

Research Article

Open Access

Bernard J. Jansen*, Joni O. Salminen, Soon-Gyo Jung

Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics

<https://doi.org/10.2478/dim-2020-0005>

received January 4, 2020; accepted February 22, 2020.

Abstract: Persona is a common human-computer interaction technique for increasing stakeholders' understanding of audiences, customers, or users. Applied in many domains, such as e-commerce, health, marketing, software development, and system design, personas have remained relatively unchanged for several decades. However, with the increasing popularity of digital user data and data science algorithms, there are new opportunities to progressively shift personas from general representations of user segments to precise interactive tools for decision-making. In this vision, the persona profile functions as an interface to a fully functional analytics system. With this research, we conceptually investigate how data-driven personas can be leveraged as analytics tools for understanding users. We present a conceptual framework consisting of (a) persona benefits, (b) analytics benefits, and (c) decision-making outcomes. We apply this framework for an analysis of digital marketing use cases to demonstrate how data-driven personas can be leveraged in practical situations. We then present a functional overview of an actual data-driven persona system that relies on the concept of data aggregation in which the fundamental question defines the *unit of analysis* for decision-making. The system provides several functionalities for stakeholders within organizations to address this question.

Keywords: persona, analytics, data-driven personas, data science, conceptual analysis

1 Introduction

Personas are imaginary persons that are employed as representing segments of real people within a population (Cooper, 2004; Pruitt & Grudin, 2003). The use of personas is well established in the field of human-computer interaction (HCI) as a method for understanding user populations. The population segment represented by a persona can be customers or users of content, a product, or a system, respectively. Traditionally, personas are usually generated from qualitative data gathered from surveys, focus groups, or related data collection methods (Miaskiewicz & Luxmoore, 2017; Pruitt & Grudin, 2003), meaning the data were manually gathered and then manually analyzed to develop the personas (Pruitt & Adlin, 2006).

From their inception in the late 1980s, refinement in the early 1990s (Cooper, 2004), and following decades of use, personas were generally “flat media” (i.e., presented on paper or in an electronic document), encompassed within a one-page or a two-page summary called a persona profile (Nielsen & Storgaard Hansen, 2014). A persona profile generally (Figure 1¹) contains an assortment of attributes concerning the persona, including name, gender, age, goals, demographic attributes, pain points, a headshot image, a quote, and so on. As such, from their initial inception through numerous decades of use, personas were primarily data structures—not interactive decision-making tools. A persona profile was a mechanism to organize the data concerning the imaginary person to present the information that decision-makers and others leverage to accomplish their goals of better targeting customers/users/audiences.²

*Corresponding author: Bernard J. Jansen, Qatar Computing Research Institute, HBKU, Doha Qatar, Email: jjansen@acm.org
Joni O. Salminen, Soon-Gyo Jung, Qatar Computing Research Institute, HBKU, Doha Qatar

1 Credit: See https://www.ibm.com/garage/method/practices/think/practice_personas/.

2 Note – For brevity, we generally use the term “user” in the manuscript but imply also customer or audience.



Figure 1. Example of a traditional flat-file persona profile.

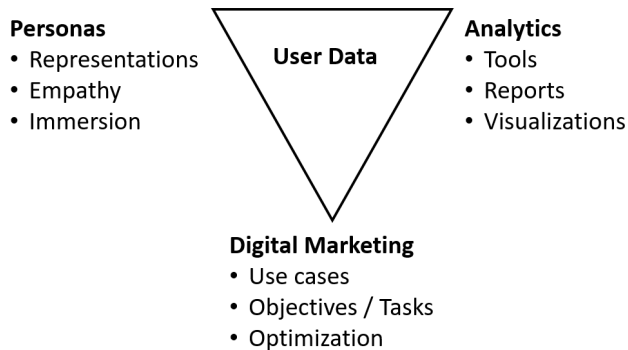


Figure 2. Conceptual positioning of the persona – analytics integration research.

However, the advent of online web, social media (Khan, Si, & Khan, 2019), and system analytics (Baig, Shuib, & Yadegaridehkordi, 2019) platforms allows for the possibility of data-driven personas that are created algorithmically (An, Kwak, Jung, Salminen, & Jansen, 2018) and enable interaction between decision-makers and the persona information. Compared to a flat file, these personas provide an interface to an integrated full-stack persona analytics system, along the conceptual lines of task integration (Byström & Kumpulainen, 2020) that contains individual user data to the persona profile (i.e., from backend user data to frontend user

conceptualization). Rather than a flat file, for these systems, the persona profile serves as a system interface, serving as both a data structure and equip the stakeholders with data interactivity. However, despite the availability of online data, there have been few data-driven persona systems to date.

Thus, in this research, we conceptually explore the relationship between personas (as *tools to understand people*) and analytics (as *systems to accomplish tasks*), using digital marketing use cases to exemplify the strengths of data-driven personas amidst analytics use cases by examining specific use cases pertaining to digital marketing. This positioning is reflected in Figure 2.

While there have been attempts to “modernize” personas by automating their creation and tying the concept to behavioral online analytics data (Hammou, Lahcen, & Mouline, 2020), the question remains: *can data-driven personas offer efficiency and/or effectiveness value relative to other analytics approaches and data representations for user understanding tasks?* This question remains largely unaddressed in the previous research, even for personas in general, and it serves as the motivation for the research presented here. Overall, there is surprisingly little research in the field of marketing, HCI, or information science about data granularity in user segmentation. Yet, this is an issue that is practically important and theoretically interesting. Notably, customer

Table 1

Attributes (Including Both Advantages and Disadvantages) of Manually and Automatically Generated Personas (Salminen, Vahlo, Koponen, Jung, Chowdhury, & Jansen, 2020)

Manual personas	Automatic personas
Give a “face” to user data	Give a “face” to user data
Low sample size (“small data”)	High sample size (“big data”)
Qualitative data	Quantitative data
Slow to create (typically taking months)	Fast to create (typically taking days)
Unresponsive to changes in user preferences	Responsive to changes in user preferences
Expensive	Affordable
Nuanced in terms of user needs, goals, and desires	Explicit in terms of user behaviors that can be algorithmically identified

segmentation and granularity are discussed by Claycamp and Massy (1968) who found aggregation beneficial for marketing practice. The optimal segmentation of user groups has also been pursued in Computer Science without a definitive solution (Jiang & Tuzhilin, 2009). While not discounting traditionally generated personas, our position is that data-driven persona has certain inherent advantages, as shown in Table 1.

Overall, the topic needs a refresher in the academic literature that we aim to provide with this research. Therefore, our research objectives are as follows:

- *Clarify how are personas positioned amidst online analytics tools*
- *Illustrate how personas lend themselves to various digital marketing use cases*
- *Demonstrate how personas might access online data to provide value for business decision-makers*

To address these questions, we frame the discussion toward automatic persona generation (APG) (Jung, An, Kwak, Ahmad, Nielsen, & Jansen, 2017), a robust full-stack persona system to illustrate the functional capabilities of a data-driven persona system at three levels of granularity: data (i.e., the user), analytics (i.e., percentages, probabilities, and weights of segments), and conceptualization (i.e., the personas profiles). Data-driven and algorithmic-based persona generation systems can drastically improve the process of persona creation, how to use persona profiles, and the conceptualization of persona profiles as design tools, along with providing enhancements to the user of analytics. The persona creation and generation process are now shifting from manual creation to data-driven generation. The result is that the created persona profile is linked to the underlying user data, ranging from the aggregated to the individual

level. As such, the persona concept is changed from a flat data structure document to an interactive interface for the underlying analytics system. This new framework offers a merging of personas and analytics; rather than competition, these two are now viewed as complements. Yet, there are still contexts for traditionally developed personas and for the employment of analytics in a pure form.

We begin with a short literature review for the positioning of data-driven personas within the broader context of personas in general and analytics. We then discuss a sampling of digital marketing use cases to show the need for data-driven personas. Then, we present a data-driven persona system that integrates personas with analytics. We end with the discussion and implications of the role of data-driven personas for digital marketing.

