UNIT 6 — DATA-DRIVEN BPR & PROCESS MINING (CONCEPTUAL)

This unit answers **one core question**:

> *How do we stop redesigning processes based on opinion and instead redesign them using evidence?*

6.1 What does "Data-Driven" mean in BPR?

Definition (important)

A **data-driven approach** in BPR means **using actual process data** (time, frequency, cost, variation) to:

- identify inefficiencies,
- locate bottlenecks,
- justify redesign decisions,
- and evaluate improvements objectively.

Data-driven ≠ technology-driven.
Technology is only a **tool**; data is the **basis for decision-making**.

Opinion-driven vs Data-driven redesign (5 differences)

Aspect	Opinion-Driven	Data-Driven
Basis	Intuition, seniority	Evidence, metrics
Bias	High	Low
Justification	Weak	Strong
Risk	High redesign failure	Controlled
Acceptance	Political resistance	Easier buy-in

Example

Opinion-driven:

> “Customers complain, so add more staff.”

Data-driven:

> “80% of delays occur at approval step; adding staff elsewhere won’t help.”

Typical exam question

Q. Why is data-driven analysis essential in BPR?
Because BPR involves radical change with high cost and risk. Data-driven analysis ensures redesign decisions are based on actual performance evidence rather than assumptions, reducing failure and resistance.

6.2 What kind of “data” do we use?

Types of process data (must know)

1. **Time data**
    - Start time
    - End time
    - Waiting time
    - Cycle time
2. **Frequency data**
    - How often steps occur
    - How often rework happens
    - How often approvals repeat
3. **Cost data**
    - Labor cost
    - Cost per transaction
    - Cost of delay
4. **Variation data**
    - Differences across cases
    - Standard deviation
    - Outliers
5. **Path data**
    - Which steps occur

- In what sequence
- Which steps are skipped or repeated

Why this matters

Processes don't fail the same way every time.
Variation across cases is often the biggest problem.

6.3 Introduction to Process Mining (Conceptual)

Definition

Process mining is a technique that uses **event data** from information systems to **discover, analyze, and improve real process flows**.

It answers:

- What actually happens?
- How long does it take?
- Where do cases diverge?
- Where are delays concentrated?

What is an "event"?

An **event** is a recorded activity with:

- Case ID (process instance)
- Activity name
- Timestamp
- Resource (optional)

Example (procurement):

- Create Purchase Requisition
- Analyze Requisition
- Request Quotation
- Approve Quotation

Each execution = **one case**.

Process Mining vs Process Mapping (6 differences)

Aspect	Process Mapping	Process Mining
Source	Interviews, workshops	System logs
Accuracy	Assumed	Actual
Bias	Human bias	Low
Variation visibility	Limited	Strong
Scalability	Small	Large datasets
Update	Manual	Continuous

Key exam insight

Process mining does **not replace** mapping.
It **validates and challenges** it.

6.4 Cases & Variants (VERY IMPORTANT)

Case

A **case** is one complete execution of a process from start to end.

Example:

- One purchase request
- One tea order
- One student registration

Variant

A **variant** is a **unique path** a case follows.

Two cases can:

- Use the same steps (same variant)
- Or follow different paths (different variants)

Why variants matter

- High number of variants = low standardization
- Rare variants often contain:
    - Exceptions
    - Rework
    - Special approvals
    - Errors

Example (from your class)

- Case A: 6 events → simple, fast
- Case B: 18 events → complex, slow

Interpretation:
The 18-event case is not “better”; it is **more complicated and more delayed**.

Variant frequency insight

- **Most common variant** = intended process
- **Rare variants** = exceptions, breakdowns, policy overload

---

Typical exam question

Q. Why are rare variants important in process mining?
Because rare variants often contain rework, extra approvals, or exception handling that significantly increase cycle time and cost, making them key targets for redesign.

6.5 Bottlenecks through Process Mining

Traditional bottleneck detection

- Observation
- Interviews
- Manual timing

Process-mining-based detection

- Waiting time spikes
- Long idle periods
- Repeated activities
- Queue buildup at same step

Types of bottlenecks revealed

1. **Approval bottlenecks** – long idle time
2. **Information bottlenecks** – missing data pauses
3. **Rework bottlenecks** – repeated steps
4. **Resource bottlenecks** – limited staff availability

Example interpretation

If:

- Step duration = 5 minutes
- Waiting before step = 2 days

→ The problem is **not execution**, it is **handoff / approval logic**.

Exam framing

Do NOT say:

> “DISCO shows red bars.”

Say:

> “Event data shows prolonged waiting time before approval steps, indicating approval-based bottlenecks.”

6.6 Waiting Time vs Active Time (CRITICAL)

Active time

Time when work is actually being done.

Waiting time

Time when nothing happens:

- waiting for approval
- waiting for data
- waiting for availability

Key insight

In many processes:

- Active time < 20%
- Waiting time > 80%

This is **where BPR targets**.

Example

A procurement process takes 13 days:

- Active work: 2 days
- Waiting: 11 days

Redesign focus = **waiting**, not working faster.

Exam question

Q. Why does BPR focus more on waiting time than active time?
Because waiting time contributes the majority of lead time without adding value. Reducing waiting through structural redesign yields far greater improvements than speeding up already efficient tasks.

6.7 Variability & Process Stability

High variability indicates:

- Lack of standardization
- Too many exceptions
- Policy overload
- Role ambiguity

Stable vs unstable processes

Aspect	Stable	Unstable
Variants	Few	Many
Cycle time	Predictable	Wide range
Quality	Consistent	Inconsistent
Control	Rule-based	Person-based

Link to Six Sigma

- Process mining reveals **where variation occurs**
- Six Sigma explains **why variation occurs**
- BPR decides **whether the process should exist in this form**

6.8 Data-Driven Decision Making (CEO Perspective)

Your class emphasized this strongly.

Typical top-down logic

1. CEO sets target (e.g., 5% cost reduction)
2. Departments list their processes

3. Each process evaluated using data:
    - Cost
    - Time
    - Complexity
    - Benefit

Cost–Benefit / Cost–Complexity Matrix

Four zones (conceptual):

1. **Low cost, high benefit** → priority
2. **High cost, high benefit** → strategic decision
3. **Low cost, low benefit** → minor fixes
4. **High cost, low benefit** → avoid

Key lesson

> Not every process improvement is worth doing.

Big ideas without measurable benefit are **waste**, not innovation.

6.9 Technology ≠ Data-Driven (VERY IMPORTANT)

Your teacher explicitly warned:

> “Don’t obliterate driven by technology.”

Technology-driven mistake

- Add software because it exists
- Automate without understanding process

Data-driven approach

- Understand the problem
- Measure it
- Then decide whether technology is needed

Example

Wrong:

> “Add AI chatbot.”

Right:

> “Data shows 70% queries are repetitive → automation justified.”

UNIT 6 — EXAM QUESTION BANK

1. What does data-driven BPR mean?
2. Differentiate opinion-driven and data-driven redesign.
3. What is process mining and what problem does it solve?
4. Define case and variant with examples.
5. Why are rare variants important?
6. Explain waiting time vs active time.
7. How does process mining help identify bottlenecks?
8. Why is technology-driven redesign dangerous?
9. How does process mining complement Six Sigma?
10. Explain data-driven decision making from a CEO perspective.