Hiring Process Analytics

Project Description:

The purpose of this project is to analyse the hiring process data of a multinational company like Google to derive actionable insights that can enhance decision making and optimise recruitment strategies. The hiring process is an important function of any company and understanding trends such as number of rejections, interviews, job type and vacancies can provide valuable insights for the hiring department.
The dataset containing records of previous hires has been provided by the company. By utilising Excel and data visualisation technique the following analysis will be done on the provided dataset to provide a comprehensive understanding of company's hiring process:

1. **Hiring Analysis:** The hiring process involves recruiting new individuals in the company for different roles. Understanding the gender distribution in hires will help to access gender diversity and to identify any potential biases.
2. **Salary Analysis:** Calculating and analysing the average salary offered by the company will help gain insights in company's compensation practices.
3. **Salary Distribution:** Class intervals represent ranges of values and in our dataset salary ranges. Understanding salary distribution through class intervals enhances the analysis on company's compensation practices.
4. **Departmental Analysis:** Visualising distribution of employees across different departments provides insights on strength and weaknesses of different departments and overall structure of the company.
5. **Position Tier Analysis:** Different positions within the company often have different tiers or levels. Analysing the distribution of positions across different levels will help in understanding the hierarchy and structure of job roles within the company.

Approach:

To accomplish the necessary tasks and to finalize the project, the approach to this project involves a structured methodology to analyse the hiring process data. The following steps were executed:

1. **Understanding the requirements of data:** Firstly, the provided dataset was imported into Excel. The raw data was converted into tabular form to handle the data efficiently for analysis.
   Then the structure of the data was observed. The dataset has the following details:
   - Number of data points: 7168
   - Number of columns: 7
   - Column details:
     1. application_id: ID of the applicant applying for job.
     2. Interview Taken on: Date and time of job interview.

3. Status: Applicant hired or rejected for the job.
4. event_name: Gender of the applicant.
5. Department: Name of the department for which job interview was conducted.
6. Post Name: Name of the post offered.
7. Offered Salary: Salary amount offered for the job.

2. **Data Cleaning and Preparation:** To summarize the findings and derive actionable insights, it is crucial to handle them appropriately.
    - **Handling missing data, duplicate data and errors:** The datasets will be checked for missing values, duplicate values and errors and the best strategy to handle and rectify them will be determined.
      In the provided dataset, column name application_id has 54 rows that have duplicate values, since application_id should be unique to each applicant hence, it needs to be removed otherwise it will create discrepancy in the statistical calculations. Here are few rows highlighting the duplicating values.

	A	B	C	D	E	F	G
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
180	55833	23-08-2014 14:11	Hired	Male	Operations Department	c5	88439
409	55833	17-07-2014 16:46	Rejected	Female	Purchase Department	c9	17750
411	201849	21-07-2014 08:23	Hired	Male	Service Department	c8	86328
590	201849	16-05-2014 15:44	Hired	Male	Operations Department	b9	17810
850	229281	08-08-2014 17:46	Hired	Male	Sales Department	i7	10461
906	229281	15-07-2014 08:38	Hired	Male	Operations Department	i6	92128
919	241980	15-08-2014 09:32	Hired	Female	Production Department	c9	40276
922	241980	23-05-2014 14:14	Hired	Female	Service Department	i7	75669
1087	287111	25-06-2014 17:41	Rejected	Female	Service Department	c9	50460
1525	287111	19-08-2014 10:46	Hired	Male	Service Department	c5	18169
1587	326373	29-06-2014 12:39	Hired	Male	Service Department	i7	72303

Column name event_name has 15 rows whose value is “- “. It does not imply any gender, so to rectify it has been replaced by “Don’t want to say” which implies the gender of the applicant is not known.

