Employment Trends Analysis in Canada

2015-2019 vs. 2020-2024

Ya-Nuo, Hsu and Kevin Johnson

December 18, 2024

Introduction

Employment trends are a key indicator of economic performance and recovery, particularly during periods of disruption. In this report, we analyze how employment in Canada has evolved over two distinct periods: the pre-pandemic years (2015–2019) and the post-pandemic years (2020–2024). The goal is to identify changes in employment trends across industries and regions, using data spanning from 2015 to 2024. In this study, we will be researching the questions: How were employment trends in the major industries in Canada affected by the COVID-19 pandemic? We will look at different factors of this data and compare employment levels across different industries and geographical regions. We will group these industries into two different groups: Goods producing industries and service producing industries. The goods producing industries are made up of Construction, Manufacturing, and Mining, and the service producing industries are made up of Accommodation and food services, Healthcare, and Transportation and warehousing.

Provenance

We found our data on Kaggle, and it is sourced from the Survey of Employment, Payrolls, and Hours. This data was collected to provide information about employment levels and includes variables such as geographical region, industry, and employment level. Each case is represented by the employment of a specific industry in a specific geographic region for a given month and year.

Literature Review

Employment trends have been extensively studied, especially during periods of economic recovery following major events like the COVID-19 pandemic. Studies have identified regional and industry-specific disparities in employment recovery post-pandemic. For instance, service-based industries often experience slower recovery compared to technology-driven sectors. This section will explore relevant literature on employment trends in Canada and similar economies, emphasizing regional and industrial differences.

The COVID-19 Pandemic had a drastic effect on the Canadian economy and labor market, causing changes in employment levels. The article from StatCan (2022) mentions that between January 2020 and May 2020, Canada saw roughly 3.4 million jobs lost (nearly 20% of employment). RBC (2022) highlights the change in employment across different sectors, and how employment seemed to rise post pandemic in low contact industries, such as finance, but fell in high contact industries, such as accommodation and food services. Environics Analytics (2021) goes into detail on how population growth decreased across regions in 2020, as

compared to the growth it experienced in 2018 and 2019. It also mentions an overall decline in mental health amongst Canadians, illustrating the fact that COVID had great impacts in Canada beyond just employment.

Methodology

We use a publicly available dataset covering monthly employment data from 2015 to 2024. The dataset includes fields for regions, industries, and employment numbers. Data were filtered to include only the years of interest (2015–2024) and were classified into two periods: Pre-Pandemic (2015–2019) and Post-Pandemic (2020–2024).

Statistical analysis and visualizations are used to identify trends and significant differences between these periods.

Our main goal is to investigate employment trends across major industry sectors in Canada and how they were affected by COVID-19. Through statistical analysis and data visualization, we will see how employment level was affected by the pandemic, and how different industries rebounded from the pandemic. We will look at factors such as geographic region to compare these trends across different regions in Canada.

