Journal of Applied Data Sciences
Vol. 5, No. 4, December 2024, pp. 2054-2066

ISSN 2723-6471
2054

Data Visualization of Climate Patterns in Indonesia Using Python and Looker Studio Dashboard: A Visual Data Mining Approach

Rina Refianti[1], Achmad Benny Mutiara[2,*], Ananda Satria Ariyanto[3]

[1,2,3]*Faculty of Computer Science and Information Technology, Gunadarma University, Margonda Raya No.100, Depok 16424, Indonesia*

(Received: June 21, 2024; Revised: August 14, 2024; Accepted: October 22, 2024; Available online: November 7, 2024)

Abstract

Climate has a significant impact on the lives of Indonesian people. Information about climate patterns, when presented visually and interactively, can greatly enhance understanding of climate conditions in Indonesia. This study aims to produce a visualization of climate pattern data in Indonesia that can be accessed online by the general public, serving as a valuable resource for climate information. The study highlights the ability to display historical trends for a 10-year period (2010-2020) through interactive visuals, which load information according to user-defined filters, enabling diverse presentations of data. The research employs the Visual Data Mining method, encompassing Project Planning, Data Preparation, and Data Analysis phases. Additionally, Exploratory Data Analysis techniques were utilized in the data analysis phase. The data was cleaned and processed using the Python programming language with libraries such as pandas, numpy, seaborn, and matplotlib. Visualizations were created using Looker Studio tools and published on a website, providing accessible climate pattern information in Indonesia via the Internet. The final results of this research indicate that the developed climate visualization dashboard successfully delivers detailed insights into sunlight duration, temperature, humidity, rainfall, and wind speed across various Indonesian regions. Users can effectively monitor climate trends and weather changes. The dashboard also demonstrates significant seasonal variations and differences in climate patterns between provinces. Performance metrics reveal that the dashboard meets Key Performance Indicators, achieving a click-through ratio of 40.1%, the average page position in search engines is 4.8 top positions, and receiving positive user experience scores. Further development and research on the Climate Pattern Dashboard in Indonesia still have room for enhancement. Important aspects include expanding data coverage to include multiple decades for observing significant climate patterns and applying sophisticated prediction methods like machine learning algorithms for future climate change projections.

Keywords: Climate, Data Visualization, Looker Studio, Python, Visual Data Mining

1. Introduction

Climate plays a critical role in people's daily lives, influencing various sectors such as agriculture, transportation, health, and the environment. As an archipelagic country surrounded by vast oceans, Indonesia is particularly vulnerable to the impacts of climate change. Shifts in climate patterns—whether through rising temperatures, increased frequency of natural disasters, or changes in rainfall—have wide-reaching effects across Indonesia’s sectors.

Given these circumstances, accurate and accessible climate pattern information is essential to raise public awareness of climate conditions in Indonesia. This awareness is crucial for fostering more informed mitigation and adaptation strategies in response to climate change. One effective way to convey climate information is through data visualization, as visual representations tend to be more comprehensible than text-heavy explanations. Additionally, data visualization has a strong potential to increase public engagement by presenting information in a visually appealing and efficient manner.

Motivated by this potential, this study aims to present climate pattern information in a user-friendly manner and provide a tool that can be accessed by a wider audience in the form of a visualization dashboard. The goal is to not only make climate data more understandable but also to enhance public access to it. Several studies have explored related topics. For example, "Visualization of Maritime Weather Data at BMKG Maritime Semarang" [1] focuses on maritime weather data but is limited geographically to Semarang. This research, however, broadens the scope to include climate

***Corresponding author: Achmad Benny Mutiara (amutiara@staff.gunadarma.ac.id)**
DOI: https://doi.org/10.47738/jads.v5i4.420

patterns across all of Indonesia. Another relevant study, "Exploration of Climate Data using Interactive Visualization" [2], emphasizes user interaction, which this research builds upon by offering deeper user engagement and exploration features. "Visualization of Climate and Climate Change Data: An Overview" [3] discusses visualization techniques for climate science, and this research extends those techniques by integrating specific test parameters to enhance clarity and accessibility.

Moreover, "Exploratory Data Analysis and Multivariate Strategies for Revealing Multivariate Structures in Climate Data" [4] applies EDA techniques to analyze climate data. This study combines Exploratory Data Analysis (EDA) with Visual Data Mining (VDM) to deliver more comprehensive insights. In "Visual Data Mining: Techniques and Tools for Data Visualization and Mining" [5], various tools for environmental data mining are reviewed, which are employed in this research to uncover significant climate patterns. Additionally, "Big Data Analysis Using the EDA Method and Visualization with Jupyter Notebook" [6] showcases the use of EDA for big data analysis before visualization, a method closely aligned with the approach in this study, which merges both EDA and VDM for enhanced data analysis and presentation.

Recognizing the importance of making climate data accessible to the public through visual means, this study aims to develop a climate pattern data visualization dashboard for Indonesia. The dashboard, designed using Python and Looker Studio, will provide an online platform for visual data mining (VDM). Thus, this research is titled "Visualization of Climate Pattern Data in Indonesia Using Python and Looker Studio Dashboard: A Visual Data Mining (VDM) Approach."

2. Literature Review

2.1. Information System

A system is defined as a set of interconnected components that work together to achieve a specific goal. Systems are typically characterized by the presence of inputs, processes, and outputs, where these components interact to perform functions that contribute to the overall objective. Systems can exist in various forms, from mechanical systems to organizational systems, and usually follow structured processes with predefined rules or methods [7].

