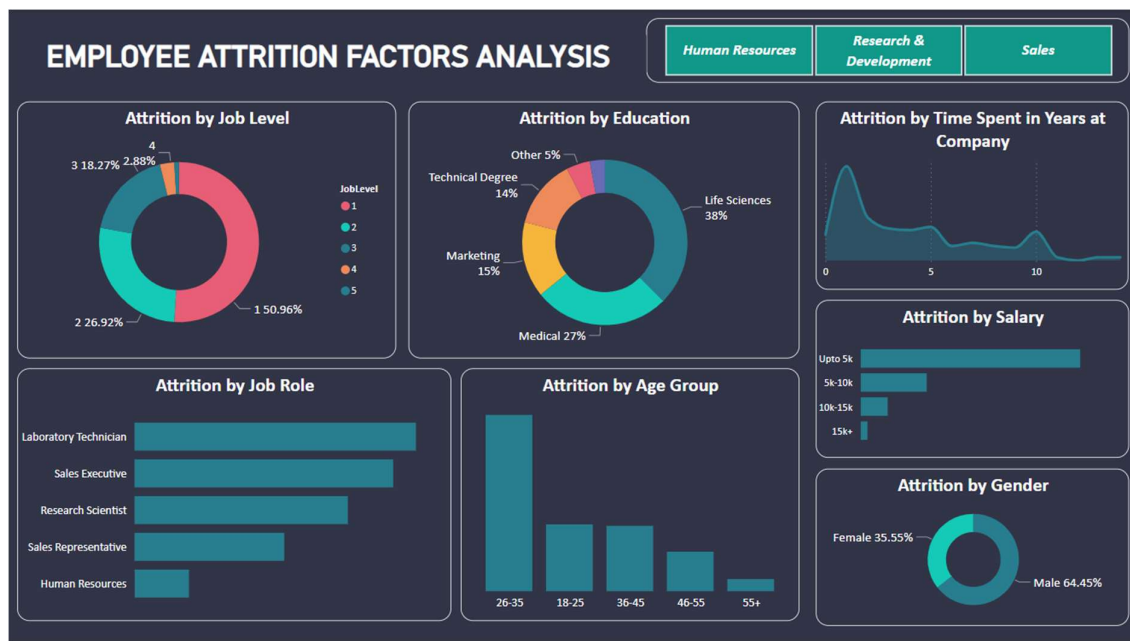**KEY INSIGHTS AND TAKEAWAYS-EMPLOYEE ATTRITION FACTORS ANALYSIS**

Detailed Analysis of Each Section:

1. **Attrition by Job Level:**

   o **Chart Type:** Donut chart.

   o **Insights:** Shows the percentage of attrition at different job.

   o **Key Takeaways:** Job level 1 has the highest attrition rate, followed by job level 2.

   o **Implications:** This suggests that employees at lower job levels (1) might be experiencing specific challenges leading to higher attrition.

2. **Attrition by Education:**

   o **Chart Type:** Donut chart.

   o **Insights:** Shows the percentage of attrition based on the employee's education background.

   o **Key Takeaways:** Life Sciences (38%) and Medical (27%) backgrounds exhibit the highest attrition, followed by Marketing (15%) and Technical Degree (14%). Employees with "Other" education backgrounds have the lowest attrition (5%).

   o **Implications:** This suggests that the company might need to address specific issues related to employees with Life Sciences and Medical backgrounds. It's worth investigating why these groups have higher turnover.

3. **Attrition by Time Spent in Years at Company:**

   o **Chart Type:** Line graph

- **Insights:** Shows the attrition rate over the years an employee has been with the company.

- **Key Takeaways:** Attrition peaks around 1 years of tenure. There's also a smaller peak around 5 and 10 years.

- **Implications:** This suggests that there might be issues related to career progression or job satisfaction around the 1,5,10-year mark. The peak around 1 years could indicate problems with onboarding or initial job fit.

4. **Attrition by Job Role:**

   - **Chart Type:** Horizontal bar chart.

   - **Insights:** Shows the attrition rate for different job roles within the company.

   - **Key Takeaways:** Laboratory Technicians, Sales Executives, Research Scientists, Sales Representatives, and Human Resources roles are listed in descending order.

   - **Implications:** The chart indicates which job roles experience higher attrition. Further investigation is needed to determine the specific reasons for attrition in each role.

5. **Attrition by Age Group:**

   - **Chart Type:** Vertical bar chart

   - **Insights:** Shows the attrition rate across different age groups.

   - **Key Takeaways:** The 26-35 age group has the highest attrition, followed by 18-25 and then 36-45. Attrition is significantly lower for older age groups (46-55 and 55+).

