Department of Computer Engineering
College of Engineering
Polytechnic University of the Philippines Sta. Mesa

CMPE 40153

Exploratory Data Analysis of Urban Flooding in Mandaue City: Contributing Factors and Risks

Submitted by:

Daza, Ronnie M.
Esguerra, Adrian R.
Lobusta, Vanica O.
Merque, John Ric C.

BSCOE 3-2

Submitted to:

EDCEL B. ARTIFICIO

I. Introduction and Purpose of the Analysis

What if a single flood could bring an entire city to its knees, causing devastation that reshapes its future? In Mandaue City, Metro Cebu, this question is more than a hypothetical scenario—it's a recurring reality. The city grapples with severe flooding that not only disrupts daily life but also poses significant health risks, particularly in low-lying and coastal barangays like Looc, Opao, Subangdako, Tipolo, and Umapad. Recent intense rainfall has exacerbated these issues, leading to property damage, increased health hazards, and strained public services.

Mandaue under state of calamity after worst flood in 30 years
(Photo taken by BFP Mandaue City, September 2022)

But why does flooding persistently plague these areas, and what are its broader implications? Our study seeks to unravel these areas of inquisition by delving into a decade's worth of data. This includes precipitation data to understand how rainfall triggers flooding, population data to uncover urbanization levels, flood area percentage, and dengue cases relative to fatality. This is to assess factors that lead to flooding, map the current levels of flooding in each Branagay area, identify areas vulnerable to flood risks, and seek answers as to how flooding impacts public health focusing on the increased risk of diseases such as dengue and the adequacy of public health care facilities during emergencies.

Our objective is clear: to provide a comprehensive analysis of Mandaue City's flood contributors, risks, and impacts. By exploring historical data and evaluating current urban infrastructure, we aim to offer actionable insights that can guide effective flood management strategies. This study will highlight the vulnerabilities of affected communities and uncover insights for better flood mitigation strategies and urban planning in Mandaue City.

II. Data Dictionary

The Project Climate Change, Health, and Artificial Intelligence (Project CCHAIN) dataset is a validated, open-source dataset that tracks 20 years (2003-2022) of climate, environmental, socioeconomic, and health dimensions at the barangay level across 12 Philippine cities. For this specific analysis, the focus is on Mandaue City and the columns related to flooding data, using only data from 2012 to 2022.

Dataset: mandaue-brgy-level-urban-flooding
Description: Flood-related data columns for Mandaue city for each barangay.

Column Name	Data Type	Description
date	string	Standard date
adm4_pcode	string	Philippine Standard Geographic Code (PSGC) of baranggay
adm3_en	string	Name of municipality
adm4_en	string	Name of baranggay
brgy_total_area	double	Total area of barangay (sqm)
pct_area_flood_hazard_100yr_high	double	Area proportion of the barangay, ranging from 0-100 percent, that is potentially affected with a high flood hazard within a 100 year rain return period
pct_area_flood_hazard_100yr_low	double	Area proportion of the barangay, ranging from 0-100 percent, that is potentially affected by a low hazard flood within a 100 year rain return period
pct_area_flood_hazard_100yr_med	double	Area proportion of the barangay, ranging from 0-100 percent, that is potentially affected by a medium hazard flood within a 100 year rain return period
pop_count_mean	double	Mean population density of all the 100m grids within the barangay.
pop_count_total	double	Sum of all population count grids within the barangay
pop_density_mean	double	Mean population density of all the 100m grids within the barangay.
pr	double	Rainfall estimates (mm/day) from rain gauge and satellite observations. Extracted from CHIRPS. Barangay Tumalutab (Zamboanga) has no data.

Number of rows: 98,631

Dataset: mandaue-city-level-urban-flooding
Description: Public health- related data columns for Mandaue city

Column Name	Data Type	Description
adm3_en	string	Name of municipality
adm3_pcode	string	Philippine Standard Geographic Code (PSGC) of municipality
date	string	Standard date
case_total_dengue	integer	Number of recorded disease (dengue) cases. May be suspected, probable, or confirmed.
death_total_dengue	integer	Number of recorded deaths associated with a disease (dengue).

Number of rows: 520

III. Analysis Process

Outlined here is the general process of how the analysis was conducted:

1. Data Prepreprocessing

Relevant columns from the Project Climate Change, Health, and Artificial Intelligence (Project CCHAIN) dataset were extracted, focusing exclusively on records of Mandaue City. Subsequently, the columns specifically related to urban flooding were identified and selected. A thorough data cleaning process was then conducted, involving the removal of null values to ensure the integrity and reliability of the dataset.

2. Data Analysis

The data analysis for urban flooding in Mandaue City was structured into three key phases: **Contributing Factors**, **Flooding Levels**, and **Risk Factors**. Each phase utilized specific statistical and graphical techniques to identify patterns, correlations, and outliers for the data exploration of urban flooding contributing factors and risks.

2.1 Urban Flooding Contributing Factors. The first phase focused on identifying the root causes of flooding. *Precipitation Time Series Analysis* was employed to explore trends and seasonal patterns in rainfall, while *Correlation Analysis* examined the relationship between population density and barangay area. The goal was to assess the level of urbanization and its impact on flooding susceptibility, determining if more urbanized areas face a higher risk of flooding.

2.2 Urban Flooding Levels. In the second phase, barangays were categorized based on flood severity. *Barangay Flood Area Analysis* was conducted to assess low, medium, high, and average flood-prone areas, with visualizations primarily using area charts to represent the distribution of flood risks. This phase provided a detailed view of flood risk across barangays, highlighting areas experiencing different levels of flooding.

2.3 Urban Flooding Risk Factors. The final phase aimed at categorizing barangays by overall flood risk using techniques like *K-means Clustering* and *Graphical Outlier Analysis*. These methods identified vulnerable areas by grouping barangays with similar flood area percentages and population counts, while also detecting extreme outliers. Additionally, health-related risk factors, such as dengue cases, were correlated with flood-prone areas, offering a broader perspective on the impact of flooding on public health.

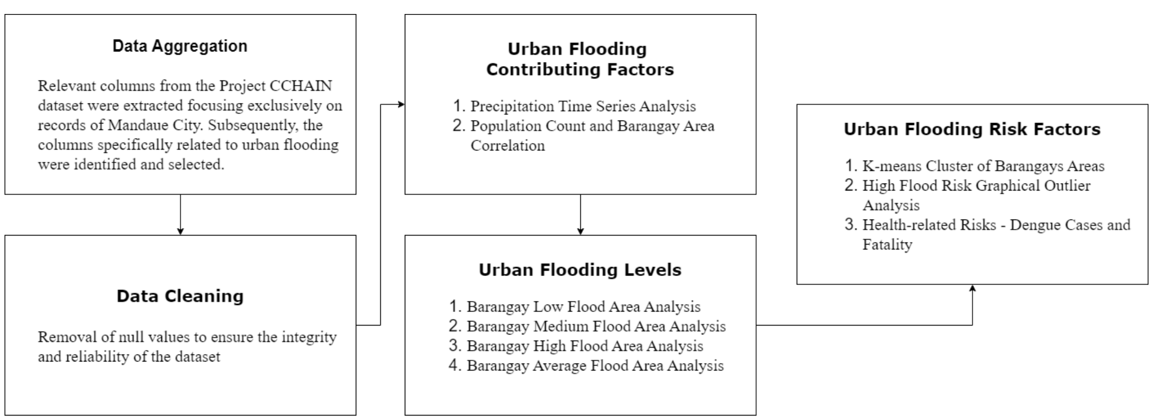

Figure 1. Analysis Process

IV. Analysis and Insights

This analysis provides key insights into the environmental factors contributing to flood risks in Mandaue City. The findings indicate that specific areas within the city are more vulnerable to severe flooding. Additionally, the analysis highlighted potential public health risks associated with these flood-prone regions. These results addressed critical questions regarding the determinants of urban flooding and identified new areas for further investigation, particularly in relation to infrastructure quality and urban planning.

1. Urban Flooding Contributing Factors

A. Precipitation Time Series Analysis

One of the primary questions explored in the EDA was identifying the key factors contributing to urban flooding in Mandaue City. According to the analysis, one of the main causes of flood events was determined to be heavy precipitation, especially in some months. An examination of the city's monthly precipitation levels is presented in Figure 2 through a time series analysis.

