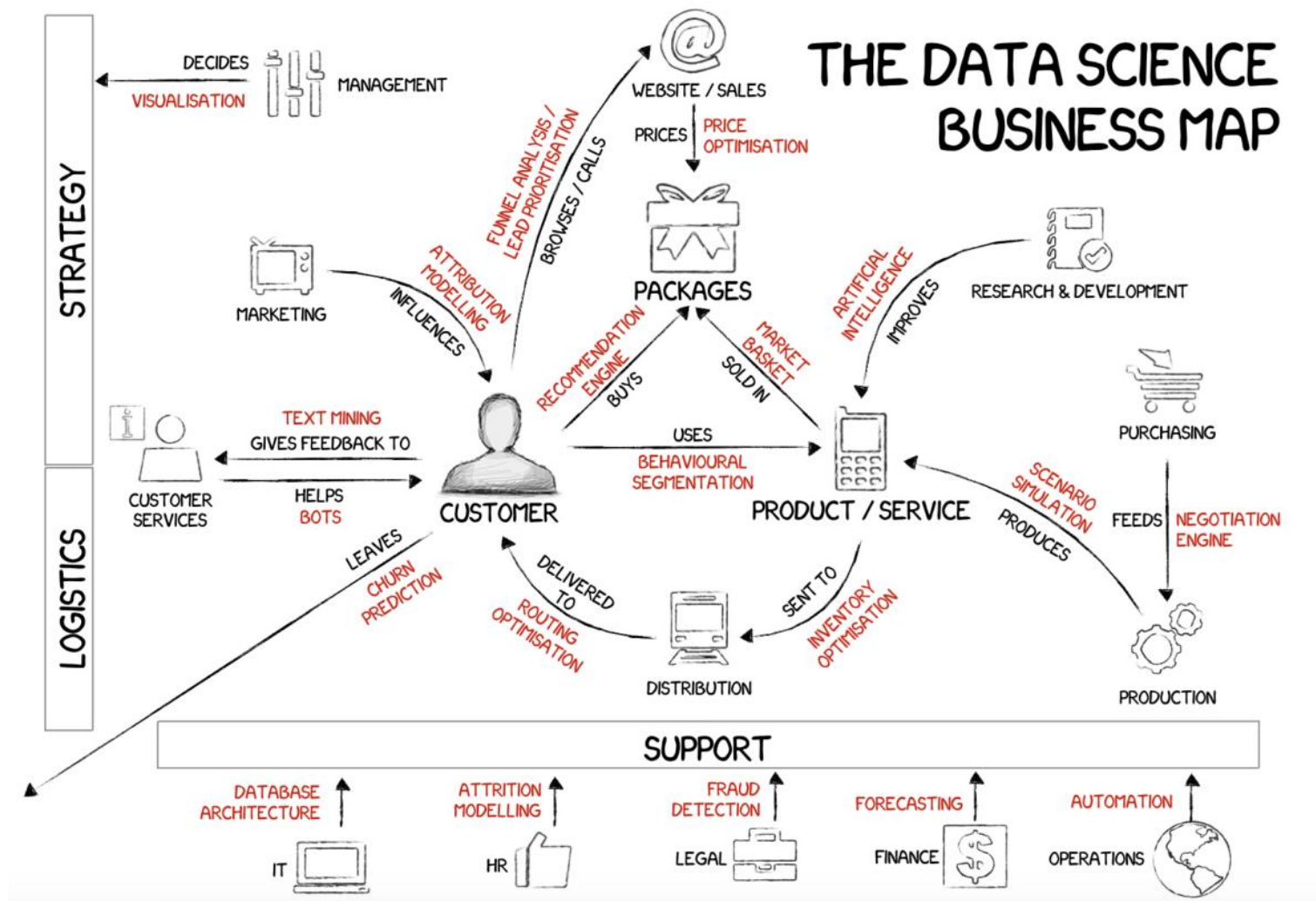


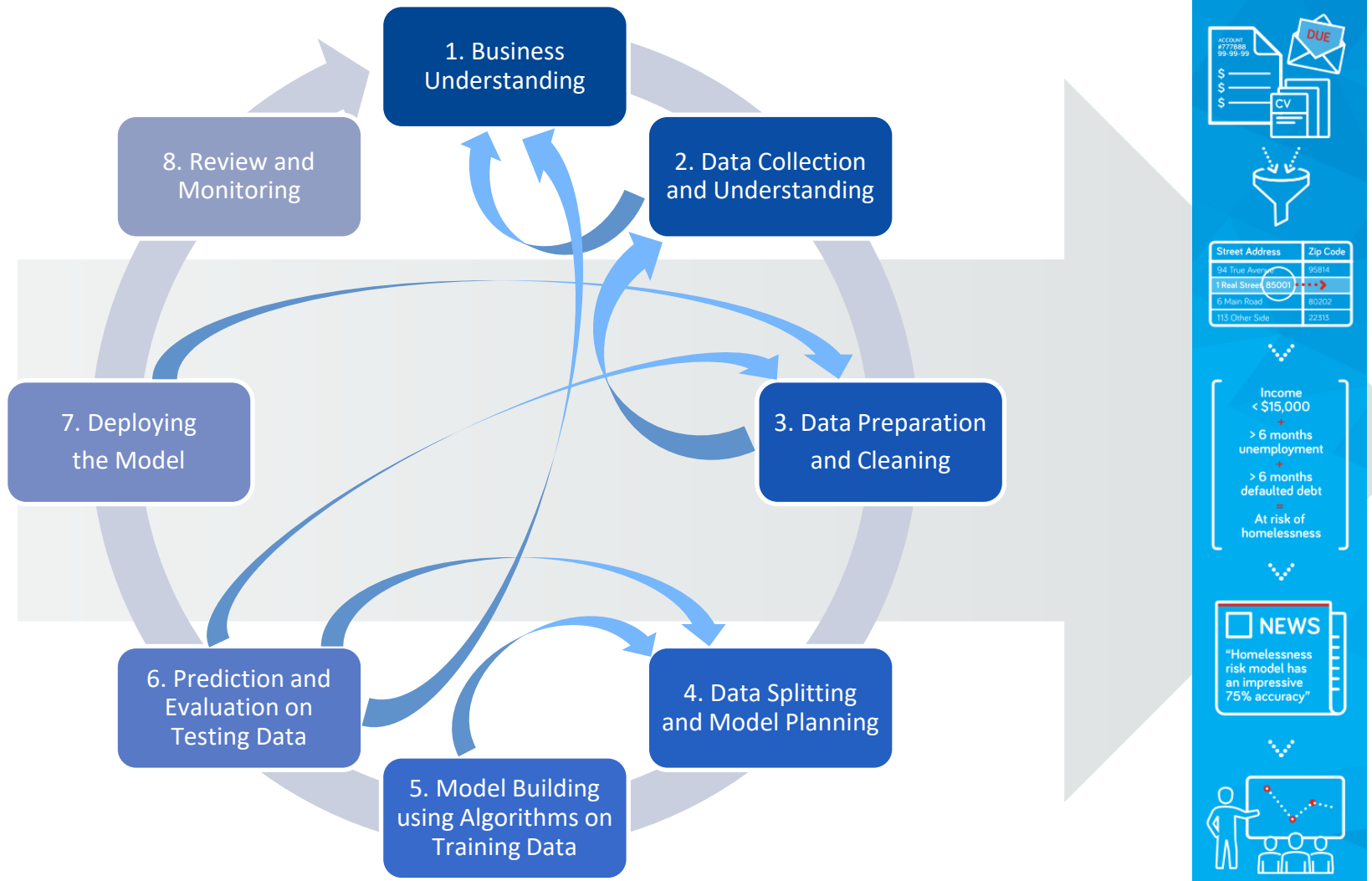
What is Data Science?

➤ The Real World Data Science



Data Science Process

➤ Getting from Raw Data to Outcomes



Who is a Data Scientist?

➤ What should we do?



Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

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
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom 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
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include marketing strategy and optimization, customer tracking and on-site analytics, predictive analytics and econometrics, data warehousing and big data systems, marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY

Who is a Data Scientist?

➤ Traditional Specialist of Data Science Team



Data Analyst (DA)

: Assist DS with domain understanding, data preprocessing and problem defining.



Data Scientist (DS)

: Prepares data, engineers features, most valuable skill: training models.



Data Engineer (DE)

: Data acquisition focus. Build data pipelines. Not uncommon to have 5:1 ratio DE:DS



Data Application Architect (DAA)

: Design complete solution; deploy and maintain models in production

Who is a Data Scientist?

➤ Typical Collaboration of Data Science Project

- Makes data science teams more productive
- Broad support for open source libraries in various languages



**Understand
Business
Objectives**

**ID Mapping
Procure
Training
Data**

**Prepare Data
and Build
Features**

**Train, Tune,
and Test
Models**

**Deploy and
Operationalize
Models**

**Update
Models**

Who is a Data Scientist?

➤ What is a Project of Data Science?



▪ TASKS

: In addition to advanced analytic skills, these individuals are also proficient at integrating and preparing large, varied datasets, architecting specialized database and computing environments, and communicating results.

▪ MISSION

: A data scientist may or may not have specialized industry knowledge to aid in modeling business problems and with understanding and preparing data.

▪ TALENT

: Creating value from data requires a range of talents from data integration and preparation, to architecting specialized computing/database environments, to data mining and intelligent algorithms.

▪ RESPONSIBILITY

: An individual responsible for modeling complex business problems, discovering business insights and identifying opportunities through the use of statistical, algorithmic, mining and visualization techniques.

WE ALL CAN BE DATA SCIENTISTS NOW!

