```
import os
```

```
# Change your directory
os.chdir("C:\Users\Claire Gebirtz\Desktop\Python Datasets")
```

```
import pandas as pd
```

```
# Set your present working directory
pwd = os.getcwd()
```

```
# Load in the dataset
dataset = pd.read_excel(pwd+"\\data - Survey Monkey Output Edited.xlsx", sheet_name="Edited Data")
```

```
# Display your present working directory
pwd
```

```
'C:\Users\Claire Gebirtz\Desktop\Python Datasets'
```

```
# Display the loaded dataset
dataset
```

	Respondent	Start Date	End Date	Email Address	Last Name	Custom Data 1	Identify which of the following best describes your job role?	Which of the following best describes your place of work?	Which of the following best describes your gender?	Please select the age group that you identify in.	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9	Question 10
0	5379342560	2021-06-01 17:51:17-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN
1	2080720268	2021-06-01 17:52:11-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Finance	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
2	4044420599	2021-06-01 17:52:36-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Department Lead	...	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
3	5030968025	2021-06-01 17:53:20-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Manager	...	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
4	1598826391	2021-06-01 17:56:24-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Pod Operations	NaN	Manager	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN
195	7940900661	2021-06-08 15:51:11-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Department Lead	...	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
196	5327709011	2021-06-08 15:52:17-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Finance	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN
197	4847659952	2021-06-08 15:53:03-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Pod Operations	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN
198	5639135540	2021-06-08 15:53:50-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Manager	...	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
199	5514655413	2021-06-08 15:54:39-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Information Technology	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN

199 rows × 100 columns

```
# Make a copy of the dataframe
dataset_modified = dataset.copy()
dataset_modified
```

	Respondent	Start Date	End Date	Email Address	Last Name	Custom Data 1	Identify which of the following best describes your job role?	Which of the following best describes your place of work?	Which of the following best describes your gender?	Please select the age group that you identify in.	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9	Question 10
0	5379342560	2021-06-01 17:51:17-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN
1	2080720268	2021-06-01 17:52:11-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Finance	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
2	4044420599	2021-06-01 17:52:36-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Department Lead	...	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
3	5030968025	2021-06-01 17:53:20-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Manager	...	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
4	1598826391	2021-06-01 17:56:24-05:00	2021-06-01 17:56:41-05:00	NaN	NaN	NaN	NaN	Pod Operations	NaN	Manager	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN
195	7940900661	2021-06-08 15:51:11-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Department Lead	...	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
196	5327709011	2021-06-08 15:52:17-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Finance	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN
197	4847659952	2021-06-08 15:53:03-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Pod Operations	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN
198	5639135540	2021-06-08 15:53:50-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Infrastructure	NaN	Manager	...	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2
199	5514655413	2021-06-08 15:54:39-05:00	2021-06-08 15:56:11-05:00	NaN	NaN	NaN	NaN	Information Technology	NaN	Staff	...	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN

199 rows × 100 columns

```
# View the column names
dataset_modified.columns
```

```
Index(['Respondent ID', 'Start Date', 'End Date', 'Email Address',
       'First Name', 'Last Name', 'Custom Data 1',
       'Identify which of the following best describes your job role?',
       'Which of the following best describes your place of work?',
       'Which of the following best describes your gender?',
       'Please select the age group that you identify in.',
       'Question 1', 'Question 2', 'Question 3', 'Question 4',
       'Question 5', 'Question 6', 'Question 7', 'Question 8',
       'Question 9', 'Question 10',
       'Please select the gender at which you identify.',
       'Please select the place of work at which you identify.',
       'Which of the following best describes your company? Response',
       'Which of the following best describes your employment type?',
       'Question 1 Response', 'Question 1 Subquestion',
       'Question 1 Open Ended Response', 'Question 1 Answer',
       'Question 2 Response', 'Question 2 Subquestion',
       'Question 2 Open Ended Response', 'Question 2 Answer',
       'Question 3 Response', 'Question 3 Subquestion',
       'Question 3 Open Ended Response', 'Question 3 Answer',
       'Question 4 Response', 'Question 4 Subquestion',
       'Question 4 Open Ended Response', 'Question 4 Answer',
       'Question 5 Response', 'Question 5 Subquestion',
       'Question 5 Open Ended Response', 'Question 5 Answer',
       'Question 6 Response', 'Question 6 Subquestion',
       'Question 6 Open Ended Response', 'Question 6 Answer',
       'Question 7 Response', 'Question 7 Subquestion',
       'Question 7 Open Ended Response', 'Question 7 Answer',
       'Question 8 Response', 'Question 8 Subquestion',
       'Question 8 Open Ended Response', 'Question 8 Answer',
       'Question 9 Response', 'Question 9 Subquestion',
       'Question 9 Open Ended Response', 'Question 9 Answer',
       'Question 10 Response', 'Question 10 Subquestion',
       'Question 10 Open Ended Response', 'Question 10 Answer',
       'Question 11 Response', 'Question 11 Subquestion',
       'Question 11 Open Ended Response', 'Question 11 Answer',
       'Question 12 Response', 'Question 12 Subquestion',
       'Question 12 Open Ended Response', 'Question 12 Answer',
       'Question 13 Response', 'Question 13 Subquestion',
       'Question 13 Open Ended Response', 'Question 13 Answer',
       'Question 14 Response', 'Question 14 Subquestion',
       'Question 14 Open Ended Response', 'Question 14 Answer',
       'Question 15 Response', 'Question 15 Subquestion',
       'Question 15 Open Ended Response', 'Question 15 Answer',
       'Question 16 Response', 'Question 16 Subquestion',
       'Question 16 Open Ended Response', 'Question 16 Answer',
       'Question 17 Response', 'Question 17 Subquestion',
       'Question 17 Open Ended Response', 'Question 17 Answer',
       'Question 18 Response', 'Question 18 Subquestion',
       'Question 18 Open Ended Response', 'Question 18 Answer',
       'Question 19 Response', 'Question 19 Subquestion',
       'Question 19 Open Ended Response', 'Question 19 Answer',
       'Question 20 Response', 'Question 20 Subquestion',
       'Question 20 Open Ended Response', 'Question 20 Answer'],
      dtype='object', 100 rows)
```

```
# Drop the columns not needed
columns_to_drop = ['Start Date', 'End Date', 'Email Address', 'First Name', 'Last Name', 'Custom Data 1']
```

```
# Start Date
# End Date
# Email Address
# First Name
# Last Name
# Custom Data 1
```

```
dataset_modified = dataset_modified.drop(columns=columns_to_drop)
```

```
dataset_modified
```

	Respondent ID	Identify which of the following best describes your job role?	Which of the following best describes your place of work?	Which of the following best describes your gender?	Please select the age group that you identify in.	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9	Question 10
0	5379342560	Infrastructure	NaN	Information Technology	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN	NaN	NaN
1	2080720268	Finance	NaN	Staff	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2	NaN	NaN
2	4044420599	Infrastructure	NaN	Department Lead	NaN	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2	NaN
3	5030968025	Infrastructure	NaN	Manager	NaN	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2	NaN
4	1598826391	Pod Operations	NaN	Manager	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN	NaN	NaN
195	7940900661	Infrastructure	NaN	Department Lead	NaN	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2	NaN
196	5327709011	Finance	NaN	Staff	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN	NaN	NaN
197	4847659952	Pod Operations	NaN	Staff	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN	NaN	NaN
198	5639135540	Infrastructure	NaN	Manager	NaN	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	Answer 2	NaN
199	5514655413	Information Technology	NaN	Staff	NaN	Answer 1	Answer 3	Answer 4	Answer 5	NaN	NaN	Answer 1	NaN	NaN	NaN

