Collaborative Toponyms in OpenStreetMap: The Case of Amsterdam

Reproduction and Replication Report for Data Mastery Challenge

June Okongo - s3473600
Jonathan Polidano - s3745643
Mashal Riaz - s3690970

Table of Contents

1. Introduction

The OpenStreetMap (OSM) is a platform that supports increasing availability of volunteered geographic information (VGI), enabling large-scale analyses of toponym representation in urban areas. The presence or absence of populated name attributes linked to OSM features are used in research as a substitute for developing an understanding of the spatial distribution of named places and their association of intrinsic qualities of the built features in urban environment. An empirical study "Collaborative toponyms in OpenStreetMap: an open-source framework to investigate the relationship with intrinsic quality parameters" by Nunes and Camboim (2025) explored these relationships by using a grid-based aggregation and spatial statistics including regression analysis. The analysis provided evidence of variations in a systematic way across the OSM tag categories and prominent spatial clusters of the toponyms.

This project approached the analysis in two steps. The first was to reproduce the analytical logic and methodological workflow of the original study for Brazil on the standard 200m grid. This was accomplished by following the same conceptual framework, definition of variables and statistical methods. This step focused mainly on implementation of the original steps with emphasis on transparency of method and reproducibility rather than any modification. The second step was the replication of the study in a new geographical location, Amsterdam using a standardized 100m grid. The aim of this replication was to assess whether the observation made in the case of Brazil can persist in the Dutch city with different urban form, mapping practices and naming conventions. Following the city-wide analysis, the project was further focused on selected sub-area within Amsterdam for a detailed spatial analysis and observation of local toponym trends.

By adopting the approach of reproducibility and then replicability of the original study, this project focuses on assessing the robustness and transferability of the original findings, while reflecting critically on the methodological and technical challenges in spatial data analysis via volunteered geographic information.

2. Hypotheses

It was anticipated that many of the main findings from the original study would prove applicable across the new study-area, despite the significant geographical differences between Amsterdam Centrum and the two Brazilian neighbourhoods in the original study. Thus, consistent with the original findings, it was hypothesised that:

1. The OSM tag categories would dispaly notable differences in the proportion of features with populated name attributes. Specifically, it was hypothesised that leisure and amenity features would be more frequently named and building features less frequently named.
2. The OSM tag categories would display varying regression relationship strengths between intrinsic quality parameters and the proportion of populated toponyms. Specifically, it was

hypothesised that ordinary least squares (OLS) regression would obtain higher R^2 values for the leisure and amenity tags and lower R^2 values for 'buildings' tags.

3. Significant spatial autocorrelation would be found for all intrinsic quality parameters, as well as for the proportion of populated toponyms per grid cell, across each OSM tag category.
4. Although there would be spatial clustering of OSM toponyms, the specific configuration of hot and cold spots would differ from that reported in the original study due to the differing study area. Specifically, it was hypothesised that hotspots would typically concentrate towards the old core (the De Wallen area), wih cold spots more prevalent along the southern, western, and eastern margins (the comparatively more recent areas such as Jordaan, the Spiegelkwartier and Plantage).

3. Method

3.1. Workflow

This study started by reproducing the results of the original study. However, while the original study analysed two separate study-areas, in the interests of time the current study reproduced the results for only one (the neighbourhood of Água Verde in the city of Curitiba, Paraná, Brazil). The analysis was then replicated by performing it again for the Centrum district of Amsterdam. Following this, the analysis was performed on a smaller study-area (the Centrum-West ward, approximately half of the Centrum district) using two grids each with different-sized cells, in order to understand the impact of grid cell size on results. Further details on grid cell sizes are given in Section 3.3.

Following the original study, the research presented here obtained OpenStreetMap data for roughly 16 years' worth of data (2007-10-08 to 2024-03-10) for the study-area for each of the five variables outlined in Table 1 below. Variable 1 in the Table served as the dependent variable in subsequent regression analysis, while the remaining variables served as independent variables. The latter are 'intrinsic quality indicators' of OSM data which emerge from the literature on the topic (Nunes and Camboim, 2025). Their usefulness is that they are characteristics of OSM data that can indicate the quality of said data. Therefore, they allow for estimation of OSM data quality even in cases where comparison to external reference data is not possible. The original research strove to analyse the relationship in their study-areas between the quantity of named OSM features (especially the relative proportion of features that are named) and the intrinsic quality parameters, striving to map areas where OSM data may be of sufficient quantity and quality to be a useful source of information to official mapping agencies.