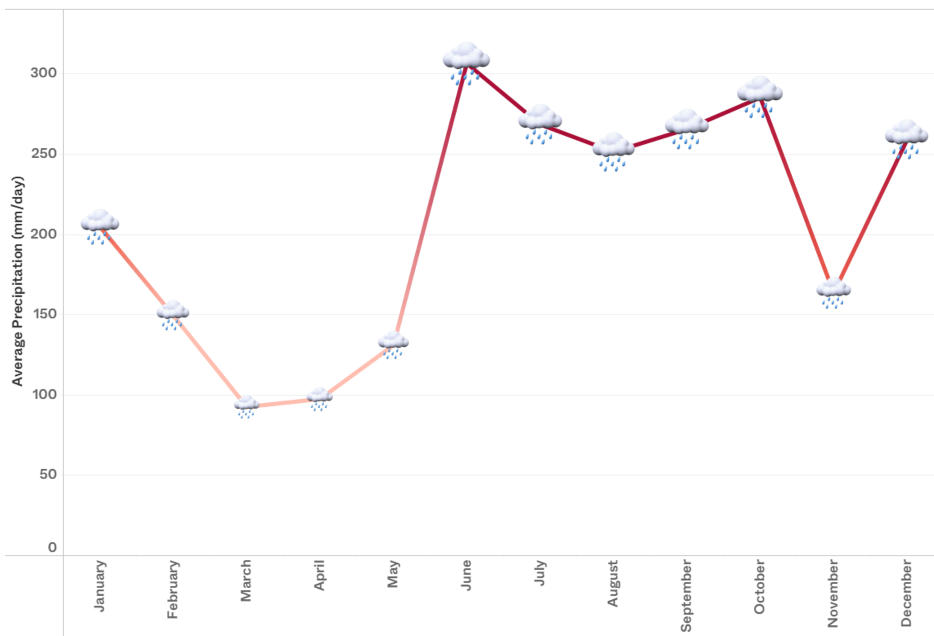

Figure 2. Precipitation Time Series Analysis

Rainfall intensity is classified according to the rate of precipitation, which is given as the amount of rainfall per unit of time. According to MANOBS (Manual of Surface Weather Observations), the following general categories are used to classify rainfall intensity, which is determined by the rainfall rate shown in Table 1.

Level	Rainfall Rate
Light	Less than 60 mm/day
Moderate	62.4 to 180 mm/day
Heavy	182.4 to 1200 mm/day
Violent	Greater than 1200 mm/day

Table 1. Rainfall Intensity Classification

The precipitation time series analysis for Mandaue City highlights a significant increase in rainfall during June, with average precipitation levels peaking at approximately 300 mm/day. This level of rainfall falls into the heavy rainfall category, which is defined as precipitation between 182.4 and 1200 mm/day. The data reveals a strong seasonal pattern, where June consistently experiences intense downpours compared to other months. This surge in heavy rainfall suggests an elevated risk of flooding during this period, especially in areas already prone to such events. The findings underline the importance of implementing targeted flood management strategies during June to mitigate the potential impacts of these heavy rains, particularly in the city's most vulnerable and densely populated areas.

B. *Population Count and Barangay Area Correlation*

Considering that Mandaue City is a rapidly urbanizing area, understanding the relationship between barangay area and population count is crucial for identifying potential flood risks. Figure 3 reveals a general positive correlation between barangay area and population count, suggesting that larger barangays tend to house more residents.

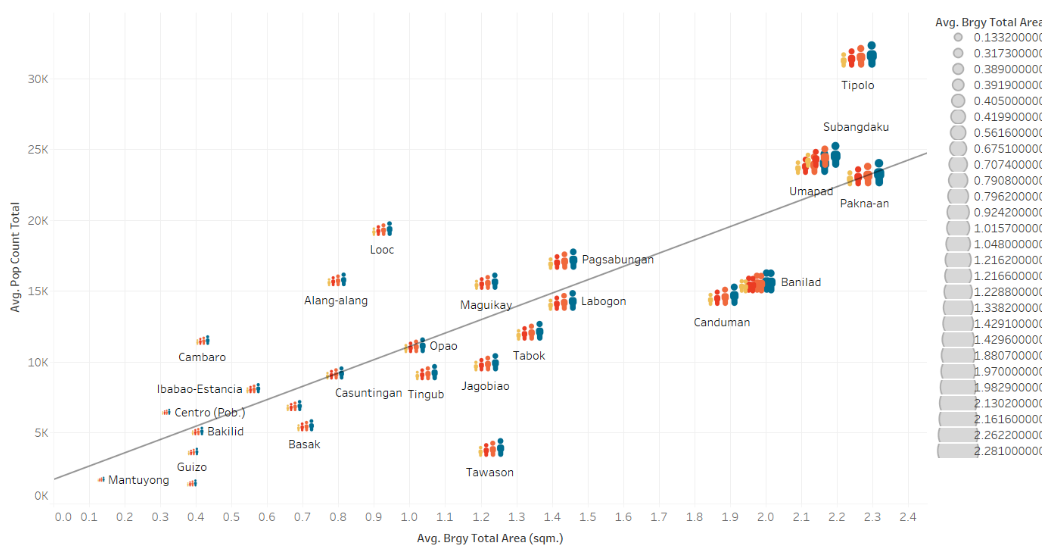

Figure 3. Population Count and Barangay Area Correlation

As shown, barangays like Tipolo and Looc, having relatively large areas, exhibit high population counts, positioning them above the trend line. This indicates that these barangays have higher-than-expected population densities compared to other areas of similar size. The elevated population density in these barangays reflects its corresponding level of urbanization, which increases their susceptibility to various challenges, including the management of resources and infrastructure. This high density makes these barangays more susceptible to flooding, particularly during periods of heavy rainfall, as indicated by the precipitation analyses. This interplay between population density and environmental factors is crucial for understanding and mitigating flood risks in urban areas like Mandaue City.

2. Urban Flooding Levels

Urban flooding is a significant concern in Mandaue City, with various barangays exhibiting different levels of flood susceptibility. This section presents a detailed analysis of the flood hazard levels—low, medium, and high—across the city's barangays, along with insights into the overall flood vulnerability.

A. Barangay Low Flood Area Analysis

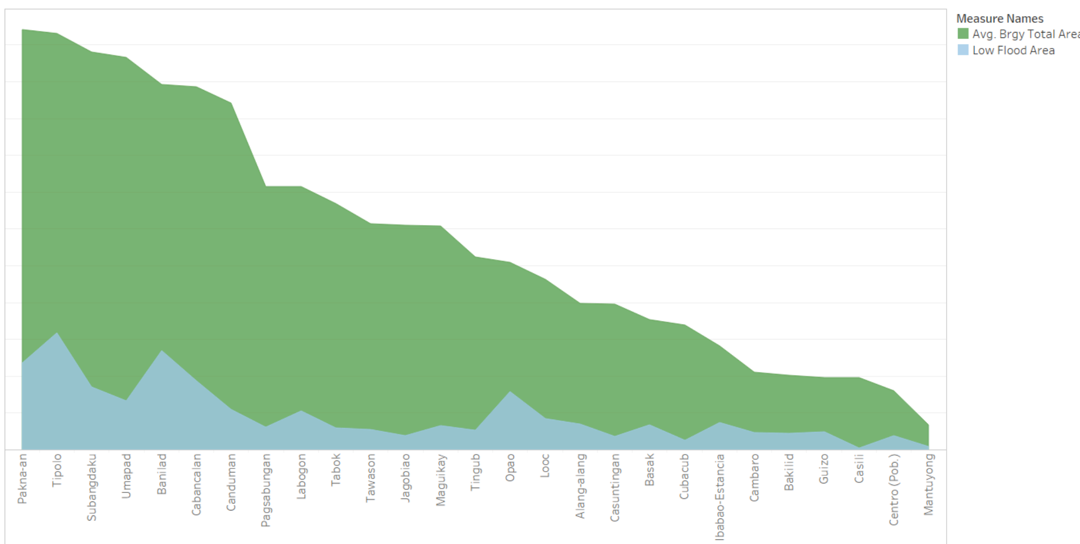

Figure 4. Barangay areas that are likely to be affected by low floods

As shown in Figure 4, Tipolo emerges as the barangay with the highest total low flood hazard area. In contrast, Opao is the most flooded barangay in this category, with an average of 31.14% of its area being prone to low flood conditions. This suggests that while Tipolo is more likely to experience frequent low-level flooding, Opao faces a more widespread impact in terms of area coverage. The large percentage of Opao's land affected by low flooding indicates that even minor rainfall events could lead to significant disruptions in this barangay.

B. Barangay Medium Flood Area Analysis

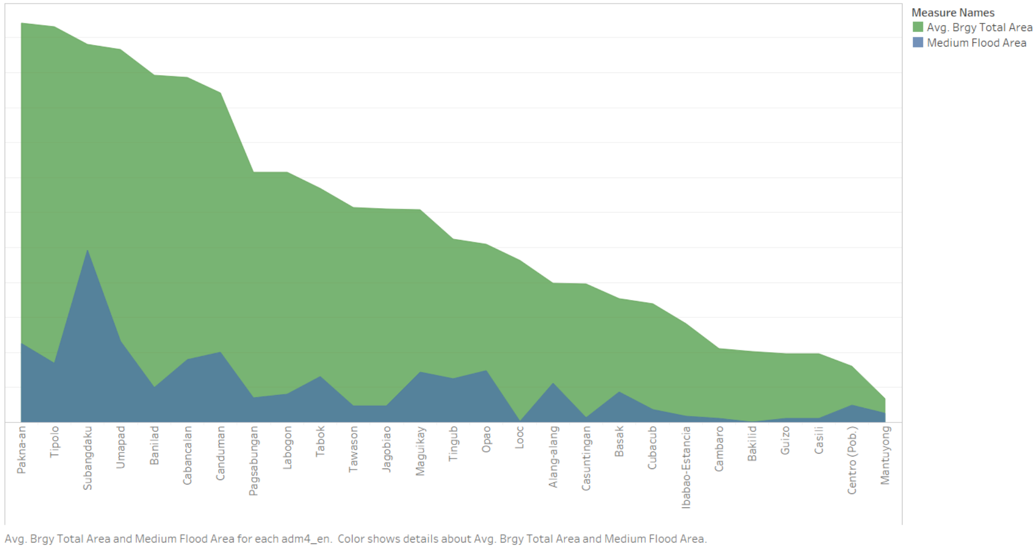

Figure 5. Barangay areas that are likely to be affected by medium floods

Figure 5 shows that Subangdaku stands out as both the most flooded barangay and the one with the highest medium flood hazard. On average, 45.57% of Subangdaku's area is susceptible to medium-level flooding, indicating a substantial portion of the barangay is at risk during moderate rainfall events. This finding highlights the urgent need for flood mitigation measures in Subangdaku, as nearly half of the area could be compromised during the rainy season, affecting both residents and infrastructure.

