DATA UNDERSTANDING AND COMMUNICATION

Summary of basic information from the DUCK Course

Spis treści

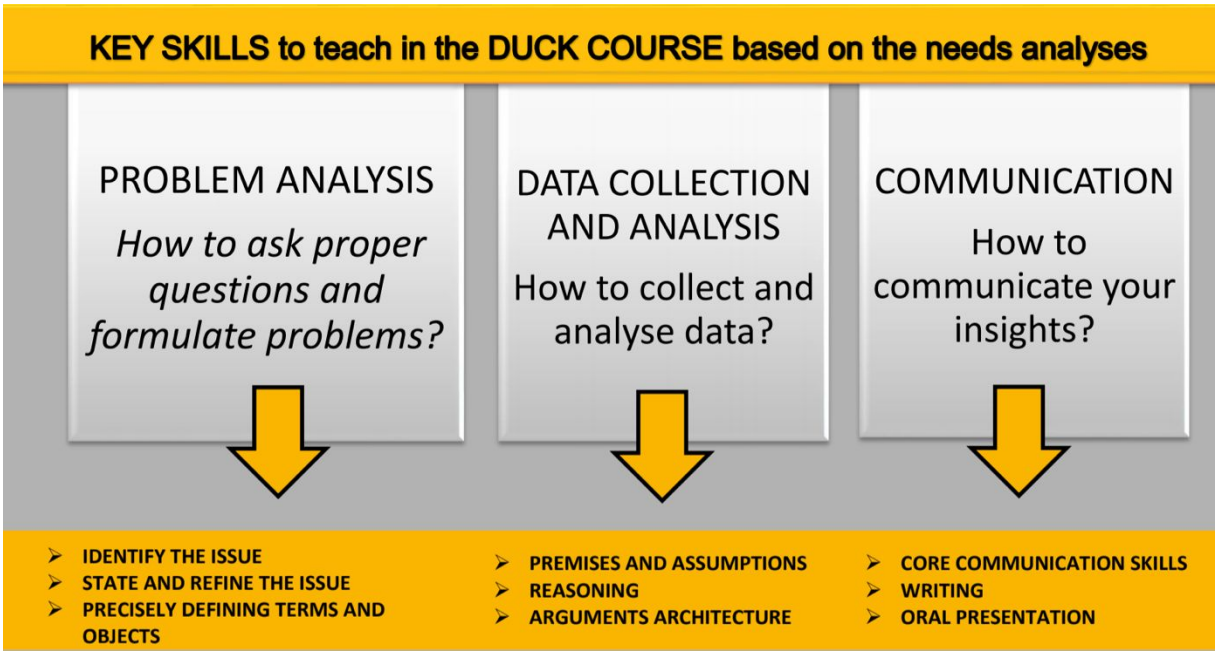
INTRODUCTION

Industries generate more and more data (see Internet of Things, BigData, etc), and along with it the requirement of not only technical, but also strong analytical skills. Business Intelligence technology makes complex data analysis easier to perform without the need of low-level programming skills. Thus, understanding and communicating data become critical for companies, and data scientists are becoming a very scarce resource, and lack of data communication and visualization skills, including data storytelling and infographics, lack of knowledge of basic tools to analyses and present data become a serious problem for the modern labour market.

The skills of the teachers and trainers to analyse and communicate data is not just a question of the quality and the presentation of their research: it is also important to note that this issue is not only concerns the data that they have to analyze for their research work but they also have to analyse data regarding their students and their students' needs.

The DUCK project developed a 30-hour multimedia, mixed-method curriculum consisting of three modules: problem analysis, data analysis, and data communication modules. The curriculum is suitable for introduction in vocational education and training, higher education and adult education. Once completed, teachers and trainers (as well as their students) will be able to analyze and present data more effectively, thereby enriching the quality of their education and improving their employability.

This document summarizes the most important parts of the curriculum. Suitable for use as a note during learning or as a reminder after completing a course.



MODULE 1: PROBLEM ANALYSIS

Learning goals:

Ability to define problem:

- Investigate the problem in order to better understand the goals of the project;
- to state and refine the question: Know different types of questions, main characteristics of a good question, translate a question into a problem;
- to understand inputs: Data, Concepts, Attributes, Examples
- to understand potential knowledge representations of the solution: Tables, Linear Models, Trees, Rules, Clusters

Skill 1: Identify the issue

The basic identification of finding the real **issue** is starting with analyzing the letter, message or the text that we have got in the following way:

1. The manager or leader directly tells it in the title: Production crisis – how do we react?

2. The issue is explicitly written in the beginning of the text.
3. There is a clue in the text: “What should a rector react in this situation?”
4. The knowledge of the background or responsibility of the sender of the text, or message: A scientific vice-rector will search for research data increase rather than enrollment or income problems.

