HUMBER INSTITUTE OF TECHNOLOGY

AND ADVANCED LEARNING

(HUMBER COLLEGE)

**Group Project:**

**Canada’s Quality of Life Framework**

COURSE: **Big Data 2 for BIA - BIA-5303-0LA**

TEAM: **Group 3**

SUBMITTED BY:

|  |  |  |
| --- | --- | --- |
| **Last Name** | **First Name** | **Student Number** |
| Castro | Francis Angelo | N01605741 |
| Balgos | Harold Raymark | N01603663 |
| Ali | Mohammed | N01590561 |
| Nayab | Rameesha | N01607151 |
| Nguyen | Thi Thach Thao | N01606717 |
| Nguyen | Tuan Huy | N01603873 |

SUBMITTED TO: Professor Ammar Al-Qaraghuli

SUBMISSION DATE: December 7, 2024

TABLE OF CONTENTS:

[INTRODUCTION 3](#_Toc184379179)

[CANADA’S QUALITY OF LIFE FRAMEWORK 3](#_Toc184379180)

[**1.** **PROSPERITY** 3](#_Toc184379181)

[**2.** **HEALTH** 6](#_Toc184379182)

[**3.** **SOCIETY** 14](#_Toc184379183)

[**4.** **ENVIRONMENT** 20](#_Toc184379184)

[**5.** **GOOD GOVERNANCE** 25](#_Toc184379185)

[OVERALL CONCLUSION 30](#_Toc184379186)

[INFOGRAPHIC 30](#_Toc184379187)

[PRESENTATION 32](#_Toc184379188)

# **INTRODUCTION**

Canada's Quality of Life framework provides a comprehensive view of Canadians' well-being across various domains like health, prosperity, society, environment, and good governance. These provide a holistic view of the country's social, economic, and environmental health.

# **CANADA’S QUALITY OF LIFE FRAMEWORK**

## **PROSPERITY**

***Dataset:***



***Code:***

# Install necessary packages

install.packages(c("tidyverse", "readxl", "tibble"))

# Load libraries

library(tidyverse)

library(readxl)

library(tibble)

# Load the Excel file

df <- read\_excel("Prosperity.xlsx")

# Transpose the data

df\_transposed <- as.data.frame(t(df))

colnames(df\_transposed) <- df\_transposed[1, ]

df\_transposed <- df\_transposed[-1, ]

df\_transposed <- rownames\_to\_column(df\_transposed, var = "Year")

# Convert the Year column to numeric

df\_transposed$Year <- as.numeric(df\_transposed$Year)

df\_long <- df\_transposed %>%

pivot\_longer(cols = -Year, names\_to = "variable", values\_to = "value")

#Filter the data for the bar graph

bar\_data <- df\_long %>%

filter(variable %in% c("Median market income", "Median after-tax income", "Median government transfers"))

# Reorder the factor levels to ensure "Median government transfers" is last

bar\_data$variable <- factor(bar\_data$variable, levels = c("Median market income", "Median after-tax income", "Median government transfers"))

# Plot the bar graph with different shades of green

bar\_plot <- ggplot(bar\_data, aes(x = Year, y = round(as.numeric(value)), fill = variable)) +

geom\_bar(stat = "identity", position = "dodge") +

scale\_fill\_manual(values = c("Median market income" = "#b2df8a", "Median after-tax income" = "#33a02c", "Median government transfers" = "#1b5e20")) +

labs(title = "Prosperity Domain - Income, Employment, and Poverty Measures", x = "Year", y = "2022 constant dollars") +

ylim(0, max(as.numeric(bar\_data$value)) \* 1.2) + # Increase y-axis range

theme\_minimal() +

theme(legend.position = "bottom", panel.grid.major = element\_line(color = "grey"), panel.grid.minor = element\_line(color = "lightgrey"), panel.background = element\_rect(fill = "white"), plot.background = element\_rect(fill = "white"), plot.title = element\_text(face = "bold"))

# Filter the data for the line graph

line\_data <- df\_long %>%

filter(variable %in% c("Poverty Rate (%)", "Unemployment rate (%)"))

# Create a secondary y-axis

secondary\_y\_axis <- sec\_axis(~ ., name = "Rate (%)", breaks = seq(0, 20, by = 2))

# Plot the combined graph

combined\_plot <- ggplot() +

geom\_bar(data = bar\_data, aes(x = Year, y = round(as.numeric(value)), fill = variable), stat = "identity", position = "dodge") +

geom\_line(data = line\_data, aes(x = Year, y = as.numeric(value) \* (max(as.numeric(bar\_data$value)) \* 1.2 / 20), color = variable, group = variable), size = 1.5) +

geom\_text(data = line\_data, aes(x = Year, y = as.numeric(value) \* (max(as.numeric(bar\_data$value)) \* 1.2 / 20), label = sprintf("%.1f", as.numeric(value))), vjust = -0.5, color = "white", size = 4, fontface = "bold") +

scale\_fill\_manual(values = c("Median market income" = "#b2df8a", "Median after-tax income" = "#33a02c", "Median government transfers" = "#1b5e20")) +

scale\_color\_manual(values = c("Poverty Rate (%)" = "red", "Unemployment rate (%)" = "#ff9900")) +

labs(title = "Prosperity Domain - Income, Employment, and Poverty Measures", x = "Year", y = "2022 constant dollars") +

theme\_minimal() +

theme(legend.position = "bottom",

panel.grid.major = element\_blank(), # Remove major gridlines

panel.grid.minor = element\_blank(), # Remove minor gridlines

panel.background = element\_rect(fill = "#def7de"), # Set background color to light salmon

plot.background = element\_rect(fill = "#def7de"), # Set plot background color to light salmon

plot.title = element\_text(face = "bold", hjust = 0.5)) +

scale\_y\_continuous(sec.axis = secondary\_y\_axis)

print(combined\_plot)

# Save the plot

ggsave("Prosperity Domain.png", plot = combined\_plot, width = 10, height = 6, dpi = 300)

***Graph:***A graph of green and orange bars

Description automatically generated

***Analysis:***

The Prosperity domain measures economic stability and growth, including income, employment, and housing. Here, we see that while median market income slowed in 2020, after-tax income, and especially government transfers, have jumped during the pandemic. We can also see here that unemployment spiked but poverty plunged. All this is due to government support -stimulus - that aimed to mitigate the negative impact of lockdowns. Recently, these measures are returning to their pre-pandemic levels. 2022 saw a decrease in incomes, a rise in poverty, and a decrease in unemployment.

