

# Post-COVID-19 Readiness in Canada: Demographics and Activity Insights

Caterina Tasinato

## DATA PROCESSING

```
# Load 'tidyverse' library
library(tidyverse)

# Load 'showtext' library
library(showtext)

# Add Google font Montserrat
font_add_google("montserrat")

# Enable showtext to use the fonts
showtext_auto()

# Import the dataset from the specified file path into a data frame
survey <- read.csv("~/Desktop/DOCUMENTS/ds_research_data.csv", header=TRUE)

# View the data to verify the results
View(survey)

# Remove the first 2 rows (question id, question name)
survey <- survey[-c(1, 2), ]

# View the data to verify the results
View(survey)
```

```
# Check data information
glimpse(survey) # 1588 rows, 129 columns
```

## DEMOGRAPHICS

```
# Count missing values for 'gender' and 'province' from the source data
missing_gender <- survey %>%
  summarise(missing_gender = sum(is.na(gender)))

missing_province <- survey %>%
  summarise(missing_province = sum(is.na(Province)))

# Inspect unique values for 'gender' and 'province' from the source data
unique_genders <- unique(survey$gender)
unique_provinces <- unique(survey$Province)

# Print results
missing_gender      # No missing values
missing_province    # No missing values
unique_genders      # 'Male', 'Female', 'I prefer not to say', 'Other'
unique_provinces    # 10 Provinces + Territories
```

```
# Transform data to create a subset
province_sample <- survey %>%
  # Rename the 'Province' column to 'province'
  rename(province = Province) %>%
  # Filter 'gender' and recode 'province'
  filter(gender %in% c("Male", "Female")) %>%
  # Recode certain province names in the 'province' column
  mutate(province = recode(province,
                           "BC" = "BC and Territories",
                           "Territories" = "BC and Territories",
                           "NovaScotia" = "Nova Scotia",
                           "NewBrunswick" = "New Brunswick")) %>%

# Group by 'province' and 'gender', then count occurrences
group_by(province, gender) %>%
  summarise(count = n()) %>%
```

```
# Calculate total and percentages of 'gender' within each province
group_by(province) %>%
mutate(total = sum(count),
       percentage = round((count / total) * 100)) %>%
ungroup()
```

