Introduction to Data Science

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

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

3. **Name(s) of instructor(s) or course coordinator(s):** 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:**
      Introduction to Data Science: introduction: data scientists work steps - numpy, pandas – data acquisition , data wrangling : data formats - pandasql, SQLite - Api, data checking, data preparation, partial deletion, imputation – exploratory data analysis : statistical significance tests, statistical rigor, t-tests , normal distribution, welch's t-test, non-normal data, Shapiro-wilk test, Mann-whitney u test, non-parametric tests, machine learning, linear regression , gradient descent, coefficient of determination – data visualization: information visualization, components of effective visualization: visual cues, coordinate systems, scale and data types, context- visualization time series data, plotting in python – big data: basics of MapReduce: Hadoop – implementation : Notebook Jupyter

   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)