HYPERSOLID

User Guide

Author: Laurent Oliversen

Version: 2026-06-28

contact @ hypersolid.app

1. Hypersolid User Guide

---
title: Hypersolid User Guide
author: Laurent Oliversen
copyright: Copyright 2026, Laurent Oliversen. All rights reserved.

This web application is a tool for exploring polytopes and analyzing multidimensional datasets, all powered by an intuitive 3D interface.
This guide explains the mathematical concepts and features of the Hypersolid application.

You can download this documentation in PDF format on the author's website:
https://www.laurentoliversen.com/docs/hypersolid-user-guide-en.pdf

1. Geometric Core

1.1. Solid Gen

The application generates geometric shapes defined by their vertices and edges in an N-dimensional space. The application supports the following shapes. Note that spheres are treated as polytopes for visualization purposes, even though they are technically smooth manifolds.
The generation and projection pipeline consists of the following steps:
- Polytope Generation (100% CPU Maths)
- ND-4D: normalization, rotation, shearing, projection (CPU Maths)
- 4D-3D: rotation, shearing, projection, normalization, scaling (CPU Maths)
- 3D-2D/1D: collapse onto a plane or a line (CPU Maths Filter)
- Scene Assembly (Construction of spheres, lines, and faces)
- Camera / GPU Rendering (Calculation of the final angle and screen display)

The order of transformations is indeed as follows: 1) rotation, 2) shearing, 3) projection.

The dimensional axes are denoted as follows:
- X, Y, Z for the usual 3D axes (right, up, and forward-oriented)
- W for the fourth dimension
- V5, V6, V7, V8, ... for all higher dimensions

1.1.1. Polygons (2D)

- Triangle: 2D simplex (3 vertices). The simplest regular polygon.
- Square: 2D hypercube (4 vertices). Regular polygon.
- Pentagon: regular polygon with 5 vertices.

- Hexagon: regular polygon with 6 vertices.
- G2 Roots: structure based on the vectors of the G2 root system.
- Circle (1-Sphere): the 2D circle.

1.1.2. Polyhedra (3D)

- Tetrahedron: 3D simplex (4 vertices). The simplest regular polyhedron.
- Cube (Hexahedron): 3D hypercube (8 vertices). Regular hexahedron.
- Octahedron: 3D orthoplex (6 vertices). Dual of the Cube.
- Dodecahedron: regular polyhedron with 12 pentagonal faces.
- Icosahedron: regular polyhedron with 20 triangular faces.
- Cuboctahedron: Archimedean solid with 8 triangular faces and 6 square faces.
- Rhombic Dodecahedron (Granatohedron): Catalan solid (dual of the Cuboctahedron) with 12 rhombic faces.
- Sphere (2-Sphere): usual 3D sphere. Treated as a polytope for visualization, using either nested layers (stratification) or a wireframe mesh.

Note: the sphere has two additional settings:
- Parallels: number of parallels appearing on the sphere
- Meridians: number of meridians appearing on the sphere

1.1.3. Polychora (4D)

- Pentachoron (5-Cell): 4D simplex (5 vertices). The simplest 4D polychoron.
- Tesseract (8-Cell): 4D hypercube (16 vertices). The 4D analogue of the cube.
- Hexadecachoron (16-Cell): 4D orthoplex (8 vertices). Dual of the Tesseract.
- Icositetrachoron (24-Cell): 4D polychoron with no 3D analogue. Self-dual.
- Hecatonicosachoron (120-Cell): 4D polychoron with 120 dodecahedral cells.
- Hexacosichoron (600-Cell): 4D polychoron with 600 tetrahedral cells.
- Glome (3-Sphere): 4D analogue of a sphere. Treated as a polytope for visualization, using either nested layers (stratification) or a wireframe mesh.
- 3-Ellipsoid (4D Ellipsoid): 4D analogue of an ellipsoid. Treated as a polytope for visualization, using either nested layers (stratification) or a wireframe mesh.

Sphere/Ellipsoid Settings

The glome and the 3-Ellipsoid feature additional settings:
- Tracing Mode: representation mode of the glome. Three main families: Mesh, Stratified (nesting) and Hopf Fibration (special case of the glome). Meshes and stratifications come in two varieties: Edges (edges like polytopes) and Circles (curve-generating functions)

---- Mesh Circles: representation based on parallels, meridians, and hypermeridians. Each component is a circle (deformed for hypermeridians).

---- Mesh Edges: same as above, but the mesh is built with edges, thus allowing vertex localization and color management. However, the curvature of the circles is lost.

---- Stratified Circles: nesting of 2-spheres; nesting is another way to represent the fourth dimension.

---- Stratified Edges: same as above, but each 2-Sphere is built with edges (and vertices).
---- Hopf Fibration: Hopf fibration, another way to represent the partition of a glome from intertwined great circles that do not touch. Note: for the ellipsoid, it is renamed Pseudo-Hopf because it is not truly a fibration, but an a posteriori deformation of the true Hopf fibration.
- H-Mode: representation of hypermeridians
---- Magnetic
---- Interlaced
---- Tilted
---- Equatorial
- Parallels: number of parallels appearing on the glome
- Meridians: number of meridians appearing on the glome
- Hypermeridians: number of hypermeridians appearing on the glome (only for Mesh Circles and Mesh Edges modes).
- Shells: number of shells (2-Spheres) constituting the glome sections (only for Stratified Circles and Stratified Edges modes).
- Shell Distance: distribution of shells within the spherical envelope. The larger the value, the more the inner shells pack against the outer shell. The smaller the value, the further apart the shells are (outermost shells contract towards the center).

Ellipsoid Settings
Finally, ellipsoids have the following scale settings: Scale X, Scale Y, Scale Z and Scale W, which allow stretching the sphere in the desired dimensions.

Note: only spheres/ellipsoids with mesh-type tracing modes are eligible for tessellation.

1.1.4. N-Polytopes (N > 4)

For all generalized polytopes, a Dimension slider allows choosing the number of spatial axes.
- N-Simplex: generalized simplex in N dimensions (N+1 vertices).
- N-Cube: generalized hypercube in N dimensions (2^N vertices). Coords are permutations of (±1, ..., ±1).
- N-Orthoplex: generalized cross-polytope in N dimensions (2N vertices). Coords are permutations of (±1, 0, ..., 0).
- N-Demicube: polytope obtained from a hypercube by removing every other vertex.
- En Roots (E8 roots): structures based on the vectors of the En root system (base for generating Gosset polytopes).
- Gosset K21: exceptional Gosset polytopes up to 8 dimensions.
- N-Sphere: spheres in N dimensions.
- N-Ellipsoid: ellipsoid in N dimensions.

Sphere/Ellipsoid Settings
N-Spheres and N-Ellipsoids feature additional settings compared to dimension 4:
- Alternate H-Modes: only available for Mesh Circles tracing in dimension > 4. This option

alternates the hypermeridian representation mode (magnetic, interlaced, tilted, equatorial) depending on the dimension.
The Tracing Mode replaces the Stratified Edges/Circles entries with a new Hybrid mode combining meshes (mesh circles) and stratification (stratified circles). This mode has specific settings:
- Suspension Depth: dimensional stretching then polar contraction; this factor models the number of dimensions represented by a glome interconnection (hypermeridian).
- Glomes: number of glomes per dimensional suspension.
- H-Mode: representation mode of the interconnections (hypermeridians) between glomes in a suspension:
---- Radial
---- Peripheral
---- Helical

Ellipsoid Settings
Finally, ellipsoids have additional scale settings in addition to X, Y, Z and W (when dimension > 4):
- Scale V Base: stretching in the 5th dimension. Serves as a reference value for all dimensions > 4.
- Scale V Factor: exponential stretch factor in dimensions > 5.

1.1.5. Special

A few non-conventional solids can be selected:
- Elysion Vault: A pointed 3D shape obtained from 4 pyramids, used as a logo by Laurent Oliversen, creator of Hypersolid.
The Elysion Vault has two specific settings, Vault Gap X and Vault Gap Y, which allow modifying the spacing between the pyramids.

1.1.6. Tessellation

1.1.6.1. Tiling Types

To fill the space by replicating polytopes, you can choose between different arrangement modes:
- Prism
- Honeycomb
- Swarm
- Custom nets (special networks)

1.1.6.2. Prism

Prisms are defined by a base polytope replicated in a single direction. The corresponding vertices of each replica are connected to one another.

- Nb Slices: number of replicas
- Flat Base: the polytope serving as the prism base is flattened in the selected plane
- Tile Rotation: each replica rotates in a plane orthogonal to the shear direction; rotation proportional to the distance to the origin
- Rotation Factor: amplitude of the rotation, as a function of the distance to the origin
- Slice Shear X/Y/Z: this is the spacing between the replicas along the X/Y/Z axes

1.1.6.3. Honeycomb

Only certain shapes are eligible for "honeycomb" tiling, such as cubes or hexagons. Unlike other arrangements (prisms and swarms), these tilings are compact and fill space perfectly without gaps.
- Nb Slices X/Y/Z: defines the number of slices in the X/Y/Z directions.

2D tilings (of the XY plane) have two additional settings to enter the third dimension (disjoint sets of slices in the Z direction).
- Gap Z: allows spacing the tiling slices along the Z axis.
- Nexus Z: allows connecting the corresponding vertices of replicas belonging to different Z slices.

1.1.6.4. Swarm

Swarms are a generalization of prisms and honeycombs. They are the most flexible and customizable. They consist of disjoint slices of replicas in all directions, which can be connected by nexus (edges linking corresponding vertices of each replica).
In terms of settings, swarms combine those of prisms and honeycombs. In other words:
- Nb Slices X/Y/Z: number of slices in all directions (same as honeycomb)
- Nexus X/Y/Z: a connection system in all directions
- Gap X/Y/Z: spacing between slices is managed in all directions

In addition, there are the settings specific to prisms: Slice Shear X/Y/Z, Flat Base, Tile Rotation, Rotation Factor.

1.1.6.5. Special (Specific Net)

Some polytopes have predefined nets that can be selected. For example:
For cubes:
- Tesseract Net: 3D cross made of N-cubes.

For the Elysion Vault:
- Elysion Seal: circular assembly of Vaults
- Elysion 2-Seal: double circular assembly of Vaults
- Elysion 3-Seal: triple circular assembly of Vaults

1.2. Solid Style

The application uses an automatic coloring system to help identify dimensions.

Edges are colored based on their alignment with the main axes in the original high-dimensional space.
By default:
- X: Green
- Y: Blue
- Z: Orange
- W: Purple
- E5, E6, E7, E8: Gray

Note that these colors are determined systematically for structures with edges parallel to the frame axes. This applies to hypercubes. For figures with edges inclined in multiple dimensions, color management is more subtle.

Stylistic settings are organized into the following sections: Vertices, Edges, Faces, Depth Effects, Background, and Layout.

1.2.1. Vertices

1.2.1.1. Basic Style

Vertices have 3 settings:
- Size
- Shape (circle, square, or cross shape)
- Color

1.2.1.2. Vertex Depth

Activating depth effects is done by checking the Apply Depth to Vertices box. You can then control the opacity ("Alpha") of geometric elements based on their depth in 4D/3D space. These commands are located in the $Depth Effects$ interface panel. Depending on the drawn object, depth is evaluated differently:
- Depth Quantile: Discretizes distances from the origin into a set of bins (e.g., 10 quantiles), rendering them in distinct alpha layers via an X-Ray approach to avoid visual collisions (Z-Fighting).
- Outer Alpha: Base opacity for the outermost elements of the objects.
- Inner Alpha: Base opacity for the innermost elements (the core) of the objects.

Note: The system automatically creates a linear opacity gradient between outer and inner elements. If Outer Alpha < Inner Alpha, the gradient is disabled and the Outer Alpha value is applied uniformly to all depths.
- Invert Alpha: a mathematical toggle that inverts the opacity. Useful for hiding the exterior to observe only the solid core.
- Alpha Cutoff: a strict rendering threshold applied after the final alpha (and inversion) calculation. Any element whose final opacity falls below this value is entirely removed from the rendering pipeline (hidden), improving visual clarity and rendering performance.

