

# Understanding Member Engagement at Wellspring:

Key Drivers of Attendance and Service Utilization

Carrie Lam, Danni Luo, Xinyue Pu, Xiaotong Shen |  
TUT0202-B

# Introduction:

## **Wellspring:**

- ▶ a Canadian charity network providing free cancer support programs.

## **Population:**

- ▶ All potential and current members of Wellspring in Canada, including cancer patients and their families.

## **Audience:**

- ▶ Wellspring staff, who are experts in cancer support services but may not have formal training in statistics.

## Research Questions:

- 1. How do age and program interests influence average monthly attendance?**
- 2. Does POC (Person of Color) status impact a member's total amount of service attended?**
- 3. Can registration date, mailing province, and cancer risk predict attendance in the past three months?**

## Question 1: Age, Program Interests, and Average Monthly Attendance

This question can help Wellspring to understand the age and number of interests of the population that needs to be focused on improving attendance numbers to more effectively plan Wellspring's future programs.

Variables Used:

- ▶ **Age** – Member's age in years.
- ▶ **Program Interests** – Number of programs the member expressed interest in.
- ▶ **Services Attended** – Total number of Wellspring sessions each member participated in.
- ▶ **Membership Duration** – Calculated from join date (year & month) to March 2024.

## Data Wrangling:

Filtered members with valid age data ( $\geq 0$ );

Categorizing members based on their number of program interests:

- ▶ 1 program of interest
- ▶ 2-4 programs of interest
- ▶ 5-7 programs of interest
- ▶ No program of interest

Calculated monthly attendance as

$$\frac{\text{Number of Services Attended}}{(2024 - \text{Start Year}) \times 12 + (12 - \text{Start Month} + 1)}$$

## Visualization:

Table 1: Member Distribution by Program Interest

Number of Program Interests	Number of Members
1 program of interest	101
2-4 programs of interest	162
5-7 programs of interest	94
no program of interest	2492

Prediction of member's average attendance/month with age and number of program interest as

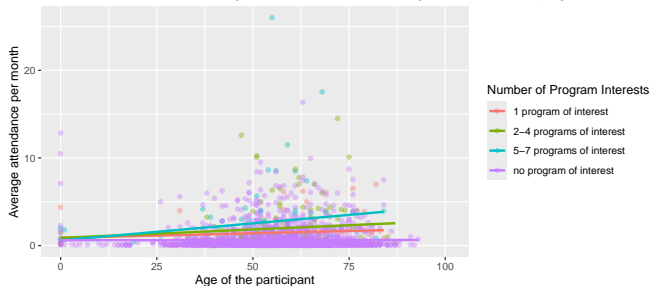


Figure 1

## Conclusion: 1. Age, Program Interests, and Average Monthly Attendance

### 1. **Age and Attendance:**

Older participants between approximately 30-80 years old tend to have higher average attendance rates, while younger participants tend to have a lower attendance rate.

### 2. **Number of Program Interests and Attendance:**

Participants with 5-7 program interests have the highest average attendance per month, while those with no program interest have the lowest, suggesting that limited interest reduces participation.

### **Wellspring could:**

Develop targeted programs for younger members to increase their engagement; Promoting cross-program participation to encourage multiple program exploration.

## Question 2: Impact of POC Status on Engagement

Understanding differences in program attendance between POC and non-POC members helps to improve engagement by ensuring fairness.

### Variables Used:

- ▶ member id: Unique ID of each member
- ▶ POC status: Whether identifies self as a Person of Color
- ▶ attendance status: Attendance records.

### Hypothesis Test:

- ▶  $H_0$ : POC status does not affect attendance.
- ▶  $H_A$ : POC status affect attendance.
- ▶ p-value: the likelihood of your data occurring under  $H_0$ .

( $p \geq 0.05$  means fail to reject  $H_0$ )

### Data Wrangling:

- ▶ Calculate total attendance by grouping member id.
- ▶ Joined attendance data with POC status.
- ▶ Filtered out members with missing POC status.