If we can find the **conclusion** we can mirror the real question or issue. To find the conclusion we are:

1. Searching for indicator words such: consequently, therefore, thus.
2. Looking at typical locations: Beginning of the message – end of the message.
3. Closing out: examples, statistics, definitions, evidence that are not conclusions.
4. Background of author here again.

We have descriptive and prescriptive issues:

Typical questions start with “Where?”, “What?”, “When?” and ``How?”.

Typical prescriptive questions start with: “Should?”, “Ought to?”, “Must? Is it good?”.

Try to avoid prescriptive issues, analyze and **formulate descriptive** issues: First we want to understand **why** things happen(ed) and only at a second step whether it is good or bad for the observer.

Skill 2: State and refine the questions

Type of questions:

1. Descriptive question. This is the most fundamental question among others. This is the most factual and easy to understand one. This question is summarizing, highlighting a characteristic of a given set of collected data.
2. Exploratory question. This is a hypothesis-generating question following a general feeling or thought, to check trends, patterns, relationships. Therefore, if you feel that exam results are somehow related to sleeping time, or you feel that dropout rate might be in relation with exam timing patterns, you pose exploratory question.
3. Inferential question. This question is a restatement of the hypothesis to see exactly in which data sets and to what extent this hypothesis is true.

4. Predictive question. Predictive question is trying to highlight whether a given behavior or fact will be observed. You are interested what predicts the behavior or fact.

5. Casual question. A casual question asks about the relation of factors. It asks whether increasing or decreasing a factor will effect change in another factor.

6. Mechanistic question. This type of question is answering how one factor is effecting the other factor.

To formulate a question there is a three steps model:

	Question 1st draft	
Expectations	Comparison: Does in fulfil expectations?	Information about the Question 1st draft
New expectations	Question 2nd draft	Information about Question 2nd draft

What kind of information are important about the question?:

a. The question should be **of interest** (target population).

b. The question is **not already answered**.

c. The question is **feasible** (likely or plausible).

d. The question is **answerable**.

e. The question is not **specific** enough (e.g. general, fuzzy).

Further characteristics:

1. The question must be **operationalized**.

2. The question should not contain **confounding.**

3. The questions should not lead to **biased** result.

Open or closed questions?

In our problem formulation (data understanding) module, it is important to say that:

1. For very specific problems we suggest so use closed questions (starting with Is it? Does it? etc.)

2. For exploratory problems we suggest to use very open questions (starting with How? etc.)

Skill 3: Precisely defining terms and objects

In order to make definitions more precise:

- **Use precise language**: A vocabulary of precise nouns and vivid verbs helps you create strong mental pictures and avoid wordiness.
- **Use concise language**: Using the fewest possible words without sacrificing meaning makes your writing more understandable.
- Before using a synonym that you are not familiar with, make sure you are aware of any connotations
- Avoid jargon: even if your audience is familiar with it, comprehension is better if you use plain language
- Define terms: if you use technical terms, define them clearly at the beginning

  Ambiguity is a natural occurrence in the language. When something or someone is ambiguous, it does not mean negative intent or improper actions.

In order to avoid ambiguities:

- It is best to narrow down our search: crucial terms and phrases are those that are present in the conclusion of the argument and in the reasons for that conclusion.
- When you have defined those, check if the author provided a definition of the terms used.
- Another good method to locate ambiguities is to look for abstract words and phrases in the conclusion or the reasoning.

MODULE 2: DATA ANALYSIS

Learning goals:

Ability to gather, clean and analyse data (find different file types; extract data; organize, merge and manage the data);

- to explore/understand statistics and probability (types of data; probability & expected values; variability, distribution, asymptotic);
- to understand data with visualization;
- to understand data with descriptive statistics (take a sneak peek at the data and identify its dimensions and data types, compute, visualize and discuss simple statistics and summarize the data);
- to use basic tools to analyse data.

SKILL 1: UNDERSTAND DATA

We have to make a distinction between data, concepts, attributes:

Data consists of observations about records in a database, notes in a logbook, images on a hard drive. These observations may prove useful or useless, accurate or inaccurate, helpful or unhelpful. At the outset, they are only observations. Observations alone are not enough to act on. When we connect observations to how a matter of concern works, we have the opportunity to make knowledge.

