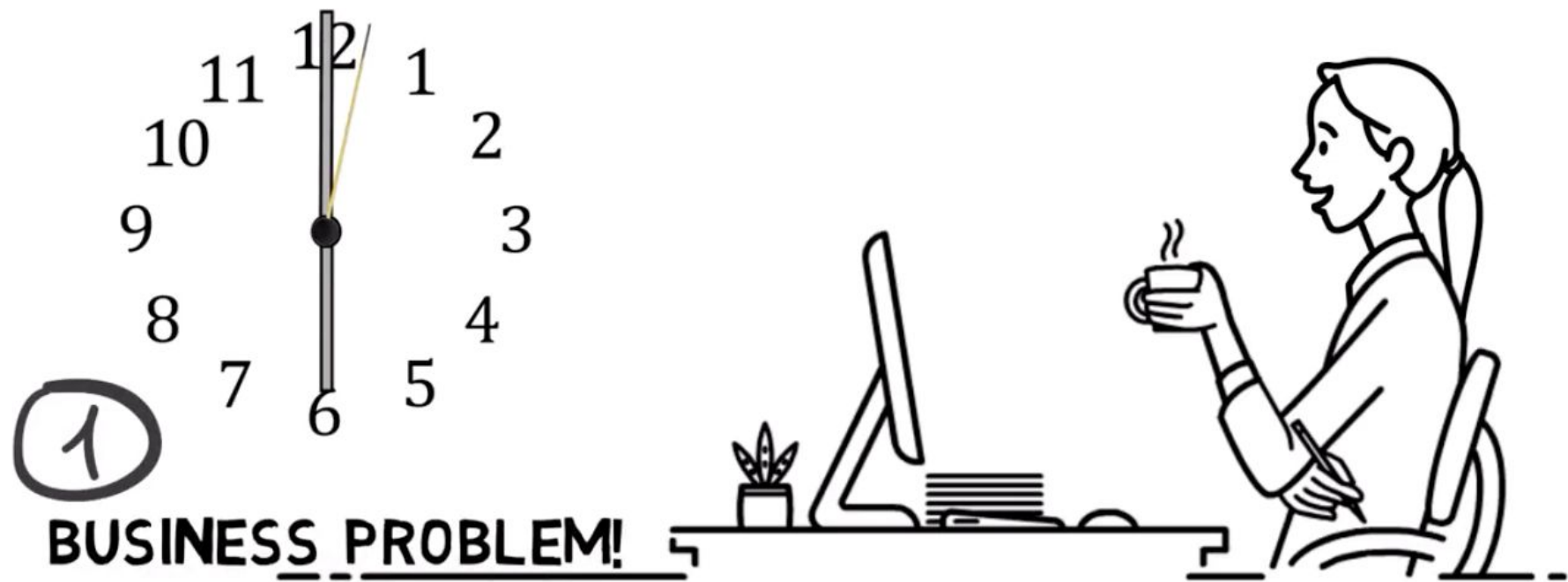


DataScience 101



1

BUSINESS PROBLEM

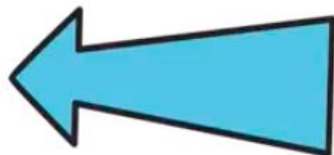
WHY?....WHY?....WHY?....



