

Data Requests & Validation Needs

What only Binzagr can provide, why research cannot fill it, and the checks that will validate the simulated data.

Version 3.0 • 22 October 2026 • Companion workbook: **04_Gaps/Binzagr_Data_Request_Templates.xlsx** (Read me + 11 sheets, 6 of them pre-filled from Binzagr's own data)

Answers drop straight into the pipeline: 04_Gaps/scripts/ingest_client_templates.py validates a filled workbook.

New in version 3: a rough head count from employee ids (202 ids scan a month; ~161 paid FTE in our base), now the biggest driver of the labour cost range.

6	24	12	11
priority-1 requests: needed for the client-facing P&L comparison	data gaps in total, each with a stand-in today	validation checks the answers will settle	template sheets ready to fill

1. The six things we need first

Everything else in the PoV already runs on cleaned and simulated data. The first five turn the simulated Jeddah P&L into Binzagr's actual one; the sixth is the quickest check of the labour model, in the unit Binzagr already uses. For each, a minimum version is enough to start.

#	Request	Why it matters	Minimum acceptable	Owner	Sheet
G1	Actual Jeddah P&L, monthly, Feb-Sep 2026	It is the baseline Salman wants the simulation compared with; today the baseline margin is 6.5% by construction.	Branch-level monthly revenue, own labour, agency labour, other opex	Salman / Finance	1
G2	Attendance: clock-in / out per person per shift, by payroll number	58.9% of simulated paid hours fall outside any scan; only attendance shows if that is unscanned work, waiting, leave or idle.	Two sites x two weeks, or the shift roster	Badr / HR	2
G3	Headcount and all-in labour cost by site, role, month	Every SAR figure scales with it; who is paid moves labour by ~SAR 2M a year either way.	Monthly total labour cost and headcount per site	Finance / HR	3
G4	Contracted rate per client and activity	Revenue today is cost plus margin (circular); Badr quoted SAR 0.50-0.80 per case.	The sample client rates Badr offered	Finance / Commercial	4
G5	Site register: name, temperature, size, status, changes	Sites can only be benchmarked against like sites (ambient vs cold, relocations such as 2100-01 -> 2100-07).	Name, ambient / chilled / frozen, active / closed per site	Badr / Operations	5
G24	Current cost per case and per pallet by site	Checks our simulated labour (SAR 0.77-1.02 per loose case, 12.10-16.06 per pallet received) in Binzagr's own unit.	One month, by site	Badr / Finance	11

2. Why research cannot fill these

- **Binzagr's own money:** actual P&L, contract rates and pay. A private group publishes no warehouse accounts or rate cards.
- **Binzagr's own people and time:** who was on shift when. Scans show work, not attendance. Payroll numbers are enough; we never need names.
- **The meaning of internal codes:** clients, sites, hierarchy codes, flags, and where Binzagr's mapping disagrees with the log.
- **Data lost in the export:** pallet ids kept to their first 6 of 12 digits (proved by Binzagr's sample) and category codes cut to 15 digits. Only a re-export brings them back.
- **History that was not sent:** 8 months hold one Ramadan; yearly Hijri seasonality cannot be learned from one cycle.

3. Gap register

Priority 1 = needed for the client-facing P&L comparison or its quickest check; 2 = replaces an inference that already works; 3 = re-export, history or future scope. "Stand-in" is what the PoV uses today (document 3).

