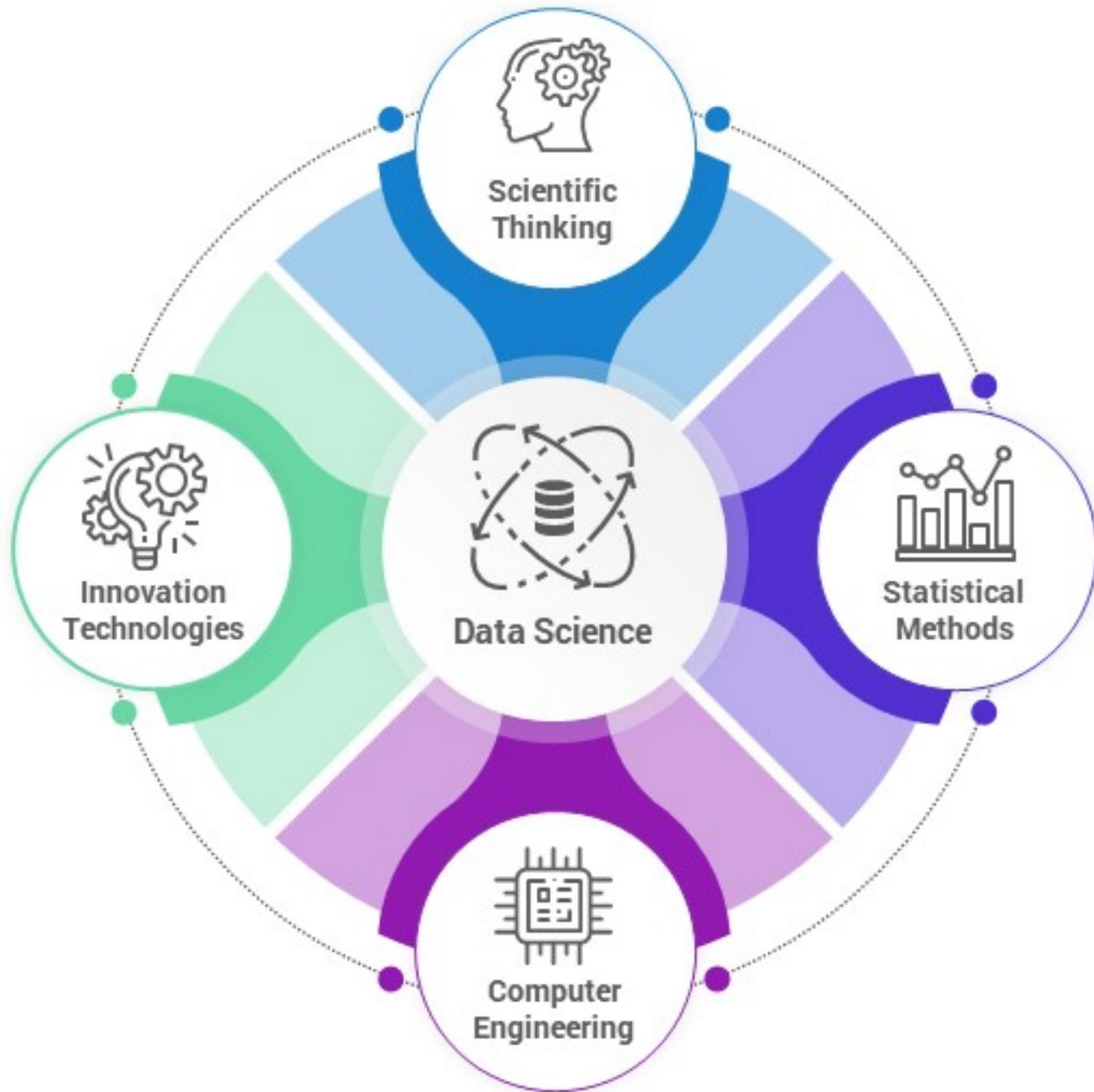


Data Science



What is Data science?

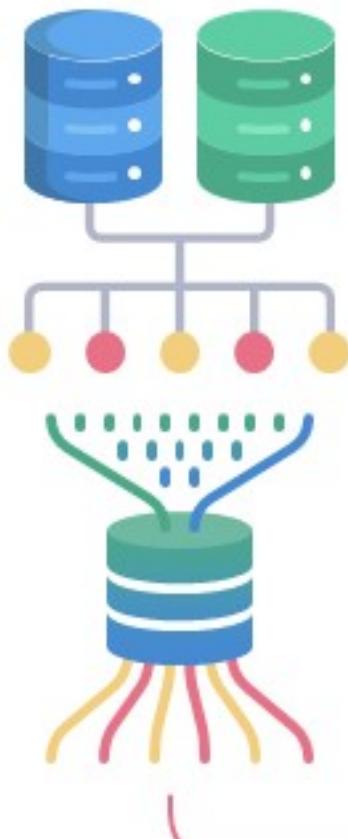
Data science is a multidisciplinary approach that brings together scientific thinking, statistical methods, computer engineering, and innovative technologies to collect, combine, and assess significant amounts of structured and unstructured data in ways that can provide better insights and predictions to help answer business questions and drive strategic decisions.



