Final Report - DATA 205

Project Title: Analysis of Engagement Trends at KID Museum

Name: Piushani Herath

Course: DATA 205

Fall 2025

Sponsoring Organization: KID Museum

1. **Introduction / Project Overview**

This project analyzes participation patterns at KID Museum to better understand how families, students, and schools engage with museum programs over time and across different access pathways. Using historical participation data from General Admissions, Field Visits and After School programs, the project examines temporal trends, attendance behavior, and equity-relevant engagement patterns to support data-informed planning and outreach decisions.

The analysis focuses on Event Date based participation, ensuring that trends reflect actual engagement rather than administrative booking activity. Monthly and weekly participation patterns are visualized through interactive dashboards to highlight seasonal variation, peak engagement periods, and differences in weekday usage. These insights help identify when demand is highest and where opportunities exist to optimize scheduling, staffing, and program delivery.

In addition to temporal analysis, the project integrates geographic and demographic context to support KID Museum's mission of expanding access. Spatial visualizations examine where participating schools and families are located relative to museum sites, while demographic indicators such as Title I, CEP, and FARMs eligibility provide context for understanding equity and access. Together, these views help distinguish between underserved and underrepresented communities and guide more targeted outreach strategies.

The final deliverable is an interactive web based dashboard that brings together engagement trends, attendance outcomes, and equity lenses in a clear and accessible format. By combining data cleaning, aggregation, visualization, and interpretation, this project demonstrates how data science can support evidence based decision making for cultural and educational institutions like KID Museum.

Data: This project uses three primary datasets provided by KID Museum, with analysis focused on General Admissions, After School programs, and Field Trips participation data.

Data Sources: The datasets were sourced from KID Museum's internal booking and attendance records and provided in Microsoft Excel format. Each dataset includes event level records with dates, participation counts, and related metadata.

Included/Excluded Data: For this analysis, records with valid Event Dates and numeric participant quantities were retained, using Event Date as the primary temporal variable and attendance counts to measure engagement. Additional derived fields such as Month, Year, and Day of Week were created to support time-based analysis, and geocoded latitude and longitude fields were included where available to enable spatial mapping and accessibility analysis. Records missing Event Dates, participant counts, or usable location information were excluded, along with administrative fields unrelated to attendance analysis (such as contact details and order metadata), to ensure that the dataset accurately reflects actual participation and spatial engagement trends.

Goals:

- Understand temporal attendance patterns across KID Museum programs by analyzing monthly and weekly participation trends.
- Support operational planning by informing staffing, scheduling, and capacity decisions for General Admissions, After School programs, and Field Trips.

- Identify schools and communities that could benefit most from KID Museum programs using geographic and demographic analysis.
- Reduce attendance attrition by examining differences between booked and actual participation across programs and time periods
- Analyze the KID Museum ladder of engagement to understand how families progress across programs over time
- Deliver clear and accessible insights through interactive dashboards and maps that support data driven decision making across the organization

Tools and Methods: This project uses Python and Pandas for data cleaning, aggregation, and analysis, with Plotly used to create interactive visualizations exported as HTML and hosted on GitHub Pages. Jupyter Notebook supports exploratory and reproducible workflows, while ArcGIS is used to create geospatial maps showing school locations, accessibility, and demographic context. Methods include Event Date based aggregation for monthly and weekly trend analysis, time series grouping across 2023 to 2025, and the creation of interactive charts and maps to support equity focused planning and outreach.

2. Data cleaning and pre processing

Data cleaning and pre processing focused on ensuring consistency, accuracy, and usability across all datasets. Column names and formats were standardized, Event Date fields were parsed and validated, and records with missing or invalid dates or participant quantities were removed. Participation measures were normalized to ensure consistent interpretation across programs, and derived temporal fields such as month, year, and day of week were created to support time-based analysis. Address and location data were cleaned and geocoded where available to enable spatial mapping. Only fields relevant to attendance, timing, and geographic analysis were retained, with full documentation and reproducible cleaning steps provided in the project's GitHub repository. (https://github.com/pherathm/KID_Museum_Web/blob/main/DATA205_KID_MUSEUM_PROJECT.ipynb)

3. Basic descriptive statistics

Basic descriptive statistics provide an overview of participation across KID Museum programs from 2023 to 2025. These summaries focus on the number of events, total and average participant counts, and temporal distributions by month and day of week for General Admissions, After School programs, and Field Trips. Monthly and weekly aggregates highlight seasonal variation, peak attendance periods, and differences in weekday engagement. These statistics establish a baseline understanding of program scale and participation patterns and support subsequent mapping, forecasting, and equity-focused analyses.