C. Barangay High Flood Area Analysis

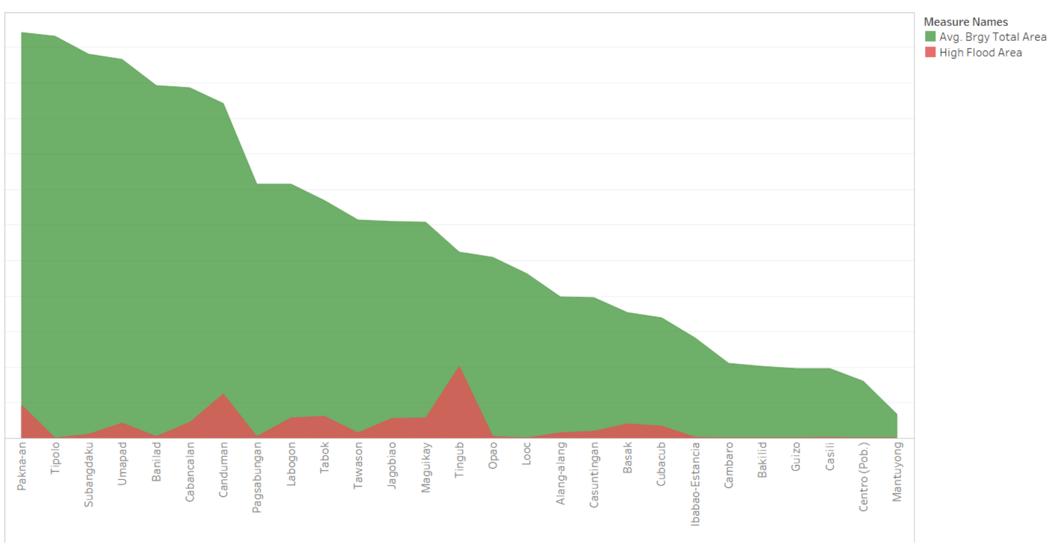

Figure 6. Barangay areas that are likely to be affected by high floods

As shown in Figure 6, for high flood hazards, Tingub is the most affected barangay, with 38.53% of its area on average likely to experience severe flooding. Tingub also holds the distinction of having the highest flood hazard level in the city. This suggests that Tingub is particularly vulnerable during extreme weather events, where intense rainfall could lead to significant flooding, posing serious risks to both property and life. The concentration of high flood hazards in Tingub calls for targeted interventions to enhance flood resilience.

D. Barangay Average Flood Area Analysis

Taking a holistic view, the overall area percentage of flood hazards for Mandaue City indicates that 5.68% of the city's area is prone to high flooding, 17.67% to medium flooding, and 16.44% to low flooding. To understand clearly, Figure 7 illustrates the three flood levels that Mandaue City is most likely to experience.

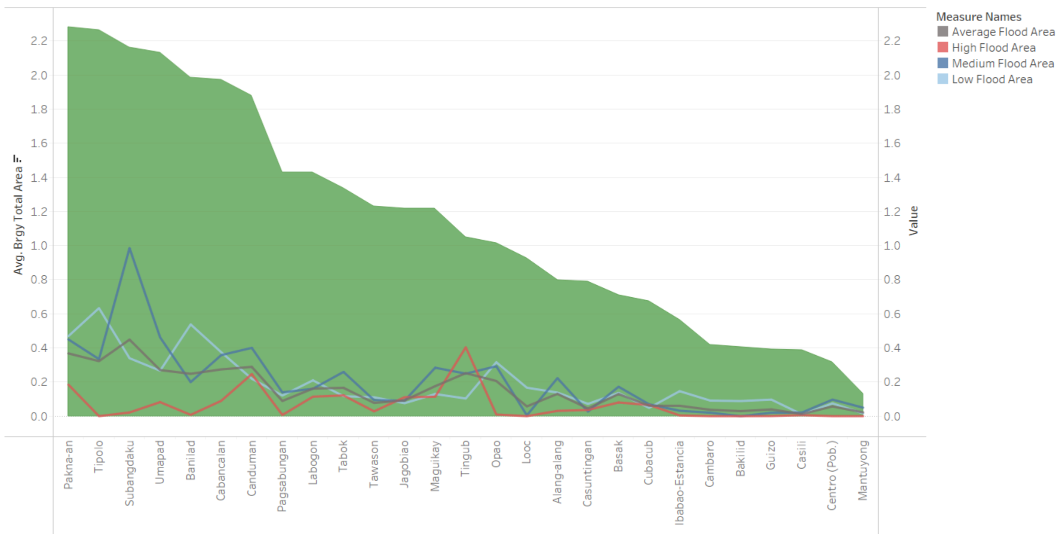

Figure 7. Barangay areas that are likely to be affected by different flood levels

Among the barangays, Tingub emerges as the most flooded in terms of the average flood hazard area, with 24.06% of its land likely to be affected by flooding across all levels. However, when considering the total flood hazard across all levels, Subangdaku presents the highest risk, with a significant portion of its area vulnerable to flooding. These findings underscore the critical importance of flood management strategies that are tailored to the specific flood profiles of each barangay, particularly focusing on areas like Tingub and Subangdaku that face the highest risks.

By understanding the varying levels of flood risk across Mandaue City's barangays, local authorities can better allocate resources and implement targeted measures to mitigate the impacts of flooding, thereby enhancing the city's overall resilience to natural disasters.

3. Urban Flooding Risk Factors

Urban flooding is a significant and growing concern, particularly in densely populated areas where the impact of heavy rainfall can be severe, like Mandaue City. As cities expand and climate change continues to influence weather patterns, the risks associated with urban flooding are becoming more pronounced. Understanding and identifying the factors that contribute to flood risk is essential for effective urban planning and disaster management. In this context, analyzing flood-prone areas and their associated population densities can provide crucial insights into which regions are most vulnerable to flood-related hazards.

A. K-means Cluster of Barangays

One method to systematically assess and categorize the levels of flood-related risks across barangays is through the use of K-means clustering, a data-driven approach that groups similar areas based on key variables. In this analysis, population count and flood area percentage are used. This analysis can help prioritize resource allocation and identify the barangay areas most vulnerable to flood-related risks based on flood area percentage and population count. The K-means process was depicted in Figure 8.

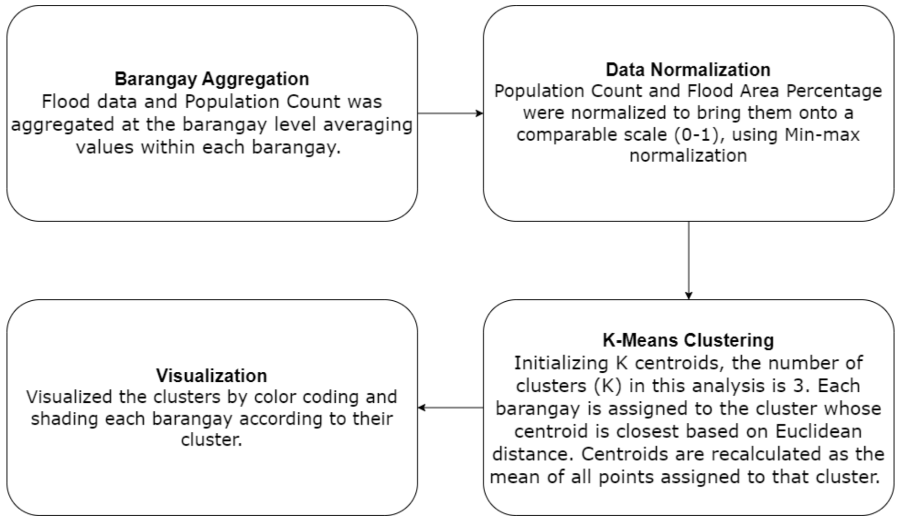

Figure 8. K-means Process

The K-means clustering process begins with the aggregation of flood data and population counts at the barangay level. These values are then normalized using min-max normalization to bring the population count and flood area percentage onto a comparable scale of 0 to 1. Following this, the K-means clustering algorithm is applied, initializing three clusters, where each barangay is assigned to a cluster based on the closest centroid using Euclidean distance. The centroids are recalculated as the mean of all points within the cluster. Finally, the results are visualized by color-coding and shading each barangay according to its cluster, highlighting the levels of vulnerability across the different areas. The K-means cluster graph is shown in Figure 9.