Data Exploration

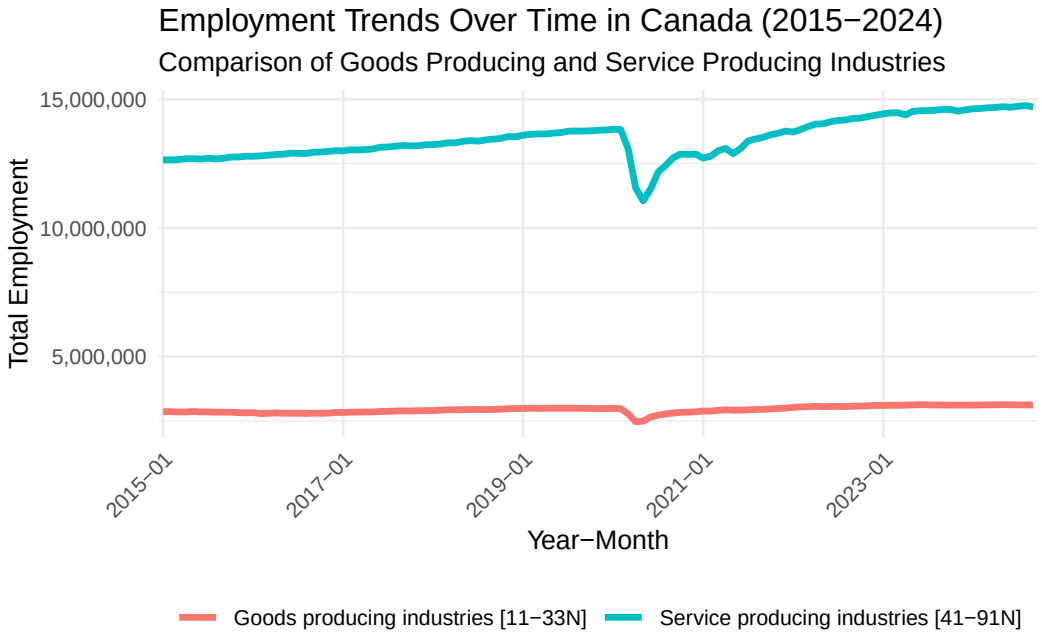
Table 1: Summary Statistics for Employment Trends in Canada

Period	Industry	Avg_Employment	Min_Employment	Max_Employment	SD_Employment
Post-Pandemic	Accommodation and food services [72]	1160107.5	577970	1346479	187254.288
Pre-Pandemic	Accommodation and food services [72]	1292906.5	1218786	1350618	38986.373
Post-Pandemic	Construction [23]	1095392.3	808909	1180205	87422.625
Pre-Pandemic	Construction [23]	1002135.4	963264	1047179	27493.754
Post-Pandemic	Goods producing industries [11-33N]	2994051.2	2468268	3132774	156344.089
Pre-Pandemic	Goods producing industries [11-33N]	2889377.2	2787853	2994154	69699.465
Post-Pandemic	Health care and social assistance [62]	2206168.8	1847387	2402880	127551.909
Pre-Pandemic	Health care and social assistance [62]	1959342.0	1815967	2105036	85755.952
Post-Pandemic	Manufacturing [31-33]	1531177.2	1316947	1573407	55408.631
Pre-Pandemic	Manufacturing [31-33]	1525439.3	1474398	1588677	37120.804
Post-Pandemic	Mining, quarrying, and oil and gas extraction [21]	201460.0	176080	214611	12008.066
Pre-Pandemic	Mining, quarrying, and oil and gas extraction [21]	201091.7	185828	229447	8556.804
Post-Pandemic	Service producing industries [41-91N]	13783696.4	11052028	14754122	915031.588
Pre-Pandemic	Service producing industries [41-91N]	13164625.0	12645274	13803575	370684.225
Post-Pandemic	Transportation and warehousing [48-49]	798430.5	686470	850482	43185.552
Pre-Pandemic	Transportation and warehousing [48-49]	746500.0	717961	795365	23103.875

This five number summary table includes key statistics regarding the employment level across different industries both before and after the COVID-19 pandemic and it provides us with an idea of the employment distribution across. The industry "Industrial aggregate including unclassified businesses" groups all the industries together and provides us with insight to the employment of Canada as a whole. The data from the aggregate grouping of industries allows us to compare individual industries to the total employment in Canada, and it sets a baseline for comparing these industries with each other.

