

# Analysis of Online Shopping Behavior

Despoina Iapona & Gerard Palomo

2024-12-08

## Contents

<b>Multidimensional Scaling</b>	<b>1</b>
<b>Advantages of MDS:</b>	1
<b>Challenges:</b>	1
Distance Metrics	2
Why the Right Distance Metric Matters:	2
Comparison of Distance Metrics	2

## Multidimensional Scaling

Multidimensional Scaling (MDS) is a powerful technique used to visualize complex data by arranging points in a low-dimensional Euclidean space. Unlike Principal Component Analysis (PCA), which works on raw data, MDS operates on a **distance matrix**, making it suitable for various types of data, including binary, categorical, and quantitative.

The goal of MDS is to find a configuration of points in a lower-dimensional space that best preserves the pairwise distances from the original distance matrix. This allows us to visually explore the relationships between observations in a way that is easier to interpret.

### Advantages of MDS:

- Works with any type of data as long as a distance measure can be computed.
- Provides a clear visual representation of complex relationships in the data.

### Challenges:

- Interpreting the principal coordinates can be more difficult than in PCA.
- MDS can be computationally expensive for large datasets.

MDS is particularly useful when we only have distance information and want to understand the structure of the data without needing to rely on raw feature values. In this section, we will apply MDS to our dataset to uncover its underlying patterns.

## Distance Metrics

When applying Multidimensional Scaling (MDS), selecting the appropriate distance metric is crucial because it directly impacts the MDS configuration and how relationships between data points are interpreted.

Different distance metrics capture different aspects of similarity or dissimilarity, which leads to varying representations of the data in the low-dimensional space. An inappropriate distance metric can distort these relationships, resulting in misleading conclusions.

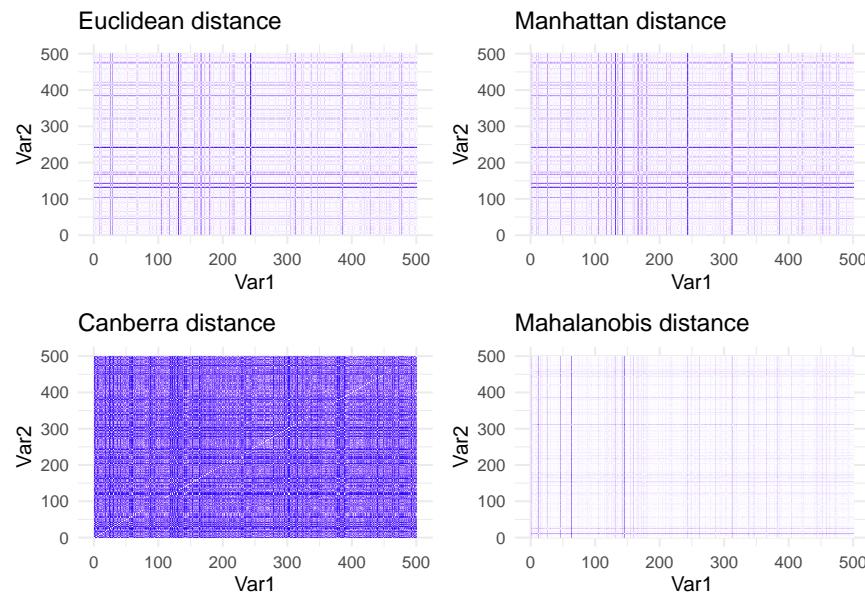
For example, using **Euclidean distance** on a dataset with both quantitative and qualitative variables may not reflect the true dissimilarities, as it assumes continuous and scale-invariant data. In such cases, a metric like **Gower's distance**, which can handle mixed data types, would be more appropriate.

### Why the Right Distance Metric Matters:

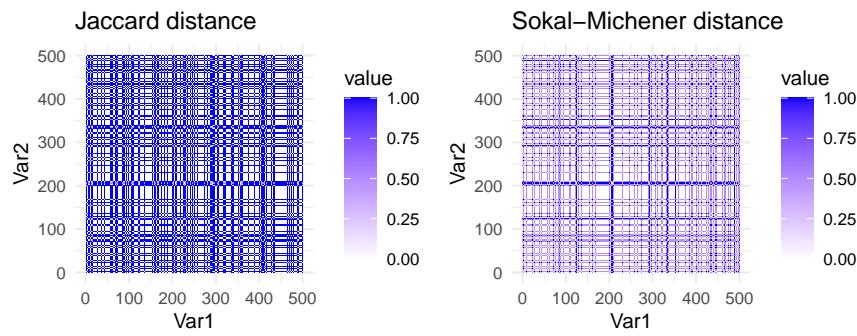
- **Data Type Compatibility:**
  - **Euclidean distance** is ideal for continuous, quantitative data.
  - **Matching coefficients** are used for binary data.
  - **Gower's distance** is versatile, handling quantitative, binary, and categorical data.
- **Scale Invariance:**
  - **Euclidean distance** is **not scale-invariant**, meaning larger-scaled variables influence the distance more.
  - **Mahalanobis distance** is **scale-invariant**, adjusting for variance and correlations between variables.

### Comparison of Distance Metrics

#### Continuous Data



## Binary Data



## Categorical Data

