HR Data Analysis Case Study

Project Overview

This case study documents an **HR Data Analysis project** performed strictly based on the provided dataset and analysis code. The project uses a workforce dataset containing **2,000,000 employee records** and focuses on exploratory data analysis (EDA) and visualization using Python and Excel.

All observations, metrics, and descriptions in this document are **directly derived from the dataset and analysis steps present in the source file**, without adding assumptions, simulations, or external business context.

Analysis Objective

The objective of this project was to analyze a large-scale HR dataset to:

- Understand employee distribution across departments, job roles, locations, and work modes
- Analyze hiring trends over time using hire date information
- Examine employee status distribution (Active, Resigned, Terminated, Retired)
- Study salary patterns in relation to experience and department

These objectives are based solely on the exploratory analysis performed in the code.

Dataset Description

- **Total Records:** 2,000,000+ employees
- **Data Size:** ~170 MB
- **Key Attributes:**
    - Employee_ID
    - Full_Name
    - Department
    - Job_Title
    - Hire_Date
    - Location

- Performance_Rating (1–5)
- Experience_Years (0–15)
- Employment Status (Active, Resigned, Terminated, Retired)
- Work_Mode (On-site, Remote)
- Salary (INR)

The dataset contained **no missing values**, ensuring high data quality for analysis.

Tools & Technologies Used

- **Programming Language:** Python
- **Libraries:**
  - Pandas – data manipulation and aggregation
  - NumPy – numerical operations and conditional logic
  - Matplotlib & Seaborn – data visualization
- **Data Handling:** CSV files
- **Additional Tool:** Microsoft Excel (Pivot Tables & summaries)

Data Preparation & Feature Engineering

- Converted Hire_Date into datetime format
- Extracted **Year** and **Month** from hire dates for time-based analysis
- Renamed and standardized columns for clarity
- Created a **Salary Hike** feature based on experience:
  - 1–5 years: 5% increment
  - 6–10 years: 10% increment
  - 11–15 years: 20% increment
- Validated data types and memory usage for large-scale processing

---

Exploratory Data Analysis (EDA)

1. Workforce Experience Analysis

- Analyzed total experience by department
- Identified departments with the most experienced workforce
- Found that **IT, Finance, HR, and Marketing** consistently had high cumulative experience

2. Hiring Trends Over Time

- Year-wise hiring analysis segmented by employee status
- Identified spikes in hiring during recent years
- Visualized trends using line charts with status-based segmentation

3. Attrition & Employment Status Analysis

- Compared Active, Resigned, Terminated, and Retired employees across departments
- Identified departments with relatively higher resignation counts
- Helped highlight potential attrition risk areas

4. Salary & Compensation Insights

- Analyzed salary distribution across departments
- Observed strong correlation between **experience and salary growth**
- Calculated projected salary hikes based on experience slabs
- Used pie charts to show department-wise salary share

5. Job Role Distribution

- Identified **top 5 job roles** by employee count (e.g., Software Engineer, Sales Executive)
- Identified **bottom 5 niche roles** with minimal headcount
- Insights useful for workforce balancing and hiring strategy

6. Work Mode Analysis

- Compared On-site vs Remote workforce across employment status
- Highlighted adoption of remote work and its distribution

7. Location-Based Workforce Distribution

- Identified top and bottom employee locations globally
- Used bar charts and pie charts to visualize geographic concentration

---

Excel Analysis

- Created **Pivot Tables** to summarize:
    - Department-wise headcount
    - Status-wise employee distribution
    - Salary summaries
- Used Excel for quick stakeholder-friendly summaries alongside Python-based analysis

---

Key Observations from Analysis

- Department-wise experience totals were calculated using grouped aggregations
- Hiring trends were visualized year-wise and month-wise using employee counts
- Salary distribution was analyzed across departments and experience levels
- Job titles were ranked based on employee count to identify high- and low-frequency roles
- Workforce distribution was examined by work mode and geographic location

All observations reflect patterns visible in the generated aggregations and visualizations.

