Exploratory Data Analysis of COVID-19 Clinical Trials

1. Objective

The objective is to explore the dataset to gain insights into the characteristics of COVID-19 clinical trials, such as their status, phases, study designs, and demographics.

2. Importing Libraries and Loading Data

First, you'll need to import the necessary libraries and load your dataset.

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import missingno as msno
```

In [3]:
```
df = pd.read_csv(r'C:\Users\user\OneDrive\Desktop\Data internship\Datasets\Proje
df.head()
```

Out[3]:

	Rank	NCT Number	Title	Acronym	Status	Study Results	Conditions
0	1	NCT04785898	Diagnostic Performance of the ID Now™ COVID-19...	COVID-IDNow	Active, not recruiting	No Results Available	Covid19
1	2	NCT04595136	Study to Evaluate the Efficacy of COVID19-0001...	COVID-19	Not yet recruiting	No Results Available	SARS-CoV-2 Infection
2	3	NCT04395482	Lung CT Scan Analysis of SARS-CoV2 Induced Lun...	TAC-COVID19	Recruiting	No Results Available	covid19
3	4	NCT04416061	The Role of a Private Hospital in Hong Kong Am...	COVID-19	Active, not recruiting	No Results Available	COVID
4	5	NCT04395924	Maternal-foetal Transmission of SARS-Cov-2	TMF-COVID-19	Recruiting	No Results Available	Maternal Fetal Infection Transmission|COVID-19...

5 rows × 27 columns

3. Initial Data Exploration

In [5]:
```
df.shape
```

Out[5]:
```
(5783, 27)
```

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5783 entries, 0 to 5782
Data columns (total 27 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Rank                      5783 non-null   int64
 1   NCT Number                5783 non-null   object
 2   Title                     5783 non-null   object
 3   Acronym                   2480 non-null   object
 4   Status                    5783 non-null   object
 5   Study Results             5783 non-null   object
 6   Conditions                5783 non-null   object
 7   Interventions             4897 non-null   object
 8   Outcome Measures          5748 non-null   object
 9   Sponsor/Collaborators     5783 non-null   object
 10  Gender                    5773 non-null   object
 11  Age                       5783 non-null   object
 12  Phases                    3322 non-null   object
 13  Enrollment                5749 non-null   float64
 14  Funded Bys                5783 non-null   object
 15  Study Type                5783 non-null   object
 16  Study Designs             5748 non-null   object
 17  Other IDs                 5782 non-null   object
 18  Start Date                5749 non-null   object
 19  Primary Completion Date   5747 non-null   object
 20  Completion Date           5747 non-null   object
 21  First Posted              5783 non-null   object
 22  Results First Posted      36 non-null     object
 23  Last Update Posted        5783 non-null   object
 24  Locations                 5198 non-null   object
 25  Study Documents           182 non-null    object
 26  URL                       5783 non-null   object
dtypes: float64(1), int64(1), object(25)
memory usage: 1.2+ MB
```

In [11]:
```
pd.options.display.float_format = '{:.2f}'.format

df.describe()
```

Out[11]:

	Rank	Enrollment
count	5783.00	5749.00
mean	2892.00	18319.49
std	1669.55	404543.73
min	1.00	0.00
25%	1446.50	60.00
50%	2892.00	170.00
75%	4337.50	560.00
max	5783.00	20000000.00

In [8]:
```
df.describe(include='object')
```

Out[8]:

	NCT Number	Title	Acronym	Status	Study Results	Conditions	Interventi
count	5783	5783	2480	5783	5783	5783	4
unique	5783	5775	2338	12	2	3067	4
top	NCT04785898	Study Assessing Vagus Nerve Stimulation in CoV...	COVID-19	Recruiting	No Results Available	COVID-19	Other: interven
freq	1	2	47	2805	5747	720	

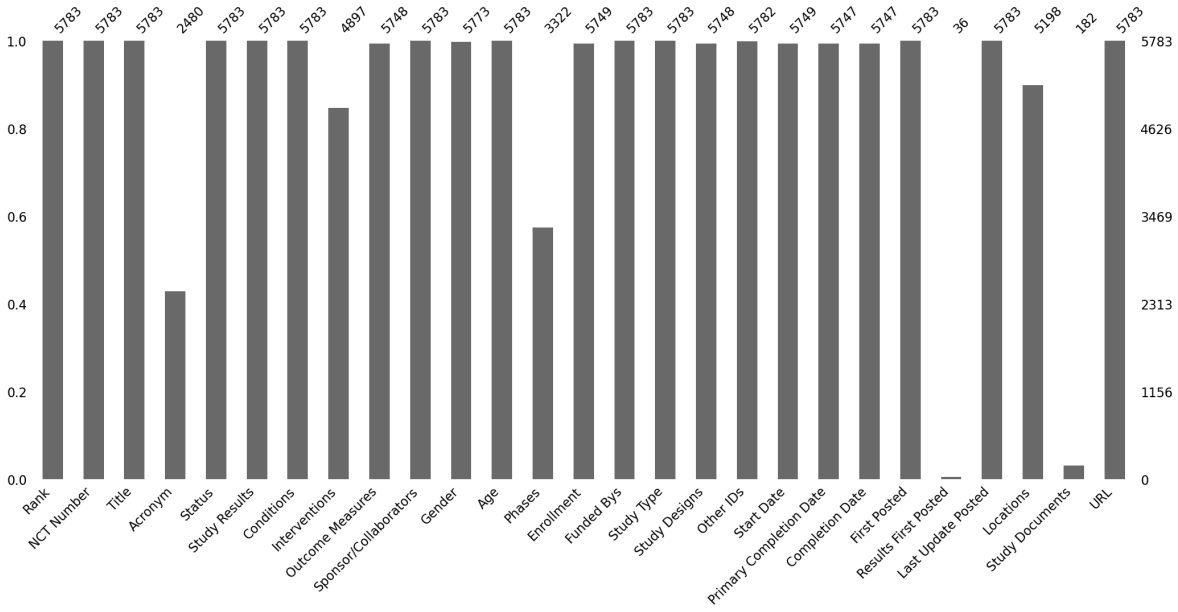
4 rows × 25 columns

4. Handling Missing Data

In [12]:
```
df.isnull().sum()
```

Out[12]:
```
Rank                        0
NCT Number                  0
Title                       0
Acronym                  3303
Status                      0
Study Results               0
Conditions                  0
Interventions             886
Outcome Measures           35
Sponsor/Collaborators       0
Gender                     10
Age                         0
Phases                   2461
Enrollment                 34
Funded Bys                  0
Study Type                  0
Study Designs              35
Other IDs                   1
Start Date                 34
Primary Completion Date    36
Completion Date            36
First Posted                0
Results First Posted     5747
Last Update Posted          0
Locations                 585
Study Documents          5601
URL                         0
dtype: int64
```

In [16]:
```
msno.bar(df)
plt.show()
```



In [18]:
```
columns_to_drop = [
    'Acronym',
    'Results First Posted',
    'Study Documents',
    'URL',
    'Other IDs',
    'Funded Bys',
    'Study Designs'
]

df = df.drop(columns=columns_to_drop)
df.head()
```

Out[18]:

	Rank	NCT Number	Title	Status	Study Results	Conditions	Interventi
0	1	NCT04785898	Diagnostic Performance of the ID Now™ COVID-19...	Active, not recruiting	No Results Available	Covid19	Diagnc Test: ID Nc COVID Screer
1	2	NCT04595136	Study to Evaluate the Efficacy of COVID19-0001...	Not yet recruiting	No Results Available	SARS-CoV-2 Infection	Drug: D COVID 0(USR|D normal sa
2	3	NCT04395482	Lung CT Scan Analysis of SARS-CoV2 Induced Lun...	Recruiting	No Results Available	covid19	Other: L CT s analys COVID pati
3	4	NCT04416061	The Role of a Private Hospital in Hong Kong Am...	Active, not recruiting	No Results Available	COVID	Diagnc Test: CC 19 Diagnc
4	5	NCT04395924	Maternal-foetal Transmission of SARS-Cov-2	Recruiting	No Results Available	Maternal Fetal Infection Transmission|COVID-19...	Diagnc Diagnos SARS-Cov F

