ChatGPT

Research Report – Understanding Mimica's Platform and Dataset (v5)

This report synthesises insights from the Mimica platform screenshots, the challenge brief and our synthetic data. It identifies the core features, user journeys, data requirements and gaps to inform the development of an MVP analytics platform that supports process standardisation across regions.

1. Problem Context and Personas

The Mimica platform helps organisations capture every click, keystroke and screen that employees interact with. Its **Mapper** module transforms this raw data into detailed **process maps** and summarises metrics like **time saved**, **automatability**, **actions**, **semi-structured inputs**, **decisions**, **applications** and **websites** used. The challenge requires building a new analytics module to enable **global process standardisation**, allowing leadership to compare how processes are executed across regions and identify the best-practice workflow.

The primary persona is the **Process Analyst / Continuous Improvement Manager**, who investigates workflow variation and proposes changes. Secondary personas include **Operations Managers** (oversee adoption of standard processes), **Transformation Leads** (drive change initiatives), **Subject Matter Experts** (contribute recordings and feedback) and **Compliance Officers** (ensure regulatory alignment).

2. Screenshot Analysis and Feature Inventory

2.1 Process List & Summary (image 1)

The left panel lists captured processes with columns for **Ease**, **Automatability** and **Time Spent** per year. Users can search and filter processes. The right panel shows a detailed summary for the selected process:

- **Metadata** – created date, recorded SMEs, total SMEs and frequency (transactions per day).
- **Time Saved** – large metric (e.g., 1301 hrs/year), accompanied by automatability rating, per SME/day time saved and number of SMEs.
- **Ease of Deployment** gauge – semi-circular chart showing relative complexity across **actions**, **virtualised actions**, **structured inputs**, etc.
- **Activity Counts** – numbers of actions, semi-structured inputs, decisions, applications, websites and decision paths.
- **Application & Website Usage Charts** – donut charts summarising percentage of time spent in each application (e.g., SAP 42 %, Outlook 26 %) and website domain.

These features provide a high-level overview and help users decide which process to explore further.

2.2 Process Map & Step Details (images 2 & 3)

The process map view visualises the end-to-end workflow as a flowchart with rectangular nodes for actions and diamond nodes for decision points. A toolbar allows zooming and toggling between **Transactions** mode (view individual executions) and aggregated map mode. Selecting a node reveals a right-hand panel displaying:

- **Step Name & ID** – descriptive label (e.g., "Read: Purchase Requisition Number") and unique identifier.
- **Automatability Indicator** – qualitative rating (Fully Automatable, Partially Automatable, etc.).
- **Metrics** – step duration (e.g., 3 s), percentage of variance across executions and automatability percentage (e.g., 77.4 %).
- **Screenshot Thumbnail** – miniature screenshot of the screen when the step was recorded.
- **Application Tag** – indicates which application was used (e.g., Chrome, File Explorer, SAP).

This view allows users to trace the sequence of steps, examine variations and identify where branches occur. The **Legend** dropdown describes node colours and shapes. The ability to switch to **Transactions** mode suggests the platform can overlay individual execution paths on the aggregated map.

3. Data Evaluation and Gaps

We generated a synthetic dataset (`new_synthetic_invoice_data.json`) to mirror Mimica's data. It contains ~4,153 step records across 500 transactions, with fields for **region**, **user_id**, **role**, **variant**, **action_name**, **application**, **duration**, **start/end times**, **decision outcomes** and **automata scores**. Aggregated metrics were derived (region metrics, variant distribution, step metrics, bottlenecks, step type counts, application usage).

Comparison with the screenshots reveals several gaps:

- **Time Saved & Per-SME/Day Metrics** – The synthetic data does not include baseline vs. improved durations to compute time saved. We approximated cycle times but cannot calculate automation savings or per-SME/day savings without additional assumptions.
- **Ease of Deployment Gauge** – The platform classifies steps into **actions**, **virtualised actions**, **structured inputs** and **semi-structured inputs**. We added step type classification (action, semi-structured, decision, virtualised) and computed counts, but further calibration is needed to produce ease-of-deployment scores.
- **Decision Paths** – The platform counts the number of decision paths; our dataset includes decision outcomes but does not summarise unique paths. We must compute path counts by analysing variants and branching points.
- **Screenshots** – Real Mimica shows screenshot thumbnails; our synthetic dataset lacks screen content. For the MVP we can either omit this feature or use placeholder images.
- **Website Usage** – The dataset tracks applications but not website domains. We can approximate websites by treating certain applications (e.g., SharePoint) as websites or adding a `website` field.
- **SME & Frequency Metrics** – The platform displays number of SMEs and frequency (transactions per day). Our dataset includes user IDs but not recording dates; to compute frequency, we need timestamps or sample frequencies.

