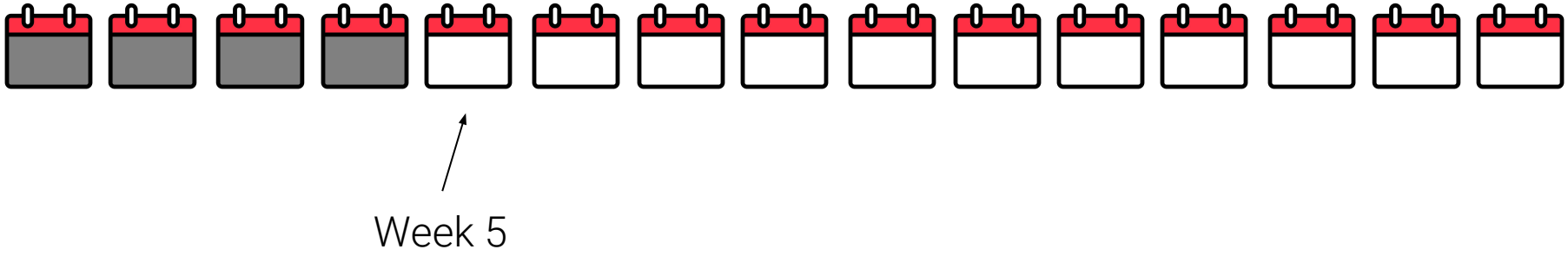
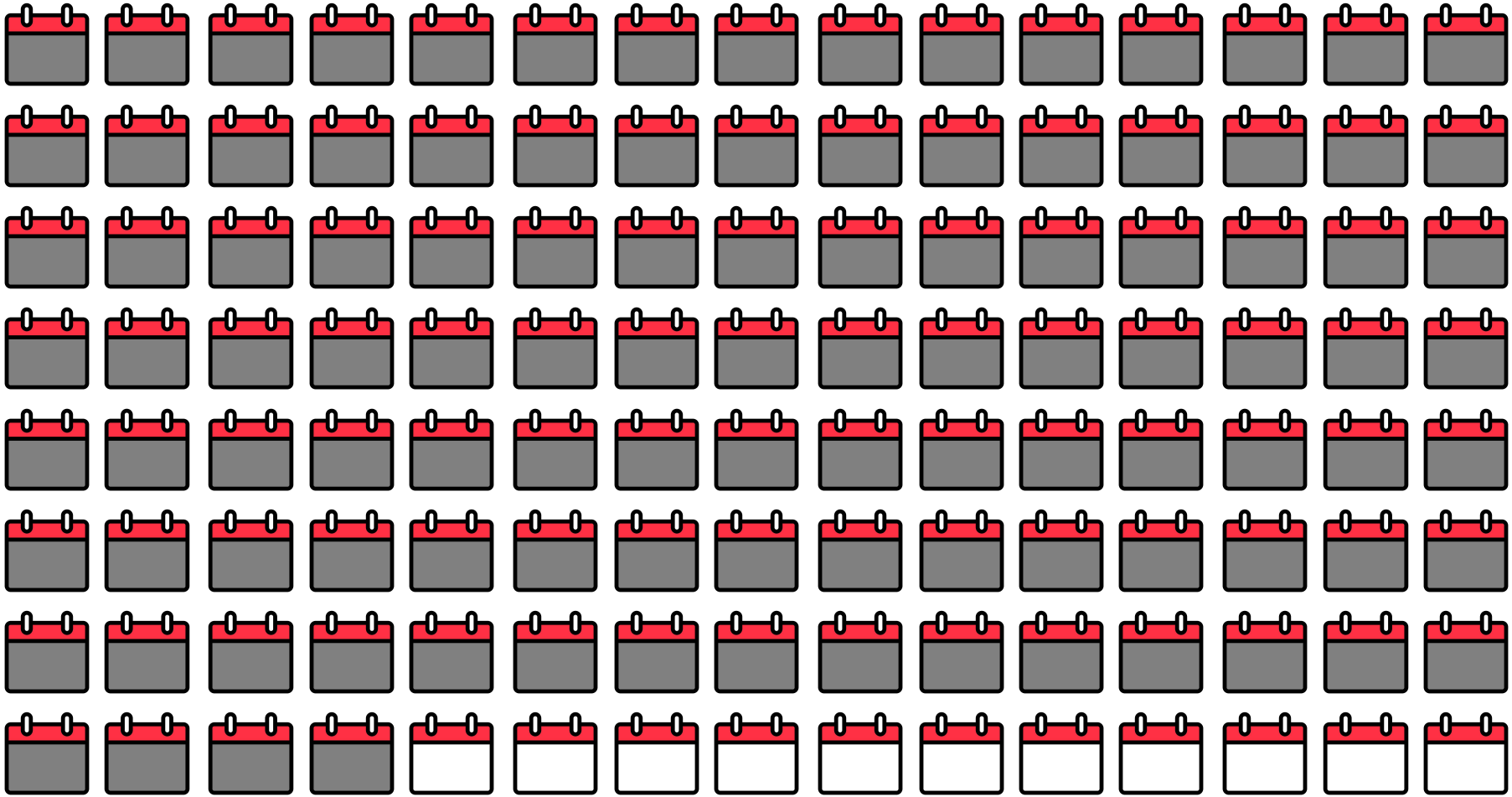
Business
Technology
Group

