Elena Porfidia “elenaporfidia”
Miguel Perez “miguelperez”
Joaquin Pinto “joaquinp”
Rodrigo Rivarola Monzon “rrivarola”

CAPP 30122 - Computer Science with Applications II - Milestone 3

Bridging the Gap: Enhancing Early Childhood Education Access in Illinois

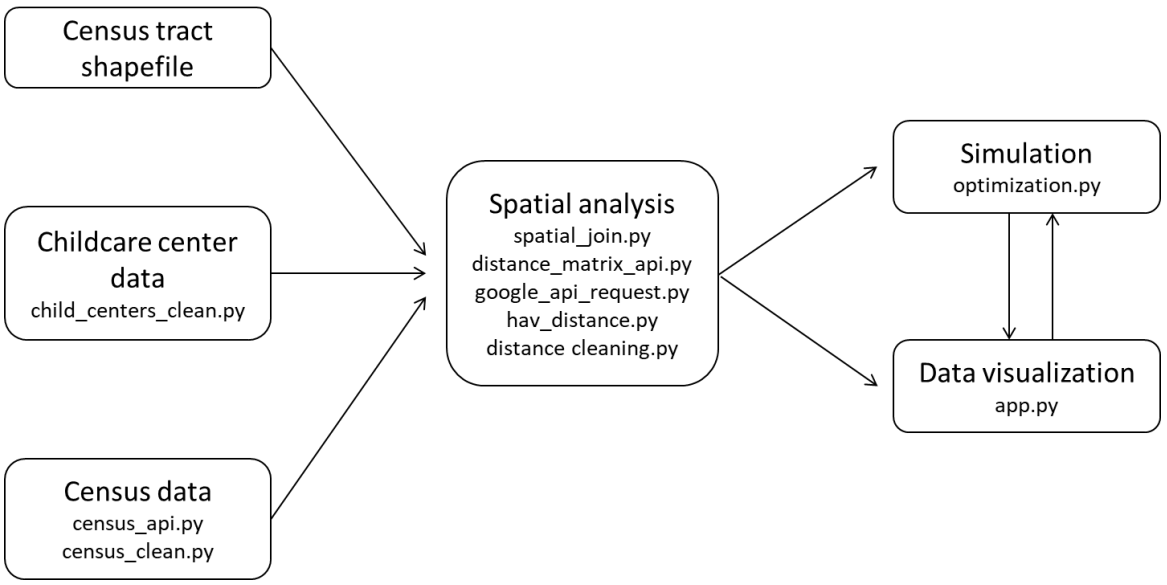
By Los CAPPos

● Project Abstract

The “Bridging the Gap: Enhancing Early Childhood Education Access in Illinois” project embarks on a pivotal analysis to understand the role of school vicinity on early childhood education (ECE) attendance rates for children between the ages of 0 and 5 in Illinois. While early childhood centers (ECCs) are more accessible and have a noteworthy capacity in Illinois, only 45% of children participate in these ECE programs. This indicates a significant disparity between availability and utilization. Our project targets this discrepancy, aiming to uncover the zip codes, counties, and census tracts, with the most pronounced shortages in licensed childcare relative to the local young population's size.

This study collects and evaluates information on ECE facility locations, capabilities, and socioeconomic factors at the census tract level using a rigorous approach. The dynamic model simulation projects the possible results of establishing more ECCs to see the real gains in accessibility, making sure that interventions are effective in reaching underprivileged communities based on data. By use of this simulation, interested parties may proactively strategize fair extensions of early childhood education centers, therefore directly addressing the distinct requirements brought to light by our study.

- **Overall structure of the software**



The project structure is pretty straightforward as illustrated in the graph above. Our data is stored in the “*data*” folder, our analysis scripts in the “*analysis*” folder, and configuration files are at the root. Specifically, in the “*analysis*” folder you will find:

1. Data Ingestion:
    a. Childcare center data: the child_centers_clean.py processes the data regarding ECC ensuring readiness for spatial analysis.
    b. Census data: the census_api.py and census_clean.py modules handle retrieval and cleaning of census data, preparing it for further analysis.

2. Spatial Analysis: merges data sources (spatial_join.py and hav_distance.py), adds distances with the Google Distance Matrix API (distance_matrix_api.py and google_api_request.py) and data cleaning (distance_cleaning.py) of the database to be ready for the Visualization and Simulation.

3. Simulation and Data Visualization: using the simulation algorithm from optimization.py and the dashboard data visualization from app.py.

Elena Porfidia “elenaporfidia”
Miguel Perez “miguelperez”
Joaquin Pinto “joaquinp”
Rodrigo Rivarola Monzon “rrivarola”

- **Description of code responsibilities for each member**

Task	Code	Responsible
Census data gathering and cleaning (Census API)	census_api.py census_clean.py	Joaquin
Designing __main__ file to run the project	__main__.py	Joaquin
Spatial join (census and childcare center data) and gathering data from Google Distance Matrix API	spatial_join.py google_api_request.py* distance_matrix_api.py distance_cleaning.py	Rodrigo
Childcare center's data cleaning	child_centers_clean.py	Miguel
Simulation algorithm	google_api_request.py* hav_distance.py** optimization.py	Miguel
Data visualization	app.py	Elena

* Developed by Rodrigo and Miguel.

