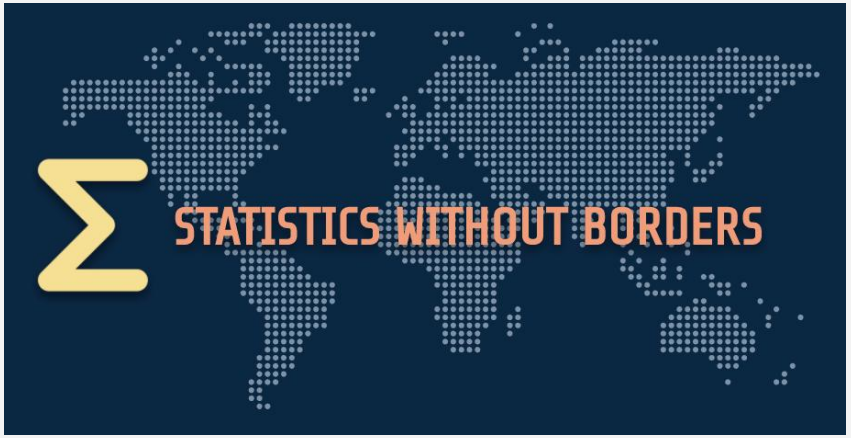
Worldwide Statistics Without Borders: Innovative Project Collaborations to Inform Decisions and Drive Discovery



August 7, 2023

Welcome to JSM 2023!

Chair: David Corliss

Presenter: David Whitford

Presenter: Dr. Michiko Wolcott

Presenter: Dr. Michelle Vanchu-Orosco

We welcome your questions throughout the session!

To start off in a good way, we are sharing the following

Land Acknowledgement

SWB Events at JSM

- Panel Presentation: Monday, 8:30am ← ***you are here!***
- SWB In-Person Meeting: Tuesday, 3:00pm ← ***tomorrow!***
- SWB Virtual Annual Meeting: Tuesday, August 15, 3:00pm EST
    - To register, email Matthew.W.Brems@gmail.com.

What should you expect today?

1. **Introduction to Statistics Without Borders (**David Whitford)
2. IMPACT Initiatives: Project Lifecycle (David Whitford)
3. Global Count Women's Right Org Directory (Dr. Michiko Wolcott)
4. Social Media Sentiment Analysis (Dr. Michelle Vanchu-Orosco)

What is SWB?

- **Started in 2008** by Steve Pierson, Gary Shapiro, Jim Cochran, Fritz Scheuren
- **Mission:** Statistics Without Borders (SWB) is committed to providing statistical services to organizations to maximize their positive impact on society.
- **Vision:** SWB envisions a world in which everyone has access to statistical expertise and uses that expertise to positively impact our world.

We’re organized in three main groups.

How can I get involved?

Find a vacancy!

Current:
- Networking Coordinator
- Content Developer
- Newsletter Developer
- Web Developer
- SharePoint Administrator
- Data Analyst

Reply to "Calls for Volunteers"

When **any vacancy opens up** (permanent roles or project-specific), we send a "call for volunteers" via email.