   - **Implications:** This suggests that younger employees are more likely to leave the company. Factors like career development, work-life balance, and compensation might be more critical for this demographic.

6. **Attrition by Salary:**

   - **Chart Type:** Stacked bar chart

   - **Insights:** Shows the attrition rate across different salary brackets (Upto 5k, 5k-10k, 10k-15k, 15k+).

   - **Key Takeaways:** low salaries have highest attrition.

   - **Implications:** This chart helps understand if salary is a contributing factor to attrition. It's worth investigating if lower salary brackets have higher attrition.

7. **Attrition by Gender:**

   - **Chart Type:** Donut chart.

   - **Insights:** Shows the percentage of attrition by gender (Female and Male).

   - **Key Takeaways:** Male employees have a significantly higher attrition rate (64.45%) compared to female employees (35.55%).

   - **Implications:** This suggests that there might be gender-specific factors contributing to attrition. Further investigation is needed to understand these factors.

EMPLOYEE ATTRITION PREDICTION USING LOGISTIC REGRESSION
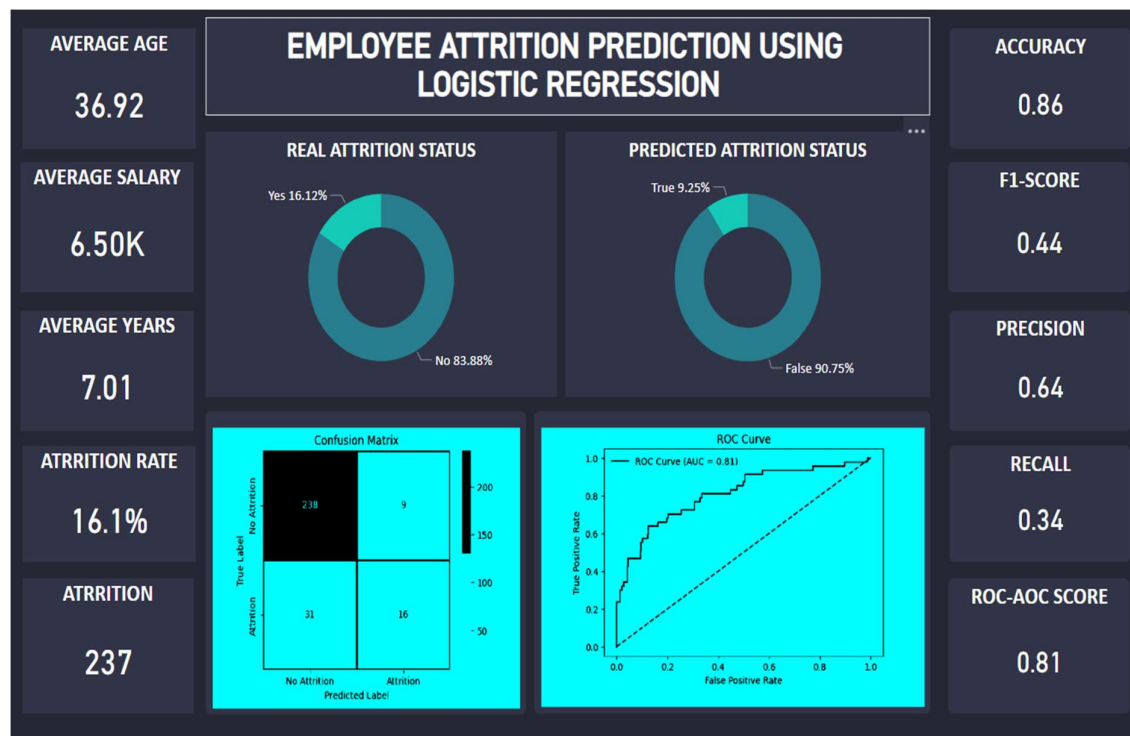
1. **KPIS**

Average Age: 36.92 (The average age of employees in the dataset or a significant age feature used in the model.)

Average Salary: 6.50K (The average salary of employees in the dataset or a relevant salary feature.)

Average Years: 7.01 (The average tenure of employees or a feature related to years of service.)

Attrition Rate: 16.1% (This is the overall attrition rate in the dataset, indicating the proportion of employees who left the company.)

Total Attrition: 237 (The total number of employees who experienced attrition in the dataset.)

2. **Real Attrition Status:**

   a. **Chart Type:** Donut Chart

   b. **Insights:** Shows the actual distribution of attrition in the dataset. 83.88% of employees did not experience attrition (No), while 16.12% did (Yes).

   c. **Significance:** This provides the baseline distribution of the target variable (attrition) in the data.

