Doing Data Science with Python

COURSE INTRODUCTION

Abhishek Kumar
AUTHOR

@meabhishekkumar

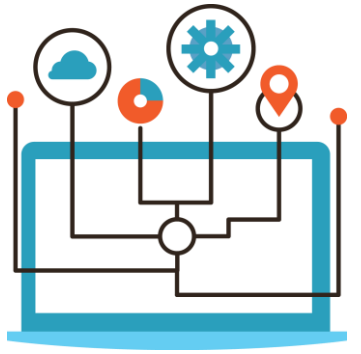
2,500,000,
000,000,
000,000
000,000
(2.5 Quintillion Bytes /
2.3 Trillion Gigabytes)
Per Day

40 Zettabytes
(43 Trillion Gigabytes)
by 2020

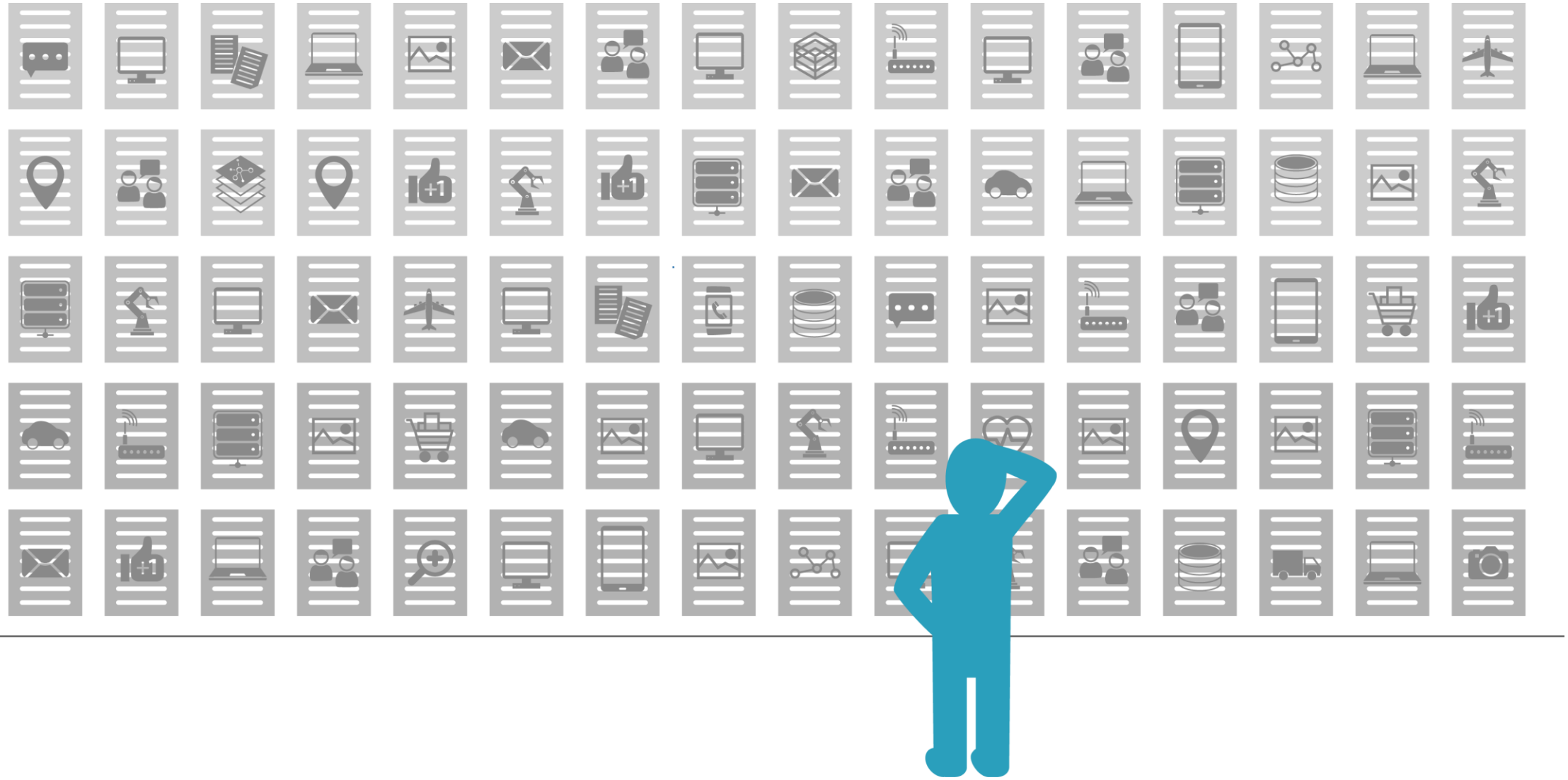
Transactions

Social Networks

Processes



Internet of Things



Data Science

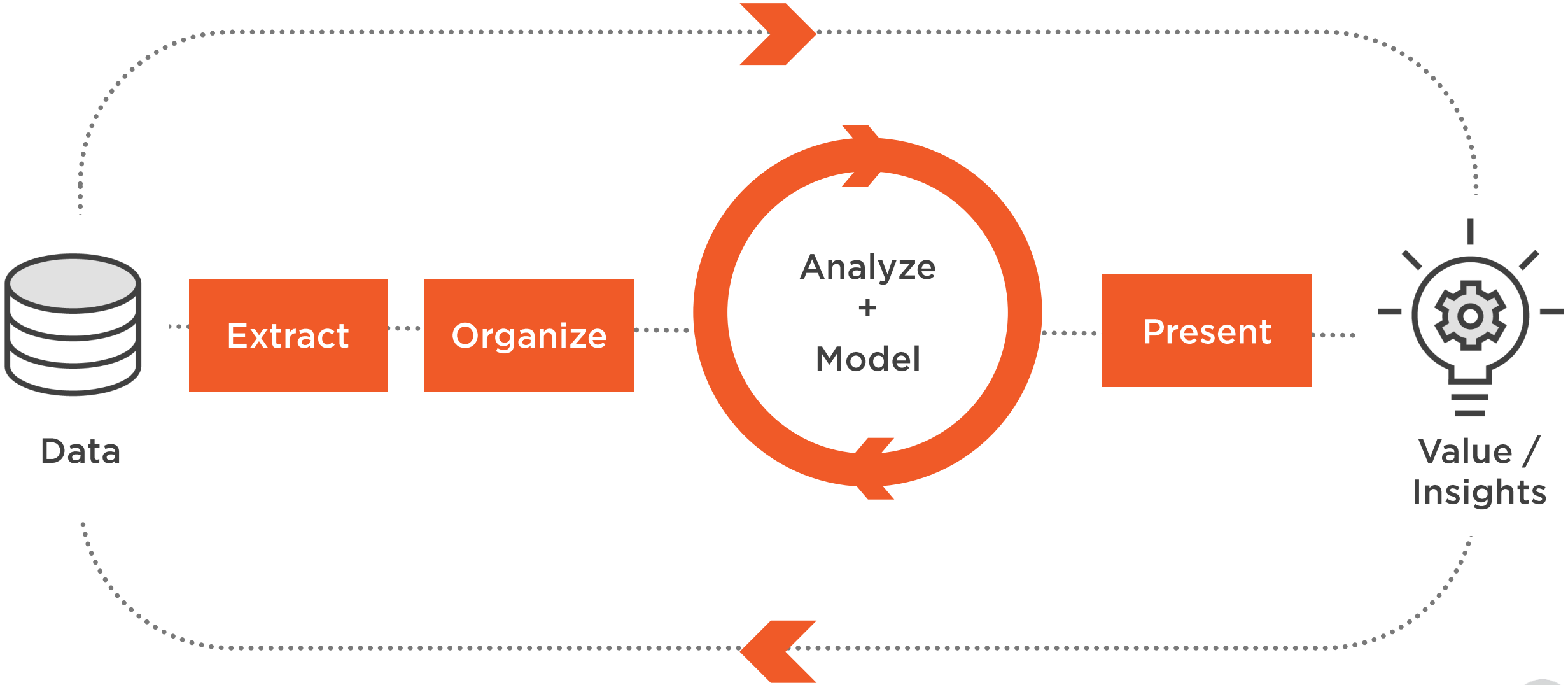
Set of fundamental principles that guide the extraction of knowledge of data.

Source : Data Science for Business, Foster Provost & Tom Fawcett

"Sexiest Job of the 21st Century."

- Harvard Business Review

What You Will Learn?



What You Will Learn?

- Data science project cycle
- Hands-on exercises
- Case study
- Best practices
- Tips & Tricks

Case Study : Titanic Disaster

- No specific domain knowledge required
- Both simple & complex

Case Study : Titanic Disaster

kaggle | Host | Competitions | Datasets | Kernels | Jobs | Community | Sign up | Login

Knowledge • 5,165 teams

Titanic: Machine Learning from Disaster

Fri 28 Sep 2012 — Sat 31 Dec 2016 (3 months to go)

