

Project 4: COVID Vaccines Analysis

Objective

To conduct an in-depth analysis of Covid-19 vaccine data, focusing on vaccine efficacy, distribution, and adverse effects. The goal is to provide insights that aid policymakers and health organizations in optimizing vaccine deployment strategies.

Design thinking phase

Data collection:

The data needed for analysis will be collected from government organisations like WHO,CDN and other institutions plus online portals like github,huggingface.io..the collection of data will be from multiple source for cross verification.

Data will be Collected based on country,vaccine types, efficacy rates, distribution metrics (e.g., doses administered, vaccination rates), and adverse event reports.

Data preprocessing

The collected data should be cleaned and unified into single database table. this step includes data cleaning, transformation and integration. the following things will be made clear:

- 1) Remove duplicate entries to maintain data integrity.
- 2) Check and correct inconsistencies in data, such as variations in date formats or units of measurement.
- 3) Create new variables or features, such as calculating vaccine coverage rates or adverse event rates per 100,000 population.
- 4) Combine data from different sources into a unified dataset if applicable.
- 5) Ensure that the data is ready for analysis, with appropriate columns and data types.

Exploratory Data Analysis (EDA):

- 1)Generating Summary Statistics like calculating basic statistics like mean, median, standard deviation, and percentiles for relevant variables
- 2)Creating histograms, box plots, and density plots to visualize the distribution of key variables.
- 3)Using scatter plots and heatmaps to explore relationships between variables.
- 4)Creating geographic maps to visualize vaccine distribution patterns across regions.
- 5)Based on initial visualizations, formulate hypotheses about trends or patterns in the data.
- 6)Identify potential outliers or anomalies .

Statistical Analysis:

This steps involves:

1) Choosing appropriate statistical tests (e.g., t-tests, chi-squared tests) to answer specific research questions.

2) Then performing hypothesis tests to compare vaccine efficacy between different brands or regions.

3) Utilizing regression models (e.g., linear regression, logistic regression) to assess the impact of factors like population density, age demographics, or vaccine supply on vaccine distribution or adverse effects.

4) Interpret coefficients and p-values to determine significance.

Visualisation

Creating interactive dashboard for visualisation using tools like tableau, power BI or other python libraries.

Variety of visualizations like bar charts, line graphs, scatter plots, and geographic maps will be used for representing summary of collected covid data. colors, labels, and annotations to enhance the clarity and informativeness of your visualizations.

Insights & Recommendations:

Interpret my finding for this project.i will provide evidence based recommendations for policymakers and health organisations.will also give suggestions based on my analysis for increasing efficiency of vaccine deployment.

x-x-x-END OF DOCUMENT-x-x-x