We talk of a concept - as formal definition of what is being researched. It is a general, basic form of a particular set of instantiations (specific though different).

Attributes are characteristic or quality of objects (person, thing, etc.) and are closely related to variables.
A variable is a logical set of attributes. Variables can "vary" - for example, be high or low.

Making a graph to visualize your data is a good way to further your understanding of your question and your data. Your graphic possibilities depend on the data you wish to present. The graph types of data visualization are the following:

- Distribution: violin, density, histogram, boxplot, ridgeline;
- Correlation: scatter, heatmap, correlogram, bubble, connected scatter, density 2d;
- Ranking: barplot, spider/radar, worldcloud, parallel, lollipop, circular barplot;
- Part of the whole: treemap, venn diagram, doughnut, pie chart, dendrogram, circular packing, sunburst;
- Evolution: line plot, area, stacked area, streamchart;
- Map: map, choropleth, hexbin map, cartogram, connection, bubble map;
- Flow: chord diagram, network, Sankey, arc diagram, edge bundling.

The process of exploring the data includes examining the structure and components of the dataset, the distributions of individual variables, and the relationships between two or more variables. The goal of exploratory data analysis is to get you thinking about your data and reasoning your question.

Exploratory data analysis checklist:

1. Formulate your question as precise as possible
2. Read in your data
3. Check the packaging
4. Look at the top and the bottom of your data.
5. Check your numbers, verify counting
6. Validate with at least one external data source
7. Make a graph or a map
8. Try the easy solution first
9. Follow up questions

Skill 2: Understand statistics and probability

We have arrived to the end of this skill. The most important topics that we covered are:

In the beginning, we presented the basic definitions which are necessary to understand the basics of descriptive statistics and the results of data statistically processed. You should remember that:

1. The median is useful when we have in a data set such values that deviate from the arithmetic mean of this set.

2. The histogram perfectly illustrates data variability in a set.
3. The probability value of an event influences the expected value of the result, and the approach works well in portfolio methods when the risk is limited by diversity, as it is when constructing an investment portfolio of shares.

In the next unit, we learned that:

1. The variability of data set can well describe the data set variability: mean, median, standard deviation.
2. Reliable sources of statistics are publicly available, and the great benefit of such sources is the harmonization of data, as is the case with data provided by Eurostat.
3. Percentiles can be useful to understand and interpret measurement data.

In the last unit of the study, we dealt with the correlation of two variables. We learned that:

1. Correlation, in other word dependence, can be identified and quantified using a correlation coefficient.
2. Value classification allows you to specify the direction (positive or negative) and the strength of the correlation.
3. Commonly used spreadsheets shown used in illustrating examples will enable you easily to do your correlation analysis of two variables.

We would like to remind you again that the scope of statistical issues presented in '*Understand statistics and probability*' is limited to basic concepts and approaches. We are aware that statistics courses are standard in high school programs and the humanities and engineering studies. The basics of statistics are essential in the context of the module '*Data Analysis*' and a completed 'skill' can be a reminder of acquired knowledge, as well as encouragement to reach for advanced tools.

Skill 3: Premises and assumptions

In the first step in making an argument, you need to know what you are trying to prove and state your reasoning so it is important to be familiar with conclusions, premises and reasons:

Conclusion is always responding to an issue, and an issue is a question or controversy responsible for the conversation or discussion.

Reasons are beliefs, metaphors, and other statements. These are inherently statements that form the basis for creating the credibility of the application. In other words, these are explanations or justifications of why we should believe in a particular conclusion.

A stated reason that supports the conclusion is a premise. A premise is any statement that the author uses to support the conclusion.

An unstated reason that supports the conclusion is an assumption. Assumption could be value (prescriptive) or descriptive. Value or prescriptive assumptions are beliefs about how the world should be. Descriptive assumptions are unstated beliefs about how the world was, is, or will become.

To make a statement or an argument, you need truly authoritative sources. Always cite yours sources and always rely on reliable, informed and impartial sources.

Skill 4: Reasoning

Categories of disputes provide a framework for understanding how to make a coherent argument.

A very powerful way to organize our thoughts is by classifying each point of dispute in our argument.

A point of dispute will fall into one of four categories: fact, definition, value, and policy.

