DATA SCIENCE BOOTCAMP

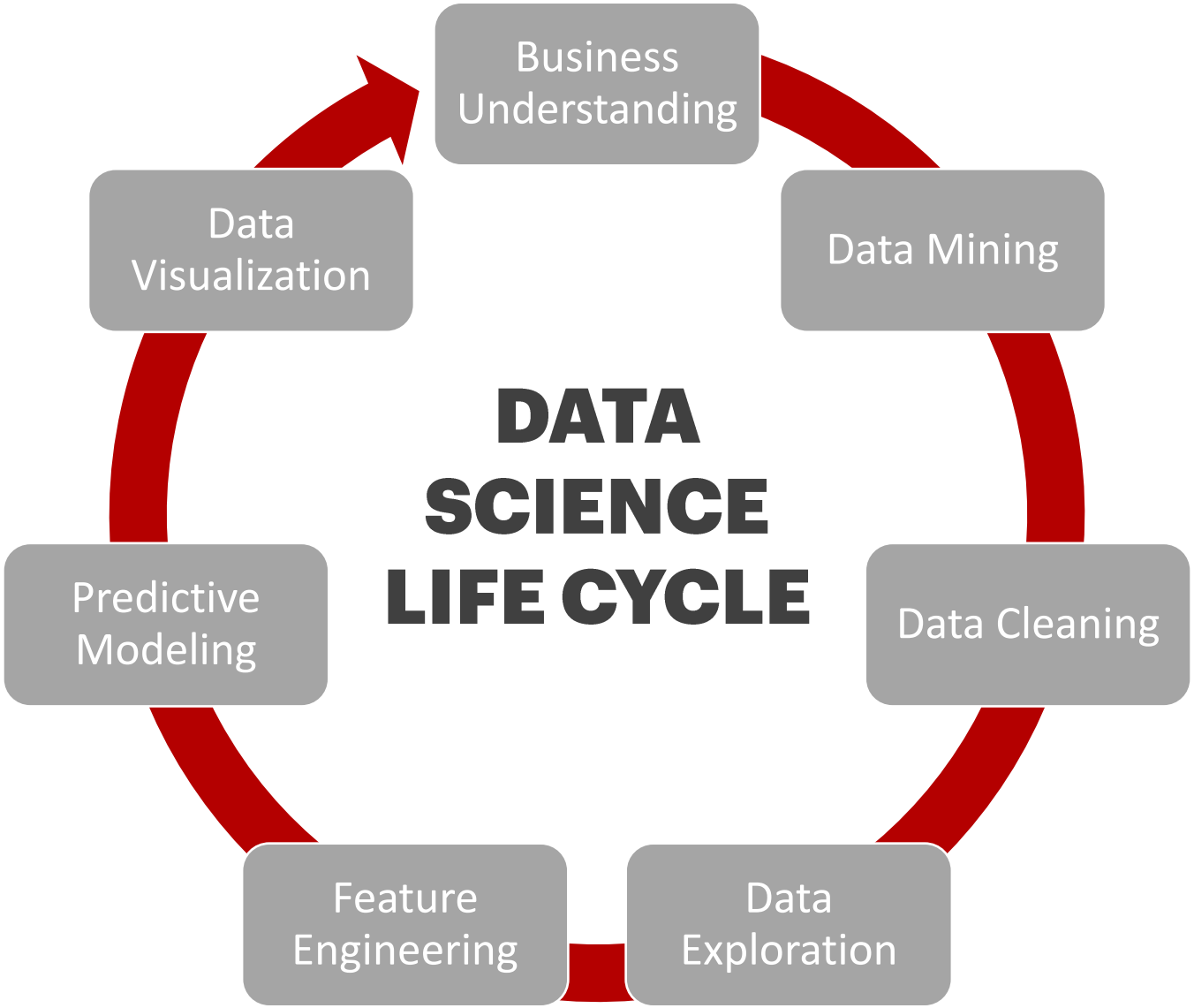
Session 2: Data Science Methodology and Python Syntax

SESSION 2 OUTLINE

- DATA SCIENCE METHODOLOGY
- PYTHON SYNTAX

SESSION 2 OUTLINE

- DATA SCIENCE METHODOLOGY
- PYTHON SYNTAX



DATA SCIENCE ROLES

Data Engineer

Key Technical Skills

- ETL/ELT Development (Informatica, Talend, Python)
- Data Management (SQL, Spark, Hadoop)
- Data Orchestration (Python, platform-specific services)
- Data Pipeline Deployment, Monitoring & Maintenance
- Business Intelligence and Data Visualization (Tableau, Power BI, Qlik)

Machine Learning Engineer

Key Technical Skills

- Statistical Programming/Data Mining (R, Python)
- Machine Learning (R, Python)
- Model Deployment, Monitoring & Maintenance

Data Scientist

Key Technical Skills

- Statistical Programming/Data Mining (R, Python)
- Machine Learning (R, Python)
- Statistics, Applied Mathematics

DATA SCIENCE PROCESS

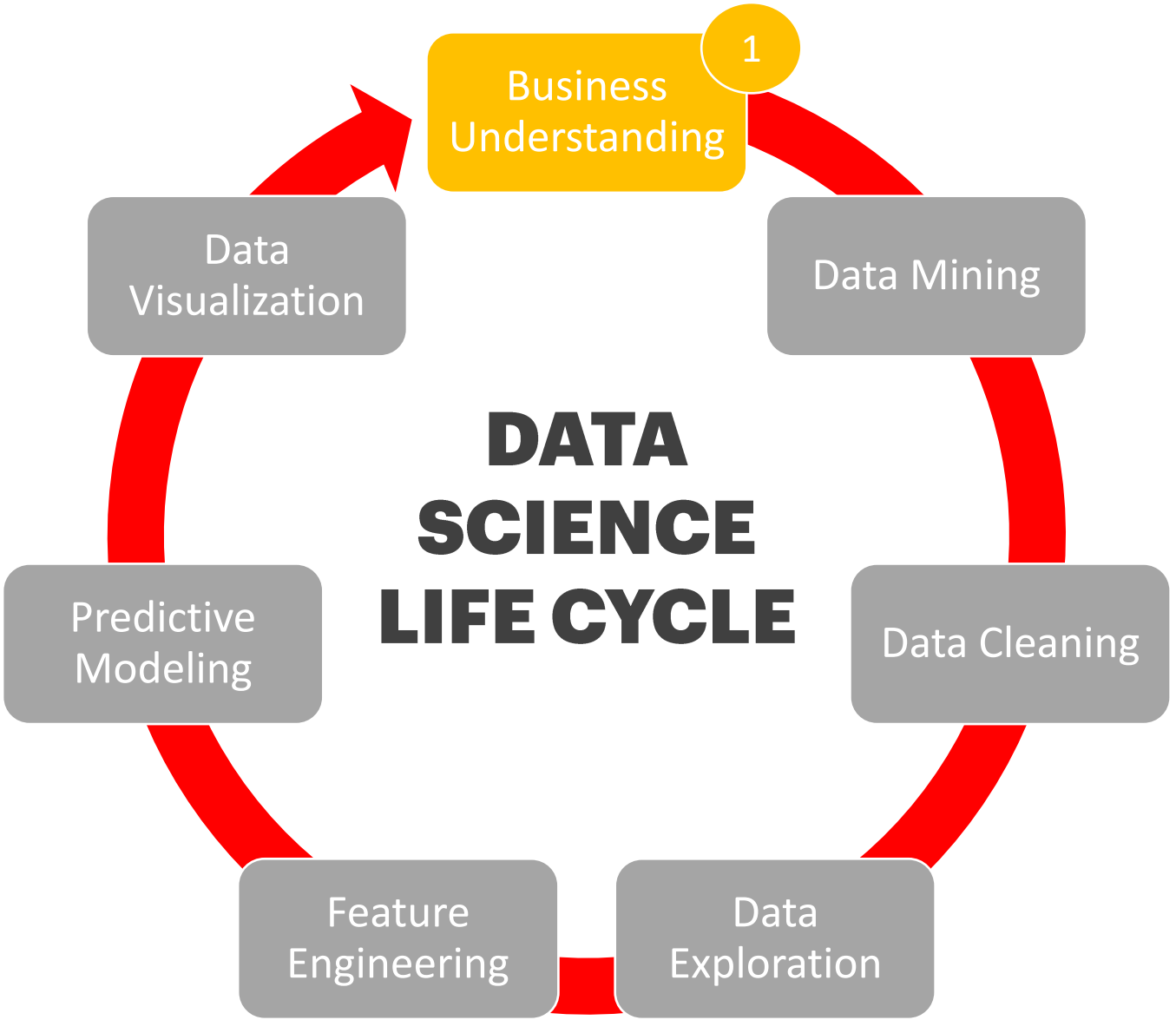
DATA ENGINEERING

COLLABORATION

ML DEVELOPMENT

BUSINESS UNDERSTANDING

Ensures that every decision made is supported by concrete data and that it is guaranteed (with a high probability) to achieve results.