②

DATA ACQUISITION

- WEB SERVERS
- LOGS
- DATABASES
- API'S
- ONLINE REPOSITORIES





DATA PREPARATION

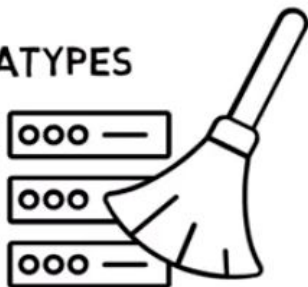
DATA CLEANING

TRANSFORMATION

INCONSISTENT DATATYPES

MISSPELLED ATTRIBUTES

MISSING AND DUPLICATE VALUES



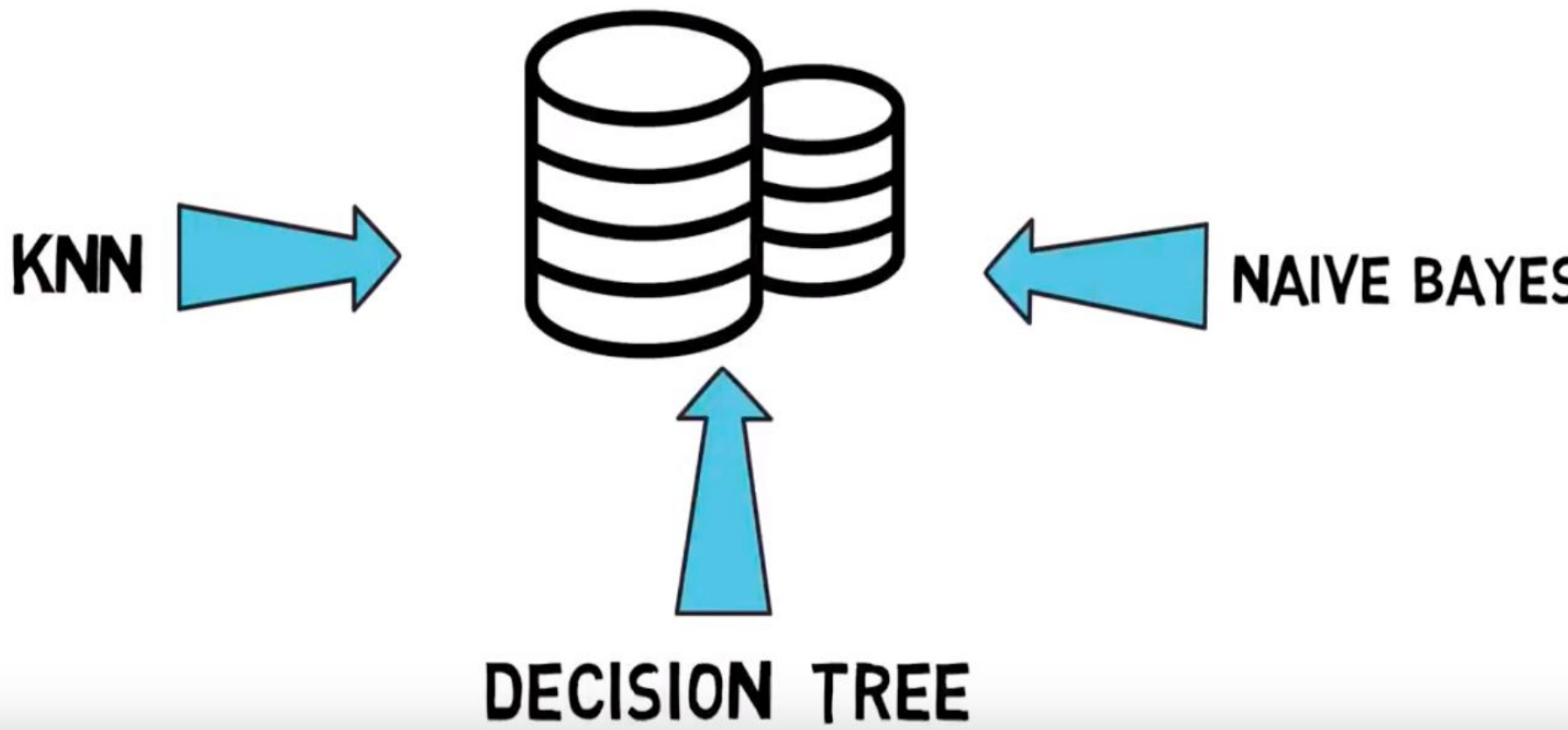


④ EXPLORATORY DATA ANALYSIS



**DEFINES AND REFINES
THE SELECTION OF FEATURE
VARIABLES THAT WILL BE USED
IN THE MODEL DEVELOPMENT**

⑤ DATA MODELING