Introduction to Data Science

February 11th, 2019

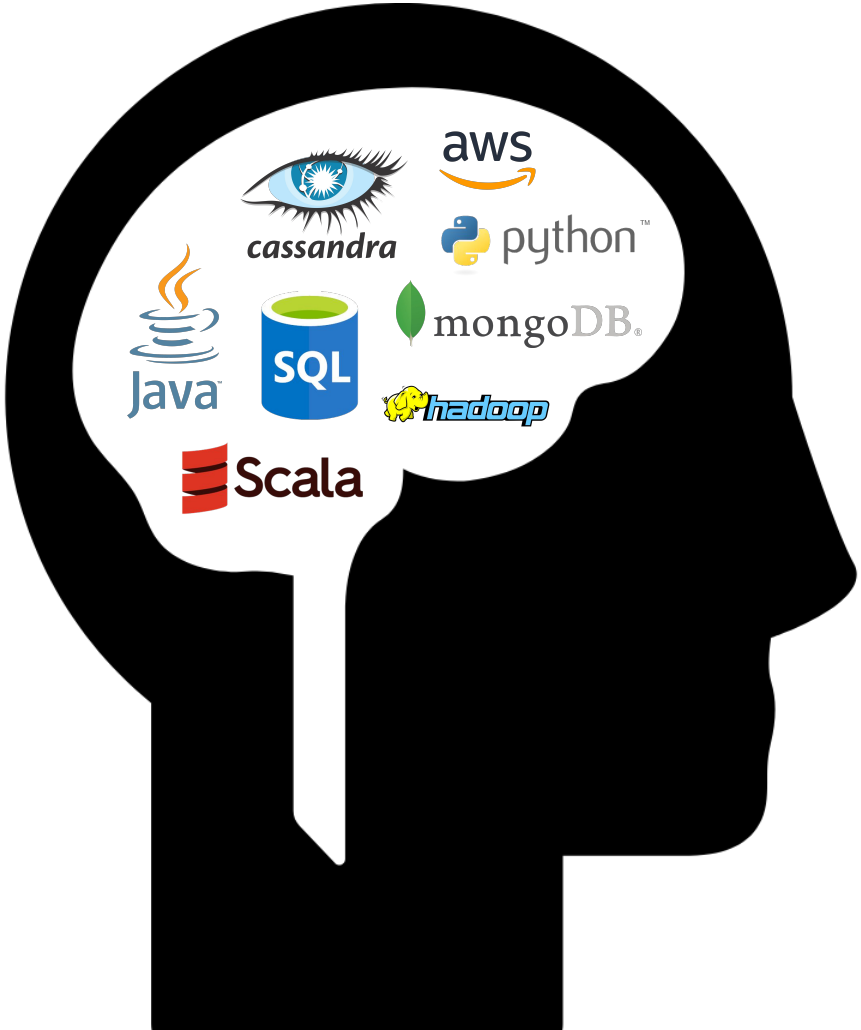

Week 5



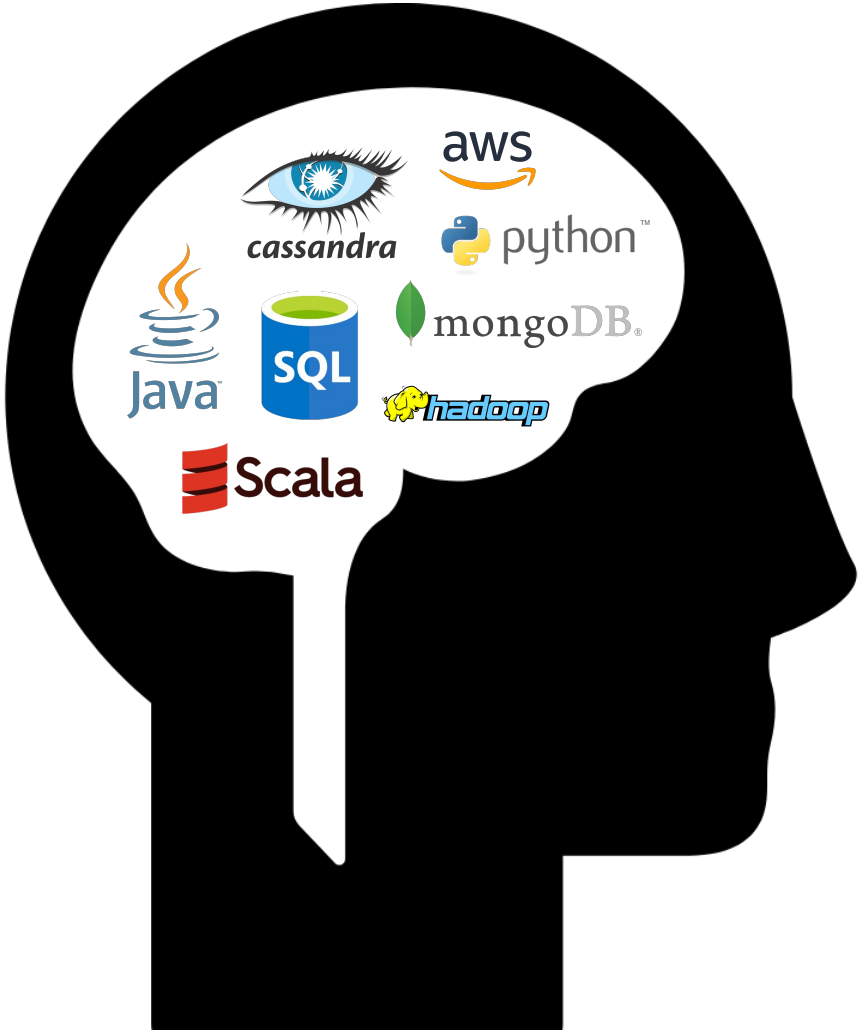
What is Data Science?

DETECTION
SOCIAL MEDIA
SERVICES
PROJECTS
WWW
BIG DATA
MULTIMEDIA
NETWORK
PREDICTIVE
PROGRAM
ANALYTICS
CONTENT
COMPUTING
PROMOTION
PROCESSING
CONSUMER
ORGANIZATION
PLANNING
EVENTS
PROGRAMMING
CODING
SOFTWARE
INFORMATION
TECHNOLOGY
MODELS
E-MARKETING
COMMUNICATION
BRANDING
COMPUTER
CONSUMER DEMAND MARKETS
WEB MARKETING
DATA MINING
DATA SCIENCE
INFORMATION
MACHINE LEARNING
VISION
ENGINEERING
RESEARCH
PROBABILITY
KDD
WEB SERVICES
BIG
DATA
SERVICE
WEB DEV
STRATEGY
WORLDWIDE
COMPUTING
MATHS
PATTERN
ENGINEERING
VISUALIZATION
BIG DATA
SOLUTIONS
PLANNING
MEDIA
PROJECTS
MOBILE
PRO
PRICING
SEGMENTATION
SOCIAL NETWORK
STATISTICS
INFORMATION
DIGITAL
CODING
SOCIAL NETWORKS
TARGET

