Optimizing Resource Allocation to Address Childhood Obesity in London: Integrating Clustering and Linear Programming

Word Count (excluding list of references):	*1831 words*

1. Introduction

Childhood obesity is a pressing public health issue in London, mirroring global urban health disparities. Despite policies like the London Environment Strategy, regional inequalities persist. Research shows obesity clusters in areas with high poverty or limited access to green spaces and fitness facilities (Singh et al., 2010).

This study addresses these disparities by developing a quantitative resource allocation model that integrates clustering and linear programming (LP). By identifying high-risk areas and optimizing resource distribution, the model offers a systematic framework for reducing obesity rates. Historical data is analyzed to evaluate the model's alignment with London's policy changes, providing actionable insights for future interventions.

1.1 Literature Review

London has implemented policies such as the 2018 *London Environment Strategy*, and the 2016 *Sport Unites Programme*, which enhances access to green spaces, and community sports facilities (Greater London Authority, 2023). However, the effectiveness of these measures varies regionally, and few studies have systematically evaluated their efficiency in resource allocation. This underscores the need for a data-driven approach to improve intervention strategies.

Extensive research using regression analysis has confirmed a strong correlation between childhood obesity and factors such as poverty rates and access to green spaces (Singh, Siahpush & Kogan, 2010). While cluster analysis has been widely applied to identify high-risk areas, and linear programming (LP) has proven effective in resource allocation for land-use planning and facility optimization (Chuvieco, 1993), their interdisciplinary application to address childhood obesity remains underexplored.

This study bridges these gaps by integrating cluster analysis to identify high-risk areas with targeted resource allocation optimization for childhood obesity. A resource optimization model based on historical data is proposed, and its decision-support value is evaluated by comparing it with policy trends and outcomes.

1.2 Research Question

a. What are the key socioeconomic and environmental factors driving high-obesity areas, and how can these areas be systematically identified?
b. How can limited public resources be optimally allocated to achieve maximum reductions in childhood obesity rates?
c. To what extent does the proposed resource allocation model align with policy trends in London?

2. Data Preparation

This study focuses on the London region at the ward level, which is the finest publicly available data source.

2.1 childhood obesity rates in London

According to the NCMP, childhood obesity rates fluctuate between 2006 and 2024 in London, peaking at 24.7% in 2013 (NCMP, 2006–2024) before gradually declining in subsequent years. To capture this pivotal moment, I utilize 2013 dataset, combining obesity rates for childhood age to calculate the total excess rate.

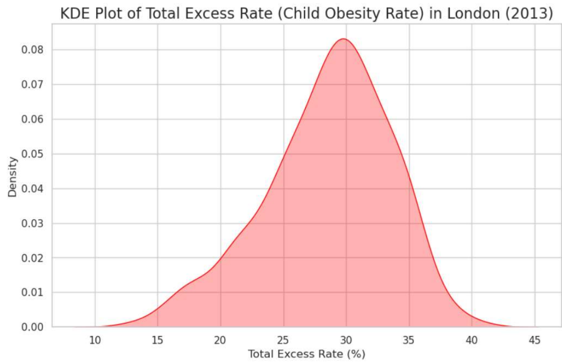

Figure 1 KDE Plot of Total Child Obesity Rate in London (2013)

As shown in **Figure 1**, the total excess rate distribution is concentrated around 30%, with the distribution skewed, further reflecting inequality in obesity influenced by underlying factors.

2.2 Poverty and Environmental Resource Factors

Key variables were selected to analyze the relationship between childhood obesity and socioeconomic and environmental factors. Data cleaning involved handling missing values and normalizing variables for clustering and optimization. All key variables are summarized in **Table 1**.

Table 1 Key Variables

Variable	Type	Description	Source
Childhood Obesity Rate	Numeric	Combined childhood obesity prevalence for ages 4–5 and 10–11.	*Prevalence of Childhood Obesity Dataset (2013).*
Childhood Poverty Ratio	Numeric	The proportion of children in out-of-work households.	*Ward Profiles and Atlas.*
Open Space Rate	Numeric	Open space per household.	*Access to Public Open Space.*
Gym Ratio	Numeric	Gym facilities per 10,000 people in each ward	*Active Places Power (Health and Fitness).*
Cycling Ratio	Numeric	Cycling facilities per 10,000 people in each ward	*Active Places Power (Cycling).*
Athletics Ratio	Numeric	Athletics facilities per 10,000 people in each ward	*Active Places Power (Athletics).*

