# CIND820 Analysis of Trends in Industries and Business Size in Mississauga – Revised Abstract, Literature Review and Data Structure

# Jennifer McNenly [jmcnenly@torontomu.ca](mailto:jmcnenly@torontomu.ca)

# Student ID: 500105361

# Supervisor: Tamer Abdou

Table of Contents

[Github Repository Details 3](#_Toc127796096)

[Revised Abstract 3](#_Toc127796097)

[Literature Review 6](#_Toc127796098)

[Project Methodology 11](#_Toc127796099)

[Data Preparation and Initial Analysis 12](#_Toc127796100)

[NAICS Clustering 12](#_Toc127796101)

[Prediction 12](#_Toc127796102)

[Data Structure 13](#_Toc127796103)

[Simplified Data Structure Table 13](#_Toc127796104)

# Github Repository Details

My private Github repository entitled “MississaugaBusinesses” has been shared with you and can be found at <https://github.com/mcnenlyj/MississaugaBusinesses>. I have uploaded the relevant Microsoft Word, Excel, html and pdf files. All coding files are in Python.

# Revised Abstract

Businesses in Mississauga have changed over the course of the Covid-19 pandemic. Some businesses closed and others have survived, and in some cases thrived, during this time period. How those businesses that survived and failed are classified and how it can be predicted which ones survived and why are key themes in my research project. The City of Mississauga has identified [6 key large knowledge-based industry sectors](https://www.thefutureisunlimited.ca/industries/)[[1]](#footnote-1) of current business growth that it believes will continue to grow and attract businesses and workers to live and work in Mississauga.

Investigating the research questions can be used to provide insight into the period under study as well as future years: What did businesses look like over the course of the data set? Did the ones that failed before 2020 look similar to the ones that failed after 2020? Which factors impacted the viability of businesses to survive during 2020, the first year of the pandemic (business size, NAICS code, location, etc…)? Do businesses that survived and thrived fall within city’s 6 key broad industry areas? Can we forecast which businesses failed in 2020 and why? Were smaller businesses more adversely affected in certain industries? Were businesses that had been in the Directory longer less likely to fail? What does the latest year of the directory tell us about the current state of businesses in Mississauga? Did business reduce their headcounts? Did industry share shrink in certain industries?

The [Mississauga Business Directory](https://data.mississauga.ca/datasets/mississauga::2021-mississauga-business-directory/explore?location=43.609143%2C-79.675702%2C12.14)[[2]](#footnote-2) is an annual directory data set published in the [City of Mississauga Open Data Catalogue](https://data.mississauga.ca/)[[3]](#footnote-3) set that lists businesses in Mississauga by postal code, size of business, NAICS code and NAICS description. There are 5 years of directories covering the period 2016 – 2021. 2020 appears to be the only year it was not collected. Each directory has a unique ID for a business that you can use to track how it has evolved from year to year. When an ID is no longer in the directory my assumption is that the business has closed and is no longer in business. When a new ID is added this denotes either a brand new business that opened in the interim or a previously unlisted business that is now formally listed in the directory. I note that the directory is voluntary and only businesses who agree are listed. I propose using the full set of directories for the years 2016, 2017, 2018, 2019 and 2021, as they represent the immediate pre-covid period and post-2020 period. The number of businesses dramatically dropped from 16518 to 14825 between 2019 and 2021. If data for 2022 is published in the coming weeks, I would propose increasing the time series to further investigate and account for the full impact of the pandemic on Mississauga businesses.

The NAICS code system is very granular and provides too many levels. Classification (including logistic regression, decision trees and random forest) will help us to group businesses that survived or failed into broader categories, particularly those that the city sees as meaningful to its future growth. Effectively forecasting businesses that closed in Mississauga can be achieved by using time series machine learning methods, including linear regression.

# Literature Review

I was not able to find any studies specifically using the data set I have chosen. I am unaware of anyone else releasing a public study of an analysis of the Mississauga Business Directory either before Covid or recently. However, from my literature review, I was able to find studies of businesses in other jurisdictions that asked pertinent questions both in the pre-pandemic and pandemic periods.

