

# Hiring Process Analytics

## Statistics

### Project Description

The hiring process is a crucial function of any organization, as it involves bringing new individuals into the company for various roles. Understanding the trends and patterns of the hiring process can provide valuable insights for the hiring department and help them improve their hiring decisions and strategies.

The objective of this project was to analyze the hiring process data of a multinational company like Google and draw meaningful insights from it. The data contained records of previous hires, such as their ***application\_id, interview date, status, event\_name, department, post name, and offered salary***. The data had **7168** records and **7** columns.

The project involved data analysis and reporting using Microsoft Excel. The results of the analysis were presented in a report that included charts, graphs, tables, and statistics to illustrate the findings. The report also discussed the insights and knowledge gained during the project and how they can contribute to the understanding of hiring process analytics.

### Approach

The approach followed in this project was to use Microsoft Excel to perform data analysis and answer certain questions that can help the company improve its hiring process. The approach consisted of the following steps:

1. **Data Cleaning and Preparation:** The first step was to check the quality and completeness of the data. This involved handling missing values, clubbing columns, and detecting and removing outliers. Missing values were replaced with the median or mode of the respective column, depending on the data type. Columns with multiple categories that could be combined were clubbed together to simplify the analysis. Outliers were identified using box plots and removed from the dataset.
2. **Data Summarization:** The next step was to summarize the data using descriptive statistics and visualizations. This involved calculating averages, medians, modes, standard deviations, and other measures for the numerical variables. It also involved creating frequency tables, pivot tables, pie charts, bar graphs, histograms, and other charts for the categorical variables. These summaries helped to understand the distribution and characteristics of the data.
3. **Data Analysis and Reporting:** The final step was to use the results of the data summarization to answer the questions mentioned in the project details. The questions

were related to hiring analysis, salary analysis, salary distribution, departmental analysis, and position tier analysis. The answers were supported by relevant statistics and visualizations from the previous step. The answers were also discussed in terms of their implications and recommendations for the company's hiring process.

## Tech-Stack Used

The software used for this project was **Microsoft Excel 365**. Microsoft Excel is a spreadsheet program that allows users to perform various data analysis and reporting tasks. It has many features and functions that can help users manipulate, summarize, visualize, and present data.

The purpose of using **Microsoft Excel** for this project was to leverage its capabilities and tools for data analysis and reporting. Some of the specific techniques and tools used in this project were:

- **Excel Functions and Formulas:** These are predefined calculations that can perform various operations on data, such as arithmetic, logical, statistical, text, date and time, etc. Some of the functions and formulas used in this project were COUNTIF, AVERAGE, MEDIAN, MODE, STDEV, FREQUENCY, IF, AND, OR, etc.
- **Excel Charts and Graphs:** These are graphical representations of data that can help users understand and communicate data better. Some of the charts and graphs used in this project were pie charts, bar graphs, histograms, box plots, scatter plots, etc.
- **Excel Pivot Tables:** These are interactive tables that can summarize large amounts of data by grouping and aggregating them based on different criteria. Pivot tables were used to create frequency for the variables in the data.

## Insights

Here are some of the insights and knowledge I gained while working on this project:

### 1. Data Cleaning and Preparation:

The first step was to check the quality and completeness of the data. This involved handling missing values, and detecting and removing outliers. The specific techniques and tools used for this step were:

#### a) Handling Missing Data:

To check if there were any missing values in the dataset, the **COUNTBLANK** function was used on each column. In some columns like *event\_name* and *Post Name*, there were many “-” values, so we

