Introducing a Financial Twin App for Exploration and Comparison of Municipal Financial Patterns

Pankaj Kumar, Sibylle Hestermann, Julia Mayer

(Pankaj Kumar, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU), Fachbereich Informatik, Gottlieb-Daimler-Straße, Gebäude 48, 67663 Kaiserslautern, sharmap@rptu.de)

(Sibylle Hestermann, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU), Pfaffenbergstraße 95, 67663 Kaiserslautern, sibylle.hestermann@ru.rptu.de)

(Julia Mayer, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Trippstadter Str. 122, 67663 Kaiserslautern, julia.mayer@dfki.de)

1 ABSTRACT

The availability of financial resources determines the extent to which municipalities can proactively manage critical aspects such as sustainable urban planning, infrastructure development, public service improvement, and regional development (Zimmermann, 2018). The foresighted management of these financial resources is, therefore, crucial for municipal decision-making. This project introduces a decision-support tool designed to analyze and visualize financial patterns among municipalities in North Rhine-Westphalia, Germany.

Drawing on municipal financial data and population records for the 22 urban municipalities in North Rhine-Westphalia from 2009 to 2022, the application is built on a Flask-based web framework integrated with Python libraries, including pandas, scikit-learn, and GeoPandas. These libraries enable robust data processing, analysis, and visualization. GeoPandas facilitates geospatial analyses and visualizations, which are integral to understanding spatial trends in municipal finance (Jordahl et al., 2014). The system identifies "financial twins" – municipalities exhibiting comparable financial behavior in terms of income and expenditures across domains such as education, public safety, and infrastructure. The inclusion of population records allows for per-capita and absolute value analyses, ensuring contextual relevance.

The tool provides a list of financial twins for all 22 urban municipalities based on cosine similarity and Euclidean distance. It offers corresponding data for each year from 2009 to 2022, with the flexibility to analyze twins either with or without factoring in the population of each municipality for a given year. The application also employs the k-means clustering technique to group municipalities into intuitive categories, highlighting shared financial characteristics.

The tool can serve as a resource for:

- Benchmarking and Collaboration: Identifying peer municipalities to share successful financial strategies.
- Exploration and Comparison: Highlighting similarities and differences, and their impacts on financial decision-making.
- Policy Development: Offering actionable insights into resource allocation and prioritization.

This initiative underscores the critical role of data-driven approaches in addressing the complexities of urban financial planning in public administration. By making sophisticated clustering and visualization accessible to planners, the tool bridges the gap between raw data and actionable policy insights. Future work will explore predictive modelling and integration with socio-economic data to further enhance its relevance for strategic urban development.

Keywords: big data, financial twin app, decision support, financial data, clustering

2 MOTIVATION

The analysis of municipal financial data is crucial for effective governance and sustainable community development. Municipalities are responsible for providing essential services such as education, public safety, infrastructure, and health care. Understanding the financial health of a municipality enables decision-makers to allocate resources efficiently, ensuring that these services are delivered effectively and equitably (Bahl & Linn, 1992). One of the primary reasons for analyzing municipal financial data is to identify trends and patterns that can inform policy decisions. By examining revenue sources, expenditure patterns, and debt levels, municipal leaders can gain insights into the fiscal sustainability of their communities. This analysis helps in forecasting future financial conditions, allowing municipalities to prepare for economic fluctuations and demographic changes (Mikesell, 2013).Moreover, the analysis of financial data fosters transparency and accountability. Citizens have a right to know how their taxes are being spent. By making financial data accessible and understandable, municipalities can build trust with their constituents. Engaged citizens are more likely to participate in local governance, advocate for necessary changes, and support initiatives that enhance community well-being (National League of Cities, 2018).Additionally, financial data analysis aids in strategic planning. Municipalities can identify areas where efficiencies can be gained or where investments are needed most. For instance, analyzing data on infrastructure maintenance costs can highlight the need for preventive measures, ultimately saving money in the long run (Gordon, 2015).

Amid growing challenges from demographic shifts and economic pressures, leveraging financial data is essential for navigating the complexities of modern governance (Bourdeaux, 2008).This study applies clustering and similarity analysis approaches to enhance the understanding of municipal financial data and enable policymakers in making better-informed decisions. First, an overview over existing research in the field of digital twins in municipal policymaking is provided,followed by a detailed description of the dataset used for this approach. Before presenting insights into the application, the methodological approaches are outlinedin greater detail. The study concludes with an outlook on potential future research steps.

2.1 Existing research approaches in the field of municipal financial data analysis

Municipal financial data analysis is essential for maintaining transparency, accountability, and efficient resource allocation in local government operations in Germany. Local governments are responsible for key public services, including education, infrastructure, social welfare, and public safety, all of which require significant financial oversight. By analyzing municipal financial data, local governments can ensure that their expenditures align with budgetary constraints, forecast future revenue streams, and monitor the long-term sustainability of investments (Löffler & Hoffer, 2015).

Research highlights that municipal financial analysis plays a crucial role in ensuring fiscal discipline and effective governance. For instance, Lehtonen (2017) emphasizes that financial transparency at the local level improves public trust, while Feld &Kirchgässner (2009) argue that rigorous financial management practices contribute to reducing public debt and avoiding fiscal crises. Such analysis also aids in identifying inefficiencies, prioritizing public spending, and assessing the economic health of municipalities (Scharpf, 2003). Furthermore, studies like Homburg (2010) suggest that financial data analysis provides a foundation for evidence-based policy decisions, enabling municipalities to navigate challenges such as economic downturns or demographic shifts. By making financial information accessible to citizens and stakeholders, municipalities foster greater civic participation and accountability (Bovens, 2005).