`summarise()` has grouped output by 'province'. You can override using the  
`.groups` argument.

```
# Create a summary for the full sample across provinces
full_sample <- survey %>%
  filter(gender %in% c("Male", "Female")) %>%
  group_by(gender) %>%
  summarise(count = n()) %>%
  mutate(total = sum(count),
         percentage = round((count / total) * 100),
         province = "Full Sample")

# Combine the province sample and full sample data into a dataset
demographics <- bind_rows(province_sample, full_sample)

# View the results
View(demographics) # 1583 values due to 5 excluded participants
```

```
# Set the desired order of 'province' levels for plotting
demographics$province <- factor(demographics$province,
                               levels = c("Saskatchewan",
                                           "Quebec",
                                           "PEI",
                                           "Ontario",
                                           "Nova Scotia",
                                           "Newfoundland",
                                           "New Brunswick",
                                           "Manitoba",
                                           "BC and Territories",
                                           "Alberta",
                                           "Full Sample"))

# Convert 'gender' to a factor
demographics$gender <- as.factor(demographics$gender)
```

```

# Plotting gender distribution across Canadian provinces and territories
ggplot(demographics, aes(province, count, fill = gender)) +

# Create bar plot for each province to show proportions
geom_bar(stat = "identity", position = "fill", color = "white", width = 0.8) +

# Add percentage value inside bars
geom_text(aes(label = paste0(percentage, "%")),
          position = position_fill(vjust = 0.5),
          size = 4.5, color = "white", family = "montserrat",
          face = "bold") +

# Add plot title and remove axis titles (set to NULL)
labs(title = "Gender distribution across Canadian provinces and territories",
      x = NULL, y = NULL, fill = NULL) +

# Remove y-axis labels and breaks
scale_y_continuous(labels = NULL, breaks = NULL) +

# Customize x-axis labels
scale_x_discrete(labels = function(x) {
  ifelse(x == "Full Sample", expression(bold("Full Sample")), x)
}) +

# Customize fill colors and order labels
scale_fill_manual(values = c("Male" = "#080757", "Female" = "#D13060"),
                  # Set the order of legend items
                  breaks = c("Male", "Female")) +

# Apply minimal theme
theme_minimal() +

# Flip the coordinates to switch the bars to a horizontal orientation
coord_flip() +

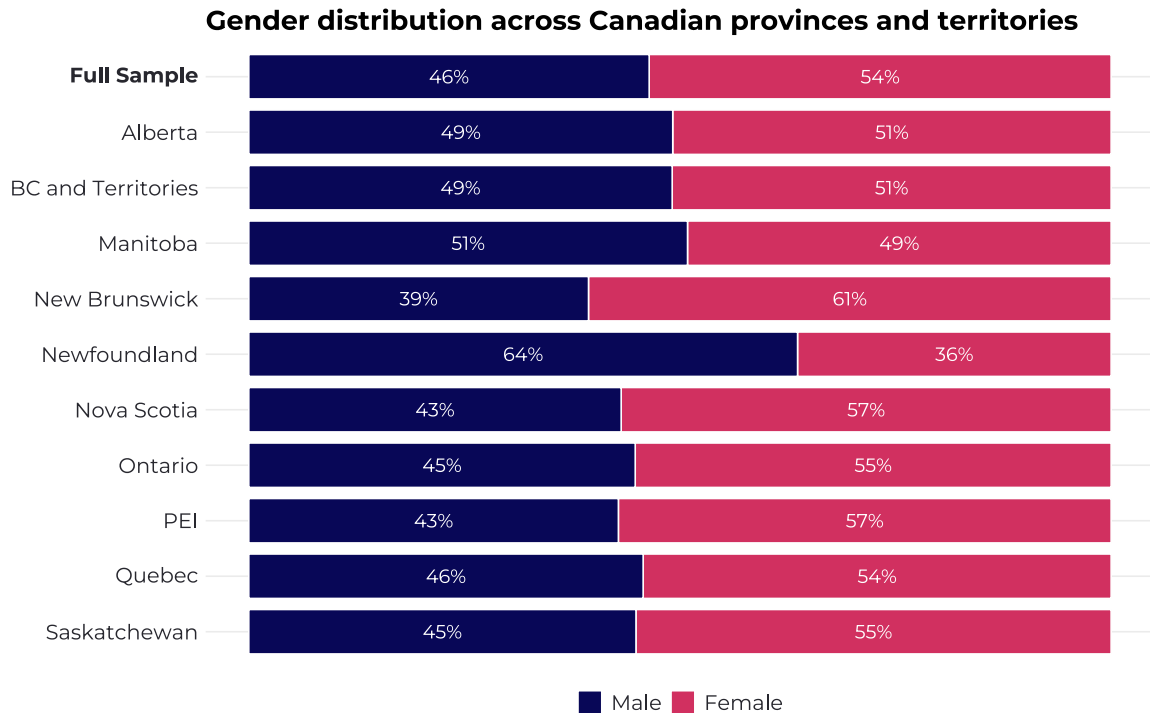
# Additional theme personalizations
theme(
  text = element_text(family = "montserrat"),
  axis.text.y = element_text(size = 14, color = "#27272e"),
  plot.title = element_text(size = 18, color = "black", face = "bold"),
  plot.margin = margin(0.5, 0.5, 0.5, 0.5, "cm"),
  legend.position = "bottom",

```

```

legend.title = element_blank(),
legend.text = element_text(size = 14, color = "#27272e")
)

```



The bar chart displays the proportions of male and female Canadians participating in the survey divided by their province of residence. The graph shows that, in most provinces, the distribution of male and female participants closely mirrors the overall sample distribution (54% female, 46% male). However, New Brunswick and Newfoundland deviate from this pattern: in New Brunswick, 61% of participants are female and 39% are male; in Newfoundland, 64% are male and 36% are female. In my view, these differences likely result from unbalanced sample sizes rather than reflecting an inherent gender imbalance in these provinces. Therefore, to correct these differences, I would simply suggest administering the survey to additional male participants in New Brunswick, and additional female participants in Newfoundland in order to balance the gender distribution.

