Defining Analytics Maturity Indicators: a survey approach

Abstract

The ability to derive new insights from data using advanced machine learning or analytics techniques can enhance the decision-making process in companies. Nevertheless, researchers have found that the actual application of analytics in companies is still in its initial stages. Therefore, this paper studies by means of a descriptive survey the application of analytics with regards to five different aspects as defined by the DELTA model: data, enterprise or organization, leadership, targets or techniques and applications, and the analysts who apply the techniques themselves. We found that the analytics organization in companies matures with regards to these aspects. As such, if companies started earlier with analytics, they apply nowadays more complex techniques such as neural networks, and more advanced applications such as HR analytics and predictive analytics. Moreover, analytics is differently propagated throughout companies as they mature with a larger focus on department-wide or organization-wide analytics and a more advanced data governance policy. Next, we research by means of clustering how these characteristics can indicate the analytics maturity stage of companies. As such, we discover four clusters with a clear growth path: no analytics, analytics bootstrappers, sustainable analytics adopters and disruptive analytics innovators.

Keywords: Analytics maturity; analytics techniques; organizational characteristics; survey research

1. Introduction to analytics maturity

Being able to derive insights from data and use them in decision-making has become more and more important in the last few years, as emphasized in a recent special issue of MIS Quarterly on transformational issues of big data and analytics in networked business (forthcoming). However, it is unclear to what extent companies are already applying analytics, also referred to as data science, nowadays as there are still a lot of challenges. Moreover, how does this influence the level of analytics maturity in companies?

One of the most well-known analytics maturity models was developed as early as in 2007 by Davenport & Harris (2007), who composed five consecutive stages of analytical competition. Their analysis is focused on analytics as a driver for competitive advantage. In 2010, this was complemented with the DELTA framework (Davenport et al., 2010) which stands for accessible, high-quality *data*; *enterprise* orientation; analytical *leadership*; strategic *targets*; and *analysts*. The authors developed several general guidelines per success factor to transition from one stage of analytical competition to the next. Saxena & Srinivasan (2013) propose a maturity model with three dimensions: capability, culture and technology. They note that companies often excel in capability but lag with regards to technology. All three dimensions should, however, be in balance. In their work, Cosic et al. (2012) aim to develop a business analytics capability maturity model. They define sixteen business analytics capabilities spread out over four capability areas: governance, culture, technology and people. Comuzzi & Patel (2016), on the other hand, developed a model specifically for big data maturity consisting of five domains with each six levels, namely strategic alignment, data, organization, governance and information technology. Other researchers aim to define maturity levels based on survey research. As such, LaValle et al. (2011) define three levels of analytical capability: aspirational, experienced and transformed. Finally, Ransbotham et al. (2015) propose three maturity levels: analytically challenged, analytical practitioners and analytical innovators. All models try to classify the stages of analytics maturity and are based on experience and interpretation of survey and interview research. Furthermore, maturity models exist concerning the related topics of data warehousing and business intelligence. Frequently, these models are developed by companies such as AMR Research (Hagerty, 2006), Gartner (IBM, 2009), and HP (2015). For an overview, the reader is referred to Muller & Hart (2016).

We aim to complement previous research by reviewing how analytics is currently applied and how these findings impact analytics maturity. For this purpose, clustering on questionnaire data was performed in order to expose underlying maturity levels.

In what follows, we first describe our research methodology. In section 3 our findings with regards to how analytics is currently applied, are presented. Then, in section 4, we discuss how these characteristics can indicate a higher analytics maturity level. Finally, the findings are validated in section 5.

2. Material and methods

In this study, we opted for descriptive survey research because this method is recommended for researching phenomena in their natural settings and it allows us to collect quantitative descriptions about the studied environment (Pinsonneault & Kraemer, 1993).

2.1. Survey development and validation

Two cross-sectional, world-wide surveys were developed, targeting middle to large companies from all types of industries, e.g. financial services, healthcare, technology, telco, utilities, pharmaceutics and HR, and various levels of analytics maturity ranging from no applications to analytics embedded throughout the whole organization. The first questionnaire is an extensive study of the organizational characteristics of analytics in the responding companies and how they report to apply analytics and use the resulting insights. In order to improve uniformity accross responses we started the questionnaire with the definition of Davenport & Harris (2007) for analytics, namely analytics is the "extensive use of data, statistical and quantitative analysis, explanatory and predictive models and fact-based management to drive decisions and actions". Before going live, this questionnaire was subjected to a pre-test by means of six interviews with analytics experts from the financial services, retail, real estate, telco and government sector. Each expert completed the questionnaire and provided extensive feedback and suggestions in order to test the survey. The findings from the first questionnaire were also validated by means of seven interviews with analytics experts from the financial services, retail, real estate and telco sector. A second, follow-up questionnaire was sent out one year later with the purpose of validating the previous findings.

During these phases, some measures were taken to improve *generalizability*. The respondents are analytics and IT experts from a variety of sectors, functions and countries which leads to a balanced and knowledgeable sample. For each question, they were given the option to select 'I do not know'. Furthermore, anonymity was guaranteed. These measures improve the external validity of the study. Nevertheless, some limitations remain. A larger sample size and better response rate would further ameliorate the generalizability. Furthermore, given the focus of the survey and the respondents targeted, the number of companies not applying analytics might be underestimated.

2.2. Data collection

Seventy-three responses were collected during the first survey[1] between March and June 2015 by contacting relevant profiles in information technology (IT) and analytics by means of e-mail (response rate = 9.27% out of 205 contacts) and social media. We reached a variety of profiles as summarized in Tables 1 and 2 for the companies' and respondents' profiles respectively. This information was gathered at the end of the questionnaire. Note that 5% of the respondents stated that analytics was not applied in their company and were thus excluded from

[1] The questionnaire can be found at [LINK omitted to preserve anonymity, please see attachment]

analytics application analyses. The other responding companies either only apply analytics for specific projects or initiatives (16%), apply analytics actively in certain departments (52%), or have already integrated analytics throughout their company (26%). Moreover, 'only' 38% of our respondents reported that they have been applying analytics for 10 years or more. Fifty-six percent of them have been applying analytics for at least 5 years, and 76% for at least 2 years.

An additional 32 responses were collected for the follow-up survey[2] during July and August, 2016, by contacting chief-level executives in data and analytics (response rate = 18.93%). On a 5-level scale from no knowledge and experience to expert in analytics, they rate on average level 4. More details about the profiles of the respondents can be found in Table 3.

