Accelerate Your Data Science Delivery with Integrated Notebooks and IBM BigInsights

Chris Snow
chris.snow@uk.ibm.com

October 27, 2016

World of Watson 2016

Please note

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice and at IBM’s sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

What's changed in the world today?

A world awash with data

The re-invention of the world in code

The advent of cognitive computing

Yet only 15% of organizations have the capability to leverage data and advanced analytics across their organization.

HBR Insight Economy Study

Digital disruption is upon all of us … and it is aided by data!

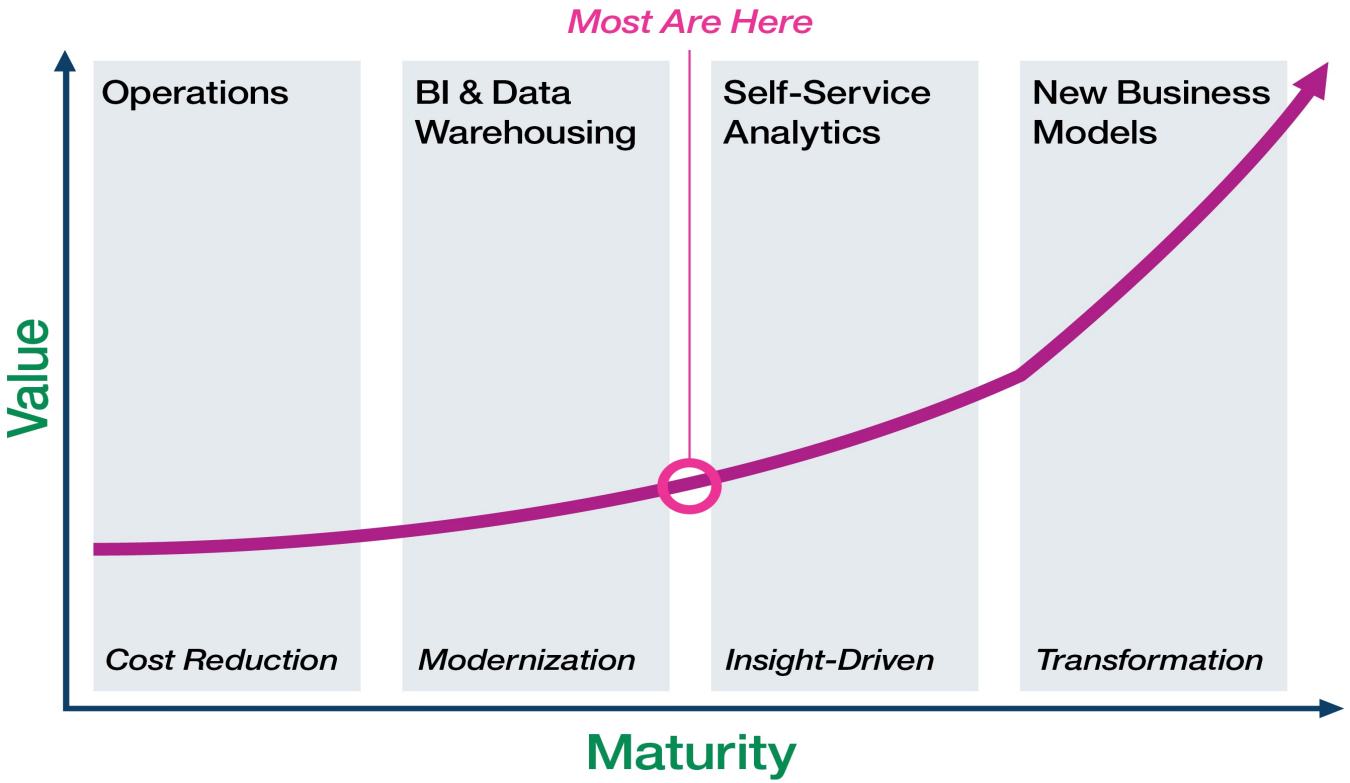
World's Largest Accommodations Company
Owns No Real Estate

World's Largest Taxi Company
Owns No Vehicles

World's Largest Retailer
Carries No Inventory

World's Largest Media Company
Creates No Content

Tools available till recently have been more focused on the initial phases of exploiting data



Data Science Life: Explore and Predict

1) Exploration: We don't have any special attribute we want to predict. Rather we want to understand the structure present in the data. Are there clusters? Non-obvious relationships?

