

**1. Course Code**

2261

**2. Course Title**

G51e: Data Science

**3. Teacher**

HAMIDULLAH, Sokout

**4. Term**

Spring 1

**5. Course Requirements (Courses / Knowledge for this course) and Important Information**

Basic Knowledge of Statistics with Excel and Computer Programming (But NOT Essential)

**6. Course Overview and Objectives**

Data Science has a big impact on the business landscape, which is constantly changing in today's world. As the quantity of human connection with digital footprint increases daily, an unimaginable large mass of fine-grained data is generated on a regular basis. Global internet users have climbed to 4.95 billion at the start of 2022, with internet penetration now standing at 62.5 percent of the world's total population. The data generated behind could be useful to analyze existing issues, reveal previously hidden opportunities and predict future insights. Companies will compete for hundreds of thousands of new workers needed to navigate the digital world. Already, experts estimate that millions of jobs in DATA SCIENCE might remain vacant for the lack of readily available talent.

This course is about the world of data science. In this course we will start with an introduction where we will discuss the world of data science. I explain why data science is important and how it helps businesses stay competitive in this day and age. Then continues by introducing important mathematical and statistical concepts that are essential to do data science. Once we have laid out a solid foundation, we will be ready to learn Data Analysis for Data Science. Starting from the very basics, we will build up your skills and soon you will be able to play with Exploratory Data Analysis with Python.

**7. Course Outline**

- 1 Course orientation and Introduction to Data Science
- 2 Major terminologies in Data Science
- 3 Data Science Tools and Methodology
- 4 Fundamentals of Data and Data Science
- 5 Descriptive Statistics
- 6 Exercises
- 7 Inferential Statistics
- 8 Exercises
- 9 Hypothesis testing
- 10 P-value
- 11 Data Analysis with Python
- 12 Data Analysis with Python Exercises
- 13 Exploratory Data Analysis with Python
- 14 Exploratory Data Analysis with Python Exercises
- 15 Final Presentation
- 16 Final Presentation

**8. Textbooks (Required Books for this course)****9. Reference Books (optional books for further study)**

- (1) Learning from Data: An Introduction to statistical reasoning, third edition. International Standard Book Number-13: 978-0-8058-4921-9.
- (2) Data Science from scratch, ISBN-978-1-4919-0142-7.
- (3) Available resources on Internet

10. Course Goals (Attainment Targets)						
(1)	Become familiar with foundations of Data and Data Science.					
(2)	Understanding the basic of statistics for Data Science.					
(3)	Be able to summarize a data set using descriptive statistics and Inferential statistics.					
(4)	Understand the Data Science Methodology (i) from problem to approach, (ii) working with the data, (iii) Deriving the answer.					
(5)	Be able to apply Data Analysis and Exploratory Data Analysis with Python for Data Science					
(6)	Be able to define data-intensive problems in data science and understand their underlying statistical and computational principles.					
(7)						
(8)						
11. Correspondence relationship between Educational goals and Course goals						
Educational goals of the school				Course Goals		
High level ICT skills	Basic academic skills			(1) (2)		
	Specialized knowledge and literacy			(1) (2) (3) (4)		
Human skill (Tankyu skill)	Ability to continually improve own strengths			(4) (5)		
	Ability to discover and resolve the problem in society	Problem setting		(6)		
		Hypothesis planning		(2) (3) (6)		
		Hypothesis testing		(2) (3) (6)		
		Practice		(6)		
	Fundamental Competencies for Working Persons	Ability to step forward				
		Ability to think through		(3)		
		Ability to work in a team				
Professional ethics						
12. Evaluation						
Goals	Evaluation method & point allocation					
	Examination	Quiz	Reports	Presentation	Deliverables	Other
(1)		○		○		
(2)		○			○	
(3)				○	○	
(4)		○				
(5)		○		○	○	
(6)		○		○	○	
(7)						
(8)						
Allocation		30		30	40	
13. Evaluation Criteria						
Examination						
Quiz	Every week multiple choice quizzes are used to evaluate the understanding of students and motivate them for further learning.					
Reports						
Presentation	In the final presentation, students will be asked to choose and summarize a scientific research journal paper or their own ideas related to the course main objectives and contents. The evaluation will be based on participation in the group presentation, the structure of presentation, the relevance of the argument, time management and relation to the course contents.					
Deliverables	Individual and group assignment will be assigned for the students with focus on learning goals (2,3, and 5). The evaluation will be based on how the students understand the exercises and participation.					
Other						

