Computer & Mathematical Sciences
UNIVERSITY OF TORONTO
SCARBOROUGH

STAA57 Course Project: Canadian Workforce Analysis

Donglin Que, Yiming Chen

Last compiled on April 07, 2024

Contents

1 Introduction

In recent years, the job landscape in Canada has grown increasingly challenging due to deteriorating economic conditions. Ontario, despite having one of the stronger provincial economies(Davies 2022), is not immune to these challenges. Nearly all sectors are experiencing hiring freezes, affecting industries integral to daily life. This trend is underscored by reports from the Ontario Federation of Labour, which highlight the impact of such freezes on already strained public services(https://www.fao-on.org n.d.). Furthermore, the Financial Accountability Office of Ontario observed a notable slowdown in the labour market in 2023, following two years of significant job growth(mperry 2018).

As international students in Ontario, we want to be deeply invested in understanding the nuances of Canada's employment environment. Drawing from a dataset curated by Immigration, Refugees, and Citizenship Canada covering a 14-year period, our study aims to map out the employment patterns of Canada's immigrant population meticulously("Labour Force Estimates by Immigration - Dataset - Ontario Data Catalogue" n.d.). By employing a comprehensive analytical framework, we aspire to contribute valuable insights into navigating the evolving job market in Ontario for university students.

Regarding the selection of measurement parameters, our data set contains two indicators: employment rate and unemployment rate. The employment-to-population ratio is a useful, general measure that simply shows the number of people currently employed as a proportion of the total working-age population. The unemployment rate measures the proportion of workers in the labour force who are currently not employed but are actively looking for work. It measures the percentage of job seekers in the labour force (employed and unemployed), not the population as a whole("Useful Definitions" n.d.). Although the unemployment rate is affected by the size of the labour force, for students, a group that belongs to the labour force, we choose the unemployment rate as the main indicator of the study here.

2 Objective

This study aims to illuminate the intricate dynamics of Canada's labour market, particularly through the lens of immigration status and its implications on employment outcomes. Our investigation is driven by a series of critical research questions, each designed to unravel different facets of the labour market's complexity:

1. **Educational Attainment and Unemployment**: How does possessing a bachelor's degree compare with having advanced educational qualifications in terms of job-finding ease?
2. **Immigrant Status and Job Market Accessibility**: Does job market accessibility differ significantly between landed immigrants and their non-landed Canadian counterparts?
3. **Gender Influence on Employment Opportunities**: Are there noticeable differences in the hiring ease between genders within the Canadian job market?

4. **Age-related Employment Dynamics**: What are the patterns of unemployment rates across different age groups, and are there any significant variations to note?
5. **Trend Analysis of Unemployment Rates**: What trends can be identified in unemployment rates over the study period, and how do these trends vary across different demographic segments?

By addressing these questions, our study seeks not only to chart the contours of employment variations across different demographic groups but also to contribute to a more nuanced understanding of the factors influencing employment prospects in Canada.

3 Result

3.1 Data Overview

Table 1: Variable Descriptions for Immigration Data Analysis. This table lists the variables used in the dataset, along with a brief description of each.

Variable	Description
Month	Month and year of data collection
Prov	indicate the data is from Canada-wide or Ontario only
Immig	Immigration status including Landed/Non-landed immigrants and Born in Canada. We will compare the different statuses and their unemployment rate
Char	Characteristics of the labour force, including overall population(Labour/Non-labour force), Employment/Unemployment rate and Participation rate. We are mainly focused on unemployment rate
Sex	The sex of the individuals (Male, Female, Both Sexes)
Age	Age groups of the individuals (e.g., 15 years and over, 15-24 years, etc.)
Total..all.education.levels	Total population across all education levels
No.PSE	Population without post-secondary education
X....No.degree..certificate.or.diploma	Population without any degree, certificate, or diploma
High.school.graduate	Population who are high school graduates
X....High.school.graduate..some.post.secondary	Population with some post-secondary education
PSE	Population with post-secondary education
Post.secondary.certificate.or.diploma	Population with post-secondary certificate/diploma
Without.high.school.graduation	Population without high school graduation
With.high.school.graduation	Population with high school graduation
University.degree	Population with a university degree or higher
Bachelor.s.degree	Population with a bachelor's degree
Above.bachelor.s.degree	Population with a degree above a bachelor's

The dataset comprises 18 variables and 424,800 observations spanning from March 1, 2006, to November 30, 2020. It offers a comprehensive foundation for examining employment status and trends across various demographic groups within Ontario(Table 1). Notably, all numerical values are presented as character strings, the population data with a unit of 1000 people, and the ratio is provided in percentage. A conversion to double precision format was done for all the digit analyses.

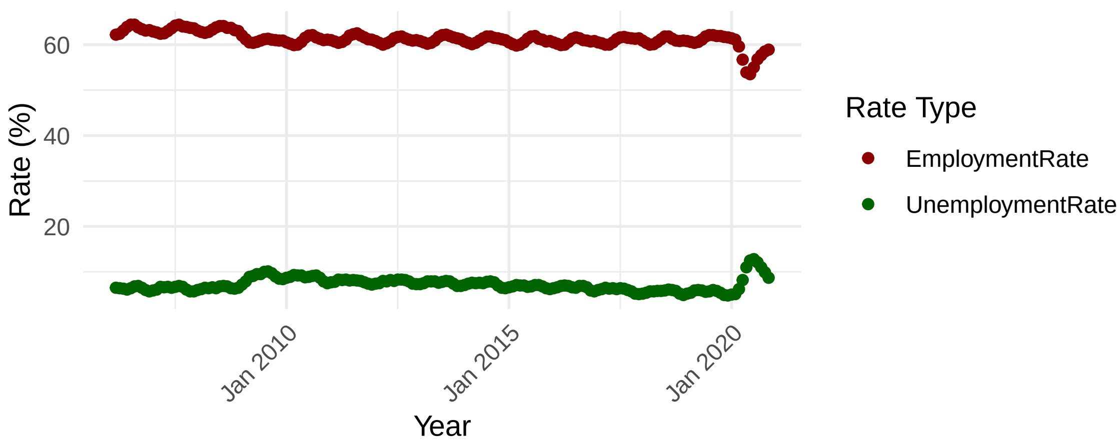


Figure 1: Unemployment and Employment Rates in Ontario Over Time.

