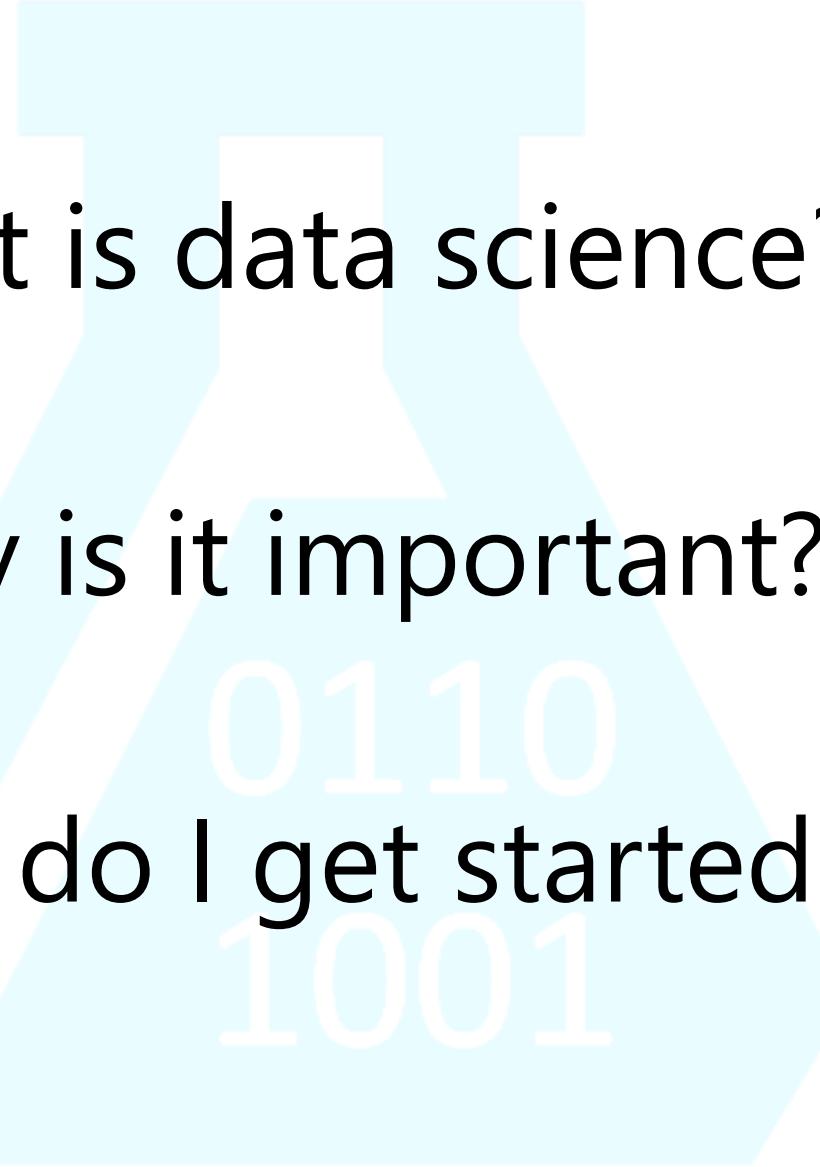


DevSci: Better Software Through Data Science

@MatthewRenze

#KCDC2018



What is data science?

Why is it important?

How do I get started?







The Economist

FEBRUARY 27TH-MARCH 5TH 2010

Gordon Brown's pitch
What went wrong at RBS
Genetically modified crops blossom
The EU woos Russia
The right to eat cats and dogs

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

The New York Times

For Today's Graduate, Just One Word: Statistics

By STEVE LOMK
Published: August 5, 2009

MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls “all the computer and math stuff” that was part of the job.

[TWITTER](#)
[LINKEDIN](#)
[COMMENTS
\(58\)](#)
[SIGN IN TO E-MAIL](#)

Data Scientist: *The Sexiest Job of the 21st Century*

**Meet the people who
can coax treasure out of
messy, unstructured data.**

by Thomas H. Davenport
and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, “It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early.”

Job Postings for Data Scientists

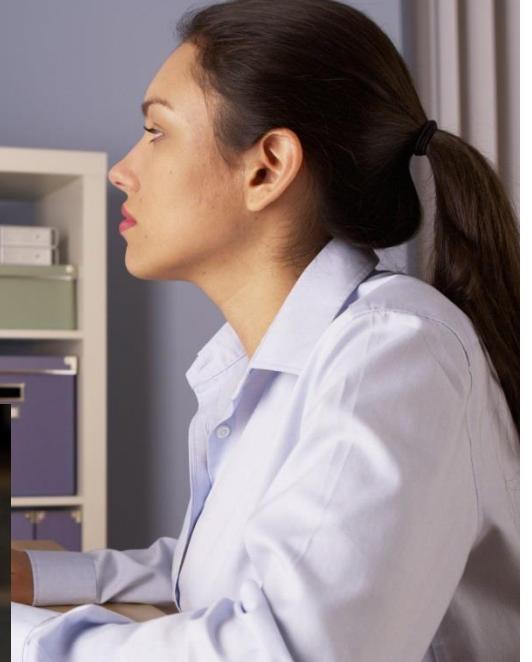


Top-paying Tech Skills

Skill	2016	Change
HANA (High Performance Analytical Application)	\$ 128,958	-3.3%
MapReduce	\$ 125,009	-0.3%
Cloud Foundry	\$ 124,038	n/a
Hbase	\$ 123,934	5.7%
Omnigraffle	\$ 123,782	-1.9%
Cassandra	\$ 123,459	2.2%
Apache Kafka	\$ 122,728	n/a
SOA (Service Oriented Architecture)	\$ 122,094	-1.9%
Ansible	\$ 121,382	n/a
Jetty	\$ 120,978	1.3%
PaaS (Platform as a Service)	\$ 120,403	-4.4%
Elasticsearch	\$ 120,002	n/a
ABAP (Advanced Business Application Programming)	\$ 119,961	0.5%
NoSQL	\$ 119,498	1.3%
CMMI (Capability Maturity Model Integration)	\$ 119,466	-0.6%
Amazon Redshift	\$ 119,197	n/a
Pig	\$ 119,118	-4.2%
Solr	\$ 119,032	0.1%
Cloudera	\$ 118,896	-9.0%
Docker	\$ 118,873	0.2%

Skill	2016	Change
Amazon Route 53	\$ 118,828	n/a
Hadoop	\$ 118,625	-2.5%
Hive	\$ 118,589	-1.3%
Korn Shell	\$ 118,273	1.4%
PMBok (Project Management Body of Knowledge)	\$ 118,233	0.7%
Dynamo DB	\$ 118,119	n/a
Groovy	\$ 117,897	-0.1%
IaaS (Infrastructure as a Service)	\$ 117,422	n/a
JAX-RS (Java API RestFUL Services)	\$ 116,997	n/a
RabbitMQ	\$ 116,909	n/a
JDBC (Java Database Connectivity)	\$ 116,833	2.0%
SOX (Sarbanes Oxley)	\$ 116,743	0.6%
Objective C	\$ 116,667	2.5%
FCoE (Fibre Channel over Ethernet)	\$ 116,145	7.2%
UML (Unified Modeling Language)	\$ 115,285	-3.6%
XSLT (Extensible Stylesheet Language Transformations)	\$ 115,089	3.5%
Redis	\$ 114,922	2.8%
ETL (Extract Transform and Load)	\$ 114,892	2.6%
SDN (Software Defined Network)	\$ 114,739	-2.3%
Informatica	\$ 114,143	1.1%

Source: Dice Salary Survey 2017







About Me

Data Science Consultant

IOWA STATE
UNIVERSITY



About Me

Data Science Consultant

Education

B.S. in Computer Science (ISU)

B.A. in Philosophy (ISU)

IOWA STATE
UNIVERSITY



About Me

Data Science Consultant

Education

B.S. in Computer Science (ISU)

B.A. in Philosophy (ISU)

Community

Keynote speaker

Pluralsight author

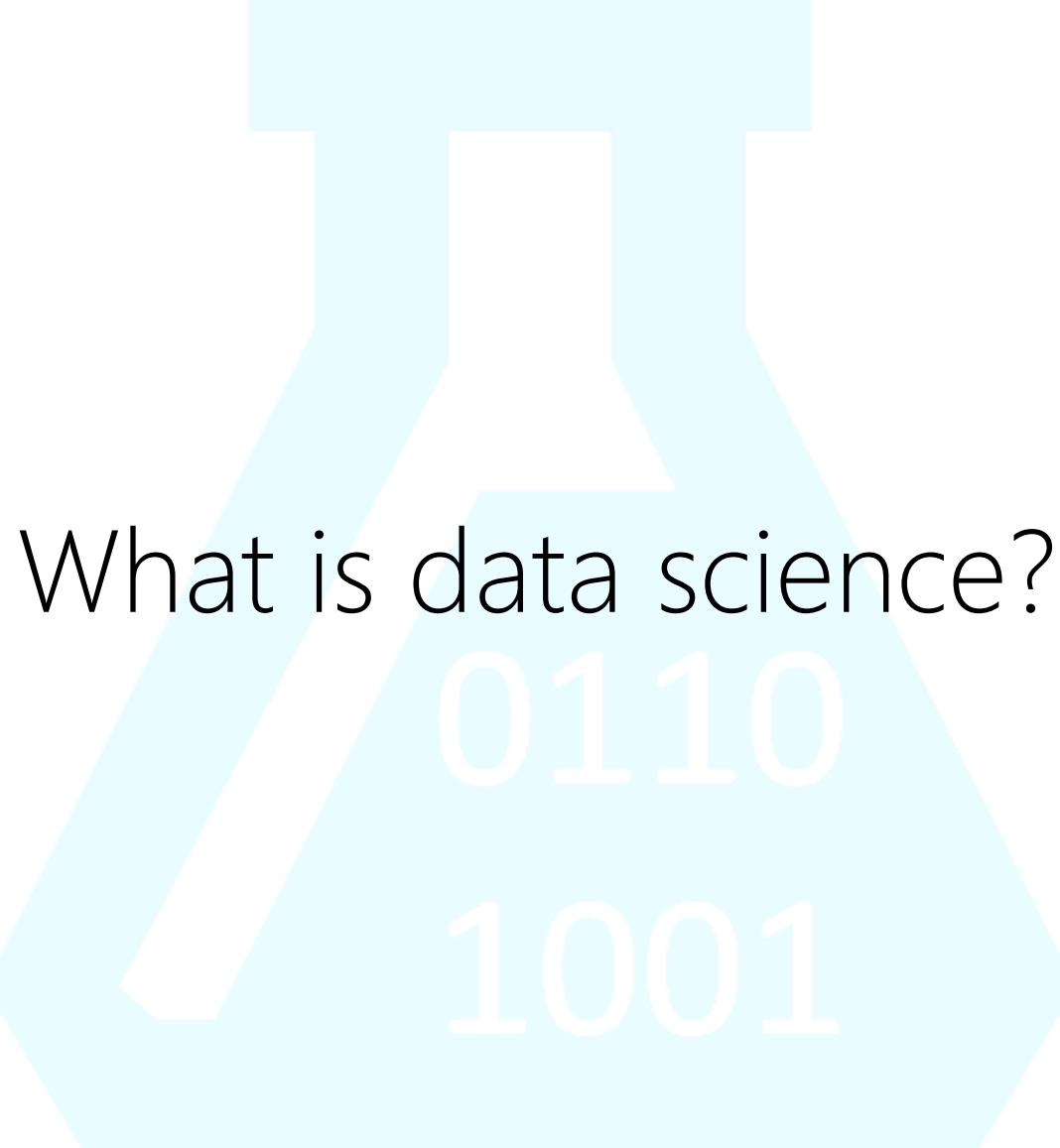
DataCamp author

Microsoft MVP AI

ASPIInsider

IOWA STATE
UNIVERSITY

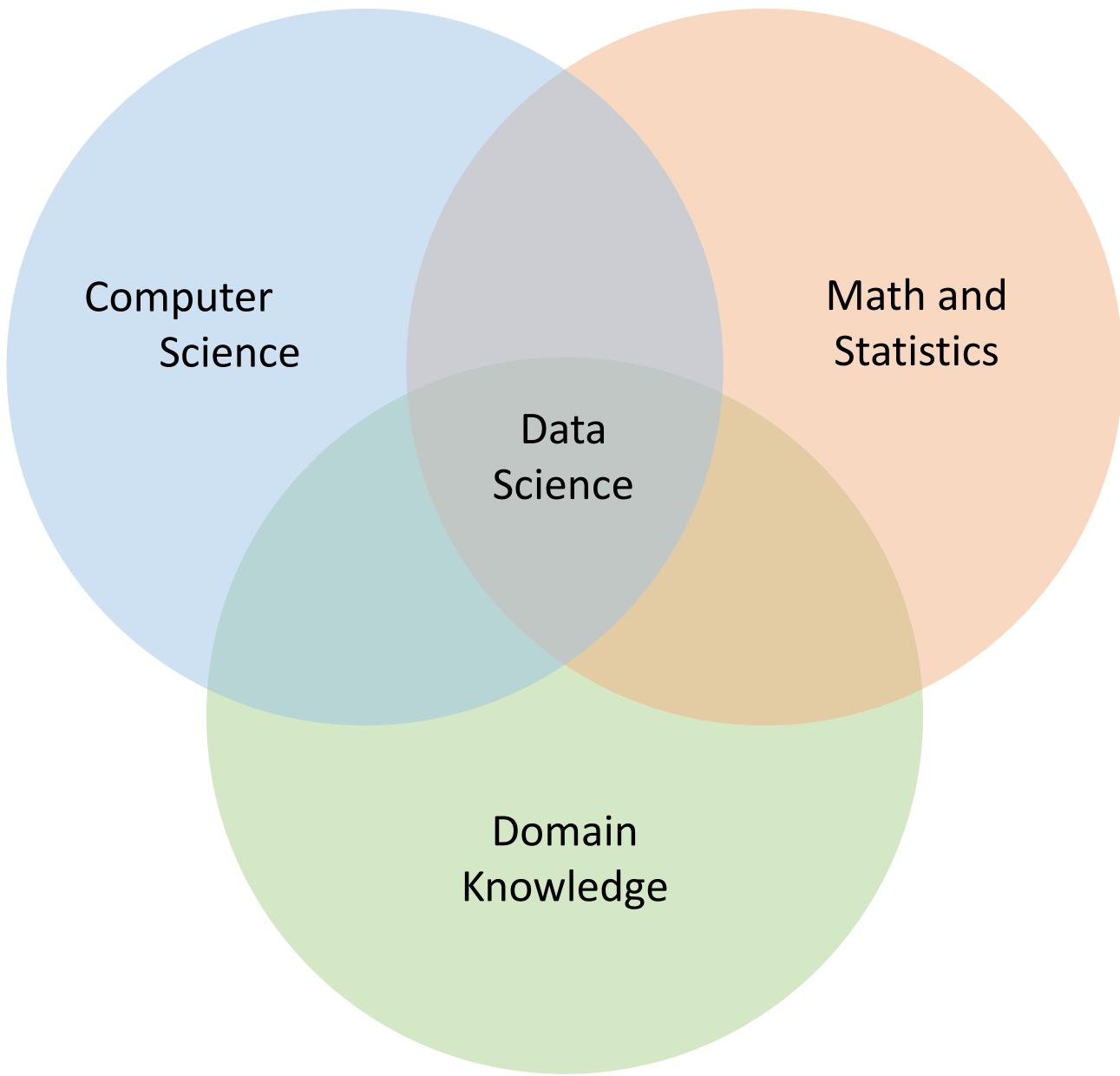


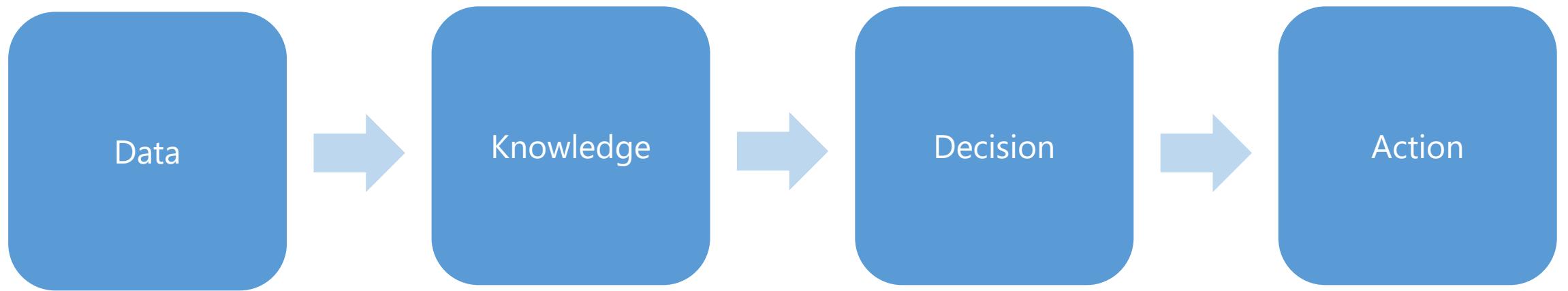


What is data science?

0110

1001





What Is a Data Scientist?

Performs data science

More than a scientist

More than an analyst

More than a developer



What skills are necessary?