- Also referred to as "unsupervised learning"
- E.g., K-means clustering

Use Cases -> Understanding categories of customers, cross-selling opportunities, etc…

2) Prediction: The data contains a particular attribute (called the target attribute) and we want to learn how the target attribute depends on the other attributes.

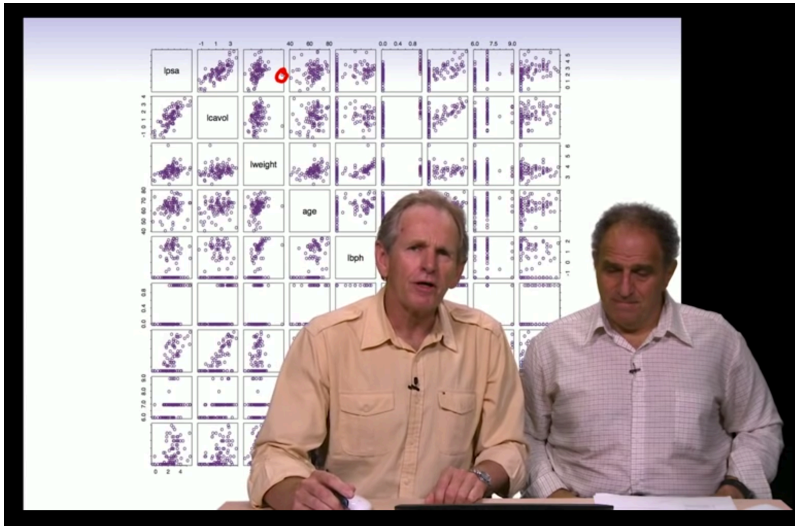
- Also referred to as "supervised learning"
- E.g., Support vector machines

