

Hiring Process Analytics

Project Description:

This project aims to provide insights into the hiring process based on the dataset provided. The insights derived will be beneficial for the hiring department in improving its hiring process. As a data analyst, we received a dataset containing records of previous hires along with specific questions to analyze the data and provide insights accordingly.

Approach:

- Understanding the Dataset: Initially, we thoroughly examined the dataset to understand its structure and the information it contained.
- Data Pre-Processing: We performed data pre-processing to ensure data integrity. This involved handling missing data, correcting errors, detecting and handling outliers, and removing duplicates, especially for the applicant ID.
- Filtering and Analysis: We directly proceeded to filter the data based on the questions provided. Utilizing Microsoft Excel, we extracted relevant information and performed analysis to answer each question accurately.
- Visualization: To enhance our understanding and communicate insights effectively, we created visualizations such as pivot tables and charts. These visualizations aided in identifying trends, patterns, and anomalies in the data.
- Conclusion and Insights: After analyzing the filtered data and interpreting the visualizations, we drew conclusions and extracted meaningful insights. These insights were aligned with the questions posed and provided valuable information for the hiring department.
- Report Generation: Finally, we compiled a detailed report summarizing our findings, insights, and recommendations. This report will be presented to the hiring department for their review and consideration in optimizing the hiring process.

Tech-Stack Used:

- Microsoft Excel: Version 2022
- Pivot tables, charts, and Excel functions: Utilized for data manipulation, analysis, and visualization.

- a. Column **event_name** has **15** rows with “-“ as its values. It can be termed as Null value. 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.

application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
195323	09-05-2014 12:48	Hired	-	Service Department	i7	81757
742283	02-05-2014 08:11	Rejected	-	Service Department	i5	100
227046	27-08-2014 18:08	Hired	-	Operations Department	b9	76730
711350	16-07-2014 13:33	Rejected	-	Operations Department	c-10	25785
835053	16-05-2014 18:34	Hired	-	Operations Department	c5	25583
444043	11-07-2014 14:52	Hired	-	Sales Department	c5	80262
352309	20-08-2014 10:38	Hired	-	Service Department	i5	4308
204014	09-08-2014 16:09	Rejected	-	Purchase Department	c5	96396
901867	18-08-2014 09:36	Rejected	-	Service Department	c5	22393
937905	08-08-2014 19:29	Hired	-	Marketing Department	c9	94032
564743	28-08-2014 10:25	Rejected	-	Production Department	c9	4076
245473	14-05-2014 18:48	Hired	-	Service Department	c5	66948

- b. Column **Offered Salary** has **1** row with **Null Value**. The corresponding value in **Department** column is “**Sales Department**” and **Post Name** is “**i7**”. So, we replaced it with median of **Offered Salary** for **Sales Department** and **i7 Post Name**. The median came out to **45400**.

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

A	B	C	D	E	F	G
application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
						45400

- c. Column **Post Name** has **1** row with “-“ as its value. It can be termed 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**”.

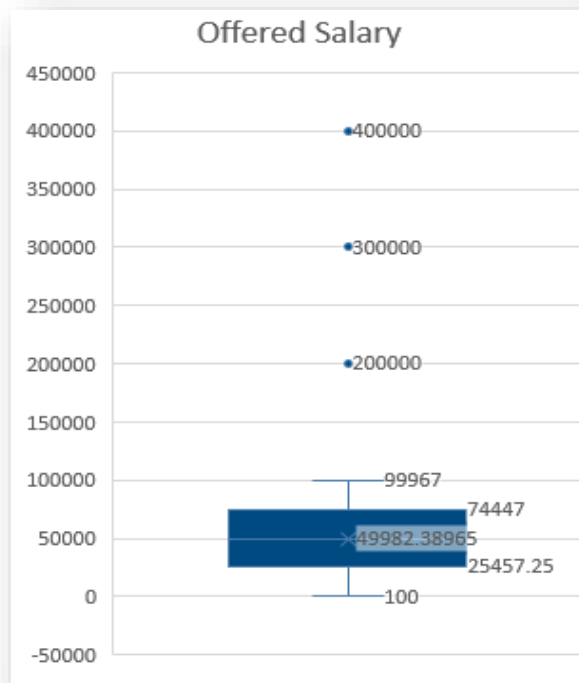
application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
289907	01-05-2014 07:44	Hired	Male	Sales Department	-	85914
955372	02-05-2014 09:16	Rejected	Female	Sales Department	b9	85176
882740	07-08-2014 13:42	Rejected	Male	Sales Department	c9	85837
230796	25-08-2014 09:32	Hired	Female	Sales Department	c9	85569
101190	09-05-2014 17:23	Hired	Male	Sales Department	c9	85057
804019	08-07-2014 16:52	Hired	Female	Sales Department	c9	85130
226229	10-07-2014 19:03	Hired	Male	Sales Department	c5	85610
788752	18-07-2014 17:37	Rejected	Male	Sales Department	c5	85392
294317	29-07-2014 17:33	Rejected	Female	Sales Department	b9	85400