2 Positioning Based on Related Literature

2.1 Relationship between Personas and User Segmentation

The concept of personas is employed within many fields, industries, and domains (Cooper, 2004; Nielsen & Storgaard Hansen, 2014; Pruitt & Adlin, 2006), such as advertising (Clarke, 2015), content creation (Nielsen, Jung, An, Salminen, Kwak, & Jansen, 2017), marketing (Revella, 2015), and software design (Marshall et al., 2015), among an abundance of others. Most user-facing organizations have personas at some level of planning and usage. There are many claimed benefits of personas (Drego, Dorsey, Burns, & Catino, 2010; Eriksson, Artman, & Swartling,

2013; Friess, 2012; Miaskiewicz & Kozar, 2011; Miaskiewicz & Luxmoore, 2017; Rönkkö, 2005), such as facilitating communication among team and group members (Pruitt & Adlin, 2006), keeping the design focus on user (Cooper, 2004), helping to avoid biases in situations where the preferences of decisions makers may deviate from those of the customer, and perhaps most importantly providing an empathetic concept (Marsden, Pröbster, Haque, & Hermann, 2017; Miaskiewicz & Kozar, 2011; Norman, 2004) that most people can relate to (i.e., another person). Salminen et al. (2020a) demonstrate that, for a user identification task, the persona has significant advantages relative to analytics, as measured by metrics such as steps and time to task completion. An et al. (2018a) demonstrate some of the advantages of data-driven persona, including representativeness and time to creation.

Practically, personas reduce a large number of user or customer segments into relatively few typical users or customers, thereby reducing (Friess, 2012) the end user's cognitive load (Sweller, 1988). Personas "humanize" data by replacing numbers that most individuals struggle to memorize with people-based data representations that individuals naturally empathize with (Friess, 2012). A well-crafted persona is a digestible and useful representation of the underlying user group. It is a rounded, believable characterization of a user type that exists in the real world (Thoma & Williams, 2009). These empathetic and communicative benefits are the core rationale behind deploying personas for tasks requiring user understanding.

Simply put, this refers to the decision-making in the organization being driven by external customer feedback, rather than internal beliefs and assumptions. However, implementing market orientation in practice has proven to be challenging, especially so that market-oriented practices permeate the whole organization (Han, Kim, & Srivastava, 1998). According to our view, personas can help by providing operationalization of the market orientation concept. It is easy to see personas could be of help here, as they provide a customer-oriented frame of reference that requires little analytical sophistication from decision-makers and can, therefore, be applied at all levels of the organization.

In organizations, personas are applied at different levels, for example, by customer service at the operative customer interface, designers and developers in product lifecycle planning, and by executives at the strategic corporate level who desire to craft strategic agendas that are aligned with underlying market trends (Kohli & Jaworski 1990). This enables the decision-makers to understand experiences and backgrounds different from

their own, supporting the realization that the preferences of the users most likely differ from those working inside the organizational bubble (Miaskiewicz & Kozar, 2011). Particularly, cross-disciplinary teams may benefit, as personas may represent a unifying compass for goal alignment (Miaskiewicz & Kozar, 2011).

2.2 Personas and Analytics: Enemies or Complements?

With online analytics tools, marketers have access to individual-level data that perhaps better capture the complex and nuanced nature of audience characteristics than personas that are based on aggregated user archetypes (Rönkkö, 2005). This problem is especially pertinent in the field of digital marketing that is shifting toward a higher degree of personalization, one-to-one recommendations, and prevalent user data from a variety of channels (Salminen, Jansen, An, Kwak, & Jung, 2018) (Figure 3 for illustration). However, personalization is not the same as understanding the user. For example, a recommendation algorithm may personalize product suggestions for a given customer in such an effective manner that the user generally accepts the recommendation. However, this does not necessarily imply that the user is happy or pleased with the overall customer service.

The machine finds the higher performance subgroups from the target audience. Prior information, such as the marketer's experience or use of personas can also be used to narrow down the search space.

Conversely, since their initial inception, personas have typically been flat structures, either in the form of actual paper or in the form of an electronic PDF-like documents, and they served as mainly data structures that presented present assorted bits of data into usable information for decision-makers or implementors. Partly due to this flat data structure, personas have come under substantial criticism for being of limited practical value within the actual design process (Chapman & Milham, 2006), with decision-makers questioning whether or not the underlying data are truly reflective of the user populations (Chapman, Love, Milham, ElRif, & Alford, 2008; Tarka, 2019). Also, the traditional persona creation process has typically relied on data collection methods, such as surveys or focus groups (Nielsen, 2019; Pruitt & Grudin, 2003), when using customer data has even been employed at all in the persona creation process. As the number of participants in focus groups is usually small and noted issues with survey data (Salminen et al., 2018), this has led to the employment of mainly qualitative data

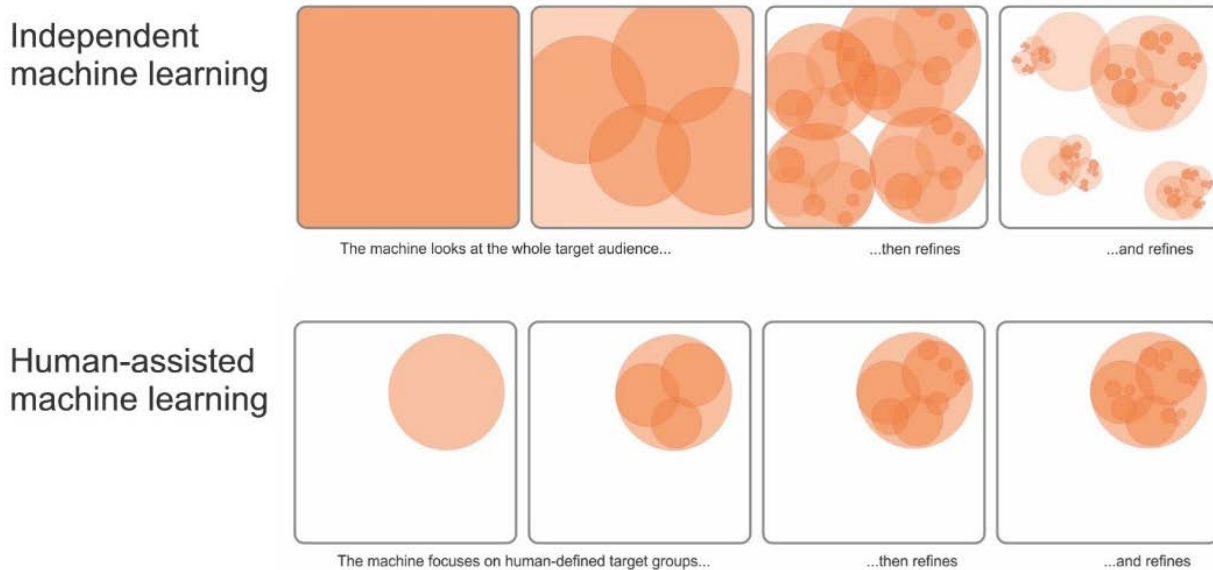


Figure 3. Algorithmic ad targeting and optimization (adapted from Smith, 2015).

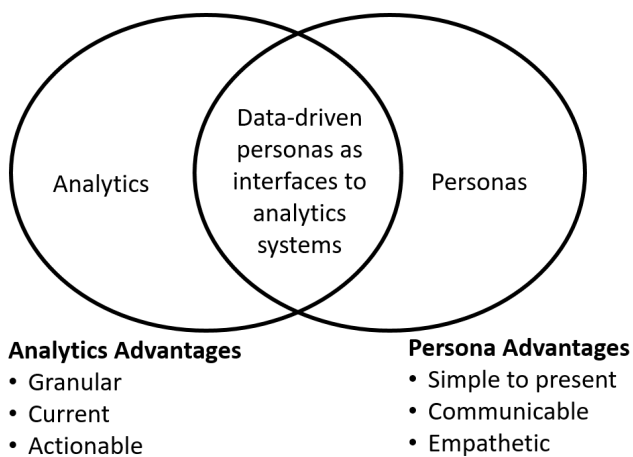


Figure 4. Attributes of manual persona generation and automatic persona generation.

analysis approaches that lead end users to question if the underlying persona data is representative, updated, or valid of the actual user population. However, the advent of data-driven personas in the age of easily available analytics data has led to a recognition of the possible integration of the personas (to give a human face) and analytics (to provide actionable numbers and precision).

Remarkably, with the combining of personas and analytics, the strengths of each approach help offset the deficiencies of the other. Conceptually, personas are easy for people to understand and generate empathy for the user, but personas are perceived as not granular and not actionable. Analytics data can be granular and actionable,

but analytics can be cumbersome for employment and difficult for end users to comprehend. Nonetheless, the combination of both leverages the strengths and limits the weakness of each, as shown in (Salminen, Guan, Nielsen, Jung, Chowdhury, & Jansen, 2020). Personas that are automatically generated from analytics data have all the strength of standard personas, and, when serving as the interface to analytics systems, they provide all of the strengths of user analytics, as illustrated in Figure 4.