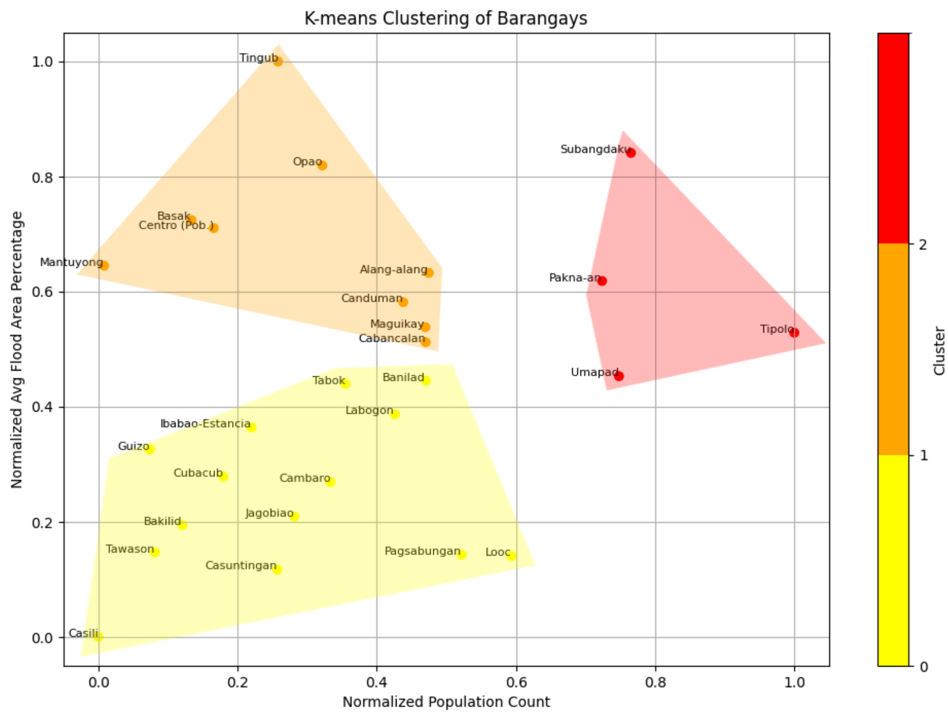

Figure 9. K-means Clustering of Barangays based on Flood Risk levels

X-axis (Normalized Population Count) represents the population count of each barangay, normalized to a scale of 0 to 1. This means the population count is scaled relative to the highest and lowest values in the dataset.
Y-axis (Normalized Average Flood Area Percentage) represents the percentage of the barangay's area that is prone to flooding, also normalized to a scale of 0 to 1.

The K-means clustering analysis of barangays in Mandaue City reveals three distinct groups based on their normalized population counts and flood area percentages, indicating varying levels of vulnerability to flood-related risks, as shown in Table 2.

Cluster	Description
Cluster 0	Barangays that generally have a lower flood area percentage and a lower to moderate population count, which makes them slightly vulnerable to flood-related risks.
Cluster 1	Barangays with a moderate flood area percentage and a moderate to high population count, which makes them moderately vulnerable to flood-related risks.
Cluster 2	Barangays that have either a high flood area percentage, a high population count, or both, which makes them most vulnerable to flood-related risks.

Table 2. Barangay Color-Coded Clusters Based on Flood Risk Levels

Barangays in Cluster 0 (Yellow), such as Casili, Casuntingan, and Bakilid, are characterized by low population counts and low flood area percentages, making them less vulnerable to flooding. These areas are relatively safer due to their lower population density and minimal flood-prone regions. Cluster 1 (Orange) includes barangays like Tingub, Opa, and Canduman, which have moderate population counts and flood area percentages. These barangays are moderately vulnerable, as their combination of moderate density and flood-prone areas requires some level of flood risk management. Finally, Cluster 2 (Red) comprises barangays such as Tipolo, Subangdaku, and Pakna-an, which exhibit high population counts and high flood area percentages. These areas are the most vulnerable to flooding, given their dense populations and significant flood-prone regions, necessitating urgent flood risk mitigation efforts. Overall, the clustering highlights the barangays that are at the highest risk, those that need moderate attention, and those that are relatively safer in terms of flood-related vulnerabilities.

B. High Flood Risk Vulnerability Areas

Understanding flood risk vulnerability is crucial for effective disaster management and urban planning. Flood risk is not solely determined by population density; other factors such as topography, infrastructure, and environmental conditions also play a significant role. This analysis aims to identify areas with high flood risk vulnerability across various barangays in Mandaue City, shedding light on how different factors contribute to flood exposure. To visualize this better, a scatter plot was provided, as shown in Figure 10.

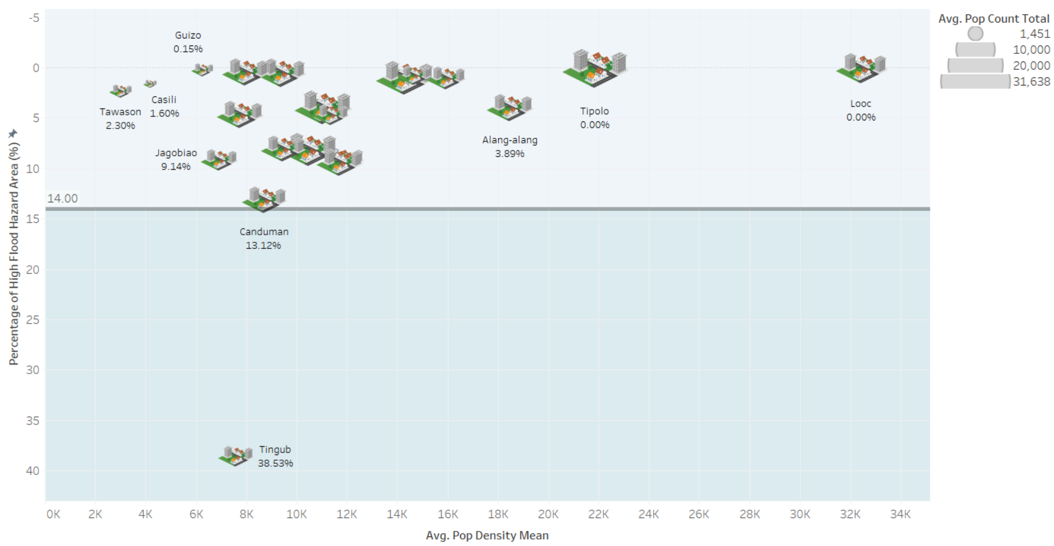

Figure 10. High flood-risk areas based on population density

X-axis (Avg. Pop Density Mean) represents the average population density across each barangay, with values increasing from left to right.
Y-axis (Percentage of High Flood Hazard Area) shows the percentage of the barangay's area that is likely to be affected by high flood levels.
A higher value indicates a greater portion of the area is at risk of significant flooding.

Figure 10 highlights the correlation between population density and flood risk vulnerability across various barangays in Mandaue City using a scatter plot. One of the key insights is the significant flood risk posed to Tingub, where 38.53% of its area is likely to be impacted by high flood levels. This barangay, despite having a lower average population density, demonstrates that flood risk is not solely dependent on how densely populated an area is. This could indicate that other factors, such as topography or infrastructure, play a significant role in flood vulnerability.

Barangay Tingub faces severe flooding issues due to several interconnected factors. Its proximity to a creek is a primary concern, as recent reports from SunStar indicate that an overflow from a nearby creek led to the evacuation of three families to the Tingub Elementary School gym. This close proximity places Tingub at high risk of flooding when the creek overflows [2]. Additionally, Tingub's low elevation exacerbates the situation. At just 22 meters above sea level, it is significantly lower than the nearby Barangay Pagsabungan, which sits at 37 meters. This elevation difference causes water from higher areas to flow down and accumulate in Tingub, making it more susceptible to flooding [3].

The name "Tingub" itself, which refers to the accumulation of water, reflects the area's tendency to flood. The creek's role in gathering water from surrounding areas such as Canduman, Pagsabungan, and Tabok further compounds the flooding risk [4]. Despite these issues, Tingub has been identified as a priority for potential retention basins, although this is not currently a top priority. This consideration underscores the need for flood mitigation solutions [5].

Moreover, substantial funds have been allocated for flood control in Tingub [6], including investments in water pump stations and other flood mitigation projects [6], as reported by the Daily Tribune. These financial commitments highlight the ongoing flood challenges faced by the area. Collectively, these factors—proximity to the creek, low elevation, water accumulation from surrounding regions, plans for retention basins, and significant financial investments—contribute to Tingub's severe flooding issues, with 38.53% of the area affected by high flood levels.

