

Navigating Economic Challenges: How Business Bankruptcies Influence Food Bank Trends in Toronto, 2018-2023*

Yingqi Pang

April 4, 2024

This study investigates the relationship between business bankruptcies and food bank usage in Toronto from January 2018 to November 2023, incorporating a multiple linear regression model that also considers the date variable. Findings indicate a significant positive correlation, where increases in business bankruptcies are associated with higher food bank usage. This relationship highlights the broader economic vulnerability and the critical role of food banks in urban social support systems. The insights from this research underscore the importance of strategic planning for food security in the face of economic downturns, enhancing the resilience of communities during financial crises.

Table of contents

Introduction	2
Data	3
Data Collection	3
Data Cleaning	3
Variables	4
Measurement	5
Data Analysis	6
Model	8
Model Setup	8

*Code and data are available at<https://github.com/pangyin2/Navigating-Economic-Challenges-How-Business-Bankruptcies-Influence-Food-Bank-Trends-in-Toronto.git>

Model Justification	8
Features	9
Model concerns	9
Results	10
Discussion	12
Findings	12
Bias and ethical concerns	12
Weakness and next steps	13
Reference	13

Introduction

In Toronto, the economic health of businesses and the welfare of its residents are threads interwoven in the complex social tapestry. The closure of a business is more than a commercial end—it reverberates through the community, often resulting in an increased reliance on social support systems like food banks. This project investigates the relationship between business bankruptcies—a stark indicator of economic downturns and the utilization of food banks within Toronto from January 2018 to November 2023, alongside date factors.

Economic downturns manifest in numerous ways, one of which is an increase in the number of business bankruptcies. This phenomenon has been associated with an increase in food insecurity, a challenge that food banks strive to mitigate. The article “The State of Food Security and Nutrition in the World” highlights how economic slowdowns can exacerbate hunger and food insecurity across the globe, suggesting that food banks play a crucial role in safeguarding against these downturns (RESILIENCE 2017). Conversely, David Cook and his colleague emphasize the complex interplay between economic indicators and health outcomes, offering a framework for understanding how macroeconomic fluctuations can influence social welfare systems (Cook and Davísdóttir 2021). Both researches underscore the broader economic implications and the need for policies that can preemptively address the socio-economic fallout from business failures.

While existing research provides broad insights into the relationship between economic indicators and food security, this project narrows the focus to the unique economic environment of Toronto. It differentiates itself by offering a micro-level analysis that considers local variables and direct measures of community impact, such as food bank usage statistics, which have been less explored in the existing body of research. The estimand of the study is the change in the number of food bank usages that would occur if there were an increase in the number of business bankruptcies within the same period. This is considered in terms of monthly data collected from the City of Toronto, focusing on the direct correlation between the economic downturns, as evidenced by business bankruptcies, and the resultant demand on food bank

services. It endeavors to quantify the extent to which business bankruptcies can predict variations in food bank usage, providing a localized understanding that can inform community planning and social policy. To achieve this aim, the project adopts a quantitative approach, leveraging multiple linear regression analysis to dissect the relationship between the economic downturns—as represented by business bankruptcies—and the demand on food banks, alongside the date factor. The methodology includes a detailed cleaning of the datasets to focus on the targeted time frame and the use of statistical software to ensure precise, robust modeling.

Data

Data Collection

The datasets come from the City of Toronto’s open data portal (Smith Year of Publication), specifically focusing on client visits to Daily Bread Food Bank member food banks per month, starting from the year 2009. The Daily Bread Food Bank dataset records the number of client visits each month, but it is important to note that the data might be an underestimate of the actual figures due to some larger agencies not reporting their numbers and North York Harvest Food Bank agencies being excluded from this data. The Daily Bread Food Bank provides these monthly client visit statistics to the City Manager’s Office, which in turn reports them on a quarterly basis in its Toronto Dashboard. The data is available in CSV format and can be exported directly from the Toronto Dashboard for analysis. It’s indicated that the data was last refreshed on May 31, 2017, and there is no indication that further refreshes will occur. This dataset is categorized under civic issues such as affordable housing and poverty reduction, fitting within broader topics of community services. The licensing falls under the Open Government License - Toronto, meaning it can be used as long as appropriate credit is given. The second dataset focuses on the number of business bankruptcies in the City of Toronto. It would include monthly or yearly counts of businesses that have declared bankruptcy within a specified period. Both datasets are likely to provide insights into community needs and economic health. When using these datasets for analysis, it will be crucial to consider any potential biases or gaps in the data, especially given the noted underestimation of food bank usage data. Additionally, ethical considerations should guide the analysis and presentation of findings, especially due to the sensitive nature of the data concerning individuals and organizations facing financial hardship.

