PRESENTATION

HIRING PROCESS ANALYTICS

Outline

01 **Problem Description**

02 **Approach**

- Performing Exploratory Data Analysis
- Answer Questions

03 **Insights**

04 **Results**

Problem Description

- The hiring process is an essential function of any company where the company learns the major underlying trends about the hiring process. such as the number of rejections, number of interviews, types of jobs, vacancies, etc

- With the given dataset of a company where the details about people who registered for a particular post in a department of this company, the task is to perform EDA of data and draw necessary conclusions about the company.

- Also, we have to answer several questions which include the number of males and females hired, the average salary offered in a company, class intervals for salary, and display charts to show the proportion of people working in different departments and represent different post tiers.

Approach

Before deriving any conclusions, we will first perform Exploratory Data Analysis to understand the data.

i) Understanding data columns and data

This step focuses on getting familiarized with the column names and their data types.

Here, all the columns are In general format except one column "Interview Taken On" which tells about the data the interview is conducted on. It is in dd:mm:yyyy hh:mm format.

EDA Steps

- Understanding data columns and data
- Check for missing data
- Clubbing coumns with multiple categiries
- Checking for outliers
- Removing Outliers
- Drawing data summary

ii) Check for missing data

In this step, we count the number of blank cells in each column and the percentage of missing data in each column using the formula:
(count of blank cells) / (total number of records) * 100

I have used **COUNTBLANK()** function to count missing values and **COUNTIF()** function because some columns have '-' values which can be well understood as missing.

- Offered Salary column contains **1** missing value.
- Post Name column contains **1** missing value.
- event_name column which holds information of gender contains **15** missing values.
  393 candidates haven't informed their gender, which is denoted by "Don’t want to say" in the spreadsheet.

iii) Clubbing columns with multiple categories

In this step, we club the multiple categories of column into more broader category to get some wonderful insights.
Here, I have clubbed the categories in the event_name column which contains the gender information into two broad categories that are "Gender Known" or "Gender Unknown".

=IF(OR(D2="Male", D2="Female"), "Gender Known", IF(OR(D2="Don’t want to say",D2="-"), "Gender Unknown", D2))

By clubbing the event_name column which tells about the gender of the candidate, we found out there are 6760 candidates who have informed their gender whereas 408 candidates have not.

Similarly, Department and Post Name can also be clubbed according to the questions which are to be answered for the case study.

iv) Checking for outliers

In this step, we check whether any column contains outliers. An outlier can be defined as a data point that significantly deviates or differs from the majority of the other data points in a dataset. can be defined as a data point that significantly deviates or differs from the majority of the other data points in a dataset.
Here, there Is only one column that contained numeric values and can have outliers which is Offered Salary.
=IF(ABS(A2-AVERAGE(A:A)) > 3*STDEV(A:A), "Outlier", "")

3 records were found which are stated as outliers from the above function because they have very high salaries - 200000, 300000, and 400000 respectively.

But, outliers are usually defined as abnormal or impossible values for a specific field. In this case, these 3 records have very high salaries, that's why it comes under the outlier category, but it is completely possible to have higher salaries in comparison to other employees. So, here we don't have to remove any outliers.

v) Removing outliers

For removing outliers,

- **Filter the dataset based on the column containing potential outliers.**
- **Remove the filtered rows from the dataset.**

Here, no such outliers are found which should be removed.

