DES MOINES UNIVERSITY

DMU Data Dojo

Community of Practice

Meeting 003 – Technologies & Terminologies of Data Analysis

BASIC DATA ANALYSIS TERMINOLOGY

"Precise language is not the problem.
Clear language is the problem."

— Richard P. Feynman

Basic Data Analysis Terminology

Storage & Preparation

- Data Steward
- Data Source
- Data Warehouse
- Dataflow
- Dataset
- Data Model

Analysis & Visualization

- On-Prem & Cloud Servers
- Paginated Reports
- Interactive Reports
- Dashboards
- KPIs, OKRs, etc.
- Alerts & Subscriptions
- Workspaces & Apps

DATA STORAGE AND PREPARATION

"Data should be transformed as far upstream as possible,
and as far downstream as necessary."

— Matthew Roche

Storage & Preparation

Data Steward: Ownership, Oversight, & Governance

- A "Data Steward" *is*:
  - Each ***Business Unit (BU)***'s data ***Subject-Matter Expert (SME)***
  - Detail-oriented, meticulous, and conscientious
  - Creator, maintainer, and enforcer of their BU's data standards
  - First responder whenever "something is wrong with the data"
  - A critical component in every healthy data culture
  - A role/responsibility, not necessarily a dedicated job title
- A "Data Steward" *is not/doesn't need to be*:
  - An "Excel Wizard," "tech-savvy," or even "good w/ computers"
  - An executive, team leader, or business decision-maker

Storage & Preparation

Data Source: Where the Data Came From

- The meaning of “Data Source” is context-dependent
  - In a Power BI report, the “Data Source” is usually the Data Warehouse, a Dataflow, Dataset, SharePoint List, Excel File, etc.
  - In the Data Warehouse, the “Data Source” refers to the system or application where the data originated
- The complete path that our data travels, from its original source system, through any transformations and intermediate storage, all the way to where it is displayed in user-facing reports, is called “Data Lineage.”

Storage & Preparation

Data Warehouse (DW): Centralized Data Storage

- Extract, Transform, and Load (ETL)
  - Extract: Download the data from its original source
  - Transform: Re-organize and optimize the data for analysis purposes
  - Load: Import the data into the Data Warehouse
- Enforce rules on incoming data
  - Data Type (whole number, decimal, text, date, etc.)
  - “Nullable” or “Not Null” (allow blank values or not?)
- Create relationships between tables
  - Dimensions: Student, Program, College, Department, Date, etc.
  - Facts: Application, Enrollment, Attendance, Exam, Rotation, etc.
- Run custom routines
  - Scheduled
  - Conditional/event-driven
  - On-demand

Storage & Preparation

Dataflow: Pre-Loaded Tables of Data in the Cloud

- Specific to Power BI
  - Power BI Desktop: ✗
  - Power BI Cloud Service: ☑
- Like a simplified, cloud-based Data Warehouse
- Strengths:
  - Automatically refreshes with the latest data every day
  - Cloud-based (a faster, more direct route to the data for cloud-based Power BI reports than pulling it from a server on campus)
- Limitations:
  - No relationships between tables
  - Limited custom routine capabilities
  - Cloud-based (inaccessible from on-premises report servers)

Storage & Preparation

Dataset: Specific Data for a Specific Purpose

- "Dataset" is mostly synonymous with "Data Model"
  - Small semantic differences; none worth worrying about
  - Most data professionals use the terms interchangeably
- These terms predate most modern data systems
  - Lots of data tools & platforms out there
  - Each has its own definition of "dataset" and "data model"
- Most Datasets / Data Models consist of:
  - One or more tables of data
  - Relationships between tables (fields w/ matching values)
  - Calculated fields
  - Aggregated (sum, avg, min/max, etc.) values

Storage & Preparation

Dataset/Data Model Example 1

This is an example of a relatively well-made dataset. Simple, straightforward, and easy to use.

