

Labour and Unemployment Rates in Canada:

*A five-decade Analysis using Big Data Tools**

Co-Authors:

Datta, Sumitro
Fahme, Fabio
Ghodosipour, Farnaz
Kaur, Jasneet
Rodriguez, Maria
Tang, Chen

(Group 1)

* This is a group project submitted as part of the requirements for completion of the Data Science 3 (Big Data Management, Systems and Tools) Professional Development course offered by the University of Toronto and the University of Waterloo, 12th of April, 2021.

Labour and Unemployment Rates in Canada

Objectives

A recession is a normal part of a business cycle. It occurs when there are two or more consecutive quarters of negative economic growth, as measured by Gross Domestic Product or other indicators of macroeconomic performance including unemployment. Technically, recession and unemployment are tied to each other.^{1,2} The official dates of recessions include a rise in unemployment as part of the definition of what constitutes a recession.

During a recession many businesses lay-off employees at the same time, and available jobs are scarce. A spike in unemployment and persistence of joblessness is one of the hallmarks of recession. Businesses lay-off workers in the face of losses and potential bankruptcies as a recession spreads, and re-employing those workers is a challenging process that takes time and faces several economic and policy-driven obstacles.²

The extent of a recession in Canada is determined by many factors, depending on which parts of the economy are in decline. For example, the Canadian economy is very sensitive to activity in natural resources such as oil and gas, mining, and lumber.¹

Recessions in Canada have occurred concurrently with that of the United States, showing that the two economies are highly synchronized. Likewise, Canada was not exempt from experiencing global recessions such as:³

The Energy Crisis Recession: (January 1980–July 1980)

The Iran/Energy Crisis Recession: (July 1981–November 1982)

The Gulf War Recession: (July 1990–March 1991)

The 9/11 Recession: (March 2001–November 2001)

The Great Recession: (December 2007–June 2009)

Covid-19 Recession (February 2020–Ongoing)

The 1980s Energy Crisis recession was considered to be the most severe recession between World War II and the current recession. It affected much of the world from 1980 to 1983. The recession was caused by a global disruption in oil supply precipitated by the Iranian Revolution. The shortage of oil led to sharp oil prices and high inflation rates, causing declines in economic output.⁴

The current Covid-19 pandemic caused a spike in unemployment in 2020 and global economic slowdown, leading the World Bank to forecast a recession that would be worse than that of the Energy Crisis.⁵ However, the pattern is atypical and does not follow the cyclical unemployment normally associated with a recession. Most pandemic-associated unemployments are viewed

as temporary, instead of permanent. Marked changes in social interactions necessitate revamping of job descriptions and rematching workers to employers.²

Due to the high public as well as personal impact of recession, it is essential that adequate efforts be performed to understand the factors in play. Conventional technology has been utilized to provide insights on causes and possible recurrence of recessions. However, with the exponential growth of data, in addition to the concept that the current recession presents an atypical pattern, the availability of increased computing power, and the increased demand for responsive applications by the public-- it is worthwhile to utilize big data tools in lieu of conventional techniques.

Big data may be defined as data that is of high volume, possibly unstructured, and processed at a fast pace. Common applications for big data include data mining, predictive analytics, and machine learning.^{6,7} Big data tools have capabilities of processing large data sets in a distributed environment. These tools include MongoDB, Hadoop, Apache Spark, Cassandra, among others.⁸

This project aimed to utilize big data techniques in exploring and providing predictive analysis on unemployment rates in Canada. The research questions were:

1. What is the distribution profile of unemployment rates in Canada for the past four decades?
2. Can we predict the future trend in unemployment rates in Canada and Ontario?
3. How long does it take to recover from a recession?

Data Preparation

The dataset was obtained from [Statistics Canada](#).⁹ It contained the annual labour force frequencies in Canada and the provinces from the time of the major questionnaire redesign in 1976¹⁰ until 2020. There were 18 attributes (including age group, sex, type of employment and industry) and 931,393 observations.

Exploratory analysis was done on the labour force and types of employment for Canada and its regions. Targeted analysis was performed on Unemployment rates.

Filtering using all industries yielded 35,640 values. Of these, 68 were blank, mostly from the earlier years (1970s). These were assumed to be zero, and values were replaced as such.

The platform used was Databricks, and the big data tool of choice was PySpark. Several challenges were met while working with the Databricks Spark Community version:

1. The non-recognition of previously working Databricks accounts, leading to loss of code and analysis. Alternatives like Zeppelin, Amazon Web Services and Google Cloud were explored, but these did not provide learner-friendly platforms, and thus were not adopted.

2. Opening a cluster is slow. It is not efficient in retaining temporary data in memory, and repeated loading is necessary to check the flow of the codes.
3. The platform could utilize both a Spark context and a SQL context in creating tables, however, the tables do not automatically translate from one to the other.
4. Some graphs are not visible when worksheets are published.

Languages used were python, pyspark and sql. The libraries utilized included request, urllib, json, pandas, numpy, matplotlib, seaborn, statsmodels, fbprophet, pyspark.ml and mllib.

Predictive analysis techniques included Time series, FBProphet, Linear Regression, Random Forest Regression and Decision Trees. The Linear Regression model used the 5 most full-time employed industries in the previous year and their unemployment rate in the previous year. To train the Random Forest and Decision Trees models, the labour force & unemployment rate were extracted as additional features. Segment, age group industry and location were also used in the training to predict the unemployment rate for a specific segment.

