

Metrics & Goals Templates

North star, OKRs, KPIs and A/B testing: what each is, why it matters, and a fill-in template

1 • North Star Metric

Value statement, inputs, guardrails

2 • OKRs

Objectives, key results, scoring, check-ins

3 • KPIs

Definitions, ranges, alerts, dashboards

4 • A/B Testing

Hypothesis, sample size, decision rules

Harshal Borole · Product Manager · October 2026

Prompts in [brackets] are placeholders to replace with your own answers.

How to use these templates

These four templates help a product team decide what to measure, what to aim for, and how to prove a change worked. Each section explains the concept, then gives a fill-in template with prompts in [\[brackets\]](#). Replace every bracket with your own answer and delete any prompt that does not apply.

The four tools work at different levels, from long-term direction down to a single experiment.

Tool	The question it answers	Time horizon	Who owns it
North Star Metric	What single measure best captures the value customers get from us?	1 to 3 years, stable	Head of Product with leadership
OKRs	What will we change this quarter, and how will we know we did it?	One quarter, sometimes one year	Each team, aligned to company OKRs
KPIs	Is the business and product healthy right now?	Ongoing, reviewed weekly or monthly	Functional owners
A/B Testing	Did this specific change cause the result we wanted?	Days to weeks per test	Product manager with data partner

How they connect. The north star is the destination. OKRs are the quarterly bets to move it. KPIs are the dashboard lights that tell you the engine is healthy while you move. A/B tests are how you prove which bets actually worked.

Layer	Example
North star	Weekly active teams completing a core workflow
Company OKR	Raise weekly active teams from 4,000 to 5,000 this quarter
Team KPI	Day-7 activation rate, currently 38%
A/B test	New guided setup vs current setup, measured on day-7 activation

Writing tips that apply to all four

- Every metric needs a precise definition: what counts, over what time window, from which data source.
- Prefer rates and ratios to raw totals, which grow just because the company grows.
- Pair every target with a guardrail so a metric cannot be gamed at the expense of quality.
- Write targets with a baseline, a target value and a date: "from 38% to 50% by June 30".
- Decide how you will measure before you launch, not after you see the data.

1. North Star Metric

What is a north star metric?

A north star metric (NSM) is the single measure that best captures the core value a product delivers to customers. When it rises, customers are getting more value and the business grows as a result. It aligns every team on one outcome, so that marketing, product and engineering pull in the same direction instead of optimizing their own numbers.

Why it matters. Teams without a shared metric optimize locally: marketing drives sign-ups that never activate, and product ships features nobody adopts. A good north star makes those trade-offs visible.

Qualities of a good north star

Quality	Test question
Reflects customer value	If this rises, are customers clearly better off?
Leads revenue	Does it predict future revenue rather than report past revenue?
Measurable and frequent	Can we see it move within days or weeks?
Actionable	Can teams influence it through their work?
Understandable	Can anyone in the company explain it in one sentence?

Examples by business type

Business type	Typical north star shape
Collaboration software	Weekly active teams completing a core workflow
Marketplace	Successful transactions per week
Media or content	Total time spent with engaged content
Operations software	Jobs or tasks completed per user per week
Subscription consumer app	Weekly users who reach the core outcome

North star versus revenue. Revenue is an outcome of value delivered, not the value itself. Choose a north star that revenue follows, then track revenue as a business KPI alongside it.

Template: North Star Metric

1. Product value statement

[One sentence: the core value customers get from the product. Example shape: "[Customer] gets [outcome] when they [key action] ."]

2. North star metric

- Metric name: *[Short name]*
- Definition: *[Exactly what counts, the time window, and what is excluded]*
- Data source: *[Where the data comes from and who owns it]*
- Current baseline: *[Value and date measured]*
- Target: *[Value and date]*

3. Why this metric

Quality	How our metric meets it
Reflects customer value	<i>[Explanation]</i>
Leads revenue	<i>[Explanation]</i>
Measurable and frequent	<i>[Explanation]</i>
Actionable	<i>[Explanation]</i>

4. Alternatives considered and rejected

Candidate metric	Why we rejected it
<i>[Candidate 1, such as total sign-ups]</i>	<i>[Reason, such as "rises without customer value"]</i>
<i>[Candidate 2]</i>	<i>[Reason]</i>

5. Input metrics (the levers)

[Break the north star into 3 to 5 inputs that teams can own. Often a formula such as reach × activation × frequency × depth.]

