Most Asked Questions

Lecture: Exploratory Data with Pandas
Lecture date: 04-08-2025

✅ Top 10 Technical Queries:

1. **Query:** What is the role of AI/ML in the process for front end to back end to database?
   Answer: AI/ML primarily operates on backend data—optimizing data handling, predictions, and automation. It enhances front-end personalization and decision-making using backend databases and models.

2. **Query:** Don't other programming languages like C++ and Java have similar data structures?
   Answer: Yes, all programming languages implement core data structures like arrays, lists, and trees, but their syntax and built-in library support may differ.

3. **Query:** Matrix can also be n-dimensional right? Or is it 2D only?
   Answer: Yes, in Python using NumPy, arrays can be n-dimensional (nD), not limited to 2D.

4. **Query:** Is there any other structure to store millions of row or column data?
   Answer: Yes, tools like NumPy arrays, Pandas DataFrames, and PySpark DataFrames are used depending on volume and structure of data.

5. **Query:** What is `cv2` and why do we import it?
   Answer: `cv2` is the OpenCV module in Python used for image and video processing. It helps with reading, manipulating, and analyzing visual content.

6. **Query:** Why are images in Python stored as arrays?
   Answer: Images are essentially a grid of pixels, which can be represented numerically. Hence, arrays are a natural way to store them for processing.

7. **Query:** What are decoder libraries and can you give examples?
   Answer: Decoder libraries convert encoded media formats (audio, video) into readable formats. Examples: ffmpeg, libavcodec, OpenCV for image/video decoding.

8. **Query:** Why use NumPy when we have Pandas?
   Answer: NumPy is faster and supports nD data. Pandas is built on NumPy but is better for tabular (structured) data manipulation with labeled columns.

9. **Query:** Can we use NumPy for nD while Pandas is limited to 1D or 2D?
   Answer: Correct. NumPy supports n-dimensional arrays; Pandas is limited to 1D (Series) and 2D (DataFrames).

10. **Query:** How do we handle missing values in DataFrame?
    Answer: In Pandas, use methods like df.fillna(), df.dropna() to handle missing values based on use case.

11. **Query:** Once we learn entire Python, are we going to make codes live?
    Answer: Yes, in later modules, you'll be working on real-time projects and deploying code.

12. **Query:** Can we get 10 AI tools every week to explore?
    Answer: This suggestion can be shared with your instructor. Meanwhile, you can explore curated lists on platforms like Futurepedia.

13. **Query:** Where can we learn and practice Data Science visually?
    Answer: Platforms like Kaggle, Google Colab, and Microsoft Azure Notebooks offer a visual interface for hands-on data science learning.

14. **Query:** Where do I find problems to practice Pandas and NumPy?
    Answer: Recommended platforms: Kaggle, HackerRank (Python/Data Science tracks), LeetCode (Data-focused questions), and the Masai LMS.

15. **Query:** Upto how much length of data is it preferable to use Pandas before switching to PySpark?
    Answer: Pandas is efficient for datasets up to 1–2 million rows. Beyond that, PySpark is more suitable for distributed data processing.

16. **Query:** How do we optimize Pandas code for performance?
    Answer: Use vectorized operations, avoid loops, use `df.query()` or `df.loc[]`, and consider chunking large files.

17. **Query:** What is the difference between a library and a framework?
    Answer: A library offers specific functionality (you control the flow). A framework provides structure and controls the flow (you fill in parts).

✅ Top 10 Non-Technical Queries:

1. **Query: I don't have access to this particular OneNote page which sir is sharing.** Can you please share it after today's class? Answer: All session materials (PPTs, PDFs, OneNote, code files) will be shared via the LMS within 24 working hours. Please check under the same session ID. If not available, raise a ticket.

2. **Query:** Recording of last lecture is missing. Is it not uploaded yet?
   Answer: Lecture recordings are typically uploaded within 24 working hours. Please check again after that time.

3. **Query:** How to login to Google Colab?
   Answer: Visit https://colab.research.google.com, sign in with Google, and start coding in Python.