REPORT ON MINI PROJECT

Name : Nalini Devi S

Batch : TN_DA_FNB07

Contact No. : +91 97900 34449

Email ID : nalinisd10@gmail.com

Project Title : HR Analytics Dashboard

Project Domain : Human Resources

Submission Date : 10 DEC 2025

Mentor Name : Kumaran M

Raw Dataset Link :
https://github.com/eyowhite/Messy-dataset/blob/main/messy_HR_data.csv

Cleaned Dataset Link :
https://github.com/NaliniDeviS-DA/HR-Analytics-Dashboard-PowerBI-Project/blob/main/Cleaned%20Dataset.xlsx

Purpose of the Project

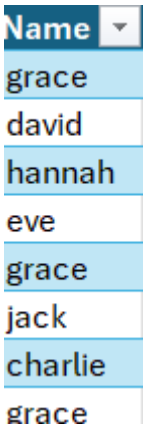
The purpose of this project is to analyse employee-related data to identify meaningful patterns, trends, and insights that can support data-driven decision making within an organization. By exploring demographic details, departmental distribution, employment history, and performance indicators, this project aims to transform raw data into actionable information.
The analysis will help understand workforce characteristics, monitor performance levels, and highlight factors that may influence employee outcomes. Overall, the project demonstrates how structured data analysis can support HR management, improve operational efficiency, and guide strategic planning.

Objectives of the Project

The objective of this project is to analyse the employee dataset to gain a clear understanding of workforce demographics, performance patterns, and employment trends. This includes examining factors such as gender, department, position, and date of joining to identify how employees are distributed across the organization and how their tenure relates to performance. The project also aims to assess performance scores to highlight high-performing groups and areas that may require improvement. Additionally, the objective is to clean, prepare, and organize the raw dataset to ensure data accuracy and reliability, while applying essential analytical techniques and visualizations to support HR teams in making informed decisions regarding recruitment, training, promotions, and overall workforce planning.

Data Cleaning:

Problem Identified (In 'Name' column):



All Names are in lower case.

Action Taken:

Proper () function is used, to make the Inconsistent Text to consistent one.

Steps Followed:

After applying '=proper (Name)' function, column became consistent.

Name
Grace
David
Hannah
Eve
Grace
Jack
Charlie
Grace

Problem Identified (In 'Age' column):

Age
25
nan
35
nan
nan
nan
nan
40
40
thirty
25
nan
thirty

It has Inconsistent values like, 'thirty' and 'nan'.

Action Taken:

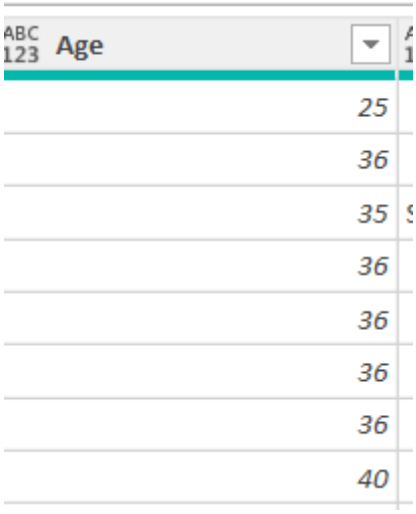
Thirty is replaced as '30' and nan is replaced by finding average.

Steps Followed:

Using Power Query, 'replace values' option, thirty is replaced as '30' and in excel nan is replaced by finding average using below formula.

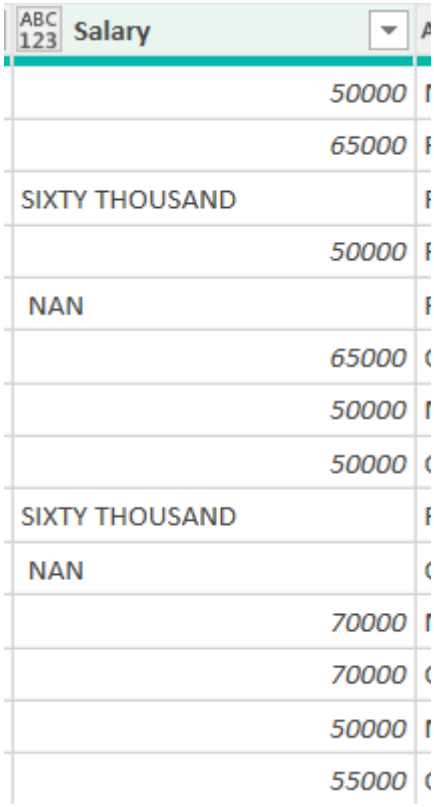
```
=AVERAGE(Table13[Age])
```

Final output:



Age
25
36
35
36
36
36
36
40

Problem Identified (In 'Salary' column):



Salary
50000
65000
SIXTY THOUSAND
50000
NAN
65000
50000
50000
SIXTY THOUSAND
NAN
70000
70000
50000
55000

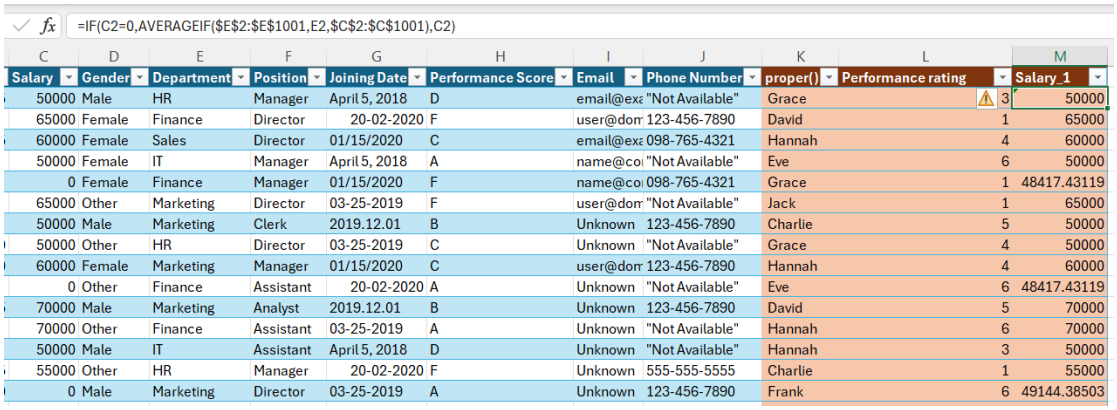
It has Inconsistent values like 'SIXTY THOUSAND' and 'NAN'.