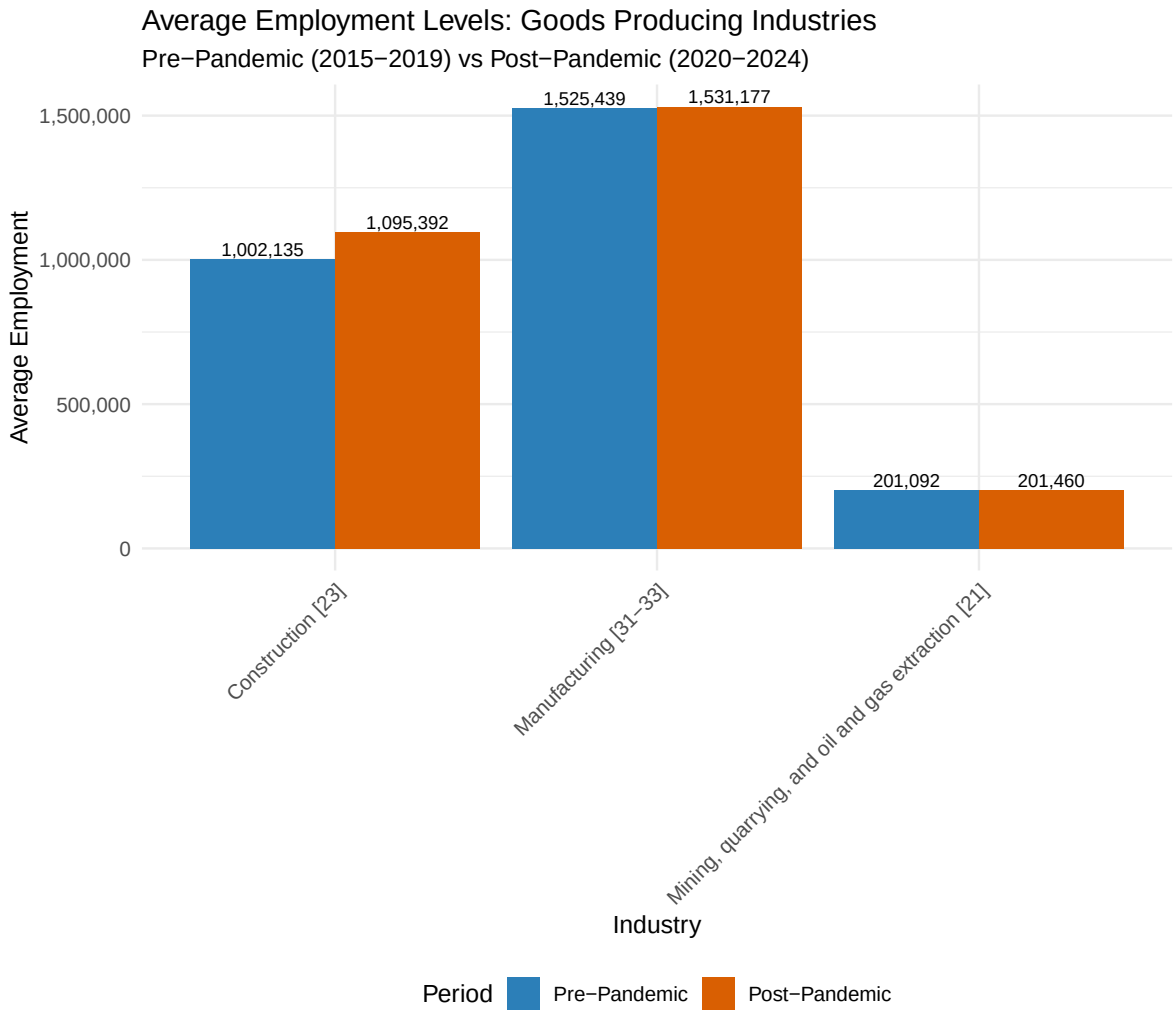
We will also use this summary table to get a basis for the major industries we will be researching: Goods producing industries (Construction, Manufacturing, and Mining) and service producing industries (Accommodation and food services, Healthcare, and Transportation and warehousing).

