doi:10.1093/iwcomp/10.1093/iwc/iwae037

Enricommender: Business Intelligence for User Interface Design

Alexis Ciarrone, Luis A. Leiva, Mateusz Dubiel*

University of Luxembourg, Luxembourg
**Corresponding author: name.surname@uni.lu*

Graphical user interface (GUI) browsing and retrieval tools are becoming essential to interaction design research and practice. These tools allow GUI designers to browse large amounts of data and recover inspiring or relevant designs for their task. Unfortunately, data-driven market analysis or business intelligence (BI) applied to GUIs have mostly been left aside. To address this research gap, we elicit designers' needs and responses regarding the development of Enricommender, a high-fidelity prototype of a market analysis recommender and reporting system. We identify and discuss key design challenges, as reported by more than 200 real-world designers, as well as their workflows and overall expectations towards such a BI system. Ultimately, this article sets the foundation for developing future GUI-oriented BI applications.

Research Highlights

- **Insights about ideation and design of graphical user interfaces.**
- **Elicited preferences and current practices from 200 real-world designers.**
- **A business intelligence prototype for comparative analysis and design recommendations.**

Keywords: Graphical User Interfaces; Comparative analysis; Requirements engineering
ACM CCS2 Keywords: Information systems – Business intelligence; Human-centered computing – Systems and tools for interaction design

Responsible Editorial Board Member: Prof. Helen Petrie

1. INTRODUCTION

Designing a User Interface (UI) is central to the development cycle of software supporting Human-Computer interaction (HCI). UI quality in large part determines the user experience (UX) and can play a crucial role in the success of a software system (Bunian et al., 2021) such as a mobile application (often shortened as *app*). In mobile apps, user interaction is primarily managed by a Graphical User Interface (GUI).

1.1. Problem statement

Although crucial, designing GUIs can end up being a complex, challenging, and time-consuming process for designers, regardless of their level of expertise or the type of system they are designing the interface for. These design issues have been identified by the research community (Arens et al., 1988; Eisenstein et al., 2002; Hashimoto and Igarashi, 2005; Deka et al., 2016) and still remain to be addressed. Considering the rapid proliferation of software systems available on the market and the paramount importance of their UIs in their usage, there is a real need for the said interfaces to be efficiently designed.

To facilitate this task, it is a common practice for designers to search for already-existing designs to achieve their ideation goals. Such methods have been recognised and studied on multiple occasions (Bonnardel, 1999; Herring et al., 2009; Gonçalves et al., 2014; Vasconcelos and Crilly, 2016; Koch et al., 2018; Leiva et al., 2020b). Designers may want to seek inspiration from their peers (Swearngin et al., 2018; Wallace et al., 2020; Bunian et al., 2021), stimulate their creativity and come up with new

ideas (Bunian et al., 2021; Ritchie et al., 2011), or gain a better understanding of the current UI design trends and practices (Deka et al., 2017), ultimately conditioning final designs (Eckert and Stacey, 2000).

To achieve these purposes, designers usually rely on online specialised platforms (Leiva et al., 2020b), such as Behance,[1] Pinterest,[2] or Dribbble,[3] which are prime locations for designers to share their work with their peers. Nonetheless, searching for design examples that are relevant to the designer's task is considered demanding (Koch et al., 2018). More specifically, the aesthetic, functional, or semantic properties of a given GUI are not entirely reflected by its pixels or associated metadata (Leiva et al., 2020b). Furthermore, such design search methods fail in some cases, e.g., when designers do not have clearly-defined purposes (Sharmin et al., 2009), once again making the search of relevant interface examples complex. Some researchers argue that this may represent a threat to design novelty and creativity (Perttula and Sipilä, 2007), as well as quality (Chan et al., 2017).

Since the concern of easily finding relevant design examples has been receiving a growing interest, researchers have recently tried to bring alternative solutions, including image-based mobile GUI retrieval tools such as Swire (Huang et al., 2019), VINS (Bunian et al., 2021) or Screen2Vec (Li et al., 2021). Based on a given input design, these systems are able to search for similar app design examples, i.e., screenshots often coupled with annotations, populated in large-scale datasets, e.g., ERICA (Deka et al., 2016), Rico (Deka et al., 2017), Enrico (Leiva et al., 2020a), or Screen2Words (Wang et al., 2021).

While the easy access to pre-existing designs is a huge asset for UI designers, it turns out that the aforementioned studies do not feature a more in-depth analysis of their results. The app designs populated in the datasets contain a wealth of information that is currently mostly used for retrieval tasks. These designs are composed not only of app screenshots, but also of real semantic aspects regarding their organisation, e.g., semantic wireframes or view hierarchies, which constitute valuable information that could be extracted from them. As the mobile app market has become such a significant and diverse industry (Inukollu et al., 2014), we can imagine that recurring GUI patterns, i.e., existing design principles and solutions that might be applied to solve a given design problem (Nilsson, 2009), could be identified and explicitly provided to designers in the form of advice or recommendations. We suppose that showing and using proven solutions to common design challenges to guide designers in the adoption of good design practices could be worthy of exploration, since is has potential to enhance design practice and consequently lead to more usable and user-friendly UIs. To the best of our knowledge, there is currently no prior work that delves into GUI design challenges and tries to mitigate them by using a market analysis recommendation-based approach, i.e., that studies the importance of design recommendations based on the analysis of real market data.

[1] https://www.behance.net/
[2] https://www.pinterest.com/
[3] https://dribbble.com/

1.2. Contributions

This article aims to address this gap in the research literature on GUI design ideation by providing:

(i) A comprehensive overview and analysis of GUI design challenges, designers' typical workflow, and expectations towards getting UI design recommendations from a data-driven perspective that relies on pre-existing material.
(ii) Enricommender, shorthand of *Enrico Recommender*, a high-fidelity interactive prototype of a Business Intelligence (BI) data-driven intermediary between mobile designers and datasets, offering them an exhaustive analysis of the market in which they wish to engage with their new application.

1.3. Research questions

The present study seeks to answer the following research questions (RQs):

RQ1: How interested would designers be in using a BI recommender tool as part of their workflow?

RQ2: What are the features of such a tool to best support designers' needs?

2. RELATED WORK

Accelerating and improving the design process of GUIs have been actively pursued goals for quite some time, and especially over the past ten years or so. Many researchers have been trying to come up with various solutions to reach these objectives, examples of which being the creation of GUI browsing (Ritchie et al., 2011; Kumar et al., 2013; Huang et al., 2019; Bunian et al., 2021; Li et al., 2021; Todi et al., 2021) and generation (Beltramelli, 2017; Moran et al., 2018; Huang et al., 2021) platforms, alongside large mobile GUI repositories (Deka et al., 2016, 2017; Leiva et al., 2020a; Li et al., 2020; Wang et al., 2021; Bunian et al., 2021). Here we provide an overview of the literature detailing these already-existing solutions, explaining how they were conceived and how they intend to improve the UI design process.

2.1. Design and layout retrieval

2.1.1. Web-centred search of GUIs

Providing designers with convenient ways to seek inspiration from pre-existing material is an efficient strategy for facilitating GUI design. The d.tour design exploration tool (Ritchie et al., 2011) is an example of such a strategy. It provides designers with the ability to browse existing webpages and to query the system with example designs, in order to obtain other similar interfaces.

Similarly, Webzeitgeist (Kumar et al., 2013) is defined as a mining tool for the web used for knowledge discovery and Machine Learning (ML) purposes, e.g., classification or metric learning. Thanks to thousands of webpage screenshots and millions of UI elements contained in Webzeitgeist, designers can query and browse the system to get inspiration from existing webpages, identify possible patterns and trends, and/or other assets such as cursors, overlays, typography, or text colour.

2.1.2. Exploration of mobile interfaces

The popularity and omnipresence of mobile phones has turned mobile app UI browsing into a real subject of interest. Consequently, many mobile interface retrieval or browsing systems, which the current work focuses upon, have been recently developed. Swire (Huang et al., 2019) is an example of such a system. It uses a deep neural network and performs nearest neighbour search to retrieve UIs that are similar to an input sketch, like a low-fidelity prototype. It also supports the recovery of UIs that resemble high-fidelity designs, to allow comparisons with alternative interfaces. According to its authors, Swire achieves a 60% relevancy, i.e., finding the example that is the most relevant within the top-10 results based on the input query. However, it has several limitations, including its lack of understanding of the sketch semantics, as Swire only focuses on their high-level layout information, and its difficulties to handle colourful components, thus recovering irrelevant examples.

VINS (Bunian et al., 2021) is another mobile UI search tool that aims to overcome Swire's limitations. It takes either an application screen's design or a more abstract wireframe as input and outputs a ranked list of similar designs, after performing computer vision tasks such as object detection, instance segmentation, and multi-modal embeddings over the input. The VINS workflow is composed of two phases, i.e., detection and image retrieval. First, during the detection phase, the different UI elements are detected and a semantic segmentation of the screen's layout is performed. Next, the image retrieval phase searches and returns designs that have similar hierarchical structures to the input. VINS has achieved promising results according to its designers, with a mean average precision (mAP) of more than 76% for UI elements detection and a 80–90% precision for similar design retrieval, using their custom dataset. However, its limited size may have caused certain UI elements not to be recognised and similar designs not to be retrieved.

Screen2Vec (Li et al., 2021) enables designers to browse mobile GUIs and retrieve ones that are similar to the input design. This process relies on comprehensive semantic embeddings of GUI screens that are ensured through a two-level architecture, i.e., GUI component and GUI screen levels. On the one hand, the component level encodes GUI components into embedding vectors, which are composed of the textual content and class type of these components. On the other hand, the GUI screen level encodes the textual content, the design, and layout patterns of the application's screen. Designers are then allowed to retrieve similar designs to an input one using more or less complex nearest neighbour queries. Nonetheless, Screen2Vec has only been trained on Android app screenshots—an issue it shares with Swire—letting an important part of mobile app design aside. There also exist inherent limitations of the dataset concerning interaction traces that Screen2Vec uses. Lastly, Screen2Vec cannot handle UI elements not containing textual information, e.g., glyphs or logos.

The previous three discussed solutions are quite similar in terms of their mechanics, in the sense that they return GUIs that are similar to an original, input design. A problem that arises with such approaches is that designers, when wanting to study designs that already exist and that are similar to their idea, may not have conceived or simply sketched an initial interface yet. To address this problem, an alternative input method consists of natural language queries. For example, Conversations with GUIs (Todi et al., 2021) introduces the 'Hey GUI' chatbot that answers mobile UI-related natural language queries. This Conversational User Interface (CUI) returns both graphical, i.e., Android app screenshot images corresponding to the query's criteria, and textual responses, i.e., information regarding apps' metadata.

Additional options that are more commonly known by designers are also worth to be considered. We can first mention Mobbin,[4] a regularly-updated web-accessible platform containing more than 260,000 screenshots that come from more than 880 iOS, Android, and web applications, together with user flows. Furthermore, GoodUX[5] populates good UX examples from various mobile apps and lists some of their features, describes them, and explicitly states why they are well designed. Finally, Simform[6] includes 'mobile patterns,' i.e., user

[4] https://mobbin.com
[5] https://goodux.appcues.com/categories/mobile-apps
[6] https://www.simform.com/mobile-patterns

flows for specific mobile app features, under the form of short screen recordings with comments and explanations.