	A	B	C	D	E	F	G
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
17	195323	09-05-2014 12:48	Hired	Don’t want to say	Service Department	i7	81757
19	742283	02-05-2014 08:11	Rejected	Don’t want to say	Service Department	i5	100
1602	227046	27-08-2014 18:08	Hired	Don’t want to say	Operations Department	b9	76730
1791	711350	16-07-2014 13:33	Rejected	Don’t want to say	Operations Department	c-10	25785
2878	835053	16-05-2014 18:34	Hired	Don’t want to say	Operations Department	c5	25583
3259	444043	11-07-2014 14:52	Hired	Don’t want to say	Sales Department	c5	80262
4018	352309	20-08-2014 10:38	Hired	Don’t want to say	Service Department	i5	4308
4126	204014	09-08-2014 16:09	Rejected	Don’t want to say	Purchase Department	c5	96396
4410	901867	18-08-2014 09:36	Rejected	Don’t want to say	Service Department	c5	22393
5560	937905	08-08-2014 19:29	Hired	Don’t want to say	Marketing Department	c9	94032
5607	564743	28-08-2014 10:25	Rejected	Don’t want to say	Production Department	c9	4076
5889	245473	14-05-2014 18:48	Hired	Don’t want to say	Service Department	c5	66948
6330	411295	22-06-2014 14:38	Hired	Don’t want to say	Operations Department	i1	98070
6658	487617	30-05-2014 16:29	Hired	Don’t want to say	Service Department	c8	12470
6998	827628	30-08-2014 15:51	Hired	Don’t want to say	Service Department	i1	3134

Column Post name has value “c-10” which does not align with the format of other post name. So, it has been replaced by “c10”.

	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
7	289907	01-05-2014 07:44	Hired	Male	Sales Department	c10	85914
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

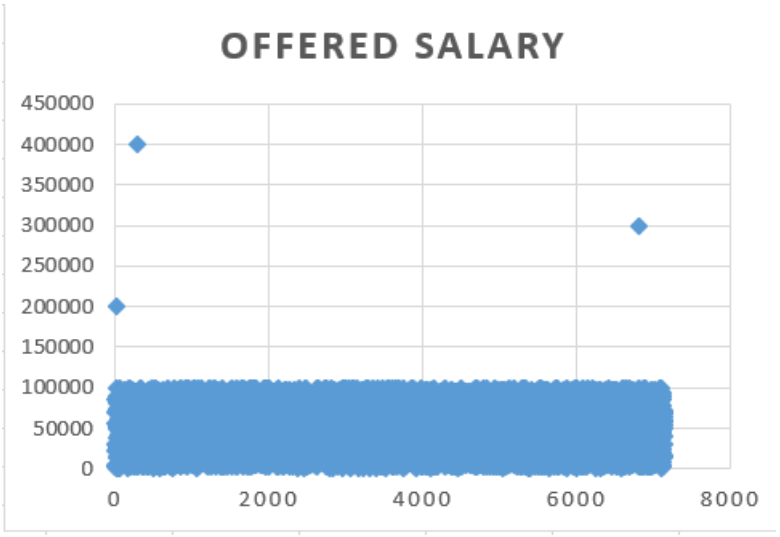
Column Post name has 1 row with value "- "which has been replaced by c10 after performing some filtration on corresponding Department and Offered Salary columns.