application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
289907	05-01-2014 07:44	Hired	Male	Sales Department	c9	85914
959124	05-06-2014 16:27	Rejected	Male	Sales Department	i7	69904

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

application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
361096	15-05-2014 09:56	Rejected	Male	Service Department	c-10	9390
691216	15-05-2014 09:56	Rejected	Male	Service Department	c-10	67066
567661	15-05-2014 09:57	Rejected	Male	Service Department	c-10	8723
382645	15-05-2014 09:57	Hired	Male	Service Department	c-10	65587
767003	15-05-2014 10:01	Hired	Male	Service Department	c-10	73396
412827	15-05-2014 15:57	Rejected	Male	Service Department	c-10	76789
105729	15-05-2014 16:13	Rejected	Male	Service Department	c-10	80817

- e. 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 can replace them with median value of **Offered Salary** for corresponding **Department** and **Post Name**.



- **Offered Salary** value of **200000**.

	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	200000

171	=AGGREGATE(12,1,G14:G7148)						
	A	B	C	D	E	F	G
	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
	418179	22-08-2014 18:06	Rejected	Female	Service Department	b9	87078
	887973	22-08-2014 00:56	Hired	Female	Service Department	b9	38985
							44819

- Offered Salary value of 300000.

application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Sala
874368	21-07-2014 15:39	Hired	Male	General Management	i7	300000

170 X ✓ fx =AGGREGATE(12,1,G592:G6028)

A	B	C	D	E	F	G
application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Sala
375556	01-05-2014 09:46	Hired	Female	General Management	i7	81434
719895	01-05-2014 09:47	Hired	Female	General Management	i7	86399
464899	14-05-2014 11:57	Hired	Female	General Management	i7	1022
						56820

- Offered Salary is 400000. For Post Name labeled as i4 it has only one row so we included all rows with Post Name value starting with 'i'. In our Excel sheet we rounded up the value to 56295.

application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Sala
795330	15-06-2014 09:45	Hired	Female	General Management	i4	400000

170 X ✓ fx =AGGREGATE(12,1,G98:G6825)

A	B	C	D	E	F	G
application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Sala
103305	18-07-2014 07:25	Hired	Male	General Management	i5	95960
573332	19-07-2014 14:00	Hired	Male	General Management	i5	46109
874368	21-07-2014 15:39	Hired	Male	General Management	i7	56820
						56294.5

- Column application_id has duplicate values which should either be removed or replaced with proper value. Here we are just highlighting the cells having duplicate values.