As discussed before, the aforementioned tools are mostly mobile UI browsing and retrieval tools used to retrieve interfaces, but without further consideration of the information that can be extracted from them. More specifically, a market-centred comparative analysis applied to these interface examples is currently lacking, constituting a gap that Enricommender is meant to address.

2.2. Repositories of mobile graphical user interfaces

To fulfil their goals, mobile interface retrieval systems have to rely on existing applications, and in particular their GUI. Therefore, large enough datasets that contain this information are required.

Rico[7] (Deka et al., 2017) is one of the first large-scale datasets containing mobile app screenshots. It contains more than 72,000 UIs belonging to 9,772 Android applications from 27 categories, together with apps' metadata recovered from Google Play, user interactions, semantic wireframes, and view hierarchies. View hierarchies textually encompass all the GUI elements and their properties of a particular app screen, in addition to also exposing relationships between these elements. Rico also supports a visual representation of view hierarchies through semantic wireframes, representing the global screen's layout and differentiating textual and non-textual elements. Rico has proven to be a convenient support for designers to seek inspiration from existing apps, as it is the underlying data source of many mobile UI mining platforms (e.g., Swire, VINS, Screen2Vec).

Nonetheless, it should be pointed out that Rico was found to be noisy, which is an issue inherent to its rather large size (Lee et al., 2020; Leiva et al., 2020a). To solve this issue, an improved and cleaned-up version of Rico has been proposed in the form of the Enrico[8] dataset (Leiva et al., 2020a). Enrico contains a representative selection of only about 2% of Rico's data, i.e., 1,460 UIs, which has been manually reviewed and selected to ensure a higher quality of its content. Furthermore, Enrico UIs have been manually annotated to assign a topic, i.e., category or purpose, to each screen, thus adding an extra level of screen semantic understanding compared to Rico.

Yet another example of a new dataset aiming at improving Rico is RicoSCA[9] (Li et al., 2020), in which some screens from Rico have been manually removed to eliminate inconsistencies between screenshots and view hierarchies, a process similar to what has been done for Enrico. This eventually resulted in a dataset containing 25,677 unique UIs, to be compared to the 72,219 that Rico contains and the 1,460 ones in Enrico.

Screen2Words[10] (Wang et al., 2021) extends RicoSCA by providing to 22,417 of its UI screens human-annotated English summarisations, i.e., short sentences explaining the content and purpose of each screen. Those annotations encompass yet another degree of semantic expression of UIs.

Last but not least, the VINS[11] dataset (Bunian et al., 2021), used for the eponymous platform that we already described in Section 2.1.2, contains 4,800 UIs, including both abstract wireframes and high-fidelity designs. These UIs correspond to not only Android application screens (partially from Rico) like the other studied datasets, but also iOS ones. This dataset therefore solves an important limitation applicable to the other datasets, in the sense that it includes app designs from the two biggest operating systems currently in use, and not only one, thus better supporting real mobile UI designers' needs.

Since the number of mobile app interface datasets is constantly increasing, this does not represent an exhaustive list of all currently-existing options. Alternative datasets include those created for other projects like ReDraw (Moran et al., 2018), a GUI prototyping software, or Rewire (Swearngin et al., 2018), a mobile interface design assistant. All things considered, this leaves us with a large choice for our own system, Enricommender.

3. METHODS

We provide here a detailed overview of the Enricommender design process. We begin with a presentation of the mobile UI dataset that we have chosen, and provide our motivation behind this selection. We then explain the process of gathering the system requirements.

3.1. Dataset

One of the goals of this work is to create an outline of a BI recommender system for mobile GUI design, based on existing interfaces in the market. As our approach is data-driven, we had to search for an appropriate mobile UI repository that Enricommender can rely on. Following the examination of the repositories, listed in Section 2.2, and a comparison between them, we decided to use Enrico (Leiva et al., 2020a) as the underlying data source for our recommender system, Enricommender.

[7] http://interactionmining.org/rico
[8] https://github.com/luileito/enrico
[9] https://github.com/google-research-datasets/seq2act
[10] https://github.com/google-research/google-research/tree/master/screen2words
[11] https://github.com/sbunian/VINS

3.1.1. Selection

The selection of Enrico was made based on several key factors, namely *popularity*, data *accuracy*, and *relevance* regarding the target task.

Popularity. The popularity of a dataset demonstrates the extent to which it is known, sought after, commonly used and, more importantly, trusted by scientists and researchers in a given field or to solve a particular problem. Among all the datasets we studied, the most popular, based on its number of citations[12], was Rico (Deka et al., 2017). At this stage, it was quite clear that Rico (or one of its extensions) should be considered as the dataset we would be using to help reach this work's goals, given the importance it has gained over the years in the field.

Accuracy. The accuracy of a dataset refers to its correctness with respect to what it is supposed to represent (Fleckenstein and Fellows, 2018; Mahanti, 2019) and constitutes one dimension of data quality, which is a multidimensional concept (Ballou and Pazer, 1985; Pipino et al., 2003). Since Rico has been proven to be noisy (Lee et al., 2020; Leiva et al., 2020a) and sometimes inaccurate (e.g., bounding boxes not representative of the screen's content, incorrectly sized, incorrect and inconsistent class labels of UI elements, etc.), our interest has turned to one of its derivatives instead.

Relevance. While all of Rico's extensions represent strong options as mobile GUI datasets, Enrico's most distinctive feature, i.e., screen topics (cf. Figure 1), particularly stands out. This constitutes an additional level of screen semantic understanding, in the sense that it clearly assigns a purpose to each screen. Such a functionality is useful since this information can be used to provide our users with more specific statistics and recommendations that concern particular types of screens expressing a clearly defined goal.

3.1.2. Enrico

As described in Section 2.2, the Enrico dataset contains 1,460 Android UIs under the form of real screenshots, semantically annotated thanks to semantic wireframes and view hierarchies. Each app screen in Enrico is assigned a design topic, and is linked to a metadata file directly extracted from Google Play. It is a filtered down version of Rico where inaccurate examples have been eliminated.

Design topics. Enrico's distinctive feature lies in its design topics, that could be summarised as the category of a given screen, emphasising its functionality and purpose. There are 20 design topics in total (cf. Table 1). Concretely, in the dataset, design topics are assigned to each screen (e.g., Figure 1), inside of a comma-separated value (CSV) file.[13] Enrico UIs were manually labelled by two human annotators with the topic that was deemed to be the most relevant.

Table 1. Enrico design topics (Leiva et al., 2020a).

Topic	# UIs	Description
Bare	76	Largely unused area
Dialer	6	Number entry
Camera	8	Camera functionality
Chat	11	Chat functionality
Editor	18	Text/image editing
Form	103	Form filling functionality
Gallery	144	Grid-like layout with images
List	265	Elements organized in a column
Login	141	Input fields for logging
Maps	9	Geographic display
Media Player	32	Music or video player
Menu	79	Items list in an overlay or aside
Modal	67	A popup-like window
News	59	Snippets list: image, title, text
Other	52	Everything else (rejection class)
Profile	63	Info on a user profile or product
Search	35	Search engine functionality
Settings	90	Controls to change app settings
Terms	39	Terms and conditions of service
Tutorial	163	Onboarding screen
Total	1,460	

Screenshot images. Screenshot images in Enrico directly correspond to the high-fidelity, production version of an application UI that is visible to users. Extracted from Rico, the screenshots were selected with respect to their correspondence with their associated semantic wireframe (and view hierarchy). Enrico contains 1,460 screenshots coming from various Android apps, one of which can be seen in Figure 2a.

Semantic wireframe images and annotations. Semantic wireframes, also recovered from Rico and for which an

[12] https://scholar.google.com/

[13] https://github.com/luileito/enrico/blob/master/design_topics.csv

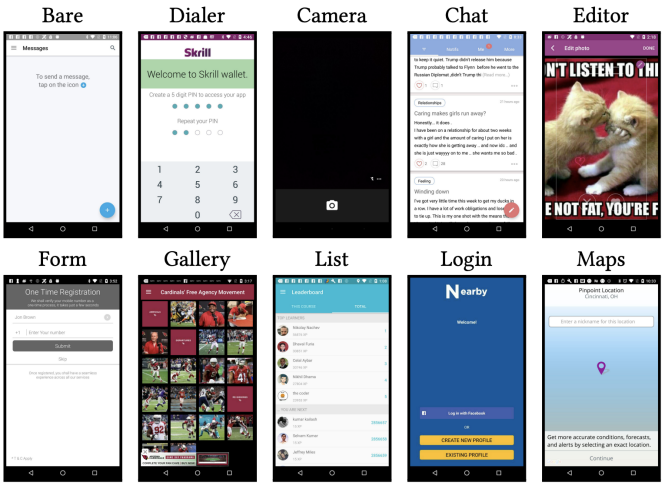
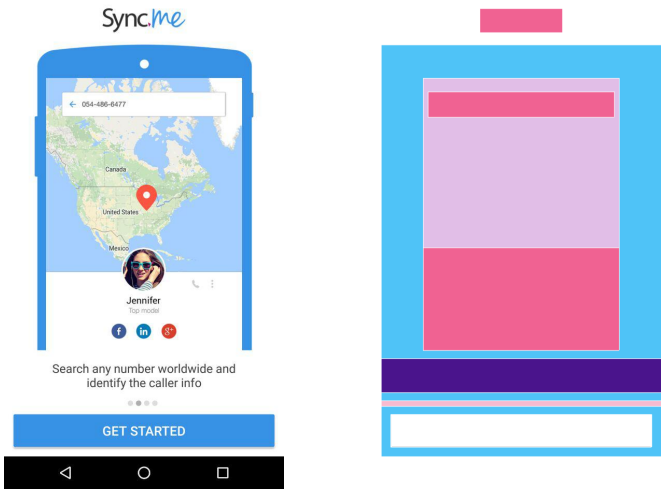


Figure 1: Examples of some Enrico screenshots assigned to their respective topic (Leiva et al., 2020a).



(a) Screenshot. (b) Semantic wireframe.

Figure 2: An application's screenshot with its associated semantic wireframe (Deka et al., 2017).

example is depicted in Figure 2b, express the semantic annotations of UI screens, i.e., the nature and function of the UI elements they contain, in visual form. To differentiate different UI elements embedded in the wireframes, colour codes are defined in three JavaScript Object Notation (JSON) files for UI components,[14] text buttons,[15] and icons,[16] and applied to colour the semantic wireframe images.

View hierarchies. The view hierarchies in Enrico represent the semantic annotation of UIs in textual format (Leiva et al., 2020a), in contrast to semantic wireframes that are purely visual. They correspond to the elements' metadata and express the structural representation of the UI (Deka et al., 2017) by listing in a JSON file[17] all the UI elements that can be found on screen. It can be seen as a textual representation of the semantic wireframes in a tree-like format.

