LEADING Boot Camp (2022):
WEEK 1

Monday, June 6, 2022	
10:30 - 11:00 AM EDT	**Welcome and Introduction (Professor Jane Greenberg/Sam Grabus, LEADING Project Manager, doctoral candidate)**
11:00 - 12:15 PM EDT	**Data Science Overview (Professor Jake Williams)**
	• What is data science and who does it?
	• Data science life cycles
	• Introduction to common tools used in data science
12:15 - 1:15 PM EDT	BREAK
1:15 - 3:30 PM EDT	**Data Management: SQL and NoSQL (Professor Il-Yeol Song)**
	• Understand SQL as a database language
	• Use SQL for data science
	• Understand differences between SQL and NoSQL
Wednesday, June 8, 2022	
10:30 - 12:15 PM EDT	**Data Cleaning and Preprocessing (Professor Shadi Rezapour)**
	• Dealing with messy data
	• Creating a data cleaning pipeline

	• Data transformation
12:15 - 1:15 PM EDT	BREAK
1:15 - 3:30 PM EDT	**Text Processing / Mining / Search (Professor Mat Kelly)**
	• Normalization, stop words, stemming, n-grams
	• Information retrieval (IR) methods
	• Similarity evaluation

LEADING Boot Camp: WEEK 2	
Monday, June 12, 2022	
10:30 - 12:15 PM EDT	**Supervised Learning (Professor Lei Wang)**
	• Overview of Machine Learning
	• Frequently utilized supervised learning methods
	• Evaluation of the supervised learning
12:15 - 1:15 PM EDT	BREAK
1:15 - 2:15 PM EDT	**Big Metadata, Ontologies, Linked Open Data (Professor Jane Greenberg)**
	• Metadata to big metadata
	• Ontologies and Protege
	• Linked open data

2:15 - 2:30 PM EDT	BREAK
2:30 - 3:30 PM EDT	**Data Preservation /Web Archiving (Professor Mat Kelly)**
	• Archival crawling
	• Access and replay
	• Formats and metadata
Wednesday, June 14, 2022	
10:30 - 11:30 AM EDT	**Data Curation (Professor Alex Poole)**
	• The data lifecycle
	• Data curation principles
	• Data curation practices
11:30 - 12:15 PM EDT	**Icebreaker Session: Get to know your cohort (Professor Jane Greenberg, Sam Grabus)**
12:15 - 1:15 PM EDT	BREAK
1:15 - 3:30 PM EDT	**Unsupervised Learning (Professor Shadi Rezapour)**
	• Intro to unsupervised learning
	• Clustering methods
	• Kmeans

Additional Asynchronous Session:

Computational thinking, Data Integration and Quality with OpenRefine

Guest: Richard Marciano, Professor and Director, AI-Collaboratory, University of Maryland iSchool
Computational thinking model
Data integration and data quality
Working with Open Refine

LEADING Boot Camp: WEEK 3	
Monday, June 19, 2022	
10:30 - 12:15 PM EDT	**Deep Learning (Professor Lei Wang)**
	• Introduction to Artificial Neural Network
	• Implementation of deep learning with Keras
12:15 - 1:15 PM EDT	BREAK
1:15 - 3:30 PM EDT	**Large-Scale Data Intensive Computing (Professor Weimao Ke)**
	• Big data volumes and velocity
	• Data-intensive processing and parallel computing
	• Spark Machine Learning libraries (MLlib) with Python (PySpark)
Wednesday, June 21	
10:30 - 12:15 PM EDT	**Graph/Network Models and Matrix Processing (Professor Jake Williams)**

	• What are graph/network models?
	• Sources of graph/network data and an introduction into types of models
	• Tools used for representing and analyzing graph/network data
12:15 - 1:15 PM EDT	BREAK
1:15 - 2:15 PM EDT	**Human-Centered Data Science and Data Ethics (Professor Alex Poole)**
	• Defining human-centeredness
	• The need for HCDS
	• Ethics for human-centered data science
2:15 - 2:30 PM EDT	BREAK
2:30 - 3:15 PM EDT	**Visual Analytics (Professor Erjia Yan)**
	• Fundamentals of visual analytics
	• Misuse of visualizations
	• Best practice of visualizing data