fundus

July 27, 2022

1 FUNDUS - The urban geography of inequalities: Budapest summer edition

The FUNDUS project of Urbanum Lab (a tech lab empowered by the interdisciplinary Urbanum Research Foundation), as its very title suggests, aims at laying the foundations for a globally applicable, open and accessible approach (including technological as well as methodological contributions) for the assessment of how economical factors of living (especially, real estate prices) are correlated with the quality of life in a certain urban environment, as captured by environmental and other social indicators.

In this notebook, the *Budapest summer edition*, we will be focusing on a simple, yet central question with manifold interpretations and consequences: do higher (average) real estate prices indicate a greener environment, less prone to the heat island phenomenon? (In simpler words: *does more expensive mean greener and cooler in the summer*?)

2 Outline

In this notebook, we are using Budapest as an example, and part of our data comes from our own scraper solution to obtain average property prices from real estate offers on the internet.

In particular, the notebook is structured as follows: + We present technical details on the data being used + we summarize the technical takeaways provided, + …

2.1 How to reproduce our results

This notebook presents the first, exploratory data analysis phase of our project. We would like to get a glimpse into the data and its possible use. First, you have to run the property price scraper. You can find more info on getting the property prices data in the repository of the project. We assume that you have a WEkEO account, if this is not the case, please register here. It is a good practice to install all the project requirements using a separated Python 3.9.10 (or higher) virtual environment. Use requirements.txt to install all dependencies. Last, you have to configure your `.hdarc` file using your WEkEO credentials. This article shows you how to do this.

If you would like to adapt our notebook to a different municipality or for a different time horizon, use WEkEO's online data exploratory tool.

You might find useful the BoundingBox tool to get latitude and longitude coordinates of a given area.

3 Data used

3.1 WEkEO datasets

Dataset queries were generated by using the WEkEO online platform. The queries can be fund in the `data/jsons` folder + Global 10-daily Leaf Area Index 333m


```json
{
  "datasetId": "EO:CLMS:DAT:CGLS_GLOBAL_LAI300_V1_333M",
  "dateRangeSelectValues": [
    {
      "name": "dtrange",
      "start": "2022-06-01T00:00:00.000Z",
      "end": "2022-06-30T23:59:59.999Z"
    }
  ]
}
```


- Level 2 Land - Sea and Land Surface Temperature Radiometer (SLSTR) - Sentinel-3


```json
{
  "datasetId": "EO:ESA:DAT:SENTINEL-3:SL_2_LST___",
  "boundingBoxValues": [
    {
      "name": "bbox",
      "bbox": [
        18.99804053609134,
        47.42120186691113,
        19.190237776905892,
        47.58048586099437
      ]
    }
  ],
  "dateRangeSelectValues": [
    {
      "name": "position",
      "start": "2022-06-01T00:00:00.000Z",
      "end": "2022-06-30T00:00:00.000Z"
    }
  ],
  "stringChoiceValues": [
    {
      "name": "productType",
      "value": "LST"
    },
    {
      "name": "timeliness",
      "value": "Near+Real+Time"
    },
    {
```

```
      "name": "orbitDirection",
      "value": "ascending"
    },
    {
      "name": "processingLevel",
      "value": "LEVEL2"
    }
  ]
}
```

- Global 10-daily Fraction of Vegetation Cover 333m

```
{
  "datasetId": "EO:CLMS:DAT:CGLS_GLOBAL_FCOVER300_V1_333M",
  "dateRangeSelectValues": [
    {
      "name": "dtrange",
      "start": "2022-06-01T00:00:00.000Z",
      "end": "2022-06-30T23:59:59.999Z"
    }
  ]
}
```

3.2 External dataset

The square meter prices dataset was collected on 12th July 2022 using our freely available scraper. The data is in the `data/aggragated` folder. **WARNING**: Check the README.md file of the scraper to get your own data.

4 Technical takeaways

The notebook provides a novel methodology for exploratory data analysis in the field of urban digital geography, instantiated using a limited, yet appropriate example (Budapest, Hungary in the summer). At the end of this notebook, you will know: + How to aggregate data using the *h3* library, + how to visualize the data using the pydeck package, and + how to investigate spatial discrepancies along property prices and environmental factors

5 Data acquisition

5.1 Library imports

Here, we import packages for the project.