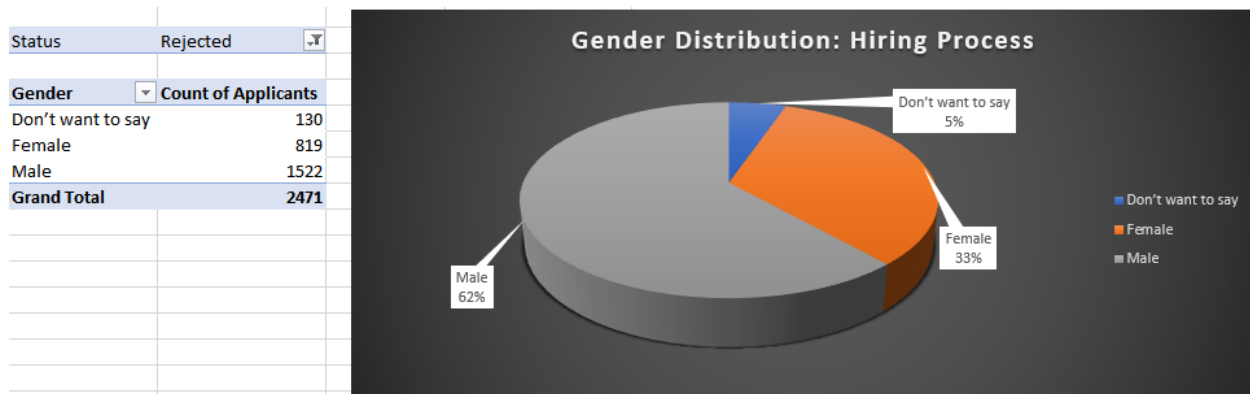
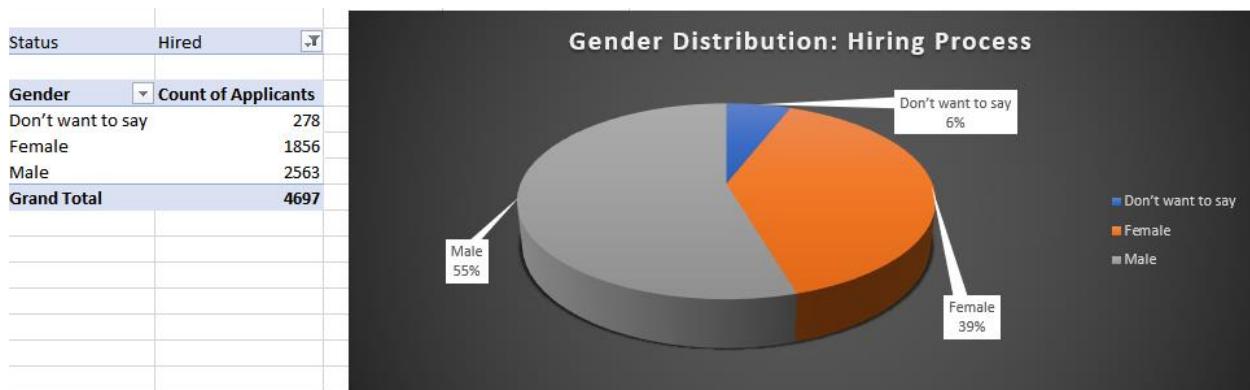
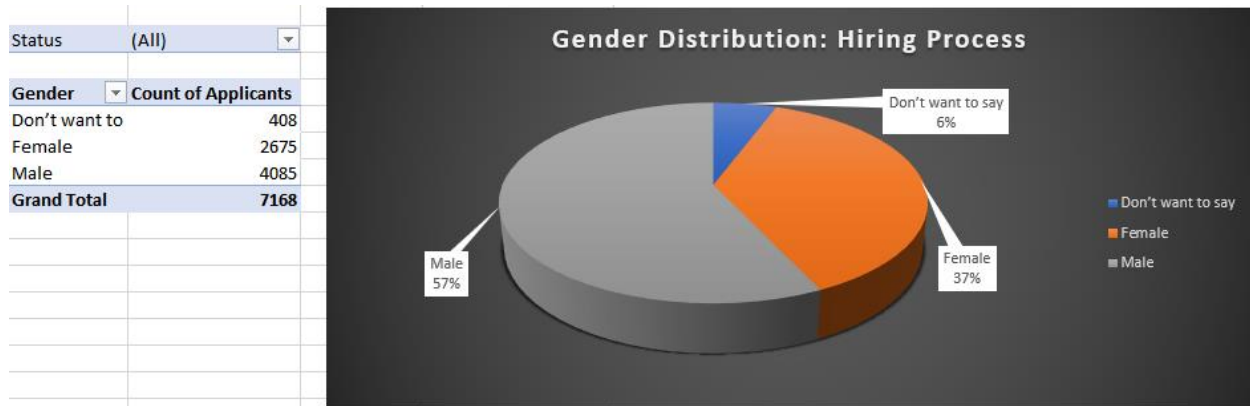
application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
861705	07-07-2014 09:35	Rejected	Male	Sales Department	i7	11184
779678	5/29/14 19:57	Hired	Male	Sales Department	c9	56913
961990	06-11-2014 10:11	Rejected	Male	Sales Department	c9	36776
229281	08-08-2014 17:46	Hired	Male	Sales Department	i7	10461

