

Data Dictionaries

Datasets for Chapters and Capstone Project (with the NorthStar Golden Thread Design)

1. Purpose

This document provides structured descriptions of the datasets used throughout the book and the capstone project.

It is designed to help you:

- understand what each variable represents
 - interpret data before modeling
 - connect data to **decision context and forecasting design**
-

Core Principle

Data are not just inputs—they are representations of how organizations operate, respond, and decide.

2. The Golden Thread: Why One Dataset Matters

Throughout this book, most chapters use a shared dataset:

NorthStar Retail Group — Everyday Essentials™ Weekly Sales

This is not accidental. It is a deliberate instructional design choice known as the **Golden Thread**.

What Is the Golden Thread?

The Golden Thread means:

- the **same business context**
- the **same core dataset**
- evolving across chapters with increasing complexity

Why This Matters

Without a Golden Thread, learning forecasting often feels like:

- disconnected techniques
- unrelated datasets
- repeated “starting from scratch”

With the Golden Thread, you instead experience:

1. Continuity of Thinking

You are not relearning new data each chapter.

You are **deepening your understanding of the same system.**

2. Layered Learning

The dataset evolves as your skills evolve:

- early chapters → simple signals
 - middle chapters → structure and diagnostics
 - later chapters → decision systems
-

3. Realistic Practice

In real organizations:

- data do not reset between problems
- understanding accumulates over time

The Golden Thread mirrors this reality.

4. Stronger Decision Context

Because the dataset remains consistent:

- decisions feel connected
- trade-offs become visible
- learning becomes **system-level, not method-level**

Design Insight

The goal is not to learn many datasets.

The goal is to learn how one system can be understood, modeled, and governed over time.

3. Dataset Overview by Chapter

Dataset	Chapters	Role in Learning Progression
essentials_sales_lite.csv	Chapters 1–5	Seeing signal and structure (Golden Thread – simplified view)
essentials_sales_residuals.csv	Chapter 6	Diagnosing behavior and trust (Golden Thread – diagnostic layer)
essentials_sales.csv	Chapters 7–8	Decision-aware forecasting (Golden Thread – full system)
healthcare_capacity_weekly.xls ^x	Capstone Project	New domain, full decision system under uncertainty

4. Dataset 1 — essentials_sales_lite.csv

Used in Chapters 1, 2, 3, 4, 5

Golden Thread — Foundational Layer

Purpose

This dataset introduces **time-based thinking** using a simplified structure.

It supports:

- signal vs. noise distinction
 - smoothing
 - decomposition
 - early understanding of temporal structure
-

Variables

Variable Name	Type	Description	Example	Decision Meaning
week	Date / string	Weekly time period	2023-W01	Time anchor
week_index	Integer	Sequential index	1, 2, 3	Ordering of observations
sales	Numeric	Weekly unit sales	12,540	Demand signal

Interpretation Guidance

- Complexity is intentionally removed
 - All insight must come from **time and sales alone**
-

Golden Thread Role

This is your **first view of the system**—simple, but foundational.

Learning Insight

Before modeling, you must learn to see.

5. Dataset 2 — essentials_sales_residuals.csv

Used in Chapter 6

Golden Thread — Diagnostic Layer

Purpose

This dataset introduces **forecast diagnostics and validation**.

It supports understanding:

- forecast errors
 - residual behavior over time
 - early warning signals of model failure
-

Variables

Variable Name	Type	Description	Example	Decision Meaning
week	Date	Time index	2023-W10	Tracking residual over time
actual_sales	Numeric	Observed sales	12,300	Ground truth
forecast_sales	Numeric	Model prediction	12,100	Expected value
residual	Numeric	Actual – Forecast	+200	Forecast error
abs_error	Numeric	Absolute error	200	Error magnitude
squared_error	Numeric	Squared error	40,000	Penalized error
rolling_residual_mean	Numeric	Smoothed residual	150	Drift detection

Interpretation Guidance

Residuals reveal:

- where the model struggles
- how behavior changes over time

Golden Thread Role

This is where the system becomes **self-aware**.

Learning Insight

Forecasting is not complete until you understand when it fails.

6. Dataset 3 — essentials_sales.csv

Used in Chapters 7 and 8

Golden Thread — Decision Layer

Purpose

This dataset introduces **decision-aware forecasting** by adding operational and contextual variables.

It supports:

- feature-based forecasting
 - AI-assisted reasoning
 - system-level design
-

Core Variables

Variable Name	Type	Description	Example	Decision Link
week	Date	Time index	2023-W15	Temporal anchor
sales	Numeric	Weekly sales	13,200	Target variable

Operational Variables

Variable Name	Type	Description	Example	Decision Link
inventory	Numeric	1-sales/3500	0.2	Supply constraints

Contextual Variables

Variable Name	Type	Description	Example	Decision Link
holiday_flag	Binary	Holiday week	1	Seasonal demand
promotion_flag	Binary	Promotion active	1	Marketing actions

Interpretation Guidance

- Some variables are **controllable**
 - Others reflect **external uncertainty**
-

Golden Thread Role

This is the **full system view**, where forecasting meets decisions.

Learning Insight

Forecasts become meaningful when they reflect how decisions shape outcomes.

7. Dataset 4 — healthcare_capacity_weekly.xlsx

Used in Capstone Project
Capstone Transition — New Domain

Purpose

This dataset introduces a **new domain** where forecasting must guide high-stakes decisions.

It represents:

- healthcare demand
 - capacity constraints
 - operational risk
-

Variables

Variable Name	Type	Description	Example	Decision Link
week	Date	Weekly time period	2022-W40	Time anchor
patient_demand	Numeric	Weekly patient volume	1,250	Demand planning
bed_capacity	Numeric	Available beds	1,100	Capacity limits
staffing_level	Numeric	Available staff	320	Resource planning
utilization_rate	Numeric (%)	Capacity usage	0.92	System stress
emergency_flag	Binary	Surge condition	1	Crisis response
policy_change_flag	Binary	Policy shift	1	Structural change

Interpretation Guidance

- Capacity constraints directly affect outcomes
 - Forecasting must support **preparedness, not just prediction**
-

Capstone Role

You must transfer your learning from the Golden Thread into a **new, unfamiliar system**.

Learning Insight

Mastery is demonstrated when you can apply design thinking beyond the original dataset.

8. Cross-Dataset Learning Progression

Stage	Dataset	Learning Focus
Early	Lite dataset	Seeing patterns
Middle	Residual dataset	Understanding behavior
Late	Full dataset	Designing systems
Capstone	Healthcare dataset	Making decisions under uncertainty

Design Insight

The Golden Thread ensures that learning progresses as:

see → model → diagnose → design → decide

9. Data Understanding Checklist

Before modeling, always ask:

1. What does each variable represent in reality?
 2. Which variables are controllable vs. external?
 3. Are there distortions (e.g., stockouts)?
 4. What patterns exist over time?
 5. What decision does this data support?
-

10. Final Reflection

Across all datasets, one principle remains:

Forecasting is not about the data you have—it is about how you interpret data to support decisions.
