Tab 1

COVID-19 Clinical Trials EDA

Data Cleaning & Preparation

- Remove duplicates, blanks, and irrelevant columns

- Fix inconsistent formatting (dates, case, spacing, etc.)

- Standardize categorical values

Exploratory Data Analysis (EDA)

- Summaries (number of studies, countries, sponsors, etc.)

- Distribution of study phases and statuses

- Timeline analysis of trials over months or years

Visualization Dashboard (Excel Charts)

- Bar/Column charts for study status

- Pie charts for study phases

- Trend chart for number of studies over time

- Country-wise heatmap

Pivot Tables & Insights

- Top sponsors or organizations

- Most common intervention types

- Studies by condition or phase

Tab 2

COVID-19 Clinical Trials Analysis Dashboard

Data Source: ClinicalTrials,gov

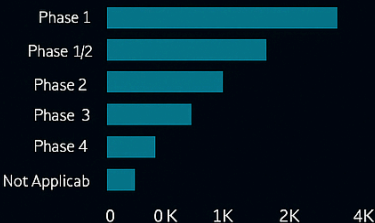| **5,783**	**118**	**2,000**	**2023**
Total Studies	Total Countries	Total Sponsors	Latest Start

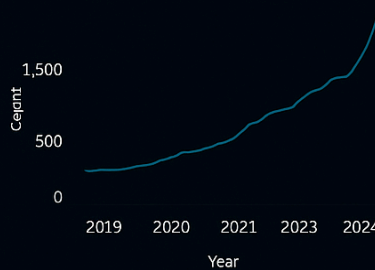## Studies by Phase

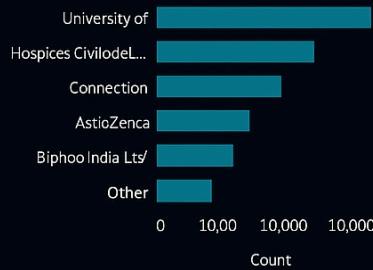## Studies by Overall Status

Geographic Distribution of

Studies by Year

Top 20 Sponsors

Top 20 Interventions

Tab 3

MY SQLQuries

Database creation : CREATE DATABASE covid_trials;
USE covid_trials;

Cleaned data table : CREATE TABLE covid_trials_clean AS
SELECT
  TRIM(`NCT Number`) AS nct_number,
  TRIM(Title) AS title,
  CASE
    WHEN LOWER(Status) LIKE '%recruit%' THEN 'Recruiting'
    WHEN LOWER(Status) LIKE '%complete%' THEN 'Completed'
    ELSE Status
  END AS status,
  STR_TO_DATE(`Start Date`, '%Y-%m-%d') AS start_date,
  STR_TO_DATE(`Completion Date`, '%Y-%m-%d') AS completion_date,
  TRIM(Country) AS country,
  TRIM(Phase) AS phase,
  TRIM(`Study Type`) AS study_type
FROM covid_trials_raw;

Add trial duration column : ALTER TABLE covid_trials_clean ADD COLUMN
duration_days INT;
UPDATE covid_trials_clean
SET duration_days = DATEDIFF(completion_date, start_date)
WHERE start_date IS NOT NULL AND completion_date IS NOT NULL;

Tab 4

🧬 COVID-19 Clinical Trials Exploratory Data Analysis

📊 Project Overview

This project provides an in-depth **Exploratory Data Analysis (EDA)** of global COVID-19 clinical trials.
Using data cleaning, transformation, and visualization techniques in **Python**, the analysis explores trends in trial phases, statuses, conditions, interventions, enrollments, and timelines.

🧠 Objectives

- Understand the **distribution of clinical trials** by phase, status, and country

- Identify the **most common conditions** and **interventions** studied

- Examine **enrollment patterns** and **trial durations**

- Track the **temporal trend** of trial initiations

- Provide a **cleaned and standardized dataset** for downstream analysis or dashboarding

🧹 Data Cleaning & Preparation

1. **Removed duplicates** based on trial ID and title

2. **Standardized phase and status** fields

3. **Normalized multi-value columns** (split countries, interventions, conditions)

