

Analytics and Applications

Lecture 1: Introduction to Data Science and Business Analytics

First, let me introduce myself...

**Prof. Dr.
Wolfgang
Ketter**



- **Chaired Professorship** at University of Cologne and Rotterdam School of Management
- **Director** of the **Erasmus Centre for Future Energy Business and Learning Agents Research Group** at **Rotterdam School of Management**
- **Visiting Professorship** at **UC Berkeley**
- **PhD** in Computer Science and Information Systems from the **University of Minnesota**

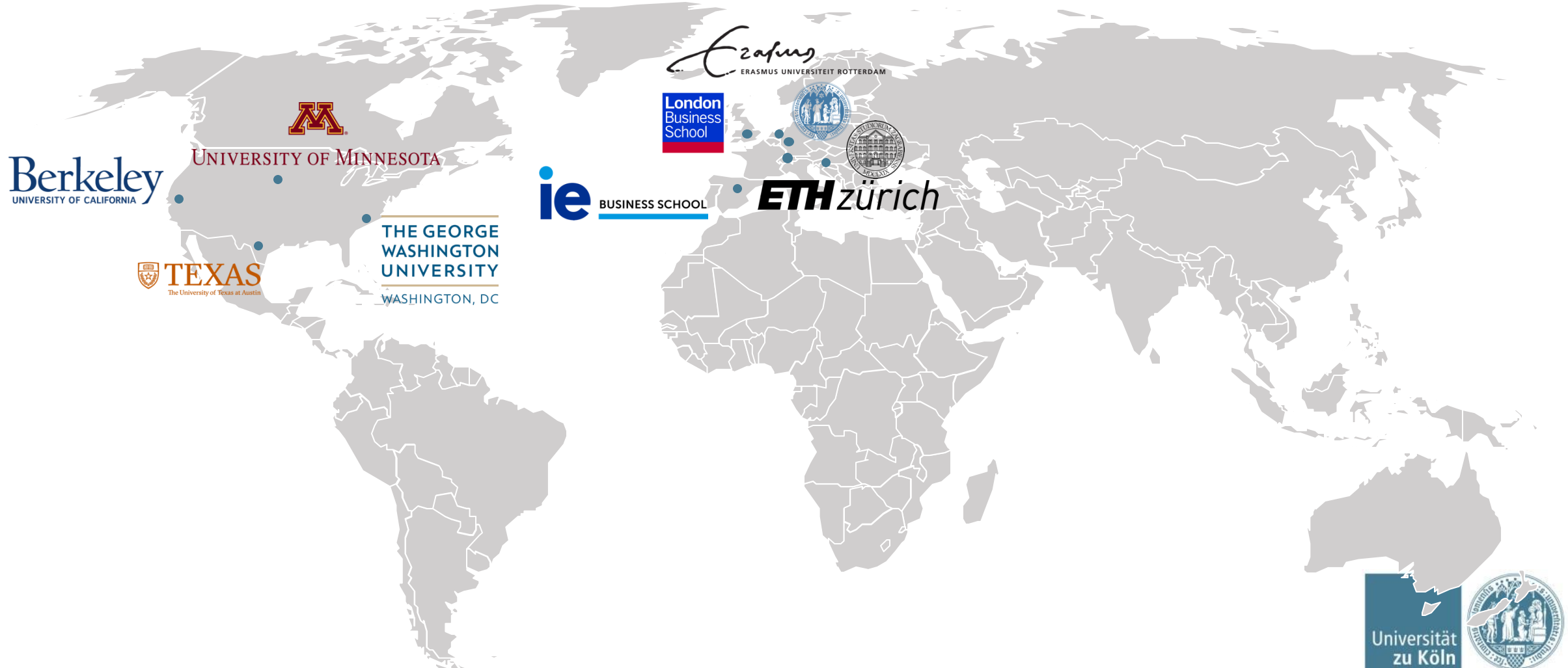
Janik Muires, the TA for this course

**Janik
Muires**

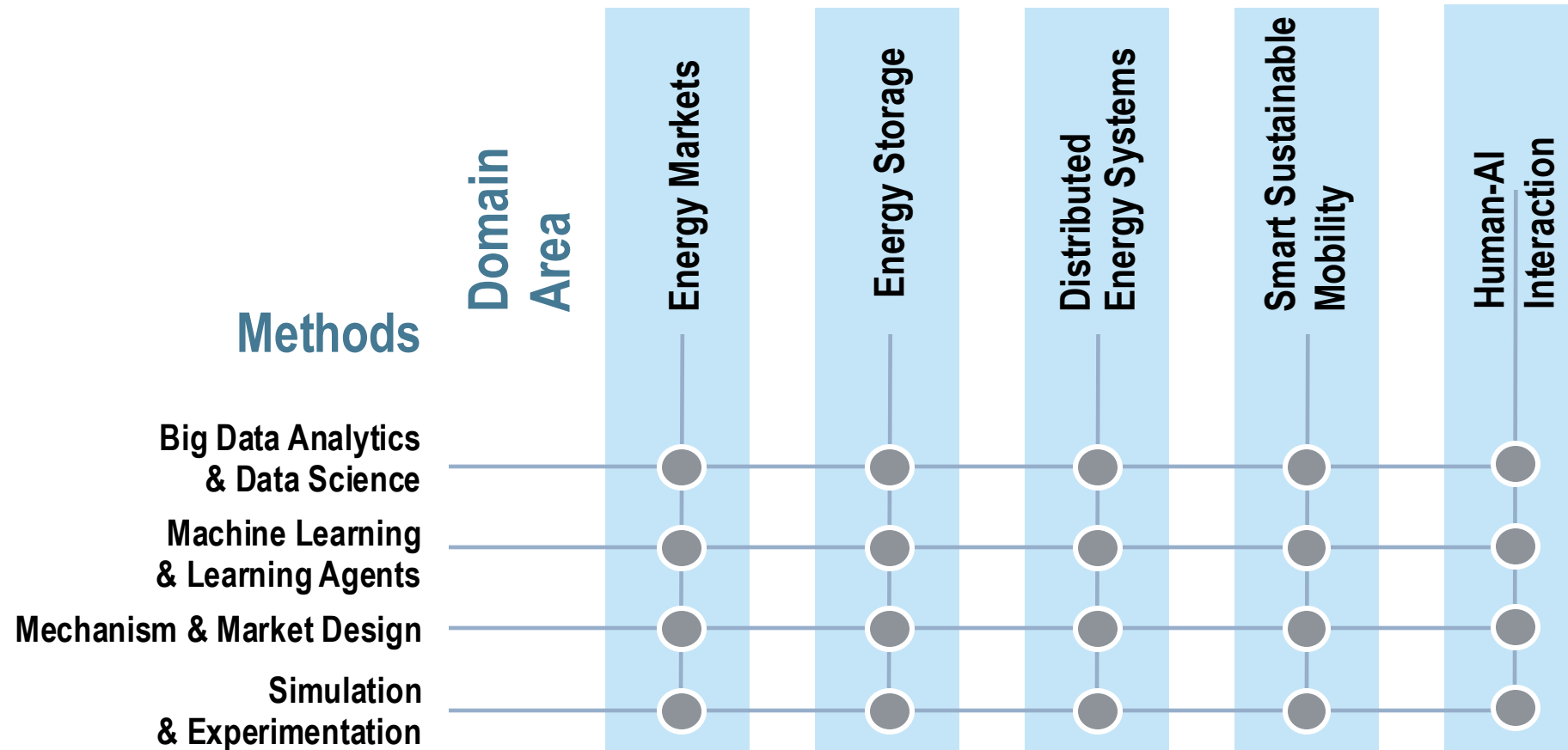


- **PhD Researcher** at the Chair for Information Systems for Sustainable Society
- B.Sc. and M.Sc. in **Information Systems** from University of Cologne – knows the ins and outs of this university
- More than five years experience in business intelligence, data science, and software development
- Research interest in smart sustainable mobility, simulation of mobility systems, and big-data driven empirics

We interface and **collaborate closely** with **international research groups** from top institutions around the world



Our research is highly interdisciplinary – We apply **big data**, **data science** and **machine learning** to societal challenges in **energy** and **mobility**



As part of our research we collaborate with research **partners** from **industry** and **public institutions** on a regular basis (examples)

Selection of recent research partners

 **DriveNow**

green | connector
con | energy gruppe

ProCom
bringt Transparenz



UBER



Selected relevant projects

- **Profitable virtual power plants** for electric fleet vehicles
- **Car fleet management** and repositioning via behavioral nudging
- Identification of **economic regimes** and ML-based **price forecasting** in electricity spot markets
- **Electric vehicle adoption forecasting** and **grid integration** study
- **Decentral marketplaces** for peak shaving of large industrial customers
- **Auction re-design** for world's largest descending-bid auction

We publish this research in **top-ranked international journals** in Information Systems and Energy

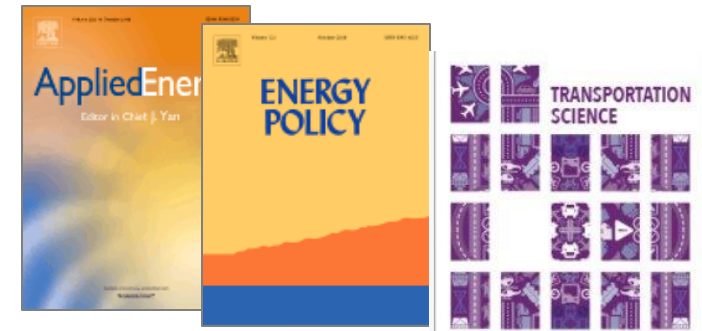
Top-notch **IS** journals



Recent examples

- Kahlen, Ketter, & van Dalen. **Electric Vehicle Virtual Power Plant Dilemma: Grid Balancing Versus Customer Mobility.** In Production and Operations Management
- Peters, Saar-Tsechansky, Ketter, Williamson, Groot & Heskes. **A Scalable Preference Model for Autonomous Decision-Making.** In Machine Learning
- Ketter., Peters, Collins, & Gupta. **A Multiagent Competitive Gaming Platform to Address Societal Challenges.** In MIS Quarterly.

High-impact **transportation & energy** journals



Recent examples

- Ahadi, Ketter, Collins, Daina. **Cooperative Learning for Smart Charging of Shared Autonomous Vehicle Fleets.** In Transportation Science
- Hentschel, Ketter, and Collins. **Renewable Energy Cooperatives: Facilitating the Energy Transition at the Port of Rotterdam.** In Energy Policy
- Fridgen, Kahlen, Ketter, Rieger, and Thimmel. **One Rate Does Not Fit All: An Empirical Analysis of Electricity Tariffs for Residential Microgrids.** In Applied Energy

Our teaching portfolio consists of both **analytics and domain-based courses** at Master and PhD level (and Bachelor)

		Courses	Seminars
Bachelor		Introduction to Data Science and Machine Learning [DSML]	Next Generation Information Systems
Master	Basic	Analytics and Applications [AA]	Advanced Seminar Information Systems for Sustainable Society
	Advanced (+ PhD)	Advanced Analytics and Applications [AAA]	Applied Mathematical Optimization [AMO]
			Information Systems Research: Analytics for a Sustainable Society

Now let's get to know you!

