Understanding participation through a data-driven approach

Androniki Pappa[1], Alexandra Paio[1], Serjoscha Duering[2], Angelos Chronis[2]

[1] Iscte – Instituto Universitário de Lisboa, Lisbon, Portugal
androniki.pappa@iscte-iul.pt; alexandra.paio@iscte-iul.pt;
[2] Austrian Institute of Technology, Vienna, Austria
serjoscha.duering@ait.ac.at; angelos.chronis@ait.ac.at

Abstract. Participatory models in urban regeneration are increasingly integrated in local agendas. Yet there is still a need for evaluation methodologies of those models and their impact. This paper presents a data-driven and computational methodology to measure the impact of the BIP/ZIP program in Lisbon. Using qualitative coding, data integration, unsupervised machine learning models for data clustering and interactive visualization dashboards the study aims to explore the large and complex dataset of the projects of BIP/ZIP and identify correlation patterns between their data and especially the areas of implementation, the networks of partners and the identified activities. Departing from the pilot-case of BIP/ZIP, the proposed methodology is a first step towards the development of a generalizable evaluation framework for participatory models in urban regeneration, that considers them as urban practices and hence evaluates them based on appropriate urban tools.

Keywords: Participatory Strategies, Participation Evaluation, Data-Driven Evaluation, Unsupervised Learning, Data Visualization.

1 Introduction

Citizen participation in urban regeneration and governance is increasingly integrated in national and local development strategies, as a tool to foster social and territorial cohesion and inclusivity. This rising turn is significantly attributed to international directives for sustainability to which national and local governments commit, that consider participatory urbanism a key for the achievement of sustainability targets. For instance, the New Urban Agenda envisions *“cities and human settlements that are participatory*” (United Nations (Habitat III), 2017, p. 5) and focus on safe, inclusive and pluralistic societies, prioritising the specific needs of those in disadvantaged situations, while the 2030 Agenda for Sustainable Development devotes its 16.7 target for sustainable development to actions that "*ensure responsive, inclusive, participatory and representative decision-making at all levels*" (Annex of the 2030 Agenda, 2017).

Nevertheless, despite the unprecedented integration of participatory processes in urban strategies at a local level, there is an imbalance between the implementation of these strategies and the assessment of their impact (Falanga, 2019). Scholars and practitioners that engage with the evaluation of urban regeneration policies that adopt participatory processes are confronted with a challenge caused predominantly by the complexity of the concept of participation and its operationalisation into practice (Falanga, 2020).

Looking at theory on participation retrospectively, the fundamental work of Sherry Arnstein (1969) "Ladder of citizen participation" focused on the conceptual categorisation of participation according to the redistribution of power between the authorities and the participants. Later, participatory theory followed the communicative turn in planning (Bahreldin, 2013), also accompanied by changes on the conceptual understanding on evaluation tools. In this regard, scholars have defined more complex frameworks grounded to the multi-dimensional reality, such as Archon Fung's (2006) *Democracy Cube* that takes into consideration multiple factors, such as who are the participants, what are the tools and how are processes linked to policymaking. Such multi-criteria approaches in the definition of participation allow for more complex correlations among the selected parameters, providing insightful considerations for the nature and operationalisation of participatory processes. In extension when applied to bodies of case studies or projects that define entire strategies, they allow for comparability and clustering and can hence be a valuable tool for the future of these strategies.

Stemming from a multi-dimensional definition and hence evaluation of participatory models, this research explores a data-driven methodology based on urban analysis rather than policy evaluation tools and highlights the importance of an assessment model that considers the data produced by the individual projects of the participatory policies.

1.1 The BIP/ZIP Program

The research is grounded on the BIP/ZIP Local Partnerships Program in Lisbon, the first participatory budget implemented at municipal level in a European capital. Launched in 2011 by the Department of Housing and Local Development of the Municipality of Lisbon, the BIP/ZIP Local Partnerships Program is part of a bigger instrument of collaborative public policy called Bairros e Zonas de Intervenção Prioritária that aims to improve the quality of life and territorial cohesion in disadvantaged neighbourhoods of the city through urban regeneration.