vi) Drawing data summary

Summarizing Categorical Variables

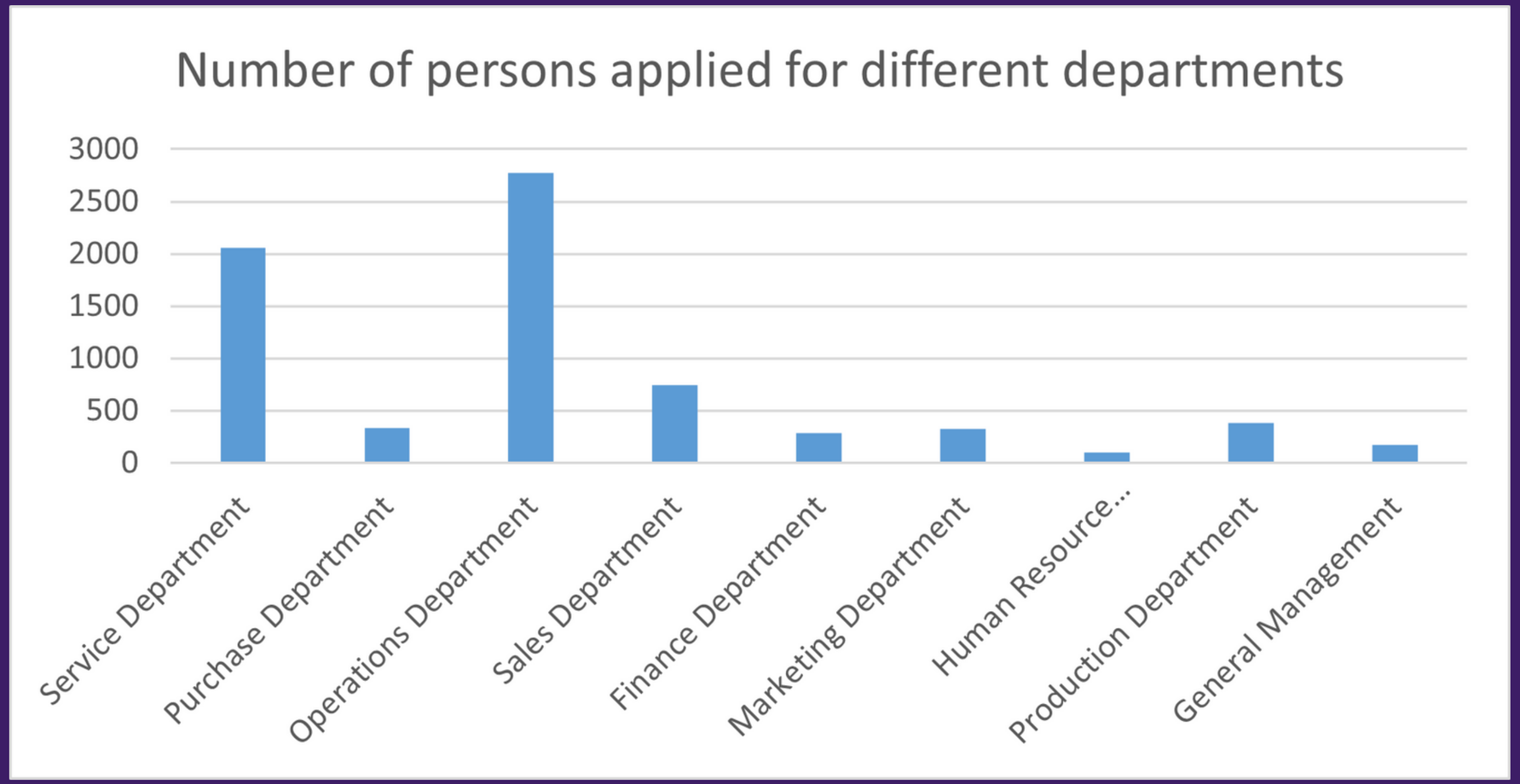
- **Status**

The table creation for plotting charts is done in the Excel sheet.

All analysis is done in this Excel workbook :
https://docs.google.com/spreadsheets/d/1K5vqQW_95e7UCZ3dkxQktkxUaUKXyzxF/edit?usp=sharing&ouid=108310215373032149971&rtpof=true&sd=true