Since personas were first introduced, a range of other online analytics platforms, services, and tools have emerged (Cooper, 2004; Jung, Salminen, An, Kwak, & Jansen, 2018; Springer & Whittaker, 2019) - including Adobe Analytics, IBM Analytics, Facebook Insights, Google Analytics—as well as integrating services (e.g., HootSuite, Tableau, etc.) that organizations can use to understand their users and user segments. These tools provide organizations with access to both individual users and aggregated users’ big data, resulting in further questioning about the value of using traditional qualitative personas for user insights.

As such, there is a critical need for the updating of the persona creation concept and the use of analytics through the employment and use of data-driven personas that are created algorithmically using large quantities of user data and then presenting these personas as interfaces to fully functional interactive analytics systems for user understanding.

Table 2
Digital Marketing Use Cases

	Use case	Definition
Measurement and optimization	Channel selection	Choosing the channels or platforms in which the organization interacts with potential and existing customers
	Targeting	Addressing particular users or user groups within the channel or platform
	Identifying needs	Using sources such as social media and online discussion forums to detect visible and latent customer needs
	Message creation	Creation of messages for advertising or organic marketing communications
	Budget allocation	Dividing the available digital marketing budget between and within channels
	Dynamic pricing	Setting the price to the optimal level in order to maximize revenue
	Recommendations	Showing the visitors content or products that is most likely interesting to them
	Personalization	Manipulating the website appearances according to known or estimated user attributes

2.3 Conceptual Analysis of Personas for Digital Marketing

As an illustration of the employment of data-driven personas, we present their employment through a use case approach in the domain of digital marketing.

Digital marketing embodies several techniques to get a response from their target customers. Drawing from prior studies on digital marketing, we define digital marketing use cases. A “use case” is defined as an activity that the marketer has control over, which is important in achieving marketing goals, and that covers the full spectrum of the first touch to purchase by the consumers (Lemon & Verhoef, 2016). Particularly, (Constantinides, 2004) (p. 113) refers to “Web experience” that “embraces elements like searching, browsing, finding, selecting, comparing and evaluating information as well as interacting and transacting with the online firm.” As such, it is essential to define the use cases of digital marketing in relation to online user behavior. Table 2 depicts digital marketing use cases found from the literature, with a more complete systemic review presented in (Salminen et al., 2018).

Briefly addressing the use cases, channel selection is central in digital marketing, since the audience attributes, competition, prices, and performance depend at least partially on the channel used (Ansari, Mela, & Neslin, 2008). Knowing customer behavior in different channels is crucial for this activity (Jansen & Schuster, 2011), as the channels need to be frequented by target customers. Variables, such as enjoyment, marketing efforts, and age, influence users’ attitudes toward a channel (Srisuwan & Barnes, 2008). Targeting is often done by demographic information (Doherty & Ellis-Chadwick, 2003). For

example, gender differences can be impactful on online shopping behavior (Wolin & Korgaonkar, 2003). However, targeting can also be based on consumer intent; such is the case with keyword advertising (Jansen & Schuster, 2011). More recently, psychographic factors have become more relevant in targeting as measurement and understanding have increased (Leong Jaafar, & Sulaiman, 2017). Overall, marketers are also encouraged to create an effective mix of online and offline communication to target consumers on different channels (Srisuwan & Barnes, 2008). The theoretical foundation of these use cases is the focus of the 4Cs (i.e., clarity, credibility, consistency, and competitiveness) of marketing communications (Jobber & Fahy, 2009).

Moreover, marketers need to decide how to split the available budget across the chosen channels. Flexibility is a well-known advantage of online media (Kiani, 1998). For example, the budget can be allocated between display, search engine, and social media advertising or across different platforms of the same channel category (Yang, Zeng, Yang, & Zhang, 2015). Budget allocation takes place between channels and within channels, for example, by setting keyword bids (Jansen & Schuster, 2011) and dividing budget across campaigns and ad sets. Using methods, such as keyword analysis (Jansen, Spink, & Saracevic, 2000) and netnography (Kozinets, 2002), marketers aim at identifying the needs of potential customers. Message creation, in turn, aims at persuading the target audience to visit the organization and take action (Li & Kannan, 2014). Messages include the meaning communicated to the target audience, format such as images and text ads, and context. As noted by Kiani (1998, p. 192), “The advertising objective is to say the right

things to the right people and have them perceive what is said.” This statement accurately portrays the relationship between targeting and message creation.

Once visitors are attracted to the target destination (e.g., website and mobile app), other tactics follow to achieve conversion, the final outcome of online interaction (Pitt, Berthon & Watson, 1996). According to (Constantinides, 2004), the decision-making processes of online users can be influenced by the creation and delivery of Web experiences, comprising information content, perceptual cues and stimuli, and offerings. The methods of dynamic pricing are used to achieve the optimal pricing strategy; for example, airlines may set their prices based on customer profiling (Elmaghraby & Keskinocak, 2003). Product recommendations and personalization are applied to improve conversion rates (Mahajan & Venkatesh, 2000; Ricotta & Costabile, 2007). Recommendation engines are used to suggest content or products to individual users: these are traditionally separated into collaborative filtering or content-enabled recommendations (Balabanović & Shoham, 1997). In personalization, the website is morphed according to known or estimated user attributes (Hauser, Urban, Liberali, & Braun, 2009). Srisuwan and Barnes (2008) highlight the need to know consumer preferences after the click, based on which the firm can prioritize activities that increase consumer enjoyment.

3 Data-driven Personas: Combining Analytics and Human Empathy

3.1 Overview of APG

Leveraging these digital marketing use cases, we present research applying the concept of “persona as interface” as the developmental foundation of APG, which is a system for the algorithmic creation of data-driven personas from a variety of online user data, such as Web, social media, or in-house customer data. In driving the advancement of persona conceptualization, development, and use, APG presents a multilayered (i.e., “full-stack”) data integration concerning customers, audiences, or users. We demonstrate three levels of data access afforded by APG, which are (a) conceptual persona-level, (b) analytics-level, and (c) foundational customer/audience/user level. Personas-as-interfaces to online data systems, such as APG, address the common concerns voiced about traditionally created personas, most notably issues of data validity and practical value.

Related works (An et al., 2018; An, Kwak, Salminen, Jung, & Jansen, 2018; An, Kwak, & Jansen, 2017; Kwak, An, & Jansen, 2017) have presented a working methodology and system for APG, which utilizes social media data that the system accesses through an application programming interface (API). APG then performs non-negative matrix computations (An et al., 2017) and outputs a set of 5–15 personas for the end user. The technical process of automatically generated personas has explained in detail in related work (An et al., 2017; Jung et al., 2018), and we, therefore, refrain from repeating it here, instead of focusing on the applicability of these data-driven personas for use cases such as those presented earlier. The APG personas are rendered by using Flask, an open-source web framework, and they can be accessed through a web browser by end users in the client organizations. However, what prior work has not specifically addressed is the direct application of the data-driven persona to specific market use cases and explicitly articulating how data-driven personas integrate into this marketing merger of empathic personas and rational analytics. We begin addressing this gap by presenting the main benefits of APG in Table 3.

Since APG obtains data directly from online platforms through APIs, this update process is completely automated. The fact we use these APIs also ensures that the privacy of individual users is kept intact. We retrieve bucketed information on user groups (e.g., Female 18–24, Brazil) and performance metrics (e.g., “Views” and “Clicks”) that we computationally transform into persona representations. For a more detailed explanation, refer to An, Kwak, and Jansen (2016). Since the system is built upon a Web framework, rapid manipulation, and experimentation with different persona layouts become possible.

