



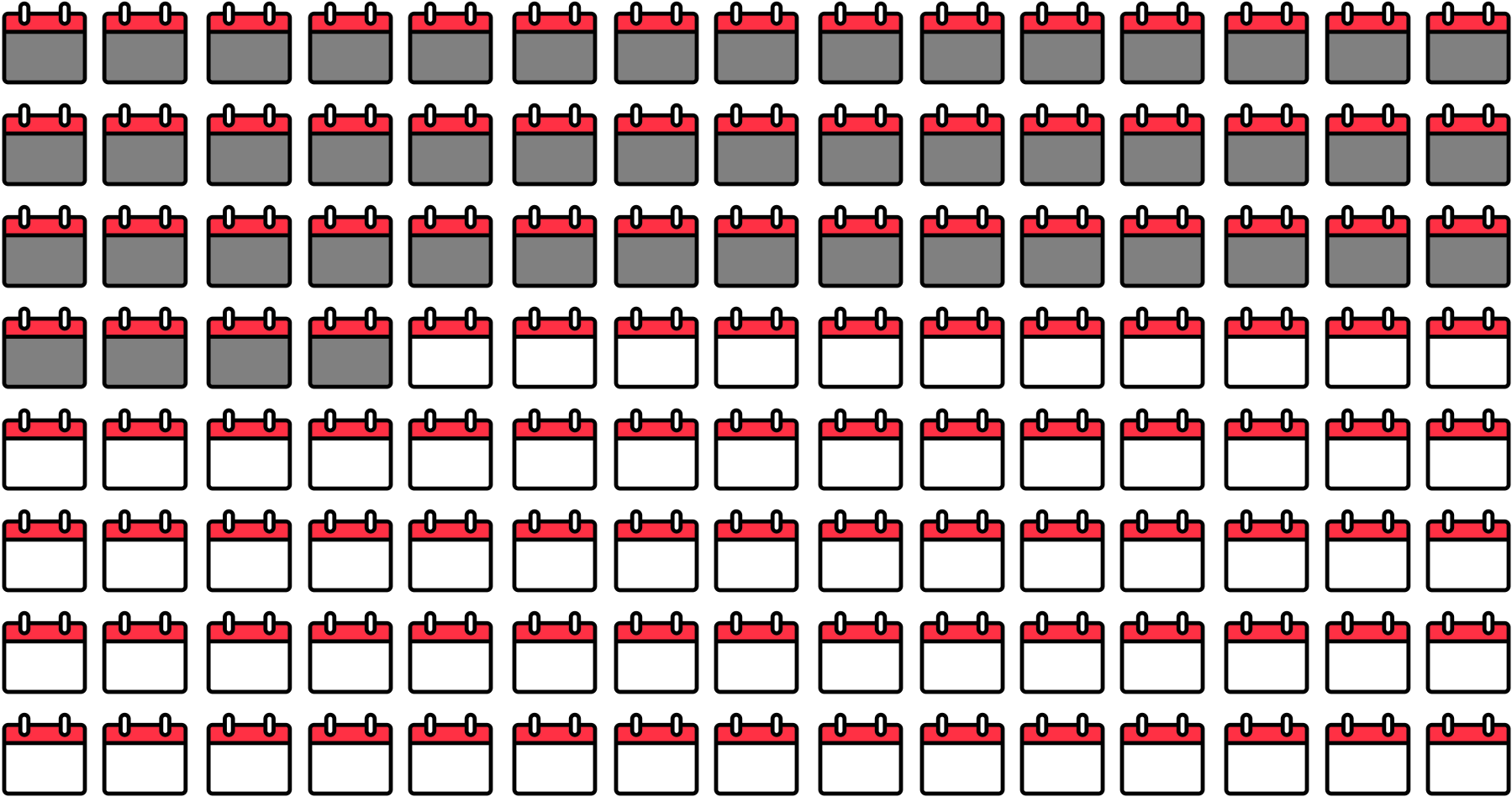
# Business Technology Group

