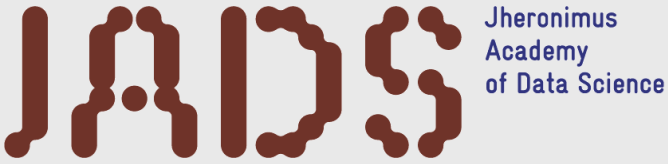
REPORT FOR THE CERTIFICATION COMMITTEE

PROFESSIONAL DOCTORATE IN ENGINEERING PROGRAM DATA SCIENCE

September 2016 – September 2019



Jheronimus Academy of Data Science

Sint Janssingel 92

5211 DA 's-Hertogenbosch

Summary

Learning environment

The PDEng program Data Science (PDEng DS) provides training, coaching, and education to its students in order to let them become successful professional data scientists with the ability to design data products that interact with data and with users at different levels of complexity. The key features of the program are:

- Professionalization enhanced learning environment
- Educational philosophy based on self-directed learning
- Integral part of the educational community of JADS Den Bosch
- Close cooperation with the ecosystem

Students learn to make sense of data and develop the ability to articulate their discoveries and recommendations to those not schooled in the world of data in the frame of industrial and business, design and decision processes.

Professional / Personal Development

The program PDEng DS design has come to completion in the past three years. Its main elements are concentrating on Professional Development and Personal Development. The time students can spend on both has a ratio of 5 : 3. Professional Development may be regarded as the obliged part of the program; Personal Development is the elective part. Here we define:

Professional Development is the assembly of all activities by which the program aims to professionalize the student to become a data scientist in industry.
The typical Professional Development activities refer to all forms of knowledge transfer and skill training from data domains and data science experts.

Personal Development is the assembly of all activities in the program that the student plans to undertake to reach his/her individual learning goals as he/she sets at the start of the program.
The typical Personal Development activities refer to all forms of E-learning, Kaggle competitions, coaching, Hackathons, summer schools and conferences, and courses and trainings shared with other PDEng programs.

Learning environment

Besides a variety of training and coaching activities in cooperation with our industrial partners, such as

- Interview workshop
- Data quick scan
- In-house data challenge
- Case study
- Data challenge week

The program offers the following nine modules

Module	Coach
Introduction to Data Science	Stef van Eijndhoven
Data Mining	Decebal Moncanu
Data Engineering	Damian Tamburri
Statistics	Rui Castro Pires da Silva
Visualization	Jack van Wijk
Analytics	Dick den Hertog
Modeling & Simulation	Tiberiu Muntean
Supply Chain Management	Geert-Jan van Houtum
Process Mining	Massimiliano di Leoni

Staff and Personnel

The staff and personnel of the program are concerned with the management of the program, the relationship and embedding within the ecosystem, the skill training and individual coaching, the training on issues related to ethics and law, and the training on technical/business report writing. The division is as follows:

Management

- Jack van Wijk
- Stef van Eijndhoven (till 01-01-2020)
- Tiberiu Muntean (from 01-06-2019)

Secretariat

- Denie Maas
- Femke Korst

Liaison

- Liesbeth Leijssen

Professional Coaching

- Sandra van Dongen
- Anouk van de Kerkhof

Ethics & Law

- Koen van Holten

Technical Writing & Editing

- Judy Strother

Final Projects

The first nine final projects were started up on 1 September 2017. Since then 45 final projects were acquired. From these, 33 were completed before 1 October 2019, and 12 started or will be started in 2019.

Industry (18)		
	Alfa Laval	1
	ASML	3
	DAF	1
	Heijmans	1
	Hendrix-Genetics	1
	Jumbo	1
	Nabuurs	1
	Omron	1
	Philips Health	1
	Philips Research	2
	Signify	1
	TE Connectivity	1
	Unilever	3
(semi)-Government (14)		
	Municipality Den Bosch	1
	GGD	2
	OM Zuid	1
	LOCC	1
	Court of Justice	1
	National Police	2
	KNVB	1
	Brabant Water	1
	Enexis	2
	Firebrigade Amsterdam Amstel	1
	Rijksmuseum	1
Commercial Service (8)		
	Netscalers	1
	Studyportals	1
	Saint Gobain	2
	Shoeby	1
	Green Gorillas	1
	CINOP	1
	Efteling	1
Financial Services (3)		
	CZ	1
	Vivat	2
University (2)		
	WUR	1
	TiU	1

Challenges

The main challenges that the program faces are:

- To maintain the program's outspoken educational philosophy offering a learning environment with a self-directed goal setting learning style
- To find motivated staff members from TU/e and TiU to step into the PDEng Data Science program adventure
- To remain an integral part of the JADS community at Mariënburg, 's-Hertogenbosch
- To build strong relationships with industrial partners so as to let acquisition of final projects become a less critical part of the program
- To be frontrunner in educating the next generation data scientists
- To keep attracting highly motivated students and young professionals from the Netherlands and from all over the world
- To efficiently manage a growing program

Table of Contents

Chapter 1 | **Introduction**

In the PDEng program Data Science, professional development is the core focusing on the development of the technical and non-technical skills in an integrated way. Professional development is directly related to the professional profile of the data scientist in action. Knowledge acquisition in the professional field is directly related to the program's educational profile.

1.1 Data Science

Data Science can broadly be defined as the study and design of computational principles, and automated methods and systems, to analyze massive and complex data from which to extract useful information. As such Data Science lies at the crossroads of computer science, applied mathematics and statistics. Large data sets are now generated by almost every activity in science, society, and commerce — ranging from molecular biology to social media, from sustainable energy to health care. Data Science asks: How can we efficiently find patterns and the dynamics of these patterns in these vast streams of data in the context of the data environment? Many research areas have tackled parts of this problem: machine learning, and data and process mining focus on finding patterns and making predictions from data; databases are needed for efficiently accessing and integrating data and ensuring its quality; algorithms and architectural models are required to build systems that scale to big data streams; natural language processing, computer vision, and speech processing are each needed for analysis of different types of unstructured data. Knowledge of law and ethics is required to understand legal and ethical implications; knowledge of management and business to enhance market decisions and explore economic consequences of choices. Recently, these distinct disciplines have begun to converge into a single field called Data Science. Data Science becomes a new frontier for design.

> *"A data driven organization acquires, processes, and leverages data in a timely fashion to create efficiencies, iterate on and develop new products, and navigate the competitive landscape."*
>
> *Patel, president LinkedIn*

1.2 Mission

The two-year interdisciplinary PDEng program Data Science combines statistics, computer science, mathematics, and design theory with the business acumen to explore data sets, gather actionable insights, visualize results, and communicate meaningful findings taking into consideration underlying legal and ethical contemplations. Graduates make sense of data and have the ability to articulate their discoveries and recommendations to those not schooled in the world of data in the frame of industrial and business design and decision processes.

1.3 Candidate Profile

The profile of the candidate students of the PDEng program Data Science is characterized by:

- Solid background in mathematics, statistics, and computer science
- Experience or at least strong affinity with data science
- High motivation, eagerness to learn, and self-propelling
- Great desire to make the next step after an academic Master education towards becoming a top level professional data scientist

1.4 Professional Development

Professional Development in the PDEng program Data Science is the assembly of all activities in the program that aim to professionalize the (PDEng) trainee to become a data scientist in industry.
Professional Development contrasts with Academic Development that focuses on knowledge acquisition with the intent to specialize as a data scientist in academia.

> "Data science professionals are characterized by a combination of technical and non-technical skills, also referred to as thinking, learning, literacy, and social skills. The main focus of Professional Development is on the integration of these skills in an industrial and/or organizational context. Working on an industrial/organizational problem in a case study context with data from a company is an example, since without skill integration such study is deemed to fail.
>
> "Typical Professional Development activities in the program refer to forms of knowledge transfer from data domain experts such as interview workshops and data quick scans, skill workshops as part of data challenge week preparations, but the coaching from a mentor and discussions with colleagues are part of the Professional Development process, as well.

1.5 Educational Philosophy

The key points that characterize the educational profile of the PDEng program Data Science are:

1. *Learning to learn*. Data Science technology is changing fast, students must be encouraged and enabled to learn new things themselves. The program uses blended learning: students should follow MOOCs, read articles and books, and scan the web to obtain the knowledge they need.
2. *Learning from each other*. Students must be encouraged to explain their colleagues what they have done and why, support each other with their own specific knowledge and experience, ask critical questions to each other, etc.
3. *Learning by doing*. To digest theory is easy, to apply it in practice is often much more difficult, and can only be learned by being confronted with strengths and limitations in a hands-on setting.
4. *Focus on real-world problems.* Students should be exposed to real-world problems, for a variety of domains, requiring different solutions and strategies.
5. *Broad expertise*. Data science is a huge field, and nobody can be expected to know all about everything. But, students should be able to understand experts in aspects of data science, and to go in depth when needed.

6. *Development of professional skills,* including team-work, planning, presentation, reporting, and the ability to apply these for data science is central. Law and ethics with respect to data privacy and governance are a prominent part of the daily practice of the professional data scientist.

Each element of the PDEng program addresses these aspects, and lecturers, coaches, and supervisors are invited to take these into account.

1.6 Final Project

The twelve month project is a full time project carried out by one PDEng-student, preferably at a company. In a preparatory phase, before the start of the project, problem context and problem specification are defined. Thus, as the project starts, deliverables, project planning, and time planning are identified. The project has a substantially scientific level and is devoted to problems of highly innovative level. The project is executed under a nondisclosure agreement. The project is executed as contract research with Eindhoven University of Technology and Tilburg University and thus based on a formal contract. Costs of a long term project are € 5,800 per month, VAT excluded.

1.7 Graduate Profile

The graduation profile of the PDEng program Data Science reveals that the student adequately developed technical and nontechnical skills and integrated these skills in the following training and learning outcomes:

- Systematically acquired an understanding of a substantial body of knowledge and experience with a feel for entrepreneurial issues and challenges that are at the forefront of the professional Data Science practices
- The ability to conceptualize, design, and implement a project for the generation of new knowledge, application or understanding at the forefront of Data Science disciplines; to adjust the project design in the light of unforeseen problems
- A detailed understanding of applicable Data Science techniques for design and analysis at academic level
- Ability to make informed judgments on complex issues in specialist domains, such as agro & food, health, financial markets, smart industry, and life science, often in the absence of complete data, and be able to communicate ideas and conclusions clearly and effectively to specialist and non-specialist audiences

The qualities and transferable skills necessary for employment requiring the exercise of personal responsibility and largely autonomous initiative in complex and/or unpredictable situations, in professional and equivalent environments.

Chapter 2 | **Profile and Assessment**

2.1 Professional Profile

The amount of data produced across the globe increases exponentially and will continue to do so in the foreseeable future. In business institutes, internet markets, and industries, servers are overflowing with usage logs, message streams, transaction records, sensor data, business operation records, and mobile device data. An efficient analysis of these huge collections of data — big data — will create significant value for any economy by enhancing productivity, increasing efficiency, and delivering more value to consumers. Studies estimate that trillions of euros of value in efficiency improvements and economic growth can be unlocked by extracting actionable knowledge from the deluge of data now being collected in almost every sector of the economy.

> "We are on the cusp of a tremendous wave of innovation, productivity, and growth, as well as new modes of competition and value capture—all driven by big data as consumers, companies, and economic sectors exploit its potential," write the authors of Big Data: The Next Frontier for Innovation, Competition, and Productivity, a comprehensive research study published by the McKinsey Global

Nowhere has the benefit of analyzing data been felt more strongly than at top technology companies. Throughout the world, many Business Analytics companies are founded to support the production and analysis of data, so that insights from that data becomes in the benefit of their users. To make use of data, companies first need to be able to reliably store, process and query its huge inflows. As a result, the data infrastructure needs to be distributed, scalable, and reliable, which is not a trivial engineering task given the terabytes of data involved. The data engineer creates and maintains these robust big data pipelines. The data scientists develops tools to analyze the data. Data scientists and data engineers form the basis of the data teams that are quickly becoming a central part of most technology companies' technical teams.

At present, data scientists and data engineers come from the traditional areas of computer science, mathematics, and engineering. They leveraged their underlying skills to enter this fast changing, dynamic area. This includes researchers or professionals coming from fields such as computer engineering, electrical engineering, embedded systems, network engineering and systems engineering. Individuals who have studied mathematics and computer science fundamentals and have worked with large code bases or managed computational infrastructure in fields such as aerospace engineering, nuclear engineering, chemical engineering, mechanical engineering, physics and mathematics have all successfully made the transition.

Engineers and PhDs coming from backgrounds like the ones listed above often take a long and winding road to get into the field, learning the tools used in big data informally over long periods of time and through chance encounters with the profession. While serendipity may be a good way for people to discover a field in their infancy, as it matures and as demand grows, there needs to be a more direct and efficient route into the profession. At this point, the PDEng program Data Science (PDEng DS) wants to step in and become the program par excellence to develop the next generation of leading big data professionals. The main challenge for PDEng DS is an educational one. There is an overwhelming number of tools,

skills, and competences needed to survive in the Big Data ecosystem. For a detailed discussion on the professional profiles of data engineers, data scientists, and business analysts we refer to Addendum A.

2.2 Candidate Profile

Candidates are creative problem solvers with excellent technical skills. They are team players, who seek interaction with a multidisciplinary environment to tackle problems, they are eager to learn, their attitude shows the responsibility for quality and quantity of knowledge gained. They have a solid background in Mathematics, Statistics, and Computer Science, and strong affinity with Data Science.

1. Candidates possess an academic master degree (MSc) in Mathematics, Statistics, Computer Science, or in a relevant application field of Data Science such as Econometrics and Bio-informatics. The MSc-degree is from an accredited institution comparable to the Eindhoven University of Technology or Tilburg University.

2. Candidates have an academic background that includes:
   - **Mathematics & Statistics –** a least two semesters of mathematics and statistics. Courses on calculus, linear algebra, statistics, optimization theory, and probability theory are required. Courses on topics as operations research, signal and time series analysis, and dynamical systems are recommended. The courses should indicate that the applicant has achieved the mathematical and statistical maturity to be expected of an upper level mathematics / statistics / econometrics graduate.
   - **Computer Science –** at least two semesters of computer science. Courses on programming, algorithms & data structures, databases, data mining/machine learning are required. Courses on topics such as object oriented programming and web development are recommended. The courses should indicate that the applicant has achieved solid knowledge and experience with the computer science aspects that are highly relevant for Data Science.
   - **Data Science –** Candidates show affinity with the field of data engineering or data analytics. They have experience with the application of Data Science technology to real world problems, via projects carried out during their education or their working experience afterwards.

3. Candidates show high **motivation and eagerness** to develop the skills to:
   - Combine generalist thinking and expert thinking in various data domains
   - Formulate opportunities and discover value in data
   - Solve complex problems in a data driven environment
   - Demonstrate sense for entrepreneurship, business and industrial processes
   - Execute projects in a well-managed, professional way, optimally using resources and obeying constraints
   - Acquire knowledge through a self-directed learning style
4. Candidates have an attitude that shows:
   - **Technical expertise:** desire to become an expert in one or more data domains with a generalist thinking attitude

- **Curiosity:** desire to discover, distill, and model a problem down to a clear set of concepts and hypotheses that can be tested
- **Creativity:** desire to look at a problem in different ways and find novel solutions
- **Communication skills:** desire to communicate effectively with all stakeholders, on all aspects of any data science project that include requirement elicitation, project progress, developed solutions, up to use of data to tell a story.
- **Social skills:** desire to collaborate with others, with similar and different backgrounds, to tackle challenging problems

In order to select the appropriate candidates for the PDEng DS, the management team of the program provides an extensive assessment. This assessment is based on the best practices of the former PDEng program Mathematics for Industry, in which a long history exists from low drop-outs, strong feeling of community, and high learning impact within this assessment.

2.3 Assessment procedure

The process of obtaining a position in the PDEng DS starts by applicants sending a motivation letter and Curriculum Vitae. The first selection is made by the program manager with focus on technical background, specialization, university of applicants' home country, and working experience. The selected candidates are invited for a Data Challenge Week, in which they are assessed on technical and non-technical aspects. The Data Challenge is introduced by a company, business or public institute. Participants work in teams of five.
The Data Challenge Week[1] is a 24/7 social event. Participants are lodged in bungalows. Data Challenge Weeks are organized twice a year. The accepted candidates join the program some months after the Data Challenge Week. The program does not recognize fixed starting dates.
First year students of the PDEng DS take part in the Data Challenge Weeks as well. Whereas this week is an assessment for the applicants, it offers a learning experience for the PDEng DS students. They are challenged to show their acquired technical knowledge, skills on teamwork, leadership and more non-technical issues. After a Data Challenge Week, the feeling of being part of a community increase significantly, which is an important indicator of success.

2.4 Assessment criteria

The criteria for assessing applicants in the Data Challenge Weeks are strongly related to the T-profile of the Data Science Professional. Besides the essential technical background applicants are selected on the following distinctive attitudes and skills: creative, self-aware, critical thinking, team and communication skills. For every attitude and skill, assessment criteria are described in more detail.

Essential elements of the assessment:

- Candidates write a report. They write a document on their personal understanding of the research problem and a personal introduction.

[1] For a typical schedule of a Data Challenge Week we refer to Appendix B.

