DSC465 – Final Project

World Happiness Report

Decoding Global Happiness: Analyzing Worldwide Happiness Trends and Unveiling the Key Factors Shaping Happiness Score

Group Name: Data Visionaries

Group Members:

- Lakshmi Sowjanya Gangumolu
- Vineeth Appikatla
- Patel Riyaben Hiteshkumar
- Vishw Sanjaykumar Shah
- Patel Charmiben Kamleshkumar

Submission Date: 11-20-2023

1. Introduction

The World Happiness Report offers insights beyond economic measures, revealing the nuances of global well-being. Our project analyzes data from 2015 to 2022, exploring the various factors influencing happiness across nations.

Dataset Name: World Happiness Report from 2015 to 2022

Source: Kaggle (https://www.kaggle.com/datasets/mathurinache/world-happiness-report/data)

Key Features:

- Rank: The ranking of countries based on their happiness scores.
- Country: The name of the country being assessed.
- Happiness Score: The national average of responses to the main life evaluation question.
- Dystopia Residual: A calculated value that contributes to the happiness score.
- GDP: The extent to which GDP per capita explains happiness.
- Social Support: The extent to which social support explains happiness.
- Healthy Life Expectancy: The extent to which healthy life expectancy explains happiness.
- Freedom to Make Life Choices: The extent to which freedom to make life choices explains happiness.
- Generosity: The extent to which generosity explains happiness.
- Perceptions of Corruption: The extent to which perceptions of corruption explain happiness.

Our primary objective is to extract meaningful narratives from this rich dataset, exploring how these factors interrelate to influence societal contentment. Through comprehensive visualizations and analyses, we strive to provide a deeper understanding of what shapes societal contentment, appealing to policymakers, researchers, and those interested in the complexities of global happiness.

2. Exploratory Analysis:

We used histograms and scatterplots to understand the distribution of variables.

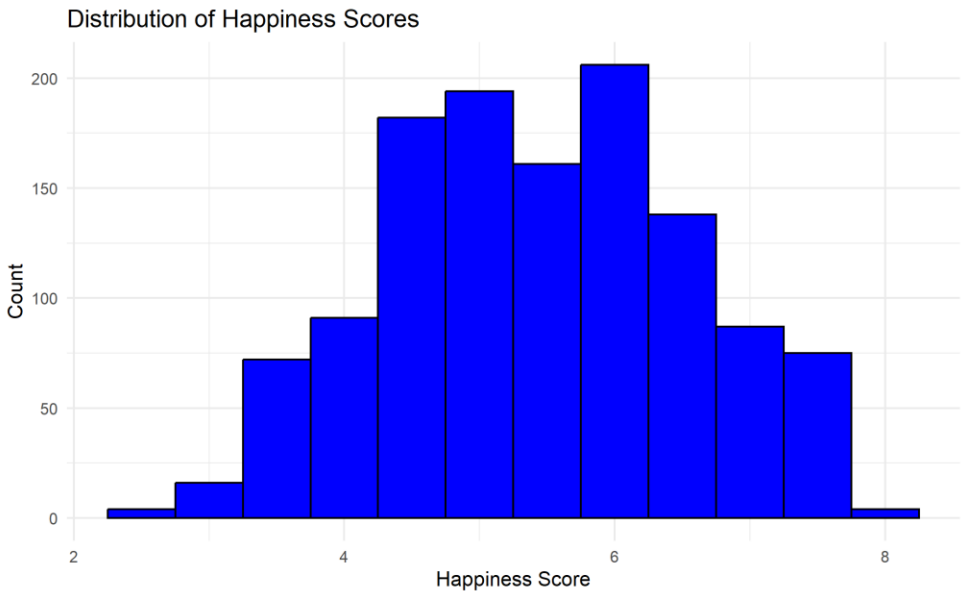

Figure 1 - Histogram of Happiness score

GDP per Capita vs. Happiness Score:

Figure 2 - Avg GDP Vs Avg Happiness score

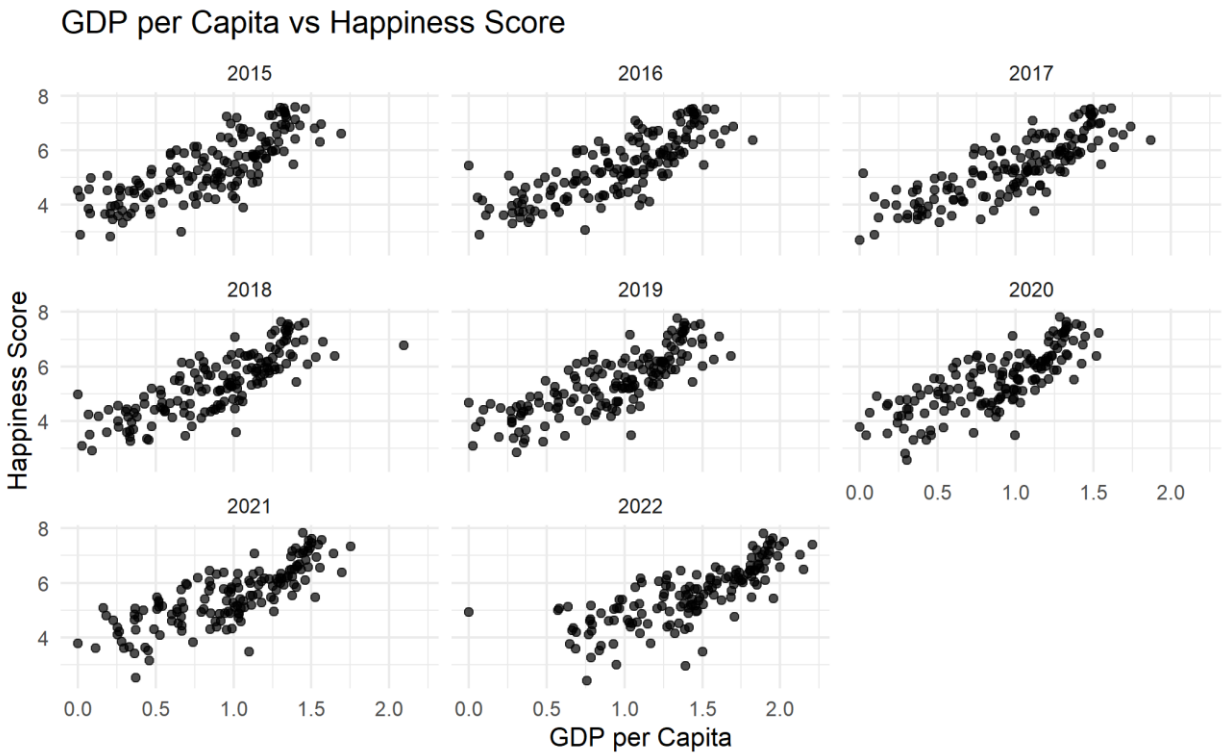

Figure 3 - Scatter plot of GDP VS Happiness score over the years

The above scatterplot shows a clear relationship between a country's GDP per capita and its Happiness Score. As the GDP per capita increases, the Happiness Score tends to rise as well during all the years, forming a diagonal trend that slopes upwards.

This suggests that countries with higher economic output per person generally report higher levels of happiness among their population. While most countries follow this pattern, there are a few outliers—countries that don't fit this trend. These outliers could offer valuable insights into other factors influencing happiness beyond economic prosperity.

Overall, the scatterplot illustrates the strong link between economic well-being, measured by GDP per capita, and a nation's happiness. Exploring outliers further could reveal additional complexities in understanding what contributes to societal contentment beyond financial factors.

Global Happiness Score Distribution by Country:

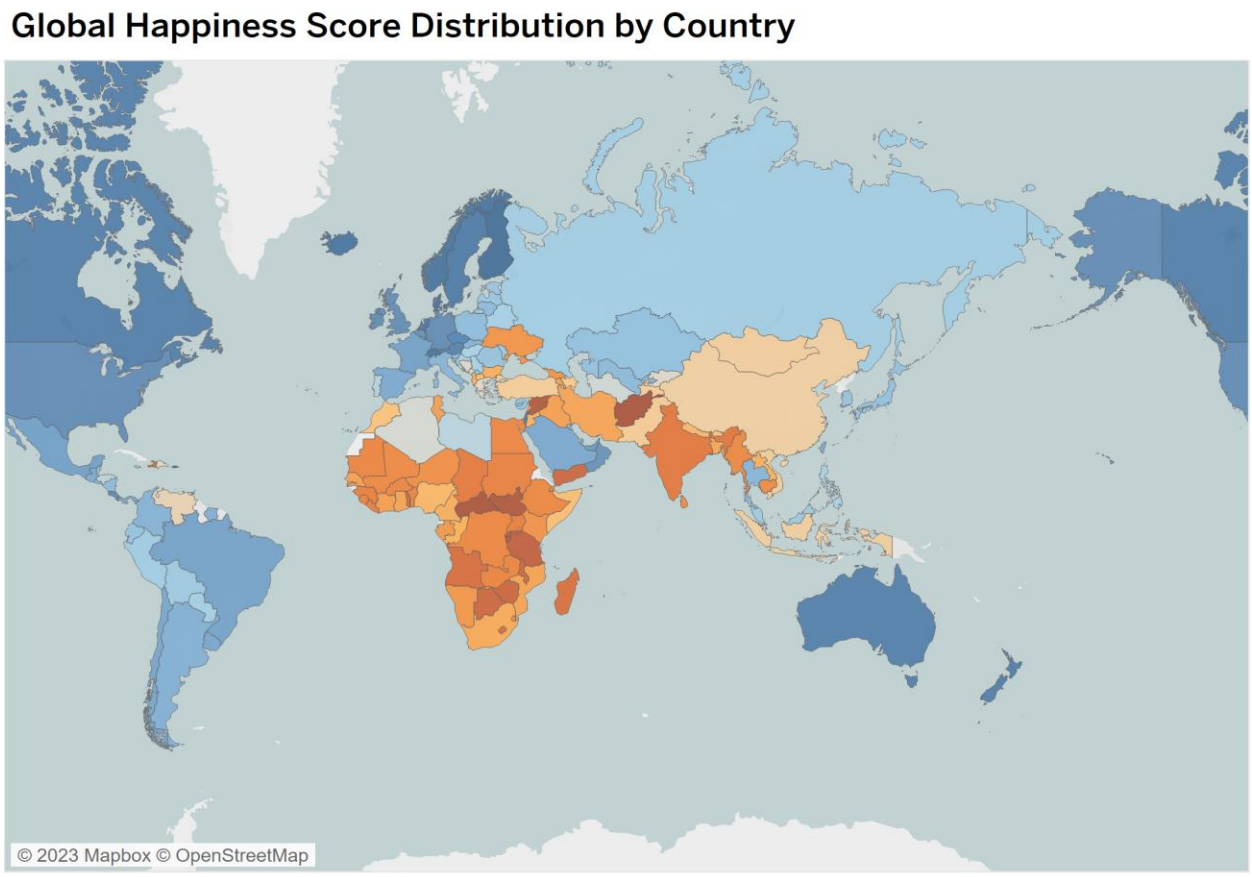

Figure 4 - Map of Happiness score

The color-coded map, representing average happiness scores across countries, provides a powerful visual representation of global well-being disparities. When countries are shaded according to their happiness scores, it becomes evident that some regions, particularly in Africa, the Middle East, and parts of South Asia, exhibit lower happiness scores in comparison to others. These lower scores can be attributed to a complex interplay of socioeconomic, political, and cultural factors. Economic challenges, income inequality, limited access to healthcare and education, and the presence of social unrest and political instability contribute to the lower happiness scores in these areas. The map serves as a reminder of the importance

of addressing these disparities through targeted policies and interventions to uplift well-being in regions where happiness scores are lower. Nonetheless, it's crucial to recognize the resilience and community support that exist in these regions, which continue to thrive despite facing adversity.

3. Visualizations:

Key Trends in Happiness Score Rankings (2015-2022) among Selected Countries:

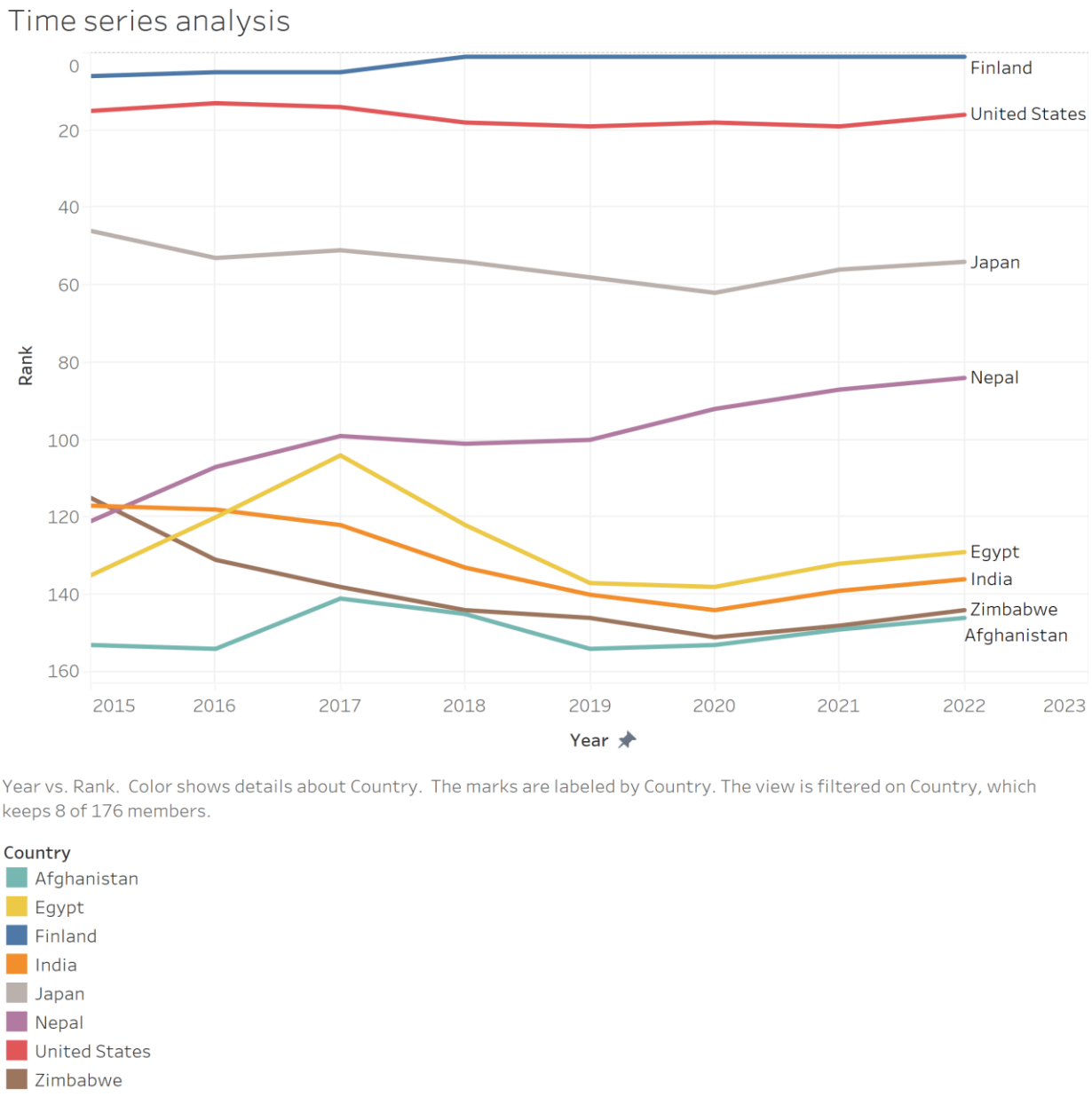

Figure 5 - Line chart of change in Rank over the years

Our selection of countries, chosen deliberately from diverse regions with varying ranks in global Happiness Scores, reveals intriguing insights into well-being trends. Finland emerges as a consistent top performer, securing a place within the top 10 rankings throughout the examined period and claiming the top spot persistently from 2018 to 2022, indicative of robust societal well-being. Conversely, nations such as India, Japan, and Zimbabwe witnessed declines in their Happiness Scores, resulting in a decrement in their global rankings, signifying potential challenges impacting overall well-being. Nepal showcased remarkable progress, ascending significantly from around 120 to 84 in its global ranking, reflective of noteworthy enhancements in multiple happiness-contributing factors. Amidst these fluctuations, the United States, deliberately included for its relative stability, exhibited consistent Happiness Score rankings with minimal fluctuations across the analyzed period. This diversified selection underscores the disparate trajectories in happiness levels among countries from varied regions, providing a nuanced understanding of the multifaceted nature of societal well-being.

Bar chart of factors contributing to Happiness score:

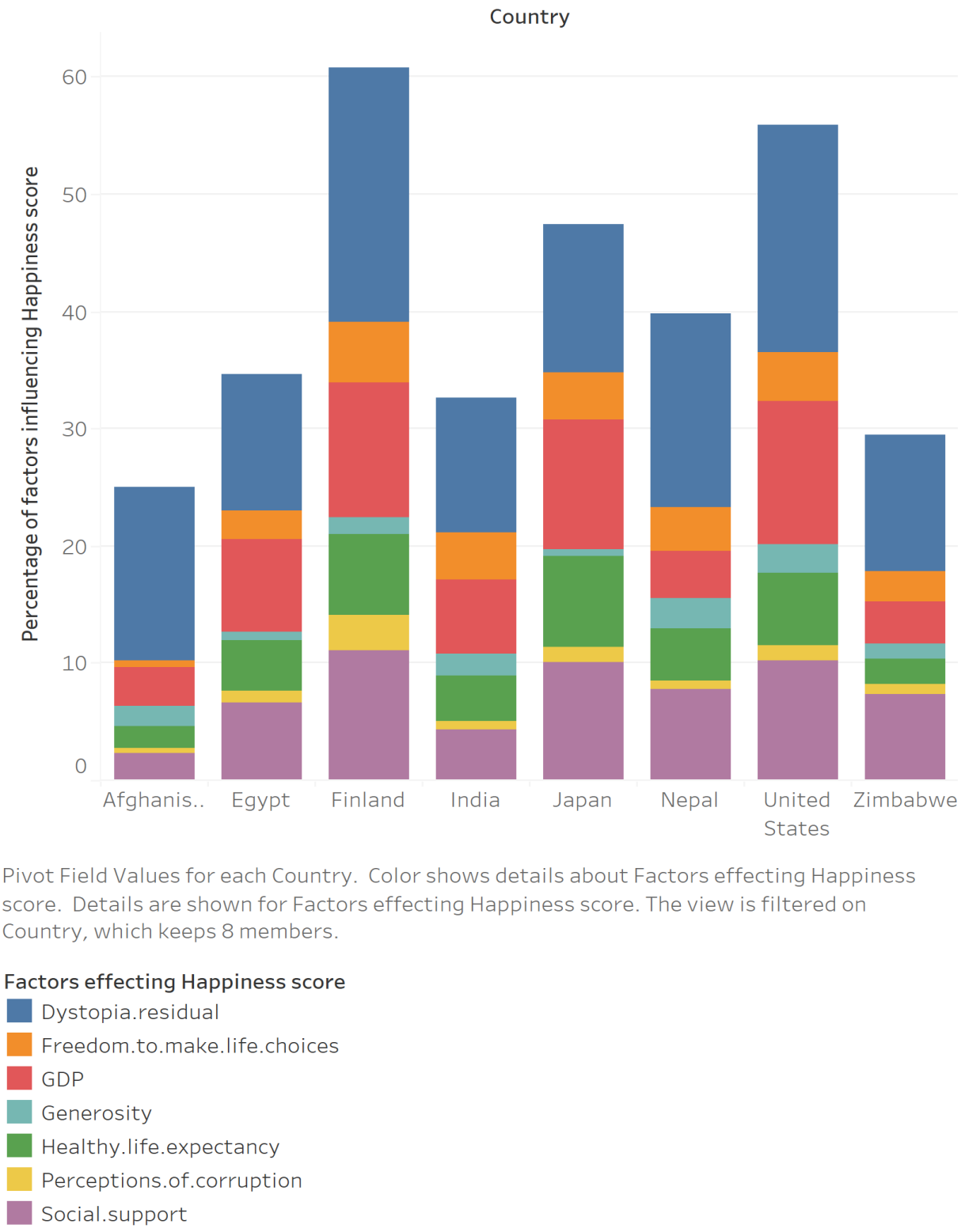

Figure 6 - Percentage of factors contributing to Happiness score

From the above chart, which shows the factors contributing to happiness score on average from 2015 to 2022 for selected countries, we can observe significant shifts in happiness scores and their influencing factors. The height of the bars indicates the happiness score. Different colors are used to distinguish factors such as GDP, life expectancy, Dystopia residual, freedom to make life choices, Generosity, social support, and corruption. Notably, Dystopia residual and GDP had a substantial impact on happiness scores over these years. We can see that some countries experienced significant changes in their happiness scores due to these factors. The chart's clear visual representation provides insights into how specific factors have influenced the well-being of these nations over time. This analysis can be valuable for policymakers and researchers seeking to understand the evolving dynamics of global happiness.

4. Analysis and Conclusion:

Figure 2 & 3, scatter plot of GDP Vs Happiness score shows a clear trend: countries with higher economic wealth, measured by GDP per person, tend to have happier populations. However, there are exceptions to this rule. Very few countries deviate from the expected pattern, indicating that happiness isn't solely determined by financial prosperity. These outliers might hold secrets about other factors that contribute to well-being, beyond just economic measures.

Figure 4, a color-coded map, paints a picture of global happiness disparities. It's evident that certain regions, especially in Africa and the Middle East, consistently have lower happiness scores. These areas face complex challenges like poverty, inequality, and social and political unrest. This map serves as a stark reminder of the urgent need for targeted interventions to uplift well-being in these regions, while also highlighting the resilience present despite these adversities.

Additionally, we can see from figure 5 & 6, how various factors influence happiness scores over time for selected countries. It's clear that factors like a country's general conditions and economic prosperity play significant roles in shaping happiness levels. The shifts observed across these factors have had notable impacts on the well-being of different nations. This visualization is invaluable for policymakers and researchers seeking to understand the ever-evolving dynamics of global happiness and the factors driving these changes.

Together, these visualizations reveal the complexity of happiness, showing that while economic factors matter, they aren't the sole determinants of a nation's well-being. They emphasize the need for comprehensive strategies that address diverse social, economic, and political challenges to foster happiness across the globe.

5. Appendices:

R Code:

```
library(dplyr)
library(readr)  # for reading CSV files
setwd("C:\\Users\\sowja\\OneDrive\\Desktop\\DSC465\\Project new")

# Read CSV files for all the years
data_2015 <- read.csv("data_2015.csv")
data_2016 <- read.csv("data_2016.csv")
data_2017 <- read.csv("data_2017.csv")
data_2018 <- read.csv("data_2018.csv")
data_2019 <- read.csv("data_2019.csv")
data_2020 <- read.csv("data_2020.csv")
data_2021 <- read.csv("data_2021.csv")
data_2022 <- read.csv("data_2022.csv")

data_2022$Happiness.score <- as.numeric(gsub(",", ".", data_2022$Happiness.score))
data_2022$Dystopia.residual <- as.numeric(gsub(",", ".", data_2022$Dystopia.residual))
data_2022$GDP <- as.numeric(gsub(",", ".", data_2022$GDP))
data_2022$Social.support <- as.numeric(gsub(",", ".", data_2022$Social.support))
data_2022$Healthy.life.expectancy <- as.numeric(gsub(",", ".",
data_2022$Healthy.life.expectancy))
data_2022$Freedom.to.make.life.choices <- as.numeric(gsub(",", ".",
data_2022$Freedom.to.make.life.choices))
data_2022$Generosity <- as.numeric(gsub(",", ".", data_2022$Generosity))
data_2022$Perceptions.of.corruption <- as.numeric(gsub(",", ".",
data_2022$Perceptions.of.corruption))
#Adding Year column
data_2015$Year <- 2015
data_2016$Year <- 2016
data_2017$Year <- 2017
data_2018$Year <- 2018
data_2019$Year <- 2019
data_2020$Year <- 2020
data_2021$Year <- 2021
data_2022$Year <- 2022
# Combine the datasets
combined_data <- bind_rows(data_2015,data_2016,data_2017,data_2018,data_2019,data_2020,
data_2021, data_2022)
# Using gsub to remove asterisks from country names
combined_data$Country <- gsub("\\*", "", combined_data$Country)
# Write the combined data to a new CSV file
write.csv(combined_data, "combined_happiness_report.csv", row.names = FALSE)
head(combined_data)
library(ggplot2)
# Histogram of Happiness Score
ggplot(combined_data, aes(x = Happiness.score)) +
  geom_histogram(binwidth = 0.5, fill = "blue", color = "black") +
```

```
43.   theme_minimal() +
44.   labs(title = "Distribution of Happiness Scores", x = "Happiness Score", y = "Count")
45. # Scatter plot of GDP per Capita vs Happiness Score with separate panels for each Year and
colored by Health Life Expectancy
46.       ggplot(combined_data, aes(x = GDP, y = Happiness.score)) +
47.         geom_point(alpha = 0.7) +
48.         facet_wrap(~ Year) +  # Creates separate panels for each year
49.         scale_color_gradient(low = "blue", high = "red") +  # Use a gradient color scale for
Health Life Expectancy
50.         theme_minimal() +
51.         labs(title = "GDP per Capita vs Happiness Score ",
52.              x = "GDP per Capita",
53.              y = "Happiness Score")
54.
```