3. **Predicted Attrition Status:**

   a. **Chart Type:** Donut Chart

   b. **Insights:** Shows the distribution of attrition as predicted by the Logistic Regression model. 90.75% of employees were predicted to have "No Attrition" (False), and 9.25% were predicted to have "Attrition" (True).

   c. **Significance:** This shows the model's prediction distribution, which can be compared to the real distribution to assess model performance.

4. **Model Evaluation Metrics:**

   a. **Accuracy:** 0.86 (86%) - This is the overall proportion of correctly predicted instances (both "Attrition" and "No Attrition").

   b. **F1-Score:** 0.44 - This is the harmonic mean of precision and recall, providing a balanced measure of the model's performance, especially when dealing with imbalanced datasets (where one class has significantly more samples than the other).

   c. **Precision:** 0.64 - This measures the proportion of correctly predicted "Attrition" cases out of all cases predicted as "Attrition". It indicates how reliable the model is when it predicts attrition.

   d. **Recall:** 0.34 - This measures the proportion of correctly predicted "Attrition" cases out of all actual "Attrition" cases. It indicates the model's ability to find all attrition cases.

   e. **ROC-AUC Score:** 0.81 - This represents the Area Under the Receiver Operating Characteristic (ROC) curve. It measures the model's ability to distinguish between the two classes (Attrition and No Attrition). A higher AUC score indicates better performance.

5. **Confusion Matrix:**

   a. **Chart Type:** Heatmap or Matrix

   b. **Insights:** Shows the number of correctly and incorrectly predicted instances.

      i. **True Negatives:** 238 (Correctly predicted "No Attrition")

      ii. **False Positives:** 9 (Incorrectly predicted "Attrition" when it was "No Attrition")

      iii. **False Negatives:** 31 (Incorrectly predicted "No Attrition" when it was "Attrition")

      iv. **True Positives:** 16 (Correctly predicted "Attrition")

   c. **Significance:** The confusion matrix provides a detailed breakdown of the model's performance for each class, helping to identify specific areas of improvement.

6. **ROC Curve:**

    a. **Chart Type:** Line Graph

    b. **Insights:** Plots the True Positive Rate (Recall) against the False Positive Rate at various threshold settings.

    c. **Significance:** The ROC curve visually represents the trade-off between sensitivity (recall) and specificity (1 - false positive rate). The AUC (0.81) indicates good overall performance.

Overall Conclusions & Recommendations:

- **Focus on Job Levels 2 and 1:** Investigate the reasons for high attrition at these job levels.

- **Address Issues Related to Life Sciences and Medical Backgrounds:** Understand the specific challenges faced by employees with these educational backgrounds.

- **Review Career Development and Job Satisfaction at the 1,5,10-Year Mark:** Implement strategies to retain long-term employees.

- **Understand Factors Contributing to Attrition in Younger Employees (26-35):** Focus on career development, work-life balance, and competitive compensation.

- **Investigate Gender-Specific Attrition Factors:** Identify and address any potential biases or issues affecting male employees.

- **Analyze Salary Data:** Determine if salary is a significant factor contributing to attrition.

Interpretation and Implications:

- **Model Performance:** The model has a relatively good accuracy (86%) and AUC (0.81), suggesting it's reasonably effective at predicting attrition. However, the F1-score (0.44) and recall (0.34) indicate that the model struggles to correctly identify all attrition cases. It might be biased towards predicting "No Attrition" due to the imbalanced dataset.

- **Class Imbalance:** The "Real Attrition Status" shows a significant class imbalance (83.88% No Attrition vs. 16.12% Attrition). This can affect the model's performance, especially its ability to predict the minority class (Attrition).

- **Focus on Recall:** Given the low recall, the model might need improvement in identifying employees who are likely to leave. This is crucial for proactive retention strategies.

- **Feature Importance:** While the dashboard provides average values for age, salary, and years, it doesn't show the importance of each feature in the model's predictions. Understanding feature importance can help identify key drivers of attrition.

Recommendations:

Address Class Imbalance: Consider using techniques like oversampling the minority class (Attrition) or undersampling the majority class (No Attrition) to improve model performance.

- **Feature Engineering:** Explore additional features or create new features from existing ones to improve the model's predictive power.

- **Model Tuning:** Experiment with different parameters of the Logistic Regression model or consider other classification algorithms to find a model with better recall.

- **Investigate False Negatives:** Analyze the 31 false negatives (employees predicted as "No Attrition" but actually experienced attrition) to understand why the model failed to identify them.

- **Feature Importance Analysis:** Determine the importance of each feature in the model to identify key drivers of attrition.