Data Cleaning

The data cleaning process for this project involved a series of methodical steps to prepare the “food bank usage” and “Number of Business Bankruptcies for the City of Toronto” datasets for analysis. Initially, necessary R libraries (R Core Team 2020), including `dplyr` for data manipulation (Wickham et al. 2021), `lubridate` (Grolemund and Wickham 2011) for date handling,

and `ggplot2` for visualization (Wickham 2016), were loaded to provide the required functionality. Subsequently, the datasets were read into R using the `read.csv` function. To standardize the date information across both datasets, a new ‘Date’ column was created by concatenating the ‘Year’ and ‘Period’ columns and converting the resulting string into a date format. This conversion was crucial to ensure consistency in temporal data representation, facilitating accurate comparisons and analysis. Following the date standardization, both datasets underwent a filtering process to retain records from January 2018 to November 2023. This specific time frame was selected to align with the project’s objectives, focusing on recent trends and excluding irrelevant historical data. The next step involved merging the filtered datasets based on the ‘Date’ column to create a single, comprehensive dataset named `merged_data`. This merger was pivotal in consolidating the information from both sources, enabling a cohesive analysis of the relationship between business bankruptcies and food bank usage over time. To streamline the merged dataset and focus on relevant variables, the dataset was reformatted to include only the ‘Date’, newly calculated ‘year’ and ‘month’ from the ‘Period’ column of the food bank usage dataset, ‘food_bank_usage’, and ‘num_of_bankruptcies’. This selection process was executed using the `select` function from the `dplyr` package, effectively reducing the dataset to the essential columns for analysis. Finally, the cleaned and consolidated dataset was exported to a CSV file named “cleaned_data.csv” using the `write.csv` function, with `row.names` set to `FALSE` to omit row names from the output. This file serves as the basis for subsequent analysis, encapsulating the processed data in a structured and accessible format.

Variables

Date: It records the specific month and year of each observation. It is essential for tracking changes over time and aligning the data points from both datasets. The ‘Date’ facilitates time-series analysis, allowing for the examination of trends and patterns in food bank usage and business bankruptcies across the specified period (January 2018 to November 2023).

Year and Month: Derived from the ‘Date’ variable, ‘Year’ and ‘Month’ provide disaggregated temporal markers. These allow for more granular analysis, such as seasonal variations in food bank usage or bankruptcy filings and are crucial for any time-based comparisons or aggregations within the analysis.

Food Bank Usage: This is a count variable indicating the number of visits or usages of food banks recorded each month. It serves as a proxy for the level of food insecurity and the demand for food bank services within the community. As the dependent variable in the linear model, it is analyzed in relation to the number of business bankruptcies to understand if economic downturns, as reflected by increased bankruptcies, correlate with greater reliance on food banks.

Num_of_Bankruptcies: Representing the number of business bankruptcies filed in Toronto each month, this variable is used as an independent variable in the linear regression model. It quantifies economic distress within the local business community and is examined for its

potential influence on food bank demand, under the hypothesis that higher bankruptcy rates could lead to increased food bank usage.

These variables were carefully selected and processed from the original datasets, which contained more extensive data, to hone in on the study’s focus. Data manipulation techniques, including the use of `dplyr` for data transformation and `lubridate` for handling date variables, were employed to prepare and refine the data for analysis.

Measurement

We meticulously measured and analyzed the relationship between business bankruptcies and food bank usage within Toronto from January 2018 to November 2023. This required a detailed examination and manipulation of the variables within our dataset. The “Date” variable was transformed into a proper datetime format to ensure temporal accuracy and facilitate time-series analysis. For measurement, “Food Bank Usage” was quantified as the count of food bank visits or services utilized each month, serving as a direct indicator of community reliance on food support services. This variable was treated as a continuous measure to observe and analyze trends over time. “Num_of_Bankruptcies”, representing the monthly count of business bankruptcies, was also treated as a continuous variable, reflecting the economic health of the local business landscape.