The Local Partnerships Program contains an annual cycle of funding for initiatives and projects ignited through local partnerships. In this sense, the program enables bottom-up initiatives to emerge and be realised in neighbourhoods of Lisbon that have been characterised as ‘priority’ after an initial diagnostic mapping of socio-economic, environmental and urban factors (CML, 2010). As of its 2021 edition, the program counts 426 realised interventions in 67 urban neighbourhoods, affecting more than 100.000 habitants per year and is characterised as an URBACT good practice (CML & DMHDL, 2017). In light of the program’s reformation and ambition to upscale its grants (An Integrated Toolbox for Deprived Neighbourhoods | URBACT, 2021), it is imperative to learn from the realised projects and inform its further development.

The need for assessment of the BIP/ZIP program has been recognised by Lisbon City Hall which in 2013 employed a team of consultants to design an evaluation model. Supporting this team and researcher that specialises in evaluation of participatory processes, Roberto Falanga (2019), published in 2019 an Index of the evaluation of the BIP/ZIP Program, focusing on the triangulation of three key elements: local partnerships, initiatives and public funding. This research aims to stem from Falanga’s fundamental work and suggest a computational data-driven approach on the analysis and evaluation of the BIP/ZIP projects, to reveal new forms of relationships between the elements of participation and their socio-territorial impact.

1.2 Data Driven evaluation

The standardisation of process through looking at data provide insights for the sustainability of participatory mechanisms and the establishment of their role in public policies (Falanga, 2019). Data-driven analysis has long been integrated and advancing in research in architecture and urban studies. For instance, KPI based evaluation of scenarios includes multiple applications, such as predictions and assessment of environmental performance or architectural design scenarios. Based on the insightful results in other domains, we explore the hypothesis of the application of such a methodology in the field of participation evaluation with the identification of the right set of KPIs. To this, we employ an already developed interactive dashboard by part of the authors (Duering et al., 2022) designed to serve the purpose of performance impact assessment for urban development. The data analysis and visualization components of the dashboard have been developed as an open-ended framework, thus allowing the integration of diverse datasets and KPIs such as the ones of the BIP/ZIP program of this study.

Beyond information that is easy to codify into KPIs for the dashboard, such as number of partners, locations and funding, a major challenge of this research has been the translation of qualitative information, such as the activities conducted within each one of the participatory projects into quantified data. To this, we followed a manual mapping process that required our expertise and review of sustainability impact indicators.

2 Methodology

2.1 Data Collection and cleaning

The first part of the methodology refers to the collection of data and their preparation, cleaning and feature extraction for the composition of a dataset. The initial phase of the evaluation methodology described in this paper is dedicated merely to understanding correlations and patterns that stem from the existing raw data. The selection of appropriate variables to measure the impact of the program in the city of Lisbon has been significantly shaped by the available public data of the projects. The main sources are the applications of the projects published at the BIP/ZIP website (*BIP/ZIP*, n.d.), the Forum Urbano website (*Forum Urbano*, n.d.) that includes a mapping of the realised projects up to the 2020 edition, as well as the annual calls for applications, called Ciclo e Regras (*BIP/ZIP*, n.d.). The dataset used in this study was informed and validated by the Municipality of Lisbon, through a research collaboration of the authors.

In the dataset, each project is treated as a data point with all the relevant attributes associated to it. The first attributes relate to the project's identity and are specifically the name of the project and an ID number, as well as the year of the project's application. These features are significant for the connections among the different datasets created for the analysis process. Furthermore, the attributes of the location of each project have been selected as a significant feature both for this preliminary step of the evaluation and most importantly for the future visualisation of the results in a geolocated maps. At this stage, the location of each project is described through the number of the BIP/ZIP priority area or areas (1-67) where it is implemented.

The first key indicator examines the generation of partnerships for the projects, containing raw data on the types of partners-institutions. The recorded types comprise 18 categories, such as neighbourhood, cultural, religious, sports associations, non-governmental organisations and non-profit cooperatives. Additionally, the theme of each project refers to the scope of the initiative and is defined during the application phase with the form of a selection among five predefined categories. The categories are consistent throughout the ten years, with minor updates on their titles and objectives and refer to projects that target to Improve Neighbourhoods Life; Skills and Entrepreneurship; Community Space; Prevention and Inclusion; and Promoting Community Dynamization and Citizenship. Initiatives can be given funds up to €50,000 each, but it is advisable that applications demonstrate complementary funding by external resources, proving the ability to sustain the operation of the project after the first year of municipal grant. Hence, the indicator of funding is described by two values, one provided by BIP/ZIP and one by external sources. Moreover, an important parameter for understanding the projects impact in the priority neighbourhoods refers to the target group attribute to which the project addresses, and which can be children; youth; elderly; families; vulnerable groups; and adults (Figure 1).

