Ethics Analysis - Survey Data

This notebook performs numeric and visual analysis on the cleaned survey data:

- Descriptive statistics
- Likert scale visualization
- Correlation heatmap
- Experience vs confidence analysis

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load cleaned data
df = pd.read_csv("../cleaned_survey.csv")
df.head()
```

Out[ ]:

	startdate	enddate	status	progress	duration_(in_seconds)	finished	recordeddate	
0	2025-10-09 18:45:00	2025-10-09 18:45:00	IP Address	100	30	True	10/9/2025 18:45	
1	2025-10-09 18:45:00	2025-10-09 18:53:00	IP Address	100	440	True	10/9/2025 18:53	R_
2	2025-10-10 10:17:00	2025-10-10 10:48:00	IP Address	100	1880	True	10/10/2025 10:48	
3	2025-10-10 11:34:00	2025-10-10 12:15:00	IP Address	100	2446	True	10/10/2025 12:15	F
4	2025-10-10 12:15:00	2025-10-10 12:17:00	IP Address	100	104	True	10/10/2025 12:17	F

5 rows × 24 columns
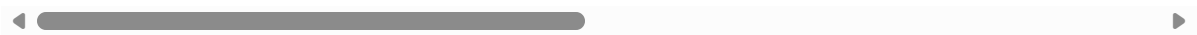
Descriptive Statistics

We start by summarizing the dataset to understand its structure and key metrics.

```python
# Ensure dataframe `df` exists; if not, load it (and import pandas only if needed)
if 'df' not in globals():
        if 'pd' not in globals():
```

```
        import pandas as pd
    df = pd.read_csv("../cleaned_survey.csv")

df.describe(include='all')
```

Out[4]:

	startdate	enddate	status	progress	duration_(in_seconds)	finished	recordeddat
count	30	30	30	30.0	30.000000	30	3
unique	28	29	1	NaN	NaN	1	2
top	2025-10-09 18:45:00	2025-11-12 19:20:00	IP Address	NaN	NaN	True	11/12/202 19:2
freq	2	2	30	NaN	NaN	30	
mean	NaN	NaN	NaN	100.0	461.533333	NaN	Na
std	NaN	NaN	NaN	0.0	589.413076	NaN	Na
min	NaN	NaN	NaN	100.0	24.000000	NaN	Na
25%	NaN	NaN	NaN	100.0	150.250000	NaN	Na
50%	NaN	NaN	NaN	100.0	268.000000	NaN	Na
75%	NaN	NaN	NaN	100.0	432.500000	NaN	Na
max	NaN	NaN	NaN	100.0	2446.000000	NaN	Na

11 rows × 24 columns

◄ ──────────────────────── ►
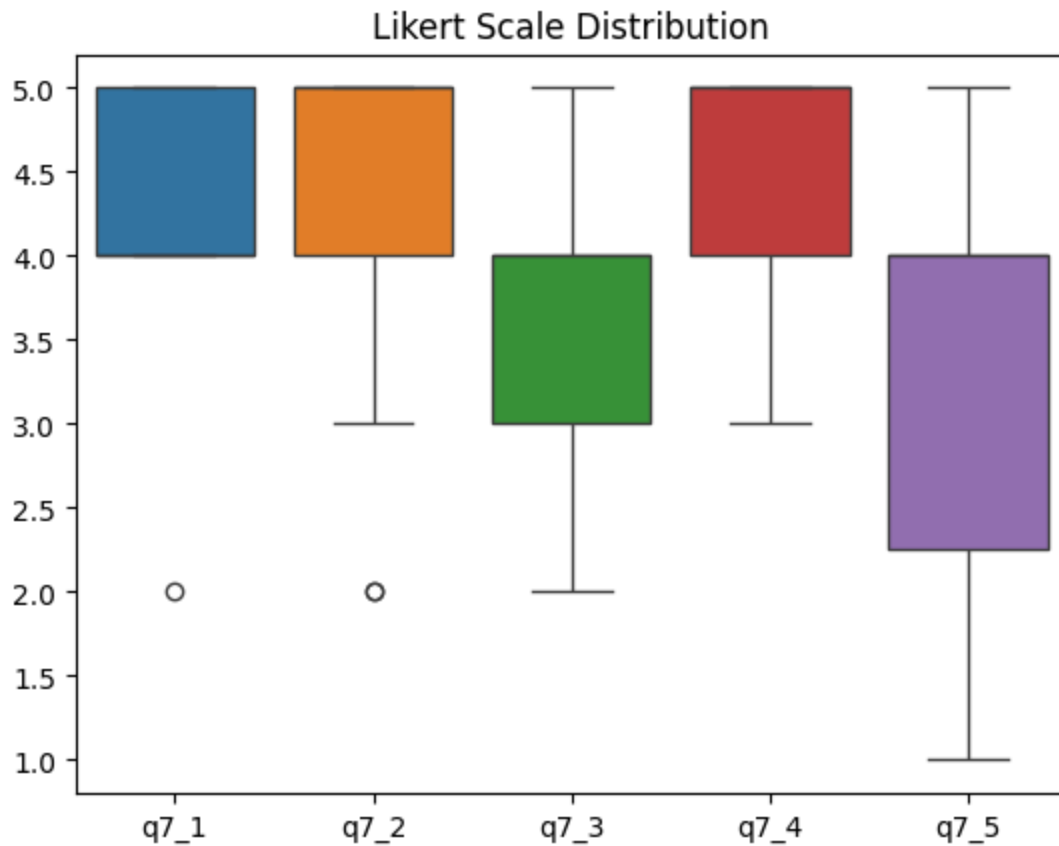
Likert Scale Analysis

Visualize the distribution of Likert scale responses (Q7_1 to Q7_5).

