Data Analytics

Lesson 05.

Data Modeling

Dr. Hai Tran

hai.tran@sbsuni.edu.vn

Scholar: https://scholar.google.com/citations?user=kHZvlTkAAAAJ&hl=en&oi=ao

Co-Founder: XAI - https://xai.foo/

Saigon Business School

In partnership with

UNIVERSITY CANADA WEST

MACQUARIE University SYDNEY·AUSTRALIA

University of Northampton

AMITY GLOBAL INSTITUTE

Saigon Business School

Learning materials

- Textbook
  - Evans, J. (2016) Business Analytics. 2nd edn. Pearson.
  - Runkler, T. (2016) Data Analytics: Models and Algorithms for Intelligent Data Analysis. 2nd edn. Vieweg+Teubner Verlag.

- Online reference materials
  - archive.ics.uci.edu/ml/
  - powerbi.microsoft.com
  - https://github.com/topics/data-analysis-python
  - https://media.pearsoncmg.com/ph/esm/esm_evans_eba3e_20/tools/eba3e_analytic_solver.html
  - https://data.imf.org/

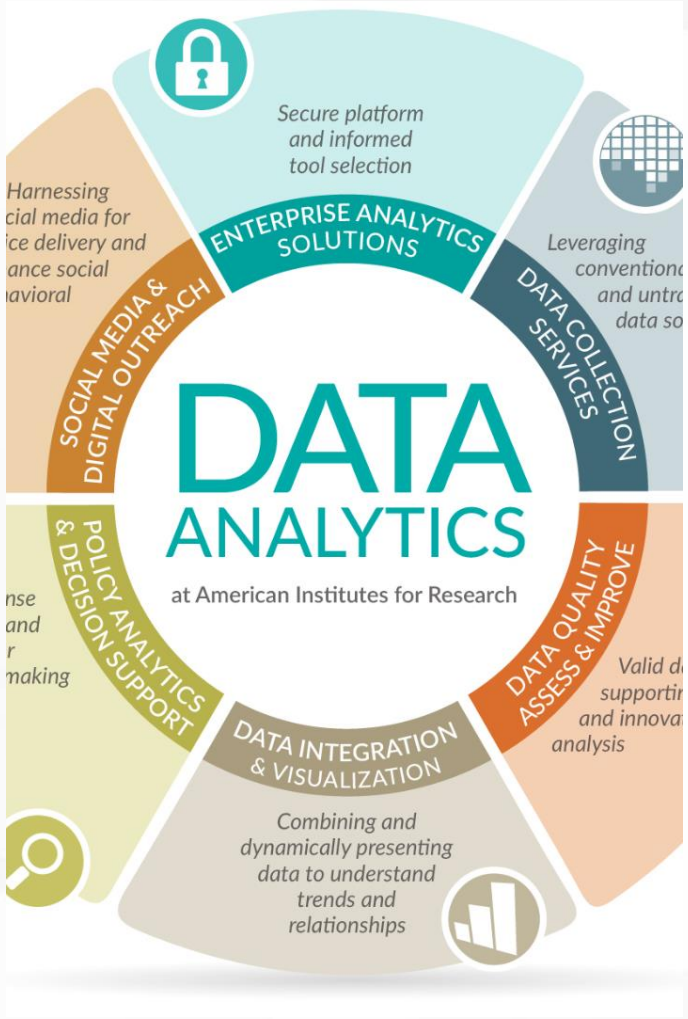
Saigon Business School

Agenda

- Lesson 1: Understanding Data Analytics Terminologies.
- Lesson 2: Foundation of Business Analytics
- Lesson 3: Visualizing and Exploring data
- Lesson 4: Applying Descriptive Analytic Techniques
- Lesson 5: Data Modeling
- Lesson 6: Predictive Analytics
- Lesson 7: Regression, Classification and Clustering
- Lesson 8: Forecasting Techniques
- Lesson 9: Investigating Predictive Analytic Techniques
- Lesson 10: Introduction to Data Mining
- Lesson 11: Demonstrating Prescriptive Analytic Methods
- Lesson 12: Recap and advanced topics

Saigon Business School

Data Modeling



Data modeling is a crucial aspect of data analytics, enabling improved decision making, effective data analysis, and enhanced data governance.

Why Data Modeling Matters

1 Improved Decision Making

By structuring data and relationships, data modeling helps uncover insights that drive informed decisions.