2.2 Activity Mapping

The final parameter set of selected attributes refers to the activities executed throughout each project, based on the application phase. These attributes are the most difficult to record in an organised way, as the available information is based on free descriptive text. The first attempt to codify the activities has been approached with semantic techniques, such as Natural Language Processing models and word clouds, however the results were not as useful as expected. A second attempt includes the method of cycles of qualitative coding (Saldana, 2021). The initial free text is codified in keywords and further assigned to categories by the authors. The categorisation is based on an index of parameters that relate to social and spatial impact for sustainability (Colantonio et al., 2009; Global utmaning, 2017; Maio et al., 2020; Van Herck et al., 2019) (Figure 2).

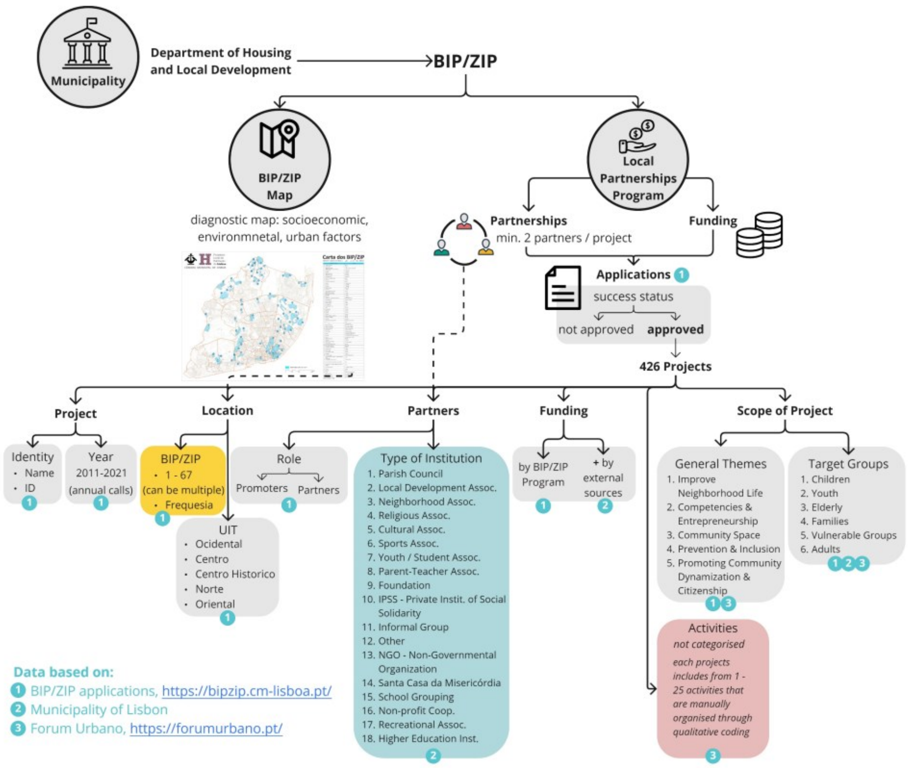

Figure 1. Data collection diagram.

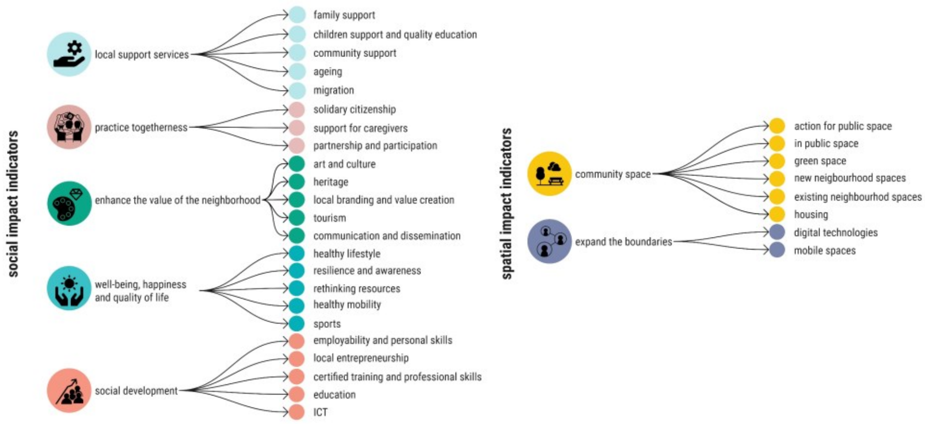

Figure 2. Activity mapping: social and spatial impact indicators.

