



# AI Urban Voids: A Data-Driven Approach to Urban Activation

Amal Algamdey<sup>(✉)</sup>, Aleksander Mastalski, Angelos Chronis, Amar Gurung, Felipe Romero Vargas<sup>(✉)</sup>, German Bodenbender, and Lea Khairallah

IAAC-Institute for Advanced Architecture of Catalonia, Barcelona, Spain  
{amal.algamdey, felipe.romero}@students.iaac.net,  
algamdeyamal@gmail.com, romerovargas.f@gmail.com

**Abstract.** With the development of digital technologies, big urban data is now readily available online. This opens the opportunity to utilize new data and create new relationships within multiple urban features for cities. Moreover, new computational design techniques open a new portal for architects and designers to reinterpret this urban data and provide much better-informed design decisions. The “AI Urban Voids” project is defined as a data-driven approach to analyze and predict the strategic location for urban uses in the addition of amenities within the city. The location of these urban amenities is evaluated based on predictions and scores followed by a series of urban analyses and simulations using K-Means clustering. Furthermore, these results are then visualized in a web-based platform; likewise, the aim is to create a tool that will work on a feedback loop system that constantly updates the information. This paper explains the use of different datasets from Five cities including Melbourne, Sydney, Berlin, Warsaw, and Sao Paulo. Python, Osmx libraries and K-means clustering open the way to manipulate large data sets by introducing a collection of computational processes that can override traditional urban analysis.

**Keywords:** Computational urban design · Artificial intelligence · Machine learning · Urban data · KMeans clustering · Data visualization

## 1 Introduction

By 2050, 68% of the world’s population will be living in cities. In order to adapt to this rapid urbanization growth while making cities more sustainable, livable, and equitable, designers must utilize qualitative and quantitative tools to make better-informed decisions about future cities [1]. In addition, big urban data is now readily available online, allowing the opportunity to utilize this information to generate new urban analyses between various features within the urban fabric [2]. A new digital layer can be added toward urban complexity through the novel perspective of data accumulation. However, despite these advancements, urban analysis and planning processes still follow the most static models that do not fit the development of today’s cities (Jordan and Mitchell 2015; Al-Garadi et al. 2020).

© The Author(s) 2023

P. F. Yuan et al. (eds.), *Hybrid Intelligence*, Computational Design and Robotic Fabrication,  
[https://doi.org/10.1007/978-981-19-8637-6\\_26](https://doi.org/10.1007/978-981-19-8637-6_26)

Moreover, the available urban data is vast and beyond human capability to handle; consequently, building new tools to understand and manage such substantial information is needed. Recently, cities are increasingly incorporating Machine Learning (ML) applications, primarily to meet economic and sustainability goals, etc. (Li et al. 2020; Choung and Kim 2019; Liu et al. 2017). Therefore, several approaches are presented to model the dynamics of urban drivers as a function of the different features of the urban form. The latest research results show that ML methods have significantly exceeded the conventional prescriptive modelling methods of urban indicators to evolve into an essential tool for urban planning decision-making (Ma et al. 2020; Hecht et al. 2013). To meet the challenges of the current urban complexity of emerging big data, the modelling of urban indicators increasingly exploits intelligent automatic methods using ML algorithms which can override traditional methods.

This paper aims to investigate the role of ML in analyzing and predicting the accessibility and proximity to urban features of the city of Melbourne. In addition, it demonstrates the advantages of building new tools for prediction-based planning to accurately define the interconnections within cities. With the availability of massive datasets, this study integrates a Machine Learning process using (K-Means clustering) to improve capabilities for urban planners to understand cities better. Also, It demonstrates how technological advances in data gathering and processing can depict more aspects of urban complexity and dynamic urban nature. However, due to the complexity of computation, these technological advances are still not accessible and lack connectivity to real-world scenarios and design practices [3]. For this reason, there is a need for urban planners and urban designers to operate and integrate computational design processes. Moreover, it is essential to develop new tools to have a deeper understanding of cities and take advantage of computational design and machine learning to improve design decisions for better cities.