Areas of Focus



Data Engineer



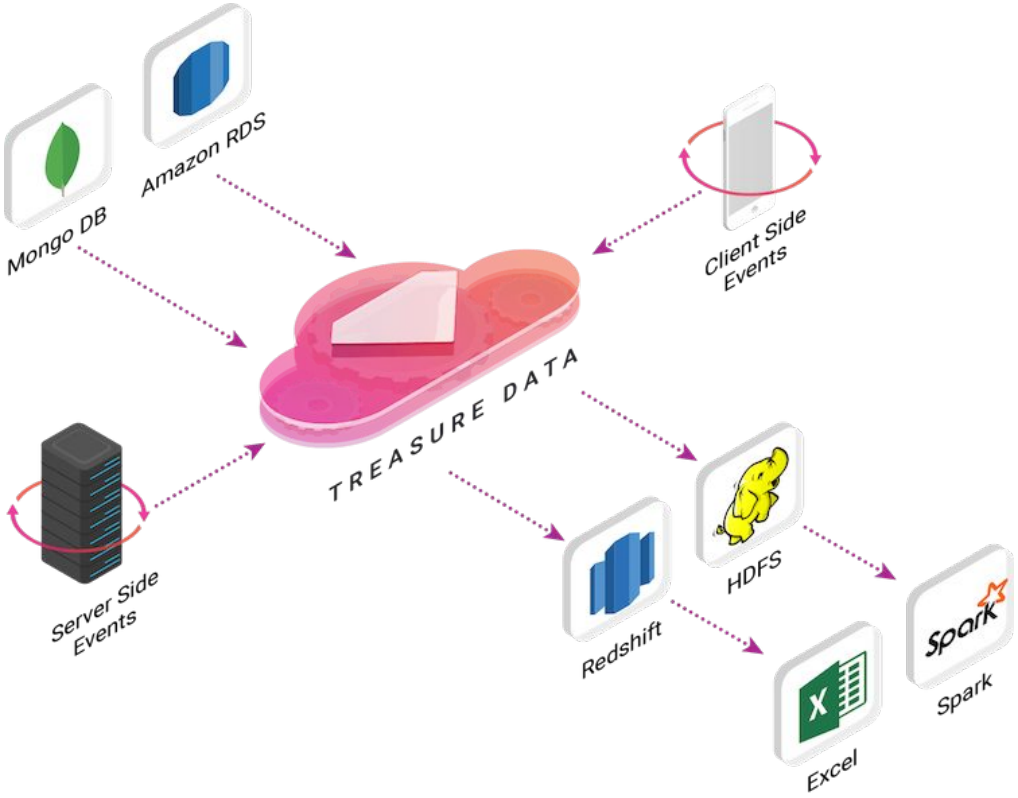
“Data Wrangling”

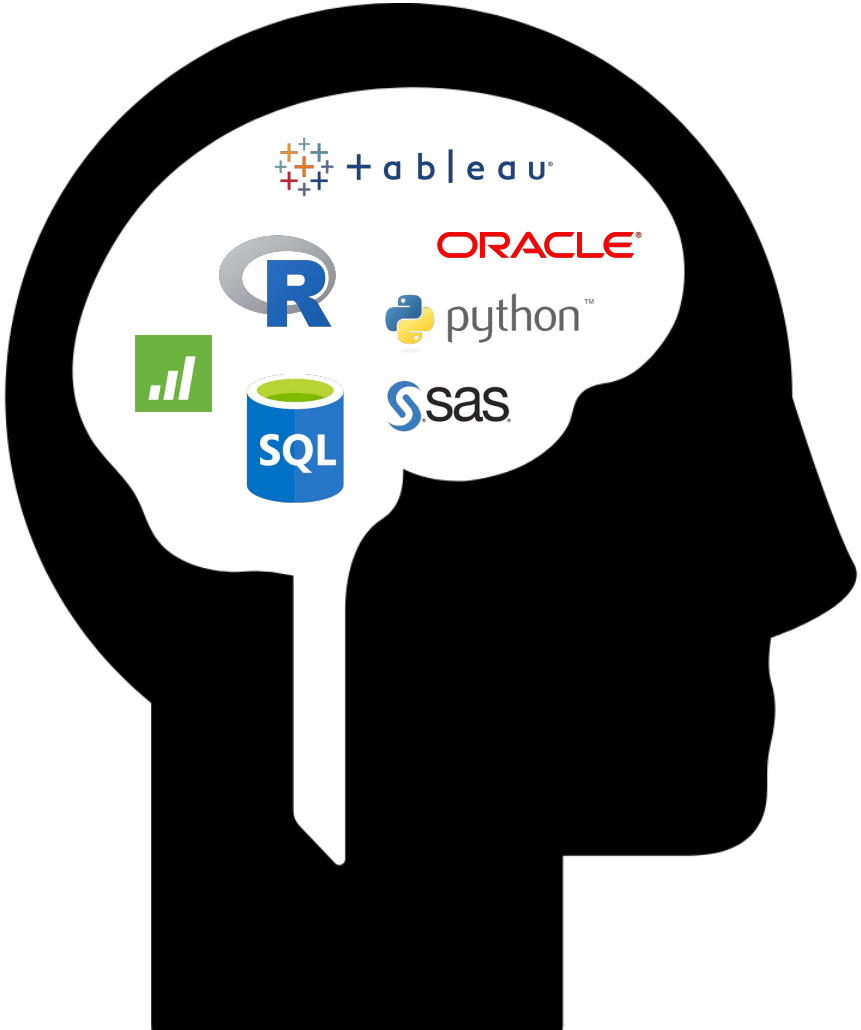
“Data Cleaning”