While most digital twin application focus on providing additional insights and scenario-based planning approaches by creating a hypothetical twin model of the real subject, our approach focuses on identifying an existing financial twin using approaches from similarity analysis. With this approach we not only aim to identify existing patterns but to also enable municipalities to discuss strategies and policy options based in similarities in very specific areas of responsibilities.

In the context of enhancing the understanding of municipal financial data, the introduction of the financial twin application serves a dual purpose: (1) applying clustering approaches to reveal patterns across municipalities based on financial data, and (2) utilizing similarity analysis approaches to provide deeper insights into the specificexpenditure categories causing financial similarity. This clustering approach extends beyond traditional approaches in grouping municipalities, which predominantly focus mostly on regional structure, overall resource availability and by demographic factors such as youth quotient. The proposed approach uses financial data to an extent to which learning processes can be deducted from the initial analysis and with furthering this research important factors influencing changes in clustering patterns can be identified. The approaches with similarity measures takes those insights further in enabling the identification of product groups and detailed levels of municipal expenditure data that drive similarities between municipalities.

3 DATA SOURCE

The municipal structure in Germany plays a crucial role in the federal system of governance, with municipalities (known as "Kommunen") being the local level of government responsible for various administrative, social, and economic functions. Municipalities in Germany vary in size, scope, and population, from small rural villages to large cities. Despite their differences, they all share a basic framework of responsibilities and governance structures. The responsibilities are vast, ranging from social welfare and public safety to urban planning, education, and infrastructure management. For each municipal responsibility expenditure (payouts) and income (payin) data can be attributed. The structure thereby is designed to promote local autonomy, ensuring that municipalities can tailor their services to the specific needs of their residents. The municipal responsibilitiesand the corresponding expenditure and income data can be sorted into many categories (see Figure 1).

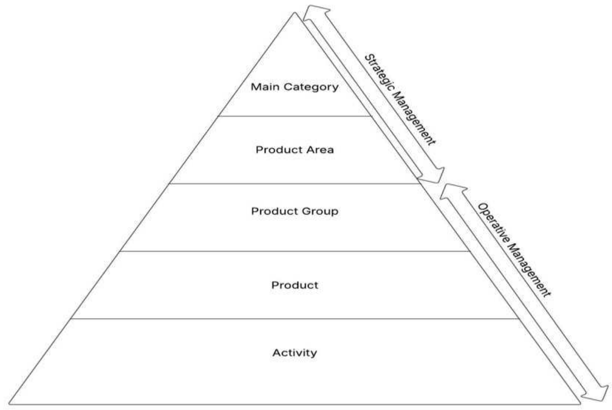

Fig. 1: Product pyramid of municipal accounting

Examples can be local government and administration, public services and infrastructure, social services, public safety and emergency services, culture, sports and leisure, economic development and local business support. Each of those main categories can be specified even further. Social services, for example, includeproduct areas such as childcare and specifying again even further children day care centers. In our data set we analyze the expenditure and income data for 102 product groups.

Municipal structures in Germany generally differentiate between (district-free) cities, towns and districts.While (district-free) cities and districts share a similar level of responsibility, towns and smaller municipalities have fewer obligations to fulfill. The partition of responsibilities between states, (district-free) cities, districts and smaller municipalities varies across federal states. For the initialimplementation of the financial twin app, the focus is therefore set on the state of North-Rhein Westfalia (NRW)to ensure comparability under uniform responsibilities and similar expenditure potentials. The study analyzes a newly composed dataset consisting of time series data from 22 urban municipalities in NRW, covering the period from 2009 to 2022. The dataset includes the cash inflow and cash outflow of those municipalities. Since the objective of this research is to identify common expenditure and income patterns, we are not focusing on overall expenditure and spending but rather on the expenditure and income within individual product groups.Focusing on the level of product group, this analysis is observing both strategic management decisions and a more detailed level of operative management decisions on how to provide sufficient service to the public. Toadd another level of depth to the analysis population data is incorporated to enable analysis of per capita values. This enhances an even better comparability of financial behaviors across municipalities.

4 METHODOLOGY

To measure the similarity between the financial structure of two municipalities based on the provided data, several preliminary steps are required. First, it is necessary to identify the relevant attributes and determine an appropriate weighting strategy. Secondly, the user needs to be provided with a preselection of the most suitable distance metrics for this specific application. The chosen clustering technique is also described in detail. Moreover, this section presents a detailed explanation of the system design of the tool, outlining its structure, key components, and their interactions to ensure a well-organized and effective implementation.

4.1 Data Preparation and Analytical approaches

The datasets available, which have been explained in detail in Section 3, provide a wealth of information that needs to be meaningfully correlated. To achieve this, the datasets themselves must be prepared, and appropriate techniques for analysis must be selected.

4.1.1 Identifying key parameters

It makes a significant difference when considering whether to analyzethe absolutefigures of a municipality or the per capita values. Therefore, the total population of the municipality provides valuable information. The idea of selecting or aggregating individual accounts to simplify understanding and enhance clarity was discarded, as it could, under certain circumstances, hinder detailed comparability by experts. Instead, users are given the option to select the accounts relevant to their specific analysis. In the default mode, a comprehensive view of all accounts is always provided. Inflows and outflows are considered separately. The user can choose whether to focus on one or the other.

The option to assign greater weight to individual accounts has not yet been implemented. It is also conceivable to take the age distribution of the population into account. In the first version of the application, municipalities are also only compared within the same year, even though multiple years are available, and the comparison could be extended.