Use Cases -> Building a model to predict customer churn, fraud, etc…

Visualizing data

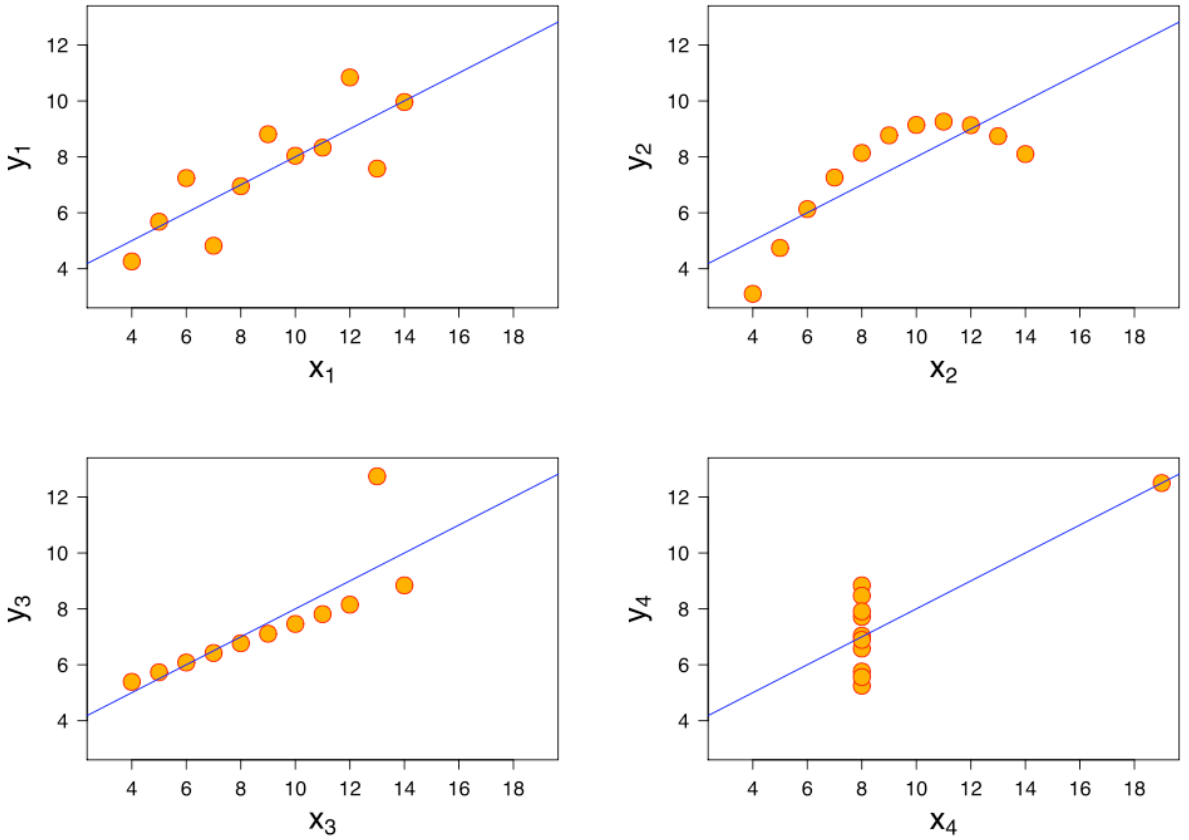
“The first thing to do when you get a set of data for analysis is not to run it through a fancy algorithm. Make some graphs, some plots. Look at the data..”

Robert Tibshirani FRSC
Professor in the Departments of Statistics and Health Research and Policy at Stanford University