Individual Report:

Name: Lakshmi Sowjanya Gangumolu

Student ID: 2132626

As the group liaison, my primary responsibilities encompassed team coordination, facilitating communication with our professor, and managing our submissions throughout the project duration. Taking the initiative, I led the preparation of documentation, ensuring clarity and coherence in our project deliverables.

A significant portion of my contributions revolved around data preprocessing, involving the consolidation of multi-year datasets and meticulous data cleaning procedures. This groundwork was pivotal in establishing a robust foundation for our subsequent analyses.

Moreover, my focus on developing explanatory visualizations aimed to enhance our insights into the dataset. Specifically, I crafted a detailed line chart illustrating the fluctuation in rankings for different countries across years and identified how the trend has changed over time for specific countries. Additionally, I created a comprehensive bar chart showcasing the percentage distribution of factors such as GDP, life expectancy, Dystopia residual, freedom to make life choices, Generosity, social support, and corruption influencing happiness scores across various countries.

Engaging in this project provided me with a profound insight into the fundamental significance of meticulous data preprocessing. Understanding the necessity of refining and consolidating

disparate datasets from different years was pivotal. The process of cleaning and standardizing data not only ensured consistency but also laid the groundwork for our subsequent analyses. Recognizing the critical role of this preparatory phase in the accuracy and reliability of our findings was an eye-opener. It underscored the indispensable nature of data preprocessing as the cornerstone of any meaningful data-driven exploration.

Furthermore, immersing myself in the realm of explanatory visualizations was a captivating journey. Crafting visual representations that elucidate complex patterns and trends within the dataset was an enlightening experience. I discovered that beyond being visually appealing, these graphics serve as powerful tools for storytelling. Through charts, graphs, and diagrams, we were able to communicate intricate data-driven narratives effectively. This process highlighted the importance of not just analyzing data but also translating it into compelling visual stories that resonate with audiences. The ability to convey nuanced insights through visuals not only enhances understanding but also fosters engagement and facilitates decision-making.

Overall, this project and course significantly deepened my appreciation for the intricacies of data visualization. It reinforced the understanding that data visualization is not merely about creating aesthetically pleasing graphics but about distilling complex information into accessible and meaningful narratives. The fusion of meticulous data preparation with compelling visual storytelling emerged as a potent means to extract valuable insights from datasets, emphasizing the pivotal role of data visualization in driving informed decision-making and understanding complex phenomena.

Name: Patel Charmiben kamleshkumar

Student ID: 2121191

Data visualization involves the creation of visual representations such as charts, graphs, maps, and other graphical elements to effectively convey information from data. It serves as a means to communicate insights, narrate stories, and influence audiences.

Our goal is The group collaboration focused on scrutinizing the World Happiness Report dataset to understand the various factors influencing global happiness. The project had a clear objective of outlining research and activities to be conducted. My role in the project encompassed diverse

tasks, including exploring the dataset and creating visual representations. The report provides a detailed summary of my contributions, offering a thorough analysis of each aspect.

The primary goal of our collaborative initiative was to analyze the World Happiness Report dataset. Our aim was to uncover trends, patterns, and correlations that impact the overall happiness scores of different countries. This involved a comprehensive examination of the data, followed by analysis, interpretation, and presentation of the findings. The report serves as a comprehensive documentation of the research process and outcomes, shedding light on the intricate dynamics of global happiness as revealed by the dataset.

As we Present in the Project,

Obtaining a dataset was among my initial contributions, where I actively participated in the Procurement process. This included researching potential sources, evaluating the relevance of the data to our project, and ensuring its appropriateness for thorough analysis. And we guys are coordinating , group meetings and as follow I consistently maintained active participation in all team meetings, recognizing the importance of effective communication and collaboration. Regular involvement allowed me to stay updated on project progress, contribute valuable insights, and work together with the team to address any challenges.

Taking on the task of obtaining presentation materials, I made sure that our project presentation not only looked visually appealing but also aligned seamlessly with the key messages we aimed to convey. This involved thorough research, meticulous selection, and methodical arrangement of visual aids, all geared towards enhancing the overall impact of our findings then A crucial aspect of our examination involved creating a histogram to illustrate the happiness scores. Leveraging my proficiency in data visualization, I adeptly produced a histogram that precisely conveyed the distribution of happiness scores in the dataset. The incorporation of this visual aid elevated the thoroughness of our presentation and facilitated a clearer understanding of the data.

The project faced various obstacles, including the necessity to efficiently synchronize group meeting schedules, manage diverse datasets, and guarantee the accuracy of visual representations. However, through transparent communication and collaborative problem-solving, the team successfully navigated and resolved these challenges.