## **HEALTH**

***Datasets:***



***Code:***

# Install necessary packages

install.packages(c("sf", "ggplot2", "ggrepel", "ggimage", "rnaturalearth", "rnaturalearthhires", "rnaturalearthdata", "readr", "tidyr", "tidyverse", "dplyr"))

# Load libraries

library(sf)

library(ggplot2)

library(ggrepel)

library(ggimage)

library(rnaturalearth)

library(rnaturalearthdata)

library(rnaturalearthhires)

library(readr)

library(tidyr)

library(tidyverse)

library(dplyr)

# Load the datasets

lifex\_data <- read.csv("LifeExpectancy.csv")

Health\_data <- read\_csv("HealthData.csv", col\_names = TRUE, show\_col\_types =FALSE)

metrics\_data <- read.csv("Health\_Metrics.csv")

unmet\_data <- read.csv("UnmetNeedsHealthcare.csv")

# Round up values

lifex\_rounded <- lifex\_data %>%

mutate(across(everything(), ~ ifelse(is.na(as.numeric(.)), ., ceiling(as.numeric(.)))))

# Load Canada's map data

canada\_map <- ne\_states(country = "Canada", returnclass = "sf")

# Join the percentage data with the map data

canada\_map <- canada\_map %>%

left\_join(lifex\_rounded, by = c("name\_en" = "Geography"))

# Calculate centroids for province labels

canada\_map <- canada\_map %>%

mutate(centroid = st\_centroid(geometry)) %>%

mutate(longitude = st\_coordinates(centroid)[, 1],

latitude = st\_coordinates(centroid)[, 2])

# Plot the Heatmap for Life Expectancy

ggplot(data = canada\_map) +

geom\_sf(aes(fill = BothSex), color = "black", size = 0.2) +

geom\_text\_repel(aes(x = longitude, y = latitude, label = name),

color = "black", size = 3, fontface = "bold") +

scale\_fill\_gradient(low = "pink", high = "red4", na.value = "snow3",

name = "Life expectancy (Years)") +

theme\_minimal() +

labs(title = "Life Expectancy per region") +

theme(plot.title = element\_text(size = 16, face = "bold",

color = "black", hjust = 0.5),

plot.background = element\_rect(fill = "#f8eceb", color = NA),

panel.background = element\_rect(fill = "#f8eceb", color = NA),

panel.grid = element\_blank(),

axis.text.x = element\_blank(),

axis.text.y = element\_blank(),

axis.title.x = element\_blank(),

axis.title.y = element\_blank(),

legend.title = element\_text(size = 11, face = "bold"),

legend.position = "right")

# Transpose Health data

Health\_data\_long <- Health\_data %>%

pivot\_longer(cols = c(Percentage\_with\_unmet\_needs, Arthritis, Diabetes,

High\_blood\_pressure, Current\_smoker),

names\_to = "Health\_Indicator", values\_to = "Value")

# Color Palette

pal <- c("indianred4", "tomato", "lightpink1", "orangered4", "hotpink2")

# Plot the chart for Health Data

ggplot() +

# Bar graph for Very\_good\_Perceived\_health

geom\_bar(data = Health\_data, aes(x = Year, y = Very\_good\_Perceived\_health, fill = Age\_group),

stat = "identity", position = "stack", color = "black", size = 0.3) +

scale\_fill\_manual(values = pal) +

scale\_color\_manual(values = pal) +

labs(title = "Population with Good Health by Age Group",

x = "",

y = "in Millions",

fill = "Age Group",

color = "Health Indicator") +

theme\_minimal() +

theme(panel.grid = element\_blank(),

panel.grid.major = element\_blank(),

panel.grid.minor = element\_blank(),

panel.border = element\_blank(),

plot.title = element\_text(size = 20, face = "bold", color = "black", hjust = 0.5),

plot.background = element\_rect(fill = "#f8eceb"),

axis.text.x = element\_text(size = 11, hjust = 1, color = "black", face = "bold"),

axis.text.y = element\_text(size = 11, hjust = 1, color = "black", face = "bold"),

axis.line.y = element\_line(color = "grey", size = 0.5),

axis.line.x = element\_line(color = "grey", size = 0.5),

axis.ticks = element\_blank(),

legend.title = element\_text(size = 11, hjust = 1, color = "black", face = "bold"),

legend.text = element\_text(size = 11, color = "black"),

legend.position = "top")

# Transpose Health Metrics data

metrics\_long <- metrics\_data %>%

pivot\_longer(cols = -Year, names\_to = "Metric", values\_to = "Value")

# Color Palette

pal2 <- c("indianred4", "darkorange3", "lightcoral", "red3")

# Plot the chart for Health Metrics Data

ggplot(metrics\_long, aes(x = Year, y = Value, color = Metric)) +

geom\_line(aes(color = Metric), linewidth = 2) +

geom\_point(aes(shape = Metric, color = as.character(Metric)), size = 5) +

scale\_shape\_manual(values = c(21,22,23, 24)) +

scale\_fill\_manual(values = pal2) +

scale\_color\_manual(values = pal2) +

labs(title = "Health Metrics Over Years",

x = "",

y = "Population in Millions",

color = "Metric") +

coord\_cartesian(ylim = c(2, 7)) +

theme\_minimal() +

theme(panel.grid = element\_blank(),

panel.grid.major = element\_blank(),

panel.grid.minor = element\_blank(),

panel.border = element\_blank(),

plot.title = element\_text(size = 15, face = "bold", color = "black", hjust = 0.5),

plot.background = element\_rect(fill = "#f8eceb"),

axis.text.x = element\_text(size = 11, hjust = 1, color = "black", face = "bold"),

axis.text.y = element\_text(size = 11, hjust = 1, color = "black", face = "bold"),

axis.line.y = element\_line(color = "grey", size = 0.5),

axis.line.x = element\_line(color = "grey", size = 0.5),

axis.ticks = element\_blank(),

legend.title = element\_text(size = 11, hjust = 1, color = "black", face = "bold"),

legend.text = element\_text(size = 11, color = "black"),

legend.position = "top")

