



Problem Statement and Objectives

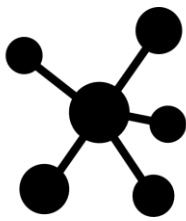
Orofacial clefts are a critical but under-addressed global health challenge, especially in **sub-Saharan Africa**, where access to timely and safe surgical care is limited by geography, infrastructure, and workforce shortages.

Despite the existence of mission-based organizations like Operation Smile and Smile Train, **millions remain underserved**, and there is no data-driven system to strategically expand surgical coverage or forecast future care needs.

Using geospatial data, facility metadata, and demographic projections, our team asked:

How can Operation Smile most effectively expand its surgical network and prepare for future cleft care demand across Africa?

Our aim is to develop a system that identifies high-impact hospital partnerships to:



Expand surgical network coverage



Deliver more cleft surgeries to underserved areas



Strengthen long-term partnerships and regional healthcare infrastructure

Data

Operation Smile Data

Partner Facility Location
Cleft Prevalence & Incident

Smile Train Data

Smile Train Partner Facility Data

WorldPop Data

High Resolution Population Density

IHME Data

2019 Global Burden of Disease Data

WHO Data

Candidate Facility Listing for future partnership

Methodology

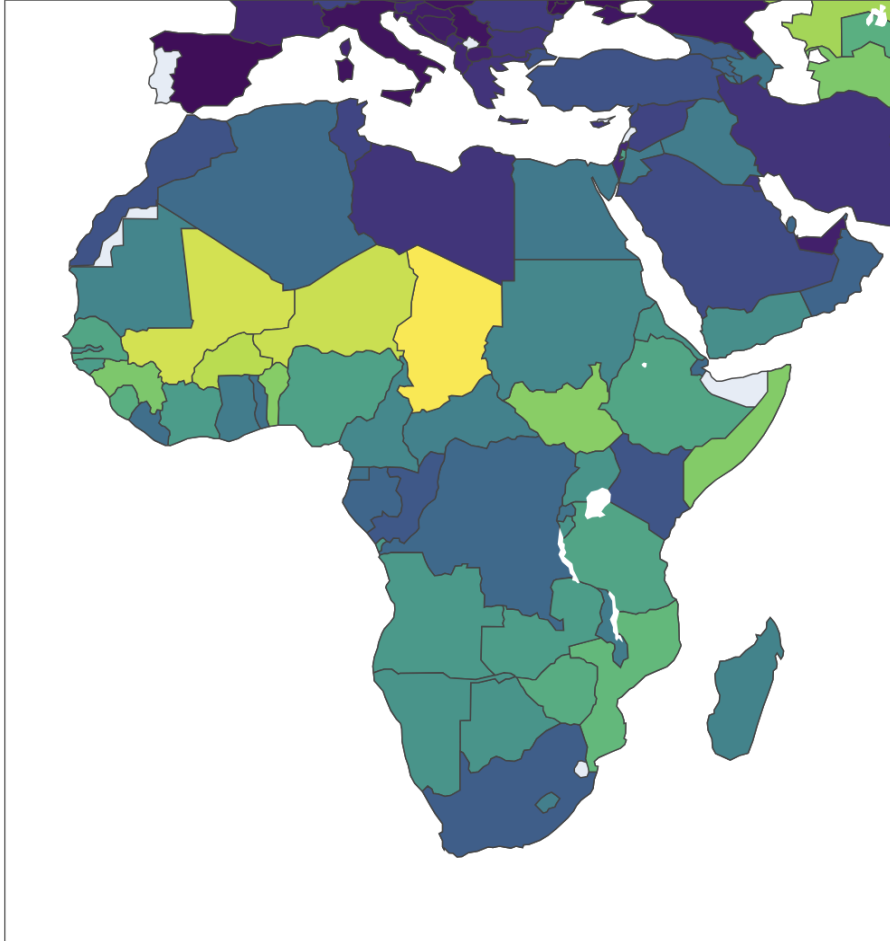
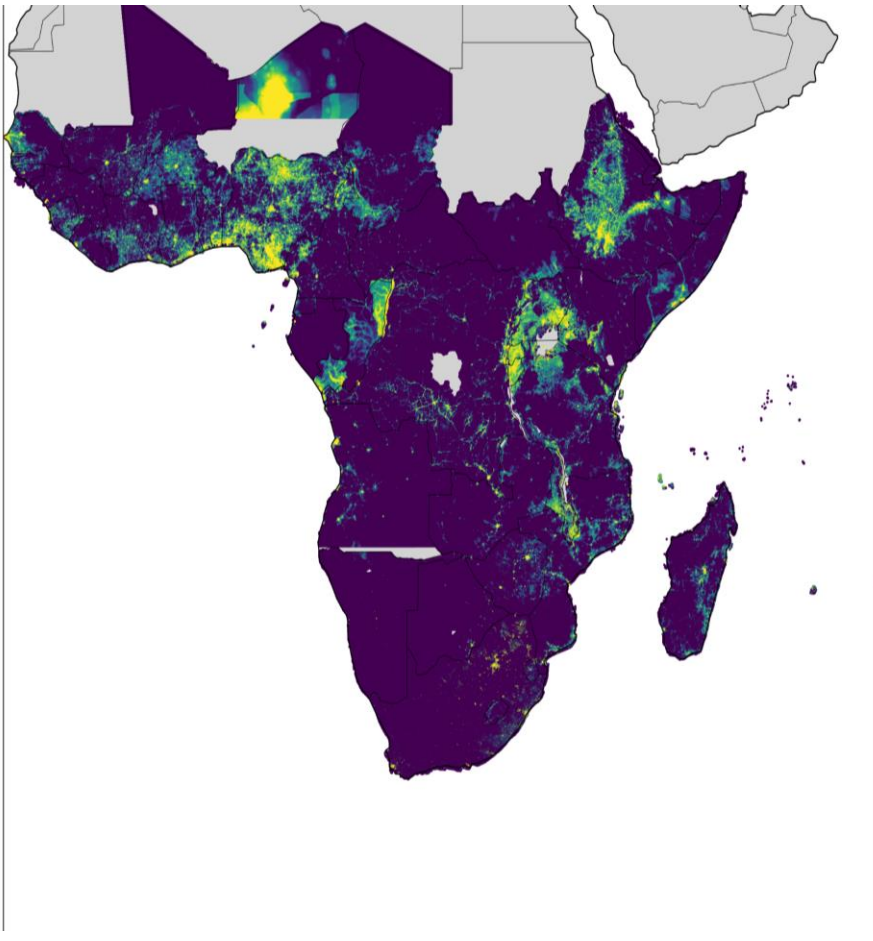
Exploratory Data Analysis/Geospatial Analysis

