Principal Component Analysis

Environmental Impact of the Fast Fashion Industry

ADIMI Dalil BEN AMOR Hassen DACHRAOUI Elyes
TIR Ayoub AZGOUNE Mustapha

ING2-NEW

Course Instructor: MOHSEN Zahraa

EFREI Paris
ST2DA-I2 | 2025-2026

December 16, 2025

Outline

Context and Motivation

The Fast Fashion Industry

- Rapid production cycles and low-cost manufacturing
- Significant environmental footprint: CO_2 emissions, water usage, textile waste
- Growing consumer awareness and demand for sustainability

Why Principal Component Analysis?

- **High-dimensional data**: 15 environmental, economic, and social variables
- **Correlated variables**: Many metrics measure related aspects
- **Goal**: Identify underlying patterns and reduce complexity

Dataset Overview

Dataset: True Cost of Fast Fashion

- **Individuals**: 3,000 observations
- **Brands**: Shein, Zara, H&M, Forever 21, Uniqlo
- **Time period**: 2015–2024
- **Countries**: Multiple production locations

Selected Variables ($p = 15$)

- Production metrics (2)
- Environmental impact (3)
- Economic factors (2)
- Sustainability indicators (5)
- Social conditions (3)

Appropriateness for PCA

Variables measured in different units require **standardization** $\Rightarrow$ PCA on correlation matrix R

Research Objectives

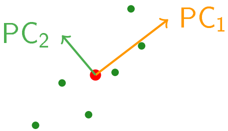
1. **Dimensionality Reduction**: Transform 15 correlated variables into fewer uncorrelated principal components
2. **Pattern Discovery**: Identify latent factors driving environmental impact
3. **Brand Comparison**: Visualize brands in reduced principal component space
4. **Variable Interpretation**: Understand relationships between environmental and sustainability metrics

Definition

PCA is a dimensionality reduction technique that transforms p correlated variables into p **uncorrelated** principal components, ordered by the amount of variance they explain.

Key Ideas:

- Find new axes (PCs) that capture maximum variance
- First PC: direction of greatest spread
- Subsequent PCs: orthogonal directions
- Reduce dimensions while preserving information



Mathematical Foundation

Standardization

Transform M to standardized M_s: $z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$
where $\bar{x}_j$ = mean, σ_j = std. dev. of variable j

Eigenvalue Decomposition

Correlation matrix: $R = \frac{1}{n} M_s^T M_s$, solve: $R \cdot u = \lambda \cdot u$

- λ_i: eigenvalue (variance by PC_i)
- u_i: eigenvector (direction of PC_i)

Key Matrices in PCA

1. **Change of Basis Matrix** P:

$$P = [u_1 \mid u_2 \mid \cdots \mid u_p]_{p\times p}$$

Columns are normalized eigenvectors defining the new axes.

2. **Principal Component Scores** F:

$$F_{n\times p} = M_s \cdot P$$

Coordinates of individuals in the principal component space.

3. **Saturation Matrix (Loadings)** S:

$$S_{p\times p} = P \cdot D^{1/2}, \quad D = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$$

Correlations between original variables and principal components.

1. Overall Quality of Explanation (OQE)

$$\mathrm{oqe}(i) = \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j} = \frac{\lambda_i}{p}$$

Proportion of total variance explained by PC_i.

2. Quality of Representation (QLT)

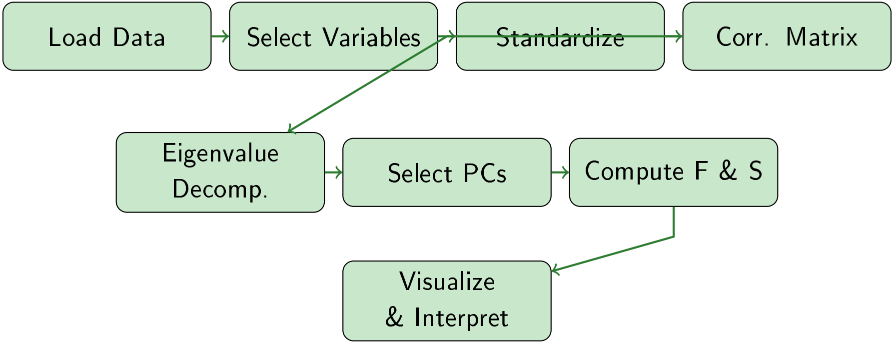
$$\mathrm{qlt}(\text{ind. } i, \text{axis } j) = \frac{f_{i,j}^2}{\sum_{k=1}^{p} f_{i,k}^2}$$

