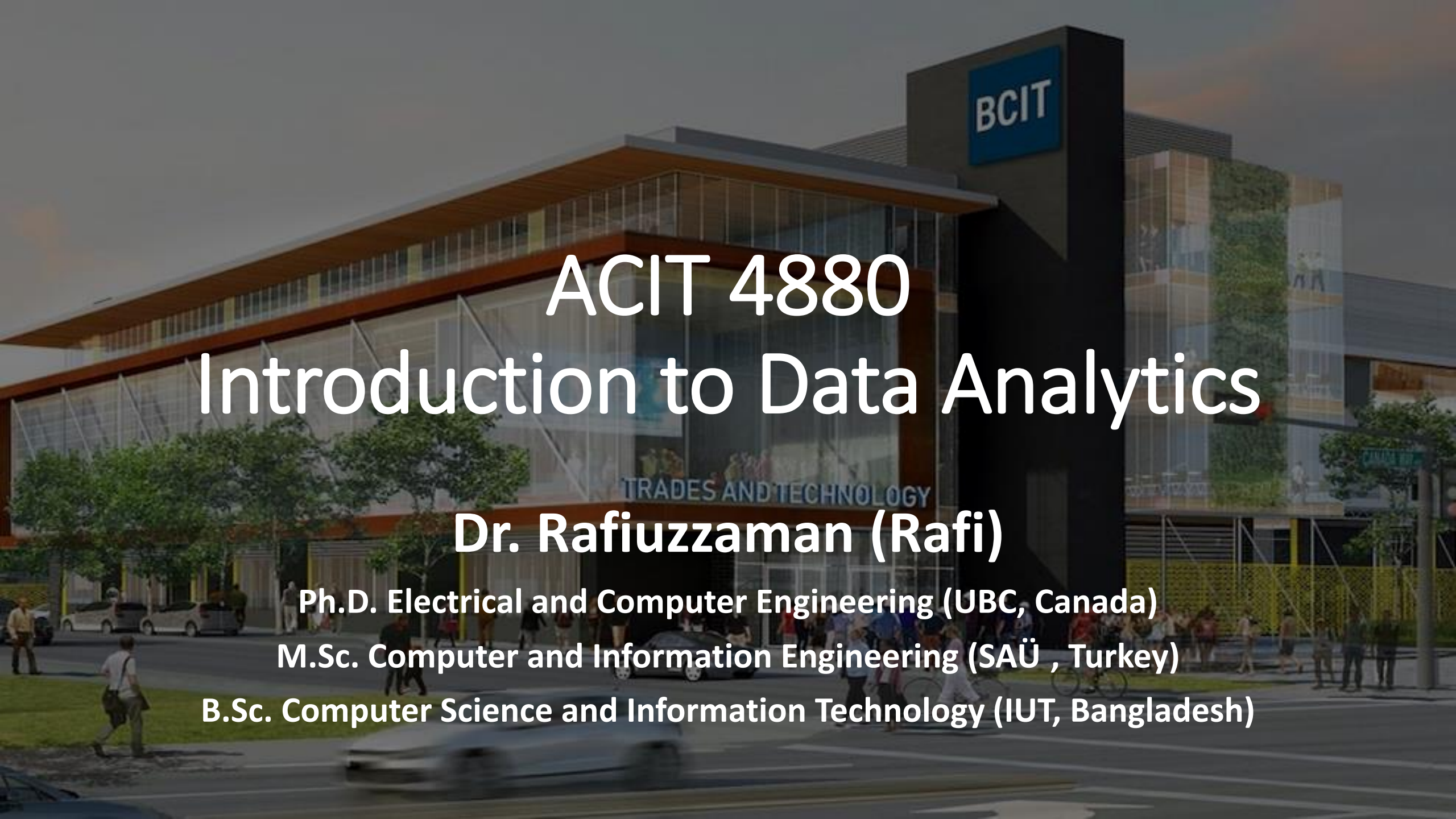


ACIT 4880
Introduction to Data Analytics

Dr. Rafiuzzaman (Rafi)