Data Science Skills

Programming

Working with data

Descriptive statistics

Data visualization

Data Science Skills

Programming

Working with data

Descriptive statistics

Data visualization

Statistical modeling

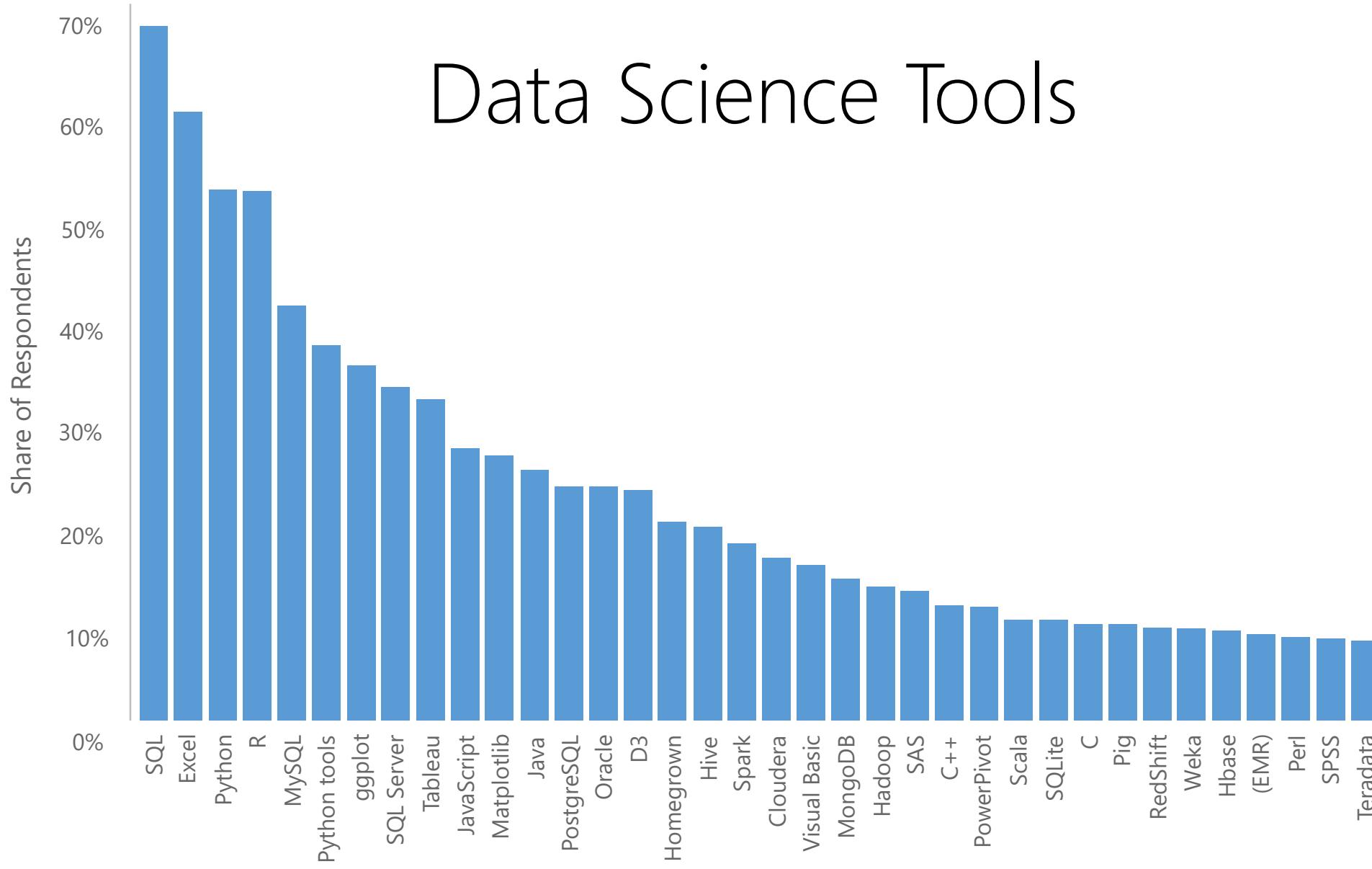
Handling Big Data

Machine learning

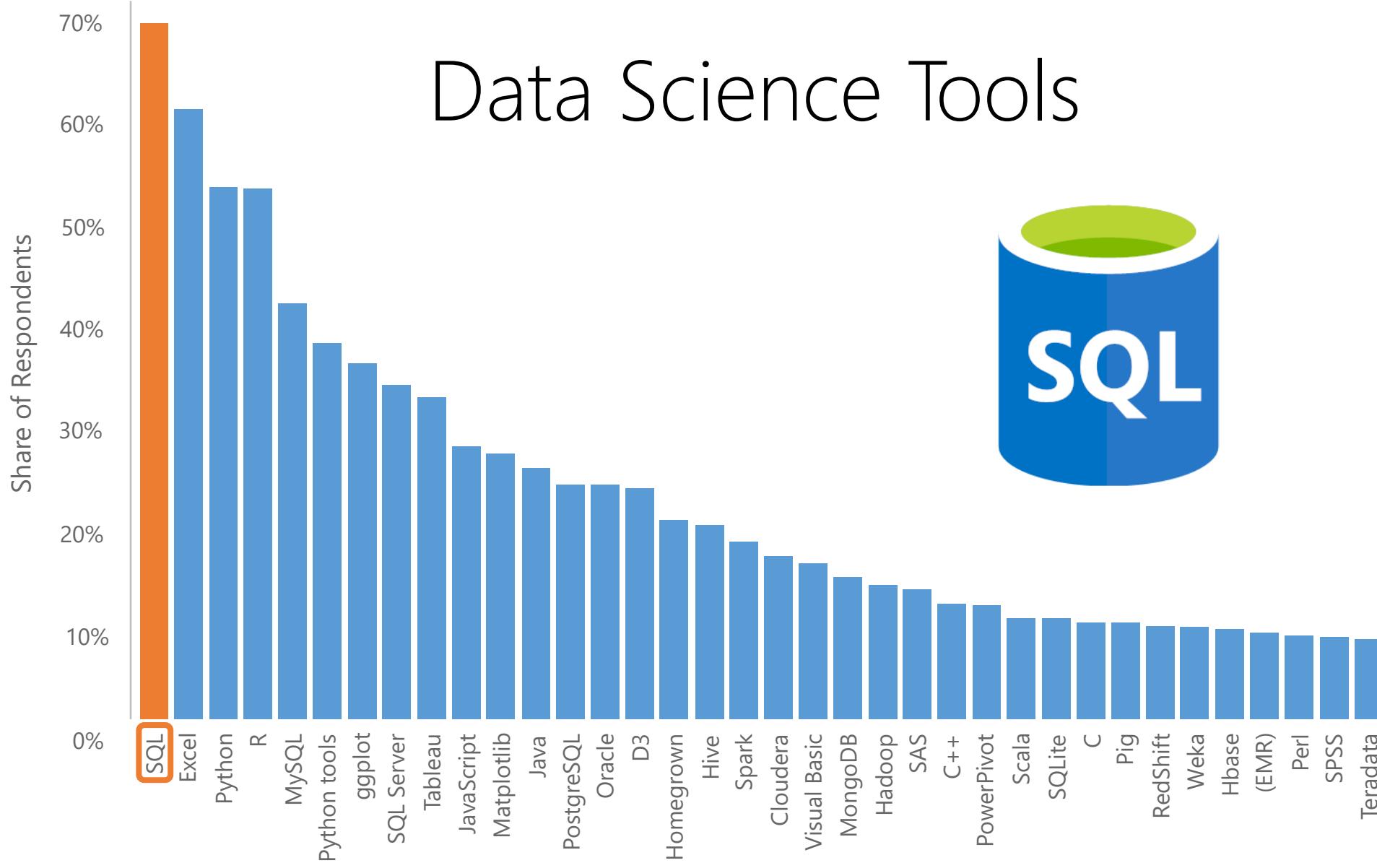
Deploying to production



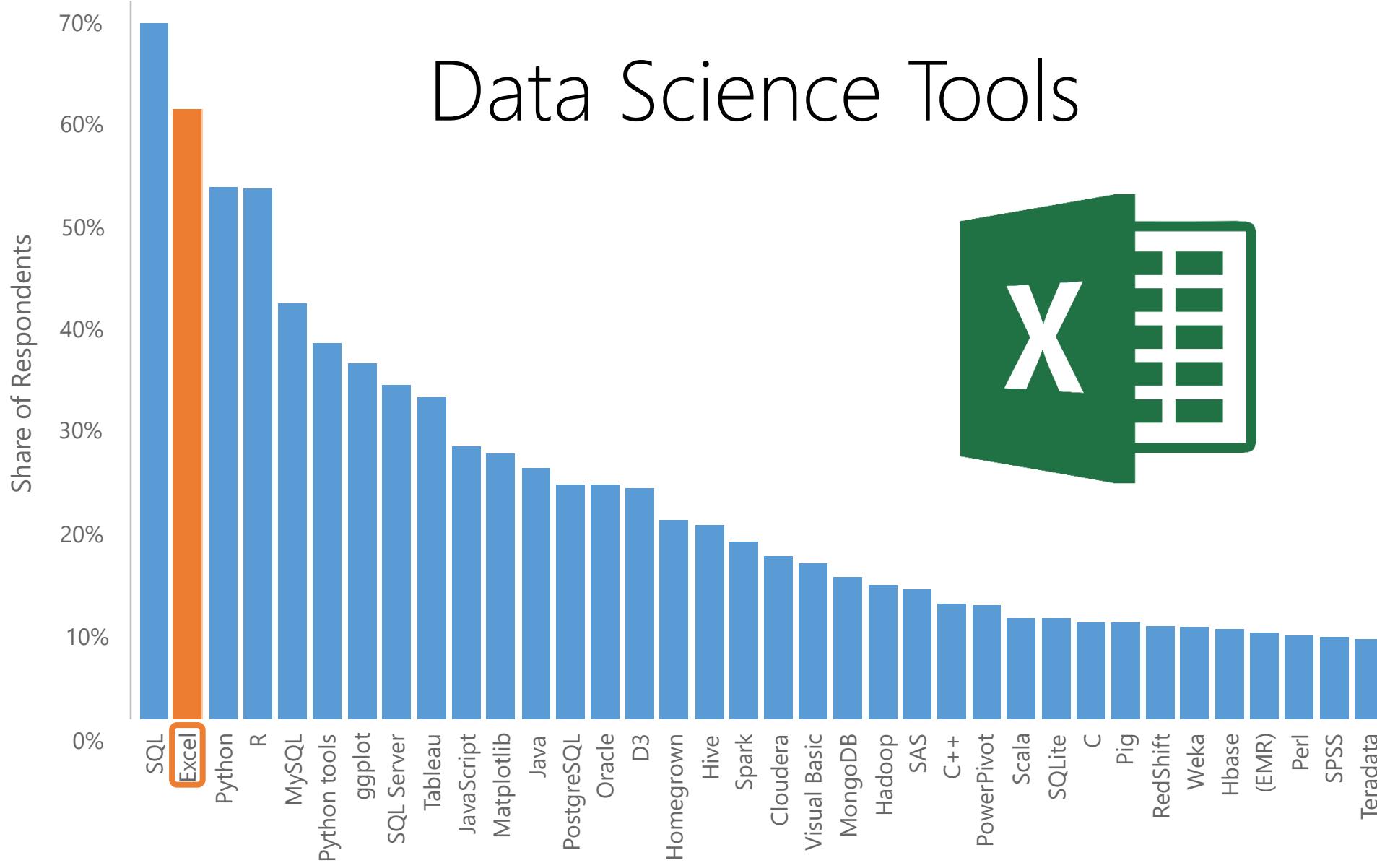
What tools are used?



Source: O'Reilly 2015 Data Science Salary Survey

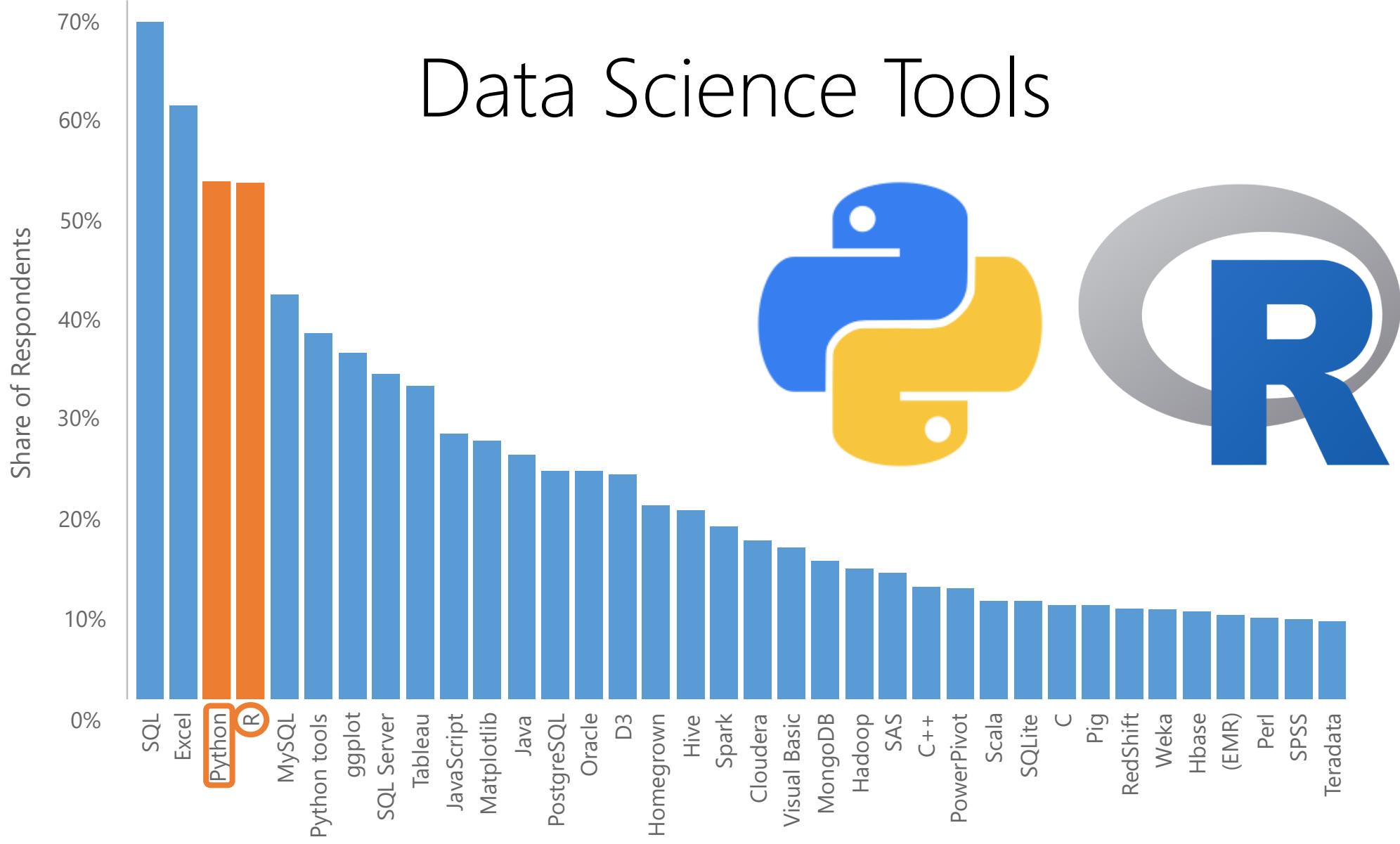


Source: O'Reilly 2015 Data Science Salary Survey

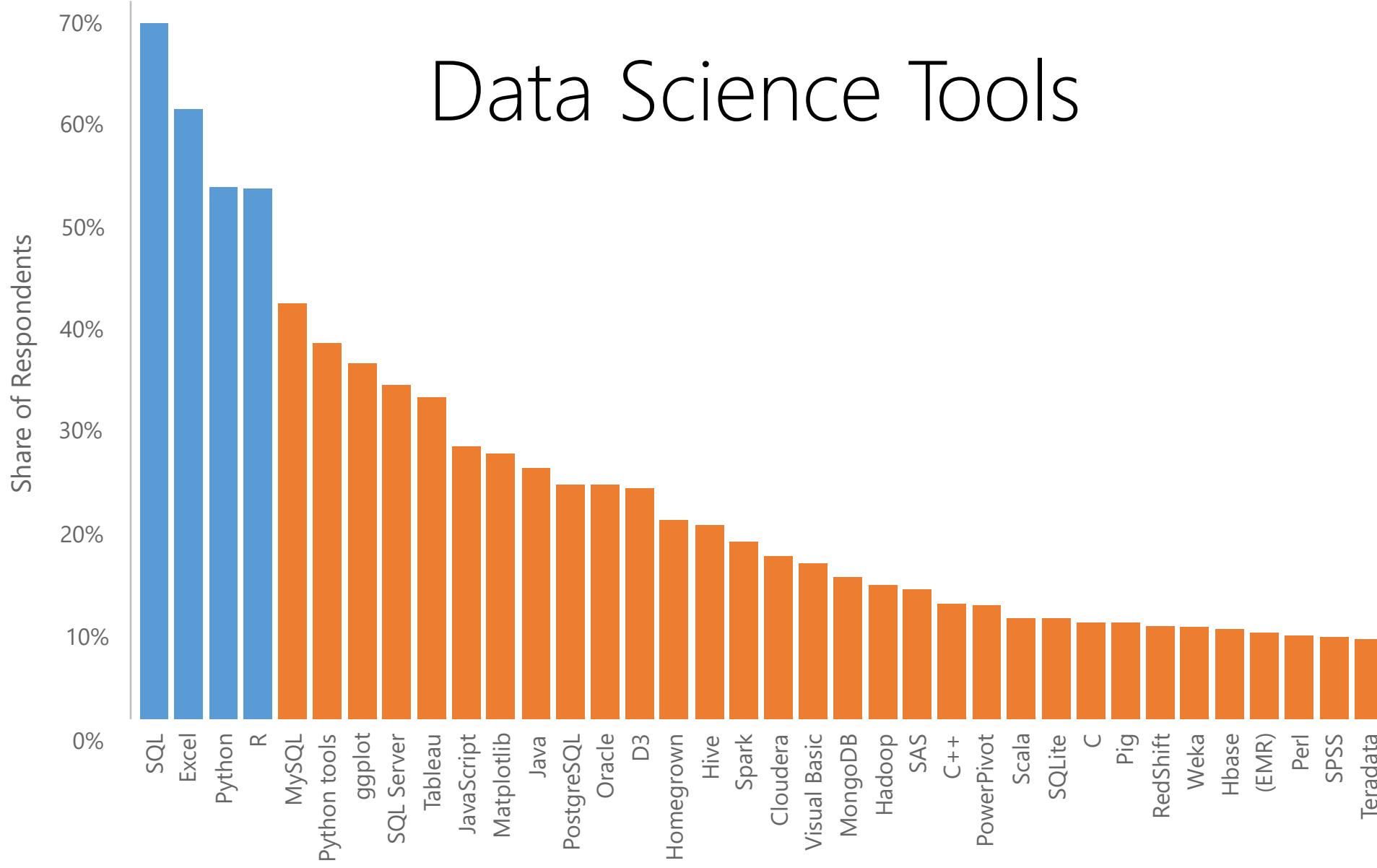


Tool: language, platform, analytics

Source: O'Reilly 2015 Data Science Salary Survey



Source: O'Reilly 2015 Data Science Salary Survey



Source: O'Reilly 2015 Data Science Salary Survey



How is data science performed?