3.2 Methodology of APG

We provide a concise overview of the APG system workflow that describes how user data for the persona profiles is retrieved, the manner the data are processed, and the process of creating the persona profiles. In general, APG produces personas from quantitative user data employing the following steps:

- **Step 1:** Generate an interaction matrix (V) with items as columns (c), demographic user groups as rows (g), and the interaction amount of each group for each item as matrix elements
- **Step 2:** Use non-negative matrix factorization (NMF) (Lee & Seung, 1999) to the interaction matrix to

Table 3
Main Benefits of Automatically Generated Personas

Advantage	Description
Speed	Personas can be created from social media data within hours. Traditional persona creation typically takes months
Accuracy	Personas are based on latent behavioral patterns of the users. Traditional persona creation typically omits the measurement of behavioral patterns
Freshness	Personas are updated at a set interval (currently, each month). Traditionally created personas are not frequently updated due to time and cost
Manipulability	The persona profile template is easily modifiable enabling rapid experimentation. Traditionally generated personas are static and do not allow for easy experimentation
Inactivity	The persona profile serves as an interface to the underlying analytics information and user data. The personas are no longer a flat file serving as solely a data structure

distinguish p latent item interaction behaviors (where p is a predetermined hyper-parameter signifying the number of personas).

- **Step 3:** Select the equivalent representative demographic attributes for each behavior by using weights from the NMF calculation.
- **Step 4:** Augment p personas by adding the corresponding representative demographic groups with other information, such as name, picture, topics of interest, and so on.

After attaining a grouped interaction matrix, the system applies NMF for classifying latent user interaction (Figure 5). We first endeavored to identify segments through clustering, but this clustering approach did not work, as the social analytics data are aggregated. We, therefore, needed a method for de-aggregating the analytics. NMF is chiefly intended for decreasing the dimensionality of large datasets by discriminating latent factors (Lee & Seung, 1999). From this initial “persona,” the APG system applies a process of improvement by adding a correct name, social media quotes, picture, and associated demographic characteristics (e.g., educational level, marital status, occupation, etc.) through querying the Facebook Marketing API. The result is a set of structural persona profiles representing the user population segments. The personas are presented to end users (i.e., the ones from the organization using the personas) through an online system operating on Flask,³ an open-source Python web framework. APG is a fully functional real system that is deployed with actual client organizations, with a demo available online.⁴ A more detailed and full technical explanation of the infrastructure

³ <http://flask.pocoo.org>.

⁴ <https://persona.qcri.org>.

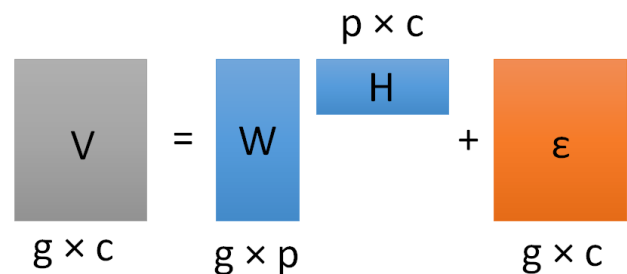


Figure 5. Matrix decomposition by means of NMF. Matrix V is decomposed into W and H . g denotes demographic groups in the dataset; c denotes product units, and p is the number of latent behaviors of demographic groups over product units and e is the error term.

of APG’s system is presented in (Jung et al., 2018; Jung, Salminen, Kwak, An, & Jansen, 2018). A more detailed narrative of the APG’s employed algorithms and methods is discussed in An et al. (2018).

We now present the employment of APG relative to our research objective of demonstrating how personas access online data to provide value for business decision-makers, with a focus on levels of granularity.

3.3 Conceptual Understanding of APG

Data-driven persona systems, similar to APG, provide the opportunity for a full-stack data methodological approach discussed earlier, with an integration of data from the customer-level through conceptual at the persona-level (see Figure 6).

As illustrated in Figure 6, the data-driven persona system employs the base level user-data that the system algorithms act upon, converting this user data into useful

Level	Classification	Artifacts
Conceptual (personas)	Abstracted	Persona as interface
Analytics (measures)	Processed	Metrics and ratios
User (numbers)	Data	Raw values and figures

Figure 6. An illustration of full-stack data integration within a persona-analytics system, from the data-level to the analytics-level to the conceptual-level with the personas.

information. The product of this algorithmic processing is actionable measures and metrics concerning the user population (i.e., weights, probabilities, percentages, etc.) of the kind that one would usually encounter in industry-standard analytics platforms. The subsequent level of abstraction engaged by the system is to use this analytics data, with the corresponding meta-tagged content, such as names and photos, to create sets of personas profiles at the conceptual level. This results in a data-driven persona system adept at presenting user insights at diverse levels of granularity, with levels that are both appropriately integrated to the task. We now discuss these three levels.

3.3.1 Conceptual Level (Personas)

The uppermost level of abstraction is the set of personas generated by the system from the user data, with a default setting of ten personas for APG, although the system user can adjust this number to generate fewer or more personas (Figure 7). The persona profile (Figure 7) has most of the standard attributes that one finds in the traditional flat persona profiles (e.g., name, photo, etc.); however, the data-driven persona is derived from real online user data, which addresses many of the major criticisms of flat personas created manually that they

- a) are not based on quantitative user data (Pruitt & Adlin, 2006)—for data-driven personas, the data is immediately underlying the personas and can be viewed by the end user,
- b) take too long to create for fast-paced industries such as online content creation (Drego et al., 2010; Rönkkö, 2005)—APG can produce a full, rich set of personas

- for an organization within a matter of hours upon accessing the data, and
- c) quickly become outdated (Jung, Salminen, & Jansen, 2019)—APG updates the personas automatically at regular intervals determined by the user organization.

In addition, data-driven personas, by relying on systematic data collection intervals, can augment the tradition persona profile with supplementary attributes such as (a) customer loyalty, (b) sentiment analysis (Tahara, Ikeda, & Hoashi, 2019), and (c) topics of interest (see Figure 7). Also, the data-driven persona profile is interactive, as it functions as an interface to the other level of data that are discussed in the following subsections.

The APG persona profiles contain the usual persona profile attributes, along with direct access provided to the underlying user data.

3.3.2 Analytics Level (Probabilities, Percentages, and Weights)

The APG persona profiles act as interactive interfaces to the underlying information leveraged to create the persona profiles. The information may somewhat vary among systems, but the analytics level reflects the specific measures and metrics generated using the foundational user data used to create the personas. In the APG system, the interactive personas profiles provide the affordances to the numerous analytics information through clickable icons in the persona interface. We discuss three examples.

As shown in Figure 8, the APG system can display the entire user population percentage that a specific persona represents (this percentage is calculated from the Facebook API audience manager database). This particular analytic insight is valuable for organization decision-makers to decide the importance of developing or designing for a given persona, and it also assists in addressing the issue of whether or not the persona is valid (i.e., does the persona represent actual people).

APG gathers overall demographic information from the online analytics platform from which it is collecting the user data, which is typically gender, age grouping, and nationality of a user segment. Then, using these three attributes and the associated interests that are derived from the results of non-negative matrix factorization (NMF), APG leverages the Facebook API (with access to data from billions of users) to determine the probability of other demographic characteristics. The persona profiles provide access to these probabilities, which one example is shown in Figure 9.

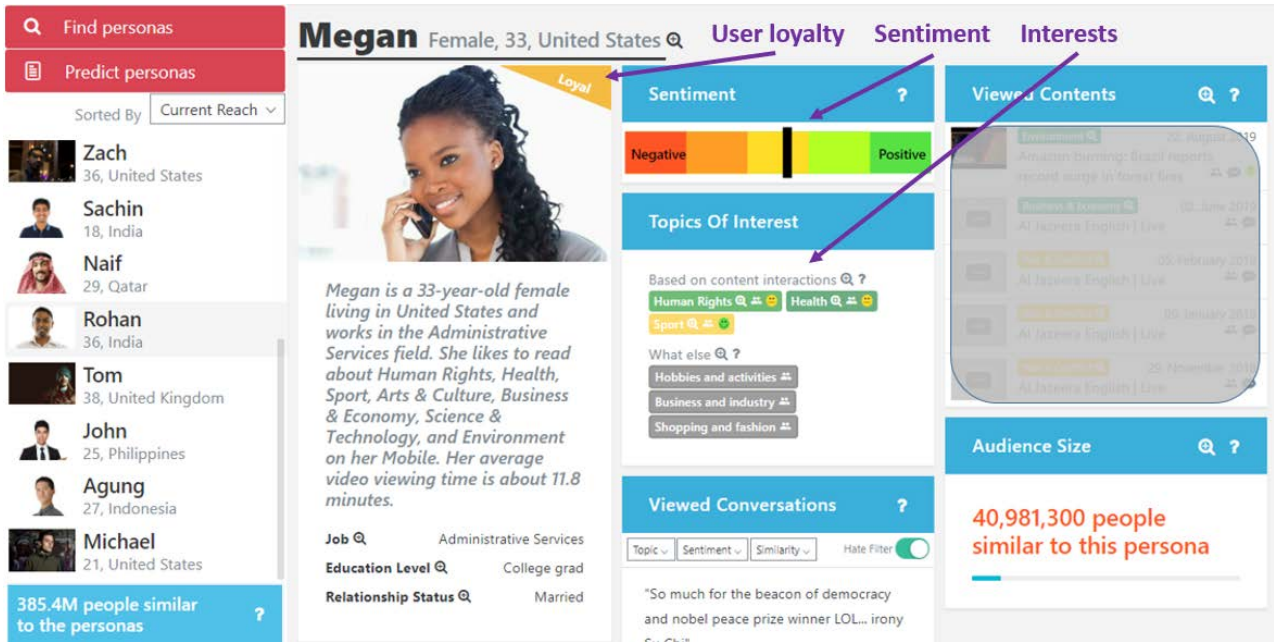


Figure 7. Example of the listing of APG personas (screen left) and a displayed data-driven persona profile.

