

# Hiring Process Data Analysis

BY POOJA

## Introduction

In today's competitive job market, an efficient hiring process is crucial for any organization's success. This project aims to analyze hiring data to provide valuable insights into recruitment trends, candidate demographics, and salary distributions. These insights can assist the hiring department in optimizing its processes and making data-driven decisions.



## Contents

1. [Tech Stack Used](#)
2. [Dataset Overview](#)
3. [Data Cleaning](#)
4. [Insights and Analysis](#)
5. [Conclusion](#)
6. [References and Links](#)

## Tech Stack Used

- **Microsoft Excel:** Used for data filtering, computations, and visualization through graphs to derive insights from the hiring data.

## Dataset Overview

The project focuses on the **hiring process** of a company, a fundamental and crucial aspect of its operations. The analysis is conducted using the provided **dataset**. Primarily, **Exploratory Data Analysis (EDA)** is performed, where we aim to **summarize, visualize, and identify missing values** in the data. This process will assist in answering various questions related to the dataset. Some of the key questions include: **how many males and females are hired? what is the average salary offered in the company?** and more.

The dataset contains records of candidates who were interviewed previously with information about hiring status, hiring department, salary etc.

- **Source:** [Google Spreadsheet](#)
- **Total Records:** 7,168 data points
- **Features:** 6 columns, including:
  1. **Application ID** – Unique identifier for each applicant
  2. **Interview Taken on** – Date and time of the interview
  3. **Hiring Status** – Indicates whether the applicant was hired or rejected
  4. **Gender** – Gender of the applicant
  5. **Department** – Department for which the interview was conducted
  6. **Post Name** – Job role applied for
  7. **Offered Salary** – Proposed salary for the position

## Data Cleaning

### Handling Missing Values

1. The **event\_name** column contains 15 rows with “-“ as the value, which we considered as Null values. These were replaced with “Don’t want to say,” as both values convey the same meaning in this context—the gender of the candidate is unknown.

**Method:** Used find and replace or applied a filter to identify and replace the values.

2. The **Offered Salary** column has 1 row with a Null value. The corresponding entry in the **Department** column is “Sales Department,” and the **Post Name** is “i7.” This missing value was replaced with the median Offered Salary for the Sales Department and i7 Post Name, which was 45,400.

**Method:** Applied a filter for **Sales Department** and calculated the **median** salary for the i7 Post Name.

3. The **Post Name** column has 1 row with “-“ as its value, which we treated as a Null value. The corresponding entry in the **Department** column is “Sales Department,” and the **Offered Salary** is “85,914.” We replaced this value with the most frequent **Post Name** for candidates in the **Sales Department** whose Offered Salary ranges from 85,000 to 96,000, which was “c9.”

**Method:** Filtered for **Sales Department** and Offered Salary between 85,000 and 96,000, then used a pivot table to find the most frequent Post Name.

## Error Rectification

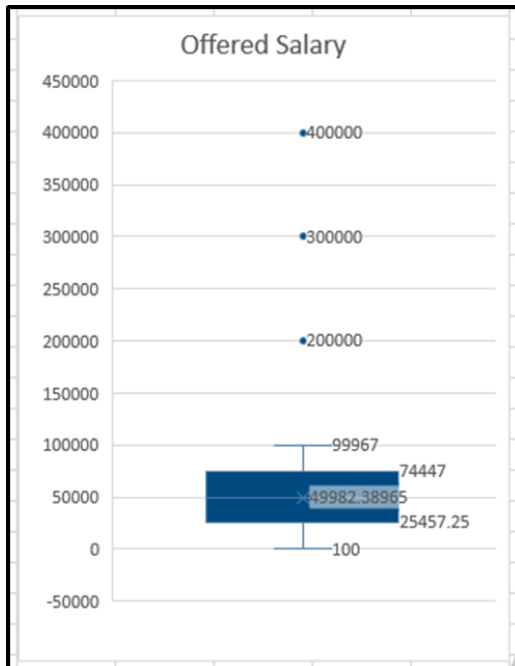
1. The **Post Name** column contained a typo, where “c-10” was used instead of the correct category, “c10.” This was rectified.

