Statistics with Data Science

Consultancy Project

Spatially Aware Modelling for Data-Driven Gym Membership Demand Estimation

---

The School of Mathematics

THE UNIVERSITY of EDINBURGH

Gleb Novikov

S2693219 – MSc Statistic with Data Science
University of Edinburgh

Contents

Executive summary

Building a gym represents a substantial investment, and as a highly localized service, location plays a critical role in determining its success. Thus, trying to understand in advance the factors driving future demand, including spatial patterns, is crucial. This project applies modern statistical and spatiotemporal modeling techniques to estimate gym demand in an attempted formal scientific way. Using Bayesian modeling via INLA, we analyze panel data provided by Simon-Kucher to compare a variety of demand estimation models. Our final model, based on the SPDE (Stochastic Partial Differential Equation) approach, outperforms simpler alternatives on key Bayesian model selection criteria such as WAIC and LCPO. However, it underperforms in 10-fold cross-validation, potentially signalizing overfitting.

Key factors associated with increased gym demand include population density and parking availability, while demand is negatively impacted by local competition and the proportion of repeat users, which aligns well with economic and operational intuition. The spatial component of the SPDE model revealed significant within-region heterogeneity: for example, demand was systematically lower than expected in Greater London and higher in areas such as the Southeast England.

Interestingly, once these key variables were controlled for, many others lost statistical and economic significance. This was especially true for pricing variables, whose effects proved unstable or counterintuitive and thus unsuitable for correct price optimization. These findings underscore the limitations of using historical observational data to estimate price elasticity. Instead, we recommend the use of randomized experiments, such as A/B testing, for more credible causal inference in pricing decisions.

Own Work Declaration

I assure the single-handed composition of this MSc dissertation project only supported by declared resources.
Edinburgh, 24th July 2025

Word count: 4986

1. Introduction

1.1 Background and Motivation

For many businesses, particularly fitness centers, the nature of their operations requires a physical presence to deliver their services or goods. Gyms must invest heavily in establishing a physical site, which typically goes through a “maturation” process over time (often a year or more), gradually building demand from nearby residents or commuters. Moreover, gyms are highly localized services: customers place significant value on proximity and often choose between options primarily based on geographic location (Wicker et al., 2012).

Therefore, management of fitness centers aims to choose the location of new gyms as wisely as possible. Traditionally, this involved physically scouting potential sites. However, for large gym chains that have accumulated substantial data, modern statistical tools combined with publicly available datasets now make it possible to take a more scientific approach. Specifically, we can investigate how demand for gyms depends on factors such as location, surrounding competition, and local demographics. If such a model proves to be predictive, it could help companies avoid poor investment decisions and improve access to fitness facilities within communities.

1.2 Data Source

The dataset was provided by Simon-Kucher & Partners and consists of anonymized internal data from a major UK gym chain with 236 operational locations. It includes each gym’s geographic location, square footage, membership numbers, and monthly pricing over a 12-month period. In addition, Simon-Kucher analysts supplemented the dataset with external sources, incorporating demographic data for areas surrounding each gym, the number of nearby competitors, local market characteristics such as transaction volumes, as well as the presence of nearby parking facilities or universities.

1.3 Aims

This project aims to develop a robust Bayesian regression model to predict total gym membership at maturity and design an algorithm that recommends the optimal monthly price to maximize expected revenue for each gym based on its unique context.

The project comprises four key components:

1. **Exploratory Data Analysis**
   We begin by examining the data through targeted visualizations and summary statistics to understand the business problem, guide feature selection and preprocessing, and identify spatial or temporal patterns.
2. **Model Development**
   A range of Bayesian models from simpler random-effects to complex spatial and temporal specifications, such as SPDE, are built and compared. Evaluation is based on multiple metrics, including LCPO, WAIC, and 10-fold cross-validation.
3. **Feature Interpretation**
   We analyse which features drive gym membership demand in our final SPDE model.
4. **Price Optimization**
   We attempt to create a separate model estimating the relationship between price and demand.

2. Exploratory Data Analysis

2.1 Illustration of the problem: Central London case study

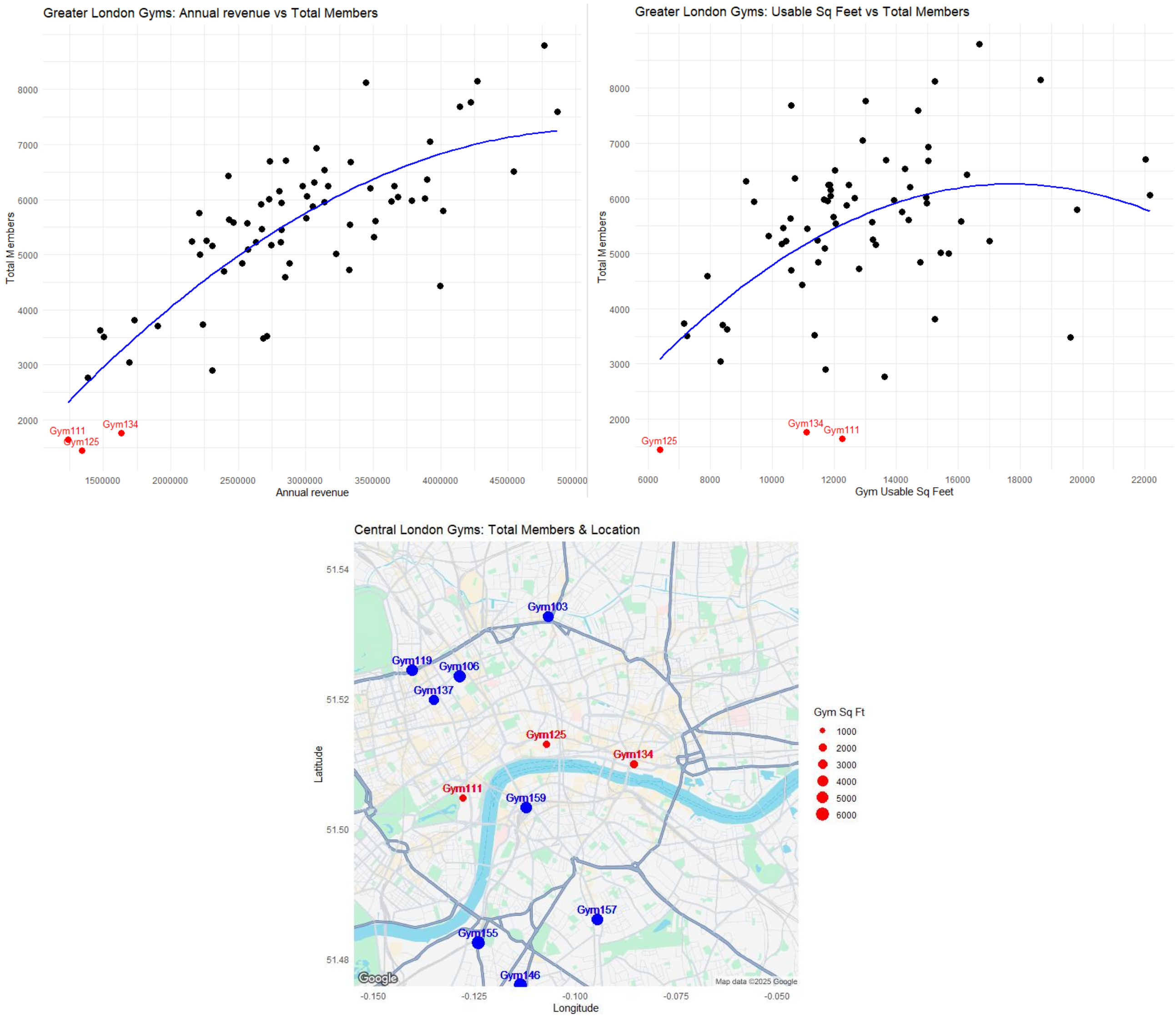


Figure 1. Identifying underperforming gyms in the dataset with the example of London gyms.

We begin our project by illustrating the core problem through a closer look at gyms in the Greater London area, with a particular focus on central London. Among these, we identify three gyms (125, 111, and 134) as clear underperformers in terms of both total members and revenue.

One possible explanation is that these gyms are simply smaller in size, leading to lower membership. To explore this, we plot members vs. size for all Greater London gyms (see, Figure 1), revealing a positive, likely quadratic, relationship. This explains Gym125's low count—it's the smallest in London. However, Gyms 134 and 111 are average in size but still underperform, so size alone doesn't explain their low membership.

Another speculative explanation is that these gyms were low-cost investments, yielding revenue typical of less competitive regions (e.g., Northeast England or Scotland; see Table 1). However, as shown in Figure 1, all three gyms are in highly central London, likely among the most expensive UK areas. If we speculatively assume that square footage costs are relatively uniform across central London, then Gyms 111 and 134 stand out as particularly poor investments.

This example highlights the importance of carefully selecting gym locations and underscores the need for a data-driven statistical approach to support investment decisions and avoid costly misallocations.

2.2 Distributions of numeric variables: non-normality of independent variables & overdispersion of count response.

At the outset, it is important to highlight two key characteristics of our data. First, the response variable (total members) is a count variable, consisting of non-negative integers. Moreover, it exhibits strong overdispersion: even when considering only the yearly averages across 236 gyms (thus omitting within-gym temporal variation), the variance is 1,844,169, which is substantially greater than the sample mean of 5,049.5.

Second, when examining a non-exhaustive selection of key variables from our dataset (see Figure 2), we observe clear signs of non-normality in their distributions characterized by heavy tails, skewness, and sometimes lack of unimodality.

Together, these two observations lead us to conclude that we should first scale and normalize the explanatory variables. For modelling the response, a negative binomial distribution is more appropriate, as standard Poisson or Gaussian assumptions are unlikely to perform well.

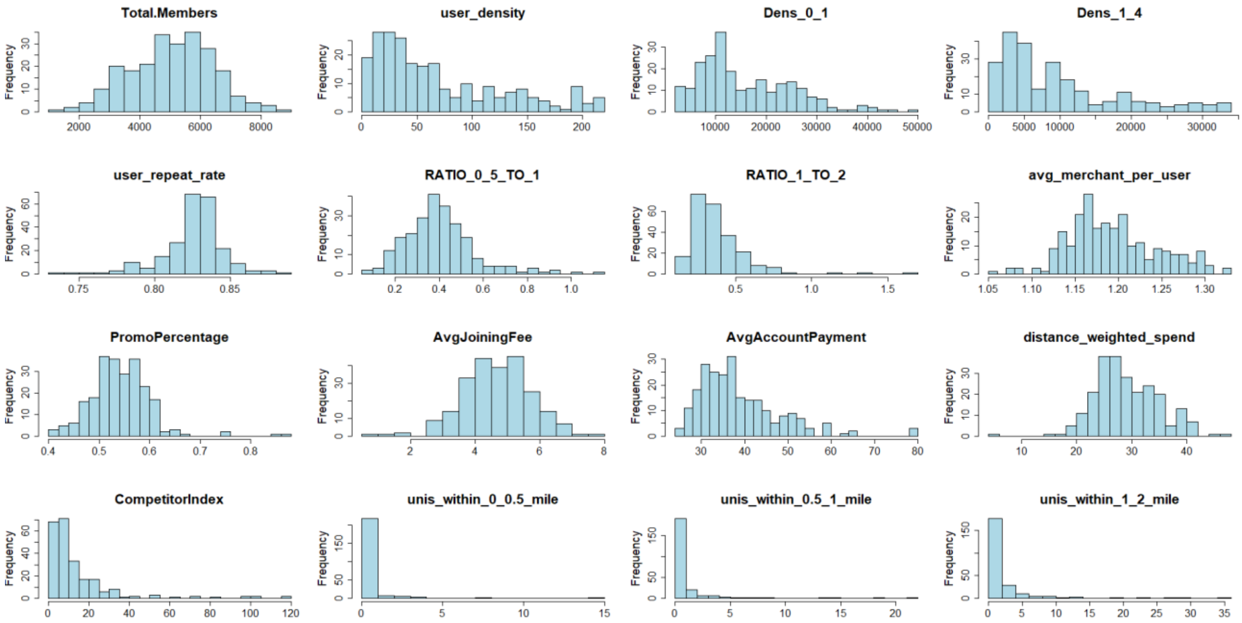


Figure 2. Histogram of several explanatory variables and response from the dataset.

