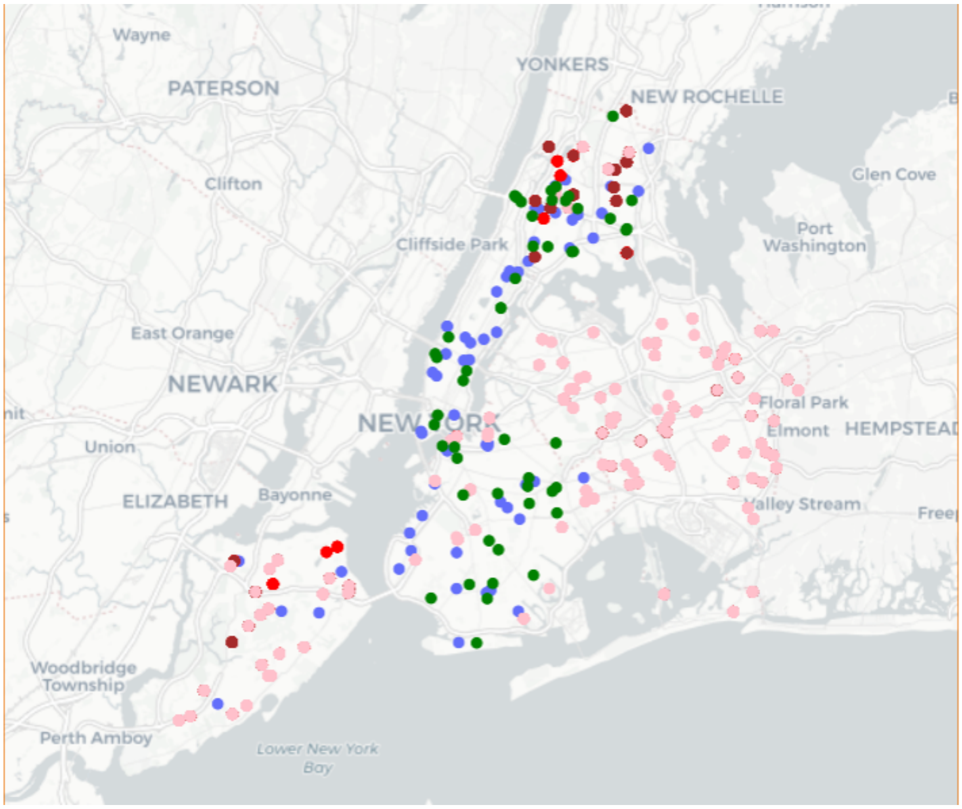


On **Data-Driven Strategies for Optimizing the Planning of Urban Health Centers**

November 2023

CS-GY 6513: Big Data

Aniruddha Jayant Karajgi	*ajk*8815
Saumya Parag Phadkar	*sp*7778
Varad Naik	*vvn*7114

This document serves as **the final report** of our project, done in partial fulfillment of the course, **CS-GY 6513** Big Data. This project is being done under the guidance of Dr. Juliana Freire.

Status Report

Milestones	Status
1. Dataset Collection	Done
2. Analysis	Done
3. Solution Implementation	Done
4. Refinements and Scope Expansion	Done

Table 1: Project Milestones and Status

Milestone 4 Sections marked with this tag were worked on post the mid term report

Introduction

Ensuring access to quality healthcare is paramount as urban areas expand. The strategic placement of healthcare facilities becomes critical to guarantee accessibility for all residents. This project addresses this concern by analyzing the location of medical centers and vaccination providers in urban areas, focusing on New York City. The proposal outlines the project's objectives, including evaluating the relationship between healthcare facility placement and population density, economic factors, demographics, and vaccination rates. Additionally, it explores the evolving nature of healthcare facility locations over time and proposes solutions for optimizing their placement.

The proper distribution of healthcare facilities significantly impacts residents' health and well-being. Factors like population density, economic disparities, and demographic diversity play a crucial role in determining healthcare service accessibility. Beyond fundamental goals, the proposal outlines objectives related to stretch goals, integrating data on weather, disease prevalence, traffic patterns, crime rates, insurance, and food scarcity. The project aims to provide urban planners, healthcare administrators, and policymakers with actionable insights to improve healthcare facility placement in urban areas, with New York City as a primary focus. This endeavor leverages data-driven insights and innovative solutions to address the intricate relationship between healthcare facility placement and various influential factors.

As we navigate through the implementation of our project, focused on optimizing the planning of urban health centers, this progress report marks a pivotal moment in our journey. Access to quality healthcare is an ever-evolving necessity, and our endeavor to strategically analyze healthcare facility locations in urban areas, particularly in New York City, is gaining momentum. This introduction provides a glimpse into the ongoing efforts and achievements outlined in this progress report, offering a holistic view of our commitment to data-driven insights and innovative solutions in the realm of urban healthcare planning.

Problem formulation

In tackling the complexities of urban health center planning, our project's central focus is on developing a comprehensive understanding of healthcare facility locations. This endeavor involves a meticulous analysis of diverse factors, including population density, demographics, education, occupation, and income. These elements serve as the foundation for creating an informed expectation of healthcare locations based on the insights derived from multiple datasets. The continuous monitoring of actual healthcare facility locations over time emerges as a critical component, allowing us to identify anomalies or deviations from the expected patterns.

Beyond our primary objectives, we have set ambitious stretch goals to enhance the depth and breadth of our project. These stretch goals encompass the integration of additional factors such as weather, air quality, crime, traffic, disease, and food availability, thereby enriching our understanding of the urban healthcare landscape. Furthermore, our project seeks to transcend mere location-based recommendations. We aspire to provide valuable insights into specialized healthcare services, explore possibilities for at-home care, and propose and implement performance improvements through more efficient data processing strategies. The comparative analysis involving health improvement indicators, such as decreasing mortality rates, coupled with an examination of healthcare clinics vis-à-vis health expenditure data, aims to uncover potential systemic issues. This multifaceted problem formulation sets the stage for a holistic exploration of urban healthcare dynamics, reflecting our commitment to delivering impactful, data-driven solutions.