	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	c10	85914

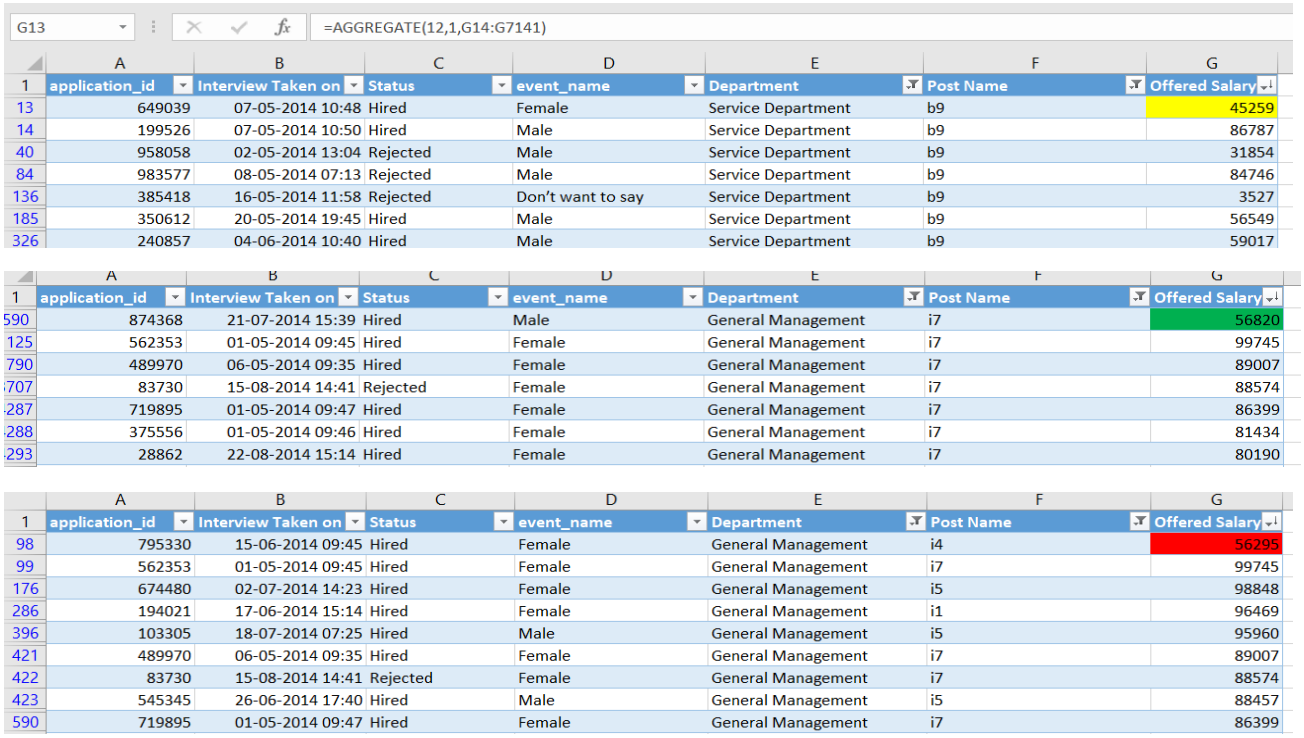
Column Offered Salary has 1 blank cell. To replace it filtration has been performed on corresponding Department and Post name columns and then median of Offered Salary has been calculated which came out to be 45400.

	A	B	C	D	E	F	G
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
077	114584	07-05-2014 08:08	Rejected	Male	Sales Department	i7	45400

- **Clubbing columns:** Columns with multiple categories that can be combined will be grouped together. Our dataset doesn't have columns that needs to be clubbed.
- **Detecting and handling outlier:** Statistical methods will be used to identify outliers and the best strategy to handle them will be determined. The scatter plot of Offered salary shown below shows three column values which are outliers and the values are 200000,300000,400000.



The outlier values 200000,300000,400000 has been replaced by the median value of column Offered Salary for corresponding Department and Post Name



G13 =AGGREGATE(12,1,G14:G7141)

	A	B	C	D	E	F	G
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
13	649039	07-05-2014 10:48	Hired	Female	Service Department	b9	45259
14	199526	07-05-2014 10:50	Hired	Male	Service Department	b9	86787
40	958058	02-05-2014 13:04	Rejected	Male	Service Department	b9	31854
84	983577	08-05-2014 07:13	Rejected	Male	Service Department	b9	84746
136	385418	16-05-2014 11:58	Rejected	Don't want to say	Service Department	b9	3527
185	350612	20-05-2014 19:45	Hired	Male	Service Department	b9	56549
326	240857	04-06-2014 10:40	Hired	Male	Service Department	b9	59017

	A	B	C	D	E	F	G
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
590	874368	21-07-2014 15:39	Hired	Male	General Management	i7	56820
125	562353	01-05-2014 09:45	Hired	Female	General Management	i7	99745
790	489970	06-05-2014 09:35	Hired	Female	General Management	i7	89007
707	83730	15-08-2014 14:41	Rejected	Female	General Management	i7	88574
287	719895	01-05-2014 09:47	Hired	Female	General Management	i7	86399
288	375556	01-05-2014 09:46	Hired	Female	General Management	i7	81434
293	28862	22-08-2014 15:14	Hired	Female	General Management	i7	80190

