



dspathways.com

This work is supported by the DS-PATH Summer Fellowship Program under the National Science Foundation Harnessing Data Revolution Data Science Corps Award #2123444, #2123271, #2123313.

# LA City-Youth Development Department ELP(Earn, Learn, Play) Project

Samarth Srinivasa, Rajeswari Pedaballi, Alexis Luevanos, Dammy Obasanjo  
University of California, Riverside  
California State University, San Bernardino



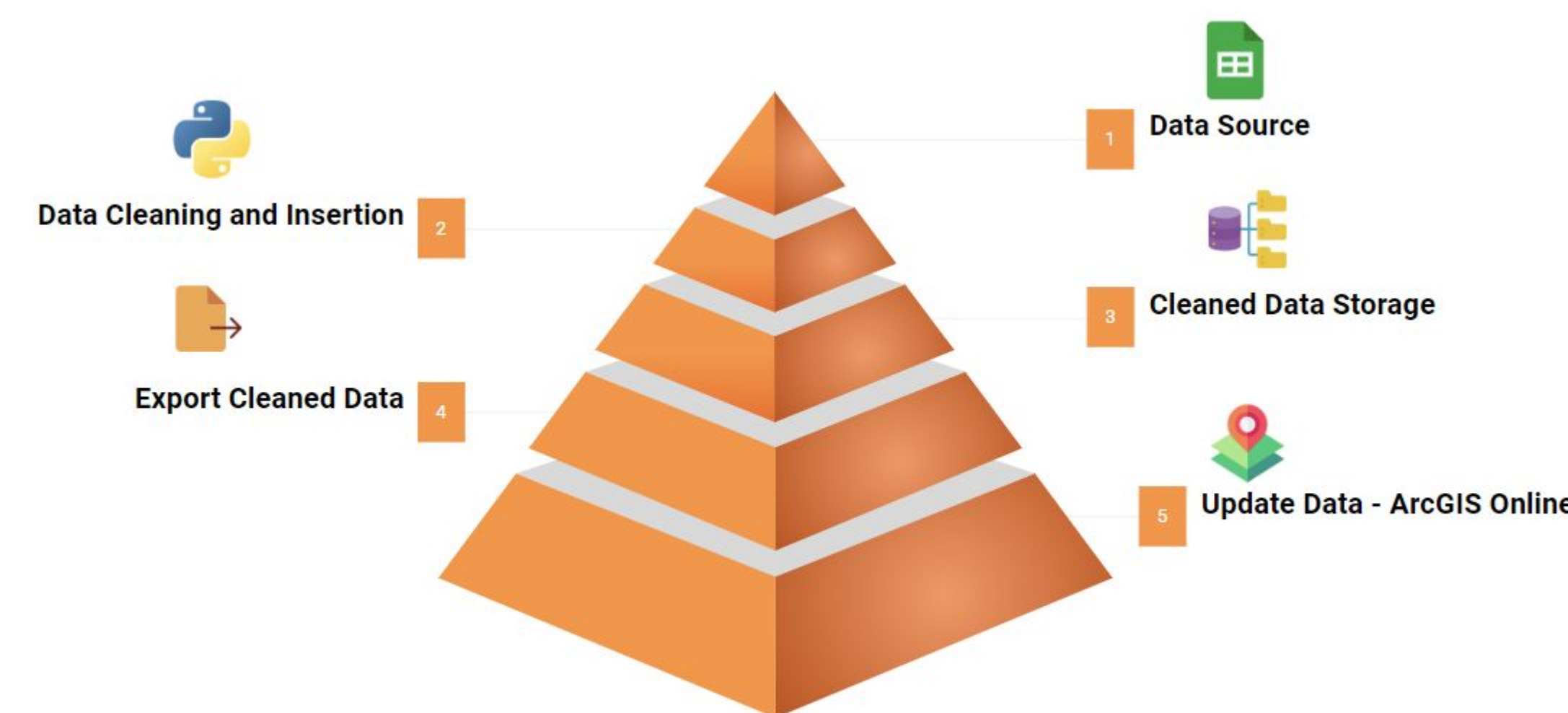
## Background

Earn Learn Play, a project spearheaded by the Youth Development Department, aims to offer summer opportunities for Los Angeles youth. This year, they are collaborating with a greater number of organizations and integrating more data than previously. They require support with data cleaning and consolidation, as well as the development of data pipelines into ArcGIS Online to ensure the map remains updated. The ultimate objective of Earn Learn Play is to become the city's central information hub for youth and their guardians to effortlessly find and enroll in a variety of free and low-cost programs, services, and events provided by the City, community non-profits, and regional partners.

## Project Scope

- The Youth Development Department tasked us with finding solutions to issues that affected the user experience on the ELP website. The ELP Website primarily collected data from other departments using google sheets, which made data standardization a challenging task with how the departments individually collect data and uploaded it to the master sheet.
- As that data was imported into ArcGIS to create the interactive map for the user, there were issues with untitled data being lost, problems with the responsiveness of the built in ArcGIS widgets, and this still required a manual handling of the data pipeline process.
- The goal of this project is to streamline and automate data collection so that for any new data that is added, whether through a database or a direct connection between google sheets and ArcGIS, it can be handled and streamed without error, and to make the widgets more intuitive for the user experience, which would help youth and parents find programs and opportunities easier.

## Methodology



### 1. Data Collection:

- Source: Data is collected from various city departments and stored in Google Sheets. Each department maintains its own sheet with information about youth programs.
- Automation: Google Apps Script is used to automate data collection processes, ensuring regular updates from each department.

### 2. Data Preprocessing:

- Standardization: All text fields are standardized to lowercase to ensure consistency.
- Validation: Each URL is validated to ensure it begins with 'http' or 'https' and points to an active webpage. Fields such as age ranges and costs are checked to ensure they contain valid numeric values.
- Handling Missing Values: Rows with missing values in non-critical fields are flagged for review or removed if they fall below a quality threshold.
- Duplicate Removal: A SHA-256 hash is generated for each row to create a unique identifier. Duplicate rows are identified by matching these hashes and are removed to ensure data integrity.

