# Data Analysis using Python

## **Introduction**

This project analyzes employee data from **ABC Company**, which consists of **458 rows and 9 columns**. The goal is to **preprocess, analyze, and visualize** the dataset to extract meaningful insights about the workforce.

The dataset was processed using **Python** with libraries such as **NumPy, Pandas, Matplotlib, and Seaborn**.Below is an overview of the key steps and findings from the analysis.

Dataset used : <https://docs.google.com/spreadsheets/d/1VP9BE_eI2yl6uUHSm4mGiiwjRdoqCqnkcIjsv5Q2ex4/edit?usp=share_link>

Below is an overview of the key steps and findings from the analysis.

Step 1: Imported necessary libraries like **numpy,pandas,matplotlib,seaborn**

Step 2 : Loaded dataset which is in the csv file format

Step 3 : Used the **info()** function to check for missing values.Filled null values in the **Salary** column with the **mean salary** and filled null values in the **College** column with the string **'Unknown'**.The **Height** column had incorrect values, so it was replaced withrandom values between **150 cm and 180 cm**.

Step 4 : The **team-wise employee distribution** was calculated both in **count** and **percentage**.and calculated the percentage split relative to the total number of employees.The **New Orleans Pelicans** had the highest number of employees . A **bar chart** is used to represent the distribution of employees across teams.

Step 5: Segregated employees based on their positions within the company . The position **"SG"** had the **maximum** number of employees. Visualized these findings using a **bar chart .**

Step 6 : Identified the predominant age group among employees.Employees were grouped based on age, and it was found that most employees are 24 years old. Visualized this using a **bar chart**.

Step 7 : Discovered which team and position have the highest salary expenditure.The **Cleveland Cavaliers** team had the **highest total salary expenditure**.The **"C" (Center) position** was the **most highly paid** among all positions. Visualized these findings using a **bar chart** .

Step 8 : Investigated if there's any correlation between age and salary. A slight positive correlation was found between age and salary. A **Seaborn scatter plot** with a regression line was used to depict the correlation.

**Conclusion :**

This analysis provides valuable workforce insights, highlighting key trends in team structure, salaries, and employee demographics.

That’s the end of the project…