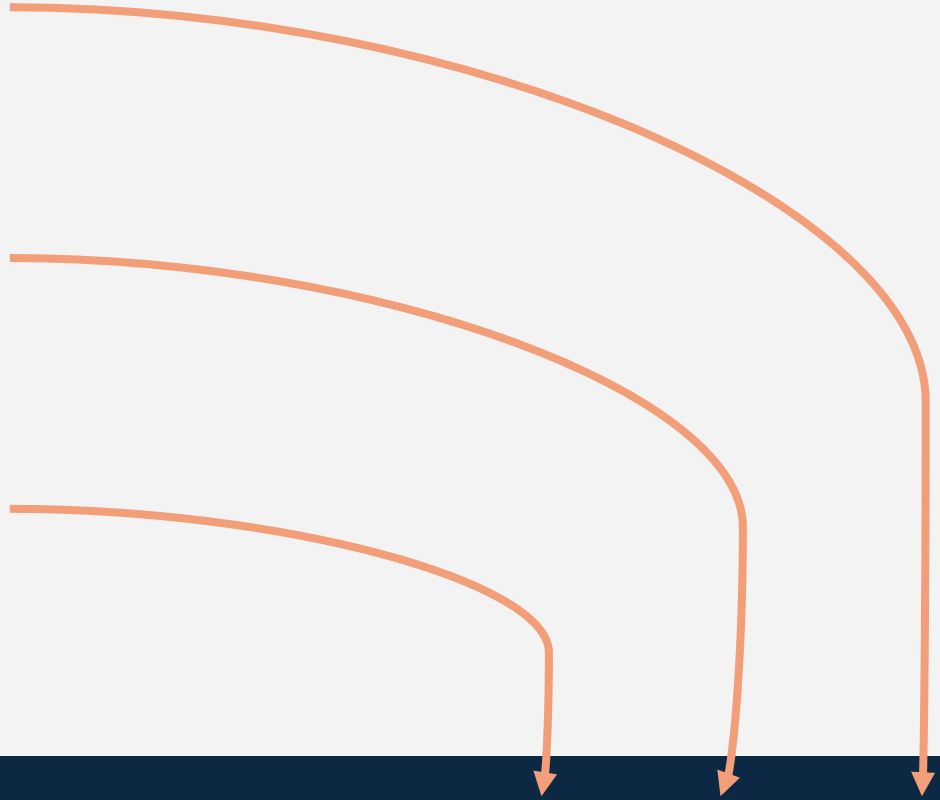
Register to be a member!

In order to receive emails, you have to register to be a member (and opt into receiving emails from SWB). **Fill out an application!**

Know a non-profit that could benefit from free data/consulting work?

Have skills & want to volunteer on projects around the globe?

Want to hear about the impactful projects we do on social media?



STATISTICS WITHOUT BORDERS

https://StatisticsWithoutBorders.org

What should you expect today?

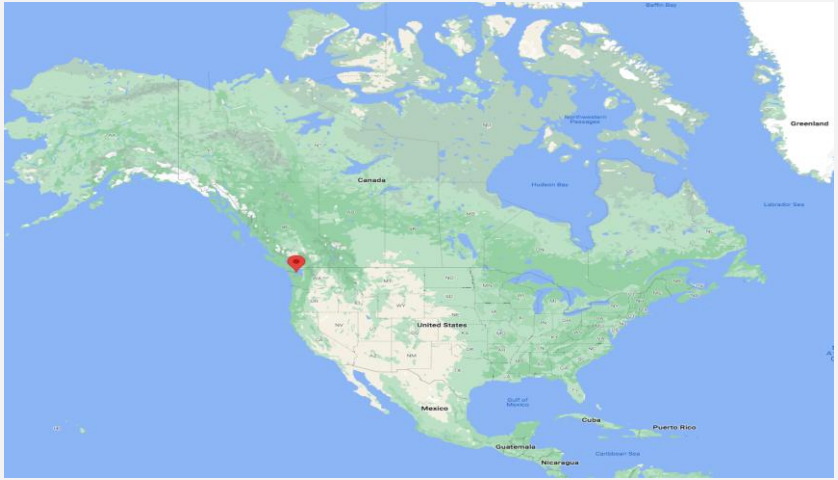
1. Introduction to Statistics Without Borders (David Whitford)
2. **IMPACT Initiatives Project Lifecycle (David Whitford)**
3. Global Count Women's Right Org Directory (Dr. Michiko Wolcott)
4. Social Media Sentiment Analysis (Dr. Michelle Vanchu-Orosco)

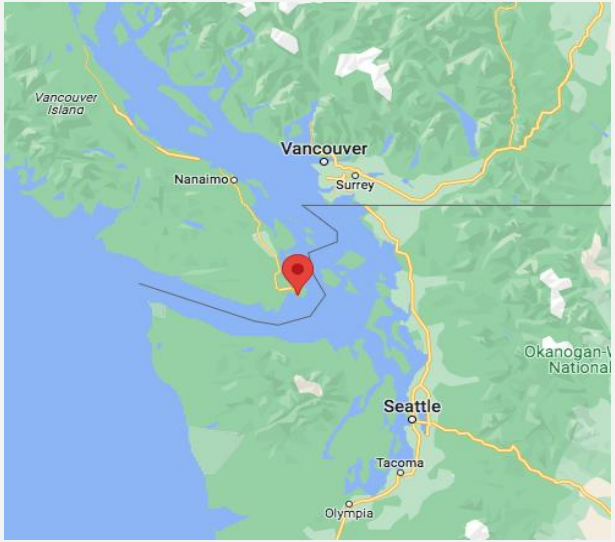
What should you expect today?

