Project Title - COVID-19 Clinical   Trials EDA Pandas

Tool-PYTHON

Dataset: Dataset is available in the given link. You can download it at your convenience.

https://drive.google.com/file/d/1WessMr8tdADxk04uVT-AaoP3_CxwJIYR/view

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.

DATA ANALYSIS LINK

https://colab.research.google.com/drive/1jlLLTwzLE8VMmLjnnWNyS4l9KtyfplRg?usp=sharing

Rank	NCT Number	Title	Acronym	Status	Study Res	Condition	Interventi	Outcome	Sponsor/(	Gender	Age	Phases	Enrollmen	Funded B(	Study Typ	Study Des	Other IDs	Start Date	Primary C(	Completi(	First Post(	Results Fi	Last Upda	Locations	Study Doc	URL
1	NCT04785898	Diagnostic Performance of t	COVID-ID	Active, no	No Result	Covid19	Diagnosti	Evaluate t	Groupe H	All	18 Years a	Not Applic	1000	Other	Interventi	Allocation	COVID-IDNow	November	December	April 30, 2	March 8, 2021		March 8, 2	Groupe Hospitalier F	https://ClinicalTrials.gov/show/NCT04785898
2	NCT04595136	Study to Evaluate the Efficac	COVID-19	Not yet rec	No Result	SARS-CoV	Drug: Dru	Change or	United Me	All	18 Years a	Phase 1	F	60	Other	Interventi	Allocation	COVID19-0001	November	December	January 25	October 20, 2020		October 2	Cimedical, Barranqu	https://ClinicalTrials.gov/show/NCT04595136
3	NCT04395482	Lung CT Scan Analysis of SA	TAC-COVI	Recruiting	No Result	covid19	Other: Lur	A qualitati	University	All	18 Years and older	À	500	Other	Observati	Observati	TAC-COVID19	May 7, 202	June 15, 2	June 15, 2	May 20, 2020		November	Ospedale Papa Giov	https://ClinicalTrials.gov/show/NCT04395482
4	NCT04416061	The Role of a Private Hospita	COVID-19	Active, no	No Result	COVID		Diagnosti	Proportio	Hong Kon	All	Child, Adult, Older A	2500	Industry	Observati	Observati	RC-2020-08	May 25, 20	July 31, 20	August 31	June 4, 2020		June 4, 20	Hong Kong Sanatorii	https://ClinicalTrials.gov/show/NCT04416061
5	NCT04395924	Maternal-foetal Transmissio	TMF-COVI	Recruiting	No Result	Maternal F	Diagnosti	COVID-19	Centre Ho	Female	18 Years to 48 Years	50	Other	Observati	Observati	CHRO-2020-10	May 5, 202	May-21	May-21	May 20, 2020		June 4, 20	CHR OrlÃ©ans, OrlÃ	https://ClinicalTrials.gov/show/NCT04395924
6	NCT04516954	Convalescent Plasma for CC	CPCP	Enrolling l	No Result	COVID 19	Biological	Evaluate t	Vinmec R	All	18 Years t	Early Phas	10	Other	Interventi	Allocation	ISC.20.11.1	August 1,	November	December	August 18, 2020		August 20	Vinmec Research Ins	https://ClinicalTrials.gov/show/NCT04516954
7	NCT04476940	COVID-19 Breastfeeding Gu	COVID-BF	Not yet rec	No Result	Covid19	Behaviora	COVID-19	Meharry M	Female	18 Years t	Not Applic	200	Other	Interventi	Allocation		Sep-20	Oct-21	Jun-22	July 20, 2020		July 20, 20	Meharry Medical Coll	https://ClinicalTrials.gov/show/NCT04476940
8	NCT04634214	The Severity of COVID 19 in [	COVID19	Not yet rec	No Result	Covid19	Type2 Diab	Severity of	India Diab	All	18 Years to 99 Years	1050	Other	Observati	Observati	IDRFARH015	November	February	May 16, 20	November 18, 2020		November	Orthomed Hospital,	https://ClinicalTrials.gov/show/NCT04634214
9	NCT04602884	Early Detection of COVID-19	COVID-19	Suspende	No Result	Covid19	Diagnosti	Correlatic	Scentech	All	18 Years t	Not Applic	50	Industry	Interventi	Allocation	Cov-2-IDF	Septembe	December	December	October 26, 2020		April 13, 2	IDF COVID 19 Isolati	https://ClinicalTrials.gov/show/NCT04602884
10	NCT04384588	COVID19-Convalescent Plas	FALP-COV	Recruiting	No Result	COVID-19	Biological	in-hospit:	Fundacio	All	15 Years a	Phase 2	F	100	Other	Interventi	Allocation	FALP 001-2020	April 7, 20	April 6, 20	April 6, 20	May 12, 2020		May 12, 20	Fundacion Arturo Lo	https://ClinicalTrials.gov/show/NCT04384588
11	NCT04355897	CoVID-19 Plasma in Treatment of COVI	Recruiting	No Result	COVID 19	Biological	Reduce m	The Christ	All	18 Years t	Early Phas	100	Other	Interventi	Allocation	20-23	April 28, 2	Jul-20	Aug-20	April 21, 2020		May 20, 20	The Christ Hospital,	https://ClinicalTrials.gov/show/NCT04355897
