
By: ANIRUDH CHAUDHARY

# PROJECT DESCRİPTİON:

The hiring process is a critical function for any organization, ensuring the recruitment of qualified talent while optimizing the company's resources and strategies. For this project, the goal was to analyze hiring process data from a multinational company like Google to identify trends, insights, and opportunities for improvement.

**The dataset provided includes details such as candidate demographics, salary data, departmental distribution, and job tier classifications.**

This project aimed to address specific areas:

**1. Understanding the gender distribution of hires to promote diversity and inclusion.**
**2. Analyzing salary trends to ensure competitiveness and equity.**
**3. Identifying the distribution of employees across departments and position tiers to highlight resource allocation and organizational structure.**
**4. Detecting outliers, handling missing data, and summarizing the dataset for actionable insights.**
**The findings from this analysis could guide data-driven decisions to enhance the hiring process, ensure equitable practices, and align recruitment strategies with organizational goals.**

# APPROACH:

**To execute this project, a systematic approach was adopted to ensure accuracy and efficiency. The steps were as follows:**

### 1. Data Cleaning and Preparation:

o Missing Data Handling: Identified columns with missing data. Missing values were addressed using strategies like filling with the mean/median for numerical data or mode for categorical data. In cases where data was irretrievable, rows were excluded to maintain dataset integrity.
o Outlier Detection and Handling: Salary data was examined for outliers using statistical methods such as the interquartile range (IQR). Outliers were visualized using box plots, and decisions were made to either remove or replace them based on their impact on analysis.

### 2. Data Summarization:

o Statistical measures (mean, median, and standard deviation) were calculated to understand the central tendencies and variability within the dataset.
o Data was visually explored through pivot tables and charts to summarize key findings.

### 3. Question-Specific Analysis:

o Gender Distribution: Filtered data to count the total hires by gender.
o Average Salary: Used Excel functions like AVERAGE to compute the mean salary across the dataset.
o Salary Distribution: Created salary intervals using Excel's frequency distribution functions, which were later represented visually.
o Departmental Analysis: Used pie charts and bar graphs to depict the proportional distribution of employees across various departments.
o Position Tier Analysis: Analyzed the number of employees at each job tier level and visualized this data using stacked bar charts.

### 4. Visualization and Reporting:

All analyses were translated into clear and intuitive visualizations. Each chart or graph provided a visual narrative of the findings, making it easier to interpret and present to stakeholders.

# TECH-STACK USED

The primary tool for this project was Microsoft Excel (Version 2022). Excel was chosen for its versatility in handling structured data, performing statistical analysis, and creating visualizations. The following Excel features and functions were employed:

**1. Data Cleaning Tools:** Sorting, filtering, and removing duplicates.

**2. Statistical Functions:** AVERAGE, MEDIAN, MIN, MAX, COUNTIF, and PERCENTILE functions for detailed analysis.

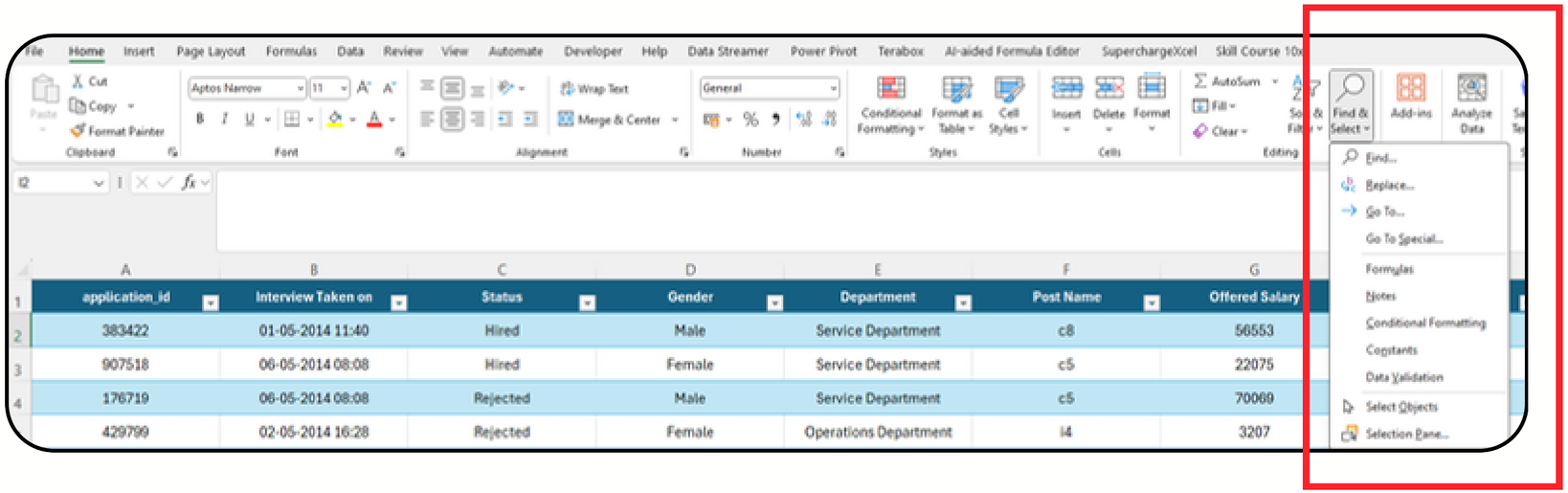
**3. Visualization Tools:** Pie charts, bar graphs, histograms, and stacked bar charts to present findings effectively.

**4. Pivot Tables:** Used to summarize large datasets and extract key insights dynamically.

**5. Conditional Formatting:** Highlighted trends and anomalies in the dataset.

# Cleaning the “Offered-Salary” Column, Identifying Blank Cells/Duplicate Values

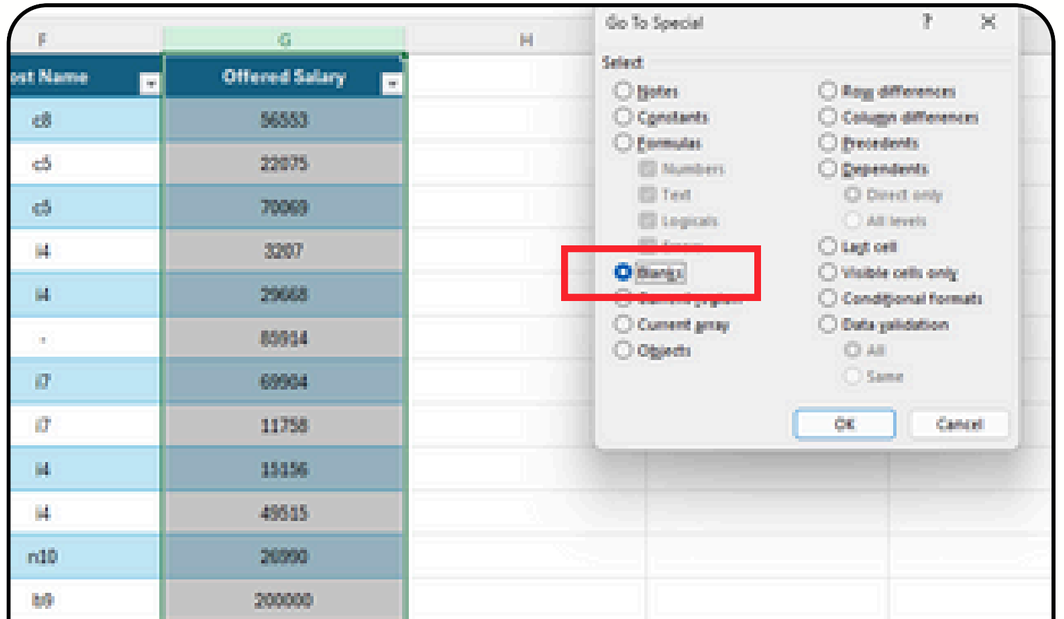
DATA-CLEANING




1

| application_id | Interview Taken on | Status | Gender | Department | Post Name | Offered Salary |
|---|---|---|---|---|---|---|
| 383422 | 01-05-2014 11:40 | Hired | Male | Service Department | c8 | 56553 |
| 907518 | 06-05-2014 08:08 | Hired | Female | Service Department | c5 | 22075 |
| 176719 | 06-05-2014 08:08 | Rejected | Male | Service Department | c5 | 70069 |
| 429799 | 02-05-2014 16:28 | Rejected | Female | Operations Department | i4 | 3207 |