Key Takeaways:

Engaging in the collaborative initiative has provided valuable insights into the effectiveness of teamwork, the importance of thorough examination of datasets, and the power of visual aids in conveying complex information. Additionally, this experience has bolstered my ability to make meaningful contributions to collaborative analytical projects

In summary, my involvement played a pivotal role in the group project by significantly contributing to the analysis and presentation of the World Happiness Report dataset. The successful collaboration with team members facilitated the resolution of challenges and resulted in the creation of a comprehensive report delving deeply into the complexities of global

happiness. This encounter has cultivated a heightened awareness of the intricacies and nuances inherent in the analysis of extensive datasets.

I gained valuable insights from this project, particularly in the creation of visual representations that unveil intricate patterns in the dataset. The process of crafting visually appealing graphics was not only enlightening but also emphasized their role as potent storytelling tools. Utilizing charts, graphs, and diagrams allowed us to effectively communicate complex data-driven narratives. This experience underscored the significance of not just analyzing data but also translating it into compelling visual stories that resonate with audiences. The capacity to convey nuanced insights through visuals not only enhances understanding but also encourages engagement and facilitates decision-making.

In essence, this project and course deepened my appreciation for the nuances of data visualization. It reinforced the idea that it goes beyond aesthetics, emphasizing the importance of distilling complex information into accessible and meaningful narratives. The synergy of meticulous data preparation with compelling visual storytelling emerged as a powerful means to extract valuable insights from datasets, highlighting the pivotal role of data visualization in driving informed decision-making and understanding complex phenomena.

Name: Patel Riyaben Hiteshkumar

Student ID: 2124701

1. **Introduction:**

   As a group member on this project, I am responsible for conducting data analysis and contributing to the overall interpretation of the World Happiness Report data from 2015 to 2022. My expertise in data analysis and statistical techniques will be instrumental in identifying trends, patterns, and correlations that shape the global happiness landscape. I am particularly interested in exploring the complex interplay of GDP, social, and Life expectancy factors that influence human well-being. Through a comprehensive analysis of the data, I aim to contribute to a deeper understanding of the factors that promote happiness across nations and contribute to the development of effective policies and programs that foster well-being around the world.

2. **Role and Responsibilities:**
   Data Collection and Preparation:
   Gathering and organizing data from various sources, including the World Happiness Report website and other relevant databases. Cleaning and preprocessing the data to ensure accuracy and consistency. Identifying and handling missing values or outliers in the dataset.
   Leadership and Coordination:
   Actively participating in team discussions and brainstorming sessions to contribute ideas and perspectives. Providing technical guidance and support to team members with less

data analysis experience. Coordinating data sharing and collaboration among team members to ensure seamless workflow and efficient progress.

3. **Contributions to Visualizations:**

Visualization 1: GDP vs Happiness Score

Purpose: This visualization aims to explore the relationship between GDP, happiness score, and life expectancy. By coloring each data point based on life expectancy, we can visualize how happiness and GDP vary across countries with different life expectancies.

Contribution: This visualization provides a clear and intuitive way to understand the complex interplay between economic prosperity, life expectancy, and subjective well-being. The color gradient allows us to identify countries with high happiness scores and long life expectancies, as well as those with lower happiness scores and shorter life expectancies.

Visualization 2: Hexbin Plot of GDP vs Generosity

Purpose: This visualization aims to examine the relationship between GDP and generosity. Hexbin plots are a type of scatter plot that use hexagonal bins to represent the density of data points. This visualization allows us to identify trends and patterns in the data that might not be apparent from a traditional scatter plot.

Contribution: This visualization provides a more nuanced understanding of the relationship between GDP and generosity. While a traditional scatter plot might suggest a positive correlation, the hexbin plot reveals a more complex pattern, with a concentration of data points in the middle range of both GDP and generosity. This suggests that there may be a threshold effect, where generosity increases with GDP up to a certain point, but then plateaus or even declines.

Reason for Hexbin Plot Failure:

While I was able to generate a hexbin plot of GDP vs generosity, it did not effectively convey the relationship between the two variables. The density of data points was too low, and the color gradient was not well-chosen. As a result, the visualization was not informative and did not contribute to our understanding of the data.

4. **Learning Experience - Data Visualization:**

Choosing the right chart type: I gained a deeper understanding of different chart types and when to use each one. I learned that the choice of chart type depends on the type of data being visualized, the purpose of the visualization, and the target audience.

Creating interactive visualizations: I explored the use of interactive visualizations to make data more engaging and accessible.

5. **Conclusion:**
   Key takeaways and lessons learned from my involvement in this project include:

   - The importance of data visualization in communicating complex information in a clear and concise way.
   - The iterative nature of data visualization and the importance of experimentation to find the most effective way to present data.
   - The need to consider the target audience when selecting the right chart type, color scheme, and layout for a visualization.

Name: Vishw Shah

student ID 2078150

Abstract

- This collaborative group project aimed to analyze the World Happiness Report dataset in order to investigate the various factors that contribute to global happiness. My contributions to the project encompassed a wide range of tasks, including dataset exploration and the development of visual representations. This report provides a comprehensive overview of the aforementioned contributions, offering a detailed analysis of each.
- The purpose of this project is to provide an overview of the research and activities that will be undertaken.
- The main aim of our collaborative endeavor was to conduct an analysis of the dataset from the World Happiness Report, with the purpose of identifying and examining trends, patterns, and correlations that are influential in determining the aggregate happiness scores of different countries. The project encompassed a comprehensive examination of data, its analysis, interpretation, and subsequent presentation of the findings.