1. Introduction to Statistics Without Borders (David Whitford)
2. IMPACT Initiatives: Project Lifecycle (David Whitford)
3. Global Count Women's Right Org Directory (Dr. Michiko Wolcott)
4. **Social Media Sentiment Analysis** (Dr. Michelle Vanchu-Orosco)

Social Media Sentiment Analysis: Using Python to Help the Unhoused

Social Media Sentiment Analysis: Using Python to Help the Unhoused





Social Media Sentiment Analysis: Using Python to Help the Unhoused

- Non-profit: Greater Victoria Coalition to End Homelessness
- Rebranded in 2023 to: Alliance to End Homelessness in the Captial Region (AEHCR)
- Community Owned Solution with the mission to end homelessness in the capital region (South Vancouver Island – 13 municipalities)
- Wanted tool to support monitoring sentiment on Homelessness in Greater Victoria

Social Media Sentiment Analysis: Using Python to Help the Unhoused

Statistics Without Borders statisticswithoutborders@gmail.com via statisticswithoutborders.org Sun, 30 Jan 2022, 16:21

to me

Hello Joshua,

The Greater Victoria Coalition to End Homelessness (GVCEH, the Coalition) was formed in 2008 with a mission to end homelessness in the capital region. GVCEH is looking to build a tool that uses Sentiment Analysis to analyze information on social media/news posts by tone (positive, neutral, and negative). As this was previously done in a labor intensive way, the new tool also allows GVCEH to reduce subjective biases and standardize the methods and processes. If GVCEH understands where the information comes from (positive, neutral, and negative) strategies to target communities to change the narrative become possible.

SWB help is sought to perform Sentiment analysis based on Twitter and Reddit posts, find a way to quantify attitudes and impact of interventions, and integrate them into a beginner-friendly graphical user interface.

We are looking for 6-8 volunteers with experience with: technical team leadership, client interview and co-design of dashboards, web scrapping data collection, natural language processing (sentiment analysis), and time series data analysis.

Project duration: 20 weeks. **Expected workload:** 6-8 hours per week

Statement of Work

Call For Volunteers

The email sent out by SWB in January 2022

Social Media Sentiment Analysis: Using Python to Help the Unhoused

SWB Team

Terry Cai
US
Project Coord

Sangeeta Jayadevan
IN
Project Coord

Ju Du
DE
Project Coord

Shelia Flood
US
Integration

Oren Dar
IL
Modeling

Shreya Datar
US
Dashboard

Josh Weissbock
CA
Scraping

AEHCR Team

Althea Divigracia
CA
Communications

Michelle Vanchu-Orosco
CA
Research & SWB member

Social Media Sentiment Analysis: Using Python to Help the Unhoused

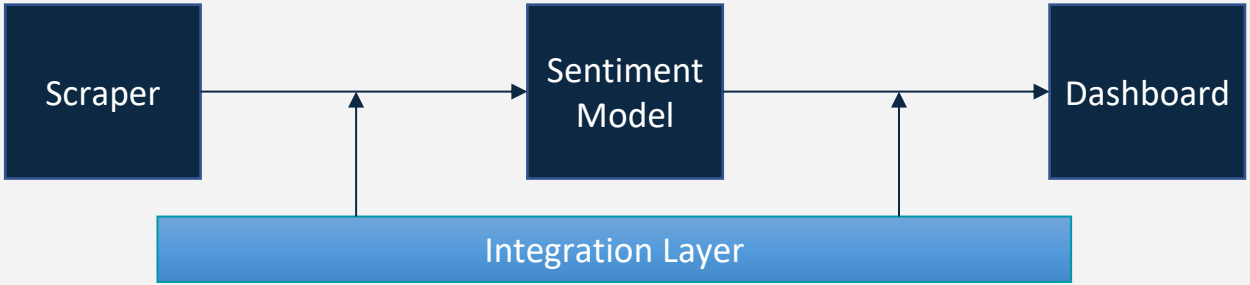
Milestone 1: The initial Deliverable

Scope

- Perform Sentiment Analysis based on Tweets using provided names and keywords
- Architecture decisions
  - Pipeline
  - All Python
  - Free

