01-results-data-prep

February 16, 2026

0.1 01-results-data-prep.ipynb

Creates the **Chapter 4.2 Participant Profile** tables and bar charts from the raw survey export (`data.csv`).

What it does: - Loads the Google Forms dataset and maps long question headers to short keys. - Generates **frequency tables (n, %)** and **bar charts** for: - demographics (status, gender, age, semester) - education background (level of study, field/department) - experience & AI background (work experience, AI awareness, AI exposure)

Output: Descriptive tables and plots used in the participant profile/results section.

[9]:
```
import pandas as pd
import matplotlib.pyplot as plt
import re

# Display tables
pd.set_option('display.max_rows', 200)
pd.set_option('display.max_colwidth', 200)

df = pd.read_csv("data.csv") #reads data from the same repository
df.head()
```

[9]:
```
            Column 1  \
0  1/8/2026 2:33:06
1  1/8/2026 2:36:28
2  1/8/2026 2:43:17
3  1/8/2026 2:45:05
4  1/8/2026 2:48:44

  Have you read the information above and do you agree to participate in this
survey?  \
0                                                  Yes, I agree to
participate.
1                                                  Yes, I agree to
participate.
2                                                  Yes, I agree to
participate.
3                                                  Yes, I agree to
```

```
participate.
4                                                    Yes, I agree to
participate.

        Which of the following best describes you?  \
0  I am currently enrolled as a university student
1  I am currently enrolled as a university student
2  I am currently enrolled as a university student
3  I am currently enrolled as a university student
4  I am currently enrolled as a university student

  Which semester are you currently in? What is your age group?  \
0                                  1-2                   20-22
1                                  3-4                   26-30
2                                  3-4                   20-22
3                                  1-2                   20-22
4                                  5-6                 Over 30

  What is your gender?  \
0                 Male
1                 Male
2                 Male
3                 Male
4               Female

  Which of the following best describes your study programme or department?  \
0                                                                        Law
1                                       Business / Management / Economics
2                                       Business / Management / Economics
3                                       Business / Management / Economics
4                                                   Engineering (non-IT)

  Which best describes your current or most recently completed level of study?
\
0                                                                    Bachelor
1                                                                    Bachelor
2                                                                    Bachelor
3                                                                    Bachelor
4                                                                    Bachelor

  Do you have any previous work or internship experience?  \
0                                    Yes, part-time work
1                                    Yes, full-time work
2                                 Yes, an internship only
3                                    Yes, part-time work
4     Yes, both internship and (part-time/full-time) work
```

```
  Before this survey, were you aware that some companies use AI-based systems in
recruitment and selection (e.g., automated CV screening, AI ranking of
candidates, AI video interview analysis)?  \
0
Yes
1
Yes
2
Yes
3
Yes
4
Yes

   …  \
0  …
1  …
2  …
3  …
4  …

  Please indicate how strongly you agree or disagree with the following
statements about AI-based recruitment systems.\n(AI-based recruitment includes
tools such as automated CV screening, AI ranking of candidates, AI video
interview analysis, etc.) [AI-based recruitment systems feel like a "black box"
to me. (reverse-coded)]  \
0
Disagree
1
Strongly agree
2
Strongly disagree
3
Agree