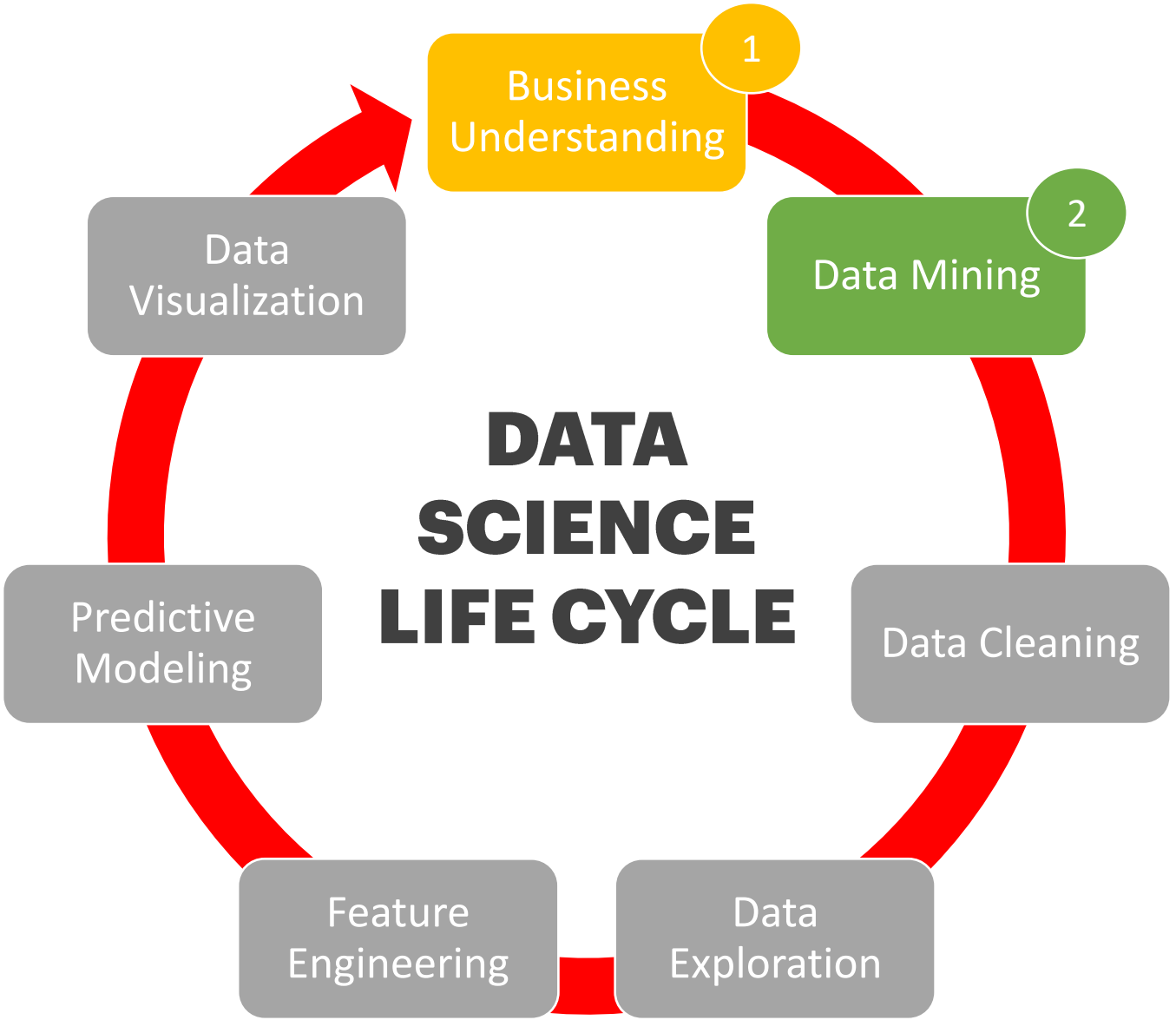
DATA ENGINEERING

COLLABORATION

ML DEVELOPMENT

DATA MINING

Data mining is the process of gathering data from different sources.



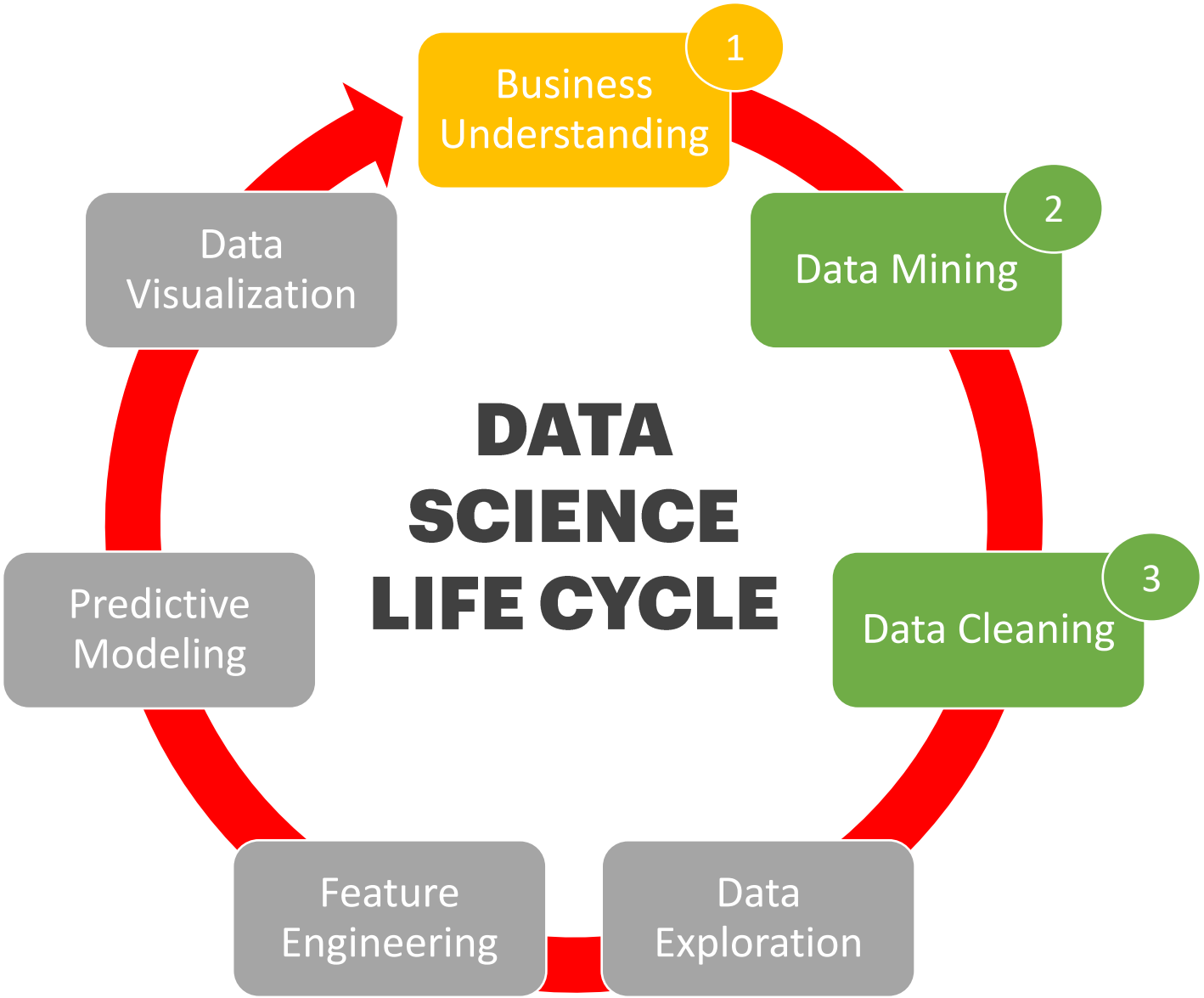
DATA ENGINEERING

COLLABORATION

ML DEVELOPMENT

DATA CLEANING

Most time-consuming step of all.



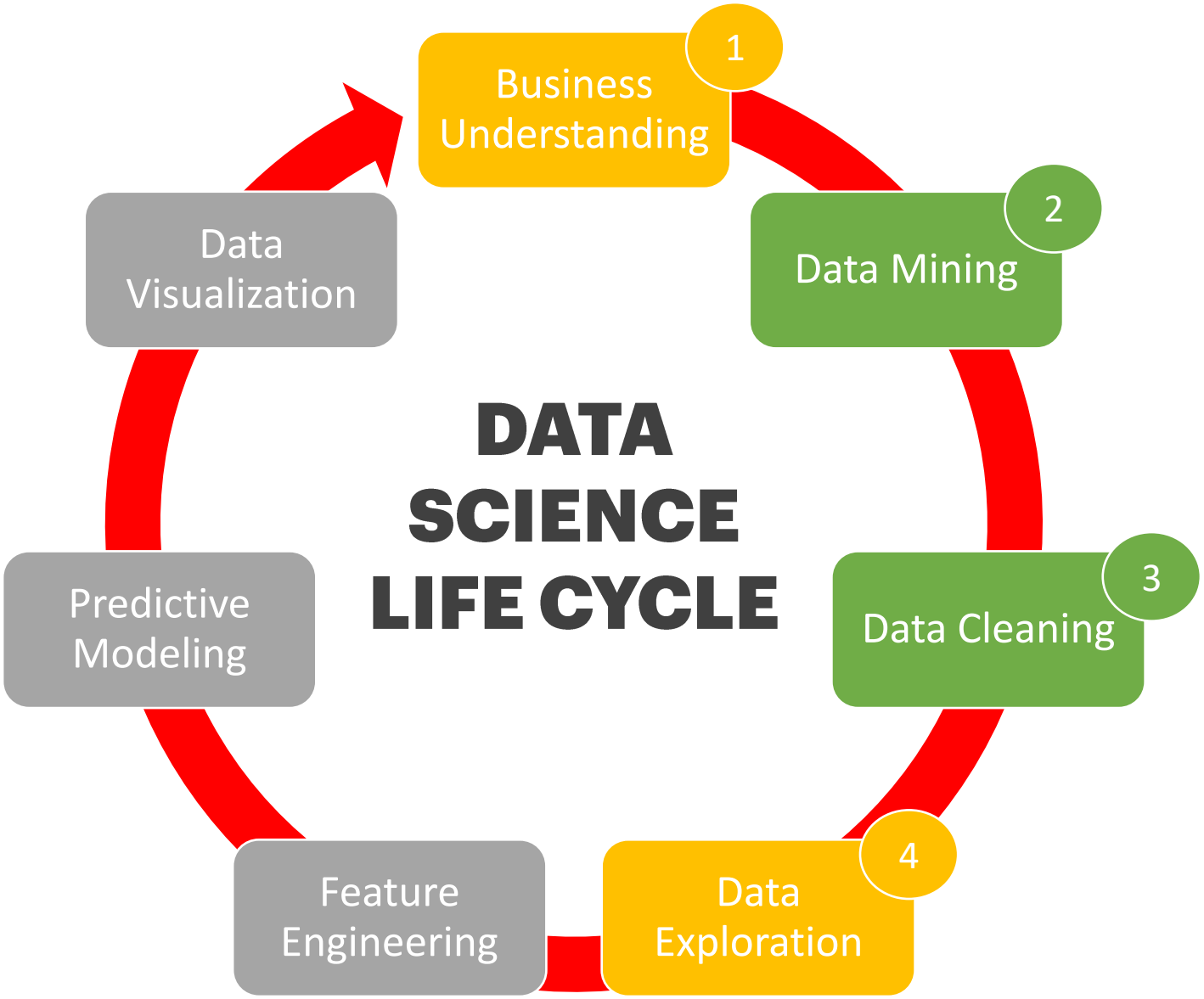
DATA ENGINEERING

COLLABORATION

ML DEVELOPMENT

DATA EXPLORATION

This stage is like the brainstorming of data analysis.