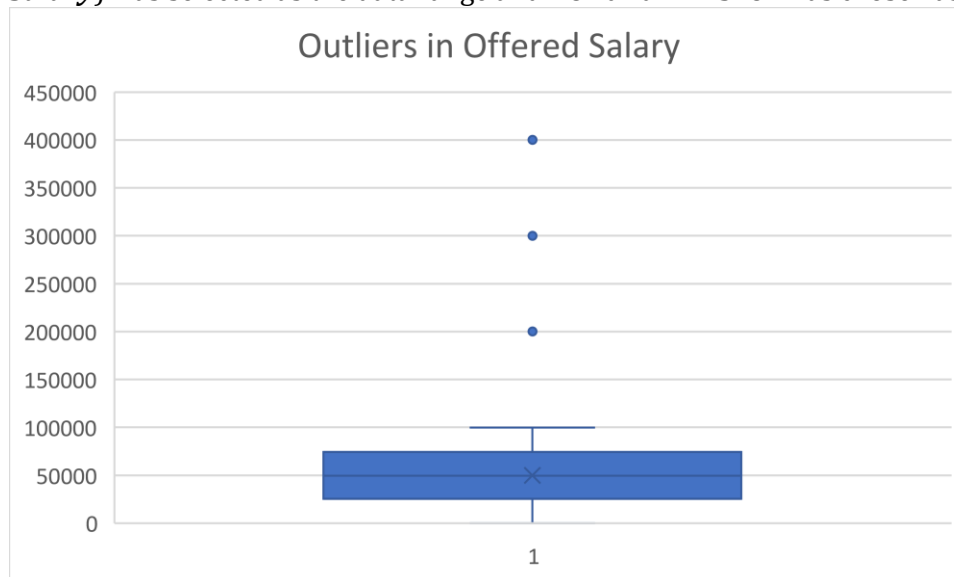
considered them as missing values too. In those columns we used the COUNTIF function. It was found that there were **15** missing values in the **event\_name** column, **1** missing value in the **Post Name** column, and **1** missing value in the **Offered Salary** column.

Number of missing values						
application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
0	0	0	15	0	1	1

To handle these missing values, the median of the Offered Salary column was calculated using the **MEDIAN** function; and for the other two columns, the mode of those columns was calculated using the **MODE** function. These values were used to replace the missing values. This was done using the **IF** and **ISBLANK** functions. The median was chosen as a replacement strategy because it is less sensitive to outliers than the mean.

#### b) Outlier Detection:

To check for outliers in the dataset that may skew the analysis, box plots were created for the numerical variables, such as **Offered Salary**. Box plots are graphical representations of data that show the minimum, maximum, median, first quartile, and third quartile of a distribution. Outliers are values that lie outside the whiskers of the box plot, which are 1.5 times the interquartile range (IQR) above or below the quartiles. To create a box plot for Offered Salary, column G (**Offered Salary**) was selected as the data range and **Box and Whisker** was chosen as the chart type.



From this output, we can see that there were 3 outliers in the **Offered Salary** column. Now we must find a way to handle these outliers.

### c) Removing Outliers:

To handle outliers in the dataset, a decision was made to remove them from the analysis. This was done because outliers can distort the results and affect the accuracy and validity of the analysis. Outliers were removed by deleting the entire row which contained the outliers.

## 2. Data Summarization:

After cleaning and preparing the data, the next step was to summarize the data using descriptive statistics and visualizations. This involved calculating sums, averages, medians, modes, standard deviations, and other measures for the numerical variables. It also involved creating pivot tables and pivot charts for the categorical variables. These summaries helped to understand the distribution and characteristics of the data. The specific techniques and tools used for this step were:

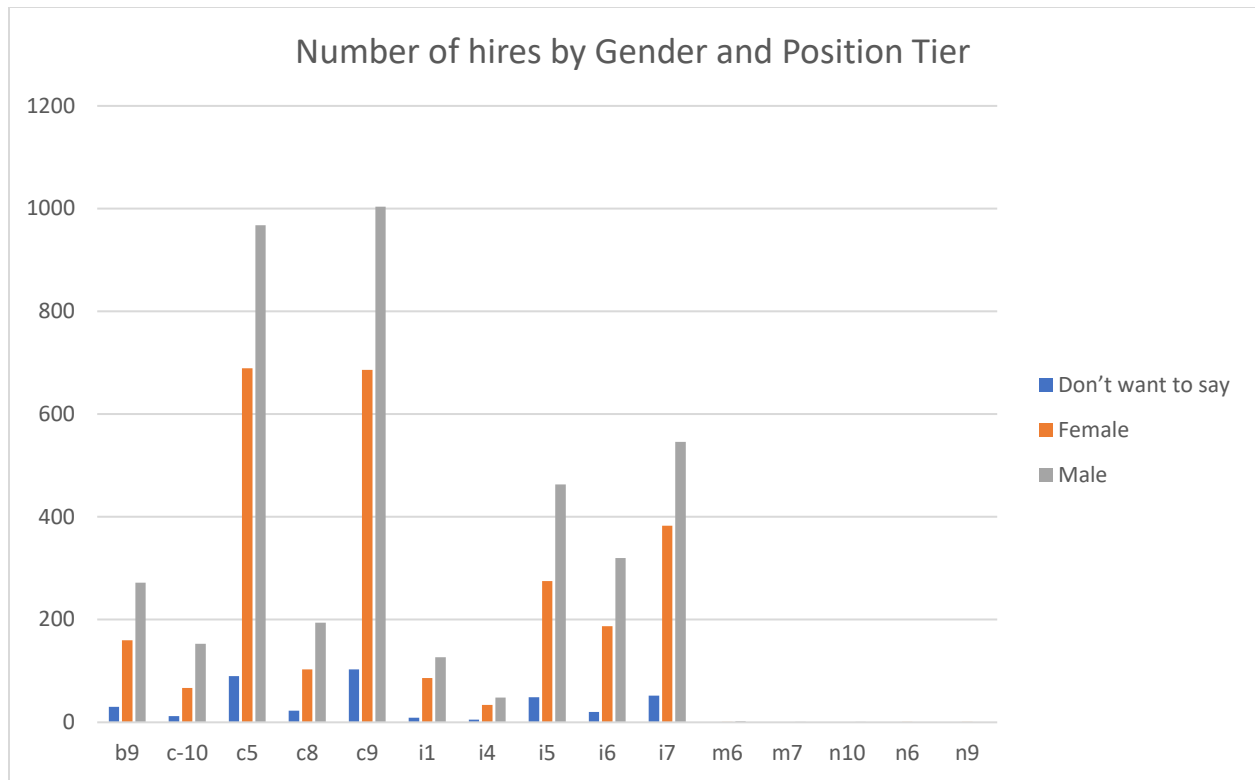
### a) Task 1: Weekly User Engagement:

These are predefined calculations that can perform various operations on data, such as arithmetic, logical, statistical, text, date and time, etc. Some of the functions and formulas used in this project were SUM, AVERAGE, MEDIAN, MODE, STDEV.

Offered Salary	Value
Sum	357377994
Average/Mean	49878.29644
Median	49625
Mode	72843
Standard Deviation	28349.69127