Dashboard

Competition Details » Get the Data » Make a submission

Target Audience

Aspirants

Professionals

Target Audience

Aspirants

Professionals

Target Audience

Aspirants

Professionals

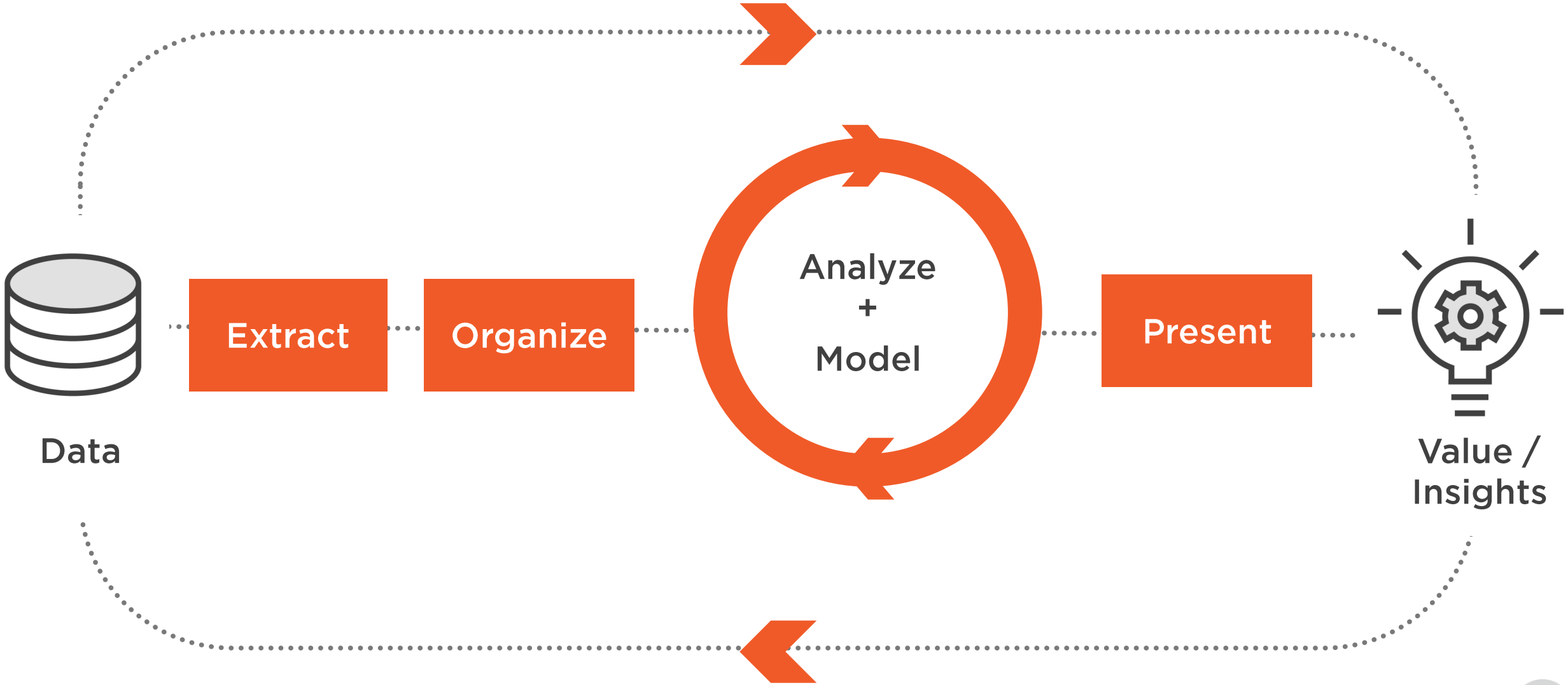
Course Prerequisites

Python

High school mathematics

Statistics & Machine Learning

Data Science Project Cycle Overview



Why Python for Data Science?

