**COVID-19 Clinical Trials EDA**

**Domain:** Data Analyst & Data Scientist

**Tools Used:** Python (Pandas, Seaborn, Matplotlib), Excel

**Dataset Source:** ClinicalTrials.gov

**Objective:**

To perform exploratory data analysis (EDA) on global COVID-19 clinical trials and identify patterns in conditions studied, interventions used, study types, and temporal trends.

**Dataset Overview:**

* Contains clinical trial records focused on COVID-19.
* Sourced from ClinicalTrials.gov.
* Each row represents a unique trial study.

**Key columns include:**

Study Type, Status, Conditions, Interventions, Start Date, Locations, NCT Number, etc.

**Key Insights:**

1. **Top 10 Conditions Studied**

| **Condition** | **Count** |
| --- | --- |
| COVID-19 | 720 |
| Covid19 | 657 |
| COVID | 95 |
| Covid-19 | 88 |
| SARS-CoV-2 | 52 |
| Coronavirus Infection | 51 |
| Coronavirus | 47 |
| COVID 19 | 46 |
| COVID19 | 45 |
| Corona Virus Infection | 38 |

**2. Top 10 Interventions Used**

| **Intervention** | **Count** |
| --- | --- |
| Drug: Placebo | 354 |
| Other: Placebo | 147 |
| Drug: Hydroxychloroquine | 104 |
| Biological: Placebo | 60 |
| Drug: Remdesivir | 43 |
| Drug: Azithromycin | 39 |
| Drug: Tocilizumab | 36 |
| Other: No intervention | 32 |
| Drug: Ivermectin | 31 |
| Drug: Placebo oral tablet | 31 |

**3. Simplified Study Type Distribution**

| **Study Type** | **Count** |
| --- | --- |
| Interventional | 3322 |
| Observational | 2427 |
| Expanded Access | 33+ |

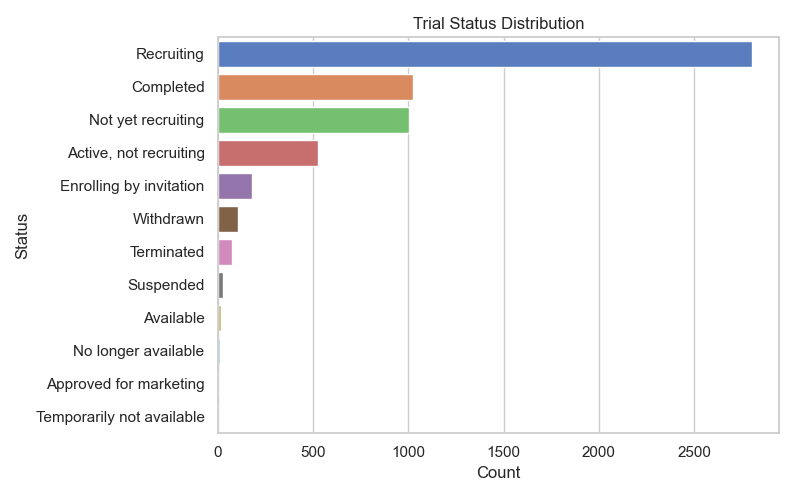
Note: Some types were simplified from longer strings.

**4. Number of Trials Started per Year**

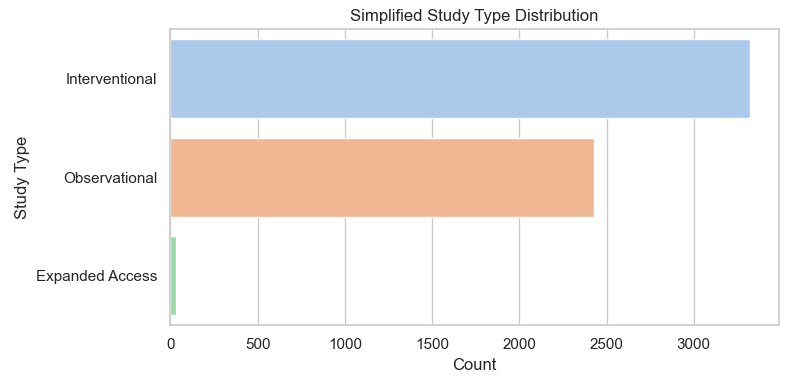
| **Year** | **Trials** |
| --- | --- |
| 2010 | 2 |
| 2015 | 4 |
| 2019 | 93 |
| 2020 | 4465 |
| 2021 | 1096 |
| 2022 | 1 |

