

Project Title: TEMPORAL ANALYSIS OF CLIMATIC INFLUENCES ON FOREST FIRE PATTERNS IN PENINSULAR MALAYSIA USING STATISTICAL METHOD

Prepared by: GRACE LING KIAN HWAI, MCS231026, GraceLKH

Chapter 4 EDA/Initial Results

4.1 Introduction

This chapter delves into the Exploratory Data Analysis (EDA) performed to comprehend the main data and obtain initial insights related to the research. Essential process in any data science project as it aids in identifying patterns, detecting abnormalities, forming hypotheses, and validating assumptions by summary statistics and visual representations. The chapter begins by describing the primary data sources, then details the visualizations and descriptive statistics used to explore the data. Initial insights gained from the EDA and feature engineering processes are also discussed. Finally, the chapter concludes with a summary that reiterates the importance of these initial findings in guiding the subsequent phases of the research.

4.2 Primary Data

The primary data for this study consists of two main datasets:

1. Forest Fire Dataset: Obtained from multiple-source remote sensing data through Google Earth Engine, this dataset provides detailed information on forest fire occurrences in Peninsular Malaysia, including the location, time, and intensity of fires.
2. Climate Change Dataset: Obtained from the IMF Climate Change Dashboard, this dataset includes various meteorological factors like temperature, rainfall, moisture, and wind velocity. These variables are crucial for analysing the climatic influences on forest fire patterns.

4.3 Exploratory Data Analysis (EDA)

EDA is an essential stage in comprehending the fundamental framework of the data, identifying patterns, detecting anomalies, and forming hypotheses. The following subsections describe the visualizations, descriptive statistics, initial insights, and feature engineering performed during the EDA phase.

4.3.1 Visualizations and Descriptive Statistics

To explore the data, several visualizations and descriptive statistics were generated. Key visualizations include:

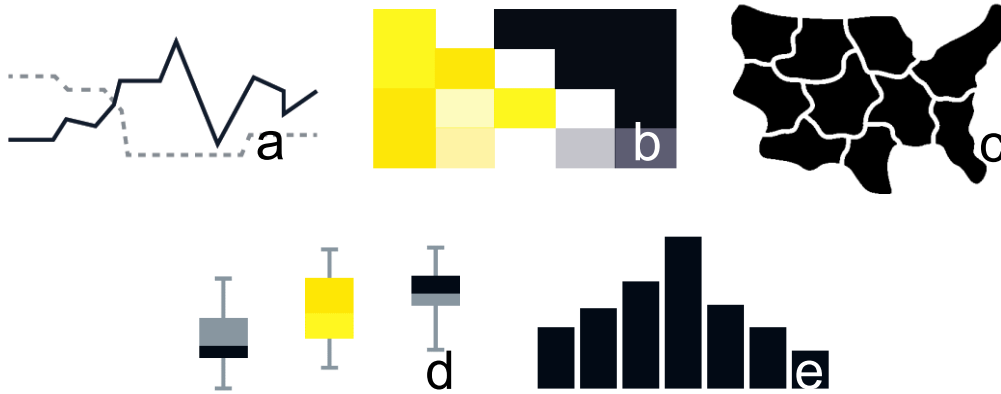


Figure 4.3.1: Key visualizations: (a) Line Graph; (b) Heatmap; (c) Map; (d) Boxplot; (e) Histogram.

1. Time Series Plots: Displaying the temporal patterns of forest fire occurrences and climatic variables.
2. Heatmaps: Showing the correlation between different climatic variables and fire incidents.
3. Geographical Maps: Highlighting the spatial distribution of forest fires across Peninsular Malaysia.
4. Box Plots: Illustrating the distribution and range of climatic variables during fire and non-fire periods.
5. Histograms: Depicting the frequency distribution of fire occurrences and climatic variables.

Descriptive statistics, which include mean, median, standard deviation, and quartiles, were used to summarize the central tendency, spread, and distribution of the data.

4.4 Initial Insights Gained from EDA

From the EDA, several initial insights were derived:

1. Seasonal Trends: A noticeable increase in fire occurrences during the dry season, with a significant correlation between high temperatures and low humidity levels.

2. **Geographical Hotspots:** Certain regions in Peninsular Malaysia are more prone to forest fires, particularly areas with dense vegetation and lower rainfall.
3. **Climatic Influences:** Strong correlations between specific climatic variables (e.g., temperature, wind speed) and the frequency of forest fires, suggesting that these factors play a critical role in fire occurrences.
4. **Anomalies and Outliers:** Identification of outliers in both the fire and climate datasets, which could indicate extreme events or data quality issues.

4.5 Feature Engineering

Feature engineering is the process of generating additional variables using the available data in order to enhance the predictive capability of the models. Key feature engineering steps include:

1. **Fire Weather Indices:** Calculation of indices such as the Fire Weather Index (FWI) and Keetch-Byram Drought Index (KBDI) to quantify fire risk levels.
2. **Temporal Features:** Extraction of temporal features such as day of the year, month, and season to capture seasonal patterns.
3. **Interaction Terms:** Generation of interaction terms between climatic variables (e.g., temperature and humidity) to capture combined effects.
4. **Lagged Variables:** Creation of lagged variables to incorporate past climatic conditions into the predictive models, acknowledging that fire risk is often influenced by previous weather patterns.

4.6 Summary

This chapter presented the primary data sources, including forest fire data from Google Earth Engine and climatic data from the IMF Climate Change Dashboard. The EDA involved generating visualizations and descriptive statistics to explore the data and gain initial insights. Feature engineering steps were detailed, focusing on generating additional variables to enhance the predictive accuracy of the models.

4.7 Expected Outcome

The expected outcomes of this research include:

1. A comprehensive understanding of the temporal and spatial patterns of forest fires in Peninsular Malaysia.
2. Identification of key climatic factors influencing fire occurrences.
3. Development of predictive models with enhanced accuracy and reliability.
4. Contributions to fire management strategies and early warning systems in the region.

These outcomes will advance knowledge in data science and environmental studies, providing valuable insights for policymakers and researchers.

4.8 Future Work

Potential avenues for future research include:

1. Application to Other Regions: Extending the analysis to other regions with similar climatic conditions to validate the findings.
2. Incorporating Additional Data Sources: Integrating more data sources, such as socio-economic factors, to explore their influence on fire patterns.
3. Advanced Modelling Techniques: Employing more advanced machine learning algorithms to further improve prediction accuracy.
4. Longitudinal Studies: Conducting longitudinal studies to monitor the long-term impacts of climate change on forest fire patterns.

By addressing these future research directions, the study demonstrates a forward-thinking approach and the broader applicability of its findings.