On the other hand Canduman, with 13.12% of its area exposed to high flood risk, also presents a moderate level of vulnerability, just below the identified threshold of 14%. The visualization underscores that while population density is a crucial factor in disaster planning, it is not the only determinant of flood risk. This is evident in the barangays with varying densities yet facing different levels of flood exposure. The analysis suggests that flood risk management strategies should be tailored not only based on population metrics but also on the specific environmental and infrastructural characteristics of each barangay the proximity to a creek, low elevation, accumulation of water from nearby areas, consideration for retention basins, and significant financial investments for flood control all contribute to the area's severe flooding problems.

C. Health-related Risks

Examining Dengue Case Rate and Fatality Rate

Dengue fever, a mosquito-borne viral infection, poses a significant public health risk in tropical and subtropical regions, including the Philippines. Mandaue City, like many urban areas, faces challenges in managing and controlling dengue outbreaks. The visualization shown in Figure 11 presents a detailed analysis of the number of dengue cases and related fatalities in Mandaue City from 2012 to 2023.

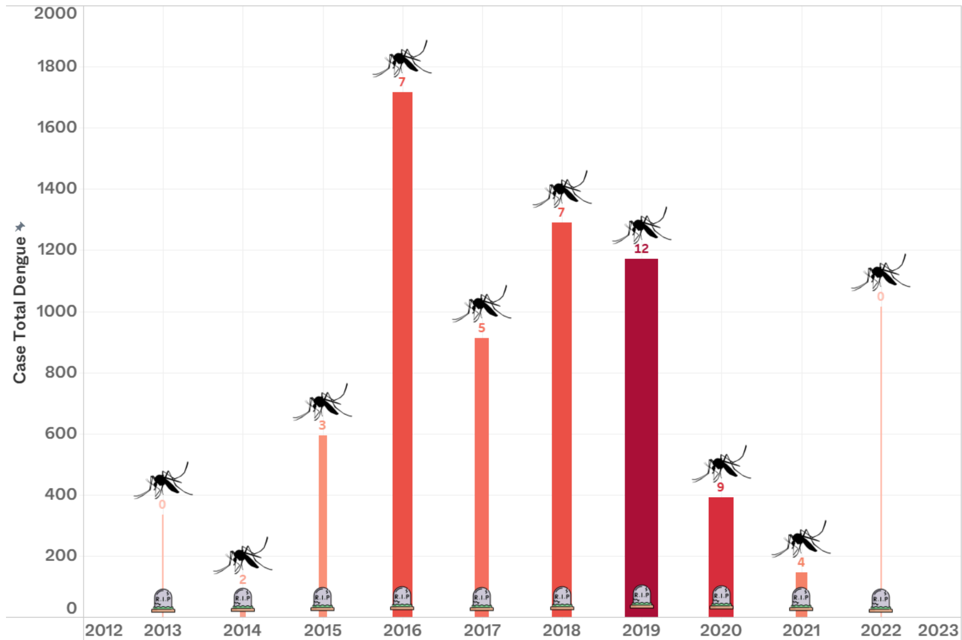

Figure 11. The number of deaths per dengue case in Mandaue City

Figure 11 illustrates the total number of dengue cases (represented by the height of the bars) and the corresponding number of deaths (denoted by the gravestones) for each year from 2012 to 2023. The color intensity of the bars indicates the severity of fatalities, with darker shades corresponding to years with higher death rates. Additionally, mosquito icons are used to represent the presence of the dengue virus, emphasizing the correlation between the disease's spread and its impact on public health.

The data indicates a fluctuating trend in both dengue cases and fatalities over the years. Notably, the year 2019 stands out with the highest number of dengue cases, reaching nearly 1,800, and also records the highest number of deaths at 12. This suggests a significant outbreak during that period, highlighting the need for improved public health measures and vector control strategies.

In contrast, the years 2013 and 2022 report zero fatalities despite having recorded dengue cases, which could be attributed to effective medical interventions or better public health infrastructure in those years. However, the relatively high death toll in 2020 (9 deaths), despite a reduction in total cases compared to previous years, suggests that the virulence of the dengue strain or other compounding factors, such as overwhelmed healthcare systems, may have played a role.

On average, 5.09% of people in Mandaue suffered from dengue, and 1.31% of them died. Generally, less than 1% of people die from dengue fever, which is usually a self-limiting illness [9]. In comparison, the mortality rate of 1.31% in Mandaue City is higher than the typical mortality rate for general dengue fever (<1%). This suggests that the impact of dengue in Mandaue City is relatively severe, possibly due to factors such as the prevalence of dengue hemorrhagic fever or challenges in timely and effective treatment. Additionally, the higher infection rate (5.09%) indicates that dengue is a significant public health concern in Mandaue City, with a mortality rate that, while still concerning, could indicate room for improved treatment and prevention efforts.

Flooding significantly elevates health risks, particularly during the rainy season [7] . The World Health Organization (WHO) underscores that floods contaminate water sources, which can lead to waterborne diseases such as typhoid fever and cholera. Additionally, the stagnant water created by flooding fosters mosquito breeding grounds, thereby increasing the risk of vector-borne diseases like dengue (PCHRD, 2024) [7]. This correlation is further supported by a review published in the National Center for Biotechnology Information (NCBI), which highlights that flooding provides ideal conditions for mosquito proliferation, exacerbating the spread of dengue and other mosquito-borne illnesses (NCBI, 2024) [8].

Further research, as detailed in various studies on flooding impacts, confirms that significant flooding events are often followed by a rise in vector-borne diseases, including dengue. However, the results are mixed and highly context-dependent. Factors such as the effectiveness of disaster response, the local landscape, and the specific mosquito species involved influence the health outcomes. The complexity and variability in these impacts underscore the challenges in drawing consistent conclusions about the relationship between flooding and disease incidence across different regions and circumstances [8].

Contributions of the Analysis to Mandaue City

This exploratory data analysis delves into the complex factors that contribute to urban flooding in Mandaue City, focusing on geographic and climatic elements. The EDA identifies key risk areas, correlates flood frequency with rainfall patterns, and examines the impact of urbanization on flooding events. This EDA suggests that certain districts within Mandaue City are more prone to flooding due to their low elevation and high population density. Additionally, the report also emphasizes the role of seasonal rainfall and typhoon events in exacerbating flooding risks.

The analyses and insights from this report contribute significantly to the body of knowledge in urban flood management, particularly for stakeholders directly involved in disaster risk reduction and urban planning. By identifying the critical factors that lead to flooding in Mandaue City, this study offers data-driven insights that can inform urban planning and infrastructure development. The primary beneficiaries of this EDA include:

City Planners and Local Government Officials - The findings provide crucial data specifically tailored to Mandaue City, enabling the optimization of urban development strategies. By identifying the most flood-prone areas, planners can prioritize infrastructure improvements, such as enhancing drainage systems and implementing flood mitigation measures in the city's high-risk zones.

Disaster Risk Reduction Agencies - The insights into the relationship between rainfall patterns, typhoon events, and flood frequency in Mandaue City can help disaster management agencies develop more effective early warning systems and preparedness plans. These agencies can use the data to allocate resources more efficiently and design targeted interventions that minimize the impact of flooding on vulnerable communities within the city.

Environmental Scientists and Researchers - The study's detailed examination of how urbanization impacts flooding risks in Mandaue City contributes to the broader academic discourse on urban resilience. Researchers can build on this EDA to explore specific environmental factors contributing to flooding in the city and develop more advanced models for predicting flood events in similar urban settings.

Residents and Community Leaders - The local communities in Mandaue City will directly benefit from the report's recommendations. By highlighting high-risk areas, the report empowers community leaders and residents with information on the necessary precautions and actions to take before and during flood events, potentially reducing loss of life and property.

Policy Makers - The EDA offers evidence-based insights that can support the development of more robust policies and regulations aimed at sustainable urban development in Mandaue City. Policymakers can leverage this data to enact legislation that addresses the root causes of flooding and promotes the development of resilient infrastructure tailored to the city's unique challenges.

By addressing the specific needs and interests of these stakeholders, this report not only advances academic knowledge but also offers practical solutions that can lead to more resilient urban environments, ultimately benefiting the people of Mandaue City and beyond.

V. Conclusions and Recommendations

Mandaue City experiences varying levels of urban flooding driven by heavy rainfall, high population densities linked to urbanization, and inadequate urban planning, with certain barangays being more at risk than others. These factors contribute to conditions that facilitate the spread of dengue, intensifying public health concerns in the most vulnerable areas. The city frequently experiences heavy rainfall especially in the months of June to October, heavily impacting widespread flooding. Moreover, Its high population density further increases its susceptibility to floods, with barangays such as Tingub, Subangdaku, and Opao being particularly vulnerable due to their high flood risks and moderate population density. Furthermore, this results in health risks, particularly Dengue to be prevalent in the city, affecting 5.09% of the population and resulting in a fatality rate of 1.31%.

Key Findings

Seasonal Flooding Linked to Heavy Rainfall:

The time series analysis of precipitation data indicates that Mandaue City experiences significantly higher rainfall in June, peaking at approximately 300 mm/day. This heavy rainfall, classified under the "heavy" category, correlates strongly with increased flood events, especially in flood-prone barangays.

High Urbanization Contributes to Flooding:

A positive correlation between barangay area and population count was identified, indicating that larger barangays like Tipolo and Looc, which also have high population densities, are more susceptible to flooding. High urbanization, reflected in dense populations, exacerbates the flooding risk.