2.3 Variables' interrelationships: curse of dimensionality, correlation matrix and initial features selection

Another key challenge in our dataset is the imbalance between the number of features (51) and unique gym observations (236), which risks overfitting and illustrates the curse of dimensionality. Many features are also highly correlated, both conceptually and numerically. In Bayesian models, such multicollinearity increases posterior uncertainty, making individual effect estimates less interpretable and potentially destabilizing inference (Wicker et al., 2019).

To address this, we applied the following simplifications and feature reductions in a consecutive and logical manner, with each step building on the previous:

1. Redundant Variable Pairs
    - Population vs. Transaction Counts/Densities: These pairs represent the same concepts and are perfectly correlated (100%). We arbitrarily choose to retain only density-based variables.
2. Reducing the number of population-density related variables
    - Population density variables at nearby radii (e.g., 0_0.5 and 0.5_1 miles; 1_2, 2_3, 3_4 miles) are highly correlated. To reduce redundancy, we aggregate them into two features: Population Density 0_1 mile and Population Density 1–4 miles.
    - User density, measured over a 4-mile radius, is strongly correlated with population density at higher radii. However, it reflects actualized fitness demand, as it depends on existing gym users (see Section 2.7). To avoid collinearity and maintain interpretability, we choose to represent actualized demand via transaction density instead. As a result, we include only the two aggregated population density variables and omit user density.
    - We drop the Density_Drop variables, as their information is, to some extent, already captured by the population ratio variables comparing different radii.
3. Highly Correlated Transaction Features
    - Transaction densities across different radii are strongly correlated and we conclude that user average spending capture most of this information.
    - Of the two spending features (average vs. distance-weighted average), we keep the weighted one as we believe it might be more informative.
4. Other Reductions
    - Between the two heavily correlated population ratio variables (0.5-mile vs. 4+ mile and 1-mile vs. 4+ mile), we drop the 0.5-4-mile ratio arbitrarily.
    - We exclude the PCA-based variables (umap2d-1, umap2d-2) to preserve model interpretability.
    - We omit the individual competitor count variables and retain only the Competitor Index, as it captures this information in a more consolidated form.
5. Notes for future model feature selection
    - Competitor Index is almost the only remaining feature with very high correlations to other variables. However, we consider it very meaningful for interpretation and will retain it despite this.
    - The current 13 remaining features set will be further reduced if some variables prove to be statistically insignificant.

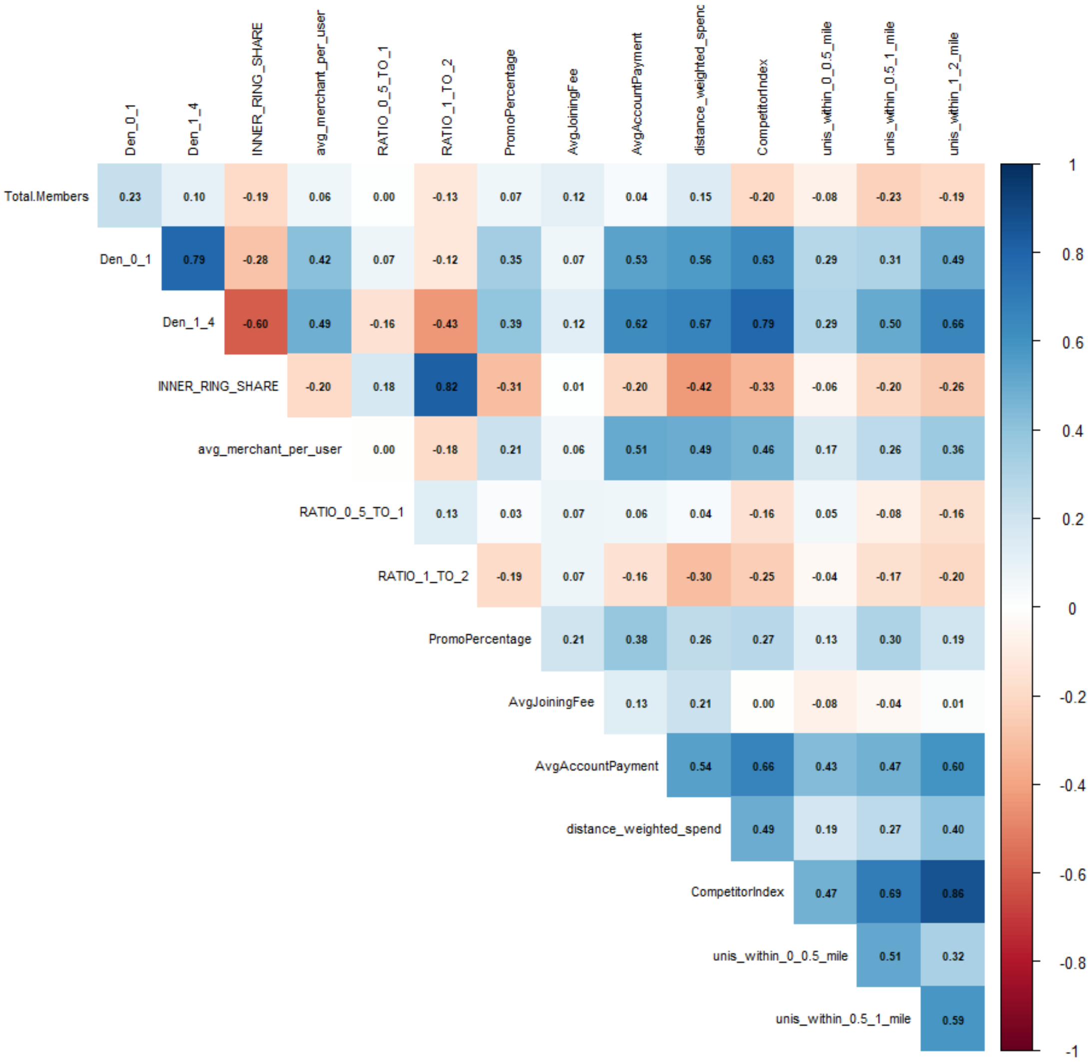


Figure 3. Correlation matrix of the remaining features after initial feature selection.

2.4 Categorical Variables

We examine two categorical variables: (1) the presence of a nearby parking lot (yes/no), and (2) the gym's area type (residential, workforce, or hybrid) defined by the average distance between users' homes and the gym. Figure 4 shows that gyms in workforce areas have lower average membership levels, and that there is no clear difference between gyms with and without parking; however, both patterns are unadjusted and may be influenced by other correlated factors. The actual effect of area type will be assessed more reliably during the modelling stage.

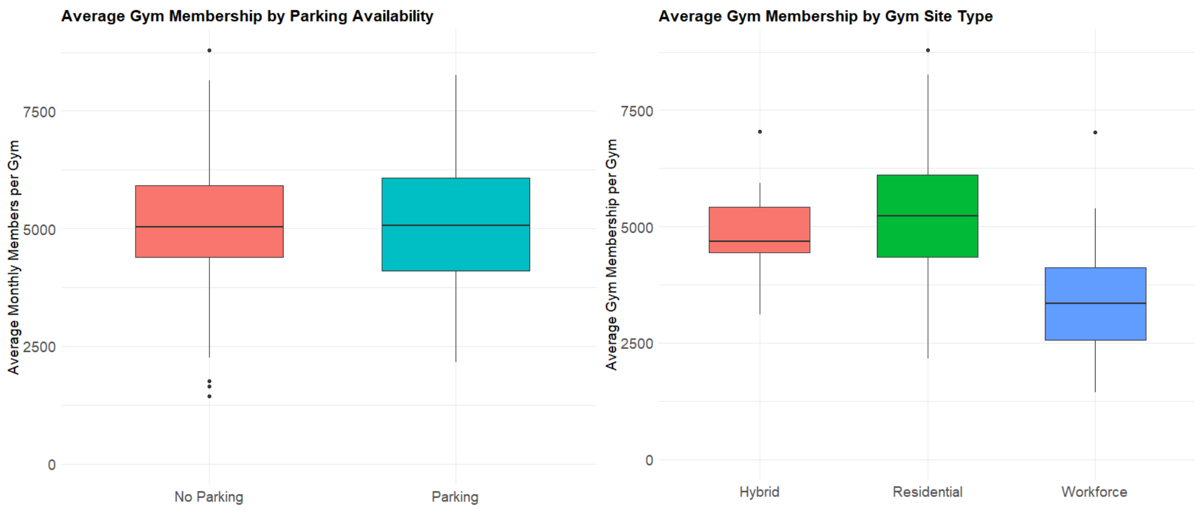


Figure 4. Distributions of average monthly gym membership for key categorical features.

2.5 Spatial distribution

One way to explore the spatial distribution of the data is by identifying the administrative region each gym belongs to (see, figure 6).

A key observation is that London is heavily overrepresented, with 70 out of 236 gyms located there. To account for this imbalance, separating London into five distinct districts is a reasonable approach (see, Table 1). Additionally, we observe substantial variation in average revenue and price across regions. London gyms tend to attract the highest number of members and charge higher prices, resulting in top-performing revenue figures. In fact, all 5 London districts are in top 5 regions in terms of revenue. This likely explains why the company is focusing so heavily on the London market.

FinalRegion	UniqueGyms	AvgMembers	AvgAccountPayment	AvgRevenue
East London	12	6033.1	41.7	251283.3
West London	14	5771.3	42.8	247176.0
South London	18	5677.1	43.4	246142.0
Central London	15	4101.6	58.9	241623.1
North London	11	5740.3	41.6	238625.7
West Wales and the Valleys	2	5683.5	41.0	232868.6
South East England	36	5107.5	38.4	196367.0
North West England	21	5192.1	34.7	179958.7
West Midlands	19	5402.7	33.2	179273.3
Yorkshire and the Humber	14	5001.0	35.6	178129.5
East Wales	1	4820.1	36.2	174445.9
East Midlands	13	4978.2	33.9	168830.5
Eastern Scotland	8	4598.5	36.2	166388.6
South Western Scotland	11	4286.8	36.3	155590.5
East England	18	4578.5	33.8	154787.6
South West England	15	4192.4	36.5	153173.7
North East England	8	4490.7	33.3	149663.1

Table 1. Summary of key gym metrics by UK regions.

Figure 5 below shows the distribution of average monthly members across regions in greater detail. While the mean membership levels differ significantly between regions, there is also substantial within-region variation. This suggests that, although regional location carries meaningful information, it is not sufficient on its own to explain membership differences.

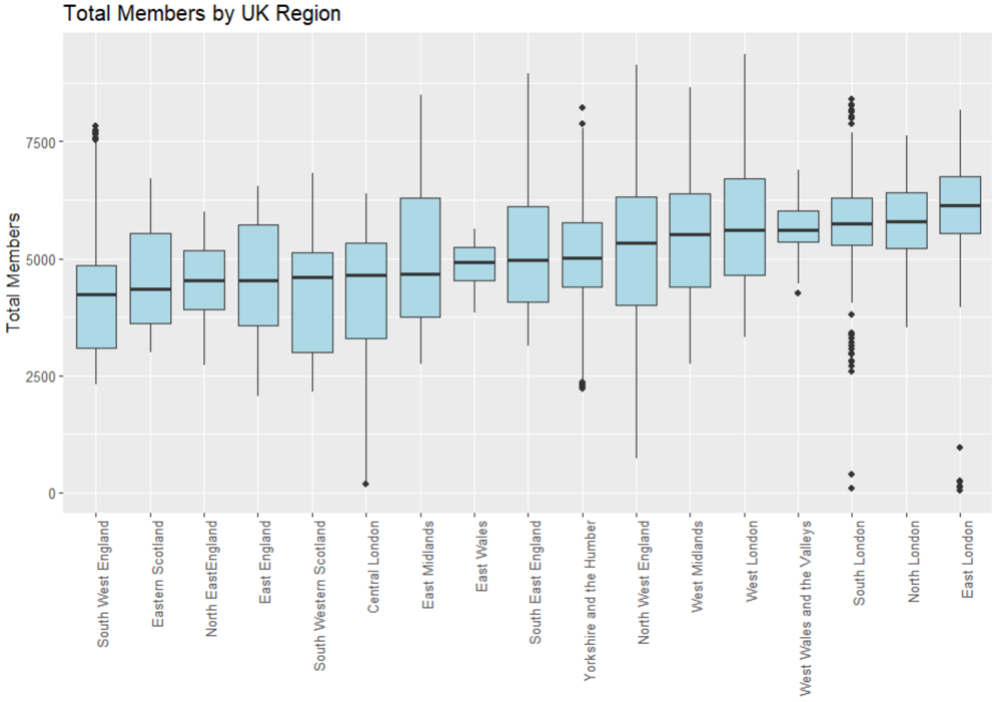


Figure 5. Distributions of average monthly members between UK regions.