1.2.2. Edges

Edges have 2 settings: Width (thickness) and Color.
The Edge Colors checkbox allows defining a color per edge type. If unchecked, all edges carry the color defined by Edge Default Color.
If checked, the following setting allows managing colors for edges whose orientation is not parallel to the axes.

1.2.2.1. Color Tie Break

You can define the 5 colors associated with each frame axis:
- X Axis, Y Axis, Z Axis, W Axis, Ei Axis (i > 4).

Depending on certain objects, these colors are used differently (typically for spheres).

The Color Tie Break list:
When an edge is not perfectly aligned with a single axis (e.g., diagonal edges of a triangle), the color is determined by a tie-breaking rule (choice between 3 rules):
- Vertex: Color based on the index of the start/end vertex.
- Edge: Color based on the index of the edge in the list.
- Hash: Consistent random color based on the properties of the edge.

This rule applies to most polytopes. The rendering results are hard to anticipate, but they generally allow distributing colors in different ways.

For spheres, other techniques are available:
For spheres/ellipsoids in Mesh/Stratified Edges mode:
- Circle: only; reproduces the same color-by-circle-type system used by Mesh/Stratified Circles tracings.

For spheres/ellipsoids in Hopf Fibration mode (complex, quaternionic, octonionic):
- Fiber: assigns a color per circle topology (associated with different imaginary numbers).
- Polar: assigns a gradient of 3 colors (from the Axis X/Y/Z colors) to distinguish circles close to the north pole, south pole, and equator.
- Bipolar: assigns two gradients of 3 colors (from the Axis X/Y/Z/W/V colors) to distinguish circles close to the north pole, south pole, and equator.
- Tripolar: same with three gradients.
- Quadripolar: same with four gradients.

1.2.2.2. Sphere Outlines

Since spheres are generated differently, they use a specific coloring logic depending on their Tracing Mode:
Mesh Edges/Circles cases: The mesh colors are defined by the colors in the Edges section:
- Parallels: color defined by X Axis

- Meridians: color defined by Y Axis
- Hypermeridians: color defined by Z Axis

Hopf Fibration Case

This advanced visualization uses the Hopf map to project the 3-Sphere (S^3}) as a bundle of circles (called "fibres") over a 2-Sphere (S^2).
- Base: A 2-Sphere where each point corresponds to a full circle (fiber) in the 3-Sphere.
- Fibers: Disjoint circles filling 4D space. A scale of 12 fibers allows visualizing the structure without clutter.
- Coloring: complex Hopf fibers (4D) are colored based on their position on the base 2-Sphere (longitude/latitude), revealing the twisting structure. They use only the colors defined by X Axis and Y Axis. Quaternionic and octonionic Hopf fibers use color systems based on circle topology (either related to the mathematical structure or their proximity to the poles).

Stratified Mode (Equator & Poles)

In this mode, the 3-Sphere is decomposed into a series of parallel 3D shells (a collection of classical spheres), in the same way a globe is sliced by latitude lines. The colors depend on the shell's position in the nesting.
- *Equator (Outer) Shell (outer equatorial shell = widest geometry)*: The "surface" of the 4D sphere. Represented by X/Y. Geometrically, this is the largest shell. The parallels and meridians of this sphere use the colors defined by X Axis and Y Axis.
- *Pole (Inner) Shell (inner polar shell = smallest geometry)*: The "core" of the 4D sphere. Represented by Z/W. Geometrically, these are points or very small circles. The parallels and meridians of this sphere use the colors defined by Z Axis and W Axis.
- *Intermediate Shells*: the colors of non-extreme shells are interpolated between the previous two: this is a gradient between the inner and outer shells.

Note: Stratification Artifact in Perspective Mode

Why do inner shells sometimes appear larger than outer shells?

When using Perspective Projection, objects closer to the 4D camera appear significantly larger. If the 4D Distance is small (camera close to the object):
- An *intermediate shell* (which is geometrically smaller than the equator) happens to be closer to the camera.
- The perspective magnification can make this inner shell appear *visually larger* than the outer equatorial shell, potentially to the point of "swallowing" it.

The colors remain correct (as they are based on rank):
- The shell colored in X/Y is always the geometric equator (Rank 1).
- The shell colored in Z/W is always the geometric pole (Rank N).

If you see a Z/W shell appearing larger than an X/Y shell, it is a pure optical illusion caused by 4D perspective. Solution: modify the 4D Distance or switch to Orthographic Projection to restore the intuitive visual hierarchy.

1.2.2.3. Edge Depth

Activating depth effects is done by checking the Apply Depth to Edges box.

Edges have the same settings as vertices to manage depth: Depth Quantile, Outer Alpha, Inner Alpha, Invert Alpha, Alpha Cutoff. See vertex depth settings for more details.

1.2.2.4. Depth in Spheres

A distinction is made between edges (straight lines) and shells (curved shells).

Objects with edges (polytopes and spheres using Mesh Edges tracing mode)
Opacity is calculated by vertex/edge distance relative to the origin *in the projected 3D space* (not the original 4D space). Distances are sorted and divided into N quantile bins. Each bin receives an alpha value linearly interpolated between Outer Alpha and Inner Alpha.

Note on the depth of the 2/3-Sphere:
Even though all points of a 3-sphere are equidistant from the 4D origin by definition, the 4D→3D projection breaks this isotropy. In *perspective* projection, circles close to the W-pole (large W) are compressed towards the 3D origin, while equatorial circles (small W) are projected further away. In *orthographic* projection, a parallel near the north pole lies close to the 3D origin, while an equatorial parallel lies at a unit distance (the largest). The depth gradient on a Glome in *Mesh Edges* mode is thus real and physically significant: it reflects the 4D depth structure as seen through the projection.

Spheres with shells (Stratified Edges and Stratified Circles)
Depth is structurally evaluated based on the mathematical index of the shell (from 1 for the geometric equator to N for the pole). Since shells are naturally nested topologies, the alpha gradient is continuous and exact. Note: V/E Depth Quantile (depth quantiles) do *not* apply to shells.

Hopf Fibration: shell-based depth (without edges)
The Hopf Fibration is rendered as a set of topologically nested *fiber circles* (closed loops), rather than discrete edges between vertices. Each circle belongs to a mathematical shell (a torus layer), numbered from the outermost torus to the innermost central fiber. The depth alpha is computed directly from the shell index; no 3D Z-distance sorting is needed. The formula is the same as the one used for Stratified/Shell objects. Since Hopf fibers are generated in a strict topological order (exterior → interior), the mathematical structure inherently encodes depth. This makes depth management exact, continuous, and perfectly consistent with the Inner/Outer Alpha controls, without any geometric heuristics.

Limitations - Mesh Circles only
In the *Mesh Circles* tracing mode for the 2/3-Sphere, all circle geometry is merged into a

single combined mesh without a per-circle depth shader. Depth alpha does *not* apply to circles, only to vertices. In contrast, the *Mesh Edges* mode for Spheres *does support* depth because a GPU shader evaluates depth = distance per vertex at render time.

1.2.3. Faces

The Show Faces toggle overlays filled polygons onto the polytope's wireframe. The engine supports two fundamental modes for detecting and generating faces:

1.2.3.1. Face Detection Methods

The application uses specialized algorithmic generators to identify the true geometric faces of the polytope. The method used depends on the polytope family. The techniques used are:
- Graph Cycle Detection (DFS): used for standard 3D/4D regular solids, the $N-Simplex$ and the $N-Cube$ (up to $N=5$). A Depth-First Search explores the edge graph to identify closed loops of 3, 4, or 5 edges.
- Axes Combinatorial Search: specific to the $N-Orthoplex$ (N-Cross) family. It identifies triplets of vertices located on distinct orthogonal axes to form perfectly regular triangular faces. This mode is active up to $N=6$.
- Parity Subspace Logic: specific to the $N-Demicube$ ($Dn\ Group$) family. It identifies the 3D cubic subspaces of the parent hypercube and extracts the tetrahedral faces that preserve the alternating parity of the vertices. This mode is active up to $N=5$.
- Stochastic Sampling (in other cases): for complex exceptional polytopes (Gosset $E6, E7, E8$) or any object exceeding the dimension thresholds mentioned above, topological detection is automatically bypassed in favor of random sampling to maintain interactive performance.

1.2.3.2. Face Rendering

Faces have different rendering types:
- Solid: faces are triangulated and rendered as filled surfaces.
- Wireframe: only the border edges and internal diagonals of the identified faces are drawn. This reveals the "tiling" of the polytope.
It is also possible to choose the face color, its transparency (Face Alpha), and the line thickness for the Wireframe mode.

1.2.3.3. Face Sampling

For complex or high-dimensional objects where topological detection is too costly, the engine can stochastically sample pairs of vertices to form triangles (or quadrilaterals/pentagones).
In the following paragraph: a face is said to be *small* (respectively *large*) when the vertices defining it are, on average, at a distance *less than* (respectively *greater than*) a given value. The notion of distance refers to the Euclidean distance between a vertex and

the polytope origin (the barycenter in most cases).

The associated settings are:
- Large/Small Cutoff: this parameter is a normalized distance used to classify faces into two categories: small or large. This threshold is between 0 and 1 because it represents the percentage of the largest distance observed among all faces. For example, a threshold of 0.5 means that the vertices of a face must be, on average, at a distance (to the origin) *greater than* (respectively *less than*) 0.5 of the maximum distance (the face considered "furthest") for the face to be considered *large* (respectively *small*).
- Large Faces: number of faces classified as *large* that will be retained (sampled) for coloring.
- Small Faces: number of faces classified as *small* that will be retained (sampled) for coloring.

Note for Regular Polytopes: For perfectly regular polytopes (Cubes, Tesseracts) where all vertices are equidistant from the barycenter, the "Small" population is mathematically empty because the normalized distance threshold is constant. In this case, only "Large" face sampling will produce results.

1.2.3.4. Face Depth

Activating depth effects is done by checking the Apply Depth to Faces box.
Z-Sorting & Alpha (Painter's Algorithm)
To apply depth to faces, the engine dynamically reorganizes the faces from back to front at each frame. This enables depth-based transparency gradients:
- Front Alpha: Opacity of the faces closest to the camera (the "skin").
- Back Alpha: Opacity of the faces furthest from the camera (the "core").

Example usage: setting Back Alpha = 1.0 and Front Alpha = 0.2 allows you to see the solid interior of the structure through a translucent outer layer.

Coverage by Polytope Type:

✅ Platonic Solids (3D): triangles, squares, pentagons

✅ Archimedean / Catalan (cuboctahedron, rhombic dodecahedron): mixed polygon types

✅ 4D Polychora: via graph cycle detection

✅ N-Simplex / N-Cube / N-Orthoplex: up to available resolution

❌ Sphere / Glome: Not applicable, rendered as circles

✅ Gosset K.21 dim=5 (D5): triangles + squares, manageable count

❌ Gosset K.21 dim≥6 (E6, E7, E8): Gosset polytopes E6, E7, E8 possess a number of 2-faces (planar faces) that grows exponentially. For E6, face calculation is fast, but projecting 2160 overlapping faces from 6D creates an unreadable opaque mesh with no visual benefit. For E7 and E8, the calculation would take several seconds and the result is just as unreadable. Only D5 (dim=5, ~40 vertices) exposes its faces, as a practical compromise between visual information and performance.

1.2.4. Background

This section allows displaying background or overlay elements.

1.2.4.1. Background Color and Gradient

To configure the background color:
- Color
- Use Gradient: allows defining a gradient with 3 parameters: Gradient Top, Gradient Bottom and Gradient Angle.
- Transparent: removes the background (transparency).

1.2.4.2. Grid

The Grid option displays a grid tangent to peripheral vertices, useful for adjusting alignments during rotations, for example.

