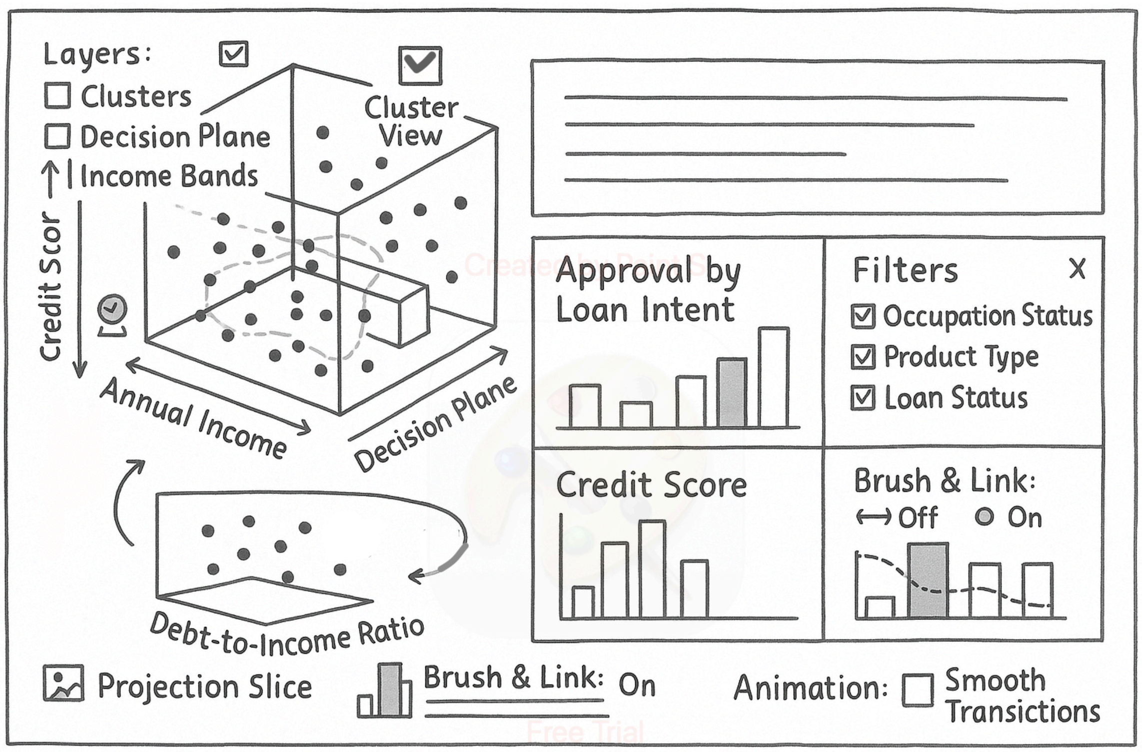
TU Wien
Information Visualization
Part 3

We created a two-part visualization system that approaches loan application data from complementary perspectives. The system combines population-level exploration with individual-level inspection, enabling both broad pattern analysis and detailed case examination.

The Population View (Sketch #1 GenAI was used to tune the image)

