A retail company wants to reduce customer churn.

Business and project management acronyms

- KPI (Key Performance Indicator): A quantifiable measure used to track and assess progress toward a specific business objective.
- MVP (Minimum Viable Product): A version of a new product that allows a team to collect the maximum amount of validated learning about customers with the least effort.
- CRM (Customer Relationship Management): The system or technology used to manage all of a company's interactions and relationships with customers and potential customers.
- NFR (Non-Functional Requirement): A quality or performance attribute that a solution must meet, such as scalability, security, or usability. It defines how a system operates, as opposed to what it does.
- AC (Acceptance Criteria): A set of predefined requirements that must be met for a user story to be considered "Done" and shippable.
- DS (Data Scientist): An expert in data modeling and analysis who uses statistical methods and machine learning to extract insights from data.
- PM (Product Manager): The person responsible for a product, from strategic planning to tactical activities, to ensure it meets market needs and business goals.
- Ops (Operations): Short for "operations," this refers to the teams and individuals who ensure the smooth functioning of business processes and systems.
- PII (Personally Identifiable Information): Any information that can be used to identify, contact, or locate an individual, either directly or indirectly.
- GDPR (General Data Protection Regulation): A European data protection law that regulates how companies protect and process the personal data of EU and EEA citizens.

Data science and machine learning acronyms

- RFM (Recency, Frequency, Monetary): A marketing analysis framework for segmenting customers based on how recently they made a purchase, how often they purchase, and how much they spend.
- AOV (Average Order Value): The average amount of money a customer spends per transaction or order.
- NPS (Net Promoter Score): A metric used to measure customer loyalty by asking how likely they are to recommend a product or service to others.
- PR-AUC (Precision-Recall Area Under the Curve): An evaluation metric for binary classification models, particularly useful for imbalanced datasets. A higher score indicates better performance in distinguishing the positive class.
- ROC-AUC (Receiver Operating Characteristic Area Under the Curve): A common evaluation metric for binary classifiers. It represents the model's ability to distinguish between positive and negative classes across all possible classification thresholds.
- GBM (Gradient Boosting Machine): A powerful machine learning algorithm for building predictive models.
- SHAP (SHapley Additive exPlanations): A game theory-based method used to explain the output of any machine learning model by showing how much each feature contributes to the prediction.

- ECE (Expected Calibration Error): A metric used to evaluate how well a model's predicted probabilities align with the actual probabilities. A low ECE means the model is well-calibrated.

1) Business Understanding

Goal: Turn the vague ask into a measurable objective + constraints.

Ask the client

- What is *churn* here? (e.g., **no purchase in 90 days** for active customers)
- Target segment(s)? (loyalty members, e-commerce, in-store, subscription?)
- What interventions are allowed? (discount size caps, outreach channels, contact frequency)
- Business KPI priority: **churn rate ↓**, **net revenue ↑**, **contact cost ↓**, **margin preserved**?
- Ethical/brand limits: no sensitive attribute targeting, opt-out rules, frequency caps.

Decisions & KPIs

- **Primary success metric:** relative churn reduction vs. control (e.g., **−10%** in 8 weeks).
- **Guardrails:** net incremental margin ≥ $0; contact rate ≤ 1 per 14 days; discount budget ≤ $X.

Deliverables

- Problem framing one-pager (definition, scope, KPIs, constraints).
- Intervention playbook menu (email/SMS/push/call/offer sizes).

---

2) Data Understanding

Goal: Inventory data, validate quality, and define the *label*.

Typical sources

- Transactions (customer_id, timestamp, sku, price, channel).
- Customer master (signup date, demographics, loyalty tier).
- Engagement (email opens/clicks, app sessions, web events).
- Service/returns/tickets; store visits; marketing exposure.

Label & cohorting (example)

- Cohort: customers **active in March**.
- Label: `churn = 1` if **no transaction in the next 90 days** (Apr–Jun).
- Feature window: **look back 180 days** (Oct–Mar) to build features.

Quick data checks

- Missing/dupe customer_ids, time-zone issues, outliers (AOV spikes).
- **Leakage hunt:** any post-label info (future returns, future emails) must be excluded.

Deliverables

- Data dictionary, label spec, timelines diagram, quality report.

---

3) Data Preparation

Goal: Build a repeatable feature pipeline.

Common features

- **Recency, Frequency, Monetary** (RFM): days since last purchase; #orders; spend.
- **Product mix** (categories %, subscription vs. one-off).
- **Engagement**: email open rate last 60d, push opt-in, site visits.
- **Price/discount sensitivity**: % orders with promo; elasticity proxy.
- **Service friction**: returns count, ticket count, NPS (if available).
- **Seasonality**: month, holiday flags; store region.