I was able to find and purchase a very useful chapter on using machine learning to predict business survival during Covid-19 from businesses listed in Yelp and other sources. This resource was not available on the Toronto Metropolitan University library and I have included the chapter in my Github repository [here](https://github.com/mcnenlyj/MississaugaBusinesses/blob/main/Bibliography_Resources/Using%20Machine%20Learning%20to%20Predict%20Business%20Survival%20in%20COVID-19.pdf) for your convenience. The researchers, Garthi and Mathur, used a variety of classifiers, namely KNN, Logistic Regression, Random Forest and Neural Network to make predictions about the survival of these businesses and assessed each model’s performance. They found that all models were better at predicting open businesses than closed ones by having higher precision, recall and *F*1 scores for Open Businesses. They noted that Random Forest is the best according to the sensitivity metric, KNN is the best according to the specificity metric and Neural Network is the best when balancing the sensitivity and specificity metrics (balanced accuracy). Their conclusion was that the presence of four Covid features suggested that the survival of a business was tied to providing services that ensured customer health and safety. [[4]](#footnote-4)

A survey of small businesses across several US states released in the first few months of the pandemic found that mass layoffs and closures had already occurred. The sample found 43 percent of businesses had temporarily or permanently closed and additionally, on average, reduced their head count of employees by 40 percent. Further, the same study found that due to the forced closure of some businesses, while others deemed essential remained open, created an “existential threat” to businesses in certain industries, most notably, *Arts and Entertainment*, *Tourism and Lodging* and *All Retailers except Grocery*. Businesses with fewer employees were most adversely affected. [[5]](#footnote-5)

A pre-pandemic study by Zhang and Stevens shows that businesses that are older and more established contribute to employment growth and thus have larger employee head counts over time. In over half of the industries studied, but most evidently in *Manufacturing*, *Wholesale Trade* and *Education Services*. Other industries like *Construction* and *Other Services* were less likely to grow and the *Information* and *Finance & Insurance* industries the age of a business was not found to be statistically significant.[[6]](#footnote-6) Hyatt, on the other hand argues that, though firm age is more important than business in driving employment growth, it is the younger new businesses (less than 10 years old) that are driving job growth. These younger businesses have a higher need for credit and borrowing compared to older businesses.[[7]](#footnote-7)

A pre-pandemic Canadian study by Dixon for Statistics Canada found little evidence to support the assertion that smaller businesses have proportionally higher employment growth rates year-on-year. Employment growth rates rose with business size for businesses with fewer than 20 employees. In businesses above the 20-employee threshold no relationship emerges between employment growth and business size. These results are consistent with the average proportionate growth condition of Gibrat’s Law; the assertion of French economist Robert Gibrat that average employment growth is independent of business size.[[8]](#footnote-8)

Another Statistics Canada survey early in the pandemic found that in Ontario staffing actions taken by businesses showed on average 38.9 percent laid off staff and only 2 percent hired more staff. This compares to nationally where on average 40.5 percent laid off staff and only 2.5 percent hired more staff. In the same survey industries by NAICS code across Canada that fared the worst included *Accommodation and Food Services [72]*, *Retail Trade [44-45]* and *Construction [23]* where 67.5 percent, 54.9 percent and 50.7 percent of businesses laid off staff respectively. Those industries impacted negatively the least included *Agriculture, forestry, fishing and hunting [11]*, *Management of companies and enterprises [55]* and *Finance and insurance [52]* where 14.4 percent, 14.6 percent and 17.8 percent of businesses laid off staff respectively. In terms of size of business, the hardest hit were those with *20-99 employees*, *5-19 employees* and *100-249 employees* where 61.7 percent, 58.5 percent and 54.5 percent of businesses laid of staff respectively. Those businesses that fared the least negatively were those with *0 employees*, *1-4 employees* and *250-499 employees* where 5.1 percent, 31.4 percent and 49.3 percent of businesses laid of staff respectively. This includes all employees who would receive a T4. Excluded from number of employees are business owners, contract workers and other personnel who would not receive a T4.[[9]](#footnote-9) Statistics Canada followed this up later in the same year with similar results for most industries and business sizes in terms of status of being open or closed.[[10]](#footnote-10)

McKinsey published a study early in the pandemic on the impact of Covid-19 on small businesses in the US and found that businesses with less than 100 employees where the hardest hit and most vulnerable. In addition, *Accommodations and food services*, *Arts, entertainment and recreation* and *Personal services* were the most vulnerable industries for businesses. *Healthcare and social assistance*, *Finance and insurance* and *Professional, scientific, and technical* industries were among the least vulnerable. [[11]](#footnote-11) But a mere month later McKinsey reported that Finance and insurance and Healthcare and social assistance had also moved into the most vulnerable category.[[12]](#footnote-12) The takeaway being the things moved rapidly at the beginning of the pandemic and accurate data was difficult to obtain.