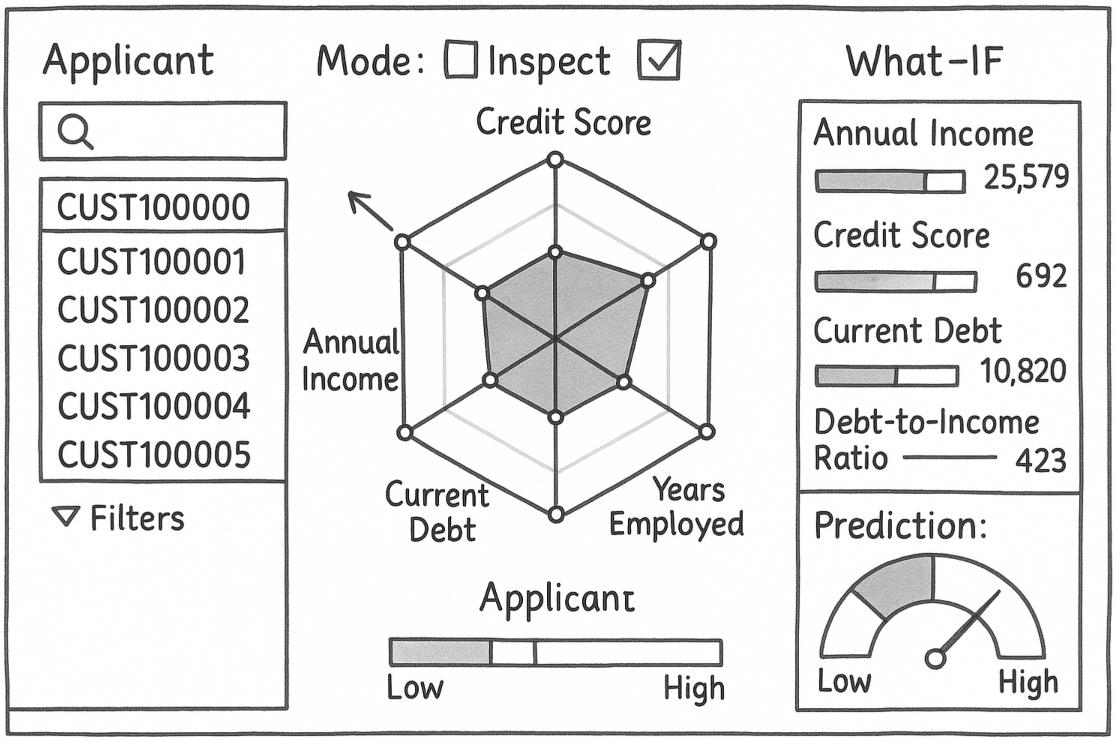


Our first visualization focuses on revealing patterns across thousands of loan applications simultaneously. We designed a 3D scatterplot centered on the three most critical risk factors:

- Credit Score (Y-axis) - Historical repayment reliability
- Annual Income (X-axis) - Financial capacity
- Debt-to-Income Ratio (Z-axis) - Existing financial obligations

Each point in this space represents one loan application, color-coded by approval status (green for approved, red for rejected). Supporting views surrounding the main scatterplot display distributions and breakdowns by categorical variables such as occupation status and loan product type. All views are linked through brushing interactions, meaning selections in one view automatically highlight the corresponding data in all other views.

The Individual Profile View (Sketch #2 GenAI was used to tune the image)



While the 3D view excels at spotting population-level trends, loan decisions ultimately depend on individual circumstances. Our second visualization addresses this need by shifting focus to single applicant profiles and scenario exploration.

The centerpiece is a radar chart displaying a selected applicant's financial profile across five key dimensions: Credit Score, Annual Income, Current Debt, Years Employed, and Debt-to-Income Ratio. The polygon formed by these values creates a distinctive "shape" that conveys the overall balance of an applicant's financial characteristics. A light reference outline shows either typical approved profiles or acceptable value ranges for comparison.

An applicant list on the left allows browsing or searching for specific cases. When an applicant is selected, the radar chart animates smoothly to display their profile, maintaining visual continuity during transitions.

The interface includes a mode selector at the top offering two distinct modes:

Inspect Mode: All values remain fixed, displaying the profile exactly as recorded in the dataset. This mode serves analytical purposes where understanding actual outcomes is the goal.

Simulate Mode: Each radar axis becomes interactive through draggable handles. Axes can be pulled outward to increase values or pushed inward to decrease them. For instance, dragging the Credit Score axis upward simulates a higher credit score, while pushing the Current Debt axis inward reduces simulated debt levels.

A What-If panel on the right provides sliders and numeric inputs for precise value adjustments. Below the controls, a prediction gauge continuously updates based on a lightweight model, showing how changes to individual attributes affect estimated approval probability.

Selection of Core Variables

We chose Credit Score, Annual Income, and Debt-to-Income Ratio as primary dimensions because they represent the fundamental components of lending risk assessment. Credit score indicates repayment likelihood based on historical behavior, income demonstrates repayment capacity, and debt-to-income ratio reveals existing financial strain. These three variables capture the essential risk equation that drives lending decisions.

The 3D Approach

The decision to use 3D visualization required careful consideration. While 3D representations introduce known perception challenge - particularly regarding depth judgment and occlusion - the ability to observe how all three risk factors interact simultaneously provides insights that multiple 2D charts cannot. Splitting these variables across separate views forces mental integration that may lead to missed patterns or incorrect conclusions about multivariate relationships.

To address the limitations of 3D visualization, we implemented rotation controls allowing the space to be viewed from different angles, and a projection slice feature that can lock one dimension to show 2D cross-sections while maintaining access to the full dimensional structure.

Layer System Design

The layer system (clusters, decision plane, income bands) provides optional interpretive overlays that can be toggled on or off. This design acknowledges that different users approach the data with different questions: loan applicants may want to understand decision boundaries, analysts may need to identify risk segments, and researchers may prefer examining raw clustering patterns. The layer approach serves these diverse needs without forcing a single interpretation onto all users.

Individual Profile Rationale

The second visualization connects population-level insights with explanations of individual outcomes. While the 3D scatterplot reveals global risk patterns, understanding why a specific applicant was approved or rejected requires examining their particular combination of characteristics.

The radar chart format works well for this purpose because it displays multiple financial attributes simultaneously, creating an intuitive visual "signature" for each profile. The goal is not precise measurement but rather showing the balance, strengths, and weaknesses across dimensions at a glance.

With real-time prediction feedback, the tool becomes both explanatory and exploratory. Counterfactual scenarios can be tested - what if this applicant had a higher credit score? What if their debt were lower? - with immediate visual and quantitative feedback showing how approval likelihood would change.

Visual Encodings

Population View Encodings

Position encodes the three main risk variables (credit score, income, debt-to-income ratio) as X, Y, and Z coordinates. We selected position because it represents the most accurately perceived visual channel according to visualization research.

Color distinguishes loan status, with green indicating approved applications and red indicating rejected ones. When cluster overlays are activated, color switches to a blue-yellow-red scale representing low-to-high risk segments.

Opacity and size indicate selection state during brushing interactions. Selected points appear at full opacity and larger size, while unselected points fade to provide clear visual distinction. Supporting charts use standard bar height encoding for frequencies and counts.

Individual View Encodings

In the radar chart, applicant attributes are encoded through polygon geometry. Each axis represents a key financial feature, and the applicant's values form a filled polygon at corresponding positions along these axes. The shape itself conveys overall balance, while the reference outline provides context for comparison.

During simulation, dragging an axis endpoint updates both the position on that axis and the polygon's geometry in real-time. The What-If panel mirrors these changes through slider positions and numeric labels.

Prediction feedback uses gauge encoding, where needle position and color intensity indicate approval likelihood. This provides both a quantitative readout and a quick qualitative assessment.

Interaction Methods

Population View Interactions

Brushing and Linking: Selecting points in the 3D view automatically highlights them across all charts. Supporting views update to display statistics for the selected subset only, eliminating the need for manual cross-referencing.

Filtering: The dataset can be narrowed by categorical variables (occupation status, product type, loan status) or continuous ranges (credit score, income, debt-to-income). Filters combine with brushing for progressive refinement from broad patterns to specific segments.

Layer Toggling: Interpretive overlays (clusters, decision plane, income bands) can be shown or hidden on the 3D view. This allows activation of only the layers relevant to the current analytical task.

3D Navigation: Mouse controls enable rotation (drag), zoom (scroll), and panning (right-click-drag). Rotation reveals occluded points and exposes patterns from different viewing angles.

Projection Slices: One dimension can be locked to display 2D cross-sections of the 3D space, providing 2D clarity while maintaining access to the full dimensional structure.

Tooltips: Hovering over individual points reveals complete applicant details, bridging the gap between aggregate patterns and specific cases.

Individual View Interactions

The individual view interactions mirror principles from the population view but adapted for single-case analysis. Selecting an applicant from the list updates all corresponding interface elements - the radar chart, numeric fields, and prediction gauge - simultaneously.

In simulation mode, dragging axes triggers what-if computation and updates both the radar polygon and prediction value dynamically. Sliders provide an alternative input method for users who prefer precise numeric entry over direct manipulation.

Task Support

The two visualizations work together to support a coherent analytical workflow. The 3D scatterplot enables examination of overall distribution patterns across approved and rejected applications, application of filters to study the effects of specific variables, and use of clustering or decision-plane overlays to identify segments and boundary patterns. Rotation and slicing help clarify multivariate interactions that remain hidden in static 2D views.

The applicant simulator extends this analysis by enabling selection of specific applicants, visualization of their financial profiles in radar format, and experimentation with hypothetical adjustments to evaluate how changes in individual features alter predicted outcomes. Through brushing, filtering, and synchronized updates, the system supports seamless movement between broad pattern analysis and detailed case-by-case reasoning.

Strengths

The simultaneous display of three core risk factors in the 3D view makes their combined effects immediately visible without requiring mental integration of separate charts. Coordinated views eliminate the need for manual cross-referencing by carrying selections across all visual components automatically.

The layer system accommodates different user needs by offering high-level cluster overviews for executives alongside detailed filtering and inspection tools for analysts reviewing edge cases. These capabilities extend naturally into the applicant simulator, where individual profiles can be examined and adjusted without losing the broader context provided by the main dashboard.

Throughout both views, all 50,000 applications remain accessible as individual data points rather than aggregated summaries. This preserves transparency and supports both population-level exploration and case-specific reasoning without information loss.

Limitations

The 3D visualization introduces several limitations. Spatial reasoning demands may challenge some users, and occlusion and depth perception issues persist despite rotation and projection aids. Rendering all 50,000 points interactively can strain performance on older systems, potentially requiring WebGL optimizations or level-of-detail techniques.

Because the main view displays only three of the eighteen predictor variables at once, exploring the full dataset requires sequential filtering or dimension swapping. Statistically derived clusters may not correspond to domain-specific risk categories, and their interpretation assumes statistical familiarity.

The overall interface, with its multiple layers, filters, and interaction modes, presents a learning curve for casual users. This complexity increases further when combined with the applicant simulation view and its dual modes of operation.

Possible Extensions

Several extensions could enhance the system further. If temporal information were available, a time-based view could reveal how lending standards and approval patterns shift across different periods.

The applicant simulator already suggests potential for a fuller predictive mode where hypothetical profiles could be input directly and estimated outcomes displayed immediately, without requiring an existing applicant as a starting point.

The 3D view could support model comparison by overlaying decision boundaries from different algorithms - logistic regression, random forests, neural networks - allowing assessment of how various modeling approaches treat the same applicant space differently. This would provide insights into model behavior and potential biases in algorithmic decision-making.