Why Data Science

Purpose of Data Science

Extracting
Data



Data Analysis &
Processing

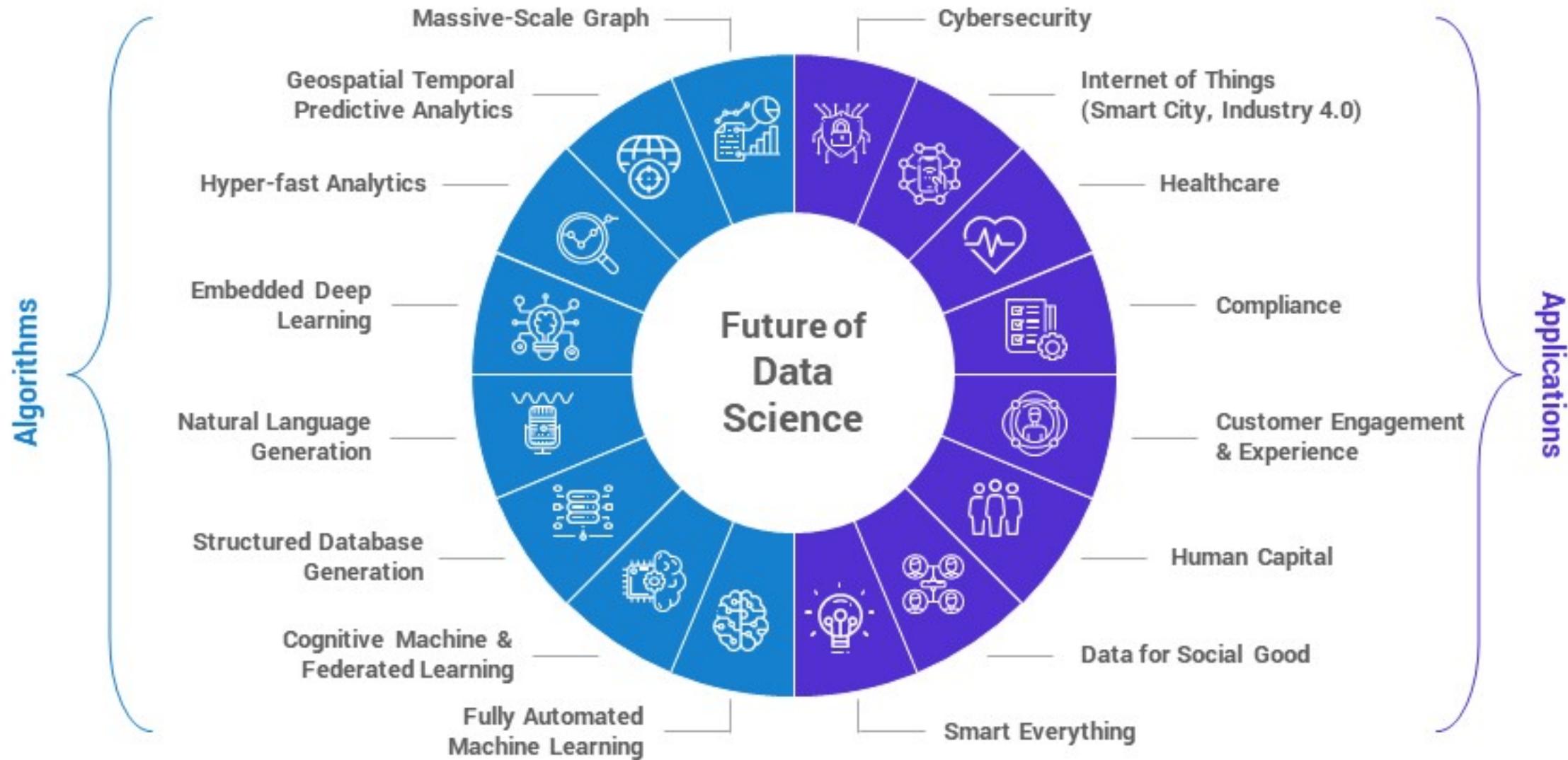


Generating Insights
from Data



Future of Data Science

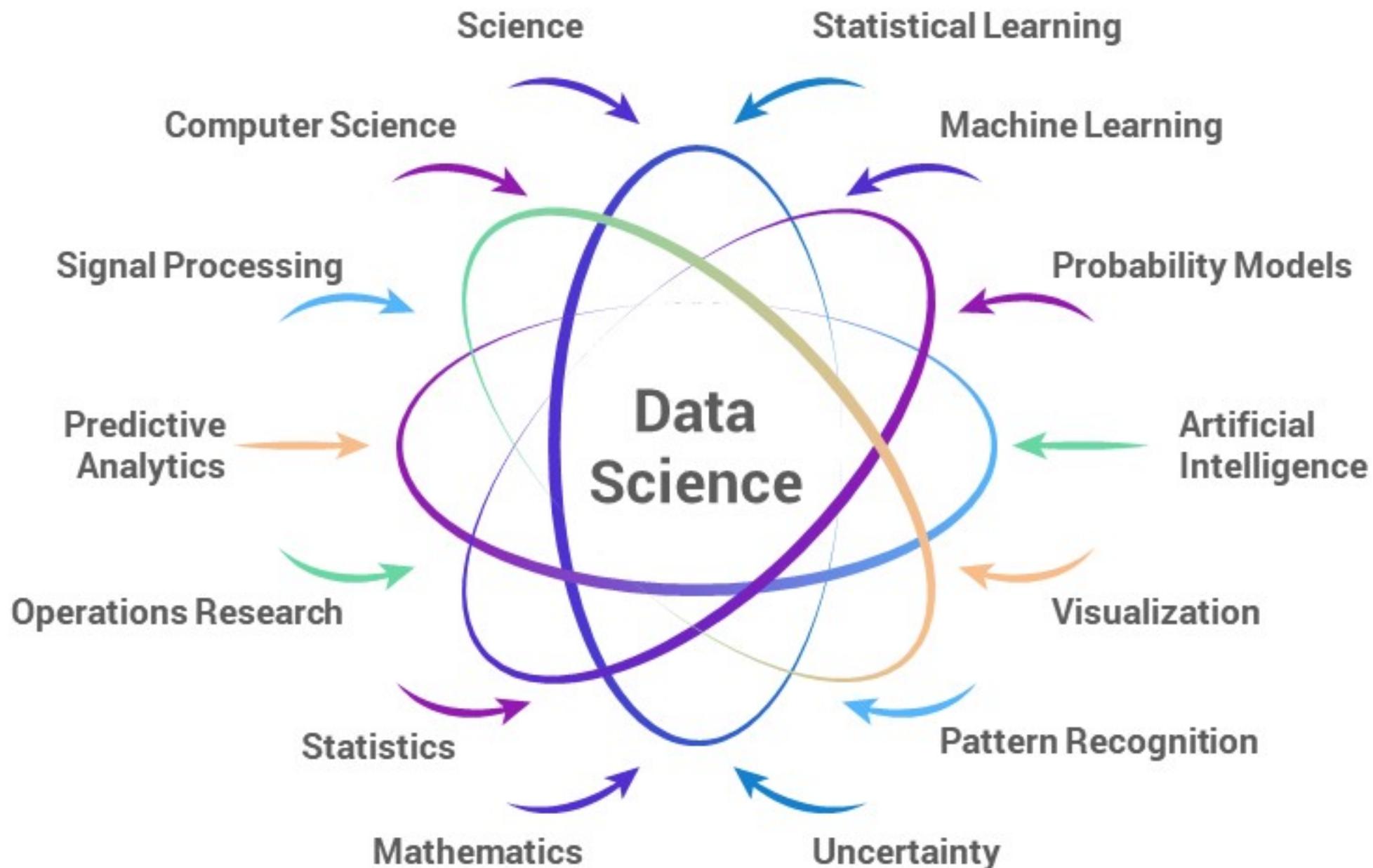
Future of Data Science



Why Learn Data Science?

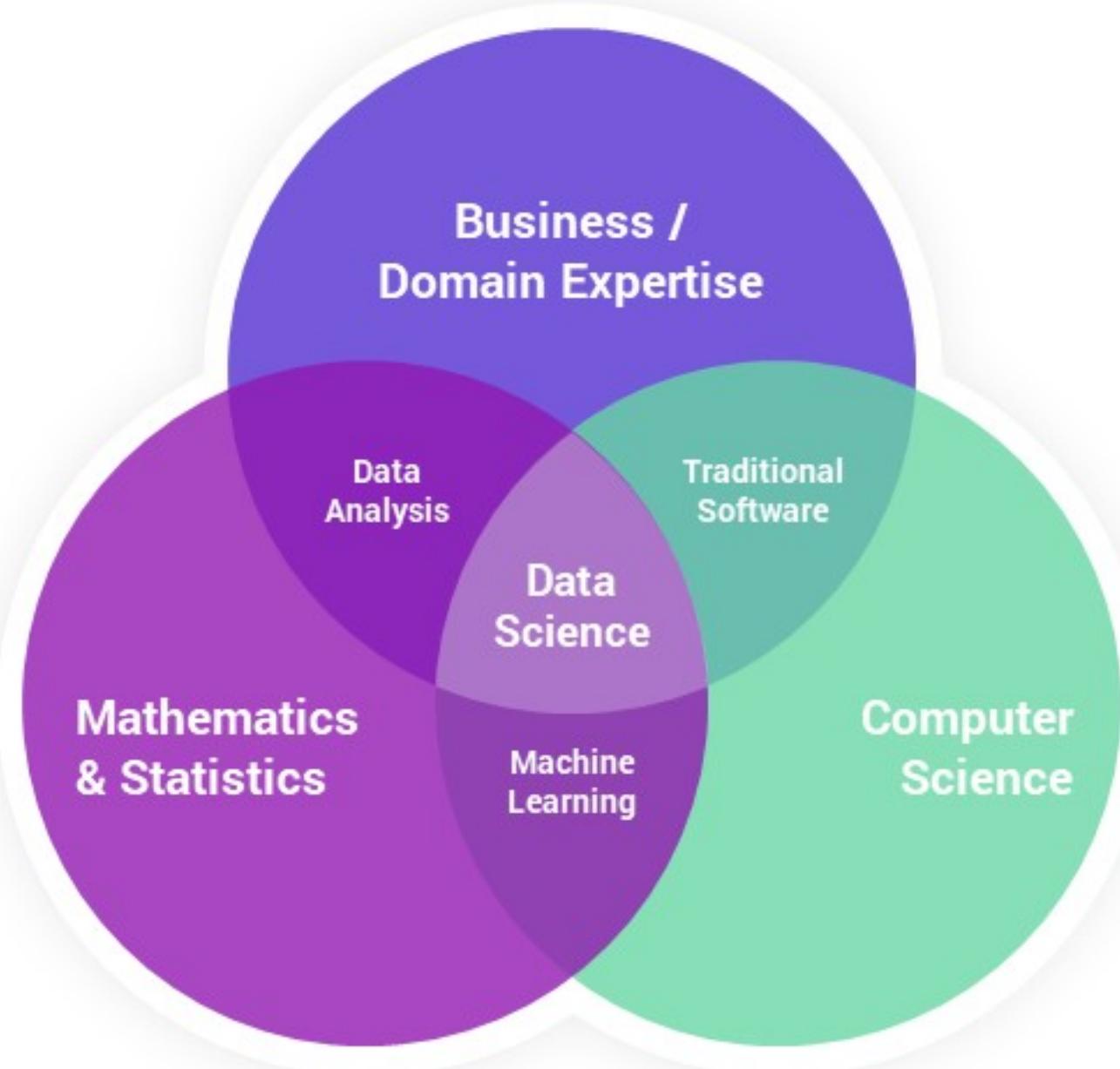
Top 10 Reasons to Learn Data Science in 2021

