



DEVELOPMENT AID

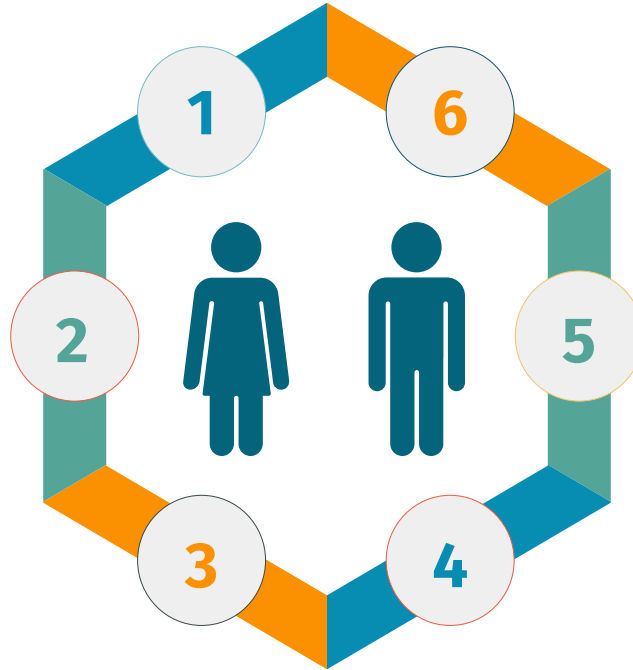
Megan Yit
Institute of Data

OUTLINE

1 PROBLEM
Stakeholders? Business
Context? Data Question?

2 DATA
Sources? Cleaning?

3 DATA EXPLORATION
Child Mortality, Exports, Health,
Imports, Income, Inflation, Life
expectancy, Fertility, GDP



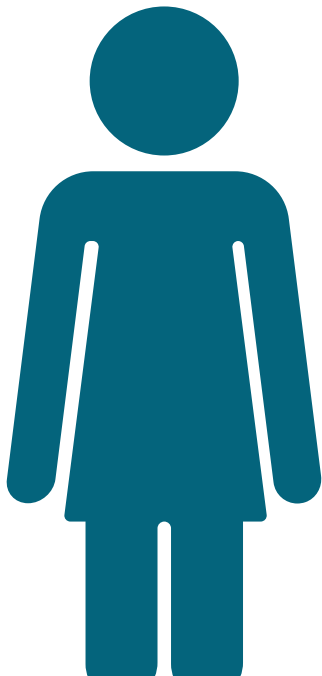
SCALING & DIMENSIONALITY REDUCTION
Normalization vs
Standardization, PCA

CLUSTERING MODELS
Distance-based,
Hierarchy-based, optimal
hyperparameters?

LIMITATIONS

1

PROBLEM



STAKEHOLDERS?

International Charity Organization that raised money for financial aid

BUSINESS QUESTION?

Which countries are in (relatively) more dire need of financial aid?

DATA QUESTION?

Looking at specific socio-economic and health factor as gauges of country development, can Machine Learning identify clusters of “development levels”?

DATA

- COUNTRY
- CHILD MORTALITY
- EXPORTS
- IMPORTS
- HEALTH
- INCOME
- INFLATION
- LIFE EXPECTANCY
- TOTAL FERTILITY
- GDP PER CAPITA



3

DATA EXPLORATION

COUNTRY

CHILD MORTALITY**EXPORTS**

IMPORTS

HEALTH

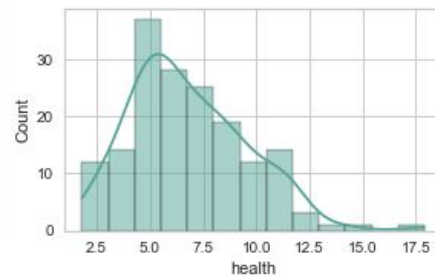
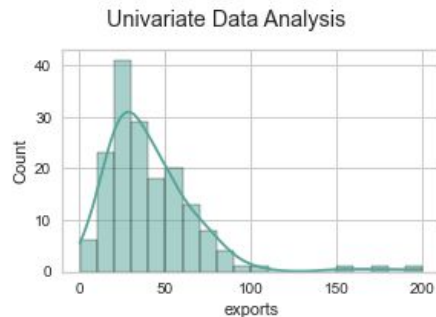
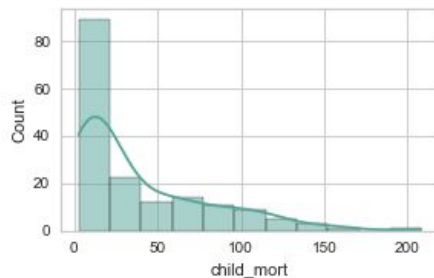
INCOME

INFLATION

LIFE EXPECTANCY

TOTAL FERTILITY

GDP PER CAPITA



3

DATA EXPLORATION

COUNTRY

CHILD MORTALITY**EXPORTS**

IMPORTS

HEALTH

INCOME

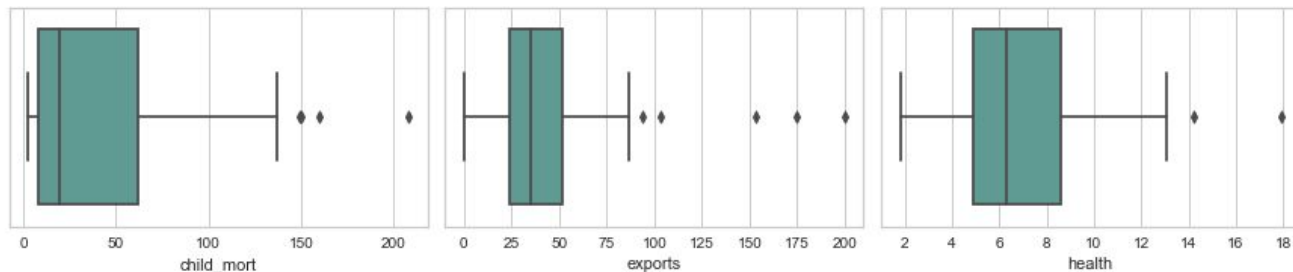
INFLATION

LIFE EXPECTANCY

TOTAL FERTILITY

GDP PER CAPITA

Univariate Data Analysis



3

DATA EXPLORATION

COUNTRY

CHILD MORTALITY

EXPORTS

IMPORTS

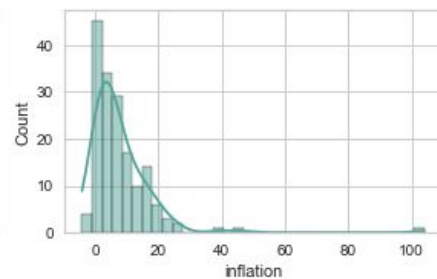
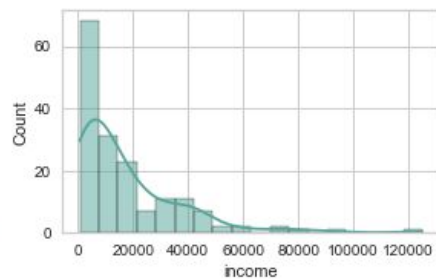
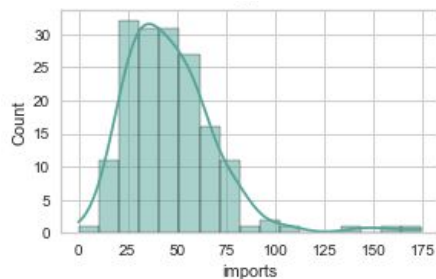
HEALTH