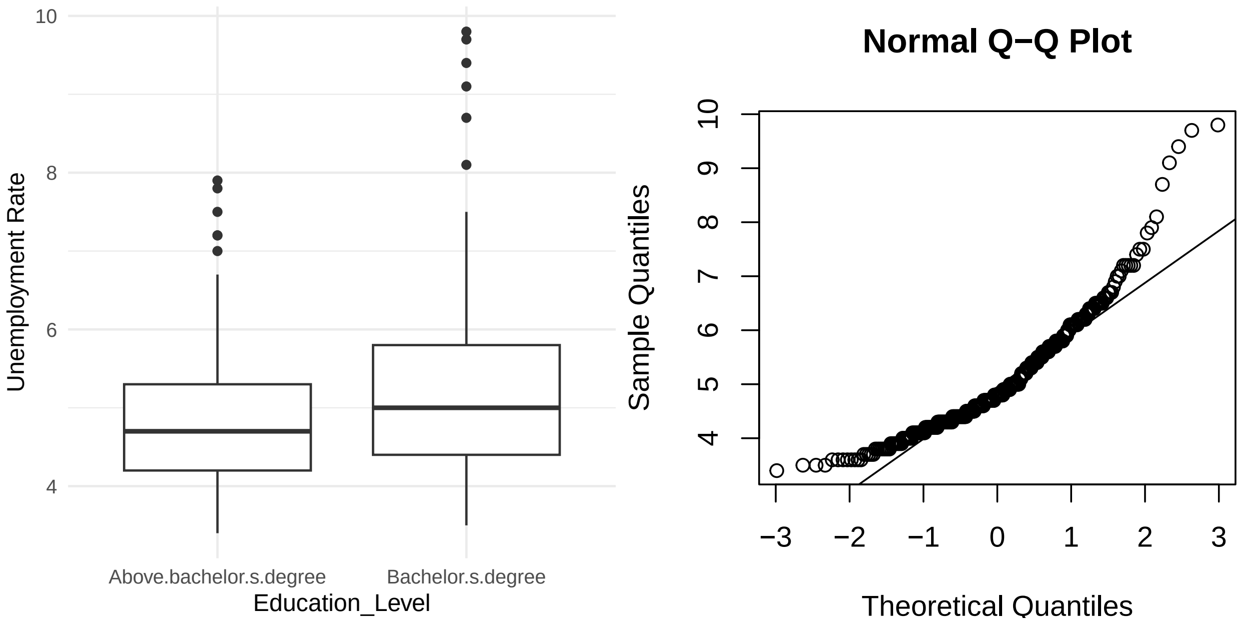
Visualizing employment and unemployment rates, it can be seen from the trends in Fig. 1 that there seems to be some connection between them. In order to ensure the reliability of the research data, here we quickly compare the correlation between the employment rate and the unemployment rate and get true correlation is -0.5238883, which shows that using the unemployment rate in this data set is effective.

3.2 Educational Attainment and Employment

We examine the impact of educational attainment on employment rates in Ontario, with a specific focus on comparing all Ontario residents older than 15 years who hold a Bachelor's degree against those with degrees above a Bachelor's.

We tried to use the test to test the difference in employment rates between different studies, but judging from the data distribution displayed by Boxplot, our sample is likely not to conform to the normal distribution (Fig.2(a)). Through the non-linear distribution of the normal QQ plot (Fig. 2(b)), we conclude that the sample does not have normality, so we abandoned the use of the t-test and switched to the Wilcoxon test. With a p-value smaller enough, we reject the null hypothesis, indicating a statistically significant difference in unemployment rates between individuals with a Bachelor's degree and those with higher educational qualifications in Ontario.

3.3 Immigrant Status and Job Market Accessibility



(a) Distribution of unemployment rate by education level

(b) Q-Q plot checking normality

Figure 2: Relationship of unemployment rate with education level.

Table 2: Variable Descriptions for Immigration Data Analysis. This table lists the variables used in the analysis, along with a brief description of each.