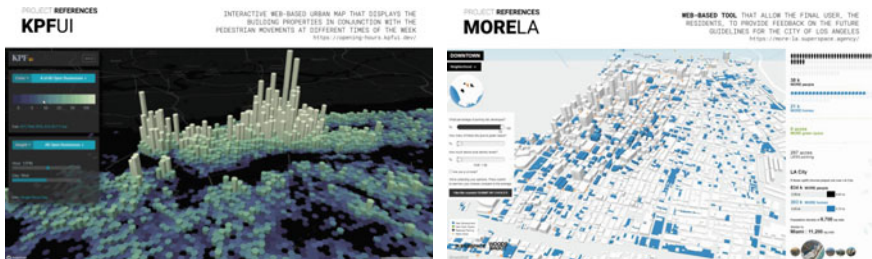
## 2 Background

Urban indicators analysis using ML has been growing with great success to address urban planning challenges in recent years, but some are still in a very early stage of development or have been developed as part of research without implementation on real-case scenarios [4]. This method previously addressed some urban issues, including smart cities, mobility, climate, density, and energy. However, the evolution of that research indicates a promising future and outcome for ML application in this area [2]. In contemporary research, the ML algorithms applied are deep learning, artificial neural networks (ANN), support vector machines, neuro-fuzzy, and decision trees. These methods are usually used to classify information and create predictive models [5]. The reviewed papers highlight applications of ML algorithms to understand trade-offs between the city, its indicators, and urban planning. Also, to support the discussion on the role of these methods, the implementation of ML strategies on a city scale, and the possibility of integrating ML with other technologies.

The design of digital, smart, and connected cities is at the core of modern urban planning (Ma et al. 2020; Middel et al. 2019; He et al. 2018). Cities must demonstrate digital transformation and integration initiatives by creating intelligent and smart cities

to become more influential. The Internet of Things is at the centre of the smart city and enables the enrichment of collected urban data into Urban Big Data. These big data capture information on all urban activities and are also the ideal source for ML algorithms. Indeed, from this big data, ML will allow the creation of AIs, allowing to make the city more intelligent (Jordan and Mitchell 2015). But also, and most importantly, the processing of big data by the rising power of ML algorithms will allow urban planners to predict the city's evolutionary trends and regulate them by orienting the city's shape towards the most sustainable, intelligent, digital, and connected form possible. Thus, urban data enriches the ML to create AI for IoT and intelligent urban planning, making the city more and more intelligent, digital, and connected, which helps enhance big urban data. Therefore, the power of ML algorithms, IoT applications, the engine of AI, and intelligent urban planning are crucial for addressing the challenges of the smart city (Al-Garadi et al. 2020; Jordan and Mitchell 2015).

Therefore, to develop our project AI Urban Voids, some projects were studied to understand existing computational design processes applied to urban environments. The Hive, for instance, is a web-based urban map that displays the building properties in conjunction with pedestrian movements at different times of the week (<https://comotionla.com/more-la-parking-to-places>). In addition, Hive proposes a 3D urban data visualization tool that helps find urban relationships between business opening hours by integrating civic open data sets. In this case, this project was the primary reference as an interface for urban indicators metrics and for understanding the potential for urban tools to analyze urban activation and economy (Fig. 1).



**Fig. 1.** Simulation Platforms: Hive, and more LA.

Secondly, in another study, RE{CODE} is an urban simulation platform developed at IAAC (Institute of Advanced Architecture of Catalonia) that modifies & understands the existing urban fabric and proposes methods to minimize the social and economic imbalances. The tools suggest a new rearranged urban pattern based on performance relationships and urban dynamics, this tool supports the urban analysis for large scale urban centers by identifying urban discrepancies and future opportunities.

Finally, MoreLA is a project developed by Superspace-Woods Bagot in the mobility framework (<https://comotionla.com/more-la-parking-to-places>). This project was developed as an interactive survey and a web-based tool created to allow the final user, the residents, to provide feedback on the future guidelines for the city of Los Angeles. MoreLA introduces the concept of participatory design by integrating users' opinions

for large urban centers; this opens the opportunity to understand people’s choices in early urban planning phases and identify urban solutions for local communities.

The overall analyses of these three projects conclude that computational design processes & machine learning applications have recently grown in urban studies to solve new challenges. Furthermore, data availability and accessibility of computational tools helped enhance this growth. However, some of these urban dynamics analyses approaches lack processes to identify the potential location for future urban amenities based on proximity and score criteria; this is to draw a relationship between urban nodes.

There is a necessity to develop computational urban design tools to manage and understand extensive data sets that cities are generating from time to time; this could lead to a more dynamic process between designers, users and urban communities [6].

