School of Information Technology and Engineering at the ADA University

School of Engineering and Applied Science at the George Washington University

AN EXPERT-BASED APPROACH FOR RECOMMENDING DATA VISUALIZATION TECHNIQUES

A Thesis
Presented to the Graduate Program of Computer Science and Data Analytics
of the School of Information Technology and Engineering
ADA University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Computer Science and Data Analytics
ADA University

By
Kanan Gafarov

April, 2025

THESIS ACCEPTANCE

This Thesis by: Kanan Gafarov
Entitled: *An Expert-Based Approach for Recommending Data Visualization Techniques*

has been approved as meeting the requirement for the Degree of Master of Science in Computer Science and Data Analytics of the School of Information Technology and Engineering, ADA University.

Approved:

Dr. Stephen Kaisler		29/04/2025
(Adviser)		(Date)
Dr. Samir Rustamov.		29/04/2025
(Program Director)		(Date)
Dr. Abzatdin Adamov		29/04/2025
(Dean)		(Date)

ABSTRACT

In the era of data-driven decision-making, effective visualization of complex datasets plays a pivotal role in enhancing analytical insights and communication. Yet, selecting the most appropriate data visualization technique remains a significant challenge for users, especially those without formal training in data analytics or visual design. This research addresses that gap by developing a rule-based expert recommendation system that guides users in selecting suitable visualization techniques based on specific dataset characteristics. The proposed system integrates expert knowledge, formalized rules, and a user-centric interface to support and streamline the visualization selection process.

The system architecture follows a Model–View–Controller (MVC) design pattern, with a robust backend built using Python and MongoDB. The Model stores a curated database of visualization techniques along with associated rules and attributes, the Controller implements a dynamic decision engine that interprets user responses, and the View presents an interactive Q&A interface built using Streamlit. The questionnaire comprises 8–10 targeted questions that gather key information about the dataset—such as data type, structure, analytical goal, and data volume—which are then evaluated by the system's reasoning engine to generate tailored recommendations.

A unique feature of this system is its expert rule base, manually constructed based on visualization theory, best practices, and authoritative resources. The database of visualization techniques was also built from scratch, requiring extensive research and careful structuring of technique-specific metadata. The rules were refined through iterative testing and feedback to ensure accuracy and contextual relevance across a variety of datasets.

The system was evaluated through a formative usability study with five participants, who tested the prototype on real datasets and provided qualitative feedback. Results indicated high usability, with participants praising the intuitive question flow, clarity of recommendations, and educational value. Users found that the system either confirmed their expectations or introduced new and contextually relevant visualization ideas. The rule-based recommendations were transparent and interpretable, contributing to user confidence and improved visualization literacy.

This research contributes a novel and practical solution to the problem of visualization technique selection by embedding expert knowledge in a usable and accessible form. It has implications for both novice users seeking guidance and experienced analysts looking for validation or inspiration. Future enhancements may include expanding the visualization database, improving rule transparency, and deploying the system as a web-based application. Overall, the system demonstrates the feasibility and value of expert-based, rule-driven approaches in supporting effective data visualization.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Explanation
AI	Artificial Intelligence
API	Application Programming Interface
DSR	Design Science Research
GUI	Graphical User Interface
JSON	JavaScript Object Notation
MVC	Model–View–Controller
NLP	Natural Language Processing
NoSQL	Not Only SQL (non-relational database model)
Q&A	Question and Answer
RDBMS	Relational Database Management System
UI	User Interface
UX	User Experience
XML	eXtensible Markup Language
TF	Term Frequency
IDF	Inverse Document Frequency
TFBC	Term Frequency Bar Chart

1 INTRODUCTION

The ability to visualize data effectively has become essential for making informed decisions in our data-driven world. Visualizations connect raw data to actionable insights in all sectors including business analytics, healthcare, education and scientific research. Visualizations reduce complicated information into simpler forms which reveal important patterns to both technical and non-technical audiences. Users face challenges when trying to select the best visualization method for their particular data characteristics and analytical requirements.

The number of visualization techniques available creates an additional challenge as the selection process becomes more complex. The various visualization methods including bar charts and network diagrams possess unique advantages and limitations which depend on their specific application context. The research aims to develop a system which guides users toward selecting the most appropriate visualization methods for their data.

1.1 Problem Statement

The process of data visualization transforms raw data into meaningful insights through the creation of clear and easy-to-understand visual representations of complex information. Through effective visualization techniques users can detect patterns and relationships in data which enables them to make better decisions across business, healthcare and scientific research fields. The selection of suitable visualization methods proves difficult for users who lack experience with the extensive range of available visualization tools.

Selecting an inappropriate representation of data could greatly impair the ability to obtain correct inferences to inform decisions from the data. If the form of representation doesn't match the level of complexity or nature of the data, you could end up with misrepresentation, or cognitive overload or false impressions. For example, consider a pie chart representing a few categories that have similar but slightly different values. If the respective slices are visually indistinguishable, then any meaning extracted from the intended comparison would be lost. By using a pie chart, we have lost the clarity of comparison between categories, and at this point our likelihood of drawing false impressions increase. Clearly defined categories along the same axis with a bar chart would allow the same comparison, but the visual comparison of magnitudes would be far more precise. This illustrates how careful evaluation of the appropriate choice of graphical data representation should be attended to in a manner which represents the structure of the data and conforms to the intent of the analysis.

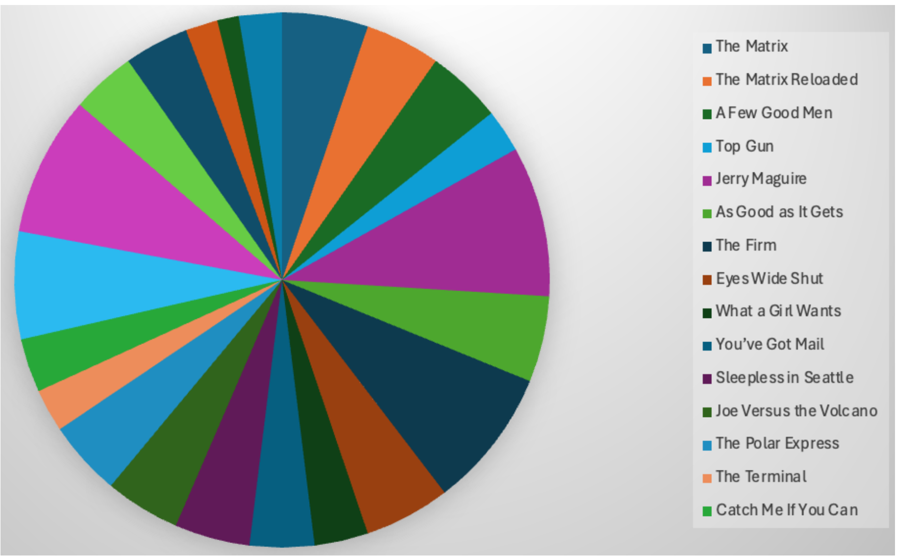

Figure 1. Example of a poor choice for data visualisation.

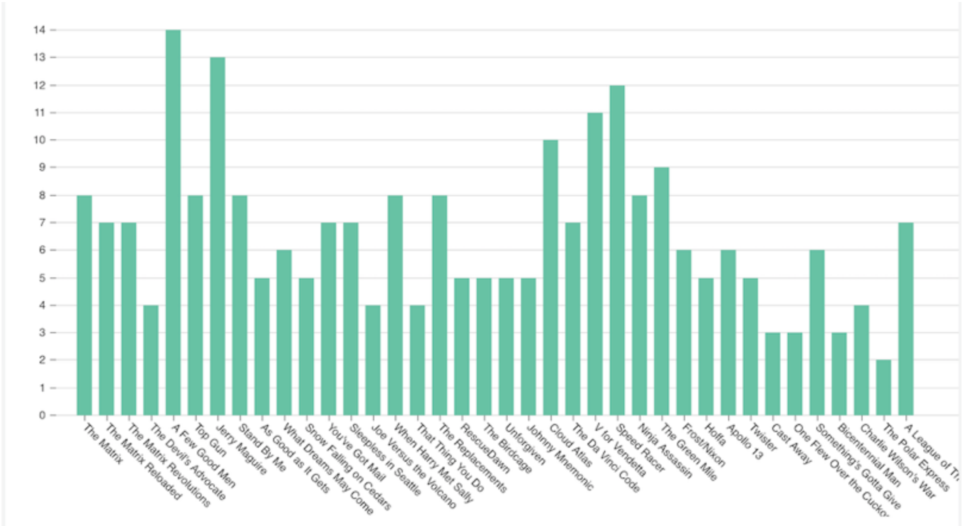

Figure 2. Example of a good choice for data visualization.

The research develops a Data Visualization Technique Recommendation System to assist users in discovering suitable visualization approaches for their datasets. The system employs a rule-based decision engine together with a structured database of visualization techniques and an interactive Q&A feature to assist users in making knowledgeable choices. The system combines expert guidelines with an easy-to-use interface to help users better understand complex data and achieve effective visualization of important insights.

The system provides users with a tool to pick visualization methods that match their dataset's specific characteristics. The system utilizes an object-oriented framework which relies on a rule-based reasoning engine that uses distinct rule IDs to validate the recommendations. The system builds confidence in its recommended visualization techniques by accumulating rule IDs that support each recommendation until it can confirm these methods represent the most suitable visualization approaches according to the decision rules.

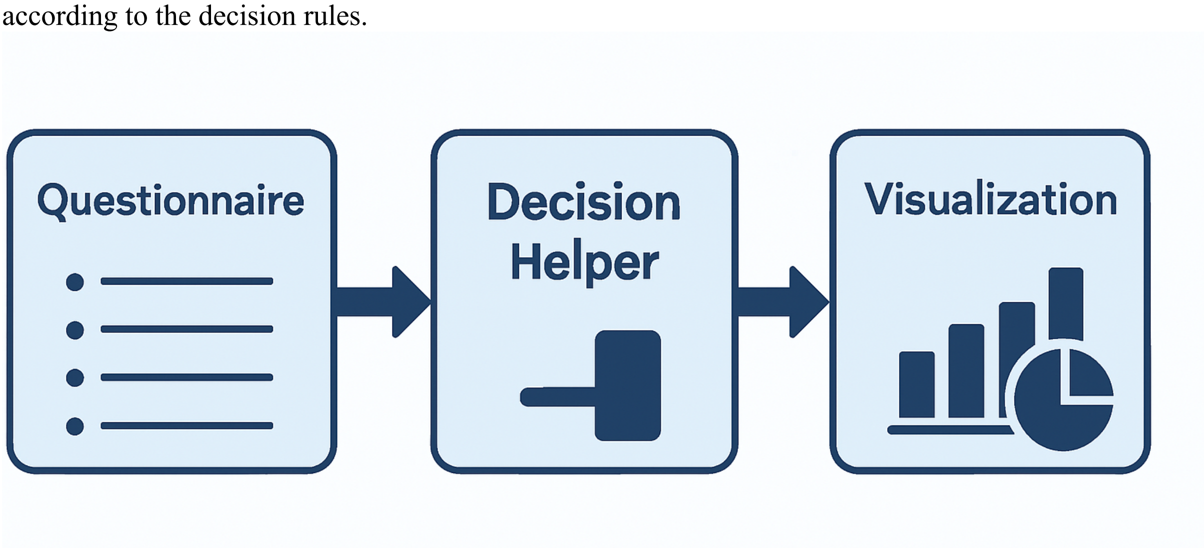

Figure 3. Workflow of the Visualization Recommendation Process.

1.2 Motivation

Data visualization continues to expand across different fields yet many users who lack analytics training face challenges when choosing suitable visualization methods [1] [2]. The selection of incorrect methods results in both failed data communication and incorrect data understanding. Users must dedicate time to read documentation while making several attempts before they can use current tools effectively.

The rule-based recommendation system addresses the current research gap by creating a simpler process. The system relies on established guidelines [3] [4] together with user-provided data characteristics to create actionable recommendations which enhance data visualization clarity and impact. The system produces dataset-specific recommendations through its rule-based decision engine and its Q&A process structure. The research findings will improve both data literacy and visualization technique usage which will result in enhanced data analysis and communication abilities.

1.3 Personal Motivation

My research stems from direct observations of data visualization obstacles which affect people who lack experience with multiple visualization approaches. The selection of inappropriate visualization methods results in unclear or misleading data presentations which fail to deliver meaningful insights. The problems I witnessed during my experience drove me to create a solution which would simplify the process while providing structure and easy accessibility.

My academic background in data analytics drives my passion to create new tools which enhance data-driven decision-making processes. The research objective supports my mission to simplify data visualization so users at every skill level can successfully present and understand their data. This project uses a rule-based reasoning engine to create a user-friendly recommendation system which adjusts its suggestions through user input to enable complete data visualization capabilities that produce superior analytical results across different fields.

1.4 Research Questions

Three research questions have been identified that will guide this research effort:

1) What is the best way to design a system that uses object-oriented programming and a rule-based system to store all the data visualization techniques in the database? This question examines the creation of a general data structure (e.g., classes and objects) that can handle and store different visualization techniques and their properties. The emphasis is on the system being able to handle a large number of techniques and their properties and the use of rule-based reasoning to aid the decision-making process.

2) How can the user interface (UI) be designed and implemented to make it easy for users to interact with the system and receive accurate recommendations for visualization techniques? This question deals with the development of a user-friendly interface that assists users in providing their data characteristics and options and ensures a smooth process that results in good recommendations. The focus is on UI/UX principles that facilitate seamless interaction and minimize user effort. The system will also include references to established visualization guidelines to support its recommendations.

3) How can a rule-based modeling approach be effectively implemented in the Q&A system to analyze user input and recommend the most suitable data visualization technique(s)? This question is on the development of a well-defined rule-based system that can understand the user's responses to the questions asked, assess the features of the data provided, and recommend one or more suitable visualization techniques.

1.5 Significance of Research

This research fills an essential void in data visualization research by providing an intelligent guided system to determine appropriate visualization methods based on dataset characteristics. The modern data-centric environment requires effective information visualization as an essential business necessity. The lack of domain knowledge and visualization literacy among users leads to inappropriate visual formats which produces poor or misleading representations. The research establishes a rule-based recommendation system to help users select visualization methods based on their specific context.

The system achieves its importance through its power to make data visualization accessible to everyone. The system provides advanced visualization guidance through a rule-driven engine combined with an intuitive Q&A interface which enables students and business analysts and educators and healthcare professionals who lack extensive data analysis experience to access expert knowledge. The system provides both simplified learning and improved visualization clarity and effectiveness for all domains.

The research demonstrates how expert knowledge and structured data can be embedded into user-facing tools through object-oriented design and reasoning mechanisms for intelligent decision support systems. The system provides a flexible framework which accepts additional visualization methods and rules and dataset attributes for future development. The model provides a starting point for integrating into bigger systems including data analytics platforms and educational tools that teach data communication effectiveness.

This research shows promise to enhance data interpretation while building user confidence in data-driven conclusions and improving decision-making through dataset property-based visualization alignment in a guided and reliable system.

1.6 Definition of Terms

To ensure clarity and consistency throughout this research, the following key terms are defined:

- **Data Visualization:** The graphical representation of data and information using visual elements such as charts, graphs, and maps. Visualization techniques help users identify patterns, trends, and outliers in large datasets.
- **Visualization Technique:** A specific method used to present data visually. Examples include pie charts, bar charts, scatterplots, heatmaps, and network diagrams. Each technique is suitable for different types of data and analytical goals.
- **Recommendation System:** A system designed to suggest the most relevant items or actions to a user based on input preferences or data. In this research, it refers to a rule-based system that suggests visualization techniques based on dataset characteristics.
- **Rule-Based System:** A decision-making system that applies a set of predefined logical rules to infer conclusions or make recommendations. Each rule in this system links data attributes to suitable visualization techniques.
- **Object-Oriented Approach:** A software design paradigm based on the concept of "objects," which encapsulate data and behavior. It is used in this research to represent and manage various visualization techniques and their associated attributes.
- **User Interface (UI):** The part of the system that allows users to interact with the recommendation engine, input data characteristics, and receive suggestions. A well-designed UI improves usability and system effectiveness.
- **Data Characteristics**: Properties of a dataset that influence visualization choices, such as data type (numerical, categorical, textual), dimensionality (univariate, multivariate), data distribution, and analytical intent.
- **Expert Guidelines:** Established best practices and heuristics derived from visualization literature or domain experts that guide the choice of appropriate visualization methods for specific data scenarios.
- **Q&A System:** An interactive mechanism that presents users with a sequence of questions to understand their dataset and context. The responses are used by the decision engine to determine suitable visualization techniques.
- **Visualization Literacy:** The ability to read, interpret, and create data visualizations. This research aims to support users with varying levels of visualization literacy by offering structured and guided recommendations.

