DS612: Interactive Data Visualization

Course Project

Winter Semester 2025

Submitted By

Meenakshi Iyer : 202418031
Bhavin Gurnani : 202418043
Nupoor Londhe : 202418040
Priyanshi Saini : 202418002

Instructor : Bhaskar Chaudhury

HUMAN RESOURCE ANALYSIS

10-04-2025–

TEAM ID : 15

DATASET : Human Resource Dataset (DS15)

INDEX

INTRODUCTION

Overview

This HR analytics project focuses on uncovering patterns in employee performance, pay equity, and termination to support more informed and strategic HR decisions.

Goals

This HR data analysis project aims to investigate key workforce trends to support data-driven decision-making. The primary objectives are:

Evaluate Termination Risks – Identify if lower performance scores or higher absenteeism correlate with employee termination..

Assess Gender Pay Equity – Determine if there is a significant pay disparity between male and female employees.

Analyze Manager Impact on Performance – Examine whether an employee's performance varies based on their assigned manager.

Explore Recruitment Influence on Performance – Investigate if the recruitment source (e.g., job boards, referrals, agencies) affects employee performance scores.

Narrative and Storytelling Approach

The report will unfold as an investigative story into workforce dynamics, structured as follows:

1. Introduction & Business Context

HR analytics plays a vital role in **enhancing workforce efficiency, performance, and strategic data-driven decision making**. It involves leveraging data to understand, predict, and improve employee-related outcomes ultimately driving organizational success.

HR teams can:

1. Improve Employee Retention

2.Ensure Pay Equity & Compliance

3. Enhance Manager Effectiveness

4. Optimize Recruitment Strategies

5.Boost Workforce Productivity

6. Improving Diversity

Hence, Asking few potential questions which can be answered using the data visualization performed can facilitate the HR analysis are as follows:

1.Does marital status influence employee attrition rates, particularly among married females?

2.How does an employee's performance score correlate with their salary, and is there a gender-based pay gap?

3.Is there a significant relationship between an employee's recruitment source and their performance score?

2. Data Visualization Key Findings

- **Termination Analysis:**
  - Visualize the relationship between performance scores, absenteeism, and termination rates.
  - Identify thresholds where employees are at higher risk.
- **Gender Pay Analysis:**

- Compare average salaries by gender, adjusting for role, experience, and performance.
- Highlight any unexplained disparities.

- **Manager Impact Analysis:**
  - Group employees by managers and compare average performance scores.
  - Identify managers with unusually high/low-performing teams.

- **Recruitment Source Analysis:**
  - Compare performance scores across recruitment channels (e.g., LinkedIn, referrals, campus hires).
  - Determine which sources yield the highest-performing employees.

3. Insights & Recommendations

- **For Retention:**
  - **Identify at-risk employees** (low performance/high absenteeism) for retention strategies and Proactive coaching for employees near performance/absence risk thresholds.

- **For Pay Equity:**
  - **Assess pay fairness** and recommend adjustments if gender disparities exist and Address any unjustified pay gaps through structured salary reviews.

- **For Leadership Development:**
  - **Evaluate manager effectiveness** to guide leadership development programs and
    Provide training for managers with underperforming teams.

- **For Hiring Strategy:**

    ○ Optimize recruitment strategies by identifying the best talent sources based on performance trends.

This analysis will help HR leaders make **evidence-based decisions** to improve retention, fairness, and overall workforce productivity.

DATASET DESCRIPTION

The raw HR Dataset contains 311 rows in which each row represents an individual employee record and 36 columns where each column is a variable associated with that employee.

Link : https://www.kaggle.com/datasets/rhuebner/human-resources-data-set

HR Dataset Column Classification

1. Categorical Columns (Object/String Types)

These columns contain non-numeric data, used for classification or grouping.

• Employee_Name: Full name of the employee.

• Position: Job title held by the employee.

• State: US state abbreviation where the employee is located.

• Sex: Gender of the employee (M/F).

• MaritalDesc: Marital status (e.g., Single, Married).

• CitizenDesc: Citizenship status (e.g., US Citizen).

• HispanicLatino: Ethnic background (Yes/No).

• RaceDesc: Racial category (e.g., White, Black).

• TermReason: Explanation for termination.

• EmploymentStatus: Current employment status (e.g., Active, Terminated).

• Department: Department where the employee works.

• ManagerName: Full name of the direct manager

• RecruitmentSource: Source from which the employee was hired (e.g., LinkedIn, Indeed)

• PerformanceScore: Text description of performance (e.g., Fully Meets, Exceeds)

2. Numerical Columns (Int/Float Types)

These columns contain numeric data, used for ranking, encoding or state value.

• EmpID: Unique numeric identifier for each employee.

• MarriedID: Binary indicator for marital status (0 = not married, 1 = married).

• MaritalStatusID: Encoded form of marital status.

• GenderID: Encoded gender (0 = Female, 1 = Male).

• EmpStatusID: Encoded employment status.

• DeptID: Encoded department.

• PerfScoreID: Encoded performance score.

• FromDiversityJobFairID: Binary indicator if hired through a diversity job fair (0/1).

• Salary: Annual salary of the employee

• TermD: Binary flag used for termination and active status (0 = active, 1 = terminated).

• PositionID: Encoded job title

• Zip: Zip code (numeric, but categorical in analysis)

• EmpSatisfaction: Self-reported employee satisfaction (1 to 5 scale)

• SpecialProjectsCount: Number of special projects the employee worked on

• DaysLateLast30: Days the employee was late in the last 30 days

• Absences: Total number of days employee was absent.

• ManagerID: Encoded manager ID .

• EngagementSurvey: Employee engagement survey score (float between 0–5).

• DOB: Date of birth

• DateofHire: The date the employee was hired

• DateofTermination: The date the employee was terminated (if applicable)

• LastPerformanceReview_Date: Last recorded date of performance evaluation.

HR Dataset New Derived Column

• work_life_blance: Estimate how well employees are managing their work alongside personal life using indirect indicators from the dataset.

• Employee_tenure: the number of years the employee has worked in the company.

HYPOTHESIS

Following are the hypothesis we will be working with:

H1: Termination Analysis

Employees with lower performance or higher absence are most likely to get terminated.

H2: Gender Pay Analysis

Females are paid less compared to male workers.

H3: · Manager Impact Analysis

Performance of an Employee can be dependent upon the manager they are working under.

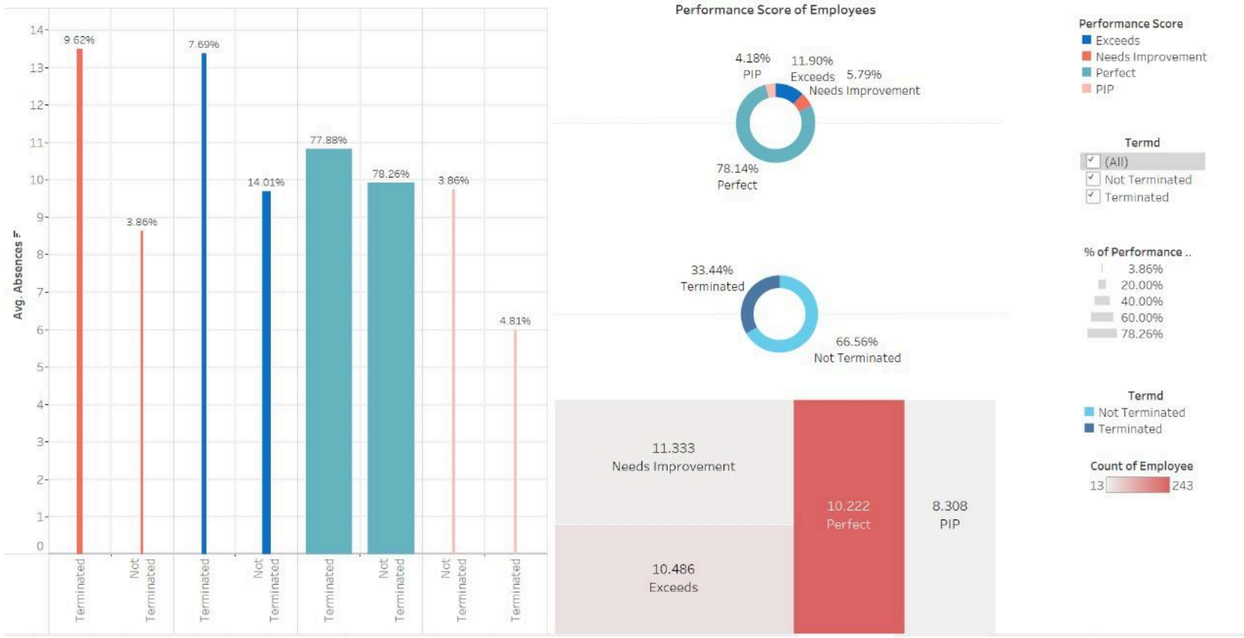
H4:· Recruitment Source Analysis

Performance score of an employee can be associated from where the employee has been recruited.