4. Results
4.1 Geographic Engagement

The Geographic Engagement component of the final data product visualizes where KID Museum participants and partner schools are in relation to museum sites. Using geospatial mapping, this section displays school locations, attendance density, and proximity to KID Museum facilities such as Fort Totten and the Bethesda Metro Center.
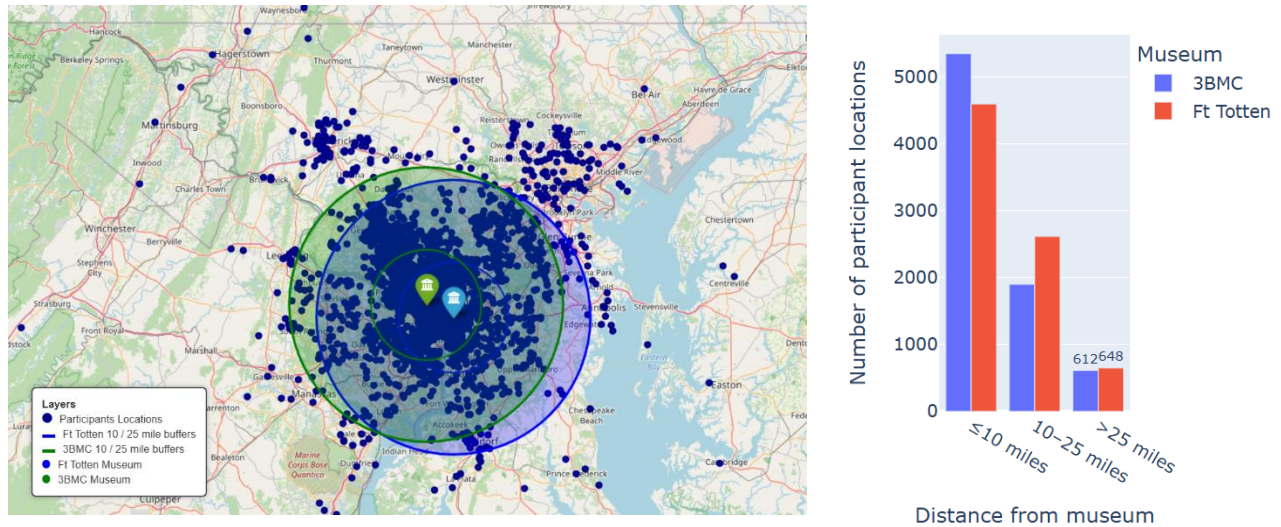

Figure 1. Spatial distribution of participant locations with distance buffers around the Bethesda Metro Center and Fort Totten sites, with a bar chart showing participation by distance category.

The spatial distribution of participants shows strong engagement within 10 miles of both KID Museum sites, with participation declining as distance increases. This distance-decay pattern indicates that geographic proximity significantly influences program access, reinforcing the importance of accessibility in outreach planning.
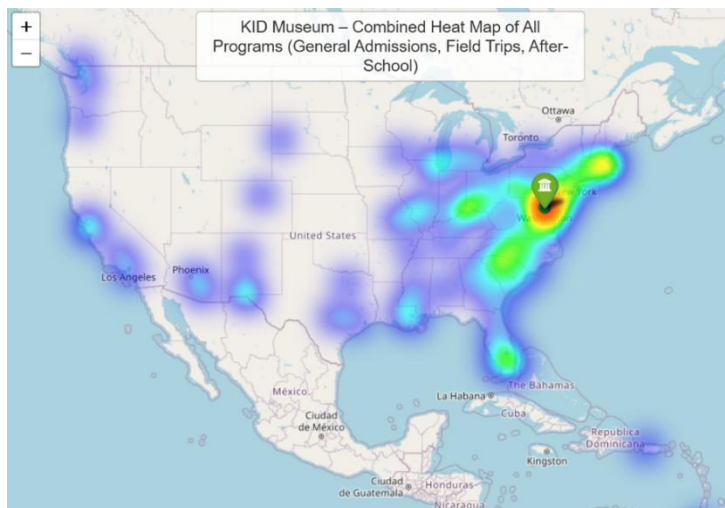

Figure 2. Geographic heat map of participation across KID Museum programs.