Input metric	Definition	Owning team	Baseline	Target
<i>[Breadth: how many users]</i>	<i>[Definition]</i>	<i>[Team]</i>	<i>[Value]</i>	<i>[Value]</i>
<i>[Depth: how much each user does]</i>	<i>[Definition]</i>	<i>[Team]</i>	<i>[Value]</i>	<i>[Value]</i>
<i>[Frequency: how often]</i>	<i>[Definition]</i>	<i>[Team]</i>	<i>[Value]</i>	<i>[Value]</i>
<i>[Efficiency or quality]</i>	<i>[Definition]</i>	<i>[Team]</i>	<i>[Value]</i>	<i>[Value]</i>

6. Guardrail metrics

[Metrics that must not get worse while the north star rises.]

- *[Guardrail 1, such as customer satisfaction stays above X]*
- *[Guardrail 2, such as error rate stays below Y]*

7. Review cadence

[How often the north star and inputs are reviewed, and by whom. Example: weekly dashboard, monthly deep dive, yearly check that the metric still fits.]

Common mistakes

- **Choosing revenue.** Revenue lags value and tells teams little about what to change.
- **Choosing a vanity metric.** Downloads, page views and sign-ups can rise while customers get nothing.
- **Picking several north stars.** More than one means no clear priority.
- **No input metrics.** A north star nobody can directly influence turns into a scoreboard, not a guide.

2. OKRs (Objectives and Key Results)

What are OKRs?

OKRs are a goal-setting framework that pairs an **objective**, a qualitative statement of what you want to achieve, with 2 to 4 **key results**, measurable outcomes that prove you achieved it. OKRs are usually set each quarter, focus a team on a few important changes, and are designed to stretch the team beyond business as usual.

Why it matters. OKRs turn strategy into focus. They force a team to choose a small number of outcomes, make those outcomes measurable, and connect each team's work to company goals.

Part	What it is	Good example	Weak example
Objective	Qualitative, inspiring, time-bound direction	Make onboarding effortless for new teams	Improve onboarding
Key result	Measurable outcome with baseline and target	Raise day-7 activation from 38% to 50%	Launch new onboarding flow
Initiative	The work you will do to move the key result	Build a guided setup checklist	Not part of the OKR itself

Outcomes, not outputs. A key result measures a change in behavior or results, not a task completed. "Ship feature X" is an output; it belongs in the initiatives list. "Increase Y from A to B" is an outcome.

Committed versus aspirational. Committed OKRs are expected to be fully met, such as reliability targets. Aspirational OKRs are stretch goals where reaching about 70% is a strong result. Label each OKR so the score is read correctly.

Scoring scale

Score	Meaning
0.0 to 0.3	Little progress; learn why and decide whether to keep the goal
0.4 to 0.6	Meaningful progress, short of the target
0.7 to 1.0	Target met or nearly met; for aspirational OKRs, 0.7 is success

Template: OKRs

1. Context

- Period: *[Quarter and dates]*
- Team: *[Team name]*
- Link to strategy: *[The north star or company OKR this team supports]*

2. OKRs

Objective 1: *[Qualitative, inspiring statement of what we want to achieve]*

Type: *[Committed or aspirational]*

Key result	Baseline	Target	Owner	Confidence (1 to 10)
KR1: <i>[Increase or decrease metric from X to Y]</i>	<i>[Value]</i>	<i>[Value]</i>	<i>[Name]</i>	<i>[Score]</i>
KR2: <i>[Metric from X to Y]</i>	<i>[Value]</i>	<i>[Value]</i>	<i>[Name]</i>	<i>[Score]</i>
KR3: <i>[Metric from X to Y]</i>	<i>[Value]</i>	<i>[Value]</i>	<i>[Name]</i>	<i>[Score]</i>

