

HOW-TO GUIDE SIX STEPS

How to calculate the ROI of your document processing

What document automation is actually worth to your function — a method you can run on your own volumes in an afternoon, with every assumption exposed rather than buried.

WHO THIS IS FOR	WHAT YOU NEED FIRST	WHAT YOU GET
Finance and operations leaders deciding whether document automation is worth doing	One real month of volume, and the loaded cost of a processing FTE	Your cost per document today, what it becomes, and the return on the difference

Every automation pitch leads with a percentage. "70% faster" tells you nothing you can act on — faster than what, and does any of it leave your cost base? This guide works the other way round: six steps, each producing a number from your own operation, ending in what the change is worth to you. It takes an afternoon and no vendor input.

01 Fix the unit of measure before anything else

Decide whether you are costing a **document** or a **page**, and hold that unit throughout. Vendors price per page; your team works per document. Record both, and convert once.

COUNT	SEPARATE	EXCLUDE
Documents per month by type, and average pages each. One real month, not a good one.	Reading and keying from judgment. Automation removes the first and leaves the second.	Anything you cannot evidence. A dropped claim beats one you defend badly.

02 Build the current cost per document in five lines

This is the number everything else rests on, and it is almost always understated because only the first line shows up in a budget. Fill all five, and leave any you cannot source at zero rather than estimating it.

COST LINE	HOW TO SOURCE IT
Direct handling	Touch time per document × loaded hourly cost
Review and approval	Second-pass checking time, at the reviewer's cost
Rework	Share of documents reopened × time to fix
Overtime and peak cover	Month-end overtime, temp and vendor spend
Error leakage	Duplicate payments, missed discounts, penalties — actuals only

Use loaded hourly cost, not salary: cost to company divided by productive hours. Five recurring errors: salary in place of loaded cost, review time left out, rework counted once when it recurs monthly, overtime averaged over the year, and leakage guessed as a share of spend.

03 Model the post-implementation cost, exceptions included

The most common inflation in these numbers is assuming the new cost is just the processing fee. It is not. Automated processing has three cost lines, and the second one — the exception queue — is the one a vendor will understate. Budget it yourself at a straight-through rate you would bet on, not the one in the deck.

Processing fee

Pages per month × per-page rate.
Predictable and the easiest line to defend.

Exception handling

Documents below threshold × review time × reviewer cost. Model at 10–20% until you have data.

One-time setup, if any

Nil for standard document types on DocXtract. Applies only where a custom model or bespoke integration is needed.

Distinguish saving from capacity. Removed hours only become cash if you redeploy the headcount or stop backfilling it. If the team stays the same size and absorbs growth instead, the benefit is avoided hiring — real, but a different line in your P&L. Decide which of the two you are buying before you approve anything.

04 The four formulas

Everything above resolves into four calculations. Keep them visible rather than buried in spreadsheet cells, so you can change one input and watch the answer move — that is how you find out which assumption you are really betting on.

1 · BASELINE COST PER MONTH

$(\text{docs} \times \text{touch mins} \div 60 \times \text{clerk rate})$
+ review + rework + overtime + leakage

2 · AUTOMATED COST PER MONTH

$(\text{pages} \times \text{per-page rate})$
+ $(\text{docs} \times \text{exception \%} \times \text{mins} \div 60 \times \text{rate})$

3 · PAYBACK PERIOD

one-time cost, if any
 $\div (\text{baseline} - \text{automated}) \text{ per month}$
nil setup ⇒ pays back month one

4 · FIRST-YEAR RETURN

$(\text{annual saving} - \text{annual fees} - \text{setup})$
 $\div (\text{annual fees} + \text{setup})$

Ramp matters in year one. If volume moves onto the platform over two months, you have ten months of benefit, not twelve. Model monthly and sum rather than annualising a steady-state figure — otherwise the variance lands in your quarter, not the vendor's.

05 A worked example, end to end

An AP function processing 20,000 supplier invoices a month, averaging two pages each. Every input is an assumption you would replace with your own — they are shown so you can check the arithmetic rather than take it on trust.

