Northeastern University

Final Capstone Project Deliverable

Analysing Pathways Out of Poverty: A Data-Driven Exploration with YSM

Professor: Dr. Ganesh Subramanian

ALY6980: Capstone

Term: Winter 2025

GROUP 4

Aditya Raj Singh,

Yasaswi Manthena,

Denish Sureshbhai Borad,

Kevin Tushar Pandya,

Punith Matta

Executive Summary

This capstone project, conducted in collaboration with Yonge Street Mission (YSM), focuses on understanding how YSM's Workforce Development (WFD) programs contribute to the journey out of poverty. Using TIMES (Tracking Individual Measures of Economic Stability) assessment scores, participant demographics, and program engagement data, we developed two key dashboards and several analyses to identify the most impactful services and indicators.

Our findings indicate that quality, personalized programming - specifically in Employment Assistance and Goal-Setting - significantly drives improvements in TIMES scores. Participation trends also revealed shifts in engagement post-2022, warranting further strategic outreach.

Table of Contents

Introduction

The Yonge Street Mission (YSM) is a community-based organization committed to transforming the lives of individuals experiencing poverty in Toronto. Through its Workforce Development (WFD) programs, YSM provides a range of services aimed at helping participants build employment readiness, economic stability, and long-term independence.

As part of this capstone project, our team partnered with YSM to conduct a comprehensive data-driven analysis of its Workforce Development initiatives. By leveraging TIMES assessment scores, demographic insights, service delivery records, and program participation data, our goal was to uncover actionable insights into the effectiveness of YSM's programming.

This report presents the findings of our exploratory data analysis (EDA), supported by interactive Power BI dashboards and statistical techniques. The insights aim to inform YSM's strategic decisions, improve program delivery, and ultimately enhance outcomes for the communities it serves. The analysis focuses on uncovering the relationships between program engagement, key psychosocial indicators (such as self-awareness and sense of power), and improvements in TIMES scores.

Through this initiative, we seek to answer critical questions about which program components deliver the most value, how different services influence participant outcomes, and how demographic factors intersect with program effectiveness. The results of our work are intended to support YSM in refining its approach to building sustainable pathways out of poverty.

YSM's Mission and Project Objective

YSM is a Toronto-based nonprofit supporting people living in poverty through holistic community services. The objective of this project was to:

- Analyse TIMES scores to uncover service effectiveness
- Track program participation trends over time
- Identify factors that best predict participant success
- Build scalable dashboards for ongoing performance monitoring

Business Questions

The project centres around five core questions:

1. What is the effect of engaging in multiple programs on TIMES scores?
2. How does the duration of participation in WFD programs correlate with TIMES scores?

3. How do indicators such as Sense of Power, Self-Awareness, Relationships, and Values influence TIMES scores?
4. Do goal-setting programs outperform non-goal-setting programs in terms of TIMES score impact?
5. Which program components best improve Employment Readiness and Economic Stability?

Methodology

The project followed an agile, cloud-integrated process consisting of:

- **Data Cleaning**: Handling missing values, outliers, and standardizing schemas.
- **Exploratory Data Analysis (EDA)**: Examining patterns, trends, and clusters.
- **Data Visualization**: Creating dashboards and visual charts using Power BI.
- **Cloud Workflow**: Using Google Cloud and BigQuery for scalable, real-time analytics.

Tools Used

- **Data Integration**: Python, SQL
- **Visualization**: Power BI
- **Cloud**: Google Cloud, BigQuery
- **Project Management**: Jira, Agile Sprint Planning

Data Cleaning

Effective data analysis begins with robust cleaning and preprocessing. The raw datasets from YSM contained a mix of structured demographic information, program participation logs, and TIMES assessment scores, each with inconsistencies and missing entries. Below are the key steps undertaken to ensure data quality and integrity before moving into analysis and dashboard creation:

1. Merging and Schema Validation

- Multiple datasets - including program initiations/terminations, service deliveries, demographics, and TIMES scores - were joined using unique participant IDs.
- Before merging, each dataset's schema was validated to ensure consistent column naming, date formats, and key structures.

2. Standardizing Column Names and Data Types

- All column headers were converted to lowercase, and spaces were replaced with underscores for consistency.

- Dates were converted into proper datetime objects (e.g., birth date, service date, program start date).
- Categorical variables such as gender, ethnicity, and program names were converted to uniform string formats.

3. Handling Missing Values

- **Categorical Columns** (e.g., *income source*, *housing type*, *ethnicity*, *gender*):
    - Missing values were filled with "Unknown" to retain the records while still indicating a lack of response.
- **Numerical Columns**:
    - For columns such as age and household_size, missing values were imputed using median values within relevant subgroups (e.g., program or gender).
    - This approach helped maintain statistical consistency without skewing the data.