	A	B	C	D	E	F	G
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
98	795330	15-06-2014 09:45	Hired	Female	General Management	i4	56295
99	562353	01-05-2014 09:45	Hired	Female	General Management	i7	99745
176	674480	02-07-2014 14:23	Hired	Female	General Management	i5	98848
286	194021	17-06-2014 15:14	Hired	Female	General Management	i1	96469
396	103305	18-07-2014 07:25	Hired	Male	General Management	i5	95960
421	489970	06-05-2014 09:35	Hired	Female	General Management	i7	89007
422	83730	15-08-2014 14:41	Rejected	Female	General Management	i7	88574
423	545345	26-06-2014 17:40	Hired	Male	General Management	i5	88457
590	719895	01-05-2014 09:47	Hired	Female	General Management	i7	86399

3. **Data Summary:** After cleaning and preparing the data, statistical measures will be calculated like median averages, etc. and visualization will be done to derive key insights.

- ## Tech-Stack used

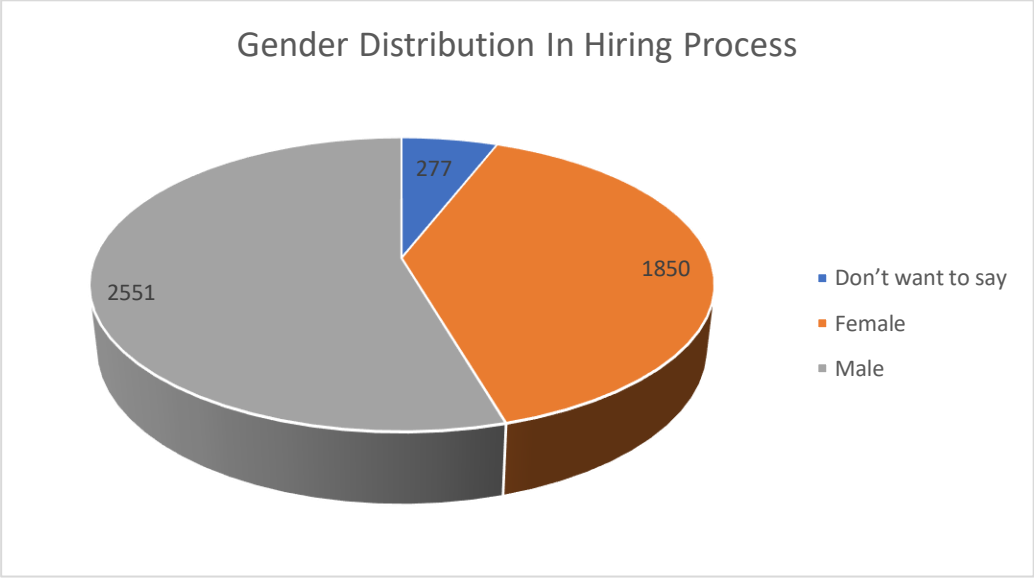
Microsoft Excel 2019 has been used to clean and prepare the dataset for analysis. It was also used for basic statistical calculations. Pivot tables were created, Excel charts were utilised for visualisation and potential outliers were identified.

The reason behind using excel was that is widely used and easy to understand. It is great tool for analysing data for large datasets and works well with other tools like Word and PowerPoint, so it is easy to create reports and presentation based on the analysis. It is cost effective as it is often installed in our computers.

- ## Insights:

1. **Hiring Analysis:** The task is to determine the representation of males and females in company's hiring process.

	A	B	C
1	Status	Hired	
2			
3	**Row Labels**	**Count**	**Percenatge**
4	Don't want to say	277	5.92%
5	Female	1850	39.55%
6	Male	2551	54.53%
7	**Grand Total**	**4678**	**100.00%**
8			



Insights: From the above data it can be observed that male applicants are hired more than female applicants whereas the remaining applicants have not disclosed their gender.

