Hiring Process Analytics

Submitted by:

Nadendla Dharani

Hyper Link to the Excel Sheet:

https://docs.google.com/spreadsheets/d/1t9RjgED_vIEho4PsYd7vKAPrGbcQrIlA/edit?usp=sharing&ouid=111691645789497796027&rtpof=true&sd=true

Project Description

The project involves analyzing the hiring process data of a multinational company, aiming to derive meaningful insights that can contribute to improving the overall hiring process. The dataset includes records of previous hires, and the analysis focuses on aspects such as gender distribution, average salary, salary distribution, departmental composition, and position tier distribution. The ultimate goal is to provide actionable insights for optimizing the hiring procedures and decision-making within the company.

Approach

The analysis of the hiring process data at a multinational company begins with a meticulous understanding of the dataset and a comprehensive data cleaning process, addressing missing values, and consolidating categories. Outliers are identified and handled strategically. Gender distribution is determined using Excel functions and visualizations, providing insights into diversity. Average salary calculations offer a snapshot of compensation practices. Salary distribution is visualized through histograms. Departmental and position tier analyses guide workforce planning. The process is summarized with additional statistical measures, presented visually, and documented for transparency. Iterative analysis ensures refinement and actionable insights, contributing to the optimization of the hiring process.

Tech-Stack Used

The project leverages Microsoft Excel as the primary tool for data analysis and visualization. Excel's versatile functions and features facilitate data cleaning,

computation of summary statistics, and the creation of various visualizations, including pie charts, bar graphs, and histograms. The familiar interface of Excel allows for efficient handling of tasks such as handling missing values, clubbing categories, and outlier detection. Additionally, standard statistical functions within Excel contribute to deriving meaningful insights, making it a powerful and accessible tech-stack for this data analytics project.

Insights

A. Hiring Analysis:

- Provided a clear breakdown of gender distribution in the hires.
- Identified the number of males and females brought into the organization.

B. Salary Analysis:

- Calculated and presented the average salary offered by the company.

C. Salary Distribution:

- Established class intervals for salary ranges, aiding in understanding the overall salary distribution within the company.

D. Departmental Analysis:

- Utilized visualizations like pie charts or bar graphs to showcase the proportion of employees in different departments.
- Identified trends or imbalances in departmental composition.

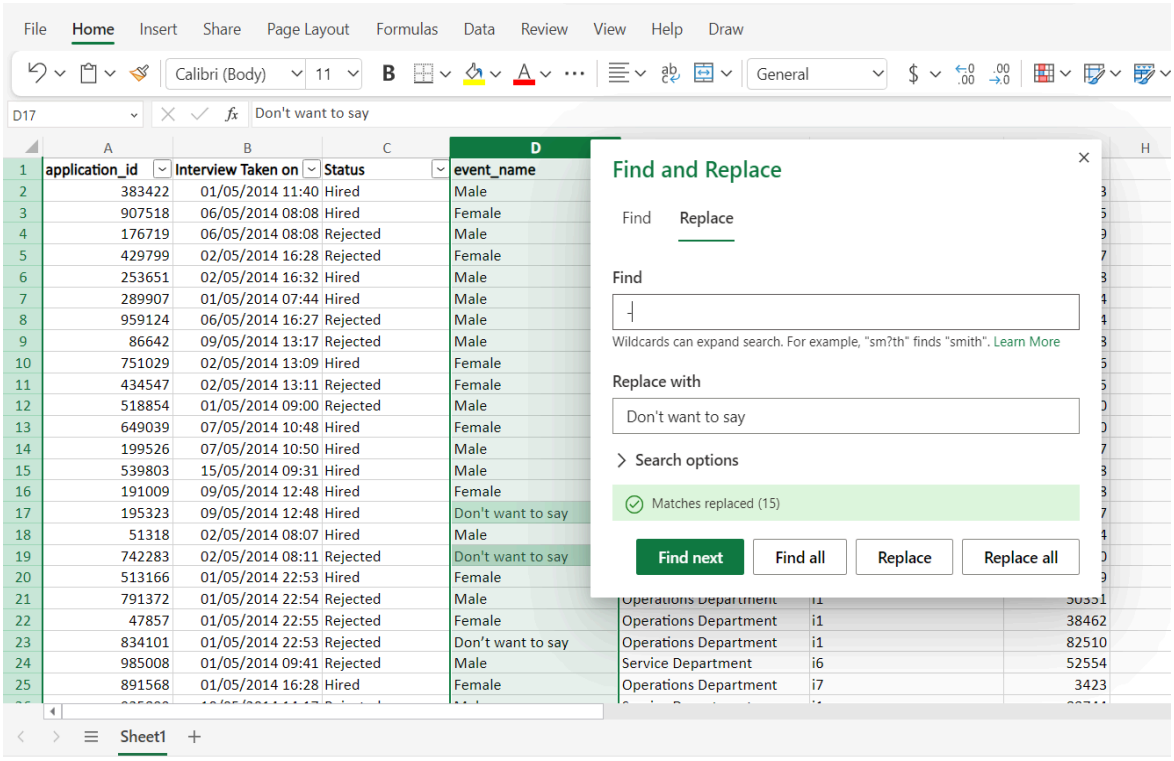
E. Position Tier Analysis:

- Represented the distribution of positions across different tiers using charts or graphs.
- Provided insights into the hierarchical structure of the organization.

Results

1. **Handling Missing Data:** Check if there are any missing values in the dataset. If there are, decide on the best strategy to handle them.

- Event_name column has 3 values, “Male, Female and Don’t want to say”. But it also has 15 rows with “-” values, these are null values. I found and replaced those null values with “Don’t want to say”.



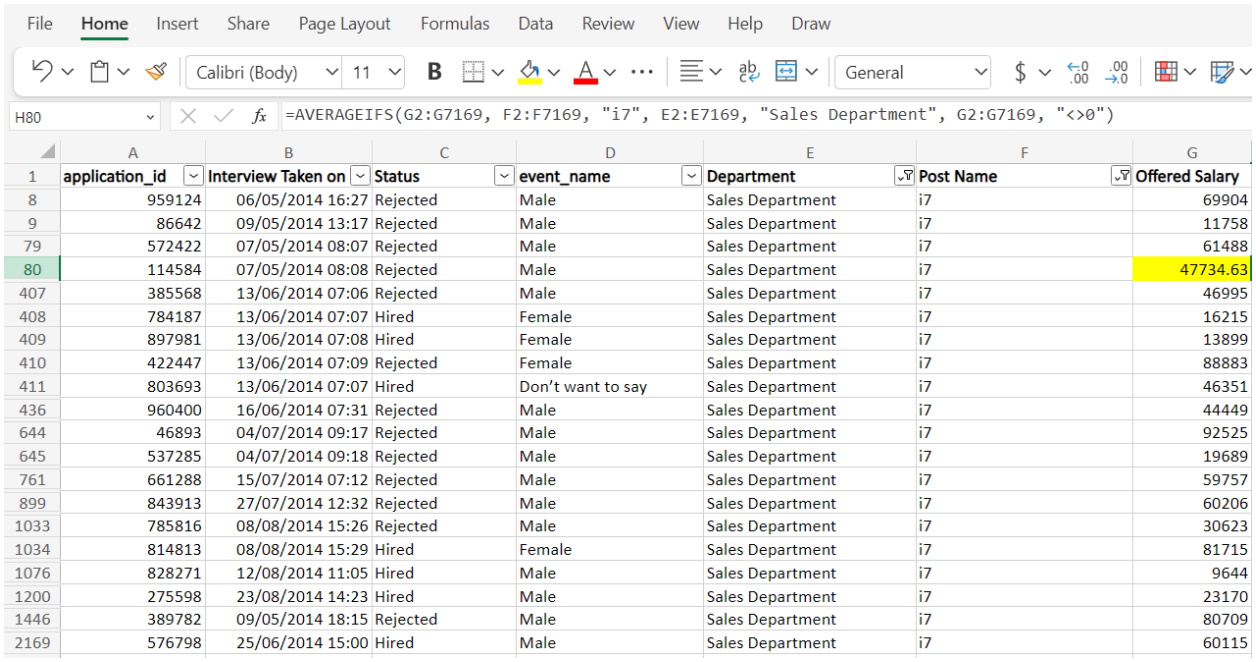
	A	B	C	D	E	F	G	H
1	application_id	Interview Taken on	Status	event_name				
2	383422	01/05/2014 11:40	Hired	Male				
3	907518	06/05/2014 08:08	Hired	Female				
4	176719	06/05/2014 08:08	Rejected	Male				
5	429799	02/05/2014 16:28	Rejected	Female				
6	253651	02/05/2014 16:32	Hired	Male				
7	289907	01/05/2014 07:44	Hired	Male				
8	959124	06/05/2014 16:27	Rejected	Male				
9	86642	09/05/2014 13:17	Rejected	Male				
10	751029	02/05/2014 13:09	Hired	Female				
11	434547	02/05/2014 13:11	Rejected	Female				
12	518854	01/05/2014 09:00	Rejected	Male				
13	649039	07/05/2014 10:48	Hired	Female				
14	199526	07/05/2014 10:50	Hired	Male				
15	539803	15/05/2014 09:31	Hired	Male				
16	191009	09/05/2014 12:48	Hired	Female				
17	195323	09/05/2014 12:48	Hired	Don't want to say				
18	51318	02/05/2014 08:07	Hired	Male				
19	742283	02/05/2014 08:11	Rejected	Don't want to say				
20	513166	01/05/2014 22:53	Hired	Female				
21	791372	01/05/2014 22:54	Rejected	Male	Operations Department	i1	50351	
22	47857	01/05/2014 22:55	Rejected	Female	Operations Department	i1	38462	
23	834101	01/05/2014 22:53	Rejected	Don’t want to say	Operations Department	i1	82510	
24	985008	01/05/2014 09:41	Rejected	Male	Service Department	i6	52554	
25	891568	01/05/2014 16:28	Hired	Female	Operations Department	i7	3423	