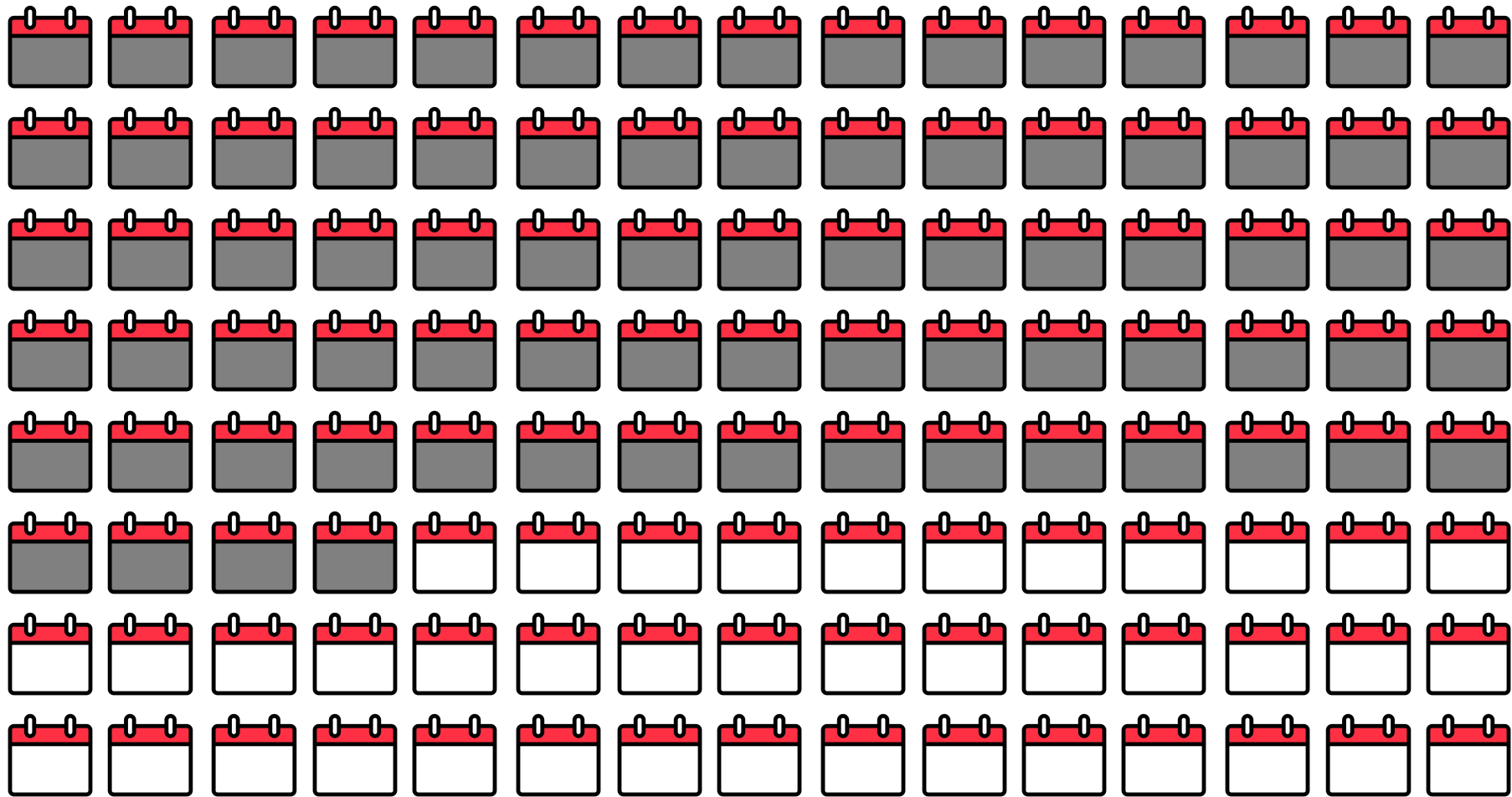
Introduction to Data Science  
February 11th, 2019