# Visualization(Diagram 1)

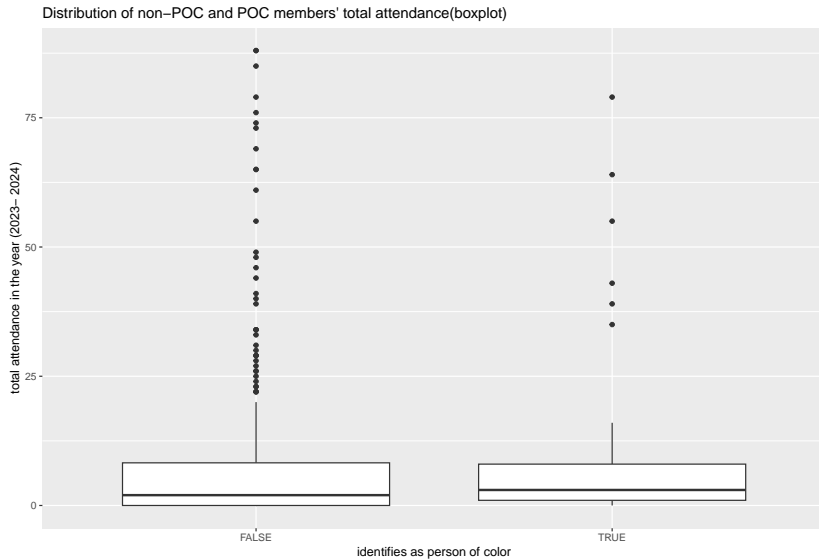


Figure 2

# Visualization(Diagram 2)

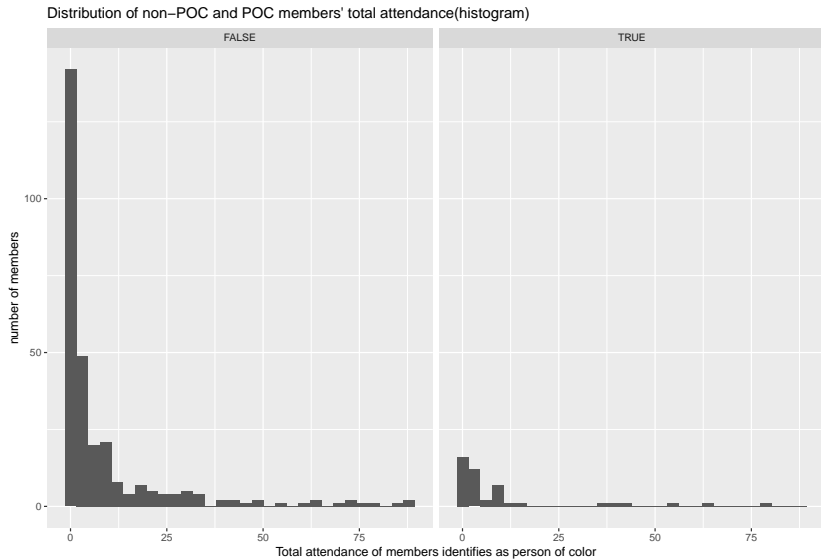


Figure 3

# Visualization(Tables)

Table 2: Table of different scenarios of inactive members in data wrangling

Scenario	People
people never sign up in 2023-2024	105
people sign up at least once but never showing up	5
all people that attend 0 event	110

Table 3: Summarise of our sample

Identified as POC	People	Mean
FALSE	296	10.63176
TRUE	46	12.34783

Table 4: Table of results

Metric	Value
Test statistic	1.716069
P value	0.636000

## Question 2 Conclusion

- ▶ No significant difference in attendance between members who identify as POC and those who do not.
- ▶ Attendance distribution was heavily skewed toward low values for both groups.
- ▶ Many members attend few or no events.

This highlights the need for:

1. Further investigation into barriers to engagement, such as accessibility or program relevance.
2. Targeted outreach to ensure all members feel welcome and supported.

# Question 3: Predicting Attendance Using Registration Date, Mailing Province, and Cancer Risk

## Goals:

- ▶ Finding trends in variable significance
- ▶ Using these variables to predict attendance/active status of an individual

## Applications:

- ▶ Adjust web-content or service depending of personal information variables.

