HR Insights Pro: Employee Retention & Performance Analytics

1. Project Overview

A comprehensive HR analytics project aimed at identifying the key factors influencing employee attrition, satisfaction, and performance. Using **Python** for cleaning and feature engineering, **PostgreSQL (SQL)** for data querying, and **Power BI** for visualization, the project provides data-driven insights to improve employee retention and workforce planning.

2. Dataset Summary

Rows: 1,470 **Columns:** 35
Key Features:
- Demographics: Age, Gender, Marital Status, Education, Education Field, Distance from Home
- Job Details: Department, Job Role, Job Level, Job Involvement, Work-Life Balance, Years at Company
- Performance & Satisfaction: Performance Rating, Job Satisfaction, Environment Satisfaction, Relationship Satisfaction
- Compensation: Monthly Income, Daily Rate, Hourly Rate, Stock Option Level, Percent Salary Hike
- Work Conditions: Overtime, Business Travel, Training Times Last Year, Standard Hours
- Engineered Features: age_group, experience_level, attrition_flag

Data Source: IBM HR Analytics Employee Attrition Dataset (Kaggle)

3. Exploratory Data Analysis (Python)

Conducted structured data preparation and feature engineering to ensure clean, reliable inputs for SQL and Power BI analysis.
Key Steps:
- **Data Loading:** Imported the *HR Employee Attrition* dataset using **pandas** for exploration and preprocessing.
- **Initial Exploration:** Used df.info() and df.describe() to inspect data structure, data types, and summary statistics for numeric columns.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
...
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

- **Missing Data Check:** Verified null values using df.isnull().sum() — found no significant missing data across columns.
- **Column Standardization:** Converted all column names to lowercase and replaced spaces with underscores for consistent formatting and easier SQL integration.

- **Feature Engineering:**
  ○ Created an **agegroup** column using `pd.cut()` to segment employees into age categories — *Young, Adult, Mid-Age, Senior.*
  ○ Created an **experiencelevel** column by binning `yearsatcompany` into stages — *New, Intermediate, Experienced, Senior, Veteran.*
  ○ Generated a clean structured dataset for further SQL-based analysis.
- **Duplicate Check:** Validated dataset integrity using `df.duplicated().sum()` — confirmed no duplicate records.
- **Database Integration:** Established a connection to **PostgreSQL** using *SQLAlchemy* and exported the cleaned DataFrame to the `employee_attrition` table for SQL analysis.
- **Export:** Saved the final cleaned dataset as `cleaned_hr_employee_attrition.csv` for future reference and Power BI visualization.

4. Data Analysis using SQL (Business Insights)

Performed advanced queries in PostgreSQL to uncover trends and validate HR hypotheses:

1. **Attrition by Department** – Compared attrition percentages across departments to identify divisions with the highest turnover (e.g., Sales, R&D).

	total_employees bigint	employees_left bigint	attrition_rate numeric
1	1470	237	16.12

2. **Attrition by Job Role** – Determined roles most affected by attrition, helping HR prioritize critical retention plans.

	department text	total_employees bigint	employees_left bigint	attrition_rate numeric
1	Sales	446	92	20.63
2	Human Resources	63	12	19.05
3	Research & Developm...	961	133	13.84

3. **Average Salary by Job Level** – Analyzed compensation fairness across levels to detect potential pay gaps.

	jobrole text	avg_income numeric
1	Manager	17181.68
2	Research Director	16033.55
3	Healthcare Representati...	7528.76
4	Manufacturing Director	7295.14
5	Sales Executive	6924.28
6	Human Resources	4235.75
7	Research Scientist	3239.97
8	Laboratory Technician	3237.17
9	Sales Representative	2626.00

4. **Overtime Impact on Attrition** – Examined how working overtime influences employee turnover rates.

	overtime text	total_employees bigint	employees_left bigint	attrition_rate numeric
1	Yes	416	127	30.53
2	No	1054	110	10.44

5. **Distance from Home vs Attrition** – Evaluated whether longer commute distances correlate with higher attrition.

	distance_range text	total_employees bigint	employees_left bigint	attrition_rate numeric
1	20+ km	204	45	22.06
2	11-20 km	240	48	20.00
3	6-10 km	394	57	14.47
4	0-5 km	632	87	13.77

6. **Attrition by Education Field** – Compared attrition rates across education backgrounds to tailor career development strategies.

	educationfield text	total_employees bigint	employees_left bigint	attrition_rate numeric
1	Human Resour...	27	7	25.93
2	Technical Degr...	132	32	24.24
3	Marketing	159	35	22.01
4	Life Sciences	606	89	14.69
5	Medical	464	63	13.58
6	Other	82	11	13.41

7. **Job Satisfaction vs Attrition** – Investigated if low satisfaction scores contribute significantly to employee exits.

	jobsatisfaction bigint	total_employees bigint	employees_left bigint	attrition_rate numeric
1	1	289	66	22.84
2	2	280	46	16.43
3	3	442	73	16.52
4	4	459	52	11.33

8. **Experience Level vs Attrition** – Analyzed attrition by total working years to understand tenure-based turnover trends.

	experiencelevel text	total_employees bigint	employees_left bigint	attrition_rate numeric
1	[null]	1	1	100.00
2	New	215	75	34.88
3	Intermediate	365	66	18.08
4	Experienced	524	58	11.07
5	Senior	273	29	10.62
6	Veteran	92	8	8.70

9. **Salary Hike vs Retention** – Compared percent salary hike groups to identify how compensation adjustments affect retention.

	percentsalaryhike bigint	total_employees bigint	employees_left bigint	attrition_rate numeric
1	25	18	1	5.56
2	24	21	6	28.57
3	23	28	6	21.43
4	22	56	12	21.43
5	21	48	5	10.42
6	20	55	7	12.73
7	19	76	9	11.84
8	18	89	13	14.61
9	17	82	14	17.07
10	16	78	14	17.95
11	15	101	18	17.82
12	14	201	24	11.94
13	13	209	34	16.27
14	12	198	33	16.67
15	11	210	41	19.52

10. **Performance Rating by Age Group** – Evaluated average performance across age groups to uncover age-based productivity trends.

	agegroup text	avg_performance numeric	avg_satisfaction numeric	total_employees bigint
1	[null]	3.20	2.20	5
2	Young	3.16	2.74	97
3	Adult	3.16	2.73	554
4	Senior	3.16	2.70	309
5	Mid-Age	3.15	2.74	505

5. Power BI Dashboard

Developed an interactive Power BI dashboard for HR decision-making:
- Department-wise Attrition & Salary Distribution
- Overtime Impact Analysis
- Job Satisfaction vs Attrition Trends
- Distance from Home vs Attrition Rate
- Dynamic Filters: Department, Gender, Age Group, Experience Level

These visuals empower leadership to monitor attrition KPIs, assess workforce health, and optimize retention strategies.

6. Business Recommendations

- **Reduce Attrition** – Introduce mentorship and recognition programs for critical roles.
- **Promote Balance** – Implement flexible schedules to minimize overtime stress.
- **Ensure Pay Equity** – Align salaries with performance and experience levels.
- **Grow Internally** – Build structured upskilling and career growth frameworks.
- **Adopt Hybrid Work** – Offer flexibility to employees with long commutes.
- **Boost Engagement** – Launch regular satisfaction surveys and wellness initiatives.
- **Leverage Data** – Use dashboards for predictive HR decision-making.

7. Professional Impact

- Strengthened HR decision-making and workforce transparency.
- Identified 20–25% potential reduction in annual attrition.
- Improved engagement tracking and pay structure alignment.
- Enhanced data-driven culture for continuous HR optimization.