H1: Termination Analysis

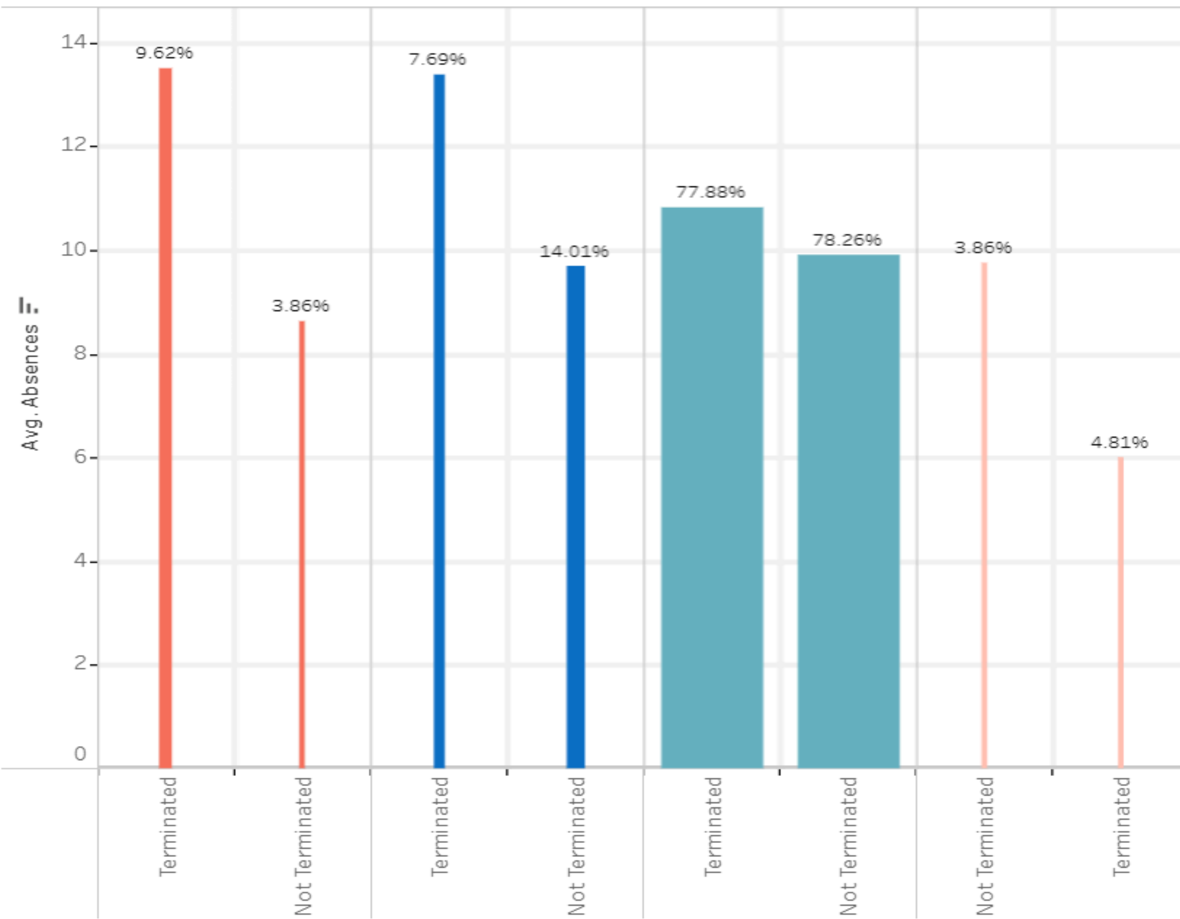
Null Hypothesis (H_0): Absenteeism and performance score are not significantly related to an employee's likelihood of termination.

Alternative Hypothesis (H_1): Absenteeism and performance score are significantly related to an employee's likelihood of termination



This dashboard presents a combined view of employee absenteeism, performance scores, and termination status using a bar chart, pie charts, and a tree map. These visual tools allow us to examine patterns that may indicate predictors of termination, aligning with our hypothesis.

1. Average Absences by Termination and Performance Score (Bar Chart)



1. The height of the bars represents the average number of absences.

2. The width of the bars indicates the proportion of employees in each performance category within the "Terminated" and "Not Terminated" groups.

Why ?

Why is it used? It offers a clear, quantitative comparison between the number of terminated vs non-terminated employees across absenteeism levels or performance brackets.

Why not something else? A stacked bar chart was cluttered with information making it hard to compare the part to whole relationship of these performances

Key Observations:

- Terminated employees consistently show higher average absences across performance categories compared to non-terminated ones.

E.g., “Needs Improvement” terminated employees have the highest absences (~9.62), versus ~3.86 for their non-terminated counterparts.

- The “Perfect” performance group is the most populated in both employee status groups, but terminated employees with this score still have higher absenteeism (~14.01) than the not-terminated ones (~13).
- A notable pattern appears in the PIP (Performance Improvement Plan) group:

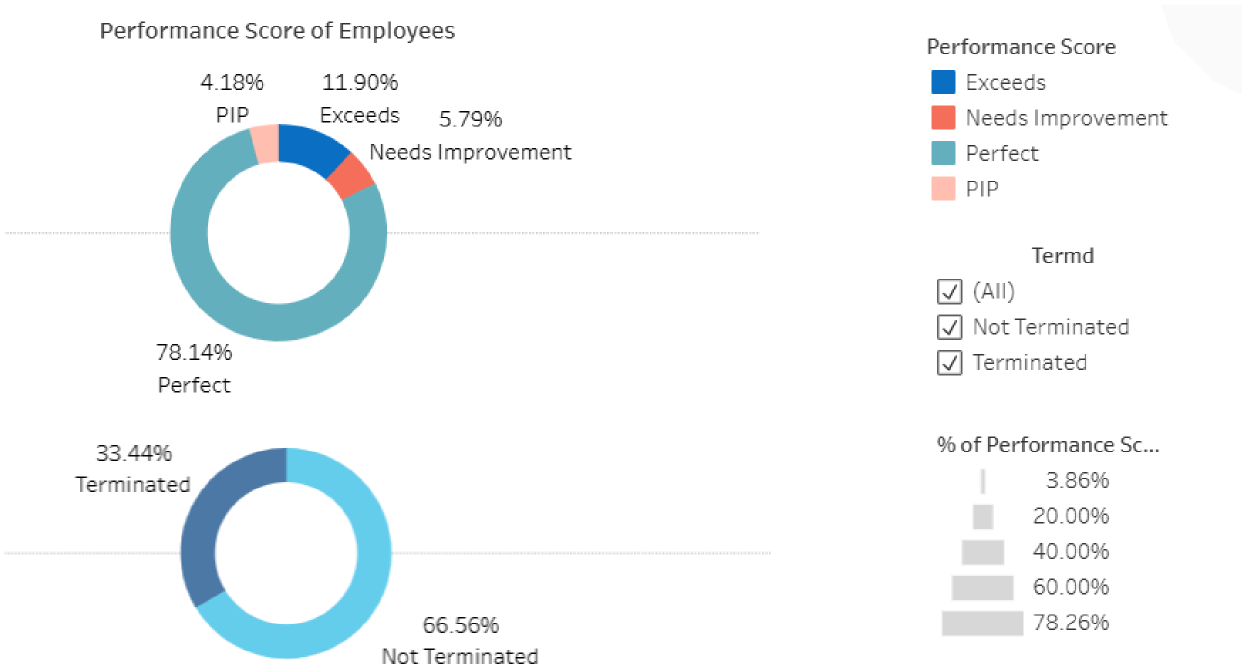
Terminated PIP employees have lower absences (~4.81) compared to non-terminated PIP employees (~8.3), suggesting that while absenteeism is a factor, poor performance alone can also lead to termination.

Marks and Channels :

Marks : 2-D, since we have used both length and width of the bar chart to encode the channels, the mark type then becomes 2-D.