Application metadata. In addition to containing element-level metadata, Enrico also contains app-level metadata in JSON format, just like the annotations and view hierarchies. They were directly recovered from Google Play in August 2020, and contain app-related information such as name, developer, description, category, etc. There are 29 attributes in total, populated[18] in Table 2 together with explanations and examples corresponding to the app from which the screen of Figure 2 is taken. In the context of Enricommender, such exhaustive data about apps is particularly useful as it allows to efficiently filter the results for users, as well as providing an opportunity to establish statistical measurements.

3.2. Requirements elicitation

Requirements elicitation is a crucial part of requirements engineering. To get a more precise sense of the scope of our study, alongside first basic user needs and expectations, we started this phase by meeting three HCI professionals within our university: a research specialist and a postdoctoral researcher from the HCI research group, and a UI/UX designer who previously worked for TikTok.[19] During these interviews, we presented initial ideas behind Enricommender, asked our interlocutors what they would demand from it and how it could possibly benefit them. The outcomes of these discussions allowed us to reshape our contribution and to make it more relevant to the target audience. We subsequently designed a low-fidelity prototype of Enricommender (cf. Figure 3) focused on the expressed and requested needs. Although we feel that this step was necessary, as it represents some

[14] https://github.com/luileito/enrico/blob/master/component_legend.json

[15] https://github.com/luileito/enrico/blob/master/textButton_legend.json

[16] https://github.com/luileito/enrico/blob/master/icon_legend.json

[17] https://github.com/luileito/enrico/blob/master/samples/10594-hierarchy.json

[18] Non-significant elements, e.g., app description in HTML format, developer's ID, email, address or URL are not included in the table, but are available.

[19] https://www.tiktok.com/

Table 2. Google Play app metadata (sample file).

Attribute	Description
`title`	App name
`icon`	URL of the app icon
`screenshots`	URLs of the app's images
`video`	URL of the app's video
`category`	App category
`score`	App's average score
`histogram`	Distribution of scores
`reviews`	Number of reviews
`description`	App description
`recent_changes`	List of recent changes
`editors_choice`	Whether the app is highlighted as Editors' Choice
`price`	App price (US dollars)
`free`	Whether the app is free
`iap`	Whether the app relies on *in-app purchases* (IAPs)
`updated`	Date of last update
`size`	App size
`installs`	Number of installs
`current_version`	App version
`required_android_version`	Android version required
`content_rating`	App's maturity level
`iap_range`	Range of IAPs (US dollars)
`developer`	Developer name
`app_id`	Unique ID of the app
`url`	Store URL of the app

preliminary formational feedback, it was not sufficient; as we cannot derive generalities from discussions with three people only. In this context, we have created a larger-scale survey directly targeting designers, whose goal is to gather information regarding their workflow, design practices, and expectations regarding Enricommender.

3.2.1. Participants

The participants of the survey were recruited through the Prolific crowdsourcing platform.[20] To ensure the quality of the answers, we specifically targeted people with a UI design or UX backgrounds, who have been using the platform for more than a year, were fluent in English, and had an approval rate of 100% in previous studies. Participants could only complete the study once.

In total, 202 people (162 males, 38 females, 2 did not specify) of 37 different nationalities participated. Participants were aged between 19 and 62 years old (median age = 30, mean age = 32.75). Participants had various backgrounds and knowledge of software

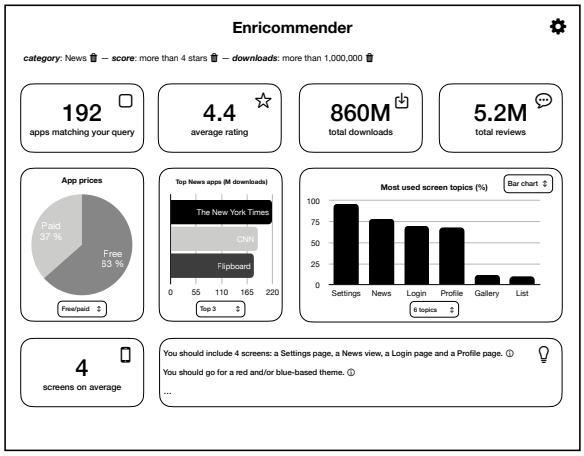


Figure 3: Early low-fidelity mockup of Enricommender's dashboard interface.

development techniques, e.g., web development, UX, UI design, responsive design, etc. out of which 117 people were professionals and 64 were students. The rest were mainly hobbyists, developers, or freelancers. Professional participants were further able to select multiple[21] areas of expertise; 79 self-identified as web designers, 62 as UI/UX designers, 49 as graphic designers, 29 as desktop app designers, and 28 as mobile app designers (7 chose another area).

Completing the survey took 27 minutes on average, and participants were paid the equivalent of 7 euros per hour upon completion. Participants were informed that all of the data collected is anonymous and solely used for the purposes of the present research. The study was covered by the Ethics Review Panel of the University of Luxembourg (ID: ERP 22-071).

3.2.2. Procedure

The questionnaire is divided in five main parts, where we aim to understand various participants' characteristics and thoughts.

(i) **Introduction and demographics**: explanation about survey's goals, expected completion time, information about anonymity, data usage, and participants' rights. Furthermore, participants were asked whether they are professionals or students;
(ii) **Areas of expertise (professionals only)**: people had to select all the domains that apply to them from a predefined list, i.e., (1) UI/UX designer, (2) graphic designer, (3) mobile app

[20] `https://www.prolific.co/`

[21] Meaning that they were able to select all options that applied.

designer, (4) web designer, (5) desktop app designer, and (6) other;

(iii) **Challenges of GUI design**: participants were asked to identify what they considered as the most difficult part(s) about designing GUIs;

(iv) **Sources of inspiration**: designers were questioned about their main sources of inspiration at the beginning of their process—namely colleagues, enterprise guidelines, or dedicated websites–with their opinion about it/them. As Enricommender aims at being a new source of inspiration, participants were then asked what they would think about receiving GUI design recommendations and how it would be useful for them;

(v) **Expectations towards the system of interest**: explanation about what Enricommender is and what it aims to achieve. Furthermore, participants were asked about their understanding of the tool, based on the provided description and an early low-fidelity prototype we have previously designed (cf. Figure 3). This was followed by the list of app metadata we have at our disposal in the Enrico dataset, from which we ask participants to derive Key Performance Indicators (KPIs). More specifically, participants had to assign a score to each attribute, between 1 (not important at all) and 5 (very important). To finish, we explicitly asked the participants what they would like a recommendation and BI-based tool like Enricommender to do, in terms of features, layout, and presentation of the information.

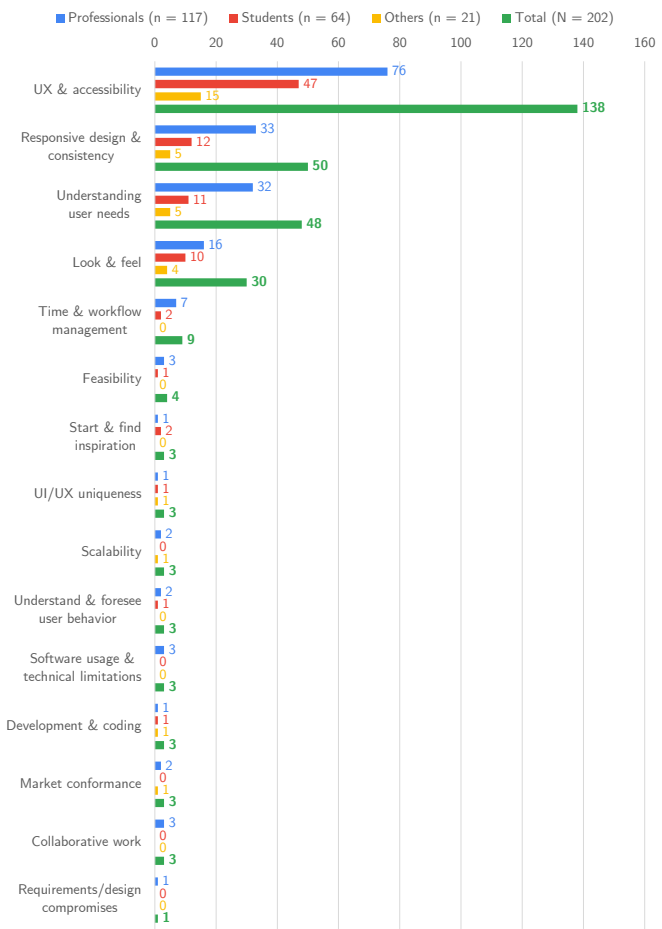


Figure 4: Identified challenges of GUI design by occurrence and participant group.

4. RESULTS

The results correspond to the two contributions we are making to the research area in order to close some of its remaining gaps. The first contribution is a comprehensive analysis of GUI designers' identified challenges, workflow, and prospects regarding a BI market analysis-based recommender system, i.e., that is based on real market data and on its analysis. Our second contribution translates the needs identified in our inquiry and features them in a high-fidelity, interactive prototype of Enricommender.

4.1. Challenges of GUI design

The first insight obtained from the participants' answers concerns what they identify as the main challenges when designing GUIs. Our goal here is to determine whether these challenges could be solved by Enricommender.

Overall, we have managed to identify a set of 15 challenges from the answers, that we have coded to group them based on the ideas they expressed (cf. Figure 4). Four types of challenges have been mentioned by at least 10% of the participants, i.e., 20 out of 202 people: (1) find an appropriate UX and accessibility level for the interface; (2) create a responsive UI and ensure consistency across different instances of the UI; (3) properly understand user needs; and (4) look and feel considerations.

4.1.1. UX and accessibility

Providing an appropriate UX and the right level of accessibility to users is the most important challenge that designers seem to face when designing GUIs. Such concerns have overall been reported by 138 participants (68.32% of the collected answers).

According to the ISO 9241-210:2019 standard,[22] UX refers to a 'user's perceptions and responses that result

[22] See Ergonomics of Human-System Interaction – Part 210: Human-centred Design for Interactive Systems, International Organization for Standardization, 2019.

from the use and/or anticipated use of a system, product or service.' Designers are especially concerned by this aspect, as they want the interface they design to provide the best possible experience for their users. Designers are paying more and more attention to accessibility, i.e., the 'extent to which products, systems, services, environments, and facilities can be used by people from a population with the widest range of user needs, characteristics, and capabilities.'[22]

In this regard, many participants described the challenge related to UX and accessibility in such terms:

> '*Ensuring that the interface is user-friendly and the users will be able to navigate without needing any guidance or support or any sort of instructions.*' ($P1$)

This definition directly employs the notions of user-friendliness and intuitiveness. The terms '*user-friendly*' and '*intuitive*,' as well as '*easy to use*,' were often expressed in the answers to describe challenging characteristics of the interfaces to be designed. In brief, these three aspects refer to an interface that can be used immediately and instinctively, thus offering a good and seamless experience to its users.