Visualization: Employment Trends Over Time



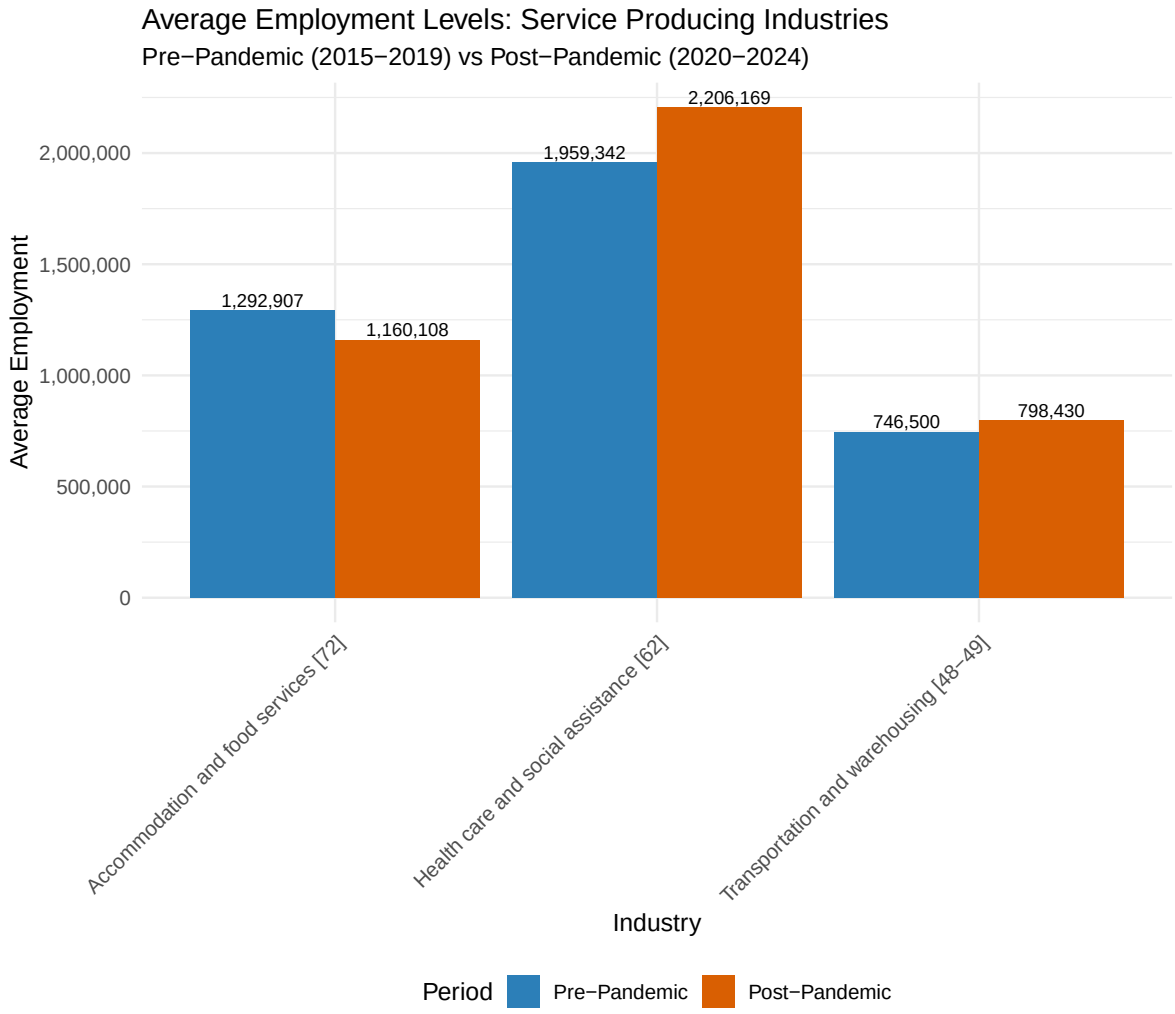
Based on this data visualization, it is apparent that from the beginning 2015 until early 2020, both the goods producing industries and service producing industries experienced stable employment growth. However, both industries experienced a decline in employment between March and May of 2020. To further investigate this detail, we performed some calculations on the data. The service producing industries peak employment of 13,803,575 (December

2019) in the pre-pandemic period experienced a 19.93% decrease, dropping to an employment level of 11,052,028 in May of 2020, its lowest employment level in the post-pandemic period. The goods producing industries saw an employment level of 2,994,154, its highest of the pre-pandemic period, drop to 2,468,268 in April of 2020, indicating a 17.56% decrease in employment. Although they experienced similar decreases in employment, based off the graph it appears that the goods producing industries recovered more steadily than the service producing industries.



