

COVID-19 Dataset EDA - Summary

1. Objective

- - Conduct EDA to understand COVID-19 patterns.
- - Use statistical and visual techniques to uncover relationships.

2. Dataset Overview

- - Size: 1,048,575 records and 21 columns.
- - Key columns: Demographics (AGE, SEX), Health Conditions (PNEUMONIA, DIABETES, COPD, OBESITY), Outcomes (PATIENT_TYPE, DATE_DIED, ICU).
- - No missing values detected; dataset is clean.

3. Key Insights

a) Patient Type Distribution

- - Majority were outpatients.

b) Gender Distribution

- - Females: ~50.1%, Males: ~49.9% (balanced).

c) Age Distribution

- - Most patients aged between 20-60 years.

d) Health Factors and Mortality

- - Higher death rates in patients with pneumonia and diabetes.
- - Older patients had higher mortality.
- - Comorbidities (e.g., hypertension, cardiovascular disease) increased death risk.

4. Correlation Analysis

- - Strong positive correlation between Total Cases and Deaths.
- - Moderate negative correlation between Recovered and Active Cases.
- - Health conditions correlated with ICU admission and death status.

5. Visualization Highlights

- - Countplot: Outpatient vs Hospitalized.
- - Pie Chart: Gender distribution.

- - Histogram: Age distribution.
- - Heatmap: Variable relationships.
- - Pairplot: Health conditions vs death outcome.

6. Conclusion

- - Age, pneumonia, and diabetes majorly influence mortality.
- - Early treatment and comorbidity management critical.
- - Insights can support better healthcare planning for future pandemics.