- Join at **www.kahoot.it** or with the **Kahoot! App**

Kahoot!

Agenda



Course Goal



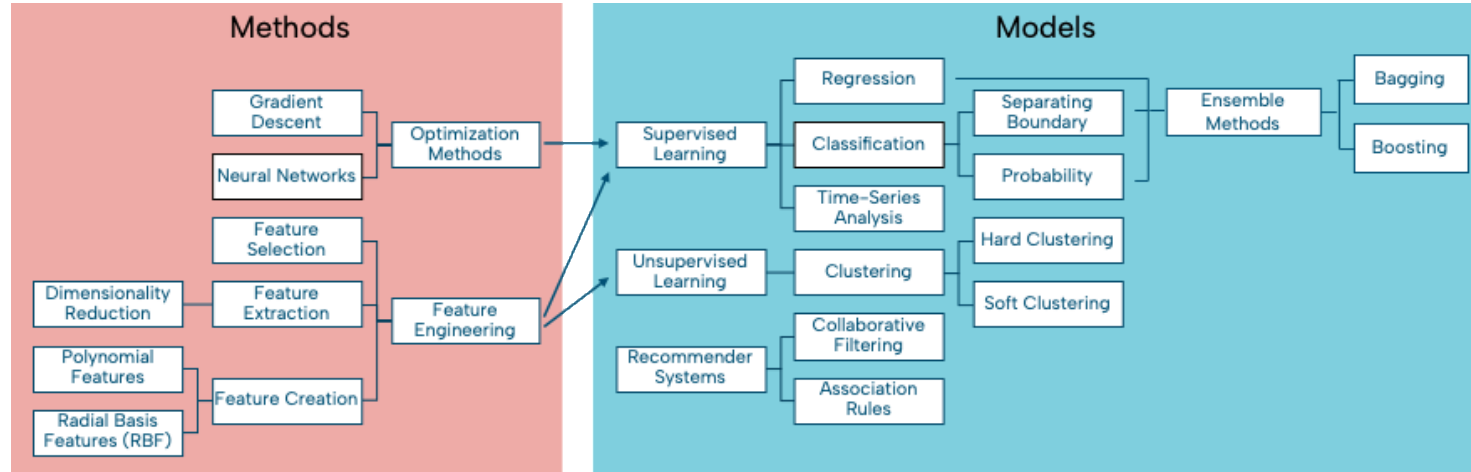
This course will help you to possess valuable practical analytical skills that will equip you with a competitive skill in almost any contemporary workplace.

Course Objectives

The course will provide participants with the following skills and knowledge:

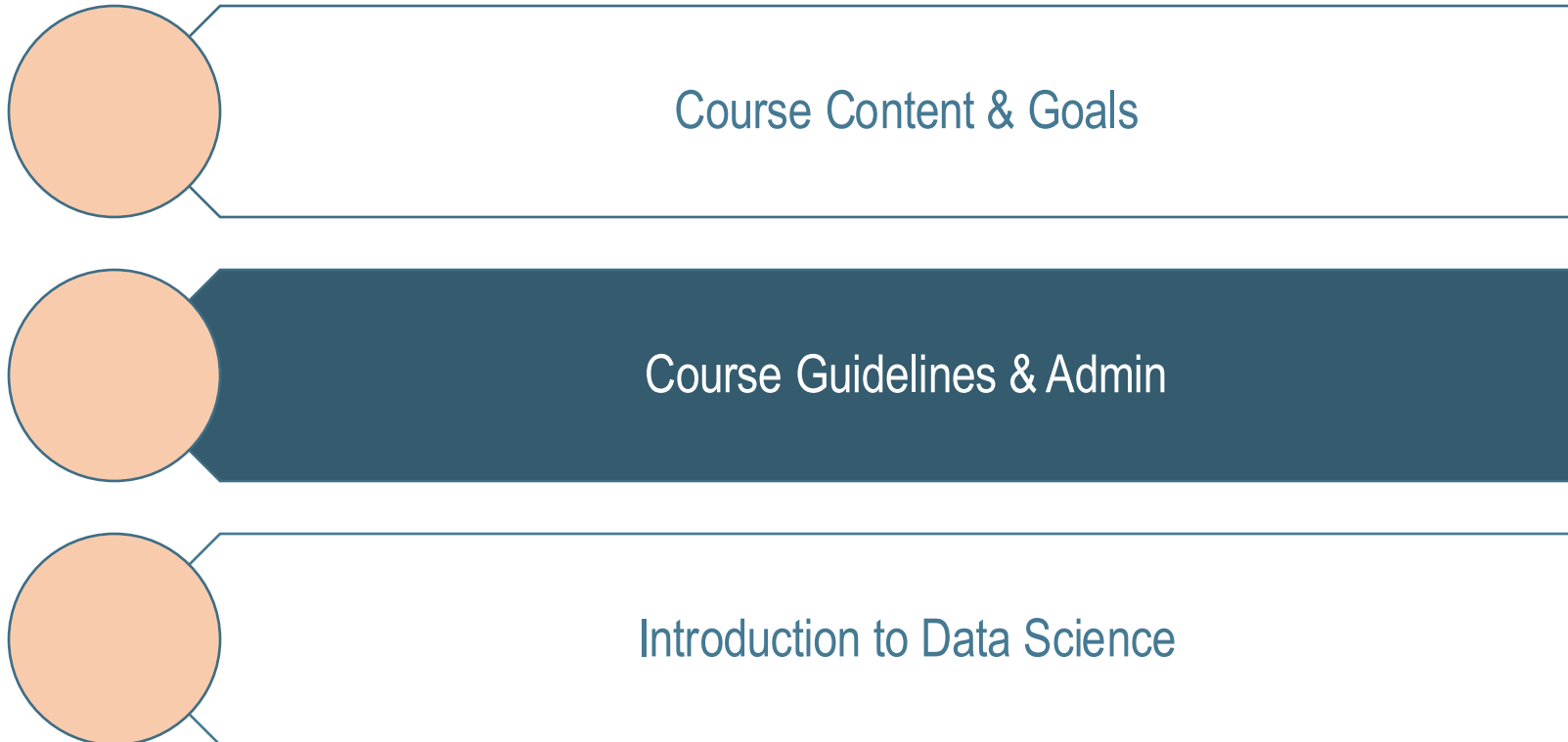
- Become familiar with the potential and dangers of data mining in today's data-rich environment
- Gain experience using key data mining methods of classification, prediction, data reduction and exploration
- Know how to decide when to use which technique
- Be able to implement major data mining techniques using software
- Become a smart and critical user and consumer of data mining techniques
- Gain the intellectual capital required to provide responsible business analytics services
- Be conscious of humanistic and societal implications of deploying data mining in organizations

Course Topics



- This is a preview of what you will get to know during this semester
- This might look overwhelming, but we will build of this complex structure of connected topics slowly over time
- We will look at **models** and **methods**
- You will see that we usually need both to solve a data science task

Agenda



Not on ILIAS? – Make sure you join!

- ILIAS Course Name: **Analytics and Application** (can be accessed [here](#))
- ILIAS Password: **AA_2024!**

We will now run through some of the key admin points of the **course syllabus** – Make sure you download a copy from ILIAS and read it thoroughly

#	Title	Topics	Recommended reading
#1 Oct 12	Kick-off	Introducti & announ	
** #2 Oct 19	Data Mining Process	Core idea diction	
#3 Oct 26	Supervised Learning: Regression (1/2)	Linear reg	
** #4 Nov 02	Supervised Learning: Regression (2/2)	Performa	
#5 Nov 09	Supervised Learning: Classification	Naive Bay Logistic r	
** #6 Nov 16	Supervised Learning: Classification and Decision Trees	Classificat evaluation	
#7 Nov 23	Supervised Learning: Combining methods	Ensembles	
** #8 Nov 30	Supervised Learning: Artificial Neural Network	Intro to A	
#9 Dec 07	Unsupervised Learning: Cluster Analysis	Hard Clu Hierarchic	
** #10 Dec 14	Unsupervised Learning: Cluster Analysis	Soft clust with PCA	
#11 Dec 21	Unsupervised Learning: Association Rules	Introducti filtering	
** #12 Jan 11	Times Series Analysis	Handling Smoothing	
#13 Jan 18	Social Network Analysis	Primer G & Eigenve	
** #14 Jan 25	Text Mining & NLP	Introducti Text Mini	
#15 Feb 01	Wrap-up	Course s exam pre	
Feb 01	Deadline: Group Projects	Submit y (12.00h)	
Feb 07	Exam 1	15:00 to inspection	
Mar 10	Exam 2	15:00 to inspection	

Table 1. Tentative Schedule (**

Analytics and Application[AA] 2021/22

Master of Science WI / IS
Information Systems II
Faculty of Management, Economics, and Social Sciences
Department of Information Systems for Sustainable Society
University of Cologne
Version - September 20, 2021

Instructor Prof. Dr. Wolfgang Ketter Term WS 2021
TA Nastaran Nasiri Class Time Wed, 14.00-17.30h
Website www.is3.uni-koeln.de and ILIAS Class Location H113 & Zoom

Welcome to **Analytics and Application [AA]**. This course and the accompanying reading materials aim to provide you with knowledge and skills required for data analytics using information systems that drive business success.

Business analytics is the use of data-driven decision making. Companies, governments, and other organizations now collect and have access to large amounts of data about suppliers, clients, employees, citizens, transactions, etc. Data Mining and predictive analytics provide a powerful toolkit for detecting actionable patterns in data and generating predictions. These methods are used in many industries: Mobile companies use their customer database to predict customer churn or to personalize SMS messages for improving customer service; Financial institutions use past loan data to predict defaulting chance for loan applicants; Charities use data from a campaign in one location to target the right people in another location; Politicians use databases of supporters to segment and best target each audience; Movie rental and e-Commerce websites provide recommendations based on users' online behavior; Renewable energy providers use weather forecasts to predict their electricity generation to better trade in different types of markets.

Most business analytics applications are geared towards benefiting the company, often at the expense of the individual, community, and society. The focus of this course is on human and socially-responsible business analytics, especially with the focus on sustainability.

In this course we will work with real business problems and real data. We will examine types of questions that data mining can answer and will develop a variety of data-driven tools to answer those questions. The emphasis is on understanding the concepts and logic behind a wide set of data mining techniques and their relation to specific business analytics situations. The course is not about mastering the theoretical underpinnings of the techniques. The gained knowledge will be applied on a business analytics team project that encapsulates the learnings, while is not expected to master the theoretical underpinnings of the techniques.

You will learn about the process of data analytics. You will learn to identify problems, to define the structure of an information system, to evaluate competing solutions for a business problem, and you will start to "speak business and analytics". You will gain an insight into how important these activities are in creating information systems that are truly aligned with business needs. Throughout the course, you will also learn about selected topics in sustainability, especially in terms of renewable electricity.

Some of the most important thoughts that we will present throughout the lecture are summarized in this reader and the references therein. This material is designed to help you find additional references, and to help you recapture what was taught in the lectures. It is not designed, however, to serve as an exclusive source of exam preparation material. In our lectures and workshops, we will present facts, examples, and teach you skills that are not on the following pages, and we reserve the right to ask you about them during the final exam of your assignments. Please make sure you take advantage of all modes of learning that we offer in this course!

