



Visual Business Analytics: Using the Example of a Call Center

¹Pascal-Philipp Nöllenburg, ²Arthur Dill

^{1,2}FOM University of Applied Science, Essen, Germany

Abstract: In this article we will examine an approach to the analysis of semi-structured log data using the example of a call center as a subsection of a central corporate service center. In such data all events of a caller passing through the routing are stored. Consequently, it is possible to trace more precisely what the customer experiences during his call. However, this information is only available in semi-structured form. A little-known approach in Anglo-Saxon literature, the Visual Business Analytics (VBA), represents a holistic concept for achieving added value from semi-structured data. In the VBA, data is initially transformed and structured and then prepared for analysis purposes. The goal is to derive recommendations for supporting management decisions in a call center. In the further course, the development of the approaches of information representation is examined first, then the VBA is presented and applied to the log protocols in a call center. Finally, other call center applications of VBA are considered, and an outlook is given on other industries in which the use of VBA offers advantages.

Keywords: Visual Business Analytics, Management decisions, Call center, Big data

1. Introduction

Progressive digitization has led to a pronounced change in people's consumer behavior in recent years. While consumers in the 1980s still made all their purchases locally and handled their banking and insurance affairs in person at their branches or with their insurance broker, this approach has changed fundamentally. Digital transformation is shifting end-customer business from local stores to the World Wide Web. Online shopping is steadily gaining importance over brick-and-mortar retail, and local services are increasingly being substituted by digital service offerings such as online banking (Jahn, 2017).

However, customers' need for advice has not decreased. Rather, the type of contact options has spread to new channels. The repertoire is now not limited to the classic telephone network but has been expanded to include e-mail, messenger and video services. In order to handle these different customer inquiries, it is necessary for companies to be able to serve all these channels with the highest possible quality. That is why the number of call center agents in Germany more than doubled from 220,000 to 520,000 between 2000 and 2012 (Herzog, 2017).

The increasing relevance of service centers also becomes apparent when looking at Commerzbank's press release of February 11, 2021. It announced that Commerzbank would be undergoing a transformation process in the future. This involves reducing the number of branches in Germany from the current 790 to 450 by 2024. Commerzbank intends to focus more strongly on a central service center. In addition, the bank will continue to focus intensively on expanding online banking and automation in the form of self-service processes. The aim is to reduce costs, increase the bank's profitability and respond to changing customer needs (Commerzbank, 2021).

But it is not only Commerzbank that will be significantly reducing its branch network. Companies such as Douglas, Depot, H&M, Zara, Kaufhof and Esprit also announced that they would be closing numerous stores permanently in 2021 and relying increasingly on online retailing. This circumstance will further accelerate the global digitization process. In the future, a key challenge for companies will be to analyze the consumer behavior of customers not only in local stores but also on digital platforms in particular. For this purpose, a vast amount of data is currently available to companies as Big Data. Many companies are already able to classify customers based on their customer profiles using common characteristics (Seufert, 2016). The goal of these evaluations is to successfully and profitably use the insights gained

from them to optimize their digital products and central service centers for their end-users. At the same time, further survey data is generated in order to use it for competitive advantage.

1.1 Problem Statement

For a long time, the half-hourly interval was considered the industry reporting standard for managing a call center with, for example, the following questions:

- How many callers were there between 9:00 and 9:30?
- How long did customers have to wait to reach a call center agent?
- How many callers hung up while on hold?
- How long was the call time?
- Were there enough call center agents available?
- What was the overall reachability of the call center?

These questions could be answered by standard reporting of the telephone system. A reporting database provided all these data aggregated as key figures. Additional transformations of the data were not necessary since the data preparation took place in the data memory of the telephone system. Initially, there was no need for the call centers to expand this, but as Big Data became more relevant, the service centers recognized the added value that could be derived from transactional data. For this reason, it became possible to store each call with all its parameters in a separate table. Other industries, however, were already able to analyze the complete click and purchase behavior of their customers based on the following exemplary questions: Which product did which customer click on and when? Which product was he still interested in and when did the customer decide to pre-order a new TV?

Inspired by such transaction analyses, the complete call has also been stored in the call center from a routing perspective. In a call center, routing is understood as a process in which incoming calls are automatically routed to a call center agent based on predefined parameters. To perform single-call analyses, all calls with their potential parameters are transferred to a separate table. The problem is that the data is only available in semi-structured form. To analyze more precisely what the customer experienced in the call center, it is necessary to structure and visualize the data.