199 rows × 94 columns

```
id_vars = list(dataset_modified.columns[0:9])
# This is equivalent to
id_vars = ['Respondent ID', 'Identify which of the following best describes your job role?',
           'Which of the following best describes your place of work?',
           'Which of the following best describes your gender?',
           'Please select the age group that you identify in.',
           'Question 1', 'Question 2', 'Question 3', 'Question 4']
```

```
# Change the DataFrame format from wide to long
dataset_melted = dataset_modified.melt(id_vars=id_vars, value_vars = value_vars, var_name="Question + Subquestion", value_name="Answer")
dataset_melted
```

	Respondent ID	Identify which of the following best describes your job role?	Which of the following best describes your place of work?	Which of the following best describes your gender?	Please select the age group that you identify in.	Question 1	Question 2	Question 3	Question 4	Question + Subquestion	Answer
0	5379342560	Infrastructure	NaN	Staff	Gen Z (Born 1997-1999)	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN
1	2080720268	Finance	NaN	Staff	Gen Z (Born 1997-1999)	NaN	Answer 1	Answer 3	Answer 4	Question 5 Response	NaN
2	4044420599	Infrastructure	NaN	Department Lead	Gen Z (Born 1997-1999)	Answer 1	Answer 2	Answer 3	Answer 4	Question 5 Response	Answer 5
3	5030968025	Infrastructure	NaN	Manager	Gen Z (Born 1997-1999)	Answer 1	Answer 2	Answer 3	Answer 4	Question 5 Response	Answer 5
4	1598826391	Pod Operations	NaN	Manager	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN
1760	7940900661	Infrastructure	NaN	Department Lead	Baby Boomer (Born 1946-1964)	NaN	Answer 2	Answer 3	Answer 4	Question 20 Response	Answer 5
1764	5327709011	Finance	NaN	Staff	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN
1768	4847659952	Pod Operations	NaN	Staff	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN
1772	5639135540	Infrastructure	NaN	Manager	Millennial (Born 1981-1996)	Answer 1	Answer 2	Answer 3	Answer 4	Question 20 Response	Answer 5
1776	5514655413	Information Technology	NaN	Staff	Prefer not to answer	NaN	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN

1760 rows × 12 columns

```
# Import the 'Questions' sheet name from the Excel file
questions_import = pd.read_excel(pwd+"\\data - Survey Monkey Output Edited.xlsx", sheet_name= "Questions")
questions_import
```

	Respondent	Start Date	End Date	Email Address	Last Name	Custom Data 1
0	Respondent ID	Text	Response ID	Text	Response ID	Text
1	Start Date	Text	Start Date	Text	Start Date	Text
2	End Date	Text	End Date	Text	End Date	Text
3	Email Address	Text	Email Address	Text	Email Address	Text
4	First Name	Text	First Name	Text	First Name	Text
...	...	...	...	...	...	...
96	NaN	Response 19	Question 20	Response 19	Question 20 Response 19	NaN
97	NaN	Question 20	Response 20	Question 20	Response 20	NaN
98	Question 20	Response 20	Question 20	Response 20	Question 20 Response 20	NaN
99	NaN	Response 1	Question 20	Response 1	Question 20 Response 1	NaN
100	NaN	Response 2	Question 20	Response 2	Question 20 Response 2	NaN

100 rows × 6 columns

```
# Make a copy of the 'Questions' sheet name from the Excel file
questions = questions_import.copy()
questions.drop(['Start Date', 'End Date', 'Email Address', 'Last Name', 'Custom Data 1'], inplace=True)
questions
```

	Questions	Question + Subquestion
0	Respondent ID	Respondent ID
1	Start Date	Start Date
2	End Date	End Date
3	Email Address	Email Address
4	First Name	First Name
...	...	...
96	NaN	Question 20 Response 19
97	Question 20	Question 20 Response 20
98	Question 20	Question 20 Response 20
99	NaN	Question 20 Response 1
100	NaN	Question 20 Response 2