Clustering of industries is another avenue of research that has been performed in many studies. Recent examples, like Wang and Wen (2021), looked at homeownership, income, education level and the relationship to the type of businesses that were opened in neighbourhoods.[[13]](#footnote-13)

The data set I have chosen does not contain demographic information on the residents (homeowners or renters) in each postal code. I will therefore not address any correlation between where people live, nor their demographic characteristics such as income, education and the types of industries and businesses that are nearby. Further, the data set does not contain any detailed demographic and financial information about the businesses, such as income, expenses, and demographic details about the owner of the business. I may be able to show clustering of certain industries but into certain areas of the city but there is no data to imply causation other than the fields I have in the combined data set.

I believe my work is worth doing as there is not a detailed study of the effect of the Covid-19 pandemic on businesses in Mississauga. The results will fit in with what has gone before in terms of answering questions about the size of businesses over time and overall effect on economic growth (or in the case of the pandemic, shrinkage). It will also answer questions about the type of industries that were affected either positively or negatively by the pandemic. Does what was reflected in the Statistics Canada’s national and provincial survey early in the pandemic mirror what happened to businesses and industries in a city like Mississauga? Covid-19 was an unforeseen emergency that strained the global economy. The ramifications of its impact will be felt for years to come.

**Hypothesis & statistical testing**

1. Company age (years listed in the business directory) has a significant positive effect on whether a company remained in the directory by 2021.
2. Company size has a significant positive effect on whether a company remained in the directory by 2021.
3. Company industry has a significant positive effect on whether a company remained in the directory by 2021.

# Project Methodology

****

# Data Preparation and Initial Analysis

Prepare combined data set and add new fields to the data set that can be created after examination of the data. The first new field would be the year of the Directory the records came from when combined. The second additional field will note if a business was no longer listed in a subsequent year of the directory and is assumed closed. The third additional field will note if a business was net new to the directory in a given year. The fourth additional field will note the age of a business, that is, the number years a business has been listed in the directory. The fifth field would be a new unique ID to provide to the newly combined data set.

Next steps would be to transforming fields, where necessary, into useful data types for analysis. Analysis by year can then be done to see the broad patterns in the data. For example, for businesses that closed each year what industries were they in and what size of business?

# NAICS Clustering

Most studies in the literature review dealt with what the early months of the pandemic looked like in terms of the kinds of industries that were negatively impacted. I would propose looking at the data in a similar way to see if Naics clustering of closed businesses yields similar results to what was seen in broader US and Canadian studies.

# Prediction

Dividing the data into Training sets before 2019 and Test 2019 & 2021 rather than 80 / 20 split would most likely not yield good results for prediction since Covid was a unforeseen, highly negative business event. Rather, using an 80 / 20 split of the complete data set for all years would likely yield better results after feature engineering. Beyond the fields I propose to add to the data set during data preparation I will then see what other feature engineering techniques can be used to parse the data set into a useful tool for creating a training and test set. I will plan to use Decision Tree, KNN, Logistic Regression, Random Forest and Neural Network to compare results.

# Data Structure

## Simplified Data Structure Table

(Current field types from EDA reports. Fields in green are fields I propose adding to the comined data set.)

|  |  |  |
| --- | --- | --- |
| Field Name | Field Type | Description |
| FID | Real number | Unique ID in the uncombined data sets |
| BusinessID | Real number | Business ID |
| NAME | Categorical | Business name |
| EmplRange | Categorical | Employee Range |
| NAICSTitle | Categorical | NAICS Title |
| NAICSDescr | Categorical | NAICS Description |
| NAICSCode | Real number | NAICS Code |
| Phone | Categorical | Business Phone |
| Fax | Categorical | Business Fax |
| TollFree | Categorical | Business Toll Free Number |
| Email | Categorical | Business Email |
| WebAddress | Categorical | Business Web Address |
| StreetNo | Real number | Address - Street Number |
| StreetName | Categorical | Address - Street Name |
| Address | Categorical | Address – Street Number and Street Name |
| PostalCode | Categorical | Address – Postal Code |
| BldgNo | Categorical | Address – Building Number |
| UnitNo | Categorical | Address – Unit Number |
| Modified | Categorical (should be time) | Date Employment Number was Updated |
| PIN | Real number | PIN code for longitude and latitude |
| X | Decimal | Latitude |
| Y | Decimal | Longitude |
| CHArea | Categorical | Character Area |
| BIA\_Name | Categorical | Business Improvement Association - Acronym |
| BIAFulName | Categorical | Business Improvement Association – Full Name |
| Ward | Real number | Ward |
| DirectoryYear | Real number | Year of the Directory the data is from |
| Closed | Binary | Did the business close in the subsequent year |
| IsNew | Binary | Did the business get listed in the current year |
| BusinessAge | Real number | How many years the business has been listed in the Directory |
| UniqueID | Real number | New unique ID for combined data set |

