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

Yingqi Pang

April 2, 2024

This study analyzes the connection between business bankruptcies and the usage of food banks in Toronto from January 2018 to November 2023, employing a linear regression model to quantify their relationship. 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

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

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.

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 linear regression analysis to dissect the relationship between the economic downturns—as represented by business bankruptcies—and the demand on food banks (Table 3). 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

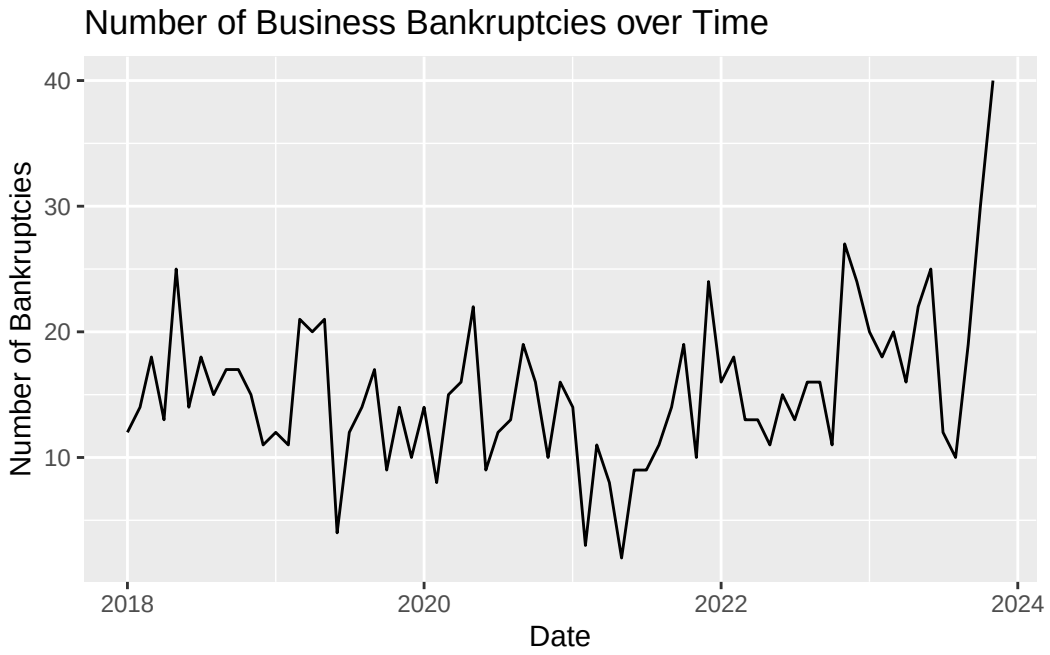

Figure 1: Food Bank Usage across time from 2018 to 2023

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.

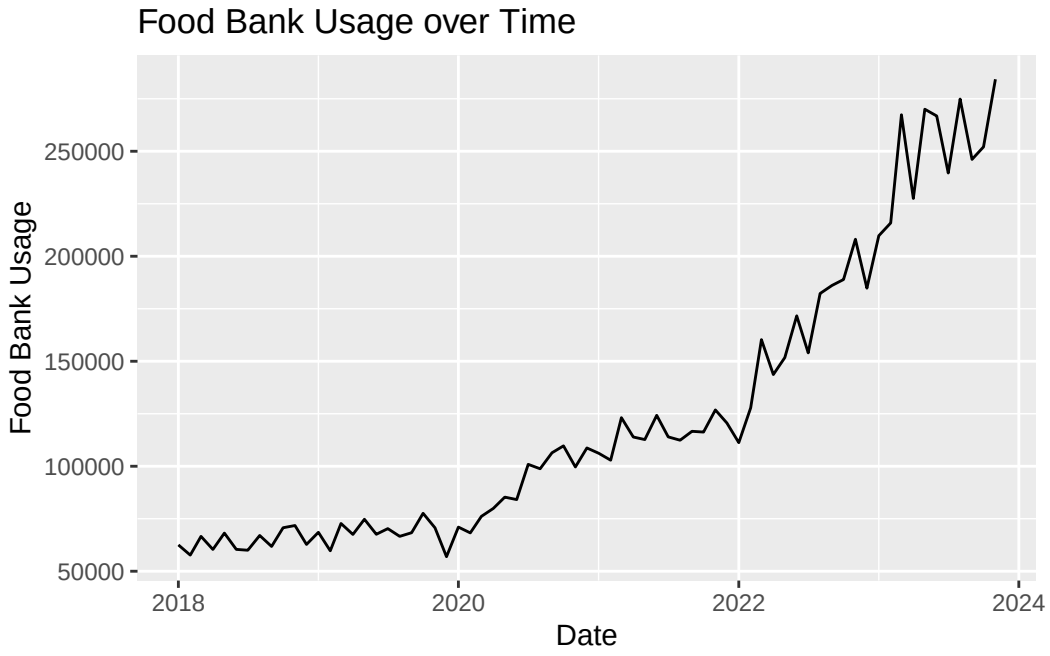

Figure 2: Number of Business Bankruptcies across time from 2018 to 2023

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.

