

# Graduation from DEPI Final Project Data Analysis Track

Congratulations on completing the DEPI Final Project Data Analysis Track! This presentation will take you through the 4 unique projects our team of 4 members undertook, exploring different data sets and honing our data analysis skills. Let's dive in and celebrate our achievements!

# Project Overview

## 4 Projects

Our team completed 4 data analysis projects, each with a unique data set.

## Different Data Sets

The data sets covered a range of topics, from spreadsheets to Python and SQL to Power BI.

## Comprehensive Skills

Through these projects, we developed a diverse set of data analysis skills and tools.

# Spreadsheet Project

## 1 Data Exploration

### Overview

- **Dataset:** Water service data from 2000 to 2022, divided into rural, urban, and all-area categories.

## 2 Data Preprocessing

- 1-no duplicated rows
- 2-there is no missing value in some column and we filled it with average
- 3-max service=100 , min service=0
- 4-no outliers from Q1,Q2,IQR and this show in box plot

## 3 Impactful Visualizations

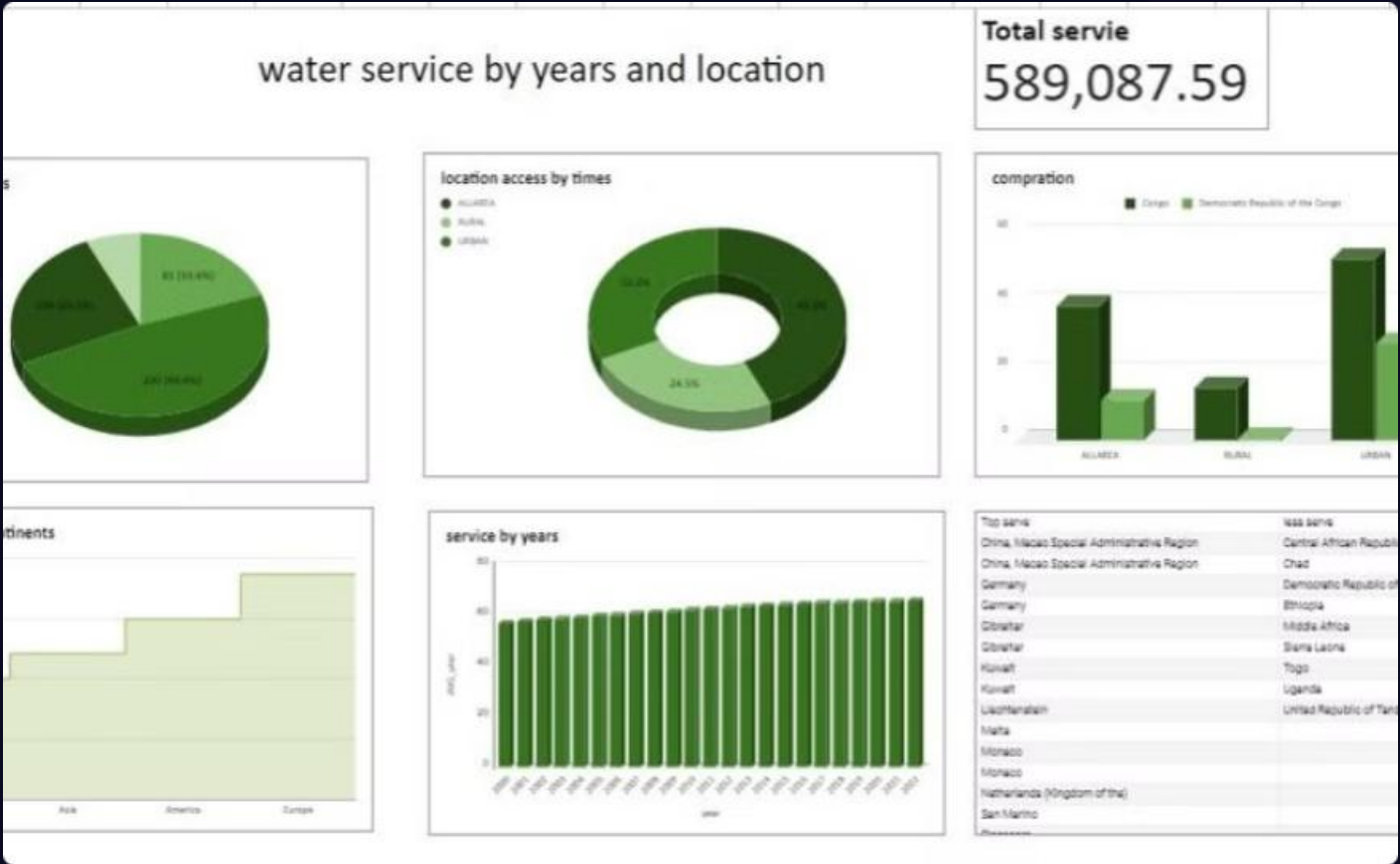
- Countries Categorized by Service Levels
- Key Findings
- Pivot Analysis:

# Insights and Visualizations for Spreadsheet Project

63.14	FALSE	FALSE	FALSE	upper	upper	Algeria	FALSE	1,406.15
80.11	FALSE	FALSE	FALSE	upper	upper	Algeria	FALSE	1,842.62
86.39	FALSE	FALSE	FALSE	upper	upper	American	FALSE	1,086.91
90.64	True	FALSE	FALSE	FALSE	top	Andorra	FALSE	2,084.72
82.90	FALSE	FALSE	FALSE	upper	upper	Armenia	FALSE	1,906.79
65.31	FALSE	FALSE	FALSE	upper	upper	Asia	FALSE	1,502.16
56.92	FALSE	FALSE	FALSE	upper	upper	Asia	FALSE	1,309.11
76.16	FALSE	FALSE	FALSE	upper	upper	Asia	FALSE	1,751.78
99.18	True	FALSE	FALSE	FALSE	top	Australia	FALSE	2,283.19
99.30	True	FALSE	FALSE	FALSE	top	Australia	FALSE	2,283.82
98.25	True	FALSE	FALSE	FALSE	top	Austria	FALSE	2,259.76
65.41	FALSE	FALSE	FALSE	upper	upper	Azerbaijan	FALSE	1,504.50
59.76	FALSE	FALSE	FALSE	upper	upper	Azerbaijan	FALSE	1,374.43
89.19	FALSE	FALSE	FALSE	upper	upper	Azerbaijan	FALSE	2,053.42
98.91	True	FALSE	FALSE	FALSE	top	Bahrain	FALSE	2,274.95
56.20	FALSE	FALSE	FALSE	upper	upper	Bangladesh	FALSE	1,292.53
60.80	FALSE	FALSE	FALSE	upper	upper	Bangladesh	FALSE	1,398.45
45.64	FALSE	FALSE	FALSE	upper	upper	Bangladesh	FALSE	1,049.82
88.83	FALSE	FALSE	FALSE	upper	upper	Belarus	FALSE	2,043.11
99.66	True	FALSE	FALSE	FALSE	top	Belgium	FALSE	2,292.28
39.61	FALSE	FALSE	under	FALSE	under	Bhutan	FALSE	911.06
38.28	FALSE	FALSE	under	FALSE	under	Bhutan	FALSE	880.40
44.75	FALSE	FALSE	FALSE	upper	upper	Bhutan	FALSE	1,029.23
92.11	True	FALSE	FALSE	FALSE	top	Bonaire, E	FALSE	2,118.42
88.22	FALSE	FALSE	FALSE	upper	upper	Bosnia and	FALSE	2,029.16
79.00	FALSE	FALSE	FALSE	upper	upper	Botswana	FALSE	1,817.03
80.58	FALSE	FALSE	FALSE	upper	upper	Brazil	FALSE	1,853.41
59.27	FALSE	FALSE	FALSE	upper	upper	Brazil	FALSE	1,363.21
84.62	FALSE	FALSE	FALSE	upper	upper	Brazil	FALSE	1,946.35
94.32	True	FALSE	FALSE	FALSE	top	Bulgaria	FALSE	2,169.29
22.75	FALSE	FALSE	under	FALSE	under	Cambodia	FALSE	523.27
15.01	FALSE	FALSE	under	FALSE	under	Cambodia	FALSE	345.14
51.61	FALSE	FALSE	FALSE	upper	upper	Cambodia	FALSE	1,187.08
98.60	True	FALSE	FALSE	FALSE	top	Canada	FALSE	2,267.70

Total servie

589,087.59



- Added a column to categorize countries by service levels:
- 90: Very good service
- 50-90: Good service, needs improvement
- 10-50: Low service, requires improvement
- <10: Worst service, needs significant work
- 17 countries have excellent service (>99.5%), and 9 countries have the worst service (<5%).
- Congo vs. Democratic Republic of Congo: Both have low service, but Democratic Congo has the worst.
- Continental Comparison: Europe offers the best service, while Africa has the worst.
- Pivot Tables created to show average service by GeoAreaName and TimePeriod.