12	NCT04412265	Frailty in Elderly Patients Wit	FRA-COVI	Recruiting	No Result	Covid19	Other: Rel	Developm	University	All	18 Years and older	À	300	Other	Observati	Observati	FRA-COVID	April 16, 2	March 1, 2	March 1, 2	June 2, 2020		January 22	Dipartimento di Geri	https://ClinicalTrials.gov/show/NCT04412265
13	NCT04659759	COVID-19 Pregnancy Relate	COVID-PF	Recruiting	No Result	Covid19	Other: CO	Maternal (	Thomas Je	Female	Child, Adult, Older A	300	Other	Observati	Observati	20F.1043	November	December	June 30, 2	December 9, 2020		March 5, 2	Thomas Jefferson Un	https://ClinicalTrials.gov/show/NCT04659759
14	NCT04427332	Smell and Taste Disorders in COVID-19	Complete	No Result	covid19	Other: Inv	Identificat	University	All	18 Years and older	À	376	Other	Observati	Observati	COVID-19 ORL	June 11, 2	October 1	October 1	June 11, 2020		January 27	ASST Monza-Ospeda	https://ClinicalTrials.gov/show/NCT04427332
15	NCT04842708	Evaluation of Anti-COVID 19	COVID-19	Recruiting	No Result	Covid19	Diagnosti	Associati	Scentech	All	18 Years t	Not Applic	50	Industry	Interventi	Allocation	Cov-2-SMC-V-2	December	October 2	October 2	April 13, 2021		April 13, 2	Shamir Medical Cent	https://ClinicalTrials.gov/show/NCT04842708
16	NCT04367805	COVID-19 Infection in Patien	COVID19-	Recruiting	No Result	Hepatoce	Diagnosti	Incidence	Centre Ho	All	18 Years and older	À	50	Other	Observati	Observati	PI2020_843_00	April 27, 2	Oct-20	Oct-20	April 29, 2020		April 29, 2	CHU Amiens, Amien	https://ClinicalTrials.gov/show/NCT04367805
17	NCT04609969	Diagnostic Performance Eva	COVID-VII	Complete	No Result	Covid19	Diagnosti	Evaluation	Centre Ho	All	18 Years a	Not Applic	200	Other	Interventi	Allocation	CHRO-2020-18	October 1	October 1	October 1	October 30, 2020		December	Centre Hospitalier R	https://ClinicalTrials.gov/show/NCT04609969
18	NCT04817553	Impact of COVID-19 on the C	IgG4-COV	Recruiting	No Result	IgG4 Relat	Other: exp	Incidence	Chinese L	All	18 Years and older	À	150	Other	Observati	Observati	IgG4-COVID	March 24,	Jul-21	Oct-21	March 26, 2021		March 26,	Prince of Wales Hos;	https://ClinicalTrials.gov/show/NCT04817553
19	NCT04632355	Musculoskeletal Pain in Pati	Covid19-1	Not yet rec	No Result	Covid19	Other: Inte	Self-repor	Fondazior	All	18 Years and older	À	100	Other	Observati	Observati	Covid19-Pain	December	July 31, 20	July 31, 20	November 17, 2020		November	Jorge Hugo VillafÃ	https://ClinicalTrials.gov/show/NCT04632355
20	NCT04407585	Testing the Accuracy of a Digital Test to	Recruiting	No Result	Covid-19	Diagnosti	SARS-CoV	King's Col	All	18 Years and older	À	1000000	Other	Observati	Observati	Covid-19 Valida	June 1, 20	May 10, 20	May 10, 20	May 29, 2020		June 24, 2	King's Col "Statistici	https://ClinicalTrials.gov/show/NCT04407585
21	NCT04657510	Femoral frACturEs and COVI	FACE COV	Recruiting	No Result	Covid19	Other: CO	Survival at	Istituto O	All	45 Years and older	À	140	Other	Observati	Observati	FACE COVID-19	November	December	December	December 8, 2020		December	IRCCS Istituto Ortop	https://ClinicalTrials.gov/show/NCT04657510
22	NCT04473170	Study Evaluating the Safety a	SENTAD-(	Complete	No Result	Coronavir	Biological	Adverse re	Abu Dhab	All	18 Years a	Phase 1	F	146	Other	Interventi	Allocation	CT.001.1.0.SEN	April 4, 20	May 20, 20	July 14, 20	July 16, 2020		July 16, 20	Abu Dhabi Stem Cell	https://ClinicalTrials.gov/show/NCT04473170
23	NCT04497298	Clinical Trial to Evaluate the S	COVID-19	Active, no	No Result	COVID-19	Biological	To assess	Institut Pa	All	18 Years t	Phase 1		90	Industry/(	Interventi	Allocation	2020-016|202(	August 10	Jun-21	Jul-21	August 4, 2020		April 13, 2	SGS Life Sciences, C	https://ClinicalTrials.gov/show/NCT04497298
24	NCT04549831	Genetic Bases of COVID-19	GEN-COV	Recruiting	No Result	COVID-19	Genetic: N	To identif	University	All	18 Years and older	À	2000	Other	Observati	Observati		16917	April 8, 20	April 8, 20	April 8, 20	September 16, 2020		Septembe	University of Siena, S	https://ClinicalTrials.gov/show/NCT04549831
25	NCT04706390	Comparison of the Immune	COVID19-	Recruiting	No Result	Covid19	Biological	immune re	University	All	20 Years to 100 Year	2500	Other	Observati	Observati	COVID-19vacc-	January 12	January 12	January 12	January 12, 2021		January 12	University of Bergen,	https://ClinicalTrials.gov/show/NCT04706390