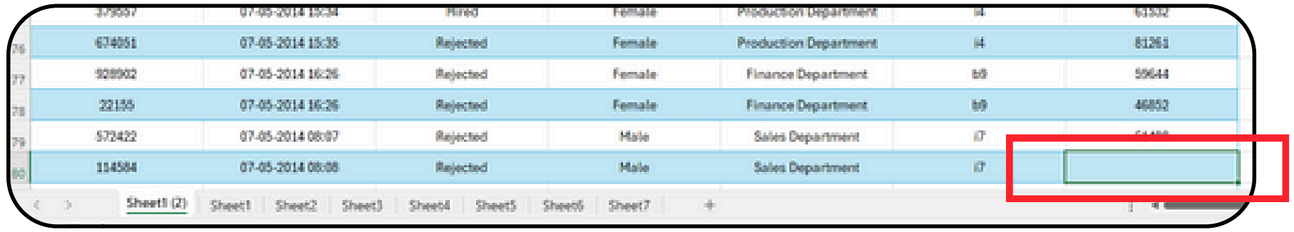
2



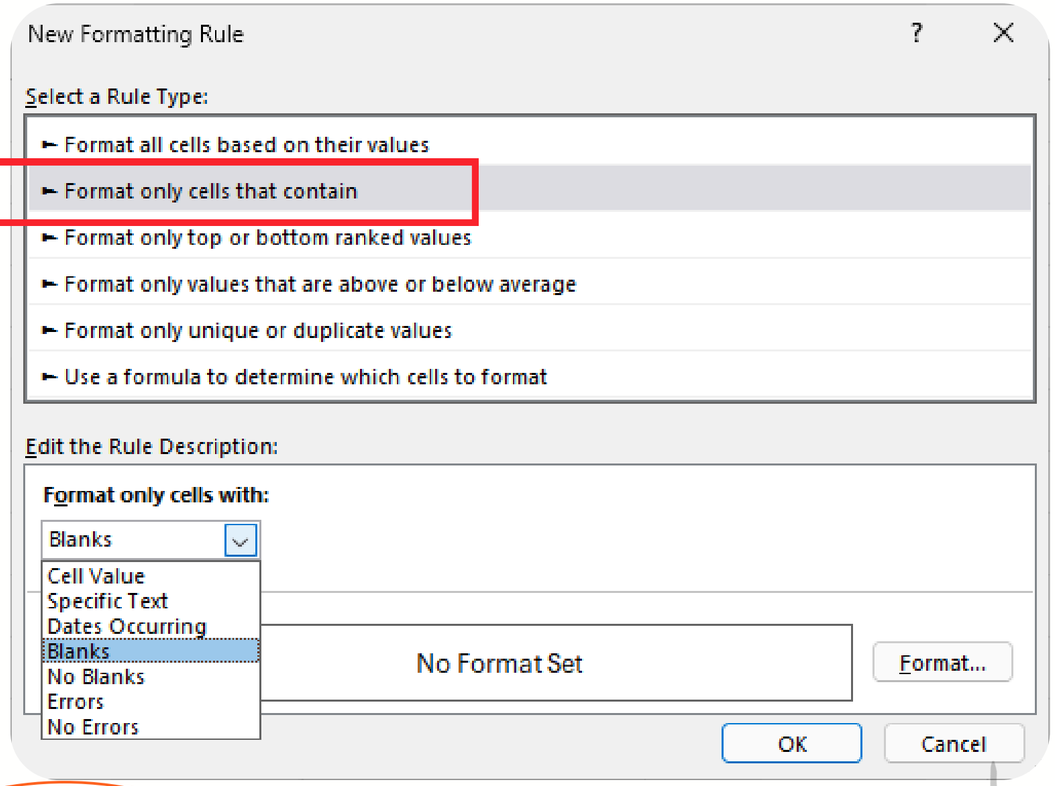
# Post Identifying the “blank cells”, Highlighting the same

DATA-CLEANING

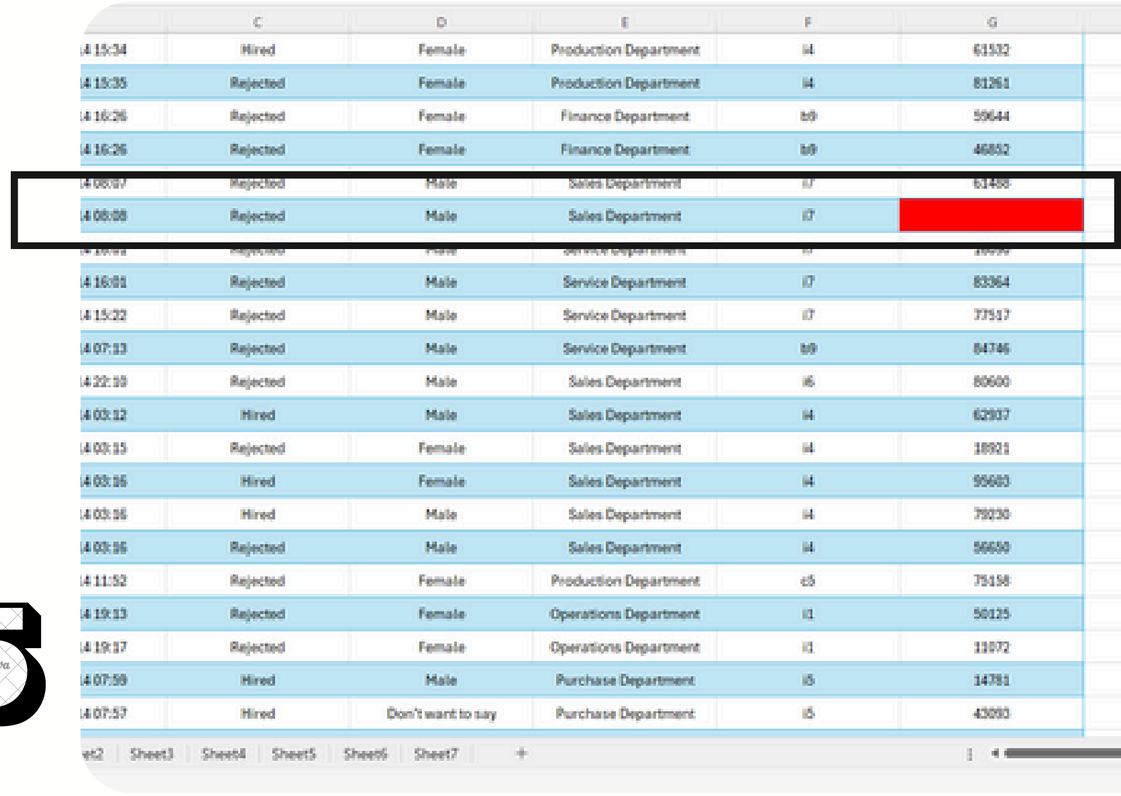
3



4




5



Post Identifying the “blank cells”, Highlighting the same

DATA-CLEANING

**Identified a cell with “BLANK-SALARY”, department “Sales”, but the male-candidate identified is “REJECECTED” in the interview**

- Since the candidate is **Rejected** we may either consider deleting the row

**OR**

- Or Use the AVERAGE-FUNCTION, to consider the average salary of the employee from all “Sales-Department Employees”, after filtering (sales department )

