Temporal analysis of fuzzy c-means clustering for global data between 2020 and 2024

Gherasim Delia-Catalina

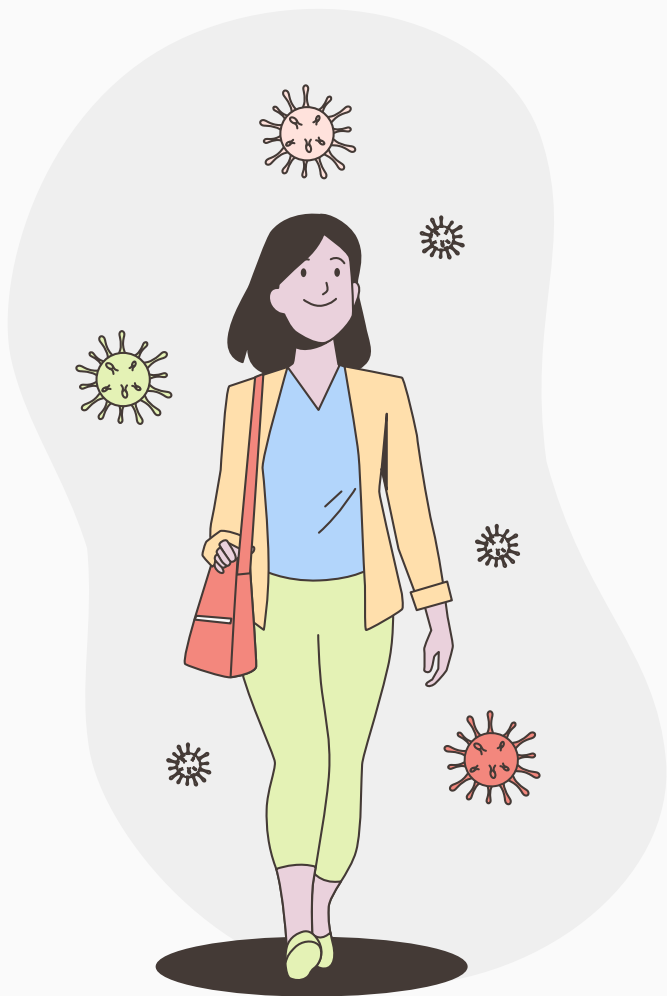Applied Computational Intelligence

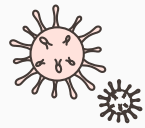Group 246-1

01

Introduction

Analyze the effects of the pandemic

Fuzzy c-means clustering

Four different datasets that contain global time-series information

Identify shared patterns and gradual transitions between countries and time periods

Fuzzy C-Means Clustering

The main goal of FCM is to find the best cluster centers and memberships that minimize the overall distance between data points and cluster centers. The algorithm starts with random assigning and updates the cluster centers and memberships until the changes are insignificant, resulting in overlapping clusters that reflect fuzzy boundaries between groups in the data.

Datasets
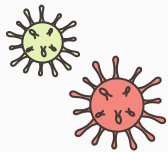
https://www.kaggle.com/datasets

Daily Covid-19 Data

World Health Organization
Daily reported Covid-19 cases and deaths by country

Global Logistics and economic indicators

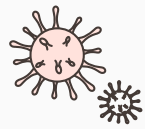Daily shipping rates, brent oil prices and global supply chain pressure index

Layoffs

Data scrapped from Layoffs.fyi
Global layoff events: date, company, location, laid off count, country

World Happiness Report

UN Sustainable Development Solutions Network
Country level happiness ranking and scores by year: 3social support, freedom, generosity

03 Data Preprocessing

```
RangeIndex: 3642 entries, 0 to 3641
Data columns (total 12 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   Company                 3642 non-null    object
 1   Location_HQ             3642 non-null    object
 2   Industry                3642 non-null    object
 3   Laid_Off_Count          2389 non-null    float64
 4   Date                    3642 non-null    object
 5   Source                  3642 non-null    object
 6   Funds_Raised            3252 non-null    float64
 7   Stage                   3642 non-null    object
 8   Date_Added              3642 non-null    object
 9   Country                 3642 non-null    object
 10  Percentage              2342 non-null    float64
 11  List_of_Employees_Laid_Off  3642 non-null  object
dtypes: float64(3), object(9)
memory usage: 341.6+ KB
```

Initial Layoffs dataset

- Min-Max Scaling

- One-Hot Encoding

- Fill missing values with column average

- Monthly aggregation

- Feature selection

```
RangeIndex: 57840 entries, 0 to 57839
Data columns (total 8 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   Date_reported       57840 non-null   object
 1   Country_code        57599 non-null   object
 2   Country             57840 non-null   object
 3   WHO_region          53502 non-null   object
 4   New_cases           39028 non-null   float64
 5   Cumulative_cases    57840 non-null   int64
 6   New_deaths          25001 non-null   float64
 7   Cumulative_deaths   57840 non-null   int64
dtypes: float64(2), int64(2), object(4)
memory usage: 3.5+ MB
```