26	NCT04632719	The MentalPlusÂ® for Assess	MP-COVI	Recruiting	No Result	Cognitive	Device: Th	Assessme	University	All	8 Years to	Not Applic	200	Other	Interventi	Allocation	MentalPlusÂ	November	December	December	November 17, 2020		November	Livia Stocco Sanche	https://ClinicalTrials.gov/show/NCT04632719
27	NCT04427345	Predictive Factors COVID-19	MI-COVID	Recruiting	No Result	covid19	Other: Pre	Identify ris	University	All	18 Years and older	À	1000	Other	Observati	Observati	MI-COVID	April 30, 2	April 30, 2	April 30, 2	June 11, 2020		November	ASST Monza-Osped:	https://ClinicalTrials.gov/show/NCT04427345
28	NCT04424004	MURDOCK Cabarrus County	C3PI	Active, no	No Result	COVID 19	Other: CO	Estimate t	Duke Univ	All	18 Years and older	À	500	Other	Observati	Observati	Pro00105703	June 9, 20	June 30, 2	June 30, 2	June 9, 2020		November	Duke CTSI Translatic	https://ClinicalTrials.gov/show/NCT04424004
29	NCT04407494	Diagnostic Value of Patient -	COVID-Ol	Complete	No Result	COVID-19	Biological	Diagnosti	University	All	18 Years and older	À	809	Other	Observati	Observati	RECHMPL20_01	March 1, 2	April 1, 20	April 30, 2	May 29, 2020		June 1, 20	Uhmontpellier, Mont	https://ClinicalTrials.gov/show/NCT04407494
30	NCT04360811	Assessment of Obstetric, Fe	COroFet	Recruiting	No Result	Pregnanc	Diagnosti	number o	University	Female	18 Years a	Not Applic	3600	Other	Interventi	Allocation	RC31/20/0123	April 17, 2	Apr-21	Apr-22	April 24, 2020		April 24, 2	University Hospital o	https://ClinicalTrials.gov/show/NCT04360811
31	NCT04403672	Performance Evaluation of RealDetect	RecDetect	Recruiting	No Result	High Sens	Device: pe	Performar	Banglade:	All	5 Years to 65 Years	À	120	Other	Observati	Observati	30005052020	May 18, 20	July 24, 20	July 24, 20	May 27, 2020		May 28, 2	Institute of Epidemic	https://ClinicalTrials.gov/show/NCT04403672
32	NCT04363099	Multicentre Observational S	COVID AN	Recruiting	No Result	Covid-19	Other: obs	clinical de	Maison de	All	up to 99 Years À (Ch	1000	Other	Observati	Observati	2020-A01059-:	March 1, 2	June 30, 2	December	April 27, 2020		April 27, 2	Msp de Creil, Creil, Fr	https://ClinicalTrials.gov/show/NCT04363099
33	NCT04377425	COVID-19 Prevalence and Co	Neuro-Co	Withdraw	No Result	Neurologi	Diagnosti	Prevalenc	Aarhus Ur	All	18 Years and older	À	0	Other	Observati	Observati	Neuro-Covid-19	May 7, 202	November	June 30, 2	May 6, 2020		Septembe	Aalborg University H	https://ClinicalTrials.gov/show/NCT04377425
34	NCT04602871	Early Detection of COVID-19	COVID-19	Complete	No Result	Covid19	Diagnosti	Correlatic	Scentech	All	18 Years t	Not Applic	108	Industry	Interventi	Allocation	Cov-2-2020	July 3, 202	January 30	January 30	October 26, 2020		April 13, 2	Soroka Medical Cent	https://ClinicalTrials.gov/show/NCT04602871
35	NCT04691895	Gastrointestinal Symptoms in COVID-1	Active, no	No Result	COVID19	Other: SAI	Prevalenc	University	All	18 Years to 85 Years	2000	Other	Observati	Observati	GI-COVID19_01	May 1, 202	Septembe	December	December 31, 2020		January 5,	IRCCS Azienda Ospe	https://ClinicalTrials.gov/show/NCT04691895
36	NCT04645498	COVID-19 and Hereditary Me	COVID19-	Not yet rec	No Result	Covid19	Metabolis	Frequenc	University	All	Child, Adult, Older A	77	Other	Observati	Observati	2020_52	2020	Jan-21	Jan-23	Jan-23	November 27, 2020		November 27, 2020		https://ClinicalTrials.gov/show/NCT04645498
37	NCT04747821	An Effectiveness Study of th	Projeto S	Active, no	No Result	Covid19	Biological	Incidence	Butantan	All	18 Years a	Phase 4		27711	Other	Interventi	Allocation	COV-04-IB	February 7	May-21	Feb-22	February 10, 2021		April 8, 20	Hospital Estadual de	https://ClinicalTrials.gov/show/NCT04747821
28	NCT04272004	Comparison of the Efficacy (	CATCH C(	Recruiting	No Result	COVID-19	Diagnosti	detection	Texas Car	All	18 Years t	Phase 1		100	Other	Interventi	Allocation	TCAL CATCH C(	May 9, 202	May-21	Jun-21	May 1, 2020		August 11	Texas Cardiac Arrhyt	https://ClinicalTrials.gov/show/NCT04272004

