📘 Project: "Chicago Migrant Resource Pulse" — A Web-Scraped, Real-Time Support Ecosystem Index (2025 Edition)

A multi-source, data-engineered platform that continuously scrapes, cleans, models, and visualizes migrant support activity across Chicago.

This project collects **public, ethical, non-intrusive** data about:

- Shelter availability
- Mutual aid activity
- Food distribution events
- Legal aid announcements
- Volunteer posts
- Trending public conversations about migrant needs

This forms a **searchable, structured, academically valuable library** for social science researchers, city planners, nonprofits, and journalists.

It is **unique**, fully aligned with Adithya's resume, and positions him as someone who can build **public-interest data engineering pipelines** at scale.

📅 6-Week Professional Timeline (Milestones, Instructions, Deliverables)

🔵 Week 1 — Scope, Architecture, and Data Governance Setup

Goals

Define the platform boundaries, identify data sources, and architect the ETL + warehousing pipeline.

Instructions

1. **Select Ethical, Public Data Sources**
    - Chicago mutual aid groups (public posts only)
    - City of Chicago open data (shelter status, 311 calls, resource centers)
    - Public Facebook pages of nonprofits
    - News sites: Block Club Chicago, WBEZ, Tribune
    - Event platforms announcing donation drives or legal aid clinics
    - Twitter/X hashtags related to Chicago shelters, support, migrant aid
2. **Architect the Data Pipeline**
   Adithya's expertise in **medallion architecture** fits perfectly here.
    - **Bronze Layer:** Raw scraped JSON + HTML
    - **Silver Layer:** Cleaned, standardized fields
    - **Gold Layer:** Analytical model with KPIs
3. **Define KPIs to Compute Weekly**
   Examples:
    - Resource availability index
    - Volume of public support posts
    - Frequency of legal-aid announcements
    - Geographic clustering of help requests
4. **Create Project Repository Structure**
5. /bronze
6. /silver
7. /gold
8. /scripts
9. /dashboards
10. /docs
11. /api

Deliverables

- **Project charter**

- **Data governance & ethics document**
- **Data model schema (entity relationship diagram)**
- **GitHub repo initialized**

🔵 Week 2 — Web Scraping Engine (Python + ETL Pipelines)

Goals

Build scrapers for each public data source and begin populating the Bronze layer.

Instructions

1. **Build Python Scrapers for:**
   - News articles (RSS + HTML parsing)
   - Public NGO Facebook pages (Graph API or HTML)
   - Twitter/X (SNScrape)
   - Public Google Events about mutual aid
   - Chicago.gov datasets (REST API)
2. **Store Raw Data in Bronze Layer**
   - JSON files
   - Raw HTML snapshots
   - SQL staging table
3. **Create Automated ETL Jobs**
   - Use Python + cron or Jenkins (skills from resume)
   - Log every extraction in an audit table
4. **Track Metadata**
   - Timestamp
   - Source platform
   - Neighborhood/geo text
   - Entity type (food, shelter, legal, donations, advocacy)

Deliverables

- **Scraping scripts**
- **Bronze layer populated**
- **ETL job schedules (cron/Jenkins)**
- **Data ingestion documentation**

🔵 Week 3 — Cleaning, Standardization & Gold-Layer Modeling

Goals

Move Bronze → Silver → Gold layers using Adithya's data quality + pipeline experience.

Instructions

1. **Data Cleaning Rules**
   - Remove duplicates
   - Standardize Chicago neighborhood names
   - Extract keywords using Python NLP
   - Convert dates into unified formats
   - Remove unrelated content ("ice" weather posts, etc.)
2. **Deduplicate Event-Type Posts**
   Example: multiple posts about the same food distribution event.
3. **Feature Engineering for Analytics**
   - Event category → {shelter, food, legal, mutual aid}
   - Sentiment score
   - Signal strength (engagement)
   - Urgency index (based on keywords like "need," "help," "urgent")

4. **Gold-Layer KPIs**
   Build SQL views:
   - Resource Pulse Score
   - Weekly new events
   - Neighborhood-level support heatmap
   - Top support providers
   - Demand vs. support signal ratio

Deliverables

- **Silver layer (clean tables)**
- **Gold layer analytical views**
- **NLP-based classification notebook**
- **Data dictionary**

🔵 **Week 4 — Predictive & Statistical Modeling Layer**

Goals

Use Adithya's skills in **regression, clustering, and forecasting** to forecast support needs.