INCOME**INFLATION**

LIFE EXPECTANCY

TOTAL FERTILITY

GDP PER CAPITA



3

DATA EXPLORATION

COUNTRY

CHILD MORTALITY

EXPORTS

IMPORTS

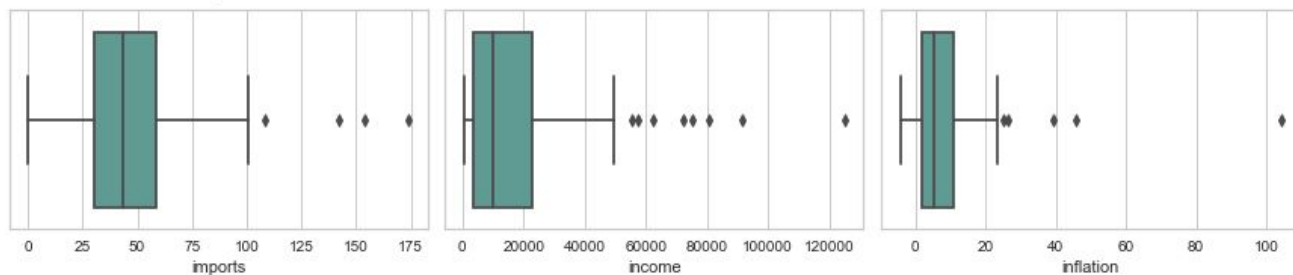
HEALTH

INCOME**INFLATION**

LIFE EXPECTANCY

TOTAL FERTILITY

GDP PER CAPITA



3

DATA EXPLORATION

COUNTRY

CHILD MORTALITY

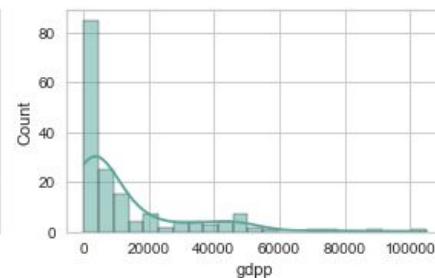
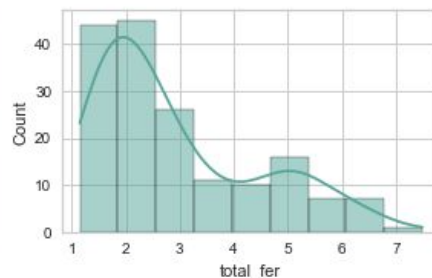
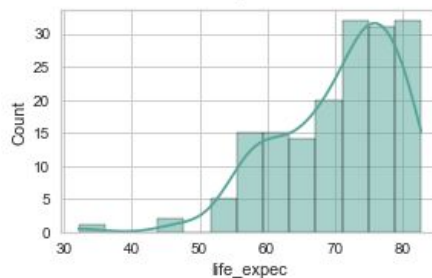
EXPORTS

IMPORTS

HEALTH

INCOME

INFLATION

LIFE EXPECTANCY**TOTAL FERTILITY****GDP PER CAPITA**

3

DATA EXPLORATION

COUNTRY

CHILD MORTALITY

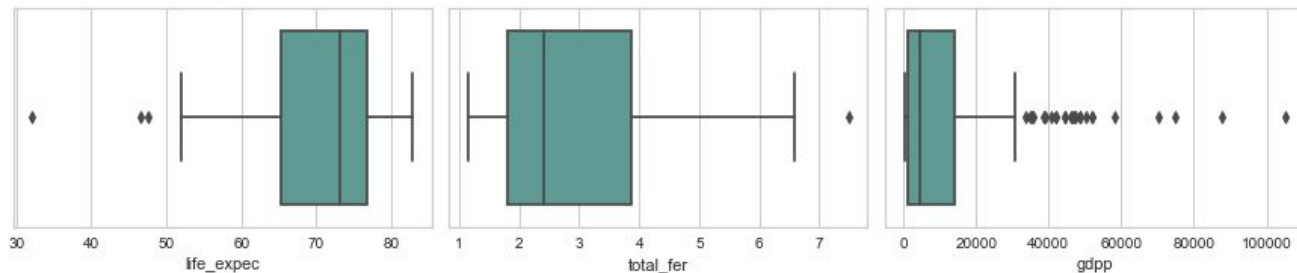
EXPORTS

IMPORTS

HEALTH

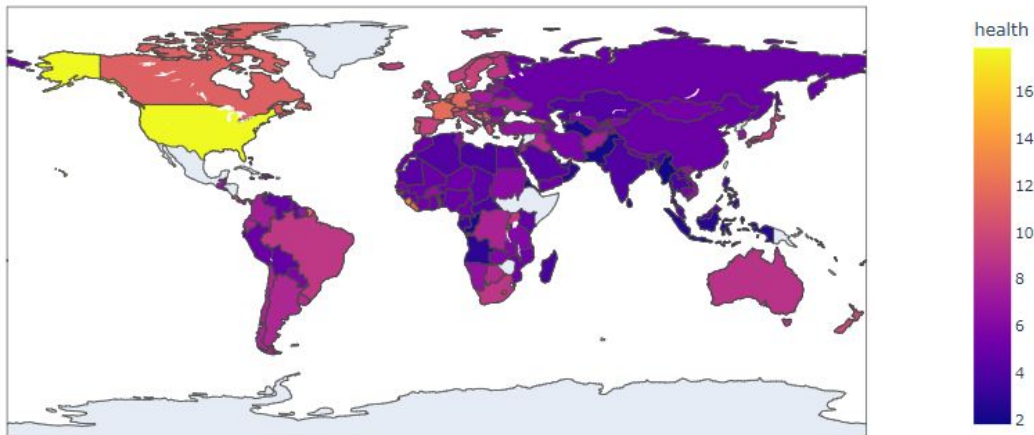
INCOME

INFLATION

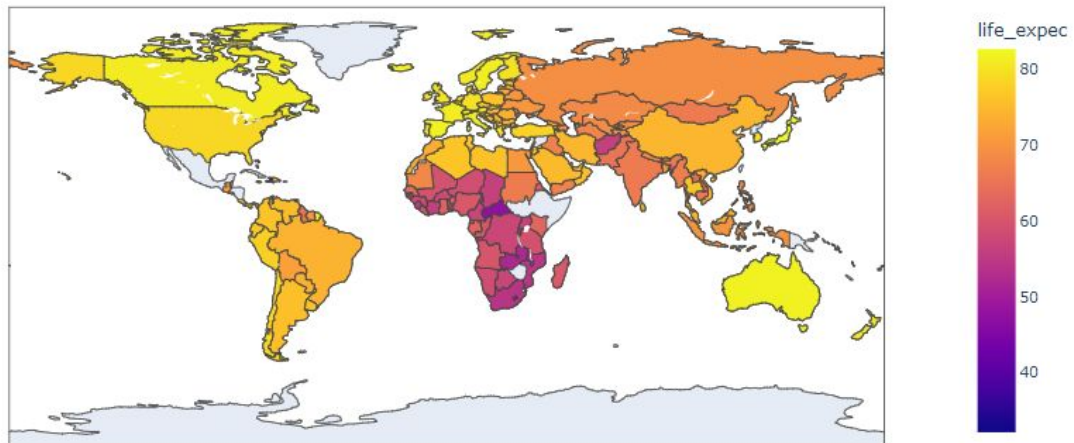
LIFE EXPECTANCY**TOTAL FERTILITY****GDP PER CAPITA**