100 rows × 2 columns

```
# Join the merge method for pointwise combination the specified Dataset
dataset_merged1=pd.merge(left=dataset_melted,right=questions, how="left", left_on="Question + Subquestion", right_on="Question + Subquestion")
print("Merged Dataset Shape: ", dataset_merged1.shape)
print("Merge Method: left")
dataset_merged1
```

```
Merged Dataset Shape: (1760, 13)
Merge Method: left
```

	Respondent ID	Identify which of the following best describes your job role?	Which of the following best describes your place of work?	Which of the following best describes your gender?	Please select the age group that you identify in.	Question 1	Question 2	Question 3	Question 4	Question + Subquestion	Answer	Questions
0	5379342560	Infrastructure	NaN	Information Technology	NaN	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN	Question 5
1	2080720268	Finance	NaN	Staff	NaN	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN	Question 5
2	4044420599	Infrastructure	NaN	Department Lead	Gen Z (Born 1997-1999)	Answer 1	Answer 2	Answer 3	Answer 4	Question 5 Response	Answer 5	Question 5
3	5030968025	Infrastructure	NaN	Manager	Gen Z (Born 1997-1999)	Answer 1	Answer 2	Answer 3	Answer 4	Question 5 Response	Answer 5	Question 5
4	1598826391	Pod Operations	NaN	Manager	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN	Question 5
...	...	...	...	...	...	...	...	...	...	...	...	...
1760	7940900661	Infrastructure	NaN	Department Lead	Baby Boomer (Born 1946-1964)	NaN	Answer 2	Answer 3	Answer 4	Question 20 Response	Answer 5	Question 20
1764	5327709011	Finance	NaN	Staff	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN	Question 20
1768	4847659952	Pod Operations	NaN	Staff	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN	Question 20
1772	5639135540	Infrastructure	NaN	Manager	Millennial (Born 1981-1996)	Answer 1	Answer 2	Answer 3	Answer 4	Question 20 Response	Answer 5	Question 20
1776	5514655413	Information Technology	NaN	Staff	Prefer not to answer	NaN	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN	Question 20

1760 rows × 12 columns

```
# Find the number of respondents in each question
respondents = questions[questions['Questions'] == 'Respondent ID'].count()
respondents=respondents.drop(labels=['Questions'])
respondents = pandas.DataFrame(data=['Respondents'], index=respondents, columns=['Questions'])
respondents.sort_index()
```

	Questions
	Respondents
Question 1	110
Question 10	110
Question 11	104
Question 12	110
Question 13	110
Question 14	110
Question 15	110
Question 16	110
Question 17	110
Question 18	110
Question 19	110
Question 2	110
Question 20	110
Question 3	110
Question 4	110
Question 5	110
Question 6	110
Question 7	110
Question 8	110
Question 9	110

```
# Merge the specified series
dataset_merged_sum = pd.merge(left= dataset_melted, right=respondents, how="left", left_on="Questions", right_on="Questions")
print("Merged Dataset Shape: ", dataset_merged_sum.shape)
print("Merge Method: left")
dataset_merged_sum
```

```
Merged Dataset Shape: (1760, 13)
Merge Method: left
```