Social Media Sentiment Analysis: Using Python to Help the Unhoused

Process Pipeline



Social Media Sentiment Analysis: Using Python to Help the Unhoused

The Scraper

- SOW defined keywords to collect relevant tweets
  - About People Experiencing Homelessness
  - In Victoria, BC (Capital Regional District)
- Python: Tweepy Package (Twitter API) & Pandas
- Collected Additional metadata on tweets
- No substantive geo/location data available
- Stored as CSV files

Social Media Sentiment Analysis: Using Python to Help the Unhoused

Keyword Logic (Subject)

About Homelessness + IN Victoria

Organizations

- District of North Saanich
- District of Saanich
- Downtown Victoria Business Association
- Fernwood Community Association
- Foundry Victoria
- Greater Victoria Acting Together
- …

Sectors

- Harm Reduction
- Safe Supply
- Service provider
- Frontline
- front line
- frontline worker
- front line worker
- …

Words

- Poverty
- Poor
- Crime
- Violence
- Thief
- Theft
- Camp
- Camping
- …

Social Media Sentiment Analysis: Using Python to Help the Unhoused

Keyword Logic (Region)

About Homelessness + IN Victoria

Neighbourhoods

- Colwood
- Langford
- Metchosin
- Sooke
- View Royal
- Esquimalt
- …

Region Identifier

- Greater Victoria (#GreaterVictoria)
- Victoria
- VictoriaBC
- Victoria B.C.
- #YYJ, #GVCEH
- Greater Victoria Coalition to End Homelessness
- …

Influencers

- @BC_Housing
- @Island Health
- @VanIslandHealth
- @Adam_Stirling
- @CharlesBodi
- @NicoleChaland
- …

Social Media Sentiment Analysis: Using Python to Help the Unhoused

Scraper Challenges & Solutions

- Limited Twitter / API support
- API Restrictions
  - 7 Day History
    - Routine collection
  - Limited # of requests per 15 minutes
    - Batch Scraping
    - Boolean Keyword Logic
    - time.sleep(2.5)

Social Media Sentiment Analysis: Using Python to Help the Unhoused

Example Tweets

Date	Tweet
2023-02-16 03:25:58.335608	@CityOfVictoria former houseless encampment needing attention. Located waterside of Galloping Goose near FolEpi and Dockside Green. Butane/propane canisters and debris.
2023-02-16 03:32:04.920188	Our team will be out tonight with hot soup/ Bannock, harm reduction, coffee cards etc . We help folks connect with service providers such as housing applications specifically M'akola housing , detox, stabilization. As always thank you Victoria for supporting our work.
2023-02-17 03:34:27.066345	@DesmondCole I imagine it's the same story all across Canada. I live in Victoria and I can't even fathom the stats about homeless people here. How many there are, how many are sick and disabled, and how many have died from the negligence of those that are supposed to be helping them.

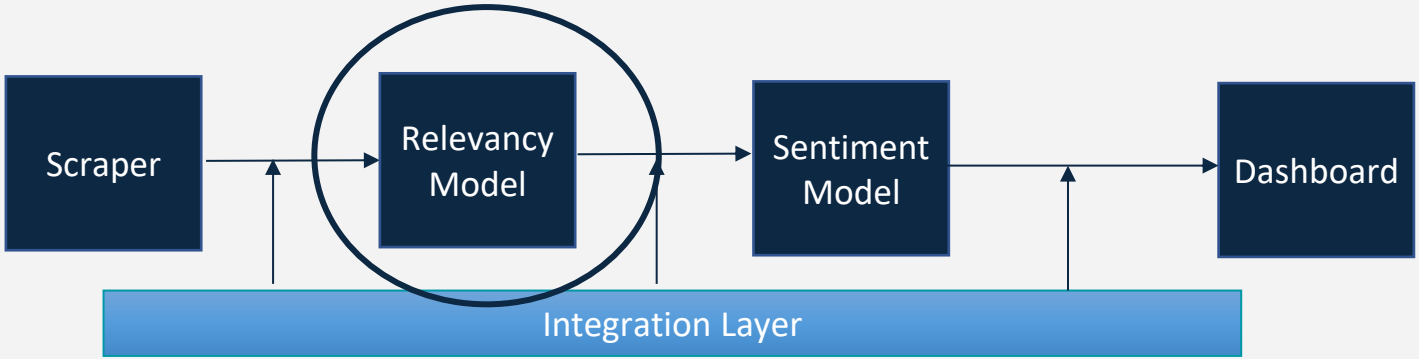
Social Media Sentiment Analysis: Using Python to Help the Unhoused

