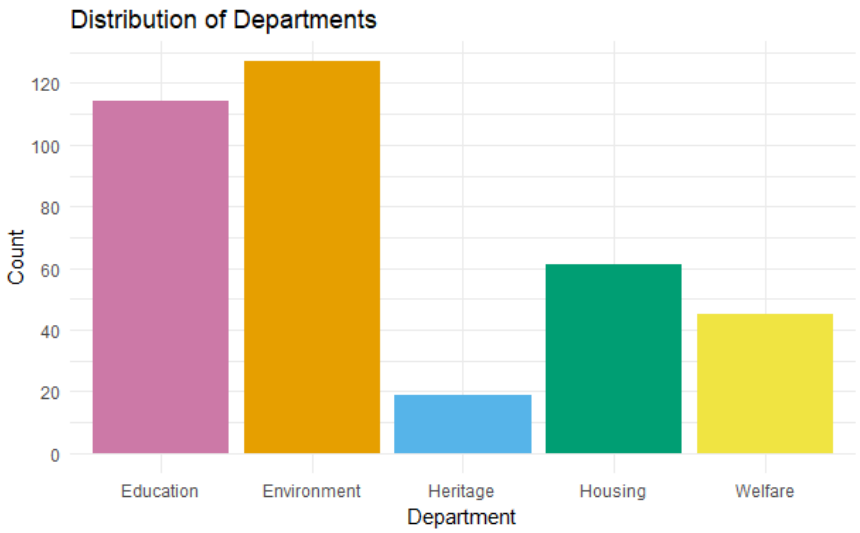
Data Analytics and Visualisation using R

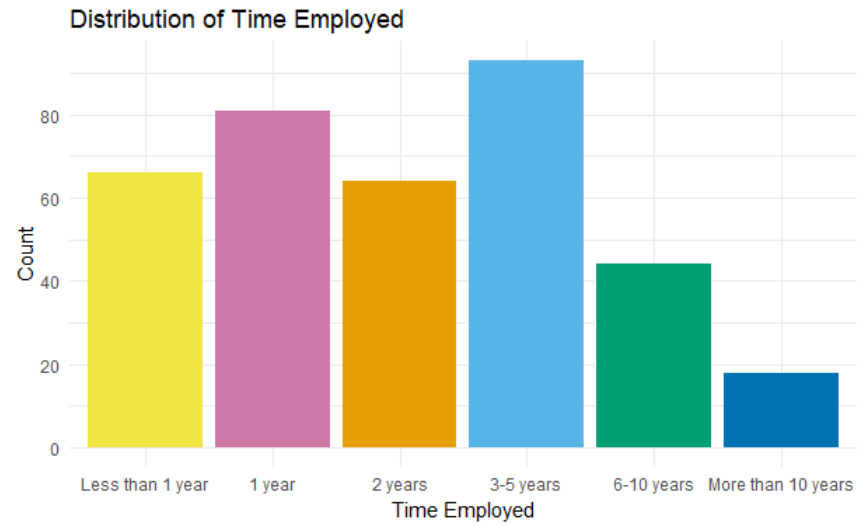
Dataset

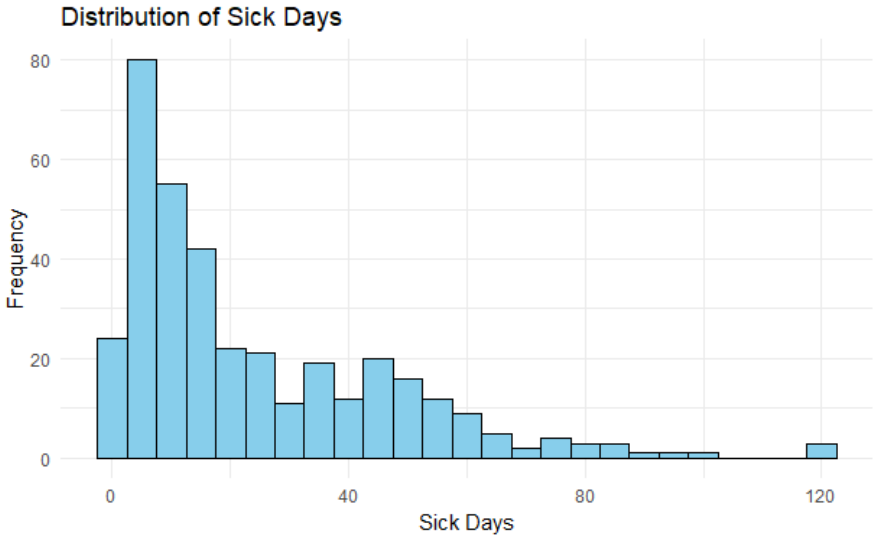
The dataset shows staff information from departments of local government. There are 366 staff members included, grouped by departments and the length of time that they have been employed.

