A Project Report

On

DESIGN AND EVALUATION OF
A B.I. DASHBOARD CREATION ASSISTANT

BY

MANAN MAYUR POPAT – 2020A7PS0029H
SAI HEMANTH ANANTHOJU – 2020A7PS0116H
NISHITH KUMAR – 2020A7PS0157H

Under the supervision of

PROF. NARASIMHA BOLLOJU

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

CS F376/377: DESIGN PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

HYDERABAD CAMPUS

(DECEMBER 2023)

ACKNOWLEDGMENTS

We would like to extend our sincere and heartfelt thanks to all those who have helped us in building this project so far. Without their active guidance, support, cooperation, and encouragement, we would not have been able to achieve this feat.

We express our sincere gratitude to our mentor, Prof. Narasimha Bolloju**, and the *Instructor-in-Charge, CS F376/377 (Design Project)*,** Prof. Dipanjan Chakraborty**, for their moral support and guidance during the tenure of this project.**

We also acknowledge, with a deep sense of reverence, our gratitude towards our parents and other faculty members of *BITS Pilani* for their valuable suggestions given to us from time to time.

Date: 22-12-2023

Manan Mayur Popat
Sai Hemanth Ananthoju
Nishith Kumar

Birla Institute of Technology and Science-Pilani,

Hyderabad Campus

Certificate

This is to certify that the project report entitled **“DESIGN AND EVALUATION OF A B.I. DASHBOARD CREATION ASSISTANT”** submitted by:

- Mr. MANAN MAYUR POPAT (2020A7PS0029H)
- Mr. SAI HEMANTH ANANTHOJU (2020A7PS0116H)
- Mr. NISHITH KUMAR (2020A7PS0157H)

in partial fulfillment of the requirements of the course CS F376/377, Design Project Course, embodies the work done by them under my supervision and guidance.

Date: **(PROF. NARASIMHA BOLLOJU)**

BITS - Pilani, Hyderabad Campus

ABSTRACT

Business Intelligence (BI) is the process of using software or systems to transform business data into insights that can improve decision-making. BI tools include reporting and query software, data visualization software, data mining tools, online analytical processing tools, and self-service business intelligence tools.

Business Intelligence Dashboards are data visualization tools that show the KPIs and other metrics of an organization, department, or process. They use charts, graphs, maps, tables, and other forms to display information in a clear and concise way. They can be tailored to suit the needs and goals of different users and audiences.

Creating effective dashboards is challenging, as it requires finding the optimal balance between functionality, aesthetics and usability. Some of the challenges are:

- Choosing the right metrics and KPIs that align with the dashboard objectives and the user needs.
- Organizing and presenting the data in a clear and consistent way that highlights the key messages and insights.
- Optimizing the performance and usability of the dashboard.
- Ensuring that the dashboard is interactive and user-friendly.

The project aims to address these challenges through the following contributions:

- To advance the knowledge and understanding of BI dashboards and dashboard design methods and tools by providing a comprehensive and systematic literature review on these topics.
- To propose a novel and innovative approach to dashboard design to provide customized and context-aware guidance and feedback to dashboard designers throughout the design process.
- To design a prototype of a BI dashboard design assistant that demonstrates its effectiveness and usability in supporting dashboard designers in different domains and scenarios and proposing evaluation strategies.
- To propose the prototype design for development of a tool that can be used by dashboard creators, and the evaluation methods suggested could be used for training the architected model.
- To provide practical implications and recommendations for dashboard designers, dashboard users, dashboard developers, dashboard researchers, and other stakeholders who are interested or involved in BI dashboards and dashboard design.

Presently, we have derived a set of guidelines that a machine learning model can follow optimally to generate logical and coherent Power BI reports. We have also proposed a flow diagram (with the steps in detail) that can be used for developing an automatic dashboard creation tool and commonly accepted dashboard design evaluation strategies.

Keywords: BI Dashboard Design, Design Assistant, Machine Learning, Usability, Reinforcement Learning, VizML, Power BI, Dashboard Optimization

CONTENTS

1. INTRODUCTION

Business Intelligence (BI) refers to the use of software or systems to collect, store, analyze and present business data in a way that enables users to make better decisions. BI can help businesses understand their Key Performance Indicators (KPIs), identify problems and opportunities and discover trends in the patterns in their data. There are various BI tools available, namely, Reporting and Query Software, Data Visualization Software (Dashboards, Maps, Infographics, etc.), Data Mining tools, Online Analytical Processing (OLAP) tools and Self-Service Business Intelligence (SSBI) tools. Using these tools, businesses can gain a competitive advantage by taking note of valuable insights that can improve their performance.

Business Intelligence Dashboards are Data Visualization tools that help organizations understand their Key Performance Indicators (KPIs) and other metrics related to their department, business or process. BI Dashboards display information using various forms such as charts, graphs, maps or tables, to help users understand, share and collaborate on their information. BI Dashboards can be customized to meet the specific needs and goals of different users and audiences.

There are multiple approaches to designing dashboards:

- Goal-Oriented Approach: Defining the goals and objectives and selecting metrics and KPIs that align with them.
- User-Centered Approach: Understanding the needs, preferences and expectations of dashboard users and designing accordingly.
- Data-Driven Approach: Analyzing the data and discovering the key messages and insights it reveals.
- Innovative Approach: Leveraging AI techniques to provide customized and context-aware guidance and feedback to Dashboard Designers.

Designing Effective Dashboards is a challenging task, as it requires a balance between functionality, aesthetics and usability. Some of the common challenges in creating effective dashboards are:

- Choosing the right metrics and KPIs that align with the dashboard objectives and the user needs. Different users may have different goals and expectations from the dashboard, so it is important to understand the audience and their requirements.
- Organizing and presenting the data in a clear and consistent way that highlights the key messages and insights. There are many types of charts, graphs, maps and tables that can be used to show data, but not all of them are suitable for every scenario. The choice of visualization techniques and tools should depend on the data type, format and distribution.
- Optimizing the performance and usability of the dashboard. The dashboard should be connected to reliable and updated data sources and have fast loading and refreshing times. It should also have a simple and intuitive interface that makes it easy to navigate and understand.