Table 1: Description of the companies' profiles for the first questionnaire

Sector	**Publicly listed**	**(Partly) governmental**	**Market**	**Regions**	**Globalization level**
Consulting: 11% Financial services: 37% Government: 5% Healthcare: 7% Marketing & communication: 3% Technology: 7% Telecommunication: 5% Utilities: 3% Other: 14% Not specified: 8%	Yes: 49% No: 41% Not specified: 10%	Yes: 18% No: 70% Not specified: 12%	Offline: 77% Online: 64% Both online & offline: 56%	Asia: 33% Africa: 19% Europe: 60% North America: 41% Oceania: 18% South America: 23% Not specified: 23%	Local: 10% National: 19% International: 37% Global: 25% Not specified: 10% *(Note that local refers to companies active within specific regions of a country, e.g. only one city or state.)*

[2]The questionnaire can be found at [LINK omitted to preserve anonymity, please see attachment]

Table 2: Description of the respondents' profiles for the first questionnaire

Function	Functional domain	Personal involvement in analytics
Senior executive: 22%	No specific domain: 15%	Function in analytics: 69%
Executive: 23%	Business analytics: 29%	No function in analytics but
Project leader: 18%	Finance: 14%	collaborate with data scientists: 12%
Manager: 7%	HR: 1%	No function in analytic but
Data scientist: 14%	IT: 14%	make decisions based on analytics: 6%
Business user: 4%	Marketing: 5%	
Other: 4%	Operations: 7%	
Not specified: 8%	Sales: 3%	
	Other: 3%	
	Not specified: 10%	

Table 3: Description of the respondents' profiles for the second questionnaire

Level of analytics experience	Analytics involvement
No knowledge: 0%	Applying analytics: 41%
Basic knowledge but no experience: 6%	Managing data scientists: 50%
Experience with classic analytics techniques: 22%	Collaborating with data scientists: 6%
Proficient experience: 38%	Basing decisions on analytics insights: 3%
Expert: 34%	

2.3. Survey testing

We ran some tests for survey bias. Non-response bias was tested by comparing the answers from the first and last quartile of respondents to e-mail invitations (Armstrong & Overton, 1977). After applying Pearson's Chi-squared test, the null hypothesis that there is no *non-response bias* could not be rejected at the 90% confidence level. *Common method bias* was tested using analysis of variance (ANOVA) (p-value > 0.80) and the Tukey method (p-values > 0.80). The propagation of analytics within their company did not correlate with the response duration of participants. Finally, *collection method bias* measures if there are differences depending on the collection method used. There is a correlation between the collection method and the response duration (ANOVA, p-value < 0.01). As such, respondents to e-mail invitations took a significant longer time to complete than respondents to social media invitations (median response times are 26.26 and 46.47 minutes for social media and e-mail respectively). The reader should thus take into account that respondents to e-mail invitations might respond differently from respondents to social media invitations. Note that the central limit theorem was applicable for common method bias testing and that for collection method bias testing the sample size was sufficient. However, non-response bias tests require cautiousness.

2.4. Data analysis

The survey data was analyzed using descriptive and predictive analytics techniques and consecutively presented using a categorization based on the DELTA model (Davenport et al., 2010). As such, the findings describe the characteristics of companies with regards to analytics on a data, organizational, leadership, techniques and applications, and analysts level. Thereafter, the impact of each characteristic on analytics maturity is analyzed using partitioning around medoids clustering (Reynolds et al., 2006), with a distance matrix calculating Gower's distance (Gower, 1971), on the analytics techniques and applications of the participating companies. We set the number of clusters to four based on the ratio of the average distance between clusters and the average distance within clusters. For more technical details about the clustering technique, the reader is referred to Appendix A.

3. How is analytics applied?

This section presents our findings with regards to different perspectives of analytics.

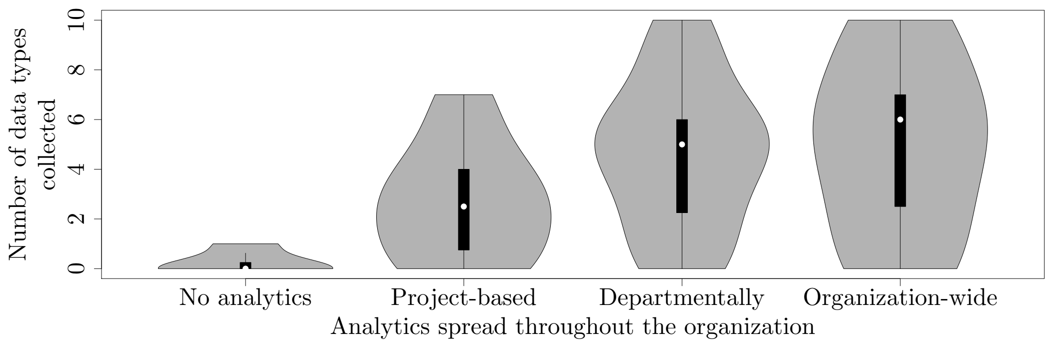
3.1. The data perspective

The competence to integrate data from multiple sources and the ability to share data is regarded as a key requirement for analytics (Kiron et al., 2012). Therefore, this section discusses different aspects of data and how companies deal with these.

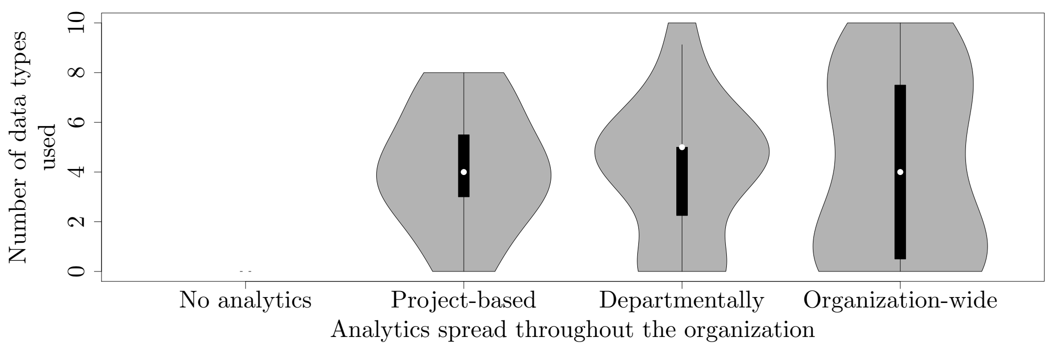
We observe that purchase data is the most commonly collected data type (59%), while demographics data is used the most (58%). This confirms that demographics are used most commonly for analytics, but also purchase data prove to be valuable for companies, as already indicated in 1996 by Rossi et al. (1996). We notice a discrepancy between the types of data collected and used which is potentially caused by excessive data collection and augmentation by external data. For instance, 42% of respondents indicate collection of text data, while only 25% actually use it. The same trend can be observed for clickstream data: 41% collect these data and 36% use it. The tendency to collect more data than one actually uses, which holds in particular for unstructured data, is illustrated by the concept of data lakes. In contrast to traditional data warehouses, data lakes collect data that is not (or only to a limited extent) transformed, cleansed and prepared for certain types of analysis. As such, possibly useful data is collected 'as is' without knowing upfront which types of analysis (if any) will be performed on it.

