# Where is the Best Place to Work in Indonesia?

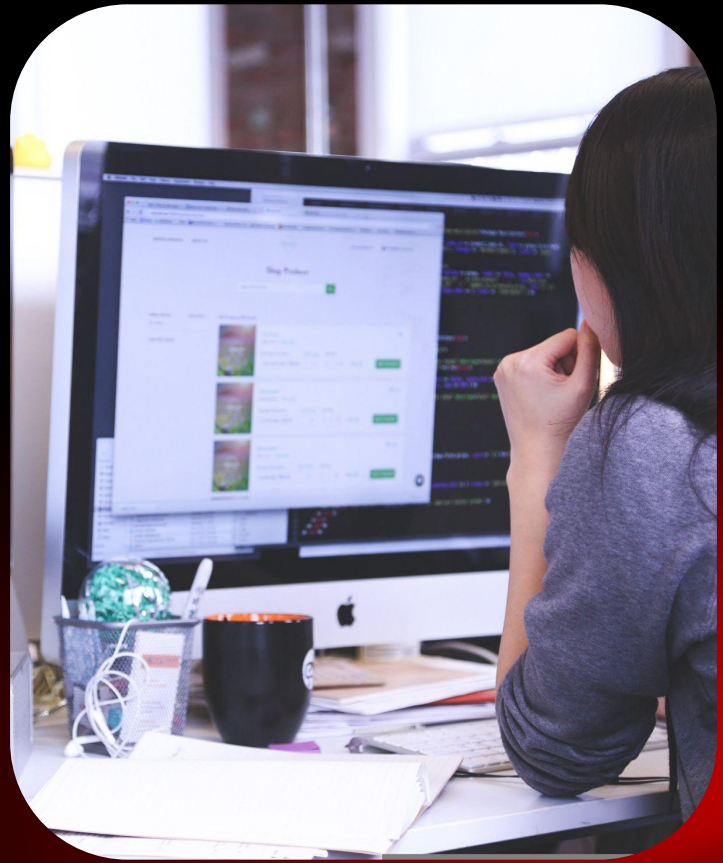
## -Analysis of Worker Welfare 2015-2022-

Bela Sisilia

# Project Summary

This project uncovers the most advantageous provinces in Indonesia that offer the best opportunities and quality of life for workers. By analyzing data from 2015 to 2022, including well-being scores, wage levels, poverty lines, and poverty gaps. This analysis highlights where talent can thrive and businesses can grow. The insights provide a clear picture of the regions with the strongest potential for a successful and rewarding workforce environment.

**Technical Stack:** Python (Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn).



# Dataset Overview

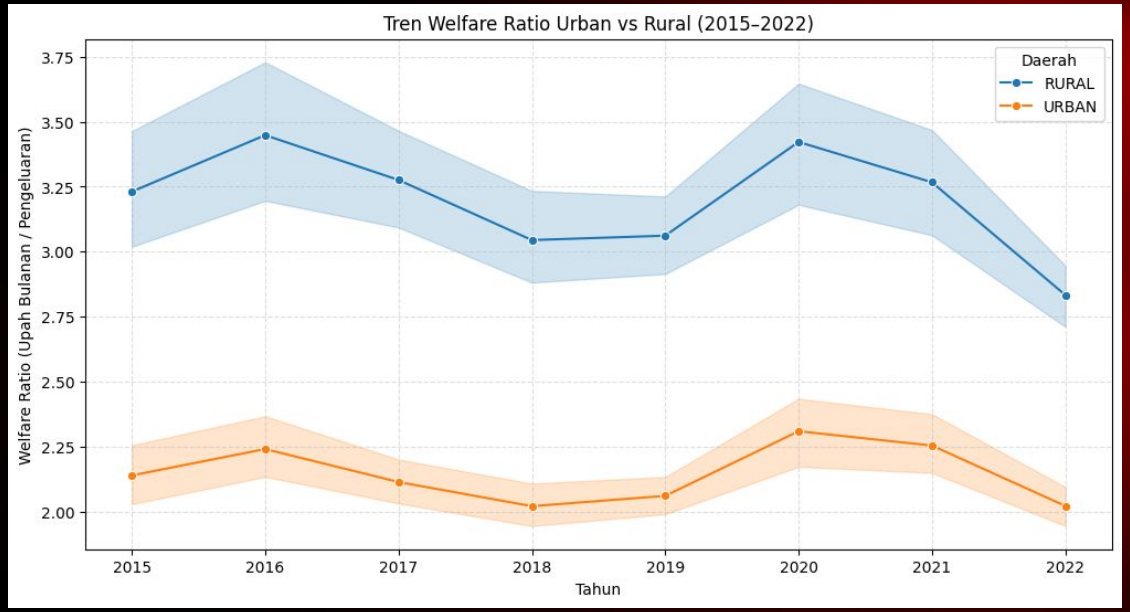
The dataset used in this project is the Indonesian Workers' Welfare dataset obtained from Kaggle.com, which is a combination of several datasets sourced from the Central Badan Pusat Statistik (BPS). There are four datasets included, which are:

- upah.df.csv
- ump.df.csv
- gk.df.csv
- peng.df.csv

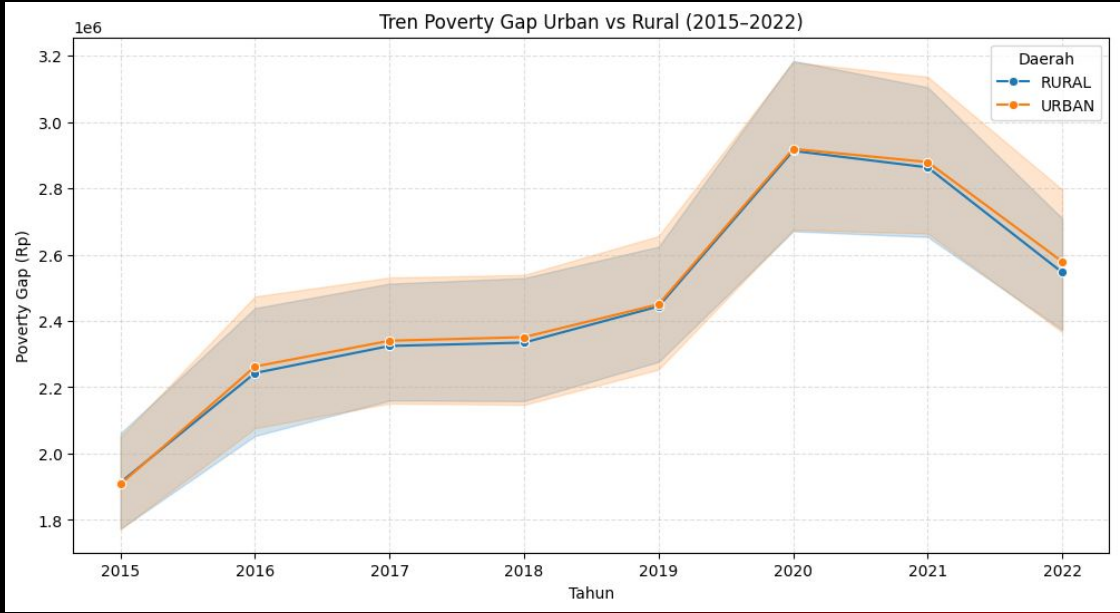
The dataset undergoes a structured data cleaning and standardization process, including the normalization of province name fields and temporal filtering to ensure consistency with the March reference period. Following this, feature engineering is performed to generate two key analytical metrics: **welfare_ratio** and **poverty_gap**.

In addition to the two feature-engineered metrics used as core variables, this analysis also incorporates the following variables:

- Hourly Wage: average income earned per hour.
- Provincial Minimum Wage (UMP): government-mandated minimum wage for each province.
- Total Poverty Line: threshold used to distinguish poor and non-poor households.
- Total Per-Capita Expenditure: average household spending level.