Ph.D. Electrical and Computer Engineering (UBC, Canada)

M.Sc. Computer and Information Engineering (SAÜ , Turkey)

B.Sc. Computer Science and Information Technology (IUT, Bangladesh)

Learning Objectives

- Overview
- Use cases
- Types of analytics
- The scientific method for analytics
- Course outline

What is Data Analytics

A Carpenter + An Electrician + A Plumber =

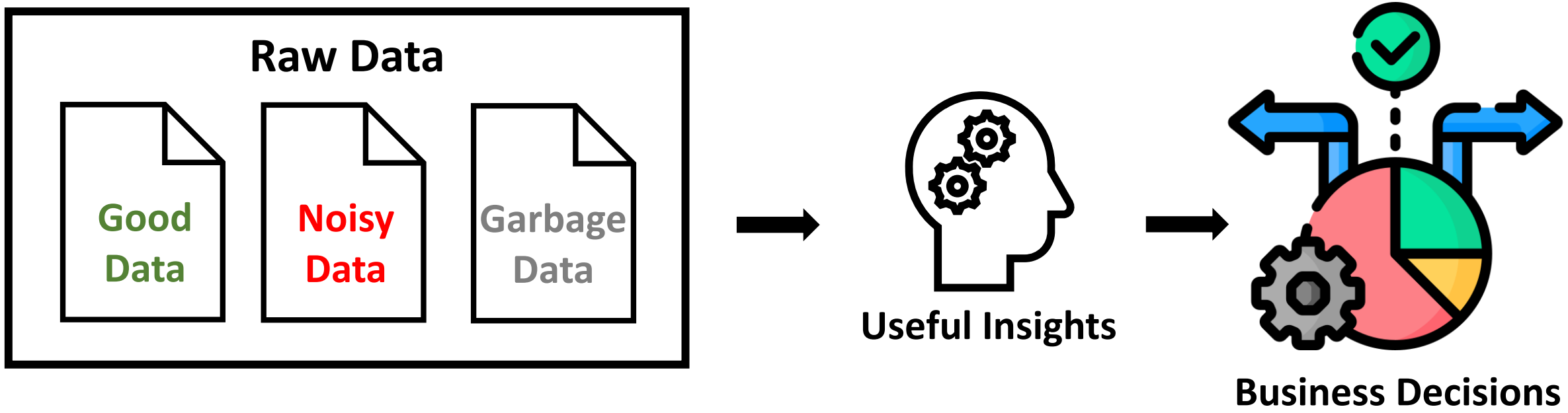
Your messed up home

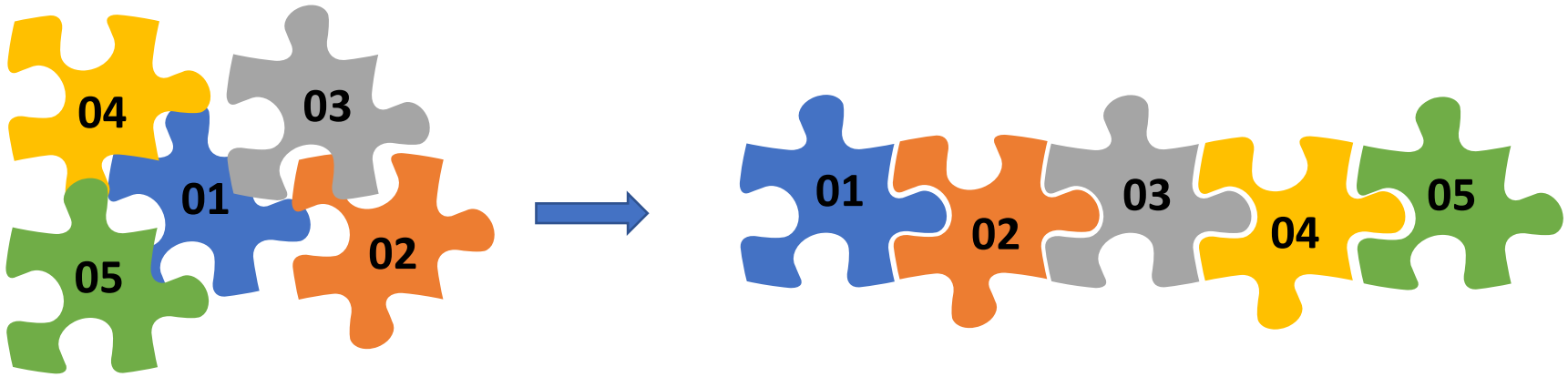
What is Data Analytics (Cont....)