Train/validation split

- Time-based split: train on older cohorts, validate on more recent to mimic reality.

Deliverables

- Feature store table spec; pipeline code skeleton; leakage tests; class-imbalance plan.

---

4) Modeling

Goal: Produce a policy the business can act on—ideally **uplift** not just risk.

Two paths

1. **Churn risk model** (e.g., gradient boosting / logistic regression) → probability of churn → threshold → treat top-N.
2. **Uplift model** (treatment effect) to pick customers who **change behavior if contacted** (best for budget/discount control).

Baselines

- Heuristic: treat customers with **Recency > 60d** and **AOV > $X**.
- Risk model: predict churn@k with PR-AUC / Recall@k; cost-sensitive thresholding.

Fairness & explainability

- Monitor performance by segments (region, tenure, language).
- Provide reason codes/SHAP top features per customer (for compliant messaging).

Deliverables

- Model card (data, features, metrics, caveats), score API/notebook, threshold policy.

---

5) Evaluation

Goal: Prove business value before full rollout.

Offline

- Metrics: PR-AUC, ROC-AUC (secondary), **Recall@K**, **Lift at K**, calibration.
- Sanity: stability across time and segments; backtests on multiple cohorts.

Online (recommended)

- **Holdout A/B test** for 4–8 weeks:
  - Control: business-as-usual (no targeted outreach or generic rule).
  - Treatment: model-selected customers receive intervention.
- **Primary outcome:** churn rate at 90d (or proxy like 30d purchase rate).
- **Secondaries:** incremental revenue, contact cost, margin, complaint rate.

Deliverables

- Test plan (sample size/power), dashboards, decision memo with go/no-go gates.

---

6) Deployment

Goal: Make it run, safe, and measurable.

Flow

1. Weekly job scores eligible customers (excludes recent contacts, opt-outs).
2. Orchestrator applies business rules (frequency cap, budget, channel eligibility).
3. Sends audience + personalized offers to CRM; logs exposures.
4. Observability: volume, response, revenue, cost, churn outcomes; drift alerts.
5. Governance: audit log, suppression lists, data retention, rollback.

NFRs (example, measurable)

- Scoring job **< 30 min**, P95; **uptime ≥ 99%** during window.
- **Cost cap**: ≤ $0.05 per scored customer; discount budget ≤ $50k/month.
- **Privacy**: opt-out honored within 24h; PII encrypted at rest; retention 12 months.
- **Fairness**: disparity ratio in contact rates across key segments ≥ 0.8.

Deliverables

- Runbook, monitoring dashboard, incidents/rollback checklist.

Product north star (for the board header)

Goal: Reduce 90-day churn by **≥10%** vs control without lowering net margin.
MVP (end of Sprint 3): A weekly job that scores customers for churn risk, applies business rules (budget + frequency caps), exports a **target list** + offer tier to CRM, and a live dashboard for outcomes & guardrails.

Epics (7)

1. **E1 -Business Definition & Experiment Design**
2. **E2 -Data Access & Labeling**
3. **E3 -Feature Engineering Pipeline**
4. **E4 -Modeling & Thresholding**
5. **E5 -CRM Integration & Offer Policy**
6. **E6 -Experimentation, Monitoring & Dashboard**
7. **E7 -Privacy, Security & Runbook**

> Tip: Keep “Done” = shippable artifact (doc, code, dashboard) not “analysis completed”.

User stories (with acceptance criteria)

(*Use these; prune to fit capacity. Est. points are just a guide for an individual student team-of-one.*)

E1 -Business Definition & Experiment Design

- **US1.1 (3 pts)** As a **Marketing Director**, I see a one-pager defining churn, target segment, KPIs, and guardrails.
  AC: Churn=90d no-purchase; primary KPI=relative churn ↓; guardrails: margin ≥ $0; contact ≤ 1/14d; budget cap documented.
- **US1.2 (2 pts)** As a **Stakeholder**, I see an A/B test design with sample size and success criteria.
  AC: Power ≥80%, duration estimate, success gate (≥10% churn reduction).

E2 -Data Access & Labeling

- **US2.1 (3 pts)** As a **Data Scientist**, I can query transactions & customers securely.
  AC: Read-only access granted; data dictionary drafted.
- **US2.2 (5 pts)** As a **DS**, I can generate a labeled table for a monthly cohort.
  AC: SQL/Notebook outputs: `customer_id`, `churn_90d`; reproducible; unit

tests catch leakage.