The processed data reveals that rural regions consistently demonstrate higher welfare metrics than urban regions. Additionally, the temporal trend indicates a downturn in 2017–2018, a recovery in 2020, and a subsequent decline during the 2021–2022 period.

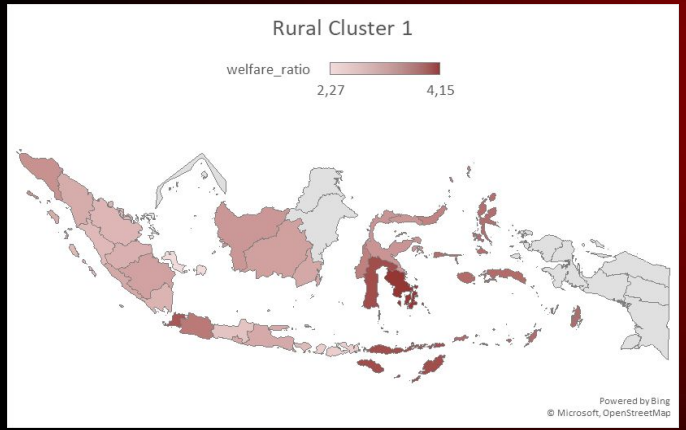


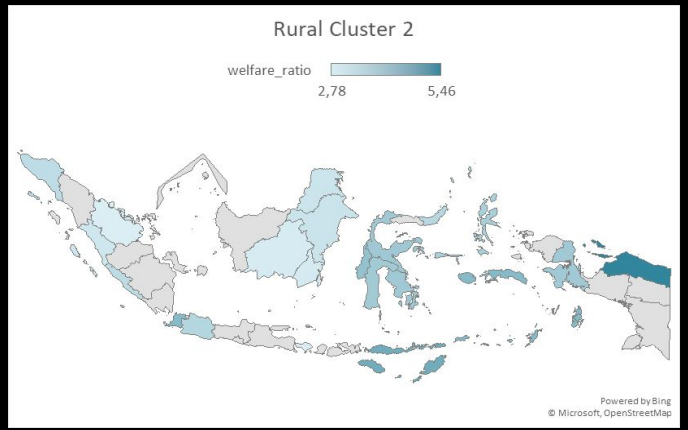

The visualization presents the poverty gap trend for urban and rural workers from 2015 to 2022.

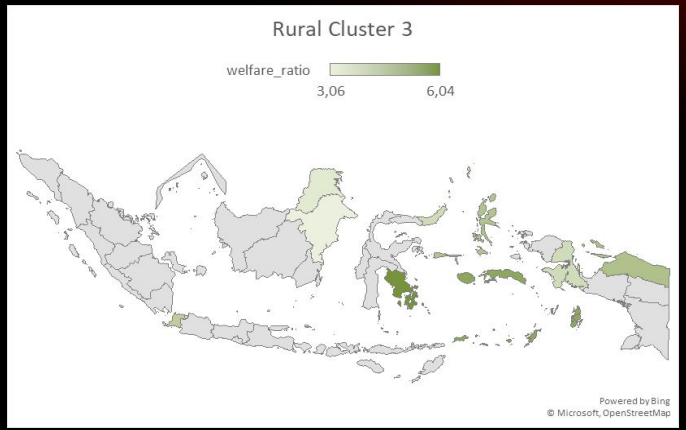
The chart illustrates the changes in the poverty gap among The poverty gap measures how far workers' incomes are positioned above the poverty line, with higher values indicating lower vulnerability to poverty. The relatively small difference between urban and rural trajectories suggests that welfare pressures are becoming more evenly distributed, affecting both regions in comparable ways.in urban and rural areas from 2015 to 2022.