We cannot talk about ease of use, intuitiveness and user-friendliness without mentioning accessibility, another aspect of the UX challenge that has gained importance over the last couple of years. Making an interface pleasant and usable for everyone is clearly an aspect that designers care about. As one participant ($P2$) aptly pointed out, one of the main challenges with usability and accessibility is that, as a designer, you are '*not designing for yourself but* [for] *the client*,' who may possibly have some disabilities. The UX and interface design have to be carefully considered, namely in terms of the size or placement of UI elements, e.g., for people with motor disabilities, or colour schemes, e.g., to accommodate colour blindness.

To summarise, the vast majority of designers usually has difficulties creating an interface that is meant to be accessible to as many people as possible, in which users can navigate easily and intuitively, while also benefiting from pleasant aesthetics. Finding the suitable UX for an app's interface is a challenging process that requires many iterative trials and extensive user feedback, as some participants recognised.

4.1.2. Responsive design and consistency

The second aspect that came up most frequently in the participants' answers (50 people, 24.75%) when it comes to design challenges is linked to the creation of responsive and consistent UIs. First and foremost, responsive (web) design denotes the capability for a user interface to take into account and properly scale to all devices' screens, independently of their size and resolution (Voutilainen et al., 2015).

One of the implied difficulties is that some UI parts cannot be simply extended or 'up-scaled' to go from a small screen, e.g., on a phone, to a larger screen, fitted for example to a computer or certain tablets. Indeed, '*what may work on a desktop browser may suffer on a mobile device browser*,' a participant ($P3$) said. This point is supported by another person ($P4$) who gives the example of '*a table*,' which '*is easily readable on a bigger screen*' but not really on a small one. According to him, the main difficulty resides in the design of an interface '*that looks great on many kinds of devices* [...]. *They all have different screen sizes, which makes it difficult to design a UI that works well on all of them while remaining consistent*' ($P4$).

Speaking of consistency, and in particular *internal* consistency as inferred by the answers, it corresponds to the level of aesthetic and functional coherence within a system, e.g., in terms of layout – including '*details, like spacing between components*' ($P5$) –, elements' size, colour, shape, functionality, and naming (Grudin, 1989). Participants also frequently mentioned difficulties to ensure consistency not only within a given system, but also between different instances of the said system, running on different devices with different screen sizes. This makes responsive design and consistency deeply connected in our context, as perceived by a participant who noted:

> '*Accommodating various screen sizes, resolutions, and device types adds complexity to the design process, as the interface needs to adapt and provide a consistent experience across different platforms.*' ($P6$)

Finding consistency in a design is a necessity, as it reduces and eases the learning task of using the system for users (Satzinger and Olfman, 1998) and significantly improves the overall UX.

4.1.3. Understanding user needs

The third challenge that designers often encounter while designing a GUI is related to the proper understanding of user and/or stakeholder needs. This issue seems to be as important as the responsive design and consistency-related one, both having a similar occurrence in the answers (48 people, i.e., 23.76% v. 50 and 24.75%, respectively).

As recognised by many participants, the fact that, as a designer, you are '*not designing for yourself*' (e.g., $P2$, $P7$), but for external users, sometimes leads to an important bias or misunderstanding of what the client and the audience they target really need. Human beings tend to favour things they know, that is, their own experiences, opinions, preferences, emotions or needs (Romanenkova, 2020), rather than others' that may appear irrelevant, distant or completely unknown

because of '*the absence of knowledge/understanding about user expectations*' ($P8$). In the design process, the implicit contribution of designers' own interpretations, background, and prejudices cannot lead to an objective mindset, but rather results in an unsuitable *self-referential* design:

> '*Self-referential design is the tendency that many designers have to instead of asking "What would a user want?" they ask themselves "What would I want if I were a user?"*' ($P7$)

This also represents an easy solution – overcoming a reported '*lack of opportunities to speak directly with real end-users*' ($P9$) – that can be seen as efficient both in terms of cost and time, as the only person involved in the decision-making process is oneself. Such a fact has been clearly identified:

> '*In order to answer the question of what a user would actually want, you have to study them, ask them questions, and get to know them. In order to answer what you would want, you need to only ask yourself.*' ($P7$)

Therefore, for designers, overcoming the self-referential design practices appears to be a major challenge, directly caused by human nature and studied many times in the literature as the *familiarity bias* phenomenon, which posits that people are more leaning towards things that are familiar to them, and are usually repelled by the unknown (Cao et al., 2011):

> '*Sometimes familiarity with a system can blind you to how users will respond and interpret elements and actions.*' ($P10$)

Thus, relying on the designer's intuition for the product's design cannot be considered as a suitable practice. Intuitions and assumptions are, by nature, highly subjective, and potentially irrelevant for the task at hand, as '*people are far more diverse than you can possibly imagine*' ($P7$).

Undecided, poorly expressed or changing requirements also play a role in the difficulty to understand user needs. Participants admitted that some of their clients '*don't know what they want*' ($P11$, $P12$, $P13$), '*do not clearly express their needs*' ($P14$) or regularly '*change their mind*' ($P14$, $P15$, $P16$). Also, some people bear witness to '*unrealistic requirements*' ($P17$) or '*innovative graphic solutions*' ($P18$) that may be difficult to adapt to '*widely accepted interface design standards*' ($P19$) or conflict with the expectations of the users who '*are not necessarily the most skilled or used to new forms of design*' ($P18$).

Finally, '*balancing the needs of a vast number of very different users*' ($P20$) has also been recognised as challenging. Survey respondents acknowledged that it is difficult '*to design a GUI that meets the needs of everyone*' ($P21$), while it is also '*sometimes impossible to please everyone*' ($P22$).

4.1.4. Look and feel

The last significant challenge that we learned from our survey corresponds to the perceived aesthetics and behaviour of UIs, that we referred to as the 'look and feel.' Part of the UX of a design, the look and feel comprises visual elements like colour, shape, style, text or layout, as well as interaction means (Lee and Sunder, 2016). The expression of look and feel-related concerns was found in 14.85% of all the answers, i.e., 30 out of 202.

Designers usually seem to experience difficulties with '*interactivity and animation*' ($P23$) or '*the arrangement, alignment, symmetry, and consistency of UI elements*' ($P24$, $P25$, $P26$). Participants also discussed how complicated it can be to choose the appropriate '*font, colour scheme or palette*' ($P16$, $P17$, $P27$, $P28$) without clashing with accessibility principles and '*people's disabilities*' ($P27$, $P29$, $P30$). In this context, almost 8% of the answers, i.e., 16 out of 202, mentioned the difficulty to find the proper balance for a design to be '*visually appealing*,' yet at the same time functional and '*intuitive for users*,' in order not to fall into the 'style over substance' problem and to pay a careful attention to '*cognitive load*' ($P31$, $P32$). Also, the strive for '*the perfect design*' is depicted as long, arduous, tiring, and '*delaying projects*,' without guaranteed results ($P33$).

Overall, according to designers, visual and interactive properties of an interface are complex to find and balance, considering that they are also related and directly imply other design considerations, from UX and consistency to the satisfaction of the users' needs and the app's functional requirements.

4.2. Sources of inspiration

The second part of our survey consisted of asking designers about their source(s) of inspiration in the context of their design process. We provided them with three predefined answers, i.e., their enterprise guidelines, their colleagues or dedicated websites that they had to cite (cf. Figure 5). For participants using other sources, an 'Other' field was provided.

4.2.1. Enterprise guidelines

A vast majority of the surveyed designers, i.e., 73.76% (149 out of 202), was found to use their enterprise guidelines as an inspiration source. According to the participants, this allows them to benefit from a solid basis that helps to make design '*more concise* [with] *less room for doing something wrong*' ($P34$), while '*maintaining consistency across the interface*' ($P35$) '*under one design framework*' ($P36$) that aligns with '*the company's visual identity*' ($P6$). Also, it has been recognised as '*the quickest way to get inspiration*' ($P37$) that usually '*focuses on what*

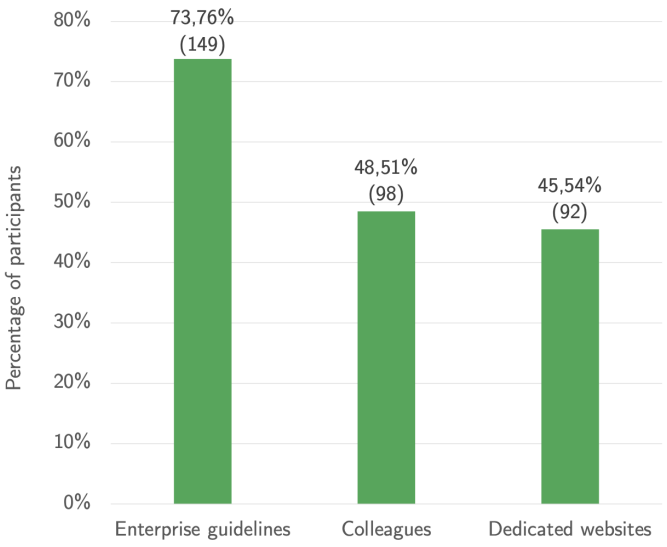


Figure 5: Designers' main sources of inspiration.

the client exactly needs' ($P38$), as these directives may already have gotten some user feedback.

Even if this method seems to be universally acclaimed, it also has its share of disadvantages, the biggest of which being the associated limitations for designers. Considered as '*a bit too strict*' ($P39$) and '*restrictive*' ($P40$), enterprise guidelines were often considered as a drag for design freedom and creativity:

> '*Enterprise guidelines sometimes constraint the things we can do in the interface, as they act as a boundary during the design. Also, you* [often] *can't deviate from those guidelines.*' ($P41$)

In addition, considering the ever-evolving nature of mobile app aesthetics and the inertia of the procedures of some companies, corporate guidelines can sometimes be '*outdated*' ($P42$) and become '*stale over time*' ($P10$), in addition to be '*very confused and not clear*' at times, by the confession of a mobile app and web designer ($P43$).

4.2.2. Colleagues

Besides following their company's guidelines, almost half of the participants (98 people – 48.51%) recognised discussions with their colleagues as an additional source of design inspiration. According to them, this practice helps designers '*to get into the right mindset*' ($P44$), as it is reportedly '*easier to draw inspiration from like-minded individuals*' ($P44$, $P45$), who are '*trustworthy*' ($P46$, $P47$), '*experienced*' ($P19$, $P48$) and '*have some ideas which sometimes are brilliant*' ($P49$).

Furthermore, it also stimulates designers' creativity and offers them new standpoints '*with methods* [they] *may not have thought of*' ($P44$) or '*fresh and innovative approaches to the interface design*' ($P6$). It further represents an opportunity to engage in productive activities such as, brainstorming sessions or collaborative work that some designers especially value:

> '*Collaborating and exchanging ideas with fellow designers brings diverse perspectives and feedback. It creates a dynamic environment where we can share experiences and ignite creativity, ultimately enhancing the design process.*' ($P45$)

However, such '*diverse perspectives*' and '*conflicting opinions*' ($P46$) can be double-edged, in the sense that it may lead to major disagreements and deadlock situations that '*might affect how the overall interface looks, or whether the design would be done right the first time*' ($P46$).

4.2.3. Market research and comparative analysis

Last but not least, looking at already-existing UIs, that serve a similar purpose to that of the task at hand, and comparing them with one another is another practice that designers often engage in during their design process. Such an inspiration seeking method is reportedly used by 45.54% of the participants, i.e., 92 people, which represents nearly one in two designers.