Both variables were considered within the predefined temporal scope of our study, which did not originally categorize data based on specific economic or social events. However, our analysis implicitly captures any fluctuations in these measures that could be associated with broader economic trends or specific policy changes affecting Toronto’s economy and social welfare systems. The accuracy of “Food Bank Usage” and “Num_of_Bankruptcies” was crucial, necessitating validation steps such as ensuring non-negative integer values and consistency across reporting periods. Efforts to ensure data integrity included rigorous cleaning protocols, such as filtering out incomplete records and standardizing date formats. Challenges in measurement, such as potential underreporting in “Food Bank Usage” or the exclusion of certain data points due to missing values in “Num_of_Bankruptcies”, were addressed through careful data preprocessing. The selection of these variables and the attention to their precise measurement allow for a robust analysis of how economic downturns, as evidenced by increased business bankruptcies, correlate with the demand for food bank services, providing insights into the socio-economic dynamics within Toronto.

Data Analysis

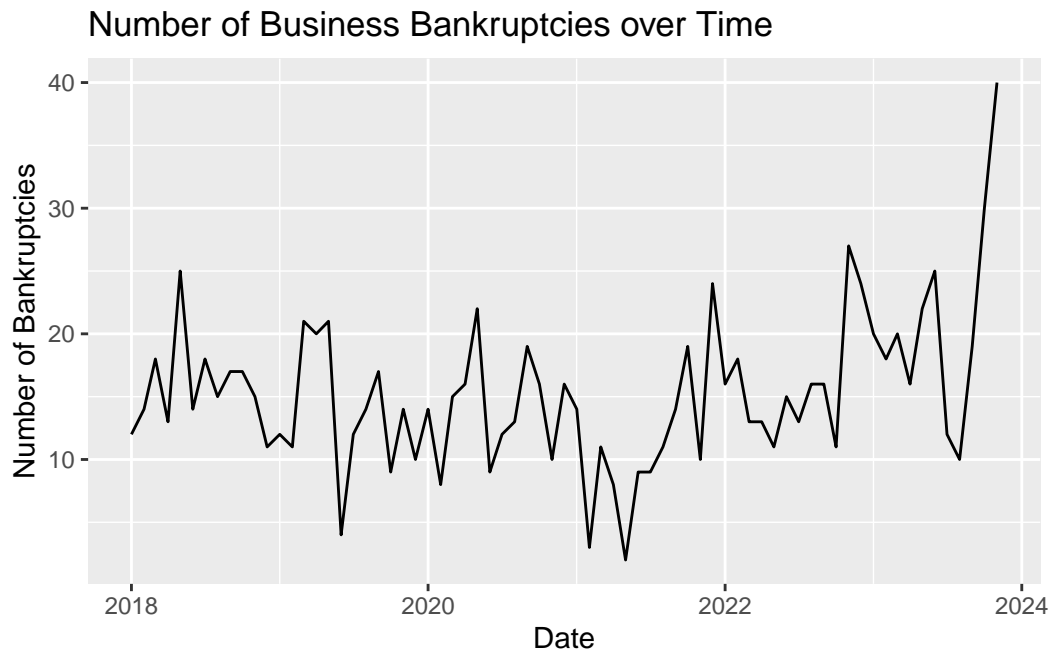


Figure 1: Food Bank Usage across time from 2018 to 2023

Figure 1 displays the trend of food bank usage in the City of Toronto from January 2018 to November 2023. The x-axis represents the date, spanning nearly six years, while the y-axis indicates the food bank usage, measured in an unspecified unit. From the graph, we can observe that food bank usage has been on an overall upward trend over the specified period. There are noticeable fluctuations throughout the years, with some periods showing sharper increases or temporary plateaus. The trend line begins at a point representing the food bank usage at the start of 2018 and ends with the last available data point in November 2023, where the usage is at its highest. The graph is marked by a series of peaks and troughs, suggesting variability in food bank usage, potentially correlated with seasonal trends, economic conditions, or policy changes affecting food security in Toronto.