This project aims to tackle these challenges by designing a prototype (architectural proposal) of an assistant for creation of BI Dashboards. The assistant can have the following features to help users create effective dashboards:

1. The assistant can help users define their dashboard objectives and select the most appropriate metrics and KPIs for their needs. The assistant could ask users a series of questions to understand their goals, preferences and constraints, and then suggest a list of relevant metrics and KPIs that match their criteria.

2. The assistant can suggest the best visualization tools for displaying the data, based on the data type, format and distribution. The assistant can analyze the data and recommend the most suitable charts, graphs, maps or tables that can effectively communicate the data.

The scope of the project includes figuring out the ***steps in the architecture (flow) for creating a prototype*** that can be used to build an intelligent assistant that assists users in designing effective and visually compelling business intelligence dashboards, which is also capable of analysing dimensional models represented as star schemas. It can provide recommendations for optimal design layouts and suitable visualizations specifically tailored for <u>operational dashboards</u>. The project will conduct a systematic literature review to provide a comprehensive understanding of BI dashboards, dashboard design methods, and relevant tools. This literature review will serve as a foundation for the development and evaluation of the assistant.

Moreover, the project aims to ***propose methods for evaluating*** the effectiveness and usability of the intelligent assistant in supporting dashboard designers across different domains and scenarios. This evaluation would employ a mixed-methods approach, combining experimental, observational, and survey methods. Thus, the project will culminate in the provision of practical implications and recommendations for various stakeholders, including dashboard designers, users, developers, and researchers, interested or involved in BI dashboards and dashboard design. At a high level, the project aims for the following outcomes:

- To advance the knowledge and understanding of BI dashboards and dashboard design methods and tools by providing a comprehensive and systematic literature review on these topics.
- To propose a novel and innovative approach to design a dashboard creation prototype for providing customized and context-aware guidance and feedback to dashboard designers throughout the design process.
- To design a prototype of a BI dashboard assistant that demonstrates its effectiveness and usability in supporting dashboard designers in different domains and scenarios.
- To provide practical implications and recommendations for dashboard designers, dashboard users, dashboard developers, dashboard researchers, and other stakeholders who are interested or involved in BI dashboards and dashboard design.

In this section, we first explained what BI dashboards are and why they matter for decision-making and discussed the difficulties in designing effective dashboards. After that, we will outline the specific objectives of our project. Following this, we would go on to review what we found in the existing literature in the background chapter. Then, our research questions are presented and how we approached them. This is followed by a summary of the guidelines extracted from the literature review. Further in the report, in the prototype design and evaluation section, we will show you what we have built, represented in the form of a flow diagram, and how it works (various steps involved, in detail). Finally, the report shall be concluded by summarizing our work, highlighting key findings, mentioning any limitations, and suggesting future areas of exploration.

2. BACKGROUND

Literature Review

Business Intelligence (BI) is a set of technologies, processes, and practices that enable organizations to collect, integrate, analyse, and present data to support decision-making and improve performance. BI is used in various domains such as finance, healthcare, marketing, and manufacturing to gain insights into business operations, customer behaviour, market trends, and other critical aspects of the business. BI involves several stages such as data extraction, data transformation, data modelling, data analysis, and data visualization. One of the key components of BI is the dashboard, which is a visual representation of key performance indicators (KPIs) that provide a quick overview of the business performance.

Dimensional modelling is a technique used in BI to organize data in a way that makes it easier for businesses to analyse and gain insights from their data. Dimensional models represent business processes or events as dimensions (such as time or geography) and facts (such as sales or revenue). The dimensional model is used as a basis for creating data marts or cubes that can be used for reporting or analysis purposes. Dimensional modelling provides several benefits such as improved query performance, simplified data access, and enhanced user experience, as highlighted in a popular publication by IBM Redbooks **[1]**.

Dashboards are an essential component of BI that provide a quick overview of the business performance using visualizations such as charts, graphs, tables, and gauges. Dashboards are designed to be user-friendly and intuitive so that users can quickly identify trends, patterns, and anomalies in the data. However, designing an effective and user-friendly dashboard is not a trivial task as it requires a balance between functionality, aesthetics, and usability. Moreover, dashboard design often involves multiple stakeholders with different needs, preferences, and expectations which adds to the complexity and challenges of the design process. An online article by NetSuite **[2]** provides 23 real-world examples of how companies are using BI to gain insights into their business operations. The examples cover various domains such as finance, marketing, sales, and operations. Similarly, another article by DataCamp **[3]** provides nine examples of Power BI dashboards that cover a variety of topics such as sales, marketing, finance, and operations. Thus, creating dashboards that cater to diverse stakeholder needs and align with the ever-evolving landscape of business intelligence demands a thoughtful and adaptable approach to design.

Dimensional modelling is commonly used in BI to organize data in a way that makes it easier for businesses to analyse and gain insights from their data. Several methods, frameworks, principles, and tools have been proposed (**[4]**, **[5]**) to assist dashboard designers in creating dashboards that meet the requirements and expectations of the users. Several research papers (**[6]**, **[7]**, **[8]**) have highlighted the use of dimensional modelling in BI and provided best practices for creating a dimensional model using dataflows. Additionally, many of these have also highlighted examples of dashboards for BI tasks in real-life industries.

Several methods, frameworks, principles, and tools have been proposed to assist dashboard designers in creating dashboards that meet the requirements and expectations of the users. A study by Few **[9]** proposed 13 principles for designing dashboards based on perceptual and cognitive psychology. The principles include displaying the most important information prominently, using appropriate visualizations, minimizing clutter, and providing context for the data. The book provides practical guidance on how to apply these principles to create effective dashboards. Yigitbasioglu *et al.* **[10]** conducted a review of dashboards in performance management and identified several implications for design and research. The authors highlighted the importance of involving users in the design process, using appropriate visualizations, providing interactivity, and ensuring data quality.

