

# EXPLORATORY DATA ANALYSIS REPORT

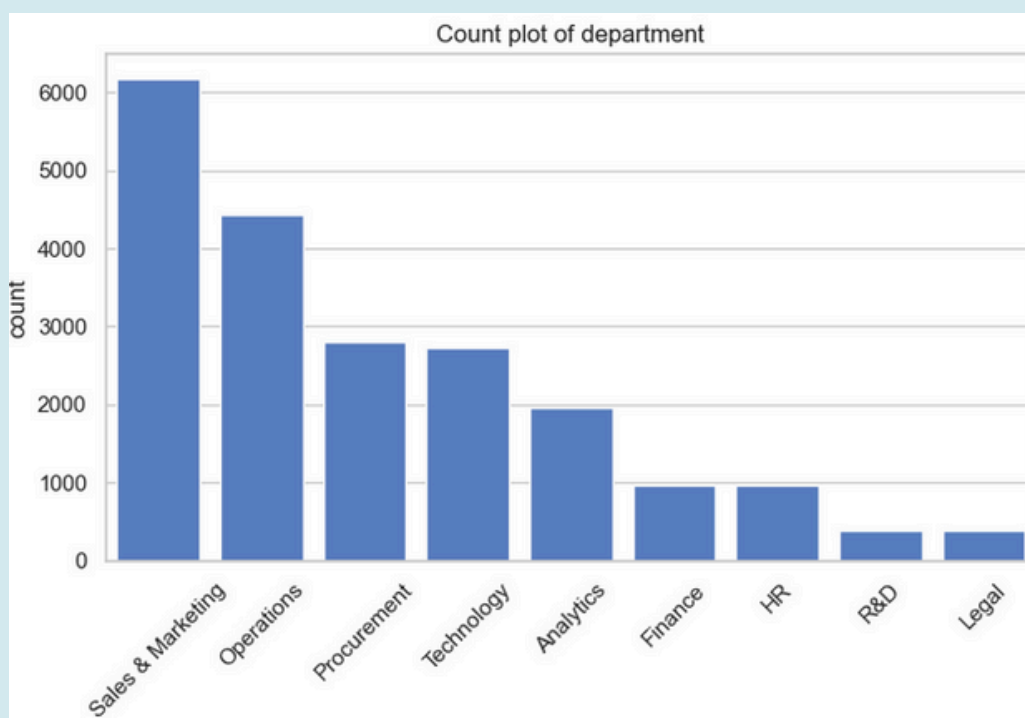
Categorical  
Features

# DEPARTMENT DISTRIBUTION

- The largest number of employees work in Sales & Marketing and Operations, together making up almost half the workforce.
- Medium-sized departments include Procurement, Technology, and Analytics.
- Legal and R&D are very small departments with fewer than 500 employees each.

## What this means:

- Promotions and attrition are likely to be driven by the larger departments, since they dominate the dataset.
- Smaller departments may not provide enough data for reliable predictions → risk of bias.