#	Gap	Kind	Stand-in today	Unknown without it	Owner	Sheet	P
G1	Actual Jeddah P&L, Feb-Sep 2026	own money	Cost-recovery P&L at 6.5% margin by construction	Simulated-vs-actual comparison; blue-collar % of revenue	Salman / Finance	1	1
G2	Attendance by payroll number	people & time	One paid shift per person session under Labour Law	58.9% of paid time unobserved (46.0% on scan days only); who is on shift: 202 ids scan a month, ~269 listed, ~161 paid FTE; badge sharing	Badr / HR	2	1
G3	Labour cost and headcount	own money	Role-tier build-up: SAR 23.86 / 24.43 per paid hour (tally clerk / storekeeper), manager 51.25; head count from employee ids (~161 paid FTE)	All SAR results (labour cost P10-P90 SAR 6.84M-11.42M)	Finance / HR	3	1
G4	Rate card per client and activity	own money	Cost-recovery rates on Binzagr's 9 revenue codes (e.g. SAR 2.17 per case picked)	Real revenue and margin by client; storage revenue	Finance / Commercial	4	1
G5	Site register	internal codes	Temperature from item mix; changes from volumes	Fair site benchmarking	Badr / Ops	5	1
G6	Client names, type, billing basis (31); in-house accounts	internal codes	26 business units / 5 external from item codes	Third-party vs internal revenue; accounts not billed	Badr / Finance	6	2
G7	What the 149 unexplained logins are	people & time	User list maps 242 of 387 badges (93.9% of scans); the rest one person per login	People vs desktops, devices or agency staff	Badr / HR	7	2
G8	Invoice-level receivables with payments	own money	FIFO allocation of van-sales balances	True ageing; wholesale receivables	Neeraj / Finance	8	2
G9	Official names of 65 hierarchy codes	internal codes	Channel from name keywords	Channel of code 10160 (SAR 18.7M)	Neeraj	9	2
G10	31 code interpretations (13 from the 22 Oct files)	internal codes	Assumptions stated in documents 1-3	e.g. code 080 listed twice; code 100 meaning; 154 items with pallet < case; ~269 listed users vs ~138 who scan	Neeraj / Badr	10	2
G11	Time-and-motion study	people & time	P50 standards by activity and line size; own top quartile	Absolute productivity	Badr / Raza	11	2
G12	WMS re-export: pallet id as text, transaction id, client on billing rows, January 2026	lost in export	Rows kept and flagged; 58 sample rows recovered	2.86M pallet ids cannot be repaired; 58,012 identical rows; 19,900 billing rows without client	Badr	11	2
G13	Product hierarchy codes as text	lost in export	Exact codes rebuilt on the product side	L5 / L6 names for 10,035 / 18,206 products	Neeraj	11	2
G14	Sales and returns history, 24+ months	history	Hijri events as features	Learned Ramadan / Eid seasonality	Neeraj	11	2
G15	WMS history, 24 months	history	8 months, one Ramadan	Workload seasonality for staffing	Badr	11	3
G16	Storage occupancy per client per day	own operations	Not modelled	Storage revenue	Badr	11	3
G17	Van loads per rep per day	own operations	Not modelled	Van-load optimisation	Neeraj	11	3
G18	Route plans and visit logs	own operations	Rep profile from sales (6 drops a day)	Route optimisation; missed visits	Neeraj	11	3
G19	Outlet closure confirmations	own operations	Churn proxy	Labels for the churn model	Neeraj	11	3
G20	Other clusters' masters	own operations	Riyadh only	Cross-cluster harmonisation	Neeraj	11	3
G21	Shelf-life master	own operations	SHELF_LIFE empty for every product	Only if the shelf-life use case returns	Neeraj	11	3
G22	HR job titles by payroll number	people & time	WMS menu level used as the role	Wage rates by job (picker, forklift driver, loader)	Badr / HR	11	2
G23	Billing volumes per client (codes 110-920)	own money	Billed units counted from the scan log	What was actually invoiced, per client and month	Finance / Badr	11	2
G24	Current cost per case and per pallet	own money	Simulated labour SAR 0.77-1.02 per loose case, 5.70-7.57 per pallet picked, 12.10-16.06 per pallet received (worked time - all paid days)	Validation of the labour model in Binzagr's own unit	Badr / Finance	11	1

Outside the PoV (no template): billing, trade spend and rebates (Salman's first priority) and logistics / transport. No data has been received for them; they would need promotion calendars, rebate agreements, invoice-level discounts and claims. Also raised on 21 Oct and not modelled: SLA penalties for missed turnaround times and outbound orders waiting in staging for late trucks (neither is tracked today). Say so explicitly rather than implying coverage.

