Project Process: Contact Tracing System Using Machine Learning

1. **Data Loading**:
    - Loaded the **contact tracing data** from a JSON file into a pandas DataFrame for further analysis.
2. **Initial Exploration**:
    - Checked for **null values** and **duplicate entries** in the dataset to ensure data quality.
    - Examined the **data types** and conducted a **statistical summary** using df.info() and df.describe() to better understand the structure of the data.
3. **Exploratory Data Analysis (EDA)**:
    - **Visualized Spatial Data**: Plotted the **latitude** and **longitude** of the individuals on a map of Bangalore to get an overview of the distribution of data points.
    - Used **scatter plots** and **heatmaps** to examine patterns in the geographical distribution and detect areas of high activity or contact.
4. **Data Preprocessing**:
    - Checked and handled **duplicate values** and **null values** where necessary.
    - **Feature Engineering**:
        - Extracted **hour** from the timestamp to analyze the data on an hourly basis.
    - Checked for **data imbalance** or skewness in any features that might affect modeling, and applied necessary transformations or preprocessing steps.
5. **Visualization and Mapping**:
    - Used **Folium** and **Branca** to visualize the geographical locations and contact points on a map, adjusting the size and positioning of markers and clusters.
    - Added interactive features like **HeatMapWithTime** to visualize spatio-temporal data, showing how the intensity of contacts changes over time.
    - **Zoomed into specific areas** of the map to analyze contacts within particular geographic regions.

6. **Modeling and Analysis**:
    - Applied clustering algorithms such as **DBSCAN** to group individuals who were in close proximity during specific time intervals.
    - Analyzed movement patterns, identifying potential hotspots of contact, and used these insights to recommend actions for controlling the spread of disease.
    - Checked for **spatial and temporal patterns** of contact, identifying moments of potential risk.
7. **Evaluation**:
    - Evaluated the performance of the contact tracing system by comparing the detected clusters and movements with ground truth (if available) or known areas of high risk.
    - Analyzed the effectiveness of different clustering algorithms and geospatial techniques used to track contacts.