Initiatives: *[The main projects we believe will move these key results]*

Objective 2: *[Statement]*

Type: *[Committed or aspirational]*

Key result	Baseline	Target	Owner	Confidence (1 to 10)
KR1: <i>[Metric from X to Y]</i>	<i>[Value]</i>	<i>[Value]</i>	<i>[Name]</i>	<i>[Score]</i>
KR2: <i>[Metric from X to Y]</i>	<i>[Value]</i>	<i>[Value]</i>	<i>[Name]</i>	<i>[Score]</i>

Initiatives: *[Projects]*

3. Alignment map

Company OKR	Team objective that supports it	Key result that connects them
<i>[Company OKR]</i>	<i>[Team objective]</i>	<i>[KR]</i>

4. Weekly check-in

Key result	Current value	Confidence this week	Blockers	Next action
<i>[KR]</i>	<i>[Value]</i>	<i>[Score]</i>	<i>[Blocker]</i>	<i>[Action]</i>

5. End-of-quarter review

Key result	Final value	Score (0 to 1)	What we learned	Carry forward?
<i>[KR]</i>	<i>[Value]</i>	<i>[Score]</i>	<i>[Learning]</i>	<i>[Yes or no]</i>

Common mistakes

- **Writing tasks as key results.** "Launch the new dashboard" measures effort, not impact.
- **Too many OKRs.** More than 3 objectives per team per quarter dilutes focus.
- **No baseline.** "Reach 50%" means little if you do not know the starting point.

- **Using OKRs for performance reviews.** Tying pay to scores pushes teams to set safe, easy targets.
- **Set and forget.** Without weekly check-ins, OKRs become a planning exercise rather than a steering tool.

3. KPIs (Key Performance Indicators)

What are KPIs?

KPIs are the ongoing measures that show whether a product or business is healthy. Unlike OKRs, which aim to change something within a quarter, KPIs are tracked continuously and alert the team when something drifts outside its normal range. A KPI can become a key result when the team decides to actively move it.

Why it matters. KPIs are the instrument panel. They catch problems early, such as rising churn or slower load times, and give leaders a shared, trusted view of performance.

KPIs versus OKRs versus north star

	North star	OKR key result	KPI
Purpose	Long-term direction	Change something this quarter	Monitor ongoing health
How many	One	2 to 4 per objective	A focused set per area, often 5 to 10
Lifespan	Years	One quarter	Ongoing
Example	Weekly active teams	Day-7 activation from 38% to 50%	Monthly churn rate

Leading versus lagging indicators. Lagging indicators report results after the fact, such as revenue or churn. Leading indicators predict them, such as activation or feature adoption. A balanced KPI set has both: leading indicators let you act before lagging ones turn bad.

Common product KPI families (the AARRR funnel and beyond)

Family	What it tells you	Example KPIs
Acquisition	Are the right people finding us?	Visitors, sign-ups, cost per acquisition, channel mix
Activation	Do new users reach first value?	Activation rate, time to first value, onboarding completion
Engagement	Do users keep using the product?	DAU/MAU ratio, sessions per user, feature adoption
Retention	Do users stay?	Cohort retention at day 7, 30, 90; logo churn
Revenue	Does the value convert into money?	MRR, ARPU, net revenue retention, LTV to CAC ratio
Referral	Do users bring others?	Referral rate, net promoter score
Quality	Is the experience reliable?	Uptime, error rate, page load time, support tickets per user

Template: KPIs

1. Scope

- Product or area: [\[Name\]](#)

- KPI owner: *[Name and role]*
- Review cadence: *[Daily, weekly or monthly, and the forum]*

