COVID-19 Clinical Trials: Exploratory Data Analysis (EDA)

---

1. Introduction

Objective:

- Analyze COVID-19 clinical trial data to uncover patterns and trends.
- Identify insights related to trial phases, sponsors, and study types.

Dataset Overview:

- Data sourced from clinical trial repositories.
- Includes trial IDs, study types, phases, locations, and outcomes.

---

2. Data Cleaning & Preprocessing

Steps Taken:

- Removed duplicate entries.
- Handled missing values (imputation or removal as necessary).
- Standardized categorical variables (e.g., trial phases, study types).
- Converted date formats for consistency.
- Filtered out irrelevant or incomplete trials.

---

3. Exploratory Data Analysis (EDA)

Key Explorations:

- Distribution of trials across different phases.
- Geographic distribution of trials.
- Frequency of study types (interventional, observational, etc.).
- Top sponsors and organizations funding the trials.
- Duration analysis of trials (start to completion).

---

4. Key Insights & Visualizations

Visuals & Findings:

- **Bar Chart:** Number of trials per phase (I, II, III, IV).
- **Heatmap:** Correlation between trial phases and study completion rates.
- **Pie Chart:** Distribution of study types.
- **Histogram:** Duration of trials in days/months.
- **World Map:** Geographic distribution of trials.
- **Box Plot:** Analysis of trial durations and outliers.

---

5. Conclusion & Next Steps

Key Takeaways:

- Majority of trials are in early phases (I & II).
- Interventional studies dominate the dataset.
- North America and Europe have the highest concentration of trials.
- Trial durations vary significantly based on study type.

Future Recommendations:

- Further analysis of patient demographics.
- Sentiment analysis on trial outcomes.
- Predictive modeling for trial success rates.
- Integration with external datasets (e.g., vaccine efficacy reports).

Thank You! 🎯