4
Strongly disagree

  Please indicate how strongly you agree or disagree with the following
statements about AI-based recruitment systems.\n(AI-based recruitment includes
tools such as automated CV screening, AI ranking of candidates, AI video
interview analysis, etc.) [I trust AI systems to evaluate candidates fairly.]  \
0
Disagree
1
Strongly agree
2
Disagree
```

```
3
Neither agree nor disagree
4
Agree

   Please indicate how strongly you agree or disagree with the following
statements about AI-based recruitment systems.\n(AI-based recruitment includes
tools such as automated CV screening, AI ranking of candidates, AI video
interview analysis, etc.) [I worry that AI recruitment systems might
discriminate against certain groups of applicants. (reverse-coded)]  \
0
Strongly agree
1
Agree
2
Neither agree nor disagree
3
Neither agree nor disagree
4
Strongly agree

  Please indicate how strongly you agree or disagree with the following
statements about AI-based recruitment systems.\n(AI-based recruitment includes
tools such as automated CV screening, AI ranking of candidates, AI video
interview analysis, etc.) [I would be uncomfortable if an AI system made the
final decision about whether I get an interview. (reverse-coded)]  \
0
Strongly disagree
1
Neither agree nor disagree
2
Neither agree nor disagree
3
Disagree
4
Agree

  Please indicate how strongly you agree or disagree with the following
statements about AI-based recruitment systems.\n(AI-based recruitment includes
tools such as automated CV screening, AI ranking of candidates, AI video
interview analysis, etc.) [Overall, I feel positive about the use of AI in
recruitment.]  \
0
Strongly agree
1
Neither agree nor disagree
2
```

```
Neither agree nor disagree
3
Strongly agree
4
Disagree

  Please indicate how strongly you agree or disagree with the following
statements about AI-based recruitment systems.\n(AI-based recruitment includes
tools such as automated CV screening, AI ranking of candidates, AI video
interview analysis, etc.) [I would apply to a company that uses AI in its
recruitment process.]  \
0
Neither agree nor disagree
1
Agree
2
Disagree
3
Neither agree nor disagree
4
Neither agree nor disagree

  Please indicate how strongly you agree or disagree with the following
statements about AI-based recruitment systems.\n(AI-based recruitment includes
tools such as automated CV screening, AI ranking of candidates, AI video
interview analysis, etc.) [Knowing that a company uses AI to screen applications
would discourage me from applying. (reverse-coded)]  \
0
Agree
1
Neither agree nor disagree
2
Neither agree nor disagree
3
Disagree
4
Agree

  If a company clearly states that it uses AI tools as part of its recruitment
process, how likely would you be to apply for a job or internship there?  \
0
4
1
4
2
2
3
```

```
4
4
2

  If you had two similar job opportunities, and the only difference was the
recruitment process, which would you prefer?  \
0                                    Company B: Recruitment handled by a
combination of AI tools and human recruiters
1                                                  Company A: Recruitment
handled mainly by human recruiters only
2                                    Company B: Recruitment handled by a
combination of AI tools and human recruiters
3                                    Company B: Recruitment handled by a
combination of AI tools and human recruiters
4                                    Company B: Recruitment handled by a
combination of AI tools and human recruiters

  For the initial screening of applications (deciding who is invited to the
first interview), which option would you feel most comfortable with?
0
AI tools fully decide which candidates are shortlisted
1                                                    AI tools review
applications first, and human recruiters review the shortlisted candidates
2                                                    AI tools review
applications first, and human recruiters review the shortlisted candidates
3                                                    AI tools review
applications first, and human recruiters review the shortlisted candidates
4
Only human recruiters review all applications

[5 rows x 28 columns]
```