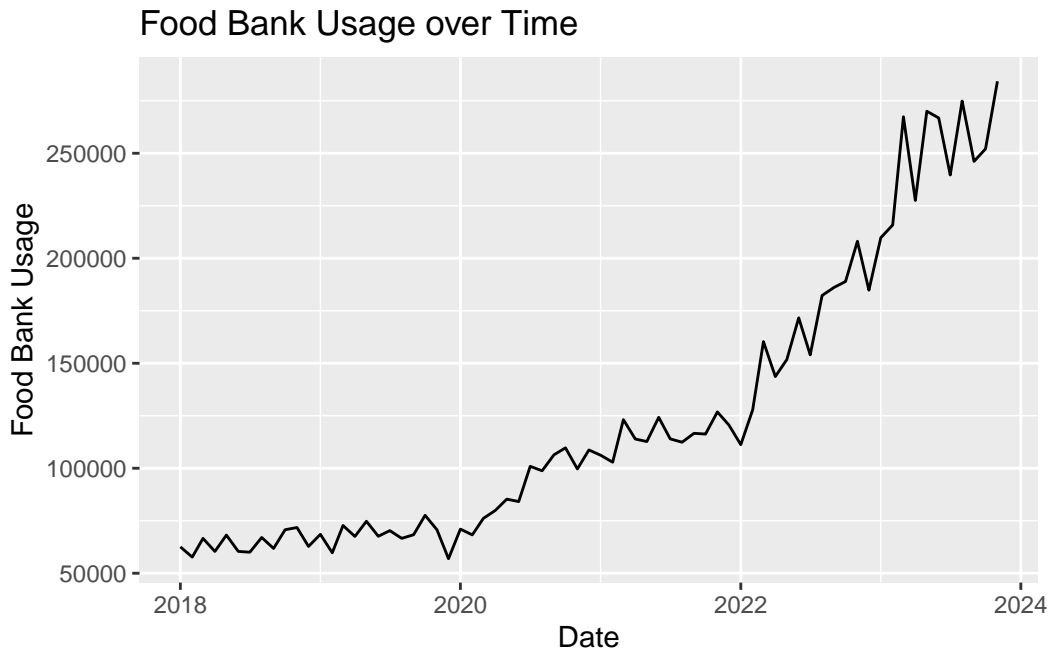


Figure 2: Number of Business Bankruptcies across time from 2018 to 2023

Figure 2 shows the trend of business bankruptcies in the City of Toronto from January 2018 to November 2023. The x-axis represents time, and the y-axis represents the number of business bankruptcies. The line graph depicts the count of business bankruptcies recorded each month. The data demonstrates fluctuations over time with varying frequencies and amplitudes. There is an observable increase in the number of bankruptcies towards the end of the period, with a significant spike in the latter part of 2023, where the figure reaches the highest point on the graph. The rest of the timeline shows periodic increases and decreases, indicating the variable nature of business bankruptcies over the observed years.

Model

Model Setup

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij} + \varepsilon_{ij}$$

Model Justification

- Y_{ij} is the Food Bank Usage for observation j in month i .
- X_{ij} is the Number of Business Bankruptcies for observation j in month i .
- Z_{ij} is the time variable, representing the date for the i th month and j th observation, which allows for capturing trends over time.
- β_0 is the intercept or constant term, which represents the expected value of Food Bank Usage when the Number of Business Bankruptcies is zero.
- β_1 is the slope coefficient or the estimated change in Food Bank Usage for a one-unit increase in the Number of Business Bankruptcies.
- β_2 is the slope coefficient for the Date, indicating the change in Food Bank Usage over time, holding the Number of Business Bankruptcies constant.
- ε_{ij} is the error term or the deviation of the actual value of Food Bank Usage from the predicted value based on the regression equation.
- **The Vector of Control Variables:** Z_{ij} acts as a control variable accounting for the temporal aspect, capturing the effect of time on Food Bank Usage. This inclusion helps to isolate the specific impact of the Number of Business Bankruptcies on Food Bank Usage. In a more comprehensive model, one would include additional control variables such as unemployment rates, other economic indicators like GDP or inflation, policy changes, and possibly weather events that could affect Food Bank Usage. These controls are essential to differentiate the effect of business bankruptcies from other factors that may also impact the dependent variable.
- **Comparison Group:** The model implicitly compares the observed Food Bank Usage against a hypothetical scenario where the Number of Business Bankruptcies is zero, as suggested by the intercept β_0 . This baseline allows for the assessment of how Food Bank Usage changes in response to business bankruptcies. Since the model does not have an explicit control group, the ‘treatment’ effect here is the variation in the number of business bankruptcies. The Date variable Z_{ij} , while not a ‘treatment’ variable, enables the examination of how Food Bank Usage changes over time, providing a nuanced understanding of the time-related trends.

