

Project Report:- Hiring Process Analytics

Project Description: - In this project, I have worked on a dataset of a company where the details about people who registered for a particular post in a department of the company were provided. The main objective was to draw insights from the data and provide the necessary information for the company to analyze before hiring new employees. The project involved tasks such as understanding the data columns, checking for missing data, clubbing columns with multiple categories, checking for outliers, removing outliers, drawing data summaries, and using Excel to answer the given questions.

Approach:-

I started by exploring the dataset and understanding the different columns of data. I then checked for missing data and outliers in the data. After that, I used Excel to answer the given questions, which involved tasks such as calculating the number of males and females hired, finding the average salary offered in the company, drawing class intervals for salary, and representing the data using different charts and graphs.

Tech-Stack Used:

I used Microsoft Excel to work on this project. Microsoft Excel is a powerful tool for data analysis and is widely used in the industry for various data-related tasks.

Insights:

Through this project, I gained insights into various aspects of data analysis, such as data cleaning, exploratory data analysis, and visualization. I also learned how to use different Excel formulas and functions to perform statistical analysis on the data. The project helped me understand the importance of data analysis in decision-making and provided me with hands-on experience in working with real-world datasets.

Result:

Through this project, I was able to offer the organisation useful information regarding the hiring process, including the proportion of employees working in various departments, the average wage offered, and the number of men and women employed. The business can utilise these data to optimise the hiring process and make well-informed decisions when employing new staff. The project also gave me practical experience dealing with real-world data and helped me improve my data analysis skills.

1)Hiring: Process of intaking of people into an organization for different kinds of positions.

Your task: How many males and females are Hired ?

Task-1	
HIRED MALE AND FEMALE	
MALE	FEMALE
2563	1856

=COUNTIFS(Sheet1!D2:D7169,"Male",Sheet1!C2:C7169,"Hired")

=COUNTIFS(Sheet1!D2:D7169,"Male",Sheet1!C2:C7169,"Hired")

2) Average Salary: Adding all the salaries for a select group of employees and then dividing the sum by the number of employees in the group.

Your task: What is the average salary offered in this company ?

Task-2	
Average salary offered in this company	
Mean	49983.02902

=AVERAGE(Sheet1!G2:G7169)

3) Class Intervals: The class interval is the difference between the upper class limit and the lower class limit.

Your task: Draw the class intervals for salary in the company ?

Column:-J5 To J14 Define Name:-Interval(Intervals)

Column:-G2To G7169 Define Name:-Salary(Salary Offered)

=FREQUENCY(Salary,Interval)

Task-3

Class Intervals for salary		
Intervals		Count of Offer Salary
100	10000	678
10000	20000	732
20000	30000	711
30000	40000	710
40000	50000	781
50000	60000	750
60000	70000	698
70000	80000	734
80000	90000	711
100000	200000	660
Sum of total offersalary		7165

4)Charts and Plots: This is one of the most important part of analysis to visualize the data.