- The Offered Salary column has 1 empty row, G80. The corresponding values in the Department and Post Name column are “Sales Department” and “i7”. So, I replaced the empty cell with average of Offered Salary values having Sales Department and i7 by using the formula

```
=AVERAGEIFS(G2:G7169, F2:F7169, "i7", E2:E7169, "Sales Department", G2:G7169, "<>0").
```

I got an average of 47735(rounded off).

	A	B	C	D	E	F	G	H
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary	
80	114584	07/05/2014 08:08	Rejected	Male	Sales Department	i7		
7170								

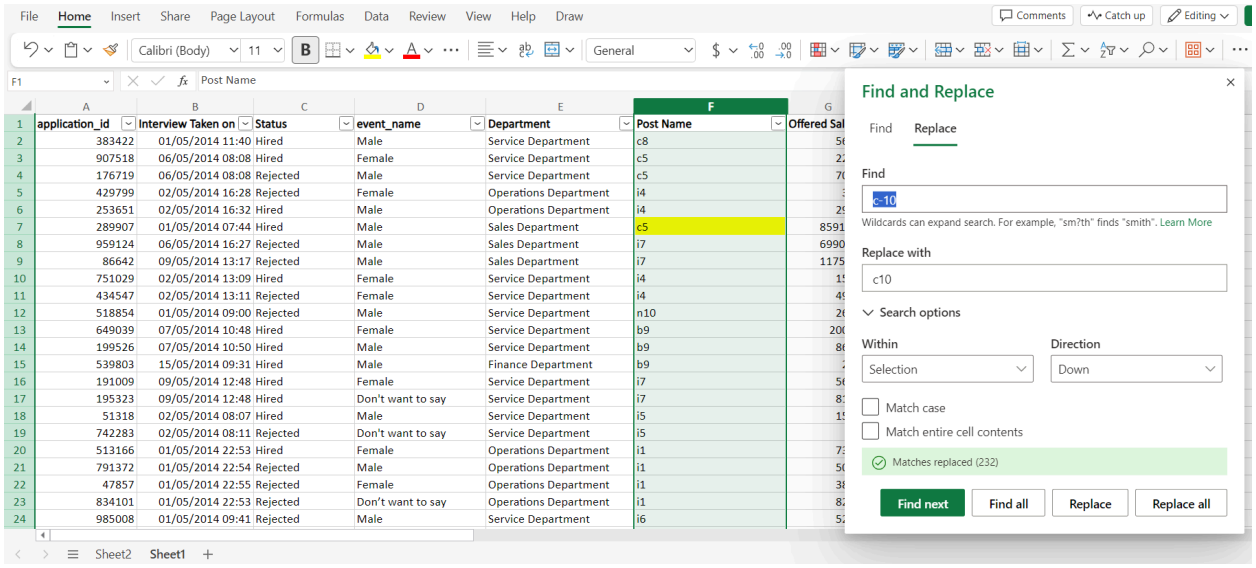


	A	B	C	D	E	F	G
1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
8	959124	06/05/2014 16:27	Rejected	Male	Sales Department	i7	69904
9	86642	09/05/2014 13:17	Rejected	Male	Sales Department	i7	11758
79	572422	07/05/2014 08:07	Rejected	Male	Sales Department	i7	61488
80	114584	07/05/2014 08:08	Rejected	Male	Sales Department	i7	47734.63
407	385568	13/06/2014 07:06	Rejected	Male	Sales Department	i7	46995
408	784187	13/06/2014 07:07	Hired	Female	Sales Department	i7	16215
409	897981	13/06/2014 07:08	Hired	Female	Sales Department	i7	13899
410	422447	13/06/2014 07:09	Rejected	Female	Sales Department	i7	88883
411	803693	13/06/2014 07:07	Hired	Don't want to say	Sales Department	i7	46351
436	960400	16/06/2014 07:31	Rejected	Male	Sales Department	i7	44449
644	46893	04/07/2014 09:17	Rejected	Male	Sales Department	i7	92525
645	537285	04/07/2014 09:18	Rejected	Male	Sales Department	i7	19689
761	661288	15/07/2014 07:12	Rejected	Male	Sales Department	i7	59757
899	843913	27/07/2014 12:32	Rejected	Male	Sales Department	i7	60206
1033	785816	08/08/2014 15:26	Rejected	Male	Sales Department	i7	30623
1034	814813	08/08/2014 15:29	Hired	Female	Sales Department	i7	81715
1076	828271	12/08/2014 11:05	Hired	Male	Sales Department	i7	9644
1200	275598	23/08/2014 14:23	Hired	Male	Sales Department	i7	23170
1446	389782	09/05/2014 18:15	Rejected	Male	Sales Department	i7	80709
2169	576798	25/06/2014 15:00	Hired	Male	Sales Department	i7	60115

- Column Post Name has 1 null value, and its corresponding department is "Sales Department" and "Offered Salary" is 85914. So I replaced it with a majority count of Posts(which is 30) for candidates in the Sales Department and whose Offered Salary is between 85,000 and 100,000, which is "c5".

