Designing a Data Science Course for Non-Computer Science Students: Practical Considerations and Findings

Yllka Velaj*, Dominik Dolezal†, Roland Ambros‡, Claudia Plant§, and Renate Motschnig¶
Faculty of Computer Science, §ds:UniVie
University of Vienna
Vienna, Austria
Email: *yllka.velaj@univie.ac.at, †dominik.dolezal@univie.ac.at, ‡roland.ambros@univie.ac.at,
§claudia.plant@univie.ac.at, ¶renate.motschnig@univie.ac.at

***Abstract*—This Full Paper in the Research-To-Practice Category illustrates what an online survey of students' opinions reveals about students' perceived learning and take-away from a course on Data Science. The course is offered at University of Vienna in the Business Analytics, Data Science, and Digital Humanities faculties as part of the masters' programs. In this work, we outline the course structures, goals, modules, and present preliminary findings gained while teaching the course. Due to the global pandemic and resulting lock-downs, the course was held online, hence, we also analyze the effects of the current situation on the students.**

The results revealed that the structure of the course is appreciated by the students. Furthermore, the students liked the open source software taught in the course where they can create visual workflows with an intuitive, drag-and-drop style graphical interface, without the need for coding. The results also confirmed our hypothesis which showed that working in groups is more complex and difficult using online tools. We learn that the instructor-generated technique for forming the groups assigning to each group students with different backgrounds, lead to teams that are able to solve problems faster as they are more cognitively diverse. These findings confirm that the approach used in the Data Science course is viable for teaching computer science skills to non computer-scientist and can be used by other educational institutions.

***Index Terms*—data science, emergency remote teaching, distance learning, 21st century skills, computer science education, student-centered learning**

I. Introduction

Data Science is one of the most popular emerging domains and it provides tools for 21st century jobs and for the most required career options nowadays.

The importance of Data Science grew with the increasing amount of available data which is a precious asset of any organization. It helps companies to understand and enhance their processes.

Data Science involves mining large datasets containing structured and unstructured data and identifying hidden patterns to extract actionable information. The importance of Data Science lies in its innumerable uses. It is a flourishing industry and an interdisciplinary field that encompasses Computer Science, Statistics, Inference, Machine Learning algorithms, Predictive Analysis, and new technologies. Moreover, it has many applications in different domains including health-related research, financial market analysis, fraud detection, energy exploration, environmental protection, and more.

Hence, the goal of this paper is to present the design strategy of a course whose goal is to provide the students with the fundamentals for setting up, managing, and conducting Data Science projects. The course is offered at University of Vienna in the Business Analytics, Data Science, and Digital Humanities faculties as part of the master's programs.

One of the main challenges in designing this Data Science course is to take into account the different knowledge that students with different backgrounds have, and to provide computer science-related skills to non-computer scientists. The main skills that the students should acquire thanks to the Data Science course are: to be able to understand the variability and uncertainty inherent in the problem, to develop clear statements of the problem/scientific research question, and to understand the purpose of the answer. Moreover, the students should be able to obtain high-quality data from the original dataset and to approach modeling as a process that requires an overall strategy, not an application of a collection of techniques or methods.

With our paper in the research-to-practice category, our goal is to propose a new method for designing a course that is suitable for teaching data science and is also widely accessible to a broad range of students. To the best of our knowledge, there are very few works that focus on the design of data science courses whose aim is to provide students with heterogeneous backgrounds the skills they need in the 21st century.

II. Background

A. Teaching Data Science

Motivated by the previous observations, many courses have been designed to prepare students to become data scientists with the correct blend of practical and academic knowledge

We thank the Austrian Federal Ministry of Education, Science and Research for providing the funding for the project "Teaching Digital Thinking".

and skills. In [1] a flipped classroom approach was adopted in a post-graduate Data Science course. The study compared student evaluations and implementation experience of traditional teaching methods with a flipped classroom approach. This approach switches the traditional activities conducted in the classroom and those conducted at home, i.e. lectures are watched at home whilst homework questions are discussed in class. The results show that great strides have been made in the adoption of the flipped method for the lectures.

The authors in [2] design a low-level undergraduate introductory Data Science course. As our Data Science course, also the course described in [2] is multidisciplinary and incorporates topics from different domain areas, such as astronomy and social networks. The need for a course that provides examples of data science applications from different fields is related to the heterogeneity of the background of the students enrolled in the course. It is of crucial importance to find interesting topics for all the students to keep them motivated. Moreover, the description of different case studies will present to the student a broad overview of Data Science application scenarios. The results in [2] are in line with our findings, the authors confirm that designing an interdisciplinary Data Science course for students with different levels of computer science knowledge is challenging.

In [3] the authors underline the importance of introducing a Data Science course in undergraduate degrees in statistics. After analyzing seven different study programs in various universities, the authors conclude that, not surprisingly, a course that incorporates Data Science can be a way to motivate students to further study statistics as the students perceive these courses as relevant and exciting. These courses, like the one presented in our work, present the students with projects and assignments that can be adapted or extended to meet the levels of the students.