DATA ENGINEERING

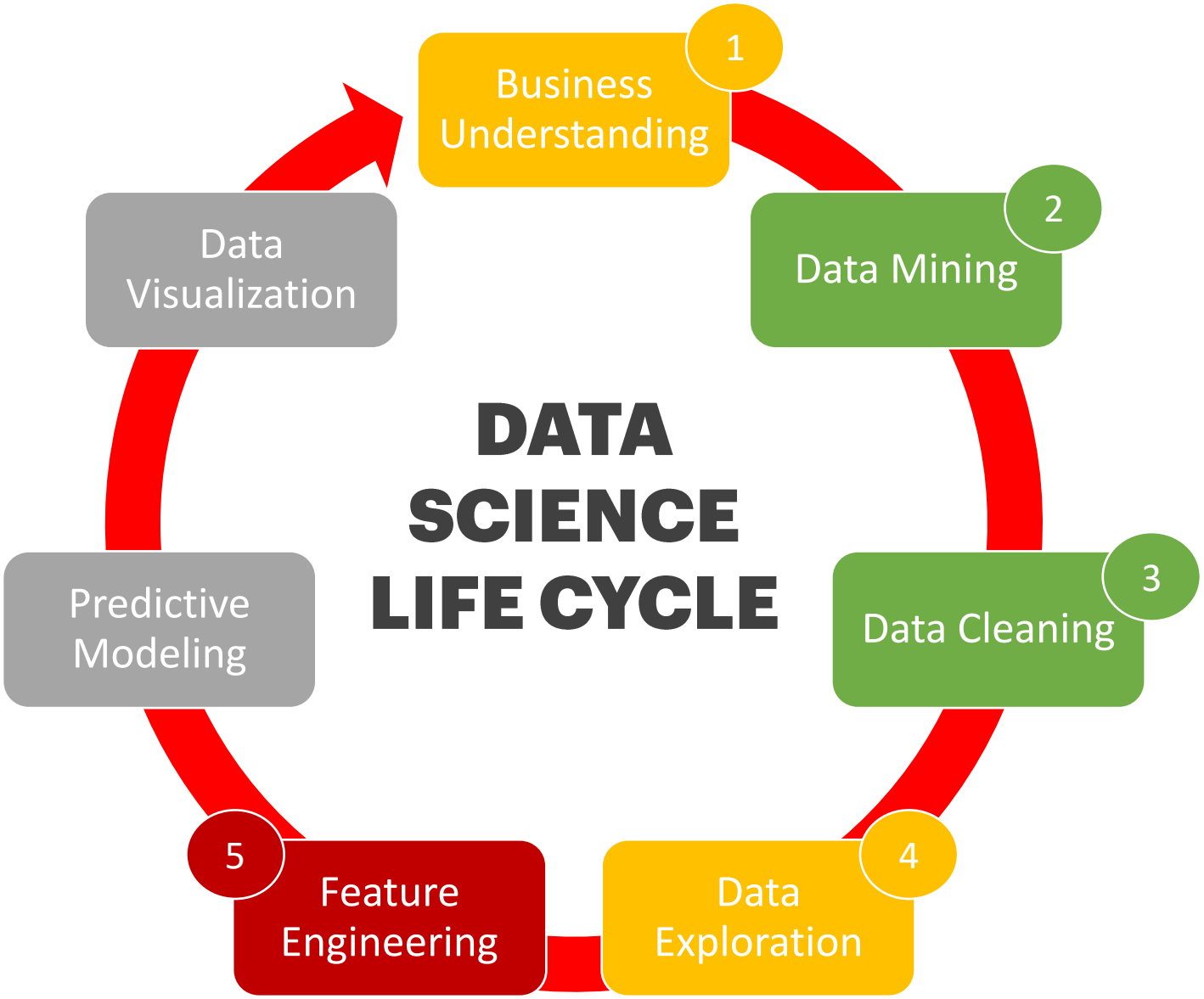
COLLABORATION

ML DEVELOPMENT

FEATURE ENGINEERING

In machine learning, a feature is a measurable property or attribute of a phenomenon being observed.

Two tasks: Feature selection and Feature engineering



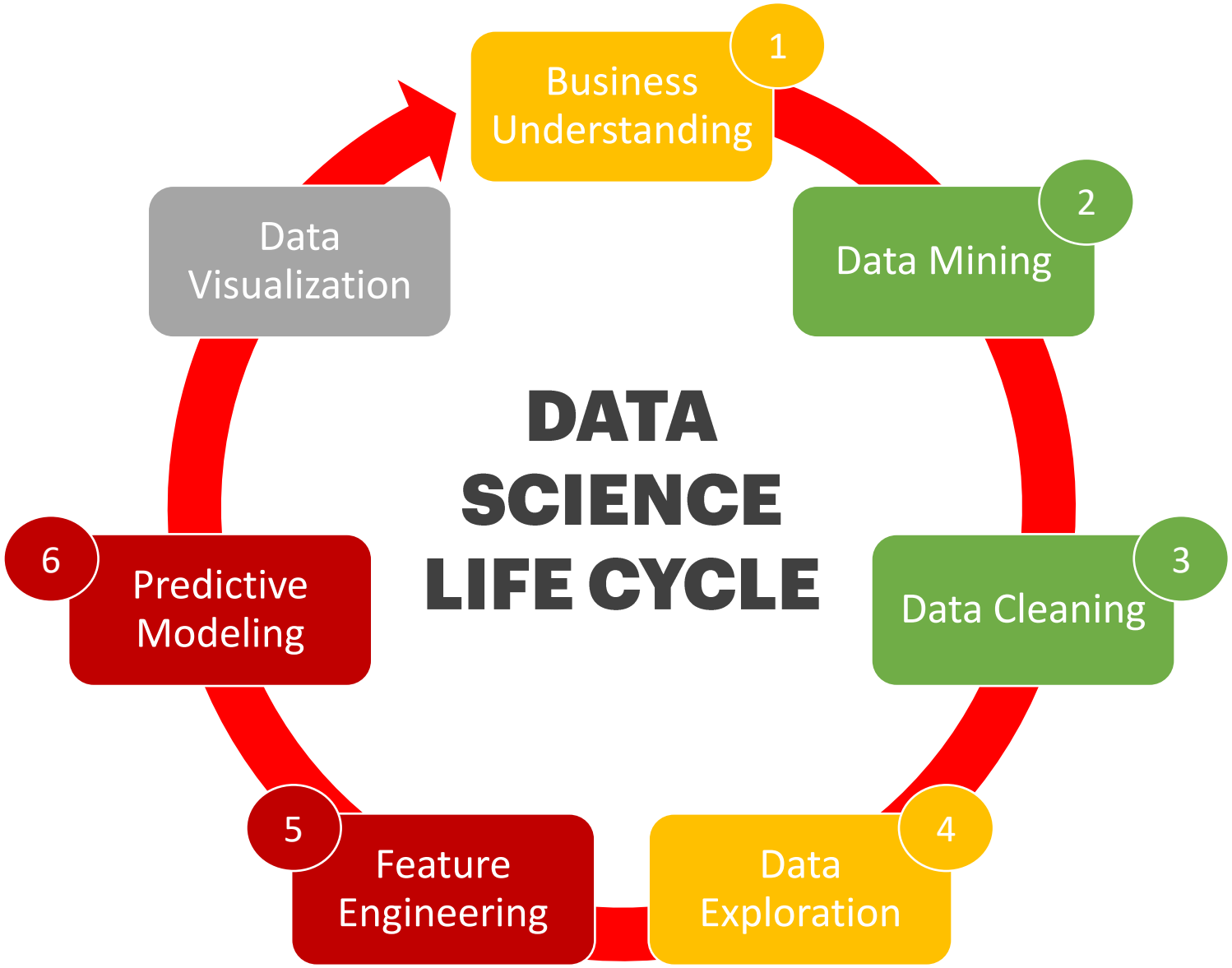
DATA ENGINEERING

COLLABORATION

ML DEVELOPMENT

PREDICTIVE MODELING

This is where machine learning finally comes into the project.



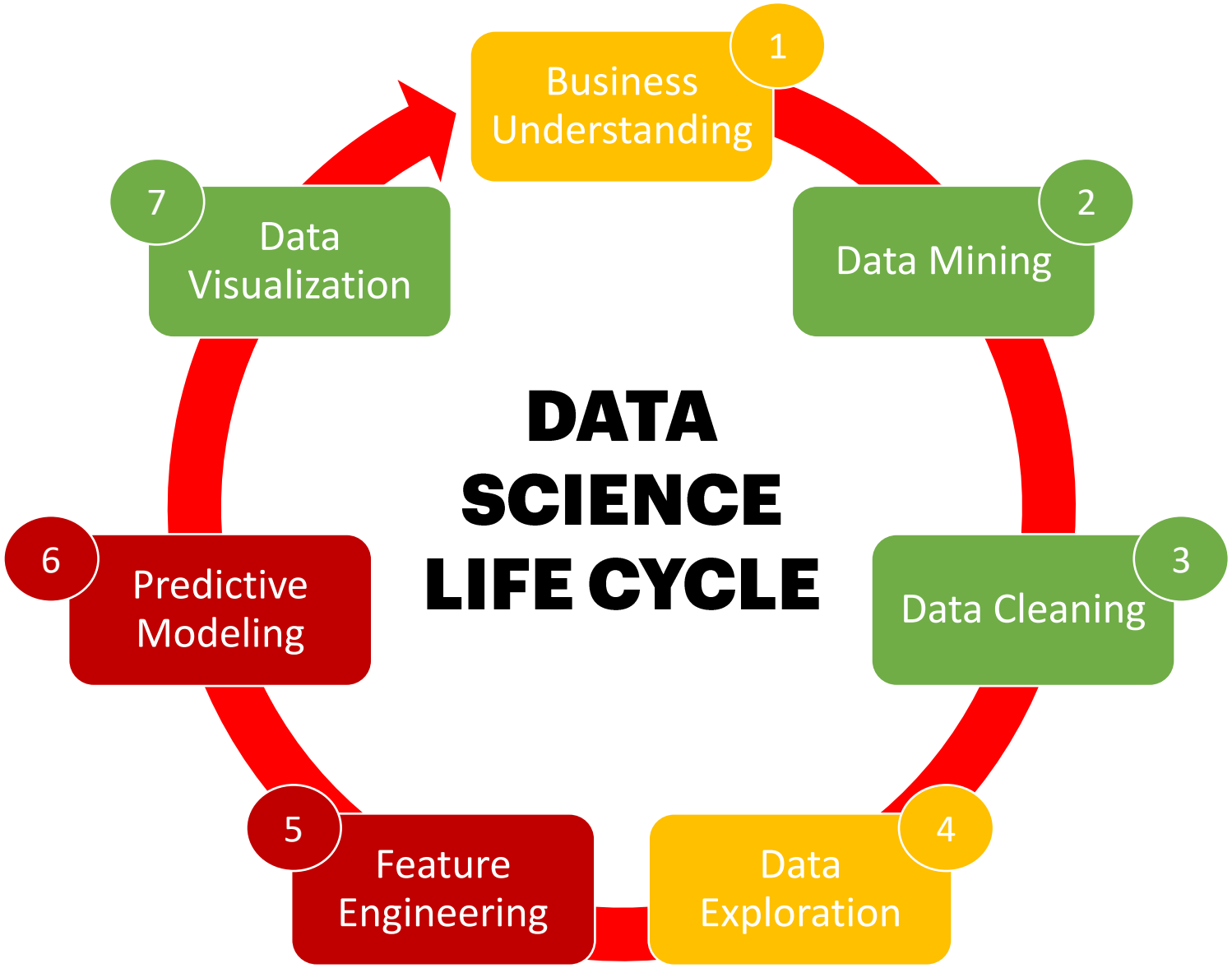
DATA ENGINEERING

COLLABORATION

ML DEVELOPMENT

DATA VISUALIZATION

This is where we derived the intended insights from the model.