1.2 Aim of the Study

The aim of the study is to present a holistic approach to the analysis of structured, semi-structured and unstructured data in companies. The application is made on the example of a call center. Consequently, it is shown which potential the findings of data analysis offer there for supporting management decisions. For this purpose, an approach developed by Kohlhammer et al. (2018), the Visual Business Analytics (VBA), is used which is largely unknown in the Anglo-Saxon literature. At the core of this approach is how data can be efficiently processed, analyzed and visualized in the company. For this reason, the following chapter will first provide an overview of the development of visualization to get an impression of how information representation has changed over time.

2. Literature Review

The visualization of data and information has become much more relevant due to the advancing digitalization of recent years. Never before has such a large amount of data been available for decision-making in a company. Complex interrelationships must be presented in a comprehensible and compact way using different visualization tools based on vast amounts of data (Jacobs and Hensel-Börner, 2020).

For this reason, it no longer seems appropriate to simply provide management with a pure data table. The graphical preparation of data enables those responsible to process the information more quickly and efficiently (Schön, 2016).

But the visual representation of information is not a phenomenon that has only been known for a few years. First visualizations existed already in the beginnings of the ancient Greek-Roman culture about 3,000 years ago. Here we can find some historical wall paintings that people used to communicate utilizing geometric shapes and colors in addition to speech and gestures. People used these to protect themselves from dangerous animals living nearby. In addition, the paintings were used as a source of information for hunting. However, these wall paintings were not only used to visually represent information but also to express certain situations and feelings in pictures (Meinel and Sack, 2004).

The first known statistical visualization is based on Michael Florent van Langren from 1644, a cosmographer and mathematician, who recorded the distance between Rome and the Spanish province of Toledo (Figure 1). The peculiarity of his visualization is that he put his own calculation of the distance in relation to twelve other known estimators within one representation. Here he chose the form of a graph and used the distance, measured in longitude, as an axis measure (Tufte, 1997).

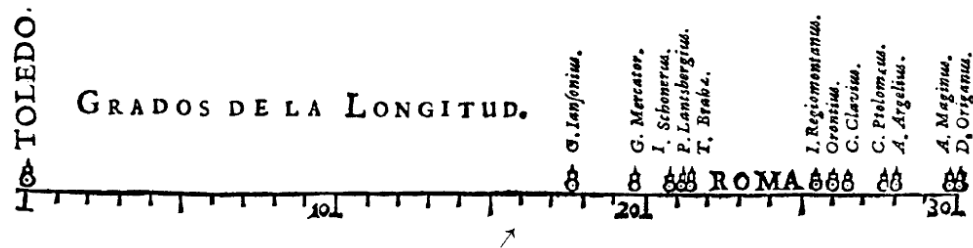


Figure 1: Visualization of the distance between Rome and Toledo (Tufte, 1997)

Almost 150 years later, William Playfair's "Commercial and Political Atlas and Statistical Breviary" appeared, offering a visualization of information that had never been seen before (Playfair, 1801).

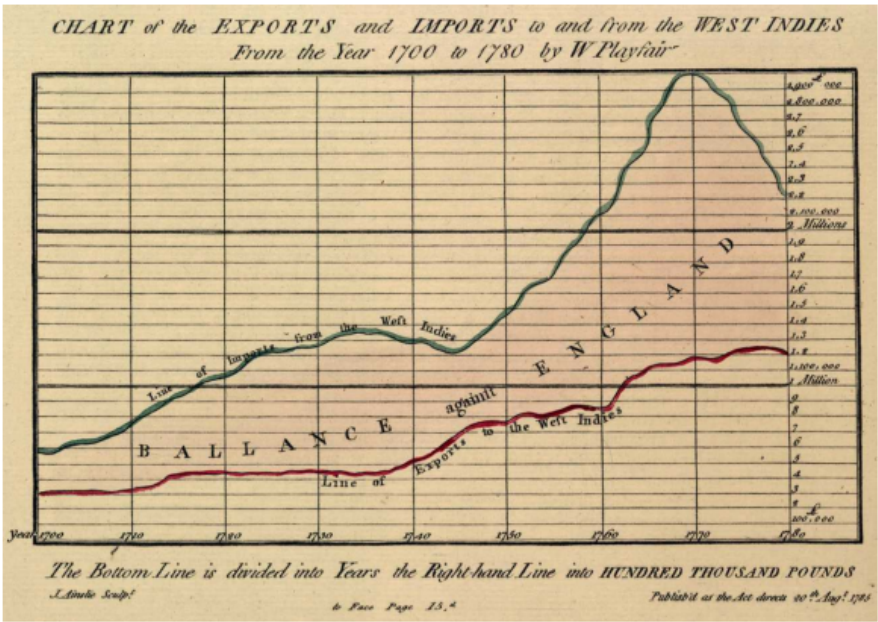


Figure 2: West Indian trade balance from 1700 to 1780 (Spence, 2006)

William Playfair was a Scottish engineer and economist. He succeeded in visualizing income and population trends using bar and line charts. He also succeeded in visualizing the trade balances of 17 countries using bar charts. Since this form of visual representation of information was considered entirely novel, William Playfair is regarded as the inventor of modern chart types for bar, line, column and pie charts (Figure 2 and 3) (Spence, 2006).