The K-Means algorithm was applied to segment the provinces into three clusters based on worker welfare indicators:

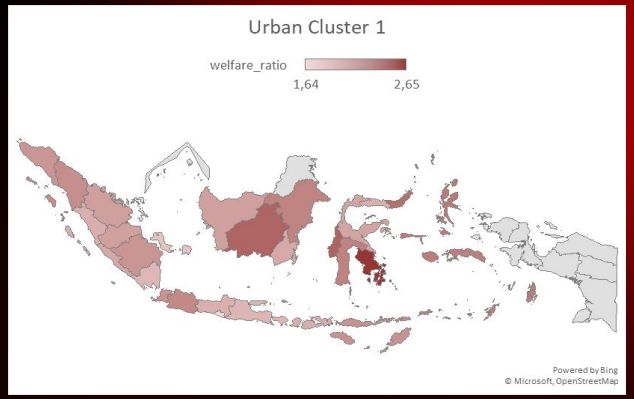
- Cluster 1 – Economically At-Risk
- Cluster 2 – Economically Stable
- Cluster 3 – Economically Secure

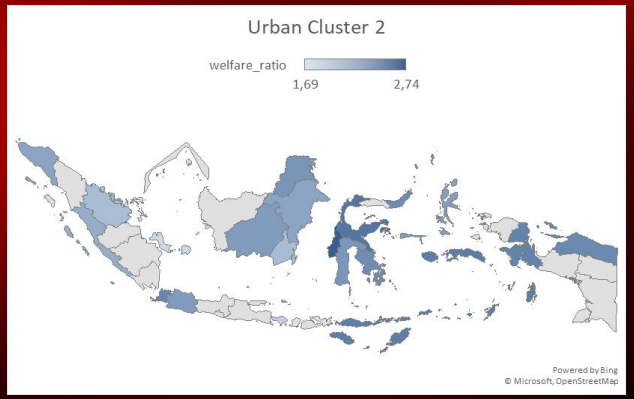


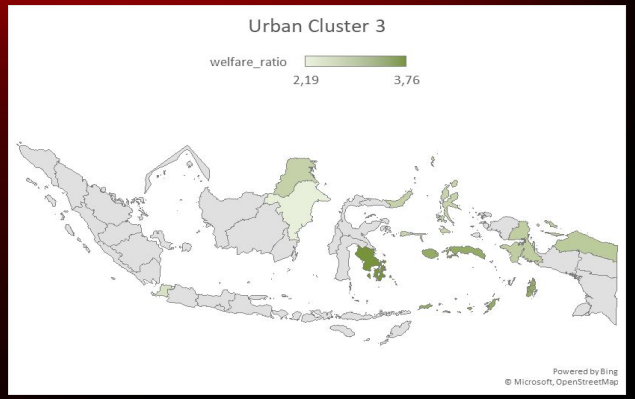




In **rural** areas, Cluster 1 includes 28 provinces with welfare ratios ranging from 2.27 to 4.15, Cluster 2 includes 22 provinces with ratios between 2.78 and 5.46, and Cluster 3 includes 10 provinces with ratios between 3.06 and 6.04.