Immig_Status	Mean	SD	N
Born in Canada	4.037853	1.023853	177
Non-landed Immigrants	7.267232	4.548170	177
Total Landed Immigrants	7.073446	1.678787	177

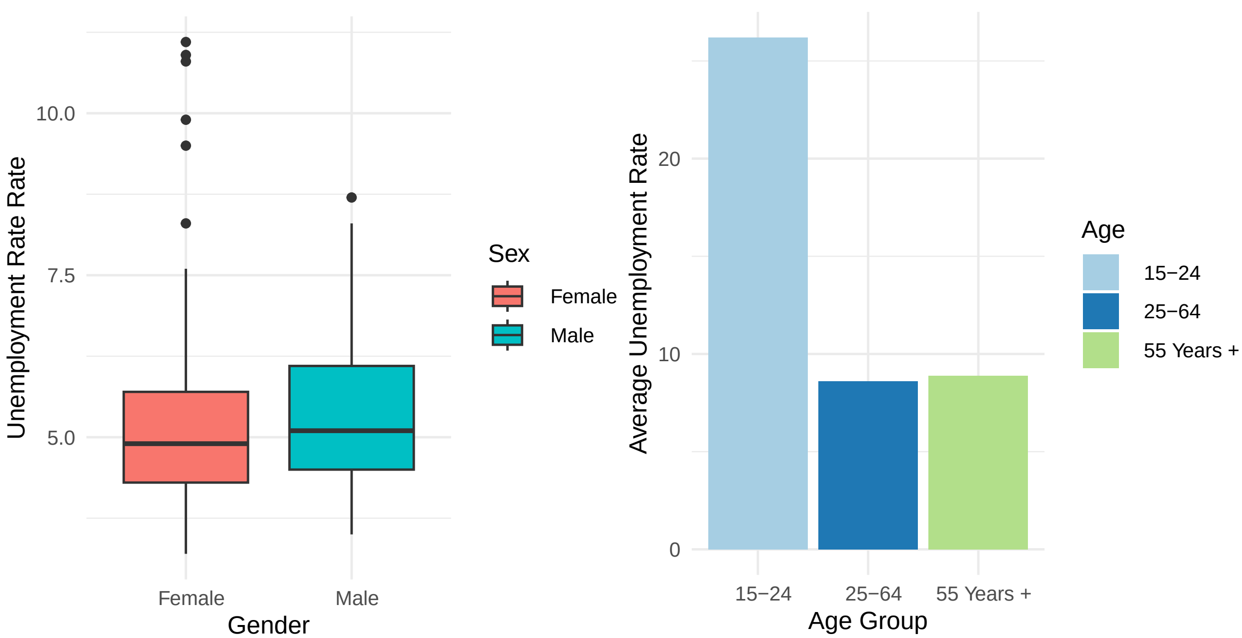
We are curious about the impact of immigration status on employment. The data distribution of different immigration statuses is shown in Table 2. On average, Canadians have the lowest unemployment rate among all bachelor's degree graduates over the age of 15 in Ontario, which has remained stable over the years. Landed immigrants have higher unemployment rates than Canadians, while non-landed immigrants have the highest unemployment rates.

The ANOVA results are statistically significant. The p-value associated with the immigration Status factor is less than the significance threshold of 0.05 indicating a very strong difference in the mean Employment Rates across the different immigrant status groups. For differences between results, we used the Turkey HSD test. The results show:

1. The employment rate of non-arrived immigrants is significantly lower than that of immigrants born in Canada, with a difference of 13.52 percentage points. The 95% confidence interval does not include 0, supporting the significance of the difference.

2. The employment rate of the total number of landed immigrants is significantly lower than that of immigrants born in Canada, with a difference of 7.16 percentage points. The 95% confidence interval range also does not include 0, indicating a significant difference.
3. The employment rate of landed immigrants is significantly higher than that of non-arrived immigrants, with a difference of 6.36 percentage points. The 95% confidence interval range is also shown to indicate a significant difference.

3.4 Gender Influence on Unemployment Rate Opportunities



(a) Unemployment Rate Figures for Males and Females

(b) Unemployment Rate Across Different Age Groups

Figure 3: Gender and age influence on Ontario unemployment rate.

We explore the potential differences in employment opportunities between the sexes in the Ontario job market. Comprehensive labour force data spanning several years on bachelor's degrees are used to determine if there are gender differences in ease of hiring.

It can be seen from the box plots(Fig. 3(a)), there is a modest degree of overlap within the interquartile ranges for both male and female groups, there were small differences in the median values, which are indicated by the horizontal lines centrally located in the middle of each box. This part of the preliminary visualization shows the nuances of gender-specific unemployment in Canada. It suggests that while the unemployment figures for males and females appear to be equal, the differences within each cohort warrant closer examination.

Through hypothesis testing(details shown in the method part), we show that there are no statistically significant differences in employment rates between male and female groups in the Canadian job market.

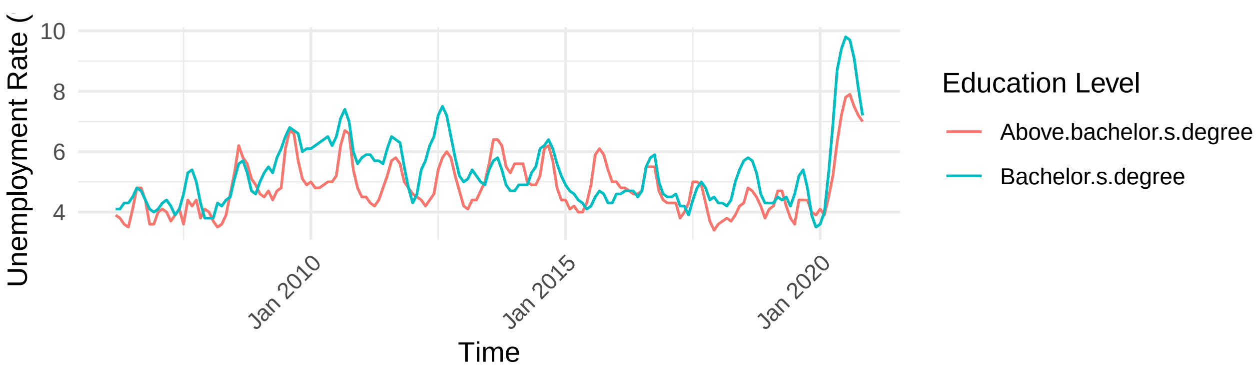
3.5 Age-related Unemployment Rate Dynamics

Exploring changes in employment rates with age is key to understanding labour market conditions. This section analyzes age-related employment dynamics and provides a statistical context for discussing age as a factor in Ontario's employment opportunities.

