***Exploratory Data Analysis (EDA) with Pandas in Covid-19***

**1. Goals of the Project**

* To analyze COVID-19 patient data and extract meaningful insights.
* To identify trends related to symptoms, test results, and severity.
* To assess the impact of comorbidities and vaccination status on patient outcomes.
* To develop visualizations for better data interpretation.
* To utilize feature engineering for improved analytical performance.

**2. Materials and Methods**

* **Dataset Used:** COVID-19 patient dataset containing 19 features.
* **Data Processing Tools:** Python, Pandas, NumPy, Matplotlib, Seaborn, Plotly.
* **Analysis Methods:** Data cleaning, feature extraction, statistical analysis, and visualization.
* **Machine Learning Techniques (if applicable):** Classification models for predicting outcomes.
* **Software:** Jupyter Notebook for coding and data exploration.

**3. General Parts**

* **Data Cleaning:** Handling missing values, removing duplicates, and correcting data inconsistencies.
* **Exploratory Data Analysis (EDA):** Distribution of symptoms, test results, and hospitalization rates.
* **Visualization:** Bar charts, histograms, correlation heatmaps, scatter plots, pie charts, and KDE plots.
* **Statistical Insights:** Mean age of infected patients, percentage of vaccinated vs. unvaccinated cases.

**4. Project Outcomes and Insights**

* **Age Factor:** Older patients had a higher rate of hospitalization and ICU admissions.
* **Symptoms Analysis:** Fever and cough were the most common symptoms.
* **Oxygen Levels:** A drop in oxygen levels correlated with ICU admissions.
* **Impact of Vaccination:** Vaccinated individuals showed lower hospitalization rates.
* **Comorbidities Influence:** Patients with pre-existing conditions had more severe outcomes.
* **Death Rate:** Higher among non-vaccinated patients and those with comorbidities.
* **Geographical Trends:** Location-based variations in hospitalization and ICU admissions.

**5. Feature Engineering**

* **Categorization of Age Groups:** Grouping patients into different age categories.
* **Severity Score:** Created a new feature based on symptoms and oxygen levels.
* **Binary Encoding:** Converted categorical variables (e.g., Vaccine Type) into numerical values.
* **Feature Scaling:** Standardized numerical features for better machine learning model performance.
* **New Variables:** Derived additional features like Age\_Group, Severity\_Score, and Comorbidity\_Factor.

**6. Key Questions and Insights to be Addressed:**

**1.How many patients tested positive for COVID-19?**

positive\_cases = df[df["Test\_Result"] == "Positive"]

print(positive\_cases)

**2.What is the distribution of patients by location?**

location\_counts = df["Location"].value\_counts()

print(location\_counts)

1. **How many patients had oxygen levels below 94?**

low\_oxygen = df[df["Oxygen\_Level"] < 94]

print(low\_oxygen)

1. **How many patients were vaccinated?**

vaccinated\_counts = df["Vaccinated"].value\_counts()

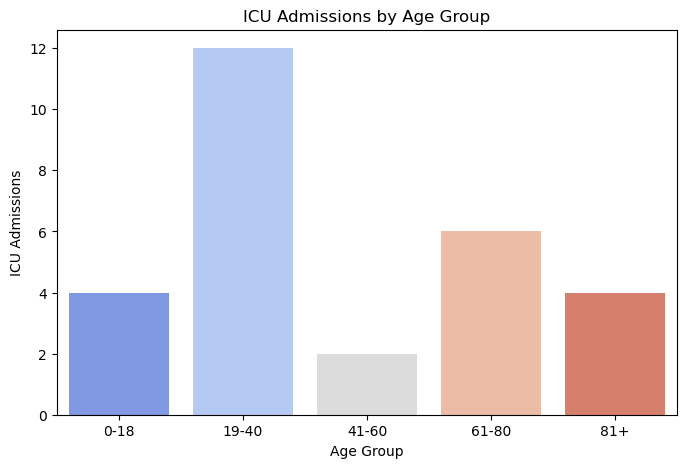
print(vaccinated\_counts)

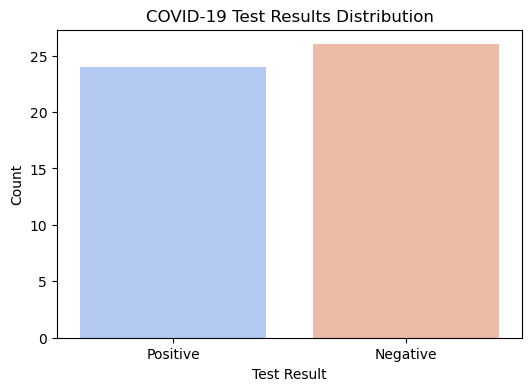
**5.What are the counts of each vaccine type?**

