
THE WORLD OF DATA SCIENCE
Kunaal Naik

Kunaal Naik

- ✓ Love teaching Data Science
- ✓ YouTube Channel teaching Data Science
- ✓ Marketing Advisor, Data Science - Dell
- ✓ Lifeaholic Evangelist
- ✓ Avid Learner
- ✓ Scuba Diver

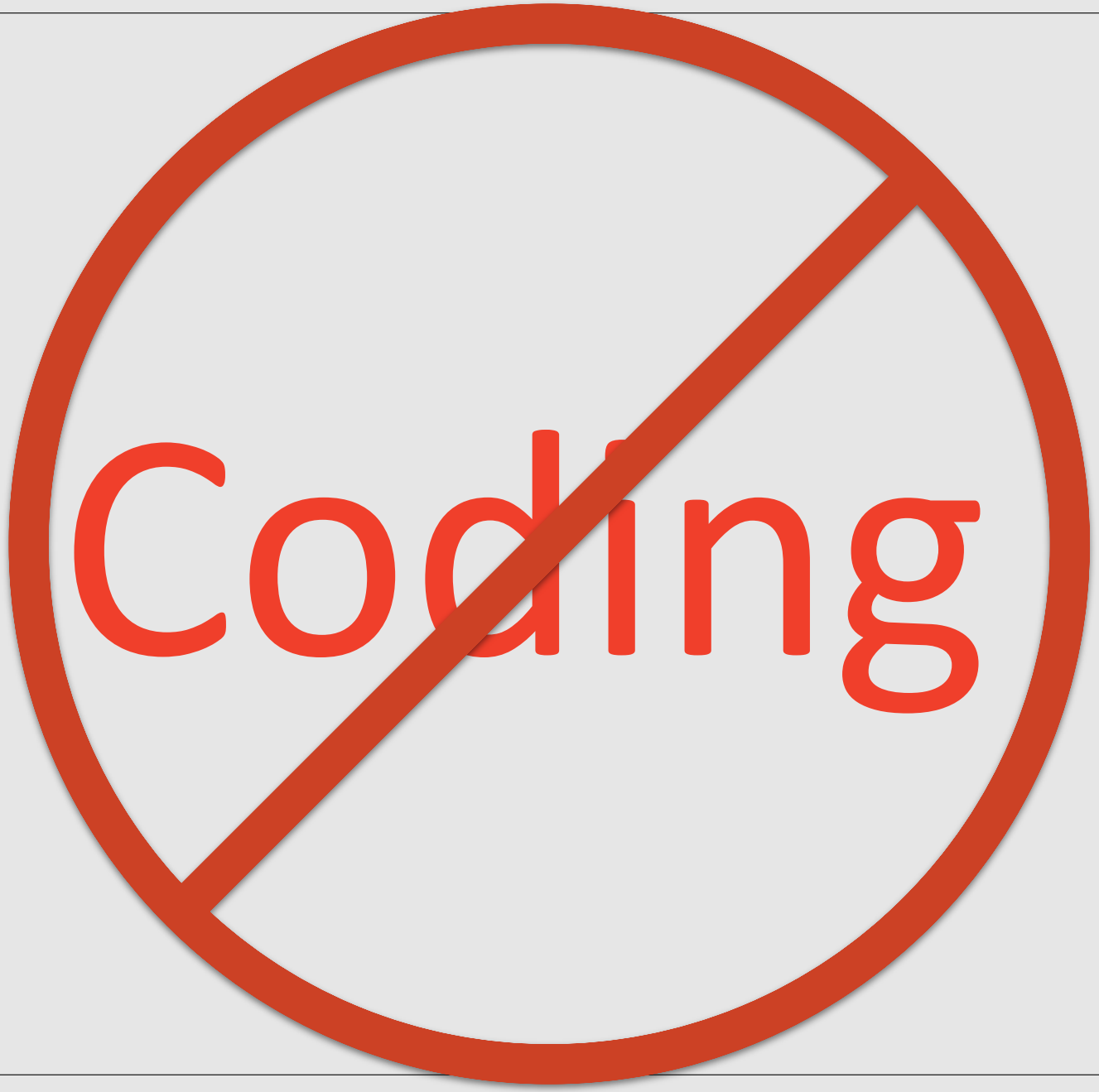
(www.youtube.com/KunaalNaik)

When I learned the

```
void main()
{
printf(“Hello World!!!”);
}
```

After 3 months

```
void main()
{
printf(“Linked List Stuff !”);
}
```


Coding

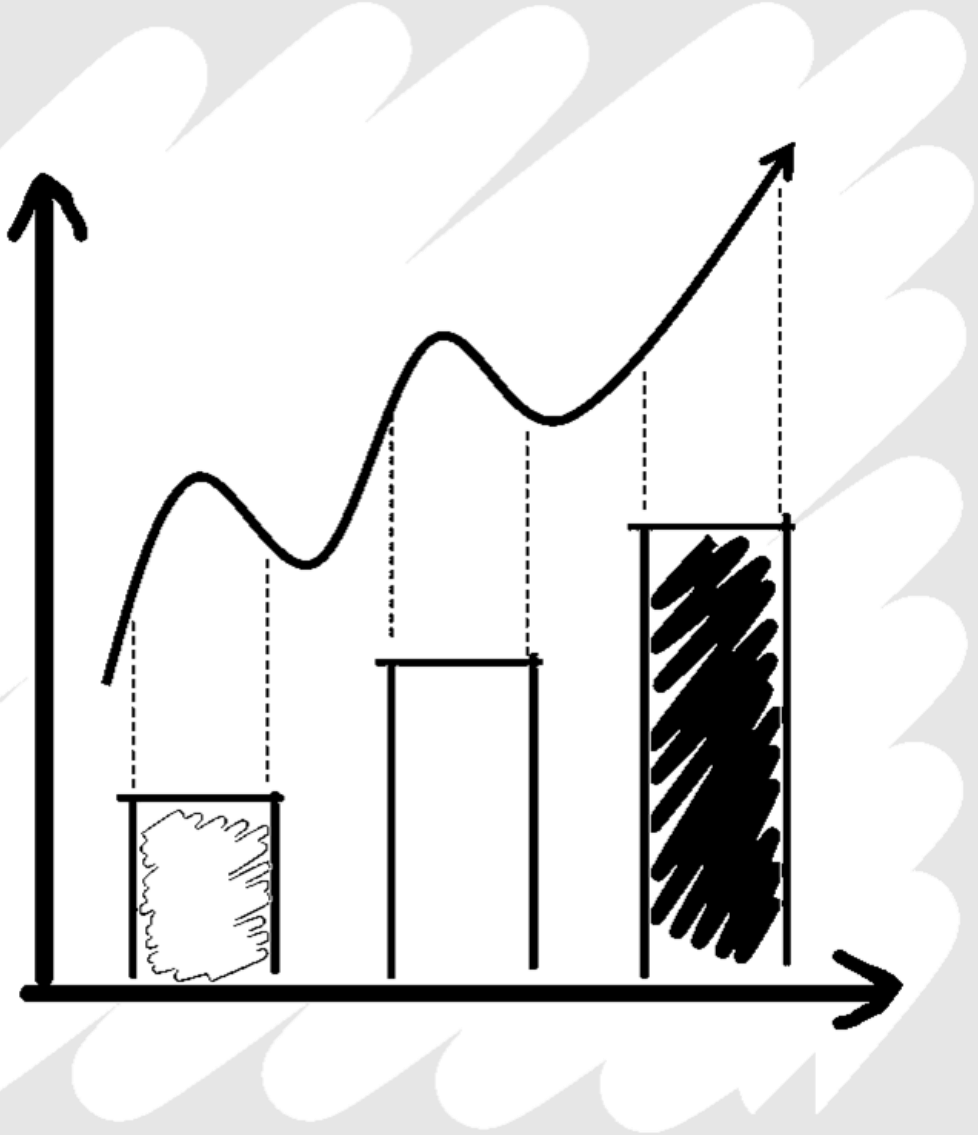
Trainer /
”Business Analyst”