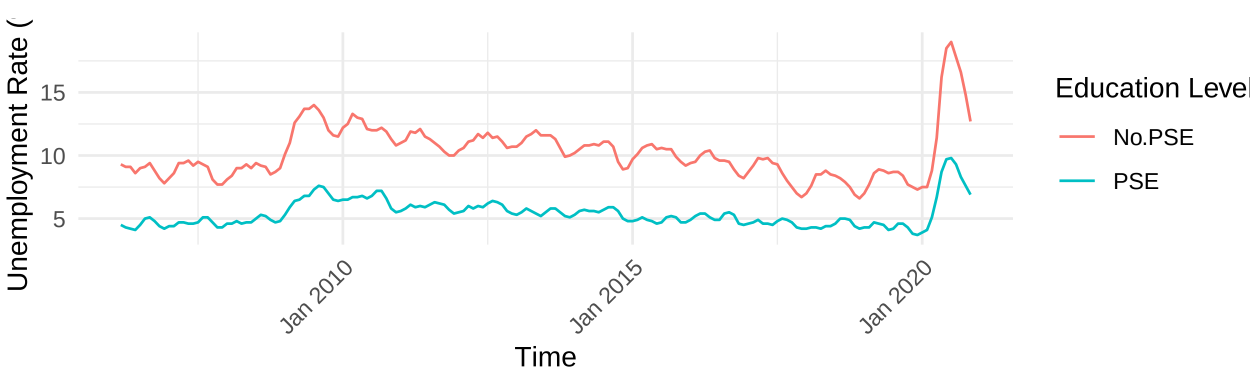
The bar chart illustrates the average unemployment rate for three different age groups, as shown in Fig. 3(b). The 25-64 age group has the lowest average unemployment, which suggests that this age group is performing strongly in the job market, which may be because they are at the peak of their careers during this period. The 15-24 age group has the highest average unemployment rate, which may reflect educational priorities, less work experience, or difficulties in finding jobs. The 55 years and over group shows a middle average unemployment rate maybe due to continuing work after retirement and out of the labour force.

To explore differences between age groups, ANOVA tests provided compelling evidence. It shows that age has a significant effect on the unemployment rate for individuals with a bachelor's degree in Ontario.

3.6 Trend Analysis of Unemployment Rates



(a) Trend of Unemployment Rates by Higher Education Level



(b) Trend of Unemployment Rates by all Education Level

Figure 4: Unemployment rate between different groups

```
## ‘geom_smooth()‘ using formula = ’y ~ x’
```

After completing the testing of multiple variables above, we have a comprehensive understanding of the data. After comparing the unemployment rates with different educational levels in Figure 4, we are curious whether there are some seasonal characteristics of the unemployment rate. We visualize this in Figure 5(a) and we can see that autumn seems to be the season with higher unemployment rates.

Figure 4 also shows other information, such as the sharp rise in unemployment in 2020 due to the COVID-19 pandemic and the fall in unemployment after a brief rise after the 2008 financial crisis. After excluding the 2020 data, we built a simple linear regression model to predict the unemployment rate.

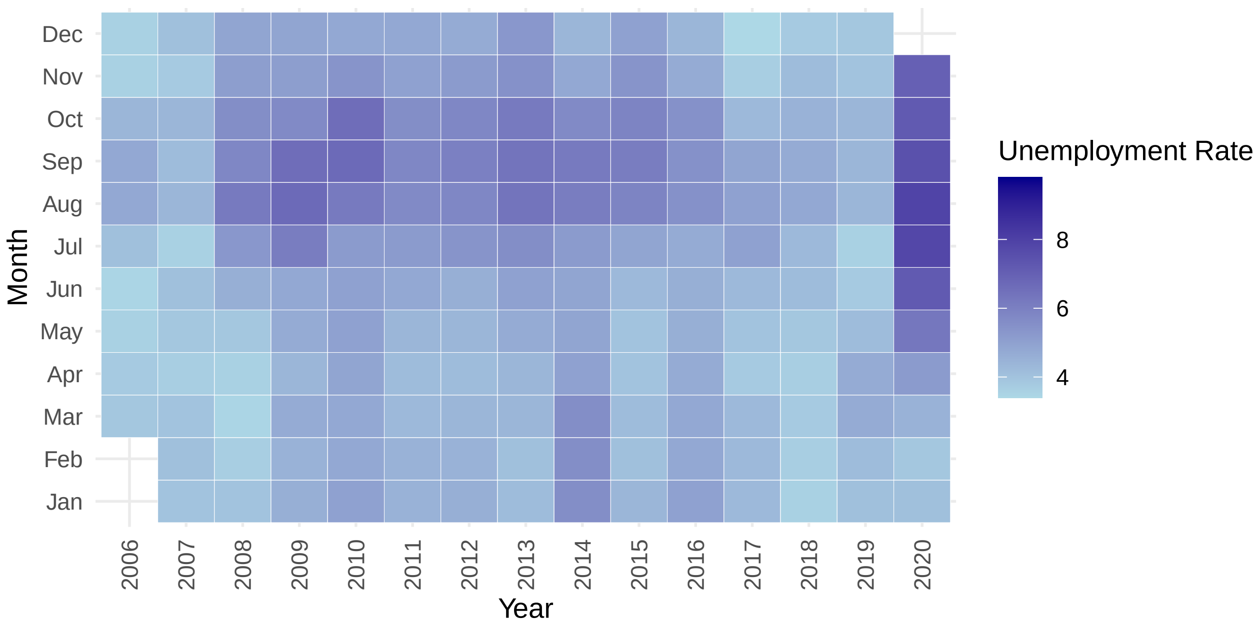
Regression analysis shows that, with the exception of 2020, there is a negative relationship between time and unemployment among individuals with a bachelor's degree in Ontario. The model shows that the unemployment rate continued to decline over time during the study period. For this group, the decrease is approximately 0.03233 percentage points per additional unit of time. However, environmental and other potential factors not included in the model must be considered. This relationship may not be linear throughout or may be affected by external events, economic cycles, or policy changes.