2 Effective Data Analysis

Data modeling allows for efficient querying, data integration, and complex analysis across datasets.

3 Enhanced Data Governance

By defining data relationships, data modeling supports data integrity, security, and compliance.

Saigon Business School

Data Modeling

Steps in Data Modeling

1

Requirement Gathering

Understand business needs, data sources, and user requirements before starting the data modeling process.

2

Conceptual Data Modeling

Create a high-level view of entities, relationships, and attributes to capture business concepts.

3

Logical Data Modeling

Translate business concepts into a detailed data structure using entity-relationship diagrams.

4

Physical Data Modeling

Implement the logical model by defining data types, constraints, and physical storage considerations.

Saigon Business School

Data Modeling

sbsedu.vn

Data Modeling Techniques and Best Practices

Entity-Relationship Modeling

Representing entities, relationships, and attributes to model real-world concepts and their associations.

Dimensional Modeling

Designing data structures that optimize query performance for analytical reporting and data exploration.

Data Normalization

Organizing data into tables with minimal redundancy, ensuring data integrity and efficient updates.

Data Validation

Implementing rules and checks to ensure the accuracy, completeness, and consistency of data.

Saigon Business School

Data Modeling

Challenges in Data Modeling

Data Quality Issues

Inconsistent, incomplete, or inaccurate data can affect the modeling process and analysis outcomes.

Compatibility with Existing Systems

Integrating the data model with legacy systems may require careful consideration and adjustment.

Changing Business Requirements

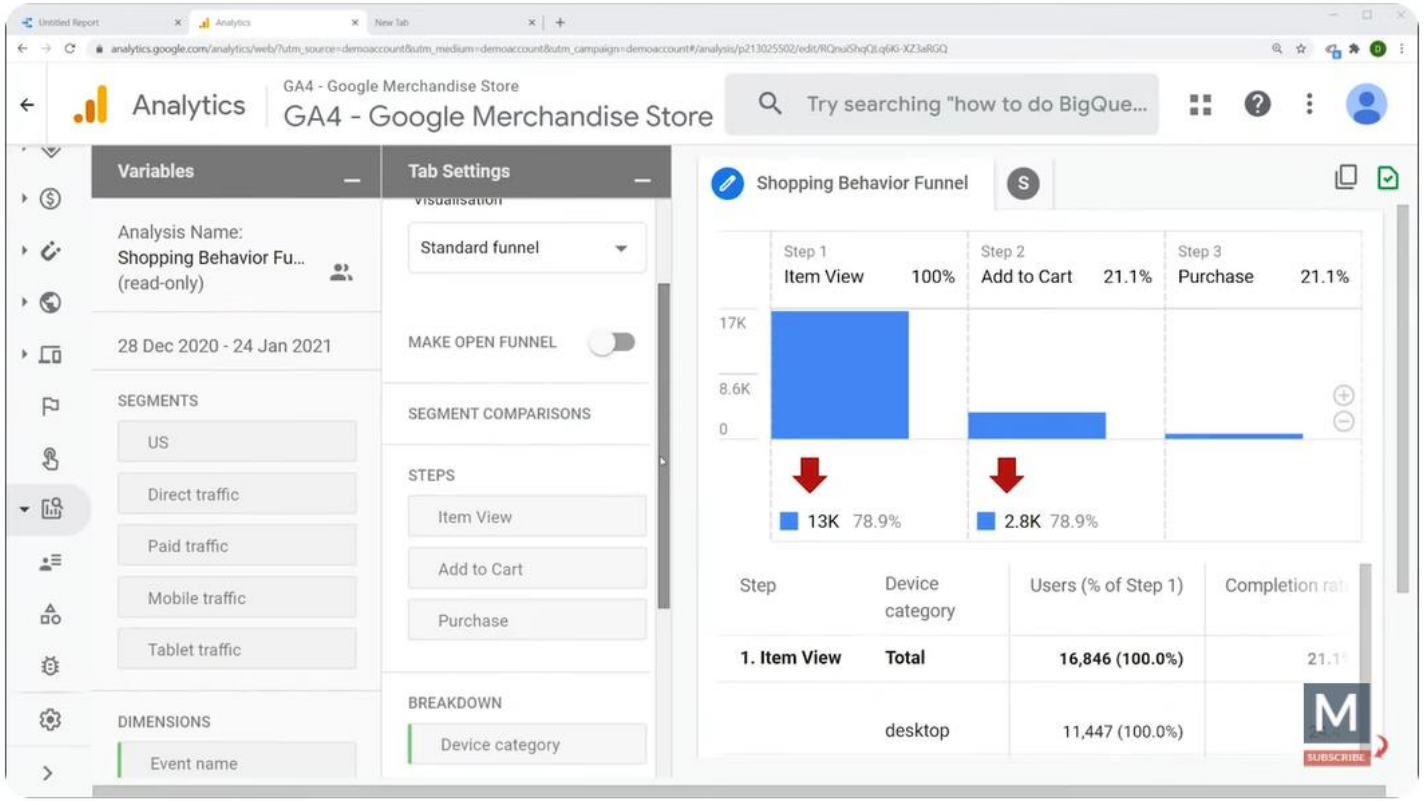
Data models need to be adaptable to evolving business needs and evolving data sources.

