



## Data Engineering for MADS

EBM213A05.2022-2023.1

# FREE EDITION\*

SUMMARY OF EVERYTHING FROM WEEK 1

LECTURES + READINGS + CHAPTERS

*Enhanced with a dynamic table of contents.*

For the full version   

# 100%

OF THE PROFIT FROM THIS  
SUMMARY IS DONATED TO  
THE FOLLOWING NGOs:



**This summary helps you study & people in need.**

Please do not share it for free, but ask your  
friends to buy it and do a good deed.

**MADS MADLAD**

**Note from MADS MADLAD:**

Thank you for checking out my free summary. When I was writing these I sometimes struggled with this program, but there were no summaries available. This is why I decided to write something that is truly complete with a lot of effort put into it.

It helped me and my friends get good grades, but I always had you in mind, the future reader. When necessary, I always went the extra mile to make my summaries, more readable, organized and complete.

If you feel like it, leave me a review of how the course is going using this summary, it will make my day to hear your feedback.

**Check out my other extensive summaries for other MADS courses:**

Statistical Learning in  
Marketing  
EBM214A05.2022-2023.1



Data Engineering for  
MADS  
EBM213A05.2022-2023.1



Companies, Brands, and  
Consumers  
EBM215A05.2022-2023.1



Retail & Omnichannel  
Marketing  
EBM880B05.2022-2023.1



Market Models  
EBM077A05.2022-2023.1



Data Science Methods for  
MADS  
EBM216A05.2022-2023.1



Digital Marketing  
Intelligence  
EBM079B05.2022-2023.1

**Contact info:**

If you need help or have an inquiry, contact me: <https://www.georgedreemer.com>

Connect with me on LinkedIn: <https://www.linkedin.com/in/georgedreemer/>

**Donations:**

By no means am I looking for fellow students to send me money! But if you feel like sending me some ETH or BTC, you can do so here:

--> **ETH:** 0x123e086c6808459e7fC6Ac7F64a77dBA1dDe0149

--> **BTC:** bc1qgwzc82vph5v8rmzef4ywechjf85772n7m2e22g

# MADS MADLAD

wishes you good luck & perseverance.



## Grades Testimony:

COURSE CODE	TITLE	SCORE	DATE	RESULT
EBS001A10	Business Research Methods for Pre-MSc	8	21-12-2021	8
FBS002A05	Mathematics for Pre-MSc	9	10-11-2021	9
EBS003A05	Organization Theory & Design for Pre-MSc	7	05-11-2021	7
EBB098A05	Contemporary Theories on Business and Management	6	11-05-2022	6
EBB649C05	Strategic Management B&M	8	15-06-2022	8
EBB617B05	Human Resource Management B&M	8	08-04-2022	8
EBB104A05	Behavioural Decision Making	7	03-11-2021	7
EBB085A05	Marketing Research for F&BE	8	04-04-2022	8
EBS008B10	Research Paper for Pre-MSc Marketing	7	05-07-2022	7
EBM043A05	Business Ethics	8	14-11-2022	8
EBB105B05	Digital Marketing Analytics	8	21-01-2022	8
EBM213A05	Data Engineering for MADS	7	01-11-2022	7
EBM214A05	Statistical Learning in Marketing	8	02-11-2022	8
EBM215A05	Companies, Brands, and Consumers	8	05-11-2022	8
EBM216A05	Data Science Methods for MADS	9	20-01-2023 <span style="color: green; font-weight: bold;">new</span>	9

## Table of Contents

Week 1 (Lectures+Readings).....	14
Lecture 0 (Intro) & 1.1 (MD->MQ->RQ) & HBR article.....	14
Management Dilemma and Questions, Research Questions .....	14
Management Dilemma.....	14
From Management Dilemma to Management Questions: .....	15
Management Question defined: .....	15
From Management Question to Research Question .....	15
Management Question vs. Research Question:.....	16
From Research Question to Analysis Questions: .....	16
7 steps of the opportunity tree .....	16
Lecture 1.2 - supplement (Wehkamp lecture 2022).....	19
7 Elements of a Data Strategy.....	19
Data Warehouse vs. Data Lake .....	19
4 Ways to work with Data.....	21
Data Science – mix of 3 fields of knowledge .....	21
Domain Expertise (Business) .....	22
Technical Data Engineering .....	22
Math & Statistical Knowledge .....	22
Reading – HBR Article (optional).....	24
4 Steps to Management Questions.....	24
Reading – Book: Verhoef et al. (11.6 required +11.1 optional) .....	25
7 Steps of Opportunity finding.....	25
Reading – Book: Business Research Methods (2.1 optional + 2.2 required)..	28
Week 2 (Readings+Lecture) .....	32
Reading – Book: Verhoef et al.(Ch.1, 2.1-2.5, 3.4, 3.9, 4, 5) .....	32
Chapter 1 (required): “Data science and big data” .....	32
Chapter 2.1-2.5 (required): “Creating value with data science” .....	35
2.1 INTRODUCTION .....	35
2.2 DATA SCIENCE VALUE CREATION MODEL.....	35

2.3	VALUE CREATION OBJECTIVES.....	36
2.3.1	Balance between V2F and V2C: .....	36
2.3.2	V2S – Value to Society .....	37
2.3.3	Metrics for V2F and V2C.....	37
2.4	DATA ASSETS.....	38
2.5	DATA ANALYTICS.....	38
2.5.1	The power of visualization and storytelling .....	39
Chapter 3.4 & 3.9 (required): “CUSTOMER METRICS” .....		39
3.4	CUSTOMER ( <i>FEEDBACK</i> ) METRICS .....	39
	First Dimension: Time span - Forward- vs. Backward-looking Metrics..	40
	Second Dimension: Measurement scale.....	41
3.4.1	Is there a silver metric? .....	41
3.4.2	Other theoretical relationship metrics .....	42
3.4.3	Customer equity drivers .....	42
3.4.4	Internal data sources.....	43
3.4.5	Online sources – Customer Reviews .....	43
3.9	CUSTOMER METRICS .....	44
3.9.1	Customer acquisition metrics .....	45
3.9.2	Customer development metrics .....	45
3.9.3	Customer value metrics.....	46
3.9.4	Customer equity.....	48
3.9.5	New big data metrics.....	49
Chapter 4 (required): “Data Assets” .....		51
4.2	DATA SOURCES AND THE DIFFERENT TYPES OF DATA.....	51
4.2.1	External data sources vs. Internal data sources .....	51
4.2.2	Structured vs. Unstructured data .....	52
4.2.3	Market data.....	53
4.2.5	Brand data.....	54
4.2.7	Customer data.....	56
4.3	USING THE DIFFERENT DATA SOURCES IN THE ERA OF BIG DATA .....	57