When browsing and assessing pre-existing UIs, designers are essentially looking for getting '*a feel*' ($P47$), know more about '*how other designers have previously approached the same requirements*' ($P9$) and, roughly speaking, '*what the clients' competition is doing, visually*' ($P48$). By doing so, designers are able to quickly '*spot design trends*' ($P49$), analyse good practices, and, using their words, '*weed out the bad stuff*' ($P50$). Such a process further permits to learn from models, '*improve on them and make a better product*' ($P51$). This research also has the effect to stimulate designers' creativity who can come up with new ideas, to the extent that a participant reveals that he '*end*[s] *up with too much inspiration with tonnes of screenshots!*' ($P52$).

The most significant platforms that were cited in the answers and that designers reported using are the following:

- Dribbble,[23] a social network grouping digital designers' portfolios.
- Behance,[24] Adobe's competitor to Dribbble.
- Pinterest,[25] an image sharing social media.
- Mobbin,[26] an iOS and Android app UIs browsing platform.
- Awwwards,[27] a web design competition platform that aims to honour and advertise the best designs;

[23] https://dribbble.com/
[24] https://www.behance.net/
[25] https://www.pinterest.com/
[26] https://mobbin.com/
[27] https://www.awwwards.com/

- Google Images.[28]
- The client's website, should it already exist and be adapted to a mobile application.

Although most of the surveyed designers declared themselves as satisfied with this inspiration seeking technique, a few saw in it some limitations. First, it can be a blow to their organisation and a designer recognised feeling '*overwhelmed with all the examples and possibilities*' ($P53$). Furthermore, as there indeed exist millions of different applications, finding relevant and interesting examples can be costly in terms of time: '*it can be rather time consuming to visit all venues when trying to come up with inspiration*,' a participant ($P54$) argued. Paradoxically, while a significant number of people considers examining already-existing UIs as an asset for their creativity, others think on the contrary that it limits originality and reinforces the '*homogenisation of designs*' ($P55$).

To finish, we have found some answers that express a need to go even further than consulting pre-existing UIs, like for example a '*design copilot*' ($P56$). It is an idea that can potentially come close to a system like Enricommender, where the GUI design recommendations can act like an outline of a copilot. Lastly, some designers are using AI generative tools, such as the Midjourney[29] image generator ($P33$, $P57$), or the now well-known ChatGPT[30] conversational agent ($P57$) that can be used by designers to get UI advice and guidelines.

Overall, designers really seem to value user and client feedback more than anything else. Other materials are considered almost unanimously as precious, but '*asking the client their opinion on what the interface should look like is also important*' ($P58$) and '*the users should also be asked if this is what they need/could use*' ($P13$). By relying too much on pre-existing UIs at the detriment of their target users, designers expose themselves to the risk of no longer meeting the needs that were to be satisfied. Other interfaces were indeed designed with other, potentially very specific requirements in mind, so one can understand that putting the client and/or end users in the loop is fundamental.

4.3. Designers' expectations

The last part of our survey, and potentially the most important, was about identifying and gathering designers' expectations regarding a BI market-based recommender system like Enricommender.

4.3.1. Help of recommendations for GUI design

We first wanted to know what the participants thought about getting UI recommendations, taking already-existing designs as a basis, and how it could help them in their design process.

Globally seen as a precious assistance and '*guidance*' ($P6$, $P53$, $P59$, $P60$), recommendations are even sometimes seen as '*a third person's opinion*' ($P61$) and can do as much as determining '*how relevant and convenient your end product will be*' ($P44$). Such advice may indeed point out '*something that you missed/neglected or help you maximise your already-existing ideas*' ($P61$). This might result in a more efficient workflow for designers, as well as significant time and cost savings that can in turn be reallocated to '*other aspects of your GUI design project*' ($P23$). Furthermore, this guidance also helps to reduce risk costs, as '*the chance of failure is exponentially higher when doing something from scratch*' ($P39$).

Additionally, market analysis-based recommendations can allow designers to better understand the users that their app is intended to target and their expectations, ultimately influencing design arrangement. That being said, a participant stated that '*market analysis helps in identifying the target audience and their specific needs*' by delivering information related to '*user demographics, preferences, and behaviour, which can guide the design decisions*' ($P35$). Such characteristics can make designers infer various design indications, concerning e.g., the app's colour scheme or aesthetic and ergonomic styles – '*if* [...] *users prefer a minimalist interface, it would be beneficial to incorporate clean and simple design elements*' ($P35$) –, in addition to current trends, '*common patterns and best practices*' ($P23$). Analysing this aspect can also give designers '*a sense of what works well and what doesn't*' ($P23$), i.e., '*successful design patterns*' ($P31$, $P38$, $P43$, $P62$, $P63$, $P64$) and '*design flaws and pitfalls*' ($P65$, $P66$, $P67$), in addition to '*valuable insights* [regarding] *industry standards*' ($P19$, $P38$, $P68$).

Moreover, being able to get inspiration from well-known or very popular apps' designs may conclusively improve the UX of a new application, as it provides a sense of familiarity that people often are comfortable with, a phenomenon that we have previously discussed. In this context, it is believed that '*users can instantly recognise the scope of interactions just from the instant impression of the layout*' ($P68$). In contrast, although implementing new design directions is an obligation in the mobile GUI field, this student ($P68$) is convinced that this can lead to lose the user in the app's navigation and significantly degrade UX:

> '*Straying too far from designs that were built with the current market in mind can lead to an interface feeling foreign or*

[28] https://images.google.com/
[29] https://www.midjourney.com/home
[30] https://chat.openai.com/

unfamiliar at worst and simply being broken and unusable by many users.' ($P68$)

Yet another argument frequently encountered among the provided responses is that the utilisation of established UI design recommendations obviates the necessity for designers to engage in permanent and often useless reinvention. By receiving advice on already established design patterns and conventions, designers therefore feel neither the need nor the interest to '*reinvent the wheel*' ($P17$, $P23$, $P39$, $P69$). Indeed, it is further argued ($P39$), in simple and meaningful terms, that '*if something works, one should draw ideas from it*,' confirming the popular adage '*if it ain't broke, don't fix it!*' (Laudan, 1989).

On the other hand, even if almost all participants recognise the already discussed benefits of GUI design recommendations, an important number of opinions are interestingly nuanced. They indeed think about the lack of creativity and uniqueness in the interfaces that this can cause, a criticism that has already been made before in the context of designers' inspiration sources:

'*Relying too much on existing designs leads to a lack of uniqueness and very stale looking designs overall because everything looks quite the same, depending on the current trend.*' ($P37$)

Some designers further recommend careful examinations of the proposed advice, in order to still '*make sure the app/site has a unique look and feel to differentiate it from competitors, and to provide "something extra" to the users*' ($P70$). What designers mean is that recommendations should only be considered as the extra help that they represent and intend to be, which reaches consensus, without '*follow*[ing] *them blindly*' ($P31$, $P63$, $P66$, $P71$, ...). One should not let these limit '*design and creativity choices*' ($P34$) that may '*suffer from* [them]' ($P70$).

4.3.2. Key performance indicators

The identification of relevant KPIs was a step further in the elaboration of Enricommender's most desired features. In the context of this article, we define KPIs as the set of data attributes that Enricommender's target users would value the most in their use of the system. The said attributes correspond to Enrico's (Leiva et al., 2020a) app-specific metadata (cf. Table 2) and were given to participants in order to be individually rated on a scale of importance, from 1 to 5, in terms of usefulness, as explained earlier in the survey's procedure.

Using this scale and the scores derived from it, we were able to define six KPIs corresponding to the top-6[31] attributes, i.e., the ones that have been given the higher scores by the participants (cf. Table 3). All of the determined KPIs have a score higher than 4 out of 5, putting into light the great importance of these attributes for the interviewed participants. The attributes' final score, denoted as $\overline{x}$, have been computed using the average method, i.e., arithmetic mean of all the unique scores x_i given by all participants (with $N = 202$, the number of participants). Furthermore, to better assess the score distribution and dispersion, we have also computed the median $\widetilde{x}$ and standard deviation σ of the series.

The list of KPIs includes the following attributes from the app's metadata:

(i) `score`: the application's score;
(ii) `installs`: the number of downloads of the application;
(iii) `free`: whether the application is free or not;
(iv) `price`: the application's selling price in Google Play, in US dollars;
(v) `category`: the application's Google Play category;
(vi) `screenshots`: the list of the application's Google Play screenshots.

4.3.3. Anticipated features

To finish, participants have been asked what they would expect from a BI recommender system like Enricommender. We asked the question in a more direct way than before in order to clearly identify target users' needs and prospects, explicitly, in terms of features or layout.

The analysis and coding of the participants' answers were performed using both Microsoft Excel and the MAXQDA[32] qualitative and quantitative data analysis software. This examination, as well as the diversity and creativity seen in the responses, permitted the identification of more than 50 requested features. We have decided to consider as features to implement in Enricommender the elements that have been cited by at least 5% of the participants, i.e., 11 people minimum. This left us with a set of 19 items, populated in Table 4.

A detailed explanation of the features' exact meanings and how they were implemented in the prototype of Enricommender is discussed in the next section.

4.4. Enricommender

In this section, we describe our high-fidelity prototype of Enricommender and detail its features, what they mean and how they are related to the expectations that the participants have expressed.

[31] The study was initially limited to five KPIs. However, as there was a tie between two items, we have included an additional one, giving six KPIs.

[32] `https://www.maxqda.com/`

Table 3. Averaged scores of the metadata elements, with the top-5 (or top-6 in case of ties) KPIs for each participant group annotated. Global averages ($\overline{x}$), medians ($\widetilde{x}$), and standard deviations (σ) are also indicated. Results are rounded to the nearest hundredth.

Metadata element	Professionals	Students	Others	$\overline{x}$	$\widetilde{x}$	σ
App description	4	3.91	4.05[5]	**3.98**	**4**	**1**
App name	3.91	4.13[4]	3.81	**3.97**	**4**	**1.16**
App size	3.38	3.28	3.62	**3.37**	**3**	**1.23**
App version	2.75	2.34	2.81	**2.63**	**2**	**1.3**
Category	4.22[4]	3.91	4.33[3]	**4.13[4]**	**4**	**0.95**
Content rating	3.69	3.66	3.81	**3.69**	**4**	**1.12**
Developer name	2.53	2.48	2.67	**2.53**	**2**	**1.15**
Editors' Choice	3.44	3.48	3.43	**3.46**	**4**	**1.05**
Free	4.19[5]	4.14[3]	4	**4.15[3]**	**4**	**0.92**
Icon	3.56	3.77	3.76	**3.65**	**4**	**1.2**
In-app purchases	3.83	3.47	3.86	**3.72**	**4**	**1.08**
IAP range	3.32	3.19	3.29	**3.28**	**3**	**1.21**
Interactive elements	3.68	3.56	4	**3.68**	**4**	**1.07**
Last updated	3.72	3.52	3.95	**3.68**	**4**	**1.12**
Number of installs	4.34[2]	4.17[2]	4.43[1]	**4.3[2]**	**4**	**0.84**
Number of reviews	4.09	3.77	4.05[5]	**3.99**	**4**	**0.94**
Price	4.3[3]	3.94	3.81	**4.13[4]**	**4**	**0.94**
Recent changes	3.51	3.45	3.52	**3.5**	**4**	**1.09**
Min. Android version	3.82	3.28	3.81	**3.65**	**4**	**1.21**
Score	4.43[1]	4.28[1]	4.38[2]	**4.38[1]**	**5**	**0.77**
Screenshots	4.04	3.97[5]	4.19[4]	**4.03[6]**	**4**	**1.04**

4.4.1. Implementation

The high-fidelity prototype of Enricommender has been designed using Figma,[33] a popular collaborative interface prototyping design platform. Enricommender's logo was created using the BrandCrowd[34] logo generator with the keywords 'letter E in a light bulb.' We wanted the logo to contain (1) the initial letter of Enricommender, in order to make the logo less generic and to link it to the product, and (2) an illustration of a light bulb, that often represents the concepts of knowledge or ideas. We think that these notions fit well with Enricommender, since giving ideas to people is one of its main goals.