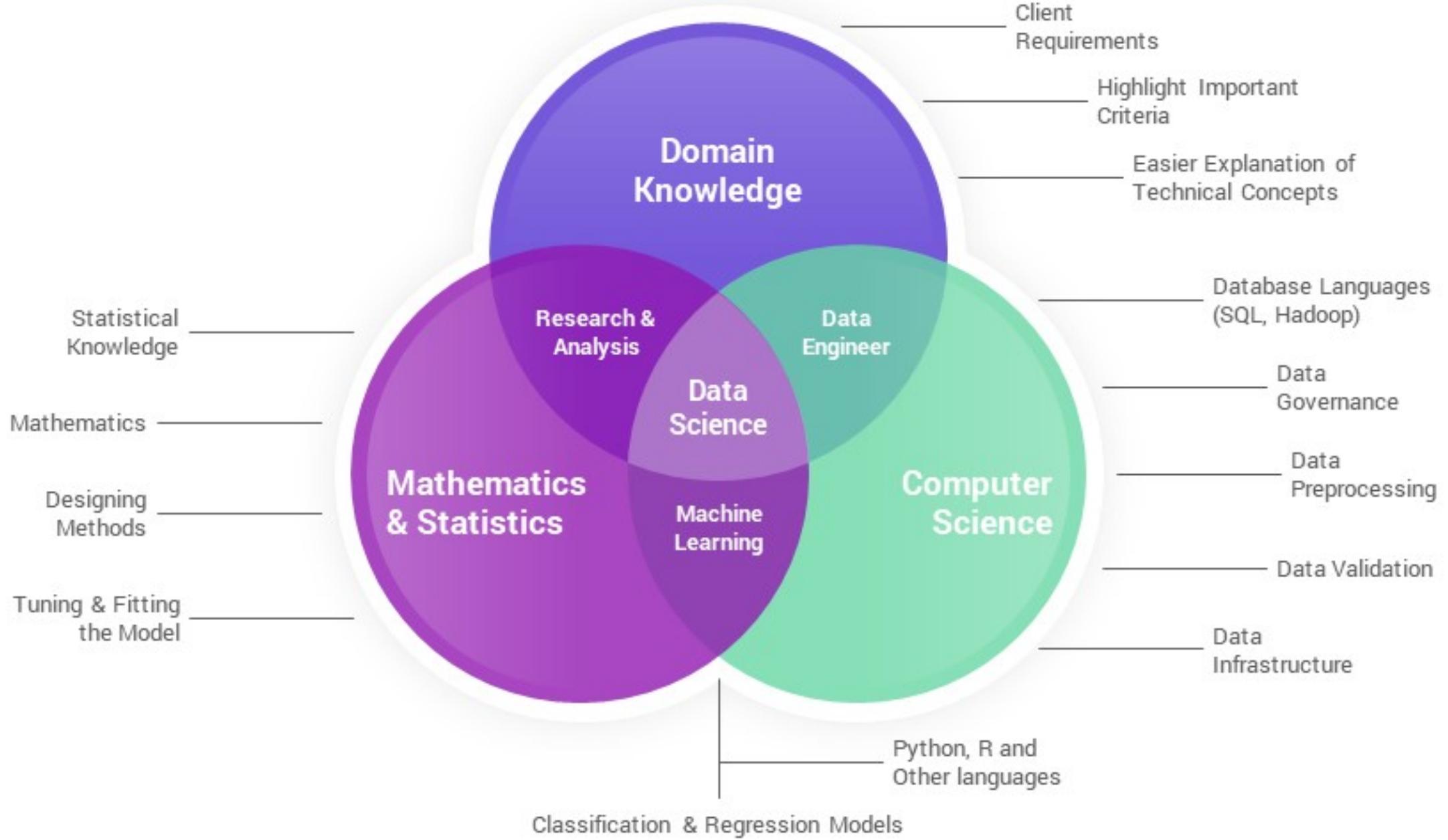




Data Science

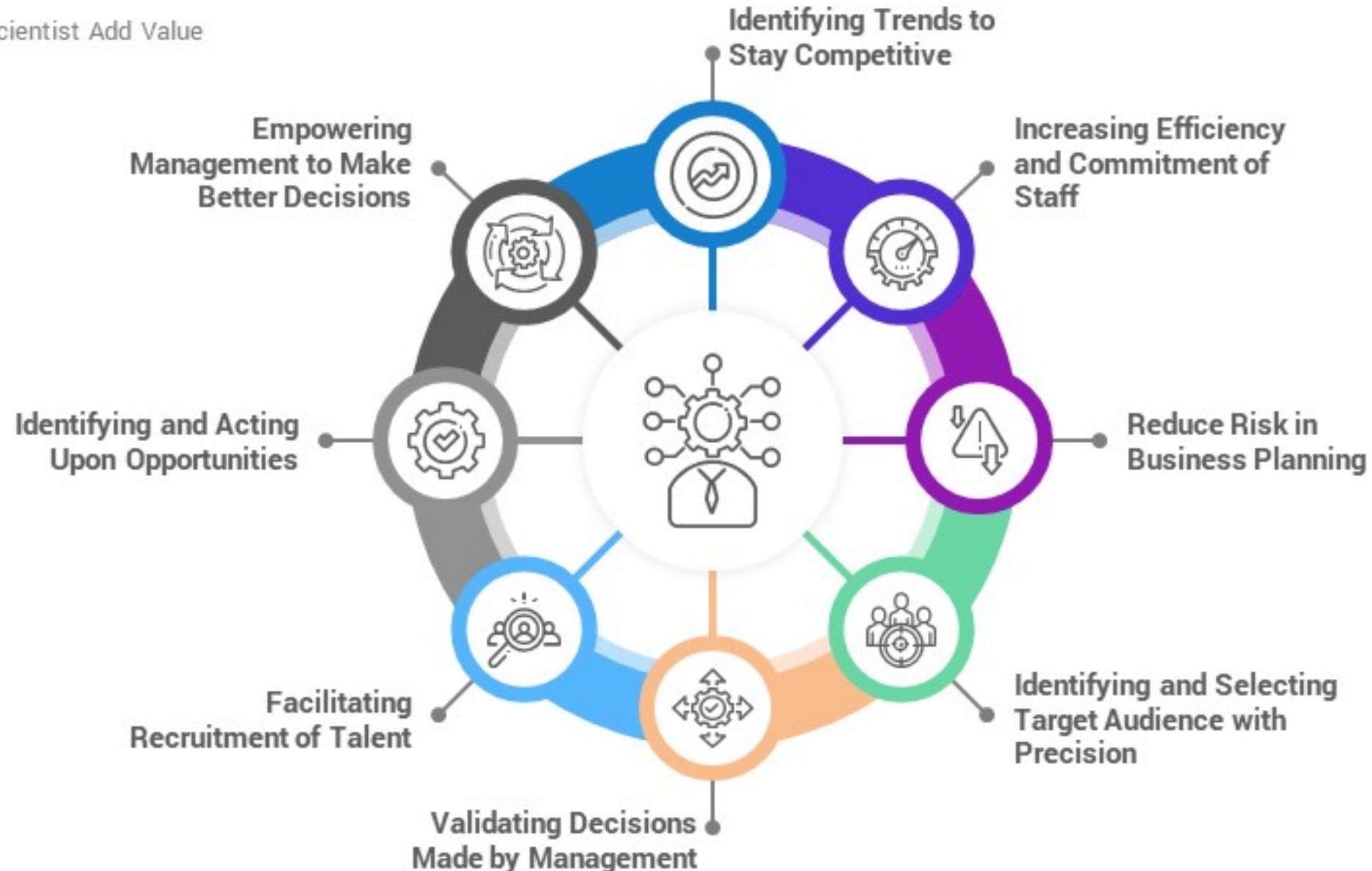
Data Science





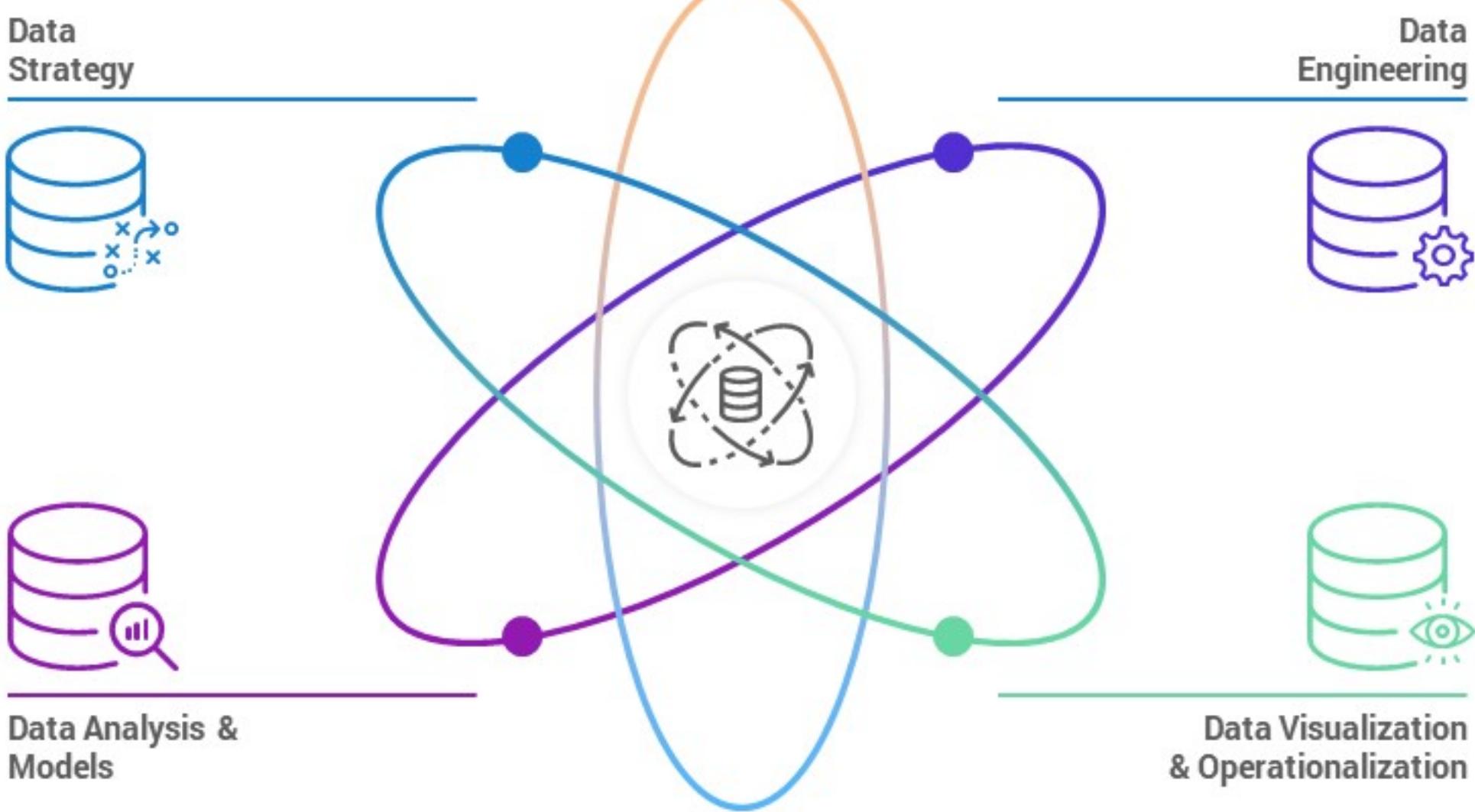
How Data Scientists Add Value

How Data Scientist Add Value



The Four Components of Data Science

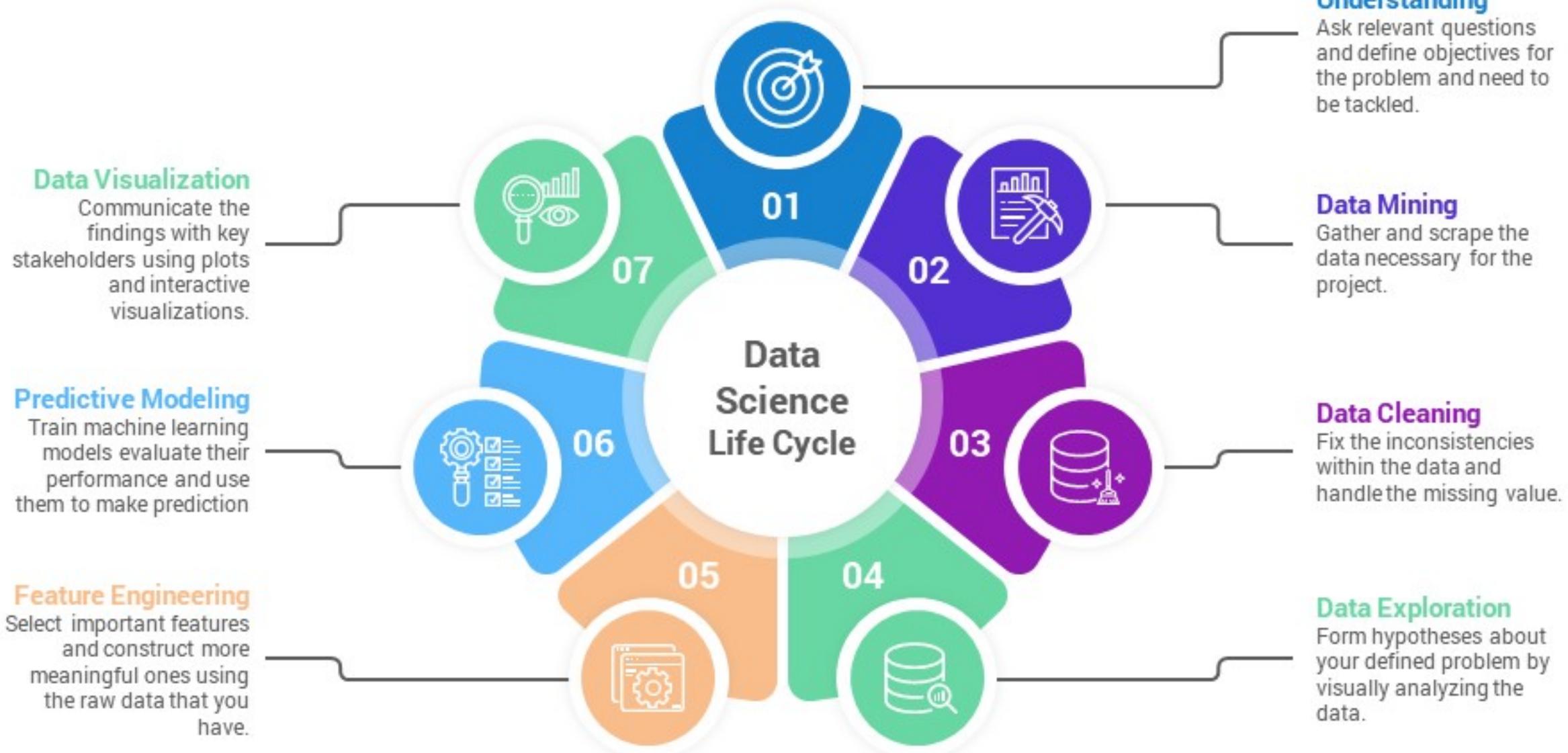
