Data and Artificial Intelligence
Cyber Shujaa Program

Week 1 Assignment
Web Scraping and Data Handling in Python

Student Name: Deborah Kwamboka Omae

Student ID: CS-DA02-25075

Table of Contents

Data and Artificial Intelligence 1
Cyber Shujaa Program 1
Week 1 Assignment Web Scraping and Data Handling in Python 1
Introduction 4
Tasks Completed 4
Step 1: connecting to data 4
Step 2: formatting sheets 4
Step 3: calculated columns 5
Step 4: created charts for visualization 6
Step 5: Building a summary dashboard 16
Step 6: building a dashboard for employee details 18
Link to code 21
Conclusion 21

Table of figures

Figure 1: screenshot showing ban total hired....5
Figure 2: screenshot showing total terminated....6
Figure 3: screenshot showing total active....6
Figure 4: line and area chart showing total employees over the years....7
Figure 5: screenshot showing line and area chart for total terminated over the years....8
Figure 6: bar chart of employees by department....8
Figure 7: bar chart of employees by job title....9
Figure 8: bar chart of employees by headquarter and branches....10
Figure 9: map showing employee distribution by state....10
Figure 10: distribution of employees by state....11
Figure 11: pie chart showing gender ratio....12
Figure 12: a heatmap to show the age groups vs group levels....12
Figure 13: bar chart showing distribution of employees by age group....13
Figure 14: bar chart showing employees by education level....14
Figure 15: corelation between educational background and performance rating....14
Figure 16: bar bell chart to show discrepancies in salaries across gender....15
Figure 17: corelation between age and salary for each department....16
Figure 18: the summary dashboard....17
Figure 19: dashboard after adding tooltips....17
Figure 20: dashboard with filters, legends and logos....18
Figure 21: sheet showing employee details....19
Figure 22: length of hire measure added....20
Figure 23: dashboard for employee details and the necessary filters....20

Introduction

This week's assignment was to perform data visualization using Tableau. Tableau is a visual analytics platform that helps users connect with and understand data to solve problems. It is a popular business intelligence (BI) tool used to create interactive visualizations, dashboards, and reports from large datasets by allowing users to drag and drop elements without needing to write complex code.

The purpose of the assignment is to gain hands-on practice:

1. Understand the business and client needs
2. Load Data
3. Transform Data
4. Create Calculated Measures
5. Visualize Dashboard
6. Publish your project as part of your portfolio collection

Tasks Completed

I created an interactive dashboard for the HR manager (our client), providing both summary views for high-level insights and detailed employee records for in-depth analysis through the following steps:

Step 1: connecting to data

I opened the tableau public and connected to my csv data source.

Step 2: formatting sheets

I formatted the sheets by changing its background colour and the measurements to 1400 by 800

Step 3: calculated columns

I created bans for total employees hired, total terminated and total active. I achieved that through the following expressions:

Total hired = COUNT([Employee ID])

Screenshot:

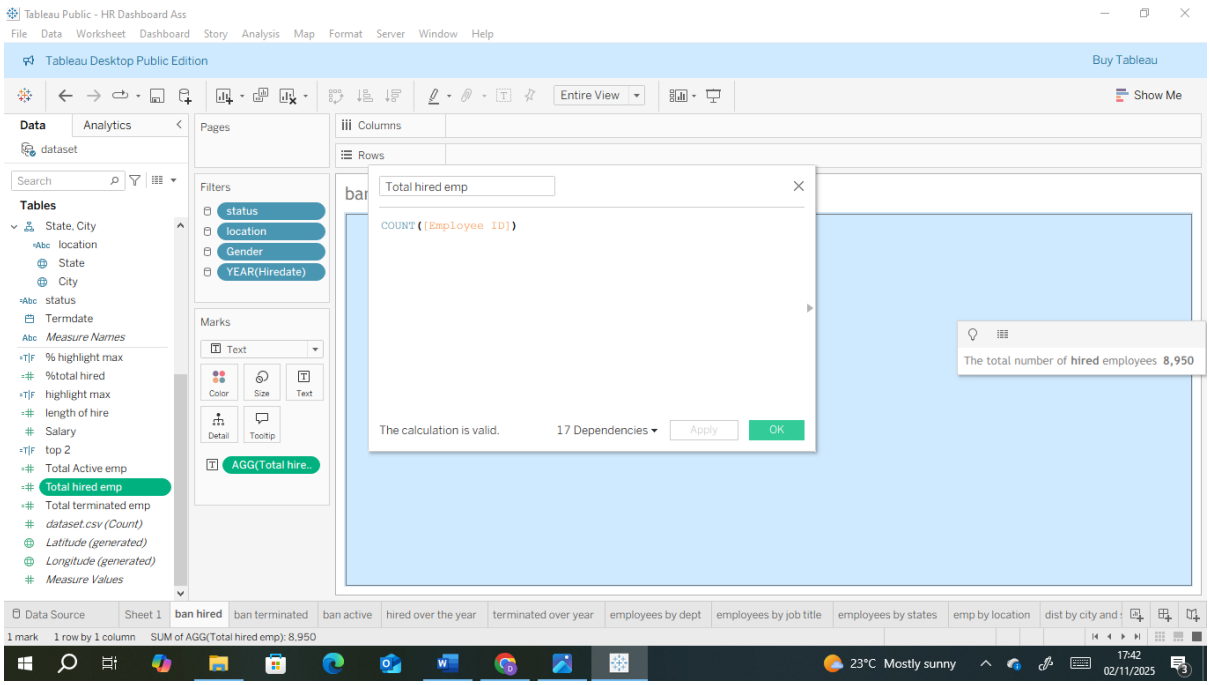

Figure 1: screenshot showing ban total hired

Total terminated = COUNT(IF NOT ISNULL([Termdate]) THEN [Employee ID] END)

Screenshot:

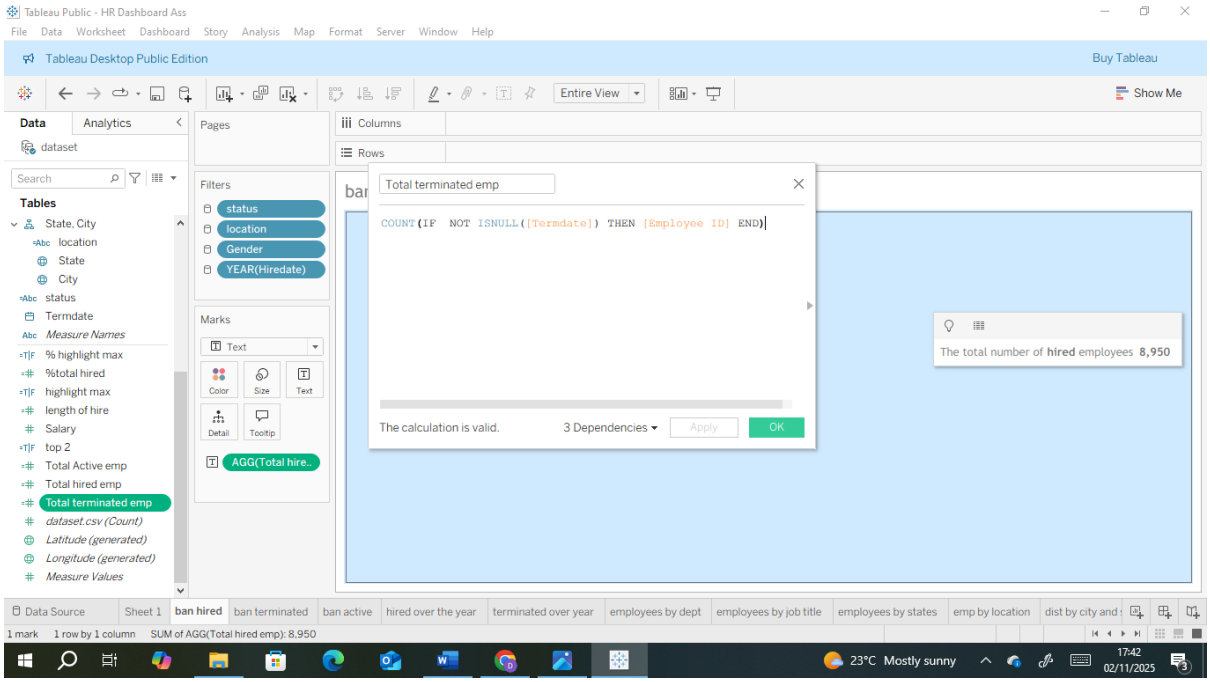

Figure 2: screenshot showing total terminated

Total active = COUNT(IF ISNULL([Termdate]) THEN [Employee ID] END)

Screenshot:

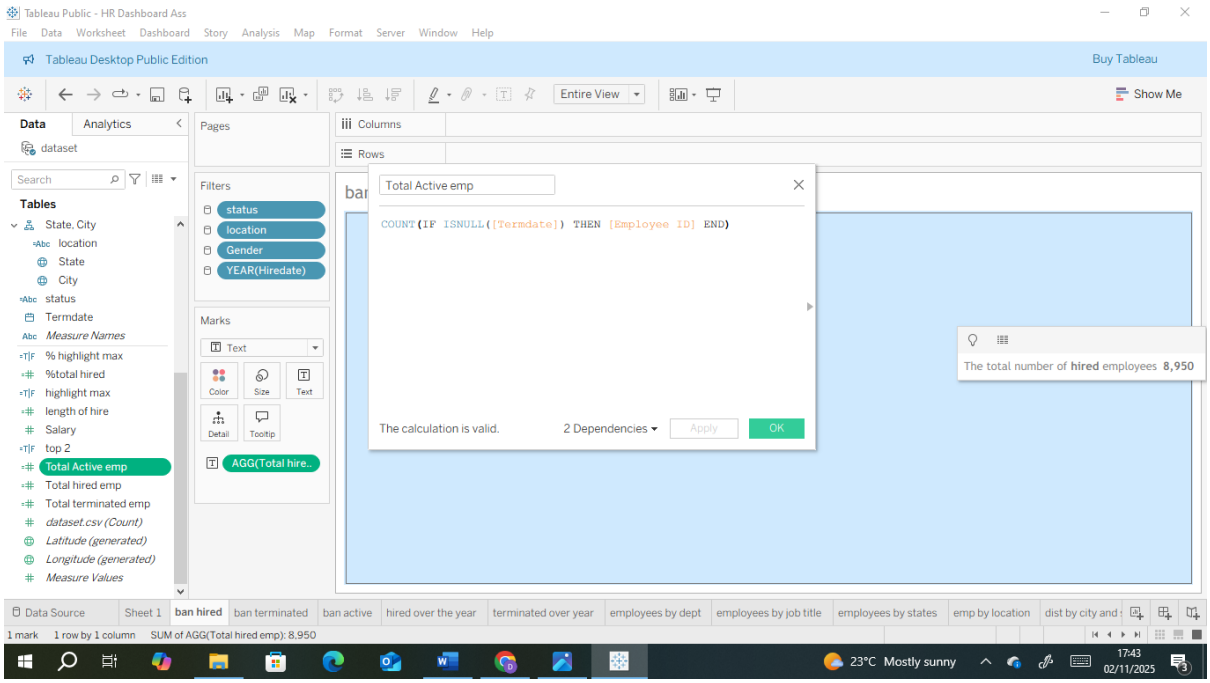

Figure 3: screenshot showing total active

Step 4: created charts for visualization

- **Overview charts:**

✓ I created both a line chart and area chart combined to show the total number of hired employees over the years.

Screenshot:

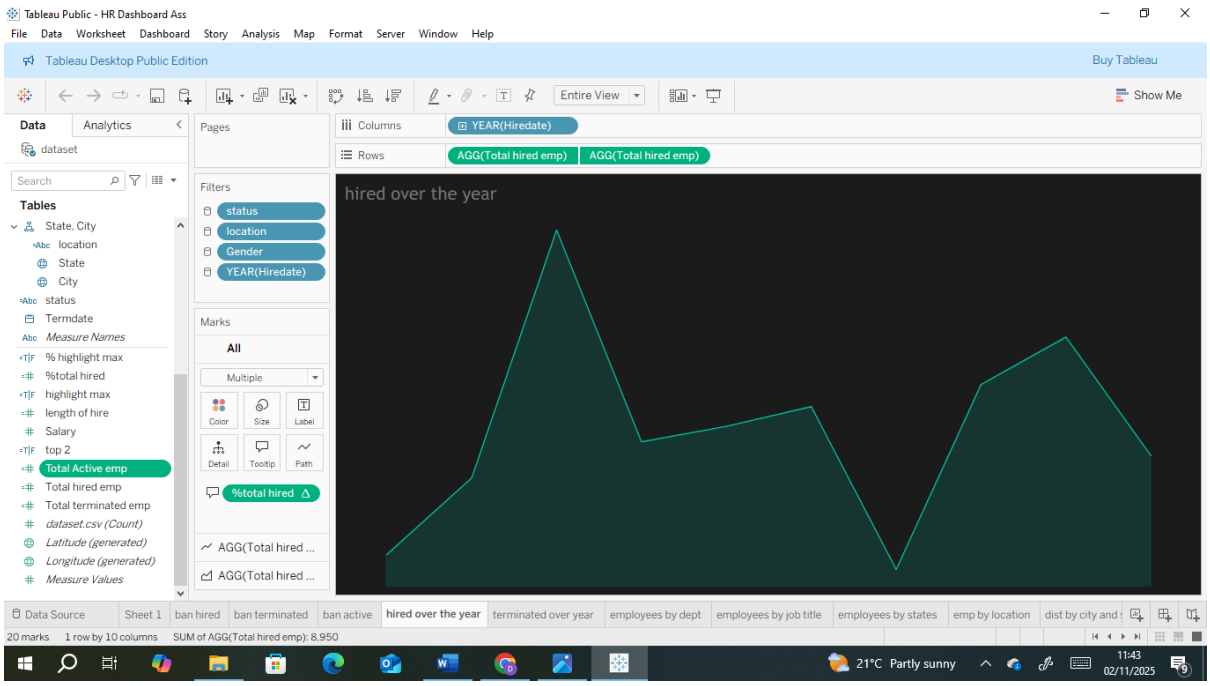

Figure 4: line and area chart showing total employees over the years

✓ I created chart for total employee over the years

Screenshot:

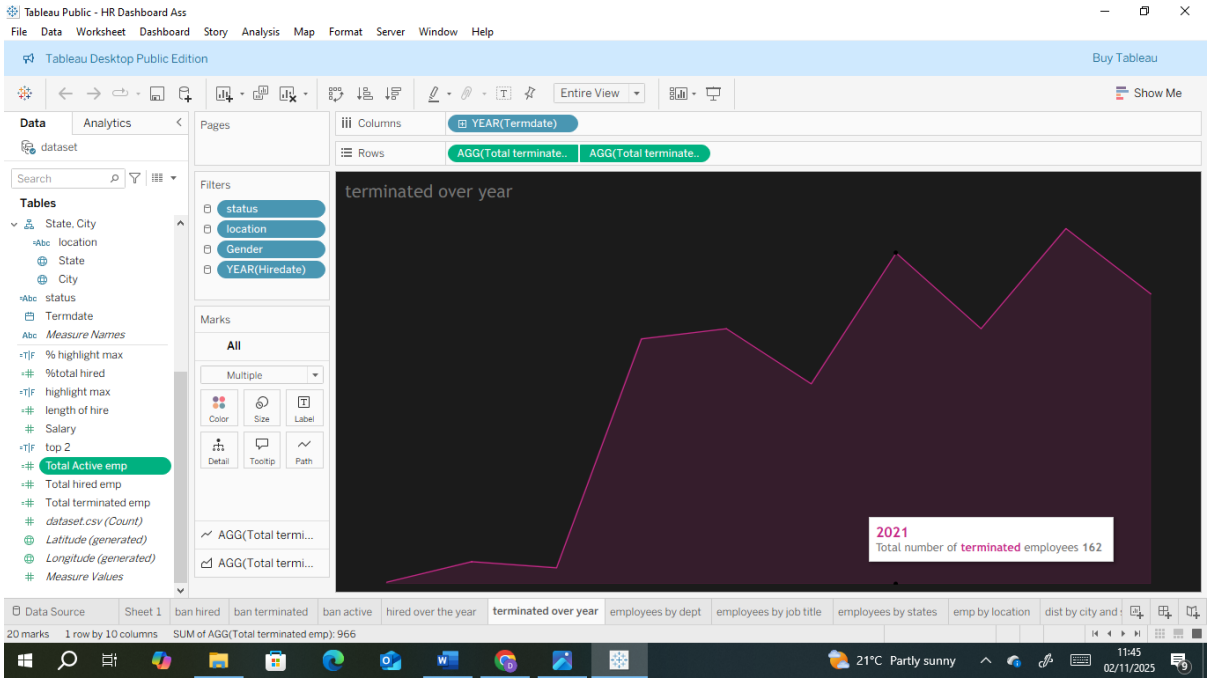

Figure 5: screenshot showing line and area chart for total terminated over the years

- ✓ I presented a breakdown of total employees by department through a bar chart.