2 LITERATURE REVIEW

The rapid increase of data throughout recent years created an urgent demand for better visualization methods to extract meaningful information. The number of potential visual encoding designs for any given dataset extends into a vast design space [3]. Through visualization we utilize human perceptual strengths to interpret data because humans excel at detecting patterns together with trends and unique data points. Selecting an appropriate visualization requires both data characteristics understanding and graphic design principles knowledge. Research in graphical perception demonstrates spatial position encoding produces more precise quantitative value decoding than color or area encoding methods. The optimal visualization for any dataset requires both knowledge about data types and human perceptual effectiveness. Two classic works establish essential principles for creating effective graphical designs [5] [1]. The author in [6] explains that data-ink ratio maximization requires the reduction of unnecessary ink while keeping data ink at a maximum level and the elimination of decorative elements that fail to enhance comprehension. Truthful data is essential in representation through proportional graphical encodings as well as clear visual communication [6]. [1] advocates for quantitative patterns in analysis through basic visual forms that he provides specific practical guidelines for exploratory visualization. The visualization should maintain its focus on the main message and target audience while selecting a technique that creates a simple narrative from the data [2]. These principles serve as foundational best practices that support various visualization recommendation methods.

Modern visualization tools receive major influence from [7]. The fundamental principal stresses how interactive exploration helps users achieve better results [7]. According to this principle, users gain maximum benefit when they start with a general view of data before moving to specific elements. A visualization recommendation system should present users with basic visualization choices first through automatic methods (such as producing an overview visualization or displaying various possible views) before enabling further customization. The effectiveness of a recommendation system requires both providing the optimal chart selection and enabling users to explore different views through quick comparison and iteration. It is important to mention that simple graphics work well for many cases, yet complex datasets need exotic visualizations and users need assistance to navigate the diverse visualization options available [3]. A sound visualization recommendation system needs to incorporate established design principles for view effectiveness and enable interactive analysis through multiple view comparison.

The principles of recommender system research offer valuable guidance to develop methods for recommending visualization techniques. The three primary categories of traditional recommender systems include content-based, collaborative and knowledge-based approaches. The two filtering methods of content-based and collaborative rely on user past preferences, but visualization lacks explicit ratings and every dataset and task remains distinct. The visualization recommendation approach shares similarities with knowledge-based recommenders because it uses "semantic user preference knowledge, item (data) knowledge, and recommendation knowledge" to provide visualizations which fulfill user requirements. The recommendation systems operate through understanding visualization design principles and data characteristics ("item") rather than through user-specific historical data learning. Most visualization recommender systems documented in literature use rules or models that encode expert knowledge for effective visual mapping between data variables. The current trend shows the development of data-driven and learning-based approaches which seek to discover recommendation strategies from extensive collections of datasets and visualizations. We evaluate the primary visualization recommendation approaches starting from rule-based methods based on human-created design guidelines through machine learning models and hybrid systems while presenting examples such as Tableau's Show Me and the Voyager system and Data2Vis project. This evaluation examines various methods for their scalability features alongside their adaptability capabilities and evaluation methods.

2.1 Rule-Based Visualization Recommendation systems

Early studies established the foundation of rule-based visualization recommendation which applied predetermined rules to determine optimal data presentation methods for specific attributes. Knowledge-driven systems depend on formal visualization grammars together with empirical design guidelines to operate. The paper [8] established a compositional algebra for visual encoding description alongside ranking criteria selection for visual encodings. The APT system implements two essential rule systems which include expressiveness to protect data integrity and effectiveness to select perceptually optimal encoding methods. The expressiveness rules in visualization would block assigning two different data fields to the same visual channel to prevent confusion and the effectiveness rules would score bar charts over pie charts because bar charts utilize position/length encoding better for proportion display [9]. Mackinlay established a design framework that integrated Bertin's semiology of graphics and psychological findings to build the first automatic chart generation system for data characteristics.

Later systems developed this fundamental concept further. Sage expanded the APT system through the implementation of advanced data and visual element taxonomies for recommendation purposes. Early academic prototypes demonstrated that logical rules could generate appropriate visualization options. The Show Me feature in Tableau established commercial prominence for this design approach [4]. Show Me delivered automatic visualization recommendations to users within a popular analytics application. The system implements heuristics and defaults that function like a built-in rule engine to generate suitable chart types according to user data field choices. The application Show Me will recommend bar charts when users choose one categorical field together with one quantitative field, but it will suggest maps when geospatial fields exist and scatter plots when two quantitative fields are selected. Show Me uses VizQL declarative language from Tableau to automatically build views that automatically create small multiples when multiple fields are selected. The system checks expressiveness rules for suggested visualizations to guarantee compatibility with data types and effectiveness heuristics select charts which demonstrate high accuracy and clarity for variable combinations.

Research data from Tableau customers demonstrated that Show Me functionality worked effectively because both beginners and experts used the feature, and the automatically generated charts met expectations in 93% of evaluations [4]. The rule-based recommender system Show Me effectively assisted users in generating effective visuals across large scales because it used best design practices to make recommendations to users.

Rule-based and knowledge-based systems continue to be developed in addition to Show Me. VizDeck [10] introduced a basic system to prioritize and suggest visualizations for relational data and SeeDB [11] implemented automatic chart searching to detect meaningful differences in data subsets for discovery support. These systems rely on statistical characteristics of the data including variance and distribution shape and the existence of outliers to shape their recommendations in addition to perceptual rules. The research takes a different direction through constraint-based approaches by defining design rules as constraints and utilizing optimization solvers to discover the best visualization [12]. These systems share a dependence on human-curated rules which derive from visualization theory and perceptual studies and expert heuristics to direct recommendations. The recommendations provided by rule-based systems have the benefit of being explainable and predictable and they provide a constructed minimum level of quality. Rule-based recommenders function as natural knowledge-based systems because they apply domain-specific knowledge to match data to visual representations [13]. The main limitations of pure rule-based systems include their failure to detect complex contextual factors and user preferences and their sensitivity to conditions outside their predetermined rule boundaries and the need for extensive manual rule development and maintenance. The expansion of data and visualization domains makes it difficult to define all required rules and adjust their importance which drives the adoption of data-driven methods.

2.2 Mixed Initiative and Hybrid Systems

The Voyager system demonstrates this mixed-initiative philosophy. Voyager introduced in 2015 provided a tool for analysts to browse visualization recommendations through faceted filtering [14]. The system produces an arranged gallery of visualization options for given datasets which includes diverse variable pairings with transformation options. Voyager's recommendation process incorporates statistical measures together with perceptual measures. The system bases its chart suggestions on statistical methods that detect meaningful patterns and employs perceptual measures along with an effectiveness ranking system which follows the expressiveness/effectiveness rules from earlier systems. The analysis presented by the authors demonstrated Voyager assists analysts in discovering new data perspectives and extends their attribute examination reach beyond manual exploration methods. The application of an automatically generated chart selection by Voyager promotes users to expand their data exploration scope beyond typical variables and visualizations.

The Voyager 2 system along with its predecessor Voyager introduces users to the process of partial specification and visualization refinement. Voyager 2 enables users to identify specific elements of their interest by choosing one variable or selecting a particular chart type before allowing the system to generate the remaining suggestions. The system completes wildcard parameters with proposed visualization options which create multiple visualization candidates that contain the user-defined element [14]. The design addresses an essential problem because analysts typically understand what they want to see but also welcome different presentation methods thus a good recommender needs to recognize their intentions. Voyager 2 provides an interface that allows analysts to switch between designing visualizations manually and retrieving recommended alternatives through a seamless user experience. The system shows recommended charts that fulfill the user's selected criteria through an interface that limits the recommendation engine based on user input and enables the engine to generate outputs that inspire user actions. The authors present this as a mixed-initiative system which enables users to perform open-ended discovery and focused inquiries simultaneously.

User-guided recommendation appears in Data-Driven Guides as one of its primary features [15]. The Data-Driven Guides tool allows designers to use freeform drawing environments while generating guide marks based on data. Designers use guide marks which originate from data values to properly position and resize their hand-drawn visuals. The system recommends boundaries instead of providing completed charts. The system generates guides as specific data points or percentile representations that designers can use to maintain accuracy when creating illustrations. The system enables users to produce innovative infographic styles that maintain data relevance through its automated constraint recommendations. The system demonstrates flexibility by adapting to designers' creative workflows to help during appropriate moments. The underlying principle demonstrated in Data-Driven Guides remains applicable to visualization recommenders because it shows a method for integrating human creativity and preference into the recommendation cycle instead of presenting pre-defined output options [15]. The hybrid system enables users to maintain control of visuals through minimal automated support which maintains data connection in the final results.

Shneiderman's mantra and other HCI principles support the concept of mixed-initiative systems by showing users should maintain control over analytic processes [7]. A recommendation system needs to function as a copilot instead of presenting unexplained results by understanding the user's current objectives. A system requires detailed interface development combined with fast interactive backend processing to produce suggestions when users adjust controls. The technical implementation includes using rule-based logic to verify suggestion validity together with dynamic ranking and machine learning methods to determine the most relevant suggestions based on current context. The adaptive user experiences functions as a core feature of these systems through features like Voyager's bookmarking system which allows users to save interesting views while hiding unimportant ones to help train the system about their interests. The system proposes visualizations after asking the user questions about their task in dialogue-based recommendation prototypes which serve as analogs to dialogue-based recommenders in different domains.

Research indicates that hybrid and mixed-initiative approaches aim to unite the wide scope of algorithmic search with the deep understanding that human expertise provides. Voyager alongside other systems demonstrate hybrid methods produce superior exploration results by revealing additional data points along with new discoveries that human operators might otherwise overlook while maintaining alignment with analytical workflows that depend on iteration and context. The approaches indicate a possible future where visualization recommendation will operate as an analytic assistant which supports analysts during their work.

2.3 Challenges and Evaluation of Visualization Recommendation Systems

The development of an effective visualization technique recommendation system faces multiple obstacles which researchers continue to address through their studies. Three essential problems need attention in visualization systems: scalability, adaptability and evaluation.

Scalability: The number of possible visualizations expands exponentially with the number of data attributes together with available encoding options. Millions of different ways exist to visualize high-dimensional datasets by examining different variable pairings with various chart formats and transformation options. The effectiveness of recommender systems depends on their ability to efficiently manage an extensive visualization space. Rule-based systems handle this by using expressiveness constraints to remove invalid combinations and effectiveness heuristics to rank the remaining options. APT uses perceptual effectiveness pruning which significantly reduces the number of choices prior to recommending a solution. Tableau's Show Me feature restricts its chart suggestions to specific chart types based on the number and type of input fields (this approach makes the problem solvable) [4]. The number of valid visualization options remains substantial when data tables expand and become more complex. Researchers point out that scalability represents their focus when developing future interactive recommendation systems that handle large high-dimensional datasets [16]. The implementation of data mining methods serves as a scalability solution by recommending only the charts that reveal statistically important or abnormal patterns which automatically eliminate numerous unhelpful combinations. The visualization tools SeeDB and DeepEye utilize recommendation mechanisms to direct users toward potentially interesting views. The alternative method involves sampling the space where Voyager produces a range of charts instead of showing all possible combinations according to its principle of displaying data variation instead of design variation. This approach enables users to explore multiple data aspects while keeping the number of visual representations reasonable. A successful system must discover the perfect number of recommendations between insufficient suggestions which might miss important findings and excessive suggestions that overwhelm users' ability to select helpful results. Current research focuses on developing chart recommendation optimization methods alongside parallel and database-based visualization generation and understanding the optimal number and type of recommendations for users.

Adaptability: The ability to customize suggestions depends on the visualization recommender system's effectiveness in responding to task requirements and user preferences while considering the unique characteristics of the current dataset. The literature identifies a significant challenge in determining the analytical purpose of users during visualization tasks. According to Mackinlay et al. the task knowledge exists outside user mental boundaries while users start with undefined goals that change as they interact with data. Automated systems encounter difficulties when determining which visualization would be relevant for each user at each specific moment. Various systems have handled this challenge through different methods. Show Me implements a standard solution through its one-size-fits-most approach by recommending specific chart types for standard data configurations which delivers acceptable results for basic purposes such as overview visualization and value comparison. Such systems fail to deliver optimal results when users need something beyond standard functionality because they are searching for specific outlying data points or verification of detailed hypotheses. Voyager and similar mixed-initiative interfaces achieve better adaptability because users can actively direct the process through their selection of fields and bookmarking of views which provides implicit feedback for subsequent recommendation guidance [14]. Some research prototypes request user preferences or ask questions to better understand their goals (e.g., “Are you exploring correlation or distribution?”). A good system needs to identify when time series data requires line charts and geospatial data needs maps automatically without needing domain-specific programming. Modern approaches use meta-learning or modular design to achieve this capability through specialized recommendation modules that can be easily added when needed.

The ability of systems to adjust to user experience levels represents another element of adaptability. Designing recommendations for novice users proves challenging because it must avoid frustrating expert users. The HCI literature including [7] and subsequent works shows that systems need to have simple learning curves which allow novice users to depend on recommendations for any visualization, but experts use them for quick shortcuts and angle verification. The system needs to operate in both modes. Show Me’s usage analysis showed that expert users found the recommendations "at least kinda good" as a starting point which indicates that proper recommendations can benefit both novice and expert users. Future systems may develop the capability to learn individual user preferences by tracking how a user creates charts and then adding recommended log-scale options to the suggestions.

Evaluation: Assessing visualization recommendation systems faces fundamental challenges during evaluation. Traditional recommenders have a clear comparison point between predicted and actual user ratings, but visualization recommendation systems lack such a definitive "correct" visualization benchmark. The evaluation methods used in published research primarily rely on qualitative user-focused approaches. Researchers evaluate systems by conducting controlled user studies together with interviews and case studies to determine their impact on data analysis processes. The evaluation of Voyager compared user data exploration between tool-enabled and tool-disabled conditions and determined that users examined more attributes and discovered more insights when using the tool [14]. The research uses exploration breadth together with task completion time and insight discovery rate and user satisfaction as primary assessment metrics. The effectiveness of recommendations can be measured by their ability to enhance user accuracy when solving analytic tasks. Users can achieve better data analysis results by using recommended visualizations because they solve their questions both faster and more accurately. Some evaluations rely on logging and usage analytics data where Tableau tracks how users interact with Show Me and whether they accept the suggested charts or modify them. A low adoption or high override rate could indicate poor recommendations, whereas a high adoption and positive feedback signals utility.

Users bring preformed biases toward specific chart types to the evaluation process which produces unreliable results. The evaluation scenario needs proper definition because a system may perform outstandingly during exploratory usage yet struggle in presentation applications (or vice versa). To evaluate learning-based approaches researchers sometimes examine how well the system replicates expert-defined "ideal" visualizations through datasets with established ground truth. The evaluation of Data2Vis would assess its ability to produce the visualization selection which human analysts chose for specific datasets. The assessment method measures accuracy but depends on the human selection being an absolute reference which visualization lacks because there is no definitive standard. The field needs stronger evaluation methods that could borrow evaluation approaches from cognitive science for measuring cognitive load and insight as well as information visualization benchmarks which have established ground truth.

3 RESEARCH DESIGN AND METHODOLOGY

The research adopts Design Science Research (DSR) methodology to develop an innovative IT artefact through creation followed by assessment and iterative improvement. The artefact consists of a rule-based intelligent system which helps users choose appropriate data visualization methods according to their dataset features. The systematic and evidence-based nature of DSR makes its structured development and testing cycles highly appropriate for this research which focuses on solving practical problems.

The main problem this study addresses involves non-expert users struggling to choose from the wide range of data visualization methods. Users who lack extensive background knowledge struggle to select optimal visualization techniques because of the complex and variable nature of visualization choices which may result in inaccurate data interpretations. The research creates a Data Visualization Technique Recommendation System prototype through iterative development cycles to achieve usability and reliability and accuracy.

3.1 System Development Approach

The proposed system was developed using Python, chosen for its robust object-oriented programming capabilities which inherently support modularity, abstraction, reusability, and scalability. Python's versatility significantly streamlined the development process, facilitating the implementation of sophisticated algorithms and clear software architecture. To enhance the clarity and maintainability of the system, a Model-View-Controller (MVC) architecture was adopted. MVC clearly separates responsibilities within the system into three distinct modules, each managing specific tasks and thereby improving maintainability and clarity:

1. **Model:** The model encapsulates the core logic related to data visualization techniques, rule management, and user input evaluation. It interacts directly with a MongoDB database, facilitating data retrieval and storage. Specifically, the model is responsible for defining the schema of visualization techniques, rules, and user inputs.
2. **View:** The user interface (UI) component guides users through an intuitive, structured questionnaire to gather relevant dataset characteristics. The UI clearly presents recommended visualization methods, detailed explanations, and supportive visual examples, enhancing user comprehension and engagement.
3. **Controller:** Serving as the intermediary, the controller manages data flow between the model and the view. It processes user interactions, invokes rule evaluation logic, and updates the interface with appropriate responses, ensuring coherent communication between system components.