[10]:
```python
# Following is the Python dictionary mapping for the columns
COL = {
    'status': 'Which of the following best describes you?',
    'semester': 'Which semester are you currently in?',
    'age': 'What is your age group?',
    'gender': 'What is your gender?',
    'field': 'Which of the following best describes your study programme or␣
 ↪department?',
    'level': 'Which best describes your current or most recently completed␣
 ↪level of study?',
    'workexp': 'Do you have any previous work or internship experience?',
    'aware': 'Before this survey, were you aware that some companies use␣
 ↪AI-based systems in recruitment and selection (e.g., automated CV screening,␣
 ↪AI ranking of candidates, AI video interview analysis)?',
```

```
    'applied_ai': 'Have you ever applied for a position (job, internship,␣
 ↪trainee, etc.) where AI was used or likely used in the recruitment process?',
    'understand': 'How well do you feel you understand, in general, how␣
 ↪AI-based recruitment systems work?',
}
```

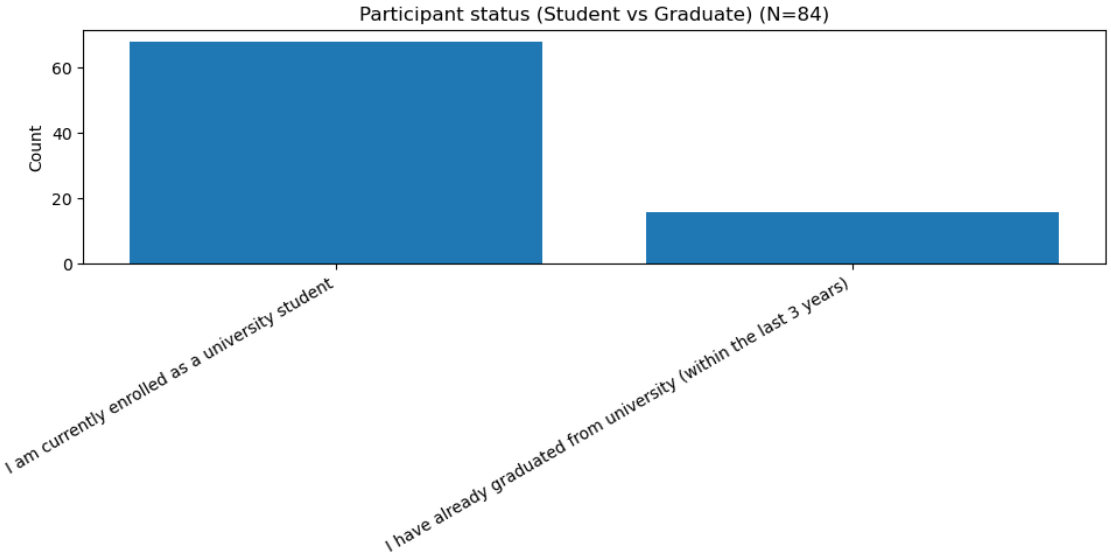
[11]:
```
# defining function for the frequency table
def freq_table(data: pd.DataFrame, col: str, dropna: bool = True) -> pd.
 ↪DataFrame:
    """Return a frequency table."""
    s = data[col]
    if dropna:
        s = s.dropna()
    counts = s.value_counts(dropna=False)
    perc = (counts / counts.sum() * 100).round(1)
    out = pd.DataFrame({
        'Category': counts.index.astype(str),
        'n': counts.values,
        '%': perc.values
    })
    return out

# defining function for the bar chart
def bar_chart(data: pd.DataFrame, col: str, title: str, dropna: bool = True,␣
 ↪rotation: int = 30):
    """Simple bar chart."""
    s = data[col]
    if dropna:
        s = s.dropna()
    counts = s.value_counts()

    plt.figure(figsize=(10, 5))
    plt.bar(counts.index.astype(str), counts.values)
    plt.title(title)
    plt.ylabel('Count')
    plt.xticks(rotation=rotation, ha='right')
    plt.tight_layout()
    plt.show()

# The following function displays the table and chart for mapped column key
def show_profile(data: pd.DataFrame, key: str, title: str, dropna: bool = True,␣
 ↪rotation: int = 30):
    """Display frequency table and bar chart for a mapped column key."""
    col = COL[key]
    t = freq_table(data, col, dropna=dropna)
    display(t)
```

```
    bar_chart(data, col, f"{title} (N={len(data)})", dropna=dropna,␣
↪rotation=rotation)
    return t
```