### b) Excel Pivot Tables and Pivot Charts:

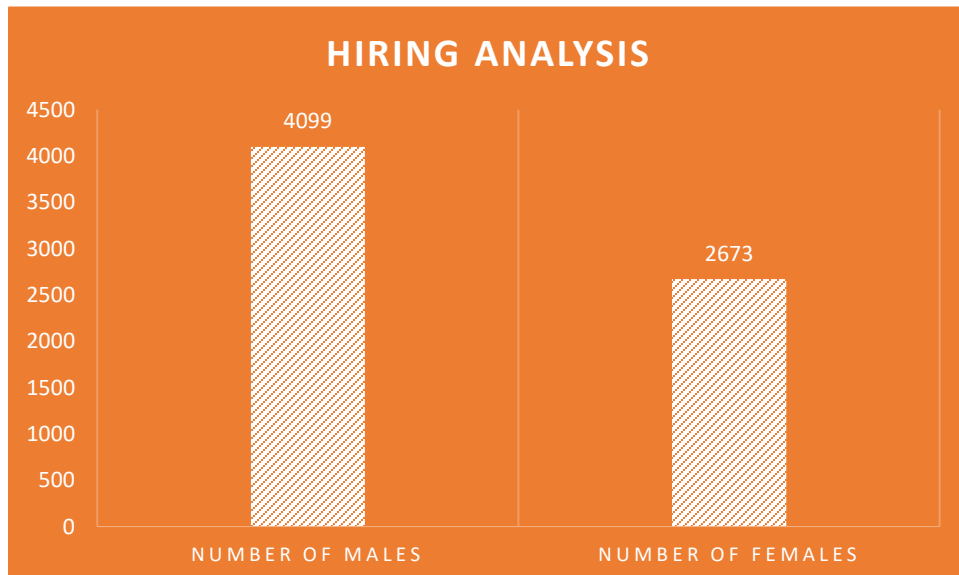
These are interactive tables and charts that can summarize large amounts of data by grouping and aggregating them based on different criteria. **Pivot tables** and **pivot charts** were used to create frequency tables and cross-tabulations for the categorical variables in the data. To create a frequency table and a pivot chart that show the number of hires by gender and position tier, select columns F (**Post Name**) and D (**event\_name**) as the data range and insert a pivot table from the Insert PivotTable option. To create a pivot chart from this pivot table, select any cell in the pivot table and click PivotChart from the Analyze tab. Excel will create a pivot chart that shows the same information as the pivot table in a graphical form.



### 3. Data Analysis:

#### c) Task 1: Hiring Analysis:

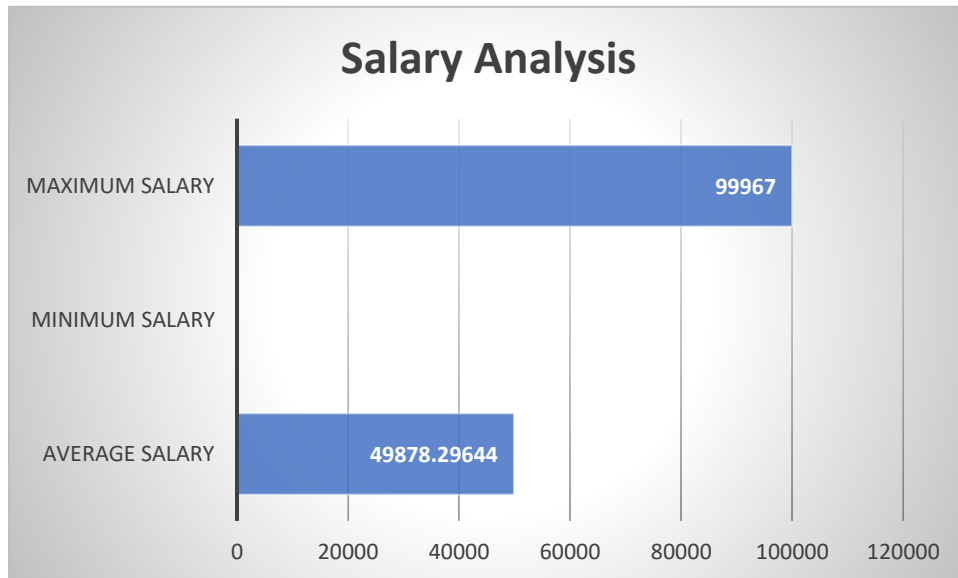
One of the tasks that the project involved was to determine the gender distribution of hires.



The analysis showed that the company has hired more males than females in the past. Out of 7165 hires, **4099** were **males** and **2673** were **females**. This could indicate a gender imbalance in the hiring process or a lack of diversity in the applicant pool.

d) Task 2: Salary Analysis:

Another task that the project involved was to calculate the average salary offered by the company.