Furthermore, we found that those companies in which analytics is more spread throughout the organization, appear to collect and use a wider variation of data types for analytics. There is a correlation between the propagation of analytics within a company and, on the one hand, the number of data types collected (ANOVA, p-value < 0.05) and, on the other hand, the number of data types used (ANOVA, p-value < 0.10), as can be observed from Figure 1. As such, there is a moderate increase of 2.4 (p-value $= 0.16$) in the number of data types collected going from no analytics to specific analytics projects, and an additional increase of 1.9 (p-value < 0.10) for departmentally organized analytics. Regarding the number of data types used, the difference is most observable going from no analytics to specific projects with a 4.2 increase in the number of data types (p-value < 0.05).

Data management and quality issues still challenge the analytics organization. In order to make significant advancements in analytics and specifically the adoption of big data, data quality is very important (Kwon et al., 2014). We found that most common big challenges for analytics are still data management issues such as the integration and sharing of data. Furthermore, a lack of adequate documentation and data quality issues such as accuracy, preciseness and consistency occur and only a minority (36%) of the companies are using standard data definitions. One year later, our follow-up survey indicated that most experts (77%) still agree that only a minority of the companies apply standard definitions and coding for data.



(a) Data collection



(b) Data usage

Figure 1: Violin plots of the number of data types (from 0 to 10) collected (a) and used (b) for companies who do not apply analytics, who apply analytics project-based, who apply analytics departmentally, and who apply analytics organization-wide.

3.2. The organizational perspective

There is some discussion on whether it is best to organize analytics centrally or locally. A centralized analytics team can offer economies of scale and scope. However, if companies outsource analytics they can profit from the provider's economies of scale and scope (Saxena & Srinivasan, 2013). Furthermore, centrally organized analytics resources can be employed more efficiently and effectively. Nevertheless, in reality dispersed models, with a total lack of centralization, are most common (Davenport et al., 2010), although they are believed to be less mature (Davenport et al., 2010; Griffin & Davenport, 2011; LaValle et al., 2011). Furthermore, entirely centralizing an analytics team is nearly impossible if companies want to establish a data-driven decision making culture (Saxena & Srinivasan, 2013). Therefore, a centralized analytics team can be complemented with localized teams in order to not give up the efficiency and effectiveness a centralized team offers (LaValle et al., 2011).

We observe that most respondents organize their analytics centrally, either in a centralized team or in a center of excellence (CoE) (47%). Another 23% combine a central team/CoE with locally organized teams. In comparison, only 6% do not organize their analytical resources in some way. However, 27% have next to some form of organized analytics also analysts not coordinated by means of central or local organization. Furthermore, 8% outsource some of their analytics and 21% hire consultants for part of the analytics work. We also observe that 24% organize analytics within a department, without any form of central coordination. In general we can state that several formats of organizing analytics are used in modern organizations and that companies frequently combine various setups. Nevertheless, we can clearly observe a preference for some form of central coordination. Moreover, companies should keep in mind that both the competitive environment and their own company might evolve regarding analytics. So the organizational format should be regularly evaluated and adapted. Upon surveying experts one year later, we found that there is still some discussion. Only 66% (strongly) agree that there is a preference for central coordination of analytics.

Furthermore, we observe that personnel holds the lion's share of the total cost of ownership (TCO) of analytics. Costs and benefits of analytics are not easy to calculate. If we want to take a realistic look at the cost of analytics, we need to consider the TCO. The TCO takes hardware, software, personnel, education and external resources into account, as is shown in Figure 2. We observe that the biggest share in TCO is personnel, with an average of 42%, followed by software (20%) and hardware (12%). Note, however, that the variance in the personnel share is relatively high ($\sigma = 33.92$) indicating a large difference in the nature of analytics spending across organizations. Upon asking one year later if the largest part of the TCO still goes to personnel, also only 63% agree. Relatively little capital is spent on education (6%) and external resources (7%).

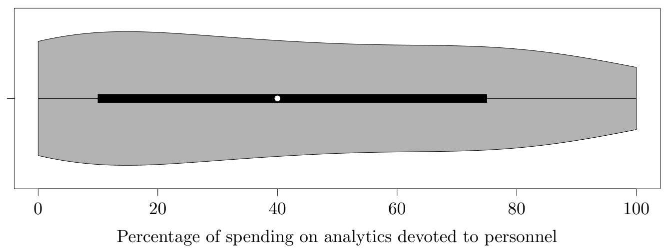
3.3. The leadership perspective

We found that analytics is conquering a seat in the board. There has been some discussion about the best way to organize your analytical talent (Davenport et al., 2010; LaValle et al., 2011; Saxena & Srinivasan, 2013). We observed that in modern organizations there are three general cases with equal occurrence: either an existing chief-level executive (CXO) takes on the responsibility for analytics (29%), a new CXO function is created (24%) or each department head is responsible for the analytics in his/her department (29%). Only a minority have a middle manager (8%), the chief executive officer (CEO) (6%) or no one (5%) responsible. This suggests that companies are increasingly organizing their analytics at a high organizational level. One year later this is still the case as 38% agree that analytics is conquering a seat in the board and an additional 16% even strongly agree. Furthermore, if we take a closer look at the reporting flow, we observe that the analytics responsible reports most frequently to the CEO (31%), closely followed by reporting to another CXO (27%). When we zoom in on new CXOs, such as a chief analytics officer, 37.5% report to another existing CXO and 62.5% report directly to the CEO.

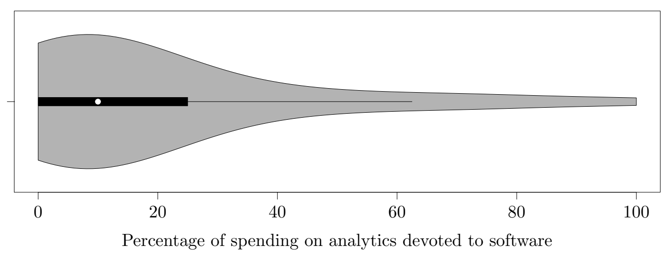
3.4. The techniques and applications perspective

In general, predictive analytics is believed to be more mature than solely descriptive analytics (Davenport et al., 2010). Therefore, we take a look at which specific application and techniques are present in companies.

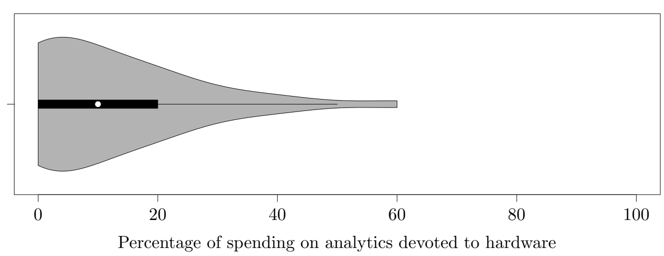
There is a preference for well-known applications of analytics such as finance and marketing analytics while HR analytics is less popular. We discovered that analytics is mainly applied for marketing (85%), financial (77%) and operations (74%) objectives. With regards to companies who do marketing analytics, descriptive analytics



(a) Personnel



(b) Software



(c) Hardware

Figure 2: Percentage of TCO of analytics devoted to personnel (a), software (b) and hardware (c).

