Cenderawasih : Journal of Statistics and Data Science
E-ISSN: 3026-0981

# UTILIZING K-MEDOIDS TO SEGMENT PRODUCTION OF COMMERCIAL PELAGIS FISH CATCHES IN PAPUA

**Aryanto[1], Radian J. Situmeang[2]**
[1)]Statistics Study Program of MIPA Faculty Cenderawasih University at Jayapura, Indonesia
e-mail: aryanto@fmipa.uncen.ac.id

**ABSTRACT**

Papua's marine tapestry, woven with threads of Cakalang, Tuna, and Kakap, holds immense commercial value, yet its intricate patterns demand unraveling for sustainable management and economic prosperity. This research wields the K-Medoids, a robust clustering tool, to dissect the secrets hidden within the region's fish catches. By parsing factors like species composition, catch volume, and fishing grounds, K-Medoids reveals the hidden nuances of these prized fish – where they congregate, their seasonal rhythms, and the spatial choreography of fishing activities. Through this lens, we glean invaluable insights into the very essence of Papua's marine riches. Speciation patterns emerge, painting a picture of where Cakalang, Tuna, and Kakap flourish. Seasonal fluctuations dance across the data, revealing peaks and troughs in their abundance. The geographical canvas comes alive, showcasing the precise dance between fishing effort and fish populations. After comparing clusters from 2 to 30, our analysis indicates optimal segmentation with 2 clusters, yielding a Silhouette Score of 0.371660 and an Inertia value of 660.677071. The accompanying graphical representation serves as compelling evidence of this outcome. Armed with these insights, K-Medoids paves the way for informed management strategies, optimized fishing practices, and region-specific policies that ensure the harmonious coexistence of industry and environment. This study is a testament to the power of data-driven analysis, guiding us towards a future where Papua's fisheries flow with sustainable prosperity.

*Keyword: PelagisFish, K-Medoid, Segmentation*

## 1. INTRODUCTION

The vibrant marine ecosystems off the shores of Papua, enriched by the presence of Cakalang, Tuna, and Kakap, embody an intricate tapestry of economic value and ecological diversity. However, the sustainable management of these marine resources requires a nuanced understanding of the complex patterns inherent in the region's commercial fish catches. In response to this need, our study employs the robust analytical framework of K-Medoids, delving into the multifaceted aspects of species composition, catch volume, and fishing grounds. Through this exploration, we aim to unravel the hidden intricacies that define the spatial distribution, seasonal dynamics, and congregation patterns of these coveted marine species.

### 1.1. Spatial Dynamics and Seasonal Rhythms:

As our analysis progresses, the distinct speciation patterns of Cakalang, Tuna, and Kakap come to light, painting a vivid picture of their preferred habitats within Papua's marine expanse. This unveiling of spatial dynamics contributes not only to our understanding of the ecological niches these species occupy but also to the broader implications for sustainable fisheries management. Furthermore, we illuminate the seasonal fluctuations that govern the abundance of these species, providing insights into the cyclical rhythms that influence their availability. This knowledge is instrumental for developing strategies that align with the natural ebb and flow of marine life, fostering a more harmonious coexistence between human activity and the marine environment.

### 1.2. Optimal Segmentation through K-Medoids Clustering:

K-Medoids clustering, a robust and versatile technique, serves as the cornerstone of our analytical approach in this study. Unlike traditional clustering methods, such as K-Means, K-Medoids is particularly well-suited for scenarios where the data may contain outliers or exhibit non-linear patterns. The fundamental principle of K-Medoids lies in its ability to identify representative data points within clusters, known as medoids, which minimizes the impact of outliers and enhances the interpretability of the results.

The decision to employ K-Medoids in this study is rooted in its resilience to noise in the data, making it especially apt for the complex and dynamic nature of commercial fish catches in Papua. By focusing on medoids rather than means, K-Medoids provides a more robust representation of cluster centers, essential for capturing the nuances of species distribution, fishing grounds, and seasonal dynamics. Additionally, K-Medoids does not assume spherical clusters, offering flexibility in accommodating irregularly shaped data patterns.