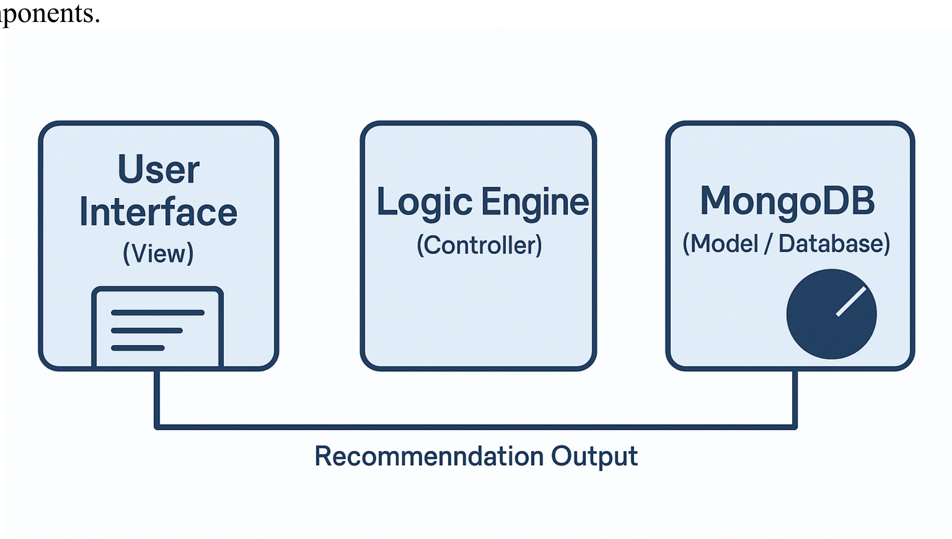

Figure 4. System Architecture of the Data Visualization Technique Recommendation System.

3.2 Rule-Based Recommendation Mechanism

The system depends on its rule-based reasoning model which was created as an essential part of the system's logic. The interactive questionnaire allows users to respond to targeted questions which are then evaluated dynamically against a large, predefined rule base. The rule base was developed through visualization best practices and theory-driven criteria to determine the most suitable visualization techniques for each user dataset characteristic.

The system provides clear explanations for each recommendation through explicit rule descriptions which help users understand the reasoning behind visualization method choices. The system's transparent nature builds user trust while simultaneously teaching users about visualization methods which leads to improved visualization literacy with time. The rule-based system has a modular structure which enables easy modifications for incorporating new visualization techniques and updating existing ones according to recent research findings and changing best practices.

3.3 Database Creation and MongoDB Integration

A relational schema initially guided the database structure development which then transitioned to MongoDB as a flexible NoSQL document-oriented database. The decision to move the database to MongoDB occurred because this NoSQL solution provides superior handling of hierarchical and semi-structured data compared to traditional relational databases.

The database development process was a complex and independent work that I performed from start to finish. Thorough research combined with detailed analysis became essential to achieve complete accuracy and usability. I followed a methodical approach to authoritative sources including the "Periodic Table of Visualization Methods" and multiple academic papers and credible online resources. The database contained a thorough evaluation of visualization techniques through which I analyzed their distinct characteristics along with their advantages and disadvantages and suitable data types and suitable application environments. The process required me to insert each entry manually while performing verification steps to guarantee information accuracy and full data consistency. The system's recommendations received thorough examination through this precise method to ensure users would find them both accurate and meaningful and applicable to real-world situations.

The rule set development followed an iterative process which was both challenging and exploratory in nature. The beginning of my work involved establishing preliminary rules through theoretical foundations found in data visualization research and cognitive load theory. The initial rules went through multiple refinement cycles after empirical experiments combined with practical application scenarios. The iterative testing involved applying each rule to different datasets to check its recommendation accuracy and usability. This task proved particularly difficult because users needed to understand recommendations easily but the system required enough complexity to maintain accurate recommendations across different data situations.

During rule creation, preliminary user testing generated feedback that led to continuous adjustments of the rules. The feedback loop served as a vital mechanism to detect weaknesses along with potential oversights in the original rule definitions. Each revision targeted to enhance both rule accuracy and widen visualization applicability to different scenarios. The iterative and detailed methodology generated a powerful adaptable rule base.

MongoDB enables each visualization method and related rules to store attributes that differ because its flexible schema matches the varied nature of visualization approaches. MongoDB contains multiple main collections that store the following data sets:

- The documents in this collection hold detailed descriptions together with visual examples (image paths) along with advantages and disadvantages and rule identifiers.
- Rules: Each document contains a unique rule ID, detailed explanation, and logical conditions governing the recommendation of visualization techniques.
- Questionnaire Data: The database contains structured questions along with potential responses which directly impact the rule evaluation process through integrated logic.

The PyMongo library enables Python to interact with MongoDB through which the application logic achieves efficient and robust database interactions. The integration allows fast data querying and modification alongside scalability which enables smooth addition of new visualization methods and rules and questionnaire elements.

3.4 Interactive Q&A Framework

The system includes an essential user interface component which uses visualization theory to develop its interactive questionnaire for dataset characteristic extraction. The questionnaire contains 8-10 multiple-choice questions which assess dataset features including data type and dimensionality and analysis intent and other factors that affect visualization selection.

The questionnaire responses activate particular rules in the rule base that subsequently control the recommendation logic. The framework reduces mental effort through brief targeted questions to obtain necessary information which leads to precise personalized recommendations based on data visualization best practices.

3.5 System Architecture Description

The architecture of the Data Visualization Technique Recommendation System consists of multiple interconnected modules that achieve a balance between user accessibility and scalability and intelligent decision-making. The View Layer functions as the main interface which provides an interactive questionnaire system that leads users through a controlled set of questions. The presentation of visualization recommendations through this layer maintains simplicity which makes it accessible to users who lack technical experience.

The Controller functions as the central component which directs user input data while connecting the interface to the system's internal operations. The system ensures that user responses are automatically transmitted to decision-making components while the interface updates dynamically based on rule evaluations. Real-time feedback through this system creates an improved user experience because it delivers prompt and context-specific responses.

The system depends on MongoDB as its main database which stores all vital data structures including visualization techniques alongside recommendation rules and questionnaire logic. The document-oriented model of MongoDB works best for this application because it effectively manages complex nested data schemas that evolve over time. The flexible data storage system supports the dynamic recommendation system operations and enables quick data access and modification.

The Model component contains a Rule-Based Decision Engine that performs decision rule evaluations on user inputs. The engine functions in real time to produce customized recommendations which derive from user-provided dataset characteristics. The system provides clear explanations for each recommendation which builds user trust and strengthens the credibility of its decision-making process. The system's components unite to create an architecture which enables intelligent adaptable user-centered visualization guidance.

3.6 Visualization Technique Repository and Dataset Description

The critical component of the expert-based recommendation system is the extensive inventory of visualization forms that enables the matching of user data characteristics with types of graphics. This inventory is based on around 40 different visualization forms ranging from basic, commonly used forms of visualizations like bar charts and scatterplots to secondary types of visualization forms like spectrograms and Tukey box plots.

The dataset was built from multiple reliable sources that were useful for the various domains of knowledge. Among these sources were the Visual Literacy Periodic Table of Visualization Methods (Figure 5), Oracle's official documentation on charts, academic research in visual perception and encoding, and practically based manuals on data visualization in both data science and business analytics that provided theoretical basis and practical knowledge on the utilization, pros-and-cons of each visualization forms.

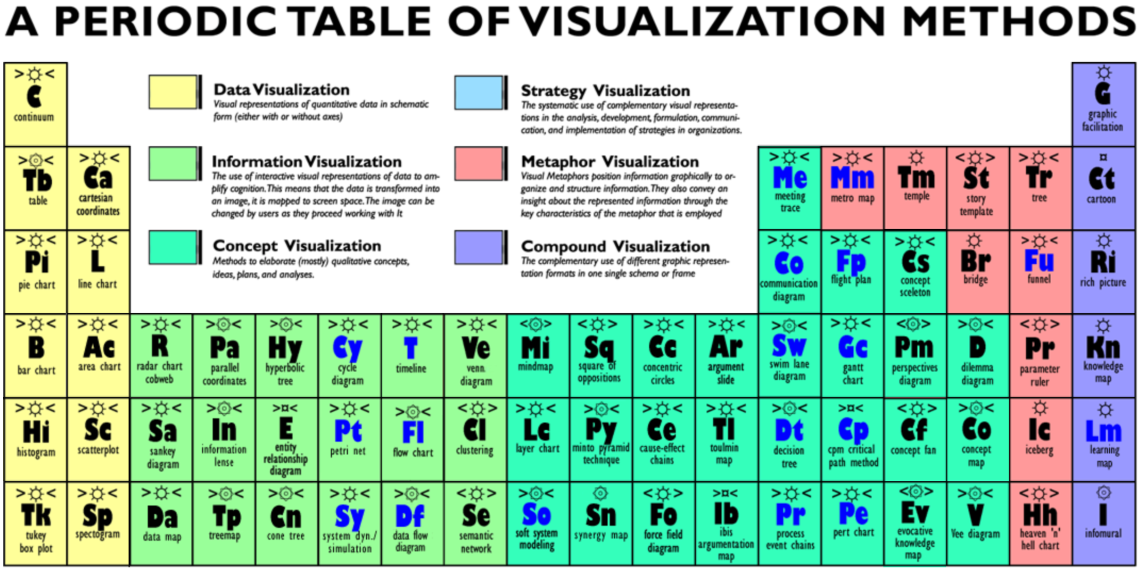

Figure 5. A Periodic Table of Visualization Methods.

Each technique in our dataset is associated with a lot of descriptors that enable recommendation and explanation. These descriptors include name of the technique; the type of data supported (e.g., categorical, numerical, textual); the intended visualization purpose (e.g., compare, detect trends, study distributions); a brief write-up; its pros and cons; and the file path to a sample image to illustrate the technique. Furthermore, additional descriptive values were developed to assist in the rule generation process. The descriptors we established include a flag to denote whether the technique supported multivariate data, as well as flags to indicate whether the technique required the data to be scaled/normalized, how well it handled outliers, the density of data, and how interpretable the technique would be to a non-technical audience.

For example, a scatterplot could be flagged as acceptable for multivariate data, appropriate for detecting trends or correlations, useful for identifying outliers, and having moderate interpretability for a non-technical audience. Each of these descriptors will serve as the conditions or targets used in the logical rules that make up the recommendation engine, allowing the system to analyze the user responses and evaluate corresponding technique profiles.

For their clarification, each of the visualization modes has a supporting example image which was shown to users when making recommendations. These images facilitate recognition and assist less experienced users in being able to visualize what the recommended chart may look like. An example of a small sampling of these visualizations in our system is displayed in Figure 6.

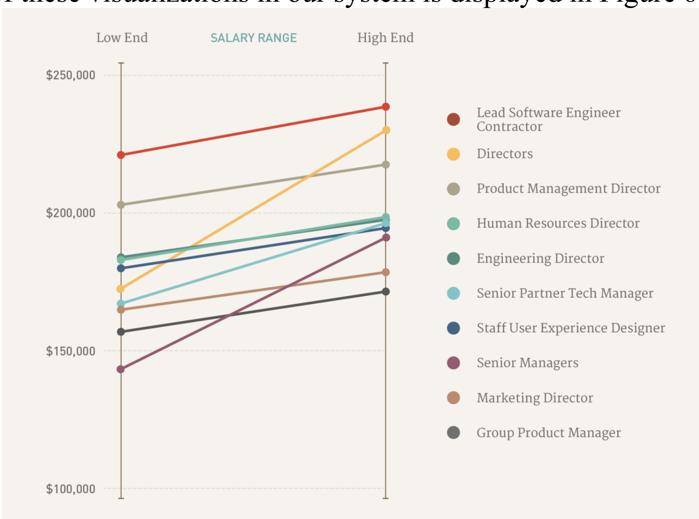

Figure 6. Example of one of the visualization technique images: Slope Chart.

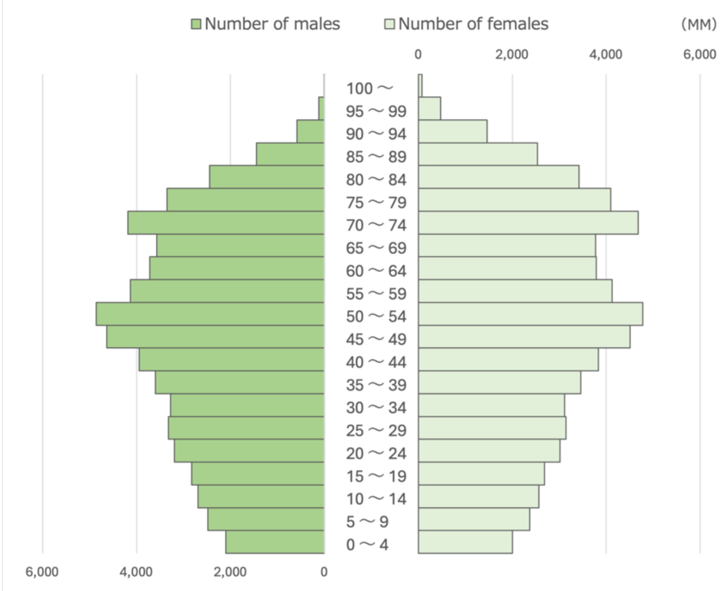

Figure 7. Figure 6. Example of one of the visualization technique images: Butterfly Chart.

Overall, this structured, attribute-rich dataset serve a dual purpose - to give accurate, interpretable recommendations with the rule-based engine, while concurrently providing a knowledge base to instruct users about a remarkable range of data visualization possibilities.

4 RESEARCH RESULTS AND ANALYSIS

The Data Visualization Technique Recommendation System was successfully developed as a fully operational interactive prototype that integrates expert knowledge into a rule-based decision-making framework. The system has been developed to enable users to choose the most suitable data visualization technique based on the characteristics of their data and objectives of analysis. The recommendation process is based on domain expertise, formalized into a structured set of rules and attribute mappings, and delivered through a guided question-and-answer interface. The final working system incorporates a knowledge base of visualization methods, dynamic interaction logic, and an intuitive graphical user interface, all orchestrated under the Model–View–Controller (MVC) architecture. This architecture was chosen because it enables clean separation of concerns, enhances system modularity and ensures ease of future maintainability and scalability.

In the Model component of the system, a curated and structured repository of visualization techniques has been developed and maintained. This repository is stored either in a collection within a MongoDB database, making it flexible for integration and expansion. Each visualization method is defined with a comprehensive set of attributes, describing the conditions under which it is suitable for use. These attributes include, but are not limited to, the data type it is intended for (e.g., numerical, textual, or mixed), whether it supports continuous or categorical data, the number of categories it handles well, whether it accommodates multivariate data structures, sensitivity to scaling requirements, suitability for high data density, responsiveness to outliers, and other relevant analytical contexts. Additional metadata includes specific textual analysis capabilities (e.g., frequency analysis vs. semantic structuring). These descriptive features allow each technique to be rigorously characterized in terms of when and how it should be applied, serving as the foundation of the system's expert knowledge base.

This knowledge base functions as the system's expert model, capturing visualization best practices, guidelines, and practical considerations in a structured, machine-readable format. These encoded rules allow the system to evaluate and rank visualization options based on the match between the user's dataset characteristics and each technique's requirements. This model-centric approach mirrors how a domain expert would reason through visualization decisions, ensuring that users receive high-quality, contextually appropriate suggestions.

The Controller serves as the brain of the system. It implements the core rule-based decision logic, as well as managing the overall flow of the interactive Q&A dialogue. As users proceed through the interface, answering each question about their data, the controller records responses, determines what information is still required, and dynamically selects the next relevant question based on a branching logic. This adaptive questioning ensures relevance at every stage of the dialogue and avoids unnecessary questions. For instance, if the user's dataset is identified as "Numerical," the system will branch into follow-up questions about whether the data is continuous or categorical, and depending on the path taken, delve into questions about scaling needs, multivariate structure, or density considerations. In contrast, if the user indicates that their data is "Textual," the system branches into questions about whether the focus is on frequency (such as term occurrence) or meaning (such as semantic relationships or embeddings). Each path through the questionnaire is uniquely constructed in real time based on user answers, emulating a guided interview with a domain expert.

In total, the interactive dialogue covers a comprehensive yet concise set of data characteristics critical to selecting visualization techniques. These include data structure, dimensionality, distribution, volume, and interpretability aspects, among others. The controller also keeps track of the user's position within the dialogue, enables Back and Next navigation for improved usability, and ultimately compiles the answers into a structured dataset profile. This profile acts as the basis for visualization recommendation. After questionnaire completion the system triggers the recommendation engine that belongs to the model layer. The recommendation engine scores every technique in the knowledge base using the dataset profile information. The scoring logic compares user-submitted dataset features with visualization technique attributes stored in the system. The scoring system gives positive scores to techniques whose attributes match the dataset profile but assigns negative weights to techniques that do not match. The Box Plot technique would obtain high scores because it specializes in displaying continuous numerical high-density data that emphasizes outlier detection. The system would strongly penalize semantic textual data techniques like dependency trees or semantic networks because they do not match the current dataset profile.