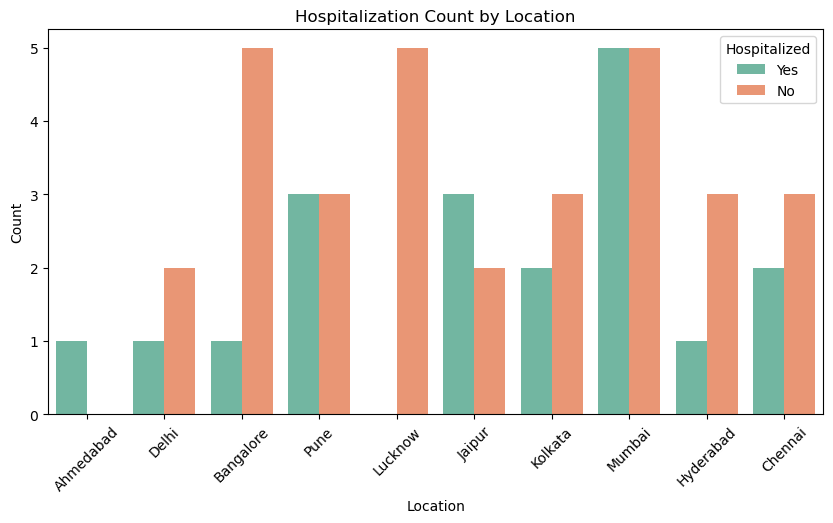
vaccine\_counts = df["Vaccine\_Type"].value\_counts()

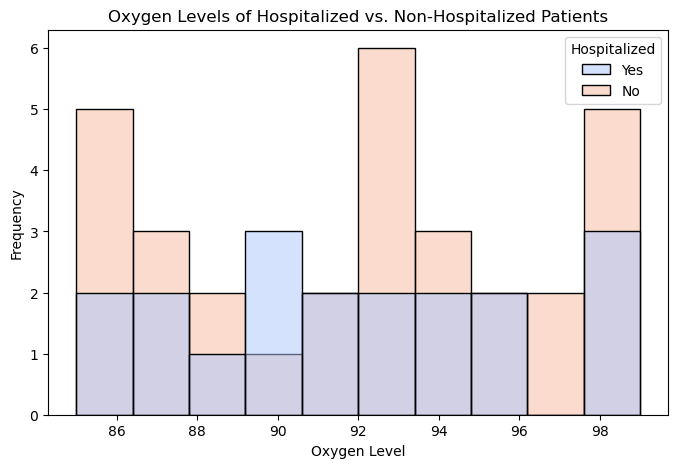
print(vaccine\_counts)