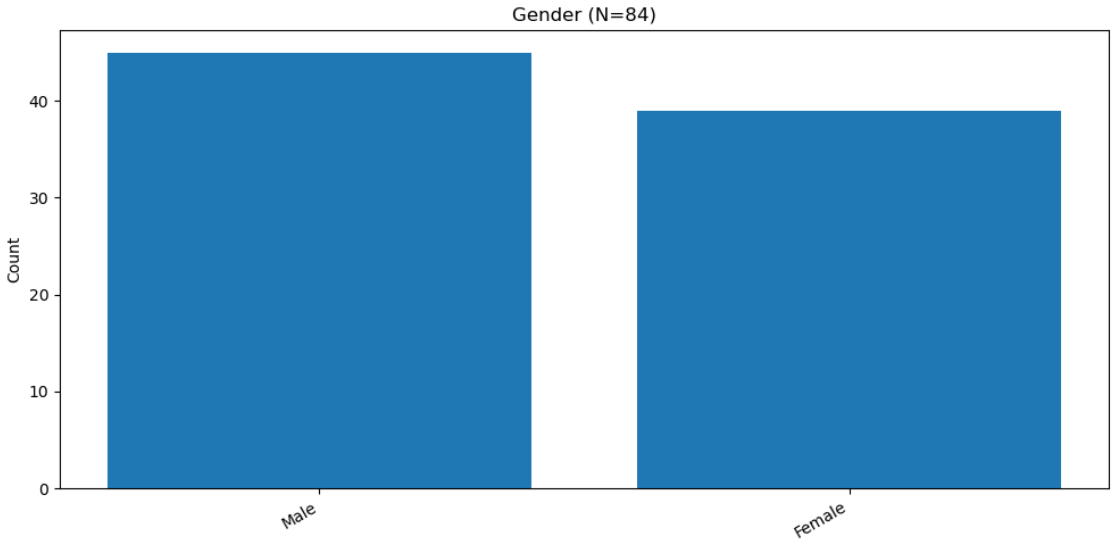
0.2 Demographics

[12]:
```
t_status = show_profile(df, 'status', 'Participant status (Student vs␣
↪Graduate)')
t_gender = show_profile(df, 'gender', 'Gender')
t_age = show_profile(df, 'age', 'Age group')
```

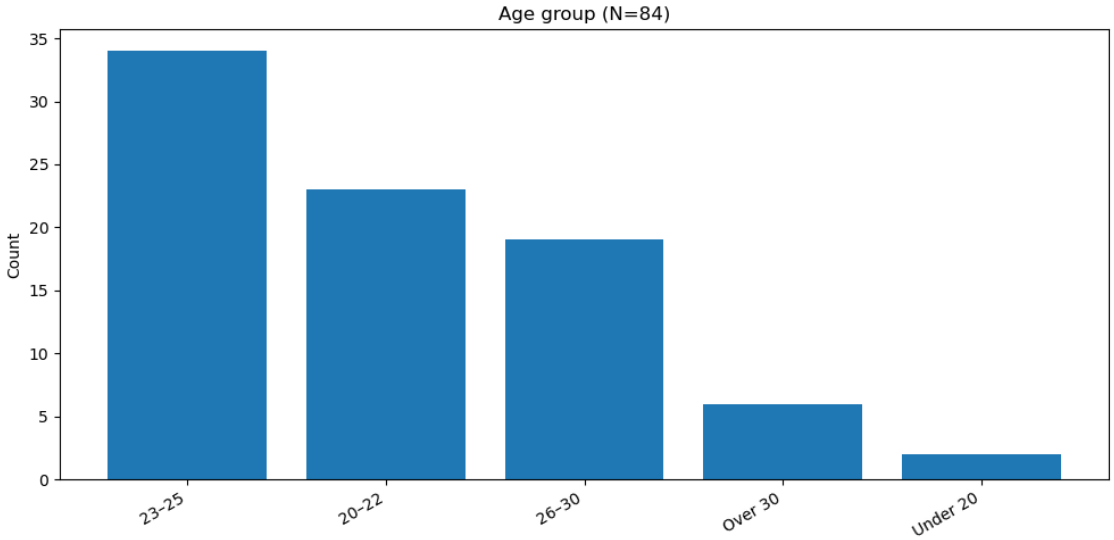
```
                                                 Category   n  \
0                    I am currently enrolled as a university student  68
1  I have already graduated from university (within the last 3 years)  16

      %
0  81.0
1  19.0
```




```
  Category   n     %
0     Male  45  53.6
1   Female  39  46.4
```