4. Priority-1 requests in detail

G1 Actual Jeddah P&L	
What we need	Monthly revenue and cost lines, Feb-Sep 2026, by site if possible: handling, storage, VAS and other revenue; own and agency labour, overtime, equipment, rent, utilities, supervision, IT, consumables, other.
Format	Template sheet 1 (one row per month x site x line item), SAR excluding VAT.
Minimum	Branch-level monthly totals of revenue, own labour, agency labour and other opex.
Replaces	The baseline of the simulated P&L (sim_jeddah_pnl_monthly). Optimised P&L = actual - recoverable labour.
Validates	The cost structure assumptions: labour share of opex (40-60%) and margin (4-10%) - check V5.
G2 Attendance / roster	
What we need	One row per person per shift, keyed by the payroll number (the WMS user list already links 242 badges, 94% of scans, to payroll numbers): date, clock-in, clock-out, role, own or agency, home site.
Format	Template sheet 2, or a CSV with the same columns (preferred for volume). No names.
Minimum	Two sites x two weeks first: Badr is unsure how reliable the attendance data is, and this sample shows it (V1). Otherwise the shift roster.
Replaces	Simulated paid shifts and hours (sim_session_attendance) and the head count inferred from employee ids (sim_person_month_labour).
Validates	Paid hours, the head count, the 58.9% unobserved time and the badge-sharing risk tiers - checks V1 and V2.
G3 Labour cost and headcount	
What we need	Monthly headcount and all-in cost by site, role, own / agency and Saudi / non-Saudi (salary, allowances, GOSI, levy, iqama, insurance, agency fees, overtime).
Format	Template sheet 3.
Minimum	One month of total labour cost and headcount per site (V4). Job titles by payroll number (G22) then let wages differ by job.
Replaces	The researched cost build-up by WMS role (sim_labour_cost_model): cost per paid hour = actual cost / paid hours; and the inferred head count (inf_jeddah_headcount).
Validates	The loaded-cost range SAR 16.25-31.81 per paid hour for operatives and the ~161 paid FTE - check V4.
G4 Rate card	
What we need	Rate per client x service x unit, with validity dates and minimum charges, for external clients and internal recharges.
Format	Template sheet 4 (services = Binzagr's Revenue Generation codes, from a dropdown).
Minimum	The sample client rates Badr offered on 21 Oct; then the 5 external clients (3069, 3030, 3024, 3017, 3099) and the recharge basis for the business units.
Replaces	Cost-recovery rates (sim_rate_card); revenue = Actual billable units x contract rate.
Validates	Units mapping and the rate card; price vs simulated labour per unit - checks V5 and V12.
G5 Site register	
What we need	Per site: name, city, ambient / chilled / frozen, area, pallet positions, operating pattern, status, and what changed when (2100-12 silent after April; 2100-01 -> 2100-07 in August).
Format	Template sheet 5 (pre-filled with what the data shows; fill the yellow columns).
Minimum	Name, temperature regime and status for each of the 9 site ids.
Replaces	inf_warehouse_profile (temperature inferred from product names, low confidence).
Validates	Site temperature inference - check V9.

G24 Current cost per case and per pallet	
What we need	Binzagr's costing as used today: cost per case and per pallet by site and month, Feb-Sep 2026 (labour, and total if available).
Format	Any export; sheet 11 tracks its status.
Minimum	One month, by site.
Replaces	Nothing: it validates sim_cost_per_unit_wh_month.
Validates	Simulated labour per loose case picked SAR 0.77 on worked time, 1.02 incl. paid days without a scan; per pallet picked 5.70 / 7.57; per pallet received 12.10 / 16.06 - check V11.

5. Validation requests

Small, targeted pieces of data that test the simulated and inferred tables before the full data arrives. Thresholds are our proposal for agreeing that a stand-in is fit for the PoV; if a check fails, the stand-in is re-based or replaced. Document 3, section 5.2, gives the method behind each check.