The Four Components of Data Science



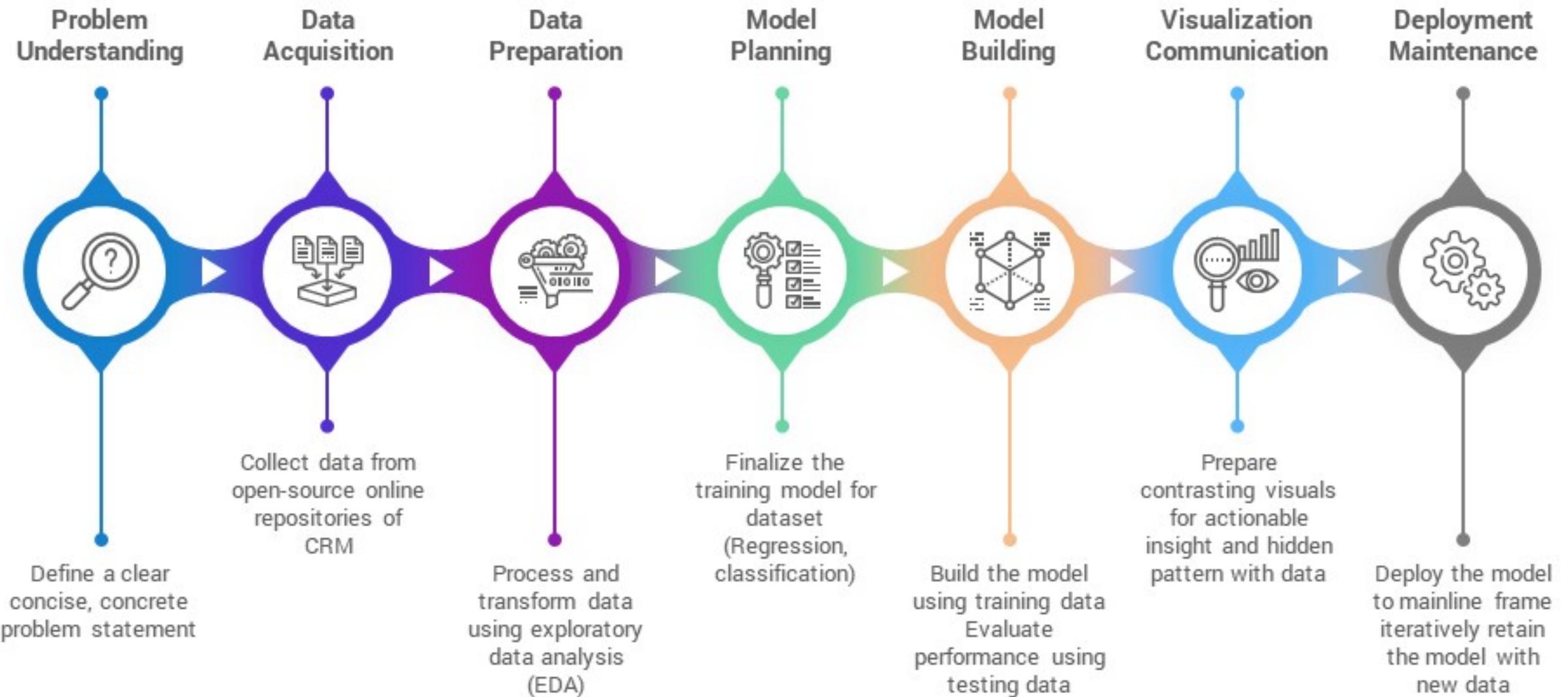
Role of Data Science in Decision Making



Data Science Life Cycle

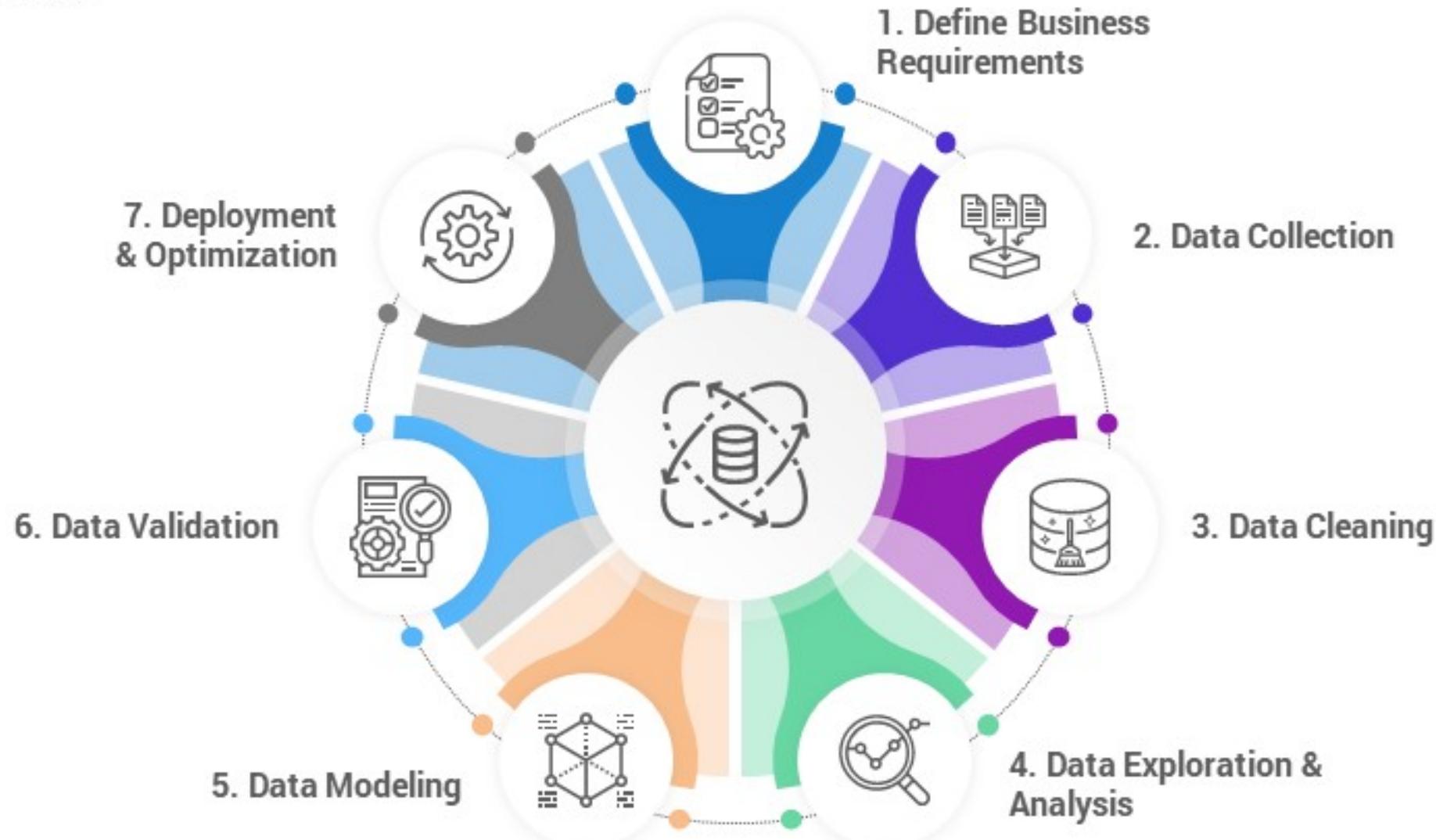


Data Science Process



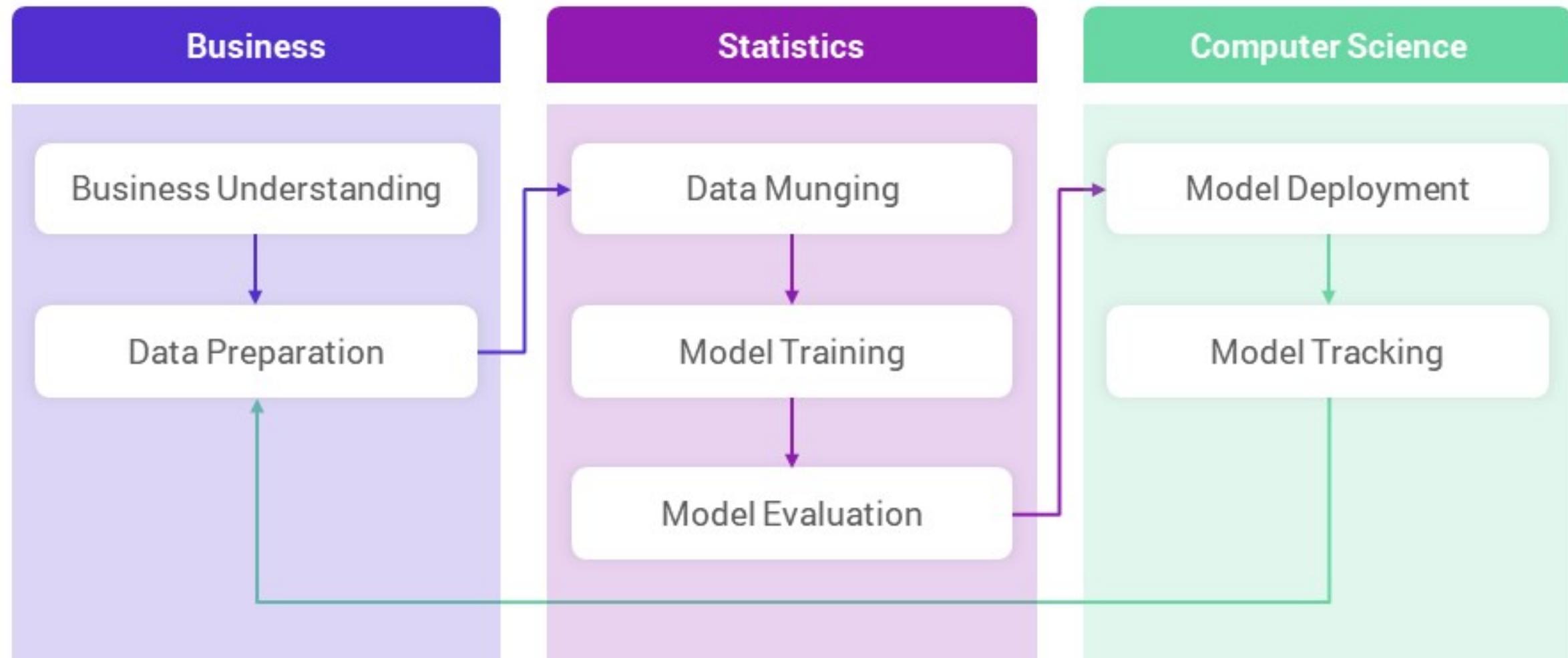
Data Science Process

