

# Hiring Process Analytics Imagine you're a data analyst at a multinational company.

You are tasked to analyze the company's hiring process data and draw meaningful insights from it. The hiring process is a crucial function of any company, and understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department. This Project is regarding giving insights about Hiring Process based on data provided which can be helpful for hiring departments.

## Contents

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

## Tech Stack Used

1. Microsoft Excel 2016 — A spreadsheet editor software used mainly by professionals to enter data in table format, perform computations, plot graphs etc. Here Microsoft Excel is used to filter data and plot graphs to get insights about hiring process of the organization.



# Dataset Overview

## Source of Data:

<https://docs.google.com/spreadsheets/d/1gAq5sK8L2e7rCP000KaNo7gqx6tfnVQk/edit#gid=1029390730>

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

### ●The Dataset details are:

- Number of Data-Points: 7,168
- Number of Features: 6
- Column Details:
  1. application\_id: ID of the applicant
  2. Interview Taken on: Date and time of the interview
  3. Status: Hired or rejected
  4. event\_name: Gender of the applicant
  5. Department: Name of the department for which interview was conducted
  5. Post Name: Name of the post offered
  6. Offered Salary: Salary offered for the job

## Data Cleaning

### Handling Missing Values

1. Column event\_name has 15 rows with “-“ as its values. These can be termed as Null values. We replaced it with “Don’t want to say” as they both implies the same thing in context of this project i.e. gender of the candidate is not known.
2. Column Offered Salary has 1row with Null Value. The corresponding value in Department column is “Sales Department” and Post Name is “i7”.

So we replaced i with median of Offered Salary for Sales Department and i7 Post Name. The median came out to 45400.

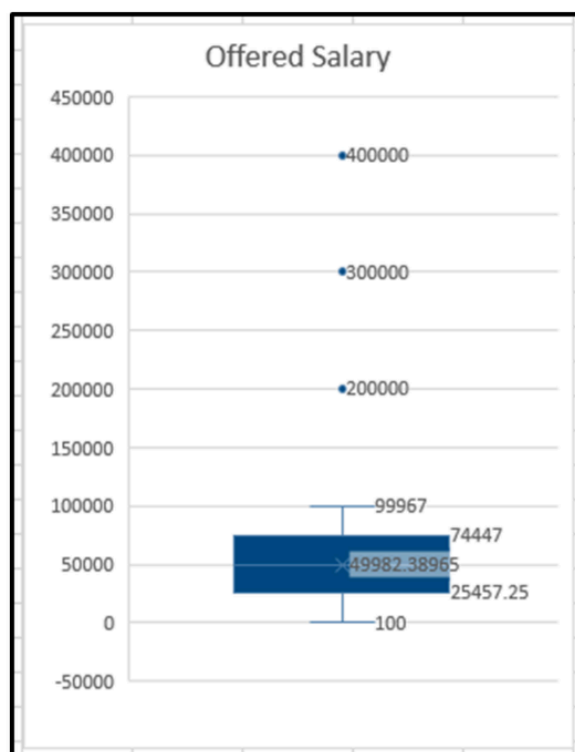
3. Column Post Name has 1 row with “-” as its value. It can be determined as Null value. The corresponding value in Department column is “Sales Department” and Offered Salary is “85914”. So we replaced it with majority count of Posts for candidates in Sales Department and whose Offered Salary is between 85,000 and 96,000, which is “c9”.

## Error Rectification

1. Column Post Name has a category “c-10” which seems to be a typo and the correct category should be “c10” which we rectified.

## Handling Outliers

1. From the below Box Plot of Column Offered Salary, we can see that there are three rows whose Column values are outliers and the values are 200000, 300000, 400000. We replaced them with median value of Offered Salary for corresponding Department and Post Name.