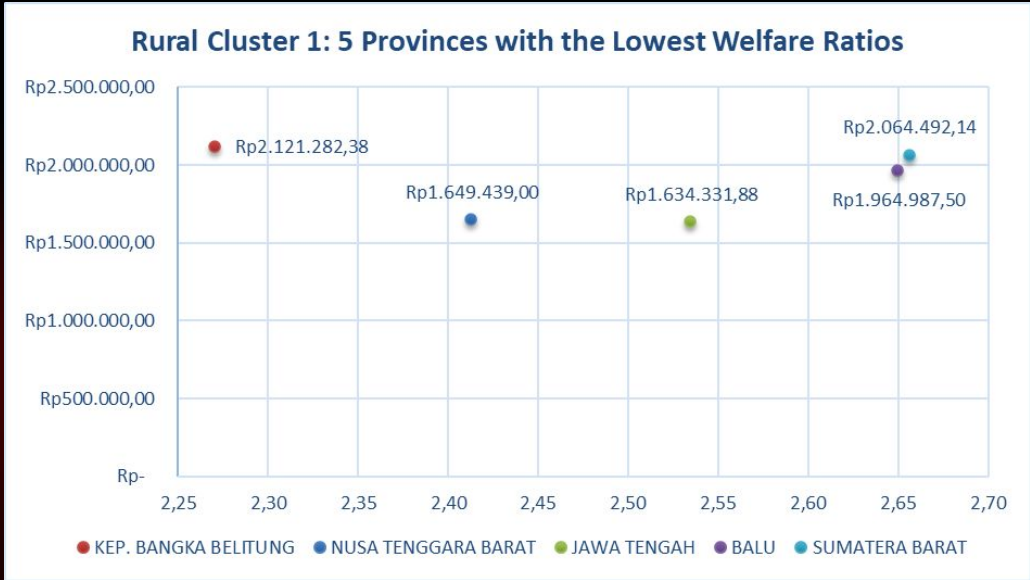




In **urban** areas, Cluster 1 includes 29 provinces with welfare ratios ranging from 1.64 to 2.65, Cluster 2 includes 24 provinces with ratios between 1.69 and 2.74, and Cluster 3 includes 11 provinces with ratios between 2.19 and 3.79.

The clustering results, both for rural and urban areas, reveal instances of cluster overlap, where certain provinces display characteristics that align with more than one cluster. This occurs because their feature values, particularly the welfare ratio and poverty gap, are positioned near the algorithm's decision boundaries. In this analysis, the poverty gap serves as the primary factor distinguishing these overlapping provinces.

# 5 Provincial Welfare Clusters (Rural Areas)

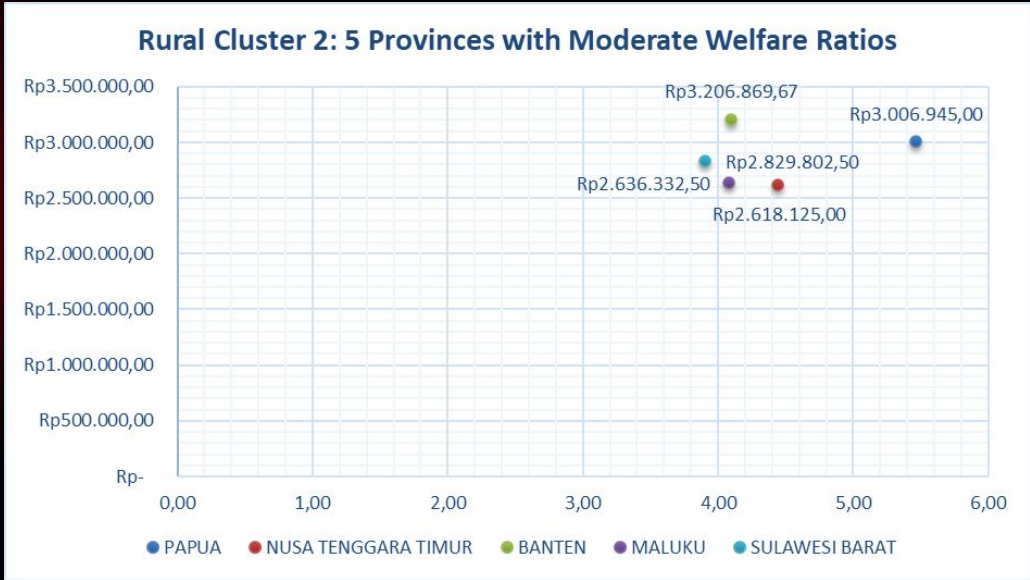
# Cluster 1




The graph presents the five rural provinces with the lowest welfare levels.

In this cluster, each province exhibits a low welfare ratio, indicating that household incomes in rural areas remain insufficient to fully meet basic living needs. Several provinces also show relatively high expenditure despite their low welfare ratios, suggesting greater pressure from rising living costs. The values across provinces in this cluster are closely distributed, causing some data points to lie near the boundaries of other clusters and appear similar to those in adjacent groups.