2. Importing Libraries and Loading Data

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [12]

```
df = pd.read_csv('covid_clinical_trials.csv')  # change name if
needed
```

In [13]

3.Initial Data Exploration

```
df.head()
```

HTML Output

Rank	NCTNumber	Title	Acronym	Status	StudyResults	Conditions	Interventions	OutcomeMeasures	Sponsor/Collaborators	...	OtherIDs	StartDate	PrimaryCompletionDate	CompletionDate	FirstPosted	ResultsFirstPosted	LastUpdatePosted	Locations	StudyDocuments	URL
0 1	NCT04785898	Diagnostic Performance of the IDNow™	COVID-IDNow	Active, not recruiting	NoResultsAvailable	Covid 19	Diagnostic Test: IDNow™ COVID-19 Screening	Evaluate the diagnostic performance of the ID...	Groupe Hospitalier Paris Saint Joseph	...	COVID-IDNow	November 9, 2020	December 22, 2020	April 30, 2021	March 8, 2021	NaN	March 8, 2021	Groupe Hospitalier Paris Saint - Jos	NaN	https://ClinicalTrials.gov/show/NCT04785898

Rank	NCTNumber	Title	Acronym	Status	StudyResults	Conditions	Interventions	OutcomeMeasures	Sponsor/Collaborators	..	OtherIDs	StartDate	PrimaryCompletionDate	CompletionDate	FirstPosted	ResultsFirstPosted	LastUpdatePosted	Locations	StudyDocuments	URL
	COVID-19...						Test											eph, Paris, ...		
1 2	NCT0459 5136	Study to Evaluate the Effic ac	COVID-19	Not yet recruitin	No Results Availab	SARS-CoV-2 Infection	Drug: Drug COVID 19-0001-	Change on viral load results from baselin	United Medical Specialties	..	COVID19-0001-U	November 2, 2020	December 15, 2020	January 29, 2021	October 20, 202	NaN	October 20, 2020	Cimedical, Barranquilla	NaN	https://ClinicalTrials.gov/show/NCT04595136

Rank	NCTNumber	Title	Acronym	Status	StudyResults	Conditions	Interventions	OutcomeMeasures	Sponsor/Collaborators	...	OtherIDs	StartDate	PrimaryCompletionDate	CompletionDate	FirstPosted	ResultsFirstPosted	LastUpdatePosted	Locations	StudyDocuments	URL
		y of COVID 19-0001...		g	Ie		USR|Drug: normal saline	e aft...			SR				0			, Atlantico, Colombia		
2 3	NCT0 43 95 48 2	Lung CTS ca	TAC-COVI	Recruitin	NoResults	covid19	Other: Lung CTsc	A qualitative analysis of par	University of Milano Bico	...	TAC-COVI	May7,202	June15,20	June15,20	May20,20	NaN	November	Ospedale P	NaN	https://ClinicalTrials.gov/show/NCT04395482

Rank	NCTNumber	Title	Acronym	Status	StudyResults	Conditions	Interventions	OutcomeMeasures	Sponsor/Collaborators	..	OtherIDs	StartDate	PrimaryCompletionDate	CompletionDate	FirstPosted	ResultsFirstPosted	LastUpdatePosted	Locations	StudyDocuments	URL
		nAnalysis of SARS-CoV2 Induced Lu	D19	g	Available		an analysis in COVID-19 patients	enchymal lung dam...	cca		D19	0	21	21	20		9,2020	apa Giovanni XXIII, Bergamo, Italy| P.		

Rank	NCTNumber	Title	Acronym	Status	StudyResults	Conditions	Interventions	OutcomeMeasures	Sponsor/Collaborators	OtherIDs	StartDate	PrimaryCompletionDate	CompletionDate	FirstPosted	ResultsFirstPosted	LastUpdatePosted	Locations	StudyDocuments	URL
		n...															..		
3 4	NCT04416061	The Role of a Private Hospital in Ho	COVID-19	Active, not recruiting	No Results Available	COVID	Diagnostic Test: COVID19 Diagnostic Te	Proportion of asymptomatic subjects|Proportion...	Hong Kong Sanatorium & Hospital	RC-2020-08	May 25, 2020	July 31, 2020	August 31, 2020	June 4, 2020	NaN	June 4, 2020	HongKong Sanatorium & Hos	NaN	https://ClinicalTrials.gov/show/NCT04416061

Rank	NCTNumber	Title	Acronym	Status	StudyResults	Conditions	Interventions	OutcomeMeasures	Sponsor/Collaborators	...	OtherIDs	StartDate	PrimaryCompletionDate	CompletionDate	FirstPosted	ResultsFirstPosted	LastUpdatePosted	Locations	StudyDocuments	URL
	ngKongAm...						st											pital, HongKong, Ho...		
4 5	NCT0 43 95 92	Maternal -fo	TMF-COV	Recruitin	NoResult	Maternal Fetal Infectio	Diagnostic Test:	COVID-19 by positive	Centre Hospitalier Régiona	..	CHRO-20	May5,202	May202	May202	May20,202	NaN	June4,20	CHROrléa	NaN	https://ClinicalTrials.gov/show/NCT043959

Rank	NCTNumber	Title	Acronym	Status	StudyResults	Conditions	Interventions	OutcomeMeasures	Sponsor/Collaborators	...	OtherIDs	StartDate	PrimaryCompletionDate	CompletionDate	FirstPosted	ResultsFirstPosted	LastUpdatePosted	Locations	StudyDocuments	URL
	4	etal Transmission of SARS-Cov-2	ID-19	g	sAvailable	n Transmission|COVID-19...	Diagnosis of SARS-Cov2 by RT-...	PCR in cord blood and/o...	l d'Orléans|Centred...		20-10	0	1	1	020		20	ns, Orléans, France		24