4. Cleaning Specific Fields

- **Age** was calculated from the date of birth and verified to ensure it was within reasonable bounds (e.g., removing records with age < 15 or > 70 if outliers).
- **Household size** entries of 0 or unusually high numbers were investigated and replaced with median values.
- **TIMES scores** with zeros or extreme outliers were dropped or capped based on the 1st and 99th percentiles.

5. Removing Duplicates

- Duplicate participant entries, especially across service records and program initiations, were identified using combinations of participant_id, program_name, and start_date.
- Only the most recent or complete record was retained per participant.

6. Standardizing Program and Service Names

- Program names were cleaned to ensure consistency — e.g., "Youth Job Connection - Summer" and "Youth Job Connection Summer" were unified.
- Similarly, variations in service names were normalized (e.g., “Job Coaching”, “Coaching - Job”, and “Job-Coaching” were standardized to a single label).

7. Dropping Irrelevant or Noisy Columns

- Highly sparse columns such as Indigenous identity, religion, and other fields with >50% missing data were dropped to reduce noise.

- Columns unrelated to analysis objectives (e.g., internal admin IDs, unused flags) were also removed.

8. Feature Engineering

- New features such as age group, income category, and program duration were created to support grouping and time-based analysis.
- A goal_setting flag was introduced to distinguish goal-setting programs from non-goal-setting ones.

Exploratory Data Analysis (EDA)

To understand the landscape of program participation, service utilization, and assessment outcomes within YSM's Workforce Development programs, we performed a comprehensive exploratory data analysis. This analysis helped identify trends, outliers, and meaningful correlations that guided our dashboard design and final recommendations.

Overview of EDA Conducted:

- **Missing Value Inspection and Imputation**: Checked and filled missing values using context-appropriate methods.
- **Data Distributions**: Explored distributions for numerical fields like age, TIMES scores, and household income.
- **Service Frequency Analysis**: Identified the most and least commonly delivered services.
- **Clustering Analysis**: Used unsupervised learning (KMeans with PCA) to segment participants based on service usage behavior.
- **Correlation Matrix**: Analyzed relationships between TIMES assessment indicators to understand interdependencies.
- **Outlier Detection**: Applied log scaling and visualization techniques to detect skewed service quantity distributions.
- **Feature Engineering**: Created derived features such as age groups, participation duration, and program flags (goal-setting vs. non-goal-setting).

EDA revealed:

1. Clustering Participants Based on Service Usage

We applied KMeans clustering (with PCA dimensionality reduction) to group participants based on their service usage patterns.

Insight:

- Three distinct clusters were identified, suggesting different types of participants - for example, Cluster 0 (purple) represents high-service users, while Cluster 2 (yellow) include lower-engagement individuals.
- This helps tailor services for different user profiles and optimize resource allocation.

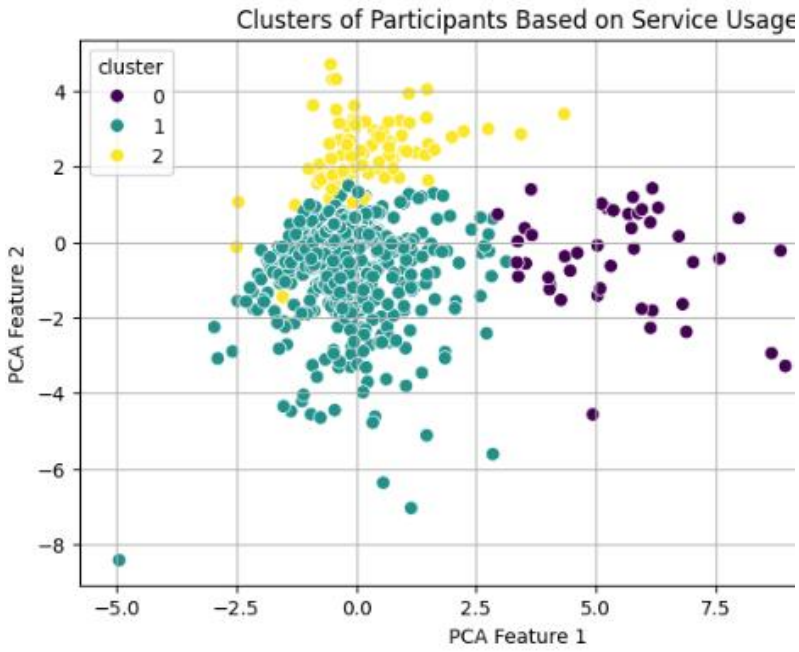


Figure (1)