Initial Covid-19 dataset

Covid-19 dataset:

- death and case growth rates

- eliminated outliers using the Interquartile Range method
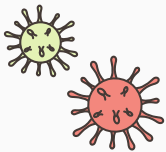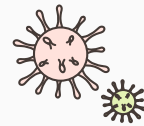
Data analysis

1st

Fuzzy clustering

scikit-fuzzy
Maximum iterations
1000
Convergence
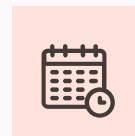tolerance 0.005

2nd

FPC score heatmap

Fuzzification
coefficient
from 1.5 to 2
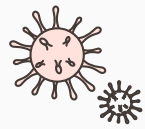number of clusters
from 3 to 7

3rd

Cluster analysis

Best model
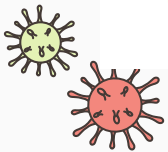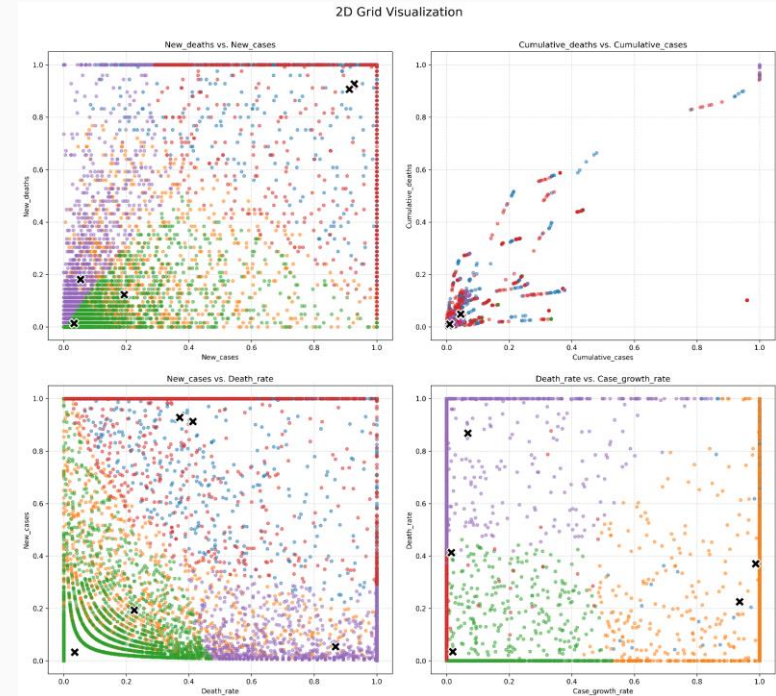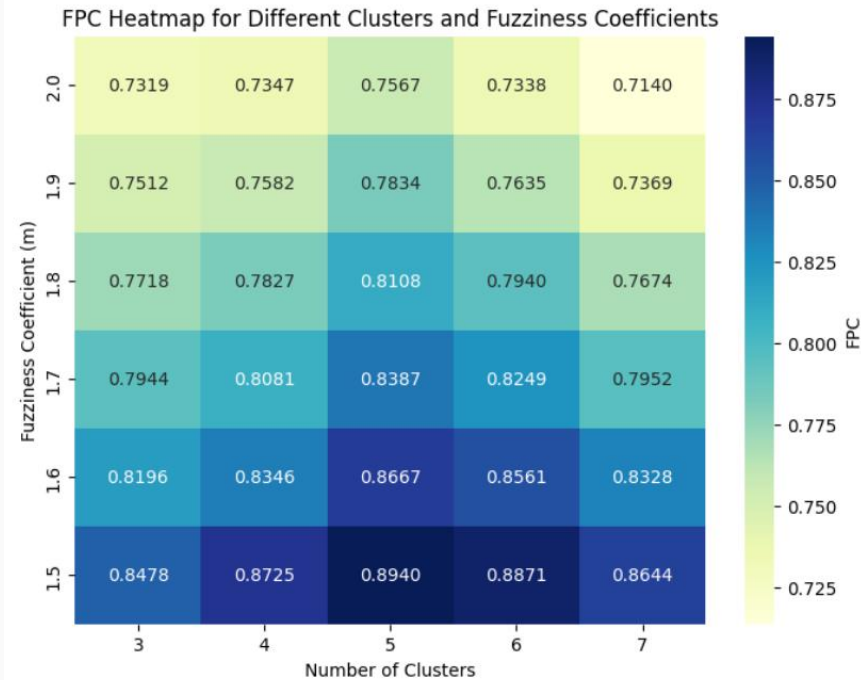Cluster average for
each column
Grid visualization