Anscombe's Quartet

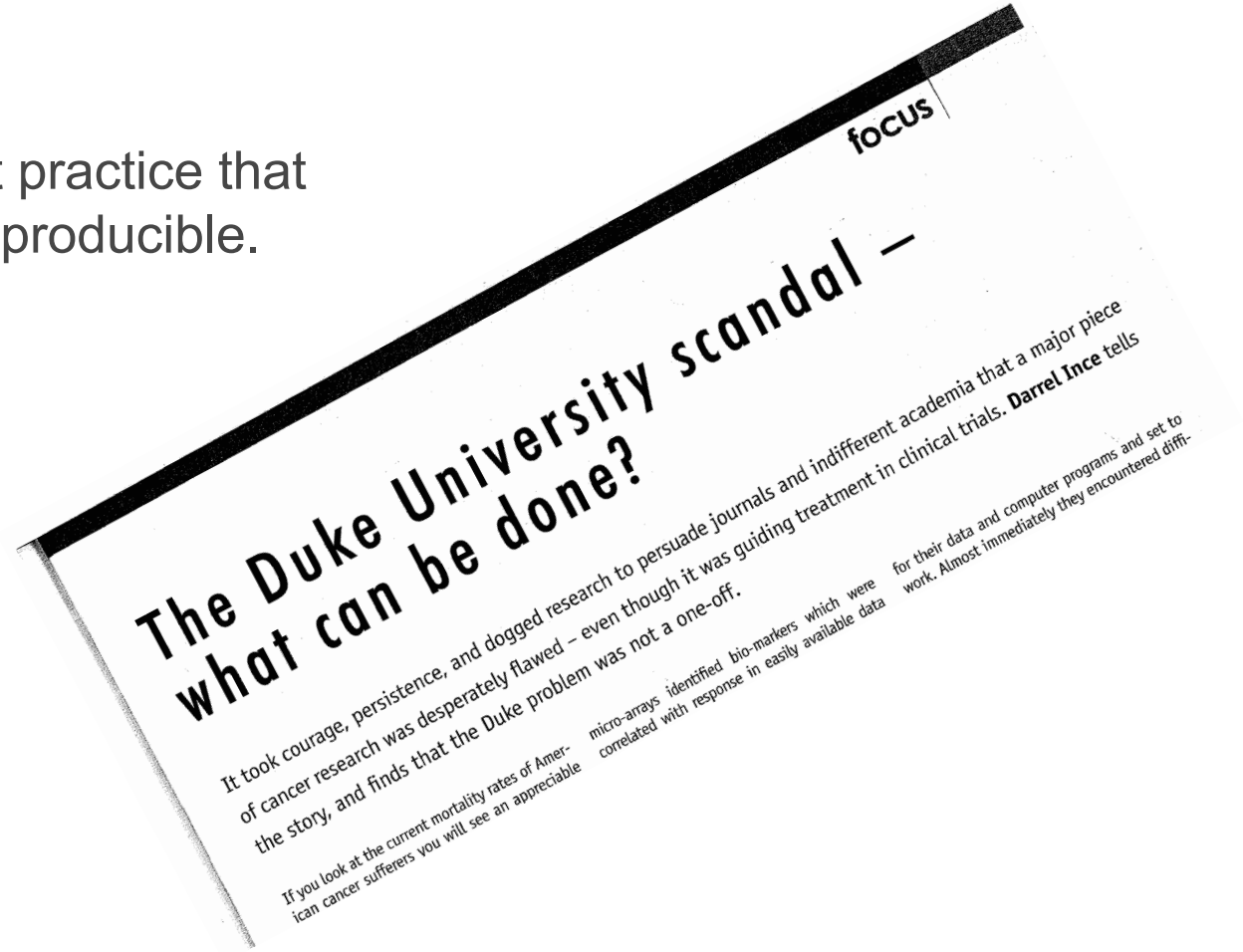
Four datasets with nearly identical statistical properties



Source: https://en.wikipedia.org/wiki/Anscombe's_quartet

Reproducible research

It is a recognised best practice that research should be reproducible.



focus

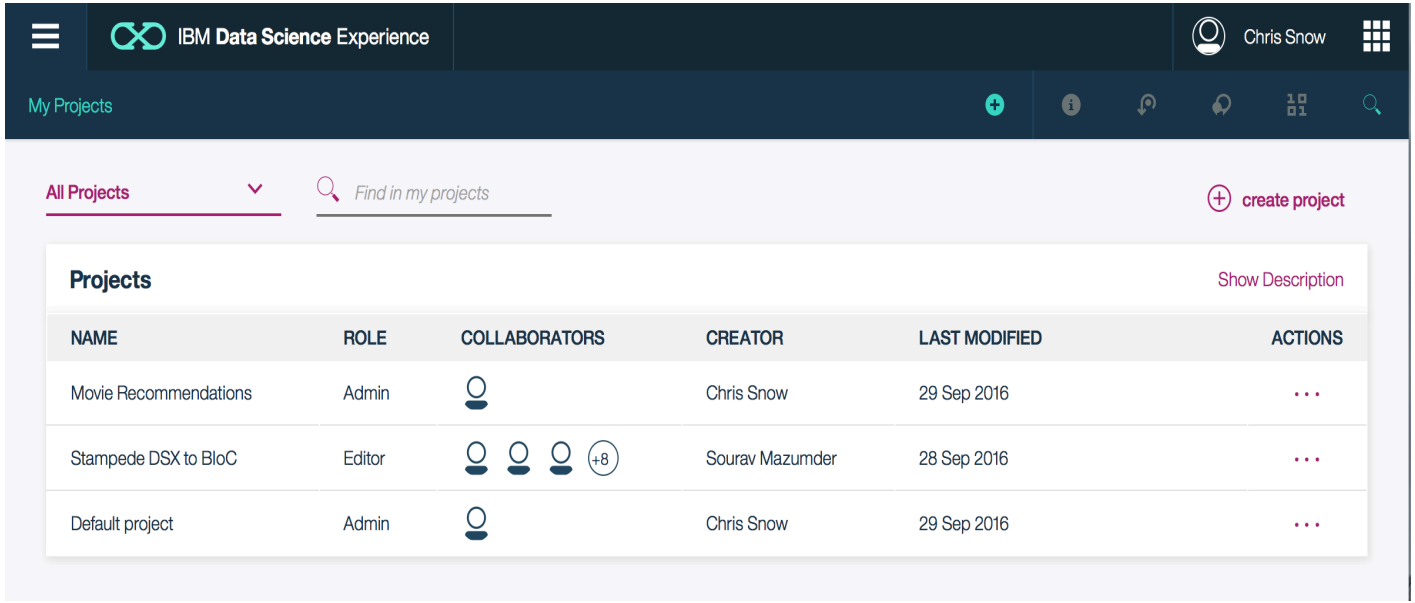
The Duke University scandal – what can be done?