The Data Science Process



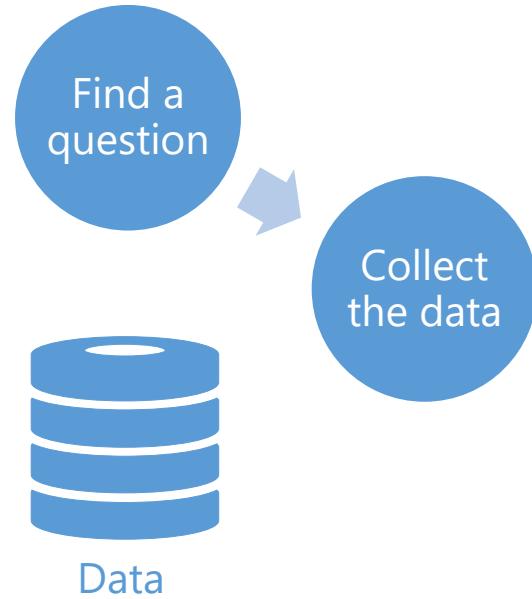
Data

The Data Science Process

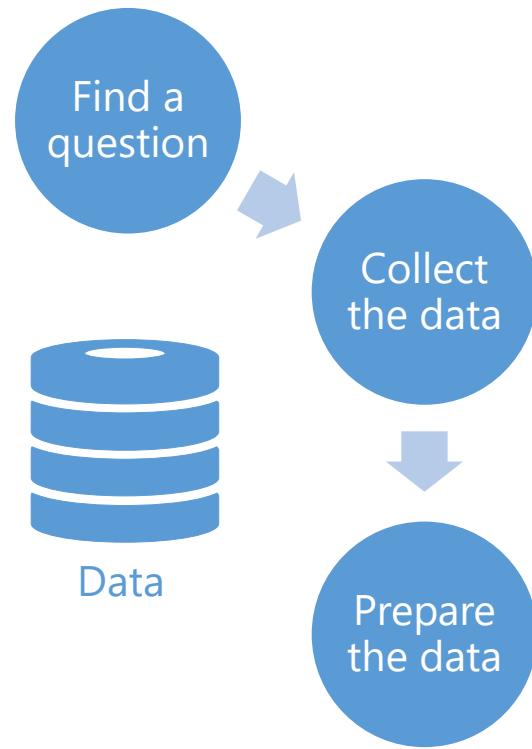


Data

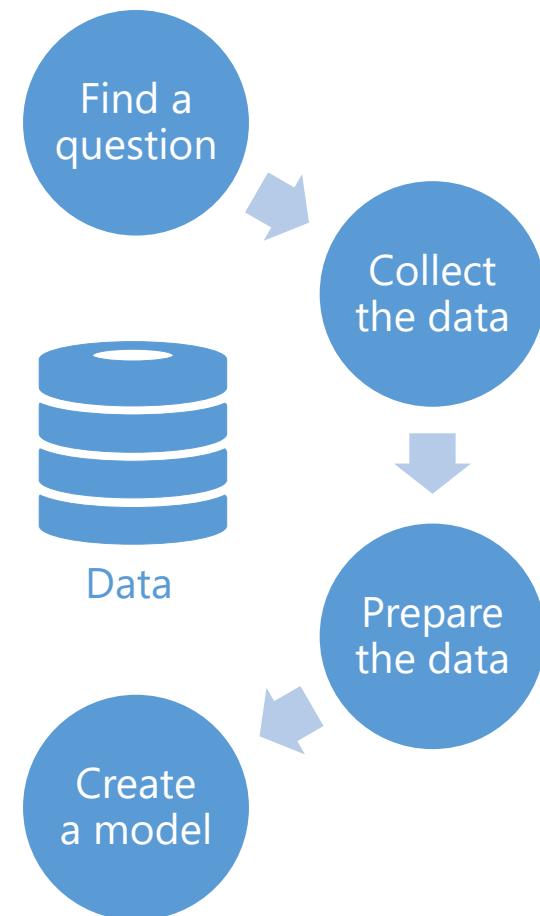
The Data Science Process



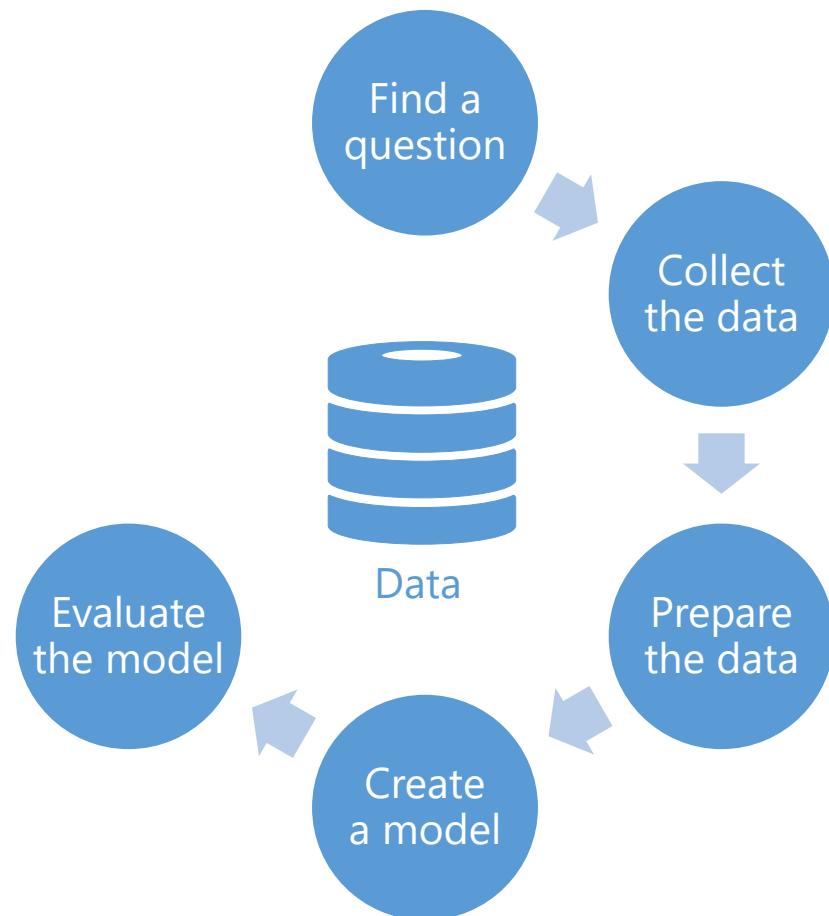
The Data Science Process



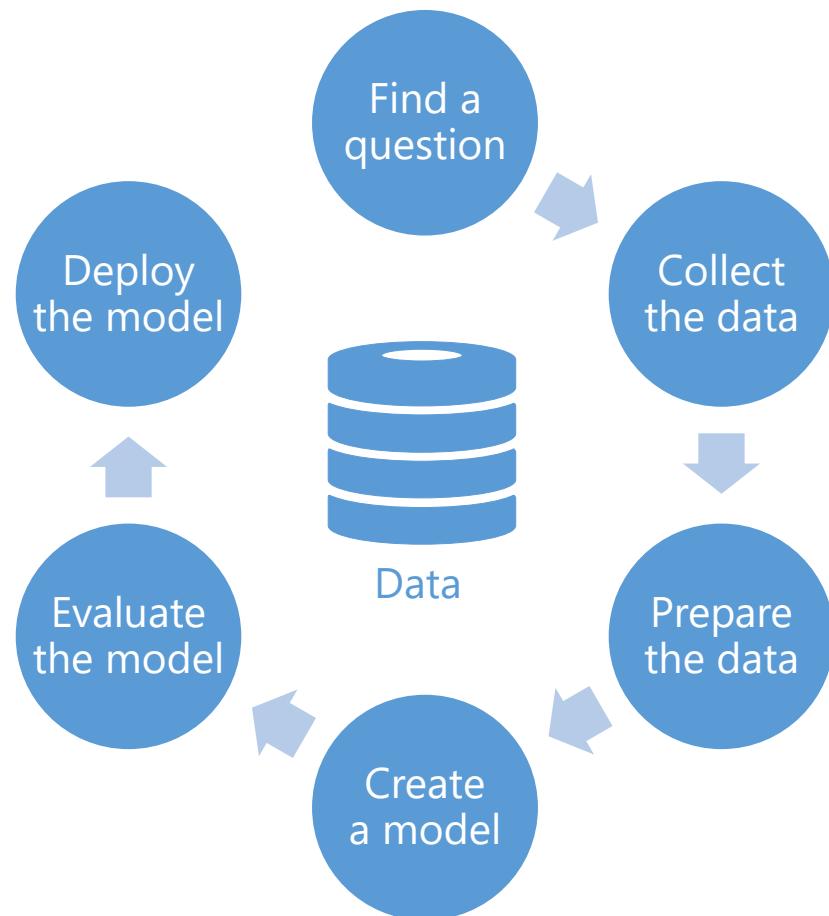
The Data Science Process



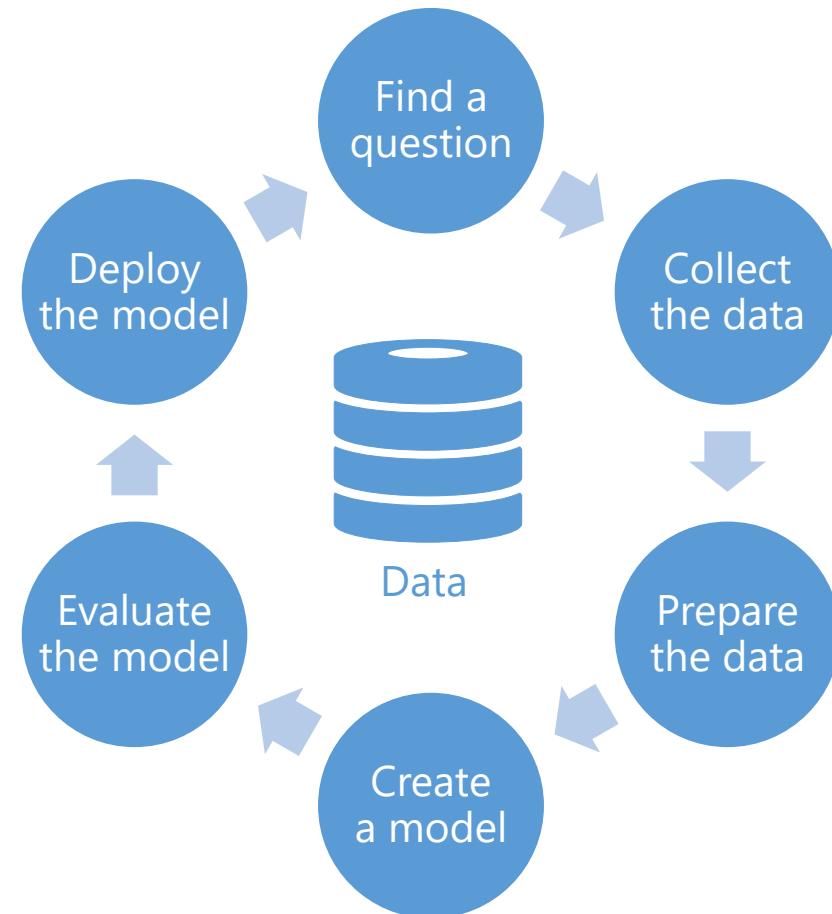
The Data Science Process



The Data Science Process

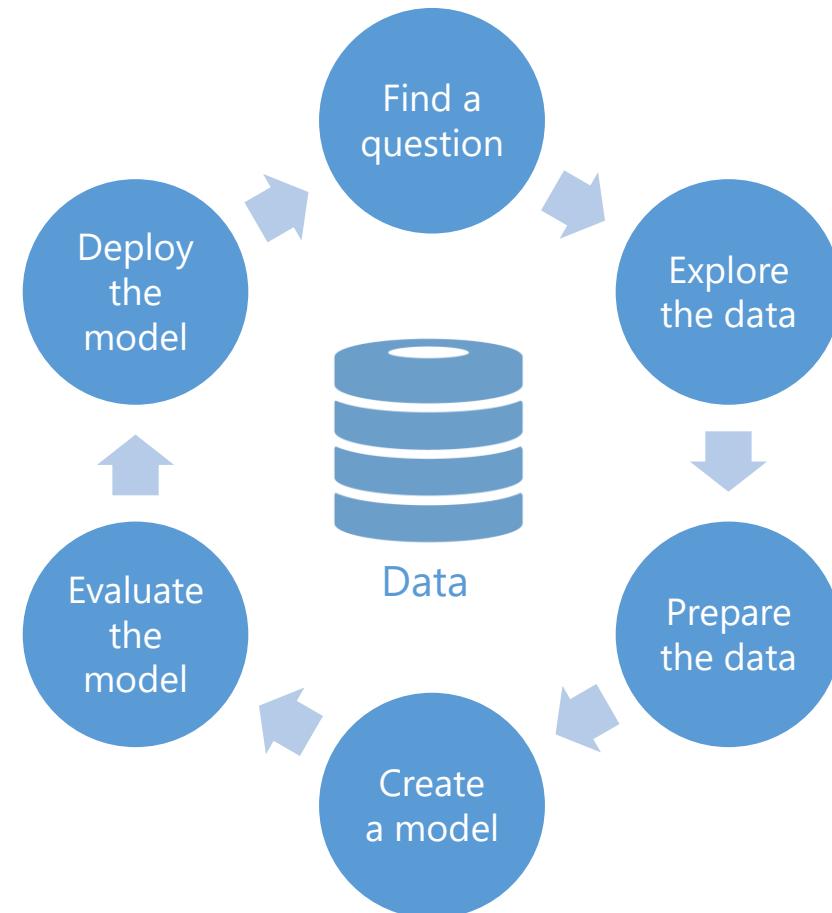


The Data Science Process



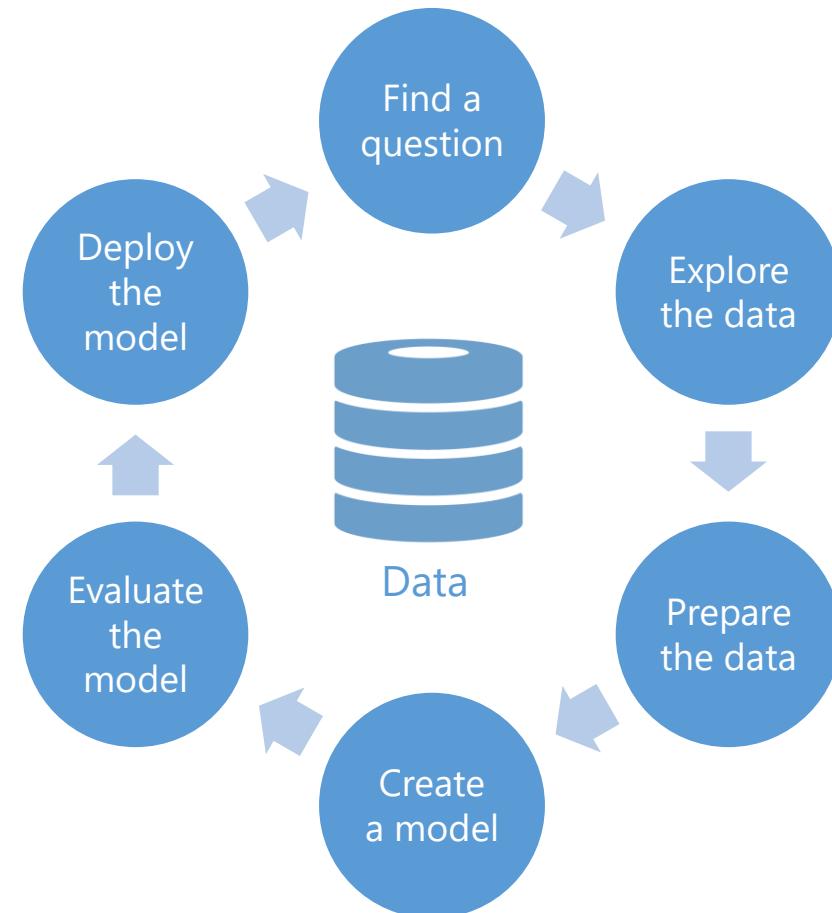
The Data Science Process

Iterative process



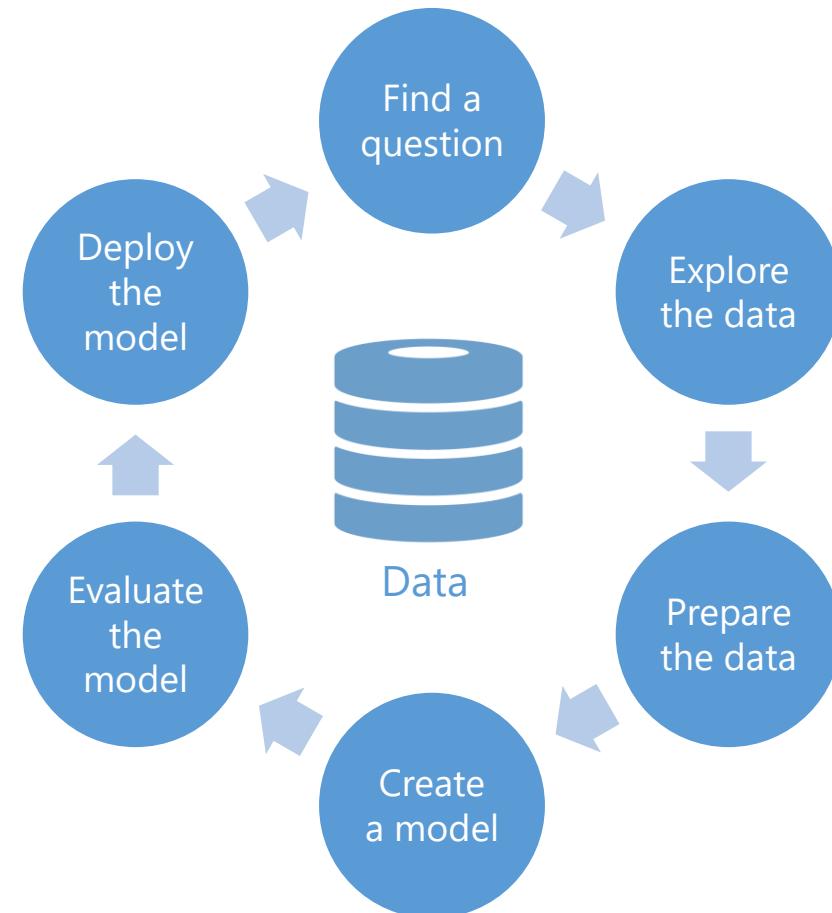
The Data Science Process

Iterative process
Non-sequential



The Data Science Process

Iterative process
Non-sequential
Early termination





Why is data science important?

0110

1001

Two Main Approaches

Build
intelligent
software

Improve
development
practices

Two Main Approaches



Build
intelligent
software



Improve
development
practices



Ask a question about the data on this dashboard

Total Stores

104

This Year's Sales

\$22.05M

This Year's Sales

BY CHAIN

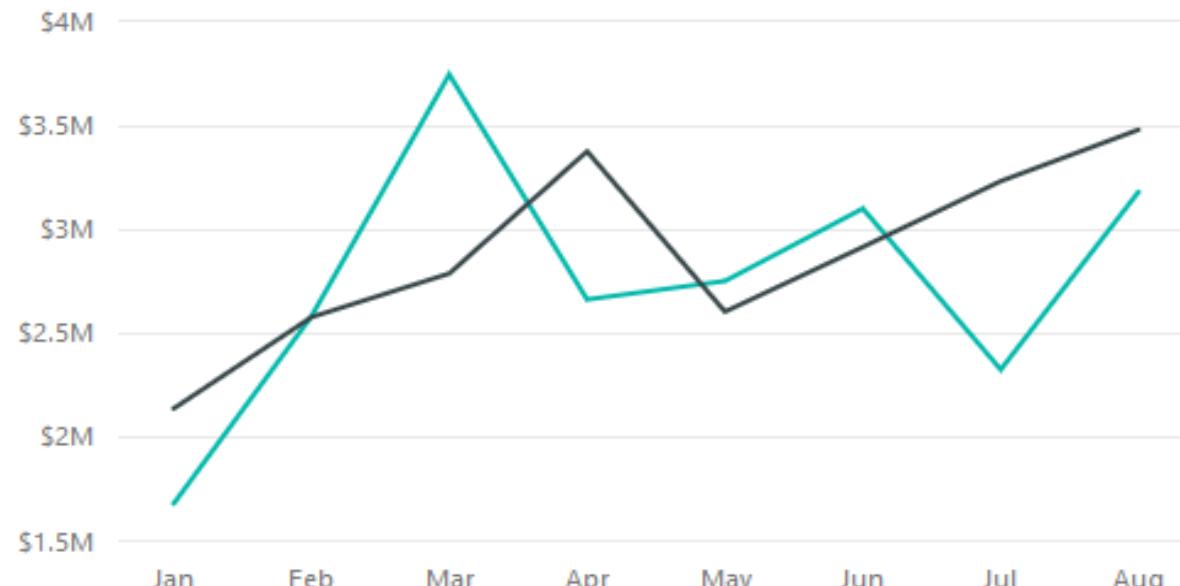


New Stores Opened This Year

10

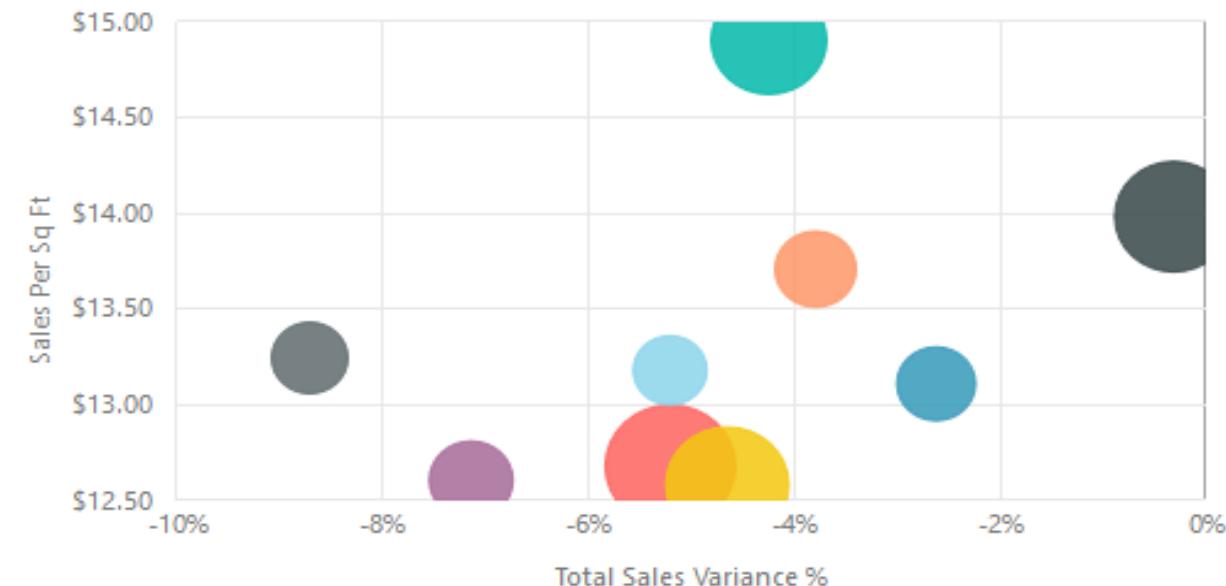
This Year's Sales, Last Year's Sales

BY FISCAL MONTH

● This Year Sales ● Last Year Sales

Sales Per Sq Ft, Total Sales Variance %, This Year's Sales

BY DISTRICT

District ● FD - 01 ● FD - 02 ● FD - 03 ● FD - 04 ● LI - 01 ● LI - 02 ● LI - 03



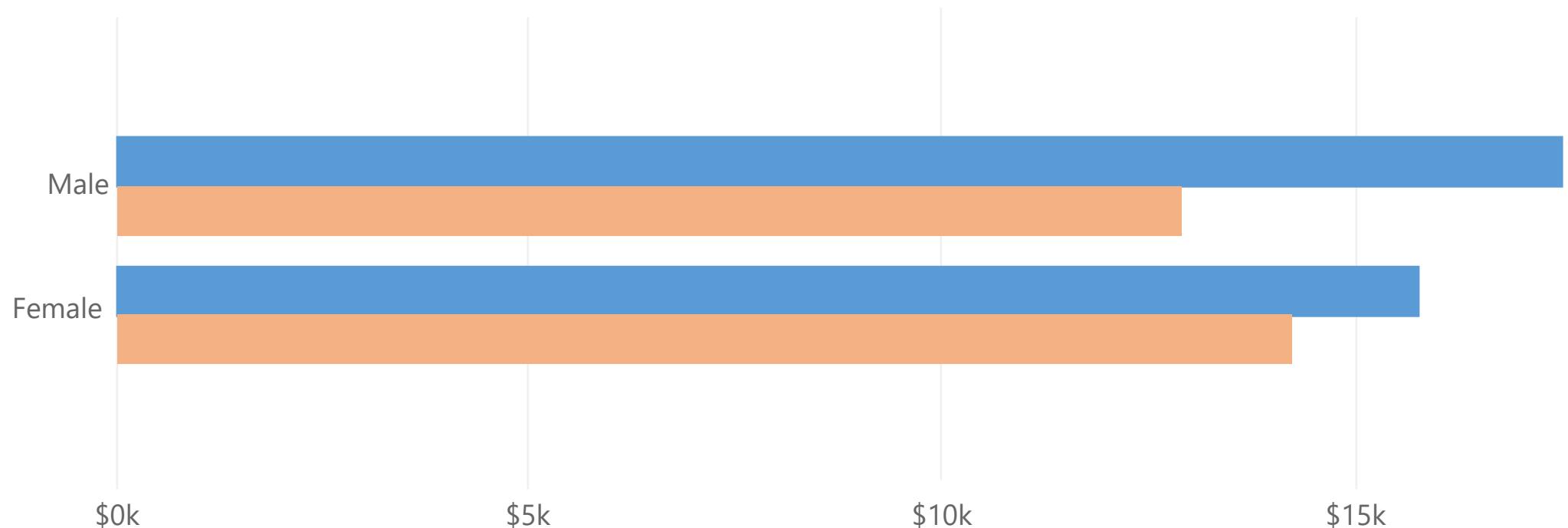
Internet Sales

Show me sales by gender and marital status.

Displaying sum of sales by gender and marital status

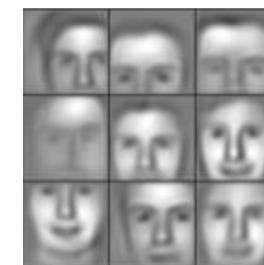
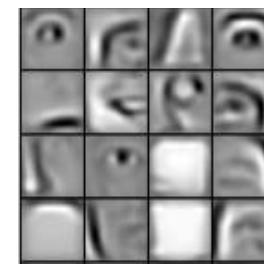
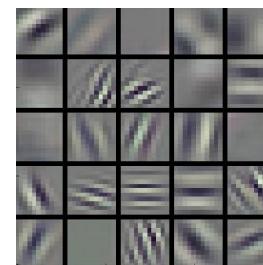
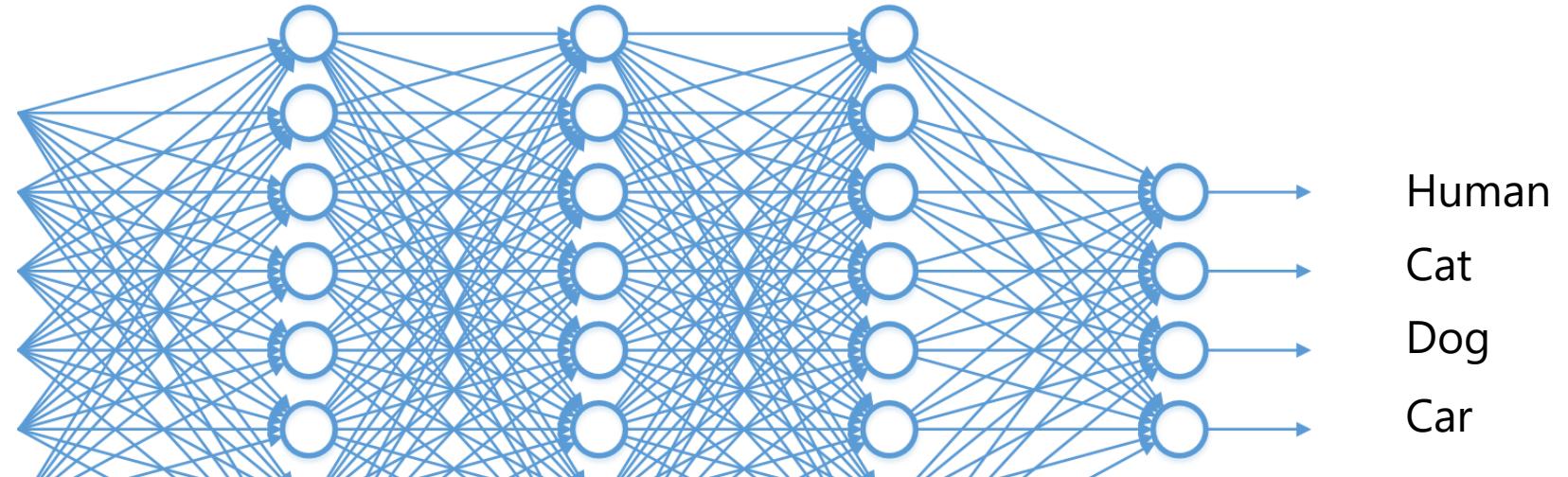
Marital Status:

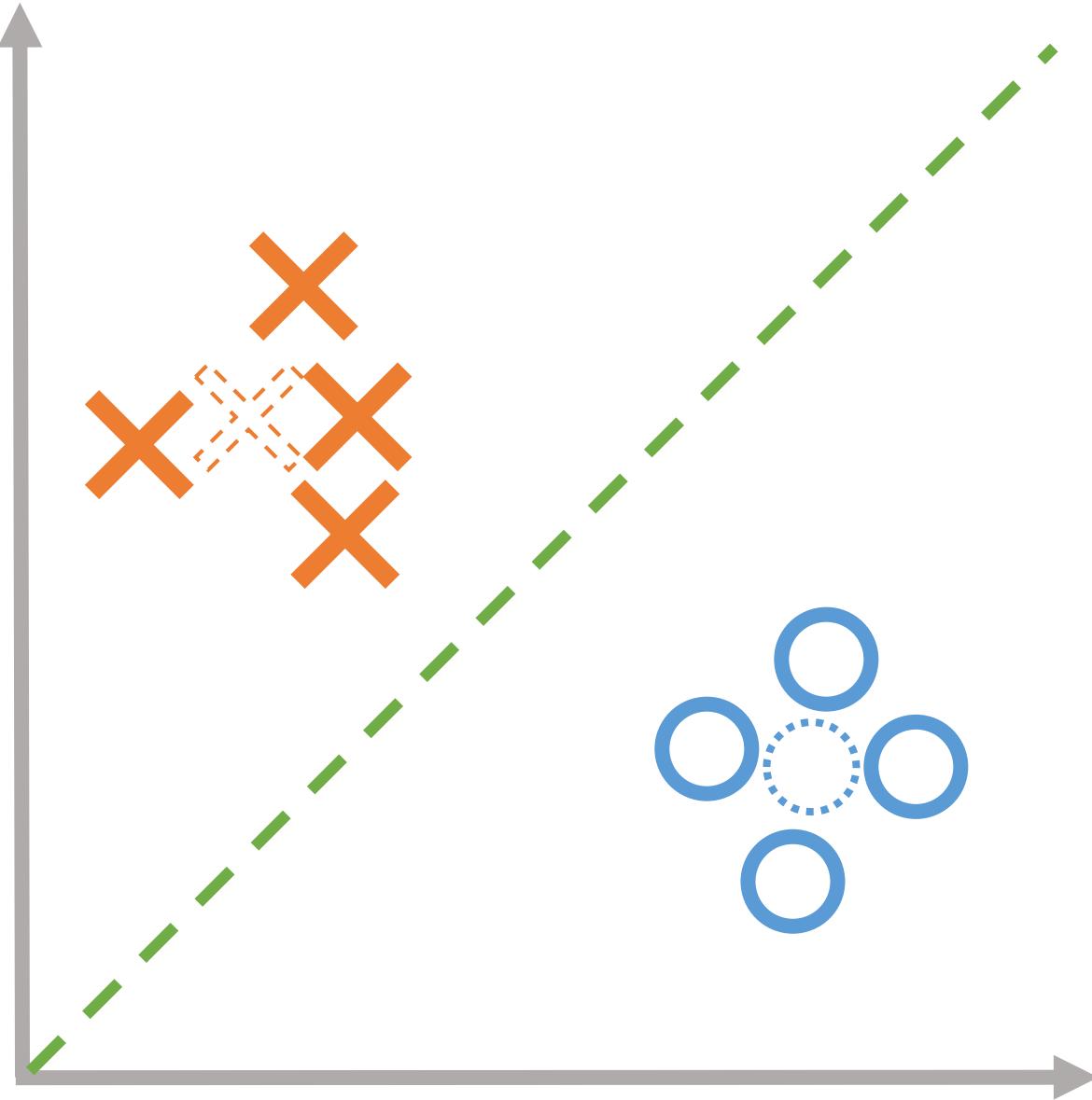
- Married
- Single

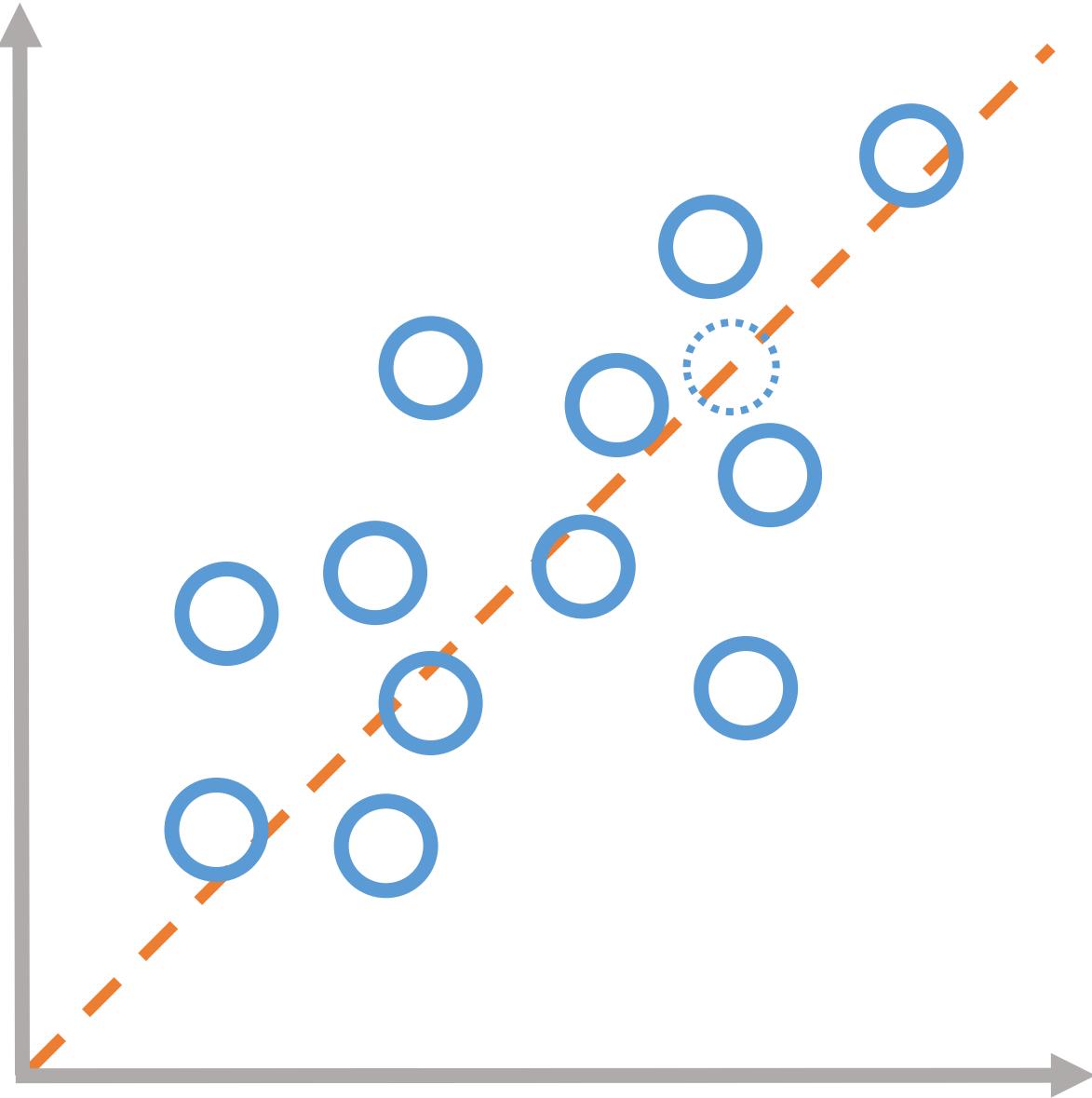


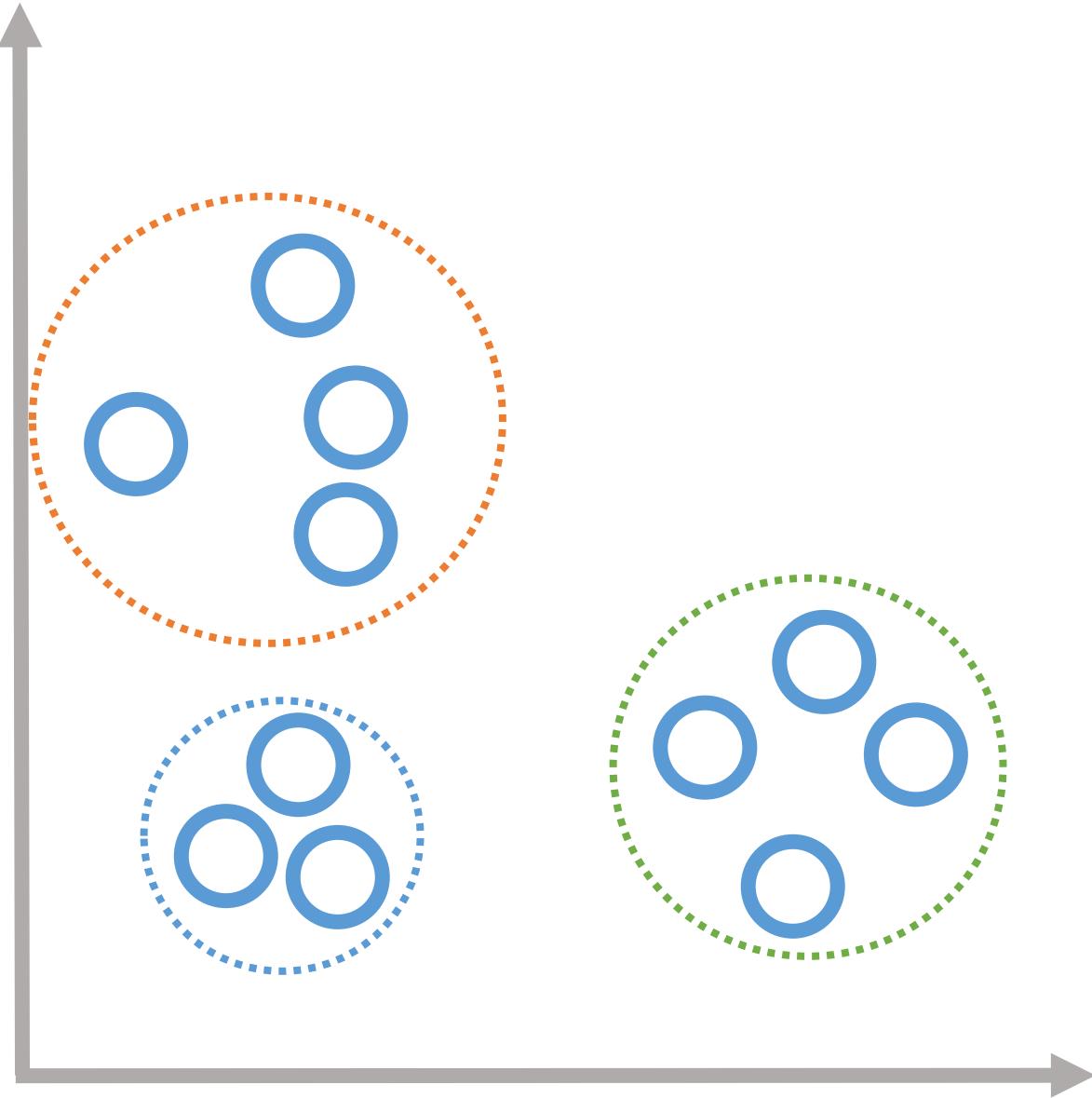
"Show me sales by
gender and marital
status."

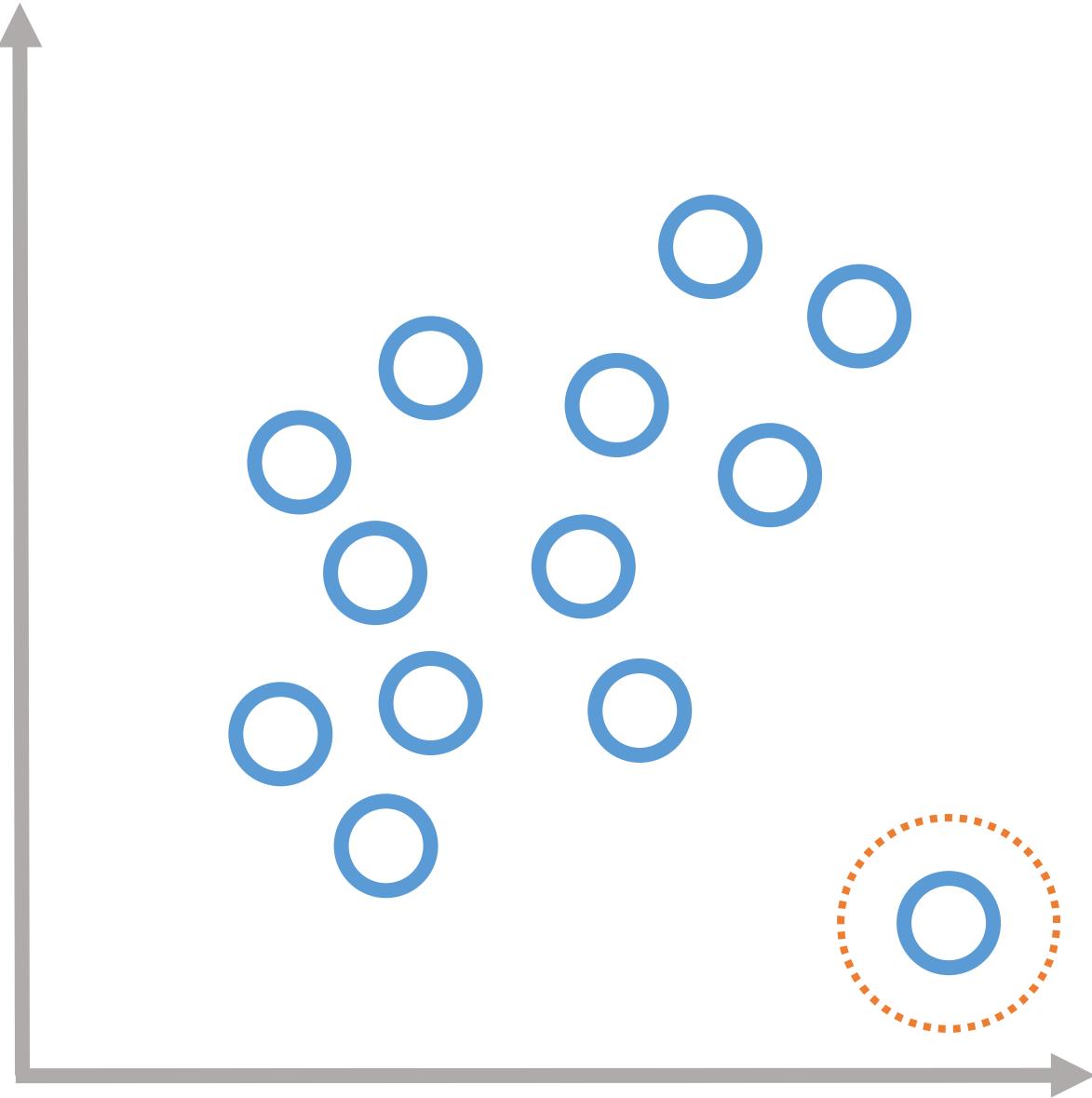
Machine Learning



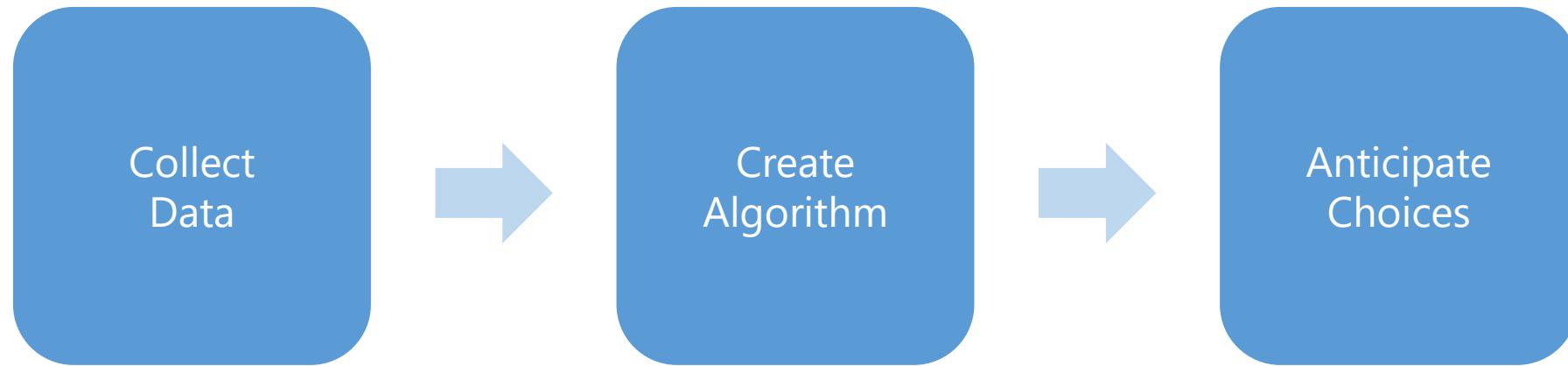








Anticipatory Design



...



Search



renze1577



Browse

Radio

YOUR LIBRARY

Your Daily Mix

Recently Played

Songs

Albums

Artists

Stations

Local Files

Videos

Podcasts

PLAYLISTS

Discover Weekly

Closer

One Man Puppet Show

A Puppet Named Julio

This Will Destroy You

Explosions in the Sky

Mastodon - Call Me

New Playlist

Your Daily Mixes

Play the music you love, without the effort. Packed with your favorites and new discoveries.



Daily Mix 1

Bonobo, The Album Leaf, Thievery Corporation and more

MADE FOR RENZE1577



Daily Mix 2

Epic45, Sigur Rós, mún and more

MADE FOR RENZE1577



Daily Mix 3

Joey Fehrenbach, The Album Leaf, The Abbasi Brothers and more

MADE FOR RENZE1577

See what your friends
are playing

FIND FRIENDS

Julia Eger Heartbeats José González The Indie Mix

Ben Khan Night Changes One Direction One Direction

Sean Aquilina 12m Flashing Lights Kanye West Graduation

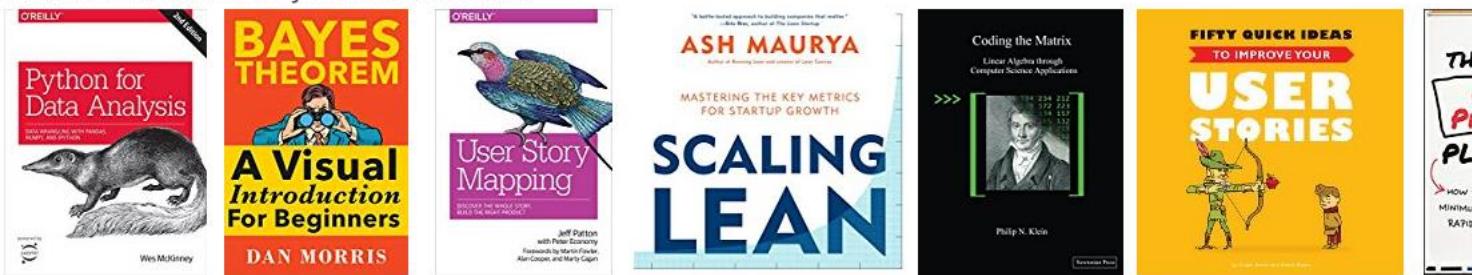
Maddie Stocker 25m Fake Plastic Trees Radiohead Radiohead Essential...

Laura Stephenson

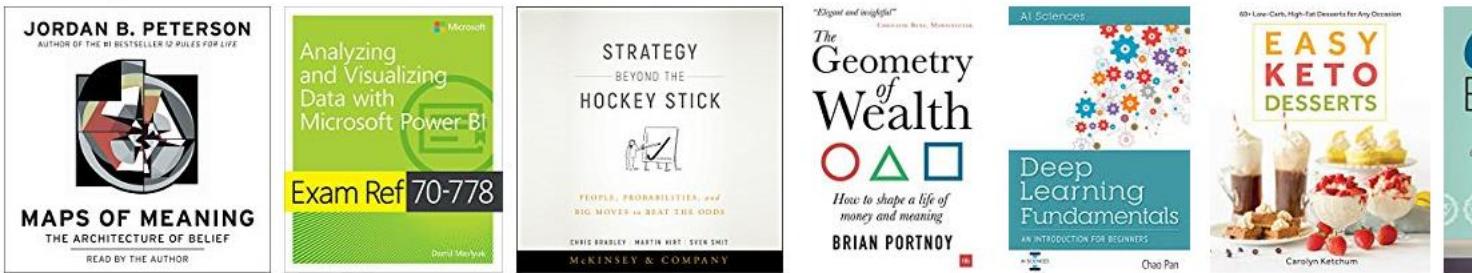
1h



Recommendations for you in Kindle Store



New for you [See more](#)



New & exclusive products
[prime day launches](#)



Shop innovative tech
[exclusives](#)



Buy it again



Amazon
Gift Cards



Millions of items,
no expiration.

> Shop now

Advertisement

Tap Below to Reorder

Your Dash Buttons

[Learn more](#)

TAP HERE for info

TAP HERE to buy



[See all your Dash Buttons](#)

Prime Members
read **free**

[prime reading](#)





SEARCH

MENU

- Suspendisse
- Velit Sit Amet
- Pellentesque
- Voluptat
- Pretium Nulla
- Eget Turpis

PROJECTS

- Nunc venenatis
- Maecenas
- Curabitur

Dashboard > Ornare Orci Velit

Consectetur Adipiscing

Donec sollicitudin, ipsum non

\$1,310

Nunc et metus magna



Aenean Porta

Donec sollicitudin, ipsum non

\$24,815

Nunc et metus magna



Aenean Porta

Donec sollicitudin, ipsum non

94.2%

Nunc et metus magna



Aenean Porta

Donec sollicitudin, ipsum non

Nunc et metus magna

Integer

Nam ultrices

Phasellus vulputate

Vivamus

< NULLAM SAPIEN >

Mon	Tue	Wed	Thu	Fri	Sat	Sun
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2

Nullam Ac Leo Vitae



Lorem Ipsum

Nunc faucibus dignissim nibh

Cras Pharetra



Aliquam

Nunc faucibus dignissim nibh

Donec orci nulla

872

Lobortis non nisi quis

650

Phasellus Interdum

Dignissim nibh

Bibendum

Nunc Leo Tortor

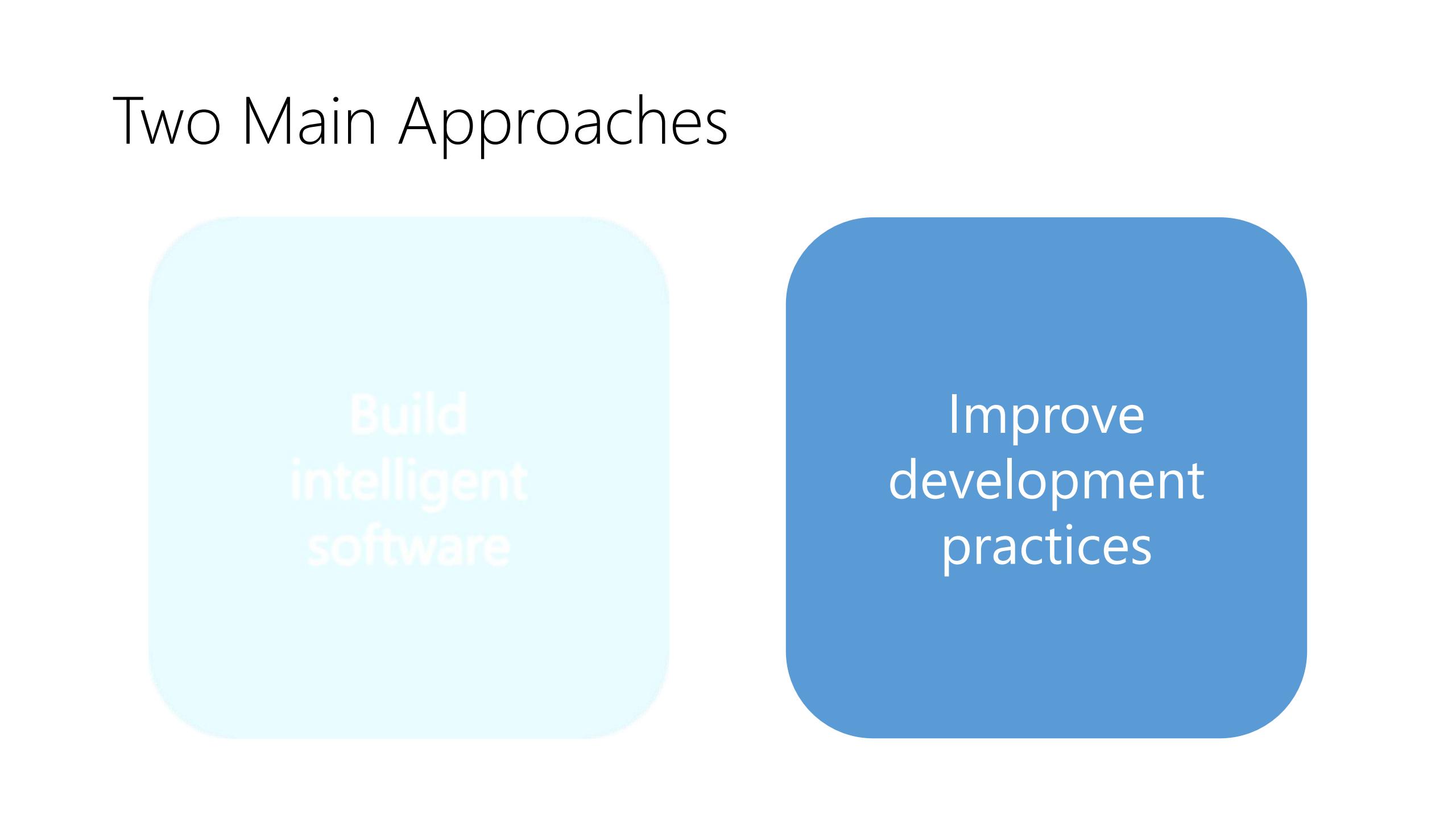
Nunc faucibus dignissim nibh

Felis ligula rutrum orci 2016

Phasellus interdum ligula eget turpis

Sed id egestas 2016

Two Main Approaches

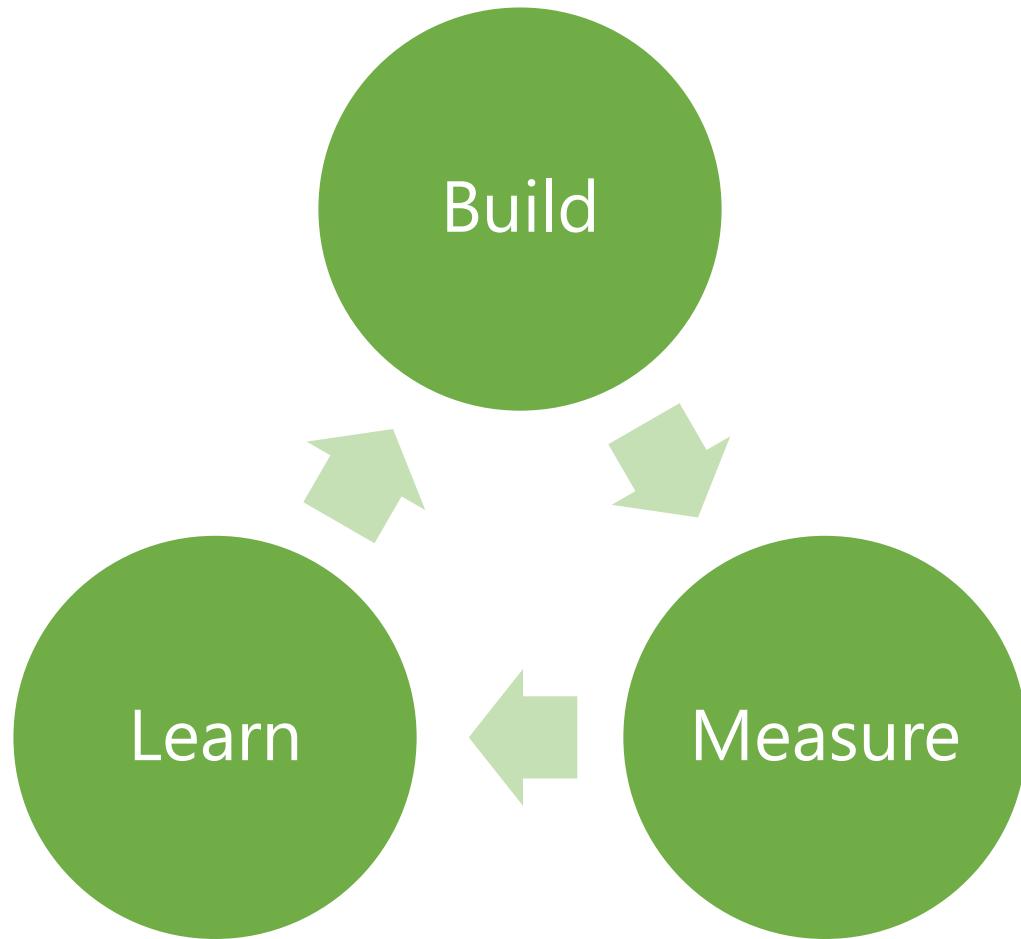


The diagram consists of a large white circle containing two rounded rectangular boxes. The left box is light blue and contains the text 'Build intelligent software'. The right box is dark blue and contains the text 'Improve development practices'.