Features

Our model utilizes a multiple linear regression framework, focusing on the “num_of_bankruptcies” variable to gauge its influence on “food_bank_usage,” suggesting that economic stress may lead to a greater need for social support services. The inclusion of the “Date” variable captures the element of time, allowing the analysis to account for temporal trends and variations that could impact food bank usage. This temporal aspect is particularly crucial for recognizing any seasonal patterns or year-over-year changes in the relationship being studied. The model’s design facilitates a clear interpretation of the results, with the straightforward inclusion of a single independent variable alongside a temporal control ensuring the analysis remains focused and interpretable.

Model concerns

The primary concern with this model is its potentially limited explanatory power due to the modest R-squared value, which indicates that a considerable amount of the variation in food bank usage is not explained by the number of business bankruptcies and the Date variable alone. This suggests the presence of other influential factors that the model does not account for, such as macroeconomic conditions, changes in social welfare policies, or demographic shifts, which could also significantly impact food bank usage. The lack of additional control variables may lead to omitted variable bias, where the effect of business bankruptcies is overestimated or underestimated because other relevant factors are not included in the model. Furthermore, the linear assumption may be an oversimplification, as the relationship between economic conditions and food bank usage is likely complex and potentially non-linear. The influence of outliers and extreme values is another concern, as they can distort the model’s predictions. Additionally, the model assumes that the relationship between the variables has been consistent over time, an assumption that may not hold true if there have been structural changes in the economy or if the behavior of food bank users and the reasons for bankruptcies have evolved. To address these concerns and enhance the model’s robustness, future iterations could expand the set of explanatory variables to include more socio-economic factors, consider alternative model specifications, and employ different statistical techniques, such as non-linear modeling or time series analysis, to better capture the complexity of the relationships between economic health, social welfare, and the use of food banks.

