

Workshop Title	Workshop for Educators - Designing and Delivering Data Science Courses at Scale
Duration	2 Days
Target Audience	Data Science Educators

Description
<p>Which COVID-19 vaccine (Pfizer or Moderna) is more effective at preventing infections in children? Do chocolates lower blood pressure? How can we improve ChatGPT to be more human-like? Answering such data-driven questions in a reliable and convincing manner is a non-trivial exercise that requires a careful understanding of how data is generated, collected, processed, analyzed, and evaluated. The systematic study of these steps, which involves both computational and inferential thinking, has come to be known as data science. Data science is now seeing profound applications in diverse fields including computer science, economics, development, business, education, engineering and law.</p> <p>This workshop is designed for educators (e.g., faculty, trainers) who have an interest in teaching and delivering data science courses/programs. Through a careful selection of case studies, datasets, and Python notebooks, this workshop will provide a pedagogical journey of teaching data science at scale. Based on experiences with teaching data science at LUMS, this workshop aims to share a rich set of perspectives about developing data science courses grounded in sound data design, exploratory data science and visualization, causal inference, statistical inference, machine learning, big data analysis, and data ethics.</p> <p>The workshop will build on the conceptualization of data science as creative problem solving, with a focus on project-based learning. Through examples and case studies, the workshop aims to highlight the benefits of incorporating societally relevant and practical applications of data science in courses to cater to diverse audiences as well as to prepare students for impactful interdisciplinary work. Finally, in this workshop we will explore ways to scale data science courses while ensuring sound hands-on and classroom learning.</p>

Learning Outcomes	
	By the end of this workshop, the participants should be able to:
LO1:	Conduct and teach sound data analysis using public datasets and appropriate examples and activities

LO2:	Develop strategies for integrating data science into existing curricula across different subject areas
LO3:	Describe a given dataset and assess its quality
LO4:	Build data pipelines (collection, cleaning, EDA, modelling, evaluation, results) for “repeatable” work
LO5:	Use Python and popular data science libraries such as Pandas, Numpy, Matplotlib, and Scikit-learn to analyze and visualize data.
LO6:	Understand the theory behind drawing inferences from data
LO7:	Communicate results effectively

Summary	
Key Concepts/Topics Covered	Timestamp
<p>[Understanding the Data Science Process] <i>What is data science? Why is it important? What are different pedagogical approaches for teaching data science?</i></p> <ul style="list-style-type: none"> • Uncovering the various shades of data • Using data as a powerful communication tool • Conceptualizing data science as creative problem solving coupled with project-based learning • Case Studies: Plastic Bottles & Waste, Education & Income, Sentiment Analysis 	Session I
<p>[Data Design, Descriptive Statistics, and Data Manipulation Using Python] <i>What are some techniques for sound data design? How can we effectively summarize data?</i></p> <ul style="list-style-type: none"> • Statistics of data and deceptive descriptions • Data design: non-probability and probability samples • Case Studies: COVID-19 Cases in Pakistan, Deworming, Predicting Election Outcomes 	Session II
<p>[Data Cleaning, Exploratory Data Analysis, and Visualisation] <i>How can we transform data for ease of analysis? What are some common data anomalies? How do we systematically conduct data explorations?</i></p> <ul style="list-style-type: none"> • Principles of exploratory data analysis and visualisation • Analysing text data • Case Study: Crime Analytics 	Session III
[Experiments, Causal Inference and Foundations of Statistical Inference]	Session IV

<p><i>How do we establish if the relationship between variable A (e.g., smoking) and B (e.g., cancer) is causal? What is the uncertainty in our data?</i></p> <ul style="list-style-type: none"> • Experiments and Observational Studies • Impact Evaluations and Evidence-based Policy Making • Hypothesis Testing, Confidence Intervals, and p-values • Case Study: Enabling Innovations at Ride-Sharing Companies 	
<p>[Machine Learning]</p> <p><i>How can we build machine learning models to predict Y (e.g., house prices) and X (e.g., size of house, city, town)? Will my predictions generalize to unseen scenarios?</i></p> <ul style="list-style-type: none"> • Simple and Multiple Linear Regression • Feature Engineering and Bias-Variance Tradeoff • Cross Validation and Regularization • Classification, Logistic Regression, Decision Trees and Random Forests • Case Study: Countering Fake News 	Session V
<p>[Big Data Analytics & Data Ethics]</p> <p><i>What is Big Data? What are Ethics? Why should we care about Ethics of Data?</i></p> <ul style="list-style-type: none"> • MapReduce and Spark • Ethics and the Data Science Lifecycle • Bias and Fairness Criterias • Case Studies: Gender Shades, Ford, Predicting Future Criminals, Conducting Ethical Experiments 	Session VI

Reading Material
Will be shared at the workshop.