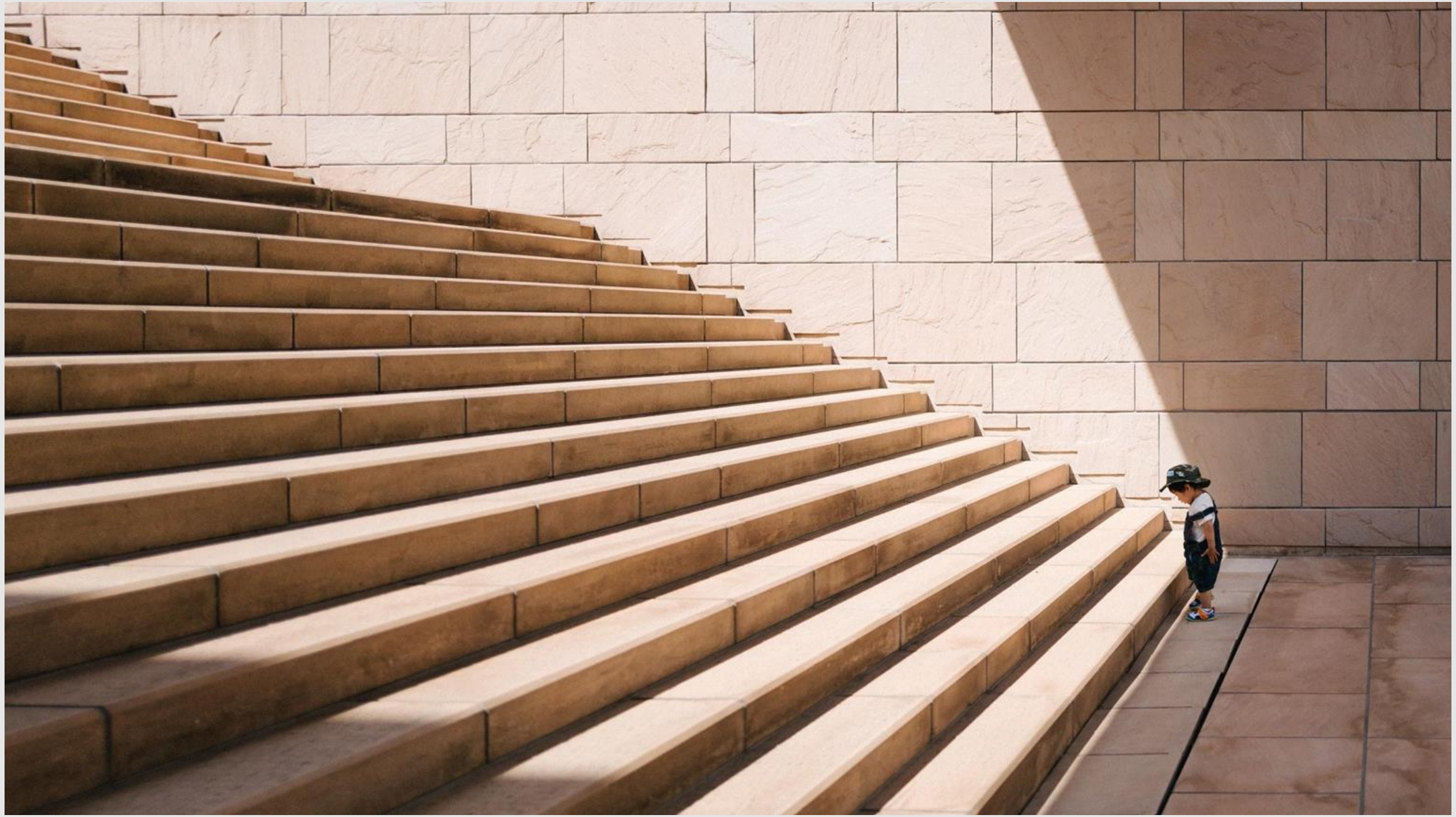
2000+

Hours of training

Over 10 years

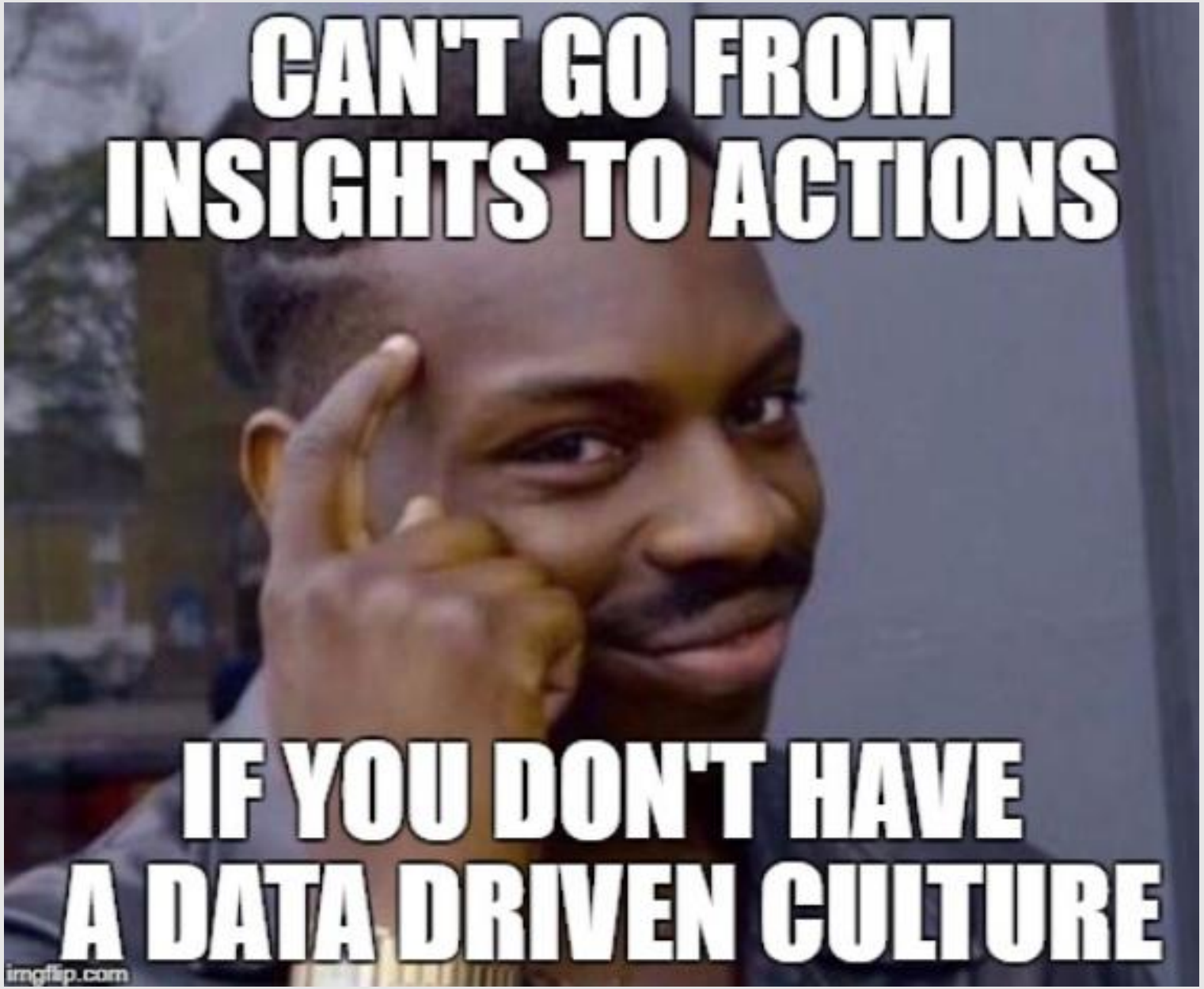
“Data Science Course”



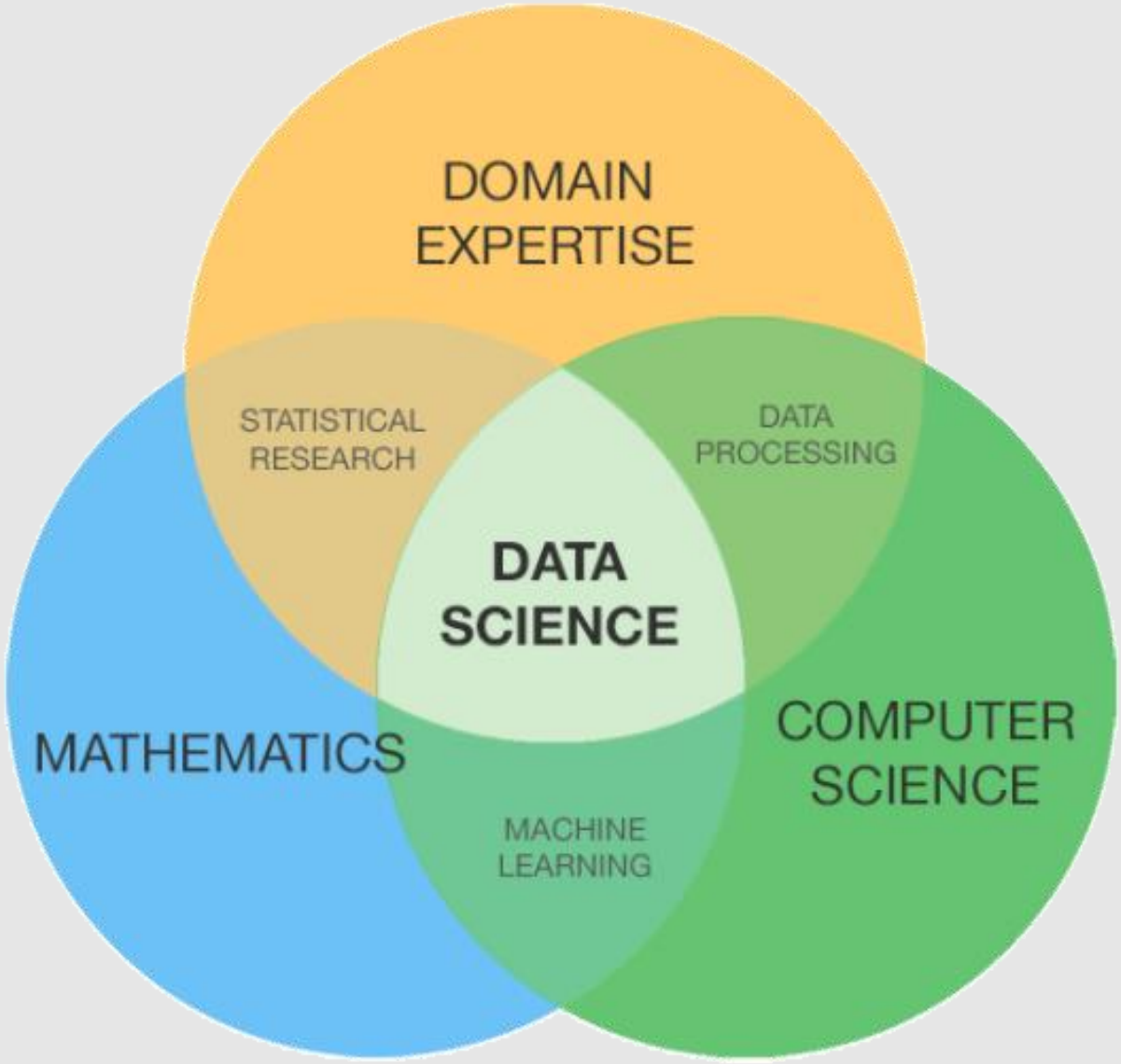



WELCOME TO THE DARK SIDE OF SCIENCE
DATA SCIENCE
memegenerator.net

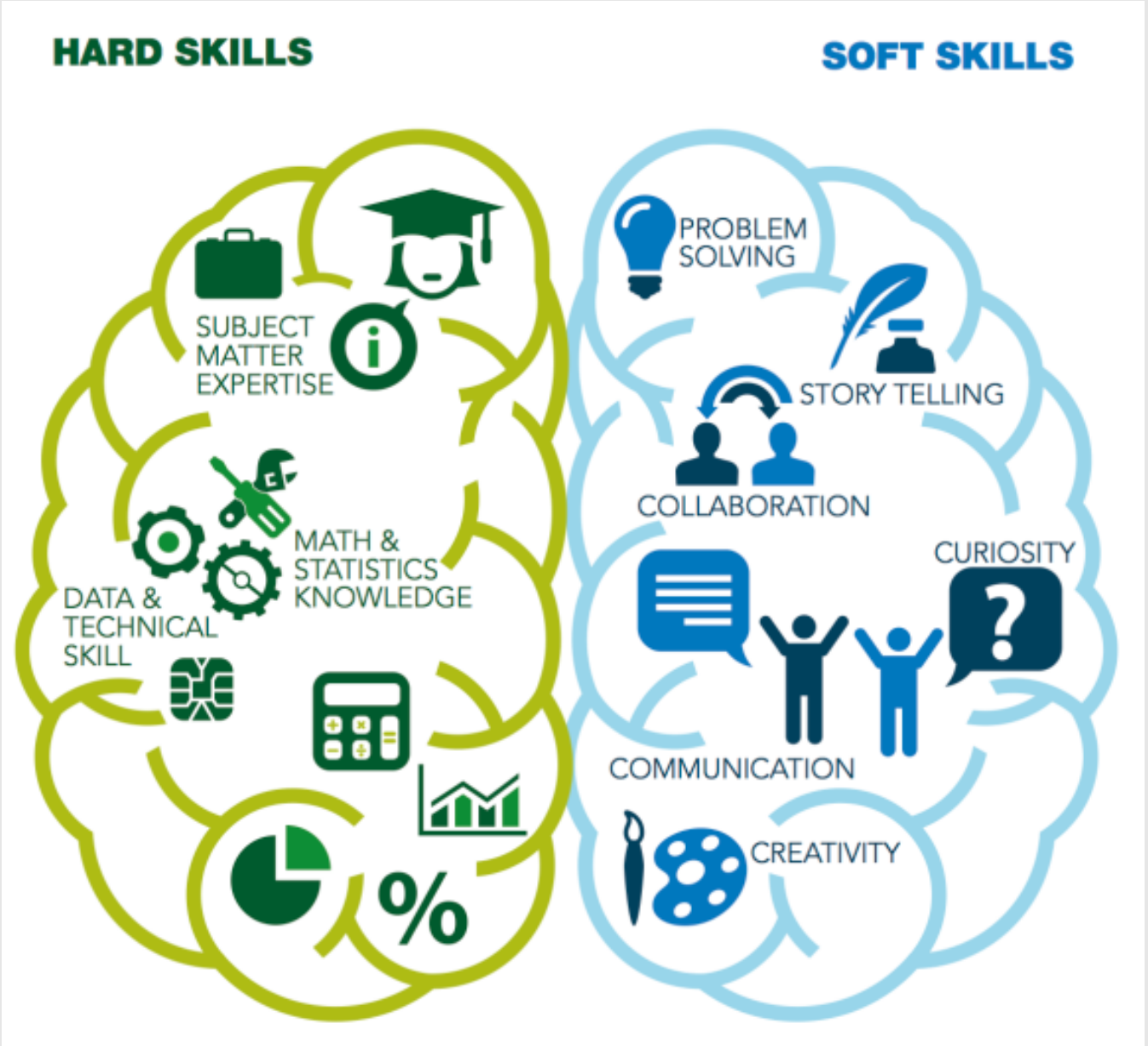
What is Data Science ?



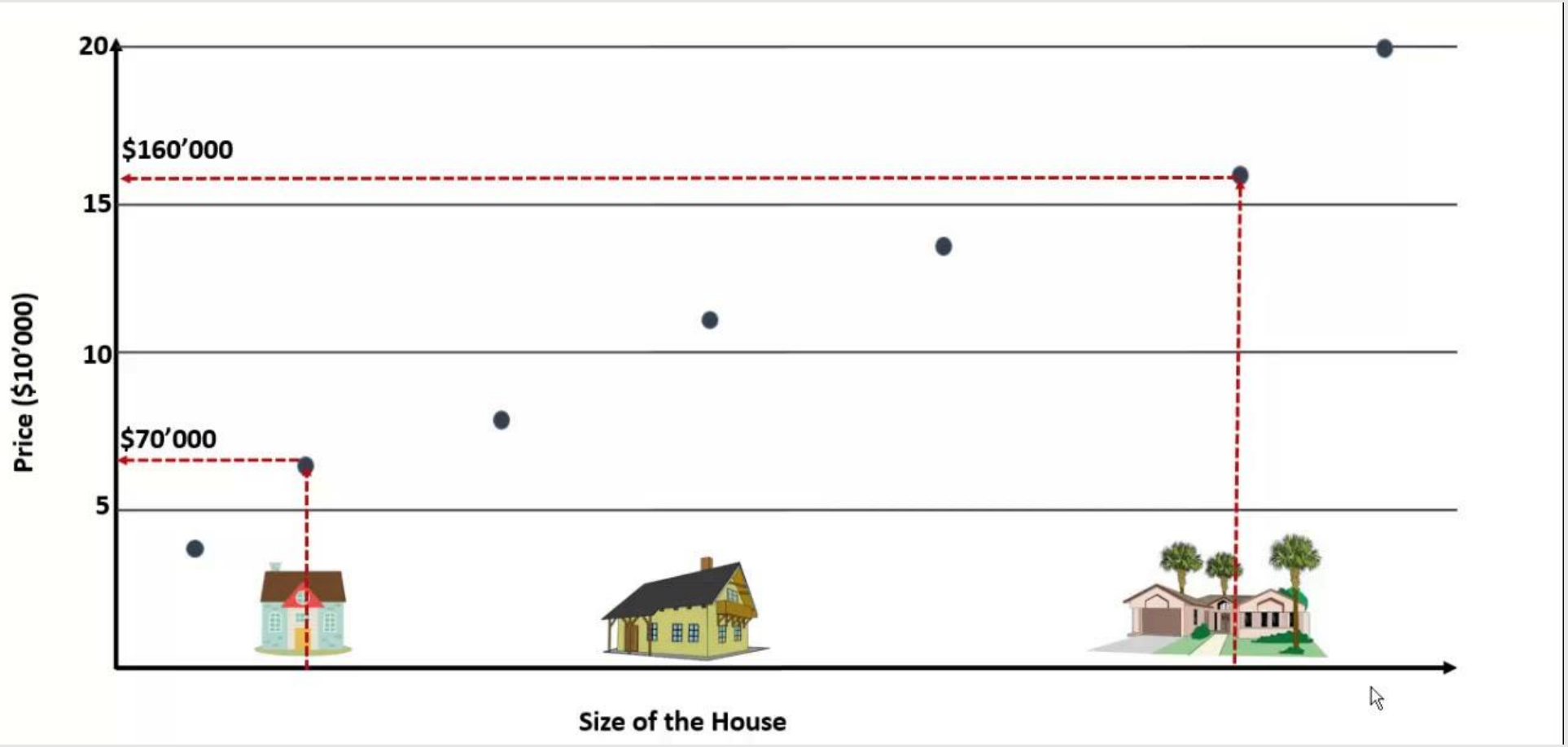
What actually is Data Science ?