2. Distribution of Service Quantities by Unit

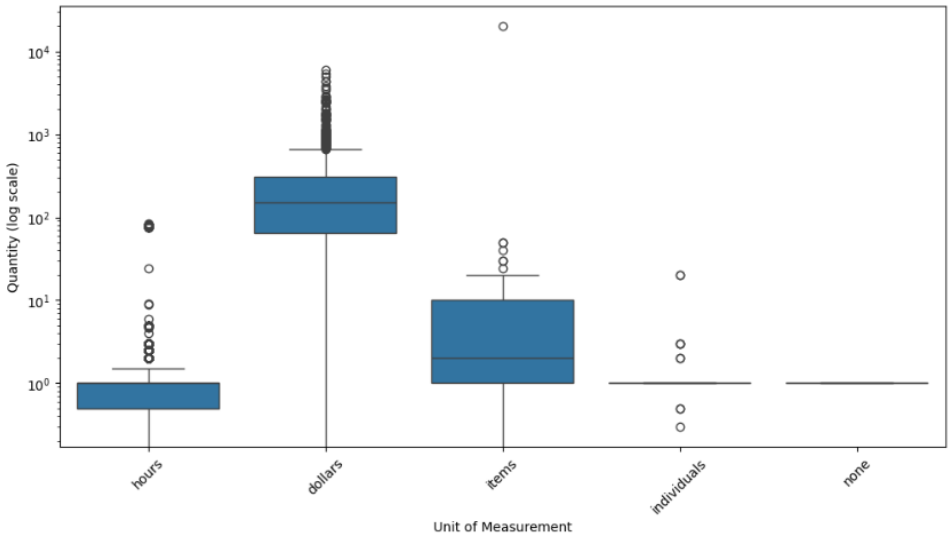


Figure (2)

This boxplot shows how different types of service units (e.g., hours, dollars, items) vary in quantity. The y-axis is log-scaled to account for wide variations.

Insight:

- Services measured in dollars show the highest variability and skewness, indicating large differences in financial support.
- Service units like “individuals” or “none” are consistently low and stable.

3. Top 10 Most Frequent Services Delivered

This bar chart reveals the top 10 services delivered by frequency.

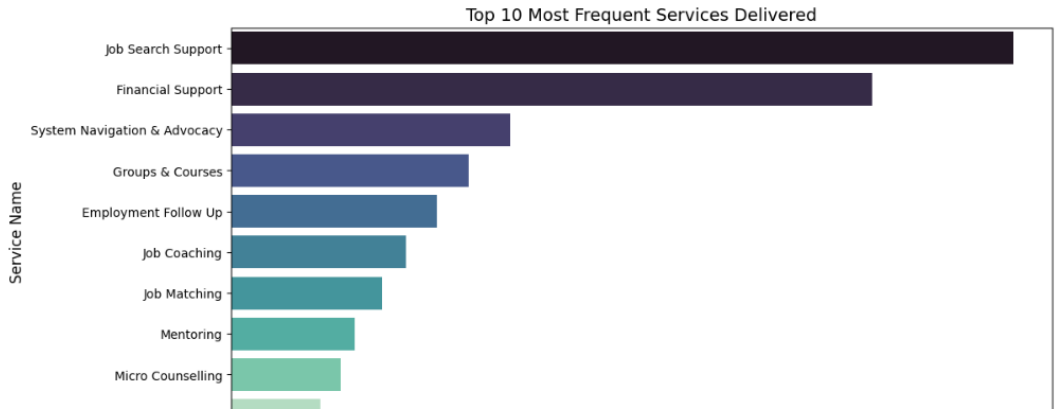


Figure (3)

Insight:

- Job Search Support, Financial Support, and System Navigation & Advocacy were the most frequently delivered services.
- These services may represent core components of WFD program offerings and have significant potential to influence TIMES scores.