Summarizing Categorical Variables

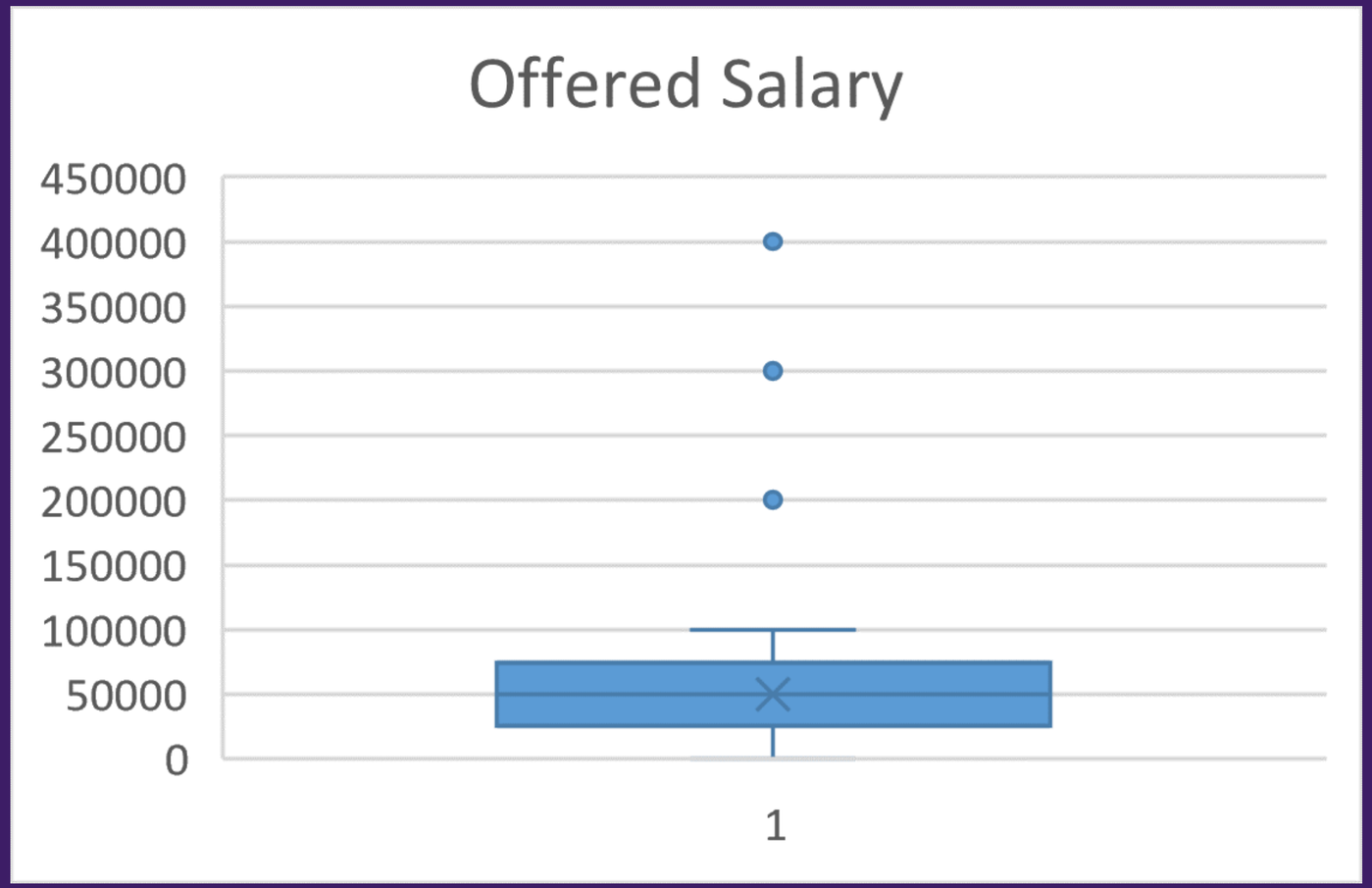
- **Department**



Summarizing Numerical Variables

- **Offered Salary :**

  Here, we can get an idea of the range of salaries.



Hiring

Process of intaking people into an organization for different kinds of positions.

How many males and females are Hired ?

We will first identify the column which contains gender information. We will use COUNTIF() function to calculate the number of males and females.

=COUNTIFS(Data!D:D, "Male", Data!C:C, "Hired")

In the given dataset, column D is labelled as event_name which contains the gender information.

In total, 2563 males and 1856 females are hired.

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.

What is the average salary offered in this company ?

We will first identify the column which contains salary information. We will use the AVERAGE() function to calculate the average salary offered.

=AVERAGE(G:G)

In the given dataset, column G is labeled as Offered Salary which contains the information on salaries of all the employees.

49983.03 is the average salary offered in the company.

Class Intervals

The class interval is the difference between the upper class limit and the lower class limit.

Draw the class intervals for salary in the company ?

We will first find the range of the Offered Salary column by using the formula :
=MAX(G2:G7169) - MIN(G2:G7169)
Here, we get the range equal to 399000.

Then, we further calculate the interval size by using formula :
=ROUNDUP(range/10, 0)
We get the interval size 39900 for dividing the salary into 10 class intervals, we can round off the interval size to the nearest whole number i.e. 40000.

