Date

COVID-19 CASES ANALYSIS

EXPLORATORY DATA ANALYSIS

Data Collection: Start by gathering the relevant data. You can use sources like the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), or your country's health department.

Data Overview: Display summary statistics to get an initial sense of the data. This may include mean, median, standard deviation, and quartiles for numeric variables. Check the data types of each variable.

Hypothesis Testing: If you have specific research questions, perform hypothesis testing to validate or refute hypotheses. For example, you might test whether there is a significant difference in vaccination rates between different regions or age groups.

Correlation Analysis: Explore relationships between variables. For example, you can check if there's a correlation between vaccination rates and infection rates or other health outcomes.

Interactive Dashboards: You can build interactive dashboards using tools like Tableau, Power BI, or Python libraries like Dash or Bokeh to provide a dynamic way for users to explore the data.

Communicate Findings: Prepare a report or presentation summarizing your EDA findings. Use visualizations and clear explanations to make your results accessible to a broader audience.

STATISTICAL ANALYSIS

Formulate Hypotheses: Start by formulating clear hypotheses that you want to test. For example, you might want to test whether the effectiveness of one COVID-19 vaccine is significantly different from another or whether vaccination rates vary significantly across age groups.

Data Preparation: Clean and preprocess the data as needed. This includes handling missing values, outliers, and transforming variables if necessary.

Descriptive Statistics: Calculate and report descriptive statistics for relevant variables. Common statistics to include are mean, median, standard deviation, quartiles, and percentages. This provides an initial overview of the data.

Regression Analysis: Perform regression analysis when you want to understand the impact of multiple variables on an outcome. For example, you might use linear regression to predict vaccination rates based on various factors like age, location, and vaccine type.

Statistical Significance: Report the results of statistical tests along with p-values. Determine the statistical significance of your findings. A common significance level is 0.05, but adjust it if needed.

VISUALIZATION

Cases Rates Over Time: Line chart or time series plot showing the cumulative or daily vaccination rates over time. This can help visualize the progress of the vaccination campaign.

Regional cases Rates: Choropleth map showing vaccination rates by region, city, or country. Color coding can highlight areas with high or low vaccination rates.

Cases adverse Events: A bar chart or stacked bar chart showing the frequency and type of adverse events reported for different vaccines.

Cases Dashboard: Interactive dashboards with multiple visualizations, filters, and drill-down capabilities for a comprehensive view of the data.