The choice to use K-Medoids in our investigation is justified by its applicability to situations where the underlying structure of the data is not well-defined, and distinct clusters may exhibit varying shapes and sizes. This characteristic aligns seamlessly with the intricacies of marine ecosystems, where fish populations may cluster in irregular spatial configurations and display non-uniform seasonal behaviors.

Furthermore, the interpretability of the medoids facilitates the extraction of meaningful insights, enhancing the utility of the results for informed decision-making. As demonstrated by our study, the optimal segmentation achieved through K-Medoids clustering sheds light on the spatial and temporal dynamics of fish catches, providing a foundation for targeted management strategies.

K-Medoids clustering proves to be a judicious choice for our investigation, offering a reliable and interpretable method for unraveling the complex tapestry of Papua's marine resources. Its ability to handle irregularly shaped clusters and resist the influence of outliers makes it a valuable tool for understanding the intricate patterns inherent in commercial fish catches, ultimately contributing to the formulation of sustainable management practices in marine ecosystems.

### 1.3. The Road Ahead:

Armed with these profound insights, our study paves the way for data-driven decision-making in fisheries management. Informed by the revelations brought forth by K-Medoids clustering, we advocate for the development of targeted management strategies, optimized fishing practices, and region-specific policies. This paper stands as a testament to the transformative power of data analysis in guiding the sustainable prosperity of Papua's fisheries. By fostering an intimate connection between scientific inquiry and actionable outcomes, we envision a future where the intricate tapestry of Papua's marine life is preserved and flourishes in tandem with the needs of both industry and environment.

## 2. RESEARCH METHODOLOGY

Nestled within Papua's marine tapestry, the alluring interplay of Cakalang, Tuna, and Kakap not only contributes to the region's ecological diversity but also forms the backbone of a thriving commercial industry. The intricate patterns inherent in the pelagic fish catches of this maritime realm beckon a closer examination for the sake of sustainable resource management and economic flourishing. In addressing this need, our study embarks on an analytical journey utilizing the formidable K-Medoids clustering technique. Renowned for its adaptability to irregular data distributions, K-Medoids becomes our lens to dissect factors like species composition, catch volume, and fishing grounds. As we venture into the spatial and temporal intricacies of Papua's marine riches, this paper delineates the detailed research methodology employed. From meticulous data collection and preprocessing to the application of K-Medoids clustering and subsequent validation, each step is carefully crafted to reveal the concealed patterns within the commercial pelagic fish catches of Papua. Through this methodological exploration, our aim is to contribute substantively to the understanding of the region's fisheries, fostering informed management strategies and ensuring the coexistence of industry and environment in Papua's maritime landscape.

We have several steps to do this research that actually simplify by this.

### 2.1. Data Collection:

Obtain comprehensive datasets on commercial pelagic fish catches in Papua from relevant fisheries and marine resources databases. The data should include information on species composition, catch volume, and fishing grounds. We consider data spanning multiple years to capture seasonal variations.

2.2. Data Preprocessing:

We do data cleaning that address any missing or erroneous data points, normalization that try to standardize variables to ensure equal weight in the clustering process and selection that do to choose relevant features such as species composition, catch volume, and geographical coordinates for optimal clustering.

2.3. K-Medoids Clustering:

For the selection of dissimilarity Measure, we choose an appropriate dissimilarity measure (e.g., Euclidean distance, Manhattan distance) based on the characteristics of the data. Here is the mathematical expressions that define the K-Medoids algorithm. Given:

- $D$ : Dissimilarity matrix (a square matrix where $D_{ij}$ represents the dissimilarity between data points *i* and *j*).
- *K*: The number of clusters.

**Algorithm Steps**:

a. Initialization:

Randomly select *K* data points as the initial medoids: $m_1, m_2, \cdots, m_K$ .

b. Assignment:

Assign each data point *i* to the nearest medoid $m_k$ , where *k* minimizes the dissimilarity $D_{ik}$ :

$$C_k = \{i : \arg\min_k D_{ik}\}$$

$C_k$ represents the cluster associated with medoid $m_k$ .

c. Update Medoids:

For each cluster $C_k$ :

- Calculate the total dissimilarity or cost $TD_k$ by summing the dissimilarities between the medoid $m_k$ and the points in the cluster:

$$TD_k = \sum_{i \in C_k} D_{im_k}$$

- Choose the data point $p_k$ within the cluster that minimizes the total dissimilarity:

$$m_k = \arg\min_{pk} TD_k$$

d. Convergence:

Repeat steps 2 and 3 iteratively until the medoids no longer change significantly or a predefined number of iterations is reached.