4.1.2 Selection of distance measures

To measure the similarity between two municipalities, pairwise distances based on either relative or total figures are utilized. Depending on the calculated value, one can make a statement about the proximity and therefore similarity of the respective municipalities. This can be done in different ways, leading to varying results. Therefore, the choice of an appropriate metric is crucial, and the correct interpretation of the results is of utmost importance. Two measures[1] have been identified as relevant:

For two given cities, P and Q, with the values of their respective accounts (in absolute or per capita values)$p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$, their *Euclidean Distance* is calculated as:

$$d_{eucl}(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2}$$

Based on the pairwise differences in the amounts of individual accounts, an overall distance is calculated. The smaller the results, the closer and thus more similar the two municipalities are.It must be noted that accounts with very high amounts can generally have a significantly greater impact. This, along with the fact that there are many zero entries, leads us to the second approach, which utilizes Cosine Similarity. In this case, the magnitude of expenditures does not matter; rather, the focus is on the pattern of accounts that are nonzero.

Again,the n-dimensional vectors p and q as above are being used, this time to first define the *Cosine Similarity*:

$$\cos(p, q) = \frac{p * q}{\|p\| \cdot \|q\|},$$

where * denotes the dot product of two vectors.

This results in the *Cosine Distance*:

$$d_{\cos}(p, q) = 1 - \frac{p * q}{\| p \| \cdot \| q \|}$$

Since only non-negative entries occur in this use case, the values of$d_{\cos}(p, q)$range between 0 and 1. A value of 0 indicates that the vectors are identical in terms of direction, while a value of 1 signifies the maximum possible difference between them.

The two measures are not completely independent of each other, but in this case, they may lead to very different interpretations. Additionally, two municipalities may be close to each other when using one measure while exhibiting a large distance when using the other (see example in Figure 2).

[1]The Python library used (scipy.spatial.distance.pdist) employs the parameter name "metric", even though Cosine Distance is not a metric in the mathematical sense, as it does not fulfill the conditions of the triangle inequality.

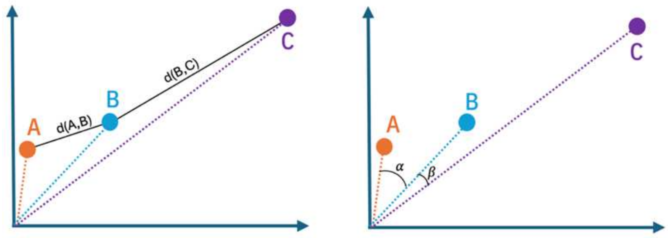

Fig. 2: The Euclidean distance d(A,B) is smaller than d(B,C) (left), whereas the cosine distance α between points A and B is greater than the cosine distance β between B and C (right). Own illustration.

4.1.3 Clustering technique and optimal number of clusters

With the aim of partitioning the given municipalities into a predefined number of disjoint groups, the well-established k-means algorithm was employed. The algorithm is provided with the desired number of clusters and, through an iterative process based on the concept of centroids, assigns municipalities to groups in a manner that minimizes their variance as much as possible.[2]

In a practical context, there is no single, universally defined true number of clusters. The optimal number of clusters must be determined based on data-driven criteria and an understanding of the specific objectives. To determine an optimal number of clusters, Silhouette Scores (Rousseeuw, 1987) were used. They provide an indication of the quality of the clustering results by calculating, for each data point, the average distance to all other points within the same cluster and comparing it to the shortest average distance to points in the nearest neighbouring cluster. The Silhouette Score is then derived by computing the ratio between these distances, resulting in a value between -1 and 1, where higher values indicate better-defined clusters.By comparing Silhouette Scores for different numbers of clusters, the optimal number of clusters can be identified (see Figure 3). In this use case, the optimal number of clusters may depend on the chosen year, the selected accounts, and whether absolute or per capita values are considered.

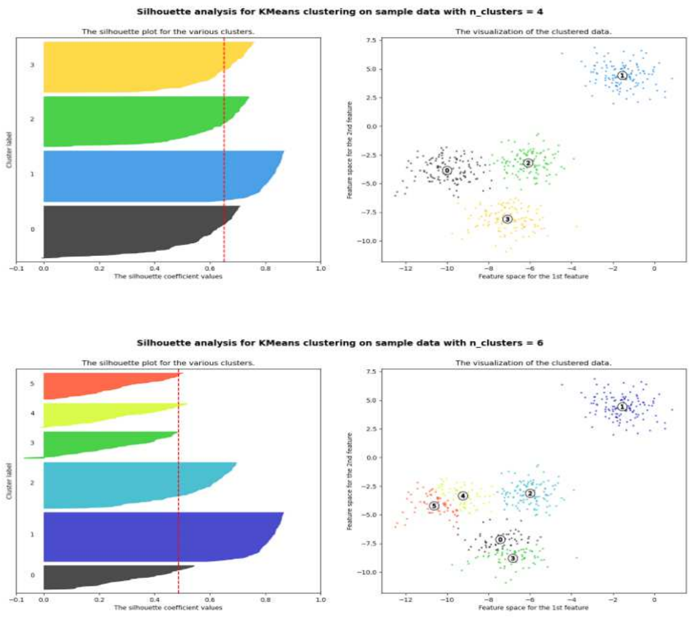

Fig. 3: Example results of two k-means clustering applications with 4 and 6 clusters. The Silhouette Score is indicated by the red dashed line. It can be observed that the 4-cluster solution achieves a higher score (Pedrogosa 2011).

