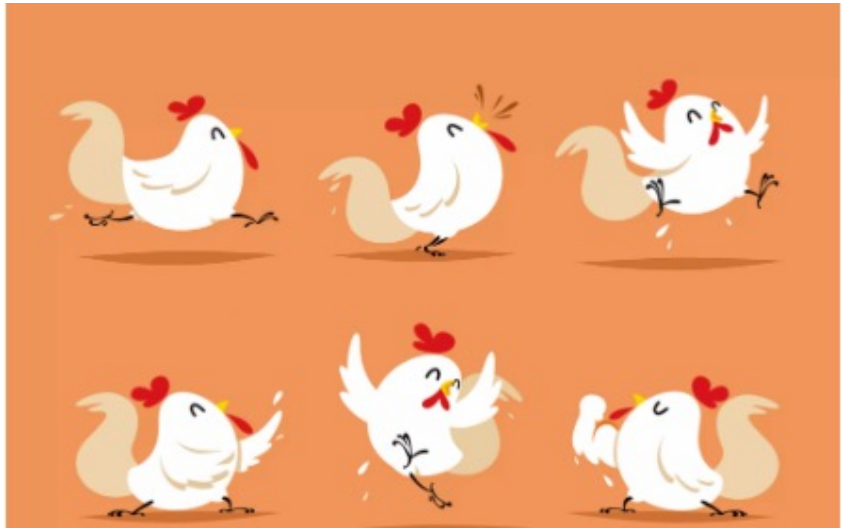
# Market research for poultry export projects



Hani Cherid

# CONTENTS

# CONTENTS

# Data Collection

Collected 4 kinds of data in each country:

1. **Poultry demand and supply**
   - Import (kg/capita year)
   - Production (kg/capita year)
   - Food intake (kcal/capita day)
   - Protein intake (g/capita day)
   - Fat intake (g/capita day)

2. **Diet conditions (Calories and animal proteins)**
   - Total food intake (kcal/capita day)
   - Total protein intake (g/capita day)
   - Proportion of animal proteins to the total(%)

3. **Population**
   - Growth rate between 2016 and 2018 (%)

4. **Number of KFC restaurants**
   - Numbers (/ capita)
   - Growth rate between 2016 and 2018 (%)

* 1-3 are available from FAO Stat

# Data Cleaning - Data from FAO Stat

Data from FAO stat

## 1. Poultry demand and supply

| | Domain Code | Domain | Area Code (FAO) | Area | Element Code | Element | Item Code | Item | Year Code | Year | Unit | Value | Flag | Flag Description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | FBS | Food Balances (2010-) | 2 | Afghanistan | 5511 | Production | 2734 | Poultry Meat | 2018 | 2018 | 1000 tonnes | 29.00 | Im | FAO data based on imputation methodology |
| 1 | FBS | Food Balances (2010-) | 2 | Afghanistan | 5611 | Import Quantity | 2734 | Poultry Meat | 2018 | 2018 | 1000 tonnes | 25.00 | Im | FAO data based on imputation methodology |

## 2. Diet conditions

| | Domain Code | Domain | Area Code (FAO) | Area | Element Code | Element | Item Code | Item | Year Code | Year | Unit | Value | Flag | Flag Description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | FBS | Food Balances (2010-) | 2 | Afghanistan | 664 | Food supply (kcal/capita/day) | 2901 | Grand Total | 2018 | 2018 | kcal/capita/day | 2270.00 | Fc | Calculated data |
| 1 | FBS | Food Balances (2010-) | 2 | Afghanistan | 674 | Protein supply quantity (g/capita/day) | 2901 | Grand Total | 2018 | 2018 | g/capita/day | 57.62 | Fc | Calculated data |

## 3. Population

| | Domain Code | Domain | Area Code (FAO) | Area | Element Code | Element | Item Code | Item | Year Code | Year | Unit | Value | Flag | Flag Description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | FBS | Food Balances (2010-) | 2 | Afghanistan | 511 | Total Population - Both sexes | 2501 | Population | 2016 | 2016 | 1000 persons | 35383.0 | * | Unofficial figure |
| 1 | FBS | Food Balances (2010-) | 2 | Afghanistan | 511 | Total Population - Both sexes | 2501 | Population | 2018 | 2018 | 1000 persons | 37172.0 | * | Unofficial figure |

- Pivot
- Join
- Null Imputations

| area_code | area | fat_gcapitaday | food_kcalcapitaday | import | production | protein_gcapitaday | country_code | total_food_kcalca | y | animal_prot_rate | pop_2018 | pop_growth_rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Armenia | 3.12 | 50.0 | 33.0 | 12.0 | 5.01 | 051 | | 0 | 49.034709 | 2952000.0 | 0.544959 |
| 2 | Afghanistan | 0.31 | 5.0 | 25.0 | 29.0 | 0.51 | 004 | | 0 | 18.552586 | 37172000.0 | 5.056100 |
| 3 | Albania | 3.83 | 50.0 | 12.0 | 16.0 | 3.52 | 008 | | .0 | 53.910887 | 2883000.0 | -0.103950 |
| 4 | Algeria | 1.59 | 23.0 | 0.0 | 286.0 | 2.08 | 012 | | 2.0 | 26.913279 | 42228000.0 | 4.135533 |
| 7 | Angola | 2.58 | 41.0 | 335.0 | 41.0 | 4.16 | 024 | | 1.0 | 30.985130 | 30810000.0 | 6.823383 |