The hiring bias towards male applicants indicate that the company unintentionally prefers men over women. This could mean unfair practices in how they select the applicants.

Since female applicants are less likely to get hired, it could mean unfair treatment that discourages them from applying. The company should make sure that everyone is given equal chance.

Since many applicants are not disclosing their gender, it raises questions on company's policies on gender inclusivity. The company should look into these policies and practices to support diversity.

2. **Salary Analysis:** The task is to use excel function to calculate the average salary offered by the company. The average salary is calculated by adding the salaries of a group of employees and then dividing the total by the number of employees.

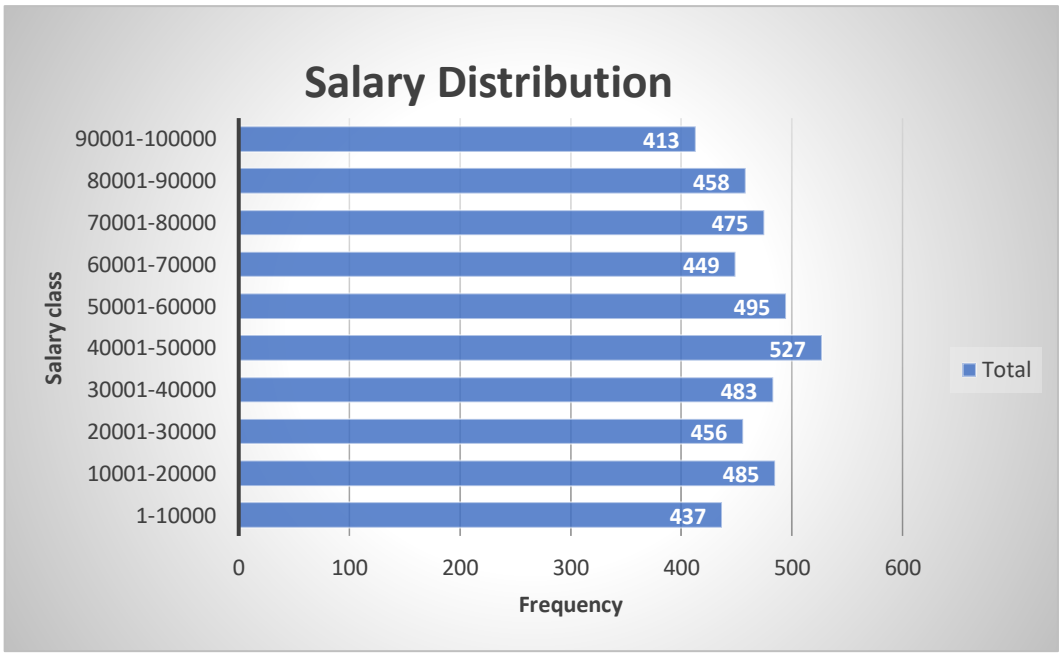
15	**Average of Offered Salary**
16	49893.32909
17	

Insights: From the above data it can be observed that the average salary offered by the company is approximately 49893. It indicates company's compensation level.

The company can compare their average salary to the industry standards and identify if the company is competitive enough in attracting applicants. Higher average salary can attract good talents and as well as encourage them to stay with the company. By analysing the average salary, the company can have understanding of company's budget as well as help in future hiring process and compensation needs.

3. **Salary Distribution:** The task is to create class intervals for the salaries in the company to understand the salary distribution. The class interval is the difference between the upper and lower limits of a class.