The combined heat map highlights dense engagement in the Washington, DC metropolitan region, with additional but less frequent participation originating from other parts of the United States. This pattern reflects KID Museum's strong regional footprint while also indicating broader national reach through field trips, special programs, and visiting families. Together, these spatial patterns help identify areas of sustained engagement as well as regions where targeted outreach could expand access.
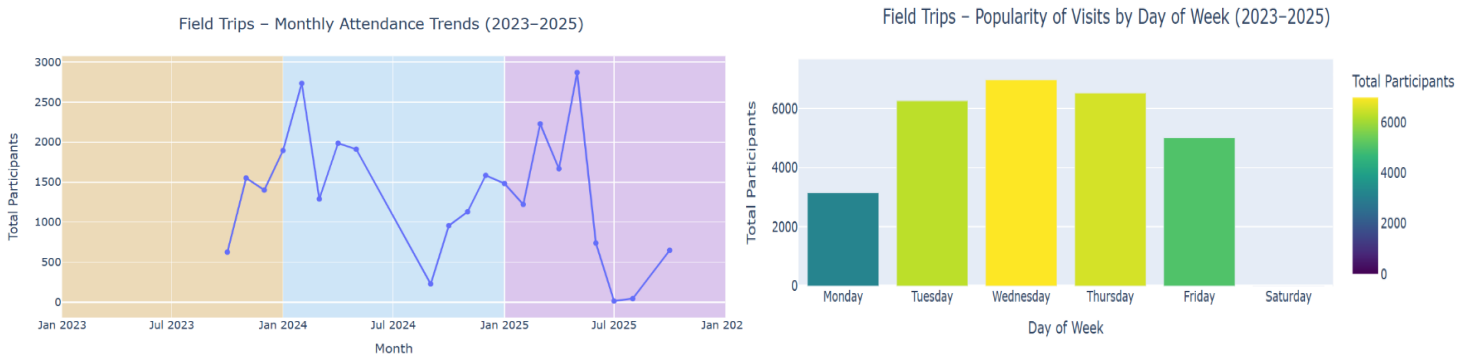
4.2 Trends Analysis

Figure 3. Monthly Field Trips attendance trends from 2023 to 2025 and total participation by day of week.

Field Trips participation shows clear temporal variation across months, with peaks during the academic year and lower attendance during summer periods. Weekday analysis indicates the highest participation occurs midweek, particularly on Tuesdays through Thursdays which reflects school-based scheduling patterns.

4.3 Monthly ARIMA Forecast
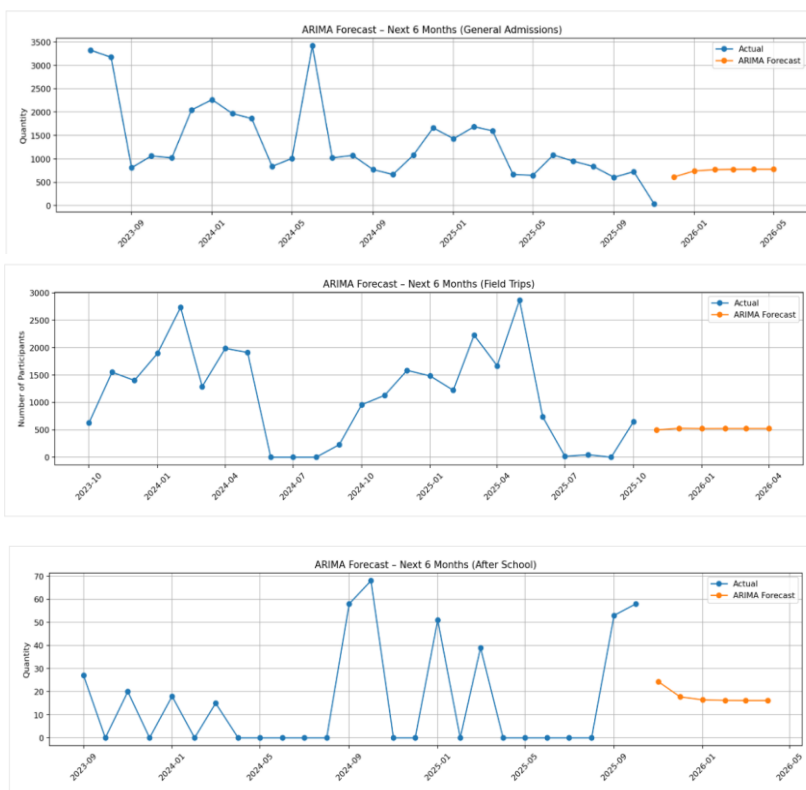
ARIMA forecasts show stable short-term attendance across General Admissions, Field Trips, and After-School programs despite past variability and seasonal spikes. General Admissions has the most consistent demand, Field Trips show school-related seasonality with moderate projected levels, and After-School programs have lower, event-based participation. These forecasts support informed decisions for staffing, scheduling, and resource allocation.