2.3 Data Analysis

The methodology followed aims to assist the evaluation of the projects in relation to the BIP/ZIP program in its entirety by identifying correlations between the partners, the locations of the projects, the activities mapped and the general attributes of each project, such as the internal and external funding, the themes and target groups and the year of implementation. The total number of selected attributes for each project in conjunction with the number of projects adds up to a very large number of data points and a complex dataset that is not manageable without automated and machine learning driven data analysis methods. Using unsupervised machine learning clustering methods, the large dimensionality of projects' dataset can be reduced and visualized to an understandable parameter space, thus allowing to abstract information and identify correlations between the different project attributes. The clustering methodology focuses on the 3 main aspects of the projects' data: locations; partners; and mapped activities.

The development of custom visualisation dashboards, based on an existing interactive dashboard for KPI-based impact assessment developed by the authors (Duering et al., 2022) allows for further exploration of the correlation of the projects' attributes, thus reducing even more the complexity and allowing for interactive interrogation of the data based on specific evaluation questions. These dashboards are informed not only by the raw project data, but also by the mapped activities per project as well as the output data of the clustering models. The 4 visualization dashboards are used interchangeably to help understand the complex relationships between the 3 main clusters of locations, partners and activities and the general attributes of the projects.

2.4 Data Clustering

The large dimensionality of the projects' data is reduced using a clustering methodology. As each project can be associated with multiple BIP/ZIP areas, partners and activities, associating and visualizing these attributes along with general attributes of the projects in one dashboard would lead to an overly complex analysis. The attributes for each clustering model are therefore reduced to relevant attributes for the 3 specific aspects of the program and then collectively used with general data of projects to identify patterns. Different unsupervised machine learning clustering models have been attempted for each analysis step with varying results, such as k-means clustering, principal component analysis (PCA), self-organizing map (SOM) and T-distributed Stochastic Neighbourhood Embedding (tSNE). Python scripts, open-source models, Google Colab collaborative code notebooks and the open-source data mining and visualization software Orange have been used to explore and visualize the clustering models. The results are produced with a k-means clustering model using an elbow variable clusters method.

For the locations clustering, the projects' BIP/ZIP areas are mapped in a table, producing a 67 dimensions vector per project. The number of areas per project is correlated with the different locations' clusters to identify correlations between project types and the number of locations involved. The locations dashboard contains locations per project, clusters and number of locations as well as activities and partners clusters.

Mapping the individual partners of each project resulted in an index that exceeds the 4700 entries. To reduce this complexity for the partners' clustering model, the partners are instead mapped in 18 institution types, as specified by the Municipality of Lisbon. Therefore, the mapping of the partner types per project is formative of an 18-dimensional vector per project which is consequently used to produce 4 clusters of partner types with common characteristics. Again, the number of partners (which can be more than 1 per type) is also used to inform the partners clustering and dashboard along with the locations and activities clusters.

Finally, the activities clustering model is based on the data produced by the manual activity mapping. The data is structured in a table of number of instances of activity types per each project. The 34 categories of activities resulted in a 34-dimensional vector per project which is used for the activities clustering. As per the other clustering methods, the number of activities as well as the locations and partners clusters are also included in the activities' dashboard. The results from all the clustering models are also directly integrated in the general dashboard, which creates an overview of all clusters along with the general data of the projects. Clustering models, such as k-means, tSNE and SOM are also inclused as interactive components of all dashboards, allowing further clustering of sub-selections of the data of each of the 4 dashboards (Figure 3).