Factor Analysis

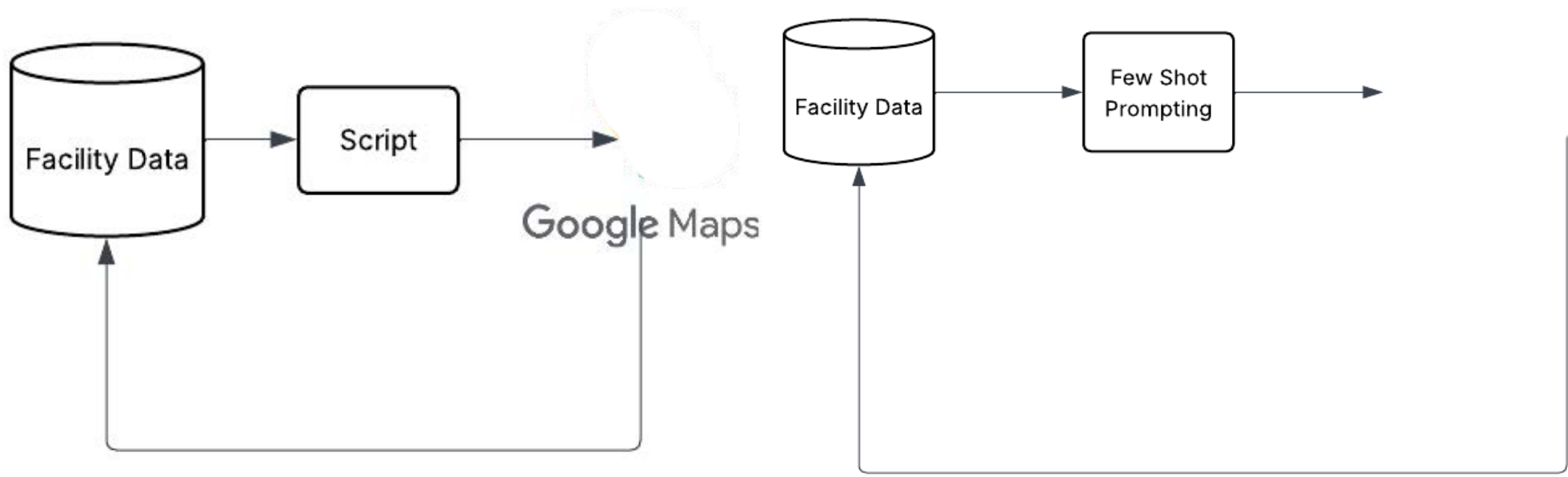
We analyzed a range of country-level indicators from IHME and the World Bank to explore potential drivers of cleft incidence. Among factors such as HAQ Index, smoking prevalence, and air pollution, we found that **GDP per capita** and **adolescent birth rate** showed the **strongest correlations with cleft condition incidence**—suggesting both economic and maternal health linkages. These insights informed the design of our optimization model.

