User Guide: OSM Scraper

# Introduction

This document is an instruction manual for using the Jupyter Notebook that collects and processing OpenStreetMap data. The purpose of this notebook is to allow for faster collection and cleaning of this data, which usually requires collection from the website and manual cleaning within QGIS, which is often labour intensive.

Currently, this notebook provides 2 working module types, and an in development custom module that will allow the user to choose any data and download it. The first working module generates heatmaps for specific categories using the amenity data of a location. The second module generates a simplified version of the street network.

All modules can be expanded, and new modules can be added when specific use cases present themselves. The custom module will also be completed in the future. One benefit of this module will be the ability to choose any specific data and download it.

# Setup

The setup is relatively simple. First, access the notebook at this link:

<https://colab.research.google.com/drive/1ucwP86PwuuXCMtD62rI0OEokmyvQ0mxV?usp=sharing>

Next, run the first cell to install the necessary libraries into the environment, and clone the Github repo containing the backend files. This should take a few seconds.

Next, run the second cell. This will begin the processor and open the UI. It should look like this:

A black rectangular object with a line

AI-generated content may be incorrect.

First, click the ‘Select Output Folder’ button. This will prompt a connection to google drive, click through. This allows google colab to ‘mount’ a folder in your drive. After clicking through the prompts, a new box will open, asking you to make a folder name. Enter your desired name and it will create this folder in your drive. From here on, everything will be saved into the folder.

Next, proceed to select a location. You can either enter the name of the location (ie Yerevan, Armenia), or upload a geojson or a zipped folder of a shapefile and its components. When entering the name, it will collect all data for that city, but using a custom area will only take what is within those bounds.

The last setup is to select the workflow you would like. Click on the option you would like to choose. Note that if you want to use Heatmap Generation, click on one of the others, then click back. That will initiate the next screen.

# Workflow 1: Heatmap Generation

A screenshot of a computer program

AI-generated content may be incorrect.

When you have chosen the heatmap generation workflow, you will first be presented with this option. You have the choice to choose some pre-generated categories, or define your own. There will be a separate document with the categories, but they are the ones that were used for analysis in Yerevan for WTCY. Templates can be added easily. If you choose to create your own, this will be possible in the next step, once the data has been gathered and cleaned.

## Template Categories

A screenshot of a computer program

AI-generated content may be incorrect.

When you continue using the template categories, it will automatically collect and clean the datasets, and prompt you to enter some information.

**Output folder:** This is the name of the folder that all rasters will be stored in, within the main output folder that was determined initially.

**Cell size:** The resolution of the raster, in degrees. The default of 0.001 is fairly suitable for city scale, however it can be adjusted.

**Bandwidth:** Determines the ‘smoothness’ of the raster. A lower bandwidth will create more focused hotspots, a smoother, more spread out raster.

**Categories:** You can choose which categories you wish to safe. Use the ‘Select All’ button to choose all, or ctrl+click to choose specific ones.

When you lick generate heatmaps, the rasters will be automatically saved into the designated output folders, and a simple visualization will appear below. From here you can go back and adjust anything as you would like. The system is fairly robust, but

## Custom Categories

A screenshot of a computer

AI-generated content may be incorrect.

When choosing custom categories, it will clean the data and provide a list of the possible facility types. From here, I would recommend inputting the list into an LLM and tell it which categories you would like. It will then sort them into those categories. You then name the category and provide the list of facility types as prompted. They must retain the quotation marks, and be separated by commas, so prompt the LLM to return them in that format. You may add or delete categories using the provided button or the trash can on the right.

Any unused facility types will be designated into the category ‘other’.

Once you choose ‘Save Categories’, the process will be the same as is described in ‘Template Categories’ above, as you will be prompted to choose the folder and raster elements.

# Workflow 2: Gather Street Network

For now, this is very straightforward. Select the network type (currently just the driving network is supported, others are more nuanced), and press Get Street Network. This will then process the data. It should take a couple of minutes, however the final and original datasets will be present within your drive folder.