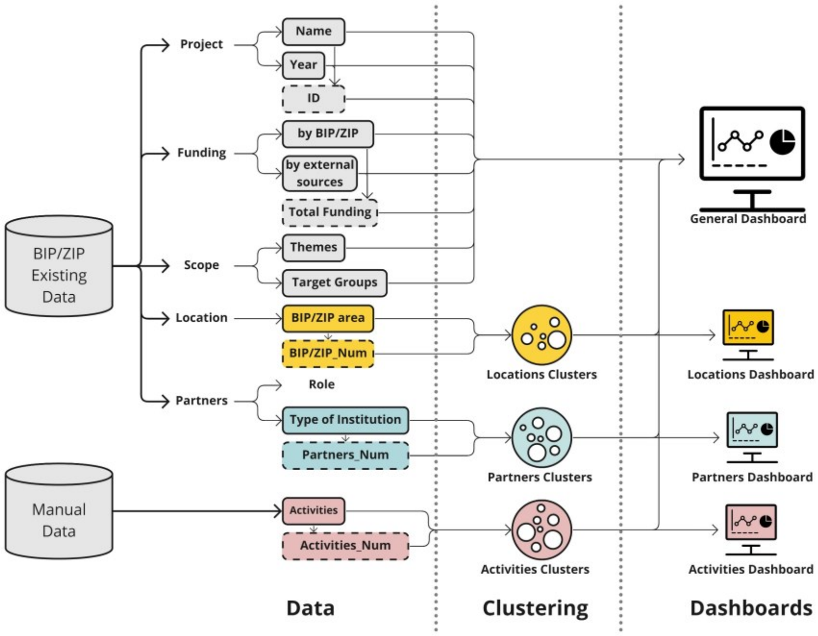

Figure 3. Methodology Diagram.

2.5 Interactive Dashboards

A key component of the data analysis is the interactive dashboard for KPI-based evaluation. Further to the ability of visualizing a large dataset, the dashboard components allow deep data exploration, using interactive graphs with associated sub-selections of data, KPIs comparison, as well as correlation matrices, selectable scatter plots and clustering models, all of which are used to understand and identify patterns in the large dataset of the BIP/ZIP program. In total, 4 different dashboards are produced using raw data, processed data and results of the clustering models. The locations, partners and activities dashboards are used individually as well as in conjunction with the general dashboard, informed by all data and clusters. The dashboards are developed in HTML using Python and Bokeh, an interactive visualization library. For the developed dashboards a local Python server is also developed to run the ML models, produce the KPIs and graphs and serve all data and visualizations to the HTML pages of the dashboards.

Using the interactive and selectable parallel coordinates tab in all dashboards, the users can select different attributes of projects, such as a specific activities cluster, the range of funding or part of the program’s timeline and interactively see updated graphs and KPI metrics showing the comparison of the sub-selection with the general distribution of KPIs for all projects. This interactive exploration of the data is instrumental for understanding the relatively complex dataset. At the same time, the dimensionality reduction tab can be used to cluster the data with different clustering models and visualize the attributes in scatter plots. Sub-selecting data points directly on the scatter plots and then investigating their KPIs distribution is valuable to understand more specific relationships between the projects' elements. Finally, the correlation matrix can be used to quickly visualize the correlations between attributes of the project both for all data points as well as very importantly for any sub-selection on the scatter plot or parallel coordinates plot. This fully connected interactivity among all graphs is invaluable for drawing conclusions from the data (Figures 4,5).

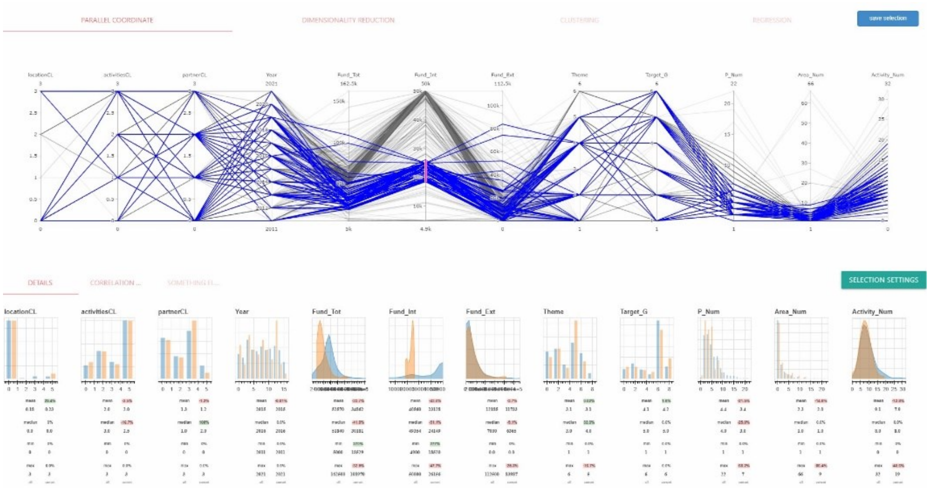

Figure 4. Interactive dashboard: parallel coordinates tab.