- Candidates take part in team activities. They are part of group meetings and technical discussions.
- Candidates give the final presentation. They present results and conclusions of their group to the representatives of the company or business.
- Candidates fill out a personality questionnaire. They fill out the MPT-BS: Multi-Cultural Personality test – Big Six.

Final part of the assessment is the job-interview, a behavioral interview, with a strong focus on technical background, motivation and the assessment criteria. The scientific director, program manager and coach professional development form the interview committee.

Chapter 3 | **Professionalization enhanced learning environment**

3.1 Introduction

Eindhoven University of Technology offers nine two-year Professional Doctorate in Engineering (PDEng) programs within a wide variety of engineering disciplines with focus on technological design and innovation. The common goal of all these programs is to professionalize its participating students for a career in industry. Starting level is a completed academic master study, so that the required knowledge is covered for the greater part.

Although professionalization is key in all PDEng programs, the definition and implementation of professionalization is very different per PDEng program. The PDEng programs educate their trainees in their own established educational environments, which varies from the classical classroom education to real-life projects in companies. In this document we want to describe the rich and unique learning environment established in the PDEng program Data Science. We want to show the positive effects of such learning environment and inspire other programs to embrace possible new educational approaches and methods of learning.

The design of the learning environment and the consequential educational approach are based on the program's vision and learning philosophy, namely, that each trainee is master of own learning and that quality is determined by the process of learning rather than by the result of learning. These fundamental principles are integrated in all activities of the program. In this inclusive, challenging, and unique environment trainees will expand their skillset in both technical and non-technical areas in an optimal way.

3.2 Educational Philosophy of the PDEng program Data Science

According to Merriam Webster's Learner Dictionary, the verb 'to professionalize' means to make (an activity) into a job that requires special education, training, or skill. To professionalize and with that professional development is key in the educational philosophy of the PDEng program Data Science (PDEng DS). The program defines and accordingly implements 'Professional Development' as the assembly of "all activities (trainings, modules, workshops, case studies) within the program that aim *to professionalize* the (PDEng) trainee to become a data scientist in industry".

In the two-year program PDEng DS, training in the first year takes place in an academic environment and in an industrial setting by a final project in the second year. The first year offers a realistic, relevant, and safe learning environment in which trainees gain experience in the technical and non-technical aspects of their future job as a professional data scientist. In the second year trainees carry out an individual project, where they can show their learnings from the first year and continue learning at the same time.

All learning is supported by coaches and data science professionals. The program has designed the learning environment such that it is attractive to work within for trainees, coaches, and data science professionals alike, where:

- Trainees have the ambition to learn, gain and apply knowledge. They actively steer their professional and personal development, coached by experienced staff and challenged by data science professionals.
- Coaches create and are part of the learning environment enabling trainees to attain knowledge and professionalize in several areas and domains within Data Science. Coaches have the ambition to guide the learning process and transfer their knowledge and experience in their expert field of Data Science.
- Data science professionals want to be part of this learning environment by contributing their expertise and skills, and offering multiple opportunities to learn from data domains.

3.3 Educational Profile of the PDEng program Data Science

The six key points that characterize the educational profile of PDEng DS are:

1. Learning to learn
2. Learning from each other
3. Learning by doing
4. Focusing on real world problems
5. Broadening expertise
6. Developing professional skills

As such the program realizes that

- Training on technical and non-technical competencies is intertwined
- Trainees work in teams on challenges triggered by practical questions
  Because of the diversity of master studies completed by the trainees, these teams are multidisciplinary by nature
- Various specific data science areas are intertwined
- Data science professionals are involved in the learning process

3.4 Key points

1. *Learning to learn*

Data Science technology is changing extremely fast. So in future jobs, data scientists should be able to adapt quickly to this challenging dynamics. The pi-shaped engineering profile expresses the need to be broadly oriented, on one hand, and the ability to deep dive into required new technologies, domains or software tools whenever the project requires this, on the other hand. This adaptive learning attitude, where the trainee is master of own learning is called ‘learning agility’. The program regards learning agility as the most important competence in their future job and actively steers on the development of that attitude.

2. *Learning from each other*

The starting point is the conviction that all trainees enter the program with a lot of knowledge already acquired from studies and work experience. Bringing all that knowledge together results in an enormous amount of knowledge to be shared and in the benefit of each generation of trainees.
In the PDEng DS learning environment activities are integrated. Trainees are encouraged to make use of each other’s expertise, background, and newly gained knowledge.

3. *Learning by doing*

To digest theory is easy, to apply it in practice is often the more difficult part, and factually can only be learned by being confronted with strengths and limitations in a hands-on setting. The educational structure of the first year of the program works mainly with real life projects. An industrial partner proposes an actual and relevant challenge for his/her company. The trainees work on the proposed challenge generally in a team in order to reach a solution that presents added value for the company. In such challenge trainees aim to consider the whole data value chain from data acquisition to data usage.

4. *Focusing on real-world problems*

Students should be exposed to real-world problems from a large variety of data domains, requiring different contexts, use cases, strategies, and solutions. The connection to all parts of Dutch Industry here is of great importance. Especially, the exposure to actual, real life industrial problems is where future data scientists can learn and train themselves.

5. *Broadening expertise*

Data science is a huge field, and nobody can be expected to know all about everything. Trainees should be able to talk with and understand experts on the various aspects of data science at the experts' level, and to go in depth when needed.

6. *Developing professional skills*

The professional skills include team-work, planning, presentation, reporting, and the ability to apply these in the context data science is central. Law and ethics with respect to data privacy and governance are a prominent part of the daily practice of the professional data scientist.

3.5 Embedding of the educational philosophy and profile

The learning environment offers a wide range of training elements that invite the trainees actively steer their own development. They define their learning goals upfront and dynamically adapt them during their stay in the program. Thus depending on motivation and the defined learning goals, trainees design their learning paths. Indeed, focus on individual learning goals is essential in this philosophy. If the trainee has the ability to define what he/she wants to learn, so for what learning he/she is intrinsically motivated, it will affect the trainee's learning attitude in a positive way, showing a pro-active productive attitude in modules, projects, and teamwork. The overall topic is defined, but trainees define, within that context, their individual learning goals. In the traditional academic learning environments, the teachers are rather lecturing than coaching. They define the learning goals of a course and so create a reactive consumptive learning attitude.
Senior and junior trainees from different generations work cooperatively together in project teams. These teams present their findings to each other and to the industrial partners, weekly review meetings include pitches where diverse topics are shared, documents run through generations, which shows the different levels of skills of people, experience, and written reports. Trainees introduce workshops on specific topics of Data Science, in which trainees get the opportunity to share their knowledge with colleagues.
With the industrial career perspective in mind, trainees are encouraged to explain their colleagues what they have done and why, support each other with their own specific

knowledge and experience, and pose critical questions to each other. In group projects, Data Challenge Weeks, review and assessment meetings, presentations, self-organized (technical) workshop for colleagues, feedback sessions with team members, trainees support each other in their professional development.
In all modules, i.e., learning blocks, a case study is integrated. These case studies are based on real life industrial problems, with which trainees have to deal and where they make the connection between the topics in the module and industry. In the most ideal case, companies provide time and data for this from their side. Besides the technical learnings, the focus on process is equally important. Communication with the client, presenting results, chairing meetings, keeping good teamwork, stakeholder management are some examples here.

3.6 Selection and assessment

The selection of trainees by the Data Challenge Weeks organized by the program is the start of introducing the basic principles of the learning environment to the next generation of trainees. In the setting of a Data Challenge week three generations are linked: Generation N+1, the senior generation that takes the lead by preparing the challenges and coaching the six teams involved, Generation N, the junior generation that has a major task the execution of the challenges together with the applicants who are invited to participate and be assessed. The applicants who are successfully assessed form the next generation, Generation N-1.
In order to recruit the proper people, the assessment of applicants should cover the main ingredients of the program, which are technical knowledge, teamwork, oral and written communication, intrinsic motivation, coach-ability, and learning attitude. The Data Challenge Week was designed such that the program management can come to a conclusion on these points.

3.7 Training and Supervision Plan (TSP)

At the start of the program the trainee draws up his/her individual Training and Supervision Plan (TSP)[2]. In this plan the trainee's learning goals are summarized. In order to reach these goals, this plan should also offer the links to the activities offered by the program (modules, workshops, case studies, skill trainings) in the first year and activities that the trainee selects individually or sets up together with peers (E-learning, Kaggle competition, summer schools, Hackathons). The training and supervision plan is input to the four month evaluation sessions. In these individual sessions, the program manager and the program coach evaluate the rate of progress of the trainee in terms of the learning goals set and define new learning goals.

[2] For a TSP template, we refer to Appendix C.

In the second year, the learning environment changes drastically. The trainee starts his/her individual twelve month project and performs his/her tasks at the company. To enhance learning also in the second year, trainees of each generation that starts up the final projects meet every month to discuss their project process, issues that they meet in the course of the project, and to address in how far they accomplished the personal learning goals that they defined for the past month.

Chapter 4 | **Learning Activities**

4.1 Workshops

Professional development is the key elements underpinning the learning philosophy adopted by the program. The program is designed such that it is attractive for trainees, coaches, and data science professionals alike, where

- Trainees have the ambition to learn and apply knowledge. They actively steer their professional and personal development, coached by experienced staff and challenged by data science professionals.
- Coaches create and are part of a learning environment enabling trainees to attain knowledge and professionalize in several areas and domains within Data Science. Coaches have the ambition to guide the learning process and transfer their knowledge and experience in their expert field of Data Science.
- Data science professionals want to be part of this learning environment by contributing their expertise and skills, and offering multiple opportunities to learn from data domains.

As such the program realizes that

- Training on technical and non-technical competencies is intertwined
- Trainees work in teams on challenges triggered by practical questions
- Because of the diversity of master studies completed by the trainees, these teams are multidisciplinary by nature.
- Various specific data science areas are intertwined
- Data science professionals are involved in the learning process

Apart from the modules that concentrate on a well-specified area of data science as described in Chapter 4, the program offers the following workshops:

- data challenge week
- case study
- in-company data challenge
- interview workshop
- data quick scan

These workshops are started up only after a carefully arranged preparation in the form of trainings on team creativity, team building, stakeholder analysis, time management, interviewing styles, team roles, and pitching.

4.1.1 Data Challenge Week

The data challenge weeks are organized twice a year for about 40 participants, 20 trainees and 20 applicants. These weeks start on a Friday and end the Friday thereafter. The challenges are executed from Friday until Wednesday; Thursday and 2nd Friday the application interviews take place. During the week the participants are accommodated in a bungalow park.

- Industrial host – providing the datasets to be challenged
- Six teams – consisting of a mix of trainees and applicants
- Team leaders – defining in preparation the data challenge themes on basis of the datasets provided
- Social event – working and having leisure time in a bungalow park
- Assessment – selecting applicants on basis of observations during the week, electronic assessment of personality and motivations and a writing assignment, presentation, and an application interview
- 1st Friday – presenting the themes by the team leaders, introducing the team division
- Wednesday – having the team presentations by the applicants in the presence of representatives of the industrial host with focus on context, results, and conclusions
- Report – being delivered two weeks after the ending of the week

4.1.2 Case Study

The case studies are organized within the scope of a module or separate from it. Case studies are executed on request of a company, business, or institute. They are characterized by:

- Duration – three to five weeks, trainees spend 25 hours per week
- Way of working – in a team of two or three trainees, integration of skills
- Scope – the full data-to-value chain
- Purpose – finding or creating value from data, clarification and ideation, getting the story clear
- Result – design of a data protocol or a prototype software application, its development and implementation
- Side effect – getting acquainted with a data domain and people within the setting of a company/business/institute

4.1.3 In-house Data Challenge

The in-house data challenges are executed at a company, business, or institute, typically, in case confidential data is involved, i.e., hospital data, tax office data, police data, OM data, GGD data, public transport data. These challenges are characterized by

- Duration – one week
- Way of working – team of three or four trainees, integration of skills
- Scope – a prepared and cleaned dataset from the company/business/institute
- Purpose – finding or creating value, clarification and ideation, getting the story clear
- Result – data exploration and data analytics
- Side effect – getting acquainted with a data domain and people within the setting of a company/business/institute

4.1.4 Interview workshop

An interview workshop is executed at a company, business, or institute. Two teams of two trainees do interviews with its representatives. The careful selection of these representatives is part of the preparation and very much related to the set purpose of the workshop.

General characteristics of the workshop are:

- Duration – three days with two days of 6 interviews per day and one day to come to conclusions and recommendations
- Way of working – two teams of two do the interviews, each interviewee is interviewed twice
- Purpose – finding the relevance of the fields of data science for company/business/institute, exploring data value chains and how they can be enriched by (external) data integration
- Results – description of data value chains with conclusions and recommendations, exploring the pros and cons concerning the use of data in decision processes, description of the current organizational structure and in what direction it should develop if opportunities of use of data is fully exploited.

4.2 Modules in the PDEng program Data Science

4.2.1 Introduction

A module consists of a study load 150 hours spread over 6 weeks. Students spend 25 hours per week on a module. The 150 hours are typically spend as follows:

- 30% tutorials
  - 20% instructions
  - 10% exercises
- 70% assignment
  - 50% execution
  - 10% feedback
  - 10% reporting

4.2.2 Tutorials

The program expects that its expert lecturers enable students to reach a next level in their understanding of the topic. A module is not a standard Bachelor program course, where the basics are presented to the class-room and students apply these to toy-problems. Also, a module is not a standard Master program course, where the latest and greatest research developments are discussed in depth, such that students can start to do research themselves. The tutorial / professional course / workshop is the better metaphor. Lecturers should provide students with:

- A broad overview of their field, including its history, the current state, and promising developments;

- Reflections on relations with other fields. How does the field fit in the data science cycle, what are limitations and constraints?
- A critical discussion on various approaches. What methods work and which do not when applied to real-world cases?
- Strategies to follow for solving problems.
- Pointers to resources, where additional information can be found: e-learning courses, websites, books, software platforms and tools, etc.
- Exercises to obtain hands-on experience.

A possible format is one-day workshops. The day starts with a 1-2 hour talk of the expert, followed by a practical exercise of 4 hours, and finishes with a presentation and discussion of results. The first week could be spend with these, or, when more convenient, each week could start with such a workshop, following the phases of a typical project.

4.2.3 Assignment

The assignment is the central part of a module.

Structure

- Team-work. Students work in teams, consisting of 3-4 students. The program aims at about 12 students per module.
- *One assignment.* In one module, teams do one assignment. Obviously, one could do many exercises in a course, here we aim at doing projects, starting from a problem and working towards a solution.
- *Different assignments per team.* Each team works on a different assignment. Teams do focus on the same domain and the same data, but here variation can be brought in by focusing on different problems, different target audiences, and different aspects of the data. We hope that this variation brings in the awareness that the different problems, target audiences, and aspects of the data lead to different solutions.
- *Planning.* Teams themselves are responsible for defining a plan, intermediate goals, and allocation of resources. Making the right choices on these is an important learning objective of the overall program. The lecturer gives feedback on the plans of the teams.
- *Schedule.* A typical schedule for a seven week assignment is:
  - *Week 1:* Student are introduced to the assignments, form teams, select problem, make a plan, decide what additional knowledge is needed, start with the assignment, get feedback by the lecturer;
  - *Weeks 2-5:* Students work and study, and get feedback by colleagues and the lecturer;
  - *Week 6:* Students work on the report and the presentation, give the final presentation, discuss and reflect on the module.

Real-world problems

The program strongly advocates the use of real-world cases, brought in by external parties. There is an important role for *The Customer*. Students are highly stimulated when such person is introduced and linked to the Module. The Customer:

- presents the domain, the data, and the problems and questions to be attacked in the first week
- represents one of the partners in the ecosystem
- is available (via email) for answering questions from students during the execution of the assignment
- is present at the final presentation, and gives feedback on the results

The problem to be attacked should:

- *be challenging and open-ended*. The answer and the approach to be followed should not be clear from the start, picking the right approach and making choices for techniques is part of the challenge.
- *be defined in terms of The Customer*. Starting point is the problem of The Customer, not a technical issue.
- *require knowledge and expertise* from the topic of the module;
- *have the right scale*. Students should be enabled to achieve concrete results at the end of the module. Bringing in different levels in the problem is a good approach here, also because it is often hard to predict in advance what is achievable.

Data

The data to be handled should be

- *Available.* This is obvious, but requires careful planning long before the start of the module. Customers are typically overoptimistic here. Also, confidentiality and privacy issues have to be dealt with. If The Customer is not able to provide data from his own organization, possibly public domain data sets can be used with similar characteristics.
- *Big.* The three V's of Big Data are Volume, Variety, and Velocity. Dependent on the aim of the module, a balance has to be found. The data should have such volume that manual processing is not possible, but it does not need to have TB scale. On one hand, one time-series consisting of time-stamps and a single measurement is not challenging, on the other hand, students should not need to spend weeks on understanding and cleaning the data. A clear description of the format of the data should be available. It is motivating for the students, if they can quickly understand and relate to the meaning of the data.
- *Relevant.* Just like the problem assignment itself, the data should provide opportunities for students to apply the technology from the topic of the module.

Feedback

The lecturer provides feedback to students during the assignment.