5.1.1 Gathering the data

WARNING: Downloading the datasets takes time! The data will be downloaded into the current working directory. Use your favorite tools to move the downloaded files to the appropriate folders.

5.1.2 SLSTR

5.1.3 LAI

5.1.4 FCOVER

Your operating system and tools might be different, below you can read tips which might be useful on a Linux machine + The notebook assumes that all data is in the `data` folder. + Move all zip and nc files to the folder (e.g. you can use the `mv` command, like `mv ../data "*.zip"`) + Make folders for the data fiels (`cd data; mkdir leaf_data temp_data fcover`) + Move the *.nc files to the corresponding folders (e.g. move Leaf Area Index data into `leaf_data`) + Move the* .zip files into the `temp_data` folder `cd temp_data; unzip "*.zip"; rm "*.zip"`

6 Cleaning and transforming data

6.1 Property square meter prices

We scraped a Hungarian real estate listing site to get property prices in Budapest. The listing entries were geocoded using the `geocoder` package. The geo-coordinates were indexed using the H3 hexagonal geospatial indexing system. You can check the resolution table of the cell areas here. For more details, you can check the repository of the scraper.

The data looks like this:

[6]:

	l7	price
0	871e020caffffff	1.219298e+06
1	871e02684ffffff	1.200000e+06
2	871e030a9ffffff	1.508301e+06
3	871e03134ffffff	9.494950e+05
4	871e03449ffffff	1.197017e+06

The hexagons listed in these files constitues our area of interest. ## Temperature data The code below aggregates the average temperature data on various levels of H3 hashing and writes the results to a tsv file.

6.2 Global 10-daily Leaf Area Index 333m

The code below computes the average LAI and assigns H3 hash codes to the values. The results will be saved into a tsv file.

6.3 Global 10-daily Fraction of Vegetation Cover 333m

The code below computes the average FCOVER and assigns H3 hash codes to the values. The results will be saved into a tsv file.

7 Visualizing the data

7.1 Maps

7.1.1 Square meter prices

[32]: <IPython.core.display.HTML object>

7.1.2 Temperature

[33]: <IPython.core.display.HTML object>

7.1.3 Leaf Area Index

[34]: <IPython.core.display.HTML object>

7.1.4 Fraction of Vegetation Cover

```
/tmp/ipykernel_67815/3755050406.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_fcover["normalized"][df_fcover["normalized"] == np.inf] = 255
```

[35]: <IPython.core.display.HTML object>

7.2 Analysis

We would like to test the common conception that wealthier neighborhoods are greener and also enjoy a lower average temperature during summer. Since we have no data on wealth at this granularity, we use property square meter prices as a proxy of wealth. This is a strong and yet not often assessed assumption. In particular, we test if + temperature and greenness, + temperature and square meter prices, + greenness and square meter prices

are connected.

We present our findings as interactive visualizations. We start with a realtively fine H3 resolution (7), giving us a fairly tight covergae of the geographical are under investigation. However, as we notice that the geographical resolution might not match the resolution of economical data, we also experiment with lower H3 resolutions.

7.3 H3 level 7

[23]:

	l7	price	celsius	lai	fcover
4	871e03449ffffff	1.197017e+06	19.883478	4.093056	0.736756
5	871e0344affffff	9.605164e+05	22.373126	3.321111	0.779938
6	871e0344bffffff	1.556710e+06	18.937453	3.557240	0.784969
7	871e03459ffffff	1.115068e+06	20.933828	3.778786	0.813342

9	871e0345dffffff	1.449016e+06	21.627415	3.191146	0.775750

[24]:

	variable	variable2	correlation	correlation_label
0	price	price	1.000000	1.00
1	price	celsius	0.082898	0.08
2	price	lai	0.130627	0.13
3	price	fcover	0.085454	0.09
4	celsius	price	0.082898	0.08
5	celsius	celsius	1.000000	1.00
6	celsius	lai	-0.064254	-0.06
7	celsius	fcover	-0.050965	-0.05
8	lai	price	0.130627	0.13
9	lai	celsius	-0.064254	-0.06
10	lai	lai	1.000000	1.00
11	lai	fcover	0.958244	0.96
12	fcover	price	0.085454	0.09
13	fcover	celsius	-0.050965	-0.05
14	fcover	lai	0.958244	0.96
15	fcover	fcover	1.000000	1.00

[25]: alt.LayerChart(…)

7.4 H3 level 6

Since our price data is collected on street name level, maybe we should use a lower resolution.