Lin *et al.* **[11]** developed 5S dashboard design principles for self-service business intelligence tool users. The principles are seeing both the forest and trees, simplicity through self-selection, simplicity through significance, simplicity through synthesis, and storytelling. These principles aim to help non-IT professionals and casual users to design insightful dashboards efficiently by using self-service BI tools. Janes *et al.* **[12]** proposed an automatic generation of dashboards for analysing business processes performance. The authors developed a tool that automatically generates dashboards based on business process models. The tool uses data mining techniques to identify relevant data sources and generate visualizations that provide insights into business processes performance.

Koopman *et al.* **[13]** proposed a methodology for evaluating dashboards based on user characteristics such as experience, expertise, and task complexity. The authors conducted a study to evaluate the effectiveness of the methodology and found that it can help identify usability issues and improve dashboard design. Abduldaem *et al.* **[14]** proposed principles for designing dashboards using goal question metrics approach based on a literature review. The authors identified several critical factors such as data quality, user needs, visual design, interactivity, and performance that should be considered in dashboard design.

In conclusion, designing an effective dashboard requires careful consideration of several factors such as functionality, aesthetics, usability, data quality, user needs, visual design, interactivity, performance, etc. Researchers have proposed several methods, frameworks, principles, and tools to assist dashboard designers in creating dashboards that meet these requirements and expectations.

However, most of the existing methods and tools for dashboard design are either too general or too specific to be applicable to different contexts and scenarios. For instance, Few's principles **[9]** are based on general guidelines that do not consider the specific characteristics and goals of each dashboard project. On the other hand, Lin *et al.*'s principles **[11]** are tailored for SSBI tool users in the health care domain which may not be suitable for other domains or tools. Therefore, there is a need for a more adaptive and personalized approach to dashboard design that can accommodate the diversity and complexity of dashboard projects.

One possible solution is to develop a BI dashboard design assistant that can provide customized and context-aware guidance and feedback to dashboard designers throughout the design process. The BI dashboard design assistant can leverage various AI techniques such as machine learning

(ML), Reinforcement Learning (RL), natural language processing (NLP), computer vision (CV), etc., to provide personalized recommendations based on user input. The BI dashboard design assistant can also use dimensional modelling techniques to organize data in a way that makes it easier for businesses to analyse and gain insights from their data.

Design Criteria for Dashboards

Designing an effective and user-friendly dashboard is a challenging task that requires a balance between functionality, aesthetics, and usability. Several researchers have proposed methods, frameworks, principles, and tools to assist dashboard designers in creating dashboards that meet the requirements and expectations of the users. One such research **[15]** introduces design patterns for dashboards to inform dashboard design processes. Based on a systematic review of 144 dashboards, the authors report on eight groups of design patterns that provide common solutions in dashboard design. The eight groups of design patterns are:

- **Overview**: Provides a high-level summary of the data.
- **Filter**: Allows users to interactively filter the data.
- **Comparison**: Allows users to compare different data points.
- **Composition**: Combines multiple visualizations into one.
- **Distribution**: Shows the distribution of data.
- **Evolution**: Shows how data changes over time.
- **Relationship**: Shows the relationship between different data points.
- **Annotation**: Provides additional information about the data.

The authors also discuss combinations of these patterns in "dashboard genres" such as narrative, analytical, or embedded dashboard. The paper aims to support dashboard designers and researchers in co-creation, structured design decisions, as well as future user evaluations about dashboard design guidelines.

A comparative analysis by Kaur *et al.* **[16]** compares different approaches to dashboard design such as traditional approach, agile approach, and user-centred approach. The authors highlight several design criteria for dashboards such as simplicity, clarity, consistency, relevance, interactivity, and aesthetics. The authors also emphasize the importance of involving users in the design process to ensure that the dashboards meet their needs and expectations.

In conclusion, designing an effective and user-friendly dashboard requires careful consideration of several factors such as functionality, aesthetics, usability, simplicity, clarity, consistency, relevance, interactivity, and involvement of users in the design process. Researchers have proposed several methods, frameworks, principles, and tools to assist dashboard designers in creating dashboards that meet these requirements and expectations. Design patterns for dashboards can provide common solutions in dashboard design and can be used to inspire designs and discuss trade-offs in screen space, interaction or information shown.

Associated Challenges

Designing an effective and user-friendly dashboard is a challenging task that requires a balance between functionality, aesthetics, and usability. Several researchers have pointed out the following challenges associated with effective dashboard design:

1. **Data quality**: Dashboards rely on data to provide insights and support decision-making. Therefore, ensuring the quality, accuracy, and completeness of the data is critical for effective dashboard design.
2. **Data overload**: Dashboards can quickly become cluttered with too much information, making it difficult for users to identify the most important insights. Therefore, it is essential to prioritize the information and present it in a clear and concise manner.
3. **User needs**: Dashboards are designed for specific user groups with different needs, preferences, and expectations. Therefore, it is essential to involve users in the design process to ensure that the dashboards meet their needs and expectations.
4. **Visual design**: Dashboards rely on visualizations such as charts, graphs, tables, and gauges to present information. Therefore, it is essential to use appropriate visualizations that are easy to read and interpret.
5. **Interactivity**: Dashboards should be interactive to allow users to explore the data and gain insights. However, too much interactivity can lead to confusion and overwhelm users.
6. **Performance**: Dashboards should load quickly and respond promptly to user interactions. Slow performance can lead to frustration and disengagement from users.

In conclusion, designing an effective dashboard requires a balance between functionality, aesthetics, and usability while addressing several challenges such as data quality, data overload, user needs, visual design, interactivity, and performance. Researchers have proposed, as identified at various places in this section, several methods, frameworks, principles, and tools to assist dashboard designers in creating dashboards that meet these requirements and expectations.

3. RESEARCH QUESTIONS AND METHODOLOGY

The project aims to develop a flow for a BI dashboard design assistant prototype that leverages various design principles highlighted in literature and patterns observed in insightful industrial dashboards, as well as AI techniques and technologies to provide customized and context-aware guidance and feedback to dashboard designers throughout the design process. The need for such an assistant arises from the complexity of BI dashboards and the need for effective design methods and tools. A comprehensive and systematic literature review on BI dashboards and dashboard design methods and tools was conducted to identify the key features, functionalities, and design principles that are essential for developing an effective BI dashboard design assistant. The expected benefits of the project include advancing the knowledge and understanding of BI dashboards and dashboard design methods and tools, proposing a novel approach to dashboard design, developing and evaluating a prototype design of a BI dashboard design assistant that implements the proposed approach, and providing practical implications and recommendations for dashboard designers, users, developers, researchers, and other stakeholders interested or involved in BI dashboards and dashboard design.