What actually is Data Science ?



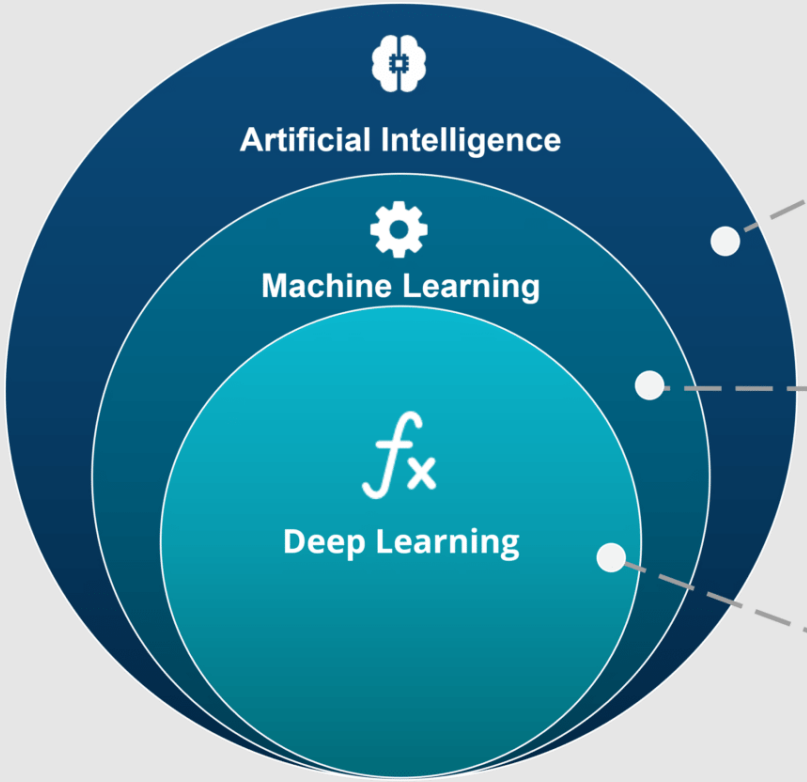
Example – Curve Fitting



THE FUSS ABOUT TERMINOLOGY

AI vs ML vs DEEP LEARNING

AI vs ML vs Deep Learning

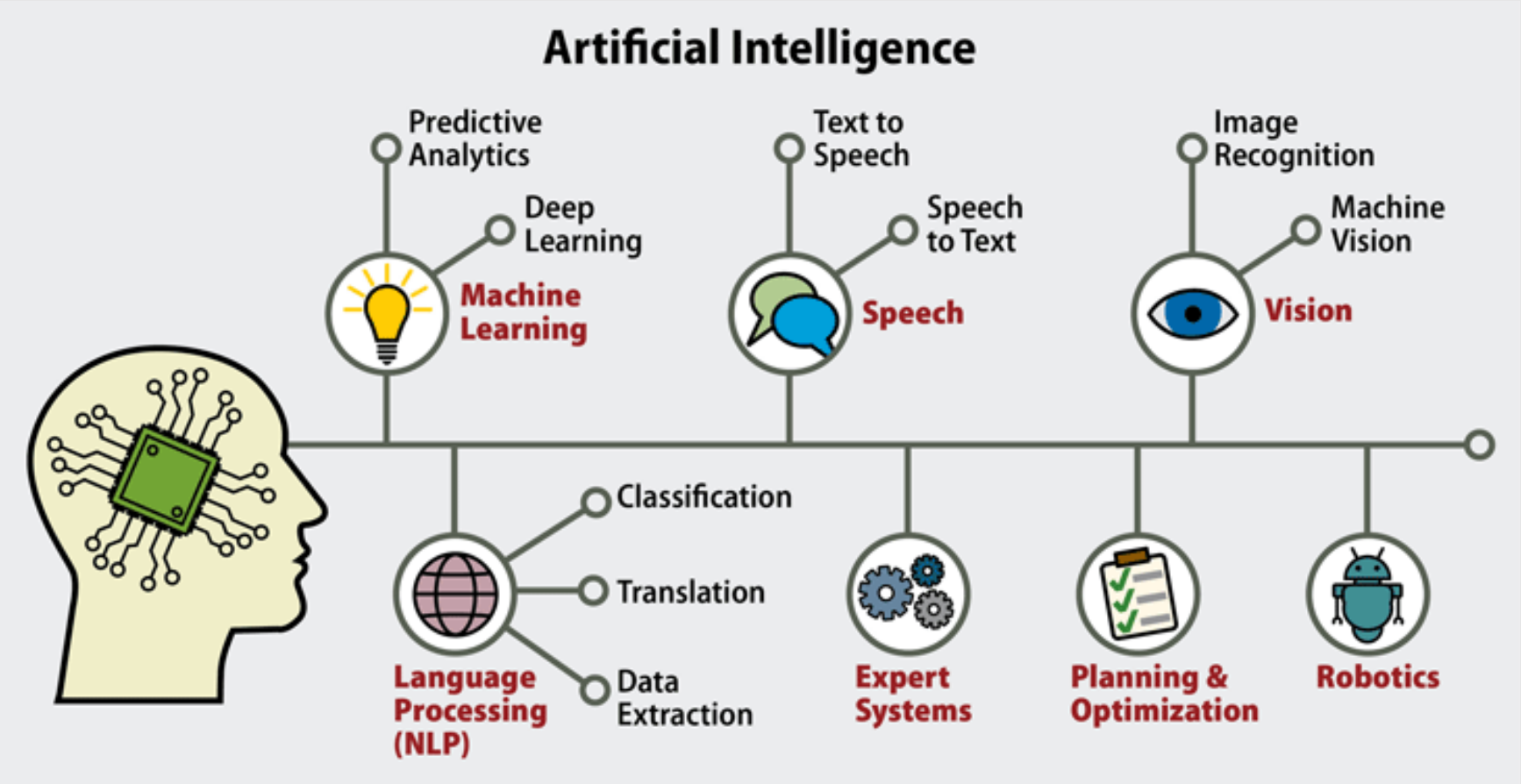


ARTIFICIAL INTELLIGENCE
A technique which enables machines to mimic human behaviour

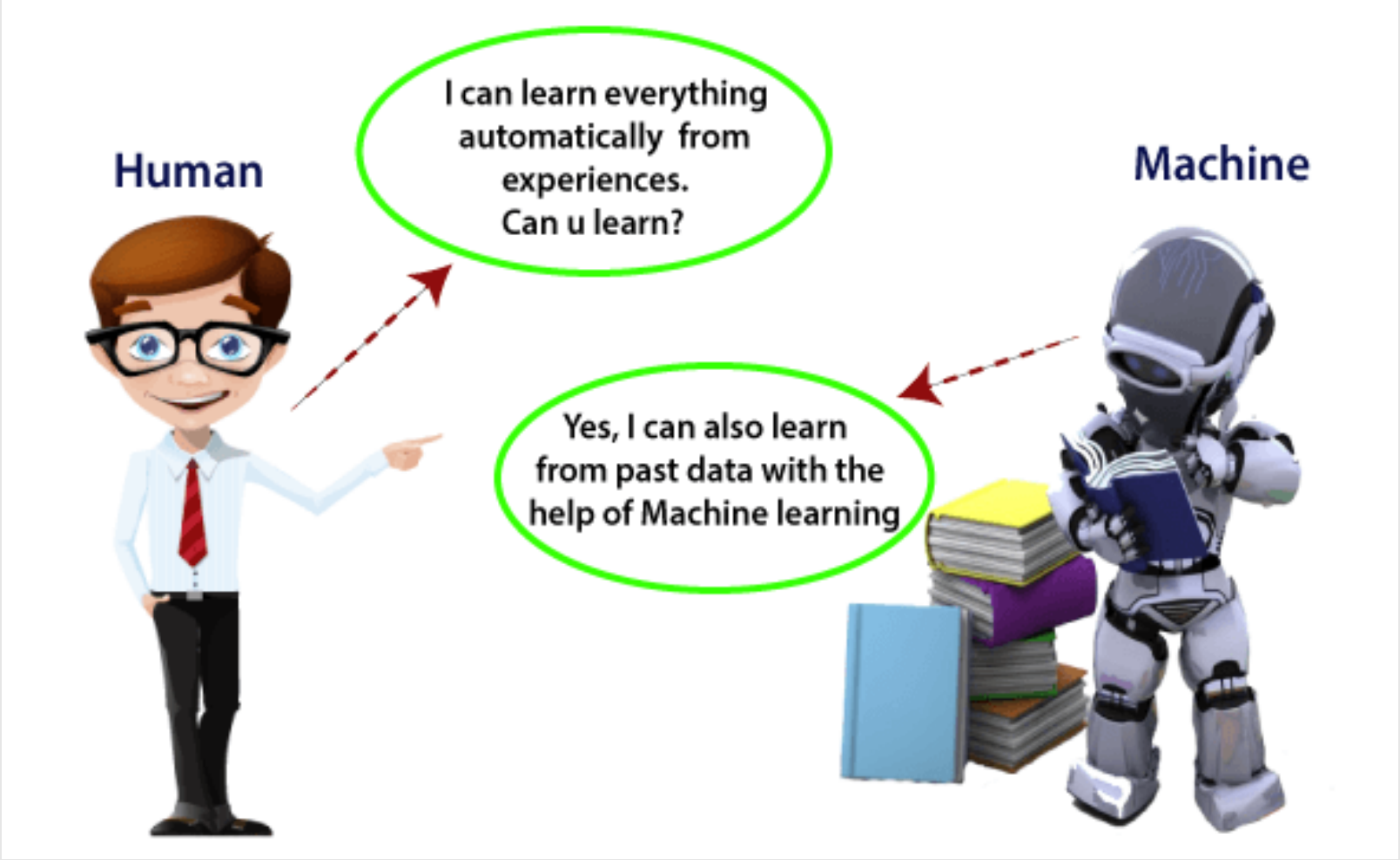
MACHINE LEARNING
Subset of AI technique which use statistical methods to enable machines to improve with experience

DEEP LEARNING
Subset of ML which make the computation of multi-layer neural network feasible