-such as marketing performance analysis (68%), customer segmentation (65%) and direct marketing (63%)- is more popular compared to predictive analytics -such as churn prediction (40%). Furthermore, we observe that personalization is still not commonplace, 33% apply analytics for personalized communication and 24% of the companies apply it for personalized online advertising. Interestingly, 30% apply social media analytics and 26% sentiment analysis. Companies who employ finance analytics, most commonly apply it for analyzing financial performance (67%) in comparison to a lower application for fraud detection (48%). Companies with operations analytics most commonly choose for process improvement practices (56%) while analytics for store lay-out is not widespread (10%). Analytics for human resources (HR) is less common (40%) and less evolved. Of this group, 21% address analytics for employee performance management compared to 15% for employee retention and 13% for employee acquisition.

Furthermore, we researched which corporate features impact these objectives by means of linear regression. Backwards variable selection was applied based on p-values (< 0.10). As such, a positive correlation can be found between, on the one hand, the number of finance analytics objectives and, on the other hand, if companies are governmental and if they operate (also) on the online market place (adjusted $R^2 = 0.54$). Marketing analytics objectives are positively correlated with (also) operating on the offline market place, with being publicly listed, and with a higher level of globalization (adjusted $R^2 = 0.60$). Possibly, marketing analytics has been found to be important or more accepted for companies competing offline, while the same takes place for finance analytics when competing online. For the details of the linear regression models, the reader is referred to Appendix B.

Next, we found a prevalence of understandable techniques such as linear regression and decision trees. We notice a higher application of better understandable and less complicated (white box) techniques such as decision trees (74%) and linear regression (74%) compared to neural networks (33%) and support vector machines (22%), as can be observed from Figure 3. We assume that the application of these more complicated, yet well-performing (Lessmann et al., 2015) techniques will still grow as we witnessed with some of the analytically more mature companies, see section 4. As indicated by a Mann-Whitney U test (p-value < 0.05), companies applying self-organizing maps, regression techniques, survival analysis, etc. started earlier with analytics. Moreover,

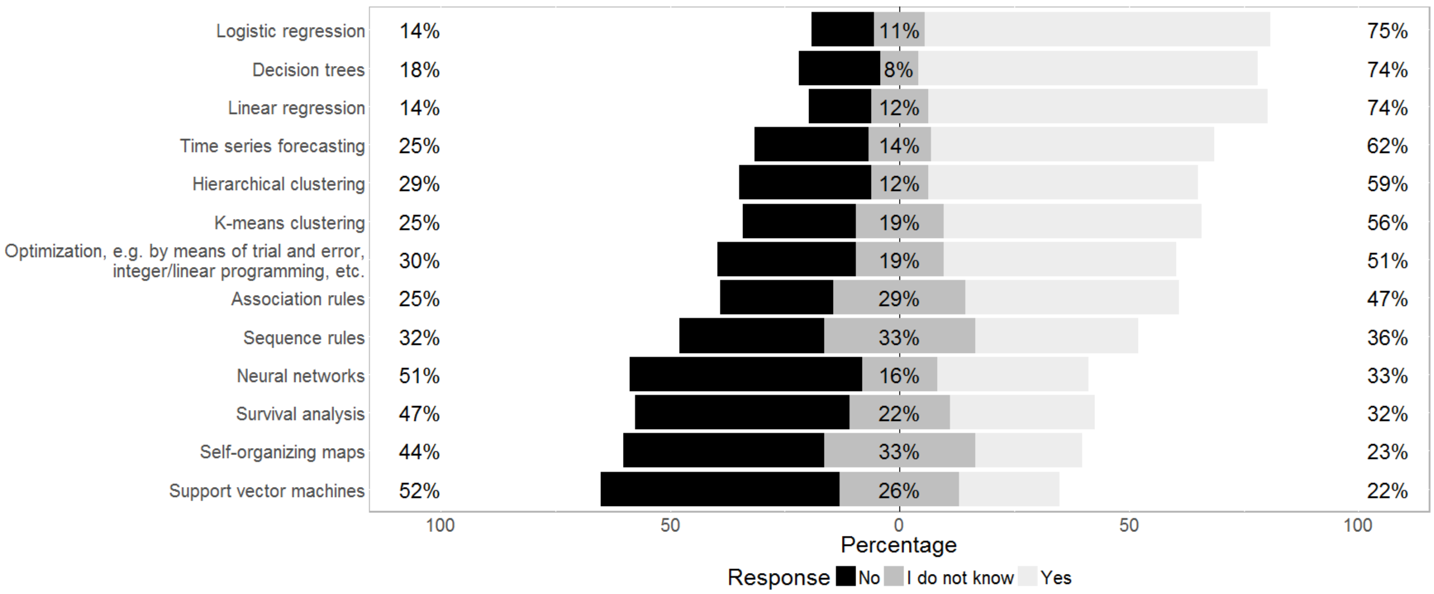


Figure 3: Percentage of respondents indicating that a particular technique is used for analytics in their organization.

we found a positive correlation of 0.29 (p-value < 0.05) between those companies who are already applying analytics for a longer time and the number of techniques they use. With regards to the tools, it is interesting to recognize a high application of spreadsheets (88%). Nevertheless, dedicated analytics and statistics tools are also commonly applied with a slight preference for commercial (77%) versus open source (64%) analytical packages. Finally, we observe that 45% apply web analytics tools.

In our second survey, one year later, experts still believe that analytics is more commonly applied for well-known applications (81% agree) and by means of well-known, understandable techniques (84% agree), while the application of HR analytics is relatively low (78% agree).

3.5. The analysts perspective

A couple of years ago, being an analyst or data scientist was called the sexiest job of the 21st century (Davenport & Patil, 2012). Although data science continues to be a popular job, it brings coaching and coordination challenges for managers. Therefore, this section takes a closer look at the analysts themselves.

We observe that the analytics team is growing. Analytics teams are relatively small, young and male. As such, 45% of the companies have up to ten data scientists while only 18% have more than fifty data scientists. However, this team has been growing for most companies. Sixty-one percent grew even with more than 25% during the last five years. We performed a Wilcoxon Signed-Rank test to compare the number of data scientists in 2010 with 2015. After outliers were removed, companies proved to have significantly more (p-value < 0.001) data scientists with a mean increase of 9 analysts. In 2016, a large majority (94%) still believes that the number of data scientists in a company increases. Note, however, that team size might depend on corporate variables. As such, by means of linear regression (adjusted $R^2 = 0.47$), we observe a positive correlation with company size, measured as the number of employees, and how long ago the company started with analytics. We also discover that companies who only operate online and have an analytics culture clearly driven by senior management positively correlate with a larger data science team. For a more detailed overview of the results of the linear regression model the reader is referred to Appendix C. When we take a closer look at gender distribution, we observe that most data scientists are male. Nevertheless, 20% of the teams have a more or less equal gender distribution and 7% have a largely female team.

