



LUMS<sup>X</sup>

# INTRODUCTION TO DATA SCIENCE

Dr. Ihsan Ayub Qazi

LUMSx is the center for online learning and professional development at LUMS. We extend LUMS' excellence in teaching and research beyond its borders by leveraging technology and innovative pedagogy. Our courses aim to bridge critical knowledge & skill gaps for Pakistani learners and to meet their diverse learning needs, we offer **Massive Open Online Courses (MOOCs)**, **Hybrid Courses**, **Synchronous (Live) Courses**, and **Free Open Online Courses (OpenCourseWare)**. We intend to harness technology for enhancing access, improving educational quality, and amplifying education's impact.



**Course Format:** Hybrid

**Language:** English, Urdu

**Starting Date:** 20th September, 2024

**Registration Deadline:** 13th September, 2024

**Duration:** 3 months

**Price:** PKR. 25,000

# DATA SCIENCE SPECIALIZATION



The **Introduction to Data Science Course** is also part of the LUMSx Data Science Specialization:

Our well-curated Data Science Specialization is designed to equip you with the technical expertise and practical skills needed to launch a successful career in this dynamic domain. You will not only gain technical knowledge through core courses Introduction to Data Science and Machine Learning but also practice through hands-on assignments and develop key non-technical skills that will enable you to thrive in any industry relevant to Data Science.

## Additional Features in the Specialization:

- Take part in both courses, Machine Learning and Introduction to Data Science for a **Discounted Price**.
- Upskill yourself through **optional add-ons** like Project Management and Business Communication at discounted prices.
- Earn both a specialization and individual course **certificates**.
- Gain a balance of technical and non-technical skills.
- Top 10% of students in the specialization will get an **Internship Opportunity**.

[LEARN MORE](#)

# ABOUT THIS COURSE



Are you curious about the world of data science, but not sure where to start? Do you want to unlock the secrets of data analysis and drive informed decisions? Are you looking to upskill and take your career to the next level? Whether you're here to upskill, gain insights into potential career pathways, or develop foundational expertise, we're thrilled for you to join our course!

Embark on an exhilarating journey into the heart of data science, where insights and innovations converge to shape the future! Are you ready to unravel the mysteries hidden within the vast realms of data?

At LUMSx, we believe in empowering you, our learners with the ever-evolving field of Data Science, where we will lay the foundation of descriptive statistics, and solve the intricacies of data biases. We'll then dive into the art of statistical inference and machine learning, where hypotheses are tested.

Together, we'll understand models and algorithms, exploring concepts like regression and classification, as we try to unlock the fundamental challenges of learning. As our journey nears its end, we'll equip ourselves with the tools and ethics needed to navigate scalable data collection and processing.

This is a cohort-based course, meaning you'll attend alongside a group of peers with a fixed start and end date, allowing for real-time interaction with the course team during live sessions. The course includes live tutorials, interactive sessions with the instructor as well as industry experts in the field of Data Science.

So, what are you waiting for? Let's get started!

# MEET YOUR INSTRUCTOR



Course Instructor

**Dr. Ihsan Ayub Qazi**

*Associate Professor, Computer Science,  
LUMS*

Dr. Ihsan Ayyub Qazi is an Associate Professor in Computer Science at LUMS, Pakistan. He earned his Ph.D. from the University of Pittsburgh in 2010 and a BSc. (Hons) from LUMS in 2005.

He was a Visiting Research Scientist at UC Berkeley in 2017. His research focuses on networked systems, including cloud computing, mobile networks, and online privacy. His work has appeared in ACM SIGCOMM and IEEE/ACM Transactions on Networking. He received the best ACM SIGCOMM CCR paper award in 2018 and multiple research awards from Google and Facebook. His research is funded by NSF, World Bank, ARC, and others.



# COURSE OUTLINE

## **Module 1: Untangling the Data Science Process**

In this module, you will get an introduction to the field of data science, the essence of data, and the importance of the data science lifecycle which covers the essential stages involved in any data science project. You will discover why data science is essential in today's data-driven world, learn how to extract meaningful insights through data analysis, and get a sneak peek of the exciting course modules that lie ahead. Whether you're new to the subject or looking to refresh your knowledge, this module is your gateway to unlocking the power of data in making informed decisions and sparking innovation.

## **Module 2: Descriptive Statistics, Data Acquisition, and Tools**

This module equips you with the skills to identify and address biases in data analysis and utilizes the Python for Data Science (Pandas) library for efficient data processing, setting a solid foundation for your data science journey. In particular, you will learn about summarizing and interpreting large amounts of data using descriptive statistics to understand measures of centrality and variability in datasets. Furthermore, you'll delve into effective data design techniques, incorporating sampling strategies and survey designs, to minimize bias in data science projects, complemented by practical case studies. Finally, you will be introduced to Pandas. You'll learn how to read data into DataFrame structures, how to query these structures, and how to index, merge, and group data effectively.