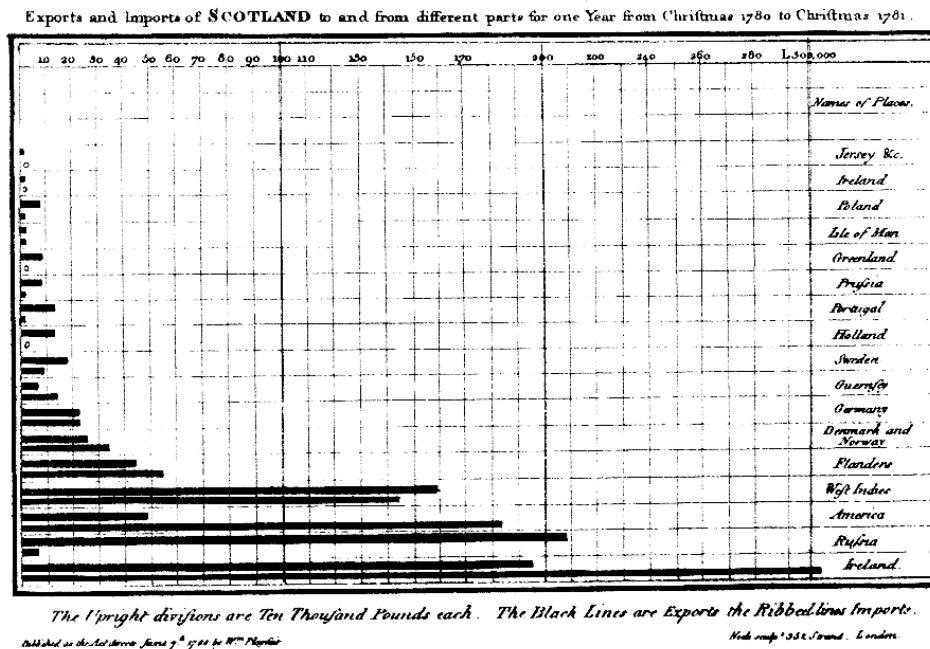


Figure 3: Illustration of 17 trade balances (Spence, 2006)

Another 50 years later, in 1854, the British physician John Snow caused a sensation with his geographical analyses. He investigated a sharp increase in cholera cases in the center of London. The prevailing theory in the mid-19th century was that cholera was an airborne infectious disease. John Snow challenged the prevalent theory of transmission of this deadly disease and therefore conducted the world's first epidemiological study of differential mortality from cholera in 32 London subdistricts. To identify the transmission routes of cholera, he drew the points of people who died on a map. What was unusual about his drawing was that he used dashes for the deceased people and arranged them parallel to the respective streets.

Thus, small bar charts were formed. In addition to the bars, he used circles to represent water pumps on the map. (Figure 4). Through this representation, it became clear that most of the deaths were in the immediate vicinity of a water pump on Broad Street. His analysis led to the conclusion that cholera is transmitted through contaminated drinking water and not through the air (Ramsay, 2006).