- **=AVERAGE(CELL- REFERANCE), Eg: =AVERAGE(H14:H25)**

# Cleaning the “Post-Name” Column, Identifying Blank Cells/Duplicate Values

DATA-CLEANING

Using Filter/Go-to special to find any blank, non-satisfactory data

1

| Department | Post Name | Offered Salary |
|---|---|---|
| Service Department | c8 | 56553 |
| Service Department | c5 | 22075 |
| Service Department | c5 | 70069 |
| Operations Department | i4 | 3207 |
| Operations Department | i4 | 29668 |
| Sales Department | - | 85914 |
| Sales Department | i7 | 69904 |
| Sales Department | i7 | 11758 |
| Service Department | i4 | 15156 |
| Service Department | i4 | 49515 |
| Service Department | n10 | 26990 |
| Service Department | b9 | 200000 |
| Service Department | b9 | 86787 |
| Finance Department | b9 | 2308 |
| Service Department | i7 | 56688 |
| Service Department | i7 | 81757 |
| Service Department | i5 | 15134 |
| Service Department | i5 | 100 |
| Operations Department | i1 | 73579 |
| Operations Department | i1 | 50351 |

**Identified a cell with “BLANK-Post Name”, department “Sales”, with cell reference “F7”**

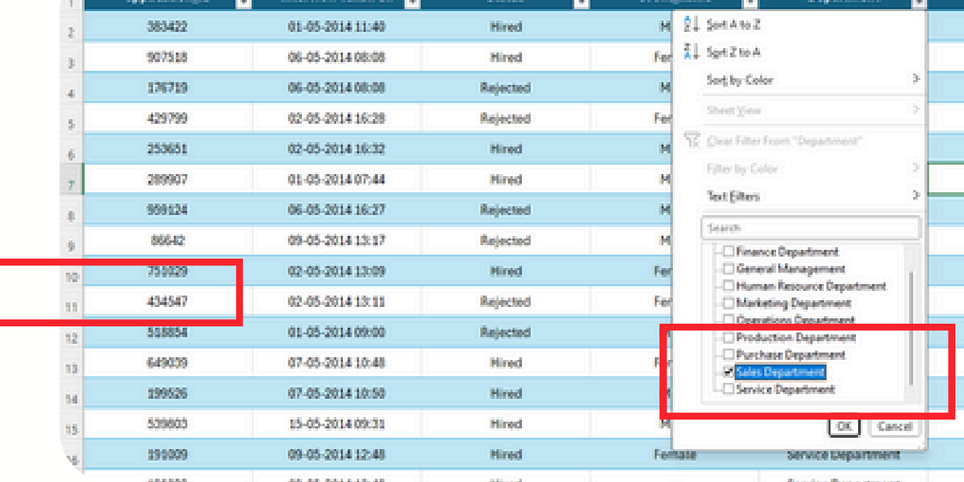
2

| | | | | | | |
|---|---|---|---|---|---|---|
| 253651 | 02-05-2014 16:32 | Hired | Male | Operations Department | i4 | 29668 |
| 289907 | 01-05-2014 07:44 | Hired | Male | Sales Department | - | 85914 |
| 969124 | 06-05-2014 16:27 | Rejected | Male | Sales Department | i7 | 69904 |
| 86642 | 09-05-2014 13:17 | Rejected | Male | Sales Department | i7 | 11758 |

To find the possible “post-name” for this application id , tried filtering the data only for “Sales-Department” & use MODE to find the most likely “Post”

3

| application_id | Interview Taken on | Status | event_name | Department | Post Name | Offered Salary |
|---|---|---|---|---|---|---|
| 383422 | 01-05-2014 11:40 | Hired | M | | c8 | 56553 |
| 907518 | 06-05-2014 08:08 | Hired | Fer | | c5 | 22075 |
| 176719 | 06-05-2014 08:08 | Rejected | M | | c5 | 70069 |
| 429799 | 02-05-2014 16:28 | Rejected | Fer | | i4 | 3207 |
| 253651 | 02-05-2014 16:32 | Hired | M | | i4 | 29668 |
| 289907 | 01-05-2014 07:44 | Hired | M | | - | 85914 |
| 969124 | 06-05-2014 16:27 | Rejected | M | | i7 | 69904 |
| 86642 | 09-05-2014 13:17 | Rejected | M | | i7 | 11758 |
| 759929 | 02-05-2014 13:09 | Hired | Fer | | i4 | 15156 |
| 434547 | 02-05-2014 13:11 | Rejected | Fer | | i4 | 49515 |
| 538858 | 01-05-2014 09:00 | Rejected | | | n10 | 26990 |
| 649039 | 07-05-2014 10:48 | Hired | F | | b9 | 200000 |
| 199526 | 07-05-2014 10:50 | Hired | | | b9 | 86787 |
| 539603 | 15-05-2014 09:31 | Hired | M | | b9 | 2308 |
| 191009 | 09-05-2014 12:48 | Hired | Female | | i7 | 56688 |



# Cleaning the “Post-Name” Column, Identifying Blank Cells/Duplicate Values

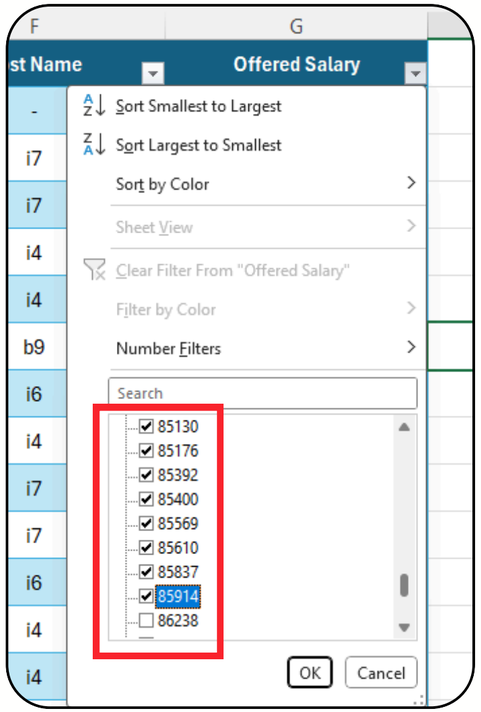
DATA-CLEANING

After applying the Filter to “Department”, and Selecting only “Sales-Department”

| application_id | Interview Taken on | Status | event_name | Department | Post Name | Offered Salary |
|---|---|---|---|---|---|---|
| 289907 | 01-05-2014 07:44 | Hired | Male | Sales Department | - | 85914 |
| 959124 | 06-05-2014 16:27 | Rejected | Male | Sales Department | i7 | 69904 |
| 86642 | 09-05-2014 13:17 | Rejected | Male | Sales Department | i7 | 11758 |
| 765579 | 01-05-2014 09:26 | Rejected | Male | Sales Department | i4 | 87884 |
| 924976 | 01-05-2014 09:26 | Rejected | Male | Sales Department | i4 | 56229 |
| 955372 | 02-05-2014 09:16 | Rejected | Female | Sales Department | b9 | 85176 |
| 15712 | 03-05-2014 00:36 | Rejected | Male | Sales Department | i6 | 11970 |
| 188820 | 07-05-2014 17:58 | Rejected | Male | Sales Department | i4 | 94869 |
| 572422 | 07-05-2014 08:07 | Rejected | Male | Sales Department | i7 | 61488 |
| 114584 | 07-05-2014 08:08 | Rejected | Male | Sales Department | i7 | |
| 720850 | 08-05-2014 22:10 | Rejected | Male | Sales Department | i6 | 80600 |
| 453498 | 08-05-2014 03:12 | Hired | Male | Sales Department | i4 | 62937 |
| 550914 | 08-05-2014 03:15 | Rejected | Female | Sales Department | i4 | 18921 |
| 37906 | 08-05-2014 03:16 | Hired | Female | Sales Department | i4 | 95603 |
| 692080 | 08-05-2014 03:16 | Hired | Male | Sales Department | i4 | 79230 |

4

5 Now Further FILTER the “Salary” column, based on the Salary around= 85914 to find the probable/most likely to be Post-Name.