INPUT	ASSUMED	WHERE IT COMES FROM
Invoices per month / pages each	20,000 / 2	ERP document count
Touch time per invoice	6 min	Timed sample, one week
Loaded cost, processing FTE	₹280 / hr	CTC ÷ productive hours
Rework rate / time to fix	8% / 12 min	Reopened-item log
Per-page processing rate	₹1.50	DocXtract published rate
Peak overtime and temp spend	₹40,000	Payroll and vendor actuals

TODAY, PER MONTH		WITH DOCXTRACT, PER MONTH	
Handling · 2,000 hrs	₹5,60,000	Processing · 40,000 pages	₹60,000
Rework · 320 hrs	₹89,600	Overtime and temp cover	Nil
Peak overtime	₹40,000		
Baseline	₹6,89,600	Run cost	₹60,000

MONTHLY SAVING	ANNUALISED	FIRST-YEAR ROI	PAYBACK
₹6.30 L	₹75.6 L	1,050%	Month one

Illustrative arithmetic on the assumptions shown, not a projection. The run cost here is the processing fee only; exception review and residual rework are excluded, so treat this as the upper bound and add both lines from step 3 to your own model. Steady state — with go-live in month one and volume moving over across the following month, first-year benefit is roughly ₹69 L. There is no one-time setup cost for standard document types; it applies only where a custom model or bespoke integration is required. Cash impact depends on redeployment, see step 3.

06 Work in a range, and know which two inputs move it

A single number is fragile — one challenged assumption and the whole figure is in doubt. Two inputs dominate this model: measured touch time and your loaded hourly cost. Run it at conservative settings for both and at your base case, and decide on the conservative one. If the return holds there, the decision is straightforward.

SCENARIO	TOUCH TIME	LOADED COST	MONTHLY SAVING	ANNUALISED	ROI
Conservative	4 min	₹250 / hr	₹3.9 L	₹46.8 L	650%
Base case	6 min	₹280 / hr	₹6.3 L	₹75.6 L	1,050%

Commit on the conservative case, not the best one. A programme approved on conservative inputs and delivered at base case is the one that gets a second phase funded. The reverse gets cancelled regardless of what it actually saved.

SIX WAYS THESE NUMBERS MISLEAD YOU	
Percentages instead of rupees "70% faster" cannot be booked. Ask for the rupee figure and the line in the budget it comes out of.	Zero exception cost Any model showing 100% straight-through is wrong. Insist on an exception line and staff it.
Headcount counted as cash Hours removed without a redeployment plan is capacity, not saving. Know which one you are buying.	No ramp in year one Twelve months of benefit claimed on a mid-year go-live. The variance lands in your quarter.
Leakage as a guess A percentage of spend is a guess dressed as a number. Use your own actuals or leave it out.	No post-go-live measurement Nobody re-runs the numbers at month six, so you never learn what it really delivered.

BEFORE YOU APPROVE IT

- ☐ Volume and average pages taken from one real month, by document type
- ☐ Touch time from a timed sample, split from judgment time
- ☐ Loaded hourly cost agreed with finance, on productive hours
- ☐ All five baseline lines filled, or set to zero and marked unsourced
- ☐ Exception rate and review time budgeted in the automated cost
- ☐ Ramp modelled monthly, not a steady state annualised
- ☐ Cash saving versus released capacity stated explicitly
- ☐ A re-measurement date set for month six after go-live

How DocXtract supports this

Per-page pricing makes the automated cost line a single multiplication rather than a licence negotiation, and template-free extraction means standard document types carry no one-time setup cost at all — new formats do not reopen it either. Field-level confidence scores let you set the exception rate deliberately instead of discovering it in month three.

- ₹1.5 per page, no per-format setup
- 98%+ accuracy, any document format
- 5,000 pages/hour, live in days

REPLACE THE ASSUMPTIONS

Run the model on your volumes

Send your document mix and one month of volume. We return the baseline, the run cost and the return on your own numbers — with every assumption shown, including the ones that work against us.

Book a Demo

All figures in this guide are illustrative assumptions used to demonstrate a method, not a forecast or a commitment. Your own baseline, exception rate, implementation cost and realised saving depend on your document mix, systems, policies and staffing, and should be measured in your environment. Costs in Indian rupees. No specific financial outcome is implied.