It is commonly accepted that businesses are looking for their data scientists to have a combination of quantitative and software skills together with business understanding. These needs still exist. Business understanding was indicated by 83% as very important; software skills, statistical knowledge and analytics skills were found very important respectively by 84%, 72%, and 63%. Noticeably, some atypical needs are also valued such as legal knowledge (72% rated this skill as very important or valuable) and data quality understanding (100% rated this skill as very important or valuable), while hardware skills are generally regarded as not important (49% rated this skill as not important). The latter might be dedicated to the observation that IT plays a supportive role for analytics in 79% of the companies. IT is encountering new opportunities with regards to data science support (DeLine, 2015) and this might contribute to a change in the responsibilities of IT and analytics. Furthermore, new skill requirements are arising due to new data types, techniques and tools as illustrated earlier. We inquired about how data scientists are trained and observed that most of them teach themselves (84%), possibly complemented by internal (57%) and external (41%)

trainings, e.g. by vendors. Moreover, 55% send their data scientists to conferences. Only a low percentage of the companies are involved with other forms of collaboration such as data science communities (28%) and academic collaborations (26%). A training format more frequently observed in business departments is rotational deployment. This is, however, not commonly used in analytics training (13%) although it can provide an interesting format.

4. Indicators of analytics maturity

Using partitioning around medoids, see section 2.4, we clustered the respondents into four groups based on their analytics techniques and applications, see Figures 4 and 5. The first cluster is relatively small, containing 5.5% of the responding companies, and is called "no analytics". The second cluster, denoted "analytics bootstrappers", contains 20.6%, the third, namely "sustainable analytics adopters", contains 38.4% and the last one, "disruptive analytics innovators", contains 35.6% of the companies. The propagation of analytics within these companies significantly (p-value < 0.001, Chi-squared test) varies among the clusters. Therefore, we will treat each cluster as a separate analytics maturity level. The key characteristics of each group and recommendations, based on the DELTA model (Davenport et al., 2010), towards how to improve the analytics maturity are summarized in Table 4. These characteristics are distinct across each maturity level and, as such, form analytics maturity indicators.

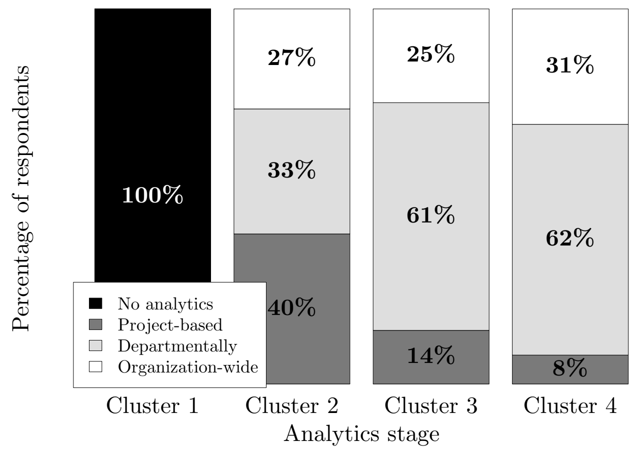


Figure 4: The propagation of analytics within a company. For each cluster, the percentage of companies with each level of analytics propagation is shown on the vertical axis.

Table 4: Characteristics of and recommendations for the different analytics maturity levels.

Cluster	Description
1: no analytics	**Key characteristics.** In our sample, they are small companies (median of 10 employees) of which none are active on the online market place and who operate on the rather local end of the globalization spectrum. **Key recommendation.** In order to grasp the potential of analytics, we recommend to *start small* with specific projects and simple techniques which can, nevertheless, contribute added value to the organization, e.g. a web analytics project. Furthermore, it is important to work on a *corporate data management strategy* such that the data are available and of high quality when one needs them.

2: analytics bootstrappers	**Key characteristics.** In our sample, they are relatively larger than cluster 1 (median of 1,200 employees) and mostly international companies. Furthermore, 67% operate on the online market place. On average, they *started with analytics 5 years ago* and employ 4.5 data scientists. Their application of analytics techniques is relatively low, with a high focus on online analytical processing (OLAP) (86%) and *basic segmentation* (71%). The latter aligns with a common existence of marketing analytics (86%). The existence of finance and operations analytics is moderate (62% and 60% respectively) but they show a very low application of HR analytics (22%). Remarkably, they do not only have an almost non-existent application of more complex techniques such as neural networks, but also a *moderate application of basic techniques* such as decision trees and linear regression. We also inquired about the governance and impact of analytics. With regards to data management, most companies do *not agree that data quality is governed* in their company (73%) although half of them agree that it is measured. Moreover, data integration and standardization are an issue. Nevertheless, 60% indicate that senior management clearly emphasizes a data-driven decision-making process although also 67% agree that decisions are still *largely based on intuition.* Only 33% claim that analytics is very mature in their company. Furthermore, 47% indicated *lack of in-house analytical skills* as a big challenge. **Key recommendation.** We recommend to work on a *corporate data governance strategy* and to focus on the *coordination* of the data scientists in order to further develop skills and techniques and to increase the number of applications.
3: sustainable analytics adopters	**Key characteristics.** In our sample, these companies are not applying analytics for a long time yet (median of 3 years). The difference in years between this and the previous cluster is, however, not significant (Mann-Whitney U test, p-value $>$ 0.85). They have on average 15 data scientists but are in general also larger (median of 3,500 employees). Seventy-five percent are active on the online market place and most companies operate internationally (43%). Slightly more of them perform finance (83%), marketing (88%) and operations (78%) analytics, but only a minority apply HR analytics (16%). We take a closer look at the specific analytics techniques and observe a *high adoption of understandable techniques* such as regression and decision trees. Nevertheless, their application of more complex techniques such as neural networks and survival analysis remains rather uncommon. Similar to the previous cluster, data integration (70% agree) and standardization (65%) are challenges but their decision-making process is less commonly impacted by intuition. Moreover, 46% agree that analytics is already very mature at their organization. **Key recommendation.** We recommend to manage analytics at the executive level and to focus on an *organization-wide coordination* and impact of analytics. A *central organization* of data scientists can improve knowledge-sharing and initiate new applications. Furthermore, performance can be boosted by a larger focus on data quality and management and by exploring more advanced techniques.

4: disruptive analytics innovators	**Key characteristics.** These companies *started* with analytics significantly *earlier* than in the previous clusters (median of 12 years). They have on average 30 data scientists but employ also more people in general (median of 10,000 employees). In our sample 62% of the companies are active on the online market place. Moreover, these companies operate more on the global end of the globalization spectrum with the lion's share operating globally (35%). This category has a larger share of companies practicing finance (91%), marketing (96%) and operations (91%) analytics. Interestingly, 76% of these companies apply *HR analytics*. With regards to the analytics techniques, they score high on both simple, understandable techniques and more *complex techniques* which allows them to discover new, disruptive insights. We also observe that a lack of adequate technologies is not really a challenge, while this is a small challenge for the other clusters. On the other hand, they identify privacy issues and inadequate documentation as a big challenge which suggests that they are more aware of their data management. Indeed, most of these companies measure (67%) and govern (57%) their data quality and are able to integrate their data (59% agree). A *lack of data standardization*, nevertheless, remains a problem (67% agree). Moreover, 57% agree that analytics is very mature at their company. Their key strength is the propagation of analytics in the organization and a *culture of data and analytics* which, in combination with advanced analytics techniques, empowers them to use analytics as a disruptive enabler in their key strategic business processes. **Key recommendation.** In order to be able to leverage analytics as a true competitive advantage, they should ameliorate their data quality and management even further. Specifically, they should also look into their documentation process.