# Cleaning the “Post-Name” Column, Identifying Blank Cells/Duplicate Values

DATA-CLEANING

After applying the filter to Salary for values around 85914, we are left with only “9” data’s,
Now we may use “MODE / MODE-Function” ( **Max. Frequency** ) **c9**=“**4-times**” to consider the most likely post for application -259907

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | application_id | Interview Taken on | Status | event_name | Department | Post Name | Offered Salary |
| 7 | 289907 | 01-05-2014 07:44 | Hired | Male | Sales Department | - | 85914 |
| 39 | 955372 | 02-05-2014 09:16 | Rejected | Female | Sales Department | b9 | 85176 |
| 1323 | 882740 | 07-08-2014 13:42 | Rejected | Male | Sales Department | c9 | 85837 |
| 1457 | 230796 | 25-08-2014 09:32 | Hired | Female | Sales Department | c9 | 85569 |
| 2256 | 101190 | 09-05-2014 17:23 | Hired | Male | Sales Department | c9 | 85057 |
| 3845 | 804019 | 08-07-2014 16:52 | Hired | Female | Sales Department | c9 | 85130 |
| 4224 | 226229 | 10-07-2014 19:03 | Hired | Male | Sales Department | c5 | 85610 |
| 4229 | 788752 | 18-07-2014 17:37 | Rejected | Male | Sales Department | c5 | 85392 |
| 4527 | 294317 | 29-07-2014 17:33 | Rejected | Female | Sales Department | b9 | 85400 |

6

After considering MODE ( Max Frequency ), we arrive at conclusion to **consider “C9” as Post-Name for ID: 259907**

7

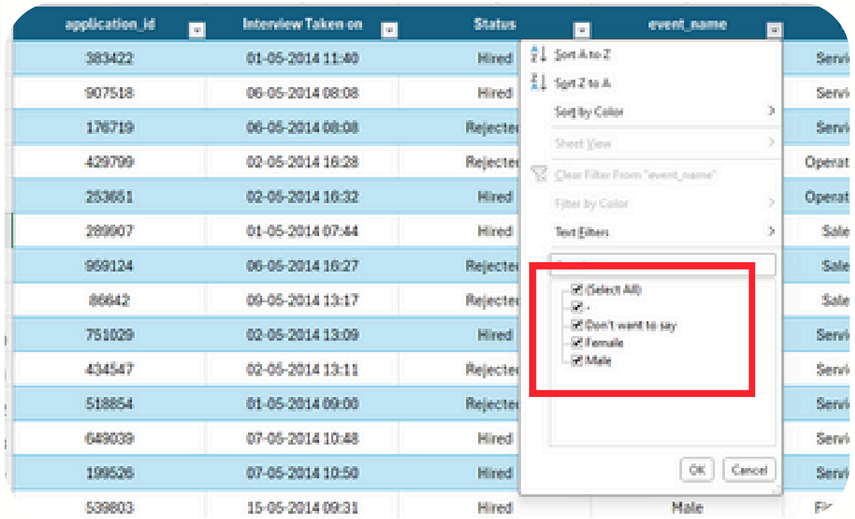
| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | application_id | Interview Taken on | Status | event_name | Department | Post Name | Offered Salary |
| 142 | 361096 | 15-05-2014 09:56 | Rejected | Male | Service Department | c10 | 9390 |
| 143 | 691216 | 15-05-2014 09:56 | Rejected | Male | Service Department | c10 | 67066 |
| 144 | 567661 | 15-05-2014 09:57 | Rejected | Male | Service Department | c10 | 8723 |
| 145 | 382645 | 15-05-2014 09:57 | Hired | Male | Service Department | c10 | 65587 |
| 146 | 767003 | 15-05-2014 10:01 | Hired | Male | Service Department | c10 | 73396 |
| 147 | 412827 | 15-05-2014 15:57 | Rejected | Male | Service Department | c10 | 76789 |
| 151 | 105729 | 15-05-2014 16:13 | Rejected | Male | Service Department | c10 | 80817 |
| 173 | 303466 | 19-05-2014 13:17 | Rejected | Male | Production Department | c10 | 81257 |
| 174 | 549934 | 19-05-2014 13:18 | Rejected | Male | Production Department | c10 | 59735 |
| 178 | 299540 | 20-05-2014 17:03 | Rejected | Female | General Management | c10 | 98404 |
| 179 | 66952 | 20-05-2014 17:04 | Rejected | Male | General Management | c10 | 58443 |
| 180 | 372707 | 20-05-2014 17:05 | Rejected | Female | General Management | c10 | 92123 |
| 181 | 703540 | 20-05-2014 17:06 | Hired | Female | General Management | c10 | 77027 |
| 182 | 838911 | 20-05-2014 17:03 | Hired | Male | General Management | c10 | 98822 |
| 183 | 51314 | 20-05-2014 17:05 | Hired | Female | General Management | c10 | 18661 |
| 184 | 716333 | 20-05-2014 17:03 | Hired | Don't want to say | General Management | c10 | 71461 |
| 199 | 826902 | 21-05-2014 12:35 | Rejected | Female | Finance Department | c10 | 23823 |
| 209 | 150277 | 22-05-2014 22:18 | Rejected | Male | Sales Department | c10 | 91928 |

# Cleaning the “Gender_Name” Column, Identifying Blank Cells/Duplicate Values

DATA-CLEANING

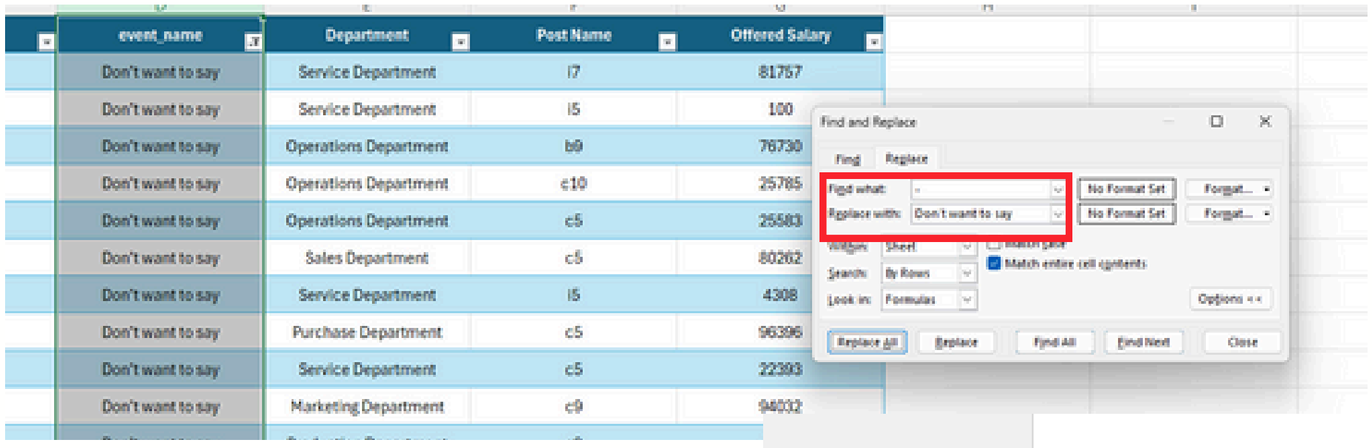
After applying the filter to “Gender name” column, found some cells containing “-”, apart from “male”, “female”, “Don’t want to say” Genders.

