algorithms

MDPI

Article

Automated Recommendation of Aggregate Visualizations for Crowdfunding Data

Mohamed A. Sharaf [1] *, Heba Helal [1], Nazar Zaki [1], Wadha Alketbi [1], Latifa Alkaabi [2], Sara Alshamsi [1] and Fatmah Alhefeiti [2]

[1] Department of Computer Science, College of Information Technology, United Arab Emirates University, Al Ain P.O. Box 15551, United Arab Emirates; 202090034@uaeu.ac.ae (H.H.); nzaki@uaeu.ac.ae (N.Z.); w_alketbi@uaeu.ac.ae (W.A.); sara.alshamsi@uaeu.ac.ae (S.A.)

[2] Department of Artificial Intelligence and Computer Vision Engineering, Abu Dhabi Autonomous Systems Investments (ADASI), EDGE Group, Abu Dhabi P.O. Box 109667, United Arab Emirates; latifa.alkaabi@adasi.ae (L.A.)

* Correspondence: msharaf@uaeu.ac.ae

Abstract: Analyzing crowdfunding data has been the focus of many research efforts, where analysts typically explore this data to identify the main factors and characteristics of the lending process as well as to discover unique patterns and anomalies in loan distributions. However, the manual exploration and visualization of such data is clearly an ad hoc, time-consuming, and labor-intensive process. Hence, in this work, we propose *LoanVis*, which is an automated solution for discovering and recommending those valuable and insightful visualizations. LoanVis is a data-driven system that utilizes objective metrics to quantify the "interestingness" of a visualization and employs such metrics in the recommendation process. We demonstrate the effectiveness of LoanVis in analyzing and exploring different aspects of the *Kiva* crowdfunding dataset.

Keywords: data exploration; visual analytics; aggregate views

check for updates

Citation: Sharaf, M.A.; Helal, H.; Zaki, N.; Alketbi, W.; Alkaabi, L.; Alshamsi, S.; Alhefeiti, F. Automated Recommendation of Aggregate Visualizations for Crowdfunding Data. *Algorithms* **2024**, *17*, 244. https://doi.org/10.3390/a17060244

Academic Editor: Francesc Pozo

Received: 14 March 2024
Revised: 24 April 2024
Accepted: 25 April 2024
Published: 6 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Crowdfunding, also called peer-to-peer lending, social lending, or crowd lending, is an internet-based fundraising mechanism soliciting small monetary contributions from crowd donors to help others in need [1,2]. The importance of crowdfunding is further underscored as economies worldwide are racing to meet the *Sustainable Development Goals* (SDGs) by 2030. Particularly, crowdfunding plays an important role in achieving those goals, especially the ones related to poverty, hunger, health, education, and gender equality [3].

Generally speaking, the mainstream crowdfunding platforms can be classified into four categories, i.e., donation-based, reward-based, equity-based, and lending-based ones [4]. Among them, the donation-based ones are becoming increasingly popular. Since the first crowdfunding platform Zopa was established in 2005, more lending platforms have emerged (e.g., Prosper, LendingClub, Kiva, Street Shares, Upstart, and Renrendai) [2,4].

Among those crowdfunding platforms, *Kiva*, which is the focus of this work, is becoming an increasingly popular platform. Kiva's goal is to develop an instrument to use charitable loans to combat and eradicate poverty [5]. Lenders provide money with the understanding that they would only recoup their initial investment and lose any profit [6]. Kiva has provided people and groups with modest incomes with loans totalling more than $1.4 billion [4].

The rapid development of crowdfunding platforms, such as Kiva, has attracted much attention from the data analytics research community. Particularly, solutions have been proposed to address some of the interesting research problems that arise in crowdfunding

platforms, such as predicting project success [7], tracking the funding dynamics [8], recommending donors [9], recommending projects for donors [10–12], etc. Orthogonal to the existing work mentioned above, our focus in this paper is to utilize visual data exploration techniques, and in particular visualization recommendation systems, to unlock valuable insights from crowdfunding data, as described next.

Visual data exploration is an essential step in the data science pipeline, in which analysts examine datasets up-close to extract valuable insights [13–16]. This process has traditionally been performed manually, where the analyst interactively applies various exploratory queries (such as SQL-based filtering, aggregation, joins, etc.). The results of those queries are presented as data-driven visualizations (e.g., bar or line charts, scatter plots, etc.). The analyst then examines those visualizations looking for insights, which are used as a springboard to decide their next analytical query.

However, unlocking those insights has been anecdotally compared to "finding a needle in a haystack" [17]. That is particularly true for big high-dimensional datasets with tens to hundreds of attributes and measures, as it is typically the case in financial data warehouses and scientific datasets [13]. Particularly, the "curse of dimensionality" leads to analysts having to manually construct a prohibitively large number of queries and visually explore their results looking for insights, which is clearly an ad hoc and labor-intensive process. That challenge motivated multiple research efforts that focused on *automatic* recommendation for data exploration. That is, recommender systems dedicated to providing the user with suggestions for specific, high-utility visualizations (e.g., [17–28]). Such systems are data-driven (also known as discovery-driven) systems, which use heuristic notions of "interestingness" and employ them in the recommendation. The main idea underlying those solutions is to automatically generate "all possible" exploratory queries of the data, generate their corresponding visualizations, and recommend the *top-k* interesting ones, where k is a user-defined parameter. Meanwhile, the interestingness of a query/visualizations is quantified using some utility metric over its result. For instance, an exploratory query, which applies filters such as: `WHERE country = 'Kenya' AND borrower-gender = 'Female';` is considered interesting if it maximizes some quantifiable metric (e.g., skewness, surprisingness, diversity, or deviation from an expected distribution [19,26,27]).

In exploring crowdfunding datasets, analysts are typically interested in identifying unique patterns, anomalies, and discrepancies across the different aspects of such data (e.g., borrower-gender, country, project-activity, etc.) [29–31]. For instance, the data visualization tool Tableau features a visualization-based case study of the Kiva dataset called Kiva Loan Story [32]. Furthermore, in an attempt to gain insights from crowdfunding data, Kiva deploys its own data visualization and analytics dashboard [30]. However, such a dashboard mainly supports basic data filtering operations and rudimentary visualizations based on the user's search parameters. That is, it is expected that the user knows exactly in advance the insights they are looking for! However, such insights become clear only in "hindsight" after spending a long time exploring the data.

Hence, in this work, we present *LoansVis*, our visual analytics platform, which employs well-studied utility metrics (e.g., [19,21,24,26,27]) to automatically recommend interesting data visualizations that reveal hidden insights (e.g., discrepancies and anomalies in the distribution of loans across different countries or by gender). Accordingly, we have designed our proposed LoanVis to automatically provide recommended visualizations in two forms: (1) *Value-based* recommendation: in which LoanVis recommends high-utility visualizations based on some specific values provided by the analyst (e.g., recommends aggregate visualizations based on the analyst manual choice of: `country = 'Kenya'`), and (2) *Aspect-based* recommendation: in which the analyst's manual interaction is further minimized and they only have to specify their aspect of analysis (e.g., `country` or `borrower-gender`), then LoanVis automatically recommends both: (i) a specific value for the user-specified aspect (e.g., `country = 'Congo'`), and (ii) high-utility insightful visualizations based on that recommended value (e.g., a visualization of loan distribution across gender in the country of Congo).

Our extensive experimental evaluation demonstrates LoanVis' ability to automatically detect and recommend interesting visualizations based on the Kiva crowdfunding dataset. Particularly in this work, we present some of the recommended visualizations that show unique and often unexpected patterns in the distribution of loans across different aspects of analysis (e.g., country, gender, sector, etc.). Our findings complement and expand on existing related works that aim at studying the discrepancy and factors affecting crowdfunding (e.g., [2,29,30,33]). For instance, the work in [33] investigates how the gender, country, and type of borrower's business affect the lenders' lending decisions. The research in [2] examines the impact of forming groups on receiving fast funds for loan requests. Both of the works in [6,34] focus on the loan characteristics, either through examining the factors that motivate lenders to contribute to donations or that lead to the success or failure of crowdfunding.

However, while those studies relied on time-consuming manual data exploration and visualization, our proposed LoanVis facilitates understanding the characteristics of crowdfunding by automatically recommending visualizations that reveal unique patterns and discrepancies.

The rest of this paper is organized as follows. In Section 2, we present our methodology, which is based on a discussion of the Kiva dataset (Section 2.1), followed by the details of our proposed LoanVis system (Section 2.2). In Section 3, we present our results based on the visualizations recommended by our LoanVis system. We conclude in Section 4.