3

health rate by countries

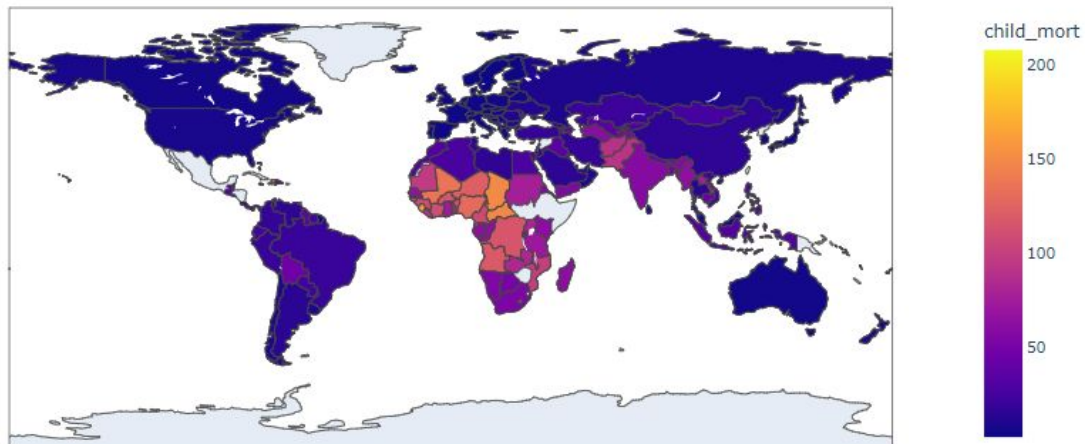


life_expec rate by countries

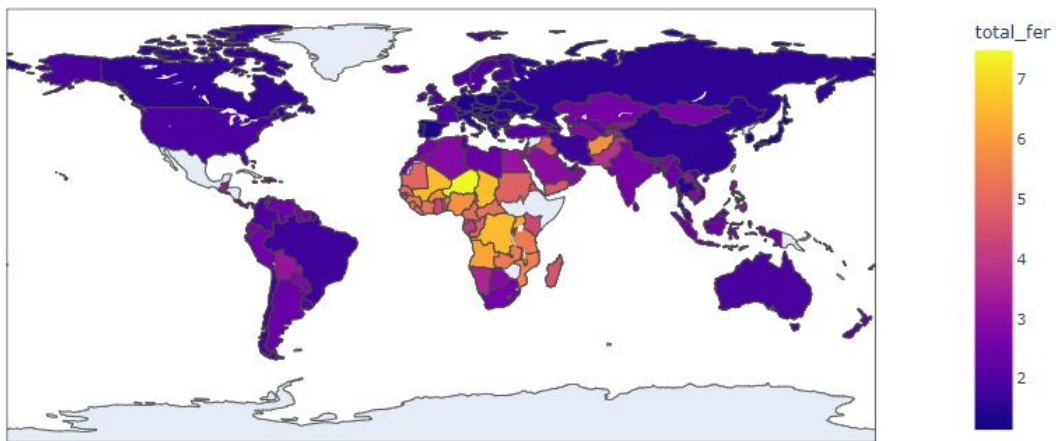


3

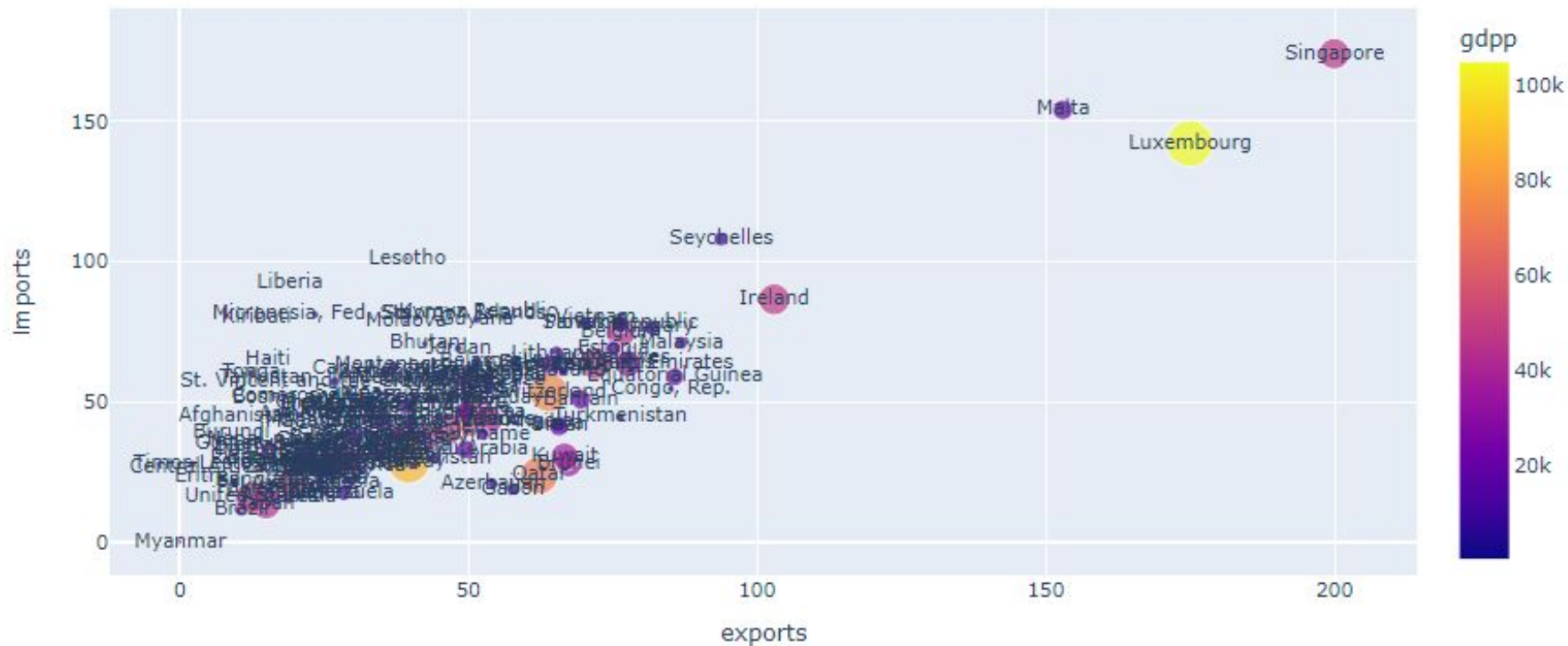