Since the algorithm depends on the pairwise distances between vectors, different metrics can again lead to different results. By default, the Euclidean distance is used in the Python library scikit-learn (Pedrogosa 2011), and this is also the case here.

[2] Since k-means is a heuristic algorithm, it may only find a locally optimal solution.

4.2 Architecture and User Interaction

The architecture of any web application serves as the foundation that supports scalability, performance and maintainability. A well-structured system facilitates seamless interaction between backend and frontend components which enables efficient data processing and subsequent visualization of the results. The decision to develop a web application was driven by the ease of use for end users and the simplicity of distribution, ensuring accessibility without the need for complex installations. This section highlights the key architectural design adopted in this Flask-based web application.

4.2.1 Overview of the Flask-Based Web Application

The application is designed with a modular and scalable architecture, ensuring efficient financial analysis for municipalities in NRW. It integrates backend processing for data computation and financial insights with an interactive frontend for clear visualizations and user-friendly interaction. Figure 4 displays the interaction between the various architectural components of the application.

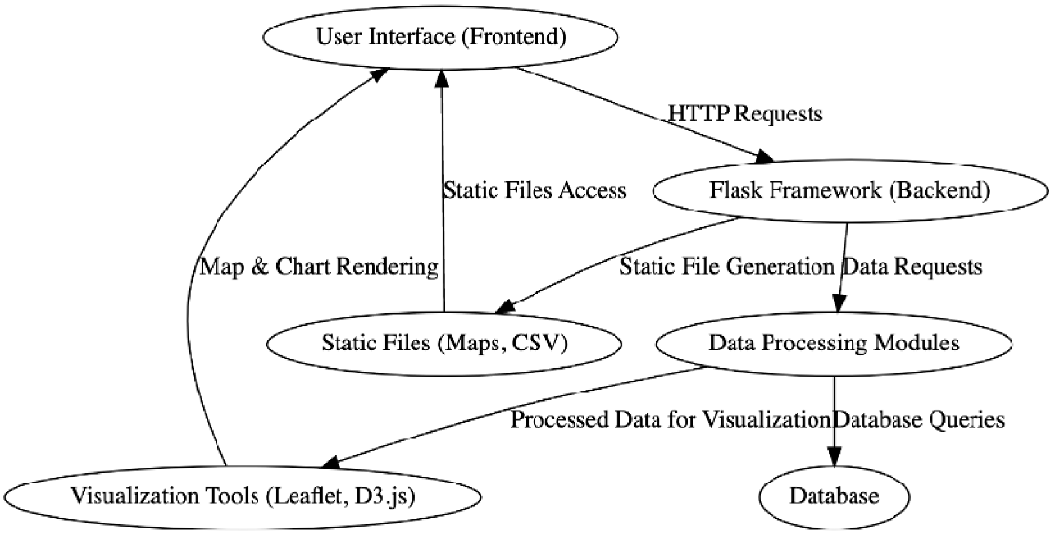

Fig. 4: Architectural representation: Illustration of the interaction of the different components of the application.

The key components of the architecture can be broadly categorized into Backendand Frontend. The Frontend components can be further differentiated between the interactive, User Interface, Flow of User Interaction and Visualization.

4.2.2 Backend: Flask Framework and Data Processing Modules

The backend forms the core of the application and is built on Flask, a lightweight yet efficient Python framework. It processes user requests, manages data interactions, and dynamically delivers results. The framework handles HTTP request management, user session control, and error handling, ensuring system reliability. Additionally, it employs the Jinja2 template engine to render dynamic content,which seamlessly integrates with the frontend.

The data processing modules utilize Python libraries such as pandas, NumPy, and SciPy to perform advanced calculations, including cosine similarity and Euclidean distance measures. Clustering algorithms like k-means from scikit-learn are applied to identify financial twin clusters. A custom module normalizes and aggregates datasets, generates insights, and prepares data for visualization. The application dynamically retrieves and processes financial data from Excel files and geospatial shapefiles, ensuring real-time updates based on user interactions with municipal and financial parameters.

4.2.3 Frontend: Interactive User Interface, Flow of User Interaction and Visualizations

The frontend is designed to provide a user-friendly and interactive environment for financial data exploration. The Jinja2 template engine dynamically renders HTML content, while Bootstrap and CSS frameworks ensure responsive and visually appealing layouts. Users can input parameters such as municipality selection, financial years, and clustering attributes through intuitive forms. The system presents results in structured tables and interactive charts, allowing users to analyze financial insights effectively.

Export options are available for downloading analysis results in Excel or ZIP formats containing maps. Navigation across the application is seamless, with dedicated pages for Startup, Analytics, and Map Generation. Users can explore financial data dynamically by selecting relevant options, and the system processes these inputs in real time, providingimmediate feedback.

The interaction begins with input parameter selection. Users choose a municipality from the Startup page by clicking on a financial twin displayed in the table. The system then selects the corresponding financial metric automatically. Additional parameters such as the year and population metric (inclusion or exclusion of per capita value in analysis) can be adjusted via dropdown menus. Once the parameters are set, the backend processes the selection to compute financial similarities, generate clustering results, and produce financial twin comparisons. The results include detailed twin pairs, key differences, and distance metrics.The visualization phase presents these findings in the form of interactive tables and maps, allowing users to explore financial twin clusters with ease. The system enables iterative refinements, allowing users to modify parameters and analyze new results dynamically. Additionally, users can export data and visualizations for offline analysis.