Screenshot:

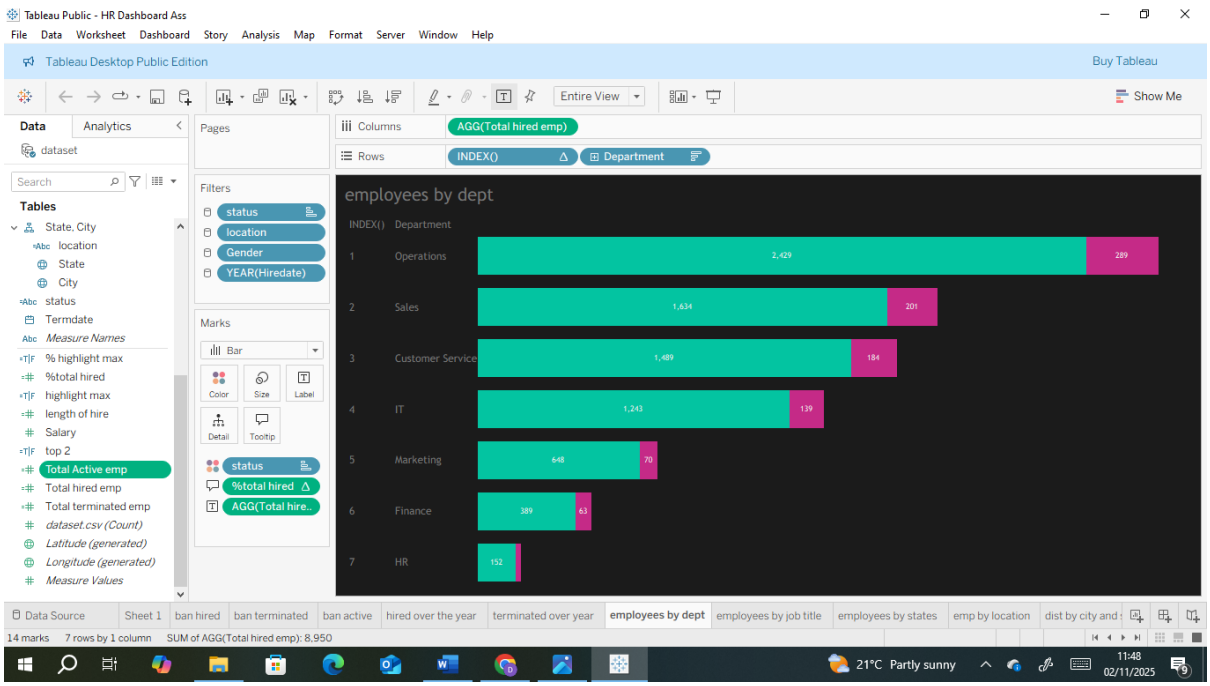

Figure 6: bar chart of employees by department

- ✓ I presented a breakdown of total employees by job title through a bar chart.

Screenshot:

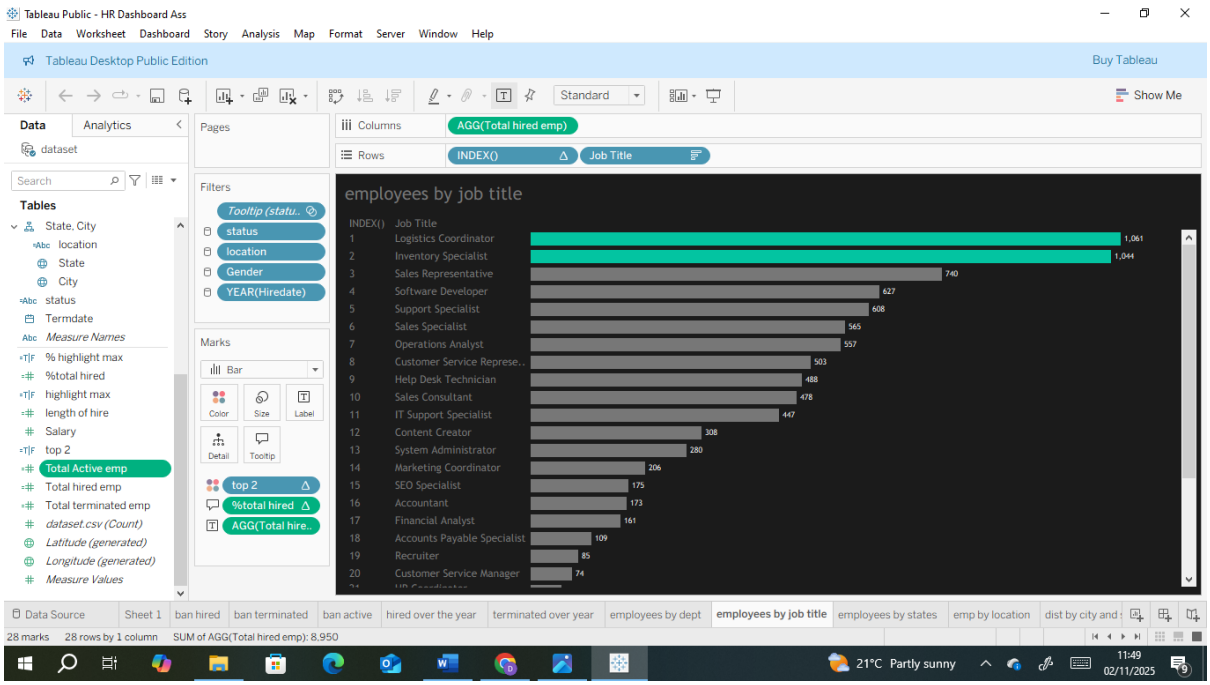

Figure 7: bar chart of employees by job title

- ✓ I compared total employees between headquarters and branches using a bar chart.

Screenshot:

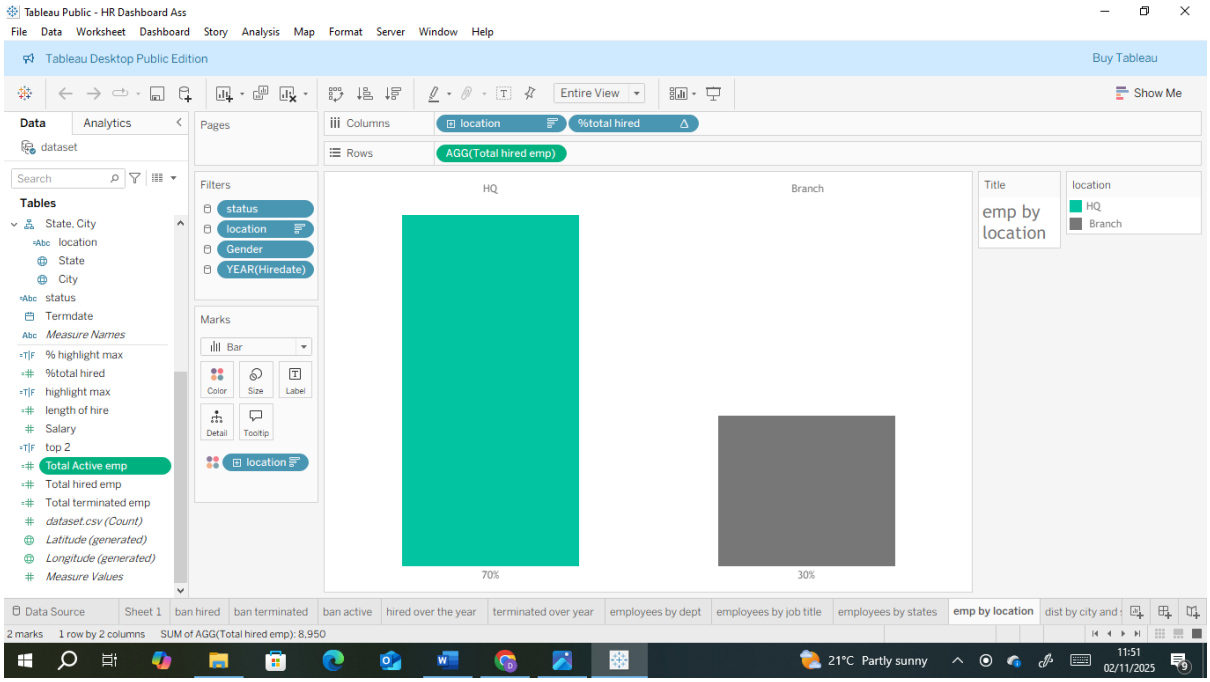

Figure 8: bar chart of employees by headquarter and branches

- ✓ I showed the distribution of employees by city and state using a map.