Instructions

1. **Cluster Neighborhoods by Support Activity**
   - K-means on event density, sentiment, resource type
   - Identify "high-need" zones
2. **Regression Model**
   Predict weekly support demand based on:
   - News volume
   - Weather
   - Neighborhood socioeconomics
   - 311 call patterns
3. **Time-Series Forecasting**
   Use ARIMA or Prophet to forecast:
   - Surge in support posts
   - Decline in resource availability
4. **Model Export**
   Produce artifacts in /models.

Deliverables

- **Cluster analysis outputs**
- **Regression model results + summary PDF**
- **Forecasting plots**
- **Gold-layer feature store tables**

🔵 **Week 5 — Dashboard Suite + Chicago Research Portal**

Goals

Create an intuitive research interface using Power BI / Tableau.

Instructions

1. **Dashboards to Build**
   - **Chicago Support Heatmap**
   - **Time-series event frequency**
   - **Source-type filters**
   - **Top 25 organizations active weekly**
   - **Must-respond neighborhoods**
   - **Forecast dashboard**
2. **Public Portal Mock-Up**

    - Search bar
    - Neighborhood filter
    - CSV download buttons
    - "What's Happening This Week?" banner
3. **Accessibility Features**
    - Use clear legends
    - Include glossary sidebar
    - Add downloadable PDF summaries

Deliverables

- **Tableau Public dashboard suite** (fully publishable)
- **Research portal prototype**
- **Visual style guide**

🔵 **Week 6 — Publishing & Academic Packaging**

Goals

Package the project as a **free, public resource** suitable for a resume, GitHub, and interviews.

Instructions

1. **Finalize GitHub Repository**
   Include:
    - /scripts
    - /bronze /silver /gold
    - /dashboards
    - /docs/methodology.pdf
    - /models
2. **Write Academic-Style Paper (4–7 pages)**
   Sections include:
    - Background
    - Data sources
    - Ethical compliance
    - Pipeline architecture
    - Key findings
    - Forecast summaries
    - How researchers can reuse the data
3. **Create Resume Bullet-Ready Summary**
   Example bullets (I can generate more):
    - "Engineered a city-scale ETL and analytics pipeline scraping 10+ public migrant resource sources across Chicago, building a 2025 'Resource Pulse Index' used by researchers and nonprofit leaders."
    - "Developed predictive models forecasting weekly migrant support needs with >85% accuracy."
4. **Publish Everywhere**
    - GitHub
    - Kaggle dataset (scrubbed + public-safe)
    - LinkedIn post announcing the tool
    - Optional: Medium article

Deliverables

- **Final GitHub repository**
- **Methodology PDF**
- **Dataset CSVs**
- **Published dashboards**
- **Public announcement post**

🎯 Why This Project Is Perfect for Adithya (Based on Resume)

:

- He has **strong ETL, ELT, data warehousing, and pipeline expertise** — perfect for Bronze/Silver/Gold architecture.
- He's skilled in **predictive modeling (Regression, ML, LSTM)** — used for forecasting migrant support activity.
- He has proven experience in **dashboarding with 20+ KPIs** — ideal for a public research portal.
- He has worked with **large-scale ingestion systems processing millions of records daily** — this project demonstrates similar engineering depth.
- He's built **consumer insights systems** — this adds a **public policy and social research dimension** to his portfolio.
- It positions him as a **data engineer + data scientist + social impact technologist** in one project.

This project is **unique**, **high-impact**, and **employer-impressive**.