Channel: Vertical Position (Avg. Absences), Horizontal width(Percentage of employee belonging to particular performance category), Hue(Performance Score)

2. Pie Charts: Performance Distribution and Termination Status:



The first pie chart shows that the majority of employees fall under the "Perfect" performance category (78.14%), followed by smaller proportions in "Exceeds," "Needs Improvement," and "PIP."

The second pie chart indicates the overall termination rate is 33.44%, while 66.56% of employees remain active.

- These charts emphasize that even among a workforce where "Perfect" performance is common, a sizable portion still face termination — indicating other contributing factors such as absenteeism.

Why ?

Why is it used? :Performance score distribution across the entire dataset.Proportion of terminated vs retained employees.

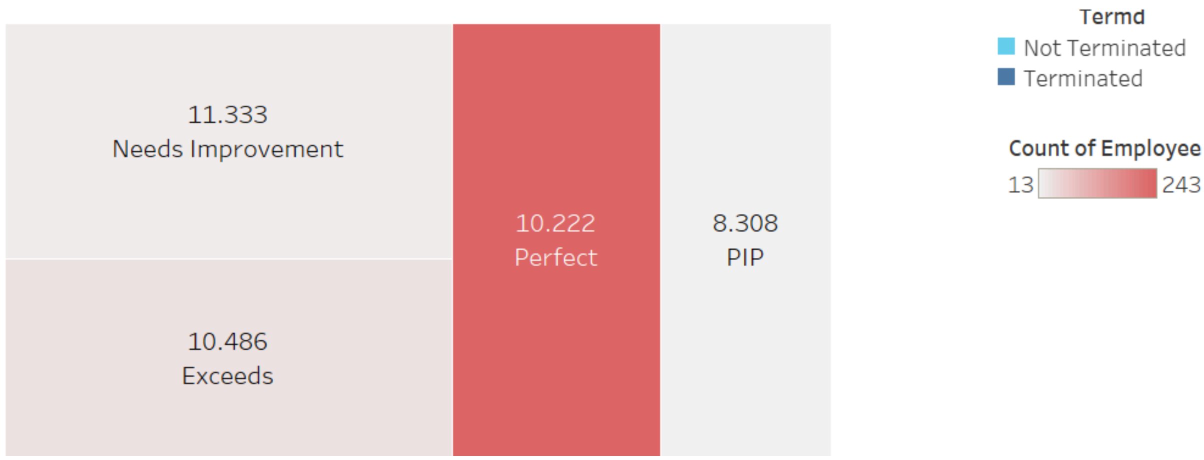
Why not a stacked bar chart? Pie charts are better since they don't take up a lot of space and since we are dealing with only 1 categorical attribute, pie charts are better preferred

Marks and Channels :

Marks : 2-D, When looking at pie chart/ donut chart instead of looking at the angles, areas covered by certain categories are focused on.

Channel: angle(percentage), hue(performance score, termination)

3. Tree Map: Absenteeism and Employee Count by Performance Category:



Each block is sized by avg. Absence and shaded by the number of employees, and labeled with the average number of absences within each performance category.

- Needs Improvement and Exceeds show relatively higher absenteeism (11.3 and 10.5 respectively), while PIP averages 8.3.
- Despite being high-performing on paper, Perfect performers average 10.2 absences, indicating that absenteeism may still play a hidden role even in this group.

Why?

Why is it used? A tree map helps visualize the hierarchical relationship and proportions between groups (e.g., performance score → termination).

Why not a stacked bar? Stacked bars get cluttered with too many categories. Tree maps provide an area-based representation, making it easier to detect dominant groups.

Uniqueness: It combines hierarchy + proportion, great for showing nested insights (like performance within termination status).

Interpretation: Does the Data Support Our Hypothesis?

Yes — the visual insights support the alternative hypothesis.

- There is a clear relationship between absenteeism and termination, particularly in the "Needs Improvement" and "Exceeds" groups.
- Higher absenteeism aligns with a greater likelihood of being terminated, especially when paired with lower performance scores.
- The anomaly in the PIP group, where non-terminated employees have higher absences, suggests that performance concerns alone can drive retention or separation decisions — meaning the company might tolerate some absenteeism if performance is at risk.

Marks and Channels :

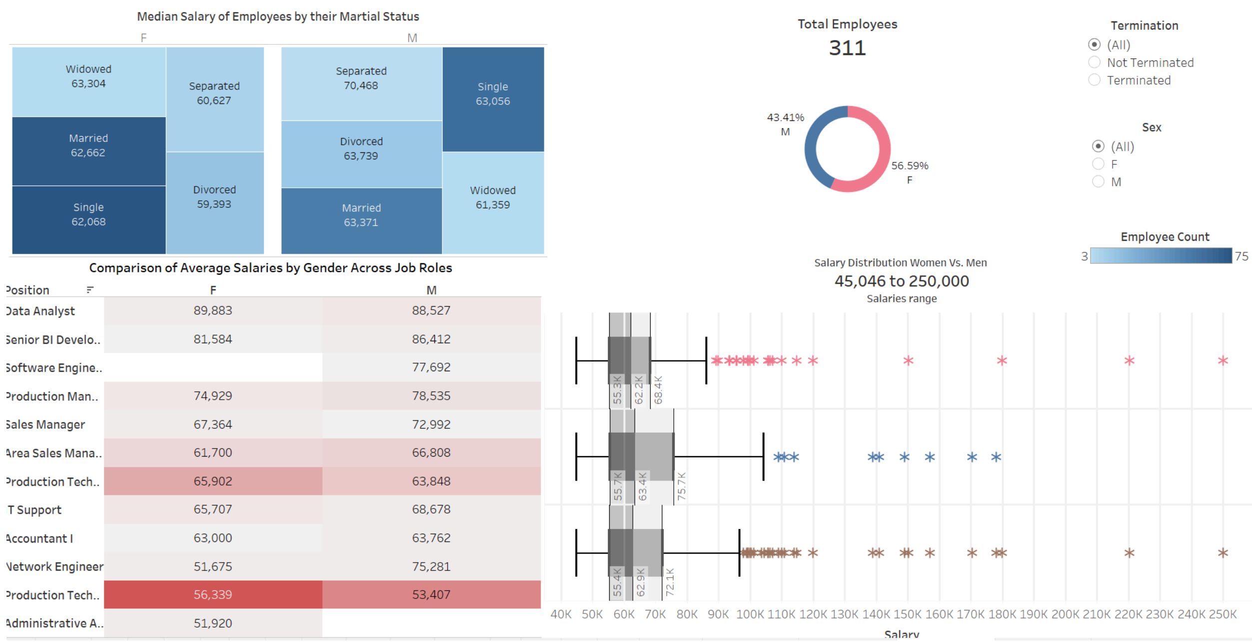
Marks : 2-D.

Channel: Saturation(Count of employee), Text(Avg. Absences, Performance score), Area (Avg. Absences)

H2: Gender Pay Analysis

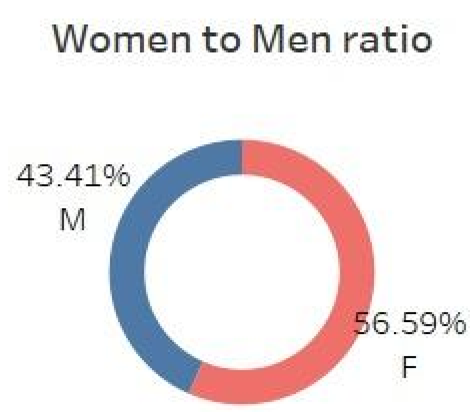
Null Hypothesis (H): There is no significant salary pay gap between men and women in the workforce.

Alternative Hypothesis (H): There is a significant salary pay gap between men and women in the workforce.



This dashboard gives a thorough look of gender based salary disparity and helps us find reasons for the same .