Data Science Process

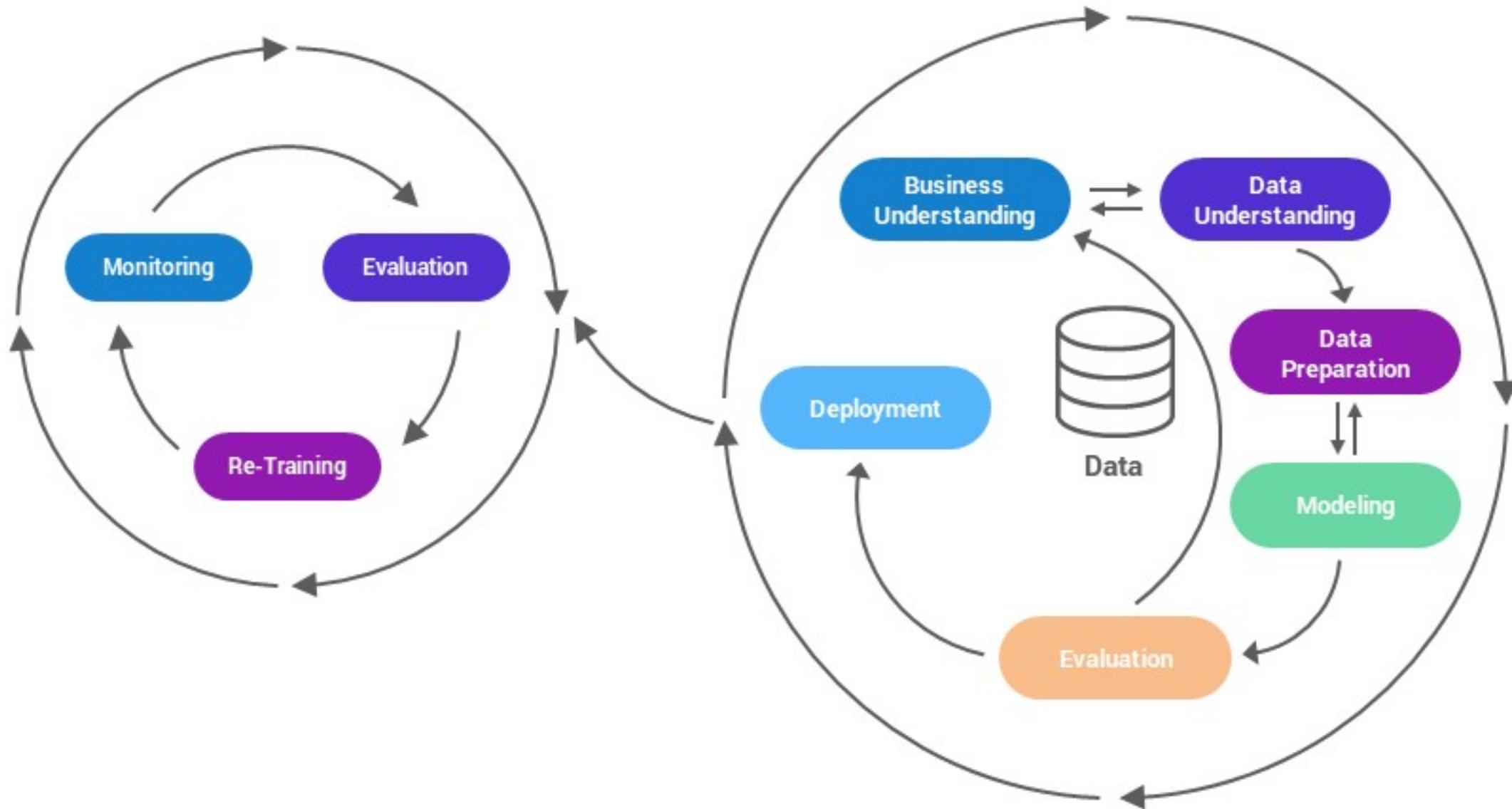


The Data Science Process

The Data Science Process



The Data Science Process (CRISP-DM)





Data Scientist Skill Set

Data Scientist Skill Set



Technical Skills

- Coding skills
- Handling skills
- Computational tools
- Basics software development
- Big data
- High performance computing
- Parallel computing

Tech-Savvy



Analytical Skills

- Advanced statistics & inference
- Modeling & simulation
- Machine learning
- Computer science
- Advanced math
- Data visualization
- Experiment design
- Research expertise