1	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salar
7	289907	01/05/2014 07:44	Hired	Male	Sales Department	c5	85914.00
31	765579	01/05/2014 09:26	Rejected	Male	Sales Department	i4	87884.00
39	955372	02/05/2014 09:16	Rejected	Female	Sales Department	b9	85176.00
72	188820	07/05/2014 17:58	Rejected	Male	Sales Department	i4	94869.00
88	37906	08/05/2014 03:16	Hired	Female	Sales Department	i4	95603.00
209	150277	22/05/2014 22:18	Rejected	Male	Sales Department	c-10	91928.00
362	894420	08/06/2014 17:29	Rejected	Male	Sales Department	c9	86412.00
410	422447	13/06/2014 07:09	Rejected	Female	Sales Department	i7	88883.00
644	46893	04/07/2014 09:17	Rejected	Male	Sales Department	i7	92525.00
646	779410	04/07/2014 03:57	Rejected	Male	Sales Department	c5	94544.00
876	467947	25/07/2014 04:48	Rejected	Male	Sales Department	c5	92438.00
878	237486	02/08/2014 13:48	Hired	Male	Sales Department	c5	87032.00
989	415817	05/08/2014 10:31	Hired	Male	Sales Department	c9	88298.00
1048	323081	11/08/2014 17:31	Rejected	Female	Sales Department	i6	90519.00
1051	351740	11/08/2014 17:32	Rejected	Female	Sales Department	i6	95852.00
1059	723807	11/08/2014 14:33	Hired	Female	Sales Department	b9	92366.00
1138	610243	19/08/2014 16:36	Rejected	Male	Sales Department	i5	88996.00
1161	367527	21/08/2014 18:28	Hired	Male	Sales Department	i5	86480.00
1177	553557	22/08/2014 11:58	Rejected	Male	Sales Department	i5	92785.00
1282	288469	30/06/2014 17:00	Hired	Female	Sales Department	c9	91349.00
1290	741579	19/08/2014 10:56	Hired	Male	Sales Department	c5	98625.00
1323	882740	07/08/2014 13:42	Rejected	Male	Sales Department	c9	85837.00
1457	230796	25/08/2014 09:32	Hired	Female	Sales Department	c9	85569.00

- Also, the Post Name column has values, "c-10" which are replaced by "c10".



2. **Outlier Detection and Removing:** Check for outliers in the dataset that may skew your analysis.

From the scatter plot of the Offered Salary Column, we can say there are 3 outliers values, 200000, 300000 and 400000. Now, I have replaced them with the mean of Offered Salary for corresponding department and Post Name.

For 200000:

Mean is rounded off to 48774.