5 rows × 27 columns

Text Output

```
    Rank   NCT Number
Title  \
0     1  NCT04785898  Diagnostic Performance of the ID Now™
COVID-19...
1     2  NCT04595136  Study to Evaluate the Efficacy of COVID19-
0001...
2     3  NCT04395482  Lung CT Scan Analysis of SARS-CoV2 Induced
Lun...
3     4  NCT04416061  The Role of a Private Hospital in Hong
Kong Am...
4     5  NCT04395924       Maternal-foetal Transmission of
SARS-Cov-2

        Acronym               Status      Study Results  \
0   COVID-IDNow  Active, not recruiting  No Results Available
1      COVID-19      Not yet recruiting  No Results Available
2   TAC-COVID19              Recruiting  No Results Available
3      COVID-19  Active, not recruiting  No Results Available
4   TMF-COVID-19             Recruiting  No Results Available

                                      Conditions  \
0                                         Covid19
1                              SARS-CoV-2 Infection
2                                         covid19
3                                           COVID
4  Maternal Fetal Infection Transmission|COVID-19...

                                   Interventions  \
0   Diagnostic Test: ID Now™ COVID-19 Screening Test
1      Drug: Drug COVID19-0001-USR|Drug: normal saline
2  Other: Lung CT scan analysis in COVID-19 patients
3          Diagnostic Test: COVID 19 Diagnostic Test
4  Diagnostic Test: Diagnosis of SARS-Cov2 by RT-...

                                 Outcome Measures  \
0  Evaluate the diagnostic performance of the ID ...
1  Change on viral load results from baseline aft...
2  A qualitative analysis of parenchymal lung dam...
3  Proportion of asymptomatic subjects|Proportion...
4  COVID-19 by positive PCR in cord blood and / o...

                            Sponsor/Collaborators  ...
Other IDs  \
0          Groupe Hospitalier Paris Saint Joseph  ...
COVID-IDNow
1                      United Medical Specialties  ...
COVID19-0001-USR
2                      University of Milano Bicocca  ...
TAC-COVID19
3                  Hong Kong Sanatorium & Hospital  ...
RC-2020-08
4  Centre Hospitalier Régional d'Orléans|Centre d...  ...
CHRO-2020-10
```

```
            Start Date Primary Completion Date   Completion Date   \
0   November 9, 2020          December 22, 2020     April 30, 2021
1   November 2, 2020          December 15, 2020   January 29, 2021
2         May 7, 2020              June 15, 2021     June 15, 2021
3        May 25, 2020              July 31, 2020   August 31, 2020
4         May 5, 2020                   May 2021          May 2021

            First Posted Results First Posted Last Update Posted  \
0      March 8, 2021                        NaN        March 8, 2021
1   October 20, 2020                        NaN     October 20, 2020
2       May 20, 2020                        NaN     November 9, 2020
3        June 4, 2020                       NaN         June 4, 2020
4       May 20, 2020                        NaN         June 4, 2020

                                        Locations Study
Documents  \
0  Groupe Hospitalier Paris Saint-Joseph, Paris, ...
NaN
1          Cimedical, Barranquilla, Atlantico, Colombia
NaN
2  Ospedale Papa Giovanni XXIII, Bergamo, Italy|P...
NaN
3  Hong Kong Sanatorium & Hospital, Hong Kong, Ho...
NaN
4                      CHR Orléans, Orléans, France
NaN

                                                    URL
0  https://ClinicalTrials.gov/show/NCT04785898
1  https://ClinicalTrials.gov/show/NCT04595136
2  https://ClinicalTrials.gov/show/NCT04395482
3  https://ClinicalTrials.gov/show/NCT04416061
4  https://ClinicalTrials.gov/show/NCT04395924

[5 rows x 27 columns]
```

In [14]

```
4.BASIC CHECKS

df.shape
```

Text Output

```
(5783, 27)
```

In [15]

```
df.info()
```

Output (stdout)

```
<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
```

```
SAVE CLEAN DATASET
df.to_csv('cleaned_covid_clinical_trials.csv', index=False)
```

```
Shape: (4112, 27)

Missing values:
 Rank                      0
NCT Number                0
Title                     0
Acronym                2229
Status                    0
Study Results             0
Conditions                0
Interventions           645
Outcome Measures         24
Sponsor/Collaborators     0
Gender                    5
```

```
Age                              0
Phases                        1865
Enrollment                      24
Funded Bys                       0
Study Type                       0
Study Designs                   25
Other IDs                        0
Start Date                      24
Primary Completion Date         24
Completion Date                 24
First Posted                     0
Results First Posted          4085
Last Update Posted               0
Locations                      402
Study Documents               3994
URL                              0
dtype: int64
```

```
5.STATUS DISTRIBUTION

status_counts = df['Status'].value_counts()
status_counts
```

HTML Output

		count
Status		
Recruiting	2805	
Completed	1025	
Not yet recruiting	1004	
Active, not recruiting	526	
Enrolling by invitation	181	
Withdrawn	107	

		count
Status		
Terminated	74	
Suspended	27	
Available	19	
No longer available	12	
Approved for marketing	2	
Temporarily not available	1	

dtype: int64

Text Output

```
Status
Recruiting                2805
Completed                 1025
Not yet recruiting        1004
Active, not recruiting     526
Enrolling by invitation    181
Withdrawn                  107
Terminated                  74
Suspended                   27
Available                   19
No longer available         12
Approved for marketing       2
Temporarily not available    1
Name: count, dtype: int64
```
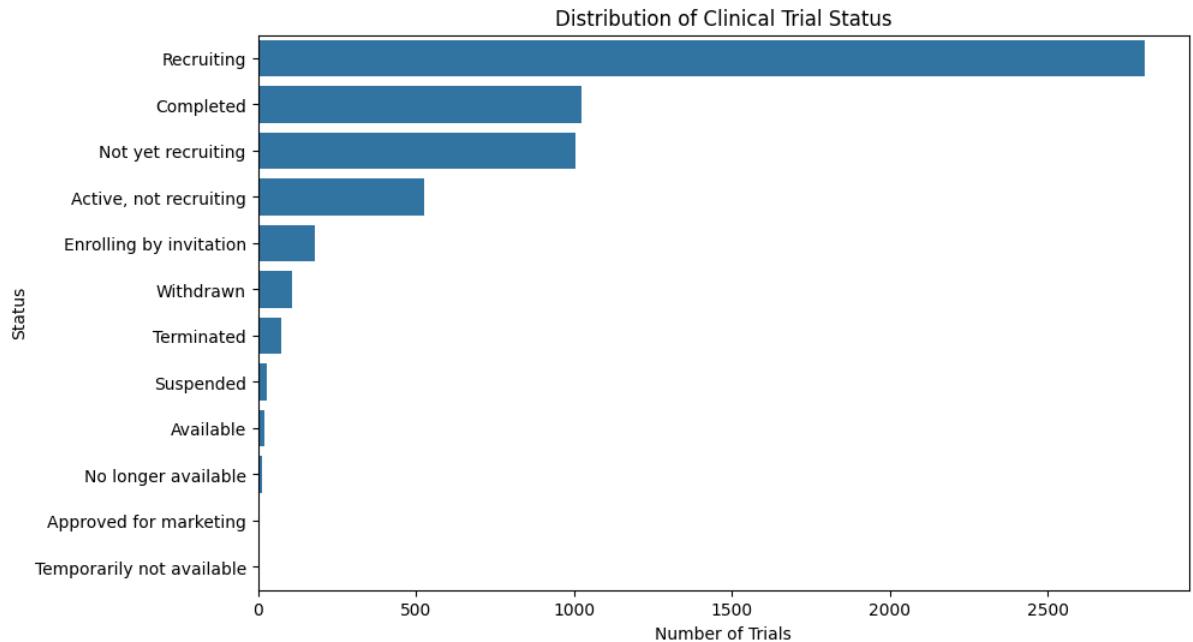
In [17]

```python
plt.figure(figsize=(10,6))
sns.barplot(
    x=status_counts.values,
    y=status_counts.index
)
plt.title('Distribution of Clinical Trial Status')
plt.xlabel('Number of Trials')
plt.ylabel('Status')
plt.show()
```