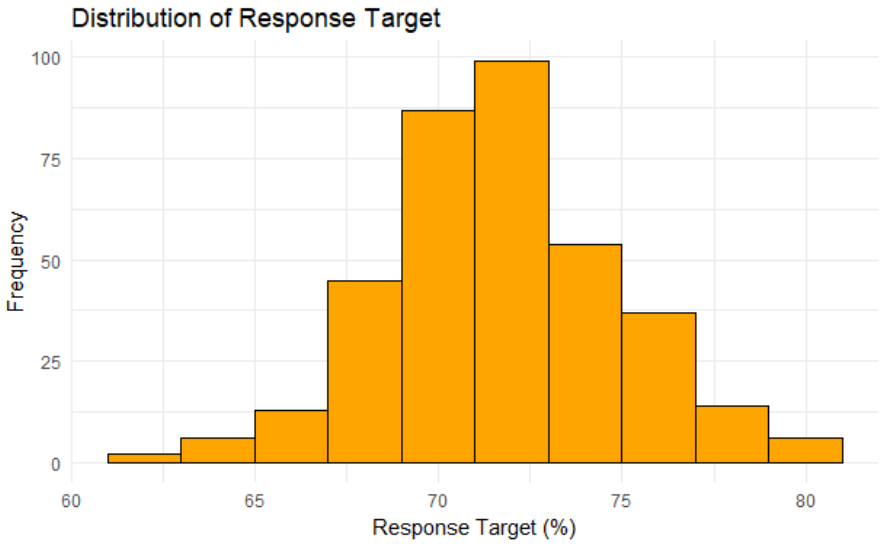
- **Dataset Features**
  - department: Department employee works in. **[Categorical]**
  - time: How long employee has been employed. **[Categorical]**
  - sick_days: No of sick days recorded during last two years of employment. **[Quantitative]**
  - response_target: Percentage of enquiries meeting 48hr response target. **[Quantitative]**
  - ftime: How long employee has been employed on in a full-time role. **[Categorical]**

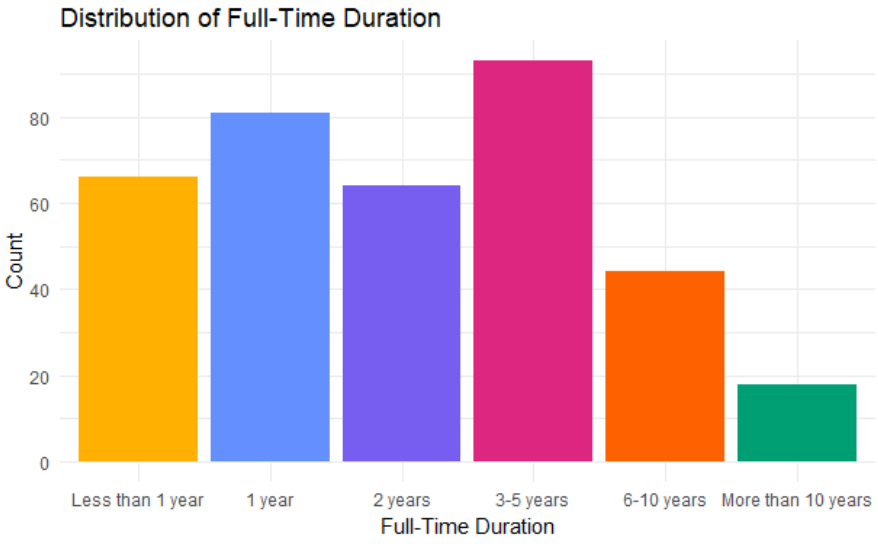
1. **Understanding the Distribution of Staff Data**




Distribution of Time Employed
Count
0
20
40
60
80
Less than 1 year
1 year
2 years
3-5 years
6-10 years
More than 10 years
Time Employed


Distribution of Sick Days
Frequency
0
20
40
60
80
0
40
80
120
Sick Days


Distribution of Response Target
Frequency
0
25
50
75
100
60
65
70
75
80
Response Target (%)



2. Summary of Key Statistics for Staff Data

	Sick Days	Response Target
Minimum	0	61.80
Average	24.10	71.68
Middle Value (Median)	15	71.60
Lower Middle (Q1)	7	69.70
Upper Middle (Q3)	37.75	73.75
Middle Range (IQR)	30.75	4.05
Maximum	120	80.40

Sick Days:

- The least number of sick days any employee has taken in the past two years is 0.
- On average, employees take about 24 sick days over two years.
- Half of the employees took fewer than 15 sick days, and the other half took more.
- Most employees took between 7 and 37.75 sick days which is where the middle 50% of employees fall.
- The employee with the highest sick days took 120 days off over the two years.