2. Materials and Methods

Differently from existing work that relies on manual data exploration for discovering insights from crowdfunding databases, in this work we propose *LoanVis*, which is an automated solution for discovering and recommending those valuable and insightful visualizations. LoanVis is a data-driven system that utilizes objective metrics to quantify the "interestingness" of a visualization and employs such metrics for recommending insightful visualizations, as shown in Figure 1. Particularly, the main idea underlying LoanVis is to automatically generate "all possible" visualizations and recommend the top-k interesting ones, where k is a user-specified parameter. However, to identify those interesting visualizations, LoanVis employs well-studied utility functions that assign each visualization a utility score [21,24,27], and then recommends the top-ranked visualizations according to that score. Notice that Figure 1 illustrates only a single iteration of that visualization recommendation process. However, that process is typically repeated multiple times throughout a data exploration session, along with other data exploration tools. For a comprehensive overview of the workflow involved in an end-to-end data exploration session, we refer the reader to [14]. In the following, we first describe the Kiva crowdfunding dataset, then we present our visualization recommendation methodology employed by LoanVis.

2.1. The Kiva Dataset

The Kiva dataset has attracted the attention of multiple research studies in the area of data science and analytics (e.g., [2,6,9,35]). The overarching goal of such research is to utilize the Kiva data set for extracting knowledge, gaining valuable insights into the crowdfunding model, and understanding its loan funding characteristics. In this paper, we expand on such research efforts and present our *LoanVis* system, which automates the process of exploring the Kiva crowdfunding dataset and provides analysts with fast data-driven recommendations of insightful visualizations.

There are mainly four types of participants in Kiva, namely: loans, borrowers, lenders, and field partners. The loans are in the form of fund-raising campaigns posted by field partners on behalf of the borrowers. Particularly, field partners are typically local non-profit organizations that act as the link between borrowers and Kiva. Lenders participate in a donation-based crowdfunding model (i.e., donors receive back their principal investment without profits).

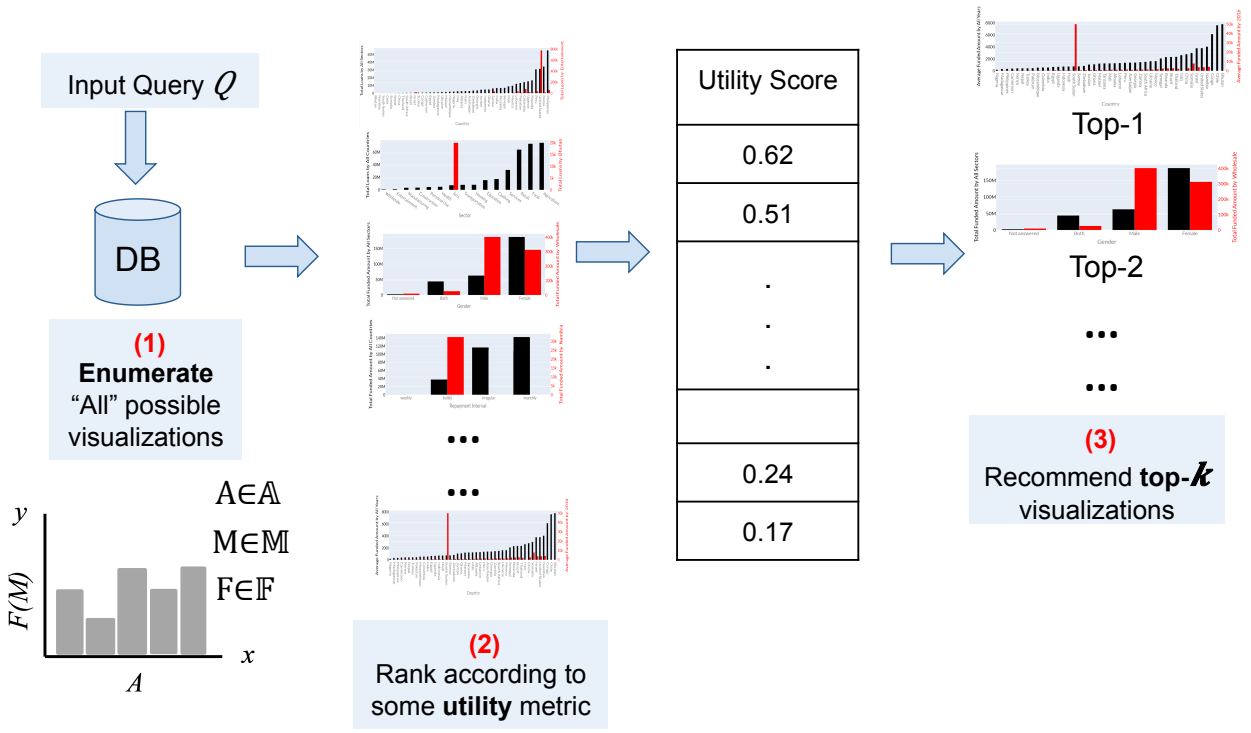


Figure 1. Ranking and Recommending Visualizations in LoanVis.

Kiva provides open public access to its data through daily snapshots and an Application Programming Interface (API). The Kiva data contains a set of heterogeneous information (i.e., data attributes) about the loans, lenders, borrowers, and field partners [34]. In particular, the main "objects" in the Kiva dataset are the *borrower*, the *lender*, and the *partner*, which are all connected to the *loan* object [1]. That is, the dataset is centered around the loan data object, namely the *"Loans"* table. Each loan listing would include information regarding the key details about that loan, such as the industry for which the loan is intended, together with information about the borrowers, as well as the loan financial information (e.g., amount, term, repayment interval, etc.). Table 1 provides a summary of the main attributes of the *Loans* dataset.

The utilized Kiva dataset contains information about more than 671,000 loans disbursed in 87 countries. A quick and simple exploration of the Kiva dataset can reveal some basic and interesting insights. For instance, out of the 15 funded sectors, projects related to the *Food* and *Agriculture* sectors receive the most loans. In terms of countries, borrowers from the Philippines and Kenya top the list of 87 countries for the total number of funded loans. Gender-wise, the data indicates that women make up the majority of borrowers, with 64% of borrowers being female.

Manually exploring the Kiva dataset and discovering some basic insights, similar to the ones mentioned above, has been the focus of multiple works (e.g., [2,29,30,33]). For instance, the work in [29] explores the Kiva dataset to understand the impact of the project sector on the distribution of loan amount (i.e., the relationship between attributes *Funded Amount* and *Sector* in Table 1). Similarly, the work in [30] attempts to understand the relationship between lending activity and features that characterize the loan, including the country of the loan, the loan sector, and the gender of the borrowers (i.e., attributes listed in Table 1 as *Country*, *Borrower Genders*, and *Sector*). In fact, Kiva provides its own data visualization and analytics dashboard [36], which allows users to explore the underlying crowdfunding data and facilitates conducting studies similar to the ones listed above. However, in this work, our goal goes beyond providing such basic statistics, and our focus is on leveraging visual analytics to automatically recommend data-driven, insightful visualization of the Kiva dataset, as described in the next sections.

Table 1. The Schema for the Kiva Dataset.

Attribute Name	Description
Borrower information	
Country	The name of the country in which the loan was disbursed.
Borrower Genders	Comma separated list of *Male*, *Female*, where each instance represents a single male/female in the group.
Loan Usage information	
Sector	High-level category of the loan usage field.
Activity	Granular category of the loan usage field.
Use	Exact Usage of loan amount.
Loan Dates	
Posted Time	The time at which the loan is posted on Kiva by the field agent.
Funded Time	The time at which the loan posted to Kiva gets funded by lenders completely.
Disbursed Time	The time at which the loan is disbursed by the field agent to the borrower.
Loan Amount	
Funded Amount	The amount disbursed by Kiva to the field agent (USD).
Loan Amount	The amount disbursed by the field agent to the borrower (USD).
Lender Count	The total number of lenders that contributed to this loan.
Loan Repayment	
Term in Months	The duration for which the loan was disbursed in months.
Repayment Interval	Loan repayment pattern - either monthly, irregular, or bullet (one time).

2.2. The LoanVis Visualization Recommendation System

The process of visual data exploration is typically initiated by an analyst specifying an exploratory query Q on a database D, as shown in Figure 1. The result of query Q, denoted as R, represents a subset of the database D, which the analyst can further transform into data visualizations that might reveal some interesting insights. For instance, an analyst exploring the Kiva dataset using our LoanVis system might pose some specific queries that are based on the following general query structure Q:

Q: `SELECT * FROM` D `WHERE T;`

Where in the exploratory query Q, D specifies the explored dataset (i.e., the Loans table) and T specifies a combination of predicates, which selects a subset of D for visual analysis.

For instance, an analyst who is reproducing the results in [29] might want to study the disparity in loan distribution for projects in the *Entertainment* sector, and in turn will pose a query Q, in which T is specified as: `sector = Entertainment`. Similarly, exploratory analysis of the other attributes and values of the Kiva dataset can be conducted using alternative settings of the predicate T (e.g., `gender = Female` or `country = Kenya AND sector = Food`, etc.).

A visual representation of the query Q is basically the process of generating different aggregate views of its result (i.e., R), which are then plotted using some of the popular data visualization representations (e.g., bar charts). An analyst typically examines those visualizations looking for insights, which are used as a springboard to decide their next analytical query. For instance, the results for a query in which T is `sector = Entertainment` might trigger further analysis based on gender, where an analyst would pose a subsequent query in which T is `sector = Entertainment AND gender = Female`.