The bar chart does not include participants who selected “Other” or “I prefer not to say” for the gender question. However, as only 5 participants fell into these categories, I would have also decided to filter the data to include only male and female participants and provide a clearer representation of the gender distribution among Canadians. Moreover, I noticed that participants from Yukon, Nunavut, and the Northwest Territories have been grouped under ‘Territories’ and combined with those from British Columbia.

## WHENREADY

```
# Transform data for the analysis by creating a subset
whenready <- survey %>%
  # Transform from wide to long format to create 'behavior' & 'response' columns
  pivot_longer(cols = starts_with("whenready_"),
               names_to = "behavior",
               values_to = "response") %>%

  # Convert behavior codes to more descriptive labels
  mutate(
    behavior = case_when(
      behavior == "whenready_1" ~ "Domestic Business Flights",
      behavior == "whenready_2" ~ "International Business Flights",
      behavior == "whenready_3" ~ "Domestic Leisure Flights",
      behavior == "whenready_4" ~ "International Leisure Flights",
      behavior == "whenready_5" ~ "In-Person Shopping",
      behavior == "whenready_6" ~ "Return to Workplace",
      behavior == "whenready_7" ~ "Eat at Restaurants",
      behavior == "whenready_8" ~ "Go to Bars or Clubs",
      behavior == "whenready_9" ~ "Public Events",
      behavior == "whenready_10" ~ "Public Transport"
    ),

    # Clean and shorten responses
    response = str_replace_all(response, "\\t+", ""),
    response = str_trim(response),
    response = case_when(
      response == "I am already ready or doing it" ~ "Already doing it",
      response == "In 1-3 months" ~ "In 1-3 months",
      response == "In 3-6 months" ~ "In 3-6 months",
      response == "In more than 6 months" ~ "In 6+ months",
      response == "When a vaccine will be ready or the virus will have disappeared"
        ~ "After vaccine or virus gone",
      response == "I don't foresee a time when I will be ready to do it again"
        ~ "No foreseeable time",
      response == "Not applicable for me" ~ "Not Applicable"
    ),

    # Calculate age, create 'age' column and create 'age group' column
    age = as.numeric(format(Sys.Date(), "%Y")) - as.numeric(birthyear),
```

```

age_group = case_when(
  age >= 18 & age <= 34 ~ "18 - 34 years",
  age >= 35 & age <= 49 ~ "35 - 49 years",
  age >= 50 & age <= 64 ~ "50 - 64 years",
  age >= 65 ~ "65+ years",
  TRUE ~ NA_character_
) %>%

# Move columns in the desired sequence
select(birthyear, age, age_group, behavior, response)

# View the data to verify the results
View(whenready)

# Count the number of missing values in the 'age_group' column
missing_age_group <- whenready %>%
  summarise(missing_age_group = sum(is.na(age_group)))

#Count the number of missing values in the 'response' column
missing_response <- whenready %>%
  summarise(missing_response = sum(is.na(response)))

# Print results
missing_age_group # No missing values
missing_response # 10 missing values due to a dormant participant

# Transform data for analysis
whenready_general <- whenready %>%
  # Filter out rows with missing 'response' values
  filter(!is.na(response)) %>%
  # Keep rows with meaningful 'response'
  filter(response %in% c("Already doing it",
                        "In 1-3 months",
                        "In 3-6 months",
                        "In 6+ months",
                        "After vaccine or virus gone",
                        "No foreseeable time")) %>%

# Group by 'behavior' and 'response' to count occurrences
group_by(behavior, response) %>%
  summarise(count = n()) %>%