['animal_protein_gcapitaday'] /['total_protein_gcapitaday']*100

( ['pop_2018'] - ['pop_2016']) / ['pop_2016'] *100

# Data Cleaning - KFC

## 4. Number of KFC restaurants

| | Country | Company | Franchise & License | Total |
|---|---|---|---|---|
| 0 | Albania | 0 | 2 | 2 |
| 1 | Angola | 0 | 9 | 9 |
| 2 | Antigua | 0 | 3 | 3 |
| 3 | Argentina | 0 | 6 | 6 |
| 4 | Armenia | 0 | 7 | 7 |

| | Country | Company | Franchise & License | Total |
|---|---|---|---|---|
| 0 | Albania | 0 | 2 | 2 |
| 1 | Algeria | 0 | 0 | 0 |
| 2 | Andorra | 0 | 0 | 0 |
| 3 | Angola | 0 | 9 | 9 |
| 4 | Antigua | 0 | 2 | 2 |

- Join
- Null Imputations

| area | kfc_2016 | kfc_2018 | kfc_growth_rate | country_code |
|---|---|---|---|---|
| Albania | 2 | 2 | 0.0 | 008 |
| Algeria | 0 | 0 | 0.0 | 012 |
| Andorra | 0 | 0 | 0.0 | 020 |
| Angola | 9 | 9 | 0.0 | 024 |
| Antigua | 3 | 2 | -33.333333 | 028 |

Converted from country names in 'area' using "pycountry"

['kfc_2018']-['kfc_2016'] /'kfc_2016']*100

# Data Cleaning - final step

**1-3**

| area_code | area | fat_gcapitaday | food_kcalcapitaday | import | production | protein_gcapitaday | country_code | total_food_kcalcapitaday |
|---|---|---|---|---|---|---|---|---|
| 1 | Armenia | 3.12 | 50.0 | 33.0 | 12.0 | 5.01 | 051 | 3074.0 |
| 2 | Afghanistan | 0.31 | 5.0 | 25.0 | 29.0 | 0.51 | 004 | 2270.0 |
| 3 | Albania | 3.83 | 50.0 | 12.0 | 16.0 | 3.52 | 008 | 3296.0 |
| 4 | Algeria | 1.59 | 23.0 | 0.0 | 286.0 | 2.08 | 012 | 3382.0 |
| 7 | Angola | 2.58 | 41.0 | 335.0 | 41.0 | 4.16 | 024 | 2451.0 |

**4**

| area | kfc_2016 | kfc_2018 | kfc_growth_rate | country_code |
|---|---|---|---|---|
| Albania | 2 | 2 | 0.0 | 008 |
| Algeria | 0 | 0 | 0.0 | 012 |
| Andorra | 0 | 0 | 0.0 | 020 |
| Angola | 9 | 9 | 0.0 | 024 |
| Antigua | 3 | 2 | -33.333333 | 028 |

- Join
- Null Imputations

| area | import_kgcapita | production_kgcapita | food_kcalcapitaday | protein_gcapita | p_growth_rate | kfc_2018_capita | kfc_growth_rate |
|---|---|---|---|---|---|---|---|
| Armenia | 11.178862 | 4.065041 | 50.0 | 5 | 0.544959 | 3.048780e-06 | 28.571429 |
| Afghanistan | 0.672549 | 0.780157 | 5.0 | | 5.056100 | 0.000000e+00 | 0.000000 |
| Albania | 4.162331 | 5.549775 | 50.0 | | -0.103950 | 6.937218e-07 | 0.000000 |
| Algeria | 0.000000 | 6.772757 | 23.0 | | 4.135533 | 0.000000e+00 | 0.000000 |
| Angola | 10.873093 | 1.330737 | 41.0 | | 6.823383 | 2.921130e-07 | 0.000000 |

Converted to per capita values

Converted to per capita values

```
Data columns (total 11 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   import_kgcapita          179 non-null    float64
 1   production_kgcapita      179 non-null    float64
 2   food_kcalcapitaday       179 non-null    float64
 3   protein_gcapitaday       179 non-null    float64
 4   fat_gcapitaday           179 non-null    float64
 5   total_food_kcalcapitaday 179 non-null    float64
 6   total_protein_gcapitaday 179 non-null    float64
 7   animal_prot_rate         179 non-null    float64
 8   pop_growth_rate          179 non-null    float64
 9   kfc_2018_capita          179 non-null    float64
```