**Method:** Applied find and replace or filtered the column to correct the typo.

## Handling Outliers

1. In the **Offered Salary** column, a Box Plot revealed three outlier values: 200,000, 300,000, and 400,000. These were replaced with the median Offered Salary for the corresponding **Department** and **Post Name**.

**Method:** To ensure a more accurate salary distribution, the **upper bound for outliers** was calculated. We also calculated the median salary for each department corresponding to its Post Name. Any extreme value exceeding the upper bound was replaced with the calculated median.



## Handling Duplicate Values

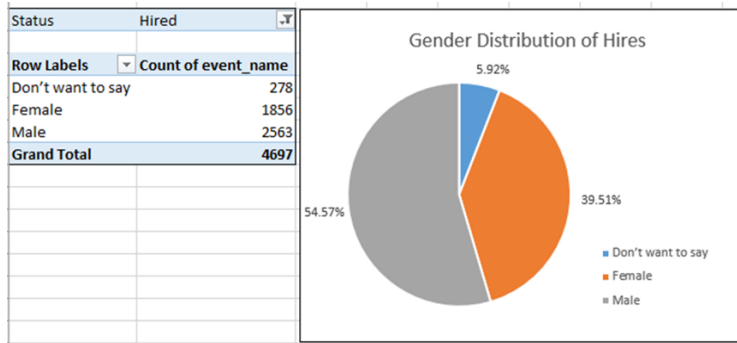
1. The **application\_id** column contains 54 rows with duplicate values. These duplicates should either be removed or replaced with the correct values.

**Method:** Applied **conditional formatting** to highlight duplicate rows and then removed or corrected them as necessary.

## Insights and Analysis

**A. Hiring Analysis:** The hiring process involves bringing new individuals into the organization for various roles.

- **Task:** Determine the gender distribution of hires. How many males and females have been hired by the company?
- **Results:**



- **Insight:**
  - More than **half** of the hired candidates are **male**, while **39.51% are female**.
  - A high gender disparity may impact the company's reputation, suggesting a need for more balanced recruitment.
  - The **data quality team** should ensure complete and accurate data entry to enhance analysis.

**B. Salary Analysis:** The average salary is calculated by adding up the salaries of a group of employees and then dividing the total by the number of employees.

- **Task:** What is the average salary offered by this company?
- **Results:**

Row Labels	Count of application_id	Average of Offered Salary
Hired	4697	49593
Rejected	2471	50420
<b>Grand Total</b>	<b>7168</b>	<b>49878</b>

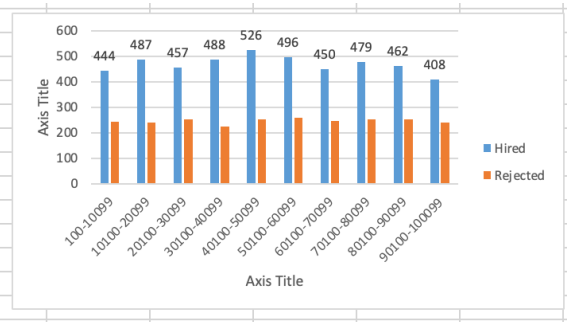
- **Insight:**
  - **Average Offered Salary:** 49,878
  - **Average Salary of Hired Candidates:** 49,593
  - Minimal variation suggests that the hiring team adheres to pre-defined salary structures.

**C. Salary Distribution:** Class intervals represent ranges of values, in this case, salary ranges. The class interval is the difference between the upper and lower limits of a class.

- **Task:** Create class intervals for the salaries in the company. This will help you understand the salary distribution.