### 3 Methodology

The project is conceived as a data-driven approach to analyzing and unveiling the urban fabric’s hidden opportunities by introducing a computational design process that calculates distances and walkability using urban points of interest and existing pedestrian networks (OSM Data & Python NetworkX). This process allows the evaluation of the results based on clustering performance, comparing distances/areas, and introducing and analyzing new urban amenities.

This research starts with gathering and cleaning from Open Street Maps (OSM) data (Fig. 2). OSM Data plays an important role based on its availability aspects; however, it also opens the questions of how reliable open-source data sets are. Therefore, the first step is collecting and cleaning data from Open Street Maps OSM; this process opens the possibility of bringing a structured OSM dataset for further computation.

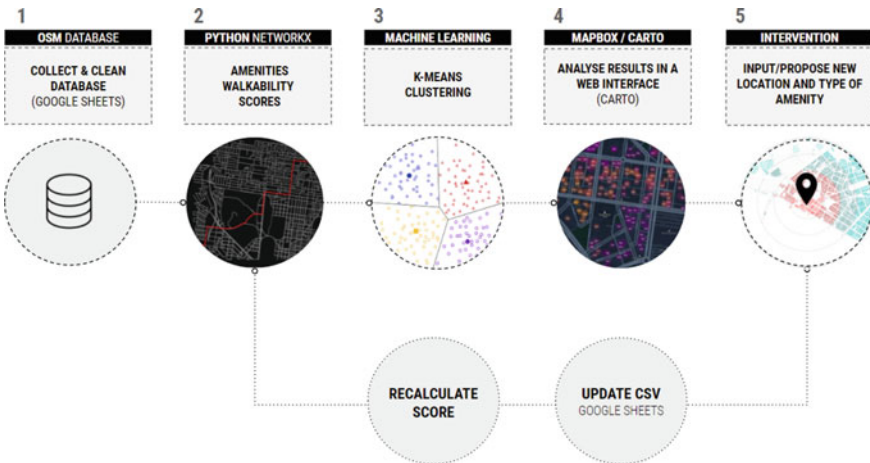


Fig. 2. Project methodology diagram.

For the intention of this project, the data set was narrowed to three key features that include:

- Points of Interest (Urban Nodes)
- Pedestrian Network (Urban)
- Locations with addresses (Urban Nodes)

Points of interest were selected based on the daily usage and importance of the services provided to cities. Some of the critical points of interest include education facilities (primary and higher), healthcare and hospitals, nightlife and entertainment facilities (Fig. 2). The second step began with Python NetworkX calculating distances between amenities and ranking them based on score criteria. NetworkX is a python library used to create, manipulate, and study the structure, dynamics, and functions of complex networks; in this case, cities.

The third step was to cluster all scores gathered from Python and NetworkX using K-means. K-means is an unsupervised machine learning process that helps identify data clusters in large datasets. The identification of these clusters was later visualized within a web interface. As a final stage, the project takes advantage of Carto as a data visualization by linking the result of the previous computation process; furthermore, this tool enables user interaction and data manipulation by a series of filters that display results instantly. Finally, for later stages, Urban Voids open the opportunity to introduce a new input where the user can introduce new urban features, whether a location or amenity type, that will feed into the database to recalculate the scores resulting in a new score and clustering based on the new input.

### 3.1 Dataset Creation (Extracting, Compiling and Clustering)

For the reliability and efficiency of the process, Python and OSMX libraries are used to extract the data from Open Street Maps to provide the script with the city-CRS, which extracts three main data frames: points of interest, pedestrian network, and address points. After this process, the data used for K-Means clustering were cleaned and adjusted (Fig. 3). The objective of the data cleaning was defined in three steps:

1. Build a Pandana Network—“Pandana is a Python library for network analysis that uses contraction hierarchies to calculate super-fast travel accessibility metrics and shortest paths.” Pandana v0.6.1 (2021).
2. Generate tags for different Points of interest.
3. Clean and provide a clear structure for different addresses in a city.

The purpose of these three objectives is to prepare the definition of processing data:

- To calculate the shortest distance from every point to an “n” number of closest points of interest.
- To evaluate a based scoring system on the average walking time of 1.2 m/s as a normal pedestrian pace.
- To introduce an optimized computation process to calculate distances between multiple points of interest tags. In this case, the calculation achieved for a tag: ‘bus stop’, for the address data frame of 320 000 rows took less than 3 s.