Status	Hired
Row Labels	**Frequency**
1-10000	437
10001-20000	485
20001-30000	456
30001-40000	483
40001-50000	527
50001-60000	495
60001-70000	449
70001-80000	475
80001-90000	458
90001-100000	413
Grand Total	**4678**

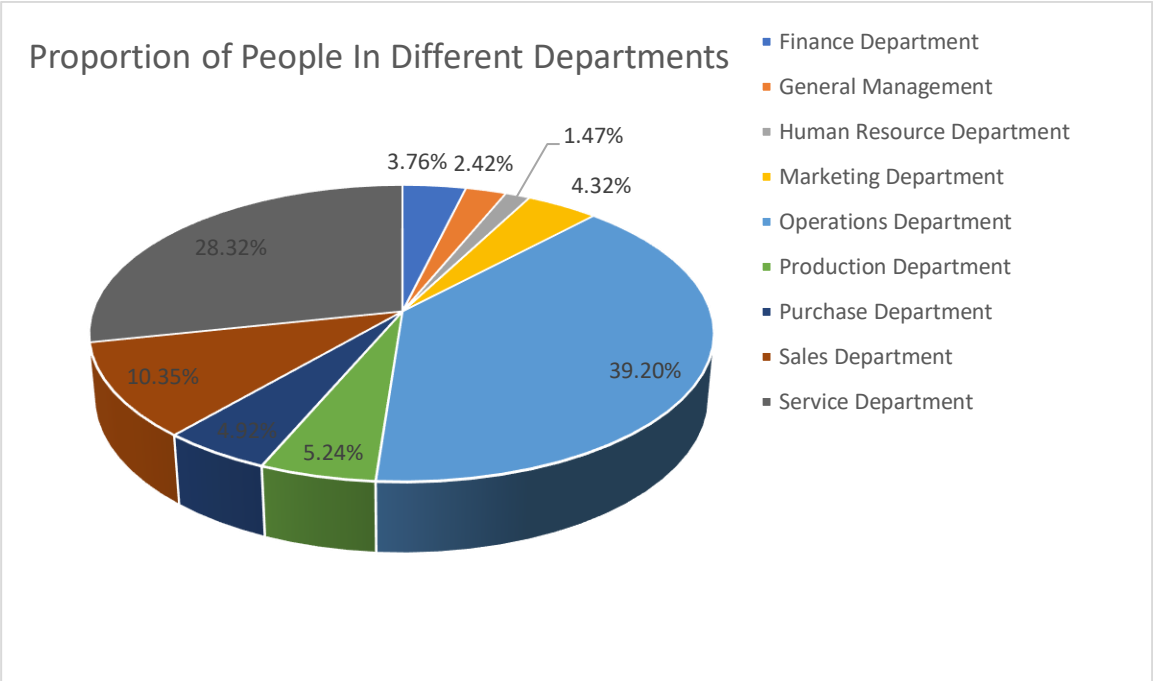


Insights: From the above data it can be observed that most of the applicants hired by the company fall between the salary range of 40001-50000. The different distribution indicates that the salary might based on different positions and department within the company.

By analysing the salary distribution, the company can focus if the salaries are fair and equal and the company pays competitively according to industry standards. The company can review if certain departments or roles are offering salary more than others and if the company is attractive to top talents.

4. **Departmental Analysis:** The task is to use pie chart to show the proportion of people working in different departments.

Status	Hired	
Row Labels	**Department wise Proportion**	**Count of Department**
Finance Department	3.76%	176
General Management	2.42%	113
Human Resource Department	1.47%	69
Marketing Department	4.32%	202
Operations Department	39.20%	1834
Production Department	5.24%	245
Purchase Department	4.92%	230
Sales Department	10.35%	484
Service Department	28.32%	1325
Grand Total	**100.00%**	**4678**