This visualization shows us the average employment levels for the three industries we are focusing on in the goods producing industries (construction, manufacturing, and mining) during both the pre-pandemic and post-pandemic periods. From the pre-pandemic to post-pandemic period, construction saw a 9.31% increase in average employment, manufacturing saw a 0.38%

increase in average employment, and mining saw a 0.18% increase in average employment. These values indicate that construction recovered strongly from the dip in employment it experienced during the pandemic, while manufacturing and mining leveled out back to their pre-pandemic levels.



This visualization shows the average employment levels between the pre-pandemic and post-pandemic years for the service producing industries (Accommodation and food services, health care and social assistance, and transportation and warehousing). Accommodation and food services saw a 10.27% decrease in average employment between the two periods, indicating that COVID had a significant impact on this industry. Health care and social assistance and transportation and warehousing saw a 12.6% increase and a 6.96% increase in average employment between the two periods, respectfully. This indicates that these sectors recovered strongly from the effects of the pandemic and are now growing.

Hypothesis Test 1: Does the mean employment level in goods producing industries in Canada between the pre-pandemic and post-pandemic periods have a statistically significant difference?

To determine if the goods producing industries experienced a significant change in average employment, we are going to conduct a hypothesis test. Our hypotheses are

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

where mu represents the average total employment in goods producing industries in Canada during a specific time period (1 = pre-pandemic and 2 = post-pandemic.) After setting our hypotheses, we have to calculate the test statistic using the formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where

$\bar{x}_1 = 2,889,377.2, \bar{x}_2 = 2,994,051.2, s_1 = 69699.465, s_2 = 156344.089, n_1 = 60, n_2 = 57$

$$t = \frac{2,889,377.2 - 2,994,051.2}{\sqrt{\frac{69699.465^2}{60} + \frac{156344.089^2}{57}}} = -4.64$$

We calculate degrees freedom using the formula

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$
$$df = \frac{2.6e + 17}{(1.11e + 14)(3.28e + 15)} = 76.55$$

We get our result of 76.55 for degrees freedom, and then calculate the p-value. We use the absolute value of our test statistic because it is negative, and multiply the p-value by 2 because it is a two tailed test.

```r
2*pt(4.64, 76.55, lower.tail = FALSE)
```

```
[1] 1.413491e-05
```

Because our p-value of about 0.00001 is less than 0.05, we are able to reject the null hypothesis at a 95% significance level. There is convincing evidence that the difference in average employment level for goods producing industries in Canada between the pre-pandemic and post-pandemic periods is statistically significant.

Because our p-value of 7.067453e-06 is less than 0.05, we are able to reject the null hypothesis at a 95% significance level. There is convincing evidence that the difference in average employment level for goods producing industries in Canada between the pre-pandemic and post-pandemic periods is statistically significant.

Hypothesis Test 2: Does the mean employment level in service producing industries in Canada between the pre-pandemic and post-pandemic periods have a statistically significant difference?