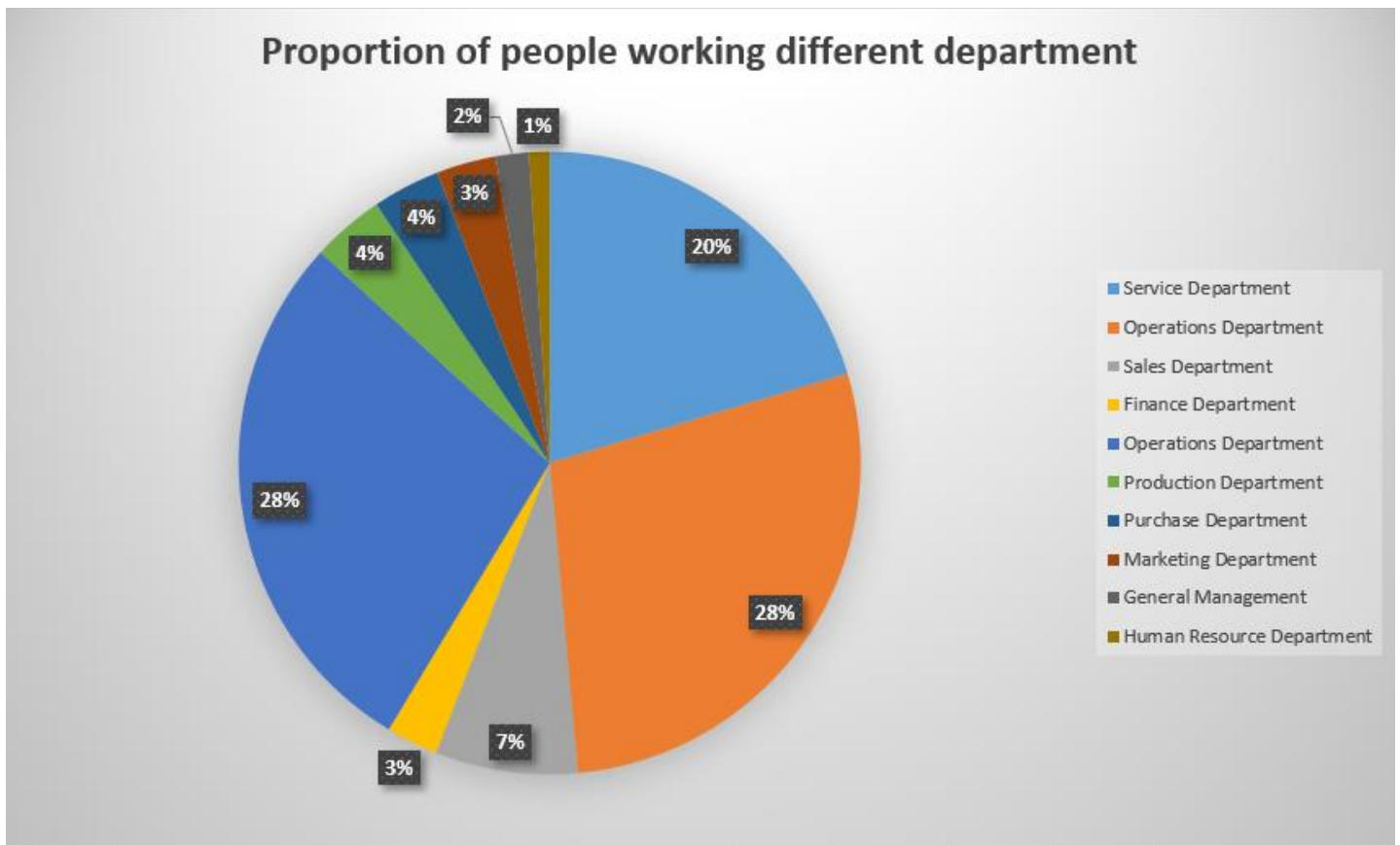
Your task: Draw Pie Chart / Bar Graph (or any other graph) to show proportion of people working different department ?

Column:-Department Define Name:-Department

Column:-Status Define Name:-Status

=COUNTIFS(Statistic.xlsx!Status,"Hired",Department,"Sales Department")

Department	Hired People
Service Department	1332
Operations Department	1843
Sales Department	485
Finance Department	176
Operations Department	1843
Production Department	246
Purchase Department	230
Marketing Department	202
General Management	113
Human Resource Department	70



5)Charts: Use different charts and graphs to perform the task representing the data.