2. KPI definitions

KPI	Family	Definition (formula, window, exclusions)	Leading or lagging	Data source	Owner
<i>[KPI 1]</i>	<i>[Family]</i>	<i>[Definition]</i>	<i>[Type]</i>	<i>[Source]</i>	<i>[Name]</i>
<i>[KPI 2]</i>	<i>[Family]</i>	<i>[Definition]</i>	<i>[Type]</i>	<i>[Source]</i>	<i>[Name]</i>
<i>[KPI 3]</i>	<i>[Family]</i>	<i>[Definition]</i>	<i>[Type]</i>	<i>[Source]</i>	<i>[Name]</i>
<i>[KPI 4]</i>	<i>[Family]</i>	<i>[Definition]</i>	<i>[Type]</i>	<i>[Source]</i>	<i>[Name]</i>

3. Healthy ranges and alerts

KPI	Current value	Healthy range	Alert threshold	Action when breached
<i>[KPI 1]</i>	<i>[Value]</i>	<i>[Range]</i>	<i>[Threshold]</i>	<i>[Who acts and how]</i>
<i>[KPI 2]</i>	<i>[Value]</i>	<i>[Range]</i>	<i>[Threshold]</i>	<i>[Action]</i>

4. Link to north star and OKRs

KPI	Which north-star input or OKR it supports	Relationship
<i>[KPI]</i>	<i>[Input or KR]</i>	<i>[How a change in the KPI moves it]</i>

5. Dashboard layout

- Top row: *[North star and its trend]*
- Second row: *[Input metrics and current OKR key results]*
- Third row: *[Health KPIs with alert status]*
- Breakdowns: *[Segments to slice by, such as plan, region, platform or cohort]*

6. Monthly KPI review notes

KPI	Change since last month	Explanation	Action
<i>[KPI]</i>	<i>[Change]</i>	<i>[Why it moved]</i>	<i>[Next step]</i>

Common mistakes

- **Tracking everything.** Fifty KPIs means nobody watches any of them. Keep a focused set.
- **Vague definitions.** Two teams computing "active user" differently will argue about numbers instead of decisions.
- **Only lagging indicators.** By the time churn rises, the cause happened months ago.
- **No owner or threshold.** A KPI with no one responsible and no alert level is just a chart.

4. A/B Testing

What is A/B testing?

An A/B test is a controlled experiment that shows two or more versions of a product experience to randomly assigned groups of users, then compares a chosen metric to learn which version performs better. Group A sees the current experience (the control); group B sees the change (the variant). Because assignment is random, a significant difference can be attributed to the change itself rather than to outside factors.

Why it matters. A/B tests replace opinions with evidence. They show whether a change caused an improvement, protect against shipping changes that quietly hurt users, and build a record of what works.

Key terms

Term	Meaning
Hypothesis	A testable prediction: "If we change X, metric Y will improve because Z"
Control and variant	The current experience and the changed experience
Primary metric	The one metric that decides the test
Guardrail metrics	Metrics that must not get worse, such as revenue, errors or load time
Minimum detectable effect (MDE)	The smallest change worth detecting; it drives the sample size
Statistical significance	How unlikely the result would be if there were truly no difference; commonly a 95% confidence level (p below 0.05)
Statistical power	The chance of detecting a real effect of at least the MDE; commonly 80%
Sample size	Users needed per group to reach the chosen power and significance

Rule of thumb for sample size. For a conversion-rate metric at 95% confidence and 80% power, users needed per group $\approx 16 \times p \times (1 - p) \div \delta^2$, where p is the baseline rate and δ is the absolute change you want to detect. Example: baseline 10% and a 1-point change (10% to 11%) needs about $16 \times 0.1 \times 0.9 \div 0.0001 \approx 14,400$ users per group. Use a proper calculator for final numbers.

The experiment lifecycle

1. **Hypothesis:** define the change, the expected effect and the reason.
2. **Design:** pick the primary metric, guardrails, MDE, sample size and duration.
3. **Build and QA:** implement the variant and verify that assignment and tracking work.
4. **Run:** launch to the planned traffic; do not stop early because results look good.
5. **Analyze:** check significance, guardrails and key segments.
6. **Decide and document:** ship, iterate or drop, and record the learning.