The application integrates various visualization tools to ensure clear and actionable financial insights through geospatial representations. Leaflet.js, a lightweight JavaScript library, enables dynamic and interactive maps, with distinct colors representing financial twin clusters for enhanced differentiation. D3.js supports custom visualizations, including graphs, sliders, and trend charts that help analyze financial patterns over time. Geospatial data integration is achieved using GeoPandas, ensuring accurate mapping of financial clusters based on municipal shapefiles. The system also includes export features that allow users to download maps as images or CSV files for further analysis.

5 IMPLEMENTATION DETAILS

The implementation of this application follows a structured approach which facilitates seamless flow during the financial twin analysis. It details the key features and functionalities, highlighting how different components work together. The following subsections describe the core features of the application and their roles in ensuring efficient financial data exploration.

5.1 Application Features

The application features are designed to provide meaningful insights in the context of financial twin analysis. The following subsections summarize the role of each page of the web application. It highlights the functionality and significance of each component.

Municipality	Payins Cosine Twin	Payins Euclidean Twin	Payouts Cosine Twin	Payouts Euclidean Twin
Düsseldorf, krfr. Stadt	Münster, krfr. Stadt	Hagen, krfr. Stadt	Köln, krfr. Stadt	Köln, krfr. Stadt
Duisburg, krfr. Stadt	Herne, krfr. Stadt	Leverkusen, krfr. Stadt	Münster, krfr. Stadt	Wuppertal, krfr. Stadt
Essen, krfr. Stadt	Mülheim an der Ruhr, krfr. Stadt	Mülheim an der Ruhr, krfr. Stadt	Hamm, krfr. Stadt	Mülheim an der Ruhr, krfr. Stadt
Krefeld, krfr. Stadt	Köln, krfr. Stadt	Köln, krfr. Stadt	Herne, krfr. Stadt	Leverkusen, krfr. Stadt
Mönchengladbach, krfr. Stadt	Dortmund, krfr. Stadt	Dortmund, krfr. Stadt	Oberhausen, krfr. Stadt	Oberhausen, krfr. Stadt
Mülheim an der Ruhr, krfr. Stadt	Essen, krfr. Stadt	Essen, krfr. Stadt	Gelsenkirchen, krfr. Stadt	Solingen, krfr. Stadt
Oberhausen, krfr. Stadt	Mönchengladbach, krfr. Stadt	Mönchengladbach, krfr. Stadt	Mönchengladbach, krfr. Stadt	Mönchengladbach, krfr. Stadt
Remscheid, krfr. Stadt	Münster, krfr. Stadt	Duisburg, krfr. Stadt	Münster, krfr. Stadt	Hagen, krfr. Stadt
Solingen, krfr. Stadt	Hamm, krfr. Stadt	Wuppertal, krfr. Stadt	Herne, krfr. Stadt	Köln, krfr. Stadt
Wuppertal, krfr. Stadt	Hamm, krfr. Stadt	Münster, krfr. Stadt	Münster, krfr. Stadt	Münster, krfr. Stadt
Bonn, krfr. Stadt	Bielefeld, krfr. Stadt	Bielefeld, krfr. Stadt	Hagen, krfr. Stadt	Hagen, krfr. Stadt
Köln, krfr. Stadt	Krefeld, krfr. Stadt	Krefeld, krfr. Stadt	Herne, krfr. Stadt	Düsseldorf, krfr. Stadt
Leverkusen, krfr. Stadt	Düsseldorf, krfr. Stadt	Hagen, krfr. Stadt	Krefeld, krfr. Stadt	Krefeld, krfr. Stadt
Bottrop, krfr. Stadt	Oberhausen, krfr. Stadt	Oberhausen, krfr. Stadt	Gelsenkirchen, krfr. Stadt	Oberhausen, krfr. Stadt

Fig. 5: Startup Pageof the application, displaying financial twin analysis in a tabular format

5.1.1 Startup Page

The startup page (Figure5) serves as the entry point to the application and provides an overview of the financial twin analysis in a tabular format. It displays financial twins based on different mathematical models, namely cosine and Euclidean similarity (refer to Section 4.1.2). The table also includes payins and payouts, representing allocated funds and incurred expenses, respectively (refer to Section 3). By default, the displayed data corresponds to the year 2019, with per capita data as basis for the calculations. Users can modify these settings through a dropdown menu, selecting a different year or opting to exclude population data from the analysis. Additionally, the startup page features interactive action buttons that enhance user navigation. The "Check Clusters on NRW Map" button redirects users to a page displaying financial twin clusters on a map, providing a geospatial perspective. The "Check Financial Twin of Individual Municipalities" button allows users to retrieve detailed financial twin data for a specific municipality each year, facilitating comparative financial analysis.

5.1.2 Analytics Page

The analytics (Figure 6) page allows users to dive deeper into the financial twin analysis by examining the closest and farthest financial parameters between municipalities.

The analytics module provides a detailed comparison of financial parameters between a municipality and its financial twin. It highlights both the closest and farthest parameters, allowing users to identify areas of similarity and divergence. By default, the analysis considers five parameters, though users have the flexibility to adjust this range between one and ten. To facilitate further exploration, an "Export as Excel" option enables users to download the results for offline analysis.

Additionally, the interface supports interactive exploration, allowing users to click on the differences displayed in the table to expand and view a detailed breakdown of the individual values contributing to the computed differences.

