ReadMe: HR Data Analysis Project

Project Overview

This project aims to **analyze Human Resources (HR) data to gain deep insights into employee demographics, salary distribution, and attrition trends**.

The project was developed as part of the **Digital Egypt Pioneers Initiative (DEPI) – Data Analysis Track**.

Primary Goal: To support HR departments with data-driven decisions that improve employee retention, workforce planning, and employee satisfaction.

Key Findings & Insights

The diagnostic analysis revealed critical risk factors impacting employee retention that require immediate action:

- **Overall Attrition Rate:** The company's overall attrition rate is **16.12%** , out of approximately 1.47K employees.
- **Overtime Risk:** Employees working overtime show a significantly higher attrition rate of **28%**, compared to **10%** for non-overtime workers.
- **Departmental Risk:**
    - The **Technology department** carries the highest attrition risk, accounting for **56.12%** of departmental attrition cases.
    - The **Sales department** is the second highest at **38.82%**.
- **Salary vs. Performance/Satisfaction:** There is **no significant correlation** between high salaries and higher performance ratings or job satisfaction.
- **Primary Retention Risk Factors:** The data highlights three key primary risk factors:
    1. Early Tenure (**0-3 years**).
    2. **Low Salary Groups**.
    3. Working **Overtime**.

Tools and Technologies

Category	Tool/Technology	Project Purpose
Core Language	**Python**	Core language for data preparation and processing.
Frameworks/Libraries	**Pandas**	Used extensively for data cleaning, duplicate removal, and mode imputation.
Visualization Tool	**Power BI**	Primary tool for data modeling, KPI creation, and developing interactive dashboards.
Development Environment	Google Colab Notebook	Used for executing Python data cleaning and preprocessing steps.
Data Source	Kaggle	Source for the raw HR dataset.

Data Preparation and Cleaning

1. Data Source

The dataset was imported from Kaggle and included multiple Excel files like 'SatisfiedLevel.xlsx' and 'Employee (E.xlsx). The final model represents approximately 1.47K employees.

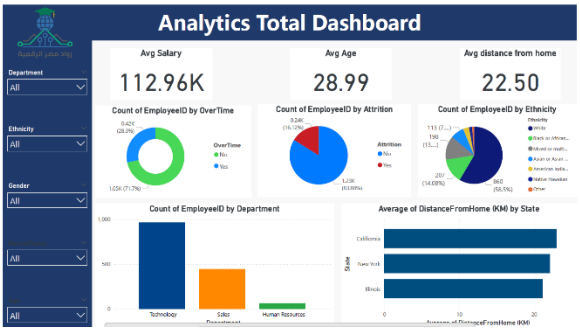
2. Cleaning Procedures

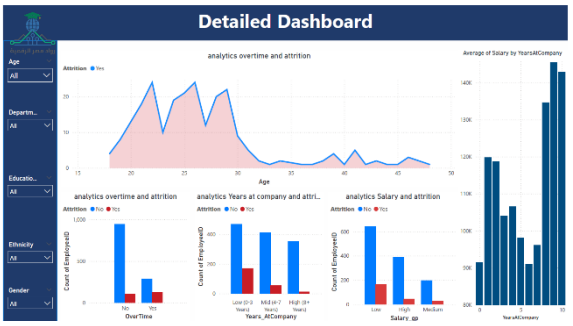
Data cleaning was performed using Python (Pandas) scripts:

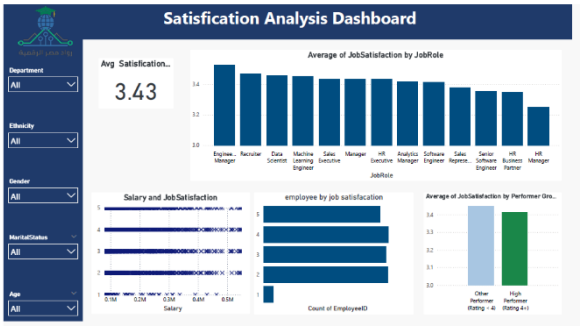
- **Duplicate Removal: 19 duplicate records** were identified and removed from the Employee dataset to ensure data integrity.
- **Missing Value Imputation:** Missing values in key categorical columns such as **'Gender'**, **'Department'**, and **'State'** were imputed using the **Mode (most frequent value)** to preserve categorical distribution and minimize bias.

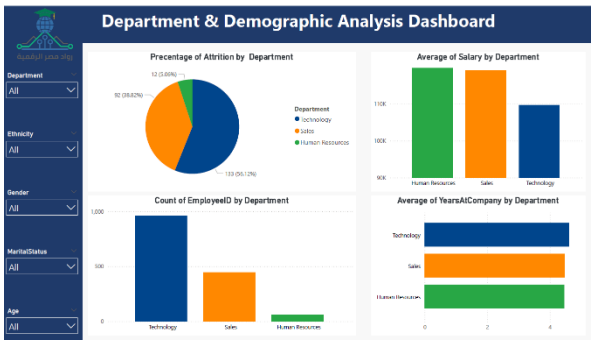
Dashboard Structure & Features

The final deliverable is a production-ready, dynamic Power BI Dashboard.









- **Key Interfaces (Views):**
    1. General Overview Dashboard (showing Avg Salary, Attrition Rate, Overtime breakdown).
    2. Department & Demographic Analysis (comparing salary, tenure, and attrition).
    3. Performance Analysis (Salary vs. Manager Rating, tenure volatility).
    4. Satisfaction and Education Analysis.
- **Usability Focus:** The interface is designed for maximum usability, allowing users to:
    - **Filter Globally:** Use slicers (Age, Department, Ethnicity, Gender) on the left panel to instantly segment data.
    - **Drill-Down:** Click on high-risk areas (e.g., Technology on the attrition chart) to dynamically filter all associated metrics across the interface.

Team and Roles

Member	Role
Arwa Kassim	Team Leader, Reporting & Documentation (Data Wrangling, Business Insights, ReadMe).
Rwan Adel	Data Cleaning and Preprocessing using Python (Pandas).
Mohamed Abdelbary	Data Cleaning and Preprocessing using Python (Pandas).
Ali Mamdouh	Designing and developing interactive Power BI dashboards and presenting key insights.
Gehad Adel	Creating PowerPoint presentations and participating in Power BI dashboard development and presentation.

Contact

For inquiries, please contact the Team Leader:

- **Arwa Kassim:** arwamac2016@gmail.com