- Each week a meeting is held, taking about one hour
- If needed, students send material they want to discuss in advance to the lecturer
- During the meeting, the progress (what has been done) and plans (what is to be done) are discussed

- The lecturer provides guidance and feedback. Students get suggestions for possible approaches, additional information, etc. Students are offered multiple suggestions, rather than directives; finding out what works best for their case is part of the challenge.
- The lecturer observes the group process, and takes action when people dominate or do not participate actively.
- After each meeting, teams prepare a short summary, describing results, plans, and action points, for the team as well as the lecturer.

Deliverables
For each team the results of an assignment are:

- *A report*. This should be concise and clear (max. 15 pages, appendixes excluded), and contain an executive summary (1 page), management introduction (max 3 pages), the main body (max 10 pages), conclusions and recommendations, suggestions for future work (max 2 pages), and appendices. Target audiences are The Customer and fellow students. What information should be given to these such that a next group can continue where they finished?
- *A presentation*. A presentation of 30-45 minutes is given, where all group members participate. The presentation should give a clear overview of the problem, the process, and results, with again The Customer and fellow students as target audience.
- *Demo's / prototypes / …* If possible, working towards products that can be demonstrated would be great.

Evaluation
After the final presentation:

- Students prepare a reflection (1-2 pages) on the course. What they learned, what they need to work on, what they liked, and what can be improved on the module.
- The lecturer reads the reports and reflections, and discusses these with the teams.
- Students are evaluated on basis of a module evaluation form that is discussed with them individually

4.2.4 Professional skills

Integration of skills is the motto of the program. Thus development and integration of professional and personal skills is an important aspect. In the writing part of each module, students are supported by the course Technical Writing & Editing. Students learn how to put their findings in a technical report. Separate workshops on aspects such as conflict management, presentation, feedback, team rolls, peer coaching, and negotiation are organized. If one topic fits well to a certain module, lecturers are encouraged to integrate such a workshop in their module.

4.2.5 Activities

The program expects the following activities from the lecturer:

- Before the module:
  - Selection of a suitable customer;
  - In cooperation with The Customer, prepare assignments and data-sets;
  - Preparation of material for workshops: slides, exercises, resources.
- During the module:
  - Provide workshops: tutorials, exercises, discussion;
  - Present (with The Customer) the assignments;
  - Give feedback to teams.
- After the module:
  - Judge the reports and give feedback to students
  - Evaluate students individually on basis of the module evaluation form.

4.3 Coaching

4.3.1 Technical Writing & Editing

Each activity in the program (module, workshop, and project) requires documentation. Thus, naturally, technical writing is integrated in all activities. The program developed report templates consisting of executive summary, management introduction, main body, conclusions & recommendations, and appendices. The goal of the coaching course Technical Writing & Editing is to let the trainees develop the ability to write a management report, a progress report, and a project proposal on basis of the proposed templates. Also, they will develop the skills to be critical on one's own writings and give feedback to peers. The course consists of 7 plenary sessions of 3 hours and 3 individual sessions of one hour per trainee. These sessions are planned throughout the first year, in November, March, and June. There are 6 writing assignments, each assignment relates to an activity within the program.

Assignment 1 – Write a self-introduction
Assignment 2 – Write a management introduction on your current case study/workshop
Assignment 3 – Write a management report with focus on executive summary and management introduction related to your project in the Data Challenge Week
Assignment 4 – Write a progress report on your current case study/workshop
Assignment 5 – Write a project proposal related to your current case study/workshop
Assignment 6 – Write a full report on your current case study/workshop

During the second year of the program the Technical Writing & Editing coach can be consulted for advice on matters related to reporting skills and feedback on progress reports to be written in the course of the final project.

4.3.2 Ethics & Law

In many of the case studies and projects carried out in the frame of the program, ethical and legal issues related to data privacy and data governance play an important major role. Thus there is an E&L coach associated to the program, who can be consulted on matters related to ethics and law. The E&L coach arranges two monthly discussion sessions with the trainees concentrating on questions as

1. Think as yourself working as a data scientist in your own business, in a few years from now. What would you like to learn from a lawyer, in order to make that you will be more efficient in your work?
2. Think as yourself working as a data scientist in a research project, at JADS or at any research institute. What would you like to learn from a lawyer, in order to make that you will be more efficient in your research?
3. Think as yourself working as a lawyer (yes!). How would you use your data science expertise in order to improve the legal system? (Don't hesitate to answer, just take the legal system as you know it now, the intention is to start a discussion on the infusion of data science in the law: what can lawyers learn from data scientists?)
4. What is the difference between ethics and law and why should we combine them?

Chapter 5 | **Module contents**

This chapter contains a description of the modules that are the backbone of the first year of the program. The focus in this chapter is on the module content and not on their organization. Since modules have their own characteristics, e.g., whereas the module on data visualization is design oriented with focus on use of applications, the module on modeling and simulation has a significant knowledge acquisition and awareness component, the description of the modules is not uniform. Yet, in each description there is reference to study material and available software.

5.1 Introduction-to-Data Science module

This module addresses the data value chain, the scoping of (big) data projects, and, more specifically, the interrelations between the topics introduced in the other data science specific modules. Typically, the data value chain for different data domains, such as Finance, Agri-Food, Health, and Logistics is explained by hands-on examples. Participants of the program are challenged to define big data projects by literature study on daily practices. The module distinguishes from the other modules in the program in that it does not address specific techniques.

A few open source references are:

1. The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches, Edward Curry, Insight Centre for Data Analytics, National University of Ireland Galway, 2016 https://link.springer.com/content/pdf/10.1007/978-3-319-21569-3_3.pdf
2. Introduction to Data Science, Wray Buntine, Monash University, 2015 https://topicmodelsdotorg.files.wordpress.com/2016/01/introds_110116.pdf
3. Setting Up a Big Data Project: Challenges, Opportunities, Technologies and Optimization, Roberto V. Zicari, Marten Rosselli, Todor Ivanov, Nikolaos Korfiatis, Karsten Tolle, Raik Niemann and Christoph Reichenbach, 2016
   Big Data Optimization: Recent Developments and Challenges, Studies in Big Data 18, Springer Verlag
4. How to run a Big Data project. Interview with James Kobielus, 2014 http://www.odbms.org/blog/2014/05/james-kobielus

5.2 Data Mining

Data Mining is the subfield within Data Science that is most strongly connected to the data modeling techniques emerging from the classical field of Statistics. The purpose of performing data mining techniques are amongst others data segmentation, data classification, predictive and rule based modeling. For most standard data mining techniques and algorithms R and Python are the most suitable software platforms. The basic techniques within Data Mining are:

1. Linear Regression, Logistic Regression, Clustering (hierarchical and non-hierarchical) and k-Nearest Neighbors
2. Decision Trees and Random Forests as segmentation-with-a-purpose strategies

3. Neural Networks for predictive classification and feature extraction, Support Vector Machines, Rule Induction Systems
4. Deep Learning, Convolutional Algorithms, Bayesian Networks, Statistical Learning, Genetic Algorithms

From the many references, we mention the one by Alex Berson et al, and the references therein: An Overview of Data Mining Techniques, excerpted from the book Building Data Mining Applications for CRM by Alex Berson, Stephen Smith, and Kurt Thearling, 2005
http://weber.itn.liu.se/~jimjo94/courses/TNM048/documents/DM-Techniques.pdf

Relevant open source courses
Mining Massive Data Sets / Stanford Coursera & Digital & Book $58
Introduction to Information Retrieval / Stanford Digital & Book $56
OSDSM Specialization Web Scraping & Crawling
Machine Learning Ng Stanford / Coursera
A Course in Machine Learning UMD / Digital Book
The Elements of Statistical Learning / Stanford Digital & Book $80 & Study Group
Neural Networks Andrej Karpathy / Python Walkthrough
Neural Networks U Toronto / Coursera

Software Packages
scikit-learn - Tools for Data Mining & Analysis
Data Science in IPython Notebooks (Linear Regression, Logistic Regression, Random Forests, K-Means Clustering)

5.3 Data Visualization

The aim of this module is to increase knowledge and experience with respect to visualization of the students participating, such that they:

- Know the basic concepts, ideas, and techniques
- Are able to find and judge more information when needed, i.e. , how to find your way in the zoo
- Are able to use visualization for the exploration of data sets
- Are able to develop prototype visualization tools for end users

Literature on visualization is extremely rich, many open source courses have been developed, and there is an abundance of visualization tools. Students are free to use what they want for implementing visualizations. There are many different options, and also, the landscape changes quickly. Understanding opportunities and limitations of visualization in general is more important than becoming an expert in tool X or technology Y. Nevertheless, to make something and gain experience, the module offers the opportunity to make dirty hands.

References
Tamara Munzner: Visualization Analysis and Design, AK Peters, 2014.
Colin Ware: Information visualization: Perception for design, 3rd edition, Morgan Kaufmann.

Edward Tufte: Envisioning Information (1990), Visual Explanations: Images and Quantities, Evidence and Narrative (1997), The Visual Display of Quantitative Information (2001), Beautiful Evidence (2006).
Stephen Few: Show me the Numbers (2012), Now you see it (2009), Information Dashboard Design (2013).

Online courses
Tamara Munzner, UBC: http://www.cs.ubc.ca/~tmm/courses/547-15/
Jeffrey Heer, Univ. Washington: http://courses.cs.washington.edu/courses/cse512/14wi/
John Stasko, Georgia Tech: http://www.cc.gatech.edu/~stasko/7450/
Miriah Meyer, Univ. of Utah: http://www.sci.utah.edu/~miriah/cs6964/

Software Tools
Tableau, https://www.tableau.com/trial/data-visualization
Spotfire, https://spotfire.tibco.com/
SynerScope, http://www.synerscope.com/
Keshif, http://keshif.me/

5.4 Data Analytics

In data analytics four types of analytics are distinguished:

- Prescriptive analytics that reveals what actions should be taken. Deliverables are recommendations what decisions best to take
- Predictive analytics that results in likely scenarios of what might happen. Deliverables are predictive forecasts.
- Diagnostic analytics that reveals past performance to determine what happened and why. The result of the analysis is often an analytic dashboard.
- Descriptive analytics that reports on what is happening now based on incoming data. To mine the analytics, you typically use a real-time dashboard and/or email reports.

In this module, participants will be confronted with a broad range of Business Analytics techniques and how to apply them to obtain smarter decisions. After completing the module students should be comfortable with:

- Using of Data Analytics methods and identifying, evaluating, and capturing business analytics opportunities that create value
- Using optimization to support decision-making in the presence of a large number of alternatives and business constraints
- Thinking critically about data and the analyses based on that data
- Deciding what analytics model or method to use when

In this module, a holistic view is presented on predictive and prescriptive analytics. Special emphasis is on optimization, the issues of scale, that is, how the size of some current large datasets prevents the direct application of some data analytics methods, and possible solutions to get around these issues.

Topics to be covered include:

- What is Data Analytics?
- The distinction and relationships between descriptive – predictive – prescriptive analytics
- How is Data Analytics related to Data Science?
- Competitive edge of Data Analytics.
- Prescriptive analytics techniques:
    - The important optimization models and techniques
    - How to cope with uncertainty in the data?
    - How to balance different competing objectives: Multi-objective optimization.

References

Competing on Analytics, Thomas H. Davenport and Jeanne G. Harris, 2007
Analytics at Work, Thomas H. Davenport, Jeanne G. Harris, Sir Robert Morison, 2009
The Analytics Edge Hardcover, Dimitris Bertsimas, Allison O'Hair, William Pulleyblank, 2016
Convex Optimization, Stephen Boyd and Lieven Vandenberghe, Cambridge University Press, http://stanford.edu/~boyd/cvxbook/
AIMMS Modelling guide
https://download.aimms.com/aimms/download/manuals/AIMMS3_OM.pdf

Video Lectures

Many lectures of Stephen Boyd on Convex Optimization
General lectures on Business Analytics:
https://www.youtube.com/watch?v=ynLs5lxI0P4
https://www.youtube.com/watch?v=lXdCnOQCCAE

Online courses

http://techcanvass.com/Courses/Business-Analytics-with-R-Programming-course.aspx
https://www.coursera.org/specializations/business-analytics

Toolkit

http://www.palisade.com/trials.asp

5.5 Modeling and Simulation

Modeling and simulation evolved from tool to discipline in less than two decades. Along with the technology boom of the 1990s, the ability to use models and simulations in nearly every aspect of life evolved rapidly. Modeling and simulation has become a capability to better understand human behavior, enterprise systems, disease proliferation, and many more. To get acquainted with the discipline, theoretical underpinnings must be understood and practical domains explored. The aim of this module is to provide students with a survey of the discipline and its applications in practice. The module is based on the book

MODELING AND SIMULATION FUNDAMENTALS
Theoretical Underpinnings and Practical Domains
Edited by John A. Sokolowski and Catherine M. Banks

http://onlinelibrary.wiley.com/book/10.1002/9780470590621

The focus of the module is on the following chapters of this book:
Chapter 1 – Introduction to modeling and simulation
Chapter 2 – Statistical concepts for discrete event simulation
Chapter 3 – Discrete-event simulation
Chapter 5 – Monte-Carlo Simulation
Chapter 6 – Systems modeling: analysis and operations research
Chapter 10 – Verification, validation, and accreditation

Additionally, the book

Simulation for Data Science with R
by Matthias Temple

https://www.amazon.com/Simulation-Data-Science-Matthias-Templ/dp/1785881167

aims to explore advanced R features to simulate data to extract insights, and provides a comprehensive coverage of several R statistical packages.

5.6 Process Mining

Process mining provides a means to improve processes in a variety of application domains. The two main drivers for this new technology are, on one hand, more and more events being recorded thus providing detailed information about the history of processes, and, on the other hand, in most organizations the need to improve process performance (e.g., reduce costs and flow time) and compliance (e.g., avoid deviations or risks). The practical relevance of process mining and related interesting scientific challenges make process mining a fascinating topic for future engineers.
The learning objectives of the module are to have a good understanding of process mining and be able to apply process mining techniques in a real-life project.
The module aims to present the following basic concepts and algorithms of process mining:

1. Petri nets
2. Process discovery
3. Conformance checking
4. BPMN modeling notation
5. Comparison of process Invariants
6. Multi-perspective process mining

Online courses
https://www.coursera.org/learn/process-mining

Open source software
http://www.promtools.org/doku.php

References
Process Mining: Discovery, Conformance and Enhancement of Business Processes by W.M.P. van der Aalst, Springer Verlag, 2011
http://www.springer.com/gp/book/9783642193453
http://wwwis.win.tue.nl/~wvdaalst/courses/howtogetstartedwithprocessmining.pdf

5.7 Design for Operations Management and Logistics

The goal of this module is to teach student to use Industrial Engineering methodologies (statistical techniques, simulation, etc.), quantitative modeling and design, when analyzing the performance of a company. Students should be able to:

- Make the translation of subjective, qualitative statements in claims on the operational characteristics and performance, which are supported by quantitative analyses
- Integrate the knowledge gained in prerequisite modules, where the student decides him/herself what to use
- Determine the applicability of several models/methods
- Determine the right functions and parameter values based on a statistical analysis
- Implement mathematical models and to understand and interpret the results.

Building on the model-based-knowledge obtained previously, this module trains the students to develop control structures for complex Industrial Engineering systems (e.g. inventory, production, transport), taking into account the technological, economical, and organizational constraints that a design has to satisfy. The students learn to evaluate the designs and select one that best satisfies the design requirements in terms of performance and costs.

The module concentrates on:

- How to construct a descriptive model (input variables, output variables)
- How to derive from this a prescriptive model (objective function, decision variables, restrictions)
- How to decide which model(s) from literature can be applied or adjusted in specific situations in practice
- Under which assumptions a model yields a correct solution to the problem
- How (empirical) input data for a model can be collected, cleaned, and analyzed, and if applicable compared with a theoretical distribution function, which can be applied in a decision support model
- How to interpret the results obtained with a model (worst case or best case)
- How to statistically test the model assumptions
- How to verify and validate a model
- How to perform a sensitivity analysis
- How to implement a model in an Excel spreadsheet

The course starts with sessions in which a design approach is explained, and a plenary session in a company in which the real-life situation is introduced. Next, trainees develop a conceptual design giving the structure of the new system, then, develop detailed designs for the subsystems, and finally integrate the detailed designs into a working concept for the whole system. In each of these stages, real-life data, and feedback from company managers and lecturer, are used to test and adapt the solutions.

Online courses
https://www.coursera.org/learn/planning
https://www.coursera.org/learn/operations
https://www.coursera.org/learn/supply-chain-logistics
https://www.coursera.org/learn/supply-chain-principles
https://www.scl.gatech.edu/education/professional-education/online-courses

- Supply Chain Fundamentals: Supply Chain Management Principles
- Supply Chain Fundamentals: Customer Service Operations
- Supply Chain Fundamentals: Transportation Operations
- Supply Chain Fundamentals: Supply Management and Procurement
- Supply Chain Fundamentals: Inventory Management
- Supply Chain Fundamentals: Demand Planning
- Supply Chain Fundamentals: Manufacturing and Services Operations

5.8 Data Engineering

The main objective of the data engineering module is to get participants acquainted with the following three main data engineering concepts:

- The relational database model & NoSQL database models.
  This first topic revolves around the classical relational database model that did arise in the 70s and has attained huge traction over the past 40+ years. The model is catered toward storing, manipulating, querying and managing structured datasets. NoSQL extends the relational model and allows for other database models including the column-, document-, key-value-, graph-, and multi-model database models.
- Data integration & quality.
  This topic addresses the issue of integration various distributed and heterogeneous data sources that may contain (semi) structured and even unstructured data. Both batch processing (ETL/data-warehousing) and streaming processing (complex events) are discussed as underpinning technologies.
- Big-data processing frameworks & analytics basis.
  Big Data frameworks cater for distributed and scalable processing of large datasets (big data) across computing clusters exploiting reasonably “simple” programming models – analytics to be designed in the scope of these frameworks will have specific features and construction style that are introduced in the scope of this module. Contemporary technologies such as Hadoop/MapReduce and Spark are introduced and used as an example.