```

```
# Calculate total responses and percentages for each 'behavior'
group_by(behavior) %>%
mutate(total = sum(count),
       percentage = round((count / total) * 100, 2)) %>%

# Remove grouping to return to a standard data frame
ungroup()
```

`summarise()` has grouped output by 'behavior'. You can override using the `.groups` argument.

```
# View the data to verify the results
View(whenready_general)
```

```
# Recode 'behavior' in the desired order for plotting
whenready_general$behavior <- factor(whenready_general$behavior,
                                   levels = c("International Leisure Flights",
                                              "International Business Flights",
                                              "Domestic Leisure Flights",
                                              "Domestic Business Flights",
                                              "Public Events",
                                              "Go to Bars or Clubs",
                                              "Public Transport",
                                              "Eat at Restaurants",
                                              "Return to Workplace",
                                              "In-Person Shopping"))

# Recode 'response' to reflect the desired order of readiness levels
whenready_general$response <- factor(whenready_general$response,
                                   levels = c("No foreseeable time",
                                              "After vaccine or virus gone",
                                              "In 6+ months",
                                              "In 3-6 months",
                                              "In 1-3 months",
                                              "Already doing it"))
```

```
# Plotting percentages of responses across behaviors
ggplot(whenready_general, aes(behavior, count, fill = response)) +

# Create a barchart for each behavior to show proportions
```



```

geom_bar(stat = "identity", position = "fill", color = "black",
         size = 0.2, width = 0.8) +

# Add percentage labels inside the bars
geom_text(aes(label = paste0(round(percentage), "%")),
         position = position_fill(vjust = 0.5),
         size = 4.5, color = "black", family = "montserrat") +

# Add a title to the plot, and set axis labels to NULL (no labels)
labs(title = "Are Canadians ready to resume their behaviors after COVID-19?",
     x = NULL, y = NULL, fill = NULL) +

# Apply a black and white theme for a clean look
theme_minimal() +

# Flip the coordinates to switch the bars to a horizontal orientation
coord_flip() +

# Remove y-axis labels and breaks
scale_y_continuous(labels = NULL, breaks = NULL) +

# Manually set the colors for each level of the 'response' variable
scale_fill_manual(values = c("Already doing it" = "#45b39d",
                             "In 1-3 months" = "#76d7c4",
                             "In 3-6 months" = "#a9dfbf",
                             "In 6+ months" = "#e7f7de",
                             "After vaccine or virus gone" = "#f5b7b1",
                             "No foreseeable time" = "#ec7063"),

                  # Set the order of legend items
                  breaks = c("Already doing it",
                             "In 1-3 months",
                             "In 3-6 months",
                             "In 6+ months",
                             "After vaccine or virus gone",
                             "No foreseeable time")) +

# Customize the appearance of the text and other barchart elements
theme(
  text = element_text(family = "montserrat"),
  axis.text.y = element_text(size = 14, color = "#27272e"),
  axis.text.x = element_text(size = 14, color = "#27272e"),