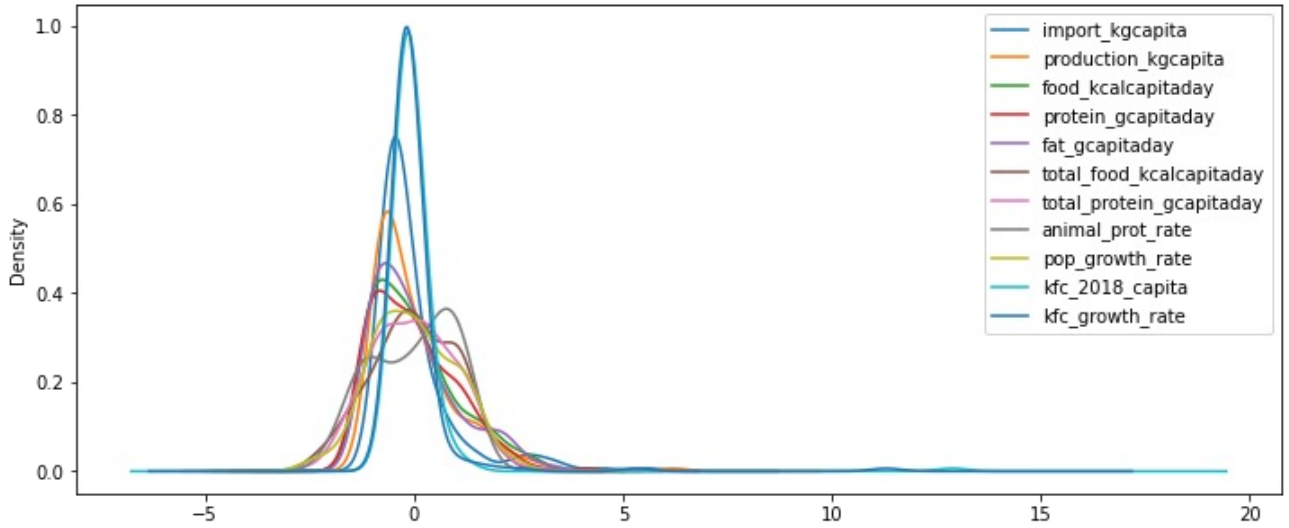
**11 variables, 179 countries**

# CONTENTS

# Hierarchical Clustering

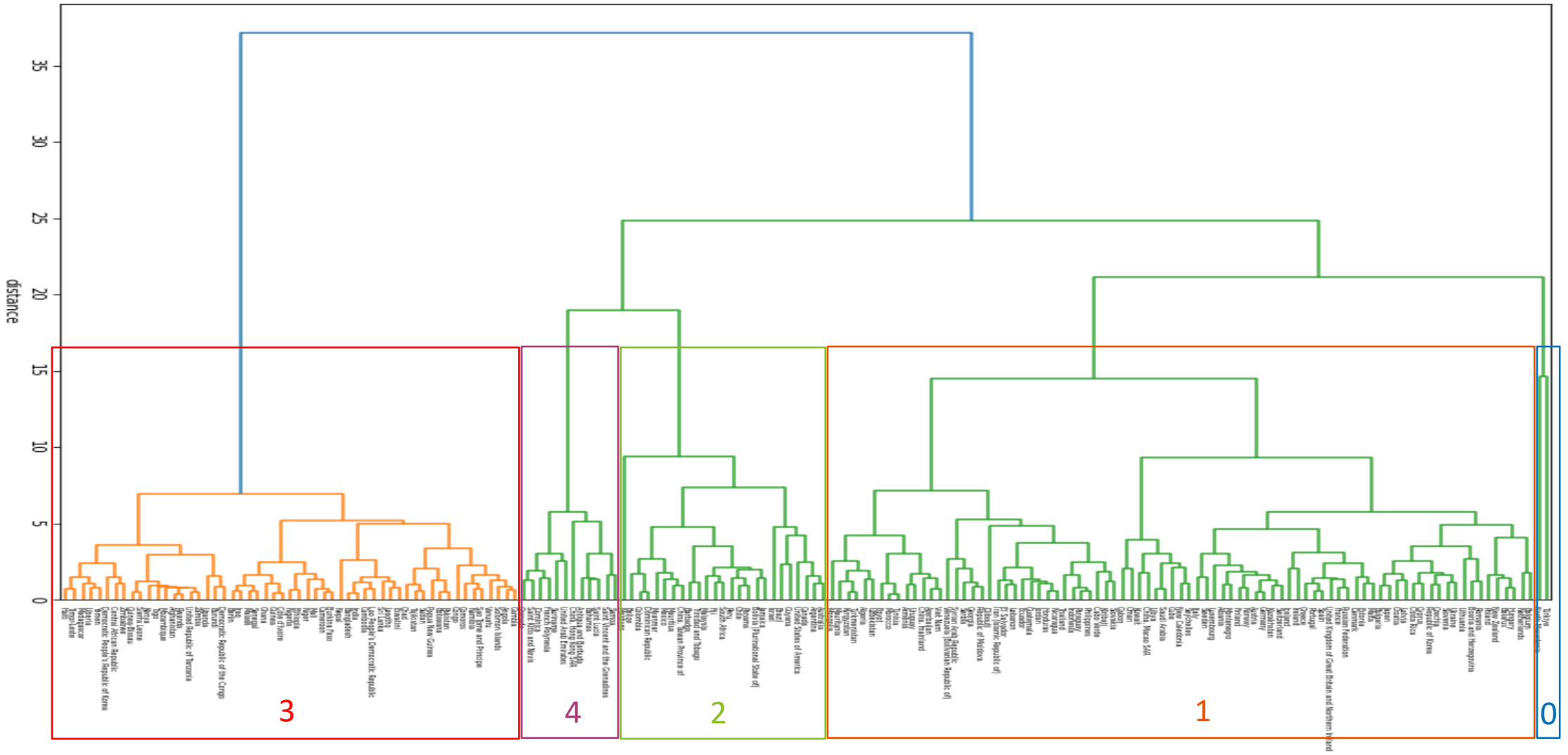
Cleaned data

- Exclude qualitative variables
- Scaling




```
# Import the hierarchical clustering algorithm
from sklearn.cluster import AgglomerativeClustering

# Create a hierarchical clustering model
hiercluster = AgglomerativeClustering(affinity='euclidean', linkage='ward',

hiercluster.set_params(n_clusters=5)
clusters = hiercluster.fit_predict(X_scaled)
```