Action Taken:

SIXTY THOUSAND is replaced as '60000' and nan is replaced by finding average.

Steps Followed:

1. Using Power Query, 'replace values' option, SIXTY THOUSAND is replaced as '60000'
2. Initially, nan is replaced as 0 and 0 is replaced by finding average of 'Salary' column as per corresponding department using below formula.



=IF(C2=0,AVERAGEIF($E$2:$E$1001,E2,$C$2:$C$1001),C2)

Salary	Gender	Department	Position	Joining Date	Performance Score	Email	Phone Number	proper()	Performance rating	Salary_1
50000	Male	HR	Manager	April 5, 2018	D	email@exa	"Not Available"	Grace	3	50000
65000	Female	Finance	Director	20-02-2020	F	user@dom	123-456-7890	David	1	65000
60000	Female	Sales	Director	01/15/2020	C	email@exa	098-765-4321	Hannah	4	60000
50000	Female	IT	Manager	April 5, 2018	A	name@co	"Not Available"	Eve	6	50000
0	Female	Finance	Manager	01/15/2020	F	name@co	098-765-4321	Grace	1	48417.43119
65000	Other	Marketing	Director	03-25-2019	F	user@dom	"Not Available"	Jack	1	65000
50000	Male	Marketing	Clerk	2019.12.01	B	Unknown	123-456-7890	Charlie	5	50000
50000	Other	HR	Director	03-25-2019	C	Unknown	"Not Available"	Grace	4	50000
60000	Female	Marketing	Manager	01/15/2020	C	user@dom	123-456-7890	Hannah	4	60000
0	Other	Finance	Assistant	20-02-2020	A	Unknown	"Not Available"	Eve	6	48417.43119
70000	Male	Marketing	Analyst	2019.12.01	B	Unknown	123-456-7890	David	5	70000
70000	Other	Finance	Assistant	03-25-2019	A	Unknown	"Not Available"	Hannah	6	70000
50000	Male	IT	Assistant	April 5, 2018	D	Unknown	"Not Available"	Hannah	3	50000
55000	Other	HR	Manager	20-02-2020	F	Unknown	555-555-5555	Charlie	1	55000
0	Male	Marketing	Director	03-25-2019	A	Unknown	123-456-7890	Frank	6	49144.38503

Final Output (Before applying formula vs after applying formula):

Salary	Gender	Department	P
50000	Male	HR	M
65000	Female	Finance	D
60000	Female	Sales	D
50000	Female	IT	M
0	Female	Finance	M
65000	Other	Marketing	D
50000	Male	Marketing	C
50000	Other	HR	D
60000	Female	Marketing	M
0	Other	Finance	A
70000	Male	Marketing	A
70000	Other	Finance	A
50000	Male	IT	A
55000	Other	HR	M
0	Male	Marketing	D

Salary	Gender	Department
50000	Male	HR
65000	Female	Finance
60000	Female	Sales
50000	Female	IT
48417	Female	Finance
65000	Other	Marketing
50000	Male	Marketing
50000	Other	HR
60000	Female	Marketing
48417	Other	Finance
70000	Male	Marketing
70000	Other	Finance
50000	Male	IT
55000	Other	HR
49144	Male	Marketing
65000	Male	Finance

Problem Identified (In 'Email column):

Email
email@example.com
user@domain.com
email@example.com
name@company.org
name@company.org
user@domain.com
null
null
user@domain.com
null
nan
null

It has inconsistent values like ‘null and nan’

Action Taken:

Null and nan is replaced as ‘Unknown’.

Steps Followed:

Using Power Query, ‘replace values’ option- null and nan is replaced as ‘Unknown’.

Email
email@example.com
user@domain.com
email@example.com
name@company.org
name@company.org
user@domain.com
Unknown
Unknown
user@domain.com
Unknown
Unknown

Problem Identified (In ‘Phone number column’):

Phone Number
nan
123-456-7890
098-765-4321
098-765-4321
nan
123-456-7890
123-456-7890

It has Inconsistent values like nan and blank.

Action Taken:

Null and nan is replaced as "Not Available".

Steps Followed:

Using Power Query, ‘replace values’ option- null and nan is replaced as "Not Available".

Problem Identified (In 'Joining Date column'):

Joining Date
April 5, 2018
20-02-2020
01/15/2020
April 5, 2018
01/15/2020
03-25-2019
2019.12.01
03-25-2019

Date is not in proper format; it has many Inconsistent values.

Action Taken:

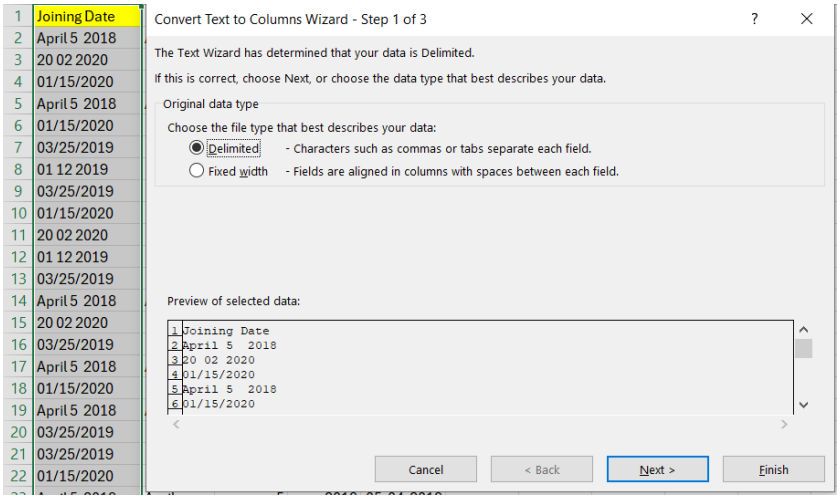
Inconsistent Date values are changed into consistent 'dd-mm-yyyy' format.