```

```

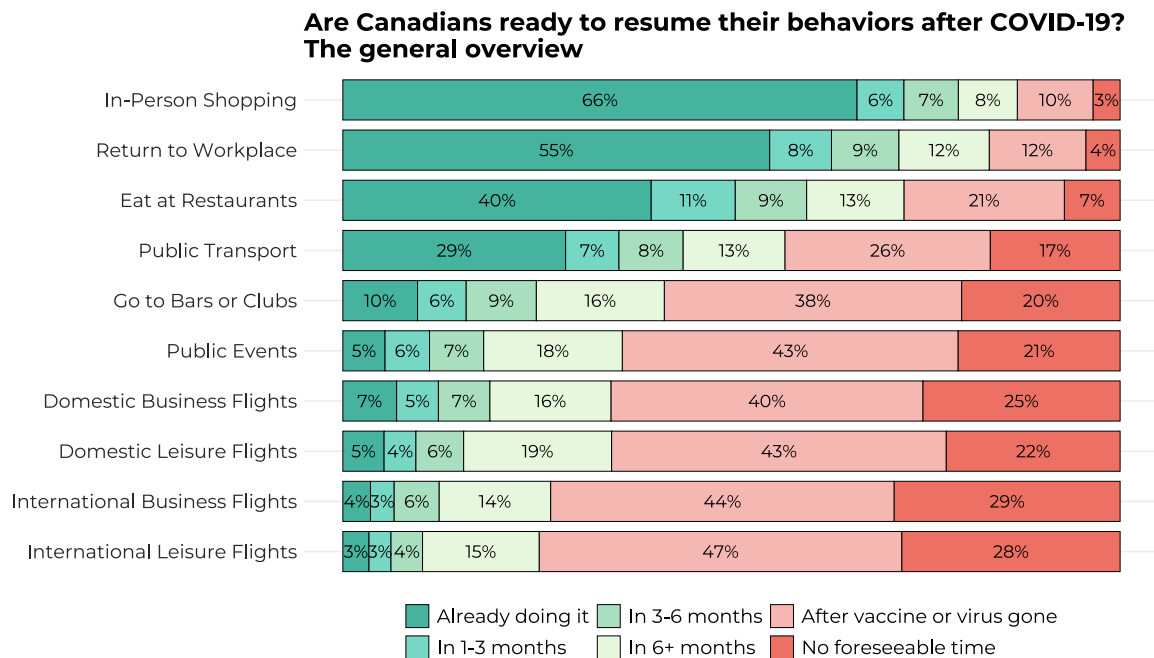
plot.title = element_text(size = 18, color = "black", face = "bold"),
plot.margin = margin(0.5, 0.5, 0.5, 0.5, "cm"),
legend.position = "bottom",
legend.text = element_text(size = 14),
legend.spacing.x = unit(0.5, 'cm'),
legend.spacing.y = unit(1, 'cm')
)+

```

```

# Configure the legend to be displayed in a single row
guides(fill = guide_legend(nrow = 2))

```



The bar chart illustrates the rate of the readiness of Canadians to resume 10 different behaviors after the COVID-19 pandemic, categorized into six different levels: ‘Already doing it’, ‘In 1-3 months’, ‘In 3-6 months’, ‘In 6+ months’, ‘After vaccine or virus gone’, and ‘No foreseeable time’.

The bar chart reveals a clear distinction between the behaviors Canadians are more ready to resume and those they are hesitant about.

- A significant majority of Canadians are ready to **resume in-person shopping** (66% already doing it), **return to their workplaces** (55% already doing it), and **eat at restaurants** (40% already doing it).

- For **public transport**, 29% are already using it, while a combined 28% are planning to resume within the next 6 months.
- **Going to bars or clubs** shows more hesitancy, with only 10% already doing it, and 38% indicating they would only do so in the next 6 months or after the virus is gone, while 20% cannot image the time they will be able to do it again.
- **Public events** like festivals and concerts have similar patterns, with only 5% already doing, a combined 31% planning within 3-6 months, but 43% waiting until safer times, and 21% not expecting to attend in the foreseeable future.
- **Domestic flights** for business or leisure also reflect high caution. Only 7% are currently taking **domestic business flights**, and 5% for **domestic leisure flights**. A substantial 40% and 43% respectively prefer waiting until a vaccine or the virus is gone, and around 25% to 29% are unsure about when they will travel again by plane.
- **International flights** for business or leisure display a similar pattern. Only 4% are currently undertaking **international business flights**, and only 3% for **international leisure flights**. Again, 44% and 47% respectively prefer to wait until the pandemic is over, and 29% feel there's no foreseeable time they'll resume this behavior.