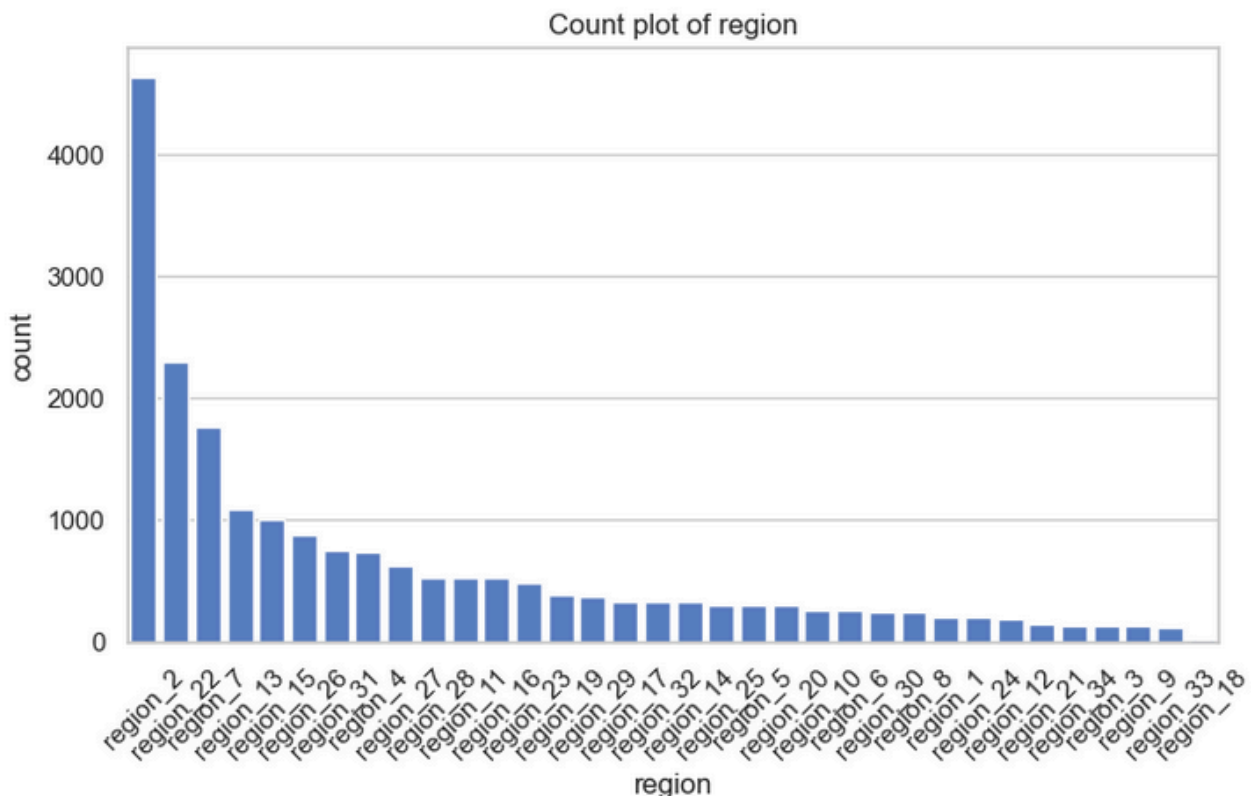


# REGION DISTRIBUTION

- A few regions (such as region\_2, region\_22, and region\_7) have thousands of employees.
- Most other regions have fewer than 500 employees.
- The distribution is highly skewed → some regions dominate while others are barely represented.

## What this means:

- Region could be an important factor for promotion (since policies and workforce size vary by location).
- However, because of the imbalance, the model may “learn more” from the larger regions and underperform for smaller ones.