Screenshot:

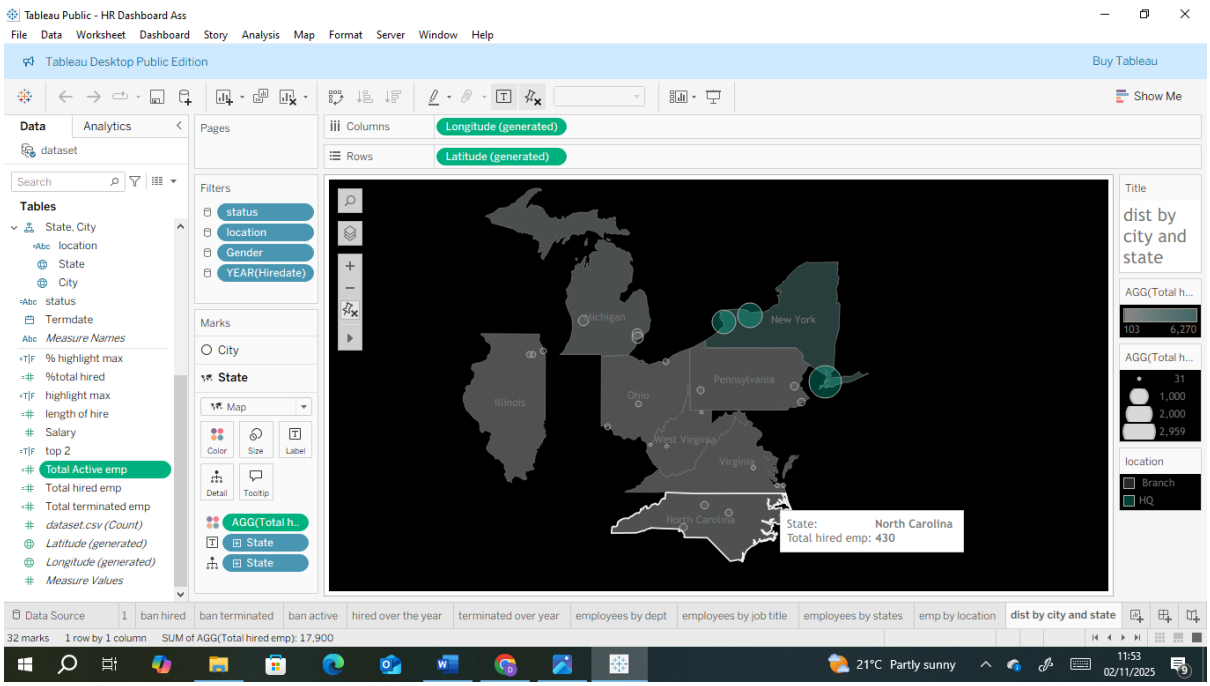

Figure 9: map showing employee distribution by state

✓ I showed the distribution of employees by state.

Screenshot:

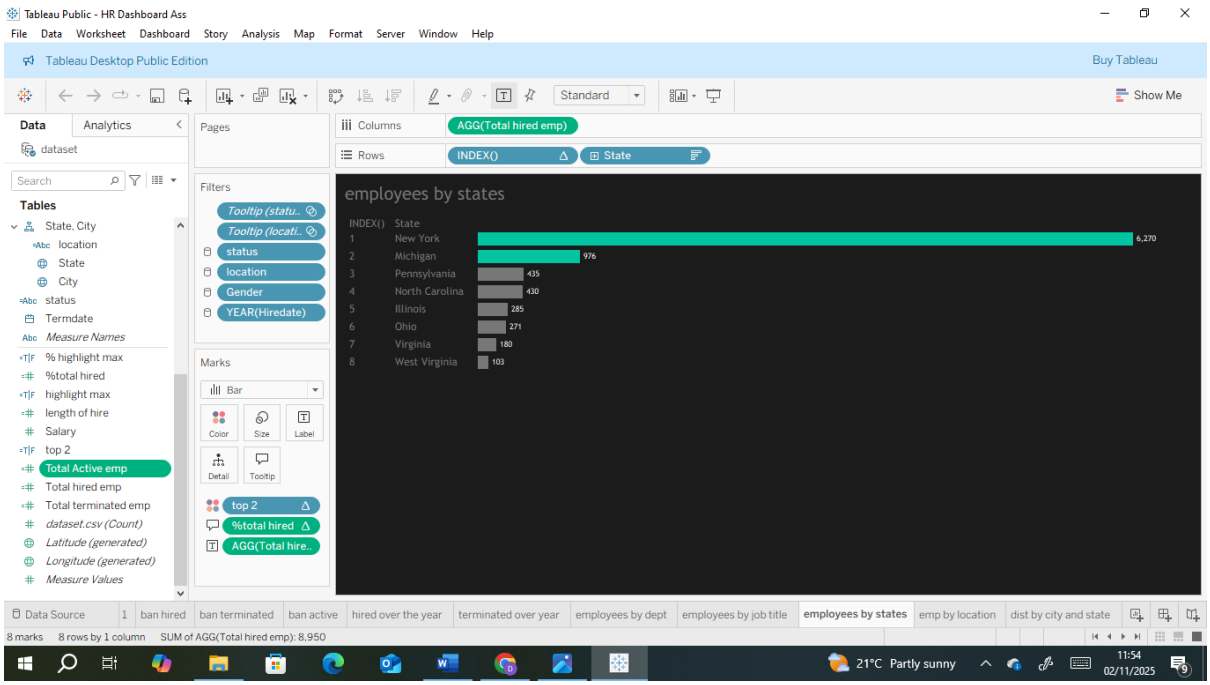

Figure 10: distribution of employees by state

- **Demographic charts:**

✓ I presented the gender ratio in the company using a pie chart.

Screenshot:

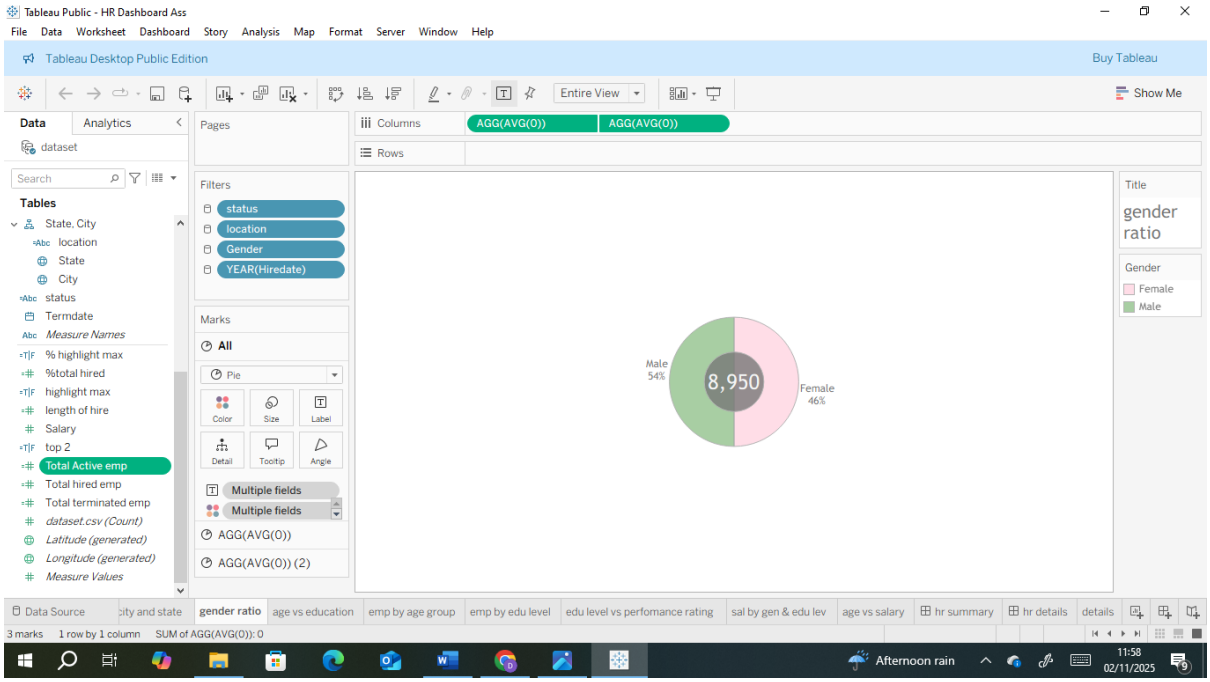

Figure 11: pie chart showing gender ratio

- ✓ I visualized the distribution of employees across age groups and group levels using a heat map

Screenshot:

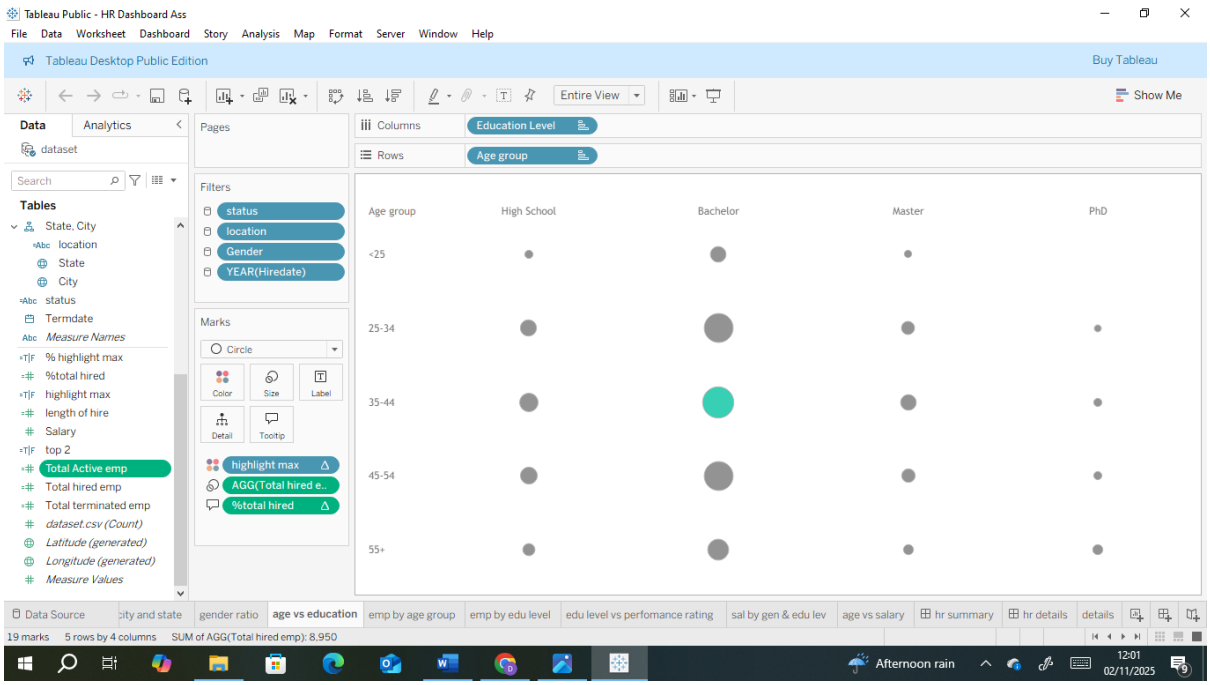

Figure 12: a heatmap to show the age groups vs group levels

- ✓ I showed the total number of employees within each age group. I first added a new measure called age group.

The expression for age group:

IF [Age] < 25 THEN '<25'

ELSEIF [Age] >=25 AND [Age] < 35 THEN '25-34'

ELSEIF [Age] >=35 AND [Age] < 45 THEN '35-44'

ELSEIF [Age] >=45 AND [Age] < 55 THEN '45-54'

ELSEIF [Age] >=55 THEN '55+'

END

Screenshot:

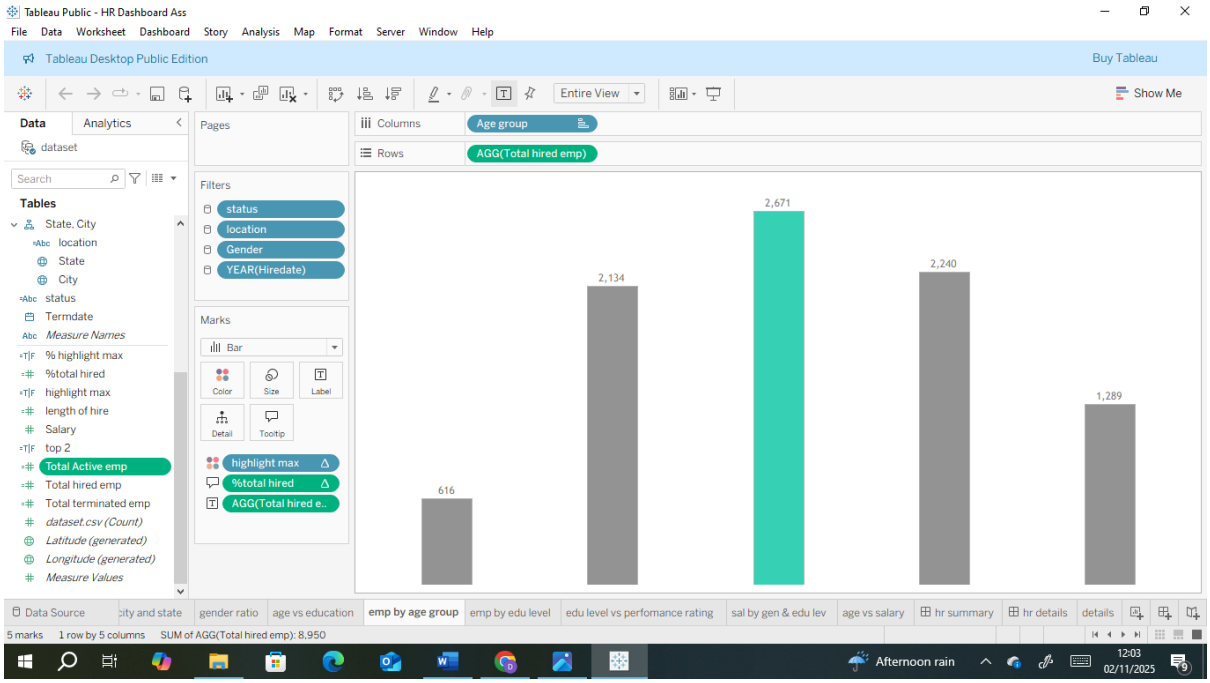

Figure 13: bar chart showing distribution of employees by age group

- ✓ I showed the total number of employees by education level.

Screenshot:

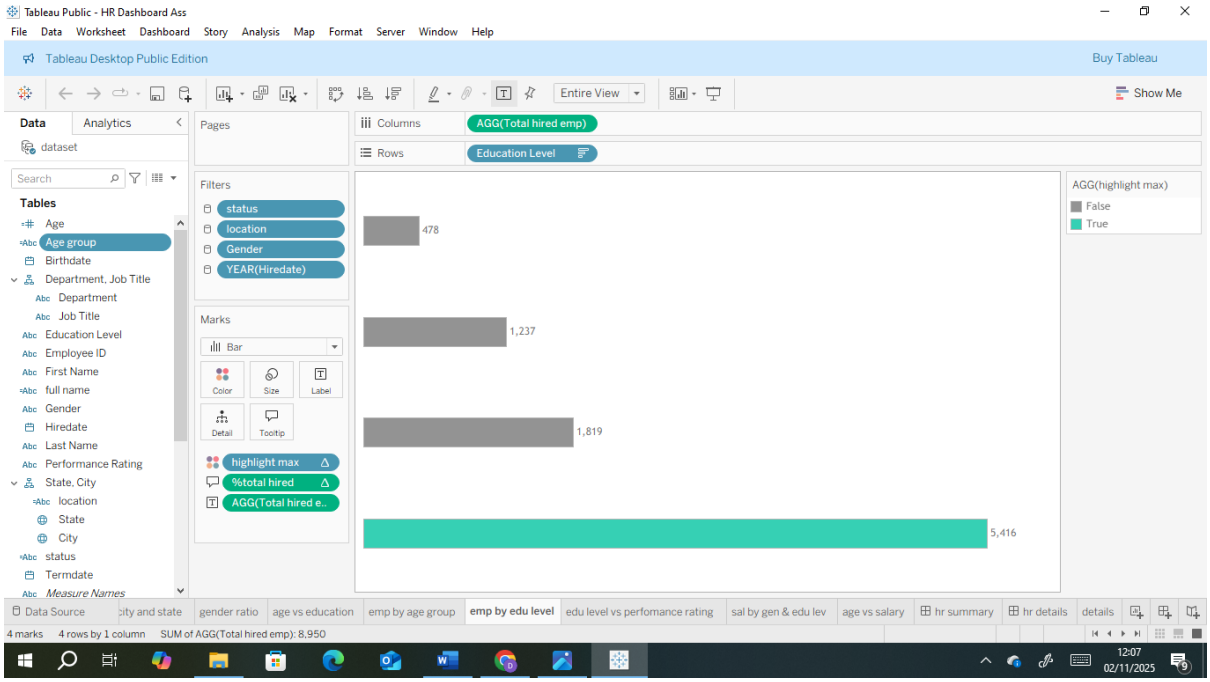

Figure 14: bar chart showing employees by education level

- ✓ I presented the corelation between employee's educational background and their performance ratings through a heat map.