Text Output

```
<Figure size 1000x600 with 1 Axes>
```

Image (PNG)

Distribution of Clinical Trial Status

In [18]

```
# Extract country from Locations column
df['Country'] = df['Locations'].astype(str).apply(lambda x:
x.split(',')[-1].strip())

# Top 10 countries
top_countries = df['Country'].value_counts().head(10)
top_countries
```

HTML Output

		count
Country		
United States	1267	

		count
Country		
France	647	
nan	585	
United Kingdom	306	
Italy	235	
Spain	234	
Turkey	219	
Canada	202	
Egypt	192	
China	171	

dtype: int64

Text Output

```
Country
United States    1267
France            647
nan               585
United Kingdom    306
Italy             235
Spain             234
Turkey            219
Canada            202
Egypt             192
China             171
Name: count, dtype: int64
```

In [19]
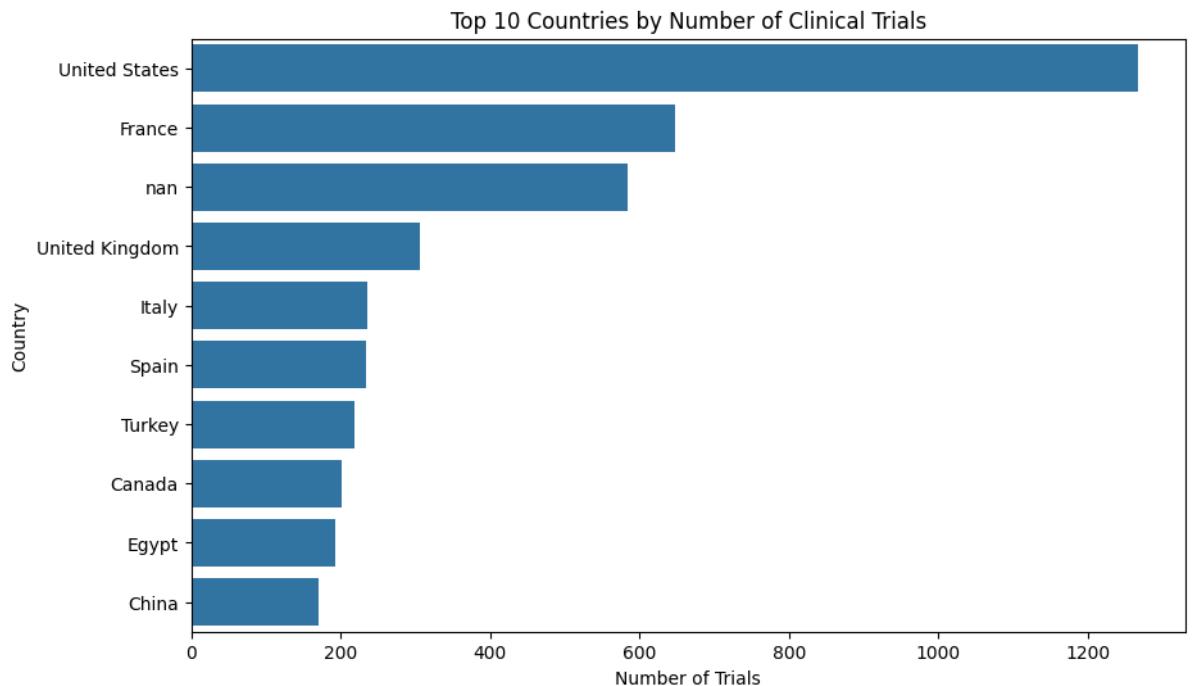
```
plt.figure(figsize=(10,6))
sns.barplot(
    x=top_countries.values,
    y=top_countries.index
```

```
)
plt.title('Top 10 Countries by Number of Clinical Trials')
plt.xlabel('Number of Trials')
plt.ylabel('Country')
plt.show()
```

Text Output

```
<Figure size 1000x600 with 1 Axes>
```

Image (PNG)

In [20]

```
PHASE DISTRIBUTION

phase_counts = df['Phases'].value_counts()
phase_counts
```