Group collaboration was facilitated by email exchanges, Google Meets and googledocs. Each group member performed individual work on exploratory and focused analyses. This encouraged hands-on learning, as well as provided a diversity of perspectives. The individual analyses were compared and integrated into one group worksheet.

Definitions from MetaData:⁹

Employed: persons having a job or business

Unemployed: persons without work, are available for work, and are actively seeking work

Labour force = employed + unemployed

Full-time employment: working \geq 30 hours per week

Part-time employment: working $<$ 30 hours per week

Unemployment rate = unemployed / labour force

The Databricks worksheet can be accessed through this link:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/313825514368579/1621046929403477/7328271394610816/latest.html>

Analyses

Data Exploration

Through the 44 years of data, the labour force has a general increasing trend with the occasional slowing down and recovery period (Figure 1). Unlike full-time employment, part-time employment is more resistant to recessions. The first three of the four visible recessions show a slow but firmly increasing part-time employment during recession.

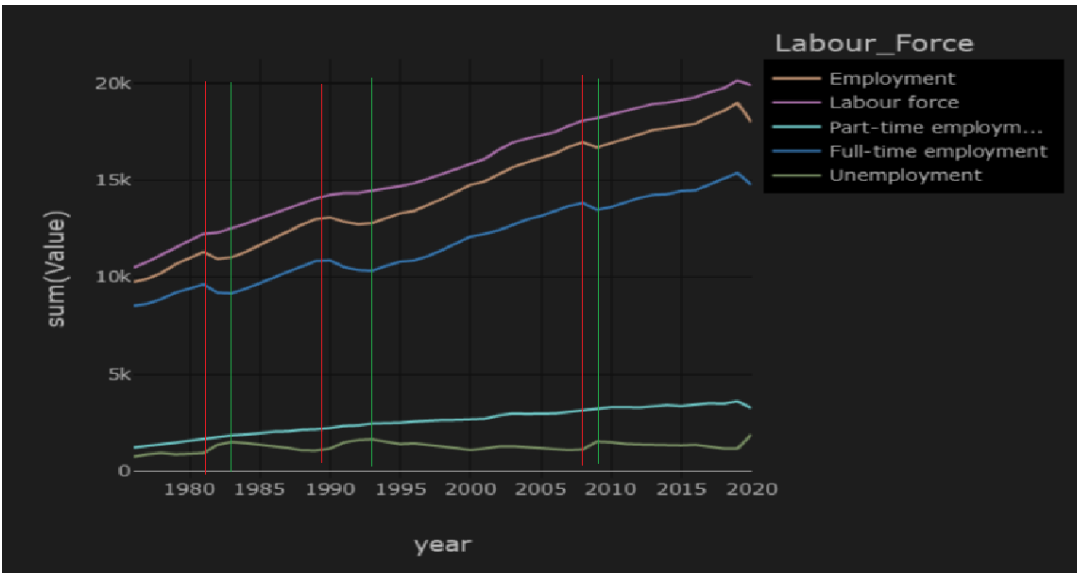


Figure 1. The labour force in Canada from 1976-2020

Age Group

Young adults were more vulnerable against economic disturbances, showing a higher proportion and a higher variability in unemployment rates (Figures 2 and 3).

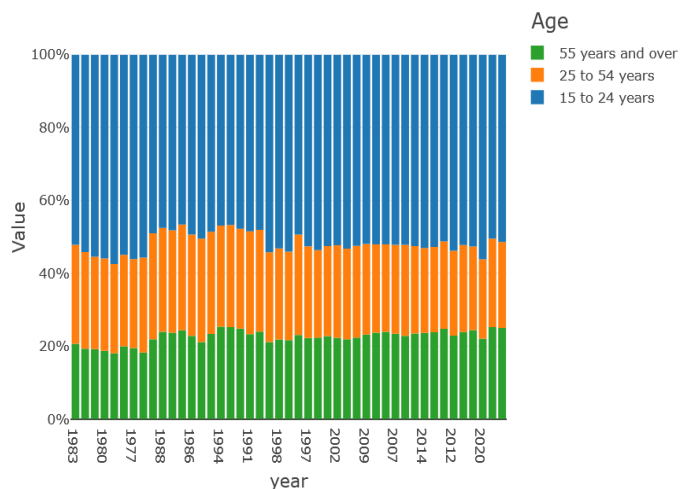
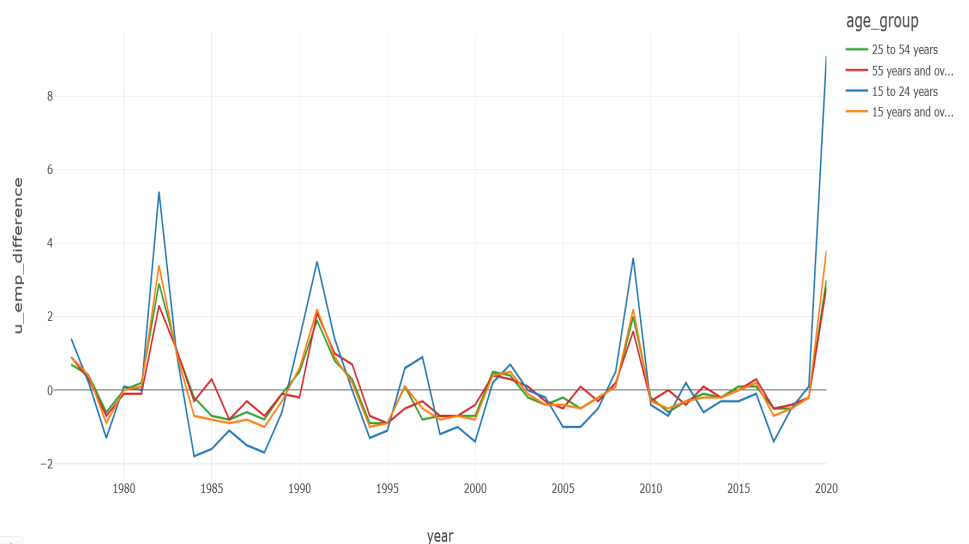


Figure 2. Proportion of age-groups' unemployment rates

Figure 3. Distribution profile of unemployment rates among the different age-groups.