At the end of the module the trainees have understanding of the theories, models, and techniques underpinning of each concept, their (dis)advantages, and understand when to apply which model for which (business) problem. The module assumes self-study and revolves around implementing real-world case studies.

Online courses
https://www.coursera.org/specializations/jhu-data-science
https://www.coursera.org/specializations/big-data
https://bigdatauniversity.com/learn/big-data/
https://bigdatauniversity.com/learn/spark/

Data Manipulation at Scale:
https://www.coursera.org/specializations/data-science
and others on an as-needed basis

5.9 Data Statistics

Data should not be confused with random variables. Random variables live in the theoretical world. Careful consideration of the connection between models (theoretical world) and data (real world) is a core component of the art of statistical practice and of the science of statistical methodology as well.

"Let us assume the data are normally distributed" is a shorthand for assuming "the variability of the data is adequately consistent with variability that would occur in a random sample."

Fisher introduced the idea of a random sample drawn from a hypothetical infinite population, and Neyman and Pearson's work encouraged statisticians to drop the word "hypothetical" and instead describe statistical inference as analogous to simple random sampling from a finite population.
The module focuses on statistical thinking and statistical procedures that are abstractly defined in terms of mathematics, but are used in conjunction with scientific models and methods to explain observable phenomena. There is a hypothetical link between variation in data and its description using statistical models. An important aspect of statistical thinking is the assumptions on the variation in the data with the underlying question whether they describe the variation in the data reasonably accurately.
In the module statistical models of regularity and variability in data are introduces as a way to express knowledge and uncertainty, via inductive reasoning. It means identification of the source of statistical inference as the hypothesized link between data and statistical models

Main contents of the module are:

1. *Statistical Inference in Data Science (with emphasis on hypothesis testing)*

Reference: Brian Caffo, A companion to the Coursera Statistical Inference Course
http://leanpub.com/LittleInferenceBook

2. *Experiment Design and Big Data*

Reference: Drovandi, Christopher C., Holmes, Christopher, McGree, James M., Mengersen, Kerrie, Richardson, Sylvia, & Ryan, Elizabeth G. (2017) Principles of experimental design for Big Data analysis. Statistical Science, 32(3), pp. 385-404
https://projecteuclid.org/euclid.ss/1504253123

3. *Bootstrap methods*

Refrence: Diego Kuonen, An Introduction to Bootstrap Methods and their Application
WBL in Angewandter Statistik ETHZ 2017/19' — January 22 & 29, 2018
https://www.ethz.ch/content/dam/ethz/special-interest/math/statistics/sfs/Education/Advanced%20Studies%20in%20Applied%20Statistics/course-material-1719/Nonparametric%20Methods/lecture_2up.pdf

4. *Bayesian Statistics*

Reference: Theo Kypraios, A Gentle Tutorial in Bayesian Statistics
https://www.scribd.com/document/132660323/A-Gentle-Tutorial-in-Bayesian-Statistics-pdf
Rens van der Schoot, et al
https://www.statmodel.com/download/introBayes.pdf

In conclusion

The above described modules contain many data science techniques, references to online courses, and open source software that help understanding, exploring, modeling, and analyzing the data value chain. Basic techniques from Data Exploration, Data Cleansing, Data Statistics, and Data Engineering are not explicitly included. All modules in some form or another touch upon these techniques.
The following *online courses* contain an inviting introduction to these:

- Harvard Video Archive & Course
- Statistics I Princeton / Coursera
- Introduction to Databases Stanford / Online Course
- SQL School Mode Analytics / Tutorials
- SQL Tutorials SQLZOO / Tutorials
- Intro to Hadoop and MapReduce UW / Lectures on MapReduce

Chapter 6 | **PDEng project Data Science**

6.1 Objective

The candidate should show his/her ability to carry out a long term project individually within an industrial or business environment. The project is related to a real life problem and is constrained by time planning, project planning, and project deliverables. The project is executed as contract research[3] with the university and thus based on a formal contract. The project has a substantially scientific level and is devoted to problems of highly innovative level. The project is executed under a nondisclosure agreement.

6.2 Organization

There is a project team, consisting of the trainee, university supervisor(s) and problem owner(s) as representative(s) of the company.
The *company* takes care of a working space for the PDEng-student.
The *project supervisor* assists in drafting the problem description, project planning and time planning, and guides the project during the twelve months. The project supervisor can involve other members from the departmental staff in the project, if their scientific expertise and support is requested and wanted.
The *program management* offers the necessary administrative support, and finalizes the contract and checks whether the contract is established in a sound way. The program management arranges evaluation meetings, peer coaching sessions, brainstorms, and progress presentations.
The company appoints a *problem owner*, who is the responsible party on behalf of the company. This way, there is the guarantee that the company keeps genuine interest in the progress and results of the project.

6.3 Project description

Each project is carried out on the basis of a contract between the company and the university. In an enclosure to the contract, the problem to be addressed in the project is described. This project description is drawn up by the trainee, project supervisor, and industrial supervisor. The final draft of the project description is the responsibility of the trainee. In preparation to the project the trainee delivers a project description that contains the following items:

- Global description of the company and of the division within the company where the project takes place together with its most important activities
- Sketch of the context of the project
- Detailed description of the project problem, goal, and deliverables
- Description of the quality of the expected final result with global qualifications that have to be satisfied.
- Time scheme with milestones and a rough project-planning schedule.

[3] The template contract is presented in Appendix D

6.4 Time scheme

Prior to the start of the twelve month project clear arrangements are made with respect to supervision both by the program administration and by the company. During the project, the student keeps a logbook, containing all decisions and agreements made at every meeting of both supervisors and trainee.
After two months, the student organizes a brainstorm session for his/her colleagues, for staff member of JADS, and representatives of the company involved in the project. After four and eight months, the trainee gives an interim presentations in the project seminar and writes an interim report. This way his/her findings are presented to the larger audience of colleague students and staff members of the department. The long term project is concluded with a presentation in the company and thereafter the project is evaluated.

- The trainee completes the final report.
- The trainee and project supervisor discuss the project regularly.
- The trainee pays a visit to the university at least once a week.
- The industrial supervisor, project supervisor and trainee meet on a regular basis. At least once every four weeks the trainee sets up a meeting, accompanied by an agenda, in which both supervisors, the trainee and, if so desired, one or more experts participate.

6.5 Final report

The report is written according to the guidelines for writing a technical report as formulated by the PDEng-program and trained in the TW&E course. Under all circumstances the final report is treated confidentially.

6.6 Evaluation

In the course of the project, three evaluations take place. The first two evaluations are scheduled after four and eight months. The trainee delivers two progress reports, which are part of the interim evaluation. The final evaluation is scheduled at the end of the project after the trainee has delivered a final report.
The interim evaluation is conducted on the basis of an evaluation interview and an evaluation form. The outcome of the interim evaluation is discussed with the trainee, and if necessary project goal and time planning are adapted.
The final evaluation is conducted by a graduation committee that is assigned for this evaluation task by the program management. The committee judges the degree in which the assignment was made a success within the set time, the way in which the trainee operated in the company, the scientific and technical quality of the work, its design features, and last but not least, the quality of the report and the presentation. Supervisors and problem owner are members of the graduation committee. The judgement of the committee is registered in an evaluation form, see Appendix E for an example.

Chapter 7 | Reflection and outlook on the future

The PDEng program Data Science offers a rich and unique learning environment to young professionals to further expand their skills in both technical and non-technical areas. The program focuses on six fundamental focal points, namely, learn to learn, learn from each other, learn by doing, real world challenges, broad expertise, and professional skills development. The PDEng program Data Science addresses these six focal points through the rich selection of activities, from courses given by TU/e and TiU experts to online courses, from soft skills coaching to real-life case studies, and from team working setups to the self-governing graduation project. To maintain this educational philosophy means we have to find solutions to several challenges.

A stable and dedicated educational team around the program is a condition to fulfill our educational philosophy. Members of the educational team are involved in the setting up and delivering the modules content, that is, the teaching activities and the supervision of the case studies attached to a module. Trainees continue their professionalization throughout the second year when the graduation project is planned. Currently, the program director is closely involved in the supervision of almost all trainees in their final project. To keep the high quality standards of project supervision and project deliverables the work load has to be shared between program managers and the educational team. Our challenge is to attract dedicated academic staff and to keep them into the educational team of our program.

To keep the quality of our educational program at high standards the management team of PDEng program Data Science continuously searches ways to improve the processes within the program. Trainees who join the PDEng program Data Science are selected on basis of an application and a week-long assignment, the Data Challenge Week. Every Data Challenge Week ends with several reflection moments, including feedback from the industrial hosts. Through the Training and Supervision Plan we establish and maintain the quality of trainee's professional training. Every module ends with a reflection moment on the content and case studies, in which trainees and educational team members are involved. Lessons learnt translate into improvements points for the next module sessions. We organize regular round table events with our business partners in which we discuss aspects such as the role of data scientist in business organizations[4], the impact of GDPR in data science projects[5], the profile of the data scientist as envisioned by the business community[6], and the role of Data Science in Research and Development[7].

In the last year the PDEng program Data Science has grown to 40-45 trainees. Trainees who start this program have diverse educational, cultural, and work backgrounds. What unites them ultimately is their drive to further professionalize in the field of Data Science. Their number, their diversity, and their drive are factors that contribute to building a strong

[4] Round table June 2019, "*What does it take to implement data science in your organization? What would your team(s) need in order to unlock the full potential of data scientists?*"
[5] Round table November 2018, "De GDPR een zorg of een zegen"
[6] Round table October 2017, "The Data Scientist Profile"
[7] Round table February 2017, "What added value do you expect from the introduction of elements of Data Science in your Research and Development activities in the near and in the long future?"

community focused on leaning from each other. Maintaining such community is a challenge. Our program must keep on attracting young professionals from all over the world while keeping a streamlined program administration. We are taking steps to improve the office administration and project acquisition process.

In line with the idea of building a multidisciplinary and multicultural program we recognize that attracting Dutch trainees is important. After all, what better way for a non-Dutch trainee is there to learn about Dutch culture and work ethic other than learning from a Dutch colleague? We are taking steps to make our program visible to Dutch students and young professionals by participating to student career days, by increasing the presence of our program on social media and professional networking platforms, and by making use of the DSconnect alumni organization.

The PDEng program Data Science has been involved with more than 50 Dutch industrial partners, government organizations, and commercial and financial services in carrying out case studies and final projects. We are proud of our wide business portfolio because we want to be forerunners in the establishment of Data Science in Dutch economy and because we believe in knowledge transfer between different economy sectors. We believe we can strive the balance between establishing long-term partnerships with Dutch companies and involving with new organizations. Acquisition and setting up the projects timely from both content and financial points of view are challenging. The program administration is continuously improving this process by working closely together with the operational management at JADS.

Appendix A | Joint statement of competence assessment and innovation

JOINT STATEMENT OF COMPETENCE ASSESSMENT AND INNOVATION

Hereby, the PDEng program Data Science, represented by its PD-coach, and the undersigning Trainee certify that with this Joint Statement of Competence Assessment and Innovation details of the competences are set out that PDEng Data Science trainees would be expected to innovate during their training.

We state that the competences may be present on commencement, explicitly taught, or developed during the course of the program. It is expected that different mechanisms will be used to support learning as appropriate, including self-direction, coach support, workshops, brainstorming sessions, seminars and Data Challenge Weeks.

The management of PDEng Data Science re-emphasizes its belief that training in competences is a key element in the development of the trainee, and that trainees are expected to make a substantial, original contribution to self-knowledge, normally leading to self-assessment. The development of wider employment-related competences should not detract from that core objective.

The statement gives the program's common view of the competences that a trainee in the program Data Science should acquire during the stay in the program in relation to:

a. General rules of conduct
b. Self-reflection
c. Learning attitude
d. Conversation skills
e. Presentation skills
f. Social skills
g. Team work effectiveness
h. Project management skills
i. Career management

Signature Trainee: Signature PD-Coach:

‘s Hertogenbosch, d.d.

See Enclosure.

Enclosure - Joint Statement of Competence Assessment and Innovation

a. General rules of conduct:
    1. Show actively one's commitment in respect to the PDEng Program Data Science
    2. Demonstrate awareness of issues relating to the rights of other data scientists, of modules and workshops, and of others who may be affected by the data science practice, e.g. confidentiality, ethical issues, integrity, attribution, copyright, malpractice, ownership of data and the requirements of the Data Protection Act
    3. Understand the processes for funding and evaluation of Personal Competence Innovation
    4. Constructively defend outcomes of applying data science techniques at seminars, and in industry and business
    5. Contribute to promoting the public understanding of the field of data science within specific data domains

b. Self-reflection - to be able to:
    1. Formulate personal learning points and strong points
    2. Be conscious of own actions and distinguishing the effect they have on others
    3. Be conscious of own opinion and being able to consciously react on the basis of this opinion
    4. Be conscious of one's own (non) verbal communication and being able to react on communication of others

c. Conversation skills – to be able to:
    1. Formulate clear goals preceding a conversation
    2. Listen, to pose relevant questions and to summarize during a conversation
    3. Negotiate and persuade
    4. Actively communicate with the problem owner

d. Presentation skills – to be able to:
    1. Construct coherent arguments and articulate ideas clearly to a range of audiences, formally and informally through a variety of techniques.
    2. Show the ability to tell the story from the data
    3. Give an informative and convincing presentation, well-tuned to the audience
    4. Present with a positive attitude
    5. Give a well-organized presentation

e. Social skills and team work effectiveness – to be able to:
    1. Take initiative
    2. Participate actively and assertively in a meeting
    3. Come to decisions
    4. Develop and maintain co-operative networks and working relationships with supervisors, industrial co-workers and peers, within the institution and the wider Research & Development industry.
    5. Understand one's behaviors and impact on others when working in and contributing to the success of formal and informal teams.
    6. Listen, give and receive feedback, and respond perceptively to others.
    7. Manage conflicts
    8. Effectively support the learning of others when involved in mentoring or modelling activities.

f. Learning styles – to be able to:
    1. Demonstrate a willingness and ability to learn and acquire knowledge.
    2. Be innovative and original in one's approach to explore, structure, visialize and analyze datasets.
    3. Demonstrate flexibility and open-mindedness.
    4. Demonstrate self-awareness and the ability to identify own training needs.
    5. Demonstrate self-discipline, motivation and thoroughness.
    6. Recognize boundaries and draw upon/use sources of support as appropriate.
    7. Show initiative, work independently and be self-reliant.

g. Project management – to be able to:
    1. Apply effective project management through the setting of modelling goals, intermediate milestones and prioritization of activities.
    2. Show leadership
    3. Summarize the essence of a meeting
    4. Summarize, document, report and reflect on progress.

h. Career management – to be able to:
    1. Appreciate the need for and show commitment to continued professional development.
    2. Take ownership for and manage one's career progression, set realistic and achievable career goals, and identify and develop ways to improve employability.
    3. Demonstrate an insight into the transferable nature of human resource skills to other work environments and the range of career opportunities within and outside academia and or industry.
    4. Present one's competences, personal attributes and experiences through effective CV's applications and interviews.