Response Target:

- The lowest percentage of time an employee responded to inquiries within 48 hours is 61.80%.
- On average, employees respond to about 72% of inquiries within 48 hours.

- Half of the employees met the target more than 71.6% of the time, and the other half did less.
- Most employees answered between 69.7% and 73.75% of inquiries within the 48-hour target.
- The best employee met the target 80.4% of the time.

Department	Count	Percentage
Education	114	31.2%
Environment	127	34.7%
Heritage	19	5.2%
Housing	61	16.7%
Welfare	45	12.3%

- **Education** has the highest number of employees, with 114 employees, making up about **31%** of the total.
- The **Heritage** department has the fewest employees, with just **19**, or about **5%** of the total.

Time Employed (time)	Count	Percentage
Less than 1 year	66	18.0%
1 year	81	22.1%
2 years	64	17.5%
3-5 years	93	25.4%
6-10 years	44	12.0%
More than 10 years	18	4.9%

- **3-5 years** is the group with the most employees, making up **25%** of the total.
- Fewer employees have been with the company for **More than 10 years (5%)**.

Fulltime Duration (ftime)	Count	Percentage
Less than 1 year	66	18.0%
1 year	81	22.1%
2 years	64	17.5%
3-5 years	93	25.4%
6-10 years	44	12.0%
More than 10 years	18	4.9%

- **3-5 years** has the highest number of full-time employees, making up **25%** of the total.
- Smaller proportions of employees have been in a full-time role for more than **10 years (5%)**.

3. **How Response Target Performance Varies by Department and Employment Time**

- Average Response Target - shows the typical response rate and helps understand the trend in the dataset.
- Middle Response Target – shows the middle value of the response rates and helps understand the distribution of the dataset.
- Range of the most Common Response Target - shows the performance of the middle 50% of employees and helps understand how spread out it is.

Response Target by Department:

Department	Average Response Target (Mean)	Middle Response Target (Median)	Range of the most Common Response Targets (1st Quartile – 3rd Quartile)
Education	71.63	71.4	69.5 - 74.20
Environment	71.39	71.3	69.6 - 72.80
Heritage	71.11	71.4	68.2 - 72.95
Housing	72.12	72.2	69.8 - 74.40
Welfare	72.24	71.8	70.1 - 75.40

Interpretation:

- **All departments perform quite similarly** when it comes to responding to inquiries within 48 hours.
- **Welfare** and **Housing** departments have a slight edge, with slightly better average response rates and higher performance from some staff members.
- The performance of staff within each department is fairly consistent, with most staff members performing similarly to each other.

Response Target by Employment Time:

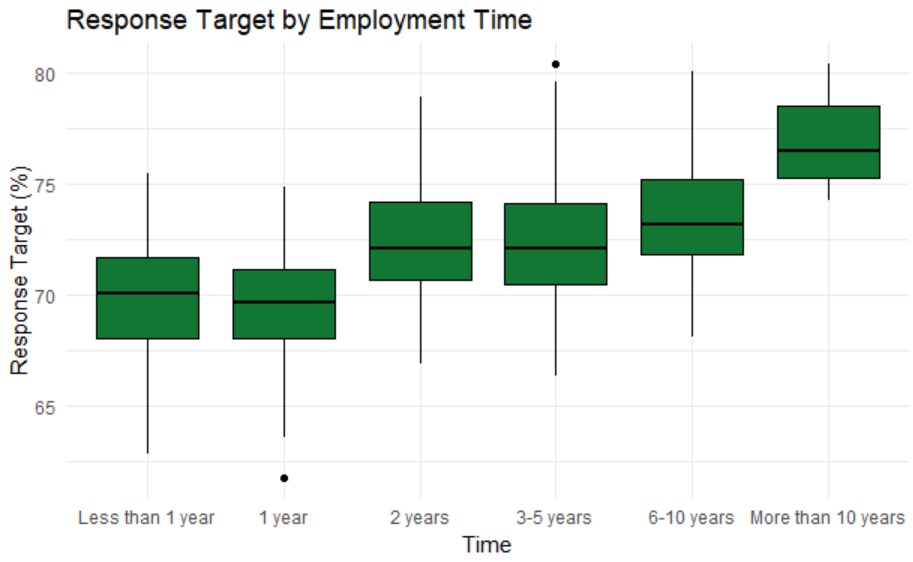
Employment Time (time)	Average Response Target (Mean)	Middle Response Target (Median)	Range of the most Common Response Targets (1st Quartile – 3rd Quartile)
Less than 1 year	69.78	70.1	68.05 - 71.70