1.2.4.3. Axes

Axes help with orientation in spatial dimensions.
- Show Scene Axes: displays the axes of absolute 3D space (as seen by the camera), |X|, |Y|, and |Z|. They are enclosed in vertical bars to avoid confusion with the relative axes below.
- Show Relative Axes: displays the relative axes, which are bound to the polytope and any projected data points. When a shear or rotation is performed, these axes shift and rotate along with the polytope and the data. Note that they no longer represent absolute spatial coordinates, but rather the coordinates after projection/transformation: X, Y, Z, W, V5, V6, ..., V10.
- Negative Extension: extends the axes to negative values as well.
- Negative Arrow: adds an arrow tip to the negative side of the axes.

The appearance of absolute (scene) and relative (solid/data) axes can be customized using:
- Axis Length: length of the axes (arrows).
- Axis Label Size: size of the labels (X, Y, Z...).
- Axis Arrow Size: size of the axis arrowheads.
- Axis Ticks: displays unit tick marks (coordinates 1, 2, 3...) for scene absolute axes and relative data/polytope axes. The ticks differ per axis, since relative coordinates are transformed by spatial projection.

You can also display a frame plane:
- Plane |XY|, Plane |XZ|, Plane |YZ|: displays the planes of the absolute frame (scene axes).

1.3. Projection

Projection is performed in this order:
- Projection of dimensions higher than 4 to the 4D space (X, Y, Z, W)
- Projection of the 4D space to 3D (X, Y, Z coordinates incorporate the remnants of the W coordinate through perspective effects)
- Optional collapse (orthographic projection) from 3D to 2D
- The entire scene is displayed in a 3D scene through the camera
In the interface, settings are categorized by mathematical operations: rotation, shear, projection (with focal distance). But conceptually, dimensions are processed in cascade.

1.3.1. Views

This section allows applying specific mathematical visualizations. These rotation values reveal characteristic projections for each polytope.
The views below capture the following elements:
- Topological: Connectivity graph (vertices, edges)
- Combinatorial: Incidences between k-faces
- Metric: Dihedral angles, relative lengths
- Global: Periodicity, envelope, stratification

1.3.1.1. Unified Projection Mode

The two buttons below allow setting all projection modes (ND-4D, 4D-3D, and Camera) at once:
- All Perspective: Unifies the entire pipeline (ND Projection, 4D Projection, and Camera) in *Perspective* mode. This guarantees a consistent focal depth throughout dimensional reduction.
- All Ortho (orthographic): Unifies the entire pipeline in *Orthographic* mode, revealing true geometric proportions without distortion.

1.3.1.2. Front and Petrie Projection

These two projections are respectively the simplest and the most complex to implement. They have a dedicated button for immediate access.
- Front: resets all rotation angles and shears (ND and 4D) to 0, and restores the camera to its default frontal orientation. This is the starting point of any visual analysis.
- Petrie: Configures rotations and projection to show the *Petrie Polygon* of the polytope. This is a specific orthographic projection where the projected polygon of the polytope forms a regular N-sided perimeter. The axis is aligned with the polygon that forms the longest Hamiltonian cycle zigzagging between all facets. This is the only view that reveals the global structure: it shows large-scale periodicity, which is invisible in local views. It remains unique in that it captures Hamiltonian periodicity, with no direct equivalent elsewhere.

1.3.1.3. Canonical Planes

Navigate between different canonical planes by clicking the Prev (previous) and Next buttons. They include:
- Vertex-first: the axis passes through a vertex. Reveals local valence and the point symmetry group: how many edges radiate, and with what angular symmetry.
- Edge-first: the axis is aligned with an edge. Reveals the dihedral angles between the two (N-1)-facets sharing this edge - the "local curvature" of the object. Often overlooked, but irreplaceable for distinguishing polytopes with similar vertex-first views. Note: an edge bounds exactly 2 facets of codimension 1 in a convex regular polytope; in general this number depends on the local structure.
- Cell-first / k-Facet-first: the axis passes through the center of a 3, 4, ..., N-1 dimensional cell or facet. Reveals concentric stratification: how cells/facets organize in successive layers around the center ("onion-like").
- Schlegel: corresponds to an (N-1)-facet-first view in perspective mode. It allows seeing all cells at once.
- Equatorial: corresponds to a Petrie projection followed by a 90° orbit in the vertical plane, as if looking at the Petrie from the edge. It allows seeing the stratification of facets around the barycenter.

Note: with the exception of the Schlegel view, all canonical plane views are only possible in orthographic projection mode.

1.3.1.4. Base Permutations

Permuting the base axes is particularly useful in factor analysis, to visualize the distribution of points along 3 simultaneous axes among the N available axes (X, Y, Z, W, V5, V6...). See the factor analysis section for more details.

Each pre-recorded view corresponds to a 90° rotation along one, two, or three axes. This visually allows replacing a base axis (X, Y, Z) with another available axis.
The 20 most common combinations (up to dimension 6) are offered. The first corresponds to the XYZ space view. The last translates the W.V5.V6 space view.

Navigate between different factor spaces by clicking the Prev (previous) and Next buttons.

To better understand rotations, it is recommended to display the relative axes (Show Relative Axes) which show the orientation of the polytope and the overlaid data.

1.3.1.5. Collapse

Optional 2D/1D flattening step.
Dedicated to the Factor visualization tab, this final mathematical step allows compressing the fully projected 3D scene (Polytope, Zodiac/Astro/MBTI/Social data, and custom data) to a 2D plane or a 1D axis. This is accomplished by forcing one or two dimensions of the 3D

output to zero immediately before rendering.

Plane Projections (2D):
- XY: Emulates a flattened view by flattening depth (z = 0).
- XZ: Flattens the vertical axis (y = 0).
- YZ: Flattens the horizontal axis (x = 0).

Axis Projections (1D):
- X, Y, Z: Flattens two dimensions simultaneously to project the entire structural cluster onto a single line. This is particularly useful for identifying correlations, variance, or standard deviations along a specific principal component.

Note: Since this projection is applied as a final filter on the 3D scene rather than in the original mathematical space, higher dimensions like W (4D depth) or V (5D+) cannot be isolated here - they are already folded into X, Y, and Z by the previous steps. Additionally, to maintain readability when points are mathematically collapsed onto identical spatial coordinates, text labels (Zodiac/Astro/MBTI/Social and custom) apply an intelligent spatial offset (derived from their original lost coordinate) to avoid visual overlapping.

1.3.1.6. Custom Views

Certain solids feature custom views (which can be modified). Navigate between views by clicking the Prev (previous) and Next buttons.

1.3.2. ND to 4D

This step reduces dimensions N > 4 to 4D (X, Y, Z, W).
The cascading pipeline of all higher dimensions is executed as follows:
- Initial rotation in all selected ND planes
- Chain of shears and projections (at each level of the dimensional cascade)

Note: by its non-destructive additive logic, shearing is much more effective than rotation for unfolding higher dimensions into the visible 3D space (projected on the first three axes X, Y, Z). This is because rotations in high dimensions can occur in planes orthogonal to the projection plane of each dimension to the next, compressing any visual unfolding attempt.

The Lock Solid checkbox allows locking the polytope and making it insensitive to rotations (note: the solid remains sensitive to shears and the focal distance of projections). This is useful for analyzing data, revealing through rotation the relative axes of invisible dimensions (W, V5, V6...) without affecting the solid itself. If a cube or sphere is displayed as a reference, it remains stationary during rotations, while the projected data (point cloud) continues to pivot with the relative axes.

1.3.2.1. ND Rotation

- Circular Rotation: performs a simultaneous rotation in XV5, YV6, ZV7, WV8, V5V9, V6V10.
- Rotation Factor: Controls the exponential decay of rotations for higher dimensions.
- Rotation Plane which contains a list of all additional rotation planes. Simply select one, then adjust the slider that appears just below it.

1.3.2.2. ND Shear

- Circular Shear: Adds a rotational geometric component (alternating directions along X, Y, Z) to the linear shear of 4D, creating spiral patterns for dimensions higher than 4.
- Shear Factor: Controls the exponential decay of shear offsets for higher dimensions.
- Shear Plane: selection of a specific plane to perform precise shearing along that plane only.

1.3.2.3. ND Projection

Perspective
Higher dimensions (D > 4) are scaled by a perspective factor. Objects with higher coordinates in these dimensions are projected closer to the 4D center, while applying a cumulative shear.
- Base Distance: Determines the reference distance (from the camera) for the "perspective" of higher dimensions. The higher the value, the closer the outer layers are to the origin (distance from the camera increases, perspective decreases).
- Distance Factor: Adjusts the exponential scale of the perspective distance across dimensions. A high value results in uniform spacing between dimensions. A low value causes wider spacing of higher dimensions.
By manipulating both the Base Distance and the Distance Factor, it is possible to customize the global perspective, as well as the spacing or clustering between dimensions (compression/dilation feel).

Orthographic
Parallel reduction to 4D. The perspective distance and depth accumulation are strictly ignored (W remains unchanged). The structure is revealed purely through linear shear. Focal distance concepts are ignored. Only the shear parameters remain to unfold the structure.

Oblique
Oblique projection is not offered in dimensions higher than 4 because it becomes very complex to configure. An orthographic projection accompanied by a rotation or shear (circular here) is equivalent to an oblique projection.

1.3.3. 4D to 3D

The projection step aims to project 4D points (x, y, z, w) into the 3D space (x', y', z'). The W coordinate represents the "4D depth".
The processing begins with an optional linear 4D shear (Shear XW, Shear YW, Shear ZW) to

the true 3D coordinates based on the W depth. This allows tilting the 4D object into our physical dimension before the final pass of the perspective or orthographic camera.

1.3.3.1. 4D Rotation

Polytope rotation can be performed in all 3D and 4D planes:
- Rotation XY, XZ, YZ, XW, YW, ZW.

1.3.3.2. 4D Shear

Polytope shearing can be performed in all 4D planes:
- Shear XW, YW, ZW: shifts the 3D position based on the 4D depth (W).

1.3.3.3. 4D Projection

Perspective
Standard 4D perspective. Objects with a higher w (further in 4D) appear smaller and converge towards the center.
The 4D Distance adjusts the focal length of the 4D camera. Smaller values create stronger perspective distortion.

Orthographic
Projection parallel to W. The W coordinate is completely ignored for point positions in 3D. Useful for preserving real geometric sizes.
Important: in the absence of rotation or shear, it becomes impossible to visualize the 4D components. You should use the rotation and shear sliders to unfold the structure.

Oblique
Unfolds the 4D structure by shifting points based on their W coordinate along a specific 2D angle on the screen. Geometrically, oblique projection is equivalent to an orthographic projection combined with a shear, but it is offered as a distinct simplified mode, typically to reproduce Cavalier/Cabinet perspectives.
- Angle: Controls the direction of the 4D shift in the 3D XY plane.
- Scale: Controls the amplitude of the shift (how far the 4D depth is "extracted").

Note: Manual Shear XW/YW/ZW is ignored in 4D Oblique mode since the oblique angle and scale fully define the 4D shear. Therefore, the manual shears are redundant.

1.3.4. Camera

The final phase renders the 3D wireframe on your 2D screen. Note that the 2D/1D projection in the Factor tab is performed before the camera rendering. This is why we can still orbit around the polytope and the collapsed points.

1.3.4.1. Camera Type

Camera Types:
- Perspective: Standard 3D camera with vanishing points. Objects far from the camera appear smaller.
- Orthographic: Parallel projection. Objects keep their size regardless of distance. Ideal for technical views.
- Oblique (cavalier/cabinet): A special mode combining orthographic projection with Z-axis (depth) shear. Oblique provides a cavaliere perspective where the front face retains its true shape, while depth is represented by diagonal lines. Two additional parameters appear:
- Angle: Direction of the axis simulating depth.
- Scale: Amplitude of the shift simulating depth.

1.3.4.2. Viewport

Viewport
- Roll: Rotates the entire view (camera roll) around the line of sight.

Note: there is no camera shearing. This would introduce unwanted tilting to the polytopes.