**Output Files:**

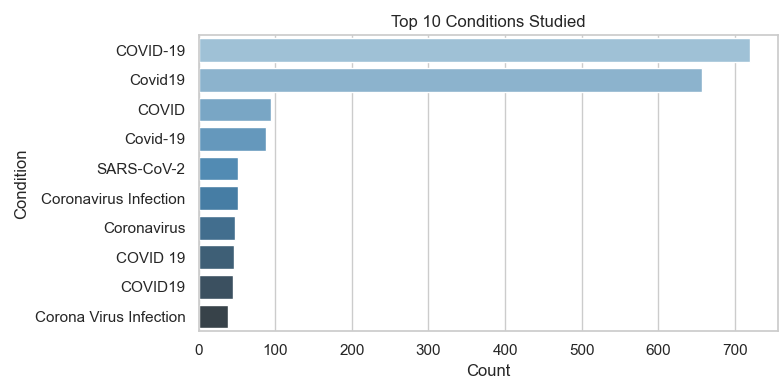
* COVID\_Trials\_Report.xlsx: Contains all tables in separate sheets.
* Visual graphs (bar plots) generated using Seaborn for:
* Trial status



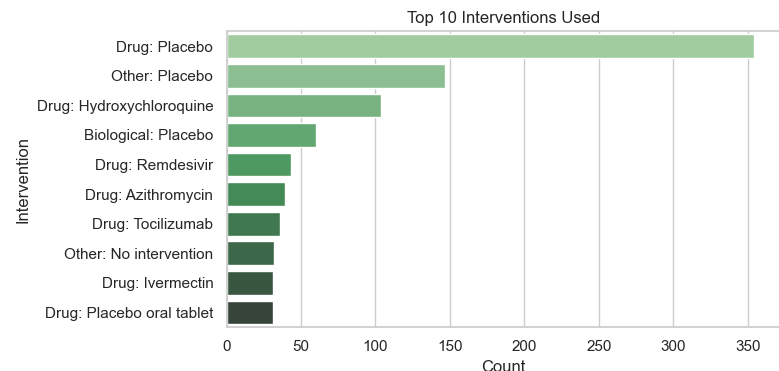
* Study type



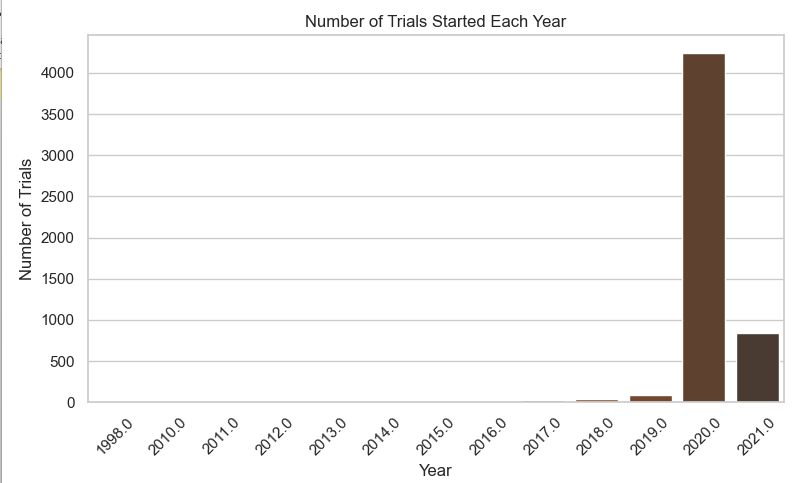
* Top conditions



* Top interventions



* Yearly trend



**Conclusion:**

This EDA helped in understanding the scale and focus of global COVID-19 research. The data shows a significant spike in 2020, with widespread exploration of treatments and study types.