Language

Easy and intuitive

Packages

Tools and Libraries

Community

Active community

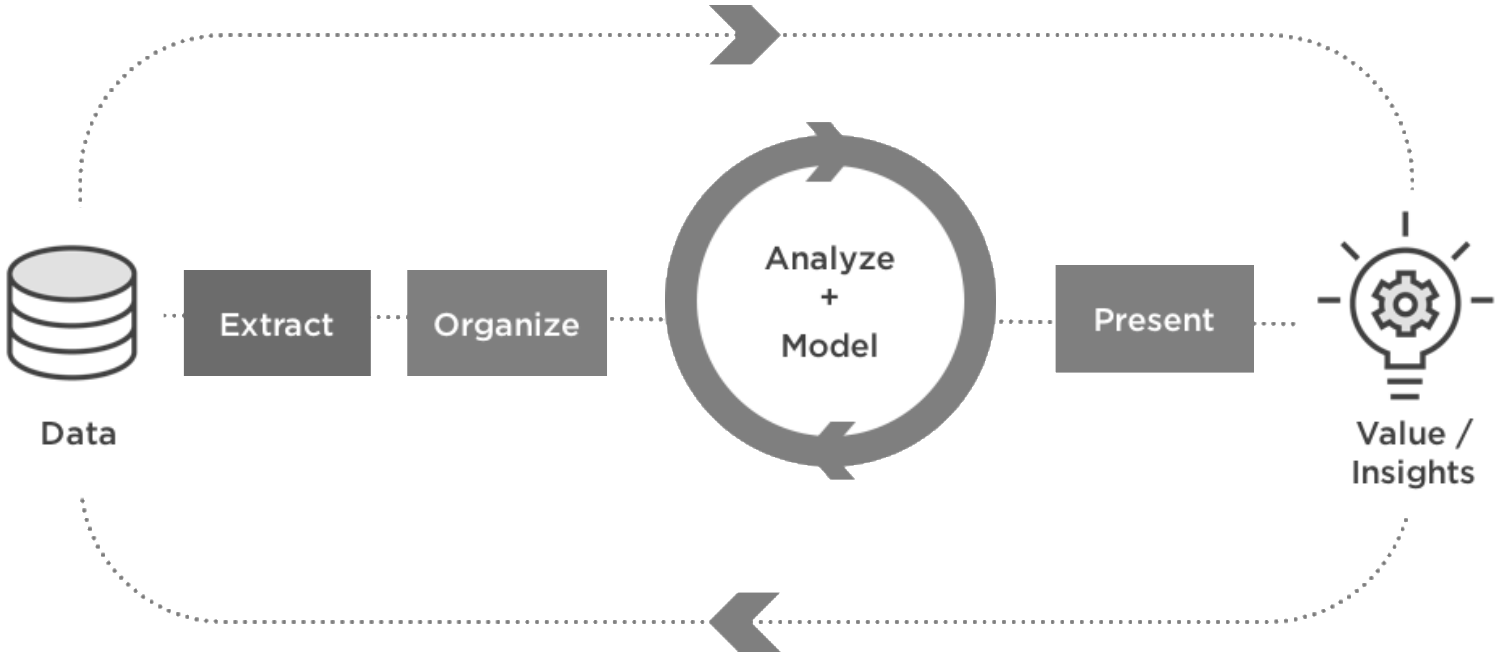
Scalability

Fast

Production

Python based application stack

Course Outline

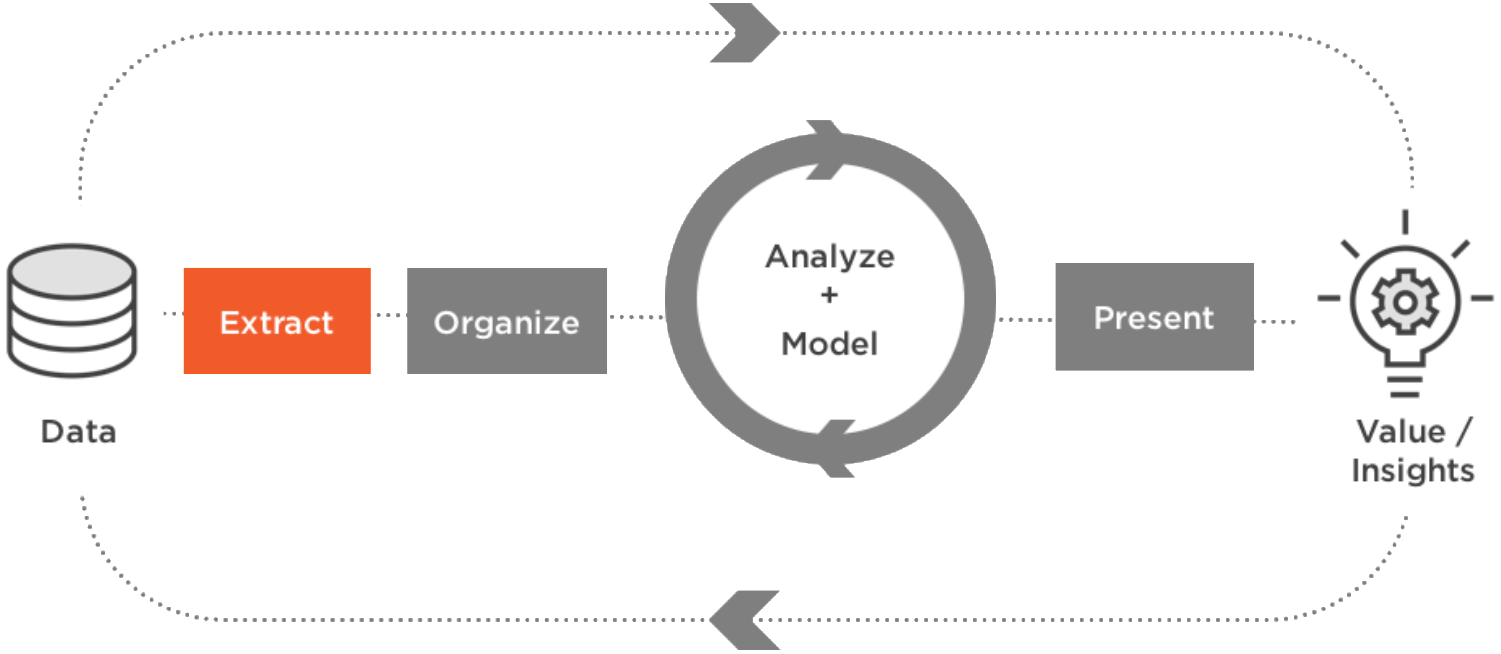


Setting up the Environment

Module 2