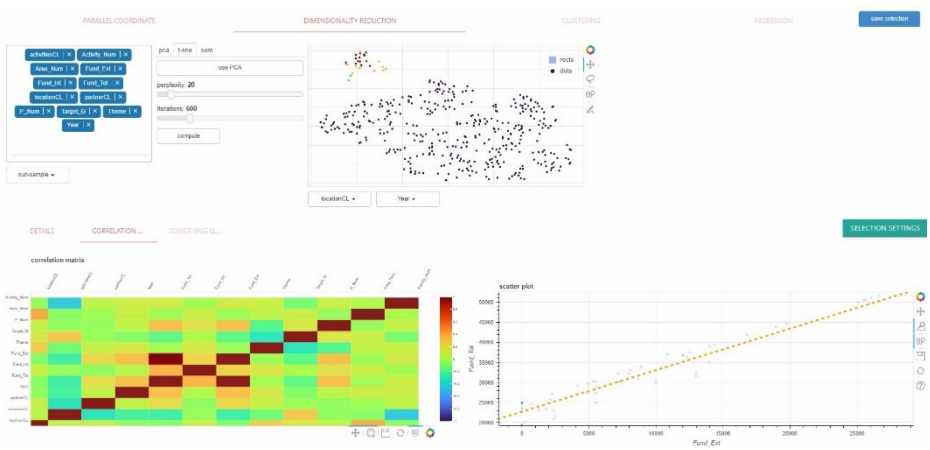

Figure 5. Interactive dashboard: dimensionality reduction tab.

3 Results & Discussion

Through the tested methodology, we have been able to collect, map, analyse and visualize the large dataset of the BIP/ZIP program and explore correlations within complex relationships of projects, partners, areas, activities, and general attributes of more than 400 projects. An exhaustive presentation of these correlations is not in the scope of this paper, as its focus is on the methodology of evaluating participator. It is important to note that for the preparation of the dataset a large amount of data points have been analysed, cleaned and manually processed. Therefore, for a generalization of the proposed methodology a robust data integration methodology would be necessary. Nevertheless, we consider that the developed methods along with their modular character can be a significant first step for data-driven evaluation of participatory models.

3.1 Locations, Partners & Activities Clusters

The ability to identify clusters with discrete characteristics and then correlate them with specific attributes is key in understanding relationships among the elements of the projects, especially because a similar exercise with all data incorporated proved very inefficient and incomprehensible. In the locations clusters results, clear patterns of project types emerged, with 1 major cluster of low number of locations (1-2 areas), and 3 clusters of large numbers of areas (1-10, 4-17, 27-67 areas). This allowed for clear correlation between smaller and larger spatial scale projects. The clustering of partner types yielded even more unpredicted patterns of partner constellations. For example, the inclusion or not of Local Development Associations as partners in projects produced 2 distinct clusters. This result correlates with how local groups such as cultural, sports and youth associations form partnerships of larger number of partners in comparison to projects with no or little local groups. These patterns can also be used to reflect on budgetary, thematic or other requirements of projects and how BIP/ZIP enables them. Furthermore, the activities clustering delved deeper in the specifics of the projects' foci, with resulting clusters of specific activities groups, translated as “arts & culture”, “support and awareness”, “identity” and “skills and entrepreneurship” clusters.

3.2 Visualization dashboards

The resulting 4 visualization dashboards accumulate the dataset of the study and their interactive components allow for a very instructive exploration of the BIP/ZIP program. Although this visualization is only a first step towards a data-driven evaluation of the program, it clearly shows how powerful the integration of interactive visualization components with unsupervised ML models can be. The interactive dashboards can produce inexhaustible associations and visualizations of data, allowing the users to examine every aspect of the projects, from constellations of partners to patterns of activities, funding and themes, providing useful feedback for the future of the project. Nevertheless, a detailed and critical evaluation of these results is needed to optimize and further develop the interactive dashboards, KPIs and ML models to better match the program's specific needs.

4 Conclusion

The study contributes to the conceptualisation of participatory models, such as the participatory budget of BIP/ZIP, as multi-dimensional systems that are defined by complex relationships of attributes. The methodology has been significantly shaped by the available data of the BIP/ZIP Program, which means that a potential transferability would require a recontextualization, or a selection of data that are meaningful for the new contexts. Despite being case-specific, this study aims to move beyond certain research outputs, and focus on the methodological contribution.