Results

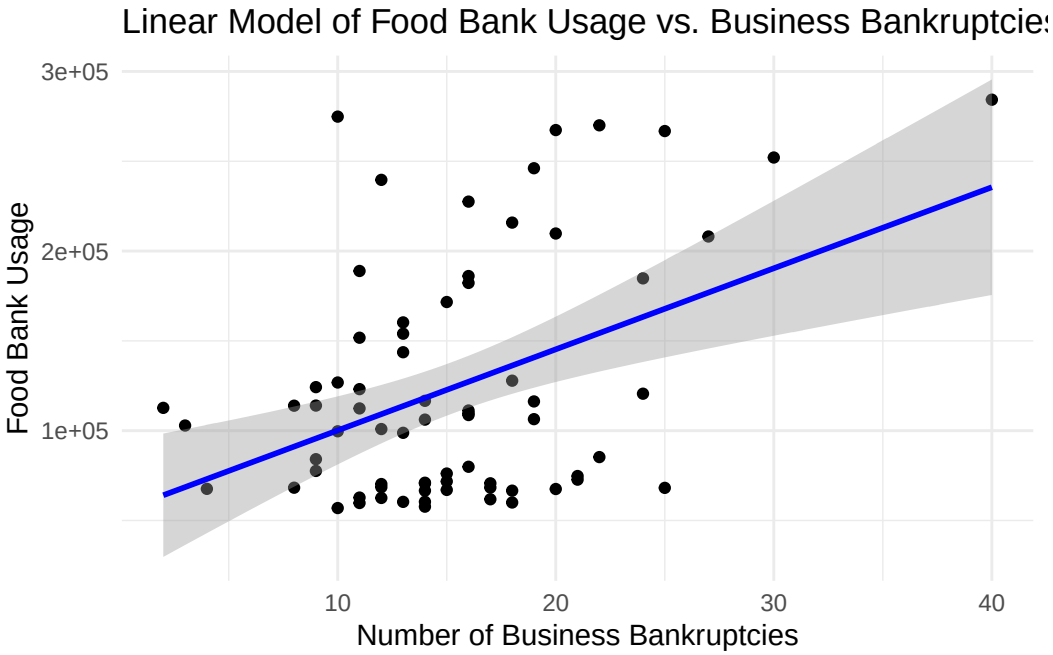

Figure 3: Linear Model of Food Bank Usage vs. Business Bankruptcies

Figure 3 is a scatter plot with a fitted linear regression line, depicting the relationship between the number of business bankruptcies and food bank usage. The x-axis represents the number of business bankruptcies, and the y-axis represents food bank usage, which appears to be in a logarithmic scale given the notation (1e+05 indicates scientific notation, commonly used for larger numbers). Each dot on the plot represents an observation of the data — a specific instance where the number of business bankruptcies corresponds to a certain level of food bank usage. The blue line represents the linear regression fit, suggesting a trend or relationship between the two variables. The shaded area around the blue line is the confidence interval, which gives an indication of the certainty of the regression estimate. From the graph, as the number of business bankruptcies increases, there appears to be a general increase in food bank usage. This suggests a positive correlation between these two variables, indicating that higher instances of business bankruptcies might be associated with increased food bank usage within the City of Toronto during the observed period. The confidence interval's width suggests variability in the data, with some points falling far from the line, indicating potential outliers or a wide range of food bank usage numbers for similar counts of business bankruptcies.

	Estimate	StdError	tValue	Pr
(Intercept)	55029.305	19382.346	2.839146	0.0059377190
num_of_bankruptcies	4513.056	1179.787	3.825313	0.0002834629

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

Table 1 presents the results from a linear regression analysis using the `lm` function in R, where `Value.x` is the dependent variable (presumably Food Bank Usage), and `Value.y` is the independent variable (presumably Number of Business Bankruptcies).

Here's a breakdown of the results:

Intercept: The estimated intercept is 55,029 with a standard error of 19,382, meaning when the Number of Business Bankruptcies is zero, the Food Bank Usage is expected to be around 55,029. The t-value is 2.839, and it is statistically significant at the 0.01 level (p-value = 0.005938).

Value.y: The coefficient for the Number of Business Bankruptcies is 4,513, with a standard error of 1,180. This suggests that for each additional business bankruptcy, there is an average increase of about 4,513 in the Food Bank Usage. This coefficient is statistically significant at the 0.001 level (p-value = 0.000283).