# Dendrogram – Definition of countries clusters




**55 countries**
incl Afghanistan

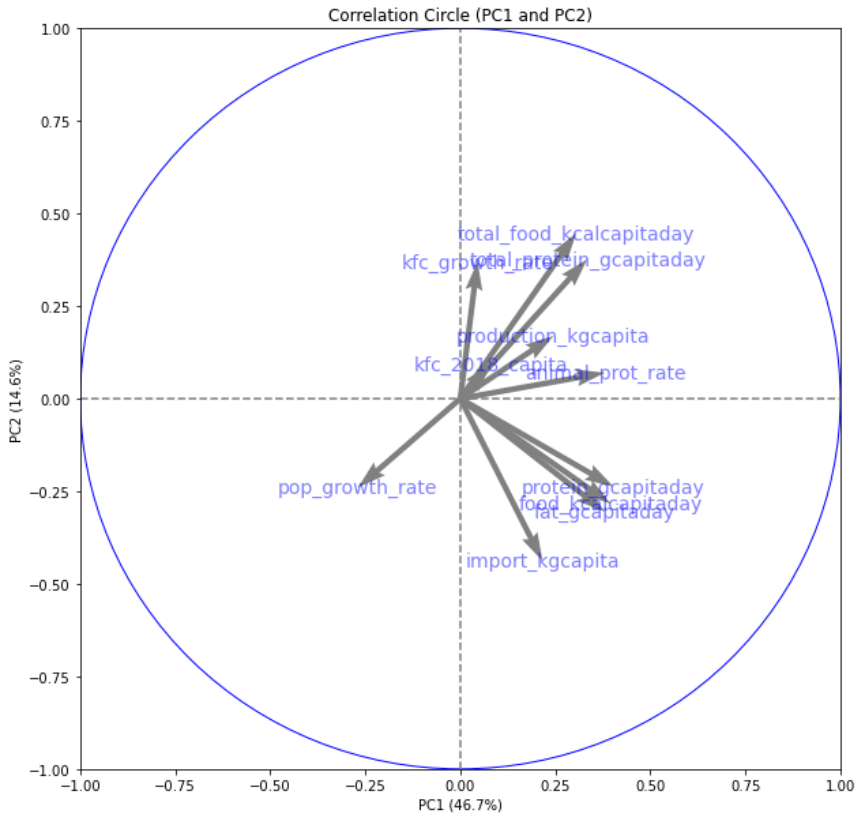
**12 countries**
incl Samoa

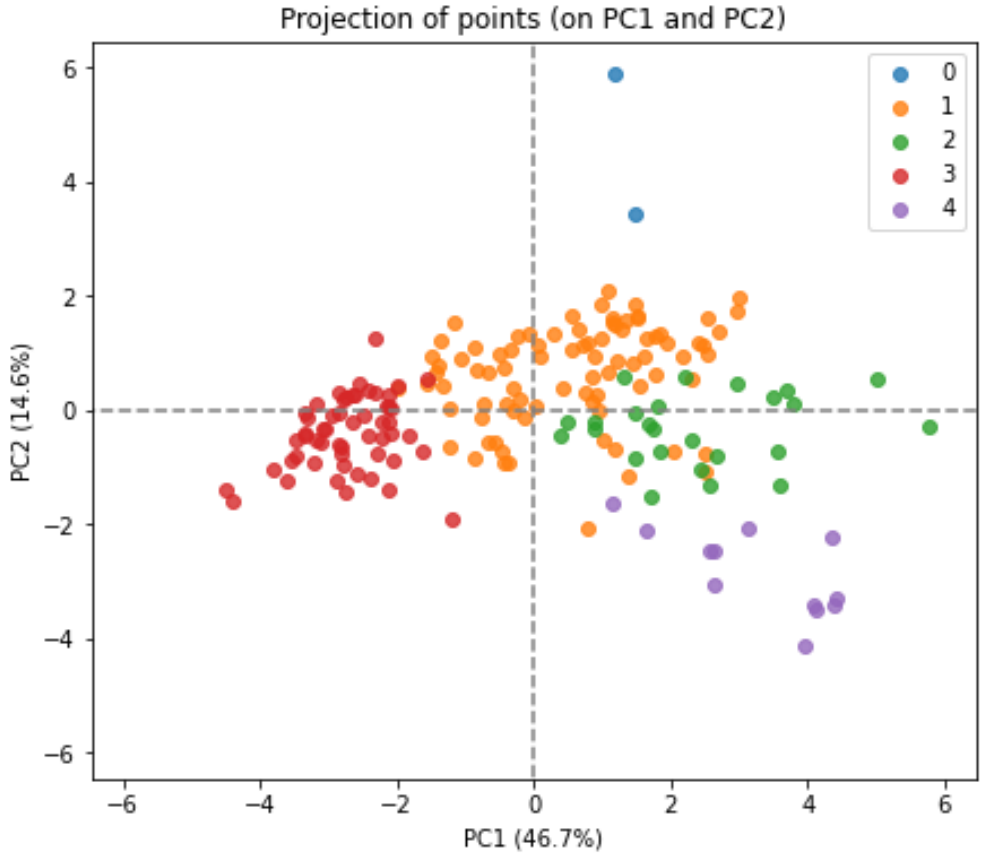