#	What we ask Binzagr for	What we compare it with	Proposed pass threshold	Effort	Gap
V1	Clock-in / out by payroll number for 2 sites x 2 weeks	Simulated paid hours per person-day and site-month; are payroll people on shift on days they do not scan?	Site-month within +/-15%; median person-day error <= 1.5 h	Medium	G2
V2	Who or what the logins behind most high / medium-risk sessions are (MDAA210004, HM1008, MDM210001 + MDM210007 used together, AHH2100004, MHA210004, BB210004, BHAS210004 ...)	Badge-sharing risk tiers	At least 70% of high-risk sessions explained by two people or a shared / system login	Low	G7
V3	The time-and-motion study, or the 3 best-run sites / shifts	Data-derived P50 standards by activity and line size; our productivity ranking	Within +/-25% for 8 of the 10 activity / line-size pairs with most hours; rank correlation >= 0.5	Low-Medium	G11
V4	One month of actual labour cost and headcount per site	Simulated cost per paid hour and paid FTE per site	Inside SAR 16.25-31.81 per hour (operatives); FTE within +/-10%	Low	G3
V5	Contract rates for the external clients + monthly P&L revenue	Simulated rate card and revenue	Revenue within +/-5%; rate gap reported per service	Medium	G4, G1
V6	Receivables ageing report at 20 Oct 2026 (bucket totals)	FIFO ageing buckets	Each bucket within +/-5 percentage points	Low	G8
V7	Official names of the 20 CUST_HIER2 codes	Inferred channel labels	Agree for codes covering >= 80% of customers	Low	G9
V8	Outlets known to have closed in 2026	Churn proxy ("lapsed")	Precision >= 60%; recall reported	Medium	G19
V9	Client list (names, type) and site list (temperature, status)	Inferred client types and site profiles	31 of 31 client types; 2100-14 / 2100-15 cold-chain confirmed	Low	G5, G6
V10	Confirm or correct the 31 code interpretations (sheet 10)	Assumptions in documents 1-3	Each confirmed or corrected	Low	G10
V11	Current cost per case and per pallet by site	Simulated labour per loose case picked (SAR 0.77-1.02), per pallet picked (5.70-7.57), per pallet received (12.10-16.06)	Within +/-20% per site-month	Low	G24
V12	The sample client rates (promised 21 Oct)	Simulated labour per unit and cost-recovery rates	Reported per activity; a price below simulated labour is checked against V11 before it is called a finding	Low	G4

Already validated without Binzagr's help: volumes in the simulated P&L equal the actual WMS counts; the official Eid and National Day dates appear as closures in the warehouse data; client types agree with item-code matching; receivables reconcile to the balances; splitting outbound labour between case and pallet picks by billed lines matches actual pick time (80.4% vs 80.1%); every random step is seeded, so the tables regenerate (document 3, section 5.1).

6. How to send the data

Use the workbook **Binzagr_Data_Request_Templates.xlsx**. Yellow cells are to be filled; grey columns are pre-filled from Binzagr's own data; row 3 of the blank sheets is an example and is ignored on import.

Sheet	What to fill	Key columns	Pre-filled
1 Jeddah P&L	Monthly P&L lines	month, site, line_item (dropdown), amount_sar	no
2 Attendance	One row per person per shift	payroll_number (as in the WMS user list), date, clock_in, clock_out, role, employment_type	no
3 Labour cost	Headcount and all-in cost	month, site, role, employment_type, nationality_group, headcount, total_cost_sar	no
4 Rate card	Rates per client and service	client_code, service (Binzagr revenue code, dropdown), unit, rate_sar, effective_from / to	no
5 Sites	Name, temperature, size, status, changes	wh_id (9 sites listed)	yes
6 Clients	Name, type, billing basis	client_code (31 listed)	yes
7 Badges	What each unexplained login is	employee_id_norm (149 listed: not in the user list, or flagged)	yes
8 Receivables	Invoice-level open items and payments	dist_cd, cust_cd, doc_no, dates, amounts	no
9 Channel codes	Official name and channel	level + code (65 listed)	yes
10 Code meanings	Correct or confirm	31 interpretations listed	yes
11 Re-exports & history	Status and date of larger extracts	15 datasets listed (incl. cost per case / pallet and sample rates)	yes

- Dates YYYY-MM-DD; timestamps YYYY-MM-DD HH:MM (24 h, Saudi time). Amounts in SAR excluding VAT.
- Codes exactly as in Binzagr's systems; code columns are formatted as text so Excel keeps long codes intact. Re-exports as CSV, with ids as text.
- Large tables (attendance, receivables, re-exports): a CSV or zip with the same column names is better than pasting.
- People: payroll numbers are enough; we never need names or phone numbers.