Data Engineer

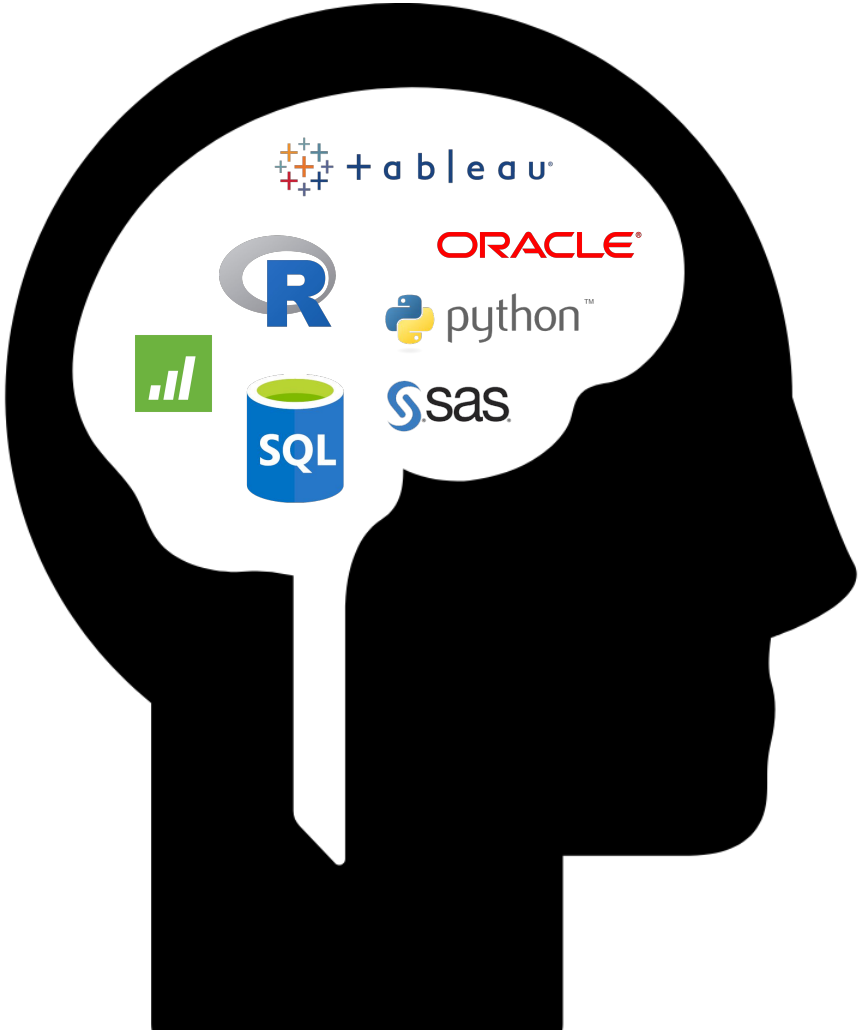
“Infrastructure”

“Pipeline”


Mongo DB
Amazon RDS
Client Side Events
TREASURE DATA
Server Side Events
HDFS
Redshift
Spark
Excel