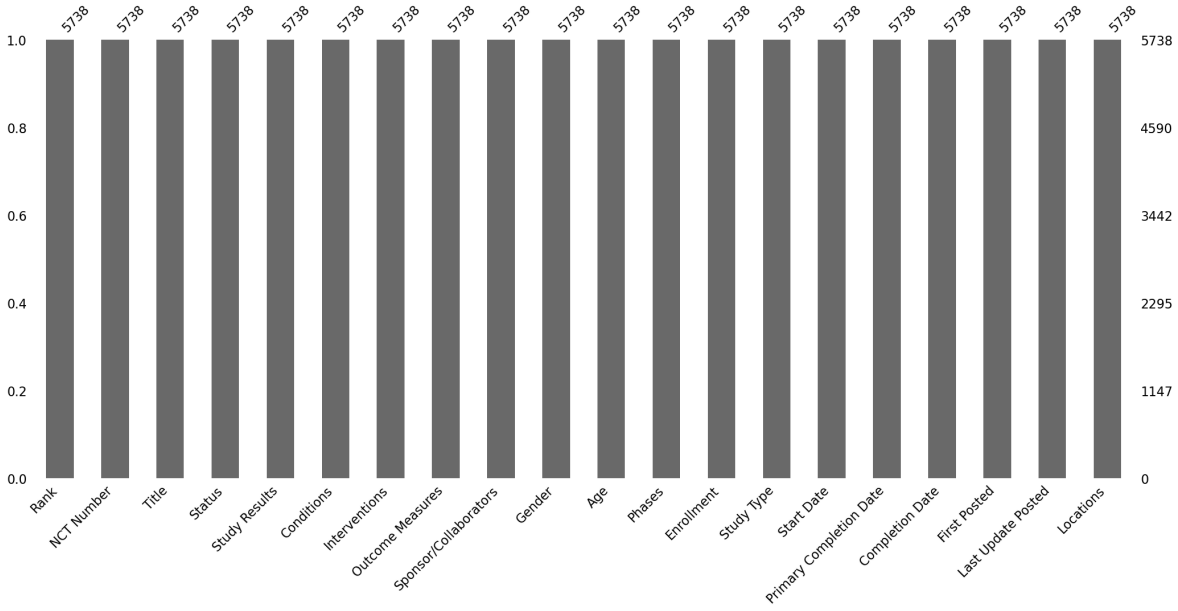
In [19]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5783 entries, 0 to 5782
Data columns (total 20 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Rank                     5783 non-null   int64  
 1   NCT Number               5783 non-null   object 
 2   Title                    5783 non-null   object 
 3   Status                   5783 non-null   object 
 4   Study Results            5783 non-null   object 
 5   Conditions               5783 non-null   object 
 6   Interventions            4897 non-null   object 
 7   Outcome Measures         5748 non-null   object 
 8   Sponsor/Collaborators    5783 non-null   object 
 9   Gender                   5773 non-null   object 
 10  Age                      5783 non-null   object 
 11  Phases                   3322 non-null   object 
 12  Enrollment               5749 non-null   float64
 13  Study Type               5783 non-null   object 
 14  Start Date               5749 non-null   object 
 15  Primary Completion Date  5747 non-null   object 
 16  Completion Date          5747 non-null   object 
 17  First Posted             5783 non-null   object 
 18  Last Update Posted       5783 non-null   object 
 19  Locations                5198 non-null   object 
dtypes: float64(1), int64(1), object(18)
memory usage: 903.7+ KB
```

In [22]:
```
df = df.dropna(subset=['Outcome Measures', 'Start Date', 'Completion Date', 'Gen

df.fillna('Unknown', inplace=True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 5738 entries, 0 to 5782
Data columns (total 20 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Rank                     5738 non-null   int64  
 1   NCT Number               5738 non-null   object 
 2   Title                    5738 non-null   object 
 3   Status                   5738 non-null   object 
 4   Study Results            5738 non-null   object 
 5   Conditions               5738 non-null   object 
 6   Interventions            5738 non-null   object 
 7   Outcome Measures         5738 non-null   object 
 8   Sponsor/Collaborators    5738 non-null   object 
 9   Gender                   5738 non-null   object 
 10  Age                      5738 non-null   object 
 11  Phases                   5738 non-null   object 
 12  Enrollment               5738 non-null   float64
 13  Study Type               5738 non-null   object 
 14  Start Date               5738 non-null   object 
 15  Primary Completion Date  5738 non-null   object 
 16  Completion Date          5738 non-null   object 
 17  First Posted             5738 non-null   object 
 18  Last Update Posted       5738 non-null   object 
 19  Locations                5738 non-null   object 
dtypes: float64(1), int64(1), object(18)
memory usage: 941.4+ KB
```

In [23]:
```
msno.bar(df)
plt.show()
```

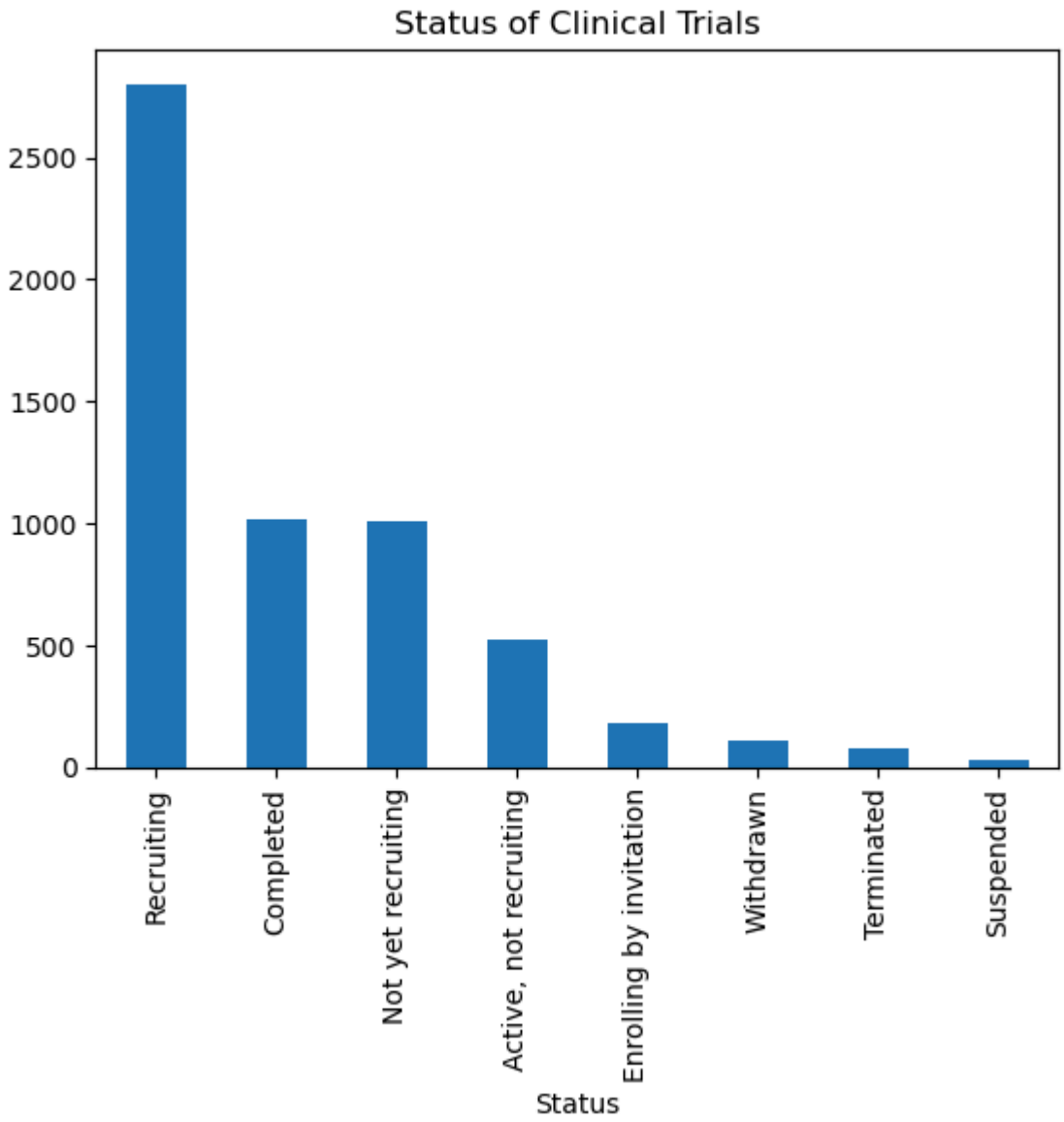


5. Univariate Analysis

In [25]:
```
# Status Distribution: Analyze the status of clinical trials.

print(df['Status'].value_counts())
df['Status'].value_counts().plot(kind='bar', title='Status of Clinical Trials')
```