**25 countries**
incl Australia

**85 countries**
incl Armenia

**2 countries**
Incl Turkey

# PCA

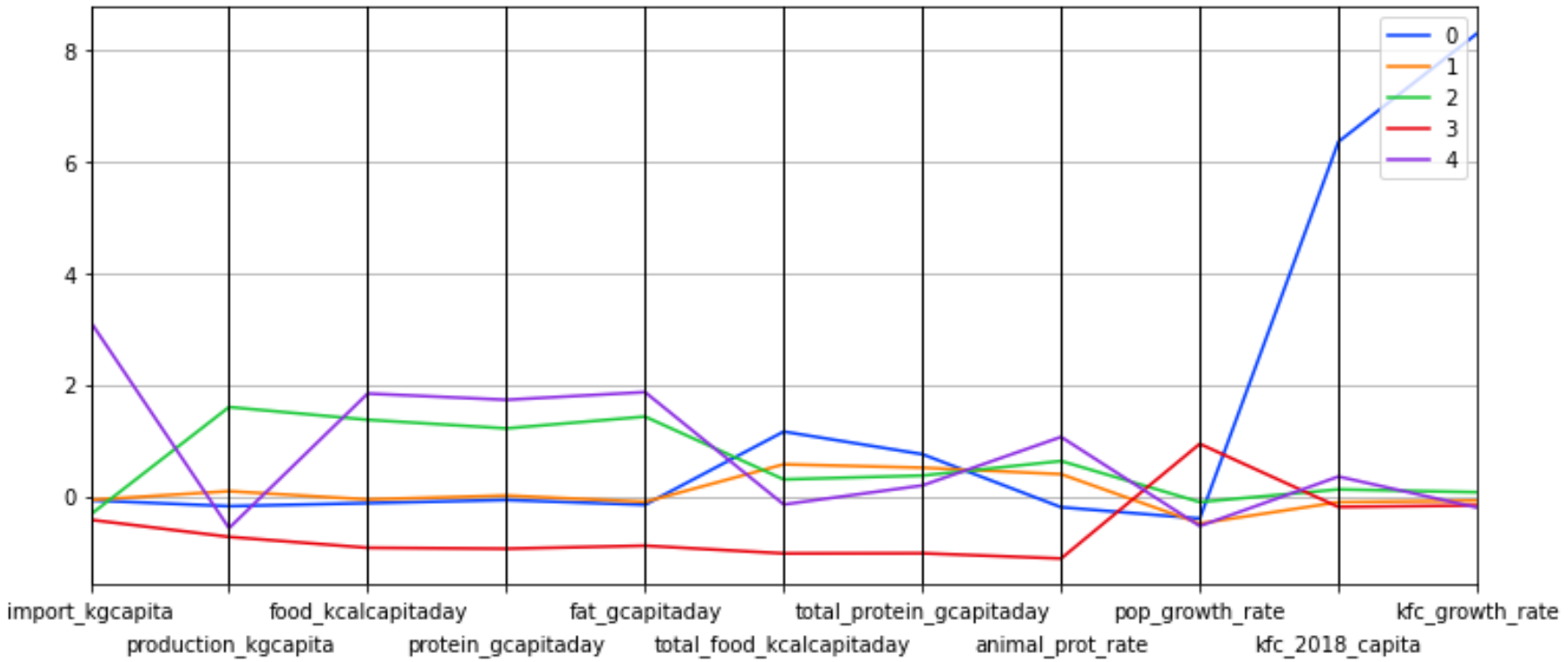



PC1(46.7%) and PC2(14.6%) **explain** about **60% of the total data**.