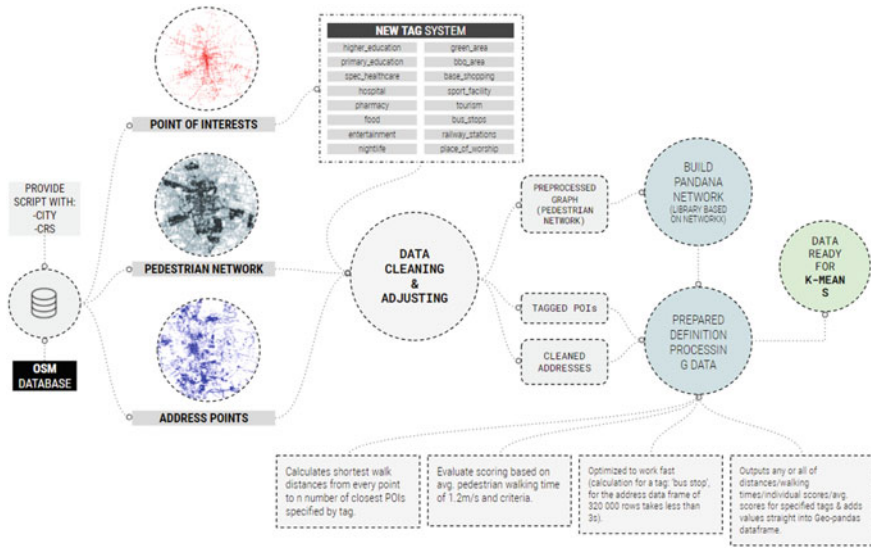


Fig. 3. Workflow of dataset creation.

- To achieve an output of any or all distances/walking times/individual scores/avg. & scores for specified tags and add values straight into GeoPandas data frame.

Building scoring and walkability analysis were developed using a scoring system as a result of the walking time from one point to another. Walking times vary between 0 min (score of 100 pts) and 70 min (score of 0 pts); using this method helped the project to identify & classify amenities for multiple addresses. As a result of this analysis, the script provided the number of X amenities closest to the subject address, based on multiple criteria, as shown in Fig. 4. Traditional Urban Design analysis strategies include using a Ped Shed Analysis that understands the proximity of urban nodes within a radius of 400, 800 and 1,200 m (5, 10 and 15 min walking distances). However, this technique can be considered a high-level analysis tool that most of the time does not consider the existing urban network and does not specify the relationship between multiple uses to the specific node.

### 3.2 Machine Learning (K-means Clustering)

“Plot pairwise relationships in a dataset. By default, this function will create a grid of Axes such that each numeric variable in data will be shared across the y-axis across a single row and the x-axis across a single column. The diagonal plots are treated differently: a univariate distribution plot is drawn to show the marginal distribution of the data in each column.” Seaborn (2021).

For this project, a pair plot was used to understand the pairwise bivariate distribution of multiple datasets and their relationships within the subject city (Fig. 5).

In addition, the Pearson correlation coefficient was used to understand which attributes are linearly related to the predicted set (Fig. 6). Furthermore, a biplot overlays

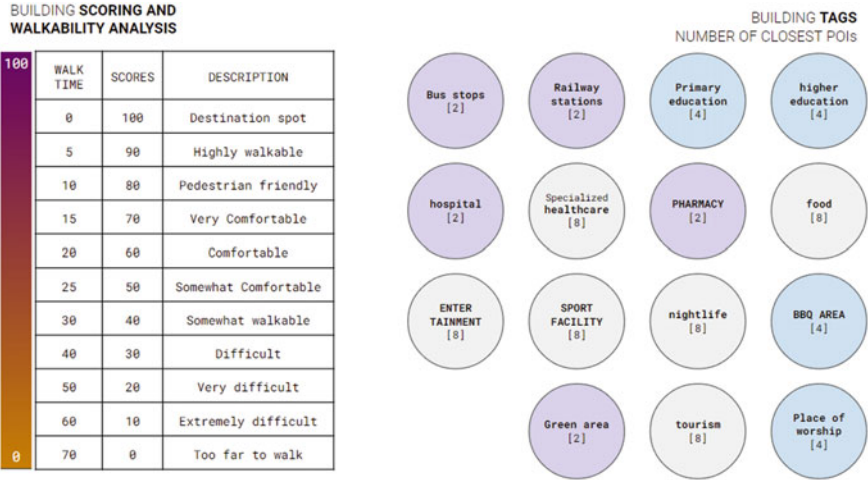


Fig. 4. Building scoring and walkability analysis.