The Sentiment Model

- Hugging Face Transformer (Python Library)
- Off-the-shelf, trained on Twitter data, can run offline
- Assigns Positive/Negative/Neutral Label
- Pre-processing Concerns
- Team verified output; results were accurate

Social Media Sentiment Analysis: Using Python to Help the Unhoused

Process Pipeline Updated



Social Media Sentiment Analysis: Using Python to Help the Unhoused

A pivot – Relevance Model (Phase 2)

Why did this happen?

Words have multiple meanings
Victoria: City in Australia, Name
Central Park: New York City
Pandora: Music service
Experimented with refining list
Experimented with keyword filter

Hand Verification and Analysis

Hand labelled 3,000 tweets
About Victoria: Yes or No
About People Experiencing Homelessness: Yes or No
Grey Area: Homelessness vs Housing
Two Rounds with updated keywords
Determined initial "relevancy"

Social Media Sentiment Analysis: Using Python to Help the Unhoused

The results…

	precision	recall	f1-score	support
0	0.96	0.95	0.95	75
1	0.90	0.93	0.91	40
accuracy			0.94	115
macro avg	0.93	0.94	0.93	115
weighted avg	0.94	0.94	0.94	115

Social Media Sentiment Analysis: Using Python to Help the Unhoused

The Dashboard

App Built & Hosted by Streamlit
Simple to use → best option
Free, open-source, python-based framework
Main Functions:
Fetching/Summary data
Filtering in Dashboard
Download data

The Homelessness in Greater Victoria Dashboard

Social Media Sentiment Analysis: Using Python to Help the Unhoused

Summary

End to End Pipeline All in Python
Free for the Alliance
Automatically collects & filters for relevant tweets
Aids Alliance's support to the unhoused in Victoria BC

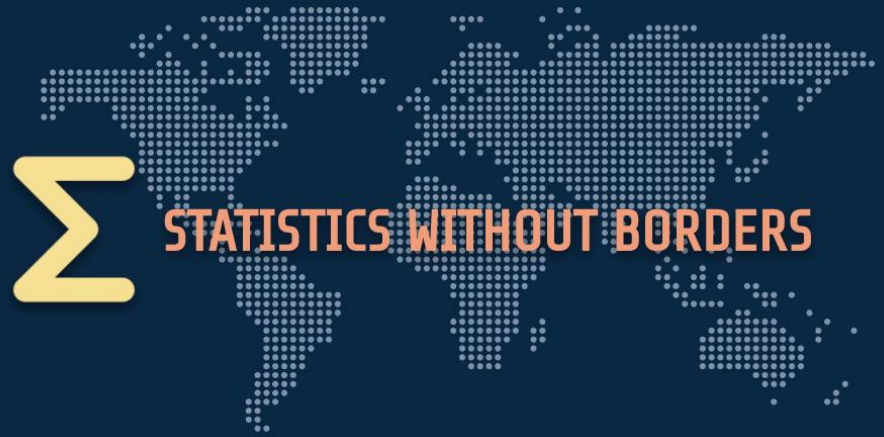
What's next?

Integration into AEHCRs Operations
Future Milestones?
Can abstract it to other sources (Reddit, News, etc.)
Future of Twitter API

SWB Events at JSM

- ~~Panel Presentation: Monday, 8:30am~~ ← ***you are here!***
- **SWB In-Person Meeting: Tuesday, 3:00pm**
  Intercontinental Toronto Centre Room: I - Ontario ← ***today!***
- SWB Virtual Annual Meeting: Wednesday, August 15, 3:00pm EST
  - To register, email Matthew.W.Brems@gmail.com.

Thank You!

https://StatisticsWithoutBorders.org

StatisticsWithoutBorders@gmail.com