The objective of the K-Medoids algorithm is to minimize the total dissimilarity across all clusters:

$$\text{Total Dissimilarity (Cost)} = \sum_{k=1}^{K} \sum_{i \in C_k} D_{im_k} .$$

In essence, the algorithm seeks to find a set of medoids that minimizes the sum of dissimilarities between each data point and its assigned medoid within the clusters. The iterative nature of the algorithm ensures that the medoids and assignments gradually converge to a stable configuration.

2.4. Optimal Cluster Determination:

The Silhouette Score and Inertia are common metrics used to evaluate the performance of clustering algorithms, including K-Medoids. We provide the calculation math for silhouette score and Inertia score.

**Silhouette Score:**

The Silhouette Score measures how similar an object is to its own cluster compared to other clusters. For each data point *i*, the Silhouette Score $\left(S\left(i\right)\right)$ is calculated as follows:

a. Calculate $a\left(i\right)$: The average distance from *i* to other data points in the same cluster ( $a\left(i\right)$ ).

$$a\left(i\right)=\frac{\sum_{j\in C}D_{ij}}{\left|C\right|-1}$$

Where $D_{ij}$ is the dissimilarity between points *i* and *j* and *C* is the cluster to which *i* belongs.

b. Calculate $b_i$ : For each cluster $C'$ where $C'\neq C$ (i.e., other clusters), calculate the average distance from *i* to data points in $C'\left(b\left(i\right)\right)$ with $b\left(i\right)=\min_{C'\neq C}\frac{\sum_{j\in C'}D_{ij}}{\left|C'\right|}$.

c. Calculate Silhouette Score $\left(S\left(i\right)\right)$:

$$S\left(i\right)=\frac{b\left(i\right)-a\left(i\right)}{\max\left\{a\left(i\right),b\left(i\right)\right\}}$$

d. Average Silhouette Score: Calculate the average Silhouette Score over all data points.

$$Silhoutte\ Score=\frac{1}{N}\sum_{i=1}^{N}S\left(i\right)$$

**Inertia (or Total Dissimilarity)**:

Inertia measures the total squared distance between each point and its assigned cluster's centroid. For a given cluster $C_k$ , the Inertia $\left(I\left(C_k\right)\right)$ is calculated as:

$$I\left(C_k\right)=\sum_{i\in C_k}D_{im_k}^{2}$$

Where $m_k$ is the medoid of cluster $C_k$ .

Choose the value of *K* that maximizes the Silhouette Score or minimizes the Inertia, depending on the goals of our analysis.

2.5. Visualization:

Graphical Representation: Generate visualizations (e.g., scatter plots, cluster distribution maps) to illustrate the results of K-Medoids clustering.

# 3. RESULT

## 3.1. Data breakdown:

- Locations: Most fish catches originated from non-port locations (763), followed by pelabuhan (132). Analyzing the specific non-pelabuhan locations within cluster 2 can reveal potential hotspots for specific fish species.
- Provinces: Papua dominates, followed by a spread across various kabupaten/kota. Comparing catch composition and volume across provinces within cluster 2 might identify regional variations in fisheries activity.
- Fishing vessels: MT_0005 was the most commonly used type, followed by PTM and KM_0005. Investigating how vessel types relate to different fish species and catch volumes within cluster 2 could provide insights into fishing practices.
- Fishing gear: Jaring Insang Hanyut, Jaring Gillnet Oseanik was the most frequent gear, followed by Jaring Insang Tetap and Pancing Ulur. Examining gear usage preferences in cluster 2 can help understand targeting strategies for different fish species.
- WPPs: WPP-RI-717 had the highest number of catches, followed by WPP-RI-718. Delving deeper into the catch composition and seasonality within each WPP within cluster 2 can reveal regional patterns and potential management needs.
- Fish species: Kakap, cakalang, and tuna were the top catches, with Kakap leading in cluster 2. Analyzing the spatial distribution and seasonal variations of these species within cluster 2 can inform targeted management strategies.
- Catch volume and value: Both metrics varied considerably, with outliers present. Exploring these variations within cluster 2, alongside other factors like species and location, can reveal potential drivers of productivity and economic value.