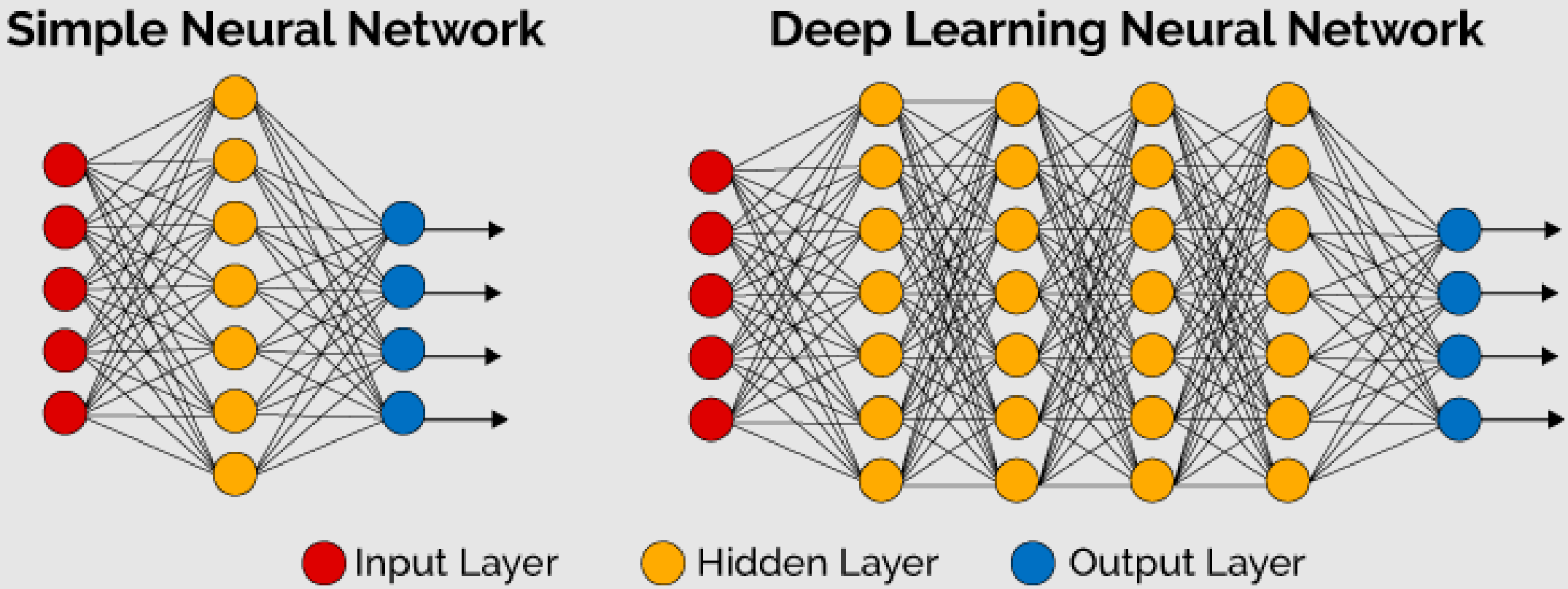
Artificial Intelligence

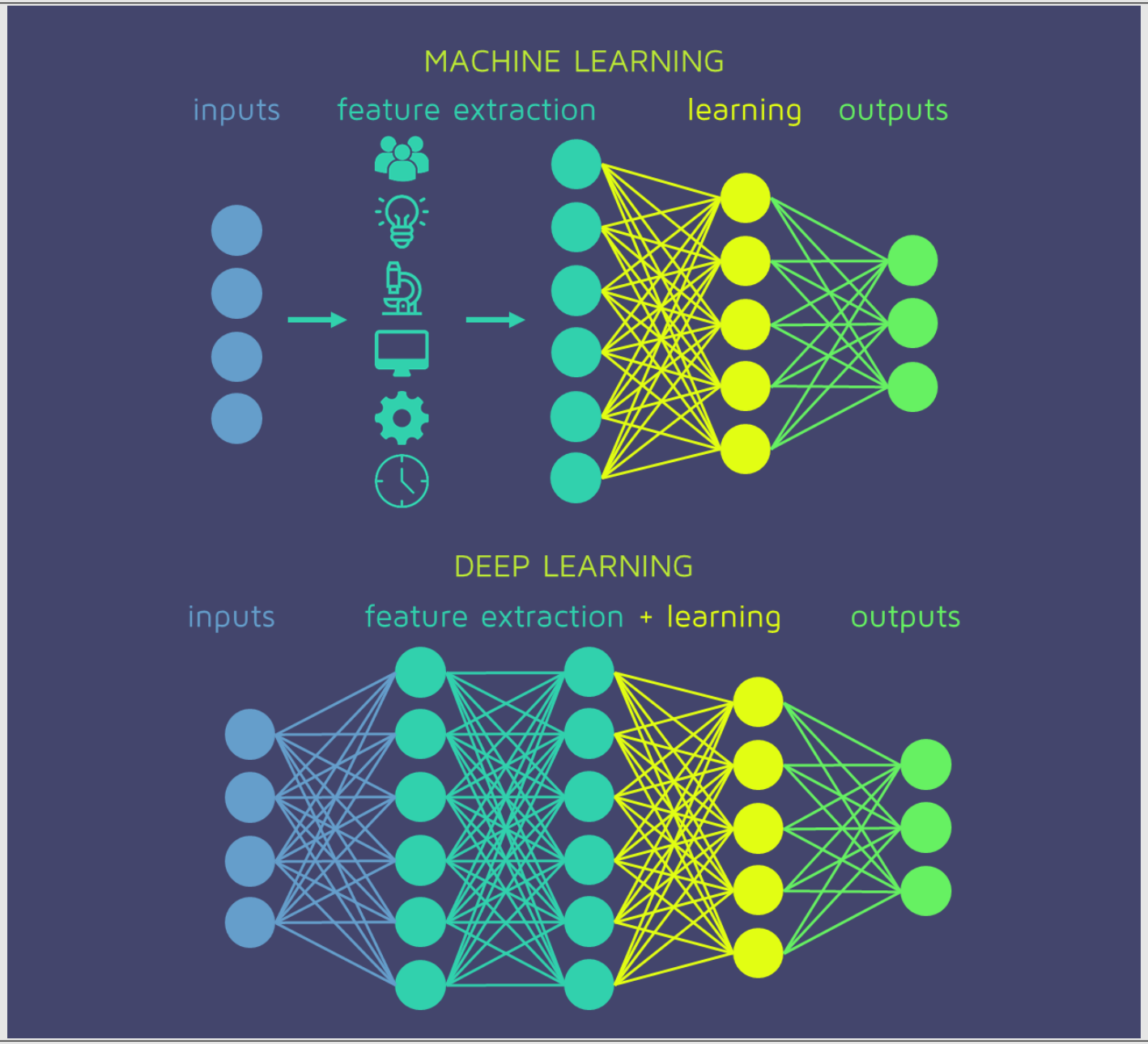


Machine Learning



Deep Learning




MACHINE LEARNING
inputs
feature extraction
learning
outputs
DEEP LEARNING
inputs
feature extraction + learning
outputs

IMPORTANCE OF DATA SCIENCE

Everything seems important!

Why does it matter currently?

28%

Growth in Jobs

417%

Growth in Jobs

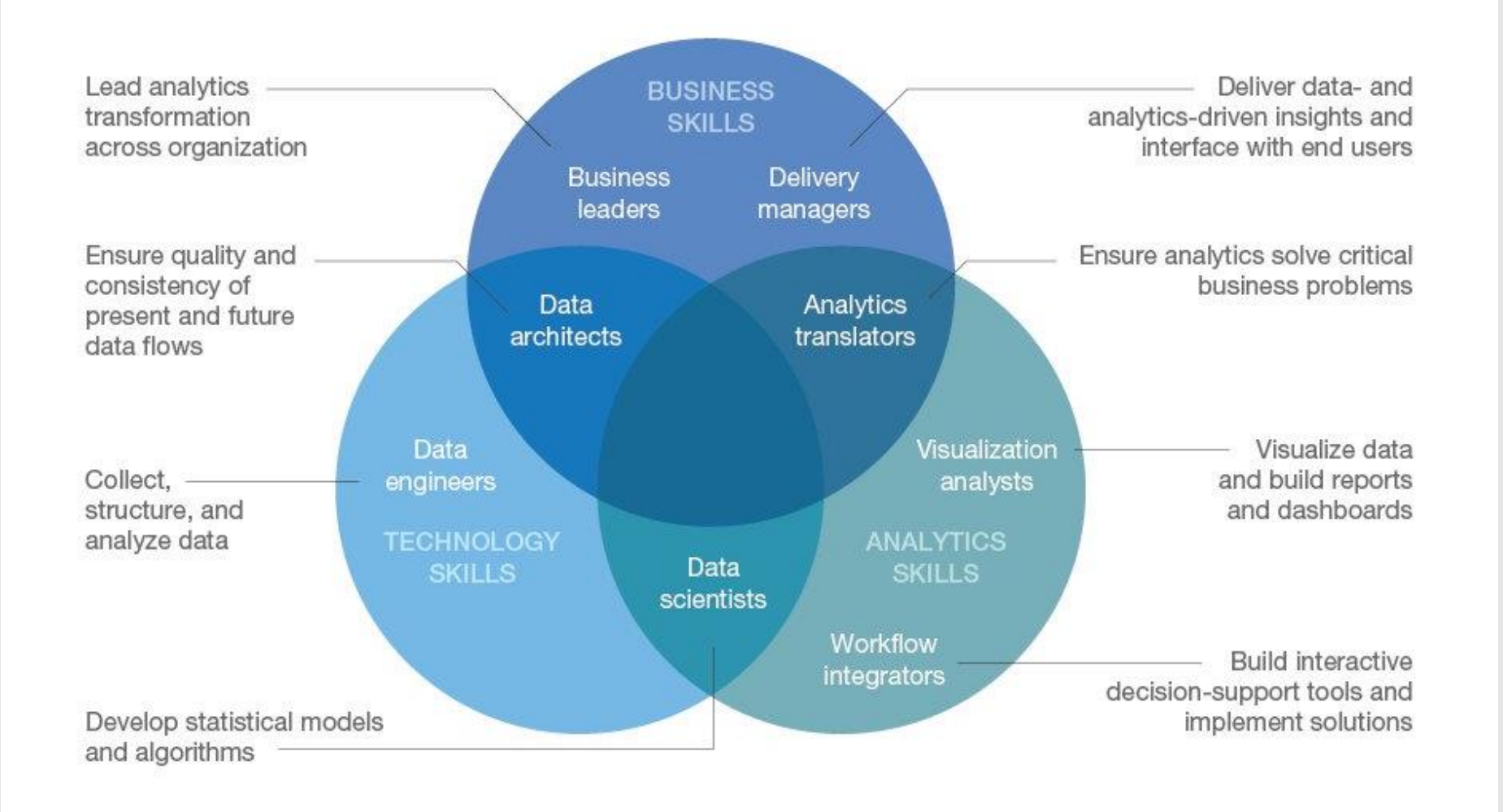
Why does it matter currently?

So what skills do you require?

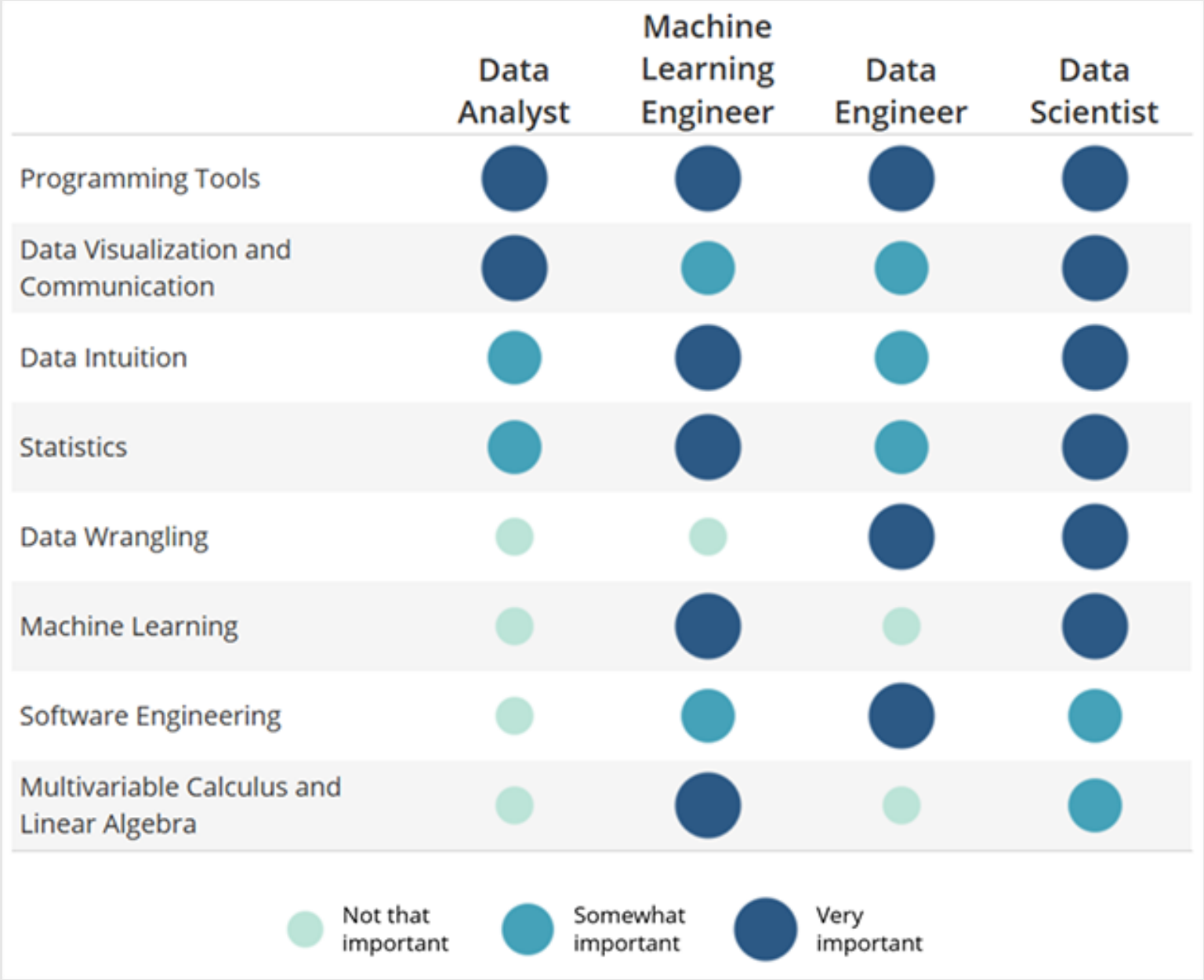
SKILLSET

It all boils down to the skills you build!