Scientific problem solving



Business Skills

- Evaluate and development of business cases
- Project management
- Business processes
- Change management
- Communication skills
- Leadership skills

Appetite for business problems

Data Scientist Skill Set

Data Scientist Skill Set



Math & Statistics

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants



Programming & Database

- Computer science fundamentals
- Scripting language e.g., Python
- Statistical computing package, e.g., R
- Database algebra
- Parallel database and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Customer reducers
- Experience with xaaS like AWS



Domain Knowledge & Soft Skills

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative innovative and collaborative

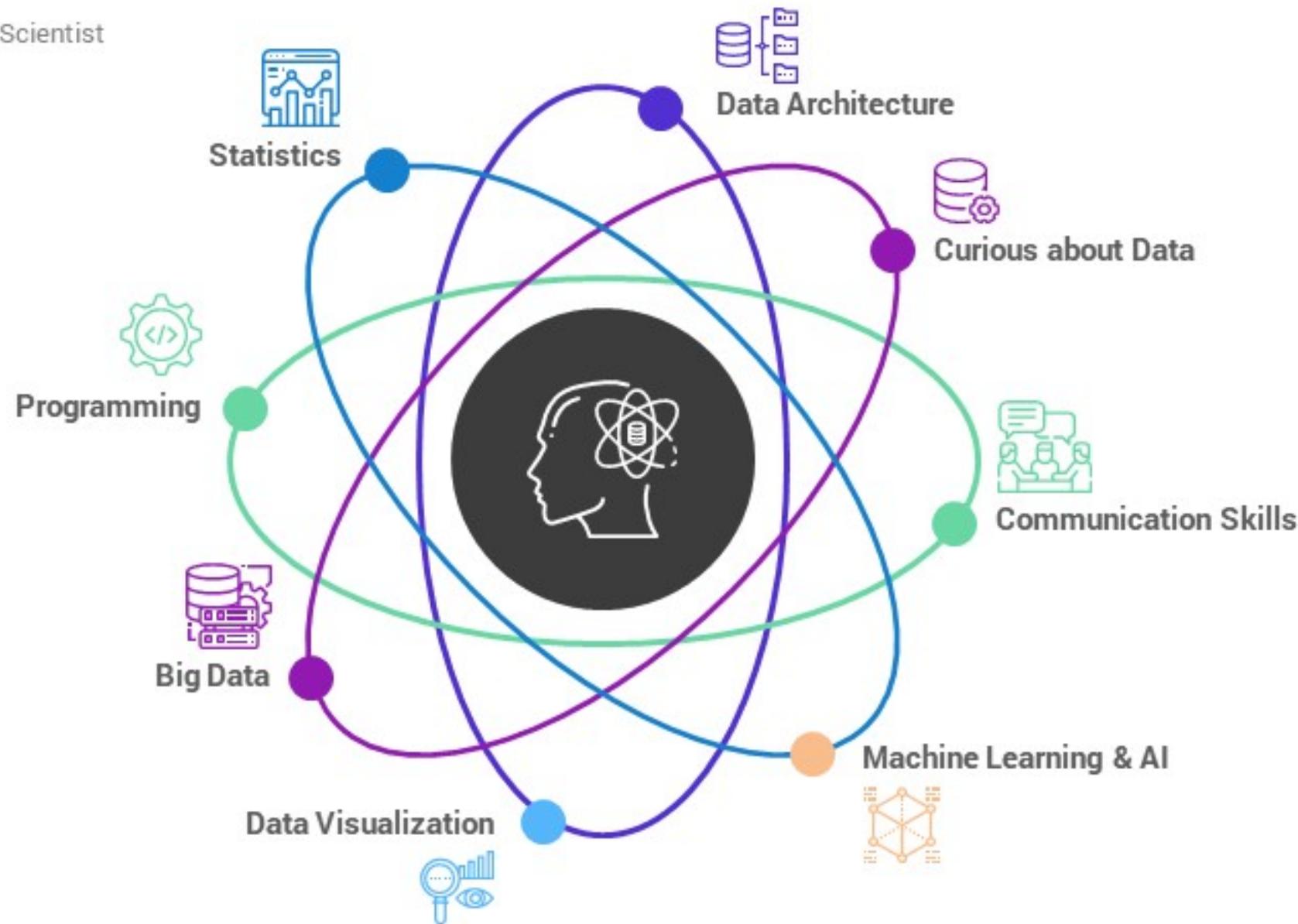


Communication & Visualization

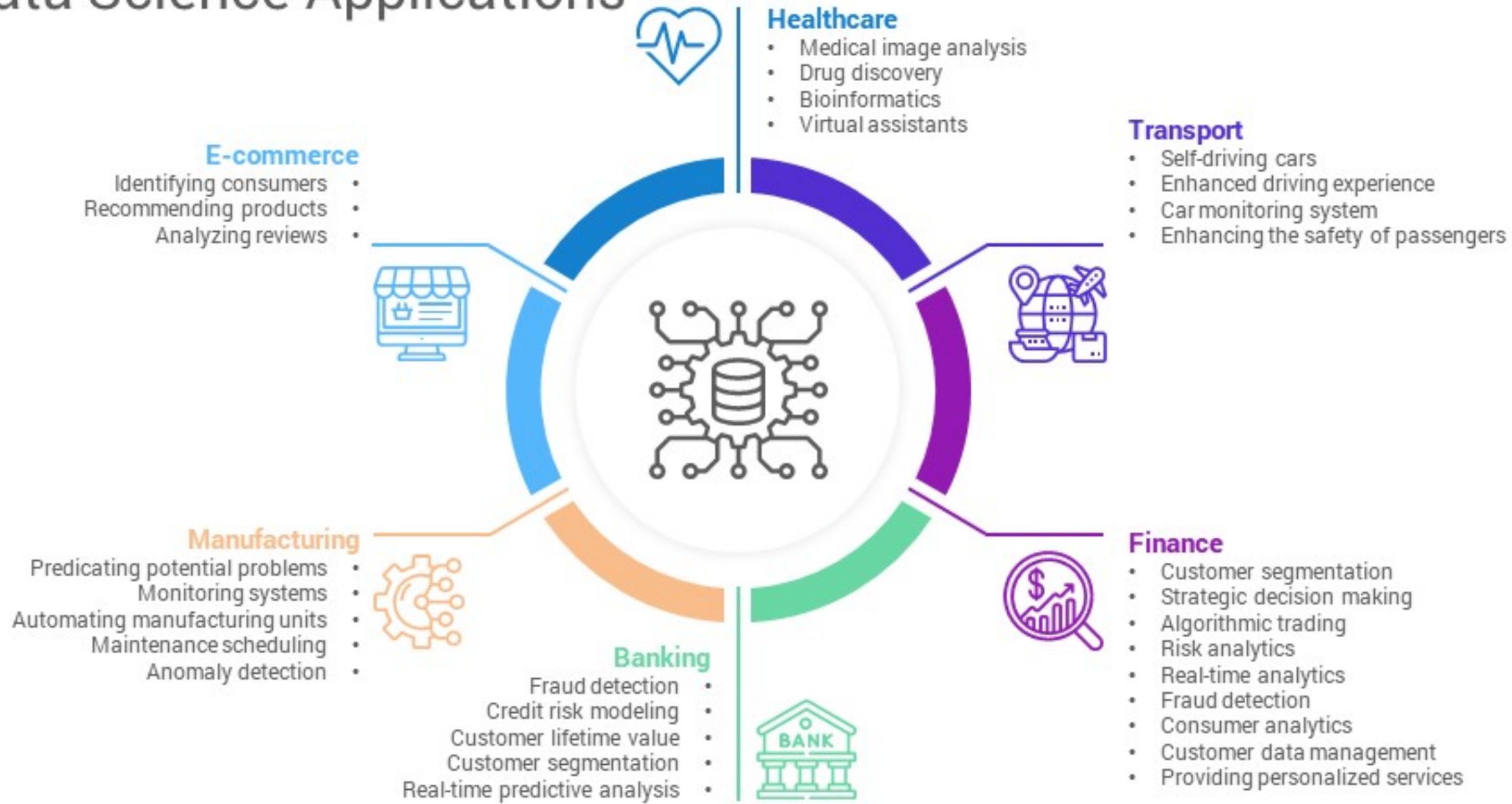
- Able to engage with senior management
- Storytelling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot2 or lattice
- Knowledge of any of visualization tools, e.g., Flare, D3.js, Tableau