Useful Insights

Business Decisions

What is Data Analytics(Cont....)

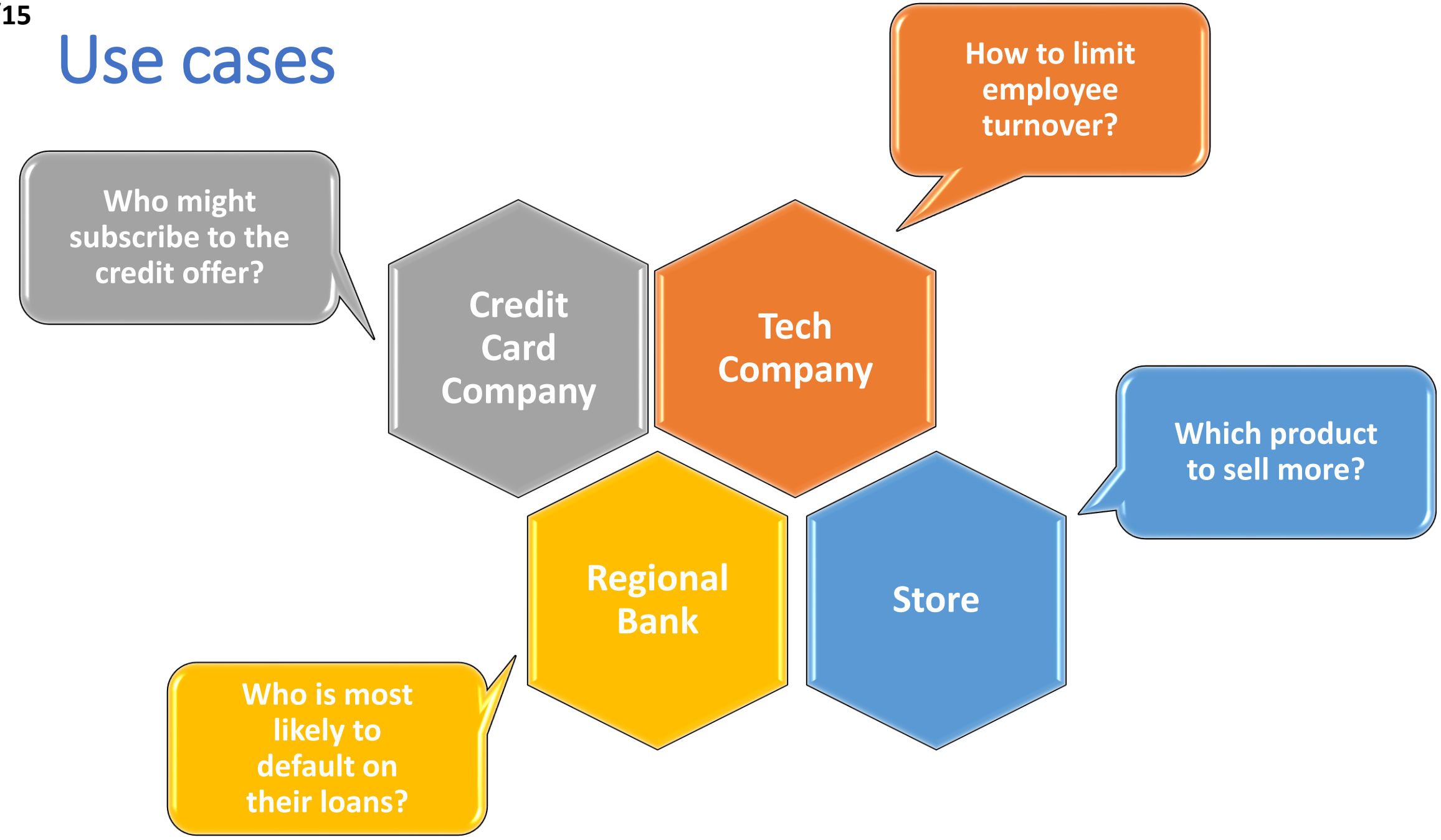




Learning Objectives

- Overview
- Use cases
- Types of analytics
- The scientific method for analytics
- Course outline

Use cases


How to limit employee turnover?
Who might subscribe to the credit offer?
Credit Card Company
Tech Company
Which product to sell more?
Regional Bank
Store
Who is most likely to default on their loans?

Use cases


facebook
Social Analytics
amazon