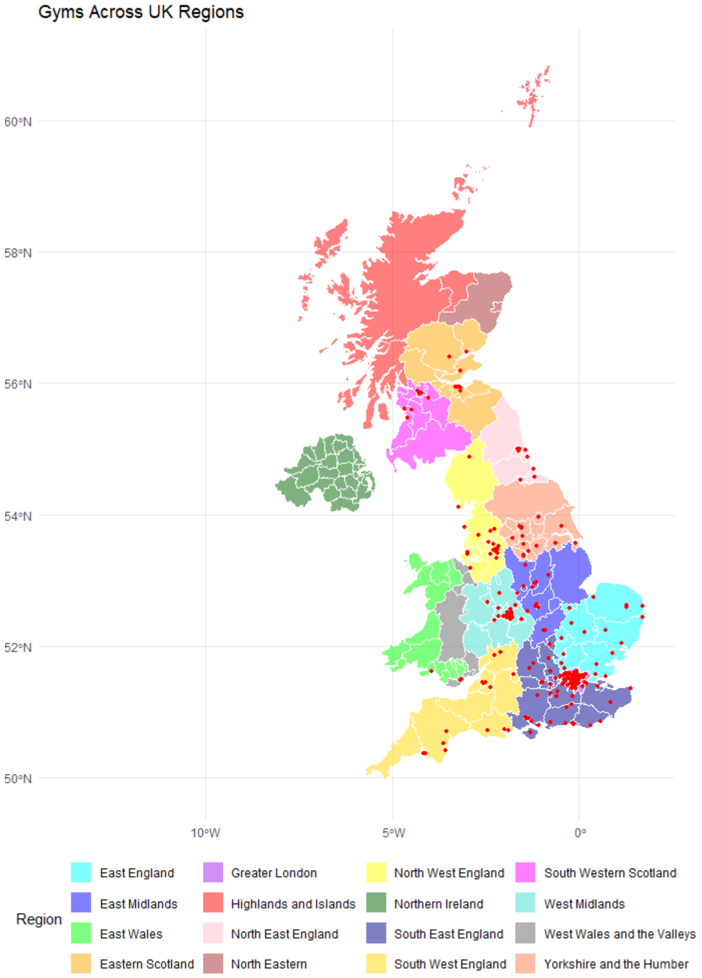


Figure 6. Illustration of selected UK regions and locations of gyms (red dots).

2.6 Temporal distribution

Another key feature of our dataset is its panel structure: we observe each gym monthly, resulting in 12 observations per gym. However, only four variables vary over time: Total Members, Average Account Payment, Promo Percentage, and Average Joining Fee.

Figure 7 shows relatively flat membership trends, though variation across gyms is high. Promotions spike in January and September, with joining fees dropping in those months. Account payments peak in September. All these three variables show similar seasonal patterns and low between-gym variability.

Two main conclusions follow:

1. We cannot conclude that membership lacks a temporal pattern without first controlling for other time-varying factors.
2. Price, promotions, and fees likely need monthly adjustments to maintain stable membership.

Figure 7. Within-year trends of time-varying variables.

2.7 Discussion of the effect of Gym size on our model and some other variables

We need to critically consider what our future model can and cannot capture, based on our understanding of the data.

As highlighted in Section 2.1, there is a clear relationship between gym size and total membership, which is conceptually intuitive: smaller gyms, even at full capacity, naturally accommodate fewer members than larger ones. Overcrowding can deter both potential and existing members, effectively capping membership levels.

However, according to the company's representatives, gym size is not a predetermined, exogenous variable. Instead, it is chosen based on an internal optimization algorithm, informed by prior demand assessments. This means gym size is not known in advance of opening a gym and cannot be treated like a standard input feature in our predictive model.

Still, we recognize that gym size influences membership outcomes, and therefore affects the quality of demand estimation. Ideally, size should be optimized jointly with other factors like price—either within the main model or in a separate decision layer. However, in line with the company's guidance, we will exclude gym size as an input in our primary model.

Other variables also require careful interpretation. For example, many of the features related to market spending are calculated based on areas where there is already some actualized fitness demand—i.e., where gyms currently operate. As a result, postcodes without existing gyms—such as remote areas or underserved urban zones—may be misrepresented in the model. In these cases, the underlying demand may not yet be realized, and our model, relying on observed market activity, could underestimate the true potential of those locations.

3. Demand Estimation

3.1 Metrics Used

In Bayesian inference, we estimate the probability distribution of unknown parameters given the observed data, rather than just single-point estimates. This results in posterior distributions that reflect both the data and prior beliefs. The Integrated Nested Laplace Approximation (INLA) is a computational method specifically designed to efficiently perform Bayesian inference for latent Gaussian models — a broad class that includes most generalized linear (and additive) models.

Our approach is to fit multiple Bayesian models with varying design and complexity and compare their performance using three key metrics: WAIC, LCPO, and 5-fold Cross-Validation.

- WAIC (Watanabe–Akaike Information Criterion)

WAIC estimates out-of-sample predictive accuracy while accounting for the entire posterior distribution, making it suitable for Bayesian inference. It balances model fit and complexity by penalizing overfitting through the effective number of parameters. The WAIC is calculated as:

$$WAIC = -2\sum_{i=1}^{n} \log \left(\frac{1}{S}\sum_{i=1}^{n} \mathrm{p}\left(y_i \middle| \theta^{(s)}\right)\right) + p_{waic}$$

Where: y_i is the observed response, θ(s) is the s-th sample from the posterior and p_{WAIC} is the effective number of parameters

Lower WAIC values indicate better predictive performance.

However, it's important to note that WAIC leaves out one observation at a time, meaning one month-level observation per gym, not entire gyms.

- LCPO (Log Conditional Predictive Ordinate)

LCPO is another Bayesian model evaluation metric based on leave-one-out predictive performance. It assesses how well the model predicts each observation when that observation is excluded from the likelihood. It is calculated as:

$$LCPO = \sum_{i=1}^{n} \log\ \mathrm{p}(\mathrm{p}(y_i|y_{-i})$$

Where: y_i is the observation being left out and y−i are all the other observations

Like WAIC, LCPO in our context leaves out a single time point per gym, not the entire gym. Thus, the model still "knows" the gym it is predicting. This makes LCPO a conditional measure and not fully appropriate when the goal is to generalize to completely new gyms.

Higher (less negative) LCPO values are better.

- 10-fold Cross Validation

To directly assess how the model performs on unseen gyms, we also conduct 10-fold cross-validation at the gym level. The gyms are split into five groups; in each fold, we train the model on four groups and validate it on the fifth, rotating the held-out group across the folds.

This method allows us to evaluate true generalization performance to new locations — which is our core modeling goal. However, this cross-validation does not use the full posterior distribution and simply evaluates predictions using point estimates.

RMSE from 10-fold CV is informative but should be interpreted as complementary to WAIC and LCPO, which do account for Bayesian uncertainty.

3.2 Priors

Throughout the modelling stage, we adopt weakly informative priors to stabilize inference while allowing the data to drive the results. For the fixed effects, we rely on default normal priors with large variance (e.g., N(0,1000)) which center the estimates around zero without overly constraining them. For random effects, such as regional or gym-level intercepts, we use default log-gamma priors on precision parameters, which translate to diffuse priors over variances, ensuring flexibility while avoiding overfitting. For the spatial SPDE model, we use penalized complexity (PC) priors for both the spatial range and marginal standard deviation. These priors regularize the spatial field by favouring smoother surfaces unless the data provides strong evidence for local variation, thus helping to control overfitting in spatial estimation. Additionally, we use a weakly informative log-gamma prior for the negative binomial dispersion parameter to allow flexible modelling of overdispersion in our count data.

3.3 Baseline vs Hierarchical Models: Gym ID and Regional Random Effects

We began our modeling with a fixed-effects-only specification as a simple baseline, fully aware that it is not well suited to our panel structure. In particular, many predictors are time-invariant across months, and the model lacks the flexibility to capture gym-specific or spatial variation — limiting both its explanatory and predictive power. Nonetheless, it served as a useful reference point for more sophisticated structures. From now on the comparisons of models are in Table 2.

To better capture unobserved heterogeneity across locations, we introduced hierarchical models starting with Gym ID as a random effect. This substantially improved model fit according to LCPO and WAIC. The estimated gym-level effects revealed consistent over- and underperformance among specific gyms, indicating meaningful variation unexplained by fixed covariates (see, figure 9). This structure is particularly useful for forecasting future performance at existing gyms, where past observations inform gym-specific adjustments.

However, this approach does not aid in predicting demand for new locations. When performing 10-fold cross-validation by holding out entire gyms, the model's performance reverted to that of the fixed-effects baseline. This is expected: for unseen gyms, the random effect is set to zero, and predictions rely entirely on fixed effects. Notably, this limitation is not captured by LCPO or WAIC, as both metrics leave out only a single monthly observation while retaining the gym's random effect, resulting in information leakage and overly optimistic evaluation for models intended to generalize to new gyms.

To introduce a spatial signal that generalizes across locations, we next fit a model with region as a random effect. This only marginally improved WAIC and LCPO relative to baseline and only roughly allowed us to detect broad geographic patterns — most regional effects were statistically insignificant.

A peculiar feature of the regional random effects model was that the credible intervals for many covariates remained notably narrower than those observed in the GymID random effects and SPDE models (see, Figure 8). This pattern was also present in the fixed-effects baseline. We may conclude that, for some reason, models that do not properly account for the panel structure of the data tend to deliver overconfident parameter estimates, likely underestimating true uncertainty. This observation further motivated the move to more flexible specifications like the SPDE model, which better accommodate the data's hierarchical and spatial complexity.

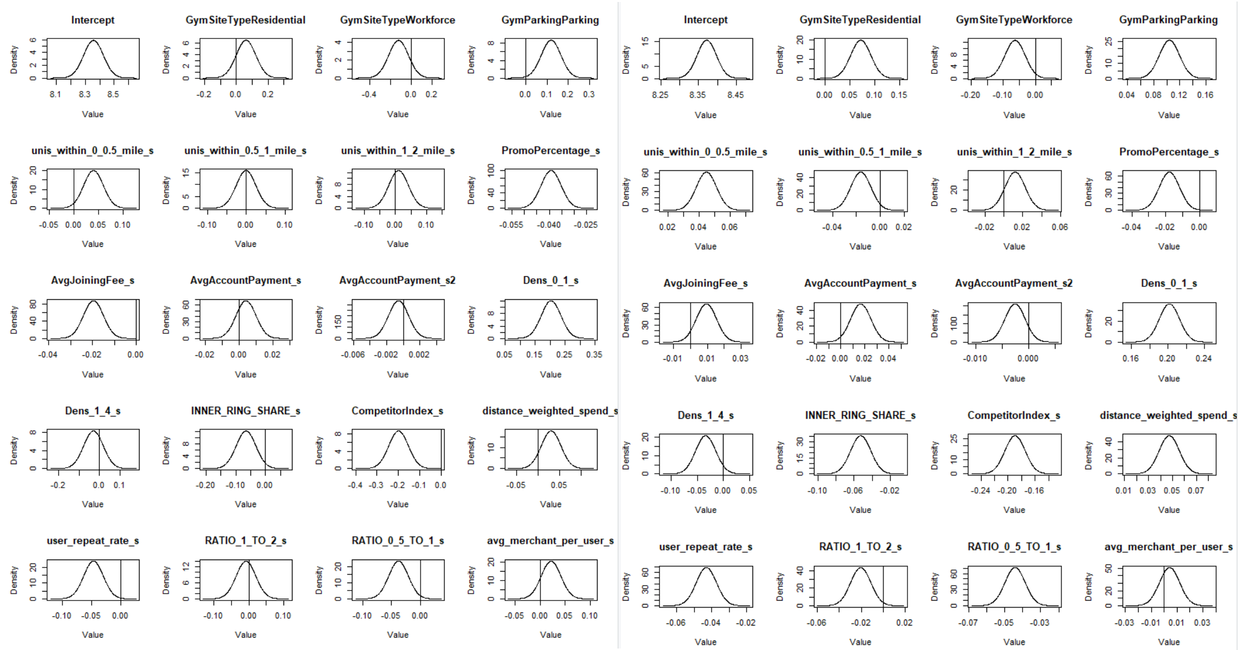


Figure 8. Comparison of estimated coefficients for 2 hierarchical models: gym ID random effects (on the left) and regional random effects (on the right). Credible intervals for the later are notably narrower (see, Appendix 7.2 &7.3).