Sex

Females presented with higher unemployment rates during 1970s-1980s, but the proportion shifted to higher unemployment rates among males from 1990s to the present (Figure 4).

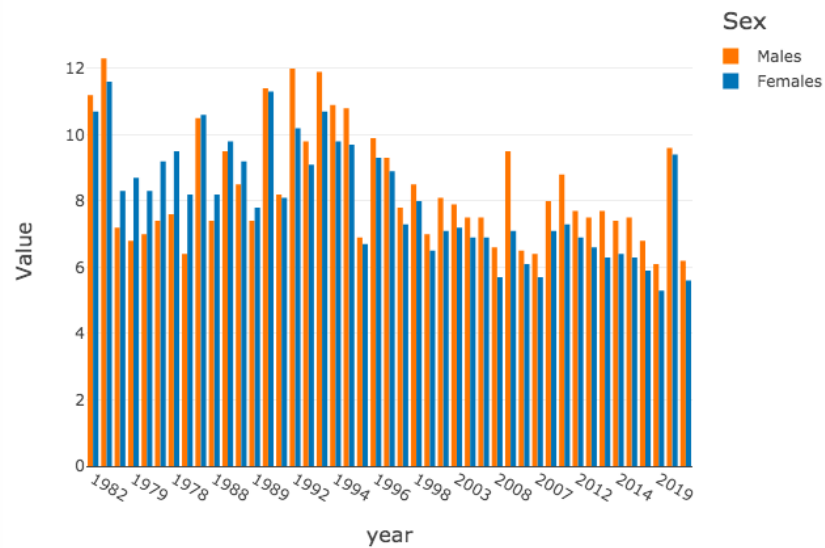


Figure 4. Distribution of unemployment rates between sexes

Location

The historical average of unemployment rates in Newfoundland and Labrador, PEI and New Brunswick were relatively higher than that of the other provinces (Figure 5).

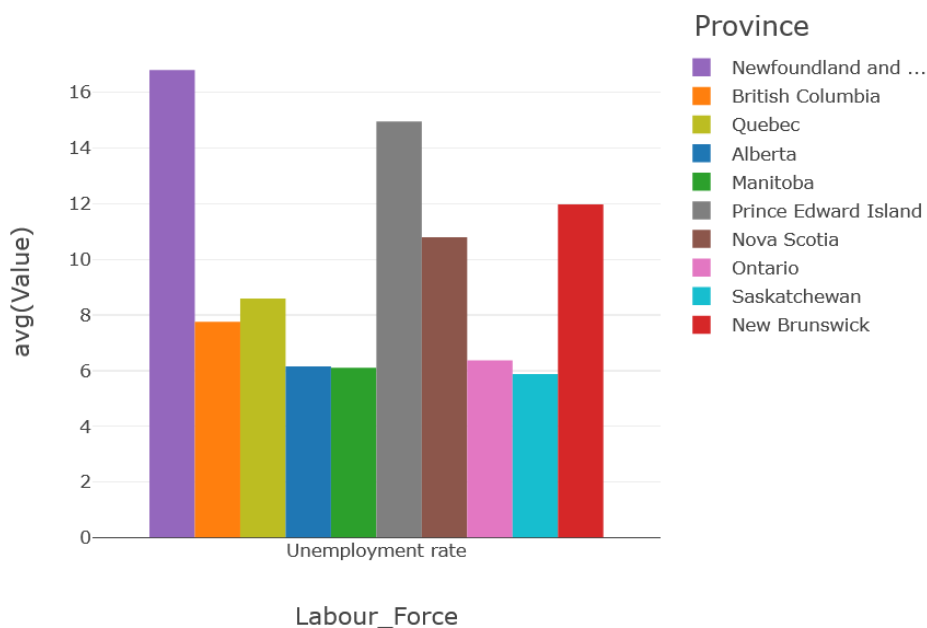


Figure 5. Distribution of unemployment rates between the Canadian provinces.

Ontario and British Columbia were more sensitive to economic disturbances, showing higher changes in unemployment rates (Figure 6).

