D2.6: Annual Report on Opportunities

<table>
<thead>
<tr>
<th>Workpackage</th>
<th>WP2 – IMPACT: Framing the European Data Economy and maximize Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editor(s):</td>
<td>Sonja Zillner (Siemens AG)</td>
</tr>
<tr>
<td>Responsible Partner:</td>
<td>Siemens AG</td>
</tr>
<tr>
<td>Contributors</td>
<td>Sonja Zillner (Siemens AG)</td>
</tr>
</tbody>
</table>
| Internal Reviewer | Nuria de Lama, ATOS  
                  | Andreas Metzger, UDE                                               |
| Status-Version: | Final                                                              |
| Due to       | 31/03/2019                                                          |
| Submission Date: | 30/04/2019                                                          |
| EC Distribution: | Public                                                             |

Abstract:
To stimulate and foster the growth of the European data economy, the continuous identification of data-driven opportunities is a critical task. In this deliverable we introduce the design and outcome of the DemoX research study investigating business opportunities of data-driven start-ups. To complement the findings of the study, the characteristics of data-driven business opportunities are described along successful start-ups by using the DemoX categories that are building upon and extending the Damian method. In addition, the evaluation of using the DemoX approach in class room and workshop settings is summarized.
## Contents

1 INTRODUCTION .................................................................................................................. 6

2 DATA DRIVEN INNOVATION MODELLING APPROACH ........................................ 6

   2.1 WHAT ARE BUSINESS OPPORTUNITIES? ................................................................. 7

   2.2 CHARACTERISTICS OF DATA-DRIVEN INNOVATION ........................................ 7

   2.3 HOW TO SCREEN DATA-DRIVEN BUSINESS OPPORTUNITIES? .......................... 7

3 DEMOX MODEL BUILDING ................................................................................................ 8

   3.1 LITERATURE REVIEW ............................................................................................... 9

   3.2 INITIAL DEMOX MODEL BUILDING ......................................................................... 9

   3.3 DATA COLLECTION AND CODING ........................................................................... 10

       3.3.1 Selection Criteria ............................................................................................... 10

       3.3.1.1 Focus on start-ups ......................................................................................... 10

       3.3.1.2 Success criteria ............................................................................................. 10

       3.3.1.3 Technology focus ......................................................................................... 11

       3.3.2 Databases .......................................................................................................... 11

       3.3.2.1 Comparison of Crunchbase and F6S ............................................................ 11

       3.3.2.2 Strategies to ensure high Data Quality ......................................................... 13

       3.3.3 Generation of Sample set .................................................................................. 14

       3.3.4 Coding of Data .................................................................................................. 16

           3.3.4.1 Coding Test Run ....................................................................................... 16

           3.3.4.2 Coding Process ......................................................................................... 17

   3.4 DATA AND PATTERN ANALYSIS .............................................................................. 18

       3.4.1 Message 1: Majority of start-ups address B2B markets .................................... 18

       3.4.2 Message 2: 75% of companies have a clear sector focus ............................... 20

       3.4.3 Message 3: Majority of start-ups harness network effects ............................. 21

       3.4.4 Message 4: Data-driven innovation relies on multiple revenue models ....... 26

       3.4.5 Main Learnings from Empirical Study ............................................................. 29

4 EXAMPLE OF SUCCESS STORIES ............................................................................... 30

   4.1 ARTOMATIX............................................................................................................... 30

   4.2 SELECTIONIST ......................................................................................................... 31

   4.3 HOOCH ...................................................................................................................... 33

   4.4 ARABLE .................................................................................................................... 34

5 EVALUATION ....................................................................................................................... 35

   5.1 DATA-DRIVEN INNOVATION COURSE AT UNIVERSITY ................................... 36

   5.2 DATA-DRIVEN INNOVATION WORKSHOPS FOR PROFESSIONALS ............ 40

6 CONCLUSIONS AND NEXT STEPS ............................................................................. 42

7 REFERENCES ..................................................................................................................... 43
List of Figures

FIGURE 1 – THE FOUR PHASES OF THE DEMOX STUDY ........................................ 8
FIGURE 2 OVERVIEW OF THE DEMOX MODEL .................................................... 9
FIGURE 3 SUMMARY OF THE COMPARISON BETWEEN CRUNCHBASE AND F6S ........ 13
FIGURE 4 OVERVIEW OF 46 CATEGORY GROUPS IN CRUNCHBASE ......................... 15
FIGURE 5 OVERVIEW OF THE GENERATION OF SAMPLE SET ............................. 15
FIGURE 6 OVERVIEW OF DATA COLLECTION, CODING AND ANALYSIS .............. 16
FIGURE 7 MAJORITY OF DATA-DRIVEN START-UPS ADDRESS B2B MARKETS ........ 19
FIGURE 8 THREE OF FOUR START-UPS HAVE A CLEAR SECTOR FOCUS ............... 20
FIGURE 9 SCREENSHOT OF INSIGHT PLATFORM .............................................. 23
FIGURE 10: 57% OF START-UPS HARNESS NETWORK EFFECTS ....................... 26
FIGURE 11 REVENUE MODEL OF BIRD.I (FREEMIUM MODEL NOT EXPLICITLY DESCRIBED) .... 28
FIGURE 12 OVERVIEW OF REVENUE MODELS USED FOR DATA-DRIVEN BUSINESSES .......... 29
FIGURE 15 ONE-PAGER ARTOMATIX ............................................................. 31
FIGURE 16 ONE-PAGER OF SELECTIONIST ..................................................... 32
FIGURE 17 ONE-PAGER OF HOOCH ............................................................... 34
FIGURE 18 ONE-PAGER OF ARABLE ............................................................ 35
FIGURE 19 RELEVANCE OF COURSE CONTENT ............................................. 36
FIGURE 20 RELEVANCE OF METHODS ......................................................... 37
FIGURE 21 IMPORTANCE OF SET UP AND INTERACTIVE ELEMENTS .................... 38
FIGURE 22 ITERATION OF DEMOX METHODS ............................................. 39

List of Tables

TABLE 1: DEFINITIONS, ACRONYMS AND ABBREVIATIONS .................................. 4
TABLE 2 QUANTITATIVE EVALUATION OF DATA-DRIVEN INNOVATION WORKSHOP .... 41
# Definitions, Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDV</td>
<td>Big Data Value</td>
</tr>
<tr>
<td>BDVe</td>
<td>Big Data Value ecosystem</td>
</tr>
<tr>
<td>BDVA</td>
<td>Big Data Value Association</td>
</tr>
<tr>
<td>DemoX</td>
<td><strong>Data-driven Ecosystem</strong> <strong>Modeling Approach</strong></td>
</tr>
</tbody>
</table>

*Table 1: Definitions, Acronyms and Abbreviations*
Executive Summary

This deliverable addresses the question of how data-driven business opportunities can be screened in an effective manner. To support the process of identifying and scoping data-driven business opportunities that are impacted by the dynamics of supply and demand trends, we introduced the DemoX approach in Deliverable 2.5. The DemoX approach guides the process of analysing demand and supply dynamics of data-driven innovations by investigating the co-evolution and interactions between the scope of the offering (supply) and the context of the market (demand).

This deliverable is the second of a series of reports that will be updated once a year. In this version, we are documenting the results and feedback we received while applying the DemoX model in a wider range of settings and contexts (verticals, horizontals, mixed): For instance, we accomplished a quantitative study of a representative set of data-driven start-ups to identify promising patterns to guide future investment decision. The results of this study provided valuable guidance in fine-tuning and updating the DemoX model. In addition, we evaluated the DemoX model with a university course and in a series of in Business Model style workshops.