1.3.4.3. Camera Monitoring

Caution: the following information is not located in the Solid Gen / Camera menu, but at the very bottom, in the Solid Gen / Camera Monitoring section.
This section displays the camera coordinates and settings:
- Orbit X (phi angle): orbit in the horizontal plane
- Orbit Y (theta angle): orbit in the vertical plane
- Roll: camera roll (clockwise or counter-clockwise)
- Zoom

This data is purely informative (non-editable). To modify the camera, simply use the mouse or trackpad.

1.4. Motion

$$Only for the Hyper package.$$
The application provides tools to visualize high-dimensional geometry through motion.

1.4.1. Auto-Rotate

Automatically rotates the polytope in up to three simultaneous planes.
- Rotation Speed: Controls the angular velocity. Supports negative values to reverse the direction.
- Rotation Planes 1/2/3: Select the rotation planes to activate (XY, XZ, XW, YZ, YW, ZW, or ND

planes for N > 4).

1.4.2. Auto-Shear

Oscillates the shear factors harmonically to make higher dimensions "pulse" in the view.
- Shear Amplitude: Amplitude of the oscillation.
- Shear Speed: Frequency of the oscillation.
- Shear Planes 1/2/3: Select the shear planes to activate (XY, XZ, XW, YZ, YW, ZW, or ND planes for N > 4).

1.4.3. Interpolation

Performs a smooth transition between global state presets.
- Start State: The initial state of the transition.
- Mid State: An optional intermediate state. If set to *None*, the transition goes directly from Start to End.
- End State: The final target state.

Interpolation Chain:
- Clicking Interpolation first moves the current view to Start State.
- It then proceeds to Mid State (if enabled).
- Finally, it settles on End State.

Non-Linear Transition Speed:
- Easing: enables a "smooth" transition, S-curve.
- Duration (s): duration of the transition.

1.4.4. Oscillation

If oscillation is enabled, the interpolation will not stop at End State. Instead, it will reverse the sequence and loop indefinitely between Start, Mid, and End.

1.5. A Few Projection Examples

1.5.1. Tesseract

Tesseract
The tesseract has a large number of famous projections.
The *two centered hexagons* view (available in the Custom Views of the Projection panel):

* Rotation: (XY, XZ, YZ) = (30°, 35.26°, 135°)
* 4D perspective projection with an orthographic camera projection
* Use the 4D distance to adjust the spacing between the two hexagons

The *double oblique perspective* view (available in the Custom Views of the Projection panel):

* 4D oblique projection with an oblique camera projection

1.6. Config

You can export and later re-import workspace configurations (parameter settings). You can export all at once (Export All Config), or only those relating to:
- Export Solid Config: settings of the styles applied to the solid
- Export Projection Config: settings of the rotations, shears, and projections effects
- Export Data Config: settings of the styles applied to projected data

You can re-import them by clicking Import Config.

2. Data Ingestion

2.1. Dataset

Location: Workspace tab, Dataset panel.

2.1.1. Synthetic Data

These datasets are random or open-source. They allow you to test the application without having to import data.
- 5D Data: data with 5 pseudo-random numerical variables
- 12D Data: data with 12 pseudo-random numerical variables
- Titanic Data: open-source data
- Titanic Categorical Data: same as above, but with discretized data; allows testing MCAs
- Titanic Mixed Data: same as above, but with mixed data; allows testing FAMDs
- MMM Data: data to test the forecast and the MMM
- Hy-Perf Single Data: data to test the Hy-Perf module, with single effect
- Hy-Perf Multi Data: data to test the Hy-Perf module, with multiple effects

2.1.2. Reference Data

These reference datasets are purely cosmetic and serve as visual landmarks.
- Zodiac Data: Astro (the 12 signs) or Malgovert (the 13 psychic clusters from the eponymous novel)
- Semiology Data: 12 neurological circuits conditioning behavior
- MBTI Data: 16 behavioral profiles used in business
- 4-Colors Data: 4 colors (red, blue, green, yellow) used in business

Fiction fact: in the Zodiac category, there is a dataset called "Malgovert" which corresponds to the psychic classification of the thirteen Zodiac clusters in the Malgovert series by Laurent Oliversen. It's somewhat like blood type classification, but with three variables. The points are defined by 3 coordinates representing the sensitivity of the zodiacal profiles to the 3 elysian serums. The spatial position of these clusters is actually used in the fiction.

2.1.3. Axes Mapping

You can assign numerical variables to the spatial axes of your choice.
Simply enter a Variable / Axis combination and click Add.

The Merge T|C columns on same axes option is only available for Hy-Perf data. It allows pooling variables suffixed with _T and those with _C. For example, instead of having two axes to display NB_T and NB_C, the option displays a single NB axis that merges the two populations (T and C). To distinguish the two populations, you must either display the labels, or go to Data Style, and check the Colorize Treatment box (Class Style section).

2.1.4. Custom Points

This section allows you to manually enter up to 10 spatial coordinates. The resulting point is displayed overlaid on the polytope.
Enter up to 10 coordinates (X, Y, Z, W, V5, ..., V10) manually. You can add a label to your point, and you can also duplicate or delete a point.

2.2. Data Import

Location: Data tab.

2.2.1. Import CSV

You can upload CSV data. Depending on your subscribed plan, data may be limited to a certain number of rows and columns (indicated in the interface).
Expected structures are:
- UTF-8 CSV format
- The first row corresponds to headers
- Separator: comma, semicolon, tab, or pipe "|"
- The following column headers are reserved for specific uses: "label" (point name), "segment" (for inferential statistics), and "target" (for supervised machine learning)

By default, if columns x, y, z, w, v5, v6, v7... are present in the files, they are automatically assigned to the corresponding spatial coordinates.
If no label matches the above convention, then numerical variables are assigned in order of appearance in the file: the first numerical column is assigned to X, the second to Y, the third to Z, etc.

In case of a mix (example: a, x, y, z, b, c), then recognized labels (x, y, z) are assigned to the 3 spatial coordinates, and unknown labels (a, b, c) are assigned to the following spatial coordinates in this order: a -> w, b -> v5, c -> v6.

2.2.2. Role Mapping

This section allows defining columns that will play a specific role for analytics. Some roles are interchangeable in certain scenarios.
The columns that can play roles are:
- label: name to be displayed above the point on the projected scene
- color: point and label color
- segment: grouping column used for descriptive or exploratory statistics (profiling)
- target: dependent variable column to be explained, used for variable selection (correlation) and supervised algorithms (LDA, Sup ML)
- treatment: binary column for calculating incremental effects and uplifts between two groups (profiling, Hy-Perf); examples: clinical trials, marketing campaigns...
- flag_test: binary column (0, 1) to define rows used to train supervised models and those for testing phases

2.2.3. Test Split

If the flag_test column is not defined, it can be created and filled automatically randomly by setting a test percentage (Test Size) and clicking on Random Split.

2.2.4. Preview

A data preview is available in the tab.
In the header are two sections.

Counting
- Total rows: number of rows in the imported dataset
- Total columns: number of columns in the imported dataset
- Numeric columns: number of numerical columns
- Categorical columns: number of categorical columns
- Rows retained: number of rows retained according to the user's plan (pricing)
- Spatial dim retained: number of spatial columns retained (according to user's plan) and thus visualizable in the scene

Reserved columns
This list of tags refers to reserved roles. It helps identify columns playing a special role: uid, label, color, target, segment, treatment, flag_test.

Column Typing
For each column, a tag indicates its type: numerical, categorical, or boolean.

2.3. Preprocessing

Location: Data tab.

This section allows performing several transformations that are sometimes required to run algorithms (factorial projections, clustering, Sup ML).

The Keep Original Column checkbox allows keeping the original column alongside the transformed one. Note: this is not always desirable, as it introduces information redundancy.

A Select Column field allows choosing the column to transform. Each transformation is applied and canceled using Apply and Undo buttons.

Finally, a Re-Apply in Inference checkbox allows keeping the settings of these steps to reuse them later on new data; a necessary step when using trained models (Factor, Clustering, Sup ML) on a transformed dataset.

2.3.1. Column Typing

5 buttons are available to change column activation and format.
- Make Numeric: transforms the column into numerical data; projects spatially in the scene
- Make String: transforms the column into qualitative data; does not project
- Make Bool: transforms the column into binary data (true/false, yes/no, 0/1); does not project in the scene; corresponds to dummy variables

String and bool data are very similar in their implication within different features. The only distinction comes from the fact that boolean variables will be properly used in linear and logistic regression models (Sup ML), whereas string data will be ignored.

Using or ignoring the column:
- Activate: allows the variable in the scene and various modeling modules
- Deactivate: excludes the column

2.3.2. Imputation

This section allows filling empty data.
- Num Imputation: value used to fill empty numerical data; choices: value 0, -1, mean of non-empty data, median of non-empty data
- Categ Numeric: string for empty categorical data

2.3.3. Winsorization

This section allows capping extreme numerical values.

- Tail Cut %: cap percentage on high values; a value of 2% means all values above the 98th percentile will be capped.
- Two-Sided: also caps extreme low or negative values

2.3.4. Discretization

This section allows discretizing numerical data into bins. They thus become categorical (qualitative). Discretization is performed into quantiles.
- Nb Bins: number of bins (quantiles)

2.3.5. Category Binning

This section allows grouping categories with low counts that induce sparsity and instability. It's the equivalent of extreme low values for qualitative data.
- Rare Category Label: label for the merged category
- Use Frequency: uses count proportions (rather than absolute counts)
- Count Threshold: threshold value (count or frequency, depending on the used option) to define a rare category that will be merged

2.3.6. One-Hot Encoding

This section transforms a categorical variable into dummy variables. Each category of the variable becomes a binary column (bool, 0/1)
- Keep all categories: keeps all categories and transforms them into binaries; if unchecked, the last category is dropped to avoid multicollinearity.

3. Data Exploration

3.1. Stats

Location: Stats tab.

3.1.1. Descriptive Statistics

Contains descriptive analyses of the data imported into the application:
- Summary table of dispersion metrics: mean, standard deviation, min and max values, quartiles (Q1, Q2, Q3), coefficient of variation (CV), semi-interquartile range (SIQR = (Q3-Q1)/2), midhinge (MH = (Q1+Q3)/2), trimean (TM = (Q2 + MH)/2), quartile coefficient of dispersion (QCD = SIQR/MH)
- Boxplots: for each numerical variable, with the option to display a distinct scale per variable (Different Scales), or a common scale for all variables (Same Scale).
- Category Distribution: count of modalities present in each categorical variable, with a list to choose the variable

3.1.2. Inferential Statistics

You can perform quick statistical tests: comparing means or proportions, as well as sample size estimations.

3.1.2.1. Settings

- Use Case: mean comparison, proportion comparison, statistical power analysis
- Value: for mean comparisons; variable acting as the numerical measure
- Outcome: for proportion comparisons; variable indicating the presence of a success/failure when evaluating the numerator of the proportion
- Group: variable acting as the categorical group

Specific settings for power analysis:
- Design: data type for the analysis: means or proportions
- Solve for: target value of the analysis: sample size or target power
- Cohen's d: desired effect size value (for two means)
- Proportion 1: value of the first proportion
- Proportion 2: value of the second proportion
- N per Group: size of each group to determine power
- Target Power: desired power to determine the size

Common settings:
- Method: parametric (based on known statistical distributions) or non-parametric (robust to outliers)
- Alpha: significance level (type I error) to reject the null hypothesis
- Alternative: one-sided (greater = right; less = left) or two-sided

3.1.2.2. Associated Statistics

The results are presented in two areas:
- On the left, test metrics and some descriptive statistics
- On the right, gauges for power, significance, and effect size; as well as a representation of the confidence interval

3.2. Data Exp (Data Exploration)

Location: Workspace tab, Data Exp panel.

3.2.1. Data Scaling

The Standardize button allows you to center-reduce the distance of the points. Subtraction of the mean followed by division by the standard deviation.