How well individual i is represented by axis j.

3. Correlation Circle

- Visualizes variable loadings
- Variables near the circle edge: well represented
- Angle between variables: correlation
- Close to 0°: positive correlation
- Close to 180°: negative correlation
- Close to 90°: uncorrelated

Analysis Workflow



Variable Selection

Category	Variables
Production	Monthly Production, Release Cycles
Environmental	Carbon Emissions, Water Usage, Waste
Economic	Avg Item Price, GDP Contribution
Sustainability	Env. Cost Index, Sustainability Score, Transparency, Compliance, Ethical Rating
Social	Worker Wage, Working Hours, Child Labor

Total: 15 quantitative variables across 5 categories

Data Preprocessing

Standardization Verification

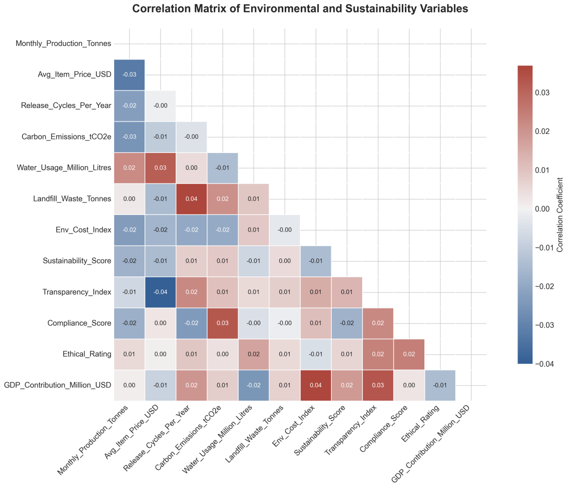
After standardization: Mean ≈ 0, Standard deviation ≈ 1

Sample Variables	Mean	Std
Monthly Production	0.000	1.000
Carbon Emissions	0.000	1.000
Sustainability Score	0.000	1.000
Transparency Index	0.000	1.000
Worker Wage	0.000	1.000

Why Standardize?

Variables have different scales $\Rightarrow$ Without standardization, high-variance variables dominate PCA

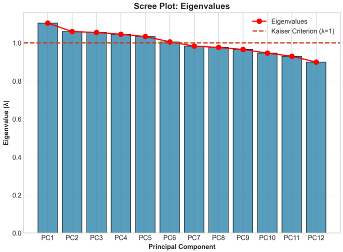
Correlation Matrix

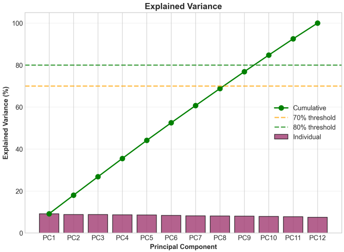


Key Observations:

- Strong correlations exist
- Environmental vars: positive
- Sustainability: negative with costs
- Justifies PCA use

Eigenvalue Analysis





Component Selection:

- **70% threshold**: 11 components
- **80% threshold**: 12 components
- **Selected**: 12 components
- **Variance explained**: 81.94%

Note

Eigenvalues are relatively uniform ⇒ No single dominant component

Eigenvalues Summary

PC	λ	Var%	Cum%
PC1	1.104	7.43	7.43
PC2	1.090	7.26	14.69
PC3	1.080	7.20	21.89
PC4	1.063	7.08	28.97
PC5	1.053	7.02	35.99
PC6	1.029	6.86	42.85