When not to A/B test. Skip it when traffic is too low to reach significance in a reasonable time, when the change is a legal or bug fix that must ship anyway, or when the decision is strategic and a test cannot capture its long-term effect. Use qualitative research, staged rollouts or before-and-after analysis instead.

Template: A/B Test Plan

1. Test overview

- Test name: *[Short name]*
- Owner: *[Product manager]* · Data partner: *[Analyst]* · Engineering: *[Lead]*
- Planned dates: *[Start and end]*
- Related OKR: *[Key result this test supports]*

2. Hypothesis

If we *[change]*, then *[primary metric]* will *[increase or decrease]* by *[amount]* for *[user segment]*, because *[reason backed by evidence such as research or data]*.

3. Variants

Group	Description	Traffic share
A: Control	<i>[Current experience]</i>	<i>[%]</i>
B: Variant	<i>[Changed experience]</i>	<i>[%]</i>

4. Metrics

Type	Metric	Definition	Baseline	Expected direction
Primary	<i>[Metric]</i>	<i>[Definition]</i>	<i>[Value]</i>	<i>[Up or down]</i>
Secondary	<i>[Metric]</i>	<i>[Definition]</i>	<i>[Value]</i>	<i>[Direction]</i>
Guardrail	<i>[Metric]</i>	<i>[Definition]</i>	<i>[Value]</i>	<i>[Must not worsen beyond X]</i>

5. Design and sample size

- Unit of randomization: *[User, account, session]*
- Target population and exclusions: *[Who is in and out]*
- Minimum detectable effect: *[Absolute or relative change]*
- Significance level and power: *[For example 95% and 80%]*
- Sample size per group: *[Number, and calculator used]*
- Expected duration: *[Days or weeks; include at least one full weekly cycle]*

6. Decision rules (agreed before launch)

Outcome	Decision
Primary metric improves significantly, guardrails hold	[Ship to all users]
No significant difference	[Keep control, or iterate on the idea]
Primary metric worsens or a guardrail breaks	[Stop and roll back]

7. Risks and QA checklist

- ☐ Users are assigned randomly and stay in their group
- ☐ Tracking events fire correctly for both groups
- ☐ Group sizes match the planned split (no sample ratio mismatch)
- ☐ No other test conflicts with this one on the same users
- ☐ Rollback plan is ready

8. Results

Metric	Control	Variant	Difference	Confidence interval	Significant?
[Primary]	[Value]	[Value]	[Change]	[Range]	[Yes or no]
[Guardrail]	[Value]	[Value]	[Change]	[Range]	[Yes or no]

Segment findings: [\[Differences by platform, new versus returning users, region\]](#)

9. Decision and learning

- Decision: [\[Ship, iterate or drop\]](#)
- What we learned: [\[Insight that informs future work\]](#)
- Next test: [\[Follow-up hypothesis\]](#)

Common mistakes

- **Peeking and stopping early.** Checking daily and stopping at the first significant result inflates false positives.
- **Too many primary metrics.** Testing many metrics makes some look significant by chance. Choose one.
- **Ignoring guardrails.** A variant that lifts clicks but hurts revenue or retention is not a win.
- **Underpowered tests.** Too few users means "no difference" may just mean "not enough data".
- **Novelty effects.** New designs can spike engagement briefly. Run long enough to see the settled effect.

5. Final review checklist

Run your metrics and goals documents through these checks before you share them.

- ☐ Every metric has a precise definition, time window and data source.
- ☐ Every target has a baseline, a target value and a date.
- ☐ Key results measure outcomes, not tasks shipped.
- ☐ Each north star, OKR and test has at least one guardrail metric.
- ☐ KPIs include both leading and lagging indicators, each with an owner and alert threshold.
- ☐ A/B test decision rules were written before launch.
- ☐ The layers connect: tests support key results, key results move north-star inputs.
- ☐ Someone outside the team can read it and explain what success looks like in one minute.