4 Discussion

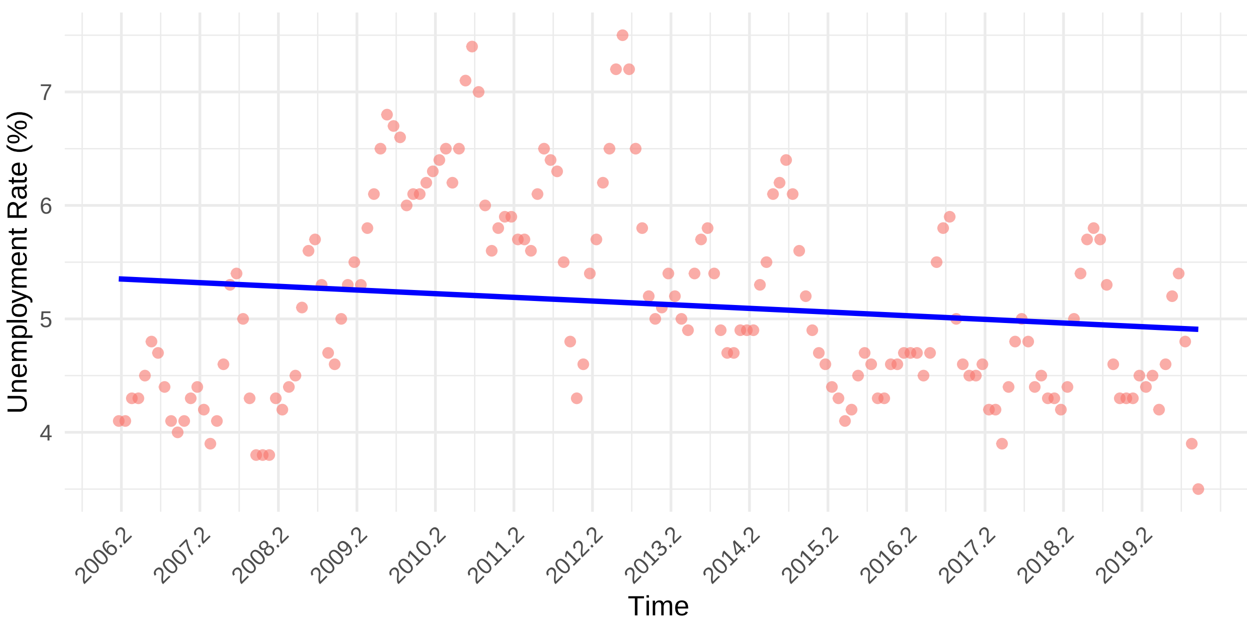
After completing the above analysis, we can see that the unemployment rate is related to various factors. First, for students in Ontario, higher academic qualifications often mean lower unemployment rates. International students (non-landed immigrants) do not have an advantage in the workplace. Gender has little impact on employment status. For a country and province where both population and economy are growing (Fig. 6(a)), generally speaking, the addition of the labour force will also lead to changes in the unemployment rate. We tried to make a fitting line of the population unemployment rate to predict the future unemployment rate (Fig. 6(b)).

According to Statistics Canada projections, Ontario's population will reach 16,020,963 by 2025("Population Projections - Dataset - Ontario Data Catalogue" n.d.). Plugging this data into the model lm, `Unemployment rate` ~ Population, ontario_model_data yields an optimistic estimate that the unemployment rate at that time will be5.811233, which is good news for Ontario residents.

It is not enough to draw conclusions from the model, we ran a bootstrap 1000 times to obtain 95% confidence intervals. The interval is expressed as (3.431796, 8.080975‘). This is a large space because there are far more variables that can determine the unemployment rate than are discussed in this article. We use 10-fold cross-validation commonly used in machine learning, and the RMSE is 1.381779*. Due to the large difference in unemployment rates, a RMSE of 1.38 is considered marginally acceptable. The R-squared is 0.1384134, which means that the model explains approximately 14% of the variance in employment rates.

