

CASE STUDY:

"Decoding Your Future: Data Storytelling in the Dynamic Job Market"

1. Background

As you prepare to graduate, you stand at the crossroads of a fast-evolving world of work. Automation, remote/hybrid policies, global talent movement, and shifting industry needs create both uncertainty and opportunity.

Decision-makers (students, universities, employers) need to understand:

- Where are opportunities growing?
- Which job roles and skills are in increasing demand?
- How do trends differ across industries, locations, and time?

You are a data storyteller, tasked with answering any (or several) of these questions. The *true challenge* is that the answers depend entirely on your data choices, analytical direction, and narrative skills.

2. Problem Statement

There is no single, correct answer.

Your mission: Use real-world job market and skills data to discover a story that helps graduates navigate their early careers.

- Pick a topic or angle you find most relevant (e.g., "Tech jobs in India," "How AI is changing marketing careers," "The future of green jobs worldwide," or "Graduate skill gaps in my city").
- Use Python for at least part of your data analysis and wrangling.
- You're encouraged to combine multiple data types, sources, and tools (Excel, Tableau, APIs, web scraping, NLP, etc.).
- The only requirement: **Justify your approach, showcase a clean and reproducible workflow, and tell a compelling story using data.**

3. Data and Tools Guidance

Data Sources

(Students may pick any, but examples include)

- **Government Portals:** Indian NCS, US BLS, UK ONS, Eurostat, etc.
- **Online Job Boards/APIs:** LinkedIn, Glassdoor, Naukri, Monster, Indeed (many have monthly reports, limited APIs, or allow scraping).
- **Open Data Portals:** [Data.gov.in](https://data.gov.in), Kaggle datasets (e.g., job posts, skill trends)
- **Survey Providers & Social Media:** Reddit/r/career guidance, Twitter, Stack Overflow Developer Survey, Coursera/Skill share trends.

Data Types

- **Structured:** CSV, Excel tables from job boards, API results, report tables
- **Unstructured:** Free text job descriptions, social media discussions, skills in resumes
- **Formats:** CSV, JSON, XML, API, Excel

Tools

- **Must Use:** Python (pandas for dataframes, optional: matplotlib/seaborn/plotly for visualization, spaCy/NLTK for text mining)
- **Can Also Use:** Excel/Google Sheets for quick table work, Tableau/Power BI for dashboards, web scraping packages (requests, BeautifulSoup).

4. Core Requirements and Process

Step 1: Formulate Your Own Real-World Question

Examples:

- "Which tech skills are emerging fastest for entry-level roles in my country in 2024–25?"
- "How do data science starting salaries vary by city and experience?"
- "What are the top soft skills employers mention in remote-friendly jobs?"

Step 2: Gather & Explore Data

- Use at least 2 different data sources.
- Try to include both structured and unstructured elements.

Step 3: Clean, Integrate, and Process

- Demonstrate handling of missing values, inconsistent formats, and outlier detection in Python (pandas/DataFrame workflows).
- Convert data types, parse/normalize fields (e.g., standardize job titles, skill keywords, salary ranges).
- For text: Use basic NLP to extract frequent skills or themes (optional but encouraged).

Step 4: Analyze, Visualize, and Tell the Story

- Visualize insights using Python (matplotlib/seaborn/plotly) or combine with Tableau/Excel.
- Aim for clear charts: trends, comparisons, word clouds for skills, or custom dashboards.
- Build a narrative: "What's surprising, counterintuitive, or notable? What *don't* you know because of data gaps?"

Step 5: Reflect and Discuss

- Justify your data and tool choices.
- Highlight limitations or ambiguities in your analysis.
- Discuss multiple possible interpretations or next steps if you had more or cleaner data.

5. Deliverables

Submit as a single report (Jupyter notebook, PDF, or dashboard link) including:

- A short intro: Explain your chosen question and why it matters to grads.
- Clean, documented code for data prep and analysis (Python a must for data wrangling).
- At least two high-quality, original visualizations.
- A written narrative (bullet points or 1–2 paragraphs) tying data to answers, uncertainties, and actionable advice.

- Reflection: What would you do next with better or more data? How did your analysis evolve as you explored?
- Short presentation (5–7 minutes) to the evaluation panel summarizing your approach, findings, and reflections, followed by a brief Q&A.

6. Rubric (For Grading)

Criteria	Description	Weight
Relevance of Question	Topic clearly motivated and focused on careers/jobs	5%
Data Diversity & Preparation	Use of mixed source/format data and depth of cleaning	15%
Analytical Depth & Coding	Solid, reproducible analysis/code in Python	20%
Narrative and Visualization	Effective storytelling, clarity of insights, creative visuals	20%
Critical Reflection & Adaptability	Bias awareness, handling ambiguity, thoughtful discussion	20%
Oral Presentation & Q&A	Clarity, confidence, response to questions, presentation visuals	20%

This assignment invites you to “think like a data scientist”—curious, critical, creative, and collaborative. There is no single path to success: your story is your own.