For our second hypothesis test, we will be testing the mean employment levels between the pre-pandemic and post-pandemic periods for service producing industries. We have the hypotheses

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Where μ is the average total employment in service producing industries in Canada during a specific time period (1 = pre-pandemic and 2 = post-pandemic). To progress with our testing, we have to calculate the test statistic.

$$t = \frac{13,164,625 - 13,783,696}{\sqrt{\frac{370684.2^2}{60} + \frac{915031.6^2}{57}}} = -4.75$$

Next, we calculate degrees freedom, and we get

$$df = \frac{2.883e + 20}{(8.86e + 16)(3.86e + 19)} = 73.13$$

After obtaining our degrees freedom of 113.85, we calculate our p-value. We use the absolute value of our test statistic because it was negative, and multiply the p-value by 2 because it is a two tailed test.

```r
2*pt(4.75, 73.13, lower.tail = FALSE)
```

```
[1] 9.883686e-06
```

Because our p-value of about 0.000009 is less than 0.05, we reject the null hypothesis at the 95% significance level. There is convincing evidence to determine that the difference in mean total employment levels in service producing industries in Canada between the pre-pandemic and post-pandemic periods is statistically significant.

Conclusion

Our analysis revealed that in both goods and service producing industries in Canada, the difference in average total employment between the pre-pandemic and post-pandemic periods was statistically significant. This finding suggests that the COVID-19 pandemic likely had a large impact on the shift in employment in industry sectors such as manufacturing, construction, mining, accommodation and food service, health care, and transportation and warehousing. For the goods producing industries, shifts in employment were likely due to factors such as disruption in production and labor availability. In the service producing industries, shifts in employment were likely due more to changes in consumer behavior and changes in demand due to protocols such as lock downs and social distancing. Despite the potential factors that caused changes in these industries being different, both goods producing and service producing industries in Canada were negatively impacted by the pandemic. This serves as an emphasis on the lasting affects of the pandemic on the economy.