Related Work

We found an example of a study related to the use of artificial intelligence in urban planning, although it's not specifically focused on hospital location. This study, titled "Spatial planning of urban communities via deep reinforcement learning," was published in Nature Computational Science. It demonstrates the use of an AI urban planning model to generate spatial plans for urban communities.

The model uses a graph to describe the topology of cities and formulates urban planning as a sequential decision-making problem. This approach was shown to outperform plans designed by human experts in objective metrics and can generate spatial plans for different circumstances and needs. You can read more about this study on the Nature Computational Science website. [1]

We additionally found many research papers that attempt to predict the optimum location for building hospitals. We have listed a few below -

1. [2] **Can fuzzy extension of Delphi-analytical hierarchy process improve hospital site selection?**

   The study explores the application of the Analytical Hierarchy Process (AHP) and its Fuzzy Extension (FAHP) in hospital site selection in rural India, areas where these methods are not commonly used. Incorporating the Delphi method, it compares AHP and FAHP in selecting three potential hospital sites, finding 'cost of land' and 'population characteristics' as key factors.

2. [3] **A Healthcare Facility Location Selection Problem with Fuzzy TOPSIS Method for a Regional Hospital**

   This study developed a decision support model using the Analytic Hierarchy Process (AHP) to identify the best location for a new hospital in Muğla, Turkey. Utilizing AHP, the study assessed all districts in Muğla based on 6 criteria and 19 sub-criteria, with districts ranked using the Saaty scale and analyzed using Super Decisions software. The results identified demand as the most crucial factor for site selection, followed by accessibility, competitors, government policies, related industries, and environmental conditions, ultimately selecting Bodrum as the optimal location.

3. [4] **A Healthcare Facility Location Selection Problem with Fuzzy TOPSIS Method for a Regional Hospital**

   This study proposes a fuzzy logic-integrated approach with multiple criteria for evaluating healthcare facility locations, utilizing a fuzzy Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). The approach is applied in a case study for selecting a regional hospital location in Adana province, with results anticipated to aid future decision-making in this field.

Methodology

Dataset Collection (Oct 24 - Oct 31) Done

The first milestone of our project, spanning from October 24 to October 31, focused on the extensive collection and compilation of datasets. This foundational step is crucial for assessing the feasibility of our objectives and refining our research hypotheses.

Dataset: Health Facility General Information
We found the Health Facility General Information dataset from the Health Facilities Information System (HFIS) to be the most suitable for our project. This dataset is integral to our project as it offers a comprehensive view

of healthcare facilities. This dataset's breadth and depth enable a multi-faceted analysis of healthcare facilities, providing valuable insights for our study. The inclusion of various attributes such as geographic location, capacity, type, and operational specifics enhances the robustness of our research, allowing for a detailed and comprehensive analysis.

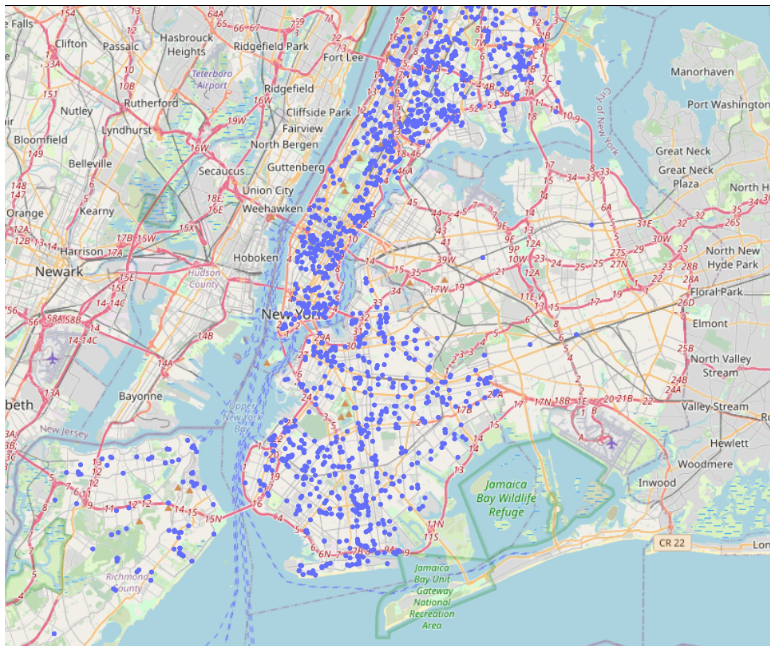


Figure 1: Hospital Locations as of 2023

The Health Facility General Information database provides a comprehensive view of healthcare facilities, encompassing key details for identification and communication, including facility names, addresses, contact numbers, and websites.

It offers extensive geographic and demographic data, such as city, state, ZIP code, county, country, and precise GPS coordinates, enabling a thorough analysis of healthcare distribution and accessibility. The database classifies facilities by type and operational status, includes capacity information, and indicates special services like trauma care and helipad availability.

Furthermore, it provides essential administrative details like data sources, validation methods, ownership, and staffing levels, along with the number of beds, making it a valuable resource for understanding healthcare infrastructure, planning capacity, and shaping public health policies. This rich dataset is instrumental for both broad-scale analyses, such as statewide healthcare capabilities, and detailed, facility-specific evaluations.

The Health Facility General Information datasets [5] are updated weekly.

Dataset: American Community Survey

The American Community Survey (ACS) is an ongoing survey conducted by the U.S. Census Bureau. It provides vital information every year about the American people and workforce. Unlike the decennial census which provides a population count every ten years, the ACS collects detailed demographic, social, economic, and housing data every year. This rich dataset covers topics such as education, occupation, language proficiency, and housing costs, offering insights into the changing dynamics of communities across the United States. The data collected is critical for government and community planning, as it helps determine how more than $675 billion in federal and state funds are distributed annually. The ACS datasets are highly valued by researchers, policymakers, and businesses for their comprehensiveness and reliability, enabling informed decision-making at various levels.