Appendix B | Program of the Data Challenge Week

Program of the Data Challenge Week

Friday

09.30 – 11.30 Reception at JADS/ Intromeeting with Program Director 5 minutes
11.30 – 11.40 Welcome & information presentation
12.30 – 13.30 Lunch
13.30 – 15.00 Work in groups
15.00 Hap & snap break
16.45 Departure to the bus
17.15 Bus departures to the bungalow park
18.15 Welcome
Dinner at the bungalows

Saturday

08.00 – 08.45 Breakfast at the bungalows
09.00 – 12.00 Work at the bungalows
12.30 – 13.30 Lunch
13.30 – 17.00 Moment with Program Director: 30 minutes per bungalow:
13.30 – 14.00 Group 1: bungalow 1 (group leader)
14.00 – 14.30 Group 2: bungalow 2
14.30 – 15.00 Group 3: bungalow 3
15.00 – 15.30 Group 4: bungalow 4
15.30 – 16.00 Group 5: bungalow 5
16.00 – 16.30 Group 6: bungalow 6
Dinner at the bungalows

Sunday

Work at the bungalows
12.30 – 13.30 Lunch
17.30 – 19.30 *Joint dinner with* Program Director and Coach bungalow 1

Monday

08.00 – 08.45 Breakfast at the bungalows
09.00 – 13.00 *Moment with Scientific Director: 30 minutes per bungalow:*
09.30 – 10.00 Group 1: bungalow 1 (group leader)
10.00 – 10.30 Group 2: bungalow 2
10.30 – 11.00 Group 3: bungalow 3
11.00 – 11.30 Group 4: bungalow 4
11.30 – 12.00 Group 5: bungalow 5
12.00 – 12.30 Group 6: bungalow 6
13.00 – 14.00 Lunch
14.00 – 18.15 Work at the bungalows
Dinner at the bungalows

Tuesday

08.00 – 08.45	Breakfast at the bungalow park
09.00 – 11.30	Assessment for applicants divided over bungalow 1, 2 & 3
09.00 – 10.00	Group evaluation trainees
10.00 – 11.00	Program Director and all trainees take a walk
11.00 – 12.00	Cake at bungalow 1
12.00 – 14.00	Work at bungalows
14.00 – 14.30	Lunch
14.30 – 19.00	Work at the bungalows

Dinner at the bungalows

Wednesday

08.00 – 08.45	Breakfast at the bungalow park
09.00 – 12.30	Preparation of presentations
12.30	Bus departures for lunch
13.00 – 14.00	Lunch
14.00 – 15.30	Presentations
15.30 – 16.00	Break
16.00 – 17.30	Presentations
18.00 – 19.00	Walk / break
19.00 – 21.00	Dinner & Data Challenge Award
21.00 – 23.00	Party
23.15	Departure to the bungalow park by bus

Thursday

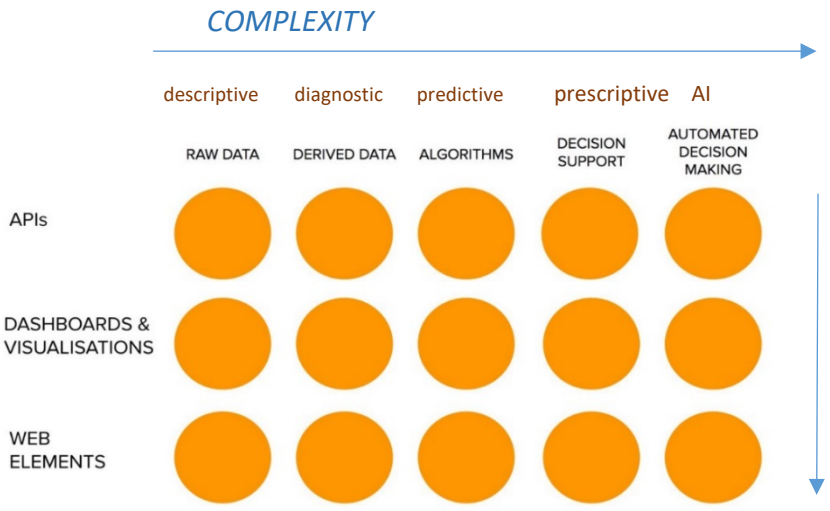
08.00 – 08.45	Breakfast at the bungalow park
09.15 – 18.30	Application interviews, bungalow 1.
	Group waits during each time frame in bungalow 2
	Social program during the day for the rest

Cleaning the bungalows

Friday

07.30 – 08.30	Breakfast at the bungalow park
09.00	Everybody departure to JADS
	Exit CenterParcs
10.00 – 17.00	Application interviews
12.15 – 12.45	Lunch

Appendix C | Training and Supervision Plan

<table>
<tr><th colspan="2">Training and Supervision Plan</th></tr>
<tr><td>Candidate</td><td></td></tr>
<tr><td>Start date</td><td></td></tr>
<tr><td>Work Experience</td><td></td></tr>
<tr><td>Master Study</td><td></td></tr>
<tr><th colspan="2">Key Learning Points</th></tr>
<tr><td>DESIGN</td><td>Reflect on your learning points related to designing and developing data science solutions for real world problems. What do you want to learn about the various steps of the design process, for what kind of data, for which parts of the data processing pipeline? What kind of products do you want to design? What is the complexity of the product, how does he data interact with that product, and how does the user? In answering these questions, feel inspired by the following item list:
• Design process - problem analysis, specification of requirements, experimentation, evaluation, process management
• Data types - multivariate data, geographic data, images, audio, text
• Data processing - collection, storage, cleaning, analysis, presentation
• Products – applications for feasibility studies, for decision support, and for automated decision making, dashboards, …
• Complexity – user product interaction, data product interaction,
Describe for each of your learning points, your starting level and the level you are aiming for according to the following example:
Learning point – ‘Perform a data scan’
Start/Aim level: novice o – S – o – A – o expert
 
</td></tr>
<tr><td>KNOWLEDGE</td><td>Reflect on your learning points related to the questions:
What knowledge do you need in order to create value from data and build data products?</td></tr>
</table>

	What knowledge do you want to gain to create value from data and build data products? What knowledge do you need to tell the story from the data? In answering these questions feel inspired by the following item list: • *Coding* – languages of databases such as SQL, Hive, and Server SQL • *Statistics* – regression, machine learning, Bayesian and descriptive statistics, software as R or Python • *Visualization* – show data in a simple, accurate, and concise way so that the audience can digest and act upon the information • *Business Analytics* – breakdown processes, outline the relevant processes, ways to improve the processes • *Modeling* – the fundamentals of mathematical modeling, the mathematical foundation of data science • *Team Work* – the fundamental principles of working in a team and be creative as a team Describe for each of your learning points, your starting level and the level you are aiming for according to the following example Learning point – '*Python in machine learning*' **o – o – S – o – A**
SKILLS	Reflect on the learning points related to the question: What skills do you want to develop to professionalize yourself as a data scientist? In answering these questions feel inspired by the following item list: • Critical thinking • Creative thinking • Design thinking • Report Writing • Creating Software • Math • Machine learning, deep learning, AI • Communication • Time management and prioritization • Data architecture • Risk analysis, process improvement, systems engineering • Problem solving and good business intuition • Continuous learning Learning point – '*Stakeholder analysis and business case*' **S – o – o – A – o**
Key Actions	

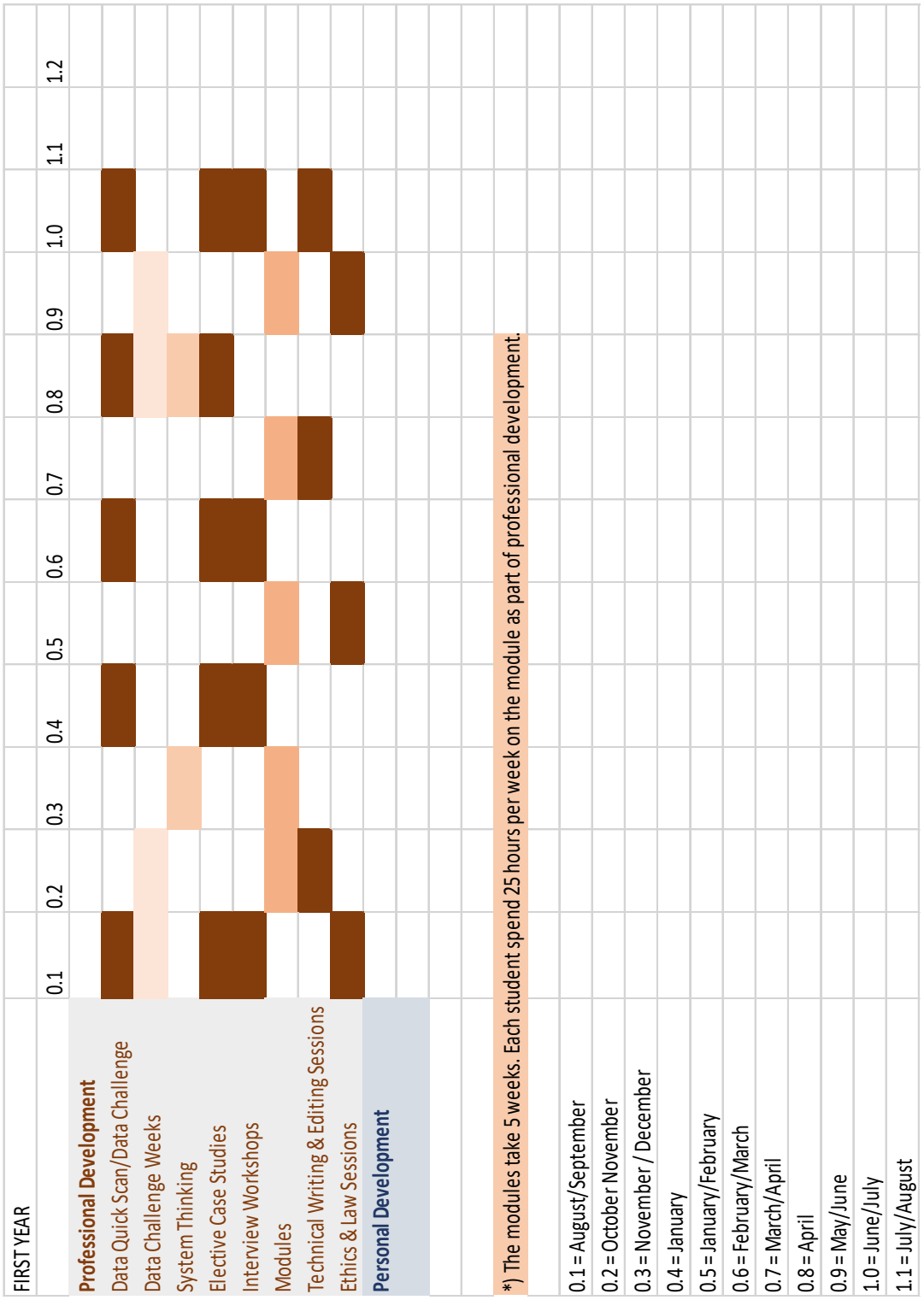
Courses	Professional Development (obliged) ➢ Modules (selection of 5 out of 9) o *Introduction to Data Science* o *Visualization* o *Statistics* o *Modelling & Simulation* o *Business Analytics* o *Data Engineering* o *Data Mining* o *Process Mining* o *Supply Chain Management & Logistics* ➢ System Thinking 1 & 2 Personal Development (elective) ➢ *Online courses (indicate the online courses you want to complete)* ➢ *Shared elective courses from other PDEng programs*
Trainings	Professional Development (obliged) ➢ Technical Writing & Editing ➢ Ethics & Law ➢ Interviewing ➢ Coaching ➢ Scientific Integrity ➢ Creative thinking
Projects	Professional Development ➢ Data Challenge Week (2) ➢ Case studies within a module Personal Development (elective) ➢ *Data Quick Scan* ➢ *Case study* ➢ *In house Data Challenge*
Others	Personal Development (elective) o Kaggle competition o Data Hackathon o Summer School o Conference, seminar

Signatures/date			
Candidate	Program Director	Program Coach	Scientific Director

Professional Development	25/40	
		Professionalizing
		Designing
		Knowledge Transfer
		Feedback & Reflection
		Technical Writing & Editing
Personal Development	15/40	
		Professionalizing
		Designing
		Knowledge Transfer
		Feedback & Reflection
		e-learning
Final Project	40/40	
		Professionalizing
		Designing
		Knowledge Transfer
		Feedback & Reflection
		E-learning
		Technical Writing & Editing

Timeline of activities

In all courses and projects, results that will be delivered are a report and a presentation. The on-line courses should always be finalized with a certificate.



Appendix D | Contract Final Project

CONTRACT
Name trainee
Data Science - 12 month project ##

Undersigning Parties:

1. **Jheronimus Academy for Data Science, (hereafter named "JADS"),** Sint Janssingel 92, 5211 DA 's-Hertogenbosch, a collaboration of the legal entities:

   Tilburg University[8]**,** having its registered seat at de Warandelaan 2, 5037 AB Tilburg (hereafter named: TiU), and
   Eindhoven University of Technology[9]**,** having its registered seat at de Groene Loper 5, 5612 AE Eindhoven, (hereafter named: TU/e)

   Both represented by A.C.L. Penners-Wouters (Eindhoven University of Technology), director operations at JADS**,**

and

2. Company, having its registered seat at Place, Country, Address (hereafter named:….) and represented by Name, Function,

Individually also denoted as the "Party" and jointly the "Parties".

[8] Tilburg University gaat uit van de stichting Stichting Katholieke Universiteit Brabant.
[9] Eindhoven University of Technology gaat uit van de publieke rechtspersoon Technische Universiteit Eindhoven.

Upon considering that:

- TUe/TiU has a post doctorate in engineering program and is responsible for the training and coaching of Professional Doctorate in Engineering (PDEng) trainees;

- Client/ Partner has an assignment that can be performed by a PDEng trainee;

- TUe/TiU is willing to designate a PDEng trainee to perform the assignment of the Client/ Partner under the terms and conditions stated below

- Parties are planning to carry out a project entitled "Project title"

- TUe/TiU are universities in the sense of the Dutch Law on Higher Education and Scientific Research within which scientific education is offered and where scientific research is conducted. Scientific integrity is an important principle for TU/e. For further support thereof, the Executive Board of TU/e has declared to apply within TU/e the "Nederlandse Gedragscode Wetenschapsbeoefening" ("The Netherlands Code of Conduct for Academic Practice") and in addition to the above code the "TU/e or TiU Gedragscode Wetenschapsbeoefening";

Have agreed as follows:

Article 1 Definitions
In this agreement the following terms are defined as follows:

Assignment:	The assignment as described in Appendix 1 and the Project Plan
Project Plan:	The intrinsic description of the Assignment, as described in Appendix 3 of this agreement.
The trainee:	The PDEng trainee named in Appendix 1 of this agreement, appointed at TU/e as a trainee of a two-year Professional Doctorate in Engineering program.
Results:	All results derived from the Assignment, such as, but not limited to, data, reports, findings, advice, conclusions, sketches, models, prototypes, materials and/or other material issues.
Confidential Information:	In any case research methods, research data, design results, designs, models, prototypes, materials, objects, intermediate reports, drawings and/or know-how;
Disclosing Party:	The Party that places Confidential Information at the disposal of the Receiving Party.
Receiving Party:	The Party that receives the Confidential Information from the Disclosing Party.

Article 2 Description

1. The project is carried out by full name trainee, trainee of the Professional Doctorate in Engineering program Data Science on basis of an agreed project plan containing milestones and time planning (see Appendix 3).
2. TUe/TiU undertakes to perform the Assignment on behalf of the Client/Partner under the terms and conditions stated in this agreement and the terms and conditions stated in Appendices 1, 2 and 3 of this agreement.

3. TUe/TiU will perform the Assignment with care and in line with generally accepted scientific standards.

Article 3 Performance

1. Name trainee is employed as a Technological Designer in Training.
2. During and at the end of the project Name trainee will deliver reports and documented software to Company.
3. The trainee, Name trainee, is responsible for the progress and reporting on the twelve month project. From JADS, the project is supervised by Name supervisor and Name supervisor. The problem owner representing the Company is Name problem owner.
4. Company is obliged to provide all information (including data) and/or materials (including equipment, samples, substances or items) to TU/e that are necessary for Tue/TiU to conduct the Assignment with care.
5. In case the trainee cannot perform his obligations due to long illness or other circumstances not attributable to the trainee and/or JADS, Parties will confer to come to a solution. If none for Company satisfactory solution has been reached within one month, Company can end the agreement by giving a notice of one month.

Article 4 Duration and term of the agreement

1. The project will be carried out full time in the period from (date-month-year) until (date-month-year).
2. The cost of the project, chargeable to Company, is € 5.800,- a month VAT excluded.
3. After acceptance by Company of the progress reports and other agreed upon deliverables, with regard to this project, invoices will be sent to Company covering the costs of the project. According to the guidelines (see Appendix 2) three invoices will be sent after each period of four months.
4. Further agreements related to the Assignment, including the work and the time commitment of all involved will be stated in the Project Plan by the trainee. This Project Plan will be approved by the Company within at least six weeks after the start of the Assignment, unless the Parties agree differently. The Project Plan is an inseparable part of this agreement, see Appendix 3
5. Each of the Parties may prematurely terminate this agreement if the Trainee is no longer employed by the TUe/TiU.

Article 5 Final report

1. TUe/TiU will ensure that within one month after the termination of the Assignment the Company receives a managerial report written according to the agreed Technical Writing & Editing standards of managerial reports. This report is confidential and will at no time be disclosed except for evaluation of the trainee by a beforehand assigned committee consisting of representatives of ‘Company’ and TU/e/TiU, see Appendix 2.
2. With respect to scientific publications derived from the Assignment, if the Company has the opinion that its commercial interest could be impaired by such scientific publication, Company may submit a request for a temporary embargo of 6 (six) months before publication

Article 6 Confidentiality

1. Both during and after the project, all directly and indirectly involved members of staff of TUe/TiU are obliged to secrecy according to a nondisclosure agreement between Company and TUe/TiU.
2. During the term of this agreement and for the duration of 2 (two) years after expiration or termination of the Assignment, Parties are obliged towards each other to keep secret each other's Confidential Information and they shall not use the Confidential Information in part or in whole for other purposes than for which it has been placed at disposal of the receiving Party. They shall not place the Confidential Information or part thereof at the disposal of third parties.
3. Information provided by Company concerning the company or the project will be treated strictly confidential. Distribution of written reports and possible resulting publications will take place only after written consent by Company.
4. Parties are obliged to sign a nondisclosure agreement.
5. Secrecy as mentioned in this Article 7 shall never affect the right and possibility of Tue/TiU to publish scientifically, which right to publish is further specified in Article 8 of this agreement.