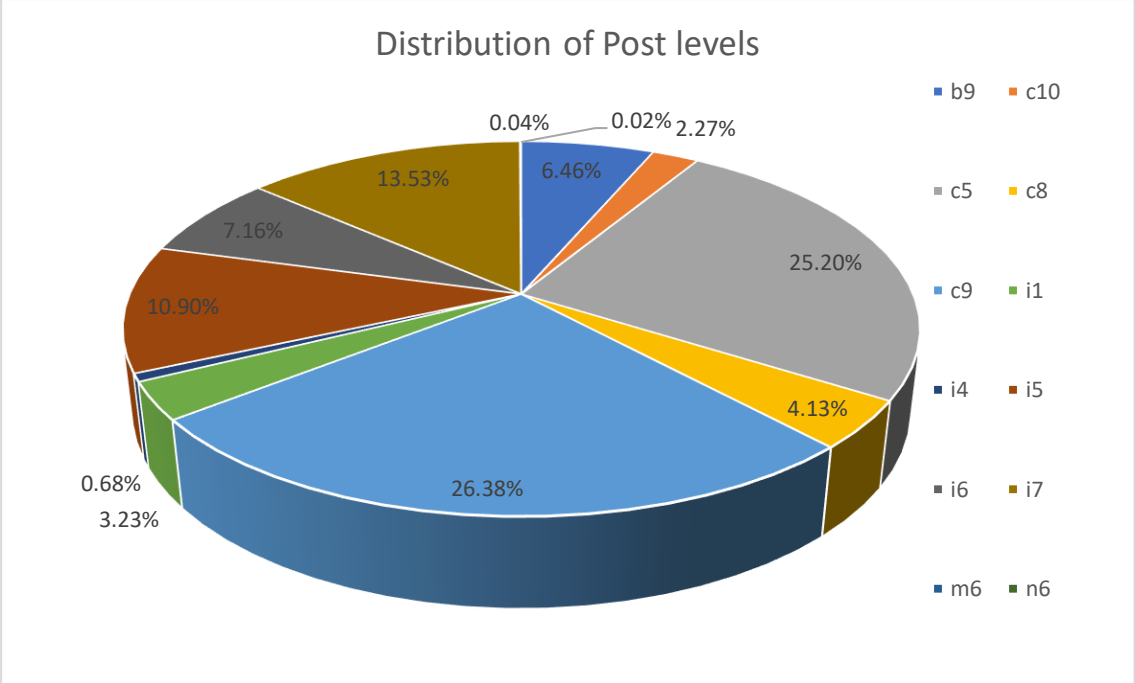


Insights: From the above data it can be analysed that the company hires most applicants in Operations Department and least in Human resource department. The above data indicates the company is seeking talent and expertise in the operation department the most. The reason can be vacant positions due to previous

employee leaving the company or necessity of new positions in the particular department.
Since the HR department has least hiring indicates that there are few positions available for the department or hiring is freeze in that particular department.
The company needs to analyse the above factors and improve their talent acquisition strategies according to company's needs and requirements.

5. **Position Tier Analysis:** The task is to use chart or graph to represent the different position tiers within the company to understand the distribution of position across different tiers.

Status	Hired	
Row Labels	**Distribution of Post levels**	**Count of Post Name**
b9	6.46%	302
c10	2.27%	106
c5	25.20%	1179
c8	4.13%	193
c9	26.38%	1234
i1	3.23%	151
i4	0.68%	32
i5	10.90%	510
i6	7.16%	335
i7	13.53%	633
m6	0.04%	2
n6	0.02%	1
Grand Total	**100.00%**	**4678**



Insights: From the above data it can be observed that the company hired most applicant for the post level c9 and the least for post level n6.
By analysing the above data, it indicates that the company is hiring for specialised role c9 indicating need for skills related to the position. Since the hiring is in bulk in this position, it might indicate that it is an entry level role which help in company's future growth.
The other post levels where hiring is not in bulk indicates they are either specialized roles or senior roles which requires certain skill sets and work experience.
The analysis will help company's talent development strategies, growth plans and leadership priorities.

- **Results:**

Through this project as a data analyst in Google, I gained significant insights that proved to be valuable. It deepened my knowledge of data analysis in organisations and process of making data driven decisions. This experience helped in improving my analytical skills and MS-Excel skills. I became comfortable in performing various functions of MS-Excel, and learned about pivot tables, charts as well as how to detect outliers. I gained experience in data pre-processing like data cleaning. I extracted meaningful insights from datasets and learnt how analysing hiring patterns and salary data can help in business decisions, strategic planning and talent management.

In conclusion this project was not just about analysing the data but also using insights to improve company's hiring process. This experience will further help in boosting my career as data analyst in uncovering valuable insights from data and making informed decisions.

Link to Excel Sheet