By relying on the DemoX model, we have a) a method in place that we can share with members of the BDV ecosystem for exploring data-driven business opportunities as well as b) comprehensive content that can be used for industrial trainings and university courses. The content and guiding questions of DemoX are shaped and complemented by the analysis of best practices of successful data-driven start-ups. In the future, we plan to engage with the stakeholders of the BDV ecosystem, with focus on SMEs and start-ups, to help them scope promising business opportunities.
1 Introduction

In this series of reports, we are investigating emerging business opportunities in the European Big Data Landscape with the goal to promote their uptake in Europe. Based on the DemoX framework developed in the first year of the project (and documented in the first report of this series (Zillner, Timan and Kotterink, 2018)), we conducted a quantitative study of a well-selected set of data-driven start-ups to identify promising patterns to guide future investment decisions. In addition, we evaluated the usage of DemoX model in practical settings within a university course and in a series of Business Model style workshops. Both, the results of the study as well as the experiences in using the model in industrial and academic trainings, provide valuable guidance for further fine-tuning and updating the DemoX model.

By relying on DemoX, we have now a method in place that we can share with members of the BDV ecosystem for exploring data-driven business opportunities. The DemoX model complemented with content slides, comprehensive set of methods and guiding questions is used for industrial trainings and university lectures. The derived characteristics and patterns of successful start-ups help entrepreneurs, innovators, and managers to scope their data-driven business opportunities in a way that industrial investment decisions are becoming more likely.

With the DemoX approach we plan to engage in the future with entre- and intrapreneurs, SMEs and start-ups to help them scope promising business opportunities. The main impact of the work described in this series of reports is to provide a continuous assessment of emerging, data-driven business opportunities by designing a tool that can be offered to EU and national programmes addressing start-ups and SMEs that would be interested to validate or further explore the data component of their value propositions. The data collected through the DemoX research study as well as applications in different contexts provide the intelligence to start monitoring and mapping an evolving landscape of European Data Value Propositions, following accepted and effective business modelling classifications.

The first part of the document (Section 2) provides reflections about how data-driven innovations are modelled. Section 3 gives an overview of our empirical study for testing the DemoX model study. Section 4 complements the DemoX study by pitching success data-driven business opportunities along the DemoX categories. In Section 5 we report about our experiences as well as qualitative and quantitative evaluation of using the DemoX model class-room settings in academia and workshop settings in industry. Section 0 concludes this document with an outlook to our future work.

2 Data Driven Innovation Modelling Approach

Finding ways and methods for the identification and scoping of data-driven business opportunities requires an understanding of the business opportunities in general as well as of the characteristics of data-driven opportunities.
As both topics have been discussed in detailed in Deliverable 2.5., we only provide a short summary to facilitate readability of the overall document.

2.1 What are business opportunities?

In general, the concept of business opportunity is a very broad concept which is used to describe the chance to address a particular market need through the creative combination of resources that allow delivering advanced value propositions (Ardichvili et al., 2003). In this way, the definition of promising business opportunities relies on the balancing of – often mainly technical – capabilities on the supply side, with the user needs and interests as well as market dynamics shaping the demand side. In addition, studies indicate that most successful entrepreneurs and investors continuously observe the demand side very carefully in order to understand what customers and marketplaces want and never lose track of it (Timmons and Spinelli, 2007). The knowledge reflecting the demand side is used to guide the scoping of offering by combining own innovative technology components with reusable and available assets from others in a way that fosters competitiveness. In addition, the development of business opportunities is described as continuous process that involves proactive efforts to explore all essential steps of a new business.

2.2 Characteristics of data-driven innovation

Data-driven innovation refers to the use of data and analytics to improve and foster new products and processes, new organizational processes and new markets and business models (OECD, 2015). The economics of data has strong impact on the development of data-driven business opportunities. For instance, data can be consumed an unlimited number of times without losing its value, data can be reused as input for the production of different goods and services and -- on the other hand -- the value of data still depends on complementary assets related to the capability of extracting information out of the data. Those fundamental economic properties of data recommend treating data as an infrastructural resource (for more details see Deliverable 2.5).

Given the mentioned economic properties, disruptions through data are becoming more likely. In particular, due to the network effects as well as the simplicity on how a variety of offerings with different value/price tags can be brought to the market, the success of data-driven innovation requires continuous alignment between needs on the demand side with the opportunities on the supply side.

2.3 How to screen data-driven business opportunities?

The data economy in general is a highly dynamic market. This is supported by the rapid growth of the European data markets, as well as recent technical breakthroughs that were made possible by the availability of large volumes of data, such as the Jeopardy demo by IBM Watson or Google Now or Siri. In addition, experts continue to highlight the wide range of commercial opportunities that can be realized by using the technology available today.
Entrepreneurs bring new offerings to the market that are confronted with this highly dynamic market and require continuous scanning of the market’s offerings to identify promising available technology component that can be reused to speed up development time of their innovation. At the same time, entrepreneurs need to continuously investigate their own unique selling point and competitiveness of their offering in a very dynamic environment. To stay competitive in this fast-moving market, entrepreneurs need to continuously reassess what is part of their core offering and in which areas they are partnering with others.

The high growth scenario\(^1\) in the comprehensive European Data market study (IDC and OpenEvidece, 2017) is based on *supply-demand dynamics that shifts from technology-push to demand pull*. In other words, any means that provide guidance in match-making between markets needs on the demand side with technical capabilities on the supply side helps to stimulate the development of data driven innovation and in consequence the growth of the European data market.

To summarize, data-driven business opportunities should be described with a clear scope of offering per market segment (*supply side*) as well as reflect the ecosystem dynamics and benefits of network effects (*demand side*).

Within the next section, we will introduce the DemoX model which guides innovators to systematically explore and analyse the supply and demand side of data-driven innovations by incorporating the particularities of data.

### 3 DemoX Model Building

We conducted an empirical study to build and test our DemoX model. The study can be divided in four phases (see Figure 1).

![Figure 1 – The four phases of the DemoX Study](image)

By first *reviewing the literature* on existing proven methods and the theoretical concepts for scoping data-driven business opportunities we could identify relevant aspects of the data driven business opportunities. The learnings of the literature review guided us in developing a *conceptual model* in the form of an ontology describing the aspects of supply and demand in data-driven ecosystems. Based on the

\(^1\) which estimated 4% of GDP growth between 2016 and 2020
D2.6: Annual Report on Opportunities

A conceptual model, data from a representative sample of data-driven start-ups could be collected and coded. Finally, the data was *analysed*, and best-practice *patterns* identified.

Please note, that the first two phases already have been covered in detail in Deliverable D2.5. (Zillner, Timan and Kotterink, 2018). We cover them very briefly in this report to ensure completeness and readability.

3.1 Literature Review

Within the literature review we aimed to identify existing business modelling methodologies that we could partially reuse or combine to not reinvent the wheel as well as to rely on established and well proven best practices. In addition, we investigated literature that provided us guidance in the meta-analysis of demand and supply aspects being very central for data-driven innovation. The detailed literature review is covered in Deliverable 2.5. ((Zillner, Timan and Kotterink, 2018)).

3.2 Initial DemoX Model Building

Based on the above-mentioned literature review, the dimension of data-driven business opportunities could be identified. The developed conceptual model in the form of an ontology aims to provide a possible set of categories and concepts to describe data-driven business opportunities in a comprehensive manner. By capturing all relevant aspects of data-driven business opportunities, the DemoX model (see Figure 2) provides means for analysing the above aspects in an effective way. Its categories are divided into supply side and demand side aspects.

![Figure 2 Overview of the DemoX model](image)
On the supply side the focus is on the development of new offerings. For a clearly defined value proposition, this includes the identification of and access to required data sources, as well as the analysis of underlying technologies. On the demand side the focus is on the dynamics of the addressed markets and associated ecosystems. The analysis includes the development of a revenue strategy, a way forward of how to harness network effects as well as an understanding of the type of business. As data-driven innovations are never done in isolation, the identification and analysis of potential development partners as well as partners in the ecosystem help to align / balance the supply and demand aspects in a way that its competitive nature will stand out.

Note that the DemoX model presented in this deliverable already incorporates the lessons learned and insights we gained by performing a coding test run on a smaller set of 20 start-ups (see Section 3.3.4.1).

3.3 Data Collection and Coding

In order to assess the theoretical findings and conceptualisation we assessed the DemoX model along a large set of successful data-driven start-ups. This empirical data helped us to improve and fine-tune the DemoX model.