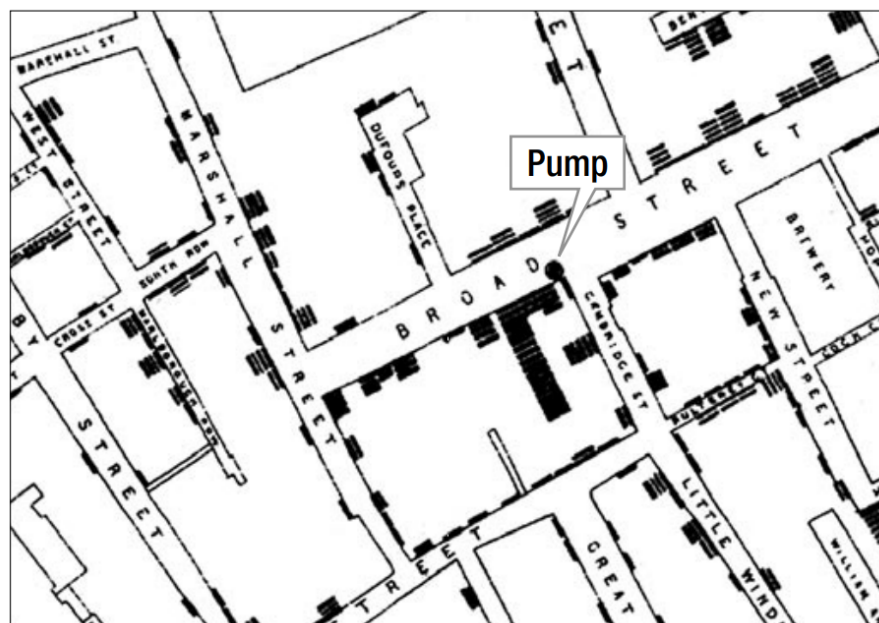


Figure 4: Development of fatalities in Broad Street (Ramsay, 2006)

Another 100 years later, the onset of digitization contributed to the development of interactive visualizations. Above all, the work of Jacques Bertin represents the cornerstone for a dynamic visualization of information. These describe the elementary structure of diagrams and contain basic rules for the visual presentation of information (Bertin 1983).

In addition to the findings of Bertin (1983), Tufte (1983) also developed a summary of principles that are still considered the cornerstone of structure in information design. Specifically, the insight that graphic elegance lies in simplicity of design as well as complexity of data is still considered the mantra of visualization today. Thus, Tufte's (1983) principles emphasize a properly chosen format and design. In addition, he gives recommendations for action, when which visualization technique should be used in which context.

Furthermore, Shneiderman shaped the modern construction of dynamic dashboards in the 1990s by defining the three main functions of interactive reporting. First of all, it is essential to get a good overview of all data. In addition, the displayed data should be dynamically changeable via zoom and filter functions. Furthermore, it must be possible to drill down into the data and perform detailed analyses (Shneiderman, 1996).

3. Research Methodology

As the amount of data has become increasingly important during digitization in recent years, new trends and sub-disciplines are emerging in the field of data visualization and analysis. One sub-discipline defined by Kohlhammer et al. (2018) is Visual Business Analytics. The VBA is divided into three areas: *Information Design*, *Visual Business Intelligence* and *Visual Analytics*, which differ from each other in terms of users, application area, data, and visualization (Figure 5). These three core areas are closely related to Business Intelligence (BI) and Big Data (Kohlhammer et al., 2018). According to Bauer and Günzel (2009), BI is understood as an overall IT-based approach integrated into the company that is used for operational decision support. As a rule, structured data is considered here.

Information Design focuses on an optimal presentation of relevant information with the help of statistical visualizations. In this process, the reporting data is created for the company management as a report or presentation. An example of this is the SAP data from the general ledger, which originates from a balance sheet and is prepared for management. Compared to Big Data, the amount of data is rather small and clearly structured. The goal of Information Design is to achieve high readability of the data. No special requirements are placed on the report user since the reports are created statically as Excel spreadsheets, PDFs or PowerPoint presentations, and the presentation cannot be changed by selecting a parameter or a dimension.

Visual Business Intelligence characterizes the visualization of business analytics and focuses on the interactive and visual use of user interfaces. Depending on the use case, the amount of data to be processed is small, whereby many BI tools are now also suitable for evaluating and visually displaying millions of data records due to in-memory technology. The prerequisite for this is a high-performance server. The data is traditionally structured. Especially when using a data cube, structured data from the data warehouse is often used. A characteristic of Visual Business Intelligence is the possibility of an interactive navigation through the data for a BI user.