Varying Flood Risks Across Barangays:

Flood risk levels vary significantly across barangays:

- **Low Flood Risk:** Tipolo faces frequent but minor flood events.
- **Medium Flood Risk:** Subangdaku is highly susceptible, with almost half its area at risk of medium-level flooding.

- **High Flood Risk:** Tingub stands out as the most flood-prone barangay, with nearly 40% of its area at risk during severe floods.

Cluster Analysis Reveals Vulnerability:

Using K-means clustering, barangays were categorized into three levels for flood-related risks:

  - **Low Risk (Cluster 0)**: Areas like Casili have low flood area percentages and population densities.
  - **Moderate Risk (Cluster 1)**: Barangays like Tingub have moderate flood areas and population counts.
  - **High Risk (Cluster 2)**: Barangays like Tipolo and Subangdaku are highly vulnerable due to their dense populations and significant flood areas and may need more urgent and effective interventions.

Recommendations

There are ways to help solve the ever-evolving problem of flooding in Mandaue City. As this is a public health concern, especially rising dengue cases from floods in high-risk areas. From this analysis, these recommendations are suggested for the mitigation of Mandaue City Urban Flooding.

Flood Mitigation Measures - Allot finances to create better drainage systems to help mitigate flood problems especially in high risk areas. Make way for research and development on how to properly solve the serious problems of floods that affect the health and economy of Mandaue City.

Urban Design and Development - Enforce zoning regulations that prohibits development of houses and establishments in areas with high risk of floods. If not possible, suggest houses and establishments to build elevated structures to prevent flood in entering their premises. Integrate flood risk assessments into urban planning, ensuring new developments in high-risk areas include adequate flood defenses and drainage solutions.

Population Density Management - Inform citizens and suggest that they live in new urban centers or satellite towns to make the density of population as low as possible to prevent overcrowding especially in high risk areas of floods. In addition, teach citizens about the risks of the floods and how to solve the flooding problems whenever they encounter them.

Targeted Flood Management Strategies - Focus on the most vulnerable barangays, particularly Tingub and Subangdaku, by enhancing drainage systems, creating flood barriers, and improving early warning systems.

Long-term Climate Adaptation Measures - Develop and implement climate adaptation strategies that address the root causes of increased flooding, such as reforestation, restoring natural floodplains, and reducing impervious surfaces in urban areas.

VI. References

[1] *J. Barani, "Rain rate intensity classification — BARANI," BARANI, Oct. 27, 2020.*

[2] *Sunnexdesk and Sunnexdesk, "'Habagat' rains cause flooding, landslides," SunStar Publishing Inc., Sep. 16, 2020.*

[3] *Admin, "MGB-7 warns residents in hazardous areas," Region 7, Jul. 03, 2017.*

[4] *Mandaue City Web Portal 'Tingub'*

[5] *Sunnexdesk and Sunnexdesk, "Mandaue needs P3B for drainage, flood control," SunStar Publishing Inc., Dec. 15, 2020.*

[6] *Rico Osmeña and Rico Osmeña, "Cortes, Ouano-Dizon united to solve flooding problems in Mandaue City," Daily Tribune, Jan. 24, 2024.*

[7] *Philippine Council for Health Research and Development, "The Risks of Flooding: Protect yourself from flood water diseases - Philippine Council for Health Research and Development," Philippine Council for Health Research and Development, Mar. 09, 2022.*

[8] *J. E. Coalson et al., "The Complex Epidemiological Relationship between Flooding Events and Human Outbreaks of Mosquito-Borne Diseases: A Scoping Review," Environmental Health Perspectives, vol. 129, no. 9, Sep. 2021*

[9] *D. S. S. M. Fidsa MSc, DTM&Amp;H, "Dengue: Practice essentials, background, pathophysiology." https://emedicine.medscape.com/article/215840-overview*

VII. Annex

Scripts and Codes

[1]:
```python
from pyspark.sql import SparkSession
# Initialize a SparkSession
spark = SparkSession.builder \
    .appName("mandaue_exploratory_data_analysis") \
    .getOrCreate()
```

[2]:
```python
# City merged Dataframe
city_merged_10_years_file_path = "city_merged_10 years.csv"
city_merged_10_years_df = spark.read.csv(city_merged_10_years_file_path, header=True, inferSchema=True)
# Display the schema of the DataFrame
city_merged_10_years_df.printSchema()
city_merged_10_years_df.describe("case_total_dengue", "death_total_dengue", "pop_count_mean", "hospital_count", "doctors_count").show()
```

```
root
 |-- adm3_en: string (nullable = true)
 |-- adm3_pcode: string (nullable = true)
 |-- date: string (nullable = true)
 |-- year: integer (nullable = true)
 |-- week: integer (nullable = true)
 |-- brgy_is_coastal: double (nullable = true)
 |-- co: double (nullable = true)
 |-- doctors_count: double (nullable = true)
 |-- hospital_count: double (nullable = true)
 |-- osm_canal_nearest: double (nullable = true)
 |-- osm_drain_nearest: double (nullable = true)
 |-- osm_water_nearest: double (nullable = true)
 |-- pct_area_flood_hazard_100yr_high: double (nullable = true)
 |-- pct_area_flood_hazard_100yr_low: double (nullable = true)
 |-- pct_area_flood_hazard_100yr_med: double (nullable = true)
 |-- pop_count_mean: double (nullable = true)
 |-- pop_count_total: double (nullable = true)
 |-- pop_density_mean: double (nullable = true)
 |-- pr: double (nullable = true)
 |-- case_total_dengue: integer (nullable = true)
 |-- death_total_dengue: integer (nullable = true)

+-------+------------------+--------------------+-----------------+------------------+------------------+
|summary| case_total_dengue|  death_total_dengue|   pop_count_mean|    hospital_count|     doctors_count|
+-------+------------------+--------------------+-----------------+------------------+------------------+
|  count|               520|                 521|              418|               470|               470|
|   mean|14.732692307692307|0.094049904030710117|2708.0798697344353| 7.998867149406383|3.3333333333340427|
| stddev| 17.31760685511655|0.32341790868571996|  91.3448893848293|3.9126787258387603|0.9420548947812599|
|    min|                 0|                   0|      2533.517159|               3.0|               3.0|
|    max|               129|                   2|      2854.069733|              13.0|               6.0|
+-------+------------------+--------------------+-----------------+------------------+------------------+
```

[4]:

```python
from pyspark.sql import functions as F

# Barangay merged Dataframe
barangay_merged_10_years_file_path = "barangay_merged_10 years.csv"
barangay_merged_10_years_df = spark.read.csv(barangay_merged_10_years_file_path, header=True, inferSchema=True)
# Display the schema of the DataFrame
barangay_merged_10_years_df.printSchema()

# Dictionary to map original column names to shorter names
column_mapping = {
    "brgy_total_area": "total_area",
    "pct_area_flood_hazard_100yr_high": "flood_hazard_high",
    "pct_area_flood_hazard_100yr_med": "flood_hazard_med",
    "pct_area_flood_hazard_100yr_low": "flood_hazard_low",
}

# Apply the renaming and select only the desired columns
brgy_health_hazard_df = barangay_merged_10_years_df.select(
    [F.col(old).alias(new) for old, new in column_mapping.items()]
)

# Display summary statistics for the selected and renamed columns
brgy_health_hazard_df.describe().show()
```

```
root
 |-- date: string (nullable = true)
 |-- adm4_pcode: string (nullable = true)
 |-- adm3_en: string (nullable = true)
 |-- adm4_en: string (nullable = true)
 |-- brgy_total_area: double (nullable = true)
 |-- brgy_distance_to_coast: double (nullable = true)
 |-- brgy_is_coastal: boolean (nullable = true)
 |-- heat_index: double (nullable = true)
 |-- pr: double (nullable = true)
 |-- osm_canal_nearest: double (nullable = true)
 |-- osm_drain_nearest: double (nullable = true)
 |-- osm_water_nearest: double (nullable = true)
 |-- pct_area_flood_hazard_100yr_high: double (nullable = true)
 |-- pct_area_flood_hazard_100yr_low: double (nullable = true)
 |-- pct_area_flood_hazard_100yr_med: double (nullable = true)
 |-- pop_count_mean: double (nullable = true)
 |-- pop_count_total: double (nullable = true)
 |-- pop_density_mean: double (nullable = true)

+-------+-----------------+-----------------+-----------------+------------------+
|summary|       total_area|flood_hazard_high| flood_hazard_med|  flood_hazard_low|
+-------+-----------------+-----------------+-----------------+------------------+
|  count|            98631|            98631|            98631|             98631|
|   mean|1.151940740740715|5.223140740740957| 16.24007037036946|16.210685185186883|
| stddev|0.657641728958141|7.719020322515324|11.416315009275278|7.4865972013974655|
|    min|           0.1332|              0.0|           0.0247|               2.7|
|    max|            2.281|          38.5303|          45.5713|           31.1449|
+-------+-----------------+-----------------+-----------------+------------------+
```