Information, on the other hand, refers to processed data that is meaningful and useful for decision-making. Raw data, often presented as a collection of facts, numbers, or statistics, becomes valuable information once organized and processed to support problem-solving or decision-making [8]. An information system is thus a combination of components such as technology, data, methods, and processes, which work together to collect, process, store, and distribute information. It facilitates the transformation of raw data into actionable information, supporting decision-making processes. The system generally comprises hardware, software, databases, networks, and human resources, working together to ensure the generated information is accurate, timely, and relevant [9].

2.2. Weather Data Set

According to the Meteorology, Climatology, and Geophysics Agency (BMKG), weather data represents real-time information about atmospheric conditions at a specific time and location. This data is gathered from observations, measurements, and analyses performed at Weather Observation Stations. Weather data includes variables such as temperature, humidity, wind speed, and atmospheric pressure, which help in understanding and predicting short-term atmospheric changes [10]. In contrast, climate data refers to the aggregation of weather data over extended periods, often decades, capturing long-term trends and patterns in atmospheric conditions. This historical data is essential for studying phenomena like global warming or regional climate variability. Unlike weather data, which is used for immediate predictions, climate data is compiled and analyzed retrospectively to detect large-scale environmental changes [10].

2.3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a key approach used by data scientists to analyze and investigate data sets in order to summarize their primary characteristics. It often employs data visualization techniques to uncover hidden patterns, detect outliers, and identify key variables within the data. EDA is typically used as an initial exploration step, focusing on what the data can reveal rather than testing specific assumptions or hypotheses [11]. Tasks involved in EDA include data cleaning, descriptive statistical analysis, and data quality assessments. Descriptive statistics provide insights into the central tendency and distribution of the data, while data visualization, such as bar charts, scatter plots, histograms, and box plots, allows for easier identification of trends, patterns, and anomalies. In this research, EDA is pivotal, as visualizing data helps present complex information in an accessible manner, forming the foundation for more advanced data analysis and decision-making.

2.4. Data Visualization

Data refers to factual information about objects or events, often represented by numbers, characters, or symbols. Over time, the definition of data has expanded to include formats such as text, graphics, sound, and video. Data serves as the fundamental element for analysis and insight extraction across various fields, especially in decision-making and business intelligence [12]. Visualization is the process of graphically representing data. It enables users to explore and present data interactively, transforming raw information into visual formats that are easier to interpret and analyze. Visualization utilizes computer-based tools to depict complex data visually, facilitating better observation and comprehension [13]. Through visualization, knowledge is systematically conveyed via images or graphical representations, revealing patterns, relationships, and insights that might not be apparent from raw data. Data visualization involves transforming data into meaningful visual representations like charts, graphs, and maps. This method allows users to intuitively understand complex data sets by visually presenting patterns, trends, and outliers quickly and effectively [14]. In fields like science, business, and journalism, data visualization is essential as it simplifies large datasets, making insights more accessible and actionable for decision-making and communication.

2.5. Weather and Climate

Weather refers to the atmospheric conditions in a specific location over a short period. It is a combination of various elements such as temperature, humidity, wind speed, and direction, which can change within hours or days [15]. For instance, morning weather conditions might differ significantly by the afternoon or evening. Weather is dynamic and fluctuates frequently over short timeframes. In contrast, climate is the long-term average of weather conditions observed over extended periods, often decades, in a specific region [16]. It provides a statistical representation of atmospheric factors over time, offering a more stable view of a region's general weather characteristics [17]. Unlike weather, which changes rapidly, climate shifts occur gradually and can influence large ecosystems and regions. The distinction between weather and climate is essential in atmospheric studies. While weather focuses on short-term fluctuations, climate emphasizes long-term patterns and trends. Climate data, accumulated over multiple decades, is crucial for understanding global phenomena such as climate change, enabling predictions of future environmental and societal impacts [18].

2.6. Python

Python is a widely-used high-level programming language, popular across industries and educational institutions. Created by Guido van Rossum in 1990, Python's simplicity and versatility have made it a preferred choice for developing various applications, from desktop to web and mobile platforms [19]. Python's extensive libraries, including NumPy and Pandas for data analysis, TensorFlow for artificial intelligence development, and Flask and Django for web development, make it applicable in diverse fields such as data science, machine learning, and system automation [20], [21]. It’s simple, readable syntax facilitates efficient development, making it an ideal tool for handling complex tasks with manageable code [22]. In the era of big data, Python has become a leading language for large-scale data analysis. Libraries like Pandas and Matplotlib provide powerful tools for data manipulation, analysis, and visualization. This has solidified Python's status as a primary tool in data science and big data applications [23]. As noted by [24], Python's simplicity allows developers to handle complex data with easy-to-understand code, further enhancing its popularity in data-driven fields.

3. Method

This research utilizes the VDM approach, which is composed of several essential stages. Each stage is carefully designed to ensure a thorough and systematic process, from initial planning to the final analysis, making use of data visualization techniques to extract valuable insights from complex datasets. As illustrated in figure 1, the VDM Method is divided into three main phases: Project Planning, Data Preparation, and Data Analysis. Each phase includes specific steps that guide the entire data mining process, ensuring a clear path from project initiation to the execution of the analysis and the creation of visual outputs.