# Plot the chart for Unmet Needs in Healthcare Data

ggplot(unmet\_data, aes(x = Year, y = Percentage)) +

geom\_line(linewidth = 1) +

geom\_point(size = 1) +

geom\_image(aes(image = icon), size = 0.1) +

labs(title = "Unmet Healthcare Needs", x = "", y = "Percentage") +

coord\_cartesian(ylim = c(4, 10)) +

theme\_minimal() +

theme(panel.grid = element\_blank(),

panel.grid.major = element\_blank(),

panel.grid.minor = element\_blank(),

panel.border = element\_blank(),

plot.title = element\_text(size = 15, face = "bold", color = "black", hjust = 0.5),

plot.background = element\_rect(fill = "#f8eceb"),

axis.text.x = element\_text(size = 11, hjust = 1, color = "black", face = "bold"),

axis.text.y = element\_text(size = 11, hjust = 1, color = "black", face = "bold"),

axis.line.y = element\_line(color = "grey", size = 0.5),

axis.line.x = element\_line(color = "grey", size = 0.5),

axis.ticks = element\_blank(),

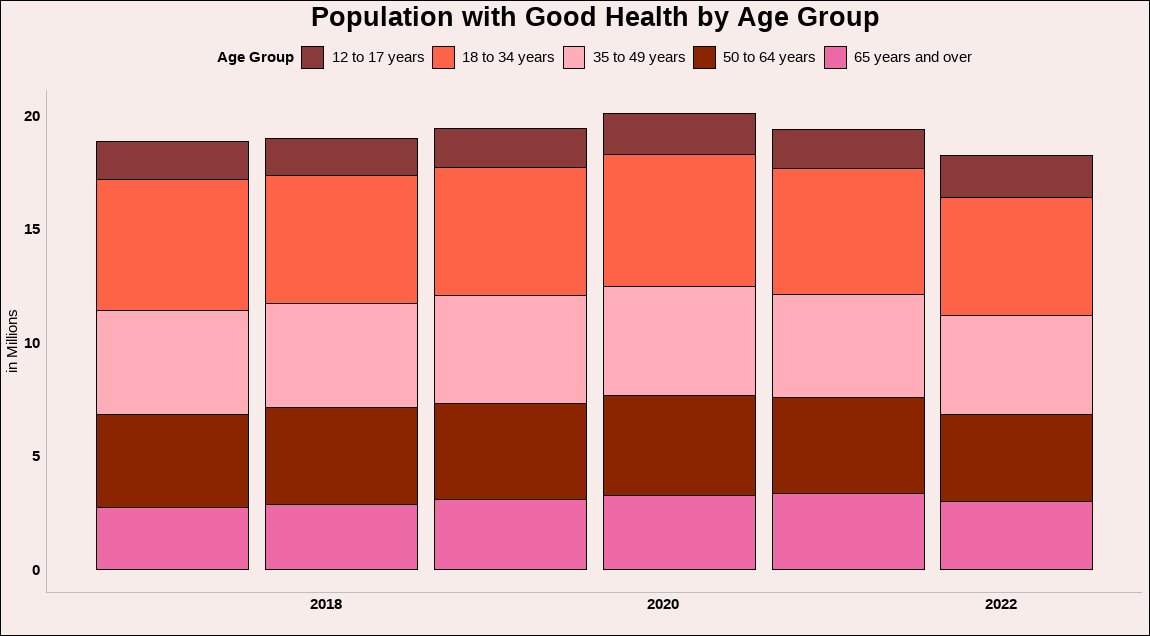
legend.title = element\_text(size = 11, hjust = 1, color = "black", face = "bold"),

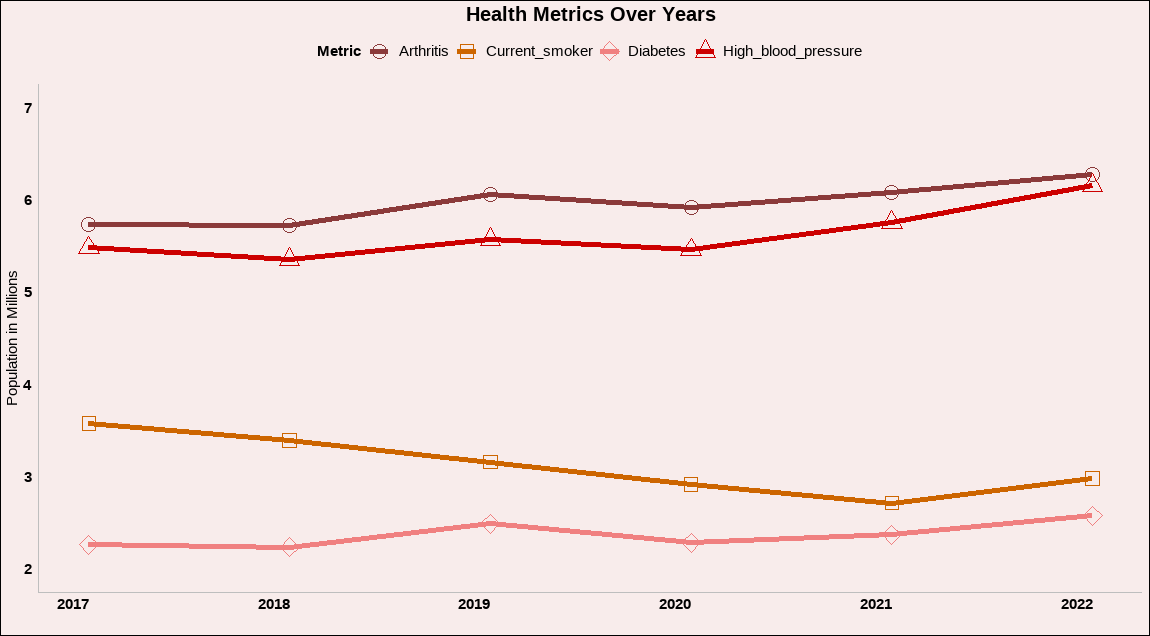
legend.text = element\_text(size = 11, color = "black"),