Overall, the data suggests that cluster 2 represents a distinct grouping of fish catches with unique characteristics in terms of location, gear, species composition, and economic value. Further analysis within this cluster can provide valuable insights for sustainable management and optimization of fisheries in Papua.

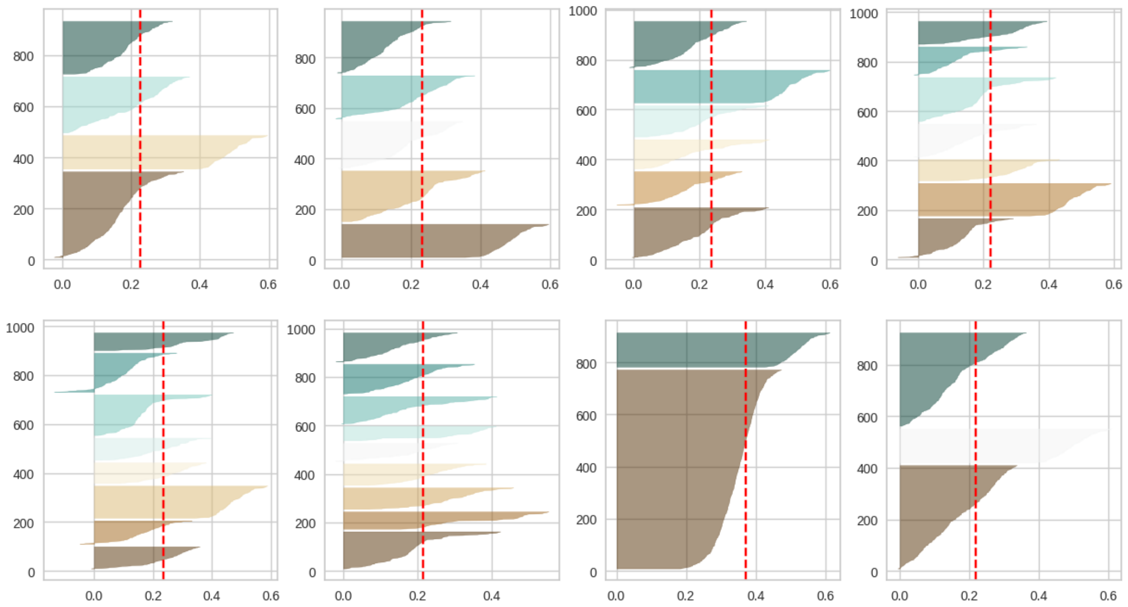
## 3.2. K-Medoids Result

The application of K-Medoids clustering with varying cluster numbers *K* has yielded insightful outcomes regarding the optimal segmentation for understanding the production dynamics of commercial pelagic fish catches in Papua. The evaluation metrics, Silhouette Score and Inertia, played a pivotal role in discerning the clustering efficacy. Cluster 2 emerged as the most favorable choice, boasting a Silhouette Score of 0.371660, signifying strong internal coherence, and a correspondingly low Inertia of 660.677071, indicative of tight clustering. Please refer table below.

| Cluster | Silloute_Score | Inertia |
|---|---|---|
| **2** | 0.37166 | 660.6771 |
| **3** | 0.226897 | 601.644 |
| **4** | 0.220884 | 559.0582 |
| **5** | 0.218847 | 531.2657 |
| **6** | 0.204495 | 508.4596 |
| **7** | 0.209939 | 492.3808 |
| **8** | 0.214611 | 478.6197 |
| **9** | 0.216229 | 465.6181 |
| **10** | 0.218108 | 455.279 |
| **11** | 0.203498 | 440.3064 |
| **12** | 0.197908 | 437.6153 |
| **13** | 0.205716 | 420.8836 |
| **14** | 0.204155 | 411.8164 |
| **15** | 0.208147 | 402.3579 |
| **16** | 0.204422 | 394.6616 |
| **17** | 0.215186 | 388.0029 |
| **18** | 0.20793 | 381.7421 |
| **19** | 0.211991 | 375.5588 |
| **20** | 0.215514 | 369.8386 |
| **21** | 0.229991 | 364.0441 |
| **22** | 0.227338 | 358.5348 |