```
Status
Recruiting                2802
Completed                 1018
Not yet recruiting        1004
Active, not recruiting     526
Enrolling by invitation    180
Withdrawn                  107
Terminated                  74
Suspended                   27
Name: count, dtype: int64
```

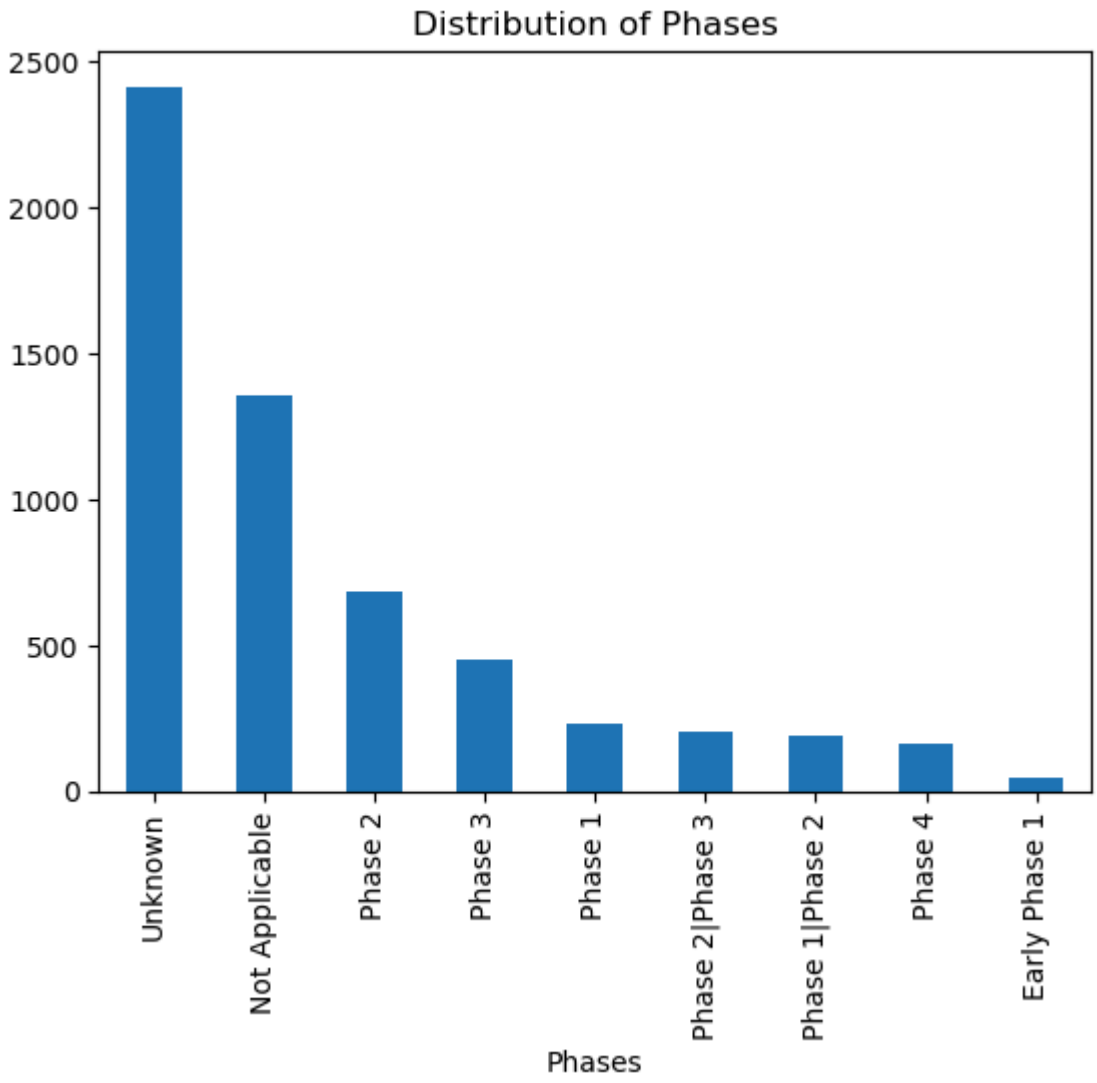
Out[25]: <Axes: title={'center': 'Status of Clinical Trials'}, xlabel='Status'>



In [26]:
```
# Phase Distribution: Understand the distribution of trial phases.

print(df['Phases'].value_counts())
df['Phases'].value_counts().plot(kind='bar', title='Distribution of Phases')
```

```
Phases
Unknown              2416
Not Applicable       1354
Phase 2               685
Phase 3               450
Phase 1               234
Phase 2|Phase 3       200
Phase 1|Phase 2       192
Phase 4               161
Early Phase 1          46
Name: count, dtype: int64
```

Out[26]: <Axes: title={'center': 'Distribution of Phases'}, xlabel='Phases'>



In [28]:
```
unique_values = df['Age'].unique()
print("Unique Values:", unique_values)

unique_count = len(unique_values)
print("Total Number of Unique Values:", unique_count)
```

```
Benzersiz Değerler: ['18 Years and older \xa0 (Adult, Older Adult)'
 'Child, Adult, Older Adult' '18 Years to 48 Years \xa0 (Adult)'
 '18 Years to 75 Years \xa0 (Adult, Older Adult)'
 '18 Years to 45 Years \xa0 (Adult)'
 '18 Years to 99 Years \xa0 (Adult, Older Adult)'
 '18 Years to 55 Years \xa0 (Adult)'
 '15 Years and older \xa0 (Child, Adult, Older Adult)'
 '18 Years to 80 Years \xa0 (Adult, Older Adult)'
 '45 Years and older \xa0 (Adult, Older Adult)'
 '20 Years to 100 Years \xa0 (Adult, Older Adult)'
 '8 Years to 88 Years \xa0 (Child, Adult, Older Adult)'
 '5 Years to 65 Years \xa0 (Child, Adult, Older Adult)'
 'up to 99 Years \xa0 (Child, Adult, Older Adult)'
 '18 Years to 85 Years \xa0 (Adult, Older Adult)'
 '18 Years to 65 Years \xa0 (Adult, Older Adult)'
 'up to 29 Days \xa0 (Child)'
 '18 Years to 70 Years \xa0 (Adult, Older Adult)'
 '18 Years to 59 Years \xa0 (Adult)'
 'up to 100 Years \xa0 (Child, Adult, Older Adult)'
 '20 Years to 60 Years \xa0 (Adult)'
 '40 Years to 80 Years \xa0 (Adult, Older Adult)'
 '23 Years and older \xa0 (Adult, Older Adult)'
 '18 Years to 120 Years \xa0 (Adult, Older Adult)'
 '16 Years and older \xa0 (Child, Adult, Older Adult)'
 '5 Years to 90 Years \xa0 (Child, Adult, Older Adult)'
 '18 Years to 90 Years \xa0 (Adult, Older Adult)'
 'up to 18 Years \xa0 (Child, Adult)'
 '2 Years and older \xa0 (Child, Adult, Older Adult)'
 '70 Years and older \xa0 (Older Adult)'
 '18 Years to 26 Years \xa0 (Adult)'
 '18 Years to 95 Years \xa0 (Adult, Older Adult)'
 '12 Years and older \xa0 (Child, Adult, Older Adult)'
 '16 Years to 55 Years \xa0 (Child, Adult)'
 '30 Years to 70 Years \xa0 (Adult, Older Adult)'
 '35 Years to 65 Years \xa0 (Adult, Older Adult)'
 '18 Years to 40 Years \xa0 (Adult)' '18 Years to 60 Years \xa0 (Adult)'
 '18 Years to 100 Years \xa0 (Adult, Older Adult)'
 '6 Years and older \xa0 (Child, Adult, Older Adult)'
 'up to 17 Years \xa0 (Child)'
 '22 Years to 72 Years \xa0 (Adult, Older Adult)'
 '16 Years to 100 Years \xa0 (Child, Adult, Older Adult)'
 '6 Months and older \xa0 (Child, Adult, Older Adult)'
 '20 Years to 65 Years \xa0 (Adult, Older Adult)'
 '14 Years to 75 Years \xa0 (Child, Adult, Older Adult)'
 '5 Years and older \xa0 (Child, Adult, Older Adult)'
 '1 Year to 100 Years \xa0 (Child, Adult, Older Adult)'
 '18 Years to 89 Years \xa0 (Adult, Older Adult)'
 '4 Years to 13 Years \xa0 (Child)'
 '16 Years to 90 Years \xa0 (Child, Adult, Older Adult)'
 '1 Year to 90 Years \xa0 (Child, Adult, Older Adult)'
 '6 Months to 100 Years \xa0 (Child, Adult, Older Adult)'
 '17 Years to 50 Years \xa0 (Child, Adult)'
 '18 Years to 69 Years \xa0 (Adult, Older Adult)'
 '19 Years and older \xa0 (Adult, Older Adult)'
 '18 Years to 50 Years \xa0 (Adult)'
 '8 Years and older \xa0 (Child, Adult, Older Adult)'
 '50 Years and older \xa0 (Adult, Older Adult)'
 'up to 20 Years \xa0 (Child, Adult)'
 '10 Years and older \xa0 (Child, Adult, Older Adult)'
 '18 Years to 49 Years \xa0 (Adult)' '18 Years to 62 Years \xa0 (Adult)'
```