The scoring-based ranking system implements the expert's decision-making behavior through quantified rules that enable multiple techniques to be compared. After scoring all techniques the View selects the highest-scoring ones (or multiple techniques in case of equal scores) to display to the user. The recommendation process is precise, interpretable, and adaptable — designed to support both singular recommendations and a broader range of suitable alternatives.

The View, which uses Streamlit for its development, offers the system's user interface. The system displays information in an easy-to-use format through its simple design. Every question in the interface includes options for selection alongside definitions of technical terms. The description under the "Multivariate Data" question explains what the term signifies while helping users recognize it within their datasets. The built-in educational elements of the interface function to serve users who need clarification about complex data concepts. The interface contains a "Back" button which lets users modify their previous responses without facing any consequences. Users can explore through trial and error while the system allows them to discover different input paths to understand how recommendations change.

Using Streamlit to implement this interface is a good option because it allows us to use pure Python to create interactive web applications. It's easy to prototype with, It can make very clean and responsive layouts but I particularly liked that it integrates seamlessly with all the data, including machine learning models, and other workflows. Streamlit is definitely intended for dynamic forms and has built-in components (e.g. sliders, buttons, select boxes) and the like that makes this implementation straightforward. Using Streamlit for building the guided questionnaire interface is a good choice because of its minimal front-end development effort which is added bonus because the system will be easier to maintain and scale.

After generating a recommendation the View displays, it through an easily consumable and interactive interface. Users can view full details about a recommended technique through a click which includes its name along with description and advantages and disadvantages and brief application scenarios and example images when available. Through this presentation users gain understanding of both the recommended option and its selection rationale along with visual examples. The detail view plays an essential role in building trust since it gives users the chance to verify the recommendation by themselves.

The system includes an additional "More Recommendations" button. The system displays additional techniques that also match the dataset but hold a lower ranking than the initial recommendation. The feature proves beneficial for practical situations because it allows selection of appropriate techniques despite presentation context variations and user preferences as well as delivery methods between static reports and interactive dashboards. The ability to present multiple valid choices aligns with the usual flexibility experts demonstrate when making decisions.

Users have access to a "Try Again" button which starts the questionnaire process from the beginning after resetting the input. The system enables users to try again through the "Try Again" feature which lets them begin the process from the start. This allows users to test various data conditions to understand how visualization recommendations modify their approach based on dataset characteristics. The system functions as both a decision-support tool and a learning platform with special benefits for students and novice analysts.

The completed system functions as an interactive system that recommends data visualizations to users. The system utilizes expert knowledge stored in a rule-based model to generate recommendations through interactive dialogue which are presented through an educational user interface. The MVC architecture serves as the foundation which supports the system's maintainability and clarity. The system operates through the coordinated effort of three components which include the Model (expert knowledge and rule engine), the Controller (logic and dialogue manager), and the View (interface and user interaction). A robust user-friendly experience results from the cohesive operation of these system components. The system delivers effective data visualization recommendations to users who possess any level of expertise in the field through its educational and context-sensitive interface.

4.1 Scenarios and Screenshots

Multiple representative scenarios were developed through sample dataset implementation to show the system's practical usage. The examples show how the interactive Q&A framework transforms user input into visualization technique recommendations through dataset characteristic analysis.

4.2.1 Numeric Scenario

A telecommunications company is analyzing the distribution of different types of customer complaints received over the past year, with the goal of identifying major areas for service improvement.

Dataset Description:

Complaint_Type: Categorical variable (e.g., Billing Issues, Network Coverage, Internet Speed, Customer Service, Mobile App, Other)

Complaint_Count: Number of complaints for each type

Screenshots:

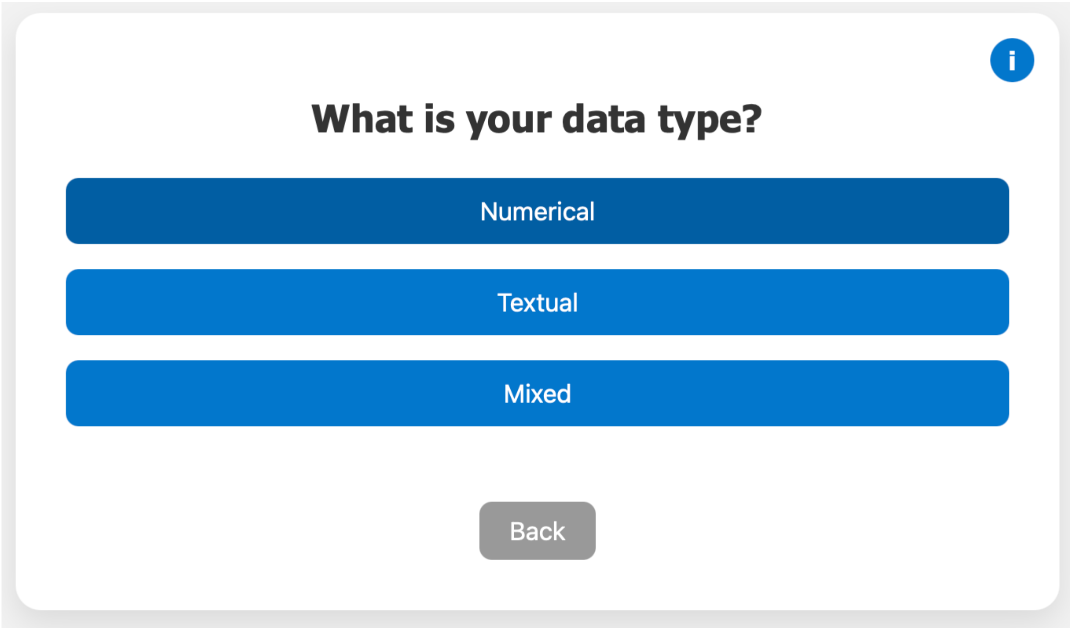

Figure 8. User interface for selecting the data type: Numerical, Textual, or Mixed.

The dataset includes numerical data representing complaint frequencies or percentages for each category. The text categories in the dataset contain numerical complaint proportion values which need to be visualized. The answer selected in Figure 3 is “Numerical”.

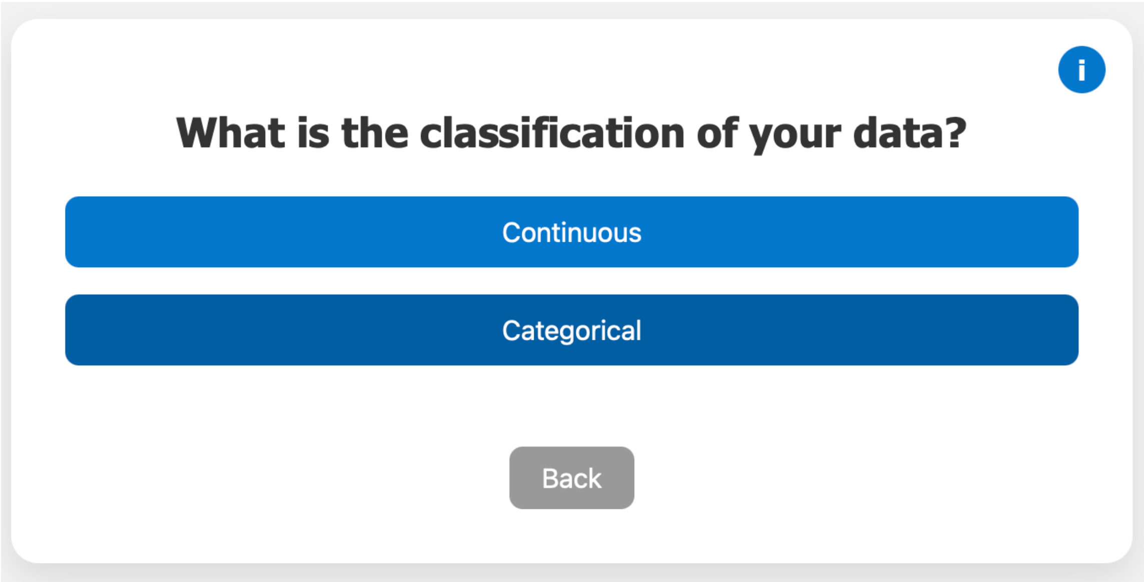

Figure 9. Interface for selecting the classification of the data: Continuous or Categorical, with explanation tooltips.

The complaint types (e.g., Billing, Network, Support) are distinct groups, making them a classic case of categorical classification. This analysis involves the evaluation of named categories instead of tracking continuous measurements similar to temperature or time measurements. A suitable answer for Figure 4 would be “Categorical”.

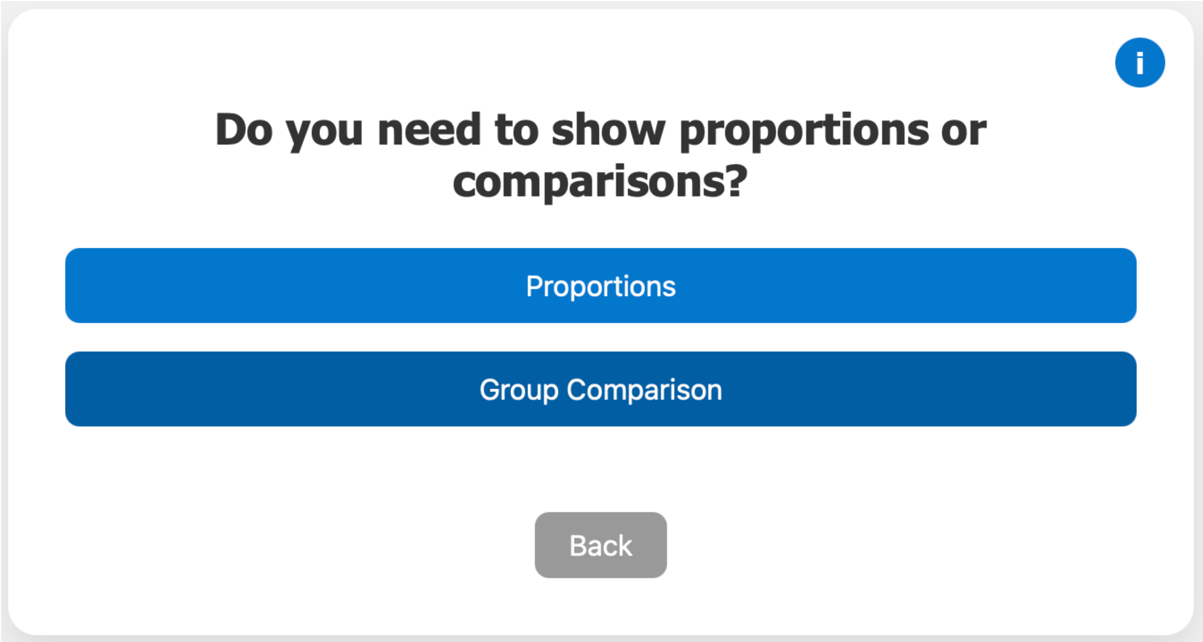

Figure 10. User prompt asking whether the goal is to show proportions or group comparisons.

The analysis aims to identify the frequency of complaints for each type to determine which types occur more or less often. The presentation focuses on parallel category analysis rather than measuring category contributions to the complete dataset. Group comparison visualizations work best to show the magnitude differences between different complaint types.

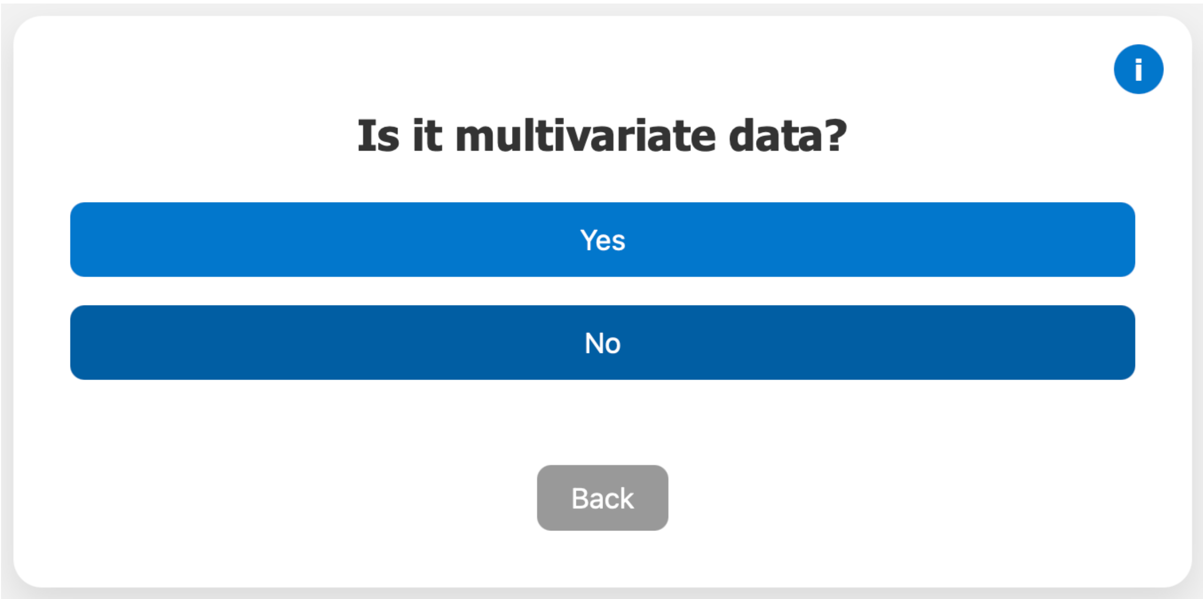

Figure 11. User interface asking whether the dataset contains multivariate data.

The dataset requires univariate visualization since you are displaying only the distribution of one variable (complaint type share) instead of examining multiple variables' interactions. “No” was chosen as an answer for Figure 6.

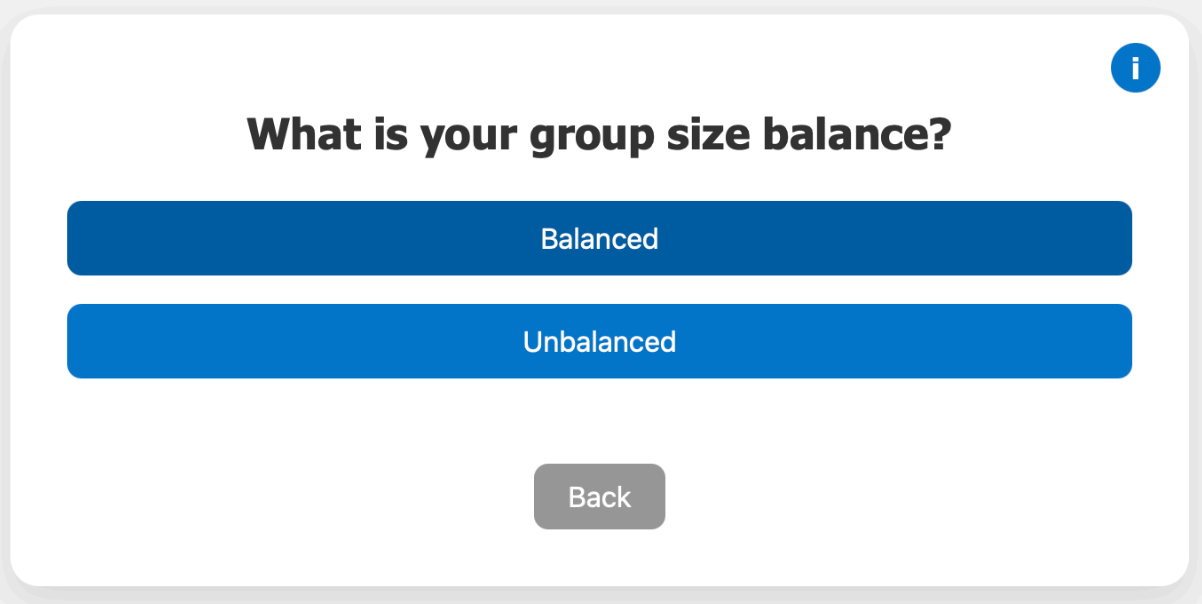

Figure 12. Interface asking about group size balance with definitions for balanced and unbalanced groups.

The distribution of categories shows approximately equal sizes because no single group prevails in the dataset. A balanced distribution enables better visual analysis of categories while maintaining group size differences visible in comparative visualization. “Balanced” is a suitable answer in Figure 7 for the dataset.