HTML Output

	count

Phases	
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

dtype: int64

Text Output

```
Phases
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
```
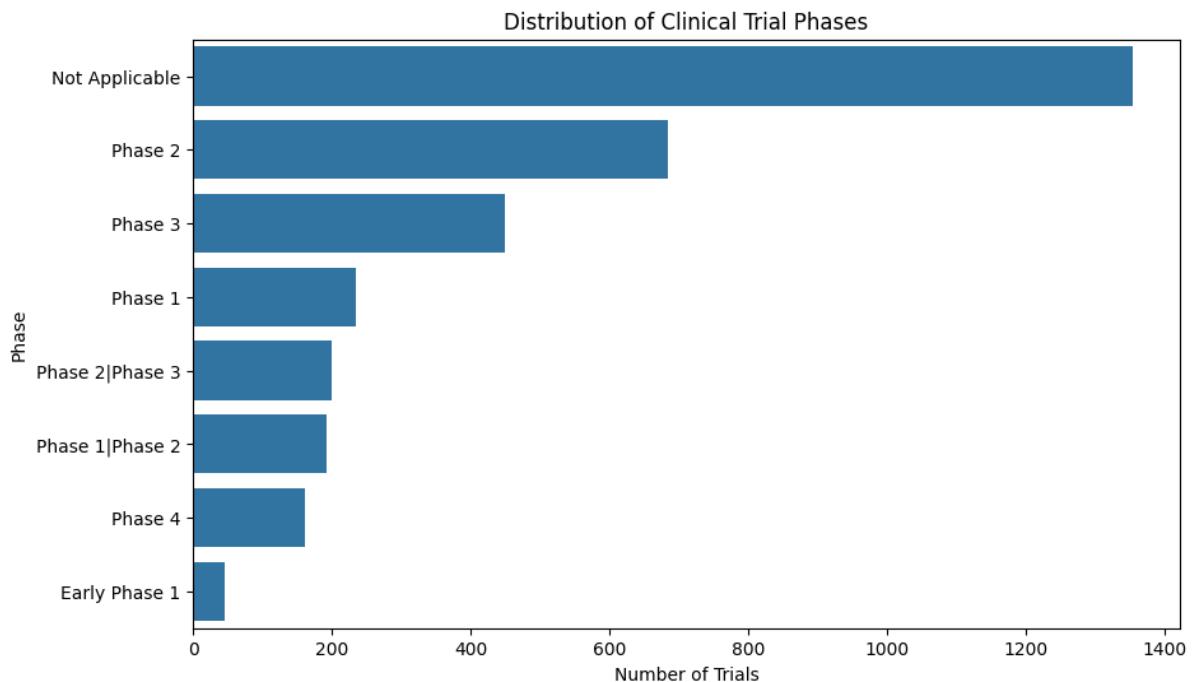
In [21]

```python
plt.figure(figsize=(10,6))
sns.barplot(
    x=phase_counts.values,
    y=phase_counts.index
)
plt.title('Distribution of Clinical Trial Phases')
plt.xlabel('Number of Trials')
plt.ylabel('Phase')
plt.show()
```

Text Output

```
<Figure size 1000x600 with 1 Axes>
```

Image (PNG)

In [22]

```
GENDER DISTRIBUTION

gender_counts = df['Gender'].value_counts()
gender_counts
```

HTML Output

		count
Gender		
All	5567	
Female	162	
Male	44	

dtype: int64

Text Output

```
Gender
All      5567
Female    162
Male       44
Name: count, dtype: int64
```
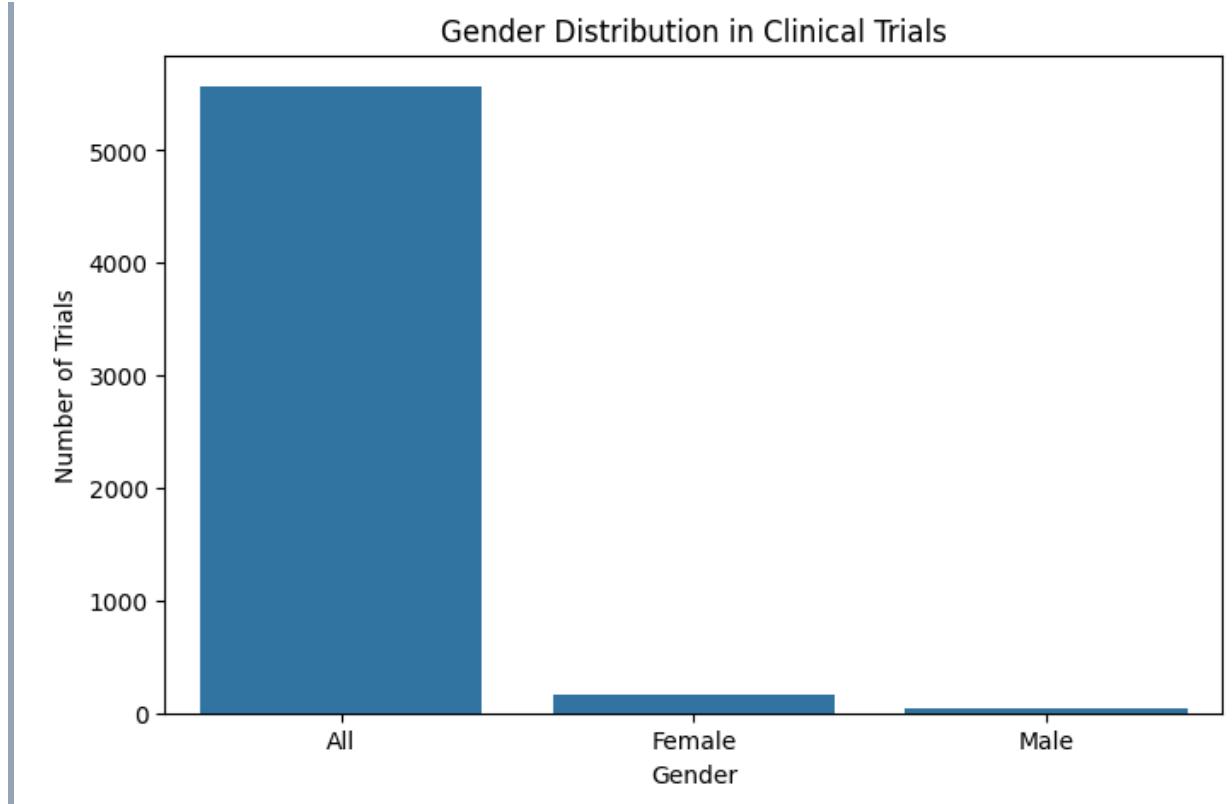
In [23]

```python
plt.figure(figsize=(8,5))
sns.barplot(
    x=gender_counts.index,
    y=gender_counts.values
)
plt.title('Gender Distribution in Clinical Trials')
plt.xlabel('Gender')
plt.ylabel('Number of Trials')
plt.show()
```