Improving E-commerce experience
UBER
Optimizing Rides
Bank of America
Increasing Customer experience
Airbnb
Improving Searches
Spotify
Music Recommendation

Learning Objectives

- Overview
- Use cases
- Types of analytics
- The scientific method for analytics
- Course outline

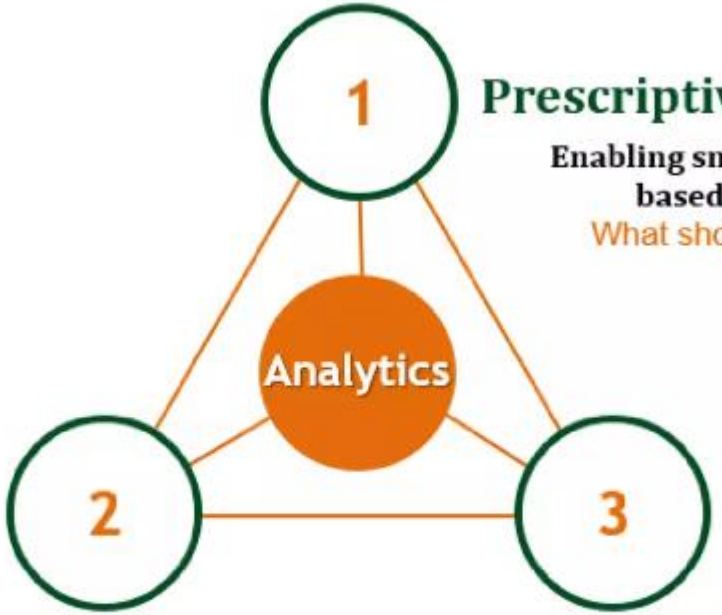
Types of Analytics

Why do airline prices change every hour?

Prescriptive Analytics
advice on possible outcomes

How do grocery cashiers know to hand you coupons you might actually use?