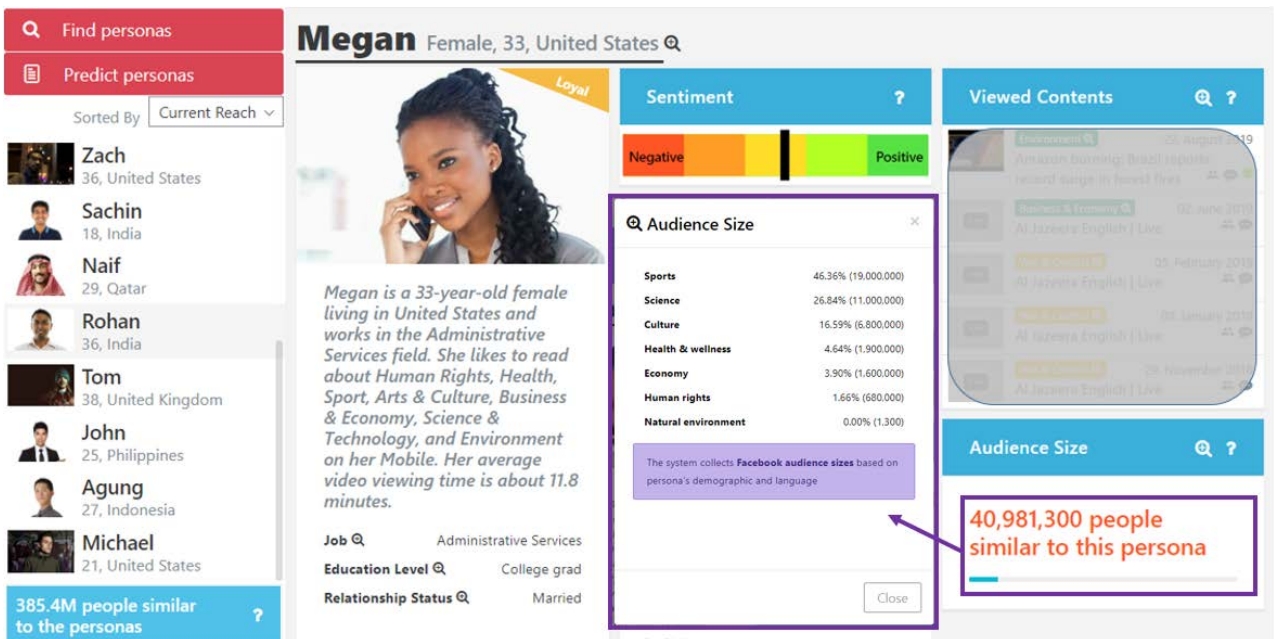


Figure 8. Screenshot of the percentage of the overall population epitomized by the persona.

APG identifies the unique behavioral patterns, and it then associates these unique patterns, using latent factors, to one or more of the demographic groups. It assigns a weight to each of these demographic groups associated with a behavioral pattern based on the strength of the association to the pattern. These demographic weights provide many insights to the end user of the

personas concerning both (a) the strength of behavior to this demographic and (b) the association of this behavior pattern with other demographic groups. Hence, APG provides access to these demographic group weights (Figure 10), which helps address concerns of what actual users that the personas represent [6].

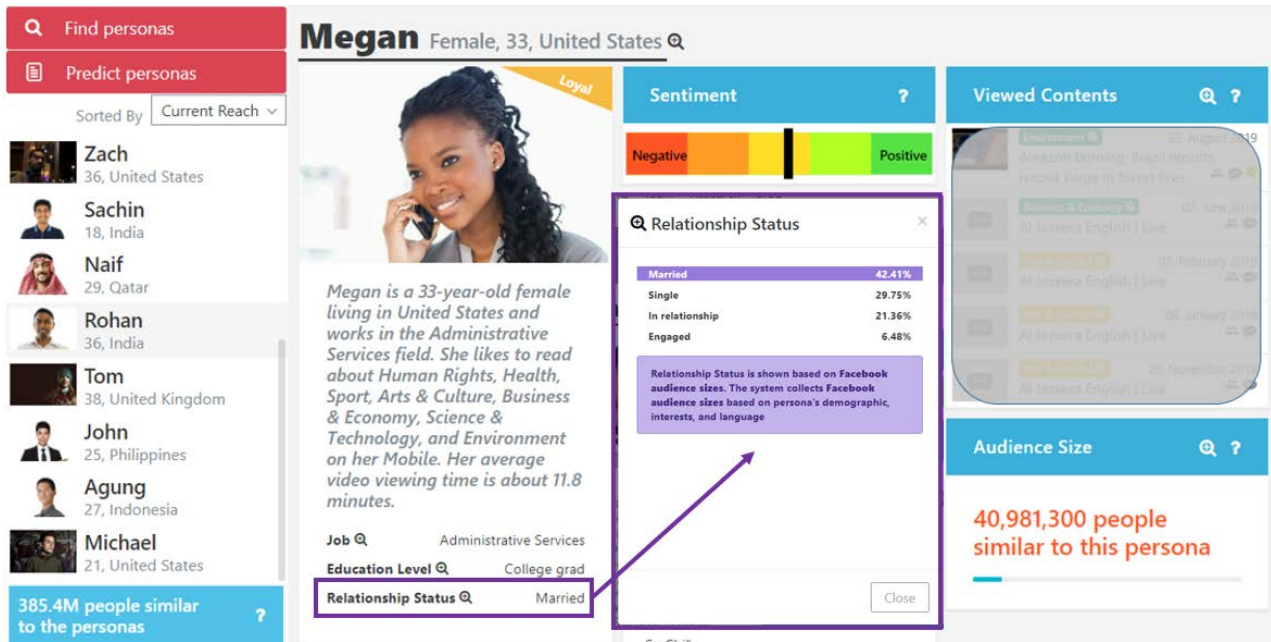


Figure 9. Probabilities of persona demographic characteristics (probability of marital status in this example).