Skills Required For Data Scientist

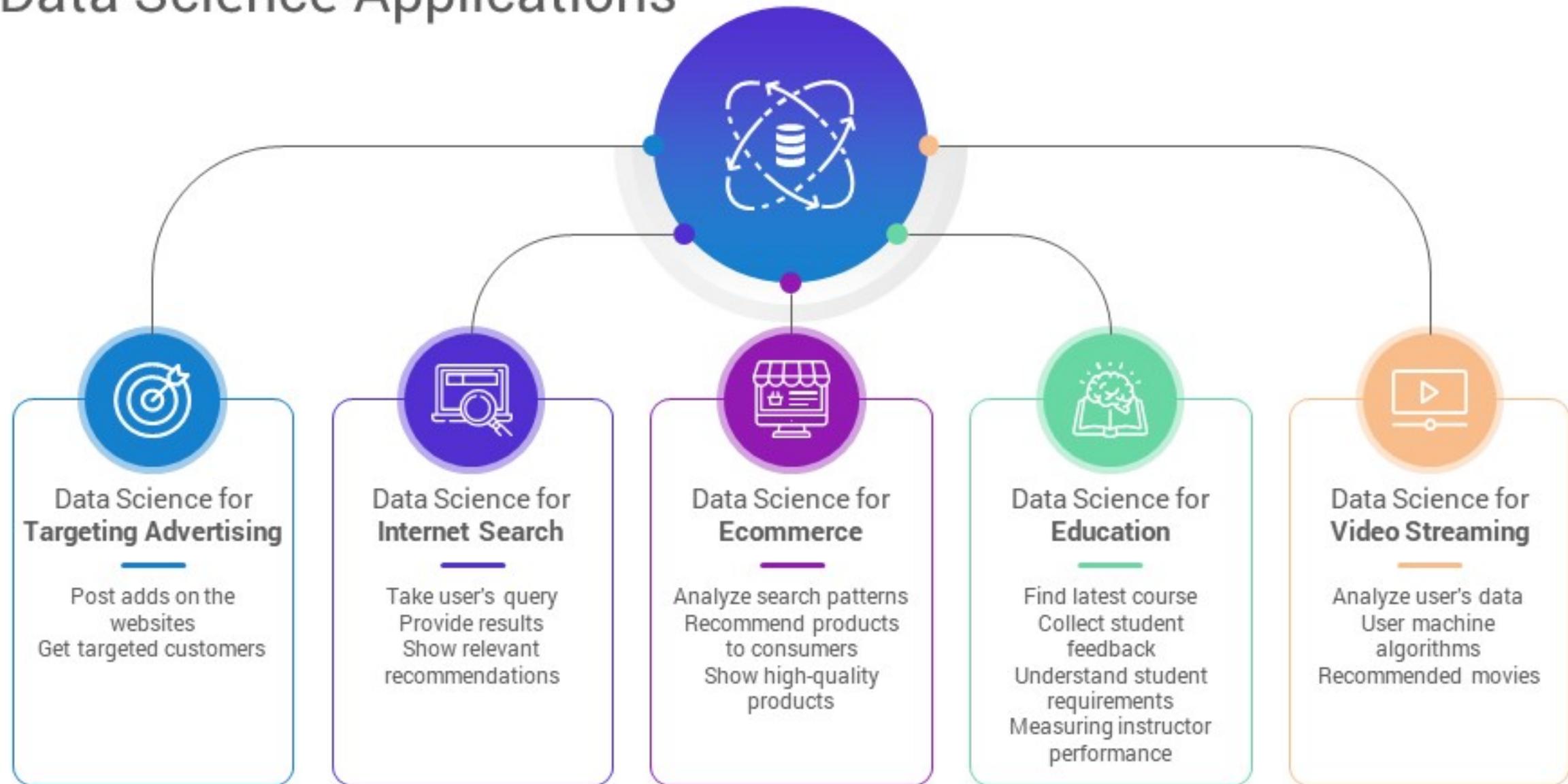
Skills Required For Data Scientist



Data Science Applications

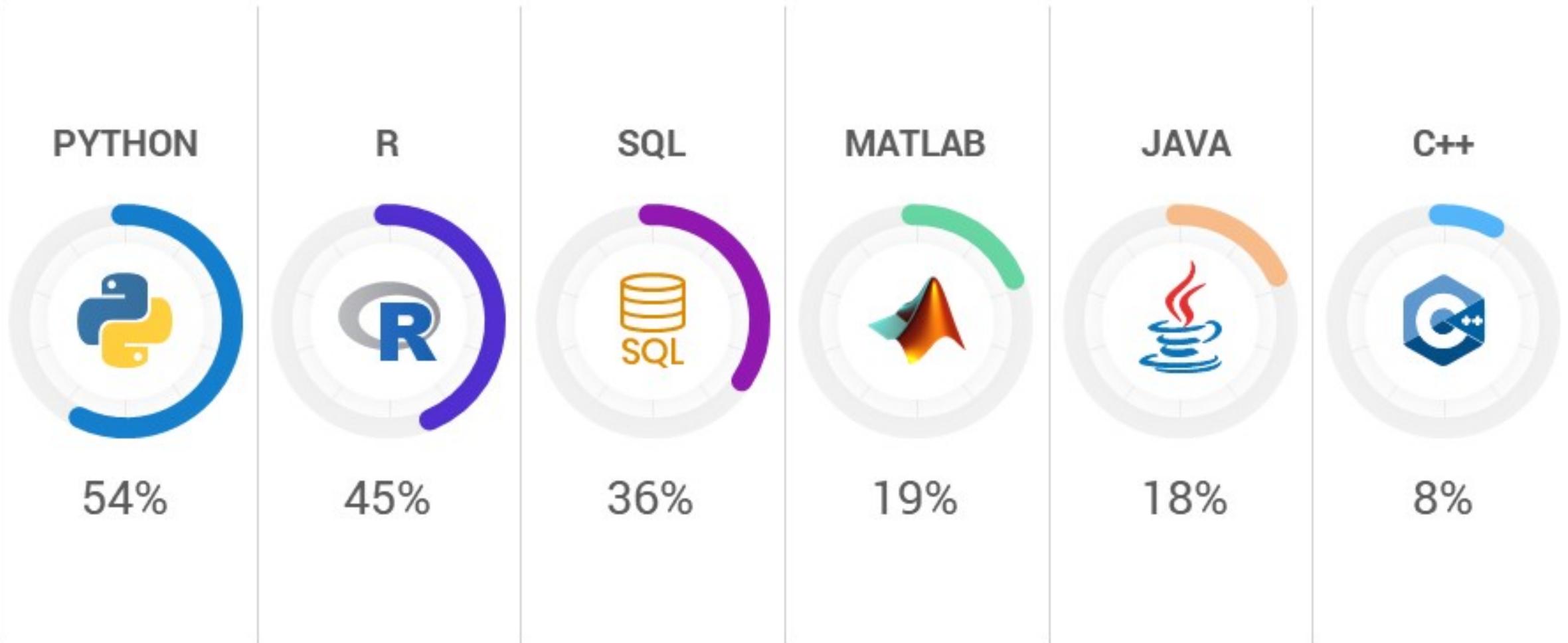


Data Science Applications



Data Science Coding Toolbox

Data Science Coding Toolbox



Top Data Science Tools

The list of 14 best data science tools that most of the data scientists used.



SAS



Apache Spark



BigML



D3.js



MATLAB



Excel



ggplot2



Tableau



Jupyter



Matplotlib



NLTK



Scikit-learn



TensorFlow



Weka