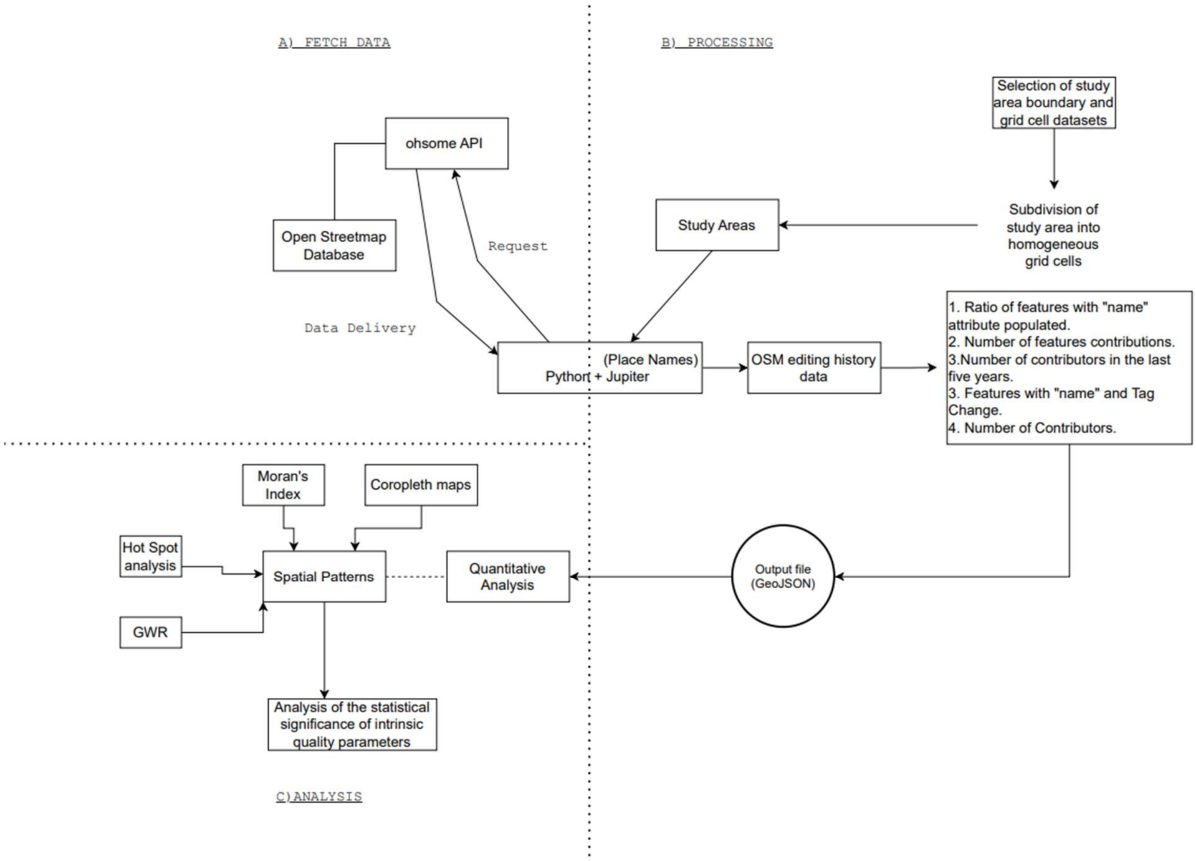


Figure 1: Workflow for the research.

Variable	Description	Example
(1) Number and proportion of named features (OSM toponyms)	Calculate the number of OSM features for the specified tags and the proportion of these features with the "name" attribute populated.	A total of 5,098 building features were retrieved, of which 759 (14.9%) had the name attribute populated.
(2) Total number of contributions	Count the total number of contributions associated with elements bearing the specified tags, distinguishing between those with and without the "name" attribute populated.	The amenity features received 2,374 contributions, 690 of which were associated with features with the name attribute populated.
(3) Recent contributions	Calculate the number of contributions over the past five years for interest	Between 2019 and 2024, building features with the name

(past five years)	tags with the "name" attribute populated.	attribute received 239 contributions, indicating continued activity.
(4) Tag change contributions	Obtain the total number of contributions to features that have the "name" attribute populated and have undergone a tag change.	A total of 114 tag change contributions were identified in named building features. These reflect direct refinements to existing attribute information.
(5) Number of distinct contributors	Quantify the number of contributors (users) who edited features with the "name" attribute populated.	For named amenity features, 94 unique contributors were identified across the study area.

Table 1: The variables used for analysis (Source: Nunes and Camboim, 2025).