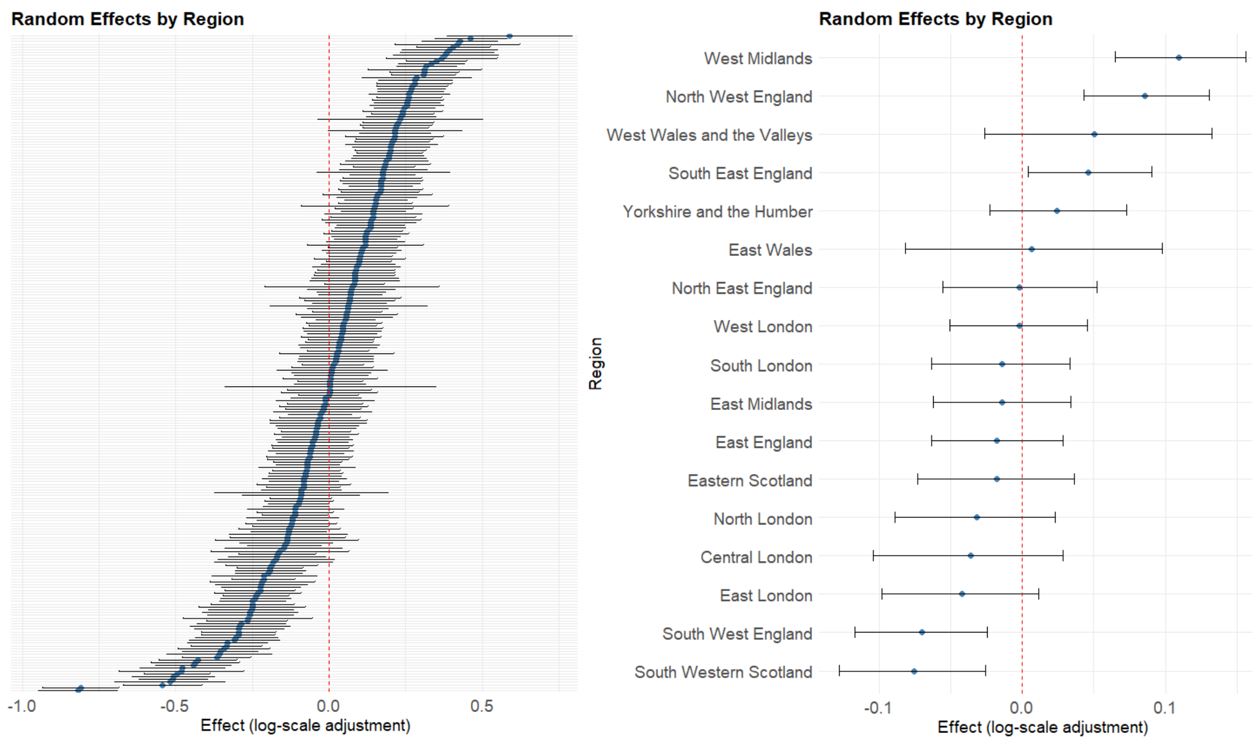


Figure 9. Illustrating direction and significance of random effects in 2 hierarchical models.

Model	DIC	WAIC	LCPO	RMSE 10-fold CV	M. Log-Likelihood
Fixed effects	48566.97	48466.50	-24287.82	1191.138	-24434.84
Fixed effects + Region random effects	48478.42	48359.83	-24241.89	1223.413	-24403.15
Fixed effects + Gym ID random effects	45376.66	44530.65	-22631.66	1203.679	-23139.25
Fixed effects + SPDE space	45257.65	45337.62	-22666.16	1423.354	-23191.55
Fixed effects (reduced) + SPDE space	45383.81	44556.16	-22631.18	1416.023	-23236.86
Fixed effects (reduced) + SPDE space + SPDE time	45385.31	44565.07	-22631.56	1421.611	-23225.42

Table 2. Comparison of attempted models.

3.4 SPDE model

Finally, we move on to our most sophisticated model, which incorporates a spatial effect using the SPDE (Stochastic Partial Differential Equation) approach. SPDE allows us to model spatially structured random effects continuously over geographic space, capturing smooth spatial correlations between gym locations. It does this by projecting the spatial process onto a mesh, efficiently approximating a Gaussian field.

The model fits well and shows a clear improvement over the fixed-effects and regional random-effects models, as reflected by better LCPO and WAIC scores (see, Table 2 & Figure 11). While its predictive performance is like the Gym ID random-effects model, the SPDE model takes a conceptually more appropriate approach to generalization: it learns spatial correlations from the data and leverages them to make informed predictions for new, unseen locations. Additionally, the issue with appropriately addressing panel data is mitigated, and credible intervals are well-calibrated and comparable to those in the GymID RE model.

Credible intervals being narrower enables a more reliable second-stage feature evaluation. Variables such as gym classification (workforce, residential, or hybrid) show little effect, with wide intervals centered around zero. Likewise, the average number of gyms used per user and the count of nearby universities (1–2-mile range) appear insignificant. Population ratio variables also lack predictive power once aggregated population density variables are included. Thus, we drop these variables which marginally increases WAIC and LCPO scores.

We also add a temporal component, modelled as a one-dimensional SPDE over the 12-month index, to capture smooth seasonal trends in gym membership. This allows the model to borrow strength across months and identify temporal autocorrelation, improving the coherence of time-related effects. While it does not substantially improve WAIC, LCPO, or CV metrics, it conceptually completes the model by accounting for both spatial and temporal structure in a unified Bayesian framework.

However, the surprising disadvantage of the SPDE-based model is that it performs significantly worse in 10-fold cross-validation, with notably higher RMSE compared to alternatives, as confirmed by a t-test (see, figure 10). Note that for all models RMSE is roughly between 1100 and 1400 which is a very high error considering that mean membership is 5050, so all our models make quite poor predictions on unseen gyms. While we acknowledge that RMSE from classical CV does not account for uncertainty (and thus isn't a perfect metric for Bayesian models) this result is still concerning. Despite its conceptual appeal and ability to model spatial correlation, the SPDE model's poor out-of-sample performance seems to be a clear drawback in this setting and a potential sign of overfitting.

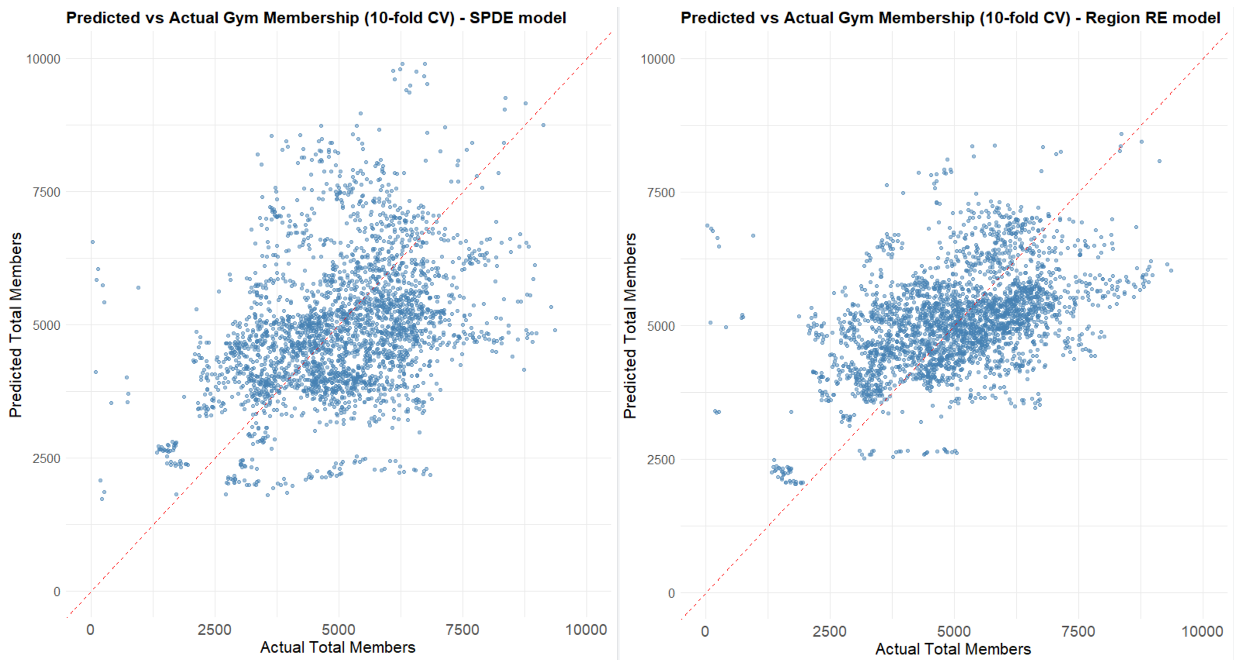


Figure 10. Comparing performance in CV of SPDE model (on the left) and regional hierarchical model (on the right) with the latter performing better (lower RMSE).

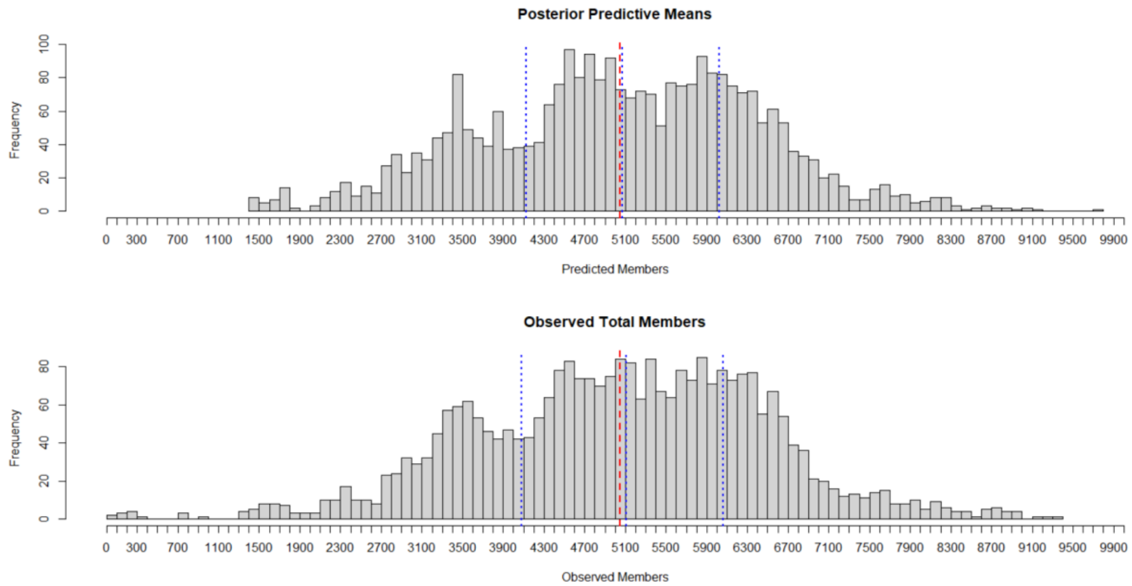


Figure 11. Model checks for the final SPDE model. Mean (red), median (blue) and quantiles (blue on the sides).

3.5 Discussion of Features Importance

Having selected the SPDE model as our final specification, we now turn to a critical evaluation of the remaining features.

One important concern is multicollinearity - particularly regarding the Competitor Index, which is strongly correlated with several key variables. For example, competitors tend to locate their gyms in areas with high population density and proximity to universities. This makes it difficult to isolate the true effect of some predictors, such as the number of nearby universities, which may appear insignificant simply due to their overlap with competitor presence.