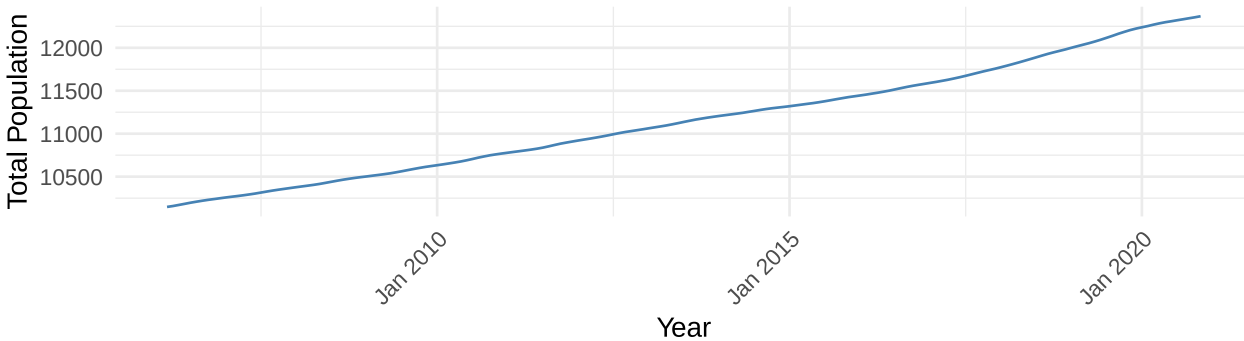


(a) Heatmap of Unemployment Rates by Month and Year

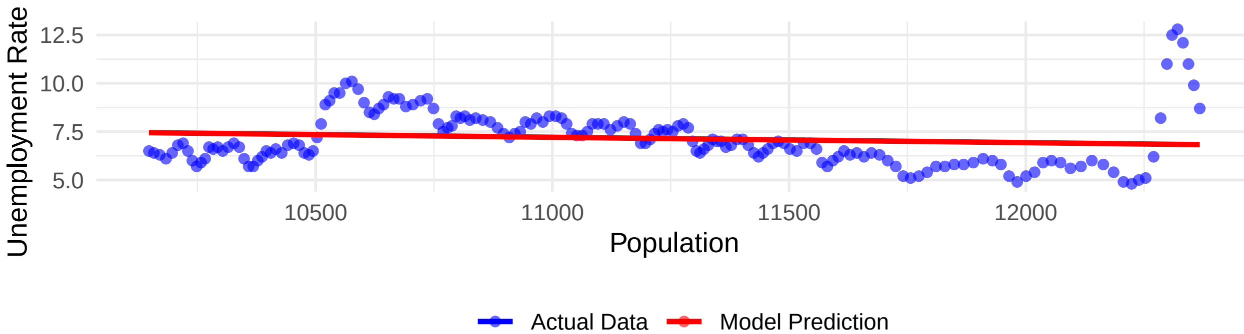


(b) Linear Model of Unemployment Rate Trend for Bachelor Degree Holders

Figure 5: Heat map and linear model of unemployment rate



(a) Ontario Population Over Time



(b) Relationship between Population and Unemployment Rate in Ontario

Figure 6: Unemployment analysis with population

This relatively low number suggests that the model cannot explain much of the variation in employment rates. These are all directions that future research can explore.

5 Method

In 3.1 when exploring employment and the unemployment rate, Pearson's product-moment correlation was used, and the resulting p-value was $7.2815815 \times 10^{-14}$. The null hypothesis was rejected, indicating that the two are statistically correlated.

Jumping to 3.2, the $3.5567207 \times 10^{-14} < \alpha = 0.05$ and the non linearity of QQ plot both indicate we can't use t-test to check the difference of unemployment rate. Wilcoxon rank sum is a non-parametric alternative to the unpaired two-samples t-test, which can be used to compare two independent groups of samples. It's result of p-value is 7.3783984×10^{-5}, where we can tell the job market is more favor to people above bachelor's degree.

ANOVA is a statistical method that separates observed variance data into different components to use for additional tests. In section 3.3, we first use ANOVA to determine whether there are differences between groups with statistically different immigration status. shows that the group differences are significant. The subsequent Tukey's honestly significant difference test results provide pairwise comparisons of employment rates between groups defined by immigration status, with all comparisons statistically significant as shown in Table 3.

Table 3: TukeyHSD result from immigration status analysis.

	diff	lwr	upr	p adj
Non-landed Immigrants-Born in Canada	3.2293785	2.5146424	3.9441147	0.0000000
Total Landed Immigrants-Born in Canada	3.0355932	2.3208570	3.7503294	0.0000000
Total Landed Immigrants-Non-landed Immigrants	-0.1937853	-0.9085215	0.5209509	0.7996373

In 3.4, we choose the Wilcox test instead of applying a two-sample t-test, because although the distributions of the two are similar, there are many outliers in women. The p-value obtained by the Wilcox test is 0.0552988, which is not less than the significance value 0.05, so the null hypothesis cannot be rejected.

Further analysis by ANOVA test was performed after the images of 3.5 were completed. From this ANOVA test output, we can observe that the p-value is <2e-16, which is very small, indicating strong statistical evidence against the null hypothesis that there is no difference in the average unemployment rate between the 3 age groups.

Table 4: Summary of unemployment model.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	70.2018894	33.5277831	2.093842	0.037812
TimeNumeric	-0.0323251	0.0166553	-1.940835	0.053993

A simple model construction of the unemployment rate in 3.6 shows that the intercept is 70.2018894, it represents the model's predicted employment rate at the starting point of the time numeric scale. A -0.0323251 change suggests that for each numeric increment in time, the unemployment rate is predicted to decrease by approximately 0.0323251 percentage points.

6 Conclusion

In our analysis of Canada's labour market, with a particular focus on Ontario, we observed that individuals holding degrees above a bachelor's level generally have an easier time finding employment compared to those with only a bachelor's degree, including international students. Notably, the study revealed negligible differences in employment opportunities between genders, suggesting commendable gender equality in the labour market. By leveraging data spanning a 14-year period, we meticulously charted employment patterns, particularly among Canada's immigrant population, utilizing a robust analytical framework aimed at providing actionable insights for navigating Ontario's evolving job market. Our regression analysis, excluding the anomalous year of 2020 due to COVID-19, unveiled a negative trend between time and unemployment among bachelor's degree holders, indicating a gradual decline in unemployment rates over the study period. This finding, however, comes with the caveat that external factors not included in the model could influence this trend. Further validated by cross-validation techniques, our model offers a glimpse into the future of Ontario's labour market, predicting a continued decrease in unemployment rates. This comprehensive study underscores the multifaceted nature of employment trends and highlights the significance of higher education and immigration status in securing employment in Ontario.

