📊 NumPy Data Explorer: Employee Productivity Analysis

🔖 Project Title

NumPy Data Explorer – Employee Productivity & Salary Analysis

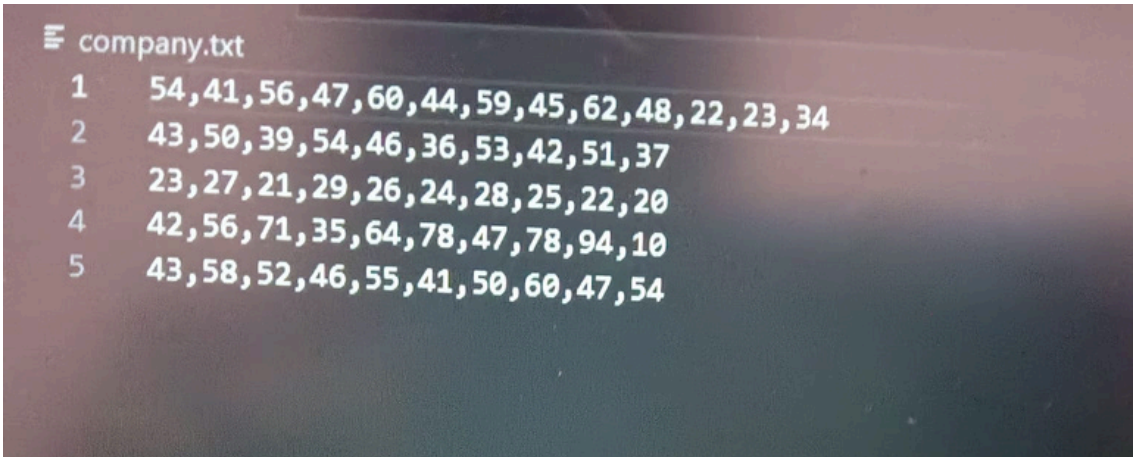
---

🧠 Project Overview

This project demonstrates how **NumPy** can be used to handle **real-world messy data**, clean it efficiently, and extract meaningful insights through **data analysis and visualization**.

The dataset represents **employee productivity and salary data** collected from multiple companies.
Since real-world datasets often contain **unequal row lengths and missing values**, this project focuses on:

- Cleaning inconsistent data
- Handling missing values using NumPy
- Converting raw text data into structured NumPy arrays
- Performing numerical analysis
- Visualizing employee productivity trends



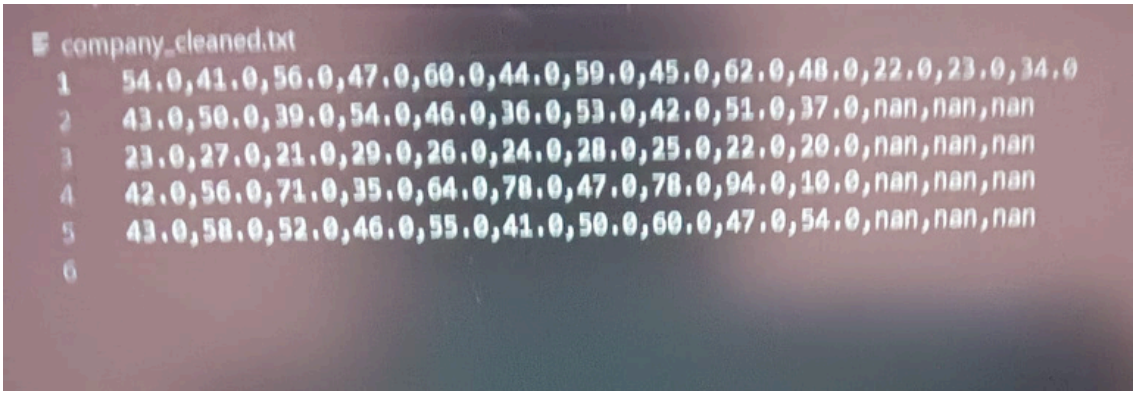
📁 Dataset Description

1️⃣ Raw Data (company.txt)

- Contains employee production data
- Each line represents a company
- Rows have **different numbers of values**
- Data is unstructured and inconsistent

📌 Example issues:

- Unequal row lengths
- Difficult to analyze directly using NumPy



2️⃣ Cleaned Data (company_cleaned.txt)

- All rows are converted to **equal length**
- Missing values are filled with NaN
- Data is converted to **float type**
- Ready for numerical computation and visualization

📌 Cleaning Techniques Used:

- Row length normalization
- Missing value handling using NaN
- NumPy array conversion

🔍 Key Features of the Project

✔ Reading raw data from .txt files
✔ Data cleaning using **NumPy only**
✔ Handling missing values with NaN
✔ Structured storage of employee productivity
✔ Salary data generation and alignment
✔ Data visualization using **Matplotlib**
✔ Real-world data handling approach

📊 Data Visualization

The project includes a **line graph visualization** titled:

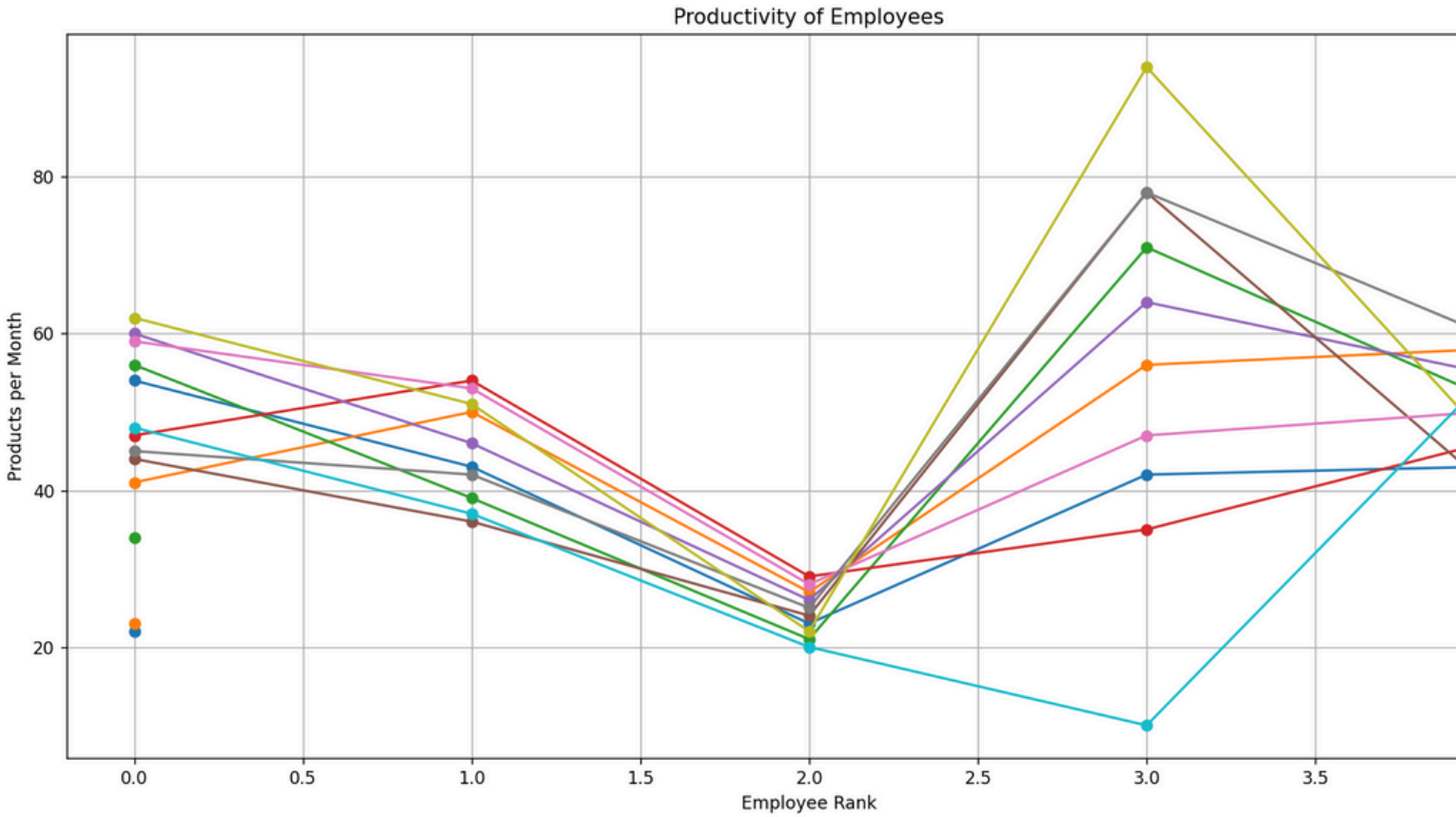
"Productivity of Employees"

Graph Details:

- **X-axis:** Employee Rank
- **Y-axis:** Products per Month
- Each line represents employee productivity across companies
- Helps compare performance patterns visually

📈 This visualization makes it easy to:

- Identify high-performing employees
- Observe productivity variations
- Compare trends across different ranks



output graph

🛠️ Technologies Used

- **Python 3.12**
- **NumPy** – Data cleaning, numerical operations
- **Matplotlib** – Data visualization
- **VS Code** – Development environment

📂 Project Structure

```
NumPy-Data-Explorer/
│
├── company.txt            # Raw employee data
├── company_cleaned.txt    # Cleaned structured data
├── main.py                # Main program file
├── README.md              # Project documentation
└── output_graph.png       # Productivity visualization
```

🚀 How the Project Works

1️⃣ Load raw data from company.txt
2️⃣ Clean inconsistent rows using NumPy
3️⃣ Replace missing values with NaN
4️⃣ Convert data into NumPy arrays
5️⃣ Generate salary data aligned with productivity
6️⃣ Visualize employee productivity using Matplotlib
7️⃣ Display cleaned output and insights in terminal

🎯 Learning Outcomes

Through this project, I learned:

- Practical use of NumPy for real-world data
- Handling messy and unstructured datasets
- Importance of data cleaning before analysis
- Efficient numerical computation with arrays
- Data visualization for better insights

📌 Conclusion

This **NumPy Data Explorer project** is a beginner-friendly yet powerful demonstration of **data cleaning, analysis, and visualization** using NumPy.
It reflects real-world data challenges and shows how Python can efficiently handle them.