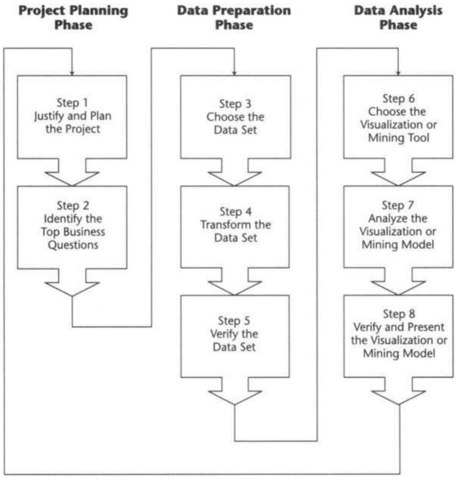

Figure 1. Visual Data Mining Method

3.1. Project Planning Phase

The first stage in this research is the Project Planning Phase, which is critical for setting the foundation of the study. In this phase, the primary objective is to clearly define the research goals and scope. This begins with identifying the core research problem or question that the study aims to solve, which in this case revolves around understanding and visualizing climate patterns in Indonesia. After defining the problem, the data requirements are outlined, including the specific types of data needed, such as weather and climate data from reliable sources like the Meteorology, Climatology, and Geophysics Agency (BMKG). A strategic plan is then developed, detailing the steps needed to collect, process, and analyze the data. Additionally, evaluation metrics are determined at this stage to measure the success of the project, ensuring that the outcomes align with the research objectives. This stage serves as a blueprint, guiding the research team in executing each subsequent phase methodically and efficiently.

3.2. Data Preparation Phase

Following the project planning, the research enters the Data Preparation Phase, where the focus shifts to handling the data itself. This phase involves several important steps, starting with data selection, which entails choosing relevant data sources that meet the research criteria. Once selected, the data is collected from the identified sources, which may involve retrieving datasets from government databases, climate monitoring stations, or other institutions. The next critical step is data cleaning, where any errors, missing values, or inconsistencies in the data are corrected to ensure its quality. Data transformation may also be performed if necessary, reformatting the data into structures suitable for analysis. Additionally, data from various sources may be combined, ensuring the dataset is comprehensive and coherent. Any irrelevant or redundant data that does not contribute to the research objectives is discarded to prevent it from distorting the results. The overall goal of this phase is to produce a clean, well-organized dataset that is ready for analysis, free from defects that could undermine the validity of the findings.

3.3. Data Analysis Phase

The third and final stage is the Data Analysis Phase, where the prepared data is explored and analyzed using various data visualization techniques. This phase plays a crucial role in uncovering hidden patterns, trends, and insights within the dataset. The process begins by applying exploratory data analysis (EDA) techniques, which involve creating a range of visual representations, such as graphs, bar charts, scatter plots, and heatmaps. These visualizations make it easier to identify relationships between variables, such as correlations between temperature and rainfall or the frequency of extreme weather events over time. Through the use of these visual tools, researchers can detect anomalies, such as outliers in temperature or precipitation patterns, which might indicate unusual climate phenomena. The visualizations also enable the identification of trends, such as gradual increases in average temperatures, which could suggest long-term climate shifts. Ultimately, this phase allows the research team to make data-driven decisions, supported by the

 ISSN 2723-6471

evidence revealed through visual exploration. The use of VDM in this phase ensures that complex climate data is transformed into meaningful insights, which are accessible and comprehensible to both researchers and the general public.

4. Results and Discussion

This section outlines the steps involved in the development of the dashboard and its web-based implementation, based on the Visual Data Mining (VDM) method.

4.1. Project Planning Phase

4.1.1. Justify and Plan the Project

The main goal of this research is to create a climate pattern visualization dashboard for Indonesia that is accessible to the public. This dashboard is designed to provide insights into daily, monthly, and annual climate patterns, presenting information on key variables such as sunshine duration, average temperature, rainfall, humidity, wind speed, and wind direction. Furthermore, it includes data on climate distribution across provinces and observation stations. A regional distribution map is also incorporated to visually highlight climate variations in different parts of Indonesia, offering users a comprehensive view of how climate conditions change across the country.

To assess the success of the project, several Key Performance Indicators (KPIs) have been defined. First, user visits are measured by tracking the number of visitors accessing the dashboard through Google search. This metric helps determine how effectively the dashboard is reaching its intended audience. Second, the number of Page Views is monitored to understand how frequently the dashboard appears in search results, indicating the visibility of the page. Third, the Click-Through Rate (CTR) is used to measure the percentage of users who click on the dashboard link after seeing it in search results. A higher CTR suggests that the dashboard is attracting users with relevant and compelling content. Finally, Search Ranking is analyzed to assess how well the dashboard ranks for relevant keywords in search engines. This ranking reflects how easy it is for users to find the dashboard when searching for related climate information.

4.1.2. Identify the Top Business Questions

To ensure the dashboard is aligned with its objectives, several key business questions were identified. These questions help guide the development of the visualization and ensure it addresses important aspects of climate analysis in Indonesia. One of the core questions focuses on how key climate variables such as sunshine duration, temperature, humidity, rainfall, and wind speed vary across different regions of Indonesia over time. By answering this question, the dashboard provides users with detailed insights into how climate conditions change in different parts of the country.

Another important question examines how climate patterns differ on daily, monthly, and yearly timescales. This exploration helps users understand the temporal dynamics of climate change, revealing both short-term fluctuations and long-term trends. Lastly, the dashboard seeks to answer whether there are significant variations in climate patterns between provinces or among different observation stations. This question allows for regional comparisons, enabling users to see how climate patterns may vary across different parts of the country, which can inform more localized climate adaptation and mitigation efforts.