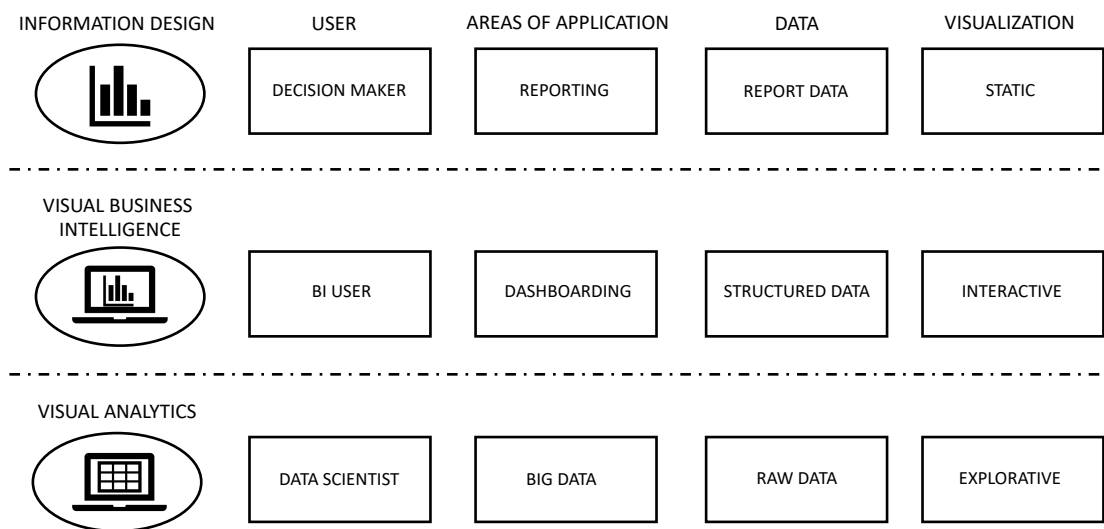


Figure 5: Core areas of the VBA (based on Kohlhammer et al., 2018)

Visual Analytics characterizes the visual use and presentation of Big Data. In contrast to visual business intelligence and information design, visual analytics is explorative. In this case, visual analytics is used by data scientists who do not yet know exactly what the result of the data analysis will be when they start their research. The volume of data is typically substantial in the case of Big Data. This data is available as raw data and can have a varying degree of structuring.

The Data Scientist uses various tools for data analysis. It is crucial that the areas of Information Design, Visual Business Intelligence and Visual Analytics are not mutually exclusive but are combined and provide added value for the company and improve decision-making. Figure 6 illustrates the interaction of these three areas of VBA (Kohlhammer et al., 2018).

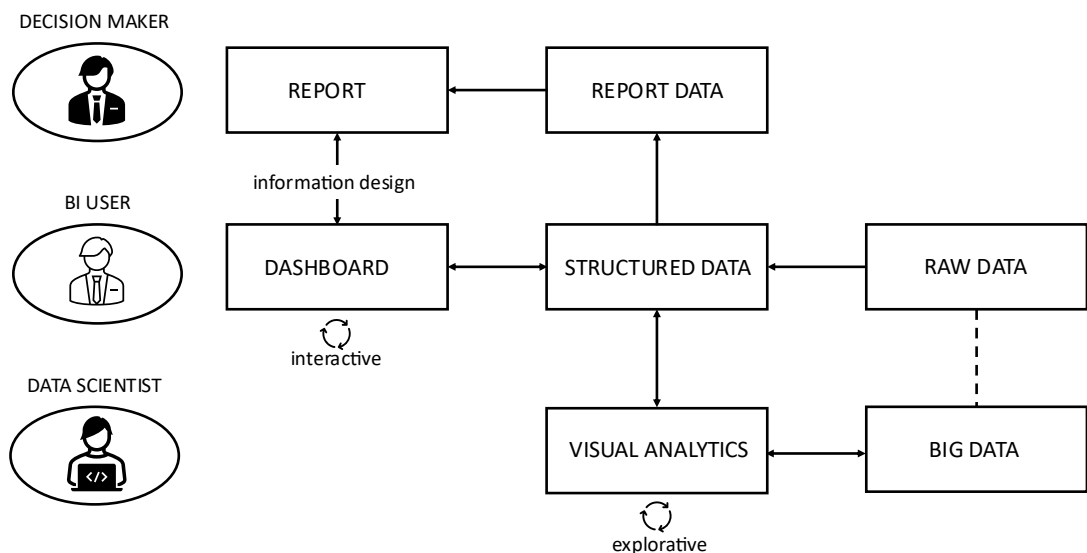


Figure 6: Interaction of the areas in the VBA (based on Kohlhammer et al., 2018)

In the context of VBA, the Data Scientist can use the Python programming language to transform the data into a structured form. Python is now used in many companies and academia in the field of Data Science. This programming language has the advantage of presenting the code as simply and clearly as possible and uses common keywords, operators, and an independent syntax for this purpose. In addition, Python contains comprehensive standard libraries. Python modules can also be integrated into other BI programs such as Qlik Sense or Tableau (Steyer, 2018).

4. Results and Discussion

If the reference is made to a call center in the context of VBA, the focus is on semi-structured log data from the telephone system. This data is processed by the Data Scientist. The unique feature of this data is that the complete log entry of a caller is stored in a single data cell. This means that it has to be transformed into a structured form for analysis purposes. Within a cell, however, the data has a basic structure. For example, there is always a routing entry combined with the associated time. However, a single routing entry is not always in the same place. Whether a caller is connected to a call center agent can be recognized by the entry *Sent to user*. However, this entry does not appear in the same place but depends on the measuring points passed through in the routing. In addition, it is also variable to which call center agent the customer is connected. Therefore, it is necessary to generate this data using different algorithms and transformations.