The Normalize button allows you to standardize the distance of points from the origin.
Once activated, an After Projection parameter can be set. It defines when normalization occurs. It only applies to normalization, NOT to standardization (centering-reducing). Two options:
- Unchecked: normalization occurs in the initial N-dimensional space, before any rotation, shear, projection.
- Checked: normalization is performed after the projection cascade, on the resulting 3D space.

It is better to filter "before" the projection to keep maximum information during the projections. However, a posteriori normalization allows obtaining a spread on the unit-sphere of all projected points, which greatly simplifies visual exploration (even if a lot of information has been lost in the process).

Finally, the Data Scale button allows spacing or contracting all projected points. It acts as a multiplier of all final 3D space coordinates, after normalization.

3.2.2. Filtering

The filtering section allows displaying projected points that meet certain conditions.
The first setting is the nature of the condition: AND (intersection) or OR (union).
The AND condition means all following rules must be met simultaneously to display a point.
The OR condition means at least one rule must be met.

3.2.2.1. Search Labels

The first part concerns the labels attached to the points:
- Case Sensitive: perform searches without considering case. "LABEL" is perceived as "label" or "Label".
- Whole Word: perform searches considering the entire label. If checked, the string "Lab" will not match the string "Label".
- Show: manual input of a string included in the label to determine the points to display; you can enter several different words separated by a comma
- Hide: manual input of a string included in the label to determine the points to hide; you can enter several different words separated by a comma

3.2.2.2. Cluster

The dropdown list displays the clusters (if a clustering has occurred) you wish to display.

3.2.2.3. Axis Value Filtering

This part concerns the properties of projected points. It requires selecting one to three axes: Axis 1, Axis 2, Axis 3. If you do not wish to populate them all, you can set them to "None".

You can then choose which values to filter: by sign, by absolute value, or by factorial metric.
- Axis 1 Sign: filter points whose coordinates are either positive or negative on axis 1.
Same for axes 2 and 3.
- Use Coords: allow filtering on coordinates
- Absolute Coords: linked to the previous check; absolute value that a point's coordinates must exceed on all previously defined axes (if AND condition) or on at least one of these axes (OR condition)
- Use Contrib: allow filtering on factorial contributions
- Contrib: same as coords, but applies to the point's factorial contribution value on the axes defined in the previous lists; only applies to "Factor" points, meaning those projected via PCA or MCA, which participated in forming the factorial axes.
- Use Cos2: allow filtering on factorial squared cosines
- Cos2: same as coords, but applies to the point's factorial squared cosine value on the axes defined in the previous lists; only applies to "Factor" points, meaning those projected via PCA or MCA, whether or not they participated in forming the factorial axes.

3.2.2.4. Combined Filtering

Combining the two previous rules
You can define conditions on strings and on axes. In this case, the AND/OR condition applies to the whole.
For example, with an AND condition, a point must respect the inputted strings and respect the values on all defined axes. With an OR condition, a point must respect the inputted strings or respect the values on one of the defined axes.

Influence of automatic projection planes (canonical and factorial)
When filters are active, if you click on an automatic projection plane, namely, the canonical planes (e.g., vertex-first, edge-first, etc. Cf. Views) or the factorial planes (Best Contrib Plan and Best Cos2 Plan), then the resulting axes are sent to the filtering axes.
This allows scanning different canonical and factorial spaces while updating filters based on the new axes displayed on the screen.

3.2.3. Orbital

The orbital is a multidimensional filtering system based on distance. Its name is not an official mathematical term, but a reference to atomistic physics-chemistry that defines the orbital as a zone of space within which an electron can be observed with a certain probability level. Here, the goal is to highlight the proportion of the population located within the zone delimited by the orbital.

In the application, by default, an orbital will hide points close to the origin and leave only peripheral points (outliers). You can reverse the masking logic (Outlier Mode) to display only the central points (inliers).

3.2.3.1. Orbital Types

Three orbitals are proposed:
- Gaussian Ellipsoid: a multidimensional ellipsoid based on normal distributions.
- Box: an n-rectangle whose lengths are defined by percentiles in each dimension; the Quantile X/Y/Z/W/V sliders allow adjusting these lengths.
- Quantile Ellipsoid: the ellipsoid inscribed within the previous n-rectangle; the Outer Orbital option allows using the ellipsoid circumscribed around the n-rectangle.

In terms of usage, the Gaussian ellipsoid is easy to handle (a single Confidence slider), but is only appropriate for point clouds with a Gaussian tendency (normal distribution in each dimension).
The "box" is a robust variant of the previous ellipsoid. But it tends to become very strict (many hidden points) for high quantile values. The quantile ellipsoid is a much softer variant of the box; it only hides points distant in several simultaneous directions. But its circumscribed variant (outer orbital) is very restrictive (even more than the box). You have to lower the thresholds to avoid hiding the entire cloud.

From a masking proportion perspective, we have roughly the following values:
- In an n-dimensional space, the 90th quantile n-rectangle in all dimensions encompasses approximately (0.9^n) % of the population (72% in 3D, 65% in 4D, 60% in 5D..., 35% in 10D).
- The ellipsoid inscribed in the previous n-rectangle only covers a rapidly decreasing fraction: in 3D, the ellipsoid orbital occupies 52% of the rectangle, in 4D 30%... and in 10D barely 0.25%!
- The ellipsoid circumscribed around the n-rectangle reverses the trend. In 3D, the orbital occupies almost 2.7 times more space, in 4D 1.2 times more, in 5D 8.3 times more... and it grows faster and faster!

From most tolerant to most strict: inscribed ellipsoid < n-rectangle < circumscribed ellipsoid.

3.2.3.2. Comparison between Orbital and Filtering

What is the difference between the orbital and filtering based on absolute coordinates?
Both objects serve to filter the visualization, but each has its nuances.
{$Number of filtered dimensions$}
Filtering based on Absolute Coords only applies to (at most) three defined filtering axes. The orbital, on the other hand, applies to all possible dimensions simultaneously.
{$Boolean condition$}
Coordinate filtering can use the AND or OR condition, whereas the orbital only uses an AND principle by construction.
{$Definition of filtering thresholds$}
Coordinate filtering requires formally stating the coordinate threshold value that will be checked on the filtering axes. The orbital works instead by a quantile logic (proportion of points having a coordinate less than or equal to the threshold) or probabilistically (Gaussian

confidence interval based on Mahalanobis distance).

3.2.4. Decision Boundary

This option displays the decision boundary of binary logistic regression or the predictive hyperplane of linear regression.
- Boundary Alpha: adjusts the transparency of the hyperplane / decision boundary
- Resolution: adjusts the mesh fineness of the hyperplane / decision boundary

Note: the decision boundary is only displayed if the target variable is binary; the prediction hyperplane is only displayed if the number of predictors (explanatory variables) is greater than or equal to 2. The degenerate case of the line (simple regression) is ignored.

3.2.5. Hy-Perf Cube

Allows quickly changing the axes displayed on the scene to analyze Hy-Performance datasets.
Cube Preset: these presets allow quickly mapping certain spatial axes to transformed dataset variables (budget, uplift, ROI, etc.).

For further customization, you need to manually map the axes in Dataset / Axes Mapping.

3.2.6. Valence

The Valence logic (connectivity links) is based on a proximity analysis in space.

Point Classification
The system groups all displayed points into three main families to filter link types:
- Reference: points from presets (Zodiac, Malgovert, DISC, MBTI, etc.).
- Data: points imported via CSV or added manually.
- Factor: points calculated by Factorial Analysis (Individuals or Modalities).

Overall logic: Data points are the mobile and analyzed elements, while Reference points, Factor points, and the Polytope serve as fixed landmarks (anchors). We thus look at how user data orients itself relative to the application's internal structure.

Valence Info
If Valence Info is checked, additional information is added to the side panel when a point is clicked: you can view the various neighbors of a point as well as their respective distances, for different projection hypotheses (ND, 3D, with or without normalization).
The Use Inverse Dist checkbox allows displaying, not Euclidean distances, but weightings by the inverse of distances. Example: with 3 neighbors, the value displayed for a given neighbor is the inverse of the distance divided by the sum (over the 3 neighbors) of the inverse distances.

Neighborhood Rank
The K-Neighbors parameter defines the number of ranks (closest distances) to display per point; it is not the exact number of links displayed (see "Ties" below).

3.2.6.1. Pipeline Stage

The concept of neighborhood depends on the projection stage. As a reminder, the projection includes all these stages:
- ND-4D: ND-standardization, ND-normalization (optional), rotation, shear, projection
- 4D-3D: rotation, shear, projection, 3D-Normalization (optional), scaling (optional)
- 3D-2D/1D: squashing

A neighborhood study can be performed in the initial ND scene before spatial projection. This can include or exclude a normalization step. Similarly, the neighborhood can be determined after projection, in the resulting 3D scene visible on screen. This scene can also include normalization.

In summary, neighborhoods can be considered:
- ND: valence is calculated on initial data, ND space without normalization
- ND Stand: valence is calculated on centered-reduced ("standardized") ND data, before projection
- ND Norm: valence is calculated on normalized ND data, before projection
- 3D: valence is calculated on projected 3D data, before 3D normalization, before scaling
- 3D Scaled: valence is calculated on projected data, after possible 3D normalizations and data scaling

3.2.6.2. Valence Directions

A connection is established between anchor points and neighborhood points. For each anchor point, we look for neighborhood points. The closest neighbors to the anchor point are called "rank 1". The next neighbors are rank 2, etc.

For each point P, all other points are sorted by Euclidean distance. Points at the same distance (with a 10^-5 tolerance) are grouped into a single rank. With K = 2, the two closest distance ranks are kept - this can generate more than 2 lines if several points share the same distance.

There are two different algorithms depending on the nature of the connection:
{$Mutual Links (Internal to a group): anchors and neighbors belong to the same point group$}
Examples: Reference->Reference, Data->Data, Polytope->Polytope
- Calculation: for two points I and J, we calculate J's position in I's neighbor ranking (rank_{IJ}) and conversely (rank_{JI}).
- Condition: the link is drawn if the average rank (rank_{IJ} + rank_{JI}) / 2 is below the K

Neighbors parameter.
- Logic: allows showing only strong and reciprocal relations within the same population.
{$Unidirectional Links: anchors and neighbors are different groups$}
Examples for: Data->Reference, Data->Factor, Zodiac->Polytope.
- Calculation: We take each "Anchor" point (e.g., a Data point) and look for its nearest neighbors in the "Neighbor" group (e.g., Polytope vertices).
- Condition: The link is drawn if the neighbor's rank is below K Neighbors.
- Logic: This is an attraction logic. We see which reference points or solid facets the data "tends" towards.

In the Valence Directions panel, you can add all desired links between anchor groups and neighborhood groups.

Handling Ties
When a rank contains more than one neighbor (a tie), the corresponding lines can be drawn dashed (click Ties Dash) instead of solid. This visually signals ambiguous neighborhoods. A line is dashed if either of its ends has a tie at its respective rank level.
- Tolerance: If two points are at an almost identical distance (10^-5), they receive the same rank.
- Visual: If a rank contains several points (tied), the lines are displayed dashed (if the Dashed Ties option is active in the Style panel).

Color Code and Distances
- Dimensions: Distances are calculated on the final coordinates (x, y, z) after all transformations (ND-4D-3D). This means zoom and 4D rotations directly impact valence links.
- Colors: each proximity level (neighbor rank) can be identified by a customizable color

3.3. Data Style

Location: Workspace tab, Data Style panel.

This section allows customizing the sizes and colors of labels, points, and valence or orbital lines. It also allows changing the number of decimal places for metrics displayed in the interface.