1. **Poverty Ratio per Child (0–18 years)**: Derived from the *Ward Profiles and Atlas* dataset, was calculated using the **Equation 1**:

$$Children\ poverty\ ratio = \frac{Number\ of\ children\ (0\text{–}18)\ \ in\ \ benefit\ claimant\ households}{Total\ Number\ of\ Children\ (0-18)} \tag{1}$$

2. **Open Space per Household**: Directly obtained from the *Access to Public Open Space and Nature* dataset, using 2013 data, representing the availability of open space per household at the ward level.

3. **Infrastructure Metrics**: Three infrastructure-related variables—Gym, cycling, and athletics facility counts—were calculated to measure the availability of health and fitness facilities per 10,000 people. These metrics were derived from the *Active Places Power* dataset (Sport England, 2023) by filtering for facilities built before 2013. The calculation for each variable is as follows:

$$Facility\ ratio = \frac{Number\ of\ facilities\ in\ a\ ward}{Ward\ population/10{,}000} \tag{2}$$

3. Methodology

3.1 Clustering Analysis

To identify high-obesity areas and their characteristics, K-Means clustering was applied, a reliable and widely used method for partitioning data based on similarity (Jain, 2010). This method was chosen for its efficiency in handling large datasets and its ability to produce interpretable groupings. The analysis incorporated childhood obesity rates, poverty rates, and infrastructure metrics to group wards into clusters, ensuring a comprehensive understanding of the underlying patterns.

The optimal number of clusters was determined via both the elbow method and silhouette scores, ensuring statistically meaningful groupings. The identified cluster with the highest obesity rates was prioritized for intervention.

3.2 Multivariate Linear Regression

To quantify the relationship between childhood obesity and potential drivers, a multivariate linear regression model was constructed. The regression model is expressed as:

$$\text{ObesityRate}_i = \sum_{j=1}^{5} a_j \cdot X_{ij} + \epsilon_i \tag{3}$$

Where:

ObesityRate_i: Childhood obesity rate in ward *i*.

X_{ij}: The value of the j^{th} explanatory variable in ward *i*. Specifically:

- X_{i1}: Childhood Poverty ratio.
- X_{i2}: Open space availability per household.
- X_{i3}: Cycling facility ratio per 10,000 people.
- X_{i4}: Gym ratio per 10,000 people.
- X_{i5}: Athletics facility ratio per 10,000 people.

α_0: Regression intercept (constant term).

α_j: Regression coefficient representing the influence of the j^{th} variable (j=1,2,3,4,5).

ϵ_i: The residual error term for ward *i*.

To ensure robustness, insignificant variables (p>0.05) were removed iteratively, and the model was re-optimized to derive the final significant coefficients used for subsequent analyses.

3.3 Linear Programming Model

To optimize resource allocation for obesity intervention, a LP model was constructed. The process consists of two key steps: conducting a sensitivity analysis to determine the optimal budget, followed by using this budget to derive the final resource allocation plan.

Step 1: Evaluating Budget

The first step is to perform sensitivity analysis by evaluating how different levels of budget impact the expected obesity reduction (Z). The LP model is expressed as follows:

$$\text{Maximize:}\, Z = \sum_{i=1}^{n} (a_1 \cdot x_i + a_2 \cdot y_i + a_1 \cdot z_i + \cdots) \tag{4}$$

Subject to:

1. Budget Constraint

$$\sum_{i=1}^{n} Cost(x_i\,, y_i, z_i \,\ldots) \leq Budget \tag{5}$$

2. Land Availability Constraint:

$$x_i, y_i, z_i \ldots \leq \text{AvailableLand}_i\ \forall_I \tag{6}$$

3. High-Obesity Area Minimum Allocation:

$$xi + yi + zi + \ldots \geq 10 \text{ for high-obesity areas} \tag{7}$$

Where:

Z: Total expected obesity reduction, calculated as a weighted sum of resource allocations across all wards.

x_i, y_i, z_i: Units of different resources allocated to ward i:

- x_i: Number of children or families lifted out of poverty through certain interventions (e.g., access to social services).
- y_i: gym facilities.
- z_i: green space resources.
- …

α_1, α_2, α_3 …: Regression coefficient representing the influence of the j^{th} variable.