```
'18 Years to 67 Years \xa0 (Adult, Older Adult)'
'40 Years and older \xa0 (Adult, Older Adult)'
'up to 15 Minutes \xa0 (Child)'
'10 Years to 50 Years \xa0 (Child, Adult)'
'11 Years to 100 Years \xa0 (Child, Adult, Older Adult)'
'up to 70 Years \xa0 (Child, Adult, Older Adult)'
'16 Years to 24 Years \xa0 (Child, Adult)'
'17 Years to 85 Years \xa0 (Child, Adult, Older Adult)'
'23 Years to 55 Years \xa0 (Adult)'
'18 Years to 84 Years \xa0 (Adult, Older Adult)'
'65 Years and older \xa0 (Older Adult)'
'21 Years to 65 Years \xa0 (Adult, Older Adult)'
'18 Years to 111 Years \xa0 (Adult, Older Adult)'
'12 Years to 65 Years \xa0 (Child, Adult, Older Adult)'
'40 Years to 76 Years \xa0 (Adult, Older Adult)'
'20 Years and older \xa0 (Adult, Older Adult)'
'25 Years to 35 Years \xa0 (Adult)' '22 Weeks to 28 Weeks \xa0 (Child)'
'18 Years to 105 Years \xa0 (Adult, Older Adult)'
'21 Years and older \xa0 (Adult, Older Adult)'
'21 Years to 50 Years \xa0 (Adult)'
'21 Years to 80 Years \xa0 (Adult, Older Adult)'
'4 Years and older \xa0 (Child, Adult, Older Adult)'
'7 Years and older \xa0 (Child, Adult, Older Adult)'
'1 Year and older \xa0 (Child, Adult, Older Adult)'
'15 Years to 85 Years \xa0 (Child, Adult, Older Adult)'
'18 Years to 64 Years \xa0 (Adult)'
'55 Years and older \xa0 (Adult, Older Adult)'
'40 Years to 90 Years \xa0 (Adult, Older Adult)'
'60 Years and older \xa0 (Adult, Older Adult)'
'20 Years to 80 Years \xa0 (Adult, Older Adult)'
'3 Years to 10 Years \xa0 (Child)'
'17 Years and older \xa0 (Child, Adult, Older Adult)'
'18 Months to 85 Years \xa0 (Child, Adult, Older Adult)'
'1 Month to 99 Years \xa0 (Child, Adult, Older Adult)'
'19 Years to 70 Years \xa0 (Adult, Older Adult)'
'8 Years to 89 Years \xa0 (Child, Adult, Older Adult)'
'18 Years to 110 Years \xa0 (Adult, Older Adult)'
'18 Years to 130 Years \xa0 (Adult, Older Adult)'
'17 Years to 65 Years \xa0 (Child, Adult, Older Adult)'
'up to 14 Years \xa0 (Child)'
'10 Years to 70 Years \xa0 (Child, Adult, Older Adult)'
'1 Year to 2 Years \xa0 (Child)' 'up to 15 Years \xa0 (Child)'
'20 Years to 70 Years \xa0 (Adult, Older Adult)'
'6 Months to 80 Years \xa0 (Child, Adult, Older Adult)'
'3 Months and older \xa0 (Child, Adult, Older Adult)'
'4 Years to 99 Years \xa0 (Child, Adult, Older Adult)'
'21 Years to 70 Years \xa0 (Adult, Older Adult)'
'22 Years to 75 Years \xa0 (Adult, Older Adult)'
'60 Years to 80 Years \xa0 (Adult, Older Adult)'
'3 Years and older \xa0 (Child, Adult, Older Adult)'
'up to 22 Years \xa0 (Child, Adult)'
'13 Years to 59 Years \xa0 (Child, Adult)'
'up to 24 Years \xa0 (Child, Adult)'
'2 Months and older \xa0 (Child, Adult, Older Adult)'
'18 Months and older \xa0 (Child, Adult, Older Adult)'
'18 Years to 79 Years \xa0 (Adult, Older Adult)'
'12 Years to 80 Years \xa0 (Child, Adult, Older Adult)'
'15 Years to 90 Years \xa0 (Child, Adult, Older Adult)'
'18 Years to 42 Years \xa0 (Adult)'
'21 Years to 100 Years \xa0 (Adult, Older Adult)'
```

```
'30 Years to 80 Years \xa0 (Adult, Older Adult)'
'25 Years to 65 Years \xa0 (Adult, Older Adult)'
'1 Year to 18 Years \xa0 (Child, Adult)'
'1 Year to 80 Years \xa0 (Child, Adult, Older Adult)'
'25 Years to 55 Years \xa0 (Adult)'
'18 Years to 101 Years \xa0 (Adult, Older Adult)'
'2 Years to 18 Years \xa0 (Child, Adult)'
'19 Years to 80 Years \xa0 (Adult, Older Adult)'
'30 Years and older \xa0 (Adult, Older Adult)'
'1 Month to 18 Years \xa0 (Child, Adult)'
'25 Years and older \xa0 (Adult, Older Adult)'
'2 Years to 27 Years \xa0 (Child, Adult)'
'16 Years to 120 Years \xa0 (Child, Adult, Older Adult)'
'60 Years to 111 Years \xa0 (Adult, Older Adult)'
'19 Years to 75 Years \xa0 (Adult, Older Adult)'
'1 Month to 17 Years \xa0 (Child)'
'20 Years to 79 Years \xa0 (Adult, Older Adult)'
'17 Years to 75 Years \xa0 (Child, Adult, Older Adult)'
'21 Years to 64 Years \xa0 (Adult)'
'25 Years to 70 Years \xa0 (Adult, Older Adult)'
'12 Years to 100 Years \xa0 (Child, Adult, Older Adult)'
'16 Years to 80 Years \xa0 (Child, Adult, Older Adult)'
'up to 21 Years \xa0 (Child, Adult)' '18 Years to 35 Years \xa0 (Adult)'
'55 Years to 120 Years \xa0 (Adult, Older Adult)'
'1 Year to 15 Years \xa0 (Child)' 'up to 45 Years \xa0 (Child, Adult)'
'90 Years and older \xa0 (Older Adult)'
'18 Years to 39 Years \xa0 (Adult)'
'15 Years to 35 Years \xa0 (Child, Adult)'
'up to 50 Years \xa0 (Child, Adult)'
'15 Years to 80 Years \xa0 (Child, Adult, Older Adult)'
'14 Years and older \xa0 (Child, Adult, Older Adult)'
'1 Year to 16 Years \xa0 (Child)' '2 Years to 15 Years \xa0 (Child)'
'17 Years to 35 Years \xa0 (Child, Adult)'
'6 Years to 85 Years \xa0 (Child, Adult, Older Adult)'
'25 Years to 40 Years \xa0 (Adult)'
'12 Years to 99 Years \xa0 (Child, Adult, Older Adult)'
'18 Years to 30 Years \xa0 (Adult)'
'75 Years and older \xa0 (Older Adult)'
'50 Years to 90 Years \xa0 (Adult, Older Adult)'
'4 Years to 18 Years \xa0 (Child, Adult)'
'up to 85 Years \xa0 (Child, Adult, Older Adult)'
'16 Years to 85 Years \xa0 (Child, Adult, Older Adult)'
'36 Years to 63 Years \xa0 (Adult)'
'4 Years to 80 Years \xa0 (Child, Adult, Older Adult)'
'25 Years to 60 Years \xa0 (Adult)'
'1 Year to 99 Years \xa0 (Child, Adult, Older Adult)'
'16 Years to 40 Years \xa0 (Child, Adult)'
'20 Years to 45 Years \xa0 (Adult)' 'up to 7 Hours \xa0 (Child)'
'4 Years to 17 Years \xa0 (Child)'
'3 Years to 67 Years \xa0 (Child, Adult, Older Adult)'
'up to 120 Years \xa0 (Child, Adult, Older Adult)'
'21 Years to 60 Years \xa0 (Adult)'
'18 Years to 76 Years \xa0 (Adult, Older Adult)'
'25 Years to 80 Years \xa0 (Adult, Older Adult)'
'19 Years to 85 Years \xa0 (Adult, Older Adult)'
'3 Years to 99 Years \xa0 (Child, Adult, Older Adult)'
'24 Years to 37 Years \xa0 (Adult)'
'12 Years to 120 Years \xa0 (Child, Adult, Older Adult)'
'20 Years to 74 Years \xa0 (Adult, Older Adult)'
'up to 28 Days \xa0 (Child)' '30 Years to 55 Years \xa0 (Adult)'
```