The variable data were obtained for each of three OSM tag categories: leisure, amenity, and building tags. Once the above variables were obtained, and again following the original study, statistical analysis was performed for each tag category as follows:

1. Preliminary exploratory analysis of the dependent variables using histograms and box plot analyses.
2. Ordinary Least Squares (OLS) regression to understand the relationship between the proportion of named features and the intrinsic quality indicators.
3. Choropleth maps of the proportion of named features, to enable visual exploration of spatial patterns.
4. Global Moran's I to calculate the spatial autocorrelation of each variable, both dependent and independent.
5. Getis-Ord Gi* hotspot analysis to identify statistically significant clusters of high/low values for the dependent variable; that is, areas where there is a significantly high or significantly low proportion of named features.
6. Geographically Weighted Regression (GWR) to understand spatial differences in the strength of the relationships that were measured globally with OLS in step (2). This was performed using Principal Component Analysis (PCA) when multicollinearity values were found to be high. First, the Variance Inflation Factor (VIF) was used to identify multicollinearity between the intrinsic quality indicators. Then, PCA was performed to reduce dimensionality of the dataset based upon VIF. Finally, GWR was performed using the retained principal components as independent variables.

3.2. Code

The code used by Nunes and Camboim (2025) was publicly available on a GitHub repository, the URL to which is listed in their paper. In this repository, the code was stored in three

Jupyter notebooks. Of these, the third generated treemaps of the OSM tag sub-categories. Since its output was not used in the original study paper, it was excluded from the current study. The remaining notebooks were run using JupyterLab, on the University of Twente's Geospatial Computing Platform (GCP).

The first notebook took a GeoJSON file containing a grid of the study-area as input and retrieved OpenStreetMap data for each grid cell using the ohsome API (Raifer et al., 2024). This API is open-source and enables retrieval of OSM historical data. The notebook divided the study-area grid cells into subsets of four cells each. For each of the five variables, it requested data one subset at a time. Given that each subset could take anywhere between two and five minutes to complete, the first notebook took a significant duration to run. The notebook output was another GeoJSON of the same study-area but with the variable data attached as attributes. Code run times depended upon factors such as the volume of requests being made to the GCP at any time. However, run time for the reproduction of the Água Verde dataset (154 grid cells of 200m x 200m each) was approximately 2.5 hours. The run time for replication to Amsterdam Centrum (818 cells of 100m x 100m each) was significantly longer, taking approximately 14 hours.

The second notebook took the output of the first notebook as its input and performed all the statistical analyses outlined in Section 3.1. While the notebook took much less time to run, it frequently generated dropdown widgets with which the user had to manually select the variables for analysis. This meant that the code could not be run from start to finish but required manual intervention for each of the total of 30 outputs generated by this notebook per study-area. However, computer processing times were minor for this notebook, with the longest-running cell lasting for less than half a minute. The outputs were a mix of .png images, grid cell maps in .html format, and data tables in .txt and .csv formats.

Prior to reproduction, some minor debugging was performed to enable the notebooks to function so that they could be understood. For example, cells functioning specifically on Google Collab were removed. Furthermore, a prior version of some libraries such as Matplotlib seems to have been used in the original, so some code had to be updated accordingly. Given the code's complexity, a full reproduction run was required for the researchers to gain understanding of the code's specifics.

Prior to replication, a more intensive debugging and editing process was conducted. Debugging was required mainly due to minor differences between the original input datasets and the replication dataset (e.g., they had a different CRS as will be mentioned in Section 3.3). Some instances of repetitive code were cleaned up, and output file names were standardised such that they were automatically populated with the study area name and other important information. Furthermore, the requests Notebook 1 made to the API often timed out, resulting in a TimeOut error and the need to restart the code. This was mitigated against by introducing a retry function that re-sent the request up to a maximum of five times instead of immediately delivering an error. Furthermore, code cells were introduced to save progress upon completion of each major step of the notebook. A more thorough simplification of the code, eliminating the use of drop-down widgets entirely and thus removing the manual input required, would have been ideal but was not possible given time constraints.

3.3. Input Data

As mentioned above, reproduction was performed using the GeoJSON grid for Água Verde, which was prepared by the original study authors and was available in their GitHub repository. The grid was 200m x 200m since it utilised the official statistical grid provided by Brazilian Institute of Geography and Statistics (Nunes and Camboim, 2025). The original study authors had extracted this grid for the study-area, with any grid cell intersecting even partially with the study-area being retained.

A similar input grid was prepared for Amsterdam Centrum, which was of comparable size to Água Verde. However, the official statistical grid for Amsterdam is 100m x 100m. Therefore, the resulting grid had roughly four times the number of grid cells as the original study grids, explaining the significantly longer run time. The projection system of the original was EPSG: 4674.