Good dashboard design principles highlighted in literature include:

- Consider your audience: Optimize the dashboard layout for your audience by considering how they use the dashboard, what key metrics help them make decisions, what learned or cultural assumptions might affect design choices, and what information they need to be successful.
- Tell a story on one screen: Dashboards are meant to show important information at a glance. Having all the tiles on one screen is best. Avoid scroll bars on your dashboard. Remove all but essential information.
- Use the right visualization for the data: Visualizations should paint a picture and be easy to read and interpret. Avoid visualization variety for the sake of variety.
- Place the most important information: Most people read from top to bottom. Put the highest level of data at the top left corner and show more detail as you move in the direction the audience uses for reading (left-to-right, top-to-bottom).
- Make it visually appealing: Use graphics, charts, and infographics to make the data easy to understand and engaging.

Requirements Gathering: The prototype of the BI dashboard design assistant is expected to provide customized and context-aware guidance and feedback to dashboard designers throughout the design process. It will leverage various AI techniques such as machine learning, natural language processing, computer vision, etc., to provide personalized recommendations based on user input. The prototype will be evaluated using various metrics such as effectiveness, efficiency, usability, user satisfaction, etc., using surveys, interviews, usability tests, etc.

Dimensional modelling is a way of organizing data in a database or data warehouse to make it easier for businesses to analyse and gain insights from their data. It involves creating a dimensional model that represents business processes or events as dimensions (such as time or geography) and facts (such as sales or revenue). The dimensional model is used as a basis for creating data marts or cubes that can be used for reporting or analysis purposes. In this project, dimensional modelling shall be used to organize data in a way that makes it easier for the BI dashboard design assistant to provide customized guidance and feedback based on user input.

The project aims to address two research questions:

1. **RQ1**: What are the requirements and specifications for developing a BI dashboard design assistant?

2. **RQ2**: How can various machine learning techniques, or openly available tools, be integrated to develop a BI dashboard design assistant?

The first research question (RQ1) aims to identify the requirements and specifications for developing a BI dashboard design assistant. This will involve conducting a comprehensive and systematic literature review on BI dashboards and dashboard design methods and tools. The review will help identify the key features, functionalities, and design principles that are essential for developing an effective BI dashboard design assistant. The results of this research question will provide a clear understanding of the scope and requirements of the project.

The second research question (RQ2) aims to explore how various open-source technologies based on machine learning can be integrated to develop a BI dashboard design assistant. This will involve identifying the most suitable machine learning models and similar techniques that can be used to provide customized and context-aware guidance and feedback to dashboard designers throughout the design process. The results of this research question will provide insights into the technical aspects of developing a BI dashboard design assistant.

The research methodology for this project involves a comprehensive and systematic literature review on BI dashboards and dashboard design methods and tools. The review was conducted to identify the key features, functionalities, and design principles that are essential for developing an effective BI dashboard design assistant. The review also helped identify the most suitable technologies that can be used to provide customized and context-aware guidance and feedback to dashboard designers throughout the design process. The literature review was conducted using various academic databases such as Google Scholar, IEEE Xplore, and online articles.

The next step will be to develop a prototype design of the BI dashboard creation assistant that implements the proposed approach. The prototype would be developed using various AI techniques and automation technologies including machine learning, natural language processing, and computer vision. The prototype can be evaluated using various metrics such as effectiveness, efficiency, usability, and user satisfaction. The evaluation could be conducted using various methods such as surveys, interviews, and usability tests. The results of the evaluation will be used to refine the prototype thus developed from the flow design and improve its effectiveness and usability.

4. GUIDELINES DERIVED FROM LITERATURE REVIEW

From the literature review conducted, we have derived a set of guidelines for a Machine Learning Model to follow optimally to create coherent Power BI reports. The guidelines are as follows:

1) **Connecting to different data sources:** The model must be able to connect to different data sources like text, CSV, Excel, databases or web services. For the purpose of our project, we are using data from MS Excel only.

2) **Preprocessing data:** The model must be able to pre-process the data using Power Query Editor, and perform operations such as cleaning, transforming, aggregating, and joining the data.

3) **Analyzing and identifying features:** The model must be able to analyze and identify relevant features, variables and dimensions required for visualization. The model should be able to categorize data types into numerical, categorical, temporal, and spatial, and choose appropriate visualizations accordingly.

   These visualizations are to be hard fed by the programmers. The current visualization scheme is:

   a) Numerical Data:
      - For continuous values, use Bar Graphs and/or Line Charts
      - For discrete values, use Violin Plots, Scatter Plots, and/or Box Plots

   b) Categorical Data:
      - Pie Charts
      - Dot Plots

   c) Temporal Data:
      - Area Charts
      - Heatmaps
      - Timelines

   d) Spatial Data:
      - Maps
      - Dot Maps and/or Bubble Charts

4) **Generating different types of Visualizations:** After choosing appropriate visualizations from the schema above, the model must be able to generate and customize the visualizations.

5) **Arranging the Visualizations:** The model must be able to arrange the visualizations in a logical and coherent manner in a Power BI report page.

For this to happen effectively, the model must be able to use grids, containers and white spaces appropriately to create visual hierarchy and alignment. This can effectively happen by using concepts from reinforcement learning.

We are expanding a bit on this topic. An optimal model must be able to:

a. Define the objective and scope of the report and select the most relevant and informative visualizations that support the goal and message of the report.
b. Use reinforcement learning to learn the optimal arrangement of visualizations on a report page, based on user feedback and business objectives. Consider factors such as readability, aesthetics, balance, and coherence when optimizing the layout.
c. Use containers to group related visualizations together and create a logical structure for the report. Containers can be rectangular, circular, or irregular shapes that enclose one or more visualizations. Containers can also have different styles, such as borders, colors, or shadows, to create contrast and emphasis.
d. Use grids to align and distribute visualizations on a report page. Grids can be horizontal, vertical, or diagonal lines that divide the page into equal or proportional segments. Grids can also have different sizes, shapes, or orientations to create variety and interest.
e. Use white spaces to create a visual hierarchy and alignment for the visualizations. White spaces are empty areas between or around visualizations that separate them from each other and from the edges of the page. White spaces can also have different widths, heights, or margins to create harmony and proportion.

