**PROJECT REPORT OBJECTIVES AND INSIGHTS**

Data Collection and Cleaning using sql

Employee Performance Analysis using sql

Productivity and Engagement Insights using python

Attrition and Retention Modeling using python

Retention Strategy Recommendations using python

Interactive Dashboard Creation using tableau

Final Report and Strategic Recommendations using tableau

**The data contains the following columns:**

* **work\_year**: Year of employment.
* **job\_title**: Specific job title of the employee.
* **job\_category**: Broad category the job belongs to.
* **salary\_currency**: Currency in which the salary is paid.
* **salary**: Annual salary amount.
* **employee\_residence**: Location of the employee's residence.
* **experience\_level**: Level of experience (e.g., Mid-level, Senior).
* **employment\_type**: Type of employment (e.g., Full-time).
* **work\_setting**: Work arrangement (e.g., Hybrid, In-person).
* **company\_location**: Location of the company.
* **company\_size**: Size of the company (e.g., Small, Medium, Large).

**OBJECTIVES**

Data Collection and Cleaning using sql

Questions:

1. Handle Missing Data: Identify and address null or incomplete entries in key fields such as salary, job\_category, or experience\_level.

**Insights:** Understand the proportion of missing values and how they affect analysis

**Outcome:** From the selected columns, none has empty value

1. Segment Data: Categorize company\_size and work\_setting for further insights (e.g., grouping company sizes as small, medium, and large).

**Insights:** Better insights into employment types (e.g., distribution of full-time vs. part-time jobs).

Create consistent reporting categories for job types.

**Outcome:** Company size column was updated by changing S to small, M to medium and L to large.

1. Add Derived Fields: Create additional columns, such as calculating employee tenure

**Insights:** Understand employee tenure trends.

1. Data Normalization

**Insights:** To normalize table name for perfect use for insights

**Employee Performance Analysis using sql**

1. Assess Salary and Performance Correlation: Investigate whether higher salaries correlate with more senior positions or experience levels.

**Insights:** Understand how experience level impacts average salaries.

Identify if higher compensation correlates with more senior roles.

**Outcome:** From the insight Executive has highest average salary followed by senior level while entry level staff has lowest average salary.

From the resulr higher compensation correlates with more senior roles.

1. Work Setting Analysis: Compare performance trends across different work settings (Hybrid, In-person, Remote).

**Insights:** Understand how experience level impacts average salaries.

**Outcome:** From the result, staff that works in-person has highest average salary while staffs that works hybrid has lowest average salary.

1. Experience Level and Job Role Insights: Evaluate which job roles and experience levels tend to have better salaries.

**Insights:** Identify high-paying job roles and the experience levels they require.

**Outcome:** From the result Machine Learning and Ai is the job role with highest average salary at executive experience level while Data Quality and Operation has the lowest average salary in job role at entry level experience level. This implies that staff with Executive experience level tends to have highest salary while staff with entry level tends to have lower salary.

### ****Productivity and Engagement Insights Using Python****

1. **Productivity by Experience Level**: Assess productivity trends based on experience\_level.

**Insights:** Determine whether senior employees are more productive (using salary as a proxy).

**Outcome:** From the result, senior level has highest count of staff while executive has lowest staff. Staffs with executive experience level has highest average salary while staff with entry level has lowest average salary. This tends that staffs with Executive experience level are more productive using salary as a proxy while staffs with entry level has lower productivity in the company using salary as proxy.

1. **Engagement Across Job Categories:** Determine which job categories have the most engagement or output

**Insights:** Identify high-paying and high-engagement job categories.

**Outcome:** From the result, Machine Learning and AI is the job category with the highest salary while Data Quality and Operations has lowest salary. Data Engineering has the highest employee count while Cloud and Database has lowest employee count.

1. **Geographical Influence on Productivity**: Analyze how employee\_residence affects productivity or engagement.

**Insights**: Highlight regions with the most productive employees.

**Outcome:** From the results**,** employee average salary and regions shows that employee who lives in QATAR has the highest average salary while employees that lives in INDONESIA receives the lowest salary.

1. **Tenure and Engagement Trends**: Explore if longer tenure leads to higher productivity or engagement

**Insights:** Explore whether longer tenure correlates with higher salaries (proxy for engagement).

**Outcome:** A new column was added to calculate the tenure by subtracting the current year from employee’s work year. The new column was use to group the average salary to get if longer tenure correlates with higher salary. From the result, employee’s with the higher tenure tends to have lower salary while employee’s with lower tenure had highest average salary.