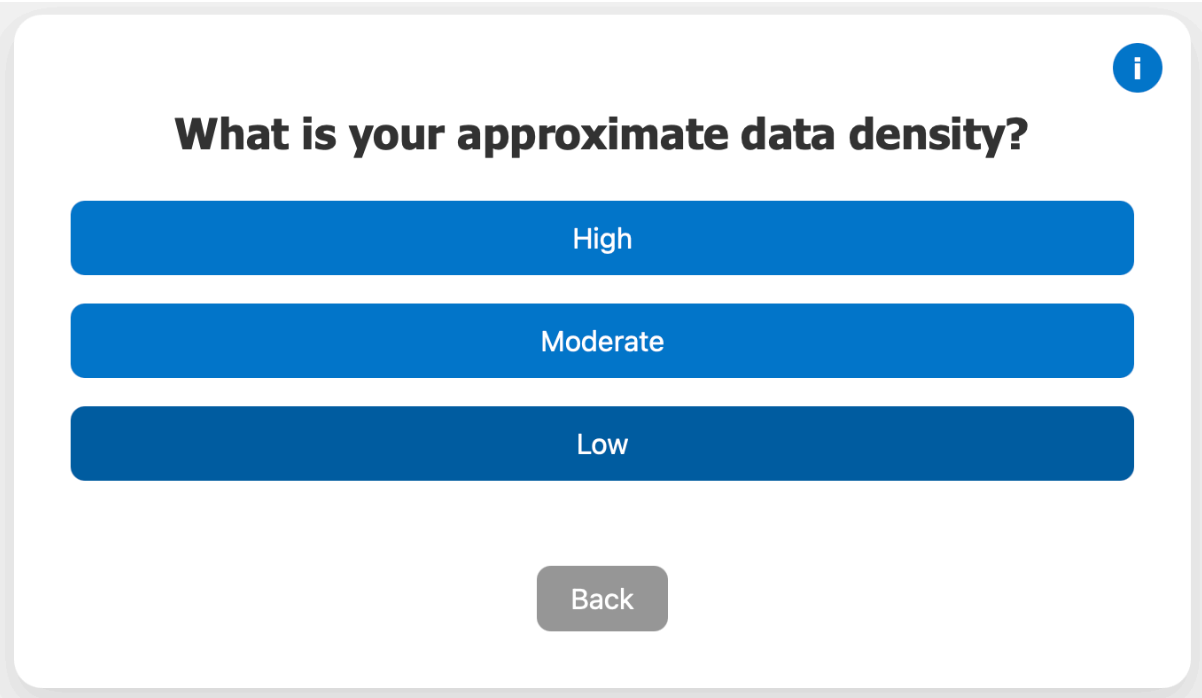

Figure 13. Interface prompting for data density selection: High, Moderate, or Low.

The dataset features a small set of categories which remains below ten in number. The visualization performs well with this dataset because a small number of categories makes it simple to understand both group distinctions and individual values. “Low” is an appropriate answer in Figure 8.

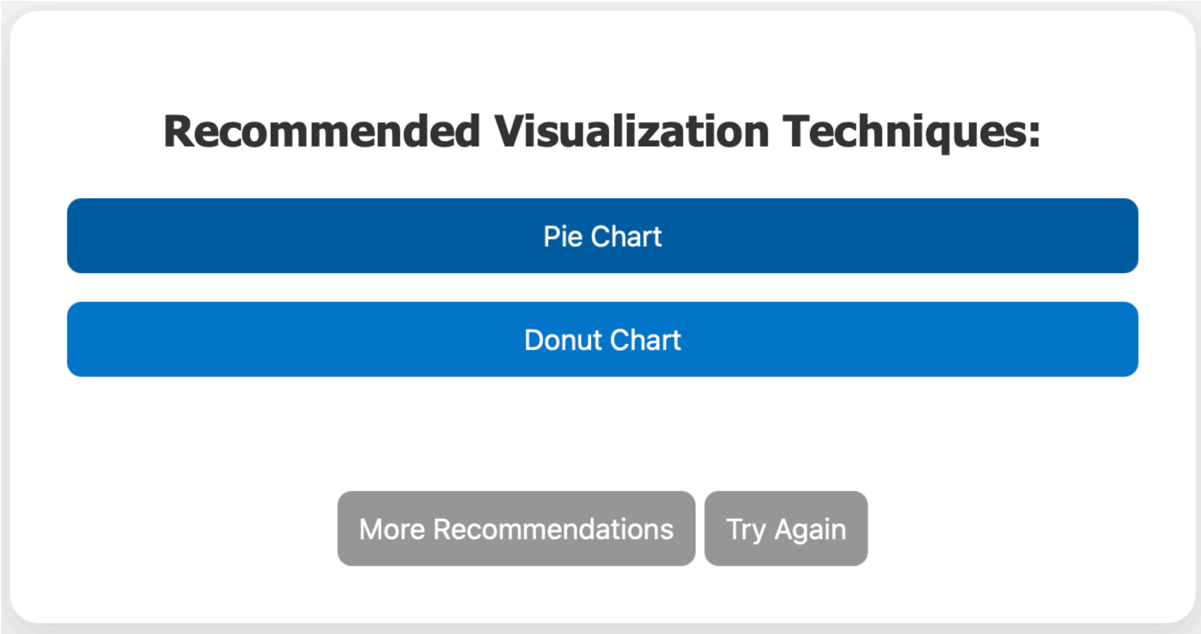

Figure 14. Final visualization recommendations based on user input: Pie Chart and Donut Chart.

The system assesses user Q&A responses against its expert rule base to produce the most appropriate visualization techniques. In the following example, the user provided the following inputs:

- Data Type: Numerical
- Classification: Categorical
- Goal: Show proportions
- Multivariate: No
- Group Size Balance: Balanced
- Data Density: Low

The system selects Pie Chart and Donut Chart as the most suitable visualization methods based on the provided inputs. The telecom complaint dataset fits these visualization techniques well because it contains a limited number of categorical groups to display part-to-whole relationships.

The telecom complaint dataset consists of numerical values showing complaint frequency for various categories such as Billing Issues, Network Problems and Customer Support. The main objective is to assess the complaint frequency distribution across categories while determining their impact on total complaints.

Both Pie Chart and Donut Chart emerge as suitable visualization options because this dataset features categorical data with numerical values and non-multivariate structure and balanced group sizes and few categories.

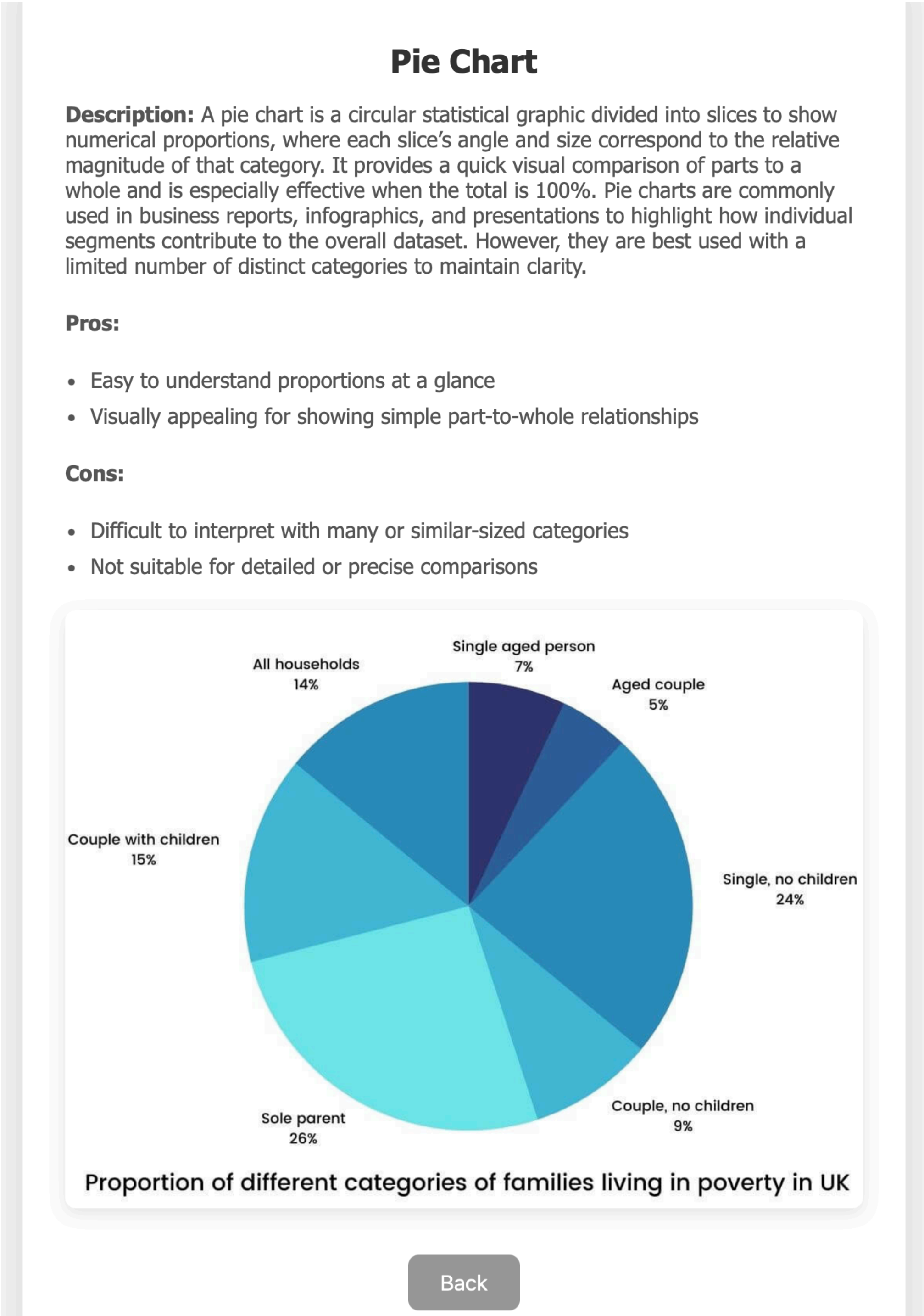

Figure 15. Pie Chart details view.

The user views detailed information about the Pie Chart technique after selecting it which includes its definition along with advantages and disadvantages and an illustration. Users gain understanding of the recommended solution together with the reasons behind its suitability for their specific dataset.

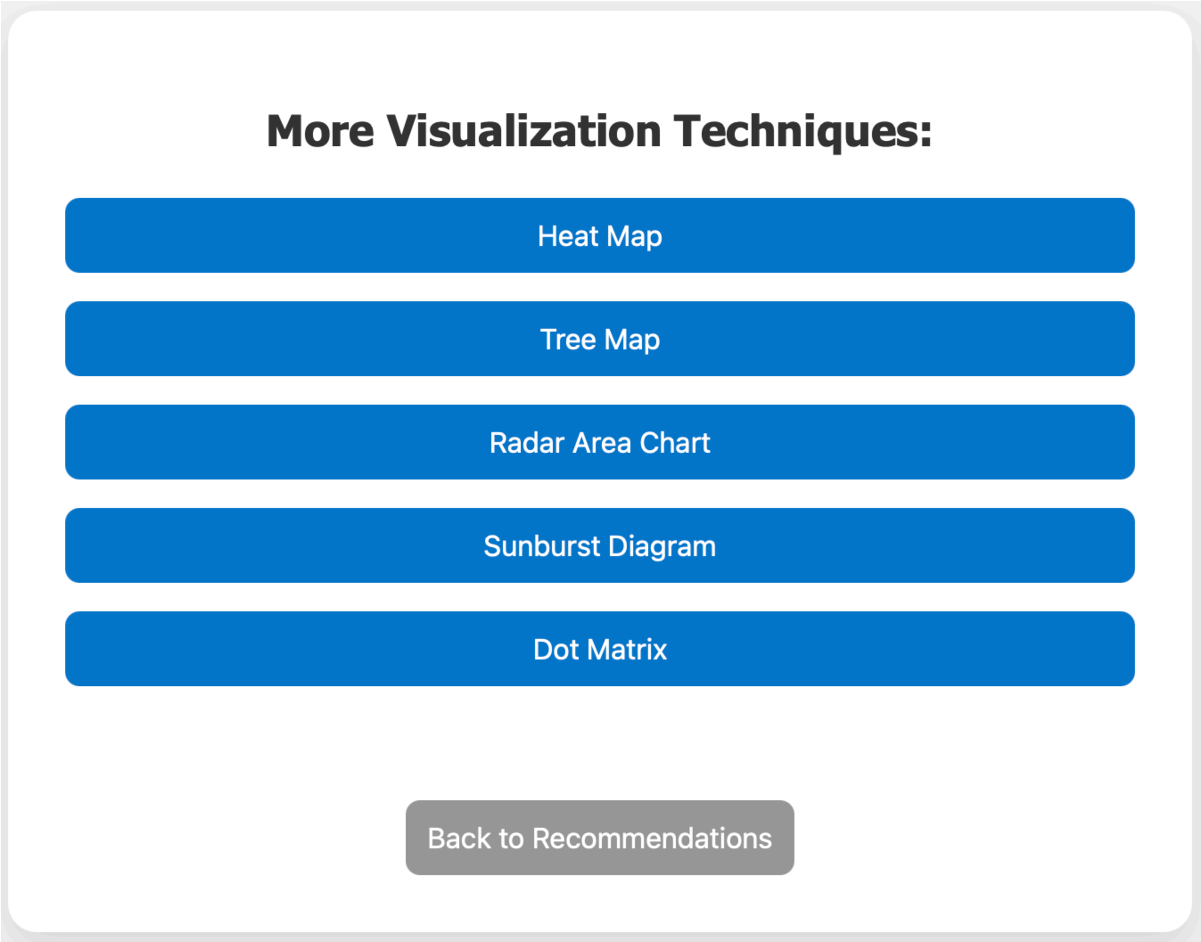

Figure 16. Additional visualization techniques presented as extended recommendations.

After viewing the primary recommendations (Pie Chart and Donut Chart), the user has the option to explore More Visualization Techniques. This feature is designed to provide broader insight and alternatives, especially useful when:

- The primary recommendations don't fully meet the user's design or interpretability preferences
- The user wants to explore less conventional or more visually rich representations of the same data
- Multiple visualizations are being considered for different parts of a report or dashboard

These additional options are still filtered based on the dataset's attributes, ensuring they remain contextually relevant.

4.1.3 Textual Scenario

Scenario: A customer support team at an e-commerce company wants to analyze open-ended feedback collected from post-purchase surveys. These responses are unstructured and vary widely in content, but the team is interested in identifying common themes and frequently mentioned terms to improve their service.

Dataset Description:

- Customer_ID: Unique identifier for each respondent
- Feedback_Text: Open-ended textual responses from customers

Screenshots:

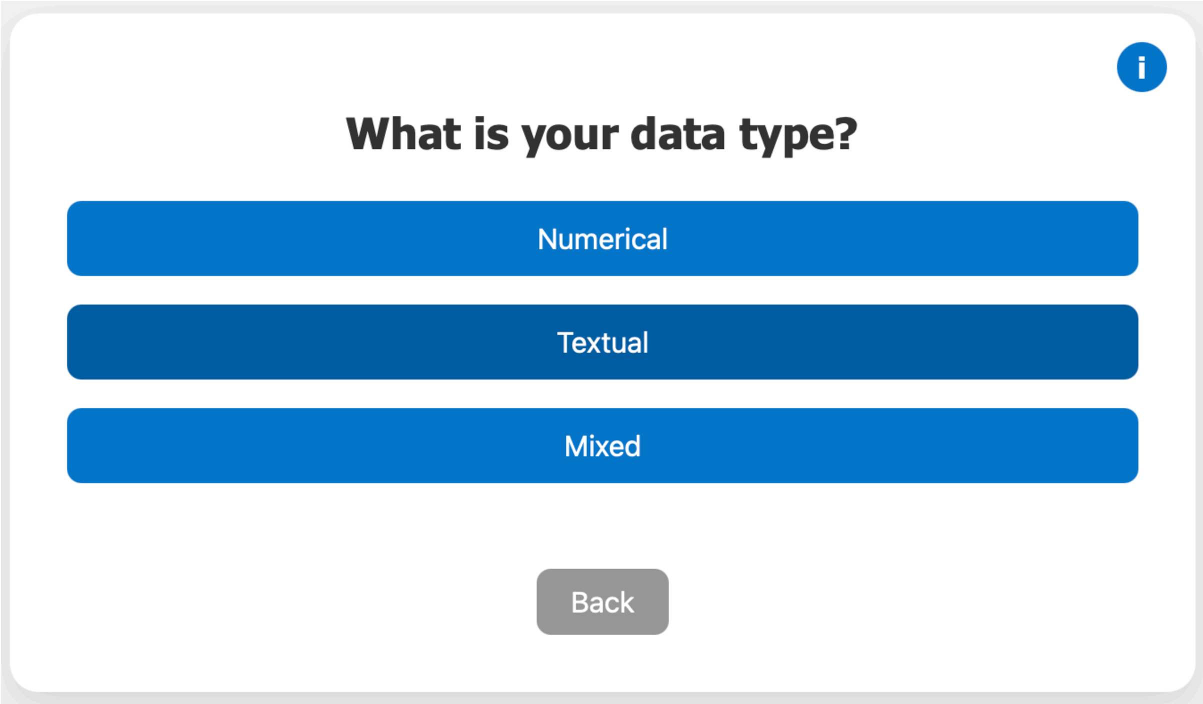

Figure 17. Interface for selecting the data type.

