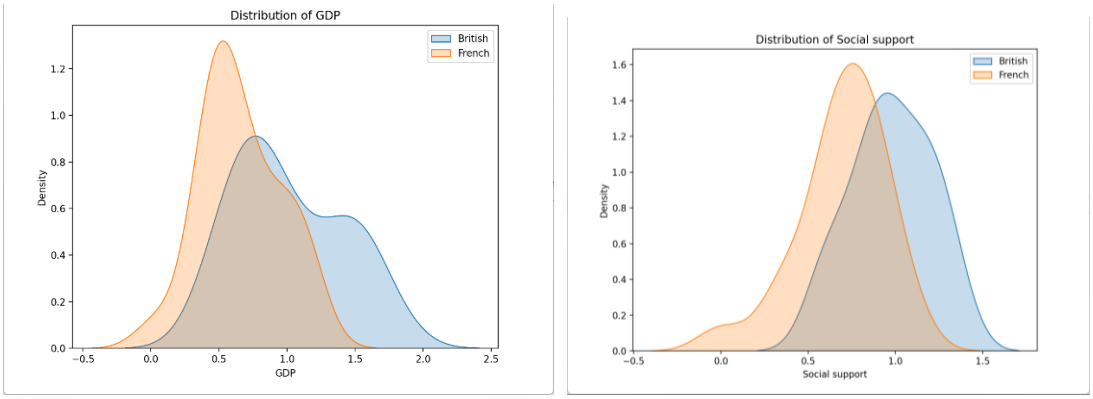

Figure 2.6 Distributions of GDP and social support

However, while this visualization method allows us to observe general trends, it falls short in clearly distinguishing the data characteristics of the two types of regions. Additionally, it is not particularly effective for identifying anomalies (regions that deviate from the expected patterns based on our hypotheses), should they exist.

Therefore, I further processed the happiness scores with the goal of amplifying the differences between them. The method applied was as follows: I calculated the average happiness score across all regions, then subtracted this average from each region's happiness score and applied a logarithmic transformation.
If British and French colonial regions exhibit distinct characteristics in terms of happiness, the visualization of the processed data with enhanced differences should allow for a clear distinction between the two types of regions.

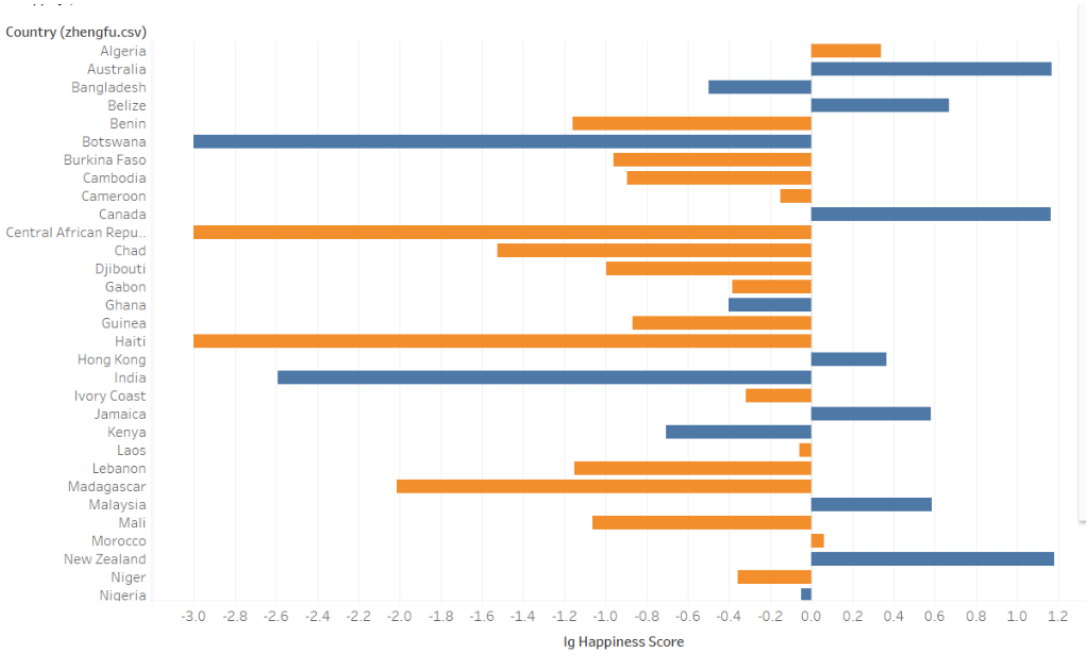

Figure 2.8 Grouped Bar Chart

3.Conclusion

The analysis confirmed that happiness scores in former British colonies tend to be higher than those in former French colonies, with British regions also exhibiting higher levels of GDP and social support. It is speculated that British colonizers may have focused more on economic development, leading to better outcomes in terms of social support and overall happiness in these regions.

However, two notable anomalies were found: India and Botswana, both of which are former British colonies but show relatively low happiness scores despite having decent GDP and social support values.

- India: Despite its relatively high GDP (0.86) and social support (0.58), India's happiness score is lower than expected. Potential factors contributing to this could include:
    - Population size and inequality: India's large population and significant economic disparities may hinder overall happiness.
    - Social issues: Challenges in healthcare, education, and employment opportunities might negatively affect well-being.
    - Cultural differences: The happiness index might not fully capture the social context in India.
- Botswana: Although Botswana has a higher GDP (1.32) and social support (0.915), its happiness score remains low. Possible reasons include:
    - Health and life expectancy: Botswana faces high rates of HIV, which negatively impacts overall well-being.
    - Resource-dependent economy: Despite high GDP, Botswana's economy relies heavily on diamonds, leading to a lack of socio-economic diversity.

These anomalies highlight the complexity of measuring happiness and suggest that further analysis, including more data and deeper exploration of socio-economic factors, is needed to fully understand the drivers of happiness in these regions.