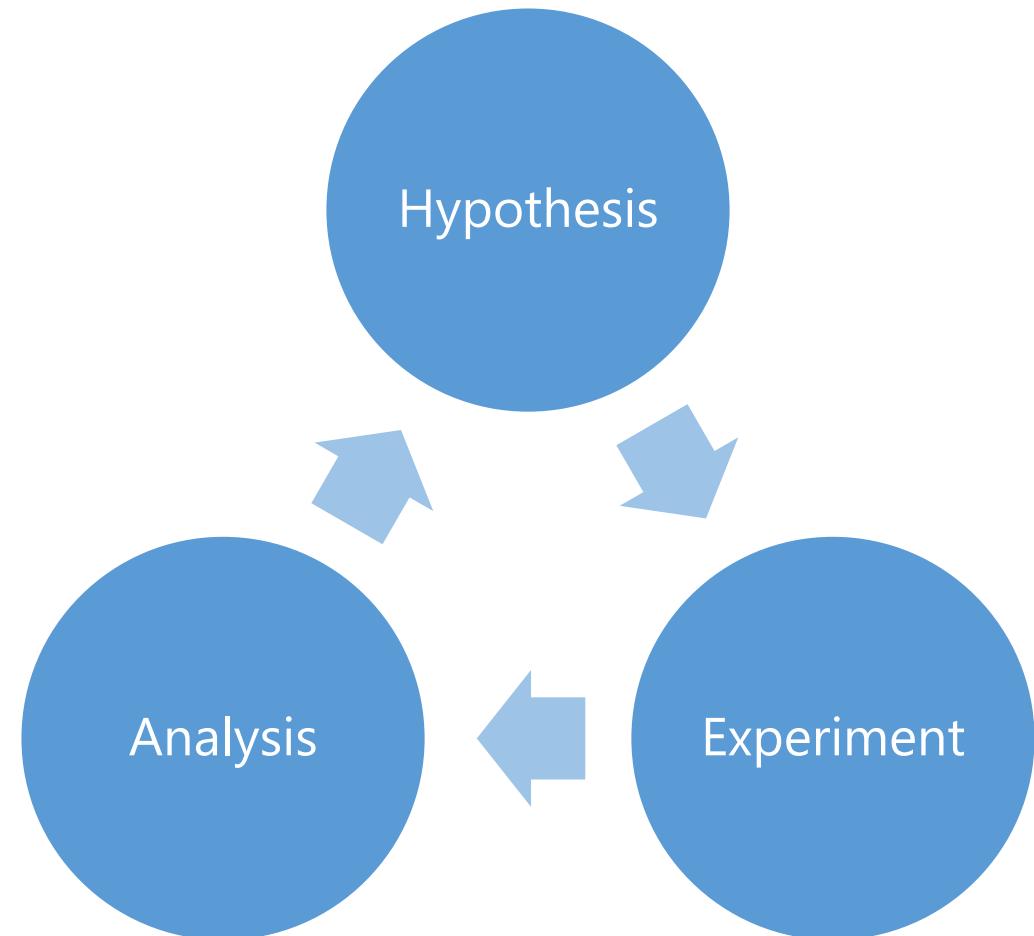
Build
intelligent
software

Improve
development
practices

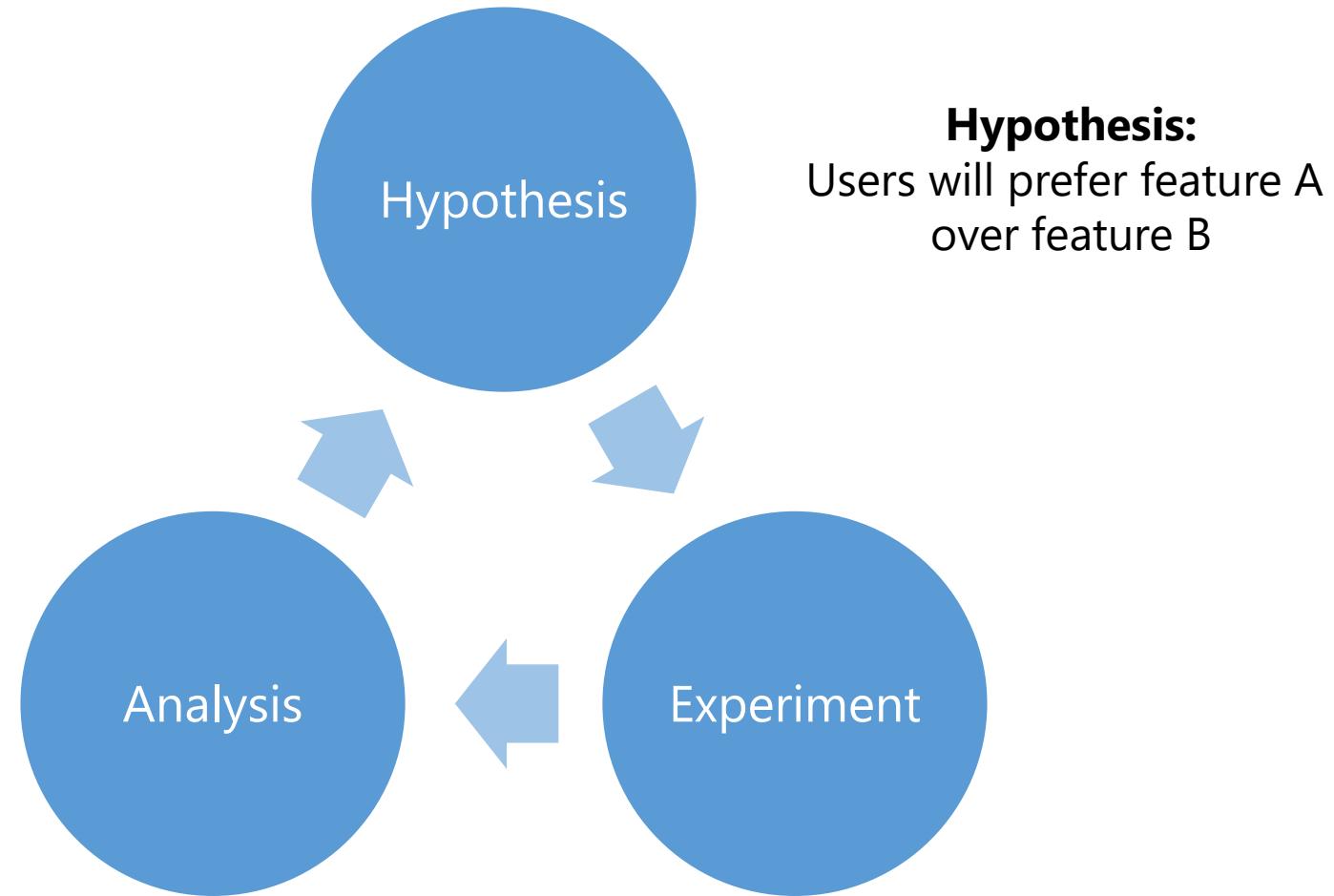
Data-Driven Decision Making



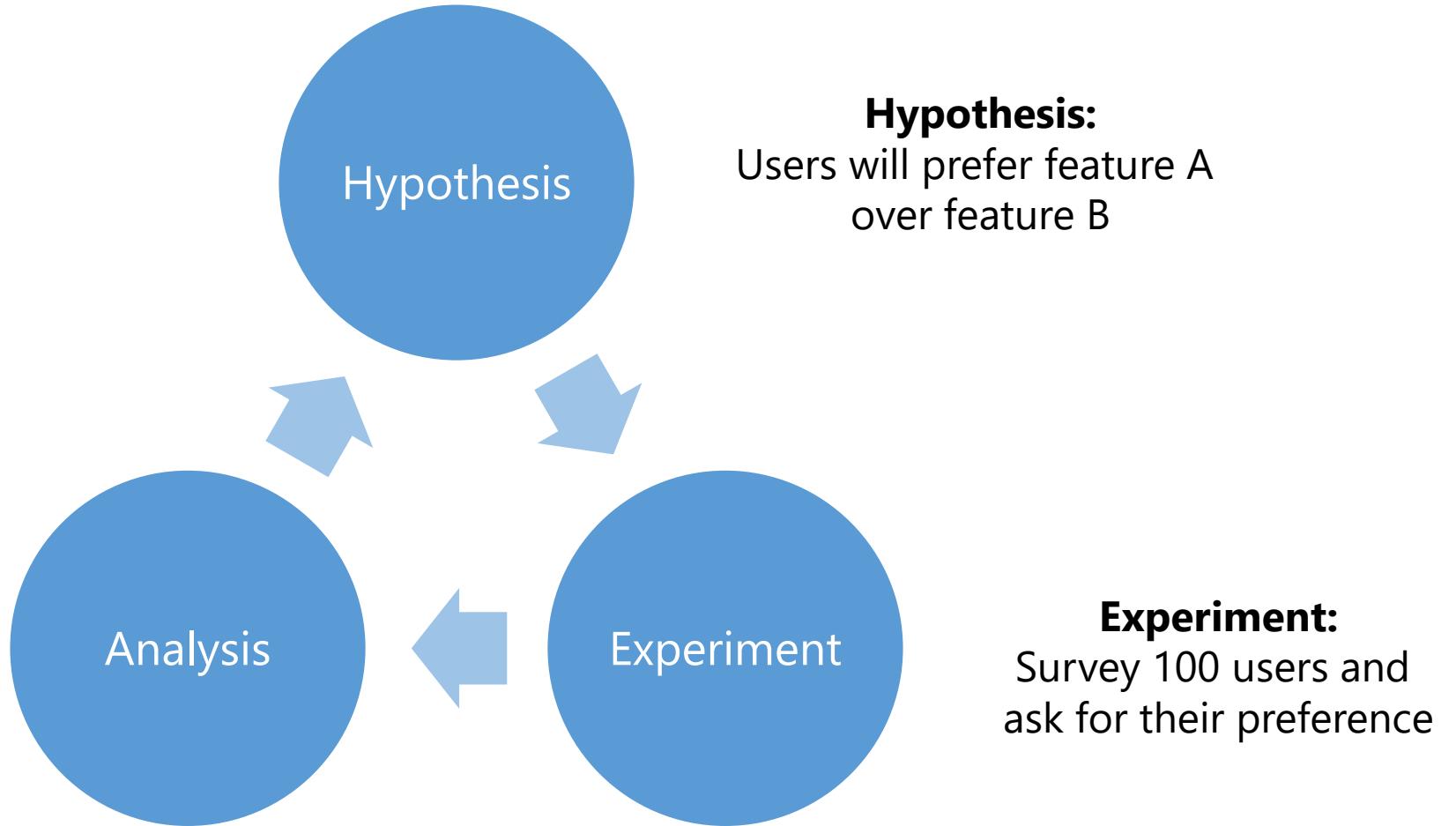
Hypothesis-Driven Development



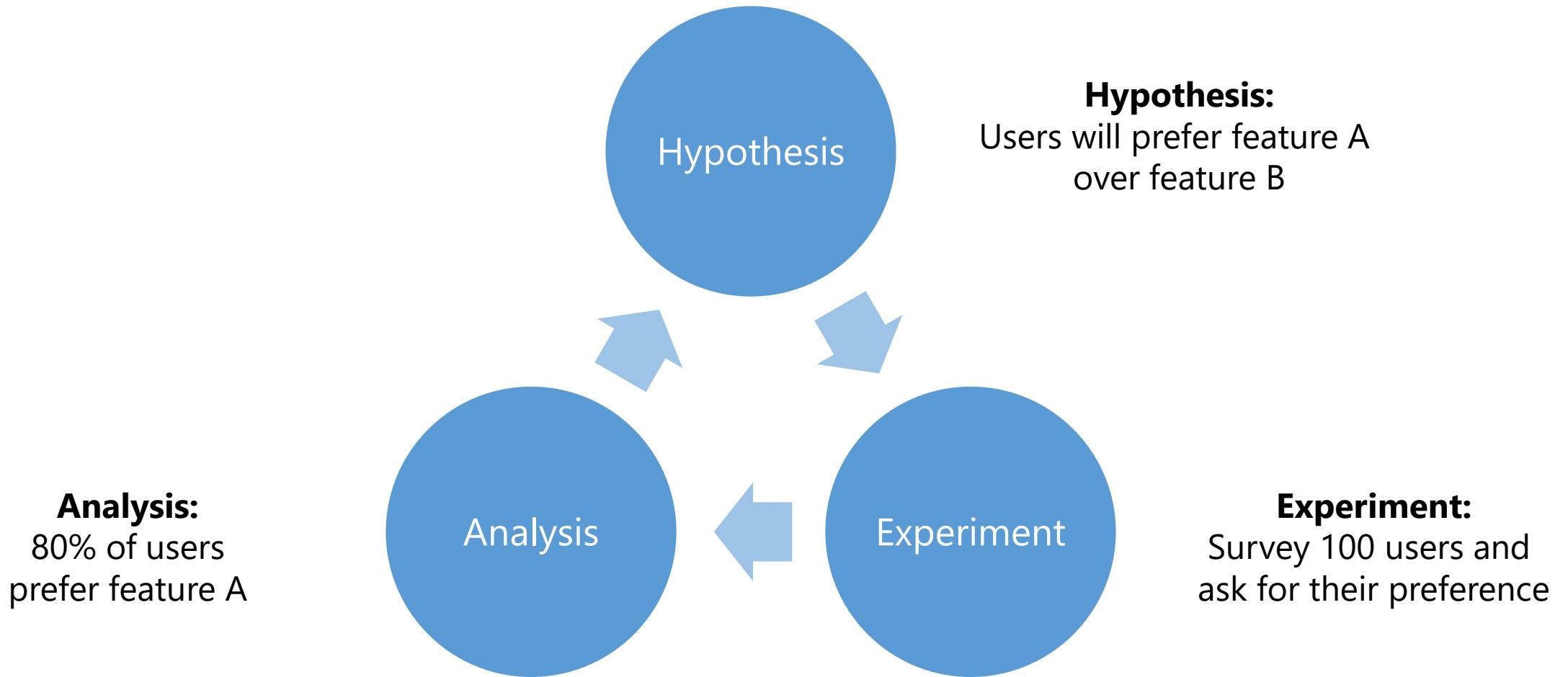
Hypothesis-Driven Development



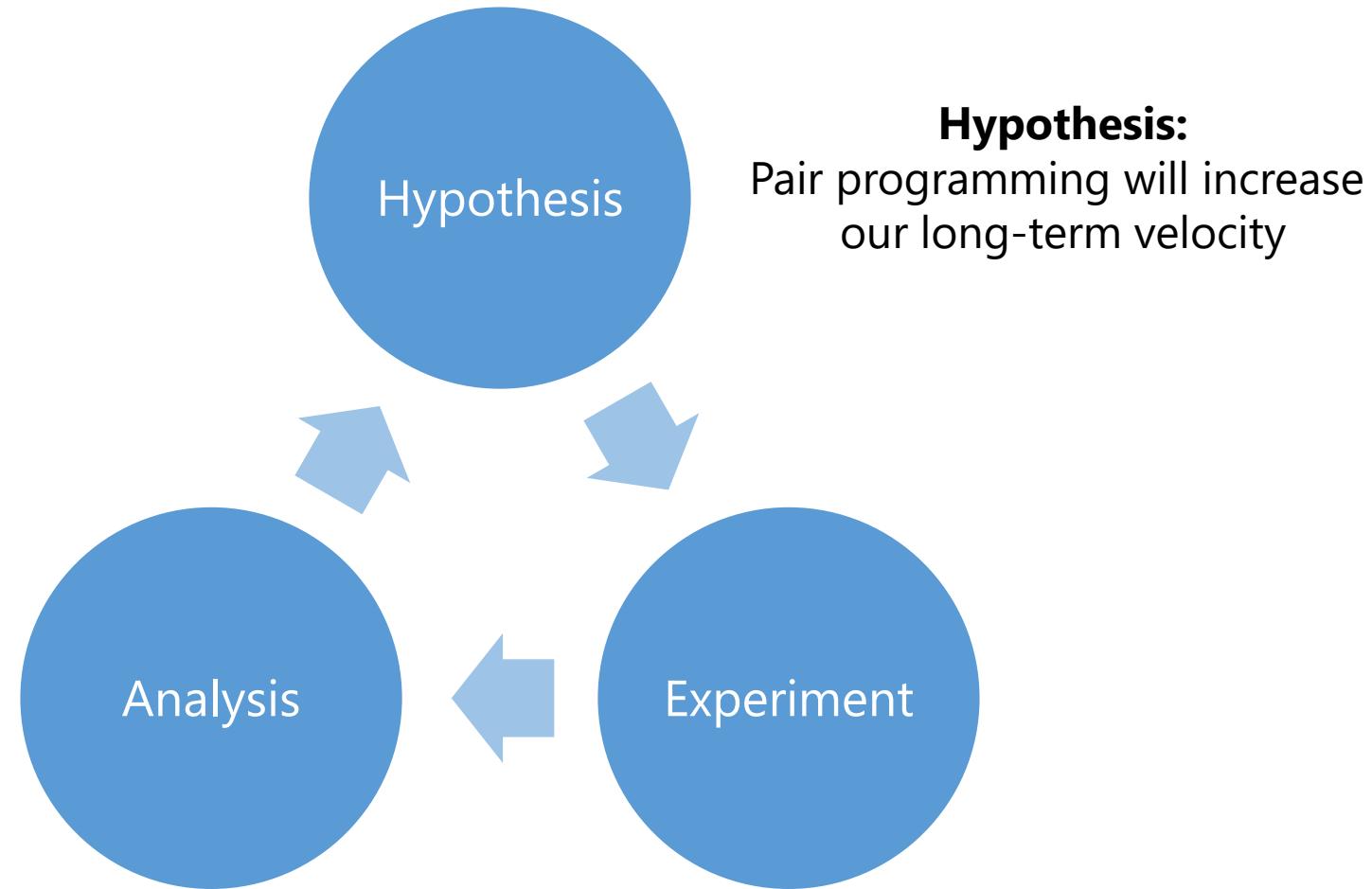
Hypothesis-Driven Development



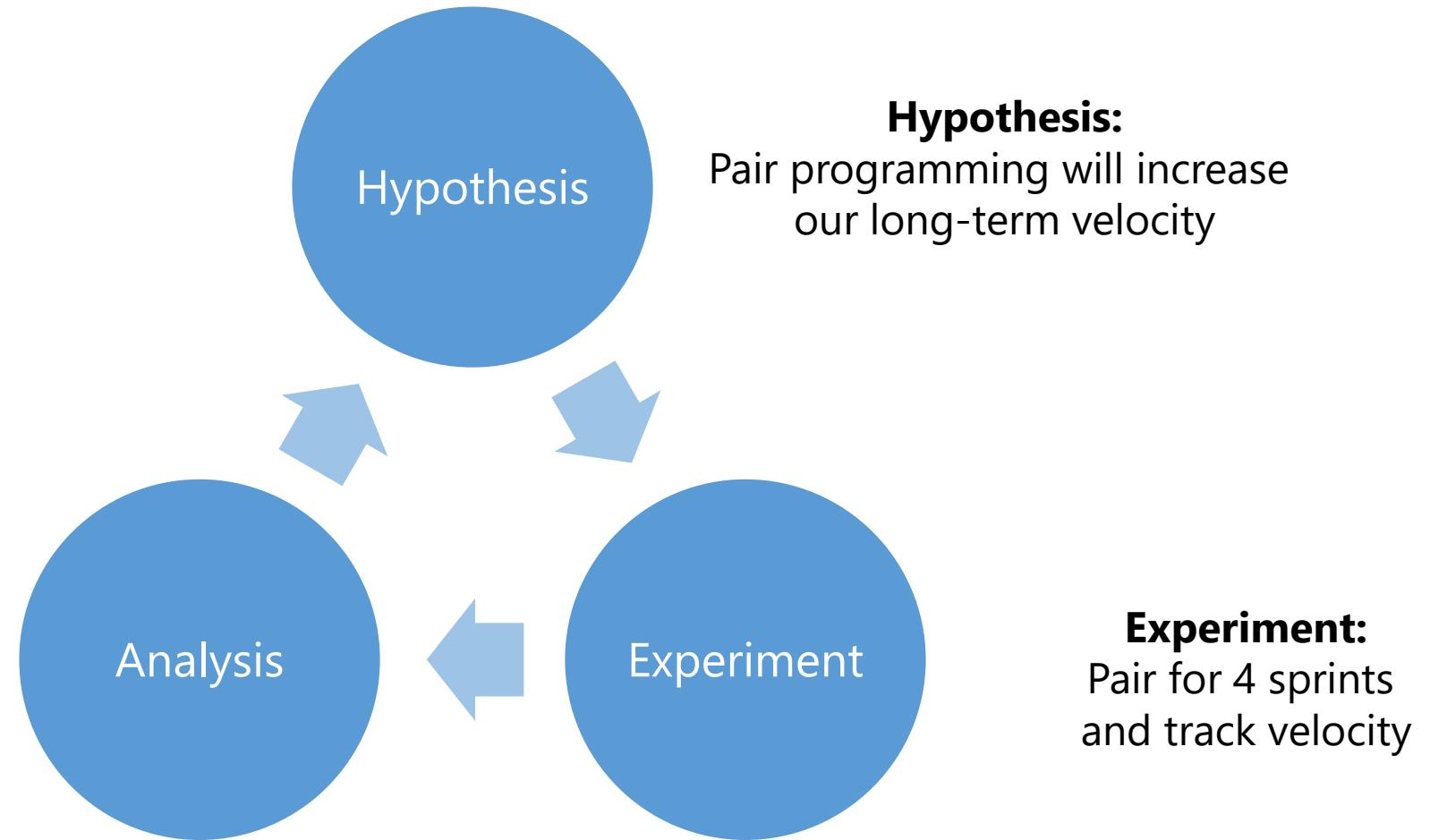
Hypothesis-Driven Development



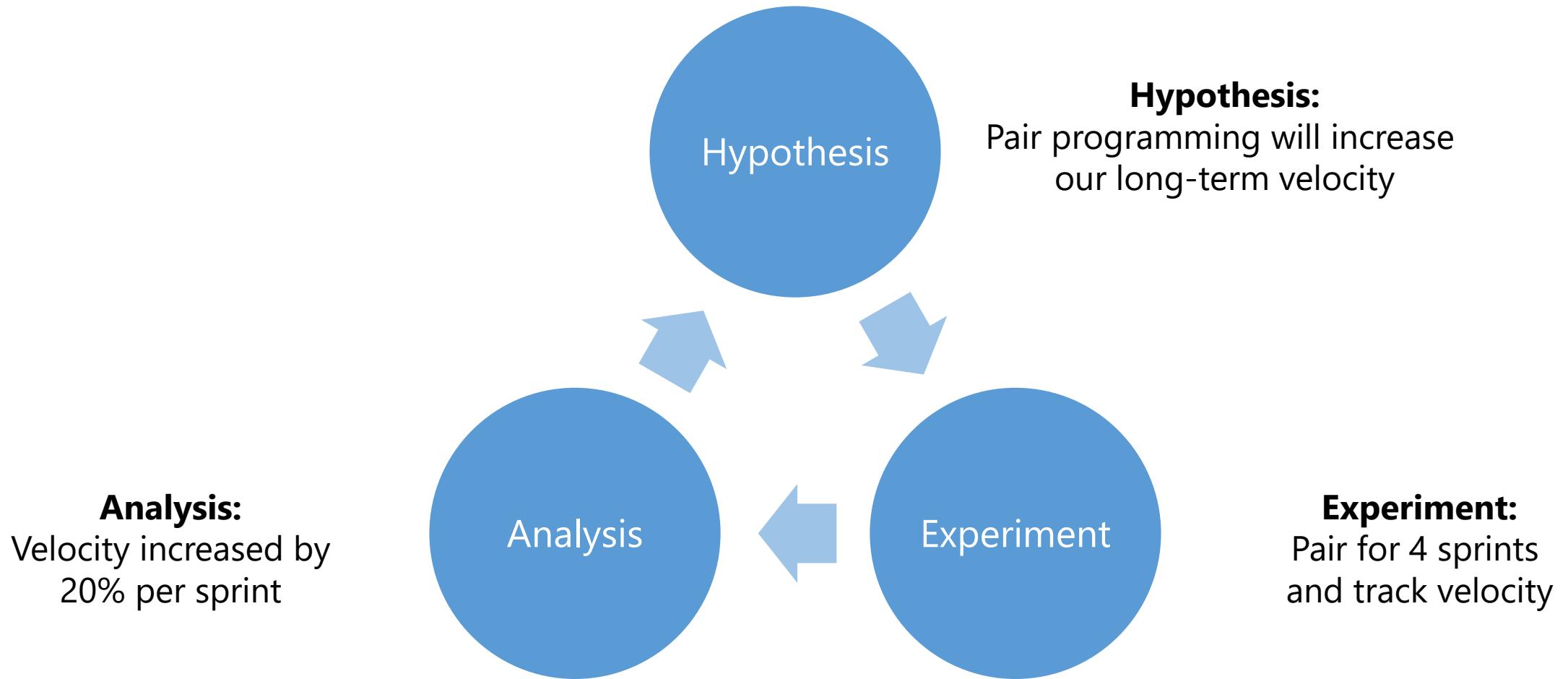
Hypothesis-Driven Development



Hypothesis-Driven Development



Hypothesis-Driven Development



Hypothesis Stories

<Hypothesis>

We assume that <hypothesis>

Will result in <outcome>

We will have succeeded when <measurable result>

Hypothesis Stories

Pair Programming Hypothesis

We assume that pair programming

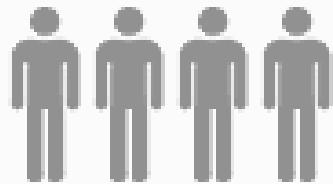