Figure 3. ARIMA forecasts for the next six months of attendance for General Admissions, Field Trips, and After-School programs, shown alongside historical participation.
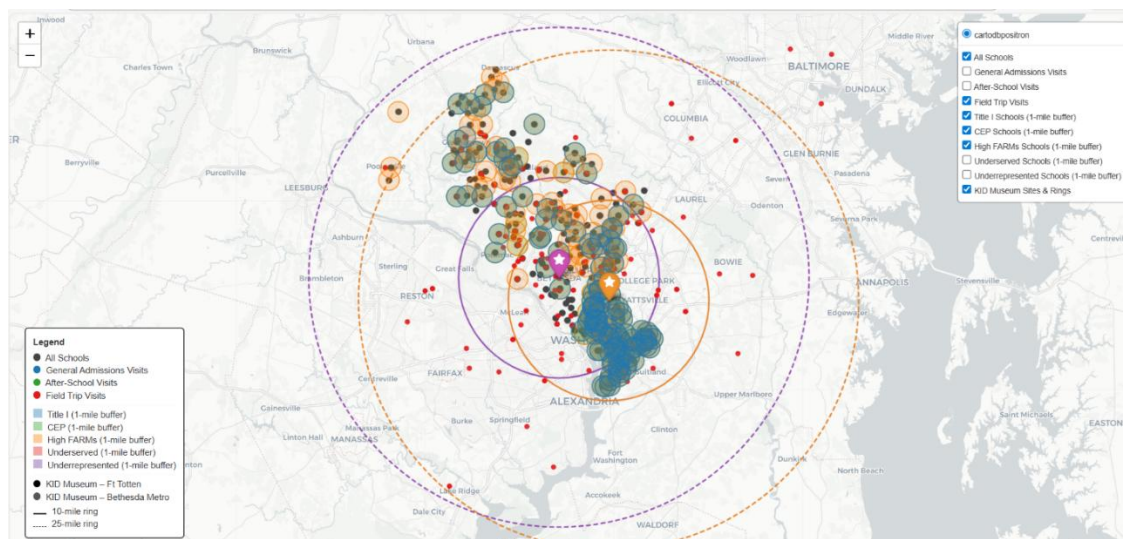
4.4. School Demography & Equity Layers