Data Science courses have been created and implemented also for secondary students. The description of the course "Introduction to Data Science" is presented in [4], which aims to develop computational and statistical thinking skills so that students can access and analyze data from a variety of sources.

Moreover, it is worth mentioning that Data Science has received attention also in the Education field, which is known as Educational Data Science [5], it works with data gathered from educational environments/settings for solving educational problems. Lastly, the work in [6] reports the results of an empirical study conducted to assess the entry competencies of students, potential candidates for a Data Science master's program. The main finding is that technical competencies are not critical for success in studying Data Science.

B. 21st Century Skills

Using digital media to acquire subject-specific competencies, autonomously working on projects, solving problems, communicating, and collaborating in a self-directed manner are some examples of the most essential 21st century skills. In particular, in [7], the authors identified three categories of 21st century skills: "learning and innovation skills", "digital literacy skills", and "career and life skills". Learning and innovation skills encompass critical thinking and problem solving, communication and collaboration, and creativity and innovation. The authors in [8], instead, conducted a comprehensive literature review and integrated the literature on digital skills with the theory of 21st century learning to develop a "framework of 21st - century digital skills". Based on the developed framework, they designed and validated a questionnaire designed to quantify these kinds of skills [9], [10].

The Data Science course provides the students with 21st century these skills using multiple approaches. During the project work, the students are presented with problems, and datasets, which they are asked to analyze and solve in groups. They learn how to collaborate and communicate and this improves their ability to work in teams and communicate effectively. In the project work, students are given the freedom to be creative in selecting the main questions they would like to answer, given the dataset thy are provided with. The goal of the course is to close the gap between the skills required in professional life and those acquired at institutions of educations [11]–[13].

C. Student-Centered Learning

Student-centered learning is a learning approach where students can self-direct themselves using non-conventional elements, such as peer review, and self-assessment. It plays an important role in student- centered approaches, which are suggested methods to enhance 21 st century skills [14].

The research in [15] describes person-centered therapy and learning providing numerous examples of successfully implemented learner-centered approaches employing self-directed learning [16], [17]. While, in [18] the authors criticized previous works for the lack of possibilities for cooperation among students, while the business sector had long recognized the importance of interpersonal skills.

In this work, we highlight how working on Data Science projects in groups, and the peer review task on the project helps students become self-responsible learners and fosters their personal and social skills.

D. Emergency Remote Teaching

This sudden and challenging switch to online learning is now known as "emergency remote teaching" in the literature [19], [20]. The SARS-CoV-2 pandemic presented educational institutes with the challenge of transforming all courses to pure online learning within a few days. On March 2020, many European universities were ordered to teach their students in a distance education setting by government decree [21], which was issued four days before taking effect.

In this paper, we provide insights into the effects of emergency remote teaching on a Data Science course offered at the Master's program of University of Vienna.

Differently from typical online teaching which provides well-prepared courses, emergency remote teaching, refers to a response to a crisis that requires to temporary switch to pure

remote teaching that otherwise would have been provided in a face-to-face setting [22].

Many researchers investigated the effects of emergency remote teaching. In [23] the authors compared the performance of students who experienced emergency remote teaching with learners who were taught in a face-to-face setting. The paper evaluates the grades for different courses such as English, business, computer programming, and communication. The conclusion is that students experiencing emergency remote teaching performed equally as those who were taught in a face-to-face setting. Other studies [24], [25] also show an increase in students' academic performance during the emergency remote learning by comparing the academic record of 2020 with those of the previous years.

In this work, we do not have data about the Data Science course held before the SARS-CoV-2 pandemic since the course was held for the first time in October 2020. We provide an analysis of the grades of two different semesters.

E. Data Science: Course Description

The Data Science course is offered at University of Vienna in the first semester of the Business Analytics and Data Science masters' programs, and in the third semester of the Digital Humanities masters' program. It provides 6 ECTS, i.e., the European Credit Transfer System. The course is denominated *Doing Data Science*.

The goal of the Data Science course is to provide the students with the skills needed to set up, manage and conduct data science projects themselves.

Students acquire knowledge of processes describing how to approach and implement data science projects. They know the steps of CRISP-DM [26], an acronym for Cross-Industry Standard Process for Data Mining, which is a standardized process that describes and codifies common approaches used by data mining experts: it is the most widely used analytical model in the industry. Moreover, students learn about various cases of how to apply this to different applications (from different areas such as business, humanities, etc..).

This class consists of different parts:

1) Part I: How to approach data science projects
2) Part II: Showcase examples for data science projects
3) Tutorial: Introduction to the data science tool KNIME
4) Group Work: Students work in an interdisciplinary group on a data science project

KNIME [1] Analytics Platform [27] is the open-source software for creating data science. Students can create visual workflows with an intuitive, drag and drop style graphical interface, without the need for coding. Moreover, other showcases are introduced by guest speakers.

An example of a workflow in KNIME is shown in Figure 1.