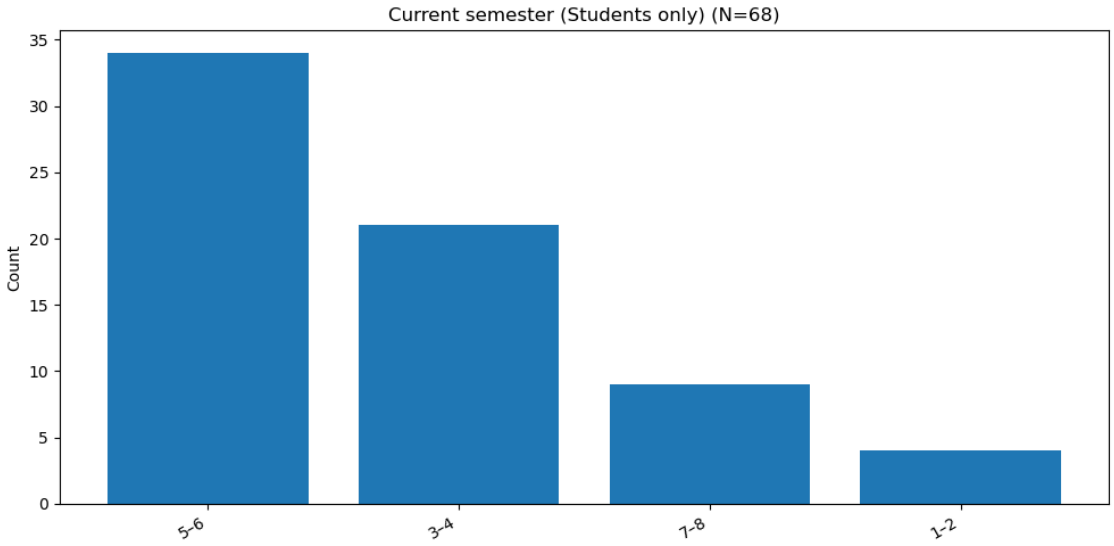
	Category	n	%
0	23-25	34	40.5
1	20-22	23	27.4
2	26-30	19	22.6
3	Over 30	6	7.1
4	Under 20	2	2.4



[13]:
```
# Semester is only meaningful for CURRENT students at universities (graduates␣
↪cannot be counted for this)
```

```
students_only = df[df[COL['status']] == 'I am currently enrolled as a␣
 ↪university student'].copy()
t_semester = show_profile(students_only, 'semester', 'Current semester␣
 ↪(Students only)', dropna=True)
```

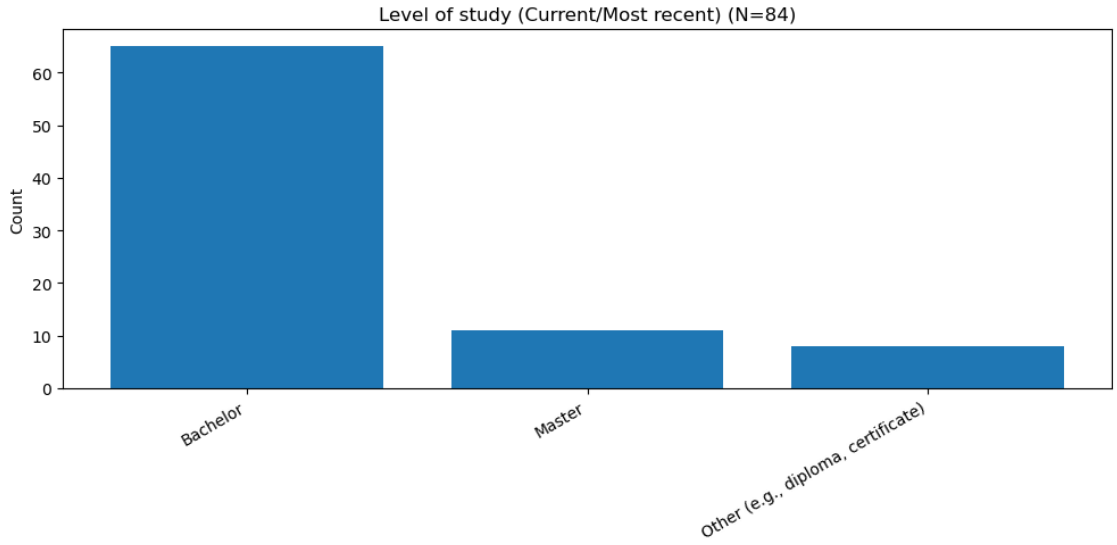
```
  Category   n     %
0      5-6  34  50.0
1      3-4  21  30.9
2      7-8   9  13.2
3      1-2   4   5.9
```