child_mort rate by countries



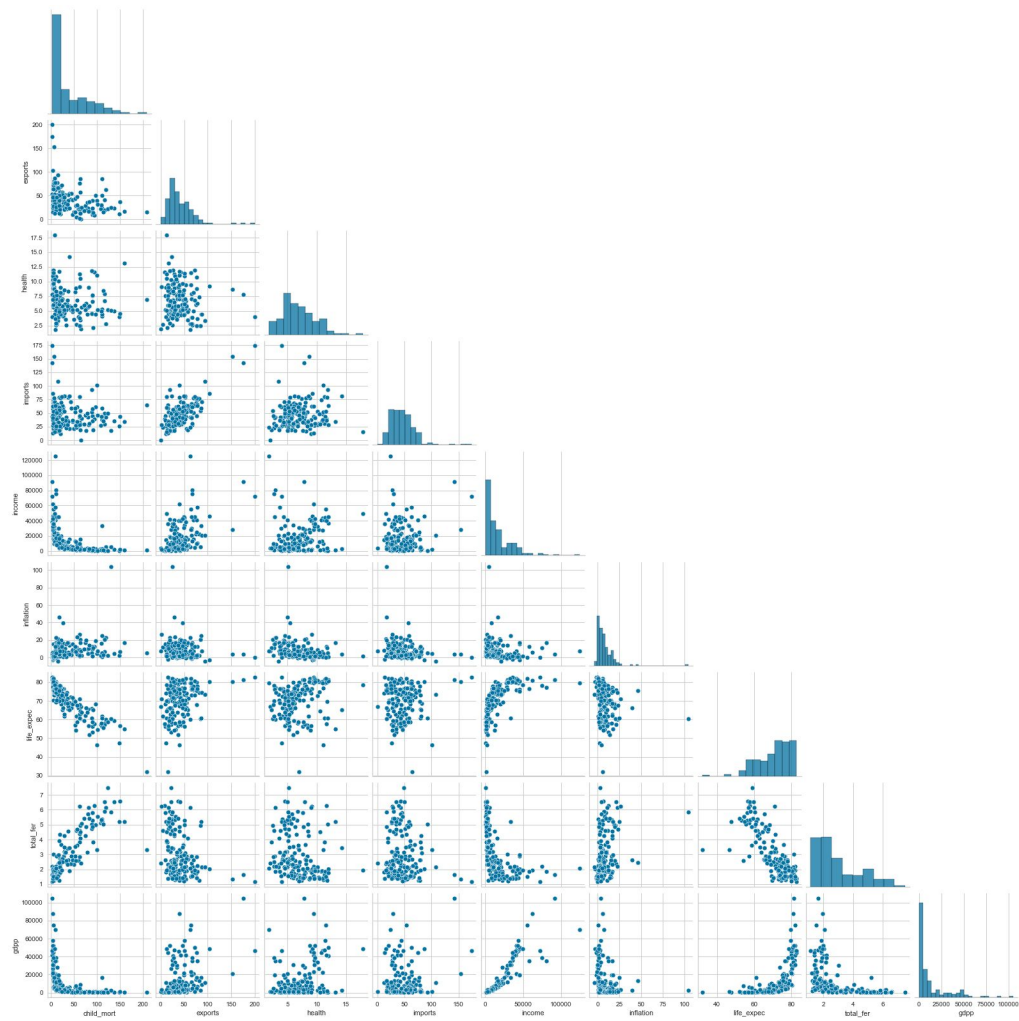
total_fer rate by countries



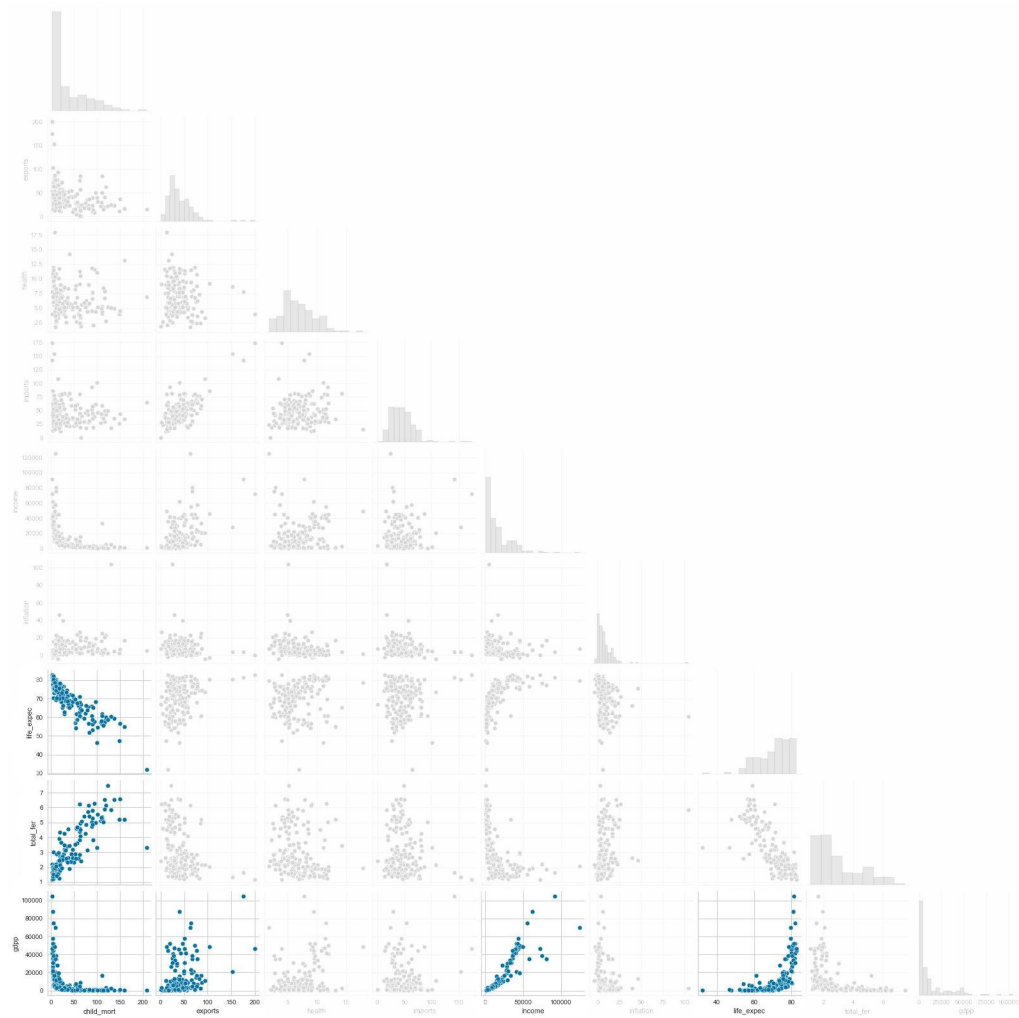
Countries by Export & Import, w corresponding GDP