Cologne, September 20, 2021

Prof. Dr. Wolfgang Ketter, Nastaran Nasiri
Department of Information Systems for Sustainable Society

Key points covered

- Lecture Schedule
- Lecture Content & Goals
- Recommended Reading
- Assessment Guidelines & Grading
- Other Course Guidelines (setup, communication, attendance, submitting work, etc.)

Weekly Lectures and Workshops

Wed

Lectures

Workshop

Lectures: Wed, 14:00 - 15:30

Workshop: Wed, 16:00 - 17:30

Check Syllabus for
detailed schedule!

Note: We might start earlier with the workshop depending on how quickly we get through with the lectures. It is advisable to not just come in for the workshops at 16:00 or you might miss parts of it!

Sometimes we might do a double lecture one week and a double workshop the week after, the syllabus will get updated in that case!

Procedure: Workshops will usually look at the previous week's lecture and we will upload material for the programming part on Fridays the previous week so you can have a look and ask questions in the workshop.

Assessment Guidelines

Written Exam

- **Scope:** 90-minutes, closed book
- **Content:** All materials provided in the Lectures and Workshops.
- **Format:** Written, with some multiple choice in the beginning and some high-level programming task at the end.
- **Date:**
 - Exam 1: February 7th
 - Exam 2: March 14th

67%

Team assignment

33%

- **Set-up:** Teams of students (team size to be announced, depending on class size)
- **Tasks:**
 - **Project report:** details the team project, from the business problem through the data mining problem and solution, to recommendations.
- **Deliverables:** Project reports as well as the Python code.
- **Milestones are passing criteria!**
- **Deadline: January 29th 23:59**

Attention: Both must be passed individually!

More on the Team Assignment

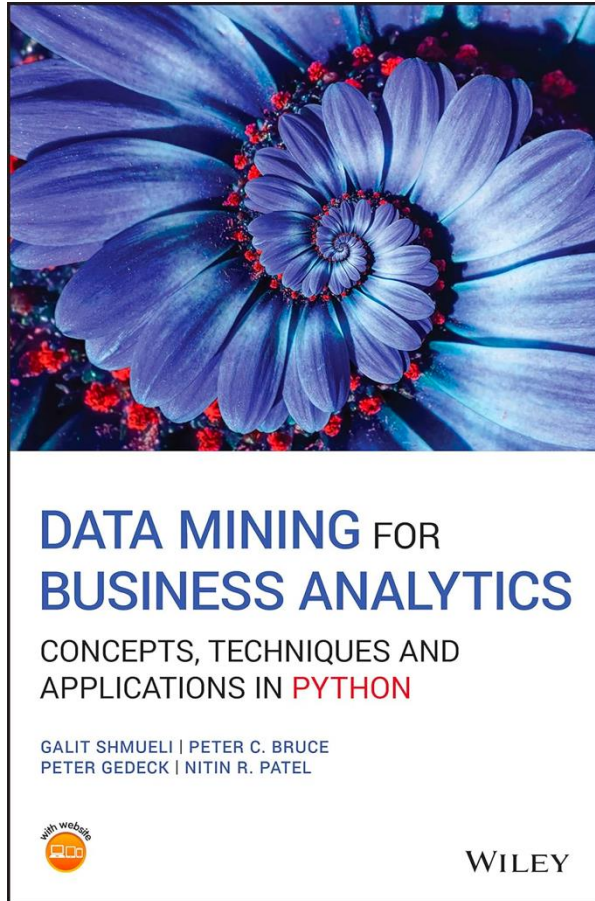
- **Milestones** will need to be handed in by every team and are passing criteria! This means, that if you fail to submit milestones, you cannot pass the team assignment and thus the course.
Don't worry: We do not grade the milestone submissions, but we want to see progress.
- Deadline Milestone 1 (Data Prep): November 27th
- Deadline Milestone 2 (Analysis): January 8th
- At the end, you hand in a **written report** and the **code** used to generate results and figures in the report (preferably you just invite the TA to a GitHub repository).

Official Pre-Text Exams (German version provided at the end of the slides)



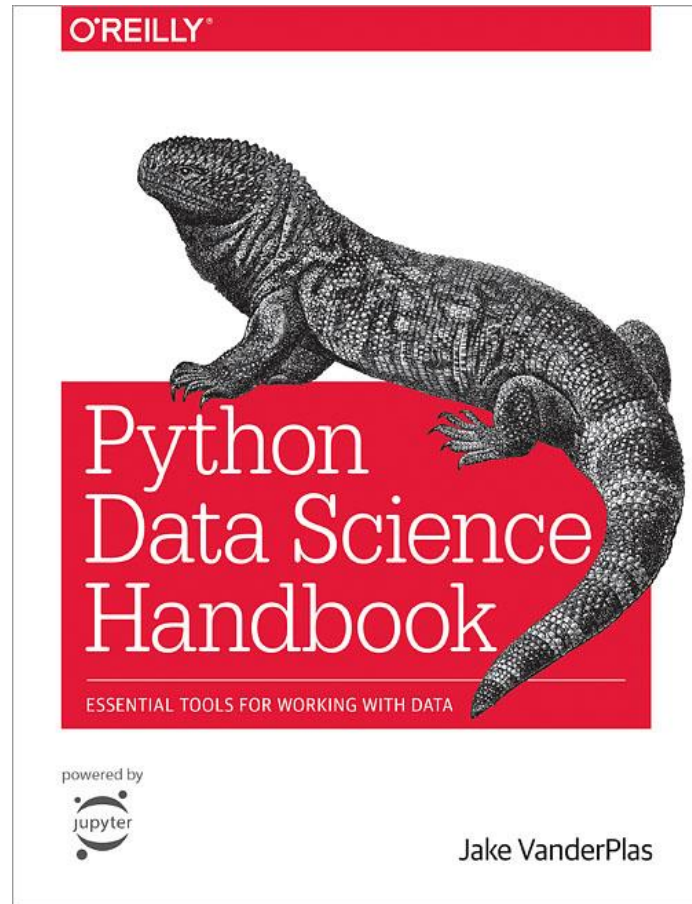
- Please refer to KLIPS for the examination dates and registration deadlines.
- Registration for the examination is required, without registration there is no entitlement to participation in or grading of the examination.
- Here you will find instructions on how to register for exams: <https://uni.koeln/87S2U>
- The deadlines have been set so that after this date it must be clear who is taking part in the examination. After the deadlines have expired, it is no longer possible to register or deregister!
- Assessment will normally take place within eight weeks of the day on which the examination was taken.
- There will be two examination dates for this course. Since the module will not be offered again in the following semester, registration for the second date will also be possible in case of a NB in the first date! The second date will take place late in this semester.

Core reading



- Data Mining for Business Analytics – Concepts, Techniques, and Applications in Python
- Galit Shmueli; Peter C. Bruce; Peter Gedeck; Nitin R. Patel
- Publisher: Wiley

Recommended reading for learning Python

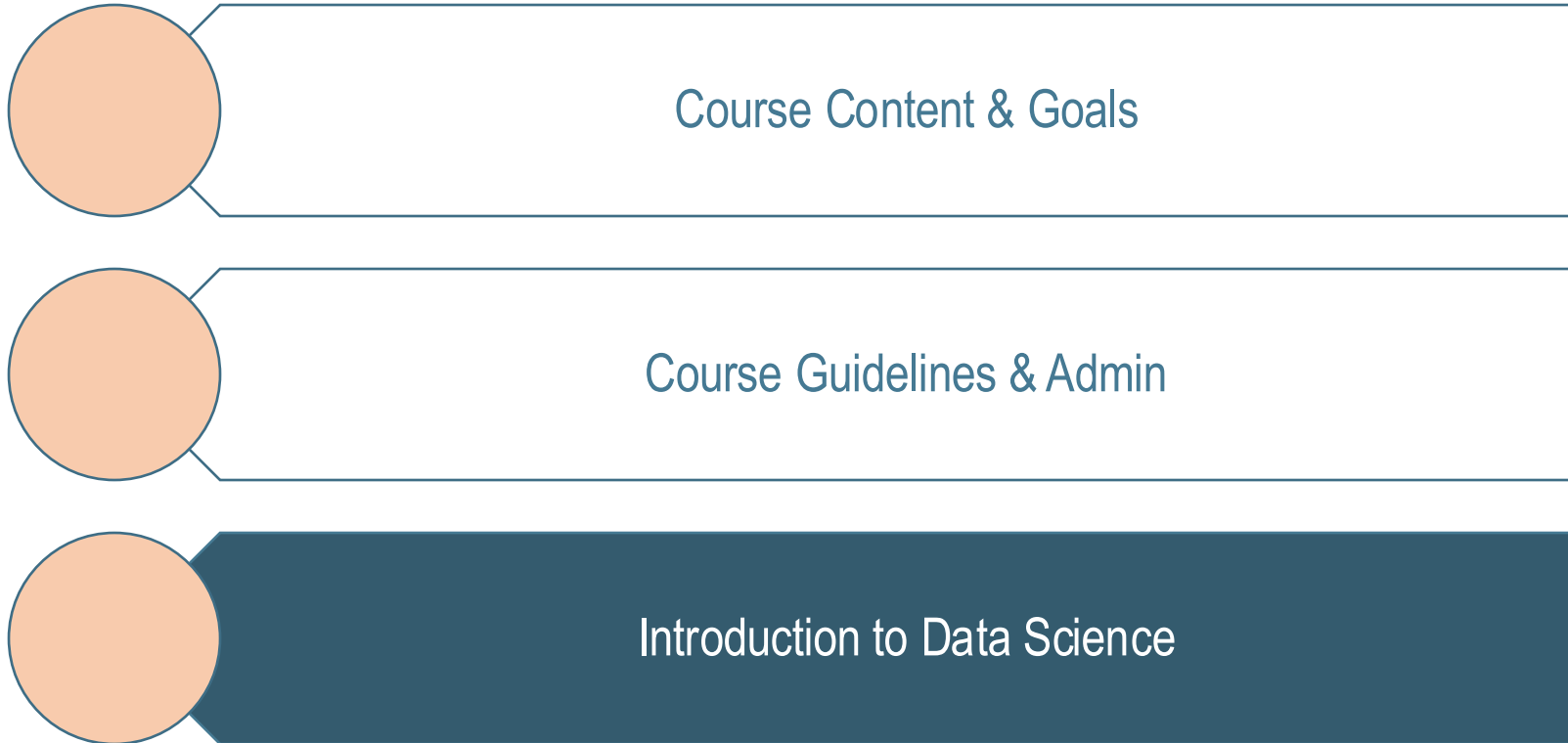


- Python Data Science Handbook – Essential Tools for Working with Data
- By Jake VanderPlas
- Publisher: O'Reilly Media
- Freely available as PDF from [here](#)



