Population Proportion to Convenient Access to Public Transit by Sex, Age and Persons with Disabilities

Michael Nwokobia
School of Electrical and
Computer Engineering
Carleton University
Ontario, Canada
Email: michaelnwokobia@cmail.carleton.ca

Jamil Dergham
School of
Software Engineering
Carleton University
Ontario, Canada
Email: jamildergham@cmail.carleton.ca

Azizul Hasan
School of Electrical and
Computer Engineering
Carleton University
Ontario, Canada
Email: azizulhasan@cmail.carleton.ca

A.J Clarke
School of Electrical and
Computer Engineering
Carleton University
Ontario, Canada
Email: jashclarke@cmail.carleton.ca

Blessing Okeme
School of Electrical and
Computer Engineering
Carleton University
Ontario, Canada
Email: blessingokeme@cmail.carleton.ca

Mingqi Feng
School of Electrical and
Computer Engineering
Carleton University
Ontario, Canada
Email: mingqifeng@cmail.carleton.ca

Keywords - GTFS (General Transit Feed Specification), CIF (Canadian Indicator Framework), GRT (Grand River Transit), CSD (Census Subdivision)

Abstract—**Canada is focused on achieving the 2030 Agenda for Sustainable Development. This project aims to provide useful data that can be used to determine if Canada is on track to meet the 2030 agenda for Sustainable Development. A big part of this initiative is ensuring a high proportion of population has convenient access to public transit in various municipalities across Canada. Additionally, most public transportation passengers walk from their starting points to stops and from stops to destinations, local spatial availability and accessibility is sometimes measured in terms of pedestrian (walk) access rather than park and ride or transfers. As a result, access to public transportation is considered convenient when an officially recognised stop is within 500 metres of a reference point such as a home, school, workplace, market, etc. to a low-capacity public transportation system (e.g. bus, Bus Rapid Transit) and/or 1 kilometre to a high-capacity system (e.g. rail, metro, ferry). We collect data from these sources in order to better understand this proportion and reach that goal. Here is the link to our GitHub repository**

I. INTRODUCTION AND BACKGROUND

The purpose of this project is to prepare an inventory of available public transit data across municipalities in Canada to determine a process to compute indicator CIF 11.4.1. The value this project will add when delivered can be split into two perspectives, Statistics Canada and the overall Canadian populace. With respect to the Canadian people, this project will aid in achieving Canada's Sustainable Development Goals, specifically goal 11.4.1 which aims to have a high percentage of the population living within 500 metres of a public transport spot. With respect to Statistics Canada, this project is valuable because it would help Statistics Canada improve the Geo-spatial data acquisition process in the context of multiple sources or data providers and help improve the stable annual production of the indicator.

The team has collected data on the transit system of CSD's with over 20,000 people. We have used the template provided to us by Alexandra Bozheva, an analyst for the Statistics Canada. The team coordinated to divide up the number of cities being investigated. The members also changed the format from phase 1 of the data set by creating individual profiles for each city, after following advice from our teaching assistant, Kevin Dick and adhering to the original template provided to us. However, due to missing data on the transit systems of a significant number of smaller cities, some of the entries do not provide useful data to be aggregated and cleaned to form a uniform dataset. Some other issues that we experienced during the investigation of the cities was the lack of a transit system or inconsistent data for the transit system. Due to the missing data and the time frame of the project, the team could not provide viable data on all the cities. Instead, all the data collected from this point will be refined and filtered so it is easy to access.

II. PROBLEM DEFINITION

The data that was collected came in a variety of formats such as Geographic Transit Feed Specifications, Grand River Transit etc. The members will find a way to display this data in a cohesive way so that it is easily accessible. The members must also assess the data closely to determine confidence lev-

els. Lastly, the members are phased with vetting the collected data to eliminate all errors that could have occurred, such as duplicates of cities or the removal of data that the members are not confident with.

III. DESIGN AND METHODOLOGY

Our data munging phase entirely consisted of manual data collection. The phase includes to create an inventory for the CSD's with a population above 20,000 which consisted of GTFS data or adjacent such as maps. We had decided to divide the   220 CSD's between the 6 of us and use our best judgement and search skills to obtain the data. Our best-case criteria was the presence of GTFS data, however, in the event of its lack there of we searched for adjacent data such as transit maps. The steps we adhered to obtain data/maps are the following:

1) Use the Statistic Canada website to select a CSD to collect its GTFS data.
2) Search the web for the GTFS data using the following group of words" (CSD) + gtfs data" to find the CSD's GTFS data
3) For the larger cities, the website 'transitfeeds.com' came in handy and we were able to obtain information about the GTFS data using the 'Feed info/Terms of Use' link which lead to the information about the GTFS dat for the particular CSD.

Fig. 1. The link lead us to the information about the GTFS data

4) In the case for the lack of GTFS data, we left it to our own best judgement on what is the best adjacent data could be. The map/data would often be on the transit website of the CSD or an online public portal.

We tried to avoid getting data from non-official sources, if a transit provider does not have an "Open Data" section on their website, we would find that data available on other websites where that public transit data was available to download. We cleaned up the data by formatting them into Excel sheets, one sheet for each specific CSD, per Stats Canada's request. In terms of the quality of our final dataset, a majority of the data is in GTFS format, with the base case being at least a route map of the CSD, our team is satisfied with the data we have collected.

The consistency of the data is difficult to judge as there is on way to double-check the data as there is only one source. Most of the data, especially for the larger cities, is relevant as they are up-to-date and can be used. For smaller cities, the data is often dated, however transit data is barely updated unless there are substantial changes in the transit infrastructure.

The steps mentioned above are used to conduct a search is the basic blueprint we followed. However, if data is still not found following the steps, we relied on the best judgement of the team member to collect best alternative.

We saved all the information about the GTFS data into separate sheets with specific template for each CSD to make sure data is easily accessible and organized.

Fig. 2. The template stores all the information about the GTFS data for the CSD

IV. RESULT AND DISCUSSION

At this point, we have collected data from 223 cities with populations greater than 20,000, indicating their public transportation routes. There are 123 cities have GTFS data, which is nearly 55.2 percent. Although this percentage seems to be high, but in this case, it won't affect the final result's accuracy that much since the cities with no GTFS data have relatively less population. We will need to investigate this issue before proceeding to the final step, which is after collecting population data around these transit routes in cities. Hence, we

can get a better estimate of whether Canada is on track to meet the 2030 Sustainable Development Agenda goal of having a high percentage of the population living within 500 meters of a public transportation stop. Furthermore, by examining this result, we can determine how Statistics Canada can improve the Geo-spatial data acquisition process in the context of multiple data providers and contribute to the indicator's stable annual production.

ACKNOWLEDGMENT

REFERENCES

[**?**] sustainable development goals - 17 goals to transform our world