```
In [5]: likert_cols = [col for col in df.columns if col.startswith("q7_")]
sns.boxplot(data=df[likert_cols])
plt.title("Likert Scale Distribution")
plt.show()
```

Likert Scale Distribution

Correlation Heatmap

Check relationships between Likert items.

```
In [6]: corr = df[likert_cols].corr()
        sns.heatmap(corr, annot=True, cmap="coolwarm")
        plt.title("Correlation Heatmap of Likert Items")
        plt.show()
```

Correlation Heatmap of Likert Items
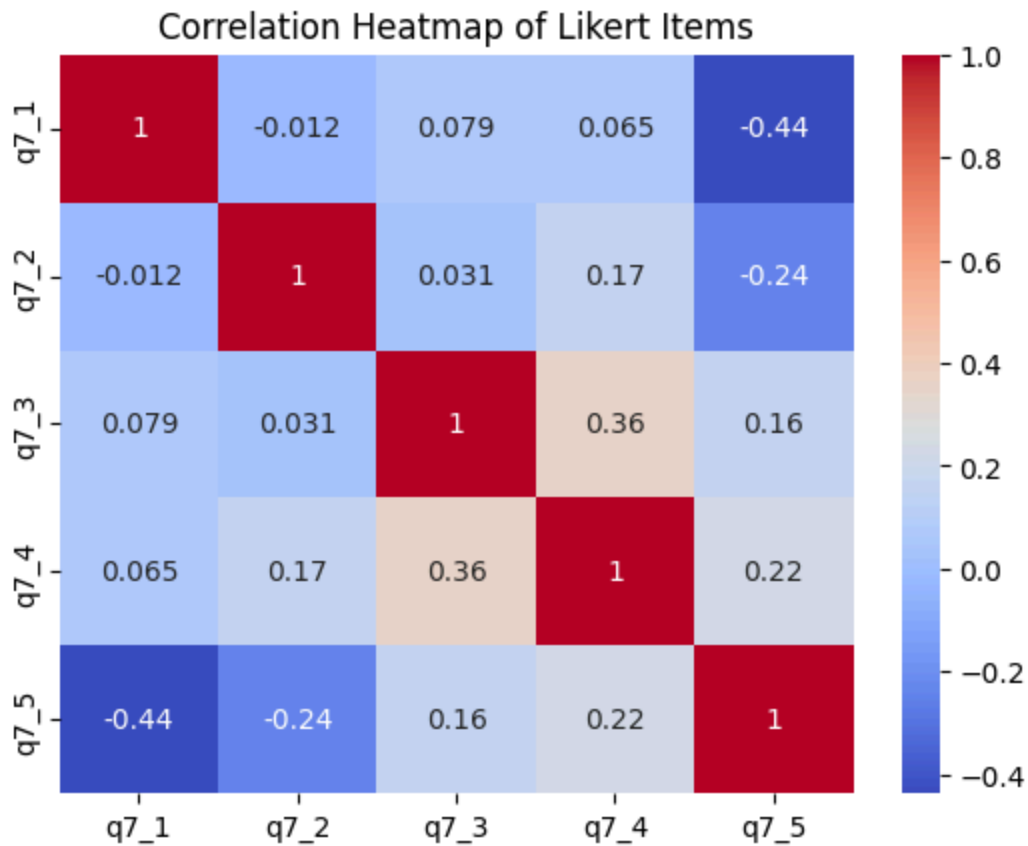
Experience vs Confidence

Convert experience ranges to numeric and plot against confidence score.

```
In [7]:  experience_map = {
             "Less than 1 year": 0.5,
             "1-3 years": 2,
             "4-7 years": 5.5,
             "8-15 years": 11.5,
             "16+ years": 18
         }
         df['years_numeric'] = df['q3'].map(experience_map)

         sns.scatterplot(x=df['years_numeric'], y=df['q7_1'])
         plt.title("Years of Experience vs Confidence")
         plt.show()
```

Years of Experience vs Confidence