# Introduction to Data Analysis Fundamentals & Tools

## Understanding the Why and How in Data

Thomas Lim

April 15, 2025

UC Irvine

# Outline

# Module 1: Data & Its Ecosystem

# Module 1: Data & Its Ecosystem

- **Data Fundamentals:**
  - **Data:** A collection of facts to describe a system.
  - **Dataset & Database:** A dataset is a manipulable collection of data; a database is where data is stored.
  - **Data Design:** How information is organized.
- **Data Environment:**
  - **Data Ecosystem:** The various elements that interact to produce, manage, store, analyze, and share data.
  - **Data Life Cycle:** The stages data experiences – plan, capture, manage, analyze, archive, and destroy.
  - **Cloud & Open Data:** Cloud storage offers online data access; open data is publicly available.
- **Data Types:**
  - **Qualitative vs. Quantitative:** Qualitative data explains qualities; quantitative data provides numerical measures.
  - **Small Data vs. Big Data:** Small data covers short-term, specific points; big data involves large, complex datasets.

# Module 2: Analytical Thinking & Business Tasks

# Module 2: Analytical Thinking & Business Tasks

- **Questions to Drive Action:**
  - **Action-oriented question:** A question whose answers lead to change.
  - **Measurable, Specific, Relevant, and Time-bound questions:** Key criteria for effective questioning (SMART methodology).
- **Analytical Approach:**
  - **Analytical skills & Analytical thinking:** Using facts to solve problems step by step. (P.A.C.E.)
  - **Reframing:** Restating a problem to redirect it toward a solution.
- **Business Context:**
  - **Business task:** The problem that data analysis addresses for a business.
  - **Relevant and Leading questions:** Questions that are significant to the problem or steer responses.
  - **Unfair question:** One that makes assumptions or is hard to answer honestly.

# Module 3: Tools & Techniques for Data Manipulation

# Module 3: Spreadsheet Tools and Functions

- **Spreadsheets:**
  - **Spreadsheet:** A digital worksheet use for working with data.
  - Ex: Excel or Google Sheets
  - **Alternatives:** Python, R, SAS, STATA, etc
- **Basic Functions:**
  - **AVERAGE, COUNT, MAX, MIN, SUM:** Common functions to calculate descriptive statistics over a range.
  - **Formula & Function:** Formulas are custom instructions; functions are preset commands.
- **Spreadsheet Tools:**
  - **Automation:** Spreadsheets allow
  - **Filtering & Sorting:** Techniques to display only relevant data or arrange data meaningfully.
  - **Operator & Order of Operations:** Symbols for calculations and grouping expressions.

# Module 3 (cont'd): Pivot Tables & Charts

- **Pivot table:** A tool to summarize and reorganize data.
- **Pivot chart:** A chart created from pivot table data.

# Module 4: Data Analysis Process & Decision-Making

# Module 4: The Data Analysis Process

- **Data Analysis Process:** The six phases – ask, prepare, process, analyze, share, and act.
- **Data-driven decision-making:** Using insights from data to guide business strategy.
- **Data-inspired decision-making:** Finding commonalities in diverse data sources.
- **Data strategy:** Managing the people, processes, and tools used in data analysis.

# Module 4 (cont'd): Metrics & Reporting

- **Metrics:**
  - **Metric & Metric goal:** Quantifiable measures to assess performance.
  - **Return on Investment (ROI):** Evaluates the success of an investment.
  - **Revenue & Turnover rate:** Indicators of business performance.
- **Reporting:**
  - **Report:** A static collection of data for stakeholders.
  - **Dashboard:** A dynamic tool for monitoring live data.

# Module 5: Querying Data & Databases

# Module 5: Querying Data

- **Query:** A request for data or information.
- **Query language & SQL:** Programming languages to communicate with databases (Structured Query Language).
- **Cell Range:** A collection of cells used in queries and functions.

# Module 6: Problem Solving & Advanced Analysis

# Module 6: Problem Solving Techniques

- **Gap analysis:** Evaluating current processes to find areas for improvement.
- **Root cause:** Identifying why a problem occurs.
- **Problem domain & Problem types:** The scope and variety of challenges in data analysis.
- **Structured thinking:** Breaking down problems logically.

# Module 6 (cont'd): Measuring and Evaluating Questions

- **Relevant question & Specific question:** Focused questions that address key issues.
- **Time-bound question:** Questions that specify a timeframe.
- **Leading question:** Questions that steer responses.

# Module 7: Visualization & Communication

# Module 7: Data Visualization & Reporting

- **Data visualization:** The graphical representation of data.
- **Report & Dashboard:** Tools to communicate findings to stakeholders.

# Module 7 (cont'd): Equations and Math Functions

- **Equation & Math expression:** Calculations involving arithmetic operations.
- **Math function:** A function used within mathematical formulas.

# Wrap-Up & Q/A

# Wrap-Up and Next Steps

- This presentation introduced key terms and processes in data analysis.
- Review these concepts and explore hands-on practice with spreadsheets, SQL, and visualization tools.
- Use these definitions as a guide for effective, data-driven decision-making.

# Questions?

Thank You!