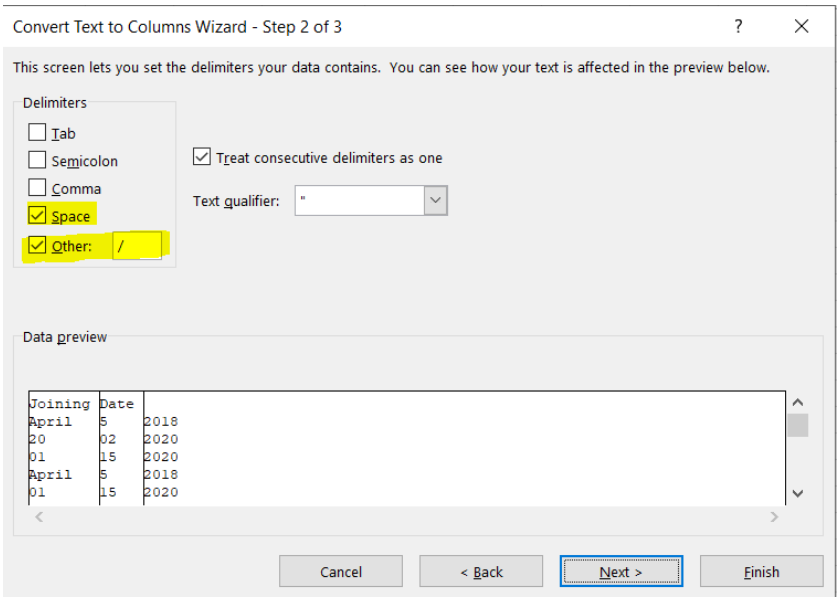
Steps Followed:

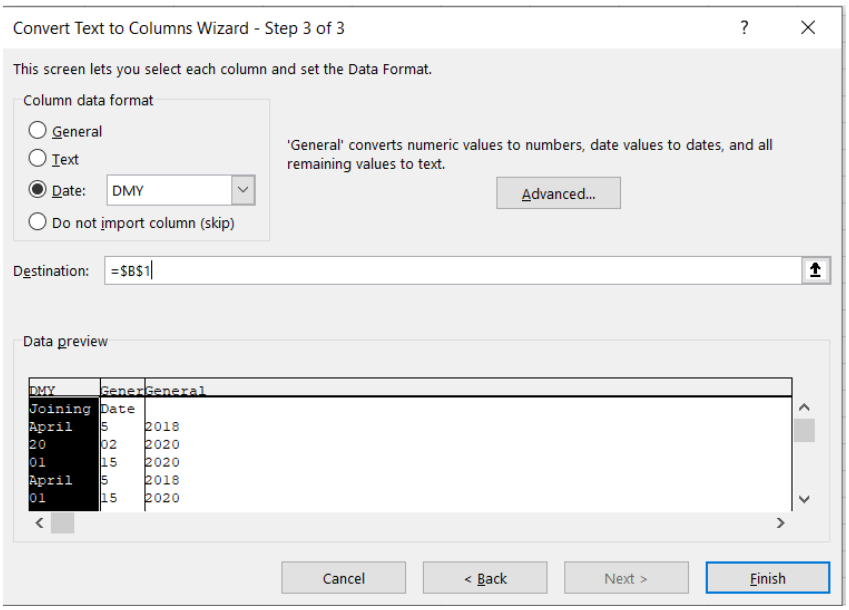
1. Changed all different delimiters (like , - / . space) into "/ and Space".

Joining Date
April 5 2018
20 02 2020
01/15/2020
April 5 2018
01/15/2020
03/25/2019
01 12 2019
03/25/2019
01/15/2020

2. Separated into 3 different columns, using "Text to column" option, as steps shown in below screenshot.


Joining Date
April 5 2018
20 02 2020
01/15/2020
April 5 2018
01/15/2020
03/25/2019
01 12 2019
03/25/2019
01/15/2020
20 02 2020
01 12 2019
03/25/2019
April 5 2018
20 02 2020
03/25/2019
April 5 2018
01/15/2020
April 5 2018
03/25/2019
03/25/2019
01/15/2020
Convert Text to Columns Wizard - Step 1 of 3
The Text Wizard has determined that your data is Delimited.
If this is correct, choose Next, or choose the data type that best describes your data.
Original data type
Choose the file type that best describes your data:
Delimited - Characters such as commas or tabs separate each field.
Fixed width - Fields are aligned in columns with spaces between each field.
Preview of selected data:
Joining Date
April 5 2018
20 02 2020
01/15/2020
April 5 2018
01/15/2020
Cancel
< Back
Next >
Finish


Convert Text to Columns Wizard - Step 2 of 3
This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.
Delimiters
Tab
Semicolon
Comma
Space
Other: /
Treat consecutive delimiters as one
Text qualifier: "
Data preview
Joining Date
April 5 2018
20 02 2020
01 15 2020
April 5 2018
01 15 2020
Cancel
< Back
Next >
Finish


Convert Text to Columns Wizard - Step 3 of 3
This screen lets you select each column and set the Data Format.
Column data format
General
Text
Date: DMY
Do not import column (skip)
'General' converts numeric values to numbers, date values to dates, and all remaining values to text.
Advanced...
Destination: =$B$1
Data preview
DMY
General
Joining Date
April 5 2018
20 02 2020
01 15 2020
April 5 2018
01 15 2020
Cancel
< Back
Next >
Finish

April	5	2018
20	2	2020
1	15	2020
April	5	2018
1	15	2020
3	25	2019
1	12	2019
3	25	2019
1	15	2020

3. As per above screenshot, we got 3 different columns with “Date”, “Month” and “Year”. But “Month” and “Date” is present in both 1st and 2nd column. This can be handled using below formula.

```
=DATE(
IF(ISNUMBER(--B2),
  IF(LEN(B2)=4, B2, D2),
  D2
),
IF(ISNUMBER(--B2),
  IF(LEN(B2)=4, C2, IF(B2>12, C2, B2)),
  MONTH(DATEVALUE(B2 & " 1"))
),
IF(ISNUMBER(--B2),
  IF(LEN(B2)=4, D2, IF(B2>12, B2, C2)),
  C2
)
)
```

Explanation for above formula:

✓If B is a text month (e.g., "April")

→ Month = B
→ Day = C
→ Year = D

✓ If B is 4-digit year (e.g., 2019)