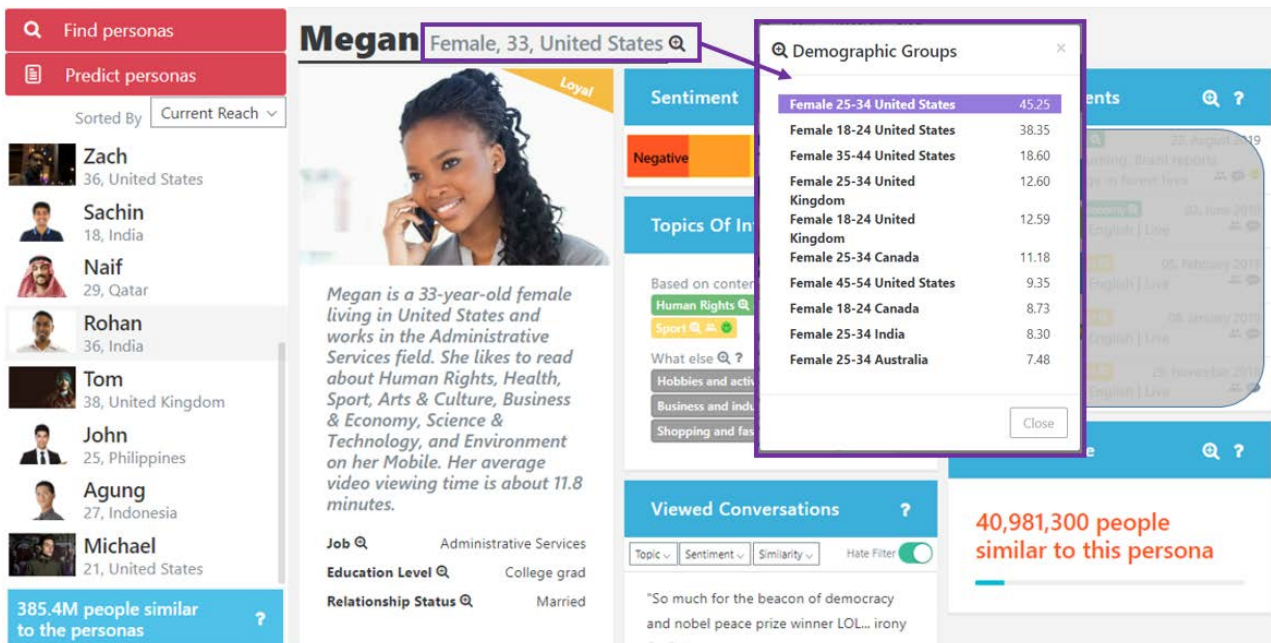


Figure 10. The weights generated by NMF for various demographic segments associated with one unique behavior pattern that used to generate the persona.

3.3.3 Customer Level (Individual Data)

By accessing the demographic meta-data output from the NMF process, the decision-maker can access the specific user level (i.e., individual or aggregate), as shown in Figure 11.

3.4 Evaluation of Personas in Digital Marketing Use Cases

Having defined the digital marketing use cases and presented APG as a system for the generation of data-driven personas, we next evaluate their compatibility

Video	Country	Count	F13-17	F18-24	F25-34	F35-44	F45-54	F55-64	F65	M13-17	M18-24
_mZfHh4_Q	US	2505	0.40	3.70	5.40	2.20	1.20	1.60	2.80	1.20	16.70
_mZfHh4_Q	ZA	697		2.30	7.20	5.00	1.80	0.90	1.80	1.80	17.30
_mZfHh4_Q	GB	398	0.60	2.80	2.80	1.10		0.60		7.20	28.30
_mZfHh4_Q	CA	337	0.70	3.40	5.40	1.40	0.70		1.40		25.00
_mZfHh4_Q	AU	126	1.40		1.40					4.30	37.10
_mZfHh4_Q	IN	126		8.60		1.70				1.70	48.30
_mZfHh4_Q	NL	113		2.60		5.10				2.60	15.40
_mZfHh4_Q	DE	97									
_mZfHh4_Q	SE	83		8.00		8.00		4.00		4.00	32.00
_mZfHh4_Q	QA	57	4.80	9.50	9.50	4.80					19.00
_mZfHh4_Q	IL	56		6.70						3.30	43.30
_mZfHh4_Q	MY	53	3.80	3.80	3.80						38.50
_mZfHh4_Q	SA	48									
_mZfHh4_Q	KE	47		9.10	4.50						27.30
_mZfHh4_Q	SG	43		3.70						7.40	14.80
_mZfHh4_Q	AE	39									
_mZfHh4_Q	IT	33									

Figure 11. The foundational user-level data from which the data-driven personas and analytics are generated. It is this numerical user data that be available in various formats, which are the foundational aspects of the data-driven persona presented in Figures 7–10.

with different levels of data aggregation. We first explain the notion of data granularity, which is crucial for understanding the conceptual position presented here. There are various levels of data aggregation granularity, from the individual customer (i.e., 1-to-1 marketing) to the entire customer base (or overall market beyond it). Segments are located between the customer base and the individual customers, summarizing information about particular sub-groups of customers (Claycamp & Massy, 1968). Individual data, therefore, refers to individual customers (e.g., “Peter Jenkinson, 45-year-old male, Portland, Maine”), aggregated data to a group (“Male, 45–55 males, Portland, Maine”). Personas are in between these two extremes, based on aggregated data but presented with individual attributes. In Table 4, we examine the digital marketing activities as potential use cases for personas to examine how online analytics data changes the applicability of personas from the decision-maker’s point of view.

In ad targeting, individual-level information is typically preferred (Blattberg & Deighton, 1991). A low level of aggregation is desirable because match-making efficiency increases with the available information. If marketers know more about individual preferences, they can match the right products with the right consumers more efficiently. As noted by Kiani (1998, p. 192), “[digital marketing] represents the opportunity to customize the interaction and tailor either the product or the marketing effort to one consumer at a time.” Mass advertising that shows one message to all segments ignores the variation of preferences and tastes in the market.

In contrast, segment-based targeting is superior because it aligns the product features with the market segments. The specificity of the target group enables marketers to tailor messages, which results in less wasted ad impressions, as the group’s marketers think are not interested become excluded. Likewise, recommendation mechanisms are optimal when automated and using individual-level data. Note that individual can mean estimates based on conjoint variables (probabilistic methods) instead of using accurate data. However, each visitor is given unique recommendations based on their class. Such approaches are also taken by newsfeed algorithms, such as Facebook’s content ranking algorithm.

Finally, for strategic planning purposes, such as channel selection, the complexity of individual-level data is excessive. While machine decision-making can handle targeting and recommendation decisions, automating strategic decision-making requires a holistic sense of understanding currently lacking from machine models. Hence, aggregated data representations seem more suitable for strategic decisions, whereas individual data are most optimal in automated decision-making. Whether the personas or numbers-based representations are “better” depends on several factors, such as the decision-making style of the individual (Thunholm, 2004), job role (Salminen, Liu, Sengun, Santos, Jung, & Jansen, 2020), and type of task (Salminen, Jung, Chowdhury, Sengün, & Jansen, 2020). However, the reported empathy benefits of personas should be widely accessible to stakeholders.

Table 4
The Suitability of Data Aggregation for Digital Marketing Use Cases.

Use case	Individual data	Aggregated data	Example of Data-Driven Implementation in APG
Channel selection		X	Personas in APG are presented across multiple online platforms
Targeting	x		Individual user data is available in APG and traceable from a given persona
Identifying needs		x	APG automatically generates interest of the personas
Message creation	x		APG offers functionality to test whether or not a message will resonate with a set of personas
Budget allocation		x	As of this date, functionality not implemented in APG
Dynamic pricing	x		As of this date, functionality not implemented in APG
Recommendations	x		APG offers analysis features for product recommendations for personas
Personalization	x		Individual user data is available in APG and traceable from a given persona

Note: For each activity, decision-makers use some form of data. When discussing personas' usefulness, we are actually asking which form of data representation is useful in which use case.

4 Discussion

The conceptual shift of persona profiles from flat data structures to data-driven personas as interfaces of systems for understanding users opens a completely new era for the implementation of personas. Using these data-driven personas that are embedded as the interfaces to user systems, executives or decision-makers can, for example, provide specific directions for an online marketing campaign, targeting one or more personas. Then, the operational managers can leverage the intermediate analytics information within the data-driven persona systems to refine and target the specific customer segments. Online marketers can then use these refined segments to address the specific customers within these segments that will receive online marketing messages.

4.1 Shift to Personalization with Analytics

Business decision-making is shifting toward more granular use of customer data. The novel method is to apply marketing automation which predicts how individual customers behave; applications include, for example, advertising and email campaigns, product recommendations, and website personalization. The more we are moving toward real-time optimization, the less useful *a priori* conceptualizations like target groups and personas become for marketing. This proposition is logically sound since messages and offerings can be even more tailored than in group-based marketing, resulting in efficiency gains.

4.2 Shift to Data-driven Personas with the Increased Availability of Online Data

Given that these alternatives are developing at a tremendous pace, the newly established criticism toward personas must be acknowledged and properly addressed. Scholars working with personas cannot simply ignore this important criticism and isolate themselves from it, but they must understand its tenets, address them, and position the use of personas in the light of new developments in the analytics sector. Throughout their existence, personas have been questioned. Yet, they have persistently maintained their appeal, drawing new practitioners and scholars in their sphere of influence. Instead of opting for general defense and justification, we should seek to analyze personas in specific use cases and to objectively assess their strengths and weaknesses. Through such intellectual integrity, HCI and marketing scholars can keep contributing to the development of personas in both theory and practice. Of course, there are many possible cases where one can leverage both manual and automatic methods of persona creation, as discussed in (Salminen et al., 2018; Salminen et al., 2020).