3.3.1 Selection Criteria

3.3.1.1 Focus on start-ups

The objective of our research study is to systematically analyse and compare successfully implemented data-driven business opportunities. Although the implementation of data-driven business opportunities is not restricted to a certain type of business or organisation, we decided in this study to focus on start-ups only due to two reasons: First, due to the lack of available public information as larger corporates and SMEs barely share information about their business or innovation designs and decisions; and second, due to high interdependence with existing operations. Innovation activities in large corporates and SMEs are often constrained by existing infrastructures, legacy systems or prior systems. Thus, any data-driven business opportunity must consider (and co-exist with) established software, IT systems, processes, support structures, and existing customer bases. In this way, any data-driven innovation design is likely to trigger some organisational implications, such as changes in the sales channels, customer bases, migration issues, pricing models and processes, customer expectations, etc. Those interdependencies with existing operations make it difficult to analyze data-driven business opportunities in isolation or to derive generic patterns.

3.3.1.2 Success criteria

For identifying successful data-driven start-ups, we needed to define a measurement for success. We decided to choose start-ups with funding between US$2M and US$10M to obtain the ones

- that had already convinced some ventures to invest into them;
that would already have their product validated;
- but that would still be a young start-up.

As we have a very specific focus for this research, which is identifying what are the data-driven business opportunities from today, it made a lot of sense limiting our data. The decision criteria for the values (between two and ten million dollars) was made in the light of venture capital theory. Although there is no consensus regarding the exact amount of money that determines each stage, we decided to follow the criteria used by Crunchbase: Angel is the first round, normally financed with less than US$10,000. The following stage is Seed, ranging from US$10,000 to US$2M. Then there are the venture rounds, that could have many series (A–Z), with the A and B series normally valued between US$1M and US$20M. It is worth to mention that we could have chosen as a criterion just A/B series, but as this could vary in terms of interpretation among start-ups, choosing their funding amount as selection / success criteria made more sense.

3.3.1.3 Technology focus
To identify data-driven start-ups, keywords / selection criteria such as Data analytics or Artificial Intelligence seem to be promising (more details on this will be covered in Section 3.3.3)

3.3.2 Databases
There exist two websites in the venture capital field that we identified as promising candidates for accessing a representative sample set of data-driven start-ups:
- Crunchbase is an American based platform for finding business information about private and public companies. The information collected includes investment and funding information, mergers and acquisition as well as a list of recent news (such as press reports) or activities of the company.
- F6S is the largest platform for founders based in Europe. Its objective is to enable founders to interact with accelerators/incubators, products/tools as well as talents on the platform to grow together.

3.3.2.1 Comparison of Crunchbase and F6S
To ensure efficient data access and high data quality, we analysed and compared both platforms. Two aspects have been in the focus of our investigations.
1) Is the information provided credible?
2) How easy is it for us to access the start-up data?

Credibility
- Crunchbase: With its main goal to collect and build a database of start-ups, Crunchbase has a user-friendly layout and wide range of data / information to be explored. In the platform, anyone can edit any Crunchbase profile, as long as it is not a big account or it is not required by the owner to be private. This allows the platform to grow fast and present any company, at variate financial stages. The platform has a team that continuously moderate editions, and only they can delete a profile, funding round, acquisition and founder information. Whenever there is a suspicion of fake profiles, users are requested to send an
e-mail to Crunchbase to initiate its review. The platform also states that over 31 millions of users use their services.

- F6S: F6S has as main goal building a platform to connect start-up and investors, as an accelerator. In terms of editing the data on the website, only the profiles' creators can edit it later, as well as delete it. There is no information on the website about dealing with fake profiles. Thus, there are currently 800 thousand profiles of start-ups in their database, more than 1 million technological founders and 10,000 start-up programs globally.

Access:

- Crunchbase: Crunchbase has a very easy and downloadable approach to extract the data when the Pro Version is selected. Crunchbase allows to select categories / filters to select the samples set of relevance. It is possible to choose diverse filters at the same time and download them as an .xsl file. This is of high relevance for the DemoX research study due to time limitation and avoidance of mistakes.

- F6S: although the platform is completely free, there is no possibility of downloading the data. Doing it manually (copy and paste) is a very time consuming and not accurate option. Besides that, it is hard to select filters and there is no information available about the criteria used in the website to classify information.
3.3.2.2 Strategies to ensure high Data Quality

To ensure high data quality, we required to consolidate a sample set of start-ups that are rich in information and well-documented allowing us to find statements in relation to the DemoX model.
Based on a sample of 60 start-ups that were manually selected\(^2\) from the F6S website by using the filter Data & Analytics, we could investigate the quality of information provided by data-driven companies listed on F6S platform only versus data-driven companies listed on both platforms. Our main objective was to investigate in which quality the start-ups were represented, i.e. to which extent we could find relevant statements and information required for coding the DemoX model. Our learnings can be summarized as follows:

- 42% of the companies that were selected on F6S are also on Crunchbase (25 out of the sample 60 start-ups)
- If the company is both on Crunchbase and F6S, the business seems to be more real
- In average every second company out of the sample that was not listed on Crunchbase looked weird in a way that there was no website or only little or outdated information available

Following the above observation, we decided to cross the data between the two platforms to ensure high data quality.

In addition, we could benefit from the convenient way of accessing the start-up sample provided by Crunchbase by keeping at the same time to interact / communicate with the start-ups via the F6S platform (a quite promising opportunity that we are currently discussing within the BDVe consortium).

### 3.3.3 Generation of Sample set

Crunchbase has 46 categories (see Figure 4) for classifying the categories of companies. Some examples are Software, Financial Services, Information Technology, Sales and Marketing and Gaming. Although Crunchbase also has more detailed categories, for example, in January 2018 there were 694 category types, we decided to not use this filter as we were not interested in limiting our area, but rather having access to different companies in the fields of Data Analytics and Artificial Intelligence. Therefore, we decided to rely on the category \textit{Data & Analytics} and \textit{Artificial Intelligence} as selection criteria for generating the sample size.

\(^2\) In F6S we did not have access to the group criteria of the platform. By setting the filter data & analytics on the F6S website, we received 14876 results. As we did not have access to the group criteria of the platform, there was no transparent way for us to shrink this large and probably too broad sample set to a suitable and specific sample set.
Figure 4 Overview of 46 Category Groups in Crunchbase

The start-up data was extracted on the 16th of January 2018 from Crunchbase using the aforementioned filters:

- Categories “Data Analytics” and “Artificial Intelligence”
- Funding between US$2M and US$10M

Based on these filters, we could extract a sample set of 2161 data-driven companies. For each start-up we could automatically extract the following data entries from Crunchbase: 1) Start-up name; 2) Organization name URL; 3) Category Group (Data & Analytics; Artificial Intelligence); 4) Headquarters location; 5) Description; 6) Website; 7) CB Ranking (Company); 8) Number of funding rounds; 9) Total funding amount (between US$2M and US$10M).

Figure 5 Overview of the Generation of Sample Set

From this larger sample set, we extracted a statistically valid sample set of start-ups with entries in both platforms. However, as the selection of start-ups with entries in both platforms could only be done manually, we decided to rely on a manual analysis of a subset of our sample to derive some estimates of the distribution that can be used for extrapolation. From reviewing a subset of 480 start-ups from the initial sample set, only 193 (40%) have been on both platforms, and approximately 20% of the start-ups being listed on both websites still had very little information.

From those numbers, we can derive a population size of 864 (40% of 2161) start-ups. Our aim was to generate a statistically valid sample set with 95% confidence level and
D2.6: Annual Report on Opportunities

10% margin of error (confidence interval) leading to an ideal sample number\(^3\) of at least 87 start-ups. To fulfil those criteria, we decided to analyse in sum 90 start-ups. For handling the issue of badly documented start-ups, we generated a reserve list of start-ups that could be used in case of sparse information.

In concrete terms, we selected from the initial sample set randomly a smaller sample set of 300 start-ups. The random sampling was done via excel by inserting an additional column and assigning a random number that could then be used for ordering entries from greatest to smallest. From this sample set we manually identified the first 90 data-driven start-ups with entries on both platforms as our sample set.

3.3.4 Coding of Data

The start-up data was coded in accordance to the categories of DemoX model. For each start-up relevant background information was manually searched and investigated to identify relevant statement(s) related to certain category of the DemoX model. Figure 6 provides an overview of the data collection, coding and analysis process.