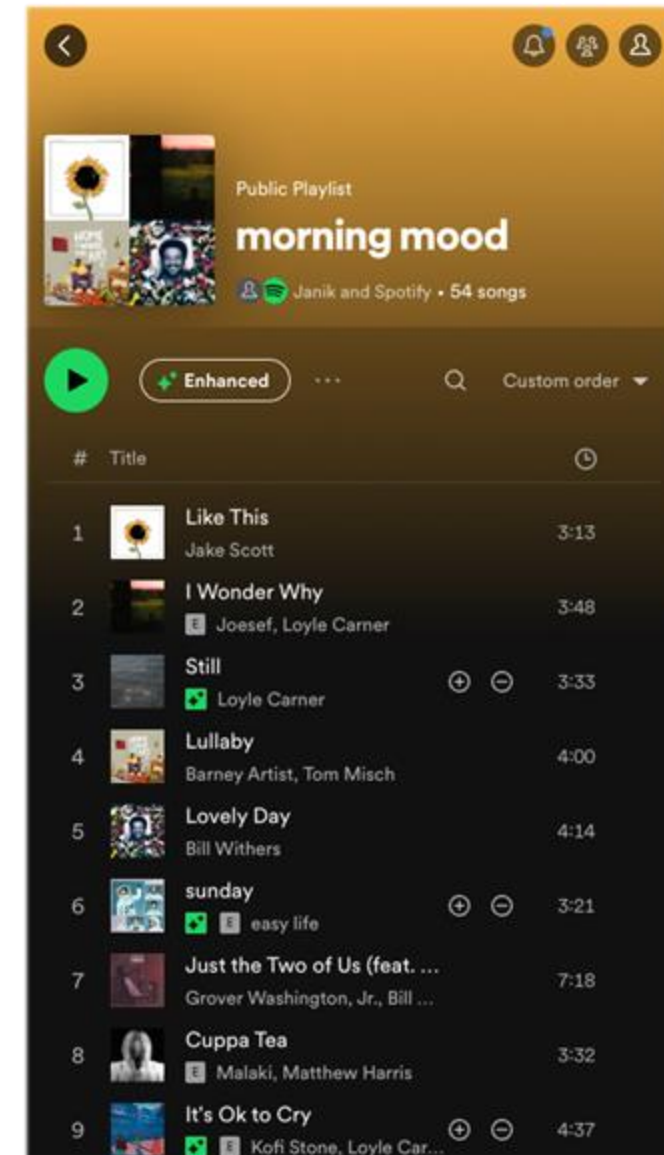
The full **syllabus** is available on **ILIAS** –
Make sure to **download** a copy and **read**
it **thoroughly!**

Agenda



Data Science in Everyday Life

- Spotify recommends songs to add to a playlist:
- This is not curated by a human, but by algorithms
- The more songs a playlist already has, the better are the recommendations
- Feedback in form of “+” and “-” increase accuracy of recommending similar songs



Some Examples of Data Science in Near Future!



<https://www.greentechmedia.com/articles/read/fully-autonomous-vehicles-decade-away-experts#gs.iwDj=Dc>

What about this?

JA Tell me in one sentence why I should attend the following graduate course: <https://is3.uni-koeln.de/en/teaching/courses/analytics-and-applications-m-sc>

You should attend the "Analytics and Applications" M.Sc. graduate course at the University of Cologne because it offers a comprehensive and practical education in analytics.

As of my last knowledge update in September 2021, Professor Wolfgang Ketter is known for his academic focus on various areas related to information systems and decision sciences, including topics such as:

1. **Business Analytics:** He has a strong emphasis on research and teaching in business analytics, which involves using data and quantitative methods to aid in business decision-making, optimization, and performance improvement.
2. **Energy and Sustainability:** Professor Ketter has conducted research in the field of energy and sustainability, particularly in the context of smart grids, energy markets, and sustainability modeling.
3. **Auction Theory and Mechanism Design:** His work extends into auction theory and mechanism design, exploring how different auction mechanisms can be designed to achieve specific economic and strategic objectives.

Regenerate

- Everybody nowadays knows about the likes of ChatGPT
- But is this Data Science?

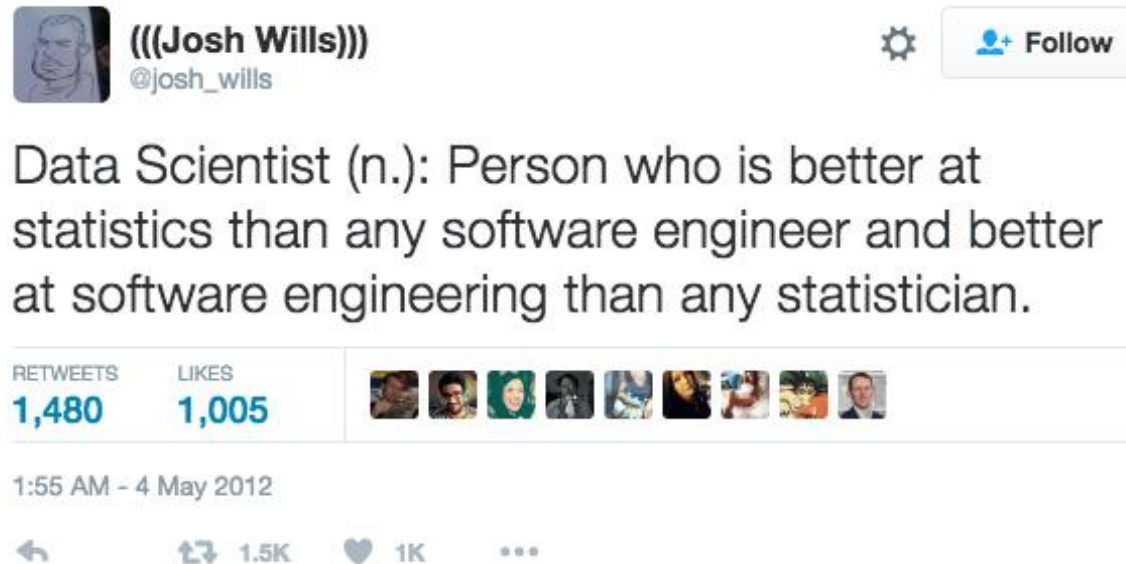
What is Data Science?

- Possible definition:

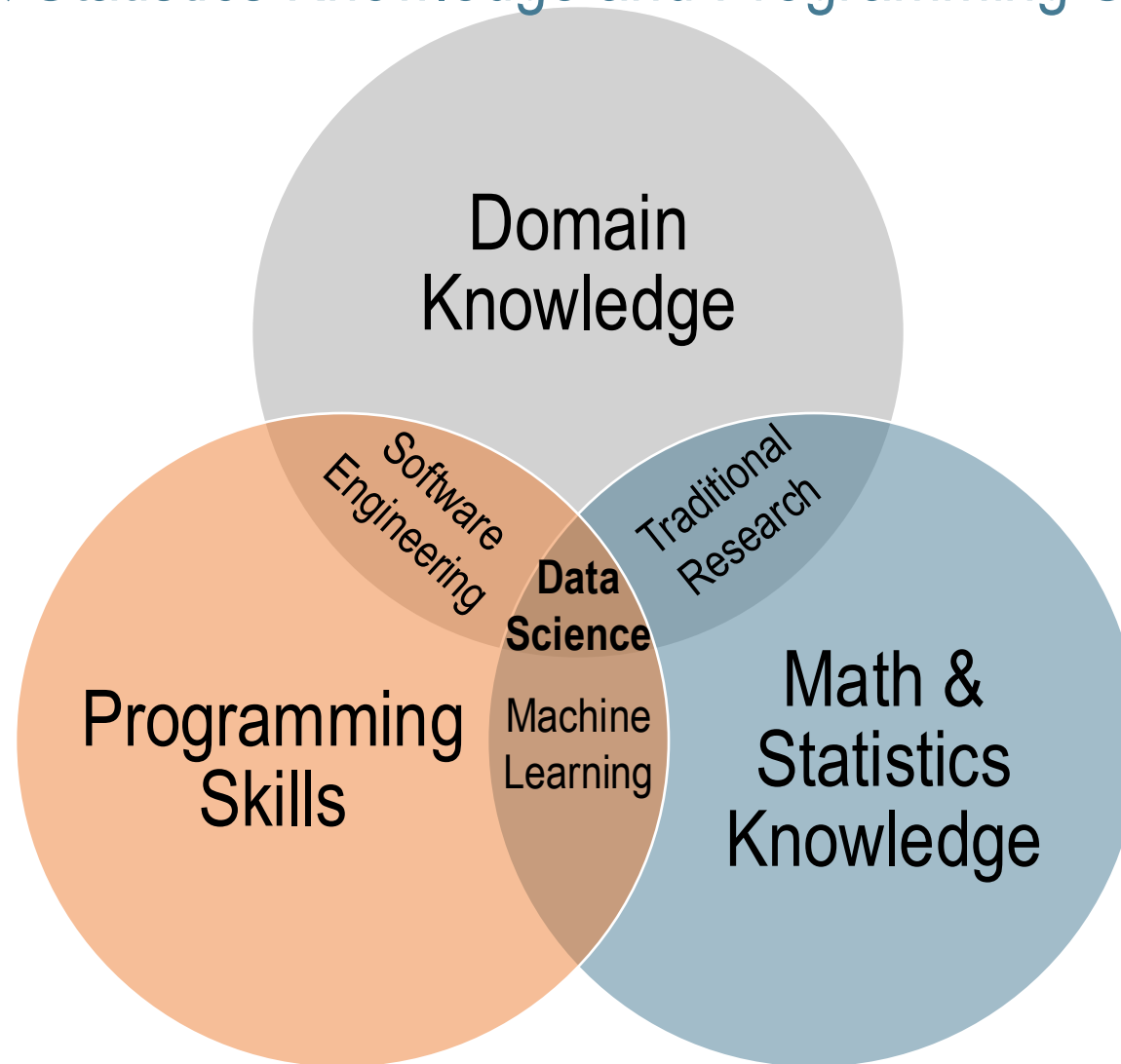
An interdisciplinary field of **computer science** and **statistics** that uses scientific methods, processes, algorithms and systems to **extract knowledge** or **insights** from **data** in various forms, either structured or unstructured

Data Science = statistics + data processing + machine learning + scientific inquiry +
visualization + business analytics + big data + ...

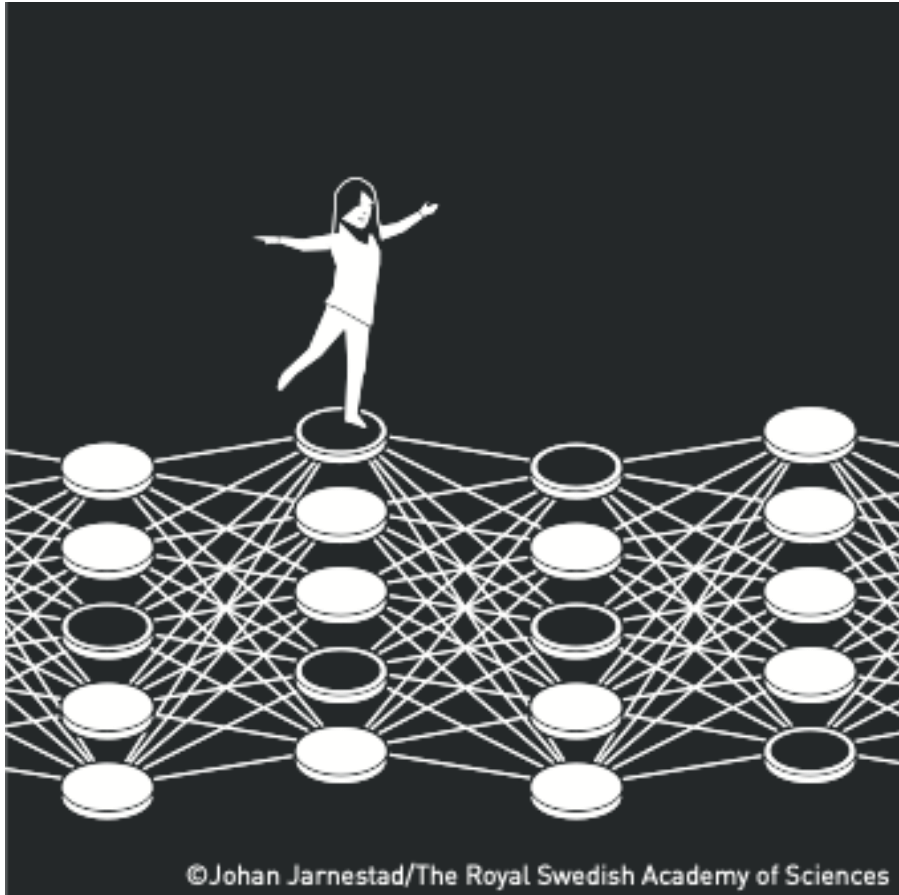
What is Data Science?



