# Credit‑Card Fraud Detection Dashboard – Prototype Testing Report

Itamar Oren‑Naftalovich  
COMM 4880/6880 – Interactive Data Visualization  
June 21, 2025

## A. Written instructions given to participants

hello! thanks for helping us improve an interactive dashboard for credit‑card fraud detection (and also improve my prototype).  
  
what will happen:  
1. i’ll give you nine short tasks to perform on the site.  
2. while you work, please say everything you’re thinking (even small reactions).  
3. when you finish the tasks we’ll have a brief interview, followed by a short survey.  
  
time needed: about 25 minutes.  
confidentiality: your answers are for a university assignment only and i will not include who said what.  
you are not being tested—the interface is. if anything feels unclear, it’s the design, not you.  
voluntary: you can pause or quit anytime.  
prototype link: see section C.

## B. Task list supplied to participants

|  |  |  |  |
| --- | --- | --- | --- |
| # | view tested | scenario (spoken to participant) | success criterion |
| 1 | time‑of‑day | “during which hour do fraud attempts peak?” | identifies 02:00–03:00 as peak |
| 2 | time‑of‑day | “hide the 'total' series; what pattern do you notice?” | legend used; observes spike 01–05 am |
| 3 | amount distribution | “are fraud amounts concentrated below €200?” | verbalises 'yes—about 75 % under €200' |
| 4 | roc curve | “what is the auc of the model?” | reads ≈0.93 |
| 5 | feature importance | “which variable contributes most to detection?” | points to V14 bar |
| 6 | threshold tool | “set threshold to 0.90; what happens to precision?” | reads ≈0.97 precision label |
| 7 | threshold tool | “with threshold 0.50, is net benefit positive?” | notes net benefit turns negative |
| 8 | impact tab | “which cost metric is highest at threshold 0.75?” | identifies review cost |
| 9 | cross‑view | “switch to data tab and then back; are previous filters remembered?” | filters persist |

## C. Link to the prototype

https://rpidatavis.itamarorenn.com/credit‑fraud

## D. Do the testing participants represent the target audience?

yes. the dashboard targets fraud‑risk analysts and data‑science stakeholders at card‑issuing banks. all five participants have daily exposure to transaction‑risk tooling or analytics.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | role & domain expertise | years analysing fraud | gender | age | daily transactions reviewed |
| P1 | junior fraud analyst | 1 | f | 23 | ≈1k |
| P2 | senior fraud analyst | 7 | m | 31 | ≈5k |
| P3 | data‑science intern (payments) | 0.5 | nb | 24 | ≈0.5k |
| P4 | risk product manager | 4 | f | 29 | ≈0 (kpi) |
| P5 | ml engineer (fintech) | 3 | m | 27 | ≈2k |

## E. Interview questions and participant responses

• most straightforward task?  
 ⦿ T1 (“filtering time chart felt natural”)

• least straightforward?  
 ⦿ T8 (“impact labels were not obvious”)

• information missing?  
 threshold tooltip lacks expected recall value (P2 & P3)

• tooltip helpful?  
 yes, but 'contrast too low' (P4)

• colour misleading?  
 total vs fraud bars share hue family (all five)

• remove anything?  
 hour‑legend arrow after first click (P1)

• add control?  
 numeric input for threshold (P5)

• notice accessibility section?  
 4 of 5 did; P3 skipped

• prefer other visuals?  
 density plot for amount (P2)

• one immediate change?  
 better contrast palette (3 votes)

## F. Additional probes

|  |  |  |
| --- | --- | --- |
| context | follow‑up probe | insight(s) |
| P1 hesitates at threshold slider | “what are you looking for?” | optimal net‑benefit sweet‑spot |
| P3 toggles legend repeatedly | “tell me what you’re checking.” | effect on tooltip numbers |
| P4 scans feature list | “what did you expect here?” | hover details on each feature bar |

## G. Link to the survey (blank)

https://forms.gle/credit‑fraud‑survey

## H. Survey results

with n = 5, all respondents rated overall satisfaction ≥ 4 out of 5; therefore a single sentence suffices per guideline.

• confidence using dashboards: 4, 4, 5, 3, 4 (median = 4)

• colour‑blind‑friendly palette: 80 % ‘sometimes’, 20 % ‘never’

• ease interpreting roc curve: all ‘easy’ (5 / 5)

• precision‑recall tool usefulness: 3 ‘very’, 2 ‘somewhat’

## I. Discussion of findings

strengths

\* overview → drill‑down flow: four participants naturally used the header tabs before filters, mirroring shneiderman’s mantra.  
\* decision‑threshold insight: the slider instantly linked model metrics to business kpis, resonating with risk analysts.  
\* perceived performance: no participant mentioned load‑time issues.

problems encountered

issue | evidence | violated guideline  
--- | --- | ---  
legend colour overlap (total vs fraud) | mis‑reads in T2 (all participants) | wcag 2.1 contrast  
low‑contrast tooltip text | comments by P4 & P5 | readability  
threshold slider lacks numeric input | P1 & P5 tried manual entry | consistency & standards

## J. Revisions

1. new palette (#83c5be for total, #e63946 for fraud) passes wcag aa and improves category distinction.  
2. tooltip background switched to #1e1f28 and font weight increased for readability.  
3. numeric input box synchronized with threshold slider, allowing precise values.