The analysis showed that the **average** salary offered by the company was **\$49,878.29**. However, there was a large variation in the salary data, as the **minimum** salary was **\$100** and the **maximum** salary was **\$99,967**. This means that most of the salaries were clustered around the middle range.

e) Task 3: Salary Distribution:

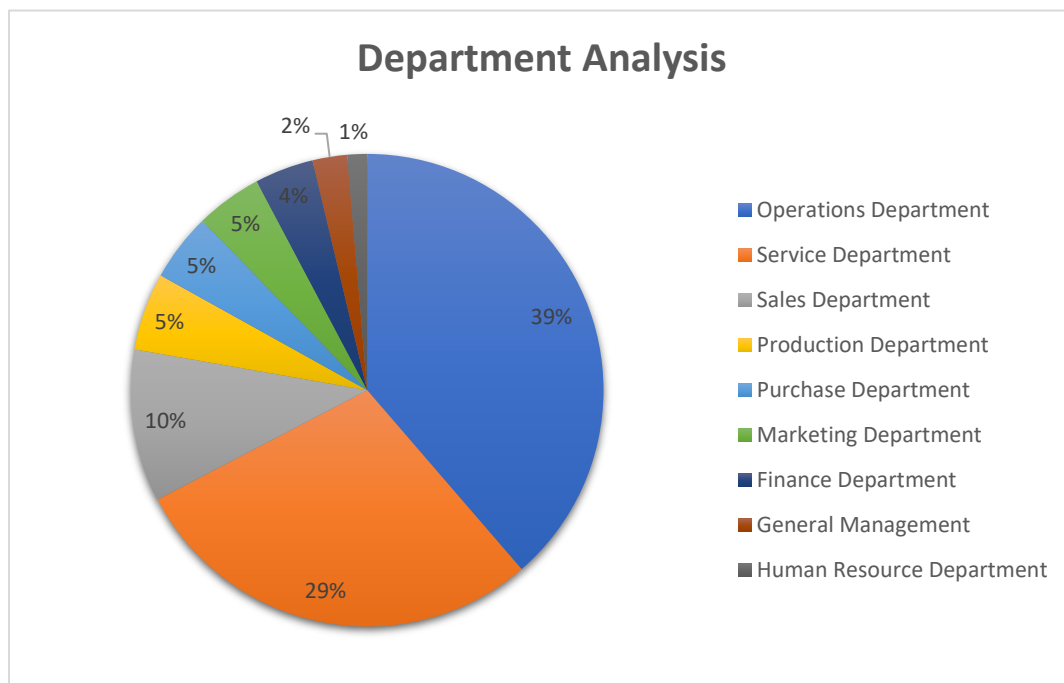
A third task that the project involved was to create class intervals for the salaries in the company.



The analysis showed that most of the salaries fell into the **\$40,000-\$60,000** range. There were **1532** salaries in this range. The next most common range was **\$60,000-\$80,000**, with **1432** salaries. The least common range was **\$80,000-\$100,000**, with **1370** salaries. This shows that the company has a perfect distribution of salaries.

f) Task 4: Departmental Analysis:

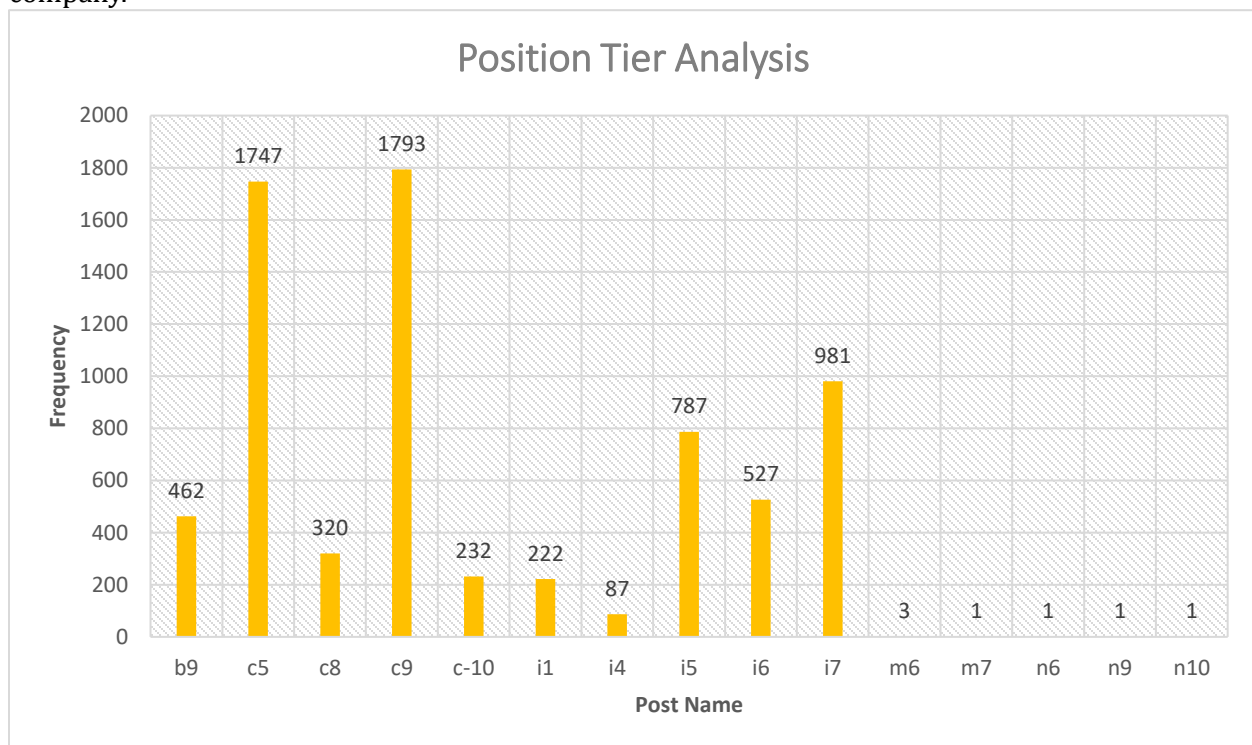
A fourth task that the project involved was to show the proportion of people working in different departments.



The analysis showed that the company had **9** different departments, with varying numbers of employees. The most populated department was **Operations**, with **39%** of employees. The least populated department was **Human Resources**, with only **1%** of employees. The other departments were **Service** (**29%** of employees), **Sales** (**10%** of employees), **Production** (**5%** of employees), **Purchase** (**5%** of employees), **Marketing** (**5%** of employees), **Finance** (**4%** of employees), and **General Management** (**2%** of employees). This shows that the company has a diverse and multidisciplinary workforce, with different functions and roles.

g) Task 5: Position Tier Analysis:

A fifth task that the project involved was to represent the different position tiers within the company.



The analysis showed that the company had many different tiers of positions, with varying levels of responsibility and salary. The most common tier was **c9**, with **1793** positions. The least common tiers were **m7**, **n6**, **n9**, and **n10**, with only **1** position each. The other important tiers were **c5** (**1747** positions), **i7** (**981** positions), **i5** (**787** positions), and **i6** (**527** positions). This shows that the company has a hierarchical structure, with more positions for higher tiers than lower tiers.



## Result

The report achieved the objective of the project, which was to analyze the hiring process data of a multinational company like Google and draw meaningful insights from it. The report contributed to the understanding of hiring process analytics by providing answers to the following questions:

- What is the gender distribution of hires?
- What is the average salary offered by the company?
- How is the salary distributed across different ranges?
- What is the proportion of people working in different departments?
- What are the different position tiers within the company?

The report also discussed the implications and recommendations for the company's hiring process based on the findings and insights. The report suggested that the company could:

- Address the gender gap in hiring and promote diversity and inclusion in its workforce.
- Review its salary structure and ensure that it is fair and competitive in the market.
- Explore new ways to attract and retain talent in different departments and position tiers.
- Monitor and evaluate its hiring performance and outcomes over time.

The report demonstrated the skills and knowledge of data analysis and reporting using Microsoft Excel. The report also showcased the ability to use data to draw meaningful conclusions and provide actionable insights for business decisions. The report was a valuable and informative resource for the leadership team of the company.

[Excel Sheet File](#)