- Group 0 is prominently higher in PC1
- Group 2 and Group 3 do not differ in PC1 but are opposite in PC2
- Group 4 is high in PC1 but low in PC2
- Group 1 is located around the center of both PC1 and PC2 axes and appears less distinctive

# Parallel Coordinates Plot



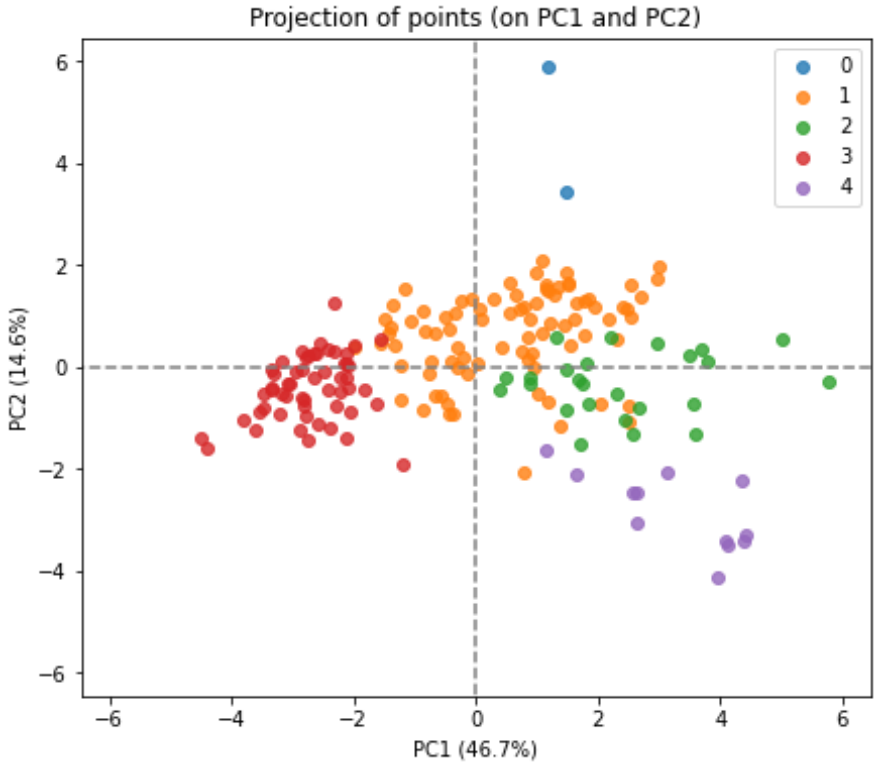

**Which group can we expect as our targets? :**

Group 0 (2 countries) : Extremely high scores of KFC per capita and its growth

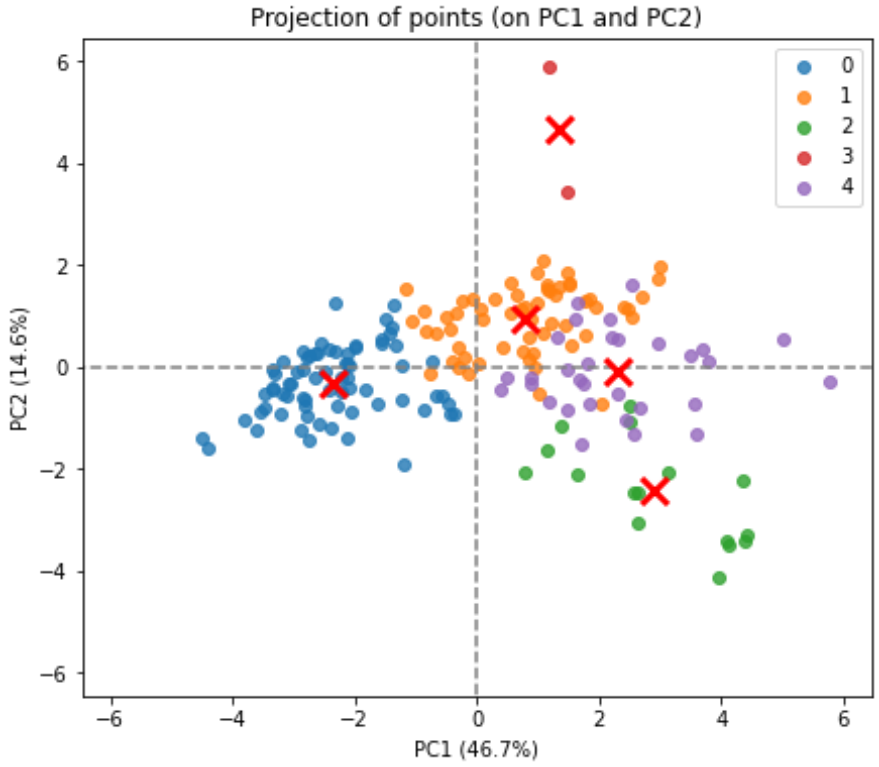
Group 4 (12 countries) : High levels of poultry meat intake but low domestic production, dependent on imports