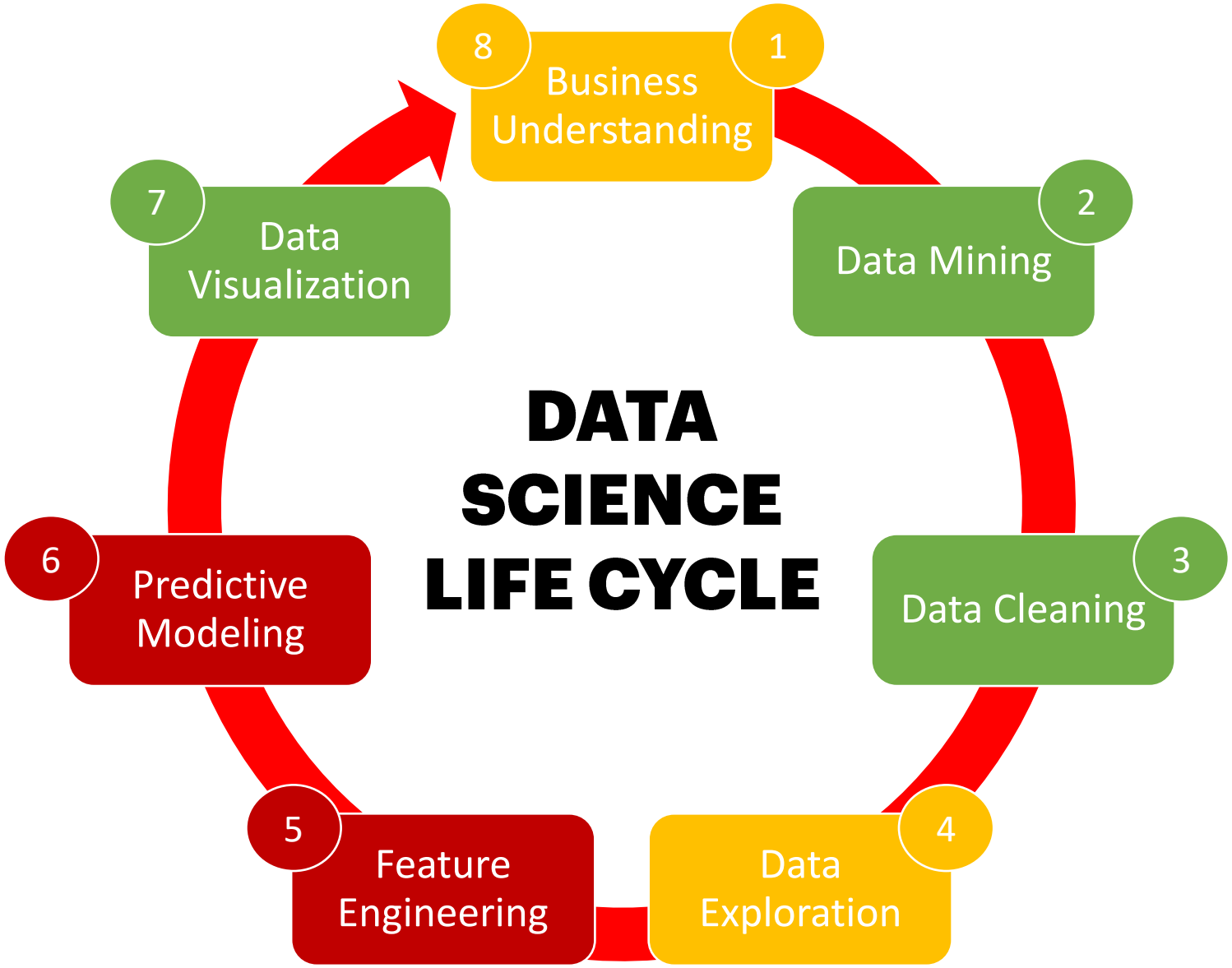
DATA ENGINEERING

COLLABORATION

ML DEVELOPMENT

BUSINESS UNDERSTANDING

This is where we evaluate the success of the model if it relates to the original business understanding.