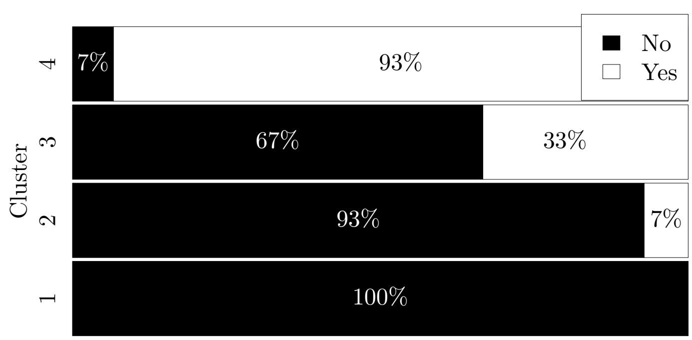
The four clusters suggest a growth path which indicates an increase in analytics maturity. Companies started significantly earlier as their level increases going from no analytics (cluster 1), to analytics on a middle-long term (clusters 2 and 3), to a long term innovative analytics implementation (cluster 4). While their analytics organization advances, also their applications and techniques do. As such, we observe that HR analytics is only common for companies in the last cluster. Although most of the companies in clusters 2 to 4 apply finance, marketing and operations analytics, we observe an increase in the variety of these applications. The most common techniques observed for analytics bootstrappers are OLAP and segmentation. Next, we already observe that sustainable analytics adopters apply simple and understandable analytical techniques. Neural networks, survival analysis, etc. are only commonplace for disruptive analytics innovators. The propagation of analytics within a company also changes, see Figure 4, as the share of companies with only project-based analytics decreases in favor of departmentally organized and organization-wide analytics. Note

(a) Decision tree

(b) Linear regression

(c) K-means clustering

(d) Neural networks



(e) Survival analysis

Figure 5: Percentage of companies who apply this analytics technique (horizontal axis), for each level of analytics maturity (vertical axis).

that this analysis indicates a general trend, and not necessarily means that companies who apply analytics only project-based are not able to apply advanced techniques. Furthermore, we remark that data quality is always a challenge, but the way companies govern it matures. As such, data integration becomes less of an issue and new challenges, such as privacy and documentation, arise for the most mature organizations. The decision-making process, moreover, becomes more data-driven. In the end, we also observe that the higher the level, the more companies who claim that analytics is mature at their organization.

One year later, we asked to what extent respondents believe these characteristics are indicators of a higher level of analytics maturity. We observe that the biggest indicators are the extent to which senior management emphasizes data-driven decision making, the extent to which decisions are based on data rather than intuition, the level of data standardization and whether companies do e-business. This can be observed from Figure 6.

This categorization suggests that there will always be challenges but that these challenges change as organizations mature. First, gathering the necessary data and skills is a key factor. Consecutively, the data governance needs to be elaborated at a corporate level. Simultaneously, the projects progress and more strategic projects develop across departments. At this point the company must work towards a central coordination model for their analytics. Senior management must discuss the decision-making process and further develop the analytics culture and data quality.

5. Validation of findings

We presented the findings in this paper to seven experts and summarize their opinions in this section. In general, they confirm that analytics is not yet widespread within organizations. More progress needs to be made in terms of 'typical' analytics objectives and techniques before turning to more advanced projects, such as working with unstructured data, social media analytics and neural networks. The lower application of neural networks compared to other techniques does not surprise experts. Neural networks are hard to understand due to their black box nature which is often not desired by the business and in some sectors not even allowed by law. Moreover, the younger the analytics organization, the more important it is that data scientists are able to explain their results. Experts also bring up two possible reasons for why HR

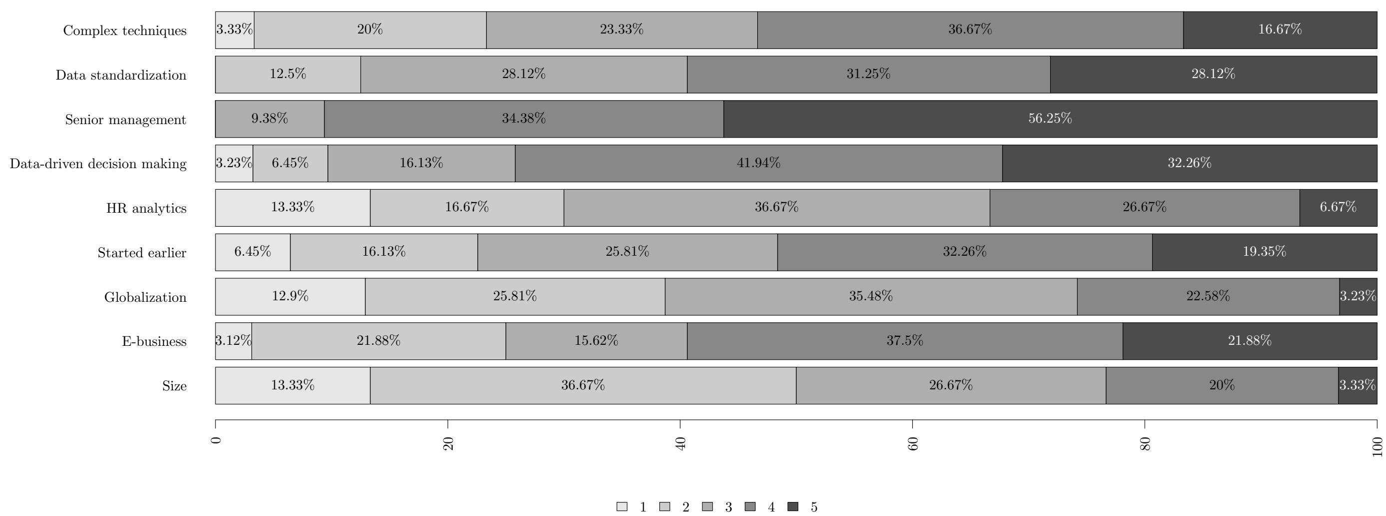


Figure 6: On the vertical axis several company characteristics can be found: the complexity of analytics techniques, the level of data standardization, the extent to which senior management emphasizes data-driven decision making, the extent to which decisions are based on data rather than intuition, whether a company performs HR analytics, how long ago a company started with analytics, the level of globalization, whether the company performs e-business and the size of a company. On the horizontal axis the percentage of respondents who believe this characteristic positively indicates analytics maturity is represented. The darker, the more respondents who agree on a scale from 1, no indication, to 5, a significant indication.

analytics has not taken off yet. First, HR analytics requires data subjected to strict privacy regulations and produces sensitive results. A sincere concern of companies and also labor unions. Secondly, it might be hard to collect the right data. Nevertheless, it is expected that HR analytics will become more popular.