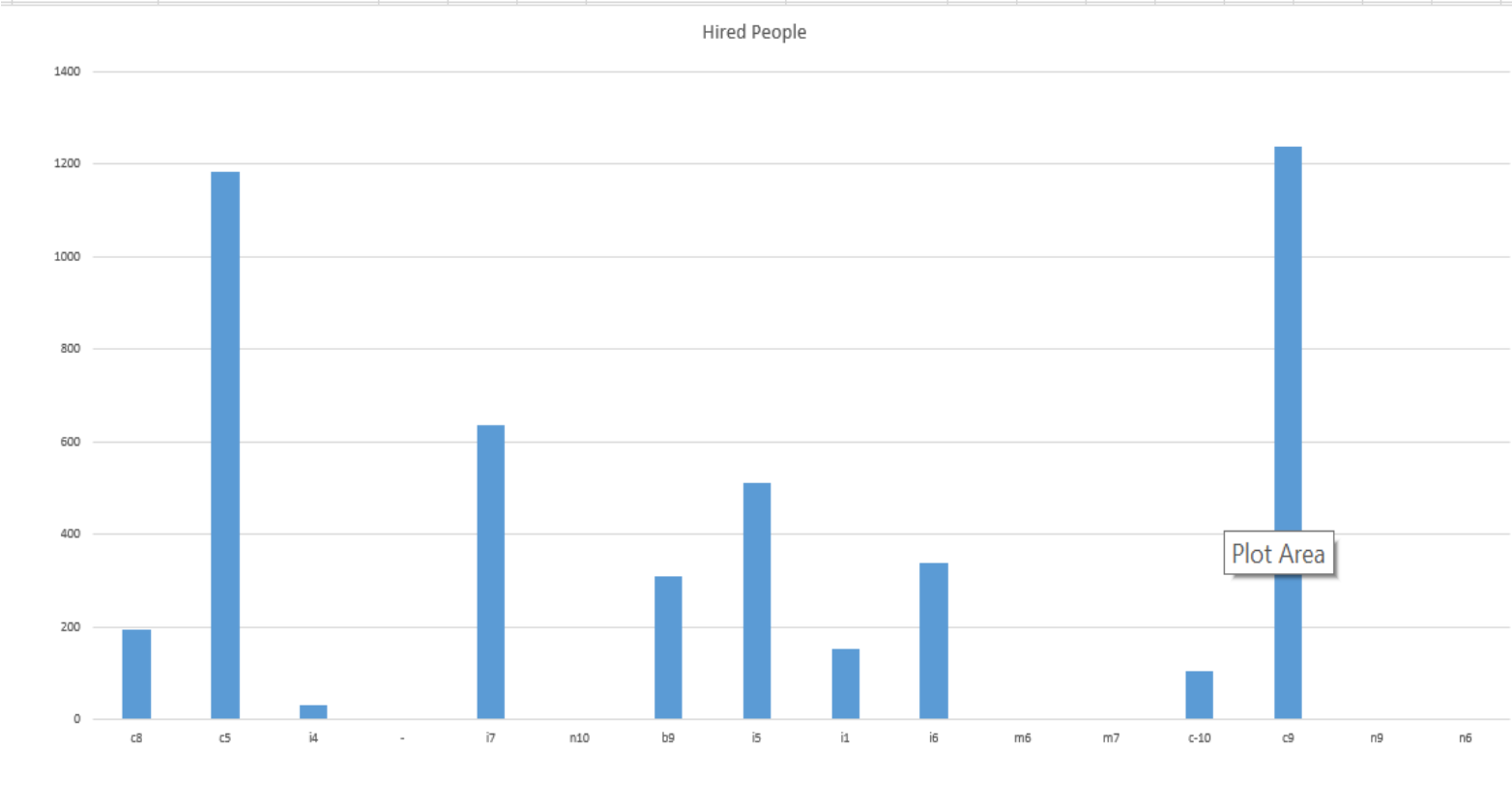
Your task: Represent different post tiers using chart/graph?

Column:-Post Name Define Name:-Post

Column:-Status Define Name:-Status

=COUNTIFS(Statistic.xlsx!Status,"Hired",Post,"c8")

Post Name	Hired People
c8	193
c5	1182
i4	32
-	1
i7	635
n10	0
b9	308
i5	511
i1	151
i6	337
m6	2
m7	0
c-10	105
c9	1239
n9	0
n6	1
Total People	4697



1)Understanding data columns and data:

To get information about the columns in the table and their data types, you can use the following SQL command:

DESCRIBE sheet1;

