

Academia de Studii Economice, București

Facultatea de Cibernetică, Statistică și Informatică Economică

Specializarea: Informatică Economică

Software Development for Data Analysis

-Project-

**Digital Economy and society-E-commerce**

Coordinating teacher, Simionescu Silviu-Robert

Conf.univ.dr.Claudiu Vinte Rosu Cristian-Codrut

1105

**Table of Contents:**

1. Data source
2. Time reference
3. Description of the variables
4. Description of observations
5. Data analysis approach
6. The motivation to opt for a particular data analysis method
7. Presentation of the results + Interpretation
8. Discussion and conclusions
9. **Data sources for each data set:**

**ICS** - Individuals - use of cloud services

<https://ec.europa.eu/eurostat/databrowser/view/isoc_cicci_use__custom_9325869/default/table?lang=en>

**ICE** - Individuals - use of collaborative economy

<https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_ce_i__custom_9325903/default/table?lang=en>

**IPI** - Internet purchases by individuals

<https://ec.europa.eu/eurostat/databrowser/view/isoc_ec_ib20__custom_9325925/default/table?lang=en>

**IGS** - Internet purchases - goods or services

<https://ec.europa.eu/eurostat/databrowser/view/isoc_ec_ibgs__custom_9325936/default/table?lang=en>

**IOS** - Internet purchases - origin of sellers

<https://ec.europa.eu/eurostat/databrowser/view/isoc_ec_ibos__custom_9325953/default/table?lang=en>

**IMS** - Internet purchases - money spent

<https://ec.europa.eu/eurostat/databrowser/view/isoc_ec_ibm__custom_9325967/default/table?lang=en>

**FAI** - Financial activities on the internet

<https://ec.europa.eu/eurostat/databrowser/view/isoc_ec_ifi20__custom_9326067/default/table?lang=en>

**PBOI** - Perceived barriers to buying/ordering over the internet

<https://ec.europa.eu/eurostat/databrowser/view/isoc_ec_inb__custom_9326102/default/table?lang=en\>

**IPCE** - Internet purchases - collaborative economy

<https://ec.europa.eu/eurostat/databrowser/view/isoc_ec_ce_i__custom_9326144/default/table?lang=en>

1. **Time reference regarding the data**

The data that we have collected is for year 2020-2021, these two years representing a significant period for the Covid-19 virus to have an impact on basically everything, but for now we are interested in the E-commerce domain.

1. **Description of the variables**

**ICS** - Individuals - use of cloud services:

Represents the degree or frequency of individuals using cloud services. Higher values may indicate more extensive use of cloud services.

**ICE** - Individuals - use of collaborative economy:

Measures the extent to which individuals participate in the collaborative economy. Higher values may suggest increased involvement in sharing resources or services through online platforms.

**IPI** - Internet purchases by individuals:

Quantifies the number or frequency of internet purchases made by individuals. Higher values indicate more frequent online shopping activity.

**IGS** - Internet purchases - goods or services:

Differentiates between the types of items individuals purchase online. It could be binary (0 or 1) where 0 represents services and 1 represents goods, or it may use numerical values to categorize the nature of purchases.

**IOS** - Internet purchases - origin of sellers:

Represents categories or numerical codes corresponding to the origin of sellers in internet purchases. It might assign different values for local, national, and international sellers.

**IMS** - Internet purchases - money spent:

Represents the amount of money individuals spend on internet purchases. It's a quantitative measure indicating the financial impact of online consumer behavior.

**FAI** - Financial activities on the internet:

Quantifies the extent of individuals' engagement in various financial activities on the internet. It could be a composite score reflecting the overall involvement in online financial transactions.

**PBOI** - Perceived barriers to buying/ordering over the internet:

Represents a scale or index measuring individuals' perceived barriers to making purchases or orders over the internet. Higher values might indicate a higher perceived difficulty or obstacles in online shopping.