Charts and graphs, i.e., the visual representation of the statistics we have implemented into our dashboard interface (cf. Figure 7), have been created using the Figma *Charts* plugin.[35] It offers multiple representation options, including line, scatter, area, bar, and pie charts. In addition, our prototype makes use of the *IBM Plex Sans*[36] font across its interface, that we have chosen for its clean, simple, and readable look. To finish, the icons we have worked with belong to the Apple *SF Symbols*[37] collection, which contains over 5,000 symbols. They are, in our opinion, neat, consistent, and easy to download and use on a Mac computer, which is what we were using to realise this article.

4.4.2. Landing page

The home page of Enricommender corresponds to a landing page that is visible in Figure 6. It introduces a quick description of what the system consists of, shows some illustrative screenshots contained in the Enrico dataset (Leiva et al., 2020a) and includes four drop-down menus. The drop-down menus correspond to the search filters that we put into place (**feature #9** in Table 4). Following what participants have asked for and the KPIs we have set, one can filter search results by app category, ratings (or scores, #**5**), number of installs (#**13**), and price. With such a system, designers would therefore have the option to decide which kind(s) of applications they want to be taken into account for the statistics and recommendations provided by Enricommender.

[33] https://www.figma.com/

[34] https://www.brandcrowd.com/maker/logos

[35] https://www.figma.com/community/plugin/731451122947612104

[36] https://www.ibm.com/plex/

[37] https://developer.apple.com/sf-symbols/

Table 4. List of features and characteristics to be implemented in Enricommender, as requested by the survey participants.

#	Feature	Occur.	Rate
1	Perform comparative analysis	62	30.69%
2	Have a simple and intuitive UI	51	25.25%
3	Provide visual elements (graphs, charts, screenshots, videos)	47	23.27%
4	Propose a customisable dashboard interface	35	17.33%
5	Show, analyse and search by user reviews and/or ratings	35	17.33%
6	Identify trends based on market analysis	34	16.84%
7	Provide user flows and behaviour analysis	32	15.84%
8	See recommendations based on users, location or context	31	15.35%
9	Propose search and filtering features	28	13.86%
10	Give suggestions and/or show characteristics of colour schemes, fonts and UI elements	24	11.88%
11	Highlight and/or propose app features	19	9.41%
12	Show the number and most used screens	18	8.91%
13	Show and search by number of downloads	17	8.42%
14	Interoperability with external services	15	7.43%
15	App sorting	15	7.43%
16	Propose an expandable and granular dashboard interface	14	6.93%
17	Give insights about apps' target users	12	5.94%
18	Split good/bad reviews and provide dos/don'ts	11	5.45%
19	Export data	11	5.45%

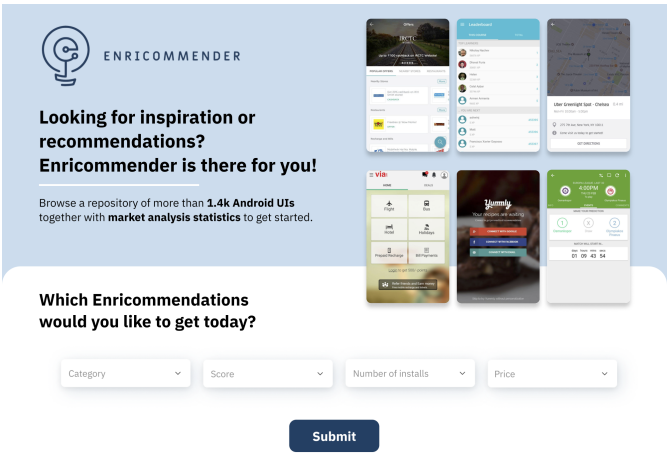


Figure 6: Enricommender's landing page.

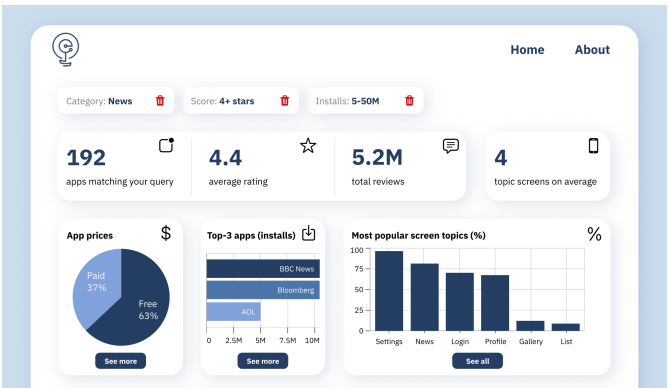


Figure 7: Overview of Enricommender's BI dashboard. This example relates to applications belonging to the 'News' category, rated at least 4 stars, and downloaded between 5 million and 50 million times.

4.4.3. Market analysis-based statistics

Once a request is processed, i.e., by selecting the applicable filters and submitting the request, users face a BI-like dashboard interface, divided in two main parts: (1) market-based statistics, with respect to the KPIs and apps in Enrico (Leiva et al., 2020a) that match the treated query and set of criteria provided by the user, and (2) GUI design recommendations, which we have named by a pun 'Enricommendations,' based on the same apps. In this part, we will only discuss the first one mentioned; the second follows in the next section.

The aforementioned statistics (cf. Figure 7) are first composed of general information, like the number of apps matching the query, their average rating (from 1 to 5 stars), the total number of reviews (#**5**), and the average number of screen topics (#**12**) across the apps. The views are expandable, which is a feature that many participants requested (#**16**), thus lightening the main interface. Maintain a relative simplicity in the UI was an important consideration for us, as it represents the second most requested characteristic for Enricommender (#**2**). The expanded view contains more detailed information, e.g., the list of matching applications with their name, score, number of reviews, and number of installs. From there, users are further able to access app-level information, e.g.,

a short description and a link to the Google Play page of the app, alongside app screenshots.

Below this first set of cards, one can find other statistics that are based on BI-associated visualisation techniques, i.e., charts and graphs, which was another highly requested aspect (#**3**). Furthermore, a more comparative analysis approach has been adopted, which is participants' most requested feature (#**1**). From the majority view, they are mainly demanding comparisons that show similarities and differences between GUIs. Such analogies and contrasts show designers what others are concretely doing, and may permit them to identify common patterns and trends (#**6**).

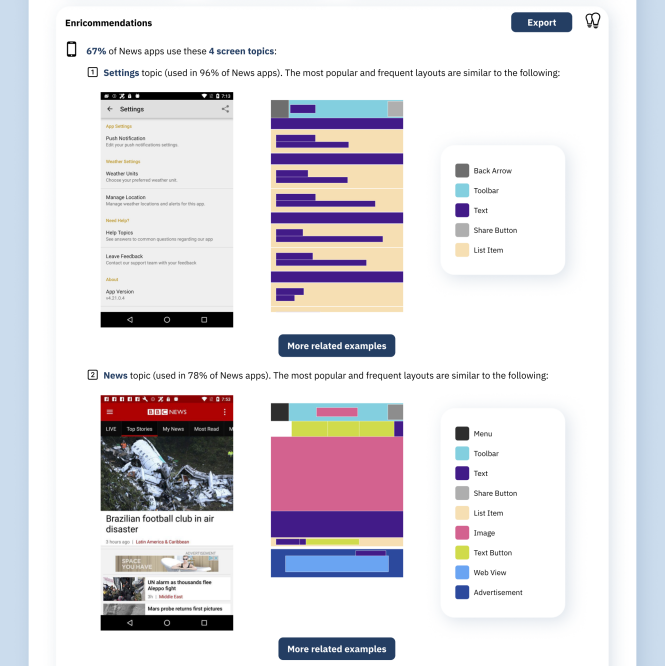
Concretely, we are showing here information like price distribution of the apps matching the specified criteria under the form of a pie chart, top apps with respect to their number of downloads and ratings, as well as the top screen topics used and their frequency of occurrence in the apps. These cards are also expandable to reveal more detailed information. This includes a more precise app prices distribution and the highlight of the presence of IAPs within the apps, aesthetic and semantic side-by-side comparisons between top apps, using screenshots and wireframes, respectively, and screen topic examples.

4.4.4. Enricommendations

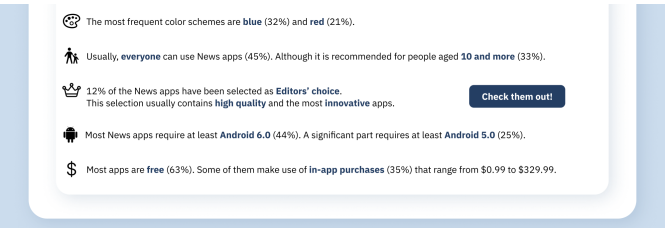
Below the market analysis figures and statistics is located another card (cf. Figure 8), which constitutes the second important part of our dashboard and the main goal of Enricommender: the Enricommendations, a pun on 'Enrico' and 'recommendations,' to be understood as 'recommendations based on Enrico.' These recommendations sometimes rely on previous statistics and are designed to provide to mobile GUI designers meaningful advice and yet additional comparative analysis statistics to support and facilitate inspiration seeking practices.

The first set of Enricommendations we have included relate to screen topics (cf. Figure 8a), as many designers expressed their need to have access to screenshots of apps that are similar to the one the interface should be designed for (#**3**). Here, we use of some of the statistics from before to specify the number of screen topics that are usually included in this category of apps, together with real screenshots and wireframes to allow designers to understand the interfaces, not only visually, but also semantically. Users are also able to expand these results further (#**16**) to see additional examples. By seeing real-life popular design examples, designers may have the opportunity to deduce what should, or should not, be done in their design, i.e., they can figure out dos and don'ts (#**18**).

Furthermore, designers can obtain aesthetic and business-oriented indications (cf. Figure 8b), e.g., popular



(a) 'Settings' and 'News' screen topic recommendation examples.



(b) Other UI and business-related recommendations.