Saigon Business School

Data Modeling

Step 1 - Sign In to Google Data Studio.
Step 2 – ????
Step 3 - Create a ??????
Step 4 - Choose ???? and data controls
Step 5 - A Template (Optional)
Step 6 - Add Data.
Step 7 – ??????????
Step 8 - ??????? Your Report.

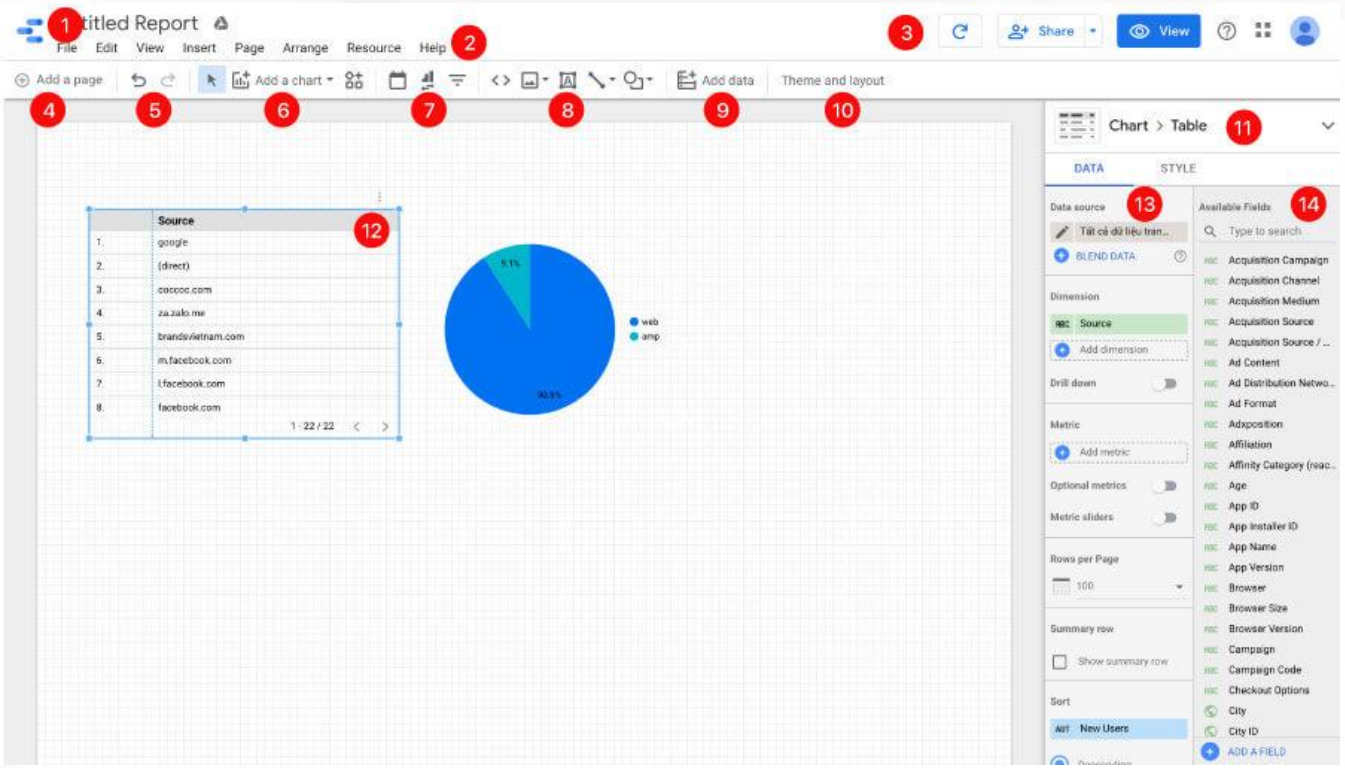
- https://www.youtube.com/watch?v=IN6VsMUGPW8



Google Data Studio Explained in 100 seconds

Saigon Business School

Data Modeling



Google Data Studio report creation page interface

Activity: Associating Numbers with Text through Proper Number Assignment.