```
'18 Years to 117 Years \xa0 (Adult, Older Adult)'
'3 Years to 17 Years \xa0 (Child)'
'12 Years to 25 Years \xa0 (Child, Adult)'
'16 Years to 70 Years \xa0 (Child, Adult, Older Adult)'
'6 Years to 18 Years \xa0 (Child, Adult)'
'30 Years to 60 Years \xa0 (Adult)' 'up to 12 Months \xa0 (Child)'
'12 Years to 57 Years \xa0 (Child, Adult)'
'15 Years to 65 Years \xa0 (Child, Adult, Older Adult)'
'3 Years to 25 Years \xa0 (Child, Adult)'
'18 Years to 88 Years \xa0 (Adult, Older Adult)'
'up to 10 Years \xa0 (Child)' '28 Years to 45 Years \xa0 (Adult)'
'13 Years and older \xa0 (Child, Adult, Older Adult)'
'up to 90 Years \xa0 (Child, Adult, Older Adult)'
'6 Years to 90 Years \xa0 (Child, Adult, Older Adult)'
'40 Years to 60 Years \xa0 (Adult)'
'30 Years to 85 Years \xa0 (Adult, Older Adult)'
'up to 89 Years \xa0 (Child, Adult, Older Adult)'
'14 Years to 100 Years \xa0 (Child, Adult, Older Adult)'
'40 Years to 85 Years \xa0 (Adult, Older Adult)'
'16 Years to 86 Years \xa0 (Child, Adult, Older Adult)'
'20 Years to 50 Years \xa0 (Adult)'
'1 Month to 100 Years \xa0 (Child, Adult, Older Adult)'
'12 Months and older \xa0 (Child, Adult, Older Adult)'
'12 Years to 50 Years \xa0 (Child, Adult)'
'50 Years to 80 Years \xa0 (Adult, Older Adult)'
'45 Years to 75 Years \xa0 (Adult, Older Adult)'
'55 Years to 85 Years \xa0 (Adult, Older Adult)'
'24 Months to 18 Years \xa0 (Child, Adult)'
'19 Years to 65 Years \xa0 (Adult, Older Adult)'
'18 Years to 71 Years \xa0 (Adult, Older Adult)'
'46 Years and older \xa0 (Adult, Older Adult)'
'35 Years to 45 Years \xa0 (Adult)'
'85 Years to 100 Years \xa0 (Older Adult)'
'45 Years to 85 Years \xa0 (Adult, Older Adult)'
'20 Years to 90 Years \xa0 (Adult, Older Adult)'
'12 Years to 85 Years \xa0 (Child, Adult, Older Adult)'
'216 Months and older \xa0 (Adult, Older Adult)'
'3 Years to 18 Years \xa0 (Child, Adult)'
'15 Years to 45 Years \xa0 (Child, Adult)'
'12 Years to 17 Years \xa0 (Child)'
'1 Month to 30 Years \xa0 (Child, Adult)'
'21 Years to 40 Years \xa0 (Adult)'
'20 Years to 85 Years \xa0 (Adult, Older Adult)'
'8 Years to 14 Years \xa0 (Child)' 'up to 25 Years \xa0 (Child, Adult)'
'11 Years and older \xa0 (Child, Adult, Older Adult)'
'19 Years to 64 Years \xa0 (Adult)'
'15 Years to 75 Years \xa0 (Child, Adult, Older Adult)'
'40 Years to 75 Years \xa0 (Adult, Older Adult)'
'16 Years to 99 Years \xa0 (Child, Adult, Older Adult)'
'4 Years to 64 Years \xa0 (Child, Adult)'
'18 Years to 54 Years \xa0 (Adult)'
'40 Years to 70 Years \xa0 (Adult, Older Adult)'
'50 Years to 85 Years \xa0 (Adult, Older Adult)'
'7 Years to 18 Years \xa0 (Child, Adult)'
'24 Years to 45 Years \xa0 (Adult)'
'40 Years to 65 Years \xa0 (Adult, Older Adult)'
'13 Years to 19 Years \xa0 (Child, Adult)'
'25 Years to 67 Years \xa0 (Adult, Older Adult)'
'18 Years to 74 Years \xa0 (Adult, Older Adult)'
'14 Years to 90 Years \xa0 (Child, Adult, Older Adult)'
```

```
'15 Years to 100 Years \xa0 (Child, Adult, Older Adult)'
'18 Years to 47 Years \xa0 (Adult)' 'up to 2 Years \xa0 (Child)'
'50 Years to 75 Years \xa0 (Adult, Older Adult)'
'31 Days to 18 Years \xa0 (Child, Adult)'
'8 Years to 80 Years \xa0 (Child, Adult, Older Adult)'
'18 Years to 125 Years \xa0 (Adult, Older Adult)'
'18 Years to 68 Years \xa0 (Adult, Older Adult)'
'up to 1 Month \xa0 (Child)' '20 Years to 38 Years \xa0 (Adult)'
'6 Months to 11 Years \xa0 (Child)' '3 Months to 16 Years \xa0 (Child)'
'26 Years to 40 Years \xa0 (Adult)'
'65 Years to 110 Years \xa0 (Older Adult)'
'7 Years to 12 Years \xa0 (Child)'
'12 Years to 90 Years \xa0 (Child, Adult, Older Adult)'
'19 Years to 55 Years \xa0 (Adult)'
'1 Month and older \xa0 (Child, Adult, Older Adult)'
'3 Years to 16 Years \xa0 (Child)'
'18 Years to 114 Years \xa0 (Adult, Older Adult)'
'35 Years to 100 Years \xa0 (Adult, Older Adult)'
'4 Years to 6 Years \xa0 (Child)' '7 Years to 9 Years \xa0 (Child)'
'12 Years to 98 Years \xa0 (Child, Adult, Older Adult)'
'20 Years to 75 Years \xa0 (Adult, Older Adult)'
'24 Years to 60 Years \xa0 (Adult)' '18 Years to 43 Years \xa0 (Adult)'
'18 Years to 24 Years \xa0 (Adult)'
'24 Years to 80 Years \xa0 (Adult, Older Adult)'
'up to 13 Years \xa0 (Child)'
'23 Years to 75 Years \xa0 (Adult, Older Adult)'
'14 Years to 70 Years \xa0 (Child, Adult, Older Adult)'
'50 Years to 110 Years \xa0 (Adult, Older Adult)'
'18 Years to 51 Years \xa0 (Adult)'
'22 Years to 65 Years \xa0 (Adult, Older Adult)'
'20 Years to 59 Years \xa0 (Adult)'
'30 Years to 75 Years \xa0 (Adult, Older Adult)'
'15 Years to 17 Years \xa0 (Child)'
'50 Years to 89 Years \xa0 (Adult, Older Adult)'
'14 Years to 80 Years \xa0 (Child, Adult, Older Adult)'
'65 Years to 100 Years \xa0 (Older Adult)'
'10 Years to 17 Years \xa0 (Child)' '48 Years to 60 Years \xa0 (Adult)'
'45 Years to 80 Years \xa0 (Adult, Older Adult)'
'20 Years to 55 Years \xa0 (Adult)'
'44 Weeks to 18 Years \xa0 (Child, Adult)'
'6 Years to 17 Years \xa0 (Child)'
'13 Years to 25 Years \xa0 (Child, Adult)'
'5 Years to 75 Years \xa0 (Child, Adult, Older Adult)'
'6 Months to 9 Months \xa0 (Child)'
'20 Years to 69 Years \xa0 (Adult, Older Adult)'
'9 Months and older \xa0 (Child, Adult, Older Adult)'
'9 Years to 12 Years \xa0 (Child)' '21 Years to 58 Years \xa0 (Adult)'
'up to 72 Hours \xa0 (Child)'
'1 Year to 95 Years \xa0 (Child, Adult, Older Adult)'
'3 Years to 13 Years \xa0 (Child)'
'30 Years to 66 Years \xa0 (Adult, Older Adult)'
'45 Years to 65 Years \xa0 (Adult, Older Adult)'
'22 Years and older \xa0 (Adult, Older Adult)'
'35 Years to 80 Years \xa0 (Adult, Older Adult)'
'14 Years to 18 Years \xa0 (Child, Adult)'
'18 Years to 22 Years \xa0 (Adult)'
'10 Years to 18 Years \xa0 (Child, Adult)'
'4 Years to 7 Years \xa0 (Child)'
'35 Years and older \xa0 (Adult, Older Adult)'
'65 Years to 85 Years \xa0 (Older Adult)'
```