Text Output

```
<Figure size 800x500 with 1 Axes>
```

Image (PNG)

Gender Distribution in Clinical Trials

In [24]

```python
# Convert Start Date to datetime
df['Start Date'] = pd.to_datetime(df['Start Date'],
errors='coerce')

# Group by month
trials_over_time = df['Start
Date'].dt.to_period('M').value_counts().sort_index()
trials_over_time
```

HTML Output

		count	
	Start Date		
1998-01		1	
2010-03		1	

		count
Start Date		
2011-02	1	
2011-03	1	
2012-01	1	
...	...	
2021-08	2	
2021-09	8	
2021-10	3	
2021-11	2	
2021-12	1	

78 rows × 1 columns

dtype: int64

Text Output

```
Start Date
1998-01    1
2010-03    1
2011-02    1
2011-03    1
2012-01    1
        ..
2021-08    2
2021-09    8
2021-10    3
2021-11    2
2021-12    1
Freq: M, Name: count, Length: 78, dtype: int64
```
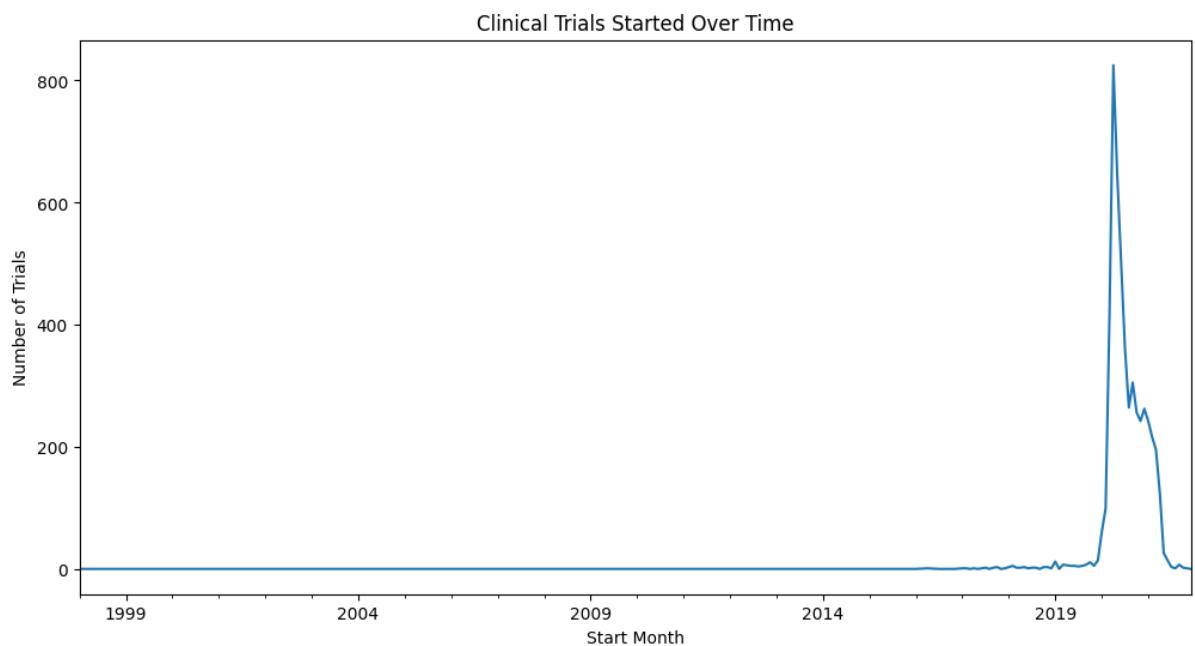
In [25]

```
plt.figure(figsize=(12,6))
trials_over_time.plot(kind='line')
plt.title('Clinical Trials Started Over Time')
plt.xlabel('Start Month')
plt.ylabel('Number of Trials')
plt.show()
```

Text Output

```
<Figure size 1200x600 with 1 Axes>
```

Image (PNG)

Project Summary & Conclusion

1. What I Did

I analyzed a dataset of over 5,000 global **COVID-19 Clinical Trials**. My goal wasn't just to look at numbers, but to understand the story behind how the world's scientists responded to the pandemic. I used **Python** (specifically Pandas for data handling and Matplotlib/Seaborn for charts) to clean the messy raw data and turn it into clear, visual insights.

2. What the Data Showed (My Findings)

- When I visualized the data, a few really interesting patterns stood out:
- **The "April Rush":** I noticed a huge spike in new trials starting in **April and May 2020**. This shows exactly when the global scientific community kicked into high gear to find a cure.
- **The US Led the Way:** My geographical analysis showed that the **United States** conducted the most trials, followed by major countries in Europe and Asia.
- **Active Research:** I found that most of the trials in the dataset are either **"Recruiting"** or **"Completed"**, which tells me that despite the pandemic slowing down, the research data is still active and valuable.
- **Inclusivity:** I was happy to see that the Gender distribution was very balanced. Most studies didn't restrict participants by gender, which is crucial for getting accurate medical results for everyone.

Conclusion

- This project was a great way for me to apply everything I've learned about Data Science in a real-world context. I successfully cleaned raw data, found meaningful trends, and even built a predictive model. It showed me how powerful data can be in tracking and understanding a global health crisis.