Screenshot:

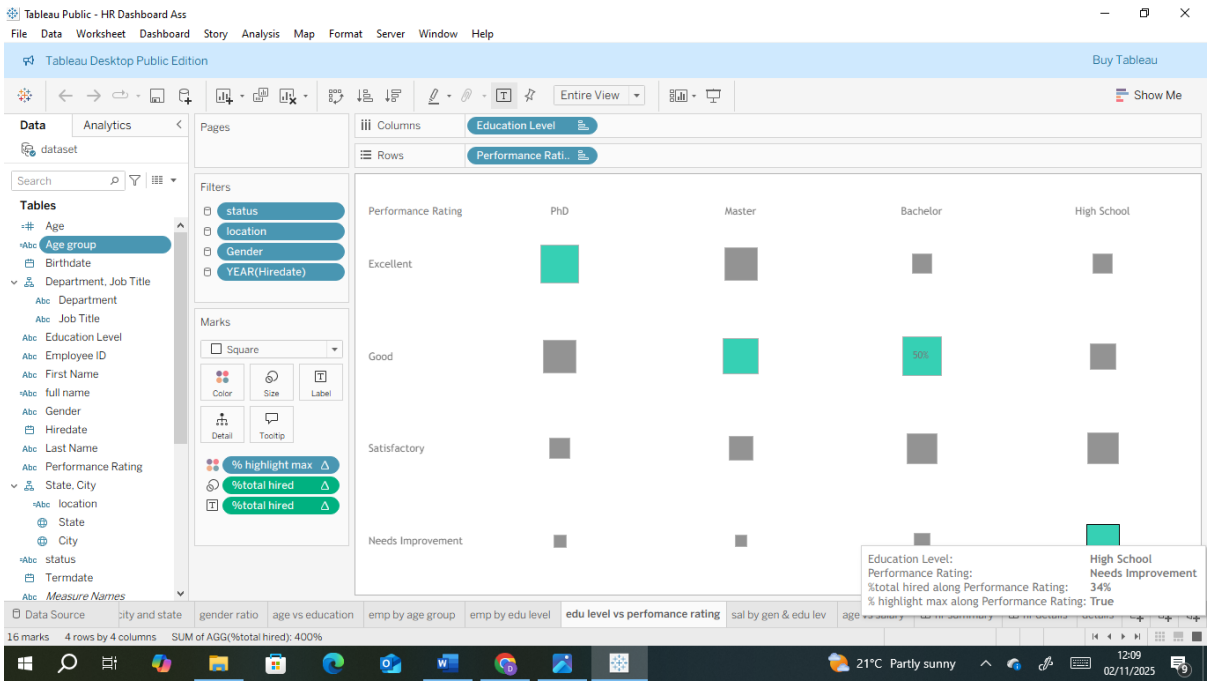

Figure 15: corelation between educational background and performance rating

- **Income analyzers:**

✓ I compared the salaries across different education levels for both genders to identify any discrepancies or patterns. I used a barbell chart.

Screenshot:

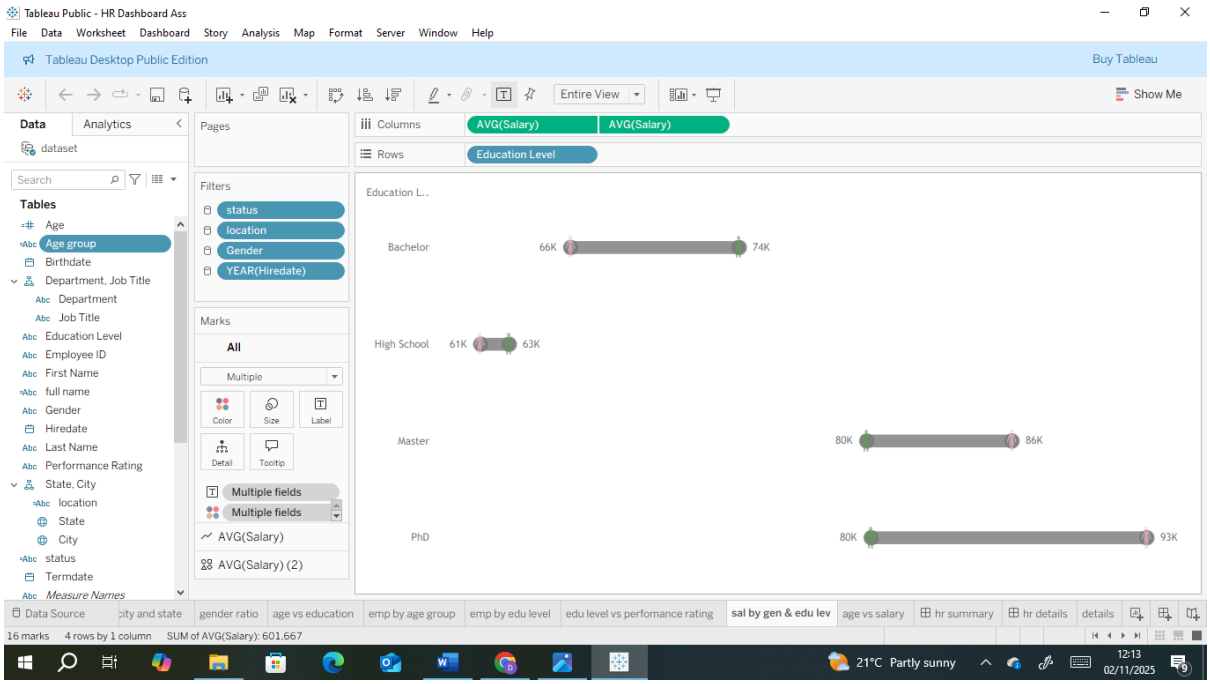

Figure 16: bar bell chart to show discrepancies in salaries across gender

✓ I presented how the age corelate with the salary for employees in each department using a scatter plot.

Screenshot:

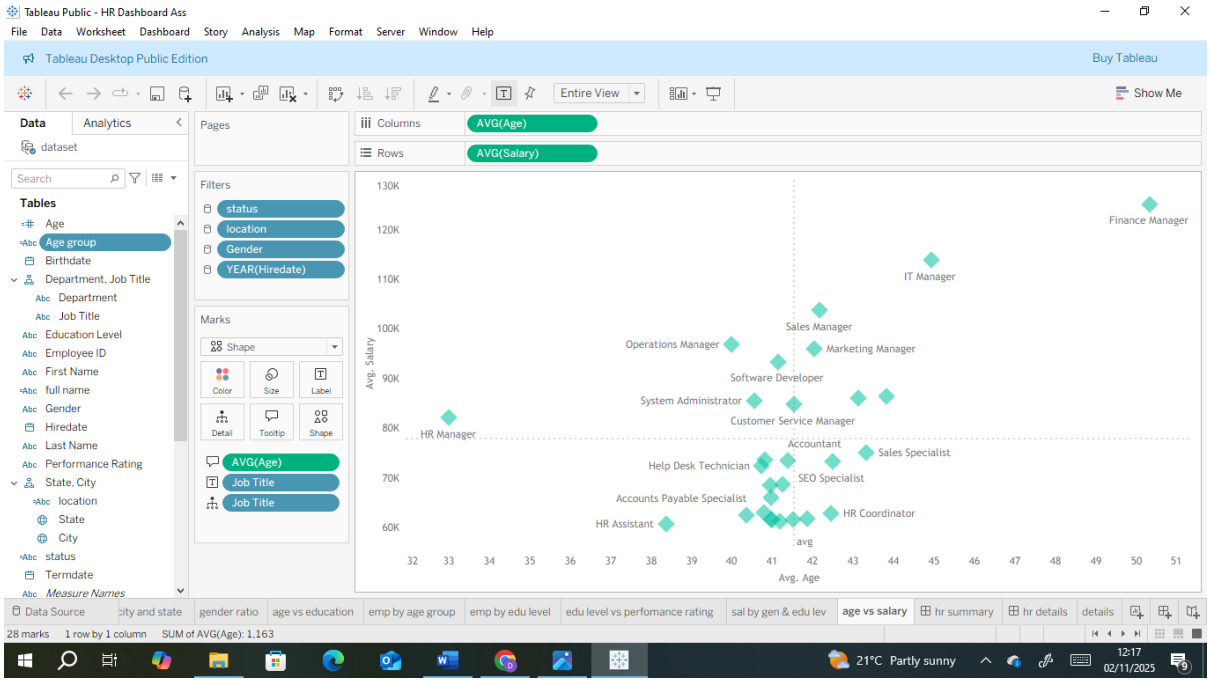

Figure 17: corelation between age and salary for each department

Step 5: Building a summary dashboard

I created an interactive dashboard that includes the overview section, demographics section and income section.

I formatted the dashboard size to 1400 by 800.

I placed the charts in the dashboards.

I did fine tuning of the dashboard to look professional.

I added background colours to the dashboard and containers.

I fixed the texts for the headings.

I added tooltips in every charts.

I added action filters and legends.

I added logos and icons.

