Case Studies in Scientific Computing [MA4306]

Project 3 – The Bikesharing Rebalancing Problem

Input Data – Readme

What's in the dataset?

	A	B	C	D	E
1		id	delta	x_coord	y_coord
2	0	881f8d4581ff	1	4.812.500.860.570.910	11.515.835.823.754.700
3	1	881f8d44c3ff	3	48.181.652.147.235.600	11.572.290.326.157.100
4	2	881f8d45d5ff	-1	481.385.379.569.946	11.521.700.265.437.700
5	3	881f8d4593ff	-4	4.812.586.145.649.900	11.560.800.501.730.200

The data provided is in .csv format. It contains 795 rows + column headers. The underlying dataset is a list of all MVG Rad trips that took place in 2021. You can find a Jupyter Notebook written in Python where the preprocessing as well as the input data in the moodle folder.
Data structure - the first column is the index. The following columns contain the following:

- **id**: the unique identifier of each hexagonal H3 cell (more information on H3 to be found here: https://h3geo.org /). The cells listed are the ones where at least one rental or return took place.
- **Delta**: the disbalance (returns – rentals) of each cell. The numbers are aggregated over one week (Monday to Friday) and rounded up (down) for positive (negative) numbers. Due to the averaging and rounding, we have a difference in total bikes needed (the sum of the column *delta* is not zero). My suggestion: only consider cells with an absolute delta greater than one. Those are still 165 and the column sum is not that skewed (difference of -15 bikes).
- **X_coord/y_coord**: the coordinates of the centers of the respective cells.

Here's a graphic representation of the data (code also in the notebook). If you want to use the notebook, unpack the input data and save it in the same directory where to notebook is.