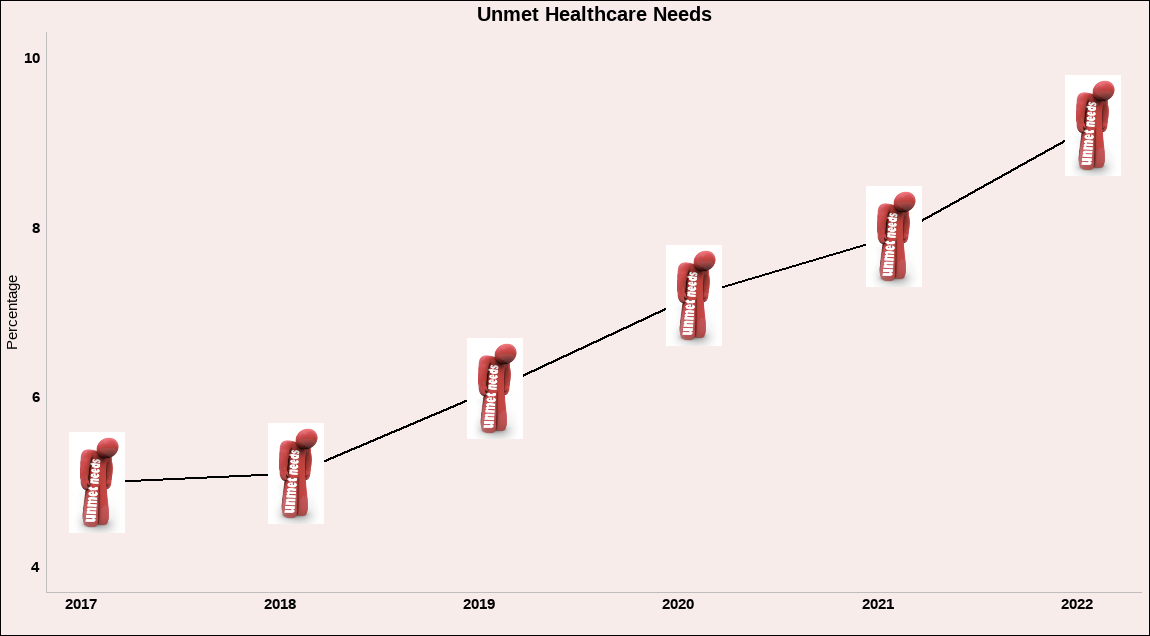
legend.position = "top")

***Graph:***A map of canada with different colored areas

Description automatically generated







***Analysis:***

The Health domain focuses on physical and mental well-being, covering aspects like life expectancy, self-rated health, and access to healthcare. Here it can be observed that there is a decrease in the population with good health after the pandemic hit. The 3 major types of chronic diseases are all trending upward from 2020. There was also a significant increase in unmet healthcare needs from 2018 to 2022. Across regions, life expectancy is highest in Quebec and Newfoundland, which is quite surprising.

## **SOCIETY**

***Dataset:***



***Code:***

# Install necessary packages

install.packages(c("readr", "tidyr", "dplyr", "ggplot2", "stringr"))

# Load libraries

library(readr)

library(tidyr)

library(dplyr)

library(ggplot2)

library(stringr)

# Load the data

Someone\_to\_Count\_On\_Data <- read\_csv("Someone to Count On Data.csv")

Community\_Belonging\_Data <- read\_csv("Community Belonging Data.csv")

# Tidy the data

Someone\_to\_Count\_On\_Data\_2024 <- Someone\_to\_Count\_On\_Data %>%

rename(Age\_group = `Sociodemographic characteristics`,

Someone\_counton\_rate = Percentage) %>%

mutate(Someone\_counton\_rate = Someone\_counton\_rate / 100) %>%

filter(Year == "2024", Gender == "Total, all persons") %>%

select( Age\_group, Indicators, Someone\_counton\_rate)

Strong\_sense\_belonging\_data\_2024<- Community\_Belonging\_Data %>%

rename(Age\_group = `Sociodemographic characteristics`,

Community\_belonging\_rate = Percentage) %>%

mutate(Community\_belonging\_rate = Community\_belonging\_rate / 100) %>%

filter(Year == "2024", Gender == "Total, all persons",

Indicators=="Very strong or somewhat strong sense of belonging to local community")%>%

select( Age\_group,Indicators, Community\_belonging\_rate)

# Wrap text

Someone\_to\_Count\_On\_Data\_2024$Indicators <- str\_wrap(Someone\_to\_Count\_On\_Data\_2024$Indicators, width = 20)

Strong\_sense\_belonging\_data\_2024$Indicators <- str\_wrap(Strong\_sense\_belonging\_data\_2024$Indicators, width = 20)

# Plot the "Someone to Count On" data

ggplot() +

# 100% Stacked bar chart for Someone\_counton\_rate

geom\_bar(data = Someone\_to\_Count\_On\_Data\_2024,

aes(x = Age\_group, y = Someone\_counton\_rate, fill = Indicators),

position = "fill", stat = "identity", alpha = 0.8) +

# Line chart for Community\_belonging\_rate referencing the same y-axis

geom\_line(data = Strong\_sense\_belonging\_data\_2024,

aes(x = Age\_group,

y = Community\_belonging\_rate,

color = Indicators,

group = Indicators),

size = 1.5) +

# Points for Community\_belonging\_rate

geom\_point(data = Strong\_sense\_belonging\_data\_2024,

aes(x = Age\_group,

y = Community\_belonging\_rate,

color = Indicators),

size = 3) +

# Define the y-axis

scale\_y\_continuous(name = "Percentage (%)",

limits = c(0, 1), # 0 to 100% (stacked bar and line share the same scale)

labels = scales::percent\_format()) + # Format y-axis labels

# Customize bar chart legend colors and line legend color

scale\_fill\_manual(values = c("lightgoldenrod", "yellow", "darkgoldenrod4"), # Specify stack colors

name = "Someone to Count On Response") + # Legend title for stack bar

scale\_color\_manual(values = c("gold4"), # Line color

name = "Community Belonging Response") + # Legend title for the line

labs(title = "2024 Distribution of Response in Someone to Count On \n & Percent of Strong Sense of Community belonging by Age group", # Add your graph title

x = "Age Group",

y = "Percentage (%)") +

theme(panel.border = element\_rect(color = "#FFF4D4", fill = NA, size = 1),

panel.grid = element\_blank(),

panel.background = element\_rect(fill = "#FFF4D4", color = NA), # White panel background

plot.background = element\_rect(fill = "#FFF4D4", color = NA), # Light gray plot background

axis.text.x = element\_text(angle = 0, hjust = 0.5, color = "black"), # Black x-axis text

axis.text.y = element\_text(color = "black"), # Black y-axis text

axis.title = element\_text(face = "bold"), # Bold axis titles

axis.ticks = element\_blank(),

plot.title = element\_text(hjust = 0.5, size = 14, face = "bold"), # Center-align title

legend.background = element\_rect(fill = "#FFF4D4", color = NA), # Light gray legend background

legend.key = element\_rect(fill = "white", color = NA), # White legend keys

legend.position = "top") #"right" # Place the legend on the right side

# "Someone to Count On" yearly data

Someone\_to\_Count\_On\_Data\_yearly <- Someone\_to\_Count\_On\_Data %>%

rename(Age\_group = `Sociodemographic characteristics`,

Someone\_counton\_rate = Percentage) %>%

mutate(Someone\_counton\_rate = Someone\_counton\_rate / 100) %>%

filter(Gender == "Total, all persons",

Indicators == "Always or often has people to depend on when needed")%>%

select(Year, Indicators, Someone\_counton\_rate)