The Amsterdam city district boundaries and the statistical grid are available for download open-source from the Municipality of Amsterdam website (Gemeente Amsterdam, n.d.a., n.d.b.). After download in GeoJSON format, they were opened in QGIS and projected to EPSG: 28992 using the Reproject Layer tool. The Centrum district was manually selected from the city districts dataset, and the grid cells that intersect at least partially with it were extracted using the Extract by Location tool (setting the city districts layer as the comparison features and ticking 'Selected features only').

For the secondary objective, the point of which was comparison of different grid cell sizes, it was decided to use only the ward of Centrum-West, which forms approximately half of Centrum, due to the long code processing time. The ward boundaries dataset was downloaded from the Municipality of Amsterdam website (Gemeente Amsterdam, n.d.a.) and the intersecting grid cells extracted using an identical procedure to the above.

A 200m x 200m grid was then generated for comparison purposes. The Create Grid tool was used on QGIS, with Grid Type being set to Polygon, and Grid Extent to the extracted 100m grid (ensuring that the resulting grid cells would fit perfectly over the 100m grid, with four 110m cells fitting perfectly in each 200m cell). The Horizontal and Vertical Spacing was set to 200m. Once generated, Extract by Location was again used, extracting features from the 200m grid by comparing to the 100m grid. This removed most of the excess 200m cells. Finally, any remaining cells from both the 200m and the 100m layer that were not perfectly covering each other were selected manually and deleted.

3.4. Software

The latest available versions were used throughout this research for all software required. The ohsome API latest version was v1.10.4. For JupyterLab, version 3.4.8 was used. The Python version used was 3.8.10. Withing JupyterLab, it was noticed that the selenium and typing_extensions libraries seemed not to be fully updated, so a code cell was added at the beginning of Notebook 2 (which requires these libraries) to install the latest upgrades for these. The QGIS version was 3.40.10 (Bratislava).

4. Results

4.1. Reproduction Results

The reproduction of the original study was largely successful, mostly because we could obtain the same results using the original data and computation workflow with a minor negligible variance on the Geographically Weighted Regression (GWR) results' section. This demonstrated that the conceptual workflow and methodological choices described in the original paper were sufficiently transparent and reusable. By following the documented data sources, processing steps, and analytical procedures, we were able to obtain outputs that closely matched the original findings. Overall, this demonstrates a high degree of reproducibility and supports the validity of the original study's hypothesis.

4.2. Results for the Full Amsterdam Centrum 100m Grid

For the statistical analysis, the results showed a consistency in spatial structure and internal coherent relationships among the name-related parameters for buildings, amenity and leisure features. The descriptive statistics and boxplots indicated a visible heterogeneity, where the name ratios were highly concentrated in the central regions, presumably the activity dense cells rather than any random distribution. This was confirmed via the result of Moran's I (≈ 0.45–0.47 across the categories), and positively sloped scatterplots, demonstrating the presence of a significant statistical positive spatial autocorrelation, as cells with high values have the tendency to be surrounded by similar high-value neighbouring cells. The Choropleth and Getis-Ord Gi* maps were significant in further reinforcement of this pattern as the results revealed spatial clusters (hotspots), that were prominent in the central regions, rather than distant outliers. Regression outputs suggested that the name contributions, recent name editing activities, tag changes and the number of contributors are all positively correlated with significant multicollinearity. This was prominent between tag-change contributions and number of contributors (often > 0.9), giving an indication that these variables have captured the related aspects of editing history and participation of users. Amenities showed the strongest general intercorrelation, followed by buildings while the leisure had the most differentiated or lower correlations between name contributions and editing history activity, suggesting that the edits were either event-driven or carried out in a phase wise manner. In conclusion, the results indicated that the naming activity at 100 m scale in OpenStreetMap is high clustered and strongly steered by the contributors or user's engagement and recent editing activities, also demonstrating that there is a structural interdependence across variables. This supports the interpretation that the observations of patterns show concentrated and socially driven mapping activity rather than any random spatial distribution.

