



Figure 1: AI-Assisted Power BI for Business Analytics

## Lab 4: Data Foundations for AI Success

### Module 4: Data Foundations for AI Success

**Duration:** 35 minutes **Dataset:** CloudRevenue (Microsoft Cloud Subscription Analytics) **Deliverable:** Before/after comparison document

#### Learning Objectives

By the end of this lab, you will be able to:

- Understand how data structure affects Copilot accuracy
- Add descriptions and synonyms to semantic models
- Test the impact of metadata on natural language queries
- Apply naming conventions that improve AI responses

#### Before You Begin

##### Required Reading

1. **CloudRevenue BRD** - View BRD
  - Review the business questions that users will ask
  - Think about how users might phrase these questions naturally
2. **CloudRevenue Data Dictionary** - View Data Dictionary
  - Note which columns have clear vs. cryptic names
  - Consider synonyms business users might use

#### Concept Review: Why Metadata Matters

Copilot interprets natural language by mapping user questions to data columns. Consider:

| User Asks         | Column Name              | Can Copilot Map? |
|-------------------|--------------------------|------------------|
| “Show me revenue” | Total_Revenue            | Yes              |
| “Show me sales”   | Total_Revenue            | Maybe not        |
| “Show me ARR”     | Annual_Recurring_Revenue | Probably not     |

**Solution:** Add synonyms so Copilot knows “sales” and “ARR” mean Total\_Revenue.

```

%%{init: {'theme':'base', 'themeVariables': {'primaryColor': '#e8f4fd', 'secondaryColor': '#e8f6ea', 'tertiaryColor': '#e8f6ea'}}}
flowchart TB
  subgraph Before[" Without Metadata"]
    A1[" 'Show me sales'"] --> B1{Copilot}
    B1 --> C1[" No match found"]
    D1[" TRev_USD"] -.->| "?" | B1
  end
end

```

```

subgraph After[" With Metadata"]
  A2[" 'Show me sales'"] --> B2{Copilot}
  B2 --> C2[" Revenue: $12.4M"]
  D2[" Total_Revenue<br/>synonyms: sales, ARR, bookings"] -->|"Match!"| B2
end

style Before fill:#fef5f5,stroke:#f0d0d0
style After fill:#edf6ee,stroke:#c2dbc8

```

*Metadata transforms cryptic column names into discoverable business terms.*

## Lab Instructions

### Part 1: Baseline Testing - Before Metadata (10 minutes)

First, let's see how Copilot performs WITHOUT enhanced metadata.

#### Step 1: Access the Semantic Model

1. Navigate to app.powerbi.com
2. Open the course workspace
3. Find the semantic model named `CloudRevenue_Lab4_Raw`
4. Create a quick report: Click **...** → **Create report** → **Start from scratch**

**Step 2: Test Natural Language Queries** Open Copilot and try these queries. Record whether Copilot understands:

| Query                          | Did Copilot Understand? | Notes |
|--------------------------------|-------------------------|-------|
| “Show me revenue”              | Yes / No / Partial      |       |
| “Show me sales”                | Yes / No / Partial      |       |
| “Show me ARR”                  | Yes / No / Partial      |       |
| “What’s our MRR trend?”        | Yes / No / Partial      |       |
| “Show customers by segment”    | Yes / No / Partial      |       |
| “Which products are churning?” | Yes / No / Partial      |       |
| “Show enterprise accounts”     | Yes / No / Partial      |       |
| “Revenue by geo”               | Yes / No / Partial      |       |

**Screenshot #1:** Capture a query where Copilot struggled

**Step 3: Document Baseline** Count your results: - Fully understood: \_\_\_\_ / 8 - Partially understood: \_\_\_\_ / 8 - Not understood: \_\_\_\_ / 8

### Part 2: Enhance the Semantic Model (12 minutes)

Now you'll add metadata to improve Copilot's understanding.

#### Step 1: Open Model View

1. In Power BI Service, find `CloudRevenue_Lab4_Editable` (a copy you can edit)
2. Click to open it
3. Click **Open data model** or access via **Model view**

**Note:** Your instructor may provide a Desktop file if Service editing is limited

**Step 2: Add Table Descriptions** For each table, add a description that explains its purpose:

1. Click on the **Subscriptions** table
2. In the Properties pane, find **Description**
3. Add:

Fact table containing all cloud subscription transactions. Each row represents one subscription with its revenue, dates, and customer information. Use for revenue analysis, churn tracking, and customer segmentation.

4. Repeat for **Products** table:

Dimension table of Microsoft cloud products including M365, Azure, and Dynamics 365 offerings. Contains product categories, pricing tiers, and SKU information.

5. Repeat for **Regions** table:

Geographic dimension with sales regions, countries, and market classifications. Use for regional performance analysis and market expansion tracking.