4. Correlation Heatmap of Assessment Indicators

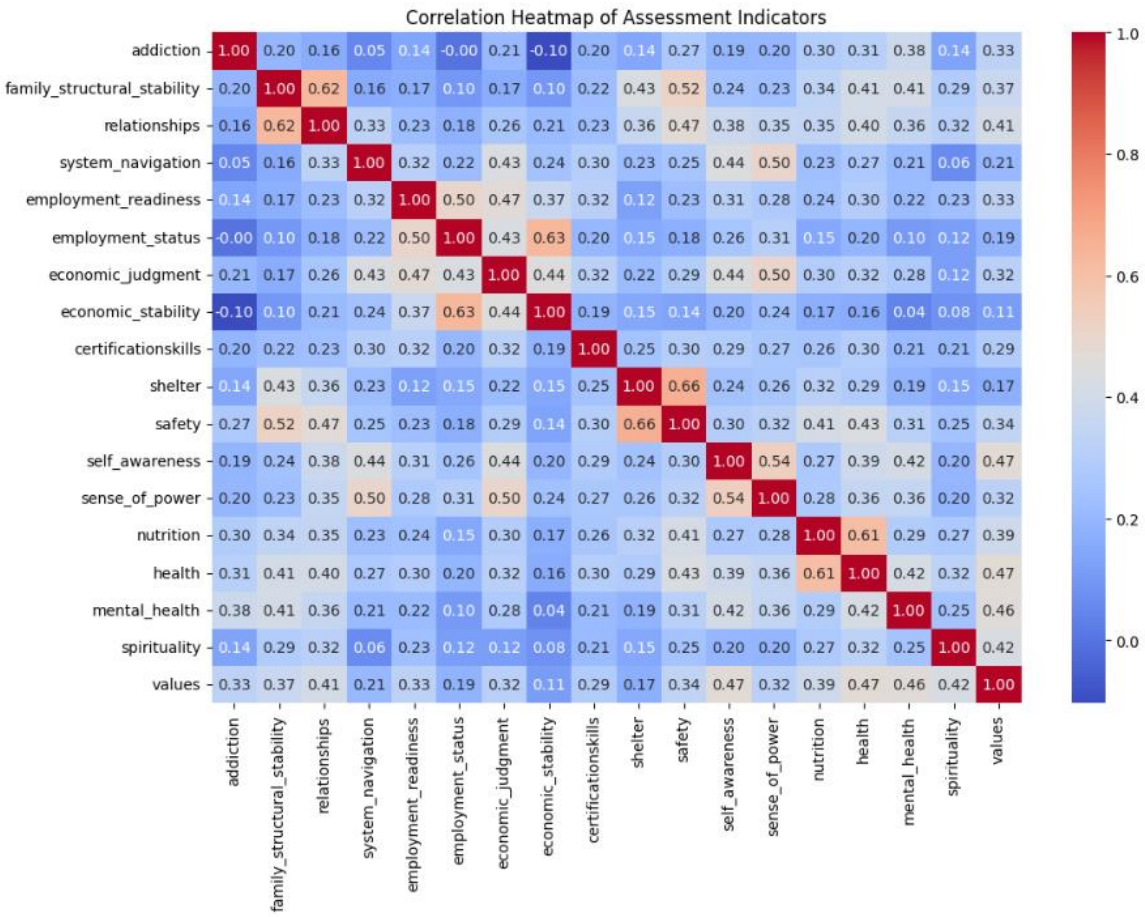


Figure (4)

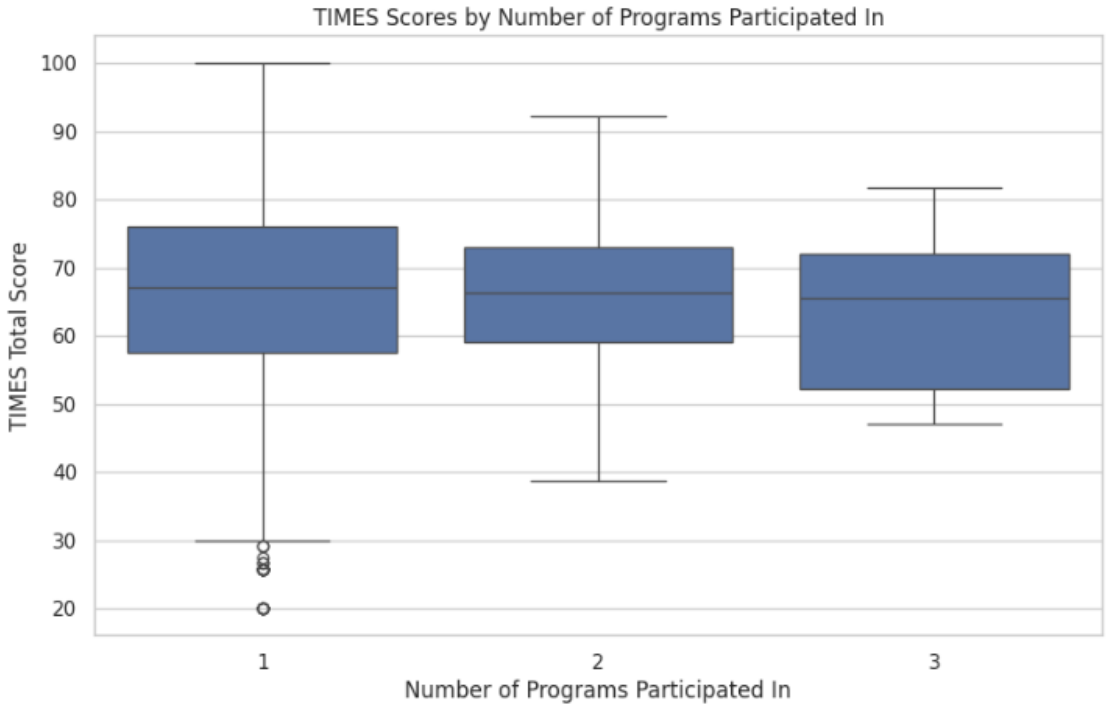
This heatmap visualizes the correlations between all TIMES assessment indicators.

Insight:

- Strong positive correlations exist between related indicators:
    - Self-Awareness, Sense of Power, and Values are interlinked.
    - Employment Readiness is moderately correlated with Employment Status, Economic Stability, and Judgment.
- Unexpected low correlations (e.g., between Addiction and most other indicators) suggest these may require isolated or focused interventions.