Someone\_to\_Count\_On\_Data\_yearly\_avg <- Someone\_to\_Count\_On\_Data\_yearly %>%

group\_by(Year) %>% # Group data by Year

summarize(avg\_Someone\_counton\_rate = mean(Someone\_counton\_rate,

na.rm = TRUE)) # Calculate the mean for each year

Someone\_to\_Count\_On\_Data\_yearly\_avg <- Someone\_to\_Count\_On\_Data\_yearly\_avg %>%

mutate(avg\_Someone\_counton\_rate\_percent = avg\_Someone\_counton\_rate \* 100) # Convert to percentage

# Plot the "Someone to Count On" yearly data

ggplot(Someone\_to\_Count\_On\_Data\_yearly\_avg, aes(x = Year, y = avg\_Someone\_counton\_rate)) +

# Line plot with a smooth curve

geom\_line(color = "#FBC02D", size = 1.5) +

# Points with bold circles

geom\_point(color = "#FBC02D", size = 6) +

geom\_point(color = "white", size = 3) + # Inner white circle to mimic design

# Add labels for each point

geom\_text(aes(label = paste0(round(avg\_Someone\_counton\_rate\_percent, 1), "%")),

vjust = -0.8, size = 5, fontface = "bold", color = "black") +

# Custom y-axis for percentages

scale\_y\_continuous(labels = scales::percent\_format(accuracy = 1),

limits = c(0.7, 0.8)) +

# Adjust x-axis breaks

scale\_x\_continuous(breaks = Someone\_to\_Count\_On\_Data\_yearly\_avg$Year) +

# Titles and labels

labs(title = "Trends in Population Reporting Dependable Support",

x = NULL, # No x-axis label

y = NULL) + # No y-axis label

# Further reduce vertical space

coord\_cartesian(ylim = c(0.725, 0.775)) +

# Theme customization

theme\_minimal() +

theme(text = element\_text(size = 14),

axis.text.x = element\_text(size = 12, color = "black"),

axis.text.y = element\_blank(), # Remove y-axis labels

axis.ticks.y = element\_blank(), # Remove y-axis ticks

panel.grid.major.x = element\_blank(), # No vertical gridlines

panel.grid.minor = element\_blank(), # No minor gridlines

panel.grid.major.y = element\_blank(), # No horizontal gridlines

plot.title = element\_text(size = 16, face = "bold", hjust = 0.5),

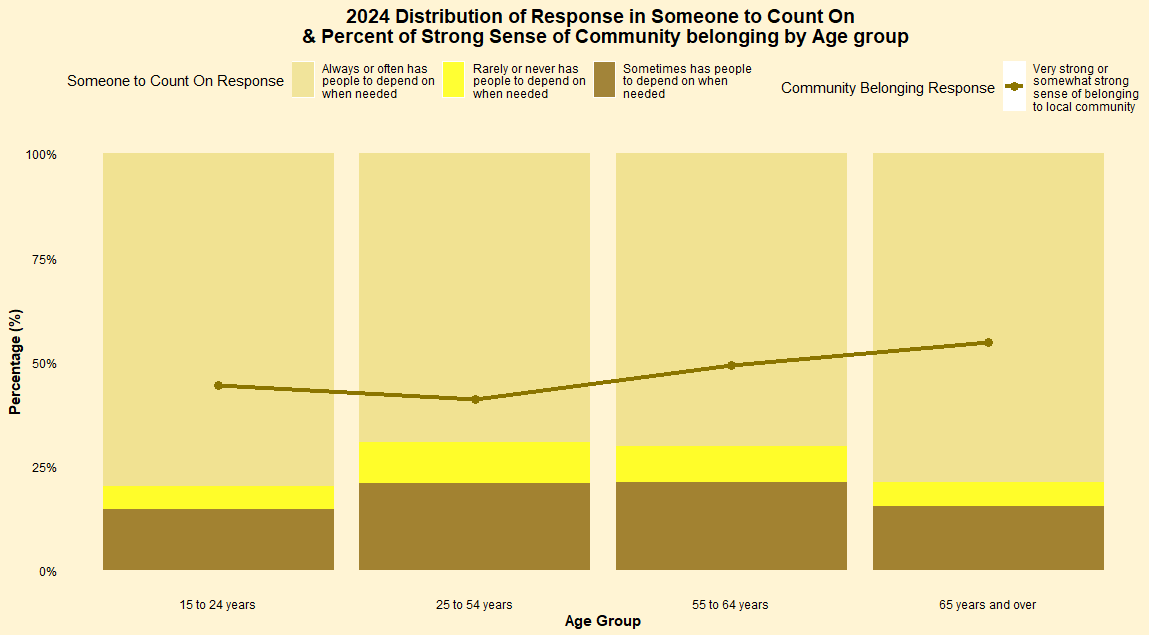
plot.background = element\_rect(fill = "#FFF4D4", color = NA), # Match background color

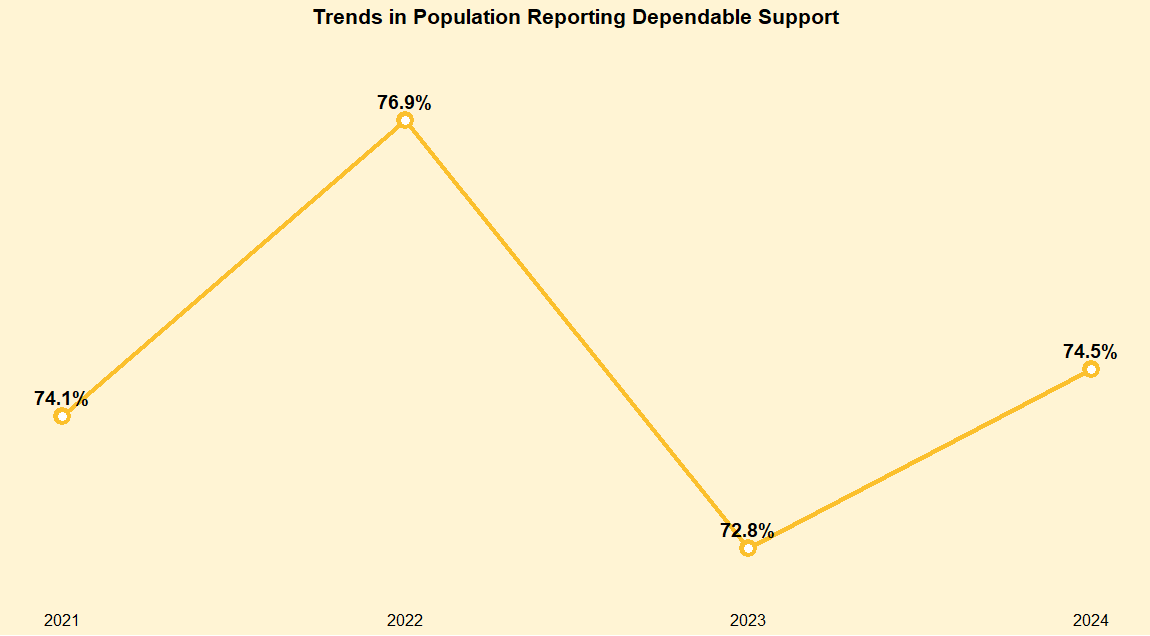
panel.background = element\_rect(fill = "#FFF4D4", color = NA))

