

Proposal for a New Course  
Department of Intelligent Systems  
Wadhvani School of AI and Intelligent Systems  
Indian Institute of Technology, Kanpur

**Course number:** ESOXXX

**Course title:** Fundamentals of Data Engineering – Part B

**Course prerequisites:** None (although familiarity with Python programming is essential)

**Course credits:** [7] (3-1-3-0)

**Course duration:** Modular (2nd half)

**Course type:** ESO (for all UG students)

**Proposing instructor:** Indranil Saha

**Others interested in teaching the course:** Purushottam Kar, Sruti Srinivasa Ragavan, Nisheeth Srivastava, and all DIS faculty members

**Course description:**

- a. *Objectives:* Proper data handling is essential for AI, ML, and data science. This course will introduce data collection, curation, and comprehension using Python libraries. The course will also offer hands-on lab experience involving digital data sources.
- b. *Content:* There will be an equivalent of 21 lectures of 50 minutes each and 7 labs of 3 hours each. A weekly breakup of lecture and lab content is given below.
- c. *Evaluation:* Evaluation will use a combination of graded lab exercises, lab exams, take-home assignments and projects, and traditional sit-down quizzes and exams.

**Lab equipment:** Sufficiently many PCs with internet connectivity, and appropriate software (browsers, IDEs such as VSCode, and Python + IPYNB runtimes with libraries) will be needed.

**Logistics:** The course will be offered in the first semester of every year. However, due to the lack of teaching resources, departments opting for this course would be required to supply the following:

- a. **Liaisons (1 liaison per department):** The liaison can be the DUGC convenor or their nominee. The liaison will ensure the availability of tutors and assistants. The liaison will be the single point of contact between the course instructor and the department.
- b. **Tutors (1 tutor per 50 students):** Tutors may be faculty members, PG students, or 3rd/4th year UG students. It is proposed that the course be granted permission to take UG students as tutors. Tutors need not have credited this course, but must have prior Python experience and be willing to learn the course contents. Tutors will conduct a weekly tutorial, help organize (e.g., set questions), and conduct weekly labs, quizzes, and examinations.
- c. **Assistants (1 assistant per 25 students):** Assistants may be PG students. Assistants will invigilate, grade weekly labs, quizzes, and exams.

**Weekly breakup of content:** Numbers in square brackets [] against each topic indicate the number of lectures/labs for that topic.

*Lecture content (21 lectures):*

1. Data storage, processing [6]:
  - a. *Formats:* a selection from CSV, Excel, parquet, SQLite, JSON, sockets, etc
  - b. *Processing:* ETL (extract, transform, load), data scraping, data preprocessing
2. Data visualization and analysis [9]:
  - a. *Visualization:* plot types (line, scatter, bar, stacked), error bars and confidence intervals, aesthetics, accessibility (colorblind-friendly color schemes, alt-text)
  - b. *Analysis:* ML primitives (e.g., classification, regression, clustering, dimensionality reduction, anomaly detection) -- brief overview only
  - c. Use of Python libraries (numpy, scipy, pandas, openpyxl, sqlite3, sklearn, matplotlib, seaborn, plotly, bokeh) to implement the above
3. Data-based app dev [6]: pipelines, data transformers, trigger/event-based programming, FastAPI, and JavaScript tools to build web apps for data visualization, data-enabled services, etc.

*Lab content (7 labs):*

1. Data storage, processing, visualization, and analysis [4]
  - a. Perform data cleanup, missing data imputation, visualization on static data from a spreadsheet or SQL database, or live data from an online source
  - b. Implement supervised and unsupervised data analysis on cleaned data
2. App development [3]
  - a. Build a web app offering ML-as-a-service
  - b. Build a live face detection system, piping camera feed into a pretrained AI model

**Short summary for inclusion in the Courses of Study booklet:** the course will introduce Python-based techniques used to collect, curate, and comprehend data.

**Textbook:** There is no textbook for this course.

**Course proposer:**

**Date:**

**Convener DUGC:**

**Date:**

The course is approved/not approved

**Chairperson, SUGC**

**Date:**