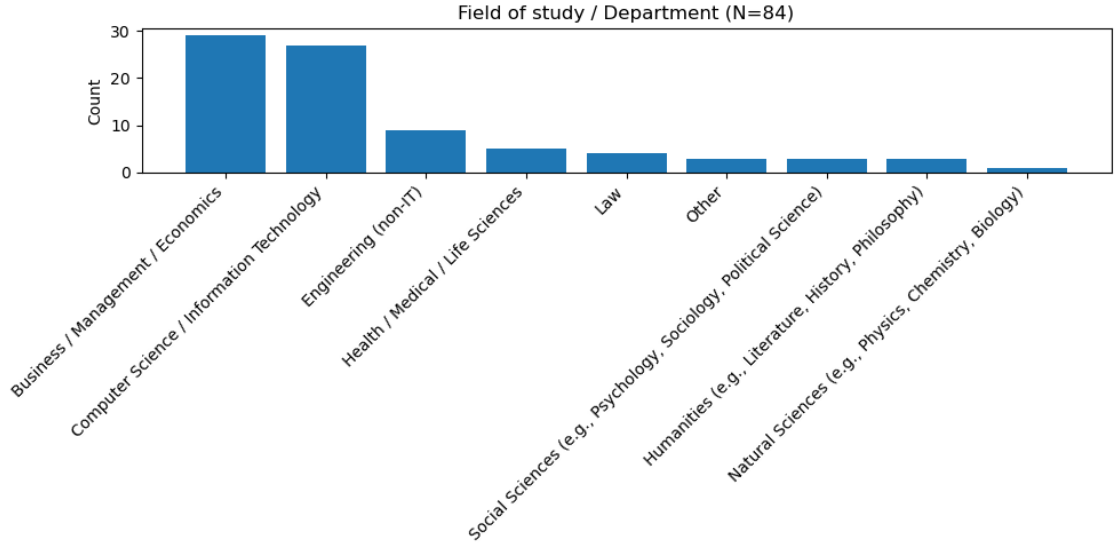
0.3 Educational background

[14]:
```
t_level = show_profile(df, 'level', 'Level of study (Current/Most recent)')
t_field = show_profile(df, 'field', 'Field of study / Department', rotation=45)
```

```
                          Category   n     %
0                         Bachelor  65  77.4
1                           Master  11  13.1
2  Other (e.g., diploma, certificate)   8   9.5
```



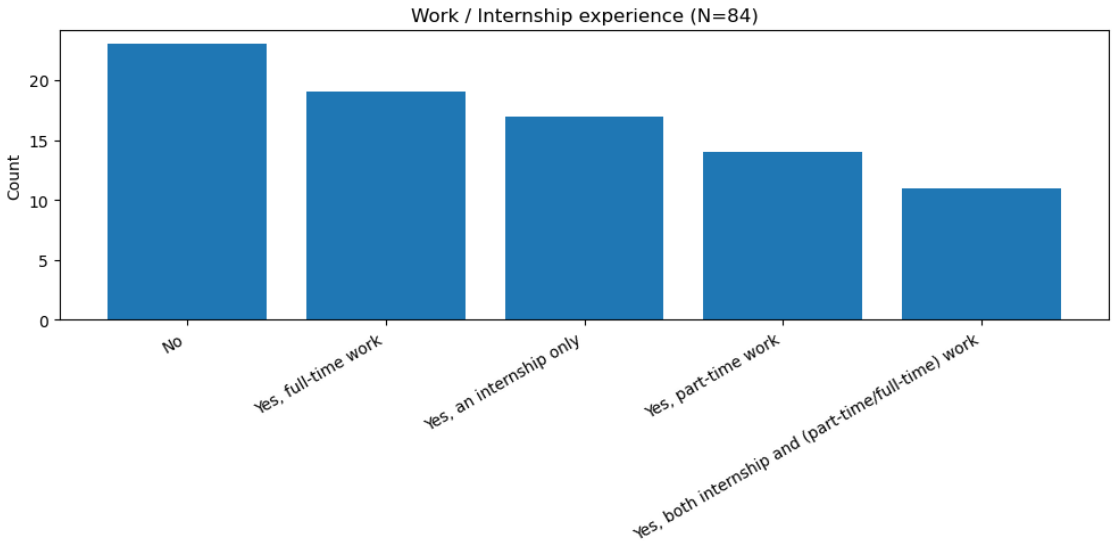
	Category	n	%
0	Business / Management / Economics	29	34.5
1	Computer Science / Information Technology	27	32.1
2	Engineering (non-IT)	9	10.7
3	Health / Medical / Life Sciences	5	6.0
4	Law	4	4.8
5	Other	3	3.6
6	Social Sciences (e.g., Psychology, Sociology, Political Science)	3	3.6
7	Humanities (e.g., Literature, History, Philosophy)	3	3.6
8	Natural Sciences (e.g., Physics, Chemistry, Biology)	1	1.2