Screenshots:

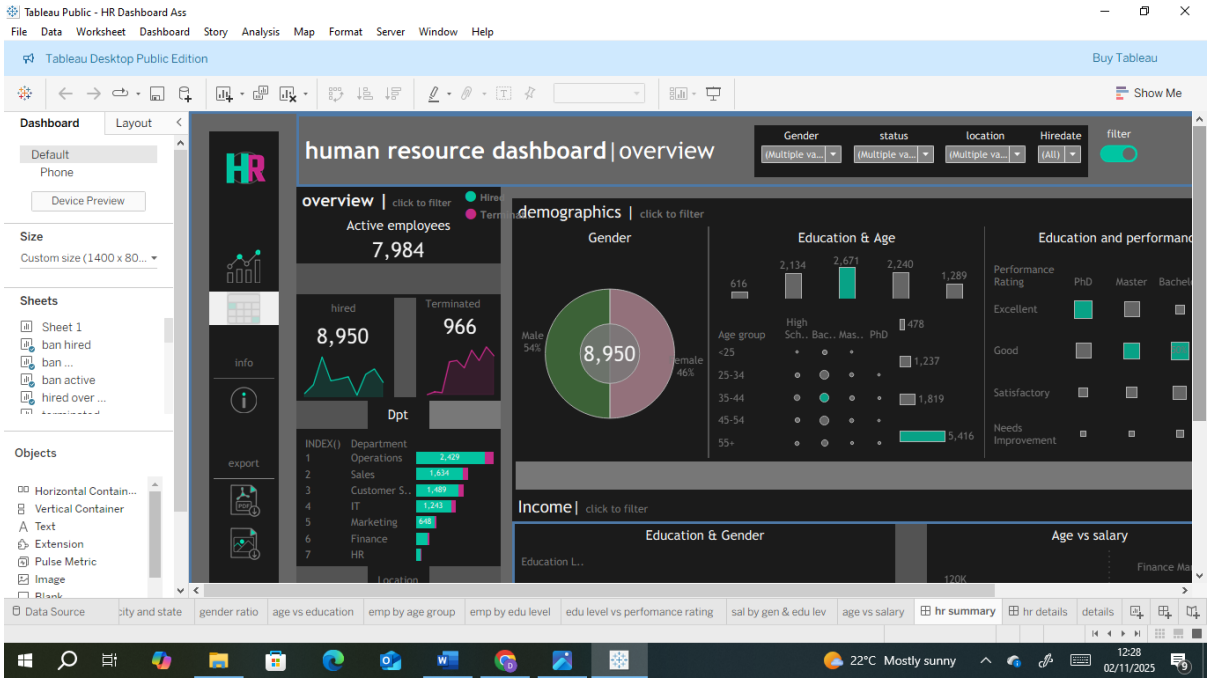

Figure 18: the summary dashboard

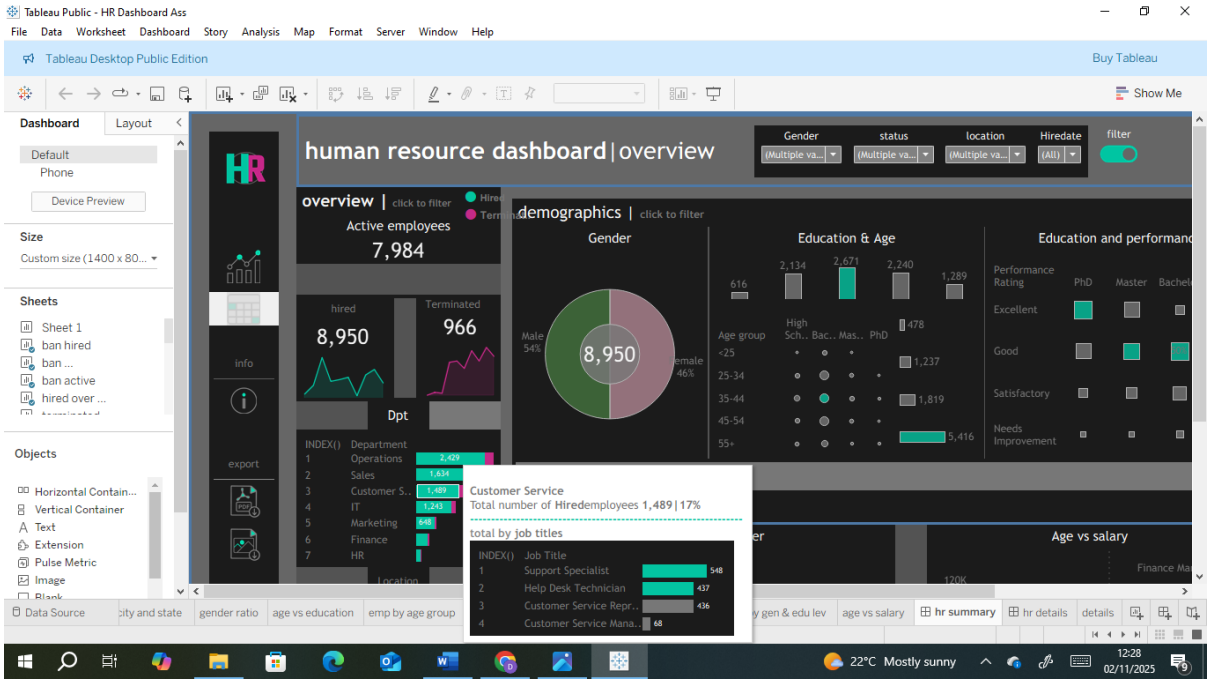

Figure 19: dashboard after adding tooltips

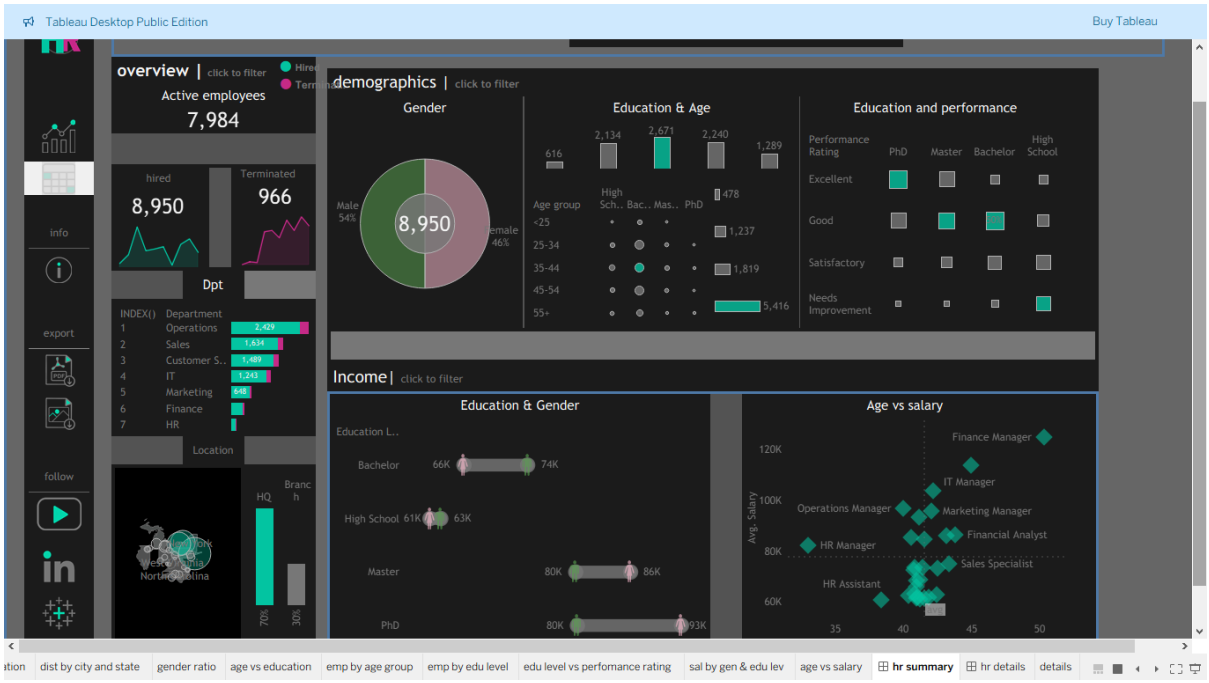

Figure 20: dashboard with filters, legends and logos

Step 6: building a dashboard for employee details

In this step I provided a list of all employees and their necessary information. I provided a way to filter the list based on the available columns.

- ✓ I first created the details sheet that includes columns for all necessary information.