Predictive Analytics
understanding the future



1 Prescriptive Analytics
Enabling smart decisions based on data
What should we do?

2 Predictive analytics
Predicting the future based on historical patterns
What could happen?

3 Descriptive analytics
Mining data to provide business insights
What has happened?

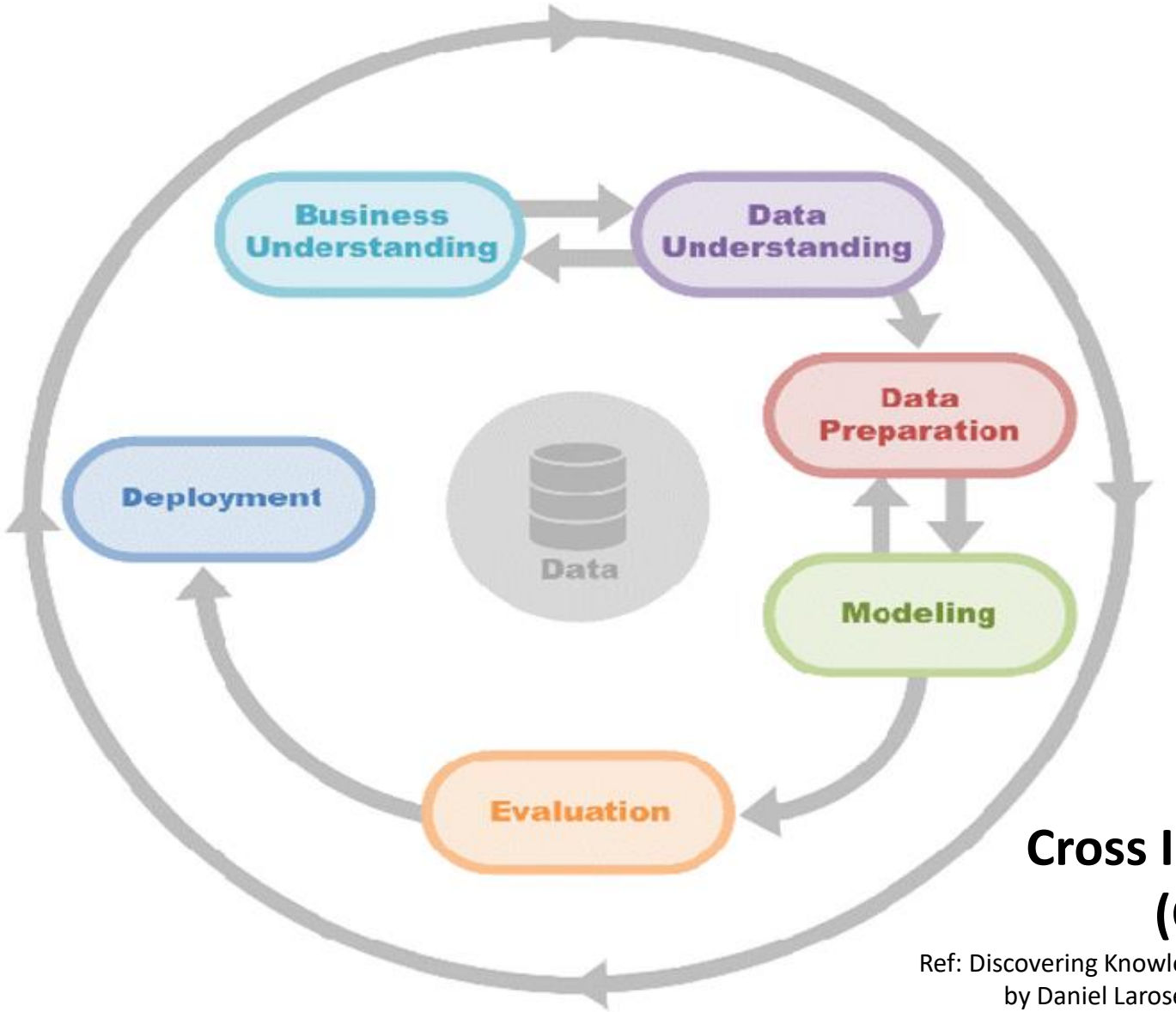
How does Netflix frequently recommend just the right movie?

Descriptive Analytics
insight into the past

Learning Objectives

- Overview
- Use cases
- Types of analytics
- The scientific method for analytics
- Course outline