1.Total Employees & Gender Ratio



Female Employees: 56.59% (~176)

Male Employees: 43.41% (~135)

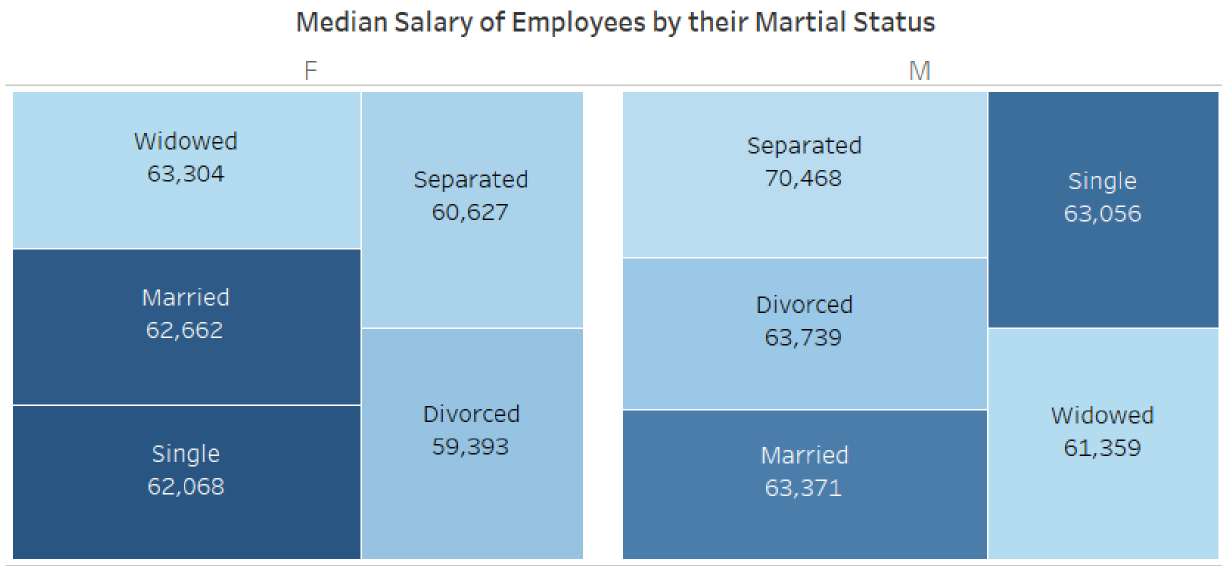
Women form the majority of the workforce. This is essential context: any salary analysis must factor in that there are more women overall, which can influence medians and distributions.

Marks and Channels :

Marks : 2-D, When looking at pie chart/ donut chart instead of looking at the angles, areas covered by certain categories are focused on.

Channel: angle(percentage), hue(Sex)

2.Median Salary by Gender and Marital Status (Tree Map)



Key Features:

Visualized as a tree map split by Gender (F & M) and Marital Status.

Box Size: Represents median salary for each subgroup.

Color Intensity: Indicates the number of employees in that group.

Why?

Why is it used? Reveals how gender-based salary patterns shift across marital statuses — a layer that's often overlooked.

Why not a group bar? Group bars would be cluttered and harder to read with too many marital status categories.

Female Employees:

Marital Status	Median Salary
Widowed	$63,304
Separated	$62,716
Married	$62,134
Single	$61,621
Divorced	$59,393

Observation: Slight upward trend with age or experience (widowed/separated earning more).

Spread is narrow → salaries among females are relatively consistent across marital statuses.

Male Employees:

Marital Status	Median Salary
Separated	$70,468
Married	$67,899
Divorced	$64,272
Single	$62,889
Widowed	$61,359

Insight:

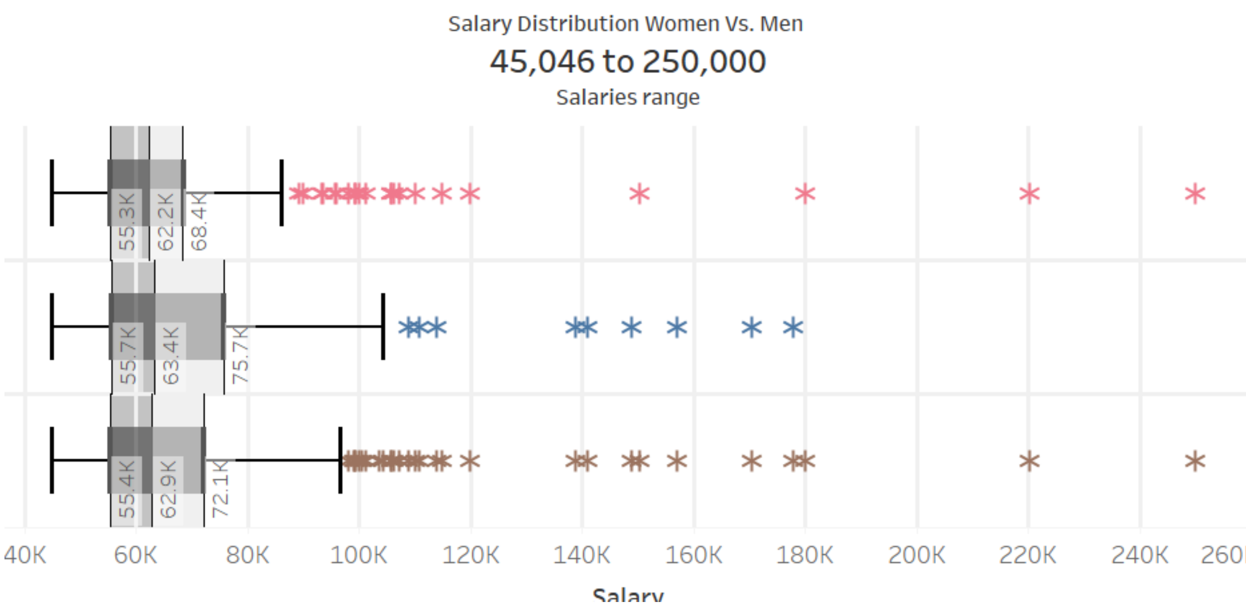
While males tend to earn slightly more, the gap across marital statuses is not extreme. However, separated males appear to earn significantly more, hinting at contextual factors like age, seniority, or role.

Marks and Channels :

Marks : 2-D, When looking at pie chart/ donut chart instead of looking at the angles, areas covered by certain categories are focused on.

Channel: Text(Marital Status and median salary), Saturation(Count of employee), Area(Median Salary)

3. Salary Distribution by Gender (Box Plot)



This visualization compares the spread, quartiles, and outliers in salaries between genders.

Women:

Q1: $55,315

Median: $62,226.5

Q3: $68,407

Min: $45,046

Max (Outlier): $250,000

Men:

Q1: $55,650

Median: $63,353

Q3: $75,655

Min: $45,115

Max: $178,000

Key Observations:

Median Salary: Men > Women by ~$1,127

IQR: Women = $13,092 vs. Men = $20,005 → Men's salaries are more spread out

Outliers:

Women have one extreme outlier ($250K)

Men have more high salaries, but none as extreme

Important Insight:

At first glance, women appear to earn less because their median salary is lower. However, this does not necessarily indicate discrimination.

Why?

Why is it used? Tables give precise numeric insight, showing mean, and comparing values side-by-side.

Why not just bar charts? Bar charts are great, but tables provide accuracy and support for detailed interpretation.

Marks and Channels :

Marks : 1-D,The length of the interquartile range matters and changing that would result in changing the underlying meaning of the channel.

Channel: Hue(Sex), Horizontal position(Salary)

Deep Dive: Why is the Female Median Salary Lower?

Explanation Using the Heatmap (Bottom-Left: Avg Salary by Role & Gender)

This visual shows the average salary for men and women in each job role. It reveals role-based salary patterns, which provide the true cause behind the median salary difference.

Roles Where Women Earn More:

Role	Women Avg	Men Avg
Data Analyst	$89,883	$88,527
Production Technician II	$65,902	$63,848
Production Technician I	$56,339	$53,407

In these roles, women slightly out-earn men — a positive indicator of role-level fairness.

Roles Where Men Earn More:

Role	Men Avg	Women Avg
Network Engineer	$75,281	$51,675
Sales Manager	$81,500	$75,800
Production Manager	$73,438	$69,838
Senior BI Developer	$91,992	$87,188

In some technical or senior roles, men earn noticeably more, especially in Network Engineering (a ~$24K gap).

Critical Finding:

The real reason for the lower female median is not unequal pay but unequal role distribution:

- More women are employed in lower-paying roles like Technician I/II.
- Fewer women are found in high-paying technical/managerial roles like Network Engineer or BI Developer.
- This skews the overall median downward for women — a composition effect, not necessarily discrimination.

Conclusion: Hypothesis & Statistical Interpretation

- Null (H_0): No significant difference in median salaries between men and women.
- Alternative (H_1): There is a significant difference.