IDENTIFY THE MODEL
THAT BEST FITS THE
BUSINESS REQUIREMENT



TRAINS THE MODELS ON THE
TRAINING DATASET AND TEST



SELECT THE BEST
PERFORMING MODEL



python™

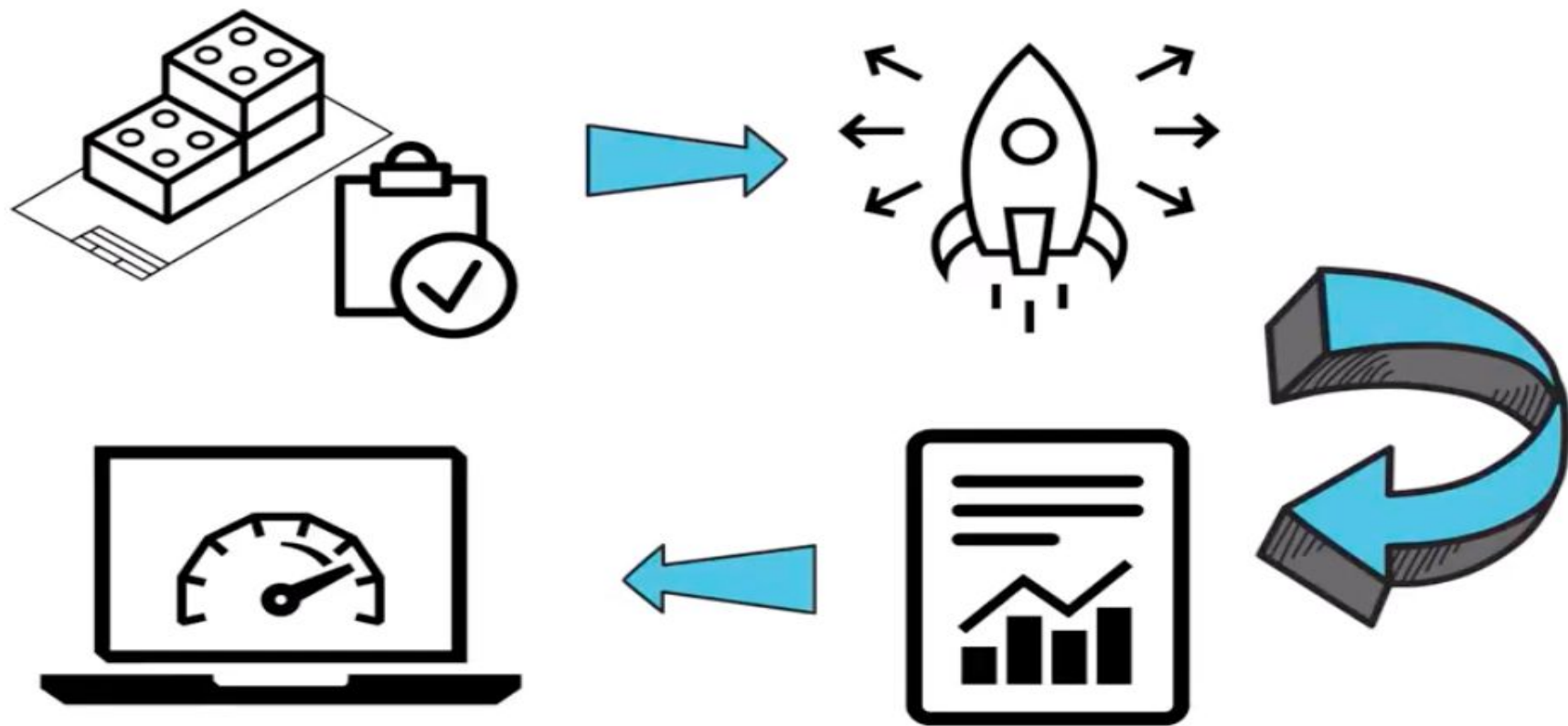


6

VISUALIZATION AND COMMUNICATION

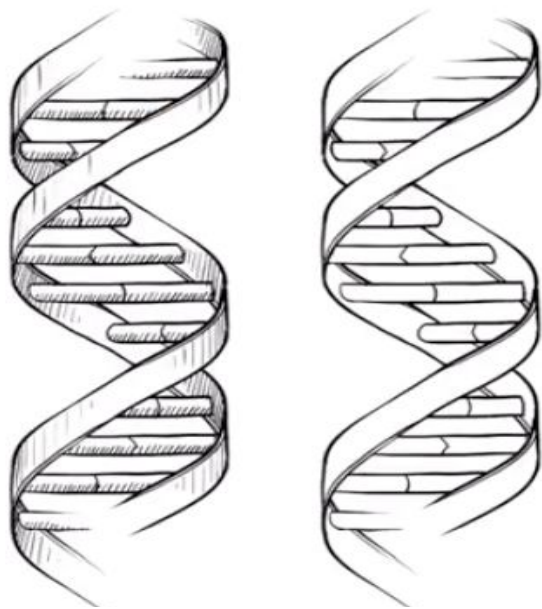


7 DEPLOYS AND MAINTAINS



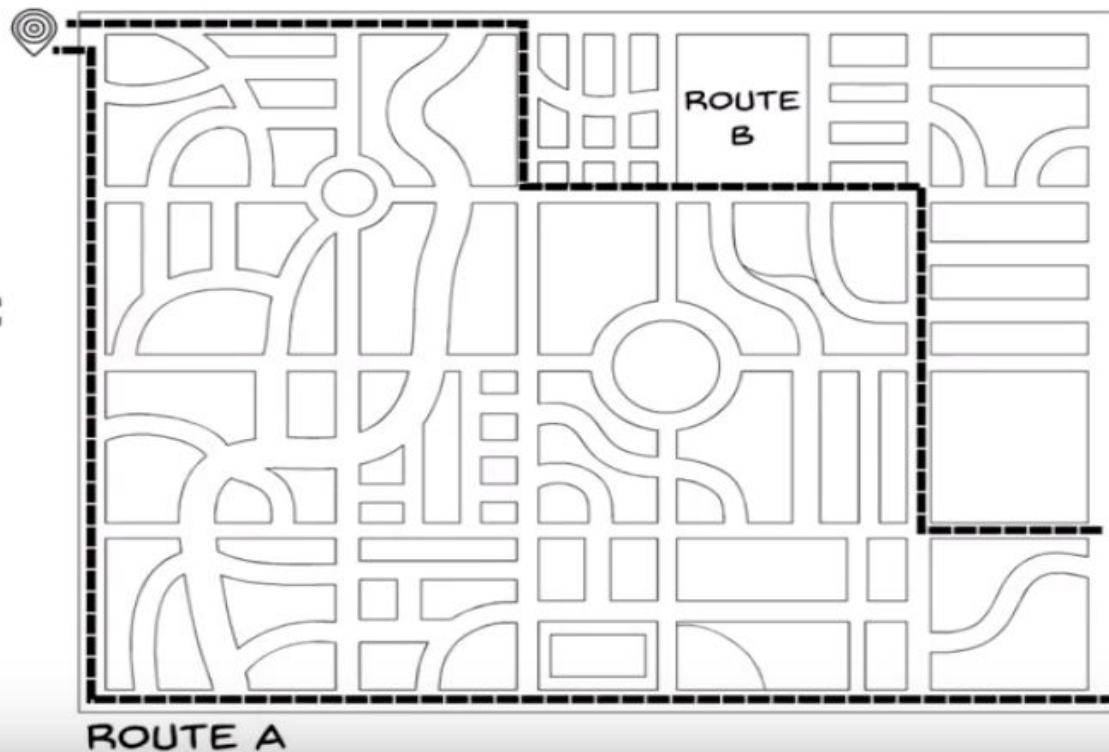
**NOW, LETS SEE HOW DATA SCIENCE
IS CHANGING THE WORLD**

GENOMIC DATA PROVIDES A DEEPER UNDERSTANDING OF GENETIC ISSUES

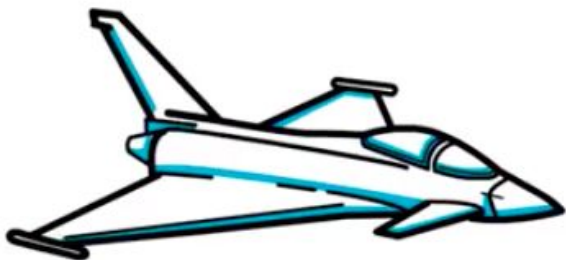


**LOGISTICS COMPANIES LIKE DHL, FEDEX HAVE
DISCOVERED THE BEST TIME AND ROUTES TO SHIP**

**BEST TIME:
12:00PM**



AIRLINE COMPANIES CAN NOW EASILY PREDICT FLIGHT DELAY AND NOTIFY THE PASSENGERS



H487	CANCELLED
PP87	DELAYED

Let's get started!

- [Introduction to Numpy](#)
- [Introduction to Pandas](#)
- [Introduction to Matplotlib](#)
- [Introduction to Scikit-Learn](#)