PC	λ	Var%	Cum%
PC7	1.013	6.75	49.61
PC8	1.006	6.71	56.31
PC9	0.985	6.57	62.88
PC10	0.978	6.52	69.40
PC11	0.944	6.29	75.69
PC12	0.937	6.25	**81.94**

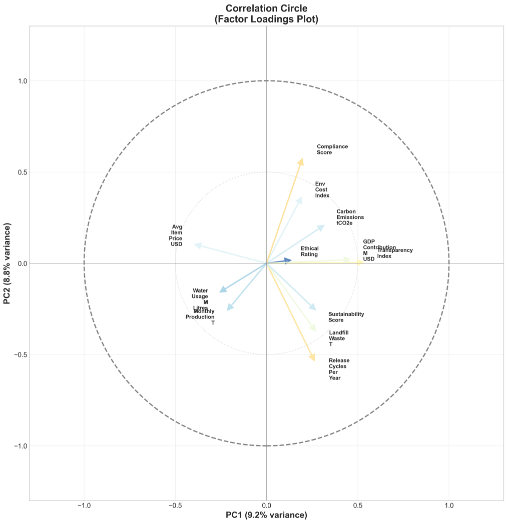
Interpretation

12 PCs retain **81.94%** of variance $\Rightarrow$ Only **18.06%** information loss

Saturation Matrix (PC1 and PC2)

Variable	PC1	PC2
Transparency Index	**+0.480**	+0.272
Carbon Emissions	**+0.426**	+0.017
Avg Item Price	**−0.398**	+0.131
Worker Wage	**+0.352**	+0.093
GDP Contribution	**+0.313**	+0.099
Working Hours/Week	−0.120	**+0.491**
Landfill Waste	+0.250	**−0.405**
Compliance Score	+0.167	**+0.403**
Child Labor	−0.104	**+0.352**
Ethical Rating	+0.128	**+0.310**

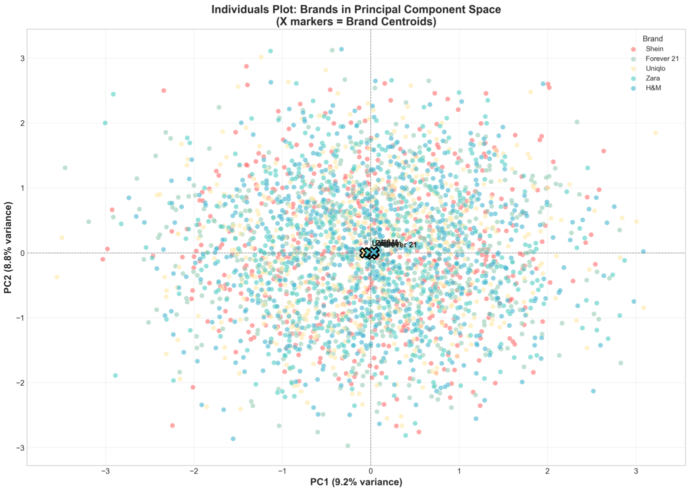
Loadings: Correlations between variables and PCs.
Green: strong positive (> 0.30), **Red**: strong negative (< -0.30)



Interpretation:

- Near edge = well represented
- **PC1**: (+) Transp., Carbon; (−) Price
- **PC2**: (+) Hours, Compl.; (−) Waste

Individuals Plot: Brands in PC Space



Points = observations, X = centroids. Brands cluster differently.

Biplot: Variables and Individuals

Red arrows = variables, Points = observations. Combines correlation circle & individuals plot.