Figure 4. *School Demography & Equity Layers Around KID Museum Sites*

This figure visualizes participating and potential partner schools overlaid with key equity indicators, including Title I, CEP, FARMs >40%, underserved, and underrepresented classifications within a one-mile buffer of each school. Distance rings (10-mile and 25-mile) around KID Museum sites at Bethesda Metro Center and Fort Totten provide geographic context for accessibility. By combining demographic layers with spatial proximity, this visualization highlights where equity-eligible schools are concentrated relative to museum locations and helps identify communities that may benefit most from targeted outreach and program expansion.
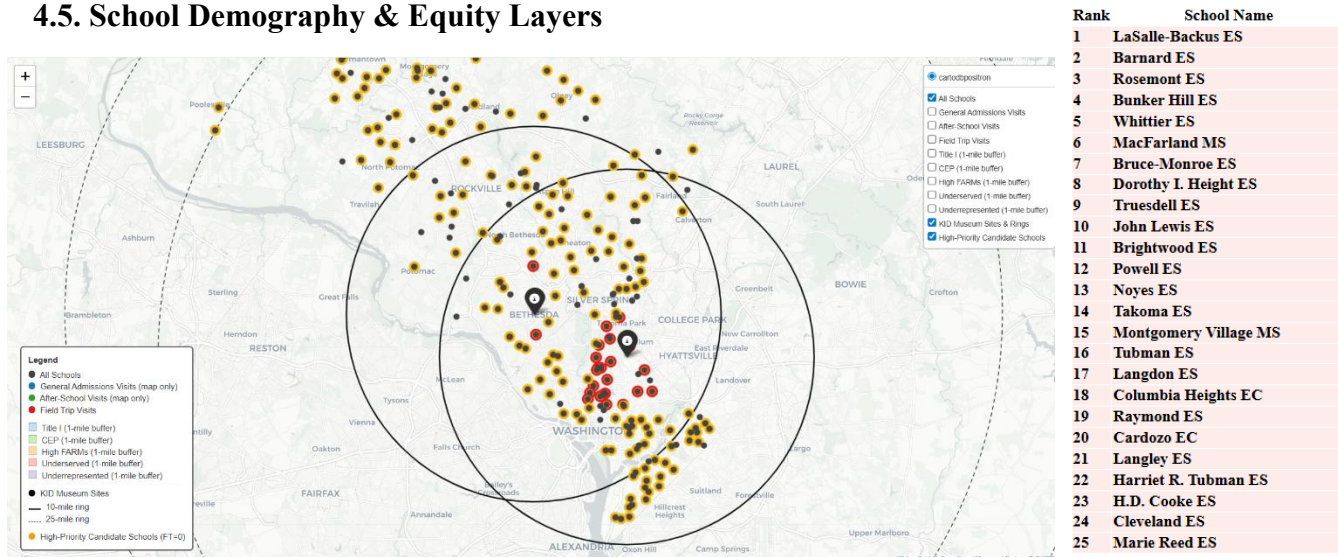
4.5. School Demography & Equity Layers

Rank	School Name
1	LaSalle-Backus ES
2	Barnard ES
3	Rosemont ES
4	Bunker Hill ES
5	Whittier ES
6	MacFarland MS
7	Bruce-Monroe ES
8	Dorothy I. Height ES
9	Truesdell ES
10	John Lewis ES
11	Brightwood ES
12	Powell ES
13	Noyes ES
14	Takoma ES
15	Montgomery Village MS
16	Tubman ES
17	Langdon ES
18	Columbia Heights EC
19	Raymond ES
20	Cardozo EC
21	Langley ES
22	Harriet R. Tubman ES
23	H.D. Cooke ES
24	Cleveland ES
25	Marie Reed ES

Figure 5. *High-Priority Candidate Schools Based on Equity Need and Geographic Accessibility*

This figure identifies high priority 25 candidate schools for outreach using a composite scoring framework that integrates equity need and geographic accessibility. Equity need is measured through the School Need Index (SNI), calculated as the sum of binary equity indicators including Title I status, CEP participation, FARMs >40%, underserved status, and underrepresented status. Geographic accessibility is represented by distance (D) to the nearest KID Museum site (Fort Totten or Bethesda Metro Center). These components are combined into a Final Priority Score (FPS) using the distance-weighted equation $FPS = 3 \times SNI - 0.1 \times D$, which prioritizes schools with higher equity need while moderately penalizing greater travel distance. The map displays schools within 10-mile and 25-mile distance rings, and the ranked list on the right presents the top 25 candidate schools based on FPS, supporting data-driven outreach and program planning.
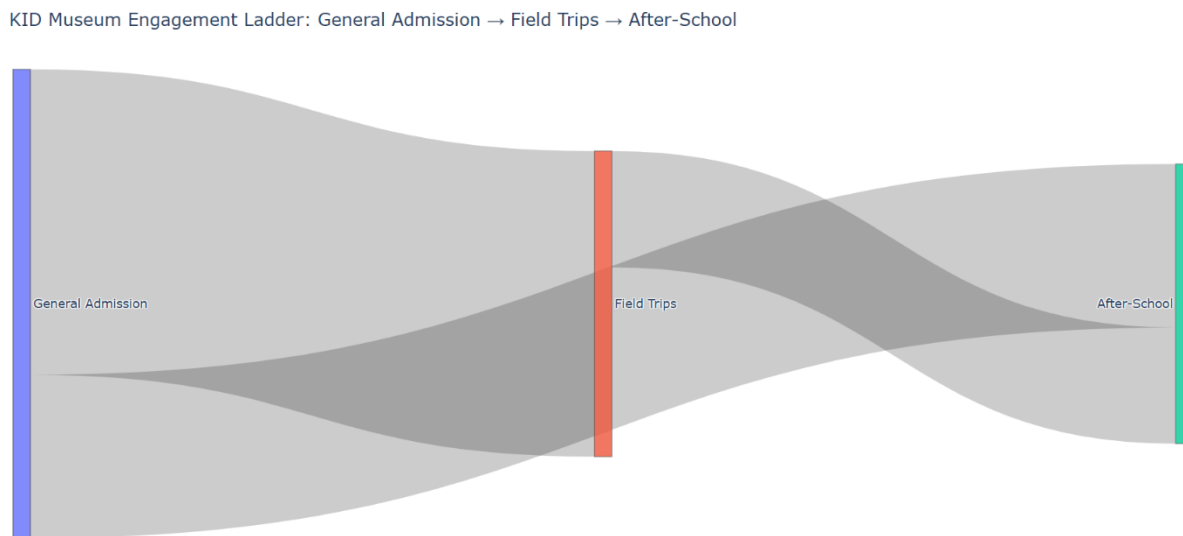
4.6. Programs Engagement Ladder

