Micro-Datathon #2 Instructions

Data Access, Result submission and communication

Data Access - Data.World

Step 1. Create data.world Account

All data you will need for this micro datathon is available on data.world. Please signup on https://data.world/ to get an account.

Step 2. Finding the project with datasets.

The data.world project Micro-datathon #2 can be accessed here.
(Last month's micro-datathon #1, Geospatial, can be accessed here)

Sharing Results

We would like participants to share their code, modified datasets and any other information that can help us improve this data as follows:

Step 1. Adding your results summary to the projects insights page

1. We would encourage all participants to add a summary of their results as an insight page on the data.world project's insight page.

2. Once on the Click on the add new insights button to create your own insight page.

3. You can add details and interesting plots here. If you have modified any files provided for this datathon i.e data cleaning and think that the changes will improve the data, please share your improved datasets via data.world and include the link in your insights page.

4. Please name the insight page as your first and last name (first_last) so that we know who has posted the results and have a way of contacting you.

Step 2. Storing related files on DataForGood's NextCloud.

If you plan to share code or any other files, please store them in our Nextcloud. Please send an email to dataops-YYC@dataforgood.ca to get credentials for login. Once you get the login details, please create a folder with the same name as your insight's page (First_Last name) and store all the files in it. Please make sure you add details about these files on the insight page.

Communication via Slack

For conversations, questions and access requests we recommend using the Data For Good Alberta slack channel **# yyc-micro-datathon**

If you are not yet a member of this Slack workspace please use the link below to join: Data For Good Alberta Invite Link

Recommended Software

All the demos shown in the Micro-datathon #1 meetup used PowerBI.

For Micro-datathon #2, we also recommend **Beyond 20/20** for IVT files (available, amongst others, from StatCan for census)

If you don't have access to PowerBI or are working on a Mac, here are other options that were also mentioned during Micro-datathon #1

- Tableau (https://public.tableau.com/s/)
- QGIS (https://qgis.org/en/site/)
- R (https://www.r-project.org/about.html),Rstudio (https://rstudio.com/products/rstudio/),leaflet package
- Python (https://www.python.org/), leaflet and folium
- Jupyter notebooks (https://jupyter.org/)

Reference Documents / Data Model

Helpful reference files can be found in our Data.World Census Project, under project files.

- Census Information Sheet (assembled by Data for Good YYC, work in progress)
- Census Profile Metadata (assembled by Data for Good YYC, work in progress)
- Micro-datathon #2 Data Model

StatCan presentation by Sarah Ahmed, at Data for Good May 28, 2020 meeting

- Slidepack from StatCan (link pending)
- Recording from StatCan (link pending)

Video Instructions

For micro-datathon **#1**, Paula Jennings produced a set of videos about spatial data and master data management, and a brief intro to using Power BI ("Web Open Data into Power BI").

A playlist of these videos can be found here.

Challenges

There are different levels of learning objectives in this micro-datahon. Start with the basic tasks and build your skills to the more advanced tasks.

Also feel free to make your own maps, based on the data we've provided (and any other data you want to link in). *Bonus if the visualizations relate to our Fall 2020 Datathon theme of homelessness.*

NOTE: See the Data.World Project Summary, for the location of the various tables within data sets (these have been provided for the Basic challenges, to help you get started).

Basic

These tasks will get you familiarized with the datasets:

- Open the StatCan / CMHC report on Core Housing. Read the report (1-pager), then play with the data using the Canada's Census Program Data Viewer (CPDV). Brief instructions are at the bottom of the web page (*search on CPDV*).
    - Starting with provincial data, find and view the 3 indicators for core housing need (unaffordable, unsuitable, inadequate housing)
    - Do comparisons across 3 Census Subdivisions (CSD, e.g. Calgary, Red Deer, Edmonton)

- In our document "Canada Census Information Sheet", go to the "Focus on Homelessness" section, and explore the various topics available

- How many shelter stays per capita, for Calgary Census Subdivision (CSD)
    - What is the difference between the Raw and the Clean shelter data file?
    - Consider what population stats you choose from Census Profile, and why (gender? include children? only 18+? …)
    - Data sources
        - Canada Census 2016 Alberta Data - ab census subdivision
        - Canada Census 2016 Homelessness - Homeless Shelter Capacity

- In your preferred visualization tool, create a tile of maps (or choose other delivery), that show % low-income (LIM-AT and LICO-AT), % insufficient housing (house in bad shape, housing too expensive), shelter stay (% bed count / population).
  The Census Profile Metadata will help you find the LIM-AT and LICO-AT,and housing indicators (see the Income and Housing categories)

- Data sources
  - Canada Census 2016 Alberta Data - ab census subdivision
  - Canada Census 2016 Homelessness - Homeless Shelter Capacity
  - Canada Census 2016 Homelessness - Core Housing Need
  - Micro-datathon #2 Project - Canada Census Profile Metadata v1

Intermediate

- Core housing Table 1 shows data across 3 census periods, for counts of core housing lack (unaffordable, unsuitable, inadequate).
  - Data underlying this table has been downloaded into a data.world data set. This data set provides rates for 2006, 2011, 2016; counts for 2016 only
  - What kinds of visualizations can you produce, how are things changing over time?

- Collective Dwellings - this statscan data is by country and province, providing counts of dwellings and population for Shelters, Nursing Homes, etc. How can you tie this is to the other data sets?

- Census Profile Data
  - Practice with Calgary vs. Edmonton CSD (later, we may be able to provide this by more-detailed geographies of Calgary Ward and Calgary Community)
  - Take advantage of our Census Profile Metadata spreadsheet, which gives some indication of rows relevant to homelessness (column N = Y). What other Census Profile rows could be relevant? Let us know your thoughts!

Advanced

- Build on micro-datathon #1
  - Given data provided in micro-datathon #1,how can you combine this with 2016 census data e.g.
    - By Census Subdivision (CSD), the Alberta Shelter Stays (matching shelter stay city to CSD),
    - By Census Tract (CT), the Kijiji rentals (which have addresses, could do spatial queries to place them in census tracts)
- Exploring the Census Profile and beyond
  - exploring the Census Profile and various StatCan data tables, reports, etc; what insights can you discover, around Homelessness Issues?

We welcome you to share your work, following the instructions above, with the rest of the Data for Good group!