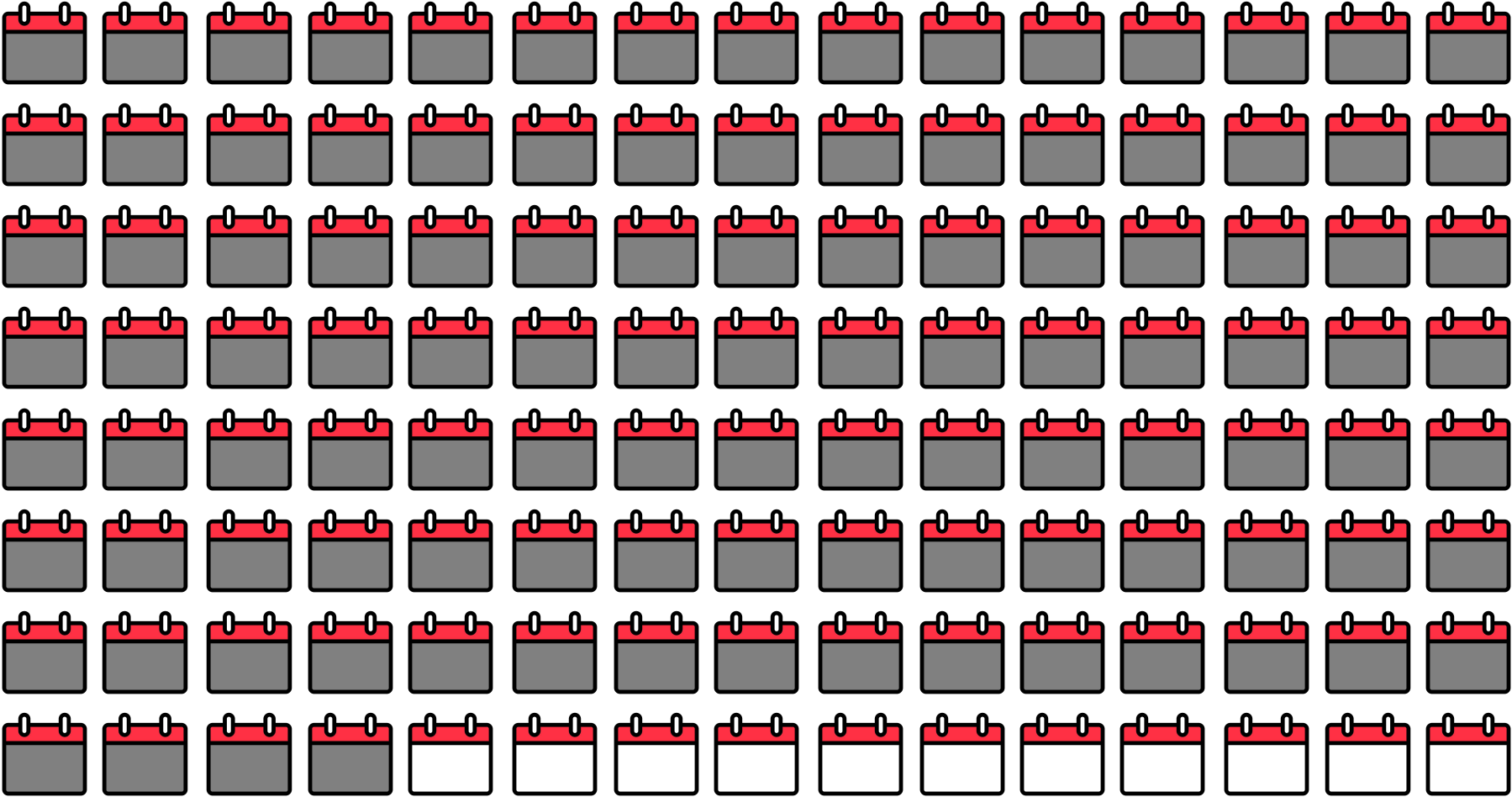


Week 5









What is Data Science?

A word cloud centered around the phrase "DATA SCIENCE", which is the largest and most prominent text. Surrounding it are numerous other terms related to data science, machine learning, and technology, arranged in various sizes and orientations. The words are all in shades of blue. The overall shape of the cloud is roughly rectangular, with words filling the space around the central title.