The system first requires users to choose the data analysis category. The input data is textual which means the dataset contains unstructured or semi-structured language data like reviews or feedback or documents. Answer for Figure 12 is “Textual”.

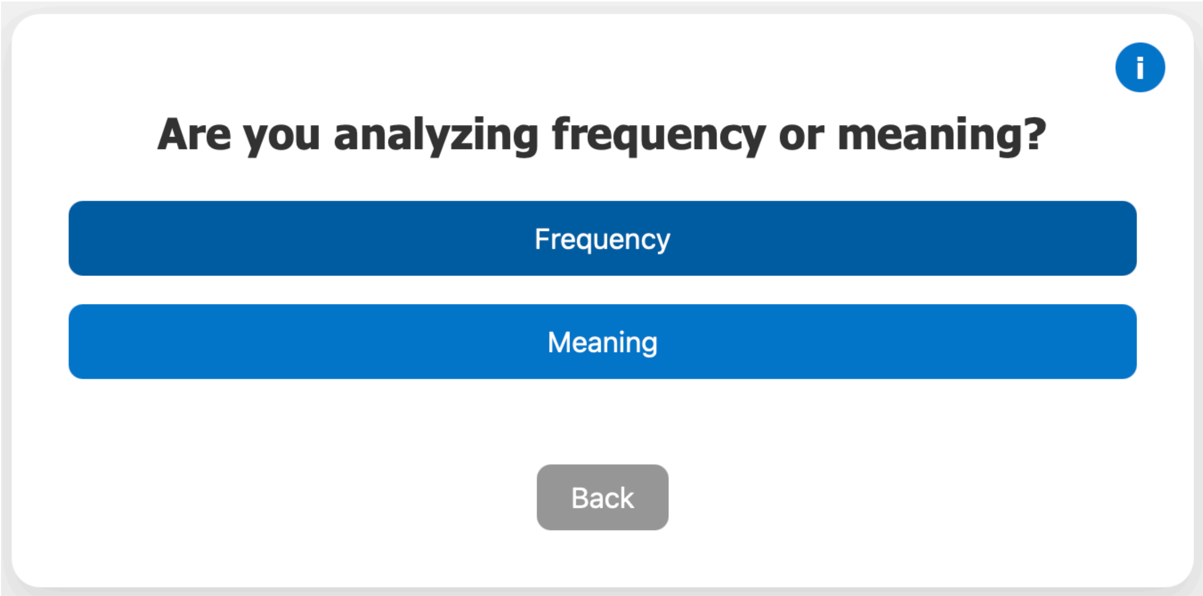

Figure 18. Question asking whether the analysis is based on word frequency or semantic meaning.

The system presents an option for users to choose either analyzing frequency or meaning in Figure 13. Selecting frequency indicates the intent to focus on how often specific words appear in the dataset, which is common when trying to identify recurring themes or dominant terms across large volumes of text.

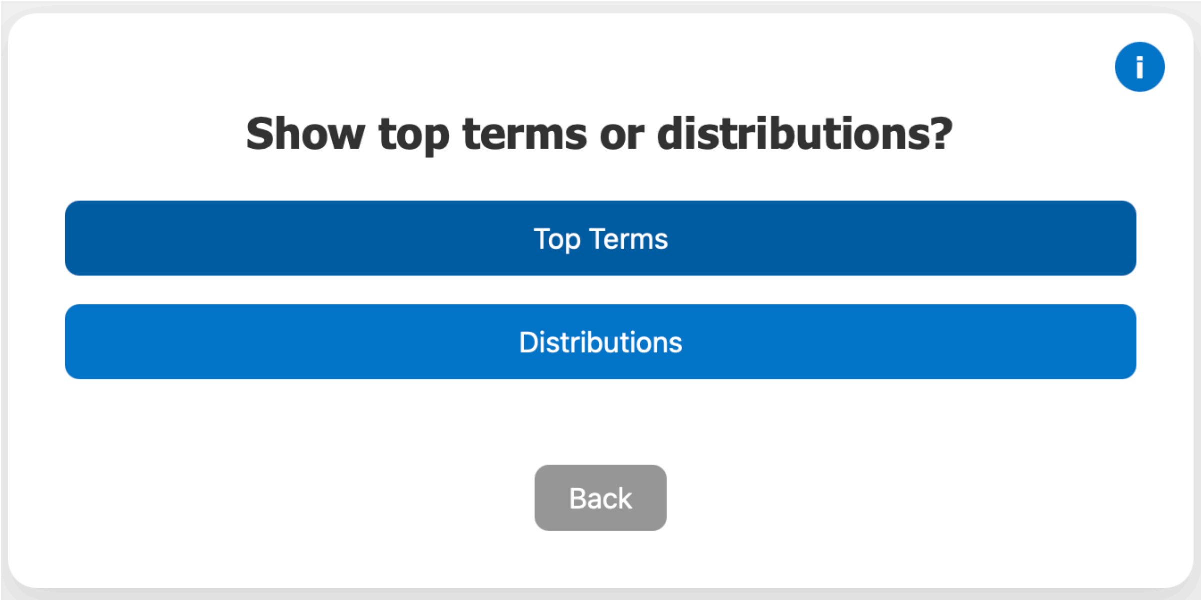

Figure 19. Interface asking whether the goal is to identify top terms or distributions.

Figure 14 of the system describes the purpose of textual analysis by specifying if users need to find main terms or check term distributions. The method of highlighting top terms works best for summarizing the most frequent content while conducting exploratory text analysis.

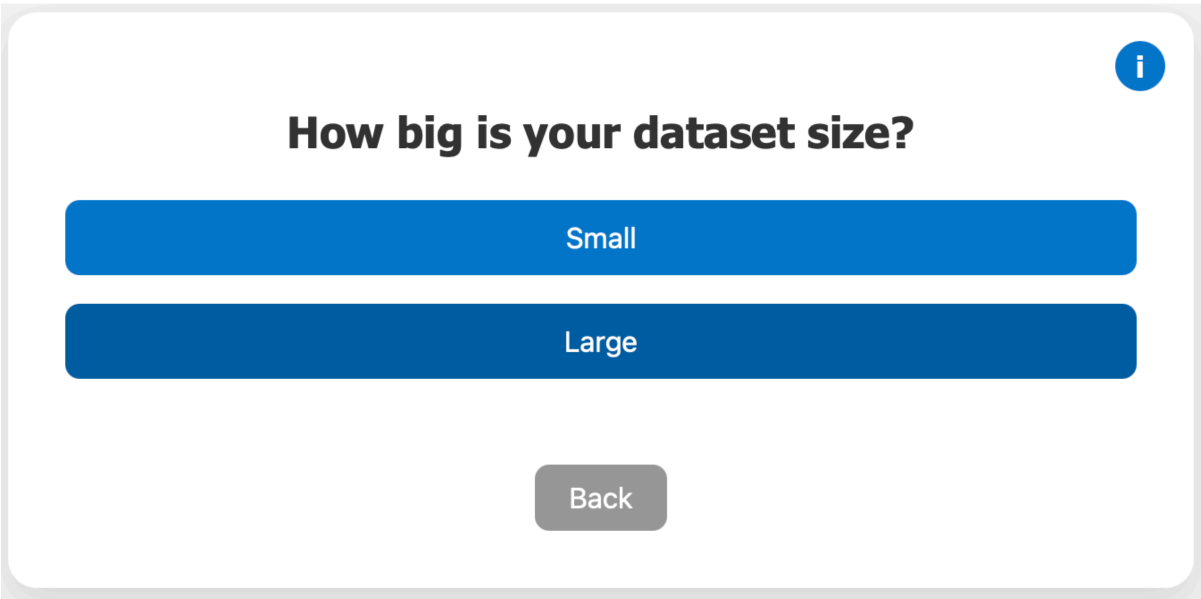

Figure 20. Dataset size selection screen.

The following step requires information about dataset dimensions. Large dataset indicates a big number of text entries including thousands of customer reviews or documents that need scalable and efficient summarization techniques. The answer to Figure 15 is “Large”.

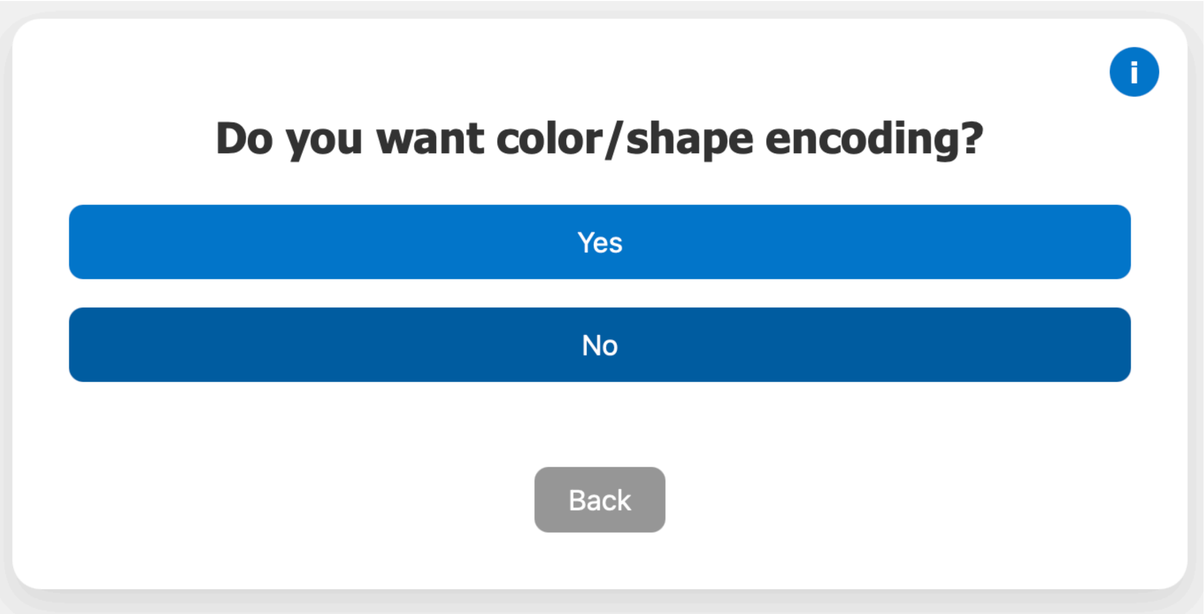

Figure 21. Interface asking whether to apply color or shape encoding.

The system asks users if they require either color or shape encoding as the final step. The selection of "No" in Figure 16 indicates that the visualization should depend solely on term prominence instead of additional visual distinctions for clear and simple presentation.

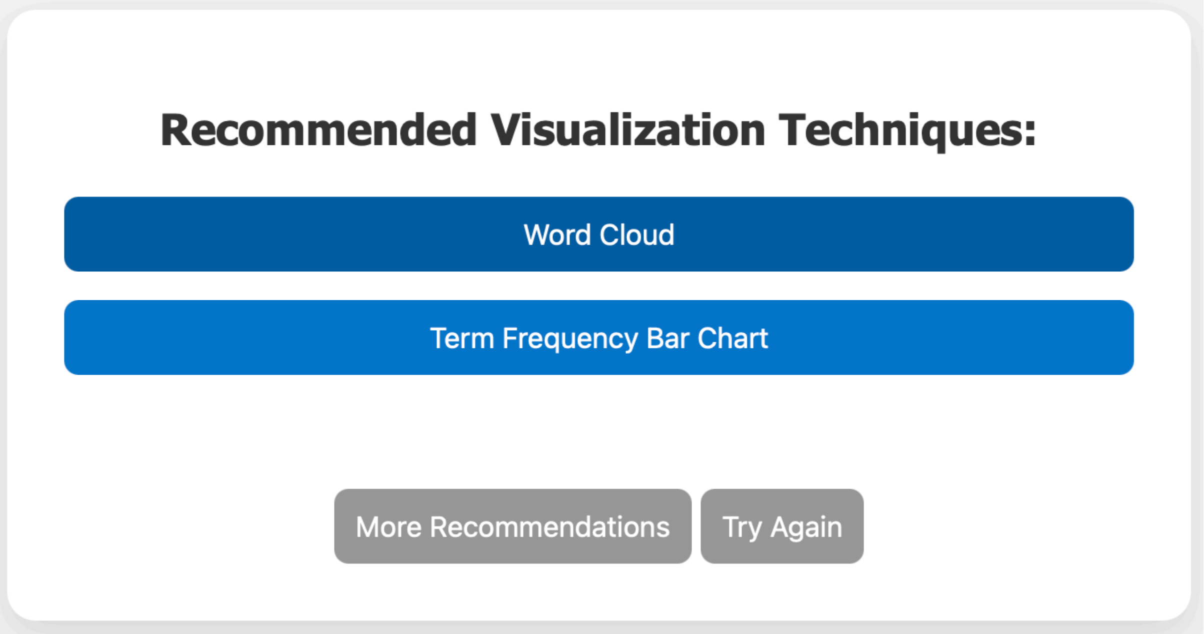

Figure 22. Final recommendation screen.

The system assesses the given inputs against its expert rule base because the data is textual and frequency-based analysis is selected for top terms from a large dataset that does not require color or shape encoding. The system selects Word Cloud and Term Frequency Bar Chart as the most suitable visualization methods.

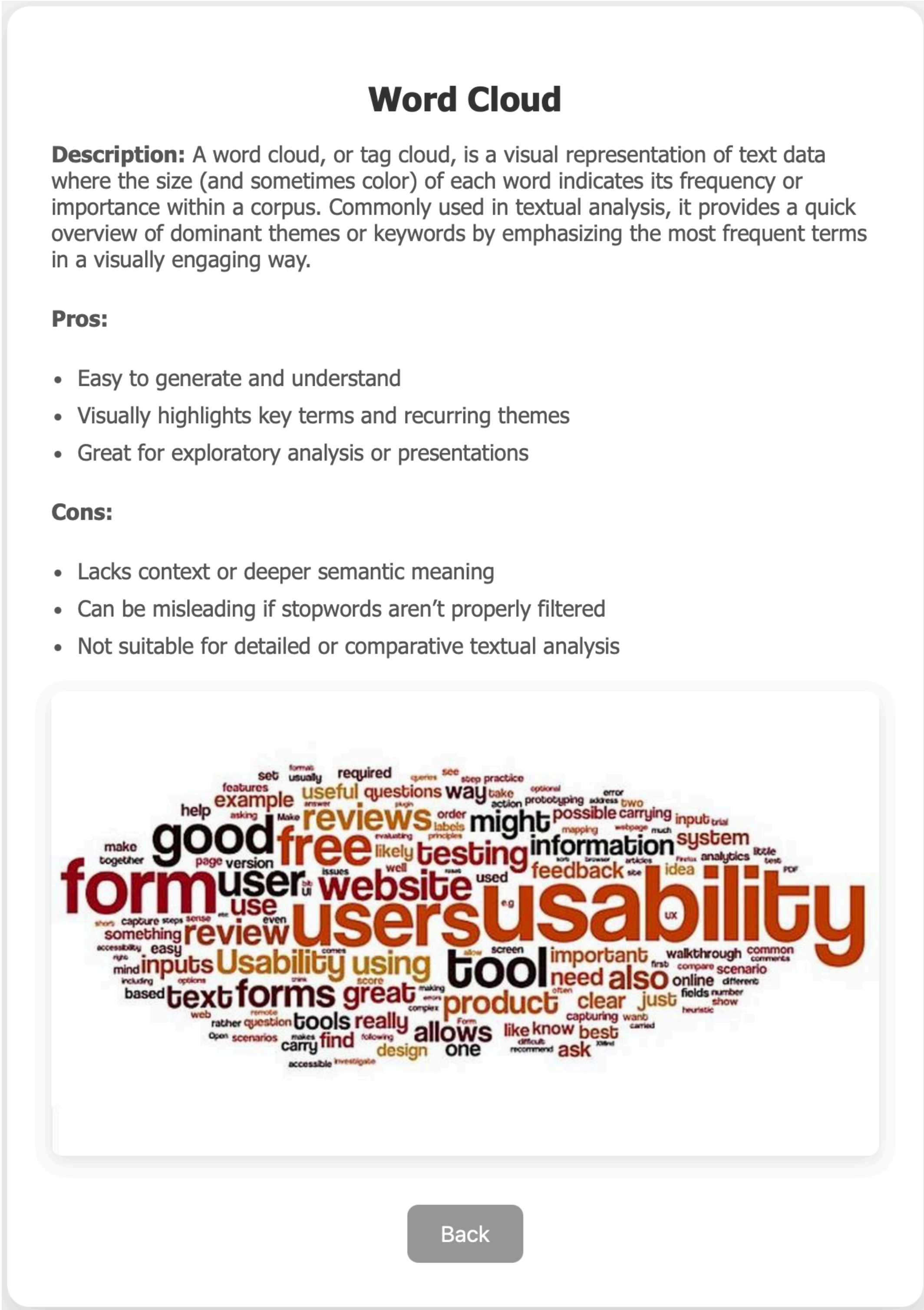

Figure 23. Detailed explanation of the Word Cloud visualization technique.