Data Roles and Skills Set

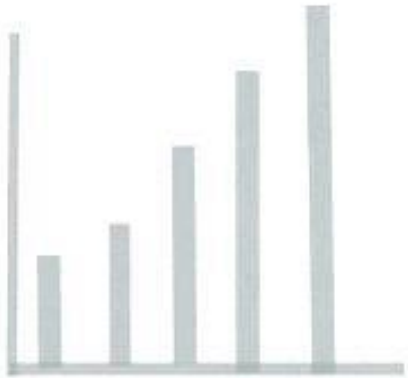


Data Roles and Skills Set



DATA ANALYST

They figure out what the data means



Technical Skills?

Data Analysts typically use computer systems and complex calculations applications to get their numbers nailed down, but there is still a lot of intellectual know-how that goes into making these systems work.

Applications

How to Data Analysts use this Data

- 13% Extrapolation and Interpretation
- 44% Projections and Advisory Responsibilities
- 32% Research and Writing Tasks
- 19% System Expertise and Troubleshooting
- 7% Training

PURPOSE

There job is to assign numerical values to different business functions, and are responsible for identifying effciencies, problem areas, and possible improvements.

DATA

What do they do with all that statistical Data?

Collect	51%
Organize	11%
Interpret	22%
Study	14%
Apply	31%

WHY?

This is a process of inspecting, cleaning, transforming, and modelling data with the goal of discovering useful information.

THE STATE OF THE DATA SCIENTIST

HIGHEST EDUCATION LEVEL

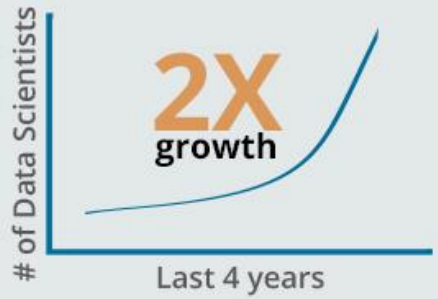
TOP PRIMARY SKILLS

1. Data Analysis
2. R
3. Python
4. Data Mining
5. Machine Learning

TOP EDUCATIONAL BACKGROUNDS

1. Computer Science
2. Business Admin
3. Statistics
4. Mathematics
5. Physics

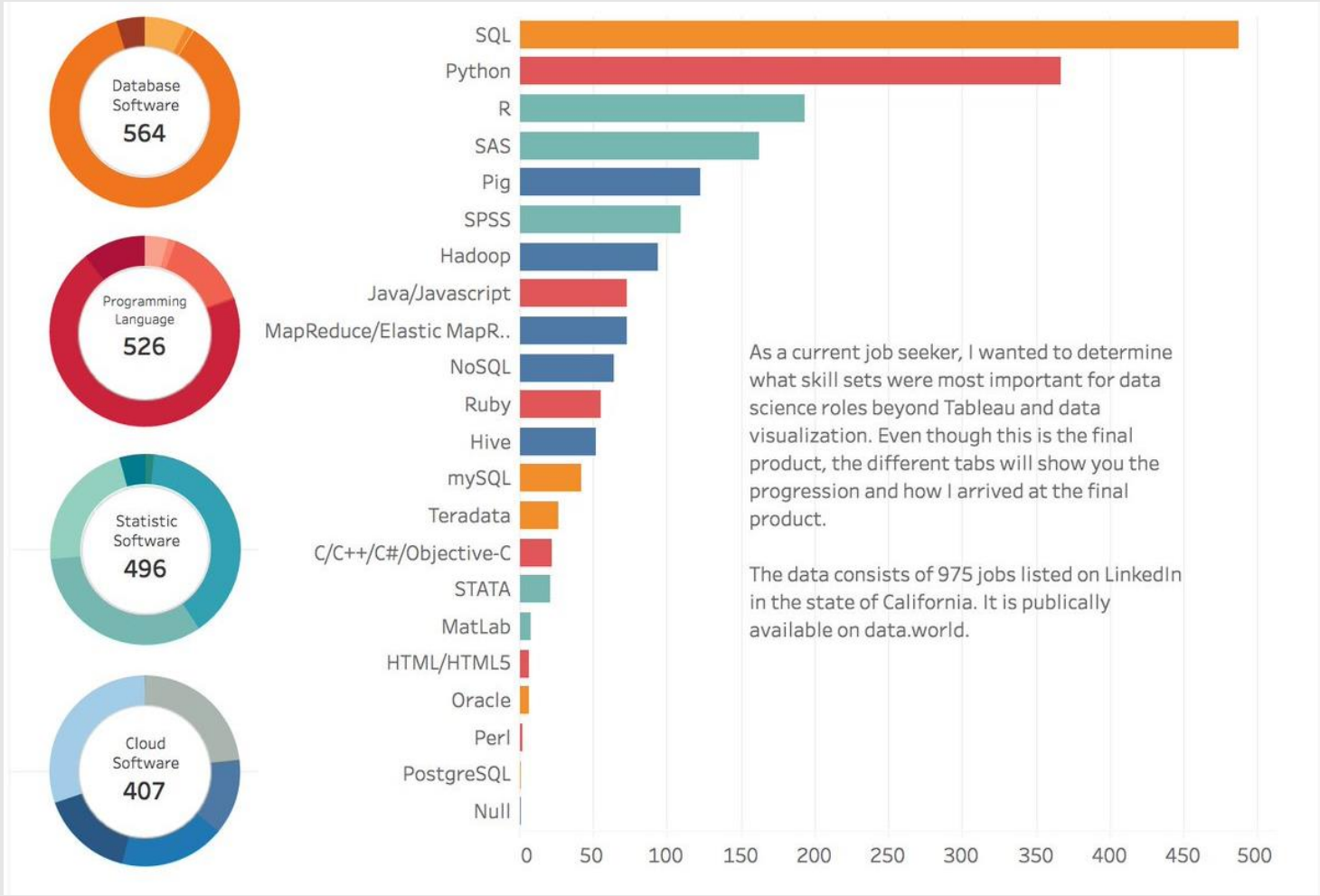
INDUSTRY GROWTH



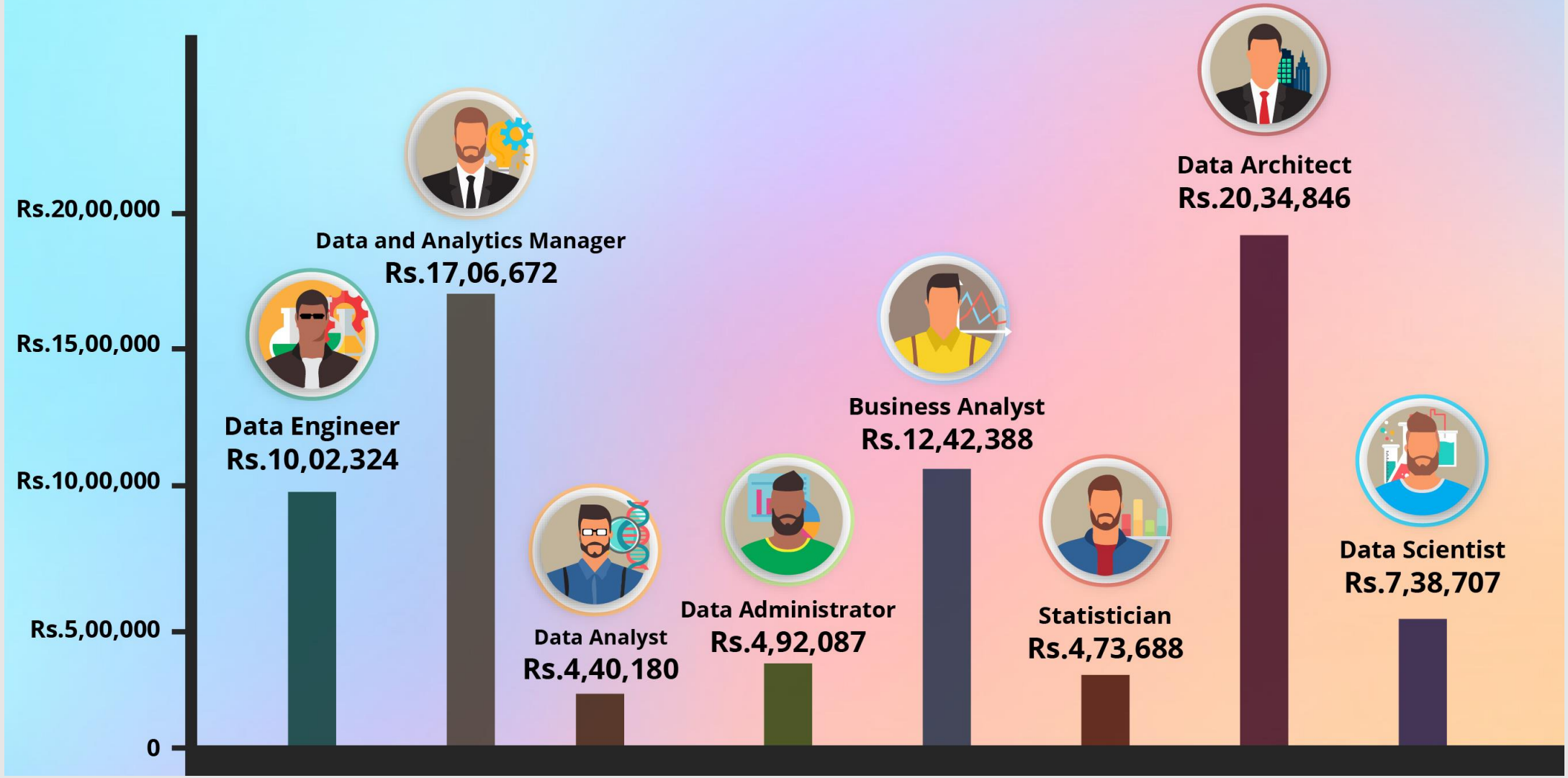
TOP INDUSTRIES EMPLOYING DATA SCIENTISTS

1. Information Technology & Services
2. Internet
3. Computer Software
4. Education
5. Banking & Financial Services

Essential Skills for Data Science



WHO EARNS WHAT SALARY - NATIONAL AVERAGE SALARIES (INDIA)



Top 10 jobs involving AI skills

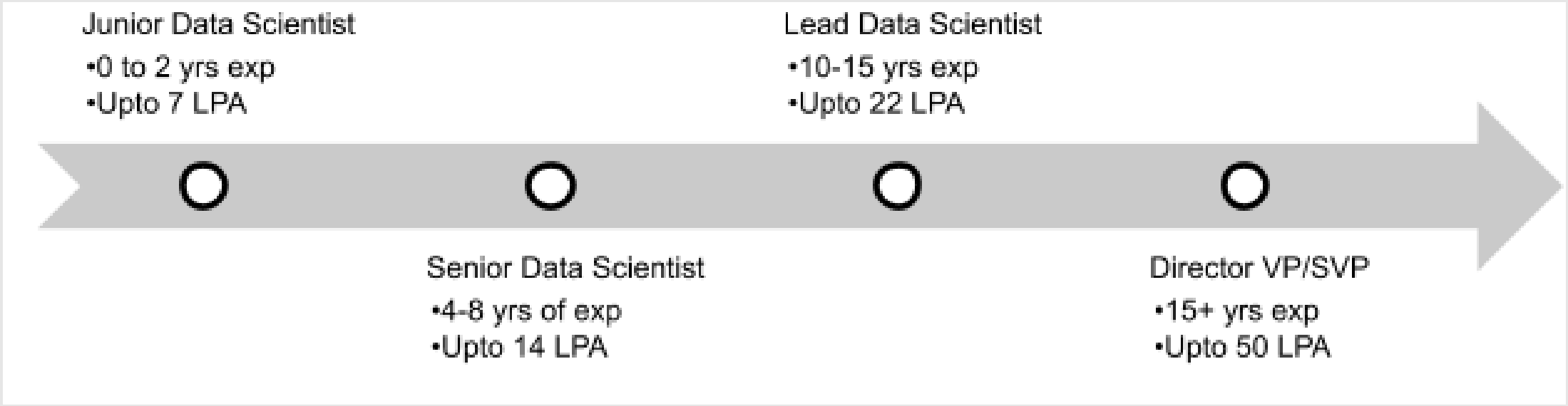
Top jobs seeking machine learning or artificial intelligence skills

Rank	Job title	% of postings containing AI or machine learning	Rank	Job title	% of postings containing AI or machine learning
1.	Machine learning engineer	75.0%	6.	Algorithm developer	46.9%
2.	Deep learning engineer	60.9%	7.	Junior data scientist	45.7%
3.	Senior data scientist	58.1%	8.	Developer consultant	44.5%
4.	Computer vision engineer	55.2%	9.	Director of data science	41.5%
5.	Data scientist	52.1%	10.	Lead data scientist	32.7%

