Hospital Dataset Analysis Goals

1. Patient Demographics & Characteristics

- What is the age distribution of patients? Are there more older or younger patients?

- How does sex vary across admissions? Any skew in male vs. female patients?

- What are the common diagnosis codes in the dataset?

2. Clinical Measurements

- How do heart rate, blood pressure, creatinine, and WBC vary by age, sex, or comorbidity score?

- Are there outliers or unusual values in lab measurements?

- Is there a relationship between comorbidity_score and vital signs or lab results?

3. Hospital Outcomes

- What proportion of patients were admitted to ICU?

- What proportion were readmitted within 30 days?

- What proportion died in-hospital?

- Does length of stay (los) vary by ICU admission, readmission, or in-hospital death?

4. Relationships Between Variables

- Are older patients more likely to be admitted to ICU or have longer stays?

- Do patients with higher comorbidity scores have higher rates of ICU admission, readmission, or death?

- Is there any association between sex and outcomes (ICU, readmission, death)?

- Are lab values (creatinine, WBC) predictive of outcomes?

5. Time-Based Analysis

- How does admission_date influence outcomes? Are there seasonal trends?

- Are patients admitted on certain months or weekdays more likely to have longer stays or adverse outcomes?

6. Derived or Composite Analysis

- Can you cluster patients by vital signs, labs, and comorbidity to identify risk groups?

- Are there patterns in readmission or death based on diagnosis code combinations or comorbidity?

- How does los correlate with other continuous variables like age, creatinine, WBC, or blood pressure?