** hav_distance.py is based on the haversine distance module used in the course CAPP 30122.

- **Guide on how to interact with the application and what it produces**

Running the Model:

- ➢ Clone the Early Education Repository to access our scripts in the analysis folder.
- ➢ Install all necessary packages via your terminal with “*poetry install*”.
- ➢ Obtain API keys for Census[1] and Google Distance Matrix[2] and insert them into “*CensusAPI_key.txt*” and “*Google_distance_API_key.txt*”.
- ➢ Execute “*poetry run analysis*” from the parent directory to fetch data and launch the Data Visualization dashboard in your browser. There are a few options you can add afterward, however, those are mostly to ensure you do not overwrite the data we have already retrieved, cleaned, and placed in the correct data folder.
- ➢ The options available for running the analysis are:
    - Run Google Distance Matrix API (there is a cost associated with this): “*--googleapi default=False*”
    - Run data cleaning and gathering: *“--gather_data default=True”*
    - Place the data cleaning and gathering output in a separate test folder, to avoid rewriting of the original databases: *“--test default=True”*

Navigating the Dashboard:

1. Interact with the US and Illinois maps to view highlighted data. Hover your cursor over the maps for detailed information.
2. Scroll down and adjust the x-axis and y-axis in the two interactive graphs to explore different data dimensions for demographic factors and different measurements of distances.
3. Run the simulation model by deciding how many new ECE centers to place in Illinois and whether to use an optimized approach or not (the optimized approach takes into account proximity benefits for neighboring tracts, rather than simply placing centers in tracts with the longest current travel times).

[1] A key request for the Census API can be done in the following link: https://api.census.gov/data/key_signup.html
[2] A key request for the Google Distance Matrix can be done in the following link: https://developers.google.com/maps/documentation/distance-matrix/overview?hl=es-419

After selecting these two input parameters, press the *“RUN SIMULATION”* button, wait for Dash to update, and scroll down to see the results.

What does the simulation produce?

The simulation process establishes what would be the optimal census tracts to locate a defined number of centers in Illinois if the only parameter that is taken into account is the driving time to the closest childcare center.[3]

For this purpose, the defined number of centers will be different iterations for the algorithm, in which every iteration will (i) generate a ranking with the census tracts with lower access to childcare centers at the beginning, and (ii) locate a new childcare center, returning the impact in minutes and kilometers for all the census tracts in Illinois in terms of the nearest childcare center for each census tract.

The location of the new center will be defined by the boolean parameter related to “Do you want to optimize?”. If the input is “No”, every iteration of the algorithm will place the new center in the census tract with the longest distance to the closest center. If the input is “Yes”, the algorithm will place the new center in the census tract with the highest estimated impact in terms of distance.

The outputs related to the simulation are the following:

1. Ranking of Census Tracts (list): previous ranking of the chosen census tract for the new center in terms of (highest) driving time to the nearest ECC.
2. Singular Impacts (list): specific impact of placing each new ECC in terms of reduced distance (in kilometers) and driving time (in minutes).
3. Benefited Census Tracts (list of lists): ID of census tracts that will benefit from the placement of each new ECC. The first ID in each list is the census tract where a new child center would be located, the others ID in each list are the census tracts that the new center will be also its nearest ECC.
4. Total Impact (float): overall reduction in travel distance and driving time related to all new childcare centers across all census tracts.

[3] We recognize that the placing decision for new childcare centers is a multifactorial decision rather than a decision that is only defined by the distance to the closest childcare center. In this context, the results of the optimization must be taken carefully and only as a reference of where new childcare centers would have the highest impact on census tracts in Illinois in terms of distance, not as a final decision or suggestion related to the best location for new childcare centers.

- **What the project tried to accomplish and what it accomplished**

The primary objective of this project was to analyze the influence of geographical proximity to early childcare centers on attendance rates for the target children that could potentially be enrolled in ECC despite it not being a mandatory requirement in the United States. Our intent was to identify and address the discrepancy in licensed childcare availability relative to the population.

The accomplished goals of the project are mainly map out ECC accessibility across Illinois census tracts, zip codes, and counties, gather and analyze data on the facilities’ locations, capacities, and the corresponding socio-economic indicators at a census tract level, utilize geospatial methods, and develop an interactive model simulation in order to forecast the impact of introducing new childcare facilities. Moreover, in addition to measuring the disparity in ECE access, our research is possibly providing policymakers and educational planners a useful tool for strategically planning and placing additional facilities. We are implementing a useful and proactive framework that can be used to forecast while meeting community needs before these communities intensify into long-lasting educational deserts.

In more detail, we were able to analyze and visually output how ECE accessibility is not just a matter of distance in terms of a target family being in a rural or urban area (our initial hypothesis) but rather it is deeply intertwined with the specific state’s demographic landscape. For this reason, since we could not access the data for all of the United States - this will be touched upon in the next paragraph - we chose Illinois as a representative State. Our analysis successfully shows that there are census tracts that would have a huge benefit of a strategic intervention to promote that a higher percentage of children can have equitable access to ECE services.

The main things we were not able to accomplish were:

- Perform the analysis for the whole United States. The main reason was that we are using the Google Distance Matrix API, so we are bound by the free credit provided by Google (the whole US would have many more requests for the API).
- Include the childcare center capacity (available seats) in the analysis. This was mainly because we decided to use census tracts as our unit of analysis, and not every census tract has a childcare center inside it (almost 50% of IL census tracts do not). We also thought about using the capacity from the closest childcare center to each census tract, but we considered this imputation could bias the analysis. Thus, we decided to leave the capacity variable out of the analysis.