ACS Age Datasets Milestone 4

We obtained the age data using the ACS 5 year estimate for "Sex by Age" datasets. For example, the dataset for brooklyn, for 2021 is[6]. The datasets were available for 2013, up till 2021. We extracted data for all 5 boroughs in New York City for these years. These datasets were at the block group granularity level, since the ACS does not publish this data at the block level. The datasets in it's raw format had the estimates of the number of people falling in various age groups, and categorized by male or female. For example, male under 5 years, male under 10 years, female under 5 years, female under 10 years, etc, till age 85 and above. We calculated the weighted mean age for each block group using this data. We also calculated the total number of people in each block group using these datasets.

ACS Income Datasets Milestone 4

We obtained the income data using the ACS 5 year estimate for "Household Income in the Past 12 Months" datasets. For example, the dataset for Brooklyn, for the year 2021 is[7] . The datasets were available for 2013, up till 2021. We extracted data for all 5 boroughs in New York City for these years. Like the age datasets, we extracted the datasets at the block group granularity level. The datasets in it's raw format had the estimates of the number of people having income in a certain range. For example, number of people with annual income below $10000, $10000 to $14999, upto $200000 and above. We then calculated the weighted mean age for each block group using this data.

One of the ACS datasets, for the year 2021, and for all block groups within Brooklyn are updated annually.

Dataset: TIGER/Line Shapefiles

TIGER/Line Shapefiles are a form of geospatial data provided by the United States Census Bureau. The acronym TIGER stands for Topologically Integrated Geographic Encoding and Referencing System. These datasets are crucial for a wide range of geographic and demographic analyses as they include detailed information about physical features such as roads, rivers, and railways, as well as legal and statistical geographic areas. TIGER/Line Shapefiles are used extensively in mapping and Geographic Information System (GIS) applications, enabling users to visualize and analyze spatial relationships and patterns. They play a key role in urban planning, transportation, and resource management. The datasets are particularly valuable for combining with demographic data from the Census Bureau, such as data from the American Community Survey, to provide comprehensive insights into the socio-economic aspects of different regions. TIGER/Line Shapefiles are updated regularly, ensuring they remain a vital resource for planners, researchers, and policymakers.

These datasets contain area divisions of various levels of granularity-

- **Block** - the smallest geographic area for which the Bureau of the Census collects and tabulates decennial census data.

- **Block Group** - Block groups are the next level above census blocks in the geographic hierarchy. A BG is a combination of census blocks that is a subdivision of a census tract or block numbering area.

- **Census Tract** - Census tracts are small, relatively permanent statistical subdivisions of a county, which are uniquely numbered in each county with a numeric code. Census tracts average about 4,000 inhabitants with populations between 1200 and 800 people.

The TIGER/Line Shapefiles datasets [8] are updated annually.

Additional Datasets Evaluated

1. Emergency Facilities Dataset by HIFLD: Considered but deemed insufficient due to coverage starting only from 2012 onwards.
2. Locations of Health Clinics in 2011 (NYC Open Data): Outdated information – last updated on July 3, 2019.
3. Hospitals in the US (HIFLD): Invalid and missing data.
4. Adult Care Facilities in NY: Missing Data.
5. Vaccine Locations: Briefly considered but not chosen due to the specific focus on vaccination rather than a broader healthcare facility perspective.

Challenges and Considerations

The major challenge encountered was ensuring the comprehensiveness and relevance of the data. Several datasets were reviewed but ultimately not included due to limitations such as data recency, geographic scope, or overlap with the primary dataset. The selection of the primary dataset was driven by its extensive coverage and up-to-date information, aligning closely with our project goals.

Exploratory Analysis (Oct 31 - Nov 6) Done

Overview

Milestone 2, spanning from October 31 to November 6, was dedicated to conducting exploratory analysis of the dataset, primarily focusing on visualizing relationships and patterns within the data. The focus of our efforts was two-fold: ensuring that the datasets we were using lent themselves to modeling, and exploring some avenues for future work we hadn't previously considered.

Assumptions and Project Constraints

1. **Geographical Focus on New York and Its Boroughs:** We narrowed our analysis to New York to ensure a more detailed and region-specific understanding, given the state's diverse healthcare landscape and population density.

2. **Inclusion of Hospitals Constructed Post-1990:** We limited our study to hospitals built after 1980 due to the low and irregular construction rate prior to this. This decision was made to focus on more recent and relevant healthcare infrastructure developments, reflecting modern healthcare needs and standards.