Despite these challenges, several intuitive relationships remain robust and statistically significant in the final model:

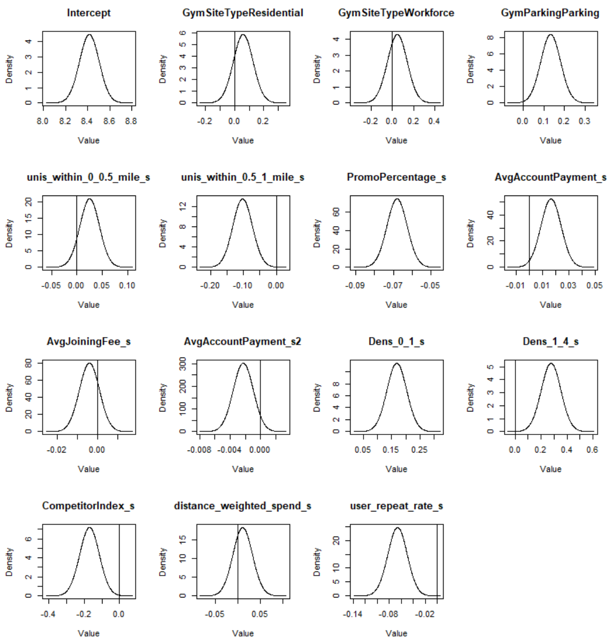
- Population density shows a strong and expected positive effect on gym membership.
- Parking availability is also positively associated with demand, which aligns with practical expectations.
- The number of competitors has a significant negative effect, consistent with basic competitive dynamics.
- User repeat rate is negatively associated with demand for new gyms, suggesting that areas where users are already loyal to existing gyms are less attractive for new openings.

Interestingly, the transaction-weighted metric, which we expected to be informative, turns out to be insignificant once other variables are controlled for. This may be due to overlapping information already captured by competition and population metrics.

In terms of pricing variables, promotion percentage is negatively associated with demand which likely reflects a reverse causal pattern where promotions are increased by management in response to falling demand. Notably, average account payment appears to have a nonlinear (quadratic) relationship with demand and when quadratic term is omitted the remaining linear relationship becomes positive. While this departs from standard economic expectations, it may simply provide a better statistical fit to the historical data rather than reflect an underlying causal mechanism.

The SPDE model also enables us to explore spatial variation within regions more deeply than the regional random effects model. SPDE model reveals substantial within-region heterogeneity. For example, in East England, we observe a distinct east-to-west gradient: gym membership is significantly lower than expected in the eastern parts, gradually increasing to above-expected levels toward the west. Additionally, Greater London area apart from its western part seems to deliver lower-than-expected membership outcomes. Similar patterns of non-uniform spatial effects are evident in other regions as well, highlighting the importance of modeling space continuously rather than relying on broad, aggregated regional categories.

On the other hand, temporal SPDE field having much shorter range and variance than spatial one suggests that after accounting for spatial and gym-level effects, little structured temporal variation remains (see, figure 12). This likely reflects the short time horizon (12 months) and relatively stable demand over that period.



Fixed effects:

	mean	sd
Intercept	8.415	0.090
GymSiteTypeResidential	0.058	0.068
GymSiteTypeWorkforce	0.048	0.093
GymParkingParking	0.133	0.048
unis_within_0_0.5_mile_s	0.025	0.019
unis_within_0.5_1_mile_s	-0.102	0.030
PromoPercentage_s	-0.068	0.005
AvgAccountPayment_s	0.017	0.008
AvgJoiningFee_s	-0.004	0.005
AvgAccountPayment_s2	-0.002	0.001
Dens_0_1_s	0.168	0.035
Dens_1_4_s	0.277	0.076
CompetitorIndex_s	-0.170	0.056
distance_weighted_spend_s	0.010	0.022
user_repeat_rate_s	-0.066	0.016

Model hyperparameters:

	mean	sd
size for the nbinomial observations (1/overdispersion)	51.847	1.521
Range for spde_space	0.108	0.007
Stdev for spde_space	0.530	0.026
Range for spde_time	0.379	0.515
Stdev for spde_time	0.044	0.011

Figure 12. Final SPDE model coefficients.

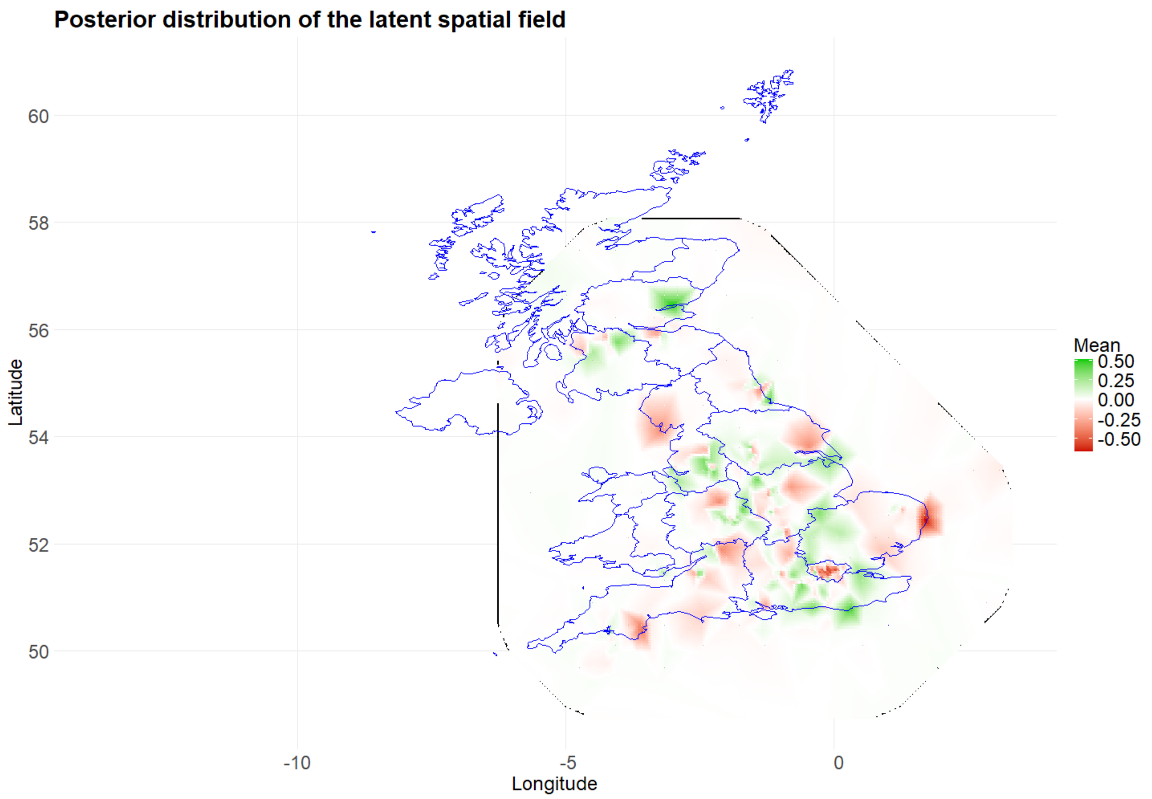


Figure 13. Map of posterior spatial effect. Note: The values are on the log-scale where 0.5 corresponds to membership being ~65% higher just due to space after controlling for all the covariates, and -0.5 corresponding to membership being ~39% lower.

4. Price Optimization

4.1 Attempt to find optimal price

Our final SPDE model appears to be optimal for explaining variations in demand; however, it performs poorly when used to identify an optimal price point. We quickly notice that the model's initial coefficients for the linear and quadratic price terms (0.017 and -0.002, respectively) suggest a highly inelastic demand, with the predicted revenue-maximizing price exceeding £500 which is highly unrealistic.

This initial model also fails to account for potential price interactions. We explored several alternative models using different combinations of variables, but most yielded similarly unrealistic results. Frequently, the coefficient for the squared price term approached zero or even became positive.

The model below was fitted with GLM with negative binomial family and was the one returning the most reasonable results, though even this model's optimal price prediction remained significantly higher than the observed average price of £38.6 (see, Figures 14 and 15).

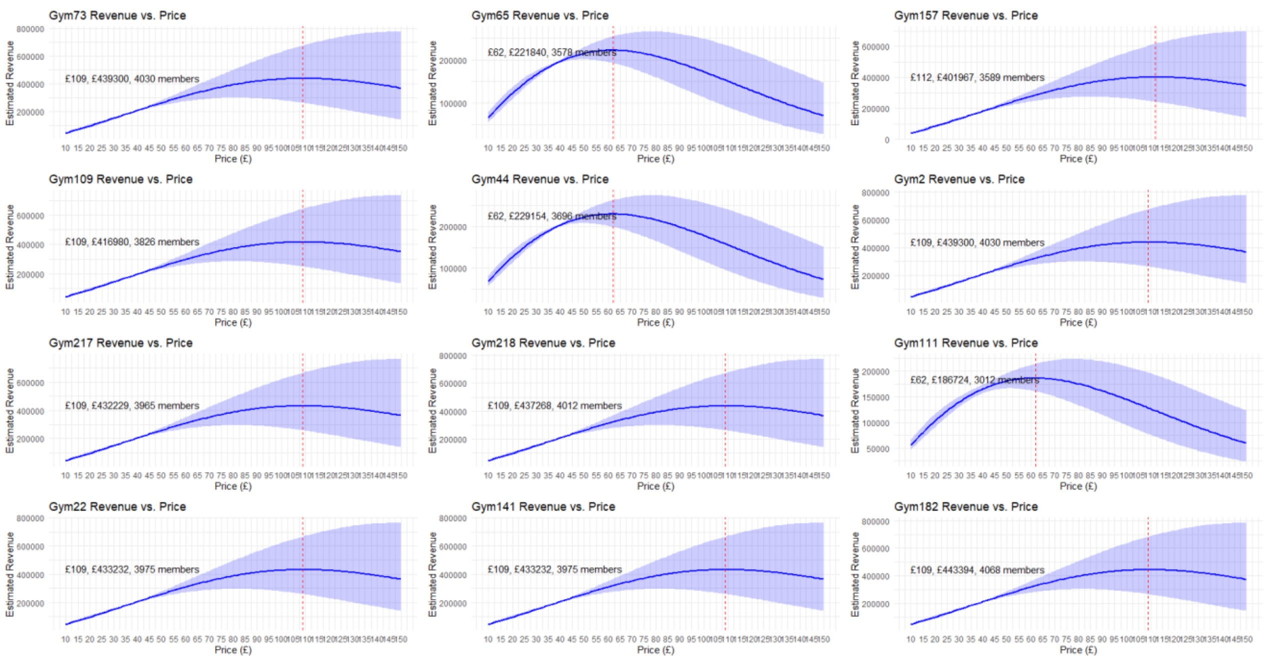


Figure 14. Simulations for 12 random gyms from the dataset.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	8.5359838	0.0065266	1307.869	< 0.0000000000000002	***
AvgAccountPayment_s	0.0205887	0.0315880	0.652	0.5145	
I(AvgAccountPayment_s^2)	-0.0136196	0.0033182	-4.104	0.00004051992302662	***
user_repeat_rate_s	-0.0452444	0.0057956	-7.807	0.00000000000000587	***
AvgAccountPayment_s:as.factor(month_index)2	0.0549533	0.0384092	1.431	0.1525	
AvgAccountPayment_s:as.factor(month_index)3	-0.0009317	0.0363482	-0.026	0.9796	
AvgAccountPayment_s:as.factor(month_index)4	-0.0048004	0.0375133	-0.128	0.8982	
AvgAccountPayment_s:as.factor(month_index)5	0.0430791	0.0386652	1.114	0.2652	
AvgAccountPayment_s:as.factor(month_index)6	0.0356896	0.0392771	0.909	0.3635	
AvgAccountPayment_s:as.factor(month_index)7	0.0158380	0.0375889	0.421	0.6735	
AvgAccountPayment_s:as.factor(month_index)8	-0.0144810	0.0364451	-0.397	0.6911	
AvgAccountPayment_s:as.factor(month_index)9	0.0701889	0.0324999	2.160	0.0308	*
AvgAccountPayment_s:as.factor(month_index)10	0.0791664	0.0324809	2.437	0.0148	*
AvgAccountPayment_s:as.factor(month_index)11	0.0336753	0.0341979	0.985	0.3248	
AvgAccountPayment_s:as.factor(month_index)12	0.0309421	0.0349352	0.886	0.3758	
AvgAccountPayment_s:GymSiteTypeWorkforce	-0.1798507	0.0196846	-9.137	< 0.0000000000000002	***
AvgAccountPayment_s:GymSiteTypeResidential	0.0125050	0.0179326	0.697	0.4856	
AvgAccountPayment_s:GymParkingParking	-0.0098791	0.0148849	-0.664	0.5069	

Figure 15. Coefficients of the GLM model for optimal price.