Based on Visual Analysis:

- Median difference (~$1.1K) is small, suggesting we might fail to reject the null in a statistical test.
- However, distributional spread (more high salaries among men) shows greater upward mobility.
- Outliers are rare but notable — a $250K female salary could skew perceptions unless statistically controlled.

Final Insights Summary

1. Female Median Salary is Lower — But Not Due to Lower Pay for Equal Work

Instead, it's because more women occupy lower-paying roles.

When compared role-by-role, some roles pay women more.

2. Role Distribution Drives Pay Gaps

Women are underrepresented in high-paying technical roles (e.g., Network Engineer).

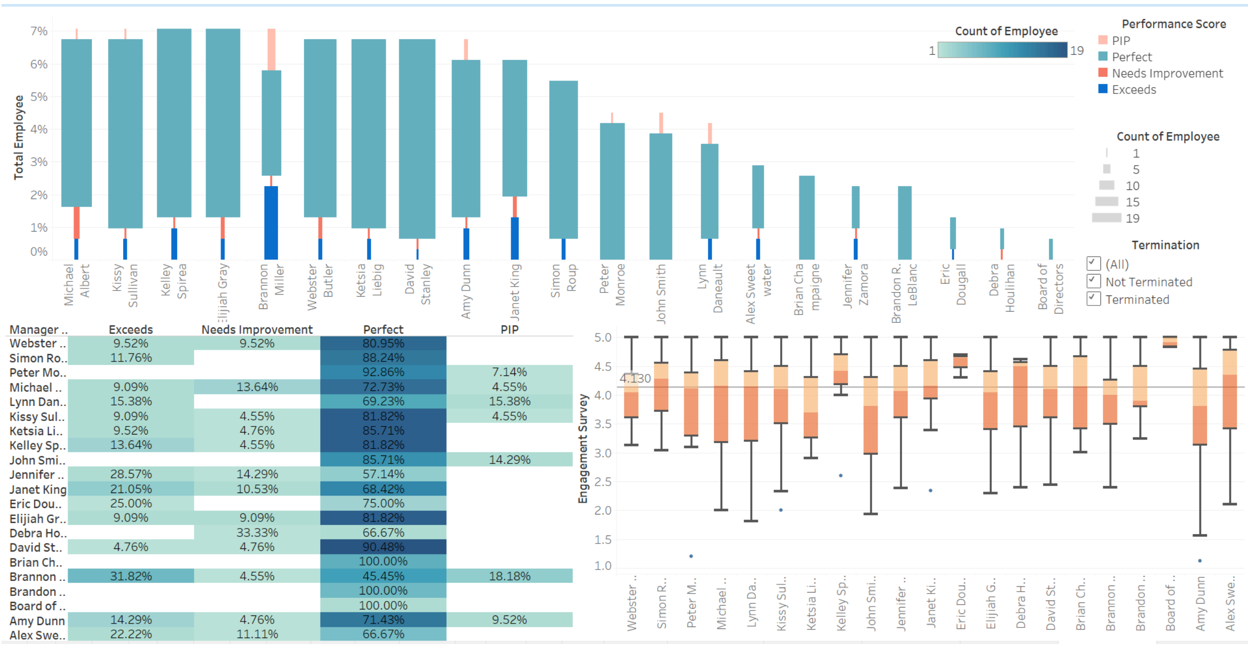
Men are more common in roles with greater salary dispersion.

3. Consistency vs. OpportunityWomen's salaries are more consistent with less variance, but also fewer high earners.Men's salaries show greater variance, indicating more mobility (but also inequality within the group).

H3: Manager Impact Analysis

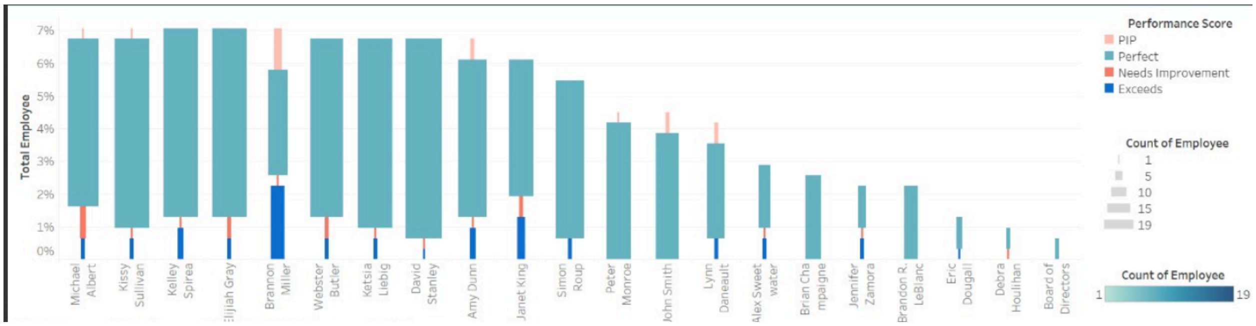
Null Hypothesis (H_0): There is no significant relationship between an employee's performance and the manager they work under.

Alternative Hypothesis (H_1): There is a significant relationship between an employee's performance and the manager they work under. (Manager does have an effect on employee performance.)



This dashboard offers a comprehensive, data-driven view into how managerial leadership correlates with employee performance and engagement across the organization.

1. Performance Score Distribution by Manager (Top Bar Chart):



- Distribution of **employee performance scores** (`Exceeds`, `Perfect`, `Needs Improvement`, `PIP`) across different **managers**.

- The bar height represents the **percentage of total employees**, and color segments show performance breakdown, while the width of the bars show the percentage of employees belonging to that performance category.

Why ?

- **Stacked bars** are ideal for comparing proportions of categories within groups (in this case, performance categories per manager).

- Makes it easy to spot managers who lead **high-performing vs. low-performing** teams.

Insight & Relevance to Hypothesis:

- Managers like **Peter Monroe**, **Michael Alberti**, and **Simon Group** have a larger proportion of "Perfect" and "Exceeds" performers.

- In contrast, **Brannon Miller** and **Brandon R. LeBlanc** show higher shares of "Needs Improvement" or "PIP".

- Suggests a **clear variation in team performance based on the manager**, supporting the hypothesis.

Marks and Channels :

Marks : 2-D,since we have used both length and width of the bar chart to encode the channels, the mark type then becomes 2-D.

Channel: Hue(Performance score), Vertical position(Percentage of employee)

2. Heatmap Table (Middle Section):

Manager ..	Exceeds	Needs Improvement	Perfect	PIP
Webster ..	9.52%	9.52%	80.95%	
Simon Ro..	11.76%		88.24%	
Peter Mo..			92.86%	7.14%
Michael ..	9.09%	13.64%	72.73%	4.55%
Lynn Dan..	15.38%		69.23%	15.38%
Kissy Sul..	9.09%	4.55%	81.82%	4.55%
Ketsia Li..	9.52%	4.76%	85.71%	
Kelley Sp..	13.64%	4.55%	81.82%	
John Smi..			85.71%	14.29%
Jennifer ..	28.57%	14.29%	57.14%	
Janet King	21.05%	10.53%	68.42%	
Eric Dou..	25.00%		75.00%	
Elijiah Gr..	9.09%	9.09%	81.82%	
Debra Ho..		33.33%	66.67%	
David St..	4.76%	4.76%	90.48%	
Brian Ch..			100.00%	
Brannon ..	31.82%	4.55%	45.45%	18.18%
Brandon ..			100.00%	
Board of ..			100.00%	
Amy Dunn	14.29%	4.76%	71.43%	9.52%
Alex Swe..	22.22%	11.11%	66.67%	

- Detailed percentage breakdown of employee performance categories **per manager** in a tabular format.

- Uses color intensity to show **higher or lower values**.

Why ?

- Heatmaps are effective for spotting patterns across dimensions — in this case, **which managers consistently have top-performing or underperforming employees**.

- Adds quantitative detail to the high-level view from the bar chart.

Insight & Relevance:

- For instance, **Eric Dougall** has a **high "Needs Improvement" (33.33%)** rate.

- **Brannon Champagne** has **31.82% "Exceeds"** — indicating strength in team performance.