	Field	Type	Null	Key	Default	Extra
▶	application_id	int	YES		NULL	
	Interview Taken on	text	YES		NULL	
	Status	text	YES		NULL	
	event_name	text	YES		NULL	
	Department	text	YES		NULL	
	Post Name	text	YES		NULL	
	Offered Salary	int	YES		NULL	

SELECT * FROM sheet1 LIMIT 10;

	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary
▶	383422	5/1/2014 11:40	Hired	Male	Service Department	c8	56553
	907518	5/6/2014 8:08	Hired	Female	Service Department	c5	22075
	176719	5/6/2014 8:08	Rejected	Male	Service Department	c5	70069
	429799	5/2/2014 16:28	Rejected	Female	Operations Department	i4	3207
	253651	5/2/2014 16:32	Hired	Male	Operations Department	i4	29668
	289907	5/1/2014 7:44	Hired	Male	Sales Department	-	85914
	959124	5/6/2014 16:27	5/1/2014 7:44	Male	Sales Department	i7	69904
	86642	5/9/2014 13:17	Rejected	Male	Sales Department	i7	11758
	751029	5/2/2014 13:09	Hired	Female	Service Department	i4	15156
	434547	5/2/2014 13:11	Rejected	Female	Service Department	i4	49515

2)Checking for missing data:

**SELECT COUNT(*) - COUNT('application_id') AS missing_count
FROM sheet1;**

	missing_count
▶	0

**SELECT COUNT(*) - COUNT('Interview Taken on') AS missing_count
FROM sheet1;**

	missing_count
▶	0

**SELECT COUNT(*) - COUNT('Offered Salary') AS missing_count
FROM sheet1;**

	missing_count
▶	0

```
SELECT COUNT(*) - COUNT('Post Name`) AS missing_count
FROM sheet1;
```

	missing_count
▶	0

```
SELECT COUNT(*) - COUNT('Department') AS missing_count
FROM sheet1;
```

	missing_count
▶	0

```
SELECT COUNT(*) - COUNT('event_name') AS missing_count
FROM sheet1;
```

	missing_count
▶	0

```
SELECT COUNT(*) - COUNT('Status') AS missing_count
FROM sheet1;
```

	missing_count
▶	0

```
SELECT COUNT(*) - COUNT('Interview Taken on`) AS missing_count
FROM sheet1;
```

	missing_count
▶	0

```
SELECT COUNT(*) - COUNT('application_id`) AS missing_count
FROM sheet1;
```

	missing_count
▶	0

3) Clubbing columns with multiple categories:

```
ALTER TABLE sheet1
```

```
ADD COLUMN `Job Title` VARCHAR(255);
```

```
UPDATE sheet1
```

```
SET `Job Title` = CONCAT(Department, ' - ', `Post Name`);
```

	application_id	Interview Taken on	Status	event_name	Department	Post Name	Offered Salary	Job Title
▶	383422	5/1/2014 11:40	Hired	Male	Service Department	c8	56553	Service Department - c8
	907518	5/6/2014 8:08	Hired	Female	Service Department	c5	22075	Service Department - c5
	176719	5/6/2014 8:08	Rejected	Male	Service Department	c5	70069	Service Department - c5
	429799	5/2/2014 16:28	Rejected	Female	Operations Department	i4	3207	Operations Department - i4
	253651	5/2/2014 16:32	Hired	Male	Operations Department	i4	29668	Operations Department - i4

4)Checking for outliers:

```
SELECT MIN(`Offered Salary`) AS min_val, MAX(`Offered Salary`) AS max_val,
       AVG(`Offered Salary`) AS mean_val, STDDEV(`Offered Salary`) AS std_dev
FROM sheet1;
```

	min_val	max_val	mean_val	std_dev
▶	100	400000	49983.0290	28852.163830905432

Syntax:

```
SELECT MIN(column_name) AS min_val, MAX(column_name) AS max_val,
       AVG(column_name) AS mean_val, STDDEV(column_name) AS std_dev
FROM sheet1;
```

5)Removing outliers:

```
DELETE FROM sheet1 WHERE `Offered Salary` < 10000 OR `Offered Salary` > 100000;
1362 row(s) affected
```