In summary, these insights reflect a cautious approach among Canadians regarding behaviors that involve larger gatherings or travel, compared to more every day-life activities.

```
# Transform and prepare data for a more specific analysis
whenready_specific <- whenready %>%
  # Filter out rows with missing 'response' values
  filter(!is.na(response)) %>%
  # Keep only rows with meaningful 'response'
  filter(response %in% c("Already doing it",
                        "In 1-3 months",
                        "In 3-6 months",
                        "In 6+ months",
                        "After vaccine or virus gone",
                        "No foreseeable time")) %>%

  # Group by 'behavior', 'age_group', and 'response' to count occurrences
  group_by(behavior, age_group, response) %>%
  summarise(count = n()) %>%

  # Calculate total responses and percentages for each 'behavior' divided by 'age_group'
  group_by(age_group, behavior) %>%
  mutate(total = sum(count),
         percentage = round((count / total) * 100, 2)) %>%
```

```
# Remove the grouping to get a standard data frame
ungroup()
```

`summarise()` has grouped output by 'behavior', 'age\_group'. You can override using the `.groups` argument.

```
View(whenready_specific)
```

```
# Recode the 'behavior' variable to specify the desired order for plotting
whenready_specific$behavior <- factor(whenready_specific$behavior,
                                     levels = c("International Leisure Flights",
                                                "International Business Flights",
                                                "Domestic Leisure Flights",
                                                "Domestic Business Flights",
                                                "Public Events",
                                                "Go to Bars or Clubs",
                                                "Public Transport",
                                                "Eat at Restaurants",
                                                "Return to Workplace",
                                                "In-Person Shopping"))

# Recode the 'response' variable to specify the desired order for plotting
whenready_specific$response <- factor(whenready_specific$response,
                                     levels = c("No foreseeable time",
                                                "After vaccine or virus gone",
                                                "In 6+ months",
                                                "In 3-6 months",
                                                "In 1-3 months",
                                                "Already doing it"))

# Recode variable 'age' group in the desired order for plotting
whenready_specific$age_group <- factor(whenready_specific$age_group,
                                     levels = c("18 - 34 years",
                                                "35 - 49 years",
                                                "50 - 64 years",
                                                "65+ years" ))
```

```
# Create a bar chart with ggplot2
ggplot(whenready_specific, aes(behavior, count, fill = response)) +

# Use a fill bar chart with specified border color, size, and width
```

```

geom_bar(stat = "identity", position = "fill", color = "black", size = 0.2, width = 0.6) +

# Add percentage labels inside the bars only if > 10
geom_text(aes(label = ifelse(percentage > 10, paste0(round(percentage), "%"), ""),
  position = position_fill(vjust = 0.5),
  size = 4.5, color = "#27272e", family = "montserrat") +

# Add a title to the plot, and set axis labels to NULL (no labels displayed)
labs(title = "Are Canadians ready to resume their behaviors after COVID-19? \nSome differer
  x = NULL, y = NULL, fill = NULL) +

# Remove y-axis labels and breaks
scale_y_continuous(labels = NULL, breaks = NULL) +

# Flip the coordinates to make the bars horizontal
coord_flip() +

# Manually set the fill colors for each level of the 'response' variable
scale_fill_manual(values = c("Already doing it" = "#45b39d",
  "In 1-3 months" = "#76d7c4",
  "In 3-6 months" = "#a9dfbf",
  "In 6+ months" = "#e7f7de",
  "After vaccine or virus gone" = "#f5b7b1",
  "No foreseeable time" = "#ec7063"),

  # Define the order of legend items
breaks = c("Already doing it",
  "In 1-3 months",
  "In 3-6 months",
  "In 6+ months",
  "After vaccine or virus gone",
  "No foreseeable time")) +

# Apply a black and white theme for a clean visual appearance
theme_bw() +

# Customize the appearance of text, axes, and other plot elements
theme(
  text = element_text(family = "montserrat"),
  axis.text.y = element_text(size = 14, color = "#27272e"),
  axis.text.x = element_text(size = 14, color = "#27272e"),
  plot.title = element_text(size = 18, face = "bold"),
  plot.margin = margin(1, 0.5, 0.5, 0.5, "cm"),