## Methodology

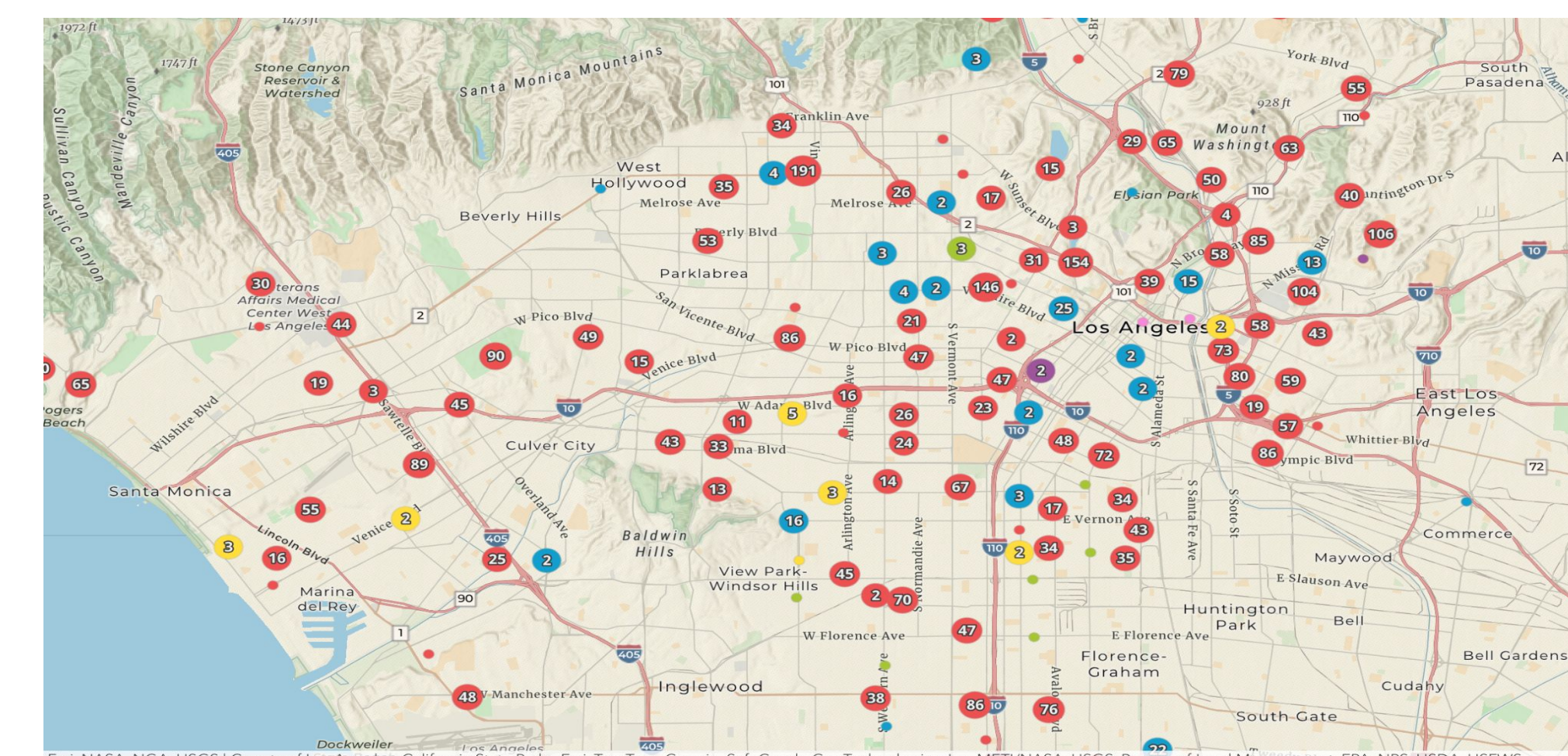
### 3. Data Consolidation and Export:

- Export Process: The consolidated master sheet is exported to a CSV file using a custom Google Apps Script.