Accordingly, we employ a multi-dimensional data model of D, which consists of a set of *dimension* attributes $\mathbb{A}$ (e.g., country, sector, etc.) and a set of *measure* attributes $\mathbb{M}$ (e.g., loan amount, lender count, etc.). Additionally, $\mathbb{F}$ is the set of possible aggregate functions over the measure attributes $\mathbb{M}$, such as `SUM`, `COUNT`, `AVG`, `MIN`, and `MAX`.

Hence, an aggregate view V_i over R is represented by a tuple (A, M, F) where $A \in \mathbb{A}$, $M \in \mathbb{M}$, and $F \in \mathbb{F}$, as shown in Figure 1. That is, R is grouped by dimension attribute A and aggregated by function F on measure attribute M.

A possible view V_i of the example query Q above would be expressed as:

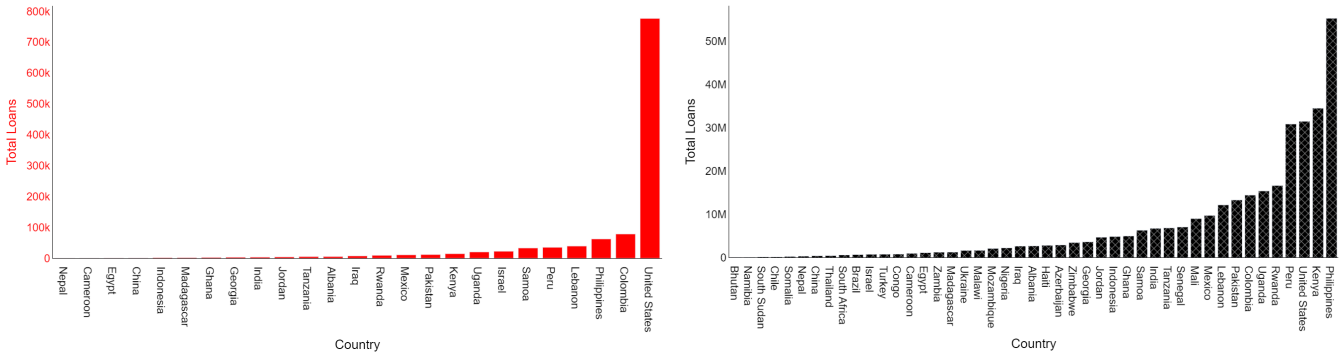
V_i: `SELECT A, F(M) FROM` D `WHERE T GROUP BY A;`

where the `GROUP BY` clause specifies the dimension A for aggregation, and $F(M)$ specifies both the aggregated measure M and the aggregate function F.

Clearly, there is a large number of possible aggregate views that can be generated from the results of each posed exploratory query Q. In fact, the number of those views/visualizations is equal to the number of all possible combinations of dimensions, measures, and aggregate functions. That is, equal to: $|\mathbb{A}| \times |\mathbb{M}| \times |\mathbb{F}|$. For instance, Figure 2a shows one of the possible visualizations of the result of the query Q above, in which `sector = Entertainment`. That visualization is equivalent to the following aggregate view V_i:

```
Vi: SELECT country, SUM (loan amount) FROM Loans
    WHERE sector = Entertainment;
    GROUP BY country;
```

Notice that to enhance the readability of a visualization, the user might include an `ORDER BY` clause in the view definition described above. For example, the visualization shown in Figure 2a is generated after extending V_i with the `ORDER BY SUM (loan amount)`. However, including such a clause would only change the order of the visualized bars, not the insights revealed by the visualization. Moreover, a `HAVING` clause might be considered to specify a condition that must be met by each group (i.e., bar) in the visualization. For instance, `HAVING SUM (loan amount) > $5000`. However, since in LoanVis those views are generated automatically, as we discuss in the next section, all the possible conditions for the `HAVING` clause must be considered during the view generation process. Clearly, there are an infinite number of such conditions, which would render the process of automated recommendation infeasible. Hence, a `HAVING` clause is excluded from our view generation model, and we adopt the basic aggregate view definition described above.



(**a**) Loans for the Entertainment Sector (**b**) Loans for All sectors combined

Figure 2. Target View (Entertainment sector) vs. Comparison View (All sectors).

2.2.1. Recommending Insightful Visualizations

Typically, a data analyst is keen to find visualizations that reveal some interesting insights about the analyzed data. For instance, to conduct studies similar to [2,29,30,33], an analyst would be exploring the Kiva dataset looking for visualizations that might reveal interesting discrepancies or anomalies in loan disbursement and distribution. That is, analysts need to manually construct a prohibitively large number of queries and visually explore their results looking for insights, which is clearly an ad-hoc and labor-intensive process. Particularly, the complexity of the manual visual data exploration process is contributed to: (1) the large number of possible visualizations, and (2) the uncertainty about the interestingness of each visualization. The challenges mentioned above motivated multiple research efforts that focused on *automatic recommendation* for visual data exploration. That is,

recommender systems that provide analysts with suggestions of interesting visualizations based on some objective, well-defined quantitative metrics (e.g., [21,24,26,27]).

For example, DeepEye is a visual insight recommendation system that employs a supervised machine learning approach to capture human perception by understanding existing examples [20]. QuickInsights [27] supports multiple types of data-driven insights for a comprehensive analysis (e.g., correlation, skewness in data distribution, diversity, etc.), and our work [37] studies the impact of data quality problems on discovering those insights. Meanwhile, SeeDB is one of the first visual insight recommendation systems that recommends top-k aggregate visualizations based on data-driven, deviation-based approach [19,21]. Other works that leverage a deviation-based approach include MuVE [23], which addresses binning problems in visualization recommendation systems. For further details, we refer the reader to a comprehensive recent survey on this topic [16].

In this work, and similar to several existing approaches (e.g., [21,23,27]), we adopt a *deviation-based* metric, which is able to provide analysts with interesting visualizations that highlight some of the particular patterns of the analyzed datasets . In particular, the deviation-based metric measures the distance between $V_i(R)$ and $V_i(D)$. That is, it measures the deviation between the aggregate view V_i generated from the subset data R vs. that generated from the entire database D. As such, $V_i(R)$ is denoted as *target* view (e.g., `sector = Entertainment`), whereas $V_i(D)$ is denoted as *comparison* view (e.g., `sector = ALL`). The premise underlying the deviation-based metric is that a view V_i that results in a higher deviation is expected to reveal some interesting insights that are very specific to the subset R and distinguish it from the general patterns in D. That is particularly important when exploring the Kiva dataset since the deviation-based metric is naturally able to capture and quantify anomalies and discrepancies in loan distribution, which has been one of the main focuses of existing work (e.g., [2,29,30,33]).

To ensure that all views are of the same scale, each target view $V_i(R)$ is normalized into a *probability distribution* $P[V_i(R)]$, and similarly, each comparison view into $P[V_i(D)]$. Particularly, consider an aggregate view $V = (A, M, F)$. A bar chart visualization of that aggregate view can be represented as the sequence of pairs: $<(a_1, f_1), (a_2, f_2), \ldots, (a_l, f_l)>$, where l is the number of distinct values (i.e., groups) in the dimension attribute A, a_i is the i-th group in attribute A, and f_i is the aggregated value $F(M)$ for the group a_i. For example, in Figure 2a, each a_i is a country, whereas each f_i is the amount of loans disbursed to that country a_i for projects related to the entertainment sector. Finally, V is scaled by the sum of aggregate values $U = \sum_{p=1}^{l} f_p$, leading to the probability distribution $P[V]$, which is computed as:

$$P[V] = < \frac{f_1}{U}, \frac{f_2}{U}, \ldots, \frac{f_l}{U} > \quad (1)$$

For an arbitrary view V_i (i.e., a specific combination of $V = (A, M, F)$), given the probability distributions of its target and comparison views (i.e., $P[V_i(R)]$ and $P[V_i(D)]$), the deviation $S(V_i)$ is computed as the *distance* between those probability distributions. Formally, for a given distance function *dist* (e.g., Euclidean distance, Earth Mover's distance, etc.), $S(V_i)$ is computed as:

$$S(V_i) = dist(P[V_i(R)], P[V_i(D)]) \quad (2)$$

In this work, we adopt *Euclidean* distance as our distance function. Hence, the deviation-based metric for a view V_i is computed as:

$$dist(P[V_i(R)], P[V_i(D)]) = \sum_{x=1}^{l} \sqrt{(P[V_i(R)]_x - P[V_i(D)]_x)^2}$$

Consequently, the deviation $S(V_i)$ of each possible view V_i is computed, and the k views with the highest deviation are recommended (i.e., *top-k*), as shown in Figure 1.