1




To correct the discrepancy, considered replacing the “-” hyphen , with “Don’t want to say” as the gender for applicant

2

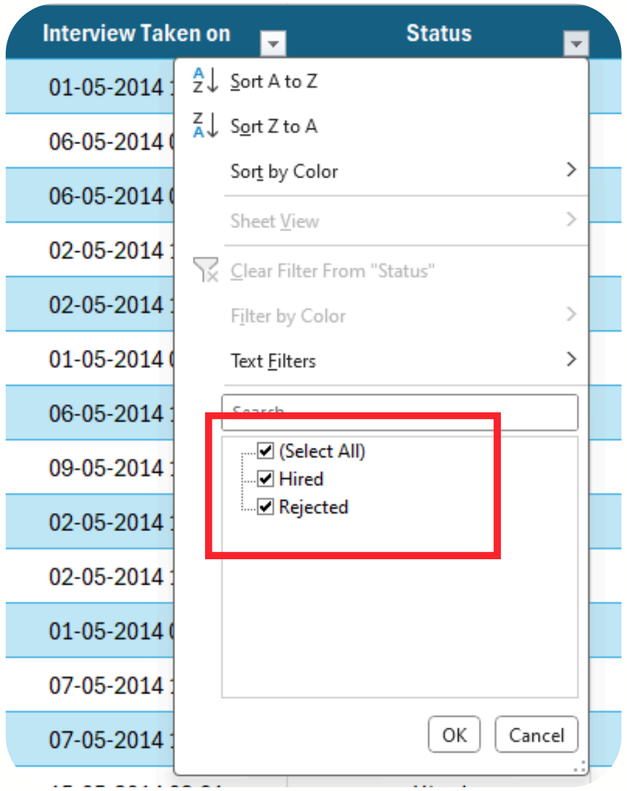


# Cleaning the “Status” Column, Identifying Blank Cells/Duplicate Values

DATA-CLEANING

After applying the filter to “STATUS” column, all the values seems correct, error-free.

1




Consider Moving with Next-Column

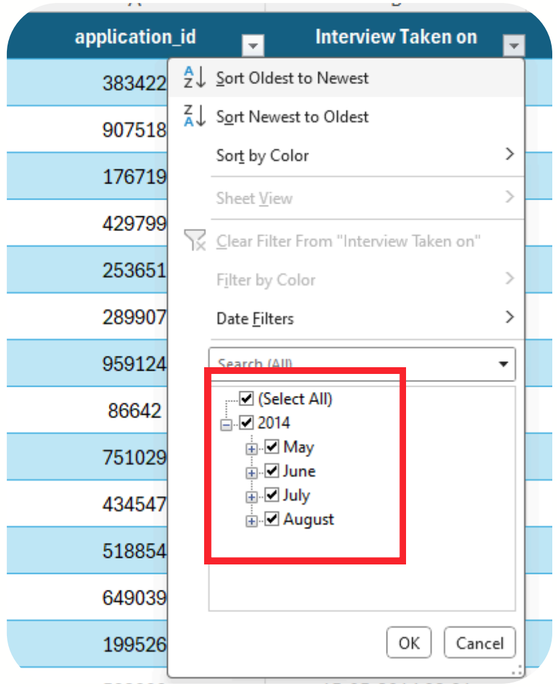
2

# Cleaning the “Interview Taken On” Column, Identifying Blank Cells/Duplicate Values

DATA-CLEANING

After applying the filter to “STATUS” column, all the values seems correct, error-free.
The column contains data for year 2014 , with months as :May, June, July, August

1




Consider Moving with Next-Column

2

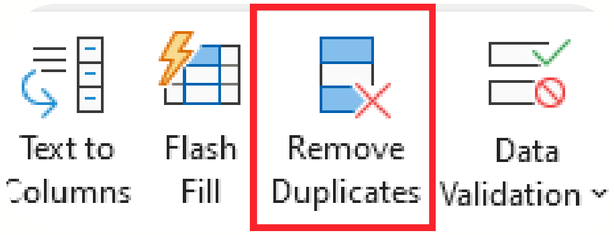
# Cleaning the “Application_id” Column, Identifying Blank Cells/Duplicate Values

DATA-CLEANING

1. After applying the Filter to “Application_Id”, check for blank-values using “Remove-Duplicate” or filter.

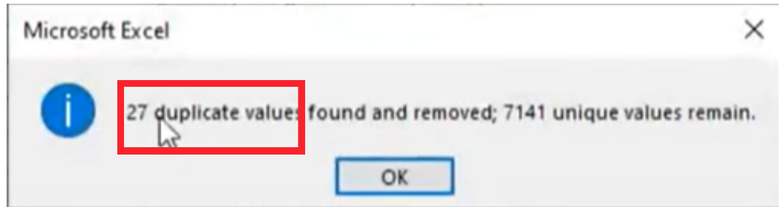
2. No Blank-Values found for Application_id.

3. Now, check for any Duplicate-Values, Under the “Data-Ribbon”, choose “Remove-Duplicates”, to find any duplicate-values.




Post checking for Duplicate-Values, Found 27 duplicate values, with 7141 unique values.

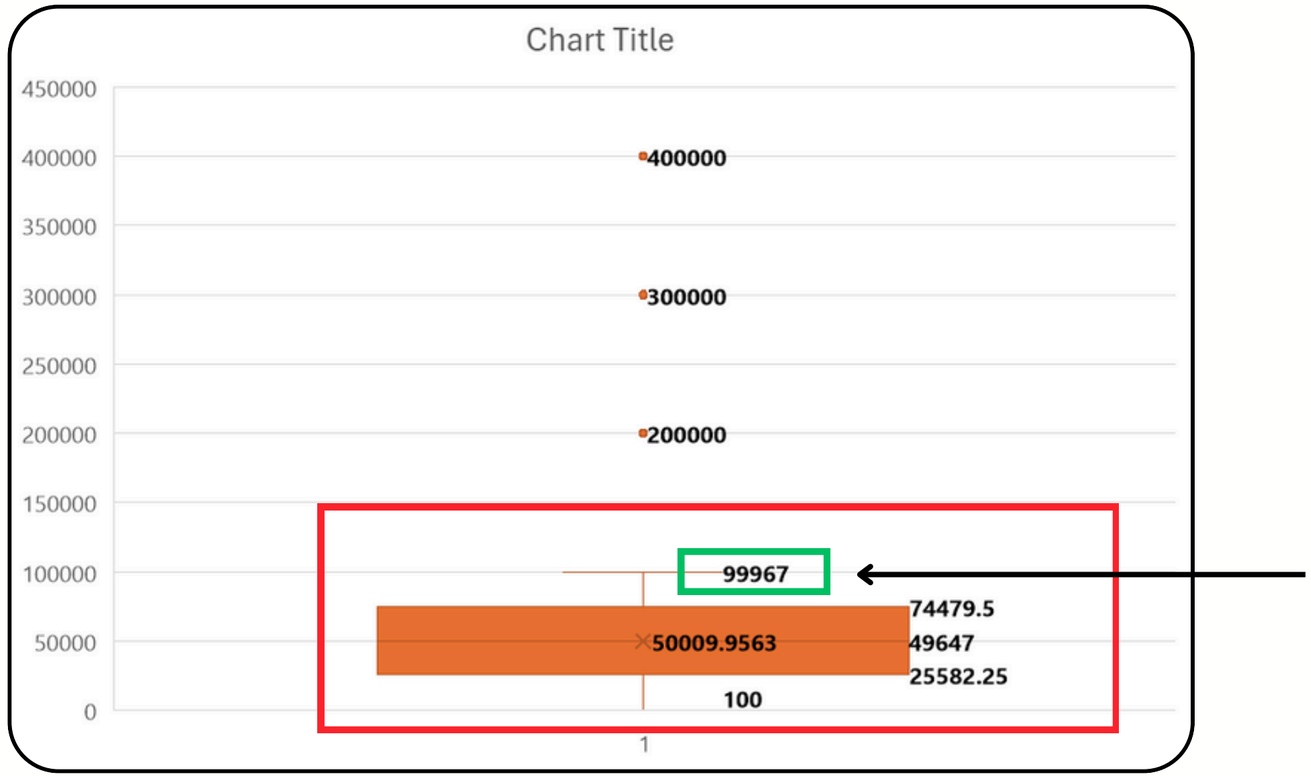
4



## Cleaning the “Application_id” Column, Identifying Blank Cells

DATA-CLEANING

1 Selected the “Offered_Salary” column, and inserted the Chart ( Box & Whisker ) for Outliers.



# HIRING ANALYSIS

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

1 2 3 4 5 6 7 8 9

1. Used “**Countifs**” Formula to determine the gender-wise Hiring

## FORMULA BASED CALCULATIONS:

| FORMULA USED | Count Ifs | FORMULA USAGE |
|---|---|---|
| | | |
| Males Candidates Hired | 2552 | =COUNTIFS('Cleaned Data'!D:D,"Male",'Cleaned Data'!C:C,"Hired") |
| | | |
| Female Candidates Hired | 1850 | =COUNTIFS('Cleaned Data'!D:D,"Female",'Cleaned Data'!C:C,"Hired") |
| | | |
| Other Genders Hired | 277 | =COUNTIFS('Cleaned Data'!D:D,"Don't want to say",'Cleaned Data'!C:C,"Hired") |
| | | |
| Total Candidates Hired | 4679 | =SUM(C23:C27) |