Addressing the Business Questions

1. **Effect of Participating in Multiple Programs:**
   Participants enrolled in one program performed better than those in two or three programs. Multiple enrollments may dilute focus or indicate fragmented service pathways.



	program_count	scaled_times_score
0	1	66.526829
1	2	65.970227
2	3	63.427287

Figure (5)

2. **Duration and TIMES Scores Correlation:**
   Programs with longer participation durations showed higher TIMES score "jumps." For example, Employment Assisted Services demonstrated significant improvements with optimal engagement around 5 months.

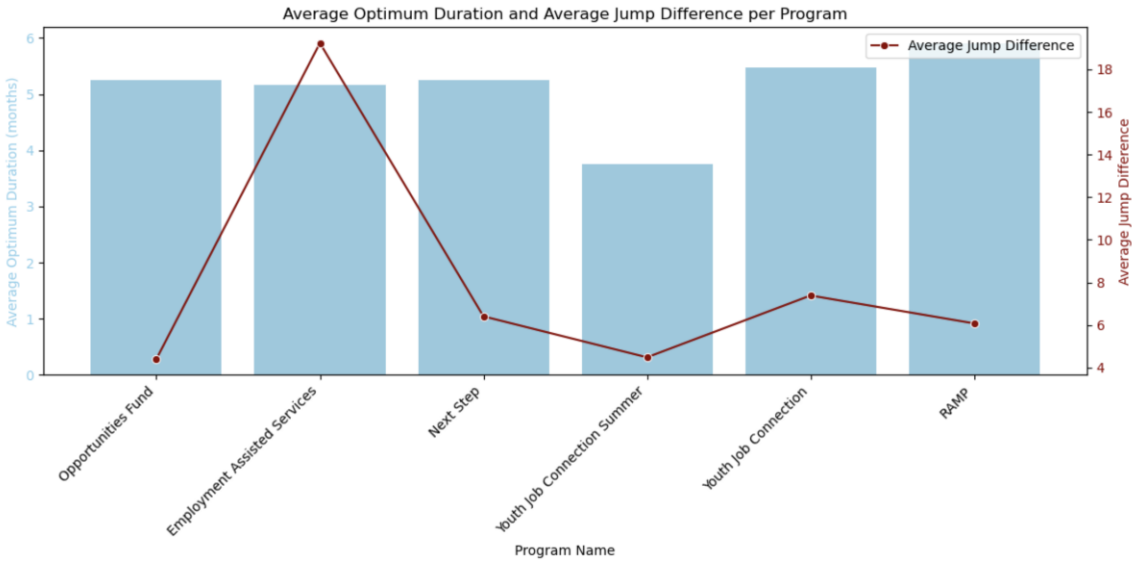


Figure (6)

3. **Impact of Key Indicators:**
   Programs that emphasize self-awareness, values, and relationships showed higher average outcomes. These “soft skills” are critical for long-term transformation.

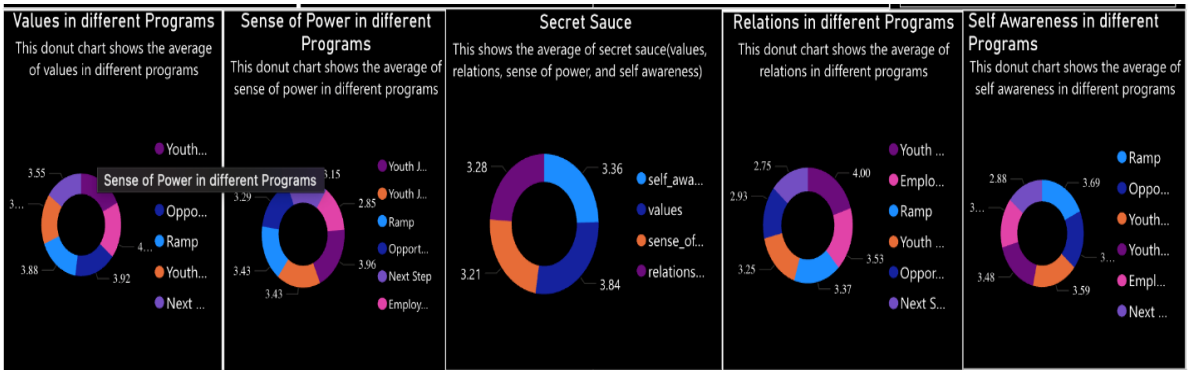


Figure (7)

4. **Goal vs. Non-Goal Setting Programs:**
   Goal-setting programs like Youth Job Connection and Ramp outperformed others, indicating the value of structured progress tracking.