4th
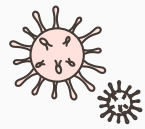
Timeline visualization
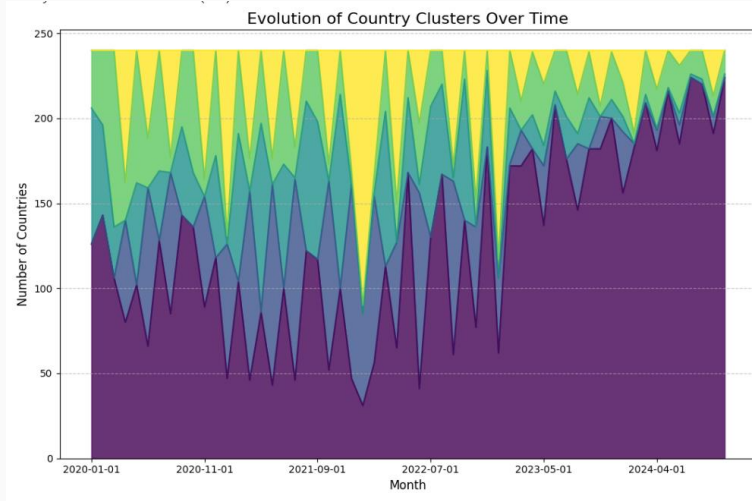
Heatmap or graph of
cluster evolution
over time

Daily COVID-19 Data

Daily COVID-19 Data

Evolution of Country Clusters Over Time

Cluster 0	
Country	
Russian Federation	0.443354
Italy	0.437393
Canada	0.429833
Portugal	0.425798
Chile	0.417501
Greece	0.413916
Poland	0.408755
United Kingdom of Great Britain and Northern Ireland	0.407794
Romania	0.385618
Australia	0.374610

Cluster 1	
Country	
Luxembourg	0.354151
Brunei Darussalam	0.344593
Afghanistan	0.300989
French Guiana	0.287949
Malta	0.282227
Qatar	0.276877
Zambia	0.262787
Democratic Republic of the Congo	0.260877
Maldives	0.258937
Bahrain	0.257842

Cluster 2	
Country	
Tokelau	0.999974
International commercial vessel	0.999973
Holy See	0.999973
Pitcairn	0.999973
International conveyance (Vanuatu)	0.999973
International conveyance (Solomon Islands)	0.999973
International conveyance (American Samoa)	0.999973
International conveyance (Kiribati)	0.999973
Democratic People's Republic of Korea	0.999973
Turkmenistan	0.999973

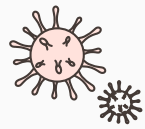| Cluster 3	
Country	
Russian Federation	0.526068
United Kingdom of Great Britain and Northern Ireland	0.440357
United States of America	0.433047
Mexico	0.430808
Italy	0.423423
Brazil	0.423270
Spain	0.420380
Chile	0.419795
Peru	0.413104
India	0.404771