3.3.1. Data Selection

In this section, you can select the data whose points and/or labels you wish to display. You will find the usual families:
- Reference data (Zodiac, MBTI, DISC, etc.)
- Imported data (called custom), synthetic or not
- Factorial data of individuals (rows)
- Factorial data of attributes (features, columns)

You can also select projected data sources (Imported Source) from:
- Imported data (CSV)
- Data projected by personality forms
- Data calculated by the Hy-Perf module

3.3.2. Valence Colors

For valence, each line connects two points, A and B, and its color translates the neighborhood rank. Point A might rank B as its closest neighbor (Rank 1), while B might rank A differently (e.g., Rank 2). To reflect this bidirectional relationship, the displayed color is determined by the integer part of the average of both ranks: Displayed Rank = $\lfloor (\text{rankA}\rightarrow\text{B} + \text{rankB}\rightarrow\text{A}) / 2 \rfloor$
This guarantees that the color reflects mutual agreement: two points considering each other close get a strong color (Rank 1), while asymmetrical relationships are naturally degraded.

3.3.3. Orbital Style

You can constrain the absolute and relative sizes of orbitals for comfort. However, manipulating these sizes is purely aesthetic and has no mathematical effect. If you want to hide the orbital, simply set its alpha (transparency) to zero.

3.3.4. Layouts

You can choose between the following layouts:
- Default Layout: standard display
- Malgovert Layout: alternative display used by the author in the context of the novel Malgovert. The X, Y, and Z coordinates are renamed K, L, and S. They stand for the Kappa (Cocytus), Lambda (Lethe), and Styx (Sigma) serums in the fictional universe of Malgovert. The colors of the corresponding axes are modified to reflect the serum colors in the novel: blue for Cocytus (K), silver for Lethe (L), and green for Styx (S).
- MBTI, DISC, Social Style, etc.: alternative display that includes labels specific to behavioral frameworks used in business.

Additionally, you have a Number Decimals setting for displaying the decimals of numbers in the various tables and panels.
Finally, there is a Label Chars setting to define the number of characters displayed for the new axis names in the different layouts.
For instance, in the "Social Style" layout, the X and Z axes become "Assertiveness" and "Responsiveness". If you set the slider to 2, the labels will become "As" and "Re".

4. Modeling

4.1. Advanced Stats

This section complements the Stats tab. Here, analyses are performed on several pairs of columns in parallel. The tab is subdivided into three parts: correlations (bivariate), profiling (over- or under-representation of certain attributes within clusters), and association rules.

4.1.1. Thresholds for Statistical Tests

It is possible to define the desired statistical tests as well as the threshold values to colorize the metrics. These thresholds apply to correlation and profiling charts.
- Num Testing: desired statistical test for numerical correlations (Pearson, Spearman, Kendall)
- Cat Testing: desired statistical test(s) for categorical dependencies (chi2 only, or a mix of chi2 and Fisher's exact test for 2x2 contingency tables)
- Mixed Testing: statistical test used for analyzing central tendency differences between two groups (ANOVA F which generalizes Student's T-test, and Kruskal-Wallis which generalizes the Mann-Whitney U test)
- Alpha: threshold value below which p-values become green in the tables.
- Num Effect Threshold: threshold value above which numerical effect sizes (in absolute value) become green; applies to R2, Spearman's rho squared, and Kendall's tau
- Cat Effect Threshold: threshold value above which categorical effect sizes (in absolute value) become green; applies to Cramer's V, Cohen's h, and Wilcoxon's r
- OR Effect Threshold: threshold value above which ratio and log-ratio type metrics (in absolute value) become green; applies to (log-)odds ratios, (log-)lifts, and Weights of Evidence
- Mix Effect Threshold: threshold value above which categorical effect sizes (in absolute value) become green; applies to Cohen's d, Hedges' g, Glass's Delta, Cliff's delta, Eta squared (H test) and Omega squared (ANOVA)

4.1.2. Correlations

4.1.2.1. Settings

Data to analyze (optional and cumulative):
- Target: target variable used to determine the most discriminant variables
- Imported Data: data imported in CSV
- Factor Data: data from a factorial projection
- Cluster Data: s-cluster label from a clustering
- Score Data: scores calculated by a Sup ML estimator

4.1.2.2. Associated Statistics

Correlation metrics table with the target
Bivariate dependency statistical tests (Pearson and equivalents, chi-square, ANOVA, etc.) with p-value and associated effect sizes (R2, omega squared, Cramer's V, etc.)

Correlation metrics table between explanatory variables
Same as above, but between explanatory variables

Effect size matrices (intra-variable correlation)
Correlation (effect size) matrices between variables.

Effect size bars vs Target
Intensity of correlation between explanatory variables and the target variable. Reflects discriminant power.

Scene correlation matrices
Correlation matrices between spatial data or data projected by a factorial method.

For these correlation matrices, several metrics are offered:
- Dot Product (normalized dot product): sum of the component-wise products, divided by the product of the norms; $dot(u, v) = (u \cdot v) / (||u|| \times ||v||)$ (cosine similarity); range: [-1, 1]. 1 = aligned, 0 = orthogonal, -1 = opposed
- Pearson: Pearson linear correlation coefficient, from -1 to 1; the coefficient is close to zero in case of non-correlation; $r(u, v) = \Sigma(u_i - \bar{u})(v_i - \bar{v}) / (\sigma_u \times \sigma_v)$ where $\bar{u}$, $\bar{v}$ are the means and σ are the standard deviations of the coordinates' components.
- Similarity (similarity index): ratio of distances between a pair and the maximum distance of the dataset; $sim(u, v) = 1 - d(u, v) / maxDist$ where d is the Euclidean distance and maxDist is the largest distance across all active points; equals 0 when two points are superimposed, equals 1 when the two points are the furthest among all pairs

Note: diagonal self-comparisons of the matrix are left empty.

4.1.3. Profiling

This section is an in-depth description of the characteristic values found in groups of interest (segment, target, cluster, score quantile), with or without treatment.

4.1.3.1. Settings

Data to analyze (optional and cumulative):
- Imported Data: data imported in CSV
- Factor Data: data from a factorial projection
- Cluster Data: s-cluster label from a clustering
- Score Data: scores calculated by a Sup ML estimator

Specific profiling settings are:
- Group Variable: variable used to define the groups to profile; corresponds to one of the variables of interest (segment, target, cluster, score quantile)

- Use Complement: to calculate ratio denominators, either the total population is used, or the complementary population to the studied group is used (= the "rest" of the population = other groups). If this option is checked, the complement is used; helps highlight metrics for small groups.
- Moment Metric: highlighted variable for displaying numerical profiling metrics
- Lift Metric: highlighted variable for displaying categorical profiling metrics

4.1.3.2. Associated Statistics

Profiling Table
Five tabs centralize all metrics:
- Profile: general summary; combines part of the lift and moment views
- Moment: focus on numerical descriptive variables
- Lift: focus on categorical descriptive variables
- Lift Net: focus on categorical descriptive variables (1 row per category) with the presence of a treatment
- Info Value: focus on categorical descriptive variables (1 row per variable) with or without the presence of treatment; it completes the Lift / Lift Net view with aggregated indicators

Profiling Bars
A graphical version of the previous results.
Charts are grouped by profiling metric. One row per variable with the ability to sort by group.
- Positive bars (one color per group) translate into characteristic high categories or values (typical, overrepresented) of the concerned groups.
- Negative bars (one color per group) translate into characteristic low categories or values (underrepresented) of the concerned groups.

4.1.4. Association

Association rules determine sets of "antecedent" categories (qualitative variables) that appear simultaneously with a "consequent" category of interest. Rules only work between categorical variables.

4.1.4.1. Settings

Specific settings for association rules are:
- Group Variable: optional variable used to partition rules: either the rules are global, or they are subdivided according to all modalities of the group variable
- Group as Consequent: rather than partitioning the rules, the group variable is this time used to directly define the Consequent values.
- Min Support: minimum support to retain a rule; it is the proportion of rows containing the set of antecedent categories; recommended 1-5%
- Min Confidence: minimum confidence to retain a rule; it is the proportion of rows (among rows having all antecedents) for which the consequent appears; allows filtering out

unreliable rules, even if it is not sufficient (the lift below is necessary to discard reliable but uninteresting rules); recommended 60-80%
- Min Lift: minimum overrepresentation to retain a rule; a rule having a lift of 1 is a random selection; a rule with a high lift indicates that the consequent indeed has a higher chance of appearing when the antecedents appear (useful rule); a negative lift allows finding the antecedents that prevent the consequent from appearing (opposite objective); recommended 1.2
- Max Antecedent Size: number of antecedents used to define a rule; the larger this number, the more the combinatorics explode
- Max Consequent Size: number of consequents used to define a rule; the larger this number, the more the combinatorics explode
- Top N: visual filtering to keep only the top N rows (but rows can be sorted or filtered independently of this value)

4.1.4.2. Associated Statistics

The result table lists the rules as well as the associated metrics (support, confidence, lift, etc.).
The symbol ∧ means intersection (logical AND).

4.2. Hy-Performance

Feature available soon.
Goal: study the effectiveness of a treatment (clinical trial or marketing action) on a population, via an A/B test integrating statistical evaluations.

4.2.1. Data Sources

Granular or aggregated data

- Keep All Factors: keep all
- Nb Factors: allows
- Inference: uses, only

4.2.2. Settings

Analysis type: absolute or incremental
- With Control Groups: keep all
- Alpha: significance level (type I error) to reject the null hypothesis
- Alternative: one-sided (greater = right; less = left) or two-sided

Analyzability thresholds (measurement reliability and volume-related instability)
- Analyz Min Conv: minimum threshold to consider an action analyzable
- Max Relative Prec (%): maximum tolerated value of the coefficient of variation, to consider

an action analyzable
- Vol Increase Method: method used to define the stability of statistical tests when volume increases: you can choose to increase the size of controls (Controls Only) or both controls and treated (All Groups)
- Vol Increase Factor: group enlargement factor to test volume stability
- Amount Formula Switch: converted value (in each group) that allows choosing the type of formulas used for uplift calculations; when a group has fewer converted entities than the number indicated here, volume-based formulas are used, otherwise amount-based formulas are used

Incremental amounts calculations
- Inlier Salvage: recovery/preservation method for low-volume actions whose means precision is sufficient
- Inlier Min Conv: minimum converted value for an action to be "recoverable"
- Inlier Max CV: maximum tolerated value of the coefficient of variation, to consider a low-volume action "recoverable"

ROI Management
- Breakeven KO -> Zero: actions that do not reach break-even (amounts from treated converted do not cover the budget) are replaced by zero (rather than negative values)
- Isolate Non-Zero IA: if the option is checked, then significant actions are isolated and are not temporally grouped; otherwise, only ROI+ actions (positive ROI) are isolated, but other actions are indeed grouped over time.