Results

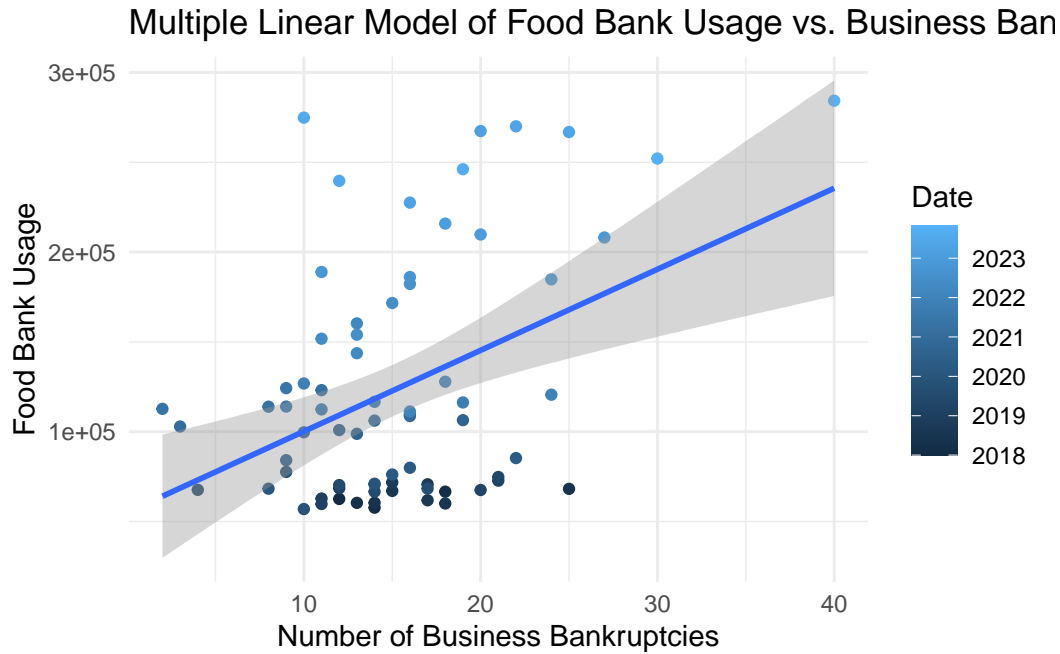


Figure 3: Multiple Linear Model of Food Bank Usage vs. Business Bankruptcies vs. Date

Figure 3 shows a scatter plot with a multiple linear regression analysis examining the effect of business bankruptcies on food bank usage, while also considering the factor of time (years from 2018 to 2023). On the x-axis, the number of business bankruptcies is displayed, and the y-axis shows the count of food bank usage. Each dot represents a data point at the intersection of these two variables, colored according to the year it represents. The blue line running through the data points is the regression line, indicating the average effect of the number of business bankruptcies on food bank usage as predicted by the model when accounting for the year. The grey area around the line is the confidence interval, providing a range where we can expect the true average effect to lie with a certain degree of confidence. The plot shows a positive trend, suggesting that as the number of business bankruptcies increases, food bank usage also tends to increase. The varying colors of the dots suggest there may be yearly trends or differences in the data across the years. The points are spread out around the trend line, which indicates variability in the data that is not fully explained by the number of bankruptcies alone. The graph implies that while bankruptcies are an important factor, other time-dependent factors not explicitly modeled may also affect food bank usage.

	Estimate	StdError	tValue	Pr
(Intercept)	-1.598891e+06	88438.847055	-18.079059	0.00e+00
num_of_bankruptcies	2.086215e+03	495.071062	4.213972	7.57e-05
Date	9.092815e+01	4.842935	18.775422	0.00e+00

Table 1: Regression Table for food bank usage vs. number of business bankruptcies vs. Date

Table 1 shows the output of a multiple linear regression model, showing the estimates, standard errors, t-values, and p-values for an intercept, the number of business bankruptcies, and a Date variable.

- **Intercept:** The model’s intercept is estimated at approximately -1.59891e+06 (or -1,598,910 when expressed in standard numerical form). The intercept represents the expected value of food bank usage when the number of business bankruptcies and the Date variable are zero. However, given that the Date variable is likely to be a time factor and not a variable that can take a value of zero, the interpretation of the intercept is not meaningful without the context of the Date encoding.
- **num_of_bankruptcies:** For the number of business bankruptcies, the model estimates an increase of approximately 2086.21 in food bank usage for each additional bankruptcy. The standard error of this estimate is 495.07, and the t-value of 4.21397 suggests that this coefficient is statistically significant (the corresponding p-value is 0.0000756913 or about 7.56e-05, indicating strong evidence against the null hypothesis of no association).
- **Date:** The coefficient for the Date variable is approximately 90.282 (indicating a time trend), with a standard error of 4.842935. The t-value for the Date variable is 18.77542, and the associated p-value is exceedingly small (1.31184e-28), showing that there is a very strong statistical significance for the time factor in the model. This suggests that over time, irrespective of the number of business bankruptcies, food bank usage is increasing.

The very low p-values for both the number of bankruptcies and the Date variable suggest that these factors are highly statistically significant predictors of food bank usage. The model output suggests that both the number of business bankruptcies and the Date are important predictors in explaining the variability of food bank usage, with both showing a statistically significant relationship with the usage of food banks. The Date coefficient’s significance also implies that food bank usage has a time trend, possibly reflecting underlying changes in the economy or social policies over time.

Discussion

Findings

Our multiple linear regression analysis from January 2018 to November 2023 indicates an increase in food bank usage in Toronto, with possible links to seasonal or economic patterns. The model reveals a statistically significant association between the rise in business bankruptcies and higher food bank usage. Specifically, each additional reported business bankruptcy is associated with an average increase of approximately 2,086 food bank visits, a figure derived from the statistically significant coefficient β_1 of 2,086.215 with a standard error of 495.071062. While the number of bankruptcies is a notable predictor of food bank usage, the moderate R-squared value suggests that other factors, which the model does not capture, also influence this usage.

