College of Computer and Information Sciences
Department of Information Technology

IT 362 Course Project
Semester-2, 1446H

Data Science
< PREDINA >

Prepared by

Student name	Student ID
Nouf AlMansour	444200525
Sara Aloqiel	444203016
Shahad Almutairi	444200935

Supervised by: Dr. Mashael Aldayel

1. Introduction

Natural disasters strike without warning, leaving behind trails of devastation. While natural disasters may not always be preventable, early warning indicators can significantly mitigate losses and enhance protection measures. This study aims to deepen the understanding of natural disasters, focusing on proactive strategies to help minimize their impact and improve preparedness.

This study aims to provide insight to the question "What are the regional patterns and financial impacts, and evolving trends of natural disasters over the past century, and what do these suggest for future risk and preparedness?"

2. Data sources

- EM-DAT: The International Disaster Database

The EM-DAT Public Table is a global disaster database maintained by CRED, tracking natural and technological disasters. It includes data on fatalities, affected populations, and economic damages, and is used for research and disaster management.

- Kaggle Dataset : ALL NATURAL DISASTERS 1900-2021 / EOSDIS

This dataset, hosted on Kaggle, provides a record of natural disasters worldwide from 1900 to 2021, sourced from NASA's Earth Observing System Data and Information System (EOSDIS). It includes details such as disaster type, location, dates, and impacts (e.g., fatalities, affected populations, and economic damages), making it useful for analyzing historical disaster trends and impacts.

Applications of Kaggle Dataset:

The **"All Natural Disasters 1900-2021"** dataset provides a comprehensive historical record of natural disasters worldwide, making it a valuable resource for various research areas. Through **Exploratory Data Analysis (EDA)**, we aim to uncover significant trends, patterns, and correlations that can offer meaningful insights. These insights contribute to multiple domains, including:

Climate Change Research

Natural disaster data is essential for understanding the impact of climate change on disaster frequency and severity. By analyzing historical trends, researchers can assess how global warming influences the occurrence of extreme events.(NASA - Climate Change Effects)

Potential Outcomes:

- Identifying correlations between rising global temperatures and the increasing frequency and intensity of disasters.
- Supporting researchers in forecasting future disaster trends and advocating for climate action policies.
- Enabling governments and environmental organizations to develop **mitigation strategies** to minimize disaster-related risks.

Economic & Policy Impact Studies

Understanding the economic and social consequences of natural disasters is crucial for effective policy-making. By leveraging this dataset, policymakers and economists can assess the financial impact of disasters and develop strategies to enhance disaster preparedness and recovery.(World Bank - Disaster Risk Management)

Potential Outcomes:

- Helping policymakers assess financial losses from disasters and optimize resource allocation.
- Assisting governments in the development of **disaster relief funds** and **insurance policies** to better support affected populations.
- Providing economists with data-driven insights to formulate **long-term economic strategies** aimed at minimizing losses from future disasters.

This dataset not only enhances academic research but also plays a crucial role in shaping **disaster risk reduction strategies, climate policies, and economic resilience efforts**.

Data Collection and Integration

- **Observations:** The combined dataset will encompass disaster records from 1900 to 2021, potentially exceeding 100,000 observations, given the extensive historical coverage.
- **Features and Data Types:**
    - **Categorical:** Disaster type, country, affected region

- **Numerical:** Fatalities, affected population, economic damages, magnitude (for earthquakes, storms, etc.)
- **Temporal:** Date of occurrence, duration
- **Geospatial:** Latitude, longitude of the disaster even

Evaluation of Potential Biases

1-Representation:

- Some disasters may be underreported, particularly in low-income regions where reporting infrastructure is limited.
- Certain regions and disaster types may be overrepresented due to better monitoring systems in developed countries.

2-Measurment Bias:

- The datasets aggregate data from multiple sources, reducing the likelihood of systematic measurement bias.
- Variations in economic damage estimation methodologies across countries may introduce inconsistencies.
- Differences in data collection methods over time and across sources may affect comparability but are mitigated through standardized reporting frameworks.

3-Historical:

- Older records may be incomplete or inconsistent due to technological and administrative limitations in past decades.
- The dataset may reflect historical inequalities, as wealthier nations tend to have more comprehensive records and faster response mechanisms, potentially skewing impact analysis.

3. Objectives

This study aims to uncover the following:

- What factors influence the intensity and impact of natural disasters in specific regions?
- Which regions are most commonly affected by specific types of natural disasters?
- What are the estimated losses associated with natural disasters based on current data and forecasts?
- Which regions have a greater chance in expecting earthquakes?

4. Method

To uncover patterns and trends in natural disasters, we need to apply a combination of data science techniques.

Analyzing correlation and feature importance can uncover what factors influence the intensity and impact of natural disasters. We will apply geospatial analysis to portray Identify regions most affected by specific natural disasters. We will apply regression analysis to estimate losses based on certain features such as regions, disaster and other factors. We will apply EDA to identify which regions have a greater chance in expecting earthquakes, as studies suggest earthquake history is a forecasting tool

These methods will help magnify insights from the data.

5.Challenges in Data Collection and Recommendations

Challenges in Data Collection:

- Missing or Incomplete Data: Some regions, especially low-income countries, may have gaps in historical records due to weak reporting systems.
- Standardization Issues: Variability in measurement units (e.g., currency for economic damages, population reporting methods) may introduce inconsistencies.
- Access Restrictions: Some high-quality disaster datasets may require paid access or institutional affiliations, limiting data availability.

Recommendations for Future Data Collection and Analysis:

- Adopt Standardized Reporting Protocols: Encourage the use of internationally recognized frameworks such as EM-DAT classification criteria to ensure consistency across sources.
- Integrate Remote Sensing Data: Utilize satellite imagery and other geospatial data sources to supplement on-the-ground disaster reports and improve coverage in underreported regions.
- Promote Open Data Initiatives: Advocate for greater transparency and accessibility of disaster-related datasets to improve data availability and completeness.

-