Data Analyst



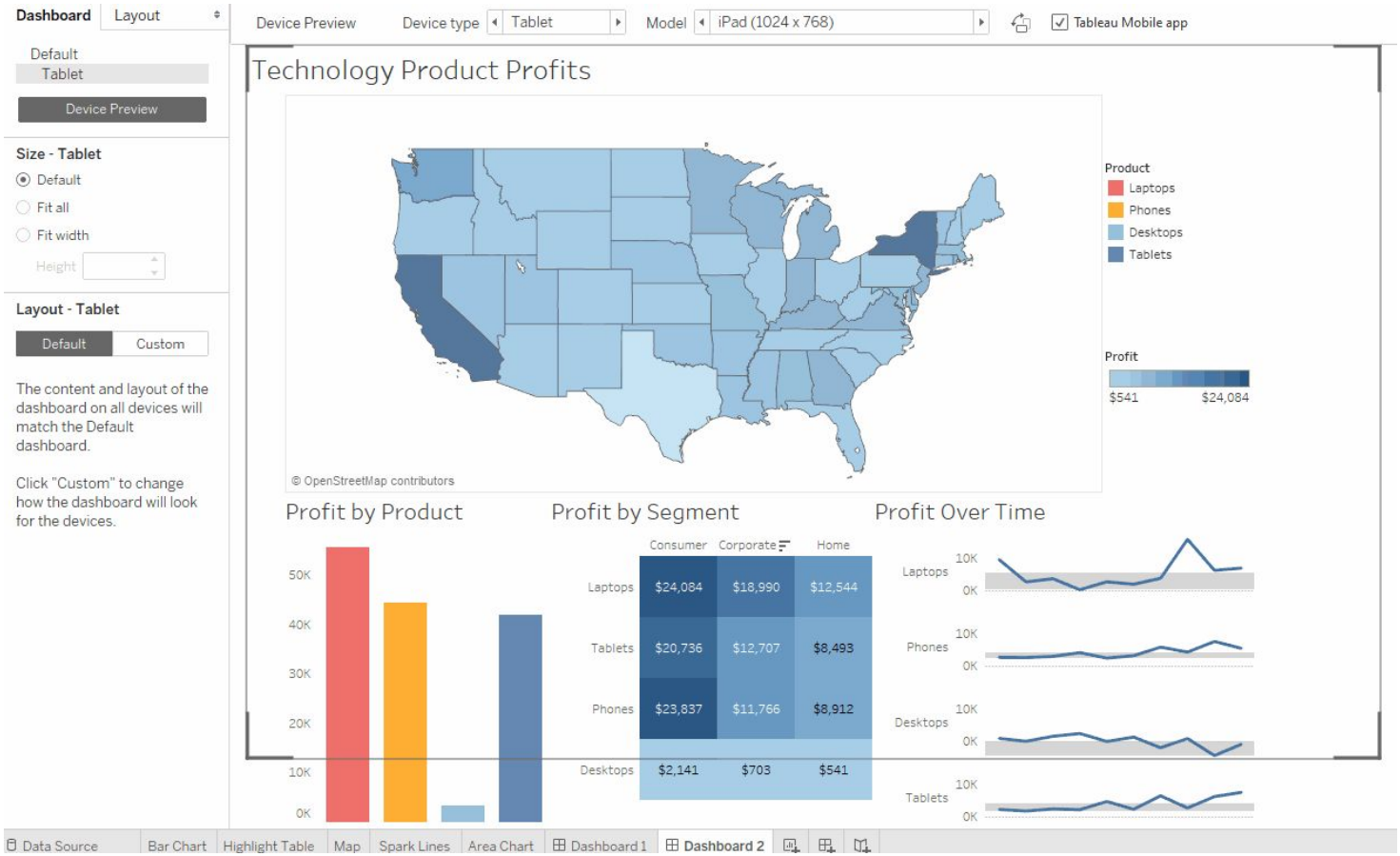
“Dashboard”

“Reports”

Data Analyst

aka Business Intelligence Analyst

“Spreadsheets”


Dashboard
Layout
Default
Tablet
Device Preview
Size - Tablet
Default
Fit all
Fit width
Height
Layout - Tablet
Default
Custom
The content and layout of the dashboard on all devices will match the Default dashboard.
Click "Custom" to change how the dashboard will look for the devices.
Device Preview
Device type
Tablet
Model
iPad (1024 x 768)
Tableau Mobile app
Technology Product Profits
© OpenStreetMap contributors
Product
Laptops
Phones
Desktops
Tablets
Profit
$541
$24,084
Profit by Product
50K
40K
30K
20K
10K
0K
Profit by Segment
Consumer
Corporate
Home
Laptops
$24,084
$18,990
$12,544
Tablets
$20,736
$12,707
$8,493
Phones
$23,837
$11,766
$8,912
Desktops
$2,141
$703
$541
Profit Over Time
Laptops
Phones
Desktops
Tablets
10K
0K
Data Source
Bar Chart
Highlight Table
Map
Spark Lines
Area Chart
Dashboard 1
Dashboard 2

Data Scientist

“Building Models”

“Data Mining”

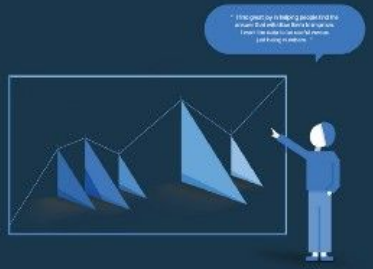
Data Scientist

“Quantitative”

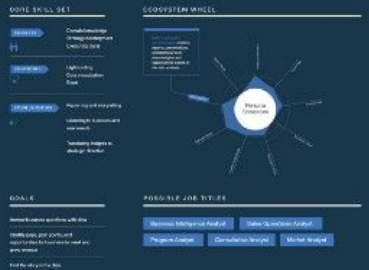
“AB Testing”

Cancel

Business Analyst

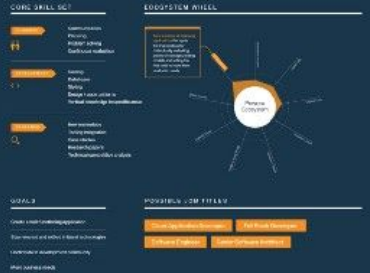


- **Eden, the Business Analyst,** approaches data from different angles and applies analytics models to confidently support or dispel assumptions.