## Dashboard

Here's a concise summary for the requested analysis:

- Line Chart: Show the yearly increase in service levels.
- Pie Chart: Display service levels categorized into Top, Upper Average, Under Average, and Less.
- Pie Chart: Compare rural vs urban vs all area access over time.
- Area Chart: Compare service levels across continents.

# Python Project

1

## Data Overview

**Dataset:** Football players' attributes (personal details, performance metrics, and physical stats).

2

## Data Cleaning

- Addressed missing values, outliers, and duplicates (52 removed from ~18,000 records).
- Median used for 'Age' and 'Overall' to handle valid outliers.

3

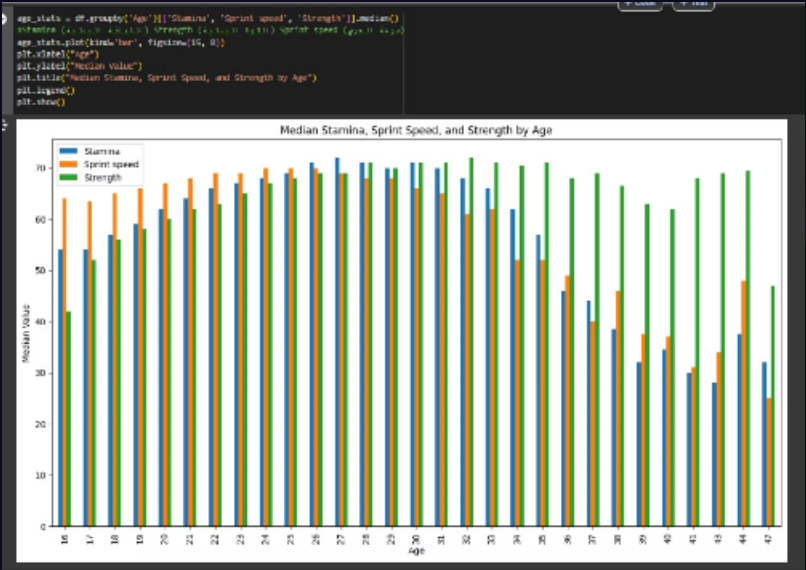
## Insights Analysis

- Distribution by nationality, age, and overall rating.
- Peak performance age: 25–39 years.
- Goalkeepers exhibit valid outliers in age and overall ratings.



# Insights and Visualizations for Python Project

Visuals showing distribution and key insights.



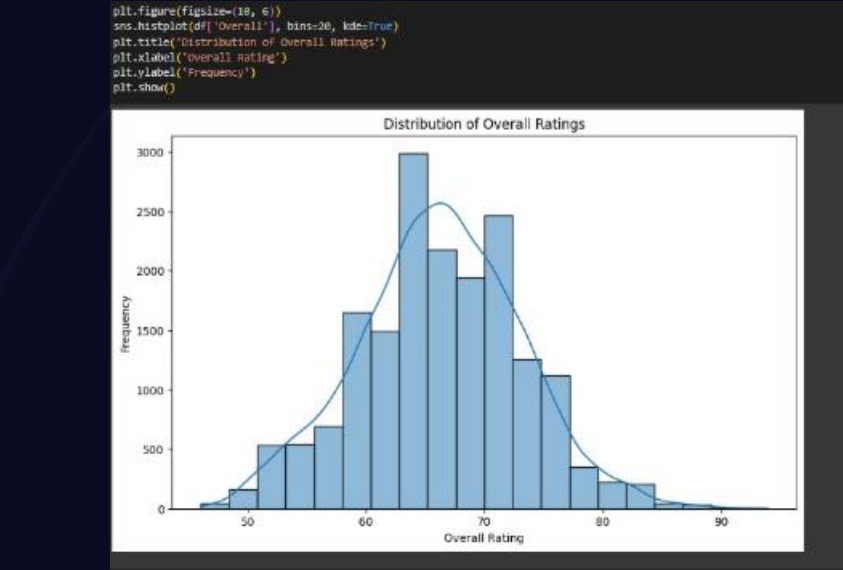
As age increases, stamina and sprint speed decrease, while strength shows the opposite trend

What are the key attributes that differentiate top-rated players from lower-rated ones?