	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salar
2	383422	01/05/2014 11:40	Hired	Male	Service Department	c8	56553
3	907518	06/05/2014 08:08	Hired	Female	Service Department	c5	22075
4	176719	06/05/2014 08:08	Rejected	Male	Service Department	c5	70069
5	429799	02/05/2014 16:28	Rejected	Female	Operations Department	i4	3207
6	253651	02/05/2014 16:32	Hired	Male	Operations Department	i4	29668
7	289907	01/05/2014 07:44	Hired	Male	Sales Department	c5	85914.00
8	959124	06/05/2014 16:27	Rejected	Male	Sales Department	i7	69904.00
9	86642	09/05/2014 13:17	Rejected	Male	Sales Department	i7	11758.00
10	751029	02/05/2014 13:09	Hired	Female	Service Department	i4	15156
11	434547	02/05/2014 13:11	Rejected	Female	Service Department	i4	49515
12	518854	01/05/2014 09:00	Rejected	Male	Service Department	n10	26990
13	649039	07/05/2014 10:48	Hired	Female	Service Department	b9	48774
14	199526	07/05/2014 10:50	Hired	Male	Service Department	b9	86787

For 300000:

Mean is rounded off to 56162, using the formula

```
=AVERAGEIFS(G2:G7169, F2:F7169, "i7", E2:E7169, "General Management")
```

6821	359954	18/07/2014 07:22	Hired	Female	General Management	i5	3947
6822	18808	18/07/2014 07:22	Hired	Male	General Management	i5	46160
6823	103305	18/07/2014 07:25	Hired	Male	General Management	i5	95960
6824	573332	19/07/2014 14:00	Hired	Male	General Management	i5	46109
6825	874368	21/07/2014 15:39	Hired	Male	General Management	i7	56162
6826	709159	09/08/2014 02:50	Hired	Female	Marketing Department	i7	98589
6827	952623	13/08/2014 10:56	Hired	Male	Marketing Department	i7	72477
6828	851730	15/08/2014 12:37	Hired	Male	Marketing Department	i7	14815
6829	412615	15/08/2014 12:39	Hired	Female	Marketing Department	i7	19732
6830	380221	15/08/2014 12:41	Hired	Female	Marketing Department	i7	76350
6831	386930	18/08/2014 17:59	Hired	Female	Marketing Department	i7	28912
6832	112091	19/08/2014 16:15	Rejected	Male	Marketing Department	i7	81226
6833	718973	27/08/2014 15:35	Hired	Male	Operations Department	i7	75131
6834	396732	27/08/2014 15:37	Hired	Female	Operations Department	i7	95417

For 400000:

For the Offered Salary value of 400000, Post Name i4 has only one row, so I considered all departments with Post Name value “i4”.
Mean is rounded off to 48878, by formula

```
=AVERAGEIFS(G2:G7169, F2:F7169, "i4")
```

258	846310	28/05/2014 08:20	Hired	Female	Operations Department	i4	74691
260	773059	28/05/2014 11:23	Rejected	Male	Operations Department	i4	25881
285	24310	29/05/2014 12:24	Rejected	Male	Service Department	i4	63559
286	795330	15/06/2014 09:45	Hired	Female	General Management	i4	48878
292	462254	30/05/2014 09:14	Hired	Female	Operations Department	i4	7210
293	413388	30/05/2014 09:14	Rejected	Don’t want to say	Operations Department	i4	77422
295	517354	31/05/2014 13:07	Rejected	Male	Purchase Department	i4	2988
297	657117	02/06/2014 22:30	Rejected	Male	Operations Department	i4	56255
298	22963	02/06/2014 22:33	Rejected	Female	Operations Department	i4	53806
299	201095	02/06/2014 22:35	Rejected	Male	Operations Department	i4	7775

Data Analytics Tasks:

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

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

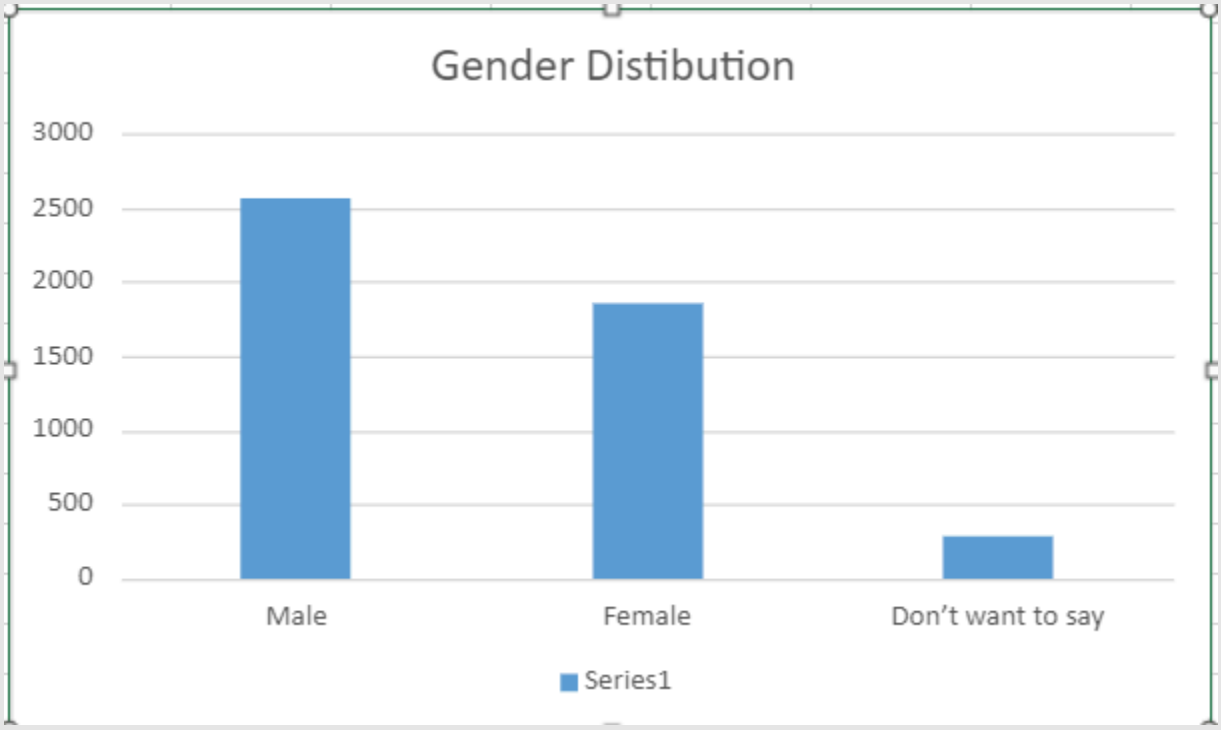
Result:

Males hired by the company: `=COUNTIFS(D2:D7169,"Male",C2:C7169,"Hired")`

Females hired by the company:
`=COUNTIFS(D2:D7169,"Female",C2:C7169,"Hired")`

“Don’t want to say” hired by the company: `=COUNTIFS(D2:D7169,"Don’t want to say", C2:C7169, "Hired")`

Male	2563
Female	1856
Don’t want to say	278



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.

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

Result:

Average salary offered by the company: `=AVERAGE(G2:G7169)`

Average salary is 49878.61579

Rounded off to 49879 using `=ROUND(G7170,0),` which round off to nearest integer.

7156		17/08/2014 12:17	Hired	Female	Production Department	c9	89565
7157		17/08/2014 07:48	Hired	Male	Service Department	c5	86162
7158		17/08/2014 07:49	Hired	Don't want to say	Service Department	c5	58900
7159		21/08/2014 07:30	Rejected	Male	Sales Department	c5	71449
7160		27/08/2014 17:36	Rejected	Male	Service Department	c5	67196
7161		22/08/2014 15:50	Hired	Male	Service Department	i7	16756
7162		27/08/2014 05:01	Hired	Female	Service Department	i7	30952
7163		28/08/2014 17:29	Hired	Male	Service Department	c9	64150
7164		28/08/2014 17:30	Hired	Male	Service Department	c9	40152
7165		28/08/2014 17:32	Hired	Male	Service Department	c9	49282
7166		31/08/2014 01:36	Hired	Female	Service Department	c5	57742
7167		31/08/2014 01:37	Hired	Male	Service Department	c5	69932
7168		31/08/2014 01:38	Rejected	Male	Service Department	c5	14489
7169		26/08/2014 12:14	Hired	Male	Operations Department	c5	54201
7170						Average Salary Offered	49878.61579
7171						Rounded off	49879
7172							

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.