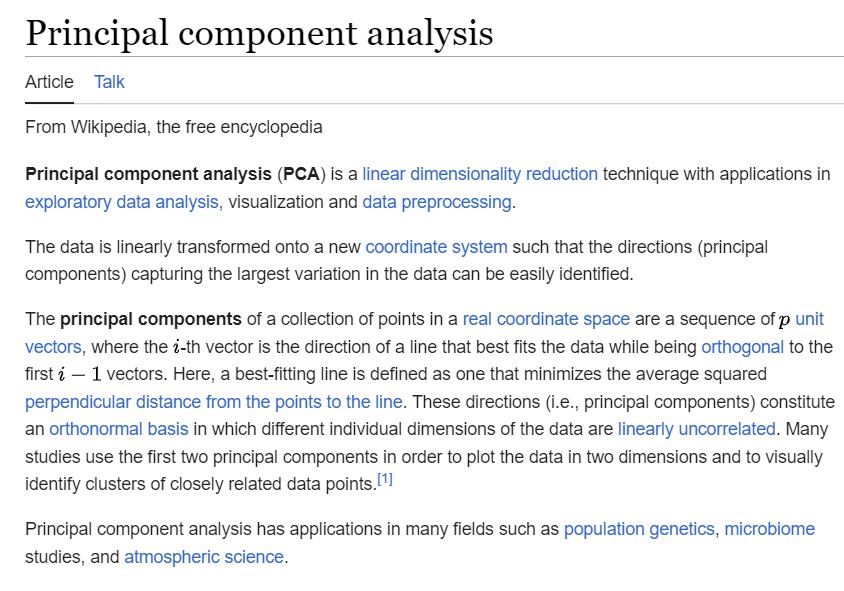
1. **Description of the observations**

The dataset contains a comprehensive array of observations spanning various European countries and aggregates, providing a detailed depiction of the Euro area dynamics for the years 2020-2021. Each row represents a specific country furnishing details on the use of **cloud services**, involvement in the **collaborative economy**, trends in **internet purchases**, and a spectrum of **E-commerce** metrics.

This extensive dataset is a useful tool for in-depth research, revealing various aspects of **digital trends**, **economic activities**, and **consumer behaviors** in the specified **regions** during the specified **time frame**.

1. **Data analysis approach**

The first data analysis method we chose for this dataset is **PCA** – Principal components analysis

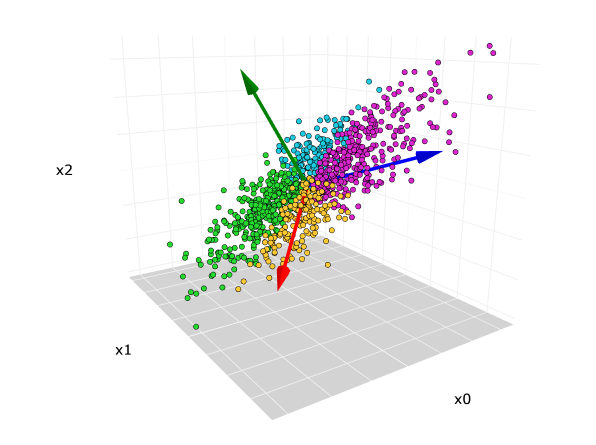


**PCA** was invented in 1901 by **Karl Pearson**, as an analogue of the principal axis theorem in mechanics; it was later independently developed and named by **Harold** **Hotelling** in the 1930s.

<https://www.datacamp.com/tutorial/principal-component-analysis-in-python>

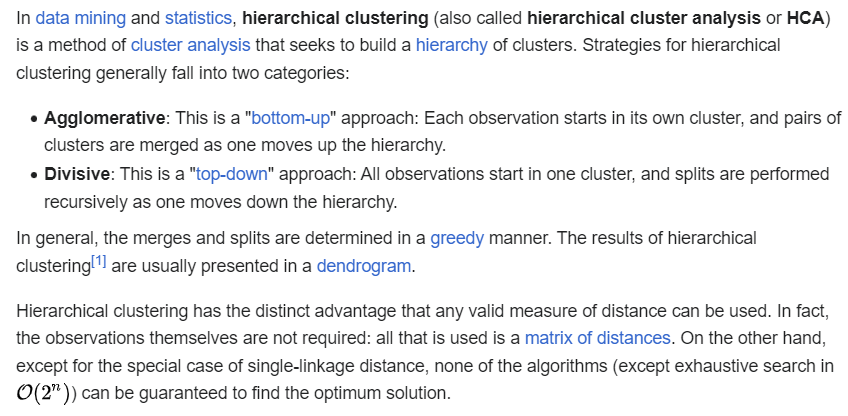
**PCA** identifies the directions (principal components) along which the data varies the most. The first **principal component** explains the maximum variance, followed by the second, and so on. By doing this, PCA enables the **reduction** of the dimensionality of the data while retaining most of its **important** information.

<https://towardsdatascience.com/principal-component-analysis-pca-explained-visually-with-zero-math-1cbf392b9e7d>



In Python, we can apply PCA to identify patterns and reduce the number of variables in your dataset. By utilizing functions provided by various libraries, we can easily transform data into its principal components. These components capture the most significant sources of variation, allowing you to focus on the key aspects of your data.

The **second** data analysis method we chose for this dataset is **HCA** – Hierarchical Cluster Analysis is a method extensively employed in Python for data analysis, particularly in **clustering** and grouping **related data points**. Python libraries like **SciPy** and sc**ikit-learn** offer functionalities to implement HCA seamlessly.



<https://en.wikipedia.org/wiki/Hierarchical_clustering#References>

Implementing HCA in Python facilitates the exploration of relationships and patterns within the given data, aiding in the identification of **natural clusters**. This method is beneficial for tasks such as taxonomy construction, pattern recognition, and understanding the inherent structures present in datasets.