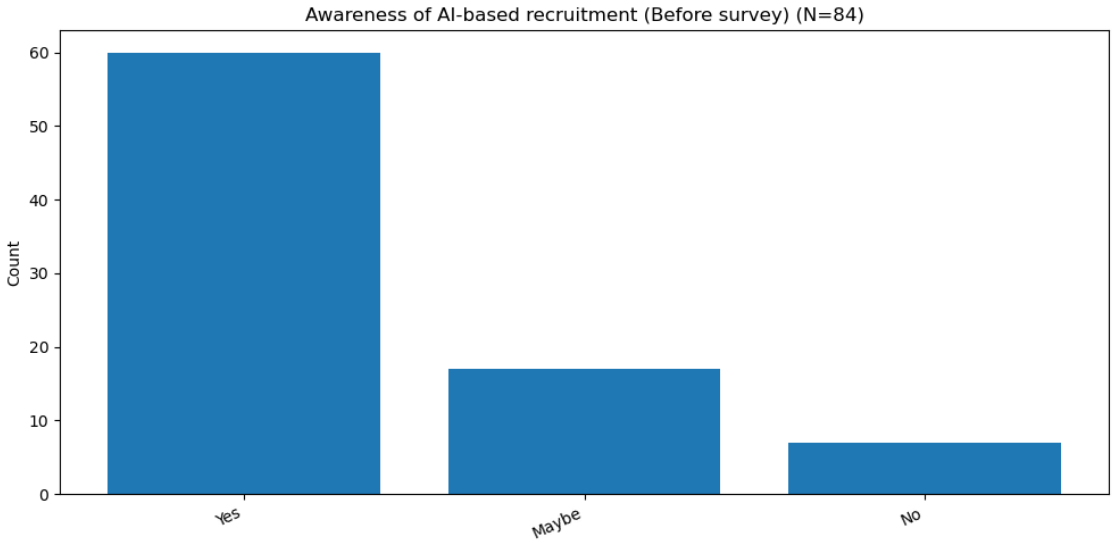
0.4 Experience and AI-related background

[15]:
```
t_workexp = show_profile(df, 'workexp', 'Work / Internship experience')
t_aware = show_profile(df, 'aware', 'Awareness of AI-based recruitment (Before␣
↪survey)', rotation=25)
t_applied = show_profile(df, 'applied_ai', 'Applied where AI was used/likely␣
↪used')
```

	Category	n	%
0	No	23	27.4
1	Yes, full-time work	19	22.6
2	Yes, an internship only	17	20.2
3	Yes, part-time work	14	16.7
4	Yes, both internship and (part-time/full-time) work	11	13.1



	Category	n	%
0	Yes	60	71.4
1	Maybe	17	20.2
2	No	7	8.3



	Category	n	%
0	No	33	39.3
1	Yes	28	33.3
2	Maybe	23	27.4