![Figure 6 Overview of data collection, coding and analysis](image)

3.3.4.1 Coding Test Run

To ensure reliability, the different categories of the DemoX model were defined before the coding exercise started (as explained in Section 3.2). To avoid coding errors, we performed a test run of the coding exercise based on a manually selected sample of 20 start-ups. After coding this initial set of start-ups by two independent coders, all categories or concept with high percentage of disagreement in coding were discussed in detail and then re-defined or removed.

**Re-defined categories:** For instance, initial definitions of the two concepts *data* and *software / analytics* related to offerings from the European data market study (IDC and OpenEvidece, 2017) lead to 30% of coding errors as offerings encompassing activities related to the pre-processing of data could often be mapped to both

\(^3\) Calculation was done via https://www.qualtrics.com/blog/calculating-sample-size/
categories. To overcome the unclear semantics, we decided to classify an offering only of type data in cases it is handled as intermediary results, i.e. the (collected) data needs to be made available in some type of plug- and play mechanism ensuring seamless access to the data. In this way, the offering of the company Scanalytics -- who is putting a lot of efforts in collecting behavioural data to feeding their algorithm and but does not provide an API or service for accessing the collected data assets – was classified as software / analytics.

**Removed categories:** For the categories related to the framework conditions, we observed 30 % coding errors by four of the five concepts. The main reason for this was that information about related framework conditions could not be found in the internet as this is an information which is explicitly discussed, announced or published. Thus, the coding result was derived by interpreting other categories, for example, open standards could be derived by the fact that semantic data have been used. As this led to the situation that the framework condition categories themselves did not convey any independent new information, we decided to remove them from the DemoX model.

### 3.3.4.2 Coding Process

The start-ups from the sample set were coded by three independent coders. For each start-up the three coders manually annotated a binary feature vector covering all DemoX categories and concepts. In case a specific feature was present it was annotated with “1”, in cases it was not present with “0” and in cases no information could be found it was indicated with “2”. This was done by searching on the internet for relevant statements indicating a specific feature of the DemoX model. The three most important websites to look for start-ups’ background information were the Crunchbase website, F6S website as well as the website of the company. The collection of “recent news and activities” on the Crunchbase website helped us to access related documentation very easily.

In cases the three main sources of information did not help us to label all categories of the DemoX model, we relied on google search and /or the following websites:

- www.techcrunch.com
- www.financeonline.com
- www.facebook.com
- www.businesswire.com
- www.nasdaq.com
- www.quora.com
- www.businessinsider.com
- techseen.com

---

4 Although the feature vector can be annotated with three values (0, 1, 2), we still treat it as binary feature vector, as the third value category “2” was only introduced for practical reasons, to indicate that for a specific feature the accomplished search did not reveal any related information. This helped us to monitor the progress of the coding exercise as well as to remove start-ups from the analysis.
For each start-up at least 3 websites (Crunchbase, F6S and company website) have been consulted. Very often additional webpages, e.g. linked press releases, have been analysed and complementary internet searches have been accomplished to ensure that all categories and concepts could be addressed.

There have been situations in which no statements to certain categories could be found. For instance, in cases when start-ups addressing emerging markets, they often do not provide any information about their revenue strategy. For instance, the start-up Iris Automation which provides robust collision avoidance systems for industrial drones is one of those examples. As the market of industrial drones is still in a very early stage, clear customer needs and opportunities for revenue generation do not yet exist. Thus, the start-up is primarily concentrating on advances in their technology development, their ecosystem partners and exit opportunities.

After having performed the manual annotations, the coder met online to compare coding results and to discuss and resolve disagreements. The result of the coding process was 90 binary feature vectors representing the presence or absence of each DemoX category or concept for each start-up.

3.4 Data and Pattern Analysis

Based on our data coding exercise we derive insights on how successful data-driven start-ups are establishing themselves on the market.

In general, this analysis is only a snapshot as start-ups in general adapt smoothly to the needs of the market and may come up with new offerings in case they observe a clear need. Whenever we observed transformations in the way the offering was positioned in the market, we tried to highlight this more explicitly.

3.4.1 Message 1: Majority of start-ups address B2B markets

A first observation is that most start-ups address B2B markets. For instance, the company Artomatix\(^5\) addresses the digital game business and provides support for 3D artist (professionals) in generating realistic 3D art of textures and texturing. Done manually, the generation of 3D art of textures and texturing is quite time consuming and tiring. By providing automatic support in the seamless generation of non-repetitive high-quality textures to build 3D worlds, characters and items, helps 3D artists to accomplish this task 10 times faster. The business customers of Artomatix are major studies or individual developers being addressed with different revenue models.

In our sample only two out of 90 start-ups focused on end-customer market solely. For example, the start-up UnaliWear\(^6\) who is building and selling smart watches for supporting elderly people in their daily activities, for instance in finding the right way or by reminding them about medications.

\(^5\) https://artomatix.com/
\(^6\) https://www.unaliwear.com/
This very low number of pure B2C start-ups might be surprising at first glance as in general start-ups are attributed to be very close to their customers as well as being highly skilled in exploring unknown user needs guided by an emphatic mindset. Is this a rumor? On a closer look, it becomes clear that start-ups accomplish the exercise of identifying user needs with many different target user groups ranging from end customers, such as individuals or households, to users in business settings, such as professionals with clear roles and tasks, to stakeholders in the sector. However, for delivering the offering to the users, start-ups seem to prefer already established channels. For instance, by integrating data-driven solutions as plug-in in established software applications the access to large customer bases can be realized. This specific strategy was for instance selected by Artomatix by giving 3D game developers access to their offering via Photoshop plug-in.

A second frequent option to deploy data-driven solutions is multi-sided market combining with freemium offerings for private users with complementary business needs. Approximately 19% of start-ups address in such a way both B2B and B2C markets by building upon multi-sided business models or digital marketplaces. An example for multi-sided business models is the data analytics platform Verv of the company Greenrunning which allows collating and analyzing real-time electricity data to generate a range of data services sourcing energy, appliances and customer analysis. End-customers can use this service to analyse the ‘energy signature’ of their own electrical appliances which helps them to explore the costs of each appliance on a real-time basis. By giving usage permission of one owns data for analytical purposes, the service is free for end-customer and can be accessed via a mobile app. This allows Greenrunning to build a unique data set as basis for offering appliance usage data and personalised insights of interest to a wide range of industries including utilities, retail, home care and insurance. Those monitoring services are offered under licence agreements to business customers.

To summarize, we can state that start-ups are very good in exploring wide range of user needs. However, we also observe that in their value capture strategy they tend to rely on established business partners to bring their offering to users. This is caused on the one side by the limited willingness of end-customer to pay (money) for an offering. With end-customers being willing to pay only very small amount of money, companies need to build large scale customer bases (long tail phenomena) which again implies a lot of effort and investments for connecting and gaining customers. A
popular shortcut for such situations is to partner or collaborate with large business players that have access to the needed customer base. The second way forward is the design of multi-sided markets that surpass the limited willingness of end-customers to pay money by introducing new currencies, such as the sharing of data.

3.4.2 Message 2: 75% of companies have a clear sector focus

75% of our start-up sample have developed a clear sector focus. Companies with clear sector-focus have a concrete customer segment(s) in mind for whom a concrete value proposition is delivered. For example, CloudMedx\(^7\) Inc. designs artificial intelligence driven software for medical analytics. Clinical partners at all levels can derive meaningful and real-time insights from their data and intervene at critical junctures of patient care. Its underlying Clinical AI Computing platform uses healthcare specific NLP and Machine learning to generate real-time clinical insights at all points of care to improve patient outcomes. By relying on evidence-based algorithms and deep learning a wide variety of structured and unstructured data being stored in clinical workflows can be understood and used for decision making.

![Figure 8 Three of four start-ups have a clear sector focus](image)

Often, due to the particularity of the data the sector focus is obvious. For instance, data intelligence solution in the medical sector are very customized solution addressing the challenge of preprocessing and annotating the wide range of heterogeneous data sources in clinical and medical settings.