Reinforcement learning is a machine learning technique that enables an agent to learn from its own actions and rewards in an environment. In the context of dashboard generation, the agent is the system that creates the dashboard, the actions are the choices of visualizations and their positions on the page, the environment is the report page and the data, and the rewards are the scores that evaluate the quality of the dashboard.

The agent can use different algorithms to learn the optimal policy, which is a function that maps each state (the current dashboard) to an action (the next visualization to add or move). One common algorithm is Monte Carlo Tree Search (MCTS), which builds a search tree that represents the possible actions and outcomes. MCTS starts from the root node (the initial dashboard) and iteratively expands the tree by selecting, simulating, evaluating, and updating nodes until a termination condition is met. The node with the highest value is then chosen as the best action.

The reward function is a crucial component of reinforcement learning, as it guides the agent to improve its policy. The reward function should reflect the design principles and user preferences for dashboard generation. For example, some possible criteria for rewarding a dashboard are:

- Expressiveness: The dashboard should use appropriate visualizations that accurately and effectively represent the data.

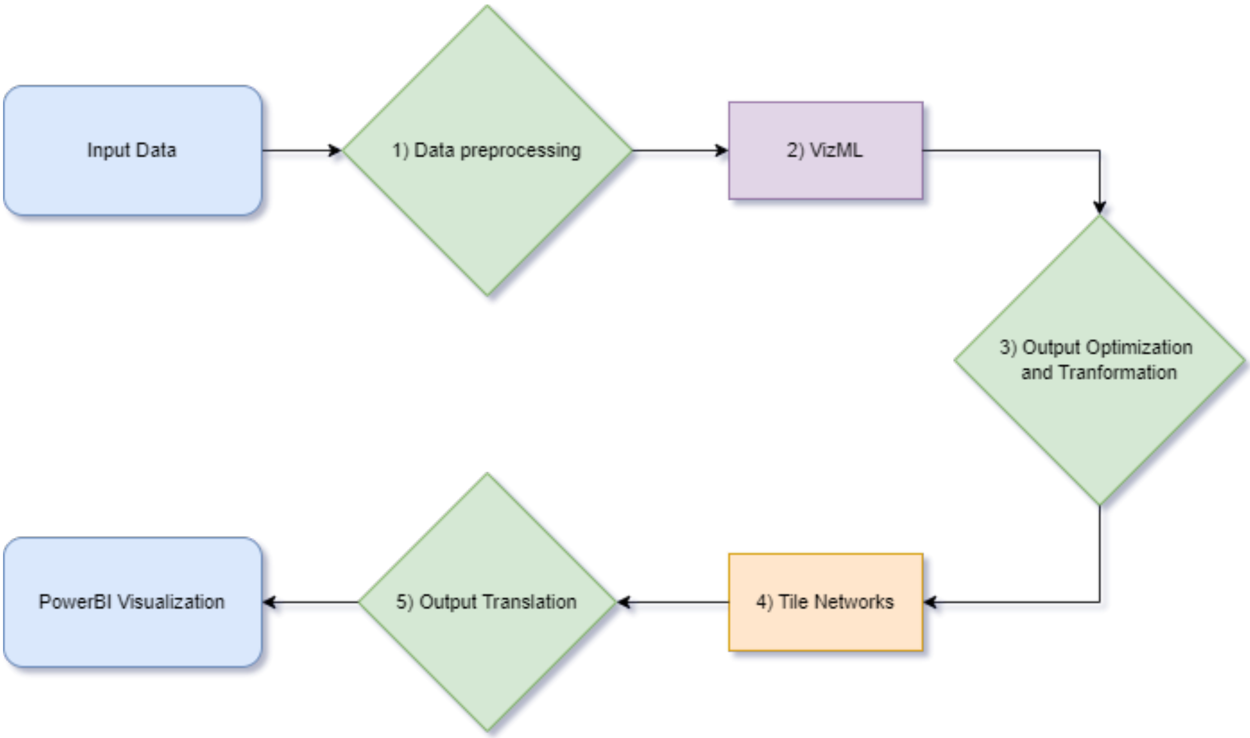
- Insightfulness: The dashboard should provide useful and meaningful insights that answer the user's questions or goals.
- Readability: The dashboard should be easy to read and understand by using clear labels, titles, legends, and annotations.
- Aesthetics: The dashboard should be visually appealing and attractive by using harmonious colours, fonts, layouts, and styles.
- Balance: The dashboard should have a balanced distribution of visualizations on the page, avoiding overcrowding or empty spaces.
- Coherence: The dashboard should have a coherent structure and flow that guides the user's attention and narrative.

These criteria can be quantified using different metrics or heuristics, such as information entropy, correlation coefficient, visual clutter, colour contrast, or alignment score. The reward function can then be defined as a weighted combination of these metrics or heuristics.

6) **Generate insights and narratives from the data:** The model must be able to generate insights and narratives from the data using captions, annotations, tooltips or drill-through actions.

7) **Publish and set up Power BI report:** The model must be able to publish the report and additionally, set up scheduled refresh for the data.

5. PROTOTYPE DESIGN AND EVALUATION



The flow diagram of the prototype design is as shown above. We go by the following steps:

1) **Data Preprocessing:**

Data must be preprocessed before it is input into VizML. The input format for VizML is a tabular data file, such as a CSV or a JSON file, that contains the data values and the metadata of the dataset. The metadata includes the name, type, and domain of each data attribute, as well as the cardinality and distribution of the data values.

The following steps are to be followed for preprocessing:

- Dataset must be converted into a tabular data type. Excel data must be converted to CSV, and Dimensional Models (in the form of Star Schemas) must be flattened into a single table or multiple tables with foreign keys.
- Dataset must be annotated by adding Metadata for each data attribute. Tools such as Data Voyager can be used.
- Ensure that your dataset is of CSV or JSON formats only. This can be fed into VizML.