Will result in higher long-term velocity

We will have succeeded when we have seen a 10%
or greater increase in velocity after 4 sprints.

A/B Testing



A/B Testing



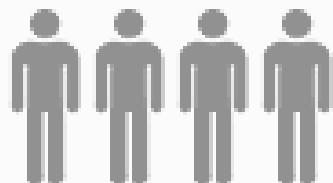
50 % visitors
see variation A



Variation A



23%
conversion



50 % visitors
see variation B

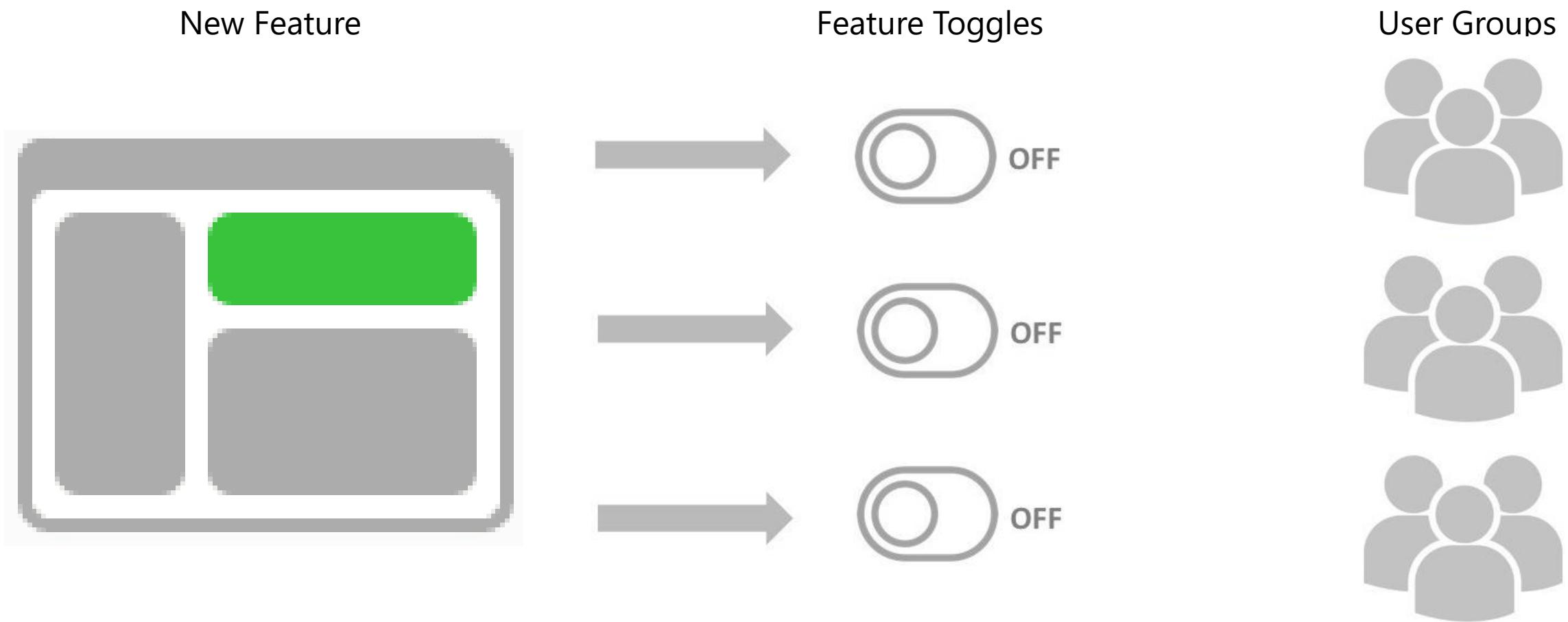


Variation B

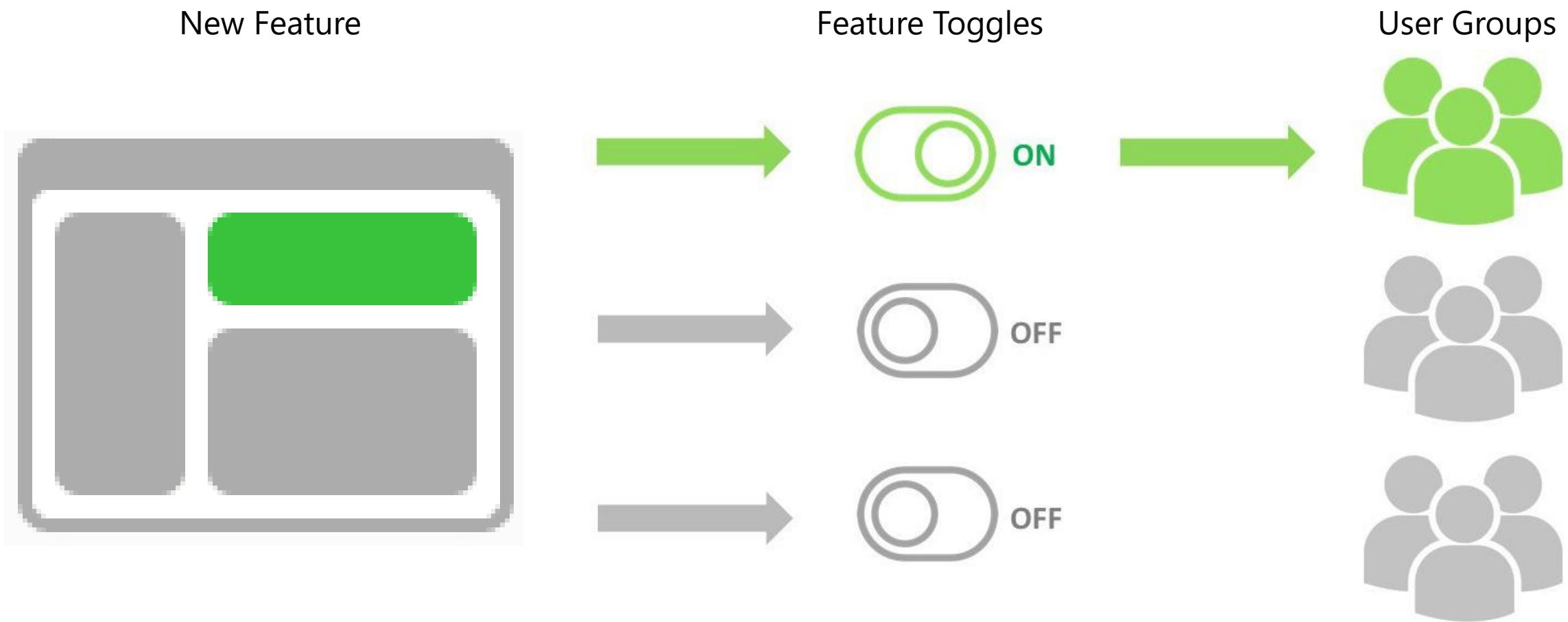


11%
conversion

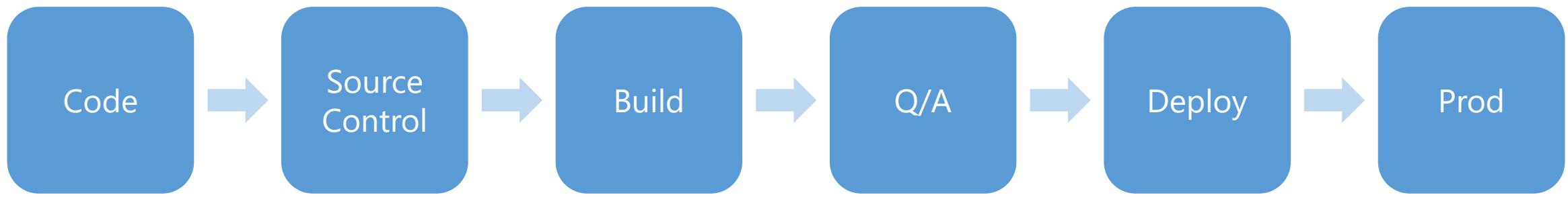
Feature Toggles



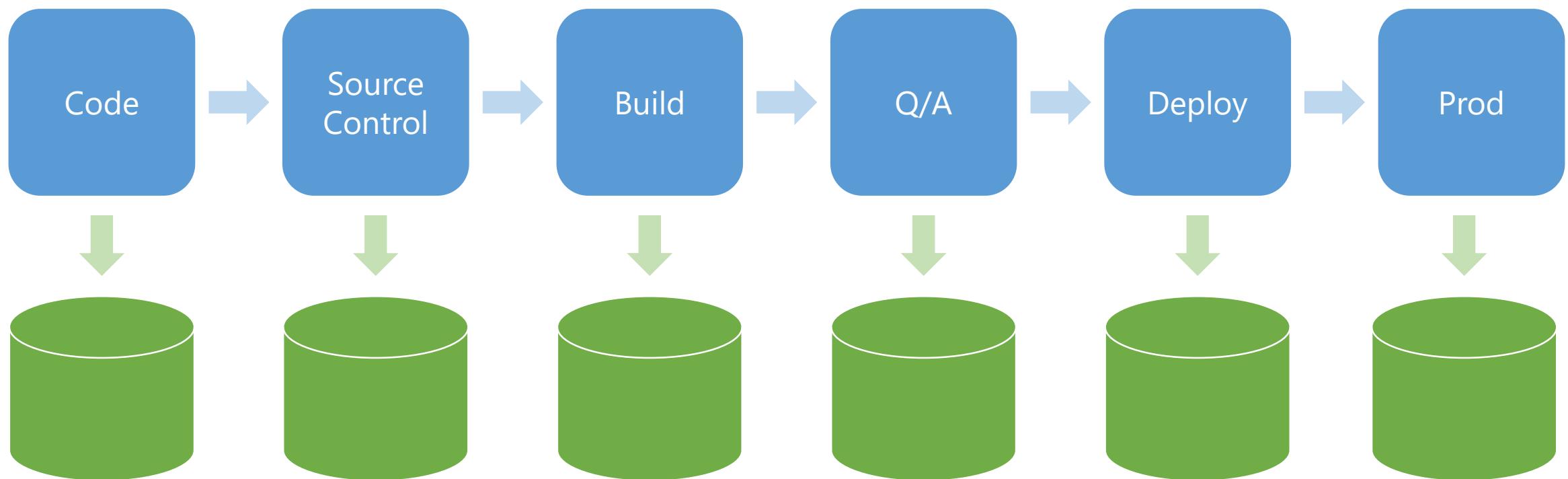
Feature Toggles



DevOps Pipeline



DevOps Pipeline



Code Quality Metrics

Source: NDepend

Source Control Metrics

Build Metrics

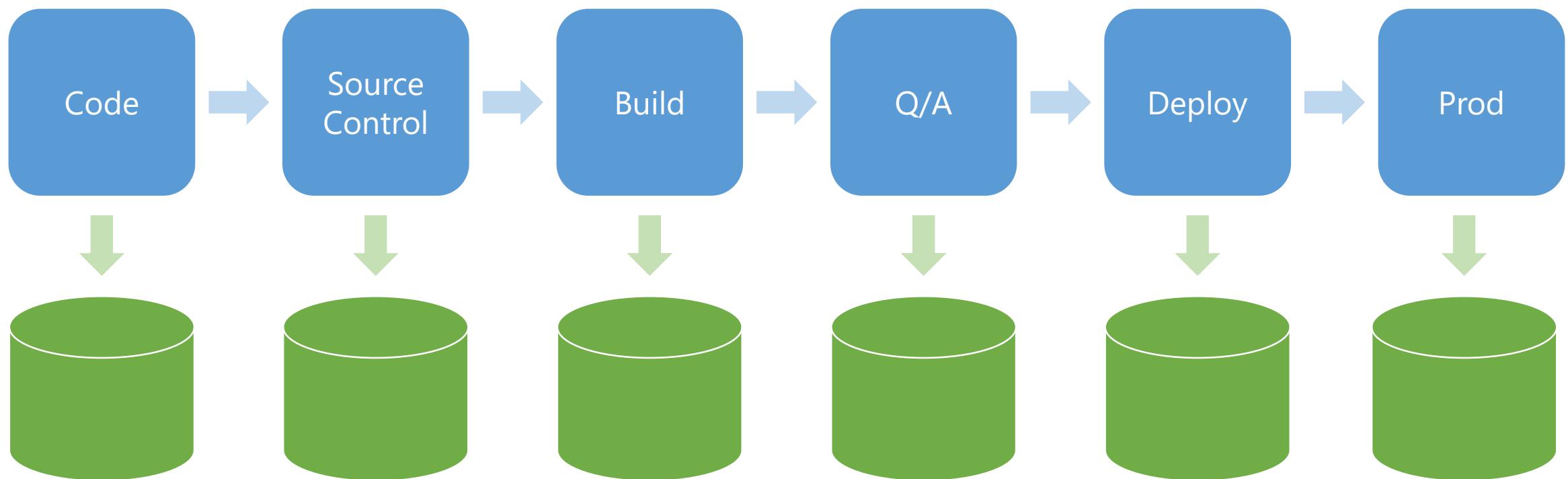
Q/A Metrics

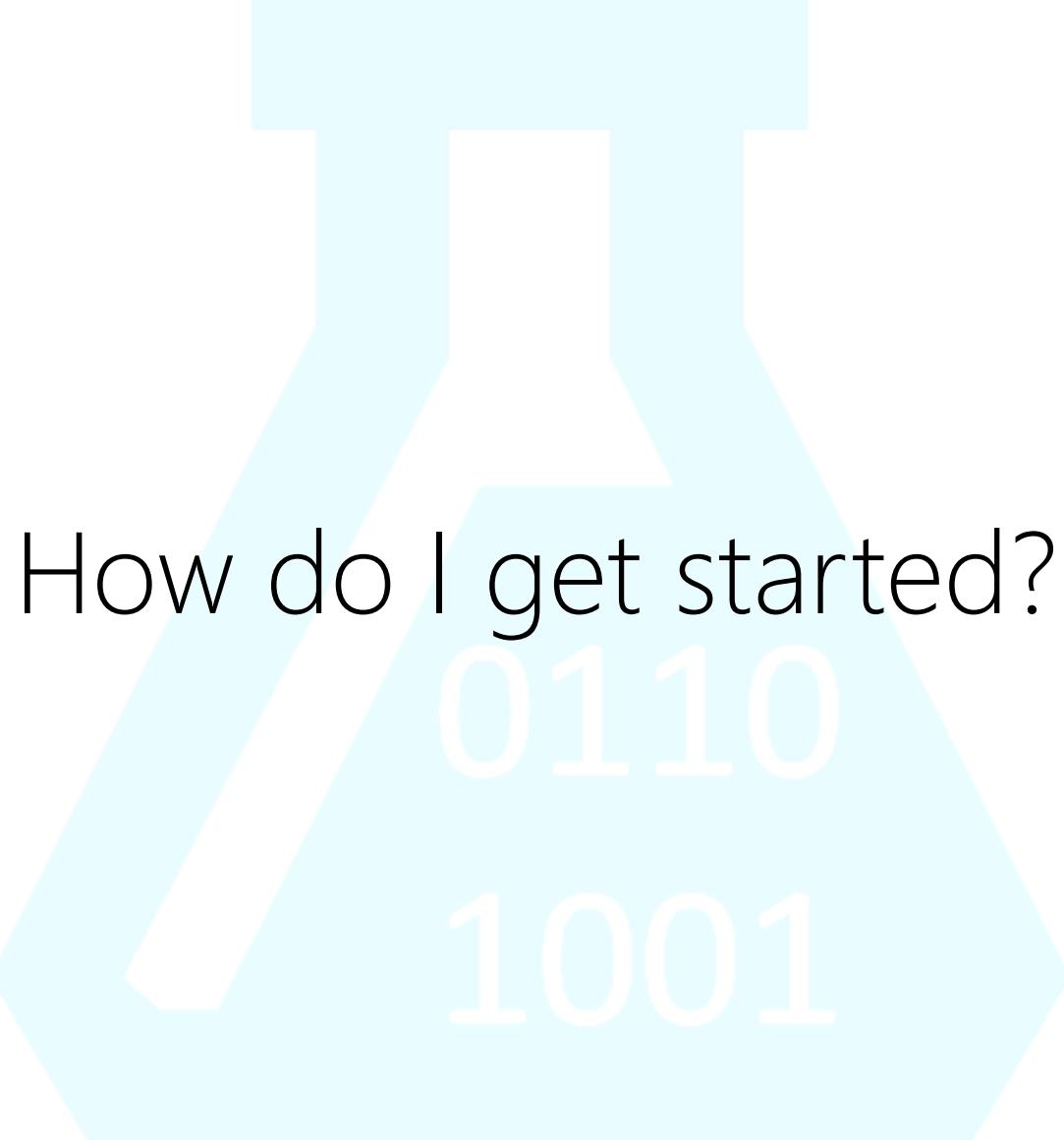


Deployment Metrics

Software Telemetry

DevOps Pipeline



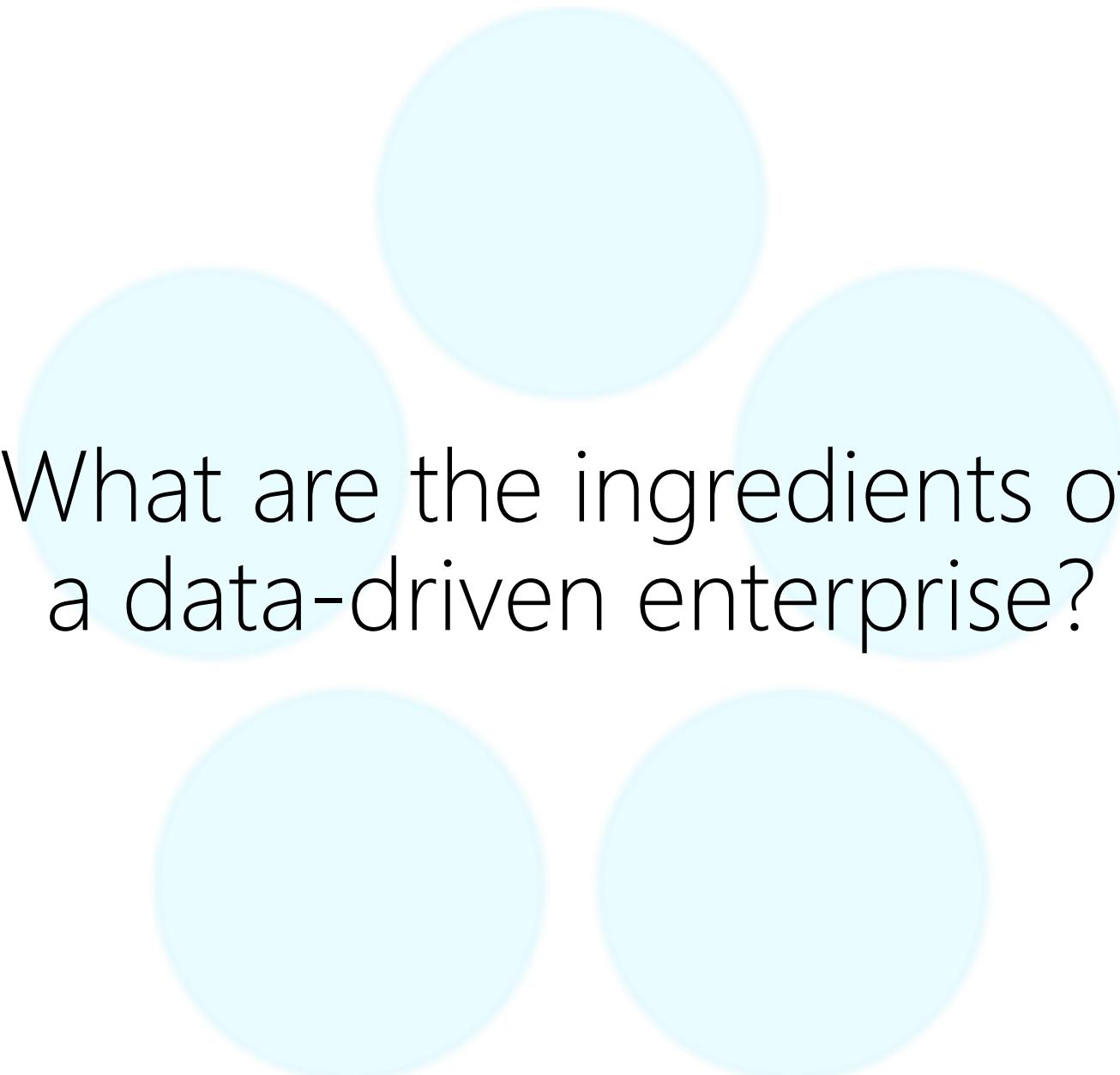


How do I get started?