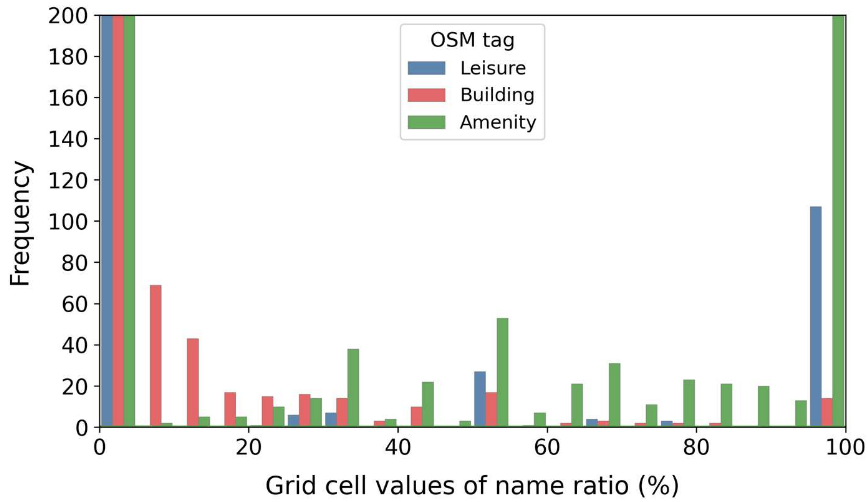


Figure 2: Histogram of name tag coverage in Amsterdam Centrum

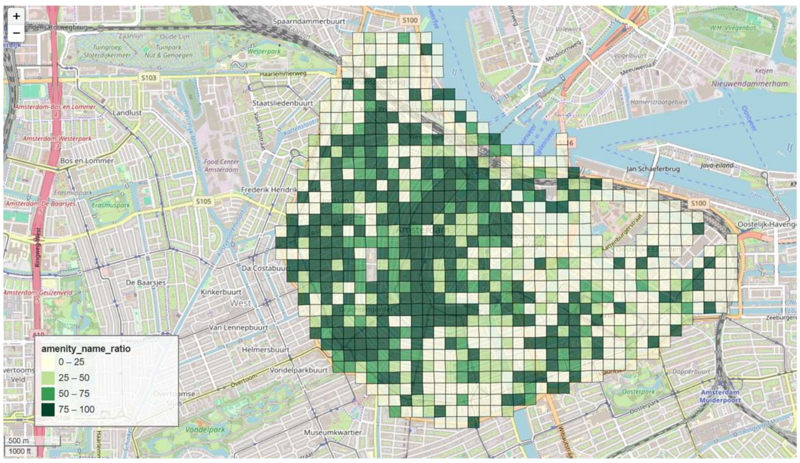


Figure 3: Choropleth map of the proportion of named Amenity tags per grid cell.

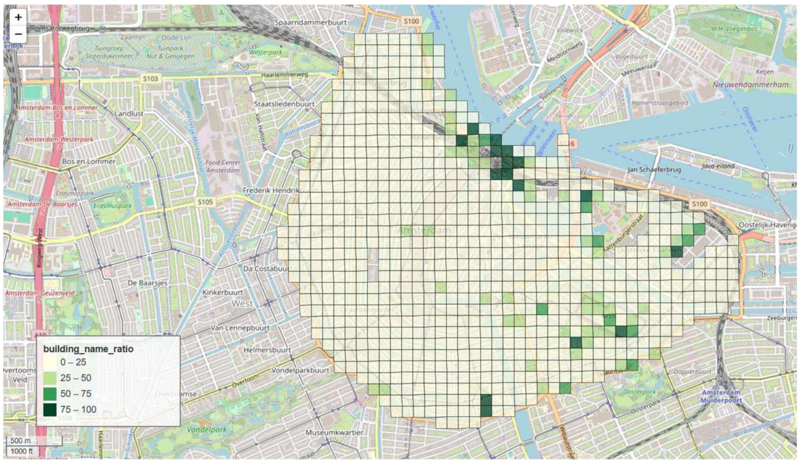


Figure 4: Choropleth map of the proportion of named Buildings tags per grid cell.

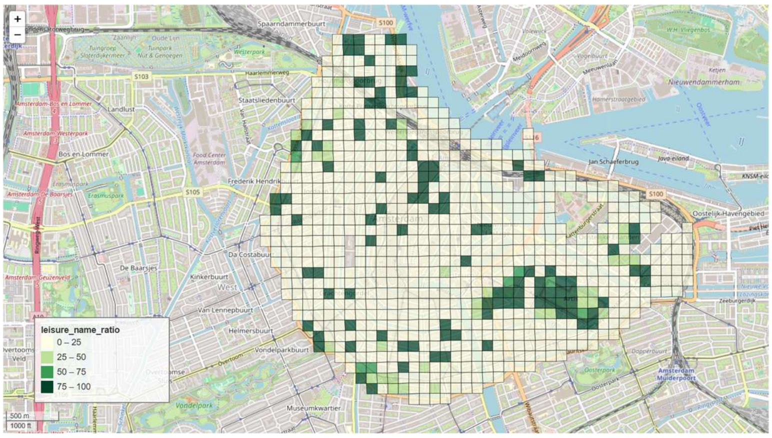


Figure 5: Choropleth map of the proportion of named Leisure tags per grid cell.