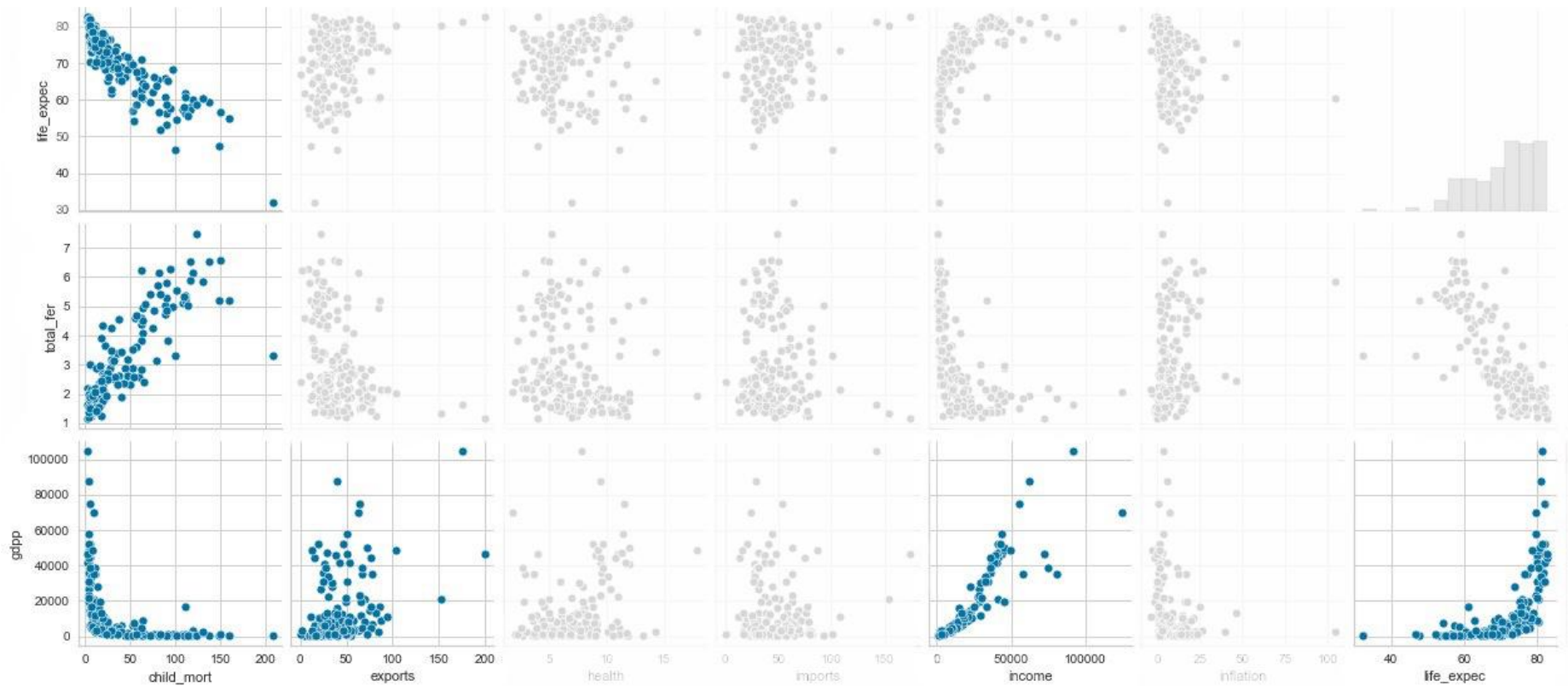


3

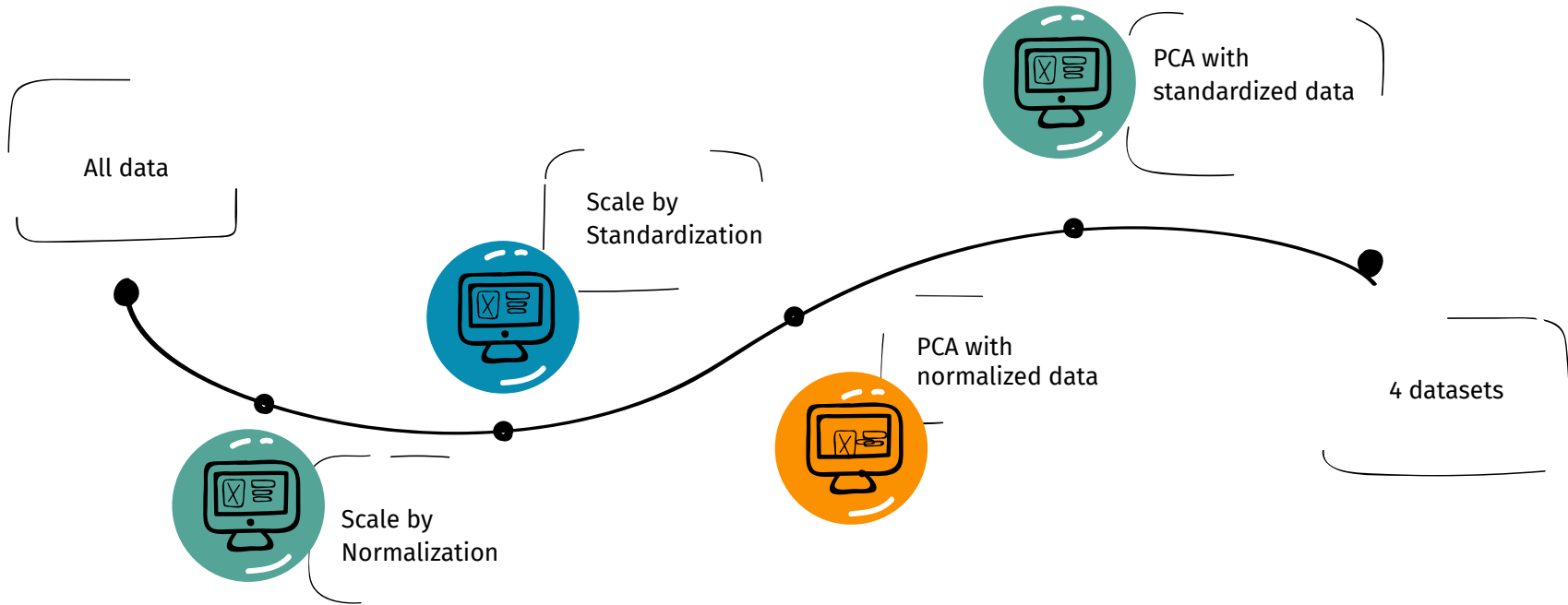


3



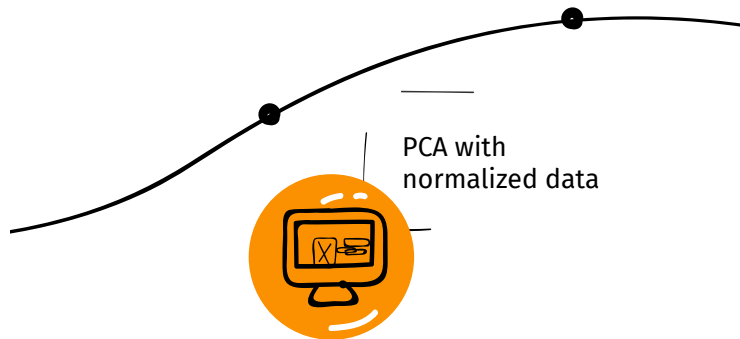


SCALING and DIMENSIONALITY REDUCTION

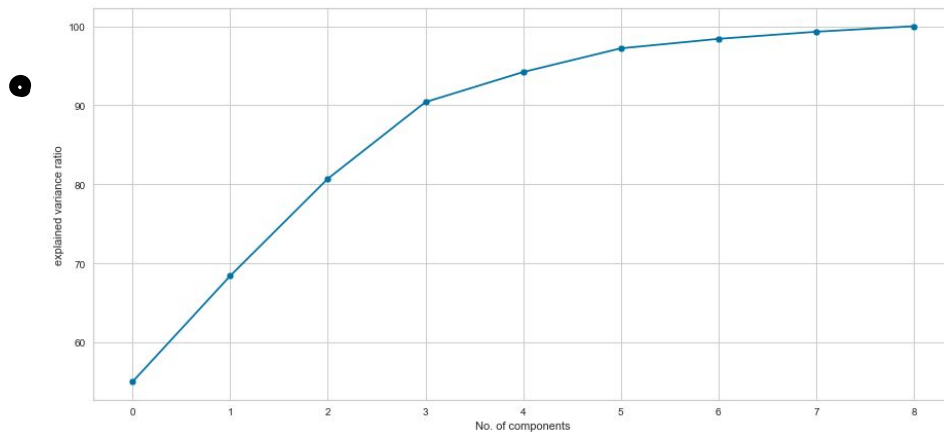
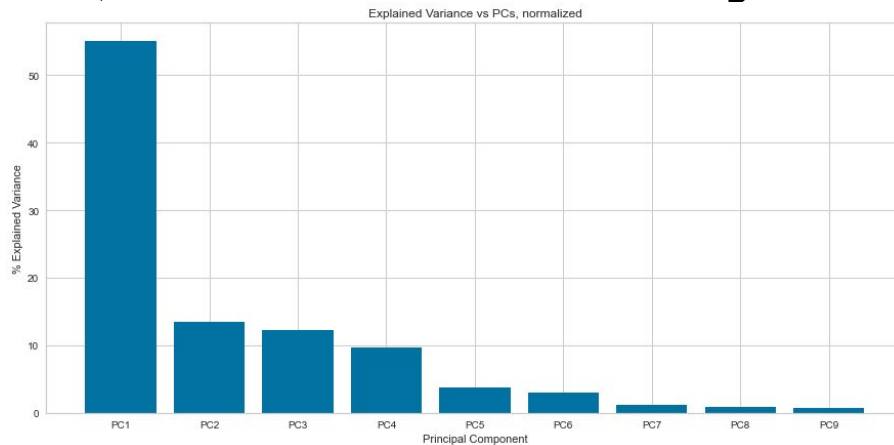


PCA (and dimensionality reduction)

- Reduces number of variables in data set
- Preserves as much information as possible