The project work is organized in four steps: in the first one, the students indicate their preferences on the topics, we try to accommodate these topic requests and assign the students to an interdisciplinary group of 3-4 members. In the second step, the students address the part "Project and Data Understanding" of the CRISP process and have the first meeting with their supervisor. In the third step, the groups work on the "Data Preparation and Modeling" phase and have a second meeting with the supervisors. The last step of the project work consists of the "Data Modeling and Evaluation" part.

In the more recent editions of the course, the second meeting with the supervisor has been substituted with a peer-reviewing process. The students are provided with a "peer-review questionnaire" that they can use to evaluate the project work of their peers, then random pairs are created so that each group can review and get a review from another group. On a given date, each pair of groups meets and discuss the state of the projects. After the meeting, the students have one week to present their feedback to the lecturers and the peer group.

At the end of this process, the students prepare a poster and a video where they describe their work.

There are three exams in total - a written mid-term and a written final exam as well as an oral exam in terms of poster pitch videos and discussions on the investigated project topics.

On the day of the oral exam, all the participants watch the poster pitch and take part in a Q&A session. Afterward, everyone is invited to join Wonder[2], an online gatherings platform, where each group stands in a "corner" and all the participants can move and join a group to ask questions or to provide feedback.

The different parts of the course contribute to the final grades as follows: the midterm and final test provide both 30% of the final points, while the project work provides 40% of the points. In particular, the points of the project are distributed as follows: poster 30%, poster video 30%, KNIME process 20%, and peer-review report 20%. When there were 2 meetings with the supervisor and without a peer-reviewing process the poster counted for 40% of the points, the poster video for 40%, and the KNIME process for 20%.

During the classes, in the first part of the course, lecturers not only provide a description and an explanation of the main theoretical concepts but also list interesting real-life examples, and share personal experiences with data science projects and problem solving. In the second part of the course, each lecture is interactive and all the data science examples are analyzed in detail, for each step possible alternatives are described together with the pros and cons. This approach helps students acquire critical thinking skills.

F. Research Questions

In detail, we address the following research questions:

- RQ1: Is the data science course provided in the masters' programs of different faculties well designed for students with different backgrounds?
- RQ2: Does the course provide the master's students with diverse backgrounds with the expected and needed skills to carry out data science projects?

[1]https://www.knime.com/

[2]https://www.wonder.me/

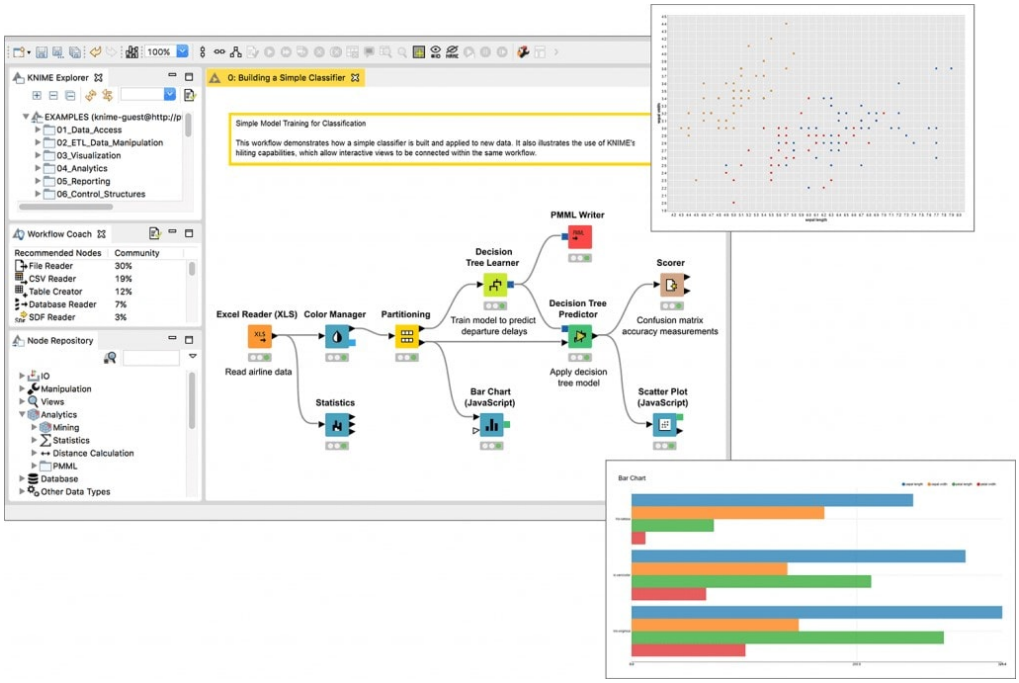

Fig. 1: KNIME software.

TABLE I: Questionnaire answers.