# Save the plot

ggsave("dependable\_support\_plot.png", width = 8, height = 2.5, dpi = 300)

***Graphs:***





***Analysis:***

The Society domain assesses social connections and community engagement, including the sense of belonging and social support. In 2024, 74.5% of respondents reported that they get dependable support. While not as high as 2022's 76.9%, this was an improvement from 2023's figure of 72.8%. Individuals who reported having someone to rely on tend to feel a strong sense of belonging within their local community. Across age groups, youth and senior levels reported a higher level of support than other age groups.

## **ENVIRONMENT**

***Dataset:***



***Code:***

# Install necessary packages

install.packages(c("tidyverse", "readxl", "ggstream", "ggtext"))

# Load libraries

library(tidyverse)

library(readxl)

library(ggstream)

library(ggtext)

# Load the data

emission\_data <- read\_csv("Enviroment\_2002\_2022.csv")

# Transform data to long format for ggplot2

emission\_data\_long <- emission\_data %>%

pivot\_longer(cols = -Year,

names\_to = "Sector",

values\_to = "Emissions") %>%

mutate(Year = as.integer(Year), # Ensure Year is numeric

Emissions = as.numeric(Emissions)) %>%

filter(Year >= 2002 & Year <= 2022) # Filter data for 2002-2022

colnames(emission\_data\_long)

#Color palette

pal <- c("olivedrab4", "mediumaquamarine", "springgreen3", "seagreen",

"darkseagreen", "lightgreen", "greenyellow")

# Stacking order

stkord <- c("Waste and others", "Agriculture", "Heavy industry", "Electricity",

"Buildings", "Transport", "Oil and gas")