[6]:
```
# Select the desired columns into a new DataFrame
brgy_climate_df = barangay_merged_10_years_df.select(
    F.col("pr"),
)

# Show summary statistics for the selected columns
brgy_climate_df.describe().show()
```

```
+-------+-----------------+
|summary|               pr|
+-------+-----------------+
|  count|            98631|
|   mean|4.741239265545317 5|
| stddev| 8.145286474348818|
|    min|              0.0|
|    max|            93.63|
+-------+-----------------+
```

[7]:
```
from pyspark.sql.functions import col, sum, desc, asc
# Assuming barangay_merged_10_years_df is your dataframe
# Find the most flooded barangay for each hazard level
most_flooded_high = barangay_merged_10_years_df.orderBy(desc("pct_area_flood_hazard_100yr_high")).select("adm4_en", "pct_area_flood_hazard_100yr_high")
most_flooded_med = barangay_merged_10_years_df.orderBy(desc("pct_area_flood_hazard_100yr_med")).select("adm4_en", "pct_area_flood_hazard_100yr_med").fi
most_flooded_low = barangay_merged_10_years_df.orderBy(desc("pct_area_flood_hazard_100yr_low")).select("adm4_en", "pct_area_flood_hazard_100yr_low").fi

# Calculate the average flood hazard percentage for each barangay
average_flood_hazard_df = barangay_merged_10_years_df.withColumn(
    "average_flood_hazard_pct",
    (col("pct_area_flood_hazard_100yr_high") + col("pct_area_flood_hazard_100yr_med") + col("pct_area_flood_hazard_100yr_low")) / 3
)

# Get the barangay with the highest average flood hazard percentage
most_flooded_overall = average_flood_hazard_df.orderBy(desc("average_flood_hazard_pct")).select("adm4_en", "average_flood_hazard_pct").first()

# Show the results
print(f"Most flooded barangay (High hazard): {most_flooded_high['adm4_en']} with {most_flooded_high['pct_area_flood_hazard_100yr_high']}%")
print(f"Most flooded barangay (Medium hazard): {most_flooded_med['adm4_en']} with {most_flooded_med['pct_area_flood_hazard_100yr_med']}%")
print(f"Most flooded barangay (Low hazard): {most_flooded_low['adm4_en']} with {most_flooded_low['pct_area_flood_hazard_100yr_low']}%")
print(f"Most flooded barangay overall: {most_flooded_overall['adm4_en']} with {most_flooded_overall['average_flood_hazard_pct']}% average flood hazard
```

```
Most flooded barangay (High hazard): Tingub with 38.5303%
Most flooded barangay (Medium hazard): Subangdaku with 45.5713%
Most flooded barangay (Low hazard): Opao with 31.1449%
Most flooded barangay overall: Tingub with 24.063% average flood hazard area
```

[8]:
```
# Assuming barangay_merged_10_years_df is your dataframe and contains only Mandaue City data

# Calculate total area
total_area = barangay_merged_10_years_df.agg(sum("brgy_total_area")).collect()[0][0]

# Calculate weighted flood hazard for high, medium, and low flood hazards
weighted_flood_hazard_high = barangay_merged_10_years_df.withColumn(
    "weighted_flood_hazard_high",
    (col("brgy_total_area") * col("pct_area_flood_hazard_100yr_high")) / 100
).agg(sum("weighted_flood_hazard_high")).collect()[0][0]

weighted_flood_hazard_med = barangay_merged_10_years_df.withColumn(
    "weighted_flood_hazard_med",
    (col("brgy_total_area") * col("pct_area_flood_hazard_100yr_med")) / 100
).agg(sum("weighted_flood_hazard_med")).collect()[0][0]

weighted_flood_hazard_low = barangay_merged_10_years_df.withColumn(
    "weighted_flood_hazard_low",
    (col("brgy_total_area") * col("pct_area_flood_hazard_100yr_low")) / 100
).agg(sum("weighted_flood_hazard_low")).collect()[0][0]

# Calculate the overall percentage of each flood hazard type
overall_flood_hazard_percentage_high = (weighted_flood_hazard_high / total_area) * 100
overall_flood_hazard_percentage_med = (weighted_flood_hazard_med / total_area) * 100
overall_flood_hazard_percentage_low = (weighted_flood_hazard_low / total_area) * 100

# Print the results
print(f"Overall percentage of high flood hazard for Mandaue City: {overall_flood_hazard_percentage_high:.2f}%")
print(f"Overall percentage of medium flood hazard for Mandaue City: {overall_flood_hazard_percentage_med:.2f}%")
print(f"Overall percentage of low flood hazard for Mandaue City: {overall_flood_hazard_percentage_low:.2f}%")
```

```
Overall percentage of high flood hazard for Mandaue City: 5.68%
Overall percentage of medium flood hazard for Mandaue City: 17.67%
Overall percentage of low flood hazard for Mandaue City: 16.44%
```

[9]:

```python
# Create a new DataFrame with additional calculated columns

from pyspark.sql.functions import avg

city_merged_analysis_df = city_merged_10_years_df.withColumn(
    "case_rate_per_100k",
    (col("case_total_dengue") / col("pop_count_total")) * 100000
).withColumn(
    "death_rate_per_100k",
    (col("death_total_dengue") / col("pop_count_total")) * 100000
).withColumn(
    "fatality_rate",
    (col("death_total_dengue") / col("case_total_dengue")) * 100
).withColumn(
    "doctors_per_100k",
    (col("doctors_count") / col("pop_count_total")) * 100000
).withColumn(
    "hospitals_per_100k",
    (col("hospital_count") / col("pop_count_total")) * 100000
)

# Remove rows with null values in the relevant columns
city_merged_analysis_df_clean = city_merged_analysis_df.dropna(subset=[
    "case_rate_per_100k",
    "death_rate_per_100k",
    "fatality_rate",
    "doctors_per_100k",
    "hospitals_per_100k"
])

# Compute average values for the entire city
average_values = city_merged_analysis_df_clean.agg(
    avg("case_rate_per_100k").alias("avg_case_rate_per_100k"),
    avg("death_rate_per_100k").alias("avg_death_rate_per_100k"),
    avg("fatality_rate").alias("avg_fatality_rate"),
    avg("doctors_per_100k").alias("avg_doctors_per_100k"),
    avg("hospitals_per_100k").alias("avg_hospitals_per_100k")
).collect()[0]

# Print the aggregated averages
print(f"Average case rate per 100k: {average_values['avg_case_rate_per_100k']:.2f}")
print(f"Average death rate per 100k: {average_values['avg_death_rate_per_100k']:.2f}")
print(f"Average fatality rate: {average_values['avg_fatality_rate']:.2f}")
print(f"Average doctors per 100k: {average_values['avg_doctors_per_100k']:.2f}")
print(f"Average hospitals per 100k: {average_values['avg_hospitals_per_100k']:.2f}")
```

```
Average case rate per 100k: 5.54
Average death rate per 100k: 0.04
Average fatality rate: 1.51
Average doctors per 100k: 0.88
Average hospitals per 100k: 2.03
```