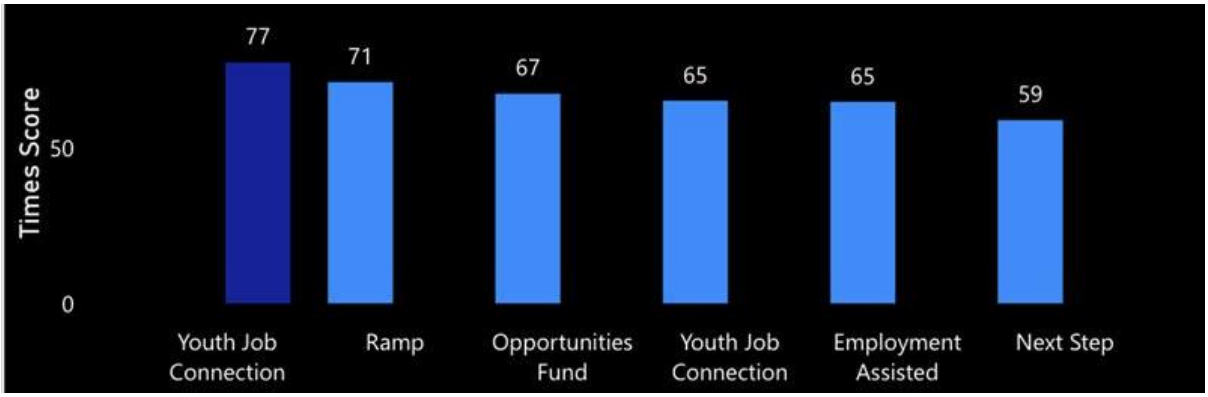


Figure (8)

5. Program Components Driving Readiness and Stability:

- **Ramp** led in Employment Readiness.

- **Workforce Development Training** led in Economic Stability. These programs offer practical training and financial empowerment.

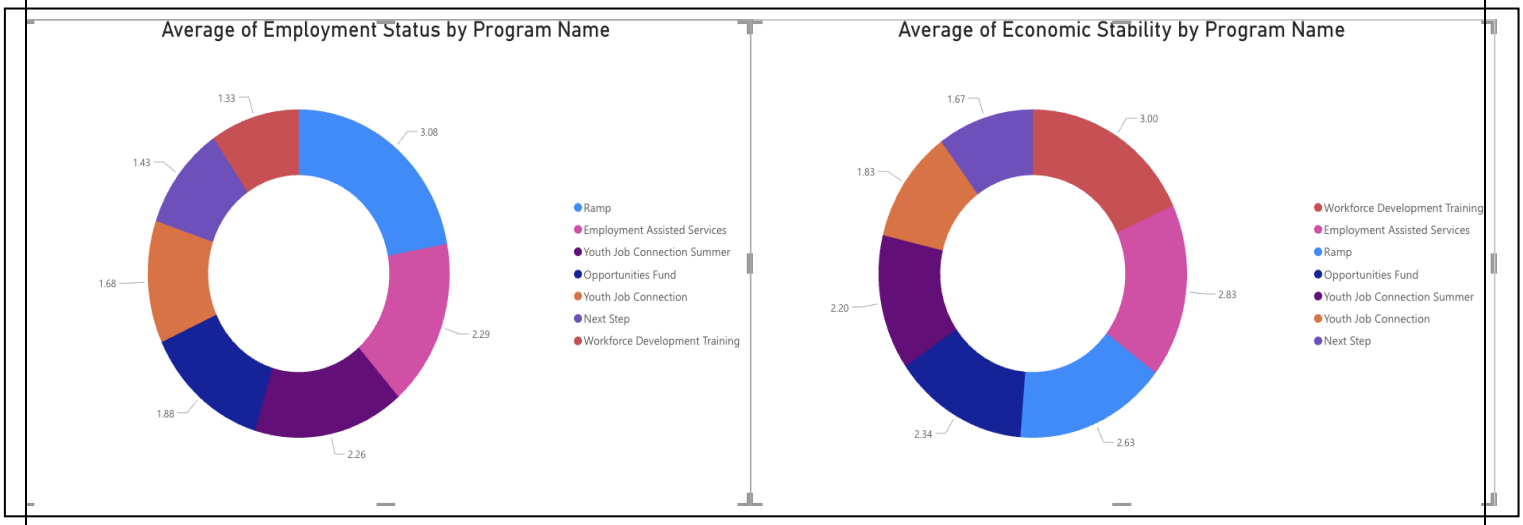


Figure (9)

Dashboards

Dashboard 1: Program and Service Impact Overview

This dashboard visualizes how different WFD programs perform across TIMES assessment indicators like Values, Sense of Power, Relations, and Self-Awareness. It also highlights the top services contributing to higher TIMES scores and compares program types (goal-setting vs non-goal-setting) based on their impact.