## **Module 3: Live Tutorial**

This module consists of one live session with our Research Assistant, one data assignment and one session with the instructor. The live tutorial will provide you with hands-on experience under the guidance of our Research Assistant. During this session the research assistant will conduct a programming demonstration. You will get a chance to interact with the course team and your peers through this live online session. The data assignment in this module is meant to reinforce practice and evaluate understanding of key programming concepts. The live session with the instructor at the end will aim to answer any course-related questions you might have, under the guidance of Dr. Ihsan Qazi.

## **Module 4: Data Cleaning, Exploratory Data Analysis & Visualization**

In this module, you'll explore the essentials of Data Cleaning, Exploratory Data Analysis (EDA), Data Visualization, Text Analysis and Relational Databases. You will learn how to prepare and examine data to reveal hidden insights. You will learn about the principles of data visualization and how to articulate data narratives visually. You'll also advance your skills with data transformations and text analysis. Finally, you will be introduced to basics of relational databases and the Structured Query Language (SQL), focusing on schemas, joins, and sampling. This streamlined journey equips you with crucial skills for insightful data analysis and management.

### **Module 5: Live Tutorial**

This module consists of one live session and one data assignment. The live tutorial will provide you with hands-on experience under the guidance of our Research Assistant. During this session the research assistant will conduct a programming demonstration. You will get a chance to interact with the course team and your peers through this live online session. The data assignment in this module is meant to reinforce practice and evaluate understanding of key programming concepts.

### **Module 6: Experiments, Causality, and Foundations of Statistical Interference**

This module offers a thorough introduction to data analysis techniques and research methodologies for making informed, data-driven decisions. In particular, you will learn about Causal Inference and Statistical Inference to uncover causal relationships between variables and assess the reliability of data. You will learn about randomized control trials, quasi-experimental methods, and evidence-based policymaking. You will learn about hypothesis testing, p-values, and model assessment. Practical exercises on A/B testing and bootstrap sampling will enhance your ability to compare data confidently. Additionally, you will grasp the importance of sample size and power in data studies.

This module also contains one session with the industry expert where you will discover how Data Science is transforming various industries and is shaping future trends. You will also gain practical insights into how Data Science drives real-world solutions. The live session with the instructor at the end will aim to answer any course-related questions you might have, under the guidance of Dr. Ihsan Qazi.

### **Module 7: Live Tutorial**

This module consists of one live session with our Research Assistant and one data assignment. The live tutorial will provide you with hands-on experience under the guidance of our Research Assistant. During this session the research assistant will conduct a programming demonstration. You will get a chance to interact with the course team and your peers through this live online session. The data assignment in this module is meant to reinforce practice and evaluate understanding of key programming concepts.

### **Module 8: Predictive Analytics and Machine Learning**

In this module, you'll embark on a journey through the fundamentals of Machine Learning and Predictive Analytics, starting with an understanding of what models are and diving into the modeling process and loss functions. You will explore regression techniques, from simple to multiple linear regression, and learn how to interpret models while understanding their accuracy through several metrics. Finally, you will delve into classification methods, focusing on logistic regression and the classification metrics.



### **Module 9: Live Tutorial**

This module consists of one live session with our Research Assistant, one data assignment and one session with the instructor. The live tutorial will provide you with hands-on experience under the guidance of our Research Assistant. During this session the research assistant will conduct a programming demonstration. You will get a chance to interact with the course team and your peers through this live online session. The data assignment in this module is meant to reinforce practice and evaluate understanding of key programming concepts. The live session with the instructor at the end will aim to answer any course-related questions you might have, under the guidance of Dr. Ihsan Qazi.

### **Module 10: Big Data and Ethics**

In this module, you will learn about Big Data, starting with its sources and the pivotal role of distributed file systems in managing it. You will be introduced to the essence of MapReduce data processing framework, with a special focus on Apache Spark, to perform large-scale data processing efficiently. In addition, you will learn about the importance of ethics in data science. You will explore how ethical considerations are integral throughout the data science lifecycle, and the pursuit of fairness in machine learning. This module equips you with a deep understanding of both the technical and ethical foundations necessary for responsible and effective data science practice.



# INTRODUCTION TO DATA SCIENCE

[ENROLL NOW](#)

As an **online learning gateway** of LUMS, we aim to extend LUMS' excellence in teaching and research **beyond its borders** by leveraging technology.