**DATA SCIENCE**

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





# Areas of Focus



Data Engineer



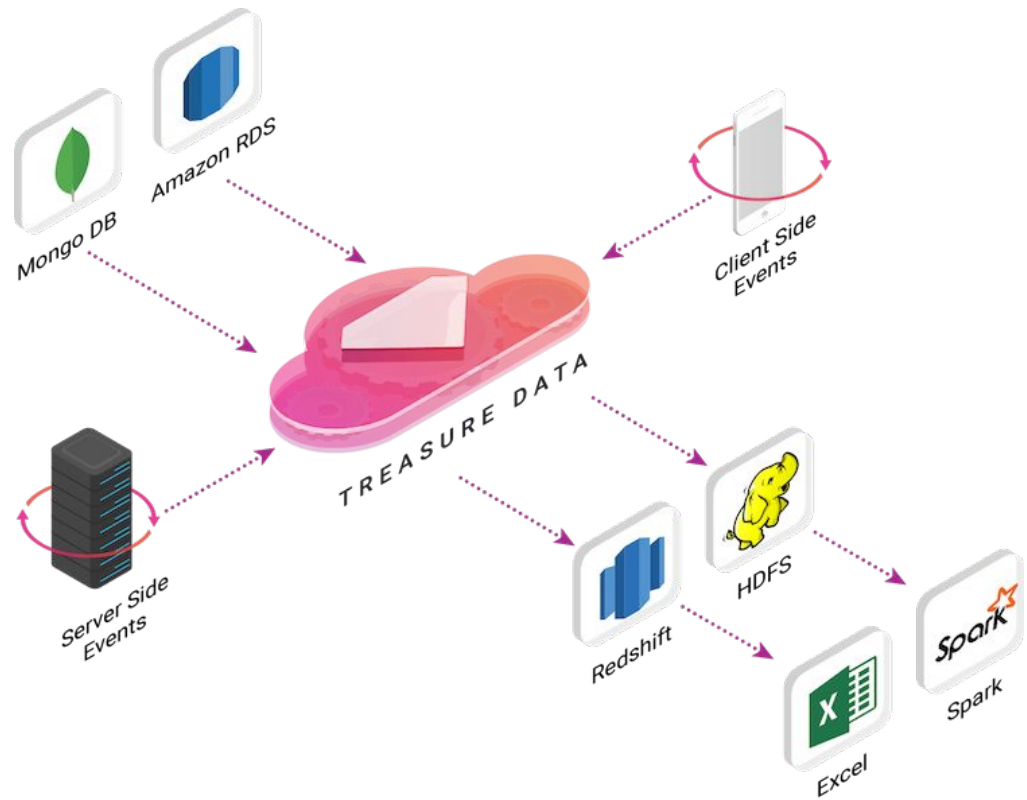
“Data Wrangling”