Despite these gaps, the synthetic data provides sufficient detail to replicate the key analytics features: cross-region comparison, variant analysis, bottleneck detection and process map overlay.

4. Process Map Generation

To replicate the process map:

1. **Sequence Extraction** – Group step records by `transaction_id` to produce ordered sequences of `action_name`. Include decision points by marking steps where `decision_outcome` is set.
2. **Graph Construction** – Create a directed graph where nodes represent unique actions and decisions. Add edges between consecutive actions in each transaction. Count how many times each edge is traversed to compute edge weights.
3. **Variant Overlay** – For each region or variant, compute edge weights separately. Use colour coding or thickness to overlay multiple graphs on the same canvas. Identify decision nodes where flow splits.
4. **Node & Edge Metadata** – Aggregate durations, automata scores and counts per node/edge to compute average times and variability. Use these to size nodes or colour them by automation potential.
5. **Layout & Rendering** – Apply a hierarchical layout algorithm (e.g., Sugiyama) to arrange nodes. Provide pan/zoom interactions and a legend explaining colours and shapes. Allow switching between aggregated and per-transaction views.

5. User Workflows & Data Flows

User workflow (existing platform):

1. **Select a Process** – The analyst navigates to the Mapper, searches for “Create Purchase Order” and clicks the row.
2. **Review Summary** – The summary panel shows high-level metrics; the analyst sees the process has “Very High” automatability and 1,301 hrs/year time saved. They note the high number of decisions (29) and applications (11).
3. **Open Map View** – Clicking “View” opens the process map. The analyst examines the flowchart, identifies a branch in the middle and selects nodes to view details. They notice a long sequence of repeated steps (e.g., 20+ identical actions) and a bottleneck where a user spends 3 s in File Explorer.
4. **Investigate Variance** – The analyst toggles to “Transactions” to compare individual executions. They view recorded screenshots and note differences between SMEs or regions.
5. **Export Map** – The analyst exports the map or metrics to share with stakeholders (functionality inferred from “Export map” button).

Data flow:

1. **Raw capture** → 2. **Processing & mapping** (task mining) → 3. **Metric computation** (time saved, counts) → 4. **Visualisation** (process list, summary, flowchart) → 5. **Export & reporting**.

6. Repository & Documentation Gaps

The `mimica-challenge` repository contains working files (conventions, CLAUDE_PROCESS.md, planning, event-stream) and a previous technical plan for Next.js implementation. However, it does not yet incorporate our enhanced data model, step classification or extended analytics features. The design system lacks detailed guidelines for replicating the Mimica aesthetic (nav layout, cards, charts) and the PRD does not cover process standardisation. These documents must be updated to reflect:

- **Complete data schema** including step types, decision path counts, baseline vs. improved durations, SME counts and frequency.
- **UI component inventory** matching Mimica's features (process list, summary cards, gauges, charts, flowchart overlay, step details panel).
- **Technical architecture** for two phases: cloning existing features using synthetic data; implementing the MVP cross-region analytics and standardisation module.
- **Design system** reflecting the look & feel (pastel palette, minimalism, 4 px grid, accessible interactions) and 21st.dev component mapping.

7. Conclusions

The Mimica platform offers deep task-level insights through an intuitive UI with summary metrics and interactive flowcharts. Our synthetic dataset approximates the structure needed to replicate many features but lacks several high-level metrics. By expanding the data with step classifications and aggregated metrics, we can build a credible prototype. The next steps are to update the PRD, design system, data guide and technical implementation plan to align with these findings and to deliver an MVP analytics dashboard that first clones Mimica's existing experience and then extends it to enable process standardisation across regions.