4.4 DATA QUALITY AND DATA CLEANSING.....	58
4.4.1 Data Quality .....	58
4.4.2 Data Cleansing.....	59
4.4.3 Missing value and data fusion .....	60
Chapter 5 (required): “Data storing and integration” .....	60
5.1 INTRODUCTION .....	60
5.2 STORING AND INTEGRATING DATA SOURCES IN DATA WAREHOUSES.60	
5.2.1 Storing data in the data warehouse.....	61
5.2.2 The data model in a data warehouse .....	62
5.2.3 Data integration into the data warehouse.....	64
5.2.3.1 Extraction .....	65
5.2.3.2 Transformation.....	65
5.2.3.3 Loading.....	65
5.3 STORING AND INTEGRATING DATA SOURCES IN DATA LAKES.....	66
5.4 CHALLENGES OF DATA INTEGRATION IN THE ERA OF BIG DATA .....	68
5.4.1 The technical challenges of integrated data .....	68
5.4.1.1 Integration at the individual level.....	69
5.4.1.2 Integration at the intermediate level.....	69
5.4.1.3 Integration at the time level.....	69
5.4.2 The analytical challenges of integrated data .....	70
5.4.3 The business challenges of integrated data .....	70
5.4.3.1 Dealing with different data types .....	70
5.4.3.2 Declared data: customer descriptors.....	71
5.4.3.3 Appended data .....	71
5.4.3.4 Overlaid data .....	71
5.4.3.5 Implied data .....	72
Lecture 2 – From ‘raw’ data to your own table: conceptual introduction .....	73
Our first choice: level of analysis (aggregation level) .....	73
Four Types of Information .....	73
Joining tables, aggregation, adding, external data .....	74

INNER_JOIN.....	74
LEFT_JOIN /RIGHT_JOIN.....	74
FULL_OUTER_JOIN.....	75
Week 3 (Lecture) .....	79
Lecture – Working with SQL, Azure Data, Creating our own table: From ‘raw’ data to your own table with SQL.....	79
Every SQL query to retrieve data consists of maximum 6 SQL instructions	79
Week 4 (Lecture) .....	80
Lecture – Combining Wehkamp data with external data in RStudio .....	80
Two dimensions: Data source & Data type .....	80
Sources for external data.....	80
Integrating external data sources .....	82
Useful commands in R Studio .....	82
Week 5 (Readings+Lecture) .....	83
Reading – Wickham (2014) “Tidy Data” .....	83
Section 2: Three characteristics of a Tidy dataset .....	83
Section 2.1: Data structure.....	83
Section 2.2: Data Semantics .....	84
Section 2.3: Tidy data .....	84
Section 3: Operations to make a messy dataset tidy.....	85
Section 3.1: Column headers are values, not variable names.....	85
3.2: Multiple variables are stored in one column .....	87
3.3: Variables are stored in both rows and columns.....	88
3.4: Multiple types of observational units are stored in the same table .....	89
3.5: A single observational unit is stored in multiple tables.....	89
Section 4: Tidy tools (tools that input & output tidy data) .....	90
Section 4.1: Data Manipulation .....	90
Section 4.2: Visualization .....	90
Section 4.3: Modeling .....	91

Reading – de Jonge, E. and van der Loo, M. (2013) "An introduction to data cleaning with R" .....	92
1 Introduction .....	92
1.1: Statistical analysis in 5 steps.....	92
1.2 Some general background in R .....	93
1.2.1 Variable types and indexing techniques .....	93
1.2.2 Special values .....	93
2 From raw data to technically correct data .....	93
2.1: Technically correct data in R.....	93
2.2: Reading text data into a R data.frame .....	94
2.2.1 read.table() and its cousins.....	94
2.3: Type conversion .....	98
2.3.1 Introduction to R's typing system .....	98
2.3.2 Recoding factors.....	98
2.3.3 Converting dates .....	99
2.4: character manipulation .....	100
2.4.1 String normalization .....	101
2.4.2 Approximate string matching .....	102
2.5 Character encoding issues .....	105
3: From technically correct data to consistent data .....	105
3.1 Detection and localization of errors .....	106
3.1.1 Missing values .....	106
3.1.2 Special values .....	107
3.1.3 Outliers.....	108
3.1.4 Obvious inconsistencies.....	109
3.2 Correction .....	111
3.2.1 Simple transformation rules .....	112
3.2.2 Deductive correction .....	114
3.2.3 Deterministic imputation.....	115
3.3 Imputation .....	116



3.3.1 Basic numeric imputation models .....	117
3.3.2 Hot deck imputation.....	117
3.3.3 kNN-imputation.....	118
3.3.4 Minimal value adjustment.....	119
Reading – Donders, A.G.T, van der Heijden, G.J.M.G, Stijnen, T and Moons, K.G.M. (2006) "Review: A gentle introduction to imputation of missing values" .....	120
Reading – Schafer, J.L. and Graham, J.W. (2002), "Missing Data: Our View of the State of the Art" .....	120
Lecture 5 – Data cleaning, outliers, missing data (1/2) .....	121
Data Quality.....	121
Data Cleaning.....	121
Steps in Cleansing (lecture) .....	122
Parsing (1) .....	122
Correcting (2) .....	123
Standardizing (3) .....	123
Matching (4).....	124
Consolidating (5) .....	124
Formatting the data (Tidy Data).....	125
Packages for Tidying Data .....	125
After Data Cleaning: Sanity Check.....	126
Outliers.....	126
What are outliers? .....	126
Why do we address outliers? .....	126
Detecting Outliers .....	127
Treating Outliers .....	128
Missing Data .....	129
MCAR, MAR, MNAR – Graphical representation (lecture) .....	130
MCAR, MAR, MNAR: Academic & Simple explanation (ext. source) ...	131