Final Dataset: After Data Processing, here is how our [dataset](#) looks:

application_id	Interview Taken on	Status	event_name	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
253651	02-05-2014 16:32	Hired	Male	Operations Department	i4	29668
289907	01-05-2014 07:44	Hired	Male	Sales Department	c9	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
751029	02-05-2014 13:09	Hired	Female	Service Department	i4	15156
434547	02-05-2014 13:11	Rejected	Female	Service Department	i4	49515
518854	01-05-2014 09:00	Rejected	Male	Service Department	n10	26990
649039	07-05-2014 10:48	Hired	Female	Service Department	b9	44819
199526	07-05-2014 10:50	Hired	Male	Service Department	b9	86787
539803	15-05-2014 09:31	Hired	Male	Finance Department	b9	2308
191009	09-05-2014 12:48	Hired	Female	Service Department	i7	56688
195323	09-05-2014 12:48	Hired	Don't want to say	Service Department	i7	81757
51318	02-05-2014 08:07	Hired	Male	Service Department	i5	15134
742283	02-05-2014 08:11	Rejected	Don't want to say	Service Department	i5	100
513166	01-05-2014 22:53	Hired	Female	Operations Department	i1	73579
791372	01-05-2014 22:54	Rejected	Male	Operations Department	i1	50351

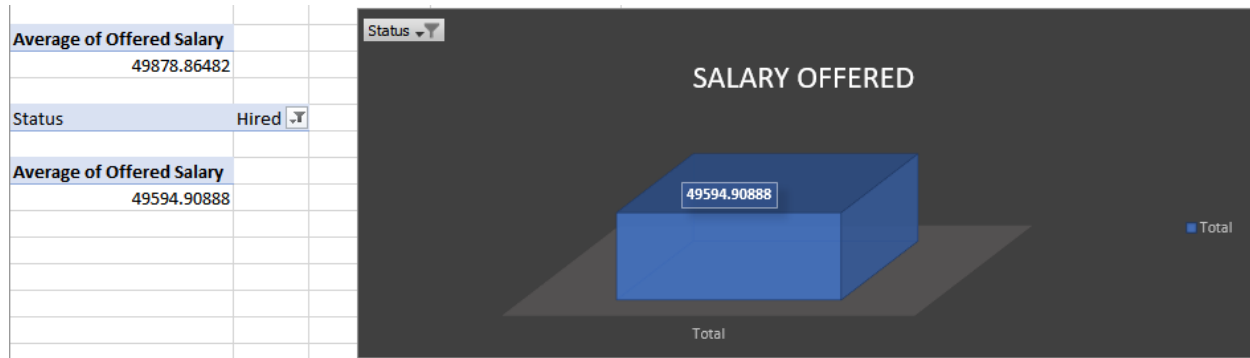
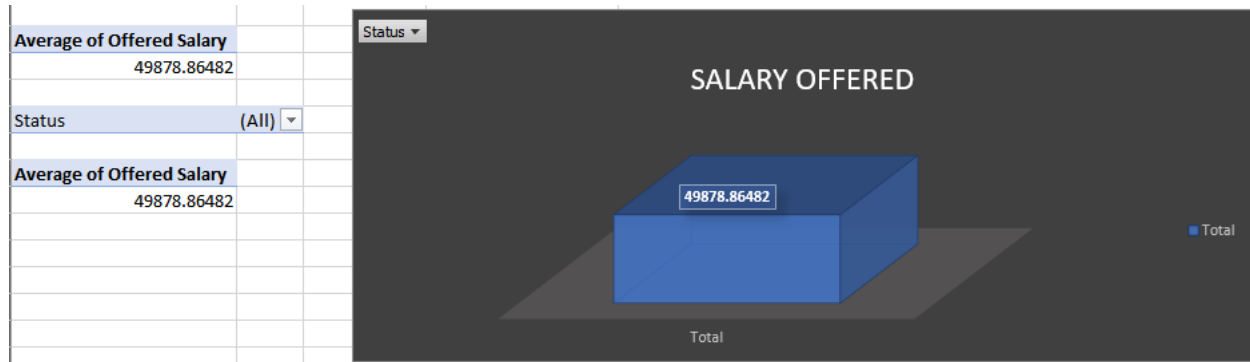
A. Hiring Analysis: The hiring process involves bringing new individuals into the organization for various roles.