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

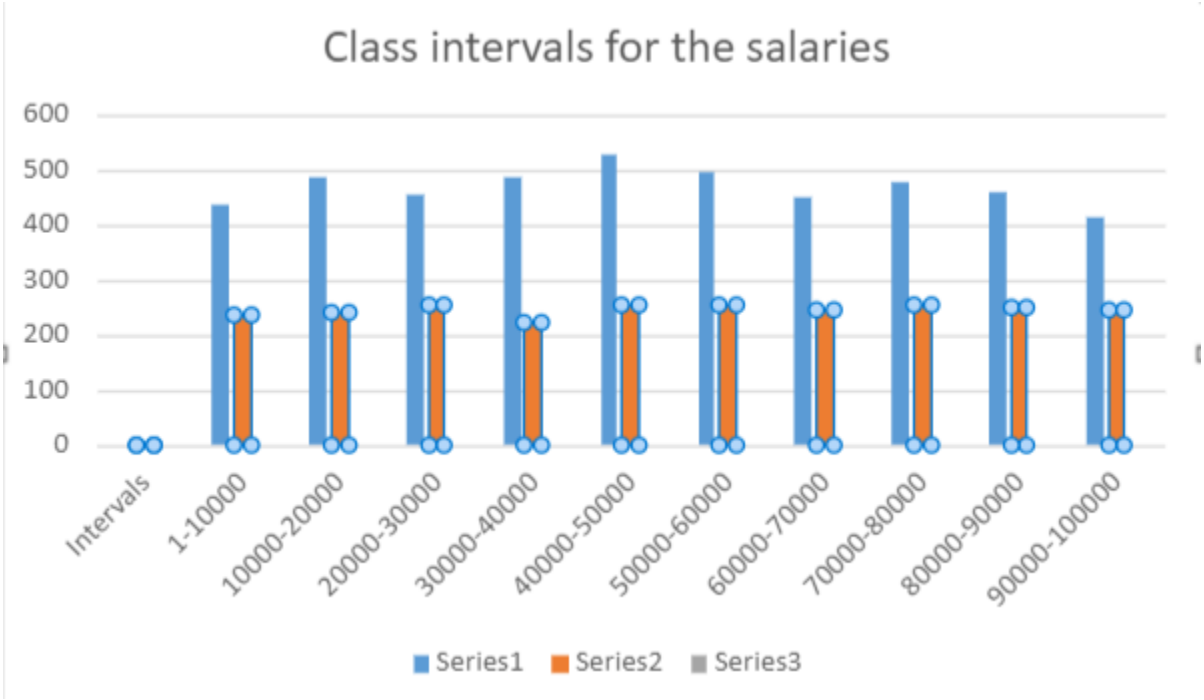
Result:

By using the below formulas:

```
=COUNTIFS(G2:G7169,">10000",G2:G7169,"<=20000",C2:C7169,"Hired")
```

```
=COUNTIFS(G2:G7169,">10000",G2:G7169,"<=20000",C2:C7169,"Rejected")
```

Intervals	Hired	Offered but rejected
1-10000	439	239
10000-20000	489	243
20000-30000	457	254
30000-40000	486	224
40000-50000	529	255
50000-60000	495	256
60000-70000	450	248
70000-80000	479	255
80000-90000	459	252
90000-100000	414	245
Total	**4697**	**2471**



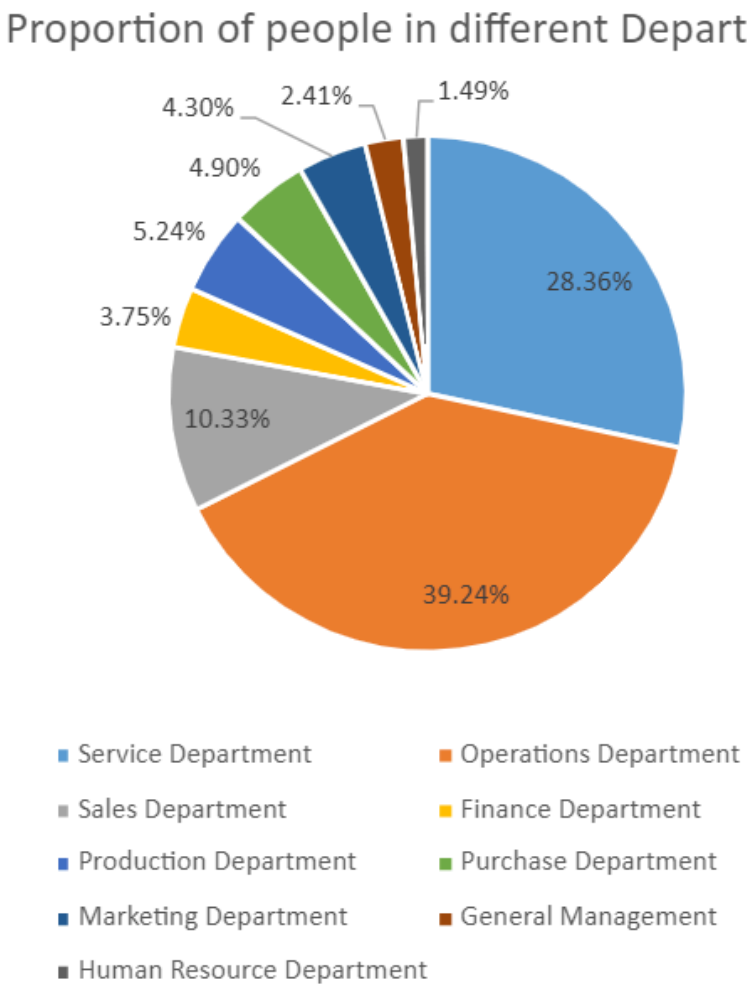
From the analysis, We can conclude that maximum people are hired with a salary range from 40000-50000. And the least number of people are hired with the salary range 90000-100000.