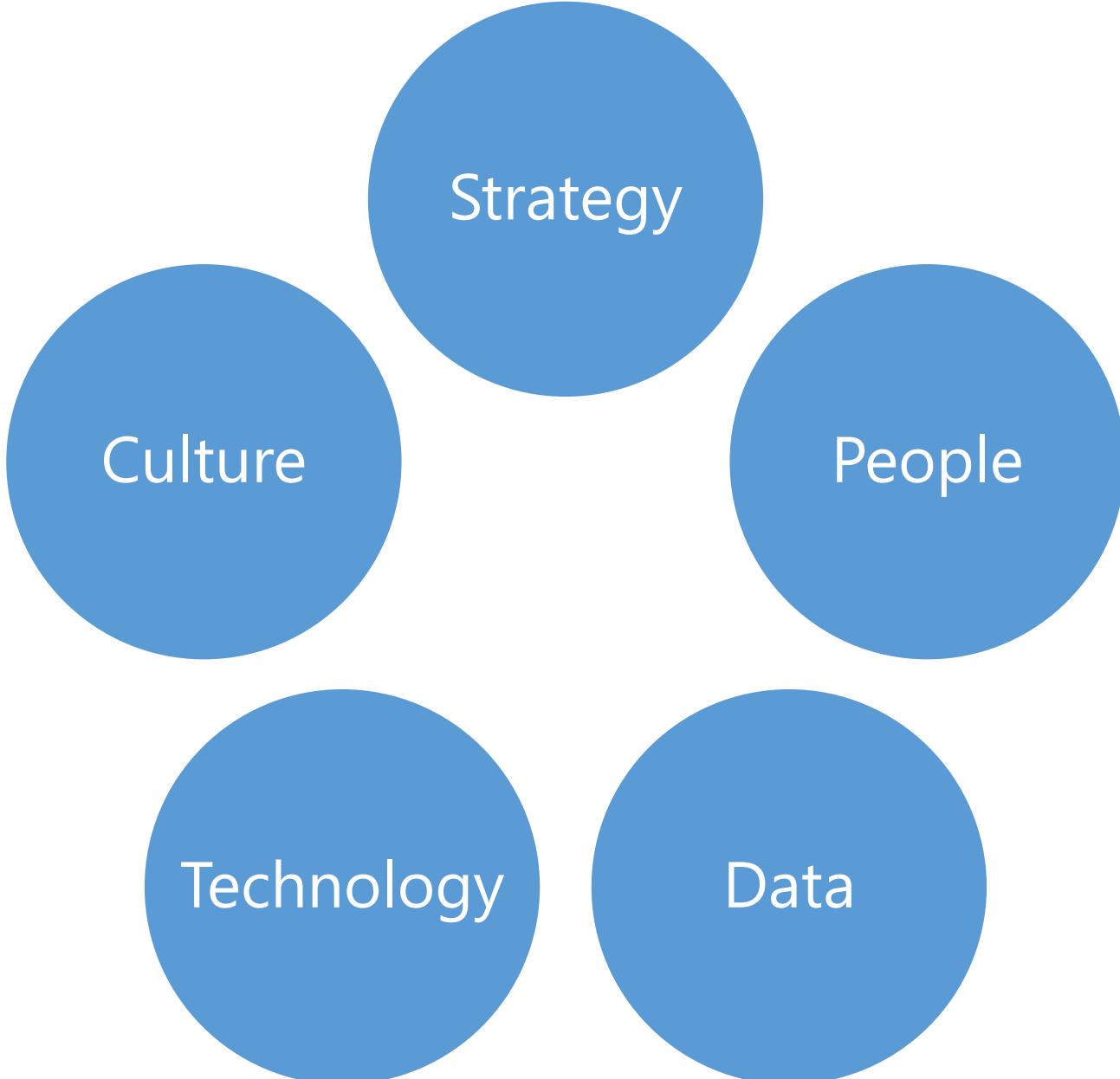
0110

1001

?



What are the ingredients of
a data-driven enterprise?



Strategy

Culture

People

Technology

Data



Strategy

A man with dark hair and a beard is sitting at a desk, looking down at a laptop screen. He is wearing a white shirt and a dark jacket. In the background, there is a large computer monitor displaying a colorful bar chart. The overall image is slightly blurred.

People

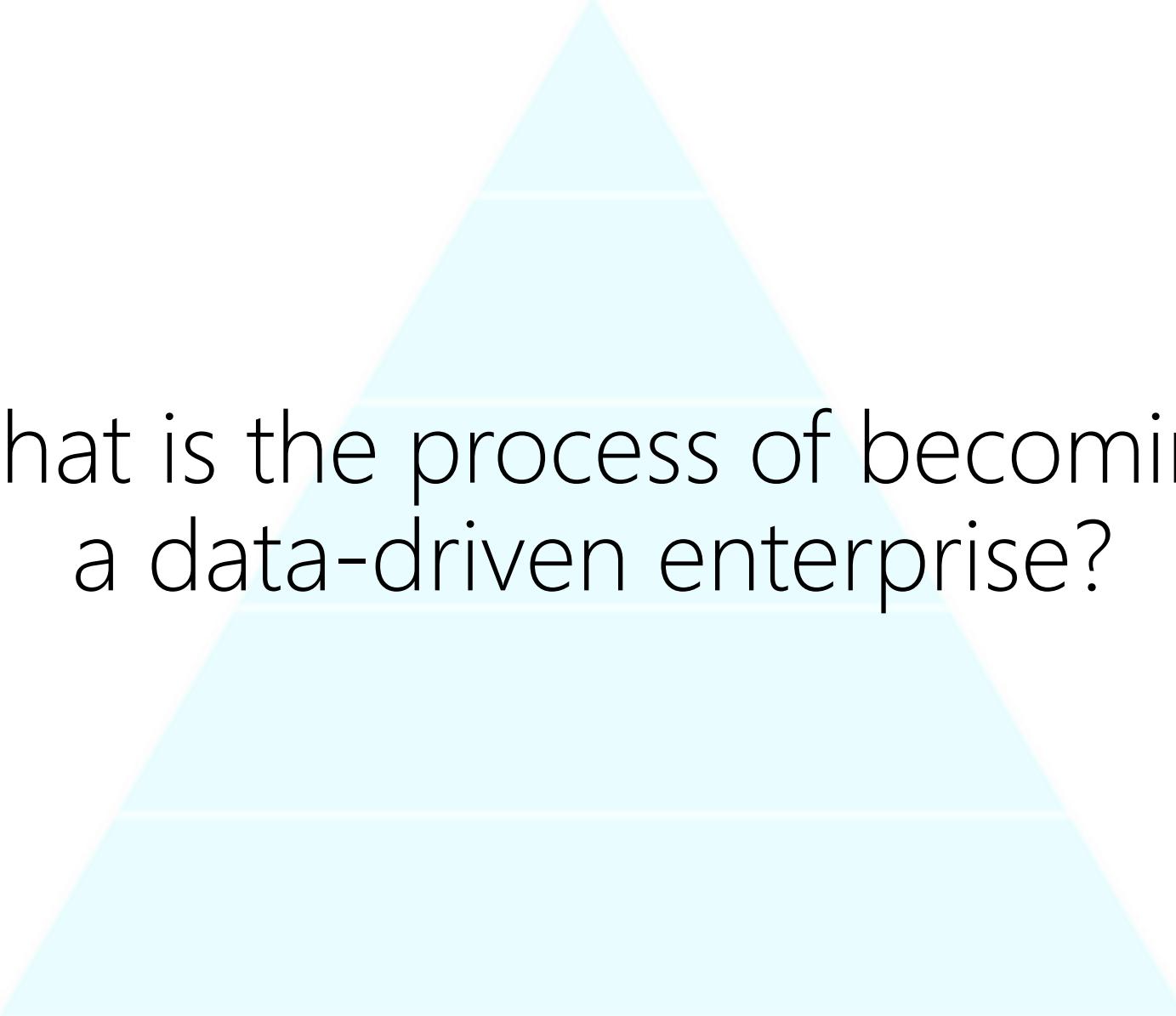
Data



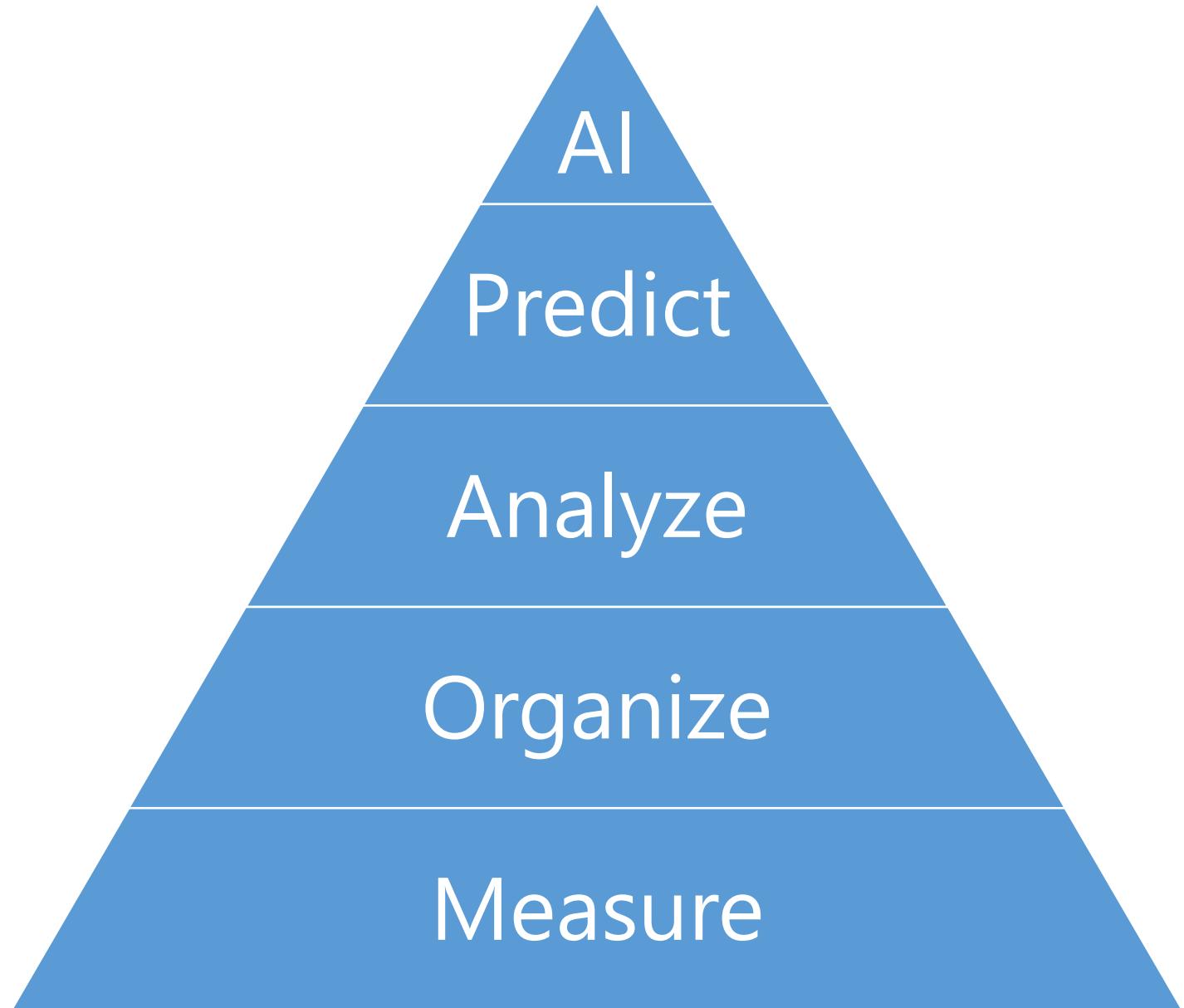
Technology

A person is sitting at a desk, working on a laptop. The background is blurred, showing what appears to be a colorful painting or mural on the wall behind them.

Culture



What is the process of becoming
a data-driven enterprise?



1. Measure

Transactions

Instrumentation

Logging

Surveys

Digitization

External data



Measure

2. Organize

Transform

Clean

Store

Data ETL

Data Warehouse

Data Lake



Organize

Measure

3. Analyze

Reports

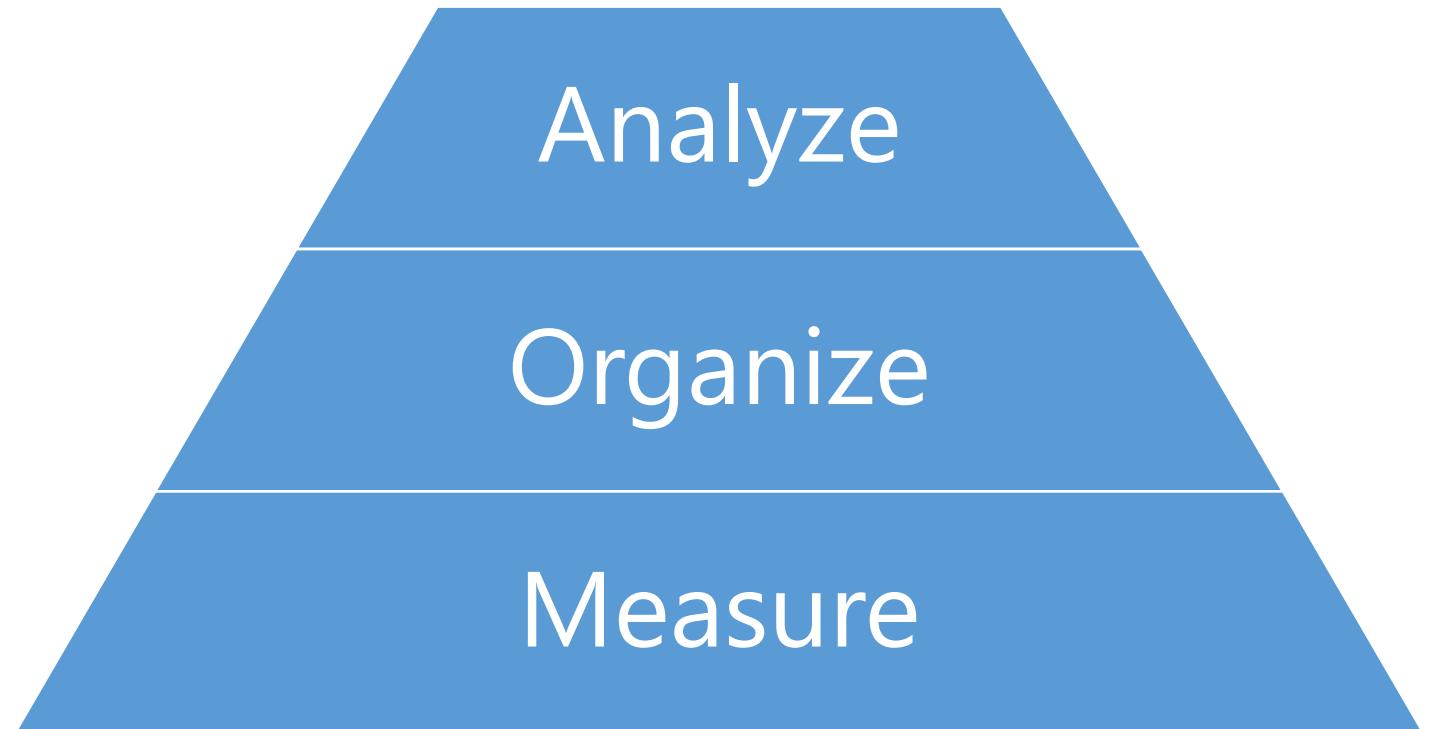
Dashboards

KPI monitors

Decision support

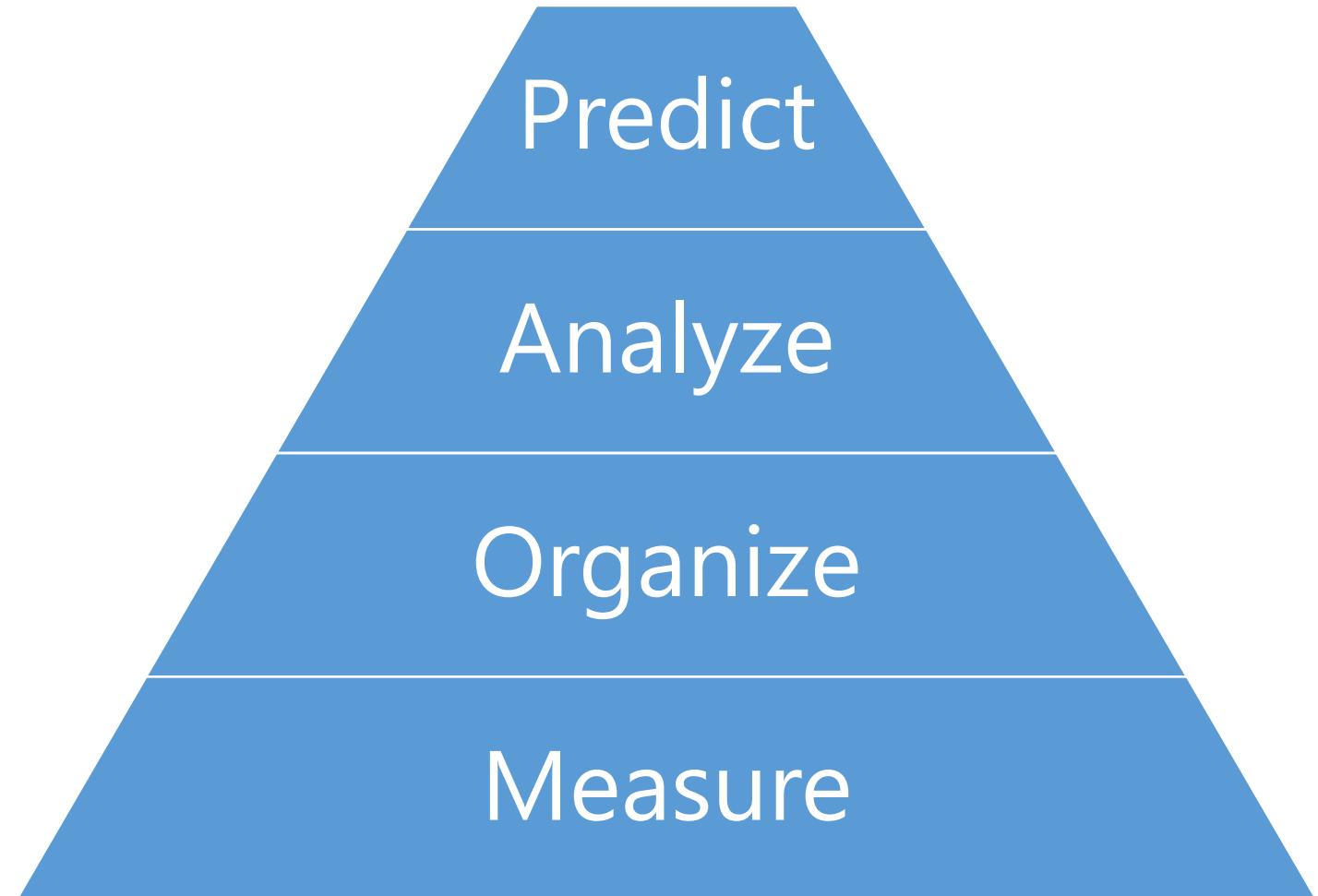
Descriptive analytics

Diagnostic analytics



4. Predict

Predictive analytics
Prescriptive analytics
Machine learning
Hypothesis testing
Experimentation