The contributions that I have made are as follows:

Dataset Acquisition

One of my initial contributions involved actively engaging in the process of acquiring a suitable dataset. The process encompassed conducting research on potential sources, assessing the pertinence of the data to our project, and verifying its suitability for comprehensive analysis.

Attendance in Group Meetings

I consistently upheld attendance in all group meetings, acknowledging the significance of proficient communication and cooperation. Consistent engagement facilitated my ability to remain informed about the project's advancements, offer valuable perspectives, and collaboratively tackle any obstacles.

Acquiring Presentation Materials

Assuming the responsibility of acquiring presentation materials, I ensured that our project presentation possessed visual appeal and was consistent with the primary messages we intended to communicate. The process encompassed conducting research, carefully choosing, and systematically arranging visual aids with the intention of augmenting the overall effectiveness of our findings.

Creation of Histogram for Dystopia Score:

One important component of our analysis entailed the construction of a histogram to represent the Dystopia Score. Utilizing my expertise in data visualization, I successfully generated a histogram that accurately depicted the distribution of Dystopia Scores within the dataset. The inclusion of this visual representation enhanced the comprehensiveness of our presentation and promoted a more lucid comprehension of the data.

Challenges Encountered

The project encountered several challenges, such as the need to effectively coordinate schedules for group meetings, handle diverse datasets, and ensure the precision of visual representations. Nevertheless, by means of transparent communication and cooperative resolution of issues, the team effectively managed to overcome these obstacles.

Key Takeaways

Participating in the collaborative endeavor has yielded significant knowledge regarding the efficacy of collaborative efforts, the significance of comprehensive examination of datasets, and the potency of visual aids in effectively communicating intricate information. Moreover, this experience has enhanced my capacity to make valuable contributions to a collaborative analytical project.

In conclusion, my contributions played a crucial role in the group project by significantly contributing to the analysis and presentation of the World Happiness Report dataset. The successful collaboration with fellow team members facilitated the resolution of obstacles and facilitated the production of a comprehensive report that extensively explored the intricacies of global happiness. This experience has engendered within me an elevated recognition of the intricacies and subtleties that are inherent in the analysis of extensive datasets.

Name: Vineeth Appikatla

Student ID: 2122039

1. Data Review

The variables in the dataset included Happiness Score, Freedom, Generosity, GDP per Capita, Health Life Expectancy, Perception, Social Support, Dystopia and Year. The commonalities among them were their contribution to understanding the determinants of happiness. Happiness Score, GDP per Capita, and Health Life Expectancy showed a strong positive correlation, highlighting them as key variables. Freedom also presented as a significant variable but with a weaker correlation to happiness.

2. Visualization Techniques

For the data, initially I have chosen Heatmaps and geographical maps as the primary visualization techniques due to their ability to effectively handle both continuous and categorical data types, as well as their ability to illustrate correlations and geographical patterns. This choice extends beyond the limitations of simple graphs and offers multidimensional perspectives that enhance data comprehension. For instance, heatmaps provide a rapid and comprehensive overview of correlations between multiple variables simultaneously, while world maps reveal spatial distributions that cannot be readily discerned from tables or simple charts.

3. Coding Effort

A statistical software like tableau and a programming language called R was used for the coding. Such scripting enables the creation of complex visualizations like heat maps and interactive elements efficiently.

4. Challenges

While heatmaps were initially considered as a potential visualization technique, they proved to be less effective for this particular dataset due to the nature of the variables and the complexity of the relationships, so I did not consider taking it to the further steps in the project. Despite this, the exploration of heatmaps provided valuable insights into the challenges of data visualization and the importance of selecting the most appropriate techniques for the data at hand.

5. Conclusions and Further Analysis

The visualizations provided compelling evidence that higher GDP and better health are strongly associated with higher happiness scores. These findings underscore the importance of economic prosperity and healthcare access in promoting well-being across nations.

The data also invites further exploration through more sophisticated analytical techniques, such as regression models to quantify the precise impact of each factor on happiness and cluster analysis to group countries based on these factors. Such investigations would deepen our understanding of the complex interplay between economic, social, and health determinants of happiness and inform the development of targeted policies that effectively promote well-being around the world.

6. Further Development

In future, I would explore the addition of interactive elements to the existing visualizations to allow users to engage with the data more deeply. Moreover, I would consider longitudinal studies to see how relationships between variables evolve over time, perhaps using animated visualizations.

7. Reflection Summary

By taking this course and doing this projected I have definitely learned the importance of selecting the right visualization technique for different types of data and relationships. It highlighted the necessity of good design practices to make complex data understandable and engaging. Additionally, it demonstrated the power of visual storytelling in conveying insights that might be opaque in raw numerical data. The use of coding not only streamlined the visualization process but also opened up possibilities for advanced and interactive elements that could lead to richer data exploration experiences.