## Variables Used:

- ▶ mailing province: Categorized as East or West Canada.
- ▶ registration date: The user's registration date
- ▶ cancer risk: Categorized as high or low risk.
- ▶ last service date before march 2024: Boolean indicating if the member attended a service in the past three months (predictor variable).

## Data Wrangling

Take a look at the observations. Number of Unique Cancer Types: 129. Number of Unique Mailing Provinces/States: 7

Wrangle: Province -> East/West, Cancer -> High Risk/Low Risk.

Table 5: Unique Types of Cancer

Type of Cancer
Breast
Brain
Leukemia
Ovarian
Other
Colorectal

Table 6: Unique Province Names

Province Names
British Columbia
Manitoba
Newfoundland
Nova Scotia
Ontario
Quebec

# Visualising Registration Distribution

Table 7: Membership Distribution

Registered Before 03/2024	Number of Users	Total Percentage
FALSE	1907	39.5
TRUE	2922	60.5

Distribution of Membership Start Dates

Red line indicates March 2024

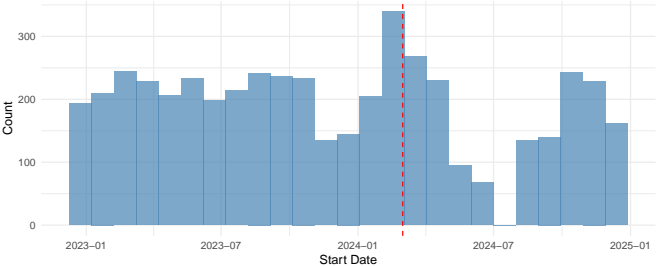


Figure 4

# Canadian Regions Distribution

Table 8: Canadian Region Distribution

Canadian Region	Number of Users	Total Percentage
East	3746	87.1
West	555	12.9

Distribution of Mailing Addresses Located in East/West Canada

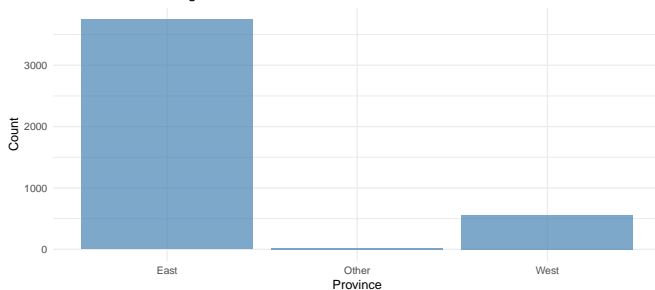


Figure 5



# Cancer Risk Distribution

Table 9: Cancer Risk Distribution

Level of Risk	Number of Users	Total Percentage
High Risk	1377	41.1
Low Risk	1973	58.9

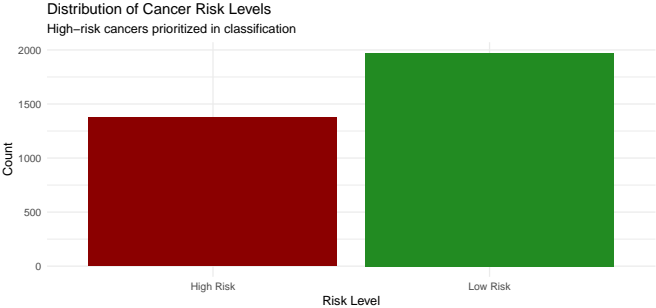


Figure 6

# Visualise Variable's Statistical Influence/Significance

Table 10: P-Values

Variable	P_Value
Cancer Risk	0.1484035
Region (Canada)	0.0008374
Before March 2024	0.0000000