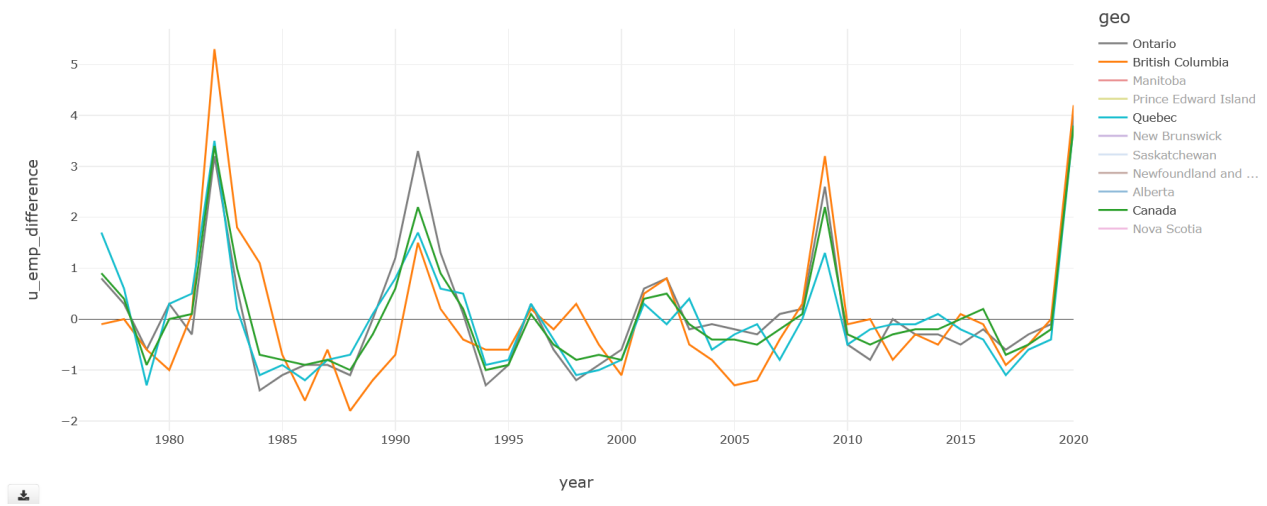


Figure 6. Annual change in unemployment rates between the different Canadian provinces.

Industry

The goods-producing services and the services producing services were the two predominant industries. The unemployment rates of goods-producing industries were more volatile, but the service-producing industry showed a steep increase in unemployment rates during 2020 (Figure 7).

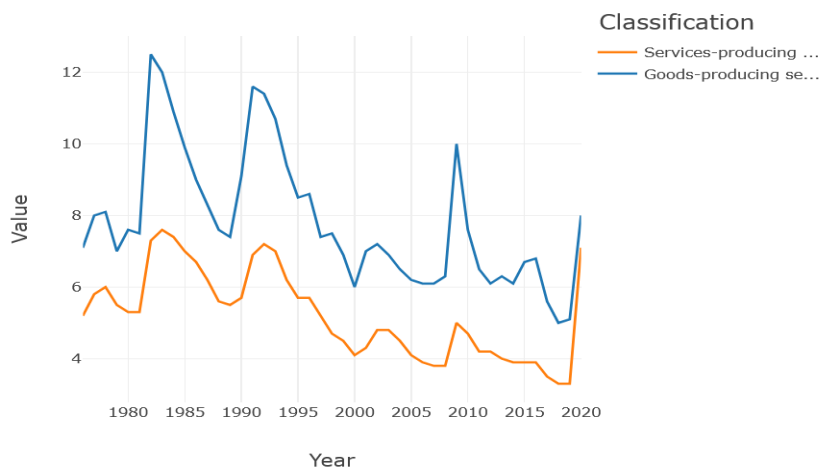


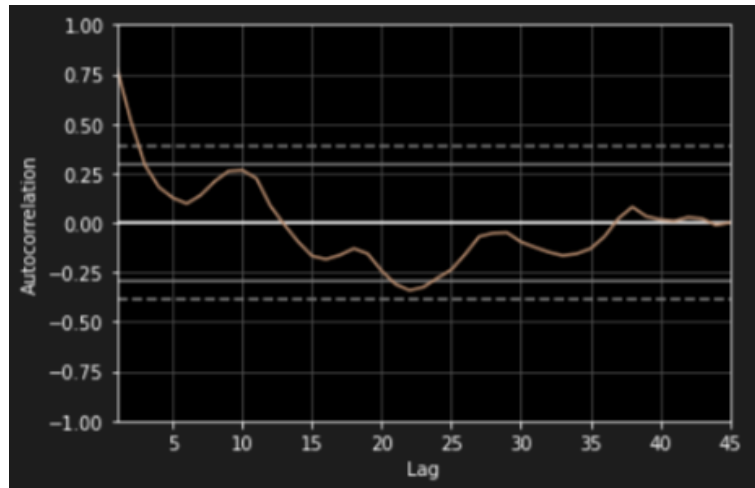
Figure 7. Unemployment rates in the two most dominant industries.

Predictive Analyses

Time Series

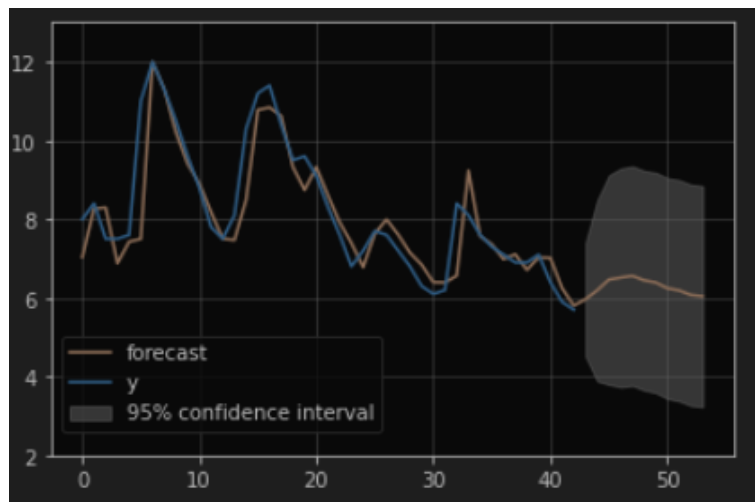
There was significant autocorrelation for unemployment rate until a lag of 2 years (Figure 8).