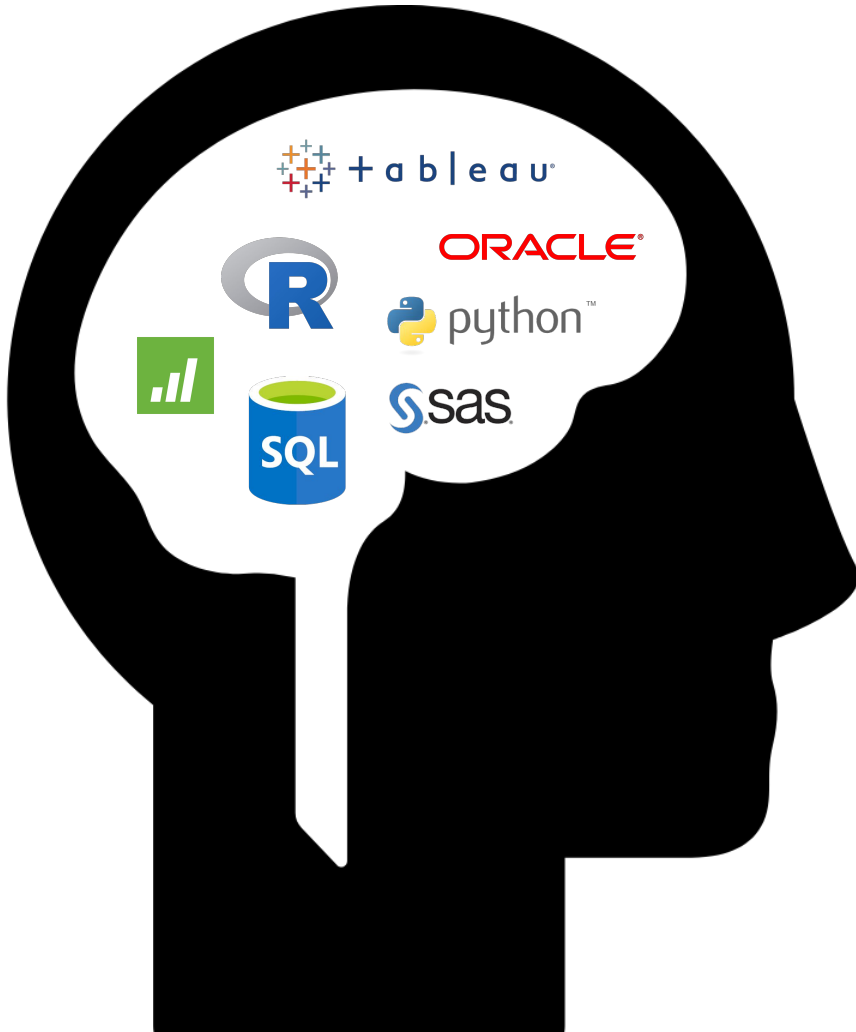
“Data Cleaning”

Data Engineer

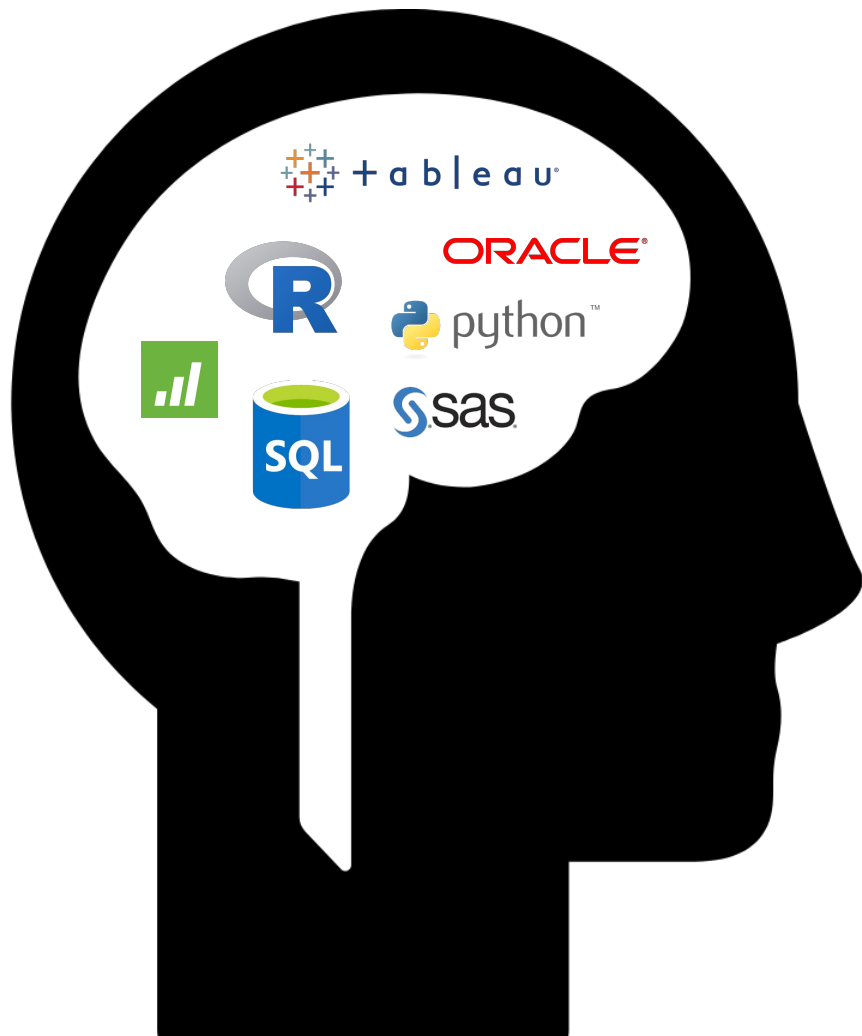
“Infrastructure”