E3 -Feature Engineering Pipeline

- **US3.1 (3 pts)** As a **DS**, I compute RFM for last 180d.
  AC: Features saved to feature table; null/dup checks pass.

- **US3.2 (3 pts)** As a **DS**, I add engagement & discount-sensitivity features.
  AC: Schema versioned; data tests for ranges/outliers.

- **US3.3 (2 pts)** As a **DS**, I create a time-based train/validate split.
  AC: Split diagram + code; leakage tests pass.

E4 -Modeling & Thresholding

- **US4.1 (5 pts)** As a **DS**, I ship a baseline logistic/GBM churn model.
  AC: PR-AUC > heuristic by ≥0.1; calibration ECE ≤0.05; model card written.

- **US4.2 (3 pts)** As a **Marketer**, I get a top-K list with **cost-aware threshold**.
  AC: Threshold policy documented (maximize expected margin under budget).

- **US4.3 (2 pts)** As a **Compliance Officer**, I see performance by segment (fairness slice).
  AC: Report includes at least 3 segments; disparity flagged if <0.8 ratio.

E5 -CRM Integration & Offer Policy

- **US5.1 (3 pts)** As a **Lifecycle Manager**, I receive a **weekly CSV** audience with offer tiers.
  AC: File delivered to CRM sandbox by 6am; schema: id, risk, tier, channel, suppression_flag.

- **US5.2 (3 pts)** As a **Marketer**, I can set frequency caps & suppression rules.
  AC: Config file/param table; logs confirm exclusions applied.

E6 -Experimentation, Monitoring & Dashboard

- **US6.1 (3 pts)** As a **PM**, I see a dashboard with volumes, risk distribution, and calibration.
  AC: Tiles: #scored, top-K count, PR curve snapshot, ECE, segment lift preview.

- **US6.2 (2 pts)** As an **Analyst**, I see a prebuilt A/B report template.
  AC: Computes churn at 90d (or 30d proxy), incremental revenue, contact cost.

- **US6.3 (2 pts)** As an **Ops**, I get drift/failed-job alerts.
  AC: Alert fires on job duration >30m or feature drift >3σ.

E7 -Privacy, Security & Runbook

- **US7.1 (2 pts)** As a **Privacy Officer**, PII handling & retention are documented.
  AC: Data flow diagram; encryption at rest; retention = 12 months.

- **US7.2 (2 pts)** As **On-call**, I have a rollback plan & runbook.
  AC: Defined Sev levels; disable switch; steps tested in dry run.

Definition of Ready (DoR) & Definition of Done (DoD)

DoR: owner named, AC written, data source known, dependencies listed.
DoD: code/tests merged; docs updated; dashboard tile or file visible; stakeholder sign-off captured.

Three-Sprint plan (from nothing → MVP)

Sprint 1 (Weeks 1–2): Define, label, baseline

Goal: Clear definition + labeled data + baseline model + test design.

- Pull in: US1.1, US1.2, US2.1, US2.2, US3.3, US3.1, US4.1 (baseline), US7.1
- Timeboxed spikes (optional): data quality profiling; quick heuristic baseline.
- **Demo:**

    - One-pager (definition & KPIs)
    - Label query + leakage tests
    - Baseline model metrics (PR-AUC, calibration)
- **Exit gates:** Stakeholder signs churn definition; PR-AUC beats heuristic; A/B test plan approved.

Sprint 2 (Weeks 3–4): Features, threshold, policy

Goal: Stronger features; cost-aware threshold; first export; dashboard skeleton.

- Pull in: US3.2, US4.2, US5.1 (sandbox), US5.2, US6.1 (skeleton), US4.3, US6.3
- **Demo:**
    - Feature store v1 (RFM + engagement)
    - Top-K policy maximizing expected margin under budget
    - Weekly CSV to CRM **sandbox** with suppression applied
    - Dashboard with volumes + calibration tile
- **Exit gates:** CSV schema frozen; P95 job time <30m; fairness slices reviewed; alert works.

Sprint 3 (Weeks 5–6): MVP wiring & experiment start

Goal: Reliable weekly scoring job, production export, monitoring, and start of A/B.

- Pull in: US6.2 (A/B template), harden US5.1 for **production**, finish US6.1 tiles, US7.2 runbook.
- **Demo:**
    - Scheduled job (cron/airflow) → production audience CSV by 6am
    - Dashboard: volumes, risk distribution, calibration, segment metrics
    - A/B launch checklist completed; randomization verified

- **Exit gates (MVP):** First production export succeeded; monitoring live; experiment launched with guardrails.