4.2. Data Preparation Phase

4.2.1. Choose the Data Set

The dataset used in this study, "Climate Data Daily IDN," was sourced from the Kaggle platform, which is a well-known repository for data science and machine learning practitioners. The dataset, uploaded by Greeg Titan and Elmajo Adriel, originates from the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG). This dataset includes essential climate variables such as daily temperature, rainfall, humidity, wind speed, and sunshine duration, making it highly relevant for comprehensive climate pattern analysis. The trusted source and data coverage across various Indonesian regions ensure the dataset’s reliability for this research.

The attributes in the dataset are outlined in table 1, which includes descriptions of variables like minimum temperature (Tn), maximum temperature (Tx), average temperature (Tavg), average humidity (RH_avg), rainfall (RR), sunshine duration (ss), wind speed, wind direction, and weather station details.

Tabel 1. Data Set climate_data.csv

Attribute	Description
Tn	Minimum temperature (°C)
Tx	Maximum temperature (°C)
Tavg	Average temperature (°C)
RH_avg	Average humidity (%)
RR	Rainfall (mm)
ss	Duration of sunshine (jam)
ff_x	Maximum wind speed (m/s)
ddd_x	Wind direction with maximum speed (°)
ff_avg	Average wind speed (m/s)
ddd_car	Mostly wind direction (°)
station_id	Weather station ID

4.2.2. Transform the Data Set

The Data Transformation Phase involves preparing the dataset for analysis by integrating and cleaning the data. The process begins by importing the necessary libraries, including pandas for managing tabular data, numpy for performing numerical operations, and seaborn and matplotlib for data visualization. Once the libraries are in place, the next step is to merge the different CSV files containing climate data, station details, and provincial information to create a more comprehensive dataset. The merging process is executed using the pd.merge() function, which combines the datasets based on shared identifiers, such as station IDs and province IDs. After the merging process is complete, the dataset's structure is verified using the .info() method to ensure that all fields have been correctly integrated and are ready for further processing.

After merging the data, the next crucial step is to clean the dataset by addressing missing values and handling potential outliers. The Data Cleaning process starts by identifying missing values in the dataset. A summary of null values is generated to detect which columns contain gaps. Based on the distribution of the data, missing values are either replaced with the mean or median. For instance, temperature-related columns, which tend to follow a normal distribution, are filled with the mean. On the other hand, humidity and rainfall columns, which might include outliers, are filled with the median to avoid skewing the data. In addition to filling missing values, the dataset is scanned for duplicate rows, which are removed to prevent redundant entries from negatively affecting the analysis results.

The next step is Outlier Detection and Handling, where boxplots are used to visualize and identify outliers in numerical columns such as temperature, humidity, and wind speed. Outliers can distort the analysis, so they are handled using the Interquartile Range (IQR) method. The IQR method identifies extreme values that fall outside the typical data range, and these outliers are removed to ensure the dataset remains clean and reliable for analysis. By eliminating extreme values, the integrity of the dataset is preserved, making it more representative of typical climate conditions across Indonesia.

Once the cleaning process is complete, the final step is to Verify the Data Set. This involves performing a final check to ensure that all missing values have been appropriately addressed. A missingno matrix is used to visually confirm that no null values remain in the dataset. Additionally, boxplots are generated once again to verify that significant outliers have been effectively removed. This final verification step ensures that the dataset is thoroughly cleaned and prepared for the subsequent stages of data analysis and visualization.