- **Results:**

Count of application_id	Column Labels		
Row Labels	Hired	Rejected	Grand Total
100-10099	444	242	686
10100-20099	487	241	728
20100-30099	457	254	711
30100-40099	488	225	713
40100-50099	526	254	780
50100-60099	496	258	754
60100-70099	450	248	698
70100-80099	479	254	733
80100-90099	462	254	716
90100-100099	408	241	649
<b>Grand Total</b>	<b>4697</b>	<b>2471</b>	<b>7168</b>



- **Insight:**

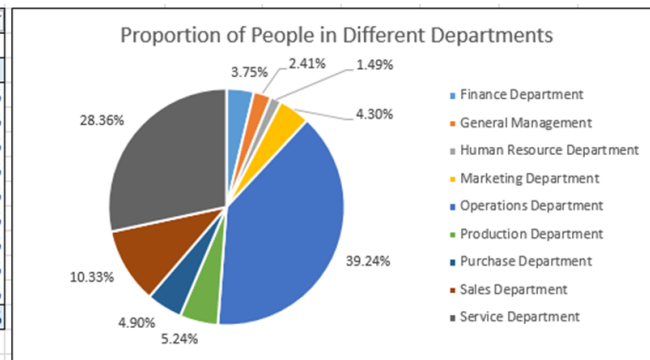
- Most salaries fall within the **40,100–50,099** range, indicating a focus on mid-level experience roles.
- Minimum salaries are offered to senior roles (90,100–100,099) and freshers (1–10,000).
- The salary distribution shows that most hires are for middle-level experience positions, with fewer hires for senior roles and freshers, assuming salary correlates with experience.

**D. Departmental Analysis:** Visualizing data through charts and plots is a crucial part of data analysis.

- **Task:** Use a pie chart, bar graph, or any other suitable visualization to show the proportion of people working in different departments.

- **Results:**

Status	Hired	
Department	Dept. wise Proportion	
Finance Department	3.75%	
General Management	2.41%	
Human Resource Department	1.49%	
Marketing Department	4.30%	
Operations Department	39.24%	
Production Department	5.24%	
Purchase Department	4.90%	
Sales Department	10.33%	
Service Department	28.36%	
<b>Grand Total</b>	<b>100.00%</b>	



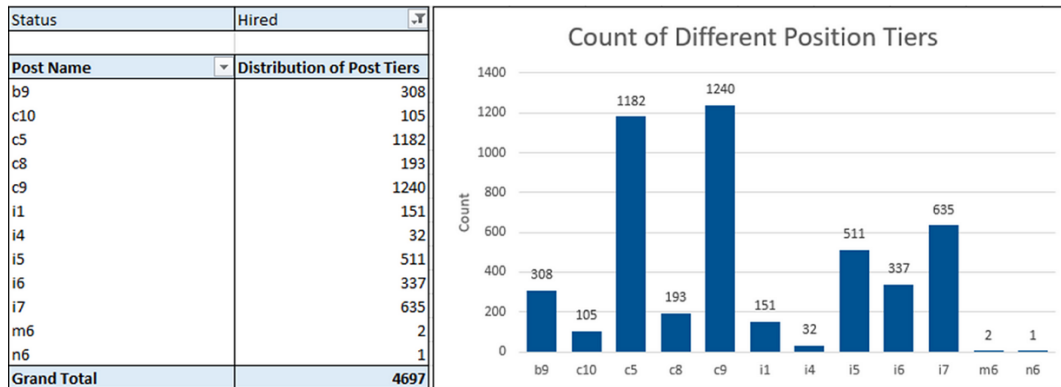
- **Insight:**

- **Highest recruitment:** Operations, Services, and Sales departments.
- **Lowest recruitment:** Human Resources.
- The trend may reflect team sizes and departmental priorities.

**E. Position Tier Analysis:** Different positions within a company often have different tiers or levels.

- **Task:** Use a chart or graph to represent the different position tiers within the company. This will help you understand the distribution of positions across different tiers.