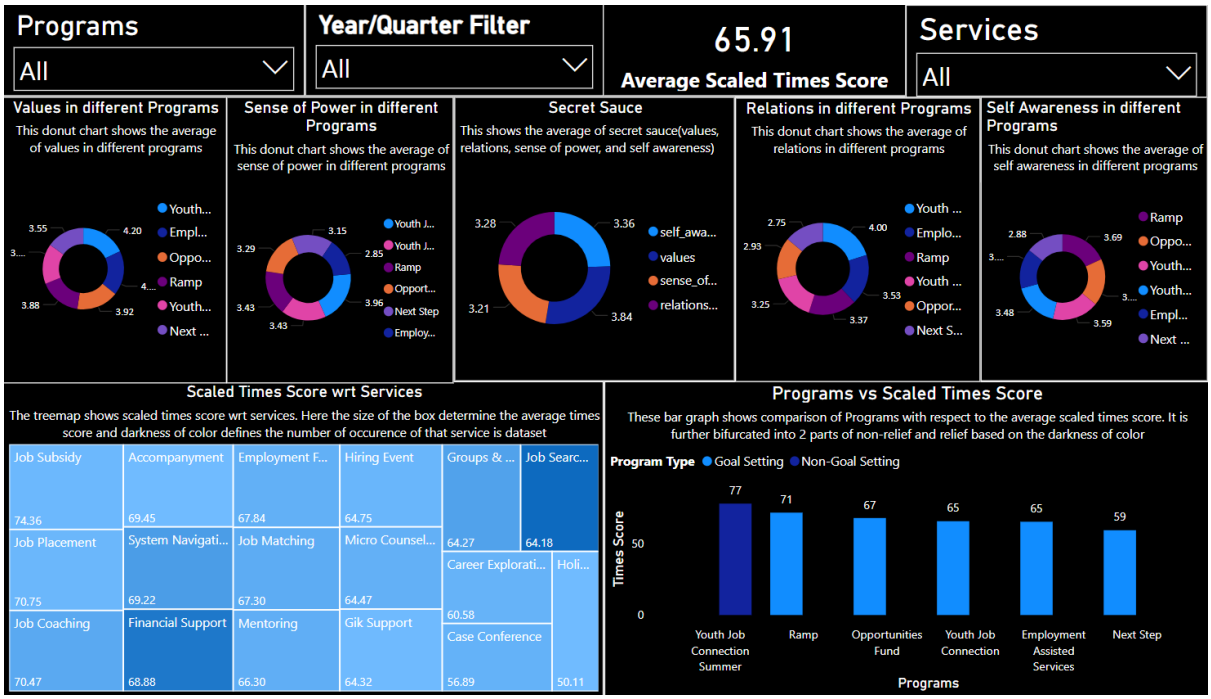


Figure (10)

Dashboard 2: Participant Demographics & Engagement

This dashboard explores participant demographics such as gender, age, income, and country of origin, alongside program-specific insights like family structure stability and certifications achieved. It helps identify who is being served and how demographic factors relate to program outcomes.

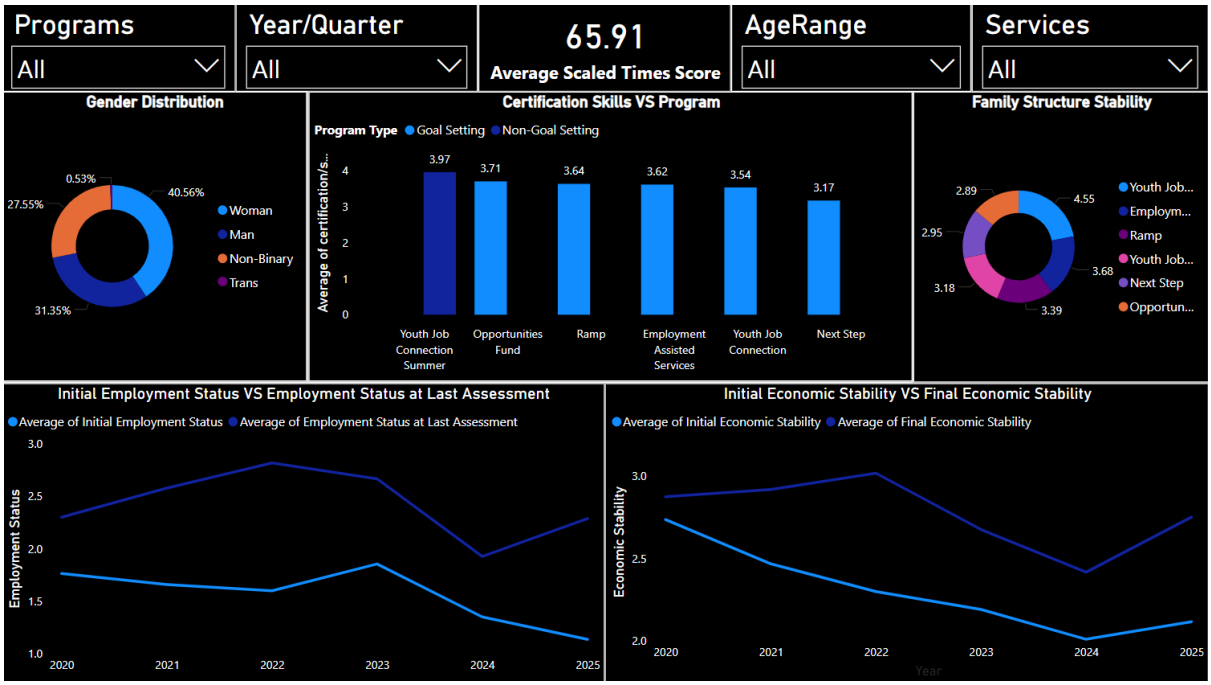


Figure (11)