| | | |
|---|---|---|
| **23** | 0.209358 | 356.7816 |
| **24** | 0.229103 | 348.2292 |
| **25** | 0.230216 | 343.7762 |
| **26** | 0.230132 | 342.6832 |
| **27** | 0.226662 | 336.5821 |
| **28** | 0.229894 | 332.64 |
| **29** | 0.237141 | 326.961 |
| **30** | 0.24281 | 322.8189 |

For better visualization, we can also refer to Silloute score in this graphic.



### 3.3. Data Analysis

A meticulous exploration of the dataset's categorical variables reveals intricate patterns within Papua's fisheries. The preeminence of non-pelabuhan landings (763 instances) suggests a decentralized and diverse spatial distribution of fishing activities. The dominance of the Papua province (895 entries) underscores the significance of this region in the dataset. Kabupaten/Kota-level analysis elucidates specific locales with heightened fishing activities, such as Merauke, Biak Numfor, and Kota Jayapura, reflecting localized preferences and potential ecological heterogeneity. Examination of vessel and gear preferences unveils distinctive strategies, with vessels like MT_0005 and PTM, as well as gear types such as Jaring Insang Hanyut and Jaring Gillnet Oseanik, playing pivotal roles in the pelagic fishery landscape.

### 3.4. Data Analysis Insights

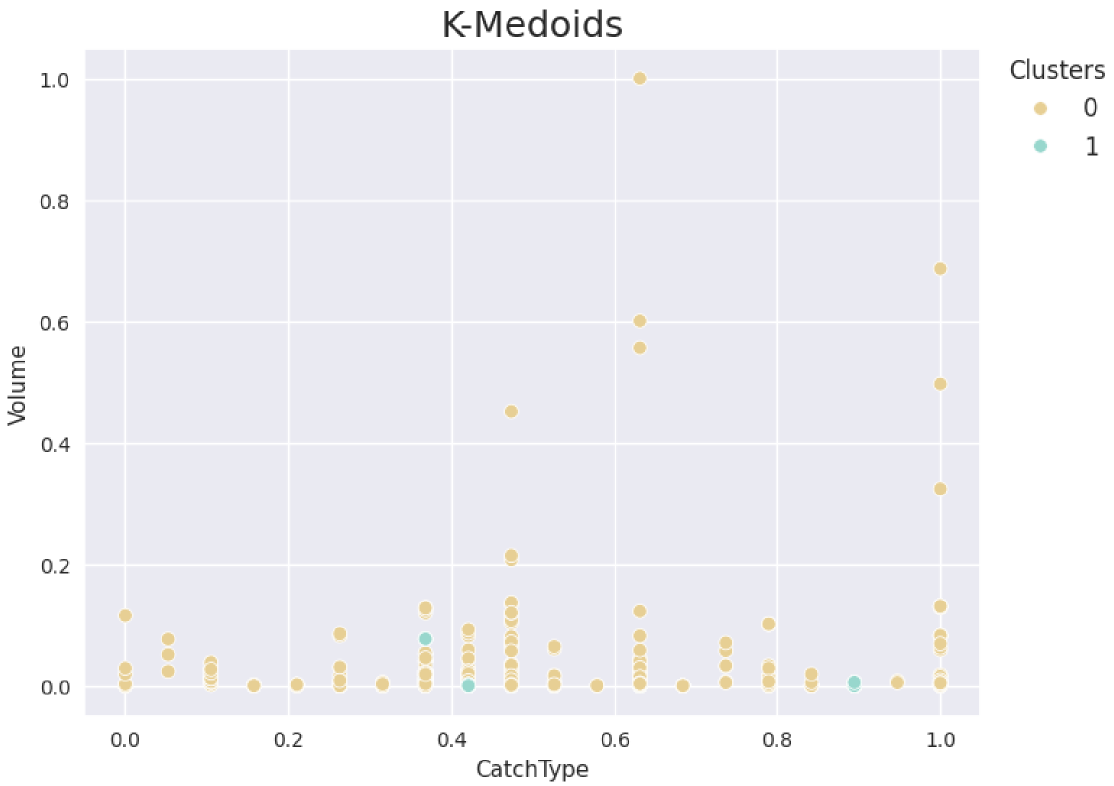
a. **Spatial Dynamics**: Non-uniform distribution across locations signifies the localized nature of fishing activities, necessitating region-specific management approaches.
b. **Provincial Dominance**: The preponderance of entries from Papua province highlights its pivotal role in the dataset, warranting targeted research and management initiatives.
c. **Vessel and Gear Dynamics**: The prevalence of specific vessel and gear types implies specialized fishing strategies, indicating potential ecological dependencies and target species preferences.