1. Industries | Unlimited Mississauga, Accessed January 22, 2023. <https://www.thefutureisunlimited.ca/industries/> [↑](#footnote-ref-1)
2. 2021 Mississauga Business Directory | Open Data Catalogue, Published January 19, 2022. Data updated January 28, 2022. <https://data.mississauga.ca/datasets/mississauga::2021-mississauga-business-directory/explore?location=43.609064%2C-79.675702%2C12.00> [↑](#footnote-ref-2)
3. City of Mississauga Open Data Catalogue, Accessed January 22, 2023. <https://data.mississauga.ca> [↑](#footnote-ref-3)
4. Garthi, G., Mathur, R. (2023) “[Using Machine Learning to Predict Business Survival in COVID-19](https://github.com/mcnenlyj/MississaugaBusinesses/blob/main/Bibliography_Resources/Using%20Machine%20Learning%20to%20Predict%20Business%20Survival%20in%20COVID-19.pdf)”, pgs 249-261 In: S. Bhattacharyya et al. (eds.), *Human-Centric Smart Computing*, Smart Innovation, Systems and Technologies 316, https://doi.org/10.1007/978-981-19-5403-0\_21 [↑](#footnote-ref-4)
5. Bartik, A. W., Bertrand, M., Cullen, Z. B., Glaeser, E. L., Luca, M., Stanton, C. T. (2020). [How are small businesses adjusting to covid-19? Early evidence from a survey (no. w26989)](https://www.nber.org/system/files/working_papers/w26989/w26989.pdf). National Bureau of Economic Research. [↑](#footnote-ref-5)
6. Zhang, T., Stevens, D. (2009). [Business Age, another Source of Employment Growth](https://nces.ed.gov/FCSM/pdf/2009FCSM_Zhang_XI-B.pdf). Merrick School of Business, University of Baltimore [↑](#footnote-ref-6)
7. Hyatt, H.R., (2022). [Firm Age and Job Creation in the US](https://wol.iza.org/uploads/articles/629/pdfs/firm-age-and-job-creation.pdf). US Census Bureau, USA, and IZA, Germany [↑](#footnote-ref-7)
8. * Dixon, D., Rollin, A.M. (2012) [Firm Dynamics: Employment Growth Rates of Small Versus Large Firms in Canada](https://www150.statcan.gc.ca/n1/en/pub/11-622-m/11-622-m2012025-eng.pdf?st=Hb3aAmdr). Catalogue no. 11-622-M, no. 025. The Canada in Transition Series. Economic Analysis Division, Statistics Canada.

   [↑](#footnote-ref-8)
9. [Staffing actions taken by businesses during the COVID-19 pandemic, by business characteristics](https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3310023101) (2020) Statistics Canada. Table: 33-10-0231-01. Release date: 2020-04-29. [↑](#footnote-ref-9)
10. [Impact of COVID-19 on business or organization status, by business characteristics, third quarter of 2020](https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=3310027601). Statistics Canada, Canadian Survey on Business Conditions, August 2020. [↑](#footnote-ref-10)
11. Dua, A., Mahajan, D., Velasco, Y.[Covid-19’s effect on jobs at small businesses in the United States](https://www.mckinsey.com/industries/public-and-social-sector/our-insights/covid-19s-effect-on-jobs-at-small-businesses-in-the-united-states). May 5, 2020. McKinsey & Company. [↑](#footnote-ref-11)
12. Dua, A., Ellingrud, K., Mahajan, D., Silberg, J. [Which small businesses are most vulnerable to Covid-19 – and when](https://www.mckinsey.com/featured-insights/americas/which-small-businesses-are-most-vulnerable-to-covid-19-and-when). June 18, 2020. McKinsey & Company. [↑](#footnote-ref-12)
13. Wang, B., Wan, B. [The spatial distribution of businesses and neighborhoods: What industries match or mismatch what neighborhoods?](https://www.sciencedirect.com/science/article/pii/S0197397521001296) Habitat International. 117 (2021) 102440. [↑](#footnote-ref-13)