# Comparison between Hierarchical- and K-Means Clustering

- Hierarchical



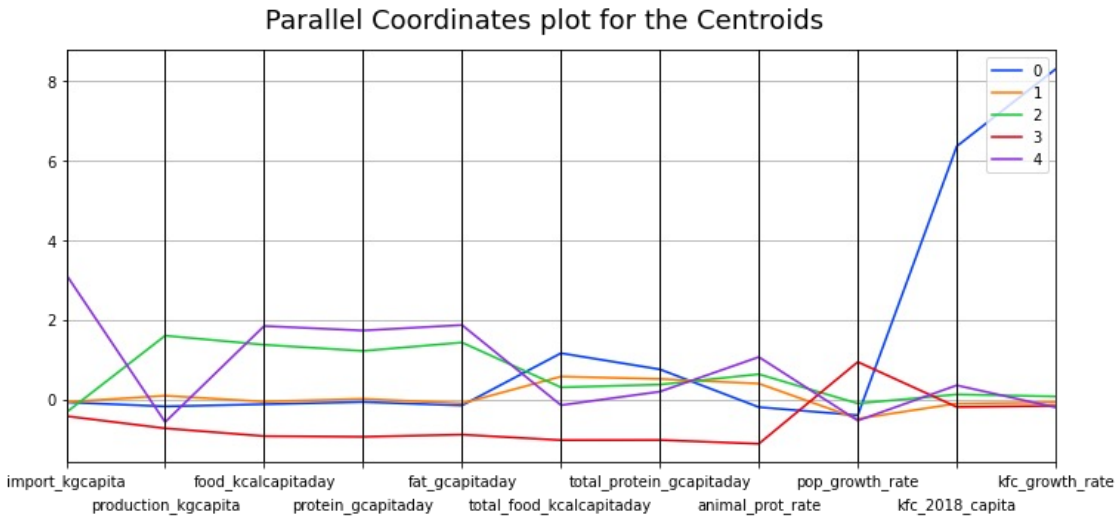

- K-Means



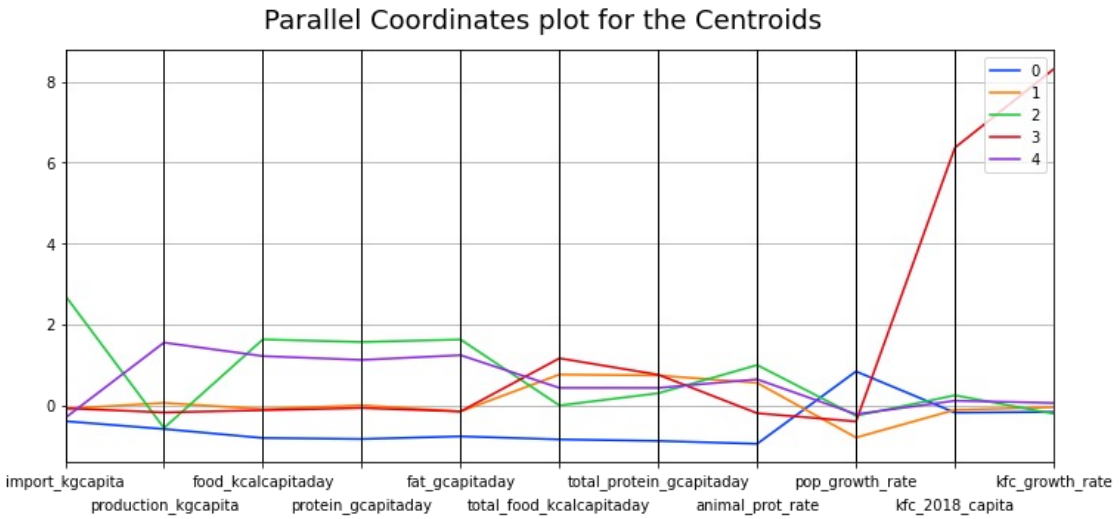

**Partitions of clusters are very similar to each other**

# Comparison between Hierarchical- and K-Means Clustering

- Hierarchical




- K-Means




Also looks similar

Compare using set()

- Group 0 countries in Hierarchical = Group 3 countries in K-means
- Group 4 countries in Hierarchical = Group 2 countries +4 more in K-means

-> **The results of hierarchical and K-means clustering are consistent**

# Conclusions

Country list of Group 0 and 4:

| Group | Country | Poultry import (kg/capita) | Poultry production (kg/capita) | Population | Pop_growth (%) | KFC_growth (%) |
|---|---|---|---|---|---|---|
| 0 | Turkey | 0.1 | 27.1 | 82,340,000 | 3.1 | 1,675 |
| | North Macedonia | 19.2 | 0.5 | 2,083,000 | 0.1 | 807 |
| 4 | United Arab Emirates | 63.2 | 4.9 | 9,630,960 | 2.9 | 7 |
| | Hong Kong | 111.4 | 3.4 | 7,372,000 | 1.8 | 11 |
| | Suriname | 38.2 | 19.1 | 576,000 | 1.9 | 50 |
| | Bahamas | 67.4 | 18.1 | 386,000 | 2.1 | 8 |
| | French Polynesia | 57.6 | 3.6 | 278,000 | 1.1 | 0 |
| | Samoa | 91.8 | 0.0 | 196,000 | 0.5 | 0 |
| | Saint Lucia | 54.9 | 11.0 | 182,000 | 1.1 | 0 |
| | Grenada | 63.1 | 9.0 | 111,000 | 0.9 | 0 |
| | Saint Vincent and the Grenadines | 72.7 | 0.0 | 110,000 | 0.9 | 0 |
| | Antigua and Barbuda | 72.9 | 0.0 | 96,000 | 1.1 | -33 |
| | Dominica | 55.6 | 0.0 | 72,000 | 1.4 | 0 |
| | Saint Kitts and Nevis | 57.7 | 0.0 | 52,000 | 0.0 | 0 |

