**Code Description**

In this file, you can read over the code files to see the logic of this project. The numbers in the file names reflect the order of the different files during the project. We have same procedure for data within any period, so we put the latest code we developed for 2020 here.

**1\_code\_for\_extracting\_information**

We will need to extract information for each of the four years from 2018~2021. The code is exactly the same except for the years and one differences from the original data: we can use safegraph\_place\_id to extract POIs before 2021 and we need to use placekey in 2021. We put the code for 2020 and 2021 here as examples.

1. 1\_Filtering\_Data\_2020.ipynb, 1\_ading Dec\_2020.ipynb; 1\_Filtering\_Data\_2021.ipynb: filter the foot traffic patterns data we need from SafeGraph.
2. 2\_Creating\_Matrix\_NYC\_2020.ipynb; 2\_Creating\_Matrix\_NYC\_2021.ipynb: we want to create the visits data before we start creating the table.
3. 3\_Creating\_final\_table\_NYC\_2020.ipynb; 3\_Creating\_final\_table\_NYC\_2020.ipynb: create the table that will be put in Particle swarm optimization (PSO).

The inputs are data from data/1\_data\_for\_extracting\_information, and the outputs will be the 4 files under data/2\_data\_for\_PSO\_Calibration.

**2\_code\_for\_PSO\_calibration**

1. 4\_4\_PSO\_2018\_NYC\_norm\_2\_PSO\_30.py, 4\_4\_PSO\_2018\_NYC\_norm\_2\_PSO\_30.sbatch: example code for running PSO. There are several lines in the .py file together with the sbatch file that are working on the large-scale computing part. Based on these two files, we create the code of each normalization and PSO combination for each year. Since the sbatch file shows settings of the server and the paper review is anonymous, we don’t share the sbatch code. Please contact us if you need it.
2. **2018/2019/2020/2021**
   1. generate\_code\_file\_2018/2019/2020/2021.ipynb: this file is to generate code of the 288 combinations and the txt.file to submit these jobs to a server.
   2. sbatches\_commands\_20182019/2020/2021.txt: this txt file is used to submit the 288 jobs to a server.
   3. **sub\_code**: this folder is to store the code for each combination. Since these follow almost same structure as the templates (4\_4\_PSO\_2018\_NYC\_norm\_2\_PSO\_30.py and 4\_4\_PSO\_2018\_NYC\_norm\_2\_PSO\_30.sbatch), we don’t upload this folder. Please contact us if you need it.

The inputs are the 4 tables under data/2\_data\_for\_PSO\_calibration, and the outputs will be the calibration results obtained by each core on the server. To make the folder less messy, we only put the aggregated results of 6493 CBGs for each combination in each year under each folder in **data/3\_PSO\_results/PSO\_calirabtion\_results**. The code will be provided in **code/3\_code\_for\_combining\_PSO\_results**.

**3\_code\_for\_combing\_PSO\_results**

1. concate\_params\_2018/2019/2020/2021.py: these files are to combine the results obtained by each core on the server for each combination of each year into an aggregated one.
2. compute\_errors\_2018/2019/2020/2021.py.ipynb: these code files are to compute the average/median costs for each combination of each year as matrices.

The inputs of this folder are the calibrated results from each core on the server, and the outputs will be the results for each combination of each year in **data/3\_PSO\_results/PSO\_calirabtion\_results**, and the average/median cost matrices of each year in **data/3\_PSO\_results/simulated\_annealing\_data\_2018~2021**.

**4\_analysis**

1. 1\_distribution\_visualization\_by\_year.ipynb: this file is to visualize the yearly distributions from 2018 to 2021 for the calibrated parameter values.
2. 2\_test\_valid\_changes\_with\_ks.ipynb: this file is to conduct Kolmogorov–Smirnov tests on the calibrated parameter values.
3. 3\_population\_param\_changes\_2019~2021.ipynb: this file is to visualize the general comparisons between the calibrated parameters of year pairs.
4. 4\_top SES\_param\_changes\_all.ipynb: this file is to visualize the comparisons in CBGs where a single socio-economic feature ranks top 5% with the calibrated parameters of year pairs.
5. 5\_top SES changes\_param\_distributions\_delta VS population\_as\_of\_2021.ipynb: this file is to visualize the comparisons in CBGs where a single socio-economic feature ranks top 5% with all residents in NYC using the delta values (as of 2020 and 2021 separately).
6. 6\_descriptive\_comparison\_store\_pairs.ipynb: this file is to pick ground truths of some chains to validate our analysis results.
7. 7\_descriptive\_comparison\_uni-dimension\_figures.ipynb: this file is to visualize the relationship between store area and store visit shares across the 4 years.
8. 8\_clusters\_param\_distributions\_all.ipynb: this file is to visualize the comparisons in each cluster identified by Kmeans with the calibrated parameters of year pairs.
9. 9\_clusters\_param\_distributions\_delta VS population\_as\_of\_2021.ipynb: this file is to visualize the comparisons in each cluster identified by Kmeans with all residents in NYC using the delta values (as of 2020 and 2021 separately).
10. 10\_DID.ipynb: This file is to conduct the DID model.
11. 11\_general\_change\_with\_top\_clusters.ipynb: this file is to get some numerical results describing the general changes in all residents, top 5% groups, and Kmeans clusters.
12. 12\_model\_performance\_analysis.ipynb: this file is to analyze the performance of the proposed model in this study.