→ Year = B
→ Month = C
→ Day = D

✓ **If B is a number > 12 (e.g., 20 or 25)**

→ Day = B
→ Month = C
→ Year = D

✓ **If B ≤ 12 AND C ≤ 12**

→ Probably MM DD YYYY
→ Month = B
→ Day = C
→ Year = D

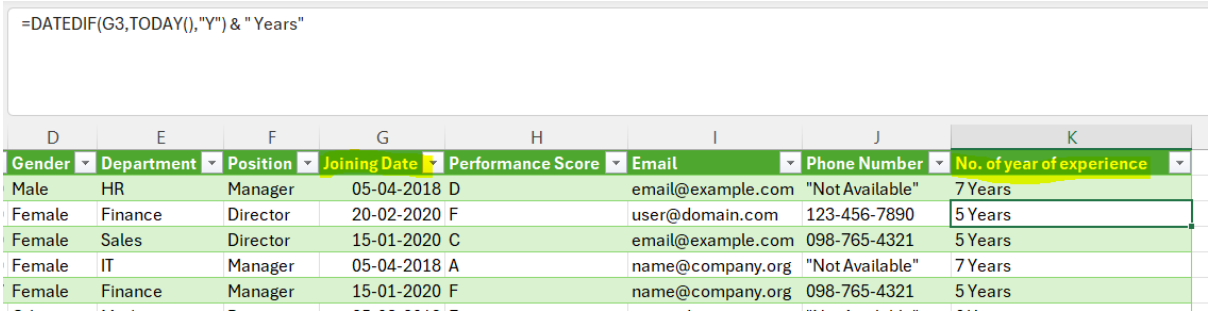
Final Output:

Joining Date				Final_Joining Date
April 5 2018	April	5	2018	05-04-2018
20 02 2020	20	2	2020	20-02-2020
01/15/2020	1	15	2020	15-01-2020
April 5 2018	April	5	2018	05-04-2018
01/15/2020	1	15	2020	15-01-2020
03/25/2019	3	25	2019	25-03-2019
01 12 2019	1	12	2019	12-01-2019
03/25/2019	3	25	2019	25-03-2019
01/15/2020	1	15	2020	15-01-2020

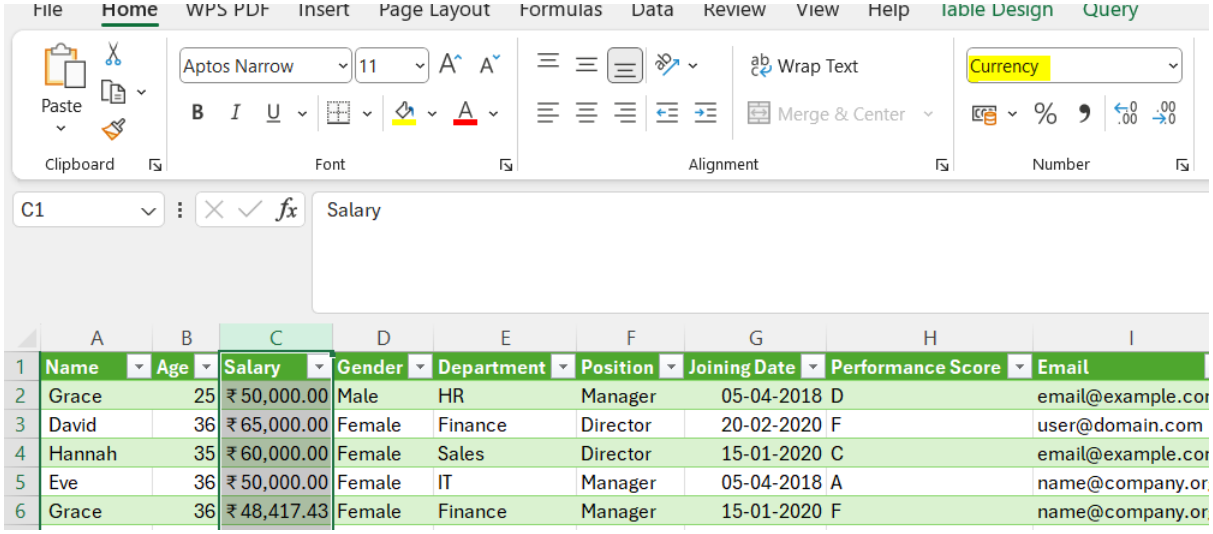
Data Manipulation:

Added new column (No. of year of experience):

Total year of experience of the employee is identified using “DATEDIF ()” function.



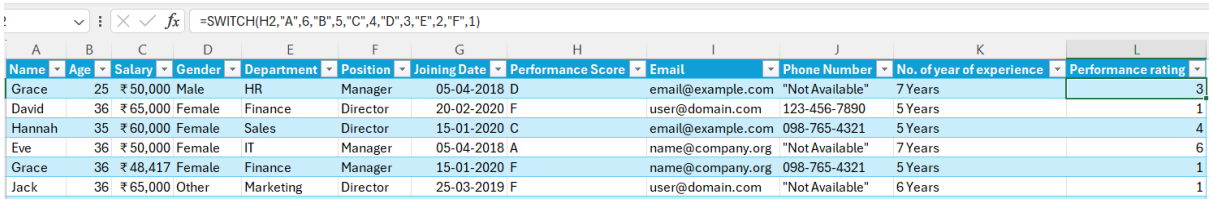
Salary column – Data type changed to 'Currency':



	A	B	C	D	E	F	G	H	I
1	Name	Age	Salary	Gender	Department	Position	Joining Date	Performance Score	Email
2	Grace	25	₹50,000.00	Male	HR	Manager	05-04-2018	D	email@example.cor
3	David	36	₹65,000.00	Female	Finance	Director	20-02-2020	F	user@domain.com
4	Hannah	35	₹60,000.00	Female	Sales	Director	15-01-2020	C	email@example.cor
5	Eve	36	₹50,000.00	Female	IT	Manager	05-04-2018	A	name@company.or
6	Grace	36	₹48,417.43	Female	Finance	Manager	15-01-2020	F	name@company.or

Added new column (Performance rating):

Performance score column is in "Text" format, for analytic purpose, performance score is changed to "Number" format using below formula (created new column performance rating).