- Python distributions
- Jupyter notebook
- Data science project template
- Versioning system

Course Outline

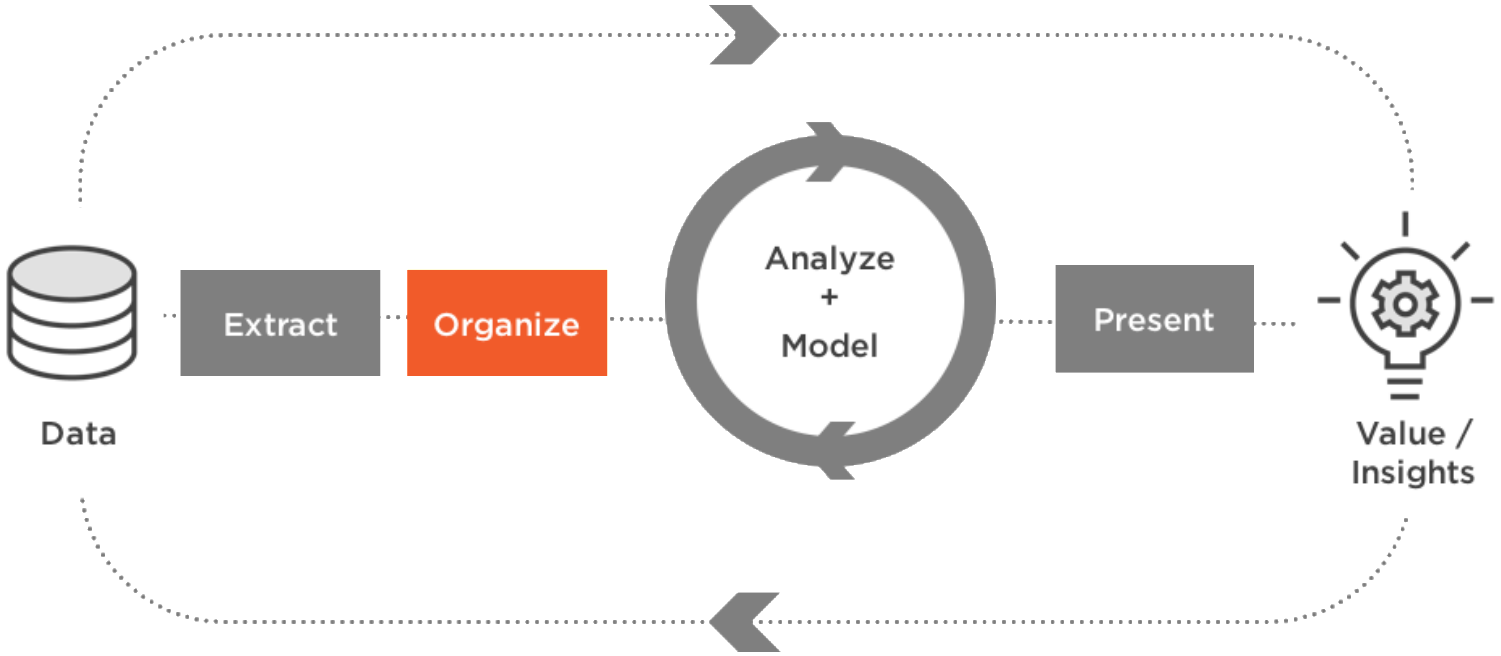


Extracting Data

Module 3

- Databases
- APIs
- Web scraping
- Titanic dataset
- Python
  - Database connectors
  - Requests
  - BeautifulSoup