2) **VizML:**

VizML's main functionality **[17]** is to automatically generate visualizations for a given dataset based on machine learning models trained on a large corpus of datasets and associated visualizations. The tool aims to lower the barrier to exploring basic visualizations by providing a search interface where users can select from a list of recommended visualizations, rather than manually specify them.

Here is an overview of the model architecture proposed in the paper:

- The input is a dataset that contains one or more columns of data. The dataset can be uploaded by the user or selected from a predefined list of datasets.
- The output is a list of recommended visualizations that are ranked by their predicted effectiveness. Each visualization consists of a visualization type (such as bar chart, scatter plot, etc.) and a set of design choices (such as which column to encode along the X- or Y-axis, whether to use color or size, etc.).
- The tool **[18]** consists of four main components: data preprocessing, feature extraction, model training, and visualization recommendation.

- Data preprocessing: This component cleans and transforms the raw datasets and visualizations into a standardized format that can be used for feature extraction and model training. This includes removing duplicates, outliers, and missing values, normalizing numerical values, and encoding categorical values.

- Feature extraction: This component extracts various features from the datasets and visualizations that can be used as inputs for the machine learning models. These features include statistical features (such as mean, standard deviation, skewness, etc.), semantic features (such as data type, column name, domain, etc.), and structural features (such as number of rows, number of columns, column order, etc.).

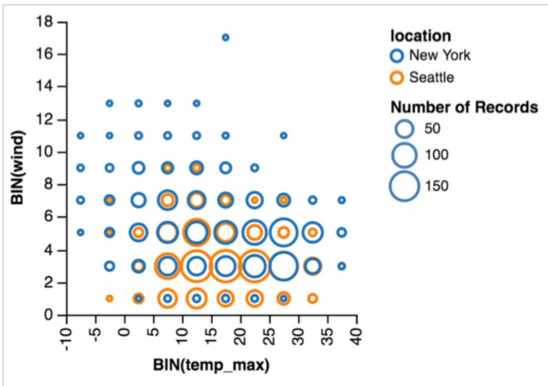
- Model training: This component trains various machine learning models to predict the visualization design choices based on the features extracted from the datasets and visualizations. These models include neural networks, random forests, gradient boosting, and logistic regression. The models are trained and evaluated using a large corpus of one million dataset-visualization pairs collected from a popular online visualization platform.

- Visualization recommendation: This component generates a list of recommended visualizations for a given dataset based on the predictions of the machine learning models. The component first selects the most suitable visualization type for the dataset based on the predicted effectiveness. Then, it generates a set of design choices for the selected visualization type based on the predicted probabilities. Finally, it ranks the recommended visualizations by their predicted effectiveness and presents them to the user.

3) VizML Output Optimization and Transformation:

The output for VizML is a set of visualization specifications that describe the visual encoding of data attributes. They are written in Vega-Lite grammar, which is a JSON based language. An example of a visualization written in Vega-Lite is given below.

```
{
  "data": {
    "url": "data/weather.csv",
    "formatType": "csv" },
  "mark": "point",
  "encoding": {
    "x": {
      "field": "temp_max",
      "type": "quantitative",
      "bin": true },
    "y": {
      "field": "wind",
      "type": "quantitative",
      "bin": true },
    "size": {
      "field": "*",
      "aggregate": "count" },
    "color": {
      "field": "location",
      "type": "nominal" }
  } }
```



{Source: Google Images}

The input for Tile Networks is a set of image files that contain different charts that show the data attributes and their relationships. Therefore, to convert the output of VizML to input for Tile Networks, the following steps are to be followed:

- Convert the visualization specifications into image files, such as PNG or JPEG files, that contain the rendered visualizations. Vega-Lite compiler and the Vega runtime, which are libraries that translate the specifications into graphical elements, can be used.
- Render the Vega specification using the Vega runtime. We can use the *vega.View* class to do this, which takes the specification as the input and creates a view object that can display the visualization on the web browser. We can also use the *view.toCanvas* or *view.toSVG* methods to export the visualization as a canvas or an SVG element, which can then be converted into an image file.
- Images must be resized to 256 x 256 resolution. Tools like ImageMagick can be used.
- Ensure that the images are saved with appropriate extensions, such as PNG or JPG.

Tile Networks extracts features from these images using a Convolutional Neural Network. Some of the features it considers are content, style and saliency of data.

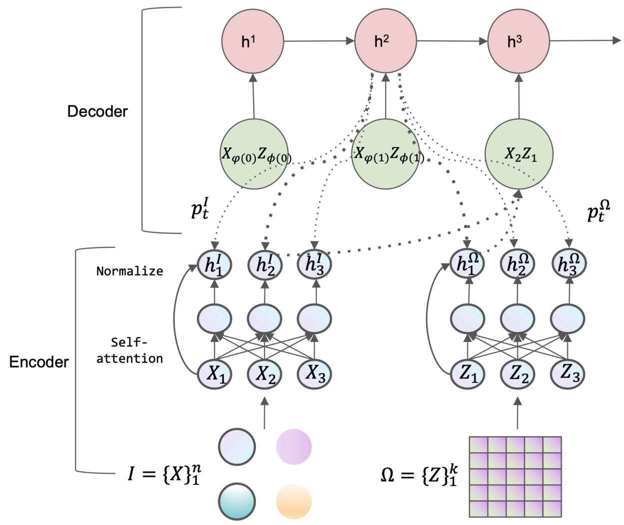
4) Tile Networks:

Tile Networks **[19]** is a type of neural network architecture that is designed to optimize 2D geometric configurations. In simpler terms, it's about arranging items in the best possible way on a 2D space or canvas.

The learning process of Tile Networks involves a method called reinforcement learning. It is a type of machine learning where an agent learns to make decisions by taking actions in an environment to maximize some type of reward. In the context of Tile Networks, the 'agent' is the network itself, the 'environment' is the 2D canvas, and the 'actions' are the placement of items on the canvas. The network iteratively selects items from a pool and places them in the most appropriate positions on the canvas.