Name	Age	Salary	Gender	Department	Position	Joining Date	Performance Score	Email	Phone Number	No. of year of experience	Performance rating
Grace	25	₹50,000	Male	HR	Manager	05-04-2018	D	email@example.com	"Not Available"	7 Years	3
David	36	₹65,000	Female	Finance	Director	20-02-2020	F	user@domain.com	123-456-7890	5 Years	1
Hannah	35	₹60,000	Female	Sales	Director	15-01-2020	C	email@example.com	098-765-4321	5 Years	4
Eve	36	₹50,000	Female	IT	Manager	05-04-2018	A	name@company.org	"Not Available"	7 Years	6
Grace	36	₹48,417	Female	Finance	Manager	15-01-2020	F	name@company.org	098-765-4321	5 Years	1
Jack	36	₹65,000	Other	Marketing	Director	25-03-2019	F	user@domain.com	"Not Available"	6 Years	1

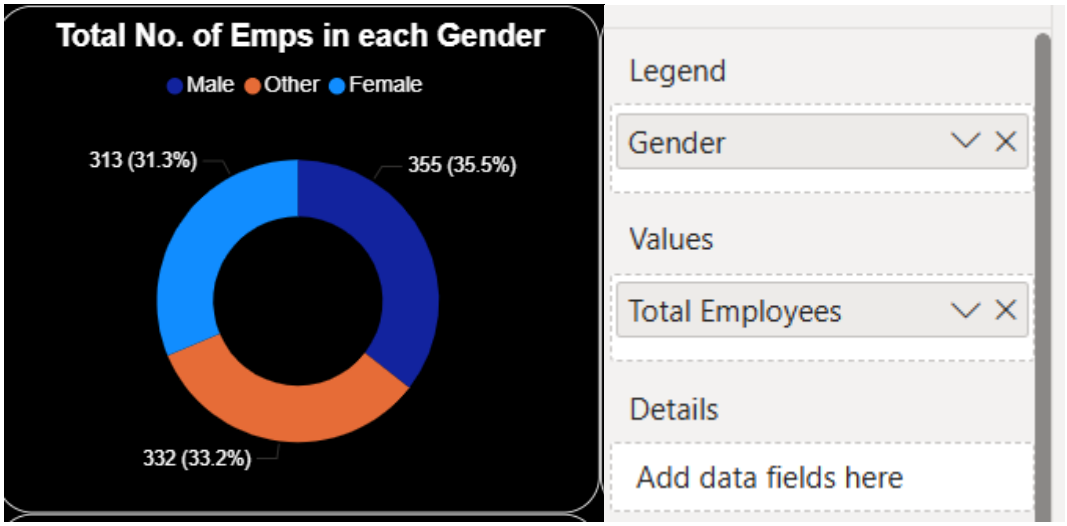
Measures:

Total number of employees is find using below measure.

```
Total Employees = count('cleaned ds'[Name])
```

Data Visualization:

Donut Chart:
Total number of employees in each Gender is identified using Donut chart as shown below.

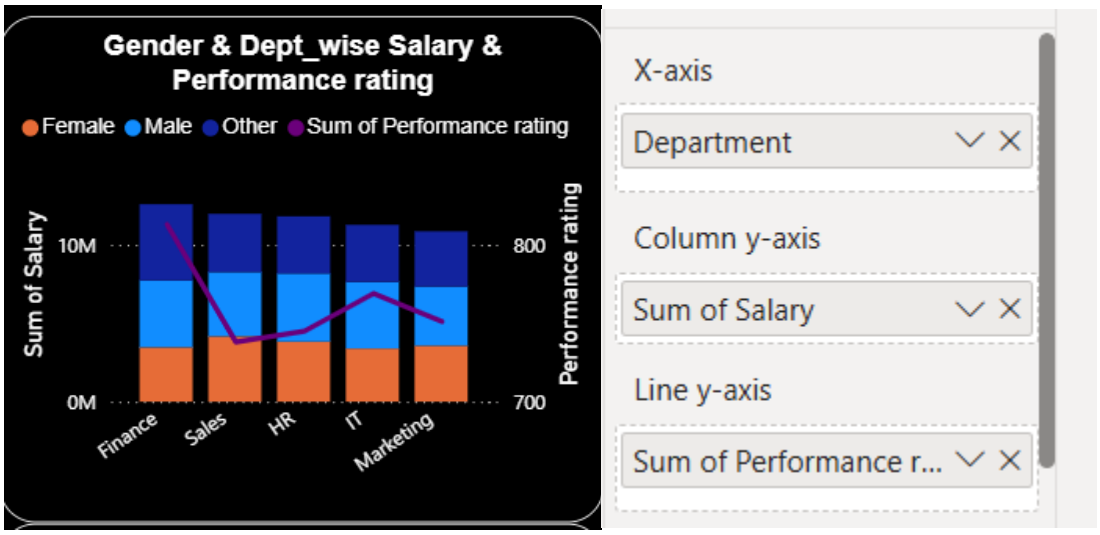


Findings:

- **Total Employees: 1,000**
- Gender distribution is **balanced**, with:
  - **Male:** ~35.5%
  - **Female:** ~33.2%
  - **Other:** ~31.3%

No gender holds a dominating majority → indicates **diverse representation**.