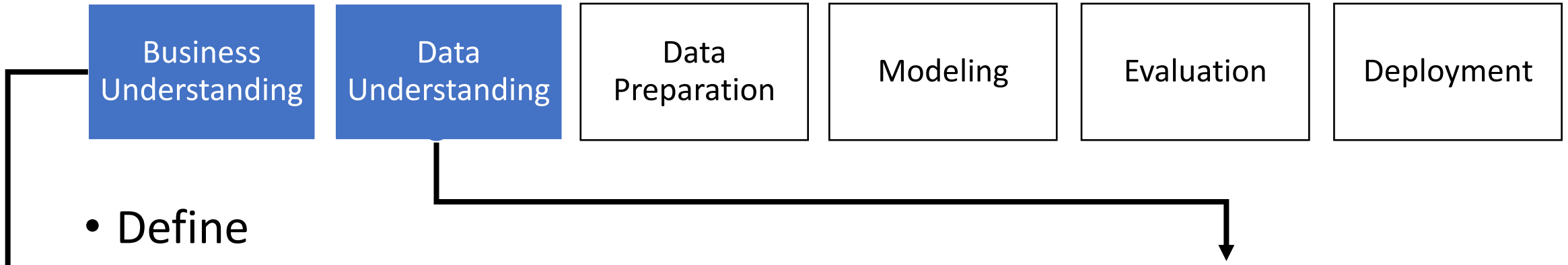
The Scientific Method for Analytics



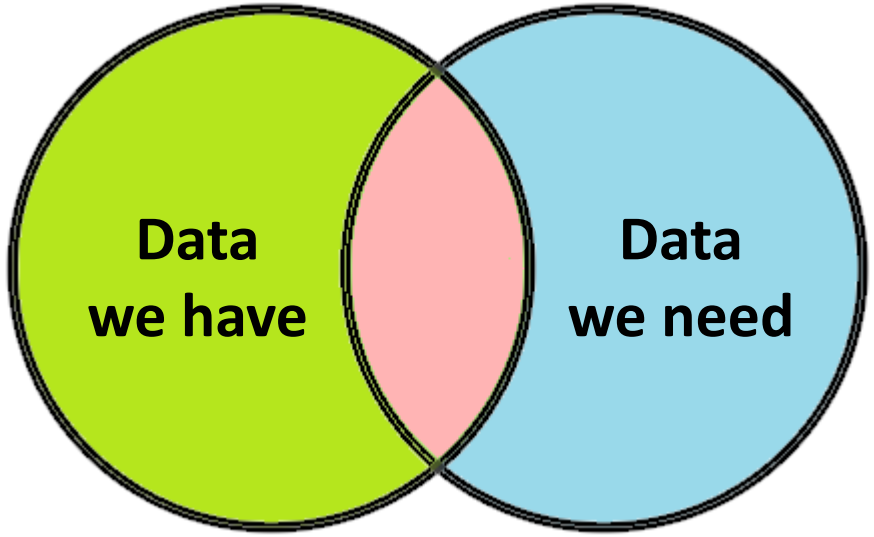
Cross Industry Standard Process (CRISP-DM) Lifecycle

Ref: Discovering Knowledge in Data: An Introduction to Data Mining, Second Edition, by Daniel Larose and Chantal Larose, John Wiley and Sons, Inc., 2014.

Method: Understanding - Business and Data



- Define
  - Project requirements and objectives
- Translate objectives
  - Into data mining problem definition
- Prepare preliminary strategy
  - To meet objectives



Method: Data Preparation

Business Understanding	Data Understanding	**Data Preparation**	Modeling	Evaluation	Deployment