Storage & Preparation

Dataset/Data Model Example 2

This is a "legacy" dataset that I inherited from my predecessor. I've been gradually cleaning up and optimizing it as time allows. Yes, it's every bit as confusing as it looks! But don't worry, that's my problem, not yours. 😅

Storage & Preparation

On-Premises & Cloud Report Servers

- On-Premises (or “on-prem” for short)
  - ServerName.dmu.edu
- Cloud
  - PowerBI.com
- Virtual Private Network (VPN)
  - *Required* for working with on-prem reports/servers
  - *Not required* for working with cloud reports/servers

DATA ANALYSIS AND VISUALIZATION

"There is a story in your data. But your tools don't know what that story is. That's where it takes ***you***—the analyst or communicator of the information—to bring that story visually and contextually to life."

— Cole Nussbaumer Knaflic

Analysis & Visualization

Report Type #1: “Static” or “Paginated” Reports

- Paginated Reports can be either:
  - SQL Server Reporting Services (SSRS) reports (on-prem only)
  - Power BI Paginated Reports (on-prem and cloud)
- Data refreshes in real-time when the user “runs” the report
- Content is static (non-interactive) after the report is rendered
- Recommended for reports and “dynamic documents” that:
  - Must contain tables and text populated directly from the data
  - Will usually be exported as a static file (Excel, PDF, etc.)
  - Must let the user filter/manipulate the data (or the query) at runtime by specifying/selecting parameter values
  - Will primarily be used for ad-hoc analysis, dynamic document generation, regulatory compliance, etc., on an infrequent basis

Analysis & Visualization

Report Type #2: “Interactive” or “Canvas” Reports

- Exclusive to Power BI (cloud and on-prem)
- Data is refreshed by an automated process or schedule
- Content is always fully interactive
- Recommended for reports and dashboards that:
  - Must contain interactive data visualizations, KPIs, etc.
  - Will mostly be used interactively, *not* printed or exported
  - Must let the user slice, filter, cross-highlight, & drill-down/thru
  - Will primarily be used to explore, inspect, and discover vital insights for data-driven decision-making on a regular/daily basis

Analysis & Visualization

KPIs, OKRs, etc. (often called “Metrics”)

- Key Performance Indicators (KPIs): Ongoing Metrics
  - Student admission/enrollment as % chg. year-over-year
  - Student retention/graduation rate as % chg. term-over-term
  - Student satisfaction as star rating (range 1-5)
- Objectives & Key Results (OKRs): Specific Goals
  - Improve student board exam scores by X%
  - Improve rotation preceptor evaluation scores by Y%
  - Increase enrollment from marginalized communities by Z%
- Most useful on dashboards and report summary pages for deans, department heads, executive leadership, etc.

Analysis & Visualization

Dashboards: *"Critical metrics in 5 seconds or less!"*

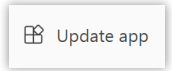
- Specific to Power BI
  - Power BI Desktop: ✗
  - Power BI Cloud Service: ☑
- Composed *only* of high-level visuals *from existing reports*
- Summary/KPIs *only*, just like the dashboard in your car
- No tables, tooltips, filters, slicers, drill-downs, etc.
- Simple, intuitive, and non-interactive by design
- ***For deeper, more detailed analysis, use a report instead***

Alerts & Subscriptions

- Alerts: Event-based delivery
  - Dashboards (Power BI Cloud Service only)
  - Scorecards (Power BI Cloud Service only)
  - "Whenever the value of ***X*** exceeds ***Y***, send an alert to ***Z***"
- Subscriptions: Scheduled delivery
  - Paginated Reports (SSRS and Power BI)
  - Interactive Reports (Power BI only)
  - "Email this report as a PDF to ***X*** every ***Y*** day at ***Z*** o'clock"

Analysis & Visualization

Power BI Workspaces & Apps (cloud only):

- Workspaces (for Report Authors)
  - Shared spaces for publishing Power BI reports, datasets, etc.
  - One for every department, college, program, office, team, etc.
  - Created upon request by ITS
  - Workspace Roles:
    - Admin, Member, Contributor, Viewer
    - Assigned to Active Directory Security Groups (and their members) by ITS
- Apps (for Report Users)
  - The user-facing “front-end” for each Workspace in Power BI
  - Reports must be published to the Workspace first, then to the App by clicking the “Update app” button, before users will see them 

  - This allows report authors to verify that their latest published report looks and behaves as it should before they make it available to users

Analysis & Visualization

Power BI Workspace & App Example

Questions?