Ad-hoc questionnaire

Semester	Participants	Survey S1	Answers %
WS 2020	38	21	55.3
WS 2021	48	13	27.1

Standardized questionnaire

Semester	Participants	Survey S2	Answers %
WS 2020	38	27	71.1
WS 2021	48	21	43.75

- RQ3: How did emergency remote teaching affect student performance and what is the lesson learned?
- RQ4: Is the software presented in the course for the project work suitable for students with heterogeneous backgrounds?
- RQ5: Is the group formation process that is adopted by the lecturers adequate for the project work?

III. Methods

To address our research questions, in addition to our observations and dialog with our students, the following data were analyzed:

Ad-hoc designed questionnaire (S1): We conducted an anonymous survey among our students to collect their feedback on the Data Science course. To find out whether students understood the questions in the way we expected, we conducted a pre-study with a few students who filled out the questionnaire and shared their reactions to the questions. Then, we submitted the final questionnaire and evaluated the answers. The surveys presented Likert-scale and open-ended questions.

Standardized course feedback questionnaire (S2): We evaluated the responses to the standardized online questionnaire proposed by the University at the end of the course. The survey consisted of several Likert-scale questions to be rated on a five-point semantic differential scale [28] from "strongly agree"=1 to "strongly disagree"=5. Students could also provide written feedback via open-ended questions asking them for feedback as well as suggestions for improvement.

The number of answers for these questionnaires is provided in Table I.

Analysis of final grades: To compare the learning outcomes and the overall performance of the students, we evaluated and compared the final grades of students upon completion of the Winter term in January 2020 and the Winter term in January 2021. The final grade is a common measurement of the overall performance [29] and is calculated by taking into account all grades given for the different parts of the course using the following mapping: 1=Very good, 2= Good, 3=Satisfactory, 4=Sufficient, 5=Failed.

Students interviews: To collect qualitative feedback from students, in addition to the surveys, students interviews have been conducted with ten students in each semester. The students were asked for their opinion on the relevance of the data science course for their future. Based on the transcriptions, a thematic content analysis was conducted to find common categories.

IV. Results

A. Ad-hoc designed questionnaire

Copies of the ad-hoc questionnaire were sent out to the students to collect feedback on the Data Science course.

TABLE II: Ad-hoc survey for the Data Science course.

Questions	Answer type
Age	Open-ended
Gender	Mult. Choice
Background-Study program	Open-ended
Full time student or working student?	Yes–No
Was your preliminary knowledge adequate to understand the topics covered for the course?	Likert scale
Are the learning materials provided adequate to study the subject?	Likert scale
Do you like the modality for the course (all the 3 parts, i.e. lectures,tutorial, project)?	Likert scale
Is it a positive aspect to collaborate with students with a different backgrounds?	Likert scale
Do you like the organization of the course (first theory/then show cases/lastly project)?	Likert scale
Were you comfortable with Moodle or the digital media used in the course?	Likert scale
What would be the best way to organize this course, in your opinion?	Open-ended
How was the course that you think was the best one so far organized? (how were the exam/ lectures/was there a project? Please mention courses attended in any university	Open-ended
How can the lecturers help you to get the best out of this course?	Open-ended
Which skills do you plan on acquiring thanks to this course?	Open-ended
Did you use any further digital media during the course (YouTube, Forums, other websites) to prepare for the exam? Or for working on the project? (e.g. to prepare the presentation, the video, learn how to code, etc.) or was the material provided enough?	Open-ended
Did you like how the groups for the projects were formed? Do you have any suggestions (other ways e.g. random, self-selected, etc.)?	Open-ended
Was the peer review phase of the project useful? if not please provide your motivations.	Open-ended

The survey was presented to the student in both Winter Semesters (WS), i.e., 2020 and 2021. The questions are presented in Table II.

The main difference between the questionnaire designed for the winter semester 2020 and that for semester 2021 is the addition of the final three open-ended questions in Table II, i.e., the questions about the use of digital media during the course, the technique used to form the groups for the project work, and usefulness of the peer review phase for the project. This phase was introduced in the Data Science course only in the winter semester 2021.

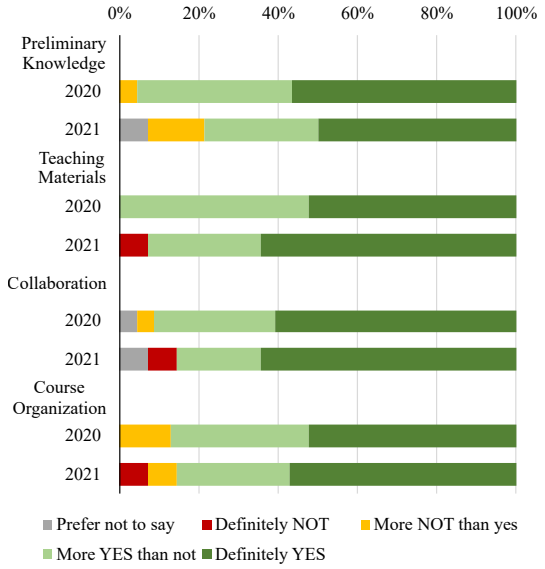

Fig. 2: Results for likert-scale questions of the ad-hoc survey.