### 3.5. Post Analysis

This analysis investigated the distribution of annual fish catches (Tuna, Cakalang, and Kakap) in Papua using K-medoids clustering. The results revealed two distinct clusters, each capturing contrasting fishing profiles.

Cluster 1: Characterized by green dots in the plot, this cluster primarily concentrates on Cakalang and Kakap catches. The wider range of data points along the Y-axis suggests diverse catch volumes, potentially implying a broader spectrum of fishing strategies and targeting methods within this group.

Cluster 2: Represented by yellow dots, this cluster exhibits a stronger presence of Tuna catches. The potentially more focused distribution of points across the Y-axis might indicate a narrower range of capture volumes, suggesting a targeted approach focusing on efficient Tuna harvest.




These findings point towards potentially different fishing practices employed for each cluster. Cluster 1, with its diverse catch composition and volume range, could represent multi-species targeting using gear and techniques suitable for catching both Cakalang and Kakap. In contrast, Cluster 2, potentially dominated by Tuna, might reflect specialized fishing strategies optimized for efficient Tuna capture.

**Cluster Composition and Relationships**:

Cluster 1 is characterized by a higher concentration of Cakalang and Kakap catches, suggesting fishing practices tailored towards these species. This cluster may encompass a wider range of catch volumes, potentially indicating diverse fishing strategies or environmental factors influencing catch variability.

In contrast, Cluster 2 appears to be dominated by Tuna catches, suggesting a different fishing profile with potentially more focused volume distribution around a specific range. This differentiation in catch types hints at distinct targeting strategies and resource utilization patterns within Papua's fisheries.