# Plot the chart

plot <- emission\_data\_long %>%

arrange(Emissions) %>%

mutate(Sector = factor(Sector, levels = stkord)) %>%

ggplot(aes(Year, Emissions, fill = Sector,

label = Sector, color = Sector)) +

geom\_stream(type = "ridge", bw = 1) +

annotate("text", x = 2005, y = 900,

label = "Greenhouse Gas\nEmissions (in Megatonnes)",

hjust=0, size=6, lineheight=.9, fontface="bold", color="black") +

#Waste and others

annotate("text", x = 2022.2, y = 800,

label = "Waste and others",

hjust=0, size=3.5, lineheight=.8, fontface="bold", color="black") +

#Agriculture

annotate("text", x = 2022.2, y = 750,

label = "Agriculture",

hjust=0, size=3.5, lineheight=.8, fontface="bold", color="black") +

#Heavy industry

annotate("text", x = 2022.2, y = 660,

label = "Heavy industry",

hjust=0, size=3.5, lineheight=.8, fontface="bold", color="black") +

#Electricity

annotate("text", x = 2022.2, y = 590,

label = "Electricity",

hjust=0, size=3.5, lineheight=.8, fontface="bold", color="black") +

#Buildings

annotate("text", x = 2022.2, y = 540,

label = "Buildings",

hjust=0, size=3.5, lineheight=.8, fontface="bold", color="black") +

#Transport

annotate("text", x = 2022.2, y = 440,

label = "Transport",

hjust=0, size=3.5, lineheight=.8, fontface="bold", color="black") +

#Oil and gas

annotate("text", x = 2022.2, y =250,

label = "Oil and gas",

hjust=0, size=3.5, lineheight=.8, fontface="bold", color="black") +

# Vertical segments

geom\_segment(aes(x = 2002, y = 0, xend = 2002,

yend = 800+100),color="black") +

geom\_point(aes(x = 2002, y = 800+100),color="black") +

annotate("text", x = 2002, y = 800+120,

label = "744.7", hjust=0.5, size=3.5, lineheight=.8,

fontface="bold", color="black") +

geom\_segment(aes(x = 2005, y = 0, xend = 2005,

yend = 700+100),color="black") +

geom\_point(aes(x = 2005, y = 700+100),color="black") +

annotate("text", x = 2005, y = 700+120,

label = "761.5", hjust=0.5, size=3.5, lineheight=.8,

fontface="bold", color="black") +

geom\_segment(aes(x = 2010, y = 0, xend = 2010,

yend = 680+100),color="black") +

geom\_point(aes(x = 2010, y = 680+100),color="black") +

annotate("text", x = 2010, y = 680+120,

label = "728.6", hjust=0.5, size=3.5, lineheight=.8,

fontface="bold", color="black") +

geom\_segment(aes(x = 2015, y = 0, xend = 2015,

yend = 690+100),color="black") +

geom\_point(aes(x = 2015, y = 690+100),color="black") +

annotate("text", x = 2015, y = 690+120,

label = "745.8", hjust=0.5, size=3.5, lineheight=.8,

fontface="bold", color="black") +

geom\_segment(aes(x = 2020, y = 0, xend = 2020,

yend = 670+100),color="black") +

geom\_point(aes(x = 2020, y = 670+100),color="black") +

annotate("text", x = 2020, y = 670+120,

label = "686.4", hjust=0.5, size=3.5, lineheight=.8,

fontface="bold", color="black") +

geom\_segment(aes(x = 2022, y = 0, xend = 2022,

yend = 720+100),color="black") +

geom\_point(aes(x = 2022, y = 720+100),color="black") +

annotate("text", x = 2022, y = 720+120,

label = "707.7", hjust=1.1, size=3.5, lineheight=.8,

fontface="bold", color="black") +

#Color scale

scale\_fill\_manual(values=pal) +

scale\_color\_manual(values=pal) +

scale\_x\_continuous(breaks = c(2002,2005,2010,2015,2020,2022),

labels = c("2002","2005","2010","2015","2020","2022")) +

scale\_y\_continuous(expand = c(0,0)) +

#Last customization

coord\_cartesian(clip = "off") +

xlab("") +

ylab("") +

theme(axis.line.x = element\_line(linewidth = .75),

panel.grid = element\_blank(),

axis.text.y = element\_blank(),

axis.text.x = element\_text(color = "black", size = 10,

margin = margin(5,0,0,0)),

plot.margin = margin(20,120,20,20),

legend.position = "none",

plot.caption = element\_markdown(hjust = 0, margin = margin(10,0,0,0),

size = 8, color = "black",

lineheight = 1.2),

plot.background = element\_rect(fill = "#aadaca", color = NA),

panel.background = element\_rect(fill = "#aadaca", color = NA))

# Display the chart

print(plot)

***Graph:***A chart of greenhouse gas emissions

Description automatically generated

***Analysis:***

The Environment domain evaluates the quality of natural and built surroundings, such as air and water quality, and green spaces. Intuitively, there has been a steady increase in emissions since the pandemic as economies opened up, with greenhouse gas emissions jumping to 707.7 megatonnes in 2022 from just 686.4 megatonnes in 2020. The top 3 culprits are oil, gas, and transport comprising 31%, 22%, and 13% of all greenhouse gas emissions. Despite this, 2022's emissions levels are still below any point before the pandemic hit. The transition to green energy should result in a continued decrease in harmful emissions in the future.

## **GOOD GOVERNANCE**

***Dataset:***



***Code:***

# Install necessary packages

install.packages(c("dplyr", "ggplot2", "tidyr"))

# Load libraries

library(dplyr)

library(ggplot2)

library(tidyr)

# Data Preparation

file\_path <- "Discrimination.csv"

data <- read.csv(file\_path)

# Combine Data for Both Periods

race\_data <- data %>%

filter(!is.na(Race\_Colour\_5yrs\_COVID), !is.na(Race\_Colour\_COVID)) %>%

mutate(Indicators = factor(Indicators, levels = unique(Indicators)),

`Before COVID` = as.numeric(Race\_Colour\_5yrs\_COVID),

`Since COVID` = as.numeric(Race\_Colour\_COVID)) %>%

select(Indicators, `Before COVID`, `Since COVID`) %>%

pivot\_longer(cols = c(`Before COVID`, `Since COVID`),

names\_to = "Period",

values\_to = "Percentage")

# Separate the data into two groups: Male and Female

df\_before <- race\_data[race\_data$Period == "Before COVID", ]

df\_after <- race\_data[race\_data$Period == "Since COVID", ]

# Sort each group by Salary in descending order

df\_before <- df\_before[order(-df\_before$Percentage, decreasing = TRUE), ]

df\_after <- df\_after[order(-df\_after$Percentage, decreasing = TRUE), ]

# Convert Name to a factor to maintain order in the plot

df\_before$Indicators <- factor(df\_before$Indicators, levels = df\_before$Indicators)

df\_after$Indicators <- factor(df\_after$Indicators, levels = df\_after$Indicators)

# Create Bar Chart for Before Covid

before\_plot <- ggplot(df\_before,

aes(x = Indicators, y = Percentage, fill = Period)) +

geom\_bar(stat = "identity", color = "black", size = 0.3) +

coord\_flip() +

geom\_text(aes(label = round(Percentage, 1)),

position = position\_dodge(width = 0.9), hjust = 1.5, vjust = 0.5,

size = 3.5, color = "white") +

scale\_fill\_manual(values = "navyblue") +

theme\_minimal() +

labs(title = "Race/Colour-Based Discrimination",

x = "",

y = "Percentage (%)") +

theme(plot.title = element\_text(size = 18, face = "bold", hjust = 0.5),

plot.background = element\_rect(fill = "#d2e2e9"),

axis.text.x = element\_text(angle = 0, hjust = 1, size = 10),

axis.title.y = element\_text(size = 12, face = "bold"),

panel.grid.major = element\_blank(),

panel.grid.minor = element\_blank(),

legend.position = "top",

legend.title = element\_blank(),

legend.text = element\_text(size = 10))

# Display the chart (Before Covid)

print(before\_plot)