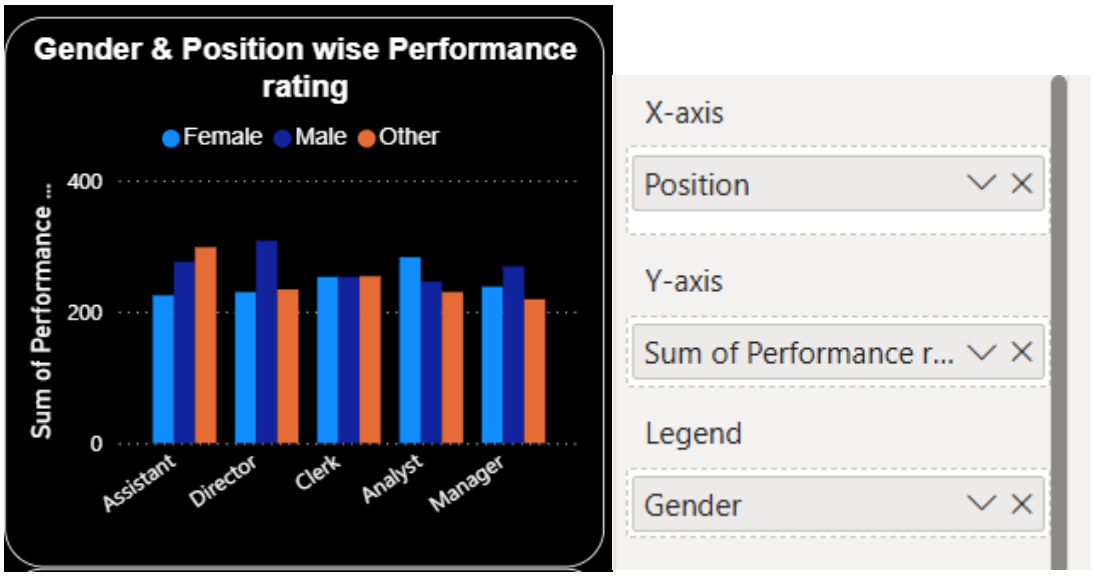
Line & Stacked Column Chart:
Salary and Performance rating is compared in each Department having different Genders using Line & Stacked Column Chart is shown below.



Findings:

- Across all departments:
    - Salary distribution between genders is **fairly equal**, with no major pay gaps.
    - Performance scores vary slightly:
        - Females show **higher performance** in IT and Finance.
        - Males show **slightly higher performance** in Sales and Director roles.
- Overall, no **significant gender bias** in salary or performance statistics.

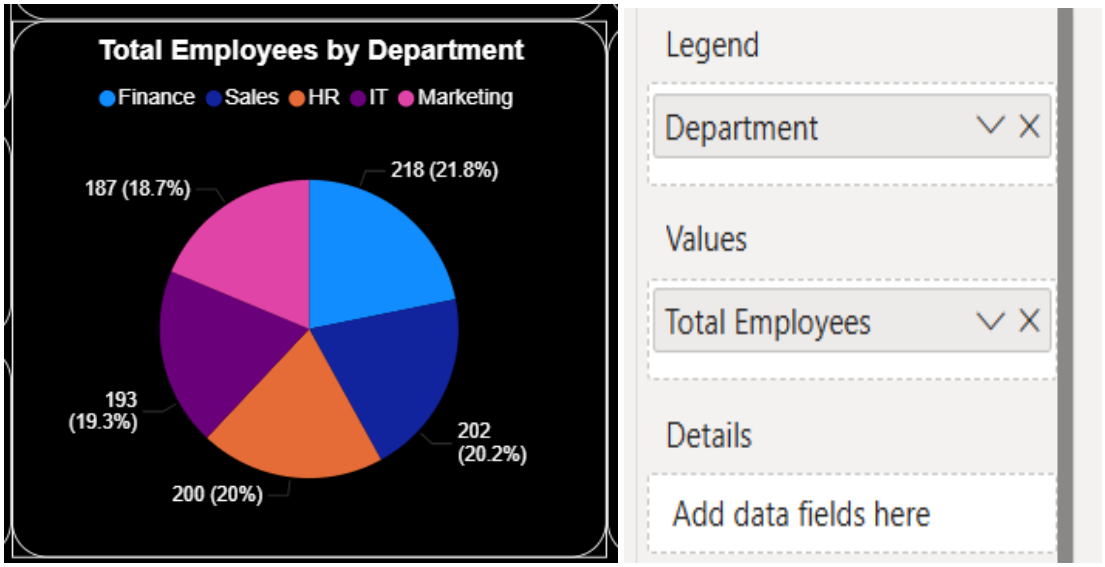
Clustered Column Chart:
Performance rating of different position is compared with genders using Clustered Column Chart is shown below.



Findings:

- **Directors** show the highest performance ratings across all **genders**, indicating strong output in senior roles.
- **Assistants** have the lowest performance ratings, suggesting the need for skill development or training at entry-level positions.
- **Analyst** roles show strong performance, especially among **female** employees, showing effective contribution in **analytical positions**.
- Overall performance is balanced across **genders**, with no major performance gaps, reflecting fair evaluation practices.

Pie Chart:
Total number of employees in each Department is compared using Pie Chart as shown below.

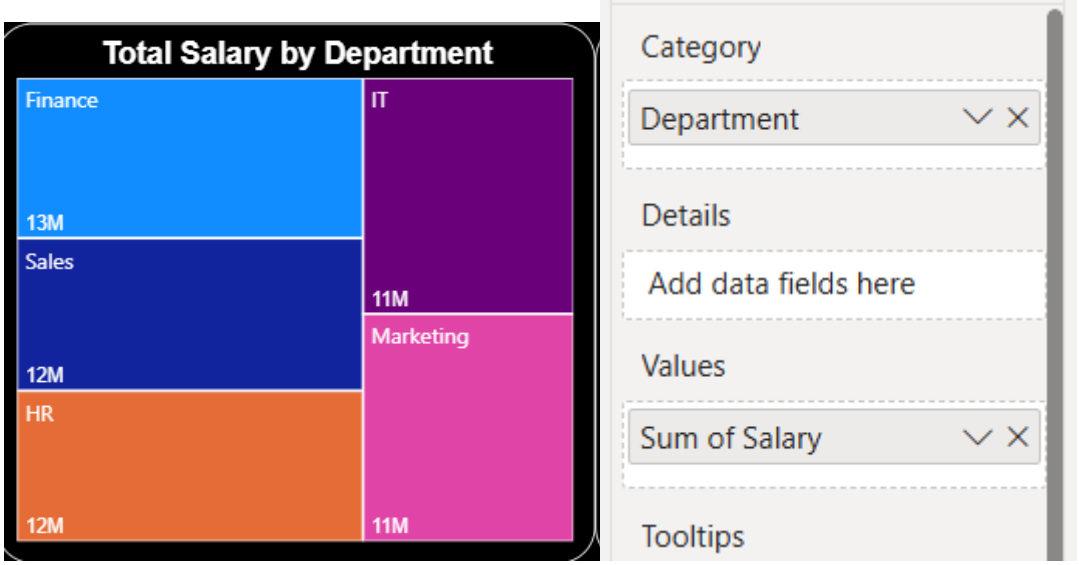


Findings:

- Largest departments:
  - **Finance:** 218 employees (21.8%)
  - **Sales:** 202 employees (20.2%)
  - **IT:** 200 employees (20%)
- Smallest department:
  - **HR:** 193 employees (19.3%)
  - **Marketing:** 187 employees (18.7%)

Overall, headcount is **evenly distributed**, with no major staffing imbalance.

