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

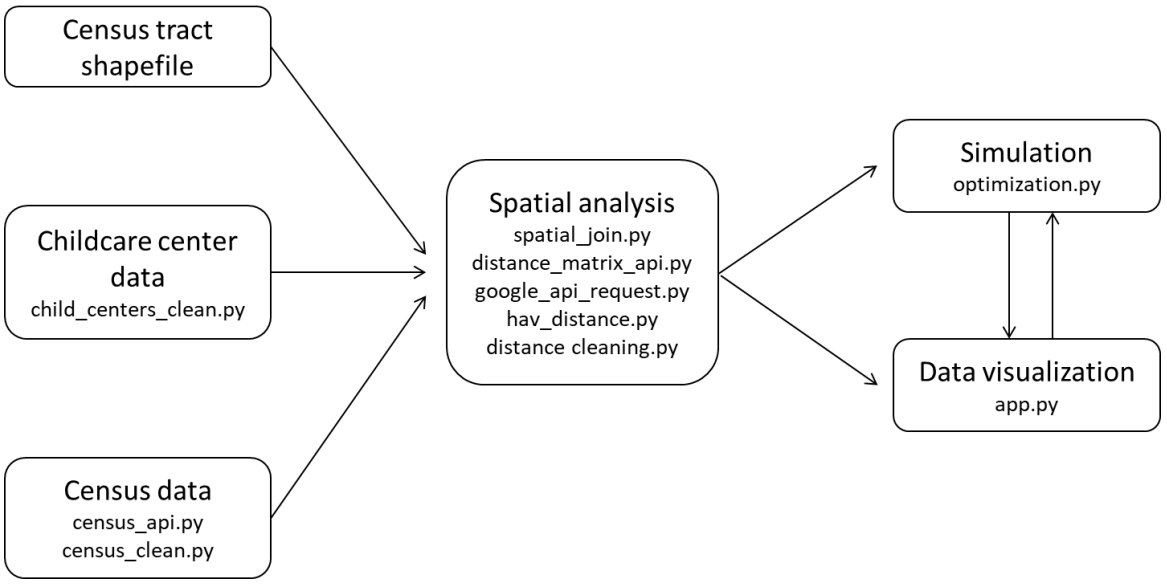
30122 – Los CAPPos – Milestone 3

- **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: consisting of 5 different modules merging our data sources together and performing a variety of different computations like for example the hav_distance.py module, and modules using the Google distance matrix.

3. Simulation and Data Visualization: with the simulation found in optimization.py and the dashboard data visualization in app.py.

- **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 distance_matrix_api.py distance_cleaning.py	Rodrigo
Childcare center's data cleaning	child_centers_clean.py	Miguel
Simulation algorithm	hav_distance.py* google_api_request.py** optimization.py	Miguel
Data visualization	app.py	Elena

*hav_distance.py is based on the haversine_distance from CAPP122
** Developed by Rodrigo and Miguel

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

- **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 and Google Distance and insert them into “`CensusAPI_key.txt`” and “`Google_distance_API_key.txt`”.

- Census API: (https://api.census.gov/data/key_signup.html)
- Google Distance API: (https://developers.google.com/maps/documentation/distance-matrix/overview?hl=es-419)

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.

```
-> Runs Google Distance API, there is a cost associated with this
--googleapi default=False

-> Runs data clean and gather, some of these take a while to run
--gather_data default=True

-> Places the data clean and gather outputs in a separate test folder, to
avoid rewriting what we already have --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). 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?

* Ranking List of Census Tracts: this shows the number in the list of census tracts ranked ordered based on the existing distance to the nearest ECC.

* Singular Impacts: The model provides the specific impact of placing each new ECC in terms of reduced travel distance (in kilometers) and time (in minutes).

* List of All Benefited Census Tracts: all the census tracts that will benefit from the placement of each new ECC.

* Total Impact: the overall reduction in travel distance and time across all benefitted census tracts.

- **What the project tried to accomplish and what it actually 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.

Accomplished goals: 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 it is a fairly representative State to accomplish and make conclusions that can be interpreted nationwide. Our analysis successfully brings to light the issue of this nation having a pressing need for some sort of strategic intervention to ensure 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 US. The main reason was that we were going to use the Google Distance Matrix API, and so we were bound by the free credit provided by Google (the whole US would have many more requests for the API, and also we are considering that it would probably run the code more than once).
- 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.