Communicational challenges remain, but the current emphasis on data quality and management is not surprising. Data quality is one way to significantly improve results, especially when other challenges dissolve and companies are collecting as much data as possible, triggered by low data storage costs. Since companies are starting to collect unstructured data, one needs to carefully think about how to store these data and how to adapt data governance rules. Furthermore, the arrival of new cloud storage solutions might decrease the popularity of central data warehouses. With regards to analytics governance, experts are not always on the same page. Decentralized teams are easier to manage and keep data scientists closer to the business. However, this should not necessarily stand in the way of centralized teams where competence can be shared across departments and projects. Remember that Davenport et al. (2010) identify centralized teams as the most mature, although they found decentralization the most prevalent format. One expert even compares the analytics function with the risk management function. The latter nowadays frequently holds a board function in the financial services sector and he predicts the same trend for the analytics function. Experts also notice a trend of sharing expertise in data science communities, e.g. Kaggle, driven by new open source solutions. This might change the modern data science team and how data scientists sharpen their skill set.

Finally, experts talked about the cost of analytics. The relative high share of personnel cost can be related to three phenomenons: (1) an increase in the number of data scientists, (2) a higher use of open source tools and (3) cheaper data storage and sharing solutions. The latter two are not widespread yet which might contribute to the variation in relative share. Furthermore, one needs to take into account when a company started applying analytics because soft- and hardware start-up costs can be large.

6. Conclusion

We studied the current characteristics of analytics from five different perspectives: the data, organizational, leadership, techniques and applications, and analysts perspective. We found that analytics is nowadays more commonly applied, mostly for well-known applications and understandable techniques. However, analytics has still many unexplored opportunities in personalization and social media as well as in HR. Next, more advanced, complex techniques, such as neural networks, remain relatively unexplored although they could offer better performance. Moreover, organizations will need to deal with new challenges. As such, data management and data quality issues are prevalent. Furthermore, it is not always clear how companies can best organize their analytics. Although there seems to be a preference for central coordination, most companies combine multiple formats. Analytics leadership is, however, being organized more and more at a higher level. New board positions, such as a chief analytics officer, are being created. Finally, the analytics team itself is also growing and new skill sets are required to meet new challenges.

We analyzed how these characteristics may form indicators for analytics maturity. We derived four analytics maturity levels from our survey: no analytics, analytics bootstrappers, sustainable analytics adopters and disruptive analytics innovators. These clusters illustrate a staging with regards to different perspectives. One can observe that more mature companies are applying a wider variety of analytics techniques and applications. Furthermore, the level of analytics and data management organization also indicates analytics maturity. In order to improve maturity, companies need to explore new opportunities and focus on analytics coordination to fully leverage each potential.

Funding

[omitted to preserve anonymity]

Vitae

[omitted to preserve anonymity]

References

Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *J. of Mark. Res.*, *14*, 396–402. doi:10.2307/3150783.

Comuzzi, M., & Patel, A. (2016). How organisations leverage big data: A maturity model. *Ind. Manag. & Data Syst.*, *116*. doi:10.1108/IMDS-12-2015-0495.

Cosic, R., Shanks, G., & Maynard, S. (2012). Towards a business analytics capability maturity model. In *Proceedings of the 23rd Australasian Conference on Information Systems (ACIS) 2012* (pp. 1–11). Geelong, Victoria.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Boston, MA: Harvard Business Review Press.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Boston, MA: Harvard Business Review Press.

Davenport, T. H., & Patil, D. (2012). Data scientist: The sexiest job of the 21st century. *Harv. Bus. Rev.*, *90*, 70–76.

DeLine, R. (2015). Research opportunities for the big data era of software engineering. In *Big Data Software Engineering (BIGDSE), 2015 IEEE/ACM 1st International Workshop on* (pp. 26–29). Florence, Italy. doi:10.1109/BIGDSE.2015.13.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biom.*, *27*, 857–871. doi:10.2307/2528823.

Griffin, J., & Davenport, T. (2011). *Organizing analytics: Building an analytical ecosystem for today, tomorrow, and beyond*. Technical Report Deloitte Global Services Limited. http://deloitte.wsj.com/cfo/files/2014/03/Organizing_Analytics.pdf.

Hagerty, J. (2006). AMR Research's business intelligence/performance management maturity model, version 2. *AMR Research. Najdeno*, *25*. http://www.eurim.org.uk/activities/ig/voi/AMR_Researchs_Business_Intelligence.pdf.

HP (2015). *Describing the BI journey: The HPE Business Intelligence (BI) Maturity Model*. Technical Report 4AA3-9723ENW Hewlett Packard Enterprise. https://www.hpe.com/h20195/v2/GetPDF.aspx/4AA3-9723ENW.pdf.

IBM (2009). *Finding your level of BI and performance management maturity: BI and planning for midsize companies - introducing IBM Cognos Express*. Technical

Report IBM. ftp://ftp.software.ibm.com/software/data/sw-library/cognos/pdfs/analystreports/ar_find_your_level_of_bi_and_pm_maturity.pdf.

Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., & Haydock, M. (2012). Analytics: The widening divide. *MIT Sloan Manag. Rev.*, *53*, 1–22.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *Int. J. of Inf. Manag.*, *34*, 387 – 394. doi:10.1016/j.ijinfomgt.2014.02.002.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Manag. Rev.*, *52*, 21–32.

Lessmann, S., Baesens, B., Seow, H., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur. J. of Oper. Res.*, *247*, 124–136. doi:10.1016/j.ejor.2015.05.030.

Muller, L., & Hart, M. (2016). Updating business intelligence and analytics maturity models for new developments. In *Decision Support Systems VI - Addressing Sustainability and Societal Challenges - 2nd International Conference (ICDSST) Proceedings* (pp. 137–151). Plymouth, UK. doi:10.1007/978-3-319-32877-5_11.

Pinsonneault, A., & Kraemer, K. L. (1993). Survey research methodology in management information systems: An assessment. *J. of Manag. Inf. Syst.*, *10*, 75–106. doi:10.1080/07421222.1993.11518001.

Ransbotham, S., Kiron, D., & Prentice, P. K. (2015). Minding the analytics gap. *MIT Sloan Manag. Rev.*, *56*, 63–68.

Reynolds, A., Richards, G., de la Iglesia, B., & Rayward-Smith, V. (2006). Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. *J. of Math. Model. and Algorithms*, *5*, 475–504. doi:10.1007/s10852-005-9022-1.

Rossi, P. E., McCulloch, R. E., & Allenby, G. M. (1996). The value of purchase history data in target marketing. *Mark. Sci.*, *15*, 321–340. doi:10.1287/mksc.15.4.321.