Principal Component 1 (7.43% variance)

Top Contributing Variables

- **Positive**: Transparency (+0.480), Carbon Emissions (+0.426), Worker Wage (+0.352)
- **Negative**: Item Price (–0.398)

Interpretation: *Transparency vs. Price Axis*

- High PC1: High transparency, higher carbon, better wages, **lower prices**
- Low PC1: Lower transparency, **higher prices**
- **Insight**: Trade-off between transparency and pricing

Principal Component 2 (7.26% variance)

Top Contributing Variables

- **Positive**: Working Hours (+0.491), Compliance (+0.403), Child Labor (+0.352)
- **Negative**: Landfill Waste (–0.405)

Interpretation: *Labor Conditions Axis*

- High PC2: Longer hours, higher compliance, more child labor
- Low PC2: Higher waste generation
- **Insight**: Contrasts labor intensity with waste management

Brand Positioning Analysis

Observations:

1. **Clustering**: Brands form distinct groups
2. **Overlap**: Some similar profiles
3. **Outliers**: Deviations from centroids

Implications:

- Brands differ in practices
- Within-brand variability
- PCs separate by transparency & labor

Key Finding

PC1 = *economic transparency*, PC2 = *social conditions*

Quality of Representation

Quality of Representation (QLT)

Measures how well individuals are represented in reduced space:

$$\text{qlt}(\text{ind. } i, \text{axes } 1{-}k) = \frac{\sum_{j=1}^{k} f_{i,j}^2}{\sum_{j=1}^{p} f_{i,j}^2}$$

For 12 components:

- Cumulative QLT: **81.94%**
- Most observations well represented
- Only **18.06%** in higher PCs (13–15)

Implication

Reduced space captures majority of variability

Summary of Findings

1. **Dimensionality Reduction Success**:
   - 15 variables $\rightarrow$ 12 PCs
   - Retained **81.94%** variance
   - Simplified complex data
2. **Identified Key Latent Factors**:
   - PC1: Transparency vs. Pricing
   - PC2: Labor conditions
   - Other PCs: Other dimensions
3. **Brand Differentiation**:
   - Brands in different PC regions
   - Multi-dimensional comparison

Advantages of PCA

For This Analysis

- **Noise reduction**: Filters redundancy
- **Visualization**: 2D/3D plots of high-dim data
- **Multicollinearity removal**: Uncorrelated components
- **Compression**: Fewer variables, minimal loss

Practical Applications

- Clustering (brand segmentation)
- Feature selection for models
- Identify sustainability drivers
- Benchmarking & tracking

Limitations

Assumptions and Constraints

1. **Linearity**: Assumes linear relationships
   - Non-linear patterns not captured
2. **Variance as info**: Prioritizes high variance
   - Low-variance patterns may be discarded
3. **Interpretability**: Linear combinations
   - Less intuitive than original variables
4. **Outlier sensitivity**: Extreme values influence results
   - Standardization helps but doesn't eliminate

Pedagogical Lessons from PCA

Key Takeaways

1. **Eigenvalue decomposition**: Linear algebra application
   - Eigenvectors = new axes, Eigenvalues = importance
2. **Variance interpretation**: Statistical perspective
   - OQE & cumulative variance guide selection
3. **Matrix operations**: Transformations
   - Change of basis: $F = M_s \cdot P$
   - Saturation: $S = P \cdot D^{1/2}$
4. **Visualization**: Geometric interpretation

Conclusion

Main Results

- Applied PCA to fast fashion environmental data
- 15 to 12 dimensions with minimal loss
- Latent factors: transparency-pricing & labor
- Brand differences visualized

Future Directions

- Cluster analysis on PC scores
- Time-series: brand evolution 2015–2024
- Predictive modeling with PCs
- Non-linear methods (t-SNE, UMAP)

References

- ST2DA Course Materials (2025–2026): Chapter 2 – Principal Component Analysis. EFREI Paris.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065), 20150202.
- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459.
- True Cost of Fast Fashion Dataset. Environmental and sustainability metrics from major fast fashion brands (2015–2024).

Thank You

Questions?

ING2-NEW

ADIMI Dalil, BEN AMOR Hassen, DACHRAOUI Elyes,
TIR Ayoub, AZGOUNE Mustapha

ST2DA-I2 | 2025–2026
EFREI Paris