Reasoning, as a way of proving arguments, comes in many different forms. The following are useful reasoning strategy:

- Reasoning by generalization
- Reasoning by analogy
- Reasoning by cause
- Deductive reasoning

Fallacy is a reasoning "trick" that an author might use while trying to persuade you to accept a conclusion

Three Common fallacies:

1. Providing reasoning that requires erroneous or incorrect assumptions, thus making it irrelevant to the conclusion
2. Distracting us by making information seem relevant to the conclusion when it's not
3. Providing support for the conclusion that depends on the conclusion being true

Other common reasoning fallacies: *Ad hominem*, *Slippery slope*, *Searching for the perfect solution*, *Equivocation*, *Appeal to popularity*, *Appeal to questionable authority*, *Appeal to emotion*, *Straw person*, *Either-or (false dilemma)*, *Explaining by naming*, *Glittering generality*, *Red herring*, *Begging the question*, *Hasty generalization*.

Skill 5: Arguments architecture

We have arrived to the end of this skill. The most important topics that we covered are:

“Epicycle of data analysis” is 3-step process including:

1. Setting Expectations,
2. Collecting information (data), comparing the data to the expectations, and if the expectations don’t match,
3. Revising the expectations or fixing the data so the data and the expectations match.

Constructing arguments means constructing specific conclusions backed by evidence and reasons. In this process you need:

- Explore the issue and identify possible positions
- Spell out basic ideas as arguments
- Develop argument in more detail
- Consider objections
- Seek informed and impartial sources

- Cross-check sources

Arguments begin by preparing reasons and organizing them in clear and fair way. In this process you need:

- Identify premises and conclusions
- Develop your ideas in a natural order
- Start from reliable premises
- Cite your sources
- Be concrete and concise

While you build the argument ask following questions:

- Does the argument make clear what a conclusion of the argument is?
- Does it present ideas in natural order?
- Are the premises reliable?
- Could the argument be clearer or more concise?
- If so, which words or expressions are unclear?
- Don't you confuse the argument by using more than one term for the same idea?

Use in your arguments words that function as premise indicators and conclusion indicators:

- Premise indicators: for, since, because, for the reason that, granted that, this follow from
- Conclusion indicators: thus, therefore, so, hence, consequently, it is shown that

MODULE 3: DATA COMMUNICATION

Learning goals:

Ability to interpret the results and communicate findings,

- to prepare presentation of results,
- to create visualizations of the results (including data storytelling and info graphics),
- to use basic tools to present data,
- to efficiently communicate the results to target audience (with appropriate content, style and attitude).

SKILL 1: CORE DATA COMMUNICATION SKILLS

From the technical point of view, **data communication** is understood as the transfer of data or information between a source and a receiver: the source transmits the data and the receiver receives it. In the context of this course, we understand data communication as the ability to communicate insights from data analysis and understanding. The basic identification of the core concepts of data communication includes distinguishing the 6 basic components in this area: sender, receiver, message, medium, protocol and feedback.

There are **many methods available for communicating your information to others**. In the frame of this course, we are presenting here some of the most important methods, including: Fact sheets, reports, online summaries and in-person presentations (summarizing other methods such as phone calls, email messages, fora…etc). How will you choose the right method(s) for your project?: you will need to choose a combination of methods that will give you the best chance at communicating your message. It is very important to think of your **target audience** and its key features before you choose.

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Key Performance Indicators (KPIs) are the most important metrics for your particular project, the keys to evaluating the crucial factors for making sure you achieve your goals. The key questions that will help you select the best visualization types:

- What story do you want to tell?
- Who do you want to tell it to?
- Are you looking to analyze particular trends?
- Do you want to demonstrate the composition of your data?
- Do you want to compare two or more sets of values?
- Is timeline a factor?
- How do you want to show your Key Performance Indicators?

In the frame of this course, we understand **tools for data communication** as the instruments to help you deliver the visuals that have been dealt in the previous Units: Unit 2 (Choosing the best communication method for a given communication goal) and Unit 3 (Identifying the best possible visualization type for a given data / problem type). Some of the most important instruments are:

- **Spreadsheets: are** computer applications for organization, analysis and storage of data in tabular form. Spreadsheets were developed as computerized analogs of paper accounting worksheets.
- **Business intelligence tools including dashboards**: Business Intelligence (often called BI) software is a type of application software designed to retrieve, analyze, transform and report data for business intelligence. In this frame, a **data dashboard** is an information management tool that visually tracks, analyzes and displays Key Performance Indicators (KPI), metrics and key data points to monitor them.