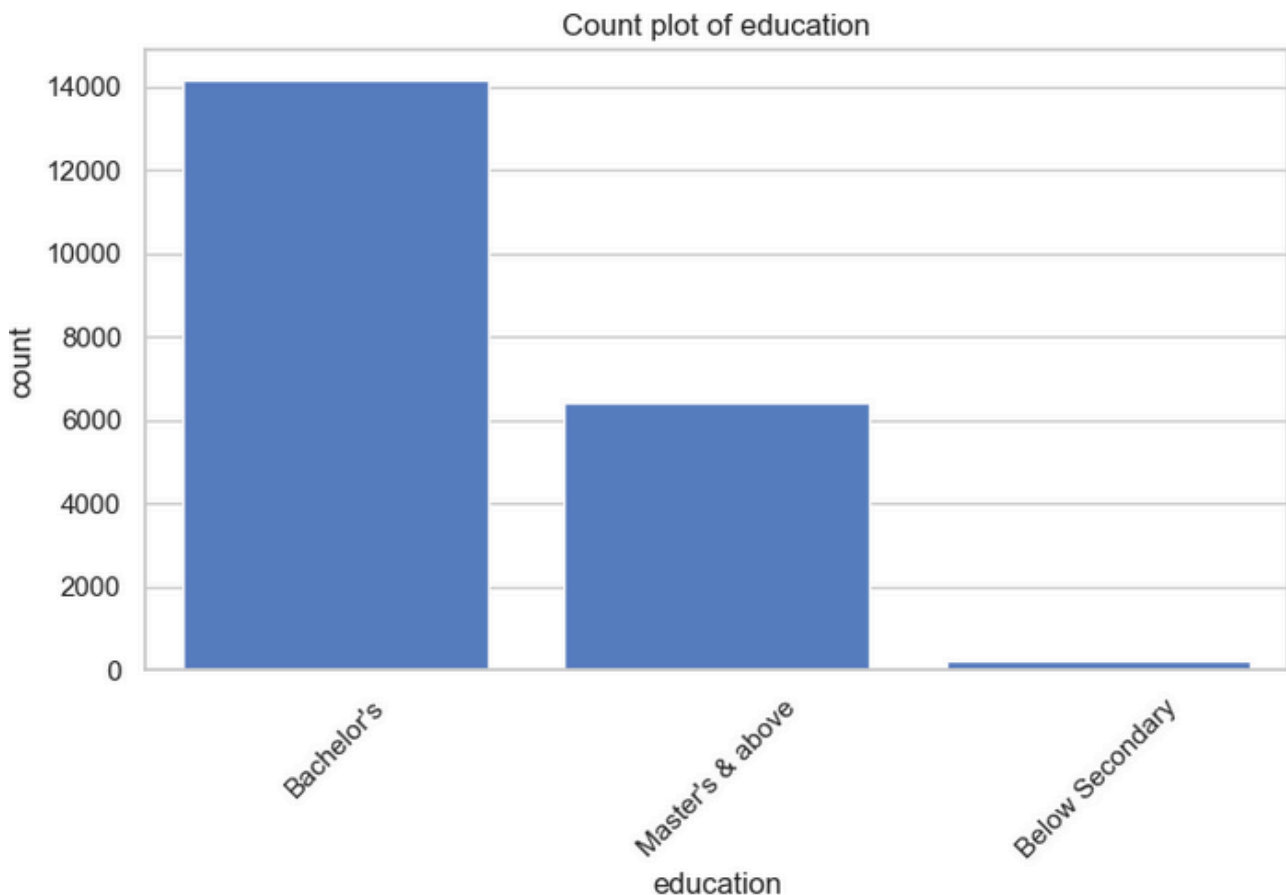


# EDUCATION LEVELS

- The majority of employees hold a Bachelor's degree (over 14,000).
- Around 7,000 employees have a Master's degree or higher.
- Below Secondary education is extremely rare (fewer than 200).

## What this means:

- Since most employees are graduates, the dataset has low variation in education.
- Education may not have strong predictive power, though higher education might still correlate with seniority.