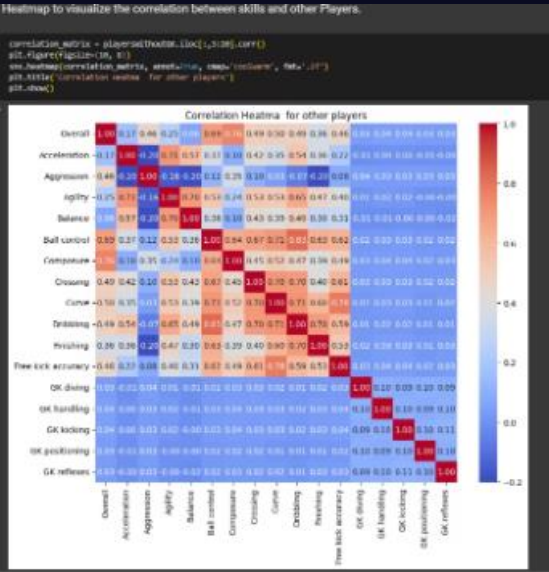
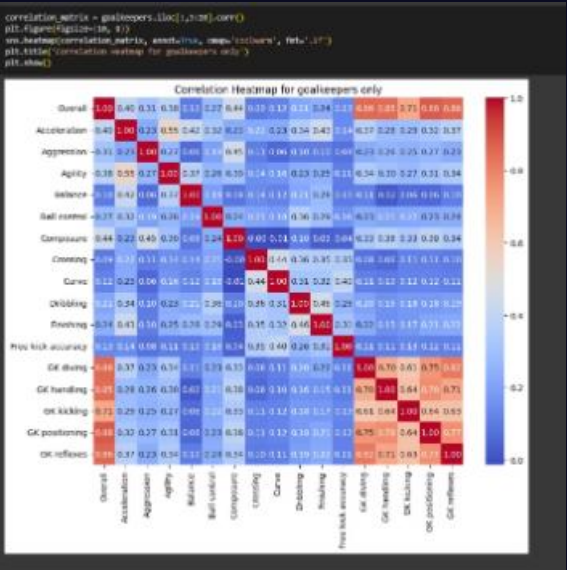
```
top_players = df.groupby('Overall Rating').agg(['max', 'min', 'std']).reset_index()
top_players = top_players[top_players['Overall Rating'] > 70]
```

Age	Overall	Acceleration	Aggression	Agility	Balance	Ball control	Composure	Crossing	Curve	Shot power	Sliding tackle	Sprint speed	Stamina	Standing tackle	Strength	Weak foot	Work rate	Wt
20	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
21	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
22	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
23	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
24	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
25	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
26	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
27	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
28	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
29	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
30	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
31	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
32	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
33	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
34	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
35	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
36	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
37	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
38	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
39	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000

Following the numbers, we conclude that the most significant attribute affecting players is age, regardless of whether it is too low or too high



Distribution by nationality, age, and overall rating.



the last two heatmaps clarify: Goalkeepers retire later than other players because their overall rating relies more on specific skills that don't decline as quickly with age.

# Python with SQL Project

1

## Data Overview

Movie genres, trends in production, revenue, and audience preferences.

2

## Data Preprocessing

Dataset analyzed for trends from 1991 to 2015.

3

## Data Analysis

Conducted advanced analytics and modeling using Python's scientific computing libraries.

### Visuals

- Bar chart, scatter plot, and line chart illustrating key trends and insights.

# Insights and Visualizations for Python with SQL Project

1

## Bar Chart:

- Most released genres: Drama, Comedy, Thriller, Action, Romance.
- Least released: TV Movies, Foreign Films, Westerns, Documentaries.