```
'13 Years to 21 Years \xa0 (Child, Adult)'
'3 Years to 6 Years \xa0 (Child)'
'74 Years to 95 Years \xa0 (Older Adult)'
'10 Years to 19 Years \xa0 (Child, Adult)'
'8 Years to 24 Years \xa0 (Child, Adult)'
'13 Years to 24 Years \xa0 (Child, Adult)'
'16 Years to 75 Years \xa0 (Child, Adult, Older Adult)'
'85 Years and older \xa0 (Older Adult)'
'18 Years to 25 Years \xa0 (Adult)'
'18 Years to 72 Years \xa0 (Adult, Older Adult)'
'17 Years to 81 Years \xa0 (Child, Adult, Older Adult)'
'50 Years to 70 Years \xa0 (Adult, Older Adult)'
'65 Years to 80 Years \xa0 (Older Adult)'
'up to 80 Years \xa0 (Child, Adult, Older Adult)'
'3 Years to 11 Years \xa0 (Child)' '8 Years to 15 Years \xa0 (Child)'
'31 Days and older \xa0 (Child, Adult, Older Adult)'
'13 Years to 99 Years \xa0 (Child, Adult, Older Adult)'
'6 Months to 6 Years \xa0 (Child)' '8 Years to 13 Years \xa0 (Child)'
'up to 110 Years \xa0 (Child, Adult, Older Adult)'
'1 Year to 10 Years \xa0 (Child)' 'up to 24 Months \xa0 (Child)'
'12 Years to 75 Years \xa0 (Child, Adult, Older Adult)'
'21 Years to 75 Years \xa0 (Adult, Older Adult)'
'11 Years to 18 Years \xa0 (Child, Adult)'
'11 Years to 17 Years \xa0 (Child)' '5 Years to 12 Years \xa0 (Child)'
'10 Years to 75 Years \xa0 (Child, Adult, Older Adult)'
'up to 16 Years \xa0 (Child)' '13 Years to 18 Years \xa0 (Child, Adult)'
'2 Years to 17 Years \xa0 (Child)'
'16 Years to 25 Years \xa0 (Child, Adult)'
'3 Years to 8 Years \xa0 (Child)'
'16 Years to 67 Years \xa0 (Child, Adult, Older Adult)'
'13 Years to 16 Years \xa0 (Child)' '20 Years to 30 Years \xa0 (Adult)'
'50 Years to 100 Years \xa0 (Adult, Older Adult)'
'60 Years to 95 Years \xa0 (Adult, Older Adult)'
'50 Years to 65 Years \xa0 (Adult, Older Adult)'
'17 Years to 40 Years \xa0 (Child, Adult)'
'4 Years to 12 Years \xa0 (Child)'
'12 Years to 18 Years \xa0 (Child, Adult)'
'5 Years to 85 Years \xa0 (Child, Adult, Older Adult)'
'20 Years to 99 Years \xa0 (Adult, Older Adult)'
'11 Years to 15 Years \xa0 (Child)' '20 Years to 40 Years \xa0 (Adult)'
'19 Years to 30 Years \xa0 (Adult)'
'52 Years and older \xa0 (Adult, Older Adult)'
'16 Years to 49 Years \xa0 (Child, Adult)'
'6 Years to 80 Years \xa0 (Child, Adult, Older Adult)'
'21 Years to 99 Years \xa0 (Adult, Older Adult)'
'21 Years to 84 Years \xa0 (Adult, Older Adult)'
'4 Years to 15 Years \xa0 (Child)'
'35 Years to 70 Years \xa0 (Adult, Older Adult)'
'55 Years to 80 Years \xa0 (Adult, Older Adult)'
'3 Years to 7 Years \xa0 (Child)' '21 Months to 47 Months \xa0 (Child)'
'18 Months to 48 Months \xa0 (Child)' '4 Years to 8 Years \xa0 (Child)'
'9 Years to 18 Years \xa0 (Child, Adult)'
'7 Years to 11 Years \xa0 (Child)'
'7 Years to 80 Years \xa0 (Child, Adult, Older Adult)'
'40 Years to 100 Years \xa0 (Adult, Older Adult)'
'4 Years to 10 Years \xa0 (Child)'
'16 Years to 50 Years \xa0 (Child, Adult)'
'55 Years to 79 Years \xa0 (Adult, Older Adult)'
'7 Years to 17 Years \xa0 (Child)' '12 Years to 16 Years \xa0 (Child)'
'48 Years to 58 Years \xa0 (Adult)'
```

```
 '60 Years to 85 Years \xa0 (Adult, Older Adult)'
 '55 Years to 110 Years \xa0 (Adult, Older Adult)'
 '17 Years to 20 Years \xa0 (Child, Adult)' 'up to 5 Years \xa0 (Child)'
 '6 Years to 12 Years \xa0 (Child)'
 '15 Years to 19 Years \xa0 (Child, Adult)' 'up to 1 Year \xa0 (Child)'
 '21 Years to 45 Years \xa0 (Adult)' '2 Years to 14 Years \xa0 (Child)'
 'up to 2 Months \xa0 (Child)' '11 Years to 16 Years \xa0 (Child)'
 '15 Years to 49 Years \xa0 (Child, Adult)'
 '70 Years to 89 Years \xa0 (Older Adult)'
 '6 Years to 7 Years \xa0 (Child)' '18 Months to 36 Months \xa0 (Child)'
 '45 Years to 59 Years \xa0 (Adult)' '6 Months to 17 Years \xa0 (Child)'
 '8 Years to 60 Years \xa0 (Child, Adult)'
 '35 Years to 85 Years \xa0 (Adult, Older Adult)'
 '62 Years and older \xa0 (Adult, Older Adult)'
 '6 Months to 24 Months \xa0 (Child)' '18 Years to 34 Years \xa0 (Adult)'
 '12 Years to 22 Years \xa0 (Child, Adult)'
 '3 Months to 18 Years \xa0 (Child, Adult)'
 '8 Years to 20 Years \xa0 (Child, Adult)']
Toplam Benzersiz Değer Sayısı: 415
```

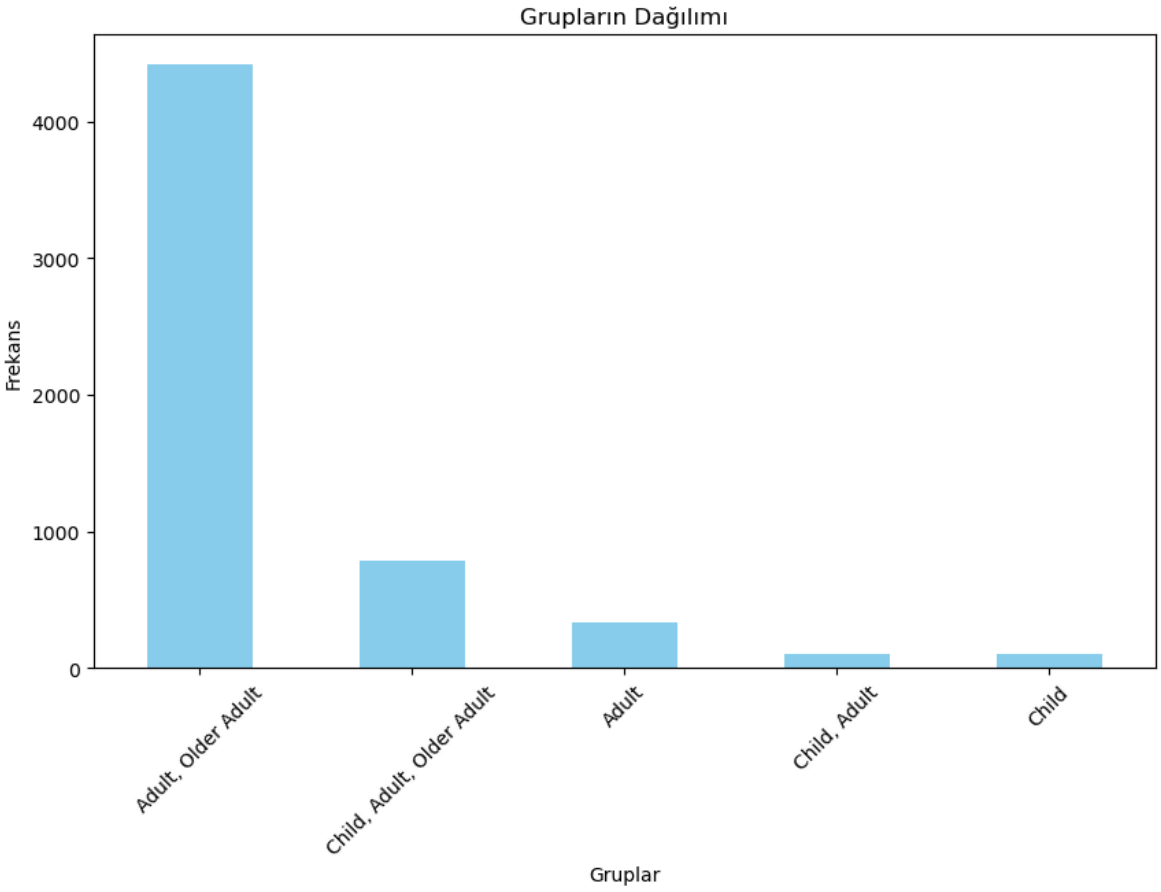
In [30]:
```
df['Age'] = df['Age'].str.replace('\xa0', '', regex=True)
```

In [48]:
```
def create_group_label(age_text):
    categories = []
    if pd.isna(age_text):
        return 'Unknown'
    if 'Child' in age_text:
        categories.append('Child')
    if 'Adult' in age_text:
        categories.append('Adult')
    if 'Older Adult' in age_text:
        categories.append('Older Adult')
    return ', '.join(categories)