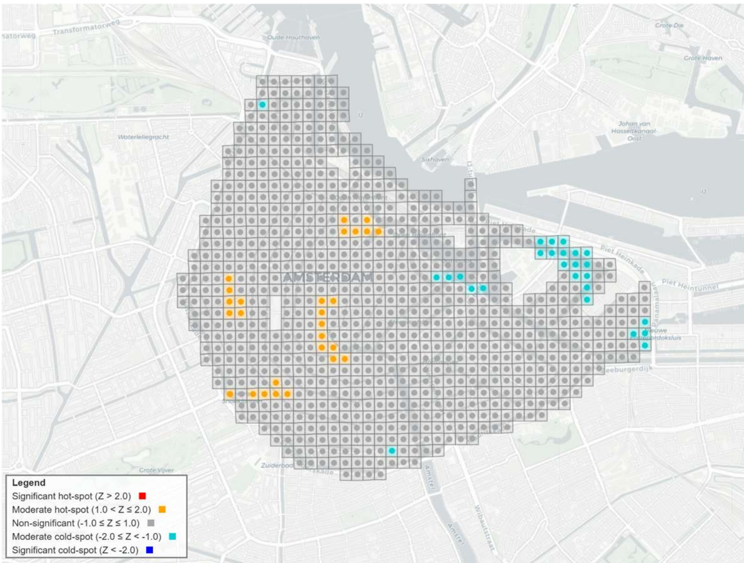


Figure 6: Getis-Ord Gi Analysis for amenity_name_ratio of Amsterdam Centrum*

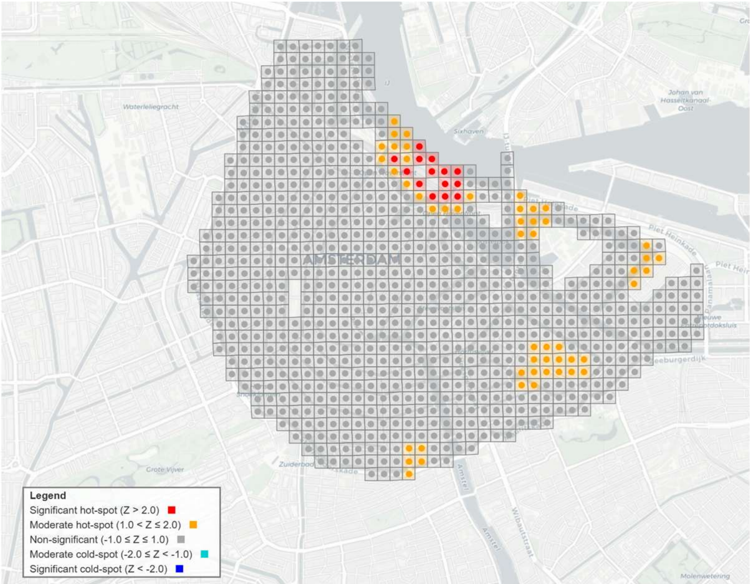


Figure 7: Getis-Ord Gi Analysis for building_name_ratio of Amsterdam Centrum*

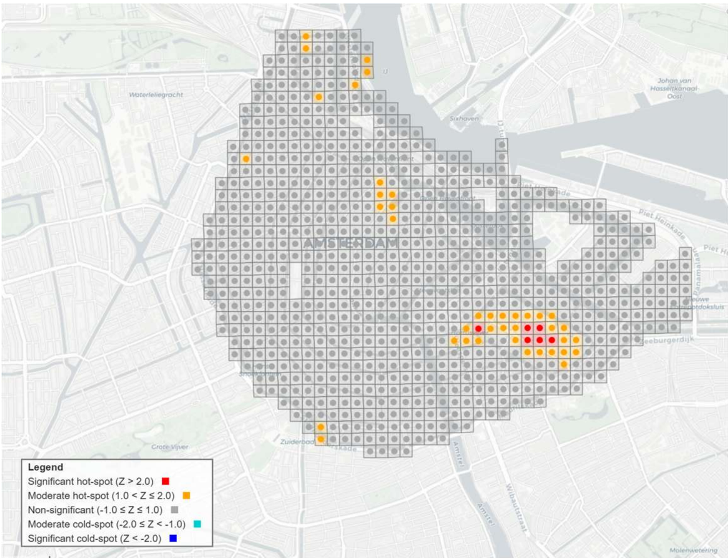


Figure 8: Getis-Ord Gi Analysis for leisure_name_ratio of Amsterdam Centrum*

4.3. The Impact of Grid Cell Size upon Results

As can be seen in Figure 5 below, the histogram for the 200m cells had a lower concentration of cells at the rightward side, indicating that coarser spatial scale diminishes the proportion of cells with high ratios of named features for all three tag categories. Perhaps in part due to this, the R^2 results obtained via OLS were lower for the 200m grid than for the 100m grid. The R^2 scores for the 100m and 200m grid are shown in Table 2 below.