Geospatial Analysis

Using WorldPop’s 1km² gridded population data and cleft incidence rates, we modeled the spatial distribution of cleft burden across sub-Saharan Africa. By overlaying this with hospital locations from Operation Smile and Smile Train, we identified **significant treatment gaps in high-density regions**. We also projected cleft burden growth to 2050, helping prioritize future areas for care expansion.



Data Scrapping & Data Labelling



Scraping Smile Train Facility Data

Used **Selenium** and **BeautifulSoup** to simulate user interaction and extract facility data from the Smile Train website, overcoming bot protection to build a complete candidate facility list.

Location Geocoding with Google Maps

Created a geocoding script using the **Google Maps API** to retrieve and append coordinates to facilities in the WHO AFRO list, enabling accurate spatial mapping and optimization.

Hub-Spoke Labeling

Used **few-shot prompting with Gemini** to classify secondary facilities as **hubs or spokes**, based on surgical capacity, location, and referral role—enhancing realism in network modeling.

Optimization Model

We develop an optimization model to help Operation Smile **choose a set of partner hospitals** to support within a given country. The goal is to **maximize cleft treatment coverage by selecting facilities located in high-burden areas**, while accounting for operational constraints such as **total budget, per-facility capacity, and the cost of capacity expansion**. The budget constraint explicitly balances between selecting more hospitals versus equipping fewer hospitals with greater capacity. In addition to maximizing coverage, the model also promotes a **geographically balanced distribution of facilities across the country**, avoiding clustering and encouraging accessibility to underserved regions.

Sets and Indices

- $I = \{1, \dots, N\}$: indices of potential facility locations
- $E = \{1, \dots, L\}$: indices of existing facilities
- $J = \{1, \dots, M\}$: indices of cleft locations
- $C_j \subseteq I$: potential facilities that can cover cleft point $j \in J$

Parameters

- $\text{covered}_j \in \{0, 1\}$: 1 if cleft point j is already covered by existing facilities
- w_j : cleft burden (e.g., number of patients) at location $j \in J$
- d_{ik} : distance between potential facilities i and k
- d_{nearest} : distance from facility i to its nearest existing facility
- $\alpha, \beta > 0$: weights for distance-based penalties
- B : total available budget
- F : base maintenance cost per facility (includes service for 100 patients)
- c : cost to expand capacity by 50 patients
- c_i : maximum physical capacity of facility $i \in I$