ISSN 2723-6471

Figure 2 provides a detailed summary of the merged dataset after integrating various data sources related to climate patterns in Indonesia. The dataset contains a total of 589,265 entries spread across 19 columns. Each column represents a specific variable, such as climate metrics, geographical information, and metadata about weather stations. For instance, columns like **Tn**, **Tx**, and **Tavg** capture the minimum, maximum, and average temperature values in degrees Celsius, respectively, while **RH_avg** denotes the average humidity and **RR** indicates rainfall in millimeters. Additional columns, such as **SS** (sunshine duration), **ff_x** (maximum wind speed), and **ddd_x** (wind direction), provide crucial insights into climate conditions.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 589265 entries, 0 to 589264
Data columns (total 19 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   date           589265 non-null  object
 1   Tn             565882 non-null  float64
 2   Tx             551529 non-null  float64
 3   Tavg           544160 non-null  float64
 4   RH_avg         541083 non-null  float64
 5   RR             463881 non-null  float64
 6   ss             545544 non-null  float64
 7   ff_x           579051 non-null  float64
 8   ddd_x          576137 non-null  float64
 9   ff_avg         579138 non-null  float64
 10  ddd_car        575526 non-null  object
 11  station_id     589265 non-null  int64
 12  station_name   589265 non-null  object
 13  region_name    589265 non-null  object
 14  latitude       589265 non-null  float64
 15  longitude      589265 non-null  float64
 16  region_id      589265 non-null  int64
 17  province_id    589265 non-null  int64
 18  province_name  589265 non-null  object
dtypes: float64(11), int64(3), object(5)
memory usage: 85.4+ MB
```

Figure 2. Info After Data Sets Are Merged

The figure also highlights that the dataset includes some missing values, with non-null counts for each column varying slightly. For example, the **Tn** (minimum temperature) column has 565,882 non-null entries, meaning some entries are missing. Similarly, **RH_avg** and **RR** have fewer non-null values than the total number of rows, reflecting the presence of missing data. The data types for the columns include both numerical (e.g., *float64* for climate measurements like temperature and wind speed) and categorical or textual types (e.g., *object* for columns like station_name, province_name, and date). Numerical values are essential for statistical analysis and visualization, while textual columns provide metadata that enrich the dataset. Moreover, the dataset is efficiently stored with a memory usage of 85.44 MB, making it manageable for analysis. This overview, generated using the .info() function from the *pandas* library, ensures that the dataset is ready for further cleaning and analysis by providing insight into the structure, quality, and completeness of the data.

Figure 3 presents a detailed summary of the missing data across various columns in the dataset. It provides two key pieces of information for each variable: the Jumlah Null, which indicates the number of missing (null) values, and the Persentase Null (%), representing the percentage of missing data relative to the total entries.

	Jumlah Null	Persentase Null (%)
date	0	0.000000
min_temperature	23383	3.968164
max_temperature	37736	6.403910
avg_temperature	45105	7.654451
avg_humidity	48182	8.176627
rainfall	125384	21.278033
duration_sunshine	43721	7.419582
max_wind	10214	1.733346
wind_direct_max	13128	2.227860
avg_wind	10127	1.718582
most_wind_direct	13739	2.331549
station_id	0	0.000000
station_name	0	0.000000
region_name	0	0.000000
latitude	0	0.000000
longitude	0	0.000000
region_id	0	0.000000
province_id	0	0.000000
province_name	0	0.000000

Figure 3. Identify Null Values

The figure reveals that certain climate variables have significant gaps in their data. The rainfall column, for instance, has the largest amount of missing data, with 125,834 null entries, accounting for approximately 21.27% of the dataset. Similarly, the avg_temperature column is missing data for about 7.67% of entries, while min_temperature and max_temperature have 3.96% and 6.40% missing data, respectively. Other key climate variables, such as avg_humidity and duration_sunshine, also show a notable percentage of missing data, with 8.17% and 7.37%, respectively.

 ISSN 2723-6471

Variables related to wind measurements, including max_wind, wind_direct_max, and avg_wind, have fewer missing values, with each showing around 2% missing data. These relatively lower percentages suggest that wind-related data is more complete compared to other climate metrics. On the other hand, several columns, such as station_id, station_name, region_name, latitude, longitude, province_id, and province_name, have no missing values at all, indicating that all geographical and metadata-related information is fully present in the dataset.

This figure is essential for diagnosing the dataset's quality, highlighting which columns will require imputation or special handling during the data cleaning process. For example, columns like rainfall and avg_temperature, which have higher percentages of missing values, will need careful attention. This might involve applying methods such as replacing missing values with the mean or median, or removing rows with excessive missing data, depending on the analysis requirements.