In comparison, we also find start-ups that focus on technology with cross-domain impact. In general, their solution will be used by other intra- or entrepreneurs to build data-driven solutions for end-user. For instance, the start-up DGraph Labs\(^8\) is offering an open source distributed graph data base. The company is planning to release an

\(^7\) [http://www.cloudmedxhealth.com/](http://www.cloudmedxhealth.com/)
\(^8\) [https://dgraph.io/](https://dgraph.io/)
enterprise version that is closed source as well as a hosted version (as it is easier to run hosted services for customers than trying to help them debug every issue on their own). Customers are using the service to build their own sector specific applications. Summarizing, sector-specific data-driven offering are much more frequent than technology-driven solutions. This is caused by the very different pre-processing challenges of the data sources in the various sectors as well as the higher possibilities to identify target groups in concrete sector settings. Most sector-agnostic offerings are intermediate functionalities addressing developers to build customized solutions.

3.4.3 Message 3: Majority of start-ups harness network effects

For digital and data-driven innovations, network effects are important phenomena to be reflected. A network effect occurs when a product or a service becomes more valuable to its users as more people use it (Shapiro and Varian, 1999). Network effects are also known as demand-side economics of scale and predominately exists in areas where networks are of importance, such as online social networks or online dating sites. A social network or dating site is more appealing to its user, the more users of interest it has. In consequence, harnessing network effects require developing a wider network of users in order to differentiate from competitors. For that reason, the critical mass of user and timing are key success factors in a network economy. Due to the high impact of the network, competitors starting from ‘ground zero’ with no users in their network will face difficulties to enter the market successfully. In this discussion we are using the expression “network effect” to highlight the positive feedback (positive network externality\(^9\)) i.e. the phenomena that already existing strength or weaknesses are reinforced, might lead to extreme outcomes. In the most extreme form, positive feedback can lead to a winner-takes-it-call market (e.g. Google).

As network effects impact the underlying economics and operation of data-driven innovation. Instead of producing products that are early on the market and differentiate from other offerings, the focus is now on scaling and scoping the demand perspective. Understanding network effects and its underlying market dynamics is crucial to position data-driven products, services and businesses well in the market. For doing so, data-driven-innovation can harness network effects on three different levels.

**On data level:** Data-driven businesses that can improve their offering the more data is available are relying on network effects on data level. Typical examples are navigation systems or recommendation engines which both become better the more — in this case behavioural data from users — can be collected and fed into the algorithm to produce more accurate traffic information or to better relate similar products with each other. In case data-driven offerings are based on network effects on data level,

---

\(^9\) For completeness we want also to mention the phenomena of negative network externalities which occurs when more users make a product less valuable (e.g. traffic congestion). Negative network effects are also referred as “congestion”.
one needs to develop strategies of how to engage with customers, users or stakeholder to get access to the required data assets. In general, we can distinguish two strategies:

- **Increasing scale of data:** In cases when the outcome of the underlying algorithm can be improved by accumulating more data assets, the strategy is to attract more users in a way that more data can be collected. For understanding how much data is needed, a detailed look at the algorithm is required. Some algorithm cannot be further improved after reaching a sufficiently large amount of data while other algorithms keep on improving continuously the more data is available without limit. In case of the former, the focus is to establish a critical amount of data traffic to ensure that the most accurate analytical model can be built. In the second case the objective of increasing the user base for building larger data assets will happen continuously. In both cases, the starting phase is the trickiest one, as one needs to attract users with a non-mature offering. In this phase a clear understanding of users’ needs and interests will help building a promising strategy to overcome the classical chicken-egg problem. With increasing numbers of users, the efforts in attracting users will decrease and in some cases disappear.

- **Increasing scope of data:** The outcome of some algorithm can be improved the more diverging aspects / perspectives the incoming data set is representing. For instance, linking data from different sources establishes the basis to generate “super additive” insights (big brother phenomena) as more aspects of a subject of interest are reflected. This is also known as process of data contextualization allowing to significantly increase the quality of input data leading to better and more transparent outcomes. In such cases, the underlying question of how to increase the range of collected data from the already existing user bases needs to be investigated. In times when data regulations have not been so strict as today in Europe, this was a very impactful strategy of the GAFA (Google, Apple, Facebook, Amazon) organizations as they based their customer analytical insights on data from different types of services.

**Example**

The platform Insights (see Figure 9) of the company Apptopia uses Big data technology to collect, measure, analyze and provide user engagement statistics for mobile apps.

---

10 [https://apptopia.com/](https://apptopia.com/)
D2.6: Annual Report on Opportunities

The more app provider are producing data being connected to the platform, the more valuable the service gets. In order to get more real-time data, they attract app developers to connect to their platform by providing free data analytics products. With this free of charge value proposition, developers benefit in registering their mobile apps to the platform while giving the platform the permission to analyze user engagement data of the mobile app. High-priced subscription fee model for business customer, including Google, Pinterest, Facebook, NBC Universal, Deloitte, and others, benefiting from real-time engagement insights of mobile apps complements the revenue strategy of this offering.

In this context, multi-sided business models are the usual way forward. Per definition, a multi-sided business model brings together two or more distinct but interdependent groups of customers. Value is only created if all groups are attracted and addressed simultaneously. The intermediary, in our example the company Apptopia, generates value by facilitating interactions between the different customer groups, whereas the value increases when more users are attracted. The more app developer register on the platform the more accurate the statistics become and with an increasing number of business customer, Apptopia has the required resources to invest into advanced functionalities for app developers.

In our study every second start-up relied on network effects on data-level. This highlights the high importance of access to data sources in the context of data-driven innovations. For that reason, we see a sustainable strategy to get access to data as important success lever for data-driven innovations.

**On infrastructure level:** Data-driven businesses that harness network effects on the infrastructure level provide a technical foundation for others, i.e. third-party companies, to build upon. Based on a layer of common components, third-party players are invited to develop and produce an increasing number of data-driven offerings. This set-up is also known as product platforms (Hagel *et al.*, 2015), a prominent example is the android platforms that provides the technical foundation for others to build apps. This includes any types of tools and services that enable the plug-and-play building of a data-driven offerings, e.g. (open) standards, de-facto standards, APIs, standardized data models, etc. The more functionalities are available
that help others to build and position innovative offerings better, faster, etc., the more attractive the offering itself becomes. The infrastructure layer itself has little value per se unless other users and partners create value on top of it.

For instance, the agricultural-robotics technology company Skyx\textsuperscript{11} is neither offering hardware nor agriculture end-customer applications, but a software that enables a modular swarm of autonomous drones for spraying. By providing a technology to plan and control the mission of drones in real-time as well as to auto-pilot the entire fleet/swarm, it addresses the need of agri-spraying application developer applicators in building their solutions in higher quality and less cost by relying on an standardized approach. In addition, as the software is compatible with any commercially available hardware, cost in connecting the wide range of drones can significantly be reduced. Thus, Skyx provides tools and connectors for agri-spraying application developer to build their own solutions. The more drone hardware can be connected, and the more spraying functionalities can be provided, the more attractive is the overall offering for applicators.

Other examples that fall under this category are anonymization services, data quality services or data acquisition services. In general, network effects on infrastructure are more likely to happen in markets ((Hagel \textit{et al.}, 2015):

\begin{itemize}
  \item that benefit from tightly integrated and standardized products technology component that can be easily be connected and combined (e.g. due high technical complexity, high dynamics of development);
  \item that benefit from third parties to contribute important assets, such as hardware, data, etc.;
  \item where third parties / customer benefit from the opportunity to implement their own sector-specific value proposition.
\end{itemize}

A critical success criterion for data-driven innovations that harness network effects on infrastructure level, is fair and balanced value creation and capturing. Third parties connecting the offering as technology providers or as application developers need to be able to generate significant value for themselves while innovators themselves need to yield strong returns at the same time. Thus, innovations with network effects on infrastructure level always impact the underlying economics and operations. As highlighted before, the scale and scope of the data-driven innovation is central to stimulate its demand side. In this context, open source software, de-facto standards, open APIs, and standardized data models are important levers (Choudary, 2015).