4.2.3. Associated Statistics

PERIOD
Period of aggregation for the groups of actions and the financial KPIs: Q1, Q2, Q3, Q4, H1, H2, FY

AGG_TYPE
Is the aggregate evaluated on actions (before grouping, coded "BG") or on groups of actions (after grouping, coded "AG")

YEAR
Year of the action

HALF_YEAR
Semester of the action

QUARTER
Quarter of the action

MONTH

Month of the action

WEEK
Week of the action

DAY
Day of the action

STRATEGY
Strategy followed by the action

FLAG_OUT
Remove from analysis

LEVEL1
Typology of the action (1st level category)

LEVEL2
Typology of the action (2nd level category)

LEVEL3
Typology of the action (3rd level category)

LEVEL4
Typology of the action (4th level category)

SEGMENT
LEVEL1 - LEVEL2 - LEVEL3 - LEVEL4

ACTION_GROUP
(Month) - Strategy - Segment - ROI Flag - (ROI Trend)

Month and Trend are only appearing for certain ROI flags and to_group flag

MARGIN
Margin Profit (in % of outstanding)

AMOUNT_FACTOR
Factor used to assess the profit from margin and amount: PROFIT = AMOUNT * AMOUNT_FACTOR * MARGIN
Useful when margins depend on two multiplied variables (e.g. monthly subscription price and number of months subscribed)

UNIT_COST

Unit communication cost

BUDGET
Budget of the action (unit cost * nb targets)

ANALYZABLE_BUDGET
Part of the budget (in €) that is analyzable (ROI_FLAG is not "NA")

ANALYZABLE_BUDGET_RATIO
Part of the budget (in %) that is analyzable (ROI_FLAG is not "NA")

POSITIVE_ROI_BUDGET
Part of the budget (in €) that is profitable (ROI_FLAG is "ROI+")

POSITIVE_ROI_BUDGET_RATIO
Part of the budget (in %) that is profitable (ROI_FLAG is "ROI+")

POS_ROI_ANA_BUDGET_RATIO
Part of the analyzable budget (in %) that is profitable (ROI_FLAG is "ROI+")

NB_T
Nb of Targets

NB_C
Nb of Controls

CG_SIZE
Control Group Size (%)

NB_CONV_T
Nb of Converted Targets

NB_CONV_C
Nb of Converted Controls

CONV_RATE_T
Conversion Rate Target = Conv Rate Target

CONV_RATE_C
Conversion Rate Control = Conv Rate Control

UPLIFT_NB
UPLIFT in volume (delta rates in percentage points)

UPLIFT_NB_RATIO
Uplift in Volume Ratio (%)

AMOUNT_T
Amount (Total Converted Amount) Target

AMOUNT_C
Amount (Total Converted Amount) Control

SQUARED_AMOUNT_T
Squared Amount (Total Squared Converted Amount) Target

SQUARED_AMOUNT_C
Squared Amount (Total Squared Converted Amount) Control

WEIGHTED_SQUARED_AMOUNT_C
Scaled Squared Amount Control

PROFIT_A
Profit All Customers (Targets + Controls)

PROFIT_T
Profit Target

PROFIT_C
Profit Control

AVG_AMOUNT_A
Average Amount per Customer

AVG_AMOUNT_T
Average Amount per Target

AVG_AMOUNT_C
Average Amount per Control

AVG_AMOUNT_CONV_A
Average Amount per Converted Customer

AVG_AMOUNT_CONV_T
Average Amount per Converted Target

AVG_AMOUNT_CONV_C
Average Amount per Converted Control

VAR_AMOUNT_T
Variance of the Amount Target

VAR_AMOUNT_C
Variance of the Amount Control

VAR_AMOUNT_CONV_T
Variance of the Converted Amount Target

VAR_AMOUNT_CONV_C
Variance of the Converted Amount Control

UPLIFT_AMOUNT_RATIO
Uplift in Amount Ratio (%)

UPLIFT_TOTAL_RATIO
Total Uplift Ratio (%)

CI_VALID_T
Is the Confidence Interval of RR valid?

CI_VALID_C
Is the Confidence Interval of RR valid?

CI_SE_NB_T
Standard Error of the Resp Rate Confidence Interval

CI_SE_NB_C
Standard Error of the Resp Rate Confidence Interval

CI_LOWER_NB_T
Lower bound of the Resp Rate Confidence Interval

CI_LOWER_NB_C
Lower bound of the Resp Rate Confidence Interval

CI_UPPER_NB_T
Upper bound of the Resp Rate Confidence Interval

CI_UPPER_NB_C
Upper bound of the Resp Rate Confidence Interval

CI_SE_AMOUNT_(CONV)_T

Standard Error of the Average Amount Confidence Interval *[1]

CI_SE_AMOUNT_(CONV)_C
Standard Error of the Average Amount Confidence Interval *[1]

CI_LOWER_AMOUNT_(CONV)_T
Lower bound of the Average Amount Confidence Interval *[1]

CI_LOWER_AMOUNT_(CONV)_C
Lower bound of the Average Amount Confidence Interval *[1]

CI_UPPER_AMOUNT_(CONV)_T
Upper bound of the Average Amount Confidence Interval *[1]

CI_UPPER_AMOUNT_(CONV)_C
Upper bound of the Average Amount Confidence Interval *[1]

VOLUME_INCREASE_FACTOR
("VOL_INC", see *[2])
Strategy to increase volumes: it impacts NB, NB_CONV, CONV_RATE.

Note: this metric is not visible in the output. It's only an intermediate within the code.

Z_STAT
Z Statistic

Z_NORM_PROBA
Cumulative probability of the Z-Stat regarding the Normal distrib

Z_NORM_P_VALUE
P-Value of the Z-Stat regarding the Normal distrib

ODDS_RATIO
Odds Ratio of the CONVancing contingency table

Z_HYPERGEOM_PROBA
Cumulative probability of the Z-Stat regarding the Hypergeometric distrib

Z_HYPERGEOM_P_VALUE
P-Value of the Z-Stat regarding the Hypergeometric distrib

P_VALUE_NB
Retained P-value for volume significance

SS_NB
Is the uplift in volume significant?

T_STAT_(CONV)
T Statistic *[1]

T_NORM_PROBA_(CONV)
Cumulative probability of the T-Stat regarding the Normal distrib *[1]

T_NORM_P_VALUE_(CONV)
P-Value of the T-Stat regarding the Normal distrib *[1]

T_STUD_PROBA_(CONV)
Cumulative probability of the T-Stat regarding the Student distrib *[1]

T_STUD_P_VALUE_(CONV)
P-Value of the T-Stat regarding the Student distrib *[1]

P_VALUE_AMOUNT_(CONV)
Retained P-value for amount significance *[1]

SS_AMOUNT_(CONV)
Is the uplift in amount significant? *[1]

SS_DOUBLE_(CONV)
Means "Significant in both amount and nb" *[1]

SS_ZERO_(CONV)
Means "Not Significant in neither amount nor nb" *[1]

SS_DIV_(CONV)
Means "Significant in amount (or nb) but not in nb (or amount)" *[1]

RAW_ANALYZABILITY
Is the action / action-group roughly analyzable regarding the CG existence and the Nb of Conversions?

CI_LOWER_NB_C_WITH_CONV_RATE_T
Theoretical value of the Control's lower Conf Interval bound if we assume that CONV_RATE_C = CONV_RATE_T

TRUE_ZERO_CONTROL
Is the zero number of Converted Control a reality or an issue due to low volume?

NET_ANALYZABILITY
Is the action / action-group definitely analyzable regarding the Conversions and the statistical significance stability?

NOTE FOR MULTI-EFFECT
- If one effect is analyzable then the global effect is analyzable

IA_PC_AMOUNT
Credit: additional Amount per customer based on amounts
Insurance: additional probability per customer based on amounts

IA_PC_NB
Credit: additional Amount per customer based on volumes
Insurance: additional probability per customer based on volumes

IA_COHERENCE_FLAG
Credit: retained formula for additional Amount
Insurance: retained formula for additional probability

"2" means amount-based formula
"1" means volume-based formula
"0" means unsignificant (IA = 0)

The "New" method is also named "coherence rule": an action is coherent if, when increasing UPLIFT_NB, we do not decrease AVG_AMOUNT_CONV_T.

The "Old" method is also named "retake rule": if we are not significant in amount, then we might "retake" the IA from the significance in nb.

IA_PC
Credit: retained additional amount per customer
Insurance: retained additional probability per customer

IA
Credit: retained additional Amount
Insurance: retained additional probability

ROI_PC
ROI per customer

Note for multi-effect
> The global-effect-ROI-PC is calculated by computing the underlying effect-IA-PC (multiplied by their respective margin/lifetime) and by subtracting the unit cost once! → it's not the simple sum of the underlying effect-ROI-PC!

- If the global effect is NOT analyzable, then the ROI is cut to zero

ADD_PROFIT_PC
Additional Profit per customer

ROI
Total ROI

ADD_PROFIT
Total additional Profit

ROI_PU
ROI per unit invested

ROI_PU_ANA
ROI per unit invested and analyzable

ROI_MAX_PC
Theoretical maximum ROI per customer (if control never gets Converted)

ROI_MAX
Theoretical maximum ROI (if control never gets Converted)

ROI_FLAG
Typology of ROI/IA (situations of profitability)

TO_GROUP
Is the action groupable?
We want to group the actions that are not analyzable or not significant

ROI_AMOUNT_WITHOUT_SIGNIF
Theoretical ROI (based on amounts) without significance

ROI_NB_WITHOUT_SIGNIF
Theoretical ROI (based on volumes) without significance

ROI_TREND
Has the action (groupable) a positive or negative trend of ROI?

4.3. Hy-Personality

Here is the "rigorous" method to determine which Zodiac signs look most like you!

A personality questionnaire integrated into the application can be filled online. It combines themes present in several business frameworks (Social Style, DISC, MBTI) as well as lexical descriptions of the astrological zodiac.

Filling out the questionnaire and projecting the results allows you to visualize similar concepts or attributes between your answers and the profiles present in the different frameworks. This approach relies on a PCA or MCA of the answers of the different profiles to the 18 questions in the form. This is a dataset provided by the application. It is free to use but cannot be modified in this interface. However, the answers you add are customizable.

Procedure: answer the 18 questions by moving the slider to the left or right (5-degree scale) according to the answer that corresponds to you the most. If you hesitate, you can leave the slider in the center (balanced or neutral answer).
Once finished, enter a label and click Submit to project the results into the factorial hyperspace!

4.3.1. Data Sources

The questionnaire can come from different sources:
- Reference Points: data already integrated into the application (behavioral frameworks and zodiac)
- User Personality Form: online questionnaire displayed on screen
- CSV: questionnaire imported in CSV format (via the Form Import section of the current tab, or the import section of the Data tab)

4.3.2. Projection Techniques

The answers to the questionnaire (from all sources) are projected using a PCA or MCA already trained with the application's reference data.
- Method: PCA or MCA
- Keep All Factors: keeps all factors (this also includes statistical noise in variance explanation)
- Nb Factors: if the previous option is unchecked, you can define here the number of factors to keep; recommended between 3 and 8

4.3.3. Reference Points Filtering

The Reference Filtering section allows you to choose which points to display: MBTI, DISC, Social Style (SS), Zodiac.
Each category is subdivided between {$Features$} which correspond to description criteria and {$Profiles$} which are the resulting combinations of several criteria.
Example: the {$Extraversion$} and {$Introversion$} points that define energy orientation are {$MBTI Features$}, while {$INTP$} is an {$MBTI Profile$} defined by the presence of the "I" (for Introvert) modality.
Likewise, {$Fire$} is a {$Zodiac Feature$}, while {$Aries$} is a {$Zodiac Profile$}.

4.3.4. Associated Statistics

Radar of answers by profiles
Each vertex of the radar represents a question (integer values from -2 to 2). Two profiles are overlaid in this radar to visualize the overall resemblance of their answers to the questionnaire.

Answers of all respondents by question
The bar chart gives the answer of all respondents to the selected question.

Scene correlation matrices
Correlation matrices between raw answers to the personality questionnaire (integer values from -2 to 2), and also between projected answers (factorial coordinates).

4.3.5. Example of use: inverted zodiac

Go to *Modeling / Personality*
In the *Personality Map* panel, leave the default settings:
- *Method*: PCA
- *Reference Points*: checked
- *User Personality Form*: checked
- *CSV*: unchecked

Fill out the questionnaire, give a name and click *Submit*. You then switch to Workspace.

Go to the *Dataset* tab of the Workspace panel, in the *Data Selection* section, check only:
- *Show Reference Pts*: these are the points already in the application
- *Show Fact Indiv Pts*: these are the answers to the online form

Go to the *Data Exp* tab of the Workspace panel.
- In *Data Scaling*, check *Normalize* and put the *Data Scale* around 2 or 3 according to your feeling.
- In *Valence*, check *Show Lines*, then check *Valence Info*, check *Use Inverse Distance*, then set the *Pipeline Stage* to *Start*.
- Still in Valence, under the K Neighbors slider, check only *Factor-Reference*

Just below, in the *Filtering* section, check Filter then in the *Personality* list, choose *Zodiac Profiles*.

Now, by clicking on the point corresponding to your answer, a side panel appears with statistics. You can view the *ND Valence* and *ND-Norm Valence* charts which indicate your "proportion" of each Zodiac sign (seen as your nearest neighbors in terms of answers).

To better realize the distances, at the very bottom of the panel, after the valences, click on the *Best Cos2 Plan* button. This will display the space in which your answer is spatially best represented. The nearest neighbors then tend to visually move closer, which is more coherent.