Decision Variables

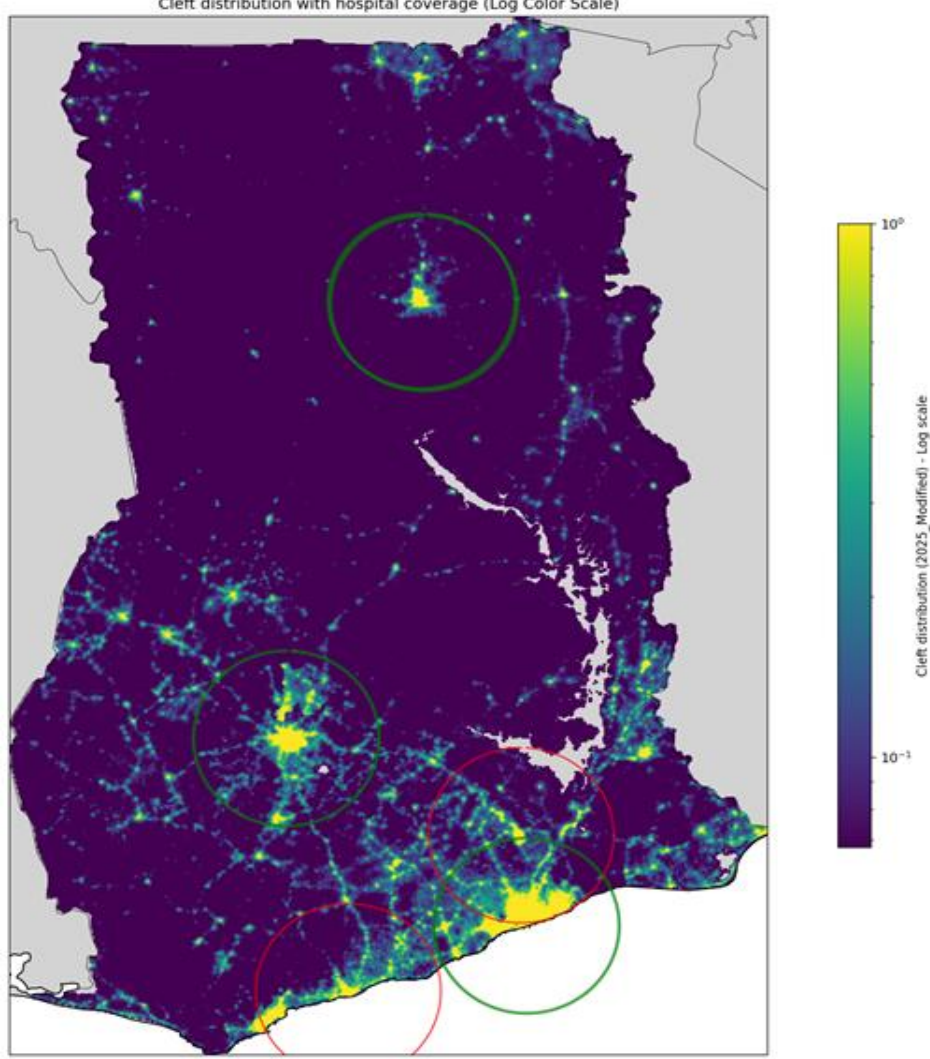
- $x_i \in \{0, 1\}$: 1 if facility i is selected
- $y_j \in \{0, 1\}$: 1 if cleft location j is newly covered
- $z_{ij} \in \{0, 1\}$: 1 if facility i covers cleft point j
- $p_i \in \mathbb{Z}_{\geq 0}$: number of 50-surgery expansions for facility i

Objective Function

$$\max_{x,y,z,p} \sum_{j \in J} w_j y_j + \underbrace{\alpha \sum_{i \in I} d_{i_4}^{\text{nearest}} x_i}_{\text{newly covered cleft mass}} + \underbrace{\beta \sum_{i < k} d_{ik} x_i x_k}_{\text{spread from existing}} + \underbrace{\beta \sum_{i < k} d_{ik} x_i x_k}_{\text{spread among new sites}}$$

Constraints

$$\begin{aligned} \sum_{i \in I} (F x_i + c p_i) &\leq B && \text{(budget constraint)} \\ \sum_{i \in I} x_i &\leq 4 && \text{(select at most 4 facilities)} \\ y_j &= 0 && \forall j \in J \text{ with } \text{covered}_j = 1 \\ y_j &\leq \sum_{i \in C_j} x_i && \forall j \in J \text{ with } \text{covered}_j = 0, C_j \neq \emptyset \\ y_j &= 0 && \forall j \in J \text{ with } \text{covered}_j = 0, C_j = \emptyset \\ z_{ij} &\leq x_i && \forall i \in I, j \in J \\ \sum_{i \in C_j} z_{ij} &\geq y_j && \forall j \in J \text{ with } \text{covered}_j = 0 \\ \sum_{j \in J} w_j z_{ij} &\leq x_i (100 + 50 p_i) && \forall i \in I \\ 100 + 50 p_i &\leq c_i && \forall i \in I \text{ (capacity constraint)} \\ x_i &\in \{0, 1\}, y_j \in \{0, 1\}, z_{ij} \in \{0, 1\}, p_i \in \mathbb{Z}_{\geq 0} && \text{(variable domains)} \end{aligned}$$



Results

Country	Current Population Covered	Current Cleft Covered	Updated Population Covered	Updated Cleft Covered	Added Hub and Spoke	Cleft coverage increase
Kenya	16.7 Million/ 55 Million	9280/23k	33.6 Million/ 55 Million	18652/23k	4 Hub, 1 Spoke	100%
Ghana	8.6 Million/ 33 Million	6155/17k	19.4 Million/ 33 Million	13831/17k	4 Hub, 1 Spoke	125%
Ethiopia	18.6 Million/ 128 Million	15.7k/59.7k	30.8 Million/ 128 Million	26.1k/59.7k	4 Hub, No Spoke	66%
South Sudan	0/ 11.5 Million	0/5.3k	4.6 Million/ 11.5 Million	2151/5.3k	4 Hub, 4 Spoke	INF
Mozambique	1.5 Million/ 33.6 Million	1289/16.7k	7.4 Million/ 33.6 Million	6149/16.7k	4 Hub, No Spoke	377 %