## Handling Duplicate Values

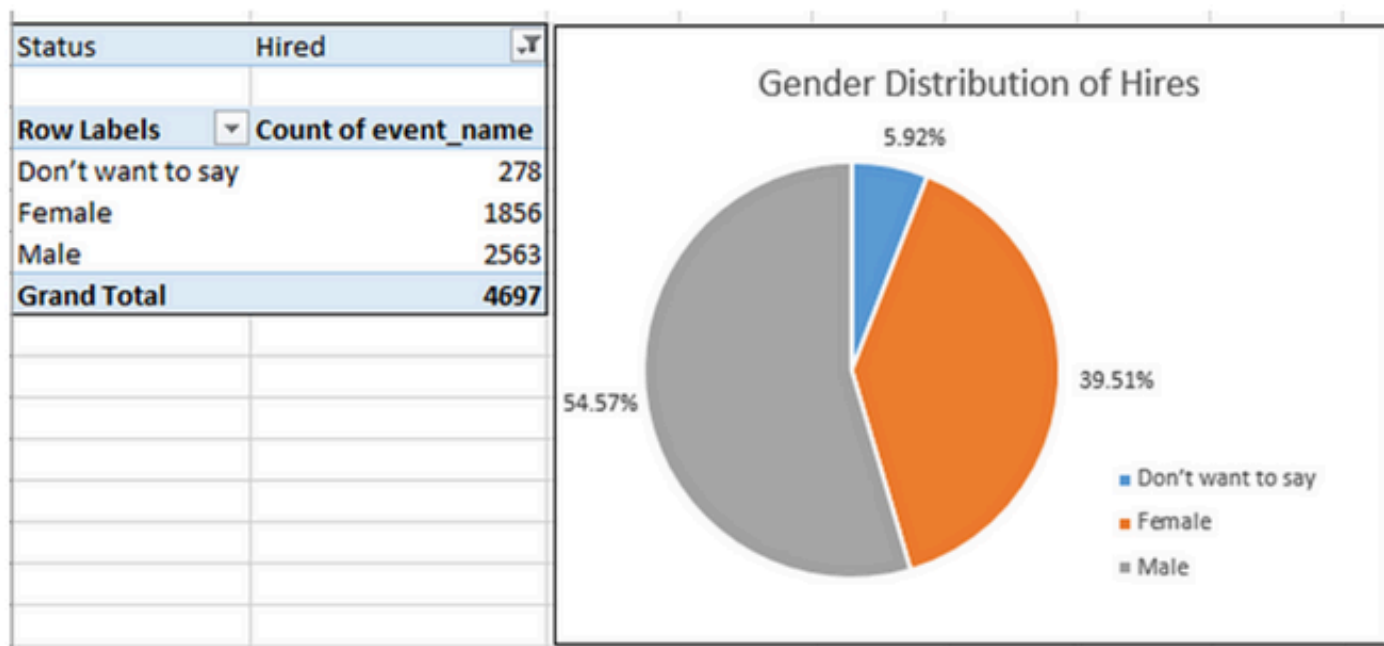
1. Column application\_id has 54 rows with duplicate values. They should either be removed or replaced with correct value.

## Insights:

**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?

●**Result:**



## ●Insight:

—More than half of the hired candidates are Male and only 39.51% of them are Female. The rest haven't disclosed their Gender. High Gender Ratio (Ratio of Male to Female) may negatively impact the Organization's image in Public domain. The Organization should therefore focus on decreasing the Gender Ratio bringing it close to 1.

—The Data Quality Team should ensure that the Data is complete and relevant as incomplete or irrelevant data may hamper the analysis process.

**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.

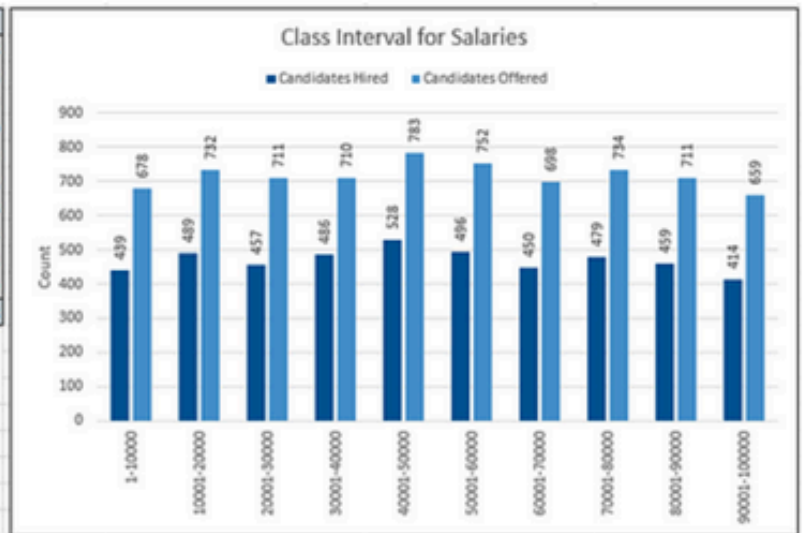
●**Task:** What is the average salary offered by this company?

Use Excel functions to calculate this.