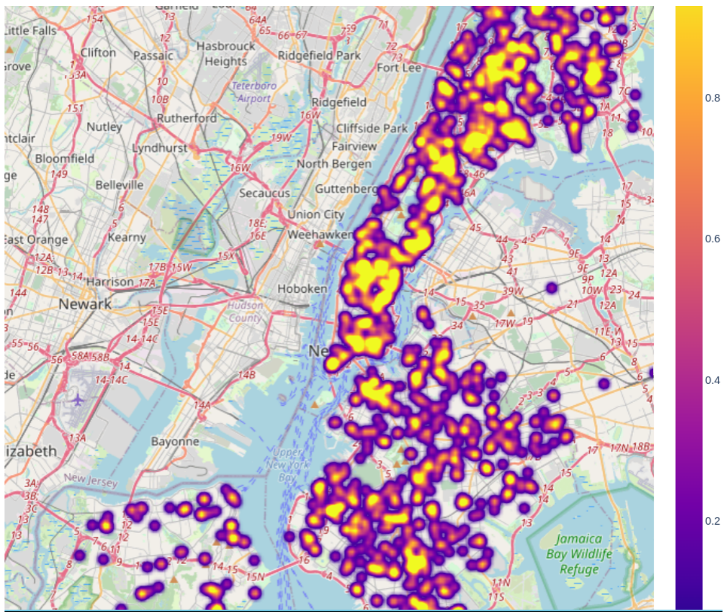


Figure 2: Density Heatmap of locations

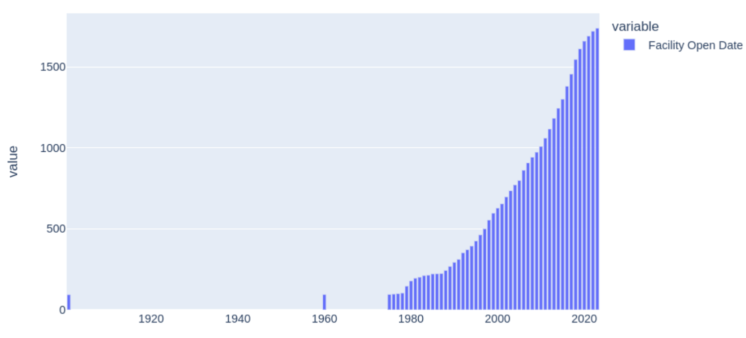


Figure 3: Cummulative Count of Hospitals over time

Preliminary Analysis Results

1. Location and Density Trends: Visual mapping of the facilities revealed distinct concentration points, suggesting these are primary areas for hospital construction, influenced by multiple factors such as population density, accessibility, and urban development. A consistent trend in hospital construction locations was observed over time, maintaining similar latitude and longitude focal points.

2. **Growth Rate of New Hospitals**: A significant increase in the rate of new hospital constructions was noted post-1975, with numbers rising from around 100 to 1691 in 2021 and 1721 in 2022. Mapping these data points across different years (1910, 1980, 2000, 2010, and 2024) demonstrated a substantial increase in hospital density in New York, evolving from sparse distribution in the early 20th century to a more robust healthcare infrastructure in recent years.

3. **Area and population**: We used the population and area data from the ACS and TIGER datasets, and created a heatmap to visualize the distribution for the same.

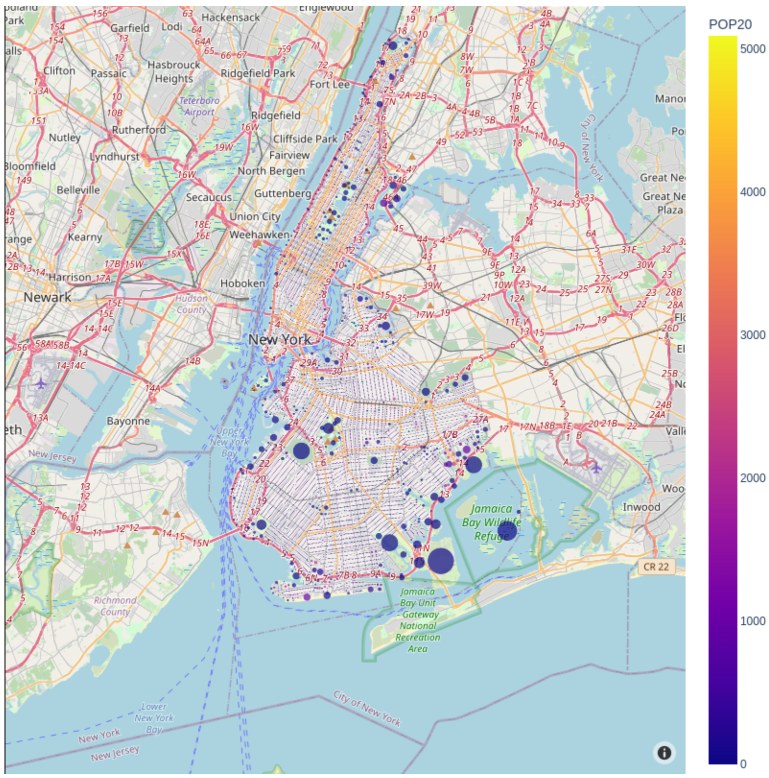


Figure 4: Size of the circle indicates the area of the block, and the colour indicates the population

Building a Solution (Nov 6 - Nov 20) Done

In this project, our central objective is to formulate a solution that optimizes the planning of urban health centers. The process begins with a detailed analysis of the existing solution, leveraging demographic factors. This involves delving into reference papers to comprehend the nuanced factors influencing hospital construction. Recognizing the inherent complexity in modeling hospital construction, a randomized Gaussian model is applied to capture the variability present in the data.

Our overarching goal is to provide an idealistic solution grounded in demographic factors, understanding the challenges of capturing the myriad variables that influence hospital construction decisions. Before proposing our solution, a rigorous comparison with real-world data from the dataset is conducted, acknowledging the multifaceted nature of determinants influencing hospital construction.

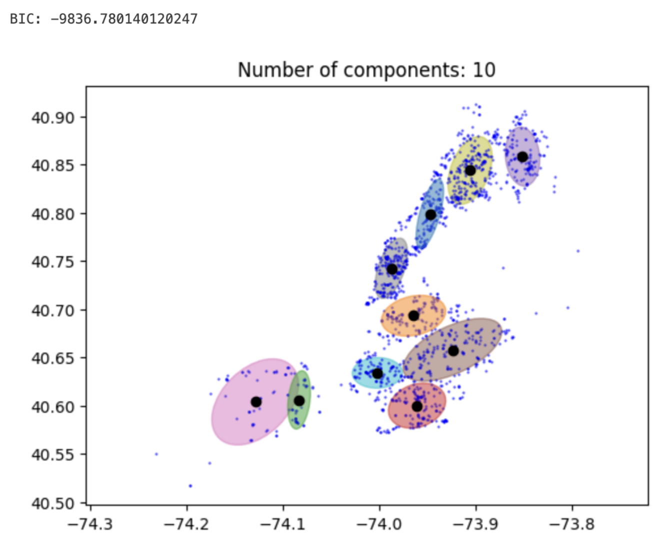


Figure 5: A 10-component Gaussian Mixture Model

Acknowledging the multitude of factors influencing hospital construction decisions, we reference relevant papers and research to gain a deeper understanding of the intricacies involved. Given the challenges of capturing all influencing factors, *Gaussian Mixture Models* (GMM) are employed to interpolate existing data, providing a more nuanced representation. The rough distribution of existing data is then utilized to predict future hospital locations, considering the extended lifespan of hospitals as a less consequential factor for our study.