[10]:

```python
from pyspark.sql.functions import col, expr, min as spark_min, max as spark_max
from pyspark.ml.feature import VectorAssembler
from pyspark.sql.functions import avg, sum as spark_sum

# Step 1: Select relevant columns and compute average flood hazard
clustering_df = barangay_merged_10_years_df.select(
    'adm4_en',
    'pop_count_total',
    'pct_area_flood_hazard_100yr_high',
    'pct_area_flood_hazard_100yr_med',
    'pct_area_flood_hazard_100yr_low'
).withColumn(
    'avg_pct_area_flood_hazard',
    expr("(pct_area_flood_hazard_100yr_high + pct_area_flood_hazard_100yr_med + pct_area_flood_hazard_100yr_low) / 3")
)

# Step 2: Remove rows with null values
clustering_df = clustering_df.dropna(subset=['pop_count_total', 'avg_pct_area_flood_hazard'])



# Show the result to verify
clustering_df.show()

# Step 1: Select relevant columns and compute average flood hazard
clustering_df = barangay_merged_10_years_df.select(
    'adm4_en',
    'pop_count_total',
    'pct_area_flood_hazard_100yr_high',
    'pct_area_flood_hazard_100yr_med',
    'pct_area_flood_hazard_100yr_low'
).withColumn(
    'avg_pct_area_flood_hazard',
    expr("(pct_area_flood_hazard_100yr_high + pct_area_flood_hazard_100yr_med + pct_area_flood_hazard_100yr_low) / 3")
)

# Step 3: Aggregate table based on 'adm4_en'
clustering_df = clustering_df.groupBy('adm4_en').agg(
    avg('pop_count_total').alias('pop_count_total'),
    avg('avg_pct_area_flood_hazard').alias('avg_pct_area_flood_hazard')
)

# Step 2: Remove rows with null values
clustering_df = clustering_df.dropna(subset=['pop_count_total', 'avg_pct_area_flood_hazard'])

# Step 3: Normalize the columns using Min-Max scaling
min_pop = clustering_df.agg(spark_min('pop_count_total')).collect()[0][0]
max_pop = clustering_df.agg(spark_max('pop_count_total')).collect()[0][0]

clustering_df = clustering_df.withColumn(
    'pop_count_total_Normalized',
    (col('pop_count_total') - min_pop) / (max_pop - min_pop)
)

min_flood = clustering_df.agg(spark_min('avg_pct_area_flood_hazard')).collect()[0][0]
max_flood = clustering_df.agg(spark_max('avg_pct_area_flood_hazard')).collect()[0][0]

clustering_df = clustering_df.withColumn(
    'avg_pct_area_flood_hazard_Normalized',
    (col('avg_pct_area_flood_hazard') - min_flood) / (max_flood - min_flood)
)
```

```python
# Step 4: Assemble features for clustering
assembler = VectorAssembler(
    inputCols=['pop_count_total_Normalized', 'avg_pct_area_flood_hazard_Normalized'],
    outputCol='features'
)
dataset = assembler.transform(clustering_df)

# Step 5: Remove rows with null values in the feature columns
dataset = dataset.dropna(subset=['pop_count_total_Normalized', 'avg_pct_area_flood_hazard_Normalized'])

# Display descriptive statistics to check the prepared dataset
dataset.describe().show()
```

```
+---------------+---------------+--------------------------------+-------------------------------+-------------------------------+---------------------
-----+
|        adm4_en|pop_count_total|pct_area_flood_hazard_100yr_high|pct_area_flood_hazard_100yr_med|pct_area_flood_hazard_100yr_low|avg_pct_area_flood_h
azard|
+---------------+---------------+--------------------------------+-------------------------------+-------------------------------+---------------------
-----+
|     Alang-alang|    14482.48828|                          3.8883|                        27.9909|                        17.3911|     16.4234333333
33335|
|         Umapad|    23250.31641|                          3.8599|                        21.6809|                        12.5203|     12.6870333333
33334|
|         Tipolo|    29959.33594|                          0.0044|                         14.772|                        28.0023|     14.2595666666
66666|
|         Tingub|    8646.967773|                         38.5303|                        23.7746|                         9.8841|                  2
4.063|
|        Tawason|    3651.513184|                          2.2985|                         7.6394|                         9.0076|      6.3151666666
66667|
|          Tabok|    11975.15039|                          9.0951|                        19.3805|                         8.7524|     12.4093333333
33334|
|     Subangdaku|    22441.55078|                          0.9995|                        45.5713|                        15.7135|     20.7614333333
33333|
|    Pagsabungan|    16310.59668|                          0.4862|                         9.7488|                         8.4596|      6.2315333333
33334|
|           Opao|    10080.33887|                          0.9649|                        28.8063|                        31.1449|     20.3053666666
66668|
|      Mantuyong|    1762.181763|                          0.0751|                        37.5041|                        12.4675|     16.6822333333
33333|
|       Maguikay|     14638.4834|                          9.3605|                        23.3542|                        10.6638|                  1
4.4595|
|           Looc|    18225.92188|                             0.0|                         0.5005|                        18.0568|      6.1857666666
66666|
|         Labogon|    13399.71289|                          7.9883|                        11.2459|                        14.6915|     11.3085666666
66666|
|       Pakna-an|    22086.93164|                          8.0926|                        19.7216|                        20.5728|                  1
6.129|
|Ibabao-Estancia|    7316.644043|                          0.7714|                         5.6427|                        26.0936|                  1
0.8359|
|       Jagobiao|    9222.175781|                          9.1363|                         7.4533|                         6.2414|      7.6103333333
33333|
|        Banilad|     15059.2959|                           0.387|                        10.0623|                         27.129|                  1
2.5261|
|          Basak|    5149.710938|                         11.3418|                         24.397|                        19.2503|                  1
8.3297|
|     Cabancalan|    14555.85938|                          4.5647|                        18.1682|                        19.0255|     13.9194666666
66667|
|        Cambaro|    10361.98047|                             0.0|                         4.8587|                        21.7246|      8.8610999999
99999|
+---------------+---------------+--------------------------------+-------------------------------+-------------------------------+---------------------
```

```
+-------+----------+------------------+-----------------------+----------------------------+-----------------------------------+
|summary|    adm4_en|    pop_count_total|avg_pct_area_flood_hazard|pop_count_total_Normalized|avg_pct_area_flood_hazard_Normalized|
+-------+----------+------------------+-----------------------+----------------------------+-----------------------------------+
|  count|        27|                27|                     27|                          27|                                 27|
|   mean|      NULL|12513.067681821836|      12.55796543209875|          0.3664473292920695|                 0.4469903208984114|
| stddev|      NULL| 7573.860098538878|      5.284563360024118|          0.2508957975259171|                 0.2540118128868994|
|    min|Alang-alang|1451.0218727259482|      3.2585999999999977|                         0.0|                                0.0|
|    max|     Umapad|31638.295526667924|      24.062999999999636|                         1.0|                                1.0|
+-------+----------+------------------+-----------------------+----------------------------+-----------------------------------+
```

[61]:

```python
import matplotlib.pyplot as plt
from pyspark.ml.clustering import KMeans
from sklearn.cluster import KMeans as skKMeans
import matplotlib.colors as mcolors

# Initialize KMeans with correct parameters
kmeans = KMeans(k=3, seed=42)

# Fit the model
model = kmeans.fit(dataset)

# Transform the data
predictions = model.transform(dataset)

# Add cluster predictions and adm4_en to the DataFrame
result_df = predictions.select('adm4_en', 'pop_count_total_Normalized', 'avg_pct_area_flood_hazard_Normalized', 'prediction')

# Show the results
result_df.show()

# Convert to Pandas DataFrame for visualization
pandas_df = result_df.toPandas()

# Elbow method to determine the optimal number of clusters
X = pandas_df[['pop_count_total_Normalized', 'avg_pct_area_flood_hazard_Normalized']]
wcss = []  # Within-cluster sum of squares
for i in range(1, 11):
    kmeans = skKMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)

# Plot the Elbow method result
plt.figure(figsize=(8, 6))
plt.plot(range(1, 11), wcss, marker='o')
plt.title('Elbow Method for Optimal K')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.grid(True)
plt.savefig('elbow_method.png')
plt.show()

# Custom color map for clusters
colors = ['yellow', 'orange', 'red']  # Color for each cluster
cmap = mcolors.ListedColormap(colors)
bounds = [0, 1, 2, 3]  # Define bounds for the color map
norm = mcolors.BoundaryNorm(bounds, cmap.N)
```

```python
# Visualize the clusters with custom shading
plt.figure(figsize=(12, 8))

scatter = plt.scatter(
    pandas_df['pop_count_total_Normalized'],
    pandas_df['avg_pct_area_flood_hazard_Normalized'],
    c=pandas_df['prediction'],
    cmap=cmap,
    norm=norm,
    marker='o'
)

# Add labels for each point
for i, row in pandas_df.iterrows():
    plt.text(
        row['pop_count_total_Normalized'],
        row['avg_pct_area_flood_hazard_Normalized'],
        row['adm4_en'],
        fontsize=8,
        ha='right'
    )

plt.xlabel('Normalized Population Count')
plt.ylabel('Normalized Avg Flood Area Percentage')
plt.title('K-means Clustering of Barangays')
plt.colorbar(scatter, label='Cluster', ticks=[0, 1, 2], format='%d')
plt.grid(True)
plt.savefig('kmeans_clustering.png')
plt.show()
```

```
+---------------+--------------------------+------------------------------------+----------+
|        adm4_en|pop_count_total_Normalized|avg_pct_area_flood_hazard_Normalized|prediction|
+---------------+--------------------------+------------------------------------+----------+
|        Banilad|        0.4706362212619998|                  0.4454586529772559|         0|
|        Tawason|       0.08133395301935935|                  0.1469192414425192|         0|
|           Opao|        0.3217892081474281|                  0.8193827587753882|         1|
|         Tipolo|                       1.0|                  0.5287807707343929|         2|
|           Looc|        0.5934948066617048|                  0.1406994033313459|         0|
|          Basak|        0.1337848864767461|                  0.7244188729307275|         1|
|        Maguikay|       0.46983710716595434|                  0.5383909173059512|         1|
|Ibabao-Estancia|        0.2201924873844392|                  0.3642162234911858|         0|
|          Tabok|       0.35509431394976043|                  0.4398460582056321|         0|
|        Bakilid|       0.12089400663954618|                  0.1945710202329020|         0|
|     Pagsabungan|         0.52210168942801|                  0.1428992584901887|         0|
|         Casili|                       0.0|                                 0.0|         0|
|        Cambaro|        0.3335077828093604|                  0.2692939955009549|         0|
|         Tingub|       0.25822825336183897|                                 1.0|         1|
|        Cubacub|       0.17966693397999872|                  0.2786750238731512|         0|
|     Alang-alang|       0.47482198303776335|                  0.6327908198906641|         1|
|         Umapad|        0.7480824827707481|                   0.453194196099545|         2|
|          Guizo|       0.07408710662773912|                  0.3258685662648292|         0|
|      Subangdaku|        0.7652539934762934|                  0.8413044035556465|         2|
|       Pakna-an|        0.7237296034921874|                   0.618638364961256|         2|
+---------------+--------------------------+------------------------------------+----------+
only showing top 20 rows
```