Course Outline

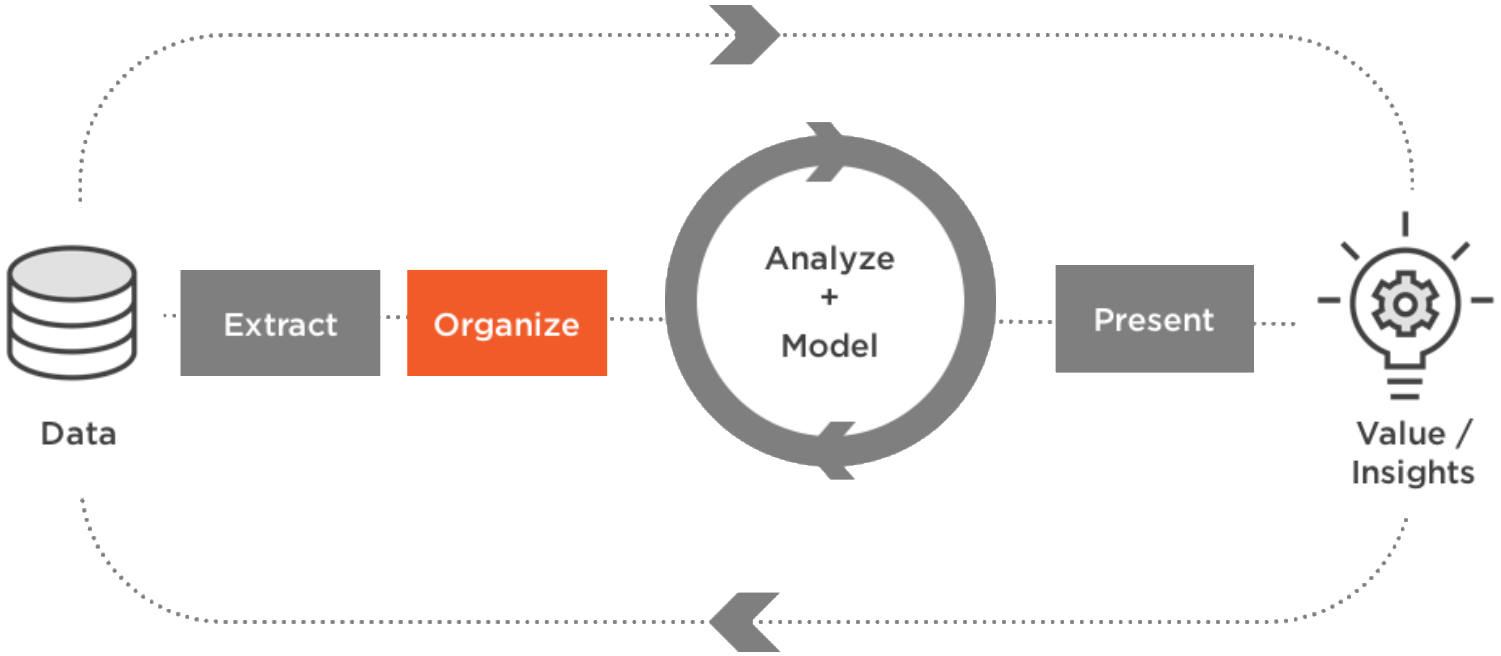


Exploring and Processing Data

Module 4 : Part 1

- Basic exploratory data analysis
- Python
  - NumPy
  - Pandas

Course Outline

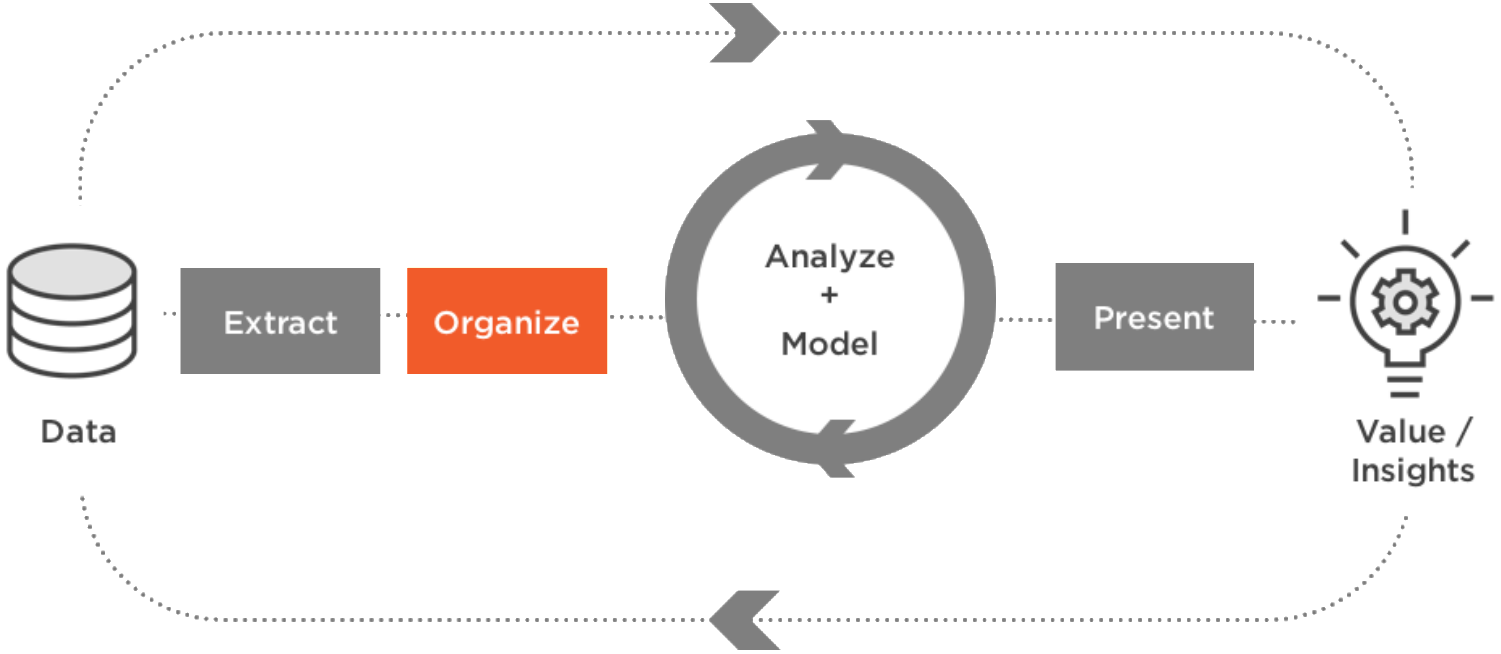


Exploring and Processing Data

Module 5 : Part 2

- Advanced exploratory data analysis
- Python
  - NumPy
  - Pandas