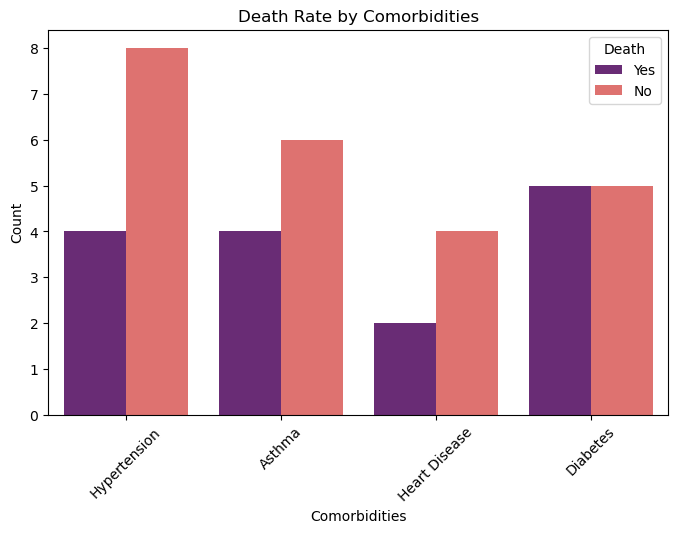
1. **Visualization**

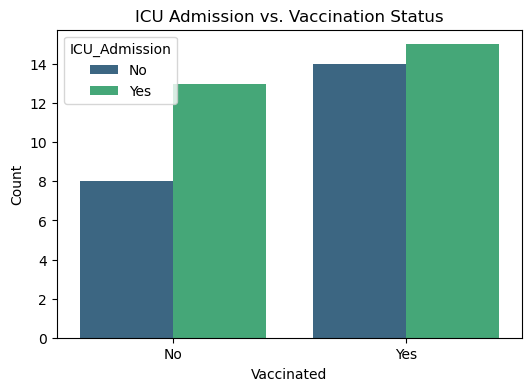


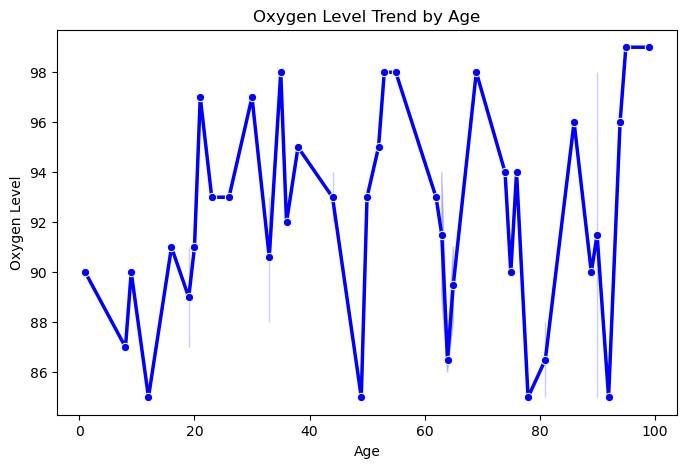
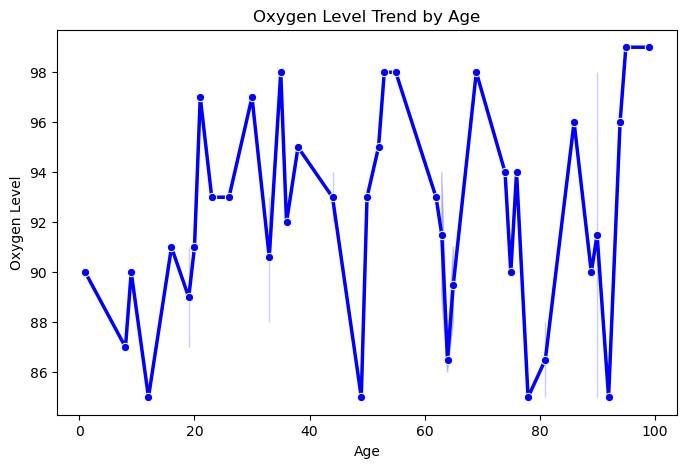


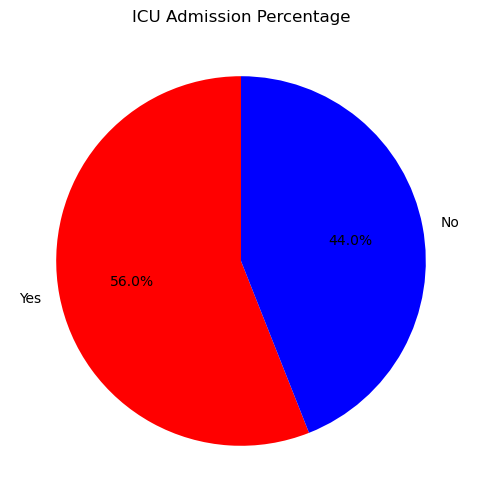


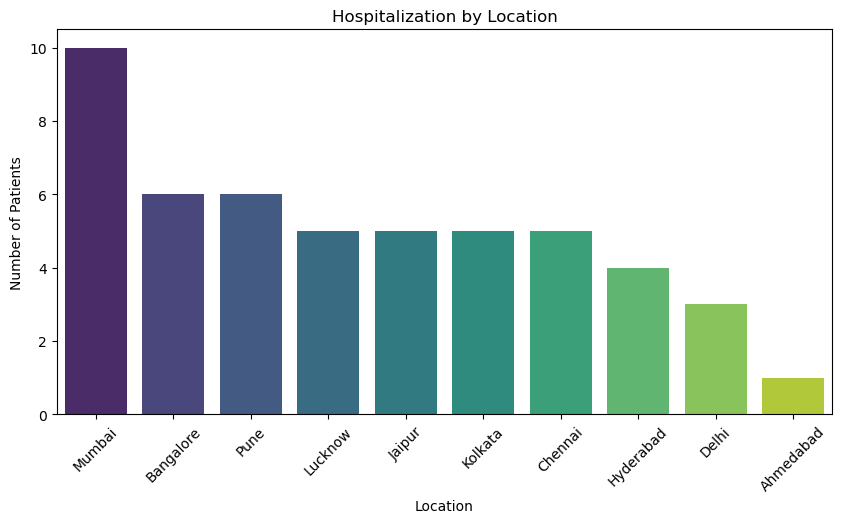












**8. Conclusion**

* The analysis provided key insights into COVID-19 patient trends and severity factors.
* Vaccination significantly reduced the severity of outcomes and mortality rates.
* Comorbidities played a crucial role in worsening patient conditions.
* The study highlights the importance of early detection and monitoring of oxygen levels.
* Various visualizations helped in understanding correlations and trends in the dataset.
* Future work can include predictive modeling to forecast patient outcomes based on symptoms and medical history.