7. Questions for the next standup

Who	Question
Salman	Can Finance share the Jeddah P&L by month for Feb-Sep, at least at branch level?
Salman	How are the 26 business-unit clients charged (internal recharge basis), and which accounts are in-house (not billed)? Who are the 5 external clients (3069, 3030, 3024, 3017, 3099)?
Badr	Can you export attendance (clock-in / out) by payroll number for two sites x two weeks, so we can check it against the scans?
Badr	What are the logins not in the user list (WA.AD.nnn, plain numbers like 751, HM1008, AMR2100, NB99)? What is BHAS210004 used for (a storekeeper login posting ~2,140 confirmations a day in batches)?
Badr	The user list shows ~269 people with an active login at the Jeddah sites each month, but only ~138 payroll people scan in a month, on 73% of their working days. Who are the others (office staff, drivers, people on leave, accounts never closed)? Is there a monthly headcount per site?
Badr	In your mapping, is code 080 In House or Duplicate, and is code 100 "Picked Volume" or "Item Import"? Where do codes 110-920 (Received HU's, Picked UOS, Shipped HU's ...) come from - a billing table?
Badr	Can we have the sample client rates, your current cost per case / per pallet by site, and the time-and-motion study results?
Badr	Can the WMS log be re-exported with the pallet id as text (as in your sample) and a transaction id, including January 2026?
Badr	What happened to 2100-12 after April, and to 2100-01 -> 2100-07 in August? Which sites are chilled or frozen?
Neeraj	Can we get the wholesale receivables ageing (and invoice-level open items)? What are CUST_HIER codes 10160 and 10161, and DIST_CD 2000?
Finance / HR	Headcount and all-in labour cost by site and job title (payroll number -> job title), own vs agency, and the Saudi share.

8. Already solved - no need to ask

Looked like a gap	How it was closed	Where
Duplicate rows at month boundaries	65,141 second copies proven at raw-text level and removed	Doc 2, 3.1
Product records swallowed by a stray quote	Rebuilt: 23,645 products, matching the sales file's product list exactly	Doc 2, 4
Hierarchy codes truncated by Excel	Exact codes recovered with a rule consistent for 100% of keys	Doc 2, 4
Client code missing on billing rows	Linked when the same-second scans show one client (88.6-96.6% coverage)	Doc 2, 3.3
Which clients are external 3PL	Item-code match: 26 business units, 5 external clients	Doc 3, 3.10

Looked like a gap	How it was closed	Where
Labour Law, statutory costs, market wages, holidays	Researched with low / base / high values (38 published sources)	Doc 3, 4 and 6
Hijri calendar and 2026 events	Computed (Umm al-Qura) and checked against warehouse closures	Doc 3, 3.16
Standard times	Derived from 4.5M scans, by activity and line size (relative standards); G11 upgrades them	Doc 3, 3.3
Which activities are billed vs cost	Binzagr's transaction mapping: 9 revenue codes in the log; loading, LP audit and QC are cost	Doc 2, 3.7; Doc 3, 3.6
Who is behind each badge	Binzagr's WMS user list: payroll number and role for 242 badges (94% of scans), month by month	Doc 2, 3.7
A rough head count	Employee ids + user list: 202 ids scan a month (138 payroll people); ~269 listed active; payroll people treated as salaried -> ~161 paid FTE (115-199). G2 / G3 make it exact	Doc 2, 3.7; Doc 3, 3.4b
Quantities in cases and pallets	Binzagr's conversion factors: 99.6% of item scans	Doc 2, 3.7
Whether pallet ids can be repaired	No: the sample shows 12-digit ids kept to 6 digits; a re-export is needed (G12)	Doc 2, 3.7
01/01/1900 end dates	All are shipment-audit events with no end time recorded; left empty, not set to 2026	Doc 2, 3 (W17)
HU_ID_2, zero-second tasks	Badr: HU_ID_2 is unused (dropped); zero-second tasks are an operator practice	Doc 2, 3 (W5); Doc 3, 3.2