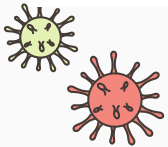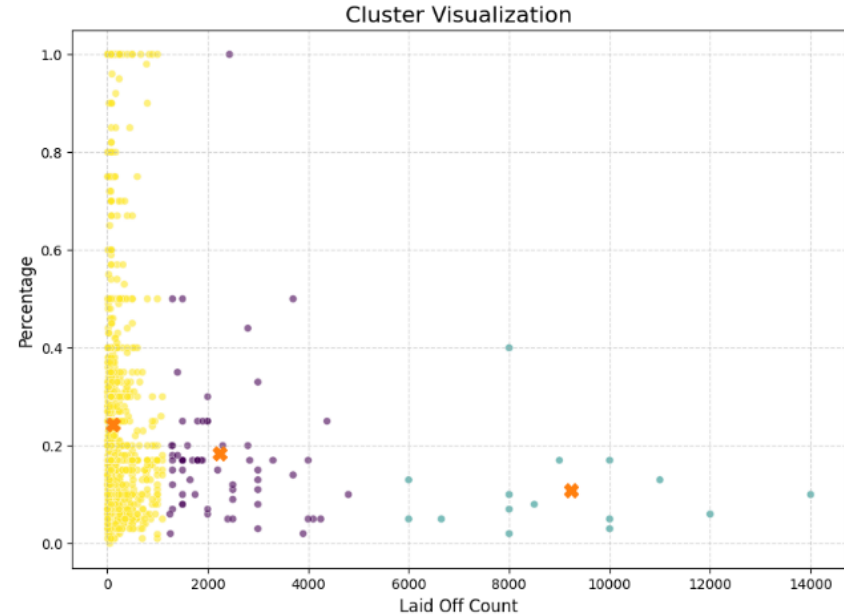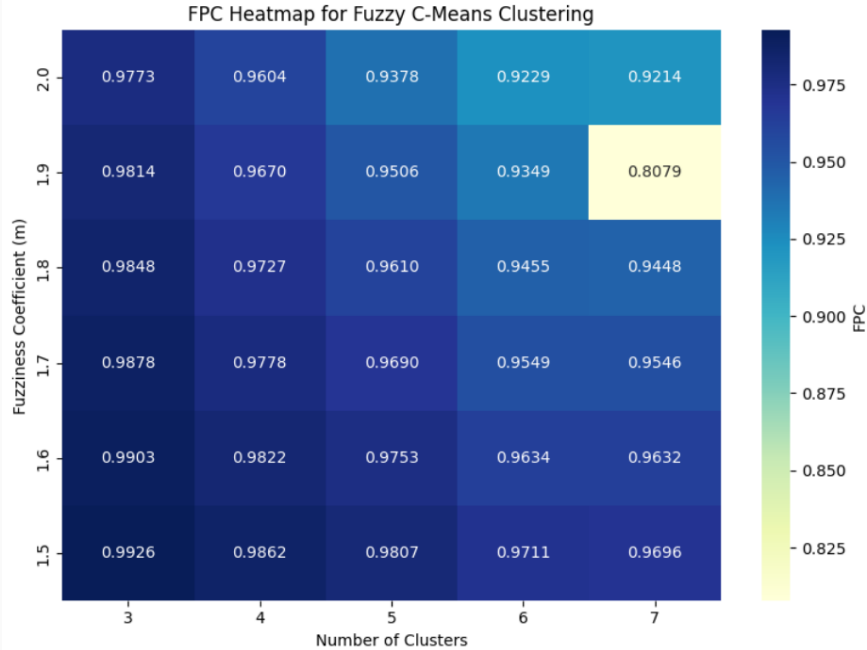| Cluster 4	
Country	
Yemen	0.470787
Jamaica	0.454516
Sudan	0.424624
Bosnia and Herzegovina	0.412714
Mali	0.399018
Guyana	0.377716
Sri Lanka	0.375112
Zimbabwe	0.373805
Slovenia	0.347362
Chad	0.339099

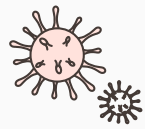|	New_cases	New_deaths	Cumulative_cases	Cumulative_deaths	Death_rate	Case_growth_rate
0	0.928115	0.926707	0.044997	0.046168	0.370545	0.987352
1	0.193172	0.122852	0.010804	0.009310	0.225573	0.936438
2	0.032929	0.013811	0.007142	0.005661	0.034895	0.019836
3	0.912638	0.906692	0.045043	0.048734	0.412658	0.015018
4	0.053884	0.181472	0.009647	0.010483	0.868594	0.067986

Layoffs

Layoffs

Cluster 0:

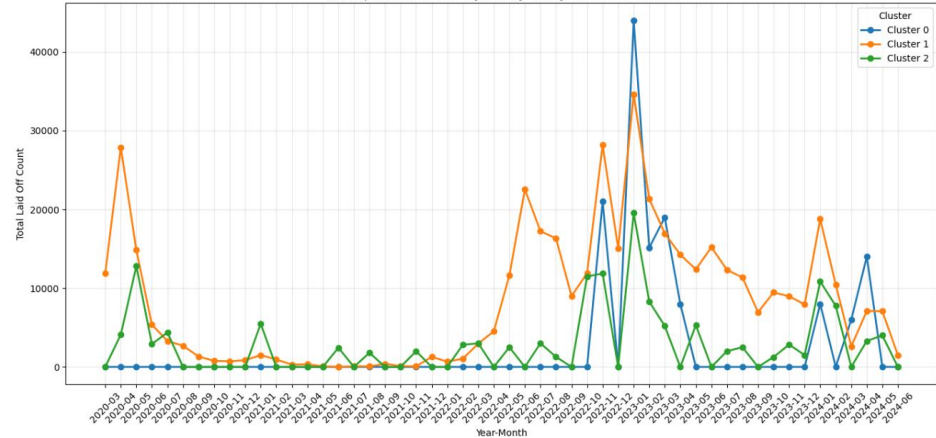|	Company	Location_HQ	Industry	Laid_Off_Count	Country
1201	Amazon	Seattle	Retail	9000.0	United States
1576	Microsoft	Seattle	Other	10000.0	United States
1218	Meta	SF Bay Area	Consumer	10000.0	United States
1959	Amazon	Seattle	Retail	10000.0	United States
1288	Ericsson	Stockholm	Other	8500.0	Sweden

Cluster 1:

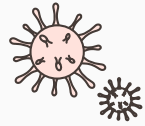|	Company	Location_HQ	Industry	Laid_Off_Count
1172	Blue Nile	Seattle	Retail	119.0
2952	Homie	Salt Lake City	Real Estate	119.0
498	Nomad Health	New York City	Healthcare	119.0
1426	Miro	SF Bay Area	Other	119.0
2687	MasterClass	SF Bay Area	Education	120.0