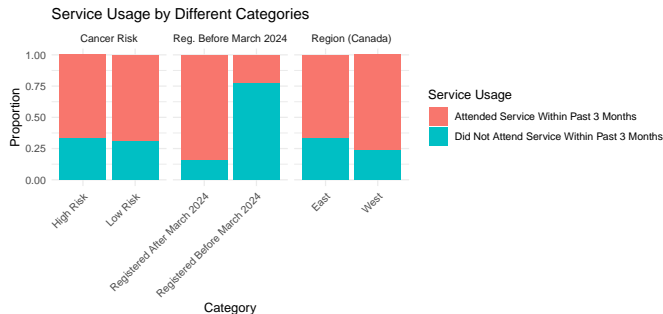
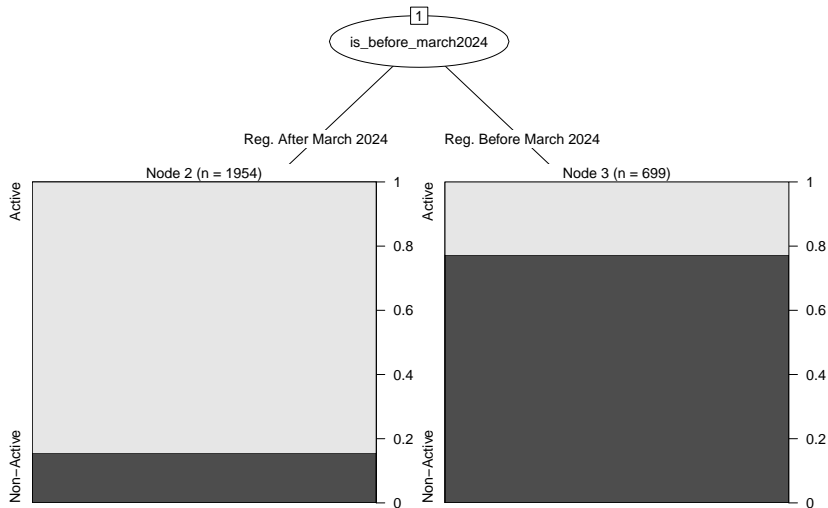


Figure 6

# Tree Model



# Tree Accuracy

Table 11: Confusion Matrix  
(Predicted/Actual)

	Active	Non-Active
Active	1649	305
Non-Active	159	540

Table 12: Accuracy Table

Metric	Percentage
Overall Accuracy	82.5
Sensitivity	77.3
Specificity	84.4

## Summary/Conclusion: Predicting Attendance Using Registration Date, Mailing Province, and Cancer Risk

Our variable significance analysis, in addition with the tree model suggests that:

- ▶ Whether or not a user registered prior or after March 2024 has the biggest impact on the user activity status (in accordance with their service attendance) by a large margin.

Interpretations:

- ▶ This means that the new registration system was largely successful, at least in regard to service attendance rate.

# Limitations(1.Data Limitations)

## 1. Limited Demographic Variables

- ▶ Misses Income, education, cultural background
- ▶ *Improvement:* Collect via surveys/registration forms

## 2. Self-Reported POC Status

- ▶ May not fully capture racial/ethnic identities
- ▶ *Improvement:* Add detailed categories

## 3. Cancer Type/Stage Excluded

- ▶ May impact attendance ability
- ▶ *Improvement:* Incorporate medical data

## 4. Geographic Oversimplification

- ▶ Only East/West Canada categorization
- ▶ *Improvement:* Use city/postal code data

# Limitations(2. Statistical Method Limitations)

## 1. Linear Regression Assumptions

- ▶ May have non-linear relationships
- ▶ *Improvement:* Test non-linear models

## 2. Randomization Test Limitations

- ▶ Ignores potential confounders (e.g., socioeconomic status)
- ▶ *Improvement:* Use multivariate regression

## 3. Classification Tree Oversimplification

- ▶ Misses variable interactions (e.g., program type)
- ▶ *Improvement:* Adopt ensemble methods (bagging/random forests)

# Overall Conclusion

The simplified registration system successfully boosted engagement, particularly among older members (30-80 years) and those with diverse program interests (5-7 programs).

However, younger participants and members with limited interests show lower attendance, while POC status, cancer risk, and location had minimal impact.

## **Next Steps and Future Analyses**

1. Target younger members with tailored programs
2. Encourage multi-program participation
3. Address broader engagement barriers (accessibility, relevance)