4. Converted **date fields** and calculated **trial duration (days)**

5. Handled **missing and inconsistent enrollments**

6. Created a **refined dataset** ready for analysis

Executive Summary

This project presents an exploratory data analysis (EDA) of global COVID-19 clinical trial data. The goal is to understand research trends, intervention types, study statuses, and global participation during the pandemic. Using Python (Pandas, Matplotlib, Seaborn), the dataset was cleaned and analyzed to derive insights into how the world responded to the COVID-19 health crisis through clinical studies.
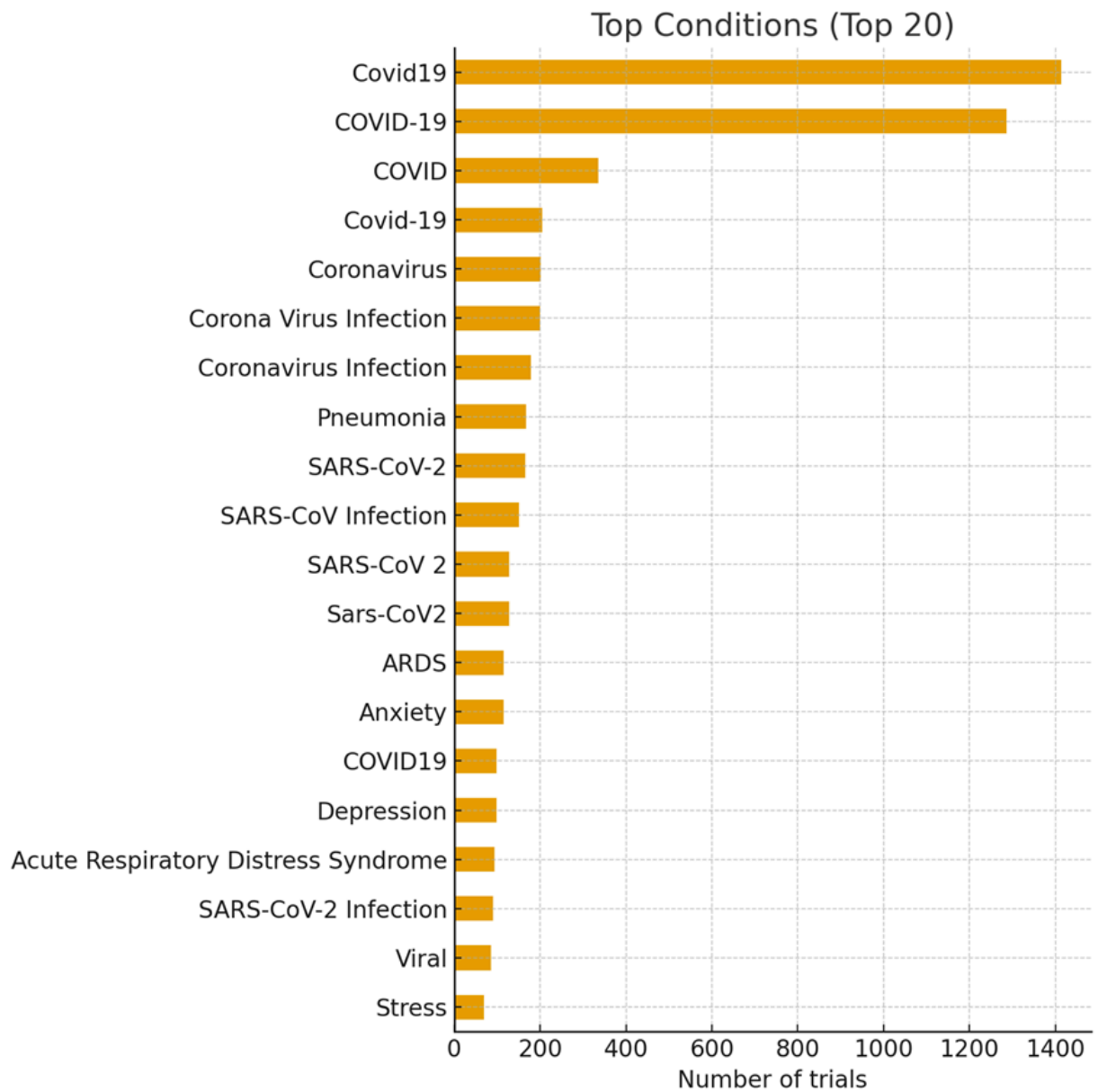
Project Overview & Objectives

The COVID-19 Clinical Trials EDA project investigates patterns in worldwide clinical studies related to COVID-19, focusing on their timelines, sponsors, interventions, and completion statuses.

Key objectives include:

- · • Analyzing trends in trial registrations and completions over time.

- · • Identifying major countries and organizations conducting trials.

- · • Examining the distribution of study statuses (Completed, Recruiting, etc.).

- · • Exploring types of interventions (Drug, Vaccine, Behavioral, etc.).

- · • Summarizing major insights and patterns from the dataset.

Trials by Status

Top Conditions (Top 20)

📜 **Author**

Project by: *Pranjal Waghmare*
 Date: November 2025