Class Intervals

100-40100
40100-80100
80100-120100
120100-160100
160100-200100
200100-240100
240100-280100
280100-320100
320100-360100
360100-400100

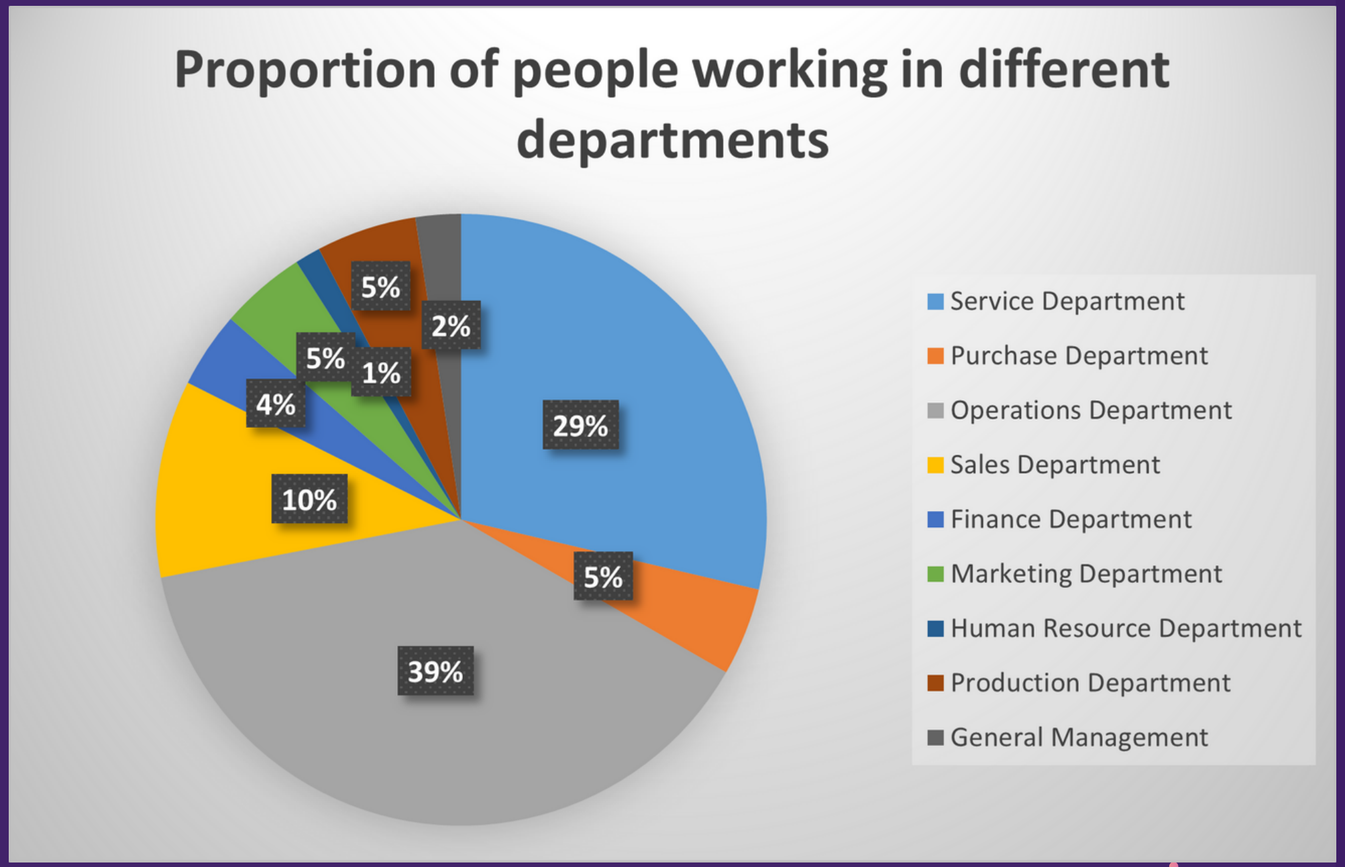
Charts and Plots

This is one of the most important part of analysis to visualize the data.

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

We create a summary table where we mention all the different categories and their corresponding counts in the subsequent column.

Then, we create a pie chart from a summary table to show the proportion of people working in different departments.