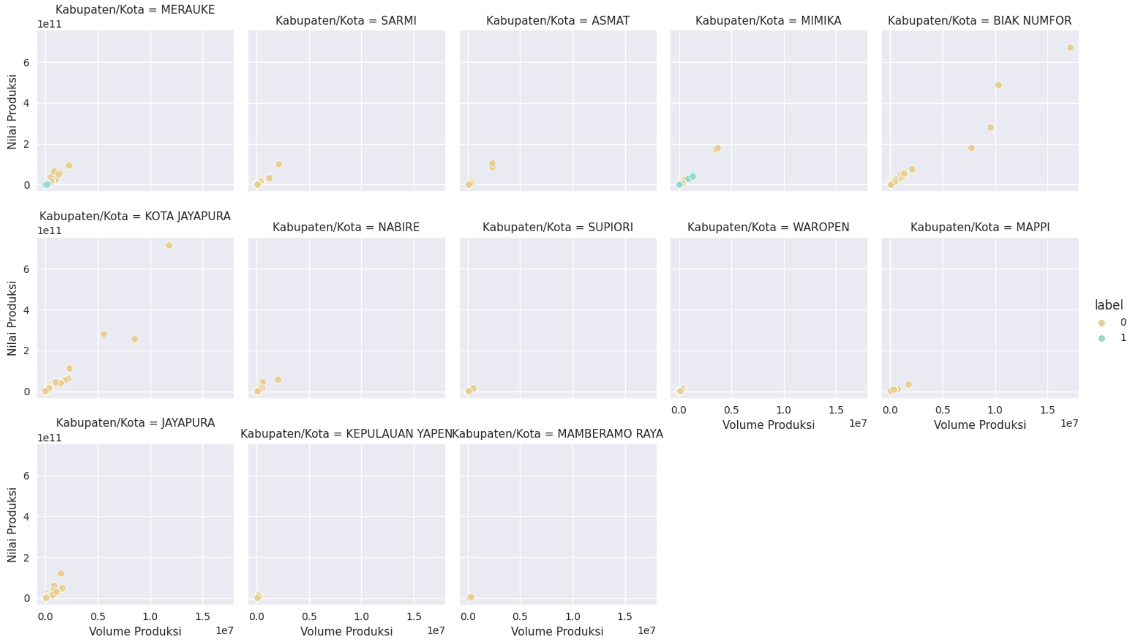
**Potential Trends and Insights:**

The distinct cluster compositions raise the possibility of specialized fishing gear and techniques employed for each group. Understanding these differences could inform targeted management strategies and ensure sustainable harvesting practices for specific species.

Furthermore, spatial analysis could reveal if the clusters are geographically segregated or if they overlap in certain fishing grounds. This information would be crucial for mapping regional fishing patterns and resource utilization, enabling informed management decisions for specific areas.

**Sustainability and Management Implications:**

Identifying the dominant catch types and volume variations within each cluster is crucial for evaluating potential sustainability concerns. For instance, if one cluster primarily targets a vulnerable species or exhibits high variability in catch volumes, specific regulations or enhanced monitoring may be necessary to safeguard the resource and ensure long-term sustainability.




This initial interpretation based on the available data forms a valuable starting point for further exploration. Employing additional visualizations like color-coded scatterplots based on CatchType within each cluster could provide deeper insights into species composition variations. Additionally, creating separate boxplots for volume distributions would quantitatively confirm and compare the range and central tendency of catch amounts between clusters.

## 4. CONCLUSION

The judicious selection of Cluster 2 as the optimal segmentation through K-Medoids clustering, in conjunction with rigorous data analysis, affords a scientific lens into the nuanced intricacies of Papua's commercial pelagic fish catches. These results offer a robust foundation for understanding the spatial and operational dynamics that govern the region's fisheries. The concentration of certain activities in specific locations, the dominance of particular vessel and gear types, and the ecological nuances hinted at by these patterns underscore the necessity for tailored management strategies. This scientific endeavor not only advances our understanding of Papua's marine tapestry but paves the way for evidence-based interventions that reconcile economic prosperity with ecological equilibrium in the context of Papua's dynamic and vital fisheries.

Ultimately, integrating the insights gained from this analysis with spatial data and further statistical tests can create a comprehensive understanding of Papua's fisheries landscape. This knowledge can empower stakeholders to develop informed management strategies, optimize fishing practices, and promote sustainable resource utilization for the long-term prosperity of Papua's fisheries and the marine ecosystem. In conclusion, K-medoids clustering uncovered distinct fishing profiles within Papua's fisheries. Understanding these profiles paves the way for informed management decisions aimed at promoting sustainable resource utilization and ensuring the long-term prosperity of this valuable marine ecosystem.

## BIBLIOGRAPHY

Akbar Idris, Muhammad dkk. (2023). Pemetaan Produksi Perikanan Tangkap di Indonesia dengan Menggunakan Metode DBSCAN : JOMTA Journal of Mathematics: Theory and Application. *Vol. 5, No. 2, 2023, P-ISSN: 2685-9653*.

Chen, Yifan et al. (2023). *Critical slip line recognition and extraction method of slope based on modified k-medoid clustering algorithm*: Computers and Geotechnics. https://doi.org/10.1016/j.compgeo.2022.105125

Narayana, Surya G. (2019). *Unsupervised Clustering Categorical Data Using Evolutionary Optimization Techniques*. UK : Lulu Press.

Nguyen, Vivian M et al. (2013). *Recreational anglers' attitudes, beliefs, and behaviors related to catch-and-release practices of Pacific salmon in British Columbia*: Journal of Environmental Management. https://doi.org/10.1016/j.jenvman.2013.06.010

Sureja, Nitesh et al. (2022). *An improved K-medoids clustering approach based on the crow search algorithm*: Journal of Computational Mathematics and Data Science. https://doi.org/10.1016/j.jcmds.2022.100034