df['Group Label'] = df['Age'].apply(create_group_label)
```

In [49]:
```
group_counts = df['Group Label'].value_counts()

print(group_counts)
```

```
Group Label
Adult, Older Adult           4413
Child, Adult, Older Adult     785
Adult                         335
Child, Adult                  104
Child                         101
Name: count, dtype: int64
```

In [50]:
```
group_counts.plot(kind='bar', figsize=(10, 6), color='skyblue')
plt.title('Grupların Dağılımı')
plt.xlabel('Gruplar')
plt.ylabel('Frekans')
plt.xticks(rotation=45)
plt.show()
```



6. Bivariate Analysis

In [52]:
```
df_filtered = df[df['Phases'] != 'Unknown']
```

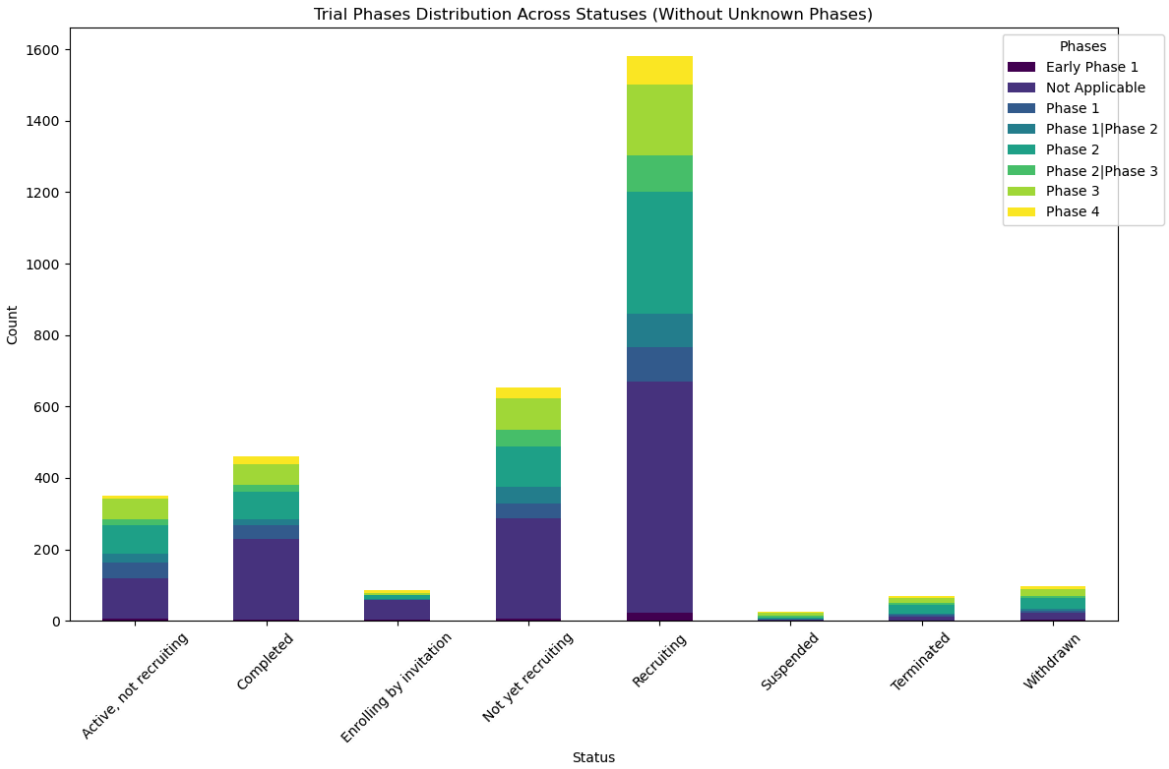
In [53]:
```
status_phase_counts = df_filtered.groupby(['Status', 'Phases']).size().reset_ind
pivot_table = status_phase_counts.pivot(index='Status', columns='Phases', values