Cluster 2:

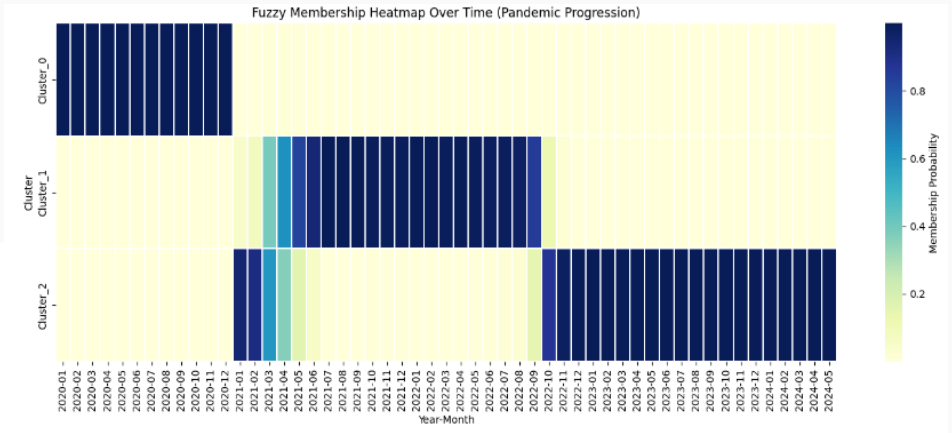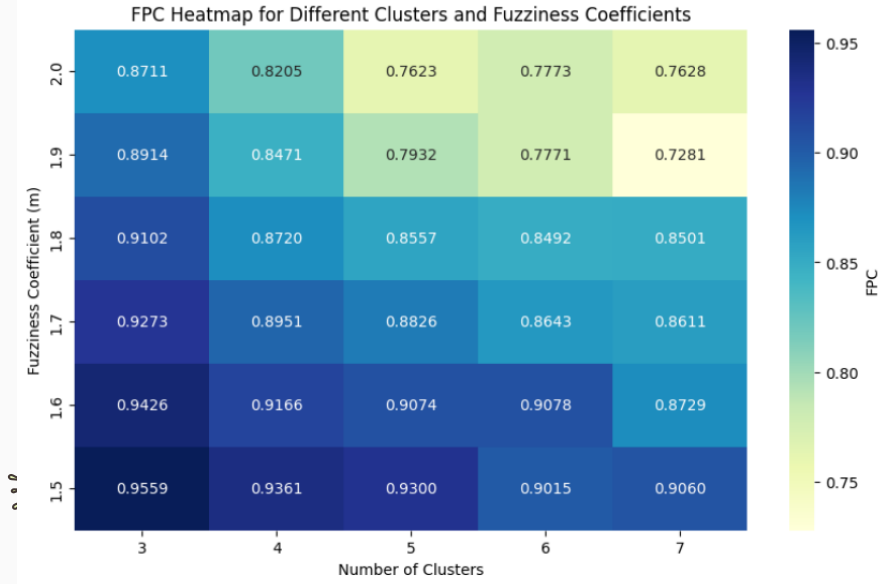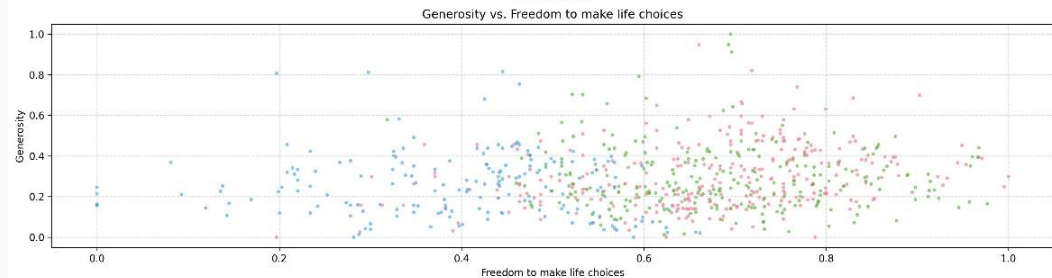|	Company	Location_HQ	Industry	Laid_Off_Count	Country
1183	Indeed	Austin	HR	2200.0	United States
1030	Shopify	Ottawa	Retail	2300.0	Canada
1325	Micron	Boise	Hardware	2400.0	United States
2979	Katerra	SF Bay Area	Construction	2434.0	United States
2878	Carvana	Phoenix	Transportation	2500.0	United States

Global logistics and economic indicators

Fuzzy Membership Heatmap Over Time (Pandemic Progression)