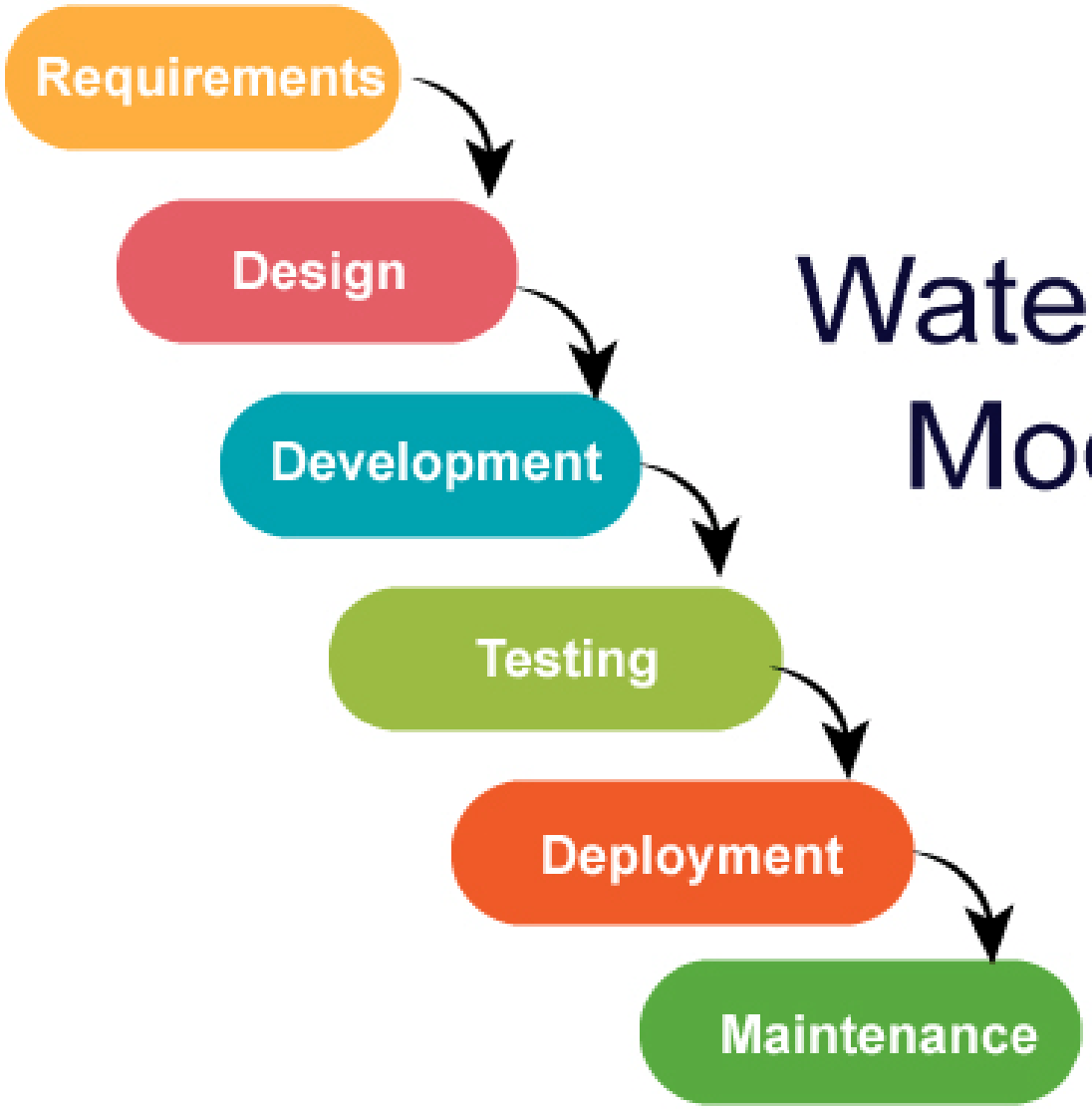


DATA ENGINEERING

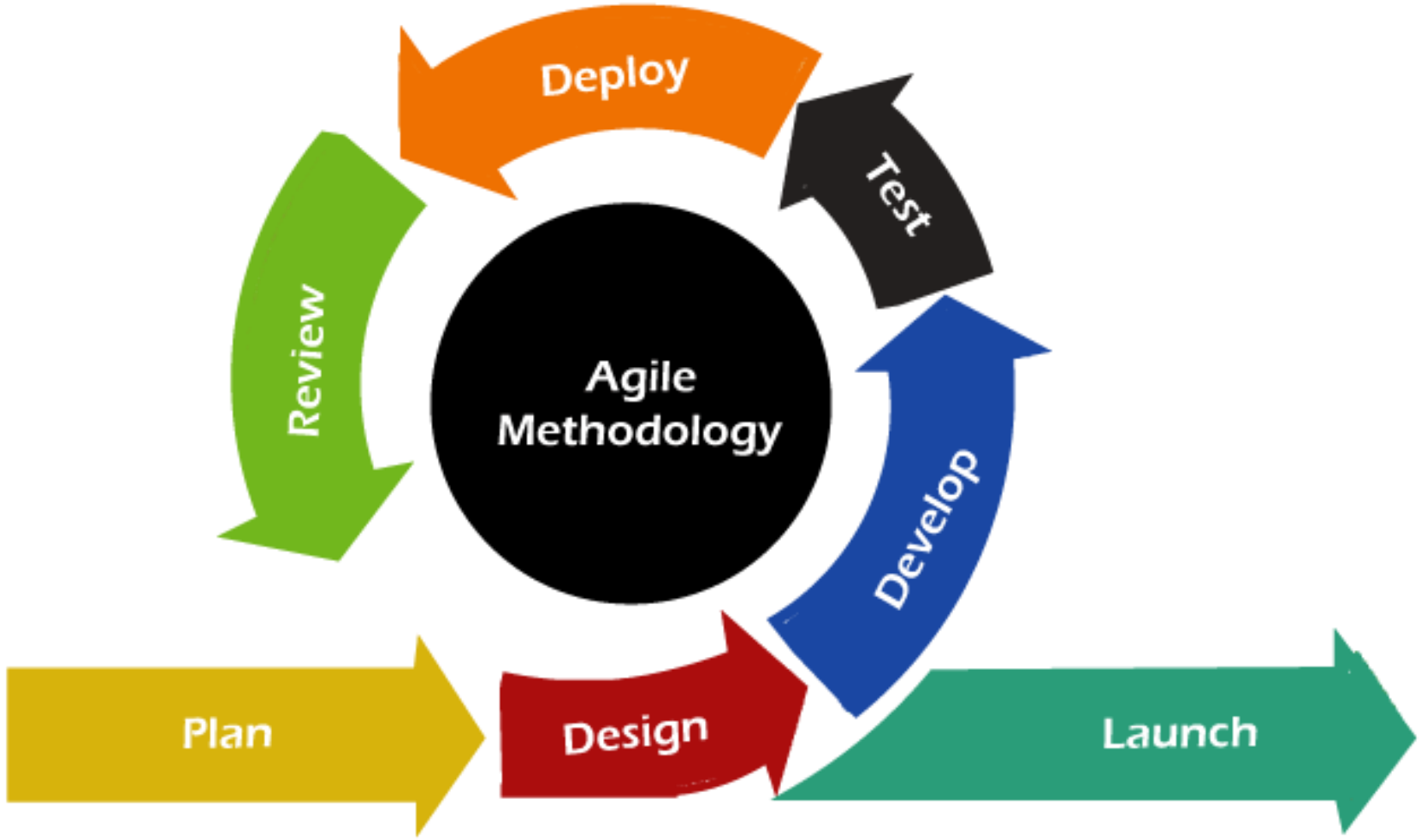
COLLABORATION

ML DEVELOPMENT

PROJECT MANAGEMENT



Waterfall Model



Waterfall Model

1 

2 

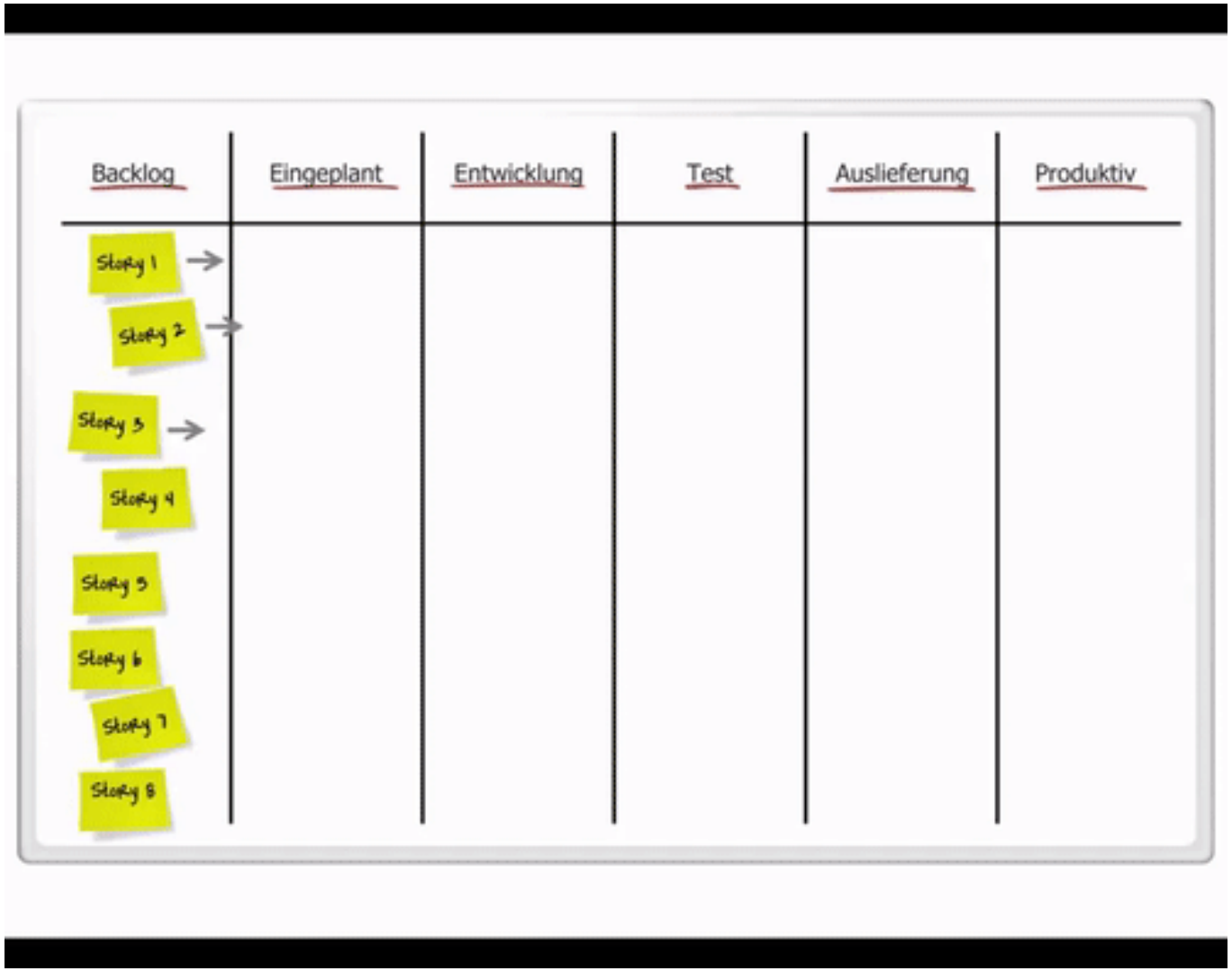
3 

Iterative (Agile methodology)

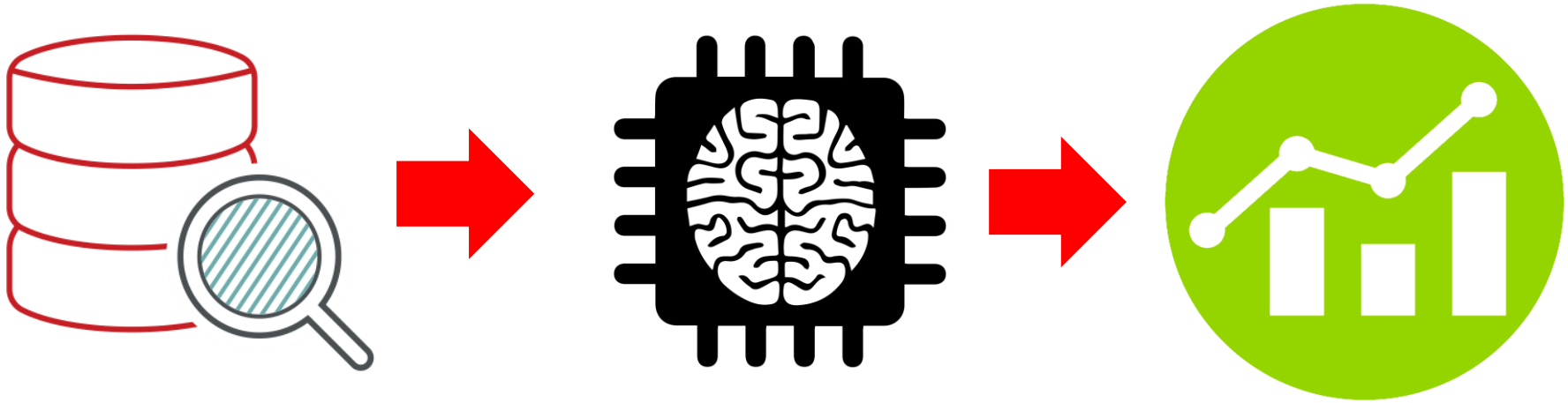
1 

2 

3 



HIGH LEVEL PROCESS



DATA SOURCE

MACHINE LEARNING

DATA CONSUMPTION

HIGH LEVEL PROCESS EXAMPLE

NETFLIX

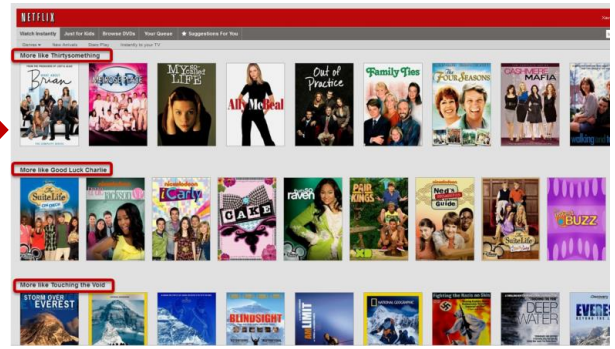
Viewing History	Actors
Ratings	Release Year
Title	Time of Day Watched
Genre	Devices
Categories	How Long You Watch

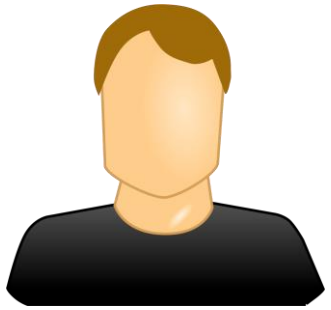
MACHINE LEARNING

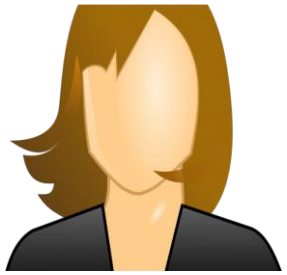
COLLABORATIVE FILTERING

CONTENT-BASED FILTERING

NETFLIX



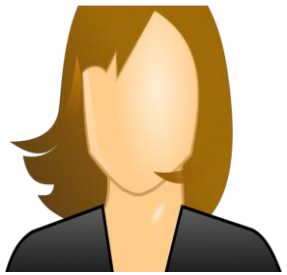




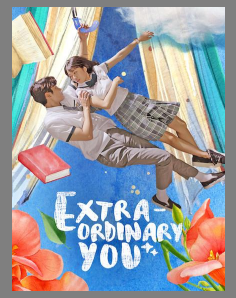
RECOMMENDATION:

COLLABORATIVE FILTERING

WHATIS.TECHTARGET.COM



RECOMMENDATION:



CONTENT-BASED FILTERING

DEVELOPERS.GOOGLE.COM

HIGH LEVEL PROCESS EXAMPLE

NETFLIX

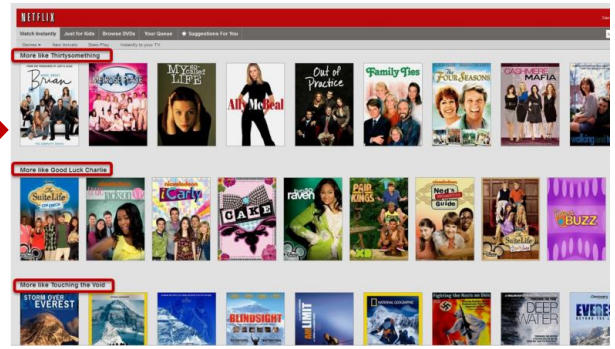
Viewing History	Actors
Ratings	Release Year
Title	Time of Day Watched
Genre	Devices
Categories	How Long You Watch

MACHINE LEARNING

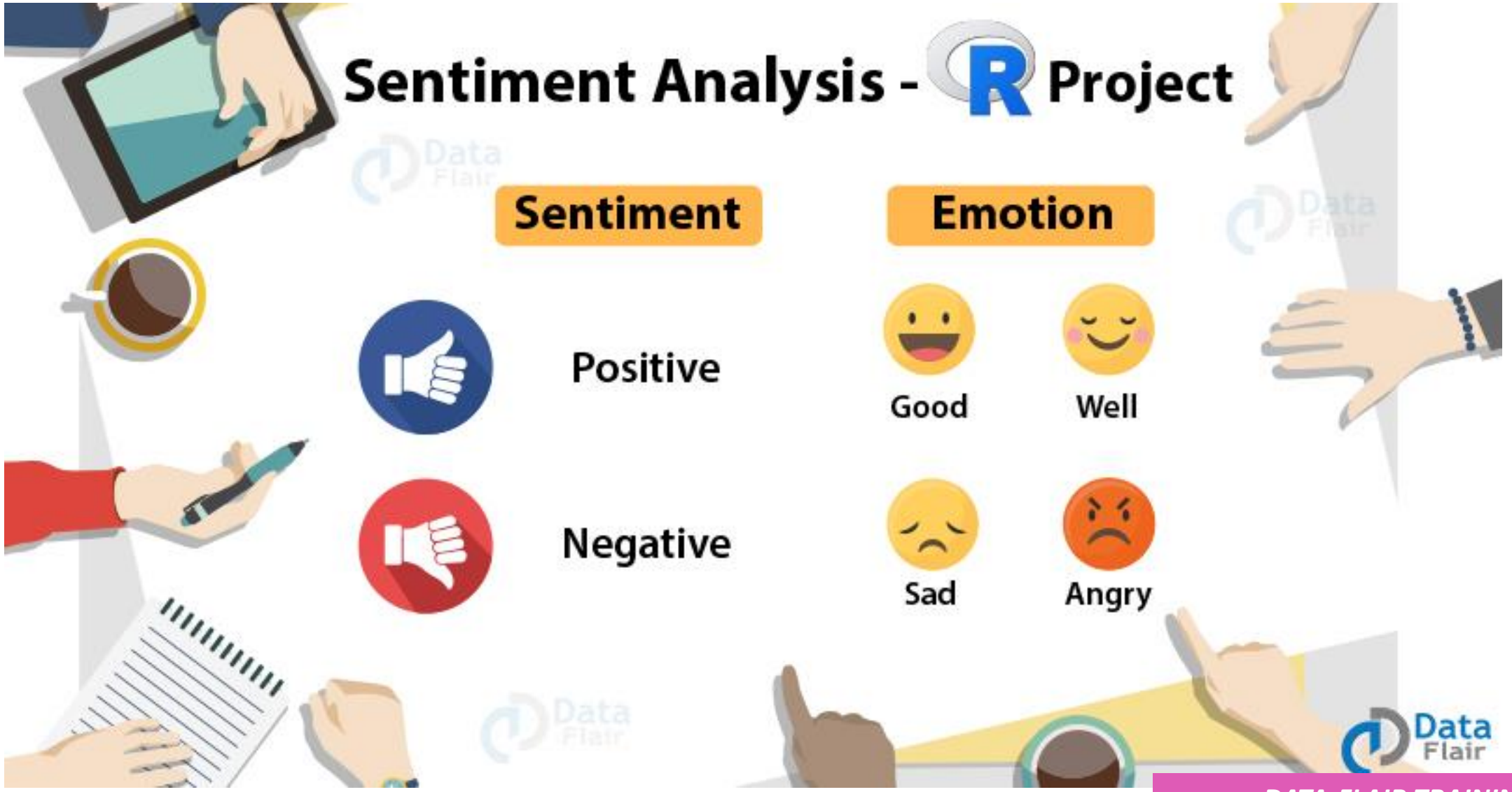
COLLABORATIVE FILTERING

CONTENT-BASED FILTERING

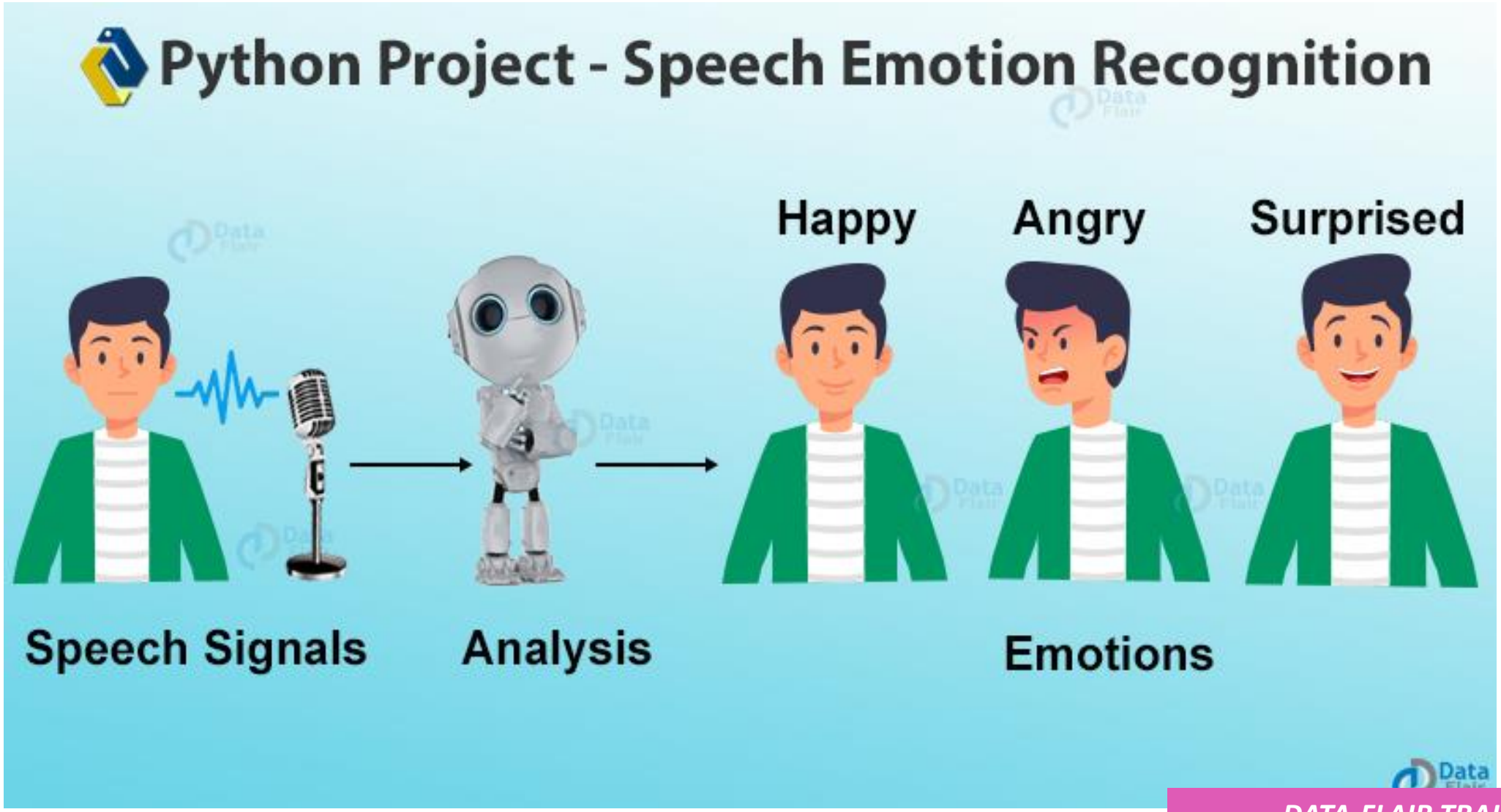
NETFLIX