Lecture 5 – Detecting outliers/Mahalanobis distance, Multiple imputation (2/2).....	132
Detecting outliers/Mahalanobis Distance .....	132
How to do it in R: .....	135
Imputation.....	137
How to do it in R.....	138
Week 6 (Readings+Lecture) .....	142
Reading – Book: Chapter 7 and 8 .....	142
Chapter 7: Data Analytics.....	142
7.2 THE POWER OF ANALYTICS .....	142
7.3 STRATEGIES FOR ANALYZING DATA .....	142
7.3.1 PROBLEM SOLVING .....	143
7.3.2 DATA EXPLOITATION .....	143
7.3.3 DATA MINING.....	144
7.3.4 COLLATERAL CATCH.....	145
7.4 TYPES OF DATA ANALYTICS .....	146
7.4.1 DESCRIPTIVE ANALYTICS.....	147
7.4.2 DIAGNOSTIC ANALYTICS .....	147
7.4.3 PREDICTIVE ANALYTICS.....	148
7.4.4 PRESCRIPTIVE ANALYTICS .....	148
7.5 HOW BIG DATA AND AI CHANGE ANALYTICS .....	149
7.5.1.1 DATA SCIENCE .....	149
7.5.1.2 AI .....	150
7.5.1.3 MACHINE LEARNING (ML) .....	150
7.5.1.4 DEEP LEARNING (DL) .....	151
7.6 ANALYTICAL METHODS AND TECHNIQUES.....	151
Chapter 8: Data Exploration.....	152
8.1 INTRODUCTION.....	152
8.2 DESCRIPTIVE ANALYSES – REPORTING .....	153

8.3 DESCRIPTIVE ANALYSES – INVESTIGATING ONE-TO-ONE RELATIONSHIPS ..... 153

    8.3.1 KPI CATEGORICAL, DRIVER CATEGORICAL ..... 154

    8.3.2 KPI NUMERICAL, DRIVER CATEGORICAL ..... 155

    8.3.3 KPI CATEGORICAL, DRIVER NUMERICAL ..... 156

    8.3.4 KPI NUMERICAL, DRIVER NUMERICAL ..... 158

8.4 SPECIAL CASES OF ONE-TO-ONE EXPLORATORY ANALYSES..... 159

    8.4.1 PROFILING AND CUSTOMER CROSSINGS ..... 159

    8.4.2 DECILE ANALYSIS ..... 159

    8.4.3 EXTERNAL PROFILING ..... 160

    8.4.4 ZIP CODE ANALYSIS..... 161

8.5 DYNAMIC ANALYSES ..... 161

    8.5.1 TREND ANALYSIS ..... 162

    8.5.2 MIGRATION ANALYSIS ..... 163

    8.5.3 LIKE-4-LIKE ANALYSIS..... 164

8.6 IDENTIFYING STRUCTURE IN DATA: UNSUPERVISED LEARNING ..... 165

    8.6.1 CLUSTER ANALYSIS ..... 165

        8.6.1.1 EXECUTION ..... 165

        8.6.1.2 SELECTION OF CLUSTER VARIABLES ..... 165

        8.6.1.3 DATA PREPARATION ..... 166

        8.6.1.4 RUNNING THE ANALYSIS..... 166

        8.6.1.5 SELECTION OF NUMBER OF CLUSTERS..... 166

        8.6.1.6 PROFILING THE CLUSTERS..... 166

    8.6.2 PRINCIPAL COMPONENT ANALYSIS (PCA)..... 167

Lecture 6 – The Power of First Insights ..... 169

    DESCRIPTIVE ANALYSES: REPORTING ..... 169

        Typical Questions of Descriptive Analyses..... 169

        Using min, max, mean, median, totals is very useful ..... 169

        Useful graphics: histograms & boxplots ..... 170

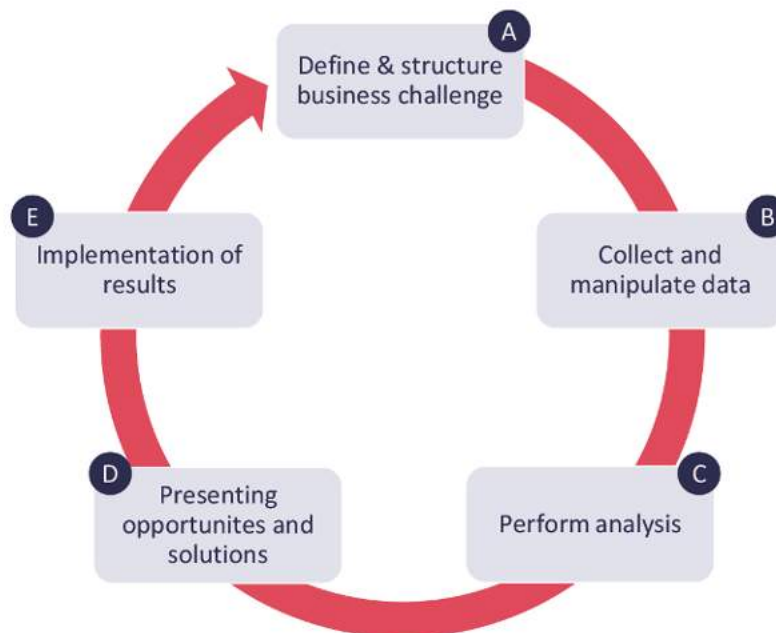
    DESCRIPTIVE ANALYSES: 1-TO-1 RELATIONSHIPS ..... 170

Chi-Square Test: Categorical KPI x Categorical drivers.....	170
Overview of tests: .....	172
Decision Tree for Choosing a Test: ANOVA vs. t-test .....	172
Numerical KPI x Categorical levels.....	173
DYNAMIC ANALYTICS.....	174
Trend Analysis .....	174
Migration Analysis.....	174
Like-4-Like Analysis .....	175
Week 7 (Readings+Lecture) .....	176
Reading – Book: Chapter 10.....	176
Chapter 10: Creating Impact with Storytelling and Visualization .....	176
10.1 INTRODUCTION.....	176
10.2 FAILURE FACTORS FOR CREATING IMPACT .....	177
10.3 STORYTELLING .....	177
10.3.1 CHECKLIST FOR A CLEAR STORYLINE .....	179
10.4 VISUALIZATION .....	179
10.4.1 CHOOSING THE CHART TYPE .....	179
10.4.1.1 SHOWING RELATIONSHIP BETWEEN DATA POINTS .....	179
10.4.1.2. COMPARING DATA POINTS.....	180
10.4.1.3 COMPOSITION .....	181
10.4.1.4 DISTRIBUTION.....	181
DECISION PROCESS FOR CHARTS (ABELA, 2008) .....	182
10.4 MISLEADING GRAPHS.....	183
10.4.3.1 TRUNCATED GRAPHS.....	183
10.4.3.2 ADJUSTED AXIS .....	183
10.4.3.3 INCORRECT SCALING .....	184
10.4.3.4 LOGARITHMIC SCALING .....	184
10.4.3.5 OMITTING DATA .....	185
10.4.3.6 SIMULATED TRENDS .....	185
10.4.3.7 REDUNDANT 3D PERSPECTIVE .....	186