**Step 3: Add Column Descriptions** Add descriptions to key columns in Subscriptions:

| Column                 | Description to Add   |
|------------------------|--|
| Total_Revenue          | Total revenue in USD from this subscription. Also known as sales, booking, or TCV (total contract value).  |
| Monthly_Revenue        | Monthly recurring revenue (MRR) from this subscription. Key SaaS metric for tracking growth.               |
| Customer_Segment       | Business segment classification: Enterprise (>1000 employees), SMB (small-medium business), or Consumer.   |
| Subscription_Status    | Current status: Active, Churned, or Pending. Use for retention and churn analysis.                         |
| Contract_Length_Months | Duration of subscription commitment in months. Longer contracts typically indicate stronger relationships. |

**Step 4: Add Synonyms** This is the most impactful enhancement. Add synonyms so Copilot recognizes alternative terms:

1. Click on **Total\_Revenue** column
2. Find **Synonyms** in Properties
3. Add: sales, revenue, bookings, TCV, total contract value, ARR
4. Add synonyms to other columns:

| Column              | Synonyms to Add                                     |
|---------------------|---|
| Monthly_Revenue     | MRR, monthly recurring, subscription revenue        |
| Customer_Segment    | segment, customer type, account type, business size |
| Region_Name         | geography, geo, territory, market, location         |
| Product_Category    | product type, offering, solution, SKU type          |
| Subscription_Status | status, state, active, churned                      |

### Step 5: Save Changes

1. Click **Save** to save the semantic model
2. Wait for changes to publish (may take 1-2 minutes)

### Part 3: After Testing - Measure Improvement (8 minutes)

Now test the same queries against your enhanced model.

#### Step 1: Create Report from Enhanced Model

1. Find `CloudRevenue_Lab4_Editable` in the workspace
2. Create a new report from it

#### Step 2: Re-Test the Same Queries

| Query                          | Before | After | Improved? |
|--------------------------------|--------|-------|-----------|
| "Show me revenue"              |        |       |           |
| "Show me sales"                |        |       |           |
| "Show me ARR"                  |        |       |           |
| "What's our MRR trend?"        |        |       |           |
| "Show customers by segment"    |        |       |           |
| "Which products are churning?" |        |       |           |
| "Show enterprise accounts"     |        |       |           |
| "Revenue by geo"               |        |       |           |

**Screenshot #2:** Capture the same query that struggled before (now working)

#### Step 3: Calculate Improvement

- Before: \_\_\_\_ / 8 fully understood
- After: \_\_\_\_ / 8 fully understood
- **Improvement:** \_\_\_\_ more queries understood

### Part 4: Advanced Enhancements (5 minutes)

If time permits, explore these advanced options:

**AI Instructions (Preview Feature)** Some semantic models support "AI Instructions" - natural language guidance for Copilot:

When users ask about "performance", they typically mean revenue growth.

When users ask about "health", they mean churn rate and retention.

"Top customers" means customers with highest total revenue.

"At-risk" customers means those with declining usage or near renewal.

#### Q&A Training

1. In Power BI Service, open your semantic model
2. Look for **Q&A setup** or **Teach Q&A**
3. Add specific question-answer pairs:
  - Q: "Who are our best customers?"
  - A: Show top 10 customers by `Total_Revenue`

**Linguistic Schema** For advanced users, you can export and edit the linguistic schema to define: - Entity relationships - Phrasings (how to talk about the data) - Complex synonyms

## Lab Deliverable

**Submit:** A before/after comparison document (1-2 pages)

### Document Structure

- 1. Baseline Assessment** - Screenshot of Copilot struggling with a query (before) - Summary of baseline test results (\_\_\_/8 understood) - List of queries that failed
- 2. Enhancements Made** - Table descriptions added (summarize) - Column descriptions added (list key ones) - Synonyms added (list the most impactful)
- 3. After Assessment** - Screenshot of same query succeeding (after) - Summary of after test results (\_\_\_/8 understood) - Improvement calculation
- 4. Key Learnings** - Which enhancement had the biggest impact? (descriptions vs. synonyms) - What surprised you? - How would you apply this to real business data?

### Submission Format

- **File Name:** Lab4\_BeforeAfter\_[YourName].docx or .pdf
- **Length:** 1-2 pages
- **Include:** At least 2 screenshots (before and after)
- **Submit via:** Canvas → Assignments → Lab 4

## Completion Checklist

Before submitting, confirm you have:

- Completed baseline testing (8 queries)
- Added descriptions to all 3 tables
- Added descriptions to at least 5 key columns
- Added synonyms to at least 5 columns
- Saved and published the enhanced model
- Re-tested all 8 queries
- Captured before/after screenshots
- Calculated improvement percentage
- Submitted comparison document via Canvas

## Key Takeaways

- 1. Metadata is the secret to Copilot accuracy:** Descriptions and synonyms dramatically improve natural language understanding
- 2. Think like your users:** Add synonyms for how business users actually talk about data
- 3. Descriptions help AI understand context:** A well-described column is easier for Copilot to map
- 4. Small investment, big payoff:** 30 minutes of metadata work saves hours of user frustration
- 5. Data preparation is “AI preparation”:** Clean, well-documented data is the foundation of AI success

## Metadata Quick Reference

Good Column Names

| Bad      | Good               |
|----------|--------------------|
| coll     | Total_Revenue      |
| dt       | Order_Date         |
| amt      | Transaction_Amount |
| cust_seg | Customer_Segment   |
| prod_cat | Product_Category   |

### Essential Synonyms by Domain

**Sales/Revenue:** - Revenue → sales, bookings, income, earnings - MRR → monthly recurring, subscription revenue - ARR → annual recurring, yearly revenue

**Customers:** - Customer → client, account, buyer - Segment → type, tier, classification - Enterprise → large, corporate, major

**Products:** - Product → offering, solution, SKU - Category → type, family, group

**Geography:** - Region → territory, geo, market, area - Country → nation, market

### Preparation for Next Lab

- Complete Prompting Journal entry #4
- Review: Storytelling with Data Blog
- Review the M365Marketing BRD and Data Dictionary
- Think about: What would an executive want to see at a glance?

*“AI is only as good as the data it understands. Your metadata is the translator.”*