The Power BI dashboards revealed:

- Majority participants are **women (46.78%)**, aged **20–30 years (44.55%)**.
- Over **52% were born in Canada**, while 22% were newcomers.
- **Low-income and unknown income backgrounds** dominate.
- **Certification skills** are highest in **goal-setting programs**.
- **Enrolment peaked in 2022** but saw a sharp drop in 2024.

Scalability & Infrastructure

Utilizing Google Cloud and BigQuery ensured:

- Scalability for growing data.
- Real-time analysis capabilities.
- Cost efficiency and futureproofing.

End-to-End Workflow

Our team followed a structured data journey:

1. Project Setup
2. Data Acquisition & Cleaning
3. EDA
4. Advanced Insights
5. Results Presentation
6. Final Delivery

Key Findings

1. **High-Impact Programs Identified**
    - Programs like Ramp and Youth Job Connection Summer consistently show higher TIMES scores, especially in areas like self-awareness and sense of power.
2. **Goal-Setting Programs Outperform Others**
    - Participants in goal-setting programs show significantly higher improvement across TIMES assessment indicators compared to non-goal-setting programs.
3. **Services Drive Score Improvement**
    - Services such as Job Subsidy, Job Coaching, Financial Support, and Employment Follow-up are highly correlated with increased TIMES scores.
4. **Demographics Influence Outcomes**
    - The largest participant age group is 20–30 years, and over 52% were born in Canada, suggesting programs can be tailored to meet the needs of younger Canadian-born individuals.
5. **Unequal Service Utilization**
    - A small set of services account for most of the service usage - highlighting the need to assess underutilized services for relevance or outreach gaps.
6. **Drop in Participation in 2024**
    - Program enrollment declined significantly in 2024, raising concerns that require further investigation and a possible review of outreach efforts.
7. **Cluster Analysis Shows Distinct Engagement Patterns**
    - Clustering participants based on service usage revealed three main groups, enabling YSM to tailor interventions to different participant profiles.

Key Takeaways

1. **Focus on Quality, Not Quantity**
    - Encouraging participants to enroll in multiple programs doesn't necessarily lead to better outcomes; focusing on high-impact programs is more effective.
2. **Strengthen and Scale Proven Services**
    - Programs should invest more in expanding Job Coaching, Financial Support, and Job Search Services, as these directly correlate with better outcomes.
3. **Use TIMES Scores as an Ongoing Monitoring Tool**
    - Regularly evaluating TIMES assessments can help track participant progress and adjust interventions in real-time.
4. **Personalize Outreach and Program Design**
    - Use clustering and demographic insights to tailor services for participant needs and increase engagement with underrepresented or low-scoring groups.
5. **Investigate Participation Drop**
    - The 2024 decline in program involvement suggests the need for enhanced engagement strategies, feedback collection, or post-pandemic program adjustments.

Conclusion

This capstone project provided a data-driven lens into the effectiveness of YSM's Workforce Development (WFD) programs, leveraging TIMES assessments, participant demographics, and service delivery records to uncover patterns that can guide strategic improvements. Through rigorous data cleaning, exploratory analysis, clustering, and visualization, we identified the key programs and services that most significantly impact participants' self-sufficiency and employment readiness.

Our findings affirm that goal-setting programs like Ramp and Youth Job Connection Summer, as well as services such as Job Coaching and Financial Support, deliver the highest value in improving TIMES scores. At the same time, the drop in participation during 2024 and the heavy reliance on a few core services point to opportunities for expanded outreach and service diversification.

As YSM continues its mission to help individuals break the cycle of poverty, we recommend adopting a more targeted, data-informed approach to program design and delivery. By regularly monitoring TIMES outcomes, focusing on high-impact interventions, and tailoring services to diverse participant profiles, YSM can amplify its positive impact and drive sustainable change in the community.

This project underscores the power of data in advancing social good - and we hope these insights support YSM in creating even more meaningful pathways out of poverty.

References

- Ridley, M., Rao, G., Schilbach, F., & Patel, V. (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. Science, 370(6522), eaay0214. https://doi.org/10.1126/science.aay0214
- Yonge Street Mission. (2023). TIMES™: Tracking Individual Measures of Economic Stability. Internal Documentation.
- OECD (2021). “Skills for Jobs: Workforce Development Policies.” https://www.oecd.org/
  Supports the importance of tailored training and workforce strategies in achieving sustainable employment.
- National Skills Coalition (2020). “The Roadmap for Racial Equity: An Imperative for Workforce Development Advocates.” https://nationalskillscoalition.org