**Problem Statement: Exploratory Data Analysis of COVID-19 Dataset**

**Objective:** Dive into the complexities of COVID-19 data to perform a comprehensive Exploratory Data Analysis (EDA). This dataset provides a rich collection of patient data, including demographic information, pre-existing conditions, COVID-19 test results, and treatment details.

**Dataset Description:**

The dataset includes the following columns:

* **sex:** 1 for female and 2 for male.
* **age:** Age of the patient.
* **classification:** COVID-19 test findings (1-3: positive with different degrees, 4 or higher: negative or inconclusive).
* **patient type:** 1 for returned home, 2 for hospitalization.
* **pneumonia:** Presence of pneumonia (Boolean).
* **pregnancy:** Pregnancy status (Boolean).
* **diabetes:** Presence of diabetes (Boolean).
* **copd:** Presence of Chronic obstructive pulmonary disease (Boolean).
* **asthma:** Presence of asthma (Boolean).
* **inmsupr:** Immunocompromised status (Boolean).
* **hypertension:** Presence of hypertension (Boolean).
* **cardiovascular:** Presence of cardiovascular disease (Boolean).
* **renal chronic:** Presence of chronic renal disease (Boolean).
* **other disease:** Presence of other diseases (Boolean).
* **obesity:** Obesity status (Boolean).
* **tobacco:** Tobacco use (Boolean).
* **usmr:** Medical unit level (Numeric).
* **medical unit:** Type of National Health System institution (Categorical).
* **intubed:** Ventilation requirement (Boolean).
* **icu:** ICU admission (Boolean).
* **date died:** Date of death or '9999-99-99' if the patient survived.

**Tasks:**

1. **Data Cleaning and Preprocessing:**
   1. Handle missing values, ensuring proper treatment of Boolean and categorical variables.
   2. Convert 'date died' to a proper datetime format, creating a binary 'death' variable indicating mortality.
2. **Demographic Analysis:**
   1. Analyse the age and sex distribution among patients, considering different COVID-19 severity levels.
   2. Investigate how demographic factors relate to mortality and hospitalization rates.
3. **Pre-existing Conditions and COVID-19:**
   1. Explore how various pre-existing conditions relate to COVID-19 classification, severity, and outcomes.
   2. Investigate patterns and trends regarding patients with multiple pre-existing conditions.
4. **Treatment and Outcomes:**
   1. Analyse the distribution of patient types, ventilation requirements, and ICU admissions.
   2. Explore how treatment variables correlate with COVID-19 severity and patient outcomes.
5. **Geographical and Institutional Analysis:**
   1. If possible, utilize 'medical unit' and 'usmr' to investigate geographical and institutional patterns in patient care and outcomes.
6. **Time Analysis:**
   1. If additional time-related data is available or can be inferred, conduct a temporal analysis to identify trends over time.
7. **Risk Factor Identification:**
   1. Utilize the dataset to identify significant risk factors for severe COVID-19 outcomes.
8. **Conclusion and Insights:**
   1. Summarize the key findings from your EDA.
   2. Provide recommendations for healthcare providers, policymakers, or patients based on your analysis.