1. **The motivation to opt for a particular**

**data analysis method**

In opting for our analysis approach, we chose Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (**HCA**) to extract **meaningful insights** from our extensive e-**commerce** dataset. **PCA** provides a robust method to **simplify our data**, revealing crucial patterns and relationships among various metrics. This choice enables us to identify the **key facto**rs influencing e-commerce **trends** across European countries efficiently.

Additionally, the application of **HCA** allows us to uncover inherent **structures** and **similarities** within the dataset. By identifying clusters of countries with similar e-commerce behaviours, we aim to categorize and understand **shared patterns**. This combined use of **PCA** and **HCA** enhances our **analytical** **depth**, enabling a more detailed exploration of the diverse e-commerce **dynamics** present in the European landscape.

1. **Presentation of the results  
   +  
   Interpretations**

**1. Eigenvalues and Explained Variance:**

Interpretation: In this analysis, **eigenvalues** are used to understand the variance explained by each principal component. **Higher** eigenvalues indicate more **significant contributions** to explaining the **variability** in the dataset. For instance, if C1 has a **high eigenvalue**, it implies that **C1** captures a substantial portion of the overall variance in the dataset. In the context of e-commerce, a high eigenvalue for a component suggests that the associated set of variables str**ongly influences the observed patterns** in the data.

**2. Correlogram of Factor Loadings (Factor Analysis):**

Interpretation: The **factor loadings** in the correlogram represent the correlation between observed variables and principal components. Variables with high loadings on a specific component are crucial contributors to that component. For example, if **C1** shows **high loadings** for **ICS** (Individuals - Use of Cloud Services), it suggests a strong correlation, indicating that cloud services usage significantly influences the definition of **C1**. Similarly, for **C2**, variables with high loadings, especially for ICE (Individuals - Use of Collaborative Economy), are influential in shaping **C2**.

Higher Loadings on **C1**: Higher loadings on **C1** indicate a potential association between cloud services usage and individuals' tendencies to make internet purchases. Individuals who actively use cloud services may exhibit specific preferences or behaviors in their online shopping habits.

**3. Intensity Map of Principal Component Scores:**

Interpretation: The intensity map helps identify patterns in the scores of observations on each principal component. Higher scores indicate a **stronger representation** of the corresponding component. For instance, individuals with **high score**s on **C1** are likely to be **active users** of cloud services in the e-commerce context. This interpretation helps in characterizing distinct user groups based on their behavior in the e-commerce domain.

High Scores on **C2**: Individuals with high scores on **C2** in the intensity map may exhibit patterns of higher spending in their internet purchases. **C2**, in this context, could represent a component related to the financial aspects of e-commerce, indicating individuals who **spend more in their online transactions**.

**4. Quality of Points Representation:**

Interpretation: This map illustrates the quality of points representation on each component. **Higher values** indicate **better representation**. If an individual has a high quality of representation on **C2**, it means that their behavior **aligns well** with the **collaborative economy component**. Identifying influential observations with high quality representation provides **insights** into key user behaviors and their impact on specific aspects of e-commerce.

High Quality Representation on C3: Individuals with **high quality representation** on **C3** may suggest a strong influence of perceived barriers to online buying. **C3**, in this context, could capture aspects related to individuals who are more cautious or hesitant about making purchases over the internet.

**5. Contribution of Observations to the Axes' Variance:**

Interpretation: The contribution map shows how much each observation contributes to the variance in each component. Focus on observations with **high contributions**, as they play a crucial role in shaping the axes' variance. If an individual contributes significantly to **C3**, for example, it implies that their behavior strongly influences the variability captured by **C3**. This analysis helps identify key user behaviors that drive variability in the e-commerce dataset.

High Contribution to **C4**: Individuals with a high contribution to **C4** may be key influencers in the variance related to internet purchases of goods or services. **C4** might represent a component highlighting diverse patterns in the types of goods or services individuals frequently purchase online.

**6. Commonalities:**

Interpretation: Commonalities represent the proportion of variance in each variable explained by all the components. Variables with **high commonalities** across **multiple** components are **crucial** **contributors** to the overall variance. Identifying these variables provides insights into the key features that **consistently** influence different aspects of e-commerce.

High Commonality across **C5** and **C6**: **IPCE** exhibiting high commonality across multiple components (**C5** and **C6**) indicates that collaborative economy-related internet purchases play a consistent role in explaining variance across different aspects of the e-commerce dataset.

1. **Discussions and conclusions**

In our analysis, we dived into the connections between various aspects of online shopping and digital activities in European countries. Using PCA and HCA, we figured out patterns that reveal the factors influencing e-commerce practices.

We gathered data from Eurostat and manipulated it in Python, having the seminars as starting points for the two analysis methods used and implemented: PCA and HCA.

Our project highlights real-world implications, especially the impact of COVID-19 on online shopping trends in 2020-2021. We broke down each element to understand how they shape people's choices in the online retail world.

To sum it up, our findings give valuable insights into the dynamics guiding the digital economy and society. We explore online shopping behaviors, internet-based financial transactions, and the challenges people face. This broad view enhances our understanding of the European e-commerce scene.