4.3 Role of Data-driven Personas in the Age of Analytics

Nonetheless, our conceptual analysis implies that the claimed irrelevance of personas seems to be mainly based on the disagreement of their advantageous purpose in

Table 5
Levels of Data Aggregation.

Level of aggregation	Data representation	Example
Low	Individual	Customer profiles
Medium	Aggregated individual	Personas
High	Aggregated numbers	Segments

Note: Aggregated individual data describes personas; keeping individualized attributes while using extensive datasets.

an age of large-scale customer data. While in certain use cases of digital marketing and online analytics, such as ad targeting and recommendation systems, personas naturally perform poorly against other techniques; there are other use cases where they are superior, simply due to the need for compression of information. Therefore, the usefulness of personas is a question of determining their applicability, defined as the scope of personas. Essentially, when the needs for aggregation range from medium to high, personas become an alternative for data representation. This needed range is illustrated in Table 5.

The concept of personas mirrors the concept of customer orientation in marketing, which has for decades been argued for, specifically in the concepts of marketing orientation (Narver & Slater 1990; Kohli & Jaworski 1990), value co-creation (Vargo & Lusch, 2008) and customer-dominant logic (Stauss, Heinonen, Strandvik, Mickelsson, Edvardsson, Sundström, & Andersson, 2010). By influencing the level of motivation and empathy, personas can also influence actions taken by individuals at different organizational levels and, thus, become a vehicle for a higher degree of market orientation within the organization employing them.

4.4 Scope of Both Personas and Analytics within the Confines of Decision-making

The question then becomes: *What decisions are the individuals making?* Obviously, the answer depends on several factors, notably the industry, organizational context, and the position of the decision-maker. Here, we have based our list of decision tasks in digital marketing. Even in this context, characterized by rapid innovation and technology as a driving force, personas seem to have their role. When dealing with a representation of a person, the interaction between personas and the users of analytics systems becomes evident. The users in these situations often use terms such as ‘she is interested in...’,

‘she likes...’. Such verbalization is not usually heard when speaking of data; for numbers, the interpretation tends to be fewer personas (“they are...,” etc.). The scope forcefully becomes more general and loses the appeal of individual characteristics. The second argument, consequently, is to define the scope of personas, more needs to be known of the decision-makers. As such, there are several areas for future research, as discussed in (Jung et al., 2018; Salminen, Jung, & Jansen, 2019). Perhaps two of the most promising areas of research including (a) validating personas and data-driven personas in a variety of contexts and setting, and (b) investigations of the theoretical aspects that make personas compelling.

5 Conclusion

As a promising use of personas as the frontend and interface to interactive systems for understanding users, there are sufficient future research opportunities, including addressing the growing levels of data granularity and also conducting user studies regarding the placement of such hybrid persona-analytics systems. Nevertheless, the conceptual enhancement of and the capabilities of the personas as an interface to full-stack systems present considerable promise for data-driven persona practices within the field of HCI and also offer considerable impact for decision-making in organizations that desire to understand their users.

References

- An, J., Kwak, H., & Jansen, B. J. (2016). Validating Social Media Data for Automatic Persona Generation. *Proceedings of 13th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, 1-6.
- An, J., Kwak, H., & Jansen, B. J. (2017). Personas for Content Creators via Decomposed Aggregate Audience Statistics. In J. Diesner, E. Ferrari, & G. D. Xu (Eds.), *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 632-635). New York, NY: ACM. doi:10.1145/3110025.3110072
- An, J., Kwak, H., Jung, S. G., Salminen, J., & Jansen, B. J. (2018a). Customer segmentation using online platforms: Isolating behavioral and demographic segments for persona creation via aggregated user data. *Social Network Analysis and Mining*, 8(1), 54.
- An, J., Kwak, H., Salminen, J., Jung, S., & Jansen, B. J. (2018). Imaginary people representing real numbers: Generating personas from online social media data. *ACM Transactions on the Web (TWEB)*, 12(4), 1-26.
- Ansari, A., Mela, C. F., & Neslin, S. A. (2008). Customer channel migration. *Journal of Marketing Research*, 45(1), 60-76.

- Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2019). Big data adoption: State of the art and research challenges. *Information Processing & Management*, 56(6). doi:10.1016/j.ipm.2019.102095.
- Balabanović, M., & Shoham, Y. (1997). Fab: Content-based, collaborative recommendation. *Communications of the ACM*, 40(3), 66–72.
- Blattberg, R. C., & Deighton, J. (1991). Interactive marketing: Exploiting the age of addressability. *Sloan Management Review*, 33(1), 5–15.
- Byström, K., & Kumpulainen, S. (2020). Vertical and horizontal relationships amongst task-based information needs. *Information Processing & Management*, 57(2). doi:10.1016/j.ipm.2019.102065
- Chapman, C. N., Love, E., Milham, R. P., ElRif, P., & Alford, J. L. (2008). Quantitative evaluation of personas as information. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (Vol. 52, No. 16, pp. 1107–1111), Los Angeles, CA: SAGE Publications. doi:10.1177/154193120805201602
- Chapman, C. N., & Milham, R. P. (2006). The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (Vol. 50, No. 5, pp. 634–636), Los Angeles, CA: SAGE Publications. doi:10.1177/154193120605000503
- Clarke, M. F. (2015). The Work of Mad Men that Makes the Methods of Math Men Work: Practically Occasioned Segment Design. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3275–3284), New York, NY: ACM. doi:10.1145/2702123.2702493
- Claycamp, H. J., & Massy, W. F. (1968). A theory of market segmentation. *Journal of Marketing Research*, 5(4), 388–394.
- Constantinides, E. (2004). Influencing the online consumer's behavior: The Web experience. *Internet Research*, 14(2), 111–126.
- Cooper, A. (2004). *The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity* (2nd ed.). Indianapolis: Pearson Higher Education.
- Doherty, N. F., & Ellis-Chadwick, F. E. (2003). The relationship between retailers' targeting and e-commerce strategies: An empirical analysis. *Internet Research*, 13(3), 170–182.
- Drego, V. L., Dorsey, M., Burns, M., & Catino, S. (2010). *The ROI Of Personas* (Research Report). Retrieved from <https://www.forrester.com/report/The+ROI+Of+Personas/-/E-RES55359>
- Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science*, 49(10), 1287–1309.
- Eriksson, E., Artman, H., & Swartling, A. (2013). The secret life of a persona: when the personal becomes private. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2677–2686). New York, NY: ACM. doi:10.1145/2470654.2481370
- Friess, E. (2012). Personas and decision making in the design process: an ethnographic case study. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1209–1218). New York, NY: ACM. doi:10.1145/2207676.2208572
- Hammou, B. A., Lahcen, A. A., & Mouline, S. (2020). Towards a real-time processing framework based on improved distributed recurrent neural network variants with fastText for social big data analytics. *Information Processing & Management*, 57(1). doi:10.1016/j.ipm.2019.102122
- Han, J. K., Kim, N., & Srivastava, R. K. (1998). Market orientation and organizational performance: Is innovation a missing link? *Journal of Marketing*, 62(4), 30–45.
- Hauser, J. R., Urban, G. L., Liberali, G., & Braun, M. (2009). Website morphing. *Marketing Science*, 28(2), 202–223.
- Jansen, B. J., & Schuster, S. (2011). Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research*, 12(1), 1–18.
- Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing & Management*, 36(2), 207–227.
- Jiang, T., & Tuzhilin, A. (2009). Improving personalization solutions through optimal segmentation of customer bases. *IEEE Transactions on Knowledge and Data Engineering*, 21(3), 305–320.
- Jobber, D., & Fahy, J. (2009). *Foundations of marketing* (3rd ed.). Maidenhead: McGraw-Hill Higher Education.
- Jung, S. G., An, J., Kwak, H., Ahmad, M., Nielsen, L., & Jansen, B. J. (2017). Persona Generation from Aggregated Social Media Data. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1748–1755). New York, NY: ACM. doi:10.1145/3027063.3053120
- Jung, S. G., Salminen, J., An, J., Kwak, H., & Jansen, B. J. (2018). Automatically Conceptualizing Social Media Analytics Data via Personas. *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2018)*, 715–716.
- Jung, S. G., Salminen, J., & Jansen, B. J. (2019). Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–6). New York, NY: ACM. doi:10.1145/3290607.3312955
- Jung, S. G., Salminen, J., Kwak, H., An, J., & Jansen, B. J. (2018). Automatic Persona Generation (APG): A Rationale and Demonstration. *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (pp. 321–324). New York, NY: ACM. doi:10.1145/3176349.3176893
- Khan, F., Si, X., & Khan, K. U. (2019). Social media affordances and information sharing: An evidence from Chinese public organizations. *Data and Information Management*, 3(3), 135–154.
- Kiani, G. R. (1998). Marketing opportunities in the digital world. *Internet Research*, 8(2), 185–194.
- Kohli, A. K., & Jaworski, B. J. (1990). Market orientation: The construct, research propositions, and managerial implications. *Journal of Marketing*, 54(2), 1–18.
- Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research*, 39(1), 61–72.
- Kwak, H., An, J., & Jansen, B. J. (2017). Automatic Generation of Personas Using YouTube Social Media Data. *Proceedings of the Hawaii International Conference on System Sciences (HICSS-50)*, 833–842.
- Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788–791.

- Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69–96.
- Leong, L.-Y., Jaafar, N. I., & Sulaiman, A. (2017). Understanding impulse purchase in Facebook commerce: Does Big Five matter? *Internet Research*, 27(4), 786–818.
- Li, H., & Kannan, P. K. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, 51(1), 40–56.
- Mahajan, V., & Venkatesh, R. (2000). Marketing modeling for e-business. *International Journal of Research in Marketing*, 17(2-3), 215–225.
- Marsden, N., Pröbster, M., Haque, M. E., & Hermann, J. (2017). Cognitive styles and personas: Designing for users who are different from me. *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (pp. 452–456). New York, NY: ACM. doi:10.1145/3152771.3156156
- Marshall, R., Cook, S., Mitchell, V., Summerskill, S., Haines, V., Maguire, M., ... & Case, K. (2015). Design and evaluation: End users, user datasets and personas. *Applied Ergonomics*, 46, 311–317.
- Miaskiewicz, T., & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? *Design Studies*, 32(5), 417–430.
- Miaskiewicz, T., & Luxmoore, C. (2017). The use of data-driven personas to facilitate organizational adoption—A case study. *The Design Journal*, 20(3), 357–374.
- Narver, J. C., & Slater, S. F. (1990). The effect of market orientation on business profitability. *Journal of Marketing*, 54(4), 20–35.
- Nielsen, L. (2019). *Personas—user focused design* (2nd ed.). London: Springer.
- Nielsen, L., Jung, S.-G., An, J., Salminen, J., Kwak, H., & Jansen, B. J. (2017). Who Are Your Users?: Comparing Media Professionals' Preconception of Users to Data-driven Personas. *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (pp. 602–606). New York, NY: ACM. doi:10.1145/3152771.3156178
- Nielsen, L., & Storgaard Hansen, K. (2014). Personas is applicable: a study on the use of personas in Denmark. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1665–1674). New York, NY: ACM. doi:10.1145/2556288.2557080
- Norman, D. (2004). *Ad-Hoc Personas & Empathetic Focus* [Personal website]. Retrieved from Jnd.Org http://www.jnd.org/dn.mss/personas_empath.html
- Pitt, L. F., Berthon, P., Watson, R. T., & Ewing, M. (2001). Pricing strategy and the net. *Business Horizons*, 44(2), 45–54.
- Pruitt, J., & Adlin, T. (2006). *The Persona Lifecycle: Keeping People in Mind Throughout Product Design* (1st ed.). San Francisco, CA: Morgan Kaufmann.
- Pruitt, J., & Grudin, J. (2003). Personas: practice and theory. *Proceedings of the 2003 Conference on Designing for User Experiences* (pp. 1–15). New York, NY: ACM. doi:10.1145/997078.997089
- Revela, A. (2015). *Buyer Personas: How to Gain Insight into Your Customer's Expectations, Align Your Marketing Strategies, and Win More Business*. Hoboken, New Jersey: John Wiley & Sons.
- Ricotta, F., & Costabile, M. (2007). Customizing customization: A conceptual framework for interactive personalization. *Journal of Interactive Marketing*, 21(2), 6–25.
- Rönkkö, K. (2005). An Empirical Study Demonstrating How Different Design Constraints, Project Organization and Contexts Limited the Utility of Personas. *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (Vol. 08, pp. 220a). NW Washington, DC: IEEE Computer Society. doi:10.1109/HICSS.2005.85
- Salminen, J., Guan, K., Nielsen, L., Jung, S., Chowdhury, S. A., & Jansen, B. J. (2020, July 19). A Template for Data-Driven Personas: Analyzing 31 Quantitatively Oriented Persona Profiles. In *Proceedings of the 22nd International Conference on Human-Computer Interaction (HCI'20)*. In-press.
- Salminen, J., Jansen, B. J., An, J., Kwak, H., & Jung, S. G. (2018). Are personas done? Evaluating their usefulness in the age of digital analytics. *Persona Studies*, 4(2), 47–65.
- Salminen, J., Jung, S., Chowdhury, S. A., Sengün, S., & Jansen, B. J. (2020a, April 25). Personas and Analytics: A Comparative User Study of Efficiency and Effectiveness for a User Identification Task. *Proceedings of the ACM Conference of Human Factors in Computing Systems (CHI'20)*. In-press.
- Salminen, J., Jung, S., & Jansen, B. J. (2019). The future of data-driven personas: A marriage of online analytics numbers and human attributes. *Proceedings of the 21st International Conference on Enterprise Information Systems*, 596–603.
- Salminen, J., Liu, Y.-H., Sengun, S., Santos, J. M., Jung, S.-G., & Jansen, B. J. (2020, March 17). The Effect of Numerical and Textual Information on Visual Engagement and Perceptions of AI-Driven Persona Interfaces. *Proceedings of the ACM Intelligent User Interfaces (IUI'20)* (357–368). Cagliari, Italy: ACM.
- Salminen, J., Şengün, S., Kwak, H., Jansen, B. J., An, J., Jung, S. G., ... & Harrell, D. F. (2018). From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas. *First Monday*, 23(6).
- Salminen, J., Vahlo, J., Koponen, A., Jung, S.-G., Chowdhury, S. A., & Jansen, B. J. (2020). Designing Prototype Player Personas from a Game Preference Survey. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. In-press.
- Smith, S. (2015). *The Benefits of Machine Learning and Programmatic Buying* [Marketing]. Retrieved from https://www.slideshare.net/Intelligent_Optimisations/the-benefits-of-machine-learning-and-programmatic-buying
- Springer, A., & Whittaker, S. (2019). Progressive disclosure: Empirically motivated approaches to designing effective transparency. *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 107–120). New York, NY: ACM. doi:10.1145/3301275.3302322
- Srisuwan, P., & Barnes, S. J. (2008). Predicting online channel use for an online and print magazine: A case study. *Internet Research*, 18(3), 266–285.
- Stauss, B., Heinonen, K., Strandvik, T., Mickelsson, K.-J., Edvardsson, B., Sundström, E., & Andersson, P. (2010). A customer-dominant logic of service. *Journal of Service Management*, 21(4), 531–548.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.

- Tahara, S., Ikeda, K., & Hoashi, K. (2019). Empathic dialogue system based on emotions extracted from tweets. *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 52–56). New York, NY: ACM. doi:10.1145/3301275.3302281
- Tarka, P. (2019). Managers' cognitive capabilities and perception of market research usefulness. *Information Processing & Management*, 56(3), 541–553.
- Thoma, V., & Williams, B. (2009). Developing and Validating Personas in e-Commerce: A Heuristic Approach. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, & M. Winckler (Eds.), *Lecture Notes in Computer Science: vol 5727, IFIP Conference on Human-Computer Interaction* (pp. 524–527). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-03658-3_56
- Thunholm, P. (2004). Decision-making style: Habit, style or both? *Personality and Individual Differences*, 36(4), 931–944.
- Vargo, S. L., & Lusch, R. F. (2008). Service-dominant logic: Continuing the evolution. *Journal of the Academy of Marketing Science*, 36(1), 1–10.
- Wolin, L. D., & Korgaonkar, P. (2003). Web advertising: Gender differences in beliefs, attitudes and behavior. *Internet Research*, 13(5), 375–385.
- Yang, Y., Zeng, D., Yang, Y., & Zhang, J. (2015). Optimal budget allocation across search advertising markets. *INFORMS Journal on Computing*, 27(2), 285–300.