

DATA 101 : INTRODUCTION TO DATA SCIENCE

Transcript title

Introduction to Data Science

Credits

4

Grading mode

Standard letter grades

Total contact hours

40

Lecture hours

40

Prerequisites

MTH 095 or higher (except MTH 098, MTH 102) or minimum placement Math Level 18.

Recommended preparation

MTH 105Z.

Course Description

Introduces students to the essential concepts of data science, emphasizing the critical thinking and analytical skills required to work with data. Develops ability to tell a story from real-world data sets, using key computational and quantitative techniques such as spreadsheets, programming and statistical inference. Engages with diverse data from various domains, applying data preparation, analysis and visual storytelling methods. Explores the ethical implications of data collection, decision-making driven by data, and privacy concerns, developing the skills to communicate these issues responsibly and effectively.

Course learning outcomes

1. Analyze data context by describing and evaluating background, including observational units, variables, their types and source.
2. Use data wrangling techniques to transform and filter data into a useable format in various environments including Structured Query Language (SQL), Python and Spreadsheets.
3. Perform exploratory data analysis within a programming environment through the use of summary statistics, including measures of central tendency and variability.
4. Create and test models within a programming environment through concepts of statistical inference, including sampling and simulation.
5. Articulate results by clearly communicating findings derived from data using graphical techniques for visualizing one, two, three and four variables simultaneously.
6. Explore ethical implications of data collection, data-driven decision making, data sharing, and privacy.

Content outline

1. Introduction to Data Science
 - a. Data science as an integral part of daily life, shaping decision-making and innovation in countless industries including business, healthcare, finance, government, and technology.
 - b. The data science workflow: data preparation, data analysis, and data storytelling.
 - c. Ethical risks in Data Science (will also be discussed throughout the course).
2. Data Wrangling
 - a. The anatomy of data as consisting of observations, variables, and values and distinguishing between different data types.
 - b. Data Preprocessing: transforming raw data into a consistent, well-organized, and structured format to facilitates effective analysis and storytelling using SQL and Python:
 - i. Applying strategies to address missing data
 - ii. Finding and addressing anomalous data
 - iii. Selecting, filtering, ordering, summarizing and merging data
3. Data Visualization (which will be ubiquitous throughout the course)
 - a. Introduce graphical techniques to visualize data:
 - i. Histograms and density plots.
 - ii. Box and violin plots.
 - iii. Scatterplots, line and bubble charts
4. Exploratory Data Analysis
 - a. Introduce numerical metrics that highlight important features of a dataset such as central tendency and distribution.
 - b. Computing and visualizing correlations to find relationships between variables and identify features.
 - c. Assessing variability and Identifying outliers.
5. Introduction to Statistical Inference and Machine Learning
 - a. Data sampling from a model or distribution using Monte Carlo simulations.
 - b. Assessing uncertainty and drawing conclusions about a population based on a sample of data:
 - i. Hypothesis testing: Null and Alternative hypothesis, t-test, Analysis of Variance
 - ii. Regression models: linear and logistic
 - iii. Clustering techniques: supervised and unsupervised
 - iv. Decision trees
6. Decision making and data storytelling

Required materials

Students are required to purchase an e-textbook. Students will also need access to a computer with internet for completing assignments.

General education/Related instruction lists

- Science not Lab