Figure 8. Autocorrelation plot for unemployment rate



The Autoregressive Integrated Moving Average (ARIMA) model showed that unemployment rates could be a value in the range of 3-9% by 2031 (95% CI) (Figure 9).

Figure 9. ARIMA forecasting for unemployment rate (using order 3,1,2)



FBProphet model

Utilizing the FBProphet library with the linear growth of multiplicative seasonality having a period of 10 and interval of 0.95, the predicted unemployment rate is approximately 6% (range 3-9%) by 2030 (Figure 10).

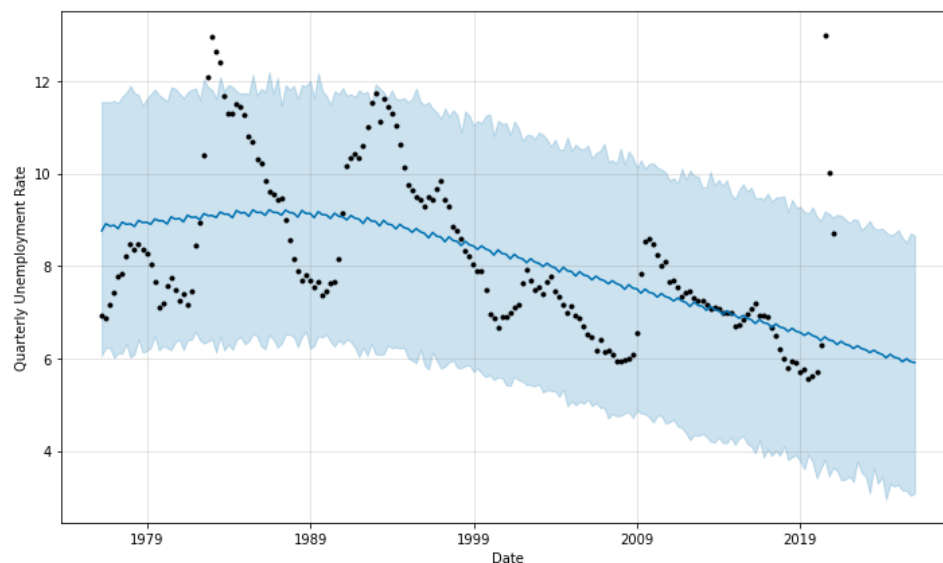


Figure 10. Predicting future quarterly unemployment rates using FBProphet

MLlib: Linear Regression model

The 2021 unemployment rate prediction outcomes were 8.84% for Canada and 8.7% for Ontario.

MLlib: Random Forest Regression & Decision Tree Regression Model:

Allowing for an error of $\pm 1\%$, the success rates for prediction by the random forest regression was 45%, and 62% for the decision tree model.

Analysis on Time to Recovery

The group reached a consensus of defining the start of a recession as an increase in unemployment rate by more than 1% compared to the previous year, and a recovery as a return to the baseline level.

Canada-wide, the 1982 recession lasted until 1987, the 1991 recession lasted until 1997, and the 2009 recession lasted until 2017 (Figure 11). In Ontario, the 1982 recession only lasted until 1985, the 1990 recession lasted until 1999, and the 2009 recession lasted until 2015. The average duration of these 3 Canadian recessions is around 6.33 years, while that of Ontario being 6 years (Figure 12).