The time factor, represented by the Date variable, also shows a significant relationship with food bank usage, with the slope coefficient β_2 of 90.92815 indicating that food bank usage has increased over time, independent of the number of bankruptcies. However, the limitations of the model, including the possibility of omitted variable bias, imply that additional variables not included—such as employment rates, policy changes, or cost of living adjustments—may also affect food bank usage. This is reinforced by the model’s residuals, which show that not all variability is accounted for, suggesting the influence of other external factors. Therefore, while the relationship between economic distress and the need for food bank services is evident, a more complex set of variables is likely at play, emphasizing the necessity for robust social support systems to mitigate the impacts of economic hardship.

Bias and ethical concerns

The issue of bias arises from the potential limitations inherent in the data collection and analysis process. Selection bias may occur if the data disproportionately represents certain demographics or time periods, leading to skewed results. Moreover, if food banks’ data collection methods differ or if the criteria for counting bankruptcies vary, this could introduce systemic bias, affecting the reliability of the study’s outcomes. A related concern is survivorship bias, which might manifest if only data from food banks that have continued operations throughout the study period are considered, neglecting those that may have closed down. This could artificially inflate the success rate or efficiency of food banks, offering a non-representative picture of the true extent of food security in the city. Data privacy is another critical ethical consideration. The project must ensure confidentiality and anonymity for individuals and businesses involved. No personal data should be identifiable in the study’s datasets or findings, respecting the privacy of those who have utilized food bank services or have been affected by business bankruptcies. Further ethical concerns include the interpretation and dissemination of the research findings. It is crucial to avoid presenting a causal relationship where only a correlation exists, as this could lead to misguided policy decisions or public misperception.

The study should not contribute to the stigmatization of food bank users or failed businesses, instead promoting a nuanced understanding of the complex economic and social factors at play. The research findings should be communicated with sensitivity to the potential socioeconomic impact on the community. Careful consideration must be given to how the conclusions might be used by policymakers, ensuring that the results are framed in a way that supports positive change and does not lead to reductions in vital services or support for those in need. The project should aspire to inform and enhance community support systems, driving solutions to mitigate food insecurity in the context of economic variability.

Weakness and next steps

The project's weakness lies in the limited scope of the variables considered, which may affect the robustness of the conclusions. The linear model assumes a direct relationship between business bankruptcies and food bank usage, potentially overlooking multifaceted socio-economic factors. The R-squared value indicates a modest fit, implying that additional variables could better capture the dynamics of food bank demand. The presence of outliers suggests that extreme values could be exerting undue influence on the regression results. For next steps, the project should consider a more comprehensive approach by including more variables that could impact food bank usage, such as unemployment rates, changes in social welfare policies, or economic indicators like inflation. Incorporating a time series analysis could also account for seasonal trends and economic cycles. Further, exploring non-linear models might provide a more nuanced understanding of the relationship between the variables. Qualitative data, such as interviews or surveys with food bank users and failed businesses, could offer deeper insights into the causal mechanisms behind the observed statistical trends. Privacy-preserving data integration methods could enrich the analysis without compromising individual confidentiality. The study could also expand geographically or longitudinally, examining different cities or extending the timeframe to bolster the findings' generalizability. Longitudinal tracking of individual businesses and food bank users, while methodologically challenging, could yield rich insights into the life course dynamics underpinning the aggregate patterns observed. Ultimately, the project should aim to provide actionable insights to policymakers and social organizations. This might involve developing predictive models that can guide resource allocation to food banks in anticipation of future economic downturns, ensuring that social support systems are responsive to community needs.

Reference

- Cook, David, and Brynhildur Davísdóttir. 2021. "An Appraisal of Interlinkages Between Macro-Economic Indicators of Economic Well-Being and the Sustainable Development Goals." *Ecological Economics* 184: 106996.
- Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.

- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- RESILIENCE, BUILDING. 2017. “The State of Food Security and Nutrition in the World.” *Rome: Building Resilience for Peace and Food Security*.
- Smith, Phillip L. F. Year of Publication. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.