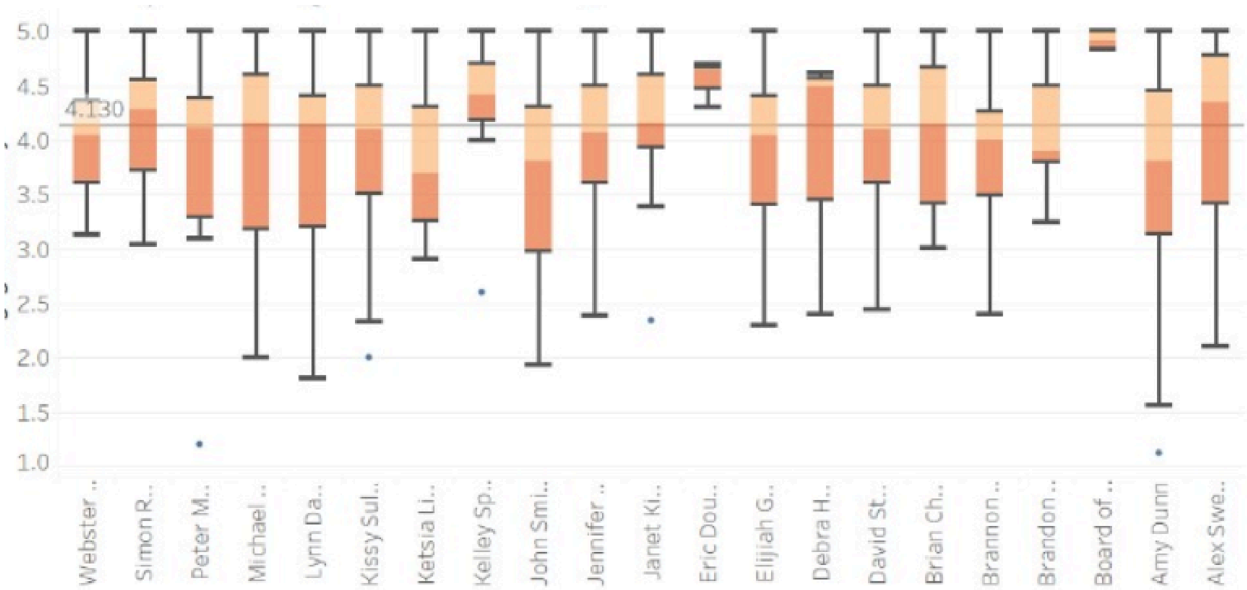
- This chart reinforces the **individual performance trends under specific managers**, further validating the hypothesis.

Marks and Channels :

Marks : 0-D.

Channel: Saturation(Count of Employees),Text(Percentage of employee working under a specific manager)

3. Box Plot (Bottom Right Section):



- **Engagement survey scores** distribution across teams managed by each manager.
- Displays **median**, **quartiles**, **range**, and **outliers** for each manager.

Why ?

- **Box plots** help visualize **variability and consistency** in engagement levels — a potential driver of performance.

- Useful to assess whether high-performing teams also report **higher engagement**.

Insight & Relevance:

- Managers like **Webster Butler** and **Board of Directors** have **higher median engagement** scores and tight spreads.

- Managers with lower-performing teams (e.g., **Amy Dunn**, **Alex Sweetwater**) show **lower and more variable** engagement.

- Demonstrates how **managerial influence may affect team engagement**, which in turn relates to performance.

Marks and Channels :

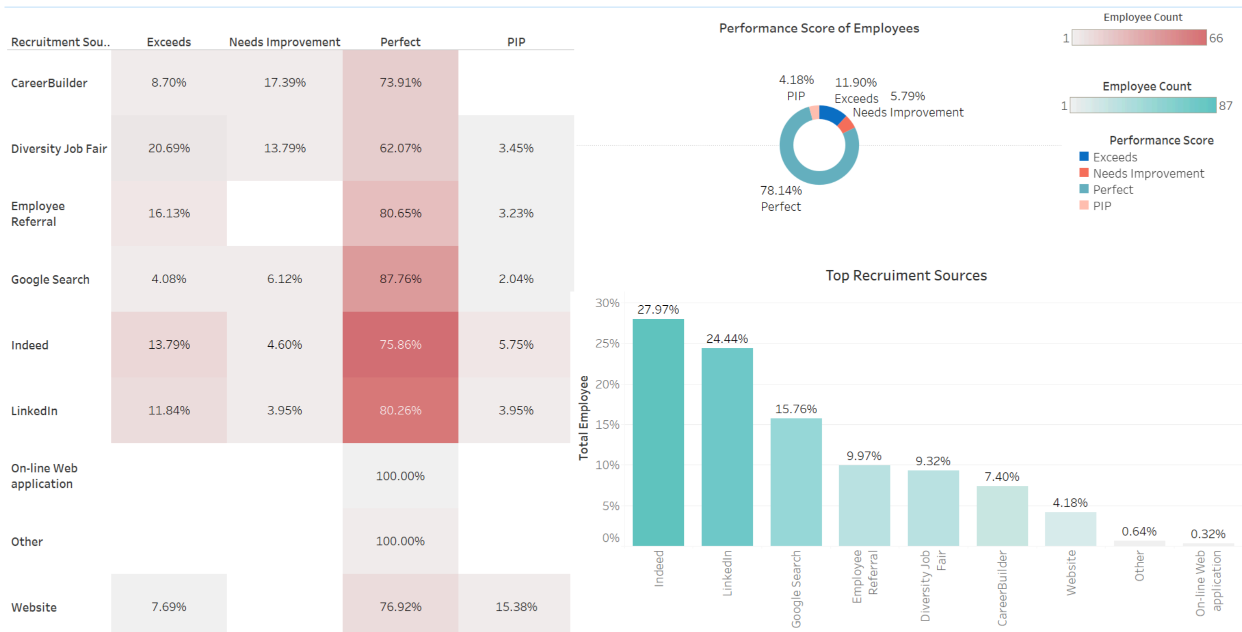
Marks : 1-D,The length of the interquartile range matters and changing that would result in changing the underlying meaning of the channel.

Channel: Vertical position(Salary)

H4: Recruitment Source Analysis

Null Hypothesis (H0): Employee performance is independent of the recruitment source. That is, the percentage of employees meeting or exceeding expectations does not significantly vary across different recruitment sources.

Alternative Hypothesis (HA): Employee performance depends on the recruitment source. That is, certain recruitment sources yield employees who perform better or worse than those from other sources.



The given dashboard is based on the distribution of sources of recruitment.

1. Table (Left): Recruitment Source vs. Performance

Recruitment Sou..	Exceeds	Needs Improvement	Perfect	PIP
CareerBuilder	8.70%	17.39%	73.91%	
Diversity Job Fair	20.69%	13.79%	62.07%	3.45%
Employee Referral	16.13%		80.65%	3.23%
Google Search	4.08%	6.12%	87.76%	2.04%
Indeed	13.79%	4.60%	75.86%	5.75%
LinkedIn	11.84%	3.95%	80.26%	3.95%
On-line Web application			100.00%	
Other			100.00%	
Website	7.69%		76.92%	15.38%

Color Coding:

The shading in each cell represents the number of employees recruited from that particular source. Darker red = more employees.

Percentages in Cells:

These represent the distribution of performance scores (Perfect, Exceeds, Needs Improvement, PIP) among only those recruited from that source. For example:

Out of all the employees recruited via LinkedIn, 80.26% are rated as having a “Perfect” performance score.

This allows us to not only see which sources perform well but also compare how consistently they produce strong performers.

Why?

Why is it used? Combines performance distribution with volume encoding via color (intensity of hiring). These are easily interpretable too.

Why not a bar plot? A bar plot wouldn’t allow us to cross-compare multiple performance categories across many sources simultaneously.

Insights per source:

- Google Search:87.76% of hires have perfect performance — highest among all.

Even though fewer people are hired from here, the quality is exceptional.

- LinkedIn:80.26% perfect, 3.95% needs improvement — very low underperformance rate.

Also one of the top hiring channels — 24.44% of total employees are from LinkedIn.

High volume + quality.

- Indeed:75.86% perfect performers.

Largest source of recruitment (27.97% of total employees).

Slightly lower quality than LinkedIn, but still effective.

- Employee Referral:80.65% perfect, low PIP.

Very reliable source though fewer hires (9.97%).

- Diversity Job Fair:62.07% perfect.

High “Exceeds” (20.69%) but also notable “Needs Improvement” (13.79%).A more mixed outcome.

- Online Web Application & Other:100% performance rating in either Perfect or Exceeds, but very low number of hires.

Can’t generalize due to small sample size.

Website & CareerBuilder:Moderate performance (~73-76% perfect).