FPC Heatmap for Different Clusters and Fuzziness Coefficients
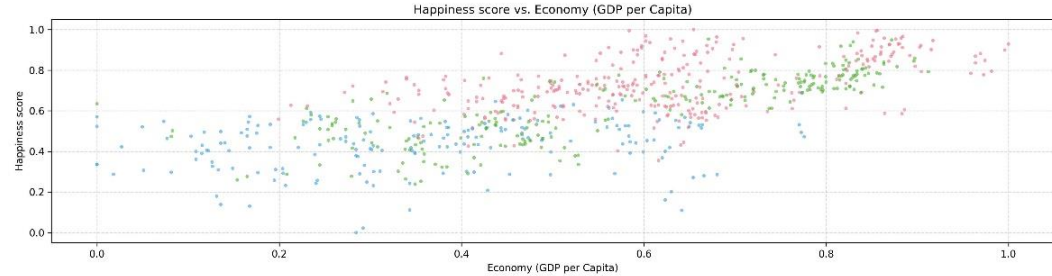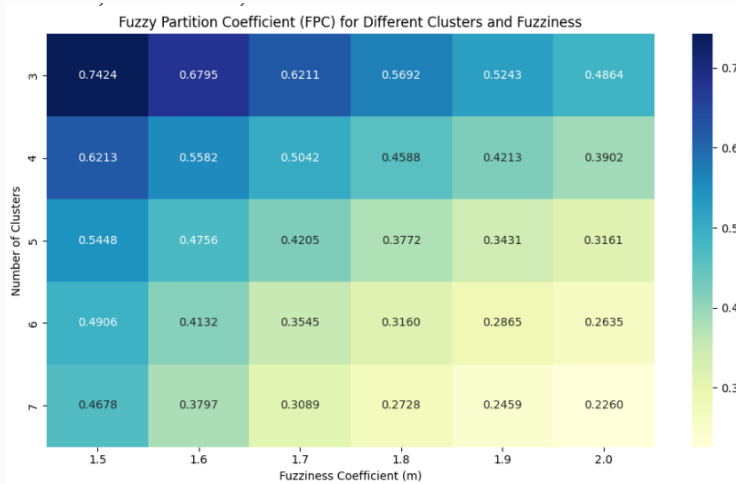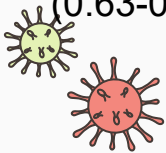
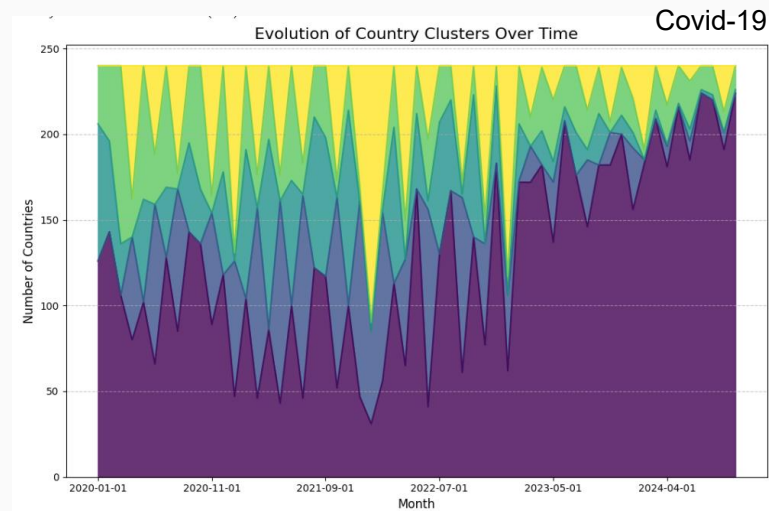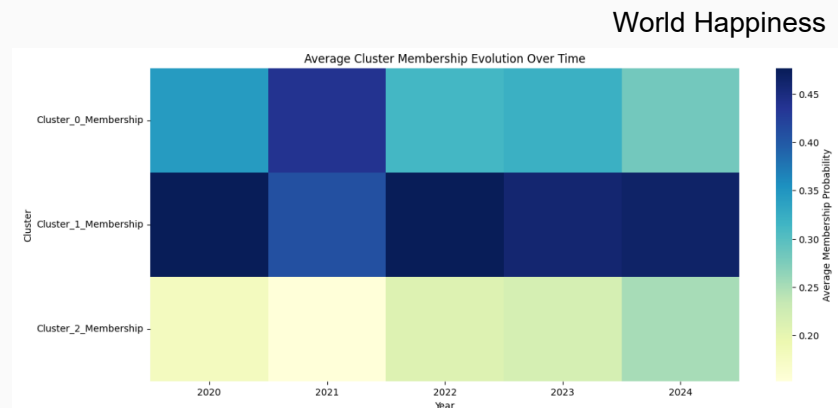| Cluster	Global_Container _Index_FBX	China_US_West _Coast_Route	US_West_ Coast_China _Route	China_North _Europe_ Route	North_ Europe_China _Route	Brent_Oil _Price_USD
0	1920.211371	2386.231183	452.000000	1960.242025	476.269789	43.517689
1	8442.407407	9197.437305	518.583029	6495.891875	1498.703015	88.153053
2	2149.753478	2224.580307	519.984841	6482.429607	1498.879883	80.576837