4.2 Why finding optimal price with historic data is problematic

Estimating optimal price from historical data is challenging due to several key limitations. First, prices in the dataset range narrowly between £20 and £60, with virtually no data above £90—where demand would likely drop drastically. Additionally, it is difficult to isolate the causal effect of price on demand, as pricing decisions in the data are likely influenced by external factors such as market conditions or managerial judgement. For example, management may lower prices in response to declining demand, introducing reverse causality into the analysis.

This endogeneity can lead to biased or unstable elasticity estimates. In particular, when excluding quadratic term, the model produces a positive association between price and demand, which contradicts standard economic theory such as the Law of Demand (Gale D., 1955).

Given these issues, we argue that estimating price elasticity from observational data is unreliable. A more credible alternative would be to run randomized controlled experiments, such as A/B tests (Dhaouadi et al., 2023), where prices are deliberately varied across groups or locations independent of demand. This allows for a cleaner, causal interpretation and provides stronger foundations for pricing decisions.

5. Limitations and Conclusion

In conclusion, we undertook a comprehensive analysis of Simon-Kucher's data, beginning with structured exploratory analysis, followed by careful variable selection focused on interpretability, multiple modeling approaches, and simulation-based price optimization.

A key finding was that there exists significant regional variation in gym membership and revenue, along with considerable within-region heterogeneity. This potentially suggests that there are indeed issues with site selection, as some gyms consistently underperform relative to their local peers. A more refined treatment of gym-level characteristics (e.g., gym size, which was excluded due to potential bias) could further clarify performance drivers. Future work could model this relationship more explicitly by incorporating financial efficiency metrics such as return on square footage or return on investment.

Among the models tested, the SPDE (Stochastic Partial Differential Equation) model provided the most nuanced explanation of spatial variation in gym membership, as reflected by strong performance in Bayesian model comparison metrics such as WAIC and LCPO. However, its performance dropped notably under 10-fold cross-validation, where full gym panels were held out. This suggests possible overfitting, raising concerns about why WAIC and LCPO failed to detect this issue. One possible explanation is that these metrics are not well-suited for panel data structures, as they leave out individual observations rather than entire grouped units (e.g., gyms), potentially masking model overconfidence. Still, we cannot fully rule out alternative explanations, and future work could explore more appropriate Bayesian validation techniques that leave out full gym panels simultaneously.

Our efforts at price optimization exposed fundamental challenges in estimating price sensitivity from historical data. The lack of variation in prices, combined with likely endogeneity, where prices may have been lowered in response to falling demand, resulted in models that inferred unrealistically high optimal prices well outside the observed range. This limitation highlights the need for experimental methods, such as A/B testing, to obtain more credible causal estimates of price elasticity (Dube, Hitsch, and Chintagunta, 2017).

Another limitation of our modelling was the presence of severe multicollinearity between several covariates. While we addressed this partly by excluding or combining some variables during feature selection, others (e.g., Competitor Index) remained highly correlated with related metrics. Future work could consider dimensionality reduction techniques like Principal Component Analysis or explore regularization via informative priors to mitigate instability and improve interpretability of parameter estimates (Jaya et al., 2019).

Overall, while our models capture several robust and interpretable relationships, these findings should be understood with an awareness of the limitations in model validation, data structure, and potential endogeneity. Nonetheless, the approach developed in this study lays important groundwork for more data-driven decision-making in gym performance analysis.

6. Bibliography

Depaoli, S., Winter, S. D., & Visser, M. (2020). The importance of prior sensitivity analysis in Bayesian statistics: demonstrations using an interactive Shiny App. *Frontiers in psychology*, *11*, 608045.

Dhaouadi, W., Johari, R., Page, O. B., & Weintraub, G. Y. (2023). Price experimentation and interference. *arXiv preprint arXiv:2310.17165*.

Gale, D. (1955). The law of supply and demand. *Mathematica scandinavica*, 155-169.

Jaya, I. G. N. M., Tantular, B., & Andriyana, Y. (2019, July). A Bayesian approach on multicollinearity problem with an Informative Prior. In *Journal of Physics: Conference Series* (Vol. 1265, No. 1, p. 012021). IoP Publishing.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., ... & Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC.

Pesaran, M. H., & Smith, R. P. (2019). A Bayesian analysis of linear regression models with highly collinear regressors. *Econometrics and Statistics*, *11*, 1-21.

Wicker, Pamela, et al. (2012) "The value of Olympic success and the intangible effects of sport events–a contingent valuation approach in Germany." *European sport management quarterly* 12.4: 337-355.

7. Appendix

7.1 Model output for fixed effects baseline model

```
Fixed effects:
                          mean    sd 0.025quant 0.5quant 0.975quant   mode kld
Intercept                8.379 0.022      8.337    8.379      8.422  8.379   0
GymSiteTypeResidential   0.059 0.020      0.019    0.059      0.098  0.059   0
GymSiteTypeWorkforce    -0.083 0.031     -0.144   -0.083     -0.023 -0.083   0
GymParkingParking        0.122 0.015      0.092    0.122      0.152  0.122   0
unis_within_0_0.5_mile_s  0.042 0.006     0.029    0.042      0.055  0.042   0
unis_within_0.5_1_mile_s -0.009 0.008    -0.025   -0.009      0.007 -0.009   0
unis_within_1_2_mile_s   0.015 0.010     -0.005    0.015      0.035  0.015   0
PromoPercentage_s       -0.017 0.006     -0.029   -0.017     -0.006 -0.017   0
AvgJoiningFee_s          0.009 0.006     -0.003    0.009      0.021  0.009   0
AvgAccountPayment_s      0.013 0.008     -0.003    0.013      0.030  0.013   0
AvgAccountPayment_s2    -0.002 0.002     -0.006   -0.002      0.002 -0.002   0
Dens_0_1_s               0.204 0.011      0.183    0.204      0.226  0.204   0
Dens_1_4_s              -0.046 0.016     -0.077   -0.046     -0.014 -0.046   0
INNER_RING_SHARE_s      -0.070 0.011     -0.091   -0.070     -0.049 -0.070   0
CompetitorIndex_s       -0.195 0.015     -0.224   -0.195     -0.166 -0.195   0
distance_weighted_spend_s  0.025 0.008    0.010    0.025      0.040  0.025   0
user_repeat_rate_s      -0.048 0.005     -0.058   -0.048     -0.037 -0.048   0
RATIO_1_TO_2_s          -0.013 0.009     -0.031   -0.013      0.005 -0.013   0
RATIO_0_5_TO_1_s        -0.037 0.006     -0.048   -0.037     -0.026 -0.037   0
avg_merchant_per_user_s  0.021 0.006      0.008    0.021      0.033  0.021   0

Model hyperparameters:
                                                     mean    sd 0.025quant 0.5quant 0.975quant  mode
size for the nbinomial observations (1/overdispersion) 14.55 0.386      13.80    14.54      15.31 14.54

Deviance Information Criterion (DIC) ...............: 48566.97
Deviance Information Criterion (DIC, saturated) ....: 2898.36
Effective number of parameters .....................: 20.99

Watanabe-Akaike information criterion (WAIC) ...: 48466.50
Effective number of parameters .................: 29.28

Marginal log-Likelihood:  -24434.84
```

7.2 Model output for Gym ID hierarchical model

```
Fixed effects:
                           mean    sd 0.025quant 0.5quant 0.975quant   mode kld
Intercept                 8.357 0.068      8.225    8.357      8.490  8.357   0
GymSiteTypeResidential    0.060 0.062     -0.062    0.060      0.183  0.060   0
GymSiteTypeWorkforce     -0.123 0.095     -0.310   -0.123      0.064 -0.123   0
GymParkingParking         0.119 0.047      0.027    0.119      0.210  0.119   0
unis_within_0_0.5_mile_s  0.040 0.020      0.001    0.040      0.079  0.040   0
unis_within_0.5_1_mile_s -0.001 0.025     -0.050   -0.001      0.049 -0.001   0
unis_within_1_2_mile_s    0.010 0.032     -0.052    0.010      0.072  0.010   0
PromoPercentage_s        -0.039 0.004     -0.047   -0.039     -0.031 -0.039   0
AvgJoiningFee_s          -0.020 0.005     -0.029   -0.020     -0.011 -0.020   0
AvgAccountPayment_s       0.004 0.006     -0.008    0.004      0.016  0.004   0
AvgAccountPayment_s2     -0.001 0.001     -0.003   -0.001      0.002 -0.001   0
Dens_0_1_s                0.203 0.033      0.139    0.203      0.268  0.203   0
Dens_1_4_s               -0.030 0.048     -0.125   -0.030      0.065 -0.030   0
INNER_RING_SHARE_s       -0.066 0.033     -0.130   -0.066     -0.001 -0.066   0
CompetitorIndex_s        -0.199 0.046     -0.290   -0.199     -0.108 -0.199   0
distance_weighted_spend_s  0.030 0.023    -0.014    0.030      0.075  0.030   0
user_repeat_rate_s       -0.047 0.017     -0.080   -0.047     -0.014 -0.047   0
RATIO_1_TO_2_s           -0.008 0.028     -0.064   -0.008      0.048 -0.008   0
RATIO_0_5_TO_1_s         -0.038 0.017     -0.072   -0.038     -0.004 -0.038   0
avg_merchant_per_user_s   0.022 0.020     -0.017    0.022      0.060  0.022   0

Random effects:
  Name    Model
    f IID model

Model hyperparameters:
                                                        mean   sd 0.025quant 0.5quant 0.975quant  mode
size for the nbinomial observations (1/overdispersion) 49.63 1.44      46.85    49.62      52.53 49.58
Precision for f                                        18.57 1.82      15.21    18.49      22.38 18.36

Deviance Information Criterion (DIC) ...............: 45376.66
Deviance Information Criterion (DIC, saturated) ....: -51378461857.14
Effective number of parameters .....................: 232.48

Watanabe-Akaike information criterion (WAIC) ...: 44530.65
Effective number of parameters .................: 279.83

Marginal log-Likelihood:  -23139.25
```

7.3 Model output for regional hierarchical model

```
Fixed effects:
                          mean    sd 0.025quant 0.5quant 0.975quant   mode kld
Intercept                8.372 0.026      8.321    8.372      8.422  8.372   0
GymSiteTypeResidential   0.072 0.020      0.032    0.072      0.112  0.072   0
GymSiteTypeWorkforce    -0.064 0.032     -0.125   -0.064     -0.002 -0.064   0
GymParkingParking        0.104 0.016      0.074    0.104      0.134  0.104   0
unis_within_0_0.5_mile_s  0.045 0.006      0.032    0.045      0.057  0.045   0
unis_within_0.5_1_mile_s -0.016 0.009     -0.033   -0.016      0.000 -0.016   0
unis_within_1_2_mile_s   0.012 0.011     -0.009    0.012      0.033  0.012   0
PromoPercentage_s       -0.018 0.006     -0.030   -0.018     -0.006 -0.018   0
AvgJoiningFee_s          0.009 0.006     -0.002    0.009      0.021  0.009   0
AvgAccountPayment_s      0.016 0.008      0.000    0.016      0.033  0.016   0
AvgAccountPayment_s2    -0.003 0.002     -0.006   -0.003      0.001 -0.003   0
Dens_0_1_s               0.201 0.011      0.180    0.201      0.223  0.201   0
Dens_1_4_s              -0.034 0.019     -0.072   -0.034      0.004 -0.034   0
INNER_RING_SHARE_s      -0.053 0.011     -0.075   -0.053     -0.031 -0.053   0
CompetitorIndex_s       -0.190 0.015     -0.219   -0.190     -0.160 -0.190   0
distance_weighted_spend_s  0.047 0.008      0.031    0.047      0.064  0.047   0
user_repeat_rate_s      -0.043 0.006     -0.054   -0.043     -0.032 -0.043   0
RATIO_1_TO_2_s          -0.021 0.009     -0.039   -0.021     -0.002 -0.021   0
RATIO_0_5_TO_1_s        -0.044 0.006     -0.055   -0.044     -0.033 -0.044   0
avg_merchant_per_user_s  0.004 0.008     -0.011    0.004      0.020  0.004   0

Random effects:
  Name    Model
    f IID model

Model hyperparameters:
                                                   mean      sd 0.025quant 0.5quant 0.975quant   mode
size for the nbinomial observations (1/overdispersion)  15.07   0.402      14.30    15.07      15.88  15.06
Precision for f                                  378.77 166.253     148.57   347.75     789.41 292.33

Deviance Information Criterion (DIC) ...............: 48478.42
Deviance Information Criterion (DIC, saturated) ....: 2910.64
Effective number of parameters .....................: 33.56

Watanabe-Akaike information criterion (WAIC) ...: 48359.83
Effective number of parameters .................: 37.78

Marginal log-Likelihood:  -24403.15
```