	Respondent ID	Identify which of the following best describes your job role?	Which of the following best describes your place of work?	Which of the following best describes your gender?	Please select the age group that you identify in.	Question 1	Question 2	Question 3	Question 4	Question + Subquestion	Answer	Questions	Respondents
0	5379342560	Infrastructure	NaN	Information Technology	NaN	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN	Question 5	110
1	2080720268	Finance	NaN	Staff	NaN	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN	Question 5	110
2	4044420599	Infrastructure	NaN	Department Lead	Gen Z (Born 1997-1999)	Answer 1	Answer 2	Answer 3	Answer 4	Question 5 Response	Answer 5	Question 5	110
3	5030968025	Infrastructure	NaN	Manager	Gen Z (Born 1997-1999)	Answer 1	Answer 2	Answer 3	Answer 4	Question 5 Response	Answer 5	Question 5	110
4	1598826391	Pod Operations	NaN	Manager	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN	Question 5	110
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1760	7940900661	Infrastructure	NaN	Department Lead	Baby Boomer (Born 1946-1964)	NaN	Answer 2	Answer 3	Answer 4	Question 20 Response	Answer 5	Question 20	110
1764	5327709011	Finance	NaN	Staff	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN	Question 20	110
1768	4847659952	Pod Operations	NaN	Staff	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN	Question 20	110
1772	5639135540	Infrastructure	NaN	Manager	Millennial (Born 1981-1996)	Answer 1	Answer 2	Answer 3	Answer 4	Question 20 Response	Answer 5	Question 20	110
1776	5514655413	Information Technology	NaN	Staff	Prefer not to answer	NaN	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN	Question 20	110

1760 rows × 12 columns

```
# Find the number of people who had same answer for the specified questions
same_answer= dataset_merged[(dataset_merged['Answer'] == 'Answer 1')]
same_answer=same_answer.groupby(['Question', 'Question + Subquestion', 'Answer'])['Respondent ID'].nunique().reset_index()
same_answer.rename(columns={'Respondent ID': 'Same Answer'}, inplace=True)
same_answer
```

	Question	Question + Subquestion	Answer	Same Answer
0	Question 1 Response	Question 1 Response	Answer 1	10
1	Question 2 Response	Question 2 Response	Answer 1	10
2	Question 3 Response	Question 3 Response	Answer 1	10
3	Question 4 Response	Question 4 Response	Answer 1	10
4	Question 5 Response	Question 5 Response	Answer 1	10
5	Question 6 Response	Question 6 Response	Answer 1	10
6	Question 7 Response	Question 7 Response	Answer 1	10
7	Question 8 Response	Question 8 Response	Answer 1	10
8	Question 9 Response	Question 9 Response	Answer 1	10
9	Question 10 Response	Question 10 Response	Answer 1	10
10	Question 11 Response	Question 11 Response	Answer 1	10
11	Question 12 Response	Question 12 Response	Answer 1	10
12	Question 13 Response	Question 13 Response	Answer 1	10
13	Question 14 Response	Question 14 Response	Answer 1	10
14	Question 15 Response	Question 15 Response	Answer 1	10
15	Question 16 Response	Question 16 Response	Answer 1	10
16	Question 17 Response	Question 17 Response	Answer 1	10
17	Question 18 Response	Question 18 Response	Answer 1	10
18	Question 19 Response	Question 19 Response	Answer 1	10
19	Question 20 Response	Question 20 Response	Answer 1	10
20	Question 5 Response	Question 5 Response	Answer 1	10
21	Question 6 Response	Question 6 Response	Answer 1	10
22	Question 7 Response	Question 7 Response	Answer 1	10
23	Question 8 Response	Question 8 Response	Answer 1	10
24	Question 9 Response	Question 9 Response	Answer 1	10
25	Question 10 Response	Question 10 Response	Answer 1	10
26	Question 11 Response	Question 11 Response	Answer 1	10
27	Question 12 Response	Question 12 Response	Answer 1	10
28	Question 13 Response	Question 13 Response	Answer 1	10
29	Question 14 Response	Question 14 Response	Answer 1	10
30	Question 15 Response	Question 15 Response	Answer 1	10
31	Question 16 Response	Question 16 Response	Answer 1	10
32	Question 17 Response	Question 17 Response	Answer 1	10
33	Question 18 Response	Question 18 Response	Answer 1	10
34	Question 19 Response	Question 19 Response	Answer 1	10
35	Question 20 Response	Question 20 Response	Answer 1	10
36	Question 5 Response	Question 5 Response	Answer 1	10
37	Question 6 Response	Question 6 Response	Answer 1	10
38	Question 7 Response	Question 7 Response	Answer 1	10
39	Question 8 Response	Question 8 Response	Answer 1	10
40	Question 9 Response	Question 9 Response	Answer 1	10
41	Question 10 Response	Question 10 Response	Answer 1	10
42	Question 11 Response	Question 11 Response	Answer 1	10
43	Question 12 Response	Question 12 Response	Answer 1	10
44	Question 13 Response	Question 13 Response	Answer 1	10
45	Question 14 Response	Question 14 Response	Answer 1	10
46	Question 15 Response	Question 15 Response	Answer 1	10
47	Question 16 Response	Question 16 Response	Answer 1	10
48	Question 17 Response	Question 17 Response	Answer 1	10
49	Question 18 Response	Question 18 Response	Answer 1	10
50	Question 19 Response	Question 19 Response	Answer 1	10
51	Question 20 Response	Question 20 Response	Answer 1	10

```
# Merge the specified series
dataset_merged_same = pd.merge(left= dataset_merged, right=same_answer, how="left", left_on="Question", right_on="Question")
print("Merged Dataset Shape: ", dataset_merged_same.shape)
print("Merge Method: left")
dataset_merged_same
```

```
Merged Dataset Shape: (1760, 15)
Merge Method: left
```