In essence a good Data Scientists needs to combine three core skills – Domain Knowledge, Math & Statistics Knowledge and Programming Skills



Just recently: Nobel Prize for Physics goes to ML Researchers!



<https://www.nobelprize.org/prizes/physics/2024/press-release/>

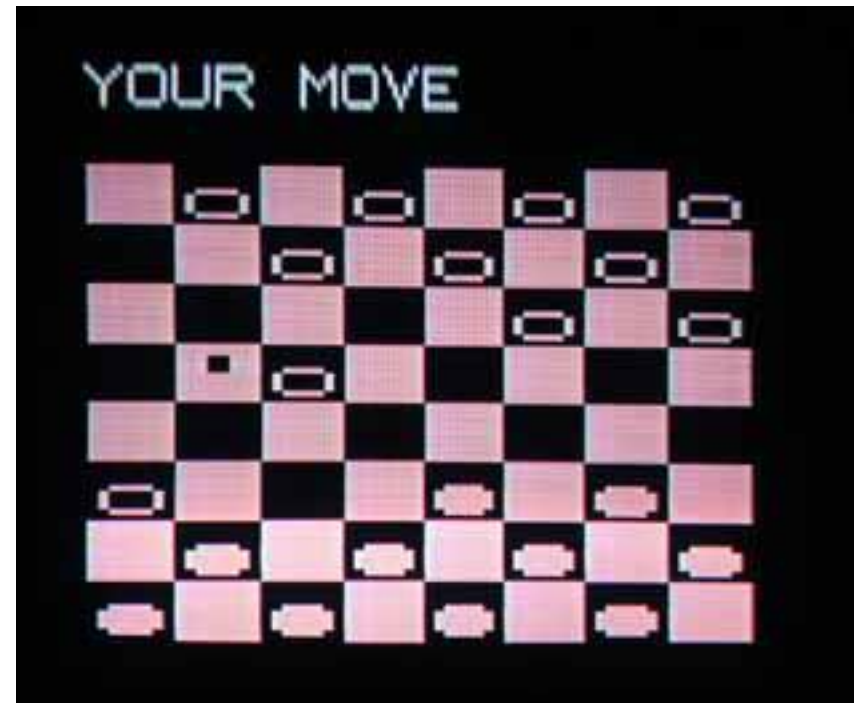
- Two Machine Learning researchers win the 2024 Nobel Prize in Physics: **John Hopfield** (Prof. at Princeton) and **Geoffrey Hinton** (Prof. in Toronto, ex-Google)
- A lot of their foundational research is in use today in modern artificial intelligence.

What is Machine Learning?

Even among machine learning practitioners, there isn't a well accepted definition of what is and what isn't machine learning. Here is some definitions of Machine learning:

- Arthur Samuel(1959): Field of study that gives computers the ability to learn without being explicitly programmed.

Example: Checkers playing program that sees tens of thousands of examples of board positions and learns over time what the good positions are.



<http://incompleteideas.net/book/first/ebook/node109.html>

What is Machine Learning?

- Tom Mitchell(1998) Well-posed Learning Problem: A computer program is said to learn from *experience E* with respect to some *task T* and some *performance measure P*, if its performance on T, as measured by P, improves with *experience E*.
- **Example:** For the checkers play, the experience E would be the experience of having the program play tens of thousands of games itself. The task T would be the task of playing checkers, and the performance measure P will be the probability that wins the next game of checkers against some new opponent.



Suppose your email program watches which emails you do or do not mark as spam, and based on that, learns how to filter spam better.

What is task T in this setting?

- a) Watching you label emails as spam or not spam
- b) Classifying emails as spam or not spam
- c) The number of emails correctly classified as spam / not spam
- d) None of the above—this is not a machine learning problem



Suppose your email program watches which emails you do or do not mark as spam, and based on that, learns how to filter spam better.

What is task T in this setting?

- a) Watching you label emails as spam or not spam
- b) **Classifying emails as spam or not spam**
- c) The number of emails correctly classified as spam / not spam
- d) None of the above—this is not a machine learning problem

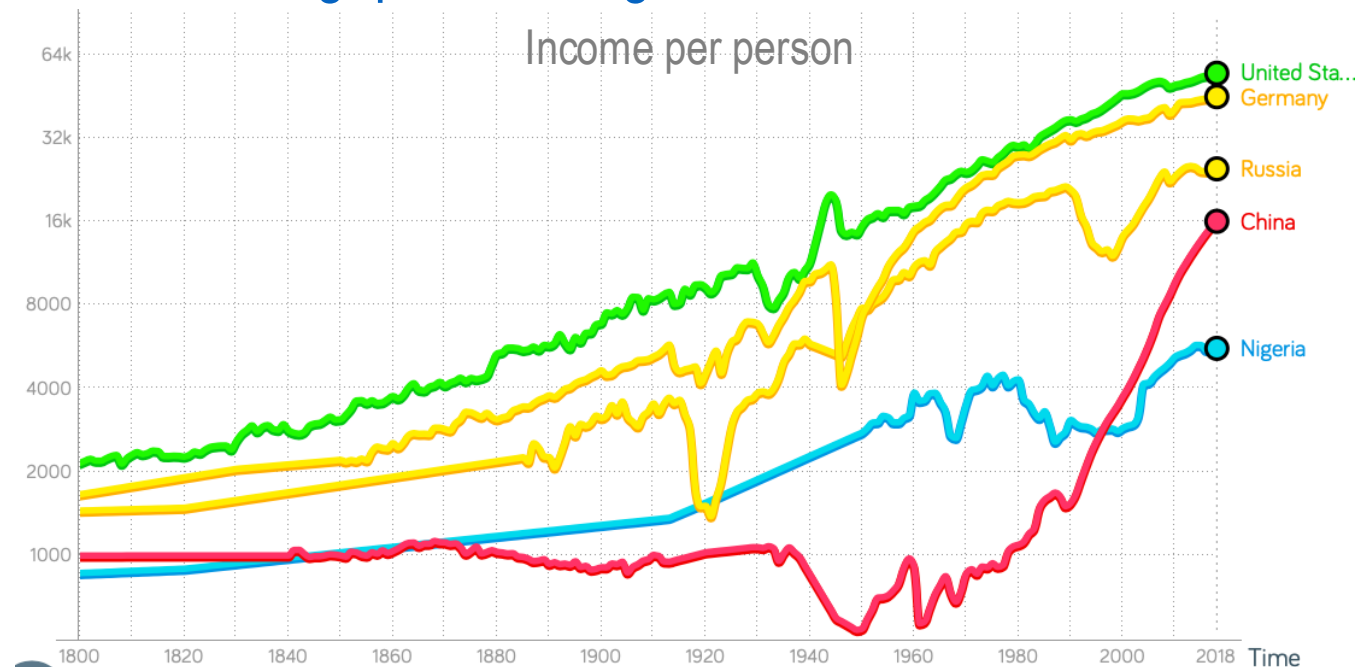
Experience E

Task T

Performance P

Business Intelligence

- *Business Intelligence (BI)* refers to extracting intelligence from data, data visualization and reporting for understanding “what happened and what is happening”.
- BI has evolved over years from static reports into more user-friendly and effective tools, such as creating interactive dashboards that allow the user not only to access the data but also interact with it. See www.gapminder.org

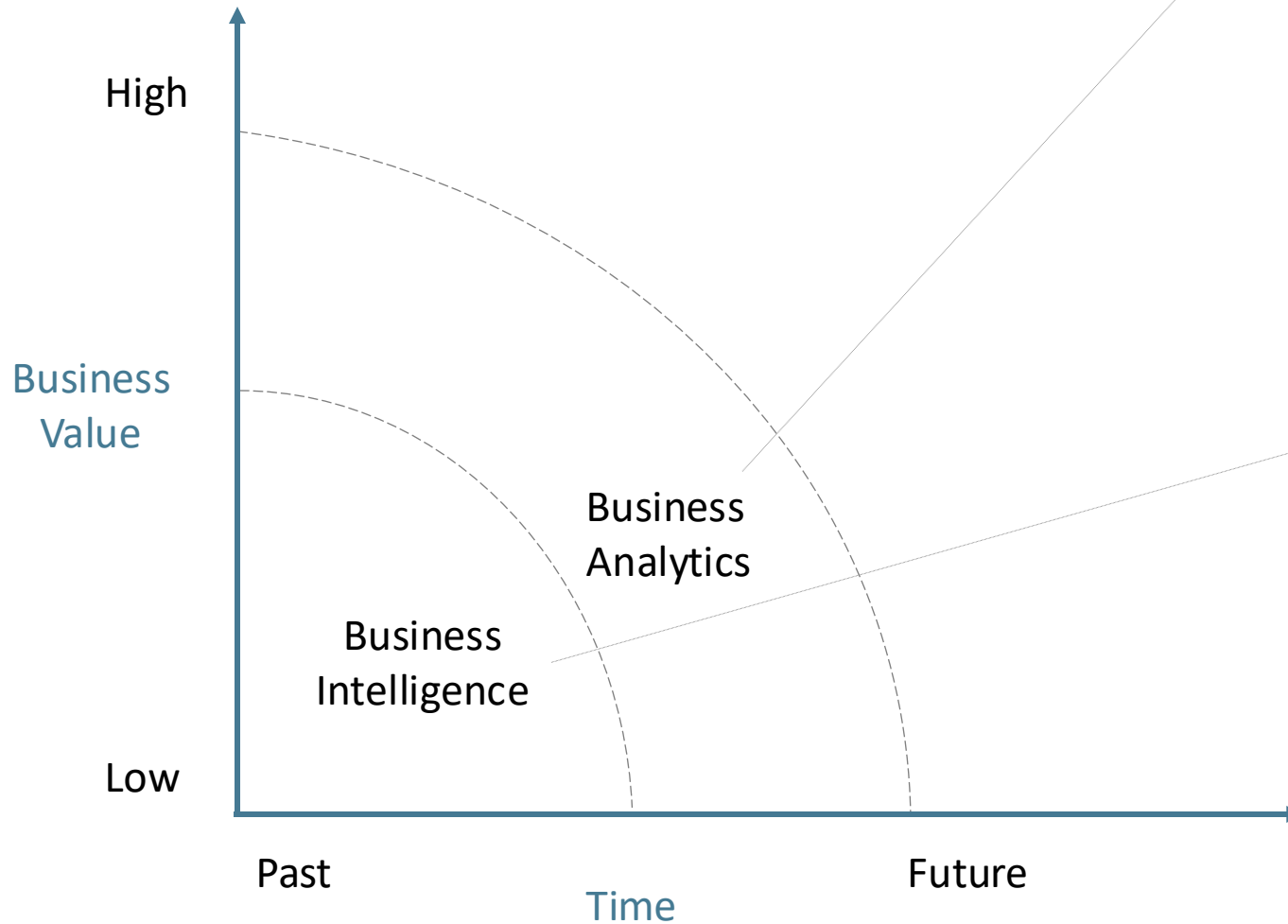


What is Business Analytics?