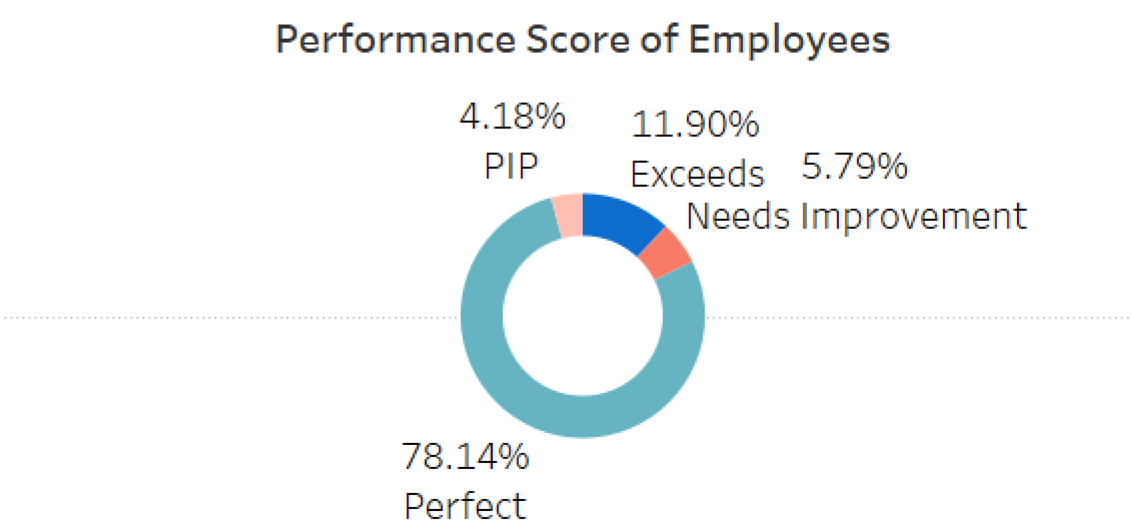
Not as strong as Indeed or LinkedIn.

Marks and Channels :

Marks : 0-D.

Channel: Text(Percentage of employee), Saturation(Count of employee)

2. Pie Chart (Top Right): Overall Performance Distribution



- 78.14% of all employees are rated as Perfect — indicating a generally high-performing workforce.
- 11.9% exceed expectations.

- 5.79% need improvement.
- 4.18% are on a Performance Improvement Plan (PIP).

This gives context to the overall baseline performance that helps us understand which sources exceed this average.

Why ?

Why is it used? Provides an overall view of performance distribution across the company.

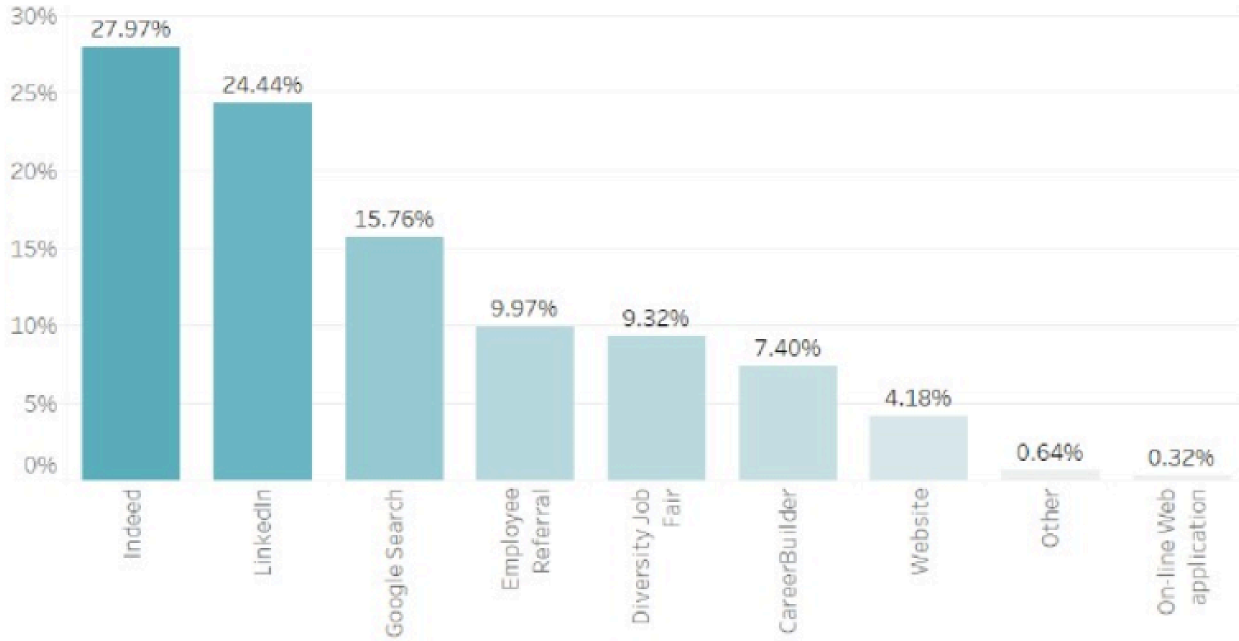
Why not a bar? Pie is better for proportional comparison when looking at a single attribute.

Marks and Channels :

Marks : 2-D, When looking at pie chart/ donut chart instead of looking at the angles, areas covered by certain categories are focused on.

Channel: angle(percentage), hue(performance score)

3. Bar Chart (Bottom Right): Volume of Recruitment by Source



Shows what percentage of the company's workforce comes from each source.

Top contributors:

1. Indeed (27.97%)
2. LinkedIn (24.44%)
3. Google Search (15.76%)

<u>This bar chart is important because it tells us where we are hiring most, and when combined with performance data, we can judge efficiency and effectiveness.</u>

For instance:

- Google Search produces the highest percentage of perfect performers (87.76%) but only accounts for 15.76% of total hires.
- LinkedIn and Indeed are strong in both quantity and quality, making them highly reliable recruitment sources.

Why ?

Why is it used? Displays how many people were hired from each recruitment source.

Why not a pie chart? Bar chart gives better granular comparison between sources, especially useful when there are many categories.

Conclusion:

<u>Performance Depends on Recruitment Source</u>

- This visualization supports the alternative hypothesis: employee performance is not independent of recruitment source.
- Sources like Google Search, LinkedIn, and Employee Referrals yield consistently better-performing employees.
- Therefore, optimizing recruitment strategies to focus more on high-yielding sources (like Google Search, LinkedIn, Referrals) can improve workforce quality.

Marks and Channels :

Marks : 1-D.

Channel: Saturation(Count of Employee), Vertical position(Percentage of employee)

Output

H1: Termination Analysis:

Interpretation: Does the Data Support Our Hypothesis?

Yes — the visual insights support the alternative hypothesis

The evidence presented aligns with our hypothesis that employees with both high absenteeism and low performance scores are more likely to be terminated. The organization seems to use a combination of attendance and performance to evaluate employee retention, which reinforces the need for proactive HR strategies addressing both dimensions.

H2: Gender Pay Analysis:

Interpretation: Does the Data Support Our Hypothesis?

NOT ASSURED— the visual insights DOESN'T SIGNIFICANTLY support the alternative hypothesis.

At first glance, women appear to earn less because their median salary is lower. However, this does not necessarily indicate discrimination.

H3: · Manager Impact Analysis:.

Interpretation: Does the Data Support Our Hypothesis?

NOT ASSURED— the visual insights DOESN'T SIGNIFICANTLY support the alternative hypothesis.