1 year	69.49	69.7	68.05 - 71.20
2 years	72.53	72.1	70.70 - 74.22
3-5 years	72.46	72.1	70.47 - 74.15
6-10 years	73.46	73.2	71.87 - 75.20
More than 10 years	76.81	76.5	75.30 - 78.55

Interpretation:

- **Employees with more experience** (especially more than 10 years) tend to have the **best performance** in responding to inquiries within 48 hours, with higher average response rates and a higher middle value.
- **Employees with less experience** (less than 1 year or 1 year) generally have lower response rates.
- **As employees gain more experience**, their ability to meet the 48-hour response target improves, and their performance becomes more consistent, as seen from the increasing median and narrowing range of most common response targets.

4. **<u>How Response Performance Improves with Time Employed</u>**



The box plot shows how the response rates (the percentage of inquiries answered within 48 hours) are distributed across different groups of employees based on their years of experience in a full-time role. It is effective for understanding not only the average performance but also the consistency and variation in response rates across different experience groups.

- The **line in the middle of each box** represents the Middle Response Target, which is the middle value of all the response rates for each group. This gives us an idea of what a "typical" employee in that group is doing in terms of response times.

- The **box** shows the range in which most of the employees' response rates fall, showing how spread out the data is.
- The **lines extending from the box** show the **minimum** and **maximum** response rates for each group excluding extreme values.
- The **dots outside the lines** represent employees whose response rates are significantly higher or lower than most others.

Interpretation:

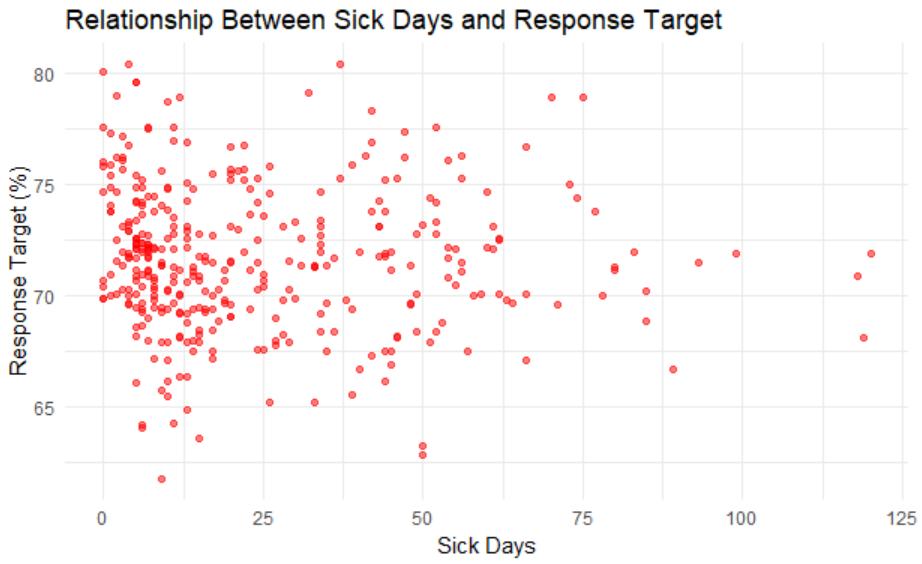
As employees spend more time in the organization, their performance seems to improve. For example, employees who have been working for **more than 10 years** typically respond to enquiries faster, meeting the target **76.5% of the time**. On the other hand, employees with **less than 1 year** of experience meet the target only about **70% of the time**. This shows that with more experience, employees tend to perform better.

Some **unusual cases** in the data were noticed:

- In the **1-year group**, there is one person who performed much worse than the others (an outlier).
- In the **3-5 years group**, there's one person who performed much better than everyone else (another outlier).

The trend suggests that staying longer in the job helps employees respond to enquiries more consistently.

5. **<u>Relationship Between Sick Days and Response Target</u>**



A scatter plot is ideal for showing relationships between two variables (in this case, sick days and response target). It helps to quickly see patterns, such as trends or clusters, without needing complex explanations.

Interpretation:

- Employees with **fewer sick days** tend to have response targets that vary widely but are generally **higher**.
- As the number of **sick days increases**, response targets **slightly decrease**, showing that employees with more sick days may struggle to meet the response target consistently.
- However, there is no clear and strong pattern, as the dots are quite scattered meaning that, sick days might have a slight effect on performance, but the relationship is not strong or definitive.