Illustrative Example: Consider a data analyst trying to gain insights into the loans disbursement to projects in the `Entertainment` sector. Particularly, the analyst poses an exploratory query:

Q: `SELECT * FROM loans WHERE sector = Entertainment;`

Clearly, the query Q above will return all the information about all the loans related to the `Entertainment` sector. Such information include different dimensions (e.g., country, repayment interval, gender, etc.), and different measures (e.g., loan amount, funded amount, lender count, etc.). Hence, the analyst can manually try creating different visualizations based on the different combinations of dimension and measure attributes, hoping that some of those visualizations would reveal interesting insights.

Alternatively, using our proposed LoanVis system, those insightful visualizations are quickly and automatically recommended to the analyst. In particular, LoanVis applies different SQL aggregate functions (i.e., $\mathbb{F}$) on the views resulting from all the possible pairwise combinations of dimensions and measures (i.e., $\mathbb{A}$ and $\mathbb{M}$), then the most interesting views are presented to the analyst (please see Figure 1). That is, the top-k views/visualizations with the highest deviation-based utility score, based on the user's setting for the value of k.

Figure 2a shows the *top-1* target view recommended by LoanVis based on the user input query Q. In particular, out of all the possible combinations of $\mathbb{A}$, $\mathbb{M}$, and $\mathbb{F}$, the view recommended by LoanVis is based on a visualization, in which: *A (x-axis) = country, M (y-axis) = loan amount, and F = SUM()*. Such view is equivalent to the following SQL query V_t.

V_t:
```
SELECT country, SUM (loan amount) FROM loans
WHERE sector = Entertainment;
GROUP BY country;
```

Essentially, LoanVis recommends that view shown in Figure 2a because it achieves the highest score according to our ranking utility function (i.e., the deviation-based metric). Specifically, the visualization based on the entertainment sector (the target view shown in Figure 2a) shows the highest deviation from the same visualization when generated for the aggregation of all sectors combined (the comparison view in Figure 2b), where the comparison view is equivalent to the following SQL query V_c:

V_c: `SELECT country, SUM (loan amount) FROM loans GROUP BY country;`

To help understand that recommendation, we combine the target view V_t and comparison view V_c in Figure 3, which reveals some very interesting observations regarding the disparity of loan distribution over the different sectors across different countries. As Figures 2b and 3 show, countries such as the Philippines, Kenya, Peru, Rwanda, Uganda, Colombia, Pakistan, Lebanon, Mexico, and Samoa are the ones that received the biggest share of total loans (summed over all the different sectors). However, when it comes to the specific loans to the entertainment sector, it is the United States of America that received most of the loans in that sector, at roughly $800,000. In fact, examining Figure 3 shows that the entertainment sector loans constitute about 2.67% of the total USA loans, whereas in the Philippines, that percentage drops to only 0.1%. That is, while projects related to the entertainment sector constitute a significant percentage in the USA, they are of lesser significance in other countries, where most loans are related to other sectors (e.g., agriculture, etc.).

2.2.2. Aspect-Based Recommendation

Notice that in the previous discussion, the analyst had to specify two inputs in their exploratory query: (1) an *aspect* for analysis (e.g., sector), and (2) a specific *value* within that aspect (e.g., entertainment). Such specification is realized using the exploratory query predicate T (e.g., `T: WHERE sector = Entertainment`). However, working with the Kiva dataset, we have learned firsthand that it is often challenging to specify those aspects and their corresponding values, which might eventually lead to some interesting visualizations being recommended!

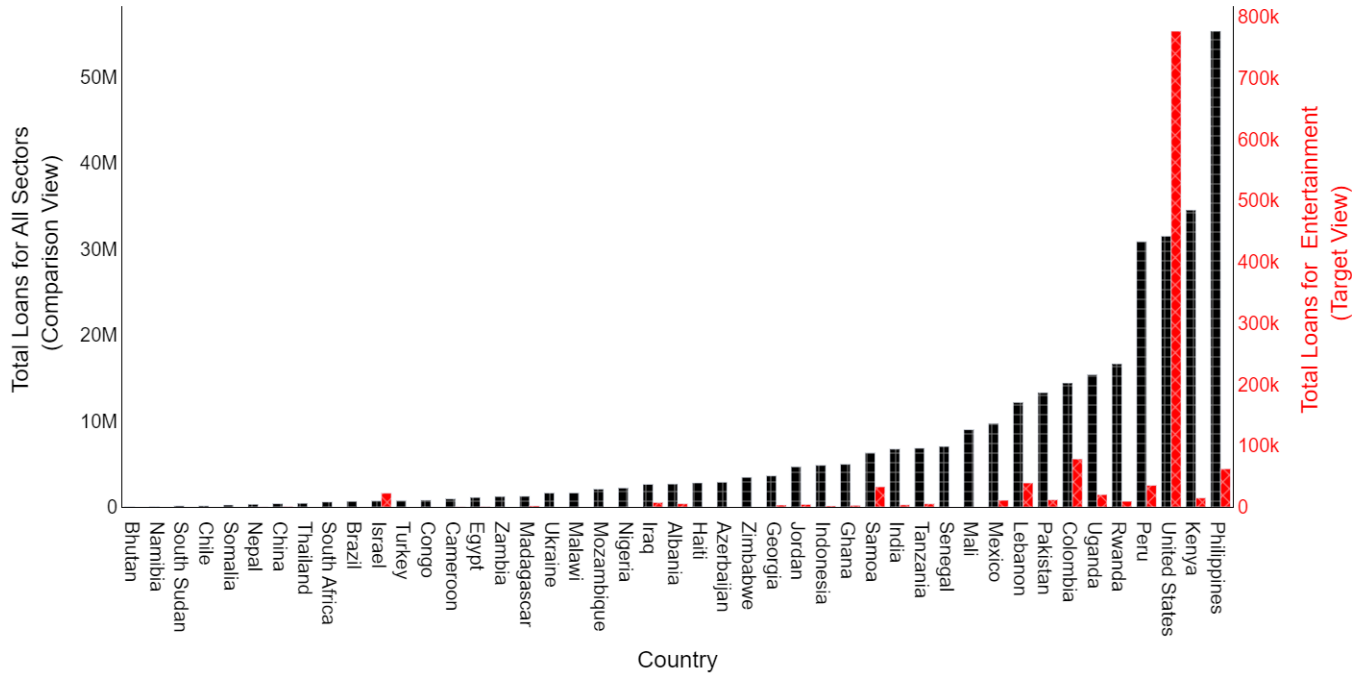


Figure 3. Deviation in loans distribution for all sectors (comparison view in black) vs. loans for the entertainment sector (target view in red).

For instance, an analyst might assume that an exploration based on `WHERE sector = education` or `WHERE sector = agriculture` would lead to some interesting recommended visualizations. However, during our analysis, we realized that all the possible visualizations, which are based on those two particular sectors, exhibit very low deviation, including the top-k ones. That is, there was nothing unique about the loans disbursed to those sectors, and their patterns followed the same pattern as that of the aggregated loans disbursed to all sectors. It was only after several rounds of experimenting that we discovered that the visualizations based on the `Entertainment` sector are the ones that reveal some interesting insights. Same for the visualizations based on the `Wholesale` and `Construction` sectors, which are presented in the next section.

However, current visualization recommendation systems (e.g., [18,26,27]) assume that the analyst is able to formulate a well-defined query that selects a subset of data, which leads to insightful visualizations being recommended (i.e., visualizations with a high utility score). That is, they are limited to only recommending interesting visualizations based on a precise exploratory query for which the analyst provides all the necessary query filters. Meanwhile, in reality, it is typically a challenging task to pose an exploratory query, which can immediately reveal some insights. Hence, it is a continuous process of trial and error, in which the analyst keeps refining their query filters manually and iteratively until some interesting visualizations are recommended. Therefore, in our design of LoanVis, we emphasize that, in addition to the existing techniques for automatically recommending interesting views, there is an equal need for additional techniques that can also automatically select subsets of data that would potentially provide such interesting views. Hence, our goal in this work is not only to recommend interesting visualizations but also to recommend exploratory queries that lead to such visualizations.

To achieve that goal, LoanVis expands and explores a larger search space of possible visualizations in order to recommend the top-k most insightful ones. In particular, we introduce the *aspect-based* recommendation, where an aspect could be any of the dimension or measure attributes of the analyzed dataset. More formally, in addition to the set of dimensions $\mathbb{A}$ and measures $\mathbb{M}$, we introduce the set of aspects $\mathbb{C}$, where $\mathbb{C} = \mathbb{A} \cup \mathbb{M}$.

Hence, for a given aspect $C \in \mathbb{C}$, LoanVis explores its distinct values searching for those that might result in high-utility visualizations. Particularly, for an aspect C (e.g.,

sector), which takes a set of distinct values: $c_1, c_2, \ldots$, LoanVis iterates through all the distinct values in C (Algorithm 1 line 1). Then, for each distinct value c_i, it generates all the possible visualizations that are based on selecting the subset of data that satisfies that value c_i (e.g., `Entertainment`) (Algorithm 1 lines 4–6). That process is repeated for all values in C, and the top-k visualizations with the highest deviation values are recommended to the analyst (Algorithm 1 line 10). That is, instead of recommending a visualization only in terms of the tuple (A, M, F), LoanVis expands the recommendation process and recommends (c, A, M, F), where c is a distinct value along an analyzed aspect C. For instance, Figure 3 shows a LoanVis recommendation, which is equivalent to the tuple $(Entertainment, Country, LoanAmount, SUM())$. Our detailed analysis presented in the next section is fully based on the aspect-based recommendation provided by LoanVis.