Word Cloud displays the most common terms in the dataset by adjusting the font size according to term frequency. Word clouds serve as an effective tool for large text collections because they help users recognize patterns quickly. The word cloud works efficiently to display major language patterns because the dataset has multiple records and focuses on main terms while omitting additional visual elements.

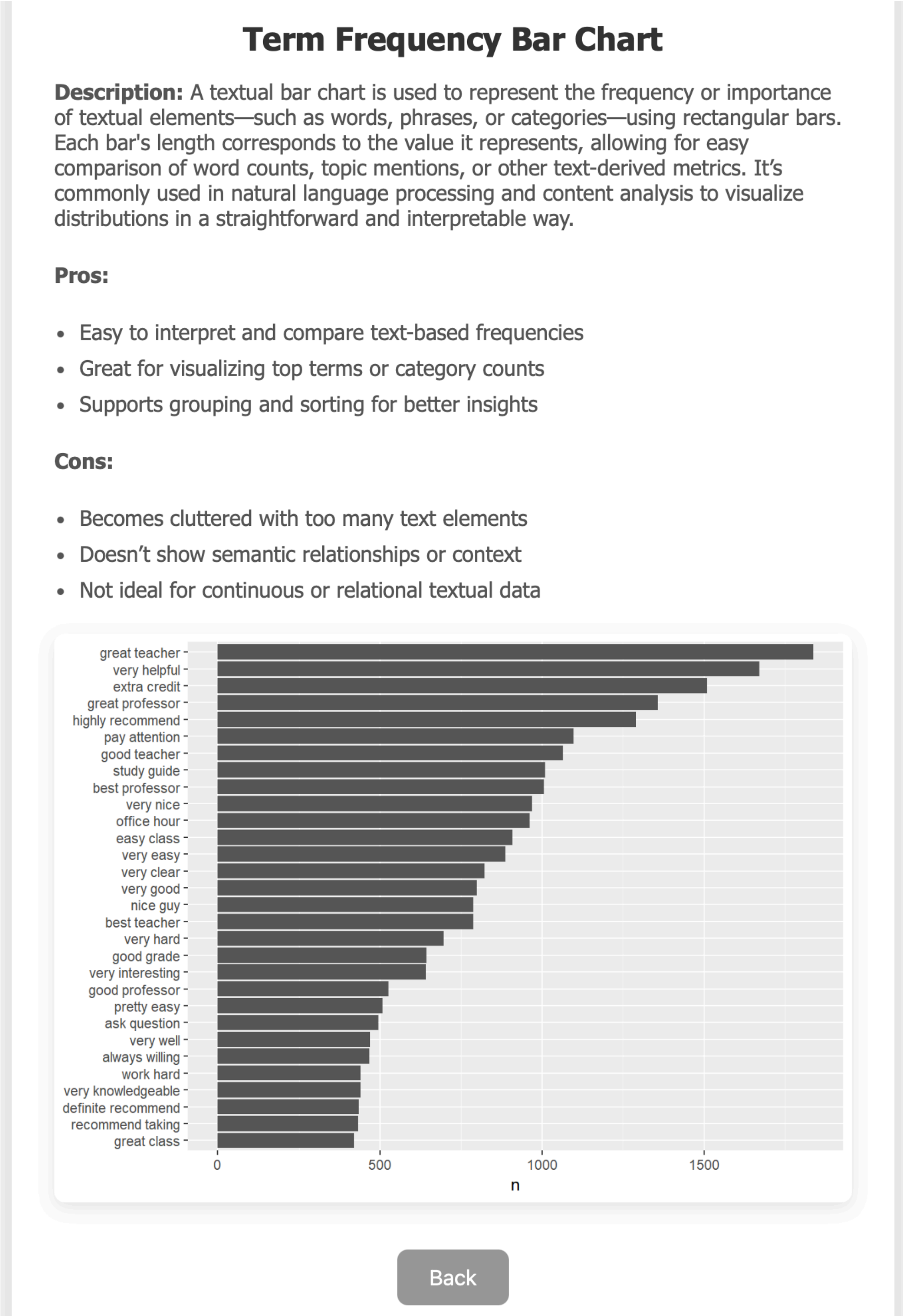

Figure 24. Term Frequency Bar Chart display with accompanying description, advantages, and limitations.

The Term Frequency Bar Chart displays words on the x-axis alongside their frequency values on the y-axis for exact term comparisons. Although it lacks the visual appeal of word clouds the bar chart presents exact frequency values and rank orders with high precision.

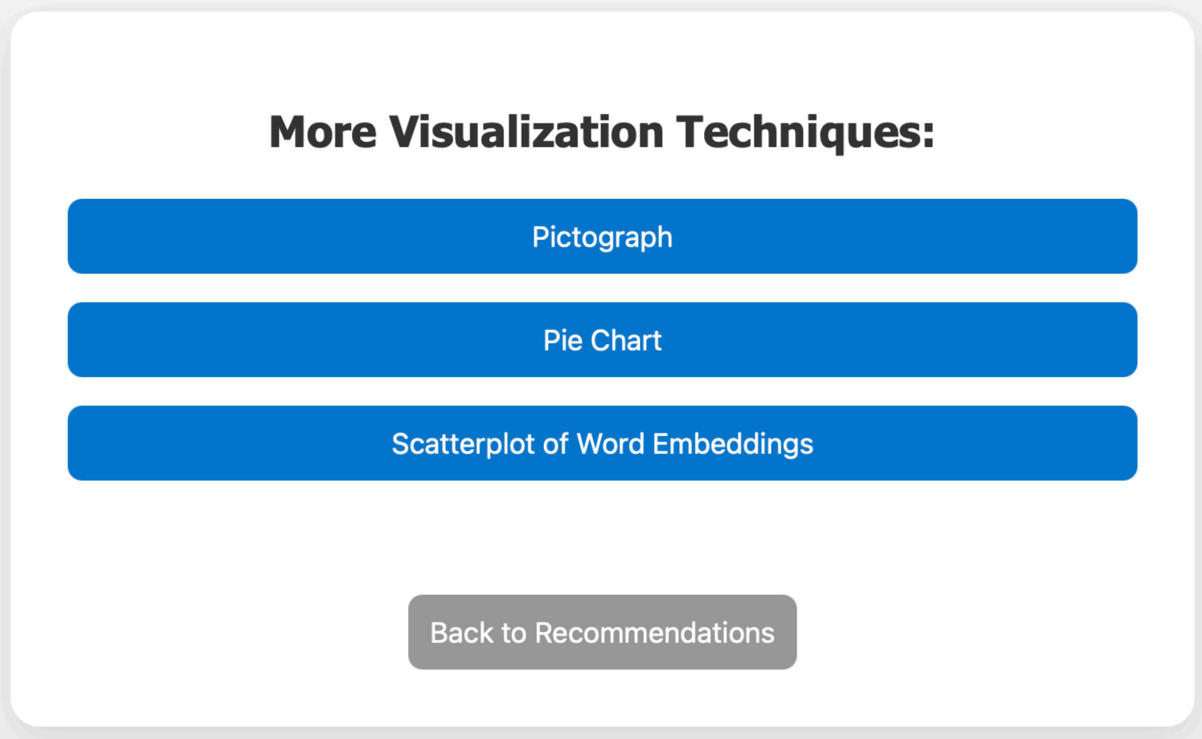

Figure 25. Alternative visualization techniques displayed after primary recommendations.

The system provides additional visualization methods which serve adjacent analytical requirements or presentation demands. The visualization methods do not optimize for the top-term frequency task but can serve as alternatives for exploratory or complementary purposes based on text characteristics and output preferences.

4.2 User Testing and Execution

A small group of end-users conducted a formative evaluation of the system after its implementation to check the system's usability and the relevance of its recommendations. A total of five users participated in the user testing. These participants were chosen to provide a mix of users – for example, graduate students and professionals who are likely to use data but not visualization experts. Each participant was asked to use the recommendation system on a dataset of their choice, or on a dataset they are familiar with. Allowing users to test the system on real, personally relevant data ensured that the feedback would be based on real use cases rather than hypothetical use cases. Before the start, users were told the purpose of the system and how the Q&A interface works, but they were not given any guidance on how to answer the questions other than what the interface itself told them. This was done in order to see how easy the system is to use for the first time.

The testing sessions were conducted individually. Participants either installed the prototype on their computer or used a provided laptop with the system pre-loaded. They then continued to work with the Q&A interface to obtain visualization recommendations for the dataset they chose. At the same time, they were told to think out loud – expressing any confusion, expectations, or reactions – in order to collect real time usability data. Each session was supposed to take about 10-15 minutes, although participants could spend more time if they needed to reflect on the questions or examine the results. The controller's "Try Again" feature was also available, and a few participants decided to run the system multiple times with slight variations of their inputs (for example, classifying their data differently) to see how recommendations changed. This gave us more insight into how the users understood the questions and logic of the system.

After they finished their interaction, participants were asked to give their feedback through a semi-structured interview. The interview focused on two main aspects: Usability – how easy and how much users liked the system and Satisfaction – whether the users found the recommendations helpful and relevant to their needs. Key questions included: "Was any part of the question-answer process confusing or difficult?", "Did the system's flow feel natural and cover the details you expected to provide?", "How do you feel about the visualization recommendation(s) you received – do they seem appropriate for your data and task?", and "Would you consider using such a recommendation tool in the future when you are unsure how to visualize data?". Users were also asked to give any additional thoughts that they had, for example, features that they liked or disliked, and suggestions for improvement. The notes were taken from each participant's responses, and with permission, their comments were recorded for accuracy in later analysis.

Since the feedback was open-ended and qualitative, the analysis of results involved thematic coding of the responses rather than numerical scoring. All interview transcripts or note sets were read to identify recurring themes or sentiments. Since the number of participants was small, a formal saturation analysis was not required; however, even with five participants, certain themes were observed. The evaluation was kept intentionally lightweight and exploratory – its purpose was to identify any major usability issues (if any) and to gauge the overall user perception of the system's value, rather than to measure precise performance metrics. This is in line with the best practices in usability testing, where five users are often enough to reveal the most of usability problems and get meaningful feedback in a qualitative study. Nevertheless, the participants' feedback offered a good understanding of how the expert-model recommendation system works in real life and how users respond to its guidance despite the small sample size.

4.3 Usability and User Experience

The qualitative feedback obtained from user testing revealed a predominantly positive response with specific themes about system usability and utility. The research data included participant comments which were sorted into categories representing their experience. This paper presents major findings grouped into two sections that analyze user system interaction (1) and evaluation of suggested visualization methods (2).

All participants found the system's interactive Q&A format to be highly intuitive and easy to follow. The participants mentioned that the step-by-step questioning resembled a dialogue with an expert who asked appropriate questions. The interface design showed users the next question one by one which helped them avoid information overload. The system achieved both efficiency and relevance by asking questions only after receiving prior responses. A system user stated that it "did not waste my time with pointless questions" because it concentrated on important aspects of my data. The successful application of branching logic was proven by the fact that users received textual-specific questions only when their data type was textual.

Multiple participants found valuable the integrated definitions for complex terms. The system provided simple definitions of technical concepts (data scaling, multivariate and preprocessing) through plain language descriptions in case of potentially unfamiliar terms. Users found these tooltips/tutorial texts informative. The participant praised the definitions for their concise and clear nature which provided sufficient information about the term. The system needed this feature because our users lacked experience with data science principles. The presence of these explanations contributed to the fact that no users experienced difficulties with any question since they understood the information requests. Users could infer the correct answer through system context or use the Back button to review previous inputs when they experienced uncertainty during testing. Users praised the Back button functionality because it allowed them to edit their answers freely while they explored the system's capabilities. The interface achieved this balance between structure and flexibility according to the participants.

Users appreciated how the system responded quickly to their interactions. Every user needed no more than a few minutes to obtain recommendations which caused them to comment about the tool's ability to integrate into their current work process. The users found the basic Streamlit GUI prototype functional, but some users mentioned that the interface design was plain. The tests revealed no major usability problems including software errors or crashes or inability to complete a task because all participants successfully received visualization recommendations on their own. The system's reliability enhanced the overall user satisfaction experience. A single user proposed future development of an interface with a more modern visual appearance through a web-based design with stylish presentation. The current design's simplicity-maintained focus on question and answer content which users appreciated despite aesthetic improvements being beyond the prototype scope. The usability testing demonstrated that the system delivers exceptional usability because it needs no training to learn and offers clear navigation with supportive guidance through the process.

4.4 Recommendation Relevance and User Satisfaction

Evaluating user satisfaction regarding the system's output of suggested visualization techniques proved to be an extremely complicated task. Subjectivity of user satisfaction exists naturally and numerous responses expressed satisfaction yet a few users provided reactions ranging from neutral to reserved. The varied feedback about the system allowed users to share their experiences with the system and its suggested visualizations.

Some users indicated that the proposed solutions matched their initial expectations regarding their data structures. During the testing phase with a financial time-series dataset (numerical, continuous data over time) the system recommended a Line Chart as the top output. This matched the visualization they had initially planned to use, and they noted that this "reassured them that the system's advice was in line with conventional practices." The positive alignment between the system's output and the user's decision reinforced their original choice but did not provide a new perspective.

A participant who analyzed survey data with categorical responses received a Bar Chart as one of their suggested visualizations. They agreed the recommendation was sensible yet added that "it's a pretty obvious option—something I already had in mind." The feedback demonstrates that the system will often support choices that expert users would select yet might fail to deliver fresh insights for people already experienced with typical visualization methods.

Some users benefited from the system by gaining knowledge of additional visualization methods. A participant who analyzed textual data through semantic relationship examination within a corpus remained uncertain about suitable visualization methods. The system provided Topic Modeling visualization through their input which revealed their interest in meaning and themes. The participant acknowledged they received the proposal as "a bit unexpected" at first yet accepted it as appropriate once they thought about it: "I wouldn't have thought of that, but it fits the task surprisingly well." Such user feedback demonstrates that the system effectively presents non-standard yet appropriate methods for data visualization.

Multiple participants expressed dissatisfaction about the system through their feedback. Several users criticized certain secondary suggestions because they appeared to be repetitive. A participant expressed doubt regarding showing both Bar Charts and Pie Charts for the same dataset because they displayed similar information through different visual structures. The system provided clear explanations about its rule-based operations to users but sometimes failed to demonstrate the reasoning behind specific technique recommendations. The system suggested the scatterplot possibly because I marked the data as dense according to the participant. The feedback demonstrates a regular pattern where users demanded more clarity about how their answers shaped the system's decision-making operations.

User satisfaction depended heavily on the amount of information provided with each suggested solution. Users who selected recommended techniques could view brief descriptions along with sample images for each visualization. Users found the provided descriptions helpful because they clarified when and why particular visualization techniques should be used. Users who were familiar with bar graphs and scatterplots received additional understanding about their function from the system descriptions. Users who received topic modeling visualizations or Treemap recommendations for their mixed categorical data gained fundamental understanding of these techniques through the provided descriptions. Multiple users described this feature through the analogy of "a mini handbook of charts" which they could access easily. The educational content increased user confidence in system suggestions which led them to properly implement recommended techniques.

Multiple recommendations available to users received positive feedback from the participants. Users preferred the functionality that allowed them to view additional recommendations after an initial suggestion failed to meet their expectations or design needs. During testing the "More Recommendations" feature displayed relevant visualization options to the survey data scenario by providing a Pie Chart after the Bar Chart recommendation. Users appreciated the secondary suggestions because they led to professional visualizations which matched their personal preferences. The system enables flexibility because it accepts that visualization can have multiple correct answers that differ based on the specific context. The system received positive user responses because it operated as a supportive assistant rather than a strict decision-making tool.

The user responses mostly showed satisfaction, but some users did not share the same level of enthusiasm. People experience satisfaction differently because it depends on their background knowledge and expectations together with their individual tastes. Users who were already familiar with visualization methods discovered that the system mainly confirmed their current choices without presenting new alternatives. The system served as a beneficial reference for users who lacked confidence in visualization techniques. All users agreed that the recommendations produced by the core rule-based engine were relevant which indicates the system's effective functioning.

The system successfully fulfills its main objective by providing helpful visualization choices to users who lack expertise according to the results obtained from the small sample measurement of subjective satisfaction. A combination of guided questions and expert-encoded rules with descriptive feedback enables users to confirm their choices and discover additional possibilities. Future development should focus on making the system more transparent and personalized.