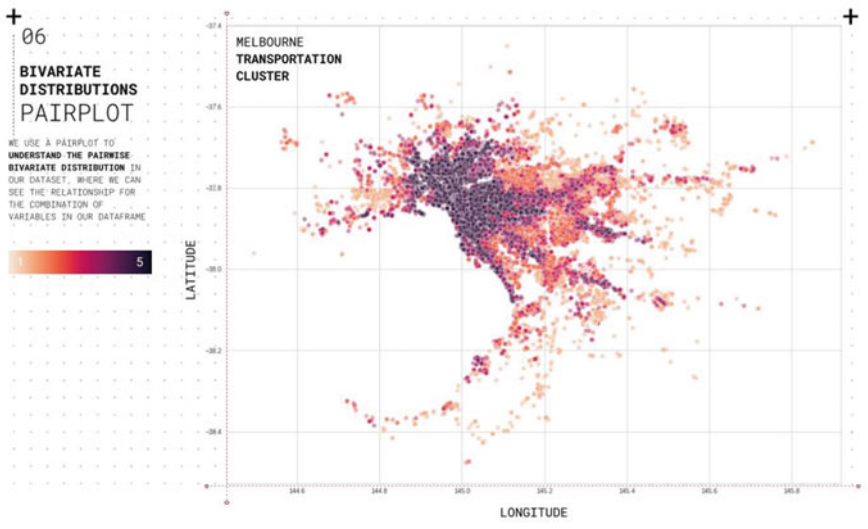


Fig. 5. Bivariate Distributions–Pair plot for Melbourne (Transportation Cluster).

both a score plot and a loading plot onto a single graph to visualize high-dimensional data onto a two-dimensional graph. Finally, the elbow method was used to understand the ideal number of clusters that should be used for the clustering based on the shape and features of the data. Then, the final clustering was plotted on a two-dimensional plot for each category based on overall performance.

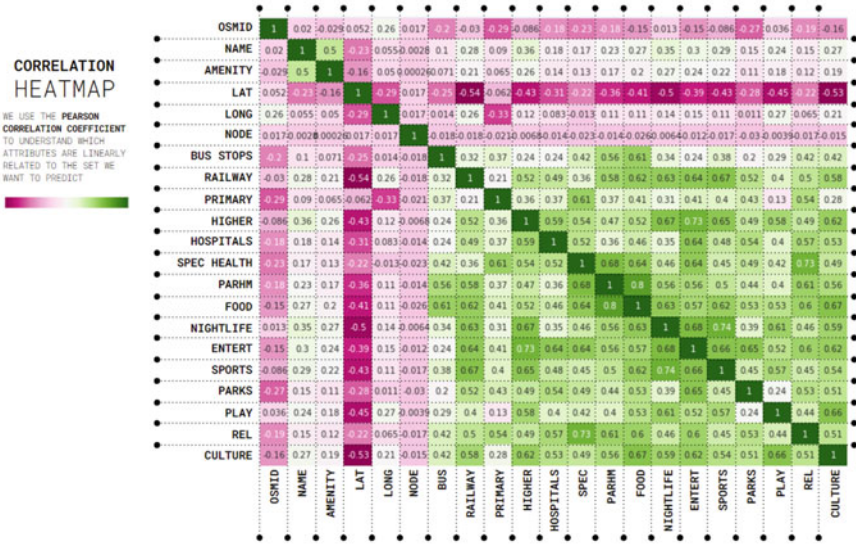


Fig. 6. Correlation Heatmap.

## 4 Case Study

Melbourne city was used as a case study to test the different clusters and their relationships to examine the described methodology. For the purpose of this project, Carto was used to create a web-based application to analyze and unveil the correlations of urban features and their opportunities for multiple cities. In this case, opportunities are measured based on the scoring system that provides a classification process based on the accessibility and direct relationship within urban areas. Five cities with different urban conditions were used as case studies for the analysis and comparison: Melbourne, Sydney, Berlin, Warsaw, and Sao Paulo. The data visualization tool allows the user to filter & compare the information based on the project criteria and display the number of clusters based on the multiple uses that are part of the data set (Fig. 7).