Algorithm 1 Aspect-based Recommendation

Input: $\mathbb{A}$, $\mathbb{M}$, C, k
1: **for** each $c_i \in C$ **do**
2: **for** each $A \in \mathbb{A}$ **do**
3: **for** each $M \in \mathbb{M}$ **do**
4: Generate target view V_t based on c_i, A, M, C
5: Generate comparison V_c view based on A, M, C
6: Calculate the deviation $dist(P[V_c], P[V_t])$
7: **end for**
8: **end for**
9: **end for**
10: Sort the generated views $\mathbb{V}$ based on their deviation score.
Output: Top-k views.

3. System and Results

We have conducted an extensive exploratory analysis of the Kiva dataset using our proposed LoanVis system. In this section, we present some of the visualizations recommended by our system LoanVis, together with some of the insights driven from those visualizations. The presented visualizations are the ones that received the highest utility score, according to the employed data-driven, deviation-based metric (please see Section 2.2).

Figure 4 shows a screenshot of LoanVis, which enables analysts to explore the Kiva dataset and recommends to the analysts visualizations that suit their exploratory analysis and are based on the deviation-based metric. LoanVis is developed using *Python* 3.7 under the *PyCharm* IDE. Our user interface is developed using the *Dash* package, which allows for the creation of an interactive web-based data application. Finally, the *Plotly* library is utilized for generating dynamic data visualizations.

As Figure 4 shows, LoanVis enables two forms of visual data exploration: (1) manual exploratory search and (2) automated recommendation-based exploration. Particularly, as shown in the top part of the interface, LoanVis allows analysts to specify parameters to manually construct different visualizations of the Kiva dataset across its different dimensions and measures. Alternatively, and as shown in the lower part of the interface, analysts can rely on LoanVis to automatically recommend insightful visualizations based on their selection of the explored aspects (e.g., sector, gender, etc.). In the rest of this section, we focus on the automated recommendations generated by LoanVis.

Notice that in this work we focus on the effectiveness of LoansVis. That is, the interesting insights discovered by LoanVis, whereas efficiency issues (i.e., query execution time) are beyond the scope of this work. Meanwhile, techniques for optimizing the query processing time of visualization recommendation systems have been proposed in some of our related work (e.g., [23–25]) and are directly applicable to our LoanVis system.

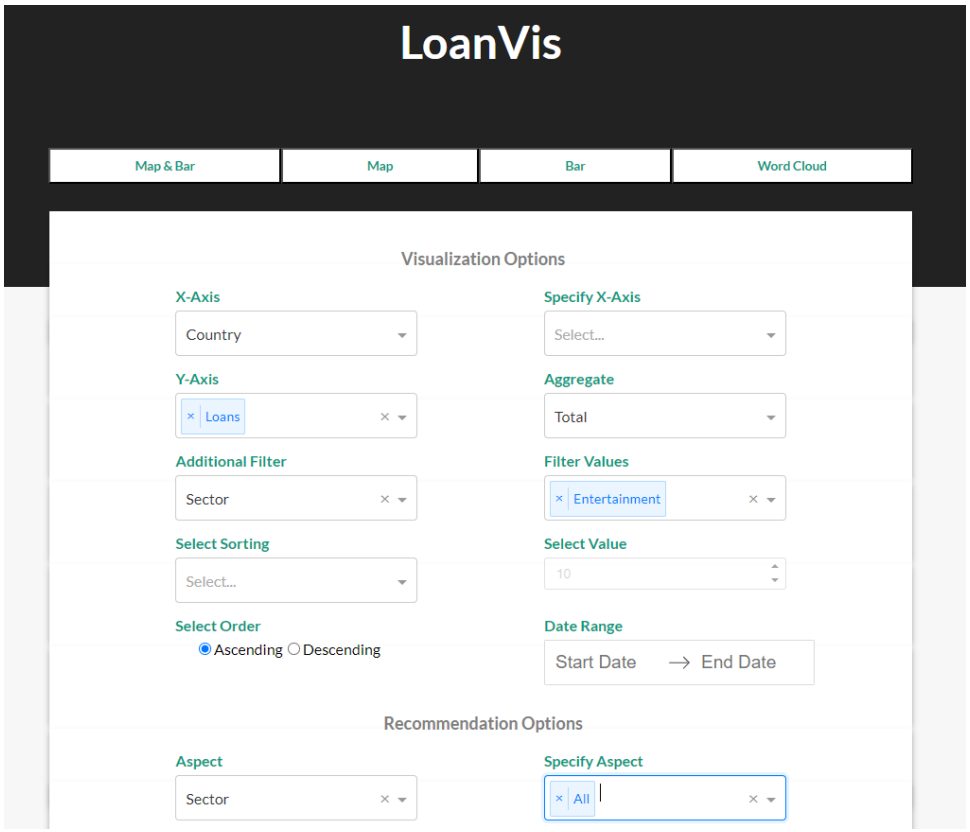


Figure 4. LoanVis: Data Exploration and Recommendation System—User Interface.

Table 2 presents a summary of all the results presented in this section (i.e., the recommended visualizations). For each result, the table shows the aspect of the Kiva dataset explored by the analyst, together with the different elements that constitute the corresponding visualizations recommended by LoanVis.

Particularly, for each recommended visualization, the table lists the following: (1) the aspect explored by the analyst (e.g., sector, country, etc.), (2) the particular value along that aspect recommended by LoanVis (e.g., *sector* = *entertainment*), (3) the dimension, measure, and aggregate function employed in the visualization recommended by LoanVis, and (4) the utility value of that recommended visualization. Notice that the maximum possible utility value for any visualization under the Euclidean distance measure is $\sqrt{2}$ [24].

Table 2. Summary of the LoanVis Recommendations.

Explored Aspect	Recommended Selection	Recommended Dimension	Recommended Measure	Recommended Aggregation	Figures	Deviation
Sector	Entertainment	Country	Loan Amount	SUM()	Figures 3 and 5	0.5761
Sector	Wholesale	Gender	Funded Amount	SUM()	Figure 6a	0.428
Sector	Construction	Gender	Lender Count	SUM()	Figure 6b	0.4204
Country	Namibia	Repayment Interval	Funded Amount	SUM()	Figure 7a	1.12
Country	Congo	Gender	Loans	Count()	Figure 7b	1.083
Year	2016	Country	Funded Amount	AVG()	Figure 8	0.4705
Year	2016	Country	Lender Count	AVG()	Figure 9	0.4903
Gender	Male	Repayment Interval	Loans	Count()	Figure 10	0.3046

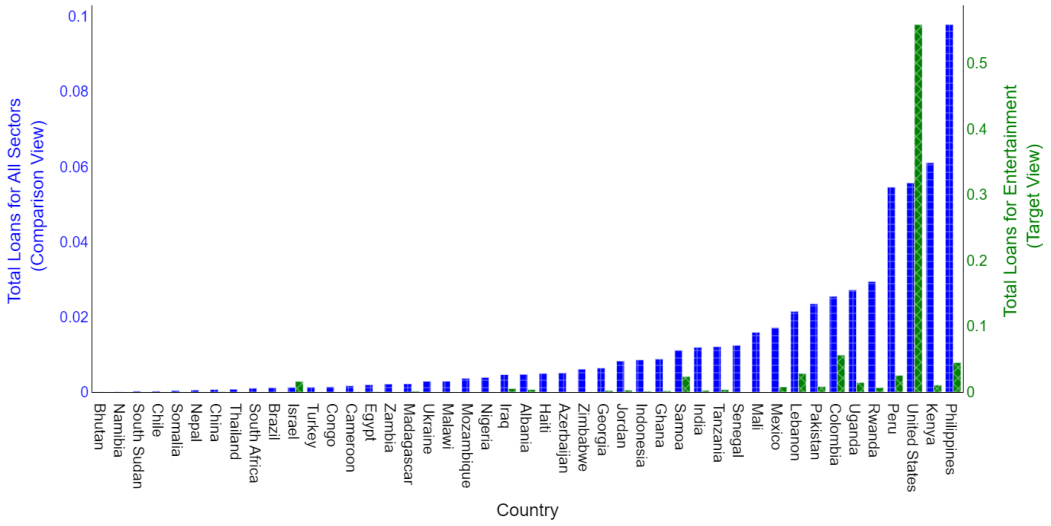


Figure 5. Normalized probability distribution in loans disbursed for all sectors (comparison view in blue) vs. loans for the entertainment sector (target view in green).

