Model Report: Analysing Latitude, Longitude, Deaths, and Case Fatality Ratio

Introduction

The purpose of this report is to explore the dataset containing information on latitude, longitude, deaths, and case fatality ratio. The report describes the analysis performed, the models applied, and the insights derived, with a focus on visual and statistical representation of the data.

Dataset Overview

The dataset comprises the following columns:

- **Latitude**: Geographic coordinate specifying north-south position.
- **Longitude**: Geographic coordinate specifying east-west position.
- **Deaths**: Number of deaths recorded at a location.
- **Case Fatality Ratio (CFR)**: The percentage of deaths relative to confirmed cases.

Data Preprocessing

The dataset was checked and pre-processed to ensure:

1. **Handling Missing Values**: Missing or null values in key columns were handled through imputation or removal of records.
2. **Data Type Verification**: Ensured all numeric columns (e.g., deaths, case fatality ratio) were in appropriate formats for analysis.
3. **Outlier Detection**: Identified and analyzed outliers in deaths and CFR using statistical methods.

Model Overview

The analysis primarily focused on understanding trends and relationships between variables:

1. **Visualization Models**:
    - **Bar Plot**: Highlighted the top 10 locations with the highest deaths.
    - **Histogram**: Displayed the distribution of case fatality ratios.
    - **Scatter Plot**: Explored the relationship between deaths and case fatality ratio.
    - **Line Plot**: Tracked cumulative deaths across latitudes.
2. **Statistical Analysis**:
    - **Correlation Analysis**: Examined the strength of relationships between variables (e.g., deaths and CFR).

- **Regression Modelling** : A linear regression model was applied to predict CFR based on deaths.

Performance and Insights

- **Bar Plot Insights**: The top 10 locations accounted for a significant proportion of deaths, emphasizing geographic hotspots.
- **Histogram Insights**: The case fatality ratio displayed a right-skewed distribution, with most values concentrated below a specific threshold.
- **Scatter Plot Insights**: A weak positive correlation was observed between deaths and case fatality ratio, indicating that higher deaths may not always imply a higher CFR.
- **Line Plot Insights**: Cumulative deaths increased consistently across latitudes, with specific spikes observed in certain regions.

Model Performance

For the regression model:

- **R-squared Value**: Measures the proportion of variance in CFR explained by deaths.
- **Mean Squared Error (MSE)**: Quantifies prediction errors.

Conclusion

The analysis provided key insights into the relationship between geographic locations and fatality data:

- Certain locations exhibited disproportionately high deaths.
- Case fatality ratios vary significantly, with most regions experiencing moderate ratios.
- A deeper analysis incorporating external factors (e.g., population density, healthcare infrastructure) could provide additional insights.