	Respondent ID	Identify which of the following best describes your job role?	Which of the following best describes your place of work?	Which of the following best describes your gender?	Please select the age group that you identify in.	Question 1	Question 2	Question 3	Question 4	Question + Subquestion	Answer	Questions	Respondents	Same Answer
0	5379342560	Infrastructure	NaN	Information Technology	NaN	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN	Question 5	110	10
1	2080720268	Finance	NaN	Staff	NaN	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN	Question 5	110	10
2	4044420599	Infrastructure	NaN	Department Lead	Gen Z (Born 1997-1999)	Answer 1	Answer 2	Answer 3	Answer 4	Question 5 Response	Answer 5	Question 5	110	10
3	5030968025	Infrastructure	NaN	Manager	Gen Z (Born 1997-1999)	Answer 1	Answer 2	Answer 3	Answer 4	Question 5 Response	Answer 5	Question 5	110	10
4	1598826391	Pod Operations	NaN	Manager	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 5 Response	NaN	Question 5	110	10
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1760	7940900661	Infrastructure	NaN	Department Lead	Baby Boomer (Born 1946-1964)	NaN	Answer 2	Answer 3	Answer 4	Question 20 Response	Answer 5	Question 20	110	10
1764	5327709011	Finance	NaN	Staff	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN	Question 20	110	10
1768	4847659952	Pod Operations	NaN	Staff	Millennial (Born 1981-1996)	Answer 1	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN	Question 20	110	10
1772	5639135540	Infrastructure	NaN	Manager	Millennial (Born 1981-1996)	Answer 1	Answer 2	Answer 3	Answer 4	Question 20 Response	Answer 5	Question 20	110	10
1776	5514655413	Information Technology	NaN	Staff	Prefer not to answer	NaN	Answer 3	Answer 4	Answer 5	Question 20 Response	NaN	Question 20	110	10

1760 rows × 15 columns

```
# Rename the columns
dataset_merged_same
dataset_merged_same.rename(columns={'Respondent ID': 'Respondent ID', 'Identify which division you work in. Response': 'Division Primary', 'Identify which division you work in. Response': 'Division Primary', 'Which of the following best describes your place of work?': 'Place of Work', 'Which of the following best describes your gender?': 'Gender', 'Which of the following best describes your employment type?': 'Employment Type', 'Question + Subquestion': 'Question + Subquestion', 'Answer': 'Answer', 'Questions': 'Question', 'Respondents': 'Same Answer', 'Same Answer': 'Respondents'}, inplace=True)
```

```
# Export the finished dataset as final dataset
output_to_excel=pd.ref("Final Dept.xlsx", index=False)
```