- *Business Analytics (BA)* is the practice and the art of bringing quantitative data to bear on decision making. However, the term can be defined differently for different organizations



Business Analytics and Business Intelligence



Business Analytics

Typical technologies Data types

- Optimization, Predictive modeling, Forecasting, Statistical analysis
- Structured/ unstructured data, internal/ external data sets (massive data sets)

Common Questions

- What if ...?
- What if these trends continue?
- What is the optimal scenario for our business?
- What will happen next?

Business Intelligence

Typical technologies Data types

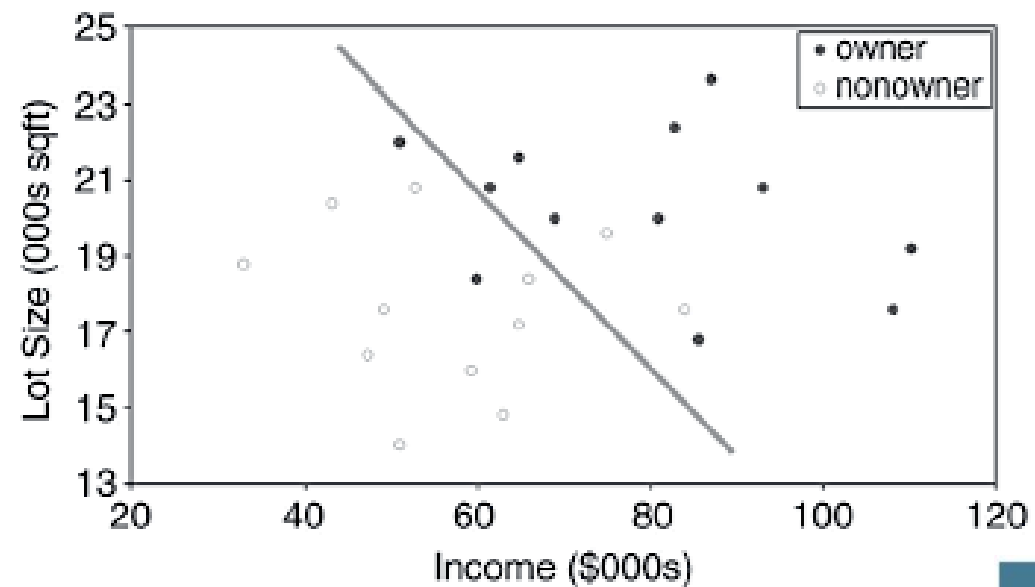
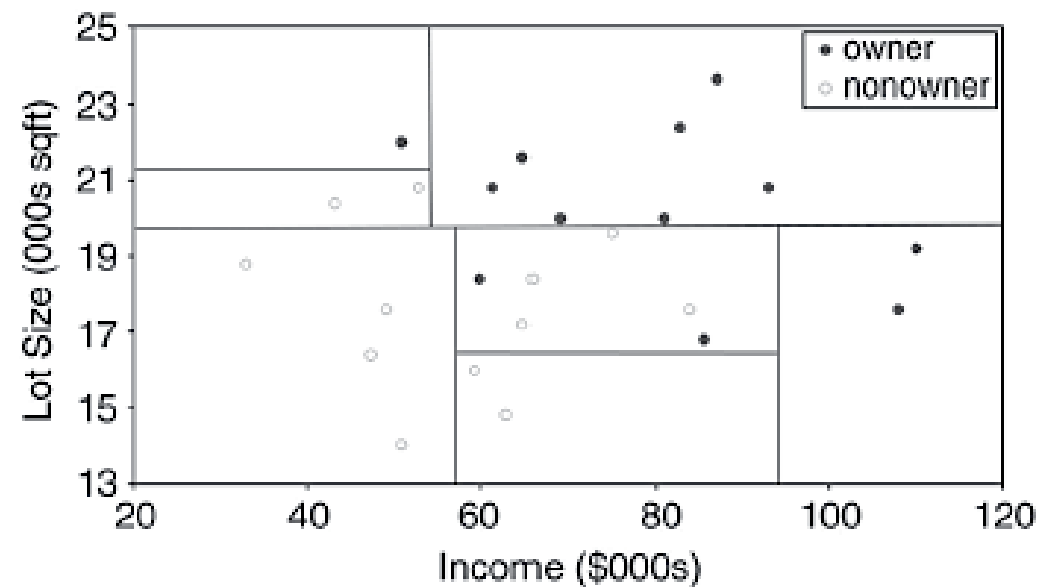
- Standard and ad hoc reporting, dashboards, alerts, queries
- Structured data, traditional sources (manageable data sets)

Common Questions

- What happened so far?
- How many did we sell?
- Where's the problem?

Why are there so many Methods?

- Each method has advantage and disadvantages. The usefulness of a method can depend on factors such as size of the dataset, the type of patterns that exist in the data, whether the data meet some **underlying assumptions** of the method and how noisy the data are, and the particular goal of the analysis.



2 cases of impactful data science – Health and Sustainability Analytics

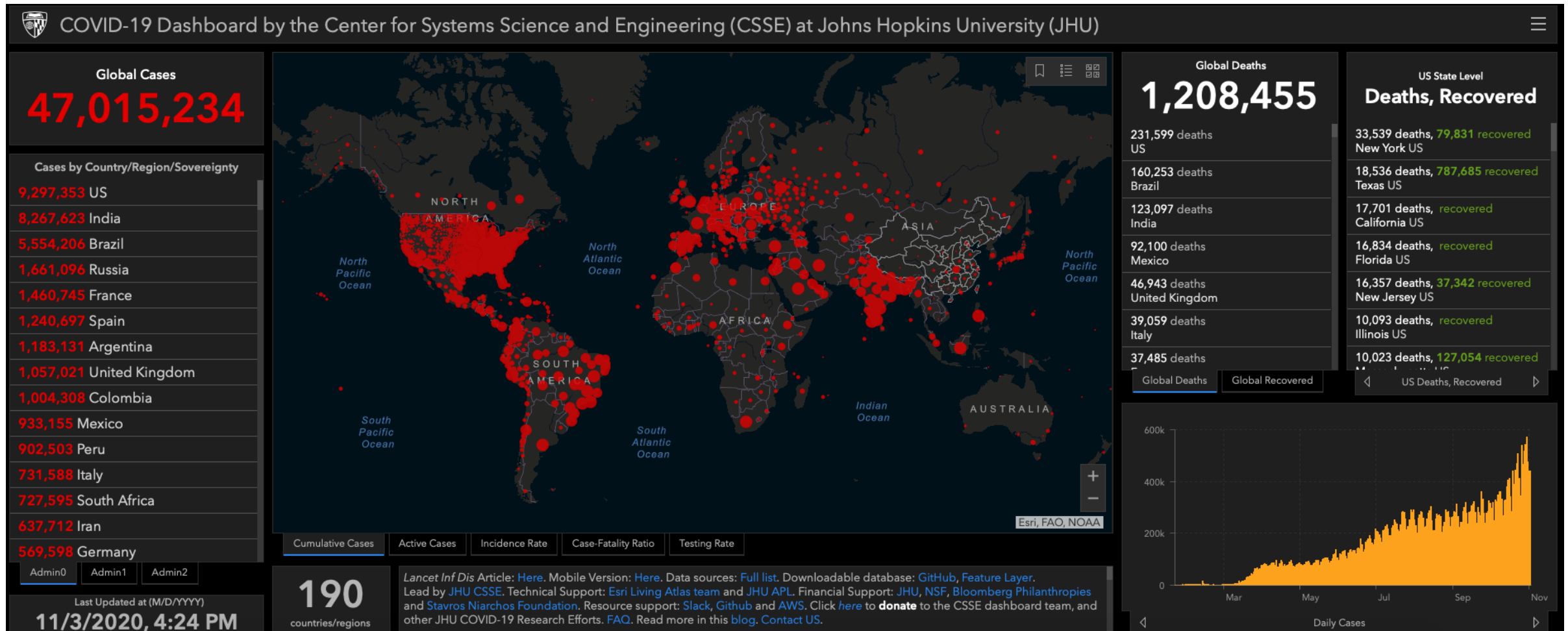
1 Mapping the spread of the Corona Virus for actionable policy recommendations

- At the onset of the Corona Virus health crisis, computational virologists and data scientists provided **actionable policy recommendations for Western countries** from data – Possibly saving countless lives
- This **world-wide effort** has resulted in cutting-edge **data-driven decision support tools** to **inform policy decisions** in real time

2 Virtual Power Plants of EVs to ensure grid stability

- Development of a **data-driven agent-based framework** to **manage EV vehicle fleet charging** in a **grid-optimal and profitable manner**
- **Battery resources** are committed in **real-time** to either the mobility or electricity services markets
- As a result the **electricity grid is stabilized** and **integration of renewable energy** carriers into the grid is facilitated