10.5 TRENDS IN VISUALIZATION .....	186
10.6 CONCLUSIONS.....	186
Reading – Berinato, S (2016).....	187
Conceptual or Data Driven.....	187
Declarative or Exploratory .....	187
The 4 Types of Visual Communication .....	187
Idea Illustration .....	188
Idea Generation .....	189
Visual Discovery .....	190
Everyday Datawiz .....	191
Reading – Cleveland et al. (1984).....	192
Reading – Swamy, P.R. (2013): Building Logic Into Communication Using the Minto Pyramid Principle .....	193
Storing and Retrieving Information.....	193
Barbara Minto’s Pyramid Principle .....	194
Structure of the Pyramid Principle .....	195
Introductory Statement.....	195
Advantages of using the Minto pyramid principle .....	197
Quizzes (1, 5 & 6) .....	198
Quiz 1.....	198
Quiz 5.....	200
Quiz 6.....	202

Week 1 (Lectures+Readings)

Lecture 0 (Intro) & 1.1 (MD->MQ->RQ) & HBR article



**FIGURE 12.4** Phases of the analytical cycle

The 5 Phases of the Analytical Cycle entail to:

- (1) define and structure the business challenge
- (2) collect and manipulate the data
- (3) perform analysis
- (4) present opportunities and solutions
- (5) implementation of results

### **Management Dilemma and Questions, Research Questions**

**Management Dilemma:** a symptom of an underlying problem.

- Profits are decreasing
- Target of improving shareholder value by 5% should be met
- Marketing efforts lack effectiveness
- Increasing handle time at the Customer Service Desk
- Underperformance of several sales force teams
- Unanticipated sudden increases in demand

***From Management Dilemma to Management Questions:***

- Discussion with relevant stakeholders
- Interviews with industry experts
- Secondary data analysis

**Goal:** restate the MD in terms of underlying problem.

Usually starting with:

- Should we...? (choice of purpose)
- How can we...?
- Why do we...?

***Management Question defined:***

- Management dilemma restated in question form
- Defined in terms of the underlying problem
- Preferably linked to a Key Performance Indicator (KPI)
- Does not specify the research that needs to be done
- Questions are still broad

**Example questions:**

- What should be done to increase conversion?
- Should we use new and promising advertising channels to improve marketing ROI?
- Why is the number of new contracts closed by several sales force teams lower than expected?

***From Management Question to Research Question (example):***

Let's assume the Management Question is:

- How can we increase sales in the Northern region this year?

Possible Research Questions:

- What causes decrease in the sales of the Northern region?
- What is the effect of the company-wide pricing strategy on the Northern region?
- **Management Question defined:**

**Management Question vs. Research Question:**

Management question	Research question
Asks what the decision maker needs to do	Asks what research should be conducted
Action oriented	Information oriented

**From Research Question to Analysis Questions:**

Using the 5W's of the opportunity finding-method helps you to come up with a good set of analysis questions.

- Who?
- What?
- Where?
- When?
- Why?

**7 steps of the opportunity tree**

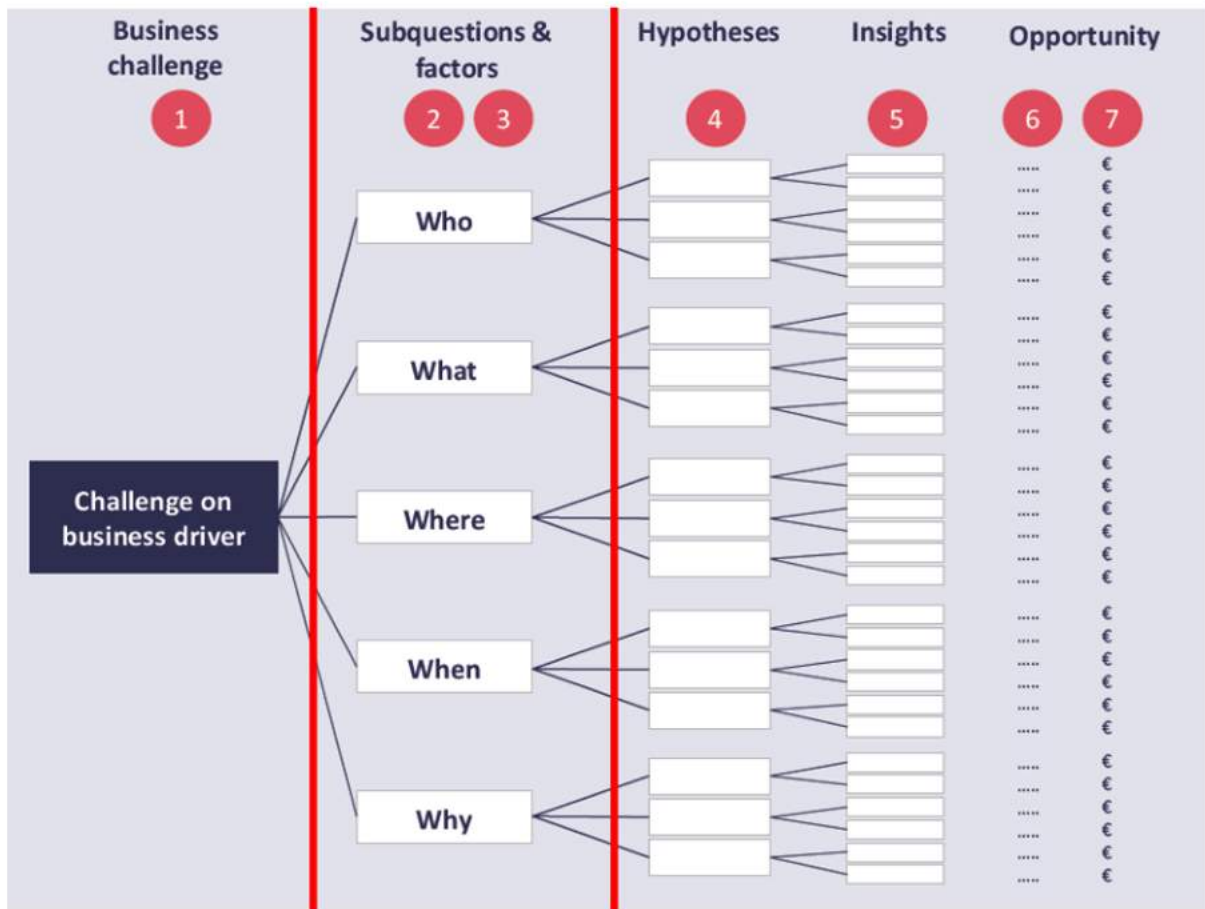
1. **Business challenge:** the starting point of the tree, defined in measurable objectives.
2. **Sub-questions:** translate business challenge into sub-questions.
3. **Factors:** define which levers you can influence or use.
4. **Hypotheses:** make a 'braindump' of all possible hypotheses.