4.4. Factor

This tab allows applying dimensional reduction methods to points imported in CSV, or to the already integrated dataset (synthetic data, answers to the personality questionnaire).
Simply choose the number of factorial axes retained and the technique used.

4.4.1. Factorial Axes

- Keep All Factors: keeps all possible factorial axes; depending on the method, this corresponds either to the number of numerical variables (PCA), the number of categorical variable modalities (MCA), the sum of the two previous (FAMD), or the number of classes - 1 (LDA).
- Nb Factors: allows choosing the number of axes to retain
- Inference: uses, on new data, the already calibrated projection matrices from a previous factorial analysis

4.4.2. Factorial Techniques

- PCA: Principal Component Analysis; adapted to continuous numerical data
- MCA: Multiple Correspondence Analysis; adapted to categorical data
- FAMD: Factor Analysis of Mixed Data; adapted to data mixing numerical and categorical variables
- LDA: Fisher's Linear Discriminant Analysis; adapted to numerical data including a target variable to separate groups
- MCA+LDA: Discriminant analysis on qualitative data (Saporta's method); allows performing an LDA on categorical data
- FAMD+LDA: Discriminant analysis on mixed data (variation of Saporta's method); allows performing an LDA on mixed data

Once projected, the data is visualizable in the Workspace tab, and appears overlaid on the polytope acting as a visual anchor. It is recommended to use a simple cube (or even a tesseract for 4D analysis) or a sphere to facilitate exploration.

4.4.3. Associated Statistics

In the Factor tab, you will find several statistics relating to the factorial method:
- Contribution of individuals (rows) to axes: a distant point on an axis will strongly "contribute" to this axis
- Contribution of features (columns) to axes: same, but from the variables' perspective

- Squared cosine of individuals: a point with a high squared cosine is well represented on this axis; its projected coordinate remains faithful
- Squared cosine of features: same, but from the variables' perspective

Note that factorial axes have a correspondence on the spatial axes:
- 1st factorial axis (PCA_1, MCA_1) corresponds to the X axis
- 2nd factorial axis (PCA_2, MCA_2) corresponds to the Y axis
And so on with the Z, W, V5, V6 axes, etc.
Thus, when analyzing contributions or squared cosines, the value of the dropdown list corresponds to the factorial axis. That is why they are displayed as "X (PCA_1), Y (PCA_2)" or "X (MCA_1), Y (MCA_2)".
But if the dataset contains spatial coordinates (also named X, Y, Z), then we find situations where axes X, Y, Z (from the input dataset before projection) contribute to axis X (which is actually the first factorial axis, after projection). It is therefore necessary to keep in mind what pertains to the input set and the result of the factorial projection in the output space (and which can share the same letters).

4.5. Clustering

Clustering allows grouping points by spatial proximity.

4.5.1. Data Sources

The data source to cluster has nuances that enable or disable certain options.
Numerical data:
- Imported Data: raw imported data (CSV); only numerical variables from the CSV are retained
- Spatial Data: data visually projected in the scene: this includes the previous numerical data as well as reference points and manually added points (Workspace / Dataset tab); it is therefore an enriched version of the previous one, but does not include the possibility of {$Inference$} mode (see below)
Factorial data:
- Factor Individuals: projected data of individuals (rows)
- Factor Features: projected data of features (columns)
Personality data (Hy-Personality):
- Personality Form: projected data of respondents to the personality questionnaire

4.5.2. Methods

Clustering methods
- K-Means: the simple method to go fast
- Hybrid: the complete method that combines K-Means and HCA (Hierarchical Clustering Analysis) and potentially a second K-Means for "peripheral cleaning"
- Partition: a special case consisting of assigning, to each spatial point, the nearest

reference point; this therefore requires displaying reference points (either the points defined in {Dataset}, or the projected points of the {Hy-Personality} questionnaire)

Specific options for Spatial Data and Personality Form sources
- Add Reference Points: if the source is {$Spatial Data$}, then this option allows adding manually added reference points (Workspace / Dataset tab); if the source is {$Personality Form$}, then this option allows adding projected points of type MBTI, DISC, SS, Zodiac.
- Reference as Center: uses reference points as seeds (initial points) for the first K-Means; it necessarily disables the choice of the number of {$Centers$}

Specific options for Imported Data and Factor Data (Individuals or Features) sources
- Inference: application of a clustering model (already trained) on new data

Common options
- Standardize: unit scaling of numerical data

4.5.3. Pre-KMeans

The first part consists of applying a K-Means method on the initial set.
- N Init: number of different attempts to determine the best partitioning
- Use Sqrt(N) Centers: uses the square root of the number of points as the number of centroids to perform the partitioning
- N Centers: if the square root rule is unchecked, then this slider sets the desired number of centroids. Generally, a number between the cube root and the square root of the initial size (number of rows) is a good starting point
- Filter Outliers: allows assigning isolated points to a separate class; based on the Mahalanobis distance
- Outlier Confidence %: confidence level used to define outliers

4.5.4. HCA (Hierarchical Cluster Analysis)

The second step is a hierarchical clustering (Ward's method) applied to the centroids of the Pre-KMeans.
- HCA Auto-Cutoff: automatically determines the number of final clusters
- Cutoff Metric: metric used to automatically determine the number of clusters; can be based on the silhouette coefficient or the acceleration curve of the dendrogram (the elbow on the inter-class inertia curve)
- Min Nb Clusters: minimum number of desired clusters (allows seeking other optima in the previous method)
- Max Nb Clusters: maximum number of desired clusters (allows restricting the search for optima in the previous method)
- Nb Clusters: if the Auto-Cutoff mode is unchecked, then this slider determines the number of final clusters

4.5.5. Post-KMeans

Depending on the spatial distribution of the data, HCA methods can suffer from questionable allocation for points located on the periphery of clusters. A K-Means performed on the barycenters of the HCA clusters allows reassigning these peripheral points to the most geometrically intuitive clusters.
- Merging Threshold %: at the very end, this percentage allows merging clusters that do not have enough individuals (proportion set by this slider)

The Clustering button starts the algorithm; the Clear button resets the results.

4.5.6. Associated Statistics

Quality metrics
The displayed quality metrics are: silhouette coefficient, Davies-Bouldin Index (DBI), Calinski-Harabasz Index, Dunn Index

Charts
The following section displays the HCA dendrogram, and the inter-class inertia acceleration curves. Then comes the population distribution by cluster, box plots by spatial variable, and contingency matrices with variables of interest (Segment, Target, Treatment).

4.6. Sup ML (Supervised Machine Learning)

Sup ML allows training regression (linear) and classification models.

4.6.1. Settings

- Source: selection of data to model: either raw imported data, or data projected by factorial method (on individuals only)
- Target: target variable to explain
- Estimator: choice of model: linear regression or logistic regression
- Score Bins: number of quantiles used to discretize the predicted score; useful for performing profiling analyses on scores (Stats tab, Profiling section)
- Inference: application of a supervised model (already trained) on new data
- Cluster Predictor: uses clustering classes (if they exist) as additional explanatory variables

The Run Estimator button starts the algorithm; the Clear button resets the results.

4.6.2. Associated Statistics

The results are available in two tabs, one per sample (Train, Test).

Linear Regression

The displayed quality metrics are: R2, MAE, RMSE.
A panel displays the regression coefficients.

Logistic Regression
The displayed quality metrics are: Balanced Accuracy, Macro Precision, Macro F1, ROC AUC, Gini.
A panel displays the confusion matrix, then the regression coefficients (and associated odds ratios).
Finally, two curves are offered: ROC and lift (cumulative gains).

Scores and quantile scores
The model scores are discretized into quantile scores. These quantiles can be used as a group variable in profiling.

4.7. Forecast

The time series forecasting model relies on Holt-Winters exponential smoothing.

4.7.1. Settings

- Time Column: time variable (can be a simple index)
- Target: target variable to predict over time
- Seasonal Period: number of periods defining a season
- Seasonality: type of seasonality: additive or multiplicative
- Auto-fit Alpha/Beta/Gamma: automatically determines the exponential smoothing coefficients
- Back-Test Holdout: number of points removed from the training set to test the model
- Inference Horizon: number of additional points to predict (data after the imported periods)

4.7.2. Associated Statistics

Metrics
The results consist of visualization curves as well as quality metrics (in the training and back-test samples): MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error).

Charts
The charts display the actual, fitted, and predicted curves.
We also have a decomposition of the 3 pillars of the built function: level (height), trend (slope), and season (periodicity).

4.8. MMM

MMM (Marketing Mix Modeling) is a powerful tool used to evaluate the contribution of investment levers to a temporal target variable.

4.8.1. Settings

- Time Column: time variable (can be a simple index)
- Target: target variable to predict over time

Investment channels
This section allows indicating to the model which columns represent the marketing levers. This assignment is essential for the model to work properly.

Baseline
The baseline is the "natural" contribution to the target variable when all other investments stop.
- Trend: adds a slope to the baseline
- Seasonal Period: to specify seasonality

Adstock
- Carryover: carryover effect (geometric law) of an investment over several periods after it occurred
- Max Delay: maximum waiting / delay time (in periods) between the moment of investment and the appearance of its effect on the target

Ridge
- L2 Penality: strength of the constraint to handle multi-collinearity; it pulls the contribution of channels towards zero to remove unnecessary noise
- Non-Negative media: constraint that forbids media from having a negative contribution

ROI Prior
- Prior Weight: global strength for ROI priors; this value is applied by default to each channel (when the channel weight value is < 0)
- ROI - <channel>: theoretical ROI value (contribution) for each channel (optional)
- W - <channel>: strength of the channel's ROI prior; this value replaces the Prior Weight value when it is >= 0; when the value is < 0, then Prior Weight is used as a replacement value
Back-Test
- Holdout: number of points removed from the training set to test the model
- Reliability Bootstrap: number of samples used to establish the 90% confidence interval of the contributions

4.8.2. Associated Statistics

Metrics
The results consist of visualization curves as well as quality metrics (in the training and back-test samples): R2, MAE (Mean Absolute Error), RMSE (Root Mean Squared Error),

MAPE (Mean Absolute Percentage Error).

Charts
The first chart represents the contribution slices.
The second table exposes the contributions of each channel, as well as some associated metrics (adstock, delay, total investment, variance inflation factor, etc.).
The third chart displays the response curves (saturation curves) that represent the evolution of the contribution as a function of the budget. Note the presence of a specific point on each curve: the elbow, i.e., the moment when the curve quickly begins to plateau.

5. Import / Export

5.1. Solid

Location: Export tab.

- View: exports the current visualization image in PNG format

5.2. Config

Location: Workspace tab, Config panel.

Export
- Export All Config: exports some of the projection settings in JSON format
You can also choose to export only a sub-part: Solid Config, Projection Config, Data Config

Import
- Import Config: re-imports previously exported configuration files

5.3. Scene

Location: Export tab

- Projected Data: exports projected spatial data in CSV format
- Personality Responses: exports all questionnaire responses (including pre-recorded responses for all profiles in all frameworks) in CSV format

5.4. Analytics

Location: Export tab, or at the top right of each analytical tab (Stats, Factor, Clustering...):

You will find all possible extractions, organized by module:
- Stats: descriptive statistics, correlation, profiling, association
- Hy-Personality: metrics related to projected responses of the personality form

- Factor: contributions, squared cosines, and variances related to the factorial projection
- Cluster: clustering metrics, curves, and contingency tables
- Sup ML: metrics, coefficients, and curves from regression/classification models
- Forecast / Time Series: metrics of the exponential smoothing (Holt-Winters)
- MMM: contributions, slice curves, and saturation curves of the MMM

5.5. Info Points

When you click on a point in the scene, some data is displayed in the right side panel. At the very bottom of this panel, there is a button to download this data.

6. Test Split

If the flag_test column is not defined, it is possible to create it and fill it automatically and randomly by setting a test percentage (Test Size) and clicking on Random Split.

* * *