7 Supplementary materials

```r
long_data <- pivot_longer(imm_data,
                          cols = c("Bachelor.s.degree", "Above.bachelor.s.degree"),
                          names_to = "Education_Level", values_to = "Count") %>%
  mutate(Count = as.numeric(Count))

employment_data <- long_data %>%
  filter(Char == "Unemployment rate", Prov == "Ontario", Immig == "Total", Sex == "Both Sexes", Age == "15 Years +") %>%
  group_by(Month, Char, Education_Level) %>% summarise(UnemploymentRate = Count, .groups = 'drop')

ggplot(employment_data, aes(x=Education_Level, y=UnemploymentRate)) + # , fill = Education_Level
  geom_boxplot() + labs(y="Unemployment Rate") + theme_minimal()

qqnorm(employment_data$UnemploymentRate)
qqline(employment_data$UnemploymentRate)
employment_canadian$UnemploymentRate <- as.numeric((employment_canadian$UnemploymentRate))
employment_lan$UnemploymentRate <- as.numeric((employment_lan$UnemploymentRate))
employment_nlan$UnemploymentRate <- as.numeric((employment_nlan$UnemploymentRate))

employment_canadian$Immig_Status <- "Born in Canada"
employment_lan$Immig_Status <- "Total Landed Immigrants"
employment_nlan$Immig_Status <- "Non-landed Immigrants"

combined_data <- bind_rows(employment_canadian, employment_lan, employment_nlan)

combined_dataci <- combined_data %>%
  group_by(Immig_Status) %>%
  summarise(Mean = mean(UnemploymentRate, na.rm = TRUE), SD = sd(UnemploymentRate, na.rm = TRUE), N = n())

kable(combined_dataci,
      caption = "Variable Descriptions for Immigration Data Analysis. This table lists the variables used in the analysis, along with a brief description of each.") %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"),
                position = "center")

anova_result <- aov(UnemploymentRate ~ Immig_Status, data = combined_data)
ano <- summary(anova_result)
tuk <- TukeyHSD(anova_result)
datag <- long_data %>%
  filter(Sex %in% c("  Male", "  Female"), Age == "15 Years +", Prov == "Ontario", Immig == "Total", Char == "Unemployment rate", Education_Level == "Bachelor.s.degree")

ggplot(datag, aes(x = Sex, y = Count, fill = Sex)) +
  geom_boxplot() +
  labs(x = "Gender", y = "Unemployment Rate Rate") +
  theme_minimal()

df_filtered <- imm_data %>%
  filter(Char == "Unemployment rate", Prov == "Ontario", Sex == "Both Sexes", Immig == "Total")

age_employment <- df_filtered %>%
  group_by(Age, Month) %>% mutate(avg_employment = mean(as.numeric(Bachelor.s.degree), na.rm = TRUE)) %>%
  ungroup() %>% mutate(Age = factor(Age, levels = c("  15-24", "  25-64", "  55 Years +"))) %>%
  filter(!is.na(Age))

ggplot(age_employment, aes(x = Age, y = avg_employment, fill = Age)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() + labs(x = "Age Group", y = "Average Unemployment Rate") + scale_fill_brewer(palette = "Paired")
employment_age <- long_data %>%
  filter(Char == "Unemployment rate", Prov == "Ontario", Sex == "Both Sexes",
         Education_Level == "Bachelor.s.degree", Immig == "Total") %>%
  drop_na(Count) %>% group_by(Month, Char, Age) %>%
  summarise(EmploymentRate = mean(as.numeric(Count), na.rm = TRUE), .groups = 'drop')

employment_age$Age <- factor(employment_age$Age, levels = c("  15-24", "  25-64", "  55 Years +"))
anova_result <- aov(EmploymentRate ~ Age, data = employment_age)
ano2 <- summary(anova_result)
employment_data <- long_data %>%
  filter(Char == "Unemployment rate", Prov == "Ontario", Sex == "Both Sexes",
        Immig == "Total", Age == "15 Years +") %>% mutate(Month = as.yearmon(Month),
         EmploymentRate = Count) %>% filter(!is.na(Count))

ggplot(employment_data, aes(x = Month, y = as.numeric(EmploymentRate), color = Education_Level)) +
  geom_line() +  theme_minimal() + labs(x = "Time", y = "Unemployment Rate (%)", color = "Education Level") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

long_data1 <- pivot_longer(imm_data, cols = c("No.PSE", "PSE"), names_to = "if_PSE", values_to = "Num") %>%
  mutate(Num = as.numeric(Num))

employment_data1 <- long_data1 %>%
  filter(Char == "Unemployment rate", Prov == "Ontario", Sex == "Both Sexes", Immig == "Total", Age == "15 Years +") %>%
  mutate(Month = as.yearmon(Month))

ggplot(employment_data1, aes(x = Month, y = Num, color = if_PSE)) +
  geom_line() +  theme_minimal() + labs(x = "Time", y = "Unemployment Rate (%)", color = "Education Level") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

employment_data2 <- long_data %>%
  filter(Char == "Unemployment rate", Prov == "Ontario", Sex == "Both Sexes",
```

```
         Immig == "Total", Age == "15 Years +") %>%
  mutate(
    Year = format(as.yearmon(Month, "%b%Y"), "%Y"),
    Month = format(as.yearmon(Month, "%b%Y"), "%b"),
    EmploymentRate = as.numeric(Count),
    Month = factor(Month, levels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))