2

## Line Chart:

- Movie production trends peaked in 2015, with a decline due to streaming platforms.

3

## Scatter Plot:

- Positive correlation between budgets and revenues.
- Suggested strategies for budget allocation in popular genres.



# Power BI project

## Data Overview

Maji Ndongo Water Project Dataset.

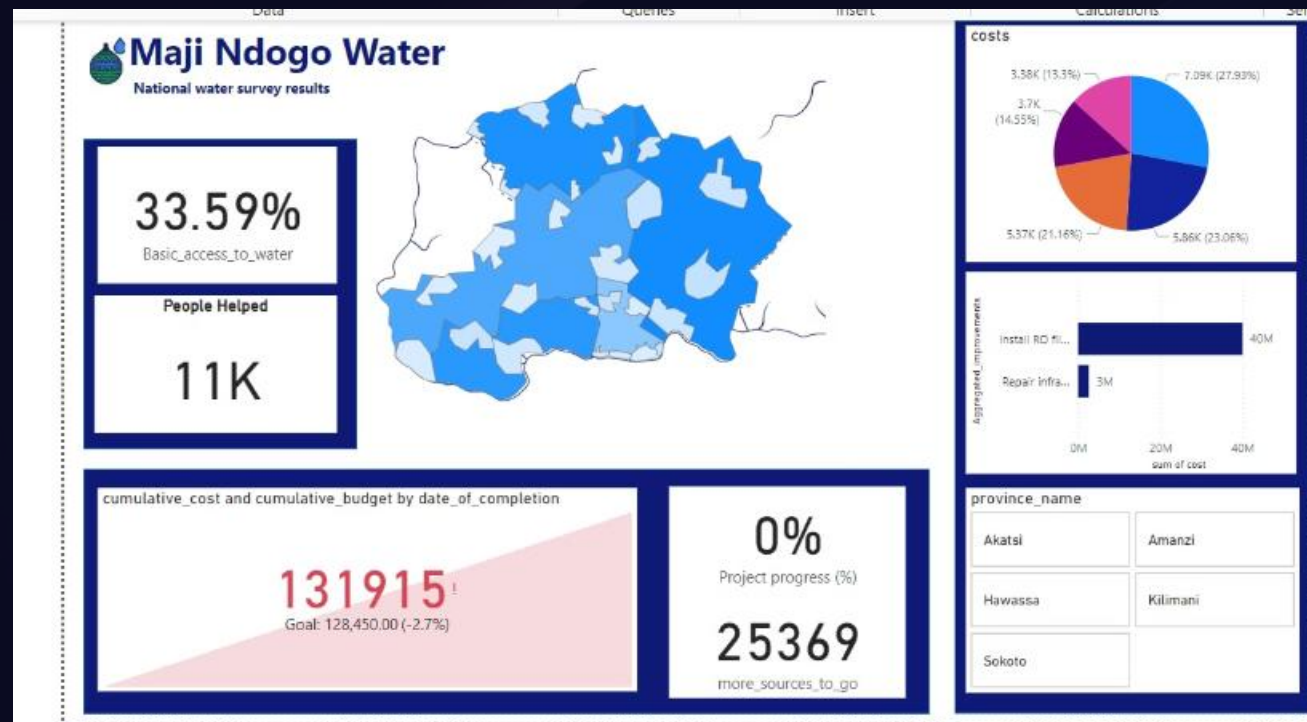
## Data Preprocessing

- Addressed missing dates and cost discrepancies.
- Verified table relationships (e.g., project and supplier schedules).

## Insights

- Cost analysis:
- Rural project costs nearly double urban ones.
- Sokoto faced the highest cost overruns due to poor infrastructure.
- Water access improvement: 34% to 48% in one year.
- Supplier performance comparison:
- Rural suppliers (e.g., mbs605) had higher costs due to challenging conditions.

# Power BI Dashboard



# Thank You for Your Attention!

## Team Members:

- Ayat Hesham
- Fatma Ahmed
- Khaled Abdullah
- Heba Reda

## Acknowledgment:

"We are proud to have completed this journey as a team, exploring diverse datasets and tools in data analysis. Each project reflects our dedication to uncovering insights and solving real-world challenges. Thank you for supporting and guiding us through this transformative experience."