Analysis Output & Business Relevance

- Generated multiple aggregated tables and visualizations using Pandas, Matplotlib, and Seaborn
- Calculated experience-based salary hike values using NumPy conditional logic
- Produced department, job title, location, and work mode–based summaries
- Used Excel Pivot Tables to support summary-level reporting

HR Strategy & Decision-Making Support

The analysis outputs are structured to support **HR strategy and data-driven decision-making**, including:

- Identifying departments and job roles with higher attrition-related status counts
- Supporting workforce planning through hiring trend analysis
- Informing compensation and salary benchmarking using experience-based salary patterns

- Assisting leadership in understanding workforce distribution across locations and work modes

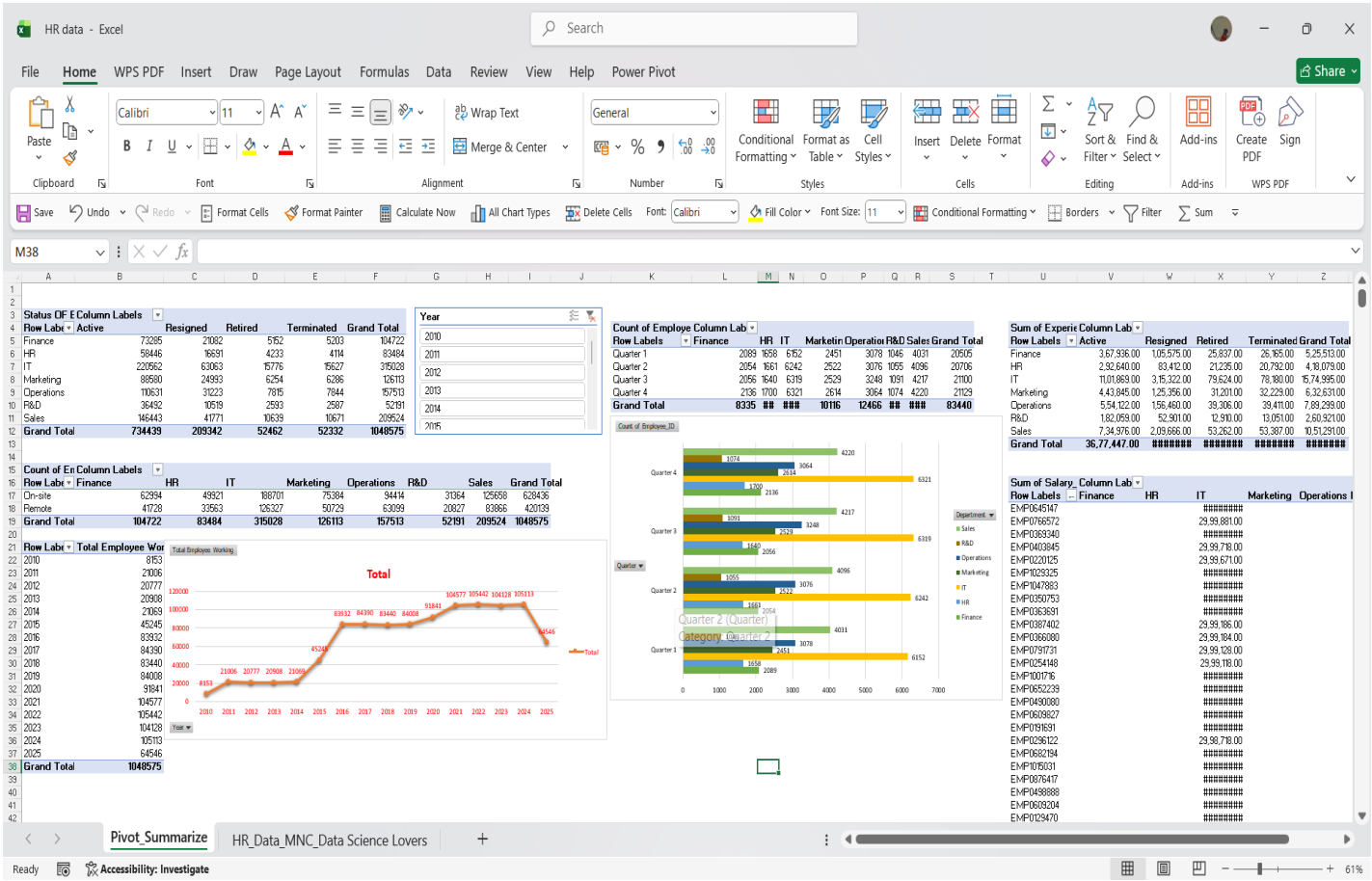
These outcomes are derived from analytical results and visual patterns observed in the dataset.