H4:· Recruitment Source Analysis

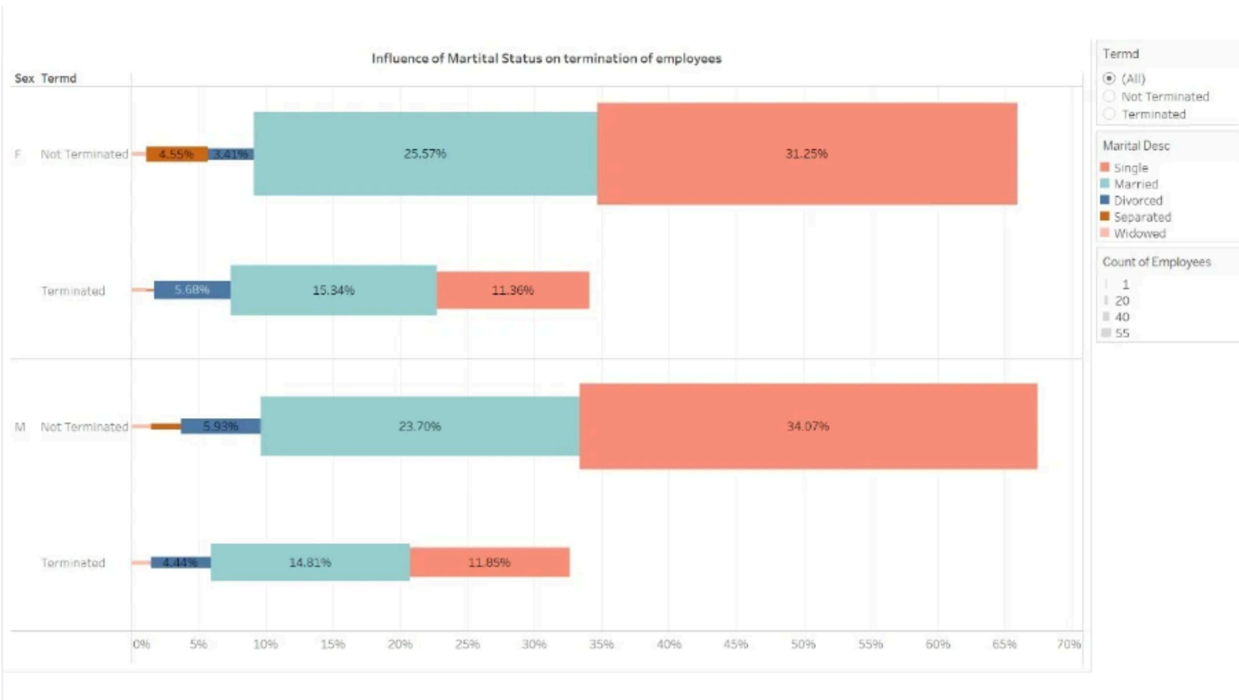
Interpretation: Does the Data Support Our Hypothesis?

Yes — the visual insights support the alternative hypothesis.

Sources like Google Search, LinkedIn, and Employee Referrals yield consistently better-performing employees.Therefore, optimizing recruitment strategies to focus more on high-yielding sources (like Google Search, LinkedIn, Referrals) can improve workforce quality

Potential Questions:

1.Does marital status influence employee attrition rates, particularly among married_females?



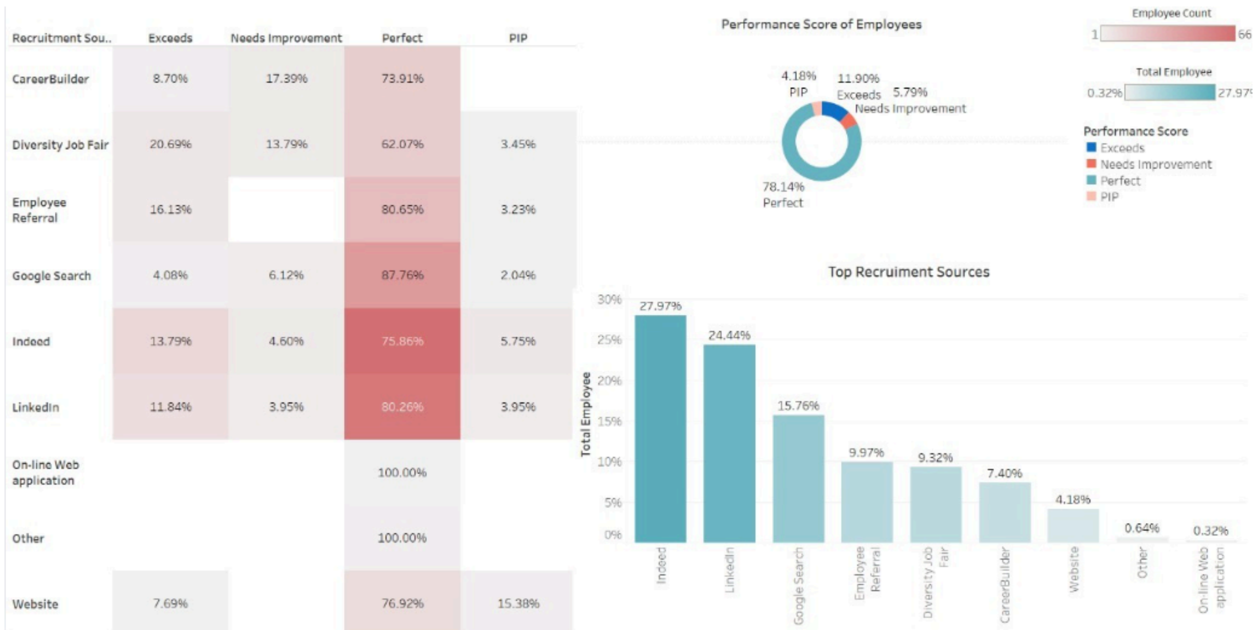
Answer **The chart analyzes termination rates by gender and marital status. Among female employees, 15.34% of married women were terminated, compared to 11.36% of single women. Similarly, 14.81% of married men were terminated, slightly higher than 11.85% of single men. While termination rates are marginally higher for married employees, the trend is similar across genders. Thus, marital status may influence attrition, but not specifically for females.**

2.How does an employee's performance score correlate with their salary, and is there a gender-based pay gap?

Answer **This visualization shows salary distributions across performance scores. Employees with "Perfect" or "Exceeds" ratings tend to have higher and more varied salaries, while those rated "PIP" have lower, tighter ranges. Although salary appears to increase with performance, the chart doesn't show gender. Therefore, no direct conclusion about a gender pay gap can be drawn without further gender-based breakdown.**

- More women are employed in lower-paying roles like Technician I/II.
- Fewer women are found in high-paying technical/managerial roles like Network Engineer or BI Developer.
- This skews the overall median downward for women, a composition effect, not necessarily discrimination.

3.Is there a significant relationship between an employee's recruitment source and their performance score?



Recruitment Sou..	Exceeds	Needs Improvement	Perfect	PIP
CareerBuilder	8.70%	17.39%	73.91%	
Diversity Job Fair	20.69%	13.79%	62.07%	3.45%
Employee Referral	16.13%		80.65%	3.23%
Google Search	4.08%	6.12%	87.76%	2.04%
Indeed	13.79%	4.60%	75.86%	5.75%
LinkedIn	11.84%	3.95%	80.26%	3.95%
On-line Web application			100.00%	
Other			100.00%	
Website	7.69%		76.92%	15.38%

Answered **Performance Depends on Recruitment Source.visualization supports the alternative hypothesis: employee performance is not independent of recruitment source.Sources like Google Search, LinkedIn, and Employee Referrals yield consistently better-performing employees.Therefore, optimizing recruitment strategies to focus more on high-yielding sources (like Google Search, LinkedIn, Referrals) can improve workforce quality**

Impact:

- **Reduce unwanted turnover** by addressing performance/attendance issues early.
- **Ensure fair compensation** to improve employee satisfaction and compliance.
- **Enhance leadership quality** by identifying strong/weak managers.
- **Improve hiring ROI** by focusing on the best recruitment channels.

SUMMARY

This analysis will help HR teams **make data-driven decisions** to boost productivity, equity, and retention.

The HR analytics study supports several key insights. Employee termination is closely tied to both high absenteeism and low performance, validating the need for early HR intervention. While women appear to earn less, this is largely due to role distribution rather than direct pay discrimination. Managerial influence is significant, with employee performance varying notably across different managers. Recruitment source also plays a role in performance, with sources like Google Search, LinkedIn, and Employee Referrals yielding better outcomes. These findings highlight opportunities to improve retention, ensure fair compensation, strengthen leadership, and refine recruitment strategies.

References

Dataset link: https://www.kaggle.com/datasets/rhuebner/human-resources-data-set

Dashboards link:

H1:

https://public.tableau.com/app/profile/meenakshi.iyer8742/viz/HRAnalyticsDashboard11/Hypothesis1

H2:
https://public.tableau.com/app/profile/meenakshi.iyer8742/viz/HRAnalyticsDashboard111/Hypothesis2

H3:

https://public.tableau.com/app/profile/meenakshi.iyer8742/viz/HRAnalyticsDashboard11/Hypothesis3

H4:

https://public.tableau.com/app/profile/meenakshi.iyer8742/viz/HRAnalyticsDashboard1111/Dashboard4

Final Dashboard

https://public.tableau.com/app/profile/meenakshi.iyer8742/viz/HRAnalyticsDashboard1_17440962605050/Dashboard1

.