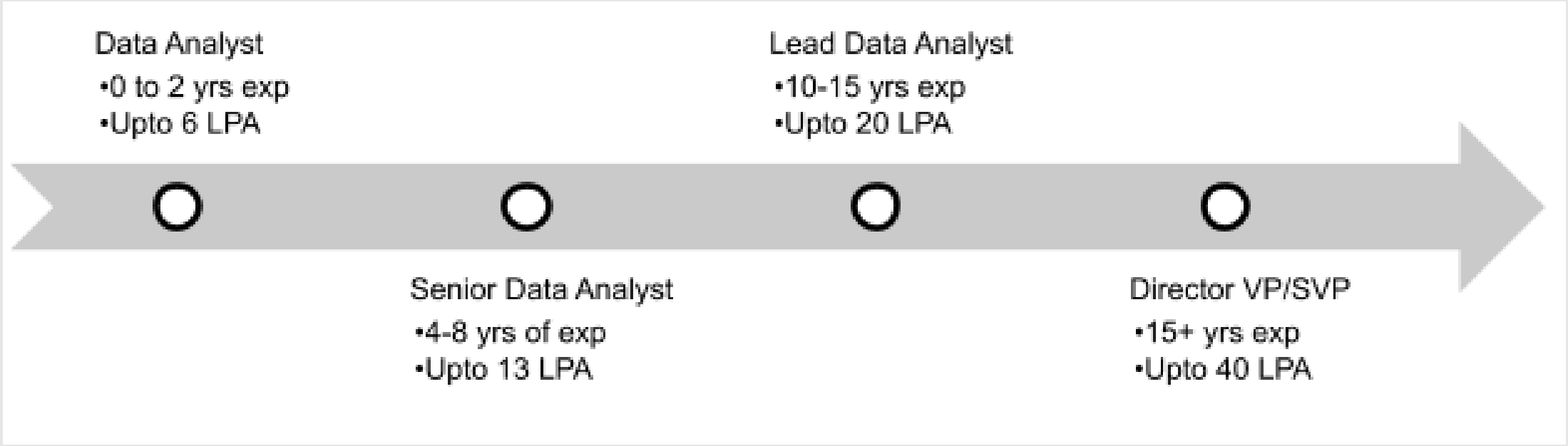
Salary by Career Paths

- Data Scientist
- Data Analyst
- Data Engineer
- Business Intelligence

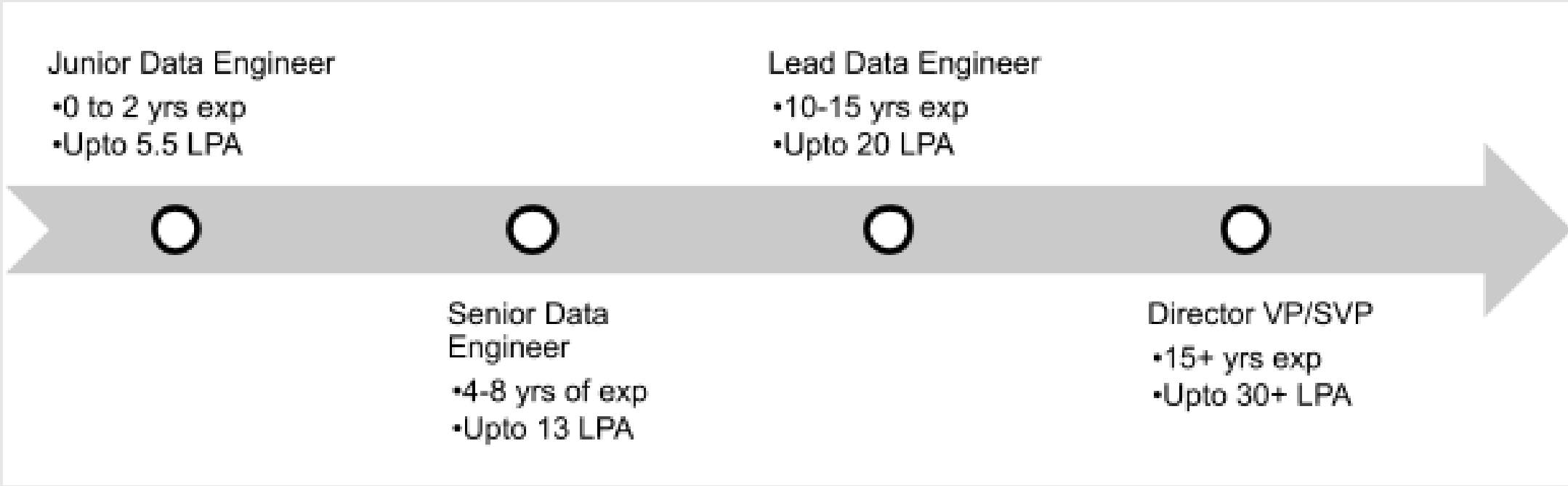
Data Scientist



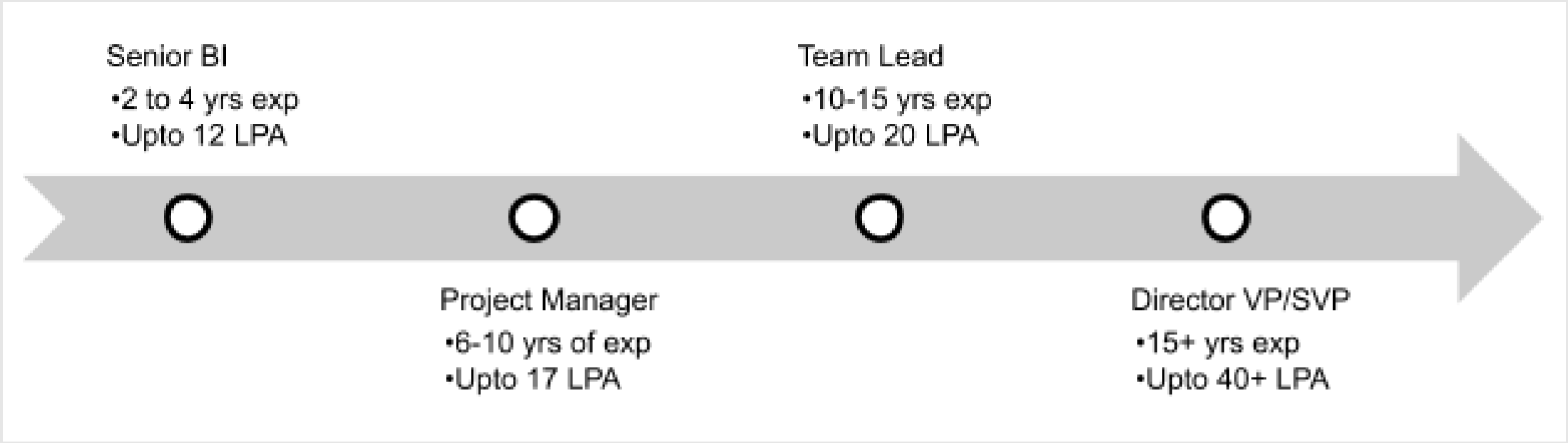
Data Analyst



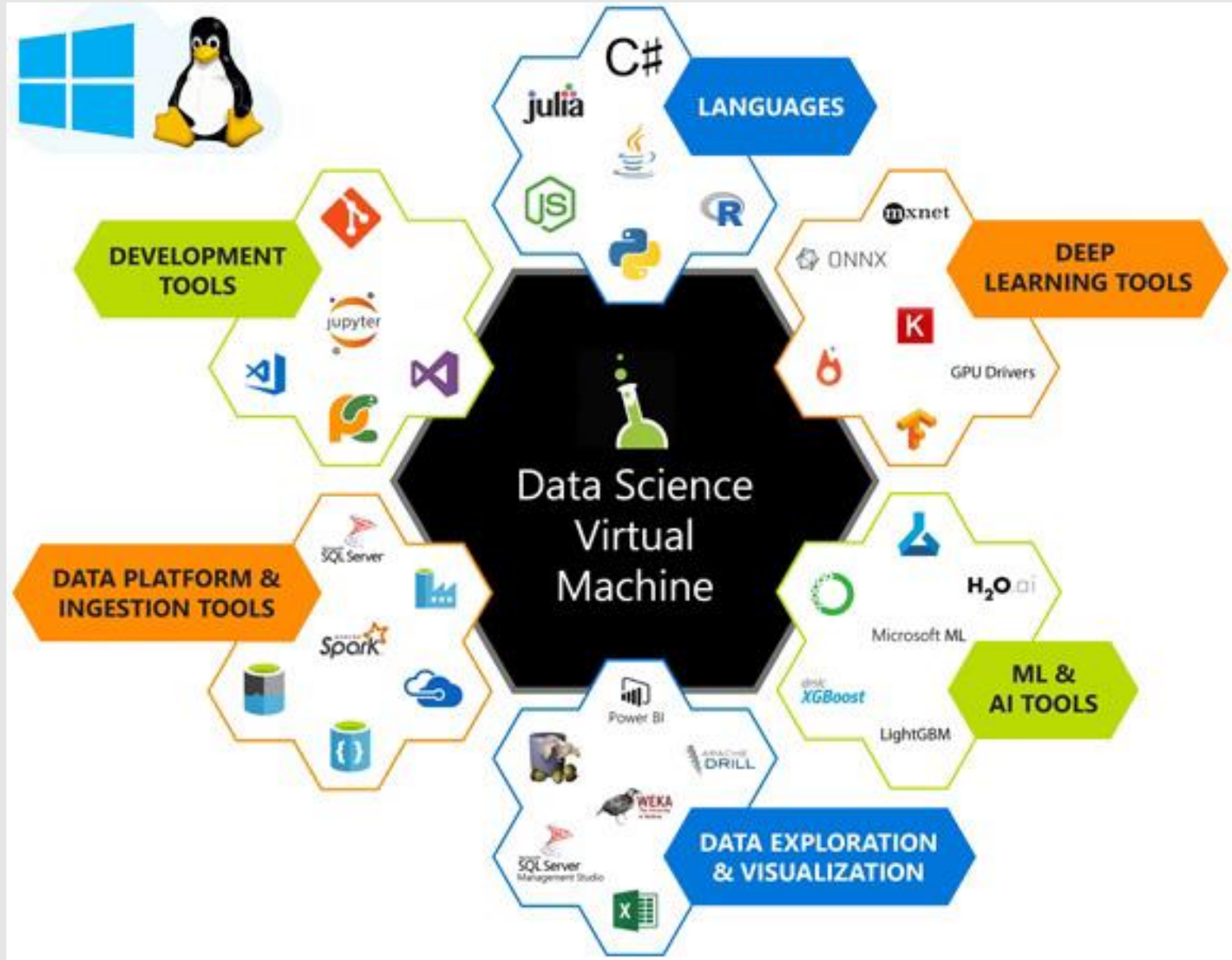
Data Engineer



Business Intelligence Developer



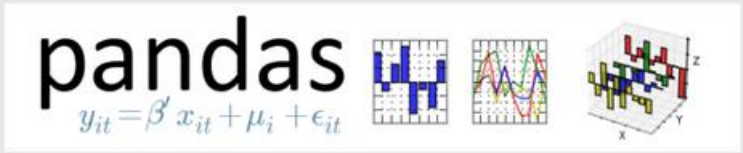
TOOLS


LANGUAGES
C#
julia
DEVELOPMENT TOOLS
jupyter
DEEP LEARNING TOOLS
mxnet
ONNX
GPU Drivers
Data Science Virtual Machine
DATA PLATFORM & INGESTION TOOLS
SQL Server
Spark
ML & AI TOOLS
H2O.ai
Microsoft ML
XGBoost
LightGBM
DATA EXPLORATION & VISUALIZATION
Power BI
DRILL
WEKA
SQL Server Management Studio