Figure 4 provides a series of boxplots that illustrate the distribution of key climate variables, such as min_temperature, max_temperature, avg_temperature, avg_humidity, rainfall, duration_sunshine, max_wind, wind_direct_max, and avg_wind. The boxplots are used to detect outliers, which are data points that deviate significantly from the general range of values. The central box in each plot represents the interquartile range (IQR), covering the middle 50% of the data, with a line indicating the median value. The whiskers extending from the box show the data points within 1.5 times the IQR, while any points outside the whiskers are classified as outliers. Notably, max_temperature and rainfall exhibit a large number of outliers, which could be attributed to extreme weather events or other anomalies in climate data.

The presence of outliers in variables like duration_sunshine, max_wind, and avg_wind is also observed, though these are less extreme compared to the temperature and rainfall data. Wind-related variables, especially wind_direct_max, show a more consistent distribution, with fewer outliers, suggesting a more stable pattern. Identifying these outliers is critical in the data preparation process, as they can influence the results of statistical analyses or machine learning models. Thus, handling outliers—whether by removing, transforming, or capping them—is necessary to ensure accurate and reliable analysis. Figure 4 highlights the importance of addressing these outliers to maintain the quality of the data used for further analysis and decision-making.

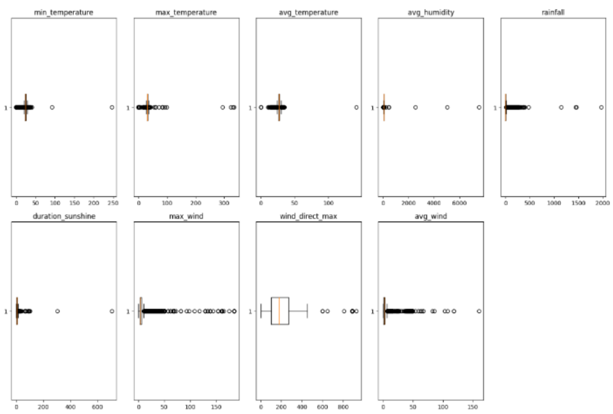

Figure 4. Identify Outliers

Figure 5 provides a clear visualization of the dataset's completeness after the data cleaning process. Each vertical bar represents a column in the dataset, with the bars indicating the presence or absence of data. In this figure, all the bars are fully filled, indicating that there are no missing values remaining in any of the 19 columns, such as min_temperature, max_temperature, avg_temperature, rainfall, and others. This suggests that the missing data identified earlier has been successfully addressed, either through imputation, removal, or another cleaning method, ensuring a complete dataset.

Figure 5. Seeing Null Values with Visual Missingno

The use of the *missingno* matrix offers an intuitive way to assess data completeness, allowing researchers to quickly confirm that the dataset is ready for further analysis. The visualization provides a high-level overview, ensuring that all columns are now fully populated with data, free of gaps or null values. This final check is crucial for verifying the integrity of the dataset before proceeding with statistical analysis or machine learning models, as missing data could otherwise distort the results.

Figure 6 showcases boxplots for key climate variables, including min_temperature, max_temperature, avg_temperature, avg_humidity, rainfall, duration_sunshine, max_wind, wind_direct_max, and avg_wind, after the data cleaning process. The cleaning methods applied, such as the removal of extreme values or transformations, appear to have successfully reduced the number of outliers for most variables. For instance, the temperature variables (min, max, and average) show compact distributions with no visible outliers, indicating that the data is now more uniform. Similarly, avg_humidity and wind-related variables (such as max_wind and avg_wind) also show clean distributions, suggesting that the cleaning process has effectively handled any irregularities in these variables.

However, the rainfall variable still displays a noticeable number of outliers, as shown by points that lie beyond the whiskers of the boxplot. This is not unusual for rainfall data, as extreme weather events can result in highly variable precipitation, leading to genuine outliers. While the cleaning process has reduced some of the more extreme values, these remaining outliers may reflect real-world phenomena and could be important for further analysis. Therefore, while most variables are free from artificial outliers, the rainfall variable should be treated with caution, as these values may provide valuable insights into unusual weather patterns. Overall, the figure confirms that the dataset has been largely cleaned and is ready for deeper analysis.

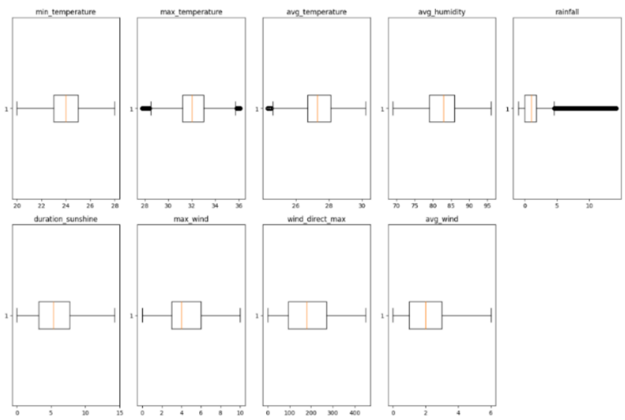

Figure 6. Seeing Outliers After Cleaning

4.3. Data Analysis Phase

4.3.1. Choose the Visualization Tool

At this stage, appropriate tools and software were selected to perform data visualization. The chosen tools were Python with the Google Colab platform, combined with Looker Studio for dashboard creation. The flexibility and versatility of Python and Google Colab in data analysis, along with their powerful visualization capabilities, made them ideal for this task. Additionally, Looker Studio was used for creating interactive visualizations that could be easily integrated into websites, making the dashboard accessible to a wider audience.

4.3.2. Analyze the Visualization

Figure 7 displays the first dashboard, which provides extensive information on climate data across Indonesia. The dashboard includes the distribution of sunlight duration, temperature, humidity, rainfall, and average wind speed across different provinces and regions. Users are provided with several interactive filters, such as the "Province Filter", "Year Filter", and a "Map" filter. These tools allow users to explore data by province, track climate trends over the years, and visualize data geographically. By using these filters simultaneously, users can perform a more specific and detailed analysis of climate conditions across Indonesia. For instance, users can compare climate variations between provinces or observe how rainfall patterns have changed over time.