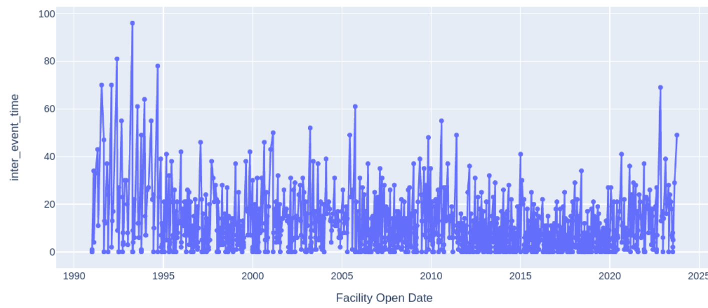


Figure 6: Inter event time of new hospitals

The prediction timing of hospital construction is achieved through the utilization of the Beta distribution, with specific details available in the project notebook for a more comprehensive understanding.

	sumsquare_error	aic	bic	kl_div	ks_statistic	ks_pvalue
halfgennorm	0.002460	1570.759717	1586.813192	inf	0.215661	4.364145e-64
ncf	0.002886	1376.684151	1403.439942	inf	0.215661	4.364145e-64
norminvgauss	0.002893	1501.018342	1522.422975	inf	0.157369	3.628371e-34
invgauss	0.002894	1498.398969	1514.452444	inf	0.157672	2.684997e-34
beta	0.003251	2659.506484	2680.911117	inf	0.215661	4.364145e-64

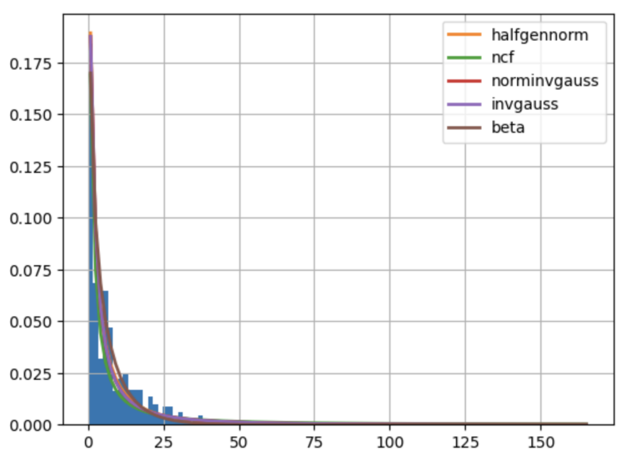


Figure 7: Distribution Fitting on inter-event time

Since these distributions are very similar, we will use the Beta-Distribution – being a more commonly used distribution – for sampling inter-event times.

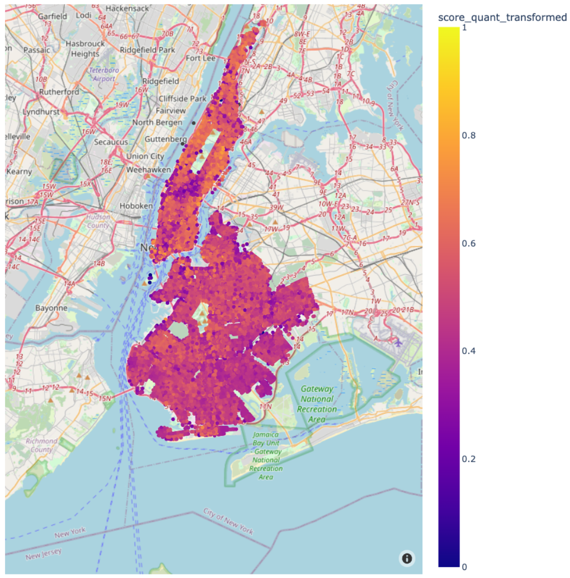


Figure 9: Enter Caption

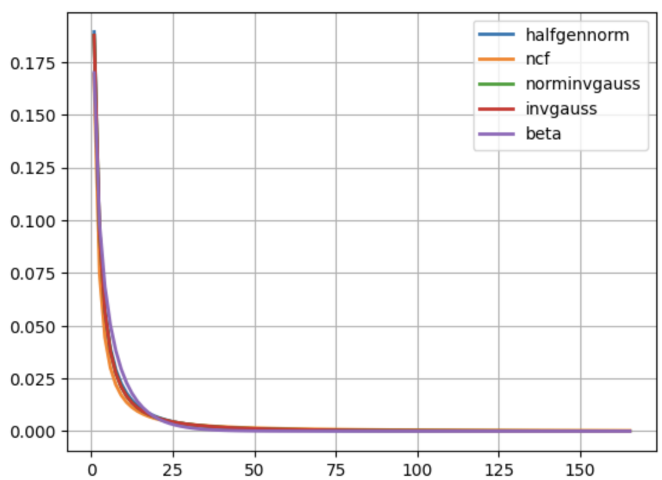


Figure 8: The best fitting distributions

Proposed solution

$$
\begin{aligned}
\mathcal{M}(\text{block\_id}) = l_1 \cdot & \frac{1}{\min(x^2, r^2)} \\
& + l_2 \cdot g(\overline{\text{age}}) \\
& + l_3 \cdot g(\overline{\text{income}}) \\
& + l_4 \cdot g(\overline{\text{hosp\_rate}}) \\
& + l_5 \cdot g(\overline{\text{cardio\_cases}}) \\
& - \lambda \left(\#\text{hospitals inside} \min(x^2, r^2)\right)
\end{aligned} \tag{1}
$$

In crafting our solution, we aim to present a statistical view of the parameters employed. The variables $l1$ and $l2$ are tweakable coefficients allowing the user to adjust the weight of the terms. g is a quadratic function of the mean age of a block. It has also been designed to be tweakable.

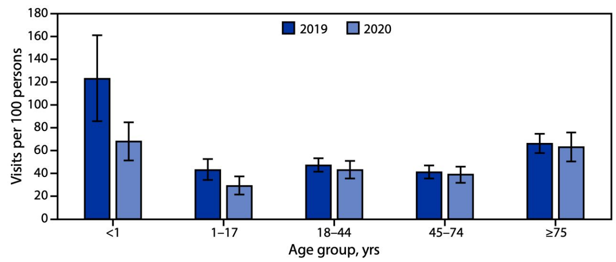
There's also a penalization mechanism for hospitals within a specified range.