DATA SCIENCE APPLICATIONS

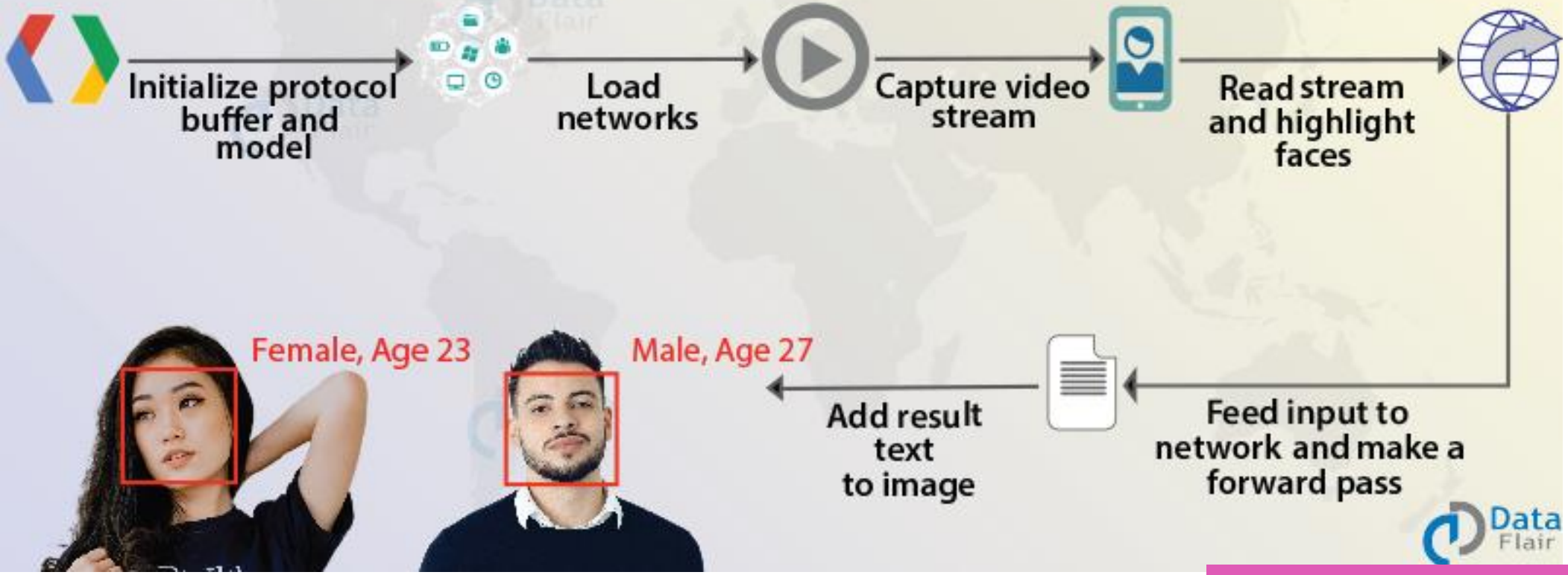


DATA-FLAIR.TRAINING

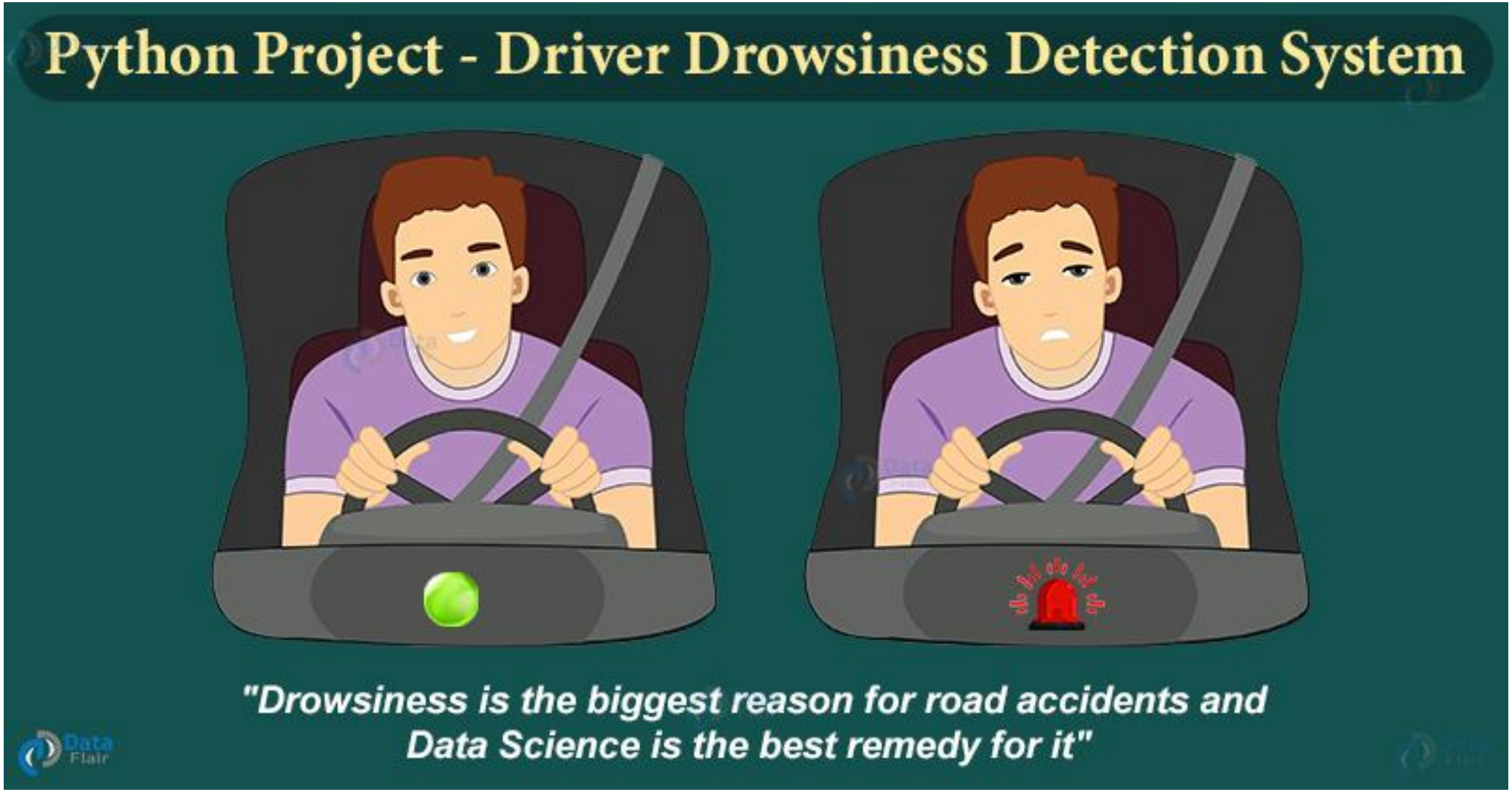


DATA-FLAIR.TRAINING

Python Project - Gender and Age Detection

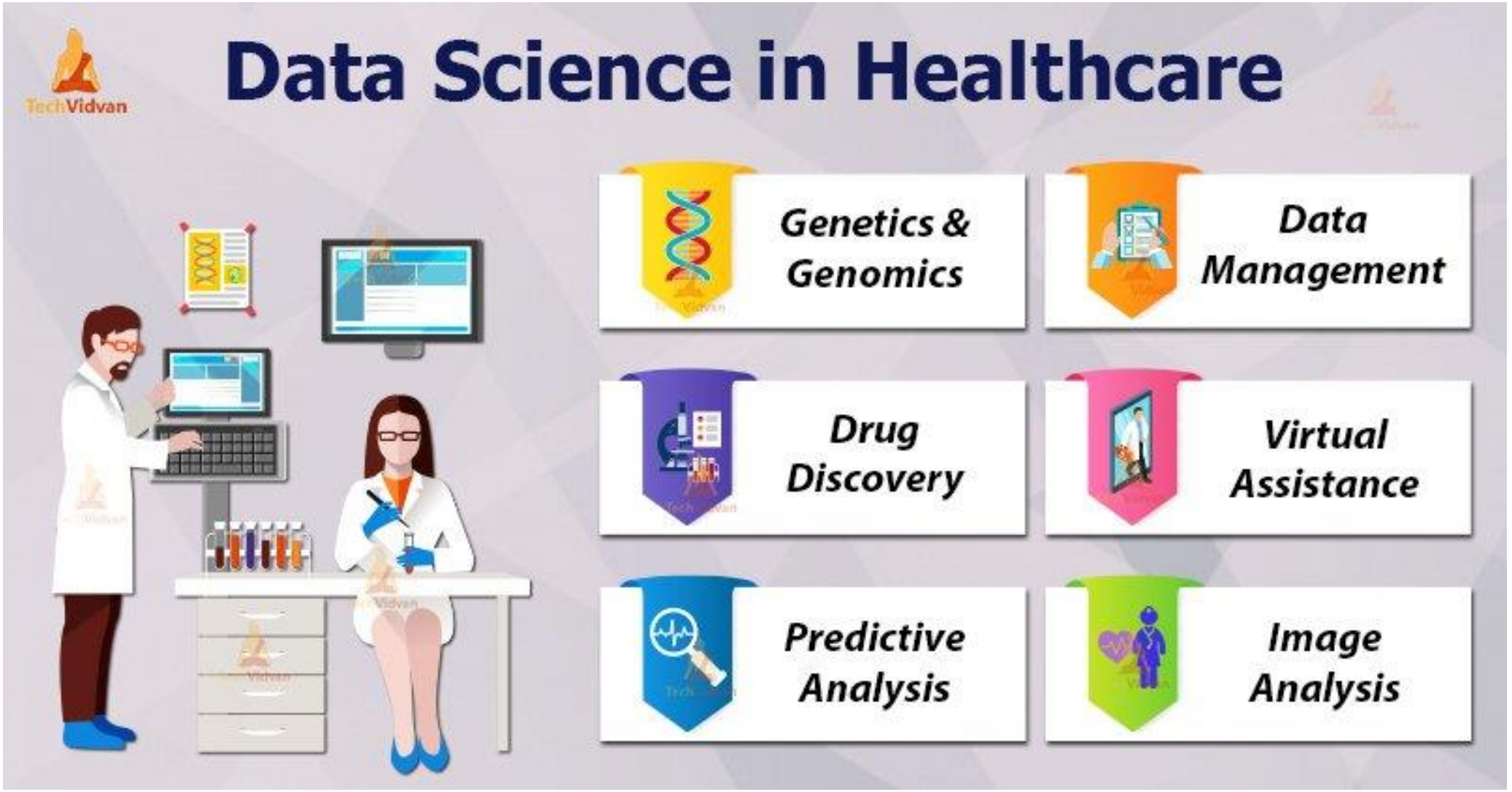


DATA-FLAIR.TRAINING

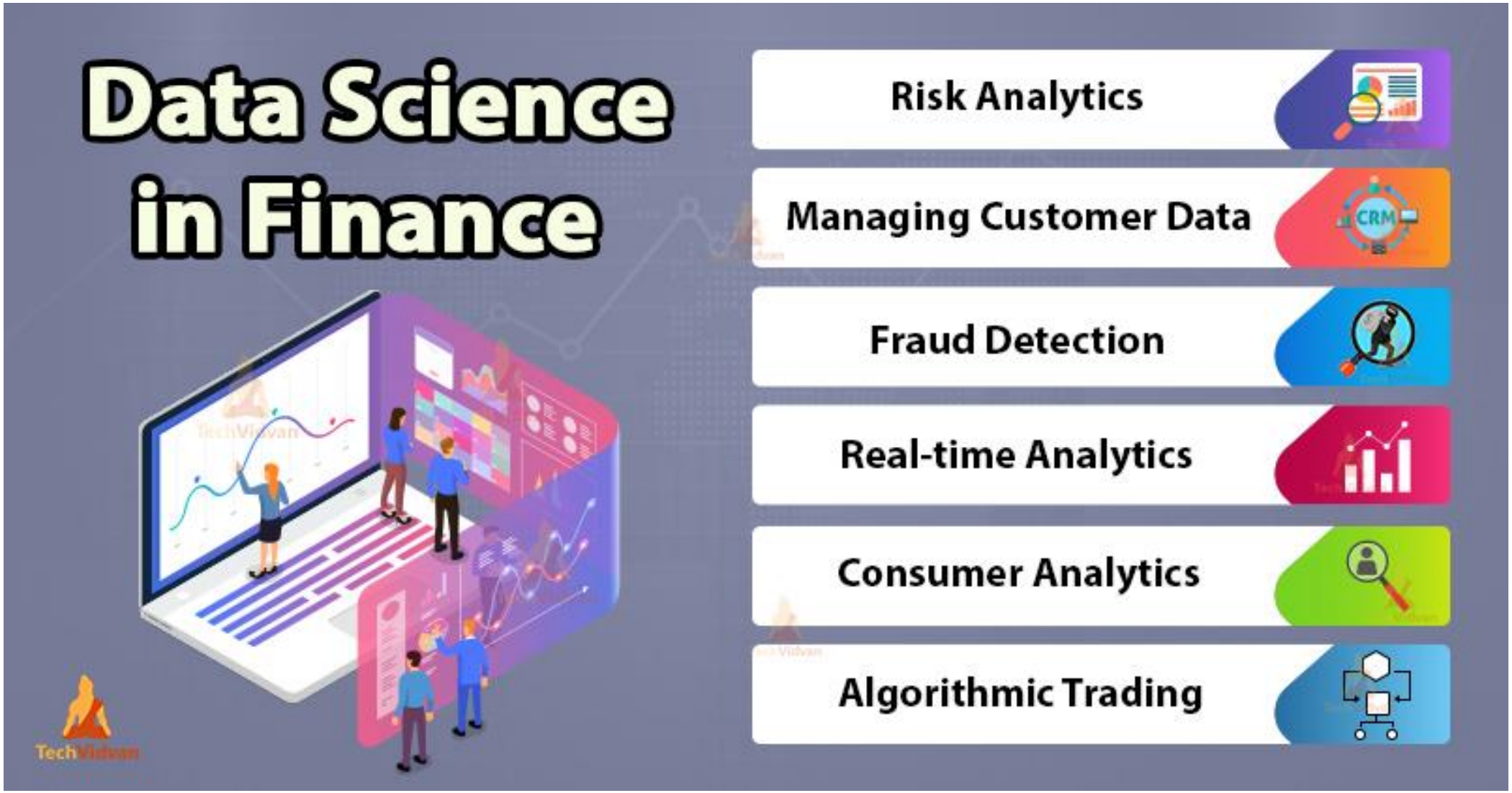


DATA-FLAIR.TRAINING

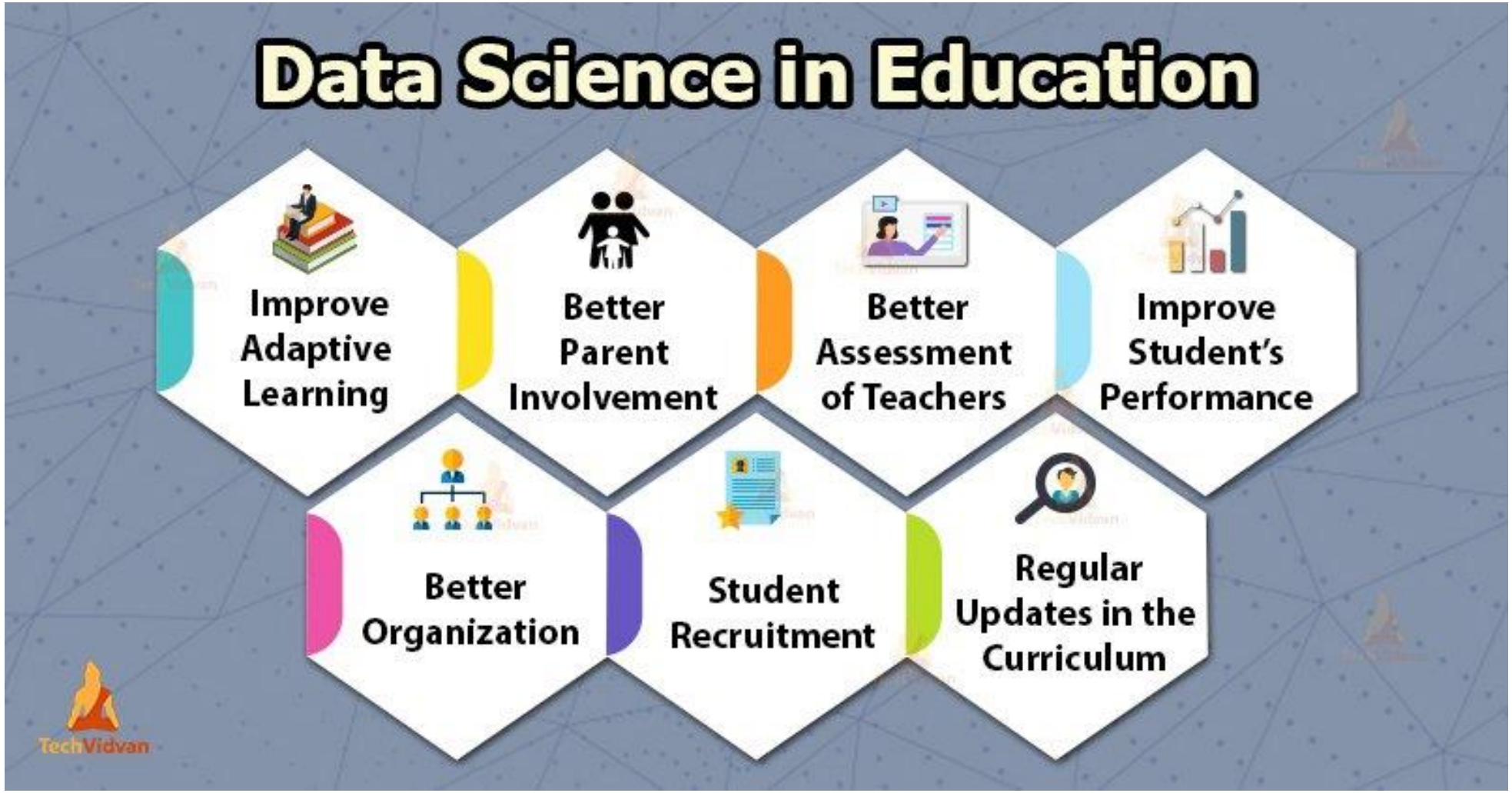
DATA SCIENCE IN DIFFERENT INDUSTRIES



