Segmentation of Statistics Canada's Proximity Measures

Weekly Meeting

Week 1

Jonah Edmundson

Hello! I'm a new mother to two beautiful babies, 2-year old Benji and a golden retriever named Carl 🐶
I love crossfit, my kids, and ice cream 🍦
Deck parties and wine baby! 🍷

Ricky Heinrich

Background: BSc in Physical Science w/ minor in Business. I dream of applying my data science knowledge and skills to practical urban planning research and improvements, and towards building sustainable ‘circle economy’-based societies.

Noman Mohammad

hmu for sum fire beats: https://www.youtube.com/@Yesuuh 🔥

Avishek Saha

Motivated Master's of Data Science student with a keen interest in applying my statistical analysis and machine learning skills to drive business development and innovation in the tech industry.

Our Client

Data Exploration and Integration Lab – Centre for Special Business Project, Statistics Canada

- Main contact: Jérôme Blanchet
  - Unit Head - Data Science Engineering - Center for Special Business Projects | Statistics Canada

Project Description

- The Proximity Measure Database (PMD) is developed by the Data Exploration and Integration Lab (DEIL) at Statistics Canada.
- PMD provides a granular measure of proximity to services and amenities for planning and policy questions.
- PMD contains continuous measures for 10 amenities at a 'dissemination block' (DB) level.
- Goal is to segment continuous proximity measures to group similar dissemination blocks based on access to amenities.
- Clusters may provide valuable insights to policymakers and urban planners.
- Insights can help prioritize efforts to improve accessibility and promote social and economic sustainability.

Methodology

- EDA
- Clustering approaches
  - Connectivity based
  - Centroid based
  - Distribution based
  - Density based
  - Grid based
- Characterizing cluster profiles
- Validate and compare clustering approaches for quality assessment

Final Deliverables

- Final Report
  - data exploration
  - clustering approaches attempted
  - sensitivity analysis
  - a chosen (best) reproducible clustering methodology
  - characteristics of the clusters
  - identification of the PMD cut-off values

- Final Presentation Slides

Team Dynamics Plan

- Daily morning meetings
- Splitting into groups of 2 (rotating)
- Each group focuses on different aspects of the project
- Rotate team members among groups to gain exposure to different challenges and skill sets

Summary of Week 1

- Reading Relevant Papers
- Proposal Drafting and Edits
- Meeting with Client

Tasks for Week 2

- EDA
  - method to deal with missing values
  - exploring additional datasets
  - summarizing characteristics of data

- Trying connectivity and centroid-based clustering approaches, recording progress

- Meeting with client