# Create Bar Chart for Since Covid

after\_plot <- ggplot(df\_after,

aes(x = Indicators, y = Percentage, fill = Period)) +

geom\_bar(stat = "identity", color = "black", size = 0.3) +

coord\_flip() +

geom\_text(aes(label = round(Percentage, 1)),

position = position\_dodge(width = 0.9), hjust = -0.5, vjust = 0.5,

size = 3.5) +

scale\_y\_reverse () +

scale\_fill\_manual(values = "lightskyblue1") +

theme\_minimal() +

labs(title = "Race/Colour-Based Discrimination",

x = "",

y = "Percentage (%)") +

theme(plot.title = element\_text(size = 18, face = "bold", hjust = 0.5),

plot.background = element\_rect(fill = "#d2e2e9"),

axis.text.x = element\_text(angle = 0, hjust = 1, size = 10),

axis.title.y = element\_text(size = 12, face = "bold"),

panel.grid.major = element\_blank(),

panel.grid.minor = element\_blank(),

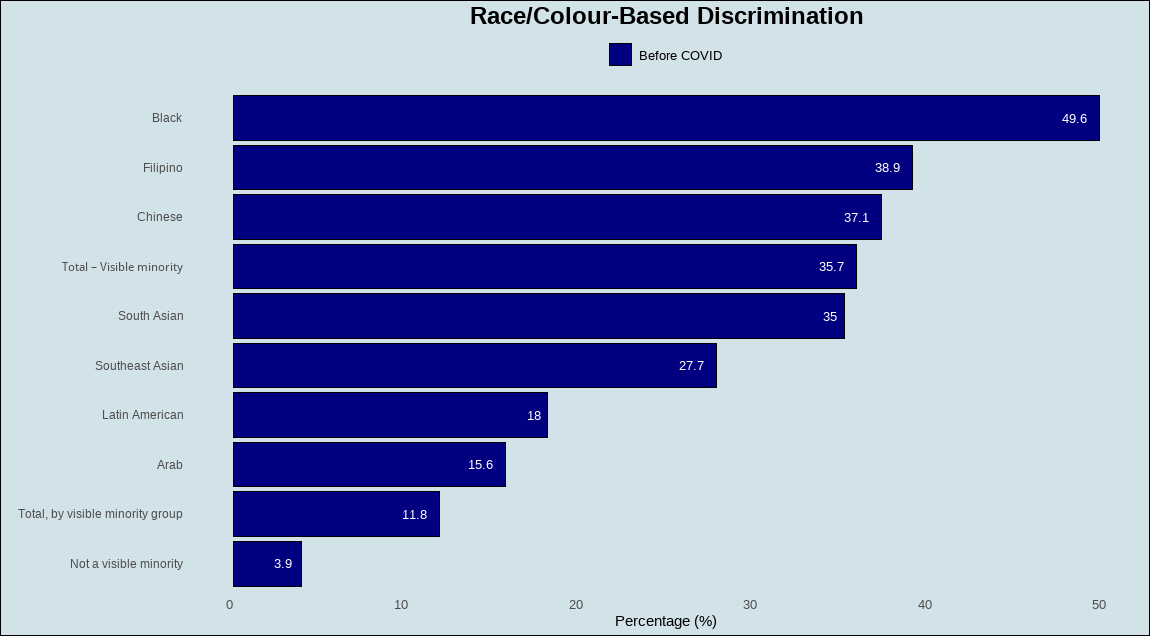
legend.position = "top",

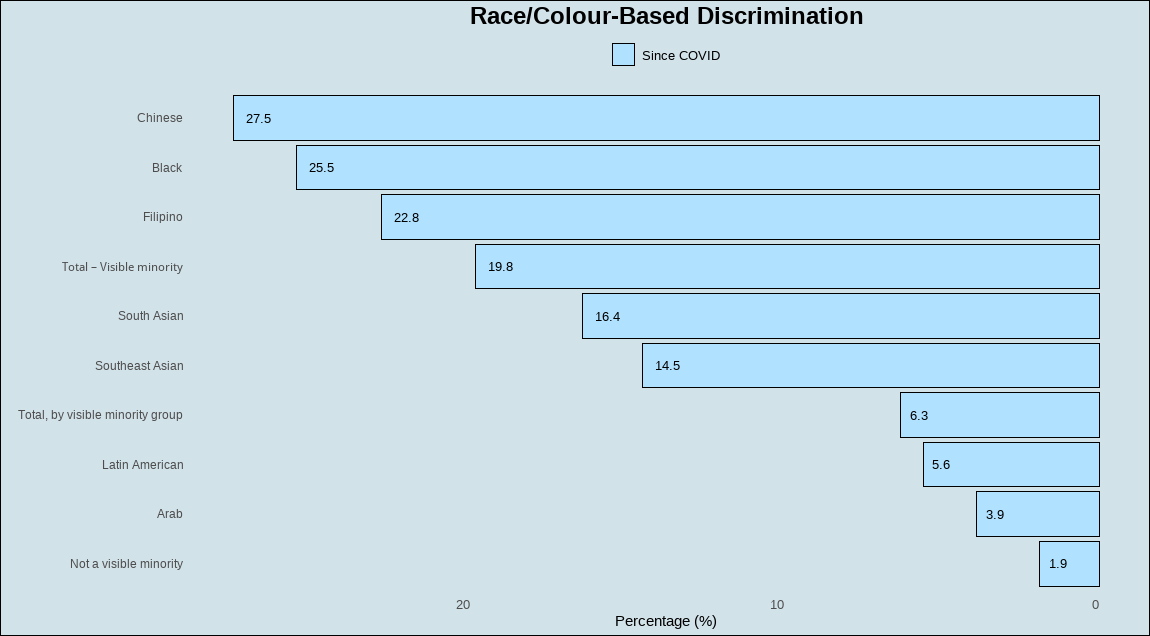
legend.title = element\_blank(),

legend.text = element\_text(size = 10))

# Display the chart (Since Covid)

print(after\_plot)

***Graph:***



***Analysis:***

The Good Governance domain measures the effectiveness and fairness of institutions, including trust in government and public services. Discrimination has significantly declined for all races/ethnic groups since the pandemic. Persons of Black, Arab, and Southeast Asian ethnicities saw the largest declines. Since the pandemic, there has also been a shift in the ethnicity that is most discriminated against - from Blacks to Chinese. This is likely attributed to the origins of the pandemic, which first spread in China.

# **OVERALL CONCLUSION**

Overall, the Prosperity, Health, and Environment domains worsened while the Society and Good Governance domains improved from the year prior. While 3 out of 5 domains deteriorated from the previous year, it is important to note that compared to pre-pandemic levels, both domains fared better. Notwithstanding data collection issues during the pandemic that may have distorted the figures for 2020, our group assesses that looking at the big picture, the Quality of Life in Canada still improved.

In conclusion, our group found this framework very useful because it provides a comprehensive and multidimensional assessment of well-being, helping policymakers and stakeholders make informed decisions to improve overall societal health and prosperity.

# **INFOGRAPHIC**



A screenshot of a poster

Description automatically generated

# **PRESENTATION**

Here’s the link to our video presentation:

[**https://humberital.sharepoint.com/:v:/r/sites/Group3-BigData2Project/Shared%20Documents/General/Recordings/Group%203%20-%20Big%20Data%202%20Project-20241205\_171124-Meeting%20Recording.mp4?csf=1&web=1&e=EaJBuQ&nav=eyJyZWZlcnJhbEluZm8iOnsicmVmZXJyYWxBcHAiOiJTdHJlYW1XZWJBcHAiLCJyZWZlcnJhbFZpZXciOiJTaGFyZURpYWxvZy1MaW5rIiwicmVmZXJyYWxBcHBQbGF0Zm9ybSI6IldlYiIsInJlZmVycmFsTW9kZSI6InZpZXcifX0%3D**](https://humberital.sharepoint.com/:v:/r/sites/Group3-BigData2Project/Shared%20Documents/General/Recordings/Group%203%20-%20Big%20Data%202%20Project-20241205_171124-Meeting%20Recording.mp4?csf=1&web=1&e=EaJBuQ&nav=eyJyZWZlcnJhbEluZm8iOnsicmVmZXJyYWxBcHAiOiJTdHJlYW1XZWJBcHAiLCJyZWZlcnJhbFZpZXciOiJTaGFyZURpYWxvZy1MaW5rIiwicmVmZXJyYWxBcHBQbGF0Zm9ybSI6IldlYiIsInJlZmVycmFsTW9kZSI6InZpZXcifX0%3D)