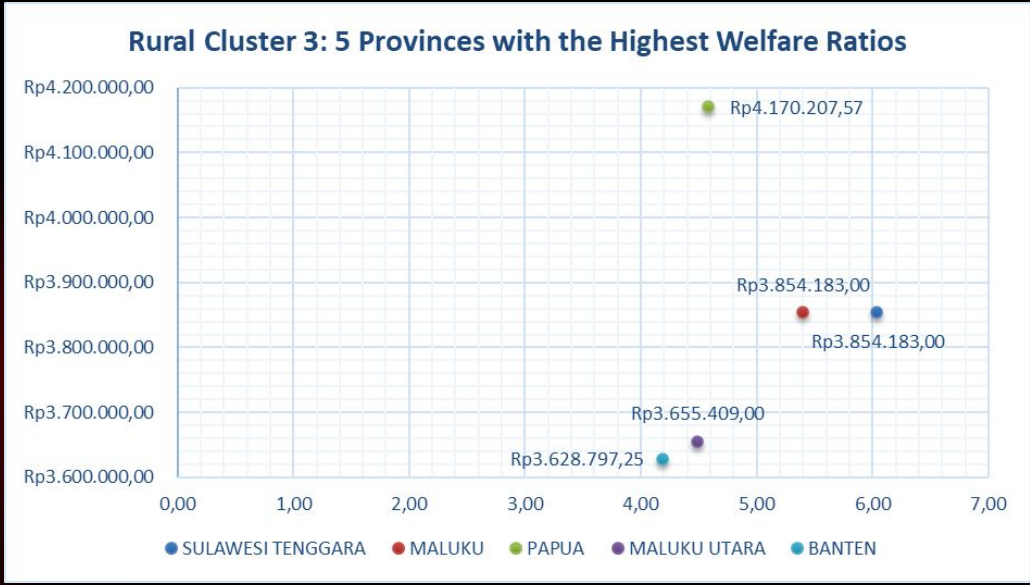
# Cluster 2




The five provinces shown in this graph represent Cluster 2, which corresponds to the group with medium welfare levels.

The welfare ratios in this cluster are higher than those in Cluster 1, yet disparities remain evident, as several provinces exhibit significantly higher expenditure while maintaining only moderate welfare ratios. These differences highlight that rural welfare conditions are not homogeneous and that some provinces share characteristics with those in other clusters. The presence of provinces with similar welfare ratios but different positions on the graph suggests that additional variables also influence the clustering outcomes.

# Cluster 3

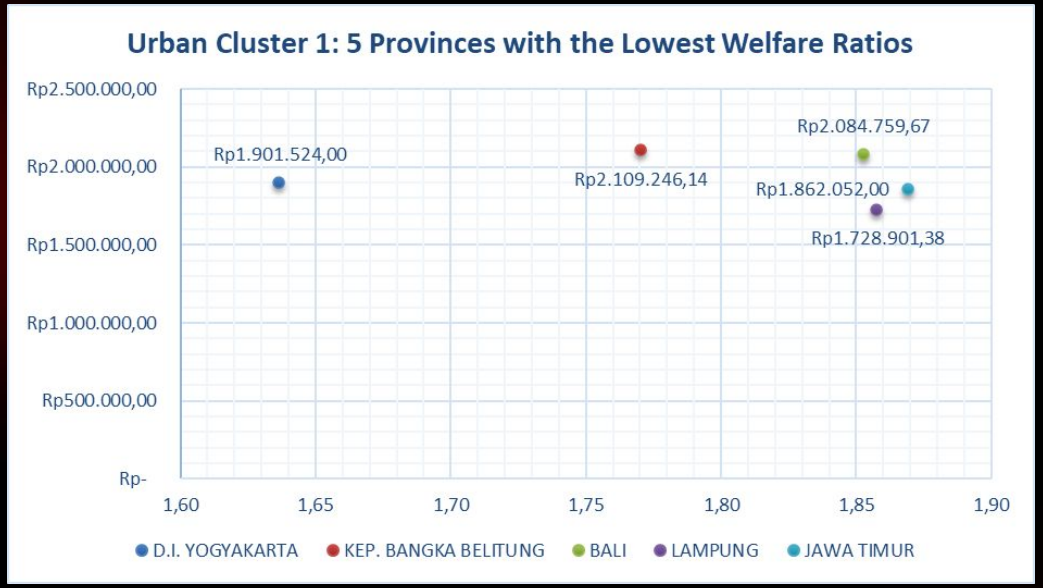



Cluster 3 consists of provinces with high welfare levels, with the graph highlighting the five provinces that have the highest welfare ratios.