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

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

Result:.

Department	Count	Proportion
Service Department	1332	28.36%
Operations Department	1843	39.24%
Sales Department	485	10.33%
Finance Department	176	3.75%
Production Department	246	5.24%
Purchase Department	230	4.90%
Marketing Department	202	4.30%
General Management	113	2.41%
Human Resource Department	70	1.49%
Total	**4697**	**100.00%**



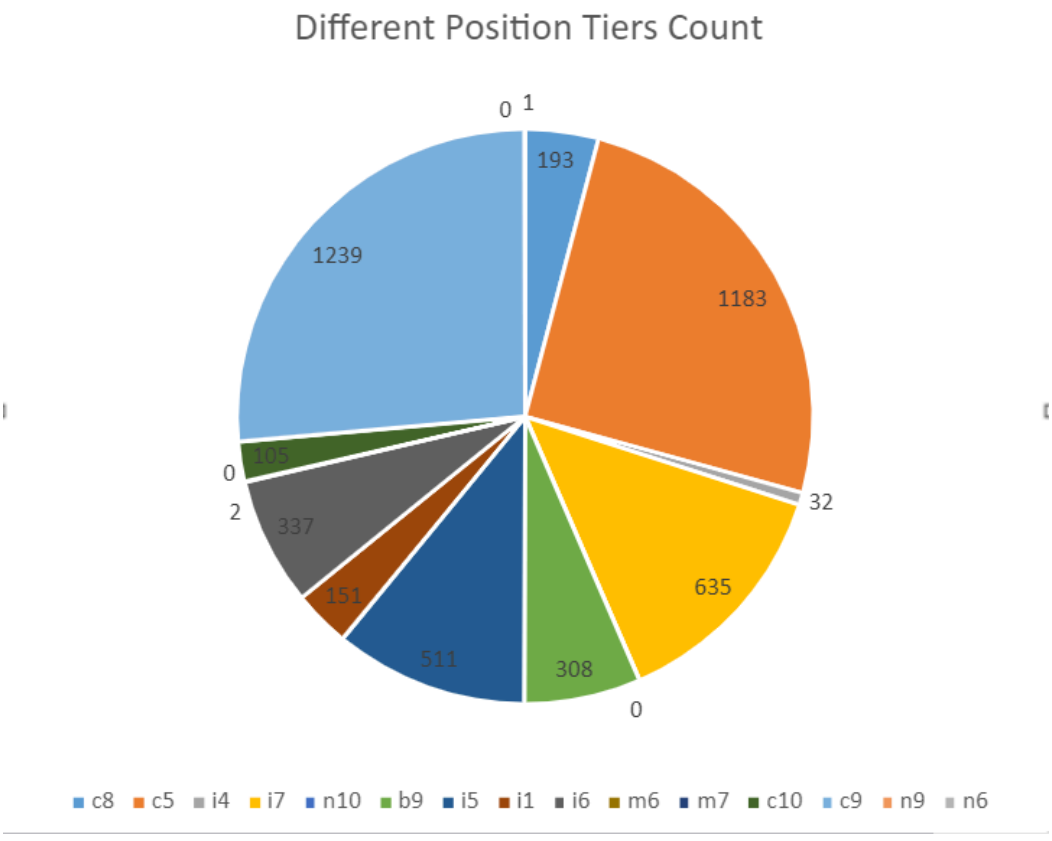
From the pie chart above, we can say that most people are hired in the Operations Department followed by the Service Department and the least number of people are hired in the Human Resource Department.

E. Position Tier Analysis: Different positions within a company often have different tiers or levels.

Your 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:

Post Name	Count
c8	193
c5	1183
i4	32
i7	635
n10	0
b9	308
i5	511
i1	151
i6	337
m6	2
m7	0
c10	105
c9	1239
n9	0
n6	1
Total	**4697**



From the pie chart above, we can say that most people are hired for c9 post tiers, followed by c5 and the least number of people are hired by n10,m7,n9 and n6.