<b>14. Active Learning</b>		
Hourly percentage of active learning within the whole class time		70%
1	Active learning such as problem solving assignment using the knowledge and skills acquired in class.	All the time
2	Active learning such as group works and discussions.	Sometimes
3	Outcome presentations and feedbacks.	Sometimes
4	Students actively make decisions on how the class should be conducted.	Sometimes
<b>15. Notes</b>		
This course is like a journey by attending it you will benefit and might enjoy. Please bring your own computer in the class.		
<b>16. Course plan</b>		
(Notice) This plan is tentative and might be changed at the time of delivery		
Lessen 1: (Course Orientation and Introduction to Data Science)		Lecture
1. Introduction, Evaluation, Scope of the Course. 2. What is Data Science. 3. Emerging Technologies shaping the modern data		
Lessen 2: ( Major Terminologies in Data Science )		Lecture
1. What is difference: <ul style="list-style-type: none"> <li>a. Big Data</li> <li>b. Data Mining</li> <li>c. Machine Learning</li> <li>d. Deep Learning</li> <li>e. Neural Networks</li> <li>f. Generative AI</li> </ul>		
Lessen 3: Fundamentals of Data Science		Lecture
1. Data <ul style="list-style-type: none"> <li>a. Traditional</li> <li>b. Big</li> </ul> 2. Data Science <ul style="list-style-type: none"> <li>a. Business Intelligence</li> <li>b. Traditional Methods</li> <li>c. Machine Learning</li> </ul>		
Lessen 4: Data Science Tools and Methodology		Lecture
1. Tools for Data Science 2. Data Science Methodology		

Lessen 5: Statistics for Data Science	Lecture + Exercises
1. Population vs Sample 2. Types of Data	
Lessen 6: Descriptive Statistics Exercises	Lecture + Exercises
1. Measure of central tendency <ul style="list-style-type: none"> <li>a. Mean</li> <li>b. Median</li> <li>c. Mode</li> </ul> 2. Measure dispersion <ul style="list-style-type: none"> <li>a. Variance, Standard deviation</li> <li>b. Covariance, Correlation</li> <li>c. Range</li> </ul>	
Lessen 7: Inferential Statistics	Lecture + Exercises
1. Distribution 2. Confidence Interval	
Lessen 8: Inferential Statistics Exercises	Lecture + Exercises
1. Hypothesis Testing <ul style="list-style-type: none"> <li>a. Null Hypothesis</li> <li>b. Alternative Hypothesis</li> <li>c. Error in Hypothesis</li> </ul>	
Lessen 9: Hypothesis Testing	Lecture + Exercises
1. Hypothesis Testing <ul style="list-style-type: none"> <li>a. Null Hypothesis</li> <li>b. Alternative Hypothesis</li> <li>c. Error in Hypothesis</li> </ul>	
Lessen 10: P-Value + Exercises	Lecture + Exercises
1. Concept of P-value 2. Exercises with P-value	

Lessen 11: Data Analysis with Python	(90 min)
1. Data Analysis with Python <ul style="list-style-type: none"> <li>a. Python Key Packages for Data Analysis</li> <li>b. Basics of Data Analysis</li> <li>c. Pre-processing of Data with Python</li> </ul>	
Lessen 12: Data Analysis with Python Exercises	Lecture + Exercises
1. Data Analysis with Python Exercises <ul style="list-style-type: none"> <li>a. Explore the dataset</li> <li>b. Deal with Missing data</li> <li>c. Formating data</li> <li>d. Data Standardization</li> <li>e. Data Normalization</li> <li>f. Data Bining</li> <li>g. Labeling data</li> </ul>	
Lessen 13: Exploratory Data Analysis with Python	Lecture + Exercises
1. Exploratory Data Analysis with Python <ul style="list-style-type: none"> <li>a. Descriptive Statistics with Python</li> <li>b. GroupBy</li> <li>c. ANOVA Test</li> <li>d. Correlation</li> </ul>	
Lessen 14: (Data Analysis with Python Exercises)	Lecture + Exercises
1. Exploratory Data Analysis with Python Exercises	
Lessen 15: (Presentation by Students)	Presentation (90 min)
Group Presentation by students	
Lessen 16: (Presentation by Students)	Presentation (90 min)
Group Presentation by students	