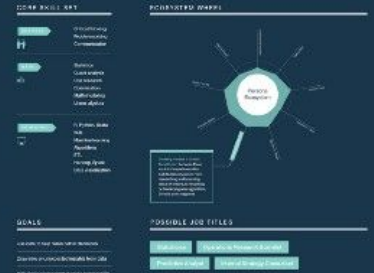
App Developer

- **Kai, the Application Developer,** builds applications that interact with data and implements data models.



Data Scientist

- **Chris, the Data Scientist** gets deep into the data to draw hidden insights and influence business decisions.

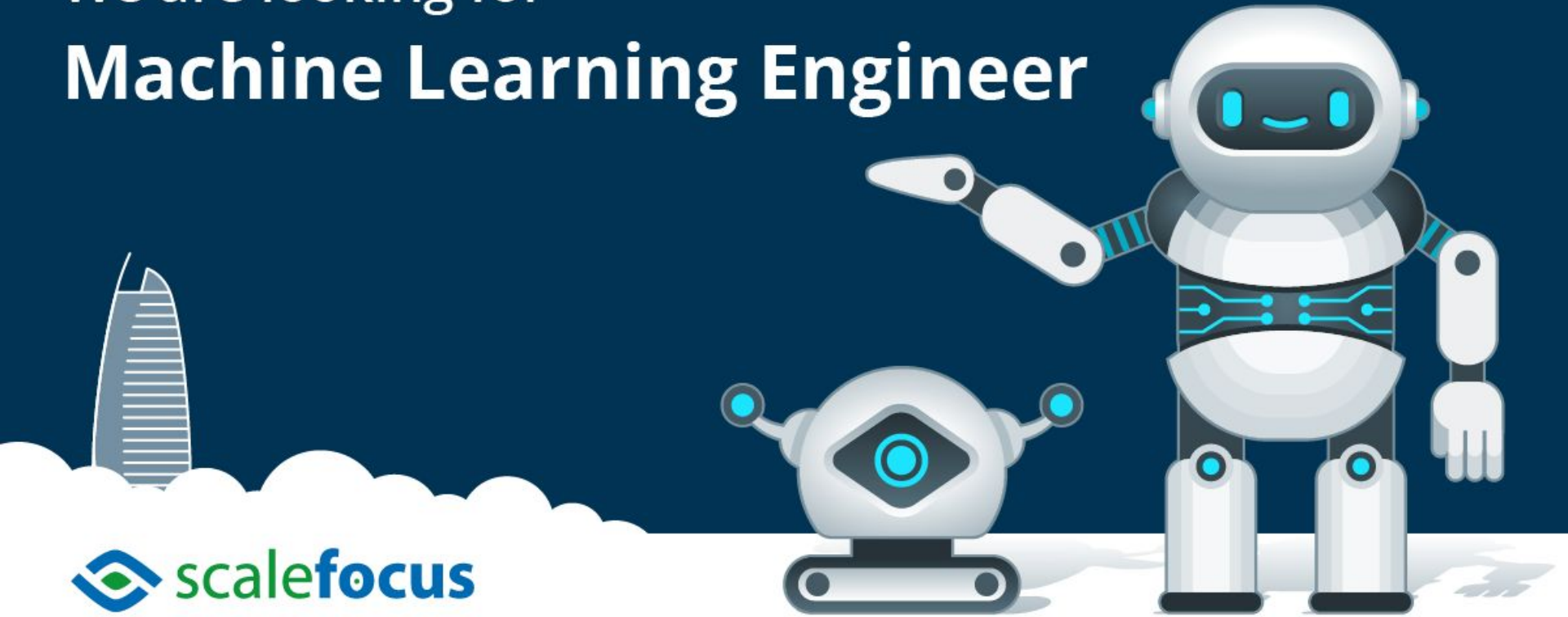


Data Engineer

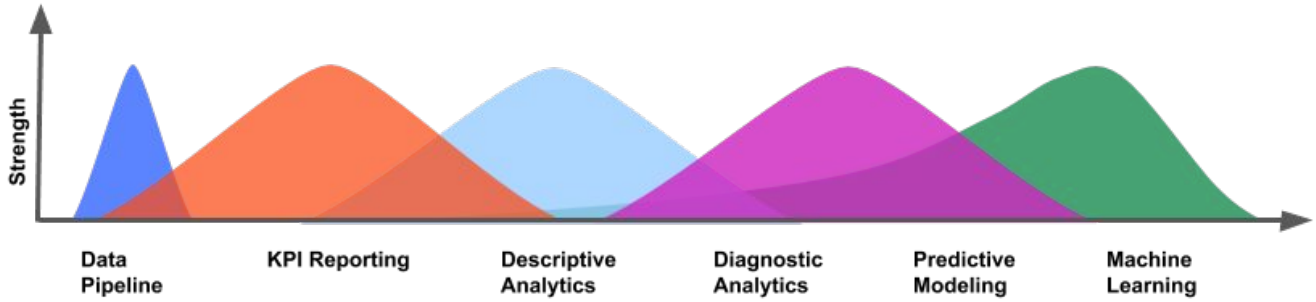
- **Harley, the Data Engineer,** builds and maintains a scalable data infrastructure to make relevant data available to teams.

We are looking for

Machine Learning Engineer



scalefocus



- **Data Engineers** perform ETL and maintain our data pipeline
- **BI Developers** deliver accurate, consistent KPIs from source to output (data pipeline)
- **BI Analysts** focus on analyses around metric performance
- **Product Scientists** use applied stats to support A/B testing and continuous product improvement
- **Data Scientists** build models and deploy their solutions to the product itself

Data Engineers BI Developers BI Analysts Product Scientists Data Scientists

indeed

Exploratory Data Analysis

30%
30%
30%
45%
30%
30%

Machine Learning

Supervised Learning

Unsupervised Learning

Machine Learning

Reinforcement Learning

Supervised Learning

With Labels

Unsupervised Learning

Without Labels