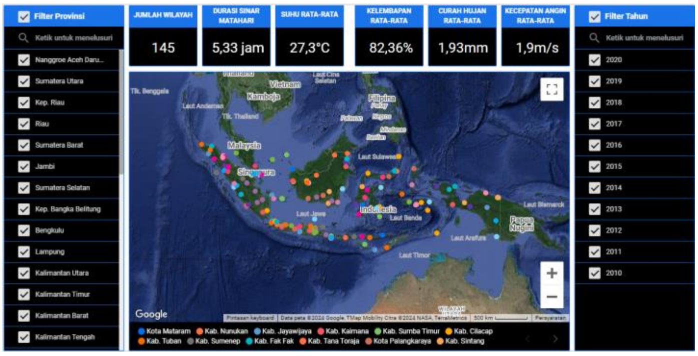

Figure 7. Dashboard 1 Climate Pattern Data in Indonesia

Figure 8 presents a visualization of wind direction patterns across Indonesia. The data can be explored on a daily, monthly, or annual basis, allowing users to analyze changes in wind direction over time. This visualization provides valuable insights into how wind patterns shift due to seasonal or yearly factors, helping users understand regional wind behavior and its potential impact on local climates.

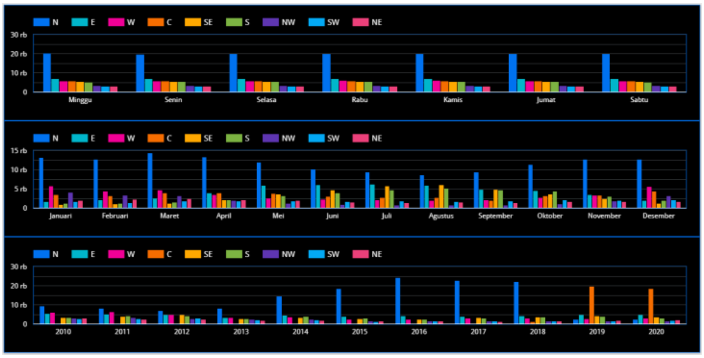

Figure 8. Wind Direction Visualization

Figure 9 displays the distribution of climate data based on province, region, and weather stations. This dashboard helps users identify which provinces have the most data, which regions are more actively monitored, and which weather stations are responsible for the largest amount of data collection. This overview allows users to understand the geographical coverage of climate data, highlighting the areas that are most frequently measured and monitored.

Figure 9. Data Distribution Visualization

Figure 10 illustrates the daily, monthly, and annual climate patterns in Indonesia. The daily dashboard visualizes variables such as temperature, humidity, wind speed, rainfall, and sunshine duration for each day of the week, from Monday to Sunday. Similarly, the monthly and annual dashboards present these variables for each month and year (from 2010 to 2020). The comprehensive charts and maps in these dashboards allow users to filter and explore data across different timeframes, offering a detailed view of climate changes over time.

Figure 10. Daily, Monthly and Yearly Climate Pattern Dashboard in Indonesia

4.3.3. Verify and Present the Visualization

After the dashboard was completed, the next step was to deploy it online via GitHub (https://anandasatriaa.github.io/visualisasi-data-iklim-di-indonesia/). This process started by creating a README.md file with the project title "visualisasi-data-iklim-di-indonesia," providing an overview of the project. A Git repository was initialized locally using the git init command, and the necessary files were added using git add. The project was committed using the git commit -m "first commit" command and then linked to a remote GitHub repository using git remote add origin <URL>. Finally, the project was pushed to GitHub using git push -u origin main, making the dashboard publicly accessible and enabling easy collaboration and updates.

To enhance the visibility of the dashboard in Google search results, several SEO (Search Engine Optimization) strategies were implemented. The website was submitted to Google Search Console for indexing, and meta tags were added to the HTML to improve ranking for relevant keywords such as "visualisasi data iklim di Indonesia." Keywords were strategically placed throughout the content, making the dashboard easier to discover in search engines. Additionally, website performance was regularly monitored to track its ranking, click-through rate (CTR), and overall visibility in search results.

5. Conclusion

The conclusion of this research shows that the climate visualization dashboard developed has succeeded in providing detailed information regarding patterns of sunlight duration, temperature, humidity, rainfall and wind speed in various regions in Indonesia at all times. This information allows users to monitor climate trends and weather changes more clearly and effectively. Data visualization also shows clearer differences in climate patterns when analyzed by month and year rather than by day. For example, the rainy season and dry season are more clearly visible in the analysis of rainfall by month and year. In addition, patterns of temperature, humidity and duration of sunlight in various regions also show significant seasonal variations. Apart from differences in climate patterns based on time, the research results also show that there are significant differences between provinces or observation stations. Some provinces or observation stations recorded higher rainfall than other provinces, while certain provinces or observation stations showed higher wind speeds, which may be caused by geographic factors or other local conditions.

In terms of effectiveness, this climate visualization dashboard has succeeded in meeting the specified Key Performance Indicators (KPIs). Based on analysis results from Google Search Console after 3 months of the website being active, it was found that this dashboard received a total of 97 clicks from 242 impressions on the Google search page, with an average click-through ratio (CTR) of 40.1%. This means that for every page appearance in search results, around 40.1% of users clicked on the dashboard link. The average page position in search engines is 4.8 top positions, and a total of 5 countries have seen impressions of the dashboard, with 152 impressions from desktop and 90 impressions from mobile devices. In terms of user experience, testing with GTmetrix shows that this dashboard website has an average score of A, with performance of 96% and structure of 83%. The main content loads in 419ms with a speed index of 2 seconds, no time block and CLS 0. Testing using PageSpeed Insights resulted in a performance score of 59 (average), accessibility 86 (average), best practices 56 (average), and SEO 100 (very good). Dashboard content loads in 800ms with a speed index of 11 seconds.