The results of the 21 and 13 responses are shown in Figure 2. The respective response rates were 55.37% and 27.1%. The overall feedback on both semesters of the course was very positive.

The students stated that: the preliminary knowledge gained during their studies is adequate to understand the topics covered in the course, and the learning materials provided (slides, lecture recordings, project template, etc.) are adequate to study the subject. In addition, the online platform, Moodle, used for teaching the course during the COVID-19 emergency was easy to use for the students. They appreciated the structure of the course consisting of 3 main parts, i.e.(i) lectures with the theory first followed by show case,(ii) tutorial with the KNIME software, and then (iii) project work. They also evaluate positively the chance to collaborate with students with a different backgrounds.

The written feedback on the open-ended question can be summarized as follows:

Course organization. Many of the feedback on this aspect were positive for both semesters: ”I think the course is very well structured and I would not change the format or the organization.”, ”The organization in three parts was great”, ”The organization was already very well executed.”, ”I think the structure is fine as it is.”. Over the 34 answers, counting both semesters, 90% of them were positive, and the others provide suggestions on how to improve the course. An example would be to have a quiz after every lecture, which might make students focus and try to get the best from every lecture.

Other courses structure. When asking the students to describe the best courses they attended during their study program, the answer were positive: ”The best course I attended

Question	Number
At the beginning of the course the learning outcomes, content and assessment criteria were clearly communicated.	1.1
The course content is very interesting.	1.2
This particular course is very important in its contribution towards reaching the objectives of the academic program.	1.3
The lecturer is very motivating.	1.4
The lecturer provided a well-structured course.	2.3
The lecturer focuses on the students and their levels of knowledge.	2.4
Complex subject matter and examples were explained in a comprehensible manner.	2.5
The lecturer gave useful feedback on questions, suggestions raised and contributions made by the students.	2.6
There were enough phases of interaction and communication. (Online or on-site)	2.13
For courses held in a hybrid or digital form: The students participating online could easily join discussions.	2.15
The students participated in the course.	2.20
I have a clear idea of what to expect in the examination.	2.21

TABLE III: Standardized survey for the Data Science course.

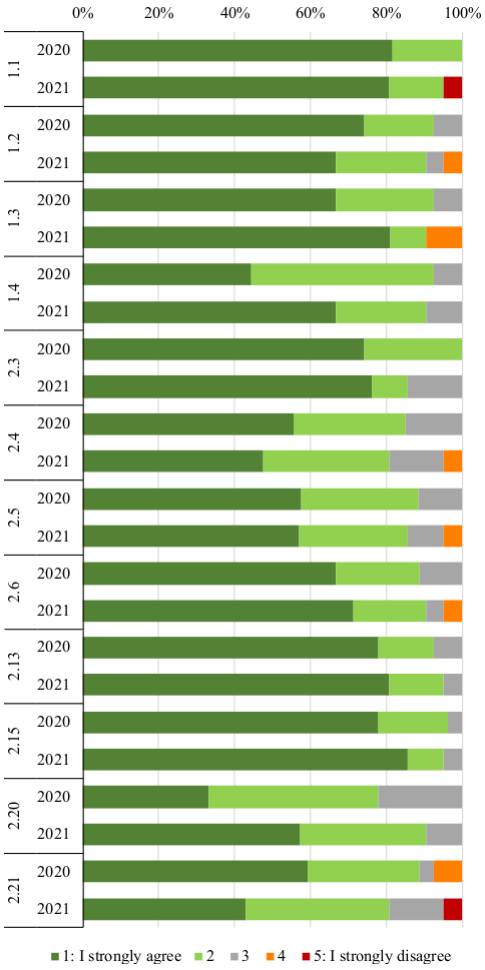

Fig. 3: Results for likert-scale questions of the standardized survey.

had almost the same structure as this one but it was focusing more on the final project and not on the examination parts.", "In terms of organization and execution, this course was the best one I have attended so far.". We notice, in their feedback, a first effect of the remote teaching: many students appreciate the possibility of having recorded lectures. This is a new opportunity offered to students and it was not exploited before.

Digital media. Even if the students reported good feedback on the material provided during the course, they accessed other websites/forums such as YouTube, Stack Overflow, and coding tutorials. The access to these resources, according to their answer, happened mainly during the project phase when they were working autonomously in groups.

Project work. The main aspects to highlight for the project work are two, the group formation and the peer review phase. The technique of forming groups with students from different study programs and backgrounds appears to be successful and is appreciated by the students who commented "I loved the group project as I would rarely work with fellow students from those very different backgrounds!...". The peer-review phase, introduced in the Winter Semester 2021, was also positively perceived by the students, but 15% of them would have preferred to receive feedback from the lecturer on the state of their project. On the project work feedback, we notice negative effects produced by the pandemic and the remote teaching. It was difficult for the students to coordinate with the other project members for online meetings, and for discussing the state of the project and how to proceed.

B. *Standardized course feedback questionnaire*

The standardized questionnaire is provided at the end of each course by the University in order to evaluate its courses.