The foundation of our solution rests on a well-defined loss function, focused on minimizing the average distance from each block to the nearest hospital. Since its expected that over time, the number of hospitals in an area would be dense enough that the value of the loss would eventually become 0, we added a constraint that we could only include those hospitals constructed within δ years of the current year in the simulation.

Age Function

We assumed the age function by analysing a visits vs age plot, and we came up with the function to be

$$
-2.31854933 \times x + 0.02567851 \times x^2 + 95.60891791326296 \tag{2}
$$



vs age.jpeg

Figure 10: Visits vs age, used to assume age function

Loss Function Milestone 4

1. Loss Function with No Time Constraint

$$\mathcal{L}(y_{\text{true}}, y_{\text{pred}}, \Delta t) = \frac{1}{N}\sum_{i=1}^{N} \min_{j=1}^{M} d(\text{block}_i, \text{hospital}_j) \tag{3}$$

The Loss Function with No Time Constraint, as defined by the above equation, measures the model's performance based on the average minimum distance between each block and the closest hospital from a set of hospitals, irrespective of the hospitals' construction timeline. This loss function does not consider when hospitals were built; it only focuses on minimizing the distance between the given blocks and any available hospitals.

The Loss Function with No Time Constraint is valuable when the temporal aspect of hospital construction is not a critical consideration. It is effective in scenarios where you want to optimize the allocation of new hospitals to minimize the distance to the nearest existing hospitals without regard to their establishment dates. This makes it suitable for various healthcare planning applications.

One drawback of the Loss Function with No Time Constraint is its failure to consider the temporal evolution of healthcare infrastructure. It does not account for the historical significance of older hospitals, potentially leading to inefficient resource allocation if newer facilities are prioritized solely based on proximity. Additionally, this loss function may overlook changes in healthcare demand patterns over time, which could result in suboptimal healthcare planning outcomes.

2. Decay Loss Function

$$\mathcal{L}(y_{\text{true}}, y_{\text{pred}}, \Delta t) = \frac{1}{N} \sum_{i=1}^{N} \min_{j=1}^{M} d(\text{block}_i, \text{hospital\_}decay_j) \tag{4}$$

The Decay Loss Function quantifies model performance based on the average minimum distance between each block and a subset of recent hospitals, giving more weight to newer healthcare facilities. It considers the temporal aspect of hospital construction, focusing on the proximity of blocks to recently established hospitals.

The Decay Loss Function is beneficial when you want to prioritize the impact of new hospitals and prefer healthcare accessibility near more recently built healthcare facilities. It aligns with scenarios where the construction timeline of hospitals plays a crucial role in healthcare planning.

While the Decay Loss Function offers advantages for prioritizing recent hospitals, it has its own set of challenges. It may not fully capture the historical importance of older healthcare facilities, potentially underestimating their role in healthcare access. Additionally, tuning the decay parameter appropriately can be crucial, as overemphasizing recency may lead to an inaccurate representation of healthcare needs in areas with older hospitals.

3. Corresponding Loss Function

$$\mathcal{L}(y_{\text{true}}, y_{\text{pred}}, \Delta t) = \frac{1}{N} \sum_{i=1}^{N} d(\text{block}_i, \text{hospital}_i) \tag{5}$$

The Corresponding Loss Function evaluates model performance based on the average distance between each hospital predicted and hospital actually built. It considers the temporal aspect by aligning each block with its corresponding hospitals.

The Corresponding Loss Function is valuable when you need to understand the healthcare accessibility specific to each block and its corresponding hospitals. It allows for a detailed assessment of the proximity of blocks to hospitals based on their construction history.

> The Corresponding Loss Function, while offering fine-grained analysis, presents certain limitations. It relies on accurate block-to-hospital correspondences, which may require additional data and matching processes. This can be particularly challenging in scenarios with complex healthcare networks. Furthermore, the computational complexity increases with a larger number of blocks and hospitals, making it resource-intensive in such cases.

This comprehensive approach not only seeks to understand the existing solution but also aims to propose an innovative and idealistic model that optimizes hospital planning based on demographic factors.

Results

We set up a simulation that would run from 2020 to 2021. The aim is to compare how different our "ideal" hospital locations are compared to the ground truth. Note that this is not a traditional machine learning task, and our aim is not to match the existing hospital locations as closely as possible. We do not concern ourselves with the realistic distribution of hospitals at all. We simply state the differences.

Hardware Specifications

Apple M2 chip
8-core CPU with 4 performance cores and 4 efficiency cores
8-core GPU
16-core Neural Engine
100GB/s memory bandwidth

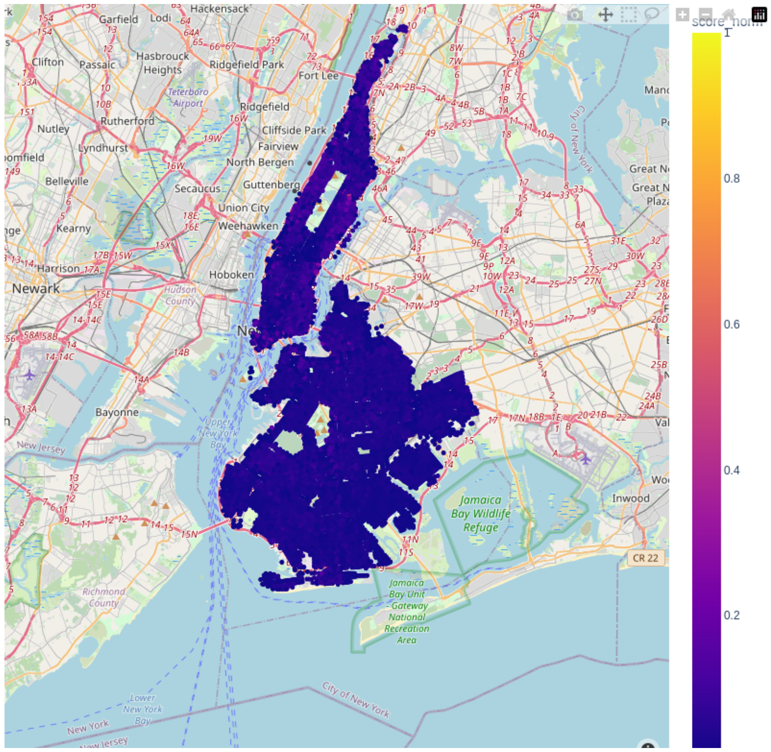


Figure 11: Block Scores without Quantile Scaling

Using the scoring function M, we compute the likelihood of placing a hospital in each census block. The result is hard to interpret due to an expected skew in the distribution of the score. We transform it using a *Quantile-Transformer* so that it is more evenly distributed.