Code Appendix

```r
# Load required libraries
library(tidyverse)
library(knitr)
library(kableExtra)
library(ggplot2)

employment_trends <- read.csv("~/Desktop/Stat 184/employment_trends.csv")

employment_trends_clean <- employment_trends %>%
  filter(REF_DATE >= "2015-01" & REF_DATE <= "2024-12",
```

```
         UOM != "Dollars") %>%
  mutate(
    REF_DATE = as.character(REF_DATE),
    Employment = replace_na(VALUE, 0),
    Period = ifelse(REF_DATE < "2020-01", "Pre-Pandemic", "Post-Pandemic")
  ) %>%
  select(-c(DGUID, SYMBOL, TERMINATED, SCALAR_FACTOR, SCALAR_ID, STATUS, DECIMALS, Est
  rename(
    Industry = North.American.Industry.Classification.System..NAICS.,
    Region = GEO,
    Date = REF_DATE
  )

write.csv(employment_trends_clean, "employment_trends_clean.csv", row.names = FALSE)

post_pandemic_employment_trends <- employment_trends_clean %>%
  filter(Date >= "2020-01" & Date <= "2024-12")

pre_pandemic_employment_trends <- employment_trends_clean %>%
  filter(Date >= "2015-01" & Date <= "2019-12")

industries_summary <- employment_trends_clean%>%
  filter(
    Region == "Canada",
    Industry %in% c("Goods producing industries [11-33N]",
                    "Service producing industries [41-91N]",
                    "Mining, quarrying, and oil and gas extraction [21]",
                    "Construction [23]",
                    "Manufacturing [31-33]",
                    "Transportation and warehousing [48-49]",
                    "Health care and social assistance [62]",
                    "Accommodation and food services [72]")
  )

# Set global options
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE, fig.align = "cent
summary_stats <- industries_summary %>%
  filter(Region == "Canada") %>%
  group_by(Period, Industry) %>%
  summarise(
```

```
    Avg_Employment = mean(Employment, na.rm = TRUE),
    Min_Employment = min(Employment, na.rm = TRUE),
    Max_Employment = max(Employment, na.rm = TRUE),
    SD_Employment = sd(Employment, na.rm = TRUE)
  ) %>%
  arrange(Industry, Period)

kable(summary_stats, caption = "Summary Statistics for Employment Trends in Canada") %
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE, font_si
  kableExtra::column_spec(2, width = "5cm")
filtered_data_canada <- employment_trends_clean %>%
  filter(
    Region == "Canada",
    Industry %in% c("Goods producing industries [11-33N]",
                    "Service producing industries [41-91N]")
  )
industry_trends_canada <- filtered_data_canada %>%
  group_by(Date, Industry) %>%
  summarise(Total_Employment = sum(Employment, na.rm = TRUE)) %>%
  ungroup()

ggplot(industry_trends_canada, aes(x = Date, y = Total_Employment, color = Industry, g
  geom_line(size = 1.2) +
  labs(
    title = "Employment Trends Over Time in Canada (2015-2024)",
    subtitle = "Comparison of Goods Producing and Service Producing Industries",
    x = "Year-Month",
    y = "Total Employment",
    color = "Industry"
  ) +
  theme_minimal(base_size = 10) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "bottom",
    legend.title = element_blank(),
    legend.text = element_text(size = 8)
  ) +
  scale_x_discrete(
    breaks = c("2015-01", "2017-01", "2019-01", "2021-01", "2023-01")
  ) +
```

```
  scale_y_continuous(labels = scales::comma)
#goods_industries
goods_industries <- c("Manufacturing [31-33]",
                      "Construction [23]",
                      "Mining, quarrying, and oil and gas extraction [21]")
goods_data <- employment_trends_clean %>%
  filter(
    Industry %in% goods_industries,
    Region == "Canada"
  )
goods_summary <- goods_data %>%
  group_by(Industry, Period) %>%
  summarise(Average_Employment = mean(Employment, na.rm = TRUE)) %>%
  ungroup()

goods_summary <- goods_summary %>%
  mutate(Period = factor(Period, levels = c("Pre-Pandemic", "Post-Pandemic")))

ggplot(goods_summary, aes(x = Industry, y = Average_Employment, fill = Period)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  geom_text(aes(label = scales::comma(round(Average_Employment, 0))),
            position = position_dodge(width = 0.9),
            vjust = -0.25, size = 3) +
  labs(
    title = "Average Employment Levels: Goods Producing Industries",
    subtitle = "Pre-Pandemic (2015-2019) vs Post-Pandemic (2020-2024)",
    x = "Industry",
    y = "Average Employment"
  ) +
  theme_minimal(base_size = 12) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "bottom"
  ) +
  scale_fill_manual(values = c("Pre-Pandemic" = "#2C7FB8", "Post-Pandemic" = "#D95F02"
  scale_y_continuous(labels = scales::comma)
#service_industries
service_industries <- c("Health care and social assistance [62]",
                        "Accommodation and food services [72]",
                        "Transportation and warehousing [48-49]")
```

```r
service_data <- employment_trends_clean %>%
  filter(
    Industry %in% service_industries,
    Region == "Canada"
  )

service_summary <- service_data %>%
  group_by(Industry, Period) %>%
  summarise(Average_Employment = mean(Employment, na.rm = TRUE)) %>%
  ungroup()

service_summary <- service_summary %>%
  mutate(Period = factor(Period, levels = c("Pre-Pandemic", "Post-Pandemic")))

ggplot(service_summary, aes(x = Industry, y = Average_Employment, fill = Period)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  geom_text(aes(label = scales::comma(round(Average_Employment, 0))),
            position = position_dodge(width = 0.9),
            vjust = -0.25, size = 3) +
  labs(
    title = "Average Employment Levels: Service Producing Industries",
    subtitle = "Pre-Pandemic (2015-2019) vs Post-Pandemic (2020-2024)",
    x = "Industry",
    y = "Average Employment"
  ) +
  theme_minimal(base_size = 12) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "bottom"
  ) +
  scale_fill_manual(values = c("Pre-Pandemic" = "#2C7FB8", "Post-Pandemic" = "#D95F02"
  scale_y_continuous(labels = scales::comma)
2*pt(4.64, 76.55, lower.tail = FALSE)
2*pt(4.75, 73.13, lower.tail = FALSE)
```