3.1. Automated Recommendations for the Sector Aspect

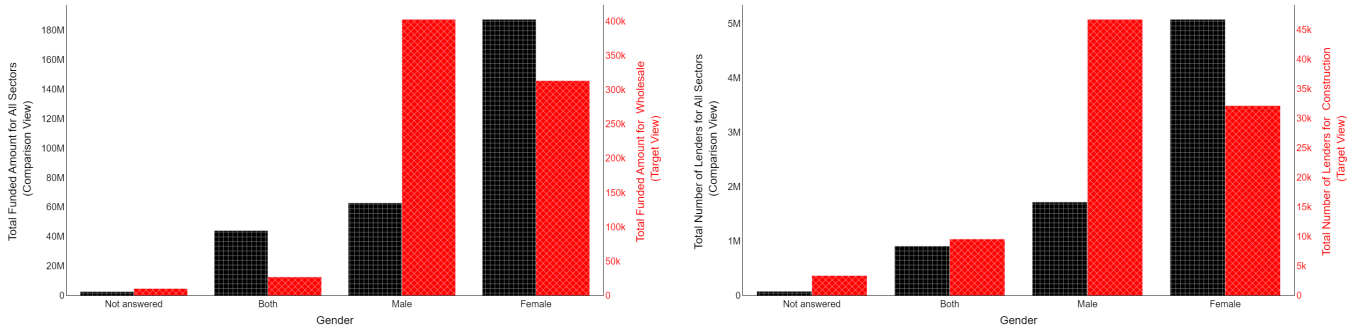
When exploring the *Sector* aspect of the Kiva dataset, the top-1 visualization recommended by LoanVis according to the employing deviation metric is the one based on the *Entertainment* sector, which has been discussed in the previous section and presented in Figure 3. Recall that the recommended visualization shows the significant discrepancy associated with the entertainment sector, where for most countries, loans provided for that sector are minimal, except for the USA.

In addition to Figure 3, that insight could be further understood by examining Figure 5. Particularly, Figure 5 shows the normalized distributions of the amount of loans for all sectors (i.e., comparison view) vs. the normalized distribution of loans directed to the entertainment sector (i.e., target view).

As expected, and as shown in the figure, the sum of all the normalized values in each of the target and comparison views adds up to 1.0 (please see Equation (1)). Further, the figure clearly emphasizes and clarifies the discrepancy highlighted earlier in Figure 3. For instance, while a country such as the Philippines receives almost 10% of the total kiva loans distributed worldwide, its share of the entertainment loans does not exceed 1%. In comparison, projects in the USA receive less than 6% of the total distributed loans, whereas its share of the loans directed to the entertainment sector is the highest among all countries at almost 10%. Similarly, as Figure 5 shows, `Israel` receives only 0.12% of the total loans worldwide but receives 3.12% of the entertainment loans.

While both Figures 3 and 5 deliver the same insights, Figure 3 utilizes two scales for plotting the absolute values on the Y-axis, whereas Figure 5 uses normalized values derived based on Equation (1). For the sake of simplicity and to enhance readability, all of the remaining visualizations are presented using absolute values, similar to Figure 3.

Figure 6a shows the top-2 visualization recommended by LoanVis along the *Sector* aspect. In contrast to the top-1 visualization, which is based on the *Country* dimension (Figure 3), this top-2 recommendation is based on the *Gender* dimension (Figure 6a). Particularly, Figure 6a shows the total amount of loans funded for projects in all sectors per gender (i.e., comparison view) vs. the loans funded in the specific `Wholesale` sector (i.e., target view). From Figure 6a, it is interesting to notice that, in general, female-led projects receive most of the funding (about \$200 M vs. \$60 M for male-led projects), as shown in the comparison view. However, for projects in the particular `Wholesale` sector, male-led projects seem to receive the higher share (about \$300 k for female-led wholesale projects vs. \$400 k for male-led ones).



(**a**) (**b**)

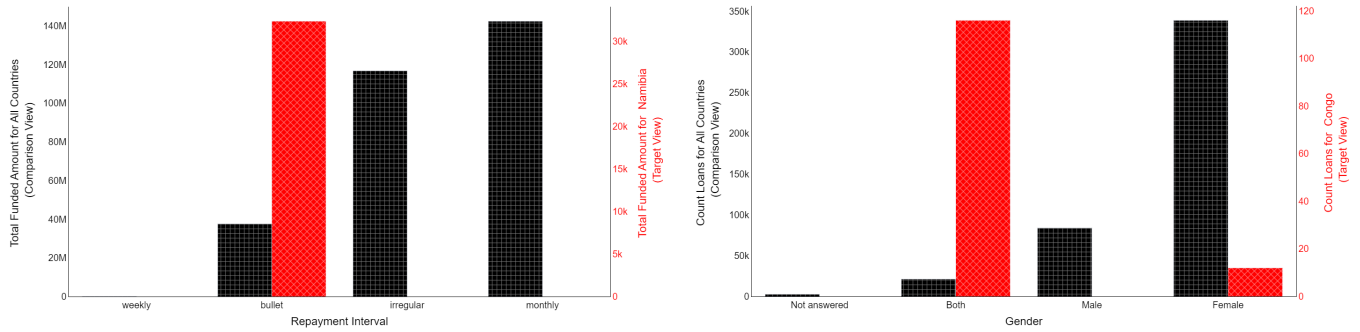
Figure 6. LoanVis top Recommendations for the *Sector* Aspect. (**a**) Deviation in loans distribution for All sectors vs. loans for `Wholesale` based on Gender. (**b**) Deviation in loans distribution for All sectors vs. loans for `Construction` based on Gender.

Interestingly, a similar discrepancy applies to projects in the `Construction` sector, which was the top-3 visualization recommended by LoanVis, and is shown in Figure 6b. In fact, during our analysis, we initially thought that male-led projects would receive the highest amount of loans in the `Construction` sector. However, as automatically discovered by LoanVis, that discrepancy is more pronounced in the `Wholesale` sector (Figure 6a), which scored a deviation value of 0.428, whereas the `Construction` sector (Figure 6b) came next with a deviation value of 0.4204.

3.2. Automated Recommendations for the Country Aspect

Figure 7a shows the top-1 recommended visualization for the *Country* aspect. That is, when the analyst utilizes LoanVis to recommend visualizations based on a country selection, LoanVis recommends selecting the country of `Namibia` and also recommends visualizing the distribution of the loans over the different repayment methods (as shown in Figure 7a). Particularly, the figure shows the number of loans paid under each of the different repayment methods. Interestingly, for all countries (i.e., the comparison view), monthly repayments are the most popular method for paying back the funded loans, followed by irregular repayments, then bullet repayments (i.e., paying back the loan all at once in full amount). However, for `Namibia` (i.e., the recommended target view), the pattern is completely different! Particularly, as the figure shows, all loans directed to projects in Namibia were paid back as bullet repayments! That significant discrepancy between how loans are paid for Namibian projects vs. the rest of the world led to that view achieving a high-deviation value of 1.12 and making it a top recommendation (recall that the maximum possible deviation under the Euclidean distance measure is $\sqrt{2}$).

Figure 7b shows another visualization that LoanVis recommended among the top ones during our analysis along the *Country* aspect. As it is already known from previous studies of the Kiva dataset, most funded projects are led by females, which is also confirmed by the general distribution of loans in the comparison view shown in Figure 7b. However, LoanVis automatically discovered an interestingly unique pattern for the country of `Congo`, which is shown as the target view in Figure 7b. Particularly, and differently from the general pattern, Figure 7b shows that in `Congo` the vast majority of funded projects are led collaboratively by both males and females! In fact, upon further analysis, we realized that those mixed-gender projects constitute 92% of the funded projects in `Congo`, while they constitute only 0.15% of the funded projects worldwide.



(**a**) (**b**)

Figure 7. LoanVis top recommendations for the *Country* aspect. (**a**) In Namibia (target view in red), most loans are payed as bullet repayments at once in contrast to the other worldwide prevailing repayment methods (comparison view in black). (**b**) In Congo (target view in red), most projects are led by mixed-gender teams vs. worldwide (comparison view in black); most projects are female-led.

3.3. Automated Recommendations for the Year Aspect

In this experiment, we focus on automatically generating recommendations based on the *Year* aspect. Figure 8 shows the top-1 visualization recommended by LoanVis for that aspect. Particularly, as the figure shows, LoanVis recommended the selection of `Year = 2016` and also recommended a visualization in which the Y-axis is the *Average Loan Amount* and the X-axis is the *Country*. Examining Figure 8, it shows the overall distribution of the average loan amount received by each country over all the recorded years (i.e., the comparison view, which is shown in black). However, the general overall distribution is significantly different in 2016, which has been recommended by LoanVis.

Specifically, looking at the distribution of loans in 2016 (i.e., target view, which is shown in red color), we notice that while the distribution of the average loan amount of most countries followed the same general pattern as in the comparison view, some discrepancies stand out, namely: (1) few countries did not receive any loans in 2016 (e.g., Bhutan, Chile, Congo, and Iraq), and (2) the country of South Sudan received loans in an average amount much higher than the loans it received in the other years. Digging deeper into that discrepancy, we realized that in 2015/2016, Kiva agreed to allow a one-year grace period for loans disbursed in `South Sudan` and also agreed to restructure repayment plans [38].