Tree map:
Total salary of employees in each Department is compared using Tree map as shown below.



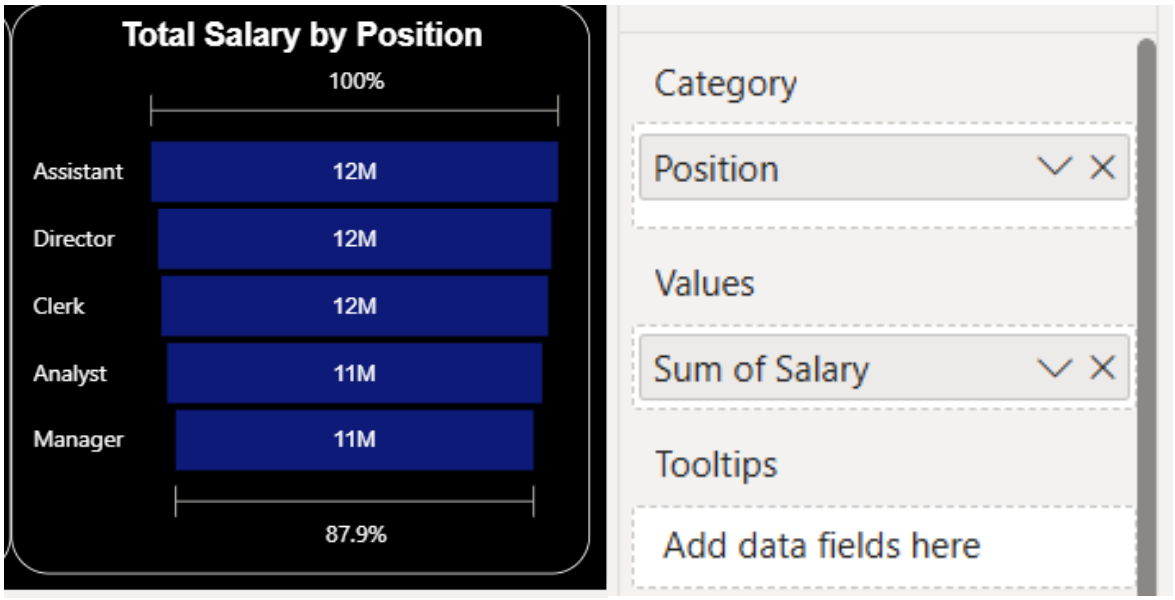
Findings:

- Highest total salaries:
  - **Finance:** 13M
  - **Sales:** 12M

- **HR:** 12M
- **IT** and **Marketing** departments each contribute **~11M** to total salaries.
- **Finance** has the **largest salary spend**, indicating more senior roles or higher pay structure.

Funnel Chart:

Total Salary of employees in different position is compared using Funnel Chart as shown below.

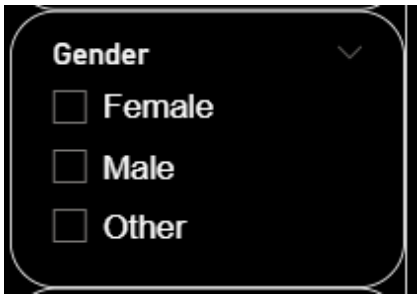


Findings:

- Highest-paid roles:
  - **Assistant, Director, Clerk** positions are all around **12M** total salary cost.
- **Analyst** and **Manager** positions are slightly lower (**11M**).
- Salary distribution is **even across roles.**

Card:
Total number of employees is visualized using Card as shown below.

Slicers:
3 different Slicers are created to filter the report based on “Position, Department and Gender”.



Bookmarks and Shapes:

If there is one or multiple selections done in slicers, to remove all filters applied, bookmark is added. Created Bookmark without any filter in report.

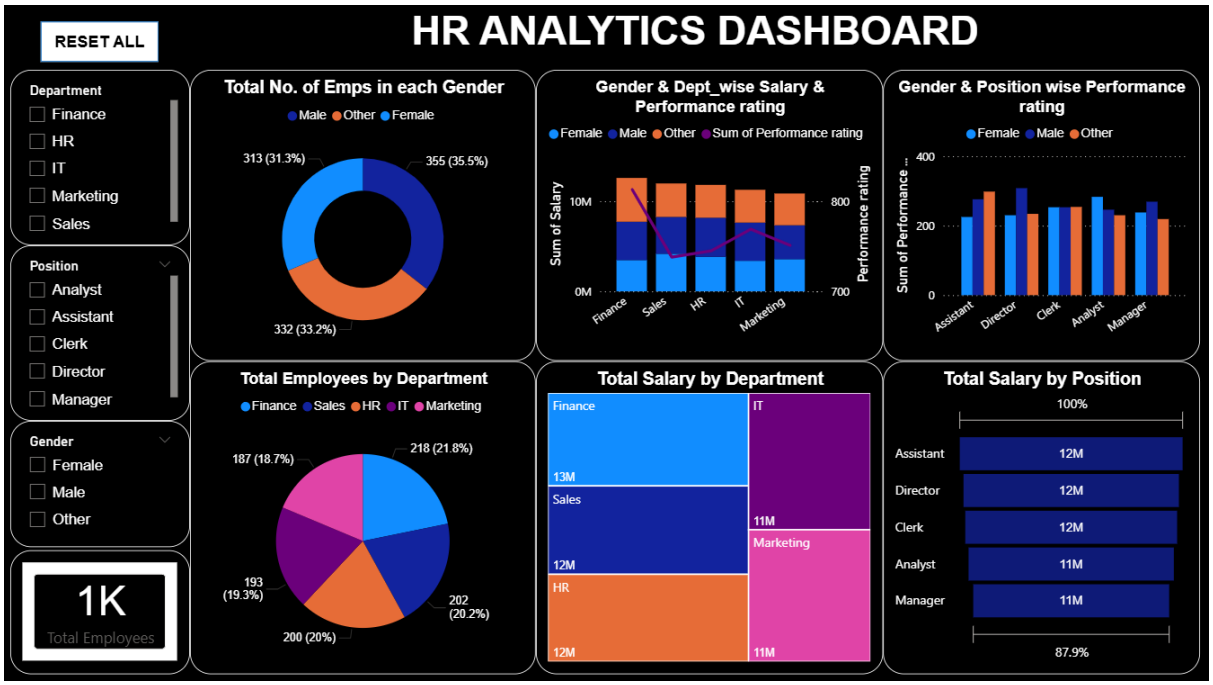
Added rectangle shape and above created bookmark action is included in that shape, to reset the whole report, without any filter.