Figure 8: Recommendations, or 'Enricommendations,' provided by Enricommender based on the processed query. Presented data is used for illustrative purposes.

colour schemes or potential high-quality designs from apps that have been selected as Editors' Choice. Such applications often stand out and are recognised as innovative. Enricommendations also provide information about typical app prices for a given category, IAPs, content rating, or the minimum versions of Android required to install apps similar to the application of interest. To finish, we have considered a way to export these recommendations, e.g., in PDF format, which is an additional feature requested by the participants (#**19**).

5. DISCUSSION

As highlighted many times in the research literature, GUI design is a challenging and time-consuming process that

often requires designers to look at other people's designs for inspiration-seeking purposes. However, the analysis of previous works showed a lack of semantic interpretation of these pre-existing interfaces. We have originally hypothesised that using these interfaces to provide GUI designers with market-based recommendations could constitute a precious help for them. Starting from that, we envisioned creating a system which serves this purpose. Specifically, we have created a questionnaire targeting designers experienced with UI design and/or UX. The collected data informed us about existing main GUI design challenges, the way designers usually seek for inspiration, how recommendations could help them in their creation process, and what they would be expecting from a BI recommender system. Such a system, that we have called Enricommender and which takes the form of a high-fidelity dashboard prototype, constitutes our answer to this set of information.

5.1. RQ1: Interest of designers towards a business intelligence design system

The answer to our RQ1 is multidimensional and should cover potential reasons that might motivate mobile GUI designers to use Enricommender. Part of the answer could be provided from the perspective of currently existing GUI design challenges that were promptly identified by our survey participants. Indeed, designers may be tempted to resort to GUI design recommendations and mobile app UI related information, that our Enricommender system aims to offer, to overcome these challenges. Statistics and recommendations provided by our tool could be a way to address them, at least partially. On the one hand, real-world app UI examples, that could be sorted to fit individual preferences and from which the best interfaces can be extracted, can give to designers valuable insights regarding how others tackled the previously mentioned challenges. Individual examples and comparisons between them can point out similarities and differences, in addition to representing, in some way, authority arguments regarding good design practices and proper ways to handle potentially challenging design concepts. On the other hand, statistics are meant to go beyond UI-related concerns and can give indications of the current state of the market or answer business-oriented questions, so that the application of interest, i.e., the one for which the interface is currently designed, can fit in as good as possible.

Furthermore, the way designers usually seek for inspiration, and especially the tendency they have to look for pre-existing designs can constitute another reason for them to use Enricommender. We recall that nearly one in two designers (45.54%) admitted to have recourse to this inspiration-seeking practice. This therefore constitutes a significantly widespread habit, albeit to a lesser extent than what we have originally anticipated. In absolute terms, the fact that nearly half of the designers look at competitors' interfaces remains substantial and still makes this practice common. However, as the literature acknowledges (Bonnardel, 1999; Herring et al., 2009; Gonçalves et al., 2014; Vasconcelos and Crilly, 2016; Koch et al., 2018; Leiva et al., 2020b) that such inspiration seeking habits are eminently frequent, we would have expected a much higher number at first. Nonetheless, this number, together with the collected answers, not only confirms, once again, that designers actually need to have access to already-existing designs, but also suggests that there exists some demand for an Enricommender-like tool, i.e., that would go further than a GUI repository. Designers recognised, in our survey, that comparative analyses, i.e., comparisons between similar apps, can end up being genuinely convenient for them to spot current design trends, identify good and bad design practices, better understand some of the needs of their target users, and, overall, to stimulate their creativity. Comparative analysis being at the core of Enricommender, we think that providing designers with market- and comparison-based recommendations represents an opportunity to make their research process both easier and, hopefully, significantly faster.

This point seems to be further corroborated by some answers we collected in our survey, and especially when designers were directly asked about their opinion towards getting comparative analysis-based UI recommendations. It seems that roughly all the participants value receiving such instructions, that they foresee as helpful to guide them in their design creation, especially in its early stages. Designers indeed saw recommendations as an helpful guidance, that they even compared to another person's opinion potentially mentoring the convenience of the end product. Thus, the conformity of an interface with the market in which it is part, induced by its respect for guidelines based on popular, effective, and established design patterns, leads to a sense of familiarity for users, allowing them to use the application more efficiently. Again, designers know, using their words, that there is no need to reinvent the wheel, at the risk of doing something wrong, losing users, and ultimately affecting UX and the success of the app altogether. Following recommendations based on pre-existing material is, in our opinion, a good way to avoid trying to fix something that does not need to be fixed in the first place.

As we have seen, there does not exist one single reason that encapsulates the interest of designers in a tool like Enricommender. In line with our initial hypothesis, the gathered data gives a clearer understanding of an enthusiasm that, certainly, exists. Recommendations can be useful for studying how others have tackled certain

design challenges and understanding how to address them oneself. Additionally, Enricommender's statistics and advice constitute a source of inspiration dedicated to designers valuing comparison-based market analyses to stimulate their creativity and study the market's state. Enricommender further plays a guiding role, helping to shape design and ideas with respect to effective and established trends and principles that there is no need to change, at least in the short term.

5.2. RQ2: Features to implement for satisfying user needs

The second research question leads us to reflect on the best possible way to satisfy the needs expressed by designers regarding a BI recommender system. With Enricommender, we have taken into account both quantitative and qualitative data, that is, the identification and assessment of KPIs, and the features designers would anticipate, respectively. On the one hand, the set of six KPIs was used in Enricommender as search filters and as a basis for the market-based statistics. For the filters, we have used five out of six KPIs, i.e., applications' score, number of installs, category, price and whether free or not, that were implemented as interactive drop-down lists in Enricommender's landing page (cf. Figure 6). App screenshots, corresponding to the last KPI, are used in multiple places throughout Enricommender's interface for illustrative and comparative purposes, as the survey participants requested. Concerning the number of KPIs to be included in Enricommender, we have judged as crucial metadata elements that obtained a mean score of 4 stars and up, as it denotes a rather high level of importance for the participants. Moreover, the study of the medians puts into light a score distribution centred on high values, i.e., 4 or 5, suggesting a sort of consensus from designers as to the importance of this data. This point is further corroborated by the observation of the standard deviations, whose values corresponding to the selected KPIs are the lowest among all the metadata elements, thus indicating a lower dispersion of the given scores. By taking such an approach, we wanted (1) to make sure that the KPIs represent an interest for designers and (2) to keep things simple, rather than potentially offering too many possibilities that may overwhelm and disengage our users. This choice should nonetheless be empirically evaluated with the users who are the main stakeholders of the system. It is also worth noting that participants' demands, expressed in free-form text, often involved the defined KPIs. Indeed, their requests often included the search, sort, display, and analysis of applications and their UIs based on the KPIs and other metadata elements, e.g., user reviews, ratings, and number of downloads.

On the other hand, we have derived from the participants' expectations a list of 19 features (cf. Table 4) that were to be included in Enricommender. We have considered to implement the capabilities that were requested by at least 5% of the participants. An initial arbitrary threshold of 10% was originally defined, to ensure the richness and representativeness of the data. However, using such an occurrence rate leads to the selection of 10 features only, which, in our opinion, is not enough and leaves some significant demands aside. Consequently, we indeed believed that choosing a threshold of 5% instead would still be reasonable, considering the time constraints we have, and achieve a satisfactory level of representativeness, as our population is fairly homogeneous. Participants have been asked regarding their expectations in terms of features, layout and presentation of the information. Obtaining a comparative information and analysis between similar apps, which are the main objectives of Enricommender, were functionalities most frequently mentioned by the participants. Comparisons are realised and presented in our system both under the form of statistics and recommendations. Statistics make use of the Enrico (Leiva et al., 2020a) metadata we have at our disposal, and connects matching apps' scores, prices, number of installs, and number of screens with each other. Requests for such app-level comparisons have been found in an important number of answers, so that '*designers can select several* [...] *apps and compare their average ratings, user reviews, and pricing to gain a better understanding of how they differ and make informed decisions in their own design process*' ($P66$). Furthermore, UI- and semantic-level comparisons are also permitted by Enricommender, as designers can have a look at different design choices for multiple apps' screens sharing similar purposes, i.e., '*lineup*[s] *of screenshots comparing analogous elements of multiple applications*' ($P55$). Side-by-side UI comparisons can also allow designers to observe and deduce design trends, even though the prototype is currently not able to identify these trends for them. By our survey participants' admission, such comparative analyses would be time saving in their research phase, hence constituting a must-have feature that could serve as a basis for new designs.

Furthermore, more than one out of four participants explicitly asked for a UI that should be '*simple*,' '*intuitive*,' and '*easy to follow*,' among other terms. Even though these criteria are subjective, we have done our best to keep a limited amount of information available at once, relying instead on a clickable and expandable approach to reveal more advanced information. This granular and expandable capabilities of the dashboard interface has also been requested multiple times. For example, general app category-level statistics are only composed, in the main dashboard view, of three elements: the number of

apps matching the query made by the user, the average rating of all these apps, and the total number of reviews. The same goes for the charts below these numbers: they only contain high-level information, like a simple free/paid distribution without precise price details, the name of the top-3 apps with their corresponding number of installs, and the most used screen topics within these apps. More in-depth information, e.g., screenshots, were not included in this view. Moreover, clickable elements are indicated by clearly recognisable buttons to facilitate navigation in the dashboard. These mentioned elements constitute some examples of how we tried to keep Enricommender's UI as simple and easy to use as possible.

The presence of visual elements to convey information is the third most requested feature of Enricommender. The use of efficient and relevant data visualisation are the essence of BI dashboards. Mentioned many times already, data can be apprehended, in our prototype, using text, pie and bar charts, app screenshots, and semantic wireframes. We have tried, for each type of information, to choose the most appropriate way to represent it. Indeed, numerical comparison data is often represented by graphs, to relate a specific number to its associated context, while app interface visualisations or comparisons are directly conveyed by screenshots and/or wireframes.

In designers' opinion, a mobile UI recommender system should be capable of providing some suggestions or showing information related to apps' colour schemes, fonts, and UI elements. Enricommender builds on this existing need by giving information about predominant colour schemes (cf. Figure 8b) and most popular screen layouts for given screen topics, together with their associated UI elements. Moreover, Enricommender is also capable of highlighting apps' number of different screens, their topics, and the most used ones for a given set of criteria.

An additional request was about adding to Enricommender contrasts between positive and negative reviews, as well as between great and poor design examples, to provide advice on things to do and things to avoid when designing mobile UIs. This feature is handled by sorting capabilities. For example, in top apps views, results can be sorted by ascending or descending number of installs, as well as by highest or lowest ratings, hence giving users the possibility to obtain UI information about the most/least downloaded and the top/lowest rated apps, and deduce associated similarities and differences.

Last but not least, the need for an export feature of Enricommender's data has been taken into consideration and implemented into an Export button at the top of the Enricommendations tile. We therefore suggest that, in a production-ready version, Enricommendations will be able to be exported, e.g., in PDF format.