World happiness report

Fuzzy Partition Coefficient (FPC) for Different Clusters and Fuzziness

Happiness score vs. Economy (GDP per Capita)

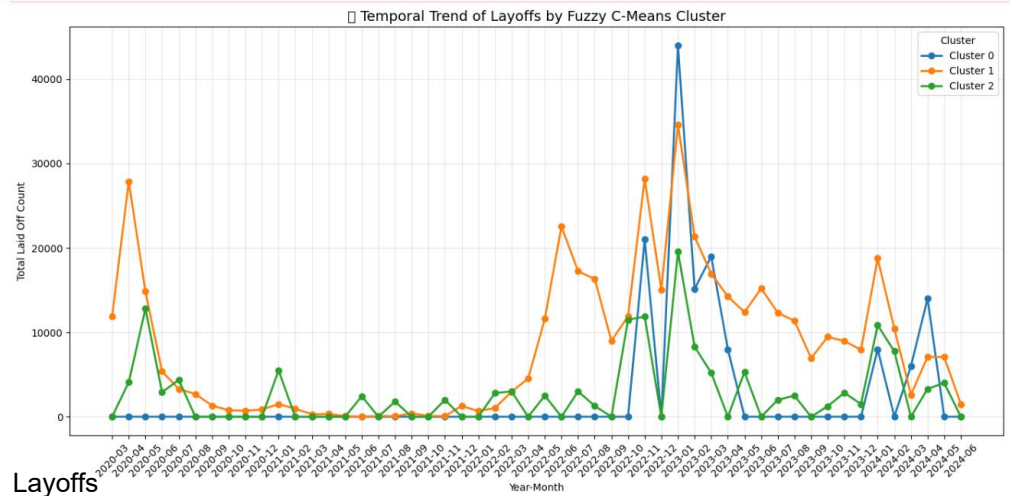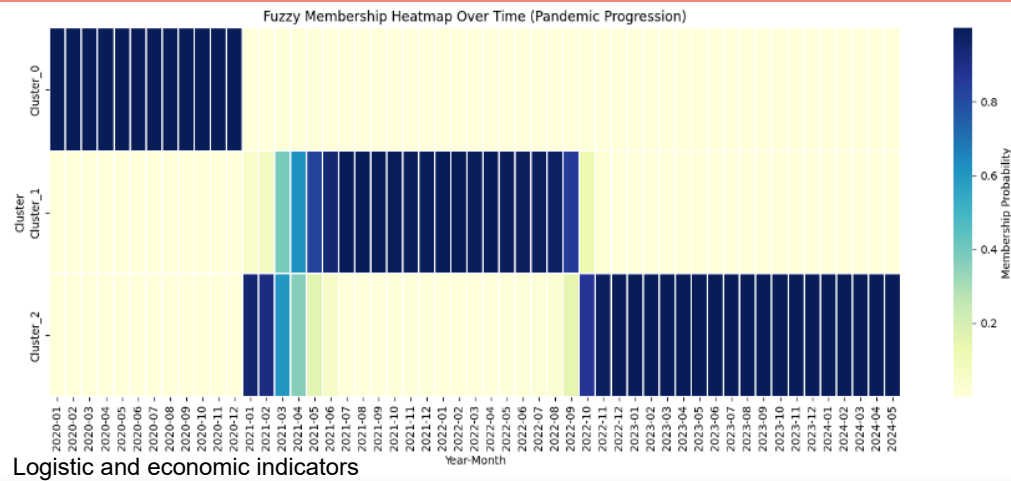Healthy life expectancy vs. Social support

Generosity vs. Freedom to make life choices
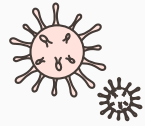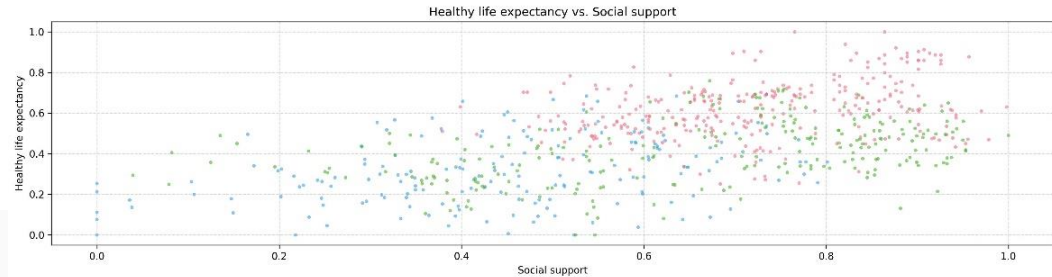
- Lower happiness scores (0.41-0.56): Burkina Faso, Liberia, Togo
- Medium happiness scores (0.63-0.72): Ecuador, Jamaica, Paraguay
- High happiness scores (0.82-0.87): Germany, Canada, Austria