The survey was presented to the students in both Winter Semesters, i.e., 2020 and 2021. The questions are presented in Table III and in Table IV. The questions in both surveys are the same from the two different semesters.

The results of the 27 and 21 responses are shown in Figure 3 and Figure 4. The respective response rates were 71.1% and 43.75%.

Question	Number
Overall, I would rate the course as	1.5
In comparison with the specifications of the curriculum (ECTS credits), the total workload is	2.7

TABLE IV: Standardized survey for the Data Science course.

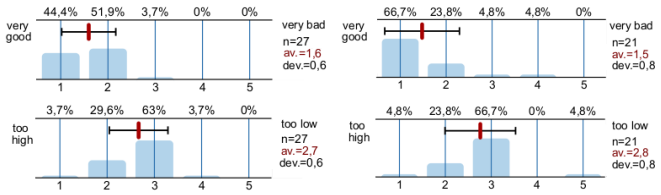

Fig. 4: Results for likert-scale questions of the standardized survey. Answer for year 2020 on the left and 2021 on the right.

The overall feedback on both semesters of the course was very positive as shown in Figure 4. The feedback is even more positive in many questions of the 2021 semester. In particular, question 2.20 was rated significantly differently in both years with p=0.02939 found using a Kruskal Wallis test. The students in 2021 felt more actively involved than in 2020. We associate this difference with the adoption of the peer review process since it was the main difference introduced in the Winter Semester 2021 compared to 2020.

Also, this survey indicates that at the beginning of the course the lecturers communicated in a clear way learning outcomes, the content, and the assessment criteria and that the lecturers provided a well-structured course focusing on the students and their levels of knowledge. Moreover, complex subject matter and examples were explained in a comprehensible manner.

This is a very important aspect of the Data Science course as it is offered in the Business Analytics, Data Science, and Digital Humanities faculties as part of the masters' programs and the students have a diverse backgrounds. It is a challenge to design a course that matches the educational background of these students. Despite this challenge, the course content was found very interesting and the course was estimated very important in contributing to reaching the objectives of the academic program.

Even if both courses were held online, the students believed that there were enough phases of interaction and, communication and that they could easily join discussions online.

C. Analysis of final grades

We collected and evaluated students' final grades in both winter semester, WS 2020 and compared them to the preceding year WS 2020. The university grading system comprises the five grades 1–5, where 1 is the best grading and 5 is the worst, which also amounts to failing a course. In Figure 5, we show the percentage of students that obtained a specific grade.

It is easy to notice that the grades for the winter semester 2021 improved compared to those of the previous year. The percentage of students who had a low grades, i.e. 4 or 5, is lower in WS 2021 while the percentage for the higher grades as 1 and 3 increased. This phenomenon could be motivated by two reasons. First, the answers to the surveys in WS 2020 were used to improve the design of the course for the following year. Second, the introduction of the peer-review process in WS 2021. This learner-centered approach drives student to be more motivated, to define their own goals, to monitor their progress autonomously, to manage their workload, decide which tasks to prioritize, to interact with each other in an open-minded and fruitful way, and to try to achieve results without constant external directions.

D. Students interviews

The students who participated in the interviews had a diverse background and 50% of them are working students. The transcripts of the interviews were analyzed and the mentioned advantages and challenges were categorized. The following listing summarizes the main findings on the relevance of the Data Science course for their future. (i) Collaboration skills: 80% of the students mention that collaborating with peers on a project is a very important skill in any environment and they acquired it thanks to this course. (ii) Real-world projects: all the interviews reveal that learning how to structure data science projects properly and how to address important questions and potential problems is very relevant for a future in the business world. (iii) Coding: for 50% of the interviewed students the challenge is to use machine learning methods without knowing coding languages but this was overcome by the use of KNIME software presented in the lectures.

V. Discussion

Several interesting significant findings were discovered. These findings confirm that the approach used in the Data Science course is viable for teaching computer science skills to non computer-scientist and can be used by other educational institutions.

The research questions could be answered as follows:

RQ1: Is the data science course provided in the masters' programs of different faculties well designed for students with different backgrounds? Yes, the surveys revealed that the Data Science course is well structured and appropriately designed for students with different knowledge gained from different study programs. With this question we analyse the general overview of the course.

RQ2: Does the course provide the master's students with diverse backgrounds with the expected and needed skills to carry out data science projects? Yes, the students acquired skills on how to handle a data science project. They are able

to assess what questions could be interesting/reasonable to answer when given a dataset, choose appropriate methods to use for answering these questions, and to communicate the results understandably and reliably. On the open-ended questions, we read that ”the course so far serves this purpose well”. Aspect: general overview of the course. With this question we analyse mainly the lectures in the first and second part of the course.

The project work and the peer reviews highlight that students acquired a deep understanding of the main topics taught in the course: they showed in the project work that they can autonomously work on a real world problem that differs from the case studies presented in the lectures, moreover, they can provide constructive feedback to their peers on topics far from their domain. Moreover, with this question we comment also on the feedback from the project work and peer review phase.

