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

**Student ID: 500105361**

**CIND820 XJH - Big Data Analytics Project - W2023**

**Abstract for Capstone Project**

**Data Analysis of the Mississauga Business Directory**

**January 21, 2023**

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.

Investing the research questions can be used to provide insight into the period under study as well as future years: 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 fail and why? Do businesses that failed predominantly fall into a certain group (e.g., small restaurants in certain areas of the city)? What does the latest year of the directory tell us about the current state of businesses in Mississauga?

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 track how it has evolved from year to year. When an ID is no longer in the directory, it is assumed, the business has closed. When a new ID is added, it is assumed, a new business that opened in the interim or is a previously unlisted business. \*The directory is voluntary and only businesses who agree are listed. I propose using just the directories for the years 2019 and 2021, as they represent the immediate pre-covid period and post-2020 period, where the number of businesses dramatically dropped from 16518 to 14825. 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. If additional data points are required I would recommend adding household income by postal code that can be incorporated from federal census data for 2021.

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 including those that the city sees as meaningful to its future growth. Effectively forecasting businesses that closed in Mississauga from 2019 to 2021 can be achieved by using time series machine learning methods, including linear regression. I will be using Python to present the code for this project.

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)