“Pipeline”





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

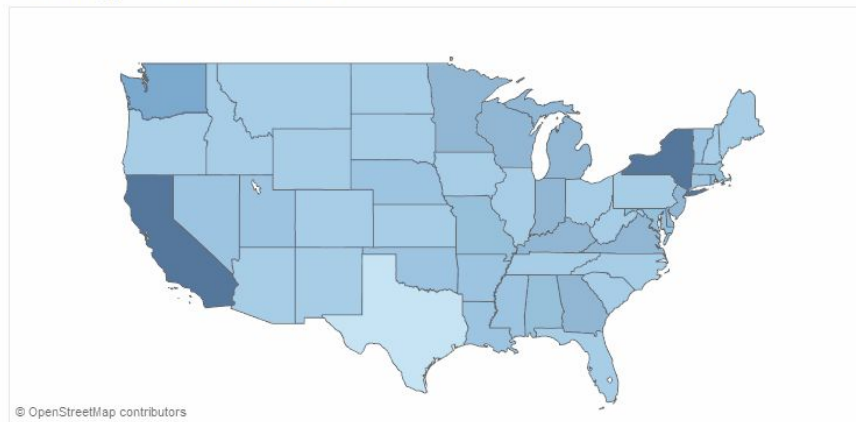
Tablet

Model

iPad (1024 x 768)

☒ Tableau Mobile app

## Technology Product Profits



Product

Laptops

Phones

Desktops

Tablets

Profit

\$541

\$24,084

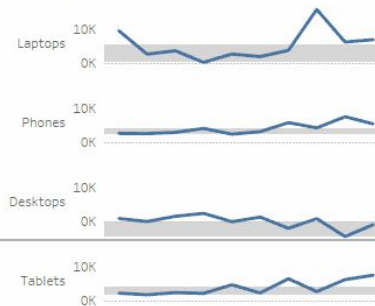
## Profit by Product



## 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



Data Source

Bar Chart

Highlight Table

Map

Spark Lines

Area Chart

Dashboard 1

Dashboard 2

Map

Table

List





Data Scientist



“Building Models”

“Data Mining”

Data Scientist

“Quantitative”

“AB Testing”



# Business Analyst

"The greater the number of people involved in an activity, the more likely it is to be successful. I have the ability to lead a team, and I am a great team player."



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

## CORE SKILL SET

- ANALYTICS**
  - Developing and implementing analytics models
  - Using data to drive decisions
- COMMUNICATION**
  - Communicating complex data to stakeholders
  - Using data to drive decisions
- STRATEGIC THINKING**
  - Identifying and analyzing business opportunities
  - Using data to drive decisions

## GOALS

- Identify business opportunities and drive growth
- Identify and analyze business opportunities
- Identify and analyze business opportunities

## ECOSYSTEM MODEL



## POSSIBLE JOB TITLES

- Business Intelligence Analyst
- Data Analytics Analyst
- Business Development
- Business Analyst

# App Developer

"I will use my creative skills to create a new app that will help people in their daily lives. I am a great team player and I am a great team player."



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

## CORE SKILL SET

- ANALYTICS**
  - Developing and implementing analytics models
  - Using data to drive decisions
- COMMUNICATION**
  - Communicating complex data to stakeholders
  - Using data to drive decisions
- STRATEGIC THINKING**
  - Identifying and analyzing business opportunities
  - Using data to drive decisions

## GOALS

- Identify business opportunities and drive growth
- Identify and analyze business opportunities
- Identify and analyze business opportunities

## ECOSYSTEM MODEL



## POSSIBLE JOB TITLES

- Mobile App Developer
- Web App Developer
- Software Engineer
- Software Engineer

# Data Scientist

"I will use my analytical skills to analyze data and identify hidden insights. I am a great team player and I am a great team player."