Time	Event
07:00:00	Initializing
07:00:02	Announcement GDPR
07:00:10	Expected waiting time 10 min
07:00:10	Entered workgroup
07:03:10	Sent to user
07:03:20	Alerting
07:03:25	Assigned
07:07:35	Disconnected

Figure 7: Log of a caller

Figure 7 shows a simplified log entry to give an impression of the primary problem. Within each cell, eight-time points and routing events are mapped. In reality, up to a hundred times the amount of data is stored. From this data, the Data Scientist can extract a lot of new information. For example, both the start and the end of the call can be determined, and thus the total time of the call. It can be determined which announcements the caller heard. In the previous example, this was the data protection instruction in the context of the General Data Protection Regulation (GDPR), the EU regulation on data protection to enhance individuals' control and rights over their personal data (European Union, 2016). It can be determined how long the expected and the actual wait time was, with which call center agent in which workgroup the caller was connected and how long the call handling time was. All in all, this information offers a high added value, which can be further processed in the next step.

After the data scientist has transformed the log data, he examines it for patterns and tries to find correlations between the data. He can use this data to cluster the individual customer concerns using the k-means method to derive commonalities that can be processed in routing. Brown et al. (2004) found that depending on the matter, callers have a different wait time. In their study, they show that customers who want to buy or sell stocks have a significantly higher willingness to wait than customers who call because of a technical concern. Furthermore, the Data Scientist can use the data to set up a machine learning process that is able to predict if and when customers hang up on hold based on wait time.

After the log data is transformed into a structured form, the BI user uses this data to create a dashboard for the decision-maker. The goal of dashboarding is to present the data so that the user can process the relevant information effectively and efficiently. Based on this, he or she derives appropriate action decisions. To facilitate decision-making, the most important call center data should be presented in the top third of the dashboard. Ideally, the overall reachability of the call center, including wait and handling time, should be displayed here. Furthermore, the dashboard should be fully dynamic and offer deeper analysis options so that the user can track each call in detail.

The dashboard created can be used for daily call center management by the various departments or controlling and offers the possibility to react quickly to potential events such as high wait time or faulty routings based on the key figures. Controlling can use this structured data not only for control purposes, but also to generate a static report for the decision-maker. Another application is that the reporting department uses the data from a dashboard and uses it for report generation. Likewise, a special dashboard for management can be constructed to show key metrics and information for decision-making. For example, a report showing low call center reachability at the monthly level may lead to a decision to hire additional call center agents. Furthermore, the data can be used in process mining to examine how long callers spent between different routing points (Figure 8).