7.4 Model output for SPDE model with all variables

```
Fixed effects:
                          mean    sd 0.025quant 0.5quant 0.975quant   mode kld
Intercept                8.392 0.089      8.217    8.392      8.566  8.392   0
GymSiteTypeResidential   0.073 0.068     -0.061    0.073      0.208  0.073   0
GymSiteTypeWorkforce     0.047 0.097     -0.144    0.047      0.237  0.047   0
GymParkingParking        0.102 0.049      0.005    0.102      0.199  0.102   0
unis_within_0_0.5_mile_s  0.028 0.022     -0.016    0.028      0.071  0.028   0
unis_within_0.5_1_mile_s -0.085 0.035     -0.153   -0.085     -0.016 -0.085   0
unis_within_1_2_mile_s  -0.011 0.049     -0.106   -0.011      0.085 -0.011   0
PromoPercentage_s       -0.068 0.005     -0.079   -0.068     -0.058 -0.068   0
AvgAccountPayment_s      0.018 0.008      0.003    0.018      0.033  0.018   0
AvgJoiningFee_s         -0.004 0.005     -0.013   -0.004      0.006 -0.004   0
AvgAccountPayment_s2    -0.002 0.001     -0.005   -0.002      0.000 -0.002   0
Dens_0_1_s               0.227 0.040      0.150    0.227      0.305  0.227   0
Dens_1_4_s               0.106 0.094     -0.077    0.106      0.290  0.106   0
INNER_RING_SHARE_s      -0.053 0.036     -0.123   -0.053      0.018 -0.053   0
CompetitorIndex_s       -0.178 0.066     -0.307   -0.178     -0.049 -0.178   0
distance_weighted_spend_s  0.010 0.023     -0.035    0.010      0.055  0.010   0
user_repeat_rate_s      -0.049 0.017     -0.083   -0.049     -0.016 -0.049   0
RATIO_1_TO_2_s          -0.002 0.027     -0.056   -0.002      0.052 -0.002   0
RATIO_0_5_TO_1_s        -0.039 0.017     -0.073   -0.039     -0.005 -0.039   0
avg_merchant_per_user_s  0.013 0.022     -0.030    0.013      0.055  0.013   0

Random effects:
  Name    Model
    spde_space SPDE2 model
   spde_time SPDE2 model

Model hyperparameters:
                                                   mean    sd 0.025quant 0.5quant 0.975quant   mode
size for the nbinomial observations (1/overdispersion) 51.848 1.521     48.919   51.825     54.909 51.777
Range for spde_space                              0.106 0.007      0.093    0.106      0.120  0.105
Stdev for spde_space                              0.521 0.026      0.471    0.520      0.575  0.519
Range for spde_time                               0.339 0.401      0.007    0.193      1.425  0.009
Stdev for spde_time                               0.044 0.011      0.027    0.043      0.071  0.040

Deviance Information Criterion (DIC) ...............: 45257.65
Deviance Information Criterion (DIC, saturated) ....: -38861947922.88
Effective number of parameters .....................: 236.03

Watanabe-Akaike information criterion (WAIC) ...: 45337.62
Effective number of parameters .................: 736.55
```

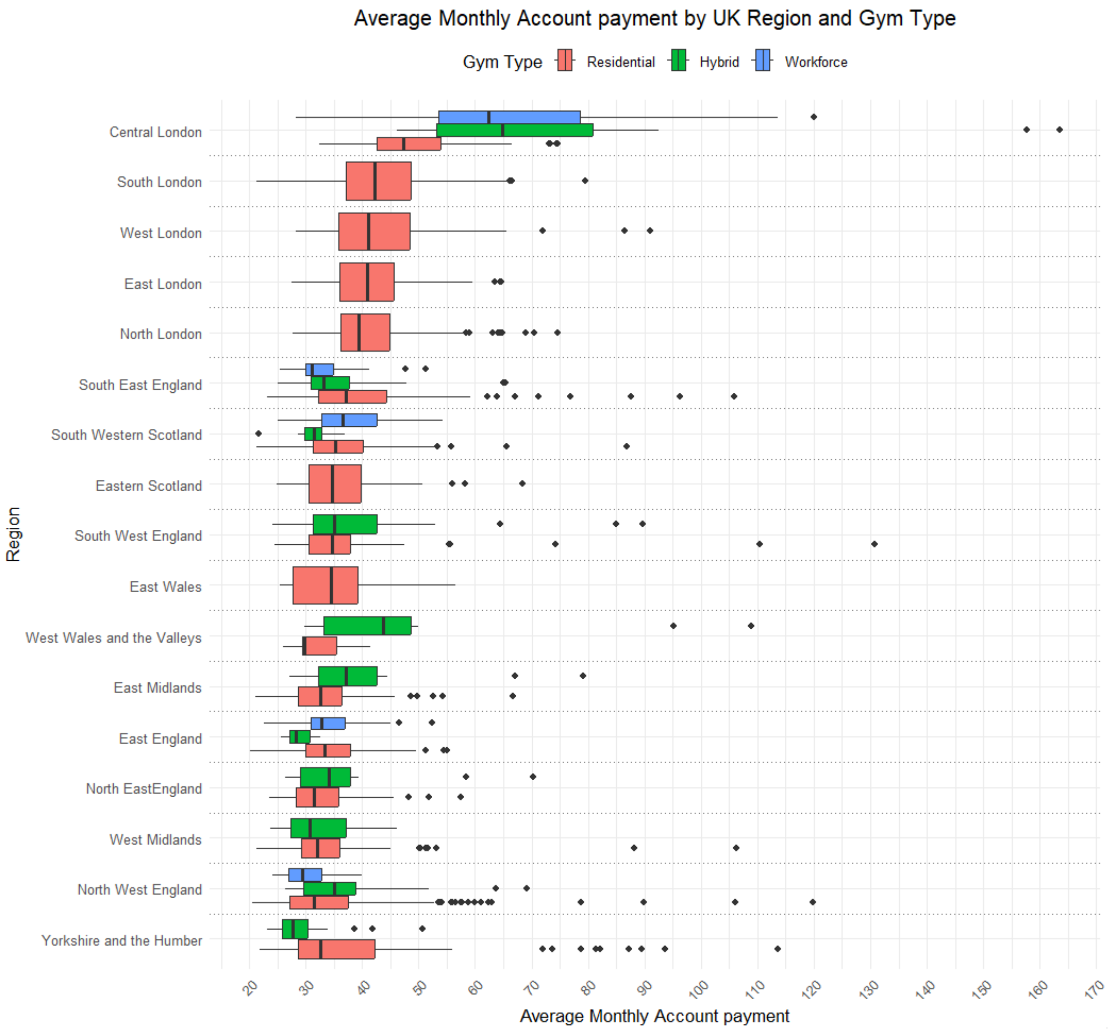
7.5 Model output for SPDE model with reduced number of variables

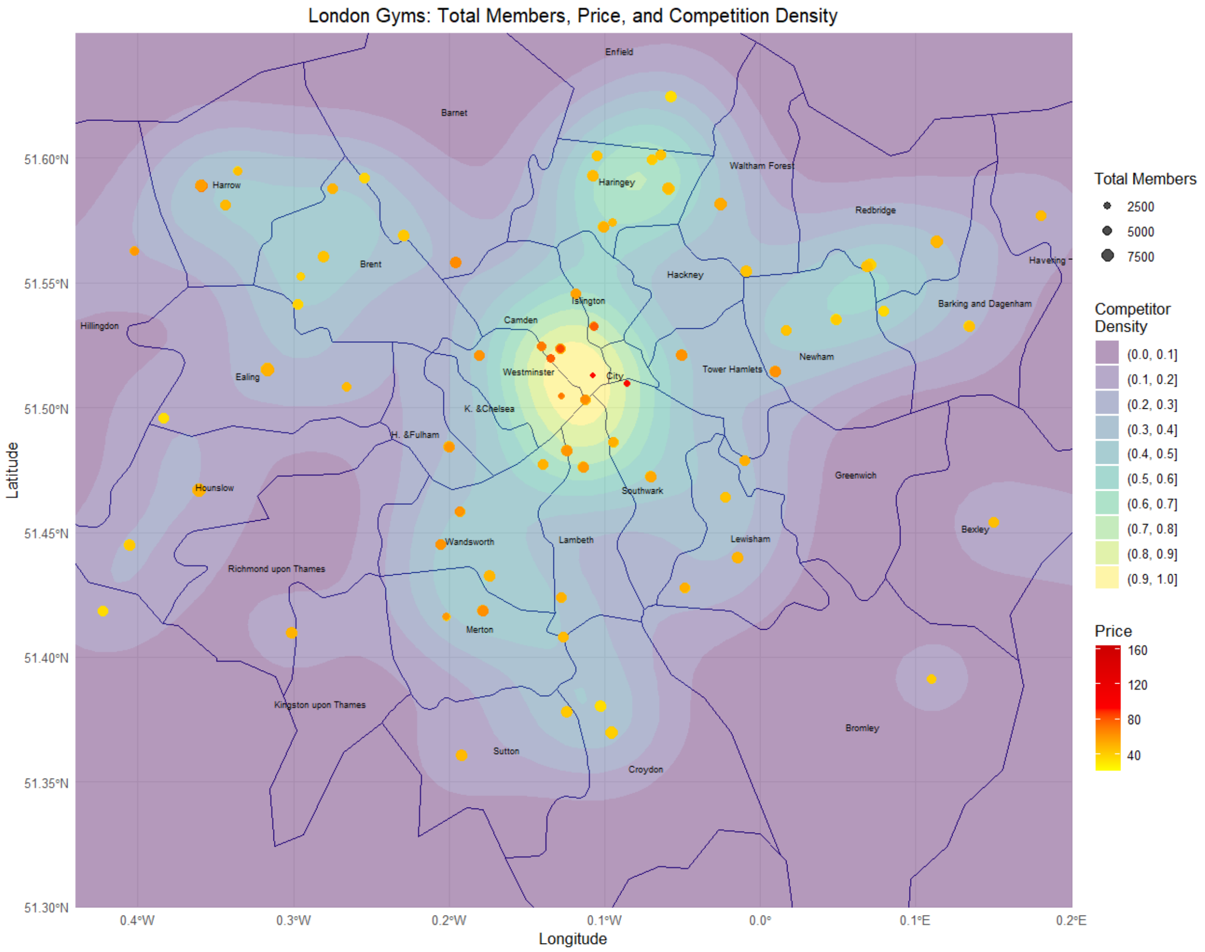
```
Fixed effects:
                            mean    sd 0.025quant 0.5quant 0.975quant   mode kld
Intercept                  8.411 0.088      8.238    8.411      8.584  8.411   0
GymSiteTypeResidential     0.062 0.068     -0.071    0.062      0.195  0.062   0
GymSiteTypeWorkforce       0.050 0.093     -0.132    0.050      0.232  0.050   0
GymParkingParking          0.134 0.048      0.041    0.134      0.229  0.134   0
unis_within_0_0.5_mile_s   0.029 0.019     -0.009    0.029      0.067  0.029   0
unis_within_0.5_1_mile_s  -0.096 0.030     -0.154   -0.096     -0.037 -0.096   0
PromoPercentage_s         -0.039 0.004     -0.047   -0.039     -0.031 -0.039   0
AvgAccountPayment_s        0.003 0.006     -0.009    0.003      0.015  0.003   0
AvgJoiningFee_s           -0.019 0.005     -0.028   -0.019     -0.010 -0.019   0
AvgAccountPayment_s2      -0.001 0.001     -0.003   -0.001      0.002 -0.001   0
Dens_0_1_s                 0.165 0.035      0.097    0.165      0.234  0.165   0
Dens_1_4_s                 0.279 0.076      0.131    0.279      0.428  0.279   0
CompetitorIndex_s         -0.187 0.056     -0.297   -0.187     -0.079 -0.187   0
distance_weighted_spend_s  0.014 0.022     -0.029    0.014      0.057  0.014   0
user_repeat_rate_s        -0.065 0.016     -0.097   -0.065     -0.033 -0.065   0

Random effects:
  Name     Model
    spde_space SPDE2 model

Model hyperparameters:
                                                         mean    sd 0.025quant 0.5quant 0.975quant   mode
size for the nbinomial observations (1/overdispersion) 49.499 1.443     46.712   49.480     52.395 49.446
Range for spde_space                                    0.108 0.007      0.095    0.108      0.123  0.108
Stdev for spde_space                                    0.526 0.026      0.476    0.525      0.580  0.524

Deviance Information Criterion (DIC) ...............: 45384.97
Deviance Information Criterion (DIC, saturated) ....: -47615429309.23
Effective number of parameters .....................: 233.16

Watanabe-Akaike information criterion (WAIC) ...: 44561.29
Effective number of parameters .................: 291.89
```