  ) %>%
  filter(!is.na(EmploymentRate))

ggplot(employment_data2, aes(x = Year, y = Month, fill = EmploymentRate)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  theme_minimal() +
  theme_minimal() +
  labs(x = "Year", y = "Month", fill = "Unemployment Rate") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

filtered_data2 <- employment_data %>%
   mutate(Year = format(as.yearmon(Month, "%b%Y"), "%Y")) %>%
  filter(Year != "2020", Education_Level == "Bachelor.s.degree")  # Exclude 2020 and focus on Bachelor's degree
filtered_data2$TimeNumeric <- as.numeric(as.yearmon(filtered_data2$Month))

ggplot(filtered_data2, aes(x = TimeNumeric, y = EmploymentRate)) +
  geom_point(aes(color = Education_Level), alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  scale_x_continuous(breaks = round(seq(min(filtered_data2$TimeNumeric), max(filtered_data2$TimeNumeric), by = 1),1)) +
  theme_minimal() + labs(x = "Time", y = "Unemployment Rate (%)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
ontario_data <- imm_data %>%
  filter(Char == "Population", Prov == "Ontario", Sex == "Both Sexes",
         Immig == "Total", Age == "15 Years +") %>%
  mutate(Year = as.yearmon(Month),
    Total_Population = as.numeric(Total..all.education.levels)
  ) %>%
  group_by(Year)

ggplot(ontario_data, aes(x = Year, y = Total_Population)) +
  geom_line(color = "steelblue") +
  theme_minimal() +
  labs(x = "Year",
       y = "Total Population") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

ontario_data1 <- imm_data %>%
  filter(Char == "Population" | Char == "Unemployment rate", Prov == "Ontario", Sex == "Both Sexes",
         Immig == "Total", Age == "15 Years +") %>%
  mutate(Year = as.yearmon(Month),
    Num = as.numeric(Total..all.education.levels)
  ) %>%
  group_by(Year)

ontario_model_data <- ontario_data1 %>%
  mutate(YearOnly = format(as.yearmon(Month, "%b%Y"), "%Y")) %>%
  pivot_wider(id_cols = Year, names_from = Char, values_from = Num)

ggplot(ontario_model_data, aes(x = Population, y = `Unemployment rate`)) +
  geom_point(aes(color = "Actual Data"), alpha = 0.6) +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE, aes(color = "Model Prediction")) +
  scale_color_manual(name = "", values = c("Actual Data" = "blue", "Model Prediction" = "red")) +
  theme_minimal() +
  labs(x = "Population",
       y = "Unemployment Rate (%)") +
  theme(legend.position = "bottom")

model <- lm(`Unemployment rate` ~ Population, data = ontario_model_data)
new_data <- data.frame(Population = 16020963/1000)
predicted_employment_rate <- predict(model, newdata = new_data)

bootstrap_predict <- function(data, indices, newdata) {
  boot_sample <- data[indices, ]
  fit <- lm(`Unemployment rate` ~ Population, data = boot_sample)
    predict(fit, newdata=newdata)
}

new_data <- data.frame(Population = 16020963/1000)
boot_results <- boot(data = ontario_model_data, statistic = bootstrap_predict, R = 1000, newdata = new_data)

conf_interval <- boot.ci(boot_results, type = "perc")$percent[4:5]

train_control <- trainControl(method = "cv", number = 10)  # 10-fold cross-validation
model_cv <- train(`Unemployment rate` ~ Population, data = ontario_model_data,
                  method = "lm", trControl = train_control)
correlation_test <- cor.test(ontario_data_Unemployment$UnemploymentRate, ontario_total_emplyt$EmploymentRate)
cor_name <- correlation_test$method
cor_estimate <- correlation_test$estimate
kable(tuk$Immig_Status,  caption = "TukeyHSD result from immigration status analysis.")
model <- lm(EmploymentRate ~ TimeNumeric, data = filtered_data2)
kable(summary(model)$coef, caption = "Summary of unemployment model.")
```

8 Reference

Davies, Naomi. 2022. "An Investor's Guide to Canada." Investment Monitor. November 15, 2022. https://www.investmentmonitor.ai/investor-tools/an-investors-guide-to-canada-2022/.

"Labour Force Estimates by Immigration - Dataset - Ontario Data Catalogue." n.d. Accessed April 7, 2024. https://data.ontario.ca/dataset/labour-force-estimates-by-immigration.

mperry. 2018. "Ford's Hiring Freeze Will Put Further Strain on Already Underfunded Public Services, Says OFL." The Ontario Federation of Labour. June 18, 2018. https://ofl.ca/fords-hiring-freeze-will-put-further-strain-on-already-underfunded-public-services-says-ofl/.

on.org, https://www.fao-. n.d. "Ontario's Labour Market in 2023." Financial Accountability Office of Ontario (FAO). Accessed April 5, 2024. https://www.fao-on.org/en/Blog/Publications/labour-market-2024.

"Population Projections - Dataset - Ontario Data Catalogue." n.d. Accessed April 8, 2024. https://data.ontario.ca/dataset/population-projections.

"Useful Definitions." n.d. Economic Policy Institute. Accessed April 6, 2024. https://www.epi.org/newsroom/useful_definitions/.