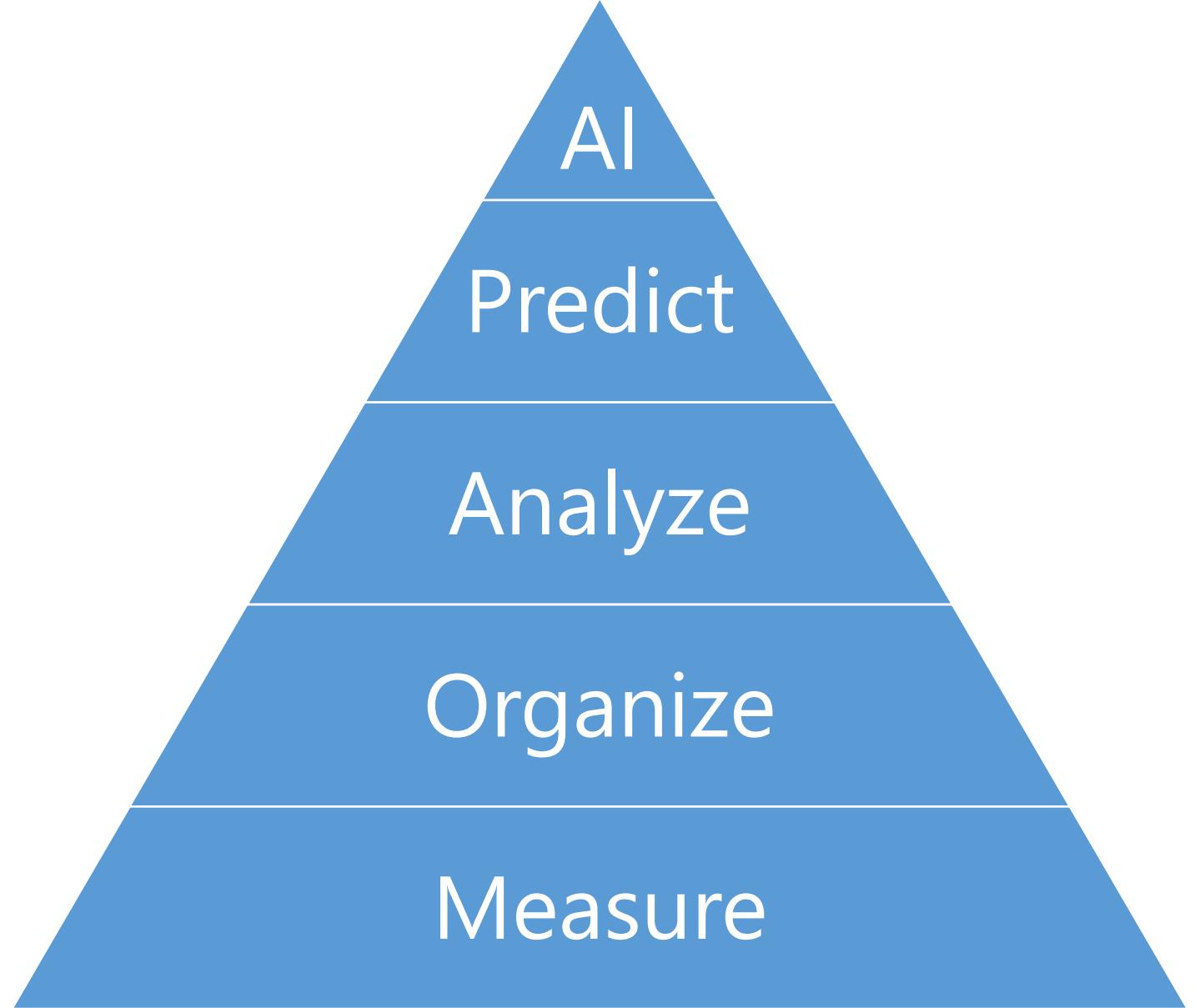


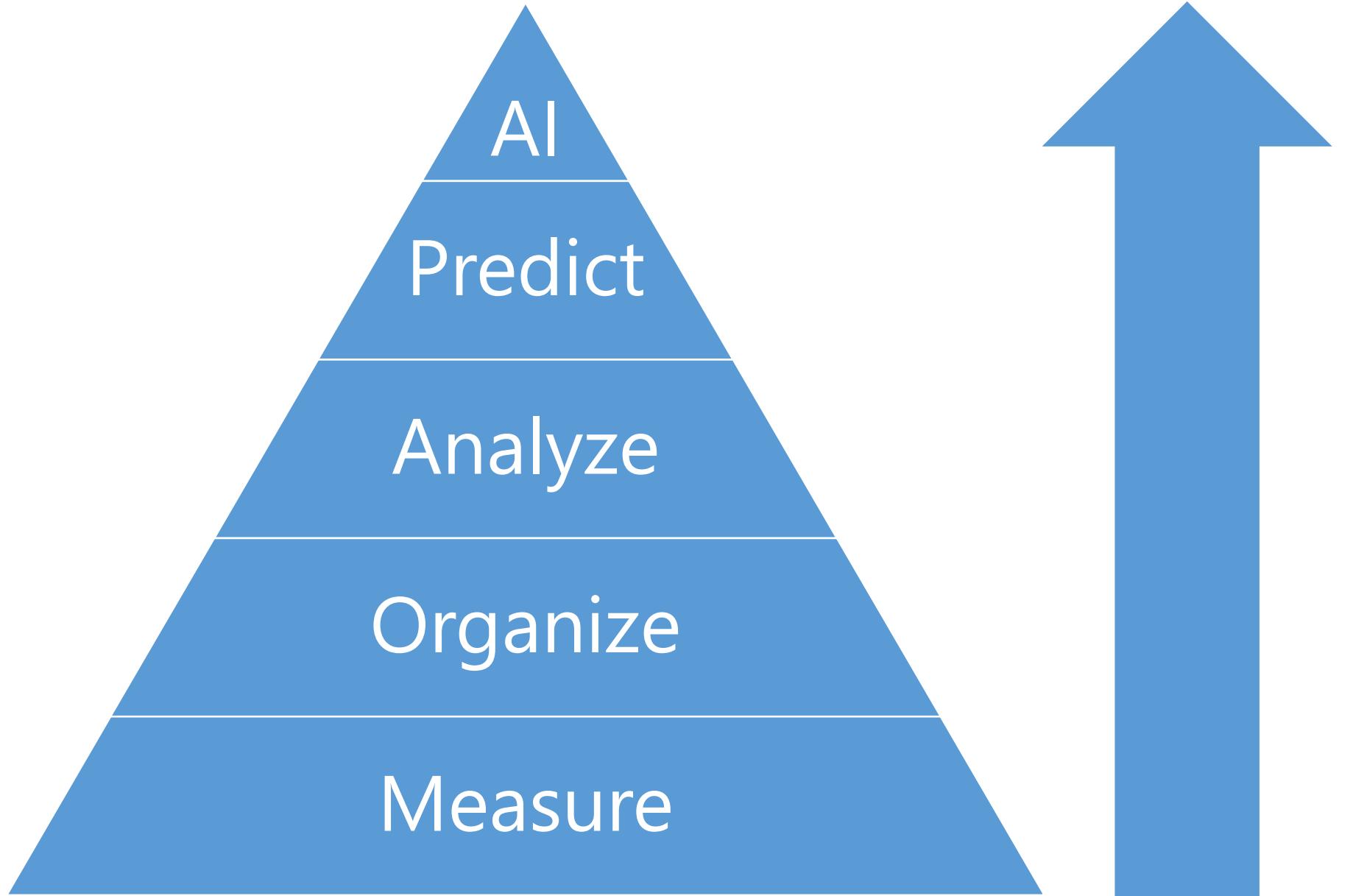
5. Automate

Artificial intelligence

Expert systems

Deep learning





Advice for Success

- Get buy-in from leadership
- Focus on low-hanging fruit
- Don't silo data science teams
- Democratize your data

Advice for Success

Embrace smart failure

Focus on feedback

Embed data collection

Avoid the Observer Effect



Where to Go Next?

0110

1001

Where to Go Next

Data Camp: <https://www.datacamp.com>

Pluralsight: <https://www.pluralsight.com>

Coursera: <https://www.coursera.org>

Pluralsight Courses

Data Science: The Big Picture

Data Science with R

Exploratory Data Analysis with R

Data Visualization with R (3-part)

Deep Learning: The Big Picture



<https://www.pluralsight.com/authors/matthew-renze>

News

2017-08-25 - Invitation to Speak at Devoxx Morocco

Very excited to announce that I've been invited to give a keynote in Casablanca at [Devoxx Morocco](#) in November. My keynote presentation will be on [Artificial Intelligence](#).



2017-08-16 - Invitation to Speak at Microsoft Ignite

I've been invited to speak at [Microsoft Ignite](#) in Orlando, Florida in September. This will be my first time speaking at Ignite. Talks will include both Data Science and Machine Learning with R.



Matthew is a data science consultant, author for [Pluralsight](#), international public speaker, a [Microsoft MVP](#), [ASPIInsider](#), and open-source software contributor.

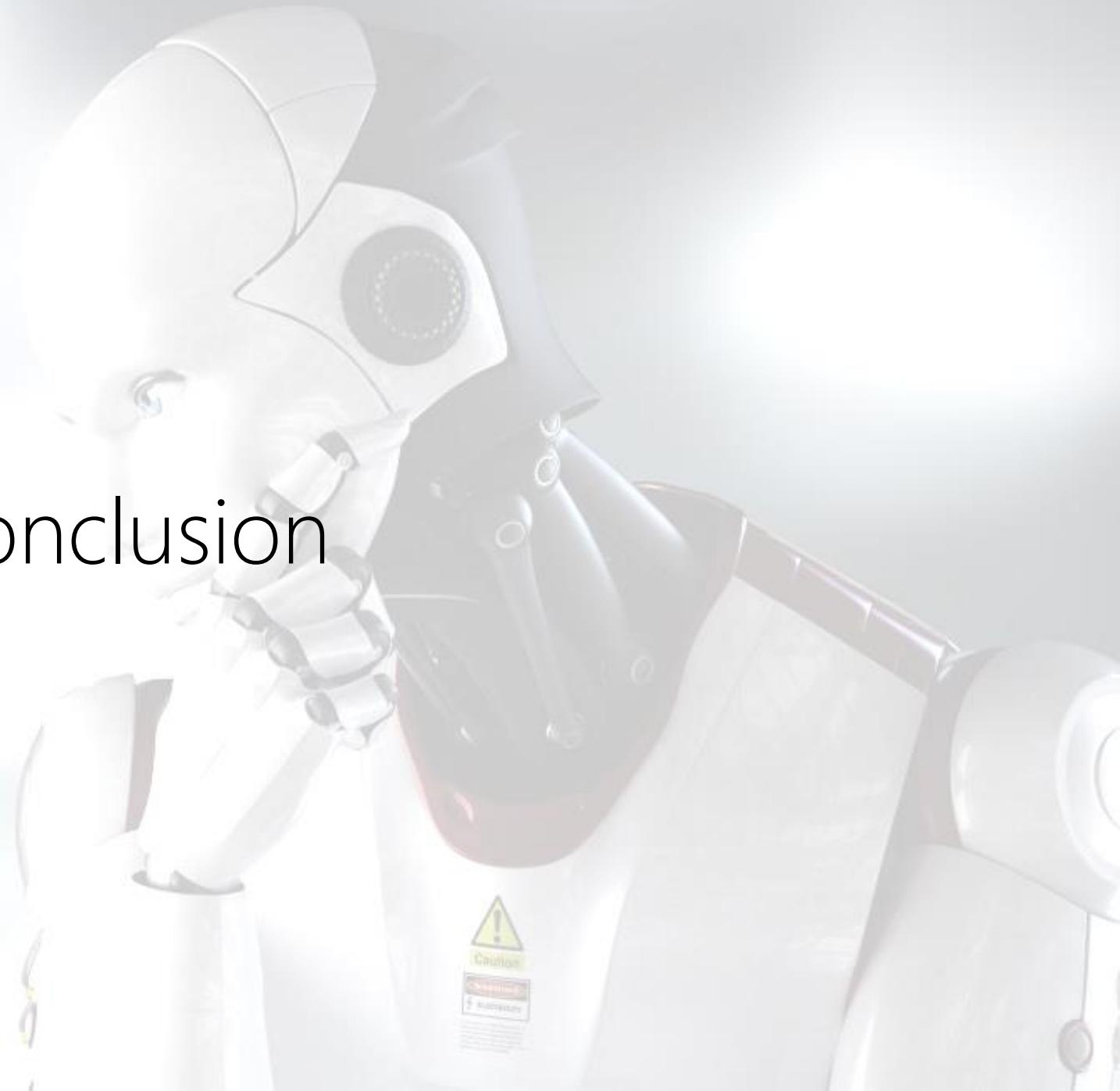
2017-08-14 - Dev on Fire Interview

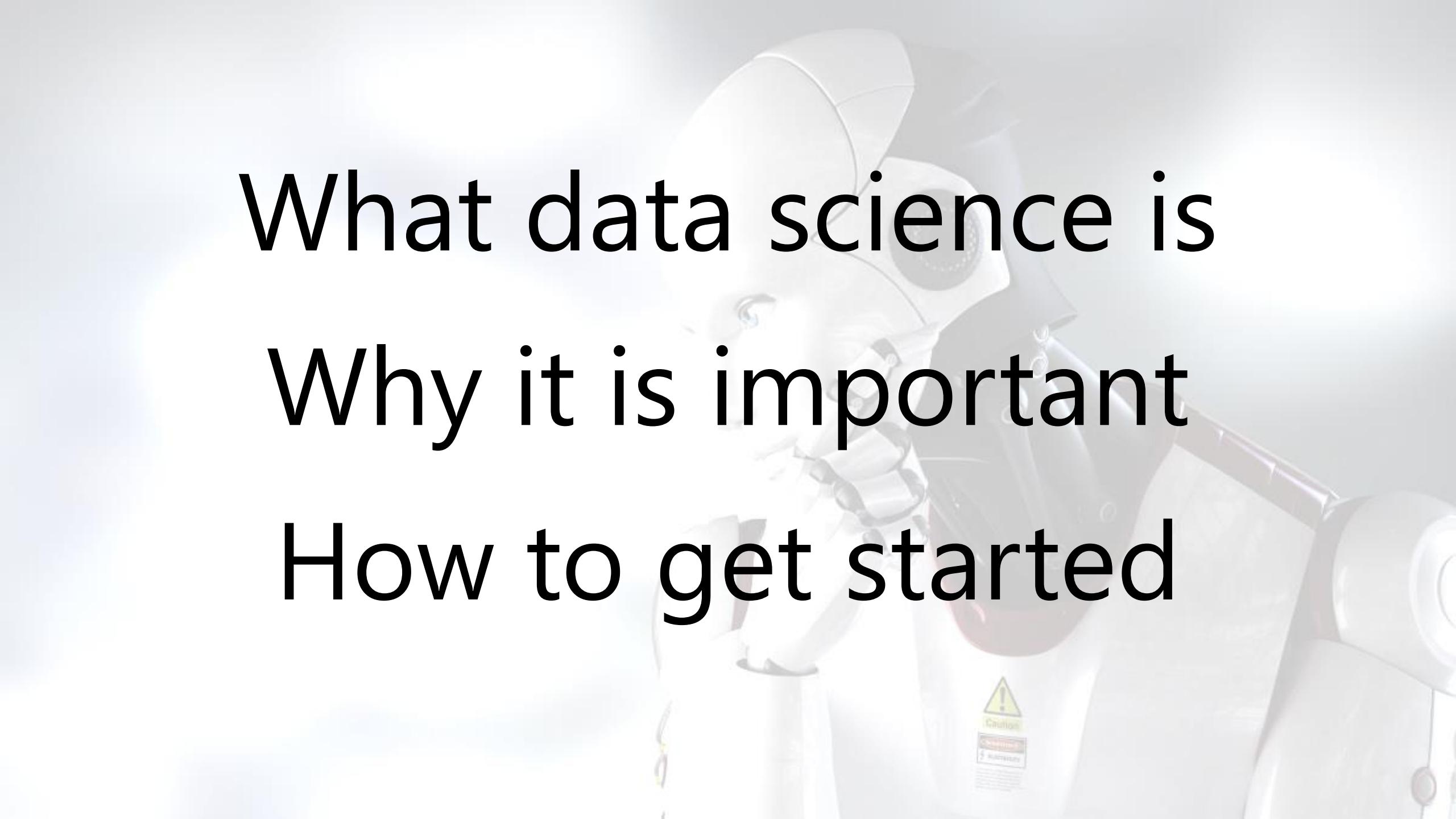
Feedback

Very important to me!
What did you like?
What could I improve?



Conclusion



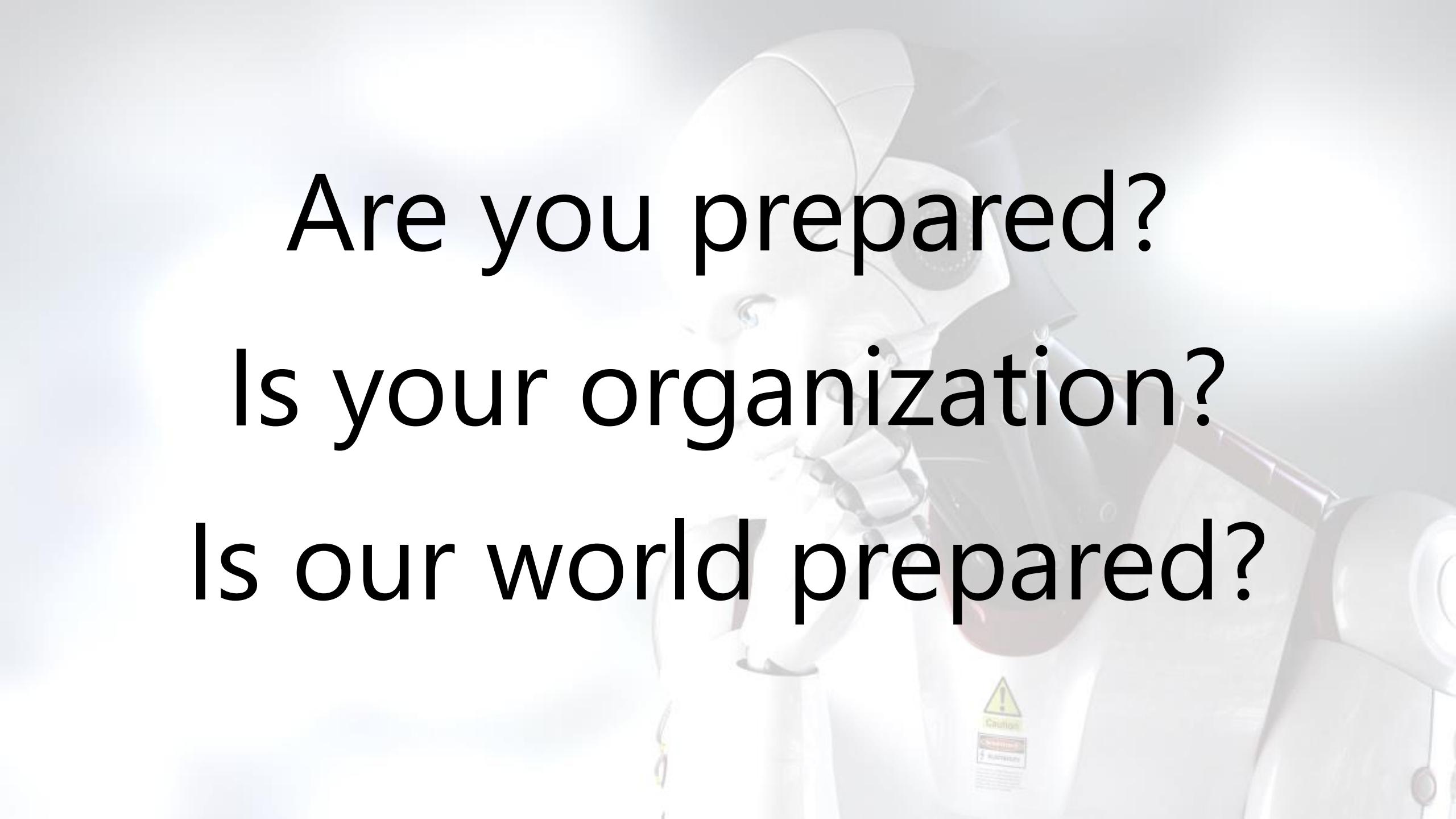
A grayscale photograph showing a person from the waist up, wearing a full-body white protective suit and a respirator mask with a circular filter. They are focused on operating a complex industrial machine with various pipes, valves, and a prominent circular component. The background is slightly blurred.

What data science is

Why it is important

How to get started

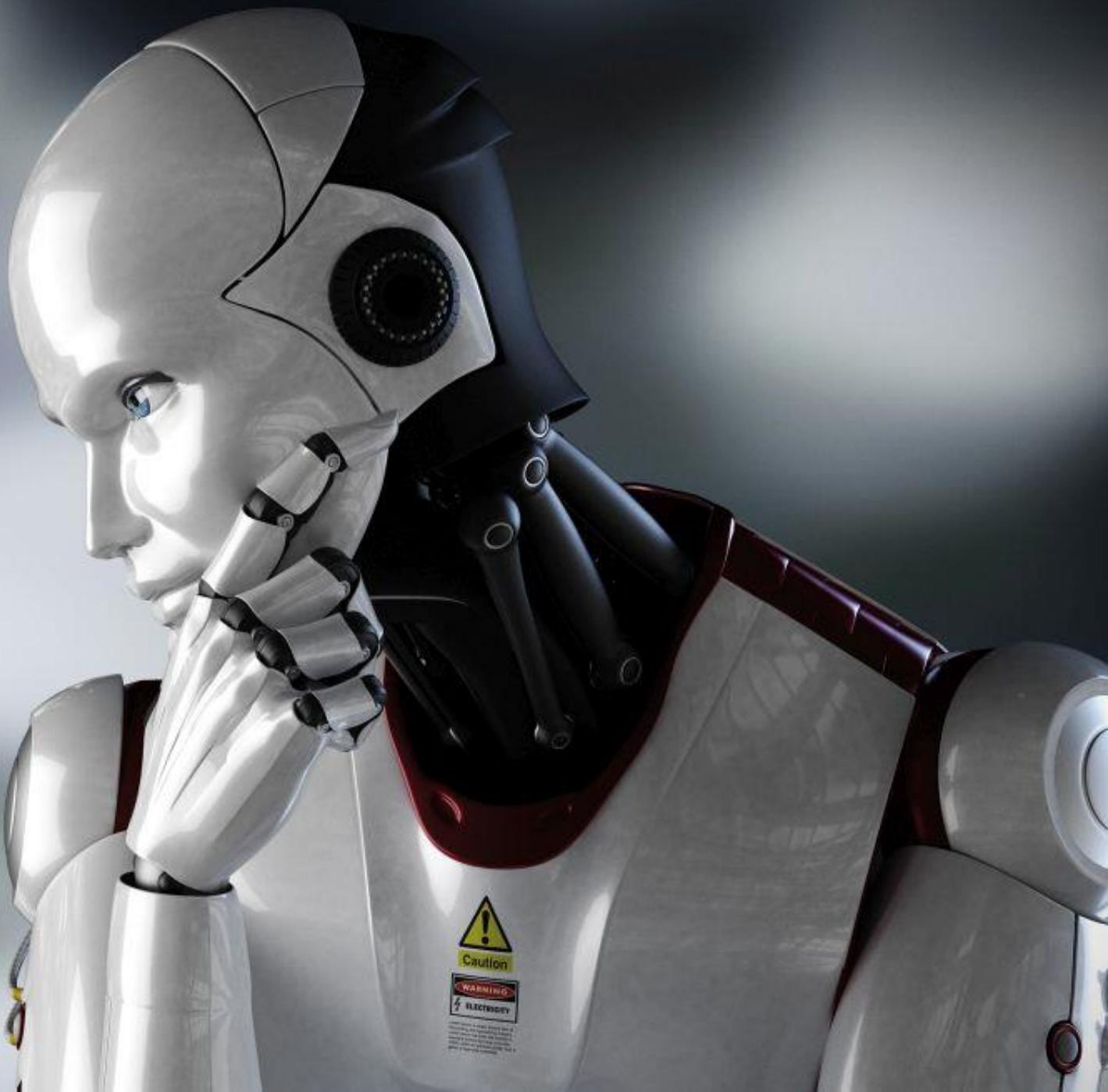




Are you prepared?

Is your organization?

Is our world prepared?



Thank You!

Matthew Renze
Data Science Consultant
Renze Consulting

Twitter: [@matthewrenze](https://twitter.com/matthewrenze)
Email: info@matthewrenze.com
Website: www.matthewrenze.com



TITANIUM SPONSORS



Platinum Sponsors



PLURALSIGHT



V M L



Gold Sponsors