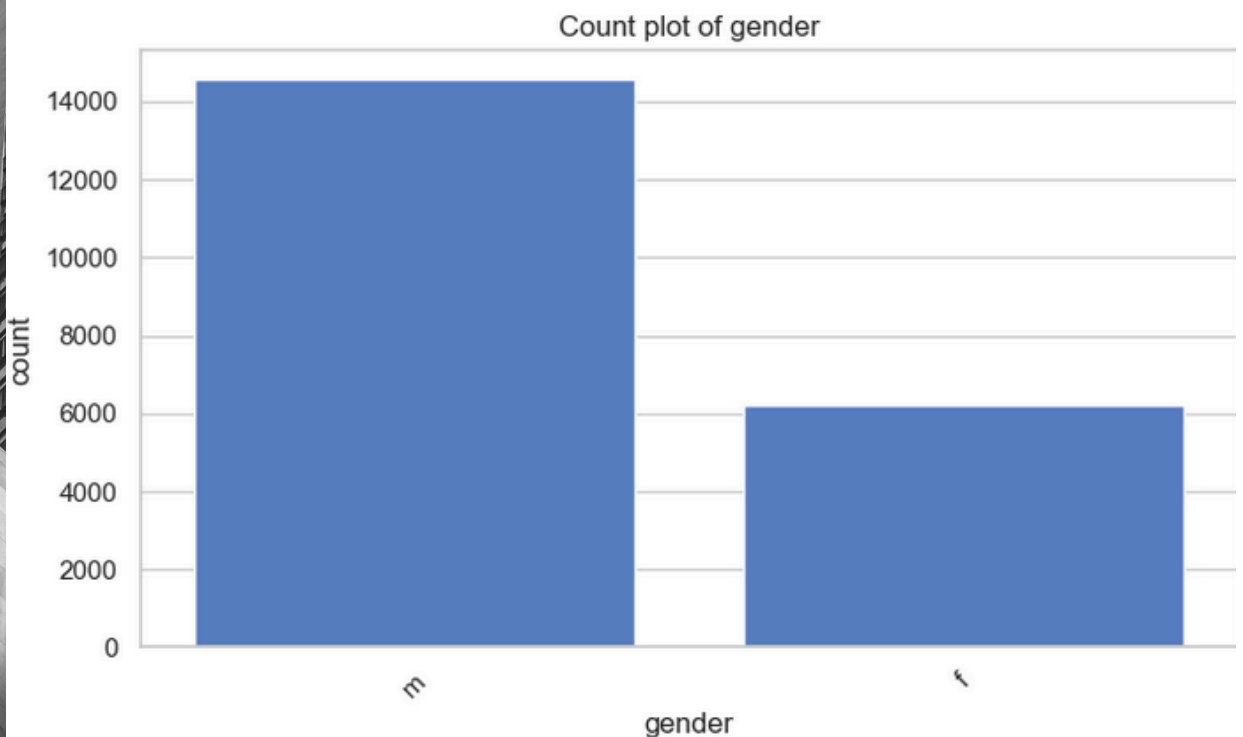


# GENDER DISTRIBUTION

- The company employs more than twice as many males as females.
- Approx. 14,000 males vs. 6,000 females.

## What this means:

- The workforce is male-dominated.
- Any analysis of promotions must consider fairness → otherwise, the model may reinforce gender imbalance.