Screenshot:

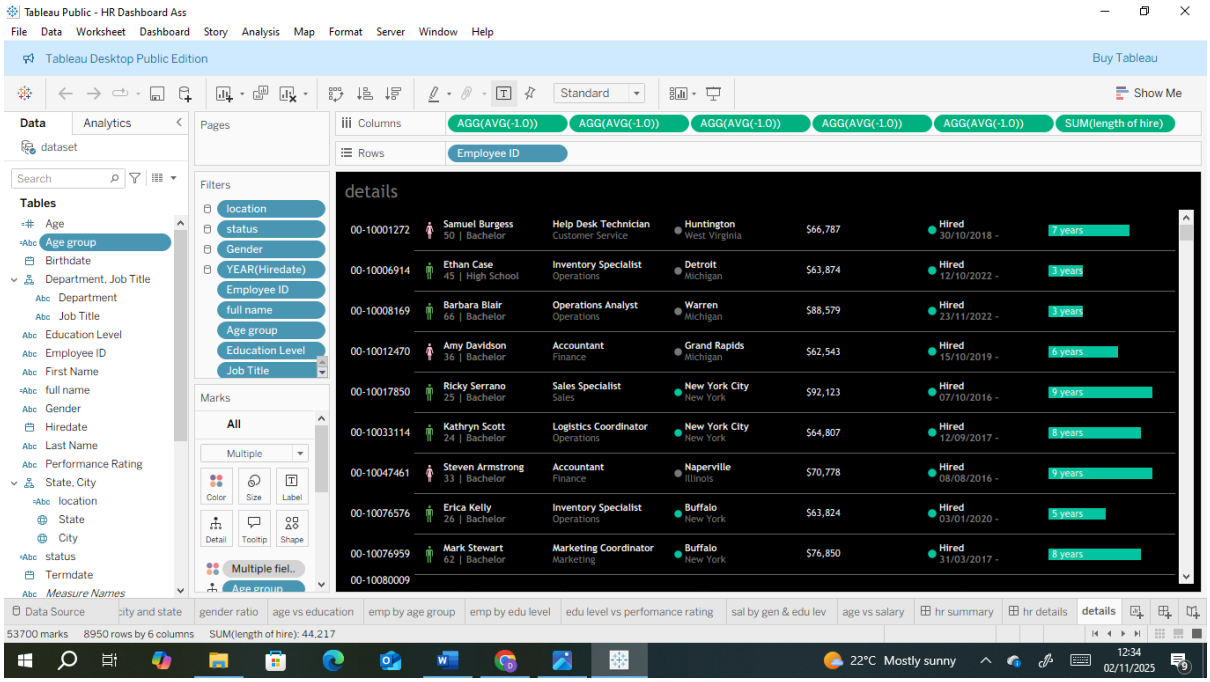

Figure 21: sheet showing employee details

To create the column for length of hire, I created a new measure for length of hire with the following expression:

IF ISNULL([Termdate])

THEN DATEDIFF('year', [Hiredate], TODAY())

ELSE DATEDIFF('year' , [Hiredate], [Termdate])

END

Screenshot:

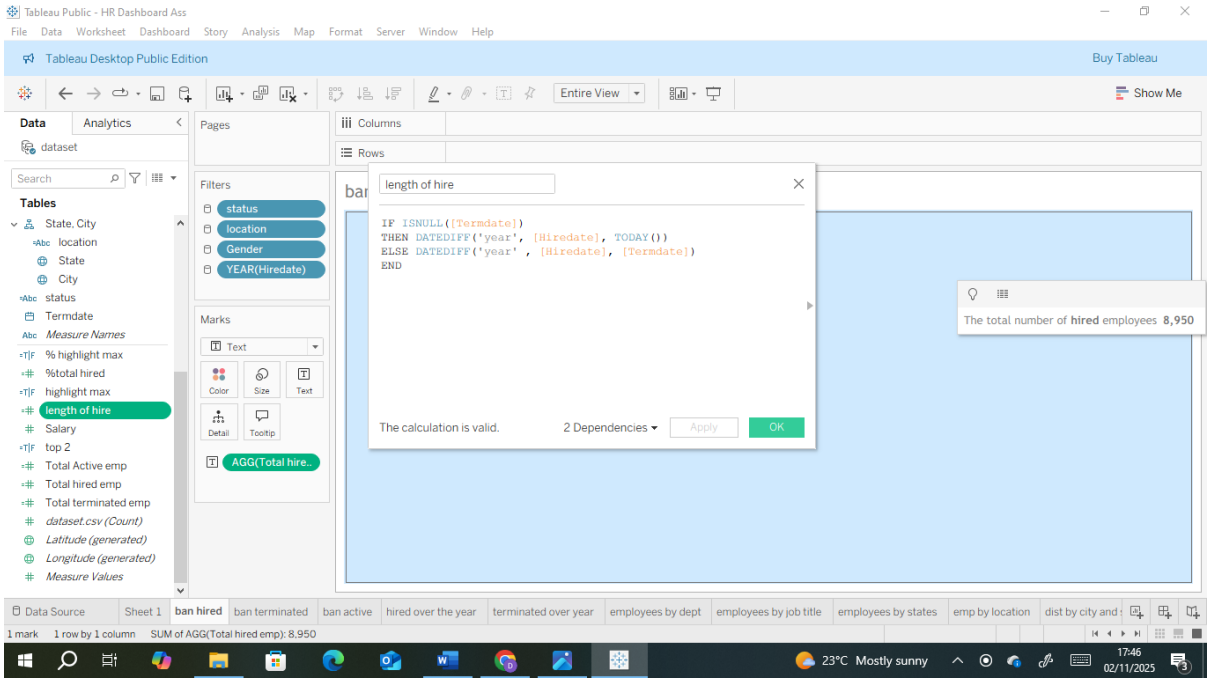

Figure 22: length of hire measure added

- ✓ I created the employee dashboard and added the necessary filters

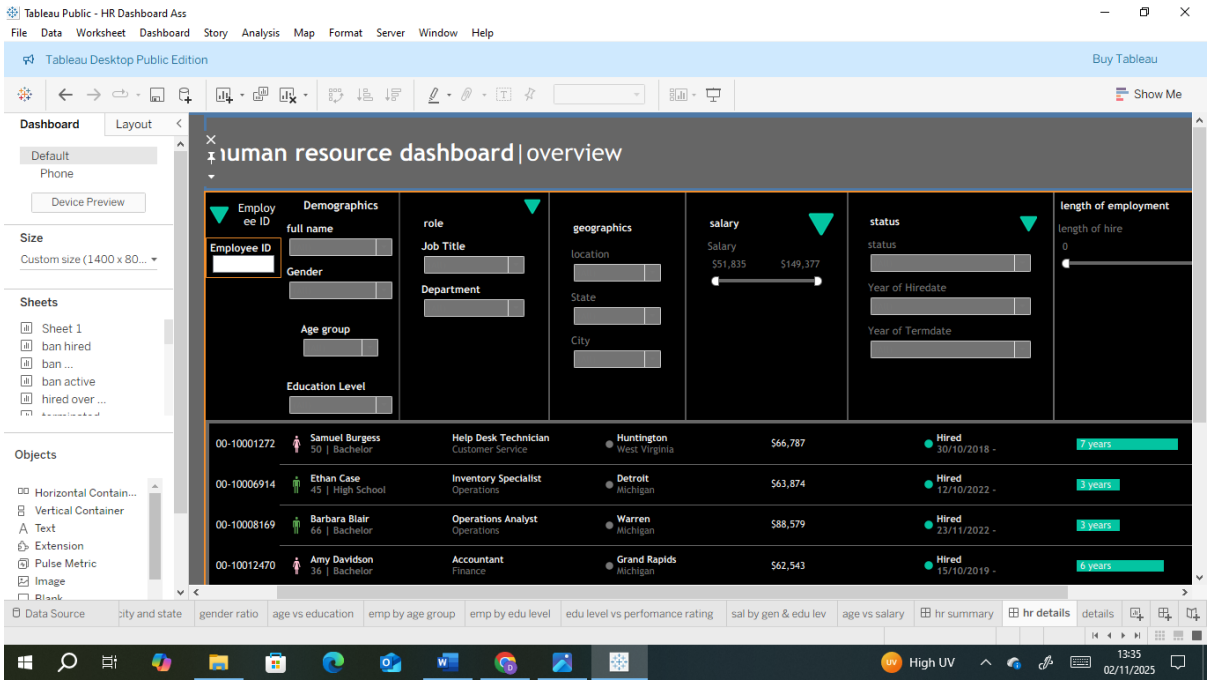

Figure 23: dashboard for employee details and the necessary filters

Link to code

Link to Code:

https://drive.google.com/drive/folders/12Dsk61nd9p7DEmPWb0NsTs4QVgiGrOKP?usp=sharing

Conclusion

In conclusion, this Tableau tutorial provided a comprehensive understanding of how to transform raw data into meaningful visual insights. I learned how to clean, organize, and visualize data using different chart types, dashboards, and interactive features. The hands-on exercises demonstrated the power of Tableau in simplifying complex datasets and enhancing data-driven decision-making. Through this experience, I developed a stronger appreciation for the role of visual analytics in communicating patterns, trends, and relationships effectively. Tableau's user-friendly interface and flexibility make it an essential tool for data analysts and professionals across various fields. Overall, this assignment strengthened my confidence in using Tableau to explore, analyse, and present data in a clear and impactful way.