It took courage, persistence, and dogged research to persuade journals and indifferent academia that a major piece of cancer research was desperately flawed – even though it was guiding treatment in clinical trials. **Darrel Ince** tells the story, and finds that the Duke problem was not a one-off.

If you look at the current mortality rates of American cancer sufferers you will see an appreciable micro-arrays identified bio-markers which were correlated with response in easily available data for their data and computer programs and set to work. Almost immediately they encountered diffi-

Essential collaboration

Work smarter using community, work faster with your team.



BigInsights provides options to suit business needs

Basic Plan

- IOP clusters within minutes
- Scale up or down based on need
- Separation of compute and storage
- Pay-as-you-go on an hourly basis

Target use cases:

- Applications development and testing
- Getting started with IOP

Basic Plan + value adds *

- IOP + BigInsights clusters within minutes
- Advanced analytics capabilities like Big SQL and Text Analytics
- Pay-as-you-go on an hourly basis
- Larger instances

Target use cases:

- Infusing advanced analytics capabilities at a low, entry price point

Enterprise Plan

- IOP + BigInsights clusters on bare metal nodes
- ISO / SOC2 / HIPAA compliance
- Dedicated clusters for performance & data privacy
- Available in multiple data centers
- Monthly subscription

Target use cases:

- Enterprise apps at production scale
- Advanced analytics use cases requiring dedicated resources
- Reserved capacity for continuous use

Enhanced capabilities, security and performance

**** Future offering***

Enter the Data Science Experience

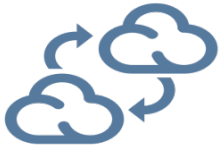
Currently in Beta

Learn

Built-in learning to get started or go the distance with advanced tutorials

Create

The best of open source and IBM value-add to create state-of-the-art data products



Collaborate

Community and social features that provide meaningful collaboration

Sign up for the beta here:
http://datascience.ibm.com

Demo

+

datascience.ibm.com

Demo

Movielens Data

The demo can be found here: https://github.com/snowch/demo_2710

a. Setup Data (SSH)

b. Import Data (WebHDFS)

f. Export Model (SCP)

g. Execute Model (SSH)