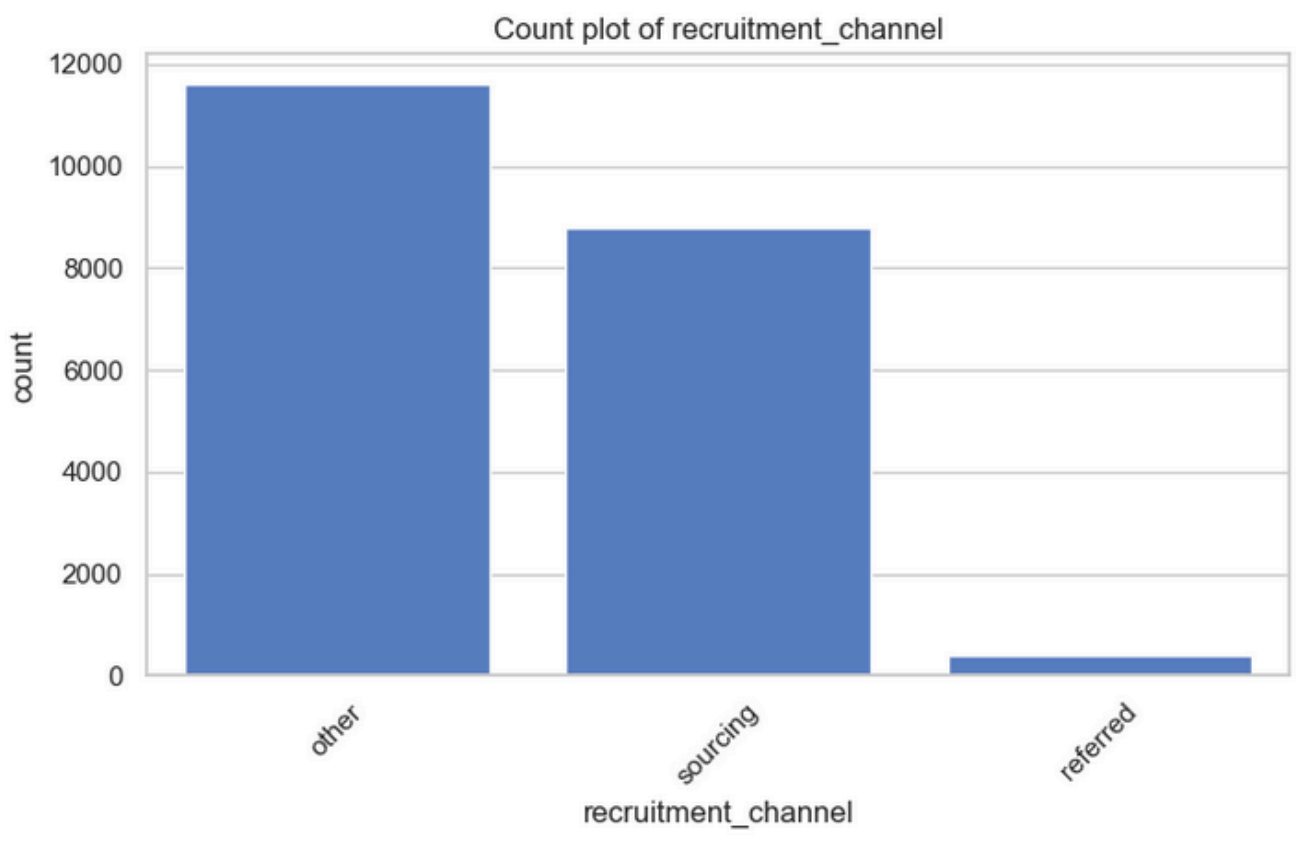


# RECRUITMENT CHANNEL

- Most employees were recruited through Other ( $\approx 11,000$ ) and Sourcing ( $\approx 9,000$ ).
- Referrals are very rare (fewer than 1,000).

## What this means:

- Recruitment channel distribution is uneven.
- This feature may still provide useful signals (e.g., referral hires might perform better), but it is not balanced.

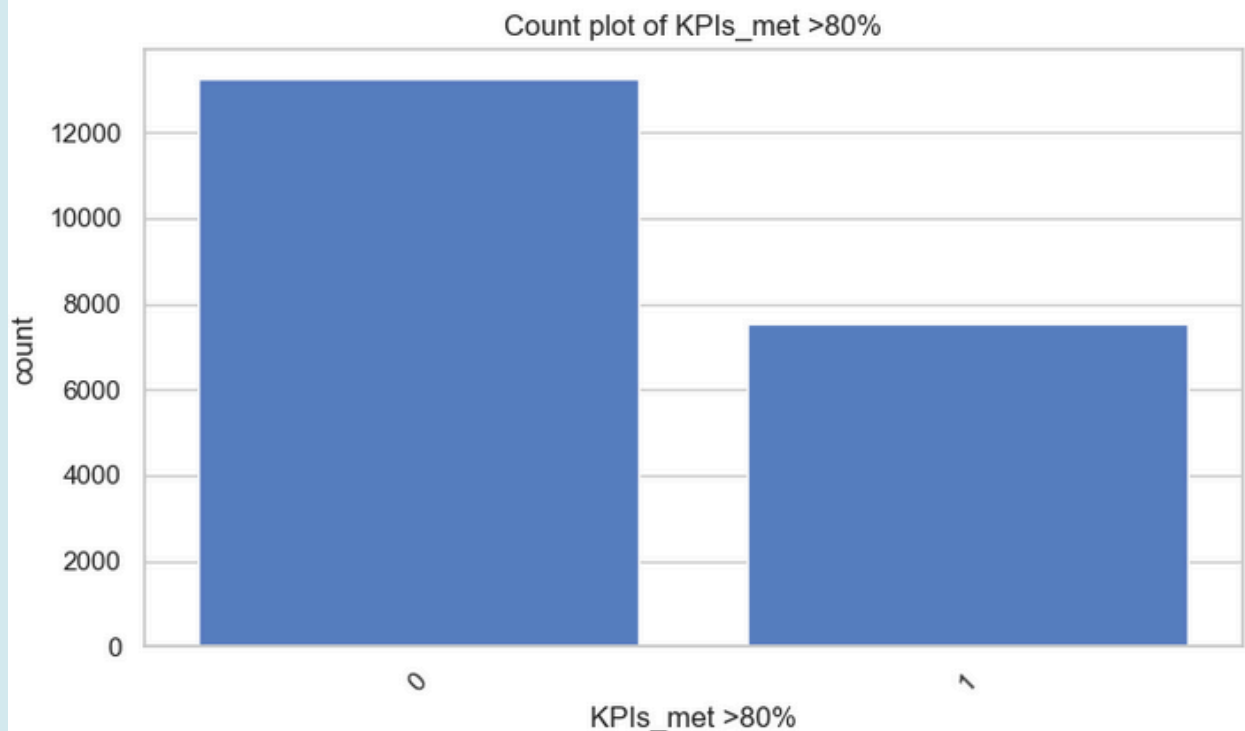


# KPI ACHIEVEMENT (>80%)

- A large number of employees did not meet KPI >80%.
- Around 7,500 employees did meet KPIs >80%, but the majority did not.

What this means:

- KPI performance is a direct measure of productivity and will likely be a key predictor for promotions.
- Imbalance shows that consistently high KPI achievement is relatively rare.