First, it suggests two ways for the simplification of complex data that would possibly contain correlations difficult to translate. These are the use of codified categorisation rather than individual free entries, such as in the case of the partners attribute that is introduced as types of institutions, and the use of clustering methods that provide unsupervised categorisation. Moverover, it examines the integration of qualitatively mapped attributes in addition to raw existing data. This step offers plenty of potential for the further development of this case and the transferability to other contexts, as it demonstrates that any additional qualitative attributes can be inserted in the dataset, as long as their integration is consistent to the data preparation process of the methodology. For example, a valuable parameter to be considered, would be the participatory assessment of the impact of each project, by the partners or residents. Also, employing the tool of a dashboard and the multiple analysis and visualisation techniques enables the correlation of selected attributes and hence the response to targeted questions. Most importantly, the interactive character of the dashboard allows for quick visualisations and experimentation.

Finally, through this data- and KPI-driven methodology for assessment of participatory programs, the study aims to denote the significance in looking at participatory budgets not merely as policies, but as formative of urban process and hence evaluated as such, using tools of urban analytics. This, as far as the authors are concerned, is a significant research innovation in the field of evaluation of participatory programs.

Acknowledgements. The work presented in has been carried out within the RE-DWELL Innovative Training Network, funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 956082.

References

An integrated toolbox for deprived neighbourhoods | URBACT. (2021). https://urbact.eu/integrated-toolbox-deprived-neighbourhoods

Annex of the 2030 Agenda. (2017). Annex: Global Indicator Framework for the Sustainable Development Goals and Targets of the 2030 Agenda for Sustainable Development. *Work of the Statistical Commission Pertaining to the 2030 Agenda for Sustainable Development*, 1–21.

Arnstein, S. R. (1969). A Ladder of Citizen Participation. *Journal of the American Planning Association*, 35(4), 216–224.

Bahreldin, I. Z. (2013). An Approach to Community Participation Evaluation Framework. *Buildings and Roads Research*, V(January 2013), 15–32.

BIP/ZIP. (n.d.). Retrieved July 19, 2022, from https://bipzip.cm-lisboa.pt/

CML. (2010). CARTA DOS BIP / ZIP Bairros e Zonas de Intervenção.

CML, & DMHDL. (2017). *Good Practice Summary – Lisbon Local Development Strategy for Neighborhoods or Areas of Priority Itervention (BIP/ZIP): an integrated toolbox.*

Colantonio, A., Dison, T., Couch, C., Fraser, C., & Percy, S. (2009). Measuring Socially Sustainable Urban Regeneration in Europe. *Creating Sustainable Environments*.

Duering, S., Fink, T., Chronis, A., & Koenig, R. (2022). *ENVIRONMENTAL PERFORMANCE ASSESSMENT-The Optimisation of High-Rises in Vienna*.

Falanga, R. (2019). Measuring citizen participation in urban regeneration: a reflection on the construction of the participation index for the Bip/Zip programme in Lisbon. *Urban Development Issues*, 62(1), 47–60.

Falanga, R. (2020). Formulating the success of citizen participation in urban regeneration: Insights and perplexities from Lisbon. *Urban Research and Practice*, 13(5), 477–499.

Forum Urbano. (n.d.). Retrieved July 19, 2022, from https://forumurbano.pt/

Fung, A. (2006). Varieties of participation in complex governance. P*ublic Administration Review* (Vol. 66, Issue SUPPL. 1, pp. 66–75). Smith and Wales.

Global utmaning. (2017). Local Implementation of the SDGs & The New Urbanism Agenda Towards A Swedish National Urban Policy. *Global Utmaning Association*.

Maio, S. de, Kuhn, S., Esteve, J. F., & Prokop, G. (2020). *Indicators for European cities to assess and monitor the UN Sustainable Development Goals (SDGs).*1–222.

Saldana, J. (2021). *The Coding Manual for Qualitative Researchers* (4th ed.). SAGE.

United Nations (Habitat III). (2017). *NEW URBAN AGENDA*. www.habitat3.org

Van Herck, B., Vanoeteren, V., & Janssen, K. (2019). *Local Indicators for the 2030 Agenda (Sustainable Development Goals).*