Case Study: Loan Application

Sena Toker

Business Information Systems

Prof. Paolo Ceravolo

a.a. 2024 - 2025

# Description of the case study

The case study focuses on a loan application process from a financial institution extracted from the BPI Challenge dataset. (BPI 2012).

**Application area**: Financial services (Load management)

**Actors:** Applicants, financial employees, automated systems

**Dataset: “**financial\_log.xes” – contains over 20.000 events and more than 13.000 cases, each representing a loan application.

**Peculiarities:** Multiple parallel human tasks-The process includes many human performed activities (validation, contacting customers ext.)

High process variability-There are 266 variants, which indicate flexibility but also raises concerns about standardization and efficiency.

Rich event attribute-Each case contains detailed data such as requested loan amount, the resources and timestamps. This detailed data facilitates the analyzing of performance, resource utilization and outcome correlations.

Different outcomes-This process does not have a single outcome, instead cases have three distinct endings: A\_APPROVED, A\_DECLINED and A\_CANCELLED. Therefore, outcome-based segmentation is needed to analyze data.

# Organisational goals

**Strategic Goal:** Increase the efficiency and reliability of the loan approval process maintaining the customer customer satisfaction.

**Operational Goals:**

1. Minimize process delays and reduce average throughout time.
2. Lower rejection rates without compromising credit standards.
3. Reduce the number of reworked activities, especially in cancelled applications.
4. Optimize resource allocation by reducing unnecessary employee interventions.

**Tactical Objectives:**

1. Identify and eliminate redundant steps.
2. Enhance process automation through smart validation and pre-screening.
3. Reorder tasks to detect early disqualification cases.
4. **Set timeout rules** to automatically cancel inactive applications and reduce process clutter.
5. **Introduce smart input validation** in the loan request form to reduce ineligible submissions.

# Knowledge Uplift Trail

The analytical workflow followed these steps:

1. **Data Preprocessing**:

1.1. Removed 0-time cases

1.2 Filtered for complete cases ending with valid outcomes

1.3 Selected top 50 most frequent variants for analysis

1. **Segmentation**:

2.1 Separated cases based on final outcome: APPROVED, DECLINED, CANCELLED

1. **Effectiveness Indicators**:

3.1 Analyzed case size, duration, requested loan amount, rework counts

1. **Complication Analysis**:

4.1 Identified high rejection rates and long processing times for specific paths

**Research Questions Addressed (BPI 2012)**:

1) What causes long throughput times?

2) Which resources are bottlenecks?

3) How can variants be improved?

.

# Project Results

Each step corresponds to annotated code in the Jupyter notebook:

**1.Filtering and Preprocessing** – [2]

Removed incomplete cases and identified valid legal ends (A\_APPROVED, etc.).

**2.Variant Filtering** – [2]

Reduced noise by selecting 50 most frequent process variants.

**3.Redundant Activities** – [3]

**Reduced redundant steps (A\_PARTLYSUBMITTED)**

**4.Segmentation** – [4]

Labeled cases by outcomes and isolated those of interest.

**5.Boxplot Analyses** – [5]

Visualized case durations, event counts, rework counts, and loan amounts.

**6.Complication Assessment** – [6]

Compare the segments to verify their significant correlations with properties that may be connected to effectiveness (case size, case duration, requested loan amount, reworked activities).

**7.Proposed Improvements** – [7]  
Recommended automation in repetitive manual validations and early rejection identification.

# Conclusions

The process shows signs of inefficiency due to:

* High rejection rate (≈84% of complete cases)
* Long delays in cancelled cases (avg >100 hours)
* Repetitive employee-driven steps consuming significant time

While the project successfully answered the main research questions, one limitation was the absence of approved cases in filtered logs, possibly due to noise filtering or rare event frequency.

**Recommendations**:

* Enhance early-stage automation to detect unqualified applications
* Reorder fraud checks and validations to earlier phases
* Simplify complex paths in low-frequency variants

**Original Contribution**:

* Custom method to identify redundant activities via transition and co-occurrence analysis (not covered in class)
* Extended segmentation analysis using outcome-based slicing

Assessment Criteria

The project is evaluated based on 4 dimensions each accounting for the 25% of the final grade.

1. Clarity. The report is clear and well organised. The oral presentation of accomplishments is organic and valid in terms of the terminology used.
2. Completeness. The techniques adopted are covering all the methods presented in the course (or more).
3. Effectiveness. The research questions presented in the BPI challenge are successfully answered or the motivations for not fully answering the questions are identified and clearly stated.
4. Originality and Complexity. The techniques adopted are not trivial or propose original methods.