# HIRING ANALYSIS

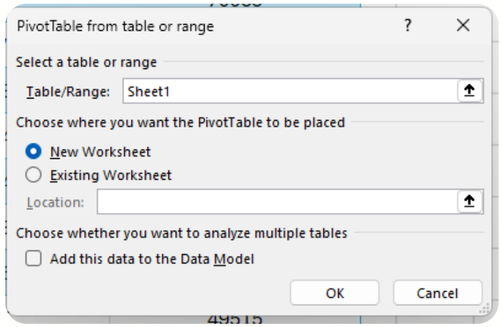
**1) Task**: Determine the gender distribution of hires. How many males and females have been hired by the company?

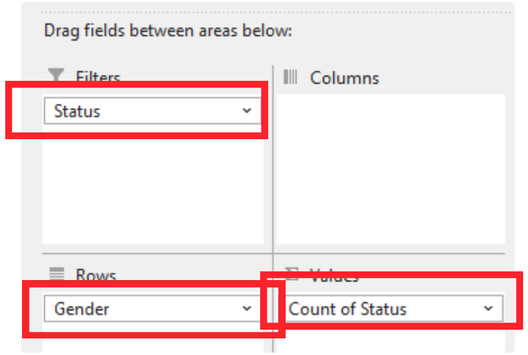
1 2 3 4 5 6 7 8 9

2 Can Use **Pivot-Table** to determine the gender-wise Hiring done.

| Status | Hired |
|---|---|

| Row Labels | Count of Status |
|---|---|
| Don't want to say | 277 |
| Female | 1850 |
| Male | 2552 |
| Grand Total | 4679 |

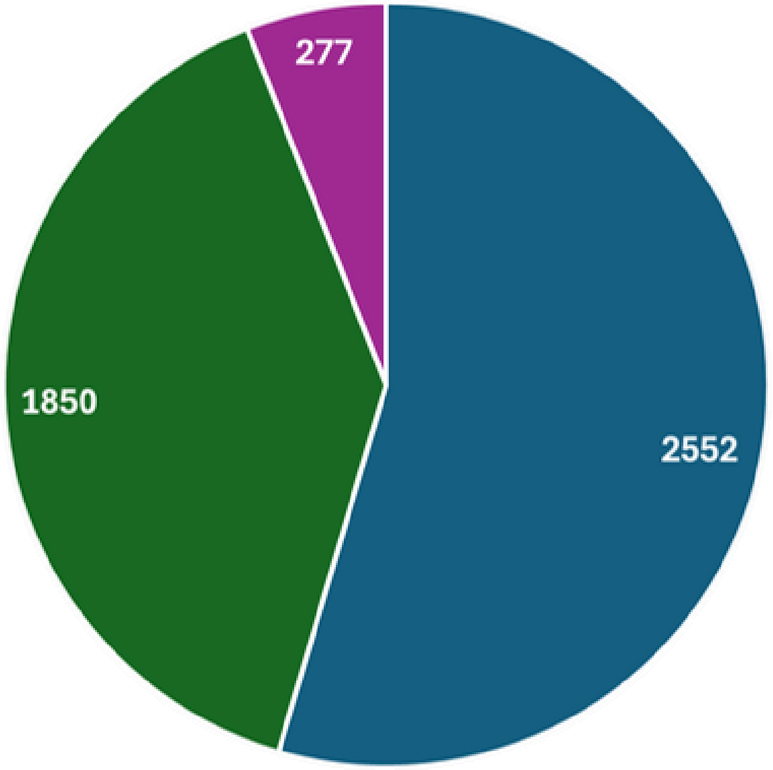


# HIRING ANALYSIS

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





# HIRING ANALYSIS

**1) Task**: What is the average salary offered by this company? Use Excel functions to calculate this.

1

Used “**Averageifs**” Formula to determine the average salary offered by this company.

## AVERAGE SALARY DEPARTMENT WISE USING FORMULA

**AVERAGE SALARY DEPARTMENT WISE**

| DEPARTMENT | RESULT |
|---|---|
| Service - Department | 50666.73839 |
| Operations - Department | 49129.95583 |
| Sales - Department | 49403.74966 |
| Human Resource - Department | 49448.63542 |
| Finance - Department | 49759.29617 |
| General - Department | 50666.73839 |
| Marketing - Department | 48489.93538 |
| Productions - Department | 49518.3562 |
| Purchase - Department | 52669.63855 |

**FORMULA USED : "AVERAGEIFS"**

| FORMULA USAGE |
|---|
| |
| =AVERAGEIFS('Cleaned Data'!H:H,'Cleaned Data'!E:E,"Operations Department") |
| |
| =AVERAGEIFS('Cleaned Data'!H:H,'Cleaned Data'!E:E,"Human Resource Department") |
| |
| =AVERAGEIFS('Cleaned Data'!H:H,'Cleaned Data'!E:E,"Service Department") |
| |
| =AVERAGEIFS('Cleaned Data'!H:H,'Cleaned Data'!E:E,"Production Department") |

# HIRING ANALYSIS

**1) Task**: What is the average salary offered by this company? Use Excel functions to calculate this.



# HIRING ANALYSIS

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

1
2
3
4
5
6
7
8
9

1 Class intervals represent ranges of values, in this case, salary ranges. The class interv
is the difference between the upper and lower limits of a class.

Your Task: Create class intervals for the salaries in the company.

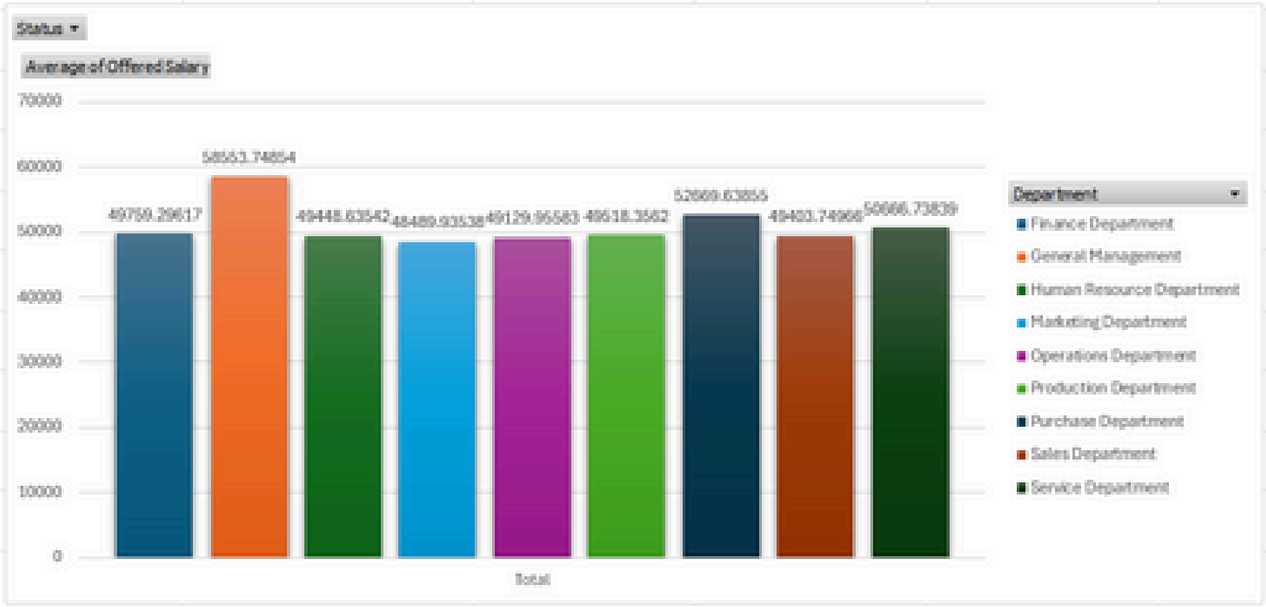
2 This will help you understand the salary distribution.