In our study, only 4\% of data-driven start-ups focused on network effects on infrastructure level. This low percentage might be caused by the deep understanding and expertise required to exploit network effects on infrastructure level. As the focus is to identify common denominator/core components required for building a larger set of offerings, deep knowledge in the similarities and difference of the various offerings is needed.

\textsuperscript{11} https://www.skyx.solutions/
**On marketplace level:** In cases where the number of marketplace participants is the key source of value, data-driven innovation offering network effects on marketplace level are central. Offerings that connect its participants in their specific roles, such as buyer and seller, consumer and producer, etc. allow that two participants can easily interact with each other. Two aspects increase the attractiveness of marketplaces:

- The number, quality and type of participants connected to the marketplace. In this context, balanced growth as well as critical mass of both participant types is required to ensure that promising counterparts can be found.
- The platform benefit from intelligent forms of matchmaking ensuring that the mapping of partners ensures nice fit of interest.

The purpose of connecting different participants via a marketplace is to exchange “something”, i.e. goods and /or services. The asset being exchanged is also called core value unit and determines the design of the platform. In this context, three different types of core value units can be distinguished:

- **Goods,** in those cases the marketplace enables the exchange of physical products that can be described along the product categories and price expectations. Prominent examples are ebay or etsy. The matchmaking algorithm is based on faceted search based on a set of well-defined product categories.
- **Standardized services** are promoted as “off-the-shelf” offerings without means for customization. Typical examples are rides on Uber or boat rentals by Zizzo¹². By standardizing the access to the service of interest, provider and consumer can be mapped automatically. The description of the service is used as input for the matching algorithm.
- **Marketplaces offering non-standardized services,** such as dating platforms, rely on the description of the service provider (also known as user profile). This description is then core value unit determining the matchmaking algorithm.

In general, in accordance to the core value unit different types of matchmaking algorithm are build. For instance, the start-up Selectionnist¹³ is based on an advanced match-making service that connects consumers and brands by means of image recognition technology. Thus, consumer that locate a product of interest in a magazine can use a mobile app to snap a picture of the product which automatically forwards the customer to the product’s online shop.

Network effects are built into the definition of a marketplace, i.e. all successful marketplaces benefit from network effects. This was also seen in our study. The low number of marketplaces in our study, indicates the high challenges of building them. The challenges are less on the technical level but more on the level of building critical size and balanced user communities. Several strategies to attract users from the different communities have been implemented by the start-ups:

- **Complementary/additional value propositions offered to engage provider:** For instance Zizzo, the marketplace for renting boats, addressed the shortage of boat charter companies by providing them a powerful inventory boat

¹² [www.zizzo.com](http://www.zizzo.com)

management tool encouraging them to digitalize their processes which again was a necessary precondition for being able to join the marketplace.

- **Building of participants communities:** the founder of Insurify\(^{14}\), the first online car insurance shopping platform Insurify\(^{15}\) (TripAdvisor for car insurances), invested several years in building relationships with car insurance companies.

- **Provisioning of incentives to attract consumer:** The Hooch\(^{16}\) (see also Section 4.3) hospitality app reward subscribed members with a voucher for one free drink every day, which helps cocktail bars to gain new customers.

![Figure 10: 57% of start-ups harness network effects](image)

To sum up the discussion about of network effects, we can state that it occurs most frequently on data level. This highlights the importance and impact of getting access to high quality data. In addition, we could observe that for businesses being based on network effects multi-sided business models are very likely to be present.

### 3.4.4 Message 4: Data-driven innovation relies on multiple revenue models

We have been interested in the question of how data-driven businesses are making money. Is this different from traditional businesses? And can we identify some dominant revenue models?

Our first finding is that often information about the type of revenue models used was difficult to find. Especially in cases when start-ups have been focusing on emerging

---

\(^{14}\) [https://insurify.com/](https://insurify.com/)

\(^{15}\) [https://insurify.com/](https://insurify.com/)

\(^{16}\) [https://hooch.co/](https://hooch.co/)
technical advances, such as drones or autonomous driving, information about revenue models was not available. As emerging technology businesses are often seen as investment or bet on the future in a market not yet established, the absence of revenue-related information is not surprising. For 10% of the companies analysed this was the case, i.e. no information about the revenue model could be found or inferred.

Our study confirmed the finding of (Attenberger, 2016) that revenue models have not changed through the usage of data technologies. The major difference to traditional businesses is that data-driven innovations rely on different types and combination of revenue streams that are continuously changing over time in order to address the specific user needs of each customer segment. On the one hand, we observe very new forms of value propositions, ranging from service offerings, the bundling and unbundling of offerings, intermediate offerings, product differentiations through versioning, etc. that allow to address the very specific user needs. On the other hand, the majority of data-driven innovations has – in comparison to traditional businesses – a different cost structure. With data and data offerings being cheap to reproduce and deliver, the typical cost structure of data-driven innovations relies on fixed costs for the development of the offerings but low variable cost. This kind of cost structure leads to substantial economics of scale as with more offerings sold, the lower the average costs of development become. In addition, as the reproduction and distribution costs are often marginal, the danger of price dumping and surplus of offerings in the competitive market is a frequent phenomenon. For instance, (Aitken and Gauntlett, 2013) counted more than 40,000 health apps in the app store being offered for free or for a very small price.

With this new cost structure for most of data-driven innovation, organisations have a new flexibility to adjust the equation between value proposition and price in accordance to the user needs of the various customer segments. In this context, companies elaborate the specific price level the targeted user group is willing to pay. The main objective for aligning the product version with pricing version for each customer segment is to attract more users and interactions as well as grow the community.
On average, data-driven companies in our study have 2.6 revenue models. The highest number of revenue models we found with the start-ups in our sample was 5. For instance the company bird.i\(^{17}\) is using five different revenue models (see Figure 11) to attract different customer segments and to ensure market growth. The company offers satellite imagery available for everybody, i.e. to all different customer segments ranging from individual to business. Through their partnerships with a wide range of satellite image providers, they offer more data in comparison to single-provider data sources. In addition, they facilitate access to the images by enhancing the original data with valuable geo-referenced information. The first revenue model is a combination of freemium and advertisement (not explicitly listed in Figure 11 as for this revenue model no contract is needed). The other revenue models range from “pay-as-you-go” to three different types of subscription models. The first one is addressing individuals and restricts usage of images for private and public applications whereas the second and third subscription models is addressing commercial usage with moderate and extensive usage. This examples also shows how the value offerings differ in in each subscription category.

In addition, due to the high market dynamics and impact of network effects, revenue models are likely to change over time due to the market dynamics as well as their own strategies to actively harness network effects.

\(^{17}\) https://hibirdi.com/
The most frequently used subscription model in our study was the subscription model.
Due to big spread and high adoption of Software as a service (SaaS) approach which
describes a software licensing and delivery model in which software is licensed on a
subscription basis and is centrally hosted\(^\text{18}\). The deployment of data-driven innovation
as SaaS-based offering brings a lot of flexibility. For instance, end-user do not require
to install the software on their own system, they are relieved from the burden of
operating the application themselves and they can access the software on-demand.
The second most frequent revenue model is the selling of services in which the
person’s time is payed for. Those revenue models are very often used for open
software offerings as well as when offerings are not

### 3.4.5 Main Learnings from Empirical Study

From the analysis of start-ups, we developed a detailed understanding how data-
driven business opportunities are scoped. First, all start-ups had very convincing value
propositions for their different customer segments in place. Start-ups addressing
emerging technology markets that were not able to address concrete end-customer
groups (as the market is not yet established) had a quite clear picture of the core
capability they are delivering for the emerging technology ecosystem. For instance,
the company Carfit who is building self-diagnostic and predictive maintenance
platform for the connected car space. This functionality will become very critical when
vehicles are moving autonomously. Carfit has a clear partner strategy to position
themselves in the strong ecosystem of Nividia. In order to ensure their future
competitive position, they are putting a lot of efforts in building the most
comprehensive library on car’s vibrations.

\(^{18}\) see [https://en.wikipedia.org/wiki/Software_as_a_service](https://en.wikipedia.org/wiki/Software_as_a_service)
Secondly, the handling of network effects requires start-ups to have dedicated partnering strategies and community or user building strategy.