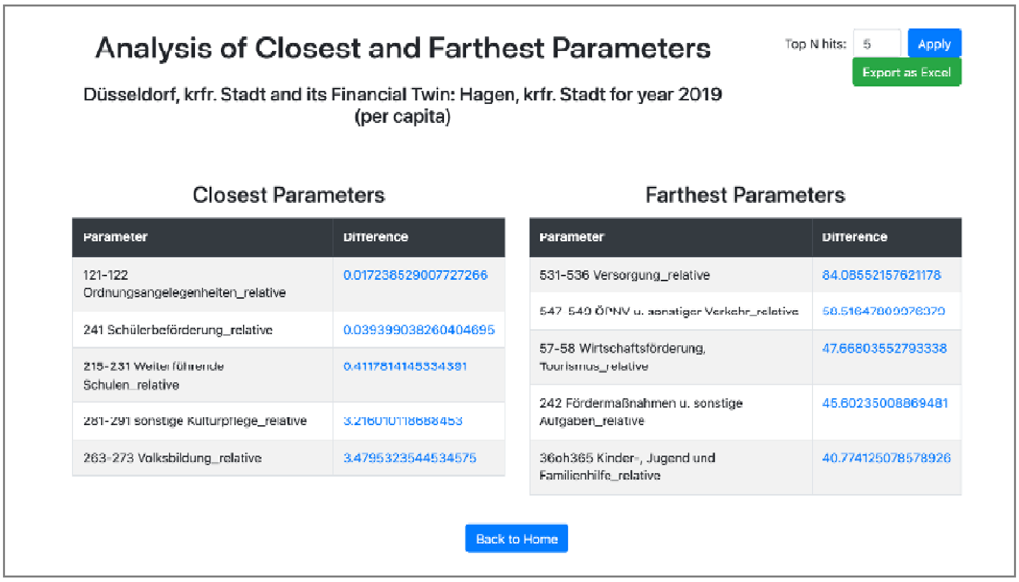

Fig. 6: The analytics page showing the closest and farthest financial parameters between a municipality and its financial twin.

5.1.3 Generate Maps Page

The generate maps page (Figure7) allows users to create visual clusters of municipalities based on specific input parameters.

The clustering module provides users with a flexible interface to customize financial data analysis. It includes input fields for selecting the financial data type, allowing users to choose between payins or payouts for clustering. The number of clusters can be defined manually, with a default value of four, or calculated dynamically based on historical data (2009–2022) using a built-in calculation function. Users can further refine their analysis by enabling the "Check Per Capita" option, which incorporates population metrics into the calculations. Additionally, a dropdown menu allows the selection of specific expense fields for clustering, with the default setting including all fields.

Upon submission, the "Generate Maps" button processes the selected parameters and visualizes the resulting clusters on an interactive map. A dedicated section enables users to compute the ideal number of clusters for a given year, considering all specified parameters. This feature is particularly useful for those seeking to explore financial distributions in greater detail and optimize cluster selection for more meaningful insights.

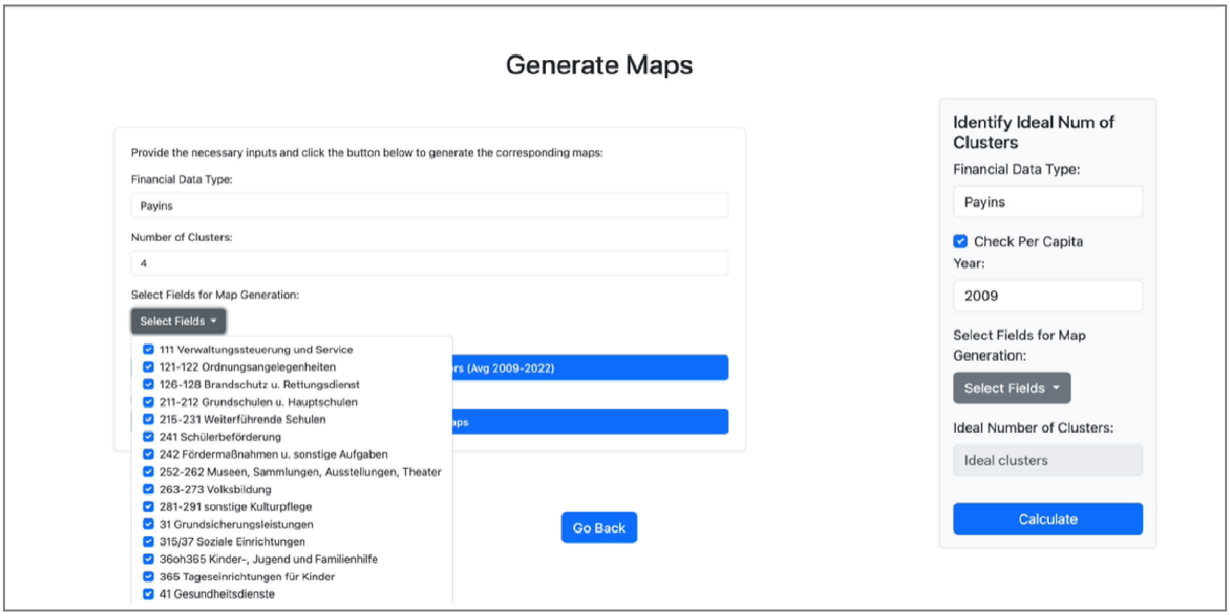

Fig. 7: The map generation interface – selecting financial data, clustering parameters, and calculating the ideal number of clusters.

