# CAP6768 Data Analytics - Dataset Documentation

---

## Dataset A: Retail Store Daily Sales

### Business Context

A local retail store wants to understand their sales patterns over the summer (3 months). They have clean, simple data and want to predict future sales.

### Dataset Info

- **File**: `data_analytics_retail.csv`
- **Records**: 90 days (June-August 2025)

### Variables

| Variable | Description | Example Values |
|---|---|---|
| date | Date of sales | 2025-06-01 |
| day_of_week | Day name | Monday, Tuesday... |
| weekend | Is it weekend? | TRUE/FALSE |
| week_number | Week of year | 23, 24, 25... |
| month | Month name | Jun, Jul, Aug |
| daily_customers | Number of customers | 50-200 |
| avg_transaction | Average sale amount | $25-45 |
| daily_revenue | Total revenue | $2,000-8,000 |
| temperature | Temperature in F | 75-95 |
| promotion | Was there a promotion? | TRUE/FALSE |

### Some Hints

- Are there days with more customers?
- Is there an important association between temperature and number of customers?
- Do promotions matter?

---

# Dataset B: Student Performance

## Business Context

A college wants to identify which students might fail, so they can offer extra help early.

## Dataset Info

- **File**: `data_analytics_students.csv`
- **Records**: 200 students
- **Difficulty**: BEGINNER FRIENDLY

## Variables

| Variable | Description | Example Values |
|---|---|---|
| student_id | Student identifier | 1, 2, 3... |
| age | Student age | 18-25 |
| study_hours | Hours studied per week | 0-25 |
| attendance_pct | Class attendance % | 20-100 |
| previous_grade | Last semester grade | 40-100 |
| works_parttime | Has part-time job? | Yes/No |
| final_grade | Final grade (0-100) | 30-100 |
| passed | Did they pass? | Pass/Fail |

## Some Hints

- What are the important variables to predict grades?
- Pass/Fail cutoff is 60
- Be careful with missing values

---