Third, successful start-ups continuously adapt their offering to market needs and demand. By investigating the user needs of one customer segments they learn how to improve which might lead them to different business type.

Finally, the supply side of offerings was always very clear and focused. By providing a very clear and focused scope of the offering, the demand side is easier to conceptualize.

4 Example of Success Stories

In the following section, we provide some examples of success stories of data-driven start-ups to give the reader an impression how clear and focused their supply and demand side can be pitched.

4.1 Artomatix

Artomatix is a Dublin-based software company founded in 2014 that uses Artificial Intelligence to create realistic 3D art creation.

Artomatix’s users are artists and developer of the video gaming industry that can benefit from a service that supports the realistic 3D art generation of textures and texturing. Before this tedious task was done manually. With the suite of tools provided by Artomatix, artists can now do the same task 10 times faster.

The technology is based on computer graphics, Deep Learning and computer vision. It uses generative neuronal networks to “imagine” new details of a texture in a way a human would do, i.e. it recognized objects in a video, can add texture and features automatically by relying on the “learned” knowledge what should be there.

The data use for training and developing the algorithm is video and image data. The software can be integrated with Photoshop and leading gaming engines like Unity and Unreal.

The company is using three different Subscription models (Indie (Revenue < $100k/year), Professional (Revenue < $1M/year) and Enterprise (Revenue > $1M/year). Enterprises can license Artomatix’s technology and build them into their existing process for an annual fee. The technology is offered as data-driven service. There are no network effects that need to be reflected.
4.2 Selectionist

Selectionist is a France-based company found in 2014 offering image recognition technology with the goal to connect readers of print journals with the world’s largest brands through an application or a chatbot. They aim to bridge the gap between offline content and online experience by offering an advanced match-making service to connect consumer and brands.

They address two different customer groups with different value proposition:

- Value proposition for consumer: they locate and potentially purchase a product they spot in a magazine just by snapping a picture of it.
D2.6: Annual Report on Opportunities

- Value proposition for brands: brands can see in real-time how readers interact with their editorial & advertising in print magazines. Their match-making algorithm is based on image recognition technology that continuously improves the more images from brands’ product are in their data bases (more brands) as well as the more user request they receive. Thus, their offering is based on network effects on data level. The service is conceptualized as marketplace based on commission fee and with network effects on marketplace-level.

Figure 14 One-Pager of Selectionist
4.3 Hooch

Hooch is a US-based company that offers a subscription-only cocktail app and discovery platform, where customers are offered a free drink each day. The platform aligns the needs and interests of two different user groups in an efficient manner:

- **End-user (Classpass for Drinks):** for a monthly subscription fee ($9.99), people can get one free drink each day at over 400+ top bars and restaurants (only selected cities).
- **Venue owners use Hooch as free marketing tool that attracts people to their venues.** As end-users do not have an incentive to “bar hop” (only one drink a day), they are likely to stay and consume more.
- **Beverage industry can benefit from insights related beverage consumption trends.**

The Hooch app is based on a simple search functionality allowing user to find interesting bars or venues. In addition, Hooch is collecting large scale real-time data on the millennials’19 behaviour, drink preferences and spirits consumptions to identify spirits trends and other related patterns. The generated insights are offered to alcohol industry as service20.

Hooch has secured national partnerships with the largest wine and spirits distributing companies (customer co-development). In addition, they announced major partnerships with Amazon to establish their own sales channel for exotic drinks branded by Hooch21.

Its revenue strategy is based on the subscription fee from members as well as the income generated by the huge data business on spirit trends. The platform relies on network-effects on data-level as with more behavioural data available the more precise insights can be generated. In addition, it relies on network effects on marketplace level as the app is only of interest for end-user in case there is a critical mass of bars and venues offered and vice versa.

---

19 Millennial is a person reaching young adulthood in the early 21st century.
20 It is not explicitly mentioned how the revenue stream is generated for this scenario.
21 E.g. the Hooch Hoola Alkoholische Maracuja und Mango Brew
4.4 Arable

Arable is an US-based company founded in 2013 offering agriculture businesses a global solution for managing weather and crop health risks, delivering real-time, actionable insights from the field.

The target users are growers, advisors, and businesses who aim to play a proactive role in the quality and longevity of their operations.

The agricultural business intelligence solution is based on in-field measurements allowing to produce real-time continuous visibility and predictive analytics in the area of crop growth, harvesting time, yield and quality. The solution is relies on field-level weather and crop monitoring devices (hardware that is part of the solution) that collect over 40 field-specific data metrics. To enable the access to data from anywhere
in real time, a cloud-based software platform based on a tiered SaaS offering (different level of services) is combined with IoT hardware

Arable is selling licenses for enterprise software to agribusinesses. As the prediction service improves with more data available, the solution of Arable is based on network-effects on data-level.

5 Evaluation

We used and evaluated the DemoX model with a university course and in a series of Business Model style workshops. The insights and learning are described in the following subsections
5.1 Data-Driven Innovation course at university

The DemoX model was evaluated with 35 students that participated in the university course “Data-driven Innovation” at the Technical University Munich in the summer semester of 2018.

The course is conceptualized as block lecture of four days complemented with three-half day session for idea pitching and feedback. The content of the course follows the logic of the DemoX model covering the value proposition, the data aspect, the technology aspects encompassing the whole data value chain, the partner and ecosystem as well as the value network and revenue model. All the course content was complemented with methods, such as the DAMIAN approach for investigating the data value chain, and guiding questions to explore the different aspects.

To evaluate the course we accomplished a quantitative and qualitative analysis based on a questionnaire the students filled out at the end of semester. The questionnaire covered questions to rate the relevance of content, methods and set-up.

For illustrating our findings, we are using in Figure 19-22 boxplots that allow to group numerical data values through their quartiles: the red line is representing the median of the data set, the box boundaries indicates the values of the upper and lower quartile (min and max values) and the dotted line the lower quartile splits of lowest 25% of data from the highest 75%, whereas the upper quartile splits of the highest 25% from the lowest 75%.

The quantitative analysis indicated that the most relevant content was the value proposition followed by the data aspects (see Figure 17). This underlines the fact, that sector-specific data-driven innovation rely on very specific customer segments and that the characteristics and nature of data has a strong influence on scoping data-driven opportunities.

---


23 Lower quartile splits of lowest 25% of data from the highest 75%, whereas the upper quartile splits of the highest 25% from the lowest 75%.
Each content block was complemented with methods to explore the theoretical content with some practical guidance (see Figure 18). The method value proposition challenges, a simplified methodological guidance derived from (Osterwalder et al., 2014) as well as the set of guiding questions related to the network strategy and revenue models have been evaluated as very relevant methods. Both methods introduce non-technical aspects that might have been unfamiliar to technical students. The DAMIAN method (see Deliverable 2.5) that provides guidance in exploring the end-to-end coverage along the whole data value chain was seen as relevant for analysing data-driven opportunities, whereas the methods for exploring the roles and interest of partners (Co-innovation Risk) and ecosystem (Adoption Chain Risk) have been assessed with slightly less relevance. This might be due to the fact, that the alignment with internal and external partners is in the beginning phase rather abstract but gains more momentum when innovations move closer to the implementation and realization phase.

![Figure 18 Relevance of Methods](image)

The course is based on four different methodical elements to ensure transparency over the innovation process and to motivate students to continuously work on further exploring the data-driven idea. In addition, the four elements are required to document the outcomes to the students’ innovation journey in a way that credits can be assigned in fair manner.