### 4. Integration with ArcGIS Online:

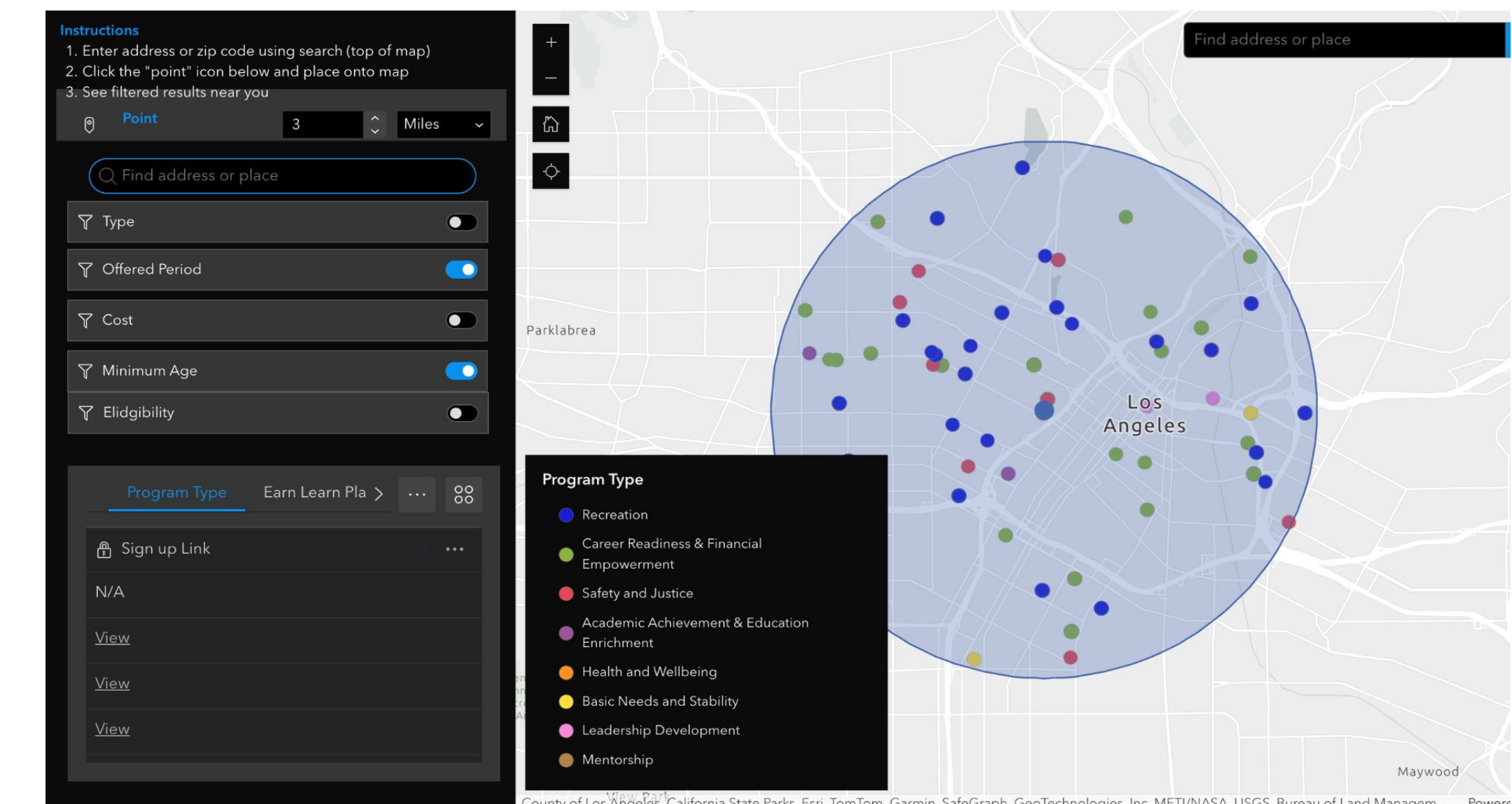
- ArcGIS API for Python: The ArcGIS API for Python is used to automate the upload of the google sheet data to ArcGIS Online as a published web page.
- Creating Feature Layers: The script creates or updates feature layers in ArcGIS Online, making the data available for visualization on maps (through the use of a GeoCode Extension).
- Filtering Logic: Advanced filtering logic is implemented to improve search capabilities. The search algorithm prioritizes key columns.

## Results (Map)



- We were able to provide an aggregation of the clusters of data points pulled directly from google sheets
- This map gives a clear view of amount of programs per location, where each color shows a different program type depending on the department, such as Basic Needs and Stability, Career Readiness & Financial Empowerment, ect

## Results (Website)



In order to make the website more intuitive, we added ArcGIS Features such as Search Filtering Widget, that filters between different data columns, along with a search bar to search for specific programs or activities. For each point, a table is shown in the bottom left with the information for every program in the area, including a link to the website.

## Conclusion

Overall, we were able to configure the website to make for easier user experience to look for programs, and we were able to create a data pipeline that allows for the data to be cleaned and verified before uploaded onto the official website. In the next few weeks, we are working with LA City to to polish and make these changes public to the ELP Project website.

## Acknowledgement

We would like to thank Dr. Salloum and Dr. Flores for giving us this opportunity to participate in the final year of the fellowship. We also thank them for all the knowledge they imparted into the members of this team throughout the fellowship and during the academic year. We would like to thank Daniel Millán, Yessi Pina, Ramon Covarrubias and Emily Niemann (our industry advisors), for all their help during this program and help getting us the resources we needed to accomplish our goals and complete our deliverables. Thank you

## Future

- Over the next few weeks, we are continuing to work with LA city to finalize our tasks:
- Continue refining filtering Logic for the search widget of the website map
- Update the python script and connecting it to the MySQL Database
- Configuring Google Cloud Console for API usage and connecting it to ArcGIS Data pipeline, which would allow for automatic updates of the data from different departments