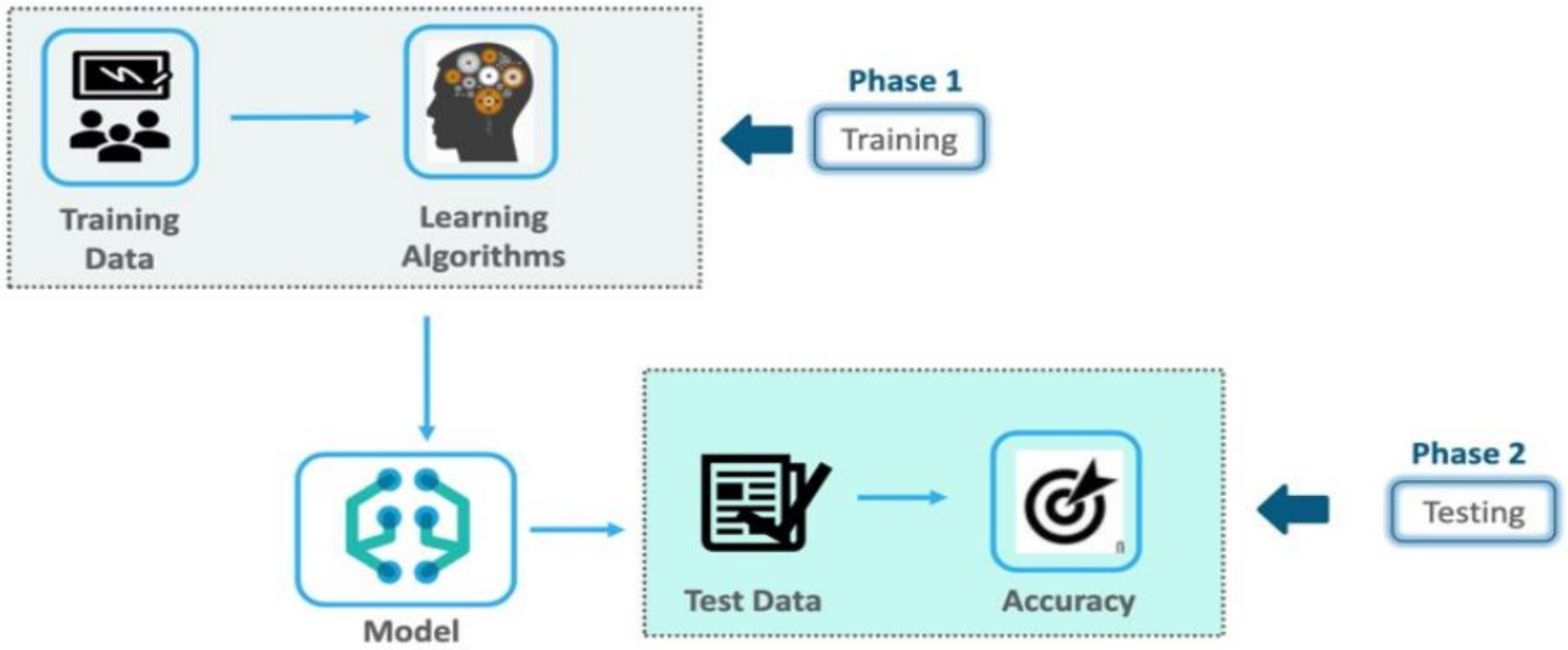
TABLE 2.1 Can You Find Any Problems in This Tiny Data Set?

Customer ID	Zip	Gender	Income	Age	Marital Status	Transaction Amount
1001	10048	M	75000	C	M	5000
1002	J2S7K7	F	—40000	40	W	4000
1003	90210		10000000	45	S	7000
1004	6269	M	50000	0	S	1000
1005	55101	F	99999	30	D	3000

Ref: Discovering Knowledge in Data: An Introduction to Data Mining, Second Edition, by Daniel Larose and Chantal Larose, John Wiley and Sons, Inc., 2014.