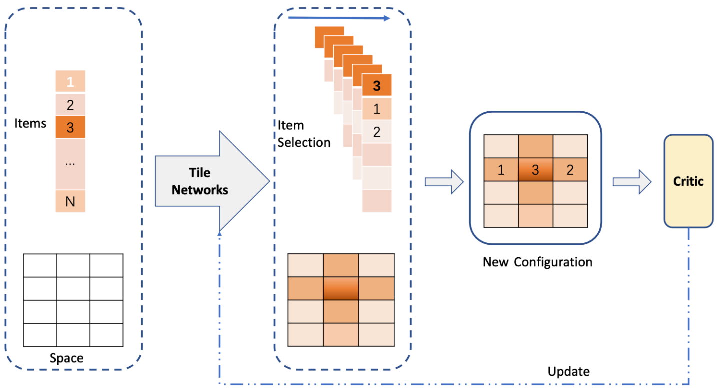
Tile Networks utilize an encoder-decoder network structure. The encoder's primary function is to transform discrete values I = {1,…, n} and spatially bound tiles Ω into a concealed representation. It ensures that each input element is cognizant of the presence of other elements during the encoding phase. As a result, our encoder employs message passing, which is facilitated by a self-attention mechanism.

The decoder's task is to choose a discrete value from I = {1,…, n} and position it within a tile in the Ω space. Simultaneously, the decoder must remember the previously chosen values and their arrangement to explicitly factor their impact into future decisions.



The goal is to find the best layout or arrangement of items on the canvas. This approach is data-driven, meaning it learns from data to make decisions, and can lead to better generalization over human-defined layouts. It's particularly useful in systems like personalized whole-page layout in recommender systems, where the aim is to present the most relevant items to the user in the most effective layout. This approach has been found to be more efficient and superior in performance compared to traditional methods.

The figure below represents the pipeline of learning algorithm, with the configurations sampled from Tile Networks that can be updated through stochastic gradient descent.



{Source: Reference [19]*}*

We found an efficient implementation of tile networks that can be efficiently used for page layout optimization **[20]**. It is a Tile-Coding package in TensorFlow that provides a sparse-coding method for generating features from real-valued data.

5) Output Translation to Microsoft Power BI:

The output of Tile Networks is a layout of items on the page, represented as J, which are the visualizations that are generated by VizML. The layout is represented by a set of coordinates and dimensions for each item, as well as the overall size and shape of the page. An example output format would be:


```
{
  "page": {
    "width": 800,
    "height": 600,
    "aspectRatio": 1.33
  },
  "items": [
    {
      "id": "item1",
      "x": 100,
      "y": 50,
      "w": 200,
      "h": 150,
      "url": "image1.png",
      "type": "image",
      "content": "",
      "style": {}
      }

   ]

}
```


We then convert these JSON objects into Power BI pages, to use them as inputs for Power BI visualizations.

- **Ensuring Optimal Working of the Prototype:**

1) Ensure that the data input is clean and valid. We can use the metadata and the *quality score* in VizML to ascertain this.

2) Incorporate User Preferences and Feedback: VizML and Tile Networks are user-centric frameworks. We need to elicit user feedback and incorporate it into our predictions accordingly.

- **Evaluation Strategies:**

Here are some measures for evaluating the appropriateness or effectiveness of a BI dashboard, as mentioned in literature:

(i) *Usage rate:* The usage rate of a dashboard is a good indicator of its quality. If a dashboard is used by stakeholders regularly, it's already reached a level of success **[21]**.

(ii) *Usability:* Usability evaluation of dashboards can be done using questionnaires such as System Usability Scale (SUS), Technology Acceptance Model (TAM), Situation Awareness Rating Technique (SART), Questionnaire for User Interaction Satisfaction (QUIS), Unified Theory of Acceptance and Use of Technology (UTAUT), and Health Information Technology Usability Evaluation Scale (Health-ITUES) **[22]**.

(iii) *Value:* The value of a BI dashboard can be assessed in terms of the business benefits that are enabled by data-driven decision-making. The value for companies who build their culture around insights generated from real, verifiable facts dwarfs the cost savings gained by organizations who deliver report and dashboard-based solutions **[23]**.

The following list is a non-exhausting list of factors that can be useful for evaluating the effectiveness of dashboard designs, which may overlap with the above criteria but can serve as handy while developing (implementing) the flow presented before:

- Ease of use
- User interface design
- Cross-platform compatibility
- Speed (how quickly does the dashboard load and respond to user interactions)
- Scalability (can the dashboard handle increasing amounts of data and users without a significant decrease in performance)
- Option for data integration
- Drill-down capability
- Allowing filtering/sorting
- Alignment with business goals
- Customizability
- Clarity of visual elements
- Appropriateness of visualizations
- User feedback and adoption
- Access control (does the dashboard have robust access controls to ensure that only authorized users can view certain data)
- Cost-benefit analysis (what is the cost-effectiveness of implementing and maintaining the BI dashboard compared to the benefits it provides)

6. FUTURE WORK

In future developments, the focus can be on achieving complete automation and integration of the proposed tool's architecture, streamlining the entire process from data preprocessing to Power BI output. This could involve scripting for diverse dataset conversions, leveraging tools like Data Voyager for metadata annotation, and extending compatibility beyond CSV and JSON formats.

Enhanced evaluation strategies can be integrated into the tool to automatically assess the quality and effectiveness of BI dashboards. Algorithms or modules could be implemented to capture and analyse usage patterns, usability metrics, and business value, providing comprehensive insights into dashboard performance.

For BI analysts, a user-friendly graphical interface can be designed and implemented to facilitate interaction with the tool. This interface can include features for customization and fine-tuning of visualizations, empowering users to tailor dashboards based on specific requirements. Additionally, implementation details for VizML can focus on efficient integration and deployment of machine learning models within the tool. Strategies for model interpretability could be explored to provide users with insights into the rationale behind specific visualization recommendations.

Efficient implementations and optimizations for Tile Networks can be explored, ensuring the tool's responsiveness and scalability. Documentation and tutorials can be provided for users to customize and extend the Tile Networks component for layout optimization. Furthermore, seamless integration with Power BI can be achieved by developing a robust module to translate the output of Tile Networks into Power BI pages. Dynamic updates and synchronization between the tool and Power BI can be explored for real-time adjustments based on evolving data or user preferences.