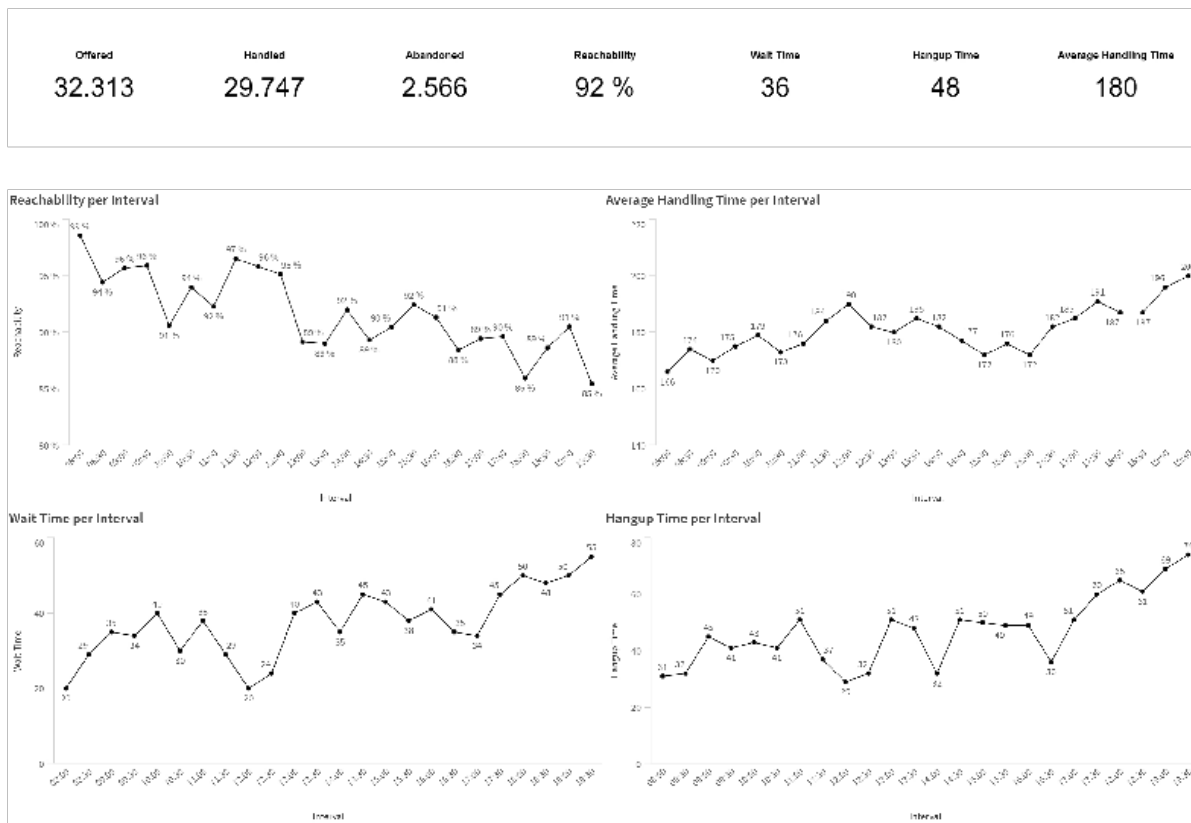


Figure 8: Call center dashboard

However, log data from the telephone system is not the only unstructured data in the call center area. The use of speech analytics offers the possibility of recording the customer conversation and converting it into text for subsequent analysis. The entire conversation can be saved as text, or recognition can be limited to particular keywords. This information can then be used for conversation optimization in the context of employee management. Karakus and Aydin (2016), for example, present the following call management characteristics of a call center agent:

- Formulation of individual relationships
- Use of positive rather than negative terms
- Frequent use of discourse particles, dragging out the conversation.

Without automated processing of this data, the manager must listen in on the call, which exposes the call center agent to more significant psychological pressure. In addition, higher personnel costs are incurred for monitoring. Emails and chat logs form a further basis for analyzing communication with customers in a central service center. Here, too, the data is not fully structured, and it is a challenge for companies to generate added value from this data. With the help of VBA, insights can also be gained here by searching the data for certain keywords (Damnati et al., 2017).

Chat protocols become even more critical when there is no agent in contact with the customer but a chatbot. With the help of chatbots, service centers can completely automate part of the customer concerns and thus reduce personnel costs. However, the chat logs should be used for qualitative reporting to ensure that a chatbot also provides the customer with the appropriate answers to their questions (Okuda and Shoda, 2018).

The situation is comparable for voice assistants, which are often switched before the call center queue and enable customers to authenticate themselves, use self-service products or express their call concerns. Again, it is important to ensure that the customer is properly understood and offered the appropriate process (Vasileanu and Ene, 2018).

5. Conclusion and Outlook

VBA, with its three areas of Information Design, Visual Business Intelligence and Visual Analytics can efficiently process, analyze and visualize data and offers a practical way to take a holistic view of call center log data to improve decision-making at the management level. Overall, the application of VBA has shown how increased added value can result from the semi-structured log data used for the planning, management and control of a call center. This can lead to reduced costs and increased profitability, making it possible to respond more flexible to changing customer needs in a call center.

Moreover, VBA can be used in more than just call centers. Many other industries lend themselves to it. Generally, it is a standard in information technology that log protocols are stored in the databases in order to be able to track error events. Here, IT dashboards offer a very good opportunity to be able to automatically detect error events and quickly correct them (Pi et al., 2019).

Even across industries, the processing of social media data on corporate platforms such as Instagram, Facebook and Twitter has become more of a focus in recent years. Likewise, this is unstructured data, the content of which is intended to inform management. In addition, this data can be used as early indicators, as the number of postings and reactions on social media, for example, can indicate a system failure.

In healthcare, VBA can be used to extract information from image data that effectively contributes to diagnosis. Currently, machine learning is already being used in some medical fields to help with diagnosis. However, VBA will take this process a step further. A dashboard will be made available to the diagnosing physician, offering not only millions of diagnostic images to filter through but also additional information such as ICD codes. The physician can then arrive at a diagnosis within a short period of time and also perform benchmark analyses. In particular, the possibility of incorporating image data into the dashboards and filtering them according to specific criteria greatly facilitates diagnosis findings (Bünthe, 2021).