4.5 Observed Strengths

Analyzing the user feedback and observing how the system was used allowed us to identify several strengths of the approach:

- **Intuitive Guided Workflow:** The Q&A expert model proved to be an intuitive way for users to articulate their data characteristics. The structured question flow acts as a built-in guide, helping users think through important aspects of their data. This not only makes the system easy to use but also has a side benefit of educating users on what factors are relevant to visualization choice (as one user put it, "answering the questions made me consider my data in a new light"). The ability to go back and adjust answers without frustration reinforced this positive experience.
- **Relevant and High-Quality Recommendations:** The rule-based engine consistently produced visualization suggestions that were appropriate for the given data scenarios. The close alignment between the system's recommendations and what experts or knowledgeable users might choose is a major strength. In cases where users were unsure how to visualize their data, the system's suggestions provided clear direction. In cases where users already had an idea, the system either confirmed that idea or offered a viable alternative. This demonstrates the effectiveness of encoding expert knowledge into the system - it can emulate expert advice reliably for a variety of inputs.
- **Knowledge Base Coverage and Detail**: The breadth of the visualization technique repository (knowledge base) ensured that the system could handle a range of common data types and goals. From basic charts (like bar, line, pie) to more advanced or specialized visuals (like heat maps or topic model diagrams), the system's database had entries spanning these needs. Each entry's inclusion of descriptive text and (where available) an example image added depth to the

recommendation. Users effectively had access to a mini library of visualization methods, which strengthened their understanding and confidence in using the recommended techniques. This richness of content is a strength that goes beyond a simple rule output - it provides context and learning material.

- **Lightweight and Accessible Implementation:** Although not a direct user-facing attribute, the choice of a Python-based MVC design with a potential for database integration (MongoDB) means the system is relatively easy to deploy and extend. From the evaluation perspective, the system ran smoothly on standard hardware and did not require any complex installation for users (aside from having a running Python environment). The speed with which recommendations were produced was essentially instantaneous after the last question, which users appreciated. This technical robustness underpins the positive user experiences described above; no participant encountered lag or crashes, reflecting the stability of the implementation.
- **Positive User Reception and Confidence:** An intangible but important strength observed was the boost in user confidence when using the system. As evidenced by their feedback, users felt more assured in their choice of visualization after consulting the system. This psychological benefit - reducing the doubt or second-guessing that often accompanies design decisions - is a key goal in developing decision-support tools. The fact that users indicated they would use such a system in the future (or recommend it to colleagues who struggle with visualization choices) is a testament to its real-world applicability and value.

5 SUMMARY AND FUTURE WORK

The final section of this thesis presents a summary of the main contributions together with essential findings and results from the Data Visualization Technique Recommendation System. The system design philosophy and technical implementation and evaluation outcomes work together to solve the visualization technique selection problem for different datasets. The system connects theoretical visualization principles to practical user-friendly applications through its design. The chapter outlines potential future development directions which include database expansion and rule transparency improvement and web deployment transition to maintain system adaptability and extensibility and alignment with changing data visualization practices and user needs in various domains.

5.1 Summary

The research began with the recognition of a persistent challenge in the field of data analytics: The selection of an appropriate visualization technique is still an underexplored problem, especially for novice users or non-specialists who often struggle to navigate the growing landscape of visual representation options. By addressing this gap, the project contributed a user-centered, knowledge-based recommendation system that integrates expert knowledge with an accessible and dynamic decision-making interface.

The system was conceptualized and implemented using a modular and scalable Model–View–Controller (MVC) architecture, ensuring that its components—knowledge base, rule engine, and user interface—function coherently yet independently. This design decision supports the long-term maintainability and expandability of the system. The backend architecture is grounded in Python and MongoDB, providing a robust and flexible structure to store and retrieve visualization rules and metadata. This configuration was essential for supporting the system's adaptive logic and enabling it to respond intelligently to a diverse range of dataset properties.

At the core of the system lies a rule-based reasoning engine. The rule base was not merely an algorithmic tool but a codified expert system that encapsulated decades of best practices from the field of data visualization. Each rule was meticulously crafted, refined, and tested using both real-world and synthetic datasets to ensure accuracy, robustness, and contextual appropriateness. Rules were designed to reflect not only data attributes—such as data type, dimensionality, or density—but also the analytical objectives of the user, whether they sought to identify outliers, visualize group distributions, or extract semantic insights from textual data.

The system's front-end was designed with human-computer interaction (HCI) principles in mind. The interactive Q&A interface guided users through a sequence of well-structured and context-sensitive questions, tailored dynamically based on their previous responses. This adaptive questioning strategy not only minimized cognitive load but also increased user engagement and reduced the likelihood of irrelevant or overly technical inquiries. Notably, the inclusion of definitions, tooltips, and examples supported users in understanding complex terminology and concepts, making the system both functional and educational.

The usability testing conducted with a sample of five users provided strong evidence of the system's value. Participants ranged from graduate students to non-expert professionals, all of whom successfully completed the interaction and expressed positive sentiment about the experience. Importantly, the system was shown to be effective both as a recommendation tool and as an educational resource. Participants not only received appropriate visualization suggestions but also learned more about when and why those visualizations are used. Some users confirmed their prior expectations, while others discovered new and unexpected—but contextually relevant—alternatives, illustrating the system's dual capacity for validation and exploration.

The interpretability of recommendations was another strong feature noted by users. Through the transparent linkage between dataset characteristics and rule-based reasoning, users were able to understand not just what visualization was recommended, but why it was recommended. The ability to explore alternative recommendations further contributed to the system's flexibility and demonstrated that it did not impose rigid decisions, but rather facilitated informed choices. Users described the system as "supportive," "confidence-boosting," and even likened it to having "a mini visualization expert" available during the data exploration process.

From a technical standpoint, the system proved lightweight and robust. It was deployable on standard computing environments, exhibited no performance bottlenecks, and required minimal user training. These properties render it a practical candidate for integration into broader analytics pipelines, educational platforms, or even enterprise-level dashboarding tools.

Importantly, the research validated its original hypotheses. It showed that expert visualization knowledge could be successfully encoded into a rule-based, object-oriented structure and retrieved dynamically through a user-driven interface. It confirmed that non-expert users could interact with the system intuitively, gain value from the recommendations, and increase their visualization literacy in the process. The alignment of system recommendations with both common visualization scenarios and less conventional yet valid alternatives demonstrates the effectiveness of the design and its underlying logic.The study uncovered additional significant effects of the research. The system connects data insights to visualization techniques which leads to better understanding of data and stronger analytical communication and data-driven decision making. These essential competencies benefit analysts and decision-makers who work in business as well as healthcare education journalism and public policy domains.

This system establishes a foundation for inclusive data communication through its democratized visualization best practices. The system establishes a basic solution that enables additional academic and practical developments. The system shows how expert models can be used to develop decision-support tools and demonstrates how software systems can combine educational and functional objectives while implementing scalable architectures for adaptable recommendation methods. The system resolves a particular technical issue while providing research opportunities for intelligent visual analytics development.

5.2 Reflective Conclusion

The findings from the system implementation and user evaluation demonstrate that a rule-based expert model proves useful for helping non-experts make suitable data visualization choices thus closing the knowledge gap between experts and non-experts in this field. The system development process fulfilled the main research targets. The research showed how object-oriented rules can maintainable track multiple visualization methods with their application rules thus fulfilling Research Question 1. The combination of structured technique data models with rule-based scoring enabled expert knowledge storage that users could retrieve dynamically through their input. The system operates as planned thus verifying that using an expert model works for this particular task.

The development of user-friendly recommendation interfaces received positive usability feedback which confirms the successful creation of guided recommendation systems (Research Question 2 regarding UI design for recommendation acquisition). Untrained users found the Q&A interface easy to navigate which proved the design successfully reduced expert-quality advice accessibility barriers for non-experts. The simple interface implemented essential interaction principles which transformed complex decision-making into a basic dialogue through guided steps and contextual help. The approach directly supports the thesis objective of simplifying data visualization selection.

The assessment results show that users benefit significantly from system recommendations during their visualization selection process (The research question investigates how useful and effective the system is). Users used the system without problems while also obtaining confidence and valuable insights from its recommendations. The system fulfilled its decision support function by raising user confidence about technique selection and extending their analysis options. The system successfully met its purpose by providing decision support so users obtained more accurate assurance about technique suitability while discovering additional options. The system proved its effectiveness at improving data literacy which aligns with the research objectives. Although the user study had a limited participant number it verifies the hypothesis about expert model recommendation systems' utility as predicted by the research.

A critical evaluation of limitations needs to occur in relation to research objectives. We need to use caution in extending generalization because of the qualitative feedback but the uniformity of responses strengthens confidence about the system meeting actual needs. The assessment of usability and satisfaction through the system yielded positive findings because users showed satisfaction and the system functioned well in its initial prototype. One objective was to validate that recommendations would be appropriate to the context which the system achieved during test scenarios to meet the requirement about user data attribute alignment.

This study establishes the concept that expert knowledge transformed into rule-based systems can be effectively transmitted through interactive applications to enhance user task performance. The successful development of the prototype demonstrates that such recommendation systems have the potential to be incorporated into data analysis pipelines and educational resources. Data analysis platforms should implement a "suggest a visualization" function which operates from a logic framework resembling the developed system to support users during analysis tasks. The positive user responses indicate that there exists a demand for assistance when selecting visualizations since users received the guidance favorably thus validating the known issue with visualization selection.

The research developed an operational expert model recommendation system and demonstrated its benefits by testing it with users. Users experienced high satisfaction and received expert-level visualization recommendations through the system because it provided an intuitive interface. The achieved results validate the approach's functionality and create foundations for upcoming improvements. The research answers all three questions positively because yes we can store visualization techniques through object-oriented rules and yes, we can build an interface for recommendation retrieval and yes users find these recommendations useful thus offering a new approach to visualization technique selection. This work enables future research about system extension as well as large-scale testing while it establishes a fundamental advancement that allows users to produce effective data visualizations both confidently and easily.

5.3 Future Work

The current prototype maintains its effective functionality through an expert-guided framework which enables multiple directions for future development. The proposed directions aim to increase system capabilities while improving usability and scalability and reach. The system will transform into a more robust tool which serves a wider user base and adapts to various use cases by implementing the following improvements.

1) **Expansion of the Knowledge Base and Domain Coverage**: The system's usability needs improvement for multiple domains and data types through an extensive expansion of its underlying database of visualization techniques. The database should include multiple data formats which extend beyond standard numeric data to include temporal data (time series) and hierarchical data (organizational structures or taxonomies) and geospatial data (maps) and multimedia or textual data. The system should include domain-specific visualization methods which include genomic data visualizations for bioinformatics and advanced financial charts for finance and network graphs for social network analysis. The recommendation engine will stay relevant for professionals in specialized fields through the integration of techniques used in those areas. The system's practical applicability and decision-support capabilities will increase through an expanded rule base which includes various data characteristics and visualization types.

2) **Enhanced Personalization of the Q&A Interface:** The Q&A Interface will receive additional personalization features through future development to achieve more adaptive and personalized interactions. The system already uses prior answers to customize questions but future versions will expand this capability by creating conditional sub-questions which probe specific details only when particular answer combinations suggest additional investigation. The system will create individualized question paths which match each user's specific situation. By choosing default answers and reordering questions based on its history of resolving questions, the system would be able to improve future sessions by identifying normal question pathways of previous user responses and common user selected answers. The front-end needs more guidance features to improve the question and answer. The front-end requires additional guidance features to enhance the Q&A process. The system would display additional tooltips and pop-up explanations and mini-tutorials which appear contextually during user question responses to enhance their understanding of their data and question reasoning. These improvements work together to create an interaction that feels like a smart assistant which learns from experience and adapts to individual needs thus enhancing user satisfaction and efficiency.

3) **Greater Transparency and Explanation of Recommendations:** In future versions, this packaged solution will be more transparent by providing more complete justifications for the reasons behind its recommendations. Features that present the rules that led to the recommendation and the rationale behind the recommended visualization will be included in the system.

4) **Web-Based Deployment for Scalability and Collaboration:** The system must migrate from its existing standalone (desktop-based Python/Streamlit) implementation to a web-based system to enhance its accessibility. Users will be able to access the recommendation system from any internet-connected device when the system is web-based deployment without needing additional software installation. The system becomes more accessible because of this change. A cloud-backed web application would support simultaneous system usage by multiple users which would enable collaborative work scenarios such as team analysts working together to select the best visualization for shared datasets. The system can be integrated with other tools through web deployment because it allows direct data retrieval from cloud storage and databases. A web architecture enables instant deployment of new features to all users while also supporting user count scalability and maintenance

and update scalability. The transformation of the prototype into a service would enable it to provide consistent visualization support to the entire organization.

5) **Extensive Evaluation and Optimization:** In addition to rigorous quantitative evaluation metrics, future system assessments should involve a far larger and more diverse set of users. A larger user study will help to verify that the system's benefits are true across different backgrounds from students to experienced professionals and across different domains of data analysis. In these evaluations, it would be useful to collect metrics such as task completion time, decision accuracy (e.g., did users choose an appropriate visualization with the system's help, as judged by experts), and changes in user confidence or knowledge before and after using the tool. The use of controlled experimentation can help isolate the attributes that contribute most to the overall performance of the system. A systematic approach to feedback will inform the design of the recommendation engine and user interface (UI). Testing the system with very large or complex data sets will also stress test the recommendation engine to see if it can maintain performance with larger sets. Such large-scale testing and iteration will not only strengthen the validity of the system in an academic sense but will also refine its functionality for real-world deployment.

6) **Multilingual Support:** Future development should include internationalization, and accessibility features to increase the system's reach. The system should include a multilingual interface to enable users who speak different languages to interact with the Q&A system in their native language thus expanding the global user base. The translation of questions and explanations and visualization technique descriptions into multiple languages and cultural adaptation of examples should be considered.

These future paths taken together seek to change the prototype into a mature, scalable, and very user-centric system. Each proposed enhancement either broadens the system's scope or deepens its capability: expanding the knowledge base and adding languages broadens who and what the system can handle; personalization, transparency, and accessibility deepen the quality of each user's experience; web deployment and extensive testing ensure the system can handle growth in users and use-cases.

References

[1] S. Few, Now You See It: Simple Visualization Techniques for Quantitative Analysis, Analytics Press, 2009.

[2] C. N. Knaflic, Storytelling with Data: A Data Visualization Guide for Business Professionals, Wiley, 2015.

[3] J. Heer, M. Bostock and V. Ogievetsky, "A Tour Through the Visualization Zoo," *Communications of the ACM,* 2010.

[4] J. Mackinlay, P. Hanrahan and C. Stolte, "Show Me: Automatic Presentation for Visual Analysis," *IEEE Transactions on Visualization and Computer Graphics,* 2007.

[5] E. R. Tufte, Visual Explanations: Images and Quantities, Evidence and Narrative, Graphics Press, 1997.

[6] E. R. Tufte, The Visual Display of Quantitative Information (2nd ed.), Graphics Press, 2001.

[7] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," *In Proceedings of the IEEE Symposium on Visual Languages,* 1996.

[8] J. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM Transactions on Graphics,* 1986.

[9] W. S. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *Journal of the American Statistical Association,* 1984.

[10] A. Key, B. Howe, D. Perry and C. Aragon, "VizDeck: Self-organizing dashboards for visual analytics," *In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data,* 2012.

[11] M. Vartak, S. Madden, A. Parameswaran and N. Polyzotis, "SEEDB: Efficient Data-Driven Visualization Recommendations to Support Visual Analytics," *In Proceedings of the VLDB Endowment,* 2014.

[12] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe and J. Heer, "Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco," *IEEE Transactions on Visualization and Computer Graphics,* 2019.

[13] D. Ricci, L. Rokach, B. Shapira and P. B. Kantor, Recommender Systems Handbook, 2011.

[14] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe and J. Heer, "Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations," *IEEE Transactions on Visualization and Computer Graphics,* 2017.

[15] Y. Kim, E. M. Peck and J. Heer, "Data-Driven Guides: Supporting Expressive Design for Information Graphics," *IEEE Transactions on Visualization and Computer Graphics,* 2016.

[16] M. Vartak, S. Huang, T. Siddiqui, S. Madden and A. Parameswaran, "Towards Visualization Recommendation Systems," *ACM SIGMOD Record,* 2018.

[17] C. Stolte, D. Tang and P. Hanrahan, "Polaris: A system for query, analysis, and visualization of multidimensional databases," *Communications of the ACM,* 2002.

[18] V. Dibia and C. Demiralp, "Data2Vis: Automatic Generation of Data Visualizations Using Sequence-to-Sequence Recurrent Neural Networks," *IEEE Computer Graphics and Applications,* 2019.

[19] K. Hu, M. A. Bakker, S. Li, C. Kraska and C. A. Hidalgo, "VizML: A Machine Learning Approach to Visualization Recommendation," *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* 2019.

[20] Y. Lu, R. Kruger, D. Thom, F. Wang, S. Koch, T. Ertl and R. Maciejewski, "DeepEye: Towards Automatic Data Visualization," *IEEE Transactions on Visualization and Computer Graphics,* 2018.