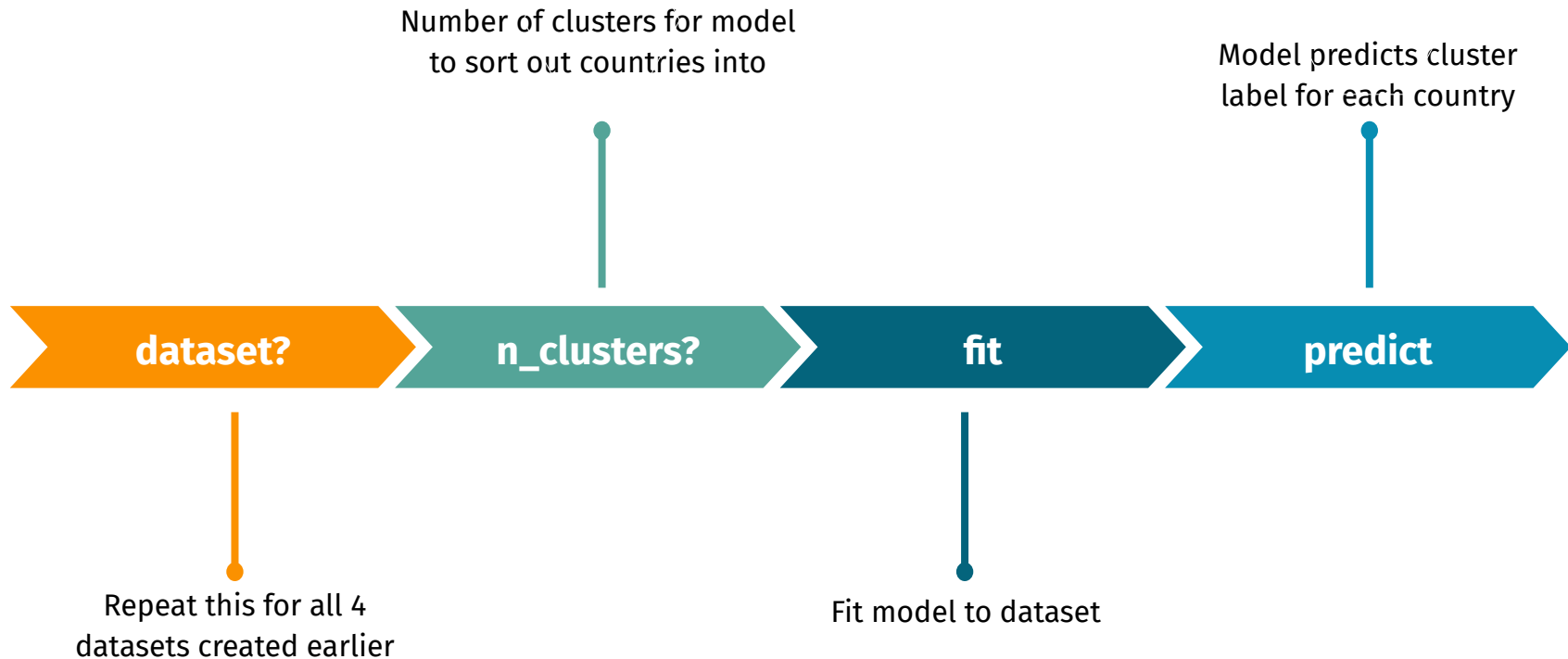
PCA (and dimensionality reduction)



PCA with
normalized data

5

MODEL (KMeans)



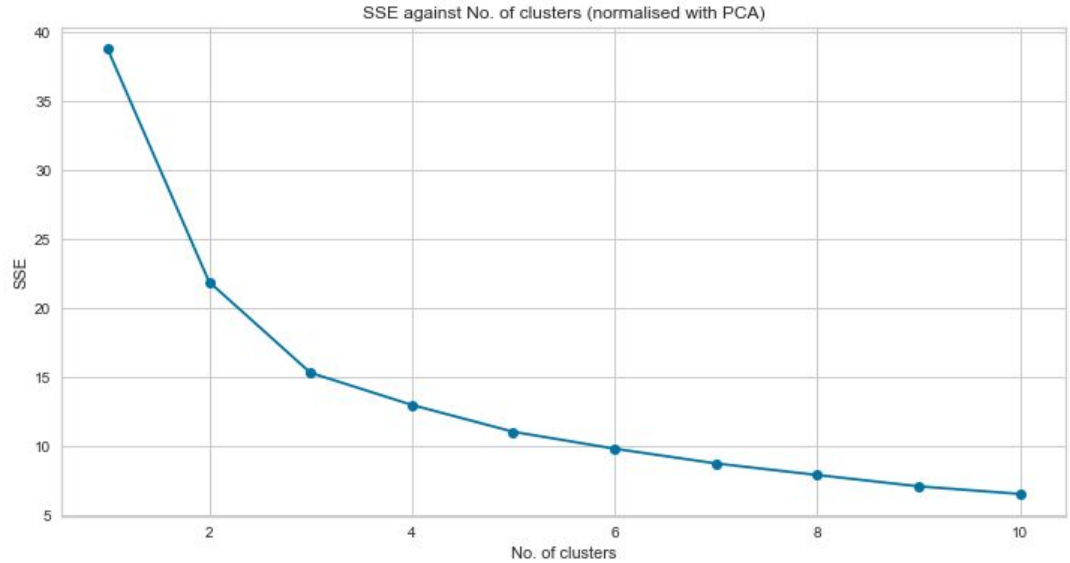
5

MODEL (KMeans)

Find optimal no. of clusters
with **ELBOW METHOD**



n_clusters?

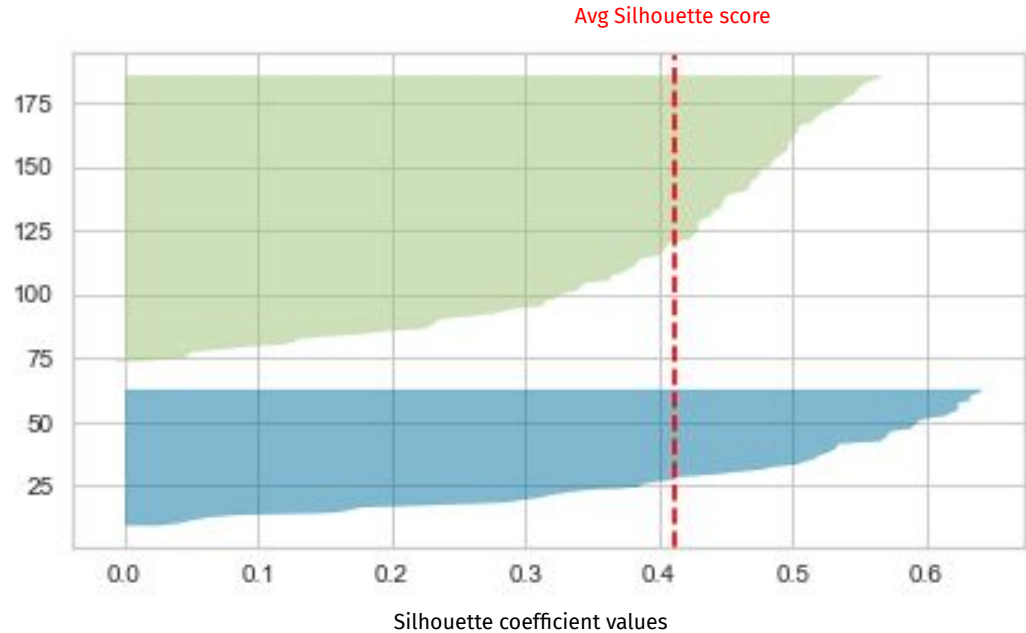


5

MODEL (KMeans)

Find optimal no. of clusters
with **SILHOUETTE METHOD**

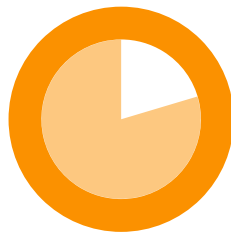
n_clusters?



MODEL (KMeans)

Internal Validation with other
scoring metrics

n_clusters?