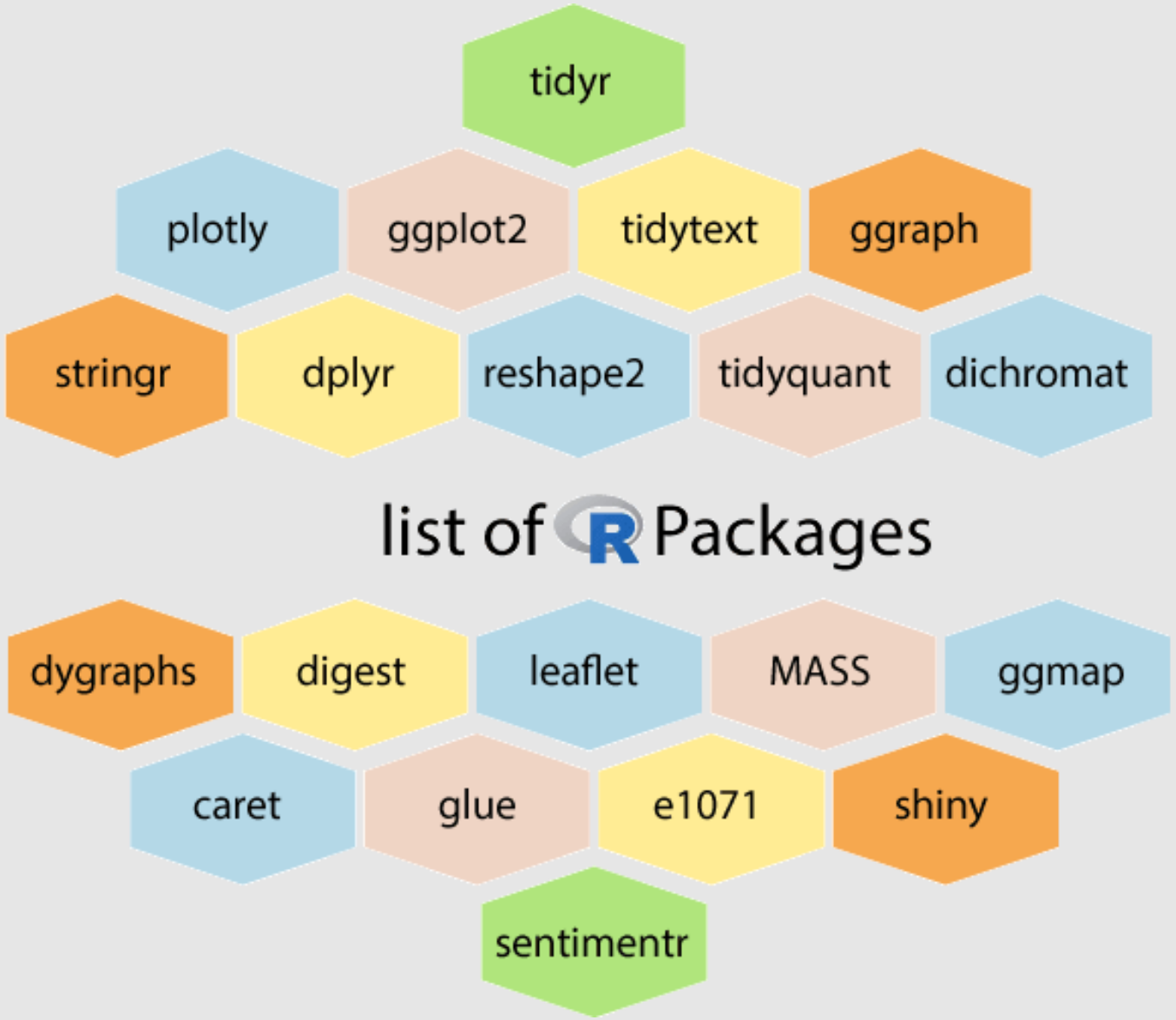
But we will choose few to start with

Python

IP[y]: IPython
Interactive Computing



R


tidyr
plotly
ggplot2
tidytext
ggraph
stringr
dplyr
reshape2
tidyquant
dichromat
list of R Packages
dygraphs
digest
leaflet
MASS
ggmap
caret
glue
e1071
shiny
sentimentr

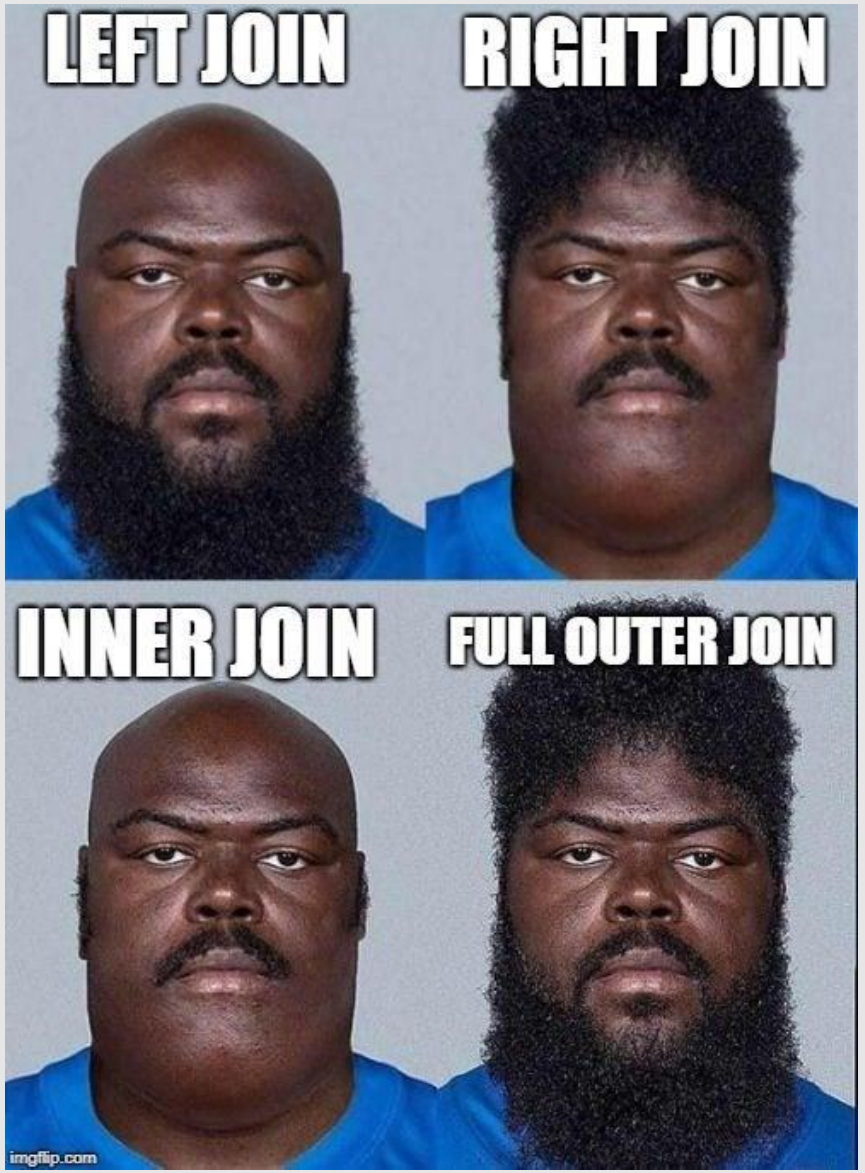
What is SQL?

SQL is an abbreviation for “Structured Query Language”.

SQL is a language used to build database applications that need to query relational databases.

SQL has statements such as CREATE, SELECT, INSERT, UPDATE, DELETE, etc., just like there are statements such as assignment statement, if statement, while statement, etc., in general purpose programming languages such as C, C++ and Java.

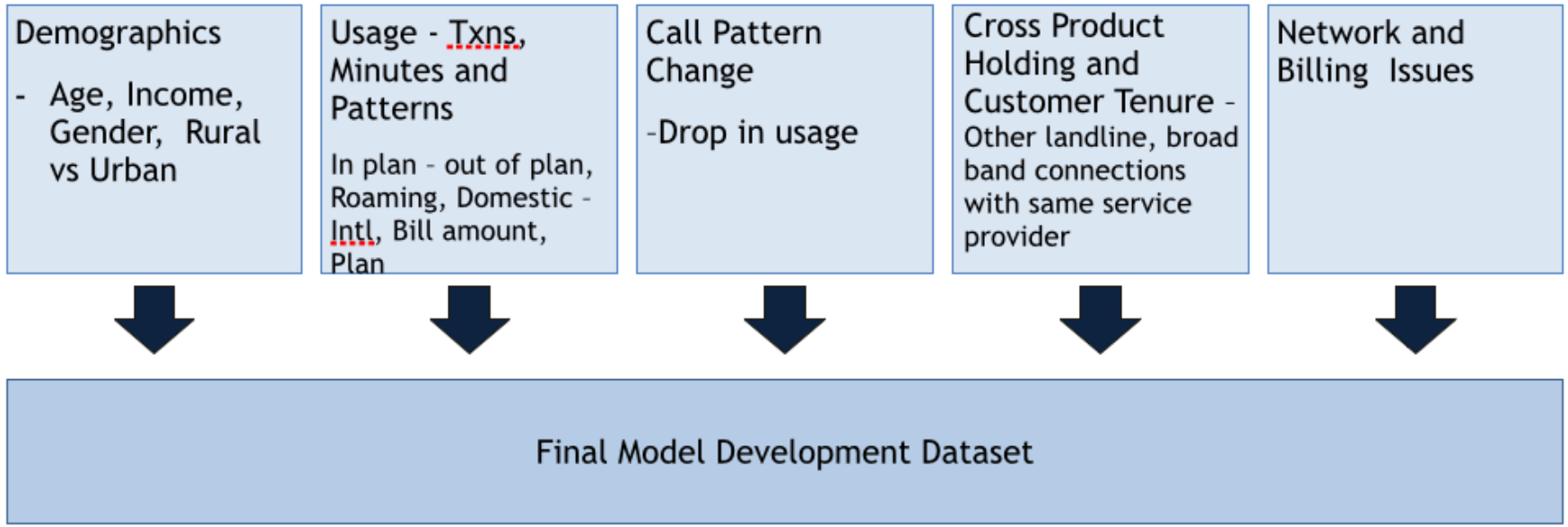
SQL is a language for databases just like C, C++ and Java are languages for general purpose programming.