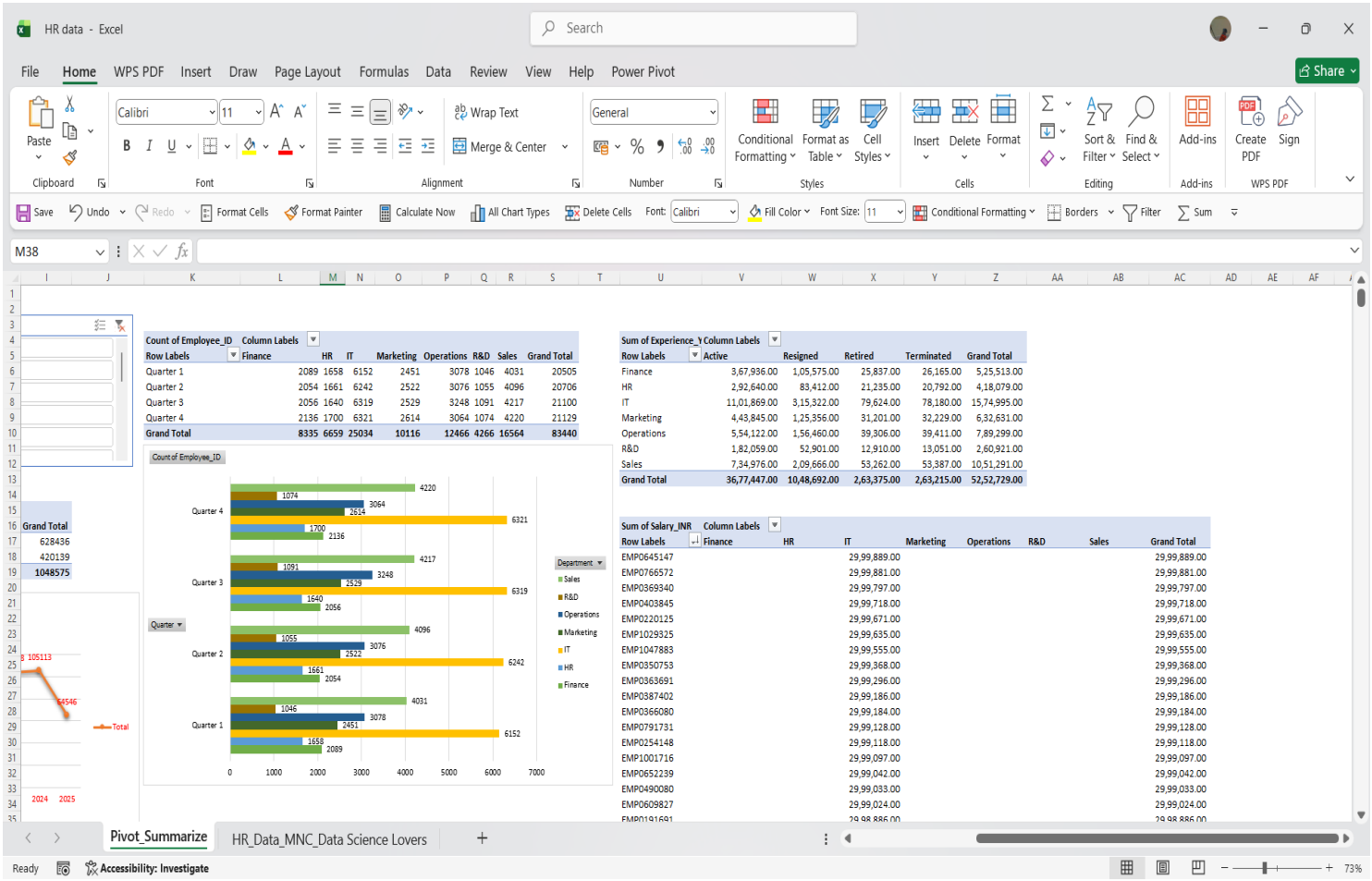
Resume-Ready Project Summary

HR Data Analysis Project | Python, Pandas, NumPy, Matplotlib, Seaborn, Excel

- Performed exploratory data analysis on a dataset of 2,000,000 employee records
- Used Pandas and NumPy for data cleaning, aggregation, and feature creation
- Built visualizations for hiring trends, salary distribution, experience, and workforce composition
- Created Excel Pivot Tables for department-wise and status-wise summaries

---





Conclusion

This project showcases strong analytical skills, real-world data handling at scale, and the ability to translate complex HR data into meaningful business insights. It is well-suited for roles in **Data Analyst, HR Analyst, and Business Analyst** domains.