Residuals: The residuals show the difference between the observed Food Bank Usage values and those predicted by the model, ranging from a minimum of -99,698 to a maximum of 174,691.

R-squared: The R-squared value of 0.175 indicates that approximately 17.5% of the variability in Food Bank Usage is explained by the model. The Adjusted R-squared, which adjusts for the number of predictors in the model, is slightly lower at 0.163, suggesting that when accounting for the number of variables, the model explains about 16.3% of the variance.

F-statistic: The F-statistic is 14.63 with a p-value of 0.0002835, indicating that the model as a whole is statistically significant and that there is a relationship between the variables.

It indicate that for every additional business bankruptcy in Toronto, food bank usage increases by an average of 4,513, with the model accounting for approximately 17.5% of the variability in food bank usage. The significant p-values for both the intercept and the slope ($p < 0.01$ and $p < 0.001$, respectively) confirm the statistical significance of these results. The intercept value of 55,029 suggests the baseline food bank usage when no bankruptcies occur. While the relationship between business bankruptcies and food bank usage is statistically significant, the R-squared value suggests that other unexamined factors may also play a significant role in determining food bank usage.

Model

Model Setup

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

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.
- 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.
- 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**: This would typically include additional variables that could influence food bank usage, such as unemployment rates, economic indicators like GDP or inflation, policy changes, or even weather events that are not part of the main analysis. These controls are important for isolating the effect of business bankruptcies on food bank usage from other confounding factors. However, based on the equation provided, it appears that this particular model does not include additional control variables beyond the primary independent variable of business bankruptcies.
- **Comparison Group**: The comparison group, or the group not treated, would be a hypothetical scenario where the number of business bankruptcies is zero. Since this is a simple regression without explicit treatment or control groups, the 'treatment' in this context refers to the occurrence of business bankruptcies. By setting the number of bankruptcies to zero, the intercept β_0 provides a baseline against which increases in bankruptcies (the 'treatment' effect) can be compared. In other words, the model assesses how food bank usage would differ with bankruptcies as compared to this baseline scenario without them.

Features

Our model employs a linear regression framework designed to quantify the impact of business bankruptcies on food bank usage within Toronto. The primary feature of our model is its direct focus on the variable "num_of_bankruptcies" as a predictor for "food_bank_usage," reflecting a hypothesis that economic distress may translate to increased demands on social support services. The simplicity of the model—using a single independent variable—ensures clear interpretability of results and straightforward statistical analysis. We also employ the "Date" variable to capture temporal dynamics, allowing for an understanding of how this relationship unfolds over time. This feature is particularly valuable for identifying trends and potential seasonal patterns that may influence food bank usage independently of bankruptcies.

Model concerns

Despite its advantages, the model raises several concerns regarding its explanatory capacity and the potential for omitted variable bias. The most pressing concern is the modest R-squared value, which indicates that a significant portion of the variation in food bank usage is left unexplained by business bankruptcies alone. This suggests that there are other factors at play not captured in the model—such as changes in social policies, fluctuations in the job market, or demographic shifts—that may also influence food bank usage. The absence of control variables to account for these factors could result in a misestimation of the effect of business bankruptcies. Additionally, the assumption of linearity may oversimplify the complex relationship between economic health and social welfare. The potential influence of outliers and extreme values also poses a risk to the model's reliability, as does the assumption that the past relationship between these variables will continue unchanged into the future. For more robust and nuanced insights, future iterations of this project would benefit from incorporating additional relevant variables, employing different model specifications, or using alternative statistical techniques that can account for the multifaceted nature of economic and social interdependencies.

Discussion

Findings

Our analysis revealed that from January 2018 to November 2023, food bank usage in Toronto showed an increasing trend with fluctuations that potentially correspond to seasonal variations or economic events. The data indicated that as business bankruptcies rose, there was a concomitant increase in food bank usage, suggesting that economic downturns tangibly affect community welfare. The regression model presented a statistically significant relationship between the number of business bankruptcies and food bank usage (Table 1). With each

additional business bankruptcy, food bank usage increased by an estimated average of 4,513 instances, as reflected by the slope coefficient (β_1) in the linear regression equation. However, the moderate R-squared value of 0.175 signified that while business bankruptcies were a significant predictor, they were not the sole factor affecting food bank usage, prompting consideration of other external influences not captured in this model. Notably, the model exhibited limitations, such as the potential for omitted variable bias, where other unmeasured factors could be influencing the relationship. For example, changes in employment rates, social welfare policies, and the cost of living may also play critical roles. The wide range of residuals suggests that the model might not fully capture the variability in food bank usage, and outliers observed in the data hint at the existence of extreme cases that could be disproportionately affecting the results. These findings have implications for policymakers and social service providers. The clear association between economic distress signals, as measured by business bankruptcies, and increased food bank demand underscores the importance of robust support systems to buffer against economic shocks.

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.