A higher welfare ratio reflects a stronger ability to meet basic needs compared to the other two clusters. However, variation still exists within this group: some provinces show high welfare ratios but relatively low expenditure, while others demonstrate higher expenditure yet remain classified within the high-welfare cluster. The proximity of values across provinces also results in several data points appearing near the cluster boundaries.

# 5 Provincial Welfare Clusters (Urban Areas)

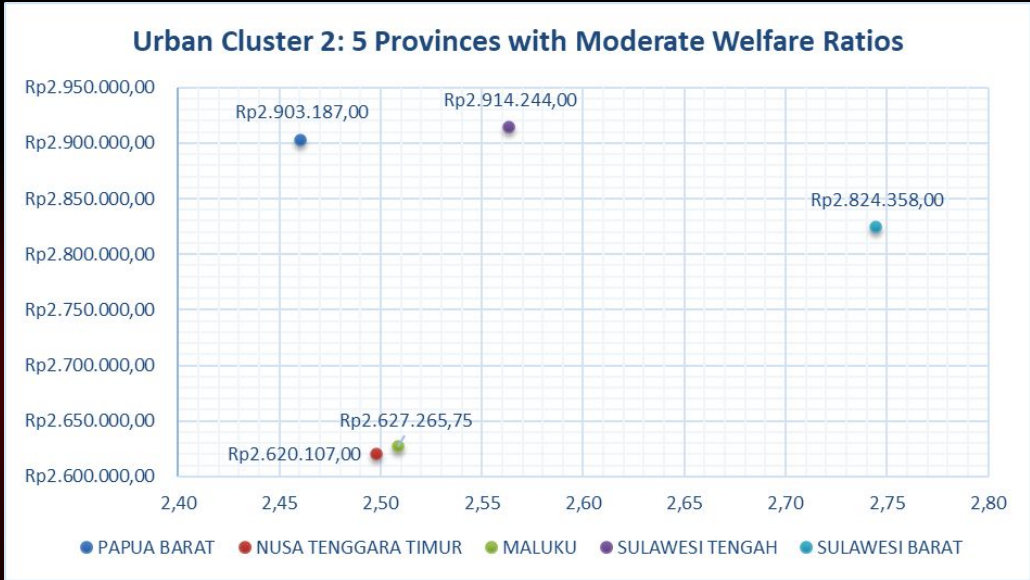
# Cluster 1




The graph illustrates the five provinces with the lowest welfare levels in urban areas.

A lower welfare ratio indicates that workers' incomes are insufficient to comfortably meet basic needs. In several provinces, the poverty gap is larger despite having similar—or even lower—welfare ratios. This suggests that vulnerable households in these regions may fall further below the poverty line.