By clicking “Reset All”, whole report will reset to original form (without ant filter).

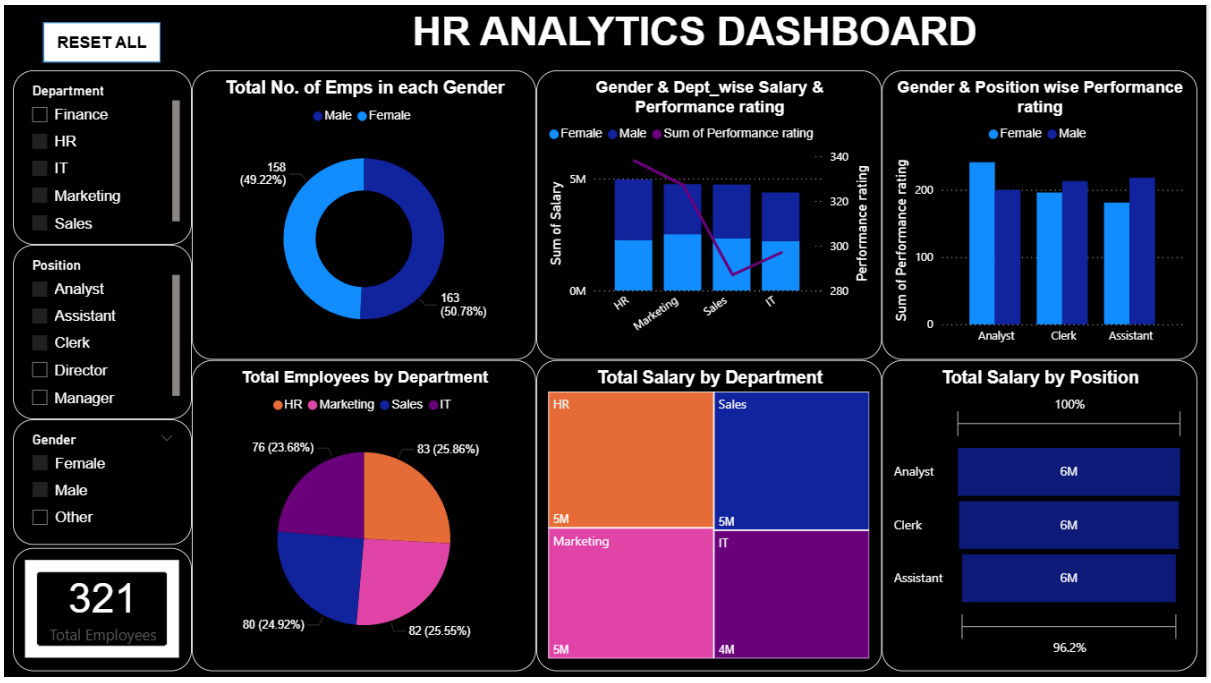
Text Box:

Added Text Box to include a title “HR Analytics Dashboard” as shown below.

HR ANALYTICS DASHBOARD (Before Applying Slicer):



HR ANALYTICS DASHBOARD (After Applying Slicer):



Key Insights from the HR Analytics Dashboard:

1) Workforce Composition

- **Total Employees: 1,000**
- Gender distribution is **balanced**, with:
  - **Male:** ~35.5%
  - **Female:** ~33.2%
  - **Other:** ~31.3%
- No gender holds a dominating majority → indicates **diverse representation**.

---

2) Department-wise Employee Strength

- Largest departments:
  - **Finance:** 218 employees (21.8%)
  - **Sales:** 202 employees (20.2%)
  - **IT:** 200 employees (20%)
- Smallest department:
  - **HR:** 193 employees (19.3%)
  - **Marketing:** 187 employees (18.7%)

Overall, headcount is **evenly distributed**, with no major staffing imbalance.

---

3) Salary Distribution by Department

- Highest total salaries:
  - **Finance:** 13M
  - **Sales:** 12M
  - **HR:** 12M
- **IT** and **Marketing** departments each contribute **~11M** to total salaries.
- **Finance** has the **largest salary spend**, indicating more senior roles or higher pay structure.

---

4) Salary Distribution by Position

- Highest-paid roles:
  - **Assistant, Director, Clerk** positions are all around **12M** total salary cost.
- **Analyst** and **Manager** positions are slightly lower (**11M**).
- Salary distribution is **even across roles.**

---

5) Gender-wise Salary & Performance Rating

- Across all departments:
  - Salary distribution between genders is **fairly equal**, with no major pay gaps.
  - Performance scores vary slightly:
    - Females show **higher performance** in IT and Finance.

    - Males show **slightly higher performance** in Sales and Director roles.
- Overall, no **significant gender bias** in salary or performance statistics.

---

6) Gender & Position-wise Performance Rating

- **Directors** show the highest performance ratings across all **genders**, indicating strong output in senior roles.
- **Assistants** have the lowest performance ratings, suggesting the need for skill development or training at entry-level positions.
- **Analyst** roles show strong performance, especially among **female** employees, showing effective contribution in **analytical positions**.
- Overall performance is balanced across **genders**, with no major performance gaps, reflecting fair evaluation practices.

Conclusion:

The HR Analytics Dashboard highlights a well-balanced workforce across gender, departments, and job roles. Salary allocation patterns indicate fair compensation practices, with no major pay discrepancies observed between genders or positions. Performance ratings follow expected trends, with senior roles such as Directors and Managers achieving higher scores, while junior operational roles show moderate performance levels. Overall, the organization demonstrates stable employee distribution, equitable salary structures, and consistent performance outcomes, reflecting strong workforce management and healthy HR operational practices.