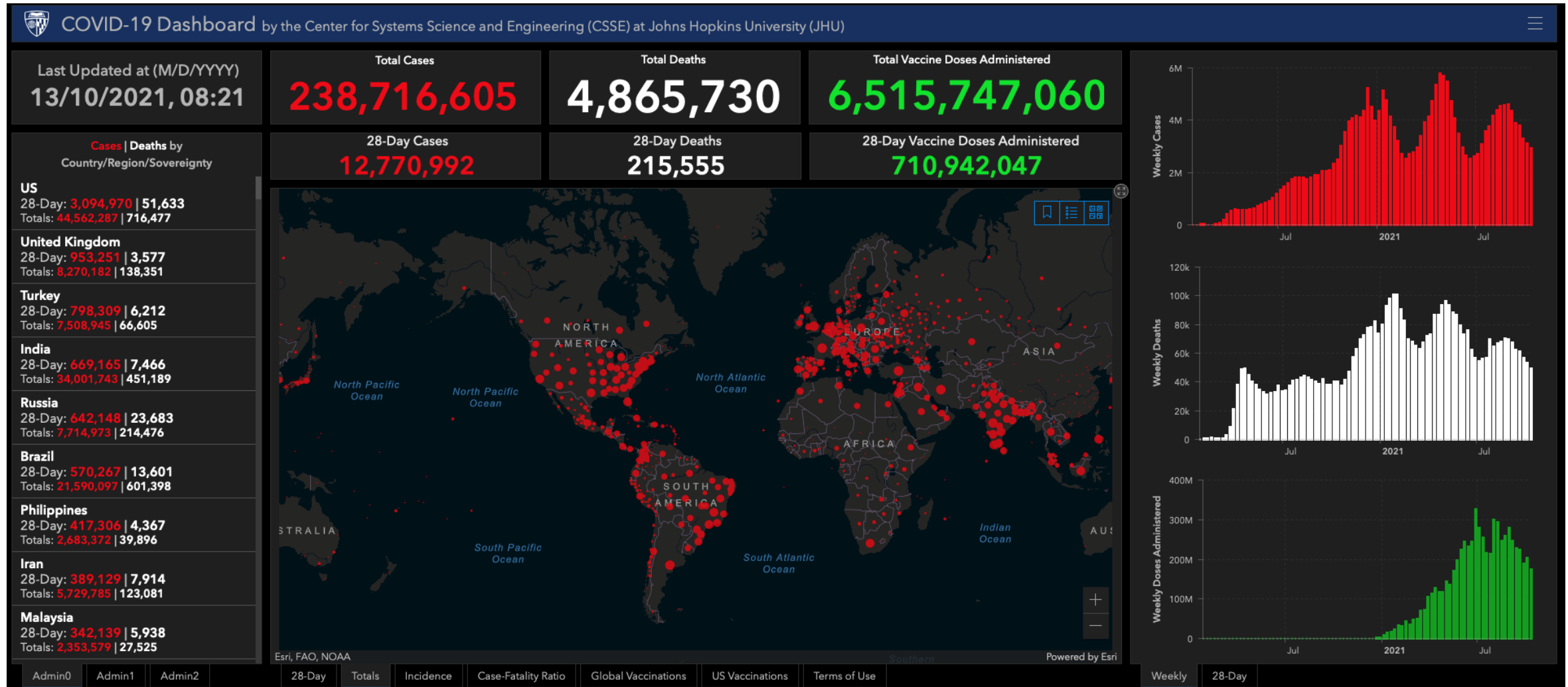
As a first step, data scientists provided close to real-time transparency on infection rates – All data available open-source for download



Source: Johns Hopkins University (<https://coronavirus.jhu.edu/map.html>)

Information Systems for Sustainable Society (is3) | WiSo Faculty | Univ.-Prof. Dr. Wolfgang Ketter | October 9th 2024

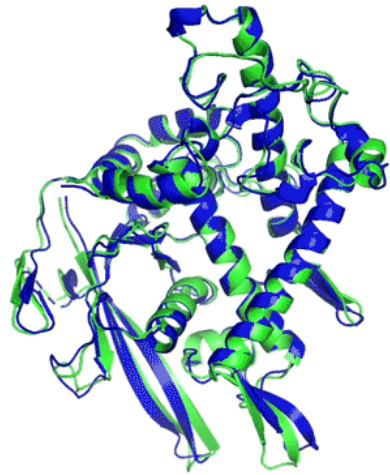
One year on the numbers look quite different but dashboards like Johns Hopkins' COVID-19 Map continue to provide decision support for the world



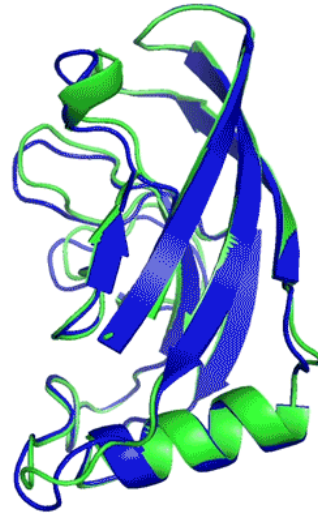
Source: Johns Hopkins University (<https://coronavirus.jhu.edu/map.html>)

Information Systems for Sustainable Society (is3) | WiSo Faculty | Univ.-Prof. Dr. Wolfgang Ketter | October 9th 2024

AlphaFold, an AI system that is recognised as a solution to "protein folding", a grand scientific challenge for more than 50 years.



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



T1049 / 6y4f
93.3 GDT
(adhesin tip)

- Experimental result
- Computational prediction

Source: <https://deepmind.com/research/case-studies/alphafold>

- AlphaFold is a groundbreaking AI tool can accurately predict 3D models of protein structures.
- AlphaFold could predict one of the coronavirus proteins with a very high accuracy (98%).
- This tool can predict the structure of millions of unknown proteins and help tackling disease and more quickly find new medicines.
- Not kidding: Yet another Nobel Prize!

<https://www.nobelprize.org/prizes/chemistry/2024/press-release/>

Data scientist took a further step: Scientists invent AI that creates COVID vaccine candidates within seconds



Source: <https://www.openaccessgovernment.org/covid-vaccine-candidates/103053/>

- Essentially, a machine-learning model can accomplish vaccine design cycles that once took months or years in a matter of seconds and minutes.
- When applied to SARS-CoV-2, the computer model quickly eliminated 95% of the compounds that could've possibly treated the pathogen and pinpointed the best options.

2 cases of impactful data science – Health and Sustainability Analytics

1 Mapping the spread of the Corona Virus for actionable policy recommendations

- At the onset of the Corona Virus health crisis, computational virologists and data scientists provided **actionable policy recommendations for Western countries** from data – Possibly saving countless lives
- This **world-wide effort** has resulted in cutting-edge **data-driven decision support tools** to **inform policy decisions** in real time

2 Virtual Power Plants of EVs to ensure grid stability

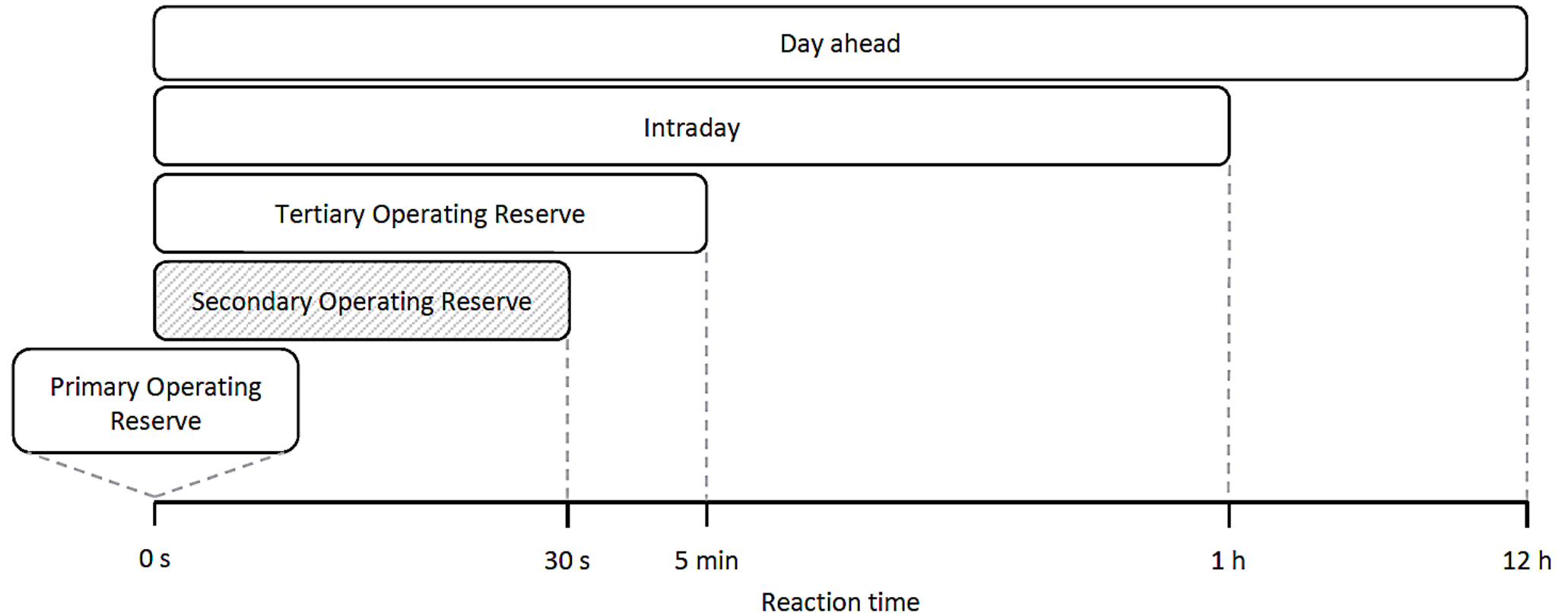
- Development of a **data-driven agent-based framework** to **manage EV vehicle fleet charging** in a **grid-optimal and profitable manner**
- **Battery resources** are committed in **real-time** to either the mobility or electricity services markets
- As a result the **electricity grid is stabilized** and **integration of renewable energy** carriers into the grid is facilitated

Electric Vehicle Virtual Power Plants for Grid Balancing

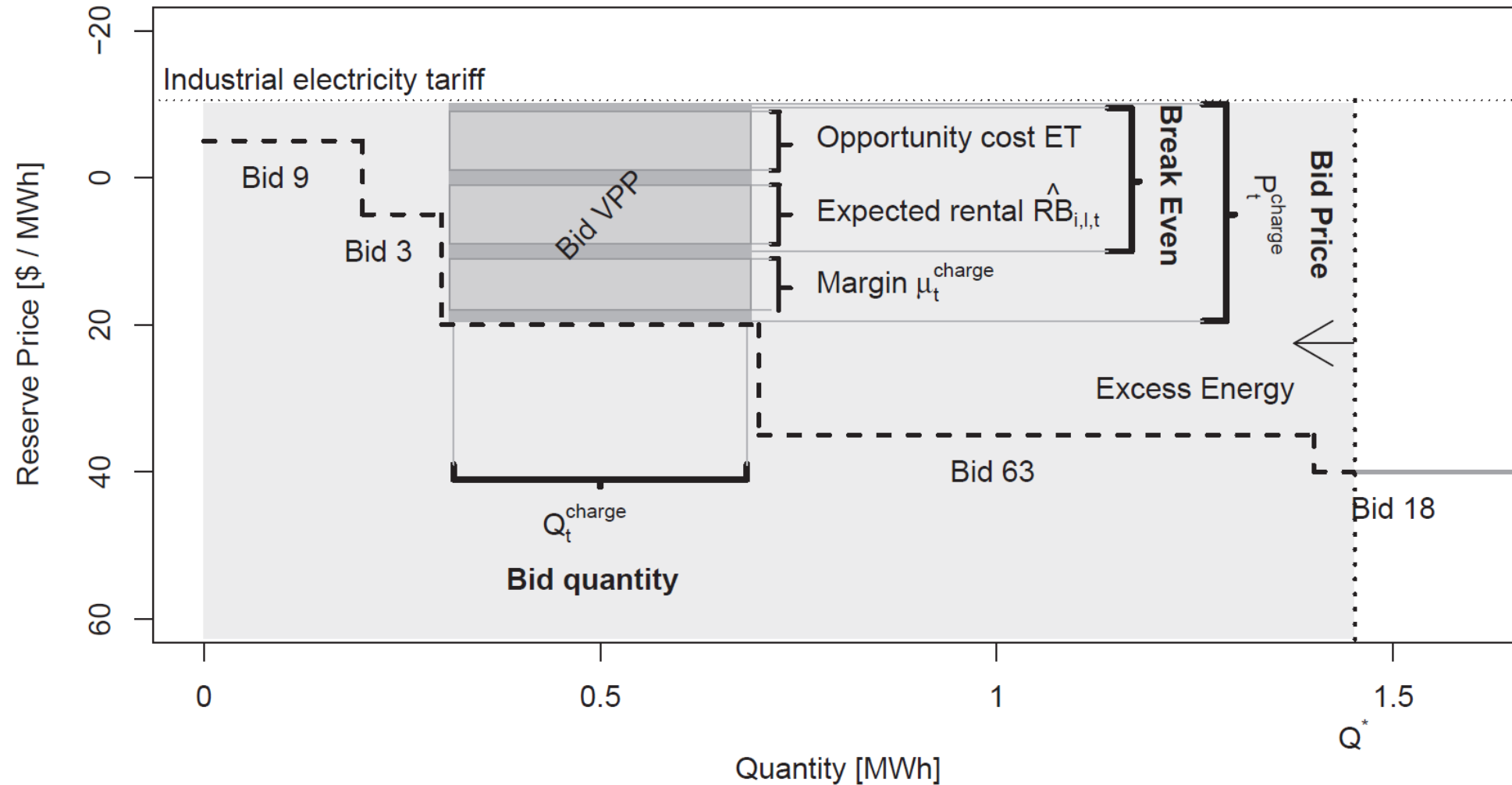
- Car Sharing costs 9 to 31ct / min
- Most of the time, cars are parked
- How to operationalize this idle time for profit?