We think that the design of a market analysis recommender system like Enricommender should be executed in consultation with the primarily targeted users, i.e., UI/UX designers. Our approach has been to question them directly using an online survey, allowing us to study their difficulties, backgrounds and needs, resulting in a list of KPIs and expected features that Enricommender should satisfy in order to be relevant. Applications, sorted and filtered by a set of user-defined criteria, are compared with each other from three perspectives: app-level, UI-level and semantic-level. Furthermore, results are displayed visually when relevant in an expandable interface that aims to be as simple as possible. Nonetheless, some of the features that designers were expecting from Enricommender could not be implemented in the current state of our prototype. Such shortcomings are discussed in the next section.

5.3. Limitations

The findings of this study are subject to some limitations. First, Enricommender's capabilities are bounded to what a high-fidelity prototype, and not a full production-ready version, allows. Indeed, our main goals were to measure designers' potential interest toward a market analysis-based recommendations for GUI design, and to gather initial feedback from them regarding their expectations towards such a recommender tool. Furthermore, we also wanted to create a prototype of an interactive BI application that would give a first sense of how such a system could look like, as well as its more important features. As a consequence, our prototype is mostly illustrative and does not feature functional logic. In addition, some of the survey participants' requested features (cf. Table 4) were especially challenging to implement; for example the customisation of the dashboard interface, requested by more than 17% of the questioned designers. In addition, a trend analysis performed without involving users' thinking, as well as the development of dos and don'ts, are interesting features to consider in future work. Lastly, a desired interoperability with external services, e.g., APIs or data sharing to other platforms, are also planned for future work.

Limitations inherent to the dataset we have chosen for our study could also be added to the previous points. For example, participants wanted Enricommender to make use of user app reviews and user flows within apps – to conduct behaviour analysis –, which unfortunately are not available in Enrico. The same applies to location-based information, e.g., top apps per country, getting insights about apps' target users, and, last but not least, app features that some participants requested to get recommendations about. The implementation of such features is therefore not feasible by relying solely on

the Enrico dataset. It is also worth noting that Enrico contains app UI screenshots coming from the Rico dataset, which dates back to 2017, and metadata information from Enrico as of 2020, which have, to our knowledge, not been updated since. Consequently, a significant amount of information in Enrico is probably outdated. Nonetheless, Enrico remains a satisfactory basis for defining the characteristics and goals of a tool like Enricommender, and was the best suited for the purpose of this study.

6. CONCLUSION

This work provides valuable insights on how additional support can be provided for the ideation and design of graphic interfaces, in response to a lack of GUI comparative-based analysis in the research literature. By surveying more than 200 designers, composed of professionals, students, developers, and more, we have been able to identify their difficulties, practices, and expectations towards getting GUI design recommendations and regarding Enricommender, a business intelligence tool for comparative analysis of GUI designs and app recommender system. The collated expectations resulted in a list of functionalities that permitted the design of a high-fidelity prototype of Enricommender. The insights gained from our analysis of all the survey's responses have highlighted a fairly high level of interest in an BI system that, to the best of our knowledge, has never been proposed before. The implementation of a production-ready version of Enricommender is left as an opportunity for future work.

ACKNOWLEDGEMENTS

This research is supported by the Horizon 2020 FET program of the European Union through the ERA-NET Cofund funding (project BANANA, grant CHIST-ERA-20-BCI-001) and Horizon Europe's European Innovation Council through the Pathfinder program (SYMBIOTIK project, grant 101071147).

DATA AVAILABILITY

The survey data underlying this article are available on reasonable request to the corresponding author.

REFERENCES

Arens, Y., Miller, L., Shapiro, S., and Sondheimer, N. (1988). *Automatic Construction of User-Interface Displays.*

Ballou, D. P. and Pazer, H. L. (1985). Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems. *Management Science*, 31(2):150–162.

Beltramelli, T. (2017). Pix2code: Generating Code from a Graphical User Interface Screenshot.

Bonnardel, N. (1999). Creativity in design activities: The role of analogies in a constrained cognitive environment. In *Proceedings of the Third Conference on Creativity & Cognition - C&C '99*, pages 158–165, Loughborough, United Kingdom. ACM Press.

Bunian, S., Li, K., Jemmali, C., Harteveld, C., Fu, Y., and El-Nasr, M. S. (2021). VINS: Visual Search for Mobile User Interface Design.

Cao, H. H., Han, B., Hirshleifer, D., and Zhang, H. H. (2011). Fear of the Unknown: Familiarity and Economic Decisions. *Review of Finance*, 15(1):173–206.

Chan, J., Siangliulue, P., Qori McDonald, D., Liu, R., Moradinezhad, R., Aman, S., Solovey, E. T., Gajos, K. Z., and Dow, S. P. (2017). Semantically Far Inspirations Considered Harmful? Accounting for Cognitive States in Collaborative Ideation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, C&C '17, pages 93–105, New York, NY, USA. Association for Computing Machinery.

Deka, B., Huang, Z., Franzen, C., Hibschman, J., Afergan, D., Li, Y., Nichols, J., and Kumar, R. (2017). Rico: A Mobile App Dataset for Building Data-Driven Design Applications. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 845–854, Québec City QC Canada. ACM.

Deka, B., Huang, Z., and Kumar, R. (2016). ERICA: Interaction Mining Mobile Apps. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 767–776, Tokyo Japan. ACM.

Eckert, C. and Stacey, M. (2000). Sources of inspiration: A language of design. *Design Studies*, 21(5):523–538.

Eisenstein, J., Vanderdonckt, J., and Puerta, A. (2002). Applying Model-Based Techniques to the Development of UIs for Mobile Computers.

Fleckenstein, M. and Fellows, L. (2018). *Modern Data Strategy*. Springer.

Gonçalves, M., Cardoso, C., and Badke-Schaub, P. (2014). What inspires designers? Preferences on inspirational approaches during idea generation. *Design Studies*, 35(1):29–53.

Grudin, J. (1989). The case against user interface consistency. *Communications of the ACM*, 32(10):1164–1173.

Hashimoto, Y. and Igarashi, T. (2005). Retrieving Web Page Layouts using Sketches to Support Example-based Web Design.

Herring, S. R., Chang, C.-C., Krantzler, J., and Bailey, B. P. (2009). Getting inspired!: Understanding how and why examples are used in creative design practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 87–96, Boston MA USA. ACM.

Huang, F., Canny, J. F., and Nichols, J. (2019). Swire: Sketch-based User Interface Retrieval. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–10, Glasgow Scotland Uk. ACM.

Huang, F., Li, G., Zhou, X., Canny, J. F., and Li, Y. (2021). Creating User Interface Mock-ups from High-Level Text Descriptions with Deep-Learning Models.

Inukollu, V. N., Keshamon, D. D., Kang, T., and Inukollu, M. (2014). Factors Influncing Quality of Mobile Apps: Role of Mobile App Development Life Cycle. *International Journal of Software Engineering & Applications*, 5(5):15–34.

Koch, J., László, M., Lucero, A., and Oulasvirta, A. (2018). Surfing for Inspiration: Digital inspirational material in design practice. In *Design Research Society Conference 2018*.

Kumar, R., Satyanarayan, A., Torres, C., Lim, M., Ahmad, S., Klemmer, S. R., and Talton, J. O. (2013). Webzeitgeist: Design mining the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3083–3092, Paris France. ACM.

Laudan, L. (1989). If It Ain't Broke, Don't Fix It. *The British Journal for the Philosophy of Science*, 40(3):369–375.

Lee, C., Kim, S., Han, D., Yang, H., Park, Y.-W., Kwon, B. C., and Ko, S. (2020). GUIComp: A GUI Design Assistant with Real-Time, Multi-Faceted Feedback.

Lee, P. and Sunder, M. (2016). The Law of Look and Feel. *SSRN Electronic Journal*.

Leiva, L. A., Hota, A., and Oulasvirta, A. (2020a). Enrico: A Dataset for Topic Modeling of Mobile UI Designs. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–4, Oldenburg Germany. ACM.

Leiva, L. A., Hota, A., and Oulasvirta, A. (2020b). Interactive Exploration of Large-Scale UI Datasets with Design Maps. *Interacting with Computers*, 32(5-6):490–509.

Li, T. J.-J., Popowski, L., Mitchell, T. M., and Myers, B. A. (2021). Screen2Vec: Semantic Embedding of GUI Screens and GUI Components. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15.

Li, Y., He, J., Zhou, X., Zhang, Y., and Baldridge, J. (2020). Mapping Natural Language Instructions to Mobile UI Action Sequences.

Mahanti, R. (2019). *Data Quality: Dimensions, Measurement, Strategy, Management, and Governance*. Quality Press.

Moran, K., Bernal-Cárdenas, C., Curcio, M., Bonett, R., and Poshyvanyk, D. (2018). Machine Learning-Based Prototyping of Graphical User Interfaces for Mobile Apps.

Nilsson, E. G. (2009). Design patterns for user interface for mobile applications. *Advances in Engineering Software*, 40(12):1318–1328.

Perttula, M. and Sipilä, P. (2007). The idea exposure paradigm in design idea generation. *Journal of Engineering Design*, 18(1):93–102.

Pipino, L., Lee, Y., and Wang, R. (2003). Data Quality Assessment. *Communications of the ACM*, 45.

Ritchie, D., Kejriwal, A. A., and Klemmer, S. R. (2011). D.tour: Style-based exploration of design example galleries. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, pages 165–174, Santa Barbara California USA. ACM.

Romanenkova, J. (2020). You are not the customer. And here's why. `https://uxdesign.cc/you-are-not-the-customer-and-heres-why-3aa94a6701fa`.

Satzinger, J. W. and Olfman, L. (1998). User Interface Consistency across End-User Applications: The Effects on Mental Models. *Journal of Management Information Systems*, 14(4):167–193.

Sharmin, M., Bailey, B. P., Coats, C., and Hamilton, K. (2009). Understanding knowledge management practices for early design activity and its implications for reuse. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2367–2376, New York, NY, USA. Association for Computing Machinery.

Swearngin, A., Dontcheva, M., Li, W., Brandt, J., Dixon, M., and Ko, A. J. (2018). Rewire: Interface Design Assistance from Examples. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Montreal QC Canada. ACM.

Todi, K., Leiva, L. A., Buschek, D., Tian, P., and Oulasvirta, A. (2021). Conversations with GUIs. In *Designing Interactive Systems Conference 2021*, pages 1447–1457, Virtual Event USA. ACM.

Vasconcelos, L. A. and Crilly, N. (2016). Inspiration and fixation: Questions, methods, findings, and challenges. *Design Studies*, 42:1–32.

Voutilainen, J.-P., Salonen, J., and Mikkonen, T. (2015). On the Design of a Responsive User Interface for a Multi-device Web Service. In *2015 2nd ACM International Conference on Mobile Software Engineering and Systems*, pages 60–63.

Wallace, S., Le, B., Leiva, L. A., Haq, A., Kintisch, A., Bufrem, G., Chang, L., and Huang, J. (2020). Sketchy: Drawing Inspiration from the Crowd. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–27.

Wang, B., Li, G., Zhou, X., Chen, Z., Grossman, T., and Li, Y. (2021). Screen2Words: Automatic Mobile UI Summarization with Multimodal Learning.