Article 7 Intellectual property

1. Developed software becomes freely available to the company. However, if software is used, to which TUe/TiU has proprietary rights, further agreements on its use within the company and on the compensation to TUe/TiU has to be made.
2. If in the framework of the project a patentable invention has been done, the claim on the patent goes to Company.
3. Results of research or knowledge generated in the frame of the project are intellectual property of Company.
4. Knowledge, owned by one of the parties before the start of the research, remains in property of that party.
5. The intellectual property right resulting from the Results lie and will lie with Company. Company hereby extends to TUe/TiU the right to use the Results of the Assignment for its own education and research as well as for research for third Parties
6. Parties cannot assign rights and obligations under this contract to third parties unless with the prior written consent of the other Party.

Article 8 Publication

1. Without the prior written consent of TUe/TiU, Company shall not communicate about the approach by TUe/TiU, its methods and the like, nor publish or make public in any other way and/or provide to third parties the final and/or intermediate reports, the reports and/or other results of the Research.
2. During the execution of the Assignment TUe/TiU shall not publish or make public in any other way the Design Results without the written permission of Company. If, within 14 days after its request for permission, Tue/TiU has not received a written response from the Company, the permission from Company will be considered to have been given. Permission from Company may not be unreasonably withheld.
3. In all cases TUe/TiU have the freedom to publish the findings within a reasonable period to be determined, a term of two (2) months shall be considered reasonable and six (6) months usually the maximum (to be calculated from the moment of delivery of the Design Results to the Company). In case of intellectual property right

issues, an exception can be made and a period with a maximum duration of twelve (12) months accepted.

Article 9 Liability

1. With respect to responsibilities, both parties agree that Company is completely responsible for the personal damage that the trainee meets through Company.
2. If the report has been accepted by Company, JADS is safeguarded from any eventual damage coming forth from the use of the deliverables
3. TUe/TiU shall not be liable for direct or consequential damage sustained by Company and/or third parties from the execution of the current agreement

Article 10 Applicable law, competent court

1. This contract shall be governed by the laws of the Netherlands.
2. Disputes between the Parties, arising out of the performance of this agreement, which cannot be reasonably resolved, shall be subjected exclusively to the competent court of 's-Hertogenbosch, The Netherlands.

As drawn up in duplicate and signed,

	Company name	**Jheronimus Academy of Data Science**
Signature		
Name		Drs ing A.C.L. Penners-Wouters
Position		Director Operations
Date		

Appendix 1: Assignment details
Appendix 2: Guidelines twelve month project
Appendix 2: Project plan

Appendix 1

Specific to matters related to the Assignment

Trainee:

University Supervisor(s):

Problem owner(s):

Start Design Assignment:

End Design Assignment:

Price per month:	€ 5,800 excluding VAT
Payment schedule:	€ …. Within xxx days after start Design Assignment € …. Within xxx days after ….. € …. Within xxx days after end Design Assignment

Appendix 2

Conditions and Guidelines twelve month PDEng project

Summary

The twelve month project is carried out by one PDEng-trainee, preferably at the company, full-time. In a preparatory phase, before the start of the project, problem context and problem specification are defined. Thus, at the start of the project, deliverables, project planning, and time planning are identified. The project is executed as contract research with the university. Costs of a long term project are € 5,800,-- per month, VAT excluded. The project has a substantially scientific level and is devoted to problems of highly innovative level. The project is executed under a nondisclosure agreement.

Objective

The candidate should show the ability to carry out a long term project individually within an industrial or business environment. The project is related to a real life problem and is constrained by time planning, project planning and project deliverables.

Responsibilities

Every long term project is preferably carried out within the company and the company takes care of a working space.

- The project supervisor assists in drafting the problem description, project planning and time planning. The project supervisor can involve other members from the department staff in the project, if their scientific expertise and support is requested and wanted.
- The program management finalizes the contract and checks whether the contract is established in a sound way and offers the necessary administrative support. The program management arranges evaluation meetings.
- The company appoints a problem owner, who is the responsible party on behalf of the company. This way, there is a guarantee that the company has a genuine interest in the progress and results of the project.

Project description and finances

Each project is carried out on the basis of a contract between the company and the university. In an enclosure to the contract, the problem to be addressed in the project is described. This project description is drawn up by the trainee, project supervisor, and problem owner. The final draft of the project description is the responsibility of the trainee. In preparation to the project the trainee delivers a project description that contains the following items:

1. Global description of the company and of the division within the company where the project takes place together with its most important activities
2. Sketch of the context of the project
3. Detailed description of the project problem, goal, and deliverables

4. Description of the quality of the expected final result with global qualifications that have to be satisfied.
5. Time scheme with milestones and project-planning schedule.

Time scheme

Prior to the start of the Twelve month project clear arrangements are made with respect to supervision both by the program administration and by the company.

- During the project, the trainee keeps a logbook, containing all decisions and agreements made at every meeting of both supervisors and trainee.
- After four and eight months, the trainee gives a progress presentation in the project seminar and writes a progress report. This way his/her findings are presented to the larger audience of colleague trainees and staff members of the department.
- The trainee completes the final report.
- The Twelve month project is concluded with a presentation in the company and thereafter the project is evaluated.
- The trainee and project supervisor discuss the project regularly.
- The trainee pays a visit to the university at least once every two weeks.
- The problem owner, project supervisor and trainee meet on a regular basis. At least once every four weeks the trainee sets up a meeting, accompanied by an agenda, in which all project stakeholders and, if so desired, one or more experts participate.

Evaluation

In the course of the project, three evaluations take place. Two evaluations are scheduled after the trainee has delivered the progress reports, likely after four and eight months; and the third one, the final evaluation, is scheduled at the end of the project.

- The progress evaluations are conducted on the basis of an evaluation interview and an evaluation form. The outcome of the progress evaluation is discussed with the trainee, and if necessary, project goal and time planning are adapted.
- The final evaluation is conducted by an evaluation committee that is assigned for this evaluation task by the program management. The committee judges:
    - the degree in which the assignment was made a success within the set time
    - the way in which the trainee operated in the company
    - the scientific and technical quality of the work, its design features
    - The quality of the report and the presentation

The supervisors and problem owners are members of the committee. The judgement of the committee is registered in an evaluation form.

Appendix 3

Project plan

Problem background

Problem background

Problem description

Problem description

Project goals

- Project goal
- Project goal
- Project goal
- Project goal

Project management

Project management

References

[1] Reference
[2] Reference
[3] Reference

Project planning

Project planning

Appendix E | Evaluation form for Final Project

Evaluation Form Final Project

Trainee:
Company:
Date

	P	M	S	G	E
Meeting the project goals					
Setting up a realistic time planning					
Focusing on meeting set goals, missions, or objectives					
Carrying out project planning and using project planning in communication					
Formulating a clear problem description					
Showing the competence to define a project related business case					
Using techniques of the various fields of data science adequately					
Working as a data science generalist combining data mining, data analytics, and modeling in a business context?					
Extracting (business) value out of data					
Designing structured and user friendly software					
Showing an active attitude in involving people with matching domain expertise					
Carrying out literature search					
Acting critically on own work					
Showing creativity in own work					
Telling the story from the data at executive, manager, and expert level					
Considering legal and ethical aspects					
Defending own findings					
Setting up regular communication between the various stakeholders in the project					
Communicating clearly and convincingly to the different stakeholders					
Working in a team in a friendly and responsible way					
Showing the attitude of a problem solver					
Delivering a well-structured, informative presentation					
Writing an industrial report					
Working as a professional data scientist					

P=Poor; M=Moderate; S=Sufficient; G=Good; E=Excellent

Addendum A | Profiles of Data Scientist, Data Engineer, and Business Analyst

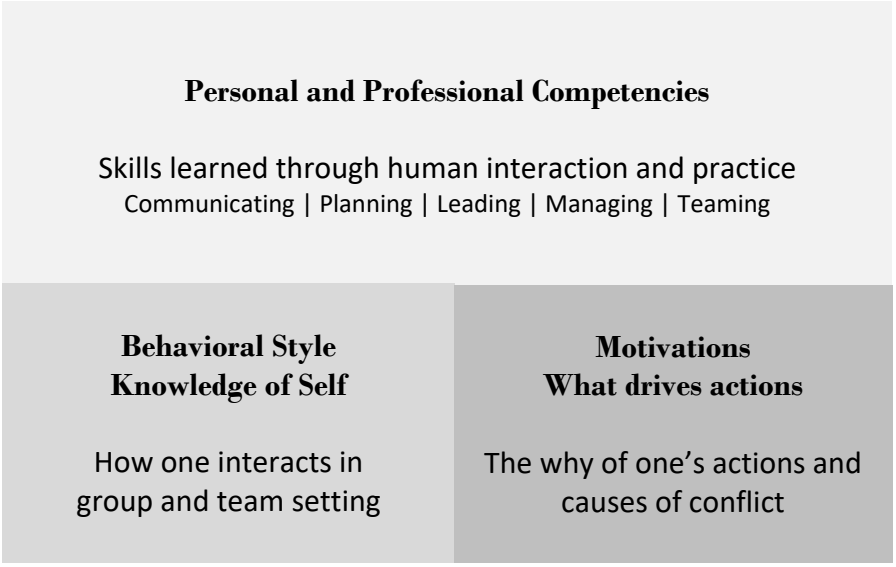
1. Introduction

The central question that we would like to address in this addendum can be phrased as follows

> *What distinguishes the performance of a Data Scientist from a Business Analyst and a Data Engineer?*

Having answered this question, we can better describe and thus assess the candidates for PDEng DS and the graduates from this program, as well. The answer helps trainees to make a well-considered choice how to shape their career as an academically skilled data science professional. We provide insight into the profiles of the data scientist, data engineer and business analyst as suggested in literature.

In order to describe trainees' profile and performance, we decided to follow the model of KEEN-TTI (see reference in the below figure caption). In that paper the authors introduced the *Performance DNA* of the entrepreneurially minded engineer (EME). Performance DNA is based on the following three latent variables: Personal and Professional Competencies, Behavioral Style, and Motivations.



Taken from Mapping the Behaviors, Motives and Professional Competencies of Entrepreneurially Minded Engineers in Theory and Practice: An Empirical Investigation, Journal of Engineering Entrepreneurship, Volume 4 number 1-2013, page 39 – 54

2. Behavioral Style (the how)

In order to categorize the behavioral style of an individual, according to the DISC model, four such styles can be distinguished.

> Twenty-four hundred years ago, scientists and philosophers, most notably Hippocrates, began to recognize and categorize differences in behavior that seemed to follow a pattern. Since then, many psychologists and scientists have explored behavioral patterns. Dr. William Marston wrote "*The Emotions of Normal People" in 1928* after earning his doctorate from Harvard University. Marston theorized that people are motivated by four intrinsic drives that direct behavioral patterns. He used four descriptive characteristics for behavioral tendencies which are represented by four letters of the alphabet: D, I, S and C. Thus the concept of "DISC" was introduced.

Dominance (challenge) – how an individual addresses problems and challenges

> Individuals with a dominant behavioral pattern put emphasis on the accomplished results as the bottom line; they show being in control and they are direct, strong-willed, and forceful. Individuals with the D – style are directed towards shaping their environment by overcoming opposition to accomplish results.

Influence (contacts) – how an individual handles situations involving people and contacts

> Persons with an influential behavioral pattern put emphasis on influencing or persuading others in the accomplishment of results; they show openness and eagerness in building relationships, and they are sociable, talkative, and lively. Persons with the I – style place are directed towards shaping the environment by influencing and persuading others.

Steadiness (consistency) – how an individual demonstrates pace and consistency

> Individuals with a steady behavioral pattern put emphasis on cooperation in the accomplishment of results; they show sincerity and dependability, and they are gentle, accommodating, and soft-hearted. Individuals with the S – style are directed towards cooperating with others within existing circumstances to carry out a task.

Conscientiousness (compliance) – how an individual responds to rules and procedures set by others

> Individuals with a conscientious behavioral pattern put emphasis on quality and accuracy in the accomplishment of results; they show having expertise and competency, and they are private, analytical, and logical. Individuals with the C – style are directed towards working conscientiously within existing circumstances to ensure quality and accuracy.

3. Attitude and value motivators (the why)

In 1928, Edward Spranger wrote the book entitled "Types of Men." He identified six major attitudes or world views. These attitudes he assumes are the type of window through which

we view the world and seek fulfillment in our lives. If we are participating in a discussion, activity, or career that is in alignment with our attitude, we will value the experience. Conversely, if we are in a conversation, activity, or career that is in conflict with our dominant attitudes, we will be indifferent or even negative toward the experience, possibly causing stress.

1. *Theoretical* - A passion to discover truth. The chief aim in life is to order and systematize knowledge for the sake of knowledge itself.
2. *Utilitarian / Economic* - Practical interest in money and a passion for what is useful. Time and resources are meted out with an eye to future economic gain.
3. *Aesthetic* - Interest in form and harmony. Life is a series of episodic events, each enjoyed for its own sake. Has a heightened sense of beauty and inner vision (not necessarily talented in creative artistry).
4. *Social* - Inherent love of people. Seeks to eliminate hate and conflict. Other persons are ends to themselves (not means). Altruistic, kind, empathetic, and generous, even to their own detriment.
5. *Individualist / Political* - The primary interest for this value is power, not necessarily politics in the traditional sense. Research indicates that leaders in most fields have a high power value. In the eyes of this person, other people may be seen and used as simply the means to an end.
6. *Traditional / Regulatory* - Unity and order. The need to be regulated or the need for structure from an outside source. Seeks to comprehend the cosmos as a whole and to relate themselves to a global totality. May alternate between the negation and affirmation of life, or seek mystical oneness. Dislikes change and chaos. May also exhibit inflexibility with regard to their convictions.

A thorough description of the six motivators is presented in the addendum of this document.

4. Skills (what)

The skills latent model is described by 23 manifest variables.

1. *Self-management* (time and priorities) - Demonstrating self-control and an ability to manage time and priorities.
2. *Customer Service* - Anticipating meeting and/or exceeding customer needs, wants, and expectations
3. *Written Communication* - Writing clearly, succinctly and understandably
4. *Goal Orientation* - Energetically focusing efforts on meeting a goal, mission or objective
5. *Flexibility* - Agility in adapting to change
6. *Persuasion* - Convincing others to change the way they think
7. *Creativity/Innovation* - Adapting traditional, or devising new approaches to, concepts, methods, models, designs, processes, technologies and/or systems.
8. *Planning/Organizing* - Utilizing logical, systematic and orderly procedures to meet objectives

9. *Interpersonal Skills* - Effectively communicating, building rapport and relating well to all kinds of people
10. *Futuristic Thinking* - Imagining, envisioning, projecting and/or predicting what has not yet been realized
11. *Presenting* - Communicating effectively to groups
12. *Continuous Learning* - Taking initiative in learning and implementing new concepts, technologies and/or methods
13. *Teamwork* - Working effectively and productively with others.
14. *Diplomacy* - Effectively handling difficult or sensitive issues by utilizing tact, diplomacy and an understanding of organizational culture, climate and/or politics
15. *Analytical Problem Solving* – Anticipating, analyzing, diagnosing and resolving problems.
16. *Personal Effectiveness* - Demonstrating initiative, self-confidence, resiliency and a willingness to take responsibility for personal actions
17. *Empathy* - Identifying with and caring about others
18. *Negotiation* - Facilitating agreements between two or more parties
19. *Decision Making* - Utilizing effective processes to make decisions.
20. *Leadership* - Achieving extraordinary business results through people.
21. *Management* - Achieving extraordinary results through effective management of resources, systems and processes
22. *Conflict Management* - Addressing and resolving conflict constructively.
23. *Employee Development/Coaching* - Facilitating and supporting the professional growth of others.

5. Profiles

The profiles of the Data Scientist, Data Engineer, and Business Analyst are characterized by the following behavior, motivation, and skills/competency patterns (the darker the more prominent):

Behavior	Data Engineer	Data Scientist	Business Analyst
Dominance			
Influence			
Steadiness			
Compliance			

Motivation	Data Engineer	Data Scientist	Business Analyst
Theoretical			
Aesthetic			
Traditional			
Individualistic			
Social			
Utilitarian			

Skills/Competency	Data Engineer	Data Scientist	Business Analyst
Self-management (time and priorities)			
Customer Service			
Written Communication			
Goal Orientation			
Flexibility			
Persuasion			
Creativity/Innovation			
Planning/Organizing			
Interpersonal Skills			
Futuristic Thinking			
Presenting			
Continuous Learning			
Teamwork			
Diplomacy			
Analytical Problem Solving			
Personal Effectiveness			
Empathy			
Negotiation			
Decision Making			
Leadership			
Management			
Conflict Management			
Employee Development/Coaching			

Addendum – Detailed description of the six attitude and value motivators

High Theoretical

People whose primary value drive is *theoretical* have a tremendous need to know, to learn, to understand. The bottom line is the accumulation of knowledge and the logical pursuit of this knowledge is where it is at for them. The primary drive with this value is the discovery of truth. Since the interests of theoretical persons are empirical, critical and rational, they are necessarily an intellectualist. The chief aim in life is to order and systematize knowledge: Knowledge for the sake of knowledge. High theoretical people are not always interested in using this knowledge, and we do find some so-called smart damn fools, absent-minded professors, etc. As far as the theoretically inclined persons are concerned, they will learn well but not always do. One of the tools for helping abstract thinkers perform is a strong standards of performance system, which can be monitored. Again, there is no interest in changing people but rather in improving performance.