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

## CORE SKILL SET

- ANALYTICS**
  - Developing and implementing analytics models
  - Using data to drive decisions
- COMMUNICATION**
  - Communicating complex data to stakeholders
  - Using data to drive decisions
- STRATEGIC THINKING**
  - Identifying and analyzing business opportunities
  - Using data to drive decisions

## GOALS

- Identify business opportunities and drive growth
- Identify and analyze business opportunities
- Identify and analyze business opportunities

## ECOSYSTEM MODEL



## POSSIBLE JOB TITLES

- Data Analyst
- Data Engineer
- Data Scientist
- Data Scientist

# Data Engineer

"I will use my technical skills to build a scalable data infrastructure that will help people in their daily lives. I am a great team player and I am a great team player."



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

## CORE SKILL SET

- ANALYTICS**
  - Developing and implementing analytics models
  - Using data to drive decisions
- COMMUNICATION**
  - Communicating complex data to stakeholders
  - Using data to drive decisions
- STRATEGIC THINKING**
  - Identifying and analyzing business opportunities
  - Using data to drive decisions

## GOALS

- Identify business opportunities and drive growth
- Identify and analyze business opportunities
- Identify and analyze business opportunities

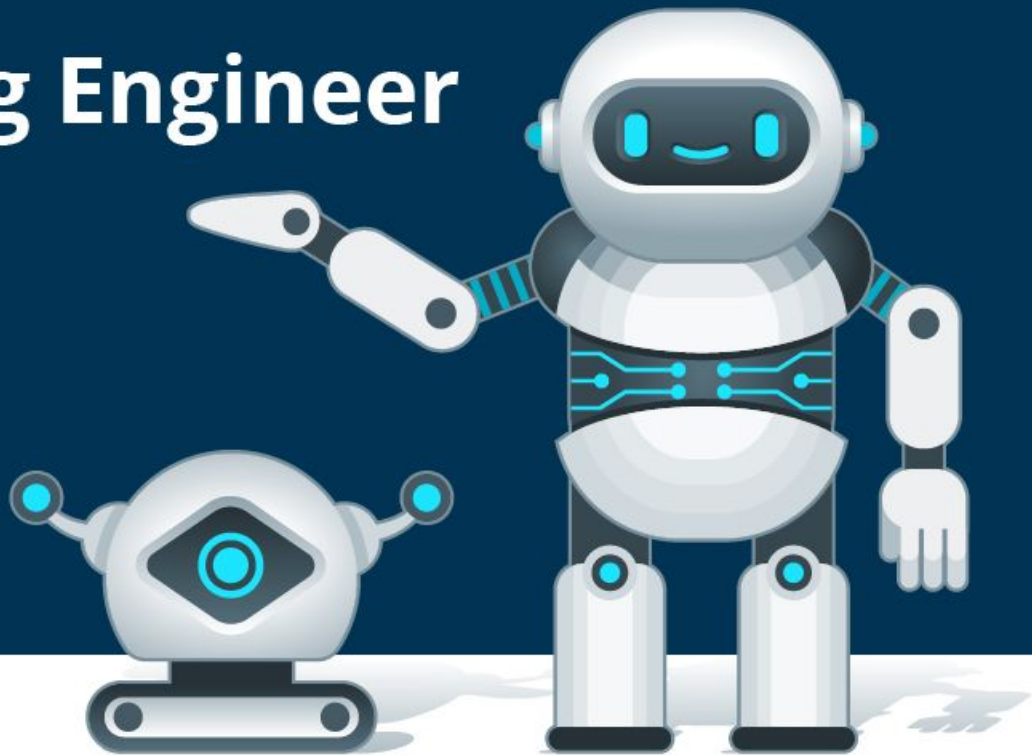
## ECOSYSTEM MODEL

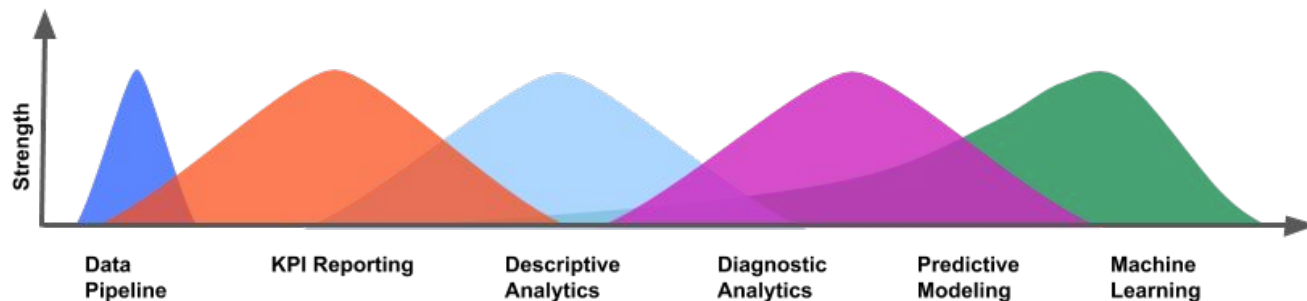


## POSSIBLE JOB TITLES

- Data Engineer
- Data Engineer
- Data Engineer
- Data Engineer

We are looking for  
**Machine Learning Engineer**





- **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

# Exploratory Data Analysis





# Machine Learning

Supervised  
Learning

Unsupervised  
Learning

# Machine Learning

Reinforcement  
Learning

Supervised  
Learning

With Labels

Unsupervised  
Learning

Without Labels

# Machine Learning



Reinforcement  
Learning

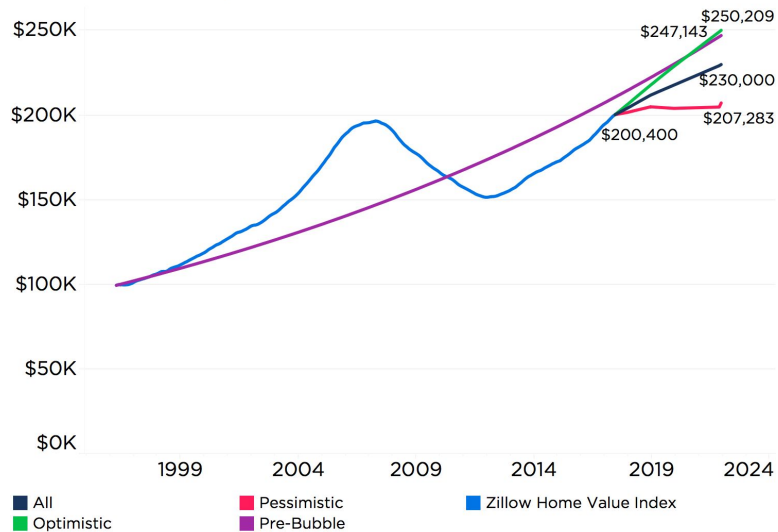


# Supervised Learning

With Labels

## Regression

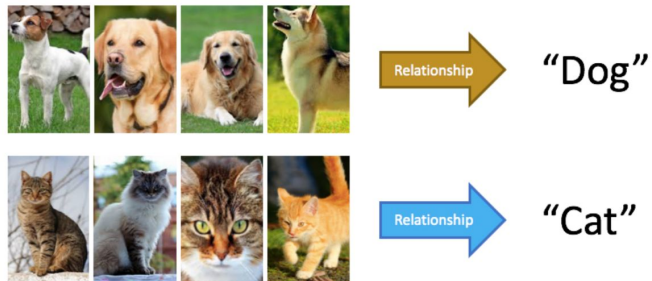
U.S. Home Prices Expectations for the Next Five Years



# Supervised Learning

With Labels

## Classification



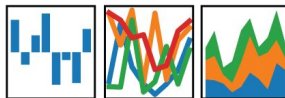
# Essential Tools



+

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



+







The Data

# **CelebFaces Attributes**

[kaggle.com/jessicali9530/celeba-dataset](https://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](https://kaggle.com/harlfoxem/housesalesprediction)



Demo

[tinyurl.com/BTG-S19-DS](https://tinyurl.com/BTG-S19-DS)