**Target areas:**
**Group 0: Turkey and North Macedonia**

# CONTENTS

# Kaiser-Meyer-Olkin (KMO) Test

Validate sample size by the Kaiser-Meyer-Olkin (KMO) Test

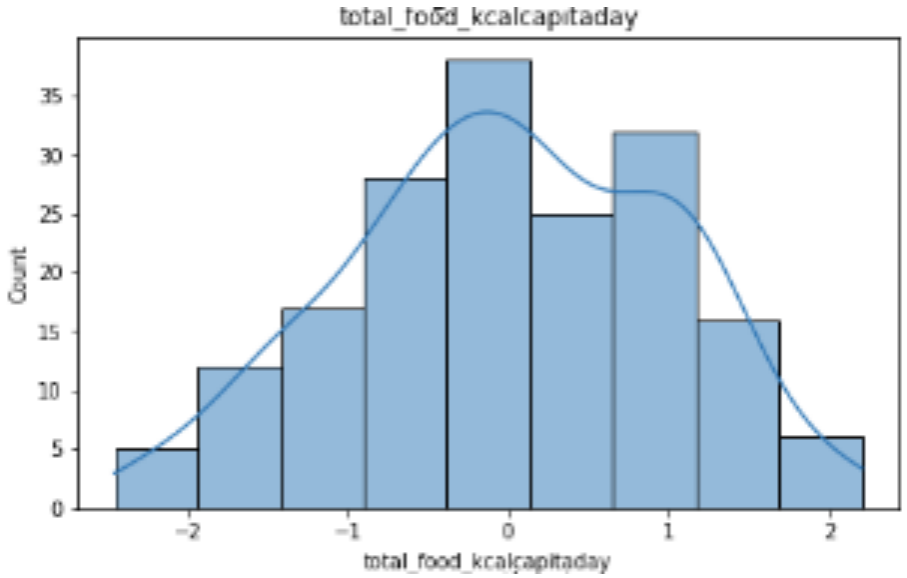
```
from factor_analyzer.factor_analyzer import calculate_kmo

kmo_all, kmo_model = calculate_kmo(X)
print(kmo_all, kmo_model)
```

```
[0.72845188 0.78681264 0.62399797 0.62857465 0.60528642 0.74971941
 0.7501516  0.86196514 0.87453552 0.5827879  0.53659464] 0.6969610497399634
```

Value of overall dataset: 0.70
-> **Sample size is adequate**

- 0.00 to 0.49 unacceptable.
- 0.50 to 0.59 miserable.
- 0.60 to 0.69 mediocre.
- 0.70 to 0.79 middling.
- 0.80 to 0.89 meritorious.
- 0.90 to 1.00 marvelous.

# Normality Test (Kolmogorov-Smirnov test)




**total_food_kcalcapitaday**

H0: Null hypothesis, the variable follows a normal distribution
H1: Alternative, the variable does not follow a normal distribution

If p value is < .05 REJECT the Null hypothesis.

kstest()

```
KstestResult(statistic=0.056383272781445615, pvalue=0.5995940995280229)
```

p value = 0.6 : CANNOT reject the null hypothesis
-> **The variable is normally distributed**

# Comparison test of two populations (t-test)

| Cluster | Avg of 'total_food_Kcalcapitaday' |
|---|---|
| Group 0 | 1.15911546 |
| Group 4 | -0.149666229 |

H0: Null hypothesis
No significant difference in average of 2 groups
H1: Alternative
There is a significant difference in average of 2 groups

If p value is < .05 REJECT the Null hypothesis.

ttest_ind()

Ttest_indResult(statistic=2.775538078535637, pvalue=0.01678868248048485)

p value = 0.02: CAN reject the null hypothesis
-> **The difference of 2 groups is statistically significant**

# CONTENTS

# Conclusion

- Divided the country list into 5 groups by the hierarchical clustering,
  resulting in the selection of a target countries:
  - Turkey and North Macedonia (high scores of KFC per capita and its growth)

- Results of K-means clustering were confirmed to be consistent with the above results.

- KMO test confirmed the validity of the sample size.

- Normality tests confirmed that one variable 'total_food_kcalcapitaday' (represents total calorie intake (kcal/capitaday)) in the sample was normally distributed.

- T-test proved that there is a significant difference between the two selected groups of total calories intake.