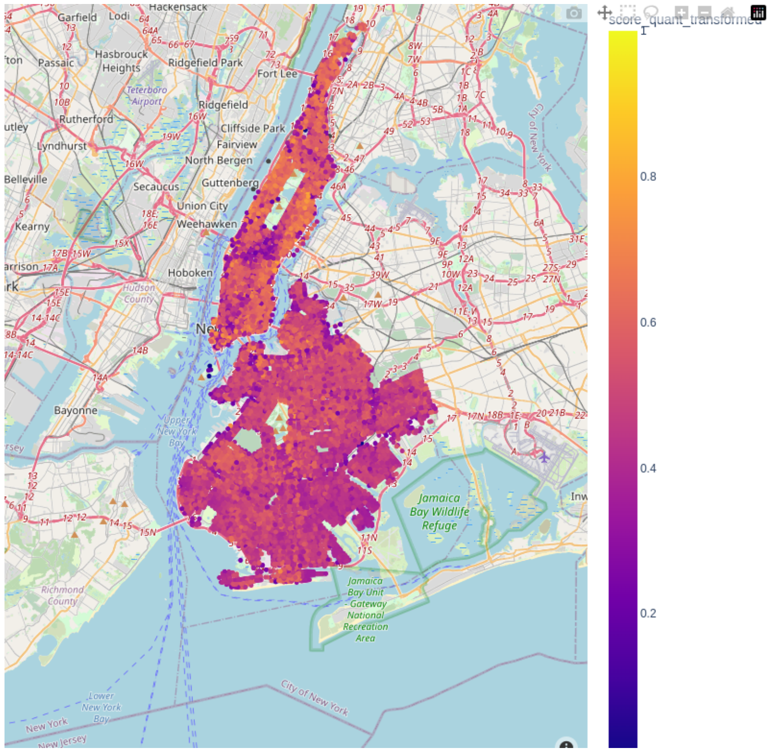


Figure 12: Block Scores after Quantile Scaling

Now that we've scaled and transformed the score distribution, we propose building a hospital at the global maxima. In case multiple blocks end up with a statistically equal score, we randomly a block.

We continue in this way, building hospitals at a uniform rate. The rate of building hospitals is simply the number of total hospitals that are built divided by the intervening years. This assumption has been made because our goal isn't to optimize the number of hospitals placed. Instead, we simply suggest good locations for future hospitals.

At the end of the simulation, we compare the distribution of hospitals that we suggested with those that were built in actuality.

Figure 13: Comparison of Predicted and True Hospital Locations on Map

We expect the value of the loss to increase with the number of hospitals, at least up to a certain point. This is because initially, there are no new hospitals being recommended. As simulated time passes, more recommended hospitals contribute to the loss. Eventually, the system should reach an equilibrium, where the loss is essentially constant.

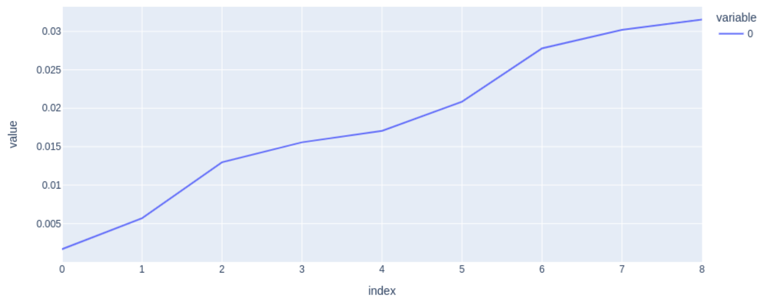


Figure 14: Line Plot of Loss Values

Comparing simulation results Milestone 4

While the entire premise of our work is that there's no ideal set of parameters, we can however look at the differences in results our parameter choices make.

Note
In each of the figures used in this section, the blue dots correspond to hospitals existing before the simulation began. Green dots, on the other hand, represent the ground truth, or hospitals that were actually built during the time the simulation ran. The remaining colors represent variations of the specific parameter being observed.

Varying the *distance_threshold* parameter
Consider the parameter *distance_threshold*. In essence, it decides how large our area of influence is when choosing to build a hospital. Choosing a larger value for this parameter is expected to create a larger spread of hospitals. In Figure 15, we can observe this effect. The lighter colors correspond to higher distance thresholds, which effectively increase the spread of hospitals.

Varying the *population_threshold* parameter
Likewise, if we look at how the parameter *population_threshold* is defined, an increase in the population threshold widens the area of influence of a hospital, meaning that we expect a larger spread. Refer to Figure 16 for this analysis. The lighter dots represent increasingly higher values of the population threshold, increasing from 100 all the way to 800. As expected, this pushes the area of influence of each hospital, leading to a larger spread of hospitals.