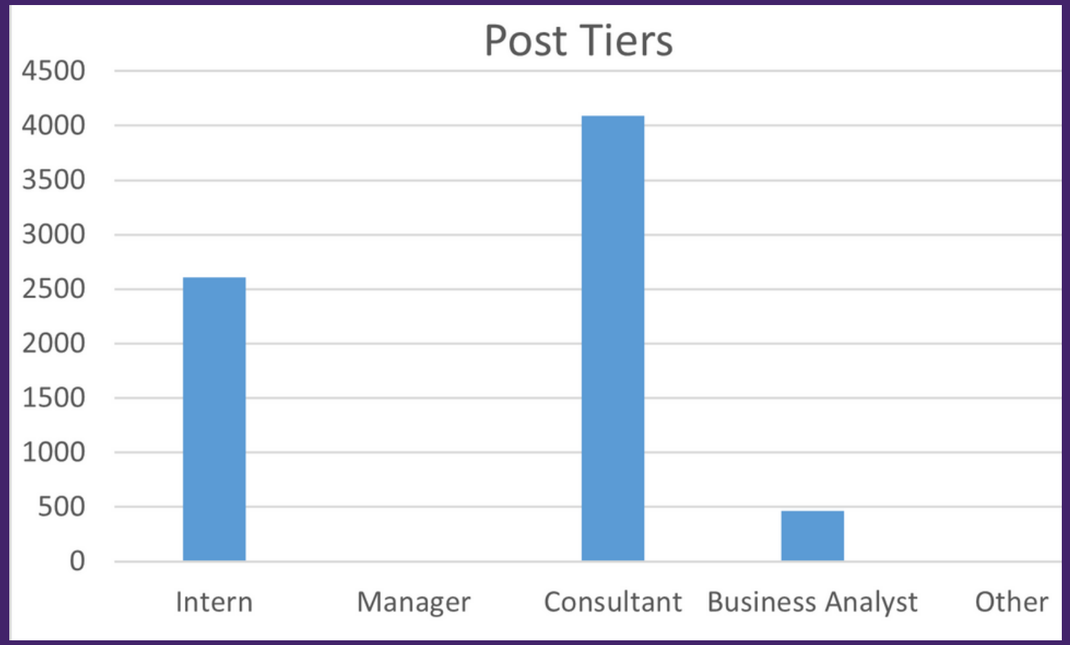
Charts

Use different charts and graphs to perform the task representing the data.

Represent different post tiers using chart/graph?

In the dataset, the Post Name column contains values like 'c9', 'm4', 'b9', 'i4' etc.
So, we have created a column Post Tier and categorized these names into different post tiers. If the name starts with c, then "Consultant", "Manager" if it starts with m, "Intern" if it starts with i, "Business Analyst" if it starts with b, and "Other" if it starts with any other letter.

```
=IF(ISNUMBER(SEARCH("m",F2)),"Manager",IF(ISNUMBER(SEARCH("c",F2)),"Consultant",IF(ISNUMBER(SEARCH("i",F2)),"Intern",IF(ISNUMBER(SEARCH("b",F2)),"Business Analyst","Other"))))
```



Then, we define a summary table for different post tiers and their corresponding counts and then create a column chart from a summary table to represent different post tiers. We got 2606 interns, 4 managers, 4091 consultants, 463 business analysts, and 4 in the others category.

TECH STACK USED

MICROSOFT EXCEL

Excel is a powerful spreadsheet software that offers a wide range of tools and features, including data analysis functions, charting capabilities, and data visualization options, making it a versatile tool for conducting data analysis and creating informative visualizations.

Insights

- Exploratory data analysis is done to understand the data.
- We can infer wonderful insights from the raw data using a powerful tool like Ms-Excel.

From this data, we have gained the following insights :

- The proportion of males and females hired is determined from the data available in the dataset.
- The average salary offered by the company is calculated from the salary data.
- The class intervals for salary is determined to better understand the distribution of salaries.
- The proportion of people working in different departments is visualized using a pie chart.
- Different post tiers are represented using a column chart to understand the distribution of posts in the company.

Results

This analysis has proved to be very beneficial. By doing this project, I got to know about the steps mentioned below.

- Data cleaning: The process of cleaning and transforming data is an important step in conducting any analysis. By working with a real-world dataset, one can improve their skills in data cleaning using Excel functions such as sorting, and filtering.

- Data manipulation: Excel offers a variety of functions and formulas that can be used to manipulate and transform data such as SUM, AVERAGE, and COUNT

- Data visualization: Excel is a powerful tool for creating charts and graphs to visualize data.