Very helpful for the students was the investor briefing, an interactive setting giving each student the opportunity to pitch the current progress of his or her idea complemented by critical questions and/or recommendation from “friendly investors”, i.e. two to three other students with the according briefing. The first investor briefing session took place during block-seminar; during the semester the investor briefings were repeated three times in three hours morning sessions. Students that attended the investor-briefings on a regular basis could progress their business ideas in systematic manner by incorporating the many ideas, thoughts and recommendations from others. The high value of the investor briefing energizing the
systematic analysis of data-driven innovations along the DemoX categories found its precipitation also in the quality of the final reports and presentations; particularly in comparison to students’ outcomes that have not been able to attend those sessions. To document all intermediate results and investigations, each student was requested to have an *innovation log-book*, a personal wiki that was used to reflect the own progress in scoping the data-driven innovation along the introduced methods and guiding questions. Concrete, each student was requested to capture the learnings, outcomes and derived decision of at least 7 applied methods by documenting all material without formalities but reusability of content in mind. Originally, the innovation log-book was introduced to give the students maximal flexibility in exploring their ideas while having transparency for later credits. In addition to this purpose, the students appreciated the approach as continuous working stream providing them guidance and triggering discipline in the overall innovation journey. For grading purposes, students have been requested to do a *final presentation* as well as consolidate a *final written report*. The final presentation was experienced as very helpful as it required the students to develop a nice convincing story about their data-driven innovation that required them to focus on the differentiating aspects of their idea. The written report was perceived as less relevant. This might be caused by the fact that the submission date was some days earlier and many of the reports have thus not been that mature and precise at this point of time.

![Figure 19 Importance of Set up and Interactive Elements](image)

In order to fully capture the high interrelation between the demand and supply trends and dynamics of the data-driven business opportunity, we encourage the students to apply the *methods in iterations*. We did not specify an exact sequence of methods but provided clear instructions for which purpose each method can be used. The evaluation shows that students used the methods several times in iterations (see Figure 20).
To complement the quantitative analysis, we added some open questions to the survey. We asked the students about the most important learning in the course. The aspect highlighted most was the systematic approach and methodical support offered by the DemoX frameworks and associated models. This includes the importance of communication (pitching and feedback), the need to go out and request interviews, and the guidance in elaborating a business idea with a clear structure in mind leading to final pitch. Based on the DemoX model, methods and guiding questions students learned how to think critically as well as were empowered to look for innovative solutions by themselves. They understood how to ask the right questions in an interview as they knew which amount and level of detail, they required for scoping the data-driven business opportunity.

One important aspect of the overall analysis was the alignment of the different perspectives of a data-driven innovation. For instance, by incorporating technical knowledge into the business idea, the supply side could be complemented with the demand side perspective as basis for positioning the start-up idea on the market. By evaluating different aspects from different angles, the student learned to get to the point of implementation.

Summarizing, students very much appreciated learning how to scope data-driven opportunities in a pragmatic manner. In particular, the development of a promising data-driven business opportunity could be demystified (“developing a start-up idea is not magic” or “data-driven innovation can be found in every aspect of our lives”). In this context it was mentioned how valuable it is to learn how data can be used to improve our lives.

We also asked the students about the most valuable aspect of the course. Most students highlighted the methods and DemoX model in general or mentioned particular aspects of the overall set-up, such as the overview of the process, the logbook, hands-on exercises, and the practical analysis of innovations. In addition, students highlighted the importance of non-technical aspects. For instance, as students mentioned that they no perceive user needs as the central aspect for scoping offerings, thus the exploration of concrete value proposition for specific target user group as well as to always consider all stakeholders gains importance. Finally, they appreciate learning how industrial business opportunities are investigated and how real-world problems using existing data technologies can be solved. To revise ideas
again and again allowing many different aspects to be considered has been experienced as new, important but easy approach.

When asking about **what should be kept** in the block course setting, the investor briefing as well as the practical approach and relaxed atmosphere have been highlighted.

When asking **what should be changed**, we learned that deliverables and tasks should be described more clearly. Being in an uncertain and new situation, any information providing guidance is seen here as very valuable. The students also highlighted that there is less need for content in frontal teaching but requested more guidance for doing the practical explorations, for instance guidance in finding a topic, guidance in continuation the idea after the course or ways to put more emphasis on the data aspects as well as the practical implementation.

### 5.2 Data-driven innovation workshops for professionals

We have designed and conducted data-driven innovation workshops in business settings as one-day trainings for professionals, as described below:

**Short Description:** Data Technologies are continuously evolving. They affect people, organizations, industries and society. The basic premise of this course is that data technologies are rapidly transforming not only the way how we live and work, but also how companies are organized and innovate. Therefore entrepreneurs / intrapreneurs, managers, and technology experts need to understand the business implications, technology paradigms, changing customer needs and management practices of data-driven innovations. The main objective of the course is to develop a basic understanding for and practical skills in implementing data-driven innovations.

**Learning objectives:** The participant

- understands the nature and characteristics of data-driven innovation
- learns how to explore user needs of selected customer segments
- understands the important role of data
- gets an overview of the main data processing steps
- gets insights in analyzing data-driven business opportunities
- understands the dynamics and impact of the associated ecosystems

**Content**

- data-driven value proposition
- data perspective
- data value chain
- data-driven network effects and value network strategy
- co-innovation and value-chain adoption risk

**Target Groups**

- Entrepreneurs,
- Intrapreneurs
- Managers
- Technology experts
D2.6: Annual Report on Opportunities

After the first workshop (on 15th July 2018 as co-located event of the internal Siemens conference on Artificial intelligence in Erlangen), a quantitative and qualitative evaluation was carried out. The quantitative evaluation is summarized in Table 2 and shows the high satisfaction of participants.

<table>
<thead>
<tr>
<th>Recommend course to other</th>
<th>Happy about learnt topics</th>
<th>Goals reached</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>1.4</td>
<td>1.6</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Table 2 Quantitative Evaluation of Data-driven Innovation workshop

In the quantitative analysis we asked about what should be kept or changed. In addition, we discussed with the participants whether they prefer the training as online training or as face-2-face workshop. And we asked about the optimal length of the workshop.

Summarizing, the participants appreciated the clear and well-structured content based on the DemoX model, wide range of concrete examples and practical exercises. We learned that a one-day workshop was perceived too short for the content provided. As the concrete business examples and content provided was perceived as very helpful, there was a clear tendency to vote for a longer workshop setting ranging from 1.5 to 2 days allowing to keeping the detail of content and give more time for practical exercises. In addition, due to the interactive character of the workshop, participants preferred a face-2-face setting as this stimulates discussion and exchange between participants.

---

24 The scale was from 1 to 5 with 1= fully agree, 2= agree, 3= neutral, 4= partly disagree and 5=completely disagree. The long description of the categories as follows: Recommend course to others: I would recommend this training to other interested Siemens employees / Happy about learnt topics: I have learnt important topics which will help me to do my job better in the future / Goals reached: I have reached my learning goals / Material: the learning material was understandable and helpful.
6 Conclusions and Next Steps

This report provides a second update of the BDVe task aimed at documenting and reviewing emerging business opportunities. We conducted a quantitative study to identify success approaches and patterns in data-driven business opportunities that will provide guidance for investment decisions in the future.

In this study we found out that the majority of data-driven businesses address B2B markets and that 75% of data-driven start-ups have a clear sector focus. In addition, we found out that data-driven start-ups rely on multiple revenue models and that their majority harness network effects.

To complement the evidence-based analysis, we tested the DemoX approach in a wide range of settings and contexts. The testing was done in Business Model Canvas style webinars and workshops in which participants were guided by DemoX models, content and guiding questions in developing data-driven business opportunities.

We want to highlight that the reason for focusing in our empirical study mainly on data-driven start-ups was because data about corporate innovation is difficult to get. The usage of the DemoX approach is not restricted to start-up innovation but can be used within all types of organisation. For instance, when using DemoX for scoping data-driven innovation projects in Siemens, we received very nice feedback.

In the next stage we will compile all material and content in a way that other trainers and teachers can reuse and easily customize the material. In addition, we plan to compile a webinar to do some marketing of the DemoX approach as well as to conduct further DemoX workshops and trainings to engage innovators, managers and experts in exploring data-driven business opportunities in well-thought through manner.

By relying on the DemoX model, we have now a proven method in place that we can share with all members of the BDV ecosystem to provide guidance in exploring and scoping data-driven business opportunities. The comprehensive content can be used for industrial workshops and educational set-ups. In the future, we plan to engage with the stakeholders of the BDV ecosystem, with focus on SMEs and start-ups, to help them scope promising business opportunities. In this way, we aim to increase the industrial investments in data-driven innovation in the PPP ecosystem.
7 References