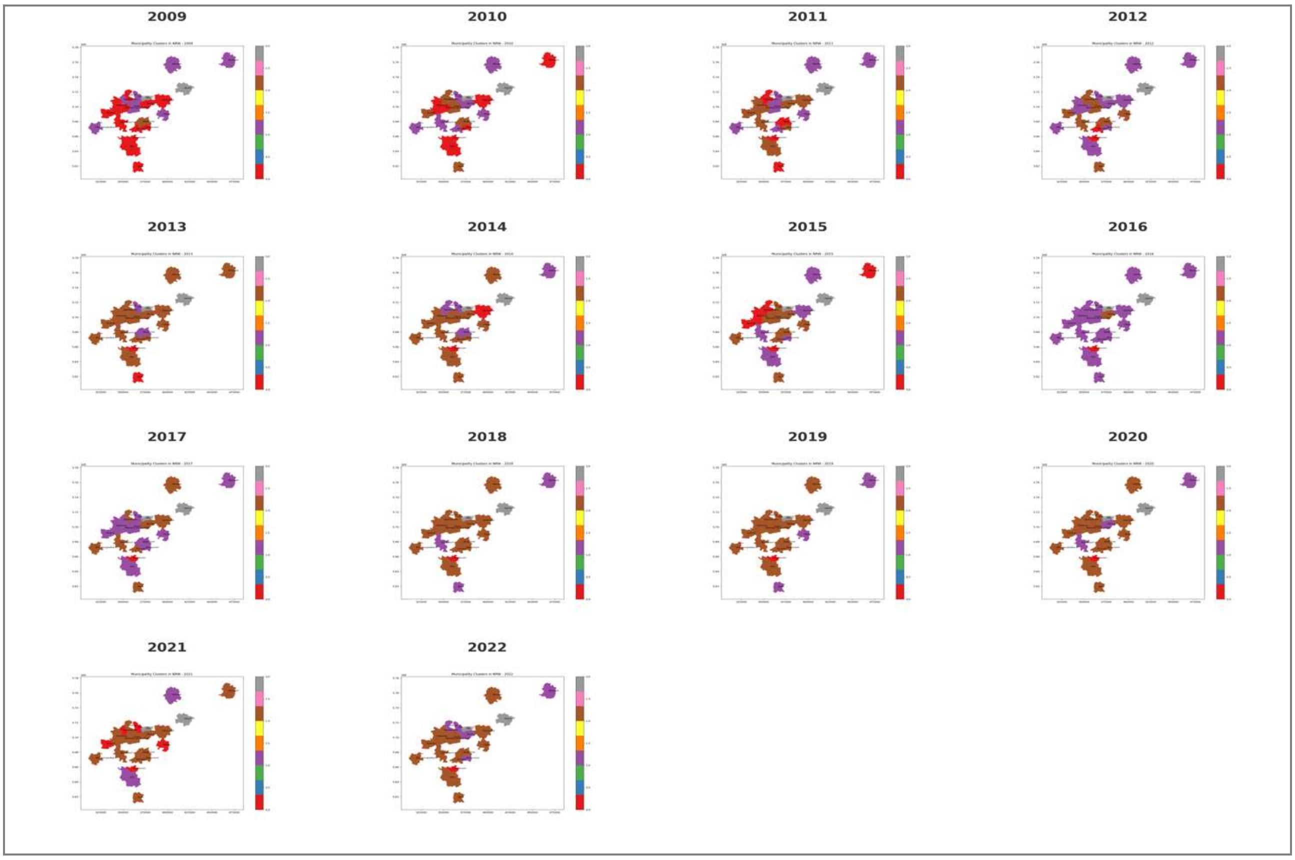

Fig. 8: Generated maps displaying municipality clusters in NRW for 2009 to 2022, with distinct colors representing different financial similarity groups.

5.1.4 Generated Maps Page

The generated maps page visualizes the financial twin clusters for NRW municipalities for all years (2009 to 2022) via a scrollable carousel display.This page provides an interactive visualization of clustered municipalities based on financial similarity, with each cluster represented by a distinct color for easy differentiation (see Figure 8). Users can navigate through different years using a slider, enabling comparative analysis over time. Additionally, the page offers export options, allowing users to download maps as images in a ZIP file or export map data as CSV files for further analysis and offline use.

6 RESULTS AND DISCUSSION

This section presents the key findings of the study, focusing on the identification of financial twins among municipalities in the German federal state of North Rhine-Westphalia (NRW) and the observed trends in financial clustering over time. The results demonstrate how demographic factors and different similarity measures influence financial analysis outcomes.

6.1 Key Insights from the Application

By leveraging mathematical models such as cosine similarity and Euclidean distance, the application effectively identifies financial twins – municipalities with comparable income and expenditure patterns. The ability to analyze financial similarity with and without population metrics allows for a more nuanced understanding of demographic influences on municipal financial structures. By implementing similarity analysis and clustering approaches the study enables both the identification of a singular twin as well as a cluster of similar municipalities.

6.1.1 Identification of Financial Twins in NRW

The application successfully pairs municipalities with similar financial behavior based on both income (payins) and expenditure (payouts). The inclusion or exclusion of population in these calculations significantly impacts the results, demonstrating the role of demographics in financial patterning.

Observations from the financial twin analysis reveal stable patterns across multiple methods. Several municipalities consistently appear as financial twins irrespective of the similarity measure used. For example, Düsseldorf is frequently linked with Leverkusen, Duisburg shares financial similarities with Herne, and Essen aligns closely with Mülheim an der Ruhr. The consistency of these twin pairings suggests robust financial similarities that persist over time.The impact of population inclusion by analyzing per capita values further refines these observations. When population metrics are incorporated, municipalities with comparable per capita income and expenses are grouped together. Conversely, when population is excluded, financial clustering is driven by absolute revenue and expenditure, leading to the grouping of larger municipalities. An example of this effect is observed in Düsseldorf, whose Euclidean twin shifts from Dortmund to Leverkusen when population is factored into the calculation. The analysis of per capita financial data is most often used since biases due to large overall spending can be neglected by focusing on per capita values. Since most of the comparative research to date focuses on using per capita values, the results presented by this study correspond in data structure with existing approaches.