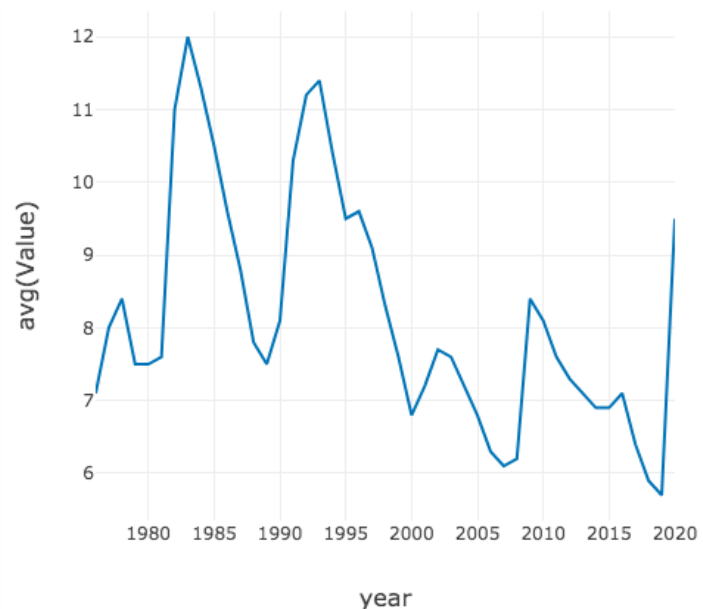


Figure 11. Canadian Annual Unemployment rates 1976-2020

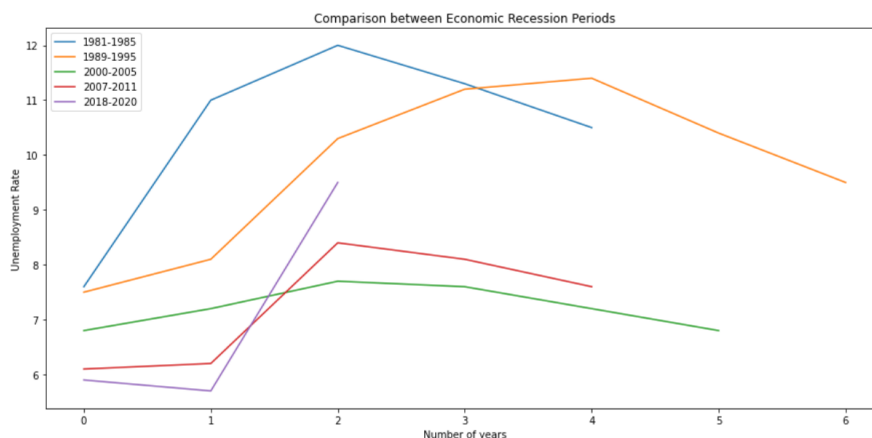


Figure 12. Duration of recession in Canada (years) and the corresponding unemployment rates

Conclusions

Analyzing Canadian labour data has provided us with several insights on the past and present economic direction. The unemployment rate results shown in our analysis correspond with the major economic recessions of the past ~40 years, particularly the 1981, 1990, 2001, 2008 and 2020 recessions. The data shows that unemployment rates have quickly risen to over 8% during the above-mentioned economic recessions, while the overall trend was decreasing.

On average, Canada required 6 years to absorb the aftermath of an economic recession (excluding 2020). There are many factors to take into consideration when trying to elaborate on a country's recovery. This analysis concentrated on the unemployment rate, looked at the long run and marked Canada as recovering from a recession when it has successfully reabsorbed employees and the unemployment rate difference is 5-10% of the value before entering a recession.

Looking at the direction of the unemployment rate curve and where it changed direction will help explain how long it took Canada to exit a recession and enter an expansion period (Figures 11-13):

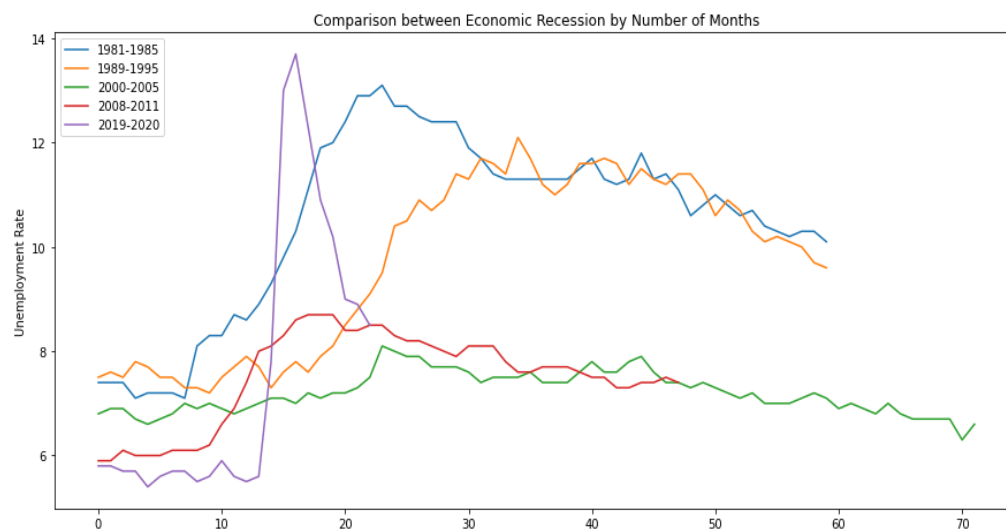
The 1980s recession took Canada 3 years to exit and enter the expansion phase while taking ~8 years to return to previous unemployment levels.

The 1990s recession took Canada 5 years to exit and enter the expansion phase while taking ~9 years to return to previous unemployment levels.

The 2001 recession was relatively shorter and took Canada 2 years to exit while taking ~3 years to return to previous unemployment levels.

The 2008 depression was short, lasting for 1 year, however the effects created took ~9 years for Canada to return to previous unemployment levels.

The 2020 recession is still developing. Currently available data, as shown in Figure 13 shows that the unemployment rate tripled in only 4 months. The recovery was also rapid with a decrease in unemployment by more than 50% in 6 months.



*Figure 13.
Monthly
distribution of
unemployment
rates*

It is difficult to predict with confidence the resolution of the present crisis because the causes are significantly different compared to the previous recessions. The past three depressions of the 80s, 90s and '08 were primarily economic or political in nature, while the current is health crisis-driven.

The 1980 recession is widely considered to have been one of the most severe recessions since modern times, and the primary cause was the 1979 energy crisis. The Iranian Revolution of 1978 caused a disruption to the global oil supply. The rise in oil prices caused an increase in inflation rates, which in turn had governments tighten their monetary policies and therefore pushing up interest rates. This is explained by the figure below.