CASE STUDIES

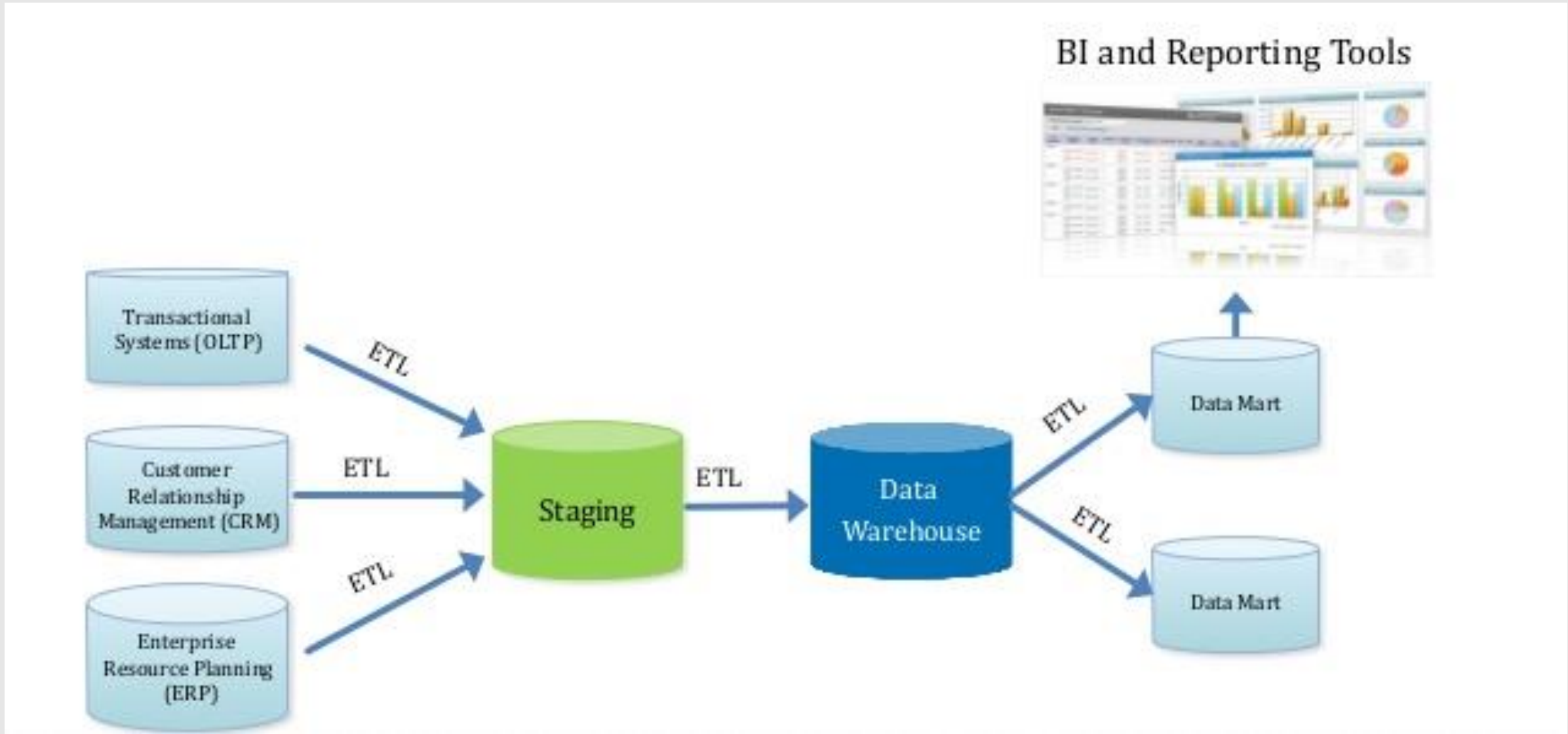
Retail Analytics

Case Study - Telecom Churn Model - Data Preparation

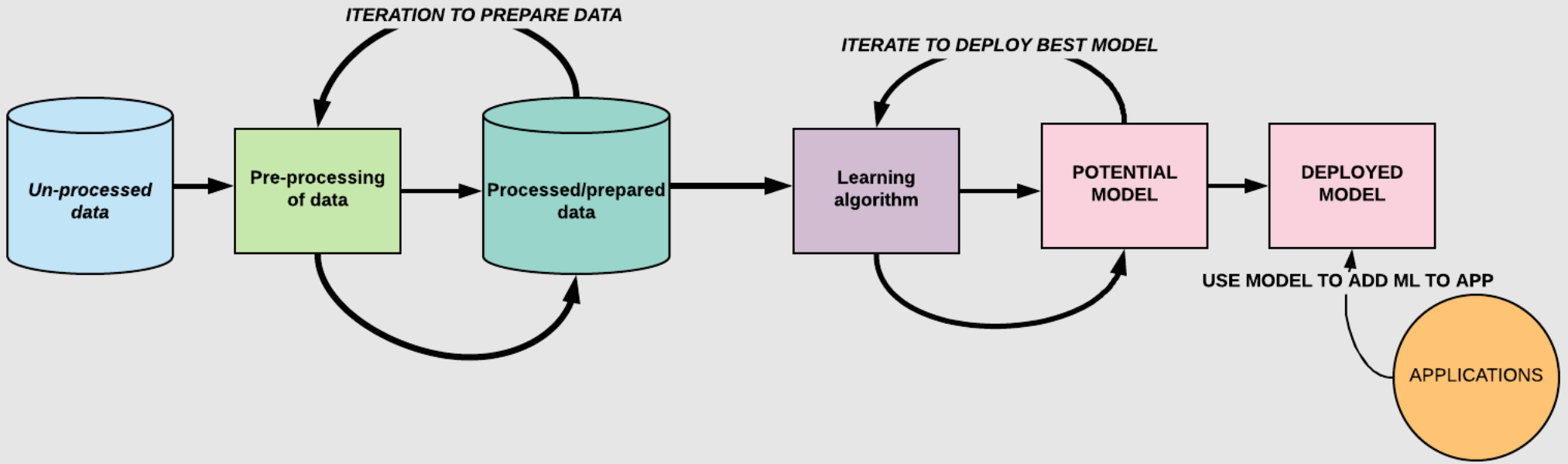


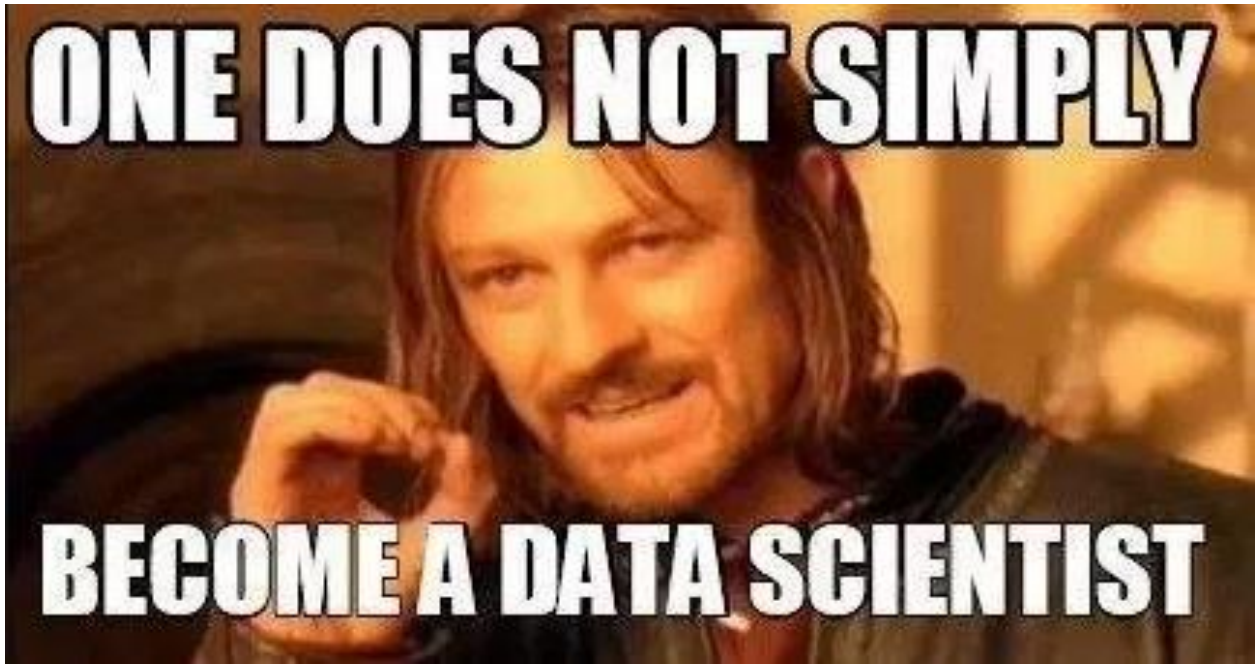
- **Target Variable is Churn**
  - = 1 if Telecom Subscriber has churned
  - = 0 otherwise
- **Monthly Churn Rate is 2%**
- Churn Rate is No. of Churners divided by sum of churners and non churners

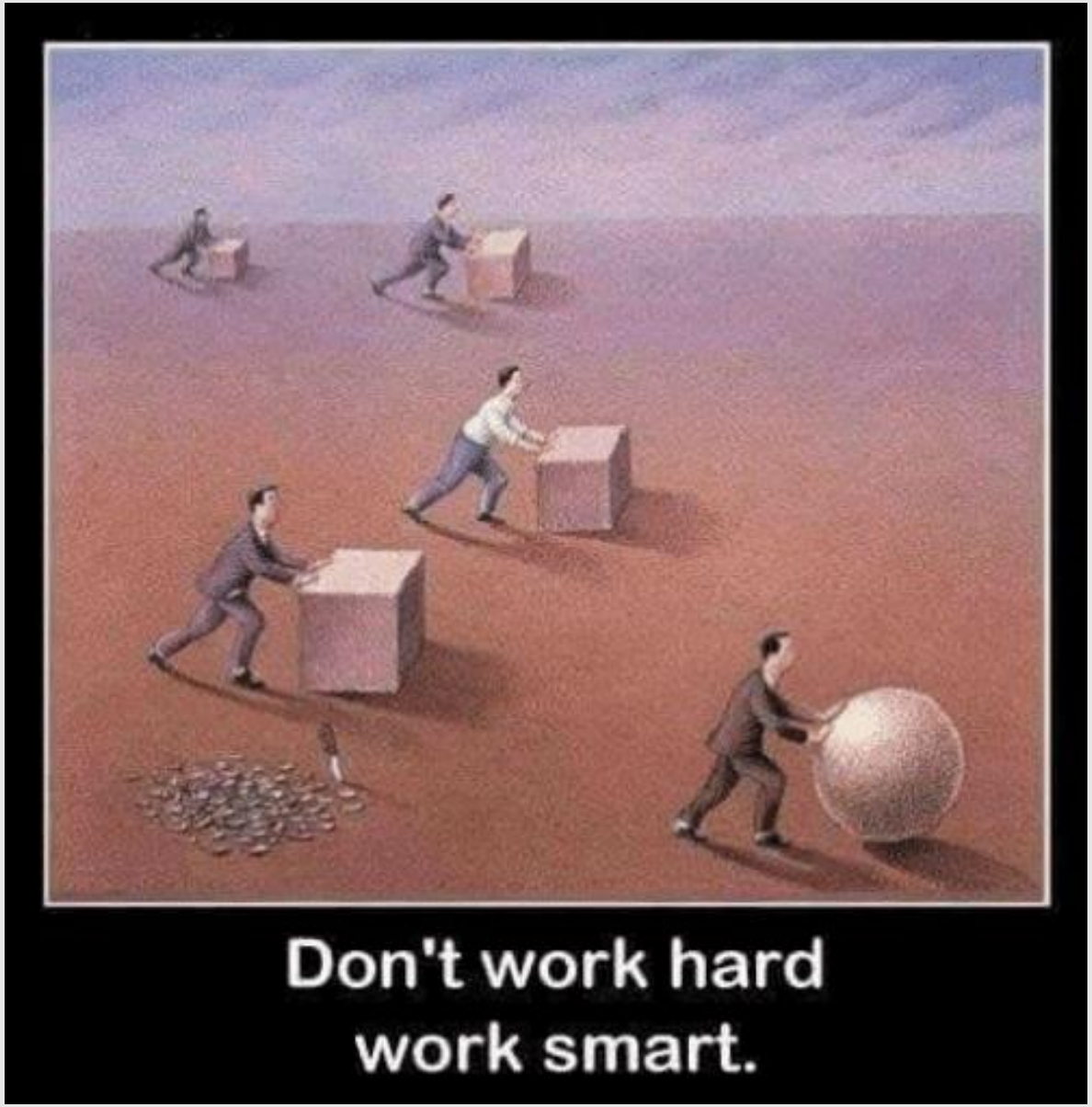
PROCESS OF DATA SCIENCE STEPS


BI and Reporting Tools
Transactional Systems (OLTP)
Customer Relationship Management (CRM)
Enterprise Resource Planning (ERP)
ETL
ETL
ETL
Staging
ETL
Data Warehouse
ETL
ETL
Data Mart
Data Mart

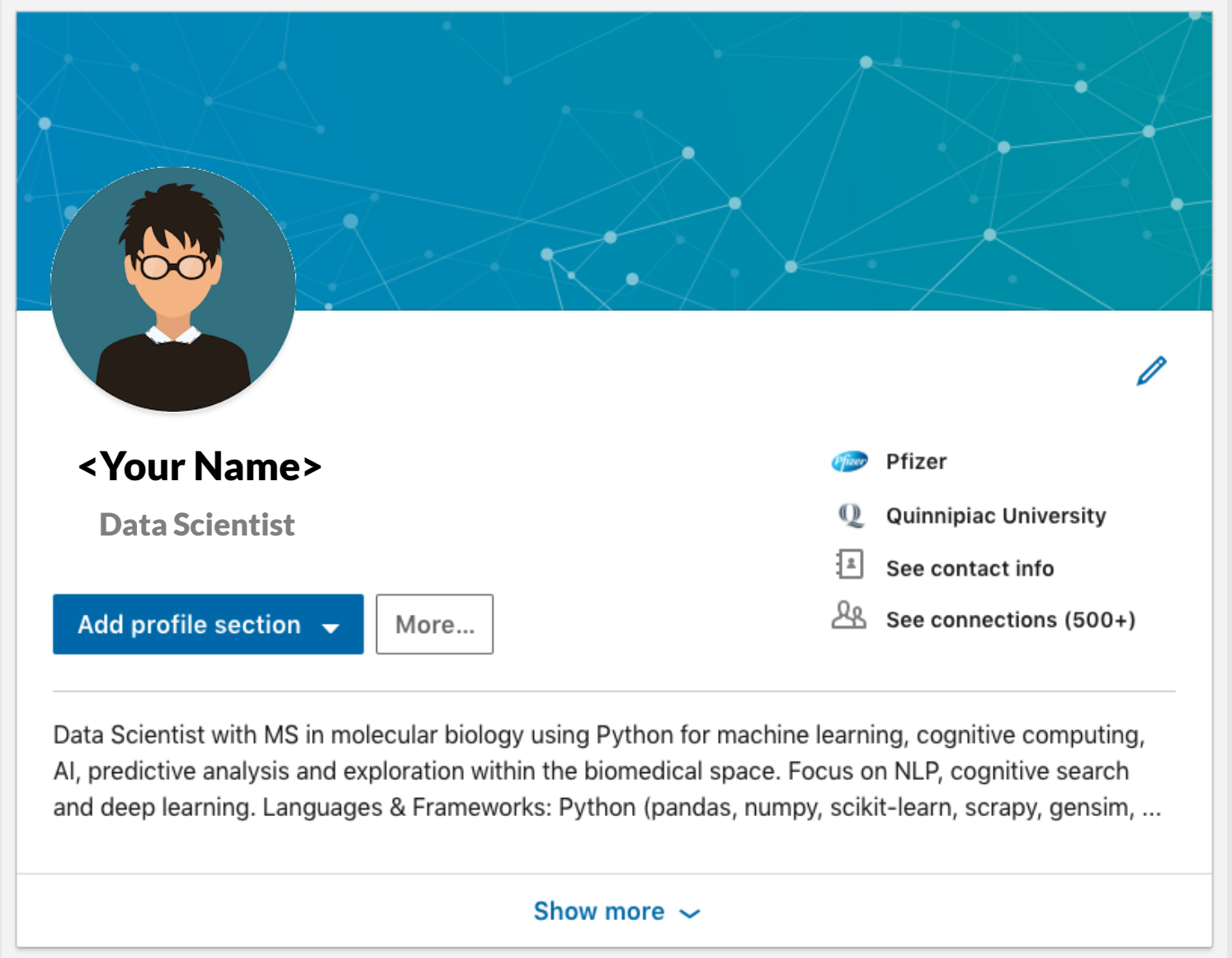
Machine Learning Modelling Process




ONE DOES NOT SIMPLY
BECOME A DATA SCIENTIST


Don't work hard
work smart.

"You can become a Data Scientist if you started your journey today"



HAVE QUESTIONS DO YOU?