General Characteristics:

Feeling for the purity of the cognitive process.
Intertwines past and present.
High interest level in solving problems, asking questions or formulating theory.
Enjoys people with convictions (knowledge) held in common.

Possible Limitations:

May have trouble dealing with practical problems.
Little time for people who see things differently - especially emotional ones with few facts.
Single mindedness at the expense of everything else.
May get bogged down in the quest for details - can lead to procrastination.

The questions one might ask regarding a person with high theoretical value are: How will the high theoretically inclined person plan, organize, direct, control and even sell? How will such a person recruit, select, train, motivate and communicate? How will they get along with others? In private life how will he or she handle social situations, play games, manage money and perform as a spouse, parent, etc.? And more importantly, what does he or she need from a manager, spouse or friend to be more effective? These same questions can and should be asked for each of the value drives.

Very Low

Practical application for the use of knowledge.

Low

Knowledge to gain results, or an advantage. Could be an avid reader regarding their needs or hobby.

Average

Need for knowledge-for-knowledge's sake is based on individual situations. If interested in a specific area, or if required for success, they will want to know everything there is to know. If not, intuition or practical information will be relied on.

High

Wants knowledge-for-knowledge's sake. Wants to become an expert. Quest for knowledge-need to know.

High Utilitarian (Economic)

As impractical as some high theoretical people are, that is just how practical and tangible-results oriented those with a high materialistic or economic value system are. Their goal is utility and what is useful. People driven by this value are achievers and want rewards and results now. Their basic interest in knowledge (theoretical) is restricted not to how much they can accumulate, but rather how they can use it. Money and possessions are the measuring tools or yardstick by which the high materialistically motivated keep track of their accomplishments. These people respond mostly to on-the-job training and a compensation system based on monetary incentives. The materialistically motivated person should not be seen as selfish but rather as practical and goal-oriented.

General Characteristics:

Very practical, can be a spender or saver.
Future oriented.

Motivated by the satisfying of needs.
Seldom or never reaches their wants, continually motivated by wants and needs.

Possible Limitations:

May be a workaholic.
Egotistical
May have a visible greed factor.
Rationalizes giving of time or resources will result in some future economic gain.

Very Low

Not overly concerned with material things or money. Wants to make sure they can keep body and soul together. Motivated toward achievement relating to their internal beliefs. Money is not a score to impress others.

Low

Not driven by a tremendous need to have great sums of money. Wants to be able to achieve the survival needs at an acceptable level as perceived by their perception of social standing. Independence is a long-term project. Will profit from economic goal-setting. Needs to meet with a mentor regularly.

Average

In specific situations they feel compelled to make the acquisition of money a very important aspect of decision-making.

High

Internally motivated by the need to have economic rewards in terms of money for security or freedom. Money in and of itself is not the end but a means to achieve that end.

Very High

Money in terms of what it can do is extremely important. Very practical. At times will be overwhelmed by the advantage that money or materialistic things can bring.

High Aesthetic

The aesthetic score indicates the relative interest in form and harmony. Each experience is judged from the standpoint of grace, symmetry, or fitness. The high aesthetically motivated are, most of all, very sensitive persons with an artistic flair for things harmonious and beautiful. While not necessarily performers or artists, per se, they do personalize beauty (as perceived by them) in the world around them. They would rather see something more charming than useful and more beautiful than practical. Finding it difficult to gain aesthetic value satisfaction in a rough and tough business climate, the high aesthetic person will usually gravitate to a nicer environment. They will also perform best in a pleasant, harmonious setting. Their goal is to experience their inner vision.

General Characteristics

Seek self-realization, self-fulfillment and self-enjoyment.
Sensitive to inner feelings.
Humorist if view of life is positive.
Sarcastic if view of life is negative.

Possible Limitations

Attempts to influence others by aesthetic beauty.
No feeling for the practical.

Perceives the world only from their inner reality.
Minimum use of logical reflections.

Very Low

Not worried about form, nor sensitive to the pleasing aspects of the environment. Very practical people. They know the sun has gone down mainly because it is dark.

Low

Does not require harmony of nature to feel fulfilled. Practically overrides sensitivity. The awareness of fine things and fine relationships is secondary. Creative problem solving vs. creative sensory pleasure. World is black and white vs. world is a colorful rainbow.

Average

Need for aesthetics (appreciation of beauty) determined on an individual basis. Specific areas could be of great interest (i.e., desiring fine things for family members, but not concerned with the depth of relationships with others).

High

Needs fine things and fine relationships. Wants a harmony that relates to enjoyment and appreciation of things that have intrinsic beauty. Internalized creative feelings. Creative designs to problem solving as they relate to the sensitivity of the relationship.

Very High

Tremendous need for a sense of balance and harmony within their environment. Desires fine things and fine relationships. More concerned with the part than the whole. Can be very creative.

High Social

Those who score very high in this value have an inherent love of people. They prize other persons and are, therefore, kind, sympathetic and unselfish. Those with a high humanitarian value system are more concerned with the welfare of others than they are for self. The humanitarian or social value drive places helping others very high on the list of personal priorities. The social person regards love as the only suitable form for human relationships. Research into this value indicates that in its pristine form the social interest is selfless. Many times this value drive rises to the top of an individual's set of values after the materialistic is satisfied. The goal of very high social value people is to eliminate hate and conflict in the world.

<u>General Characteristics:</u>

See their own value in helping others.
Real concern for others.
Ability to be empathetic.
Generous with time, talent and resources.

<u>Possible Limitations:</u>

Self-sacrifice at times and may override self-preservation.
Have difficulty saying no.
Help others even to their own detriment.

Will avoid confrontation if there is an unbearable truth that will harm a relationship.

Very Low

Concerned about the needs of others. Will help others to better themselves, but out of pity. Compassionate only for those who have either physical or mental disability. Winning is the most important necessity. Does not appreciate weakness. Strangers are strangers.

Low

Willing to help others if they are working as hard as possible to achieve their goals. Won't help others if it would be detrimental to themselves. Does not promote a welfare state. Hard work and example can motivate others.

Average

Desire to help others or not is reviewed on an individual basis. If an internal chord has been touched they would definitely attempt to help.

High

Very concerned about the welfare of others, even to their own detriment. Wants others to have the opportunity to succeed. Sometimes people become projects for saving. More evident when money needs are met.

Very High

May neglect own family and friends. Interested in humankind in general. Truly unselfish. Good team player. Gives others many chances for success and then still gives more. Efforts are sufficient proof of worth. Social concerns are of the highest value. Will be a leader of social reform.

High Individualistic (Political)

Persons with high power-seeking value drive are very easily spotted and understood. What they want is power and control and an arena in which to play where there is ample opportunity for public ego-satisfaction. The climb up the so-called corporate ladder in terms of title, recognition, and power is quite symptomatic of the need for power-seeking value drive satisfaction. The attitude here tends to be move me up or watch me move on. The leaders in most fields have high power value. The goal of this value is to assert self and have their causes victorious. A person motivated by the individualistic value drive is primarily interested in independence. The individualistic seeks personal expression and demonstrates disdain for rules and authority per se.

<u>General Characteristics:</u>

The effect of power upon others appears in the form of determination.
Control their own destiny and the destiny of others.
Power and control will usually be expressed in some other form or value
Theoretical (superiority), Economic (wealth) or Regulatory (religious).

<u>Possible Limitations:</u>

The end justifies the means.
May break rules in order to rule or control.
May be Machiavellian in their approach to others.
Need for self-assertion. Can come across as feeling superior.

Very Low

Does not need to be in the limelight. Does not need to be seen as a leader. Does not need to control others. Keeps conflict and hostility at a minimum. Ego satisfaction and praise are not necessarily success measurements. Needs stability.

Low

Positions of power and control are not an intrinsic motivating factor. Willing to allow others to set the tone and direction of their destiny. Much more patient and less ego-involved than others may be. Will participate as a team member for the team's sake, not their own sake. Does not attempt to control the destiny of others, but wants to achieve within the framework of their own area of specialty.

Average

Will evaluate each situation individually and determine how much or how little control they want to exercise. If there are strong feelings about issues, control increases.

High

Tremendous need to show that they can take charge and be the leader. Competition and struggle are part of daily routine. Wants to be the person in the forefront and seen as a mover and shaker.

Very High

Wants to control situations, as well as the destiny of themselves and the destiny of others. Being in the limelight will have them work extremely hard and for long hours. Perks and strokes are important. Titles are important. Being in charge is important. Wants to be seen as a winner and they won't play if there is not a chance to win.

High Traditional (Regulatory)

This value reflects a spiritual commitment and/or a preference for rules and authority. When the spiritual value appears low, we are seeing individuality. Those who are motivated by a high spiritual value system are primarily concerned with unity or order. They are deeply committed to belief in Supreme Being. Their goal is to search for the highest value of life. The spiritually motivated literally are faith oriented. Those who evidence a high ritualistic value drive, on the other hand, may or may not be religious in the formal sense of the word. They will demonstrate, however, a need for a dependence on authority and a clearly defined career path. They may be niche seekers. These people will be comfortable in highly structured, well defined environments and generally will be more comfortable in large companies as opposed to smaller, more entrepreneurial situations.

<u>General Characteristics:</u>

Mental structure to create the most important or satisfying value experience.
May view life positive, negative, or mixed.
Will seek power on a big scale, if political is very high.
Will seek the richest revelations of beauty, if aesthetic is very high.
Belief in their belief is so great that they will champion their beliefs.

<u>Possible Limitations:</u>

Overly rigid.

Comes across as always right.
Rarely changes mind even if logic dictates they are wrong.

Very Low

Tradition will not place limits or boundaries. Many things to see and many to try. Will experiment with different belief systems. Hard to manipulate when it comes to setting guidelines-they have very few to begin with. In many cases they want to set their own rules and allow their own intuition to guide and direct them. Can be highly organized in a very unstructured approach to any rules and procedures set by others. Can be very creative. Does not rebel-just ignores.

Low

Not bound by traditions and customs. The way things have been done before are not necessarily dismissed, but they are always exploring new ideas and new methods of doing things.

Average

Need to be able to pick and choose the traditions and set of beliefs to which they will adhere. Strong beliefs within a system that feels comfortable. Will not be so strong in beliefs if there is a lack of interest. Pick and choose whether they follow traditional ideas or deviate. Interest in the subject matter will determine following or breaking with tradition.

High

Driven by need for a traditional approach to their lifestyle. Attempts to find the guideline or the rulebook which will allow for the long-term direction. Internally driven to discover their place in the scheme of things.

Very High

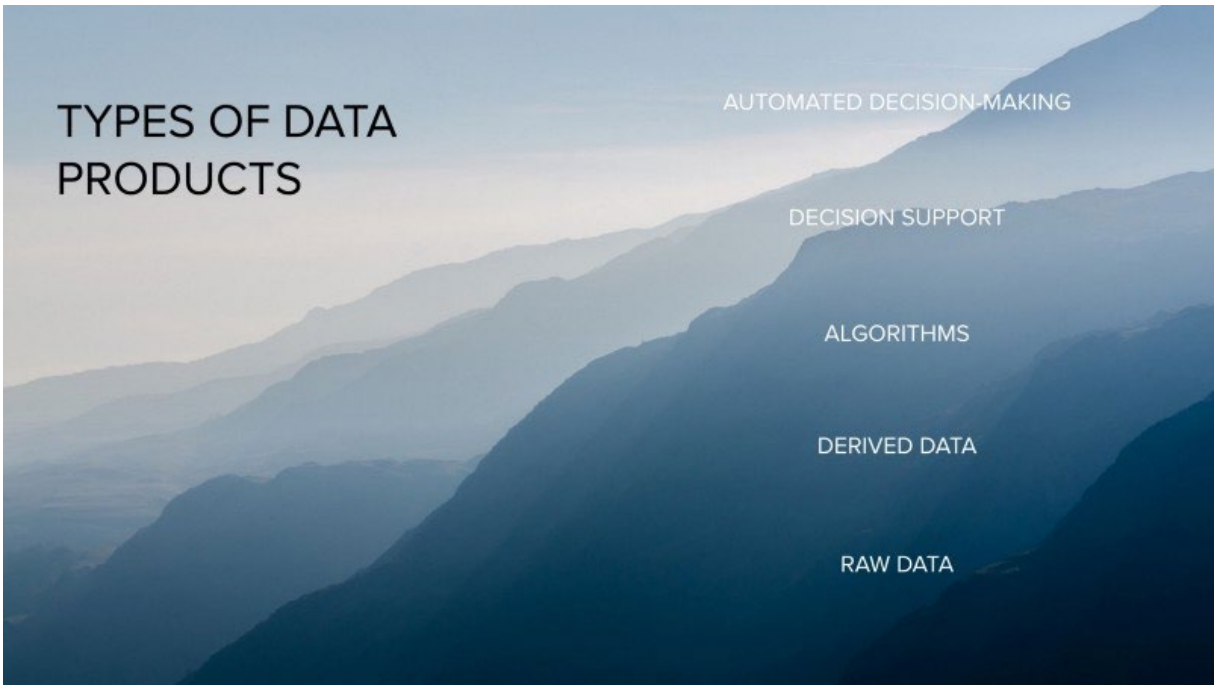
Believes in doing things the traditional way. Wants the rules and regulations of society to be a closed loop. There is only one way to do things, and that is by the rule book they have chosen to follow. Can become quite determined about their beliefs. See themselves as becoming very moralistic, with discipline and conviction. Little need to experiment with other ways of doing things. When they find a leader in whom they can believe, they will follow that leader almost absolutely.

Addendum B | Data Products and Data Interactions

Data Products and Data Interactions

Simon O' Regan

Clearly there exists a wide range of different types of data products. Even narrowing down the field of possible products to those that satisfy our definition, there is still considerable variety amongst these products. With this variety comes further subtleties in product development.



We can organize these data products into 5 broad groups: raw data, derived data, algorithms, decision support and automated decision-making.

Generally speaking these product types are listed in terms of increasing complexity. More specifically, they are listed in terms of increasing internal complexity and (should have) less complexity on the user's side.

Put another way, the more computation, decision-making or "thinking" the data product does itself, the less thinking required by the user.

Typically (but not exclusively) raw data, derived data and algorithms have technical users. Most often they tend to be internal products in an organization but counter-examples would include Ad Exchanges, or API suites. Decision support and automated decision-making products tend to have a more balanced mix of technical and non-technical users; though for any given product, the user group tends to be one or the other.

Raw data Starting with raw data, we are collecting and making available data as it is (perhaps we're doing some small processing or cleansing steps). The user can then choose to use the data as appropriate, but most of the *work* is done on the user's side.

Derived data In providing users with derived data, we are doing some of the processing on our side. We could, in the case of customer data, add additional attributes like assigning a customer segment to each customer, or we could add their likelihood of clicking on an ad or of buying a product from a certain category.

Algorithms Next we have algorithms, or algorithms-as-a-service. We are given some data, we run it through the algorithm — be that machine learning or otherwise — and we return information or insights. A good example would be Google Image: the user uploads a picture, and receives a set of images that are the same or similar to the one uploaded. Behind the scenes, the product extracts features, classifies the image and matches it to stored images, returning the ones that are most similar.

Decision support Here we are looking to provide information to the user to help them with decision-making but we are not taking the decision ourselves. Analytics dashboards such as Google Analytics, Flurry, or WGSN would fall into this category. We are doing most of the heavy lifting on our side; our intention is to give the user relevant information in an easy-to-digest format to allow them to take better decisions. In the case of Google Analytics, that could mean changing the editorial strategy, addressing leaks in the conversion funnel, or doubling down on a given product strategy. The important thing to remember here is as follows: while we have taken design-decisions in data collection, derivation of new data, in choosing what data to display and how to display it, the user is still tasked with interpreting the data themselves. They are in control of the decision to act (or not act) on that data.

Automated decision-making Here we outsource all of the intelligence within a given domain. Netflix product recommendations or Spotify's Discover Weekly would be common examples. Self-driving cars or automated drones are more physical manifestations of this closed decision-loop.

We allow the algorithm to do the work and present the user with the final output (sometimes with an explanation as to why the AI chose that option, other times completely opaque).

Data Interactions

So far we've discussed *functional* data product types.

Each of these data products can be presented to our users in a variety of ways — with clear implications for their design. What are these interfaces or interactions?