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



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.

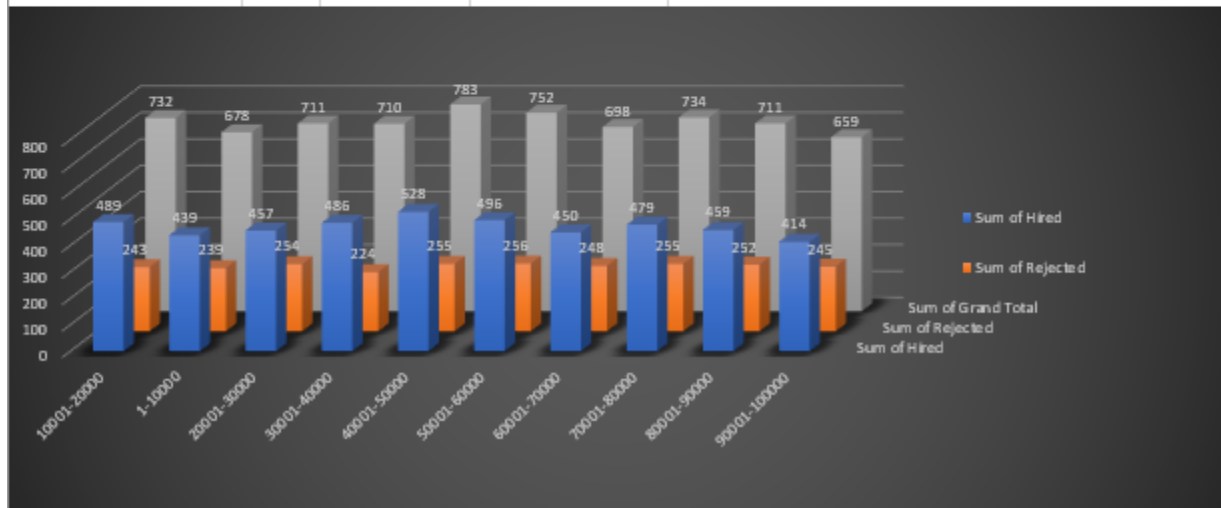
Task: What is the average salary offered by this company? Use Excel functions to calculate this.



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.