## 5 Results

As a result, this study can be divided into three categories, including statistical analysis, visualization, and web application. As shown in Fig. 8, this tool allows users to compare five different cities and visualize K-means values, this step demonstrates the ability of the tool to integrate different urban data sets and provide the results. However, it is essential to highlight that the results are driven by the quality & accessibility of the data. Figure 9, the web application, shows the AI Urban Voids as an interactive app. The overall idea is to provide a platform where the user can manipulate, filter, and understand the geographical relationship and connections within the city by using K-means clustering. This web interface can help policymakers, architects, and urban designers to evaluate and understand the city's behaviour and provide a better design response. It is essential



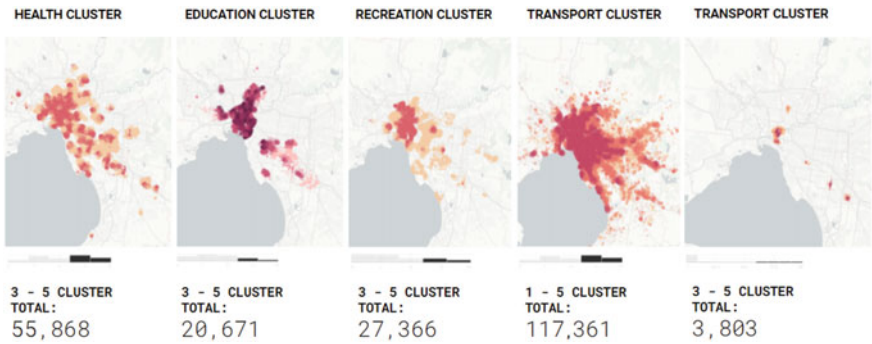


Fig. 7. Data visualization and analysis.

to highlight that the results can vary based on the quality and the data input; this is what is essential to improving and maintaining open data sources. Finally, this web interface can be used in strategic urban Planning & Urban Design phases by private or government entities to understand current urban conditions better and provide strategic projects for multiple urban contexts.

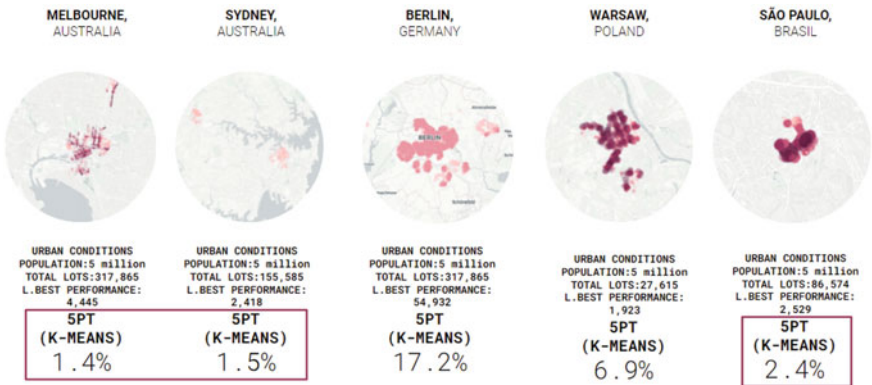


Fig. 8. K-means results from comparison and visualization.

## 6 Future Research Direction

With the evolution of today’s cities and with Urban big data, there is a need for new efficient models more than the traditional static models. Moreover, more efficient ML techniques allow for more helpful data processing. Thus, intelligent urban planning could benefit from creating a smart city. The future work coming out of this research applies to the K-means clustering method of proximity to healthcare. For instance, the correlation approach considering road networks and trips from each park can correlate parks’ mutual attractiveness rather than generic distance. Additionally, developments might come from creating 3D urban forms based on the results of clusters.

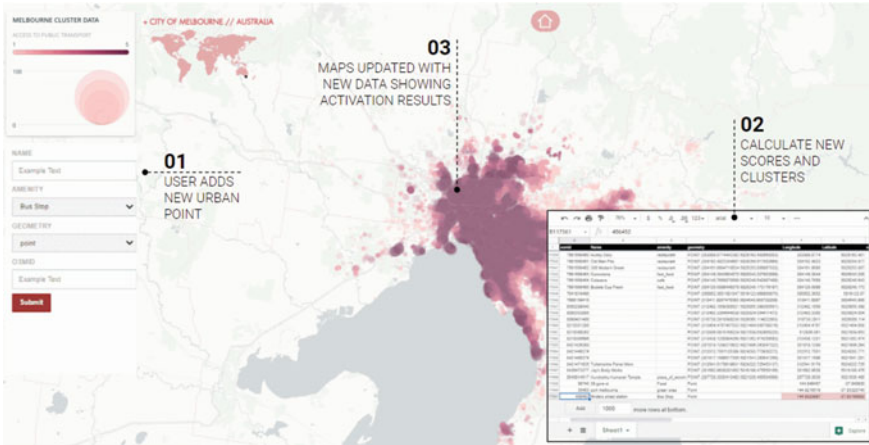


Fig. 9. Web application AI urban voids. (<https://city-lab.wixsite.com/urbanvoids>).