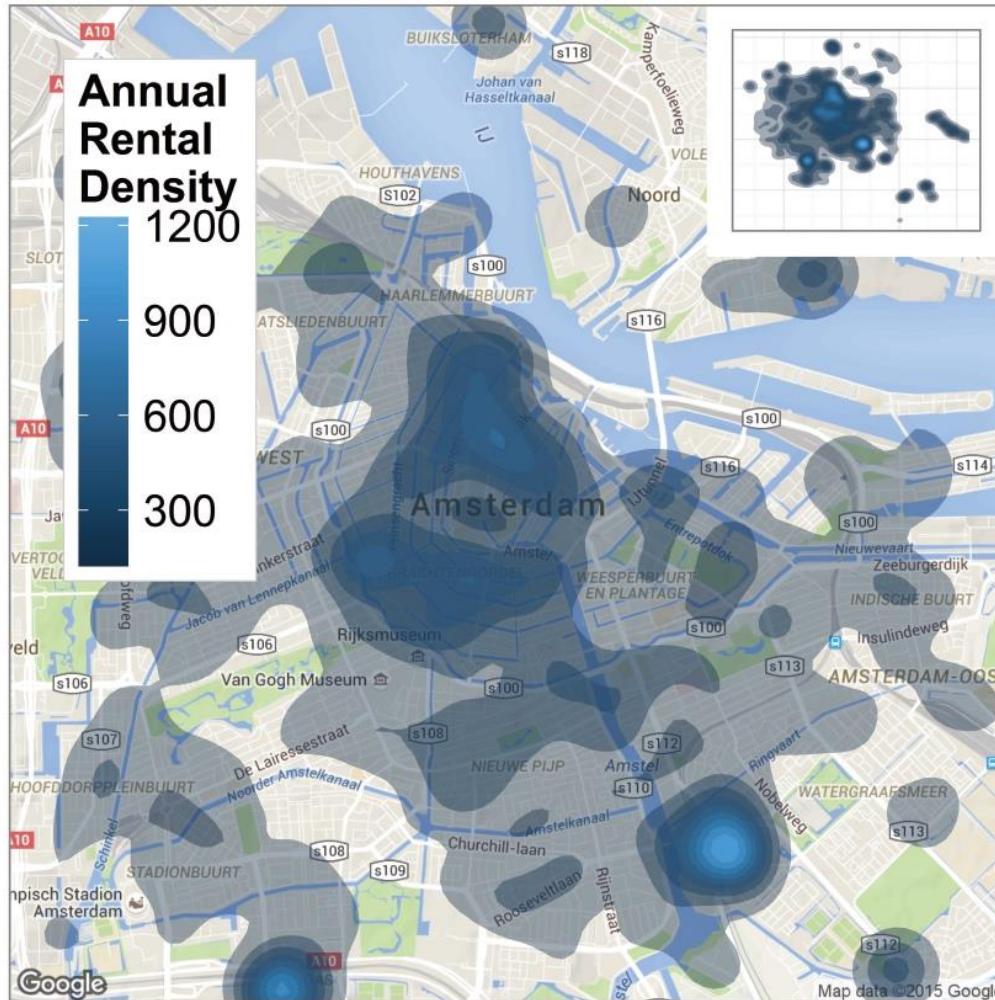
The flexibility that the car batteries offer can be marketed for



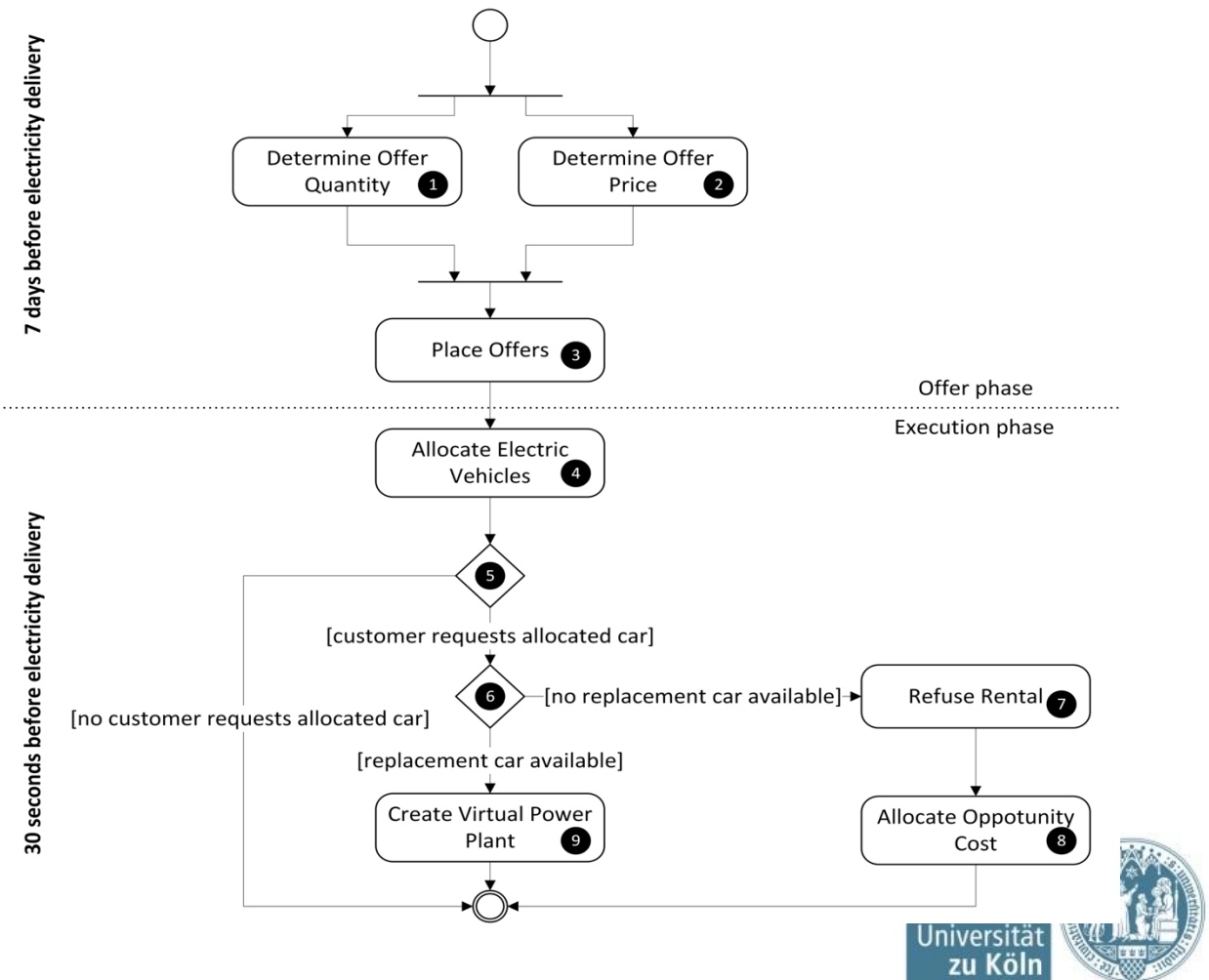
The VPP operator issues bids for upwards and downwards regulation – Example Charging



We predict rental demand and use this as input to allocate fleet vehicles to the virtual power plant



Activity Diagram Electric Vehicle Allocation Process



The VPP business model provides an additional revenue stream to the fleet operator – Overview of financial results

	Stuttgart	Amsterdam	San Diego
Battery technology 2015 (0.1 \$/kWh)	4.4% Gross profit increase	3.4% Gross profit increase	3.7% Gross profit increase
Battery technology 2020 (0.05 \$/kWh)	5.0% Gross profit increase	3.9% Gross profit increase	4.1% Gross profit increase
Battery technology 2025 (0.025 \$/kWh)	6.1% Gross profit increase	4.4% Gross profit increase	4.8% Gross profit increase

Contact



For general questions and enquiries on **research**, **teaching**, **job openings** and new **projects** refer to our website at www.is3.uni-koeln.de



For specific enquiries regarding this course, first try to utilize the ILIAS forum, and only after that contact us by sending an email to the **IS3 teaching** address at is3-teaching@wiso.uni-koeln.de

To help us process your request efficiently, use the following subject line format:

[AA] <Your request subject>

Wesentliche Informationen zur Prüfung

- » Das Modul wird mit einem **Portfolio**, bestehend aus den schriftlichen Teilen **Klausur** (**67%** der Gesamtwertung) & **Projektarbeit** (**33%** der Gesamtwertung) abgeschlossen.
Die Prüfung ist bestanden, wenn:
 - » alle Teile der Prüfung auch einzeln bestanden wurden.
- » Zu dieser Veranstaltung wird es **zwei Prüfungstermine** geben. Da das Modul im Folgesemester nicht erneut angeboten wird, wird die Anmeldung zum zweiten Termin auch im Falle eines NBs im ersten Termin möglich sein! Der Zweittermin wird noch in diesem Semester stattfinden.

Wesentliche Informationen zur Prüfung

- » Die Prüfungstermine und die Meldefristen entnehmen Sie bitte KLIPS.
- » Es ist eine Anmeldung zur Modulprüfung erforderlich, ohne Anmeldung besteht kein Anspruch auf Teilnahme an oder Bewertung der Prüfungsleistung.
- » Hier finden Sie ein Video-Tutorial zur Prüfungsan- und abmeldung:
<https://uni.koeln/87S2U>
- » Die Meldefristen wurden so gelegt, dass nach diesem Zeitpunkt klar sein muss, wer an der Prüfung teilnimmt. Nach Ablauf der Fristen ist keine An- bzw. Abmeldung mehr möglich!
- » Die Beurteilung erfolgt in der Regel innerhalb von acht Wochen nach dem Tag, an dem die Prüfung abgelegt wurde.