Cost (x_i, y_i, z_i …): The total cost of resources x_i, y_i, z_i …

AvailableLand$_i$: The maximum amount of land that can be used for resources in ward ii.

Through testing, try to find the optimal budget:

1. Define a range of budgets, e.g., Ranging from 10 million to 1 billion pounds.
2. For each budget value, solve the LP model to maximize Z, keeping within the current budget.
3. Plot a Budget vs. Obesity Reduction (Z) curve to visualize the relationship. The inflection point on the curve represents the optimal budget, where the allocation is most efficient.

Step 2. Resource Allocation with Optimal Budget

Once the optimal budget is determined, the LP model is solved again with the modified budget constraint:

$$\sum_{i=1}^{n} Cost(x_i\,, y_i, z_i \,\ldots) \leq Optimal\ Budget \tag{8}$$

Applying the final allocation of x_i, y_i, z_i … across all n wards, balancing resource priorities, spatial constraints, and funding limits. The remaining constraints (land availability, prioritization of high-obesity areas) remain unchanged.

This integrated optimization approach ensures that public resources are utilized effectively to maximize impact on obesity reduction, aligning with spatial and budgetary constraints.

4. Results

Using the elbow method **(Figure 2)**, we selected k=5 as the optimal number of clusters based on the point where the within-cluster sum of squares shows a noticeable decline before leveling off. The silhouette scores further confirmed k=5 with the highest average score, indicating well-separated clusters.

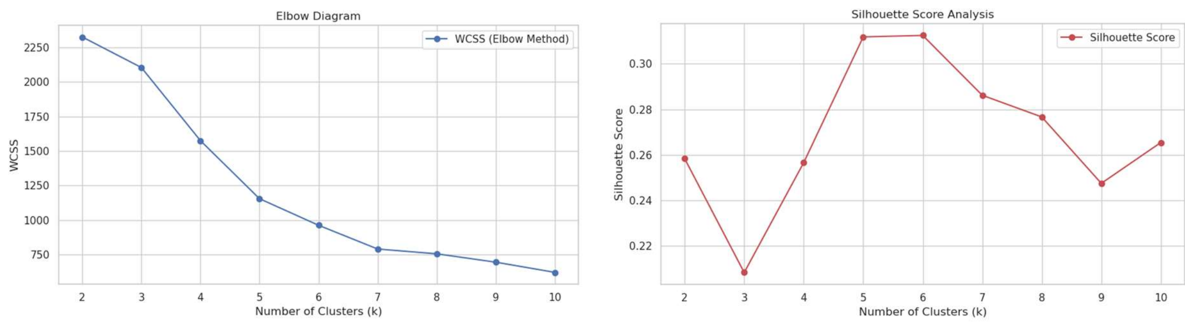

Figure 2 Elbow Diagram and Silhouette Scores of K means Clustering

The box plot in **Figure 3** further demonstrated Cluster 4 has the highest obesity rates, with a median significantly above other.

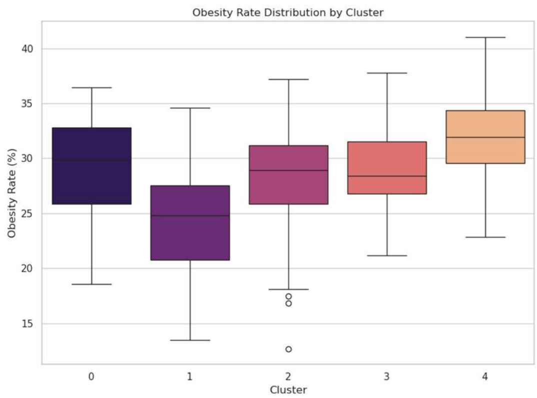

Figure 3 Obesity Rate Distribution by Cluster

In multivariate linear regression analysis, initially, all five variables were included (see **Table 2**, Model 1). However, cycling ratio and athletics ratio were statistically insignificant and excluded from the final model. The refined regression model (Model 2 in **Table 2**) retained three significant predictors: gym ratio (α=−0.3317), childhood poverty ratio (α=3.9198), and open space (α=−0.3371). These results highlight the strong positive relationship between child poverty and obesity rates, as well as the mitigating effects of gym and open space availability.