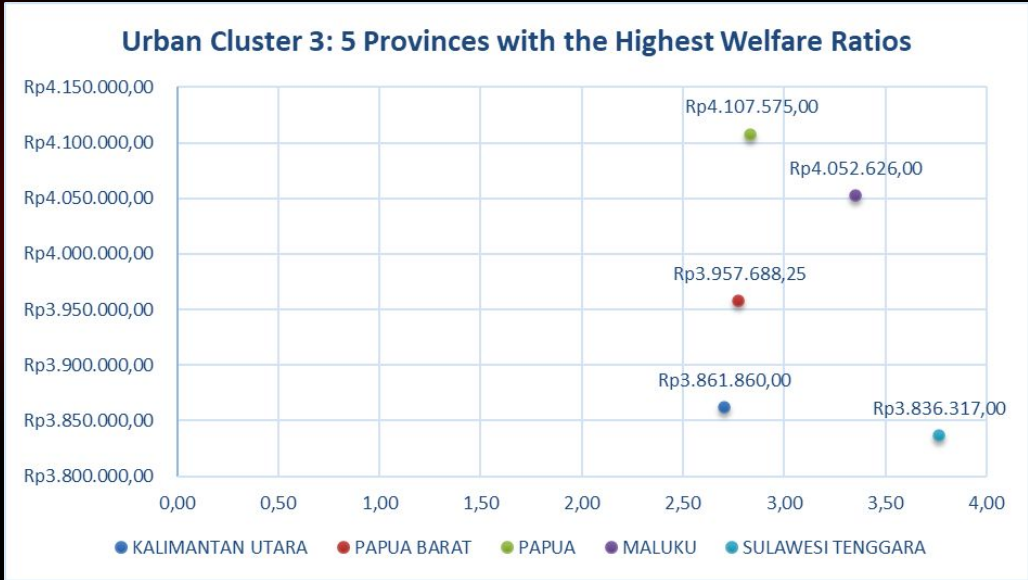
# Cluster 2




Cluster 2 represents the five provinces with medium welfare levels.

The welfare ratios in this cluster are higher than those in Cluster 1 but not sufficient to place them in Cluster 3. As a transitional group, several provinces display unique indicator combinations—for instance, a moderately strong welfare ratio paired with a high poverty gap, or vice versa. These mixed characteristics cause certain provinces to closely resemble those in other clusters, and in some cases, appear to overlap between two clusters due to their proximity to the decision boundary.

# Cluster 3




The graph highlights the five provinces with the highest welfare levels.

Provinces within Cluster 3 show stronger income capacity relative to living costs, indicating a more comfortable level of economic security. Several provinces in this cluster also exhibit a lower poverty gap, suggesting that vulnerable groups are not significantly distant from the poverty line. However, some provinces display high welfare ratios alongside relatively larger poverty gaps, indicating that despite overall strong welfare performance, disparities still persist within certain populations.

# Executive Summary

## Where is the Best Place to Work in Indonesia?

# Key Insight

This analysis identifies the provinces that offer the strongest working conditions in Indonesia by evaluating two key indicators:

- Welfare Ratio — the ability of income to cover basic living needs.
- Poverty Gap — how far vulnerable groups remain from the poverty line.

Using these metrics, a K-Means clustering approach was applied to group provinces into distinct welfare categories.

# Clustering Results

1. **Cluster 3 (High Welfare)**
   Provinces with the best overall work conditions—income sufficiently covers living costs and poverty vulnerability is relatively low.
2. **Cluster 2 (Medium Welfare)**
   Transitional provinces showing mixed patterns. Some have moderate welfare ratios but larger poverty gaps, leading to overlap with other clusters.
3. **Cluster 1 (Low Welfare)**
   Provinces where workers face the highest welfare pressure due to low income adequacy and higher vulnerability to poverty.

# Overlap & Variability

- Several provinces appear near or across cluster boundaries because the welfare ratio and poverty gap are not perfectly correlated.
- Similar income levels can lead to very different welfare pressures depending on living costs and poverty conditions.
- This explains why some provinces are visually close to more than one cluster.

# Conclusion

The question “Where is the Best Place to Work in Indonesia?” was explored using analytical methods and clustering. Cluster 3 highlights provinces where workers are better able to meet their basic needs with lower economic vulnerability. Among them, Southeast Sulawesi ranks highest based on the welfare ratio for both rural and urban areas. This is largely because workers’ monthly wages are recorded to be around three times higher than their average expenditure.

Therefore, Southeast Sulawesi can be considered one of the best provinces to work in due to its strong income–expenditure balance. However, this finding should be interpreted carefully, as a high welfare ratio does not necessarily indicate low poverty levels. For that reason, both income adequacy and poverty rates should be evaluated together when assessing the overall quality of working conditions.