----- **exhaustive opportunity tree includes these** -----

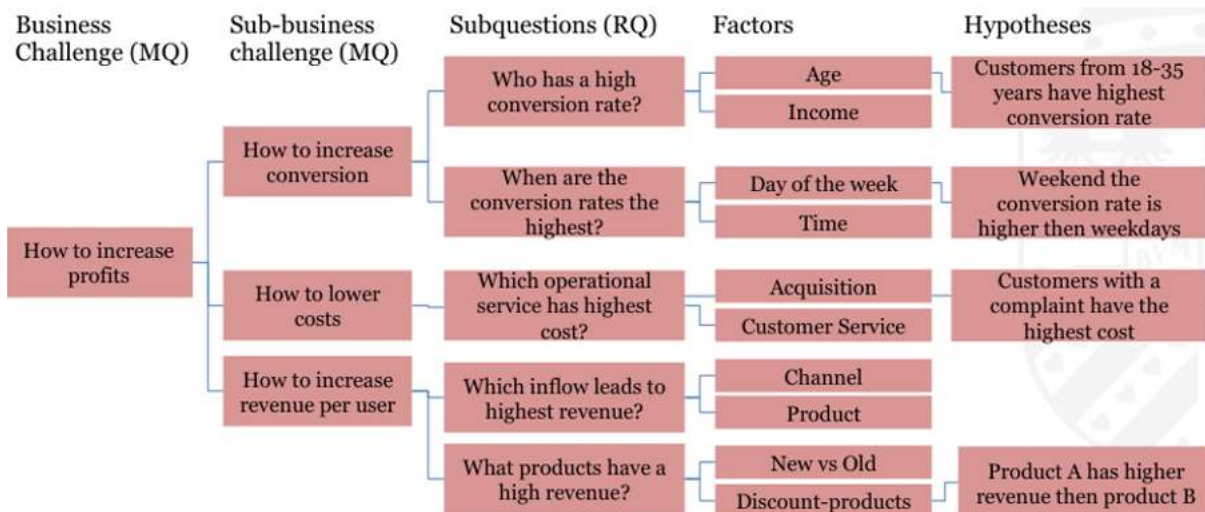
5. **Insights:** determine the analyses questions to check the hypotheses and to identify areas with high potential.
6. **Initiatives:** come up with potential initiatives to realize the targets/objectives.
7. **Impact:** calculate the monetary impact (+ or -) of initiatives and identify the most promising ones.



**Exhaustive opportunity tree:**

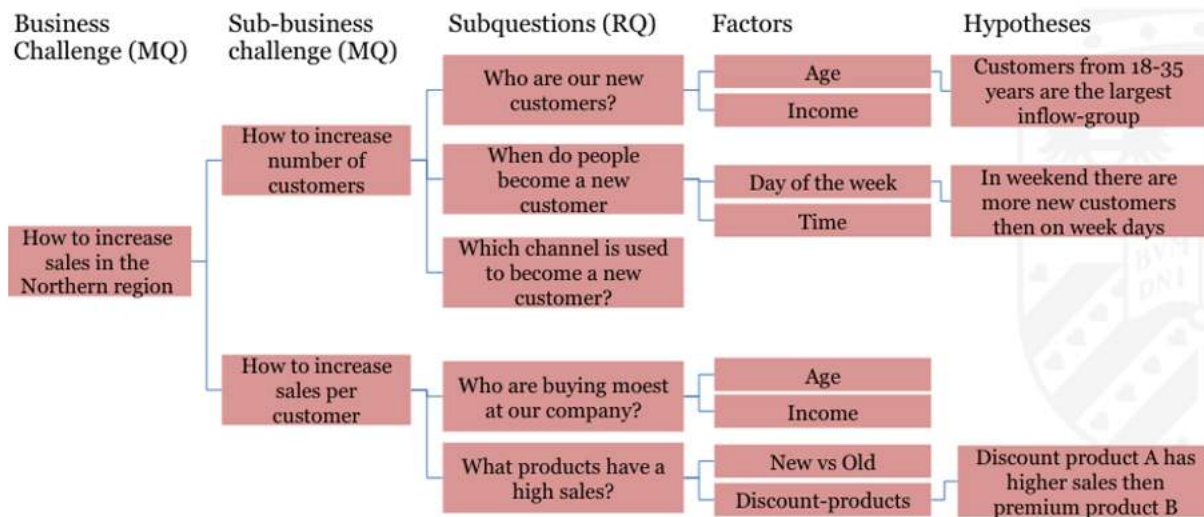


**Example of non-exhaustive opportunity tree:**



## Tips on how to fill in the 5 W's, factors and hypothesis:

- Based on hypothesis of the stakeholder
- Based on literature
- Based on steps in a customer journey
- Imagine you're the customer



## Next step – write a research proposal:

Every proposal includes two basic sections – statement of the research question (1) and brief description of the research methodology (2). (Blumberg et al., 2014, Section 2.4) In other words, **what** are we researching and **how** are we going to research.

## Lecture 1.2 - supplement (Wehkamp lecture 2022)

**7 Elements of a Data Strategy:**

1. Business Requirements
2. Sourcing and Gathering Data
3. Technology Infrastructure Requirements
4. Turning Data into Insights
5. People and Process
6. Data Governance
7. The Roadmap

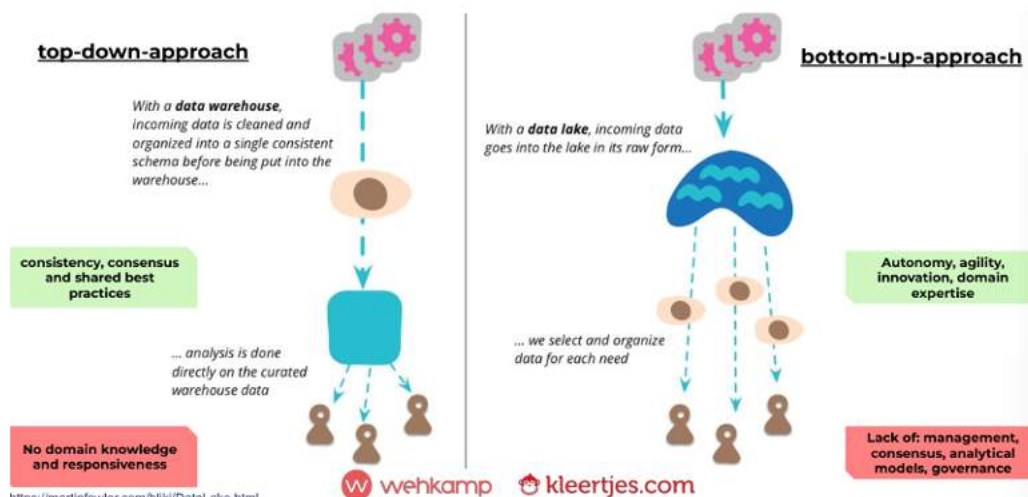
**Data Warehouse vs. Data Lake**

1. **Data Warehouse:** Incoming data is cleaned and organized into a single consistent schema before being stored. Analysis is done directly on the curated data.