Tag	100m Grid	200m Grid
Amenity	0.27	0.24
Building	0.21	0.29
Leisure	0.61	0.45

Table 2: R^2 scores for the 100m and 200m grids of Centrum-West.

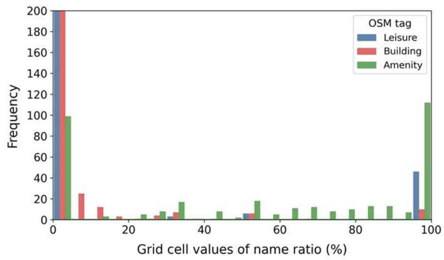
Similarly, the choropleth maps for both grid sizes were similar but revealed that some contrast between areas with high and low proportions of named features had been lost in the 200m grid for the Amenity and Leisure tags, as can be seen in Figure 9. However, the 200m Building choropleth sharpened the contrast between the hotspot of named features to the east of the study-area and the surrounding areas with very low proportions of named features.

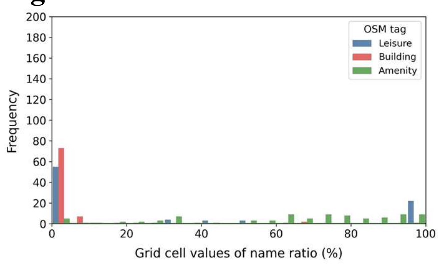
The Moran's I results generally revealed weak to moderate positive spatial autocorrelation for all 15 variables (five per tag) measured. For the 100m grid, the spatial autocorrelation scores ranged from 0.12 (total number of Leisure name contributions), 0.57 (total Building contributions). For the 200m results, spatial autocorrelation scores were slightly lower, ranging from 0.06 (Leisure tag name changes in past 5 years) to 0.55 (two variables: Amenity tag total users and Building total contributions). The generally lowered spatial autocorrelation scores once again suggest that distinct spatial patterns are degraded as grid cell size is coarsened. In similar vein, Getis-Ord Gi* hotspot maps were consistently weaker at 200m scales, with no statistically significant hotspots at all being identified for the proportion of named Amenity and Leisure features. GWR patterns were broadly similar for the 100m and 200m grids, though the maximum regression scores obtained were generally lower for the 200m grid.