Overall, the Dashboard is not only effective in presenting informative climate data, but also performs quite well in search engines as well as in terms of user experience. Future development of the Climate Pattern Dashboard in Indonesia can focus on expanding data coverage beyond the current decade (2010-2020) to capture longer-term climate trends. Incorporating machine learning algorithms for climate projections could enhance its predictive capabilities. Improving the interface and adding features like dynamic maps and customizable tools would make the dashboard more engaging. Additionally, optimizing SEO with relevant keywords like "climate visualization in Indonesia" and adding backlinks from high-traffic websites can increase its visibility in search engines.

6. Declarations

6.1. Author Contributions

Conceptualization: R.R., A.B.M.; Methodology: A.S.A., A.B.M.; Software: R.R., A.S.A.; Validation: A.B.M., A.S.A.; Formal Analysis: R.R., A.S.A.; Investigation: A.B.M., A.S.A.; Resources: R.R., A.B.M.; Data Curation: A.S.A., A.B.M.; Writing—Original Draft Preparation: R.R., A.S.A.; Writing—Review and Editing: A.B.M., A.S.A.; Visualization: A.S.A., A.B.M.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K. Wang, M. Liang, Y. Li, J. Liu, and R. W. Liu, "Maritime traffic data visualization: A brief review," *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, pp. 67-72, 2019, doi: 10.1109/ICBDA.2019.8713227.

[2] F. Ladstädter, A. K. Steiner, B. C. Lackner, G. Kirchengast, and H. P. Schmidt, "Exploration of climate data using interactive visualization," *J. Atmos. Oceanic Technol.*, vol. 27, no. 4, pp. 667–679, Apr. 2010.

[3] T. Nocke, T. Sterzel, M. Böttinger, and M. Wrobel, "Visualization of climate and climate change data: An overview," *in Proc. Digital Earth Summit on Geoinformatics*, Jan. 2008, pp. 226–232.

[4] C. M. Igwenagu, "Exploratory data analysis and multivariate strategies for revealing multivariate structures in climate data," *J. Exp. Res.*, vol. 4, no. 2, Oct. 2016, ISSN: 2502-0524.

[5] T. Soukup and I. Davidson, *Visual Data Mining: Techniques and Tools for Data Visualization and Mining*, John Wiley & Sons, Sep. 2002.

[6] M. Fiorini, A. Capata, and D. Bloisi, "AIS data visualization for maritime spatial planning (MSP)," *International Journal of e-Navigation and Maritime Economy*, vol. 5, pp. 45-60, 2016, doi: 10.1016/J.ENAVI.2016.12.004.

[7] D. Kroenke, M. Boyle, "Evaluating the data visualization for demanding marine operations," *2016 Techno-Ocean (Techno-Ocean)*, pp. 471-477, 2016, doi: 10.1109/TECHNO-OCEAN.2016.7890700.

ISSN 2723-6471

[8] N. Silva, B. Webber, A. Matthews, M. M. Feist, T. Stein, C. Holloway, M. F. A. B. Abdullah, "Validation of GPM IMERG extreme precipitation in the maritime continent by station and radar data," *Earth and Space Science*, vol. 8, 2021, doi: 10.1029/2021EA001738.

[9] K. C. Laudon and J. P. Laudon, *Management Information Systems: Managing The Digital Firm (16th ed.)*, Pearson Educación, 2020.

[10] V. Capozzi, C. De Vivo, Y. Cotroneo, C. Paciello, and F. Mignone, *"Measurements for meteorology," in Measurement for the Sea: Supporting the Marine Environment and the Blue Economy, Cham: Springer Int. Publ.*, Jan. 2022, pp. 27–49.

[11] A. Jariwala, A. Chaudhari, C. Bhatt, B. Patel, and A. Shah, "Data quality for AI tool: Exploratory data analysis on IBM API," *Int. J. Intell. Syst. Appl.*, vol. 1, no. 13, pp. 42–56, Feb. 2022, doi: 10.5815/ijisa.2022.01.04.

[12] C. Claramunt, C. Ray, E. Camossi, A. Jousselme, M. Hadzagic, G. Andrienko, N. Andrienko, Y. Theodoridis, G. Vouros, and L. Salmon, "Maritime data integration and analysis: recent progress and research challenges," *EDBT 2017*, vol. 459, pp. 192-197, 2017, doi: 10.5441/002/edbt.2017.18.

[13] K. Börner, A. Bueckle, and M. Ginda, "Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments," *Proc. Nat. Acad. Sci*., vol. 116, no. 6, pp. 1857–1864, Feb. 2019, doi: 10.1073/pnas.1807180116.

[14] A. J. Cardno, "Data visualization system and method," U.S. Patent 9,418,456, Aug. 16, 2016.

[15] Patil Prasad Vinayak, C. Sree Krishna Prabu, N. Vishwanath, and S. Prakash, "Numerical simulation of ship navigation in rough seas based on ECMWF data," *Brodogradnja*, vol. 72, no. 1, pp. 1-20, 2021, doi: 10.21278/BROD72102.

[16] IPCC, "Climate change 2007: The physical science basis," *Agenda*, vol. 6, no. 7, pp. 333–345, May 2007.

[17] P. S. Quinan, M. D. Meyer, "Visually comparing weather features in forecasts," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 389-398, 2016, doi: 10.1109/TVCG.2015.2467754.

[18] Z. Liu, C. L. Shie, A. Li, J. Smith, L. Yang, and M. Brown, "NASA global satellite and model data products and services for tropical meteorology and climatology," *Remote Sens*., vol. 12, no. 17, p. 2821, Aug. 2020, doi: 10.3390/rs12172821.

[19] A. Diehl, L. Pelorosso, C. Delrieux, C. Saulo, J. Ruiz, E. Gröller, and S. Bruckner, "Visual analysis of spatio-temporal data: Applications in weather forecasting," *Computer Graphics Forum*, vol. 34, no. 3, pp. 1-10, 2015, doi: 10.1111/cgf.12650.

[20] M. Lutz, *Learning Python: Powerful Object-Oriented Programming*, O'Reilly Media, Jun. 2013, ISBN: 978-1-449-35573-9.

[21] I. Idris, *Python Data Analysis Cookbook*, Packt Publishing, Jul. 2016, ISBN: 978-1-78528-228-7.

[22] K. J. Millman and M. Aivazis, "Python for scientists and engineers," *Comput. Sci. Eng*., vol. 13, no. 2, pp. 9–12, Apr. 2011, doi: 10.1109/MCSE.2011.36.

[23] W. McKinney, "Data structures for statistical computing in Python," *in SciPy Conf.*, Jun. 2010, vol. 445, no. 1, pp. 51–56.

[24] L. Perkovic, *Introduction to Computing Using Python: An Application Development Focus*, John Wiley & Sons, Apr. 2015, ISBN: 978-1-118-89094-3.