- **Top-Down Approach**
- **Pros:** Consistency, consensus & shared best practices
- **Cons:** No domain knowledge, no responsiveness

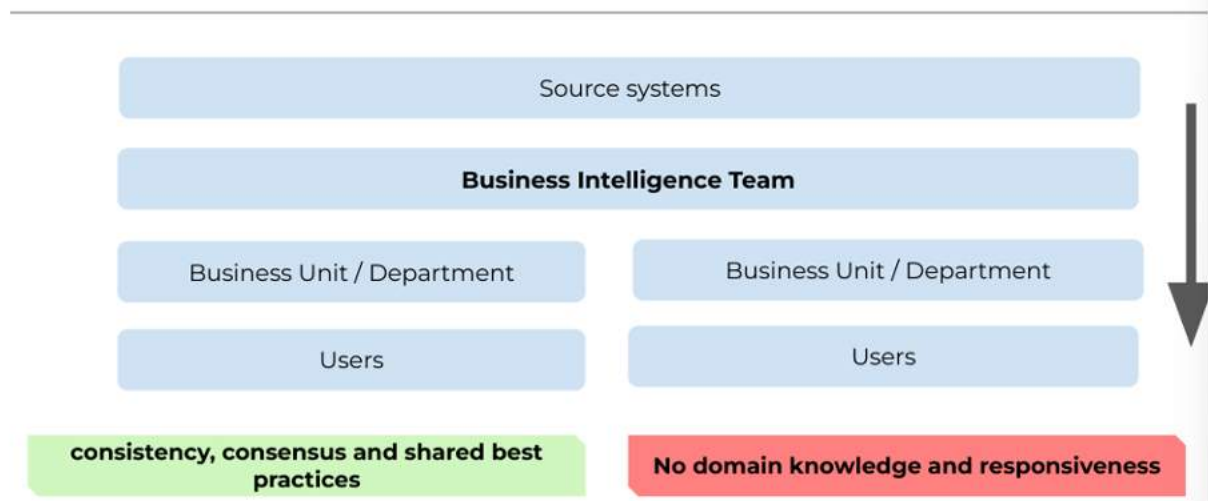
2. **Data Lake:** Incoming data goes into the lake in its raw form. We select and organize data for each need.

- **Bottom-Up Approach**
- **Pros:** Autonomy, agility, innovation, domain expertise
- **Cons:** Lack of management, consensus, governance

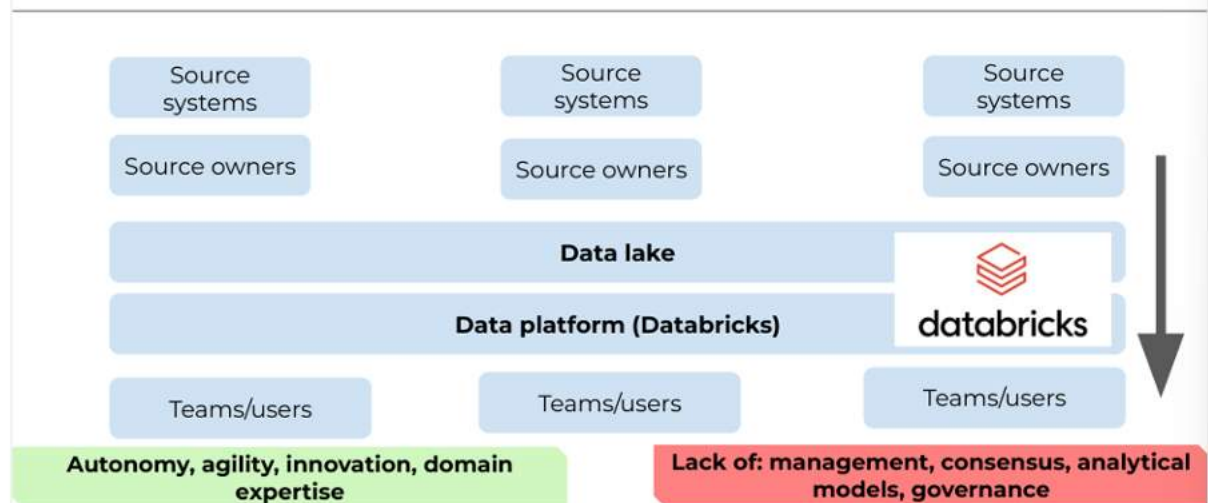


**Top-down vs. Bottom-up:**

top-down approach



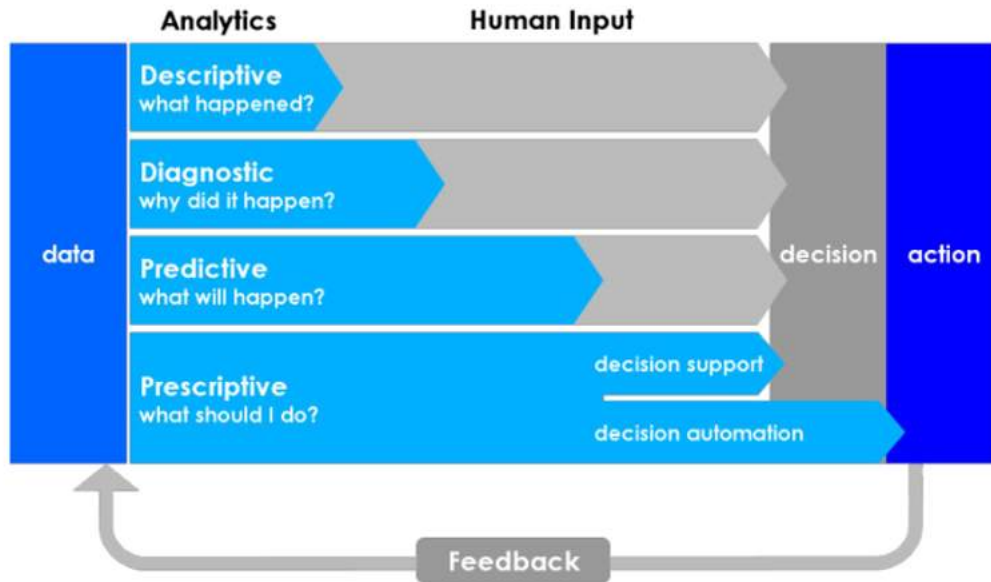
bottom-up approach



Wehkamp is now been experimenting with a hybrid approach, which entails both governed and non-governed ways of work.