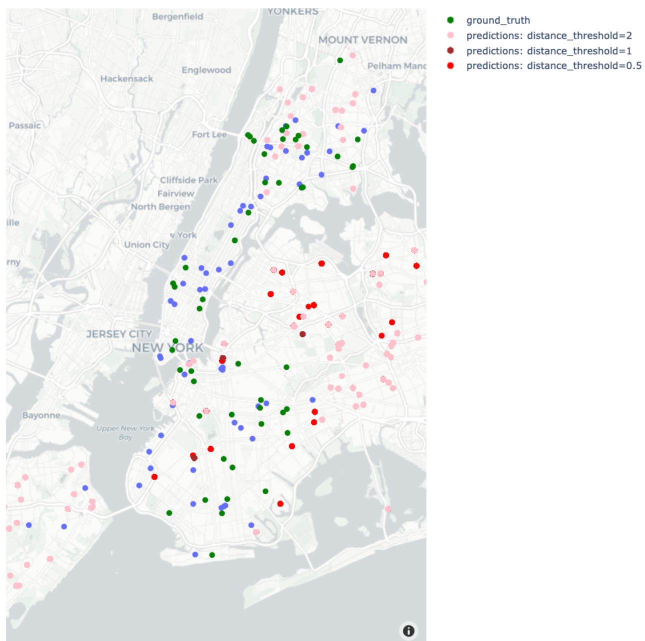


Figure 15: Increasing distance thresholds creates a larger spread of hospitals

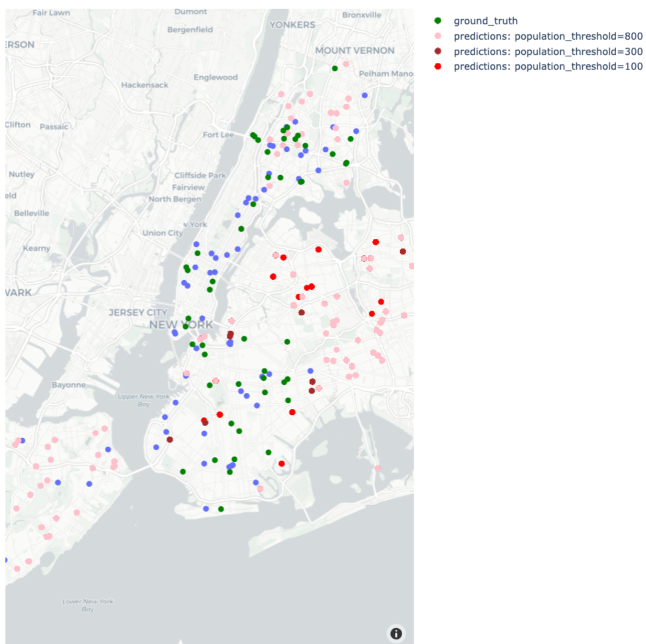


Figure 16: Increasing population thresholds leads to a wider spread of hospitals

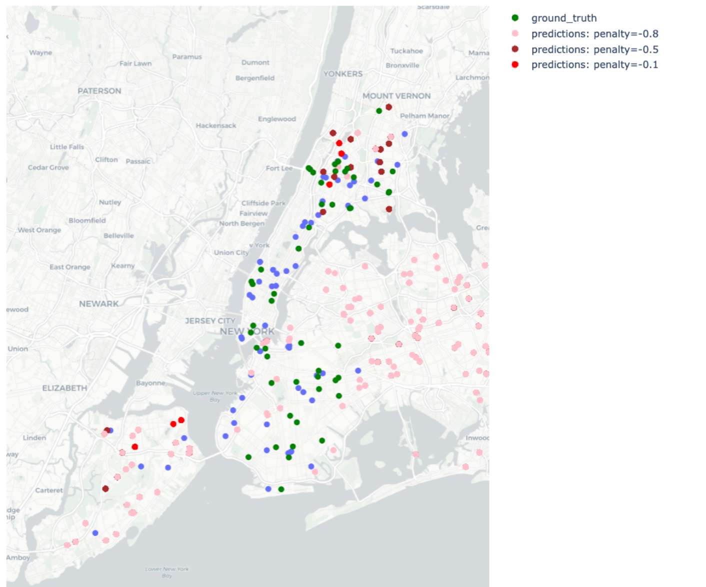


Figure 17: Reducing the penalty increases the main idea of having

Varying the *penalty* parameter
Finally, let's look at another parameter and an arguably important one: the *penalty*. Going back to the definition, it quantifies how much to penalize building hospitals near existing hospitals. Essentially, this means that setting a high penalty would lead to very few hospitals clustering together.

In Figure 17 on page 26, this is observed. Just as before, the green points represent the ground truth, or the hospitals that were actually built, while the blue ones represent previously existing hospitals from before the simulation started. What is of interest are the pink, red, and brown dots. As the saturation decreases, the magnitude of the penalty increases, meaning that lighter dots – corresponding to a higher penalty – are more spread out, while darker ones are clustered together.

In conclusion

In our pursuit of optimizing urban hospital planning, we have conducted a thorough analysis of current healthcare facility locations and their determinants. Our initial models have provided valuable insights into the complex interplay of factors that influence where hospitals are built. However, this is only the beginning of a more comprehensive journey.

Our next steps involve expanding upon the work done thus far, **basing our hypothesis and solution on additional factors** such as weather, air quality, other health indicators, disease prevalence, and food availability. By considering these elements, we expect to capture a more complete picture of the urban health landscape, identifying areas that are not just deserts in healthcare but also in other critical resources.

Furthermore, we aim to **recommend the placement of specific types and sizes of hospitals**. This stretch goal will enable us to tailor our recommendations to the nuanced needs of different neighborhoods, accounting for varying population densities, demographics, and health profiles.

In tandem, we will focus on **improving the performance of our solution** by incorporating more efficient data processing strategies. These enhancements will allow us to refine our model further, providing more accurate predictions that can keep pace with the rapid changes in urban environments.

Ultimately, our work serves as a foundation for a data-driven approach to urban hospital planning. As we build upon this foundation, we remain committed to the goal of providing accessible, high-quality healthcare to urban populations, supporting the development of more resilient and equitable healthcare systems.

References

1. https://www.nature.com/articles/s43588-023-00503-5/
2. https://www.inderscience.com/info/inarticle.php?artid=67428
3. https://www.sciencedirect.com/science/article/pii/S2211883718301540
4. https://dergipark.org.tr/en/pub/ejosat/article/584217
5. https://health.data.ny.gov/Health/Health-Facility-General-Information/vn5v-hh5r
6. https://data.census.gov/table/ACSDT5Y2021.B01001?g=050XX00US36047$1500000
7. https://data.census.gov/table/ACSDT5Y2022.B19001?g=050XX00US36047$1500000
8. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

. . .