References

- Bauer, A. and Günzel, H. (2009), Data-Warehouse-Systeme: Architektur, Entwicklung, Anwendung, dpunkt.verlag, Heidelberg, 2009.
- Bertin, J. (1983), Semiology of graphics: diagrams, networks, maps, UMI Research Press, 1983.
- Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S. and Zhao, L. (2004), Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective, in Journal of the American statistical association, 100 (2005), Nr. 469, pp. 36-50. [CrossRef](#)
- Bünthe, C. (2022), Künstliche Intelligenz – Ein Überblick über die aktuelle und zukünftige Bedeutung von KI in der Wirtschaft und im Gesundheitswesen in Europa, in Pfannstiel, M. A. (eds.) Künstliche Intelligenz im Gesundheitswesen. Springer Gabler, Wiesbaden, 2022, pp. 81-100. [CrossRef](#)
- Commerzbank (2021), Commerzbank beschließt neue Strategie bis 2024, Pressemitteilung, Frankfurt am Main: Group Communications (2021). https://www.commerzbank.de/de/hauptnavigation/presse/pressemitteilungen/archiv1/2021/1__quartal_1/presse_archiv_detail_21_01_93834.html.
- Damnati, G., Guerraz, A. and Charlet D. (2016), Web chat conversations from contact centers: a descriptive study. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016.
- European Union (2016), Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>.
- Herzog, A. (2017), Callcenter-Analyse und Management, Springer, 2017 [CrossRef](#)
- Jacobs, B. (1994), Der Einfluß von Graphtyp und Graphanordnung auf das Graphverstehen bei der Analyse von Verläufen, Universitätsklinikum des Saarlandes, Saarbrücken, 1994.
- Jahn, M. (2017), Einzelhandel in Läden–Ein Auslaufmodell?, in Handel 4.0, Springer, Berlin, 2017, pp. 25-50. [CrossRef](#)

- Karakus, B. and Aydin G. (2016), Call center performance evaluation using big data analytics, 2016 International Symposium on Networks, Computers and Communications (ISNCC), IEEE, 2016. [CrossRef](#)
- Kohlhammer, J., Proff, D. U. and Wiener, A. (2018), Visual Business Analytics: Effektiver Zugang zu Daten und Informationen, dpunkt.verlag, Heidelberg, 2018.
- Meinel, C. and Sack, H. (2004), Kommunikationsmedien im Wandel—von der Höhlenmalerei zum WWW, in Meinel, C and Sack, H. (eds.), WWW, Springer, Berlin and Heidelberg, 2004. pp. 55-89. [CrossRef](#)
- Okuda, T. and Shoda, S. (2018), AI-based chatbot service for financial industry, in Fujitsu Scientific and Technical Journal, 54 (2) (2018), pp. 4-8.
- Pi, A., Chen, W., Zeller, W. and Zhou, X. (2019), It can understand the logs, literally, 2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), IEEE, 2019. [CrossRef](#)
- Playfair, W. (1801) The commercial and political atlas: representing, by means of stained copper-plate charts, the progress of the commerce, revenues, expenditure and debts of England during the whole of the eighteenth century, T. Burton, 1801.
- Schön, D. (2016), Planung und Reporting, Springer, Berlin, 2016. [CrossRef](#)
- Seufert, A. (2016), Die Digitalisierung als Herausforderung für Unternehmen: Status Quo, Chancen und Herausforderungen im Umfeld BI & Big Data, in Fasel, D. and Meier, A. (eds.), Big Data, Springer, Berlin, 2016, pp. 39-57. [CrossRef](#)
- Shneiderman, B. (1996), The eyes have it: a task by data type taxonomy for information visualizations, in Proceedings 1996 IEEE Symposium on Visual Languages, Boulder: IEEE Computer Society, 1996, pp. 336-343.
- Spence, I. (2006), William Playfair and the psychology of graphs, in Proceedings of the American Statistical Association, Section on Statistical Graphics, 2006, pp. 2426–2463.
- Steyer, R. (2018), Programmierung in Python, Springer, Berlin, 2018. [CrossRef](#)
- Tufte, E. R. (1983), The Visual Display of Quantitative Information, Bd. 2, Graphics Press, Cheshire, 1983.
- Tufte, E. R. (1997), Visual explanations, Graphics Press, Cheshire, 1997.
- Vasilateanu, A. and Ene, R. (2018), Call-center virtual assistant using natural language processing and speech recognition, in Journal of ICT, Design, Engineering and Technological Science, 2 (2), 2018, pp. 40-46. [CrossRef](#)