## 7 Conclusion

This research has presented very positive results in breaking down the workflow between urban indicators analysis and machine learning. Moreover, this study can be considered a workflow for further exploration by linking open data sources to a web interface to unveil hidden urban relationships that can improve urban accessibility for certain uses.

Open data sources and computational skills are the major limitations to developing urban tools that can integrate large data sets. Nevertheless, despite the disadvantages above, understanding urban centres and open data sources provide a benefit that opens the possibility to integrate Computational Design & Artificial Intelligence processes.

Moreover, this workflow demonstrates the different limitations tools can have when handling large data sets. Python and Osmx libraries open the way to manipulate large data sets that benefit multiple urban communities. There is an opportunity to complement, improve and create new/existing large open-source datasets that can test and inform design processes. Google places + OSM datasets are often driven by commercial applications, neglecting non-marketable areas and spaces still important for the city.

The process allows users a new series of opportunities, but the question about real-case applications in the urban area remains, who could benefit from this, and how can we make better cities with it? The deployment of the application provides an understanding of a new possibility of making AI accessible for urban designers, planners, developers, and policymakers. However, sufficient back-end work has to be done to integrate the urban data, machine learning processes and its definition in data visualization platforms. Since the data used now is entirely open source and the workflow completely established, the next step is to make this available for every city and minimize the computation to be accessible. Finally, this study has demonstrated the importance of using computational design methods in urban matters; this allows the creation of new tools that can improve urban analysis phases that will result in better design decisions & urban policies that can

improve urban communities. The resulting data from this project can be used as a base to identify & allocate future urban density by providing an interconnected city; moreover, it can be used as a planning tool to re-interpret land zoning by local governments.

## References

1. Moustaka V, Vakali A, Anthopoulos LG (2018) A systematic review for smart city data analytics, vol 51, no 5
2. Ulusar UD, Ozcan DG, Al-Turjman F (2019) Open source tools for machine learning with big data in smart cities. *Smart cities performability. Cogn Secur* 153–168
3. Haldorai A, Ramu A, Murugan S (2019) Machine learning and big data for smart generation. *Comput Commun Syst Urban Dev* 185–203
4. Nosratabadi S, Mosavi A, Keivani R, Ardabili S, Aram F (2019) State of the art survey of deep learning and machine learning models for smart cities and urban sustainability, no Aug 2019
5. Nikolaeva A, Adey P, Cresswell T, Lee JY, Nóvoa A, Temenos C (2019) Commoning mobility: towards a new politics of mobility transitions. *Trans Inst Br Geogr* 44(2):346–360
6. Gao X, Lee GM (2019) Computers & Industrial Engineering Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning. *Comput Ind Eng* 128:60–69. (no Dec 2018)
7. Psyllidis A, Bozzon A, Bocconi S, Bolivar CT (2015) A Platform for urban analytics and semantic data integration in city planning, 21–36
8. Derix C (2019) Paradigm reversal–connectionist technologies for linear environments. In: *Research culture in architecture*, vol 21
9. Woods B (nd) Woods Bagot-sponsored study More LA anticipates major transportation changes for Los Angeles. <https://www.woodsbagot.com/global-studio/news/ideas/woods-bagot-sponsored-study-more-la-anticipates-major-transportation-changes-for-los-angeles/>. Accessed 21 Mar 2022. (re of modern genomics. Blackwell, London)
10. Podrasa, D., Zeile, P. and Nepl, M., 2021, September. Machine Learning for Land Use Scenarios and Urban Design. In *CITIES 20.50–Creating Habitats for the 3rd Millennium: Smart–Sustainable–Climate Neutral. Proceedings of REAL CORP 2021, 26th International Conference on Urban Development, Regional Planning and Information Society* (pp. 489–498). CORP–Competence Center of Urban and Regional Planning.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