Fuzzy Membership Heatmap Over Time (Pandemic Progression)

World Happiness

Average Cluster Membership Evolution Over Time

Logistic and economic indicators

Temporal Trend of Layoffs by Fuzzy C-Means Cluster

Covid-19

Evolution of Country Clusters Over Time

Layoffs

05

Findings

Covid 19

Captured real world transitions caused by vaccination, virus variants and national policies

Layoffs

Revealed that peak layoffs occurred between Sept 2020 and March 2023

The shift from before the pandemic to the beginning of the virus in December 2020 corresponds with a small wave of layoffs across different industries

The September 2022 pandemic recovery corresponds with multiple layoffs across all sectors

Happiness clusters only show subtle shifts corresponding to peak of the pandemic
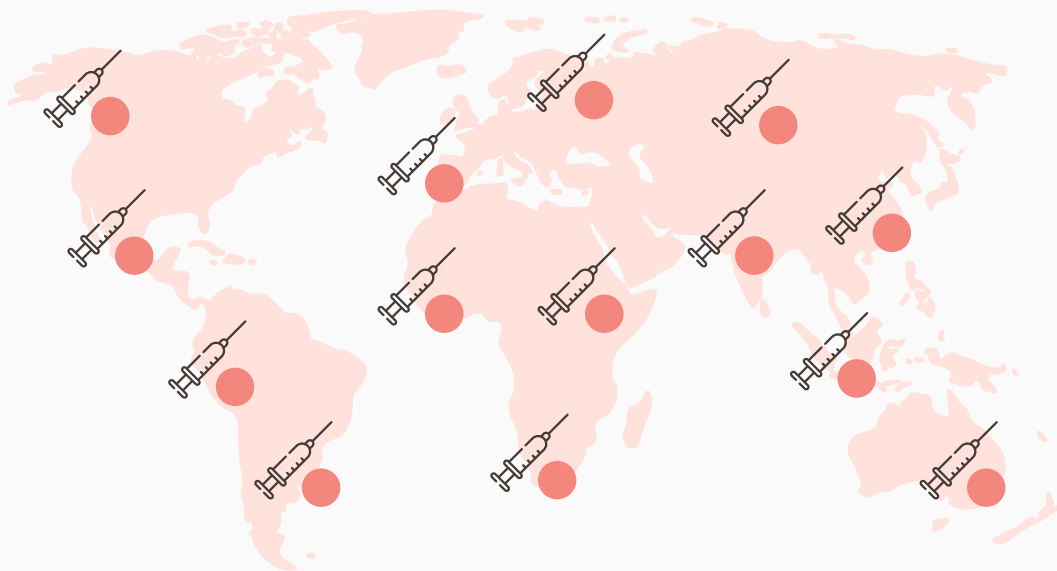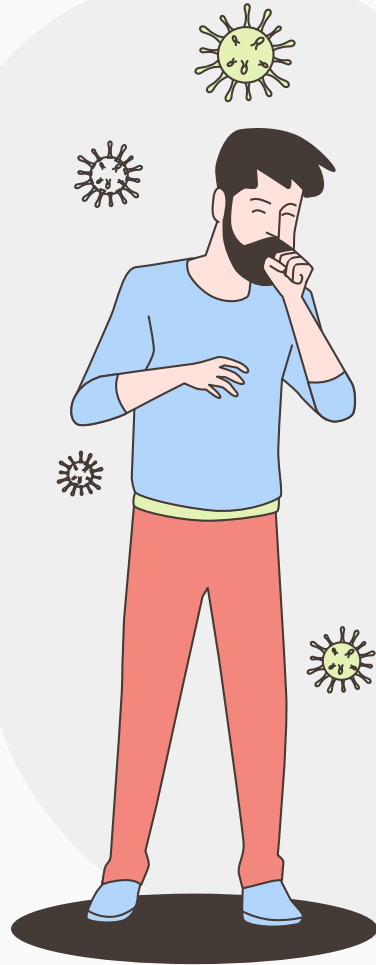
Logistics and economic

Revealed 3 economic stages

World Happiness

The clusters separated countries by socio-economic resources and wellbeing

Conclusions

The fuzzy clustering method captured variations, highlighting distinct pandemic trajectories and overlapping periods, such as the uneven economic recovery and varying pandemic intensities.

Thank you