Interestingly, our observation was further emphasized by the top-2 visualization recommended by LoanVis, which is shown in Figure 9. In that recommended visualization, the comparison view shows the average number of lenders per project over the different years for all countries, whereas the target view shows the average number of lenders per project for all countries in 2016 (in red). Looking at the general distribution captured by the comparison view, we notice a skewed distribution, in which projects in Bhutan got the highest number of lenders per project (about 200 lenders per funded project), followed by Chile, Namibia, and Nigeria comes last in terms of average number of lenders per project. However, the distribution captured by the target view for 2016, while it still shows a skewed pattern, the details of that pattern are significantly different from the general one. Particularly, as Figure 9 shows, in 2016, the country with the maximum number of lenders per project was South Sudan, followed by Somalia, then Namibia. Noticing that projects in South Sudan received the highest number of lenders in 2016 (as shown in Figure 9) might provide a potential explanation for those projects receiving loans in high amount (as shown in Figure 8).

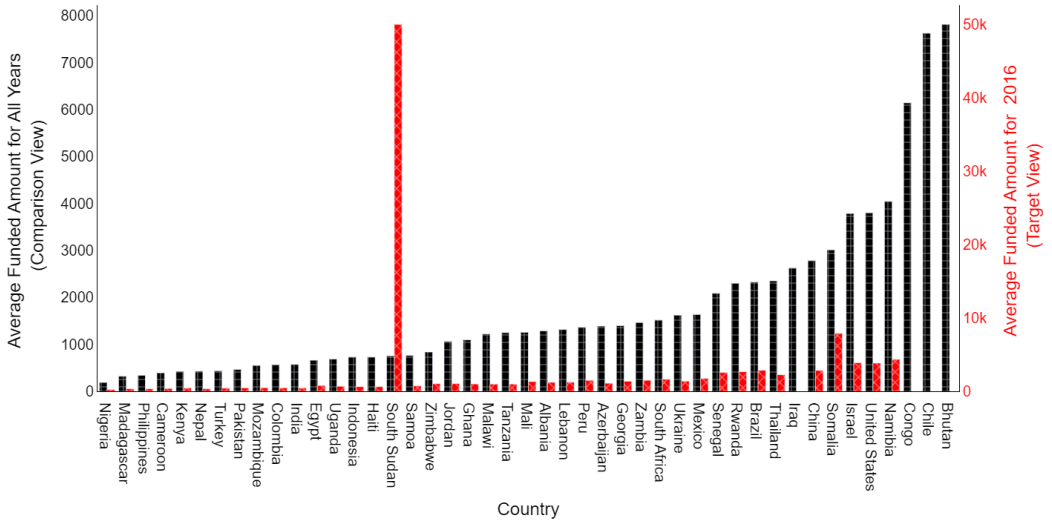


Figure 8. Deviation in loan distribution for all years (comparison view in black) vs. loans for 2016 (target view in red).

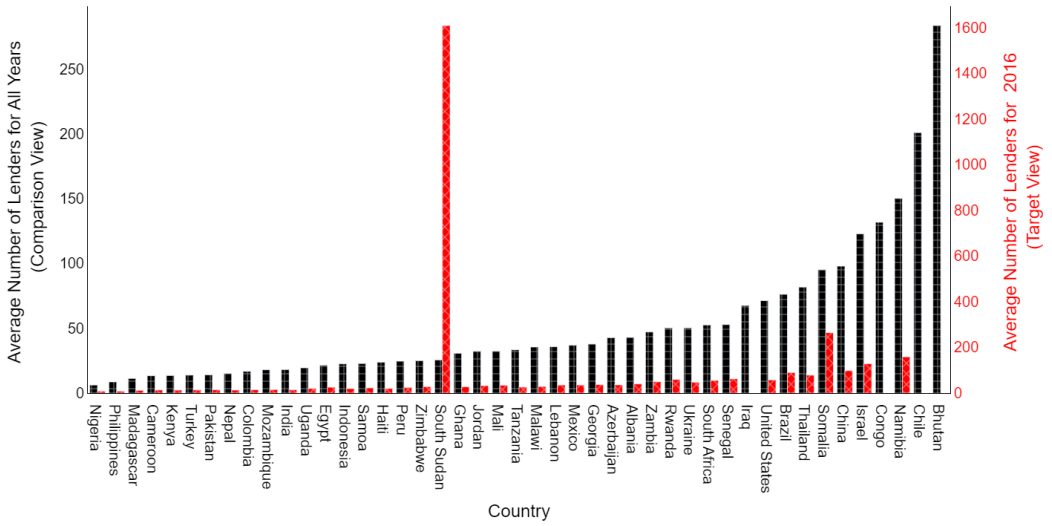


Figure 9. Deviation in number of lenders for all years vs. 2016.

3.4. Automated Recommendations for the Gender Aspect

In this last analysis, we examine generating recommendations based on the *Gender* aspect. Figure 10 shows LoanVis' recommendation, in which it automatically selected `Gender = Male` and also recommended the Y-axis to be the number of loans (i.e., `count()`) and the x-axis as the *Repayment Interval*. As the figure shows, in the general pattern (i.e., comparison view), `irregular` payment is the most popular method for paying back loans, followed by monthly, then bullet payment. However, the figure also shows that male-led projects exhibit a different pattern, in which `monthly` payment is the most popular, followed by bullet, then irregular. Discovering and studying that discrepancy could potentially assist decision-makers and lenders in structuring the repayment options for their loans.

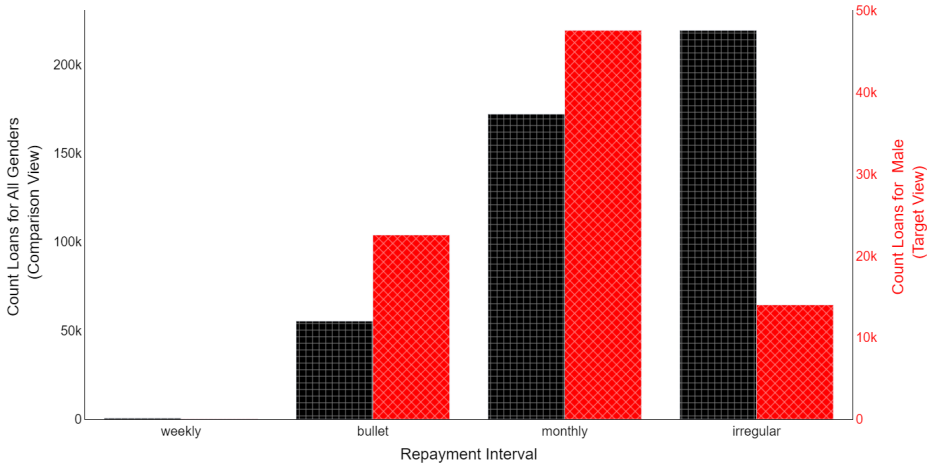


Figure 10. Distribution of repayment interval for all projects (comparison view) vs. male-led projects (target view).

4. Conclusions and Future Work

Motivated by the need for unlocking valuable insights from crowdfunding data, in this paper we propose our *LoanVis* solution for visualization recommendation. Unlike existing work that relies on manual data exploration for discovering insights from crowdfunding databases, LoanVis is an automated solution that utilizes objective metrics to quantify the utility of the recommended visualization. Our experimental evaluation demonstrated the effectiveness of LoanVis in recommending high-utility visualizations that reveal some interesting insights into some of the crowdfunding loan distribution patterns, based on the Kiva dataset.

Currently, LoanVis relies only on the deviation-based metric for capturing the interestingness of a visualization. To address that limitation in the future, in addition to the deviation-based measure, we plan to explore an expanded set of utility metrics and incorporate them into our LoanVis solution. Examples of such data-driven metrics include correlation, skewness in data distribution, diversity, etc., [26,37]. Moreover, we will investigate combining and integrating several of those data-driven utility metrics into hybrid multi-objective functions so that we can recommend visualizations that satisfy different requirements and expectations.

Furthermore, notice that while our data-driven approach has its clear advantages in visualization recommendation, it suffers from a lack of personalization, as recommendations may not be tailored to individual user preferences and needs. As such, in the future, we will investigate learning the user preference for visualization recommendations. That is, what makes a certain visualization interesting from the user perspective using ML-based classifier techniques. In turn, we also plan to conduct a real-world user study to further assess the effectiveness of our proposed LoanVis.

Author Contributions: Conceptualization, all authors; methodology, all authors; software, H.H., W.A., L.A., S.A. and F.A.; validation, all authors; formal analysis, all authors; investigation, all authors; resources, M.A.S. and N.Z.; data curation, H.H., W.A., L.A., S.A. and F.A.; writing—original draft preparation, all authors; writing—review and editing, all authors; visualization, all authors; supervision, M.A.S. and N.Z.; project administration, M.A.S. and N.Z.; funding acquisition, M.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by UAE University grant number G00004512.