Course Outline

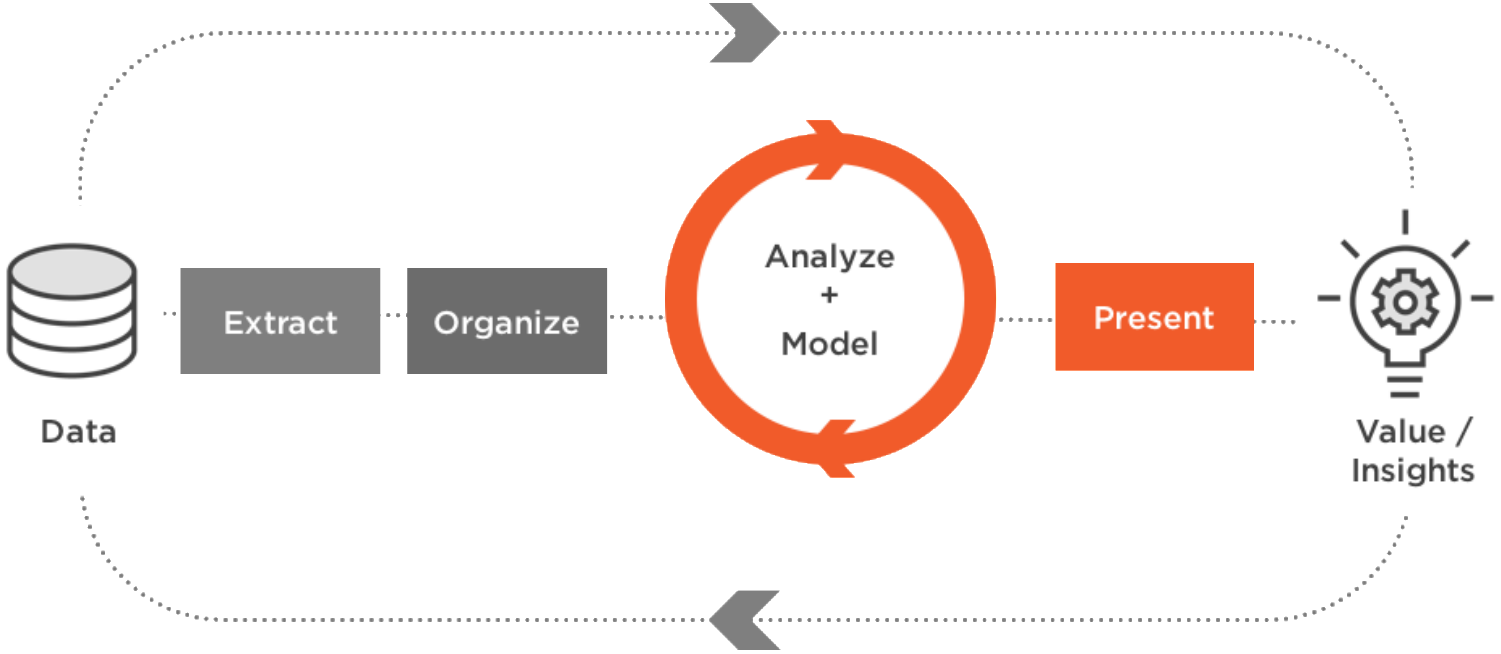


Exploring and Processing Data

Module 6 : Part 3

- Data munging
- Feature engineering
- Visualization
- Python
  - NumPy
  - Pandas
  - Matplotlib

Course Outline

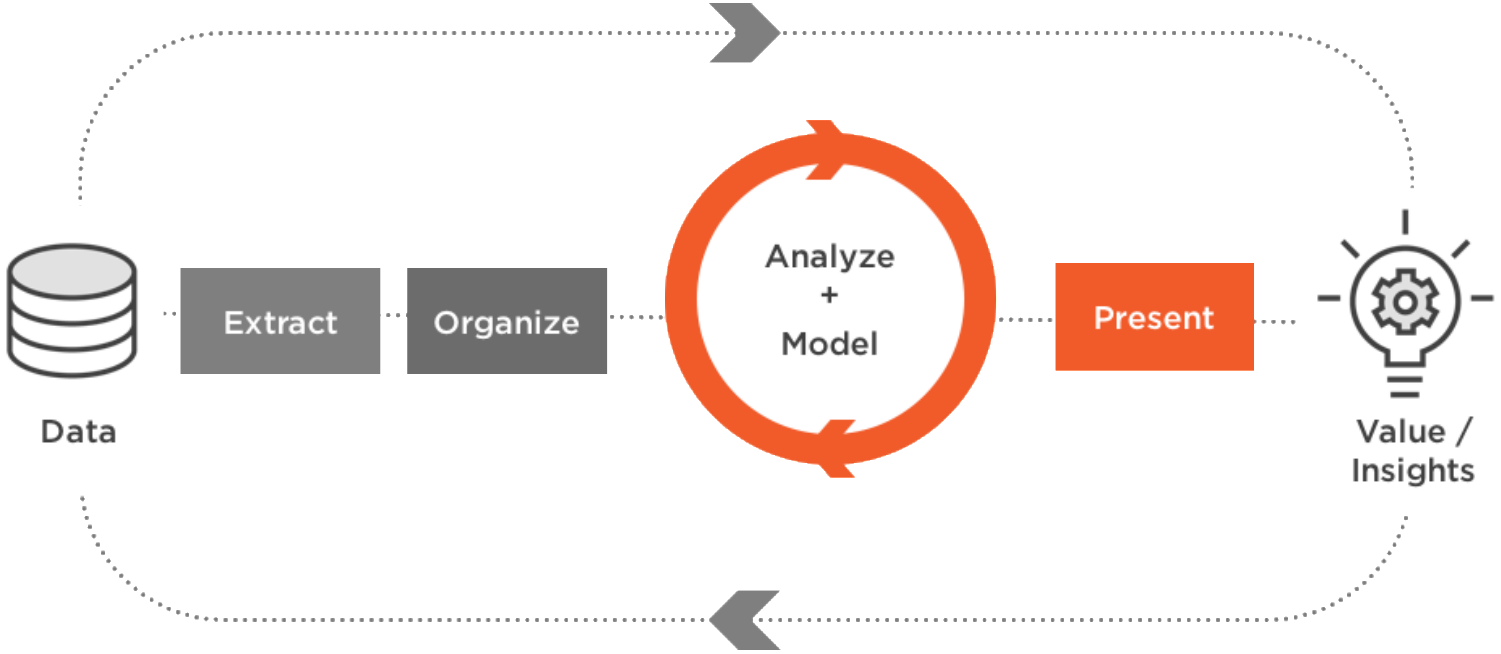


Building and Evaluating Predictive Model

Module 7 : Part 1

- Machine learning
- Build and evaluate model
- Kaggle submission
- Python
  - Scikit-Learn

Course Outline



Building and Evaluating Predictive Model

Module 8 : Part 2

- Model tuning
- Model persistence
- Machine learning API
- Python
  - Scikit-Learn
  - Pickle
  - Flask

Summary

Data science introduction

Data science project cycle

Why python for data science?

Course outline