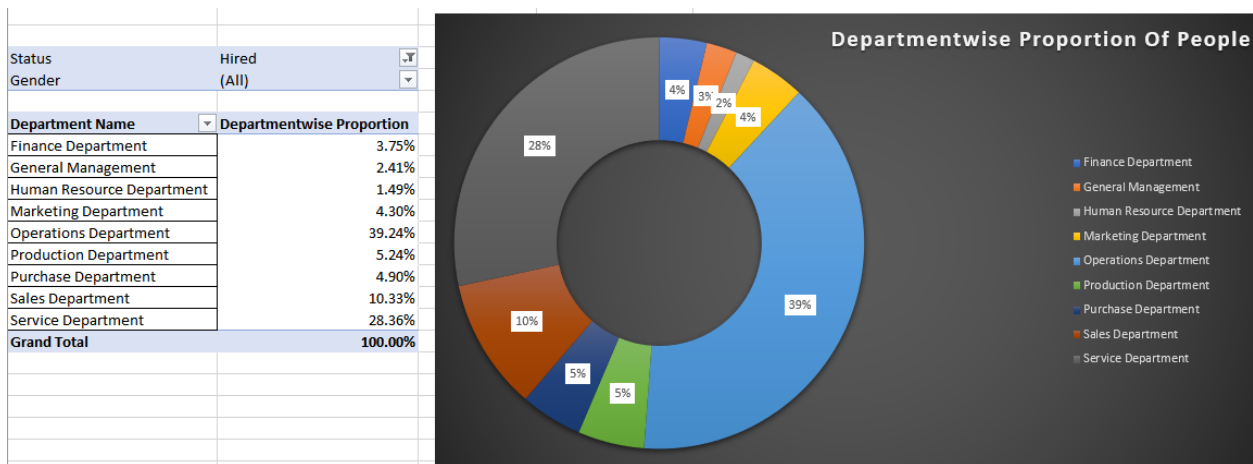
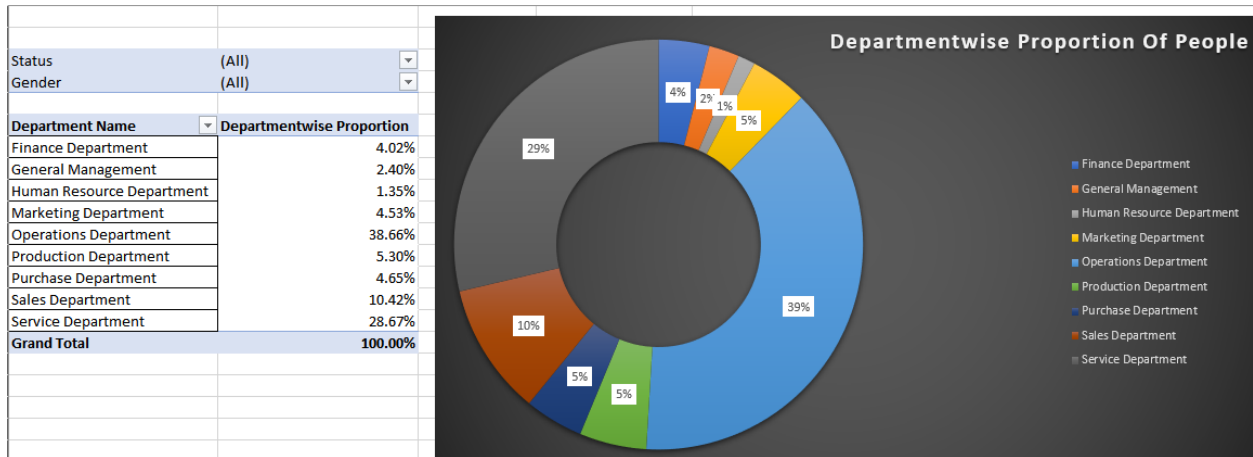
Task: Create class intervals for the salaries in the company. This will help you understand the salary distribution.

Salary Class	Sum	Sum of Rejected	Sum of Grand
10001-20000	489	243	732
1-10000	439	239	678
20001-30000	457	254	711
30001-40000	486	224	710
40001-50000	528	255	783
50001-60000	496	256	752
60001-70000	450	248	698
70001-80000	479	255	734
80001-90000	459	252	711
90001-100000	414	245	659
Grand Total	4697	2471	7168