# HIRING ANALYSIS

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

1
2
3
4
5
6
7
8
9

| | Column Labels | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Finance Department | General Management | Human Resource Department | Marketing Department | Operations Department | Production Department | Purchase Department | Sales Department | Service Department | Grand Total |
| Average of Offered Salary | 49759.29617 | 58553.74854 | 49448.63542 | 48489.93538 | 49129.95583 | 49518.3562 | 52669.63855 | 49403.74966 | 50666.73839 | 50009.9563 |



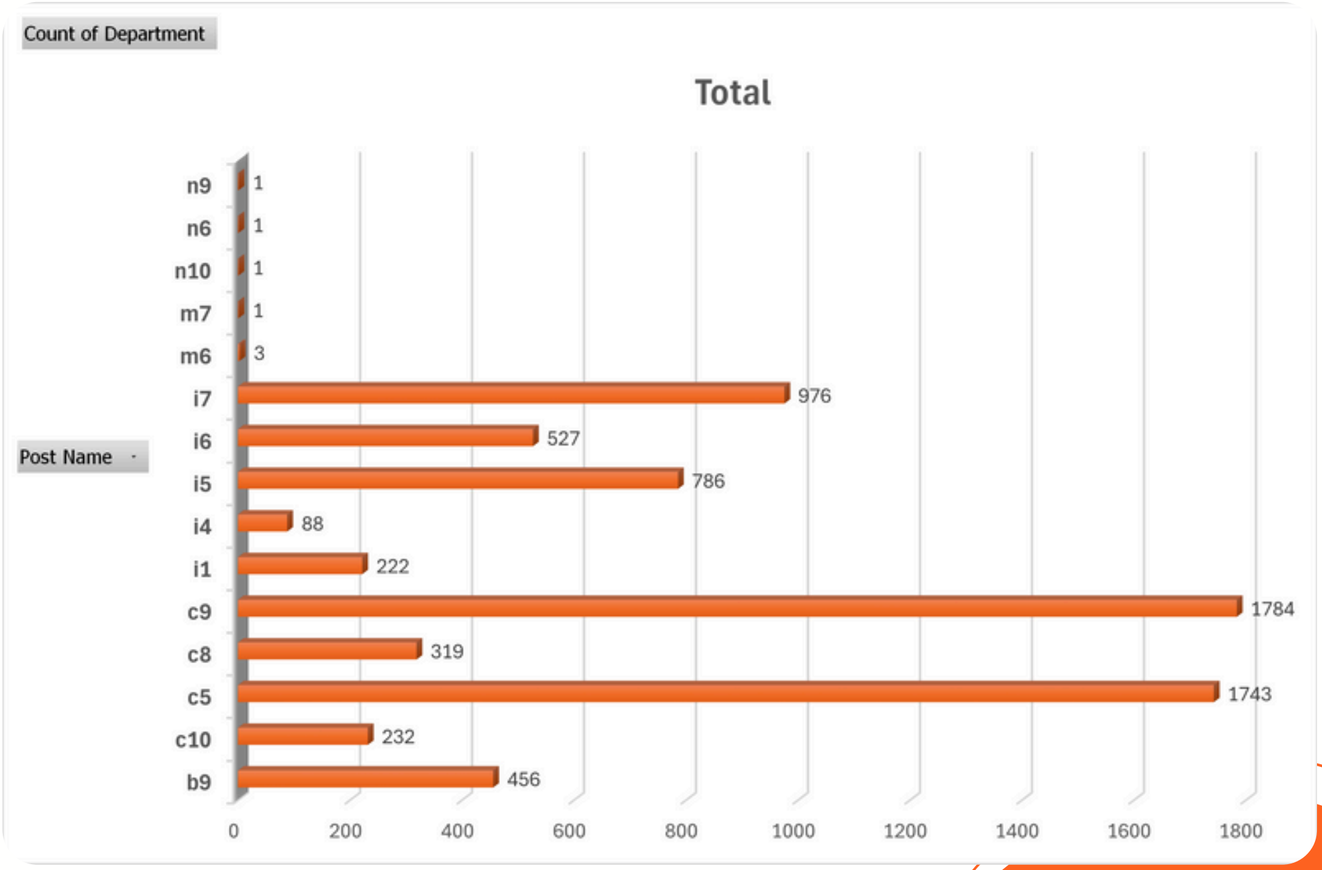
# HIRING ANALYSIS

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

1 2 3 4 5 6 7 8 9

1 Used “Pivot-table” to find the proportion of people working in different departments.

| Row Labels | Count of Department |
|---|---|
| b9 | 456 |
| c10 | 232 |
| c5 | 1743 |
| c8 | 319 |
| c9 | 1784 |
| i1 | 222 |
| i4 | 88 |
| i5 | 786 |
| i6 | 527 |
| i7 | 976 |
| m6 | 3 |
| m7 | 1 |
| n10 | 1 |
| n6 | 1 |
| n9 | 1 |
| Grand Total | 7140 |