print(pivot_table)
```

Phases	Early Phase 1	Not Applicable	Phase 1	\
Status				
Active, not recruiting	7.00	111.00	44.00	
Completed	3.00	226.00	38.00	
Enrolling by invitation	4.00	54.00	1.00	
Not yet recruiting	5.00	282.00	42.00	
Recruiting	22.00	647.00	98.00	
Suspended	2.00	2.00	0.00	
Terminated	0.00	13.00	4.00	
Withdrawn	3.00	19.00	7.00	

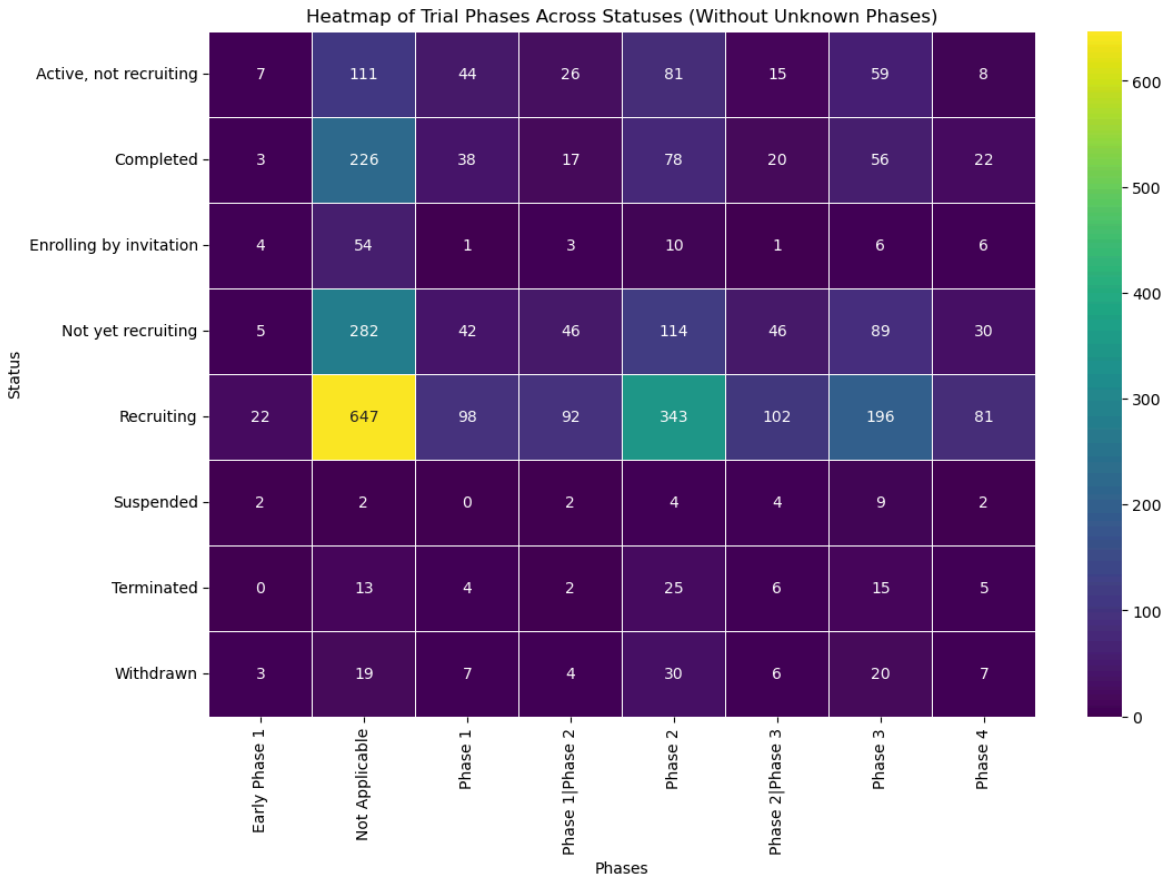
Phases	Phase 1|Phase 2	Phase 2	Phase 2|Phase 3	Phase 3	\
Status					
Active, not recruiting	26.00	81.00	15.00	59.00	
Completed	17.00	78.00	20.00	56.00	
Enrolling by invitation	3.00	10.00	1.00	6.00	
Not yet recruiting	46.00	114.00	46.00	89.00	
Recruiting	92.00	343.00	102.00	196.00	
Suspended	2.00	4.00	4.00	9.00	
Terminated	2.00	25.00	6.00	15.00	
Withdrawn	4.00	30.00	6.00	20.00	

Phases	Phase 4
Status	
Active, not recruiting	8.00
Completed	22.00
Enrolling by invitation	6.00
Not yet recruiting	30.00
Recruiting	81.00
Suspended	2.00
Terminated	5.00
Withdrawn	7.00

In [54]:
```
pivot_table.plot(kind='bar', stacked=True, figsize=(12, 8), colormap='viridis')
plt.title('Trial Phases Distribution Across Statuses (Without Unknown Phases)')
plt.xlabel('Status')
plt.ylabel('Count')
plt.legend(title='Phases', bbox_to_anchor=(1.05, 1))
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



In [55]:
```
plt.figure(figsize=(12, 8))
sns.heatmap(pivot_table, annot=True, fmt='g', cmap='viridis', linewidths=0.5)
plt.title('Heatmap of Trial Phases Across Statuses (Without Unknown Phases)')
plt.xlabel('Phases')
plt.ylabel('Status')
plt.show()
```



In [56]:
```
# Conditions vs. Outcome Measures: Understand the common outcome measures for di
```

```
conditions_outcomes = df.groupby('Conditions')['Outcome Measures'].apply(lambda
print(conditions_outcomes)
```

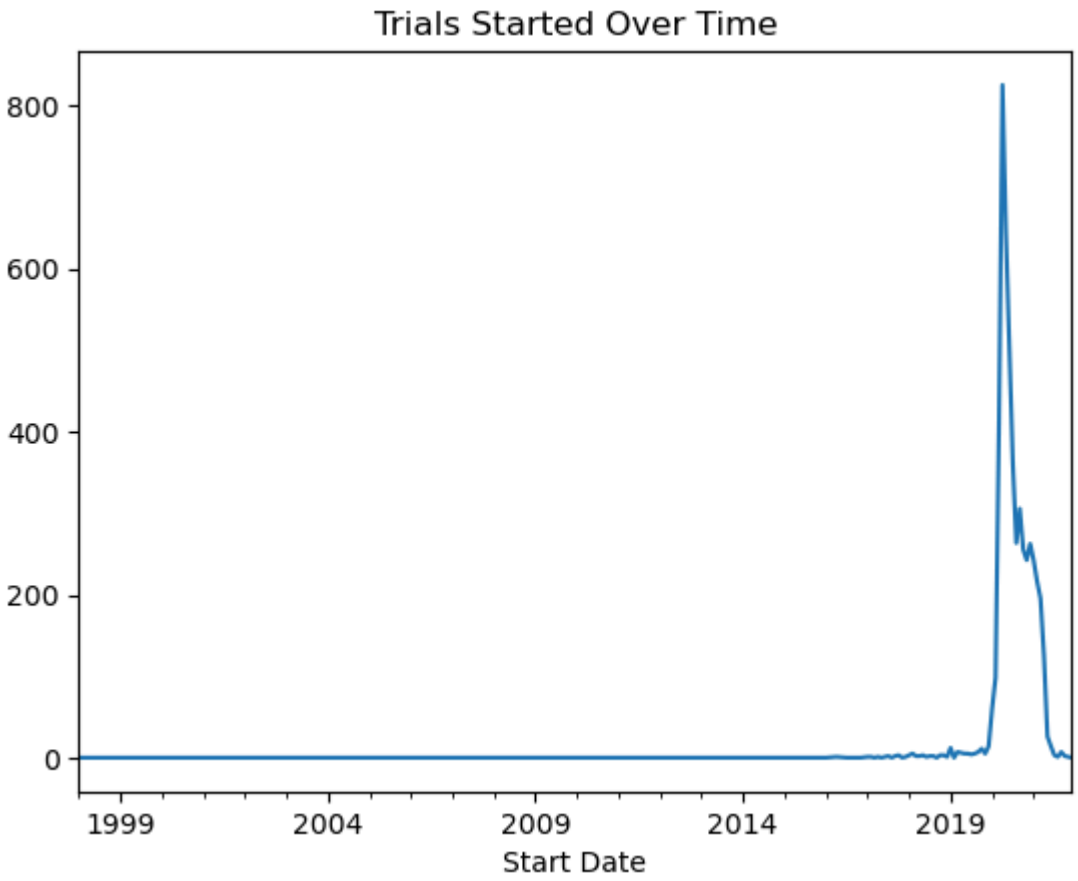
```
                                            Conditions  \
0                                2019 Novel Coronavirus
1                      2019 Novel Coronavirus Infection
2     2019 Novel Coronavirus Infection|COVID-19 Viru...
3                      2019 Novel Coronavirus Pneumonia
4              2019 Novel Coronavirus Pneumonia|COVID-19
...                                                 ...
3034                  the Lung Complication of COVID-19
3035  the Prognostic Value of Ferritin|Glycosylated ...
3036  the Study Focus on the Uses of Telephone and O...
3037  the Use of Modern Technology Applications in H...
3038  to Predict an Unfavorable Evolution of Covid-1...

                                         Outcome Measures
0     Proportion of participants who improve by at l...
1     new-onset COVID-19|Number of Participants with...
2     Number of participants with treatment emergent...
3     Clinical recovery time|Complete fever time|Cou...
4     Pneumonia severity index|Oxygenation index (Pa...
...                                                 ...
3034  lung injury score|Angiotensin 1-7 (Ang 1-7) ch...
3035  assessment of the prognostic value of ferritin...
3036  - To provide an overview about the pros and co...
3037  rate of reassurance delivered from doctors to ...
3038  Need of mechanical ventilation, transfer to an...

[3039 rows x 2 columns]
```

7. Time Series Analysis

Analyze the trends over time, such as the number of trials started over the months.

In [66]:
```
# Convert date columns to datetime
df['Start Date'] = pd.to_datetime(df['Start Date'], errors='coerce')
df['Primary Completion Date'] = pd.to_datetime(df['Primary Completion Date'], er
# Plot the number of trials started over time
df['Start Date'].dt.to_period('M').value_counts().sort_index().plot(kind= 'line'
```

Out[66]:
```
<Axes: title={'center': 'Trials Started Over Time'}, xlabel='Start Date'>
```



8. Conclusion

Summarize the findings from your EDA. For example:

● The majority of trials are in the "Completed" phase.

● Most trials target adult populations.

● There's a steady increase in the number of trials over time.

10. Output and Visuals

After running the code, you should observe:

● Bar charts showing the distribution of trial statuses, phases, and age groups.

● A time series plot illustrating the trend of trials over time. This project will provide a solid foundation in EDA using Pandas, with practical insights into the clinical trials landscape for COVID-19.