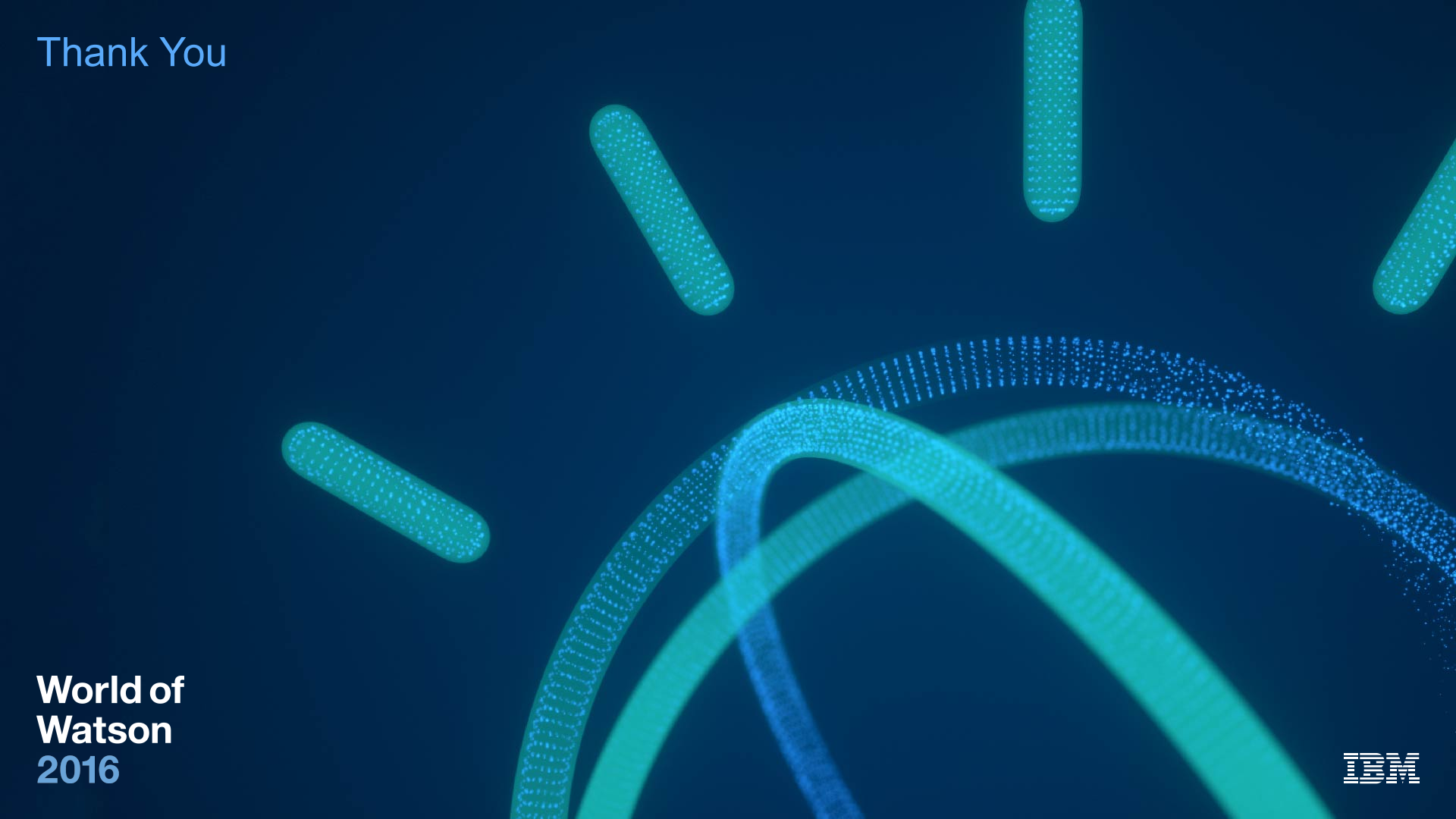
datascience.ibm.com

c. Explore

d. Train Model

e. Save Model

Thank You



World of
Watson
2016

IBM

Notices and disclaimers continued

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. IBM EXPRESSLY DISCLAIMS ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

Notices and disclaimers

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IN NO EVENT SHALL IBM BE LIABLE FOR ANY DAMAGE ARISING FROM THE USE OF THIS INFORMATION, INCLUDING BUT NOT LIMITED TO, LOSS OF DATA, BUSINESS INTERRUPTION, LOSS OF PROFIT OR LOSS OF OPPORTUNITY. IBM products and services are warranted according to the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law.