Table 2 Key variables

Variable	α (Model 1)	*p-Value* (Model 1)	**α (Model 2)**	***p-Value* (Model 2)**
Constant	28.5239	<0.001	28.5239	<0.001
Gym Ratio	-0.3390	0.013	-0.3317	0.014
Cycling Ratio	-0.1585	0.243	—	—
Athletics Ratio	0.0070	0.959	—	—
Poverty Ratio	3.9306	<0.001	3.9198	<0.001
Open Space	-0.3449	0.015	-0.3371	0.017
Adjusted R^2	0.591		0.591	

A LP model was then applied to optimize resource allocation based on these regression-derived coefficients. The unit costs were derived as follows: gym facility cost at £447,800 per 10,000 population (Sport England, 2020), poverty-targeting intervention cost at £8,000 per child based on Universal Credit eligibility criteria (Gov.uk, n.d.), and green space expansion cost at £25 per child (Ashford Borough Council, 2011). These costs were dynamically adjusted across wards based on population size and age distribution.

Sensitivity analysis, shown in **Figure 4**, reveal an inflection point at approximately £200 million, beyond which additional funding yields negligible improvements. This budget level was selected as the optimal amount for resource allocation.

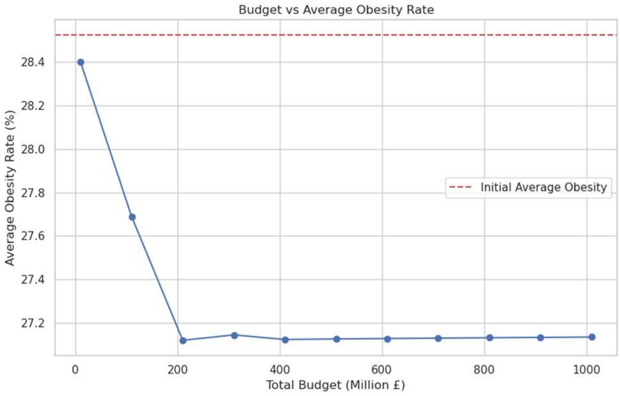

Figure 4 Budget vs Average Obesity Rate

Using the £200 million budget, the optimal resource allocation emphasizes prioritizing large-scale investment in health and fitness facilities (gyms), across London, as shown in **Figure 5 (middle)**. This approach addresses infrastructure gaps critical to reducing childhood obesity.

Expanding green space, particularly in high-risk areas like Cluster 4 (highlighted in red), is identified as the second priority (**Figure 5, left**).

Poverty-targeting interventions, despite their high impact as indicated by regression weights, were deprioritized due to their relatively higher costs (**Figure 5, right**).

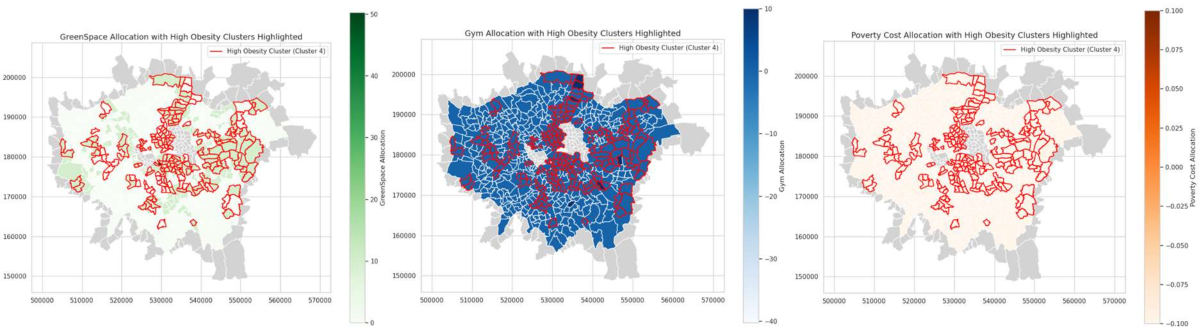

Figure 5 Allocation Strategies with High Obesity Cluster Highlighted

5. Discussion

The model prioritizes gyms and green spaces in high-obesity areas, particularly in East and South London (e.g., Southwark, Lambeth), where high poverty and limited fitness infrastructure make them critical targets for reducing health disparities.