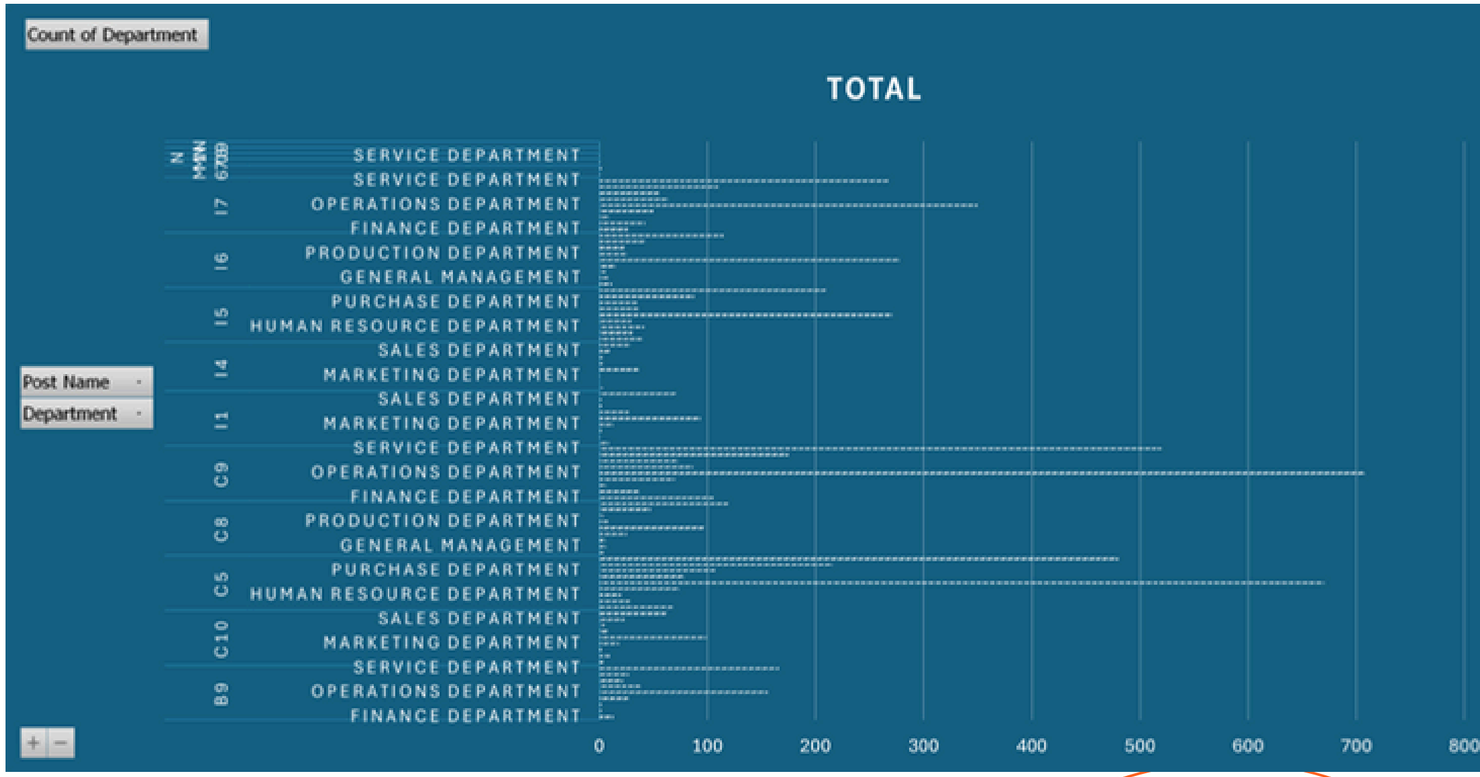
# HIRING ANALYSIS

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

1

Further filtered the pivot table , to find the department wise proportion.

| Row Labels | Count of Department |
|---|---|
| b9 | 456 |
| Finance Department | 13 |
| General Management | 2 |
| Human Resource Departme | 2 |
| Marketing Department | 28 |
| Operations Department | 156 |
| Production Department | 39 |
| Purchase Department | 22 |
| Sales Department | 28 |
| Service Department | 166 |
| c10 | 232 |
| Finance Department | 4 |
| General Management | 10 |
| Human Resource Departme | 2 |
| Marketing Department | 18 |
| Operations Department | 99 |
| Production Department | 8 |
| Purchase Department | 5 |
| Sales Department | 23 |
| Service Department | 63 |
| c5 | 1743 |
| Finance Department | 68 |



# Findings

## 1. Gender Distribution of Hires

- Analysis: Using Excel, we analyzed the dataset for gender distribution in the hiring records. A pivot table was used to summarize the number of male and female hires.

- Finding: Out of the total hires:
  - Males: 54.54% of the total hires.
  - Females: 39.53% of the total hires.

- Insight: There is a gender imbalance in the hiring process. This may indicate the need to focus on diversity initiatives.

## 2. Average Salary Analysis

- Analysis: We used the =AVERAGE() function in Excel to calculate the mean salary.
- Insight: This data point provides a benchmark for understanding the company's compensation trends and helps evaluate if the salaries align with market standards.

***Excel File:***

https://docs.google.com/spreadsheets/d/1J6XKIpgtr5qcg0dEN2KIeFg8kCreyxTk/edit?usp=sharing&ouid=103428047773693985368&rtpof=true&sd=true

# Findings

### 3. Salary Distribution

- Analysis: Salary data was grouped into class intervals using Excel's histogram feature .

- Insight: The data shows a concentration of salaries in the mid-range, with fewer employees in higher salary brackets. This indicates potential growth limitations or senior-level hiring restrictions.

### 4. Departmental Analysis

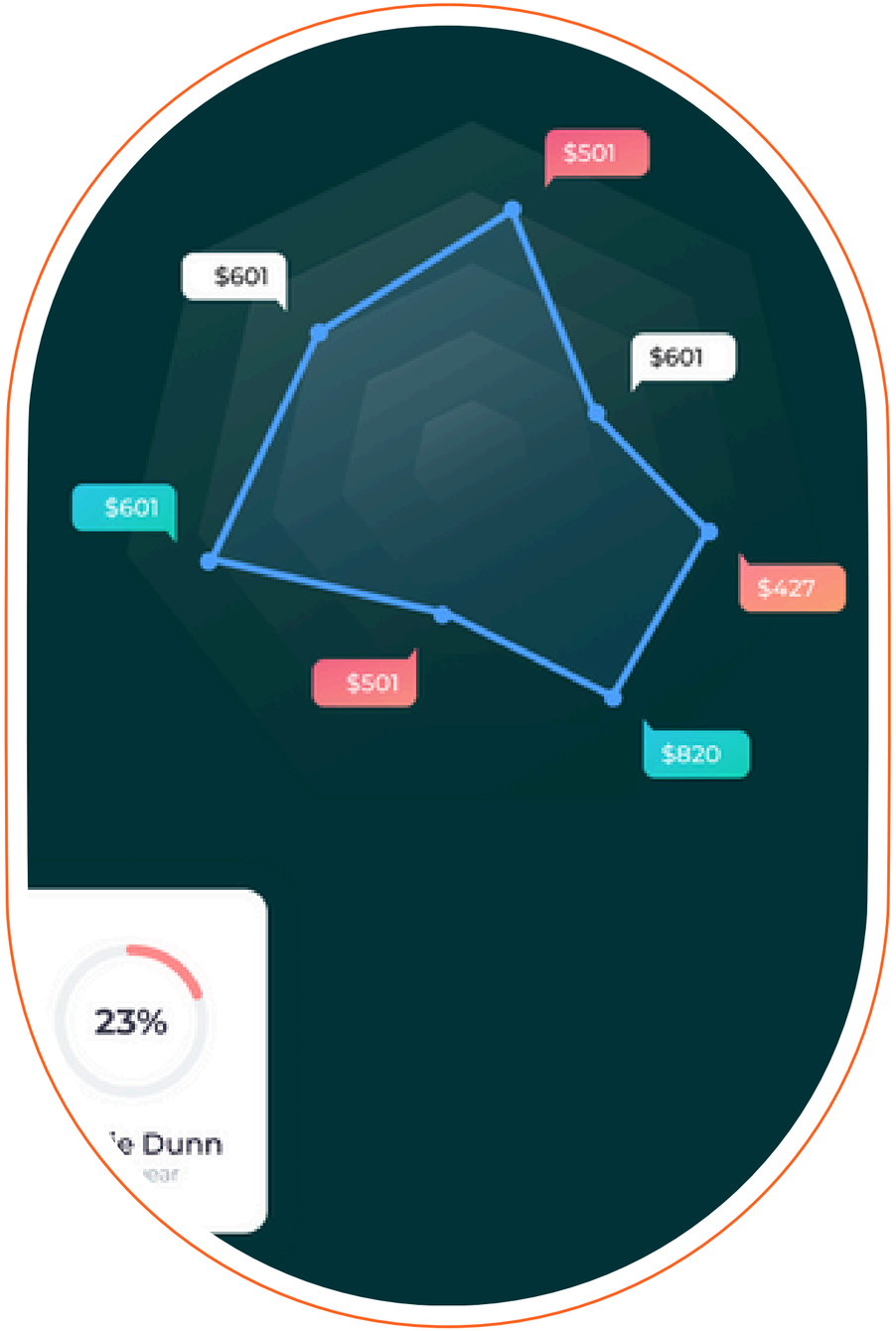
- Analysis: A pie chart was created to visualize the proportion of employees across various departments.
- Finding:
  - Operations Dept: 38%
  - HR: 1.34%

- Insight: Operations is the most dominant department in terms of headcount, while HR and others have the least representation.

# Findings

**The Hiring Process Analytics project provided a comprehensive understanding of the company's recruitment trends and identified several actionable insights:**

- **Gender Diversity:** The company’s hiring process is skewed towards male candidates. Introducing diversity initiatives and programs may help achieve a more balanced workforce.
- **Departmental Distribution:** Operations dominates the workforce, while other departments like HR have a smaller presence. This distribution aligns with the company’s core focus but also suggests room for growth in support and strategic departments.
- **Position Distribution:** The pyramid structure shows a strong focus on hiring at junior and mid-level positions. There might be potential for enhancing leadership development programs to prepare employees for senior-level roles.
- **Data Integrity:** Addressing missing data and outliers helped improve the accuracy and reliability of the analysis.




# *Thank you*