Saxena, R., & Srinivasan, A. (2013). *Business analytics: a practitioner's guide*. New york, NY: Springer-Verlag.

Appendix A. Clustering details

In this paper, clustering was performed on a total of 28 variables indicating the analytics techniques and applications used within the responding companies. Partitioning around medoids (PAM), a non-hierarchical clustering method, was specifically chosen because it provided the best results after we tested agglomerative hierarchical clustering with Ward's distance, k-means clustering and PAM. We note that in comparison to the well-known non-hierarchical k-means clustering method, PAM is less sensitive to outliers (Reynolds et al., 2006). The distance matrix was always calculated using Gower's distance (Gower, 1971). The number of clusters was set to four based on the ratio of the average distance between clusters and the average distance within clusters, see Figure A.1. This ratio is an internal validation measure addressing two criteria of a good clustering solution: compactness, measured in average distance within clusters, and separation, measured in average distance between clusters. Additionally, Table A.1 summarizes whether the characteristics gathered from the survey proved to differ significantly across the clusters.

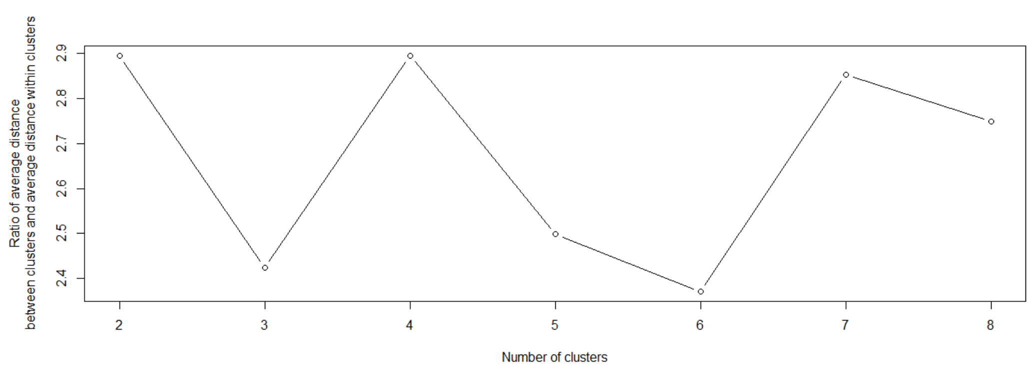


Figure A.1: The optimal number of clusters is determined based on the ratio of average distance between clusters and average distance within clusters. A higher ratio is preferred.

Table A.1: Significance of corporate characteristics across clusters.
*p-value < 0.1, **p-value < 0.05, ***p-value < 0.01, ****p-value < 0.001

Variable	Type of test
Propagation of analytics****	Pearson's Chi-squared test
Data integration from several internal and/or external sources****	Pearson's Chi-squared test
Data cleansing****	Pearson's Chi-squared test
Data aggregation****	Pearson's Chi-squared test
Sampling****	Pearson's Chi-squared test
Outlier detection and treatment****	Pearson's Chi-squared test
Data visualization****	Pearson's Chi-squared test
Hypothesis testing by means of statistical analysis****	Pearson's Chi-squared test
OLAP****	Pearson's Chi-squared test
Association rules****	Pearson's Chi-squared test
Sequence rules****	Pearson's Chi-squared test
Segmentation and clustering****	Pearson's Chi-squared test
Hierarchical clustering****	Pearson's Chi-squared test
K-means clustering****	Pearson's Chi-squared test
Self-organizing maps****	Pearson's Chi-squared test
Classification****	Pearson's Chi-squared test
Regression****	Pearson's Chi-squared test
Linear regression****	Pearson's Chi-squared test
Logistic regression****	Pearson's Chi-squared test
Decision trees****	Pearson's Chi-squared test
Neural networks****	Pearson's Chi-squared test
Support vector machines***	Pearson's Chi-squared test
Survival analysis****	Pearson's Chi-squared test
Time series forecasting****	Pearson's Chi-squared test
Optimization, e.g. integer/linear programming, etc.****	Pearson's Chi-squared test

Number of months performing analytics: cluster 1 vs. 2***	Mann-Whitney U test
Number of months performing analytics: cluster 2 vs. 3	Mann-Whitney U test
Number of months performing analytics: cluster 3 vs. 4****	Mann-Whitney U test
Size of the analytics team	Pairwise Mann-Whitney U Tests
Number of analytics applications: cluster 1 vs. 2. Finance*, marketing**, operations, HR.	Mann-Whitney U Test
Number of analytics applications: cluster 2 vs. 3. Finance**, marketing, operations, HR*.	Mann-Whitney U Test
Number of analytics applications: cluster 3 vs. 4. Finance**, marketing, operations, HR****.	Mann-Whitney U Test
Listed*	Pearson's Chi-squared test
Governmental	Pearson's Chi-squared test
Active offline*	Pearson's Chi-squared test
Active online**	Pearson's Chi-squared test
Globalization level	Pearson's Chi-squared test

Appendix B. How corporate functions impact the number of analytics objectives for finance and marketing: modeled by means of linear regression

Table B.1: Linear regression model for the number of financial analytics objectives with adjusted $R^2 = 0.5415$. The number of objectives ranges from 0 to 6.
*p-value < 0.1, **p-value < 0.05, ***p-value < 0.01, ****p-value < 0.001

Variables	Estimate
Intercept	0.1095
Governmental	1.8133**
Active online	1.6954**
Automotive sector	4.0927**
Consulting sector	3.5206****
Financial services sector	3.2459****
Healthcare sector	1.9666*
Marketing & communication sector	3.9404***
Internationally active	-1.8976***
No application of analytics	-2.5440*

Table B.2: Linear regression model for the number of marketing analytics objectives with adjusted $R^2 = 0.6008$. The number of objectives ranges from 0 to 20.
*p-value < 0.1, **p-value < 0.05, ***p-value < 0.01, ****p-value < 0.001

Variable	Estimate
Intercept	-0.6162
Listed	4.6554***
Active offline	6.3750***
Automotive sector	8.0117*
Consulting sector	-3.7356*
Technology sector	-9.4562***
Globally active	6.5409***
Internationally active	3.2295*

Appendix C. How the size of the analytics team is impacted by corporate characteristics: modeled by means of linear regression

Table C.1: Linear regression model for the size of the analytics team with adjusted $R^2 = 0.4714$. The size of the analytics team was normalized to $[0, 1]$.
*p-value < 0.1, **p-value < 0.05, ***p-value < 0.01, ****p-value < 0.001

Variable	Estimate
Intercept	0.08504*
Active offline	-0.12349**
Number of months applying analytics (normalized to $[0, 1]$)	0.16639*
Size (in number of employees) (normalized to $[0, 1]$)	0.54793****
Analytics culture clearly driven by senior management: strongly disagree	-0.03516
Analytics culture clearly driven by senior management: disagree	-0.04466
Analytics culture clearly driven by senior management: strongly agree	0.07456*