Employee Dataset

WorkForce Insights: Python-Based Employee Data Analysis

Project Overview:

This project focuses on performing Exploratory Data Analysis (EDA) on an Employee Dataset using Python. The aim is to understand the dataset, clean missing and inconsistent data, generate new metrics, perform statistical tests, and create visualizations to derive insights for decision-making.

Learning Objectives:

- Import, explore, and summarize datasets using Python.
- Clean and preprocess data by handling missing values, outliers, and duplicates.
- Create derived metrics for deeper analysis.
- Perform univariate, bivariate, and multivariate analysis.
- Conduct statistical hypothesis testing.
- Visualize data patterns using Python libraries.

Selection of Dataset:

The dataset selected for this project is an Employee Dataset containing details such as salary, age, gender, experience, city, department, and performance score. This dataset is suitable for EDA because it includes both categorical and numerical features, allowing for rich analysis and visualization.

Project Tasks

- Introduction and dataset description
- Data Understanding (shape, types, summary, missing values)
- Data Cleaning (missing values, duplicates, outliers, inconsistencies)
- Derived Metrics creation
- Filtering data for analysis
- Statistical analysis and hypothesis testing
- EDA through visualizations (univariate, bivariate, multivariate)
- Insights and conclusion

Tools & Libraries

- Python
- pandas – Data handling
- numpy – Numerical operations
- matplotlib – Data visualization
- seaborn – Advanced visualizations
- scipy – Statistical testing

Deliverables

- Python code (.ipynb or .py)
- Full project documentation PDF
- Visualizations and insights