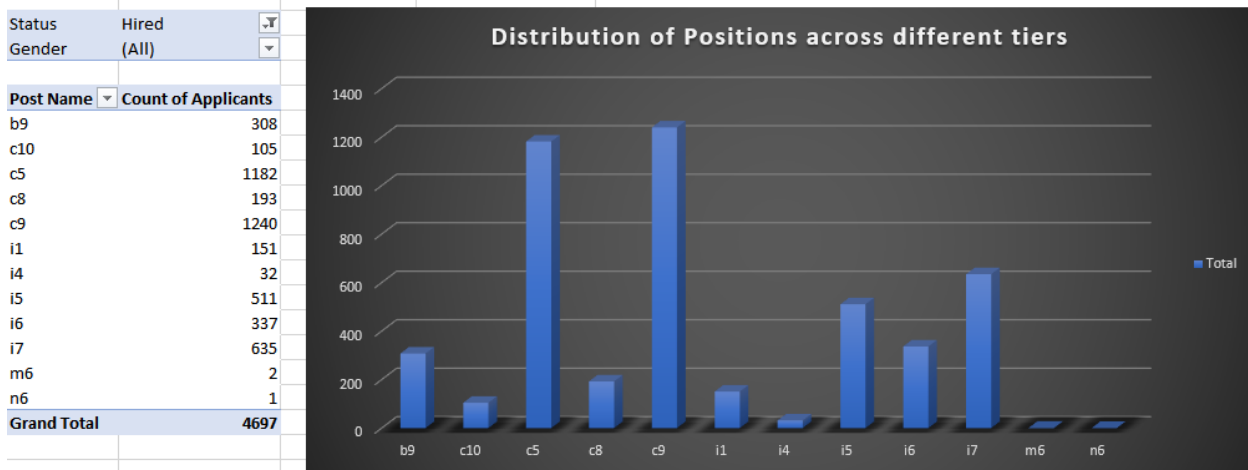
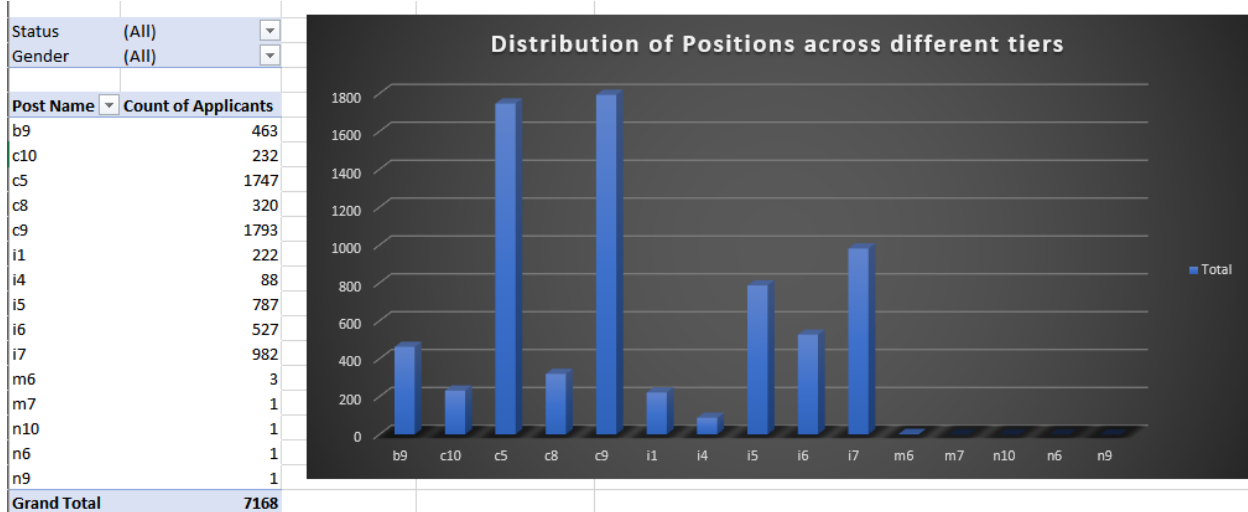
D. Departmental Analysis: Visualizing data through charts and plots is a crucial part of data analysis.

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



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.



Insights:

- The insights derived from the analysis include gender distribution of hires, average salary offered, and salary distribution across different ranges.
- The gender distribution analysis revealed a balanced hiring pattern, which could indicate a diverse and inclusive hiring approach.
- Calculating the average salary offered provided insights into the company's compensation structure and competitiveness in the job market.
- Creating salary distribution class intervals helped visualize the distribution of salaries, identifying any concentration or disparities in salary ranges.

Result:

- Through this project, we gained a deeper understanding of the hiring process analytics and its significance in organizational decision-making.
- The insights provided valuable information for the hiring department to optimize their recruitment strategies and make informed hiring decisions.
- The detailed report generated will serve as a guide for the hiring department in implementing improvements to the hiring process based on data-driven insights.