Salary Class	Candidates Hired	Candidates Offered
1-10000	439	678
10001-20000	489	732
20001-30000	457	711
30001-40000	486	710
40001-50000	528	783
50001-60000	496	752
60001-70000	450	698
70001-80000	479	734
80001-90000	459	711
90001-100000	414	659
<b>Grand Total</b>	<b>4697</b>	<b>7168</b>

\* The above table is obtained by establishing a Relationship between two tables of the same pivot table. The work is done in a separate sheet called Q-C. P.T. Conn.



### ●Insight:

—We can observe that maximum offered salary is in the interval of Currency-Unit 40001–50000 while minimum offered salary are in intervals of Currency-Unit 90001–100000 and Currency-Unit 1–10000. That is most of the job requirement was for middle experience posts and least for senior most posts and for freshers (assuming salary is directly proportional to work experience).

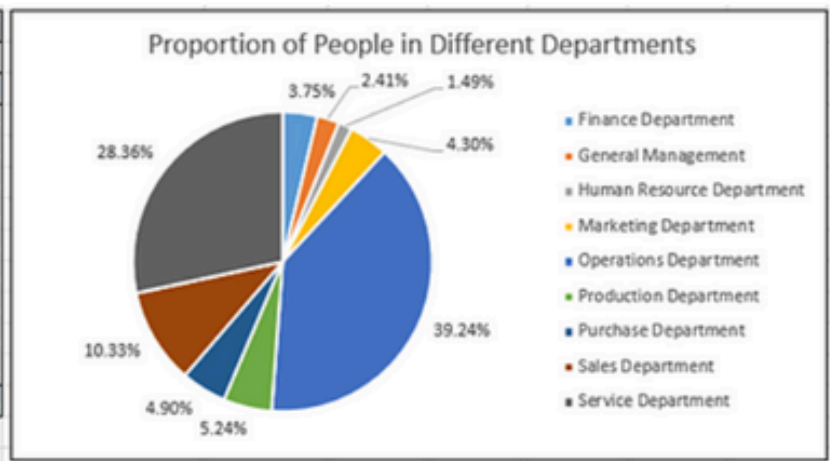
—We can also observe that the distribution of Salaries of hired candidate also follows the same pattern i.e. most candidates hired for middle experience posts and least for senior most posts and for freshers (assuming salary is directly proportional to work experience).

**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.

●**Result:**

Status	Hired	T
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%	
Grand Total	100.00%	



### ●Insight:

—From the above pie chart, we can observe that most candidates are hired in Operations Department followed by Services Department and Sales Department and the least candidates are hired in Human Resource Department.

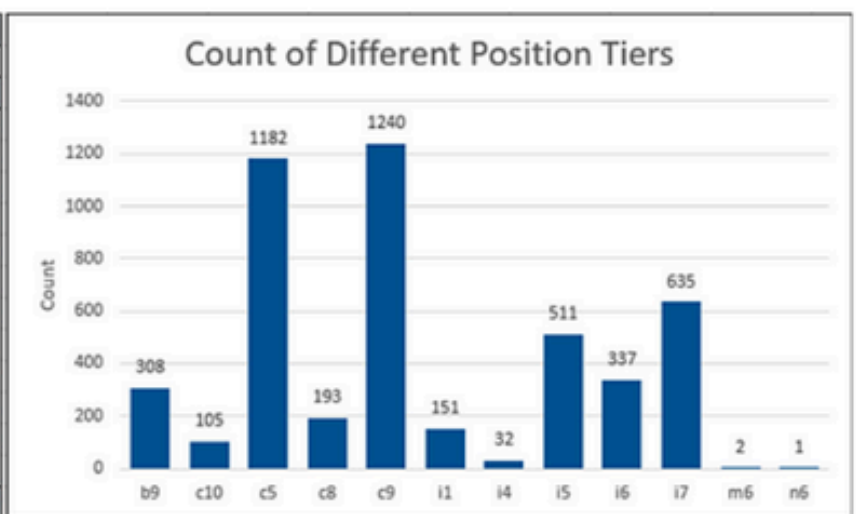
—These numbers may indicate the size of teams and importance of the departments in the organization.

**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.

### ●Result:

Status	Hired	T
Post Name	Distribution of Post Tiers	
b9	308	
c10	105	
c5	1182	
c8	193	
c9	1240	
i1	151	
i4	32	
i5	511	
i6	337	
i7	635	
m6	2	
n6	1	
Grand Total	4697	



### ●Insight:

—Here, we can observe that the organization has hired most candidates for post tier c9 followed by c5 and then i7 at distant third.

## Conclusion

This project helped me in understanding how important Data Analytics is for Hiring Process of an organization as it provides valuable insights such as number of rejections, reason for rejections, profile of applicants, vacancies etc. which helps the hiring department to take Data-Driven Decisions.