The tool's robustness can be ensured through rigorous testing under various scenarios and diverse datasets. Mechanisms for soliciting and incorporating continuous user feedback can be implemented to adapt the tool to evolving user needs. Additional evaluation measures can be considered, including cross-platform compatibility, speed, scalability, data integration capabilities, and alignment with business goals. Automated testing modules can be developed to assess ease of use, interface design, and other specified criteria.

Security features can be strengthened, implementing robust access controls, encryption mechanisms, and authentication protocols to protect sensitive data and prevent unauthorized access. Moreover, a module for conducting a cost-benefit analysis can be developed, providing insights into the economic implications of utilizing the tool for dashboard creation and optimization. Users can be equipped with information on return on investment and cost-effectiveness.

7. CONCLUSION

In summary, this project endeavours to address the challenges inherent in Business Intelligence (BI) dashboard design by proposing a comprehensive solution in the form of a BI dashboard design assistant. The research journey began with an exhaustive literature review, extracting valuable insights into effective dashboard design principles and methodologies. The two central research questions guided the exploration of requirements and the integration of machine learning techniques for personalized guidance throughout the design process.

The guidelines derived from the literature review offer a structured approach for a Machine Learning (ML) model to optimize Power BI reports. From connecting to diverse data sources to arranging visualizations using reinforcement learning, these guidelines provide a roadmap for creating coherent and impactful dashboards.

The prototype design, involving the sequential application of VizML and Tile Networks, represents a significant step forward. VizML automates the visualization generation process, while Tile Networks optimizes the layout for enhanced coherence. The translation of these outputs into Microsoft Power BI ensures practical applicability and seamless integration into existing BI workflows. Throughout this endeavour, the emphasis on user-centric design, usability, and continuous feedback incorporation emerged as crucial. By leveraging reinforcement learning and user preferences, the proposed assistant aims to evolve iteratively, aligning with the dynamic nature of data and user requirements.

The evaluation strategies outlined, encompassing usage rates, usability assessments, and value measurements, serve as vital benchmarks for gauging the effectiveness of BI dashboards. These measures, combined with a comprehensive list of factors for evaluation, ensure a holistic approach to assessing the quality and impact of the proposed dashboard design assistant.

Looking ahead, future work could focus on achieving complete automation, refining evaluation strategies, and enhancing user interaction through a graphical interface. The robustness of the tool will be tested under diverse scenarios, and security features will be strengthened to safeguard sensitive data.

In essence, this project not only contributes to the advancement of BI dashboard design but also sets the stage for future developments in automated design assistants. By combining AI techniques, reinforcement learning, and user-centric design principles, the proposed solution strives to empower BI analysts and designers, fostering a more intuitive, efficient, and user-friendly dashboard creation process.

REFERENCES

[1] Ballard, C., Farrell, D.M., Gupta, A.K., Mazuela, C., & Vohník, S. (2006). Dimensional Modelling: In a Business Intelligence Environment.

[2] https://www.netsuite.com/portal/resource/articles/business-strategy/business-intelligence-examples.shtml

[3] https://www.datacamp.com/blog/9-power-bi-dashboard-examples

[4] https://www.datapine.com/blog/dashboard-design-principles-and-best-practices/

[5] https://www.qlik.com/us/dashboard-examples/dashboard-design

[6] https://learn.microsoft.com/en-us/power-query/dataflows/best-practices-for-dimensional-model-using-dataflows

[7] https://richiebachala.medium.com/modern-data-modeling-74ca5b8d5e44

[8] https://www.projectpro.io/article/power-bi-dashboard-examples/941

[9] Few, S. (2006). Information dashboard design: The effective visual communication of data. O'Reilly Media.

[10] Yigitbasioglu, O., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. Int. J. Account. Inf. Syst., 13, 41-59.

[11] Ching-Yi Lin, Fu-Wen Liang, Sheng-Tun Li, Tsung-Hsueh Lu (2018). 5S Dashboard Design Principles for Self-Service Business Intelligence Tool Users. Journal of Big Data Research - 1(1):5-19.

[12] Janes, Andrea & Sillitti, Alberto & Succi, Giancarlo. (2013). Effective dashboard design. Cutter IT Journal. 26. 17-24.

[13] Koopman, Richelle & Kochendorfer, Karl & Moore, Joi & Mehr, David & Wakefield, Douglas & Yadamsuren, Borchuluun & Coberly, Jared & Kruse, Robin & Wakefield, Bonnie & Belden, Jeffery. (2011). A Diabetes Dashboard and Physician Efficiency and Accuracy in Accessing Data Needed for High-Quality Diabetes Care. Annals of family medicine. 9. 398-405. 10.1370/afm.1286.

[14] Abduldaem, A., & Gravell, A.M. (2019). Principles For the Design And Development Of Dashboards: Literature Review.

[15] Bach, Benjamin & Freeman, Euan & Abdul-Rahman, Alfie & Turkay, Cagatay & Khan, Saiful & Fan, Yulei & Chen, Min. (2022). Dashboard Design Patterns. IEEE Transactions on Visualization and Computer Graphics. PP. 1-11. 10.1109/TVCG.2022.3209448.

[16] Kaur, A. (2020). A Comparative Study and Review of Principles for Dashboard Design. Academia.edu.

[17] Hu et al. (2019). "VizML: A Machine Learning Approach to Visualization Recommendation". Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

[18] https://github.com/rischanlab/visualization_recommendation_systems/tree/master/vizml

[19] Xiao, S., Jiang, Z., & Yang, S. (2023). Tile Networks: Learning Optimal Geometric Layout for Whole-page Recommendation. ArXiv. /abs/2303.01671

[20] https://github.com/criteo-research/tf-tile

[21] https://uxdesign.cc/how-to-measure-the-success-of-dashboards-511559929109

[22] https://www.hindawi.com/journals/bmri/2023/9990933/

[23] https://www.tableau.com/learn/value-of-tableau/measuring-bi-value