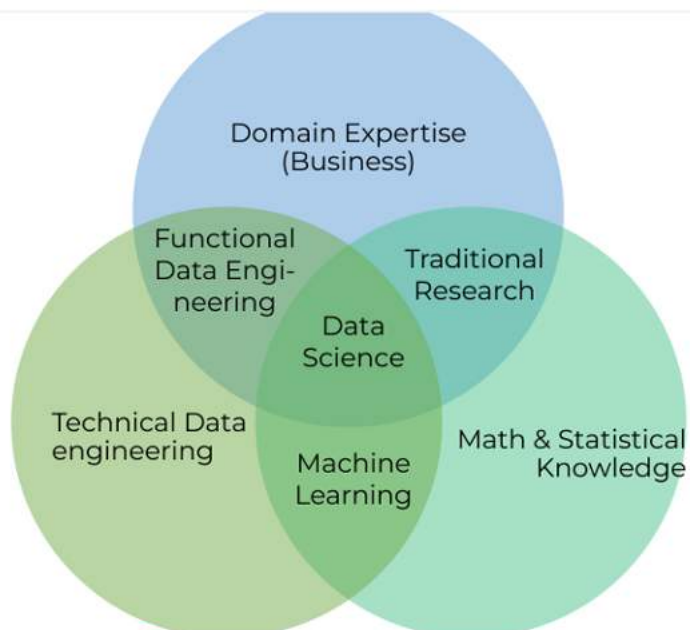
### 4 Ways to work with Data

- **Descriptive:** What happened?
- **Diagnostic:** Why did it happen?
- **Predictive:** What will happen?
- **Prescriptive:** What should be done?



### Data Science – mix of 3 fields of knowledge

1. Technical Data Engineering
2. Domain Expertise (Business)
3. Math & Statistical Knowledge

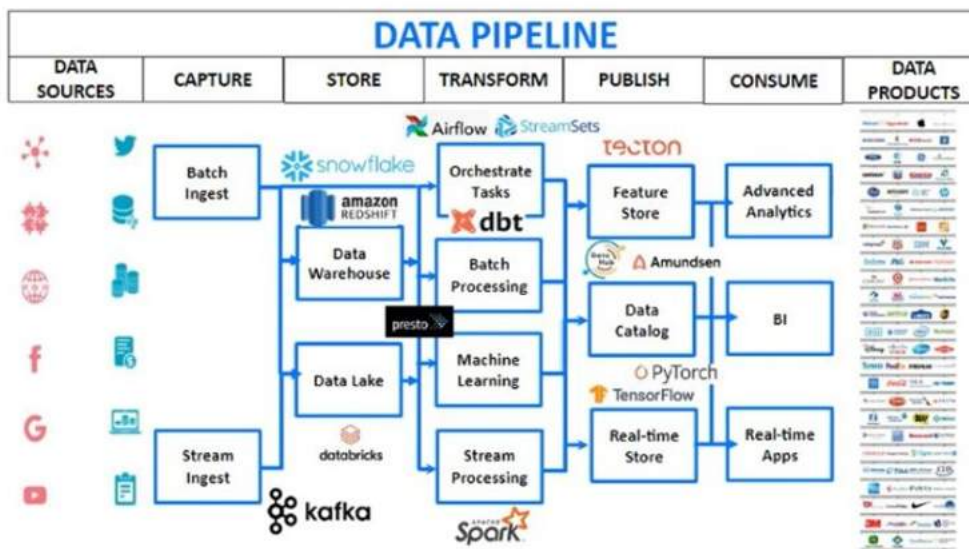


**Domain Expertise (Business)**

- Understanding the customer & process
- Make it actionable
- From business intelligence to visual analytics
- Tableau (visualisations) enable the business to use data

**Technical Data Engineering**

- **Data engineers:** develop constructs (1), test and maintain data architectures (2), such as databases and large-scale processing systems.
- Build functional data models and products
- Build ETL (Extract, Transform, Load) processes
- Responsible for definitions and structures of data