GOOGLE IMAGES: "DATA SCIENCE MEDICINE"

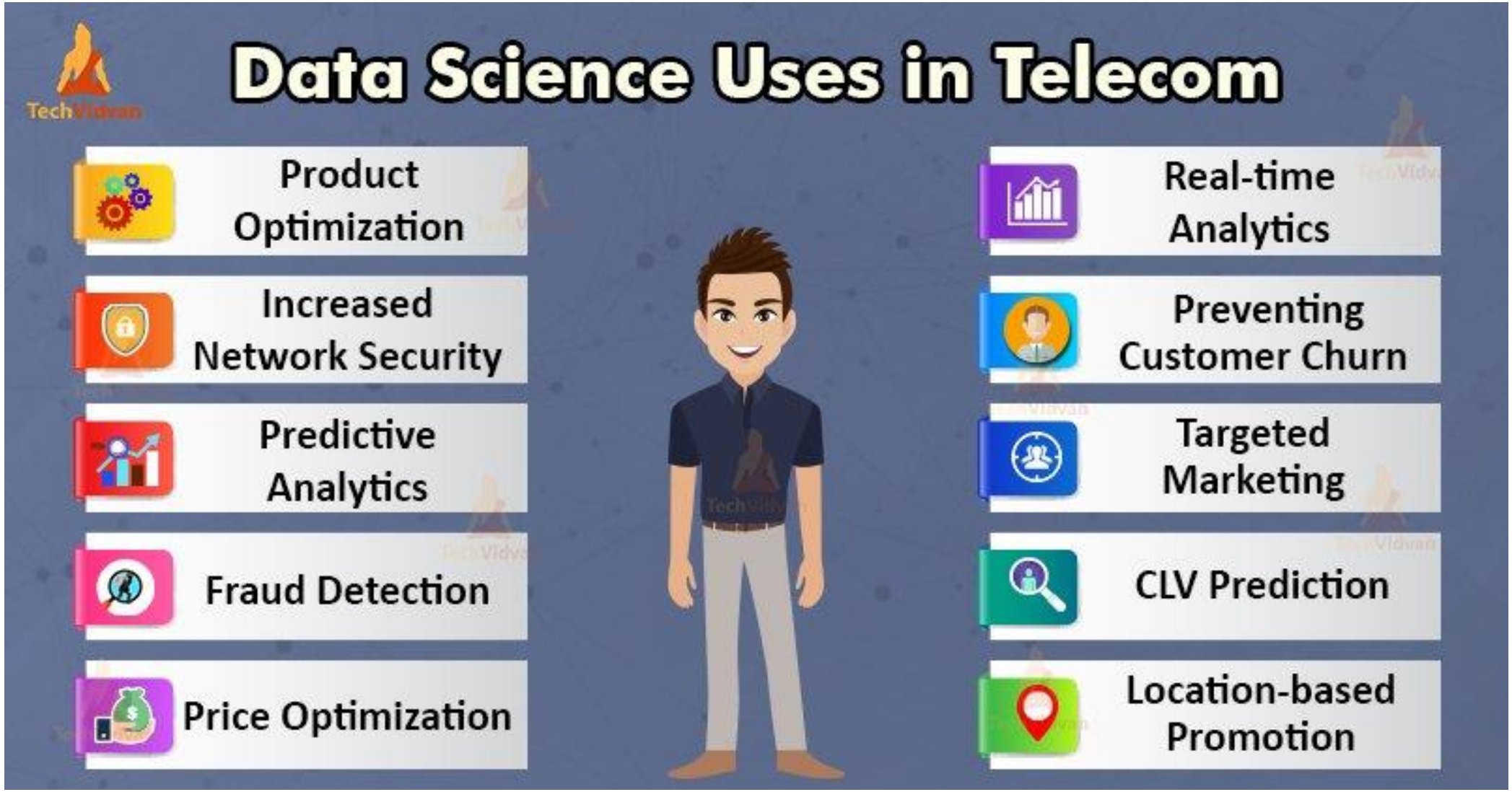


GOOGLE IMAGES: "DATA SCIENCE FINANCE"



GOOGLE IMAGES: "DATA SCIENCE EDUCATION"

GOOGLE IMAGES: "DATA SCIENCE MARKETING"



GOOGLE IMAGES: "DATA SCIENCE TELCO"

SESSION 2 OUTLINE

- DATA SCIENCE METHODOLOGY
- PYTHON SYNTAX

CARE AND SHARE!

DATA SCIENCE BOOTCAMP

Session 2: Data Science Methodology and Python Syntax