```

```

    legend.position = "bottom",
    legend.text = element_text(size = 14),
    legend.spacing.x = unit(0.1, 'cm'),
    strip.text = element_text(size = 14),
    panel.spacing = unit(0.75, "cm")
) +

# Configure the legend to display in a single row
guides(fill = guide_legend(nrow = 2)) +

# Use facet_wrap to create separate panels for each age_group
facet_wrap(~ age_group, ncol = 2, scales = "free_x")

```

### Are Canadians ready to resume their behaviors after COVID-19? Some differences across age groups



The bar chart illustrates the readiness of Canadians to resume 10 different behaviors post-COVID-19 divided across four age groups: 18-34 years, 35-49 years, 50-64 years, and 65+ years. Here are resumed some key points.

- **In-Person Shopping:** Readiness is consistently high across all age groups, with 64% to 67% already comfortable returning to in-person shopping. This suggests that shopping is overall viewed as a relatively safe activity.
- **Return to Workplace:** Middle-aged adults (50-64 years) show the highest readiness to return to the workplace, with 62% already doing so or planning to do so soon. Other age

groups show slightly more caution, with readiness ranging from 51% to 54%. This may reflect the potential for remote work among younger adults or health concerns among older ones.

- **Eat at Restaurants:** Both younger and middle-aged adults (18-64 years) rated to be ready to eat at restaurants, with 41% to 43% already doing so. In contrast, 36% of adults aged 65+ rated to be ready to resume this activity.
- **Public Transport:** Public transport usage is more common among younger age groups, with about 33% of 18-34 year-old, 31% of 34-49 year-old and 31% of 49-64 year-old using it. However, this drops to 21% for those aged 65+, indicating concerns over virus exposure in these settings.
- **Go to Bars or Clubs:** Younger adults (18-34 years) show the highest inclination, with 13% already going to bars or clubs and a combined 24% planning to do so in the next 3-6 months. In contrast, the other age groups are more cautious about resuming this activity.
- **Public Events:** There is a common caution towards attending public events, such as concerts or festivals, across all age groups, reflecting shared concerns over large and outdoor gatherings. However, younger adults (18-34 years) and middle-aged adults (35-49 years) have a combined proportion of “After vaccine or virus gone” and “No foreseeable time” at 56% and 59%, respectively. This is lower compared to the older groups (50-64 years and 65+ years), which have higher combined proportions of 66% and 69%. This may indicate that younger and middle-aged adults are slightly more optimistic about resuming this activity compared to older adults.
- **Domestic Travel:** A cautious approach towards domestic travel is evident across all age groups. However, younger adults (18-34 years) and middle-aged adults (35-49 years) show a combined proportion of “After vaccine or virus gone” and “No foreseeable time” at 54% and 62%, respectively. In contrast, older groups (50-64 years and 65+ years) exhibit higher combined proportions of 69% and 73%, respectively. This indicates that younger and middle-aged adults are again more optimistic about resuming domestic travel sooner, whereas older adults are more hesitant.
- **International Travel:** There is significant caution towards resuming international travel across all age groups. Younger adults (18-34 years) and middle-aged adults (35-49 years) have a combined proportion of “After vaccine or virus gone” and “No foreseeable time” at 62% and 72%, respectively. In contrast, older adults (50-64 years and 65+ years) show even higher combined proportions of 77% and 80%, respectively. These numbers indicate that while there is a general hesitation towards international travel, younger and middle-aged adults are somewhat more willing to consider it sooner compared to older adults, who prefer to wait until conditions are much safer.

In summary, while all age groups of Canadians are gradually moving towards resuming normal activities, younger and middle-aged adults generally feel slightly more ready to engage in



activities involving social interaction, travel, and public spaces. In contrast, older adults tend to be more cautious, preferring to prioritize health and safety.