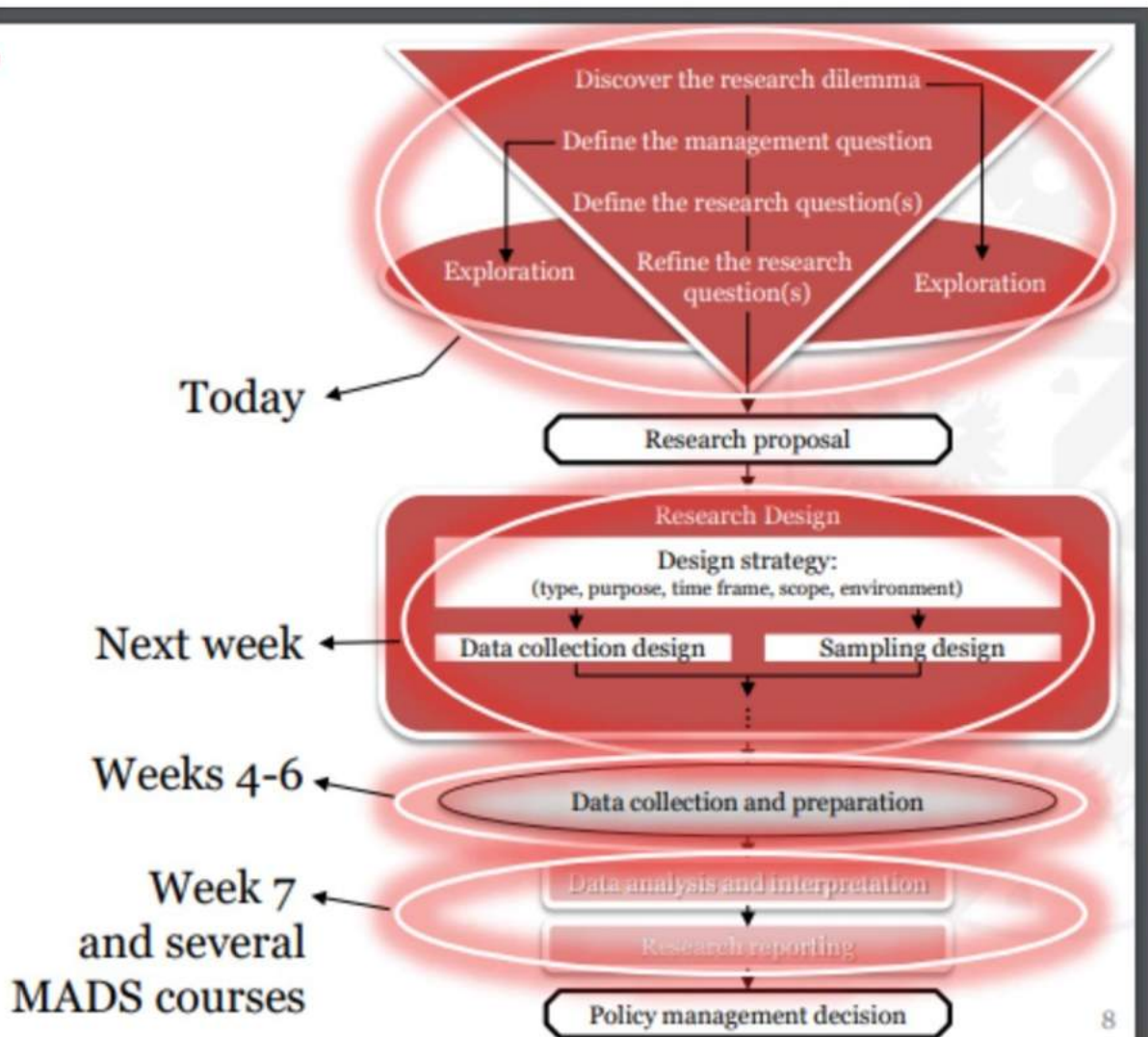


**Math & Statistical Knowledge**

- Experimental design building, Model Fitting
- No failures = no learning
- The challenge is not to get a working model, but to get the model to add value to the business

*Note from MADS Madlad: Wehkamp lecture continues with examples and description of the Wehkamp dataset. I think there is no more theory to summarise. Please refer to the original lecture for further materials on this lecture.*

## Recap of the course &amp; assignment(s):



Reading – HBR Article (optional)

#### **4 Steps to Management Questions**

**Step 1:** Articulate the problem simply.

“We are looking for X in order to achieve Z as measured by W.”

- What is the basic need?
- What is the desired outcome?
- Who stands to benefit and why?

**Step 2:** Justify the need.

- Does it align with organization’s strategic goals?
- What are the desired benefits & how do we measure them?
- How will we ensure that a solution is implemented?

**Step 3:** Contextualize the problem in terms of the past tries.

- What approaches have we tried in the past?
- What have others tried?
- What are in/external constraints on implementing a solution?

**Step 4:** Write the problem statement.

- Is the problem actually many problems?
- What requirements must a solution meet?
- Which problem solvers should we engage?
- What do solvers need to submit?
- What are incentives for the solvers?
- How will solutions be evaluated and success measured?



Reading – Book: Verhoef et al. (11.6 required +11.1 optional)

## **Section 11.6 (required): “OPPORTUNITY FINDING AS A METHODOLOGY TO CREATE MORE VALUE”**

*Note from MM: Opportunity finding is basically the opportunity tree from lecture 1.*

Data Scientists could immediately start analyzing the data and start data science projects, but that is not the point. The goal is to have **value impact** in the organization.

For that reason, this chapter discusses a methodology used to detect opportunities: **opportunity finding**.

**Opportunity finding:** structured way to identify solutions and breaking them down into initiatives.

It helps put a focus on the most valuable initiatives and facilitates monitoring progress on objectives for the defined business challenge.

### **7 Steps of Opportunity finding**

#### **1. Business Challenge (BC)**

- a. Linked to the phases of the customer lifecycle (e.g. customer acquisition) or the phases of the customer journey.
- b. Quantified in terms of the necessary potential to be realized. In other words, the delta from current to desired state (e.g. increase X from 3% to 8%).
- c. Expressed in a monetary value and measurable by at least one or more defined KPIs.

#### *Example*

“We aim to reduce the customer churn from 8.5% to 7%, to achieve an extra EBIT of 2.5m euro.”



## 2. Sub Questions

- a. Break BC down into sub questions. Use the 5 W's: Who, What, Where, When & Why.
- b. Specify BC for every W, when possible. Focus on the most important W's based on situation.

*Example*

*(continued)*

**Who:** Who are the customers that show an above or below avg. churn percentage?

**What:** What products have a higher or lower product churn compared to others?

## 3. Factors

- a. The factors behind the sub questions define the levers that can actually be used to start defining initiatives for the business challenge.

*Example (continued)*

In our "Who": we could define **age** as a possible factor that might be relevant in finding groups with a higher or lower churn percentage. Another angle could be to break down the customer base by value, creating **value segments** to explore whether different segments have different churn.

## 4. Hypotheses

- a. The hypothesis step is crucial in guiding the analysis process. It is helpful to make a "brain dump" of all possible hypotheses.

Example (continued)

Who (two factors: age and value) + Why (reason of churn):

"High value, young customers, show a high churn rate, due to cheaper alternatives and low switching barriers."



## 5. Insights

- a. Determine analyses questions to check the hypothesis and identify areas with high potential.
- b. Potential is identified in the differences in performance we can uncover along the primary KPI of our business challenge (in our case this is churn).

## 6. Initiatives

- a. Initiatives are specified as ideas to realize by addressing a target group, via one or more channels, at a specified time, with a well-targeted proposition.

## 7. Impact (in euros/dollars/...)

- a. The financial impact is calculated per initiative to identify the most promising ones, but also to validate that the initial potential of the business challenge is feasible.

*Tip (from book):*

We suggest plotting the initiatives in a matrix with the two dimensions: necessary effort to realize an initiative and the potential value of every initiative. Then start with the low effort, high value ones first.

## Section 11.1 (optional): “OPPORTUNITY FINDING FOR SURE.COM”

*Note from MADS Madlad: This is an assignment/exercise in the book for further material please refer to the original book. In short it presents you with information about the business and your task is to write down each step of the opportunity finding process summarized above (Section 11.6).*

Reading

Book:

Business Research Methods (2.1 optional + 2.2 required)

**Section 2.1 (optional): The research process**

