**Data Science Bootcamp: Nuclear Explosions Dataset**

Welcome to the Data Science Bootcamp! This program will teach you how to analyze, clean, and visualize data using a dataset of nuclear explosions conducted globally between 1945 and 1998. You’ll learn to extract insights, identify trends, and create interactive dashboards.

**Key Learning Objectives**

1. **Data Preprocessing**
   * Clean and format raw data.
   * Handle missing or inconsistent values.
2. **Exploratory Data Analysis (EDA)**
   * Discover trends and patterns in the dataset.
   * Perform time-series and geographical analysis.
3. **Data Visualization**
   * Create charts and maps to represent findings.
4. **Interactive Dashboard**
   * Build a user-friendly app with filters and dynamic visualizations.

**Dataset Overview**

|  |
| --- |
| **Column Name Description** |
|  |
| |  |  | | --- | --- | | **WEAPON SOURCE COUNTRY** | The country that conducted the nuclear explosion (e.g., USA, USSR). | |
| |  |  | | --- | --- | | **WEAPON DEPLOYMENT LOCATION** | The specific location or test site of the explosion (e.g., Alamogordo, Bikini). | |
| |  |  | | --- | --- | | **Data Source** | The organization or source that recorded the data (e.g., DOE - Department of Energy). | |
| |  |  | | --- | --- | | **Location.Cordinates.Latitude** | The latitude coordinate of the explosion's location. | |
| |  |  | | --- | --- | | **Location.Cordinates.Longitude** | The longitude coordinate of the explosion's location. | |
| |  |  | | --- | --- | | **Data.Magnitude.Body** | The body wave magnitude of the explosion, measured in seismic terms (if available). | |
| |  |  | | --- | --- | | **Data.Magnitude.Surface** | The surface wave magnitude of the explosion, measured in seismic terms (if available). | |
| |  |  | | --- | --- | | **Location.Cordinates.Depth** | The depth of the explosion: negative values for above-ground, positive for underground. | |
| |  |  | | --- | --- | | **Data.Yeild.Lower** | The estimated lower bound of the explosion's yield (measured in kilotons). | |
| |  |  | | --- | --- | | **Data.Yeild.Upper** | The estimated upper bound of the explosion's yield (measured in kilotons). | |
| |  |  | | --- | --- | | **Data.Purpose** | The purpose of the explosion (e.g., Combat, Weapon Experimentation, Peaceful Nuclear Explosion). | |
| |  |  | | --- | --- | | **Data.Name** | The unique name or code assigned to the explosion (e.g., Trinity, Littleboy). | |
| |  |  | | --- | --- | | **Country** | The country associated with the explosion (e.g., USA, USSR). | |
| |  |  | | --- | --- | | **Average Yield** | The average yield of the explosion (calculated for some records). | |
| |  |  | | --- | --- | | **Date.Month** | The month of the explosion. | |
| |  |  | | --- | --- | | **Date.Year** | The year of the explosion. | |

**How We Will Approach This Dataset**

**Step 1: Data Cleaning**

* Rename columns for better readability (e.g., converting Data.Yeild.Lower to Yield\_Lower).
* Fix inconsistencies in WEAPON DEPLOYMENT LOCATION and Data.Purpose using dictionaries.
* Handle missing values and incorrect data entries.

**Step 2: Exploratory Data Analysis (EDA)**

* **Time Analysis**:
  + How did nuclear testing activity vary by year?
* **Geographical Analysis**:
  + Where were the majority of nuclear explosions conducted?
* **Purpose Analysis**:
  + What were the most common purposes for nuclear tests?
* **Magnitude and Yield**:
  + Which tests had the highest magnitude and yields?
  + Identify trends in explosions categorized as “High,” “Moderate,” or “Low” magnitude.

**Step 3: Interactive Visualization**

* **Bar Chart**:
  + Compare the number of explosions across different locations.
* **Line Plot**:
  + Show trends in explosions over time for different countries.
* **Pie Chart**:
  + Display the distribution of explosion purposes.
* **Map**:
  + Plot all explosion locations on an interactive world map with color-coded markers for above-ground and underground tests.

**Step 4: Building an Interactive Dashboard**

* Use **Streamlit** to build a user-friendly dashboard.
* Enable filters to allow users to:
  + Select specific locations, years, or depth ranges.
  + Adjust yield ranges for analysis.

**Examples from the Dataset**

Below are sample rows to help you understand the dataset:

| **Country** | **Location** | **Latitude** | **Longitude** | **Yield (kt)** | **Purpose** | **Date** |
| --- | --- | --- | --- | --- | --- | --- |
| USA | Alamogordo | 32.54 | -105.57 | 21 | Weapon-Related | July 1945 |
| USA | Hiroshima | 34.23 | 132.27 | 15 | Combat Detonation | August 1945 |
| USA | Bikini | 11.35 | 165.20 | 21 | Weapon Experimentation | June 1946 |
| USSR | Semipalatinsk | 48.00 | 76.00 | 22 | Weapon-Related | August 1949 |

**Interactive Features**

1. **Filters**:
   * Choose specific deployment locations or countries.
   * Filter by year, yield range, or depth (above ground vs. underground).
2. **Visualizations**:
   * Bar charts to compare deployment locations.
   * Line graphs to analyze yearly trends.
   * Pie charts to show distributions of purposes.
   * Interactive maps to view explosion sites.
3. **Export Results**:
   * Download filtered data as a CSV file.

**Why This Dataset?**

This dataset is rich with features that allow us to practice real-world data science skills. From cleaning messy data to uncovering historical trends, it provides an opportunity to apply core concepts and tools in data science, including:

* Data wrangling.
* Feature engineering.
* Data visualization and storytelling.