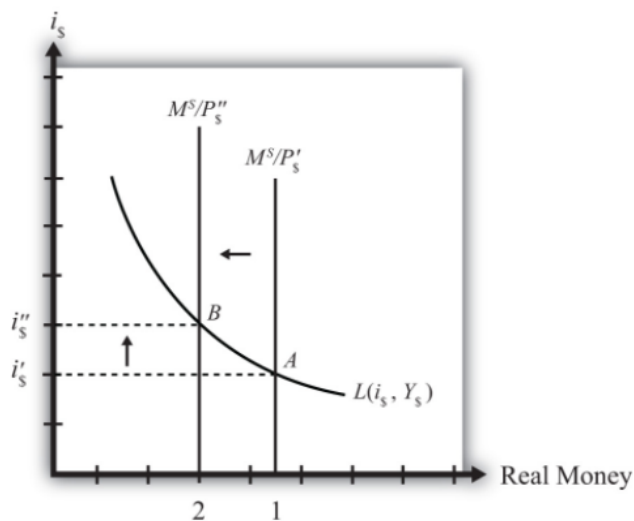


Figure 13. Price increase on monetary policy¹¹

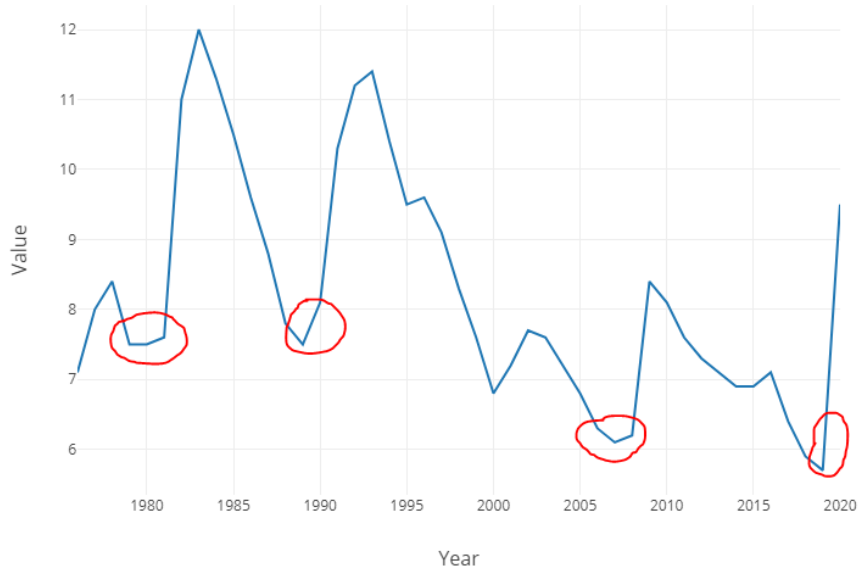
Figure 13 shows the effects of price increase on monetary policy, with interest rates on the y-axis and real money on the x-axis. When prices rise due to inflation and everything else remains the same, this leads to a decrease in the real money supply. Despite the

constant supply of money, the purchasing power of that same amount of money has decreased. The fall in real money supply is shown by a shift along the x-axis to point 2. The demand for money will exceed the supply of money causing an increase in interest rates to reach a new equilibrium in point B.

The 90s recession was mainly caused by the persistence of high levels of inflation that Canada had been experiencing since the 80s. This produced extremely high levels of Consumer Price Index which tripled compared to the levels of the previous decade.

The 2008 great depression was also caused by economic factors. It started with the burst of the US housing bubble which generated a large amount of losses in mortgage-backed securities. A year after, the US economic disturbance started manifesting in the Canadian economy due to their tight trading relationships.

In contrast to these recessions, the 2020 recession was not caused by economic or political activity, but by a global health crisis. The atypical pattern is evident in the results: previous recessions showed a slight increase in the unemployment rates before the sharp spike up, while the spike in unemployment rates in 2020 was immediate (Figure 15).

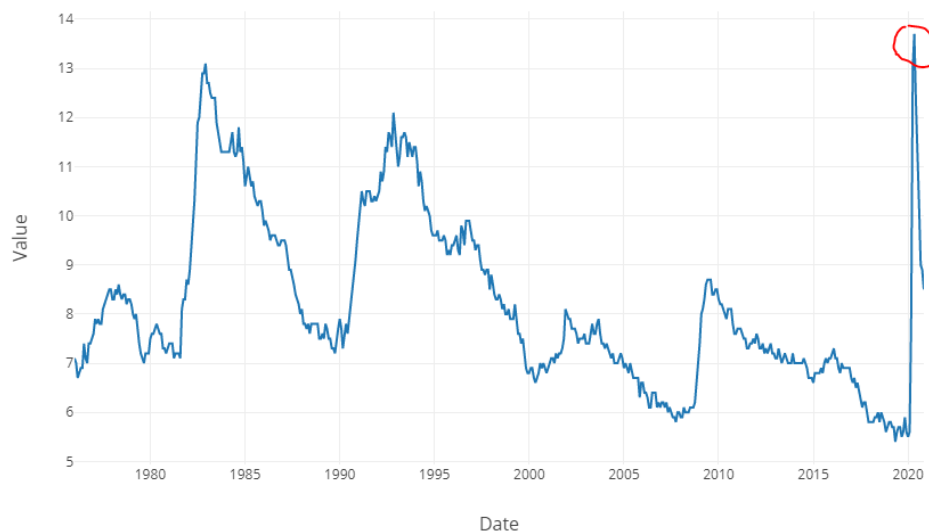


*Figure 15.
Unemployment
rates with focus
on initiating
periods
(encircled)*

The percent changes in unemployment rates between the minimum unemployment value before the economic recession, and the peak unemployment value during the recession were as follows:

- The 80s recession resulted in a 45% change in unemployment rate levels.
- The 90s recession resulted in a 36% change in unemployment rate levels.
- The early 00s resulted in a 16% change in unemployment rate levels.
- The '08 recession resulted in a 31% change in unemployment rate levels.
- The '20 recession resulted in a 60% change in unemployment rate levels.