KID Museum Engagement Ladder: General Admission → Field Trips → After-School

General Admission   Field Trips   After-School

Figure 6. *Figure 4.6. KID Museum Programs Engagement Ladder*

This figure shows how participants move across KID Museum programs, from General Admissions to Field Trips and then to After-School programs. The width of each flow shows the relative number of participants who move between programs, which helps identify common pathways and points where participation decreases. General Admissions serves as the main entry point, and a portion of participants continue to deeper engagement through Field Trips and After-School programs. This engagement ladder explains how early participation can lead to longer-term involvement and supports planning to improve retention across programs.
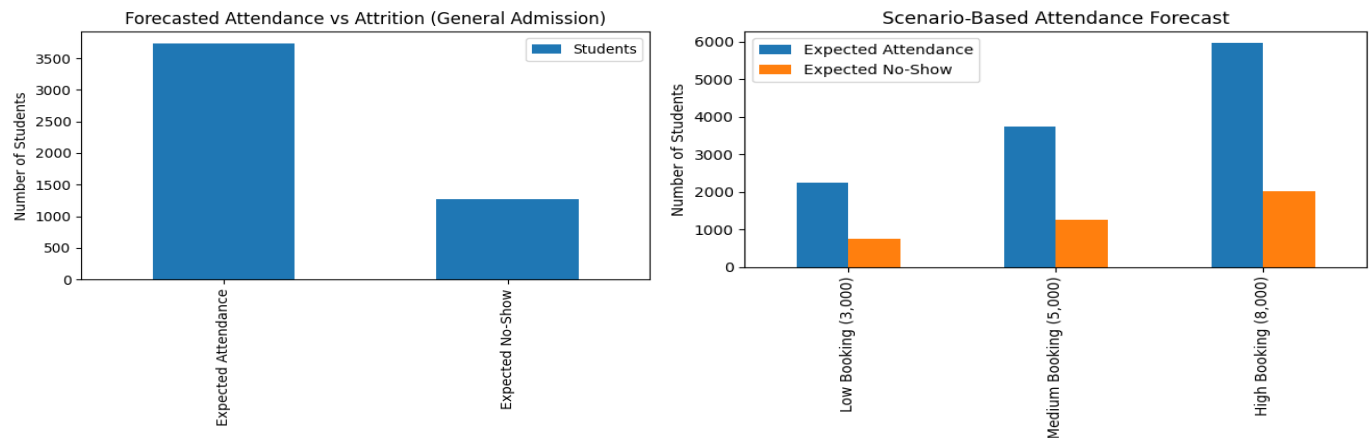
4.7 Attendance Vs. Attrition