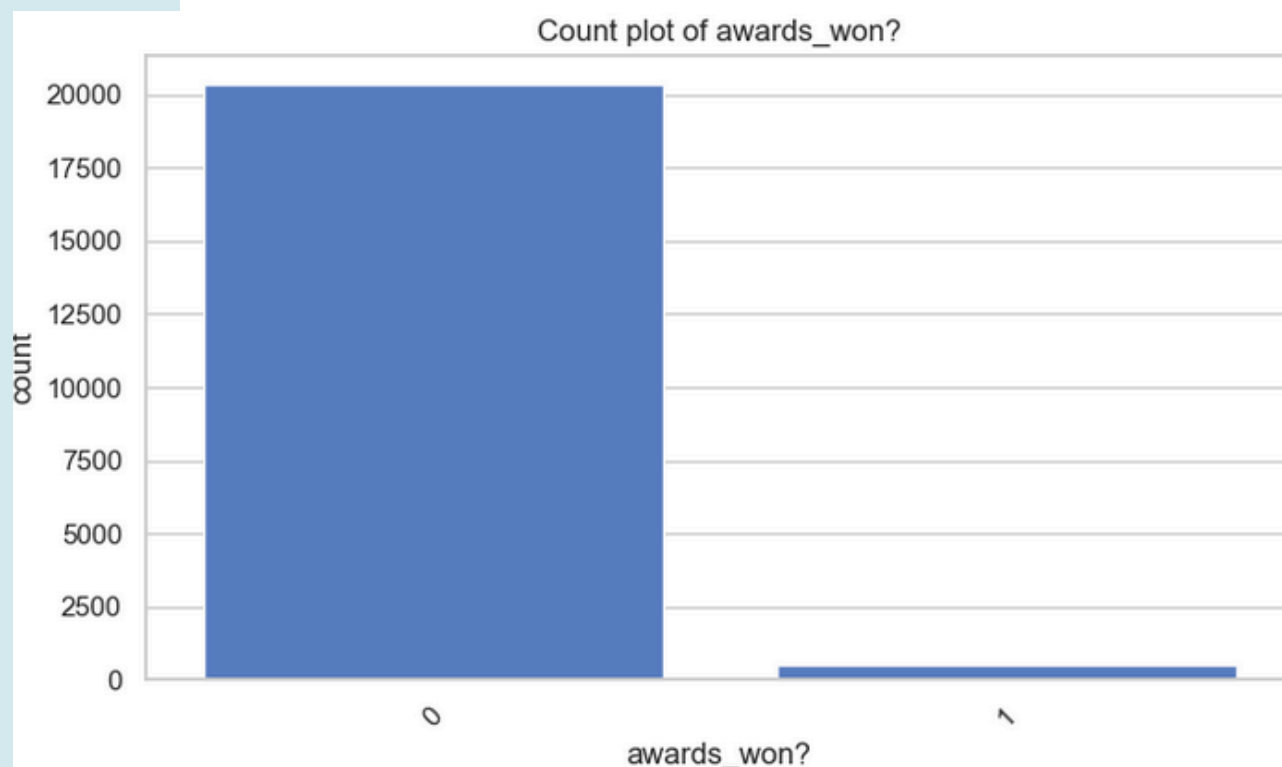


## 7. AWARDS WON

- Almost all employees have not won an award.
- Fewer than 1,000 employees have won awards.

### **What this means:**

- Awards are a very strong signal of excellence, even though rare.
- This feature is highly imbalanced, but those who have awards likely have much higher promotion chances.





# OVERALL INSIGHTS & RECOMMENDATIONS



*“Balancing key features and focusing on strong predictors like KPI achievement can improve model accuracy and fairness.”*



## 1. Imbalanced Features:

- Department, region, gender, and awards show clear imbalance.
- These imbalances may bias the ML model if not handled (e.g., through balancing techniques or careful feature engineering).

## 2. Strong Predictors:

- KPI Achievement and Awards Won → directly linked to performance, strong signals for promotions.

## 3. Moderate Predictors:

- Department and Recruitment Channel → important for context, but heavily imbalanced.

## 4. Weak Predictors:

- Education → limited variability, most are Bachelor's.
- Gender → may be more of a fairness check than a true predictor.

## 5. Business Implications:

- Company is heavily reliant on Sales & Marketing and Operations.
- Workforce is male-heavy and regionally concentrated.
- High performance (KPIs + awards) is rare but crucial for promotions.