[26]:

	l6	price	celsius	lai	fcover
4	861e0344fffffff	1.149972e+06	21.141896	3.661387	0.775320
5	861e0345fffffff	1.071476e+06	21.493805	3.573483	0.794305
7	861e03607ffffff	8.155039e+05	20.747444	0.731667	0.253750
9	861e03617ffffff	7.243512e+05	20.387555	1.173775	0.331828
10	861e0361fffffff	8.734682e+05	22.131759	0.756120	0.290160

[27]:

	variable	variable2	correlation	correlation_label
0	price	price	1.000000	1.00
1	price	celsius	-0.056756	-0.06
2	price	lai	0.205356	0.21
3	price	fcover	0.187743	0.19
4	celsius	price	-0.056756	-0.06
5	celsius	celsius	1.000000	1.00
6	celsius	lai	-0.218399	-0.22
7	celsius	fcover	-0.248033	-0.25
8	lai	price	0.205356	0.21
9	lai	celsius	-0.218399	-0.22
10	lai	lai	1.000000	1.00
11	lai	fcover	0.967857	0.97
12	fcover	price	0.187743	0.19

13	fcover	celsius	-0.248033	-0.25
14	fcover	lai	0.967857	0.97
15	fcover	fcover	1.000000	1.00

[28]: alt.LayerChart(…)

7.5 H3 Level 5

[29]:

	l5	price	celsius	lai	fcover
4	851e0347fffffff	1.126057e+06	21.151154	3.594296	0.789122
6	851e0363fffffff	7.993279e+05	20.890202	1.467126	0.421782
7	851e036bfffffff	7.341340e+05	20.802957	1.497615	0.481584
8	851e0373fffffff	9.410714e+05	20.786614	3.598063	0.761485
10	851e037bfffffff	1.072051e+06	21.231511	1.795492	0.469863

[30]:

	variable	variable2	correlation	correlation_label
0	price	price	1.000000	1.00
1	price	celsius	0.828449	0.83
2	price	lai	0.637665	0.64
3	price	fcover	0.538340	0.54
4	celsius	price	0.828449	0.83
5	celsius	celsius	1.000000	1.00
6	celsius	lai	0.115429	0.12
7	celsius	fcover	0.033319	0.03
8	lai	price	0.637665	0.64
9	lai	celsius	0.115429	0.12
10	lai	lai	1.000000	1.00
11	lai	fcover	0.967581	0.97
12	fcover	price	0.538340	0.54
13	fcover	celsius	0.033319	0.03
14	fcover	lai	0.967581	0.97
15	fcover	fcover	1.000000	1.00

[31]: alt.LayerChart(…)

8 Discussion and conclusion

Here, we checked if there is a connection between property prices and environmental factors. In particular, we wanted to see if having a more expensive property in Budapest, Hungary correlates with a greener environment and a cooler summer. We used WEkEO data to get information on the environment and we scraped property prices to supplement the data. Interestingly, using lower (but still not too coarse) geographical resolutions provided us with some notable (even if not very surprising) correlations between more expensive properties and the greenness of its environment (using both the LAI and the FCOVER measure for greenness. However, we could not verify any significant direct connection between higher prices and more acceptable temperature conditions! This might be a first step towards an important future investigation: if consequences of climate change can be effectively tackled on individual level by spending more money. (Maybe they cannot.)

9 Future directions

In the future, we would like to increase the time window to get more LAI and FCOVER data. We investigate the possibility to incorporate other datasets like OSM and get official statistics on house prices and/or income, widening the scope of the FUNDUS project, to finally turn it into a full-fledged, environmentally and economically conscious qualitiy-of-life assessment approach. As already mentioned, a particular and potentially very interesting research question arising from the present notebook: can money effectively battle the effects of climate crisis in our direct surroundings? In order to learn more, we shall primarily proceed by expanding the geographical scope of the present notebook, and maybe even considering further economical factors. Here, the scarcity of open data is still a major impediment, but we hope we can contribute to the improvement of the situation with our endeavor.