Introduction to Data Science

1. **Course number and name:** 020ISDES3/020IDSES3 Introduction to Data science

2. **Credits and contact hours:** 4 ECTS credits, 2x1.15 hours

3. **Name of course coordinator:** Bassel Nasr

4. **Instructional Materials:** PowerPoint slides; videos; course handouts; in class exercises

   References:
   – https://learn.udacity.com/courses/ud359
   – https://www.simplilearn.com/data-science-free-course-for-beginners-skillup?tag=data%20sciences

5. **Specific course information**
   a. **Catalog description:**
      This course offers a practical introduction to the core concepts, tools, and workflows of data science. Students will explore the typical steps in a data scientist's workflow, including data acquisition, wrangling, analysis, modeling, and visualization. Key programming tools such as NumPy and pandas are introduced for handling structured data, along with SQL (SQLite, pandasql) and APIs for data access. Students will learn data cleaning techniques such as type handling, outlier detection, partial deletion, and imputation. The course covers exploratory data analysis (EDA) using statistical significance tests and introduces both parametric and non-parametric testing methods, including t-tests, Welch's test, the Shapiro-Wilk test, and the Mann–Whitney U test. Fundamental machine learning techniques are introduced, with a focus on linear regression, gradient descent, and model evaluation using the coefficient of determination. Students will also study the principles of effective data visualization, including visual cues, coordinate systems, scaling, data types, and contextualization, with practical plotting exercises using Python. Finally, the course offers an introduction to big data concepts, including the basics of MapReduce and Hadoop, and integrates hands-on experience using Jupyter Notebooks.

   b. **Prerequisites:** None

   c. **Selected Elective** for CCE students

6. **Educational objectives for the course**
   a. **Specific outcomes of instruction:**
      – Learn what data scientists do, receive an introduction to Python's most common data science libraries, and get hands on practice working with data.
      – Access, process, and clean different types of data.

- Get an introduction to important data analysis tools used by data scientists, including statistical tests and machine learning models.
- Learn when to use different plot types, how to encode different types of data, and get practical advice on how to visually represent data in meaningful ways.
- Learn how to work with massive amounts of data using MapReduce and other big data tools.

b. PI addressed by the course:

PI	1.1	1.2	1.3	2.1	2.2	2.3	2.4	2.5	5.2	6.3	6.4	7.1
Covered	x	x	x	x	x	x	x	x	x	x	x	x
Assessed	x		x	x	x	x			x	x	x	

7. **Brief list of topics to be covered**
   - Introduction: Learn what data scientists do, receive an introduction to Python's most common data science libraries, and get hands on practice working with data Practice using NumPy and Pandas to make predictions (2 lectures)
   - Data Wrangling: Learn to access, process, and clean different types of data Practice data wrangling skills with some datasets (3 lectures)
   - Data Analysis: Get an introduction to important data analysis tools used by data scientists, including statistical tests and machine learning models Apply statistics and machine learning to some datasets (3 lectures)
   - Data Visualization: Learn when to use different plot types, how to encode different types of data, and get practical advice on how to visually represent data in meaningful ways Apply data visualization techniques to explore and explain some datasets (2 lecture)
   - MapReduce: Learn how to work with massive amounts of data using MapReduce Practice using MapReduce to work with some datasets (2 lectures)