??.Logo, you can click to return to the Data Studio home page

??. Menu bar: many functions you can also quickly access here

???Edit chart type

12. Select a chart to display the property sheet for that chart.

??Configure the data properties of the selected chart

Fields available. You can drag dimensions and metrics to the data attribute table.

From left to right:

- Refresh data
- Share reports
- Preview the report
- Get help from Google or feedback
- Quickly access other Google products
- Account management

Add text, draw lines and shapes to the report

Add a data source to the report

Open the themes and layouts panel

Manage reporting pages

Mode selection: undo or redo

Add charts to your report

Add interactive controls (data display time selection, filters...)

Saigon Business School

Google classroom task

- Find a short YouTube video (less than 6 mins) talking about Descriptive Data Analytics.
  - Watch and investigate.
  - Submit your answer:
    - YouTube Link
    - List down 10 keywords
    - Summary text with your comments..

Saigon Business School

Data Modeling

- Match the left column with its best right one:

Visualization Type	Best Use Case
A. Bar Chart	i. Showing trends over time
B. Scatter Plot	ii. Comparing categories and values
C. Line Chart	iii. Displaying the distribution of a variable
D. Histogram	iv. Representing relationships between variables

Statistical Term	Definition
A. Mean	i. Middle value of a dataset
B. Median	ii. Sum of values divided by the count of values
C. Mode	iii. Most frequently occurring value
D. Range	iv. Range of values in a dataset

Saigon Business School

Learning Mission

30 mins

Reading Business Analytics textbook: Chapter 5, page 131 – 180.

Discussion and answer:

Make a brief side glossary table using terms below:

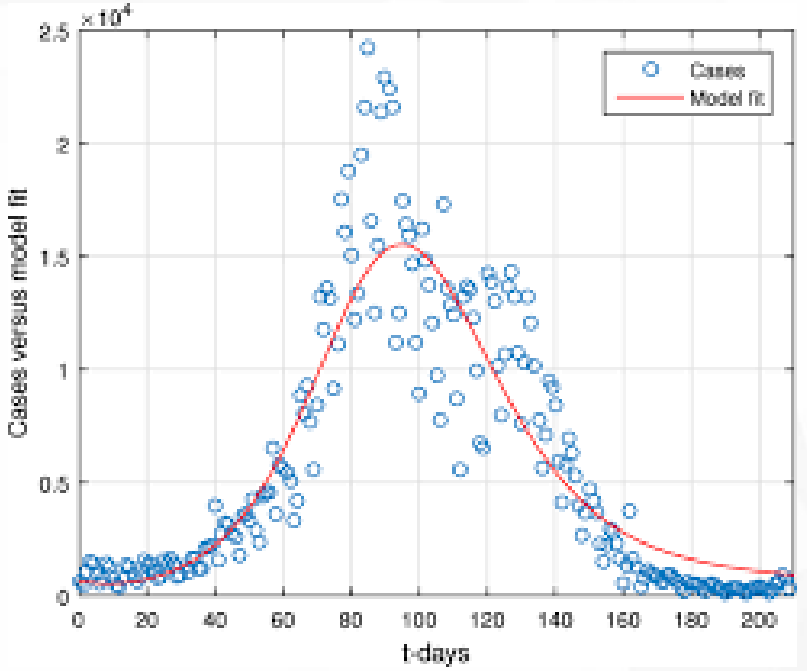
1. Bernoulli distribution
2. Binomial distribution
3. Complement
4. Conditional probability
5. Continuous random variable
6. Cumulative distribution function
7. Discrete random variable
8. Discrete uniform distribution

Saigon Business School

Conclusion and Questions

- Data Modeling

- Data modeling involves creating a visual representation of the structure and organization of data within a system or a database. It includes defining entities, attributes, relationships, constraints, and rules to ensure accurate and efficient data storage, retrieval, and management.
- Data distribution refers to how data is spread or replicated across different locations or nodes in a distributed system. In a distributed database, for example, data may be stored on multiple servers or nodes rather than in a centralized location.



Saigon Business School

Thank you