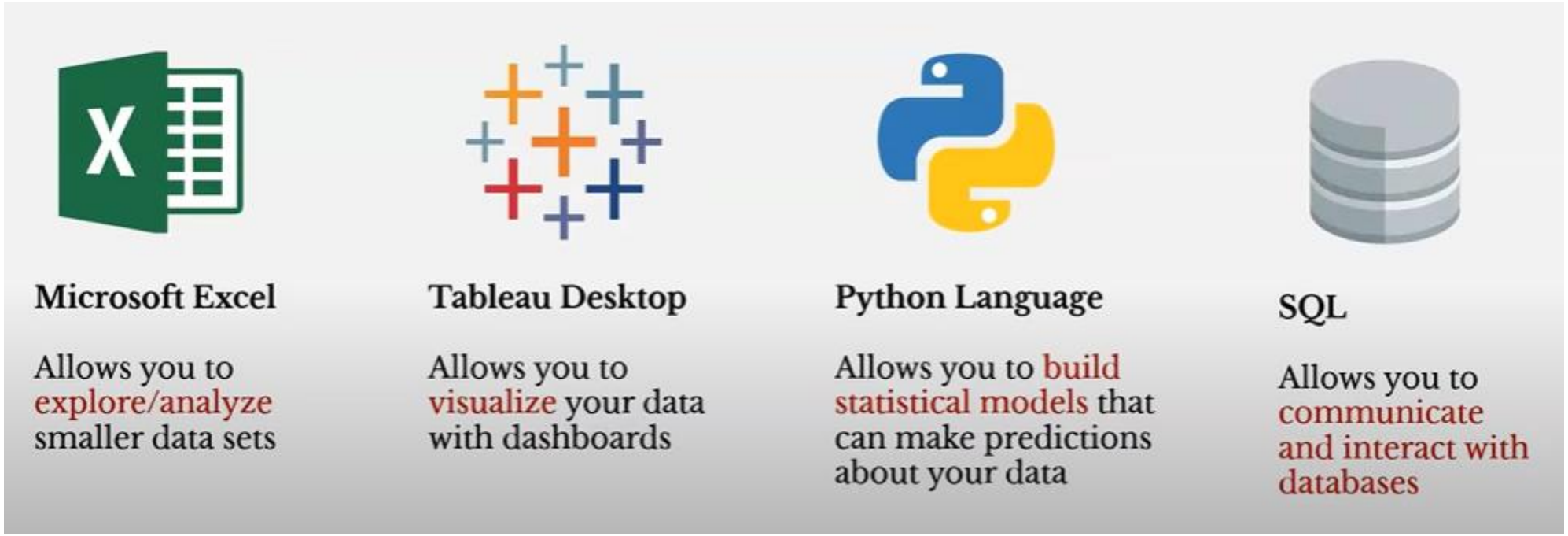
Method: Modeling and Evaluation

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment



Method: Deployment and Tools

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment



Learning Objectives

- Overview
- Use cases
- Types of analytics
- The scientific method for analytics
- Course outline

Evaluation Criteria

Criteria	%	Comments
In-class Labs	20	Open book/resource. Most of the labs will be done and graded in the same class.
Quizzes	10	Three quizzes that take place at the beginning of any three classes to cover the flipped component.
Projects	30	Two projects occur - one before the midterm and another after the midterm.
Exam	40	Closed book/resource exam covering the entire course materials.

Course Outline

Week	Week starts on	Topics	Activities	Flipped
1	08-Jan	Introductory Week	Installation	
2	15-Jan	Introduction to Python / Python Data Structures	Lab1	Reading: Chap1 from Ref2
3	22-Jan	Numpy/ Pandas	Lab2	Reading: Chap2 from Ref2
4	29-Jan	Data Preprocessing/ Visualization	Lab3	Reading: Chap1 & Chap2 from Ref1
5	05-Feb	Exploratory Data Analysis	Lab4	Reading: Chap3 from Ref1
6	12-Feb	Statistical Analysis	Lab5	Reading: Chap4 from Ref1
7	**19-Feb**	**Project1 Presentations**		
8	**26-Feb**	**Midterm Week**	NA	
9	04-Mar	Regression Analysis	Lab6	Reading: Chap6 from Ref1
10	11-Mar	Spring Break	NA	
10	18-Mar	k-Nearest Neighbor Algorithm Support vector machine	Lab7	Reading: Chap7 from Ref1
11	25-Mar	Decision Trees	Lab8	Reading: Chap8 from Ref1
		Random Forest		
12	01-Apr	Hierarchical & k-Means Clustering	Lab9	Reading: Chap9 from Ref1
14	**08-Apr**	**Project2 Presentations**		
15	**15-Apr**	**Final Exam**		