Figure 7.*. Attendance vs. Attrition for General Admissions*

Analysis of General Admissions bookings shows that about 74.6% of bookings result in attendance, while approximately 25.4% result in no-shows. This means that roughly three out of four booked participants attend, and about one in four do not convert to attendance. This level of attrition affects capacity planning, staffing, and resource allocation. Understanding attendance and no-show patterns helps improve forecasting and supports strategies such as reminder systems, waitlists, or limited overbooking to increase overall attendance rates.
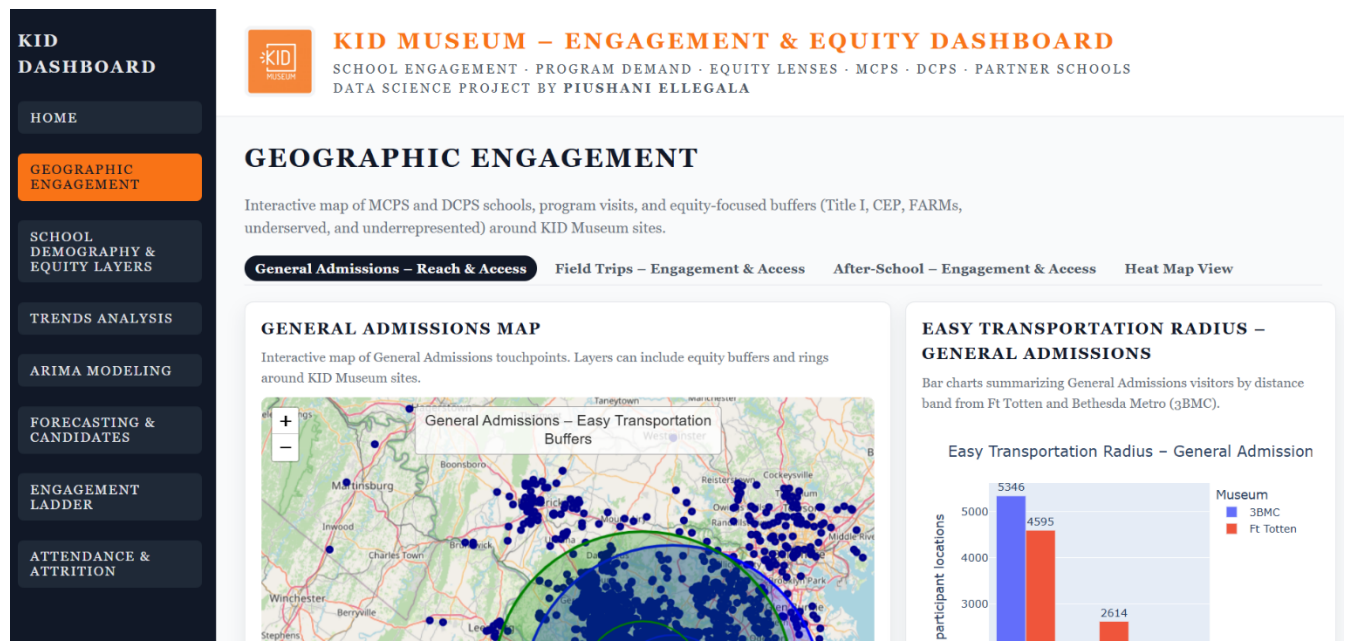
4.8 Dashboarding

Figure 8. *KID Museum Engagement & Equity Dashboard*
https://pherathm.github.io/KID_Museum_Web/

This is the final interactive dashboard developed for the KID Museum project. The dashboard integrates geographic engagement maps, school demography and equity layers, trend analysis, forecasting results, the engagement ladder, and attendance versus attrition summaries. Interactive filters and layered views allow users to explore participation patterns across programs, locations, and equity indicators, supporting data-driven planning, outreach, and decision making.

References and Acknowledgements

This project uses internal KID Museum program data, including General Admissions, Field Trips, and After-School participation records, along with publicly available demographic and geographic data from Montgomery County Public Schools (MCPS) and District of Columbia Public Schools (DCPS). Analysis and visualization were conducted using Python (Pandas, NumPy), Jupyter Notebook, Folium for interactive mapping, ARIMA time-series modeling, Excel for data validation, and GitHub/GitHub Pages for code management and public dashboard hosting. I would like to thank KID Museum staff and mentors for their guidance and data access, my DATA 205 instructor for academic support and feedback, and Montgomery College for providing the resources and learning environment that supported this project.