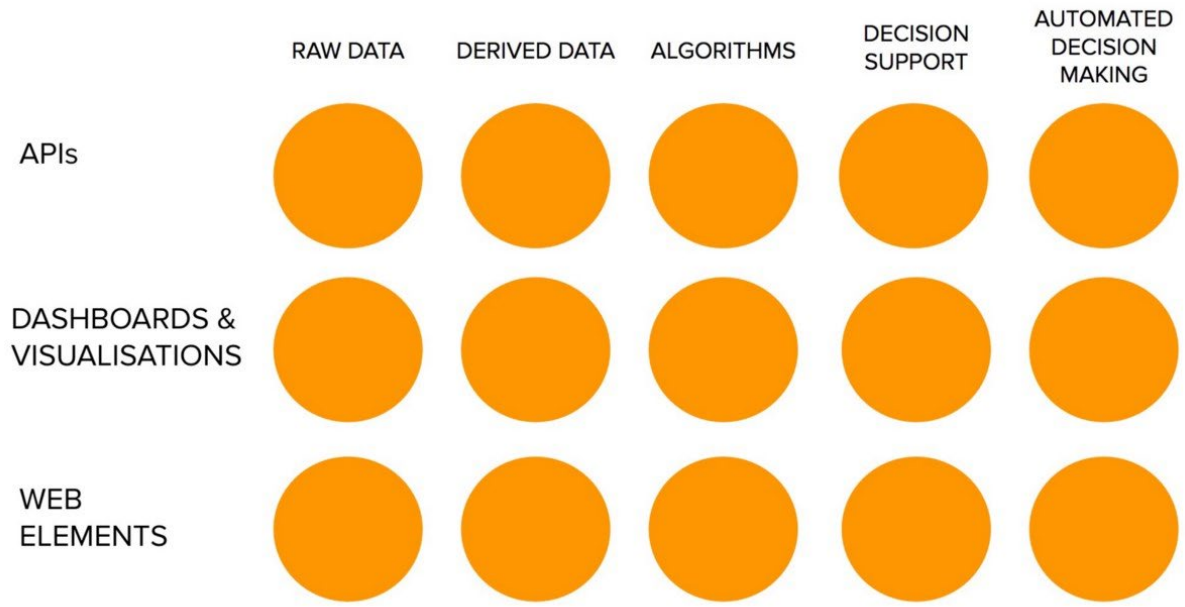
APIs. In the case of APIs, we assume a technical user. We should still follow good Product practices and ensure that the API is intuitive to use, well documented, can do what its user's need and is desirable to work with.

Dashboards & visualizations. For dashboards, and visualizations we're assuming some statistical literacy or competence in dealing with numbers. In its most extreme we can do a lot of the heavy-lifting for our users and work hard to ensure that we only present the most pertinent information in an easy-to-understand format. By choosing what information to display, we are influencing decision-making, but it still leaves interpretation and decision-making in the hands (or minds) of the user.

Web elements. For the past 5 years or so the least technical interface for data products that have been commonly seen by users has been web elements. More recently, these interfaces have been broadly extended to include voice, robotics and augmented reality, amongst others. While the design details for each of these newer interfaces are clearly distinctive, there is considerable overlap, in that they revolve around presenting the results of a decision to the user, and perhaps also communicating why or how the AI reached that decision.

Understanding what we're building

Plotting the types of data products against possible interfaces, we get a matrix of orange dots with each dot representing a different data product variant.



Data Product Matrix — different products require different approaches

Each element of the matrix demand design considerations that can differ substantially, both in terms of what the user needs and in terms of what design process we use to get there.

Moving diagonally from the top-left circle (Raw data-API) toward the bottom-right circle (Automated decision-making-Web elements) is to move from technical, engineering-driven products towards those that are more typical software products (i.e. products that are more intuitive to product managers and designers, those that tend to appear in books, magazines and articles).

Difficulties & Methodologies

In my experience the biggest problems that teams encounter with data products happen when they apply methodologies like human-centered design on more technical data products. This is not to say that engineers are not human. Most are, and those that aren't often have an uncanny likeness. But HCD is a holistic approach to product development that excels when the designer understands the motivations and behavior of the user. For technical data products, the product boundary is often artificially constrained by functional organization considerations, and the product and UX team is often insufficiently technical to either a) understand the intricacies of technical user behavior or b) insufficiently inclined to explore these intricacies.

To assume then that the Design-Thinking or Lean methodologies that we've been reading about should be applied out-of-the-box is naive.

This is not reason to panic, however.

Though the outputs from the user research may be considerably different to those experienced with consumer-facing or indeed typical SaaS products, and the definition of KPIs may err on the side of the technical, both Design Thinking and Lean are sufficiently malleable to allow us to tailor our approach to this new domain.

My advice then, when applying these methodologies to data products is to ensure that the problem-space is defined in terms of the end user, rather than just the user of the immediate data output. In all likelihood, this will mean expanding the team to include adjacent products and their managers.

Similarly, if the user is a technical one, it is on us to adapt to that context. To empathize with a user experiencing an engineering problem might just mean we have to open an IDE and get coding.

Addendum C | Nine Principles for Designing Great Data Products

Nine Principles for Designing Great Data Products
Ricky Hennessy

Data products are everywhere. From wayfinding apps to recommender systems, data enables experiences that are increasingly intelligent and personalized. However, they come with a unique set of challenges.

From data collection to analysis, to the usage and communication of results, using data to deliver quality customer experiences takes careful consideration. Here are nine principles for designing engaging data products that your users will love.

1. Collect data passively

Collecting user data should never interfere with the quality of the user's experience. While privacy is certainly an important issue, consumers have growing expectations for how their shared data can be used in exchange for better experiences. This is especially true for millennials—80 percent say they have some or a lot of trust in the companies they do business with to keep their personal information secure. Smartphones present an enormous opportunity for passive data collection: accelerometer data, GPS data and app usage data can all be used to learn about your users and to provide better user experiences. Google Maps uses GPS data from its millions of users to provide the fastest routes and help users avoid traffic.

Passive data collection also unlocks hidden business value. For example, most companies collect clickstream data, but few go beyond the use of standard analytical solutions. At frog, we helped a Fortune 100 client assess one year's worth of clickstream data from its web application to better understand their users' behaviors. What we found were ways to redesign the interface to help more customers complete transactions digitally, rather than contact the call center. To do this, we generated User Interaction Flow Diagrams, which visualized the flow of users through the experience and allowed us to find pain points. Then, the data was used to generate user behavior models able to estimate the impact a UI change would have on the overall digital completion rate. The result is a better experience for customers—and call center cost savings of over $600k per year.

2. Don't Exhaust the User

Until a user interacts with your product, no data exists and personalization is impossible. Active data collection techniques can help your product overcome the cold start problem, but they have to be a natural part of the experience. Successful data products get around this by giving users an easy and engaging onboarding experience capable of collecting the necessary data without being overly burdensome. Apple Music asks new users to tell us

what you're into and presents a few bubbles containing genres to select. Stitch Fix guides users through a questionnaire that helps ensure their first Fix contains items they'll love. Netflix asks new users to select three movies they like at sign up; they handle the rest over time.

Onboarding surveys aren't the only solution. Frog worked with a Fortune 500 financial client to design a web application for helping customers find the perfect car. We built a sophisticated recommender system that utilized the attributes of every car on the market, as well as the preferences of users that interacted with the application. However, the application was unable to provide recommendations to new users. We could have designed an onboarding survey to get around this problem, but it would have been a tedious experience. Instead, we used existing survey data to assign attributes such as Sporty, Family Friendly, Luxury or Unique to each automobile. Then, we presented users with an interface that allowed them to toggle between these attributes and view recommended vehicles with a single click. The aspirational nature of the attributes kept users engaged—and we were easily able to collect large amounts of user data that could be fed into the recommender system.

3. Constantly Validate with Data

Launching your data product is only the beginning. Once users start to engage, it's important that you are continually validating your data product by tracking important, quantifiable metrics. The world is always changing, and a model that works well today will not work well forever. Additionally, tracking important metrics gives you the ability to perform experiments, or A/B tests that can help you improve the performance of your data product. Airbnb is constantly running hypothesis driven experiments by iterating on the user experience and product offerings. This includes anything from changes to the appearance of the website to optimizations for their smart pricing algorithm. By leveraging an internal tool used to perform A/B testing, Airbnb can measure the impact changes have on important metrics like click-through-rate or the number of bookings.

While collecting user feedback to improve the user experience is important, the best data products can instantly and automatically incorporate feedback in to the overall experience. For the vehicle recommender application, frog added a button to recommended vehicles that allowed users to place the vehicle in their Garage. This let users view all their favorite vehicles on a single page, but also provided an excellent mechanism for us to collect feedback. This feedback was stored in our database, which we used to calculate the recommendations in real time. By storing user feedback in the same database that was used to make recommendations, the performance of our vehicle recommendation system improved as the number of users grew.

4. Give Users Control

An overly eager machine learning system that makes too many decisions, however accurate, will leave users feeling bewildered and frustrated. Yet, striking that perfect balance between anticipating needs and giving users the right amount of control can be challenging. Designers at Nest Labs learned this principle through experience: Making users fight against temperature schedules they did not select or want caused not only irritation and discomfort, but also thermostat usage that resulted in higher energy usage than before. By nature, people don't like being told what to do. For the Nest Thermostat, letting users feel in control led to a better experience and to increased energy efficiency. Their initial Auto-Scheduler algorithm was optimized to reduce energy costs, but because they failed to take the end user experience in to account, this algorithm led to higher energy usage. The Nest designers listened to their users and updated the Auto-Schedule algorithm to ensure comfort and respect user inputs.

In a frogVentures collaboration with Heatworks to bring the MODEL 3 connected tankless water heater and app to market, one of the primary goals was energy conservation. Part of this conservation is achieved through increased heating efficiency. However, the majority comes through encouraging users to use less hot water. The straightforward solution to increased energy efficiency would be to put strict limits on the amount of hot water a household could use in a day, but that would lead to frustration and attrition. Instead, the data collected by the MODEL 3 provides the user with historical savings, goals and recommendations that encourages the user to take control of their own water conservation.

5. Meet Unexpressed Needs

Collecting user behavior data, passively or actively, is only part of creating great data products. It's also crucial to understand how to use that data to anticipate the needs of users and respond accordingly. Tracking clickstream data, purchase data and any other user behavior data gives us the opportunity to create models of customer behavior that can be used to predict future behaviors. It also helps segment users into groups to develop personalized recommendations. Predictive texting on your iPhone, Netflix's personalized recommendations or Mint's budgeting advice all rely on massive amounts of user behavior data to provide you with timely information that meets your needs. Despite the wide variety of uses, these predictive applications all rely on the same approach: finding correlations in historical user data that can be used to predict unmet needs.

Connected vehicles represent a new opportunity to collect massive amounts of user behavior data that can be utilized to anticipate the wants and needs of the user. Working with a Fortune 500 insurance client, frog designed and built a mobile application that collected driving behavior data from GPS and accelerometer smartphone data, as well as an Automatic ODB port reader. The mobile app used this data to offer timely, location-relevant promotional deals and financial advice, as well as encourage safe driving behavior.

6. Invoke Discovery and Delight

Recommender systems are one of the most common types of data products. Providing your users with high quality recommendations keeps them engaged by providing personalized content and product recommendations. But what is a quality recommendation? Simply providing the most relevant recommendations can lead to obvious or boring results. To truly capture the attention of users, we need recommendations that invoke discovery and delight—serendipitous content users will enjoy but wouldn't have thought of on their own.

Out-of-the-box recommender system solutions provide a passable experience, but it will take a truly bespoke solution to create a sense of discovery for your users. At frog, we developed a web application to help users find the perfect college. Behind the interface is a hybrid recommender system that combines content recommender (in this case, a list of schools) and collaborative recommender (schools similar to ones you like) systems. A hybrid approach offers relevant and serendipitous results, and leads to a richer customer experience.

7. Build Thrust with Transparency

Even if your data product is working properly, users will be skeptical to engage with your product if they don't have any understanding of how a decision was made. Providing transparency into the inner workings of your data product can help earn the trust of users. For example, Spotify will make recommendations with the tagline Because you listened to… By providing this information, users will have a better understanding of what to expect and can make a better-informed decision about what to listen to next. Many machine learning algorithms generate a probability or confidence score in addition to a prediction. Sharing these confidence scores with users can help them make informed decisions. This is commonly used in weather forecasting, where users are given a percent chance of rain instead of a binary result of rain or no-rain.

Transparency is especially important in areas such as healthcare and finance, where decisions can have major consequences. Working with one of the largest banks in Mexico, frog created a dashboard to be used by bank employees assisting customers. This bank served primarily low-income customers, who tended to complete transactions in person. With this insight, we created a dashboard for bank employees that displayed relevant customer information. It then presented a set of recommended actions and financially responsible guidance personalized to the customer that a bank employee could then offer. For each recommendation, the system displayed the rationale, citing relevant information such as recent life events or payment history. Including this extra layer of transparency allowed bank employees to have confidence in their recommendations.

8. Visualize the Complex

Making data easy to interpret is essential when designing great data products. Wayfinding apps help commuters easily avoid thick red lines of heavy traffic. Fitness tracker apps show simple charts and trend lines that can be understood at a glance. News sites like FiveThirtyEight use data visualization to help readers understand complex stories and concepts. Data visualization is everywhere, yet visualizing data without becoming overly complicated or busy is a difficult balance to strike. Making careful use of location, shape, color, size, weight, motion and other means of visually encoding information draws attention to important information.

Data visualization becomes increasingly challenging as the size and complexity of the data to be visualized grows. This challenge is especially common in IoT applications, where massive amounts of streaming data need to be visualized in real time. At frog, we worked with a leader in the oil and gas industry that was collecting sensor data from the instrumentation in their drilling equipment to assess the integrity of their wells. By employing a user-centered design approach, frog created a data visualization dashboard that allowed both technical and non-technical employees to make decisions using this complex data.

9. Blend In

Sci-fi movies portray a future where machine learning and AI exists among us as robots, intelligent chat bots and fully autonomous vehicles. While many articles about machine learning focus on these far-fetched realities, the truth is that machine learning is already in our lives today in much more subtle ways. Often, users are unaware of the sophisticated machine learning that powers their favorite products. They simply notice the improvements these algorithms inform. The lesson here is that the best data products will be ones that work with the way people live today by integrating with the products they already use.

At frog, user experience is always top of mind, and we perform extensive research to ensure the products we design fit in to the users' lives. When designing data products, it's tempting to create futuristic products that force users to change their behavior. While it's important to push the boundaries, the best way to keep users engaged is to create intelligent, responsible data products that fit seamlessly in to customers' lives.

Addendum D | Ten Skills to become a data scientist

Ten skills to become a data scientist
Chris Howsett

1. SQL — SQL is the coding language of databases. There are variations like Presto SQL, Hive and Server SQL which keep things interesting. This language has been around for a long time and is a basic foundational skill set needed by every analyst. I use SQL every day, without fail. It's the workhorse of the analytics industry.

2. Statistics —Basic statistics skills are essential for analysts. Without them, analysts are restricted to data munging and basic metrics. A good knowledge of tools like regression, machine learning, Bayesian statistics and descriptive statistics give analysts a strong foundation to perform solid analyses.

3. R or Python or Statistics Software —Knowledge of a statistics software tool is essential in today's world of Big Data and open source mania. Analysis of big data sets is just not possible without knowing one of these tools. Most analysts I meet are in either the R or Python camp. SAS also has a big community of analysts. Personally, I prefer R because of the brilliant, cooperative community that contributes packages to the R codebase. I suspect Python is just as flexible. Other options include SAS, MatLab, SPSS, Sage and Mathematica.

4. Visualization — Data visualization is a fundamental skillset for analysts in order to communicate results with the wider organization. This isn't about creating beautiful charts. This is about visualizing data in a simple, accurate and concise way so that your audience can digest and act on the information. This is harder than it sounds and is a skill I'm continually trying to improve.

5. Writing — Writing is probably one of the most under-estimated skills for the modern analyst. The ability to write insights in a concise and simple way is a must-have. I also strongly recommend analysts learn and use active-voice as much as possible. It was make your writing clearer and less verbose. This is also a skillset I try to continually improve and still have a lot to learn.

6. Presenting — Presentations are a regular part of communicating results. A good presentation can mean the difference between the organization using the data or not. A good tip I once received was to always use a specific example (e.g. this city on these dates showed X, Y and Z) when explaining results. I try to give presentations to different audiences when I can to help build this muscle.

7. Design Thinking — The practice and methods behind Design Thinking have a lot of practical uses in the world of analytics. First, it's solution-based meaning the focus in practical solutions. Secondly, the critical driver of Design Thinking is the user. So Design Thinking can have be beneficial for analysts designing dashboards, developing visualizations and creating measurement tools . For more on Design Thinking, I highly recommend this short course by IDEO and Acumen.

8. Business Analysis — Business Analysis is a skill-set whereby you learn how to breakdown business problems, outline relevant business processes and think about ways to improve/make the process more efficient. Business Analysts are a professional body in their own right but a basic understanding of business analysis techniques has a number of benefits for analysts. Ultimately, an analysts job is to provide insights that will improve a component of the business whether it is marketing efficiency, product launch or branding. Analytics professional can draw on business analysis techniques to think through ways to apply insights to a business problem. More on business analysis here.

9. Time Management and Prioritization— Most analysts have more work than time available in the day. So time management and prioritization is a crucial skill to be productive and maintain sanity. At the moment, I use a task sheet and two-week Sprints to manage my projects. The task sheet is available to all of my stakeholders and partners so that they have a clear line of sight into my workload. My partners also use this to help me prioritize, if necessary. I try to pro-active with updates so that stakeholders keep up-to-date with progress so that everyone is clear on priorities.

10. Learning — Analytics is ever-changing industry. While this ten skills are great foundations, there's also a need to continue learning and professional development. Even as analysts move into managers, it's important to stay in touch with the industry and common tools. A healthy dose of ongoing professional development has always worked in my favor and something I highly recommend to other analysts.

These are 10 skills that I've come to rely on daily throughout my career. They are tried and tested — each one standing the test of time. And depending on your background and experience, some of these skills may come quicker than others.

I've found these skills have been crucial throughout my analytics career. Each one of them has been universally relevant and necessary regardless of where I'm working or the projects I'm asked to support.