Graphic design software: a program or collection of programs that enable a person to manipulate images or models visually on a computer.

Skill 2: Writing skills

The **lead** in writing is the opening hook that pulls anyone in to read a story. A well-written lead (and article) should answer most or all of the 5Ws and H: 1. Who is it about?, 2. What happened (what's your argument)? 3. When did it take place?, 4 Where did it take place?, 5 Why did it happen? and the H: How did it happen?. The things to avoid include: flowery language, unnecessary words or phrases, formulaic leads, and avoid starting your lead with “It”, it is not precise and disorients the reader.

A **claim or a proposal is the primary argument and could be the most important aspect of the writing**. The basic tips to elaborate claims and proposals are: t**hink of your Audience (**you have to convince a reader that your idea is worth pursuing, so who are you addressing? ; **research (**make sure your support all your claims with evidence); f**ocus on one main idea (**Strong claims and proposals are clear, concise, direct, and even provocative); p**re-Write (**spend some time brainstorming ideas and then spend some time thinking about how you'd like to organize them) and **Revise, Revise, Revise (**Never turn in a first draft: have a trusted peer or colleague reading your paper and giving you feedback. Then take some time to incorporate that feedback into a second draft).

An **argument in an essay** is usually a main idea, often called a "claim" or "thesis statement," you can also use evidence to refute the statement. It has to be convincing. To write an argument we advise you to follow these 3 initial steps:

1. to understand the question; what do you have to do?, what issues you need to cover?
2. to do your research: what do we know about this?, what do other researchers say?, what are the key facts of these debates and problems we are researching about?
3. then, go back to the question and see what you think: what you think is the answer. This will be your claim. and you need to argue for this position in an academic context.

An **objection** is a kind of an argument. Since all arguments have both premises and conclusion, every objection therefore also needs premises and conclusion.

Always **consider the strongest objections**. It may be tempting to pick a weak objection, one that's easy to refute, and reply to that. But doing this will strengthen your argument, because it will not satisfy a thoughtful skeptic. A **good argumentative essay is always going to look for the strongest possible objections to its main argument, present them accurately and fairly, and then attempt to systematically respond to those objections**.

When you make a concession, you actually give in to part of the other person's objections or views. You admit that he is half right, perhaps, or that s/he has a valid concern. Then you overcome that concern by logic and/or a solution.

What to avoid: overusing some word; focusing on different ideas or repeating your ideas and slang and informal speech. Plus: avoid the temptation of choosing a weak objection and do not be too rigid: making concessions to possible objections is a good practice.

Skill 3: Oral presentation skills

Oral Presentation skills are the skills you need in delivering effective and engaging presentations to a variety of audiences. Audience activation is key to the success of a presentation: it is using about using real-time data to deliver personalized and relevant experiences and stories that engage your audience.

Involving your audience is essential to making an impact. Your presentation should pull them in, get their attention and stimulate their thoughts and understanding. This can be done, working hard in a number of ways:

- The way that you plan your presentation will be critical in terms of **using language and ideas that your audience will understand**.
- You must also ensure that there is sufficient **time for questions and discussion**.
- The way that you deliver your presentation should **create a bond with your audience**.
- Your use of **eye contact, gesture, spoken language and energy should communicate effectively and enthusiastically** with all areas of the room, thus ensuring that the audience receives positive messages about you and your material.

Audience members have short attention spans and as a public speaker you need to work hard to continually capture their attention. By taking audience members on a journey, as well as letting them know where about on the journey they are allows you to maintain their attention so you can continue to deliver your message.

Visual aids may help you deliver a better presentation: they can help you explain information more coherently which makes presenting easier for you and learning easier for the audience. The different kinds of visual aids you can use in a speech are objects, models, photographs, drawings, graphs, charts, transparencies, video, multimedia presentations, and even yourself.

Receiving feedback from your presentations is the best way to grow as a presenter: when we think of critique, it often makes us uncomfortable but critique is not censured criticism, it's often helpful assessment. So if you are ready to grow your oral presentation skills, you need to be open to collecting and using genuine feedback.

Ending your presentation in style is important. The role of a conclusion in a speech is to signal to the audience that the speech is coming to a close and help them remember the most important points from the speech. The end of your speech is going to form your audience's lasting impression of everything you've said. Just as there is not a single way to reach out to your audience, there is no single way to end up a presentation. Find something that suits your personal voice and style (we have included several tips on our resources) and figure out how to make it work.