## TITLE

EMPLOYEE DATA

ANALYSIS

AND

INSIGHTS

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Prepared by: NIRMAL KUMAR**

**Date:**11/03/2025

**Institution/Organization:** KIET GROUP OF INSTITUTIONS

**Course/Department:** CSEAI

University Roll No. : 242801100300162

Submitted to : Mr. Mayank Lakhotia

**🔍 Background & Purpose**

In modern organizations, **employee data analysis** is essential for making informed decisions regarding **salaries, experience, and department-wise performance**. Understanding salary trends, departmental variations, and experience-based compensation structures helps organizations optimize workforce management, improve employee satisfaction, and ensure fair pay structures.

This project aims to **analyze and visualize employee data** to extract meaningful insights, such as:  
✔ Identifying **salary trends** across departments  
✔ Finding **correlation between salary & experience**  
✔ Detecting **outliers in salaries**  
✔ Understanding **departmental salary distribution**  
✔ Exploring **employee demographics**

By leveraging **Python, Pandas, Seaborn, and Matplotlib**, this study provides a **data-driven approach** to employee analytics, helping HR professionals and management make strategic workforce decisions.

METHODOLOGY:

**🛠 Tools & Technologies Used**

* **Python** – For data manipulation and analysis
* **Pandas** – For handling structured data (CSV file)
* **Matplotlib & Seaborn** – For visualizing salary distribution, correlations, and trends
* **Statistical Analysis** – Using correlation, outlier detection, and department-wise insights

**📂 Steps Followed**

**Step 1: Data Collection & Preprocessing**

* The employee dataset (CSV format) was loaded into Python using Pandas.
* Initial checks were performed to inspect missing values and data types.

**Step 2: Exploratory Data Analysis (EDA)**

* **Summary statistics** were computed, including min, max, mean, and standard deviation of salary, experience, and other numerical attributes.
* **Department-wise average salary & experience** were analyzed.
* **Correlation analysis** was performed between salary and experience.

**Step 3: Advanced Analysis & Outlier Detection**

* **Outliers in salary** were detected using the **Interquartile Range (IQR) method**.
* Employees with **highest & lowest salaries** were identified.

**Step 4: Data Visualization**

* **Salary Distribution** (Histogram with KDE)
* **Experience vs Salary Scatterplot** (Color-coded by department)
* **Department-wise Salary Boxplot** (To show salary spread)
* **Salary Growth Trendline** (Regression plot)
* **Employee Distribution Pie Chart**

**Step 5: Interpretation & Insights**

* The results were analyzed to identify key patterns, correlations, and department-wise salary trends.

CODE TYPED:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load the employee data CSV file

file\_path = "employee\_data.csv"  # Update if needed

df = pd.read\_csv(file\_path)

# Display basic dataset information

print("📌 Basic Information:\n", df.info())

# Display first few rows

print("\n📌 First Few Rows of the Dataset:\n", df.head())

# -------------------- SUMMARY STATISTICS --------------------

print("\n📌 Summary Statistics:\n", df.describe())

# -------------------- DEPARTMENT-WISE ANALYSIS --------------------

# Average salary and experience by department

dept\_analysis = df.groupby("Department")[["Salary", "Experience"]].mean()

print("\n📌 Department-wise Average Salary and Experience:\n", dept\_analysis)

# Count of employees in each department

dept\_count = df["Department"].value\_counts()

print("\n📌 Employee Count per Department:\n", dept\_count)

# -------------------- CORRELATION ANALYSIS ---------------# Compute correlation between Experience & Salary

correlation = df[["Experience", "Salary"]].corr()

print("\n📌 Correlation Between Experience & Salary:\n", correlation)

# -------------------- HIGHEST & LOWEST SALARY EMPLOYEES --------------------

highest\_salary = df.loc[df["Salary"].idxmax()]

lowest\_salary = df.loc[df["Salary"].idxmin()]

print("\n📌 Employee with Highest Salary:\n", highest\_salary)

print("\n📌 Employee with Lowest Salary:\n", lowest\_salary)

# -------------------- OUTLIER DETECTION --------------------

# Identify outliers in salary using IQR method

Q1 = df["Salary"].quantile(0.25)

Q3 = df["Salary"].quantile(0.75)

IQR = Q3 - Q1

lower\_bound = Q1 - 1.5 \* IQR

upper\_bound = Q3 + 1.5 \* IQR

outliers = df[(df["Salary"] < lower\_bound) | (df["Salary"] > upper\_bound)]

print("\n📌 Outliers in Salary:\n", outliers if not outliers.empty else "No outliers detected.")

# -------------------- DATA VISUALIZATION --------------------

# Plot Salary Distribution

plt.figure(figsize=(8, 5))

sns.histplot(df["Salary"], bins=10, kde=True, color="blue")

plt.title("Salary Distribution")

plt.xlabel("Salary")

plt.ylabel("Frequency")

plt.show()

# Plot Experience vs Salary

plt.figure(figsize=(8, 5))

sns.scatterplot(x=df["Experience"], y=df["Salary"], hue=df["Department"], palette="coolwarm")

plt.title("Experience vs Salary")

plt.xlabel("Years of Experience")

plt.ylabel("Salary")

plt.show()

# Boxplot of Salary by Department

plt.figure(figsize=(10, 6))

sns.boxplot(x="Department", y="Salary", data=df, palette="Set2")

plt.title("Salary Distribution by Department")

plt.xlabel("Department")

plt.ylabel("Salary")

plt.xticks(rotation=45)

plt.show()

# Salary Growth with Experience (Trendline)

plt.figure(figsize=(8, 5))

sns.regplot(x=df["Experience"], y=df["Salary"], scatter\_kws={"color": "red"}, line\_kws={"color": "black"})

plt.title("Salary Growth with Experience")

plt.xlabel("Experience (Years)")

plt.ylabel("Salary")

plt.show()

# Department-wise Employee Distribution (Pie Chart)

plt.figure(figsize=(6, 6))

plt.pie(dept\_count, labels=dept\_count.index, autopct="%1.1f%%", colors=sns.color\_palette("pastel"))

plt.title("Employee Distribution by Department")

plt.show()

SCREENSHOTS:

