TU Wien
Information Visualization
Part 2

To guide our visualization design, we analyzed the structure of the dataset, the domain of consumer lending, and the kinds of users who would realistically interact with such a tool. This helped us understand which visual encodings would be meaningful, which interactions users would expect, and what tasks our interface should ultimately support.

Data Characterization

The loan approval dataset is fundamentally multidimensional, containing a mixture of continuous numerical features (such as annual income, credit score, loan amount, debt-to-income ratio), discrete counts (delinquencies, derogatory marks), and several nominal categories (loan intent, occupation status, product type). As we described earlier in Part 1 , the dataset includes 50,000 rows and 20 attributes, with no missing values and a realistic distribution based on lending practices in the US and Canada.

Because the dataset includes many predictors of different types, it cannot be adequately understood through a single chart or a simple summary. The relationships among the variables, especially income, credit score, and debt-to-income ratio are inherently multivariate. For this reason, our design relies on multiple coordinated views and on a 3D scatterplot for the core representation. Although the data is not hierarchical, temporal, or network-based, the large number of interacting variables makes it an ideal candidate for visual analytics techniques discussed in the course, especially those related to perception and multivariate visualization.

In addition, several variables are domain-meaningful “risk signals.” For example, past defaults, very high DTI, and extremely low credit scores are not data errors but intentional edge cases generated to reflect realistic banking scenarios . Understanding how these combinations affect approval outcomes is one of the central analytical challenges the data presents.

User and Domain Analysis

The domain in which this dataset lives is consumer lending and credit risk assessment. This domain has its own terminology, conventions, and expectations. For example, financial analysts are accustomed to working with ratios such as debt-to-income and loan-to-income, and they are aware that approval decisions depend on multiple factors simultaneously. Because of this domain context, our visualizations must support comparisons across both numerical and categorical attributes, and they must communicate risk in an interpretable way.

We imagine three primary user groups for our tool.
First, credit analysts and underwriting staff, who already have a solid understanding of lending criteria and would use the tool to explore approval patterns or identify

unusual cases. These users are comfortable with data visualization but benefit from interactive filtering, layer toggles, and the ability to compare subgroups quickly.

Second, we consider students or newcomers to finance, who may not fully understand how all the variables interact. For them, visualization can make the credit approval process more transparent something we emphasized in the introduction of Part 1, where we described lending decisions as feeling like a “black box” to most applicants . These users need intuitive interactions, clear labels, and the option to hide more advanced layers.

Finally, managers or decision-makers represent a third audience. They typically need high-level overviews clusters, trends, and distributions rather than detailed individual profiles. For them, features like the cluster layer or the approval-by-intent view provide a quick sense of overall patterns.

Across all user groups, familiarity with visualization varies, but the tasks remain similar: exploring the structure of the dataset, identifying risk patterns, and understanding how approval decisions relate to different applicant characteristics. To accommodate this range, we designed the interface so that it can be simple at first glance but expandable through toggles and layers for experienced users.

Tasks and Goals

From the combination of data characteristics and user needs, we identified several key tasks that our visualization should support. One of the central goals is enabling users to explore correlations between major numerical risk factors and approval outcomes. Because variables like income, credit score, and debt-to-income ratio jointly influence decisions, it is important for users to see how these dimensions interact rather than analyzing them in isolation.

A second major task is comparing categories such as loan intent, occupation status, or product type. Different groups of applicants have different approval rates, and these distinctions often reveal important business insights. For example, “Education” loans have significantly higher approval rates than “Debt Consolidation” loans in the dataset; a bar chart helps users immediately spot these differences.

A third task is detecting anomalies or edge cases, such as applicants with extremely high DTI but very high income, or individuals who appear “borderline” relative to typical approval boundaries. Since the dataset intentionally includes such cases to mimic real banking situations , supporting anomaly detection helps users understand rare but important patterns.

Finally, users should be able to perform subgroup exploration, for example by filtering applicants by employment status or brushing a region of the 3D scatterplot. Once a subset is selected, all supporting views update to reveal its unique distribution across loan intents, credit scores, and other attributes.