Amsterdam Centrum – West results

100 x 100 grid results | **200 x 200 grid results**

1. Histograms





2. Choropleth maps

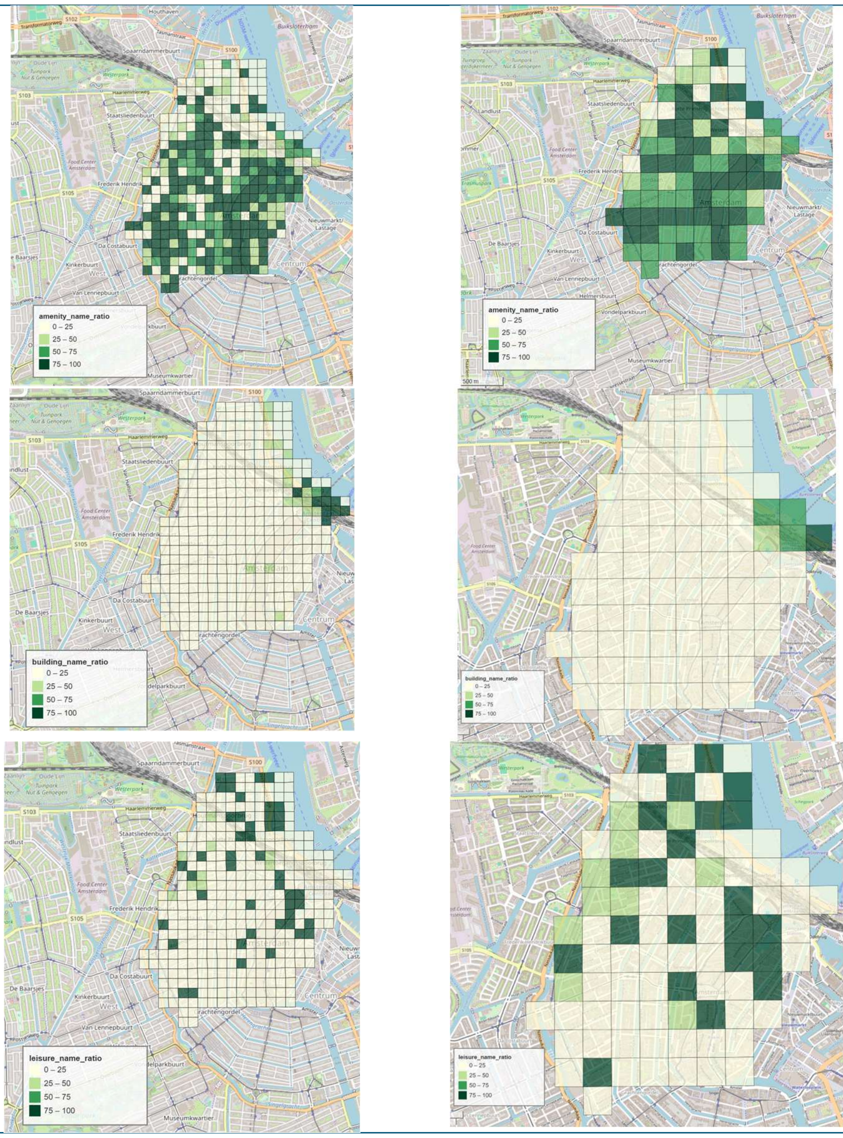


Figure 9: comparison of results

4.4. Evaluation of Hypotheses

Hypothesis 1

This hypothesis was overall validated. The histogram revealed that, although many grid cells have very low proportions of named features, it was more common for grid cells to have high proportions of named features for the Leisure and Amenity tags and less common for the Building tags.

Hypothesis 2

This hypothesis was validated. For the full study-area, the Leisure and Amenity tags (0.44 and 0.37 respectively) did indeed have higher R^2 scores than the Building tags (0.26).

Hypothesis 3

This hypothesis was validated. For the full study-area, spatial autocorrelations ranged between 0.29 (Building tag contributions in last 5 years) and 0.63 (Leisure tag contributions), indicating generally weak-to-moderate spatial autocorrelation. Interestingly, spatial autocorrelations for Centrum-West specifically were notably weaker, ranging from 0.12 to 0.57 at the comparable 100m grid cell size.

Hypothesis 4

This hypothesis was partially validated and partially falsified. There was in fact a distinct pattern of statistically significant hot and cold spots of proportions of named features. For Amenity tags, two moderate hotspots were located within or on the edges of the old city core, though another two were located on the south-western margins of Centrum. Cold spots were, as hypothesised, on the study-area boundaries, specifically to the north-east.

For Building tags, the only highly significant hotspot was located within the old city core. However, various hotspots of moderate significance were found on the eastern margins of the study-area.

The hypothesis held least for Leisure tags, with a highly significant hotspot to the east outside the old city core. Some moderately significant hotspots were identified across the rest of the study-area, one in the old core and others along the southern and western study-area margins.

5. Conclusion

This project successfully demonstrated the reproduction and replication of original study for Brazil and Amsterdam respectively. However, through the course of analysis, it was established that successful replication for a new location demands considerable interpretation of data dependencies, project related concerns, interpretation of assumptions and technical constraints. The iterative nature of debugging process, specifically related to CRS handling, data validation, API reliability with due consideration to downtime for maintenance by the service providers highlighted the significance of transparent workflow and proactive debugging approaches. The replication showed a high degree of success, where the core relationships for intrinsic quality parameters were replicated in a completely new urban context, where significant spatial difference were observed. The project also contributed towards strengthening the understanding of the limitations of VGI-based analysis, specifically the impact of processing pipelines and data completeness data completeness on the results.

References

Gemeente Amsterdam (n.d.a.). Gebieden. *Amsterdam Datadiensten*. Accessed 30 January 2026. https://api.data.amsterdam.nl/v1/docs/datasets/gebieden.html

Gemeente Amsterdam (n.d.b.). Beheerkaart CBS Grid. *Amsterdam Datadiensten*. Accessed 30 January 2026. https://api.data.amsterdam.nl/v1/docs/datasets/beheerkaart/cbs_grid.html

Nunes, D.M., and Camboim, S.P. (2025). Collaborative Toponyms in OpenStreetMap: An Open-Source Framework to Investigate the Relationship with Intrinsic Quality Parameters. *Cartography and Geographic Information Science*, pp. 1-22. doi: https://doi.org/10.1080/15230406.2025.2566807.

Raifer, M., Troilo, R., Mocnik, F.B., and Schott, M. (2024). OSHDB - OpenStreetMap History Data Analysis. *Zenodo*. Accessed 30 January 2026. https://zenodo.org/records/12580179.