Project Proposal: Network Analysis of Missing Children Data

Project Overview:
In this project, I aim to conduct a network analysis of missing children data obtained from the National Center for Missing and Exploited Children (NCMEC) dataset (available at: https://data.world/jamesgray/missing-children-in-the-us).
I would like structure the network to explore relationships between missing children based on shared geographic locations, demographic characteristics, and case types, ultimately seeking to uncover patterns and trends in missing children cases.

Project Objectives:

1. **Data collection and preparation:**
   - Collect and preprocess the NCMEC dataset.
   - Extract relevant attributes, including child details, location information, demographics, case types, and timeline.

3. **Possible research questions:**
Are there communities of missing children with shared attributes? Can we identify some patterns? For instance:

   - Are there communities of missing children based on shared demographic attributes, location, or case type?
   - Who are the central missing children in the network, and what do their connections reveal?
   - Can we identify periods or locations with a high concentration of missing children, maybe with the same demographic attributes or case types?

2. **Network Structure:**
   Nodes:
   Missing children: each individual missing child from the dataset will be a node the network.
   Geographic entities: nodes for cities and states can also be included.

   Edges:

   1. Location edges: connect missing children nodes to city or state nodes based on the city or state where they went missing. For example, if multiple missing children are from the same city or state, there will be edges connecting them to the corresponding city or state nodes. These edges represent a shared location.

   2. Demographic edges: create edges between missing children nodes based on shared demographic characteristics like:
      - Shared race/ethnicity.
      - Shared sex (male or female).
      - Shared age group (categorize children into age groups, e.g., <5, 5-10, 10-15, etc.).

   3. Temporal edges: edges can connect missing children nodes if they went missing on the same date or within a specified time period (e.g., same month or year).

4. Case type edges: edges can also connect missing children nodes based on the same case type (e.g., abduction, runaway). This can help identify patterns related to the type of disappearance.