Machine Learning

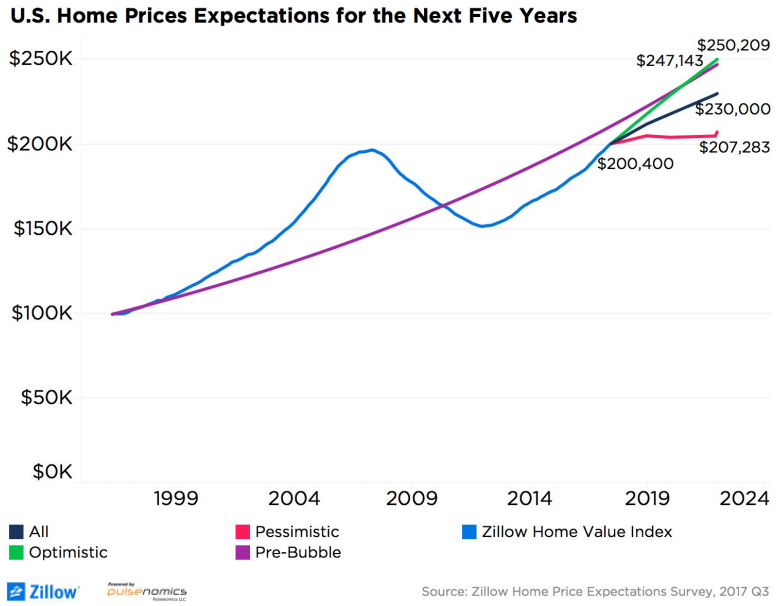
Reinforcement Learning

AlphaGo

Supervised Learning

With Labels

Regression



Supervised Learning

With Labels

Classification

Relationship

“Dog”

Relationship

“Cat”

Essential Tools

+

pandas
$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$

+

jupyter

R
Studio®

The Data

CelebFaces Attributes

kaggle.com/jessicali9530/celeba-dataset

CelebFaces Attributes

* For binary variables, 1 = yes
-1 = no

- 202,599 face images of celebrities
- 10,177 unique identities
- 5 landmark locations
- 40 binary attributes per image (facial features, pose, expression)

# 5_o_Clo...	# Arched...	# Attracti...	# Bags_U...	# Bald	# Bangs
-1	1	1	-1	-1	-1
-1	-1	-1	1	-1	-1
-1	-1	-1	-1	-1	-1
-1	-1	1	-1	-1	-1
-1	1	1	-1	-1	-1
-1	1	1	-1	-1	-1
1	-1	1	1	-1	-1
1	1	-1	1	-1	-1
-1	1	1	-1	-1	1
-1	-1	1	-1	-1	-1
-1	-1	1	-1	-1	-1
-1	-1	1	1	-1	-1

Applications

Various applications for computer vision, deep learning

- Logging into your phone with your face
- Searching through surveillance images for a particular suspect

Questions

- Can you train a model that can detect specific facial attributes?
- Which images contain people that are smiling?

https://tinyurl.com/jndemo

Housing Prices

kaggle.com/harlfoxem/housesalesprediction

Demo

tinyurl.com/BTG-S19-DS