RQ3: How did emergency remote teaching affect student performance and what is the lesson learned? The effects of remote teaching are twofold. On the one hand, it facilitates the learning process by making available material that would not be available in face-to-face teaching, i.e., recordings of the lectures. On the other hand, it makes the collaboration among students more difficult, limiting the interaction in the case of project works.

RQ4: Is the software presented in the course for the project work suitable for students with heterogeneous backgrounds? For the project work, the student are free to use any programming language, such as Python or R, if they are already familiar with it, but they are also introduced to the software KNIME, presented to them in weekly tutorials. This software is easy to use and does not require programming skills therefor id suitable for the students of this Data Science course. From the answers to the questionnaires, we learn that the material provide was enough to familiarize with this tool and to be able to work on the project. With this question we analyse the project work phase.

RQ5: Is the group formation process that is adopted by the lecturers adequate for the project work? The technique used for creating the groups for the project work consists on collecting the students' preferences on the proposed dataset, and splitting the participant of the course in groups with 3 or 4 persons from different study programs. This process was positively evaluated, the students commented ”I think it was good formed. I would not like random selection of the group members. Every student has his/her personal interests that could help the Group and also bring something new. The random selection would destroy this. ”. Moreover, they highlighted that this method solves the problem they would have in finding a group themselves as they do not know the other participant of the course: ”It is difficult to find a good group if you don't know the majority of the people.”. With this question, as for the previous one, we analyse the project work phase.

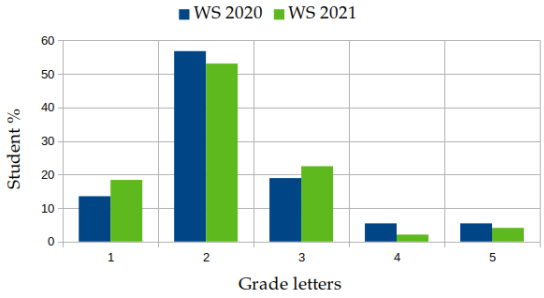

Fig. 5: Grades for WS 2020 and WS 2021.

VI. Conclusion

This paper presented our findings on the designing of a new course in Data Science and depicts the course setup. The target audiences of this paper are instructors, curriculum designers, educational researchers, and everybody who is interested in promoting the professional skills of students with different backgrounds, from computer science to digital humanities. We presented the professors' teaching approach in order to help readers imagine and potentially replicate (part of) the course. We compared the results obtained with two different anonymous questionnaires from two different semesters and we found that the students like the course structure, the presented software, and the teaching style. Moreover, the evaluation of the course and the grades of the students improved from Winter Semester 2020 to Winter Semester 2021.

We learn that (i) the course was well designed, (ii) it helped students with different backgrounds to acquire skills for handling data science projects, and (iii) the emergency remote teaching had both positive and negative effects facilitating the access to the course material and lectures but had also made the project work phase come complicated to coordinate.

We think that the structure of the Data Science course described in the paper could be used also for engineering or computer science students, maybe in a course in bachelor's programs. The topics can be presented in more detail in this case, and the use of other programming languages is also possible.

Our future work encompasses further evaluations and we are looking forward to switching to face-to-face teaching to provide our students with the essential and necessary social constructs and interactions to make their learning experience whole again. Moreover, we plan on comparing the results we have for the course held online to those of face-to-face lectures in the future semesters.

Acknowledgment

We thank our students for putting their trust in us, lecturers, and their willingness to help us improve our ways of teaching. We thank the other lecturers involved in the organization and teaching of the course: Univ.-Prof. Dr. Jan Fabian Ehmke and Steffen Elting, M.Sc.

REFERENCES

[1] S. Eybers and M. Hattingh, "Teaching data science to post graduate students: A preliminary study using a" flip" class room approach." *International Association for Development of the Information Society*, 2016.

[2] C. Dichev, D. Dicheva, L. Cassel, D. Goelman, and M. Posner, "Preparing all students for the data-driven world," in *Proceedings of the Symposium on Computing at Minority Institutions, ADMI*, vol. 346, 2016.

[3] J. Hardin, R. Hoerl, N. J. Horton, D. Nolan, B. Baumer, O. Hall-Holt, P. Murrell, R. Peng, P. Roback, D. T. Lang, and M. D. Ward, "Data science in statistics curricula: Preparing students to "think with data"," *The American Statistician*, vol. 69, no. 4, pp. 343–353, 2015. [Online]. Available: https://doi.org/10.1080/00031305.2015.1077729

[4] R. Gould, S. Machado, C. Ong, T. Johnson, J. Molyneux, S. Nolen, H. Tangmunarunkit, L. Trusela, and L. Zanontian, "Teaching data science to secondary students: The mobilize introduction to data science curriculum," *Iase-Web. Org*, 2016.