**Calinski and
Harabasz score**

n_clusters 2 or 3
dataset: PCA, norm

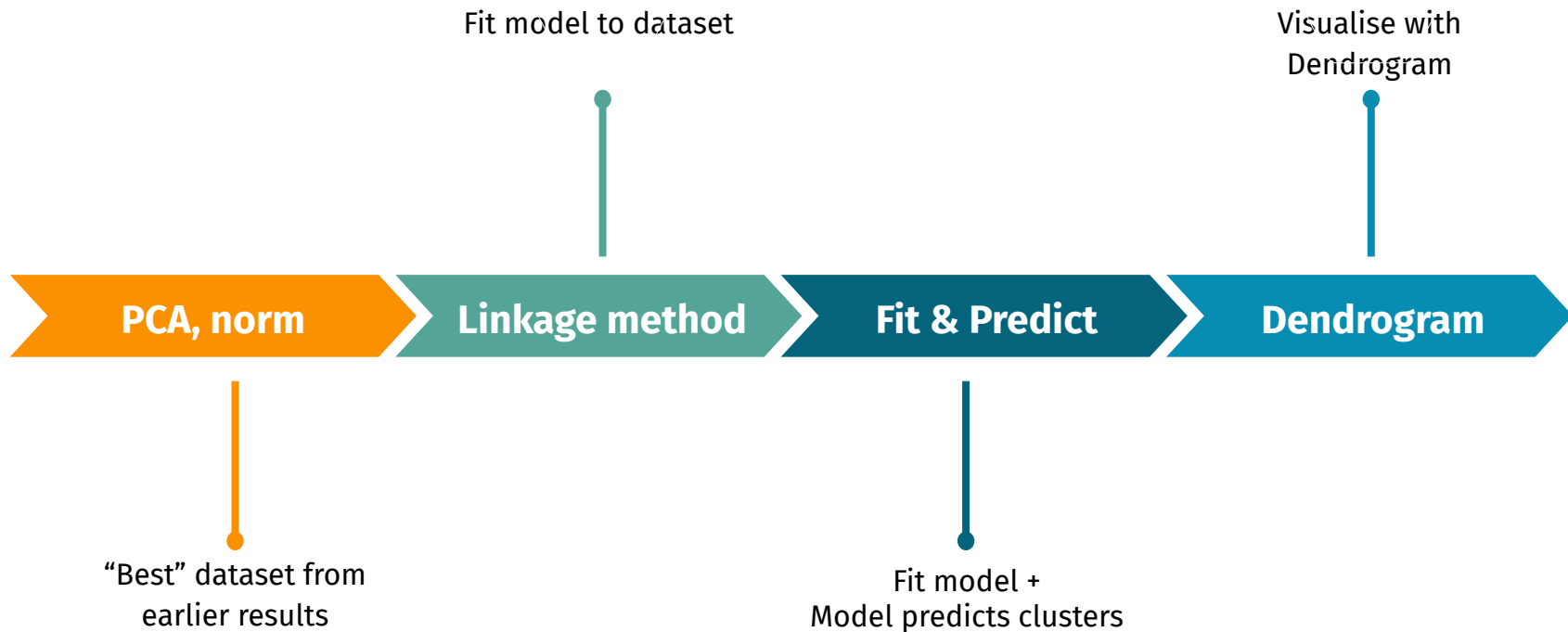


**Davies-Bouldin
score**

n_clusters 2
dataset: norm (w and w/o PCA)

5

MODEL (Hierarchical, Agglomerative)

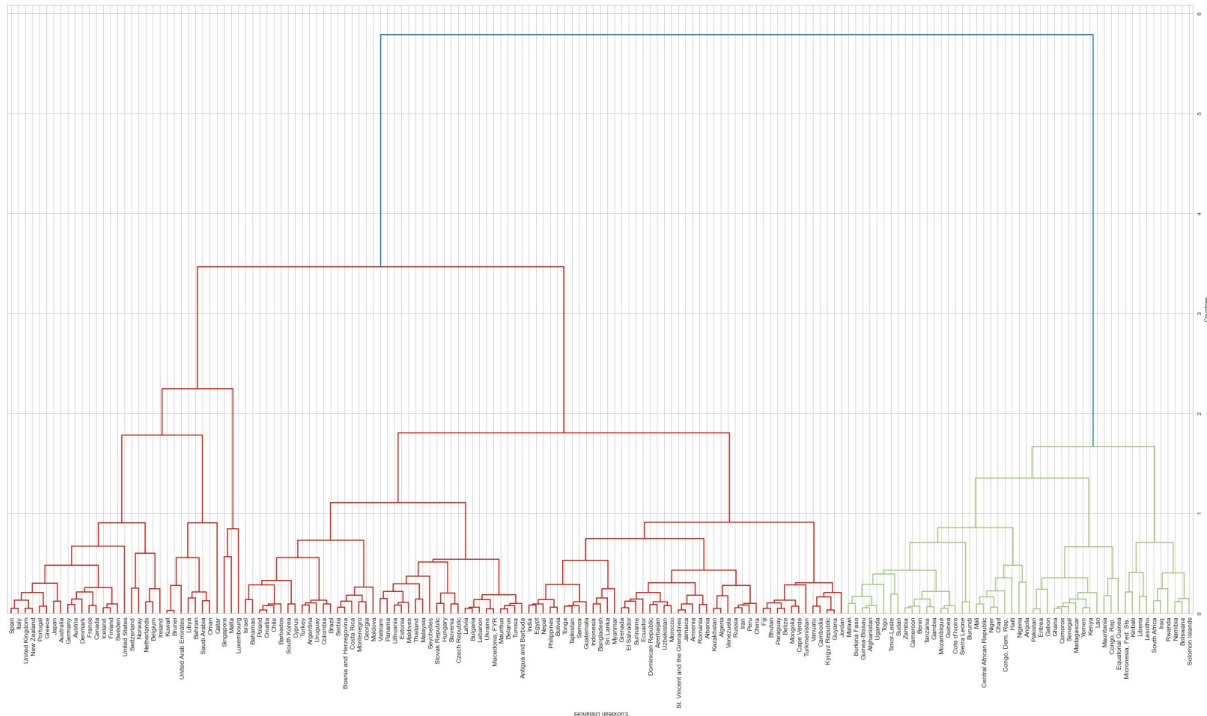


MODEL (Hierarchical, Agglomerative)