Comparing the model's results with policy changes and resource distribution since 2013 highlights its reliability and policy relevance. For instance, while the *2018 London Environment Strategy* and *2016 Sport Unites Programme* aimed to expand green spaces and community sports facilities (Greater London Authority, 2023), their resources were distributed uniformly across London, lacking a systematic focus on high-risk areas. The *Health Profile for England 2018* (Public Health England, 2018) further indicates a uniform allocation trend across regions, which may have diluted the impact in high-risk wards. In contrast, the model's targeted approach provides a more effective strategy for addressing health inequities.

Additionally, poverty alleviation policies since 2013 have seen limited expansion, which aligns with the model's deprioritization of poverty-targeting measures.

However, the study also has limitations. It relies on static historical data, ignoring dynamic changes in demographics and infrastructure. Additionally, discrepancies between the model and real-world policies may arise from operational or political constraints, as resource distribution often prioritizes equity over targeting high-risk areas.

Future research should incorporate longitudinal data and additional factors, such as transportation access and community engagement, to develop dynamic models that better align policy decisions with practical challenges and maximize impact.

6. Conclusion

This study demonstrates the effectiveness of combining clustering and linear programming to optimize resource allocation for childhood obesity in London. By identifying high-risk areas, such as parts of East and South London, and prioritizing interventions like gym facilities and green space expansion, the model offers a targeted, cost-effective alternative to uniform resource distribution and actionable insights for policymakers.

However, the reliance on static historical data and the exclusion of dynamic factors, such as demographic shifts and infrastructure changes, limit the model's adaptability to evolving conditions. Future research should incorporate longitudinal data and additional variables, such as transportation access and community engagement, to enhance the model's accuracy and practical relevance.

Overall, this study highlights the value of data-driven, targeted strategies to combat childhood obesity and reduce health inequalities, offering a valuable framework for policymakers to design more effective interventions.

Additional Resources

Li, V. (2025) *007QM coursework*, GitHub. Available at: https://github.com/VeraLi0710/007QM-coursework (Accessed: 20 January 2025).

References

NCMP (2006–2024) *National Child Measurement Programme: Trends in child obesity, 2006 to 2024*. NHS Digital. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/national-child-measurement-programme (Accessed: 19 January 2025).

Singh, G.K., Siahpush, M. & Kogan, M.D. (2010). Neighborhood socioeconomic conditions, built environments, and childhood obesity. *Health Affairs,* 29(3), pp.503-512.

Greater London Authority (2023). *London Environment Strategy*. Available at: https://www.london.gov.uk/programmes-and-strategies/environment-and-climate-change/london-environment-strategy

Chuvieco, E. (1993) 'Integration of linear programming and GIS for land-use modelling', International Journal of Geographical Information Science, 7(1), pp. 71–83.

Greater London Authority (2014) *Prevalence of Childhood Obesity by Borough, Ward and MSOA: 2013–14 data*. Data derived from the National Child Measurement Programme (NCMP). Available at: https://data.london.gov.uk/dataset/prevalence-childhood-obesity-borough (Accessed: 19 January 2025).

Greater London Authority (2014). *Ward Profiles and Atlas*. Available at: https://data.london.gov.uk/dataset/ward-profiles-and-atlas (Accessed: 19 January 2025).

Greater London Authority (2014). *Access to Public Open Space and Nature*. Available at: https://data.london.gov.uk/dataset/access-public-open-space-and-nature-ward (Accessed: 19 January 2025).

Sport England (2013). *Active Places Power*. Available at: https://www.activeplacespower.com/pages/downloads#download (Accessed: 19 January 2025).

Jain, A.K. (2010). *Data clustering: 50 years beyond K-means.* Pattern Recognition Letters, 31(8), pp. 651–666.

Gov.uk (n.d.) Universal Credit eligibility. Available at: https://www.gov.uk/universal-credit/eligibility (Accessed: 18 January 2025).

Ashford Borough Council (2011) Green space cost calculations: Evidence base to Green Spaces and Water Environment SPD. Available at: https://www.ashford.gov.uk/media/lsbi1qfc/green-space-cost-calculations-2011-evidence-base-to-spd.pdf (Accessed: 18 January 2025).

Sport England. (2020) *Affordable Sports Centre: 25m pool summary.* Sport England. Available at: https://www.sportengland.org (Accessed: 18 January 2025).

Public Health England (2018) *Health Profile for England 2018*. Available at: https://www.gov.uk/government/publications/health-profile-for-england-2018 (Accessed: 18 January 2025).