7.6 Model output for SPDE model with reduced number of variables and SPDE time

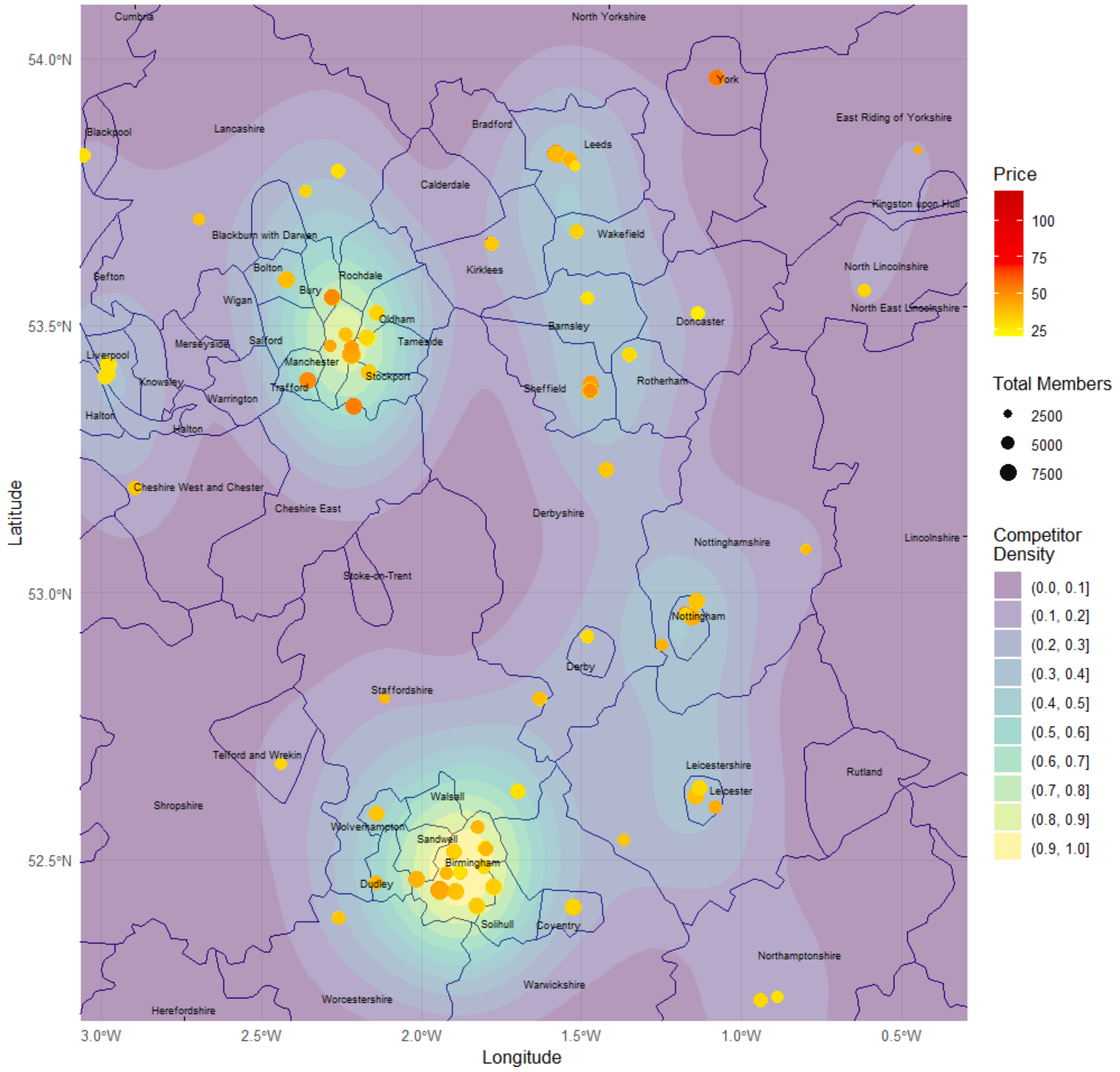
```
Fixed effects:
                            mean    sd 0.025quant 0.5quant 0.975quant   mode kld
Intercept                  8.415 0.090      8.238    8.415      8.592  8.415   0
GymSiteTypeResidential     0.058 0.068     -0.076    0.058      0.192  0.058   0
GymSiteTypeWorkforce       0.048 0.093     -0.135    0.048      0.231  0.048   0
GymParkingParking          0.133 0.048      0.039    0.133      0.228  0.133   0
unis_within_0_0.5_mile_s   0.025 0.019     -0.013    0.025      0.063  0.025   0
unis_within_0.5_1_mile_s  -0.102 0.030     -0.161   -0.102     -0.043 -0.102   0
PromoPercentage_s         -0.068 0.005     -0.079   -0.068     -0.057 -0.068   0
AvgAccountPayment_s        0.017 0.008      0.001    0.017      0.032  0.017   0
AvgJoiningFee_s           -0.004 0.005     -0.014   -0.004      0.006 -0.004   0
AvgAccountPayment_s2      -0.002 0.001     -0.005   -0.002      0.000 -0.002   0
Dens_0_1_s                 0.168 0.035      0.099    0.168      0.237  0.168   0
Dens_1_4_s                 0.277 0.076      0.128    0.277      0.427  0.277   0
CompetitorIndex_s         -0.170 0.056     -0.280   -0.170     -0.060 -0.170   0
distance_weighted_spend_s  0.010 0.022     -0.033    0.010      0.054  0.010   0
user_repeat_rate_s        -0.066 0.016     -0.099   -0.066     -0.034 -0.066   0

Random effects:
  Name     Model
    spde_space SPDE2 model
   spde_time SPDE2 model

Model hyperparameters:
                                                         mean    sd 0.025quant 0.5quant 0.975quant   mode
size for the nbinomial observations (1/overdispersion) 51.847 1.521     48.906   51.829     54.892 51.802
Range for spde_space                                    0.108 0.007      0.095    0.108      0.123  0.108
Stdev for spde_space                                    0.530 0.026      0.480    0.529      0.584  0.527
Range for spde_time                                     0.379 0.515      0.006    0.193      1.759  0.007
Stdev for spde_time                                     0.044 0.011      0.026    0.042      0.070  0.040

Deviance Information Criterion (DIC) ...............: 45257.44
Deviance Information Criterion (DIC, saturated) ....: -39858899733.23
Effective number of parameters .....................: 236.03

Watanabe-Akaike information criterion (WAIC) ...: 45336.25
Effective number of parameters .................: 736.15
```

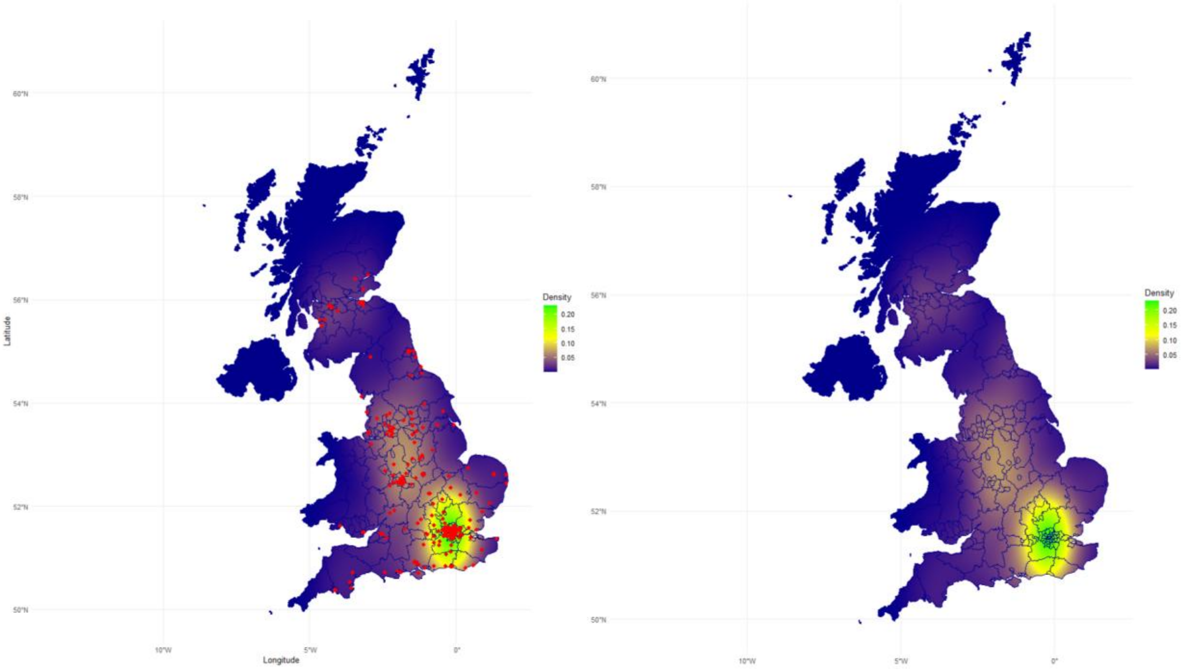
7.7 EDA extra plots


Average Monthly Account payment by UK Region and Gym Type
Gym Type
Residential
Hybrid
Workforce
Region
Central London
South London
West London
East London
North London
South East England
South Western Scotland
Eastern Scotland
South West England
East Wales
West Wales and the Valleys
East Midlands
East England
North EastEngland
West Midlands
North West England
Yorkshire and the Humber
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
Average Monthly Account payment


London Gyms: Total Members, Price, and Competition Density
Latitude
Longitude
51.60°N
51.55°N
51.50°N
51.45°N
51.40°N
51.35°N
51.30°N
0.4°W
0.3°W
0.2°W
0.1°W
0.0°
0.1°E
0.2°E
Enfield
Barnet
Harrow
Haringey
Waltham Forest
Redbridge
Havering
Brent
Hackney
Islington
Camden
Barking and Dagenham
Hillingdon
Newham
Westminster
City
Tower Hamlets
Ealing
K. &Chelsea
H. &Fulham
Greenwich
Hounslow
Southwark
Bexley
Wandsworth
Lambeth
Lewisham
Richmond upon Thames
Merton
Kingston upon Thames
Bromley
Sutton
Croydon
Total Members
2500
5000
7500
Competitor Density
(0.0, 0.1]
(0.1, 0.2]
(0.2, 0.3]
(0.3, 0.4]
(0.4, 0.5]
(0.5, 0.6]
(0.6, 0.7]
(0.7, 0.8]
(0.8, 0.9]
(0.9, 1.0]
Price
160
120
80
40

Midlands Gyms: Total Members, Price, and Competition Density





KDE of total members and gyms locations (red dots).