**Retention Strategy Recommendations using python**

### Preprocessing Data for Attrition Modeling using different python libraries

#### Build A Model to Predict Attrition

**Model Report:**

Confusion Matrix:

 [[2807]]

Classification Report:

 precision recall f1-score support

 0 1.00 1.00 1.00 2807

 accuracy 1.00 2807

 macro avg 1.00 1.00 1.00 2807

weighted avg 1.00 1.00 1.00 2807

The confusion matrix shows the model's performance:

* **Matrix Interpretation**:
	+ **True Positives (2807)**: The model accurately predicted attrition for all employees.
	+ No False Positives or Negatives were detected, indicating perfect accuracy.
* **Model Performance**:
	+ **Accuracy**: 100%
	+ **Precision, Recall, and F1-score**: All scores are 1.00, demonstrating excellent predictive capability. However, the absence of other classes (like attrition cases) indicates potential class imbalance issues.

### Feature Importance Analysis



The feature importance chart highlights the impact of each variable on attrition prediction:

* **Top Influencers**: Variables like *experience\_level*, *salary*, and *job\_category* might be critical in predicting attrition.
* **Insights**:
	+ Employees with executive experience levels or higher salaries may have unique attrition behaviors.
	+ Jobs with lower engagement (e.g., *Data Quality and Operations*) might show higher attrition rates, warranting deeper analysis.

**TABLEAU**

### ****Insights from the Dashboard****:

1. **Salary Distribution**:
	* Identify job categories with the highest salaries.
2. **Time-Based Trends**:
* Identify work year by Salary
1. **Employment and Salary Distribution:**
* Identify employment type with the highest distribution
1. **Work setting insights:**
* Identify the work setting with the highest salary
1. **Employment type, Experience level salary distribution:**
* Exploring the employment type, experience level with the highest salary to derive meaningful insights
1. **Creating slicers to filter the records and make the dashboard responsive**
* Using Year, Experience level, Job category, Employment type as slicer to filter the dashboard and make it responsive

**Charts, Interpretations, and Dashboard Insights**



The Tableau dashboard provides the following insights:

**Salary Distribution by Job Category**

* Chart type: Bar Chart

This chart gives detailed insights on Job category distribution and salary. Out of all ten Job Category Datta Science and Research tend to have the highest salary while Cloud and database job category has lowest salary.

**Time-Based Trends: Work year by Salary**

* Chart type: Line chart

This chart shows salary trend by work year, four years were present in the data and the insights gives that 2020 is the year with the lowest salary, it increased drastically when it got to 2021, 2022 and 2023 has the highest salary. Hence, 2020 has the lowest salary distribution while 2023 has highest salary distribution.

**Employment and Salary Distribution**:

* Chart type: Packed Bubble

From the report Full-time employment types dominate higher salary while Freelance employment type dominates lowest salary.

**Work Settings and salary distribution**:

* Chart type: Tree Map

In-person arrangements show higher salaries compared to hybrid or remote roles.

**Experience level, Employment type and Salary Distribution:**

* Chart type: Highlight Table

From the result Full time employment type who has Senior experience level has highest salary while Freelance employment type who has Entry level experience level has lowest salary distribution. This implies that employee’s with full time employment type tends to have higher salary as well as employee’s with senior experience level tends to have higher salary, this implies that the higher the experience level the higher the salary.

**Filtering records with Work year, Experience level, Employment type, Job category**:

* Chart type: Slicer

Slicers enable targeted insights by filtering data based on year, job category, employment type, and experience level.

**Conclusion**

The analysis provides a multi-dimensional understanding of employee performance, salary structures, and engagement trends:

1. **Performance and Salary**: Higher compensation correlates with seniority and executive roles, while entry-level positions remain undercompensated.
2. **Geographical Insights**: Salaries differ significantly across regions, with Qatar showing the highest and Indonesia the lowest.
3. **Tenure Dynamics**: Shorter-tenure employees have higher average salaries, suggesting a competitive market or proactive hiring strategies.

**Recommendations**

1. **Retention Strategies**:
	* Introduce competitive salaries for critical roles in lower-paying regions like Indonesia.
	* Target high-engagement job categories (*Data Science and Reasearch*) with retention incentives to maintain productivity.
	* Implement robust policies to support entry-level staff, fostering growth and reducing attrition.
2. **Geographical Expansion**:
	* Use insights to optimize hiring and engagement in underperforming regions.
3. **Work Setting Re-evaluation**:
	* Consider hybrid or remote work incentives to balance salary disparities and improve overall employee satisfaction.
4. **Attrition Model Refinement**:
	* Address class imbalance in the dataset to ensure a comprehensive understanding of attrition patterns.
5. **Performance Monitoring**:
	* Leverage dashboards to track year-over-year salary and engagement metrics, aligning strategy with trends.