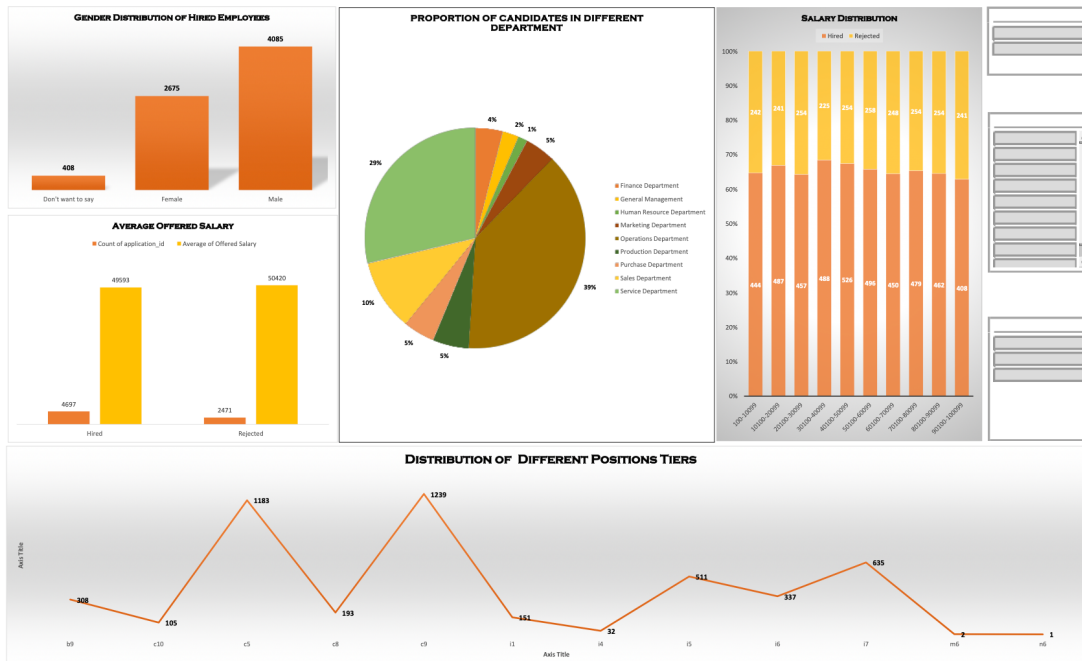
- Results:



- Insight:

- The most hired job positions: **c9**, followed by **c5**, and then **i7**.

#### HIRING PROCESS DATA ANALYSIS



## Conclusion

This analysis highlights the **significant role** that **data analytics** plays in **optimizing** the hiring process. By identifying **trends** in **recruitment**, **salary structures**, and **department-wise hiring**,

organizations can make **informed, data-driven** hiring decisions. These **insights** help improve **diversity, standardize salary offerings**, and ensure **fair recruitment practices**.

Through conducting **Exploratory Data Analysis (EDA)** on the **Human Resource dataset**, I gained a comprehensive understanding of key steps, including **data cleaning** and extracting **insights** through **statistical analysis**. **Excel** simplified the statistical analysis process with its **built-in functions**, making tasks more efficient. **Visualizing the data** further enhanced my understanding, as these visualizations greatly simplify analysis and interpretation. For example, **scatter plots** are particularly useful for quickly identifying **outliers** by simply viewing the data.

This project provided valuable **hands-on experience** with a **real-life dataset**, where I learned how to **clean, manipulate, visualize**, and extract meaningful **insights**. I addressed the dataset's questions to the best of my ability, creating the required **plots** and **charts** based on my understanding. The **EDA** phase emphasized the importance of thoroughly examining the data before applying further **analytical techniques** or building **models**. It reinforced that EDA is crucial for eliminating **errors** and **biases**, ensuring that the insights drawn are **reliable** and suitable for further **statistical** and **analytical work**.

---

**End of Document**