The monthly labour data shows that the unemployment rate has declined significantly at the start of 2020 then rapidly improved (Figure 16). From this, it can be inferred that the present crisis may resolve at a faster pace compared to previous economic depressions.



*Figure 16.
Monthly
unemployment
rates*

It is hard to confidently predict economic recessions because the causative factors could be external (non-economic) and varied. The overall trend for the past 40 years showed a diminishing unemployment rate and an increasing number of employed individuals. There was an increase in unemployment rate and economic crisis approximately every ten years (Figure 17). Given that the current disturbance was not produced by economic or political causes we can hypothesize that a possible depression is due soon but will have less impact than the 1980 and 1990 recession because of the lower unemployment rate levels.

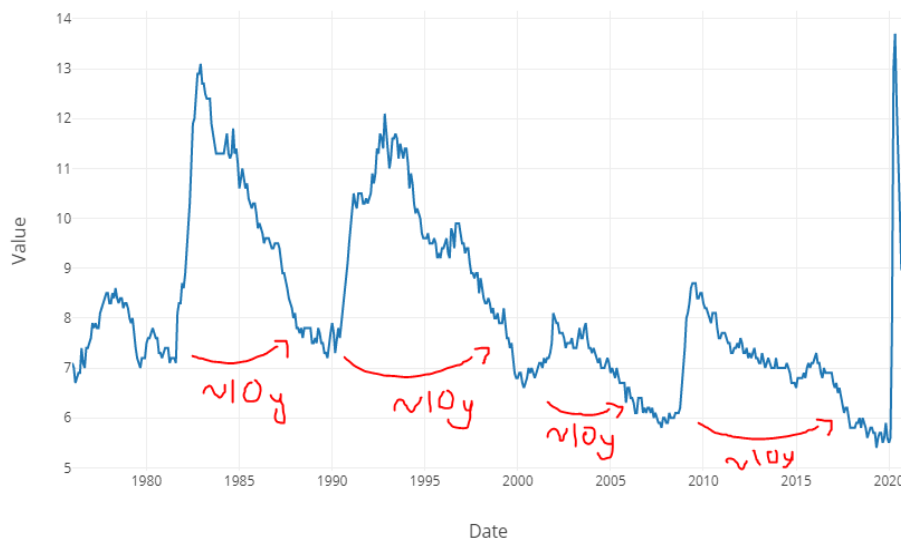


Figure 17. Pattern of economic crises.

Big data tools can help analyze the huge amounts of economic data a country produces. This, in turn, will help make better policies. The Databricks community platform presented us with several issues which would likely not occur while using a paid full account. Big data tools are recommended to help process massive economic data, develop predictive tools and applications appropriate for government, private, public and personal uses.

Summary

From 1976- 2000, the Canadian labour force had a general increasing trend, with decreasing unemployment rates. Unemployment rates were higher among young adults, males, and populations living in Atlantic Canada. Fluctuations were prominent in the top two sectors of goods-producing and service-producing industries. The unemployment rate is predicted to be 8.8% for 2021 and between 3-9% by 2031.

Increasing unemployment rates correspond to recessions, which occur approximately every 10 years. On average, recovery from a recession was achieved within 6 years. However, the extent and direction of the current recession is hard to predict because of its atypical pattern.

This project effectively utilized big data tools, particularly Spark, in exploring 4 decades of data and performing predictive analyses to provide values 10 years in the future.

References

- 1 Bonham M. (2006, Feb 7, last edited 2017, Aug 8). Recession in Canada. The Canadian Encyclopedia. <https://www.thecanadianencyclopedia.ca/en/article/recession>
- 2 Anderson S. (2021, Mar 21). Unemployment and recession - What's the relation? Investopedia. <https://www.investopedia.com/ask/answers/032515/why-does-unemployment-tend-rise-during-recession.asp>
- 3 Potters C. (2021, Jan 26). A review of past recessions. Investopedia. <https://www.investopedia.com/articles/economics/08/past-recessions.asp>
- 4 (2020 June). The early 1980s recession. Wikipedia. https://en.wikipedia.org/wiki/Early_1980s_recession
- 5 Press Release (2020, June 8). COVID-19 to plunge global economy into worst recession since World War II. The World Bank. <https://www.worldbank.org/en/news/press-release/2020/06/08/covid-19-to-plunge-global-economy-into-worst-recession-since-world-war-ii>
- 6 IBM Data Analytics. <https://www.ibm.com/analytics/hadoop/big-data-analytics>
- 7 (2021). Big data defined. Oracle. <https://www.oracle.com/ca-en/big-data/what-is-big-data/>
- 8 The top 10 open source tools for big data. University of Toronto School of Continuing Studies. <https://bootcamp.learn.utoronto.ca/blog/top-10-open-source-big-data-tools/>
- 9 (2021 April 10). Labour force characteristics by industry, annual (x1,000). Statistics Canada. <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410002301>
- 10 Usalcas J and Kinack M. (2017, Jan 6). History of the Canadian labour force survey, 1945 to 2016. Statistics Canada. <https://www150.statcan.gc.ca/n1/pub/75-005-m/75-005-m2016001-eng.htm>
- 11 (2015). Text-international-economics-theory-and-policy. Github. https://saylordotorg.github.io/text_international-economics-theory-and-policy/section_21/53662d042195320aa3bf6f21a4dcb5d8.jpg. https://github.com/saylordotorg/text_international-economics-theory-and-policy