Data Availability Statement: https://www.kiva.org/build/data-snapshots, accessed on 1 April 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhao, H.; Jin, B.; Liu, Q.; Ge, Y.; Chen, E.; Zhang, X.; Xu, T. Voice of charity: Prospecting the donation recurrence & donor retention in crowdfunding. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1652–1665.
2. Pham, T.T.; Shen, Y. A deep causal inference approach to measuring the effects of forming group loans in online non-profit microfinance platform. *arXiv* **2017**, arXiv:1706.02795.
3. Kim, M.J.; Hall, C.M.; Han, H. Behavioral influences on crowdfunding SDG initiatives: The importance of personality and subjective well-being. *Sustainability* **2021**, *13*, 3796. [CrossRef]
4. Zhao, H.; Ge, Y.; Liu, Q.; Wang, G.; Chen, E.; Zhang, H. P2P lending survey: Platforms, recent advances and prospects. *ACM Trans. Intell. Syst. Technol. (TIST)* **2017**, *8*, 1–28. [CrossRef]
5. Grant, S. Communicating Online Microfinance as an Effective Poverty Alleviation Tool: A Case Study of Kiva. Master's Dissertation, Malmo Universitet, Malmö, Sweden, 2018.
6. Moleskis, M.; Canela, M.A. *Crowdfunding Success: The Case of Kiva.Org*; IESE Business School Working Paper No. 1137-E; University of Navarra: Barcelona, Spain, 2016.
7. Lu, C.T.; Xie, S.; Kong, X.; Yu, P.S. Inferring the impacts of social media on crowdfunding. In Proceedings of the 7th ACM International Conference on Web Search and Data Mining, New York, NY, USA, 24–28 February 2014; pp. 573–582.
8. Zhao, H.; Zhang, H.; Ge, Y.; Liu, Q.; Chen, E.; Li, H.; Wu, L. Tracking the dynamics in crowdfunding. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 625–634.
9. An, J.; Quercia, D.; Crowcroft, J. Recommending investors for crowdfunding projects. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Republic of Korea, 7–11 April 2014; pp. 261–270.
10. Zhang, H.; Zhao, H.; Liu, Q.; Xu, T.; Chen, E.; Huang, X. Finding potential lenders in P2P lending: A hybrid random walk approach. *Inf. Sci.* **2018**, *432*, 376–391. [CrossRef]
11. Liu, J.; Xiao, Y.; Zheng, W. LCW: A Lightweight Recommendation Framework for Non-profit Crowdfunding Projects. In Proceedings of the 4th International Conference on Computer Science and Software Engineering, Singapore, 22–24 October 2021; pp. 238–242.
12. Zhuang, K.; Wu, S.; Liu, S. CSRLoan: Cold Start Loan Recommendation with Semantic-Enhanced Neural Matrix Factorization. *Appl. Sci.* **2022**, *12*, 13001. [CrossRef]
13. Andrienko, G.; Andrienko, N.; Drucker, S.M.; Fekete, J.D.; Fisher, D.; Idreos, S.; Kraska, T.; Li, G.; Ma, K.L.; Mackinlay, J.D.; et al. Big data visualization and analytics: Future research challenges and emerging applications. In Proceedings of the 3rd International Workshop on Big Data Visual Exploration and Analytics, Copenhagen, Denmark, 30 March 2020.
14. El, O.B.; Milo, T.; Somech, A. Towards Autonomous, Hands-Free Data Exploration. In Proceedings of the Conference on Innovative Data Systems Research (CIDR), Amsterdam, The Netherlands, 12–15 January 2020.
15. Luo, Y.; Qin, X.; Chai, C.; Tang, N.; Li, G.; Li, W. Steerable self-driving data visualization. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 475–490. [CrossRef]
16. Qin, X.; Luo, Y.; Tang, N.; Li, G. Making data visualization more efficient and effective: A survey. *VLDB J.* **2020**, *29*, 93–117. [CrossRef]
17. Ma, P.; Ding, R.; Han, S.; Zhang, D. MetaInsight: Automatic Discovery of Structured Knowledge for Exploratory Data Analysis. In Proceedings of the International Conference on Management of Data, Virtual Event, 20–25 June 2021; pp. 1262–1274.
18. Sharaf, M.A.; Ehsan, H. Efficient query refinement for view recommendation in visual data exploration. *IEEE Access* **2021**, *9*, 76461–76478.
19. Vartak, M.; Rahman, S.; Madden, S.; Parameswaran, A.; Polyzotis, N. Seedb: Efficient data-driven visualization recommendations to support visual analytics. In Proceedings of the VLDB Endowment International Conference on Very Large Data, Kahola Coast, HI, USA, 31 August–4 September 2015; Volume 8, p. 2182.
20. Luo, Y.; Qin, X.; Tang, N.; Li, G. Deepeye: Towards automatic data visualization. In Proceedings of the IEEE 34th International Conference on Data Engineering (ICDE), Paris, France, 16–19 April 2018; pp. 101–112.
21. Vartak, M.; Madden, S.; Parameswaran, A.; Polyzotis, N. SEEDB: Automatically generating query visualizations. *Proc. VLDB Endow.* **2014**, *7*, 1581–1584. [CrossRef]
22. Vartak, M.; Huang, S.; Siddiqui, T.; Madden, S.; Parameswaran, A. Towards visualization recommendation systems. *ACM Sigmod Record* **2017**, *45*, 34–39. [CrossRef]
23. Ehsan, H.; Sharaf, M.A.; Chrysanthis, P.K. MuVE: Efficient multi-objective view recommendation for visual data exploration. In Proceedings of the IEEE 32nd International Conference on Data Engineering (ICDE), Helsinki, Finland, 16–20 May 2016; pp. 731–742.
24. Ehsan, H.; Sharaf, M.A.; Chrysanthis, P.K. Efficient recommendation of aggregate data visualizations. *IEEE Trans. Knowl. Data Eng.* **2017**, *30*, 263–277. [CrossRef]
25. Ehsan, H.; Sharaf, M.A. Materialized view selection for aggregate view recommendation. In Proceedings of the 30th Australasian Databases Theory and Applications Conference, Sydney, NSW, Australia, 29 January–1 February 2019; pp. 104–118.
26. Demiralp, Ç.; Haas, P.J.; Parthasarathy, S.; Pedapati, T. Foresight: Recommending Visual Insights. *Proc. VLDB Endow.* **2017**, *10*, 1937–1940. [CrossRef]

27. Ding, R.; Han, S.; Xu, Y.; Zhang, H.; Zhang, D. Quickinsights: Quick and automatic discovery of insights from multi-dimensional data. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Amsterdam, The Netherlands, 30 June–5 July 2019; pp. 317–332.
28. Zhang, X.; Ge, X.; Chrysanthis, P.K.; Sharaf, M.A. Viewseeker: An interactive view recommendation framework. *Big Data Res.* **2021**, *25*, 100238. [CrossRef]
29. Sarkar, S.; Alvari, H. Mitigating bias in online microfinance platforms: A case study on Kiva.org. In Proceedings of the ECML PKDD 2020 Workshops, Ghent, Belgium, 14–18 September 2020; pp. 75–91.
30. Paruthi, G.; Frias-Martinez, E.; Frias-Martinez, V. Peer-to-peer microlending platforms: Characterization of online traits. In Proceedings of the IEEE International Conference on Big Data, Washington, DC, USA, 5–8 December 2016.
31. Austin, T.; Rawal, B.S. Model Retraining: Predicting the Likelihood of Financial Inclusion in Kiva Peer-to-Peer Lending to Promote Social Impact. *Algorithms* **2023**, *16*, 363. [CrossRef]
32. Available online: https://www.tableau.com/solutions/gallery/kiva-loan-story (accessed on 1 April 2024).
33. Paruthi, G.; Frias-Martinez, E.; Frias-Martinez, V. *Understanding Lending Behaviors on Online Microlending Platforms: The Case for Kiva*; Association for the Advancement of Artificial Intelligence: Washington, DC, USA, 2015.
34. Choo, J.; Lee, C.; Lee, D.; Zha, H.; Park, H. Understanding and promoting micro-finance activities in kiva.org. In Proceedings of the 7th ACM International Conference on Web Search and Data Mining, New York, NY, USA, 24–28 February 2014; pp. 583–592.
35. Burtch, G.; Ghose, A.; Wattal, S. Cultural differences and geography as determinants of online prosocial lending. *MIS Q.* **2014**, *38*, 773–794. [CrossRef]
36. Available online: https://www.kiva.org/team/data_analysis/graphs (accessed on 1 April 2024).
37. Mafrur, R.; Sharaf, M.A.; Zuccon, G. Quality matters: Understanding the impact of incomplete data on visualization recommendation. In Proceedings of the International Conference on Database and Expert Systems Applications, Bratislava, Slovakia, 14–17 September 2020; pp. 122–138.
38. Available online: https://www.kiva.org/about/where-kiva-works/partners/206 (accessed on 1 April 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.