[5] C. Romero and S. Ventura, "Educational data science in massive open online courses," *WIREs Data Mining and Knowledge Discovery*, vol. 7, no. 1, p. e1187, 2017. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1187

[6] K. Rasheva-Yordanova, S. Toleva-Stoimenova, I. Kostadinova, D. Christozov, and B. Nikolova, "Assessing the entry competences of data science master program potential students," *EDULEARN19 Proceedings. Academic Press. Retrieved from https://library. iated. org/view/RASHEVAYORDANOVA2019ASS*, 2019.

[7] B. Trilling and C. Fadel, *21st Century Skills: Learning for Life in Our Times*. San Francisco, CA: Jossey-Bass, 2009, oCLC: 1014283266. [Online]. Available: http://neworleans.lib.overdrive.com/ContentDetails.htm?ID=8DE5C066-D07D-4D61-BF37-6C472DF70393

[8] E. van Laar, A. J. Deursen, J. A. Van Dijk, and J. Haan, "The relation between 21st-century skills and digital skills: A systematic literature review," *Computers in Human Behavior*, vol. 72, 03 2017.

[9] E. van Laar, A. J. van Deursen, J. A. van Dijk, and J. de Haan, "21st-century digital skills instrument aimed at working professionals: Conceptual development and empirical validation," *Telematics and Informatics*, vol. 35, no. 8, pp. 2184–2200, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0736585318305239

[10] ——, "Determinants of 21st-century digital skills: A large-scale survey among working professionals," *Computers in Human Behavior*, vol. 100, pp. 93–104, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563219302432

[11] C. Cobo, "Mechanisms to identify and study the demand for innovation skills in world-renowned organizations," *On The Horizon - The Strategic Planning Resource for Education Professionals*, vol. 21, 05 2013.

[12] H. Soulé and T. Warrick, "Defining 21st century readiness for all students: What we know and how to get there," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 9, pp. 178–186, 05 2015.

[13] C. Sin and G. Neave, "Employability deconstructed: perceptions of bologna stakeholders," *Studies in Higher Education*, vol. 41, 11 2014.

[14] N. Mohamad Arsad, K. Osman, and T. M. Tuan Soh, "Instrument development for 21st century skills in biology," *Procedia - Social and Behavioral Sciences*, vol. 15, pp. 1470–1474, 12 2011.

[15] C. R. Rogers, "A theory of therapy , personality , and interpersonal relationships , as developed in the client-centered framework," 2010.

[16] ——, *Freedom to learn: A view of what education might become*, ser. Studies of the person. Columbus, Ohio: Merrill, 1969.

[17] ——, *Freedom to Learn for the 80's*, 2nd ed. Columbus, OH: Merrill Publishing, 1983.

[18] C. R. Rogers and J. H. Freiberg, *Freedom to Learn*, 3rd ed. Columbus, OH: Merrill Publishing, 1994.

[19] A. Mohmmed, B. Khidhir, A. Nazeer, and V. Vijayan, "Emergency remote teaching during coronavirus pandemic: the current trend and future directive at middle east college oman," *Innovative Infrastructure Solutions*, vol. 5, 12 2020.

[20] O. Misirli and F. Ergulec, "Emergency remote teaching during the covid-19 pandemic: Parents experiences and perspectives," *Education and Information Technologies*, vol. 26, 03 2021.

[21] W. u. F. Bundesministerium fur Bildung, "Umgang des bildungssystems mit dem coronavirus - erlass." [Online]. Available: https://www.bmbwf.gv.at/dam/jcr:224ef638-49dc-44b1-b58e-e6420a8e11bd/corona_erlass.pdf

[22] C. Hodges, S. Moore, B. Lockee, T. Trust, and M. Bond, "The difference between emergency remote teaching and online learning," 03 2020. [Online]. Available: https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

[23] P. Bawa, "Learning in the age of sars-cov-2: A quantitative study of learners' performance in the age of emergency remote teaching," *Computers and Education Open*, vol. 1, p. 100016, 12 2021.

[24] S. Iglesias-Pradas, Ángel Hernández-García, J. Chaparro-Peláez, and J. L. Prieto, "Emergency remote teaching and students' academic performance in higher education during the covid-19 pandemic: A case study," *Computers in Human Behavior*, vol. 119, p. 106713, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563221000352

[25] T. González, M. De la Rubia, K. Hincz, M. Comas-Lopez, L. Subirats, S. Fort, and G. Sacha, "Influence of covid-19 confinement on students' performance in higher education," *PloS one*, vol. 15, p. e0239490, 10 2020.

[26] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 01 2000.

[27] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, and B. Wiswedel, "KNIME: The Konstanz Information Miner," in *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Springer, 2007.

[28] C. Osgood, G. Suci, and P. Tenenbaum, *The Measurement of meaning*. Urbana:: University of Illinois Press, 1957.

[29] D. Dolezal, A. Posekany, L. Vittori, G. Koppensteiner, and R. Motschnig, "Fostering 21st century skills in student-centered engineering education at the secondary school level: Second evaluation of the learning office," in *2019 IEEE Frontiers in Education Conference (FIE)*, 2019, pp. 1–9.