Fit model to dataset

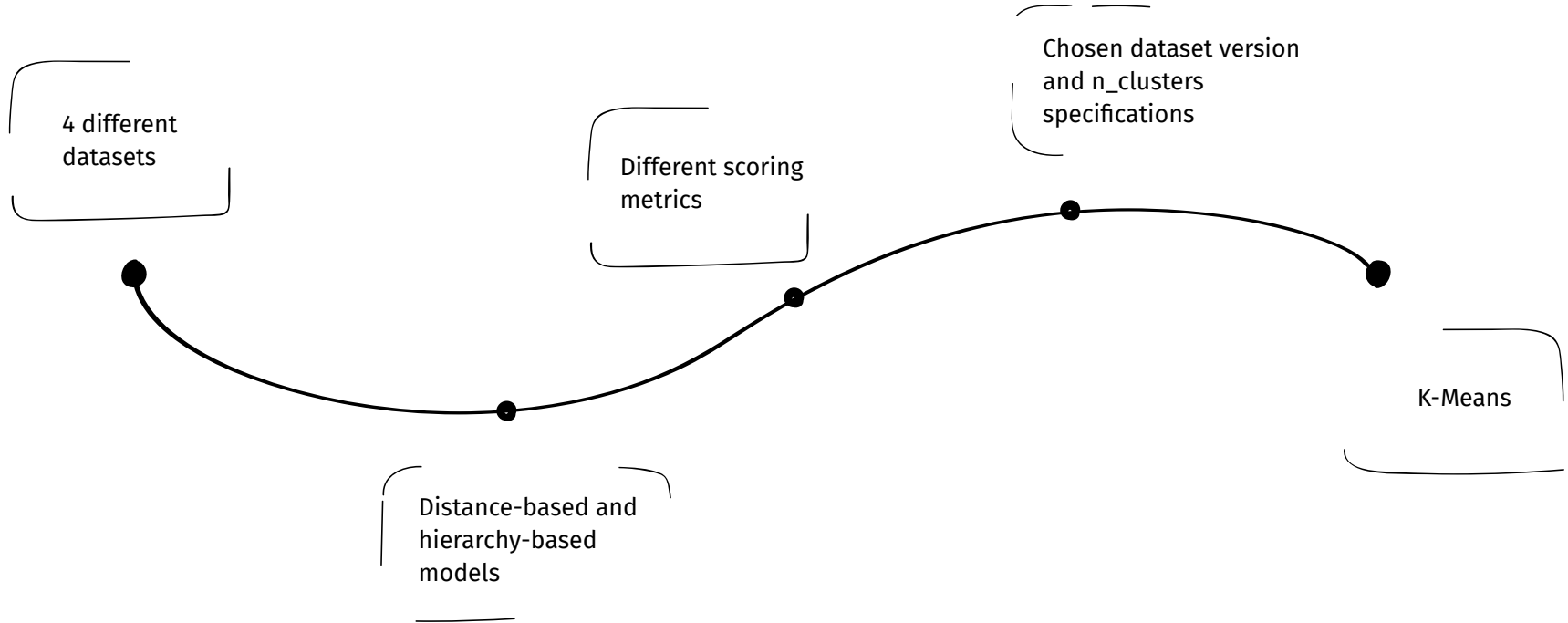
Linkage method

Hierarchical Clustering - Dendrogram



5

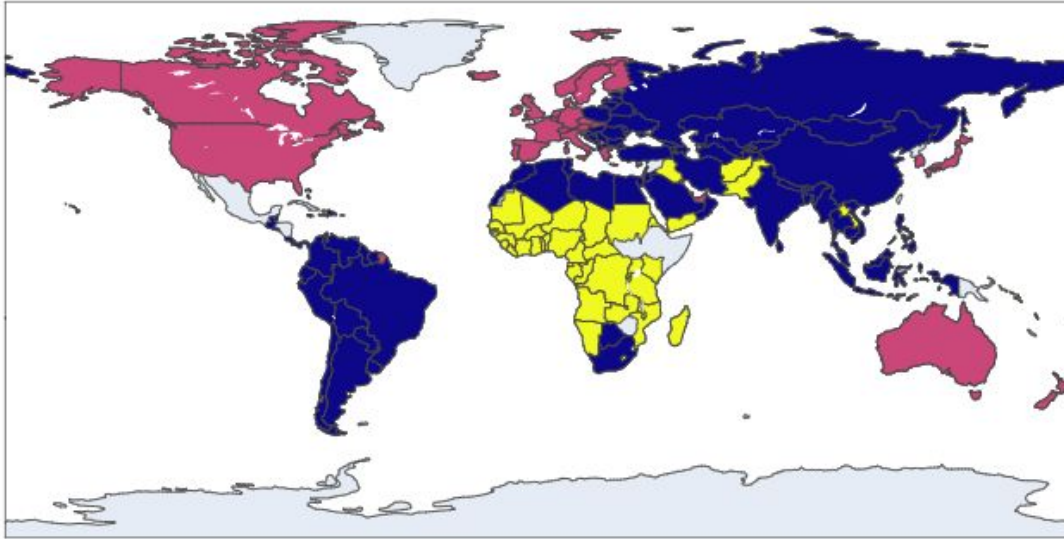
FINAL MODEL



5

FINAL MODEL

Countries by clusters

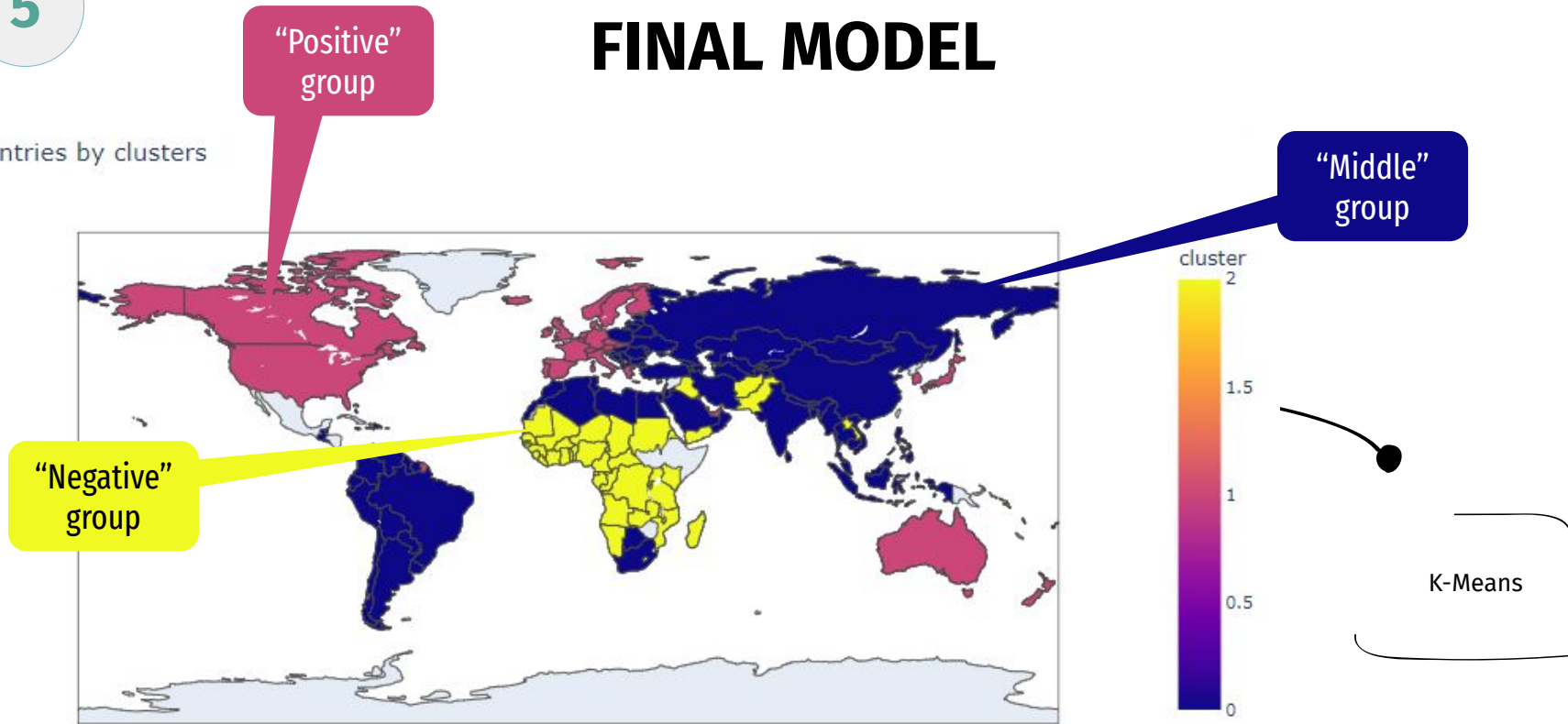


K-Means

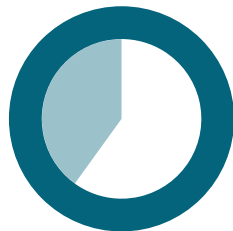
5

FINAL MODEL

Countries by clusters

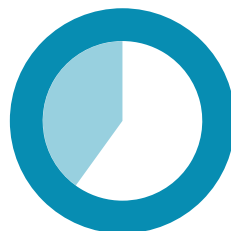


LIMITATIONS



Further clustering

Preliminary clustering identified 3 large clusters but there are still many differences within clusters



t-SNE

High-dimensional data involved



END