6.1.2 Trends Observed in Cluster Maps

The clustering maps generated by the application provide a visual representation of financial similarities across municipalities from 2009 to 2022. These maps help identify patterns in financial behavior and highlight key trends over time.

Impact of Population on Clustering: The consideration of per capita values significantly affects cluster formation. When population is considered, clusters display a more balanced distribution, with municipalities grouped based on financial efficiency rather than sheer size. This leads to smaller municipalities forming distinct clusters that reflect their financial management practices. In contrast, when population is not considered, clusters become dominated by larger municipalities, overshadowing smaller ones. Financially strong cities such as Düsseldorf, Cologne, and Dortmund consistently form distinct clusters under this approach, reinforcing the influence of absolute financial metrics in cluster formation.

Evolution of Financial Clusters Over Time: Financial clustering is not static; municipalities shift between clusters over different years, reflecting changing economic conditions. These shifts may be influenced by factors such as industrial growth, policy changes, or economic fluctuations. Certain municipalities, however, display persistent clustering behaviors. For instance, Düsseldorf and Leverkusen frequently appear in the same financial cluster across multiple years, suggesting long-term financial stability and structural similarity. The evolving nature of financial clusters highlights the need for continuous monitoring of municipal finances. Understanding these trends allows policymakers and analysts to anticipate economic shifts, optimize resource allocation, and develop strategies that enhance financial resilience at the municipal level.

7 OUTLOOK AND FUTURE WORK

This study advances existing research on municipal financial data analysis by demonstrating how clustering and similarity analysis approaches can uncover valuable insights into the financial structures of municipalities. By identifying broader groups with financial similarities through clustering and pinpointing singular financial twins using similarity analysis, the proposed methods offer practical support for various urban stakeholders in their decision-making processes.

The current research focuses on urban municipalities in the German federal state of North Rhine-Westphalia, allowing for in-depth comparative analysis within a uniform framework of responsibilities. Expanding the scope of the study to include municipalities from other federal states requires the availability of data in a comparable structure. Only with standardized and harmonized financial data across states can the analysis maintain its validity and enable meaningful comparisons.

Future research should address inter-state differences in municipal responsibilities, enabling a broader dataset and facilitating more comprehensive analyses through data assimilation strategies. Additionally, future work will explore intertemporal comparative approaches to help decision-makers identify financial twins over time. Such longitudinal analysis can foster a learning process and serve as a catalyst for communication among municipal financial decision-makers.

Beyond its analytical applications, the financial twin app could also serve as a practical tool to demonstrate to municipalities and statistical offices the potential benefits of well-structured financial data. By showcasing concrete use cases and analyses, the app can highlight how structured financial data can optimize decision-making processes and promote data-driven strategies. This approach would not only enhance the transparency and comparability of municipal financial data but also emphasize the added value of systematic and structured data collection.

Expanding the methodological toolkit to include alternative clustering techniques and involving experts in evaluating these methods will further strengthen the robustness of the analysis. This approach will help mitigate potential technical limitations of specific methods, ensuring that the most suitable approach is applied to the given data.

8 REFERENCES

BAHL, R. W., & LINN, J. F. (1992). Urban Public Finance in Developing Countries. Oxford University Press.

BOURDEAUX, C. (2008). The Role of Financial Data in Local Government Accountability. Public Budgeting & Finance, 28(1), 1-20.

BOVENS, M. (2005). Public Accountability. Public Administration, 83(2), 25-26.ZIMMERMANN, H. (2018): Kommunalfinanzen. In: ARL – Akademie für Raumforschung und Landesplanung (Hrsg.): Handwörterbuch der Stadt- und Raumentwicklung. Hannover 2018. 1159-1170.

FELD, L., & KIRCHGÄSSNER, G. (2009). Direct Democracy and Local Government Performance. Public Choice, 139(1), 91-111.

GORDON, T. (2015). The Role of Financial Data in Local Government Decision-Making. Public Administration Review, 75(3), 345-356.

HOMBURG, V. (2010). Financial Management and the Role of Local Government. Local Government Studies, 36(2), 203-220.

JORDAHL, K., & contributors (2014): GeoPandas: Python tools for geographic data. Retrieved from https://geopandas.org/en/stable/about/citing.html (accessed: 16.12.2024)

LEHTONEN, M. (2017). Financial Transparency and Governance in Local Government. Journal of Public Administration Research and Theory, 27(3), 483-505.

LÖFFLER, E., & HOFFER, R. (2015). Public Sector Financial Management in Europe: Local Government. International Journal of Public Sector Management, 28(3), 201-222.

MIKESELL, J. L. (2013). Fiscal Administration: Analysis and Applications for the Public Sector. Cengage Learning.

NATIONAL LEAGUE OF CITIES (NLC). (2018). City Fiscal Conditions 2018.

PEDROGOSA et al. (2011): Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830.

ROUSSEEUW, Peter J. (1987): Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Computational and Applied Mathematics 20: 53-65. DOI: https://doi.org/10.1016/0377-0427(87)90125-7

SCHARPF, F. W. (2003). Governance and Public Management. Oxford University Press.